Diversité structurale des Glutathion Transférases fongiques des classes Oméga et Xi et identification de leurs ligands par des approches cristallographiques

Mathieu Schwartz

▶ **To cite this version:**

Mathieu Schwartz. Diversité structurale des Glutathion Transférases fongiques des classes Oméga et Xi et identification de leurs ligands par des approches cristallographiques. Cristallographie. Université de Lorraine, 2018. Français. NNT : 2018LORR0124 . tel-01894523

HAL Id: tel-01894523
https://theses.hal.science/tel-01894523

Submitted on 12 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Collégium Sciences et Technologies
Ecole Doctorale C2MP

UNIVERSITE DE LORRAINE

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Spécialité Chimie

Présentée par

Mathieu SCHWARTZ

Thèse dirigée par Claude DIDIERJEAN et Frédérique FAVIER

Diversité structurale des Glutathion Transférases fongiques des classes Oméga et Xi et identification de leurs ligands par des approches cristallographiques

Soutenance publique prévue le 25 septembre 2018

Rapporteurs :

Mme Mirjam CZJZEK	Directrice de recherche CNRS, Station Biologique de Roscoff, Roscoff (France)
M. Jean-Michel RONDEAU	Senior Investigator NOVARTIS, Novartis Institute for Biomedical Research, Bâle (Suisse)

Examinateurs :

M. Jean CAVARELLI	Professeur, IGBMC, Université de Strasbourg
Mme Catherine CORBIER	Professeur, ENSAIA, Université de Lorraine
M. Eric GELHAYE	Directeur de recherche INRA, Université de Lorraine
Mme Frédérique FAVIER	Maître de conférences, FST, Université de Lorraine
M. Claude DIDIERJEAN	Maître de conférences, FST, Université de Lorraine

Laboratoire de Cristallographie, Résonance Magnétique et Modélisations, UMR CNRS n° 7036
Faculté des Sciences et Technologies de Nancy – Université de Lorraine
BP 70239 - 54506 Vandœuvre-lès-Nancy Cedex

Diversité structurale des

Glutathion Transférases fongiques des classes Oméga et Xi

et identification de leurs ligands

par des approches cristallographiques

Table des matières

REMERCIEMENTS 9

ABRÉVIATIONS 11

INTRODUCTION 15

ETAT DE L'ART 21

I LES GLUTATHION TRANSFÉRASES, UNE FAMILLE D'ENZYMES UBIQUITAIRES ET VARIÉES CHEZ LES CHAMPIGNONS ... 23

I.1 Description de la famille des glutathion transférases ... 25

I.1.a.Le rôle des GST dans la détoxication des composés xénobiotiques ... 25

I.1.b.Les autres rôles biologiques connus des GST ... 27

I.1.c.Structure des principaux types de GST ... 33

I.2 Les GST chez les organismes fongiques ... 40

I.2.a.Les champignons dans le règne du vivant ... 40

I.2.b.Structure et fonction des GST fongiques ... 43

II L'IDENTIFICATION DE LIGANDS DE MACROMOLÉCULES PAR CRISTALLOGRAPHIE ... 54

II.1 Etapes préliminaires à l'étude structurale haut débit de complexes protéines-ligands ... 55

II.1.a.Mise au point de la construction protéique adaptée à une étude protéine-ligand ... 55

II.1.b.Mesure de l'affinité protéine-ligand et méthodes expérimentales de criblage ... 55

II.1.c.Stratégies d'obtention de cristaux de complexes et leur analyse ... 60

II.2 La cristallographie comme méthode centrale de criblage ... 63

II.2.a.Intérêt de la cristallographie comme méthode initiale de criblage ... 63

II.2.b.Méthodes pour adapter la cristallographie à des besoins « haut-débit » ... 64

II.2.c.La cristallographie d'affinité : identification de ligands potentiels parmi un mélange complexe tel qu'un extrait naturel ... 67

III OBJECTIFS DE CE TRAVAIL DE RECHERCHE ... 69

MATÉRIELS ET MÉTHODES 73

I INTRODUCTION ... 75

II CLONAGE, PRODUCTION ET PURIFICATION DES GST ... 76

III CRISTALLOGENÈSE DES GST ... 78

III.1 Techniques de cristallisation ... 79

III.2 Technique de micro-ensemencement ... 81

III.3 Interprétation du phénomène de cristallisation à l'aide du diagramme de phase ... 81

IV PRÉPARATION DES CRISTAUX DE COMPLEXES GST-LIGAND ... 83

IV.1 Technique de trempage classique ... 84

IV.2 Techniques de co-cristallisation ... 85

IV.3 Technique de « trempage à sec » ... 86
V CONGÉLATION DES CRISTAUX, TESTS DE DIFFRACTION ET COLLECTE DES DONNÉES DE DIFFRACTION ... 87
V.1 Cryoprotection ... 87
V.2 Tests de diffraction ... 88
V.3 Collecte des données de diffraction ... 89
VI TRAITEMENT DES DONNÉES ... 91
VII PHASAGE PAR REMPLACEMENT MOLÉCULAIRE ... 93
VII.1 Introduction au problème de la phase ... 93
VII.2 Coefficient de Matthews ... 94
VII.3 Principe du remplacement moléculaire ... 95
VIII CONSTRUCTION, AFFINEMENT ET VALIDATION DU MODÈLE ... 96
VIII.1 Construction avec COOT ... 96
VIII.2 Affinement par maximum de vraisemblance avec PHENIX ... 97
VIII.3 Validation du modèle avec MOLPROBITY et PDB-Validation server ... 98
IX BIO-INFORMATIQUE ... 99
IX.1 Obtention des séquences de GST et alignements ... 99
IX.2 Phylogénie avec MEGA ... 100
IX.3 Alignements de structure ... 100
IX.4 Docking avec GOLD ... 100

RÉSULTATS ET DISCUSSION 103
I ETUDE DES GST OMÉGA DE *TRAMETES VERSICOLOR* ... 105
I.1 Etude phylogénétique et biochimique des GSTO de *T. versicolor* ... 105
I.2 Etude structurale de neuf isoformes de TvGSTO ... 108
I.3 Etude structurale des complexes GST-ligands issus du criblage ... 118
I.3.a.Ligands de type « dérivés du glutathion » ... 119
I.3.b.Ligands aromatiques ... 122
I.4 Publications associées au projet de criblage de ligands des TvGSTO ... 130
I.4.a.Publication n°1 : Identification des hydroxybenzophénones et des flavonoïdes comme ligands des isoformes TvGSTO3S et TvGSTO6S ... 130
I.4.b.Publication n°2 : Identification de molécules issues d'extraits de bois exotiques comme ligands de l'isoforme TvGSTO2S ... 164
I.5 Comparaison structurale des sites de reconnaissance des polyphénols chez les TvGSTO ... 179
II ETUDE DES GST XI DE *SACCHAROMYCES CEREVISIAE* ET *TRAMETES VERSICOLOR* ... 185
II.1 L'isoforme ScECM4 de la levure *Saccharomyces cerevisiae* ... 185
II.1.a.Résultats de l'étude cristallographique et enzymatique de ScECM4 ... 187
II.1.b.Publication n°3 : Etude structurale de ScECM4 ... 190
II.1.c.Etude prédictive des isoformes ScGTO1 et ScGTO3 de *S. cerevisiae* ... 214
II.2 Les isoformes GSTX1 et GSTX3 de *Trametes versicolor* ... 217
II.2.a.Résultats de l'étude biochimique et structurale de TvGSTX1 et TvGSTX3 ... 217
II.2.b.Publication n°4 : Etude structurale et fonctionnelle des isoformes TvGSTX1 et TvGSTX3 ... 221

DISCUSSION GÉNÉRALE ET PERSPECTIVES 255

RÉFÉRENCES BIBLIOGRAPHIQUES 267

I OUVRAGES DE RÉFÉRENCE 269

II ARTICLES ET REVUES 269

ANNEXES 287

I STRUCTURES TOPOLOGIQUES DES SUBSTRATS ET LIGANDS UTILISÉS AU COURS DE CE PROJET 289

Remerciements

Je tiens à remercier Mirjam Czjzek, Jean-Michel Rondeau, Catherine Corbier, Jean Cavarelli et Eric Gelhaye d'avoir accepté de juger ce travail de thèse.

Mes remerciements les plus sincères reviennent à Claude Didierjean et à Frédérique Favier qui m'ont encadré depuis mon stage de M2 jusqu'à l'achèvement de ces trois années de thèse. Votre rigueur scientifique aussi bien que votre bonne humeur ont grandement contribué à la réussite de ce travail.

Je tiens à remercier le Professeur Dominik Schaniel pour m'avoir accueilli au laboratoire CRM2. Mes remerciements vont également à Sandrine Mathiot, Guillermo Mulliert et Thomas Roret pour leur contribution technique ainsi que nos divers échanges. Merci également à toute l'équipe BioMod - Théo, Benoît, Christian - et aux nombreux stagiaires que j'ai rencontrés lors de ces trois années.

Je remercie le Professeur Eric Gelhaye ainsi que tous nos collaborateurs du laboratoire Interactions Arbre-Microorganismes. Mes remerciements vont aussi au Professeur Stéphane Dumarçay du Laboratoire d'Etude et de Recherche sur le Matériau Bois. Je remercie particulièrement Thomas Perrot pour nos échanges scientifiques et notre travail de concert, pour toutes ses molécules aux noms difficilement prononçables qu'un jour il m'a fallu faire diffuser dans mes cristaux, mais aussi pour notre collaboration dans les projets en formation pédagogique.

Je remercie Gilles Labesse et Jean-Luc Ferrer pour m'avoir initié à la cristallographie *in situ* et au criblage de ligands au sein de leurs laboratoires. Merci également à Jean Cavarelli et l'équipe pédagogique du réseau ReNaFoBis pour l'initiation aux différentes techniques de la biologie structurale que j'ai suivie à Oléron. Je remercie David Cobessi pour son aide précieuse lors des campagnes synchrotron à l'ESRF de Grenoble.

Mes remerciements vont à Clarisse Perrin et toute l'équipe pédagogique du département Génie Biologique Agro-Alimentaire de l'IUT Nancy-Brabois, pour m'avoir

accueilli au département, initié à l'enseignement universitaire, et pour la confiance qu'ils m'ont accordée dans le cadre de ces trois années d'enseignements en tant que DCCE.

Une petite dédicace à mes amis musiciens - Chris, Gio et Antoine – ainsi qu'à Alex et Jordhan. Merci à toute ma famille et en particulier mes parents, qui m'ont toujours soutenu durant toutes ces études, et me soutiennent encore. Sans vous ce travail n'aurait jamais vu le jour.

Enfin, ces dernières lignes reviennent à ma femme, Cassandra, pour sa patience et son amour inconditionnel. Je t'aime ma chérie, ce travail t'est dédié.

Abréviations

ANS	1-Anilino-8-Naphthalène Sulfonate
CDNB	1-Chloro 2,4-DiNitroBenzène
Cys-GST	Glutathion Transférase à Cystéine catalytique
DHAR	DéHydroAscorbate Réductase
DLS	Diffusion Dynamique de la Lumière
DMSO	DiMéthylSulfOxyde
DNDGIC	Complexe de Fer Dinitrosylé Diglutathionylé
DNGIC	Complexe de Fer Dinitrosylé Glutathionylé
DRX	Diffraction des Rayons X
EDTA	ÉthylèneDiamineTétraAcétique
EF1Bγ	sous-unité γ du Facteur d'Elongation 1B
ESRF	Installation Européenne de Rayonnement Synchrotron
ETP	EpipolyThiodioxoPiperazine
FAA	4-Fumaryl-Acétoacétate
FBDD	Découverte de Médicaments Basée sur la recherche de Fragments
GGE	Guaiacylglycérol-β-Guaiacyl Ether
GHR	Glutathionyl-Hydroquinone Réductase
GS-DNB	Glutathionyl-DiNitroBenzène
GSH	Glutathion
GS-hexane	Glutathionyl-hexane
GS-MEN	Glutathionyl-Ménadione
GS-PAP	Glutathionyl-PhenylAcétoPhénone
GS-PEITC	Glutathionyl-PhenEthyl IsoThioCyanate
GSSG	Dimère de glutathion oxydé
GST	Glutathion Transférase
GSTO	Glutathion Transférase Oméga
GSTOC	Glutathion Transférase Oméga à Cystéine catalytique
GSTOS	Glutathion Transférase Oméga à Sérine catalytique
GSTU	Glutathion Transférase Tau
GSTX	Glutathion Transférase Xi
GSTZ	Glutathion Transférase Zeta
HED	Hydroxy Ethyl Disulfure
HEPES	acide 4-(2-HydroxyEthyl)-1- Pipérazine Ethane Sulfonique
IPTG	IsoPropylThioGalactoside
JGI	Centre de séquençage américain 'Joint Genome Institute'
MAA	4-maleyl-acétoacétate
MAPEG	Protéines Associées aux Membranes et impliquées dans le métabolisme des Eicosanoïdes et du Glutathion
MES	acide 2-(N-Morpholino) Ethane Sulfonique
MMT	Tampon constitué de DL-Acide Malique, MES et TRIS
mPGES1	ProstaGlandine E2 Synthase 1 microsomale
MS	Spectrométrie de Masse
Ni-NTA	Ni^{2+}- nitrilotriacétate
NO	Oxyde Nitrique
PCR	Réaction de Polymérisation en Chaine
PDB	Base de Données de Structures expérimentales de Protéines

PEITC	PhenEthylIsoThioCyanate
RMN	Résonance Magnétique Nucléaire
RX	Rayons X
Ser-GST	Glutathion Transférase à Sérine catalytique
Site G	Site de reconnaissance du glutathion chez les GST
Site H	Site de reconnaissance du substrat hydrophobe chez les GST
Site L	Site de fixation "ligandine" chez les GST
TRIS	2-Amino-2-(hydroxyméthyl)propane-1,3-diol
TSA	Thermal Shift Analysis (variation de la stabilité thermique)
TvGSTO	Glutathion Transférase Oméga de *Trametes versicolor*
Tyr-GST	Glutathion Transférase à Tyrosine catalytique

Introduction

La cristallographie est la science qui s'intéresse à l'étude des cristaux et à leur structure chimique. Les cristaux, solides périodiques dont le motif de base est répété dans les trois dimensions, fascinent les hommes depuis des siècles. Bien souvent, on pense que ces objets sont le fruit de la nature, comme c'est le cas pour de nombreux minéraux, les cristaux de sel de mer (chlorure de sodium) ou encore le diamant (cristal de carbone dont les atomes ont une organisation particulière). Les plus spectaculaires sont probablement les cristaux géants de gypse des grottes mexicaines de Naica (Figure 1). Cependant, les cristaux peuvent aussi être générés en laboratoire, et de nombreuses molécules chimiques et biologiques sont cristallisables.

Figure 1 **Cristaux de gypse ($CaSO_4$) géants de la mine de Naica**
Région de Chihuahua, Mexique ('The Cave of Giant Crystal', par Paul Williams, image libre sous licence CC BY-NC 2.0).

L'intérêt des scientifiques pour les cristaux remonte à l'Antiquité, période au cours de laquelle les philosophes grecs portent en admiration la beauté de leurs formes géométriques. Les solides cristallins suscitent également l'intérêt des mathématiciens et des architectes au fil du temps, puis des minéralogistes à partir de la Renaissance. Plus tard, les avancées en physique avec notamment la découverte des rayons X par Wilhelm Röntgen (1895) ainsi que

la diffraction des rayons X par un cristal par Max von Laue (1912) posent les bases d'une discipline qui permet d'étudier la forme tridimensionnelle des molécules.

Les premières contributions de la cristallographie à la compréhension des mécanismes moléculaires biologiques ont été la détermination de la structure de l'ADN, support de l'information génétique, en 1959 (Watson et Crick) et la détermination de la structure de protéines globulaires : la myoglobine et l'hémoglobine par John Kendrew et Max Perutz (1962). Citons également le travail de Dorothy Hodgkin qui a déterminé la structure de l'insuline en 1969. Dès lors, la cristallographie macromoléculaire a connu un essor considérable qui a permis de comprendre le fonctionnement de nombreuses protéines et enzymes, avec d'énormes retombées dans les domaines pharmaceutiques et médicaux.

Les glutathion transférases (GST) sont des enzymes constituant une superfamille déclinée en plusieurs classes et qui sont retrouvées chez la plupart des organismes vivants. Ces enzymes participent à la détoxification des composés néfastes endogènes ou issus de l'environnement en catalysant le transfert du glutathion (GSH) sur les molécules à éliminer. Depuis leur découverte au début des années 60, de nombreux travaux ont élucidé le rôle de ces protéines dans le métabolisme des mammifères principalement. Néanmoins, ces enzymes sont ubiquitaires et les travaux de séquençage génomique débutés depuis les années 2000 révèlent que les GST sont présentes en très grand nombre chez certains organismes fongiques. Parmi ceux-ci sont retrouvés des champignons décomposeurs de bois présentant un intérêt écologique et biotechnologique de par leurs systèmes enzymatiques leur permettant de dégrader la matière organique. Une classe de GST en particulier est exprimée en de nombreuses isoformes chez les champignons basidiomycètes tel que l'agent de la pourriture blanche *Trametes versicolor* : la classe des GST Oméga. Une autre classe présente chez les champignons et largement méconnue est la classe Xi. Le rôle précis de ces deux classes de GST n'est encore pas clairement défini. L'objet de ce manuscrit est de présenter une étude cristallographique détaillée des GST Oméga et Xi d'organismes fongiques.

Après un état de l'art actuel sur les GST fongiques et les méthodes permettant de déterminer la structure de complexes protéines-ligands par cristallographie, ce manuscrit présente l'étude cristallographique de GST des classes Oméga et Xi chez le basidiomycète *Trametes versicolor*, organisme modèle pour l'étude de la dégradation du bois, ainsi qu'une structure de GST Xi chez l'organisme modèle eucaryote *Saccharomyces cerevisiae*. Ce travail a également impliqué la détermination structurale de complexes GST-ligands afin d'éclaircir

les bases d'une éventuelle spécificité vis-à-vis de certaines molécules. Les résultats obtenus ont donné lieu à quatre publications acceptées et parues dont les textes sont joints à ce manuscrit.

Etat de l’art

I Les glutathion transférases, une famille d'enzymes ubiquitaires et variées chez les champignons

Les glutathion transférases, appelées historiquement glutathion-S-transférases et abrégées 'GST', forment une large famille d'enzymes présentes chez la plupart des organismes. La première GST a été mise en évidence par la découverte d'une activité de transfert de glutathion dans le foie de rat (Booth et al, 1961). Dès lors de nombreuses GST ont été caractérisées biochimiquement mais aussi structuralement avec l'avancée des techniques de détermination structurale en cristallographie des protéines. Au moment de la rédaction de ce manuscrit, 582 structures de GST étaient disponibles dans la base de données de structures de protéines (Protein Data Bank). La majorité des GST cytosoliques est constituée de protéines dimériques. Il leur correspond finalement des rôles beaucoup plus variés que la seule fonction de transfert du glutathion dont elles tirent leur nom.

La famille des GST est très hétéroclite. Actuellement, près d'une vingtaine de classes de GST est décrite dans la littérature. Ces classes ont été nommées en majorité à partir d'une lettre grecque (Mannervik et Danielson, 1988). Quelques fois certaines classes sont spécifiques à un règne biologique (ex : les GST Beta sont retrouvées chez les bactéries uniquement) alors que d'autres classes sont conservées chez plusieurs clades biologiques (ex : les GST Xi sont retrouvées chez les bactéries, les plantes et les champignons). L'acronyme d'une glutathion transférase d'une classe donnée commence par les initiales de l'organisme, est suivi des trois lettres GST et d'une lettre pour la classe. Le numéro de l'isoforme dans l'organisme termine souvent l'acronyme (par exemple, l'acronyme de l'isoforme 1 de la GST Xi de *Trametes versicolor* est TvGSTX1). A l'origine, ces classes ont été définies à partir des propriétés physico-chimiques et immunologiques des GST correspondantes. La séquence primaire est le critère le plus utilisé aujourd'hui. Ainsi, lorsque l'identité de séquence de deux GST est supérieure à 60 %, elles appartiennent à la même classe (Sheehan et al, 2001). Si leur identité de séquence est inférieure à 30 %, elles appartiennent à des classes différentes. Ce mode de classification, initialement basé sur les GST de mammifères, présente aujourd'hui des limites face aux nombreuses classes de GST existantes (Figure 2), et la présence de certaines classes dans des organismes évolutivement très différents (Mashiyama et al, 2014).

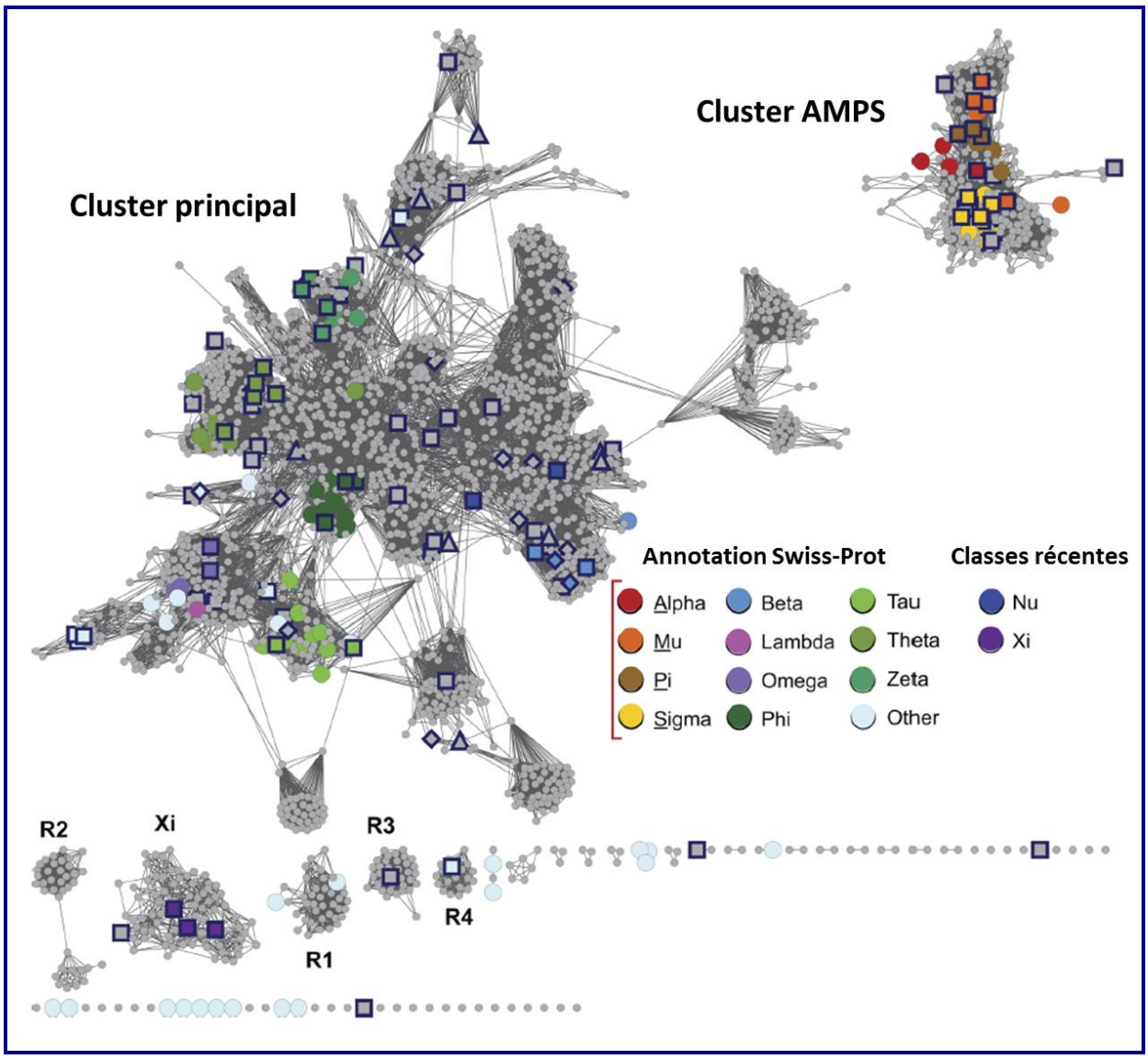


Figure 2 **Vue globale des différentes classes de GST cytosoliques existantes et de leurs relations sur la base des séquences (adapté et traduit de Mashiyama et al, 2014)**

Les 2 190 nœuds des différents clusters correspondent à 13 493 séquences putatives de GST. Chaque nœud regroupe des séquences partageant plus de 50% d'identité. La longueur de connexion entre deux nœuds représente leur degré de similarité. Si cette similarité est inférieure à un *E* de 10^{-13} (voir définition de ce score en partie IX.1 du chapitre Matériel et Méthodes), la connexion n'est pas représentée. Différents clusters sont retrouvés, notamment le cluster principal regroupant la majorité des classes, le cluster AMPS (regroupant les classes Alpha, Mu, Pi et Sigma), le cluster des GST Xi et d'autres clusters appelés R1 à R4. La classification est basée sur la définition des classes selon la banque de donnée SwissProt (https://www.uniprot.org/) ou selon la littérature pour les classes Nu et Xi. Les nœuds sont colorés si et seulement si 50% au moins des séquences présentes ont été assignées aux classes correspondantes selon la SwissProt. Les symboles (triangle, carré ou losange) avec des lignes épaisses correspondent aux séquences dont les informations de structure sont disponibles.

Dans un premier temps l'état des connaissances principales acquises dans le domaine des GST est réalisé, sans toutefois être exhaustif compte tenu de la grande quantité d'informations acquises sur cette famille d'enzymes depuis les années 60. Un accent particulier est donné sur la structure des GST ainsi que leurs propriétés de fixation et de catalyse. Dans une deuxième partie, j'ai choisi de focaliser mon discours sur les GST fongiques en les décrivant du point de vue de leurs fonctions et de leurs structures.

I.1 Description de la famille des glutathion transférases

I.1.a. Le rôle des GST dans la détoxication des composés xénobiotiques

i) Les étapes de la détoxication

De par leur activité de transfert de glutathion et leur large spécificité de substrat, le rôle initialement attribué aux GST est la détoxication des xénobiotiques (Sheehan et al, 2001). Les organismes biologiques sont exposés au cours de leur vie à tout un panel de molécules exogènes pouvant être toxiques voire oncogènes. La capacité à pouvoir neutraliser ces composés est donc une adaptation biologique fondamentale à la survie. Pour cela, les organismes vivants utilisent une stratégie de biotransformation (Figure 3) afin de modifier la molécule pour réduire sa toxicité et améliorer sa solubilité dans un premier temps (phases I et II), afin de l'exporter hors de la cellule dans un second temps (phase III).

La phase I implique une modification chimique (le plus souvent une oxydation) de la molécule cible qui a pour but d'exposer un groupement fonctionnel. Cette phase est assurée essentiellement par une famille de protéines microsomales : le système cytochrome P450. La phase II fait intervenir des enzymes dont le rôle est de faciliter la solubilisation en milieu aqueux du xénobiotique. Pour cela, les réactions mises en jeu impliquent l'association à un substrat comme la glycine, l'acide UDP-glucuronique ou, dans la majorité des cas, le glutathion. Ainsi les principales enzymes de phase II sont les glutathion transférases qui peuvent catalyser des réactions de substitutions nucléophiles, d'additions ou encore d'ouvertures de cycle époxyde par le transfert d'une molécule de glutathion (GSH) sur le xénobiotique. Durant la phase III, le conjugué GS-X formé peut être pris en charge par des systèmes membranaires d'export tels le MRP (pour 'multidrug-resistance-associated protein'), qui ont pour but d'exporter le conjugué hors de la cellule (Hayes et al, 2005).

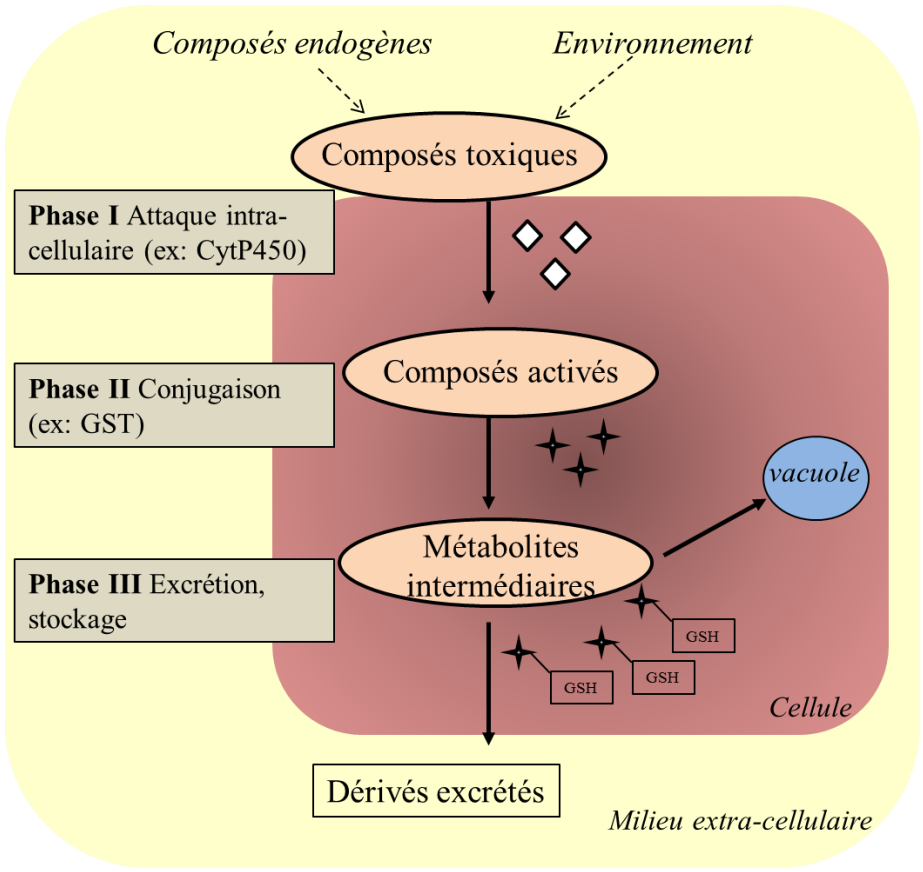


Figure 3 **Etapes clés du métabolisme de la détoxication (adapté de Morel et al, 2013)**

Le métabolisme de la détoxication peut être décomposé en trois étapes. La première étape vise à faire apparaître un groupement fonctionnel sur le composé. Ce dernier est ensuite pris en charge par des enzymes de conjugaison afin d'améliorer sa solubilité et de diminuer sa toxicité. Enfin, le composé est soit excrété de la cellule, soit stocké dans la vacuole.

ii) Le rôle des GST dans la phase II

La réaction enzymatique de conjugaison au glutathion fait intervenir soit un résidu sérinyl- (Ser-GST), soit un résidu tyrosinyl- (Tyr-GST) de la GST qui est responsable de l'activation de la cystéine du glutathion par la diminution du pKa de sa fonction thiol (Oakley et al, 2011 ; voir l'ouvrage « Enzymatic Basis of Detoxication » de Jakoby et Habig). Le glutathion ainsi activé peut réaliser une attaque nucléophile sur un substrat portant un groupement électrophile et il se forme un conjugué glutathionylé portant une liaison C-S, N-S ou S-S suivant la nature du substrat (Figure 4, Hayes et al, 2005). Les GST étant pour la plupart d'entre elles homodimériques, les réactions peuvent avoir lieu simultanément dans les deux monomères.

Plusieurs substrats synthétiques dits « de laboratoire » ont été utilisés pour caractériser cette activité de conjugaison des GST. Le substrat le plus utilisé est le 1-chloro 2,4-dinitrobenzène (CDNB, Figure 4). Lors de la réaction catalytique, le soufre nucléophile attaque le carbone électrophile portant le groupement chloro- par un mécanisme de substitution nucléophile aromatique. Le produit de réaction est le conjugué glutathionyl-dinitrobenzène dans lequel le soufre est en liaison covalente S-C avec le groupement dinitrobenzène. La majorité des Ser-GST et Tyr-GST présentent une activité enzymatique avec ce produit, détectable par spectrophotométrie UV/visible.

Une autre famille de substrats largement utilisés est la famille des isothiocyanates. Ces molécules volatiles sont retrouvées chez certains végétaux de la famille des Brassicacées (exemples : chou, raifort). Le glutathion peut réagir spontanément sur ces composés et former des conjugués de type dithiocarbamate. Il a été montré que certaines GST sont capables de catalyser cette réaction dans le sens de la conjugaison mais également dans le sens inverse (Kolm et al, 1995).

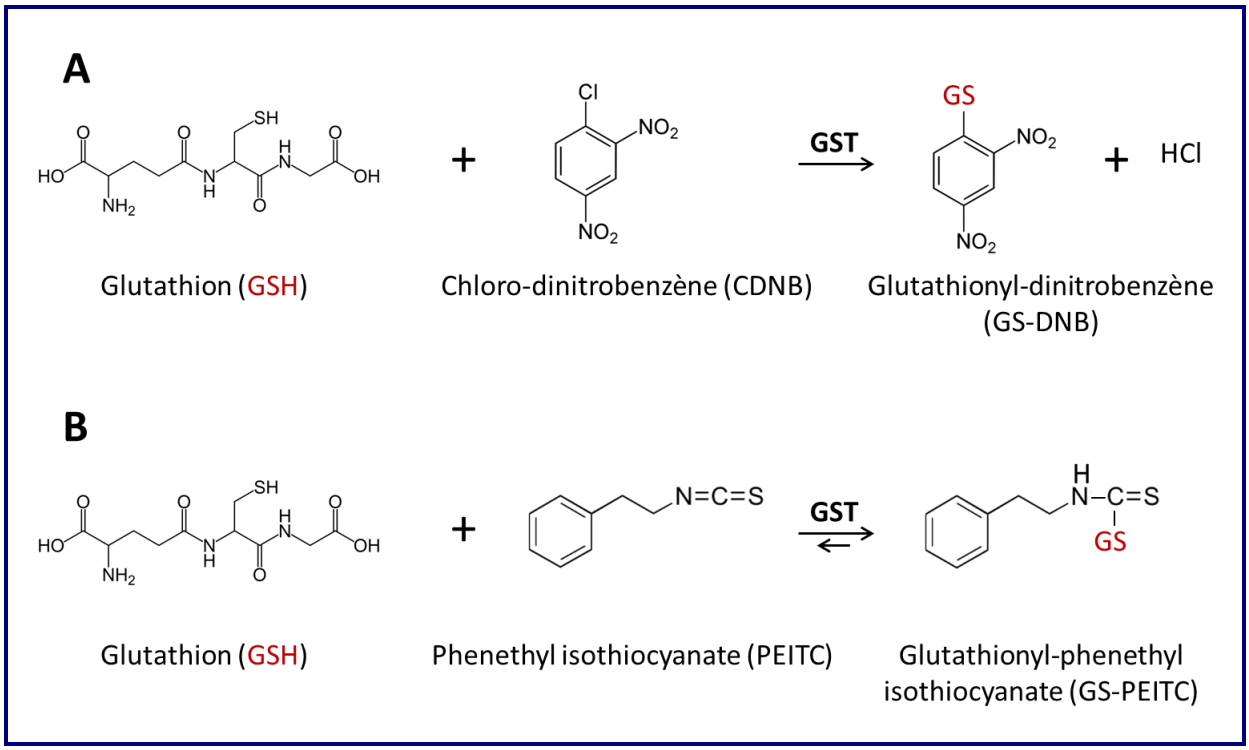


Figure 4 **Réactions de conjugaison catalysées par les GST**
A. Conjugaison enzymatique du glutathion au CDNB. B. Conjugaison enzymatique et réversible du glutathion au PEITC.

I.1.b. Les autres rôles biologiques connus des GST

En marge de leur rôle dans la détoxication, les GST présentent également d'autres fonctions biologiques, parfois selon la classe à laquelle elles appartiennent. Ainsi, par

l'utilisation du glutathion en tant que cofacteur, elles peuvent remplir des fonctions dans la protection contre un stress oxydant, le catabolisme de la tyrosine, la biosynthèse des stéroides ou encore la régénération du déhydroascorbate.

i) Rôle de protection dans le cadre d'un stress oxydant

Un stress oxydant survient lorsqu'une accumulation d'espèces réactives de l'oxygène (anion superoxyde $O_2^{\cdot -}$, peroxyde d'hydrogène H_2O_2 ou radical hydroxyle HO^-) intervient à l'intérieur des cellules des organismes aérobies. Ces espèces peuvent causer des dommages aux molécules biologiques (lipides, protéines et acides nucléiques) et sont dans la majorité des cas neutralisées par des systèmes de défense enzymatiques (superoxyde dismutase, catalase ou glutathion peroxidase) ou non enzymatique (alpha-tocophérol, acide ascorbique, glutathion). En particulier, la peroxydation des lipides membranaires est problématique car elle provoque une série de réactions en chaîne qui peuvent avoir un effet dramatique sur les membranes. Certaines GST ont la capacité à conjuguer le glutathion à certains intermédiaires peroxydés formés et ainsi interrompre les dommages causés (Hayes et al, 2005). L'activité de telles GST peut notamment être évaluée par le test de réduction du cumène hydroperoxyde.

ii) Rôle dans la dégradation des acides aminés aromatiques

Une classe particulière de GST, la classe Zeta (GSTZ) est retrouvée chez les animaux (Tong et al, 1998 ; Fernandez-Canon et al, 2002), les plantes (Dixon et al, 2000) et certains champignons (Fernandez-Canon et al, 1998 ; Edwards et al, 2011) et impliquée dans le catabolisme des acides aminés aromatiques. En effet, ces GST aussi appelées maleylacétoacétate isomérases (MAAI) catalysent la réaction d'isomérisation du 4-maleyl-acétoacétate (MAA) en 4-fumaryl-acétoacétate (FAA) dans l'avant-dernière étape de la voie de dégradation de la tyrosine (Polekhina et al, 2001). La réaction fait intervenir un intermédiaire glutathionylé qui permet l'isomérisation de la liaison *cis* du MAA en *trans* pour produire le FAA. Il est important de rappeler que des défauts dans la voie de dégradation des acides aminés aromatiques sont à l'origine de maladies telles que la phenylcétonurie ou la tyrosinémie héréditaire de type I pouvant conduire à un hépatocarcinome (Fernandez-Canon et al, 1998).

iii) Rôle dans le métabolisme des hormones stéroides

Chez l'homme, la GST Alpha A3-3 est impliquée dans l'isomérisation de liaisons doubles sur des hormones stéroïdes dans les voies de biosynthèse de ces composés (Johansson

et Mannervik, 2001). En effet, cette GST est capable d'isomériser la liaison Δ5 en liaison Δ4 sur les 3-cétostéroides afin de produire *in fine* les hormones stéroides testostérone ou progestérone. De manière intéressante, seule l'isoforme A3-3 présente cette activité de manière significative, contrairement aux autres isoformes A1-1, A2-2 et A4-4. Cette différence d'activité serait expliquée par des différences retrouvées au niveau du site de liaison du substrat hydrophobe (site H) des GST (Johansson et Mannervik, 2001).

iv) Rôle dans les processus d'inflammation

Des GST microsomales, rassemblées dans la famille des MAPEG (pour Membrane-Associated Proteins in Eicosanoids and Glutathione metabolism) et évolutivement différentes des GST cytosoliques sont décrites dans tous les règnes du vivant sauf les Archées (Morgenstern et al, 2011). Chez l'homme, 6 membres de cette famille sont décrits, présentant différents rôles biologiques. 3 membres sont impliqués dans la production de médiateurs chimiques intervenant dans les processus d'inflammation : la protéine activant la 5-lipoxygénase, les leukotriène C4-synthase et prostaglandine E2 synthase 1 microsomale (mPGES1). Deux autres membres (MGST2 et -3) présentent des activités glutathion peroxydases. MGST1 présente des activités de conjugaison du GSH sur les xénobiotiques similaires aux GST cytosoliques, quoique plus faibles cependant. Elle a la particularité d'avoir une préférence d'activité envers les substrats très hydrophobes tels que les hydrocarbures halogénés (Dekant et al, 1990) ou les phospholipides hydroperoxydés (Mosialou et al, 1993), renvoyant à une fonction de biotransformation d'espèces réactives lipophiles présentes aux membranes.

v) Rôle des GST végétales dans le transport des flavonoïdes

Il a été démontré que chez certaines plantes, des GST sont impliquées dans le métabolisme des produits naturels endogènes. En effet, des études chez le pétunia montrent qu'une GST appartenant à la classe Phi et appelée AN9 est requise pour l'export des anthocyanines de leur site de synthèse au cytoplasme jusqu'à la vacuole qui est leur site de stockage permanent (Mueller et al, 2000). Le même rôle est assuré par la GST Tau Bz2 chez le maïs. Des résultats similaires ont été obtenus à partir d'une étude sur l'homologue de AN9 chez la vigne (Gomez et al, 2011). Ce transport se fait sans réaction catalytique avec le glutathion et les affinités de liaison entre GST et ligand sont de l'ordre du micromolaire. Dixon et Edwards (2010) ont suggéré un rôle similaire pour les GST Lambda en montrant que l'isoforme de blé TaGSTL1 a la capacité de fixer certains flavonols.

vi) Rôle dans l'homéostasie de l'oxyde nitrique (NO)

Les GST interviennent dans le stockage de l'oxyde nitrique (NO), toxique à forte concentration. En effet, les molécules de NO peuvent s'accumuler dans la cellule en cas de stress oxydant et il est fondamental de contrôler leur concentration cellulaire. En 2001, Lo Bello et collaborateurs ont montré que la GST Pi humaine est impliquée dans l'homéostasie de l'oxyde nitrique. Celui-ci peut réagir non enzymatiquement avec le glutathion et le fer intracellulaire pour former un complexe appelée DNDGIC (pour 'dinitrosyl diglutathionyl iron complex'). La GST Pi est capable de lier ce composé en se « substituant » à un des groupements glutathion (le produit est appelé DNGIC pour 'dinitrosyl glutathionyl iron complex'). La Tyr7 vient stabiliser l'atome de fer à la place du glutathion (Figure 5). Plus tard, il a été démontré que la capacité à lier le DNGIC est une propriété conservée chez les Tyr-GST essentiellement, et dans une moindre mesure chez les Ser-GST (Bocedi et al, 2013). De manière intéressante, une coopérativité négative intervient de façon à maintenir l'activité de conjugaison au GSH à l'un des monomères lorsque l'autre monomère lie le DNGIC (Bocedi et al, 2016). Lorsqu'une molécule de DNGIC est fixée au premier monomère de GST Pi, l'affinité pour ce ligand diminue au deuxième monomère, de manière à conserver l'activité GSH-transférase nécessaire à la détoxication. Cette propriété permettrait à la GST Pi d'assurer son double rôle de manière optimale dans la cellule.

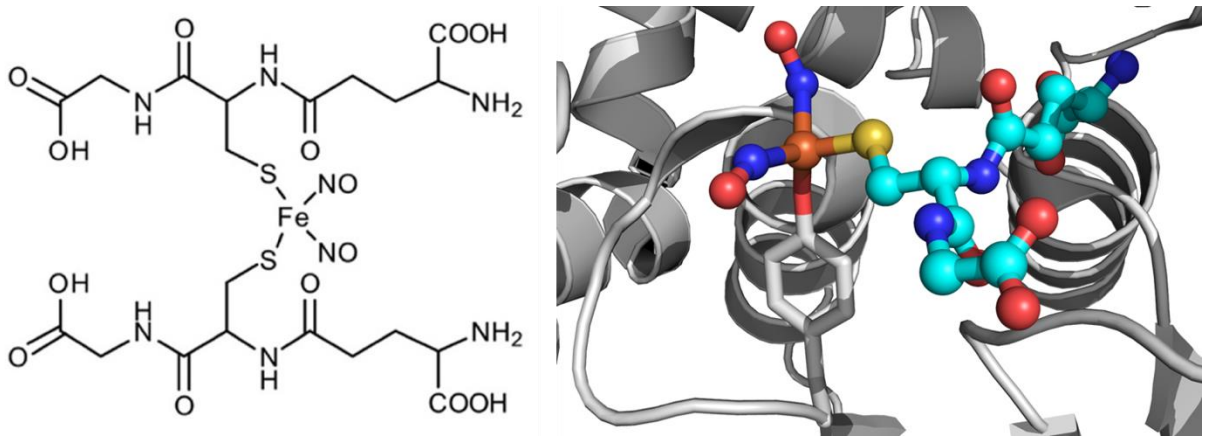


Figure 5 **« Dinitrosyl-diglutathionyl iron complex » (DNDGIC, gauche) et « dinitrosyl-glutathionyl iron complex » (DNGIC, droite) fixé au site actif de la GST Pi humaine (code PDB 1ZGN)**
La chaîne latérale de Tyr7 participe à la sphère de coordination de l'atome de fer et se substitue à l'un des glutathion lors de la fixation à la protéine.

vii) Les GST impliquées dans la dégradation de la lignine chez les bactéries du genre Sphingobium

Récemment, le rôle de GST dans le métabolisme des composés issus de la dégradation du bois a été suggéré chez certaines bactéries du sol du genre *Sphingobium*. En effet, en 2003 il a été montré que les GST dénommées LigE, LigF et LigG seraient impliquées dans la dégradation de la lignine chez les bactéries de genre *Sphingobium.* LigE et LigF catalysent une réaction beta-aryl étherase sur un substrat modèle de la lignine par conjugaison avec le glutathion (Figure 6). Le conjugué au glutathion est réduit ensuite par LigG (Masai et al, 2003). LigG est une GST à cystéine apparentée à la classe des GST Oméga (Meux et al, 2012), classe qui sera détaillée dans la partie I.2.b. Récemment, il a été montré un rôle similaire chez la bactérie *Novosphingobium aromaticivorans* pour une GST appelée NaGSTNu (Kontur et al, 2018), appartenant à la classe Nu (apparentée à la classe Ure2p chez les champignons, également détaillée dans la partie I.2.b).

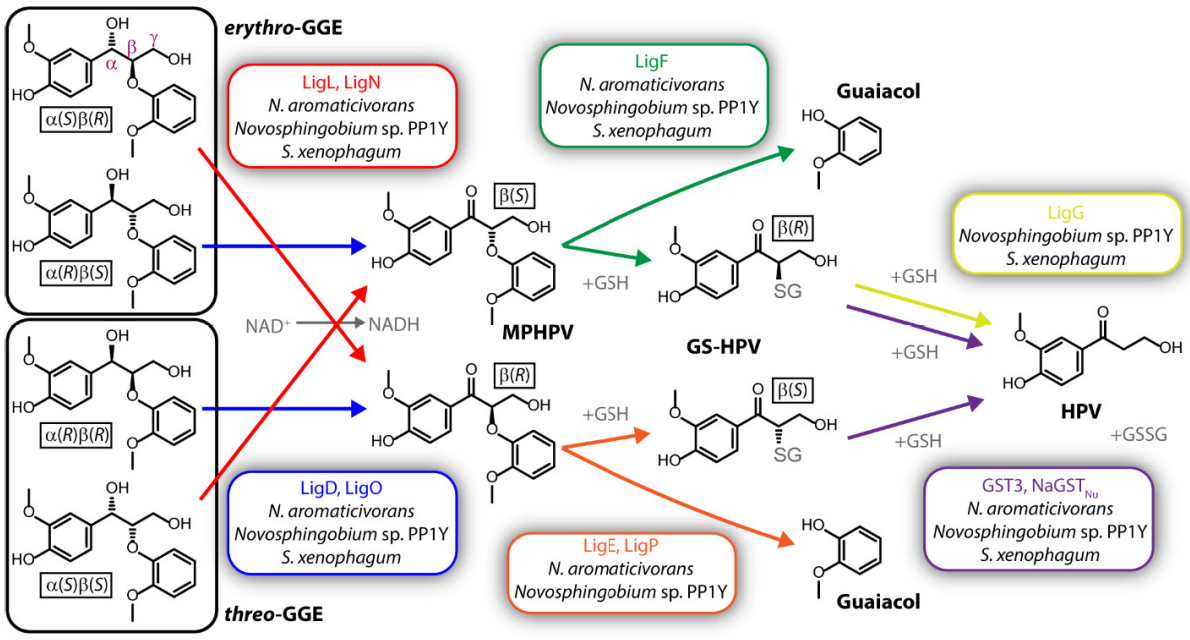


Figure 6 **Implication des GST de l'opéron Lig dans le clivage de la liaison β-aryl ether du modèle monomérique de lignine (guaiacylglycérol-β-guaiacyl ether, GGE) chez les bactéries du genre Sphingobium (d'après Kontur et al, 2018)**

viii) Les GST à cystéine et les rôles associés

Les GST présentent le plus souvent un résidu catalytique sérine ou tyrosine nécessaire à l'activation du glutathion. Néanmoins, certaines classes présentent un résidu cystéine comme résidu catalytique et de ce fait ont une chimie totalement différente. Ces GST à cystéine (Cys-GST) catalysent des réactions de type thiol-transférase avec l'hydroxy-

éthyldisulfide (HED) suivant un mécanisme similaire aux glutarédoxines (Board et al, 2000 ; Lallement et al, 2014). Les Cys-GST catalysent également la régénération du déhydroascorbate en ascorbate avec un rôle avéré dans la protection contre le stress oxydant pour la classe des déhydroascorbate réductases (DHAR) notamment. L'isoforme GSTO1 humaine est également responsable de la régulation par déglutathionylation de certaines protéines et serait impliquée dans la réaction inflammatoire (Board et Menon, 2016). De manière générale, ces GST sont capables de réduire les substrats glutathionylés qu'il s'agisse de protéines ou de petites molécules (exemple : la glutathionyl-ménadione, GS-Men ou la glutathionyl-phénylacétophénone, GS-PAP) (Figure 7). A l'issue de la réduction, la cystéine catalytique est en pont disulfure avec le glutathion et doit être régénérée par l'apport d'une deuxième molécule de glutathion et formation de glutathion oxydé (GSSG). Ces Cys-GST seront décrites plus en détail dans la partie « GST à cystéine » page 47 de ce manuscrit.

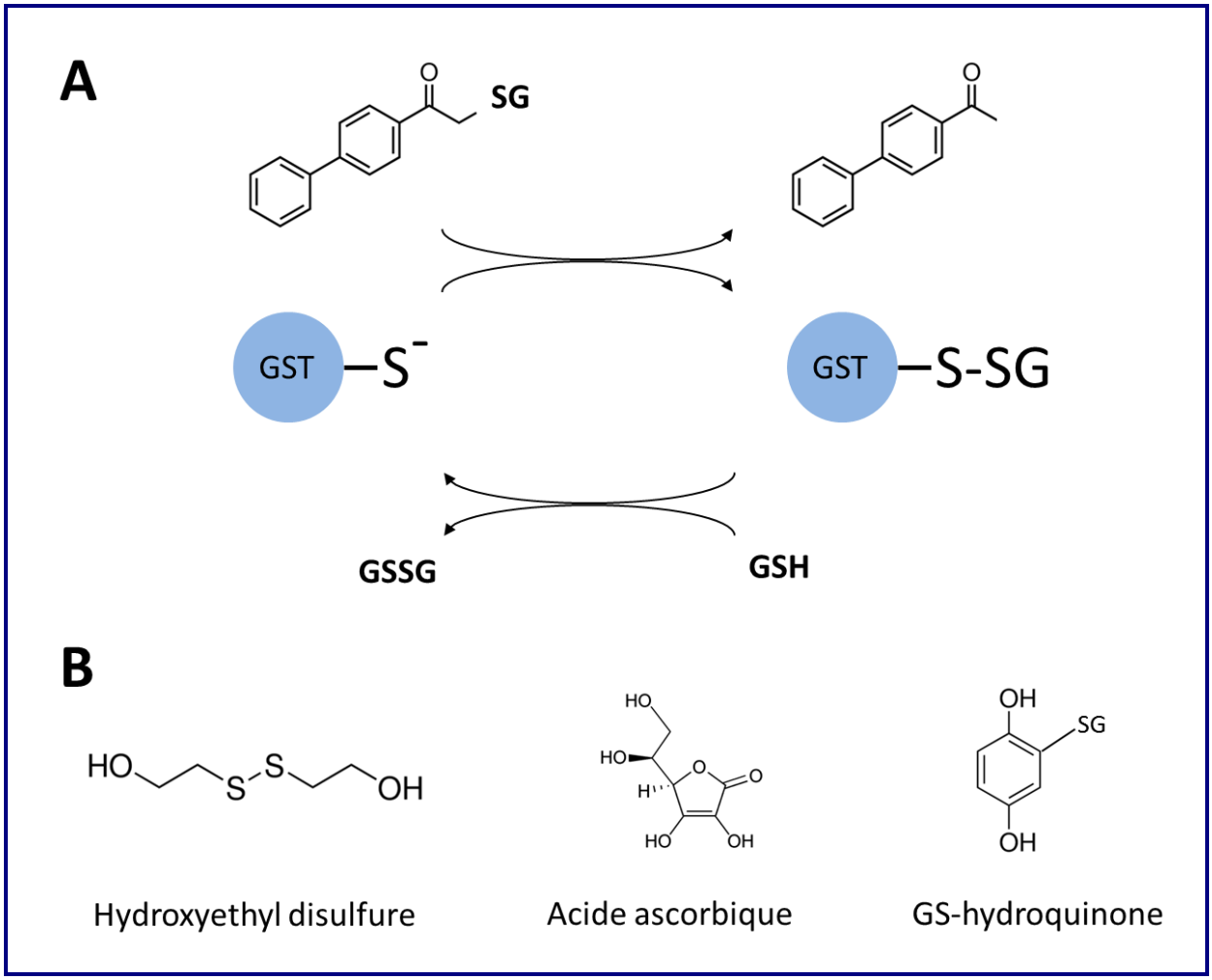


Figure 7 **Mécanisme catalytique des GST à cystéine**

A. Réduction du glutathionyl-phenylacetophenone GS-PAP catalysée par les GST Oméga. B. Autres substrats réduits par les GST à cystéine (adapté de Lallement et al, 2014).

I.1.c. Structure des principaux types de GST

Trois grandes familles de GST sont actuellement décrites et différenciées sur la base de leur localisation dans la cellule : les GST cytosoliques, les GST mitochondriales et les GST microsomales (Oakley, 2011). La famille des GST cytosoliques présente certainement le plus de diversité, tant par les fonctions assurées que par les structures tridimensionnelles adoptées par ces protéines. Les GST mitochondriales sont essentiellement représentées par la classe Kappa, ayant un repliement similaire aux GST cytosoliques avec une variation dans la succession des structures secondaires. Les GST MAPEG se distinguent des deux autres familles par un repliement et une structure quaternaire propres.

i) Les GST cytosoliques

- Structure globale

Les GST cytosoliques forment des homo-dimères d'environ 220 résidus d'acides aminés par monomère, avec des variations de longueur de séquence suivant les classes. Le repliement classique des GST cytosoliques comporte deux domaines (Figure 8). Le domaine N-terminal, fortement conservé, est constitué d'un repliement de type « thiorédoxine » car retrouvé chez d'autres protéines (familles des thiorédoxines et protéines disulfures isomérases notamment). Ce domaine (β1 α1 β2 α2 β3 β4 α3) est formé par un feuillet β mixte de 4 brins entouré de 3 hélices α. Le domaine C-terminal est rattaché au domaine N-terminal par une région linker d'une dizaine de résidus. Ce domaine C-terminal, tout alpha, contient au moins 4 hélices (α4 α5 α6 et α7) et présente plus de variabilité suivant les classes, avec notamment des hélices supplémentaires de localisation variable (Sheehan et al, 2001).

Les GST forment des homo-dimères présentant une symétrie 2. Dans la plupart des cas, le domaine N-terminal d'une sous-unité interagit avec le domaine C-terminal de l'autre sous-unité et vice-versa. Chez certaines classes de GST, cet arrangement est permis par une interaction impliquant un résidu hydrophobe phenylalanine de la boucle α2-β3 s'insérant entre les hélices α4 et α5 de l'autre monomère et appelé 'lock-and-key' (Sheehan et al, 2001) (par exemple dans la GST Mu de code PDB 6GST, Xiao et al, 1996). La nature des interactions inter-monomériques varie d'une classe à l'autre, en accord avec la variabilité du domaine C-terminal. Ainsi, la surface enfouie à l'interface peut varier de 2000 $Å^2$ à 3400 $Å^2$ selon les classes. A noter qu'un autre type de dimère a été observé chez les GST Xi (voir figure p.49).

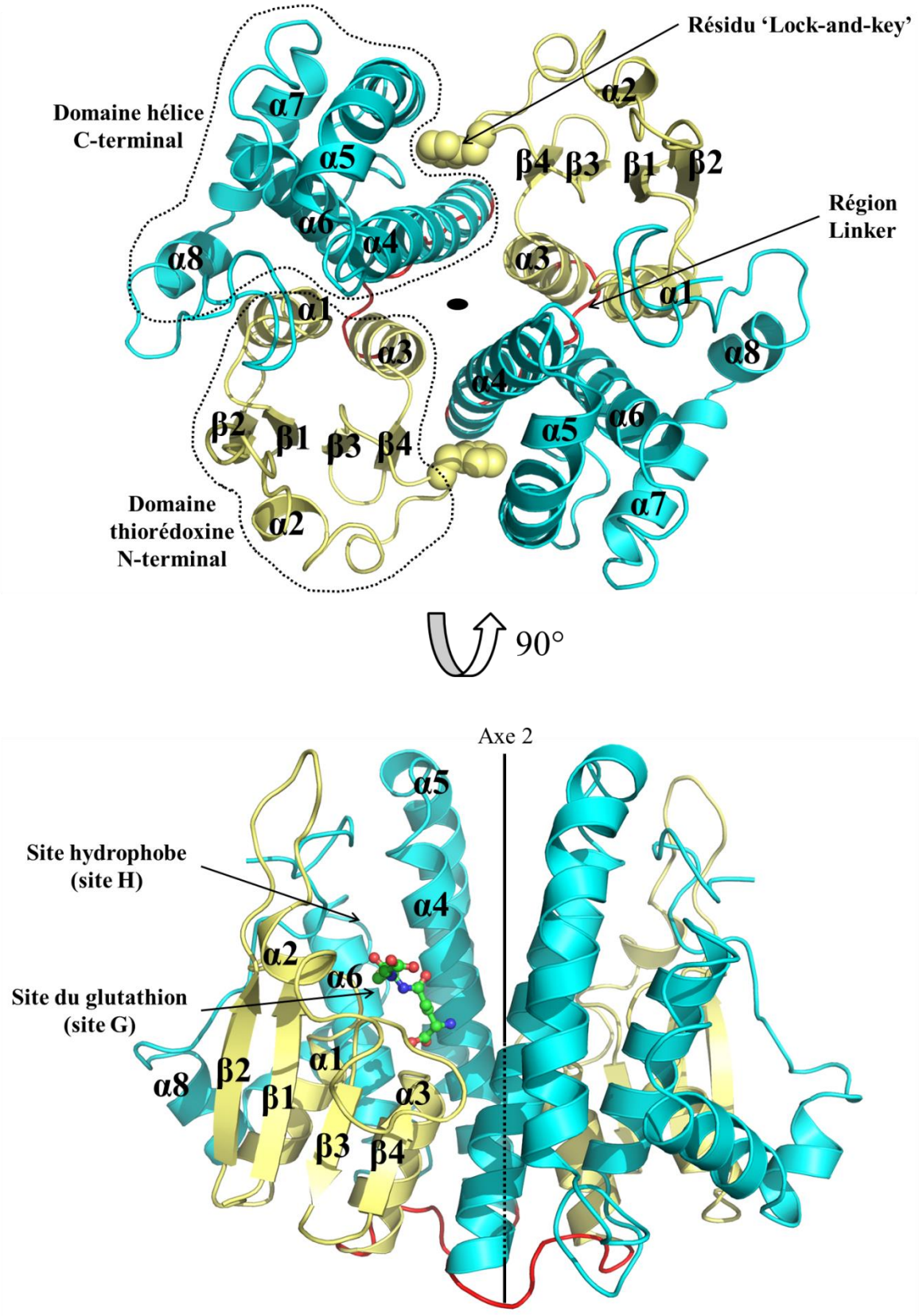


Figure 8 **Structure du dimère classique de GST : exemple de la GST Mu de rat (code PDB 6GST)**

Haut : vue dans la direction de l'axe 2 du dimère. Bas : vue perpendiculaire à l'axe 2 du dimère. Le domaine N-terminal est coloré en jaune pâle et le domaine C-terminal est coloré en cyan. La région linker est colorée en rouge. Les éléments de structure secondaire sont annotés.

➢ Site de liaison au glutathion (site G)

Le site de liaison du glutathion est situé au niveau du domaine thiorédoxine. Ce site est constitué de résidus stabilisant majoritairement le glutathion par des interactions polaires (Figure 9). Le site G contient, à l'extrémité C-terminale de la boucle α2-β3, une *cis*-proline qui est le seul résidu invariant dans la superfamille des GST (Meux et al, 2012). Le résidu précedent cette *cis*-proline, qui est souvent une valine, interagit via sa chaîne principale avec la chaîne principale de la cystéine du glutathion. Deux résidus contigus, présents en N-terminal de l'hélice α3, le plus souvent un glutamate et une sérine, sont également conservés et participent à la stabilisation de la partie γ-glutamyle du glutathion. La région de l'hélice α2, plus variable, porte des résidus généralement basiques stabilisant la partie C-terminale du glutathion. La chaîne latérale de la cystéine du glutathion est systématiquement retrouvée à proximité de l'extrémité N-terminale de l'hélice α1. Ainsi le soufre de la cystéine du GSH est en interaction polaire avec le groupement hydroxyle du résidu catalytique de la GST (le plus souvent une sérine située en N-terminal de l'hélice α1 ou une tyrosine située dans le brin β1), mais il peut également former un pont disulfure avec la protéine dans le cas des Cys-GST.

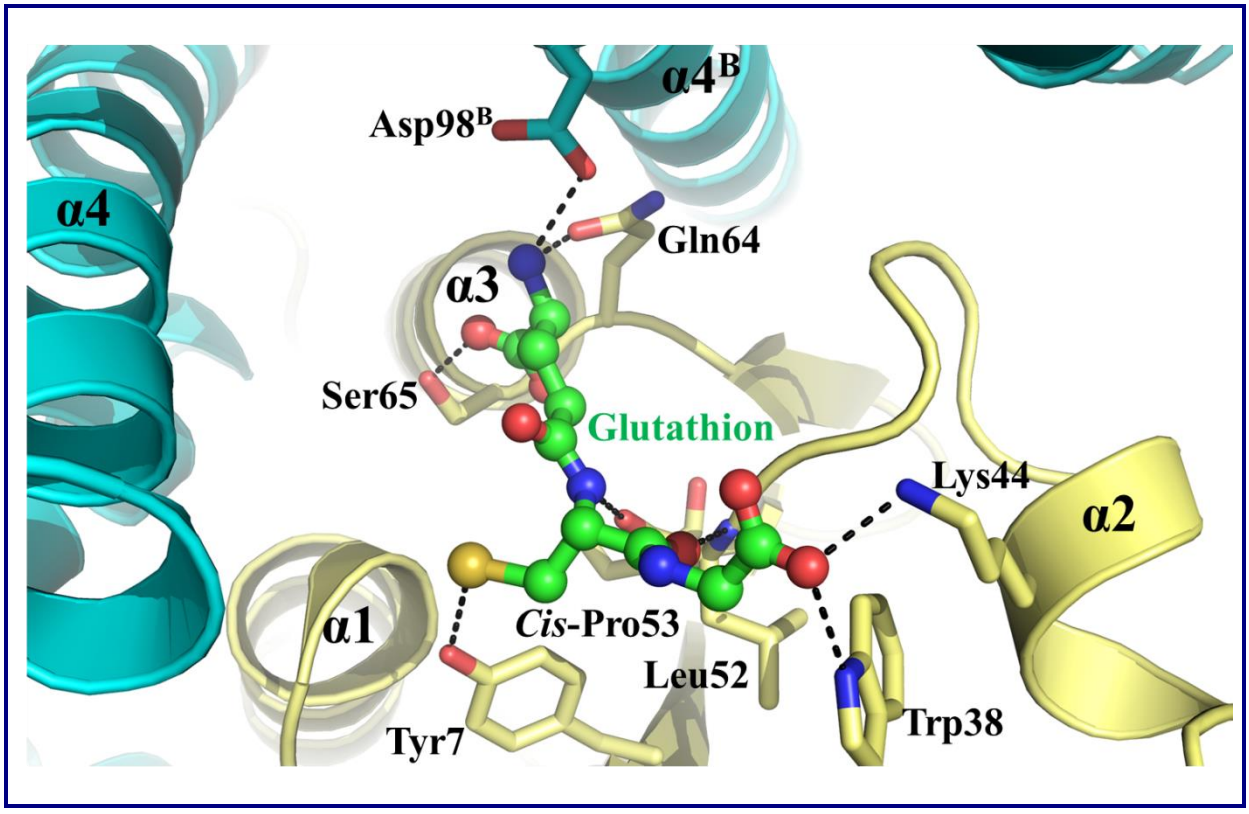


Figure 9 **Site de liaison du glutathion (site G) dans la structure de la GST Pi humaine (code PDB 1AQW)**

Les résidus participant aux interactions polaires sont annotés et représentés en batonnets. Les interactions sont représentées en pointillés.

➢ Site de liaison hydrophobe (site H)

Le site de liaison du substrat hydrophobe a été initialement mis en évidence à partir de la structure de la GST Pi humaine en complexe avec l'inhibiteur compétitif glutathionyl-hexane (Reinemer et al, 1991, code PDB 1GSS). Situé à proximité du site G, ce site est le plus souvent formé de résidus provenant des deux domaines N-terminal et C-terminal. Dans le cas de la GST Pi humaine, le site H prend la forme d'une crevasse formée par des résidus aromatiques (Figure 10) pouvant accommoder des groupements hydrophobes encombrants (exemple : benzo(a)pyrene diolepoxyde dans la structure de la GST Pi humaine de code PDB 3PGT, Ji et al, 1999). Alors que des variations dans la nature des résidus constituant le site H sont retrouvées selon les classes de GST, les interactions en jeu sont le plus souvent de nature hydrophobe.

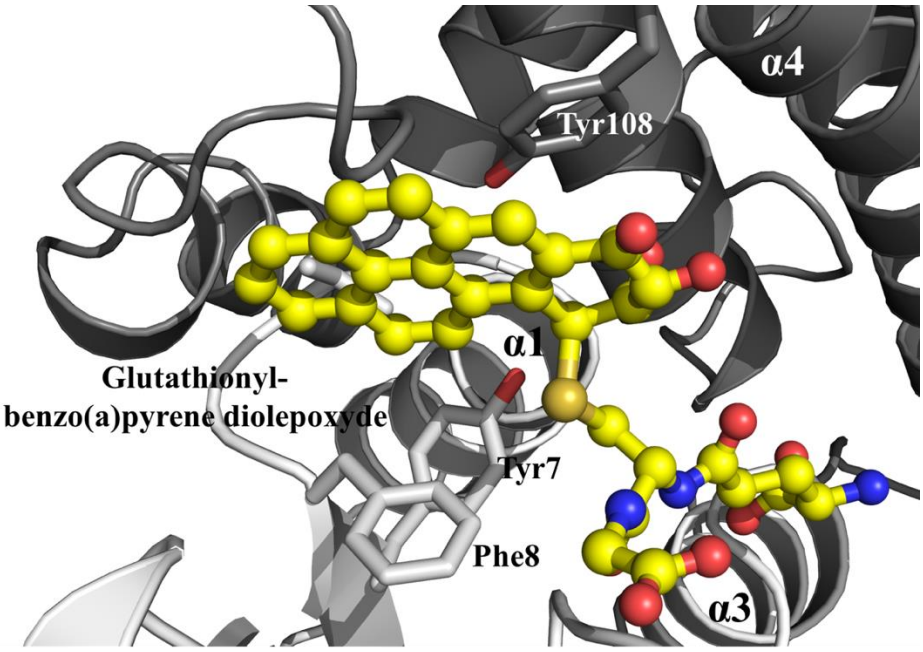


Figure 10 **Site de liaison hydrophobe (site H) de la GST Pi humaine en complexe avec le glutathionyl-benzo(a)pyrène diolepoxyde (code PDB 3PGT)**
Le site hydrophobe est une vallée composée de résidus aromatiques ou hydrophobes à proximité du site de liaison du glutathion.

➢ Sites ligandines (sites L)

D'autres sites de fixation sont décrits chez les GST. Ces sites sont appelés « ligandine » et se réfèrent à la fixation de molécules xénobiotiques sans mécanisme catalytique. Dans ce cas, le rôle supposé de la GST est de séquestrer de manière intracellulaire la molécule toxique ou de la transporter vers un autre site de détoxication (Hayes et al, 2005). Selon la classe de GST et la nature de la molécule, différents sites ligandines ont été identifiés

par cristallographie. La première structure de GST complexée à un ligand se fixant ailleurs qu'au site actif est la GST Mu de *Schistosoma japonica* en complexe avec l'anthelminthique praziquantel (McTigue et al, 1995). Cette molécule se fixe au niveau de l'interface du dimère de la GST, à proximité des hélices α3 et α4 de chaque monomère (code PDB 1GTB, Figure 11). Un site ligandine de localisation similaire à la GST Mu a été identifié à partir de la structure de la GST Oméga humaine mutée (C32A) en complexe avec le substrat GS-nitroacétophénone (Brock et al, 2013). Ce substrat ne se fixe pas au site actif de la GSTO mais à proximité de l'interface dimérique, le long des hélices α3 et α4 (code PDB 4IS0).

Dans le cas de la GST Pi humaine, a été reporté un site ligandine de localisation similaire au site H. Celui-ci est capable de fixer de larges molécules polyaromatiques telles que la sulfasalazine ou le bleu cibacron (code PDB 13GS). Malgré la fixation de ces molécules à proximité du site G, aucune réaction de glutathionylation n'a été détectée, suggérant un rôle non catalytique (Oakley et al, 1999).

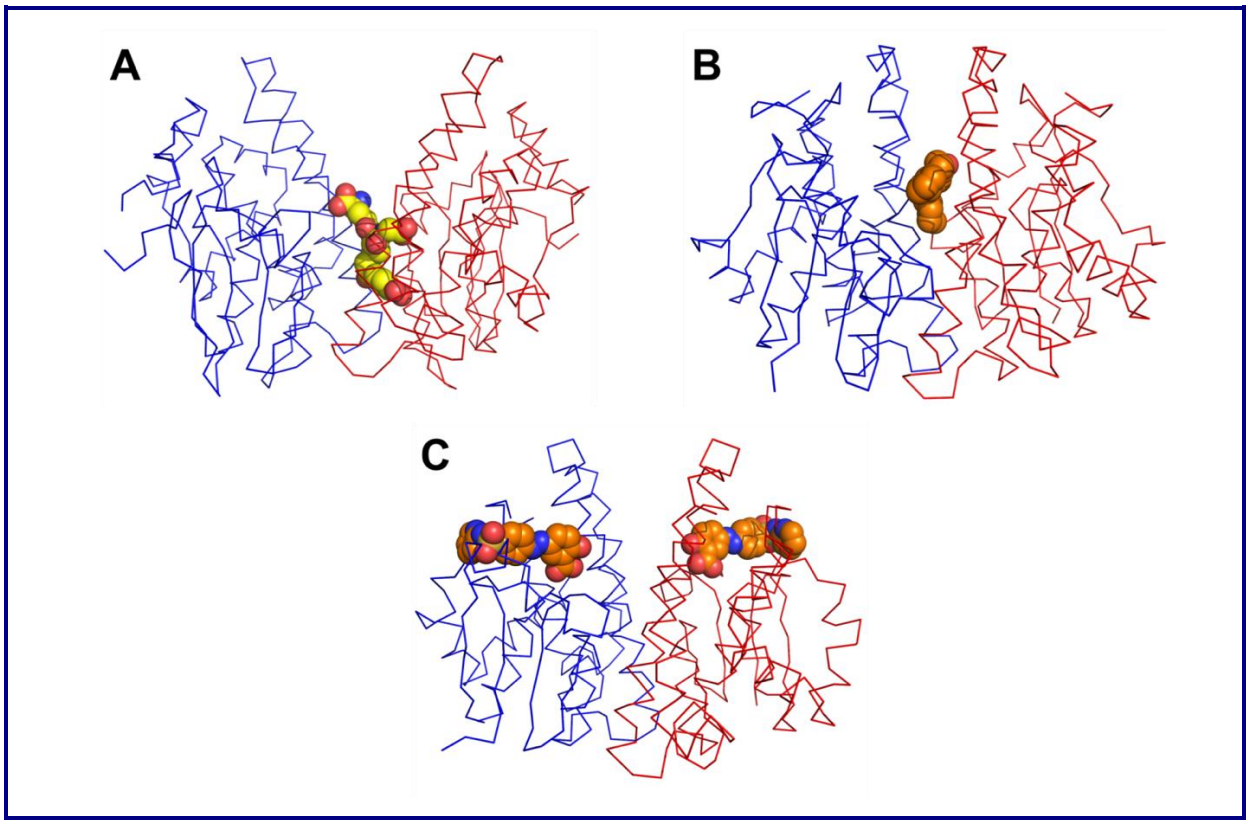


Figure 11 **Sites ligandines des GST**

A. GST Oméga 1 humaine (mutant C32A) en complexe avec le GS-nitrophénacyle (code PDB 4IS0), B. GST Mu de *Schistosoma japonica* en complexe avec le praziquantel (code PDB 1GTB), C. GST Pi humaine en complexe avec le bleu cibacron (code PDB 13GS).

Plus récemment, différents sites ligandines ont été mis en évidence pour la GST Phi 2 d'*Arabidopsis thaliana*. Cette GST est capable de lier plusieurs métabolites végétaux dont des

flavonoïdes à deux sites de fixation différents appelés site L1 (en périphérie du dimère) et L2 (à l'interface du dimère) (Ahmad et al, 2017). Ces résultats appuient l'hypothèse d'un rôle de transport de flavonoïdes dans la vacuole pour les GST végétales (Petrussa et al, 2013).

ii) Les GST mitochondriales Kappa

Tout comme les GST cytosoliques, les GST mitochondriales représentées par la classe Kappa forment des homodimères d'environ 220 résidus d'acides aminés. Alors que les GST cytosoliques sont constituées d'un domaine tout-hélice suivant le domaine thiorédoxine, les GST Kappa se caractérisent par l'insertion du domaine hélice au milieu de la séquence du domaine thiorédoxine (Figure 12). Leur structure globale les rapproche plus de la famille des disulfures isomérases (DsbA) que des GST cytosoliques (Ladner et al, 2004 ; Wang et al, 2011). Néanmoins, leur fonction GSH-transférase, sur le substrat CDNB notamment, est similaire à celle des GST cytosoliques, avec un site de fixation du glutathion structuralement proche de celui retrouvé pour les GST cytosoliques. Ces GST sont présentes dans la matrice des mitochondries et leur rôle biologique serait lié à la biosynthèse de l'adiponectine et à l'aide au repliement d'autres protéines (rôle de chaperone) (Morel et Anina, 2011).

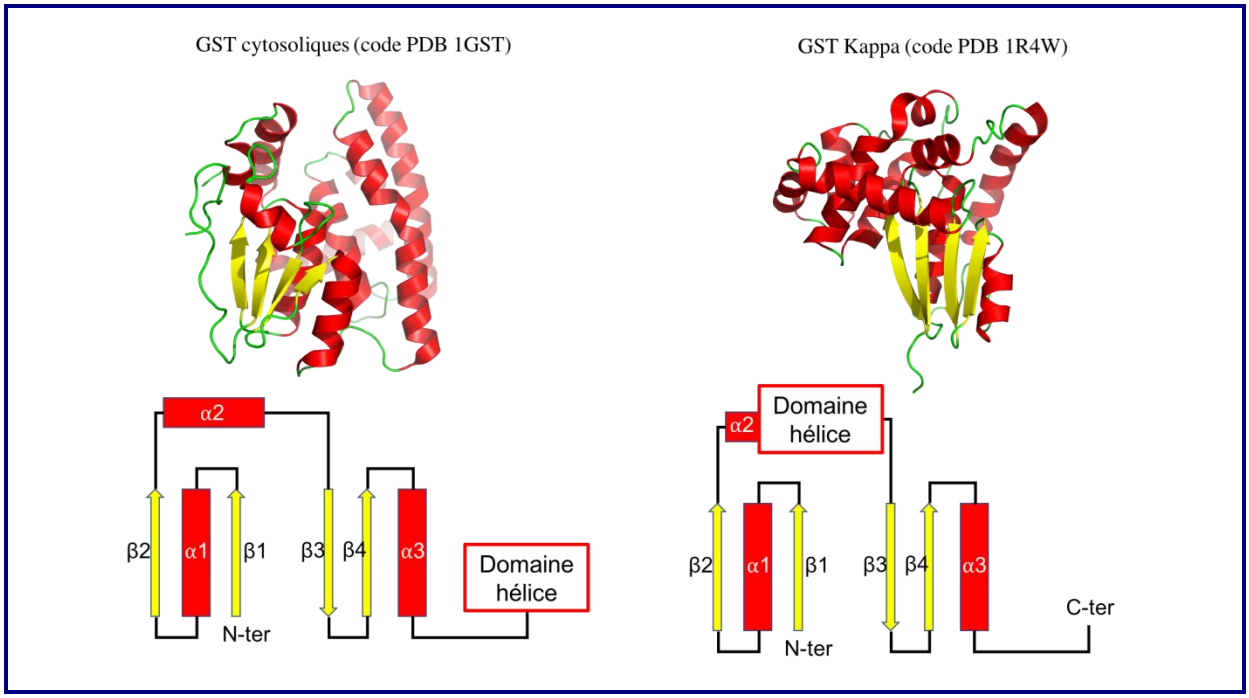


Figure 12 **Comparaison du motif structural adopté par les GST cytosoliques avec celui des GST mitochondriales de la classe Kappa (adapté de Board et al, 2013)**

La différence principale entre les GST cytosoliques et les GST Kappa est la localisation du domaine hélice : en C-terminal chez les GST cytosoliques et inséré entre les brins β2 et β3 chez les GST Kappa.

iii) Les GST MAPEG

Les GST MAPEG (pour 'Membrane Associated Proteins in Eicosanoid and Glutathione Metabolism) sont des protéines ayant une activité glutathion transférase mais évolutivement distinctes des GST cytosoliques. Celles-ci sont retrouvées parmi le sous-groupe I de la famille des MAPEG. La structure générale de ces protéines membranaires a été élucidée par cristallographie électronique (Holm et al, 2006) puis cristallographie aux rayons X (Molina et al, 2007). Ces GST forment des homotrimères (Figure 13). La sous-unité se caractérise par un fagot de 4 hélices s'insérant dans la bicouche lipidique. A partir de la structure de la leucotriène C4 synthase humaine en complexe avec le glutathion, le site actif de ces enzymes a été identifié. Celui-ci est retrouvé à l'interface entre deux sous-unités, dans le creux formé par les deux fagots d'hélices. De manière intéressante, le site de fixation du glutathion induit une conformation du tripeptide en « fer à cheval » (Molina et al, 2007), par opposition à la conformation étendue retrouvée chez les GST cytosoliques et mitochondriales.

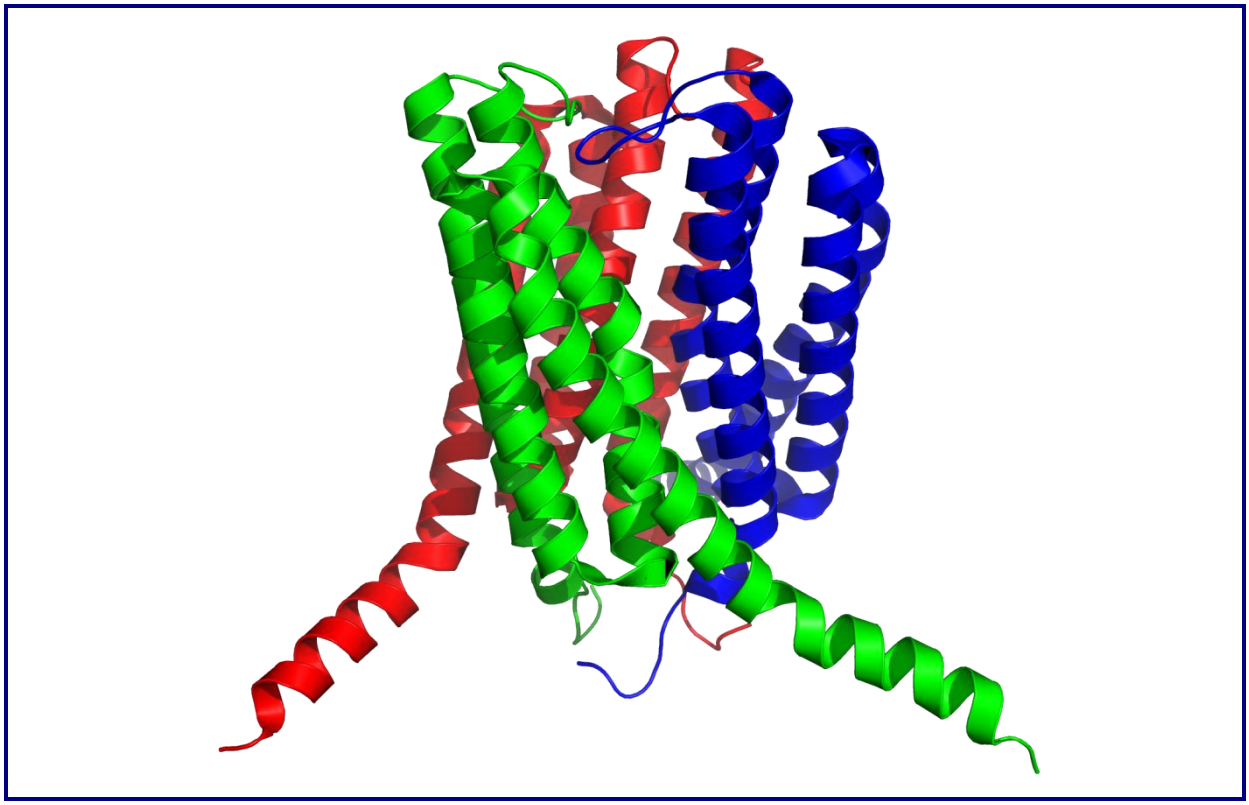


Figure 13 **Trimère biologique de la GST membranaire leucotriène C4 synthase humaine (code PDB 2UUH)**
La structure du trimère est représentée en mode « rubans » et les différents monomères sont colorés en vert, rouge et bleu.

I.2 Les GST chez les organismes fongiques

I.2.a. Les champignons dans le règne du vivant

i) Classification

Les organismes biologiques représentant le vivant sont regroupés en trois domaines distincts : les Archées, les Bactéries et les Eucaryotes. Un quatrième domaine situé hors de l'arbre du vivant est constitué par les virus, entités biologiques primitives. Les Eucaryotes représentent un domaine varié constitués de 4 règnes principaux regroupant des organismes très différents morphologiquement : les animaux, les plantes, les protistes et les champignons. Les champignons sont des organismes hétérotrophes (ayant besoin d'une source externe de carbone pour se nourrir) actuellement séparés en 6 phyla : Basidiomycota, Ascomycota, Glomeromycota, Chytridiomycota, Blastocladiomycota et Neocallimastigomycota (voir les ouvrages « The Fungi » de Watkinson et al et « Campbell Biology » de Reece et al). Cette classification est actuellement sujette à des études de taxonomie et est continuellement adaptée à la lumière des données génomiques. Cette étude est notamment rendue possible grâce au projet de séquençage de plusieurs espèces de champignons par le Joint Genome Institute (JGI), un centre de séquençage américain situé en Californie et sollicitant des scientifiques du monde entier, notamment ceux du site de l'INRA à Champenoux. Le projet du JGI appelé SAP pour Saprotrophic Agaricomycotina Project vise à séquencer le génome entier de plusieurs champignons saprotrophes afin de déterminer leurs relations évolutives mais également afin de mieux comprendre les mécanismes impliqués dans la dégradation du bois chez les décomposeurs (Eastwood et al, 2011 ; Floudas et al, 2012).

ii) Les basidiomycètes

Les quelques 30 000 organismes décrits dans le phylum Basidiomycota interviennent dans une multitude de processus écologiques (voir l'ouvrage de Watkinson et al, 2016). En effet, certains de ces champignons sont responsables de la dégradation et du recyclage de la matière morte constituée par le bois mort et les feuilles dans les forêts, alors que d'autres participent à des symbioses mycorrhiziennes avec des végétaux. D'autres peuvent également être pathogènes d'animaux ou de végétaux, ce qui est le cas notamment pour les agents en cause dans les rouilles ou les charbons (voir l'ouvrage de Watkinson et al, 2016). Les basidiomycètes se caractérisent par des traits génétiques communs et en particulier par la formation de spores appelés basidiospores à l'extérieur des cellules. Au sein des

Basidiomycètes est retrouvé un sous-phylum appelé Agaricomycotina, regroupant les champignons produisant des fructifications ainsi que certaines levures. Ce sous-phylum regroupe différents ordres : Agaricales (agarics), Boletales (bolets), Russulales, Polyporales, Phallales et Auriculariales. Parmi ces règnes sont retrouvés de nombreux décomposeurs de bois, notamment *Phanerochaete chrysosporium* et *Trametes versicolor* qui sont des modèles d'étude de la dégradation du bois et appartiennent à l'ordre des Polypores.

iii) Les décomposeurs du bois

La dégradation du bois par les champignons saprotrophes est un processus biochimique complexe dont le but est de fournir aux organismes la source carbonée nécessaire à leur métabolisme. Dans un premier temps, le champignon libère des enzymes de première phase (glyoxal oxydase, aryl-alcool oxydase) générant des espèces réactives de l'oxygène. Ces dernières sont utilisées par des enzymes de seconde phase (peroxydases et laccases) qui agissent sur les polymères présents dans le bois (lignine et cellulose) pour les dégrader en composés phénoliques et glycosidiques. Ces produits de dégradation sont ensuite internalisés par le champignon pour être pris en charge par d'autres systèmes enzymatiques intracellulaires (Schlosser et al, 1997 ; Hiscox et al, 2010). La dégradation du bois a pour conséquence la libération, au niveau de l'interface bois/champignon, de molécules extractibles. Ces molécules de faible poids moléculaire sont des constituants minoritaires vis-à-vis des polymères principaux constituant le bois mais sont importantes car lui conférant certaines caractéristiques : couleur, odeur et dureté notamment. On y retrouve les familles de molécules suivantes : glycérides et acides gras, terpènes, composés phénoliques (incluant phénols simples, stilbènes, lignanes et flavonoïdes), tannins.

iv) Le saprotrophe Trametes versicolor

L'agent de la pourriture blanche *Trametes versicolor* est un basidiomycète retrouvé couramment sur le bois mort au niveau de l'aubier (en périphérie), également capable de survivre dans le sol si les substrats y sont disponibles (Figure 14). Son rôle écologique est primordial dans la dégradation et le recyclage des composés du bois, et son action comme parasite des arbres est considérée mineure. L'intérêt porté à ce champignon depuis plusieurs siècles est multiple. En effet, des préparations pharmaceutiques à base d'extraits de *T. versicolor* sont particulièrement utilisées en Asie en médecine traditionnelle. Plus récemment, des études ont démontré le potentiel anti-cancer des polysaccharopeptides, molécules identifiées chez certaines souches de *T. versicolor* et possédant des propriétés de modulation

de la réponse immunitaire (Cheng et Leung, 2008). D'autre part, ce champignon présente également un intérêt biotechnologique puisque les laccases de *T. versicolor* sont utilisées depuis plusieurs décennies comme alternative à l'oxydation chimique nécessaire dans le process de fabrication de la pâte à papier (Camarero et al, 2007). Dans le même domaine, des études plus récentes montrent le potentiel du champignon dans le prétraitement du bois comme matière première (van Beek et al, 2007) et certains travaux démontrent des applications possibles en bioremédiation (Baldrian, 2008).

Figure 14 **Fructification du dégradeur de lignine *Trametes versicolor* sur un tronc de bois (photographie personnelle)**

L'étude moléculaire des organismes fongiques est animée par de nombreux enjeux aussi divers que les problématiques médicales impliquant des parasites, l'amélioration ou la protection des cultures en agriculture ou encore la mise au point de procédés biotechnologiques pour les industries papetières ou pour la dépollution des sols. L'étude de la capacité des champignons à pouvoir dégrader un large panel de molécules chimiques issues de leur environnement implique la caractérisation des systèmes intracellulaires disponibles, et notamment les GST.

I.2.b. Structure et fonction des GST fongiques

Actuellement, neuf classes de GST sont définies chez les champignons : Ure2p (Bousset et al, 2001a), Omega (Board et al, 2000), GTT (Choi et al, 1998), Xi (Meux et al, 2011), FuA (Mathieu et al, 2012), EF1Bγ (Jeppesen et al, 2003), MAK16 (McGoldrick et al, 2005), Zeta (Fernandez-Canon et Penalva, 1998) et Phi (McGoldrick et al, 2005). Les premières études chez la levure ont permis de mettre en évidence que certaines GST fongiques présentent des similarités avec les GST de mammifères (ex : les GST Oméga, Board et al, 2000) alors que d'autres sont évolutivement différentes et quelques fois spécifiques au règne fongique (ex : les Ure2p, Bousset et al, 2001a).

Au cours des années 2000, les études se sont élargies aux champignons supérieurs, Ascomycètes ou Basidiomycètes, et ont montré que la plupart de ces classes sont conservées chez l'ensemble des organismes fongiques (McGoldrick et al, 2005 ; Morel et al, 2009). En plus d'être conservées, certaines classes sont « étendues » chez certains organismes, ce qui signifie la présence de plusieurs gènes codant pour des isoformes d'une même classe de GST. Cette « redondance » génomique soulève question et interroge sur un lien potentiel avec la diversité chimique rencontrée par les champignons dans leur environnement naturel (Deroy et al, 2015). Par ailleurs, des études parallèles ont montré l'existence de GST fongiques non classées, dont le rôle peut être retrouvé dans le métabolisme secondaire notamment.

D'un point de vue structural, malgré la conservation du repliement typique des GST cytosoliques, des différences qui opposent les GST fongiques aux GST d'autres organismes ont été rapportées. Celles-ci peuvent être locales, telles des variations dans la nature du résidu catalytique au site actif (par exemple chez les classes Ure2p et GTT), ou conduire à des variations au niveau quaternaire (par exemple chez les GST FuA).

i) Les Ure2p

- <u>L'isoforme ScUre2P chez la levure</u>

Initialement identifiée chez la levure *Saccharomyces cerevisiae*, ScUre2p est une protéine présentant deux domaines. Alors que le domaine N-terminal (1-80) présente un domaine 'prion-like' riche en résidus asparagine, le domaine C-terminal (81-354) présente le repliement typique des GST (Umland et al, 2001). Chez la levure, Ure2p est un régulateur du catabolisme de l'azote par son interaction avec la protéine Gln3p liant l'ADN. Son rôle est de réprimer les protéines impliquées dans l'utilisation de sources pauvres en azote lorsque les

conditions extérieures permettent l'acquisition d'azote à partir de sources riches. Ure2p peut être retrouvée *de novo* sous forme de prion, appelé URE3. Sous cette forme, la protéine est agrégée dans la cellule et perd sa fonction régulatrice. De manière intéressante, le domaine C-terminal GST est suffisant pour assurer les fonctions de régulation.

La structure de la protéine ScUre2p tronquée au niveau N-terminal et contenant le domaine GST (95-354) a été résolue par cristallographie (Bousset et al, 2001a ; Umland et al, 2001). Le fragment protéique est dimérique et présente le repliement classique des GST. Plusieurs complexes entre ScUre2p et des dérivés du glutathion ont été déterminés (Bousset et al, 2001b). Le motif catalytique en début d'hélice α1 est APNG avec un résidu asparagine interagissant par liaison hydrogène avec le soufre du glutathion.

- Les Ure2p chez les champignons saprotrophes

Des GST de la classe Ure2p ont été identifiées chez d'autres champignons et notamment chez les décomposeurs avec quelques fois des extensions du nombre de gènes codant pour des Ure2p (Morel et al, 2009). Par exemple, le basidiomycète *Phanerochaete chrysosporium* possède 9 gènes codant pour des GST Ure2p (Morel et al, 2013) et l'analyse biochimique et structurale des PcUre2p a mis en évidence deux sous-classes A et B (Roret et al, 2015). Aucune de ces sous-classes ne possède l'extension N-terminale amyloidogène retrouvée dans ScUre2p. Alors que les Ure2pA sont spécifiques des Polypores et catalysent des réactions de transfert de glutathion, les Ure2pB sont retrouvées chez l'ensemble des champignons et catalysent des réactions de déglutathionylation de substrats glutathionylés. D'un point de vue structural, les deux sous-classes sont très proches. Des différences subtiles sont retrouvées au niveau des motifs catalytiques : alors que la sous-classe A présente un motif GPN, la sous-classe B présente un motif TPN. De manière intéressante, les Ure2p présentent une meilleure affinité pour le glutathion oxydé (GSSG) que pour le glutathion réduit, avec un site de liaison à l'interface entre les deux monomères (Figure 15, Roret et al, 2015).

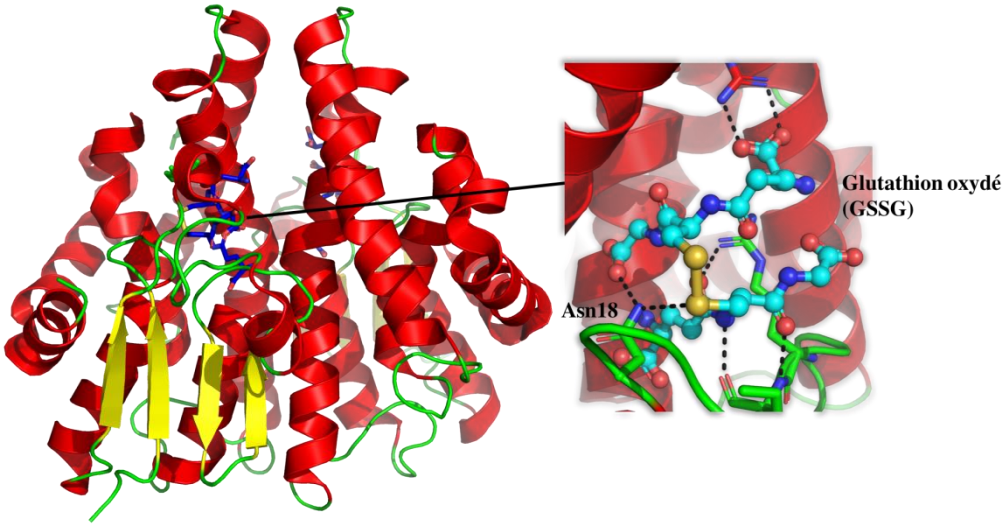


Figure 15 **Structure cristallographique de la GST Ure2p1 de *Phanerochaete chrysosporium* en complexe avec le glutathion oxydé (code PDB 4F0B)**

ii) Les GTT1 et 2

La classe GTT a été initialement définie à partir des deux isoformes de GST appelées GTT1 et GTT2 et retrouvées chez la levure *Saccharomyces cerevisiae* (Choi et al, 1998). Ces deux séquences ne présentent que peu de similarité avec les membres des autres classes de GST décrites. Chez la levure, GTT1 présente une activité glutathion transférase avec le CDNB et une activité peroxydase avec le cumène hydroperoxyde (Garcera et al, 2006). D'un point de vue physiologique, GTT1 serait associé au réticulum endoplasmique et impliquée dans la thermotolérance (Choi et al, 1998). Plus récemment, la classe GTT a été mise en évidence dans la plupart des champignons sur la base de résultats phylogénétiques (Morel et al, 2009).

En 2009, la structure de GTT2 de *S. cerevisiae* a été résolue par cristallographie (Ma et al, 2009). Le repliement adopté par GTT2 est le repliement typique des GST cytosoliques. Une particularité structurale est l'absence de résidu catalytique classique au niveau du site de reconnaissance du glutathion. En effet, l'activation nucléophile est assurée par une molécule d'eau en contact avec le soufre du glutathion. Cette molécule d'eau est stabilisée par deux résidus du domaine tout-hélice, à savoir Ser129 et His133, dont le rôle est indispensable à la catalyse (Ma et al, 2009).

iii) Les GST FuA (ou GTE)

En 2012, une nouvelle classe de GST appelée FuA pour 'Fungal Specific class A' a été décrite chez *Phanerochaete chrysosporium* (PcGSTFuA) (Mathieu et al, 2012). Cette classe a été initialement appelée etherase-like (GTE) car son plus proche homologue de GST caractérisée était la GST LigE de la bactérie *Sphingobium sp*. SYK6 capable de couper le lien β-ether d'un substrat modèle de la lignine dimérique (voir partie sur les GST impliquées dans la dégradation du bois, p.31). Les résultats biochimiques et structuraux suggèrent que cette enzyme possède trois sites de liaison : le site G, le site H et un troisième site S (proche du site G) qui lie les composés sulfatés. Les expériences de compétition montrent de plus que la GST FuA1 est capable d'interagir avec des composés issus du bois tels que la vanilline (Ki d'environ 160 µM), le coniferaldehyde ou encore le syringaldehyde avec des valeurs d'IC50 de l'ordre du µM, suggérant un rôle ligandine lié au site G (Mathieu et al, 2013).

Récemment, deux études indépendantes de GST FuA de champignons dégradeurs de bois montrent que ces enzymes ont la capacité de catalyser des réactions de type ethérase (chez *Ceriporiopsis subvermispora*, Osman et al, 2018, et *Dichomitus squalens*, Marinovic et al, 2018). En particulier pour cette dernière étude, il a été montré que DsGST1 clive sélectivement la liaison β-O-4 aryl ether d'un modèle de lignine dimérique aboutissant à un dérivé glutathion. Ces résultats montrent que la présence spécifique de GSTFuA chez les dégradeurs de bois pourrait être liée à leur mode de vie saprotrophe.

D'un point de vue structural, PcGSTFuA1 présente un mode de dimérisation atypique caractérisé par une crevasse particulièrement ouverte au niveau de l'interface du dimère (Figure 16). Cet arrangement est induit par un motif beta hairpin β2'- β2'' inséré dans la boucle α2-β3 et prévenant la formation du dimère classique. De plus, une hélice α4' supplémentaire est retrouvée entre les hélices α4 et α5 (Mathieu et al, 2012). Les structures des isoformes PcGSTFuA2 et FuA3 ont été résolues en 2013 (Mathieu et al, 2013), et présentent des caractéristiques similaires. Des variations subtiles sont retrouvées au niveau des sites actifs et des interfaces dimériques, en accord avec les profils d'interaction spécifique de chaque protéine, capables de lier de petits aldéhydes mais aussi des flavonoïdes tels que la catéchine.

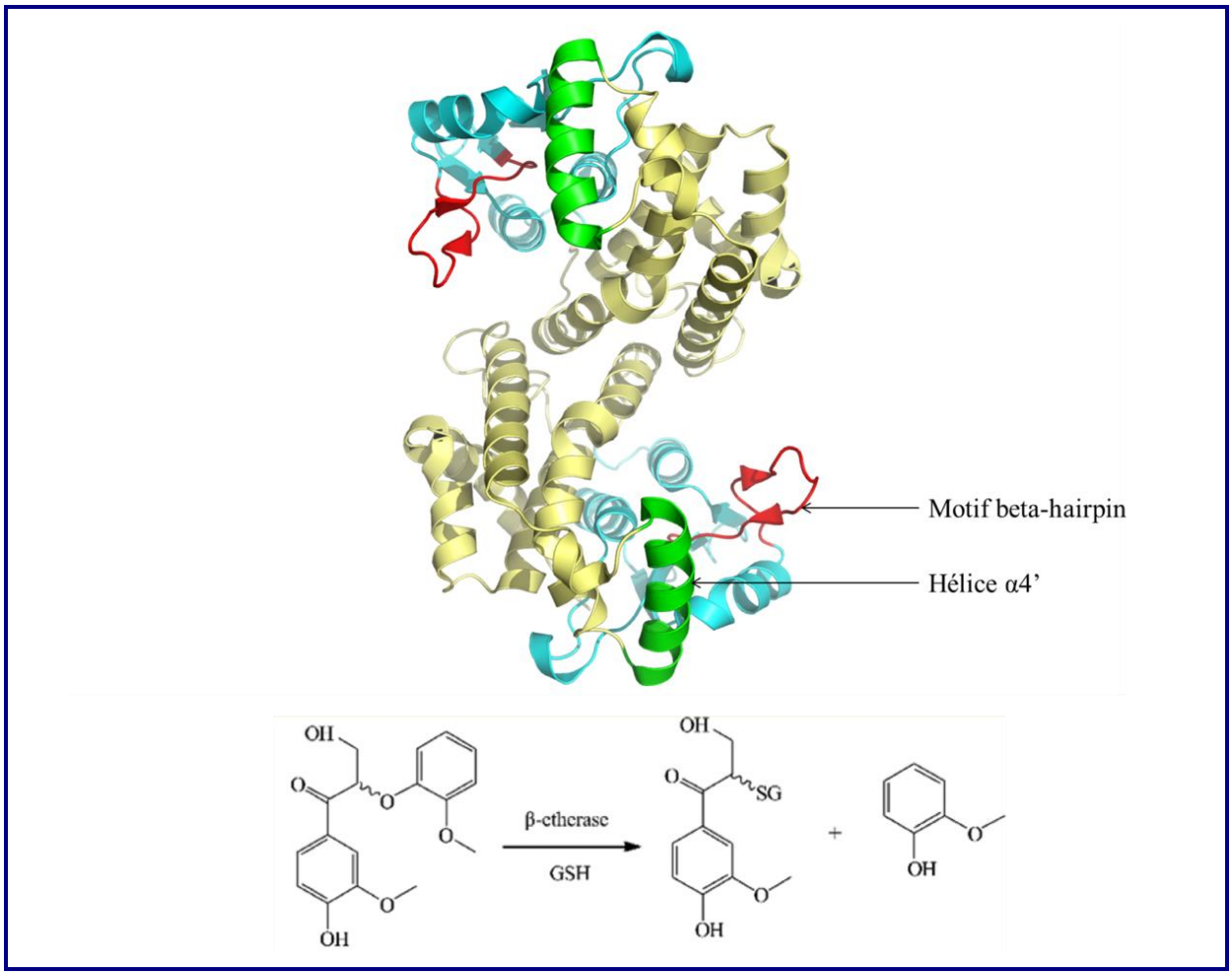


Figure 16 **Structure cristallographique de la GST FuA1 de *P. chrysosporium* (code PDB 4F03) et réaction β-étherase catalysée par la GST FuA de *Dichomitus squalens***

iv) Les GST à cystéine dans le règne fongique

En 2006, Garcera et collaborateurs ont mis en évidence chez la levure *S. cerevisiae* trois gènes codant pour des GST dont le résidu catalytique est une cystéine. Ces Cys-GST ont immédiatement été assimilées à la classe des GST Oméga, présente chez les Animaux et l'homme notamment (Board et al, 2000), et nommées ScGTO1, ScGTO2 et ScGTO3. En 2009, Morel et collaborateurs ont montré que la classe des GST Oméga, largement représentée chez les champignons, est déclinée en trois types (I, II et III) sur la base de leur séquence. La plupart de ces GST présentent une cystéine à la position habituelle du résidu catalytique (un résidu sérine est également retrouvé chez certaines GSTO de type III). Plus tard et sur la base d'études phylogénétiques, biochimiques et structurales, les GST Oméga de type I (parmi lesquelles les trois Gto de levure) se sont avérées appartenir à une classe distincte appelée Xi (Meux et al, 2011) alors que les GST Oméga des types II et III sont

phylogénétiquement plus proches des GSTO d'animaux et des GST Tau et Lambda de plantes.

- Les GST Xi

Bien qu'initialement identifiées chez les champignons, ces GST sont également présentes chez les plantes, les bactéries et les archées. Elles présentent un résidu catalytique cystéine leur permettant de catalyser des réactions de type thiol-transférase (à l'instar des glutarédoxines monothiol). Aucune activité GSH-transférase n'a pu être détectée à ce jour. De par leur fonction, les GST Xi sont également appelées GHR pour Glutathionyl-(Hydro)quinone Réductase. En effet, ces GST catalysent la réduction spécifique de dérivés quinones glutathionylées (Lam et al, 2012). En 2011, la première structure cristallographique résolue de GST Xi (PcGSTX1 de *Phanerochaete chrysosporium*, Meux et al, 2011, Figure 17) a mis en évidence des caractéristiques uniques. En effet, les GST Xi présentent des particularités structurales inédites :

- un motif catalytique CPWA comprenant la cystéine catalytique située en début d'hélice α1.
- des extensions N- et C- terminales en regard des autres familles de GST, ce qui en fait des protéines plus massives (environ 320 résidus contre 220 résidus pour les autres classes de GST).
- une triade de tyrosine intervenant dans la catalyse (Green et al, 2012).
- une insertion entre l'hélice α2 et le brin β2 gênant la formation du dimère canonique des GST (Meux et al, 2011).
- un mode de dimérisation particulier par les domaines hélices des deux monomères.

Alors que la fonction de ces protéines est bien définie, le rôle biologique des GST Xi chez les champignons reste incertain mais pourrait être lié au métabolisme des composés phénoliques (Belchik et Xun, 2011), dans la régulation rédox de protéines du métabolisme du soufre (Barreto et al, 2006) ou dans la biosynthèse de la paroi fongique (Lussier et al, 1997). A noter que la fonction biologique de la GST Xi PcpF chez la bactérie *Sphingobium chlorophenolicum* est retrouvée dans la dégradation GSH-dépendante du pentachlorophénol (Huang et al, 2008).

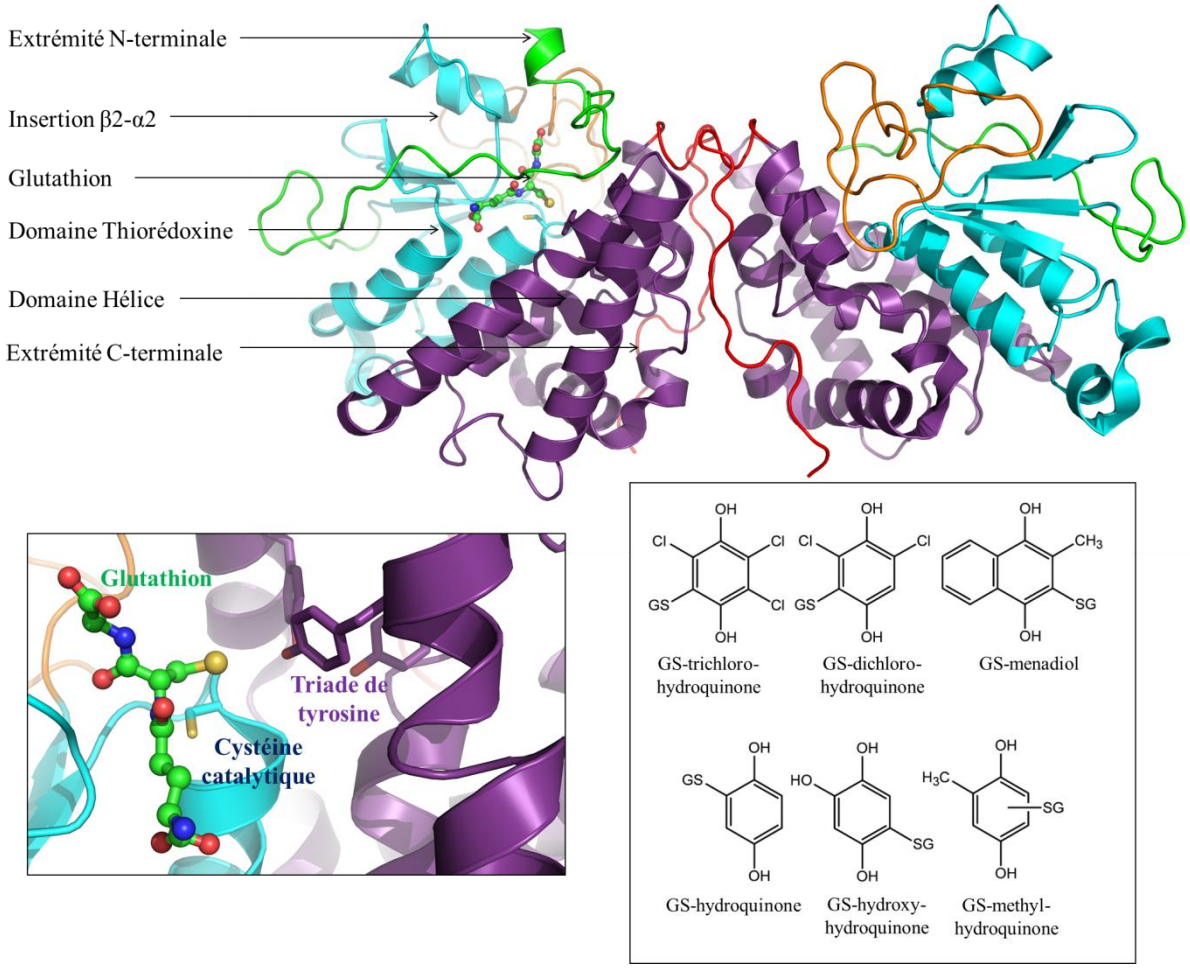


Figure 17 **Structure cristallographique de la GST Xi de *P. chrysosporium* (code PDB 3PPU) et substrats de type « glutathionyl-hydroquinone » réduits par les GST Xi**

➢ Les GST Oméga

Bien que la classe Oméga ait été identifiée en 2000 chez l'homme, la présence de « véritables » GSTO (non GST Xi) chez les champignons n'a été évoquée qu'en 2009 sur la base d'analyse de génomes (Morel et al, 2009). Ces GSTO des types II et III ont été identifiées chez les Basidiomycètes et certains Ascomycètes avec pour certains organismes des extensions de classes (ex : 7 isoformes chez *Phanerochaete chrysosporium*). Des GSTO des types II et III de *P. chrysosporium* ont été caractérisées biochimiquement en 2013 (Meux et al, 2013). Alors que PcGSTO3 et PcGSTO4 (type II) présentent une cystéine catalytique et une activité thiol-transférase, PcGSTO7 (type III) présente une sérine catalytique et une activité GSH-transférase. PcGSTO3 et 4 sont également capables de réduire des substrats glutathionylés tels que le GS-phenylacétophénone (Figure 7) ou la glutathionyl-tétralone. A

noter que Board et collaborateurs ont mis au point le substrat GS-nitroacétophénone qui est réduit spécifiquement par les GSTO (Board et Anders, 2007 ; Board et al, 2008). De manière intéressante, l'activité GSH-transférase de PcGSTO4 peut être restaurée en mutant le résidu catalytique C37 en S37. En revanche, l'activité thiol-transférase de PcGSTO7 ne peut être restaurée en mutant le résidu S15 en C, suggérant des différences en termes de catalyse entre les deux types. Une approche de screening a permis de déterminer que les GSTO de type II sont capables d'interagir avec des molécules de la classe des terpènes avec des affinités de l'ordre de 10 à 100 µM. Concernant les GSTO de type III, une étude a permis de déterminer que les isoformes TvGSTO1S à TvGSTO6S issues du champignon *T. versicolor* sont capables d'interagir avec des molécules issues d'extraits de bois, parmi lesquelles potentiellement les flavonoïdes (Deroy et al, 2015). Ces données biochimiques suggèrent que les GSTO de champignon pourraient avoir un rôle dans le transport ou le métabolisme des composés issus de la dégradation du bois, connue pour libérer des molécules potentiellement toxiques.

En l'absence de données structurales obtenues sur la classe Oméga chez les champignons, les lignes suivantes décrivent succinctement la structure de la GST Oméga humaine déterminée en 2000 par Board et collaborateurs (Board et al, 2000). Cette structure a été déterminée par cristallographie à 2,0 Å de résolution (Figure 18). Le glutathion présent dans la structure est lié de façon covalente à la cystéine catalytique (C59). Le repliement typique des GST cytosoliques est retrouvé avec des particularités structurales. D'une part, la structure est caractérisée par la présence d'extensions en N-ter et C-ter (notamment une hélice α9 supplémentaire) se rejoignant et formant un nouveau motif structural. D'autre part, la présence de la cystéine catalytique en début d'hélice α1 est en cohérence avec l'activité thiol-transférase de la GST. De plus, le dimère formé par l'assemblage des deux monomères présente une crevasse particulièrement large et accessible à des ligands chimiques (Brock et al, 2013) mais aussi potentiellement à des protéines (Menon et Board, 2013). A noter que la structure de l'isoforme GSTO2 humaine (Zhou et al, 2012), ainsi que les structures des GSTO1 et GSTO3 du ver à soie *Bombyx mori* (Chen et al, 2011 : Yamamoto et al, 2013) ont été résolues et présentent une structure similaire à la GSTO1 humaine.

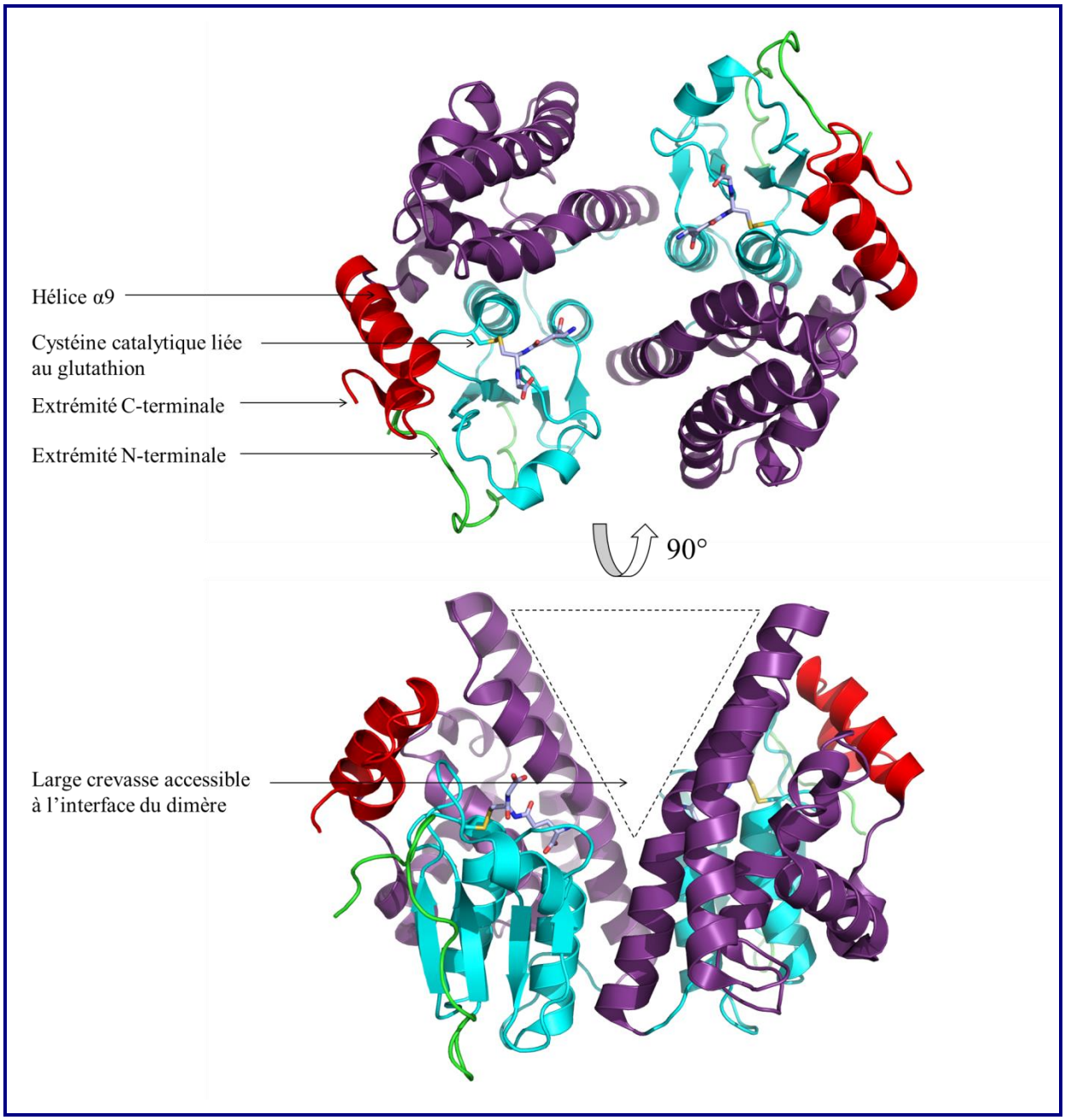


Figure 18 Structure cristallographique de la GST Oméga 1 humaine (code PDB 1EEM)

v) Les GST Zeta

Les GST de la classe Zeta ont été identifiées à l'origine chez l'ascomycète *Aspergillus nidulans* de manière concomitante avec le gène homologue chez l'homme (Fernandez Canon et al, 1998). Ces GST semblent présenter la même fonction biologique dans les deux organismes : le catabolisme de la phénylalanine. Cependant les GST Zeta ne sont pas retrouvées systématiquement chez tous les champignons, suggérant l'existence d'autres systèmes de dégradation des acides aminés aromatiques. En 2011, une structure de GST Zeta

provenant du champignon pathogène *Coccidioides immitis* a été déterminée sous les deux formes apo et complexée au glutathion (Edwards et al, 2011). La structure de cette GST est proche de ses homologues chez l'homme (Polekhina et al, 2001), les plantes (Thom et al, 2001) et les bactéries (Marsh et al, 2008), avec probablement une fonction similaire.

vi) Les GST Phi

En 2013, Morel et collaborateurs ont rapporté la présence de la classe Phi des GST chez certains champignons notamment les Basidiomycètes (Morel et al, 2013). Cette classe est habituellement retrouvée chez les plantes et peu d'études rapportent leur rôle chez les champignons. Une étude protéomique présente l'expression accrue de GST Phi chez *P. chrysosporium* en milieu supplémenté en vaniline, un intermédiaire issu de la dégradation de la lignine (Shimizu et al, 2005). Aucune étude biochimique ou structurale de GST Phi fongique n'est actuellement publiée.

vii) Les protéines appartenant à la famille des GST et impliquées dans la traduction ou la biogenèse des ribosomes

- Le facteur d'élongation eEF1Bγ

En 2003, la structure du domaine N-terminal de la sous-unité γ du facteur d'élongation 1B a été résolue chez la levure (Jeppesen et al, 2003). Ce domaine présente le repliement typique des GST mais la protéine ne présente pas d'activité GST connue. Le dimère de GST formé par eEF1Bγ est typique des GST cytosoliques et participerait à l'assemblage du facteur d'élongation 1 en interagissant avec les sous-unités 1A et 1Bα. Les GST EF1Bγ sont ubiquitaires chez les champignons (Morel et al, 2009) compte tenu de leur rôle important joué dans le processus de traduction.

- La protéine MAK16p

En 2006, Pellett et Tracy ont mis en évidence chez la levure *S. cerevisiae* une protéine nucléolaire impliquée dans la biogenèse du ribosome. Cette protéine est appelée MAK16p et présente une homologie de séquence avec les GST cytosoliques. Plus précisément, son rôle serait retrouvé dans la maturation des ARN ribosomiques 5.8S et 25S dans le noyau. Aucune structure expérimentale n'est à ce jour décrite dans la littérature, de même qu'aucune activité GST n'est décrite pour cette classe.

viii) Les GST fongiques non classées et liées au métabolisme secondaire

En marge des études classiques se focalisant le plus souvent sur la caractérisation d'un gène de GST au niveau moléculaire puis structural, des études centrées sur des voies métaboliques ont mis en évidence des GST. Dans ce cas, l'approche employée inclut la caractérisation exhaustive de gènes impliqués dans la biosynthèse de métabolites souvent d'intérêt pharmacologique. Parmi ces gènes, certains codent pour des GST de classe non déterminée, principalement chez les Ascomycètes du genre Aspergillus.

- Les GST liées à la synthèse des ETP (EpipolyThiodioxoPiperazine)

En 2004, un cluster de plusieurs gènes impliqués dans la biosynthèse de la sirodesmine a été mis en évidence chez le pathogène de plante *Leptosphaeria maculans* (Gardiner et al, 2004). La sirodesmine est un composé de la famille des ETP qui sont des molécules produites par des pathogènes de plante et de mammifères. Parmi ces neuf gènes est retrouvé sirG, un gène présentant une homologie de séquence avec les GST. Il a été proposé que cette GST participe à la biosynthèse GSH-dépendante, compte tenu de la présence de soufre dans le produit final. Une étude analogue a été effectuée en 2011 sur le cluster de gènes impliqués dans la biosynthèse de la gliotoxine chez *Aspergillus fumigatus* (Davis et al, 2011). La protéine GliG est une glutathion transférase responsable de l'incorporation de soufre dans le produit final par le biais d'une double glutathionylation.

- Les GST liées à la synthèse de polycétides

Chez l'ascomycète *A. fumigatus*, a été reporté le rôle de la GST PsoE dans la biosynthèse des pseurotines, métabolites présentant des activités biologiques d'intérêt médical (Yamamoto et al, 2016). PsoE intervient en conjuguant le glutathion à l'intermédiaire presynerazol. Dans un second temps, PsoF libère le conjugué en réalisant une isomérisation d'une double-liaison. Une étude analogue rapporte le rôle similaire d'une GST impliquée dans la biosynthèse de l'hypothemicine chez *Hypomyces subiculosus* (Reeves et al, 2008). La structure cristallographique de la GST PsoE a été résolue en complexe avec l'intermédiaire glutathionyl-presynerazol. L'examen du site actif révèle que les principales interactions stabilisant la molécule sont retrouvées au niveau de la partie glutathion, suggérant une libération facile du produit formé. Par ailleurs, PsoE présente le repliement canonique des GST.

Le but de cette partie était de répertorier les GST retrouvées chez les champignons en les analysant du point de vue de leur structure et de leurs fonctions et rôles. De nombreuses classes sont décrites, recoupant quelques fois les GST décrites chez les mammifères, et parfois étant spécifiques au règne fongique. De manière similaire aux GST d'animaux ou de plantes, les GST fongiques présentent de nombreux autres rôles que le transfert de glutathion dans la détoxification. La fonction ligandine de ces protéines est notamment proposée dans différentes classes (Oméga et FuA) et suggère un rôle important. La présence redondante du repliement GST pour assurer des fonctions différentes, parfois sans activité GSH-transférase (ex : EF1Bγ) indique que ce motif structural est particulièrement adaptable. Enfin, la présence redondante de gènes codant pour des GST dans les clusters géniques de biosynthèse de métabolites suggère une conservation de la fonction GSH-transférase comme outil de modification de composé.

II L'identification de ligands de macromolécules par cristallographie

En cristallographie macromoléculaire, un ligand est une molécule en interaction avec la protéine (ou l'acide nucléique), visible et modélisable dans la densité électronique. Dans la PDB, plus de 75 % des structures cristallines déposées contiennent au moins un ligand (Müller, 2017). Ces molécules peuvent se fixer de manière fortuite, dans le cas des composés chimiques présents dans la solution de cristallisation par exemple. Dans d'autres cas, leur présence est délibérément provoquée par leur ajout à l'expérience de cristallogenèse avant ou après cristallisation.

Les complexes protéine-ligands déterminés à haute résolution permettent d'accéder à des informations précises sur le fonctionnement des molécules en solution (Müller, 2017). Ces données sont notamment nécessaires aux laboratoires pharmaceutiques pour la mise au point d'un médicament ('drug design'). La recherche de complexes implique de cribler des chimiothèques pour identifier quels ligands potentiels interagissent avec la protéine. Le plus souvent, le criblage cristallographique consiste à mesurer au préalable l'affinité de la protéine pour des ligands potentiels par des méthodes biophysiques, puis de déterminer les complexes par cristallographie. Une alternative plus récente suggère d'utiliser la cristallographie comme méthode initiale d'identification de ligand, sans prérequis concernant l'affinité des complexes

en solution. Ce chapitre décrit les techniques et approches existantes et détaille les techniques utiles à la détermination d'une structure de complexe protéine-ligand par cristallographie.

II.1 Etapes préliminaires à l'étude structurale haut débit de complexes protéines-ligands

II.1.a. Mise au point de la construction protéique adaptée à une étude protéine-ligand

Un des prérequis indispensables à la détermination d'une structure de complexe protéine-ligand est la mise au point d'une « construction génique » permettant de produire la protéine en quantité suffisante pour la cristallogénèse. Cette étape est cruciale car elle détermine certaines caractéristiques essentielles de la protéine recombinante qui sera utilisée pour les essais de cristallisation. Parmi les paramètres expérimentaux sont notamment retrouvés le choix des plasmides et du système d'expression (bactérien, levure, cellules d'insectes ou de mammifères), la méthode de clonage, l'incorporation du/des domaine(s) d'intérêt de la protéine étudiée puis une stratégie efficace de purification. Cette mise au point doit d'une part permettre d'obtenir un échantillon protéique amenant à des cristaux, et d'autre part permettre un packing cristallin adapté à la fixation de ligands au niveau des sites de liaisons, sans problèmes d'encombrements.

Une stratégie efficace évoquée par Müller (2017) et employée par la plupart des laboratoires de biologie structurale est de démarrer un projet d'étude structurale par le design de plusieurs constructions protéiques différentes afin d'optimiser les chances de succès. Cette méthode, couplée au criblage d'un nombre modéré de conditions de cristallisation serait plus fructueuse que l'emploi d'une seule construction protéique et de nombreux tests de cristallisation. A titre d'exemple, le groupe de génomique structurale d'Oxford rapporte la résolution fructueuse de structure par cristallographie d'environ 40 % des protéines eucaryotes purifiées, par l'emploi de 9 à 20 constructions géniques différentes par cible (Savitsky et al, 2010), soulignant l'importance de cette étape.

II.1.b. Mesure de l'affinité protéine-ligand et méthodes expérimentales de criblage

Lorsque la protéine d'intérêt a été exprimée et purifiée avec succès, le criblage expérimental de chimiothèques peut être entrepris. Le plus souvent, cette étape est effectuée

par l'utilisation de techniques biochimiques ou biophysiques adaptées au criblage systématique de molécules en solution. En effet, la connaissance préalable de l'affinité de la protéine pour un ligand augmente significativement les chances d'obtenir un complexe protéine-ligand au sein du cristal. D'après une grande compagnie pharmaceutique, même pour les systèmes d'étude de complexe par cristallographie bien établis, les chances d'obtenir une structure d'un complexe désiré ne sont que de 65 % (Chung, 2007), d'où l'importance de rassembler le maximum d'informations sur ces complexes avant d'entreprendre une étude cristallographique.

i) Techniques biochimiques

Lorsque la cible étudiée présente une activité enzymatique et que les ligands recherchés sont des inhibiteurs de cette activité, un test d'inhibition d'activité peut être utilisé comme crible. Cette méthode a notamment été utilisée pour déterminer des ligands potentiels de GST Oméga, par l'utilisation de l'activité estérase de ces enzymes et d'une sonde fluorescente (Meux et al, 2013). Sur la base des résultats du criblage, les inhibiteurs les plus prometteurs peuvent être retenus pour effectuer des mesures enzymatiques visant à déterminer leur constante d'inhibition (K_i) et renseignant par-là sur leur affinité pour le site actif de la protéine.

ii) Techniques biophysiques

En l'absence d'activité enzymatique, d'autres techniques biophysiques peuvent être employées pour mesurer ou estimer l'affinité protéine-ligand. Quelques-unes des techniques les plus couramment associées à une étude cristallographique (Chung, 2007) sont détaillées dans les lignes suivantes.

- La titration calorimétrique isotherme (ITC)

La méthode principale permettant de déterminer l'affinité d'une protéine pour un ligand est l'ITC (pour 'Isothermal Titration Calorimetry'). Cette méthode repose sur l'injection répétée d'une solution de ligand dans une cuve isolée thermiquement et contenant la protéine. Les échanges de chaleur mesurés permettent de déterminer de nombreuses informations telles que la constante de dissociation, la stœchiométrie et les variations d'enthalpie et d'entropie de réaction. Néanmoins, cette méthode repose d'une part sur la détermination préalable de la concentration adéquate de protéine (qui est fonction de l'affinité

de liaison pour le ligand) et d'autre part s'avère consommatrice en échantillon, deux facteurs pouvant être limitants pour une étude haut-débit.

- La résonance plasmonique de surface (SPR)

La technique de résonance plasmonique de surface (SPR) repose sur un principe différent : la protéine est fixée sur une surface métallique et une variation de l'indice de réfraction de la surface est mesurée lors de la fixation du ligand. Dans le cadre d'un criblage de chimiothèques, bien que riche en informations, cette méthode présente le désavantage de nécessiter un équipement coûteux et un investissement en temps pour la mise au point du système expérimental.

- La spectroscopie RMN

Une technique particulièrement puissante pour étudier les systèmes biomoléculaires est la spectrométrie RMN (résonance magnétique nucléaire). Cette technique repose sur la mesure de la résonance entre noyaux magnétiquement actifs (de spin > ½) tels que ceux retrouvés dans les atomes ^{1}H, ^{13}C, ^{15}N. En plus de son potentiel pour déterminer des structures de petites protéines en solution, la RMN peut être utilisée pour des études d'interactions. Deux ensembles de techniques existent : les techniques basées sur le ligand et les techniques basées sur la protéine.

Dans le cas des expériences centrées sur le ligand, les différences de déplacements chimiques observés entre le ligand libre et le ligand en présence de la protéine renseigne sur son mode de liaison dans la protéine. Une difficulté réside dans la capacité à mesurer le signal du ligand parmi celui de la protéine. Pour surmonter ce problème, le ligand peut être marqué avec des noyaux actifs magnétiquement et non présent dans la protéine (ex : ^{31}P ou ^{19}F), ce qui peut être fastidieux en cas de criblage d'un grand nombre de molécules.

Dans le cas des expériences centrées sur la protéine, celle-ci est le plus souvent enrichie en ^{15}N, ce qui permet de mesurer des spectres 'empreintes' de la protéine appelés spectres HSQC. Ces spectres représentent les corrélations entre ^{15}N et ^{1}H des fonctions amides au niveau de la chaîne principale pour chaque résidu. Lors de la fixation d'un ligand, les perturbations des déplacements chimiques permettent de localiser dans la séquence quels résidus sont impliqués dans la liaison. De plus, des tests avec plusieurs concentrations en ligands permettent de dériver des affinités de liaison. Bien que le marquage de la protéine en

^{15}N représente une première étape nécessaire, le système est adapté pour mesurer l'affinité potentielle de nombreux ligands, rendant cette technique particulièrement intéressante.

- La diffusion dynamique de la lumière (DLS)

Cette technique repose sur la mesure de l'intensité de la lumière diffusée par un échantillon en fonction du temps. Les fluctuations mesurées reflètent les mouvements Browniens des molécules en solution et permettent de déterminer les différentes populations de particules dans l'échantillon. Cette technique est couramment utilisée pour mesurer la monodispersité d'un échantillon avant d'initier des essais de cristallogenèse. De plus, dans le cas de ligands conduisant à un changement significatif de la taille globale de la biomolécule (ex : oligomérisation), cette technique rapide, réservée à certains systèmes particuliers, peut se révéler utile.

- La variation de la stabilité thermique lors de la fixation d'un ligand ou 'Thermal Shift Assay' (TSA)

Cette technique repose sur la mesure de la température de demi-dénaturation ou température de fusion Tm d'une protéine en l'absence ou la présence d'un ligand potentiel (Cimmperman et al, 2008). L'addition d'une molécule qui se fixe à une protéine peut aussi bien avoir un effet stabilisateur (augmentation de Tm) qu'un effet détabilisateur (diminution de Tm). Toutefois, il est admis qu'un ligand qui se fixe spécifiquement à un site d'une protéine repliée augmente sa stabilité thermique, comme dans le cas d'un substrat dans le site actif d'une enzyme. En conséquence, un test a été mis au point pour estimer la capacité d'une molécule à modifier la température de fusion d'une protéine d'intérêt (Figure 19). La différence de température de demi-dénaturation entre la protéine seule et la protéine en présence d'un ligand potentiel est donnée par la valeur ΔTm. La technique repose sur l'utilisation d'une sonde moléculaire (ex : la sonde ANS pour 1-anilino-8-naphthalene sulfonate) dont la fluorescence est déclenchée lors de son interaction avec des zones hydrophobes de la protéine. Ainsi la fluorescence augmente lors de la dénaturation thermique de cette dernière.

Même si le TSA permet de mesurer l'affinité pour un ligand, il est rarement utilisé dans ce cas. Cette méthode est plutôt une méthode préliminaire pour identifier des ligands capables de stabiliser la macromolécule d'intérêt. En cristallographie elle peut se révèler utile pour déterminer des conditions permettant d'augmenter la cristallisabilité de la protéine. Le

TSA permet aussi de hiérarchiser les ligands à tester par cocristallisation ou trempage lorsque qu'un grand nombre de molécules potentielles sont étudiées (Chung, 2007). En effet, les molécules présentant les plus grandes valeurs de ΔTm sont les plus stabilisatrices, et donc potentiellement les plus prometteuses à conduire à un complexe protéine-ligand. Inversement, les molécules présentant les ΔTm les plus petits interagissent plus faiblement et peuvent être plus difficiles à cocristalliser. Néanmoins, ces généralités ne doivent pas faire état de règle absolue. En effet, l'échec à co-cristalliser un ligand malgré un ΔTm élevé peut par exemple s'expliquer par la présence de molécules dans la solution de cristallisation pouvant interférer avec la liaison protéine-ligand.

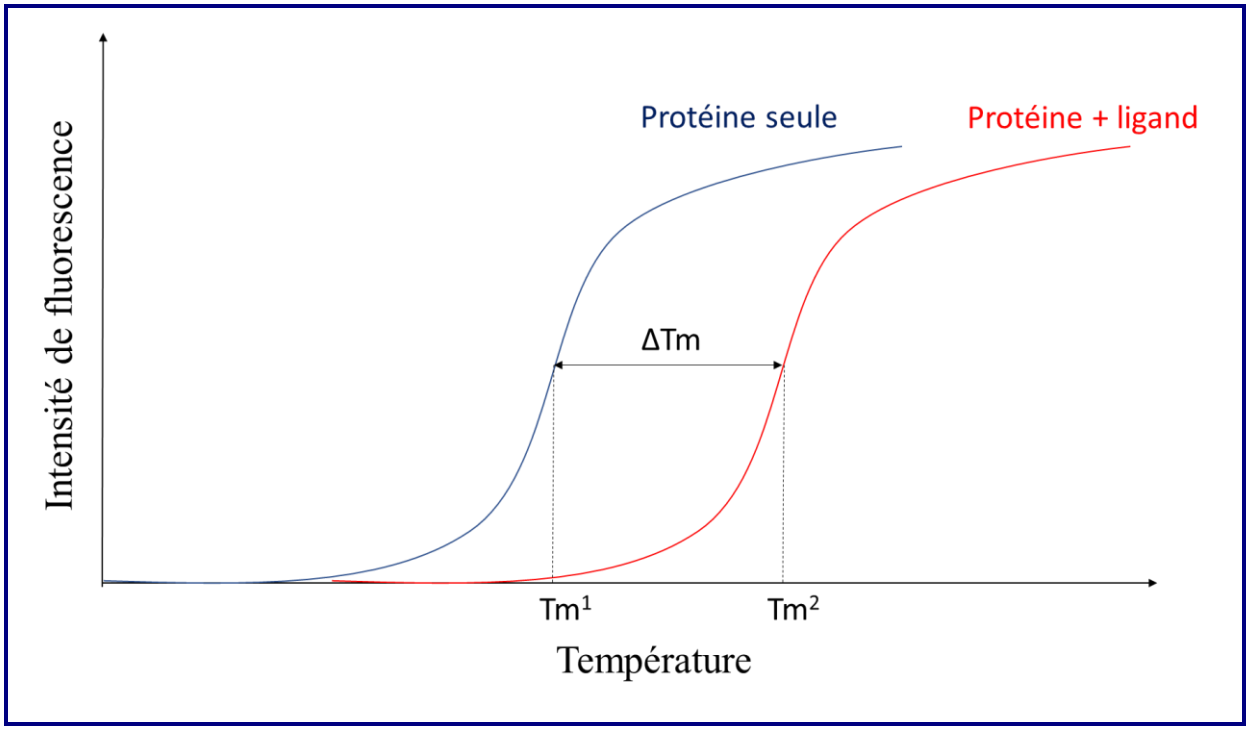


Figure 19 **Amélioration de la stabilité thermique d'un complexe protéine-ligand par rapport à la protéine seule lors d'une expérience de 'thermal shift'**
La valeur de Tm correspond au point d'inflexion de la transition entre les états natifs et dénaturés.

Le TSA présente l'avantage majeur d'être adapté au criblage d'un grand nombre de molécules. En effet, les tests peuvent être réalisés en plaque de 384 puits. Elle ne nécessite pas d'équipement coûteux puisque le seul matériel nécessaire est un thermocycleur tel qu'un appareil à PCR. De plus, le TSA est capable de détecter aussi bien des interactions fortes que faibles (K_d jusqu'à 10 mM, Cimmperman et al, 2008). Le TSA est donc utile pour fournir des points de départ notamment dans le cadre de 'drug design' par les laboratoires

pharmaceutiques. Un désavantage de la méthode est la présence potentielle de faux positifs lors d'interaction non spécifique de molécules à la protéine.

II.1.c. Stratégies d'obtention de cristaux de complexes et leur analyse

Une étape importante pour une étude structurale protéine-ligand est l'introduction du ligand dans le système. En cas de difficulté à produire la protéine d'intérêt, Hassell et collaborateurs (2007) suggèrent d'introduire le ligand dès l'étape d'expression protéique, ce qui peut conduire à une augmentation drastique du niveau d'expression, reporté notamment pour les récepteurs nucléaires (Hassell et al, 2007). Alternativement, le ligand peut aussi être introduit lors de l'étape de purification, ce qui peut résoudre les problèmes d'homogénéité de l'échantillon et favoriser la cristallisation. En dehors de ces cas problématiques, le ligand est généralement introduit avant ou après l'étape de cristallogenèse. Deux méthodes sont communément employées pour obtenir un complexe protéine-ligand en vue d'une étude cristallographique.

i) Technique de trempage des cristaux (soaking)

La première méthode la plus rapide et facile à mettre en place est un trempage des cristaux de protéine apo dans une solution de ligand (soaking). Cette technique peut être employée par l'ajout d'une solution de ligand à l'intérieur de la goutte ayant conduit à cristalliser la protéine apo. Un désavantage est le risque de détruire les cristaux suite à la modification des conditions physico-chimiques du milieu de cristallisation. Hassell propose pour remédier à ce problème de baigner le cristal au préalable dans 25 % de glutaraldéhyde, un agent crosslinker qui permet d'effectuer des pontages entre molécules du cristal pour le stabiliser. Comme autre solution, un cristal prélevé peut être baigné successivement dans des solutions de composition identique à celle de la liqueur mère et croissantes en concentrations de ligands de manière à optimiser la concentration à laquelle les cristaux restent stables (Müller, 2017). Le deuxième paramètre à optimiser est le temps d'incubation, habituellement de quelques heures à quelques jours et dépendant de l'affinité protéine-ligand. Un autre aspect important à prendre en compte est la solubilité du ligand. Bien souvent les ligands de nature hydrophobes sont solubilisés dans des solvants organiques, non miscibles avec les solutions aqueuses et compromettant les chances de créer un complexe protéine-ligand. Une possibilité est de solubiliser ces ligands dans une solution d'un agent précipitant ayant servi à cristalliser la protéine (ex : une solution pure de PEG, Danley, 2006).

ii) Techniques de co-cristallisation

La deuxième méthode employée pour obtenir des cristaux de complexes est appelée co-cristallisation. Dans ce cas, la protéine est incubée avec son ligand pendant un temps déterminé, puis l'expérience de cristallisation est effectuée. La différence majeure par rapport au trempage est la formation des complexes avant leur cristallisation. Il est à noter que bien souvent les conditions de cristallisation de la protéine complexée au ligand varient par rapport à celles de la protéine apo, si bien qu'une nouvelle recherche des conditions de cristallisation peut s'avérer nécessaire. Quand une co-cristallisation est envisagée dans un projet, il peut ainsi s'avérer judicieux d'inclure le ligand à l'expérience dès les premiers cribles de conditions de cristallisation. Une astuce utile pour accélérer la détermination des nouvelles conditions est l'utilisation d'un stock de micro-cristaux ('seed stock', Müller, 2017) préparé à partir de cristaux de la protéine apo (protocole détaillé en partie Matériel et Méthode, section III.2) et facilitant la cristallisation. Enfin, une nouvelle méthode récemment publiée (Gelin et al, 2015) suggère d'utiliser la co-cristallisation sèche : le ligand en solution est préalablement déposé au fond des puits des boites de cristallisation et son solvant est évaporé pour l'avoir sous forme solide. La cristallisation est ensuite mise au point de façon classique. Cette méthode est particulièrement avantageuse pour cribler plusieurs ligands potentiels, à condition que l'agent cristallisant reste le même pour tous les complexes.

iii) Détection et modélisation d'un ligand dans la densité électronique

La dernière étape d'une étude cristallographique de complexe protéine-ligand (Figure 20) est la modélisation du ligand éventuel dans le site de liaison de la protéine. Après enregistrement et traitement des données de diffraction, les coordonnées atomiques du modèle de la protéine apo sont généralement utilisées directement pour calculer une carte de densité électronique « différence » *Fo-Fc* dans le cas de cristaux isomorphes, ou après une étape de remplacement moléculaire dans le cas de cristaux non isomorphes (Carvalho et al, 2009). Une inspection minutieuse de la carte différence *Fo-Fc* doit ensuite être effectuée pour identifier un éventuel pic positif de densité correspondant à la présence du ligand. Afin de rendre la carte différence la plus informative possible, ou autrement dit la moins biaisée possible, une reconstruction complète des régions altérées dans la protéine, ainsi que l'incorporation des molécules d'eau doivent être effectuées. Alors seulement, la construction du ligand dans la densité peut débuter. Pour cela, les coordonnées atomiques du ligand au format PDB doivent être obtenues (récupérées dans la Protein Data Bank) ou construites si le ligand n'est pas

encore présent dans la PDB. De plus, un fichier de restreintes cristallographiques (fichier cif) doit être fourni au logiciel d'affinement pour guider l'optimisation de la géométrie de la petite molécule. Enfin, un contrôle final doit permettre de vérifier que le ligand se trouve dans un environnement favorable (liaisons hydrogènes, interactions hydrophobes) et valide la structure du complexe prête à être interprétée.

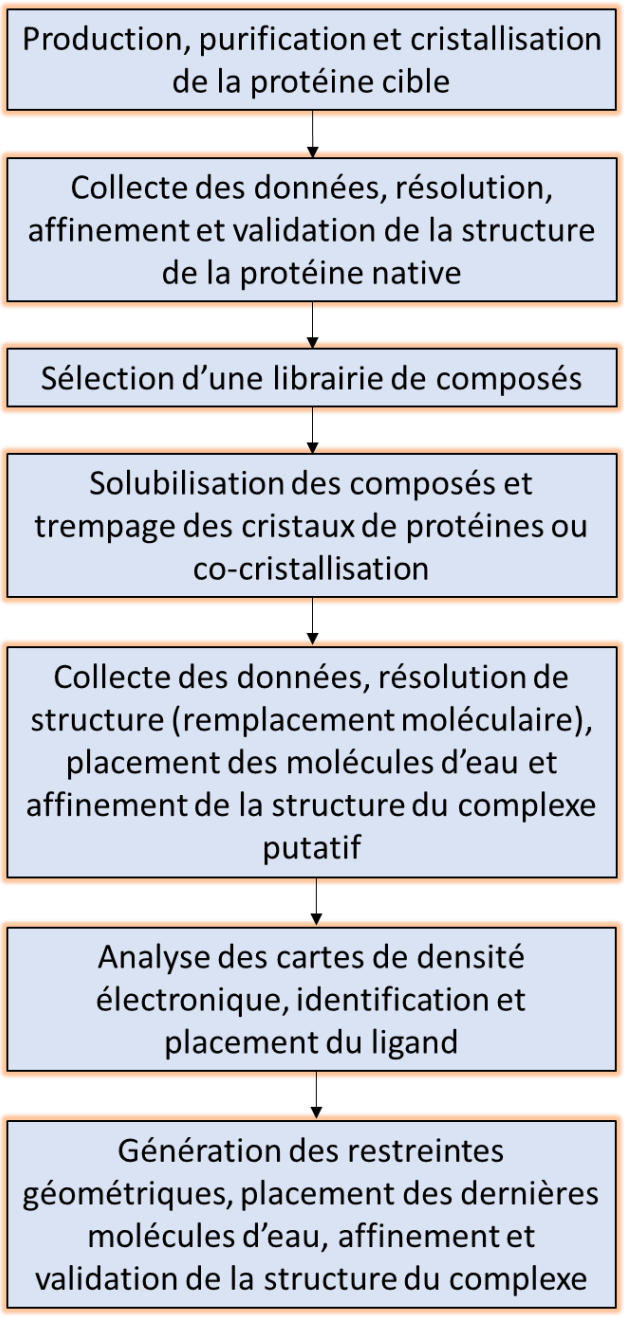


Figure 20 **Etapes clés d'un projet de criblage de ligands par cristallographie (d'après Jhoti et al, 2007)**

II.2 La cristallographie comme méthode centrale de criblage

II.2.a. Intérêt de la cristallographie comme méthode initiale de criblage

Le FBDD (pour 'Fragment-Based Drug Discovery') peut se traduire par « la méthode basée sur des fragments pour la découverte de molécules à effet thérapeutique ». Cette stratégie consiste à étudier en premier lieu la capacité d'une protéine à lier des fragments (des molécules organiques dont la masse moléculaire est généralement inférieure à 300 Da), puis à se servir de l'information structurale obtenue pour générer de nouvelles molécules de meilleure affinité. A l'origine, la preuve de concept a été avancée par Shuker et collaborateurs (1996) qui ont mis au point un nouvel inhibiteur de la protéine de liaison FK506 (qui inhibe la calcineurine) en adoptant une approche novatrice par RMN. En combinant les données structurales obtenues à partir de deux fragments différents se liant à des sites séparés, cette équipe a mis au point un inhibiteur avec une constante de dissociation de l'ordre du nanomolaire. Aujourd'hui, un large panel de techniques est utilisé en FBDD dont notamment la RMN, la SPR et la cristallographie aux rayons X.

En particulier, les avancées dans le domaine de la cristallographie ont grandement participé à l'essor du FBDD dans la recherche pharmaceutique comme en recherche académique. La cristallographie permet en effet de déterminer précisément le mode de fixation de la molécule en offrant une image en trois dimensions du complexe. Cette information est primordiale pour guider le développement de nouveaux fragments intégrant les informations obtenues des expériences précédentes. De plus, l'avantage de la cristallographie par rapport aux autres méthodes biophysiques est la détection de fragments d'affinité faible, parfois de l'ordre du millimolaire. En effet, cela permet de déterminer des molécules de départ, certes à améliorer, mais dont la détection n'aurait pas été possible par les autres méthodes (Erlanson et al, 2016), soulignant l'intérêt de la méthode dès l'étape de criblage. Certaines compagnies pharmaceutiques ont fondé leur réussite sur l'utilisation de la cristallographie dans le développement de nouveaux médicaments (ex : Astex Pharmaceuticals). Parmi les médicaments approuvés et développés grâce au FBDD, on peut citer le Vemurafenib (développé par Plexxikon) inhibant l'enzyme BRAF-V600E dans le cadre d'un mélanome, et le Venetoclax (développé par AbbVie et Genentech) inhibant BCL-2 dans le cadre de leucémie lymphocytique chronique (Erlanson et al, 2016).

Dans une étude récente, Schiebel et collaborateurs ont comparé la cristallographie à six autres méthodes biophysiques (test biochimique avec sonde fluorescente, résonance magnétique nucléaire, test de déplacement de sonde, spectrométrie de masse native, thermophorèse micro-échelle et augmentation de stabilité thermique (TSA)) dans leur capacité à prédire des ligands potentiels de l'endothiapepsine (une protéase à acide aspartique, Schiebel et al, 2016b). De manière surprenante, leurs résultats indiquent que parmi les 71 résultats positifs identifiés par cristallographie sur une banque initiale de 361 molécules, 44 % de ces « positifs » n'auraient été détectés par aucune autre méthode biophysique. La cristallographie devrait ainsi être utilisée dès les premières étapes de criblage, en parallèle des autres méthodes biophysiques. Celles-ci restent indispensables pour déterminer les affinités des ligands potentiels mais aussi car le criblage cristallographique génère beaucoup de faux négatifs (non détectés car les conditions liées au système - cristal, pH, solubilité des fragments - sont défavorables à la détection) (Schiebel et al, 2016b). Un frein majeur à cette utilisation primaire demeure le débit relativement lent des résultats issus des expériences de cristallographie (Erlanson et al, 2016). Actuellement, de nouvelles techniques émergent pour tenter de réduire ce goulot d'étranglement lié aux nombreuses collectes nécessaires d'une part, et au traitement inhérent du grand nombre de données.

II.2.b. Méthodes pour adapter la cristallographie à des besoins « haut-débit »

i) Méthode in situ facilitant la collecte de cristaux

Dans le but de réduire les manipulations et d'automatiser les collectes de données de diffraction, le Maire et collaborateurs ont présenté en 2011 une nouvelle stratégie appelée cristallographie « *in situ* » (le Maire et al, 2011). Cette méthode consiste à soumettre à un faisceau de rayons X un cristal resté dans sa liqueur mère, sans son prélèvement manuel (Figure 21). Pour cela, la boite de cristallisation utilisée doit présenter des spécificités :

- format SBS (standard des microplaques 96 puits) compatible avec la plupart des robots de manipulation.
- les puits présentent des dimensions compatibles avec l'enregistrement de données de diffraction issues d'un cristal de la boite. Idéalement, une collecte peut être planifiée sur une gamme de 80 degrés (-40 à +40 degrés, voir Figure 21).
- le matériau utilisé (copolymère d'oléfine cyclique) est choisi pour limiter l'absorption des rayons X et ainsi réduire le signal lié à la présence de la boite.

En cristallographie *in situ*, les collectes sont effectuées à température ambiante, sans cryoprotection. Les cristaux ne sont ainsi pas perturbés par l'ajout de cryoprotectant ou par le changement de température. En revanche, les radicaux libres générés par les rayons X diffusent librement dans la solution mère et peuvent potentiellement altérer les cristaux de la goutte lors de la collecte. Un autre détail à prendre en compte est la gamme angulaire restreinte accessible par cette méthode (80 degrés). Alors que cette gamme est suffisante pour un cristal présentant une haute symétrie, plusieurs collectes de cristaux peuvent s'avérer nécessaires pour obtenir un jeu de données complet dans le cas de mailles tricliniques ou monocliniques.

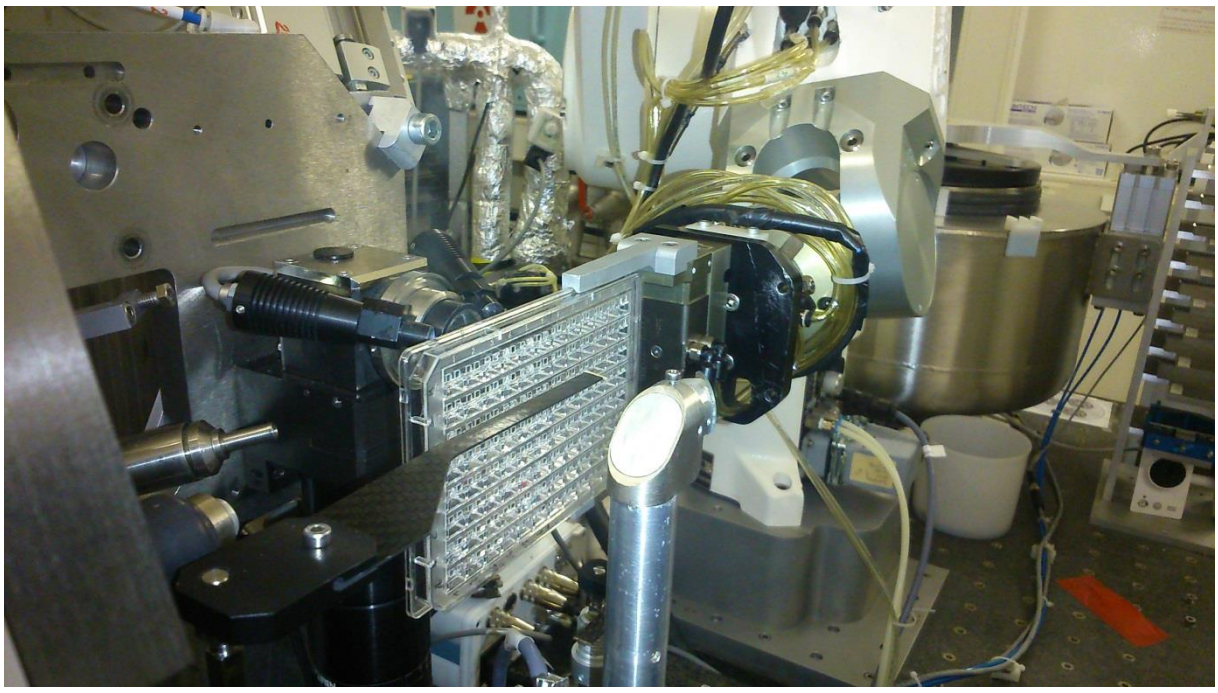


Figure 21 **Dispositif expériemental *« in situ »* présent sur la ligne BM30 de l'ESRF de Grenoble (photographie personnelle)**

Cette méthode est notamment utile à l'étape de recherches des conditions de cristallisation d'une nouvelle protéine : les cristaux de protéine peuvent être rapidement distingués des cristaux de sel par exemple. Néanmoins, le principal avantage avancé par Gelin et collaborateurs est la compatibilité de cette méthode pour la recherche de ligands par cristallographie (Gelin et al, 2015). En effet, couplée avec les méthodes de soaking ou de co-cristallisation, les cristaux peuvent être rapidement collectés et analysés en vue de la détermination structurale des complexes protéines-ligands potentiels. Les résultats issus de collectes « *in situ* » sont en effet comparables aux collectes classiques en cryoloop, bien que

les résolutions et les complétudes issues de l'expérience *in situ* soient d'une qualité inférieure. Les données collectées sont cependant suffisantes pour identifier un résultat positif potentiel.

ii) Utilisation de cocktails de molécules pour diminuer le nombre de collectes lors d'une campagne de criblage

Un usage courant en criblage de fragment par cristallographie est de tremper les cristaux de protéine dans des cocktails contenant plusieurs composés (cristallographie dite « cocktail »). Cette méthode a l'avantage principal de réduire le nombre de collectes nécessaires à une campagne de criblage de chimiothèques de plusieurs centaines de molécules. Généralement, les cocktails sont réalisés à partir de 3 à 10 molécules ayant des formes différentes, de manière à pouvoir distinguer un résultat positif potentiel dans la densité électronique (Chilingaryan et al, 2012). Les molécules sont également sélectionnées sur la base de leur solubilité (coefficient de partition cLogP <3), de leur masse (<300 Da) et le nombre de donneurs ou accepteurs de liaisons hydrogènes (<3), constituant « la règle de trois » (Congreve et al, 2003). Si une densité électronique correspondant à un ligand est identifiée, une collecte doit être réalisée avec un cristal trempé dans une solution pure du fragment identifié, ou à défaut plusieurs collectes pour chacune des molécules du cocktail si des fixations multiples sont suspectées. Cette étape est appelée « déconvolution ». Si un cristal est fragile, un plus petit nombre de fragments contenus dans les cocktails doit être utilisé, car le risque de perturber le cristal augmente avec le nombre de fragments.

Patel et collaborateurs (Patel et al, 2014) résument les différentes stratégies employées par certaines compagnies pharmaceutiques pour développer des cocktails. La compagnie Abbott Laboratories a développé une approche appelée CrystaLEADS dans laquelle une centaine de molécules différentes sont incorporées dans une solution. Cette approche a permis de développer des inhibiteurs de la dihydroneoptérine aldolase de l'ordre du micromolaire (Sanders et al, 2004). Astex a développé un système propriétaire appelé Pyramid, dans lequel le nombre de fragments est limité à quatre et l'accent est mis sur la diversité chimique des molécules d'un même cocktail. Cette approche permet une déconvolution efficace et réduit la probabilité de ligands simultanés. En contraste, l'approche employée chez Johnson and Johnson's est de réaliser des cocktails composés de molécules de formes similaires. Cette approche a l'avantage de renforcer le signal dans la densité électronique par la fixation simultanée de plusieurs ligands.

Récemment, Schiebel et collaborateurs discutent de l'utilisation des cocktails en comparaison des solutions pures de fragment, à la lumière de leurs résultats sur une campagne de criblage de 384 ligands (Schiebel et al, 2016b). Une diminution significative (environ 50 %) du nombre de molécules identifiées est constatée en utilisant des cocktails de deux molécules. Les principaux arguments avancés sont la diminution de la concentration pouvant réduire l'occupation du ligand et des interactions/réactions possibles entre les fragments.

II.2.c. La cristallographie d'affinité : identification de ligands potentiels parmi un mélange complexe tel qu'un extrait naturel

En 2002, Chen et collaborateurs ont déterminé la structure RX d'une petite molécule autoinductrice AI-2 impliquée dans le quorum-sensing chez les bactéries, en complexe avec son récepteur la protéine LuxP (Chen et al, 2002). Le jeu de données à haute résolution (1,5 Å) a permis d'identifier clairement la structure de la petite molécule dans la densité électronique (Figure 22). Ce résultat a montré notamment que AI-2 contient du bore, un élément inhabituel dans les systèmes biologiques. Ainsi, la cristallographie s'est révélé être la méthode cruciale pour révéler la nature du ligand physiologique de LuxP.

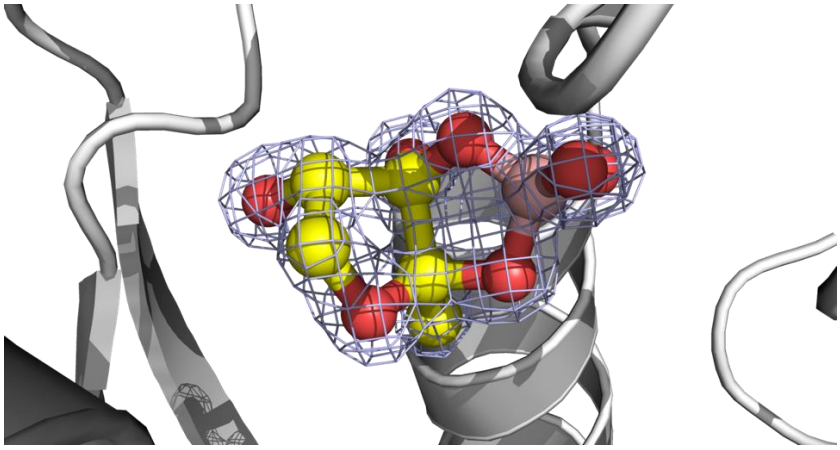


Figure 22 **Identification de la structure de l'autoinducteur AI-2 contenant un atome de bore (en rose) dans le site de liaison de la protéine LuxP (code PDB 1JX6) (carte *2Fo-Fc* contourée à 1,5 σ)**

Dans cet esprit, Aguda et collaborateurs (2016) ont présenté une méthode élégante permettant d'identifier des ligands potentiels parmi un mélange complexe, typiquement un extrait naturel (Figure 23). Cette méthode consiste à co-cristalliser (ou réaliser un trempage après cristallisation) une protéine avec un extrait naturel contenant des inhibiteurs ou ligands potentiels. Le jeu de données collectées à une résolution suffisante (<2 Å) permet d'identifier le ligand potentiel à partir de sa densité électronique au site de liaison dans la protéine. Aguda

et collaborateurs ont appliqué cette méthode à l'identification de nouveaux inhibiteurs naturels de la cathepsine K (une protéase à cystéine d'intérêt dans le traitement de l'ostéoporose) à partir d'extraits d'actinomycètes (bactéries du genre *Streptomyces* isolées d'échantillons de sol et de lichens). Cette stratégie leur a permis d'identifier une nouvelle molécule, le lichostatinal, un inhibiteur peptidique covalamment fixé à la cystéine catalytique. Pour l'identification, deux jeux de données ont été utilisés. Le premier jeu a été collecté à partir d'un co-cristal protéine-extrait brut. La densité électronique au site actif était quelque peu diffuse, reflétant une occupation incomplète du ligand. L'extrait brut a donc été purifié partiellement par l'ajout de trois étapes chromatographiques (phase inverse, échangeuse d'anion et colonne Sephadex LH 20). Un deuxième jeu a ensuite été collecté à partir d'un co-cristal protéine-extrait semi-purifié, et a révélé une densité électronique mieux définie. Après synthèse chimique de l'inhibiteur, des études cinétiques ont permis de déterminer la constante d'inhibition (K_i) du lichostatinal qui est de 11 nM. Les auteurs présentent la cristallographie d'affinité comme alternative intéressante à la méthode traditionnelle (chromatographie, résonance magnétique nucléaire, spectrométrie de masse) facilitant l'identification du ligand grâce à l'information structurale acquise (Aguda et al, 2016).

En 2017, la cristallographie d'affinité a été utilisée par Stravodimos et collaborateurs pour identifier des inhibiteurs de la glycogène phosphorylase (une cible d'intérêt dans le traitement du diabète) à partir d'extraits de grenade (*Punica granatum*). Des études précédentes leur avaient déjà permis d'identifier la quercétine à partir d'extraits de vigne *Vitis vinifera* en utilisant une approche de cristallographie d'affinité (Kantsadi et al, 2014) avant que cette méthode ne soit décrite par Aguda et al. Dans leurs travaux de 2018, leur stratégie a été de tremper les cristaux de protéine dans des extraits de grenade afin de cribler les ligands potentiels. La composition moléculaire des extraits ayant déjà été définie, la collecte des données et l'affinement leur a permis de déterminer les molécules qui se fixent au site de liaison de la glycogène phosphorylase. Ces molécules sont l'acide chlorogénique et l'acide ellagique, qui ont des affinités de l'ordre du millimolaire.

Ces exemples illustrent l'importance des données structurales dans le criblage de ligands. La cristallographie d'affinité permet d'explorer la diversité des molécules naturelles existantes d'une part et facilite d'autre part l'étape d'identification de ligands, rendant cette méthode particulièrement intéressante dans un projet de criblage de ligands.

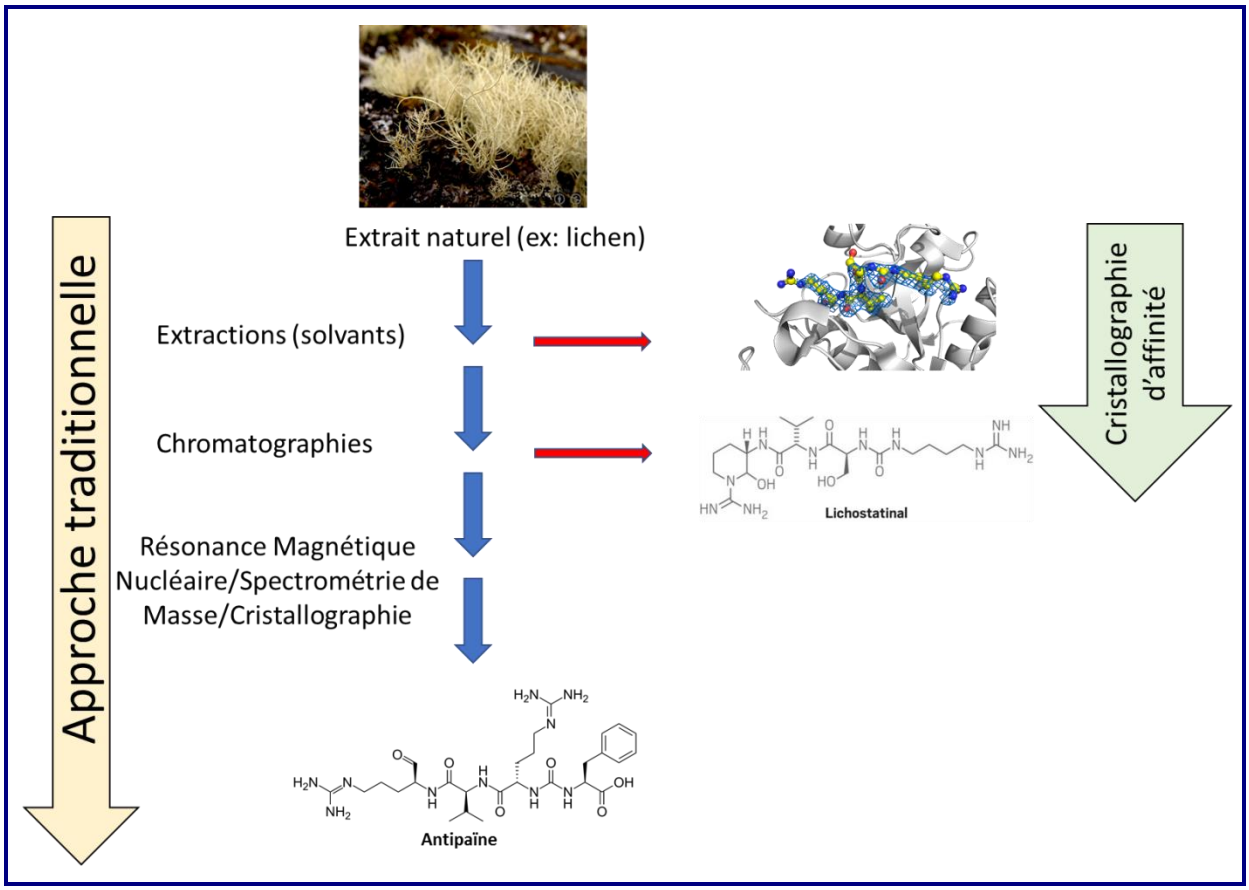


Figure 23 **Cristallographie d'affinité comme alternative à la méthode traditionnelle d'identification de ligands à partir d'extraits naturels (d'après Aguda et al, 2016)**

III Objectifs de ce travail de recherche

Alors que les GST sont bien caractérisées chez les mammifères et notamment chez l'homme, les GST fongiques demeurent largement inexplorées. Le séquençage génomique du dégradeur de bois *Trametes versicolor* a révélé que ce champignon possède seize isoformes de GST prédites dans la classe Oméga (TvGSTO) et trois isoformes appartenant à la classe Xi (TvGSTX). Par ailleurs, des études précédentes suggéraient que certaines TvGSTO présentaient une sérine catalytique (TvGSTOS) à la place d'une cystéine habituellement observée (TvGSTOC). Ces études révélaient pour ces GST des propriétés d'interaction avec des extraits de bois. Une étude approfondie des GST de *T. versicolor* s'avérait donc particulièrement intéressante. Ma participation à l'étude structure-fonction systématique pour plusieurs TvGST, au laboratoire CRM2 sous la direction des Drs. Claude Didierjean et Frédérique Favier et en collaboration avec l'équipe du Pr. Eric Gelhaye du laboratoire

Interactions Arbre-Microorganismes (IAM, UMR UL-INRA 1136), s'est articulée autour de plusieurs axes.

Le premier objectif de ce travail était de déterminer les structures de plusieurs isoformes de TvGSTO et TvGSTX de manière à obtenir des représentants structuraux pour chacune de ces deux classes, ceci dans le but de conduire une analyse structurale complète de cette famille d'enzymes. Pour cela, mes travaux ont consisté à déterminer leurs conditions de cristallisation, à effectuer les tests de diffraction puis les collectes (au laboratoire ou à l'ESRF de Grenoble) nécessaires à la résolution de leurs structures. Les structures de six TvGSTO (TvGSTO1C, TvGSTO2C, TvGSTO3C, TvGSTO2S, TvGSTO3S et TvGSTO6S) et deux TvGSTX (TvGSTX1 et TvGSTX3) ont été résolues et analysées.

Notre deuxième objectif était de déterminer les substrats et ligands ciblés par les TvGSTO. Pour cela, nous avons adopté une stratégie de recherche de ligand impliquant en particulier le criblage cristallographique d'une librairie de molécules pures, identifiées à partir de tests biochimiques et biophysiques. Mes travaux ont consisté à déterminer les conditions optimales pour obtenir des complexes GST-ligands, pour les quatre TvGSTO présentant les meilleures propriétés d'interaction (TvGSTO2C, TvGSTO2S, TvGSTO3S et TvGSTO6S). Ainsi nous avons démontré que les TvGSTOS présentent des spécificités d'interaction avec les polyphénols (flavonoïdes et benzophénones).

Le dernier objectif était d'identifier avec précision les molécules naturelles interagissant avec les TvGSTO à partir d'extraits de bois. Notre collaboration étroite avec les laboratoires IAM et LERMAB (Laboratoire de Recherche sur le Matériau Bois) et ma participation au développement d'une approche de cristallographie d'affinité ont permis d'identifier la dihydrowogonine comme ligand des GST. Ce flavonoïde est la première molécule partiellement isolée d'un extrait naturel de merisier qui se fixe au site actif de TvGSTO3S. A partir d'un extrait de bagasse, le stilbène oxyresvératrol a été identifié comme ligand de TvGSTO2S puis caractérisé structuralement. Ces résultats corroborent et précisent les précédents résultats qui suggéraient que les TvGSTOS ont des propriétés d'interaction avec les extractibles (Deroy et al, 2015).

En parallèle de l'étude des TvGST, j'ai également contribué à la caractérisation structurale de la GST ECM4 de la levure *Saccharomyces cerevisiae*, appartenant à la classe Xi et homologue des TvGSTX. Cette GST avait déjà été caractérisée biochimiquement et son

étude structurale s'avérait nécessaire puisqu'aucune structure expérimentale d'ScECM4 n'avait encore été déterminée chez cet organisme eucaryote modèle.

Dans le chapitre B suivant, une explication détaillée des méthodes expérimentales employées est rapportée. Le chapitre C détaille et discute les résultats obtenus sur chacune des deux classes de GST étudiées : les TvGSTO dans un premier temps, et les TvGSTX et ScECM4 dans un second temps. Enfin, une discussion générale des résultats obtenus durant cette thèse ainsi que les perspectives associées à ce travail de recherche sont présentées en chapitre D.

Matériels et Méthodes

I Introduction

L'objectif de ce travail de thèse était de déterminer les structures tridimensionnelles des glutathion transférases (GST) des classes Oméga et Xi par cristallographie et d'identifier leurs ligands par la résolution de structures de complexes. Ce chapitre décrit pas à pas les méthodes employées pour obtenir les structures étudiées au cours du projet, avec un accent sur les techniques de cristallogenèse et de co-cristallisation/trempage. Dans leur majorité, ces méthodes sont classiques et très bien décrites dans la littérature (soit dans des livres, soit dans des revues). Néanmoins, il a fallu en adapter quelques unes pour obtenir certaines des structures de GST étudiées en complexe avec des ligands. J'ai fait le choix de lister l'ensemble des méthodes avec parfois une description succincte de celles-ci. Les protocoles détaillés correspondant à chacun des résultats obtenus ne figurent donc pas dans ce chapitre, mais pourront être retrouvés dans les parties « Matériels et Méthodes » des publications correspondantes et jointes à ce manuscrit.

La détermination d'une structure de macromolécule par diffraction des rayons X se fait en plusieurs étapes. Tout d'abord, la molécule biologique doit être obtenue en quantité suffisante (quelques milligrammes) et de manière pure et homogène. Dans le cas des GST, l'expression chez *E. coli* s'est produite de manière aisée en quantité suffisante. La protéine doit ensuite être cristallisée grâce à l'utilisation d'agents précipitants appropriés. Puis, un monocristal est exposé à un faisceau de rayons X pour acquérir son signal de diffraction. Enfin, l'analyse des intensités diffractées permet d'obtenir des cartes de densité électronique à condition que le problème de la phase soit résolu, puis de remonter à la structure de la molécule (méthode détaillée retrouvée dans l'ouvrage « Crystals, X-rays and Proteins », de Sherwood et Cooper). Chacune de ces étapes peut être source de difficultés, voire d'échecs. Aujourd'hui encore, la capacité à cristalliser une protéine ou l'obtention de phases initiales sont des étapes qui restent des goulots d'étranglement dans la détermination d'une structure cristallographique.

Ce chapitre présente la démarche méthodologique complète qui a été employée pour produire des cristaux des glutathion transférases étudiées et déterminer leurs structures tridimensionnelles à partir de leurs figures de diffraction (Figure 24). Les aspects méthodologiques qui ont permis d'obtenir des complexes GST-ligand sont également

détaillés. Les méthodes bio-informatiques, incluant les analyses de séquence et la modélisation moléculaire sont présentées en fin de chapitre.

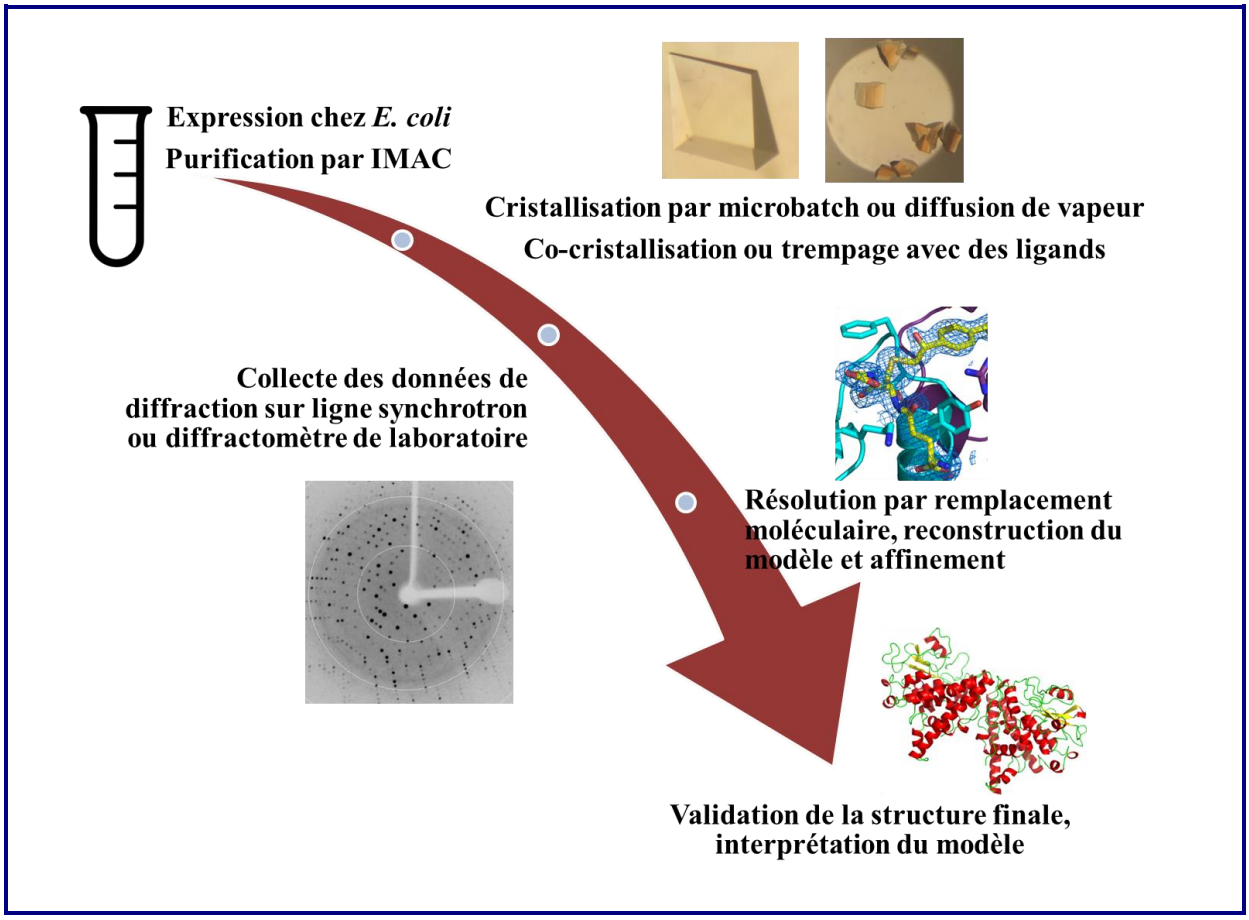


Figure 24 **Etapes successives du projet d'étude cristallographique des GST fongiques**

II Clonage, production et purification des GST

La démarche employée par nos collaborateurs du laboratoire IAM pour exprimer et purifier des GST recombinantes est explicitée brièvement dans les lignes suivantes (méthode générale détaillée dans l'ouvrage « Heterologous Gene Expression in *E.coli* » de Xu et Evans). Les plasmides recombinants pET 26b portant les gènes GST d'intérêt entre les sites de restriction NdeI et NotI ont été obtenus sous forme de gènes synthétiques. Une étiquette hexa-histidine a été ajoutée en C-terminal des séquences pour une purification par affinité sur ions métalliques immobilisés ('IMAC', pour 'immobilized metal affinity chromatography'). L'expression des protéines est effectuée à 37 °C en souches d'*Escherichia coli* Rosetta2 (DE3) pLysS (Novagen) en milieu LB (Luria-Bertani) contenant 50 µg/mL de kanamycine et

34 µg/mL de chloramphenicol. Lorsque la culture cellulaire atteint une densité optique de 0,7, l'expression des protéines est induite par l'ajout de 0,1 mM d'IPTG (Isopropyl-β-D-1-thiogalactopyranoside) et les cellules sont cultivées 4 heures supplémentaires. Les cellules sont ensuite récoltées par centrifugation et re-suspendues dans le tampon de lyse (30 mM Tris-HCl, pH 8,0, 200 mM NaCl), puis stockées à -20 °C. La lyse des cellules est effectuée par sonication (3 cycles d'une minute espacés d'une minute). L'extrait cellulaire est ensuite centrifugé à 35 000 g (25 minutes, 4 °C) afin d'enlever les débris cellulaires et les protéines agrégées. 10 mM d'imidazole sont ajoutés dans la fraction soluble puis celle-ci est déposée sur résine Ni-NTA (Ni^{2+}-nitrilotriacétate)-agarose connectée à un système FPLC de purification ÄKTA (GE Healthcare). Après une étape de lavage avec le tampon de lyse supplémenté de 20 mM d'imidazole, les protéines sont éluées en utilisant un gradient d'imidazole de 20 à 500 mM. Les fractions d'intérêt sont réunies, dialysées, concentrées par ultrafiltration sous pression d'azote (membrane YM10, Amicon), et stockées dans un tampon Tris-HCl, pH 8,0, 200 mM NaCl à -20 °C. L'homogénéité des échantillons est vérifiée par gel SDS-PAGE (Figure 25). La liste des protéines qui ont été étudiées dans le cadre de ce projet est fournie dans le Tableau 1.

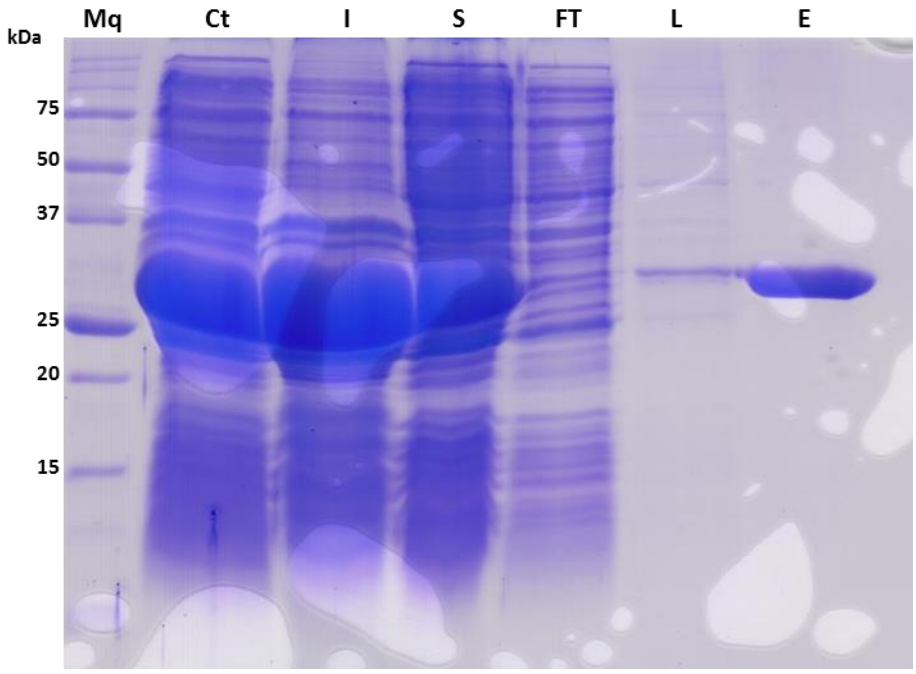


Figure 25 **Gel SDS-PAGE de séparation des protéines après purification de la TvGSTO1S**
Mq : marqueur de taille, Ct : culture totale, FT : fraction non retenue sur la colonne, S : fraction soluble, I : fraction insoluble, L : éluat après lavage, E : protéine éluée.

Organisme	Protéine étudiée	N° d'accession	Classe de GST
Trametes versicolor	TvGSTO1S	Tv75639	Oméga à sérine
	TvGSTO2S	Tv56280	
	TvGSTO3S	Tv48691	
	TvGSTO4S	Tv65402	
	TvGSTO5S	Tv54358	
	TvGSTO6S	Tv23671	
	TvGSTO1C	Tv157166	Oméga à cystéine
	TvGSTO2C	Tv67635	
	TvGSTO3C	Tv60698	
	TvGSTX1	Tv66368	Xi
	TvGSTX3	Tv73942	
Saccharomyces cerevisiae	ScECM4	NP_013002	

Tableau 1 **Glutathion transférases étudiées au cours de ce projet**
Les numéros d'accession donnés correspondent à la base de données de séquences protéiques du JGI pour *T. versicolor* et à la base de données de séquences protéiques du NCBI pour *S. cerevisiae.*

III Cristallogenèse des GST

Cette étape est nécessaire pour la détermination d'une structure cristallographique d'une macromolécule biologique. Pour cela, l'échantillon de protéine est mis en contact avec l'agent précipitant (composé souvent d'un mélange de plusieurs molécules) qui va induire un état particulier dans lequel les macromolécules biologiques sont insolubles mais forment un réseau régulier : le cristal. La détermination des conditions physicochimiques favorables à la nucléation et à la croissance de cristaux se fait le plus souvent par la méthode de la matrice d'échantillonnage. Cette stratégie « essai-erreur » consiste à tester un grand nombre de conditions différant par leur composition (en solvant, pH, tampon, sels, additifs). Ces solutions sont réunies dans des kits commerciaux développés à partir de conditions ayant précédemment provoqué la cristallisation d'autres protéines. Les kits commerciaux qui ont été utilisés sont les screens I et II de Hampton (https://www.hamptonresearch.com/), les screens I à IV de Wizard (https://www.moleculardimensions.com/) et les screens JBS 1 à 10 de Jena BioSciences (https://www.jenabioscience.com/). A l'issue de cette première étape d'exploration de conditions variées, les résultats les plus encourageants sont retenus et optimisés par une démarche rationnelle variant les différents paramètres graduellement et isolément.

III.1 Techniques de cristallisation

Après les étapes de production et de purification effectuées par nos collaborateurs, les GST sont débarassées du NaCl qu'elles contiennent par dialyse dans un tampon 30 mM Tris-HCl, pH 8,0, 1 mM EDTA sur cellules Amicon 10 kDa (Millipore) par 5 cycles de centrifugation (3 minutes, 4000 g). A ce stade, l'échantillon est prêt pour la recherche des conditions de cristallisation.

Deux méthodes ont été utilisées pour générer des cristaux : la méthode batch et la méthode de diffusion de vapeur. Dans chaque cas, les screens ont été réalisés à la main ou à l'aide d'un robot de pipetage Oryx8 (Douglas Instruments). L'utilisation d'un robot de pipetage permet l'utilisation d'un volume réduit de solution protéique (typiquement, gouttes de 100 nL au lieu d'1 µL en mode manuel) tout en gagnant en reproductibilité. En méthode batch, la solution protéique et la solution de cristallisation sont déposées au fond d'un puits, clos par l'ajout d'huile de paraffine (Figure 26). Les essais à la main sont réalisés en plaques microbatch 72 puits (Type Terasaki, Greiner) puis centrifugés 1 minute à 3000 g, 4 °C pour assurer le mélange des deux gouttes.

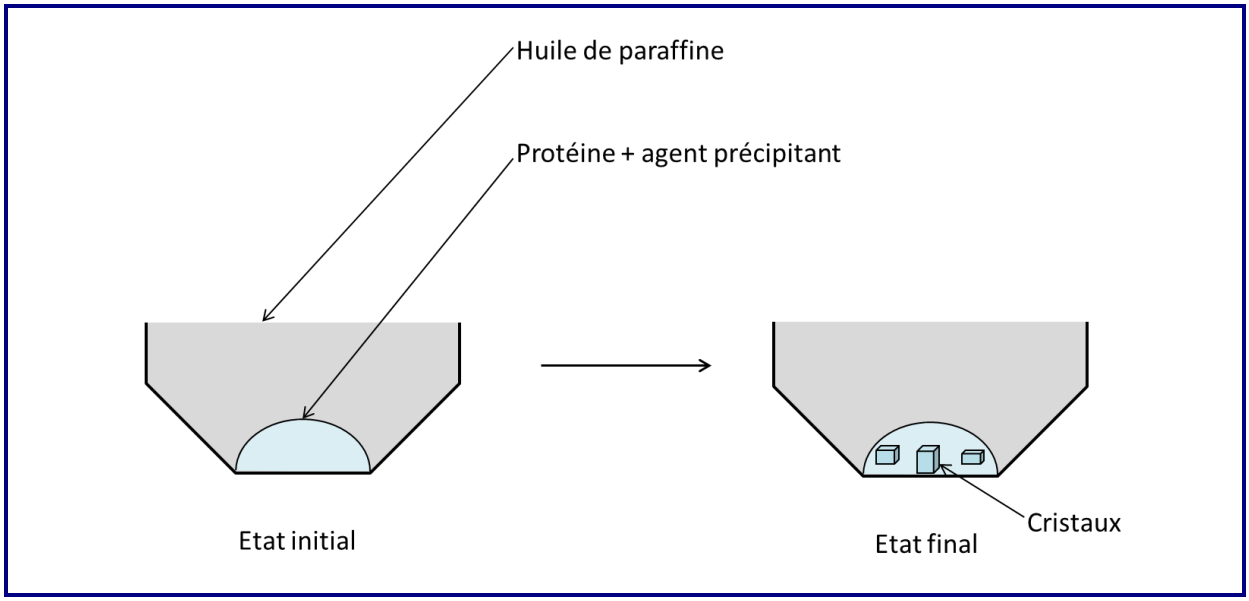


Figure 26 Schéma montrant la cristallisation en méthode batch à l'huile

La méthode de diffusion de vapeur a également été utilisée avec le robot, à l'étape de criblage des conditions de cristallisation, au cas où les protéines ne présentent aucune piste prometteuse à l'issue de la méthode batch. Les plaques utilisées sont des SwissCi 2-gouttes 96 puits (Molecular Dimensions) en goutte assise. La diffusion de vapeur a également été utilisée manuellement en plaque Linbro 24 puits (Molecular Dimensions) en goutte suspendue pour

l'étape d'optimisation et de production de cristaux en plus gros volumes (Figure 27). Brièvement, le système est constitué d'un réservoir (partiellement rempli par un volume d'agent précipitant) recouvert hermétiquement par une lamelle graissée sous laquelle est déposé un mélange équivolumique de la protéine et de l'agent précipitant, ainsi moins concentré que dans le puits. Cette différence de concentration génère des échanges de vapeur entre réservoir et goutte de cristallisation, a contrario du batch où le système est clos. Pour les deux méthodes, les boites sont placées à 4 ou 20 °C et la cristallisation est ensuite surveillée par des observations régulières à la loupe binoculaire.

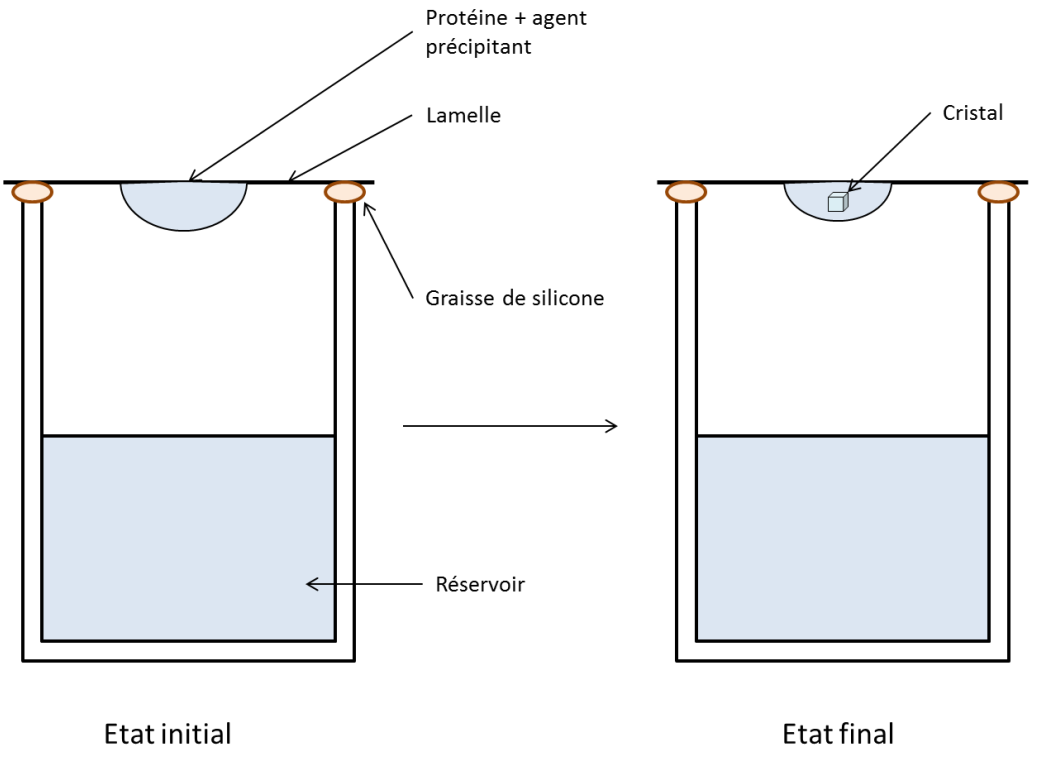


Figure 27 **Schéma montrant la cristallisation en méthode de diffusion de vapeur en « goutte suspendue »**

A partir des résultats des premières expériences, les conditions initiales sont optimisées en variant les ratios protéine/agent précipitant, leurs concentrations, la température (4 °C ou 20 °C), la présence d'additifs et éventuellement la méthode (diffusion de vapeur ou batch). Des « Grids » peuvent également être mis en place, c'est-à-dire une expérience permettant l'exploration des conditions en faisant varier graduellement deux paramètres (typiquement le pH et la concentration en agent précipitant).

III.2 Technique de micro-ensemencement

L'ensemencement est une technique d'optimisation qui consiste à préparer un stock de germes microcristallins à partir de cristaux obtenus initialement, impropres à une mesure dans l'état (trop petits, polycristallins). Si des germes sont introduits dans une goutte de cristallisation, le niveau de sursaturation requis pour provoquer la nucléation et la croissance cristalline ultérieure est bien moindre, augmentant la probabilité de cristallisation (D'arcy et al, 2007 ; Bergfors, 2003). Ainsi, la stratégie peut être d'ajouter ces germes en tant qu'additif dès les premiers tests de cristallisation, ou de les utiliser lors de l'étape d'optimisation, en combinaison d'un Grid par exemple. Cette approche s'est révélée fructueuse pour la cristallisation de TvGSTO2S et TvGSTO6S. Ces protéines produisaient des cristaux de manière peu reproductibles. L'utilisation de germes a permis de provoquer leur cristallisation de manière quasi systématique, augmentant considérablement la reproductibilité expérimentale. Pour réaliser un stock de germes, le kit Seed Bead de Hampton est utilisé (Hampton Research). Les étapes sont :

- Broyage des cristaux initiaux en germes à l'aide du concasseur
- Récupération des germes à l'aide d'une pipette et resuspension dans la solution de cristallisation (20 µL environ, 5 cycles)
- Homogénéisation 1 minute par vortex à vitesse minimale avec de la glace
- Préparation de différentes dilutions (1/10, 1/100, 1/1000)
- Ajout dans la goutte de cristallisation (environ 10 % du volume de la goutte)
- Stockage des germes à -20 °C pour réutilisation ultérieure

III.3 Interprétation du phénomène de cristallisation à l'aide du diagramme de phase

Le diagramme de phase d'une protéine explique l'état des molécules lors d'une expérience de cristallogenèse (Figure 28). Dans le cas du batch, le mélange doit placer les macromolécules dans la zone de nucléation, où des germes microcristallins peuvent se former. Si les conditions sont réunies, les molécules de protéine s'assemblent de manière ordonnée en nucléons, faisant diminuer la concentration en protéine. Les cristaux grandissent ensuite au niveau de la zone de croissance jusqu'à leur taille maximale : le système est alors à l'équilibre. Dans le cas de la diffusion de vapeur, et quelle que soit la technique employée (goutte suspendue ou goutte assise), le système est plus complexe car constitué de deux

compartiments. Au début de l'expérience, la goutte contenant protéine et agent précipitant est plus faiblement concentrée en agent précipitant que le réservoir. Les échanges de vapeur entre les deux compartiments vont entrainer une concentration des macromolécules dans la goutte. En effet, l'eau s'évapore du compartiment le moins concentré (la goutte) vers le réservoir plus concentré afin d'équilibrer les concentrations en agent précipitant au cours du temps, et également augmenter la concentration en protéine de la goutte. Cette dynamique a pour objectif d'entrainer la formation de germes microcristallins dans la zone de nucléation si les conditions sont adaptées, puis de faire grandir les cristaux au niveau de la zone de croissance. L'utilisation du micro-ensemencement permet d'introduire des germes micro-cristallins dès le début de l'expérience, ce qui peut faciliter la détermination de nouvelles conditions de cristallisation, sous réserve que les concentrations ne conduisent pas à la fonte immédiate des germes (sous-saturation).

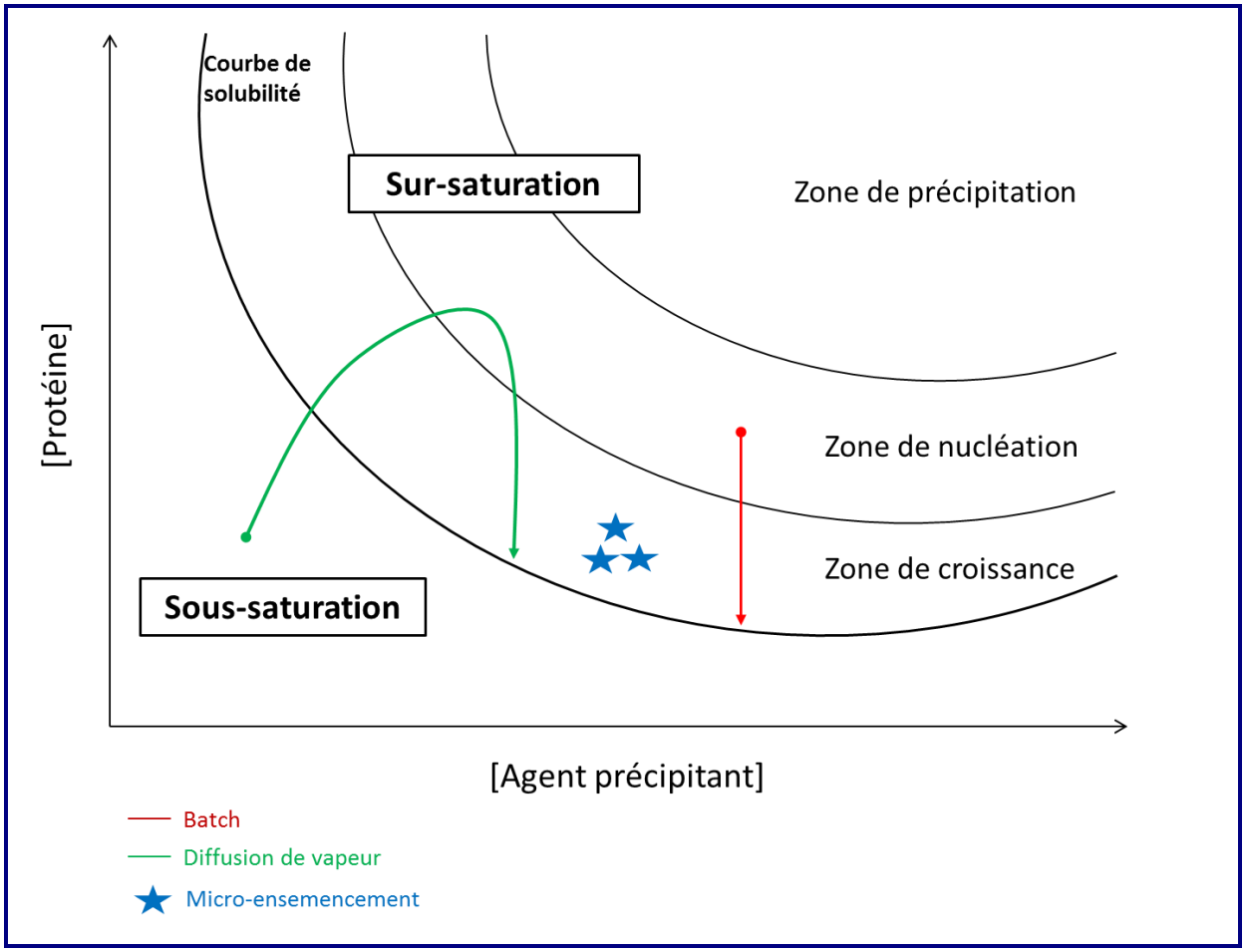


Figure 28 **Diagramme de phase de la cristallisation d'une protéine suivant la méthode utilisée**

IV Préparation des cristaux de complexes GST-ligand

Un des objectifs de ce projet était d'identifier des ligands naturels de GST, à partir de banques de molécules (Tableau 2). Afin de confirmer les résultats biochimiques de nos collaborateurs, incluant tests par inhibition d'activité enzymatique et tests de décalage thermique (ou TSA pour 'thermal shift assay'), l'ensemble des ligands potentiels ont été investigués par cristallographie de manière systématique dans le but de déterminer la structure du complexe produit.

Type de composé	Ligand
Dérivés du glutathion	GS-hexyl
	GS-dinitrobenzène
	GS-phenylacétophénone
	GS-ménadione
	GS-phenéthyl isothiocyanate
	GS-benzyl isothiocyanate
Hydroxy-benzophénones	Benzophenone
	2,4-hydroxybenzophenone
	2,3,4-hydroxybenzophenone
	2,4,4'-hydroxybenzophenone
	3,4-hydroxybenzophenone
	4,4'-hydroxybenzophenone
	2,2'-hydroxybenzophenone
	2,2',4,4'-hydroxybenzophenone
Flavonoïdes	Chrysine
	Wogonine
	Wogonoside
	Naringénine
	Naringine
	Quercétine
	Scutellaréine
	Catéchine
	Pinocembrine
	Baicaléine
Stilbènes	Oxyresvératrol

Tableau 2 Librairie de ligands testés cristallographiquement au cours de ce projet

Plusieurs méthodes existent pour obtenir une structure protéine-ligand. Si elles sont relativement simples à mettre en place, leur réussite est en grande partie conditionnée par la connaissance préalable de l'affinité existant entre la protéine et le ligand, et la solubilité du

ligand en milieu aqueux. Au cours de ce projet, différentes méthodes ont été employées pour aboutir avec succès à l'obtention d'un complexe : le trempage a été testé systématiquement dans un premier temps pour sa simplicité de mise en œuvre, puis la co-cristallisation (liquide ou sèche) a été tentée pour les ligands plus difficiles à solubiliser en milieu aqueux. Enfin, une technique innovante de « trempage sec » a été utilisée, permettant une solubilisation partielle des ligands dans la solution mère de cristallisation et particulièrement adaptée à un projet de criblage.

IV.1 Technique de trempage classique

La technique de trempage consiste à incuber un cristal de protéine dans sa liqueur mère supplémentée d'une concentration maximale de ligand afin d'obtenir une occupation la plus élevée possible du site de fixation (au sens cristallographique), tout en ne perturbant pas le cristal incubé (Schiebel et al, 2016b). Pour cela, une concentration d'environ 10 mM s'est révélée adaptée pour la plupart des cas. En pratique, la stratégie a été de produire des cristaux en batch de manière classique, puis de faire diffuser précautionneusement le ligand dans la goutte contenant les cristaux (habituellement 0,2 µL soit 10 % de la goutte initiale à une concentration stock de 100 mM, soit environ 9 mM dans la goutte).

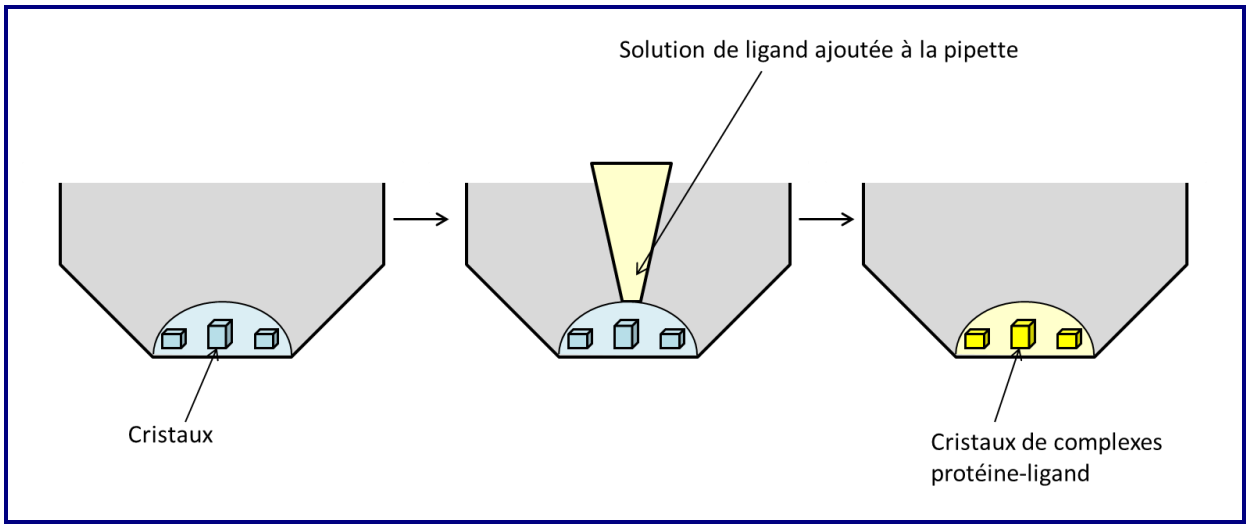


Figure 29 **Schéma montrant la mise en œuvre d'une expérience de trempage de cristaux dans une solution contenant un ligand coloré**

Cette stratégie présente l'avantage d'éviter la manipulation des cristaux hors de leur liqueur mère initiale, tout en étant pratique pour tester différents volumes de ligand jusqu'à trouver le volume maximal ne faisant pas fondre les cristaux. La technique de trempage implique la solubilisation du ligand dans la liqueur mère qui est une solution aqueuse dans la plupart des cas. Cette technique a été utilisée pour les ligands solubles dans l'eau, en

particulier le glutathion ou les molécules glutathionylées, solubilisés facilement à des concentrations de l'ordre du millimolaire en milieu aqueux. Dans le cas d'un ligand coloré présentant une longueur d'onde d'absorption dans le spectre visible, la concentration de ce ligand dans le cristal peut conduire à obtenir un cristal coloré (Figure 29).

IV.2 Techniques de co-cristallisation

En cas d'échec de la première technique, la seconde stratégie est la co-cristallisation. Cette technique consiste à cristalliser la protéine en présence de son ligand fixé préalablement par une incubation de quelques minutes à quelques heures (Figure 30). Cette méthode présente l'avantage de provoquer la rencontre entre molécules de protéine et ligand avant l'étape de cristallisation. En particulier, cette approche est appropriée pour les ligands hydrophobes dont la solubilisation en milieu aqueux est insuffisante à l'obtention de complexes où l'occupation cristallographique des sites de ces ligands est totale. En pratique, une incubation de 30 minutes est effectuée entre l'échantillon de protéine et le ligand (notamment flavonoïdes et hydroxy-benzophénones) à une concentration de 10 mM (stock 100 mM dans du DMSO). En revanche, la condition de cristallisation n'est plus nécessairement celle de la protéine native puisque la composition de la goutte est légèrement modifiée par la présence du ligand et de son solvant. Une recherche des nouvelles conditions de cristallisation peut s'avérer nécessaire. Dans la mesure où cette démarche est coûteuse en protéine et en temps, et donc limitante dans le cadre d'une étude haut-débit où de nombreux ligands sont testés, d'autres pistes de travail nous ont semblé importantes à explorer.

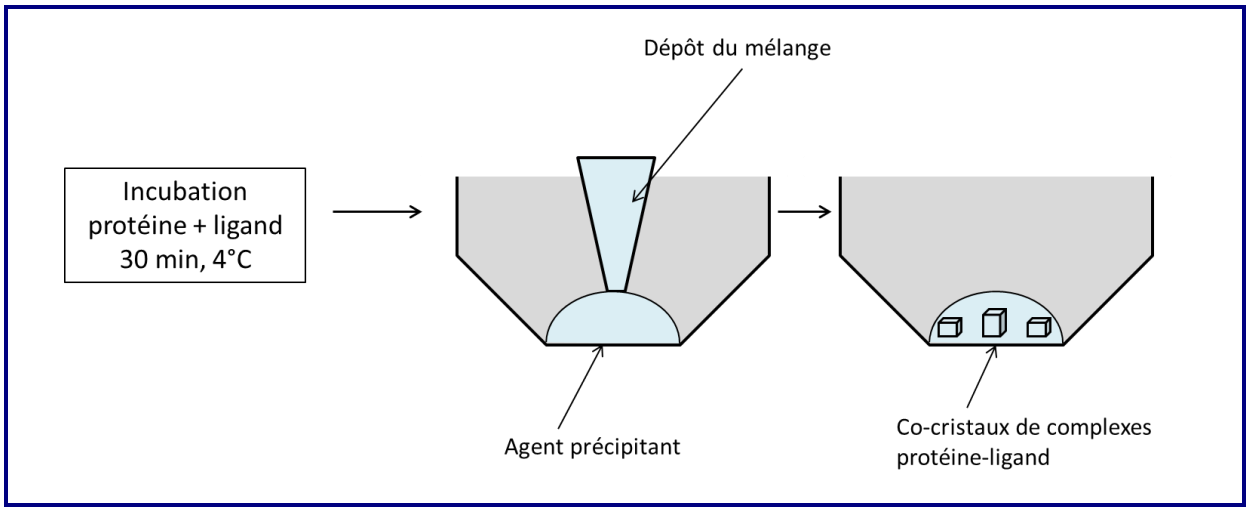


Figure 30 Schéma montrant la technique de co-cristallisation liquide

Alternativement, une co-cristallisation « sèche » peut être tentée : le ligand solubilisé dans du DMSO est déposé au fond du puits et après quelques jours d'évaporation la goutte de cristallisation est réalisée sur le ligand solide (Figure 31). Le ligand se solubilise ainsi partiellement dans la solution mère. Cette méthode est particulièrement appropriée dans le cadre d'une étude haut-débit puisque plusieurs tests de ligands peuvent être mis en place avec une même condition de cristallisation (Gelin et al, 2015).

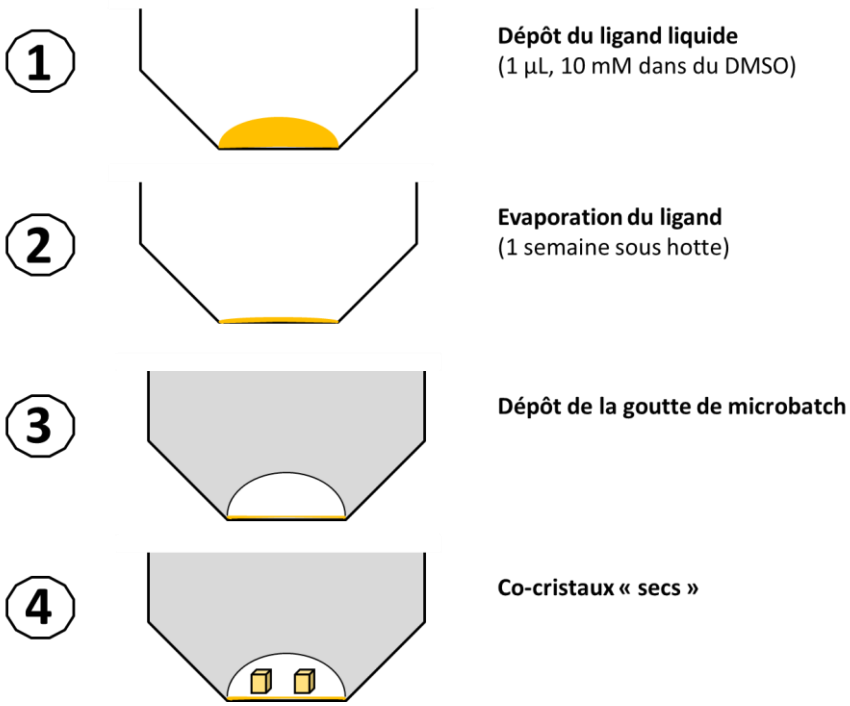


Figure 31 **Schéma montrant les étapes successives pour effectuer une co-cristallisation « sèche »**

IV.3 Technique de « trempage à sec »

La dernière technique utilisée a été mise au point pour l'obtention du complexe TvGSTO6S-naringénine. En effet, les cristaux ne supportaient pas la diffusion de ligands dans la goutte de cristallisation, et aucune stratégie de co-cristallisation (ni sèche, ni liquide) n'avait permis l'obtention de cristaux à partir des conditions de cristallisation initiales. Afin d'éviter une recherche de nouvelles conditions de cristallisation, une stratégie de « diffusion sèche » a été mise au point (Figure 32). Pour cela, 0,2 µL du ligand dans son solvant (ici la naringénine à 100 mM dans du DMSO) sont déposés sur une lamelle et évaporés (environ 1 jour sous hotte). Un cristal ainsi que 1 µL de sa liqueur mère sont déposés sur le ligand séché. La lamelle est ensuite scellée au-dessus du réservoir contenant l'agent précipitant, de manière à éviter l'évaporation totale de la goutte. Le ligand est ainsi progressivement solubilisé dans la

liqueur mère et peut diffuser dans le cristal. Un jour après, le cristal peut être prélevé et congelé. Ce procédé a également été utilisé pour la méthode du microbatch, pour les cristaux de TvGSTO2C en complexe avec des polyphénols (naringénine et oxyresvératrol). Dans ce cas, la goutte est close par l'ajout d'huile de paraffine pour éviter son évaporation.

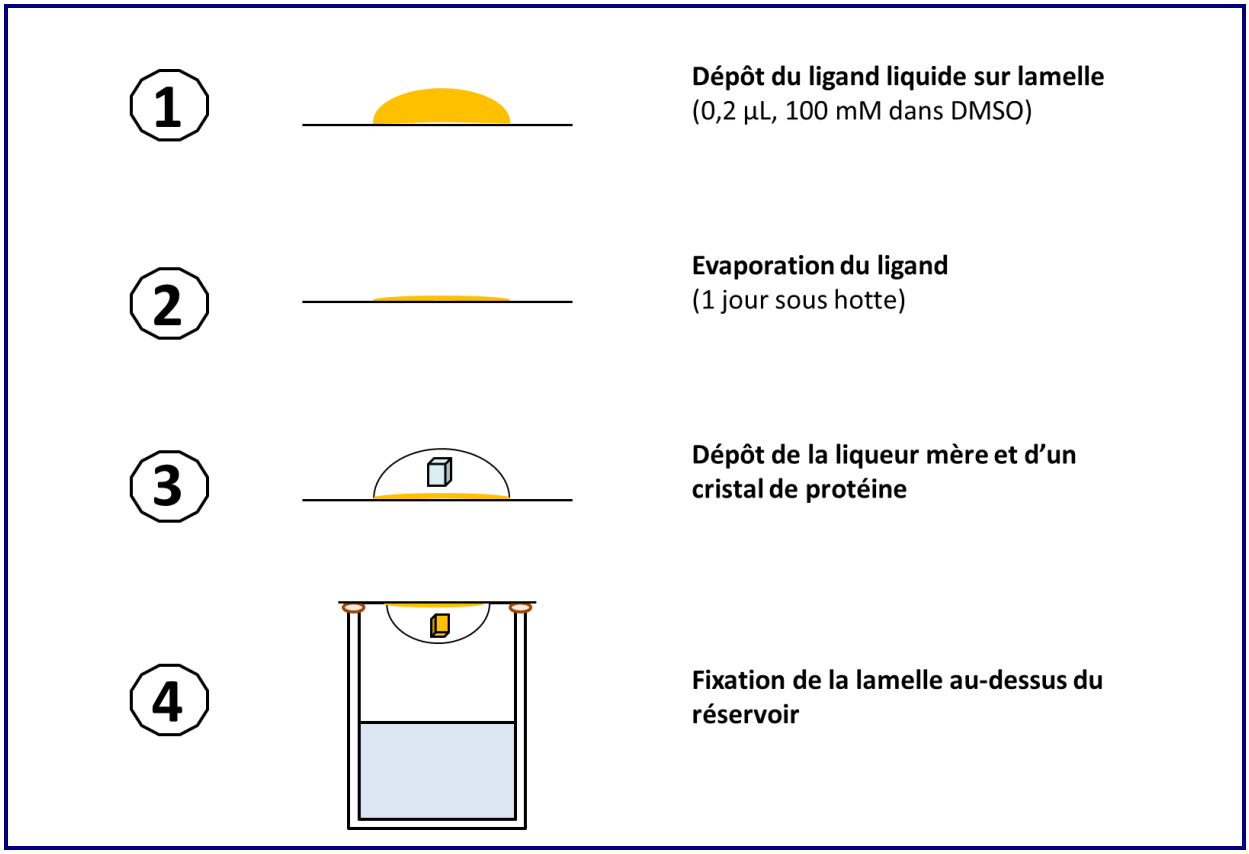


Figure 32 **Schéma montrant les étapes successives pour effectuer un trempage « à sec »**

V Congélation des cristaux, tests de diffraction et collecte des données de diffraction

La prochaine étape après l'obtention de cristaux de protéines consiste à tester le pouvoir diffractant de ces cristaux soumis à un rayonnement X. Ce paramètre est fondamental pour obtenir des données exploitables permettant la construction d'un modèle. Cette étape permet également de vérifier que les cristaux obtenus sont bien des cristaux protéiques, et non des cristaux de sels qui présentent des clichés de diffraction caractéristiques.

V.1 Cryoprotection

Aujourd'hui, la plupart des mesures effectuées sur des cristaux macromoléculaires sont réalisées à une température de 100 K maintenue par un flux d'azote (Figure 33) : c'est la

technique de cryo-cristallographie (Pflugrath, 2004). Cela permet de limiter fortement la diffusion dans le cristal des radicaux libres induits par son irradiation et ainsi d'augmenter significativement sa durée de vie. En revanche, le refroidissement à des faibles températures provoque le plus souvent la formation de glace microcristalline autour et à l'intérieur du cristal. En effet, d'une part les cristaux de protéine ne sont jamais prélevés secs et sont entourés d'une couche de liqueur mère. D'autre part les cristaux contiennent en moyenne 50 % de solvant, dû à la présence de larges canaux entre les macromolécules. Cette glace microcristalline peut détériorer le cristal (perte de son pouvoir diffractant) et nuire au signal de diffraction mesuré (anneaux de diffraction caractéristiques de l'eau microcristalline). Sa formation doit donc être empêchée. Pour cela, le cristal est refroidi instantanément en présence d'un agent cryoprotectant qui doit permettre la formation de glace amorphe tout en conservant le pouvoir diffractant du cristal. La nature de l'agent cryoprotectant dépend de la nature de la protéine et de la condition de cristallisation. Dans le cadre de ce projet, le glycérol dilué à 20 % V/V dans la liqueur mère s'est avéré efficace. Afin de limiter le choc osmotique lors du transfert des cristaux dans la solution cryoprotectante, celle-ci est systématiquement préparée en même temps que la cristallisation. Ainsi, quelques puits de la plaque de cristallisation sont enrichis de 20 % de glycérol. Dans quelques cas, des cristaux sont apparus dans ces puits.

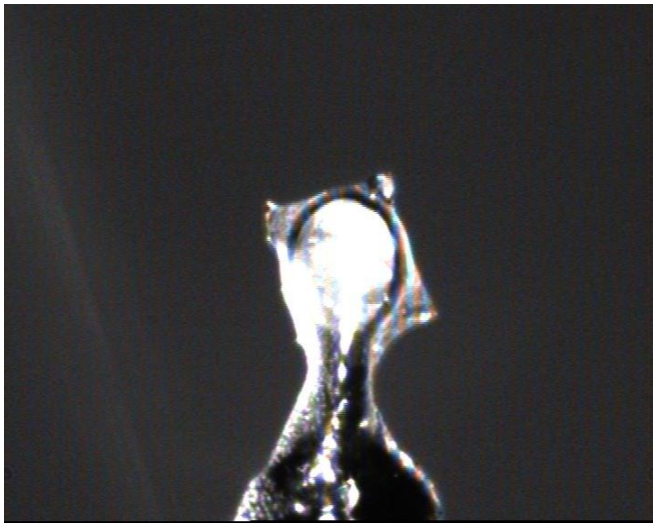


Figure 33 Cristal de ScECM4 monté sur une CryoLoop et soumis à un flux d'azote

V.2 Tests de diffraction

Les tests de diffraction ont été réalisés à l'aide du diffractomètre automatique 4 cercles (oméga, phi, kappa et thêta) SuperNova (Agilent) équipé d'une microsource au cuivre

et d'un détecteur du type CCD ('Charge Coupled Device'). Cet appareil fait partie du parc de diffractomètres de la plateforme de mesures de diffraction X gérée par le laboratoire CRM2 (Figure 34). Tous les cristaux ont été systématiquement pêchés à l'aide d'une CryoLoop (Hampton Research ou Molecular Dimensions). Un cliché de diffraction préliminaire est réalisé avec la solution cryoprotectante uniquement pour vérifier que celle-ci est bien adaptée : aucun anneau de diffraction de glace microcristalline ne doit être présent sur le cliché. Le cristal monté doit être centré sur le faisceau X grâce à une tête goniométrique. Pour un test classique, le cristal est exposé pendant 20 secondes au flux de rayons X pendant une oscillation de 1° autour de l'axe oméga du diffractomètre qui est perpendiculaire à la direction du faisceau. Un deuxième cliché est ensuite réalisé selon les mêmes paramètres avec le cristal tourné de 90° par rapport à l'axe oméga pour s'assurer de l'isotropie du signal de diffraction. Si les clichés sont satisfaisants, deux options sont possibles. Si la résolution est suffisante (au moins égale à 4 Å), une recherche de la maille peut être tentée en prenant plusieurs clichés afin d'indexer le réseau réciproque du cristal pour caractériser ses paramètres de maille. A l'issue de cette expérience, une collecte des données de diffraction peut être planifiée au laboratoire. Sinon, le cristal est conservé dans l'azote liquide pour être testé ou mesuré ultérieurement sur une ligne synchrotron où le faisceau de rayons X est plus brillant.

Figure 34 **Diffractomètre Supernova équipé d'un détecteur CCD Atlas (gauche) et dispositif expérimental de la ligne BM30A à l'ESRF de Grenoble (droite)**
Le diffractomètre provient de la société Agilent et est installé sur la plate-forme de mesures de diffraction X de l'Université de Lorraine (tiré de crm2.univ-lorraine.fr).

V.3 Collecte des données de diffraction

Classiquement au laboratoire, une stratégie de collecte est calculée par le logiciel CrysalisPro (CrysAlisPro, Oxford Diffraction/Agilent Technologies UK Ltd, Yarnton,

England) à partir des paramètres de maille obtenus par une expérience préliminaire (Figure 35). La stratégie est optimisée de manière à collecter un maximum de pics de Bragg (taches de diffraction) en un minimum de temps suivant la résolution, la complétude et la redondance désirées. La symétrie du cristal et l'estimation du pouvoir diffractant du cristal sont pris en compte dans le calcul. Pour cela, le programme CrysalisPro détermine, pour différentes orientations du cristal une plage angulaire à couvrir. Sur la base des données collectées durant l'expérience préliminaire, le programme propose l'angle d'oscillation et le temps d'exposition pour chaque image. La distance cristal-détecteur est optimisée « manuellement » pour mesurer les données à la meilleure résolution tout en ayant des taches bien séparées. Le temps d'acquisition est également ajusté. La durée d'une collecte, variable suivant la symétrie du cristal, prend en général une dizaine d'heures au plus.

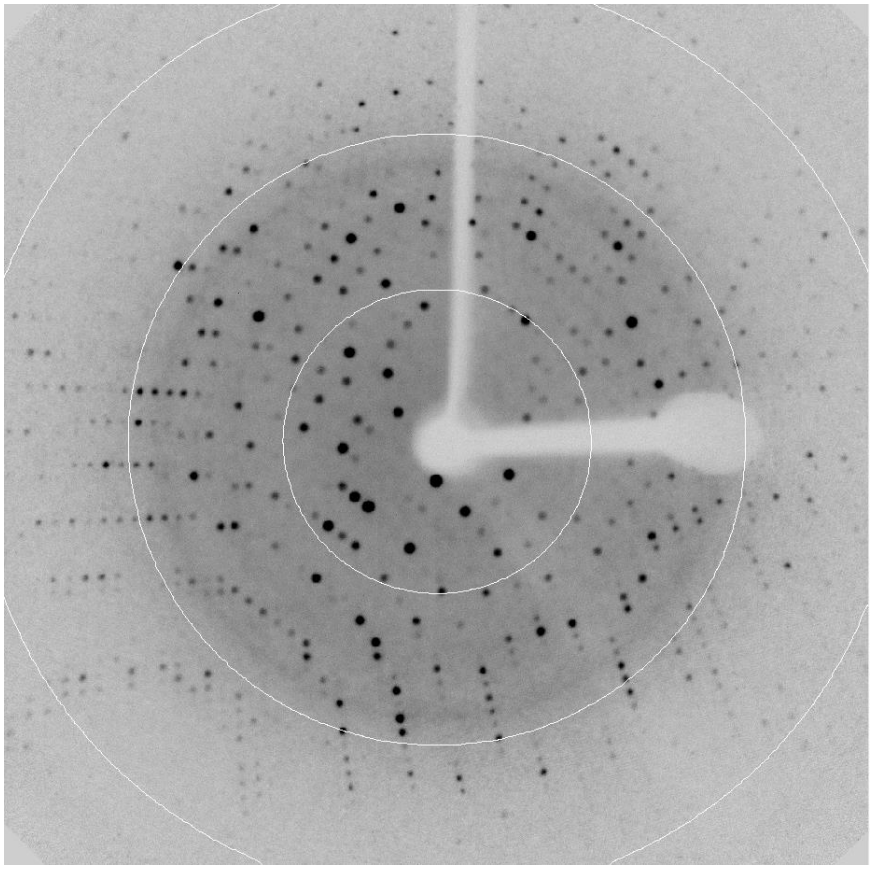


Figure 35 Cliché de diffraction d'un cristal d'ScECM4 mesuré au laboratoire

Les collectes ont été effectuées sur les lignes de lumière BM30A (Roth et al, 2002 ; Figure 34) et ID30B du synchrotron ESRF de Grenoble. Les collectes sont planifiées automatiquement selon les mêmes critères, à l'aide du logiciel MxCube pour la ligne ID30B (Gabadinho et al, 2010). La différence réside principalement en la rapidité des collectes

effectuées sur ligne synchrotron : de quelques heures à quelques secondes seulement pour les lignes les plus performantes. La brillance du faisceau permet en effet de réduire considérablement le temps d'exposition à quelques secondes par exemple sur la ligne de lumière ID30B (Mueller-Dickmann et al, 2015). Cette ligne de lumière placée à la sortie d'un onduleur est équipée d'un détecteur de rayons X Pilatus (Henrich et al, 2009) autorisant des collectes de données sans obturateur (en mode 'shutter less').

Des collectes préliminaires de plusieurs cristaux de GST ont été réalisées dans la mesure du possible au laboratoire, notamment pour les cristaux de TvGSTO3S, TvGSTO1C, TvGSTO2C, TvGSTO3C et ScECM4 qui présentaient des tailles supérieures à 200 µm dans la plus petite dimension et de forts pouvoirs diffractants. Pour l'obtention des jeux de données à haute résolution ainsi que pour les autres GST (TvGSTO4S, TvGSTO6S, TvGSTX1 et TvGSTX3) des collectes sur les lignes synchrotron ont été effectuées (Tableau 3).

Protéine		ScECM4	TvGSTX1	TvGSTX3	TvGSTO1C	TvGSTO2C
Source	Supernova	**x**			**x**	**x**
	BM30A	**x**	**x**	**x**	**x**	**x**
	ID30B	**x**				
Protéine		TvGSTO3C	TvGSTO2S	TvGSTO3S	TvGSTO4S	TvGSTO6S
Source	Supernova	**x**		**x**		
	BM30A		**x**	**x**		**x**
	ID30B	**x**		**x**	**x**	**x**

Tableau 3 Résumé des collectes effectuées et sources de rayons X utilisées

VI Traitement des données

A l'issue d'une collecte de données cristallographiques, une série de fichiers d'images correspondant aux clichés de diffraction est obtenue. La prochaine étape est alors de traiter ces images pour indexer et extraire les intensités des différents pics de Bragg mesurés (appelées également réflexions).

Différents logiciels m'ont permis d'intégrer les données au cours de ce projet : CrysalisPro (Oxford Diffraction 2003, CrysAlis CCD and CrysAlis RED, Oxford Diffraction Ltd, Wrocław, Poland) pour les collectes au laboratoire, MOSFLM de la suite CCP4 (Winn et al, 2011) et XDS (Kabsch, 2010) pour les collectes au synchrotron. Les lignes suivantes explicitent la démarche employée par XDS qui est de loin le logiciel que j'ai le plus utilisé. XDS est constitué d'une suite de programmes fonctionnant sous Linux et MacOS. A partir des fichiers images, XDS est capable d'effectuer les tâches suivantes :

- XYCORR : Application de facteurs de correction à chaque pixel pour corriger les défauts du détecteur.
- INIT : Estimation du niveau de bruit.
- COLSPOT : Identification des pics les plus intenses.
- IDXREF : Indexation des pics et premières estimations des paramètres de maille et du groupe ponctuel, de l'orientation du cristal et de la mosaïcité.
- DEFPIX : Identification des régions « utiles » sur le détecteur et mise à l'écart des régions de mauvaise qualité.
- XPLAN : Etablissement d'une stratégie de collecte à partir des informations. Cette étape est facultative mais s'est avérée particulièrement utile pour l'optimisation du temps de chaque collecte lors des mesures sur la ligne BM30A.
- INTEGRATE : Estimation de l'intensité de chaque réflexion dans les régions définies par DEFPIX.
- CORRECT : Application de facteurs de correction aux intensités des réflexions et à leurs déviations standards.

A la fin du premier tour d'intégration des données, des indicateurs statistiques évaluent la qualité du jeu de données et estiment la résolution, permettant le cas échéant d'affiner l'intégration (Wlodawer et al, 2013). Ces indicateurs sont :

- Le rapport moyen entre l'intensité et l'écart-type de la mesure (I/sig(I)).
- Le facteur R_{merge} qui évalue pour l'ensemble des réflexions, le désaccord entre l'intensité moyenne d'une réflexion *h,k,l* et les intensités de chaque mesure de la réflexion ou de ses équivalentes par symétrie.
- Le facteur R_{meas} qui est similaire au facteur R_{merge} mais corrigé pour prendre en compte la redondance.
- La complétude qui représente la quantité de pics de Bragg mesurés par rapport au maximum mesurable à une résolution donnée.
- La redondance qui correspond au nombre moyen de mesures pour chaque réflexion indépendante.
- Le facteur de corrélation de Pearson appelé $CC_{1/2}$ qui mesure la corrélation entre deux ensembles de données de diffraction divisés aléatoirement. A partir de ce facteur Karplus & Diederichs (Karplus et Diederichs, 2012) ont défini un facteur CC* comparable au facteur de corrélation entre les facteurs de structure observés et

calculés. Ainsi, une dernière tranche de résolution avec un CC* supérieur ou égal à 50 % contient de l'information utile à l'affinement du modèle.

Les réflexions sont ensuite mises à l'échelle et réduites grâce au programme XSCALE (Kabsch, 2010), qui permet également de fusionner plusieurs jeux de données. XDSCONV permet de générer un fichier de format mtz contenant l'ensemble des réflexions associées à leur intensité et déviation standard. Le programme Pointless (Evans et Murshudov, 2013) a également été utilisé pour effectuer cette étape de mise à l'échelle et de réunion des données.

VII Phasage par remplacement moléculaire

VII.1 Introduction au problème de la phase

A la fin du traitement des données, l'information liée à la structure est dissimulée dans la liste des intensités enregistrées. Pour décrire la structure de l'objet contenu dans la maille, on exploite la relation :

$$\rho(x,y,z) = \left(\frac{1}{V}\right)\sum_h\sum_k\sum_l F_{hkl}\ e^{-2i\pi(hx+ky+lz)}$$

Avec :

- $\boldsymbol{\rho}$ la densité électronique en tout point du cristal de coordonnées fractionnaires *x*, *y* et *z*.
- ***V*** le volume de la maille du cristal.
- ***h, k*** et ***l*** les coordonnées des nœuds du réseau réciproque.
- F_{hkl} le facteur de structure décrivant l'onde diffractée par la maille dans une direction définie par les indices *h*, *k* et *l*.

Néanmoins, seuls les modules $|F_{hkl}|$ sont connus à partir des intensités mesurées, alors que l'information sur la phase de ces ondes est perdue. Or pour remonter à la densité électronique, il est nécessaire de retrouver ces phases.

La méthode du remplacement isomorphe nécessite l'ajout d'atomes lourds, facilement localisables dans la maille en raison de leur facteur de diffusion élevé, pour en déduire ces phases. Une autre méthode repose sur l'utilisation de diffuseurs anomaux (incorporés dans la protéine sous forme de sélénométhionine par exemple) présentant une diffraction particulière (diffusion anomale) suivant la longueur d'onde utilisée.

En contraste de ces méthodes expérimentales dites méthodes de perturbation, la méthode du remplacement moléculaire consiste à utiliser la structure déjà déterminée d'une molécule proche de la molécule cible pour en dériver des phases initiales partiellement correctes. Le remplacement moléculaire a pour objectif de placer dans la maille du cristal à caractériser une molécule de structure similaire et connue qui sert de point de départ pour l'obtention des phases initiales. Celles-ci ne sont qu'approchées, mais suffisantes pour démarrer la reconstruction progressive du modèle.

Dans la mesure où de nombreuses structures de GST étaient déjà déposées dans la Protein Data Bank au début du projet, le remplacement moléculaire se présentait comme la méthode de choix pour résoudre le problème de la phase. De manière générale, cette méthode sera à l'avenir très certainement privilégiée par rapport aux méthodes *de novo* utilisant des dérivés lourds ou anomaux en raison du nombre croissant de structures déposées dans la PDB (Scapin, 2013). De plus, le développement de 'pipelines' (suites de programmes fonctionnant en série) permet d'une part de travailler en utilisant plusieurs « points de départs » si différents modèles sont disponibles, et d'autre part de tester les fonctionnalités de plusieurs programmes de remplacement moléculaire, de modification de modèle voire d'affinement, augmentant les chances de succès de cette méthode (Keegan et Winn, 2008).

VII.2 Coefficient de Matthews

Parallèlement à la recherche d'une protéine de structure similaire, une caractérisation du contenu du cristal étudié doit être faite. La maille élémentaire est le motif de base qui, en se répétant indéfiniment par translation, constitue le cristal. Suivant la symétrie du réseau, la maille élémentaire est subdivisée en unités asymétriques qui représentent le plus petit volume du cristal ne contenant pas d'éléments liés par symétrie cristallographique. Différents objets reliés par symétrie non cristallographique peuvent néanmoins y être présents (exemple : un homodimère relié par une symétrie axiale). Le contenu de la maille, Z_a, représentant le nombre de chaînes, peut être estimé grâce au coefficient de Matthews (V_M) exprimé par la formule suivante (Matthews, 1968 ; Kantardjieff et Rupp, 2003) :

$$V_M = \frac{V}{M.n.Z_a}$$

Avec :

- ***V*** le volume de la maille élémentaire

- ***M*** la masse moléculaire de la protéine
- ***n*** le nombre de positions équivalentes
- $\boldsymbol{Z_a}$ le nombre de molécule dans l'unité asymétrique

Sur la base de 116 cristaux de protéine, B. W. Matthews a remarqué que le coefficient V_M variait de 2 à 4 $Å^3.Da^{-1}$ correspondant à un taux de solvant d'environ 27 à 65 % en prenant 0,74 $cm^3.g^{-1}$ comme volume spécifique de la protéine. Il est ainsi possible d'estimer le nombre de molécules à rechercher et placer dans l'unité asymétrique, connaissant la masse moléculaire de la protéine et la symétrie du cristal. La résolution peut être un paramètre à prendre en compte car une statistique sur près de 16 000 structures cristallographiques a permis de montrer que les cristaux diffractant à la plus haute résolution présentent les V_M les plus petits (Kantardjieff et Rupp, 2003).

VII.3 Principe du remplacement moléculaire

Après qu'une structure de protéine similaire a été choisie (idéalement avec un pourcentage d'identité de séquence supérieur à 30 %), il peut être utile de préparer la structure modèle en « mutant » tous les résidus différant de la protéine d'intérêt par des résidus alanine. La procédure du remplacement moléculaire tente de créer un cristal virtuel de la protéine modèle similaire au cristal de la protéine étudiée (Scapin, 2013). Pour cela, les programmes de remplacement moléculaire utilisés (Molrep (Vagin et Teplyakov, 2010) et Phaser (McCoy, 2007) par exemple) fonctionnent à partir des cartes de Patterson des molécules. Une carte de Patterson est une carte centrosymétrique représentant tous les vecteurs interatomiques intra- et intermoléculaires par des pics de positions et d'intensités variables. Elle est générée par transformée de Fourier à partir des modules au carré des facteurs de structure et peut donc être calculée directement après l'enregistrement des clichés de diffraction.

Dans un premier temps, la fonction de rotation permet de déterminer l'opérateur (composé de trois angles) à appliquer au modèle pour l'orienter correctement dans la maille. En pratique, le logiciel superpose les cartes de Patterson du modèle et de la protéine cible sous différentes orientations afin de les faire coïncider au mieux, en tenant compte des vecteurs intramoléculaires uniquement (avec une sphère dont le diamètre est inférieur à celui de la protéine).

Dans un second temps, un maximum de corrélation est recherché entre la carte de Patterson du cristal et celle du modèle par la fonction de translation. Celle-ci prend en compte

les vecteurs interatomiques intermoléculaires et détermine la translation (composée de trois vecteurs) à appliquer au modèle préorienté dans la maille du cristal.

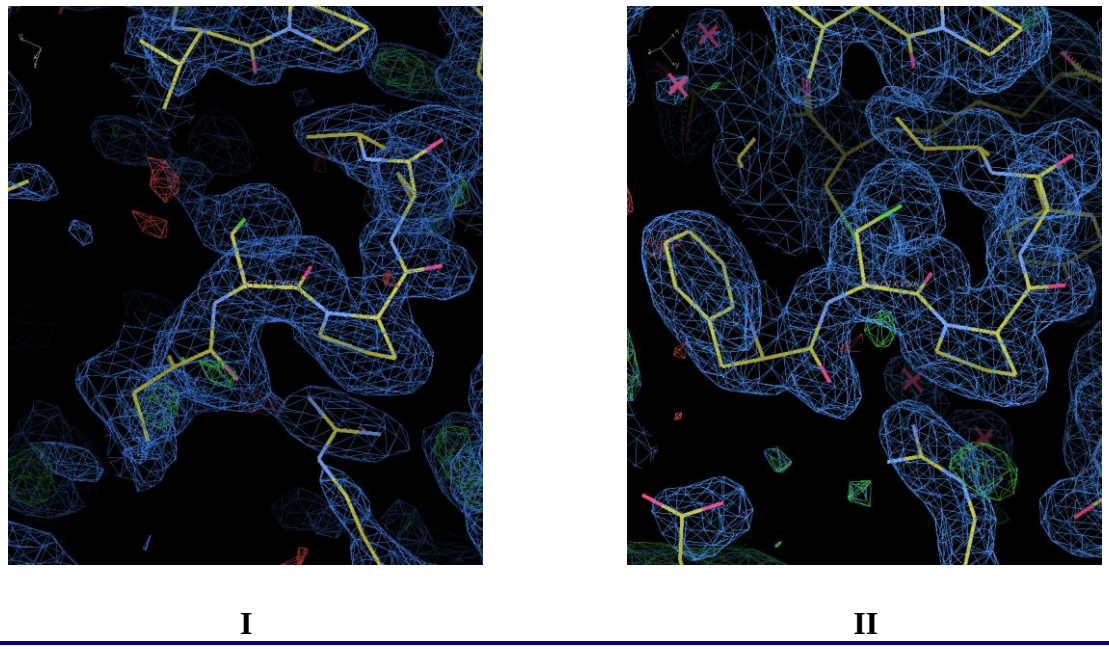
A l'issue d'un remplacement moléculaire et selon l'algorithme utilisé, différents scores sont calculés pour évaluer la fiabilité du résultat. Par ailleurs, une inspection visuelle vérifie la cohérence de l'empilement de la structure modèle dans la maille du cristal de la protéine d'intérêt. Si les molécules du modèle s'interpénètrent, la solution du remplacement moléculaire n'est pas physiquement réaliste et doit être rejetée. Dans une situation de remplacement moléculaire réussi, ce cristal virtuel du modèle permet de calculer des modules et phases qui, associés aux modules des facteurs de structure F_{khl} du cristal étudié, permettent d'obtenir différentes cartes de densité électronique de la molécule étudiée. Une reconstruction du modèle est ensuite effectuée.

VIII Construction, affinement et validation du modèle

VIII.1 Construction avec COOT

Après l'étape de remplacement moléculaire, il convient de reconstruire la structure de la protéine d'intérêt à l'aide des cartes de densité électronique. Cette reconstruction a été réalisée à l'aide du logiciel graphique Coot ('Crystallographic Oriented Object Toolkit', Emsley et Cowtan, 2004). Différentes fonctionnalités sont offertes par ce logiciel intuitif : affinement local des positions des atomes dans la densité, mutation de chaînes latérales, régularisation de la stéréochimie, etc. Le but est d'obtenir un modèle de stéréochimie correcte, en cohérence avec les données enregistrées (via les cartes de densité électronique). Le logiciel permet également d'ajouter les molécules d'eau liées à la protéine, les ions, ainsi que les éventuels ligands de la protéine. Deux types de cartes de densité électronique « différence » sont utilisées conjointement. La carte *Fo-Fc* est une carte dont le facteur de structure *Fc* « calculé » à partir du modèle est soustrait du facteur de structure *Fo* « observé » lors de l'expérience. Ainsi cette carte montre les erreurs du modèle sous la forme de pics positifs (verts, Figure 36) là où des électrons sont manquants et pics négatifs (rouges) dans les régions du modèle comportant un surplus d'électrons. Cette carte qui met l'accent sur l'erreur du modèle présente deux défauts. D'une part elle n'a pas l'apparence de la surface moléculaire et d'autre part elle est très bruitée et donc difficile à interpréter lorsque le modèle présente de

nombreuses erreurs. La deuxième carte différence *2Fo-Fc* est beaucoup plus intuitive car elle peut être interprétée comme une carte représentant la localisation réelle des électrons (surface bleue, Figure 36). Ces cartes sont souvent difficiles à interpréter car elles sont biaisées par les erreurs sur les positions atomiques, les intensités mesurées et les phases. Il est possible de prendre en compte ces erreurs en calculant des cartes pondérées appelées également cartes sigmaA (voir l'ouvrage de Sherwood et Cooper). Ces cartes sont du type *mFo-DFc* ou *2mFo-DFc* où *m* représente la qualité des phases pour chaque réflexion (figure de mérite) et *D* peut être vu comme la fraction correcte du modèle.



I **II**

Figure 36 **Région du site catalytique dans la structure de TvGSTO3C au début des étapes de reconstruction (I) et en fin d'affinement (II)**
Carte *2Fo-Fc* (contourée à 1,2 σ) en bleu et carte *Fo-Fc* (contourée à 3 σ) en rouge et vert.

Dans le cas des GST, qui sont constituées de deux domaines successifs, le domaine N-terminal est généralement bien conservé parmi les classes, alors que le domaine C-terminal est constitué de plusieurs hélices de longueur et position variables suivant les classes. Le travail au niveau du domaine N-terminal a donc surtout consisté en la mutation des chaînes latérales, alors qu'au niveau du domaine C-terminal un travail plus conséquent a parfois été nécessaire (suppression de régions/hélices du modèle du remplacement moléculaire et reconstruction complète).

VIII.2 Affinement par maximum de vraisemblance avec PHENIX

Après plusieurs étapes de reconstruction, un affinement de la structure peut être entrepris. Cette étape vise à affiner les coordonnées atomiques du modèle ainsi que les facteurs d'agitation (facteurs B) de chaque atome en tenant compte à la fois des données initiales et de la stéréochimie du modèle sous la forme de contraintes (par rapport aux longueurs de liaisons, angles de liaison, planéité, angles de torsion, etc…). Le logiciel utilisé pour affiner les structures de GST a été phenix.refine de la suite PHENIX (Adams et al, 2010) qui fonctionne en utilisant la méthode du maximum de vraisemblance. Le concept du maximum de vraisemblance repose sur la probabilité de mesurer les facteurs de structure observés F_{obs} compte tenu du modèle à affiner. La recherche du maximum de cette probabilité est ensuite effectuée lors de l'affinement.

Pour surveiller le bon déroulement de l'affinement, deux indicateurs sont utilisés : les facteurs *R* et R_{free}. Le facteur *R* (aussi appelé R_{cryst}) est un facteur d'accord exprimé selon la relation :

$$R = \sum_{hkl} \left| |F_{obs}| - k.|F_{calc}| \right| / \sum_{hkl} |F_{obs}|$$

Ce facteur mesure ainsi le désaccord entre les modules des facteurs de structure dérivés du modèle et les modules des facteurs de structure mesurés expérimentalement. Ce facteur est en général voisin d'une valeur de 40 % en début d'affinement pour diminuer à 20 % voire moins en fin de cycle de construction/affinement si la résolution est suffisante.

Le deuxième facteur R_{free} est exprimé par la même relation que le facteur *R* à partir d'une petite portion (habituellement 5 %) du jeu de données complet qui n'est pas utilisée lors de l'affinement. Cet indicateur, dit non-biaisé, sert de garde-fou pour ne pas sur-affiner les données. Ainsi la qualité du modèle s'améliore lorsque que les facteurs *R* et R_{free} diminuent tous deux au cours de l'affinement.

En pratique, plusieurs cycles d'affinement/reconstruction manuelle sont nécessaires pour aboutir à une structure correctement affinée. Lorsque les facteurs R et R_{free} ne diminuent plus, que le modèle est correct d'un point de vue stéréochimique et que l'affinement ne permet plus d'améliorer les cartes de densité, on admet que la structure est finalisée.

VIII.3 Validation du modèle avec MOLPROBITY et PDB-Validation server

La dernière étape de la résolution d'une structure est la validation de celle-ci en termes de stéréochimie et d'accord entre les données expérimentales et le modèle. Le programme MOLPROBITY (Davis et al, 2004) est implémenté dans PHENIX et permet constamment de surveiller le bon déroulement du processus après chaque étape de construction/affinement. Ce module permet notamment de vérifier les points suivants :

- diagramme de Ramachandran pour évaluer les angles de torsion
- rotamères préférentiels
- clashes entre atomes voisins
- chiralité
- planarité
- corrélation entre le modèle et la densité électronique

L'ensemble des paramètres est évalué à travers un score global qui doit être le plus bas possible (habituellement inférieur à 2) pour une structure correctement affinée, et qui reflète la résolution cristallographique à laquelle les valeurs de ces paramètres seraient attendues (Chen et al, 2010).

A l'issue de cette première évaluation, le modèle peut ensuite être vérifié par le serveur en ligne PDB-Validation Server (https://validate-rcsb-2.wwpdb.org/) qui permet d'évaluer la structure et les données à travers le pipeline menant au dépôt de la structure dans la PDB. Le cas échéant, la structure peut être modifiée avant le dépôt définitif. Le serveur de validation vérifie globalement les mêmes paramètres que MOLPROBITY mais s'avère complémentaire car il compare la structure nouvellement produite aux autres structures déjà déposées dans la banque, permettant par exemple de juger la qualité du modèle vis-à-vis de structures de résolution similaire.

IX Bio-informatique

IX.1 Obtention des séquences de GST et alignements

Les différentes séquences de GST utilisées pour réaliser les études phylogénétiques ont été obtenues sur le site web du Joint Genome Institute (https://jgi.doe.gov/), via le module de recherche Mycocosm répertoriant les séquences en acides aminés de protéines fongiques. Les séquences sont récupérées par l'algorithme de BLAST à partir d'une séquence de GST correspondant à la classe étudiée. Les séquences sont filtrées suivant la valeur *E* (représentant

la probabilité que l'alignement produit soit le fait du hasard, généralement inférieure à 10^{-3}), puis manuellement suivant leur longueur, et les séquences identiques présentes en plusieurs exemplaires sont réduites à un seul exemplaire. Les séquences sont ensuite alignées grâce au logiciel d'alignement multiple MUSCLE implémenté dans le logiciel MEGA (Tamura et al, 2013). Alternativement, les logiciels d'alignement en ligne MAFFT (Katoh et Standley, 2013) et PROMALS3D (Pei et Grishin, 2014) ont également été utilisés. Ces deux programmes présentent en effet l'avantage de permettre l'utilisation d'informations structurales pour contraindre l'alignement multiple.

IX.2 Phylogénie avec MEGA

Pour effectuer une étude de phylogénie, l'alignement multiple est soumis au logiciel MEGA (Tamura et al, 2013), qui permet de produire un arbre phylogénétique en choisissant la méthode la plus appropriée et le nombre d'itérations à effectuer pour affiner le modèle d'évolution. Pour l'analyse des séquences de GST, un modèle d'évolution de type Neighbor-Joining avec au moins 1000 amorces indépendantes a été utilisé. L'arbre phylogénétique généré est ensuite analysé et modifié dans le logiciel MEGA.

IX.3 Alignements de structure

Les alignements de structure ont été réalisés grâce au module d'alignement structural présent dans le logiciel Chimera (Pettersen et al, 2004). Ce module permet de superposer des structures proches et de produire l'alignement de séquence correspondant. Le module d'alignement structural présent dans le logiciel Pymol (Delano Scientific) a également été utilisé lors des analyses de structure.

IX.4 Docking avec GOLD

En cas d'échec à obtenir les structures de complexes GST-ligand bien que de fortes présomptions de fixation existent compte tenu d'autres expériences, des calculs de docking ont été menés avec le logiciel GOLD (Verdonk et al, 2003). Ce logiciel utilise un algorithme génétique inspiré de l'évolution pour générer une collection de positions possibles du ligand sur une protéine dont le site de fixation est connu, puis il sélectionne les plus probables sur la base d'un score évaluant leur justesse. Pour les calculs de docking du GS-PAP dans le site actif de TvGSTX3, des contraintes de distance ont été appliquées entre les résidus du site G

de la protéine et la partie glutathionyle du substrat GS-PAP. Des contraintes de distance ont également été appliquées sur la partie phenylacetophénone, afin de maintenir une stéréochimie plausible en regard des structures de petites molécules similaires présentes dans la Cambridge Structural Databank (https://www.ccdc.cam.ac.uk/).

Résultats et discussion

I Etude des GST Oméga de *Trametes versicolor*

Ce premier chapitre rapporte les résultats obtenus sur les GSTO du champignon saprotrophe *Trametes versicolor* (TvGSTO). Tout d'abord sont exposées l'analyse phylogénétique ainsi que les données enzymatiques acquises par nos collaborateurs du laboratoire IAM en partie I.1. Les parties I.2 et I.3 visent à replacer ensemble les résultats structuraux ainsi que les propriétés d'interaction des TvGSTO, ces résultats étant décrits en détail dans deux publications données en partie I.4. Une discussion sur l'aspect structural des résultats obtenus est proposée en dernière partie I.5. Les implications biologiques des résultats acquis sont discutées dans les publications associées et seront également repris dans la partie « Discussion générale et Perspectives » de ce manuscrit.

I.1 Etude phylogénétique et biochimique des GSTO de *T. versicolor*

Le champignon dégradeur de bois *Trametes versicolor* possède dans son génome 46 gènes codant pour des GST putatives. Sur la base de leurs séquences en acides aminés, ces GST peuvent être définies parmi les classes FuA (6 isoformes), Phi (4 isoformes), GTT (4 isoformes), MAK16 (une isoforme), EF1Bγ (2 isoformes), Ure2p (10 isoformes formant une extension), Xi (3 isoformes) et Oméga (16 isoformes formant une extension). Une extension du nombre d'isoformes des classes Oméga et Ure2p est aussi prédite chez le champignon de la pourriture blanche *Phanerochaete chrysosporium.* Un alignement des séquences de GSTO de ces deux organismes suivi d'une analyse phylogénétique met en évidence la présence de deux groupes distincts : les GSTO de type II et les GSTO de type III (Figure 37), correspondant à deux sous-classes précédemment identifiées (Morel et al, 2009). Les GSTO de type II (2 isoformes chez *T. versicolor* et 3 isoformes chez *P. chrysosporium*) présentent toutes une cystéine dans la région appelée « motif catalytique » (Figure 38). Les GSTO de type III (14 isoformes chez *T. versicolor* et 4 isoformes chez *P. chrysosporium)* présentent soit une cystéine, soit une sérine dans leur motif catalytique. Chez *T. versicolor,* une majorité de ces GSTO de type III présentent une sérine dans le motif (12 isoformes).

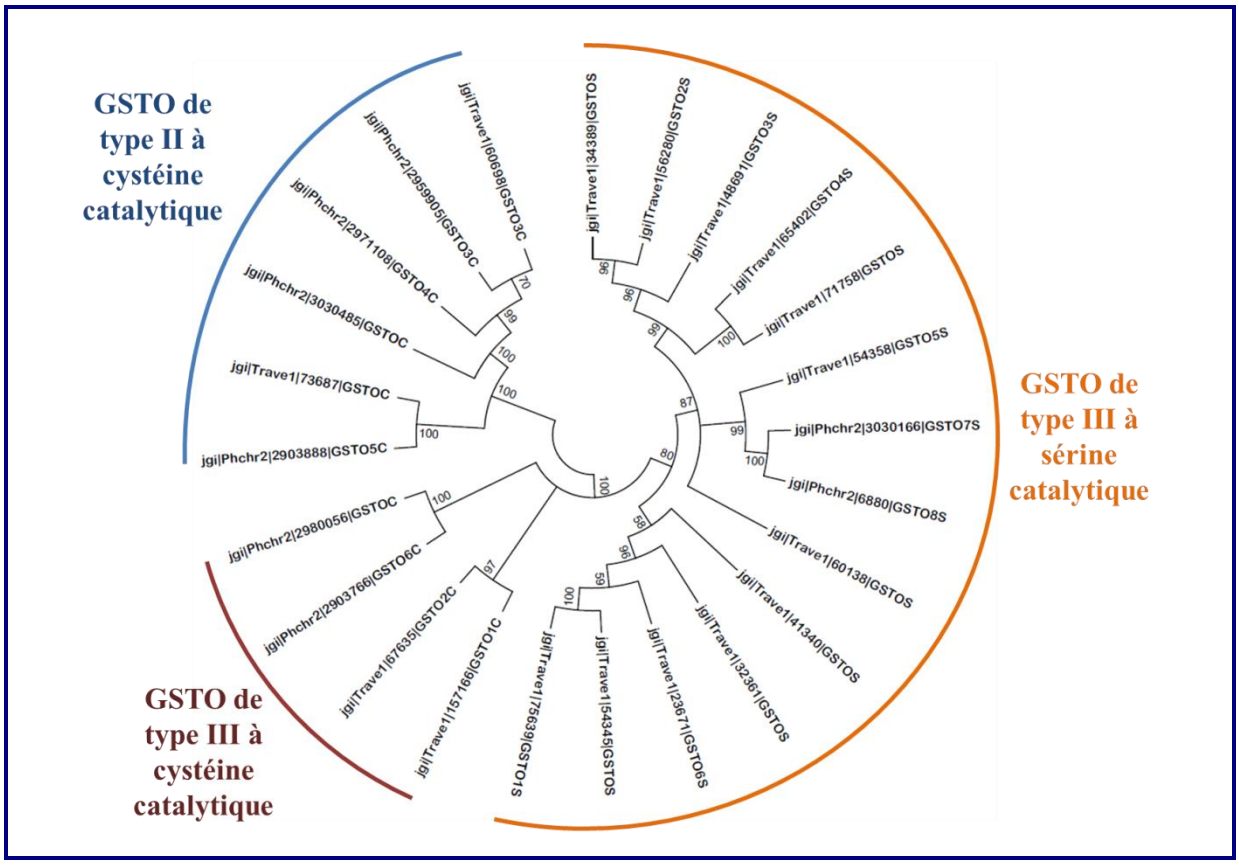


Figure 37 **Phylogénie des GST de la classe Oméga chez les dégradeurs de bois *P. chrysosporium* et *T. versicolor***

L'arbre phylogénétique a été réalisé avec le logiciel MEGA par la méthode 'Neighbor-Joining', avec 1000 amorces indépendantes.

Neuf isoformes de GSTO (représentatives des seize isoformes au total) provenant de *T. versicolor* (TvGSTO), parmi lesquelles une isoforme de type II et huit isoformes de type III, ont été choisies pour effectuer une étude biochimique et structurale. Un alignement des séquences protéiques des TvGSTO étudiées est montré en Figure 38. TvGSTO1S, TvGSTO2S, TvGSTO3S, TvGSTO4S, TvGSTO5S et TvGSTO6S sont des GSTO de type III présentant une sérine dans leur motif catalytique alors que TvGSTO1C et TvGSTO2C sont des GSTO de type III présentant une cystéine. TvGSTO3C quant à elle appartient au type II et présente donc une cystéine. Cette isoforme possède une extrémité N-terminale allongée de 22 résidus par rapport aux TvGSTO de type III. Les isoformes de TvGSTO sélectionnées ont été clonées, purifiées et caractérisées biochimiquement. Globalement, les six TvGSTO à sérine catalysent des réactions GSH-transférase et peroxydase, alors que les GSTO à cystéine catalysent des réactions thiol-transférase et GSH-lyase (Tableau 4). Ces données biochimiques permettent de supposer que selon le cas, la sérine ou la cystéine du motif joue un rôle catalytique. Une étude structurale de ces neuf isoformes a été entreprise de manière systématique.

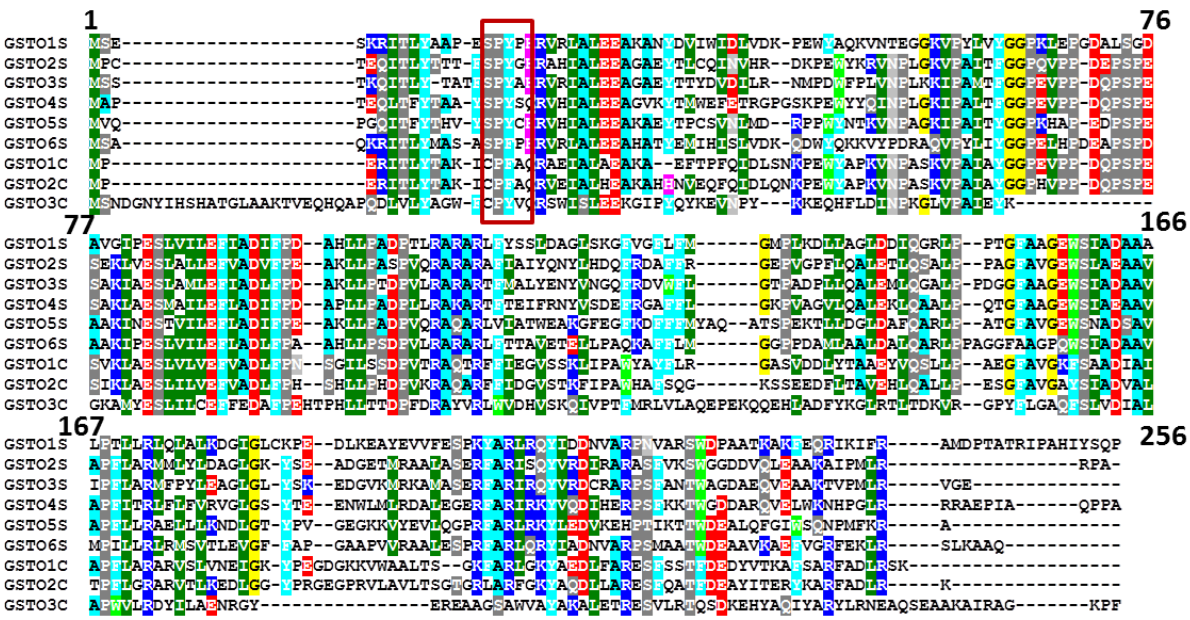


Figure 38 **Alignement des séquences de GSTO de *Trametes versicolor* étudiées dans le cadre de ce projet**

Les séquences de six GSTO à sérine et trois GSTO à cystéine ont été alignées avec le programme Promalsd3D (Pei et Grishin, 2014) en utilisant comme contrainte les structures de TvGSTO3S et TvGSTO3C résolues dans le cadre de ce projet. La numérotation correspond à la séquence de TvGSTO1S. Le motif catalytique contenant le résidu catalytique (S ou C) est encadré en rouge.

		Type d'activité					
		GSH-transférase			Peroxydase	Thiol-transférase	GSH-lyase
Résidu catalytique	**Isoforme**	CDNB	Phenethyl ITC	Benzyl ITC	Cumène hydroperoxyde	HED	GS-PAP
Sérine	GSTO1S	(2.97 ± 0.13) E+03	(7,98 ± 0,23) E+04	(2,37 ± 0,05) E+05	(2,47 ± 0,07) E+03	ND	ND
	GSTO2S	276,30 ± 9,25	(4,05 ± 0,09) E+05	(2,13 ± 0,04) E+06	887,40 ± 39,44	ND	ND
	GSTO3S	(2,97 ± 0,11) E+04	(2,53 ± 0,42) E+05	(2,72 ± 0,06) E+05	ND	ND	ND
	GSTO4S	145,90 ± 11,42	(1,80 ± 0,04) E+04	(1,17 ± 0,03) E+05	111,20 ± 10,02	ND	ND
	GSTO5S	407,50 ± 14,31	(9,67 ± 0,22) E+03	(9,02 ± 0,24) E+03	331,20 ± 9,83	ND	ND
	GSTO6S	(5,43 ± 0,18) E+03	(4,65 ± 0,15) E+05	(8,69 ± 0,49) E+05	(6,59 ± 0,20) E+03	179,50 ± 6,59	ND
Cystéine	GSTO1C	ND	ND	ND	ND	381,90 ± 43,76	(2,29 ± 0,13) E+04
	GSTO2C	ND	ND	ND	ND	(4,42 ± 0,17) E+03	(3,74 ± 0,07) E+04
	GSTO3C	ND	ND	ND	ND	(1,21 ± 0,04) E+04	(4,19 ± 0,13) E+04

Tableau 4 **Bilan des réactions enzymatiques catalysées par les TvGSTO (communication personnelle, T. Perrot, IAM)**

Les valeurs données correspondent aux efficacités catalytiques k_{cat}/K_m données en $M^{-1}.s^{-1}$. « ND » signifie qu'aucune activité n'a été détectée. Les formules développées des substrats utilisés sont données en annexe.

I.2 Etude structurale de neuf isoformes de TvGSTO

La recherche des conditions de cristallisation pour l'ensemble des neuf TvGSTO étudiées a été entreprise. A l'issue de cette étude systématique, des cristaux ont été obtenus pour toutes les isoformes à l'exception de l'isoforme TvGSTO5S. Pour les huit isoformes cristallisées, six structures ont pu été résolues : TvGSTO1C, TvGSTO2C, TvGSTO3C, TvGSTO2S, TvGSTO3S et TvGSTO6S. Dans chaque cas, des cristaux conduisant à des données de diffraction de résolution suffisante à une étude structurale de complexe enzyme-ligand ont éte obtenus (de 1,3 Å à 2,3 Å). Un résumé des conditions de cristallisation, des statistiques des données de diffraction ainsi que des affinements structuraux est donné en Tableau 5 et Tableau 6. Les données de diffraction collectées pour TvGSTO1S et TvGSTO4S avaient dans les deux cas une basse résolution (Tableau 7). Il n'a pas été possible de résoudre les structures correspondantes car des problèmes de macles importants ont été décelés lors du traitement des données. Les structures des six autres TvGSTO résolues ont été obtenues par la méthode du remplacement moléculaire. La structure de TvGSTO3S a été résolue à partir du modèle de la GST Lambda 3 de Peuplier (code PDB 4PQI) présentant 28 % d'identité de séquence. La démarche employée a été d'utiliser le pipeline MrBUMP (Keegan et Winn, 2008) qui a permis d'améliorer le modèle de remplacement moléculaire initial. La structure de TvGSTO3S a ensuite été utilisée pour résoudre par remplacement moléculaire les structures des autres isoformes de TvGSTOS et celle de TvGSTO2C. Enfin, cette dernière a été utilisée comme modèle pour résoudre les structures de TvGSTO1C et TvGSTO3C.

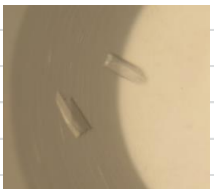
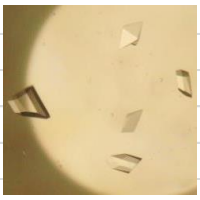
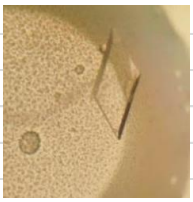
	Isoforme de GSTO		
	GSTO1C	**GSTO2C**	**GSTO3C**
Aspect des cristaux			
Conditions de cristallisation	10% PEG8000 0,1M HEPES pH 7,5	20% PEG4000 10% 2-propanol 0,1M HEPES-Na pH 7,5	15% PEG8000 0,2M sulfate d'ammonium
Méthode	Diffusion de vapeur	Microbatch	Microbatch
Collecte			
Source de rayons X	BM30A	BM30A	ID30B
Longueur d'onde (Å)	0,98014	0,97977	0,97625
Groupe d'espace	$P2_12_12_1$	C2	C2
Paramètres de maille			
a, b, c (Å)	62,7 86,4 92,1	68,9 142,2 65,4	132,5 54,3 75,2
α, β, γ (°)	90	90 116,2 90	90 94,1 90
Résolution (Å)	63,04 - 1,61 (1,64 - 1,61)	45,2 - 1,90 (1,90 - 1,95)	47,9 - 1,65 (1,69 - 1,65)
R-merge	0,070 (0,596)	0,055 (0,254)	0,034 (0,432)
Réflexions uniques	65 189 (3 134)	42 824 (3 095)	59 292 (4 551)
Complétude (%)	99,9 (97,7)	97,1 (95,1)	92,2 (96,4)
Redondance	6,8 (6,5)	3,7 (3,7)	2,6 (2,5)
I/sig(I)	15,4 (2,5)	13,3 (5,3)	14,6 (1,9)
CC1/2	1,00 (0,87)	1,00 (0,96)	1,00 (0,96)
Phasage par MR			
Modèle de remplacement	TvGSTO2C	TvGSTO3S (6F43)	TvGSTO2C
Affinement			
R-work / R-free	0,19 / 0,21 (0,24 / 0,29)	0,20 / 0,26 (0,44 / 0,42)	0,21 / 0,24 (0,30 / 0,31)
RMSZ - longueurs de liaison	0,41	0,53	0,52
RMSZ - angles de liaison	0,55	0,61	0,69
Facteur B moyen (Å²)	17,6	36,8	32,14
Ramachandran			
Favorable (%)	98,7	97,2	98,2
Outliers (%)	0,0	0,0	0,5
Code PDB	6HJS	/	/
Ligands déterminés en complexe	GSH	GSH, naringénine, oxyresvératrol	GSH

Tableau 5 **Résumé des résultats de cristallogenèse, collecte des données et affinements de structures pour les TvGSTO à cystéine catalytique**

Les valeurs entre parenthèses sont données pour la couronne de résolution la plus haute. Dans chaque cas, l'unité asymétrique contient un dimère de GSTO.

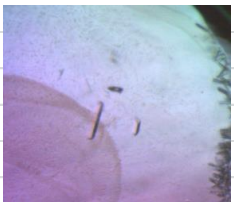
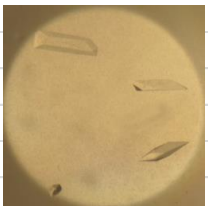
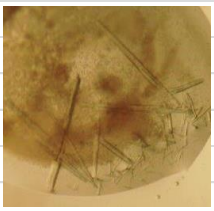
	Isoforme de GSTO		
	GSTO2S	**GSTO3S**	**GSTO6S**
Aspect des cristaux			
Conditions de cristallisation	11% PEG4000 0,1M HEPES-MES pH 7,0 0,05M acétate de Na 0,05M chlorure de Mg	30% PEG400 0,1M acétate pH 4,5 0,2M acétate de Ca	25% PEG1500 0,1M MMT pH 6,5
Méthode	Diffusion de vapeur	Microbatch	Diffusion de vapeur
Collecte			
Source de rayons X	BM30A	BM30A	ID30B
Longueur d'onde (Å)	0,97975	0,97975	0,97625
Groupe d'espace	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Paramètres de maille			
a, b, c (Å)	59,0 93,0 96,0	50,5 104,5 107,7	53,1 78,5 67,9
α, β, γ (°)	90	90	90 106,6 90
Résolution (Å)	48,02 - 2,19 (2,33 - 2,19)	47,88 - 1,35 (1,39 -1,35)	42,69 - 1,48 (1,52 - 1,48)
R-merge	0,104 (0,615)	0,071 (1,13)	0,039 (0,363)
Réflexions uniques	27 630 (4 309)	125 602 (9 203)	85 232 (6 051)
Complétude (%)	99,7 (98,4)	100 (99,9)	96,0 (92,7)
Redondance	7,2 (7,2)	8,2 (6,4)	2,7 (2,5)
I/sig(I)	16,5 (3,7)	16,5 (1,7)	14,5 (2,6)
CC1/2	1,00 (0,92)	1,00 (0,58)	1,00 (0,68)
Phasage par MR			
Modèle de remplacement	TvGSTO3S (6F43)	PtGSTL3 (4PQI)	TvGSTO3S (6F43)
Affinement			
R-work / R-free	0,20 / 0,26 (0,27 / 0,34)	0,13 / 0,17 (0,25 / 0,31)	0,15 / 0,18 (0,27 / 0,30)
RMSZ - longueurs de liaison	0,56	0,66	0,62
RMSZ - angles de liaison	0,63	0,78	0,76
Facteur B moyen (Å²)	34,5	20,3	22,7
Ramachandran			
Favorable (%)	98,3	98,3	99,0
Outliers (%)	0,2	0,0	0,0
Code PDB	6GIB	6F43	6F70
Ligands déterminés en complexe	oxyresvératrol	GSH, GS-DNB, GS-PAP, GS-hexyl, GS-PEITC, 2,4-HBP, 3,4-HBP, 2,4,4'-HBP, 2,3,4-HBP, dihydrowogonine	GSH, naringénine

Tableau 6 **Résumé des résultats de cristallogenèse, collecte des données et affinements de structures pour les TvGSTO à sérine catalytique**

Les valeurs entre parenthèses sont données pour la couronne de résolution la plus haute. Dans chaque cas, l'unité asymétrique contient un dimère de GSTO.

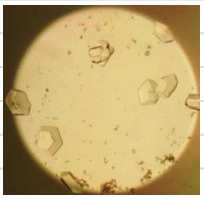
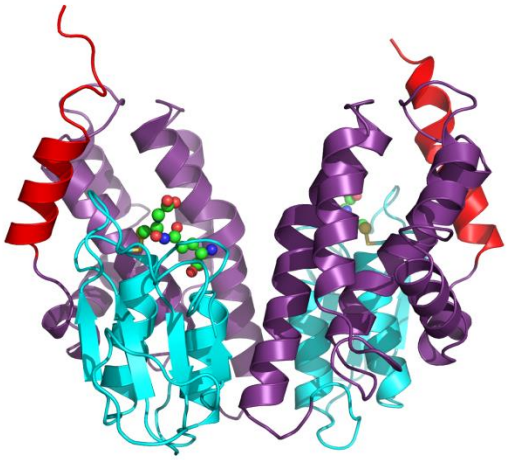
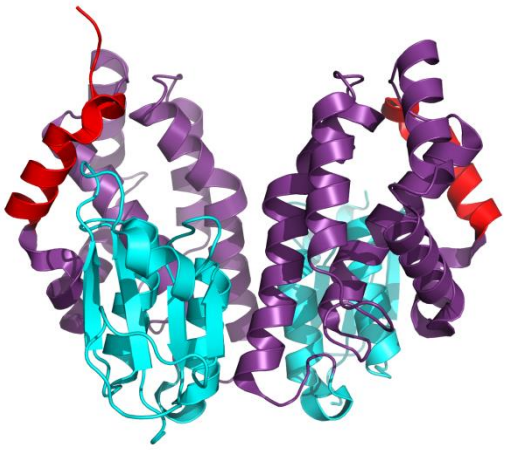
	Isoforme de GSTO	
	GSTO1S	**GSTO4S**
Aspect des cristaux		
Conditions de cristallisation	18% PEG8000	20% PEG3000
	0,1M cacodylate de Na pH 6,5	0,1M Tris pH 7,0
	0,2M acétate de Ca	0,2M acétate de Ca
		30% 1,5 diamino pentane Di-HCl
Méthode	Microbatch	Microbatch
Collecte		
Source de rayons X	BM30A	ID30B
Longueur d'onde (Å)	0,97978	0,97625
Groupe d'espace	$P6_4$	P3
Paramètres de maille		
a, b, c (Å)	173,1 173,1 165,5	155,8 155,8 57,1
α, β, γ (°)	90 90 120	90 90 120
Résolution (Å)	47,8 - 6,0 (6,7 - 6,0)	46,0 - 2,80 (2,94 - 2,80)
R-merge	0,105 (1,072)	0,104 (0,321)
Réflexions uniques	7 150 (2001)	34 882 (4 521)
Complétude (%)	99,6 (99,3)	91,5 (89,2)
Redondance	11,0 (11,3)	2,2 (2,1)
I/sig(I)	17,6 (2,5)	4,6 (1,7)

Tableau 7 **Résultats de cristallogenèse et collectes des données des TvGSTO1S et TvGSTO4S**
Les valeurs entre parenthèses sont données pour la couronne de résolution la plus haute.

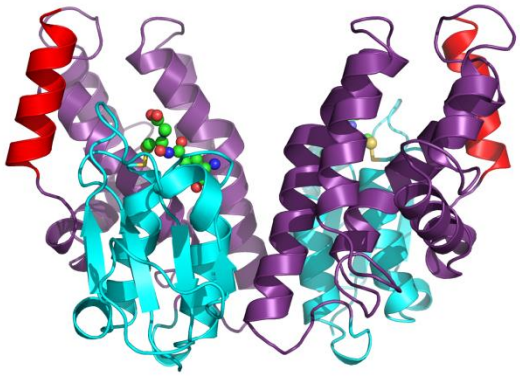
Les six structures de GSTO résolues adoptent le dimère canonique des GST à l'état cristallin (Figure 39). Les structures des différentes isoformes sont proches, avec un RMSD moyen inférieur à 1,0 Å de monomère à monomère. Néanmoins, des différences sont retrouvées au niveau de l'hélice α9 à l'extrémité C-terminale de la chaîne polypeptidique (Figure 39). Cette hélice située à proximité du site actif a une longueur variable dans les six structures (de trois tours dans les structures de TvGSTO2S, TvGSTO3S, TvGSTO1C et TvGSTO2C, à 4 tours dans la structure de TvGSTO3C et jusqu'à 5 tours dans la structure de TvGSTO6S). De plus, au sein d'une même isoforme il peut y avoir des différences au niveau de cette région : dans la structure de TvGSTO3C, cette extrémité C-terminale est repliée en hélice α9 dans le monomère A alors que dans le monomère B elle est désordonnée. L'analyse de l'empilement cristallin permet d'expliquer ces différences. De façon intéressante, la variabilité de l'extrémité C-terminale intra- et inter-isoformes a déjà été notée pour les GST Lambda du peuplier et les GST Oméga humaines (Lallement et al, 2014). Concernant les TvGSTO, un premier examen global des dimères suggère des différences subtiles entre la TvGSTO3C de type II et les cinq TvGSTO de type III.



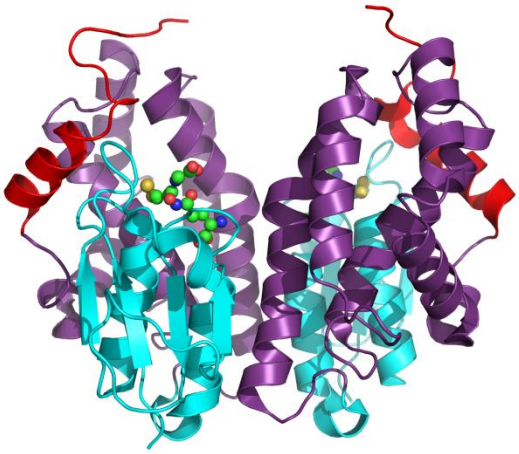
GSTO1C



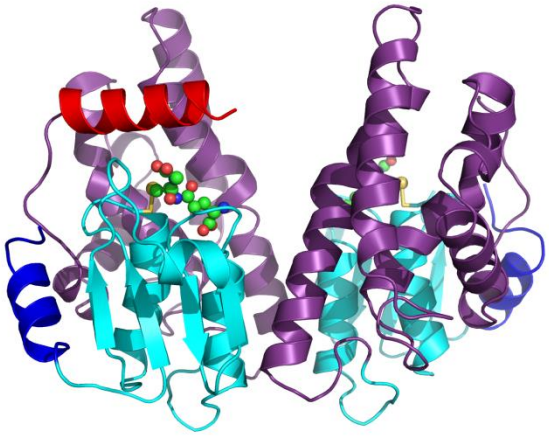
GSTO2S



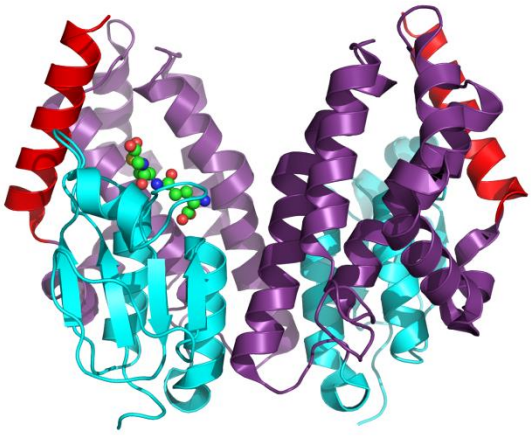
GSTO2C



GSTO3S



GSTO3C



GSTO6S

Figure 39 **Vue globale des structures des six isoformes de TvGSTO résolues lors de ce projet**

Les dimères sont représentés vus de profil, avec le domaine N-terminal en cyan et le domaine C-terminal en violet. L'hélice α9, de taille et position variable entre les structures, ainsi que l'extrémité C-terminale sont colorées en rouge. L'extrémité N-terminale ainsi que l'hélice α0 présentes chez TvGSTO3C sont colorées en bleu. Le glutathion est coloré en vert.

- Différences structurales entre les TvGSTO de type II et les TvGSTO de type III

La comparaison de la structure de TvGSTO3C (isoforme de type II) aux cinq autres structures (isoformes de type III) a permis de définir les différences structurales entre ces deux sous-classes (Figure 40) précédemment distinguées sur une base phylogénétique. La première différence est située en N-terminal de la structure de TvGSTO3C, dans laquelle une hélice supplémentaire appelée α0 est retrouvée. Cette hélice, absente des TvGSTO de type III, borde le côté extérieur du monomère et joint les domaines thiorédoxine et tout hélice. La deuxième différence est une insertion de 15 résidus entre les brins β3 et β4 dans les structures de TvGSTO de type III (TvGSTO2S, TvGSTO3S, TvGSTO6S, TvGSTO1C et TvGSTO2C). Cette insertion se rabat sur le feuillet β et est absente de la structure de TvGSTO3C où seul un coude de trois résidus est retrouvé entre les deux brins. La troisième différence est située à l'extrémité N-terminale de l'hélice α5 de TvGSTO3C, allongée de deux tours d'hélice par rapport à l'hélice équivalente chez les TvGSTO de type III. Une dernière différence structurale est observée pour ce dernier type, qui présente une hélice supplémentaire α6' à la suite de l'hélice α6, absente dans TvGSTO3C.

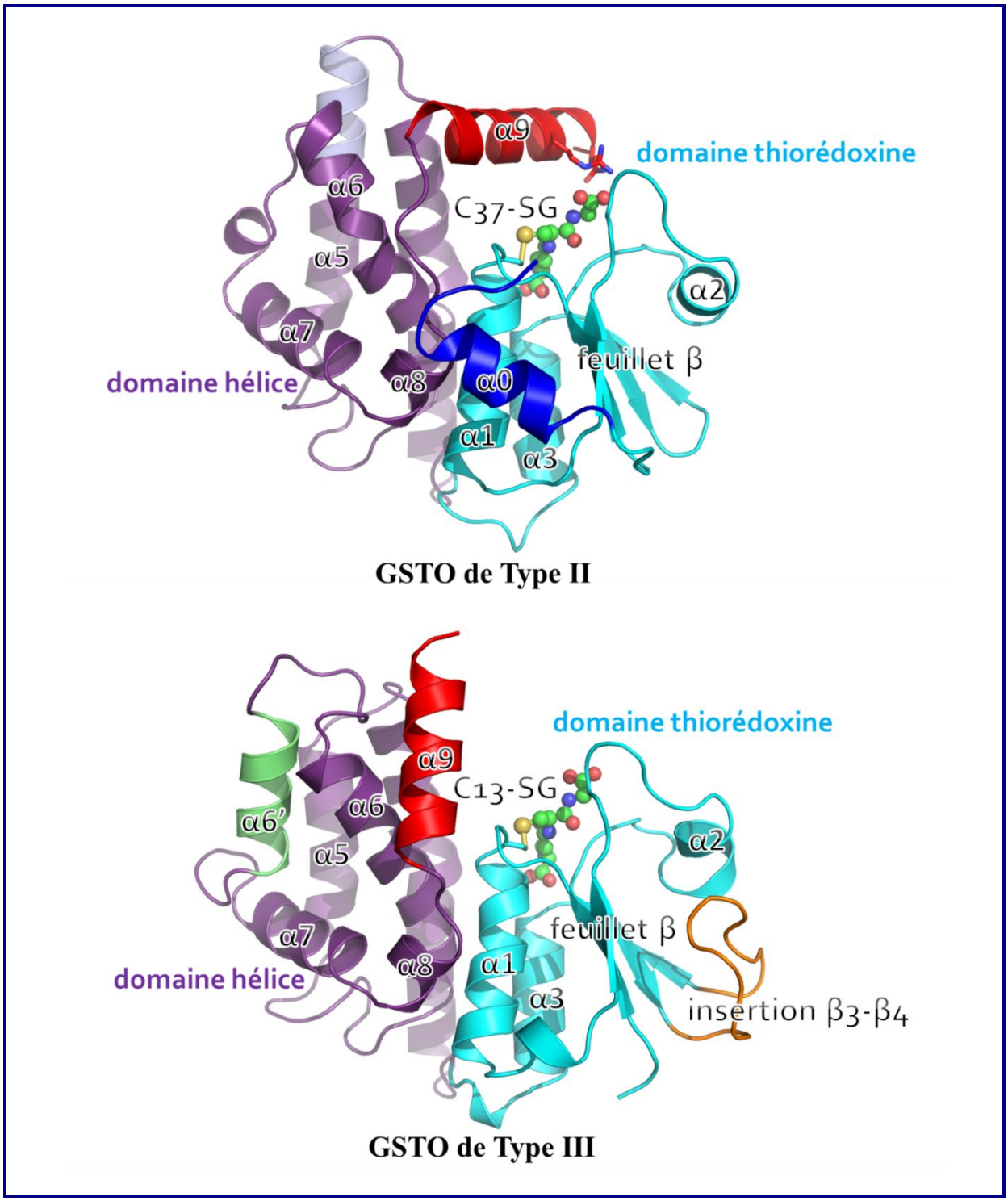


Figure 40 **Différences structurales entre les GSTO de type II et les GSTO de type III, en prenant l'exemple des structures de TvGSTO3C (type II, haut) et TvGSTO2C (type III, bas)**

Le domaine N-terminal et le domaine C-terminal sont colorés en cyan et violet, respectivement. Les spécificités structurales propres à chaque type sont colorées : pour TvGSTO3C (type II), l'hélice α0 et l'extension de l'hélice α5 sont colorées en bleu foncé et bleu clair, respectivement ; pour TvGSTO2C (type III) l'insertion entre les brins β3 et β4, et l'hélice α6' sont colorées en orange et olive, respectivement. L'hélice α9, de position différente entre les deux types, est colorée en rouge et les résidus participant à la stabilisation du glutathion dans TvGSTO3C sont représentés en bâtonnets (R238 et Q242).

➢ Site de liaison du glutathion

Les structures des trois TvGSTOC présentent un résidu cystéine en début d'hélice α1, dont l'atome de soufre gamma (Sγ) est systématiquement en pont disulfure avec l'atome Sγ du tripeptide glutathion. A la même position dans la chaîne polypeptidique, les structures des trois TvGSTOS natives présentent un résidu sérine pouvant être en liaison hydrogène avec le soufre du glutathion, lorsque celui-ci est présent dans la protéine native (TvGSTO6S) ou ajouté lors de la cristallisation, dans le cas de TvGSTO3S notamment (Figure 41). Les structures confirment ainsi le rôle catalytique présumé du résidu sérine ou cystéine retrouvé à cette position dans les TvGSTO. L'analyse structurale des sites actifs des six TvGSTO a mis en évidence leur site de reconnaissance du glutathion (site G). Celui-ci est conservé avec les chaînes latérales des résidus E80 et S81 stabilisant la partie γ-glutamyle du glutathion, la chaîne principale de V56 (I56 pour TvGSTO3S) interagissant avec la chaîne principale de la cystéine du glutathion, et une interaction électrostatique entre un résidu chargé positivement porté soit par la boucle β2-α2 (TvGSTO1C, TvGSTO2C, TvGSTO3C, TvGSTO6S), soit par la boucle α2-β3 (TvGSTO3S), et le groupement carboxy terminal du glutathion. A noter que dans le cas de l'isoforme TvGSTO3C, des interactions supplémentaires ont été observées avec les chaînes latérales des résidus R238 et Q242 provenant de l'hélice α9 au monomère B.

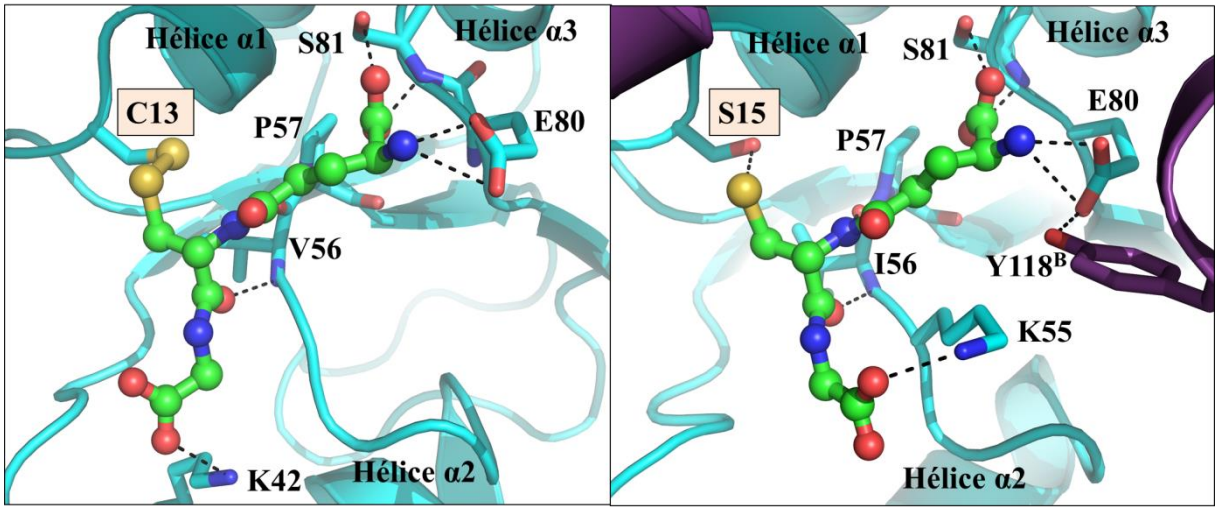


Figure 41 **Sites de reconnaissance du glutathion (sites G) dans les GSTO à cystéine (gauche, TvGSTO2C) et dans les GSTO à sérine (TvGSTO3S, droite)**

Dans le cas des GSTO à cystéine, le soufre du glutathion est lié covalemment au soufre de la cystéine catalytique (distance de 2,2 Å entre les deux atomes de soufre). Dans le cas des GSTO à sérine, le glutathion est lié non covalemment par liaison hydrogène avec l'hydroxyle de la sérine catalytique (distance de 3,2 Å entre l'atome S_γ et l'atome O_γ). Les interactions polaires sont représentées en pointillés. Le domaine N-terminal est coloré en cyan, le domaine C-terminal est coloré en violet. Le glutathion est coloré en vert.

➢ Ligands de cristallisation

Dans la structure de TvGSTO1C, une molécule de tampon HEPES a été identifiée et modélisée dans la densité électronique à proximité de la cystéine catalytique C13 en pont disulfure avec le glutathion (Figure 42). Cette molécule interagit avec la chaîne latérale de R229 portée par l'hélice α9, interagissant par liaison électrostatique via son groupement sulfonate. Cette molécule d'HEPES pourrait mimer la deuxième molécule de glutathion (GSH^2) nécessaire à la régénération des GSTO à cystéine catalytique. L'atome Sγ du GSH^2 pourrait être activé par catalyse électrostatique par R229. GSH^2 ainsi déprotoné pourrait alors réaliser l'attaque nucléophile de l'atome Sγ du premier GSH lié à la cystéine, aboutissant à la formation de glutathion oxydé (GSSG) et à la régénération de la GSTO. Un rôle similaire pour un résidu arginine a été relevé pour un mutant de la GST Mu humaine (Patskovsky et al, 2000). A noter que ce résidu arginine positionné en hélice α9 est conservé dans TvGSTO2C et TvGSTO3C, appuyant son rôle potentiel dans la catalyse des TvGSTO à cystéine.

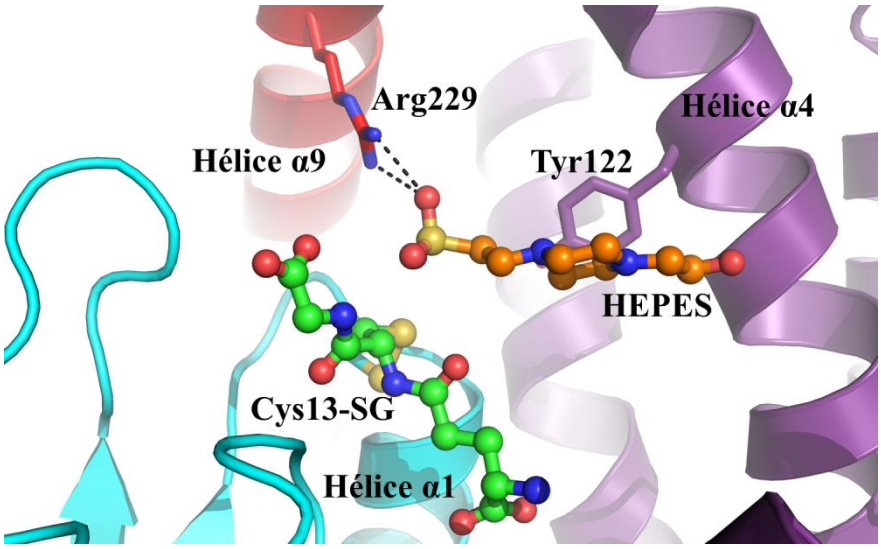


Figure 42 **Structure de TvGSTO1C en complexe covalent avec le glutathion et liant une molécule de tampon HEPES dans son site actif**

Le groupement sulfonate de la molécule d'HEPES est retrouvé à proximité du soufre du glutathion (3,0 Å), mimant ainsi la deuxième molécule de GSH intervenant dans la régénération de TvGSTO1C glutathionylée. La chaîne latérale de R229, retrouvée à proximité du groupement sulfonate de l'HEPES, interviendrait dans l'activation de cette molécule de GSH par catalyse électrostatique en déprotonant la cystéine du GSH.

Dans la structure de TvGSTO3S, deux molécules de glycérol ont été identifiées et modélisées dans la densité électronique, au niveau de la base du dimère et proche de son axe de symétrie (Figure 43). Ces molécules sont fixées lors de l'étape de cryo-protection des cristaux avant congélation. Chaque monomère porte une molécule de glycérol, ces molécules étant stabilisées par les chaines latérales des résidus R105 et D90. De manière intéressante, un

site de fixation de localisation similaire a été relevé pour la GST Phi 2 *d'Arabidopsis thaliana* (AtGSTF2) en complexe avec la quercétine (Ahmad et al, 2017).

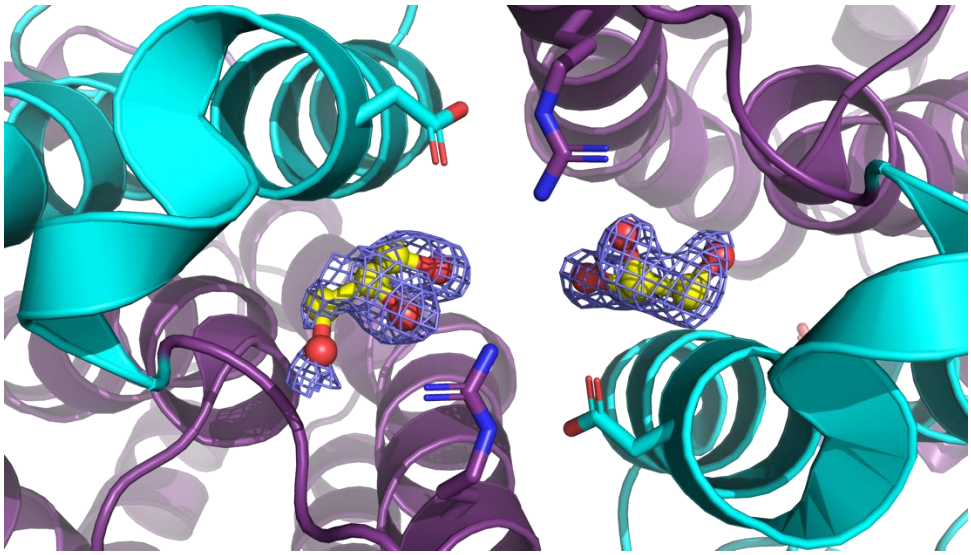


Figure 43 **Structure de TvGSTO3S native liant deux molécules de glycérol à proximité de l'axe 2 du dimère**

La carte de densité électronique montrée est une carte *2Fo-Fc* contourée à 1,0 σ. A proximité des molécules de glycérol sont retrouvées les chaînes latérales des résidus R105A,B et D90A,B.

- Ligand inconnu

Dans la structure de TvGSTO3S obtenue à 1,35 Å de résolution, les cartes de densité électronique de type *2Fo-Fc* et *Fo-Fc* ont révélé la présence d'une petite molécule organique non identifiée probablement fixée lors de la surexpression chez *E. coli* ou lors des étapes de purification (Figure 44). Cette molécule est stabilisée au niveau du site hydrophobe de TvGSTO3S, qui sera détaillé dans la partie « Ligands de type dérivés du glutathion » p. 119 de ce manuscrit. Son identification par spectrométrie de masse a échoué. Une raison pourrait être une concentration insuffisante de la molécule. Alternativement, une identification basée sur la densité électronique a été tentée par le programme phenix.ligand_identification compte tenu de la résolution. Ce programme place dans la densité et affine successivement les coordonnées de chaque molécule d'une banque constituée de 180 ligands de protéine les plus fréquents (Terwilliger et al, 2007). Malgré tout, aucun ligand potentiel n'a clairement été identifié par cette méthode et d'autres approches seront nécessaires pour identifier précisément cette molécule.

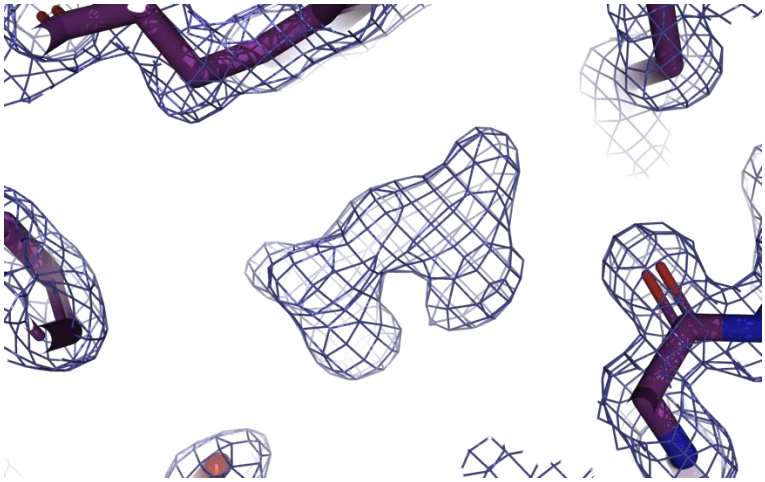


Figure 44 **Densité électronique correspondant à une molécule non identifiée au niveau du site H de TvGSTO3S**

La carte de densité électronique montrée est une carte *2Fo-Fc* contourée à 1,5 *σ*. Bien qu'aucune méthode d'identification par spectrométrie de masse ou basée sur la densité n'ait été fructueuse, le ligand est probablement une petite molécule organique.

I.3 Etude structurale des complexes GST-ligands issus du criblage

Les structures des formes apo ayant été déterminées pour trois TvGSTO à sérine et trois TvGSTO à cystéine, la suite du projet a consisté à identifier des ligands potentiels des différentes isoformes ainsi que leurs modes de fixation respectifs. Dans un premier temps, notre étude structurale s'est focalisée sur l'isoforme TvGSTO3S dont les cristaux ont été les premiers obtenus. Le trempage des cristaux avec des dérivés glutathionylés, puis la co-cristallisation avec les hydroxy-benzophénones identifiées par les expériences de TSA ont permis d'effectuer une caractérisation approfondie du site de liaison hydrophobe de cette isoforme. Dans un second temps, les polyphénols végétaux (stilbènes et flavonoïdes) identifiés par TSA ont été étudiés par cristallographie sur les trois isoformes résolues de TvGSTO à sérine, puis également sur l'isoforme TvGSTO2C. Pour chaque ligand identifié ou étudié, la valeur du TSA et/ou de l'inhibition enzymatique mesurée est donnée afin d'évaluer la stabilité du complexe formé.

I.3.a. Ligands de type « dérivés du glutathion »

Dans un premier temps, nos efforts se sont portés sur la caractérisation du site actif et notamment du site hydrophobe de l'isoforme TvGSTO3S. L'utilisation de dérivés glutathionylés avait permis précédemment d'identifier le site hydrophobe dans d'autres structures de GST, selon la littérature (Reinemer et al, 1991 ; Ji et al, 1999). Pour notre étude, deux types de dérivés glutathionylés ont été utilisés. D'une part, des analogues du glutathion ont été testés : le GS-hexane qui est un inhibiteur commun à toutes les classes de GST, et le GS-phenylacétophénone (GS-PAP) qui est un substrat dans le cas des GSTO à cystéine (Meux et al, 2013). D'autre part, des produits de conjugaison du glutathion ont aussi été utilisés : le GS-dinitrobenzène (GS-DNB) et le GS-phenethyl-isothiocyanate (GS-PEITC) qui sont les produits de conjugaison du glutathion respectivement avec le chlorodinitrobenzène (CDNB) et le phenethyl-isothiocyanate (PEITC). Ces quatre molécules solubilisées dans du tampon Tris-EDTA ont été diffusées à une concentration de 10 mM (0,5 mM pour le GS-PAP) dans la goutte contenant les cristaux de la forme apo de TvGSTO3S. Après collecte des données, la résolution de chaque structure et l'examen des cartes de densité électronique ont révélé la présence d'un ligand au site actif pour les quatre dérivés testés.

- <u>GS-PAP, GS-hexane et GS-DNB</u>

Pour ces trois dérivés, le groupement glutathionyle se fixe au site de reconnaissance du glutathion précédemment identifié à partir de GSH libre (Figure 45). Les groupements hydrophobes hexyl et phenylacétophénone se logent dans une cavité hydrophobe portée par le domaine C-terminal, à proximité du site G. Ce site peut être défini comme le site hydrophobe (site H) de TvGSTO3S, formé par les chaînes latérales de F128 (hélice α4), F168 et Y175 (hélice α6), ainsi que par les résidus R124 (hélice α4) et Y17 (hélice α1) à l'entrée du site. Ce site, particulièrement large, est capable d'accueillir deux molécules de GS-hexane (Figure 45) et est parfaitement adapté à la stabilisation des deux cycles aromatiques du groupement PAP, comme en témoigne également la forte valeur de TSA mesurée (plus de 19 °C). Pour ces deux dérivés, le ligand inconnu précédemment évoqué n'est plus présent au site H de TvGSTO3S, probablement déplacé lors de la fixation des dérivés du glutathion. Dans le cas du GS-DNB, le groupement dinitrobenzène n'est pas stabilisé à l'intérieur de cette poche mais demeure à l'entrée du site H, stabilisé par la chaîne latérale de R124 et en interaction de stacking π-π avec le cycle phényle de F128.

- GS-PEITC

Le dernier dérivé étudié adopte un mode de liaison clairement différent. En effet, le GS-PEITC a la partie carboxy terminale de son groupement glutathionyle orientée vers la cavité hydrophobe et stabilisée par la chaîne latérale du résidu S15 (Figure 45). En conséquence, l'atome Sγ du groupement glutathionyle n'est plus à proximité de S15 comme pour les autres dérivés et le GSH libre, mais orienté vers le domaine thiorédoxine. Le groupement PEITC est stabilisé au niveau d'un nouveau site de liaison offert par les chaînes latérales des résidus hydrophobes V37, M43 et F47 portés par la boucle β2-α2, changeant de conformation lors de la fixation. Ainsi, ce produit de conjugaison du glutathion adopte un mode de fixation différent de celui des analogues du glutathion (GS-hexane et GS-PAP) ou de celui du GS-DNB, se logeant tous au site H. Par ailleurs, la comparaison avec les structures des complexes des GST Alpha et Pi humaines avec le même dérivé suggère également un mode de fixation différent. Pour ces GST, le dérivé adopte plutôt une conformation telle que le groupement PEITC est stabilisé au niveau du site H porté par le domaine hélice (Kumari et al, 2016). Ces résultats pourraient suggérer un mode de fixation différent dans le cas de substrats de type isothiocyanates par TvGSTO3S et il est intéressant de rappeler que les isothiocyanates sont des composés pouvant présenter des propriétés anti-fongiques (Sellam et al, 2007).

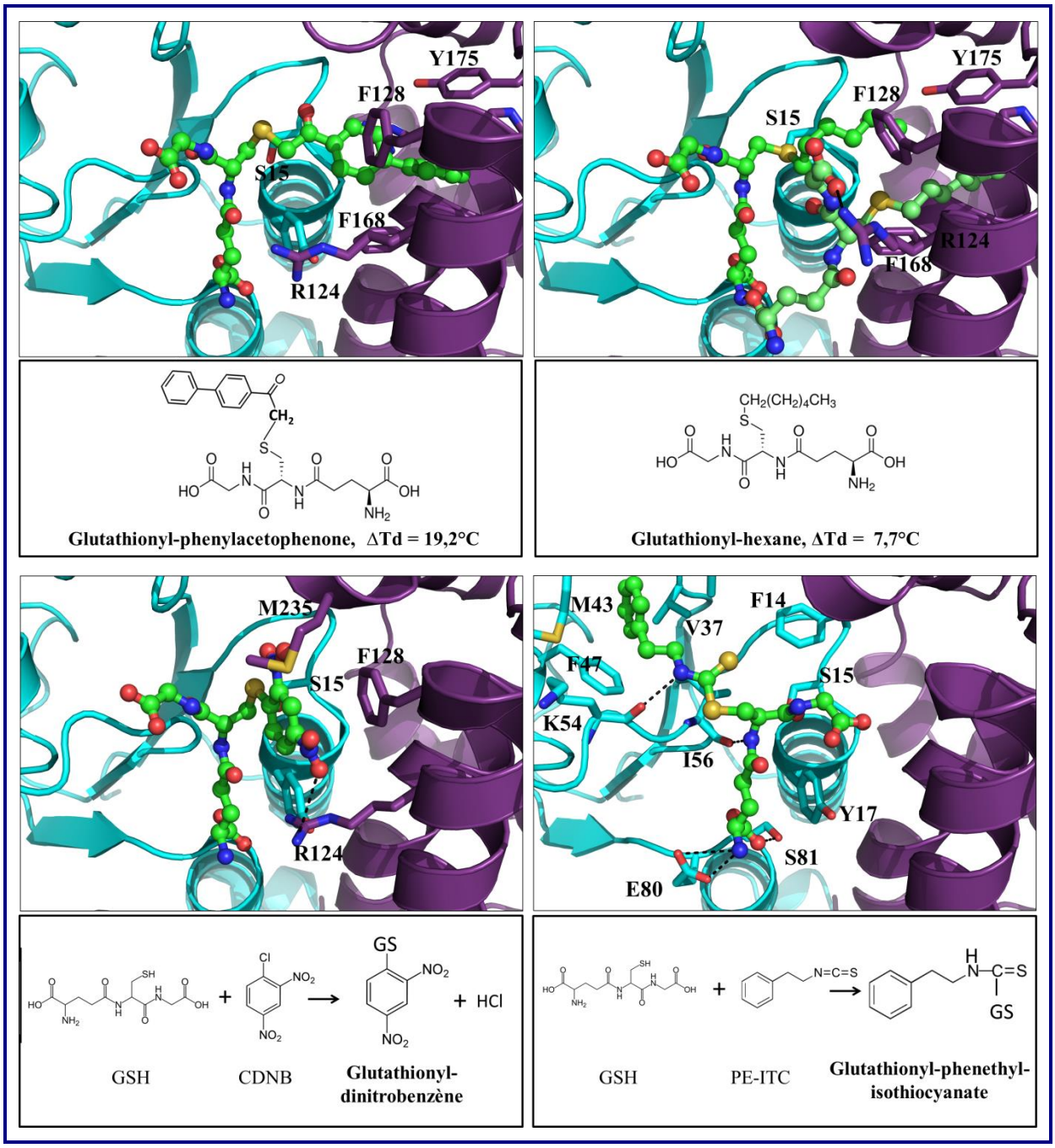


Figure 45 **Structures des complexes « TvGSTO3S-dérivés glutathionylés » déterminés par trempage des cristaux de la forme apo**

Haut : complexes de TvGSTO3S avec les analogues du glutathion GS-PAP (gauche) et GS-hexyl (droite). La valeur de TSA mesurée à partir de 80 µM de ligand est donnée. Bas : complexes avec les produits de réaction de conjugaison GS-DNB (gauche) et GS-PEITC (droite). Les interactions polaires sont représentées en pointillés. Pour raison de clarté, les interactions avec le groupement glutathion (cf Figure 41) ne sont pas représentées. Sont indiquées sous les structures concernées, les structures topologiques des analogues, ou les réactions de conjugaison impliquées dans la formation des produits identifiés dans la structure.

I.3.b. Ligands aromatiques

Pour la suite de l'étude, une stratégie de criblage par TSA a été mise en place en collaboration avec Eric Gelhaye et Thomas Perrot du laboratoire IAM (détaillée dans la publication « Schwartz et al, 2018 », *Scientific Reports*, p.130). Brièvement, la stratégie employée a été d'utiliser d'une part une librairie constituée de molécules connues pour leur réactivité avec les GST et d'autre part une librairie d'extraits de bois de forêts tempérées et tropicales. Les deux approches ont permis d'identifier globalement et indépendamment les polyphénols. A partir de la librairie de molécules chimiques ont été identifiées les hydroxybenzophénones et les flavonoïdes. De la librairie d'extraits de bois ont été identifiées précisément la dihydrowogonine (un flavonoïde issu du merisier *Prunus avium*) et l'oxyresvératrol (un stilbène issu de la bagasse guyanaise *Bagassa guianensis*). La caractérisation structurale de plusieurs complexes a ensuite été effectuée pour les isoformes présentant les propriétés d'interaction les plus prometteuses sur la base du TSA (TvGSTO2S, TvGSTO3S, TvGSTO6S et TvGSTO2C). Dans la plupart des cas où des structures de complexe ont été obtenues, les propriétés inhibitrices des molécules ont été étudiées. Ainsi, 21 essais de cristallisation de complexes ont été réalisés, parmi lesquels 9 ont aboutis à la structure d'un complexe.

i) Hydroxybenzophénones

La famille des hydroxybenzophénones (HBP) est celle qui s'est révélée la plus intéressante en TSA en particulier pour l'isoforme TvGSTO3S. Ces molécules sont composées de deux cycles benzéniques liés par un groupement cétone central (Figure 46). Un cycle ou les deux peuvent être hydroxylés. Alors que la benzophénone sans hydroxylation n'a aucune propriété stabilisatrice envers TvGSTO3S, les HBP présentant les meilleures propriétés stabilisatrices sont celles substituées sur un seul des deux cycles, ou sur les deux, avec une seule hydroxylation en position para sur le deuxième cycle substitué. Des structures cristallographiques en complexe avec TvGSTO3S ont été obtenues pour les quatre HBP (2,4-HBP, 2,3,4-HBP, 3,4-HBP et 2,4,4'-HBP) présentant les meilleurs déplacements de température de demi-dénaturation (ΔTm). Ainsi leur activité inhibitrice de la réaction GSH-transférase sur le PEITC a été évaluée afin d'estimer leur affinité pour TvGSTO3S.

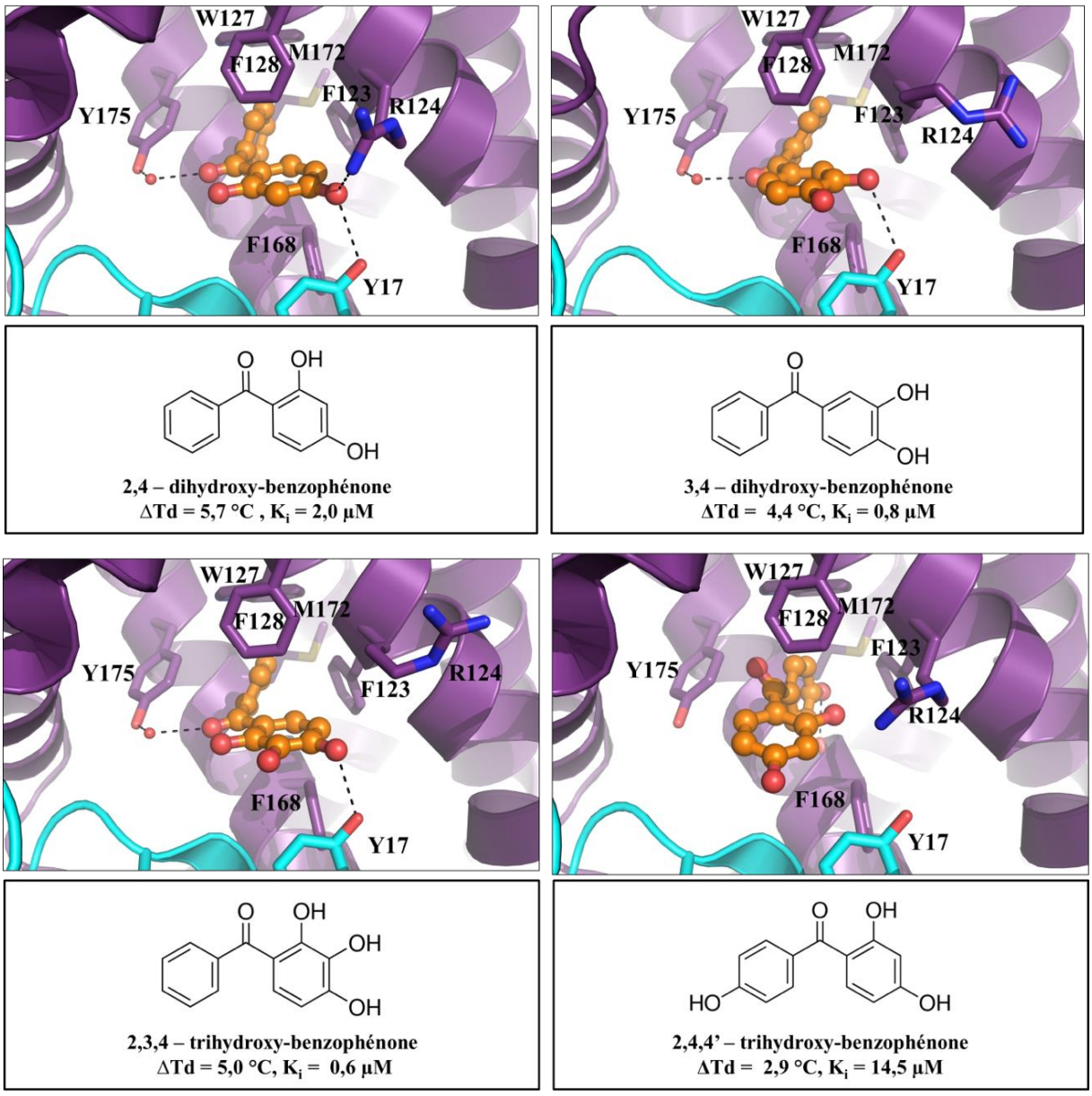


Figure 46 **Structures des complexes « TvGSTO3S-hydroxy-benzophénones » déterminées par co-cristallisation**

Les sites de fixation entre TvGSTO3S et les différentes hydroxybenzophénones identifiées par TSA sont montrés. La valeur de TSA mesurée à partir de 100 µM de ligand est donnée. La constante d'inhibition Ki correspond à l'inhibition mesurée pour la réaction GSH-transférase sur le substrat PEITC. Les interactions polaires sont représentées en pointillés.

Les HBP se fixent au niveau de la cavité abritant le site hydrophobe (Figure 46) précédemment identifié avec les dérivés glutathionylés. Alors que le cycle non substitué se loge au fond de cette poche hydrophobe dans le cas des 2,4-HBP, 2,3,4-HBP et 3,4-HBP, le cycle substitué est retrouvé à l'entrée du site, stabilisé par les chaînes latérales de Y17 et R124. La cétone centrale établit une liaison hydrogène avec une molécule d'eau cristallographique stabilisée par Y175. Dans le cas du complexe TvGSTO3S-2,4,4'-HBP, cette dernière molécule adopte une conformation différente. En effet, le cycle portant une hydroxylation en position para (4') est stabilisé au fond de la poche par l'oxygène de la chaîne principale de F168. Aucune liaison hydrogène n'est retrouvée avec le cycle à l'entrée du site H suggérant une stabilisation moins optimale, en accord avec une constante d'inhibition légèrement inférieure à celles des trois autres HBP (voir publication « Schwartz et al, 2018 », *Scientific Reports*, p.130). Concernant la 2,2',4,4'-HBP portant deux groupements hydroxyles sur chacun de ses deux cycles, aucun complexe n'a pu être obtenu, en accord avec une absence de TSA significatif avec cette molécule.

ii) Flavonoïdes et stilbènes

Le criblage de la librairie de flavonoïdes et stilbènes par TSA a mis en évidence que chaque isoforme possède un profil spécifique d'interaction (voir Tableau 8 et publications en partie I.4). Alors que l'isoforme TvGSTO6S interagirait avec la majorité des flavonoïdes testés, l'isoforme TvGSTO2S présente des interactions potentielles avec quelques flavonoïdes mais également avec le stilbène oxyresvératrol (Tableau 8). De plus, un extrait partiellement purifié de bois de merisier, co-cristallisé avec l'isoforme TvGSTO3S, a permis l'identification d'un flavonoïde par cristallographie d'affinité, la dihydrowogonine. Les différents polyphénols identifiés peuvent être séparés en deux catégories. Alors que certains se fixent aux sites hydrophobes (sites H) des GSTO (Figure 47), d'autres se fixent à un site de liaison présent à l'interface dimérique, appelé site ligandine (site L, Figure 48). Dans le cas de l'oxyresvératrol, deux configurations de la molécule sont présentes dans la structure, et interagissent au niveau de sites distincts. L'ensemble des résultats de TSA, des données structurales et d'inhibition enzymatique à partir des isoformes de structure résolue est résumé dans le Tableau 8. Dans les deux paragraphes qui suivent sont détaillés successivement les deux modes de fixation mis en évidence (site H puis site L).

Ligand	Isoforme											
	GSTO2S			GSTO3S			GSTO6S			GSTO2C		
oxyresvératrol	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	3,13	✔ HL	3,8	1,17	ND	ND	0,77		ND	0,70	✔ H	
baicaléine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	ND			1,81			5,03	ND		0,96		
scutellaréine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	ND			ND			2,51			0,81		
naringénine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	ND		ND	ND	ND	ND	1,92	✔ L	ND	1,76	✔ L	ND
wogonine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	4,44			4,40	ND	2,9	2,78		ND	-0,06		
chrysine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	3,45			2,41	ND		5,85	ND		0,27		
catéchine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	ND			ND	ND		ND			0,47	ND	
quercétine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	6,31			2,63	ND		5,72			0,32		
dihydrowogonine*	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	8,00			4,80	✔ H		5,50					
pinocembrine	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	1,95			1,47			0,94			-0,23		
wogonoside	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i	TSA	XRD	K_i
	2,49			0,48	ND	ND	1,17		ND	0,09	ND	

Tableau 8 **Bilan des résultats de criblage de la banque de flavonoïdes et stilbènes végétaux à partir des TvGSTO de structure résolue**

Pour chaque ligand, les résultats issus des expériences de stabilisation thermique (TSA, différence de température en °C), diffraction des rayons X (XRD, présence d'un ligand identifié dans le site L ou H de la GSTO) et inhibition enzymatique (Ki, valeur de la constante en µM) sont donnés. ND signifie qu'aucune interaction avec le ligand n'a été détectée. Les cases rayées correspondent aux expériences non testées. La valeur de TSA mesurée à partir de 100 µM de ligand est donnée. La constante d'inhibition Ki correspond à l'inhibition mesurée pour la réaction GSH-transférase sur le substrat PEITC. *Pour la dihydrowogonine, les tests ont été effectués à partir de la fraction purifiée de l'extrait de merisier ayant conduit à l'identification du flavonoïde. Tous les cristaux de complexes ont été préparés par la méthode du trempage à sec, à l'exception des complexes avec l'isoforme TvGSTO3S réalisés par co-cristallisation.

- Liaison au site H

Trois structures d'isoformes de TvGSTO ont été résolues avec des polyphénols présents aux sites H : TvGSTO2C et TvGSTO2S en complexe avec le *trans*-oxyresvératrol, et TvGSTO3S en complexe avec la dihydrowogonine isolée d'un extrait de merisier (Figure 47). Les modes de fixation du *trans*-oxyresvératrol adoptés par TvGSTO2C et TvGSTO2S sont différents. Alors que TvGSTO2C stabilise l'oxyresvératrol par des interactions impliquant des résidus venant du domaine thiorédoxine (Q140) et du domaine hélice (H124) chevauchant ainsi les deux sites G et H, TvGSTO2S stabilise le polyphénol dans une fente hydrophobe présente dans le domaine tout hélice (impliquant les chaînes latérales d'Y174 et D121 notamment). Dans les structures des deux isoformes, le *trans*-oxyresvératrol est retrouvé à proximité du résidu catalytique (C13 pour TvGSTO2C et S15 pour TvGSTO2S). En particulier pour TvGSTO2C, les carbones de la liaison double centrale sont retrouvés à moins de 4,0 Å du soufre du glutathion, questionnant un rôle catalytique hypothétique. Une telle proximité pourrait également être retrouvée dans le cas de TvGSTO2S en complexe avec l'oxyresvératrol si le glutathion était présent au site G. La structure de TvGSTO3S en complexe avec la dihydrowogonine révèle la fixation de cette molécule dans le site H précédemment décrit à partir des complexes avec les HBP et les dérivés glutathionylés. La large cavité hydrophobe de TvGSTO3S permet au phényle non substitué d'être stabilisé au fond de la poche, alors que le noyau chromanone est stabilisé à l'entrée du site avec des liaisons hydrogènes impliquant les résidus R124 et Y175, similairement aux HBP.

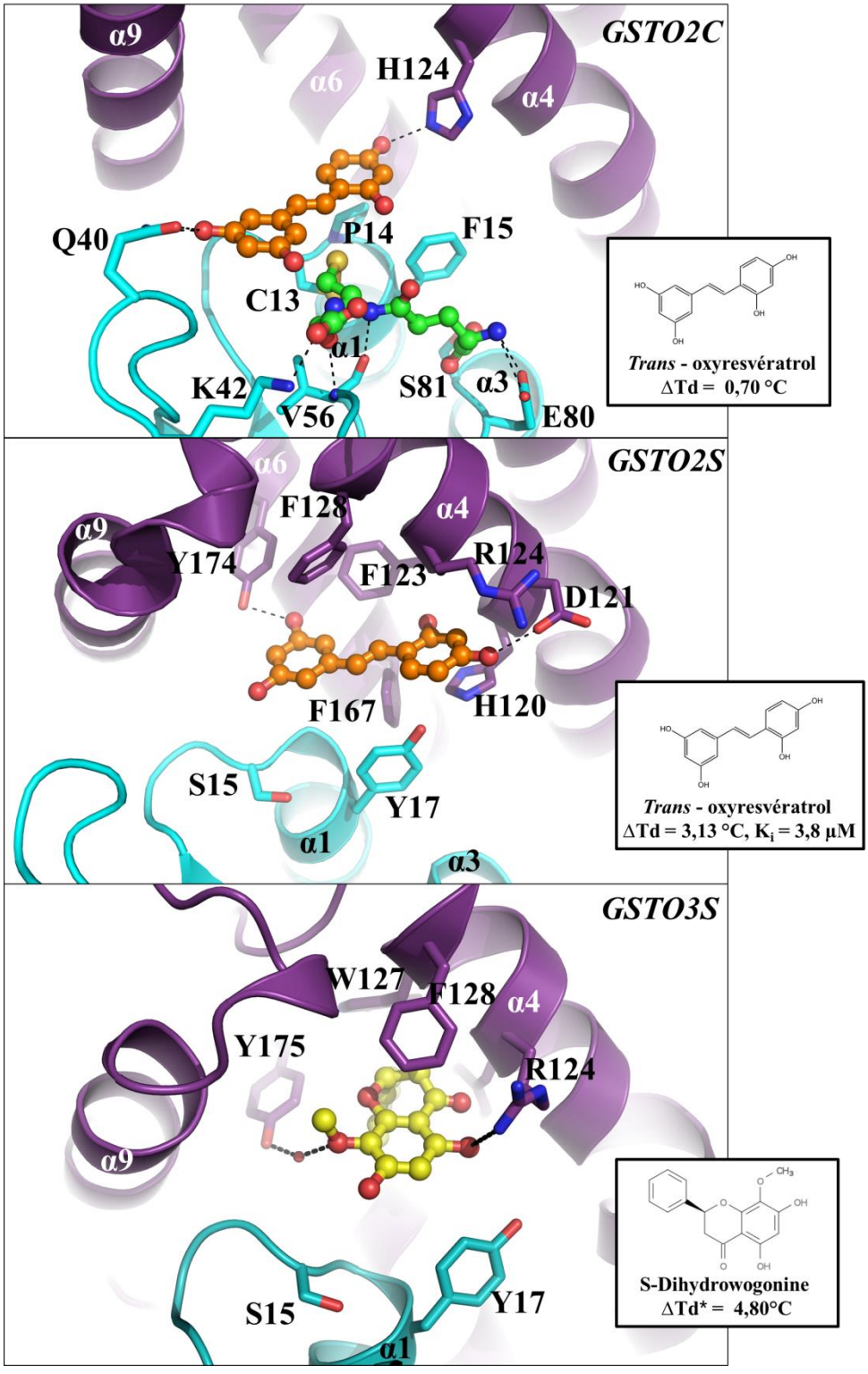


Figure 47 **Structures des complexes de TvGSTO avec les polyphénols végétaux identifiés au site H**

Les sites de fixation hydrophobes entre les TvGSTO et les différents polyphénols identifiés par TSA sont montrés. La valeur de TSA mesurée à partir de 100 µM de ligand est donnée. La constante d'inhibition Ki correspond à l'inhibition mesurée pour la réaction GSH-transférase sur le substrat PEITC. Les interactions polaires sont représentées en pointillés. *Pour la dihydrowogonine, les tests ont été effectués à partir de la fraction purifiée de l'extrait de merisier ayant conduit à l'identification du flavonoïde.

- Liaison au site L

Trois structures d'isoformes de TvGSTO ont été résolues avec des polyphénols présents aux sites L : TvGSTO2C et TvGSTO6S en complexe avec la naringénine et TvGSTO2S en complexe avec le *cis*-oxyresvératrol (Figure 48). Le site L est un site de fixation présent à l'interface entre les deux monomères, entre les hélices α3 et α4 des deux monomères ce qui constitue un site délimité par quatre hélices. La base du site est formée par deux résidus acides de l'hélice α3 et deux arginines de l'hélice α4 formant un réseau de ponts salins (soit huit résidus, E86A,B, D90A,B, R105A,B et R109A,B dans TvGSTO2C). Dans le cas des structures des complexes de TvGSTO2C et de TvGSTO6S avec la naringénine, ces sites sont capables d'accueillir deux molécules de naringénine simultanément. Ces sites sont sélectifs pour le stéréoisomère R présent dans le mélange racémique utilisé lors du trempage. Dans chaque cas, la GSTO interagit *via* liaison hydrogène par la chaîne latérale d'un glutamate (E86 pour TvGSTO2C, E88 pour TvGSTO6S) avec un groupement hydroxyle du noyau chromanone au fond de la poche. Les autres interactions sont essentiellement de nature hydrophobe. Dans les deux isoformes, les deux molécules de naringénine adoptent des conformations voisines. A noter aussi que la proximité des deux molécules dans le site favorise leur stabilisation. Un calcul par DFT[*] de l'énergie d'interaction entre les deux molécules confirme en effet cette interaction favorable dans le cas de TvGSTO6S (-13,4 kcal/mol). Des différences sont retrouvées entre les résidus constituant l'entrée du site L (S116/GSTO2C contre E118/GSTO6S, D113/GSTO2C contre T115/GSTO6S), expliquant les légères variations dans les modes de liaisons entre chaque isoforme. La structure du complexe TvGSTO2S-oxyresvératrol a révélé la présence du *trans*-oxyresvératrol au site H mais également du *cis*-oxyresvératrol au site L, provenant probablement de la photoisomérisation de l'isomère *trans* à la lumière (Maafi et Al-Qarni, 2018). Le *cis*-oxyresvératrol interagit au site L avec les chaînes latérales des résidus E86 et N117 portées par les deux monomères. Le site L de TvGSTO2S présente une architecture similaire aux sites L de TvGSTO2C et TvGSTO6S, suggérant une conservation de ce site au sein des GSTO.

[*]« DFT :'Density Functional Theory'. Il s'agit d'une méthode de calcul quantique pour étudier des densités électroniques. Le but ici était de confirmer la présence simultanée des deux molécules de naringénine sur l'axe 2 du dimère, en validant leur possibilité d'interactions simultanées. Réalisé par E. Aubert, CRM2 »

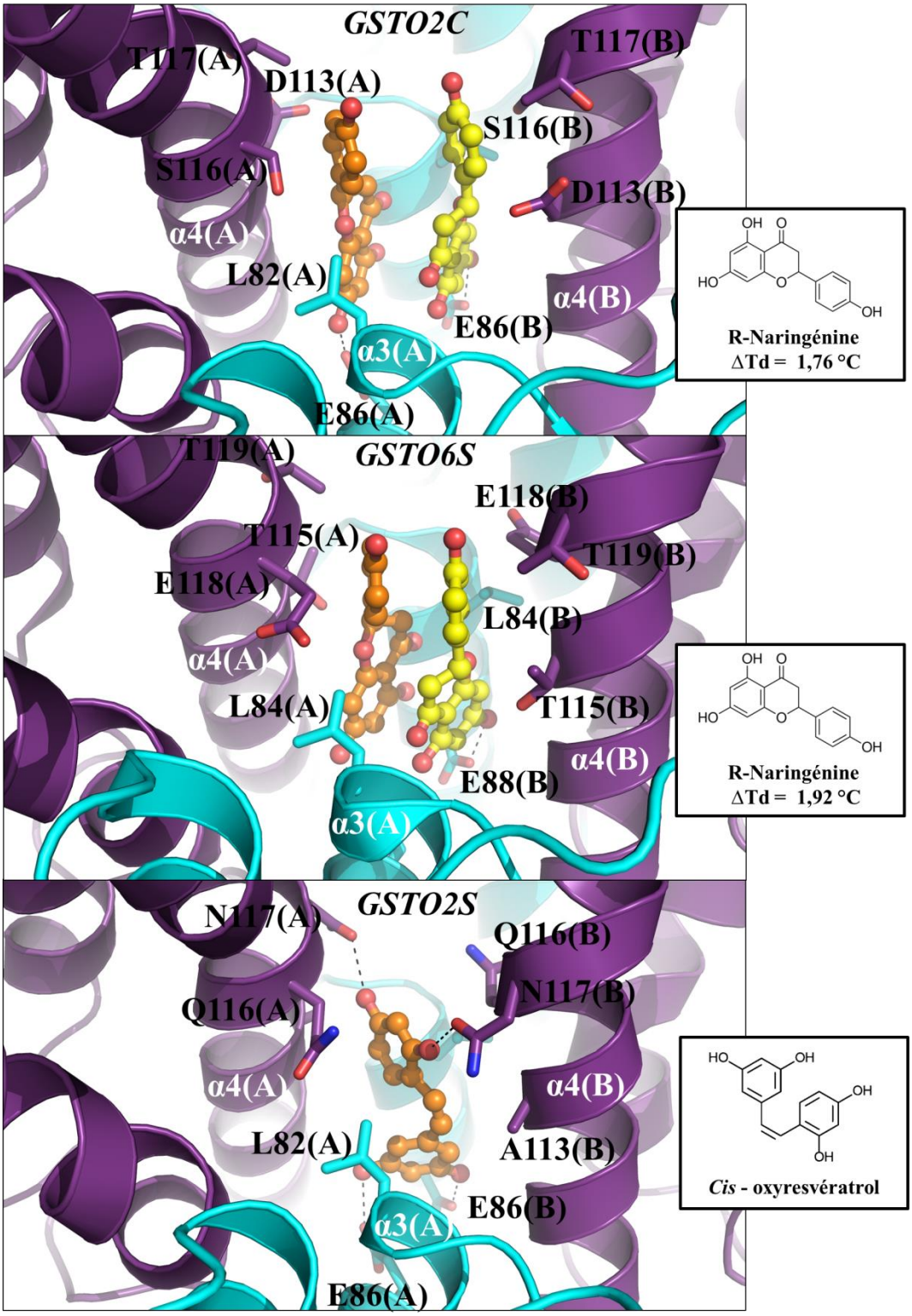


Figure 48 **Structures des complexes de TvGSTO avec les polyphénols végétaux identifiés au site L**

Les sites de fixation ligandine entre les TvGSTO et les différents polyphénols identifiés par TSA sont montrés. La valeur de TSA mesurée à partir de 100 µM de ligand est donnée. Les interactions polaires sont représentées en pointillés. La configuration *cis* de l'oxyresveratrol dans la structure de TvGSTO2S a été obtenue par photo-isomérisation probable de la forme *trans* lors du séchage du ligand avant trempage des cristaux; aussi, aucune expérience de TSA n'a été réalisée à partir de cette forme.

I.4 Publications associées au projet de criblage de ligands des TvGSTO

I.4.a. Publication n°1 : Identification des hydroxybenzophénones et des flavonoïdes comme ligands des isoformes TvGSTO3S et TvGSTO6S

i) Résumé

Les champignons de pourriture du bois ont des systèmes complexes de détoxication qui leur permettent de faire face aux métabolites secondaires produits par les plantes. Bien que le nombre de gènes codant pour les glutathion transférases soit particulièrement élevé dans les champignons lignolytiques, peu d'information sont disponibles sur leurs molécules cibles. Dans cette étude, en combinant des approches biochimique, enzymatique et structurale, les interactions entre des polyphénols et six glutathion transférases du champignon *Trametes versicolor* ont été mises en évidence. Deux isoformes, appelées TvGSTO3S et TvGSTO6S, ont été plus amplement étudiées au niveau structural. En plus du site de fixation du glutathion (site G), chaque isoforme présente deux sites distincts de fixation pour les ligands aromatiques. Le site L est un site étroit et situé à l'interface du dimère et le site H est hydrophobe et particulièrement profond. Dans TvGSTO3S, ce dernier semble optimisé pour la fixation des ligands aromatiques tels que les hydroxybenzophénones. La cristallographie d'affinité a révélé que ce site H retient le flavonoïde dihydrowogonine d'un extrait de merisier sauvage partiellement purifié. En outre, TvGSTO6S lie deux molécules du flavonoïde naringénine dans le site L. Ces données suggèrent que les isoformes de TvGSTO pourraient interagir avec les polyphénols végétaux libérés lors de la dégradation du bois.

ii) Publication

www.nature.com/scientificreports

OPEN

Received: 5 February 2018
Accepted: 4 May 2018
Published online: 31 May 2018

Molecular recognition of wood polyphenols by phase II detoxification enzymes of the white rot *Trametes versicolor*

Mathieu Schwartz[1], Thomas Perrot[2], Emmanuel Aubert[1], Stéphane Dumarçay[3], Frédérique Favier[1], Philippe Gérardin[3], Mélanie Morel-Rouhier[2], Guillermo Mulliert[1], Fanny Saiag[2], Claude Didierjean[1] & Eric Gelhaye[2]

Wood decay fungi have complex detoxification systems that enable them to cope with secondary metabolites produced by plants. Although the number of genes encoding for glutathione transferases is especially expanded in lignolytic fungi, little is known about their target molecules. In this study, by combining biochemical, enzymatic and structural approaches, interactions between polyphenols and six glutathione transferases from the white-rot fungus *Trametes versicolor* have been demonstrated. Two isoforms, named TvGSTO3S and TvGSTO6S have been deeply studied at the structural level. Each isoform shows two distinct ligand-binding sites, a narrow L-site at the dimer interface and a peculiar deep hydrophobic H-site. In TvGSTO3S, the latter appears optimized for aromatic ligand binding such as hydroxybenzophenones. Affinity crystallography revealed that this H-site retains the flavonoid dihydrowogonin from a partially purified wild-cherry extract. Besides, TvGSTO6S binds two molecules of the flavonoid naringenin in the L-site. These data suggest that TvGSTO isoforms could interact with plant polyphenols released during wood degradation.

The microbial degradation of wood has been extensively studied due to its importance in organic matter recycling and its potential valorisation in many industrial domains. This degradation is mainly mediated by fungi and in particular by white-rot fungi which are able to degrade and mineralize all the wood components. Indeed, as early as in the middle of last century, this functional trait has been correlated to the ability of these fungi to secrete extracellular enzymatic systems able to degrade wood polymers[1]. Thanks to the recent release of more than fifty fungal genomes, comparative genomic approaches have confirmed this correlation[2,3]. Beyond these extracellular systems, recent studies have also confirmed the importance of intracellular detoxification systems in the fungal wood degradation process[4]. These systems were thought to play essential roles in wood degradation. They allow fungi to (i) catabolize the oxidized compounds that result from lignin oxidation[5], and (ii) cope with wood anti-microbial compounds, such as flavonoids, stilbenes or terpenes[6,7]. The efficiency of these intracellular detoxification systems seems to be linked to the expansion of multigenic families involved in the oxidation phase such as cytochrome P450 mono-oxygenases and in the conjugation phase such as glutathione transferases (GSTs)[6,8]. Similarly, such expansions are also found in herbivorous insects, where these multigenic families play key functions in the detoxification of plant defense chemicals and also in the evolution of metabolic resistance to chemical insecticides[9–11].

Until now, in wood-decaying fungi, comparative genomic, biochemical, structural or physiological approaches gave only few insights into the function and specificity of these enzymes in the wood degradation process. This lack of knowledge is mainly due to the absence of specific substrates that would allow discrimination between the isoforms. The expansion of the GST family in these fungi mainly concerns three phylogenetic classes, named GSTFuA, Ure2p and GST Omega[12]. We had suggested that the fungal-specific GSTFuA class could be involved in the catabolism of lignin derived molecules[13]. A recent study confirmed that an isoform from *Dichomitus squalens*

[1]Université de Lorraine, CNRS, CRM2, Nancy, France. [2]Université de Lorraine, INRA, IAM, Nancy, France. [3]Université de Lorraine, LERMAB, Nancy, France. Mathieu Schwartz and Thomas Perrot contributed equally to this work. Correspondence and requests for materials should be addressed to C.D. (email: claude.didierjean@univ-lorraine.fr) or E.G. (email: eric.gelhaye@univ-lorraine.fr)

(*Ds*-GST1) selectively cleaves the β-O-4 aryl ether bond of a dimeric lignin model leading to a glutathione derivative[14]. Concerning the Ure2p class, it can be divided in two subclasses with distinct structural and biochemical properties. Ure2pAs possess the classic GSH transferase activity while Ure2pBs display a deglutathionylation activity (Ure2pB)[15]. Interestingly, bacterial orthologs of Ure2pB (named GST Nu) act as glutathione lyases in breaking the β-aryl ether bond of lignin[16]. GSTOs are involved in detoxification pathways via deglutathionylation reactions[17] and two GSTO isoforms of *Phanerochaete chrysosporium* (PcGSTO3 and PcGSTO4) bind terpenes[18]. In *Trametes versicolor*, some GSTO isoforms interact with different wood extractives[19]. These interactions could give insights into the chemical composition of the extracts.

To further explore this issue, we have conducted biochemical and crystallographic studies on the interactions of six GSTOs from *T. versicolor* with chemical libraries. We showed that these GSTOs exhibit distinct affinity patterns, particularly with benzophenones and flavonoids. An affinity crystallography approach allowed the isolation of a flavonoid from a partially-purified wild-cherry tree extract. This ligand specific to one GSTO isoform was characterized as dihydrowogonin using multiple approaches. All these results and the recent literature support the conclusion that GSTs of this class interact with wood polyphenolic compounds.

Results and Discussion

TvGSTO3S interacts with hydroxybenzophenones. Thermal shift assay (TSA) is a high-throughput ligand-screening method based on the modification of protein thermal denaturation. According to a gradient of temperature, the denaturation is followed by monitoring fluorescence enhancement of a probe (SYPRO Orange) that binds to protein hydrophobic patches upon denaturation process. This TSA method has been successfully used to detect interactions between proteins and libraries of molecules[20]. It allowed us to identify chemical families of compounds that interact with TvGSTOs and prompted us to investigate more deeply the case of TvGSTO3S with hydroxybenzophenones (HBPs), in particular by conducting a structural analysis of protein-ligand complexes.

First, the interactions between a chemical library of 27 compounds and six TvGSTOs were explored using TSA (Supplementary Fig. S1). The tested compounds were chosen either for their presence in wood or their reactivity with GSTs[19]. The six TvGSTOs (named TvGSTO1S to TvGSTO6S) used in this study are representatives of the twelve TvGSTOs that have a catalytic serine, while four others have a cysteine instead. The obtained results show patterns of interaction that distinguish each isoform from the others. Indeed, a few compounds significantly increased the stability of TvGSTO1S, 3S and 6S, *i.e.* they show a variation of the denaturation temperature $\Delta Td > 5\,°C$. On the contrary, most chemical compounds induced a destabilizing effect on TvGSTO5S and 2S ($\Delta Td < 0\,°C$), whereas the chemical library had little impact on TvGSTO4S thermal stability. In the case of TvGSTO3S, the compounds that increased protein stability all belong to the same chemical family, namely hydroxybenzophenones (HBPs). Benzophenones are present in plant extracts[21] and also in wood extractives, for example in oak heartwood[22].

A set of commercially available HBPs with various numbers and positions of hydroxylation on their rings A and B (Table 1) was tested by using TSA and the six TvGSTOs. It confirmed that TvGSTO3S has a significant affinity for HBPs. For this isoform, a group of molecules (2,4-, 3,4-, 2,3,4- and 2,4,4′- HBPs) corresponding to compounds with at least two hydroxyl groups on ring A and no more than one hydroxylation on ring B stand out for their causing large increases in the melting temperature of TvGSTO3S. Oppositely, 2,2′-, 4,4′-, 2,2′,4,4′- HBPs and the unsubstituted benzophenone had little or no effect on the thermal stability of TvGSTO3S. The replacement of the 4-hydroxyl group of 2,4-HBP by a methoxy group nullified the observed thermal shift.

For further insights into these findings, a structure/function relationship study focused on the analysis of TvGSTO3S-ligand interactions was initiated. The structure of apo TvGSTO3S was determined by X-ray crystallography at a resolution of 1.35 Å (Supplementary Table S1). It has a typical GST fold where the N-terminal thioredoxin domain of one monomer cross-interacts with the C-terminal all-helical domain of the second one, and vice-versa (Fig. 1). TvGSTO3S displays the highest sequence identity with Omega GSTs, though its closest structural homologs identified by PDBeFold[23] are Tau GSTs. The resemblance between these two classes was already discussed for a wheat Tau GST[24]. However, TvGSTO3S has unique features that distinguish it from previously described GSTs. They mainly include an elongation of the loop between β_3 and β_4 and an additional helix α_6' (Supplementary Fig. S2).

The N-terminal end of TvGSTO3S helix α_1 harbors a serine as the catalytic residue instead of the cysteine found in the other GSTOs structurally characterized so far[17,25]. It enables efficient GSH-transferase activity towards usual synthetic substrates and disables reductase activity (Supplementary Table S2). In order to get a detailed picture of the ligand binding sites of TvGSTO3S, we determined the structures of several complexes with free GSH, or some glutathionylated derivatives GS-R (glutathionyl-dinitrobenzene GS-DNB and glutathionyl-phenylacetophenone GS-PAP), or some of the HBPs identified by TSA (2,4-, 3,4-, 2,3,4-, 2,4,4′-HBPs) (Supplementary Table S1). TvGSTO3S harbors a canonical GSH binding site (G-site) made up of polar residues from the N-terminal domain which stabilizes the glutathionyl moiety (GS-) of the tested ligands (Fig. 1, Supplementary Figs S3 and S4). On the contrary, the hydrophobic binding site (H-site) which hosts the -R group of the GS-R ligand has a singular shape relative to most GSTs. While a large open valley is usually observed in a cleft between the two domains[26], TvGSTO3S exhibits a well-delineated cavity deeply inserted between helices α_4 and α_6 of the C-terminal domain. The crystal structures of the different complexes show that this pocket is perfectly suited to accommodate polyaromatic ligands, due to its strong hydrophobic character, while two polar residues are found at the entrance close to the G-site (Supplementary Fig. S3). The phenylacetophenone group of GS-PAP fully fills the cavity, on the contrary to the more polar -R group of GS-DNB that does not enter it and is only slightly stabilized at its entrance, at the interface with the G-site (Supplementary Fig. S4).

The four HBPs (2,4-, 2,3,4-, 3,4- and 2,4,4′- HBPs) that were previously identified by TSA also bind in the TvGSTO3S H-site (Fig. 2 and Supplementary Fig. S5). Overall, their phenyl ring B (Table 1) sits deep at the bottom

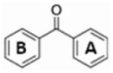
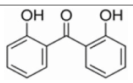
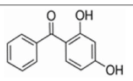
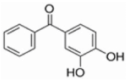
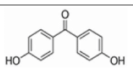
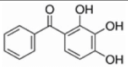
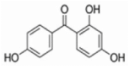
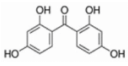
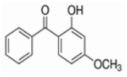
Molecules	Structures	TvGSTO1S	TvGSTO2S	TvGSTO3S	TvGSTO4S	TvGSTO5S	TvGSTO6S
Benzophenone		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2,2′-Dihydroxy benzophenone		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2,4-Dihydroxy benzophenone		2.05 °C	2.56 °C	5.69 °C	n.s.	n.s.	1.63 °C
3,4-Dihydroxy benzophenone		0.98 °C	n.s.	4.36 °C	2.77 °C	n.s.	n.s.
4,4′-Dihydroxy benzophenone		n.s.	n.s.	1.67 °C	n.s.	n.s.	n.s.
2,3,4-Trihydroxy benzophenone		1.13 °C	2.92 °C	4.96 °C	1.76 °C	−12.89 °C	4.15 °C
2,4,4′-Trihydroxy benzophenone		2.74 °C	2.03 °C	2.87 °C	n.s.	n.s.	1.28 °C
2,2′,4,4′-Tetrahydroxy benzophenone		3.74 °C	0.74 °C	n.s.	n.s.	n.s.	3.40 °C
2-Hydroxy-4-methoxy benzophenone		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Table 1. Summary of the results obtained for TvGSTOs with benzophenone compounds by thermal-shift assays. A ΔTd value is only given if the denaturation temperature is significantly modified in the presence of compounds, with respect to incubation with DMSO only. "n.s." means that the denaturation temperature has not changed significantly.

of the pocket and interacts via π-stacking with aromatic amino acid side chains. Their di- or tri-hydroxylated phenyl ring A is closer to the entrance of the cavity and forms hydrogen bonds with polar side chains and water molecules. In more detail, two distinct HBP conformations are observed (*i*) for 2,4-, 2,3,4- and 3,4-HBPs, and (*ii*) for 2,4,4′- HBP, respectively (Fig. 2). Both conformations are common for this family of molecules in the solid state[27]. In (*i*), the central ketone group of HBPs is stabilized via a well conserved water molecule present in all structures, while this interaction no longer exists in (*ii*). For the latter, a new hydrogen bond allows accommodation of the 4′ hydroxyl group at the bottom of the cavity. The presence of this additional 4′ hydroxyl group probably accounts for the distinct conformation (*ii*) observed in the H-site. Finally, another attempt was made to get the structure of a complex between TvGSTO3S and the compound 2,2′, 4,4′-HBP, but no electron density was observed for this putative ligand. These results not only correlate with TSA, but also with the binding affinities of HBPs for TvGSTO3S which were assessed from their capacity to inhibit the GSH-transferase activity towards phenethyl isothiocyanate (PEITC). Indeed, the same four HBPs differentiate from the others in the sense that only their K_i values were measurable and found in the μM range (Supplementary Table S3). Altogether, our crystallographic and enzymatic results suggest that the TvGSTO3S H-site is selective for HBPs whose ring B is totally hydrophobic and stabilized at the bottom of the pocket or bears one hydroxyl group in para position whereas ring A is di- or tri- hydroxylated and stabilized at the entrance of the H-site. As shown with TSA, other TvGSTO isoforms also interact with HBPs, each with its own selectivity. The discovery of HBPs as potential ligands for TvGSTOs echoes previous results concerning metabolization of these small molecules by the fungus. For instance, the sunscreen agent BP3 (2-hydroxy, 4-methoxy benzophenone) is metabolized by *T. versicolor* into various HBPs (2,4-, 4,4′- and 4- HBPs) by cytochrome P450 with no oxidation by extracellular laccases[28] suggesting the importance of intracellular systems to detoxify benzophenones, including phase II enzyme GSTs.

TvGSTOs interact with flavonoids. In a recent paper, we suggested that GSTs from saprotrophs could interact with flavonoids, which are phenolic compounds as HBPs and present in wood extracts[19]. In order to refine our results, TSA was used to test a set of commercial flavonoids for their interaction with TvGSTOs. Some of the putative complexes were further investigated by X-Ray crystallography. For the first time in GSTs, one structure revealed a symmetrical ligandin site (L-site) filled with a pair of interacting ligands.

As observed for HBPs, TSA shows that each TvGSTO has its own interaction profile with flavonoids and that these interactions are largely related to the number and positions of the hydroxyl groups on the aromatic rings (Supplementary Table S4). While TvGSTO4S and 5S show variable or weak responses to flavonoids, TvGSTO1S,

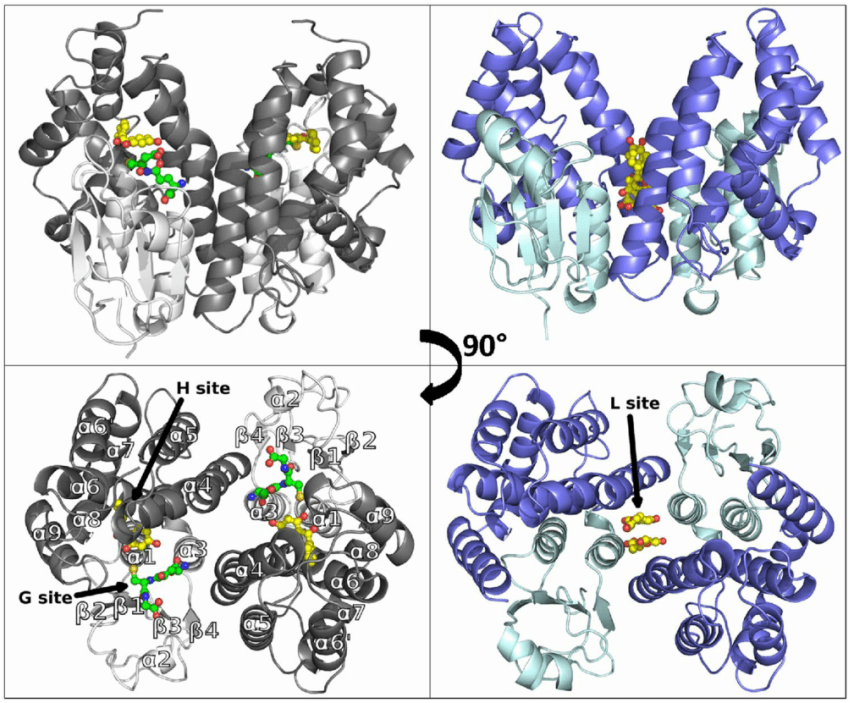


Figure 1. Overall views of *Trametes versicolor* GSTO3S structure in complex with glutathione and 2,4 hydroxy benzophenone (left panels) and GSTO6S structure in complex with naringenin (right panels). In each case, structures are depicted in cartoon mode with ligands shown as spheres and sticks (glutathione in green, 2,4-HBP and naringenin in yellow). N-terminal domains are shown in light colors (white for GSTO3S, cyan for GSTO6S) and C-terminal domains are shown in deeper colors (grey for GSTO3S, blue for GSTO6S). Black arrows indicate positions of glutathione binding site (G-site), hydrophobic binding site (H-site) and ligandin site (L-site). In each case is represented one physiological dimer that typifies the structure of GSTs where the N-terminal domain (secondary structure $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$) of one monomer cross-interacts with the C-terminal domain ($\alpha_4\alpha_5\alpha_6\alpha_6'\alpha_7\alpha_8\alpha_9$) of the second one, and vice-versa.

2S, 3S and 6S mainly show positive shifts of their denaturation temperature. For instance, all molecules but catechin affected TvGSTO6S. Interestingly, these flavonoids were recently detected by mass spectrometry in *T. versicolor* fructifications. Their presence might arise from trees or soils on which the fungi grow[29].

The crystal structure of TvGSTO6S was solved at 1.45 Å resolution (Supplementary Table S1). As expected, it reveals the same overall fold than TvGSTO3S (rmsd 0.78 Å, 183 aligned Cα per monomer). While G-sites are made of similar polar residues, differences are observed at the H-site and at the interface between monomers (Fig. 3). The lack of a structure with a ligand that binds in TvGSTO6S H-site precluded its fine description. However, when compared with the equivalent region of TvGSTO3S, significant differences are readily apparent in TvGSTO6S. Most of the residues that line the benzophenone binding site of TvGSTO3S are different. In addition, the extension of the C-terminal α-helix by two extra turns leads to an open and shallow H-site, contrary to the buried H-site in TvGSTO3S covered and closed by the loop corresponding to the extra turns. There are also slight differences between the dimer-interfaces of both isoforms. Most residues at the interface are conserved, except for a few ones situated in helix α4 (Fig. 3). The larger residues of TvGSTO6S (L112 and T115 instead of T110 and A113 in TvGSTO3S) and the loss of the inter-monomer interaction between Y118 (replaced by E120 in TvGSTO6S) and E80 present in TvGSTO3S probably account for a more open dimer in TvGSTO6S. Its interface residues form a pocket centered on the two-fold axis of the dimer (rectangular section of 8×6 Å^2), larger than the equivalent region found in TvGSTO3S (rectangular section of 7×4 Å^2, Supplementary Fig. S6). This region hosts a third binding site named the ligandin site (L-site) in the human GST Omega 1 (hGSTO1)[30].

In order to determine the structure of a complex between TvGSTO6S and a flavonoid, several candidate molecules were chosen for soaking experiments. Crystals of TvGSTO6S were unstable and brittle, and required some tricks to prepare complexes. Among the trials performed, one was successful where the droplet that gave rise to TvGSTO6S crystals was deposited for a few hours on a surface previously coated with naringenin (chemical structure in Supplementary Table S4). The structure of the complex was solved at a resolution of 2.3 Å (Figs 1 and 4, Supplementary Table S1). Two naringenin molecules bind in the pocket described above, which constitutes the L-site of TvGSTO6S (Figs 1 and 4). Each naringenin aromatic ring stacks with its symmetrical equivalent in an energetically-favored association (−13.64 kcal/mol calculated by DFT). A pair of stacked flavonoids was already observed bound to dihydroflavonol 4-reductase, however in a head to tail arrangement in the active site of the

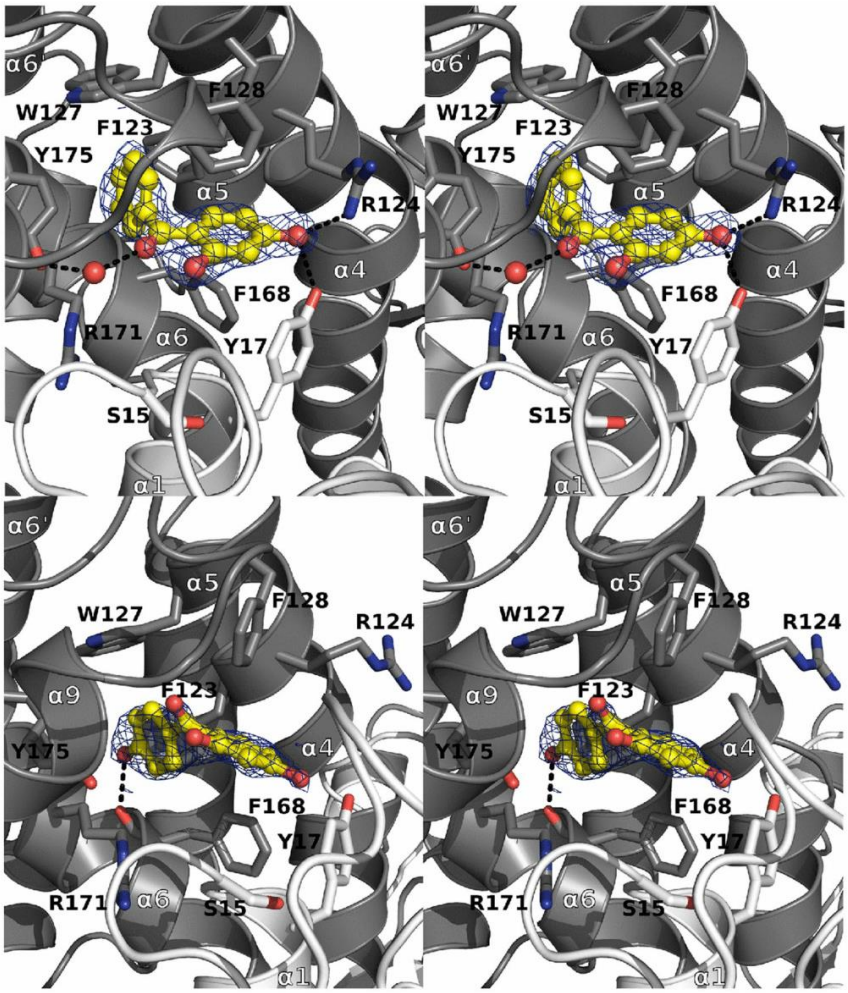


Figure 2. Binding of 2,4- and 2,4,4′-hydroxy benzophenones in GSTO3S hydrophobic binding site. Stereoviews of sections of the GSTO3S complex structures with 2,4-HBP (top view) and 2,4,4′-HBP (bottom view) are shown. TvGSTO3S H-site is a well-delineated cavity deeply inserted in between helices α_4 and α_6 of the C-terminal domain. It is perfectly suited to accommodate polyaromatic ligands, due to its strong hydrophobic character given by the aromatic side chains of F123, W127, F128, F168, the aliphatic parts of R171 and Y175, completed with polar residues (Y17 and R124) at the cavity entrance close to the G-site. Polar intermolecular contacts are materialized as dashed lines. Surrounding side chains are represented in sticks. HBPs are shown as yellow sticks and spheres. 2mFo-DFc composite omit maps shown at 1.0 σ around HBPs were calculated by PHENIX.

enzyme, with no dimer symmetry[31]. In TvGSTO6S, the naringenin pair fits the apolar environment formed by the neighboring aliphatic part of the side chains of both monomers (L84, V85, R111, T115, E118 and T119) while a hydrogen bond with E88 stabilizes one of the flavonoid hydroxyl groups. L-sites at the dimer interface were previously described for hGSTO1 (pdb code 4is0)[30] and for *Arabidopsis thaliana* GST Phi 2 (AtGSTF2, pdb code 5a4v)[32] however with one ligand only. In hGSTO1, the L-site takes place at a location similar to TvGSTO6S and shows conserved patches of hydrophobic residues along helix α3 (Supplementary Fig. S2). Interestingly, E88 that stabilizes naringenin in TvGSTO6S is conserved in hGSTO1 (E91) where its side chain is found near the nitrophenacyl moiety of the glutathione adduct. The case of AtGSTF2 is quite different. A quercetin molecule sits at the opposite side of the dimer interface, near the C-terminal end of helix α3 and the N-terminal end of helix α4.

Structural comparisons suggest that Omega GSTs have two hydrophobic ligand binding sites: an H-site near α4 and α6 and an L-site at the dimer interface. In the case of TvGSTO6S, the L-site can host flavonoids. The alignment of the solved structures with the sequences of TvGSTO1S, −2S, −4S and −5S shows variations at the putative H- and L-sites (Supplementary Fig. S2), which could explain the isoform-specific TSA patterns.

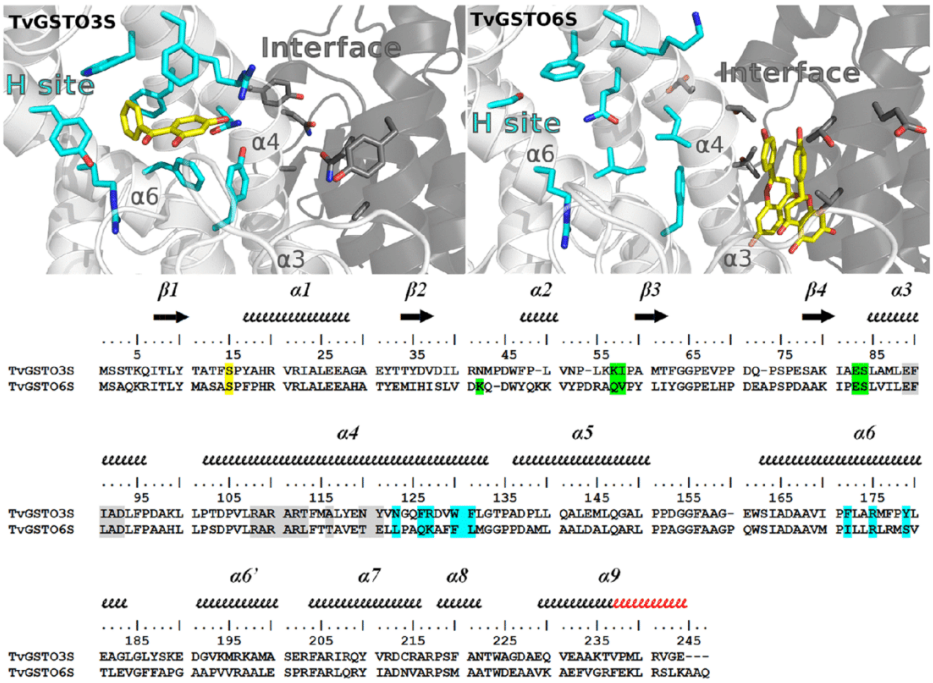


Figure 3. Structural comparison of TvGSTO3S and −6S isoforms showing differences between H-sites and L-sites. The structures shown are TvGSTO3S in complex with 2,4-HBP (top left) and TvGSTO6S in complex with naringenin (top right). In both cases, helix α9 and C-terminal tail were removed from structures for clarity. Monomers A and B are colored white and grey, respectively. 2,4-HBP and naringenin molecules are represented as yellow sticks. H-site and L-sites residues are colored respectively cyan and grey in the structures (highlighted respectively cyan and grey on the structural alignment). G-site residues are highlighted green in the sequences and are not shown on structures for clarity. Catalytic serine is highlighted yellow in the sequences. Secondary structures are labelled and shown using arrows (β-strands) and squiggles (helices). Secondary structure elements are based on TvGSTO6S. Helix α9 extra turns in TvGSTO6S are colored red on the sequence alignment.

Fishing of a flavonoid from a wild-cherry heartwood extract by affinity crystallography. One of the biggest challenges in GST characterization is the identification of natural ligands[33]. Affinity crystallography is a very new method that was recently conceived to select and identify new inhibitors from natural crude extracts as potent drug scaffolds for pharmaceutical targets[34]. We successfully applied this method on TvGSTO3S crystals to isolate the flavonoid dihydrowogonin from a partially purified wild-cherry extract.

Acetonic extracts of wild-cherry heartwood were fractioned by reverse chromatography. The potential inhibition of the various fractions on TvGSTO esterase activity was analyzed using both chloromethylfluorescein diacetate (CMFDA) and methylumbelliferyl acetate (MUA) as substrates (Fig. 5). In previous studies[13,19], these substrates had increased detection sensitivity of fluorescence and avoided quenching effects of wood extracts. The fractions that eluted after 46 and 47 min induced a strong inhibition of both measured activities. The pooled mixture was tested for its ability to interact with TvGSTO3S using TSA. A positive 4.8 °C shift of the observed ΔTd confirmed the presence of potential TvGSTO3S ligands. Analysis of this solution by LC-MS revealed the presence of at least two compounds with molecular masses of 254 and 286 g.mol^{-1} respectively (Fig. 5). The 46–47 min eluate also analysed by H^{1}NMR (spectrum shown in Fig. 5) exhibited characteristic signals of a flavanone skeleton with, in addition, typical methyloxy group singlets at 3.75 or 3.81 ppm. At this stage, several compounds already described in the extractives composition could correspond to such data (*e.g.* dihydrowogonin or sakuranetin)[35–37] so that it was impossible to definitively assign the molecular structures, particularly concerning the MeO group position either on the ring A or B (Fig. 5).

Affinity crystallography was then used to further elucidate the interactions between TvGSTO3S and compounds present in the 46–47 min eluate. This approach assumes that a complex can form when the protein is mixed with a partially purified mixture of molecules containing potential ligands. We re-suspended the dried eluate with a minimal volume of DMSO to obtain a concentrated mixture suitable for TvGTO3S crystallization. The addition of 2.5% of the concentrated solution in the mother liquor produced co-crystals as observed from the electron density peak in the H-site. High quality of the signal together with the NMR and MS data led to unambiguous identification of dihydrowogonin (Fig. 5). This flavonoid was previously identified in *Prunus avium* heartwood and described as a flavanone[35]. Interestingly, its corresponding flavone (wogonin) was shown to strongly inhibit the GSH-transferase activity of TvGSTO3S towards PEITC ($K_I = 2.89 \pm 1.24$ μM, Supplementary

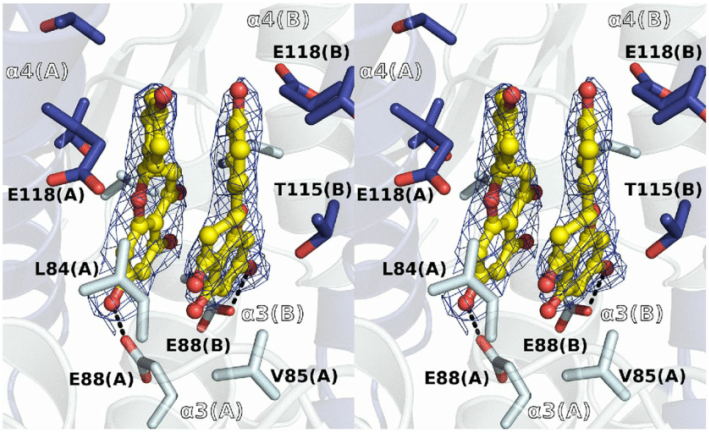


Figure 4. Binding of naringenin in GSTO6S non-catalytic ligandin site. Stereoview of a section of the GSTO6S complex structure with a pair of naringenin molecules is shown. GSTO6S ligandin site is an hydrophobic pocket inserted at the interface of the dimer, between helix α3 and α4 of both monomers around the dimer 2-fold axis. The apolar environment is formed by the aliphatic part of the side chains of both monomers (L84, V85, T115, E118) while a hydrogen bond with E88 stabilizes one of the flavonoid hydroxyl groups. Polar intermolecular contacts are materialized as dashed lines. Surrounding side chains are represented in sticks. Naringenin is shown as yellow sticks and spheres. 2mFo-DFc composite omit map shown at 1.0 σ around the ligand was calculated by PHENIX.

Table S5), while no inhibition was detected for TvGSTO6S. Since in the latter a stabilizing effect is still observed by TSA, the ligand should not bind in its catalytic site (*i.e.* H-site). In TvGSTO3S, the unsubstituted ring of dihydrowogonin sits in the bottom of the H-site like the ring B of HBPs (see above) (Supplementary Fig. S7). The opposite ring points towards the entrance of the cavity the polar residues of which stabilize its methoxy and hydroxyl groups by direct or water-mediated hydrogen bonds. The hydroxyl group situated at position 7 is only 2.6 Å from the glutathione sulfur atom. This short distance questions a possible catalysis with a related substrate. The combined use of X-ray diffraction, NMR and MS succeeded in identifying dihydrowogonin as a natural molecule that originates from wild-cherry heartwood and that tightly binds TvGSTO3S.

Conclusion

In this study, we demonstrated at the biochemical and structural level that *T. versicolor* GSTs that belong to the Omega class interact with polyphenolic compounds found in wood, and in particular with flavonoids such as dihydrowogonin and naringenin. Indeed, white-rot fungi such as *T. versicolor* have to cope with potentially toxic tree secondary-metabolites mainly constituted of polyphenolic compounds, which accumulate in different parts of the wood (heartwood, knots). The molecular interactions between TvGSTOs and polyphenols appear to be very diverse. They potentially involve two structural sites for each isoform. It is tempting to establish a correlation between the diversity of these interactions and the extension of the Omega class (at least 16 isoforms) in *T. versicolor*, which encounters a large diversity of polyphenols in its natural environment. However, the exact function of this GST network remains unclear since no catalytic activity against the tested polyphenols has been detected. The ability of TvGSTOs to bind polyphenols through different sites suggests that fungal GSTs could be involved in the transport of various polyphenols. In plants, GSTs were shown to facilitate the flavonoid transport from the cytoplasm into the vacuole[38]. These GSTs could act as flavonoid and glutathione carriers, providing ABC transporters with both molecules for their co-transport. This transport requires free glutathione but no glutathionylation activity[39]. Our study suggests that fungal GSTs as their plant homologues could be involved in the transport and sequestration of flavonoids.

Methods

Reagents. All pure molecules together with the fluorescent marker SYPRO® Orange used in TSA were purchased from Sigma-Aldrich (St. Louis, MO, USA), except for wogonin provided by Extrasynthese (Genay, France).

Production and purification of proteins. The production in *E. coli* (Rosetta2 DE3 pLysS strain, Novagen) and purification of the six selected TvGSTOs (accession number in the JGI database: TvGSTO1S: Tv75639; TvGSTO2S: Tv56280; TvGSTO3S: Tv48691; TvGSTO4S: Tv65402; TvGSTO5S: Tv54358 and TvGSTO6S: Tv23671) were performed as explained previously[19].

Study of the thermostability of TvGSTOs. The experiments were performed in 96 well microplates (Harshell, Biorad) and the measurements carried out using a real time PCR detection system (CFX 96 touch, Biorad). The assays were performed as follows: 5 µL of Tris-HCl (150 mM) pH 8 buffer, 2 µL of pure molecules dissolved in DMSO (final concentration: 0.8 mg/mL for molecules of the chemical library; 100 µM for benzophenones and flavonoids), 5 µL of proteins (contained in Tris-HCl 150 mM, pH 8; final concentration: 40 µM for

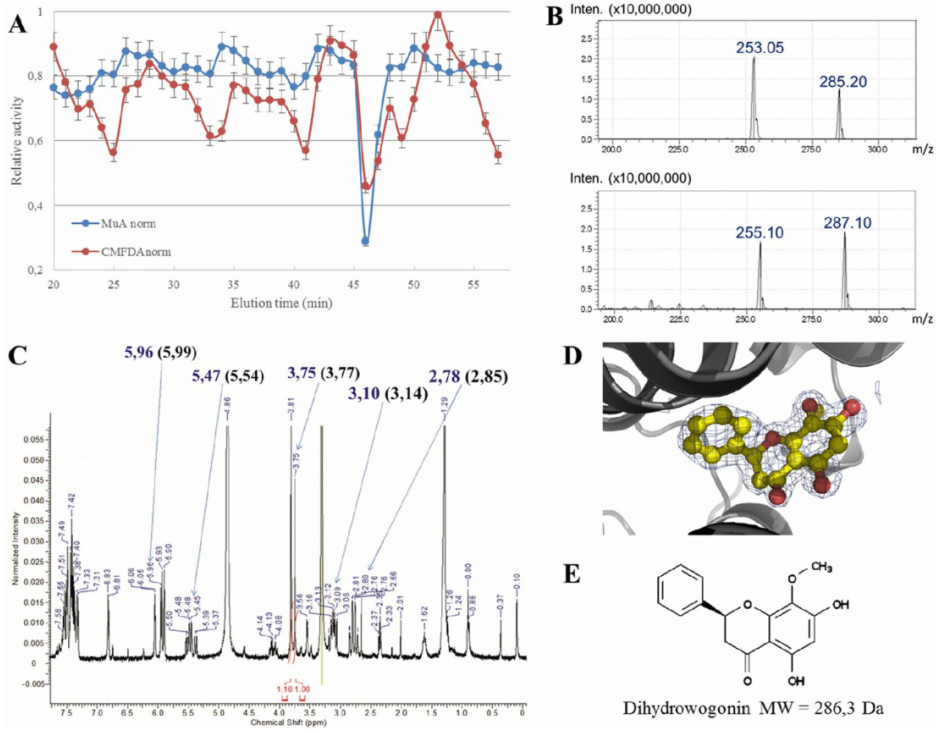


Figure 5. Combined approach including affinity crystallography revealed dihydrowogonin bound to GSTO3S hydrophobic site. (**A**) Normalized inhibition of esterase activity with substrates CMFDA and MUA is shown. Fractions 46–47 min that inhibited both activities were selected for further analysis. (**B**) MS analysis in positive (bottom panel) and negative (top panel) modes revealed two major compounds. (**C**) ^{1}H-NMR spectrum showed the structural features of flavanones. Affinity crystallography allowed the elucidation of the flavanone dihydrowogonin. Its ^{1}H-NMR data[37] are indeed found on the spectrum of the mixture: numbers in parenthesis are the typical chemical shifts for dihydrowogonin and numbers in blue correspond to the values obtained in the present study. Integration values of 2 methoxy groups allowed to evaluate maximum abundance of dihydrowogonin in the fractions (numbers in red). (**D**) Electron density of dihydrowogonin in structure of GSTO3S crystallized in presence of 46-47 eluate. The map shown is a 2mFo-DFc composite omit map contoured at 1 σ. (**E**) Chemical structure of dihydrowogonin.

the tests with the chemical library; 10 μM for the other tests), 2 μL of SYPRO orange (previously diluted 80−fold in ultra-pure water) and 11 μL of ultra-pure water for a total volume of 25 μL per well. The plate was centrifuged 30 seconds at 4000 g. Fluorescence was measured (excitation: 485 nm; emission: 530 nm) every minute starting from 3 minutes at 5 °C while increasing temperature from 5 to 95 °C with a step of 1 °C per minute. The denaturation temperature (Td), which corresponds to the temperature where 50% of the total fluorescence is measured, was determined by using the non-linear regression Boltzmann sigmoidal model in GraphPad Prism 6 software for data obtained in presence of potential ligands, while the reference was considered for similar experiments conducted by adding DMSO only.

Inhibition Kinetics. Competition tests between phenethyl isothiocyanate (PEITC) and benzophenones or flavonoids were performed in a final volume of 500 μL. Inhibition tests of the glutathionylation activity were assayed with various concentrations of PEITC (25–300 μM) and a fixed concentration of GSH (1 mM) in presence or not of inhibitors. These experiments were carried out in 100 mM pH 6.4 phosphate buffer at 25 °C by analyzing the glutathionylated product which absorbs at 274 nm. Basal activity of samples containing GSH, PEITC and benzophenones was subtracted from the enzyme catalyzed rates. The inhibition constants (Ki) were calculated using the GraphPad software with the nonlinear regression based on the mixed model inhibition.

Inhibition tests of the esterase activity of TvGSTO3S with CMFDA or MUA as substrates in presence of extract fractions were performed in 96-well microplates. The emission of fluorescence at 517 nm and 460 nm after an excitation at 492 nm and at 355 nm for CMFDA and 4-MUA respectively was followed using microplate reader (2030 Multilabel Reader Victor X5, PerkinElmer). Reactions were performed in 200 μL of Tris-HCl pH 8 (30 mM), EDTA (1 mM) buffer, with 0.5 μM of CMFDA, 5 μM of 4-MUA and 1 mM of GSH. Using both fluorescent substrates, TvGSTO3S activity was measured in presence or in absence of 2 μL of the tested fractions. The ratio between the two slopes (ΔRFU/min, Relative Fluorescence Unit) has been used to determine the potential inhibition.

Fractionation of wild-cherry hardwood acetonic extracts by HPLC. Wild-cherry hardwood acetonic extracts were fractioned by high-performance liquid chromatography. 100 µL of extract at 10 mg/mL were fractioned by reverse chromatography using Kinetex biphenyl column (250 × 4,6 mm) previously equilibrated with H_2O/Formic acid 0.1% buffer. The molecules adsorbed to the column were eluted with the help of a gradient of methanol (from 0 to 100%). Collected fractions were evaporated using SpeedVac™ (UniEquip) and finally dissolved in DMSO.

Crystallogenesis experiments. Crystallization of TvGSTO3S was assayed by the microbatch under oil method at 278 K. TvGSTO3S (13 mg/mL) crystallized by mixing 1 µL of protein with 3 µL of commercial solution consisting in 30% (w/v) PEG 400, 0.2 M calcium acetate in 0.1 M pH 4.5 acetate buffer (Wizard Classic Screen 1, Rigaku). Crystals of the complexes of TvGSTO3S with glutathione (TvGSTO3S - GSH), TvGSTO3S with glutathionyl-dinitrobenzene (TvGSTO3S - GS-DNB), TvGSTO3S with hexyl-glutathione (TvGSTO3S - GS-hexyl) and TvGSTO3S with glutathionyl-phenylacetophenone (TvGSTO3S - GS-PAP) were obtained by soaking apo TvGSTO3S crystals during one hour in the mother liquor containing 10 mM ligand (0.5 mM in the case of GS-PAP). Crystals of the complexes of TvGSTO3S with HBPs (2,3,4-trihydroxybenzophenone, 3,4-dihydroxybenzophenone, 2,4-dihydroxybenzophenone and 2,4,4′-trihydroxybenzophenone) were obtained by co-crystallizing the protein pre-incubated for 30 min with 10 mM ligand. Complex of TvGSTO3S with dihydrowogonin was obtained in a similar way by using 10 mg/mL of partially-purified wild-cherry hardwood extract.

First screening of the TvGSTO6S crystallization conditions was performed by using an Oryx 8 robot (Douglas Instruments) to implement sitting drops with commercial kits of various solutions. Crystals were then optimized by the hanging drop method. The droplet was prepared by mixing 1 µL of TvGSTO6S (26 mg/mL) with 0.2 µL of a crystal seed stock obtained by crushing the droplet content of the best hits of the screening step and with 1 µL of a crystallization solution consisting in 25% (w/v) PEG 1500, 0.1 M pH 6.5 MMT buffer (containing DL-malic acid, MES and Tris base in the molar ratios 1:2:2, respectively). The reservoir contained 1 mL of the same crystallization condition. Attempts using classical co-crystallization or soaking experiments failed to prepare crystals of TvGSTO6S complexes with flavonoids. A new strategy was developed based on the dry co-crystallization method[40]. In this original 'dry soaking' technique, 0.2 µL of naringenin (100 mM) solubilized in DMSO was deposited on a cover slide and left to complete evaporation. Then, one TvGSTO6S crystal together with 1 µL of its mother liquor was dispensed on the dried naringenin allowing partial ligand resolubilization. The cover slide was then reinstalled above the reservoir that initially allowed crystallization until crystal harvest (ca. 1 day).

Data collection, processing and refinement. TvGSTO3S and −6S crystals were flash-frozen after a quick soaking in their mother liquor complemented with 20% (v/v) glycerol as cryoprotectant. Primary X-ray diffraction experiments were carried out in-house on a laboratory diffractometer (Agilent SuperNova with CCD detector). Data collection up to 2.5 Å resolution allowed preliminary analysis, especially for ligand screening in TvGSTO3S active site. High resolution diffraction experiments were carried out on the ESRF beamlines FIP BM30A and ID30B (Grenoble, France). TvGSTO3S and −6S native crystals diffracted up to 1.35 Å and 1.48 Å, respectively. Data sets were indexed, integrated and scaled with XDS[41]. The structure of TvGSTO3S was solved by molecular replacement using MR BUMP automated pipeline from CCP4 suite[42] with the coordinates of poplar GST Lambda 3 (PDB code 4PQI) as the search model. The electron density of a small molecule that we failed to identify was observed in the H-site. It probably bound to the enzyme during the purification process and was not modelled in the electron density. Some of the structures of TvGSTO3S in complex with ligands displayed electron density corresponding to glutathione remaining from the purification process. When present, this residual GSH was modelled, sometimes with partial occupancy. The structure of TvGSTO6S was solved by molecular replacement using PHASER with the coordinates of TvGSTO3S. All structures were refined with PHENIX[43] and built with COOT[44]. Restraint files for ligands were generated with phenix.elbow and the grade server (URL http://grade.globalphasing.org/cgi-bin/grade/server.cgi). In all of the concerned structures, the occupancies of the ligands added by co-crystallization or soaking techniques were set to 1 and the corresponding B factors were compatible with full presence of the molecules in their binding sites (except for 2,3,4-HPB where the occupancy is 0.8). Validation of all structures was performed with MolProbity[45] and the wwPDB validation service (http://validate.wwpdb.org). Coordinates and structure factors have been deposited in the Protein Data Bank. Data-collection and refinement statistics of all structures are shown in Supplementary Table S1. Stereo images of a portion of 2mFo-DFc electron density maps are shown in Supplementary Table S6 to assess quality of the structural data. All figures were prepared by using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Density functional theory (DFT) calculation of naringenin dimer stabilization in TvGSTO6S (Gaussian09 software). The molecular structure of the dimer of naringenin molecules was extracted from the experimental X-ray structure of TvGSTO6S. Geometry optimization of the hydrogen atoms alone was then performed[46] at the DFT level of theory in vacuum (*i.e.* no environment effect were taken into account), employing the B3LYP functional[47] completed with Grimme's D3 dispersion correction[48]; the 6–111 G(d,p) basis set was used, and basis set superposition errors were corrected by the counterpoise method of Bernardi[49]. The complexation energy at the converged geometrical structure was −13.64 kcal/mol, indicating a strong stabilization of the naringenin molecules within their dimer as found in TvGSTO6S.

Data Availability. The atomic coordinates of the crystal structures from this publication have been deposited to the Protein Data Bank (https://www.rcsb.org) and assigned the PDB codes 6F43, 6F4B, 6F4F, 6F4K, 6F51, 6F66, 6F67, 6F68, 6F69, 6F6A, 6F70 and 6F71.

Accession codes. Accession numbers of TvGSTOs in the JGI database are as follows: TvGSTO1S: Tv75639; TvGSTO2S: Tv56280; TvGSTO3S: Tv48691; TvGSTO4S: Tv65402; TvGSTO5S: Tv54358 and TvGSTO6S: Tv23671.

References

1. Tien, M. & Kirk, T. K. Lignin-Degrading Enzyme From The Hymenomycete Phanerochaete-Chrysosporium Burds. *Science* **221**, 661–662, https://doi.org/10.1126/science.221.4611.661 (1983).
2. Floudas, D. *et al*. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719, https://doi.org/10.1126/science.1221748 (2012).
3. Eastwood, D. C. *et al*. The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science* **333**, 762–765, https://doi.org/10.1126/science.1205411 (2011).
4. Nagy, L. G. *et al*. Genetic Bases of Fungal White Rot Wood Decay Predicted by Phylogenomic Analysis of Correlated Gene-Phenotype Evolution. *Molecular biology and evolution* **34**, 35–44, https://doi.org/10.1093/molbev/msw238 (2017).
5. Harms, H., Schlosser, D. & Wick, L. Y. Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. *Nature Reviews Microbiology* **9**, 177–192, https://doi.org/10.1038/nrmicro2519 (2011).
6. Morel, M. *et al*. Xenomic networks variability and adaptation traits in wood decaying fungi. *Microbial biotechnology* **6**, 248–263, https://doi.org/10.1111/1751-7915.12015 (2013).
7. Valette, N., Perrot, T., Sormani, R., Gelhaye, E. & Morel-Rouhier, M. Antifungal activities of wood extractives. *Fungal Biology Reviews* **31**, 113–123, https://doi.org/10.1016/j.fbr.2017.01.002 (2017).
8. Syed, K., Shale, K., Pagadala, N. S. & Tuszynski, J. Systematic Identification and Evolutionary Analysis of Catalytically Versatile Cytochrome P450 Monooxygenase Families Enriched in Model Basidiomycete Fungi. *Plos One* **9**, https://doi.org/10.1371/journal.pone.0086683 (2014).
9. Enayati, A. A., Ranson, H. & Hemingway, J. Insect glutathione transferases and insecticide resistance. *Insect molecular biology* **14**, 3–8, https://doi.org/10.1111/j.1365-2583.2004.00529.x (2005).
10. Schuler, M. A. & Berenbaum, M. R. Structure and Function of Cytochrome P450S in Insect Adaptation to Natural and Synthetic Toxins: Insights Gained from Molecular Modeling. *Journal of Chemical Ecology* **39**, 1232–1245, https://doi.org/10.1007/s10886-013-0335-7 (2013).
11. Rane, R. V. *et al*. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Current Opinion in Insect Science* **13**, 70–76, https://doi.org/10.1016/j.cois.2015.12.001 (2016).
12. Morel, M., Ngadin, A. A., Droux, M., Jacquot, J. P. & Gelhaye, E. The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete Phanerochaete chrysosporium. *Cellular and molecular life sciences: CMLS* **66**, 3711–3725, https://doi.org/10.1007/s00018-009-0104-5 (2009).
13. Mathieu, Y. *et al*. Diversification of Fungal Specific Class A Glutathione Transferases in Saprotrophic Fungi. *Plos One* **8**, https://doi.org/10.1371/journal.pone.0080298 (2013).
14. Marinović, M. *et al*. Selective Cleavage of Lignin β-O-4 Aryl Ether Bond by β-Etherase of the White-Rot Fungus Dichomitus squalens. *ACS Sustainable Chemistry & Engineering* **6**, 2878–2882, https://doi.org/10.1021/acssuschemeng.7b03619 (2018).
15. Roret, T. *et al*. Evolutionary divergence of Ure2pA glutathione transferases in wood degrading fungi. *Fungal genetics and biology: FG & B* **83**, 103–112, https://doi.org/10.1016/j.fgb.2015.09.002 (2015).
16. Kontur, W. S. *et al*. Novosphingobium aromaticivorans uses a Nu-class glutathione-S-transferase as a glutathione lyase in breaking the beta-aryl ether bond of lignin. *The Journal of biological chemistry*, https://doi.org/10.1074/jbc.RA117.001268 (2018).
17. Board, P. G. *et al*. Identification, characterization, and crystal structure of the Omega class glutathione transferases. *The Journal of biological chemistry* **275**, 24798–24806, https://doi.org/10.1074/jbc.M001706200 (2000).
18. Meux, E. *et al*. New substrates and activity of Phanerochaete chrysosporium Omega glutathione transferases. *Biochimie* **95**, 336–346, https://doi.org/10.1016/j.biochi.2012.10.003 (2013).
19. Deroy, A. *et al*. The GSTome Reflects the Chemical Environment of White-Rot Fungi. *Plos One* **10**, e0137083, https://doi.org/10.1371/journal.pone.0137083 (2015).
20. Lo, M. C. *et al*. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* **332**, 153–159, https://doi.org/10.1016/j.ab.2004.04.031 (2004).
21. Wu, S. B., Long, C. L. & Kennell, E. J. Structural diversity and bioactivities of natural benzophenones. *Natural product reports* **31**, 1158–1174, https://doi.org/10.1039/c4np00027g (2014).
22. Mosedale, J. R. & Puech, J. L. Wood maturation of distilled beverages. *Trends in Food Science & Technology* **9**, 95–101, https://doi.org/10.1016/s0924-2244(98)00024-7 (1998).
23. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2256–2268, https://doi.org/10.1107/s0907444904026460 (2004).
24. Thom, R. *et al*. Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification. *Biochemistry* **41**, 7008–7020, https://doi.org/10.1021/bi015964x (2002).
25. Yamamoto, K., Suzuki, M., Higashiura, A. & Nakagawa, A. Three-dimensional structure of a Bombyx mori Omega-class glutathione transferase. *Biochem Biophys Res Commun* **438**, 588–593, https://doi.org/10.1016/j.bbrc.2013.08.011 (2013).
26. Oakley, A. Glutathione transferases: a structural perspective. *Drug Metab Rev* **43**, 138–151, https://doi.org/10.3109/03602532.2011.558093 (2011).
27. Cox, P. J., Kechagias, D. & Kelly, O. Conformations of substituted benzophenones. *Acta crystallographica. Section B, Structural science* **64**, 206–216, https://doi.org/10.1107/S0108768108000232 (2008).
28. Badia-Fabregat, M. *et al*. Degradation of UV filters in sewage sludge and 4-MBC in liquid medium by the ligninolytic fungus Trametes versicolor. *Journal of environmental management* **104**, 114–120, https://doi.org/10.1016/j.jenvman.2012.03.039 (2012).
29. Janjusevic, L. *et al*. The lignicolous fungus Trametes versicolor (L.) Lloyd (1920): a promising natural source of antiradical and AChE inhibitory agents. *Journal of enzyme inhibition and medicinal chemistry* **32**, 355–362, https://doi.org/10.1080/14756366.2016.1252759 (2017).
30. Brock, J., G., B. P. & Oakley, A. J. Structural insights into omega-class glutathione transferases: a snapshot of enzyme reduction and identification of a non-catalytic ligandin site. *Plos One* **8**, https://doi.org/10.1371/journal.pone.0060324.g001 (2013).
31. Trabelsi, N. *et al*. Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. *Acta crystallographica. Section D, Biological crystallography* **D64**, 883–891, https://doi.org/10.1107/S0907444908017769 (2008).
32. Ahmad, L., Rylott, E. L., Bruce, N. C., Edwards, R. & Grogan, G. Structural evidence for Arabidopsis glutathione transferase AtGSTF2 functioning as a transporter of small organic ligands. *Febs Open Bio* **7**, 122–132, https://doi.org/10.1002/2211-5463.12168 (2017).
33. Mashiyama, S. T. *et al*. Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS biology* **12**, e1001843, https://doi.org/10.1371/journal.pbio.1001843 (2014).
34. Aguda, A. H. *et al*. Affinity Crystallography: A New Approach to Extracting High-Affinity Enzyme Inhibitors from Natural Extracts. *Journal of natural products* **79**, 1962–1970, https://doi.org/10.1021/acs.jnatprod.6b00215 (2016).
35. McNulty, J. *et al*. Isolation of flavonoids from the heartwood and resin of Prunus avium and some preliminary biological investigations. *Phytochemistry* **70**, 2040–2046, https://doi.org/10.1016/j.phytochem.2009.08.018 (2009).
36. Kebbi-Benkeder, Z., Colin, F., Dumarcay, S. & Gerardin, P. Quantification and characterization of knotwood extractives of 12 European softwood and hardwood species. *Annals of Forest Science* **72**, 277–284, https://doi.org/10.1007/s13595-014-0428-7 (2015).
37. Vinciguerra, V., Luna, M., Bistoni, A. & Zollo, F. Variation in the composition of the heartwood flavonoids of Prunus avium by on-column capillary gas chromatography. *Phytochemical Analysis* **14**, 371–377, https://doi.org/10.1002/pca.730 (2003).

38. Dixon, D. P. & Edwards, R. Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. *The Journal of biological chemistry* **285**, 36322–36329, https://doi.org/10.1074/jbc.M110.164806 (2010).
39. Zhao, J. Flavonoid transport mechanisms: how to go, and with whom. *Trends in Plant Science* **20**, 576–585, https://doi.org/10.1016/j.tplants.2015.06.007 (2015).
40. Gelin, M. *et al.* Combining 'dry' co-crystallization and *in situ* diffraction to facilitate ligand screening by X-ray crystallography. *Acta crystallographica. Section D, Biological crystallography* **71**, 1777–1787, https://doi.org/10.1107/S1399004715010342 (2015).
41. Kabsch, W. X. *Acta crystallographica. Section D, Biological crystallography* **66**, 125–132, https://doi.org/10.1107/S0907444909047337 (2010).
42. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235–242, https://doi.org/10.1107/S0907444910045749 (2011).
43. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213–221, https://doi.org/10.1107/S0907444909052925 (2010).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126–2132, https://doi.org/10.1107/S0907444904019158 (2004).
45. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615–619, https://doi.org/10.1093/nar/gkh398 (2004).
46. Frisch, M. J. *et al.* (Wallingford CT, 2009).
47. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98**, 5648–5652, https://doi.org/10.1063/1.464913 (1993).
48. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **132**, 154104, https://doi.org/10.1063/1.3382344 (2010).
49. Boys, S. F. & Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. *Some procedures with reduced errors. Molecular Physics* **19**, 553–566, https://doi.org/10.1080/00268977000101561 (1970).

Acknowledgements

We thank Solène Telliez, Tiphaine Dhalleine, Jean-Michel Girardet and Sandrine Mathiot for technical assistance. A sincere thank you to Pr. Jean-Pierre Jacquot for constructive criticism of the manuscript. This work was granted access to the HPC resources of CCRT/CINES/IDRIS under the allocation A0030807449 made by GENCI. The authors would like to thank ESRF for beamtime, and the staff of beamlines BM30A and ID30B for assistance with crystal testing and data collection. The authors appreciated the access to the 'Plateforme de mesures de diffraction X' of the Université de Lorraine. This study was funded by the French National Research Agency (ANR-11-LAS-0002-01), the Centre National de la Recherche Scientifique, the University of Lorraine and the Région Grand Est (MS and TP Grants, PEPS-Mirabelle 2016, CPER 2014-2020, Program "Equipement mi-lourd 2016").

Author Contributions

F.F., E.G. and C.D. developed the concept and supervised this study. E.A., T.P., M.S., S.D., M.M.R., F.S., G.M. performed the experiments and interpreted the data. All the authors participated in manuscript writing. E.G., C.D. and P.G. acquired the funding. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-26601-3.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

Supplementary Information

Title: Molecular recognition of wood polyphenols by phase II detoxification enzymes of the white rot *Trametes versicolor*

Mathieu Schwartz[1†], Thomas Perrot[2†], Emmanuel Aubert[1], Stéphane Dumarçay[3], Frédérique Favier[1], Philippe Gérardin[3], Mélanie Morel-Rouhier[2], Guillermo Mulliert[1], Fanny Saiag[2], Claude Didierjean[1*] and Eric Gelhaye[2*]

1 Université de Lorraine, CNRS, CRM2, Nancy, France

2 Université de Lorraine, INRA, IAM, Nancy, France.

3 Université de Lorraine, LERMAB, Nancy, France

*Corresponding authors: claude.didierjean@univ-lorraine.fr; eric.gelhaye@univ-lorraine.fr

†These authors contributed equally to this work.

Table of content

Supplementary Figures

Figure S1. Effects of different compounds from the chemical library on the thermostability of six TvGSTOS isoforms 5

Figure S2. Multiple sequence and structure alignment of GSTs from *T. versicolor*, human GSTO1 and wheat GSTU 6

Figure S3. GSTO3S active site composed of the glutathione binding site (G-site) and the hydrophobic binding site (H-site) 7

Figure S4. Binding of glutathionyl-derivatives in the GSTO3S active site 8

Figure S5. Binding of 3,4- and 2,3,4-hydroxy benzophenones in the GSTO3S hydrophobic binding site (H-site) 9

Figure S6. Slice view of the L-sites of TvGSTO3S (left) and TvGSTO6S (right) located at the dimer interface 10

Figure S7. Binding of dihydrowogonin in the GSTO3S hydrophobic binding site (H-site) 11

Supplementary Tables

Table S1. Diffraction and refinement statistics 12

Table S2. Kinetic parameters of TvGSTO3S and TvGSTO6S toward CDNB (1-chloro-2,4-dinitrobenzene), PEITC (phenethyl isothiocyanate) and GS-PAP (glutathionyl phenylacetophenone) 16

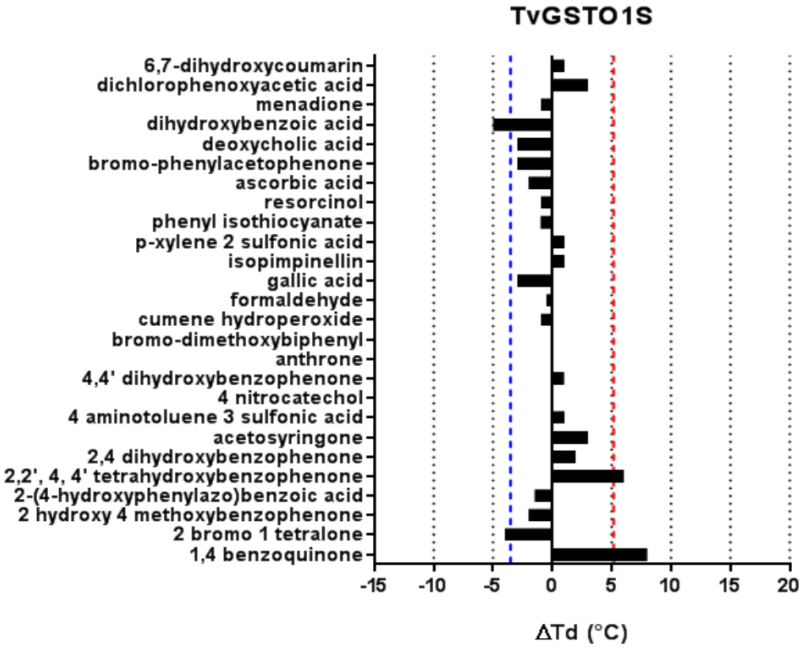
Table S3. Summary of the results obtained with thermal-shift assays and inhibition constants (Ki) obtained for TvGSTOSs with hydroxybenzophenones (HBPs) 17

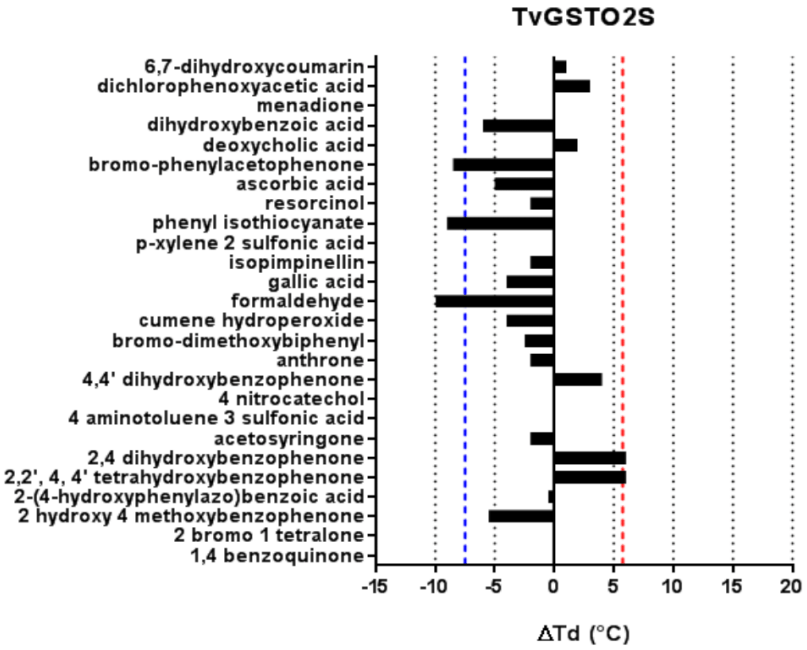
Table S4. Effects of several flavonoids on the thermostability of TvGSTOSs 18

Table S5. Inhibition constants (Ki) of TvGSTO3S and TvGSTO6S by wogonin and naringenin..... 19

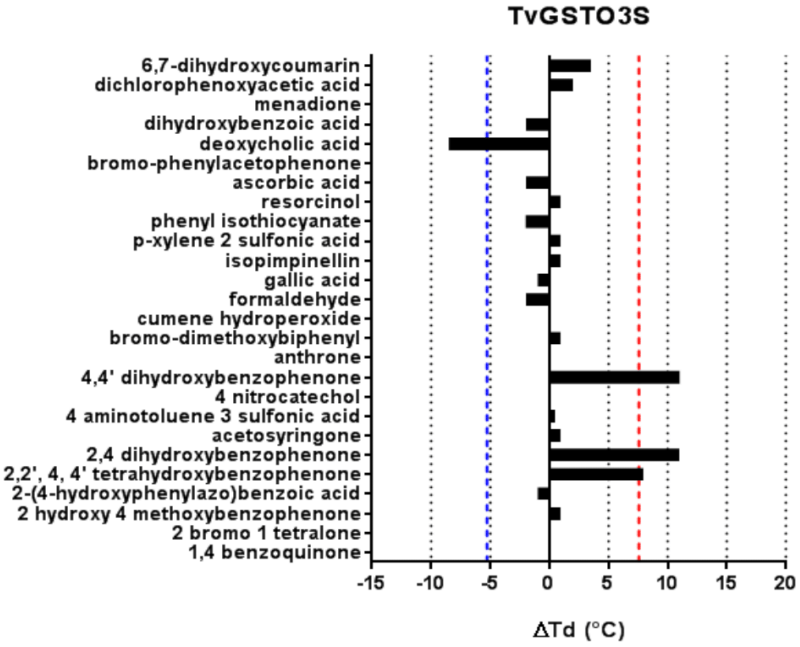
Table S6. Stereo images of 2mFo-DFc electron density maps for each of the eleven structures solved in this study 20

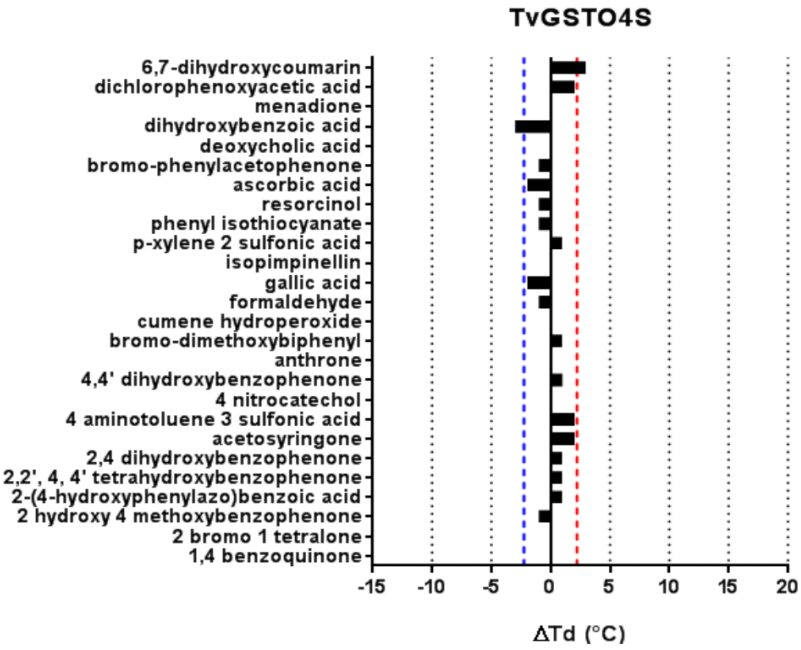
References 22


TvGSTO1S
6,7-dihydroxycoumarin
dichlorophenoxyacetic acid
menadione
dihydroxybenzoic acid
deoxycholic acid
bromo-phenylacetophenone
ascorbic acid
resorcinol
phenyl isothiocyanate
p-xylene 2 sulfonic acid
isopimpinellin
gallic acid
formaldehyde
cumene hydroperoxide
bromo-dimethoxybiphenyl
anthrone
4,4' dihydroxybenzophenone
4 nitrocatechol
4 aminotoluene 3 sulfonic acid
acetosyringone
2,4 dihydroxybenzophenone
2,2', 4, 4' tetrahydroxybenzophenone
2-(4-hydroxyphenylazo)benzoic acid
2 hydroxy 4 methoxybenzophenone
2 bromo 1 tetralone
1,4 benzoquinone
-15 -10 -5 0 5 10 15 20
ΔTd (°C)


TvGSTO2S
6,7-dihydroxycoumarin
dichlorophenoxyacetic acid
menadione
dihydroxybenzoic acid
deoxycholic acid
bromo-phenylacetophenone
ascorbic acid
resorcinol
phenyl isothiocyanate
p-xylene 2 sulfonic acid
isopimpinellin
gallic acid
formaldehyde
cumene hydroperoxide
bromo-dimethoxybiphenyl
anthrone
4,4' dihydroxybenzophenone
4 nitrocatechol
4 aminotoluene 3 sulfonic acid
acetosyringone
2,4 dihydroxybenzophenone
2,2', 4, 4' tetrahydroxybenzophenone
2-(4-hydroxyphenylazo)benzoic acid
2 hydroxy 4 methoxybenzophenone
2 bromo 1 tetralone
1,4 benzoquinone
-15 -10 -5 0 5 10 15 20
ΔTd (°C)

3


TvGSTO3S
6,7-dihydroxycoumarin
dichlorophenoxyacetic acid
menadione
dihydroxybenzoic acid
deoxycholic acid
bromo-phenylacetophenone
ascorbic acid
resorcinol
phenyl isothiocyanate
p-xylene 2 sulfonic acid
isopimpinellin
gallic acid
formaldehyde
cumene hydroperoxide
bromo-dimethoxybiphenyl
anthrone
4,4' dihydroxybenzophenone
4 nitrocatechol
4 aminotoluene 3 sulfonic acid
acetosyringone
2,4 dihydroxybenzophenone
2,2', 4, 4' tetrahydroxybenzophenone
2-(4-hydroxyphenylazo)benzoic acid
2 hydroxy 4 methoxybenzophenone
2 bromo 1 tetralone
1,4 benzoquinone
-15 -10 -5 0 5 10 15 20
ΔTd (°C)


TvGSTO4S
6,7-dihydroxycoumarin
dichlorophenoxyacetic acid
menadione
dihydroxybenzoic acid
deoxycholic acid
bromo-phenylacetophenone
ascorbic acid
resorcinol
phenyl isothiocyanate
p-xylene 2 sulfonic acid
isopimpinellin
gallic acid
formaldehyde
cumene hydroperoxide
bromo-dimethoxybiphenyl
anthrone
4,4' dihydroxybenzophenone
4 nitrocatechol
4 aminotoluene 3 sulfonic acid
acetosyringone
2,4 dihydroxybenzophenone
2,2', 4, 4' tetrahydroxybenzophenone
2-(4-hydroxyphenylazo)benzoic acid
2 hydroxy 4 methoxybenzophenone
2 bromo 1 tetralone
1,4 benzoquinone
-15 -10 -5 0 5 10 15 20
ΔTd (°C)

4

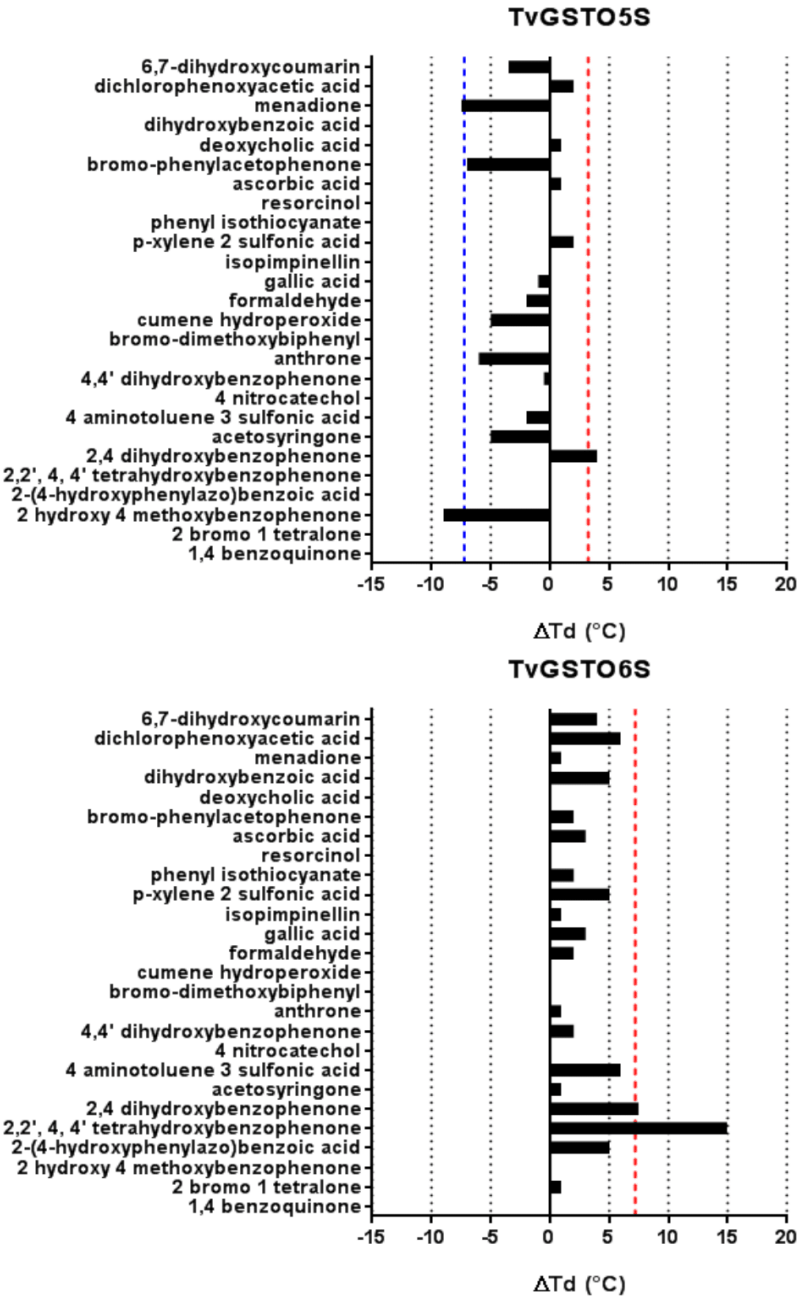


Figure S1. Effects of different compounds from the chemical library on the thermostability of six TvGSTOS isoforms

Red and blue lines correspond to the addition of the average with the positive and negative standard deviations respectively. Thermostability has been analysed by using 40 µM of protein; the final concentration of molecules tested was of 0.8 mg/mL in each well.

5

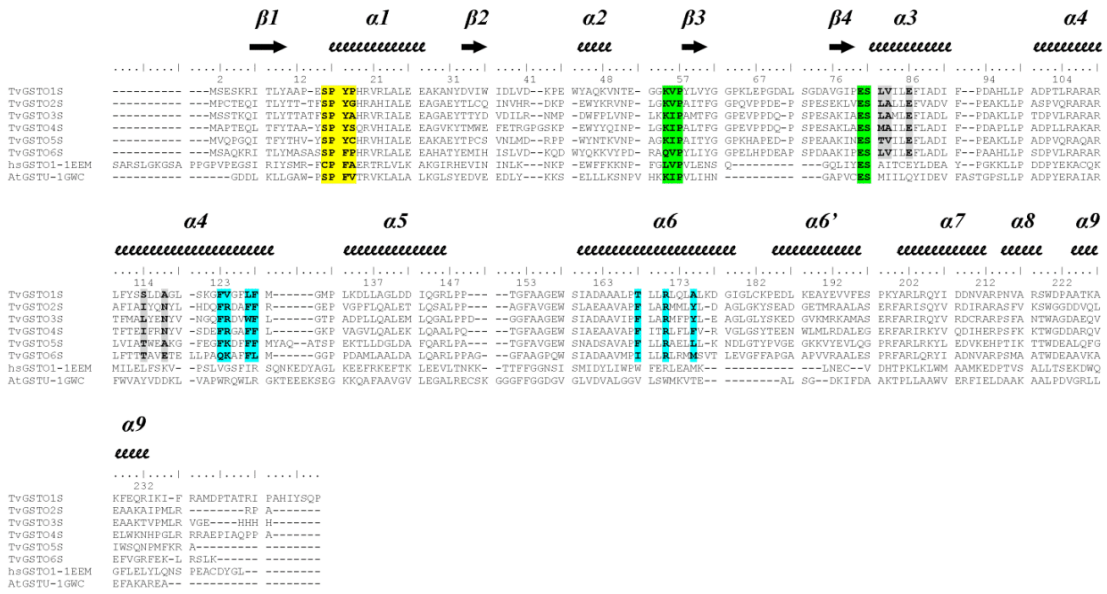


Figure S2. Multiple sequence and structure alignment of GSTs from *T. versicolor*, human GSTO1 and wheat GSTU

Sequences of TvGSTOs were retrieved from the Joint Genome Institute (URL https://jgi.doe.gov/) database (accession numbers: TvGSTO1S 75639, TvGSTO2S 56280, TvGSTO3S 48691, TvGSTO4S 65402, TvGSTO5S 54358 and TvGSTO6S 23671). Structures of TvGSTO3S and TvGSTO6S solved in this study and structures of human GSTO1 (PDB code 1eem) and wheat AtGSTU (PDB code 1gwc) were used as input structures in Promals3D [1]. Catalytic motifs, conserved glutathione binding site residues and conserved hydrophobic binding site residues of TvGSTOS isoforms are highlighted in yellow, green and blue, respectively. Residues of the non-catalytic ligandin site of TvGSTO6S and corresponding positions in TvGSTOS isoforms are highlighted in grey. Secondary structures identified in TvGSTO3S structure are reported above the alignment. Residues are numbered according to the TvGSTO3S sequence.

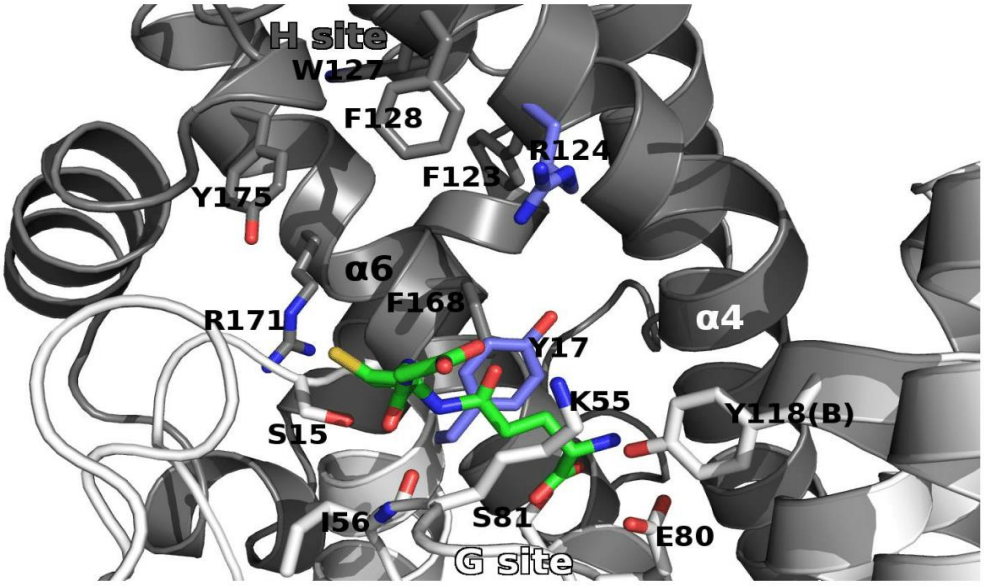


Figure S3. GSTO3S active site composed of the glutathione binding site (G-site) and the hydrophobic binding site (H-site)

Section of the TvGSTO3S active site is shown. N-terminal domain bearing the G-site residues is displayed in white and C-terminal domain bearing the H-site residues is shown in black. Glutathione is represented as green sticks. Polar residues at the entrance of the H-site are colored in violet.

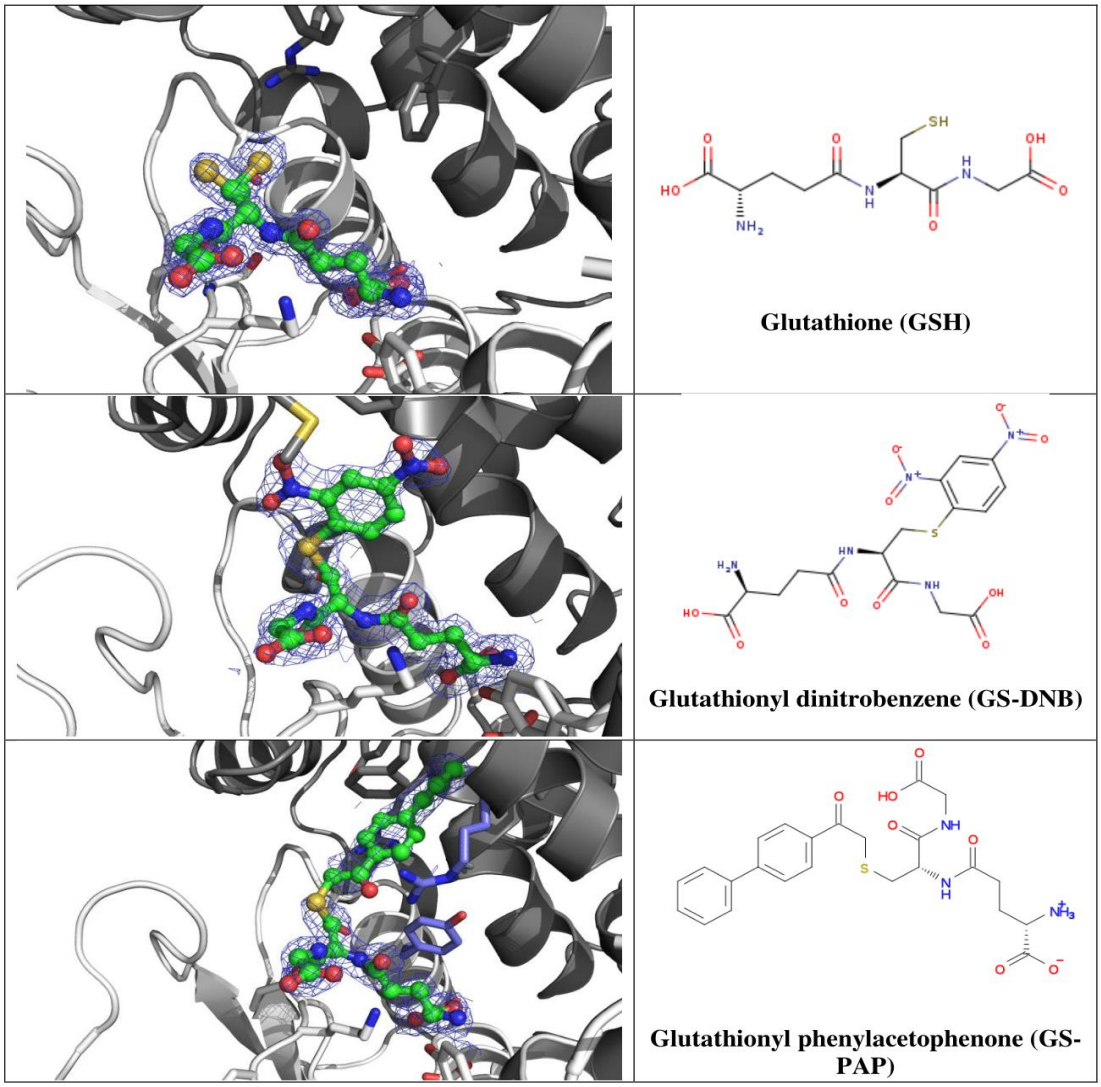


Figure S4. Binding of glutathionyl-derivatives in TvGSTO3S active site

Sections of TvGSTO3S complex structures with glutathionyl-derivatives are shown in the left column. The corresponding chemical structures of the glutathionyl-derivatives are shown in the right column. Surrounding side chains are represented by sticks. Glutathionyl-derivatives are shown as green sticks and spheres. 2mFo-DFc composite omit maps shown at 1.0 σ around glutathionyl-derivatives were calculated by PHENIX.

8

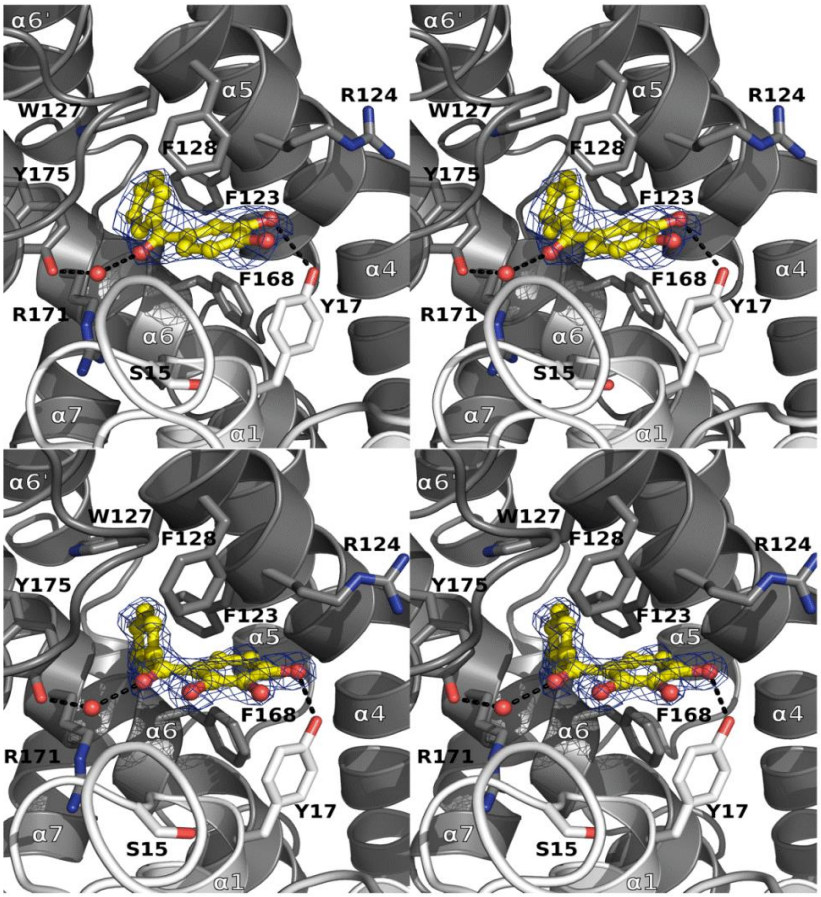


Figure S5. Binding of 3,4- and 2,3,4-hydroxy benzophenones in the GSTO3S hydrophobic binding site (H-site)

Stereoviews of sections of the TvGSTO3S complex structures with 3,4-HBP (top view) and 2,3,4-HBP (bottom view) are shown. Polar intermolecular contacts are materialized as dashed lines. Surrounding side chains are represented by sticks. HBPs are shown as yellow sticks and spheres. 2mFo-DFc composite omit maps shown at 1.0 σ around HBPs were calculated by PHENIX.

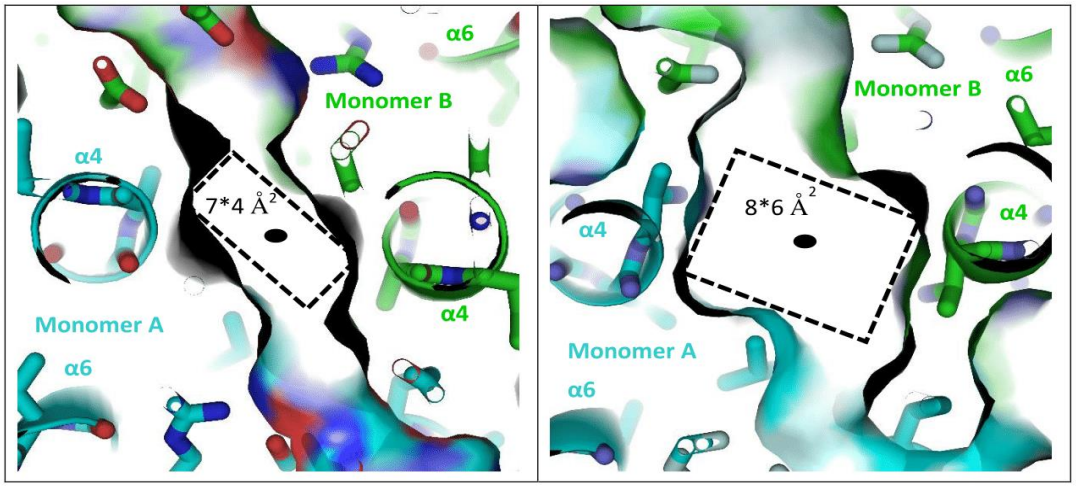


Figure S6. Slice view of the L-sites of TvGSTO3S (left) and TvGSTO6S (right) located at the dimer interface

The monomers A (cyan) and B (green) are related by a two-fold axis depicted as a black ellipsoid. Both L-sites exhibit a rectangular section with dimensions 7 Å by 4 Å for TvGSTO3S and 8 Å by 6 Å for TvGSTO6S. The protein surface is represented by a Connolly surface generated with PyMol.

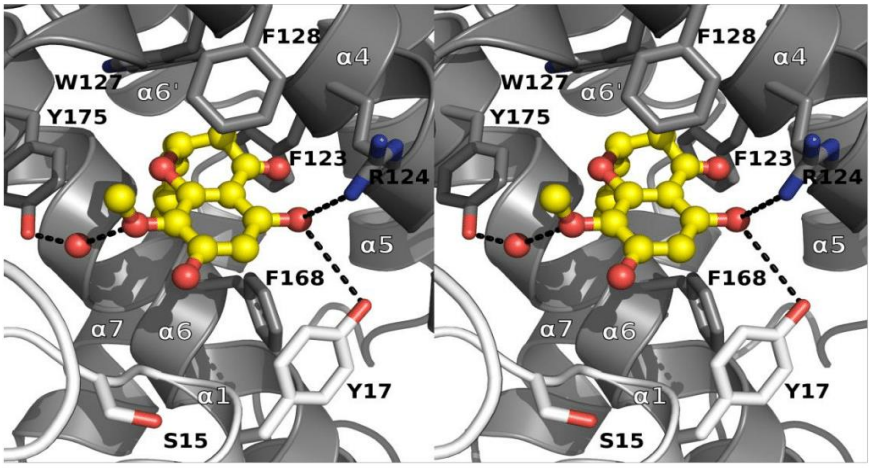


Figure S7. Binding of dihydrowogonin in the GSTO3S hydrophobic binding site (H-site)
Stereo view of the TvGSTO3S complex structure with dihydrowogonin is shown. Polar intermolecular contacts are materialized as dashed lines. Surrounding side chains are represented in sticks. Dihydrowogonin is shown as yellow sticks and spheres.

Table S1. Diffraction and refinement statistics

$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

$R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

$CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data [2].

$R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl}|F_{obs}|$. 5 % of reflections were selected for R_{free} calculation. R.m.s.z. : root mean square Z-score [3]. The molprobity clashscore is the number of serious clashes per 1000 atoms [4]. The molprobity score is a log-weighted combination of the clashscore, percentage Ramachandran not favoured and percentage bad side-chain rotamers [4]. Values in parentheses are for highest resolution shell.

		TvGSTO3S apo	**TvGSTO3S - GSH**	**TvGSTO3S - GS-DNB**
Diffraction data				
	Diffraction source	FIP-BM30A, ESRF	FIP-BM30A, ESRF	FIP-BM30A, ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97975	0.97973	0.97974
	Unit-cell parameters			
	a, b, c (Å)	50.5 104.5 107.7	50.5 104.2 107.9	50.4 103.7 107.3
	α, β, γ (°)	90	90	90
	Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
	Resolution range (Å)	47.88 - 1.35 (1.39 -1.35)	47.92 - 1.55 (1.59 -1.55)	41.78 - 1.75 (1.80 - 1.75)
	Total No. of reflections	1,024,564 (58,855)	586,955 (30,824)	186,029 (13,780)
	No. of unique reflections	125,602 (9203)	82,933 (5,773)	56,918 (4,185)
	Average redundancy	8.2 (6.4)	7.1 (5.3)	3.3 (3.3)
	Mean I/σ(I)	16.5 (1.7)	22.2 (1.9)	13.2 (1.5)
	Completeness (%)	100 (99.9)	99.5 (94.4)	98.8 (99.5)
	R_{merge}	0.071 (1.13)	0.062 (0.869)	0.081 (0.821)
	R_{meas}	0.076 (1.23)	0.067 (0.964)	0.097 (0.977)
	$CC_{1/2}$	0.99 (0.58)	1.00 (0.63)	0.97 (0.54)
Refinement				
	Resolution range (Å)	47.88 - 1.35 (1.36 - 1.35)	47.92 - 1.55 (1.57 - 1.55)	41.78 - 1.75 (1.78 - 1.75)
	R_{work} / R_{free}	0.13 / 0.17 (0.25 / 0.31)	0.16 / 0.18 (0.27 / 0.28)	0.16 / 0.21 (0.30 / 0.35)
	No. of protein atoms	7,566	7,566	7,597
	No. of waters	482	499	461
	Average B factor (Å²)	20.3	20.3	22.2
Model quality				
	RMSZ bond lenghts	0.66	0.67	0.62
	RMSZ bond angles	0.78	0.75	0.74
	Ramachandran favoured (%)	98.3	98.5	97.9
	Ramachandran outliers (%)	0.0	0.2	0.4
	Molprobity rotamer outliers (%)	0.0	0.0	0.3
	Molprobity clashscore	0.39	1.03	1.93
	Molprobity score	0.64	0.80	0.98
PDB entry		**6F43**	**6F4B**	**6F4F**

		TvGSTO3S - GS-PAP	**TvGSTO3S - 2,4 HBP**	**TvGSTO3S - eluate 46-47**
Diffraction data				
	Diffraction source	FIP-BM30A, ESRF	FIP-BM30A, ESRF	FIP-BM30A, ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97973	0.98142	0.98140
	Unit-cell parameters			
	a, b, c (Å)	50.5 104.4 106.2	50.5 104.6 107.7	50.5 104.5 108.0
	α, β, γ (°)	90	90	90
	Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
	Resolution range (Å)	46.84 - 1.92 (1.97 - 1.92)	47.89 - 1.75 (1.79 - 1.75)	47.97 - 1.70 (1.74 - 1.70)
	Total No. of reflections	303,701 (15,533)	242,605 (9,897)	461,368 (34,288)
	No. of unique reflections	43,310 (2,912)	57,643 (4,022)	63,676 (4,684)
	Average redundancy	7.0 (5.3)	4.2 (2.5)	7.2 (7.3)
	Mean I/σ(I)	17.1 (1.8)	10.8 (1.4)	19.8 (2.4)
	Completeness (%)	99.2 (91.4)	98.6 (94.4)	100 (99.9)
	R_{merge}	0.093 (1.03)	0.090 (0.614)	0.074 (0.883)
	R_{meas}	0.100 (1.14)	0.103 (0.758)	0.080 (0.950)
	$CC_{1/2}$	1.00 (0.54)	1.00 (0.57)	1.00 (0.75)
Refinement				
	Resolution range (Å)	46.84 - 1.92 (1.97 - 1.92)	47.89 - 1.75 (1.78 - 1.75)	47.97 - 1.70 (1.73 - 1.70)
	R_{work} / R_{free}	0.17 / 0.22 (0.28 / 0.34)	0.19 / 0.22 (0.29 / 0.31)	0.16 / 0.19 (0.23 / 0.25)
	No. of protein atoms	7,522	7,545	7,565
	No. of waters	327	463	311
	Average B factor (Å²)	28.2	20.5	21.5
Model quality				
	RMSZ bond lenghts	0.59	0.74	0.63
	RMSZ bond angles	0.68	0.76	0.71
	Ramachandran favoured (%)	98.5	98.5	98.8
	Ramachandran outliers (%)	0.0	0.0	0.2
	Molprobity rotamer outliers (%)	1.6	0.0	0.0
	Molprobity clashscore	1.56	2.20	1.93
	Molprobity score	1.05	1.00	0.96
PDB entry		**6F51**	**6F66**	**6F6A**

13

		TvGSTO3S - 3,4 HBP	**TvGSTO3S - 2,4,4' HBP**	**TvGSTO3S - 2,3,4 HBP**
Diffraction data				
	Diffraction source	FIP-BM30A, ESRF	FIP-BM30A, ESRF	ID30B, ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD	PILATUS 6M-F
	Wavelength (Å)	0.97992	0.98142	0.97625
	Unit-cell parameters			
	a, b, c (Å)	50.6 105.7 105.4	50.7 104.8 108.3	50.4 104.2 107.1
	α, β, γ (°)	90	90	90
	Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
	Resolution range (Å)	47.23 - 2.40 (2.46 - 2.40)	47.20 - 1.70 (1.74 - 1.70)	47.62 - 1.80 (1.85 - 1.80)
	Total No. of reflections	124,297 (7,915)	306,585 (20,269)	169,481 (10,890)
	No. of unique reflections	22,688 (1,536)	64,275 (4,473)	49,835 (3,534)
	Average redundancy	5.5 (5.2)	4.8 (4.5)	3.4 (3.1)
	Mean I/σ(I)	12.4 (2.6)	19.3 (3.45)	8.0 (2.1)
	Completeness (%)	99.5 (94.3)	99.4 (95.0)	94.3 (91.9)
	R_{merge}	0.146 (0.740)	0.050 (0.413)	0.089 (0.387)
	R_{meas}	0.161 (0.822)	0.056 (0.467)	0.103 (0.458)
	$CC_{1/2}$	1.00 (0.76)	1.00 (0.89)	0.99 (0.62)
Refinement				
	Resolution range (Å)	47.23 - 2.40 (2.51 - 2.40)	47.20 - 1.70 (1.72 - 1.70)	47.62 - 1.80 (1.84 - 1.80)
	R_{work} / R_{free}	0.18 / 0.21 (0.22 / 0.28)	0.16 / 0.19 (0.20 / 0.26)	0.22 / 0.27 (0.28 / 0.32)
	No. of protein atoms	7,530	7,582	7,567
	No. of waters	161	350	182
	Average B factor (Å²)	38.8	22.5	24.9
Model quality				
	RMSZ bond lenghts	0.71	0.59	0.69
	RMSZ bond angles	0.74	0.71	0.78
	Ramachandran favoured (%)	96.4	98.5	97.3
	Ramachandran outliers (%)	0.4	0.2	0.4
	Molprobity rotamer outliers (%)	1.8	0.3	1.0
	Molprobity clashscore	2.76	0.64	3.14
	Molprobity score	1.49	0.71	1.25
PDB entry		**6F67**	**6F68**	**6F69**

		TvGSTO6S apo	**TvGSTO6S - naringenin**
Diffraction data			
	Diffraction source	ID30B, ESRF	FIP-BM30A, ESRF
	Detector	PILATUS 6M-F	ADSC Q315r CCD
	Wavelength (Å)	0.97625	0.97994
	Unit-cell parameters		
	a, b, c (Å)	53.1 78.5 67.9	73.8 78.9 93.6
	α, β, γ (°)	90 106.6 90	90 101.5 90
	Space group	$P2_1$	$P2_1$
	Resolution range (Å)	42.69 - 1.48 (1.52 - 1.48)	49.35 - 2.30 (2.36 - 2.30)
	Total No. of reflections	226,748 (14,990)	171,767 (8,812)
	No. of unique reflections	85,232 (6,051)	46,447 (3,139)
	Average redundancy	2.7 (2.5)	3.7 (2.8)
	Mean I/σ(I)	14.5 (2.6)	8.9 (1.8)
	Completeness (%)	96.0 (92.7)	99.0 (91.3)
	R_{merge}	0.039 (0.363)	0.135 (0.599)
	R_{meas}	0.047 (0.457)	0.158 (0.748)
	$CC_{1/2}$	1.00 (0.68)	0.99 (0.68)
Refinement			
	Resolution range (Å)	42.69 - 1.48 (1.50 - 1.48)	49.35 - 2.30 (2.35 - 2.30)
	R_{work} / R_{free}	0.15 / 0.18 (0.27 / 0.30)	0.17 / 0.23 (0.22 / 0.30)
	No. of protein atoms	7,609	7,484
	No. of waters	572	472
	Average B factor (Å²)	22.7	33.0
Model quality			
	RMSZ bond lenghts	0.62	0.64
	RMSZ bond angles	0.76	0.72
	Ramachandran favoured (%)	99.0	97.7
	Ramachandran outliers (%)	0.0	0.2
	Molprobity rotamer outliers (%)	0.5	1.7
	Molprobity clashscore	1.43	2.71
	Molprobity score	0.88	1.30
PDB entry		**6F70**	**6F71**

15

Table S2. Kinetic parameters of TvGSTO3S and TvGSTO6S toward CDNB (1-chloro-2,4-dinitrobenzene), PEITC (phenethyl isothiocyanate) and GS-PAP (glutathionyl phenylacetophenone)

The kinetic parameters were calculated using the GraphPad software with the nonlinear regression based on the Michaelis-Menten. "ND" means that no activity has been detected.

TvGSTO3S	**CDNB**	**PEITC**	**GS-PAP**
K_M (µM)	897.80 ± 85.20	12.20 ± 1.44	ND
Kcat (s^{-1})	26.64 ± 0.95	3.09 ± 0.05	ND
kcat/K_M ($M^{-1}.s^{-1}$)	(2.97 ± 0.11)E+04	(2.53 ± 0.04)E+05	ND
TvGSTO6S	**CDNB**	**PEITC**	**GS-PAP**
K_M (µM)	910.40 ± 86.77	24.51 ± 2.99	ND
Kcat (s^{-1})	4.94 ± 0.17	11.38 ± 0.36	ND
kcat/K_M ($M^{-1}.s^{-1}$)	(5.43 ± 0.18)E+03	(4.65 ± 0.15)E+05	ND

Table S3. Summary of the results obtained with thermal-shift assays and inhibition constants (Ki) obtained for TvGSTOSs with hydroxybenzophenones (HBPs)

"NS" means that temperature denaturation has not been significantly modified. "ΔTd" corresponds modification of the denaturation temperature which is significantly different when the protein is incubated with only DMSO. "NI" means that no inhibition was detected. Ki have been determined with GraphPad Prism software using the mixed model inhibition. For asymmetric molecules, the less substituted phenyl ring (*i.e.* the left one on the drawing) is named **B** and the more substituted one (right) is named **A**.

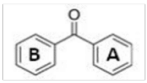
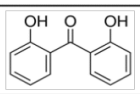
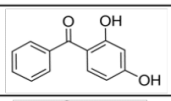
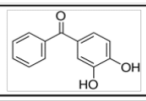
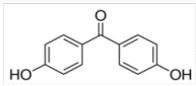
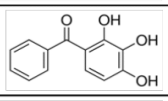
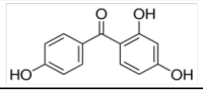
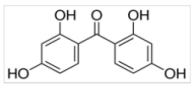
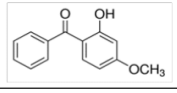
Molecules	Structures	Thermal shift (°C)	Inhibition constant (µM)
Benzophenone		NS	NI
2.2′- Dihydroxybenzophenone		NS	NI
2.4- Dihydroxybenzophenone		ΔTd = 5.69°C	2.00 ± 0.67
3.4-Dihydroxybenzophenone		ΔTd = 4.36°C	0.81 ± 0.29
4.4′-Dihydroxybenzophenone		ΔTd = 1.67°C	NI
2.3.4- Trihydroxybenzophenone		ΔTd = 4.96°C	0.58 ± 0.18
2.4.4′- Trihydroxybenzophenone		ΔTd = 2.87°C	14.51 ± 7.05
2.2′.4.4′- Tetrahydroxybenzophenone		NS	NI
2-Hydroxy-4- methoxybenzophenone		NS	NI

Table S4. Effects of several flavonoids on the thermostability of TvGSTOSs

Thermostability was analysed by using 10 µM of protein. The final concentration of the tested molecules was 100 µM in each well. A ΔTd value is only given if the denaturation temperature is significantly modified in the presence of compounds, with respect to incubation with DMSO only. "NS" means that the denaturation temperature has not changed significantly.

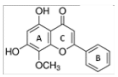
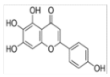
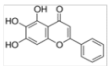
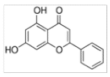
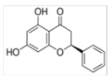
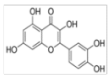
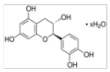
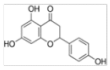
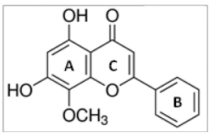
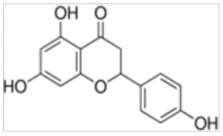
Molecules	Structures	TvGSTO1S	TvGSTO2S	TvGSTO3S	TvGSTO4S	TvGSTO5S	TvGSTO6S
Wogonin		NS	4.44°C	4.40°C	NS	NS	2.78°C
Scutellarein		3.85°C	NS	NS	NS	-11.49°C	2.51°C
Baicalein		3.39°C	NS	1.81°C	1.31°C	-3.91°C	5.03°C
Chrysin		2.10°C	3.45°C	2.41°C	NS	NS	5.85°C
Pinocembrin		2.32°C	1.95°C	1.47°C	NS	NS	0.94°C
Quercetin		2.72°C	6.31°C	2.63°C	2.85°C	5.57°C	5.72°C
Catechin		NS	NS	NS	NS	NS	NS
Naringenin		NS	NS	NS	NS	NS	1.92°C

Table S5. Inhibition constants (Ki) of TvGSTO3S and TvGSTO6S by wogonin and naringenin

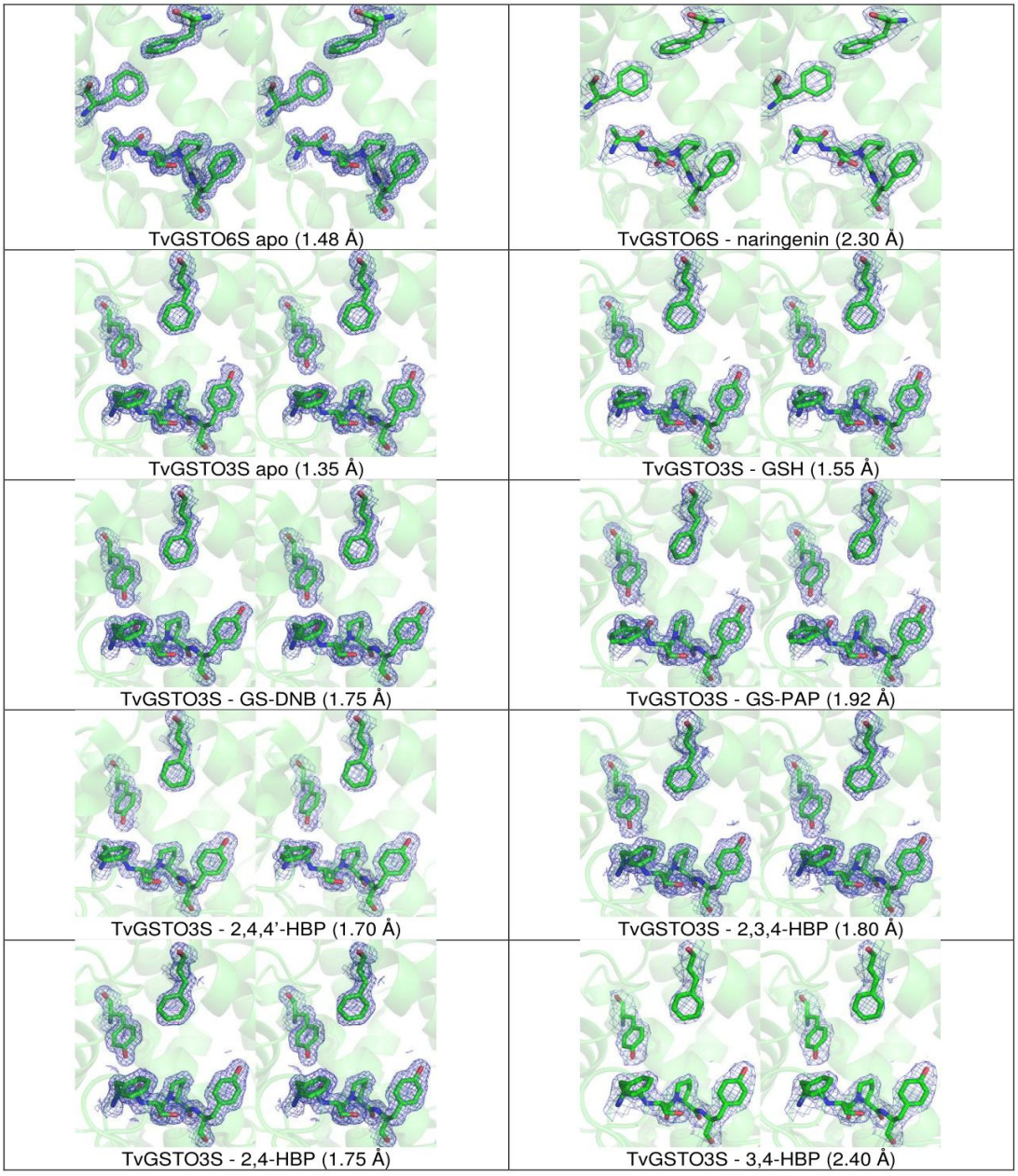
Ki have been determined with GraphPad Prism software using the mixed model inhibition. "NI" means that no inhibition was detected.

Molecules	Structures	TvGSTO3S Inhibition constant (µM)	TvGSTO6S Inhibition constant (µM)
Wogonin		2.89 ± 1.24	NI
Naringenin		NI	NI

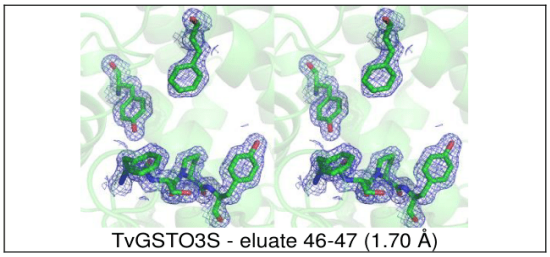
19

Table S6. Stereo images of 2mFo-DFc electron density maps for each of the eleven structures solved in this study

For each panel, stereo images of a portion of 2mFo-DFc electron density maps are shown to assess quality of the structural data (level at 1.0 σ). Residues from TvGSTO3S structures (F14, S15, P16, Y17, F128 and Y175) and TvGSTO6S structures (A14, S15, P16, F17, F219 and F232) are shown as green sticks.



20



21

References

1 Pei, J. M. & Grishin, N. V. in *Multiple Sequence Alignment Methods* Vol. 1079 *Methods in Molecular Biology* (ed D. J. Russell) 263-271 (2014).

2 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).

3 Tickle, I. J. Experimental determination of optimal root-mean-square deviations of macromolecular bond lengths and angles from their restrained ideal values. *Acta crystallographica. Section D, Biological crystallography* **63**, 1274-1281; author reply 1282-1273, doi:10.1107/S0907444907050196 (2007).

4 Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615-619, doi:10.1093/nar/gkh398 (2004).

I.4.b. Publication n°2 : Identification de molécules issues d'extraits de bois exotiques comme ligands de l'isoforme TvGSTO2S

i) Résumé

La durabilité naturelle du bois est liée à sa composition chimique et, en particulier, à la présence de métabolites appelés extractibles qui possèdent souvent une réactivité chimique. Pour faire face à ces composés, les dégradeurs de bois ont mis au point des systèmes de détoxication qui impliquent généralement des familles d'enzymes. Parmi ces enzymes, les glutathion transférases (GST) sont impliquées dans la diminution de la réactivité des composés toxiques. Dans cette étude, nous avons testé l'hypothèse selon laquelle les systèmes de détoxication des champignons dégradeurs du bois pourraient être des indicateurs de la réactivité chimique d'extraits de bois. Cette approche a été évaluée à partir de trente-deux extraits de bois provenant d'espèces originaires de Guyane française, en testant leur capacité antimicrobienne, leurs propriétés antioxydantes et leur réactivité vis-à-vis de six GST du champignon de la pourriture blanche *Trametes versicolor*. D'après les données obtenues, il existe une corrélation significative entre d'une part les propriétés antimicrobiennes et les propriétés antioxydantes des extraits de bois testés et d'autre part leur interaction avec les GST. De plus, l'analyse chimique de l'un des extraits les plus réactifs (extrait acétonique de bagasse *Bagassa guianensis*) a démontré la présence d'un composé majeur dans cet extrait, l'oxyresvératrol. Nous avons pu co-cristalliser une GST (TvGSTO2S) avec ce composé intéressant d'un point de vue commercial. L'ensemble des données présentées confirment l'hypothèse que les enzymes de détoxification pourraient être utilisées pour identifier la présence de molécules d'intérêt industriel dans les extraits de bois.

ii) Publication

ACS Sustainable Chemistry & Engineering
Research Article
Cite This: ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX
pubs.acs.org/journal/ascecg

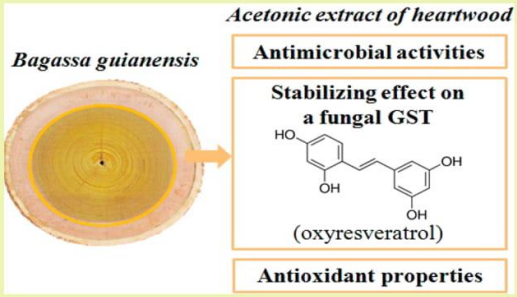
Fungal Glutathione Transferases as Tools to Explore the Chemical Diversity of Amazonian Wood Extractives

Thomas Perrot,[†,∥] Mathieu Schwartz,[‡,∥] Fanny Saiag,[†] Guillaume Salzet,[†] Stéphane Dumarçay,[§] Frédérique Favier,[‡] Philippe Gérardin,[§] Jean-Michel Girardet,[†] Rodnay Sormani,[†] Mélanie Morel-Rouhier,[†] Nadine Amusant,[⊥] Claude Didierjean,[‡] and Eric Gelhaye*[,†]

[†]Université de Lorraine, INRA, IAM, F-54000 Nancy, France
[‡]Université de Lorraine, CNRS, CRM2, F-54000 Nancy, France
[§]Université de Lorraine, INRA, LERMAB, F-54000 Nancy, France
[⊥]CIRAD, UMR Ecofog (AgroParisTech, CIRAD, CNRS, INRA, UA), BF701−97310 Kourou, Cedex, France

Supporting Information

ABSTRACT: The natural durability of wood is linked to its chemical composition and in particular the presence of metabolites called extractives that often possess chemical reactivity. For dealing with these compounds, wood degraders have developed detoxification systems usually involving enzyme families. Among these enzymes, glutathione transferases (GSTs) are involved in the decrease of the reactivity of toxic compounds. In this study, the hypothesis that the detoxification systems of wood decaying fungi could be indicators of the chemical reactivity of wood extracts has been tested. This approach has been evaluated using 32 wood extracts coming from French Guiana species, testing their antimicrobial ability, antioxidative properties, and reactivity against six GSTs from the white rot *Trametes versicolor*. From the obtained data, a significant correlation between the antimicrobial and antioxidative properties of the tested wood extracts and GST interactions was established. In addition, the chemical analysis performed on one of the most reactive extracts (an acetonic extract of *Bagassa guianensis*) has demonstrated oxyresveratrol as a major constituent. We were able to cocrystallize one GST with this commercially interesting compound. Taken together, the presented data support the hypothesis that detoxifying enzymes could be used to identify the presence of molecules of industrial interest in wood extracts.



KEYWORDS: *Trametes versicolor, Amazonian wood species, Bagassa guianensis, Glutathione transferase, Stilbene*

INTRODUCTION

Wood is a major renewable resource with many fields of application as energy, building, or polymers production. Beyond the presence of structural polymers (cellulose, hemicellulose, and lignin), wood also contains secondary metabolites called extractives. The chemical composition of extractives is various and includes terpenes, fatty acids, simple phenols, flavonoids, tannins, and stilbenes. Although the functions of these wood molecules often remain unclear, they usually possess properties of interest such as antimicrobial and antioxidative activities.[1,2] Because of their properties, these compounds could be used for several industrial purposes in particular for wood preservation, crop protection, medicinal treatments, or cosmetics.[3] However, they could also be a problem for lignocellulosic biomass valorization, limiting its enzymatic digestibility.[4]

The diversity and the potential toxicity of these extractives suggest that wood degraders and in particular wood decaying fungi are adapted to the presence of these potential toxic molecules.[5] Beyond their extracellular systems, which allow them to degrade lignocellulosic substrates,[6,7] wood-decaying fungi indeed possess extended detoxifying enzyme families, such as cytochrome P450 monooxygenases (P450s) and glutathione transferases (GSTs).[5] Such extensions are also found in herbivorous insects with these multigenic families playing key functions in the detoxification of plant defense chemicals and also in the evolution of metabolic resistance to chemical insecticides.[8] Concerning GSTs, their activity or even their expression are widely used to evaluate physiological and environmental stress of diverse organisms from molluscs to humans.[9−11] In wood-decaying fungi, the extension of the GST family concerns mainly specific phylogenetically based classes named Ure2p, GST Omega (GSTO), and GSTFuA.[12−14] GSTFuA is able to cleave lignin β-O-4 aryl ether bond in the white-rot fungus *Dichomitus squalens*.[15] In addition, GSTOs from *Trametes versicolor* interact with wood extracts from

Received: June 5, 2018
Revised: July 31, 2018
Published: August 25, 2018

ACS Publications © XXXX American Chemical Society

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

Downloaded via UNIV DE LORRAINE on September 17, 2018 at 08:53:37 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

temperate forest[16] and more particularly with polyphenols such as hydroxylbenzophenones and flavonoids.[17] In this context, we postulated that GSTs could be used as tools to identify wood extracts that possess interesting biological properties. To test this hypothesis, we investigated the biochemical interactions between six GSTOs from *Trametes versicolor*[16] and an environmental collection of wood extracts. *Trametes versicolor* is common and widespread in boreal and temperate northern hemisphere and also occurs in tropical areas of both hemispheres.[18] The six studied GST isoforms, which contain a serinyl residue in the active site, belong to the fungal type-III Omega class. The tridimensional structures of two of them, TvGSTO3S and TvGSTO6S, have been recently solved.[17] The extracts came from French Guiana woody species known to be naturally durable against fungi.[1,2,19] Using high-throughput biochemical methods, we show here that the chemical interactions of one isoform with the tested wood extracts are highly correlated with the chemical and biological properties (antioxidative and antimicrobial activities) of these compounds. Additional biochemical and structural experiments demonstrated the interactions between TvGSTO2S and oxyresveratrol, a stilbene largely used by the cosmetics industry, supporting the feasibility of our approach.

EXPERIMENTAL SECTION

Chemicals. The solvents used for the extraction step and for the chromatographic fractionation by HPLC are provided from Honeywell and Carlo Erba, respectively.

Wood Extracts. Heartwoods from *Peltogyne venosa*, *Dicorynia guianensis*, *Bagassa guianensis*, *Hymenaea courbaril*, *Tabebuia serratifolia*, *Sextonia rubra*, *Andira coriacea*, and *Eperua falcata* were from commercial origin (Degrad Saramaca's sawmill, Kourou, French Guiana). All species are well known and used in the building industry because of their durability against wood-rotting fungi (i.e., with natural durability against fungi rated 1−3 on a scale of 5 with class 1 being the most durable).

Each conditioned sample (12% relative humidity) was ground to fine sawdust with particle size between 0.2 and 0.4 mm before extraction. The obtained sawdust was Soxhlet-extracted successively during 24 h using the following solvents: dichloromethane, acetone, toluene/ethanol (2/1, v/v), and water. After each extraction, organic solvents were evaporated under vacuum using a rotary evaporator. Dried extractives were stored in a freezer (−18 °C) before analyses.

Chromatographic Analysis of the *Bagassa guianensis* Heartwood Acetonic Extract. *Bagassa guianensis* heartwood acetonic extract was fractioned by reverse chromatography (Shimadzu Prominence HPLC system) as previously described.[17] Twelve injections of 80 μL of the extract at 92 mg mL^{-1} were fractioned onto a Kinetex Biphenyl column (250 × 4.6 mm internal diameter, 5 μm particle size, 10 nm porosity; Phenomenex) previously equilibrated in water containing 0.1% formic acid. A linear gradient of methanol from 0 to 100% in the presence of 0.1% formic acid was applied for 45 min at 1 mL min^{-1}. Collected fractions (1 mL) were dried using a vacuum centrifuge concentrator SpeedVac (UniEquip) and then solubilized in 1 mL of dimethyl sulfoxide (DMSO).

Chemical Characterization of the Wood Extracts. Gas chromatography coupled to mass spectrometer (GC-MS) allowed the identification and relative quantification of the different substances present in the wood extracts. Samples were analyzed as trimethylsilyl derivatives using the following procedure. In a screw-capped vial, a sample of ∼1 mg of dry extract was dissolved in 100 μL of BSTFA/TMCS 1% (N,O-Bis(trimethylsilyl)trifluoroacetamide with 1% of trimethylsilyl chloride from Sigma-Aldrich). The solution was vortexed-stirred for ∼1 min and heated at 70 °C for 20 h. After evaporation of the silylating reagent, the residue was diluted in 1 mL of ethyl acetate. Silylation is performed to ensure as far as possible the extractives thermal stability in the gas-chromatography temperature conditions, mainly to avoid any decomposition or dehydration side-reactions possibly occurring when hydroxyl groups are free. The GC-MS analysis was performed on a Clarus 600 GC gas chromatograph coupled to a SQ8 mass spectrometer (PerkinElmer). Separation was carried out on a 5% diphenyl/95% dimethyl polysiloxane fused-silica capillary column (J&W Scientific DB-5MS, 30 m × 0.25 mm × 0.25 μm). The injection was performed at 250 °C in the splitless mode with helium as carrier gas at a constant flow of 1 mL min^{-1}. Chromatographic conditions were as follows: initial temperature 80 °C, 2 min isothermal conditions, 10 °C min^{-1} to 190 °C, 15 °C min^{-1} to 280 °C, 10 min isothermal conditions, 10 °C min^{-1} to 300 °C, 14 min isothermal conditions. The component ionization was performed by electron impact (70 eV ionization energy) to achieve their identification by mass spectra comparison with the NIST library. Samples relative compositions in extractives were obtained by the area of each peak determined on the total ion current (TIC) chromatogram divided by the sum of all of the detected peak areas.

Glutathione Transferases. GSTs that belong to the omega class have been heterogeneously produced using synthetic genes as described in Deroy et al.[16] The synthetic genes have been designed from the sequenced genome and transcriptomic studies of the white-rot *Trametes versicolor* from the following genes: TvGSTO1S, whose accession number in the JGI database is Tv75639; TvGSTO2S: Tv56280; TvGSTO3S: Tv48691; TvGSTO4S: Tv65402; TvGSTO5S: Tv54358; TvGSTO6S: Tv23671.[17]

Fluorescence-Based Thermal Stability Assay. This assay was performed as described in Deroy et al.[16] The denaturation temperature (T_d), which corresponds to the temperature where the protein is 50% unfolded, was determined using the first derivative of the obtained data in the presence or absence of potential ligands. As reference, experiments were conducted by adding DMSO only, allowing the determination of T_d ref. Then, the difference between the denaturation temperature of the protein incubated with wood extracts and with DMSO only (T_d ref) were calculated to obtain the thermal shift (ΔT_d). The sum of the absolute values of ΔT_d of the six TvGSTOS studied, $\Sigma\Delta T_d$, was determined for each wood extract.

Inhibition Kinetics. Glutathione transferase activity of TvGSTO2S (10 nM) has been tested using phenethyl isothiocyanate (PEITC) (25−250 μM) and reduced glutathione (GSH) (1 mM) in 100 mM phosphate buffer pH 6.4. The appearance of the glutathionylated product was followed measuring the absorbance at 274 nm in the presence and absence of oxyresveratrol. The catalytic constants (k_{cat}, K_M, and K_I) were calculated using the GraphPad software with the nonlinear regression based on the Michaelis−Menten model and the mixed model inhibition.

Real-Time Molecular Interaction Study. The binding of oxyresveratrol onto TvGSTO2S was investigated in real-time with an MPC-48-2-R1-S biochip placed in a biosensor analyzer switchSENSE DRX (Dynamic Biosensors GmbH, Planegg, Germany) available at the ASIA platform (Université de Lorraine). The GST was solubilized at 200 nM in 10 mM sodium phosphate buffer, pH 7.4, containing 40 mM NaCl, 0.05% Tween20, 50 μM EDTA, and 50 μM EGTA (PE40 buffer). The chip is composed of four independent channels, each containing six electrodes (for more information, see Langer et al.).[20] Real-time measurements of kinetics responding to changes to the molecular environment upon analyte binding (oxyresveratrol) gave the association kinetics (k_{on}) of the interaction. Oxyresveratrol at 100 μM in PE40 buffer containing 2% DMSO was injected in the fluidic at 5 μL min^{-1} for 5 min to a chosen channel of the chip (association kinetics), and then PE40 buffer containing 2% DMSO was injected at 50 μL min^{-1} for 30 min (dissociation kinetics). The real-time measurements were determined at 25 °C. Slight release of the GST from the dsDNAs was observed when only buffer was injected in the fluidic, and the corresponding signal was subtracted to normalize the signal of the interaction. All curves were analyzed by nonlinear fitting of single-exponential functions with the switchANALYSIS software from Dynamic Biosensors.

Microbial Growth. The growth of *Phanerochaete chrysosporium* RP78 (genome available on the Web site of the Joint Genome Institute, https://genome.jgi.doe.gov/Phchr2/Phchr2.home.html) has

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

been followed by laser nephelometry.[21] For inoculum preparation, spores were collected from 8 day old solid cultures by adding broth followed by gentle scraping of the agar plates. The wells of the microplates were filled as previously.[22] Growth was automatically recorded for ≥30 h at 37 °C using a nephelometric reader (NEPHELOstar Galaxy, BMG Labtech, Offenburg, Germany). The maximal growth rates (μ_{max}) were determined from the growth curves measuring the maximal slope as described by Joubert et al.[21]

The bacterial strains *Collimonas pratensis* Ter91 and *Burkholderia fungorum* LMG 16225 have been isolated from *Phanerochaete chrysosporium* mycosphere during a microcosm experiment on beech.[23,24] Growth was automatically recorded for ≥48 h at 25 °C using a nephelometric reader. The growth rates have been determined from the growth curves measuring the maximal slope during the exponential phase.

The growth of *Saccharomyces cerevisiae* (strain 23344C) was studied in the presence of fractions from the acetonic extract of *Bagassa guianensis*. For each fraction, the growth tests have been performed in 198 μL of culture of *S. cerevisiae* in liquid YPD medium (Yeast, Peptone, and D-glucose). Two microliters of fraction was added to the culture, and the growth was analyzed by following the turbidimetry at 600 nm during 36 h (one read per 30 min). A control condition by testing the effect of the solvent (2 μL of DMSO) was also performed.

The inhibition index (I) of the microbial growth was calculated using the formula $I = 1 - (\mu_{max}/\mu_{max}\text{control})$ (1), where μ_{max} and μ_{max}control correspond to the maximal growth rate obtained in the presence and absence of wood extracts, respectively.

Antioxidative Properties. *Reducing Power Activity.* To determine the reducing activity of each wood extract, the method of Oyaizu[25] and Yen and Chen[26] was used as described in Canabady-Rochelle et al.[27] The reducing power is expressed according to a calibration curve based on the using of ascorbic acid (equivalent of ascorbic acid in $\mu g\ L^{-1}$).

Phenolic Content. Total phenolic content of the tested wood extract was estimated using the Folin−Ciocalteu method adapted to a 96-well microplate.[28] The absorbance was measured at 735 nm by using a microplate reader (EnSight Multimode Plate Reader, PerkinElmer). The phenolic content is expressed according to a calibration curve based on the use of gallic acid (equivalent of gallic acid in $\mu g\ L^{-1}$).

Statistical Analysis. All statistical analysis including Pearson correlations between variables, principal component analysis, ANOVA, and Tukey's test have been obtained from the XLSTAT software (XLSTAT 2017 Microsoft Excel, Addinsoft, France, 2017). For performing the Pearson correlations, the antimicrobial properties were represented by the inhibition index (I, as described above); the absolute values of thermal shifts ($|\Delta T_d|$) were required, and the values of the phenolic content and reductase activity were not modified (corresponding to equivalents of gallic acid and ascorbic acid in $\mu g\ L^{-1}$, respectively). These data have been also used for the principal component analysis.

Crystallographic Study. *Crystallogenesis.* A first screening of 288 crystallization conditions was carried out at the CRM^2 crystallogenesis platform (Université de Lorraine) by using the vapor diffusion method with an Oryx 8 crystallogenesis robot (Douglas Instrument). Crystals were optimized in Linbro plates using the hanging-drop method at 4 °C. The best crystals of TvGSTO2S were obtained by mixing 1 μL of protein (18 g L^{-1}) with 0.2 μL of crystal seed stock (prepared by crushing small crystals obtained during the screening step) and 1 μL of a solution containing 10.7% PEG 4000, 0.1 M HEPES-MES buffer at pH 7.0 (at a ratio of 4:6, respectively), 0.05 M sodium acetate, and 0.05 M magnesium chloride. The reservoir contained 1 mL of the same crystallization condition. Crystals of TvGSTO2S−oxyresveratrol complex were prepared by the "dry soaking method" as explained previously.[17] Briefly, 0.1 μL of commercial oxyresveratrol (100 mM in DMSO) was deposited on a cover slide and left to complete evaporation. Then, 1 TvGSTO2S crystal together with 1 μL of its mother liquor was dispensed on the dried oxyresveratrol. After one-day incubation, the crystal did not show any damage and was flash frozen after a quick soaking in its mother liquor supplemented with 20% glycerol.

Data Collection, Processing, and Refinement. Preliminary X-ray diffraction experiments were carried out in house on an Agilent SuperNova diffractometer (Oxford Diffraction) equipped with a CCD detector. Data collections were carried out at the ESRF on beamline FIP BM30A (Grenoble, France). TvGSTO2S crystals diffracted up to 2.19 Å. Data sets were indexed and integrated with XDS[29] and scaled and merged with Aimless from the CCP4 suite.[30] The structure of TvGSTO2S was solved by molecular replacement using MOLREP with the coordinates of *Trametes versicolor* GSTO3S (PDB ID: 6F43) as the search model. A file of restraints for oxyresveratrol was generated with the GRADE server (http://grade.globalphasing.org/cgi-bin/grade/server.cgi). Structures were refined with PHENIX[31] and manually improved with COOT.[32] Validation of all structures was performed with MolProbity[33] and the PDB validation service (http://validate.wwpdb.org). Coordinates and structure factors have been deposited in the Protein Data Bank under accession IDs 6GIB and 6GIC.

RESULTS

Interactions between Wood Extracts and TvGSTOs. Heartwoods of *Peltogyne venosa*, *Dicorynia guianensis*, *Bagassa guianensis*, *Hymenaea courbaril*, *Tabeuia serratifolia*, *Sextonia rubra*, *Andira coriacea*, *Eperua falcata*, and *Eperua grandiflora* have been sequentially extracted using four solvents exhibiting different polarities (dichloromethane, acetone, toluene/ethanol, and water). The interactions between the 32 obtained wood extracts and the six GSTs from *Trametes versicolor* were studied using the thermal shift assay (TSA), a high-throughput ligand-screening method based on the modification of protein thermal denaturation. Along a gradient of temperature, the denaturation is followed by monitoring fluorescence enhancement of a probe (SYPRO Orange) that binds to protein hydrophobic patches upon denaturation process. This TSA method has been successfully used to detect interactions between proteins and libraries of molecules.[16,17] The shift of the thermal denaturation temperature (ΔT_d) induced by the tested wood extract dissolved in DMSO was measured in comparison to a control performed with pure DMSO. Each enzyme displays a specific pattern of interactions with the tested extracts (Table S1). The interactions between each isoform and the tested extracts are not significantly related to the extraction solvent used (Tukey's test, $p > 0.05$). Nevertheless, the global reactivity of the six isoforms, named $\Sigma|\Delta T_d|$ (corresponding to the sum of the absolute values of ΔT_d, which have been measured for all isoforms), was significantly correlated to the polarity of the used solvent. The dichloromethane extracts indeed induced a significantly higher $\Sigma|\Delta T_d|$ than the other extracts (Tukey's test, $p = 0.015$), suggesting that the hydrophobic compounds extracted with this solvent have a significant effect on the thermal stability of the GSTs. As shown for *Eperua falcata* dichloromethane extract, these hydrophobic molecules could increase or decrease the thermal denaturation temperature depending on the considered proteins. This extract indeed decreased the thermal stability of TvGSTO3S (ΔT_d = −7.5 °C) and TvGSTO5S (ΔT_d = −2.1 °C) and had a positive effect on the thermal stability of TvGSTO1S (ΔT_d = +3.7 °C) and TvGSTO6S (ΔT_d = +3.0 °C). We chose not to consider these potential opposite effects, and the following analysis was performed using the absolute value of the detected ΔT_d ($|\Delta T_d|$).

Properties of Wood Extracts. The antimicrobial activity of the extracts was tested against a wood decaying

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

Table 1. Pearson Coefficients Showing the Correlation between Variables[a]

variables	(I) *Bf*	(I) *Cp*	(I) *Pc*	$\|\Delta T_d\|$ 1	$\|\Delta T_d\|$ 2	$\|\Delta T_d\|$ 3	$\|\Delta T_d\|$ 4	$\|\Delta T_d\|$ 5	$\|\Delta T_d\|$ 6	A_P	A_R
(I) *Bf*	1	**0.805**	**0.606**	−0.199	**0.606**	−0.119	0.178	0.319	0.072	0.194	0.201
(I) *Cp*		1	**0.486**	−0.239	**0.464**	−0.067	0.303	0.248	−0.072	0.026	0.038
(I) *Pc*			1	−0.132	**0.519**	−0.029	0.167	0.225	0.041	0.034	0.005
$\|\Delta T_d\|$ 1				1	0.181	**0.730**	0.091	0.309	**0.418**	−0.088	−0.073
$\|\Delta T_d\|$ 2					1	−0.002	0.052	**0.529**	0.213	**0.347**	**0.375**
$\|\Delta T_d\|$ 3						1	0.327	0.134	**0.437**	−0.313	−0.321
$\|\Delta T_d\|$ 4							1	0.083	0.176	−0.096	−0.072
$\|\Delta T_d\|$ 5								1	−0.139	−0.137	−0.035
$\|\Delta T_d\|$ 6									1	0.265	0.195
A_P										1	**0.953**
A_R											1

[a]Values in bold have *p*-values lower than the significant threshold ($p < 0.05$) and are significantly different. Symbol (I) corresponds to the inhibition index on the studied microorganisms: *Bf* for *Bacillus fungorum*, *Cp* for *Collimonas pratensis*, and *Pc* for *Phanerochaete chrysosporium*. The absolute values of thermal shift ($|\Delta T_d|$) for the six isoforms TvGSTO1S (1), TvGSTO2S (2), TvGSTO3S (3), TvGSTO4S (4), TvGSTO5S (5), and TvGSTO6S (6) have been required for performing this Pearson correlation. For the phenolic content (A_P) and reductase activity (A_R), the same values were used (equivalent of gallic acid and of ascorbic acid in $\mu g\ L^{-1}$).

basidiomycete *Phanerochaete chrysosporium* and two bacterial strains isolated from degraded wood, *Collimonas pratensis* Ter91 and *Burkholderia fungorum* LMG 16225.[23,24] The three organisms were cultivated in the presence or absence of the tested wood extracts. The potential antimicrobial effect was calculated from maximal growth rates measured in the presence or absence of the tested extracts. A statistical analysis of the obtained data (Table S1) shows that these "antimicrobial" variables are significantly correlated (Table 1) with a Pearson correlation coefficient higher between both bacterial variables ($r = 0.805$, $p = 1.7 \times 10^{-8}$).

For investigating the antioxidant activity, two distinct methods were used: determination of phenolic content and reductive abilities. These "antioxidative variables" were revealed to be strongly correlated ($r = 0.953$, $p = 10^{-12}$) (Table 1).

Relationship between Thermal Responses of TvGSTOs and the Biological Properties of Wood Extract. Potential correlations between "TvGSTOS" variables (defined as the absolute value of the detected ΔT_d for a specific isoform) and antioxidative and antimicrobial variables were investigated (Table 1). No significant correlation was detected with five isoforms (TvGSTO1S, 3S, 4S, 5S, and 6S). In contrast, interactions between the "TvGSTO2S" variable and the tested wood extracts are clearly related to the chemical properties of the latter (Table 1). The "TvGSTO2S" variable is indeed correlated both to the "antioxidative variables" (phenolic content and reductase activity; $r = 0.347$, $p = 0.048$, and $r = 0.375$, $p = 0.031$, respectively) and to the "antimicrobial variables" ($r = 0.519$, $p = 0.002$; $r = 0.606$, $p = 0.001$; and $r = 0.464$, $p = 0.007$ for *P. chrysosporium*, *B. fungorum*, and *C. pratensis*, respectively) (Table 1).

A principal component (PC) analysis was also performed using all these data (Figure 1).

Three PCs explained ~69.25% of the total variance among the 32 tested extracts. Of these, the first two PCs explained ~50.76% of the total variance. The overall distribution of the wood extracts on the graph does not show a clear clustering, although all of the samples are clearly distinguishable from the control (DMSO). Briefly, the PC analysis showed that the coordinates of several wood extracts obtained with water, mix toluene/ethanol, and acetone were near among them. In contrast, it appeared that the coordinates of wood extracts

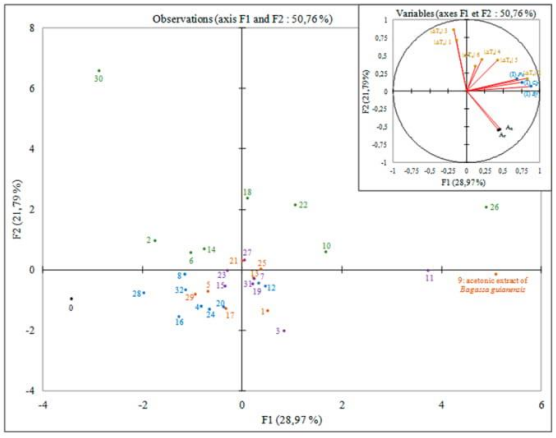


Figure 1. Principal component (PC) analysis performed using "$|\Delta T_d|$ TvGSTO1-6S", "antimicrobial", and "antioxidative" variables (Table S1) obtained with the following wood extracts. For each wood extract, absolute values of thermal shifts for the six isoforms were required for this PC. According to the protein (TvGSTO1-6S), the variables are represented by "$|\Delta T_d|$ 1", "$|\Delta T_d|$ 2", "$|\Delta T_d|$ 3", "$|\Delta T_d|$ 4", "$|\Delta T_d|$ 5", and "$|\Delta T_d|$ 6". The inhibition indexes on the growth of the three microorganisms were used, and the variables are represented in blue: (I) *Pc* for inhibition on the growth of *P. chrysosporium*, (I) *Cp* for *C. pratensis*, and (I) *Bf* for *B. fungorum*. Concerning both antioxidative variables, the phenolic content (A_P) and reductase activity (A_R) are symbolized in black. Each number corresponds to a wood extract (see Table S1). According to the solvent, the color is different: green for dichloromethane, orange for acetone, purple for the mix toluene/ethanol, and blue for water. The number 0 corresponds to the control condition (for dimethyl sulfoxide).

obtained with dichloromethane were different and were heterogeneously distributed. Interestingly, the acetonic and the toluene/ethanolic extracts of *Bagassa guianensis* differ from the others (Figure 1, nos. 9 and 11), displaying both antifungal and antibacterial properties and modifying the thermal stability of TvGSTO2S (ΔT_d = +5.25 °C, ΔT_d = +2.00 °C respectively).

Taken together, these data suggested that the detected interactions between wood extracts and TvGSTO2S could be used to detect molecules with biological properties. The

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

acetonic extract of *Bagassa guianensis* was then chosen to focus this work on the potential interactions with TvGSTO2S.

Study of the Chemical Composition and Properties of the Acetonic Extract of *Bagassa guianensis*. The chemical composition of the acetonic extract of *Bagassa guianensis* was first analyzed by GC-MS, showing the presence of oxyresveratrol as the main product (peak at 21.82 min; Figure S1) in accordance with previous studies.[34] The other peaks (19.42, 20.81, 21.10, 22.46, and 25.72 min) revealed the presence of derivatives of oxyresveratrol: resveratrol and cyclization forms of oxyresveratrol. After separation of the extract by reversed-phase high-performance liquid chromatography (HPLC), the harvested fractions were evaporated. The fractions were solubilized in DMSO and then tested for the content of phenolic compounds, the reducing power, their antimicrobial activity (in this case, the inhibition of *Saccharomyces cerevisiae* growth rate), and their interactions with TvGSTO2S (Table S2). As shown in Figure 2, a strong correlation between the antioxidative properties of the fractions and their interactions with TvGSTO2S was observed ($r = 0.957$, $p = 10^{-12}$).

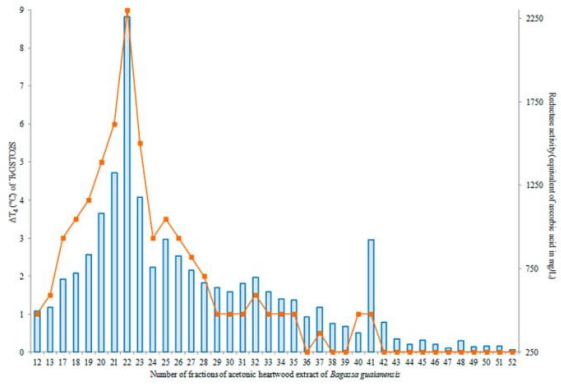


Figure 2. Thermal effects on TvGSTO2S and reductase activity of fractions from acetonic heartwood extract of *Bagassa guaianensis*. Blue bars represent the reductase activity (converted in equivalent of ascorbic acid in mg L^{-1}), and the orange line corresponds to the thermal shifts (ΔT_d, in °C) induced by each fraction.

Interactions with TvGSTO2S revealed mainly the presence of two pools of fractions centered around fractions 22 and 41, respectively, that induced an increase in the thermal stability of the protein (Figure 2). The GC-MS analysis revealed that fraction 22 mainly contained oxyresveratrol and dihydromorin and that fraction 41 contained 6-O-methyl-moracin N. All these compounds have been previously detected in *Bagassa guianensis* extract.[34] 6-O-Methyl-moracin N has been shown to exhibit antimicrobial activity, and dihydromorin shows a tyrosinase inhibitory effect.[19,35] Because of the major presence of oxyresveratrol in this extract, we chose to test the pure compound. For the next part of this work, a biochemical and structural study has been conducted to investigate the interactions between TvGSTO2S and oxyresveratrol at the molecular level.

Characterization of the Interaction between TvGSTO2S and Oxyresveratrol. To better understand the effect of oxyresveratrol on the denaturation temperature (ΔT_d) of TvGSTO2S, we tested several concentrations of stilbene with the same method (Figure 3).

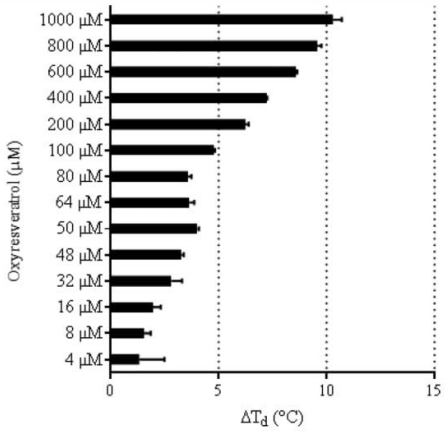


Figure 3. Stabilizing effects of oxyresveratrol on TvGSTO2S. Thermal shifts (ΔT_d) are given in °C. For each tested concentration of oxyresveratrol, the final concentration of protein was 10 μM.

ΔT_d of TvGSTO2S is significantly modified with at least 8 μM oxyresveratrol (ΔT_d = +1.55 °C). This approach did not allow an accurate determination of the apparent affinity of TvGSTO2S for oxyresveratrol.[36] Then, an enzymatic approach was performed using the ability of TvGSTO2S to transfer glutathione on PEITC. In the absence of oxyresveratrol, the following parameters have been obtained: K_M = 29.71 ± 2.84 μM and k_{cat} = 12.01 ± 0.26 s^{-1}, confirming the ability of TvGSTO2S to efficiently transfer glutathione on this substrate.[16] This activity is strongly inhibited in the presence of oxyresveratrol (K_I of 3.79 ± 1.51 μM), confirming that the stilbene is able to bind to the studied protein. Next in the study of interactions between TvGSTO2S and oxyresveratrol, an association constant k_{on} of 141 ± 18 M^{-1} s^{-1} was determined using the switchSENSE technology (Figure S2). However, the dissociation constant k_{off} was not determined as the oxyresveratrol and the protein formed a highly stable complex (no dissociation observed, Figure S2). In addition, no modification of the stilbene mediated by the TvGSTO2S was observed in the presence or absence of glutathione, suggesting that the stilbene is not a substrate for the protein.

The interactions between TvGSTO2S and oxyresveratrol were also studied at the atomic level. The TvGSTO2S crystal structure was solved at 2.19 Å (Table S5). The protein adopts a dimeric arrangement with its monomers exhibiting the characteristic traits of the cytosolic GST fold (Figure 4).

Besides its conserved two-domain organization (N-terminal domain, $\beta1\alpha1\beta2\alpha2\beta3\beta4\alpha3$, residues Cys3-Ala95; C-terminal domain, $\alpha4\alpha5\alpha6\alpha6'\alpha7\alpha8\alpha9$, Lys96-Arg236), TvGSTO2S shows the specificities of the fungal type-III Omega class, an extended loop between strands $\beta3$ and $\beta4$, a potential catalytic serine at the N-terminus of helix $\alpha1$, and an additional helix $\alpha6'$.[17] TvGSTO2S shares the same overall structure as its previously studied isoforms TvGSTO3S (rmsd of 1.366 Å for 416 aligned Cα, PDB ID: 6F43)[17] and TvGSTO6S (rmsd of 2.223 Å for 460 aligned Cα, PDB ID: 6F70).[17]

The TvGSTO2S crystal soaked with oxyresveratrol resulted in a structure without major conformational changes with

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

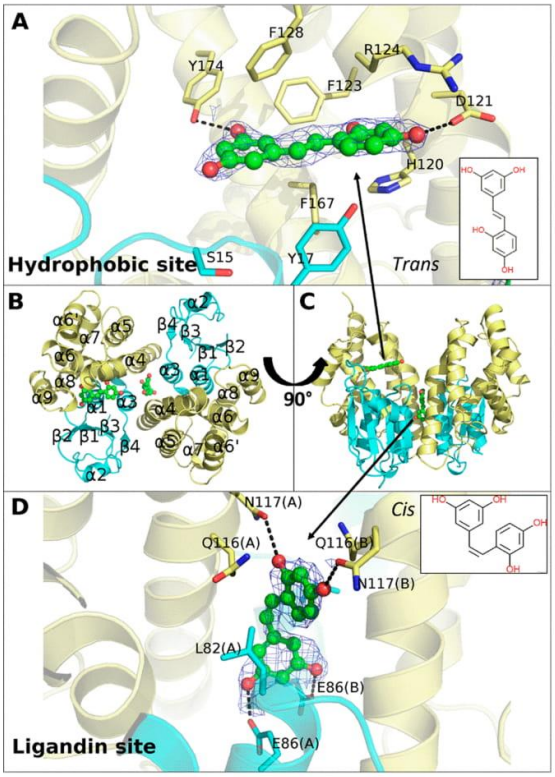


Figure 4. Crystal structure of TvGSTO2S complexed with *trans*- and *cis*-oxyresveratrol. (A) Oxyresveratrol in *trans* configuration bound to the TvGSTO2S hydrophobic site. (B,C) Overall views of TvGSTO2S structure complexed with *trans*- and *cis*-oxyresveratrol. Secondary structure elements are displayed in (B). N-terminal domain is colored cyan, and C-terminal domain is colored pale yellow. (D) Oxyresveratrol in *cis* configuration bound to the TvGSTO2S ligandin site at the dimer interface. Topological structures of *trans*- and *cis*-oxyresveratrol are shown in corresponding panels. Ligands are represented as green sticks and spheres. Polar interactions are represented as dashed lines. 2mFo-DFc omit maps calculated by PHENIX are displayed at 1.0σ around ligands.

respect to the apo form. However, two main continuous regions were detected in the Fo-Fc electron density map corresponding to two oxyresveratrol molecules. The first one binds in the hydrophobic site (H site) of monomer B of TvGSTO2S (Figure 4). In this case, the ligand is in a native *trans* configuration and is stabilized in a slit formed by the hydrophobic side chains of Phe123, Phe128, and Phe167. Oxyresveratrol molecule in *trans* configuration interacts by H-bonding with Asp121 and Tyr174 from helix α4 and α6, respectively. The second molecule of oxyresveratrol binds in the site known as ligandin site (L site) at the TvGSTO2S dimer interface. Here, it adopts a *cis* configuration according to the aromatic groups positioned on the same side of the central double bond (Figure 4). Stilbenes, including oxyresveratrol, are known to undergo *trans*/*cis* isomerization, which can be achieved by light irradiation.[37] This likely occurred in our experiment during the drying step before crystal dry-soaking. Oxyresveratrol molecule in *cis* configuration is well stabilized by four H-bonds established between the ligand hydroxy groups and the side chains of Glu86 (monomer A), Glu86 (B), Asn117 (A), and Asn117 (B) located at the dimer interface. These structural results support the strong interaction measured between TvGSTO2S and oxyresveratrol and suggest that this GST is able to interact with both *cis* and *trans* isomers at different sites.

DISCUSSION

Forests, and in particular tropical forests, are reservoirs of amazing biodiversity, where trees play a major role in interacting with a large number of biological partners. These interactions are often due to the exchange of molecular signals. The molecular diversity of the involved compounds is linked to the extraordinary variability of plant secondary metabolism. For centuries, humans have used this diversity to obtain compounds or mixtures of compounds with antimicrobial and insecticidal activity. This is the case, for instance, of wood extractives. Depending on the considered species, these compounds may have functions in wood durability and may also exhibit antimicrobial and insecticidal activities. Numerous studies have indeed been devoted to the identification and characterization of these molecules.[1,3,38−40] On another hand, the wood decayers, microorganisms and insects, have developed efficient detoxification systems to deal with these compounds and also to use wood as a source of nutriments. From comparative genomic and previously from functional studies, it appears that these systems are mainly constituted of multigenic families.[5,7,41] The resulting proteins are involved in either oxidation of wood metabolites (e.g., cytochrome P450 monooxygenases) or a decrease of their chemical reactivity by addition of sugars or glutathione. These latest reactions are catalyzed by glycosyl or glutathione transferases depending on the chemical nature of the considered adduct.

The relationship between both diversities, i.e., wood extractives and detoxification enzymes, seems obvious and suggests that the latter could be used to study the chemical diversity of wood extracts and then to identify molecules that possess interesting properties for industry such as antimicrobial or antioxidative properties. In the present study, we have validated this hypothesis studying interactions between six GSTs from the white-rot *Trametes versicolor* and a library of wood extracts. The thermal shift assay has been used; this method allows fast screening of hundreds of samples using a small amount (μg) of extracts. The obtained results demonstrated that TvGSTOSs were more prone to interact with hydrophobic compounds because dichloromethane extracts induced the highest shifts of the thermal denaturation temperatures. The six proteins show a specific pattern of interactions as previously shown with libraries of molecules.[17] Interestingly, the thermal stability of TvGSTO2S in the presence of the extracts is significantly correlated to antimicrobial and antioxidative properties of the latter. The interaction could occur at different sites of the protein. A *trans*-oxyresveratrol molecule occupies the active site (H-site), and a *cis*-oxyresveratrol molecule occupies the dimer interface (L-site). These two sites were already described as binding sites for noncatalytic ligand in TvGSTOs.[17] These observations agree with the assumed ligandin function of GSTs that can facilitate the sequestration and transport of hydrophobic molecules.[42,43] This ligandin property could be due to particular sites found at the dimer interface as demonstrated here for TvGSTO2S.[44,45] The presence of this L-site increases in this way the number of potential interactions of GSTs with

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

ligands. In accordance with the obtained biochemical data, binding of oxyresveratrol at the H-site could be responsible for the inhibition of TvGSTO2S glutathionylation activity. Nevertheless, further experiments are required to evaluate the affinity of the L-site for oxyresveratrol and the consequences of this binding on the activity of the enzyme.

As detoxification enzymes, we postulated that the substrates or ligands of GSTs could have interesting properties with potential industrial applications. This is the case for oxyresveratrol, which has been studied for its multiple properties such as anti-inflammatory activity,[46] antioxidative activity,[47] and antibacterial activity.[48] Through the example of TvGSTO2S and oxyresveratrol, the obtained data support the hypothesis that the detoxification systems of wood decayers, and in particular glutathione transferases, could be used to identify and characterize wood molecules with potential interest. At the same time, this approach should give insights on the chemical environment encountered by the studied organisms and then increase our understanding of their adaptation to their trophic features.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssuschemeng.8b02636.

Glutathione transferases, antimicrobial and antioxidative variables obtained with wood extracts; TvGSTO2S, antimicrobial, and antioxidative variables obtained with fractionated *Bagassa guianensis* acetonic extract; diffraction and refinement statistics; GS-MS analysis of acetonic extract of *Bagassa guianensis*; and real-time molecular interaction between TvGSTO2S and oxyresveratrol determined by switchSENSE analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: eric.gelhaye@univ-lorraine.fr.

ORCID

Thomas Perrot: 0000-0002-5592-7941

Author Contributions

‖T.P. and M.S. contributed equally to this work.

Author Contributions

E.G., C.D., F.F., and N.A. developed the concept and supervised this study. T.P., M.S., S.D., M.M.R., F.S., R.S., and G.S. performed the experiments and interpreted the data. All the authors participated in manuscript writing. E.G., C.D., P.G., and N.A. acquired the funding. All authors read and approved the final manuscript.

Funding

This study was funded by the French National Research Agency (ANR-11-LAS-0002−01), the Centre National de la Recherche Scientifique, the University of Lorraine, and the Région Grand Est (MS and TP Grants, PEPS-Mirabelle 2016, CPER 2014−2020, Program "Equipement mi-lourd 2016"). The authors acknowledge financial support from the "Impact Biomolecules" project of the "Lorraine Université d'Excellence" (Investissements d'avenir−ANR).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Solène Telliez, Tiphaine Dhalleine, and Sandrine Mathiot for technical assistance. We thank Sophie Mieszkin for helpful discussion concerning the bacterial strains. A sincere thank you to Jean-Pierre Jacquot for constructive criticism of the manuscript. The authors thank ESRF for beamtime and the staff of beamline BM30A for data collections. The authors appreciated the access to the "Plateforme de mesures de diffraction X" of the University of Lorraine with crystal testing.

■ REFERENCES

(1) Anouhe, J.-B. S.; Adima, A. A.; Niamké, F. B.; Stien, D.; Amian, B. K.; Blandinières, P.-A.; Virieux, D.; Pirat, J.-L.; Kati-Coulibaly, S.; Amusant, N. Dicorynamine and Harmalan-N-Oxide, Two New β-Carboline Alkaloids from *Dicorynia guianensis* Amsh Heartwood. *Phytochem. Lett.* **2015**, *12*, 158−163.

(2) Rodrigues, A. M. S.; Theodoro, P. N. E. T.; Eparvier, V.; Basset, C.; Silva, M. R. R.; Beauchêne, J.; Espíndola, L. S.; Stien, D. Search for Antifungal Compounds from the Wood of Durable Tropical Trees. *J. Nat. Prod.* **2010**, *73* (10), 1706−1707.

(3) Valette, N.; Perrot, T.; Sormani, R.; Gelhaye, E.; Morel-Rouhier, M. Antifungal Activities of Wood Extractives. *Fungal Biol. Rev.* **2017**, *31* (3), 113−123.

(4) Frankó, B.; Carlqvist, K.; Galbe, M.; Lidén, G.; Wallberg, O. Removal of Water-Soluble Extractives Improves the Enzymatic Digestibility of Steam-Pretreated Softwood Barks. *Appl. Biochem. Biotechnol.* **2018**, *184* (2), 599−615.

(5) Morel, M.; Meux, E.; Mathieu, Y.; Thuillier, A.; Chibani, K.; Harvengt, L.; Jacquot, J.-P.; Gelhaye, E. Xenomic Networks Variability and Adaptation Traits in Wood Decaying Fungi. *Microb. Biotechnol.* **2013**, *6* (3), 248−263.

(6) Riley, R.; Salamov, A. A.; Brown, D. W.; Nagy, L. G.; Floudas, D.; Held, B. W.; Levasseur, A.; Lombard, V.; Morin, E.; Otillar, R.; et al. Extensive Sampling of Basidiomycete Genomes Demonstrates Inadequacy of the White-Rot/Brown-Rot Paradigm for Wood Decay Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (27), 9923−9928.

(7) Nagy, L. G.; Riley, R.; Bergmann, P. J.; Krizsán, K.; Martin, F. M.; Grigoriev, I. V.; Cullen, D.; Hibbett, D. S. Genetic Bases of Fungal White Rot Wood Decay Predicted by Phylogenomic Analysis of Correlated Gene-Phenotype Evolution. *Mol. Biol. Evol.* **2017**, *34* (1), 35−44.

(8) Rane, R. V.; Walsh, T. K.; Pearce, S. L.; Jermiin, L. S.; Gordon, K. H.; Richards, S.; Oakeshott, J. G. Are Feeding Preferences and Insecticide Resistance Associated with the Size of Detoxifying Enzyme Families in Insect Herbivores? *Curr. Opin. Insect Sci.* **2016**, *13*, 70−76.

(9) Capolupo, M.; Valbonesi, P.; Kiwan, A.; Buratti, S.; Franzellitti, S.; Fabbri, E. Use of an Integrated Biomarker-Based Strategy to Evaluate Physiological Stress Responses Induced by Environmental Concentrations of Caffeine in the Mediterranean Mussel *Mytilus galloprovincialis*. *Sci. Total Environ.* **2016**, *563*−*564*, 538−548.

(10) Hollman, A. L.; Tchounwou, P. B.; Huang, H.-C. The Association between Gene-Environment Interactions and Diseases Involving the Human GST Superfamily with SNP Variants. *Int. J. Environ. Res. Public Health* **2016**, *13* (4), 379.

(11) Mohanty, D.; Samanta, L. Multivariate Analysis of Potential Biomarkers of Oxidative Stress in *Notopterus notopterus* Tissues from Mahanadi River as a Function of Concentration of Heavy Metals. *Chemosphere* **2016**, *155*, 28−38.

(12) Roret, T.; Thuillier, A.; Favier, F.; Gelhaye, E.; Didierjean, C.; Morel-Rouhier, M. Evolutionary Divergence of Ure2pA Glutathione Transferases in Wood Degrading Fungi. *Fungal Genet. Biol.* **2015**, *83*, 103−112.

(13) Meux, E.; Morel, M.; Lamant, T.; Gérardin, P.; Jacquot, J.-P.; Dumarçay, S.; Gelhaye, E. New Substrates and Activity of *Phanerochaete chrysosporium* Omega Glutathione Transferases. *Biochimie* **2013**, *95* (2), 336−346.

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

(14) Mathieu, Y.; Prosper, P.; Favier, F.; Harvengt, L.; Didierjean, C.; Jacquot, J.-P.; Morel-Rouhier, M.; Gelhaye, E. Diversification of Fungal Specific Class a Glutathione Transferases in Saprotrophic Fungi. *PLoS One* **2013**, *8* (11), e80298.
(15) Marinović, M.; Nousiainen, P.; Dilokpimol, A.; Kontro, J.; Moore, R.; Sipilä, J.; de Vries, R. P.; Mäkelä, M. R.; Hildén, K. Selective Cleavage of Lignin β-O-4 Aryl Ether Bond by β-Etherase of the White-Rot Fungus *Dichomitus squalens*. *ACS Sustainable Chem. Eng.* **2018**, *6* (3), 2878−2882.
(16) Deroy, A.; Saiag, F.; Kebbi-Benkeder, Z.; Touahri, N.; Hecker, A.; Morel-Rouhier, M.; Colin, F.; Dumarcay, S.; Gérardin, P.; Gelhaye, E. The GSTome Reflects the Chemical Environment of White-Rot Fungi. *PLoS One* **2015**, *10* (10), e0137083.
(17) Schwartz, M.; Perrot, T.; Aubert, E.; Dumarçay, S.; Favier, F.; Gérardin, P.; Morel-Rouhier, M.; Mulliert, G.; Saiag, F.; Didierjean, C.; et al. Molecular Recognition of Wood Polyphenols by Phase II Detoxification Enzymes of the White Rot *Trametes versicolor*. *Sci. Rep.* **2018**, *8* (1), 8472.
(18) Carlson, A.; Justo, A.; Hibbett, D. S. Species Delimitation in *Trametes*: A Comparison of ITS, RPB1, RPB2 and TEF1 Gene Phylogenies. *Mycologia* **2014**, *106* (4), 735−745.
(19) Royer, M.; Rodrigues, A. M. S.; Herbette, G.; Beauchêne, J.; Chevalier, M.; Hérault, B.; Thibaut, B.; Stien, D. Efficacy of *Bagassa guianensis* Aubl. Extract against Wood Decay and Human Pathogenic Fungi. *Int. Biodeterior. Biodegrad.* **2012**, *70*, 55−59.
(20) Langer, A.; Kaiser, W.; Svejda, M.; Schwertler, P.; Rant, U. Molecular Dynamics of DNA-Protein Conjugates on Electrified Surfaces: Solutions to the Drift-Diffusion Equation. *J. Phys. Chem. B* **2014**, *118* (2), 597−607.
(21) Joubert, A.; Calmes, B.; Berruyer, R.; Pihet, M.; Bouchara, J.-P.; Simoneau, P.; Guillemette, T. Laser Nephelometry Applied in an Automated Microplate System to Study Filamentous Fungus Growth. *BioTechniques* **2010**, *48* (5), 399−404.
(22) Tien, M.; Kirk, T. K. Lignin-Degrading Enzyme from the Hymenomycete *Phanerochaete chrysosporium* Burds. *Science* **1983**, *221* (4611), 661−663.
(23) Hervé, V.; Le Roux, X.; Uroz, S.; Gelhaye, E.; Frey-Klett, P. Diversity and Structure of Bacterial Communities Associated with *Phanerochaete chrysosporium* during Wood Decay. *Environ. Microbiol.* **2014**, *16* (7), 2238−2252.
(24) Hervé, V.; Ketter, E.; Pierrat, J.-C.; Gelhaye, E.; Frey-Klett, P. Impact of *Phanerochaete chrysosporium* on the Functional Diversity of Bacterial Communities Associated with Decaying Wood. *PLoS One* **2016**, *11* (1), e0147100.
(25) Oyaizu, M. Studies on Products of Browning Reaction. *Eiyogaku Zasshi* **1986**, *44* (6), 307−315.
(26) Yen, G.-C.; Chen, H.-Y. Antioxidant Activity of Various Tea Extracts in Relation to Their Antimutagenicity. *J. Agric. Food Chem.* **1995**, *43* (1), 27−32.
(27) Canabady-Rochelle, L. L. S.; Harscoat-Schiavo, C.; Kessler, V.; Aymes, A.; Fournier, F.; Girardet, J.-M. Determination of Reducing Power and Metal Chelating Ability of Antioxidant Peptides: Revisited Methods. *Food Chem.* **2015**, *183*, 129−135.
(28) Taga, M. S.; Miller, E. E.; Pratt, D. E. Chia Seeds as a Source of Natural Lipid Antioxidants. *J. Am. Oil Chem. Soc.* **1984**, *61* (5), 928−931.
(29) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (Pt 2), 125−132.
(30) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; et al. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67* (Pt 4), 235−242.
(31) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66* (Pt 2), 213−221.
(32) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60* (Pt 12 Pt 1), 2126−2132.
(33) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MOLPROBITY: Structure Validation and All-Atom Contact Analysis for Nucleic Acids and Their Complexes. *Nucleic Acids Res.* **2004**, *32* (Web Server issue), W615−619.
(34) Royer, M.; Herbette, G.; Eparvier, V.; Beauchêne, J.; Thibaut, B.; Stien, D. Secondary Metabolites of *Bagassa guianensis* Aubl. Wood: A Study of the Chemotaxonomy of the Moraceae Family. *Phytochemistry* **2010**, *71* (14−15), 1708−1713.
(35) Hu, S.; Zheng, Z.; Zhang, X.; Chen, F.; Wang, M. Oxyresveratrol and Trans-Dihydromorin from the Twigs of *Cudrania tricuspidata* as Hypopigmenting Agents against Melanogenesis. *J. Funct. Foods* **2015**, *13*, 375−383.
(36) Vivoli, M.; Novak, H. R.; Littlechild, J. A.; Harmer, N. J. Determination of Protein-Ligand Interactions Using Differential Scanning Fluorimetry. *J. Visualized Exp.* **2014**, *91*, 1.
(37) Maafi, M.; Al-Qarni, M. A. Φ-Order Spectrophotokinetic Characterisation and Quantification of Trans-Cis Oxyresveratrol Reactivity, Photodegradation and Actinometry. *Spectrochim. Acta, Part A* **2018**, *188*, 64−71.
(38) Bossu, J.; Beauchêne, J.; Estevez, Y.; Duplais, C.; Clair, B. New Insights on Wood Dimensional Stability Influenced by Secondary Metabolites: The Case of a Fast-Growing Tropical Species *Bagassa guianensis* Aubl. *PLoS One* **2016**, *11* (3), e0150777.
(39) Doussot, F.; De Jéso, B.; Quideau, S.; Pardon, P. Extractives Content in Cooperage Oak Wood during Natural Seasoning and Toasting; Influence of Tree Species, Geographic Location, and Single-Tree Effects. *J. Agric. Food Chem.* **2002**, *50* (21), 5955−5961.
(40) Rodrigues, A. M. S.; Amusant, N.; Beauchêne, J.; Eparvier, V.; Leménager, N.; Baudassé, C.; Espindola, L. S.; Stien, D. The Termiticidal Activity of *Sextonia rubra* (Mez) van Der Werff (*Lauraceae*) Extract and Its Active Constituent Rubrynolide. *Pest Manage. Sci.* **2011**, *67* (11), 1420−1423.
(41) Syed, K.; Shale, K.; Pagadala, N. S.; Tuszynski, J. Systematic Identification and Evolutionary Analysis of Catalytically Versatile Cytochrome P450 Monooxygenase Families Enriched in Model Basidiomycete Fungi. *PLoS One* **2014**, *9* (1), e86683.
(42) Cummins, I.; Dixon, D. P.; Freitag-Pohl, S.; Skipsey, M.; Edwards, R. Multiple Roles for Plant Glutathione Transferases in Xenobiotic Detoxification. *Drug Metab. Rev.* **2011**, *43* (2), 266−280.
(43) Zhao, J. Flavonoid Transport Mechanisms: How to Go, and with Whom. *Trends Plant Sci.* **2015**, *20* (9), 576−585.
(44) Brock, J.; Board, P. G.; Oakley, A. J. Structural Insights into Omega-Class Glutathione Transferases: A Snapshot of Enzyme Reduction and Identification of a Non-Catalytic Ligandin Site. *PLoS One* **2013**, *8* (4), e60324.
(45) Ahmad, L.; Rylott, E. L.; Bruce, N. C.; Edwards, R.; Grogan, G. Structural Evidence for Arabidopsis Glutathione Transferase AtGSTF2 Functioning as a Transporter of Small Organic Ligands. *FEBS Open Bio* **2017**, *7* (2), 122−132.
(46) Chen, Y.-C.; Tien, Y.-J.; Chen, C.-H.; Beltran, F. N.; Amor, E. C.; Wang, R.-J.; Wu, D.-J.; Mettling, C.; Lin, Y.-L.; Yang, W.-C. *Morus alba* and Active Compound Oxyresveratrol Exert Anti-Inflammatory Activity via Inhibition of Leukocyte Migration Involving MEK/ERK Signaling. *BMC Complementary Altern. Med.* **2013**, *13*, 45.
(47) Lorenz, P.; Roychowdhury, S.; Engelmann, M.; Wolf, G.; Horn, T. F. W. Oxyresveratrol and Resveratrol Are Potent Antioxidants and Free Radical Scavengers: Effect on Nitrosative and Oxidative Stress Derived from Microglial Cells. *Nitric Oxide* **2003**, *9* (2), 64−76.
(48) Joung, D.-K.; Mun, S.-H.; Choi, S.-H.; Kang, O.-H.; Kim, S.-B.; Lee, Y.-S.; Zhou, T.; Kong, R.; Choi, J.-G.; Shin, D.-W.; et al. Antibacterial Activity of Oxyresveratrol against Methicillin-Resistant *Staphylococcus aureus* and Its Mechanism. *Exp. Ther. Med.* **2016**, *12* (3), 1579−1584.

DOI: 10.1021/acssuschemeng.8b02636
ACS Sustainable Chem. Eng. XXXX, XXX, XXX−XXX

Fungal glutathione transferases as tools to explore the chemical diversity of Amazonian wood extractives

Thomas Perrot, Mathieu Schwartz, Fanny Saiag, Guillaume Salzet, Stéphane Dumarçay, Frédérique Favier, Philippe Gérardin, Jean-Michel Girardet, Rodnay Sormani, Mélanie Morel-Rouhier, Nadine Amusant, Claude Didierjean and Eric Gelhaye

Supporting Information

Table of content (page S1)

Supplementary Tables

Table S1: Glutathione transferases, antimicrobial and antioxidative variables obtained with wood extracts (page S2)

Table S2: TvGSTO2S, antimicrobial and antioxidative variables obtained with fractionated *Bagassa guianensis* (*Bg*) acetonic extract (page S3)

Table S3: Diffraction and refinement statistics (page S4)

Supplementary Figures

Figure S1: **GS-MS analysis of acetonic extract of *Bagassa guianensis* (page S5)**

Figure S2: Real-time molecular interaction between TvGSTO2S and oxyresveratrol determined by switchSENSE® analysis (page S6)

Table S1 : Glutathione transferases, antimicrobial and antioxidative variables obtained with wood extracts. Solvents are summarized as follows, DCM; dichloromethane, ACE; acetone, T/E; toluene/ethanol and WAT; water. Wood species names are summarized as follows, « *Pv* » for *Peltogyne venosa*, « *Dg* » for *Dicorynia guianensis*, « *Bg* » for *Bagassa guianensis*, « *Hc* » for *Hymenaea courbaril*, « *Ts* » for *Tabebuia serratifolia*, « *Sr* » for *Sextonia rubra*, « *Ac* » for *Andira coriacea*, « *Ef* » for *Eperua falcata*. The condition named « DMSO » corresponds to the control condition. For the first three columns, the values correspond to the inhibition index (I) described in the Material and Methods section. From the column « TvGSTO1S » to « TvGSTO6S » the data correspond to the thermal shifts (ΔT_d). The values of phenolic content and the reductase activity are given in equivalent of gallic acid and of ascorbic acid (in $\mu g.L^{-1}$) respectively.

Extractive	N°	Inhibition index			ΔT_d (°C)						$\Sigma\Delta T_d$ (°C)	Phenolic content (eq. in gallic acid in $\mu g.L^{-1}$)	Reductase activity (eq. in ascorbic acid in $\mu g.L^{-1}$)
		B. fungorum	*C. pratensis*	*P. chrysosporium*	TvGSTO1S	TvGSTO2S	TvGSTO3S	TvGSTO4S	TvGSTO5S	TvGSTO6S			
DMSO	0	0.00	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.03	0.00
Pv ACE	1	0.27	0.37	0.26	0.25	1.00	0.25	-0.25	0.67	1.25	3.67	3250.80	8018.11
Pv DCM	2	0.16	0.19	0.11	2.00	0.25	-1.00	-0.50	-0.59	0.50	4.84	1248.24	2627.05
Pv T/E	3	0.29	0.31	0.25	0.50	1.50	0.00	0.00	0.24	1.75	3.99	4071.97	10382.32
Pv WAT	4	0.13	0.14	0.21	0.25	0.75	0.00	-0.50	-0.08	1.25	2.83	2759.52	6779.27
Dg ACE	5	0.18	0.36	0.11	0.25	0.50	0.00	0.50	-0.29	1.00	2.54	1814.96	4931.14
Dg DCM	6	0.21	0.21	0.36	0.50	-0.50	-1.00	-0.25	-2.49	0.25	4.99	673.72	1174.42
Dg T/E	7	0.14	0.09	0.91	0.75	0.75	0.50	0.50	1.01	1.75	5.26	2874.77	7743.69
Dg WAT	8	0.10	0.20	0.25	0.50	0.25	0.00	0.50	-1.65	1.25	4.15	1883.94	3245.43
Bg ACE	9	0.47	0.72	0.92	0.25	5.25	1.00	0.75	-0.67	1.50	9.42	4071.48	11167.99
Bg DCM	10	0.32	0.51	0.97	0.75	1.75	-0.25	0.25	-1.51	1.50	6.01	1728.06	2647.27
Bg T/E	11	0.64	0.62	0.89	0.25	2.00	1.00	0.50	0.49	1.75	5.99	3147.94	7523.99
Bg WAT	12	0.24	0.44	0.33	0.50	2.25	0.00	0.00	0.78	1.50	5.03	1841.80	4314.29
Hc ACE	13	0.20	0.20	0.00	1.75	2.25	0.50	0.50	0.79	2.00	7.79	3695.20	10095.23
Hc DCM	14	0.15	0.20	0.68	1.75	1.25	-1.50	0.00	-0.77	0.50	5.77	1354.79	3606.66
Hc T/E	15	0.11	0.00	0.57	0.75	1.25	0.50	0.50	-0.58	2.00	5.58	3345.74	7551.98
Hc WAT	16	0.18	0.03	0.14	0.25	0.50	0.00	0.00	-0.78	1.50	3.03	2516.53	6500.58
Ts ACE	17	0.17	0.22	0.43	0.25	0.75	0.00	0.25	-0.41	1.00	2.66	2258.94	7492.20
Ts DCM	18	0.28	0.18	0.52	1.25	2.50	1.50	0.50	0.57	4.00	10.32	722.65	1495.72
Ts T/E	19	0.29	0.28	0.52	0.50	0.75	0.00	0.50	-0.28	1.50	3.53	1882.93	6629.83
Ts WAT	20	0.16	0.17	0.14	-0.50	0.25	0.00	-0.50	-2.40	0.50	4.15	2744.20	10243.87
Sr ACE	21	0.24	0.46	0.54	0.00	0.50	0.00	-0.50	-1.66	0.50	3.16	536.81	879.04
Sr DCM	22	0.34	0.53	0.63	-0.25	1.25	-1.00	-1.25	-2.07	0.50	6.32	110.70	0.00
Sr T/E	23	0.19	0.28	0.49	-0.25	1.00	0.00	-0.50	-1.59	0.50	3.84	1184.61	2648.66
Sr WAT	24	0.08	0.00	0.25	0.50	1.50	0.75	0.50	-0.05	1.50	4.80	3956.42	9293.62
Ac ACE	25	0.26	0.47	0.17	0.25	0.50	1.00	0.50	-1.70	1.75	5.70	2357.74	5907.61
Ac DCM	26	0.46	0.44	0.74	-2.00	-6.50	0.50	-0.25	-8.28	1.00	18.53	2599.54	8607.19
Ac T/E	27	0.27	0.37	0.26	0.75	-0.25	1.00	0.50	-1.37	1.75	5.62	2259.03	5407.53
Ac WAT	28	0.11	0.13	0.26	0.25	0.25	-0.25	0.00	-0.71	0.50	1.96	1044.70	1999.97
Ef ACE	29	0.00	0.27	0.17	0.75	0.75	0.25	0.50	-0.47	1.50	4.22	2678.17	8307.65
Ef DCM	30	0.02	0.12	0.14	3.75	0.50	-7.50	0.75	-2.08	3.00	17.58	162.33	0.00
Ef T/E	31	0.14	0.17	0.67	0.75	-2.00	0.50	0.25	0.58	1.50	5.58	2798.49	6239.21
Ef WAT	32	0.07	0.09	0.02	2.00	1.00	0.25	0.25	0.61	2.00	6.11	3707.96	9129.08

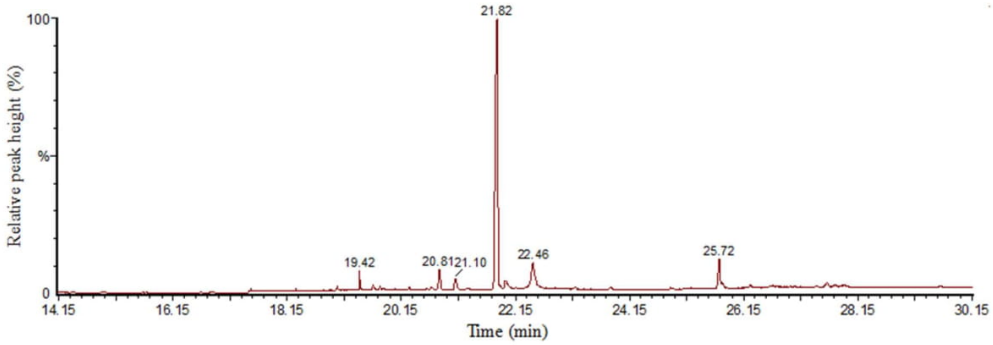
Table S2: TvGSTO2S, antimicrobial and antioxidative variables obtained with fractionated *Bagassa guianensis* (*Bg*) acetonic extract. The *Bg* acetonic extract was fractionated by reversed-phase HPLC and each harvested fraction (1 mL) was analyzed as described in the Materials and Methods section.

Fractions	ΔT_d TvGSTO2S (°C)	Reductase activity (eq. in ascorbic acid in $\mu g.L^{-1}$)	(I) *S.cerevisiae*	Phenolic content (eq. in gallic acid in $\mu g.L^{-1}$)
0	8.5	1091696.62	0.59	8599.51
1	7.0	1201733.29	0.56	9296.06
2	0.0	348171.80	0.53	133.29
12	1.0	496821.07	0.53	432.89
13	1.5	516598.95	0.39	478.20
17	3.0	687840.81	0.43	1726.34
18	3.5	721729.08	0.36	2387.13
19	4.0	832270.21	0.28	2211.36
20	5.0	1078902.05	0.20	7896.85
21	6.0	1323845.45	0.29	8108.74
22	9.0	2260937.25	0.00	13329.15
23	5.5	1178092.33	0.24	6224.87
24	3.0	756854.08	0.33	3874.90
25	3.5	925429.56	0.25	4955.32
26	3.0	824309.98	0.24	3372.01
27	2.5	739274.67	0.20	2671.77
28	2.0	666352.25	0.22	1704.17
29	1.0	635007.36	0.25	1535.30
30	1.0	610308.55	0.31	1921.17
31	1.0	659624.94	0.31	1351.16
32	1.5	696470.49	0.22	1760.18
33	1.0	610491.69	0.28	1145.47
34	1.0	568871.70	0.27	1198.26
35	1.0	561492.57	0.25	625.85
36	0.0	461916.40	0.36	476.95
37	0.5	516819.63	0.58	975.69
38	0.0	422270.92	0.44	374.54
39	0.0	404487.45	0.30	226.42
40	1.0	363146.77	0.23	124.57
41	1.0	921292.26	0.82	1231.63
42	0.0	429598.88	0.61	238.40
43	0.0	327514.90	0.31	124.04
44	0.0	293940.48	0.22	50.27
45	0.0	320253.36	0.26	34.42
46	0.0	296747.57	0.16	37.24
47	0.0	274838.32	0.19	38.72
48	0.0	318173.52	0.21	57.90
49	0.0	280902.09	0.21	35.15
50	0.0	284986.63	0.18	26.50
51	0.0	285778.25	0.25	6.35
52	0.0	262533.13	0.35	0.00
DMSO	0.0	259824.19	0.17	14.44

Table S3: Diffraction and refinement statistics

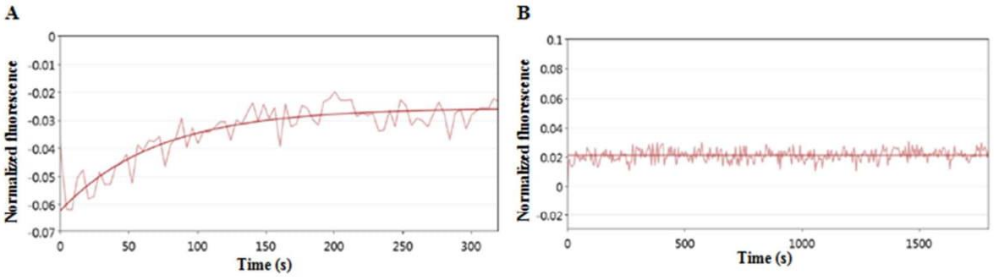
		GSTO2S	
		apo	**oxyresveratrol**
Diffraction data			
	Diffraction source	FIP-BM30A, ESRF	FIP-BM30A, ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97975	0.97975
	Unit-cell parameters		
	a, b, c (Å)	59.0 93.0 96.0	58.6 93.9 95.8
	α, β, γ (°)	90	90
	Space group	$P2_12_12_1$	$P2_12_12_1$
	Resolution range (Å)	48.02 - 2.19 (2.33 - 2.19)	47.9 - 2.30 (2.44 - 2.30)
	Total No. of reflections	198 770 (31 164)	174 134 (27 922)
	No. of unique reflections	27 630 (4 309)	24 016 (3 819)
	Average redundancy	7.2 (7.2)	7.2 (7.3)
	Mean I/σ(I)	16.5 (3.7)	13.8 (2.8)
	Completeness (%)	99.7 (98.4)	99.9 (99.9)
	R_{merge} (%)	10.4 (61.5)	13.7 (75.7)
	R_{meas} (%)	11.2 (66.1)	14.7 (81.4)
	$CC_{1/2}$ (%)	99.9 (92.0)	99.8 (85.7)
Refinement			
	Resolution range (Å)	48.02 - 2.19 (2.28 - 2.19)	47.91 - 2.30 (2.40 - 2.30)
	R_{work} / R_{free}	0.20 / 0.26 (0.27 / 0.34)	0.19 / 0.25 (0.24 / 0.33)
	No. of protein atoms	7 264	7 252
	No. of waters	67	101
	Average B factor (Å²)		
	All atoms	34.5	34.1
	Ligands		45.7
Model quality			
	RMSZ bond lengths	0.56	0.56
	RMSZ bond angles	0.63	0.63
	Ramachandran favoured (%)	98.3	98.0
	Ramachandran outliers (%)	0.2	0.2
	Molprobity rotamer outliers (%)	1.3	1.4
	Molprobity clashscore	4.13	3.28
	Molprobity score	1.29	1.22

<u>Figure S1</u> : **GS-MS analysis of acetonic extract of *Bagassa guianensis*.**



S5

<u>Figure S2:</u> Real-time molecular interaction between TvGSTO2S and oxyresveratrol determined by switchSENSE® analysis. The His-tagged TvGSTO2S was immobilized on fluorescent dsDNA nanolevers and oxyresveratrol was injected into the fluidic at 100 µM. (A) Association and (B) dissociation step.



S6

I.5 Comparaison structurale des sites de reconnaissance des polyphénols chez les TvGSTO

Cette partie dresse un bilan des différents résultats obtenus sur les TvGSTO et compare les différents sites de reconnaissance identifiés à partir des structures de complexes afin de dégager les points communs et les différences entre les isoformes étudiées. Au cours de ce projet, la famille des glutathion transférases Oméga retrouvées chez le dégradeur de bois *Trametes versicolor* a été décrite au niveau biochimique et structural. A partir des neuf isoformes de TvGSTO caractérisées, nos collaborateurs ont montré que les six isoformes de GSTO à sérine catalysent principalement des réactions GSH-transférase alors que les trois GSTO à cystéine catalysent des réactions GSH-lyase. Les structures cristallographiques de trois isoformes de TvGSTOS (TvGSTO2S, TvGSTO3S et TvGSTO6S) et des trois isoformes de TvGSTOC (TvGSTO1C, TvGSTO2C et TvGSTO3C) ont été résolues. L'analyse structurale a mis en évidence les spécificités structurales différenciant TvGSTO3C appartenant au type II des autres GSTO appartenant au type III, ces deux types ayant été précédemment distingués sur une base phylogénétique. TvGSTO3C possède notamment une hélice supplémentaire α0 en N-terminal ainsi qu'une hélice α9 recouvrant le site actif et pouvant interagir avec le glutathion lié covalamment à la cystéine catalytique. Les GSTO de type III peuvent présenter une sérine ou une cystéine comme résidu catalytique, indépendamment d'autres caractéristiques structurales conservées. Celles-ci incluent notamment une insertion dans le domaine thiorédoxine entre les brins β3 et β4 et une hélice supplémentaire α6' portée par le domaine tout hélice. De manière intéressante, les homologues structuraux les plus proches des TvGSTO identifiés par le serveur PDBeFold sont les GST de la classe Tau chez les plantes. En effet, les GSTU partagent des caractéristiques structurales et biochimiques avec les TvGSTO (Figure 49), en particulier avec les GSTO de type III : absence d'extension N-terminale, présence d'une hélice α9 recouvrant le domaine C-terminal, présence d'un résidu sérine catalytique (pour la majorité des TvGSTO de type III). Le rapprochement des deux classes Oméga et Tau a déjà été documenté auparavant lors de la comparaison structurale entre la GSTO humaine (code PDB 1EEM) et la GSTU de blé (code PDB 1GWC, Thom et al, 2001).

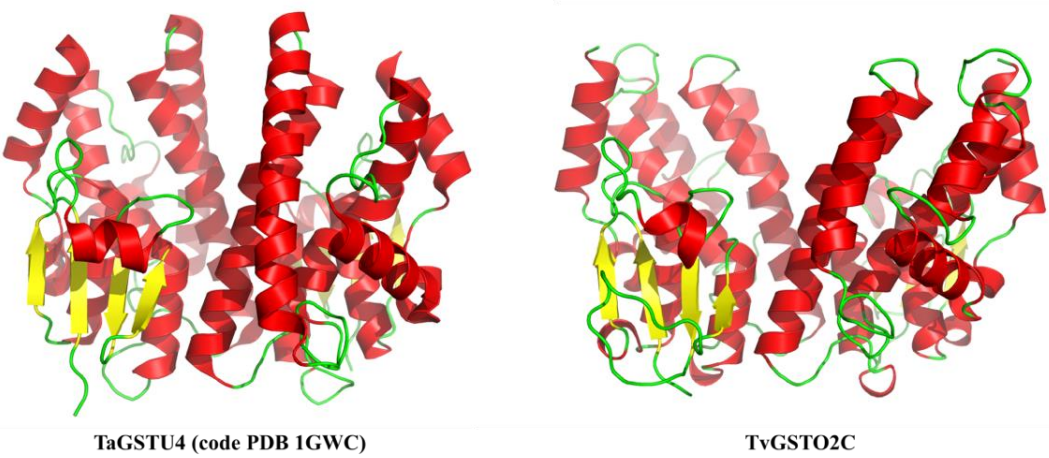


Figure 49 **Dimères de GST Tau de blé (gauche) et de TvGSTO (droite)**
Les structures sont représentées en mode rubans et flèches. Les deux structures partagent un RMSD de 1,55 Å pour 279 Cα alignés.

La détermination structurale de complexes GSTO-ligands a permis d'obtenir au niveau atomique des informations sur la capacité de ces protéines à reconnaître leurs cibles. L'isoforme TvGSTO3S, qui a été étudiée de manière approfondie, présente un site hydrophobe (site H) particulièrement large et profond capable de fixer des molécules polyaromatiques telles que le groupement phenylacétophénone du GS-PAP ou les hydroxy-benzophénones. Le site H de TvGSTO3S situé au cœur du site actif est constitué essentiellement par les chaînes latérales de résidus aromatiques, ainsi que d'une arginine et d'une tyrosine à l'entrée du site, capables d'interagir par liaison hydrogène avec les groupements polaires des ligands. Ce site H particulièrement enfoui est notamment capable de reconnaître un flavonoïde, la dihydrowogonine, présente dans un extrait partiellement purifié de bois de merisier. La comparaison structurale du site H de TvGSTO3S avec les sites H des autres isoformes de GSTO montre que l'isoforme TvGSTO3S est la seule à posséder un site avec de telles caractéristiques (Figure 50). L'isoforme TvGSTO2S possède plus de 60 % d'identité de séquence avec TvGSTO3S. Néanmoins les résidus portés par l'hélice α4 sont de nature différente (Figure 51). En conséquence, le site H de TvGSTO2S est moins profond et a une forme d'entaille. Cette poche est notamment capable de fixer le *trans*-oxyresvératrol. Concernant l'isoforme TvGSTO2C, la détermination structurale du complexe TvGSTO2C-oxyresvératrol révèle la fixation du *trans*-oxyresvératrol au niveau du site actif, à cheval entre les sites G et H. La proximité d'un groupement hydroxyle de l'oxyresvératrol avec le carboxy terminal du glutathion lié covalemment à la TvGSTO2C suggère une stabilisation par le cofacteur. En effet, les mesures de TSA montrent une meilleure interaction du *trans*-oxyresvératrol avec TvGSTO2C lorsque cette dernière est glutathionylée. De manière

générale, l'analyse des séquences montre que les positions des résidus constituant ce site sont le plus souvent de nature hydrophobe, bien que variables entre les différentes isoformes. En effet, aucun consensus n'est clairement visible pour le site H des TvGSTO. Cette particularité pourrait expliquer pourquoi chaque TvGSTO présente ses propres propriétés d'interaction. Ces nouvelles données confirment la diversité des sites hydrophobes retrouvés entre les différentes isoformes de GSTO. En revanche, aucune activité GSH-transférase ou GSH-lyase n'a été détectée sur les différentes molécules identifiées dans le site H. Cette observation pourrait être expliquée soit par l'absence à ce jour d'un substrat connu et accepté par les TvGSTO, soit par un rôle non catalytique joué par ces GST. Ces aspects seront discutés dans la partie « Discussion générale et Perspectives » de ce manuscrit.

Nos résultats structuraux sur les complexes GSTO-ligands montrent que certaines isoformes de GSTO (TvGSTO2S, TvGSTO6S et TvGSTO2C) sont capables de lier des molécules au niveau de leur interface dimérique abritant le site ligandine (site L). Ce site L est distinct du site H et accueillerait des molécules sans modification catalytique car distant du site actif de l'enzyme. Les isoformes TvGSTO6S et TvGSTO2C sont capables de fixer une paire de molécules de naringénine à ce site L alors que TvGSTO2C est capable d'y fixer l'isomère *cis* de l'oxyresvératrol. En revanche, la structure de TvGSTO3S nous montre que le dimère est plus refermé pour cette isoforme, si bien qu'aucune fixation ne parait possible à l'interface dimérique en l'absence de changement conformationnel (Figure 50). Les structures des différents complexes obtenus ont permis de définir les résidus de ce site L (Figure 51). Tout d'abord, la conservation stricte du résidu E86 (numérotation basée sur TvGSTO2S) dans l'hélice α3 suggère son importance dans la reconnaissance au site L. Cet acide glutamique intervient notamment au fond de la poche pour stabiliser la naringénine ou le *cis*-oxyresveratrol *via* leurs groupements hydroxyles. D'autres résidus situés au fond de la poche et portés par l'hélice α3 et la partie N-terminale de l'hélice α4 sont également bien conservés. Par ailleurs, une partie des résidus situés à l'entrée du site L et portés par la partie C-terminale de l'hélice α4, est plus variable. Dans la structure de TvGSTO2S, les chaînes latérales des résidus N117A,B participent notamment à la stabilisation des groupements hydroxyles du *cis*-oxyresvératrol. Ainsi, nos résultats montrent que les isoformes de TvGSTO sont capables d'interagir avec des polyphénols végétaux et que chacune d'elles présente des spécificités de résidus dans leurs sites d'interaction (H ou L). Ces résidus pourraient expliquer les propriétés spécifiques d'interaction de chaque isoforme observées lors des expériences de TSA.

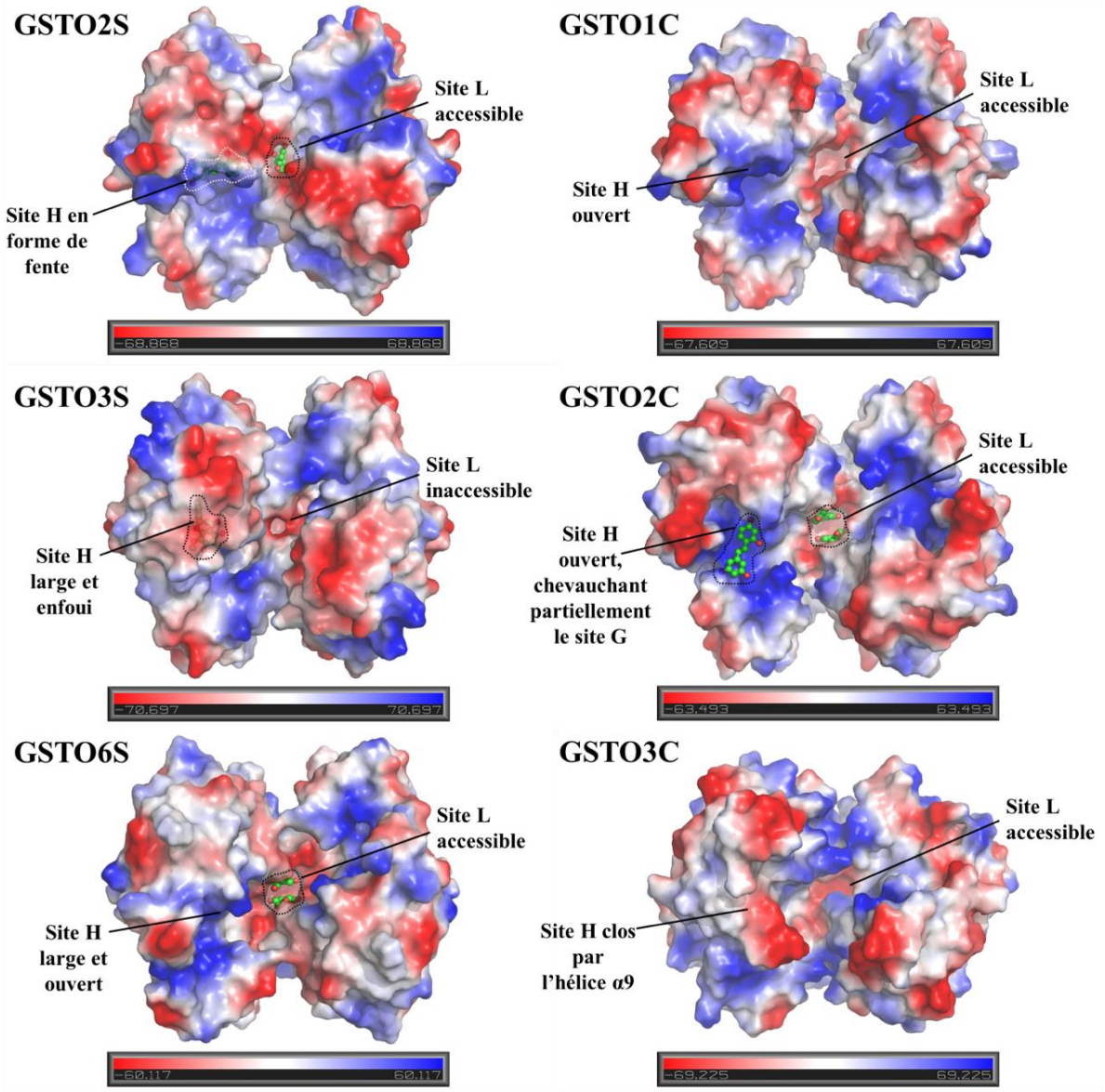


Figure 50 **Sites de liaison hydrophobes et ligandines des différentes isoformes de GSTO de structure résolue**

Les surfaces électrostatiques ont été calculées *in vacuo* à l'aide du logiciel Pymol. La position des sites H et L est rappelée et dans le cas d'un ligand identifié, celui-ci est représenté dans la structure et entouré en pointillés afin de visualiser le site.

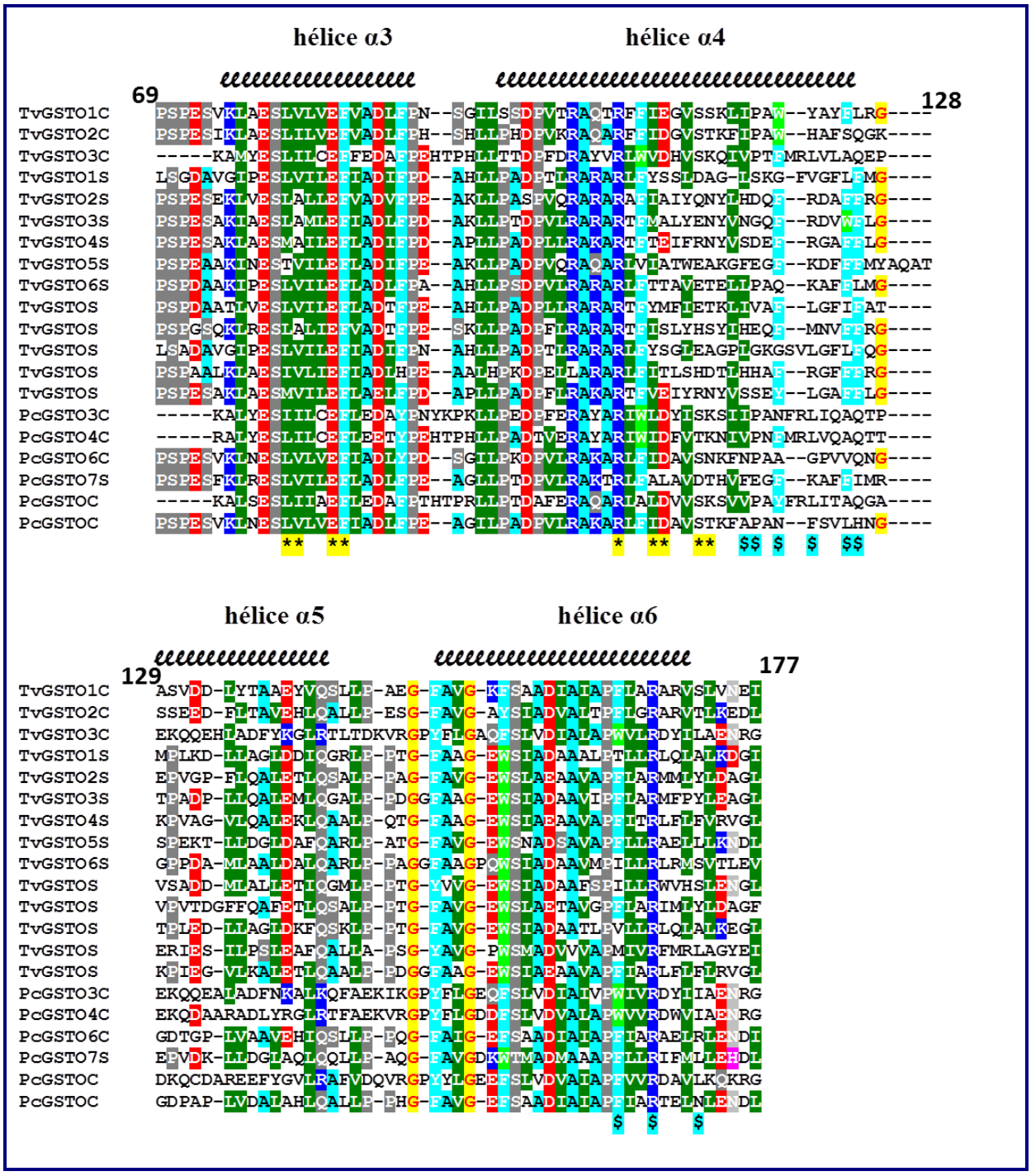


Figure 51 **Alignement des séquences partielles de GSTO de *Trametes versicolor* et *Phanerochaete chrysosporium* et mise en évidence des résidus participant aux sites H et L**

Les séquences ont été alignées avec le programme ClustalW implémenté dans BioEdit (Hall, 1999). Les séquences ont été tronquées de manière à ne représenter que la portion correspondante aux hélices α3, α4, α5 et α6. La numérotation indiquée correspond à la séquence de TvGSTO1C. Les positions des résidus du site L, définies à partir des structures résolues des complexes avec un ligand au site L sont marquées par une astérique * surlignée en jaune. Le site L peut être défini avec une partie constante portée par l'hélice α3 et la première moitié de l'hélice α4, et une partie variable portée par la deuxième moitié de l'hélice α4. Les positions des résidus du site H sont marquées par le signe $ surligné en bleu. Le site H est constitué de résidus aromatiques ou hydrophobes portés par les hélices α4 et α6 ainsi qu'un résidu arginine conservé en fin d'hélice α6.

II Etude des GST Xi de *Saccharomyces cerevisiae* et *Trametes versicolor*

Dans cette seconde partie sont présentés les résultats des études structurales et biochimiques d'une GST de classe Xi de la levure *Saccharomyces cerevisiae* d'une part, et de deux isoformes du basidiomycète saprophyte *Trametes versicolor*. La classe Xi est une classe de GST dont le résidu catalytique est une cystéine. Avant la détermination de la structure cristallographique de la GST Xi de *Phanerochaete chrysosporium* (Meux et al, 2011), cette classe était confondue avec la classe Oméga divisée phylogénétiquement en trois sous-classes dans le règne fongique (sous-classe I pour les actuelles GST Xi, sous-classe II pour les GST Oméga uniquement à cystéine et sous-classe III pour les GST Oméga à cystéine ou sérine) (Morel et al, 2009). Dans un premier temps, sont reportés les résultats obtenus à partir de l'étude biochimique et structurale de l'isoforme ECM4 de *S. cerevisiae*. Cette étude a été publiée dans la revue *PLOS One* et son manuscrit est intégré dans cette partie. La modélisation moléculaire par homologie des deux autres isoformes de *S. cerevisiae* est ensuite présentée. Une discussion structurale termine cette première partie consacrée aux GST Xi de *S. cerevisiae*. Dans un second temps, les résultats obtenus à partir de l'étude de deux GST Xi de *Trametes versicolor* sont présentés et discutés. Cette étude biochimique et structurale a fait l'objet d'une publication dans la revue *FEBS Letters*. La version finale et acceptée est incorporée à la suite d'un résumé reportant les principaux éléments discutés.

II.1 L'isoforme ScECM4 de la levure *Saccharomyces cerevisiae*

La levure *Saccharomyces cerevisiae* est un organisme eucaryote fongique modèle pour de nombreuses études de biologie moléculaire (Karathia et al, 2011). D'après Morel et collaborateurs (2009), 11 gènes de GST, répartis en 6 classes, sont présents dans le génome de *S. cerevisiae*, un nombre bien inférieur au GSTome de champignons saprotrophes tels *Trametes versicolor* (dont le génome en contient 46) (Morel et al, 2013). Sur les 6 classes présentes chez *S. cerevisiae* (GTT1, GTT2, Ure2p, EF1Bγ, MAK16, Xi (initialement classée Oméga)), seules les classes GTT2, Ure2p et EF1Bγ avaient un représentant structural dans la PDB au début de ce projet (Ma et al, 2009 ; Bousset et al, 2001a ; Jeppesen et al, 2003). De façon intéressante ces trois GST présentent des propriétés particulières. L'EF1Bγ et l'Ure2p sont des protéines multidomaines dont seule la structure du domaine glutathion transférase a

été déterminée. De plus, l'activité glutathion transférase de la GTT2 n'est pas assurée par les résidus le plus souvent observés (tyrosine, sérine ou cystéine). Une molécule d'eau stabilisée par deux résidus du domaine tout hélice assisterait la déprotonation du groupement thiol du glutathion.

Par ailleurs, l'analyse des séquences de GST chez *S. cerevisiae* a révélé trois isoformes présentant une similarité avec la GST Oméga humaine 1 (HsGSTO1) : ScGTO1, ScGTO2 (aussi appelé ScECM4) et ScGTO3 (Figure 52) (Garcera et al, 2006). Les premières études de biologie moléculaire suggèrent que le gène ECM4 (GTO2) est impliqué dans la biosynthèse de la paroi cellulaire (Lussier et al, 1997), alors que la protéine ScGTO1 serait impliquée dans la régulation de la biosynthèse des acides aminés soufrés dans le peroxysome (Barreto et al, 2006). Une première étude biochimique a permis de montrer que les trois isoformes (Garcera et al, 2006) présentent une activité thiol-transférase faisant intervenir une cystéine catalytique à l'instar de HsGSTO1. Une activité de réduction des GS-hydroquinones a ensuite été mise en évidence pour l'isoforme ScECM4 (Lam et al, 2012). Cette activité et l'analyse des séquences primaires ont permis à Lam et collaborateurs (2012) de reclasser ScECM4 parmi les glutathionyl-hydroquinone réductases (GST de classe Xi présentant une activité GHR). Nos collaborateurs (laboratoire IAM) ont produit et purifié l'isoforme ScECM4 pour comparer son activité à celle de la GHR de *P. chrysosporium* (PcGSTX1). En effet, Lam et collaborateurs (2012) n'ont pas observé pour ScECM4 d'activité avec la glutathionyl-ménadione alors que PcGSTX1 a la capacité de réduire cette molécule. Au laboratoire CRM2, une étude cristallographique a été entreprise sur ScECM4 pour tenter d'élucider les facteurs moléculaires à l'origine des différences d'activité.

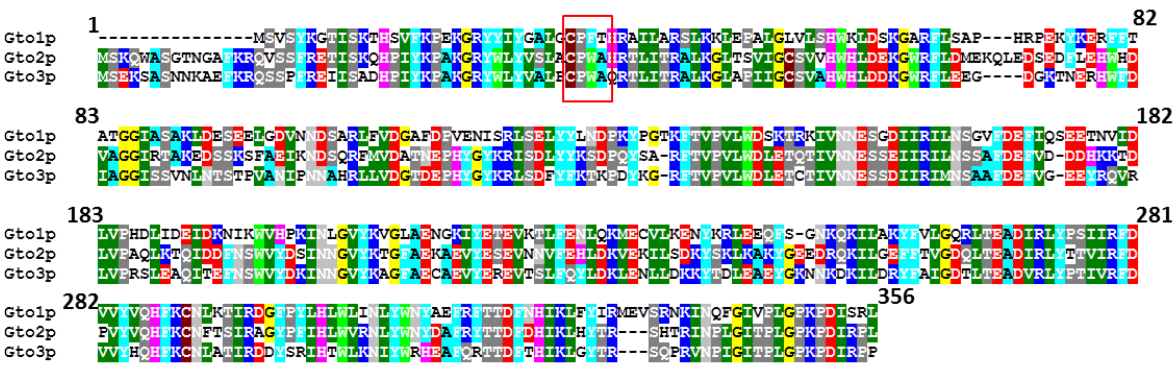


Figure 52 **Alignement des séquences de GST Xi potentielles chez la levure *Saccharomyces cerevisiae***
Les séquences ont été alignées avec ClustalW implémenté dans BioEdit. Le motif catalytique prédit est entouré en rouge. La numérotation des résidus correspond à la séquence de ScGTO1.

II.1.a. Résultats de l'étude cristallographique et enzymatique de ScECM4

La protéine ScECM4 portant une étiquette hexa-histidine a été produite, purifiée par nos collaborateurs (laboratoire IAM) puis cristallisée au CRM2 par la méthode de microbatch en goutte à l'huile. Des cristaux d'une taille appréciable (prismes jusqu'à 0,5 mm de côté, Figure 53) ont été testé sur le diffractomètre automatique de la plateforme de mesure du laboratoire. Un jeu de données a été collecté à 2,5 Å de résolution permettant une étude préalable. Des données haute résolution (1,45 Å) ont été obtenues sur la ligne BM30A du synchrotron ESRF). Le problème de la phase a été résolu par la méthode du remplacement moléculaire à partir des coordonnées de la GST Xi de *Phanerochaete chrysosporium* (PcGSTX1, code PDB 3PPU). Le dimère physiologique a été reconstruit dans l'unité asymétrique.

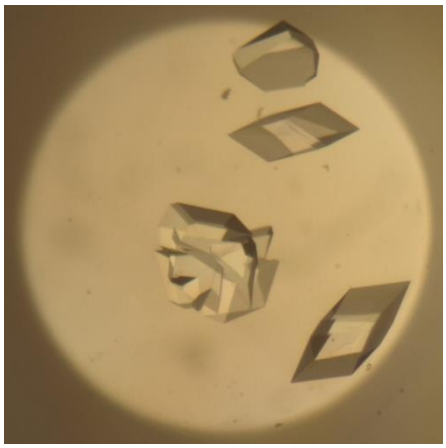


Figure 53 **Cristaux de ScECM4 obtenus par la méthode du microbatch**
Les cristaux ont été obtenus dans 20 % de PEG 2000 monomethylether et 0,1M de tampon Tris pH 7,0.

L'analyse de la structure a révélé qu'ScECM4 adopte le repliement caractéristique des GST avec les particularités de la classe Xi (Figure 54): une longue extrémité N-terminale, une boucle étendue entre le brin β2 et l'hélice α2 et un mode de dimérisation original impliquant uniquement les domaines tout hélice. En plus des caractéristiques des GST Xi, ScECM4 présente plusieurs particularités structurales :

> Une extrémité N-terminale ne se structurant pas pour participer à un petit feuillet à trois brins comme cela a été observé dans les structures des GST Xi bactériennes EcYqjG et ScPcpF (Green et al, 2012). Cette absence de structuration est également retrouvée dans les GST Xi de *P. chrysosporium* et de *Populus trichocarpa* (Meux et al, 2011 ; Lallement et al, 2015).

> Une boucle β2-α2 inédite car particulièrement longue (71 résidus) et correspondant à la plus grande insertion relevée lors des alignements de séquence. Cette boucle est impliquée dans la structure quaternaire de la protéine en interagissant avec le monomère voisin.
> Une extension de l'hélice α5 et une petite hélice α5' supplémentaire dans la boucle précédent l'hélice α6.

De manière surprenante, ces deux dernières caractéristiques sont uniquement présentes dans les GST Xi de levure fortement apparentées à *S. cerevisiae*, au sein du clade *Saccharomycetaceae*. Cependant, aucun rôle biologique de ces extensions, s'il en existe un, n'est à ce jour connu. De plus, aucun partenaire protéique d'ScECM4 n'est actuellement connu d'après la base de données répertoriant l'interactome de *Saccharomyces cerevisiae* (http://interactome.dfci.harvard.edu/S_cerevisiae).

Le site actif d'ScECM4 est localisé à l'interface entre les deux domaines du monomère avec la cystéine catalytique (C46) réduite dans la forme native de l'enzyme. Dans le but d'obtenir une structure en complexe avec la GS-ménadione, supposée analogue de substrat (Lam et al, 2012), des expériences de trempage et de co-cristallisation avec cette molécule ont été menées. Quelle que soit la méthode, les jeux collectés donnent le même résultat. Aucune densité électronique correspondant à la GS-ménadione n'est retrouvée au site actif, en revanche, les cartes de densité électronique révèlent la présence de glutathion lié covalemment à la cystéine catalytique. Ce premier constat suggère que la GS-ménadione n'est pas un analogue de substrat mais un véritable substrat puisque l'enzyme se retrouve dans un état glutathionylé correspondant à la première étape de réduction de la GS-ménadione (sans régénération de la cystéine catalytique oxydée en l'absence de glutathion réduit). Dès lors, des mesures d'activité avec la GS-ménadione ont été entreprises au CRM2 et au laboratoire IAM et ont confirmé l'activité de ScECM4 avec ce substrat (K_m de 249 ± 85 µM, k_{cat} de 3,4 ± 0,7 s^{-1}). Le seul changement de conformation majeur relevé entre les deux structures d'ScECM4 apo et glutathionylée est une absence supplémentaire de résidus dans la densité électronique à l'extrémité N-terminale, suggérant une certaine flexibilité lors de la liaison au glutathion. Les expériences de cristallographie associée aux mesures enzymatiques ont donc mis en évidence l'activité réductrice d'ScECM4 envers la GS-ménadione, similaire à l'activité réductrice envers les GS-(hydro)quinones précédemment décrites (Lam et al, 2012).

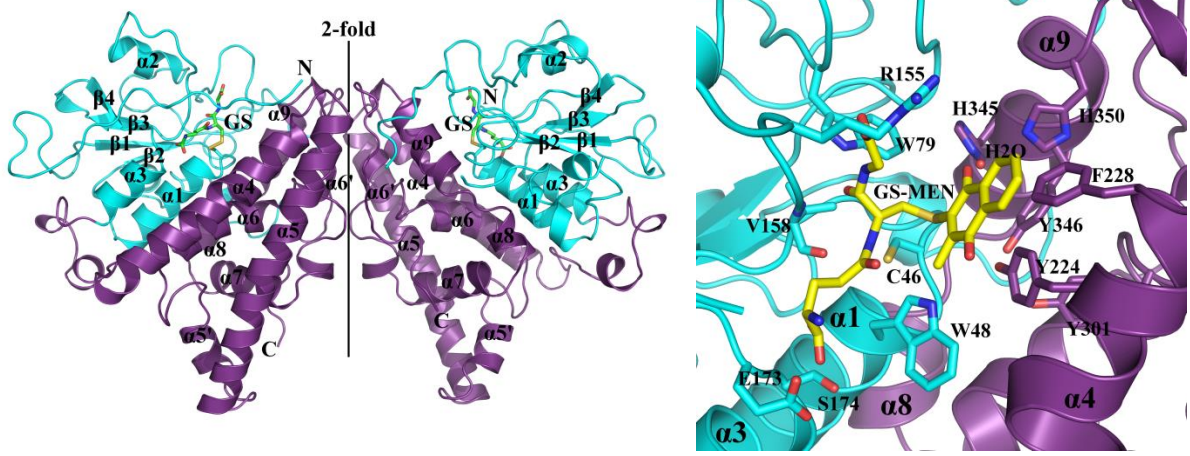


Figure 54 **Structure cristallographique de ScECM4 et modèle de son site actif en complexe avec la GS-ménadione**

Gauche, représentation en mode « rubans » de ScECM4 en complexe avec le glutathion (bâtonnets vert). Le domaine thiorédoxine est coloré en cyan et le domaine tout-hélice est coloré en magenta. Droite, structure du site actif de ScECM4 en complexe avec la GS-ménadione (en jaune), obtenue à partir de la superposition avec la structure de ScYqjG (Green et al, 2012).

Afin d'obtenir un modèle de l'enzyme en complexe avec la GS-ménadione, une superposition de structure a été réalisée : la structure de ScECM4-SG a été superposée à la structure de EcYqjG en complexe avec la GS-ménadione (code PDB 4G0K) précédemment déterminée (Green et al, 2012). La superposition des groupements glutathionyles étant satisfaisante, ce modèle nous a permis de décrire avec confiance les résidus du site actif d'ScECM4 intervenant dans la catalyse. Comme la plupart des GST, ScECM4 a un site actif constitué d'un site de liaison au groupement glutathionyle (site G) et un site hydrophobe (site H), qui dans le cas des GST Xi permet de stabiliser le groupement (hydro)quinone. La cystéine catalytique C46 est localisée à l'extrémité N-terminale de l'hélice α1, à l'interface entre le site G et le site H. A proximité de la cystéine catalytique sont retrouvés les groupements hydroxyles de la triade de tyrosines conservées (Y224, Y301 et Y346 respectivement des hélices α4, α6 et α9). Le site G est constitué de résidus conservés chez les GST et identifiés précédemment chez l'isoforme PcGSTX1. Alors que les résidus du site H sont habituellement peu conservés à la fois entre les classes et dans une même classe de GST, ils le sont relativement bien dans les GST Xi. Ce site H est constitué d'une phénylalanine (F228 de l'hélice α4) stabilisant par interactions π-π le cycle aromatique de la GS-ménadione, et d'un tryptophane (W48 de l'hélice α1) du motif catalytique (^{46}CPWA^{49}) stabilisant le premier carbonyle du cycle quinone. Par ailleurs, une paire d'histidines (H345 et H350 provenant de l'hélice α9) coordonnant une molécule d'eau stabilise le second carbonyle du

cycle quinone. Ces résidus, tout comme la triade de tyrosines précédemment mise en évidence et présente également chez ScECM4, sont conservés chez les GST Xi fongiques, bactériennes et végétales caractérisées à ce jour.

L'ensemble de ces résultats a donné lieu à une publication parue dans le journal *PLOS One*, le 13 octobre 2016.

II.1.b. Publication n°3 : Etude structurale de ScECM4

PLOS ONE

RESEARCH ARTICLE

Crystal Structure of *Saccharomyces cerevisiae* ECM4, a Xi-Class Glutathione Transferase that Reacts with Glutathionyl-(hydro) quinones

Mathieu Schwartz[1,2], Claude Didierjean[1,2]*, Arnaud Hecker[3,4], Jean-Michel Girardet[3,4], Mélanie Morel-Rouhier[3,4], Eric Gelhaye[3,4], Frédérique Favier[1,2]*

1 Université de Lorraine, CRM2, UMR 7036, F-54500 Vandœuvre-lès-Nancy, France, **2** CNRS, CRM2, UMR 7036, F-54500 Vandœuvre-lès-Nancy, France, **3** Université de Lorraine, Interactions Arbres–Microorganismes, UMR1136, F-54500 Vandœuvre-lès-Nancy, France, **4** INRA, Interactions Arbres–Microorganismes, UMR1136, F-54280 Champenoux, France

* Frederique.favier@univ-lorraine.fr (FF); claude.didierjean@univ-lorraine.fr (CD)

CrossMark click for updates

OPEN ACCESS

Citation: Schwartz M, Didierjean C, Hecker A, Girardet J-M, Morel-Rouhier M, Gelhaye E, et al. (2016) Crystal Structure of *Saccharomyces cerevisiae* ECM4, a Xi-Class Glutathione Transferase that Reacts with Glutathionyl-(hydro) quinones. PLoS ONE 11(10): e0164678. doi:10.1371/journal.pone.0164678

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: July 22, 2016

Accepted: September 28, 2016

Published: October 13, 2016

Copyright: © 2016 Schwartz et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Coordinates and structure factors of ScECM4 and ScECM4-SG have been deposited in the Protein Data Bank under accession codes 5LKB and 5LKD, respectively.

Funding: Financial support was received from CNRS/University of Lorraine (PEPS Mirabelle 2016) and the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Glutathionyl-hydroquinone reductases (GHRs) belong to the recently characterized Xi-class of glutathione transferases (GSTXs) according to unique structural properties and are present in all but animal kingdoms. The GHR ScECM4 from the yeast *Saccharomyces cerevisiae* has been studied since 1997 when it was found to be potentially involved in cell-wall biosynthesis. Up to now and in spite of biological studies made on this enzyme, its physiological role remains challenging. The work here reports its crystallographic study. In addition to exhibiting the general GSTX structural features, ScECM4 shows extensions including a huge loop which contributes to the quaternary assembly. These structural extensions are probably specific to *Saccharomycetaceae*. Soaking of ScECM4 crystals with GS-menadione results in a structure where glutathione forms a mixed disulfide bond with the cysteine 46. Solution studies confirm that ScECM4 has reductase activity for GS-menadione in presence of glutathione. Moreover, the high resolution structures allowed us to propose new roles of conserved residues of the active site to assist the cysteine 46 during the catalytic act.

Introduction

Glutathionyl-hydroquinone reductases (GHRs) belong to the glutathione transferase (GST) family of enzymes and catalyze efficiently the reduction of hydrophobic, bulky glutathionyl-hydroquinones [1]. These proteins are clustered within the structural Xi-class [2] well established in bacteria, archaea, plants, protozoa and fungi including the yeast *Saccharomyces cerevisiae* [3–5]. Although the physiological role of the bacterial GHR PcpF from *Sphingobium chlorophenolicum* (ScPcpF) is well established in the pentachlorophenol degradation pathway [6], the role of these enzymes in other prokaryotes and in eukaryotes is still challenging. They could have a function in biochemical pathways that involve quinone reduction [1].

Competing Interests: The authors have declared that no competing interests exist.

One of the most studied GHRs from a biochemical point of view is the protein ECM4 from the yeast *Saccharomyces cerevisiae* (ScECM4). This enzyme was characterized as a thiol-transferase [7]. Site-directed mutagenesis experiments allowed the authors to highlight the residues important for the thiol-transferase activity of the enzyme including the catalytic cysteine 46. Later, it was demonstrated that this cytosolic enzyme could reduce various glutathionyl-hydroquinones (GS-hydroquinones) such as GS-hydroxy-p-hydroquinone, GS-methyl-p-hydroquinone, GS-hydroquinone and GS-menadiol [1]. Molecular biology studies suggested that this enzyme could be involved in cell wall biosynthesis [8]. Moreover, ECM4 gene expression was induced by genotoxic [9] and oxidative agents [10]. Likewise, a cytotoxicity study of menadione in *S. cerevisiae* pointed to ScECM4 as a GST that might be involved in cellular protection [11]. According to sequence similarity, ScECM4 was initially classified in the Omega-class GST and named Gto2 [7]. Later, according to phylogenetic studies and for its ability to reduce GS-hydroquinones, ECM4 was moved to the GHR family [12].

Resolution of the first GHR structure (*Phanerochaete chrysosporium* PcGHR1, pdb code 3PPU) revealed unique structural properties [2]. Its canonical GST fold showed a new dimerization mode by the helical C-terminal domain and a new active site including a triad of tyrosines. These features led to establish a new structural Xi class (GSTX) for these GHRs. Up to now, structures of GHRs from Poplar (PtGHR1) [13] and from the bacteria *Escherichia coli* (EcYqjG) and *Sphingobium chlorophenolicum* (ScPcpF) [14] have been described. In addition, the Xi-class also contained proteins for which GHR activity was not established (PDB entries 3M1G and 4PTS).

In this study, we described the X-ray structures of ScECM4 and glutathionylated ScECM4 (ScECM4-SG). These structures obtained at near atomic resolution constituted the first yeast GHR models. Comparison of ScECM4 with other GSTXs from the PDB revealed additional features, the most striking ones being a huge loop β2-α2, an extension of the helix α5 and the presence of an additional helix α5'. According to sequence alignment, these specificities seemed to be conserved in the genus *Saccharomyces* and closely related genera. Moreover, the structures allowed a fine analysis of the active site and provided complementary information for the comprehension of the catalytic mechanism of this enzyme. Activity measurements associated with structural results showed the reduction of GS-menadione to menadione catalyzed by ScECM4.

Material and Methods

Cloning in bacterial expression vector

The sequence encoding ScECM4 (YKR076W) was amplified by polymerase chain reaction (PCR) from *Saccharomyces cerevisiae* genomic DNA as template and 5'-GGGGGGCATATG TCGAAACAGTGGGCGAG and 5'-GGGGGGCTCGAGTAAAGGACGAATATCTGGCTT G as specific forward and reverse primers, respectively. Amplified sequence was subsequently digested and cloned into pET-26b plasmid (Novagen) between *Nde*I and *Xho*I restriction sites, allowing the fusion of a His-Tag at the C-terminal part of ScECM4.

Expression and purification of recombinant ScECM4

The expression of recombinant ScECM4 was performed at 37°C using *E. coli* Rosetta2 (DE3) pLysS strain (Novagen) transformed with pET-26b::ScECM4 in LB medium supplemented with kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL). When the cell culture reached an OD_{600nm} of 0.7, 0.5% (v/v) of ethanol was added to the culture medium and cells were cooled to 4°C for 3 h. Protein expression was then induced with 0.1 mM isopropyl β-D-1-thio-galactopyranoside (IPTG) and cells further grown for 18 h at 20°C. Cells were then harvested

by centrifugation, resuspended in a 30 mM Tris-HCl buffer, pH 8.0, containing 200 mM NaCl and stored overnight at -20°C. Cell lysis was performed by sonication. The cell extract was centrifuged at 20,000 *g* for 25 min at 4°C to remove cellular debris and aggregated proteins. After the addition of 10 mM imidazole, ScECM4 was purified from the soluble extract by gravity-flow chromatography on a nickel nitrilotriacetate (Ni-NTA) agarose resin according to the manufacturer's recommendations followed by a gel filtration step at 4°C on a Superdex^{TM}200 16/600 column connected to an ÄKTA-Purifier^{TM} (GE Healthcare) and finally stored in a 30 mM Tris-HCl buffer, pH 8.0, containing 200 mM NaCl at -20°C. Concentration of ScECM4 recombinant protein was determined at 280 nm using a theoretical molar absorption coefficient of 84800 $M^{-1}.cm^{-1}$.

Crystallization

Crystallization of ScECM4 was assayed by the microbatch under paraffin oil method at 278 K. ScECM4 (41 mg/mL) crystallized by mixing 1 μL of protein with 1 μL of commercial crystallization solution (Wizard I n°10 (Rigaku), consisting of 20% (w/v) polyethyleneglycol 2000 monomethylether in 0.1 M Tris-HCl buffer, pH 7.0).

ScECM4-SG crystals were equally obtained either by soaking technics (by adding 5 mM GS-menadione to the drop containing crystals of the apo enzyme) or by co-crystallization (by mixing 5 mM GS-menadione to the drop containing the protein and the precipitating agent before crystallization).

Data collection, processing and refinement

ScECM4 and ScECM4-SG crystals were flash-frozen after a quick soaking in the mother liquor containing 20% (v/v) glycerol as a cryoprotectant. X-ray diffraction experiments were carried out at 100 K on the ESRF beamline FIP BM30A (Grenoble, France). ScECM4 and ScECM4-SG crystals diffracted respectively to 1.45 Å and 1.68 Å according to the $CC_{1/2}$ correlation coefficient [15]. Data sets were indexed and integrated with XDS [16] and scaled with XSCALE [16]. The data sets were complete to 1.5 Å and 1.8 Å resolution for ScECM4 and ScECM4-SG, respectively. The structure of ScECM4 was solved by the molecular replacement method using MOLREP of the CCP4 package [17] with the coordinates of PcGHR1 (PDB code 3PPU) [2] as the search model. Structures were refined with PHENIX [18] and built with COOT [19]. After the final refinement/building cycle, atoms that were unobserved in the electron density notably for side chains of surface residues were set with a null occupation and with a B-factor fixed to the respective mean B-factor of the structure. Validation of both structures was performed using MolProbity [20] and the wwPDB Validation server [21, 22].

Coordinates and structure factors of ScECM4 and ScECM4-SG have been deposited in the Protein Data Bank under accession codes 5LKB and 5LKD, respectively. Their PDB Validation reports are provided as supplementary materials (S1 and S2 Files, respectively).

Activity measurements

GS-menadione was prepared according to Lallement *et al.* [13]. Catalytic reduction of GS-menadione to menadione by ScECM4 was evaluated according to two previously described protocols [13]. In a first experiment, 1 μM ScECM4 was added to 500 μL of a 30 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA, 100 μM GS-menadione and various concentrations of GSH (0.05 to 1 mM). Reduction of GS-menadione to menadione was followed by monitoring absorbance from 280 to 480 nm. GS-menadione exhibited a maximum of absorption at 430 nm that disappeared upon reduction within 10 minutes. A control performed without ScECM4 or without glutathione resulted in no change of absorbance.

In a second experiment, menadione released by the reaction was isolated by HPLC and detected spectrophotometrically. ScECM4 was added at 10 μM in a reaction volume of 600 μL of 30 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA, 2 mM GSH and 1 mM GS-menadione. After 10 minutes of reaction at room temperature, 40% (v/v) ethanol was added to the mixture and the tubes were vigorously shaken during 30 s to stop the reaction. After centrifugation (10 min, 14 000 *g*), products were separated by HPLC onto a Gemini C18 column (150 x 3 mm internal diameter, 5-μM particle size, 10-nm porosity; Phenomenex) equilibrated with 30 mM acetic acid buffer, pH 4.16 containing 10% (v/v) acetonitrile. A stepwise gradient of acetonitrile in the acetic acid buffer was performed at flow-rate of 1 mL/min: 10–50% linear gradient for 5 min, then 50–100% linear gradient for 4 min, and finally isocratic elution at 10% (v/v) of acetonitrile for 9 min. Detection was monitored at 250 nm.

Bioinformatics

For the structural alignment, atomic coordinates of five distinct Xi GSTs were selected (PDB codes 3PPU, 4USS, 4G0K, 4FQU, 4PTS). The structural alignment was performed with the module Multiseq [23] integrated in VMD [24] and manually annotated.

For the phylogenetic analysis of yeast ECM4-like proteins, the sequence of ScECM4 (accession number NP_013002) was used in BLASTP on the Joint Genome Institute (JGI) Website to query the fungal genome database [25], or in BLASTP on the National Center for Biotechnology Information (NCBI) website. Multiple sequence alignment was made using Promals3D [26] with the structure of ScECM4 as input. The sequence alignment was manually annotated to highlight the secondary structure element and the putative residues of the active site.

Results and Discussion

ScECM4 crystals soaked with GS-menadione results in a glutathionylated enzyme

Crystallization of ScECM4 occurred quickly in about 4 h. Although polycrystals were frequently observed, several monocrystals suitable for X-ray diffraction experiments have been obtained (Table 1). One of our purposes was to obtain a high resolution structure of the complex between ScECM4 and GS-menadione, which at the moment was thought to be an analog of the substrate GS-menadiol [1]. A previous study reported the complex obtained between the bacterial GHR EcYqjG and GS-menadione, however the electron density corresponding to the menadione moiety was found to be diffuse [14]. To achieve our goal, diffusion experiments were performed in ScECM4 crystals. After a one-day soaking with GS-menadione, crystals appeared unaltered and exhibited a yellowish color due to a probable absorption of GS-menadione. Crystals with the same aspect were obtained by the cocrystallisation method. Three X-ray diffraction data sets were collected (ScECM4, ScECM4 soaked with GS-menadione and ScECM4 cocrystallized with GS-menadione) and analyzed for each experiment.

Structures of ScECM4, ScECM4 soaked with GS-menadione and ScECM4 cocrystallized with GS-menadione were solved at 1.45 Å, 1.68 Å and 1.72 Å, respectively (Fig 1). All crystals were isomorphous and belonged to space group $P2_1$. The asymmetric unit contained two monomers related by a 2-fold axis. ScECM4 exhibited no GSH in the active site but rather glycerol molecules. Surprisingly, both soaked and cocrystallized forms with GS-menadione resulted in a clear electron density that was unambiguously interpreted as glutathione covalently bound to C46 of ScECM4 (ScECM4-SG), while no GS-menadione was detected in the electron density (Fig 1). Since no differences were found between both structures in complex with glutathione, we decided to use the one with the highest resolution, corresponding to the soaking experiment

Table 1. Diffraction and refinement statistics.

		ScECM4	ScECM4-SG
Diffraction data			
	Diffraction source	FIP-BM30A, ESRF	FIP-BM30A, ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97977	0.97977
	Unit-cell parameters		
	a, b, c (Å)	55.51 82.96 80.65	55.44 82.44 80.58
	β (°)	95.1	95.3
	Space group	$P2_1$	$P2_1$
	Resolution range (Å)	46.01–1.45 (1.49–1.45)	47.59–1.68 (1.72–1.68)
	Total No. of reflections	442,863 (19591)	299,262 (16693)
	No. of unique reflections	124,109 (7335)	78,066 (4463)
	Average multiplicity	3.6 (2.7)	3.8 (3.7)
	Mean I/σ(I)	19.7 (1.9)	17.9 (1.9)
	Completeness (%)	96.4 (77.0)	95.0 (74.3)
	R_{merge} (%)	3.5 (51.1)	5.5 (70.6)
	R_{meas} (%)	4.2 (64.3)	6.4 (82.4)
	$CC_{1/2}$ (%)	99.9 (76.1)	99.9 (75.8)
	Wilson B factor (Å^2)	15.1	16.9
Refinement			
	Resolution range (Å)	46.01–1.45 (1.50–1.45)	47.59–1.68 (1.74–1.68)
	R_{work} (%)	16.7 (30.4)	15.9 (27.5)
	R_{free} (%)	19.4 (30.3)	20.1 (32.3)
	No. of protein atoms	11 305	11 284
	No. of waters	785	638
	Ligands	3 GOL	2 GSH
	Average B factor (Å^2)	22.9	25.7
Model quality			
	RMSZ bonds lengths	0.76	0.9
	RMSZ bond angles	0.84	0.89
	RSRZ outliers (%)	1	0
	Ramachandran favored (%)	98.4	98.1
	Ramachandran outliers (%)	0	0
	Molprobity rotamer outliers (%)	0	0.5
	Molprobity clashscore	0.82	1.28
	Molprobity score	0.75	0.85
PDB entry		5LKB	5LKD

$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data. $R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}|| / \sum_{hkl}|F_{obs}|$. 5% of reflections were selected for R_{free} calculation. Ligands PDB codes GOL: glycerol, GSH: glutathione. R.m.s.z.: root mean square Z-score. R.s.r.z.: real space refinement Z-score (indicate the fit between residue and electron density derived from experimental data, according to the wwPDB validation server). The molprobity clashscore is the number of serious clashes per 1000 atoms. The molprobity score is a log-weighted combination of the clashscore, percentage Ramachandran not favored and percentage bad side-chain rotamers. Values in parentheses are for highest resolution shell.

doi:10.1371/journal.pone.0164678.t001

(Table 1). Due to absence or weakness of electron density in some regions, all monomers built in the structures lacked about thirty residues at similar positions, namely the ten first N-terminal ones (fifteen in monomer B of ScECM4-SG), the 6 histidine residues from the C-terminal tag and residues between D83 and L95. For the latter, a possible explanation was the high

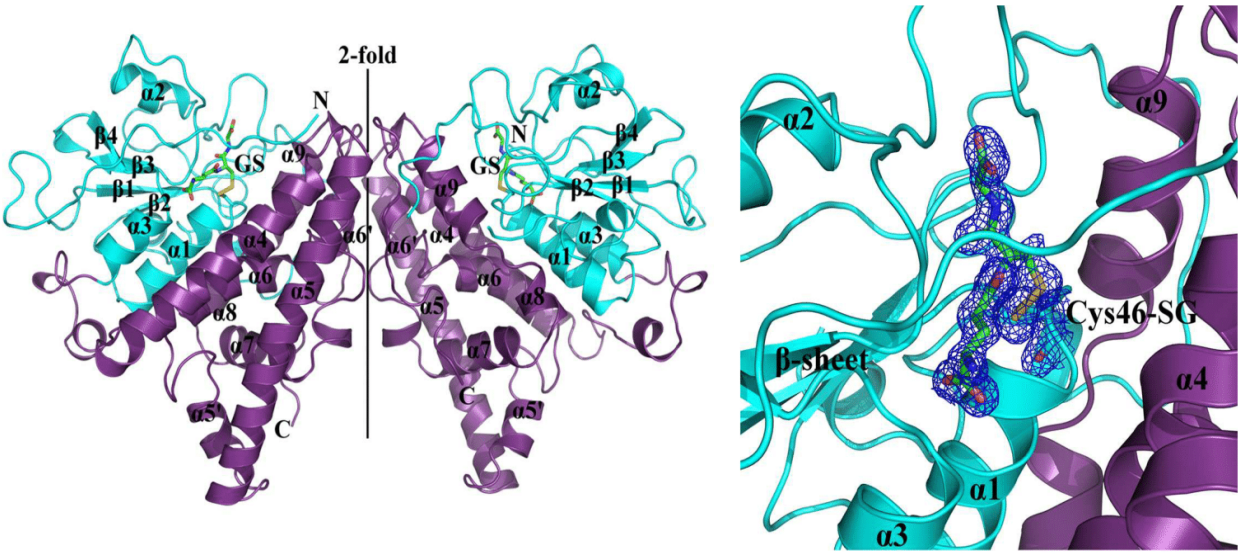


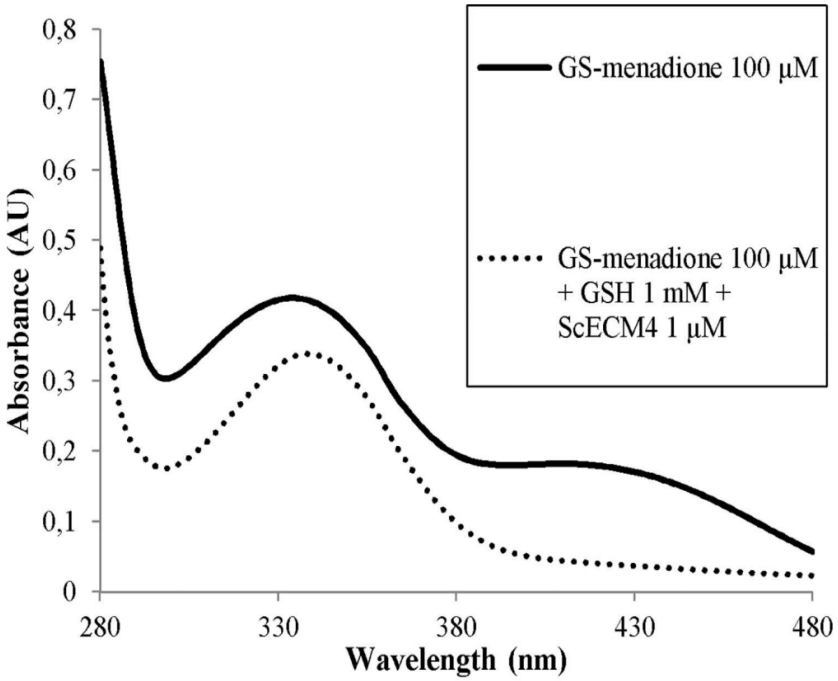
Fig 1. Crystal structure of glutathionylated ScECM4. Left: Cartoon representation of the dimer of glutathionylated ScECM4 (cyan: N-terminal domain, violet: C-terminal domain). The monomers are related by a two-fold axis symbolized as a solid line. Glutathione is represented as green sticks. **Right**: Electron density around the glutathione and cysteine 46 highlighting the disulfide bond. The 2mFo-DFc map contoured at 1.0 sigma is a composite omit map calculated by Phenix [18].

doi:10.1371/journal.pone.0164678.g001

amount of charged residues in this region. Residues R106 to S115 were also absent with the exception of monomer A of ScECM4-SG where a poor electron density was observed, leading to high B-factors for these atoms with respect to the overall B-factor.

Since the soaking and the cocrystallization experiments led to a mixed disulfide bond between C46 and glutathione, we investigated whether ScECM4 could have broken up the GS-menadione bond. First, we eliminated the possibility of a direct reaction of ScECM4 with glutathione. Soaking experiments of crystals with reduced (GSH) or oxidized (GSSG) glutathione only did not result in a covalent complex (data not shown). Then, activity of ScECM4 with GS-menadione was assessed either by direct spectrophotometry measurement at 430 nm or by HPLC separation where menadione was eluted at retention time of 12 min [13] and GS-menadione, more hydrophilic, at 10 min (Fig 2). A mixture of 1 μM ScECM4 and 100 μM GS-menadione failed to consume GS-menadione. The addition of 1 mM GSH in the mixture allowed a continuous catalytic cycle and resulted in solution fading (*i.e.* disappearance of the yellowish color due to GS-menadione). Assays with decreasing concentrations of GSH (1 to 0.05 mM) revealed a concomitant decrease of activity (S1 Fig). The need for an excess of glutathione could be explained by the high K_m value of ScECM4 for GSH (350 μM previously determined with 600 μM of GS-methylhydroquinone [1]). Altogether, these ScECM4 activity measurements stated for reduction of GS-menadione to menadione in presence of glutathione. Earlier, Lam *et al.* [1] observed a good activity of ECM4 with the reduced form of GS-menadione (*i.e.* GS-menadiol). Thus, ScECM4 resembled YqjG from *E. coli*, GHR1 from poplar and GHR1 from *P. chrysosporium* in their ability to catalyze the reduction of glutathionyl-(hydro)quinones [13].

These results explained why enzyme glutathionylation was observed after soaking experiments of ScECM4 crystals with GS-menadione. The protein was likely active in the crystals grown at pH 7.0. On the contrary, Green and colleagues succeeded to obtain crystals of the complex between EcYqjG and GS-menadione [14]. Since EcYqjG displayed optimal activity with GS-menadione at pH 8.0 [13] while the cysteine pKa was estimated to be 7.3 [14], crystals


GS-menadione 100 µM
GS-menadione 100 µM + GSH 1 mM + ScECM4 1 µM
Absorbance (AU)
Wavelength (nm)
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
280
330
380
430
480

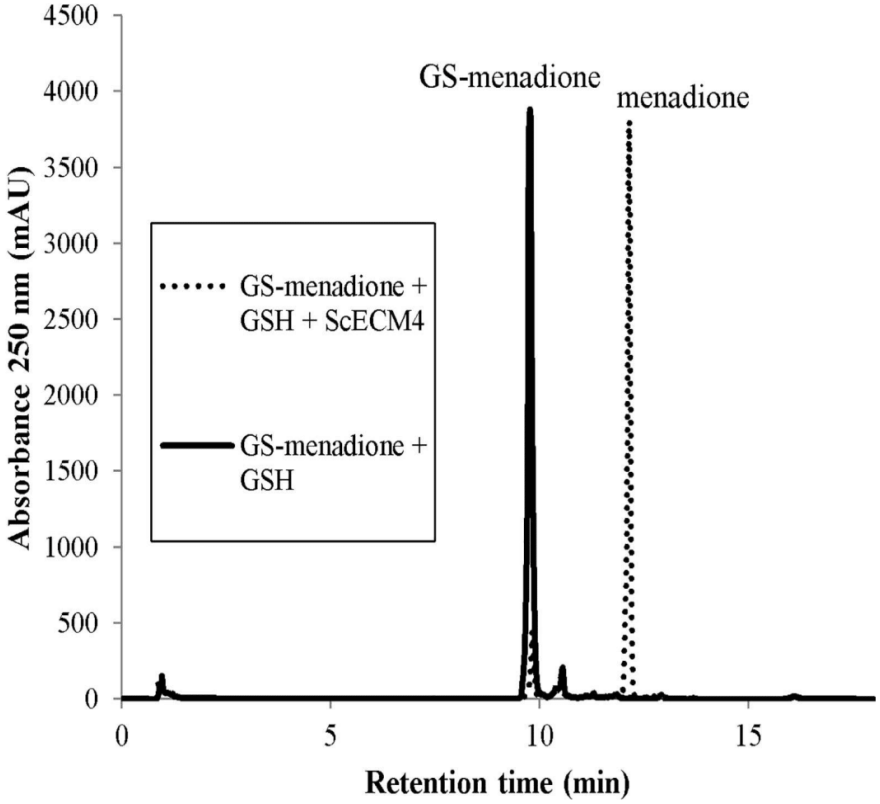

GS-menadione
menadione
GS-menadione + GSH + ScECM4
GS-menadione + GSH
Absorbance 250 nm (mAU)
Retention time (min)
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
5
10
15

Fig 2. Activity of ScECM4 with GS-menadione. Top, UV/visible spectrum of 100 μM GS-menadione in Tris-EDTA buffer, as recorded before addition of 1 μM ScECM4 and 1 mM GSH (solid line), and then after 10 min of reaction with these compounds (dotted line). **Bottom,** HPLC profile at 250 nm after 15 min incubation of ScECM4 with GS-menadione and GSH. GS-menadione shows a typical peak at a retention time of 10 min (solid line) and menadione at 12 min (dotted line).

doi:10.1371/journal.pone.0164678.g002

of complexes were probably obtained thanks to crystallization conditions (pH 6.5, 1.6 M ammonium sulfate) which could have blocked catalysis.

ScECM4 is a GST Xi with features specific to *Saccharomycetaceae*

The structures of ScECM4 and ScECM4-SG were highly similar according to a RMSD of 0.120 Å between the two monomers. As expected, the monomer of ScECM4 exhibited the GST fold and dimerized with the pattern specific to GSTXs (Figs 1 and 3). The N-terminal domain (G11-K196) adopted the thioredoxin fold (β1α1β2α2β3β4α3) and the C-terminal domain (K197-E372) was found to be all helical (α4α5α5'α6α7α8α9). The typical features of GSTXs were also present: a long N-terminal end (37-residues), an extended loop between β2 and α2 (V70-R140) and a unique dimerization mode by the helical domain [2]. However, comparison of ScECM4 with other GSTX structures from the PDB highlighted some distinctive features detailed in the following lines (Table 2, Fig 4).

The visible N-terminal end starting from G11 was found to be partially surrounding the active site in ScECM4 and in monomer A of ScECM4-SG. G11 was hydrogen-bonded with E230, at the C-terminal end of helix α4. The lack of electron density upstream G11 suggested a relative flexibility of the first residues. A similar position of the N-terminal end was described for monomer A of GHR1 from *P. chrysosporium* (PcGHR1) although in this case it was fully observed and covered the active site almost completely (S2A Fig). However, an N-terminal that surrounds the active site could be a particularity of fungal GHRs, since no similar fold was reported for the other GHRs. In the bacterial GHRs EcYqjG and ScPcpF, the N-terminal end adopted rather a characteristic β-hairpin conformation (S2B Fig) [14]. In poplar (PtGHR1), the N-terminal end adopted a more open conformation, although this end was not fully observed in the crystal structure [13].

The loop β2-α2 (residues V70-R140) was the second distinctiveness of ScECM4, and probably the most striking one. With 71 residues, it was far longer than the one of any other GSTXs (~30 residues; S2C Fig). As a comparison, it may be very shorter in non-Xi GSTs, such as in the human GST Omega 2 (6 residues). The usual conformation observed in the available models of GSTXs was a loop composed of three turns globally conserved. However variations were possible, such as a β-hairpin observed instead of the first turn in PtGHR1 (S2D Fig). In ScECM4, the first and the third turns were present. On the contrary, the second one was replaced by a long extension. It first deported the polypeptide chain from the core of the protein to the exterior side of the β-sheet opposite to the helix domain. Then, it returned towards the helix domain before it leaned on the partner monomer of the dimer. The side chain of S115 and the main chain of A117 were hydrogen-bonded to the side chain of E236 in the helix α5 of the other monomer (S2C Fig). To our knowledge, this was the first observation of a loop β2-α2 involved in the dimer interface in GSTs.

The third feature of ScECM4 was observed at helix α5 (A232-Y264). This helix was the largest secondary structure element (~45 Å), extending to the full height of the monomer (Fig 3). Indeed, its C-terminal end was elongated by two spiral turns with respect to the 6 ones usually observed in the other GSTX structures. Moreover, a little helix α5' (R269-F275) absent in previous structures of GHRs was found after helix α5. α5' redirected the polypeptide chain towards the protein core. Interestingly, such a helix was also found at the same place in EcYqjG

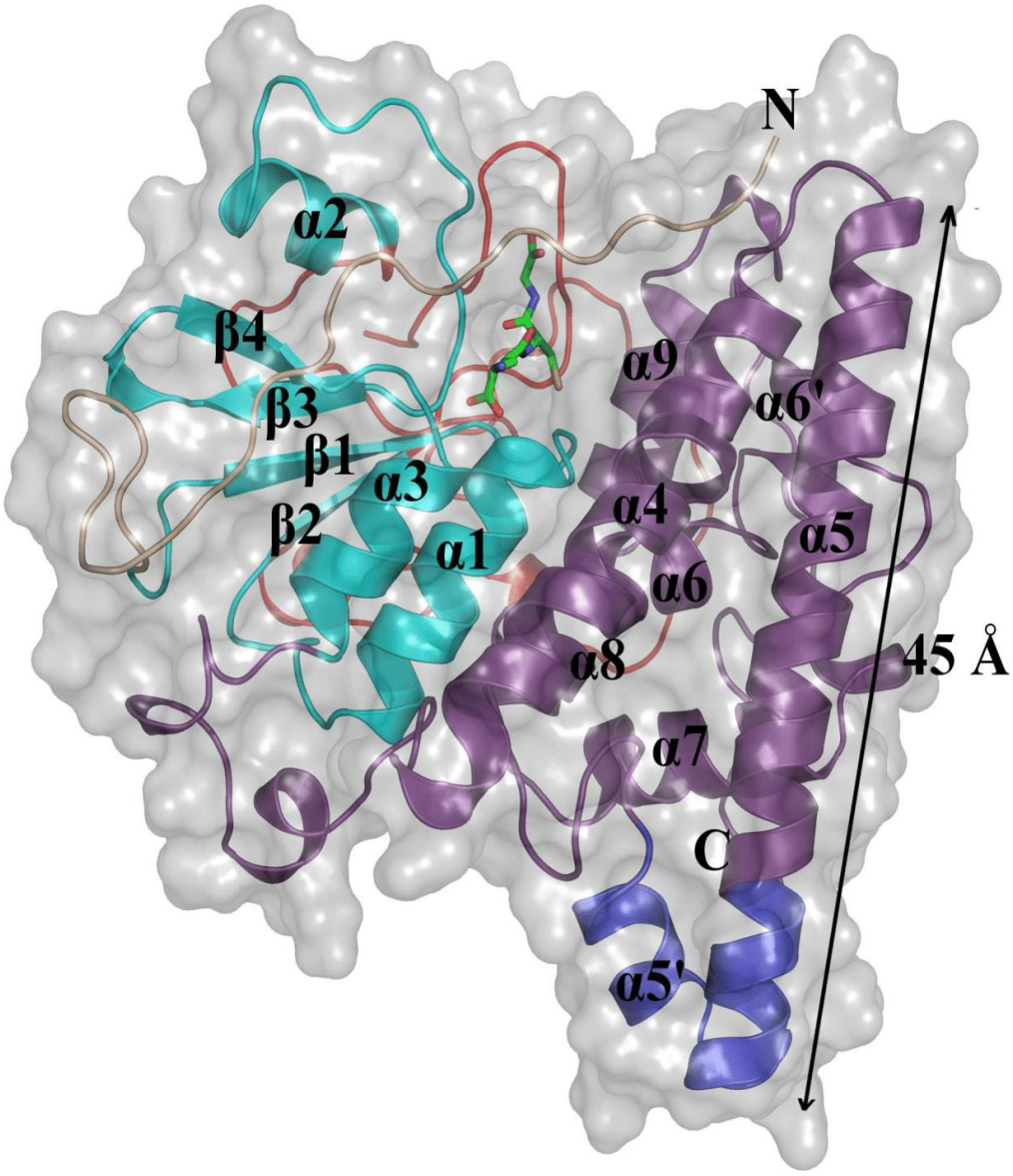


Fig 3. ScECM4 monomer with its specific features. The monomer of ScECM4 is represented as cartoon (N-terminal domain in cyan, C-terminal domain in violet). The N-terminal end is colored in wheat, the loop β2-α2 is colored in red and the extension of helix α5 along with helix α5' are colored in blue.

doi:10.1371/journal.pone.0164678.g003

and a putative GHR from *Gordonia bronchialis* but it was formed by residues situated at the end of the polypeptide sequence.

At last, we focused on D287 in helix α6 because it was shown to be crucial for the thiol-transferase activity, according to previous mutagenesis experiments [7]. This quasi-invariant aspartyl residue has a role in folding and stability [27] of human GST Pi and even in activity of human GST Alpha [28]. As observed in most known GST structures, D287 is a key residue in stabilizing helix α6 via an N-capping box [29]. In ScECM4, D287 was also hydrogen-bonded

Table 2. Characterized GSTX structures in the Protein Data Bank (July 2016).

Available GSTX structures in the Protein Data Bank				
PDB code	**Protein**	**Organism**	**Resolution (Å)**	**Ligands**
5LKB, 5LKD	ScECM4	*Saccharomyces cerevisiae*	1.45	glutathione
3PPU	PcGHR1	*Phanerochaete chrysosporium*	2.3	glutathione
4USS	PtGHR1	*Populus trichocarpa*	2.5	glutathione
4G0I, 4G0L, 4G0K	EcYqjG	*Escherichia coli*	2.05	glutathione, GS-menadione
4FQU	ScPcpF	*Sphingobium chlorophenolicum*	3.0	none
4PTS	putative GHR	*Gordonia bronchialis*	2.83	none
3M1G	putative GHR	*Corynebacterium glutamicum*	2.1	none

doi:10.1371/journal.pone.0164678.t002

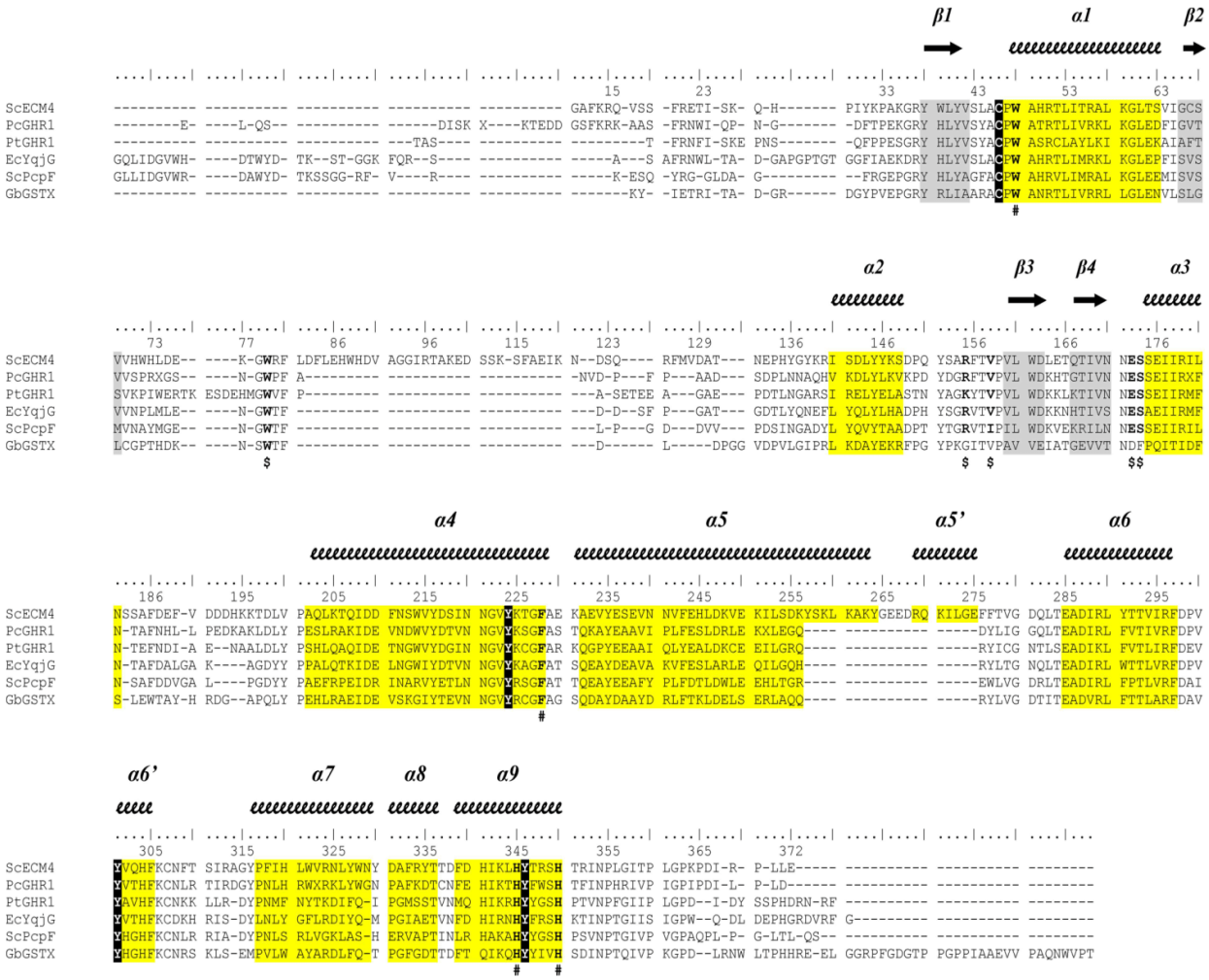


Fig 4. Structural alignment of the GSTXs available in the PDB and corresponding secondary structure elements. Structural alignment of 5 GSTXs available in the Protein Data Bank (PDB codes PcGHR1 3PPU, PtGHR1 4USS, EcYqjG 4G0K, ScPcpF 4FQU, GbGSTX 4PTS) with highlight of the secondary structure elements (helices in yellow, sheets in gray). Catalytic residues (the conserved catalytic cysteine and the triad of tyrosines) are highlighted in black. The symbols **$** indicate the positions of the residues that form interactions with the glutathione and the symbols **#** indicate the positions of the putative (hydro)quinone binding site residues. PtGHR1: GHR from *Populus trichocarpa*, PcGHR1: GHR1 from *Phanerochaete chrysosporium*, EcYqjG: YqjG from *Escherichia coli*, ScPcpF: PcpF from *Sphingobium chlorophenolicum*, GbGSTX: putative GHR from *Gordonia bronchialis*.

doi:10.1371/journal.pone.0164678.g004

with 2 additional partners (T278 and W322), making it likely important for maintaining the tridimensional structures of the enzyme and thus its activity (S2E Fig).

S. cerevisiae harbors three homologous genes encoding for GSTs with a cysteine residue as the potential catalytic residue [7]. They were originally named Gto1, Gto2 (or ECM4) and Gto3 on the basis of their sequence homology with human GST Omega [7]. The present study demonstrated that ScECM4 rather was a GST Xi as most likely Gto1 and Gto3. These isoforms could reduce glutathionyl-(hydro)quinones as ScECM4 and/or could exhibit slightly different substrate spectrum since *S. cerevisiae* contained no GST Omega in its genome. A phylogenetic approach was used in order to evaluate whether the structural extension of ScECM4 (loop β2-α2 and helix α5) were found in other organisms (S3 Fig). Unambiguously, the analysis revealed that the elongations observed in ScECM4 were restricted to *Saccharomyces* and related genera (*Kazachstania*, *Zygosaccharomyces*, *Torulaspora*, and *Kluyveromyces)*. Interestingly a recent multigene phylogenetic analysis of the subphylum *Saccharomycotina* clustered these genera in the clade *Saccharomycetaceae* [30]. GSTX from yeast genera that did not belong to this clade did not show ScECM4 extensions such as *Candida albicans* (S3 Fig). The latter had only one gene encoding for a GSTX that displayed thioltransferase activity similar to ScECM4 [31]. Therefore, the ScECM4 structural extensions would have no influence on yeast GSTX activity.

ScECM4 active site explained the ability of GHRs to reduce GS-(hydro) quinones

GHRs are GSTs and as such classically exhibit a glutathionyl moiety binding site (G site) and a (hydro)quinone-moiety binding site (H site) [2]. Attempts to obtain crystals of complexes of ScECM4 with GS-menadione were not successful despite soaking or co-crystallization experiments. However, analysis at atomic resolution of Fo-Fc and 2Fo-Fc maps derived from these crystals allowed unambiguous identification of a GSH molecule in each monomer of ScECM4 (Fig 5). The sulfur atoms of GSH and C46 were found at a distance of 2.2 Å, indicating that the protein was covalently bound to GSH. C46 was previously identified as the catalytic cysteine of ScECM4 through monitoring the thiol-transferase activity [7]. These observations were consistent with activity observed with GS-menadione and described in the previous dedicated section. The backbones of V158, S174 and the side chains of R15, W79, R155, E173 and S174 were in direct interaction with GSH (see details in Fig 5 legend). Interestingly, the only conformational change between ScECM4 and its glutathionylated form was the side chain of R155 which stabilized the C-terminal moiety of GSH. This overall rigidity of the enzyme was already reported according to fluorescence emission experiments after mutation of the catalytic cysteine [7]. On the contrary, local conformational changes were described for the bacterial homolog EcYqjG upon binding with GSH or GS-menadione [14]. Our atomic resolution models of ScECM4 did not show such flexibility.

The position of GSH in the G-site of ScECM4 orientated the GSH cysteinyl side chain towards the hydroxyl groups of three tyrosines Y224 in α4, Y301 in α6' and Y346 in α9 (Fig 5). These observations were consistent with the identification of a triad of tyrosine residues proposed to be involved in the catalytic mechanism of GHRs according to the bacterial model EcYqjG [14].

The structures of ScECM4 and ScECM4-SG raised questions about binding of GS-(hydro) quinones in the active site, and about the entailed catalytic mechanism. Unexpected absence of the menadione moiety in the active site did not allow precise description of the H site. Consequently, our analysis was based on the sole available structure of a complex between GHR and a glutathionylated compound. Overall superimposition of one monomer of EcYqjG in complex with GS-menadione (PDB code 4g0k, [14]) onto one monomer of ScECM4-SG resulted in

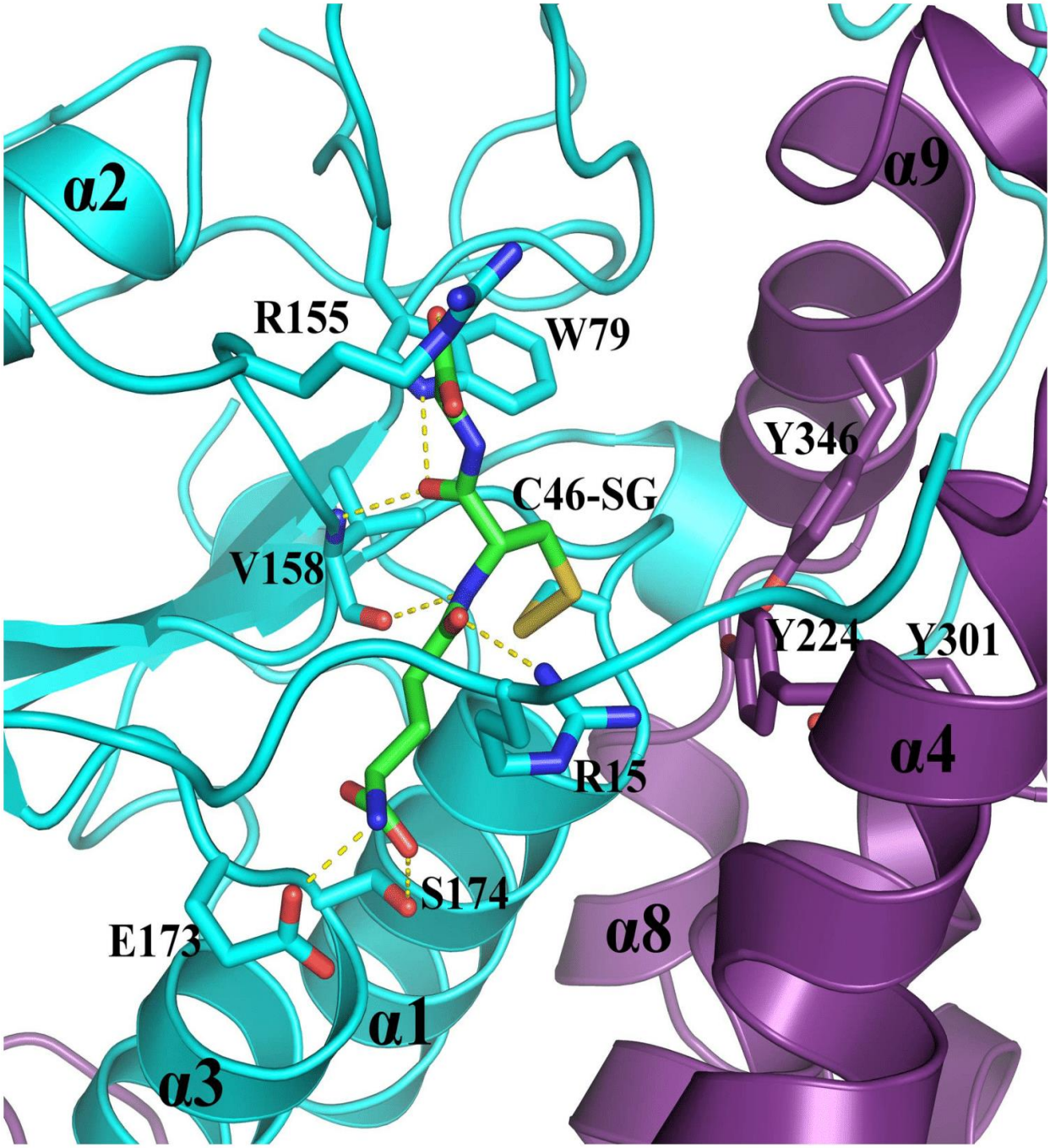


Fig 5. ScECM4 glutathione binding site. The catalytic cysteine C46, the residues involved in polar interactions with glutathione (GS) and the tyrosines Y224, Y301 and Y346 from the triad are represented as sticks. Polar interactions (yellow dashed lines) were observed between: E173 carboxylate group and GSH N-terminal amine group; NH of S174 and carboxylate group of GSH γ-glutamyl residue; S174 hydroxyl group and carboxylate group of GSH γ-glutamyl residue; R15 guanidinium group and carbonyl group of GSH γ-glutamyl residue; W79 indole group and carbonyl group of GSH cysteine; NHs and COs of V158 and of GSH cysteine; R155 guanidinium group and carboxylate group of GSH glycinyl residue.

doi:10.1371/journal.pone.0164678.g005

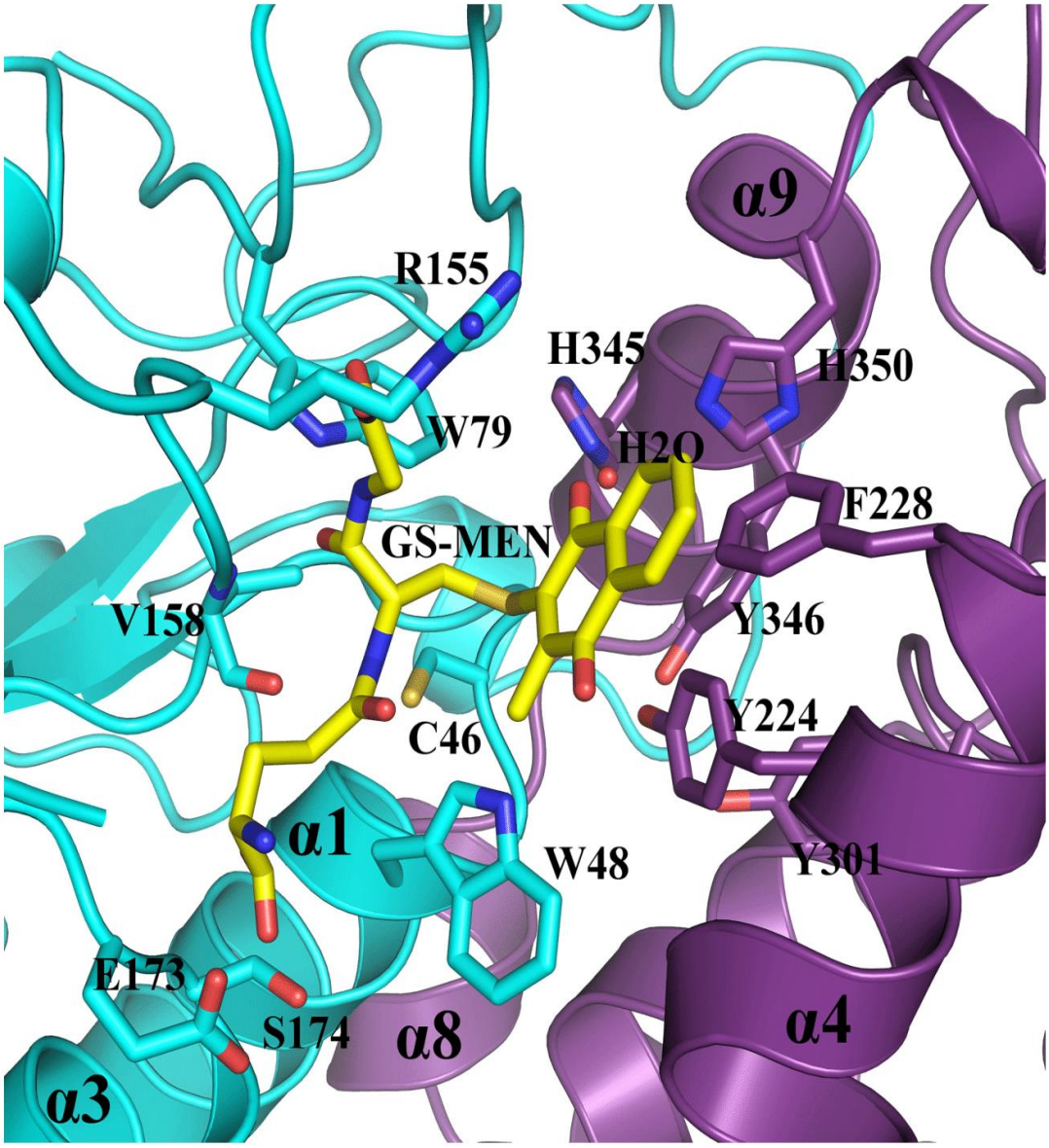


Fig 6. ScECM4 active site with superimposed GS-menadione. Residues of ScECM4 that form putative interactions with GS-menadione are labelled and represented as sticks. In addition to the G-site residues described in Fig 5, the drawing highlights the putative roles of W48, F228, H345, H350 and a water molecule (H_2O) for the binding of the (hydro)quinone moiety, depending on its orientation.

doi:10.1371/journal.pone.0164678.g006

satisfactory superposition of the residues that bordered the active site and of the glutathione (S4 Fig). Hence it provided a putative model of the way GS-menadione could bind in the active site of ScECM4.

Besides local rearrangements at the side chains of residues F14 and R15 to adapt to the presence of the menadione moiety of the substrate, most residues appeared in a suitable position to receive the substrate (Fig 6). The putative roles of five residues, F228, Y224, H345, H350 and W48 emerged from the active site analysis. As already proposed for EcYqjG, F228 could

participate in substrate binding through π-stacking [14], however only in the case of voluminous polyaromatic substrates because of its remote position. The particular orientation of the side chain of Y195 almost parallel to the menadione could offer a similar interaction to smaller substrates. Furthermore, the dipole moment of its phenol group would be placed along the S-C bond of GS-menadione that was to be broken. It could induce an antiparallel dipole at this bond, enhancing reactivity of the GS-menadione sulfur towards the thiolate group of C46. Conserved in all known GHR structures, H345 and H350 from α9 formed hydrogen bonds with a water molecule. Although not modeled in EcYqjG, this water molecule was materialized at the same position by a 3.4σ peak in the corresponding Fo-Fc map. It interacted with the menadione carbonyl group that was positioned ortho to the glutathione. Such an interaction could be maintained with any (hydro)quinone that would orient this oxygen atom towards the glycyl end of GSH. Provided that some GS-(hydro)quinones could orient this oxygen atom towards the glutamyl end of GSH, an alternative interaction would be offered by the nitrogen of the indole group of W48, adequately positioned in all but EcYqjG GHR structures. In the latter, the high concentration of ammonium sulfate used to crystallize the enzyme probably explained the 180° rotation of the tryptophan side chain and therefore the lack of reactivity with GS-menadione. To summarize, H345 and H350, F228, Y224 from the conserved triad of tyrosines, and W48 that belonged to the sequence CPWA strictly conserved in GHRs could play crucial roles in binding of GS-(hydro)quinones, and in assisting C46 during catalysis.

Conclusion

This study focused on the glutathionyl-(hydro)quinone reductase of *S. cerevisiae* named ScECM4. Two high resolution X-ray structures, with and without glutathione covalently bonded to the enzyme, were compared to previously known GHR structures. Besides the features that defined these GSTs as members of the class Xi, ScECM4 possessed a tremendously elongated loop between strand β2 and helix α2 involved in the dimer interface, two more spiral turns in helix α5 that make it 45 Å long and an additional helix α5' with two helical turns. Phylogenetic analysis seemed to confine these specificities to GHRs of the genus saccharomyces and closely related genera. On the contrary, the enzyme possessed in its active site several residues that were conserved through all known GHR structures. They could provide GHRs with their reactivity towards GS-(hydro)quinones as observed both in solution and in the solid state. The present study supports the hypothesis deduced from previous biochemical and cytotoxicity studies that point to ScECM4 as a general protective factor involved in organism resistance to quinone toxicity.

Supporting Information

S1 Fig. Activity of ScECM4 with GS-menadione and various concentrations of GSH. UV/Vis spectrum of 100 μM GS-menadione with 1 μM ScECM4 was recorded after 10 min of reaction with various concentrations of GSH (0 to 1 mM).
(PDF)

S2 Fig. Comparison of ScECM4 structure with known GHRs structures in the PDB. A, fungal GHRs structures with their N-terminal extensions bordering the active sites. N-terminal ends of ScECM4 (N-ter in violet, structure in light gray) and PcGHR1 (N-ter in cyan, structure in medium gray). The arrow indicates the position of the active site. **B,** bacterial GHR (EcYqjG) structure with its N-terminal extension colored in green. The arrow indicates the position of the active site. **C,** loop β2-α2 of ScECM4 (loop in blue, monomers in violet and cyan). Residues involved in polar contacts are labelled and represented as sticks. **D,** loop β2-α2 of PtGHR1

(loop in orange, monomer in salmon). **E,** N-capping residues of helix α6 of ScECM4. Residues involved in polar contacts are labelled and represented as sticks. In all panels, glutathione is represented as yellow sticks.
(PDF)

S3 Fig. Multiple sequence alignment of ECM4-like proteins from yeasts. Sequence of *Candida albicans* CaGTO1 was retrieved from UniprotKB (UniprotKB ID: C4YL44). Sequences from yeasts except *Candida albicans* were retrieved from MycoCosm on the JGI website (JGI protein IDs are as follow: ScECM4, YKR076W; ScGTO1, YGR154C; ScGTO3, YMR251W; *Kazachstania africana* putative GST, KAFR_0H00400; *Zygosaccharomyces rouxii* putative GST ZYRO0B09922g; *Torulaspora delbrueckii* putative GST, TDEL_0B02560; *Kluyveromyces lactis* putative GST, KLLA0F12056g; *Pichia pastoris* putative GST, PAS_chr2-1_03). Sequences were aligned using Promals3D with the structure of ScECM4 as input and were manually annotated. Catalytic residues (the conserved catalytic cysteine and the tyrosines of the triad) are colored in white and highlighted in black. The symbols **$** indicate the positions of the residues that form interactions with the glutathione and the symbols **#** indicate the positions of the putative (hydro)quinone binding site residues. ScECM4 specific extensions are highlighted in yellow.
(PDF)

S4 Fig. Superimposed active sites of EcYqjG and ScECM4. EcYqjG structure in complex with GS-menadione was superimposed onto ScECM4-SG structure in order to identify residues that form putative interactions with the (hydro)quinone moiety. Residues of both active sites, along with the glutathionyl moieties showed good superimposition, thus allowing the identification of ScECM4 putative H site residues. ScECM4-SG is colored in cyan (N-terminal domain) and violet (C-terminal domain) with GSH as green sticks, EcYqjG is colored in tints of gray with GS-menadione as yellow sticks.
(PDF)

S1 File. PDB Validation report for ScECM4
(PDF)

S2 File. PDB Validation report for ScECM4-SG
(PDF)

Acknowledgments

Access to the 'Plateforme de mesures de diffraction X' of the Université de Lorraine is appreciated. We are grateful to the staff members of the BM30A beamline at ESRF (Grenoble, France). Financial support was received from CNRS/University of Lorraine (PEPS Mirabelle 2016) and the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01).

Author Contributions

Conceptualization: CD AH FF.

Data curation: MS CD.

Formal analysis: MS CD FF AH JMG.

Funding acquisition: CD EG MMR.

Investigation: MS CD FF AH JMG.

Methodology: MS CD AH FF.

Project administration: CD EG.

Validation: MS CD FF AH.

Writing – original draft: MS FF CD.

Writing – review & editing: MS JMG FF MMR CD AH EG.

References

1. Lam LK, Zhang Z, Board PG, Xun L. Reduction of benzoquinones to hydroquinones via spontaneous reaction with glutathione and enzymatic reaction by S-glutathionyl-hydroquinone reductases. Biochemistry. 2012; 51(25):5014–21. doi: 10.1021/bi300477z PMID: 22686328.
2. Meux E, Prosper P, Ngadin A, Didierjean C, Morel M, Dumarcay S, et al. Glutathione Transferases of Phanerochaete chrysosporium: S-GLUTATHIONYL-p-HYDROQUINONE REDUCTASE BELONGS TO A NEW STRUCTURAL CLASS. Journal of Biological Chemistry. 2010; 286(11):9162–73. doi: 10.1074/jbc.M110.194548 PMID: 21177852
3. Allocati N, Federici L, Masulli M, Di Ilio C. Distribution of glutathione transferases in Gram-positive bacteria and Archaea. Biochimie. 2012; 94(3):588–96. doi: 10.1016/j.biochi.2011.09.008 PMID: 21945597
4. Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, et al. Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. PLoS biology. 2014; 12(4):e1001843. doi: 10.1371/journal.pbio.1001843 PMID: 24756107; PubMed Central PMCID: PMC3995644.
5. Belchik SM, Xun L. S-glutathionyl-(chloro)hydroquinone reductases: a new class of glutathione transferases functioning as oxidoreductases. Drug Metabolism Reviews. 2011; 43(2):307–16. doi: 10.3109/03602532.2011.552909 PMID: 21425927
6. Huang Y, Xun R, Chen G, Xun L. Maintenance Role of a Glutathionyl-Hydroquinone Lyase (PcpF) in Pentachlorophenol Degradation by Sphingobium chlorophenolicum ATCC 39723. Journal of Bacteriology. 2008; 190(23):7595–600. doi: 10.1128/jb.00489-08 PMID: 18820023
7. Garcerá A, Barreto L, Piedrafita L, Tamarit J, Herrero E. Saccharomyces cerevisiaecells have three Omega class glutathione S-transferases acting as 1-Cys thiol transferases. Biochemical Journal. 2006; 398(2):187–96. doi: 10.1042/bj20060034 PMID: 16709151
8. Lussier M, White AM, Sheraton J, diPaolo T, Treadwell J, Southard SB, et al. Large scale identification of genes involved in cell surface biosynthesis and architecture in Saccharomyces cerevisiae. Genetics. 1997; 147(2):435–50. PMID: WOS:A1997XY91900008.
9. Caba E, Dickinson DA, Warnes GR, Aubrecht J. Differentiating mechanisms of toxicity using global gene expression analysis in Saccharomyces cerevisiae. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2005; 575(1–2):34–46. doi: 10.1016/j.mrfmmm.2005.02.005 PMID: 15878181
10. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of Yeast cells to environmental changes. Molecular Biology of the Cell. 2000; 11:4241–57. PMID: 11102521
11. Castro FA, Mariani D, Panek AD, Eleutherio EC, Pereira MD. Cytotoxicity mechanism of two naphthoquinones (menadione and plumbagin) in Saccharomyces cerevisiae. PLoS One. 2008; 3(12):e3999. doi: 10.1371/journal.pone.0003999 PMID: 19098979; PubMed Central PMCID: PMC2600608.
12. Xun L, Belchik SM, Xun R, Huang Y, Zhou H, Sanchez E, et al. S-Glutathionyl-(chloro)hydroquinone reductases: a novel class of glutathione transferases. The Biochemical journal. 2010; 428(3):419–27. doi: 10.1042/BJ20091863 PMID: 20388120; PubMed Central PMCID: PMC2997670.
13. Lallement P-A, Meux E, Gualberto JM, Dumarcay S, Favier F, Didierjean C, et al. Glutathionyl-hydroquinone reductases from poplar are plastidial proteins that deglutathionylate both reduced and oxidized glutathionylated quinones. FEBS Letters. 2015; 589(1):37–44. doi: 10.1016/j.febslet.2014.11.021 PMID: 25455804
14. Green AR, Hayes RP, Xun L, Kang C. Structural Understanding of the Glutathione-dependent Reduction Mechanism of Glutathionyl-Hydroquinone Reductases. Journal of Biological Chemistry. 2012; 287(43):35838–48. doi: 10.1074/jbc.M112.395541 PMID: 22955277
15. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science. 2012; 336(6084):1030–3. doi: 10.1126/science.1218231 PMID: 22628654; PubMed Central PMCID: PMC3457925.
16. Kabsch W. XDS. Acta Crystallographica Section D-Biological Crystallography. 2010; 66:125–32. doi: 10.1107/s0907444909047337 PMID: WOS:000273820800003.

17. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D-Biological Crystallography. 2011; 67:235–42. doi: 10.1107/s0907444910045749 PMID: WOS:000288532800002.

18. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography. 2010; 66:213–21. doi: 10.1107/s0907444909052925 PMID: WOS:000273820800013.

19. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography. 2004; 60:2126–32. doi: 10.1107/s0907444904019158 PMID: WOS:000225360500002.

20. Davis IW, Murray LW, Richardson JS, Richardson DC. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Research. 2004; 32:W615–W9. doi: 10.1093/nar/gkh398 PMID: WOS:000222273100124.

21. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature Structural Biology. 2003; 10(12):980. doi: 10.1038/nsb1203-980 PMID: 14634627

22. Tickle IJ. Experimental determination of optimal root-mean-square deviations of macromolecular bond lengths and angles from their restrained ideal values. Acta crystallographica Section D, Biological crystallography. 2007; 63(Pt 12):1274–81; author reply 82–3. doi: 10.1107/S0907444907050196 PMID: 18084075.

23. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. Bmc Bioinformatics. 2006; 7. doi: 10.1186/1471-2105-7-382 PMID: WOS:000240941700001.

24. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. Journal of Molecular Graphics. 1996; 14:33–8. PMID: 8744570

25. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research. 2014; 42(D1):D699–D704. doi: 10.1093/nar/gkt1183 PMID: WOS:000331139800103.

26. Pei JM, Grishin NV. PROMALS3D: Multiple Protein Sequence Alignment Enhanced with Evolutionary and Three-Dimensional Structural Information. In: Russell DJ, editor. Multiple Sequence Alignment Methods. Methods in Molecular Biology. 10792014. p. 263–71. doi: 10.1007/978-1-62703-646-7_17 PMID: 24170408

27. Dragani B, Stenberg G, Melino S, Petruzzelli R, Mannervik B, Aceto A. The conserved N-capping Box in the Hydrophobic Core of Glutathione S-Transferase P1-1 Is essential for Refolding. Journal of Biological Chemistry. 1997; 272(41):25518–23. PMID: 9325266

28. Cocco R, Stenberg G, Dragani B, Rossi Principe D, Paludi D, Mannervik B, et al. The folding and stability of human alpha class glutathione transferase A1-1 depend on distinct roles of a conserved N-capping box and hydrophobic staple motif. The Journal of biological chemistry. 2001; 276(34):32177–83. doi: 10.1074/jbc.M104057200 PMID: 11423556.

29. Presta L, Rose G. Helix signals in proteins. Science. 1988; 240(4859):1632–41. doi: 10.1126/science.2837824 PMID: 2837824

30. Kurtzman CP, Robnett CJ. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS yeast research. 2003; 3(4):417–32. doi: 10.1016/s1567-1356(03)00012-6 PMID: WOS:000182999800009.

31. Garcera A, Casas C, Herrero E. Expression of Candida albicans glutathione transferases is induced inside phagocytes and upon diverse environmental stresses. FEMS yeast research. 2010; 10(4):422–31. doi: 10.1111/j.1567-1364.2010.00613.x PMID: 20337724.

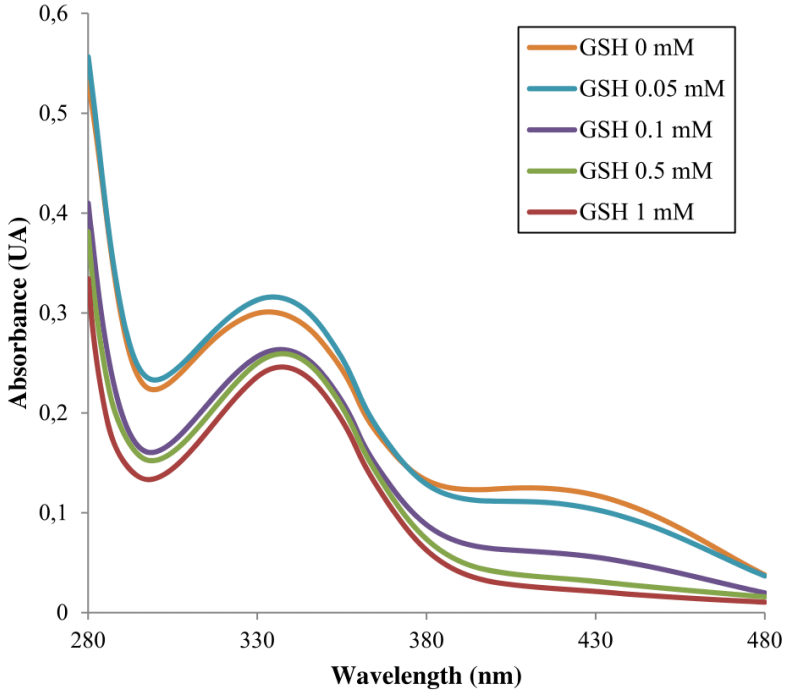
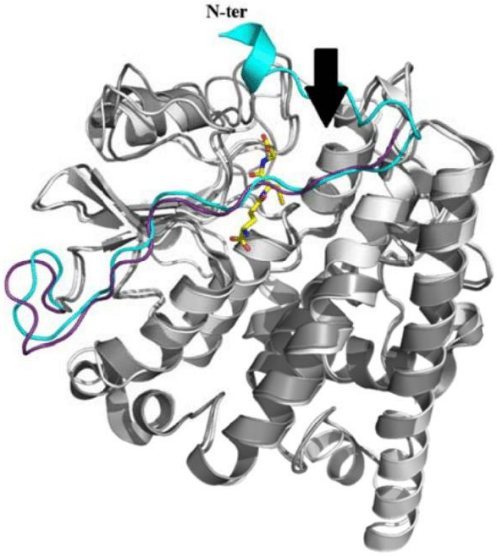
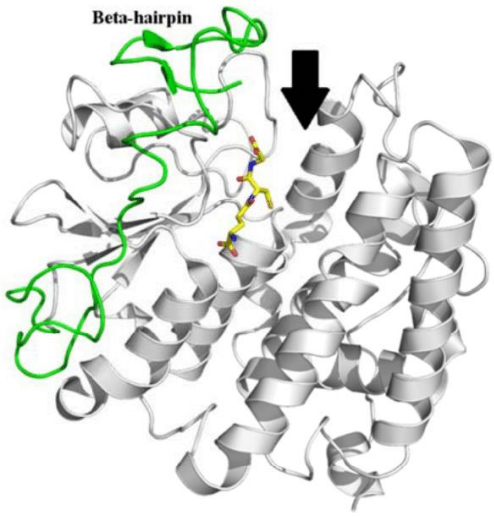


Figure S1. Activity of ScECM4 with GS-menadione and various concentrations of GSH

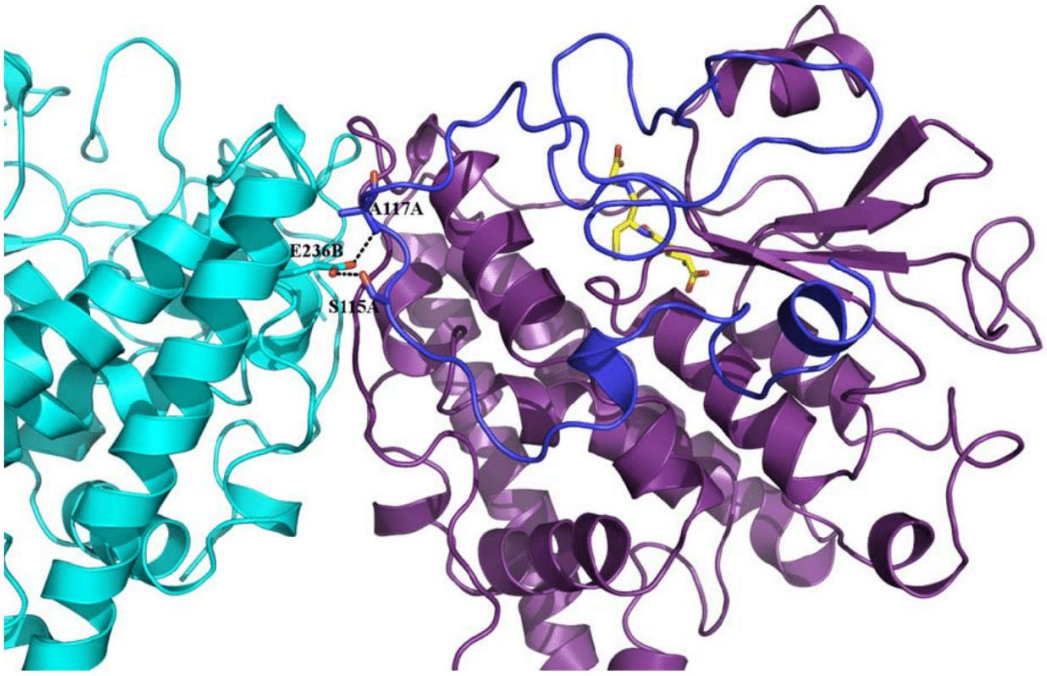
UV/Vis spectrum of 100 µM GS-menadione with 1 µM ScECM4 was recorded after 10 minutes reaction with various concentrations of GSH (0 to 1 mM).



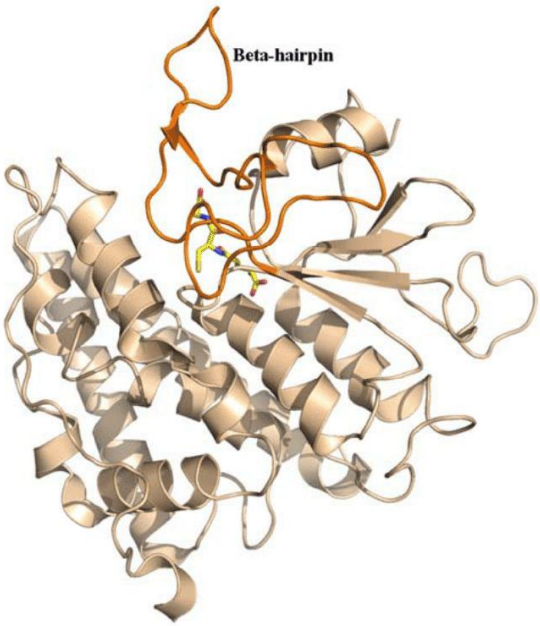
A.



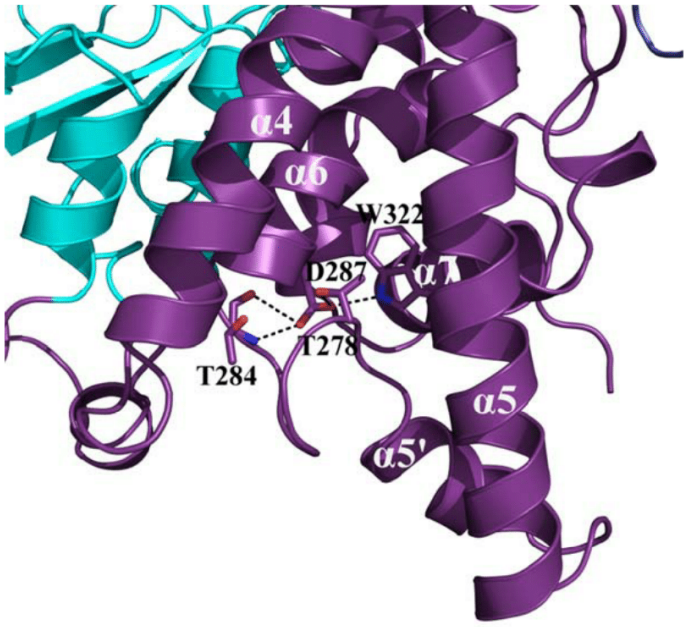
B.



C.



D.



E.

Fig S2. Comparison of ScECM4 structure with known GHRs structures in the PDB

A, fungal GHRs structures with their N-terminal extensions bordering the active sites. N-terminal ends of ScECM4 (N-ter in violet, structure in light gray) and PcGHR1 (N-ter in cyan, structure in medium gray). The arrow indicates the position of the active site. **B,** bacterial GHR (EcYqjG) structure with its N-terminal extension colored in green. The arrow indicates the position of the active site. **C,** loop β2-α2 of ScECM4 (loop in blue, monomers in violet and cyan). Residues involved in polar contacts are labelled and represented as sticks. **D,** loop β2-α2 of PtGHR1 (loop in orange, monomer in salmon). **E,** N-capping residues of helix α6 of ScECM4. Residues involved in polar contacts are labelled and represented as sticks. In all panels, glutathione is represented as yellow sticks.

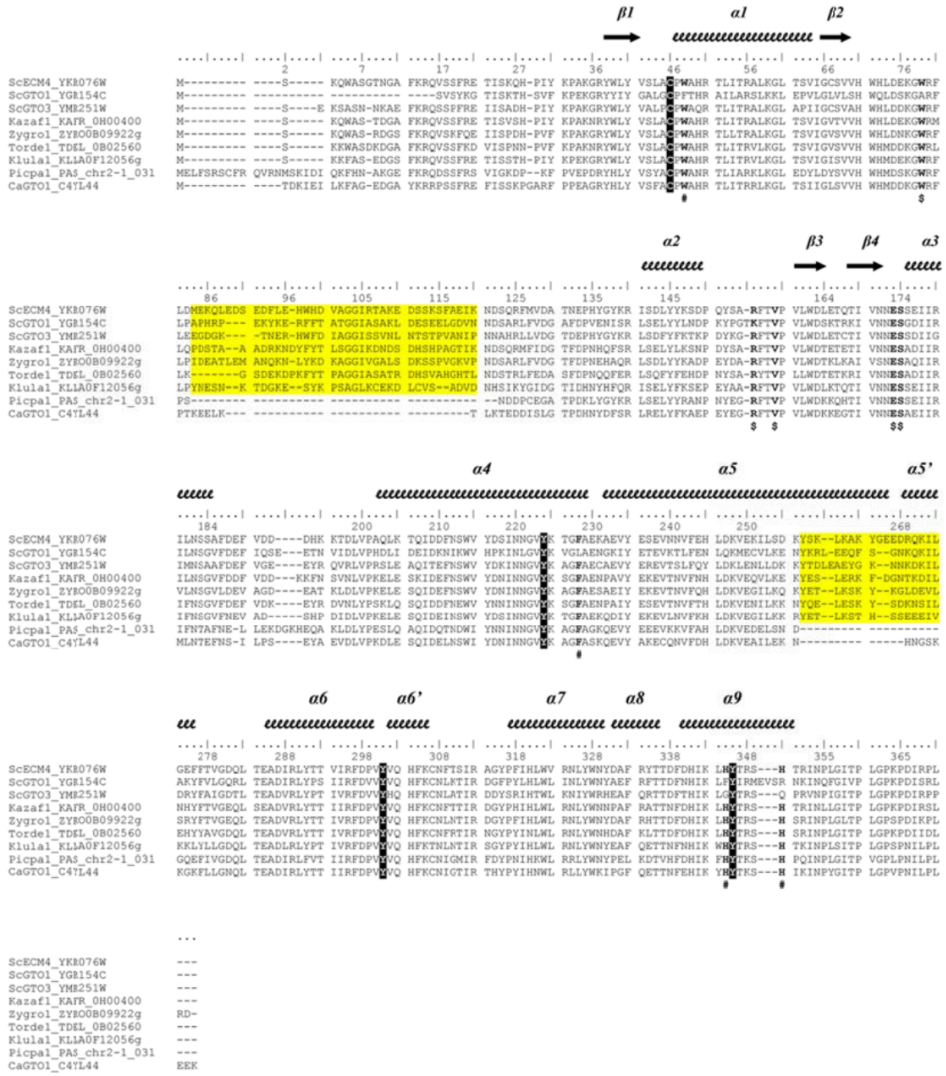


Fig S3. Multiple sequence alignment of ECM4-like proteins from yeasts.

Sequence of *Candida albicans* CaGTO1 was retrieved from UniprotKB (UniprotKB ID: C4YL44). Sequences from yeasts except *Candida albicans* were retrieved from MycoCosm on the JGI website (JGI protein IDs are as follow: ScECM4, YKR076W; ScGTO1, YGR154C; ScGTO3, YMR251W; *Kazachstania africana* putative GST, KAFR_0H00400; *Zygosaccharomyces rouxii* putative GST ZYRO0B09922g; *Torulaspora delbrueckii* putative GST, TDEL_0B02560; *Kluyveromyces lactis* putative GST, KLLA0F12056g; *Pichia pastoris* putative GST, PAS_chr2-1_03). Sequences were aligned using Promals3D with the structure of ScECM4 as input and were manually annotated. Catalytic residues (the conserved catalytic cysteine and the tyrosines of the triad) are colored in white and highlighted in black. The symbols **$** indicate the positions of the residues that form interactions with the glutathione and the symbols **#** indicate the positions of the putative (hydro)quinone binding site residues. ScECM4 specific extensions are highlighted in yellow.

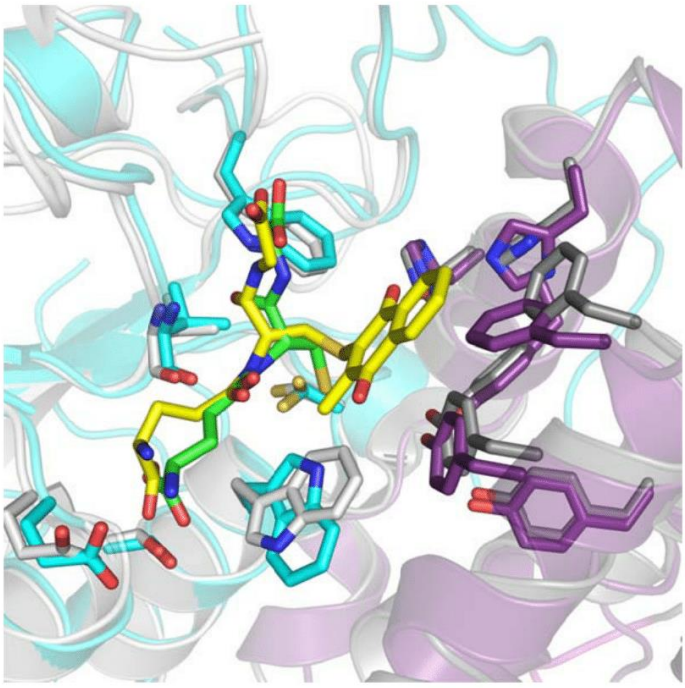


Fig S4. Superimposed active sites of EcYqjG and ScECM4.

EcYqjG structure in complex with GS-menadione was superimposed onto ScECM4-SG structure in order to identify residues that form putative interactions with the (hydro)quinone moiety. Residues of both active sites, along with the glutathionyl moieties showed good superimposition, thus allowing the identification of ScECM4 putative H site residues. ScECM4-SG is colored in cyan (N-terminal domain) and violet (C-terminal domain) with GSH as green sticks, EcYqjG is colored in tints of gray with GS-menadione as yellow sticks.

II.1.c. Etude prédictive des isoformes ScGTO1 et ScGTO3 de *S. cerevisiae*

Les données biochimiques disponibles au moment de cette étude étant centrées principalement sur l'isoforme ScECM4, peu de données étaient disponibles sur les deux autres isoformes codées par le génome de *S. cerevisiae*. ScGTO1 et ScGTO3 présentent respectivement 50 % et 66 % d'identité de séquence avec ScECM4. Ces protéines ont une activité thiol-transférase similaire à ScECM4, mais même si une activité glutathionyl-(hydro)-quinone réductase peut être attendue, celle-ci n'a pas été testée. Compte tenu de leur homologie avec ScECM4, dont la structure était déterminée à haute résolution, une étude structurale prédictive pour ces deux isoformes s'avérait intéressante.

Peu après le dépôt des coordonnées des structures d'ScECM4 dans la PDB, une modélisation comparative automatique des monomères de ScGTO1 et ScGTO3 a été effectuée par le site web ModBase (https://modbase.compbio.ucsf.edu/, Tableau 9). Ce site web contient une base de données de modèles théoriques calculés régulièrement à partir des nouvelles séquences et structures expérimentales déterminées. Les modèles sont générés et leur fiabilité est évaluée par le pipeline ModPipe (Eswar et al, 2006) utilisant entre autres le logiciel MODELLER (Pieper et al, 2004). La forte conservation du repliement tridimensionnel des GST rend la méthode de modélisation comparative particulièrement intéressante pour des isoformes dont les séquences sont proches, ce qui est le cas pour les trois isoformes en question (> 50 % d'identité).

Les modèles de ScGTO1 et ScGTO3 prédits à partir du monomère A d'ScECM4 (Tableau 9) présentent le repliement typique des GST, ainsi que les caractéristiques spécifiques de la classe Xi. Les superpositions des deux modèles avec la structure cristallographique d'ScECM4 révèlent des RMSD de 0,204 Å et 0,137 Å pour ScGTO1 et ScGTO3, respectivement. Aucune déviation majeure par rapport à la structure cristallographique d'ScECM4 n'est retrouvée dans les régions de l'extrémité N-terminale, dans les insertions β2-α2 ainsi que dans les régions entre les hélices α5 et α6. Les positions de la cystéine catalytique, des résidus du site G ainsi que celle de la triade de tyrosines sont conservées (Figure 55). Des variations sont néanmoins retrouvées au site H. Pour ScGTO1, le site H est totalement différent : le tryptophane W48 du motif catalytique ^{46}CPWA^{49} retrouvé chez ScECM4 et les autres GST Xi caractérisées est absent, remplacé par une phénylalanine au sein du motif ^{31}CPFT^{34}. De plus, la paire d'histidines précédemment identifiée dans le site

actif d'ScECM4 (H345 et H350) est remplacée par une phénylalanine et une arginine, respectivement (F328 et R336). Enfin, la phénylalanine F228 supposée interagir avec le substrat via une interaction π-π est remplacée par une leucine (L212). Pour ScGTO3, seule la paire d'histidines H345 et H350 est absente, remplacée par une glycine et une glutamine (G341 et Q346).

Séquence cible	Nom	**Gto1p**	**Gto3p**
	N° Uniprot	C7GQ62	Q04806
	Taille (résidus)	356	366
Structure matrice	Nom	ScECM4	ScECM4
	Code PDB / chaine	5LKB / A	5LKB / A
	Région	16-370	11-369
	Résolution (Å)	1,45	1,45
	Identité de séquence (%)	50	66
Modèle généré	Segment modélisé	1-356	11-365
	Chaines modélisée	1 monomère	1 monomère
	Programme	ModPipe version SVN.r1602	ModPipe version SVN.r1602
Evaluation	Valeur E issue de Psi-Blast	0	0
	Score GA341	1	1
	Score de qualité ModPipe	1,64	1,83
	Z-DOPE	-1,13	-1,48
	Ramachandran favorisé (%)	93,8	95,75
	Score Molprobity	3,21	2,9

Tableau 9 **Données issues de la modélisation comparative par ModPipe des modèles théoriques de ScGTO1 et ScGTO3**
Les modèles ont été générés automatiquement par le pipeline ModPipe et sont disponibles sur la base de données de modèles théoriques ModBase (https://modbase.compbio.ucsf.edu/).

La conservation des résidus du site G, de la cystéine catalytique et de la triade de tyrosines est en cohérence avec l'activité thiol-transférase faisant intervenir le glutathion. Le K_m pour le glutathion est en effet de l'ordre du millimolaire pour les trois isoformes (Garcera et al, 2006). En revanche, le site H typique des GHR mis en évidence chez ScECM4 présente des variations légères pour ScGTO3 (2 résidus différents) et importantes pour ScGTO1 (4 résidus différents). Les résidus en question sont supposés être impliqués dans la stabilisation du groupement quinone. Ce constat interroge sur l'activité GHR potentielle de ScGTO1 et ScGTO3 et une mesure expérimentale serait nécessaire pour confirmer ou infirmer cette activité.

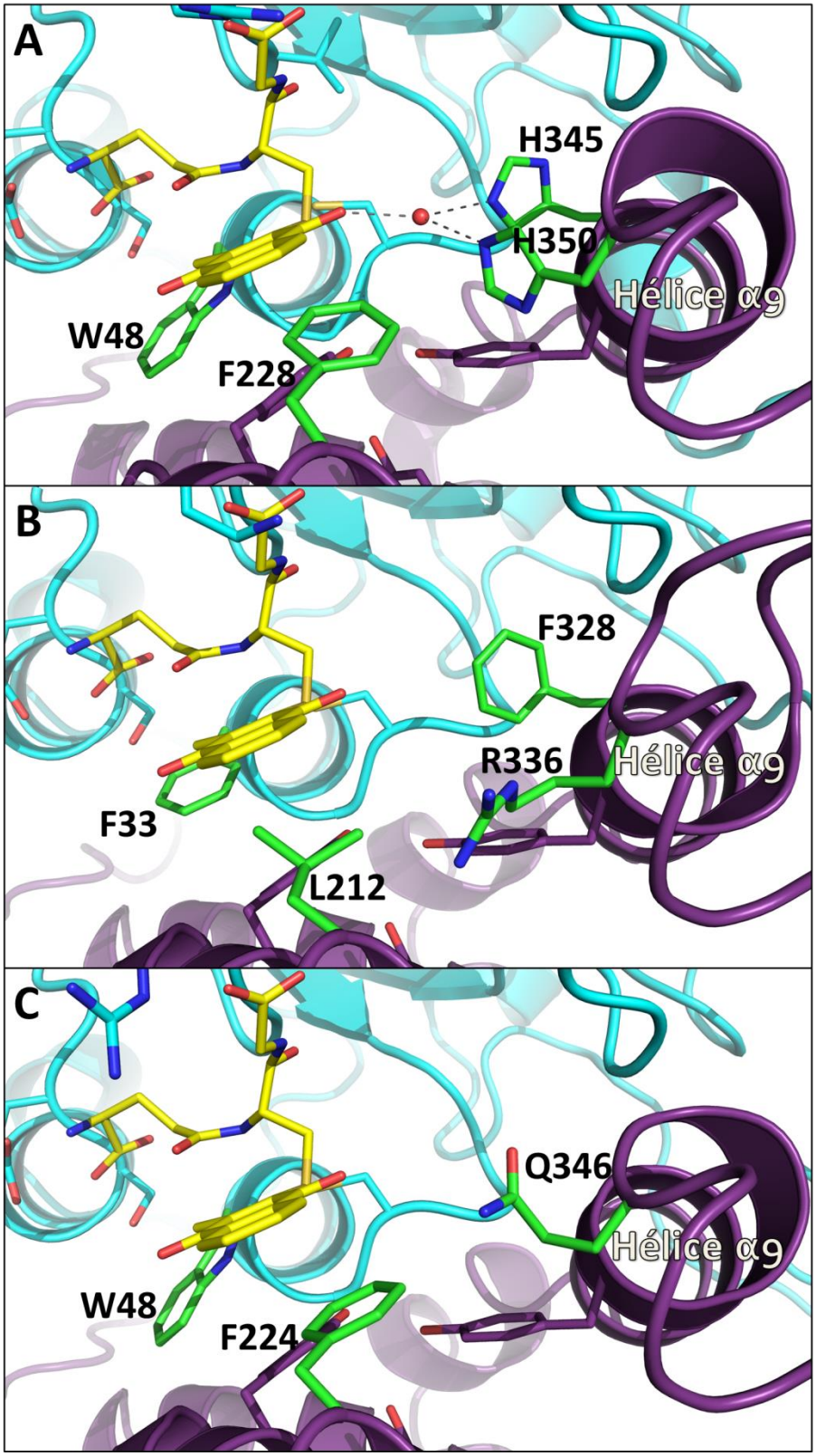


Figure 55 **Comparaison des sites actifs des trois isoformes de GST Xi retrouvées chez *S. cerevisiae***

A. Site actif de ScECM4 (code PDB 5LKD). B. Site actif de ScGTO1 (modèle théorique). C. Site actif de ScGTO3 (modèle théorique). Les résidus putatifs du site H sont colorés en vert et annotés. Le substrat GS-Men, issu de la superposition avec la structure de code PDB 4G0K (Green et al, 2012), est indiqué avec ses contacts potentiels au site H dans la structure de ScECM4.

II.2 Les isoformes GSTX1 et GSTX3 de *Trametes versicolor*

Trametes versicolor possède trois gènes de GSTX dans son génome, codant pour trois protéines appelées TvGSTX1, TvGSTX2 et TvGSTX3. Dans la continuité des travaux effectués sur les GSTX de *S. cerevisiae* et de *P. chrysosporium*, une étude biochimique des isoformes TvGSTX1 et TvGSTX3 a été effectuée au laboratoire IAM. Cette étude a révélé une activité GHR classique envers le substrat GS-ménadione pour les deux isoformes. En revanche, l'isoforme TvGSTX3 est aussi capable de réduire le GS-phenylacétophénone (GS-PAP), qui appartient à la famille de substrats de type « glutathionyl-acétophénone » supposée spécifique des GST de la classe Oméga (Board et Anders, 2007). Une étude structurale complète a dès lors été effectuée sur ces deux isoformes pour tenter d'identifier les facteurs moléculaires à l'origine de l'activité « Oméga-like » de TvGSTX3. Dans une première partie sont résumés la démarche de cette étude ainsi que les principaux éléments de réponse obtenus, puis dans une seconde partie est fourni l'article paru dans le journal *FEBS Letters*.

II.2.a. Résultats de l'étude biochimique et structurale de TvGSTX1 et TvGSTX3

En collaboration avec le laboratoire IAM, les isoformes TvGSTX1 et TvGSTX3 ont été produites et purifiées. Ces deux isoformes présentent 76 % d'identité de séquence (83 % entre TvGSTX1 et TvGSTX2 ; 73 % entre TvGSTX2 et TvGSTX3). La caractérisation enzymatique ayant révélé la dualité d'activité pour l'isoforme TvGSTX3 capable de réduire les substrats GS-ménadione et GS-PAP (Tableau 10), une analyse structurale pour les deux isoformes de TvGSTX a été entreprise.

TvGSTX1 et TvGSTX3 ont été cristallisées et leurs structures ont été résolues par remplacement moléculaire à partir des coordonnées du modèle de PcGSTX1 de *Phanerochaete chrysosporium* (Figure 56). Ces deux isoformes ont des structures proches avec un RMSD de 0,569 Å de dimère à dimère (pour 488 Cα alignés). De plus, les caractéristiques structurales de la classe Xi sont retrouvées pour les deux isoformes, confirmant leur appartenance à cette classe structurale. Toutefois ni l'une ni l'autre ne possède les extensions particulières d'ScECM4 qui semblent spécifiques au clade *Saccharomycetaceae.*

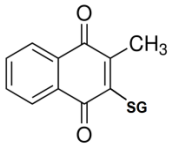
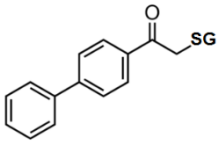
	GS-MEN	GS-PAP
		
K_m (µM)		
TvGSTX1	921 ± 200	ND
TvGSTX3	2509 ± 257	207 ± 45
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	1055 ± 189	4861 ± 1576
k_{cat} (min^{-1})		
TvGSTX1	1966 ± 201	ND
TvGSTX3	5619 ± 361	1895 ± 257
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	205 ± 16	99 ± 26
k_{cat}/K_m ($mM^{-1}.min^{-1}$)		
TvGSTX1	2196 ± 198	ND
TvGSTX3	2240 ± 143	9152 ± 1236
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	195 ± 15	15 ± 5

Tableau 10 **Paramètres cinétiques des TvGSTX déterminés pour les réactions de réduction du GS-MEN et du GS-PAP**

Une comparaison minutieuse des sites actifs a été effectuée (Figure 56). Les deux isoformes présentent les résidus catalytiquement importants pour l'activité GHR des GSTX : la cystéine catalytique (C57 pour TvGSTX1) ainsi que la triade de tyrosines (Y194, Y252 et Y297 pour TvGSTX1). Notre étude a également mis en évidence des différences entre les deux isoformes, retrouvées à proximité de la cystéine catalytique. Alors que TvGSTX1 a un site H classique similaire aux structures de GSTX précédemment résolues, TvGSTX3 présente une large crevasse accessible entre l'hélice α9 et la boucle β2-α2, due à la présence de résidus originaux au niveau de l'hélice α9 (Figure 56). En particulier, la sérine 295 de TvGSTX3 remplace une histidine du site H présente chez TvGSTX1 et dans toutes les autres structures de GSTX déterminées. Rappelons que cette histidine est supposée stabiliser, avec l'aide d'une autre histidine, une molécule d'eau et pourrait ainsi participer à la reconnaissance des GS-(hydro)quinones, comme cela avait été suggéré pour ScECM4.

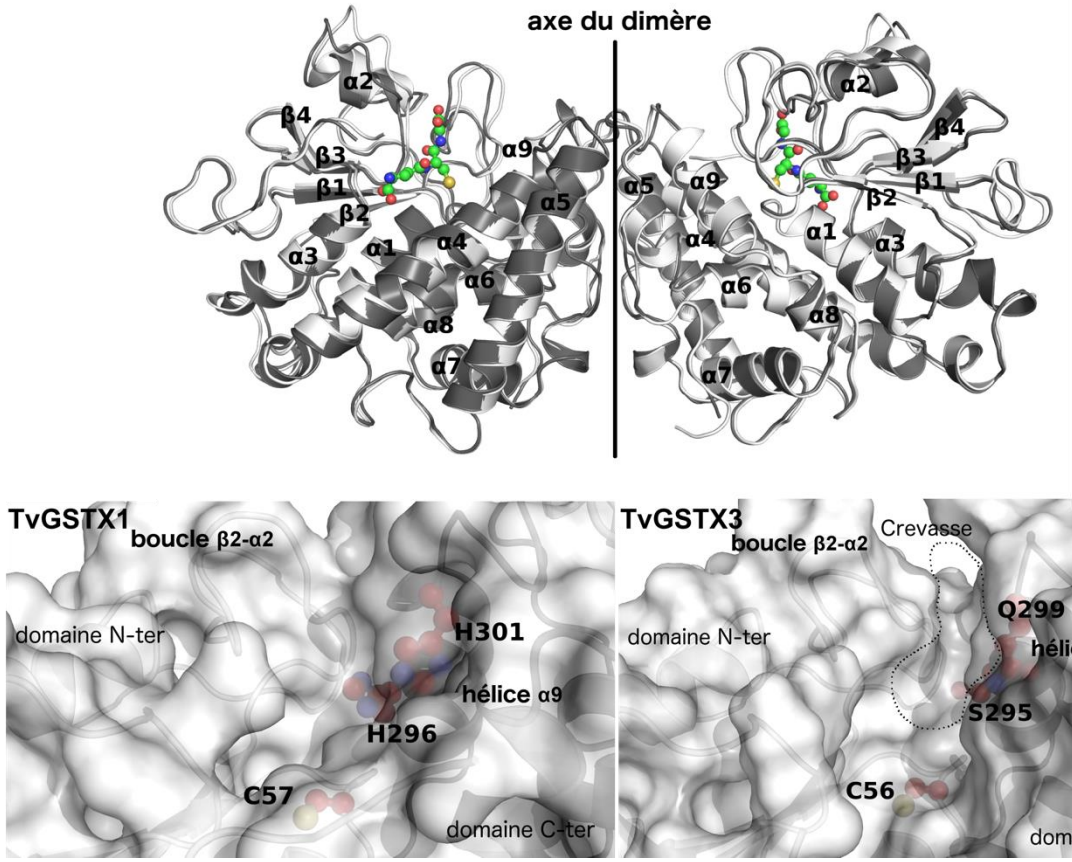


Figure 56 **Structures cristallographiques de TvGSTX1 et TvGSTX3 et comparaison de leurs sites actifs**

Haut, superposition de la structure de TvGSTX1 (représentée en rubans blancs) avec la structure de TvGSTX3 (représentée en rubans noirs). Le glutathion est représenté en bâtonnets verts. Bas, représentation des surfaces des sites actifs de TvGSTX1 (gauche) et TvGSTX3 (droite).

Afin de vérifier si la nature du résidu en position 295 (sérine dans TvGSTX3 et histidine dans TvGSTX1) pouvait avoir un lien avec la capacité de TvGSTX3 de réduire le GS-PAP, la sérine 295 a été mutée en histidine. Le mutant TvGSTX3^{S295H} ne présente qu'une faible activité réductrice avec le GS-PAP, tout en conservant une activité de réduction significative avec la GS-ménadione (Tableau 10). Cela suggère l'importance du résidu présent à cette position 295 pour les propriétés catalytiques de TvGSTX3. Des études de docking ont été menées à partir de la structure de TvGSTX3 et du substrat GS-PAP en contraignant la position du glutathion dans le site G. Cette simulation suggère trois positions possibles pour le groupement phenylacétophénone (Figure 57). Ce groupement PAP pourrait être stabilisé soit dans la nouvelle crevasse identifiée, soit le long des hélices α4 et α9, soit perpendiculairement à l'hélice α4 dans le site de fixation putatif du groupement ménadione. La première position du GS-PAP dans la crevasse hydrophobe pourrait être son mode de fixation réel car il est en

accord avec les données de mutagenèse dirigée. Cette hypothèse n'a pu être vérifiée par l'obtention de la structure cristallographique d'un complexe enzyme/GS-PAP. En effet, les essais de co-cristallisation ou de trempage avec le mutant TvGSTX3^{C56S} de la TvGSTX3 ont échoué.

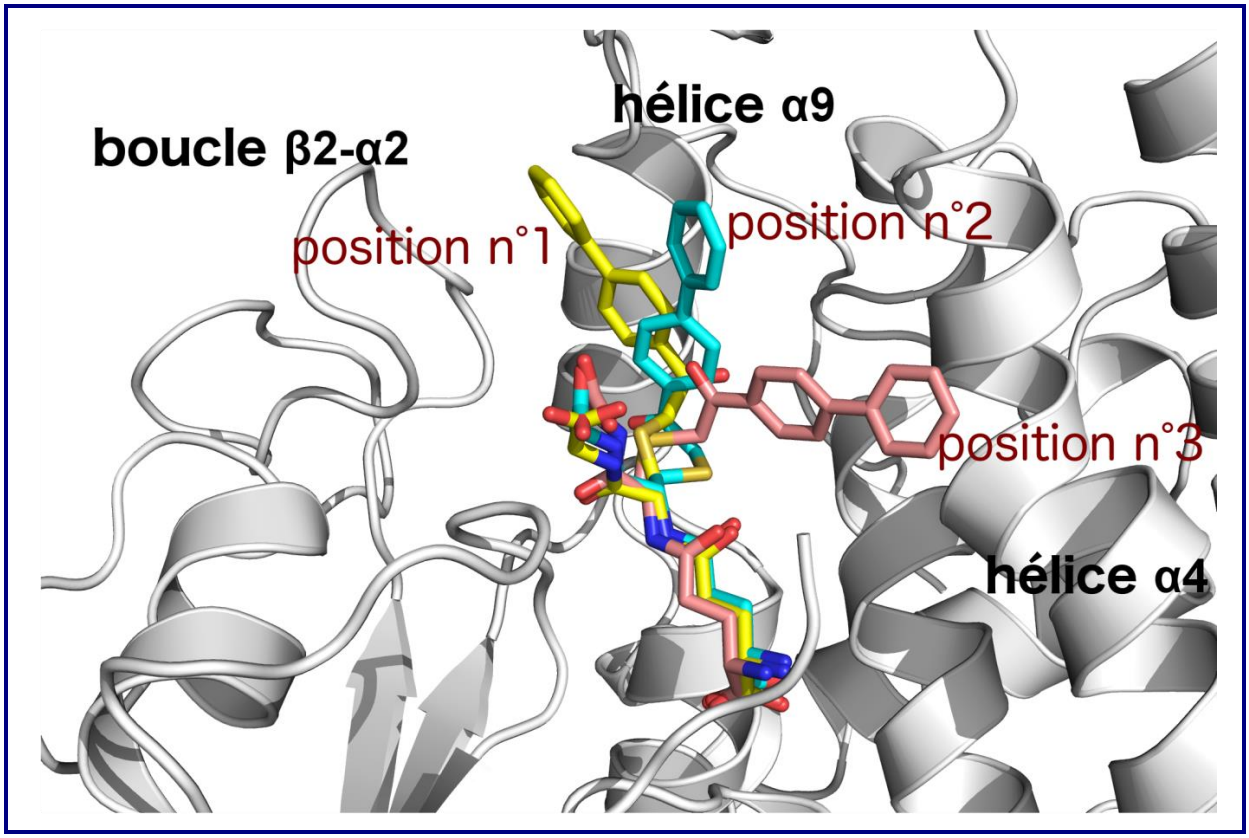


Figure 57 **Positions possibles du GS-PAP prédites par docking dans le site actif de TvGSTX3**
Le docking a été réalisé avec le logiciel GOLD (Verdonk et al, 2003) à partir de la structure cristallographique de TvGSTX3. Trois modes de fixation sont prédits.

Afin de vérifier si d'autres GSTX fongiques pouvaient présenter des variations au niveau de l'hélice α9, une approche phylogénétique a été employée (Figure 58). Cette étude indique que tous les organismes possèdent au moins une isoforme dont les résidus du site actif (cystéine catalytique, triade de tyrosines, résidus des sites G et H) sont identiques à ceux observés dans PcGSTX1, ScECM4, TvGSTX1 et les autres GSTX caractérisées d'un point de vue biochimique. Lorsque plusieurs isoformes sont présentes dans un organisme, des variations sont souvent retrouvées au niveau de l'hélice α9. Notamment la paire d'histidine du site H est fréquemment modifiée (exemple : H296 et H301 dans TvGSTX1, et S295 et Q299 dans TvGSTX3). Des résidus moins encombrants sont observés au niveau de la première histidine (glycine, thréonine ou sérine). Ainsi cette analyse suggère que cette hélice α9 pourrait être la région du site actif la plus propice aux mutations et que ces mutations sont

susceptibles d'entraîner de nouvelles propriétés catalytiques comme c'est le cas dans TvGSTX3.

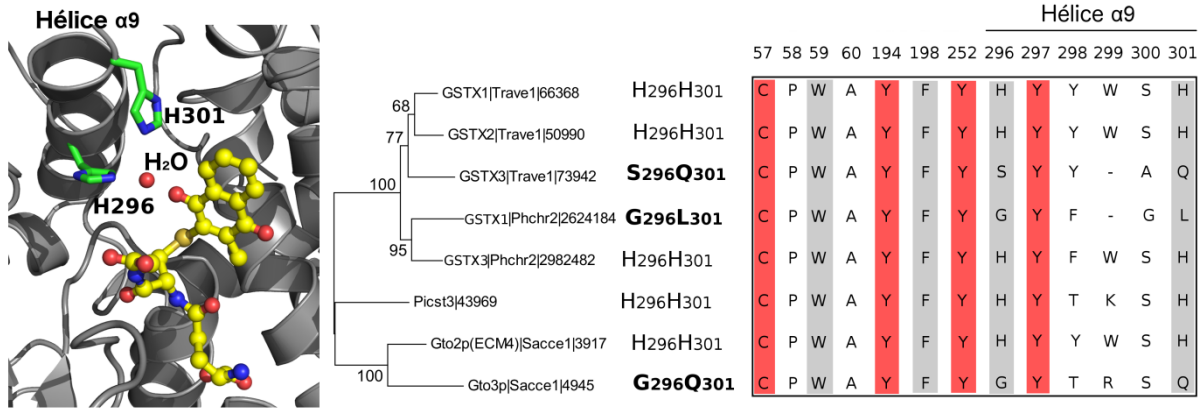


Figure 58 **Analyse phylogénétique des GSTX fongiques et variations prédites au niveau de l'hélice α9**

A gauche, est représenté le site actif de TvGSTX1 en complexe avec la GS-ménadione obtenue à partir de la superposition avec les coordonnées d'EcYqjG (code PDB 4G0K). Au milieu, est montré un arbre phylogénétique des GSTX de *T. versicolor*, *P. chrysosporium*, *P. stipitis*, et *S. cerevisiae.* A droite, est représenté l'alignement des séquences correspondant avec les résidus catalytiques (rouge) ou les résidus du site H (gris).

L'article rapportant les résultats de cette étude et accepté pour publication le 11 août 2018 dans le journal *FEBS Letters* est proposé ci-dessous.

II.2.b. Publication n°4 : Etude structurale et fonctionnelle des isoformes TvGSTX1 et TvGSTX3

FEBS Letters

FEBS PRESS science publishing by scientists

Trametes versicolor glutathione transferase Xi 3, a dual Cys-GST with catalytic specificities of both Xi and Omega classes

Mathieu Schwartz[1], Thomas Perrot[2], Aurélie Deroy[2], Thomas Roret[3], Mélanie Morel-Rouhier[2], Guillermo Mulliert[1], Eric Gelhaye[2], Frédérique Favier[1] and Claude Didierjean[1]

1 CNRS, CRM2, Université de Lorraine, Nancy, France
2 INRA, IAM, Université de Lorraine, Nancy, France
3 CNRS, LBI2M, Sorbonne Université, Roscoff, France

Correspondence
C. Didierjean, CNRS, CRM2, Université de Lorraine, Nancy 54506, France
Fax: +33 383684300
Tel: +33 372745642
E-mail: Claude.didierjean@univ-lorraine.fr

Mathieu Schwartz and Thomas Perrot contributed equally to this work.

(Received 18 May 2018, revised 9 August 2018, accepted 11 August 2018)

doi:10.1002/1873-3468.13224

Edited by Stuart Ferguson

Glutathione transferases (GSTs) from the Xi and Omega classes have a catalytic cysteine residue, which gives them reductase activities. Until now, they have been assigned distinct substrates. While Xi GSTs specifically reduce glutathionyl-(hydro)quinones, Omega GSTs are specialized in the reduction of glutathionyl-acetophenones. Here, we present the biochemical and structural analysis of TvGSTX1 and TvGSTX3 isoforms from the wood-degrading fungus *Trametes versicolor*. TvGSTX1 reduces GS-menadione as expected, while TvGSTX3 reduces both Xi and Omega substrates. An in-depth structural analysis indicates a broader active site for TvGSTX3 due to specific differences in the nature of the residues situated in the C-terminal helix α9. This feature could explain the catalytic duality of TvGSTX3. Based on phylogenetic analysis, we propose that this duality might exist in saprophytic fungi and ascomycetes.

Keywords: crystallography; dual enzyme activity; glutathione transferase Xi; glutathionyl-acetophenone reductase; glutathionyl-hydroquinone reductase

Glutathione transferases (GSTs) are widespread enzymes, which use glutathione (GSH) through several different types of reactions. Most GSTs that harbor a catalytic serine residue (Ser-GSTs) or a catalytic tyrosine residue (Tyr-GSTs) exhibit glutathione-transferase activity, and GSTs with a catalytic cysteine residue (Cys-GSTs) mainly possess glutathione-lyase activity (i.e. deglutathionylation) [1]. Cys-GSTs biochemically characterized so far are subdivided into at least six classes, Beta [2], CLIC [3], dehydroascorbate reductases [4], Omega [5], Lambda [6,7], and Xi [8]. GSTs adopt a conserved fold composed of two domains, namely an N-terminal thioredoxin-like domain, and a C-terminal α-helical domain. The active site, located in a cleft at the interface between domains, consists of a GSH-binding site (G site) and a hydrophobic moiety-binding site (H site). While Ser-GSTs and Tyr-GSTs are mainly dimeric, Cys-GSTs show more structural diversity. Indeed, GSTLs, DHARs, and CLICs are monomeric, GSTBs and GSTOs assemble through the canonical dimer, and Xi class GSTs (GSTXs) form an atypical dimer.

Omega class GST (GSTO) isoforms from human (HsGSTO1) and from the white rot fungus

Abbreviations
CDNB, 1-chloro 2,4-dinitrobenzene; ESRF, European synchrotron radiation facility; GHR, glutathionyl-hydroquinone reductase; GSH, glutathione; GS-MEN, glutathionyl-menadione; GS-PAP, glutathionyl-phenylacetophenone; GST, glutathione transferase; GSTO, Omega class GST; GSTX, Xi class GST; JGI, joint genome institute; PDB, protein data bank; rmsd, root mean square deviation.

Phanerochaete chrysosporium (PcGSTO3 and PcGSTO4) have been shown to catalyze specific reduction of glutathionyl-acetophenones [8–10] (Fig. 1). Board and coworkers proposed a reaction mechanism for human isoform HsGSTO1[9]. The catalytic thiolate is assumed to attack the thioether sulfur atom to give an acetophenone enolate and a mixed disulfide between GSH and C32 of HsGSTO1. The enolate is readily protonated to produce the acetophenone. Then a molecule of GSH reduces the active site disulfide bond and achieves enzyme regeneration. The crystal structure of an enzyme–substrate complex provided no insight into the reaction mechanism. Indeed, the structure of the mutant HsGSTO1^{C32A} in complex with glutathionyl-nitroacetophenone (GS-NAP) revealed the substrate bound in a site too far from the active site to be of catalytic relevance [11].

GSTs of the Xi class constitute a distinct phylogenetic group [12] that has been characterized through multiple studies in bacteria (PcpF from *Sphingobium chlorophenolicum* [13] and YqjG from *Escherichia coli* [14]), fungi (ECM4 from *Saccharomyces cerevisiae* [15,16], PcGSTX1 from *P. chrysosporium* previously named PcGHR1 [8]), and plants (PtGHR1 and PtGHR2 from poplar [17]). Contrary to GSTOs, GSTXs do not reduce glutathionyl-acetophenones but exhibit specific activity in reducing glutathionyl-(hydro)quinones [18,19] (Fig. 1). GSTXs have been initially named glutathionyl-hydroquinone reductases (GHRs) and such activity was not detected for GSTO isoforms [18,20]. Green and coworkers [14] proposed that three tyrosine residues and a cysteine within the active site contribute to the catalytic activity of Xi GSTs. First, the attack of the substrate by the catalytic thiolate together with an acid catalysis by the tyrosine cluster result in a mixed disulfide bond Cys-S-SG and a protonated hydroquinone.

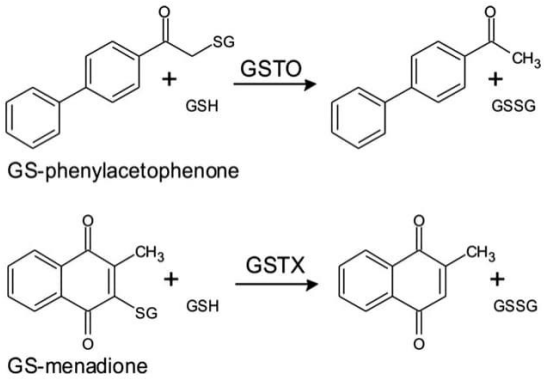


Fig. 1. Reactions catalyzed by GSTOs and GSTXs.

Then, the tyrosine triad acts as a base to activate a second GSH molecule during the regenerative process. Structural comparisons based on the crystal structure of *E. coli* YqjG isoform in complex with glutathionyl-menadione (GS-MEN) enabled a description of the residues involved in the hydroquinone stabilization [15,21]. The putative roles of five residues, F228, Y224, H345, H350, and W48 (ECM4 numbering) emerged from the active site analysis. The tyrosine and phenylalanine residues could stabilize the aromatic moiety of the substrate. The nitrogen atom of the W48 indole group and a water molecule coordinated by both histidine residues could interact with the carbonyl–quinone groups.

In this study, we present the biochemical and structural analysis of the GST isoforms X1 and X3 from the white rot *T. versicolor*. While both possess a GHR function as expected, TvGSTX3 has a surprising additional activity so far thought to be specific to GSTOs [9]. A crystallographic study helped in determining the molecular features that could explain this original profile. Then, a phylogenetic analysis focused on fungal GSTXs investigated the possible distribution of this catalytic diversity.

Materials and methods

Cloning, mutagenesis, expression, and purification

Synthetic genes encoding TvGSTX1 [Joint Genome Institute (JGI) accession number: Tv66368], TvGSTX3 (Tv73942), mutant TvGSTX3^{S295H} were purchased from GeneCust (Luxembourg). A site-directed mutagenesis experiment using two complementary mutagenic primers was performed for obtaining the catalytic mutant TvGSTX3^{C56S}. These genes were cloned into the recombinant plasmid pET26b in order to add a His-tag at the C-term for facilitating their purification (Table S1). Production step was made at 37 °C by using a heterologous system with *E. coli* Rosetta2™ (DE3) pLysS strain (Novagen) in a Luria Bertani Broth medium. The recombinant proteins were purified and eluted by using a column containing a Ni^{2+}-nitriloacetate-agarose resin, as described previously [22]. In addition, the expected molecular masses of the purified proteins were checked by mass spectrometry (electrospray ionization, for more details see the legend of Table S2).

Enzymatic measurements

The reductase activity of the studied TvGSTXs was tested against glutathionyl-phenylacetophenone (GS-PAP) and

FEBS Letters (2018) © 2018 Federation of European Biochemical Societies

GS-MEN, the substrates being synthetized as described previously [23]. The appearance of the expected products (phenylacetophenone and menadione) was followed by reverse phase chromatography. The assays were carried out as developed previously [7] and the concentration ranges were 5–200 μM and 500–2500 μM for GS-PAP and GS-MEN, respectively. Catalytic parameters were determined using varying substrate concentrations at saturating GSH concentration (4 mM) by fitting the Michaelis–Menten equation to the data using GraphPad Prism 5 software (GraphPad Software, La Jolla, CA, USA) (nonlinear regression).

The GST activity of the recombinant proteins toward 1-chloro 2,4-dinitrobenzene (CDNB) was also investigated by using a spectrophotometric method described previously [24].

Crystallization, X-ray data collection, processing, and refinement

Crystallization of TvGSTXs was conducted with the microbatch under oil method at 278 K. TvGSTX1 was crystallized by mixing 1 μL of protein (24 mg·mL^{-1}) with 1 μL of solution consisting in 16% (w/v) polyethylene glycol 4000, 0.1 M pH 8.5 Tris-HCl buffer and 0.2 M magnesium chloride. TvGSTX3 wild-type was crystallized by mixing 2 μL of protein (7.5 mg·mL^{-1}) with 1 μL of solution consisting in 20% (w/v) polyethylene glycol 8000, 0.1 M pH 6.5 sodium cacodylate buffer, and 0.2 M magnesium acetate. TvGSTX3-GSH crystals were obtained by cocrystallization using the same conditions for the native protein and 10 mM GSH. Monocrystals of TvGSTX3^{C56S} were obtained under the same conditions as those for the wild-type protein plus 0.2 μL of 30% dextran sulfate sodium salt. All crystals were cryoprotected by a quick soaking in their mother liquor plus 20% glycerol. X-ray diffraction experiments were carried out on the ESRF beamline FIP BM30A (Grenoble, France). Data sets were indexed and integrated using XDS [25] and scaled with XSCALE or Aimless of the CCP4 suite [26]. The structure of TvGSTX3 was solved by molecular replacement with MOLREP [27] using coordinates of PcGSTX1 (Protein Data Bank code 3PPU) [8] as the search model. Structures were refined with PHENIX [28] and manually built with COOT [29]. All structures were validated with MOLPROBITY [30]. All figures were prepared by using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Docking study

Molecular docking calculations of GS-PAP in TvGSTX3 were performed by using GOLD software (CCDC, Cambridge, UK) [31]. GS-PAP binding site was defined by specifying the sulfur atom of catalytic cysteine C56 (TvGSTX3 numbering) as the center of the cavity. Distance constraints were applied in order to (a) maintain the GSH moiety in its binding site (G site) and (b) to keep plausible stereochemistry of the phenylacetophenone moiety, as defined from crystal structures of similar molecules from the Cambridge Structural Database [32]. All other parameters were kept at their default values.

Phylogenetic analysis

GST Xi sequences were retrieved from the JGI database (https://jgi.doe.gov/) by the Protein Blast method using the sequences of TvGSTX1 and TvGSTX3 as templates. Putative GTSX sequences were curated manually. All sequences were aligned using the MAFFT multiple sequence alignment algorithm [33]. Phylogenetic analysis was performed using the neighbor-Joining method implemented in the MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets [34].

Results and Discussion

TvGSTX1 reduces GS-MEN while TvGSTX3 reduces both GS-MEN and GS-PAP

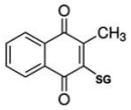
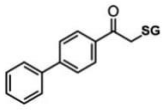
Trametes versicolor possesses three GSTX-coding genes and the corresponding proteins are named TvGSTX1, TvGSTX2, and TvGSTX3. We purified recombinant TvGSTX1 and TvGSTX3, which shared 76% sequence identity (82% between TvGSTX1 and TvGSTX2, and 73% between TvGSTX2 and TvGSTX3). Mass spectrometry experiments conducted with the two His-tagged proteins revealed masses of 37 129 and 36 982 Da (Table S2) respectively, consistent with native proteins in their reduced state (devoid of any covalent adduct, e.g., Cys-S-SG). TvGSTXs were both tested with different substrates to assess their reductase activity (classic GSTX substrate GS-MEN; classic GSTO substrate GS-PAP), and GSH-transferase activity (classic GST substrate CDNB) (Table 1; Fig. S1).

Expectedly, no GSH-transferase activity was detected for any isoform, which is a typical trait for most Cys-GSTs [8,20]. TvGSTX1 and TvGSTX3 both have glutathionyl-hydroquinone reductase activity with the substrate GS-MEN (k_{cat}/K_m around 2200 mM^{-1}·min^{-1}) in accordance with data obtained on GSTXs from various organisms (PcGSTX1 from *P. chrysosporium*, 600 mM^{-1}·min^{-1}; PtGHR1 from poplar, 1000 mM^{-1}·min^{-1}; YqjG from *E. coli*, 1100 mM^{-1}·min^{-1}, Table 2). To complete the enzymatic profile of each isoform, the substrate GS-PAP (usually reduced by GSTOs but not GSTXs [17,20])

was tested. While TvGSTX1 behaves like previously studied GSTXs and is inactive with GS-PAP, TvGSTX3 surprisingly reduces this substrate with a catalytic efficiency of 9152 $mM^{-1}\cdot min^{-1}$, below but comparable to PcGSTOs (Table 2).

In order to further characterize the peculiar isoform TvGSTX3, we studied the mutant in which a serine residue replaced the catalytic cysteine C56 ($TvGSTX3^{C56S}$). As expected, $TvGSTX3^{C56S}$ no longer reduced GS- derivatives consistently with data obtained with $PcGSTX1^{C86S}$ in a previous study [8]. However, $TvGSTX3^{C56S}$ acquired GSH-transferase activity with CDNB (Fig. S1). The same property was observed for $PcGSTO4^{C37S}$ [20] but not for $PcGSTX1^{C86S}$ [8]. As a conclusion to these experiments, TvGSTX3 isoform has the singular catalytic property to behave both like a GHR and a GSTO. It prompted us to determine its crystal structure.

Table 1. Kinetic parameters of wild-type and mutant TvGSTXs. Data are represented as mean ± SD (n = 3). ND, not detected.

	GS-MEN	GS-PAP
		
K_m (μM)		
TvGSTX1	921 ± 200	ND
TvGSTX3	2509 ± 257	207 ± 45
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	1055 ± 189	4861 ± 1576
k_{cat} (min^{-1})		
TvGSTX1	1966 ± 201	ND
TvGSTX3	5619 ± 361	1895 ± 257
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	205 ± 16	99 ± 26
k_{cat}/K_m ($mM^{-1}\cdot min^{-1}$)		
TvGSTX1	2196 ± 198	ND
TvGSTX3	2240 ± 143	9152 ± 1236
TvGSTX3 C56S	ND	ND
TvGSTX3 S295H	195 ± 15	15 ± 5

TvGSTX1 and TvGSTX3 share similar structures with differences in the active sites

A crystallographic study was undertaken on TvGSTX1 and TvGSTX3. We solved four structures: wild-type $TvGSTX1^{WT}$, wild-type $TvGSTX3^{WT}$, wild-type $TvGSTX3^{GSH}$ cocrystallized with GSH, and $TvGSTX3^{C56S}$ mutant bound with dextran sulfate used as a crystallization additive (Table S3). The three structures that correspond to isoform TvGSTX3 are almost identical (largest rmsd of 0.267 Å for 508 Cα aligned between $TvGSTX3^{WT}$ and $TvGSTX3^{C56S}$, Table S4). The only difference lies in the N-terminal end visible in electron density starting from Asp21 (monomer A) in $TvGSTX3^{WT}$ and from Ala28 (monomers A) in $TvGSTX3^{GSH}$ and $TvGSTX3^{C56S}$. This suggests some flexibility upon ligand binding. Superimposed structures of $TvGSTX1^{WT}$ and $TvGSTX3^{WT}$ indicate a high degree of likeness (rmsd of 0.569 Å for 488 Cα). Both isoforms display their highest structural homology with PcGSTX1 (rmsd of 0.546 Å for 563 Cα and sequence identity of 70% between TvGSTX1 and PcGSTX1) (Fig. 2). In brief, TvGSTX monomers adopt the GST fold (N-terminal domain β1α1β2α2β3β4α3, and C-terminal domain α4α5α6α7α8α9) with the specific features of the Xi class, which include a long insertion loop between β2-α2, extended N- and C-terminal coils, and an atypical dimerization mode by the helical domains [8].

Table 2. Reductase activities of characterized GSTXs and GSTOs toward GS-hydroquinones or GS-acetophenones. ND means that reductase activity was not detected. For all GSTXs, glutathionyl-hydroquinone reductase activity was assessed by using GS-MEN substrate. For all GSTOs, glutathionyl-acetophenone reductase activity was assessed by using GS-PAP substrate [20], except for HsGSTO isoforms for which GS-nitroacetophenone was used [9].

			Catalytic efficiency (k_{cat}/K_m, $mM^{-1}\cdot min^{-1}$)		
Class	Organism	Enzyme	GS-hydroquinone	GS-acetophenone	References
GSTX	*Escherichia coli*	EcYqjG	1100	ND	Xun (2010) [18]; Green (2012) [14]
	Saccharomyces cerevisiae	ScECM4	812	ND	Schwartz (2016) [15]
	Phanerochaete chrysosporium	PcGSTX1	600	ND	Meux (2011) [8]
		TvGSTX1	2196	ND	This study
	Trametes versicolor	TvGSTX3	2240	9152	This study
		$TvGSTX3^{S295H}$	195	15	This study
GSTO	*Phanerochaete chrysosporium*	PcGSTO3	ND	2427	Meux (2011) [8]
		PcGSTO4	ND	52 173	Meux (2013) [20]
	Homo sapiens	HsGSTO1-1	ND	780	Board (2007) [9]; Xun (2010) [18]
		HsGSTO2-2	ND	ND	Board (2007) [9]; Xun (2010) [18]

A detailed structural comparison of the active sites was done in an attempt to explain the catalytic profile of TvGSTX3 with respect to the X1 isoform. Both crystallographic structures indicate active sites that share many features in common. The determination of $TvGSTX3^{GSH}$ structure enabled the identification of residues essential for glutathione recognition (G site) (Fig. S2). GSH interacts mainly through polar contacts with the side chain or the main chain of the residues W89, R125, V128, E143, and S144. This binding site is strictly conserved in TvGSTX1, as in GSTX structures determined so far (Fig. S3). The Cys-GSTs and the tyrosine triad are arranged in the same way in both isoforms. Significant differences are observed at the H site. On one hand, the residues that line the putative quinone-binding site (H site) of TvGSTX1 are identical to those suggested from the structures of PcGSTX1[8], ScECM4[15], and EcYqjG [14]. They include residues W58 from helix α1, Y194 and F198 from α4, H296, and H301 from helix α9. On the other hand, both histidine residues are not conserved in TvGSTX3 and helix α9 is one residue shorter at its C-terminal end (Fig. 2). Indeed, the ^{295}SYYA-Q^{299} sequence in TvGSTX3 replaces the ^{296}HYYWSH^{301} sequence of TvGSTX1. These differences create an additional solvent accessible cleft in the active site of TvGSTX3, which is located between helix α9 and loop β2-α2. Its presence is mainly due to the replacement of the first histidine (H296 in TvGSTX1) by the smaller S295. The next section discusses the potential of this cleft as a binding site for the phenylacetophenone moiety of the TvGSTX3 substrate GS-PAP.

In $TvGSTX3^{C56S}$ crystal structure, inspection of electron density maps revealed the presence of a crystallization additive bound to the active site (Fig. S4). This ligand was identified as three units of dextran sulfate (sulfated glycosidic polymer with α-1,6 bonds). Dextran sulfate fully occupies both G and H sites. Sidechains of residues F197, S295, and N301 move to accommodate the ligand that is stabilized mainly by its

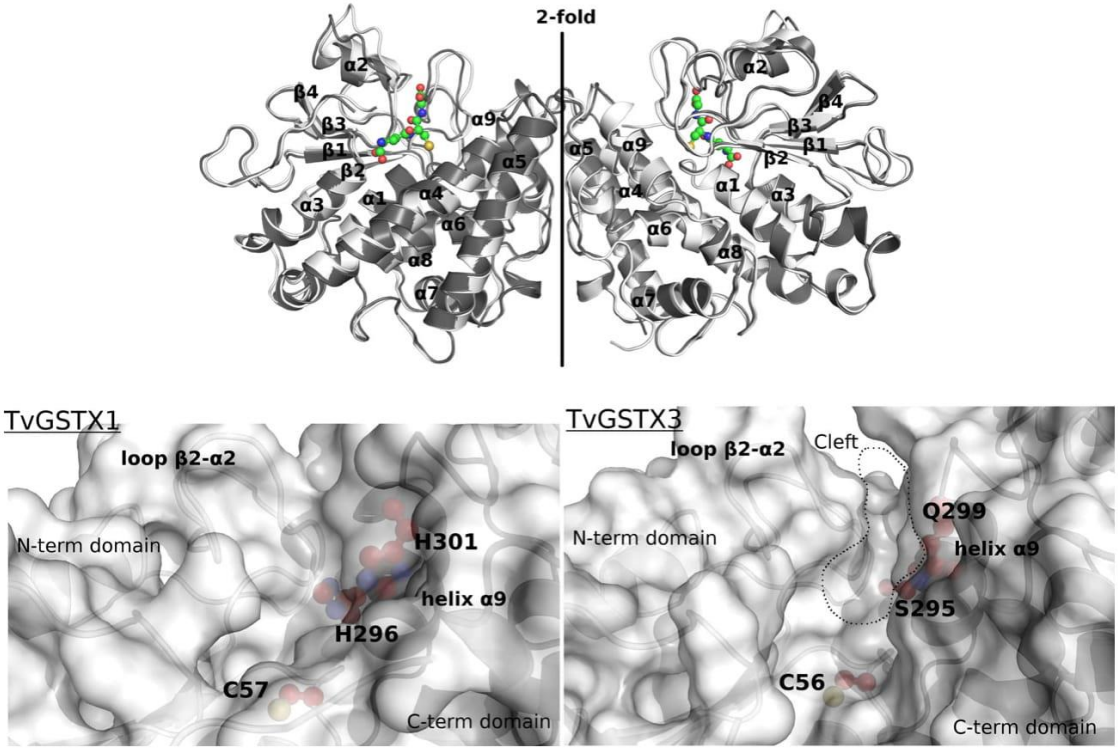


Fig. 2. Crystal structures of TvGSTX1 and TvGSTX3. Top panel: Superimposition of TvGSTX1 dimer (white) and TvGSTX3 dimer (black) complexed with GSH. Secondary structure elements are labeled and GSH is represented as green sticks and spheres. Bottom panels: Active sites of TvGSTX1 (left) and TvGSTX3 (right). Differences present in TvGSTX3 helix α9 create an accessible cleft (circled by a dotted line) near the catalytic cysteine C56. Molecular surfaces are represented in white transparency. Side chains are displayed as red sticks and spheres.

sulfate groups. Five hydrogen bonds are established between the dextran and side chains from R112 (α2), R125 (α2-β3, involved in GSH binding), Y193 (α4, which belongs to the tyrosine cluster), Q299 (α9), and N301 (C-terminal tail). The presence of a polysaccharide in the active site may be of biological relevance. Indeed, it was shown that GSTX ScECM4 would be involved in the biosynthesis of the fungal cell wall [35], which is mainly constituted by carbohydrate polymers (chitin, β-1,3- and β-1,6- glucans)[36].

TvGSTX3 has a new cleft suited for GS-PAP binding

Enzymatic assays showed that TvGSTX1 and TvGSTX3 reduced GS-MEN as expected and unveiled an additional activity for TvGSTX3 in reducing GS-PAP. Crystallographic study revealed an additional accessible cleft in the active site of TvGSTX3 when compared to TvGSTX1 and known GSTX structures. A less bulky serine residue in helix α9 of TvGSTX3 (S295) partly explains the presence of the new cleft. A histidine residue occupies this position in TvGSTX1 (H296). We generated the mutation S295H to evaluate the importance of this position in the catalytic diversity of TvGSTX3. The catalytic efficiency of $TvGSTX3^{S295H}$ with GS-PAP dropped to 15 $mM^{-1}{\cdot}min^{-1}$ (resulting in a 623-fold loss with respect to the wild-type protein) by decreasing both the apparent affinity for GS-PAP (K_m value multiplied by 24) and the turnover number (k_{cat} value divided by 19, Table 1). The GS-MEN activity did not show such a change. $TvGSTX3^{S295H}$ was still active with GS-MEN though with a 10-fold drop in k_{cat}/K_m caused by a decrease of the catalytic rate only. This suggests that the histidine side chain does not cause a decrease in the ability to bind GS-MEN, but that the conformation of the bound substrate is not ideal with respect to the catalytic mechanism. Our results show that a single mutation in TvGSTX3 helix α9 leads to the near loss of the GSTO activity while retaining the GSTX activity.

We performed a molecular docking study to evaluate if the new cleft in the active site of TvGSTX3 and its activity against GS-PAP could be related. We assumed that the glutathionyl moiety of GS-PAP would bind to the G site in a conformation similar to reduced GSH in $TvGSTX3^{GSH}$ and we applied atomic position constraints accordingly during molecular docking. Additionally, distance constraints based on small molecule crystal structures were also applied to the phenyl-acetophenone group to maintain plausible stereochemistry. The constrained docking suggests three regions in the active site, which could be competent for phenylacetophenone moiety binding (Fig. 3). Indeed, we found a mean distance between the sulfur atom of the catalytic cysteine (C56) and the thioether sulfur atom of GS-PAP near 4 Å. The first region corresponds to the new cleft described in the previous section, between helix α9 and the adjacent loop β2-α2. The side chains of Phe84, Tyr193, Ser295, Ala298, and Gln299 ensure a hydrophobic environment and a snug fit for the phenylacetophenone moiety. These residues interact with the ligand dextran sulfate in the $TvGSTX3^{C56S}$ crystal structure (see above). The second region is located along helices α4 and α9 with Phe197 and Gln299 as the residues potentially engaged in stabilization of the GSH adduct. The last region is perpendicular to the helix α4 with the side chains of Trp58, Asn190, and Tyr193 as the closest neighbors of the phenylacetophenone moiety. This third region corresponds to the menadione moiety binding site in the crystal structure of EcYqjG in complex with GS-MEN [14]. Taken together, the mutagenesis study and the docking simulation support the reduction of GS-PAP with its phenylacetophenone-moiety bound in the cleft present in TvGSTX3 but not in TvGSTX1. In the next section, we investigate the differences between these two isoforms from a phylogenetic point of view.

Helix α9 is prone to mutations in the active site of fungal GSTXs

Putative GSTX sequences were retrieved from the 49 fungal genomes that were used in two previous evolutionary studies [37,38]. TvGSTX1 and TvGSTX3 were

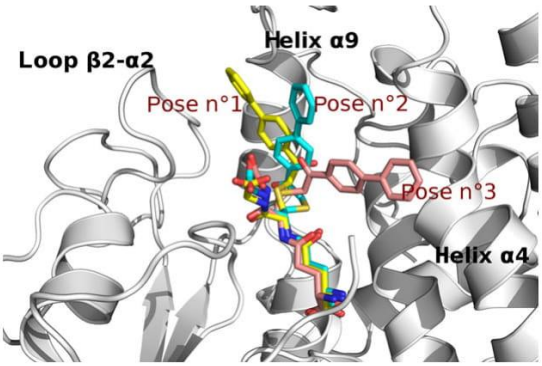


Fig. 3. Docking poses of GS-PAP in TvGSTX3 active site. Representative binding poses (*n* °1 in yellow, *n* °2 in cyan and *n* °3 in salmon) of GS-PAP were predicted by GOLD in TvGSTX3 active site. Region of pose *n* °1 is located in the cleft between loop β2-α2 and helix α9.

FEBS Letters (2018) © 2018 Federation of European Biochemical Societies

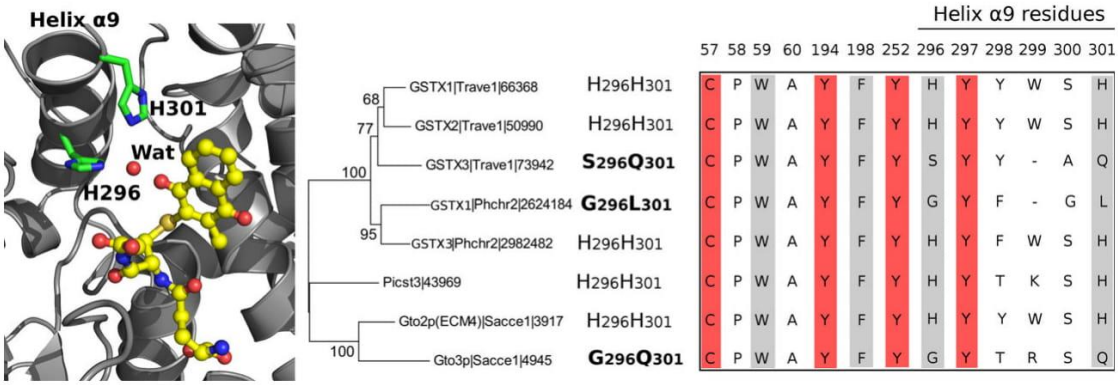


Fig. 4. Neighbor-joining analysis of fungal GSTX revealed variations at helix α9. GS-MEN in TvGSTX1 active site (model based on the superimposition with YqjG complexed with GS-MEN) is shown in left panel. The histidine pair (side chains in green sticks) that coordinates a water molecule in the vicinity of GS-MEN (yellow sticks) is labeled. Neighbor-Joining analysis of representative GSTX sequences is displayed in the middle panel. Sequences for which the histidine pair is absent are in bold and original residues are indicated. Bootstrap values are indicated at nodes. Trave: *Trametes versicolor*, Phchr: *Phanerochaete chrysosporium*, Picst: *Pichia stipitis*, Sacce: *Saccharomyces cerevisiae*. GSTX important residues are indicated in the right panel : catalytic residues are colored red, putative (hydro) quinone-binding site residues are colored gray. Numbering is based on TvGSTX1 sequence.

used as the queries for Protein BLAST searches. This set of sequences was aligned and used to perform a phylogenetic study (Fig. S5). The results suggest the presence of at least one GSTX isoform in each organism. Larger numbers of GSTX isoforms (up to five in *Oidiodendron maius*) are found in two groups of organisms: ascomycetes and saprotrophic basidiomycetes (three isoforms in *T. versicolor*). Extensions of the GST classes Ure2p and GSTFuAs have been previously reported in wood decayers, with more than 10 isoforms for some of them [39,40]. In these classes, the sequences cluster in an organism-specific manner, which could reflect a recent diversification [40,41]. We observed a similar trend for GSTXs.

The aligned sequences reveal the invariance of the catalytic cysteine as well as of the three tyrosine residues that form the catalytic triad (Fig. S3). Most residues that constitute the substrate-binding sites (G and H sites) are also well conserved. Interestingly, variations are found at C-terminal helix α9, which is a part of the H site bottom (Fig. 4). Position 296 (TvGSTX1 numbering) is occupied by a His residue in 61% of the sequences or by a smaller residue in 28% (a Gly residue in most cases, S295 in TvGSTX3). This residue is critical in providing a large cleft in TvGSTX3 potentially related to its catalytic diversity. An additional difference is found at position 301 with a histidine residue in 55% of the sequences (H301 in TvGSTX1 and Q299 in TvGSTX3). The sequences show that residues at positions 296 and 301 (TvGSTX1 numbering) tend to coevolve. Indeed, when the first position is occupied by a non-His residue, so does the second one (81% of the cases). Thus, the helix α9 region is probably prone to mutations as previously reported for solvent-exposed helices [42,43]. This variable region could promote catalytic diversity in the GSTX family, as shown for TvGSTX3 isoform.

Conclusion

In this study, we show that GSTX1 and GSTX3 from *T. versicolor* have the usual GS-(hydro)quinone reductase activity while TvGSTX3 has an additional glutathionyl-acetophenone reductase activity, a feature of GSTOs. Both have the GSTX fold and exhibit structural differences in the active site located at the C-terminal end of helix α9. In TvGSTX3, the presence of smaller residues creates a new cleft that could offer favorable environment for GS-PAP and could explain the substrate diversity observed for this isoform, in accordance with mutagenesis and docking experiments. This peculiar H site makes TvGSTX3 a dual Cys-GST, which shares the substrate specificities of Xi and Omega classes. An examination of fungal GSTX sequences shows that the case of TvGSTX3 should not be unique. This suggests additional functions of GSTXs with respect to their initially assigned GHR activity. Further studies are necessary to clarify whether isoforms from different organisms also exhibit substrate diversity.

Acknowledgements

The authors would like to thank ESRF for beamtime, and the staff of beamline BM30A for data collections. The authors appreciated the access to the 'Plateforme de mesures de diffraction X' of the University of Lorraine with crystal testing.

Funding

This work has been funded by the Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ARBRE grant ANR 11 LABX 0002 01). The authors acknowledge financial support from the 'Impact Biomolecules' project of the 'Lorraine Université d'Excellence' (Investissements d'avenir – ANR).

Author contributions

CD, FF, EG, and MMR conceived and supervised the study. CD, EG, MS, and TP designed experiments. MS, TP, AD, GM, and CD performed experiments. MS, TR, TP, CD, GM, EG, and MMR analyzed data. MS, FF, and CD wrote the manuscript. All authors made manuscript revisions.

Data accessibility

Coordinates and structures factors of TvGSTX1^{WT}, TvGSTX3^{WT}, TvGSTX3^{GSH} and TvGSTX3^{C56S} have been deposited in the Protein Data Bank under accession codes 6GC9, 6GCA, 6GCB and, 6GCC, respectively.

References

1 Lallement PA, Brouwer B, Keech O, Hecker A and Rouhier N (2014) The still mysterious roles of cysteine-containing glutathione transferases in plants. *Front Pharmacol* **5**, 192.

2 Rossjohn J, Polekhina G, Feil SC, Allocati N, Masulli M, Ilio CD and Parker MW (1998) A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure* **6**, 721–734.

3 Cromer BA, Morton CJ, Board PG and Parker MW (2002) From glutathione transferase to pore in a CLIC. *Eur Biophys J* **31**, 356–364.

4 Lallement PA, Roret T, Tsan P, Gualberto JM, Girardet JM, Didierjean C, Rouhier N and Hecker A (2016) Insights into ascorbate regeneration in plants: investigating the redox and structural properties of dehydroascorbate reductases from Populus trichocarpa. *Biochem J* **473**, 717–731.

5 Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermiin LS, Schulte GK, Danley DE, Hoth LR, Griffor MC, Kamath AV *et al.* (2000) Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem* **275**, 24798–24806.

6 Dixon DP and Edwards R (2010) Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. *J Biol Chem* **285**, 36322–36329.

7 Lallement PA, Meux E, Gualberto JM, Prosper P, Didierjean C, Saul F, Haouz A, Rouhier N and Hecker A (2014) Structural and enzymatic insights into Lambda glutathione transferases from Populus trichocarpa, monomeric enzymes constituting an early divergent class specific to terrestrial plants. *Biochem J* **462**, 39–52.

8 Meux E, Prosper P, Ngadin A, Didierjean C, Morel M, Dumarçay S, Lamant T, Jacquot JP, Favier F and Gelhaye E (2011) Glutathione transferases of Phanerochaete chrysosporium: S-glutathionyl-p-hydroquinone reductase belongs to a new structural class. *J Biol Chem* **286**, 9162–9173.

9 Board PG and Anders MW (2007) Glutathione transferase omega 1 catalyzes the reduction of S-(phenacyl)glutathiones to acetophenones. *Chem Res Toxicol* **20**, 149–154.

10 Morel M, Meux E, Mathieu Y, Thuillier A, Chibani K, Harvengt L, Jacquot JP and Gelhaye E (2013) Xenomic networks variability and adaptation traits in wood decaying fungi. *Microb Biotechnol* **6**, 248–263.

11 Brock J, Board PG and Oakley AJ (2013) Structural insights into omega-class glutathione transferases: a snapshot of enzyme reduction and identification of a non-catalytic ligandin site. *PLoS ONE* **8**, e60324.

12 Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, Jagessar K, Kim J, Patskovsky Y, Seidel RD *et al.* (2014) Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere. *PLoS Biol* **12**, e1001843.

13 Huang Y, Xun R, Chen G and Xun L (2008) Maintenance role of a glutathionyl-hydroquinone lyase (PcpF) in pentachlorophenol degradation by Sphingobium chlorophenolicum ATCC 39723. *J Bacteriol* **190**, 7595–7600.

14 Green AR, Hayes RP, Xun L and Kang C (2012) Structural understanding of the glutathione-dependent reduction mechanism of glutathionyl-hydroquinone reductases. *J Biol Chem* **287**, 35838–35848.

15 Schwartz M, Didierjean C, Hecker A, Girardet JM, Morel-Rouhier M, Gelhaye E and Favier F (2016) Crystal structure of saccharomyces cerevisiae ECM4, a Xi-class glutathione transferase that reacts with glutathionyl-(hydro)quinones. *PLoS ONE* **11**, e0164678.

16 Garcera A, Barreto L, Piedrafita L, Tamarit J and Herrero E (2006) Saccharomyces cerevisiae cells have three Omega class glutathione S-transferases acting as 1-Cys thiol transferases. *Biochem J* **398**, 187–196.
17 Lallement PA, Meux E, Gualberto JM, Dumarcay S, Favier F, Didierjean C, Saul F, Haouz A, Morel-Rouhier M, Gelhaye E *et al.* (2015) Glutathionyl-hydroquinone reductases from poplar are plastidial proteins that deglutathionylate both reduced and oxidized glutathionylated quinones. *FEBS Lett* **589**, 37–44.
18 Xun L, Belchik SM, Xun R, Huang Y, Zhou H, Sanchez E, Kang C and Board PG (2010) S-Glutathionyl-(chloro) hydroquinone reductases: a novel class of glutathione transferases. *Biochem J* **428**, 419–427.
19 Lam LK, Zhang Z, Board PG and Xun L (2012) Reduction of benzoquinones to hydroquinones *via* spontaneous reaction with glutathione and enzymatic reaction by S-glutathionyl-hydroquinone reductases. *Biochemistry* **51**, 5014–5021.
20 Meux E, Morel M, Lamant T, Gerardin P, Jacquot JP, Dumarcay S and Gelhaye E (2013) New substrates and activity of Phanerochaete chrysosporium Omega glutathione transferases. *Biochimie* **95**, 336–346.
21 Belchik SM and Xun L (2011) S-glutathionyl-(chloro) hydroquinone reductases: a new class of glutathione transferases functioning as oxidoreductases. *Drug Metab Rev* **43**, 307–316.
22 Deroy A, Saiag F, Kebbi-Benkeder Z, Touahri N, Hecker A, Morel-Rouhier M, Colin F, Dumarcay S, Gérardin P, Gelhaye E *et al.* (2015) The GSTome reflects the chemical environment of white-rot fungi. *PLoS ONE* **10**, e0137083.
23 Nickerson WJ, Falcone G and Strauss G (1963) Studies on quinone-thioethers. I. mechanism of formation and properties of thiodione*. *Biochemistry* **2**, 537–543.
24 Habig WH, Pabst MJ and Jakoby WB (1974) Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. *J Biol Chem* **249**, 7130–7139.
25 Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
26 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
27 Vagin A and Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25.
28 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
29 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
30 Davis IW, Murray LW, Richardson JS and Richardson DC (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615–W619.
31 Verdonk ML, Cole JC, Hartshorn MJ, Murray CW and Taylor RD (2003) Improved protein–ligand docking using GOLD.. *Proteins* **52**, 609–623.
32 Groom CR, Bruno IJ, Lightfoot MP and Ward SC (2016) The Cambridge structural database. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **72**, 171–179.
33 Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
34 Kumar S, Stecher G and Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874. https://doi.org/10.1093/molbev/msw054
35 Lussier M, White AM, Sheraton J, di Paolo T, Treadwell J, Southard SB, Horenstein CI, Chen-Weiner J, Ram AF, Kapteyn JC *et al.* (1997) Large scale identification of genes involved in cell surface biosynthesis and architecture in Saccharomyces cerevisiae. *Genetics* **147**, 435–450.
36 Osumi M (1998) The ultrastructure of yeast: Cell wall structure and formation. *Micron* **29**, 207–233.
37 Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS *et al.* (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719.
38 Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A *et al.* (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* **47**, 410–415.
39 Morel M, Ngadin AA, Droux M, Jacquot JP and Gelhaye E (2009) The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete Phanerochaete chrysosporium. *Cell Mol Life Sci* **66**, 3711–3725.
40 Mathieu Y, Prosper P, Favier F, Harvengt L, Didierjean C, Jacquot JP, Morel-Rouhier M and Gelhaye E (2013) Diversification of fungal specific class a glutathione transferases in saprotrophic fungi. *PLoS ONE* **8**, e80298.
41 Roret T, Thuillier A, Favier F, Gelhaye E, Didierjean C and Morel-Rouhier M (2015) Evolutionary divergence of Ure2pA glutathione transferases in wood degrading fungi. *Fungal Genet Biol* **83**, 103–112.
42 Goldman N, Thorne JL and Jones DT (1998) Assessing the impact of secondary structure and solvent

accessibility on protein evolution. *Genetics* **149**, 445–458.
43 Abrusan G and Marsh JA (2016) Alpha helices are more robust to mutations than beta strands. *PLoS Comput Biol* **12**, e1005242.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Nucleotide sequences of TvGSTX1^{WT} and TvGSTX3 $^{WT,\ C56S\ and\ S295H}$ designed to contain a His-tag at the end of the C-terminal part.

Table S2. Mass spectrometry analysis of recombinant TvGSTXs.

Table S3. Diffraction and refinement statistics.

Table S4. Superimpositions of the TvGSTX structures.

Fig. S1. Glutathione transferase activity toward CDNB for the mutant TvGSTX3^{C56S}.

Fig. S2. Glutathione bound to TvGSTX3 G site.

Fig. S3. Multiple sequence alignment of putative fungal GSTX.

Fig. S4. Dextran sulfate complexed in TvGSTX3 C56S active site.

Fig. S5. Neighbor-Joining analysis of putative GSTX sequences from various fungi.

Supplementary Material

***Trametes versicolor* Glutathione Transferase Xi 3, a dual Cys-GST with catalytic specificities of both Xi and Omega classes**

Mathieu Schwartz[1†], Thomas Perrot[2†], Aurélie Deroy[2], Thomas Roret[3], Mélanie Morel-Rouhier[2], Guillermo Mulliert[1], Eric Gelhaye[2], Frédérique Favier[1] and Claude Didierjean[1]

[1]Université de Lorraine, CNRS, CRM2, Nancy, France

[2]Université de Lorraine, INRA, IAM, Nancy, France.

[3]Sorbonne Universités, CNRS, IBMM, Roscoff, France.

*Corresponding author: Claude.didierjean@univ-lorraine.fr

[†]These authors contributed equally to this work.

Protein	Sequence
TvGSTX1^{WT}	CATATGAGCTCTACTACGCGCGACGTCTCGCACTTGTCCGACATCTCCAAGAGCATCTCT GAGCCAGACGGCTCGTTCAACCGCAAGCCCTCCGTGTTCAGGAACTTCGTGGAAAAGG GCGGCGCGTTCGCGCCCGAGAAGGGCCGGTACCACCTCTATGTGTCCTACGCATGCCCC TGGGCGACGCGCACGCTCATCGTGCGCAAGATCAAGGGCCTCGAGGAGTTCATAGACG TTAGCGTCGTCTCGCCGCGCATGGGCGCGCACGGCTGGCCGTTTGCGAACGTCGACGCG TTCCCCGGCGCCGACGCGGACCCCCTCTTCGGCGCGGAGCACGTCAAAGATTTGTACAT GCGCGTGAACCCCGACTACGAGGGCCGCTTCACGGTTCCTATCCTCTGGGACAAGAAG ACCTCGGCGATCGTGAACAACGAGAGCTCGGAGATCATCCGGATGTTCAACACCGCGT TCAATGACCAGCTCCCCTCGGACAAGGCCGCGATCGACCTCTACCCGGAGGCGCTCCGC GGGGAGATCGACGGGATTAACGAGTGGGTGTACGACACCGTCAACAACGGCGTGTAC AAGTCCGGCTTCGCGACATCACAGAAGGCATACGAGGCCGCGGTCGTCCCCCTGTTCG AGTCGCTCGACCGGCTCGAGAAGATCCTCACTGGCAAGGACTACCTCGTCGGCGACCA GCTCACAGAGGCAGATATTCGCCTCTTTGTGACCATCGTGCGCTTCGACCCCGTCTACGT CGGCCACTTCAAGTGCAACATCCGCACCATCCGCGATGGCTACCCTGCGATTGACCTGT GGATGCGGAAGCTGTACTGGAACAACTCCGCGTTCTCGTCGACGTGCAACTTTGAGCAC ATCAAGACGCATTACTACTGGTCGCATCCCCAGATTAACCCGCACCGCGTCGTGCCCGT GGGCCCGATCCCGAACATTCTTCCTCTGCACCATCACCATCACCATTAAGCGGCCGC
TvGSTX3^{WT}	CATATGGCAGCCACACGCGACGTCACCCACCTCACCGACATCAGCAAGGGCCTCGCGG AGGCGGACGGCTCGTTCAAGCGCAAAGCGTCGACCTTCCGCAGGTTCATCGAGAAGGG CGGCGAGTTCGAGCCTGAGAAAGGCCGGTACCATCTCTACGTCGCGTATTCCTGCCCGT GGGCGACGCGAACTCTGATCGTGCGGAAGATTAAGGGCCTCGAAGAGATCGTAGGGG TCACCATCGTGTCTCCGCTGTTCAGCGCGCATGGGTGGCCGTTCGGGGACGTCTCGCCCT TCCCTGGTGCCGAAGCCGACCCCTTCTACAACGCCCAGTACGTCAGGGATCTGTATCTC CGCGCGGACCCCAAATACGAGGGCCGCTTCACCGTCCCCGTGCTCTGGGATAAGAAGA CGGAGACGGTGGTGAACAACGAGAGCTCGGAGATCATCCGCATCTTCAACACCGCCTT CAACGAGTTCCTGCCCGCAGACAAAGCTGCGATCCACCTCTACCCGGAGGCTCTCAAG AGTGAGATCGACGAGATTAACGAGTGGGTGTACGATACCGTGAACAATGGTGTATACA AGGCCGGGTTTGCGACGACTCAGCAGGCGTACGAAGCGGCGGTCATCCCGCTCTTCGA GTCGCTCGACCGCCTCGAGAAGATTCTCACTGGCAAGGATTACCTGGTCGGAGATCAGC TTACCGAGGCAGACGTCCGGCTCTTCGTGACGATCATACGCTTCGATCCGGCGTACGTC GGACACTTCAAGTGCAACCTTCGCACTATCCGCGACGGGTACCCTGCGATCCACCTATG GCTCCGCAAGCTGTACTGGAACAACTCGGCATTCTCGGAGACGTGCAAGTTTGACCATA TCAAGGCGAGCTATTACGCGCAGAAGAATGTCAACCCTACCCTGGTTGTCCCTTTGGGT CCCATCCCGAACATCTTGCCTTTGCACCATCACCATCACCATTGAGCGGCCGC

Protein	Sequence
TvGSTX3^{C56S}	CATATGGCAGCCACACGCGACGTCACCCACCTCACCGACATCAGCAAGGGCCTCGCGG AGGCGGACGGCTCGTTCAAGCGCAAAGCGTCGACCTTCCGCAGGTTCATCGAGAAGGG CGGCGAGTTCGAGCCTGAGAAAGGCCGGTACCATCTCTACGTCCGCTATTCCTCCCCGT GGGCGACGCGAACTCTGATCGTGCGGAAGATTAAGGGCCTCGAAGAGATCGTAGGGG TCACCATCGTGTCTCCGCTGTTCAGCGCGCATGGGTGGCCGTTCGGGGACGTCTCGCCCT TCCCTGGTGCCGAAGCCGACCCCTTCTACAACGCCCAGTACGTCAGGGATCTGTATCTC CGCGCGGACCCCAAATACGAGGGCCGCTTCACCGTCCCCGTGCTCTGGGATAAGAAGA CGGAGACGGTGGTGAACAACGAGAGCTCGGAGATCATCCGCATCTTCAACACCGCCTT CAACGAGTTCCTGCCCGCAGACAAAGCTGCGATCCACCTCTACCCGGAGGCTCTCAAG AGTGAGATCGACGAGATTAACGAGTGGGTGTACGATACCGTGAACAATGGTGTATACA AGGCCGGGTTTGCGACGACTCAGCAGGCGTACGAAGCGGCGGTCATCCCGCTCTTCGA GTCGCTCGACCGCCTCGAGAAGATTCTCACTGGCAAGGATTACCTGGTCGGAGATCAG CTTACCGAGGCAGACGTCCGGCTCTTCGTGACGATCATACGCTTCGATCCGGCGTACGT CGGACACTTCAAGTGCAACCTTCGCACTATCCGCGACGGGTACCCTGCGATCCACCTAT GGCTCCGCAAGCTGTACTGGAACAACTCGGCATTCTCGGAGACGTGCAAGTTTGACCAT ATCAAGGCGAGCTATTACGCGCAGAAGAATGTCAACCCTACCCTGGTTGTCCCTTTGGG TCCCATCCCGAACATCTTGCCTTTGCACCATCACCATCACCATTGAGCGGCG
TvGSTX3^{S295H}	CATATGGCAGCCACACGCGACGTCACCCACCTCACCGACATCAGCAAGGGCCTCGCGG AGGCGGACGGCTCGTTCAAGCGCAAAGCGTCGACCTTCCGCAGGTTCATCGAGAAGGG CGGCGAGTTCGAGCCTGAGAAAGGCCGGTACCATCTCTACGTCGCGTATTCCTGCCCGT GGGCGACGCGAACTCTGATCGTGCGGAAGATTAAGGGCCTCGAAGAGATCGTAGGGG TCACCATCGTGTCTCCGCTGTTCAGCGCGCATGGGTGGCCGTTCGGGGACGTCTCGCCCT TCCCTGGTGCCGAAGCCGACCCCTTCTACAACGCCCAGTACGTCAGGGATCTGTATCTC CGCGCGGACCCCAAATACGAGGGCCGCTTCACCGTCCCCGTGCTCTGGGATAAGAAGA CGGAGACGGTGGTGAACAACGAGAGCTCGGAGATCATCCGCATCTTCAACACCGCCTT CAACGAGTTCCTGCCCGCAGACAAAGCTGCGATCCACCTCTACCCGGAGGCTCTCAAG AGTGAGATCGACGAGATTAACGAGTGGGTGTACGATACCGTGAACAATGGTGTATACA AGGCCGGGTTTGCGACGACTCAGCAGGCGTACGAAGCGGCGGTCATCCCGCTCTTCGA GTCGCTCGACCGCCTCGAGAAGATTCTCACTGGCAAGGATTACCTGGTCGGAGATCAG CTTACCGAGGCAGACGTCCGGCTCTTCGTGACGATCATACGCTTCGATCCGGCGTACGT CGGACACTTCAAGTGCAACCTTCGCACTATCCGCGACGGGTACCCTGCGATCCACCTAT GGCTCCGCAAGCTGTACTGGAACAACTCGGCATTCTCGGAGACGTGCAAGTTTGACCAT ATCAAGGCGCACTATTACGCGCAGAAGAATGTCAACCCTACCCTGGTTGTCCCTTTGGG TCCCATCCCGAACATCTTGCCTTTGCACCATCACCATCACCATTGAGCGGCCGC

Table S1. Nucleotide sequences of TvGSTX1^{WT} and TvGSTX3 $^{WT, C56S and S295H}$ designed to contain a His-tag at the end of the C-terminal part.

Sequences coding for TvGSTX1^{WT}, TvGSTX3^{WT} and TvGSTX3^{S295H} were obtained from gene synthesis whereas TvGSTX3^{C56S} was generated by a site-directed mutagenesis experiment.

	Mass expected (Da)	Mass of native proteins (Da)
TvGSTX1^{WT}	37,129	37,129
TvGSTX3^{WT}	36,982	36,982
TvGSTX3^{C56S}	36,966	36,966
TvGSTX3^{S295H}	37,032	37,033

Table S2. Mass spectrometry analysis of recombinant TvGSTXs

Molecular masses of mutant and wild-type TvGSTXs were analyzed after purification. Mass-expected values correspond to proteins devoided of the first Met residue and bearing a hexa-histidine tag. Mass spectrometry analysis have been done by using a Bruker microTOF-Q spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an Apollo II electrospray ionization source with an ion funnel and operated in the negative ion mode. A concentrated sample (around 100 µL at 100 µM) in formic acid has been injected. The potential between the spray needle and the orifice was set to 4.5 kV. Before each run, the instrument was calibrated externally with the TunemixTM mixture (Agilent Technologies) in quadratic regression mode. Data were analyzed with DataAnalysis software (Bruker).

		TvGSTX3^{WT}	**TvGSTX3^{GSH}**
		apo form	co-crystallized with GSH
Diffraction data			
	Diffraction source	FIP-BM30A. ESRF	FIP-BM30A. ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97974	0.97972
	Unit-cell parameters		
	a. b. c (Å)	146.4 57.5 94.6	72.4 72.4 316.7
	α. β. γ (°)	90 107.3 90	90
	Space group	C2	P43212
	Resolution range (Å)	48.68 - 2.28 (2.34 - 2.28)	48.69 - 1.80 (1.84 - 1.80)
	Total No. of reflections	125,406 (8,906)	650,332 (12,729)
	No. of unique reflections	33,851 (2,465)	78,924 (4,842)
	Average redundancy	3.7 (3.6)	8.2 (2.6)
	Mean I/σ(I)	14.0 (2.0)	13.7 (1.8)
	Completeness (%)	98.1 (97.8)	98.6 (83.7)
	R_{merge}	0.10 (0.69)	0.11 (0.39)
	R_{meas}	0.12 (0.81)	0.11 (0.48)
	R_{pim}	0.06 (0.37)	0.05 (0.29)
	$CC_{1/2}$	1.00 (0.76)	1.00 (0.76)
Refinement			
	Resolution range (Å)	48.68 - 2.28 (2.35 - 2.28)	48.69 - 1.80 (1.82 - 1.80)
	R_{work} / R_{free}	0.16 / 0.23 (0.25 / 0.36)	0.16 / 0.19 (0.25 / 0.26)
	No. of protein atoms	9,583	9,455
	No. of waters	358	492
	Average B factor (Å²)	26.4	14.8
Model quality			
	RMSZ bond lengths	0.57	0.46
	RMSZ bond angles	0.65	0.60
	Ramachandran favoured (%)	97.1	98.8
	Ramachandran outliers (%)	0.0	0.0
	Molprobity rotamer outliers (%)	0.6	0.2
	Molprobity clashscore	1.77	0.63
	Molprobity score	1.09	0.71
PDB code		6GCA	6GCB

		TvGSTX3^{C56S}	**TvGSTX1^{WT}**
		co-crystallized with dextran-sulfate	apo form
Diffraction data			
	Diffraction source	FIP-BM30A. ESRF	FIP-BM30A. ESRF
	Detector	ADSC Q315r CCD	ADSC Q315r CCD
	Wavelength (Å)	0.97977	0.97973
	Unit-cell parameters		
	a. b. c (Å)	72.3 72.3 317.6	65.0 65.0 316.4
	α. β. γ (°)	90	90
	Space group	P43212	P43212
	Resolution range (Å)	46.01 - 1.90 (1.94 - 1.90)	45.45 - 3.20 (3.42 - 3.20)
	Total No. of reflections	242,903 (9,766)	165,198 (30,309)
	No. of unique reflections	65,928 (3,578)	12,131 (2,132)
	Average redundancy	3.7 (2.7)	13.6 (9.7)
	Mean I/σ(I)	15.5 (3.4)	11.0 (4.2)
	Completeness (%)	97.6 (85.0)	100 (100)
	R_{merge}	0.06 (0.25)	0.25 (0.78)
	R_{meas}	0.07 (0.31)	0.26 (0.81)
	R_{pim}	0.03 (0.16)	0.07 (0.21)
	$CC_{1/2}$	1.00 (0.93)	1.00 (0.96)
Refinement			
	Resolution range (Å)	46.01 - 1.90 (1.93 - 1.90)	45.45 - 3.20 (3.52 - 3.20)
	R_{work} / R_{free}	0.16 / 0.19 (0.40 / 0.38)	0.22 / 0.25 (0.26 / 0.33)
	No. of protein atoms	9,489	4,787
	No. of waters	740	8
	Average B factor (Å²)	22.6	56.5
Model quality			
	RMSZ bond lengths	0.55	0.30
	RMSZ bond angles	0.63	0.45
	Ramachandran favoured (%)	98.4	94.4
	Ramachandran outliers (%)	0.0	1.2
	Molprobity rotamer outliers (%)	0.0	0.0
	Molprobity clashscore	0.94	10.28
	Molprobity score	0.78	1.91
PDB code		6GCC	6GC9

Table S3. Diffraction and refinement statistics

$R_{merge} = \sum_{hkl} \sum_i \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum_i I_i(hkl)$.

$R_{meas} = \sum_{hkl} \{N(hkl)/[N(hkl)-1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

$R_{pim} = \sum_{hkl} \{1/[N(hkl)-1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

$R_{work} = \sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right| / \sum_{hkl} |F_{obs}|$.

$CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data[1]. 5% of reflections were selected for R_{free} calculation. RMSZ : root mean square Z-score [2]. The Molprobity clashscore is the number of serious clashes per 1000 atoms[3]. The Molprobity score is a log-weighted combination of the clashscore, percentage Ramachandran not favored and percentage bad side-chain rotamers[3]. Values in parentheses are for highest resolution shell.

	TvGSTX1^{WT}	TvGSTX3^{WT}	TvGSTX3^{GSH}	TvGSTX3^{C56S}
TvGSTX1^{WT}	-	**488**	**508**	**509**
TvGSTX3^{WT}	*0.569*	-	**507**	**501**
TvGSTX3^{GSH}	*0.565*	*0.244*	-	**552**
TvGSTX3^{C56S}	*0.566*	*0.267*	*0.143*	-

Table S4. Superimpositions of the TvGSTX structures

The rmsd values are in italics (Å, below diagonal) and the number of superimposed Cα atoms is in bold (above diagonal).

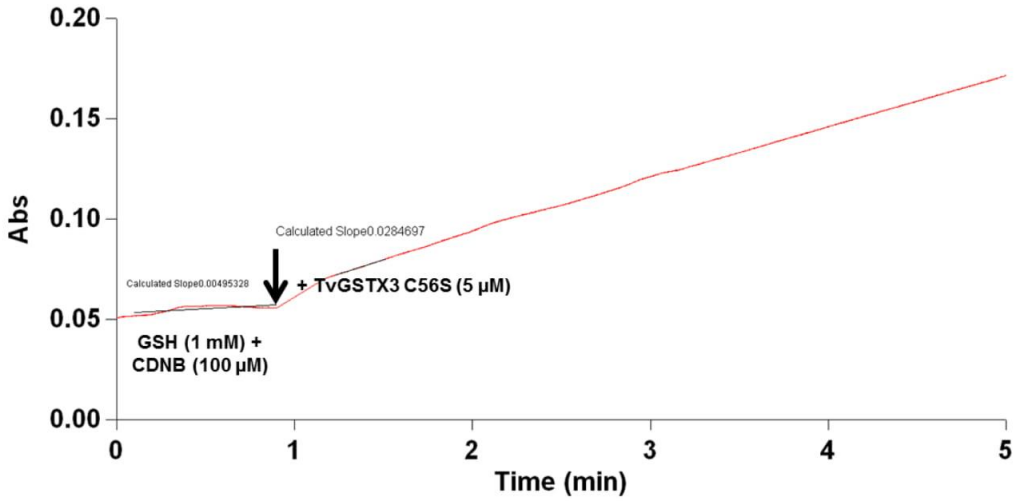


Figure S1. Glutathione transferase activity towards CDNB for the mutant TvGSTX3^{C56S}

Glutathione transferase activity was assessed by using 1-Chloro 2,4-dinitrobenzene (CDNB) and glutathione (GSH) in a 30 mM Tris-HCl pH 8.0, 1 mM EDTA buffer at room temperature. Enzymatic reaction was triggered by adding 5 µM of TvGSTX3^{C56S} to the mixture (black arrow). Absorbance of the product was monitored at 340 nm.

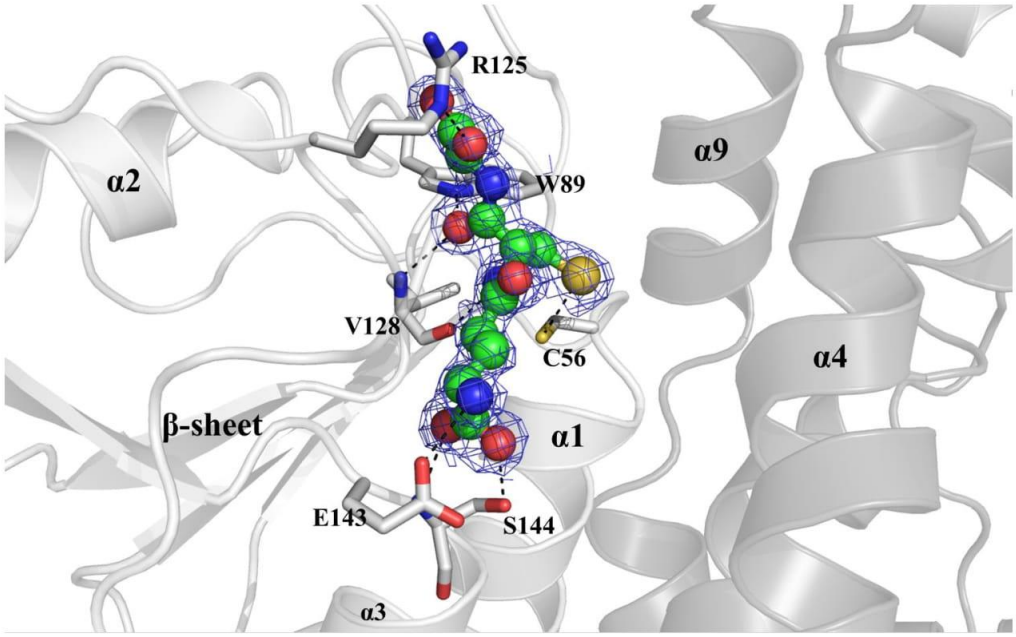
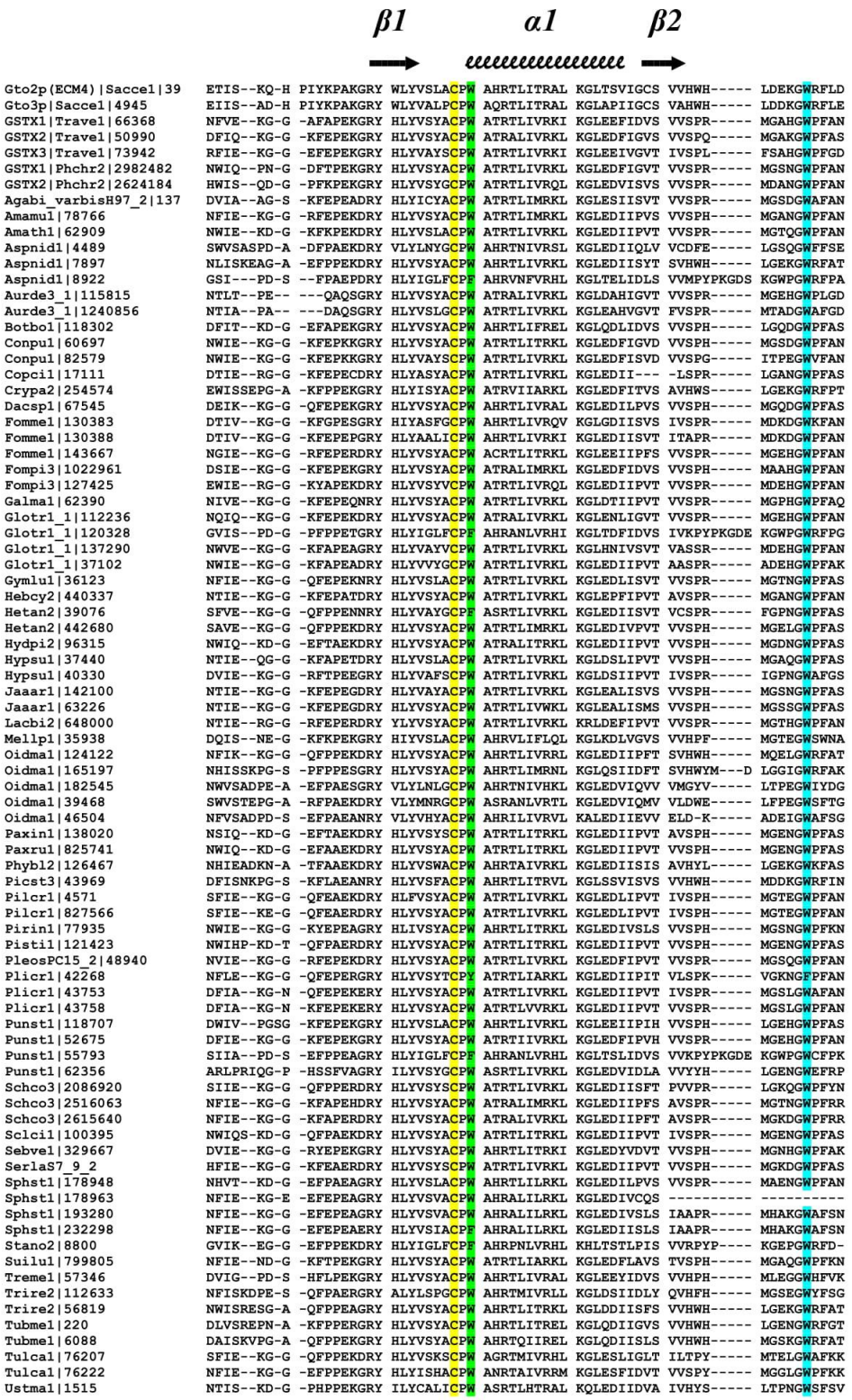


Figure S2. Glutathione bound to TvGSTX3 G site

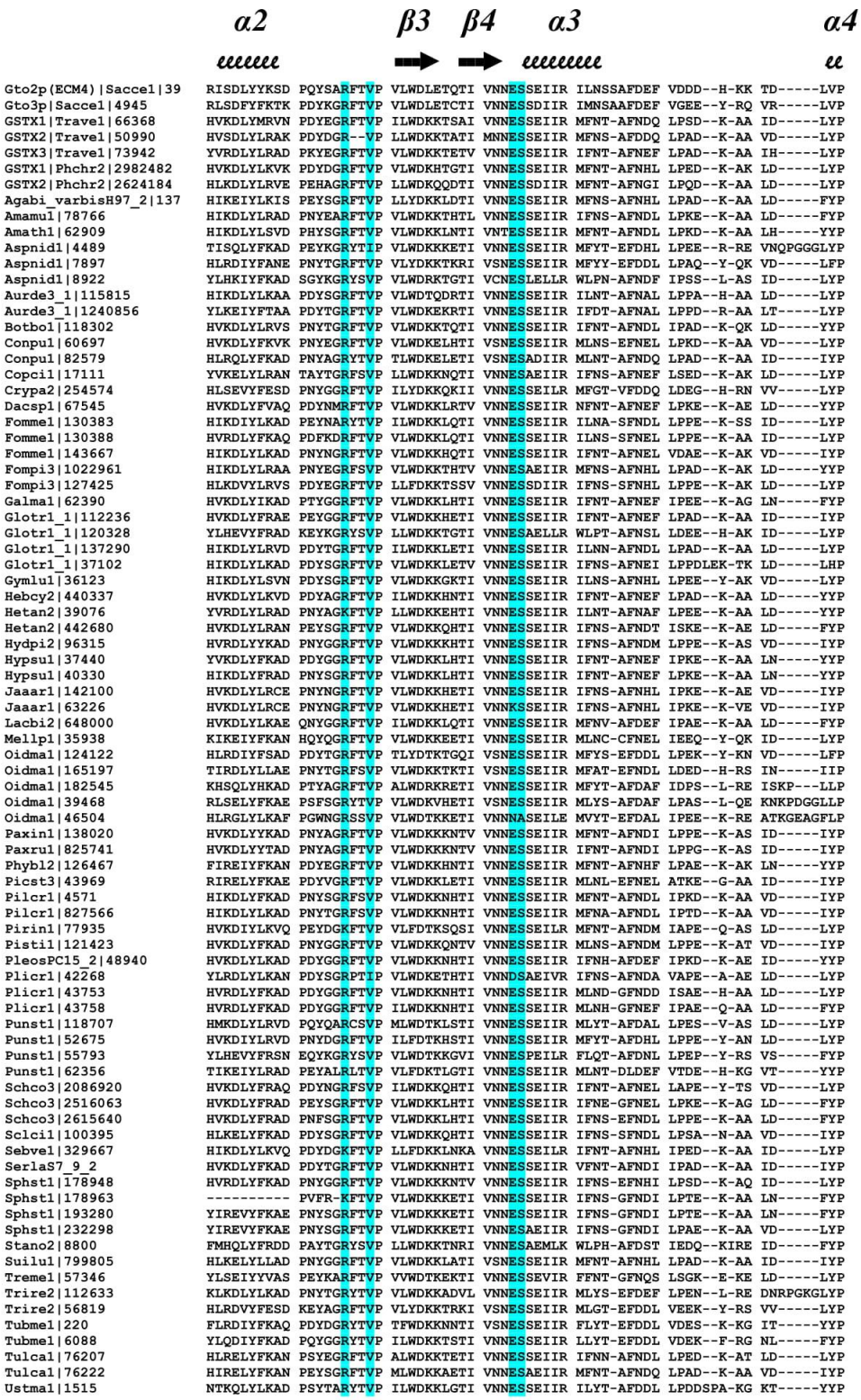
2Fo-Fc electron density omit map around glutathione is shown at 1.0 σ. Glutathione binding site residues are displayed as sticks and polar contacts are represented as dashed lines.

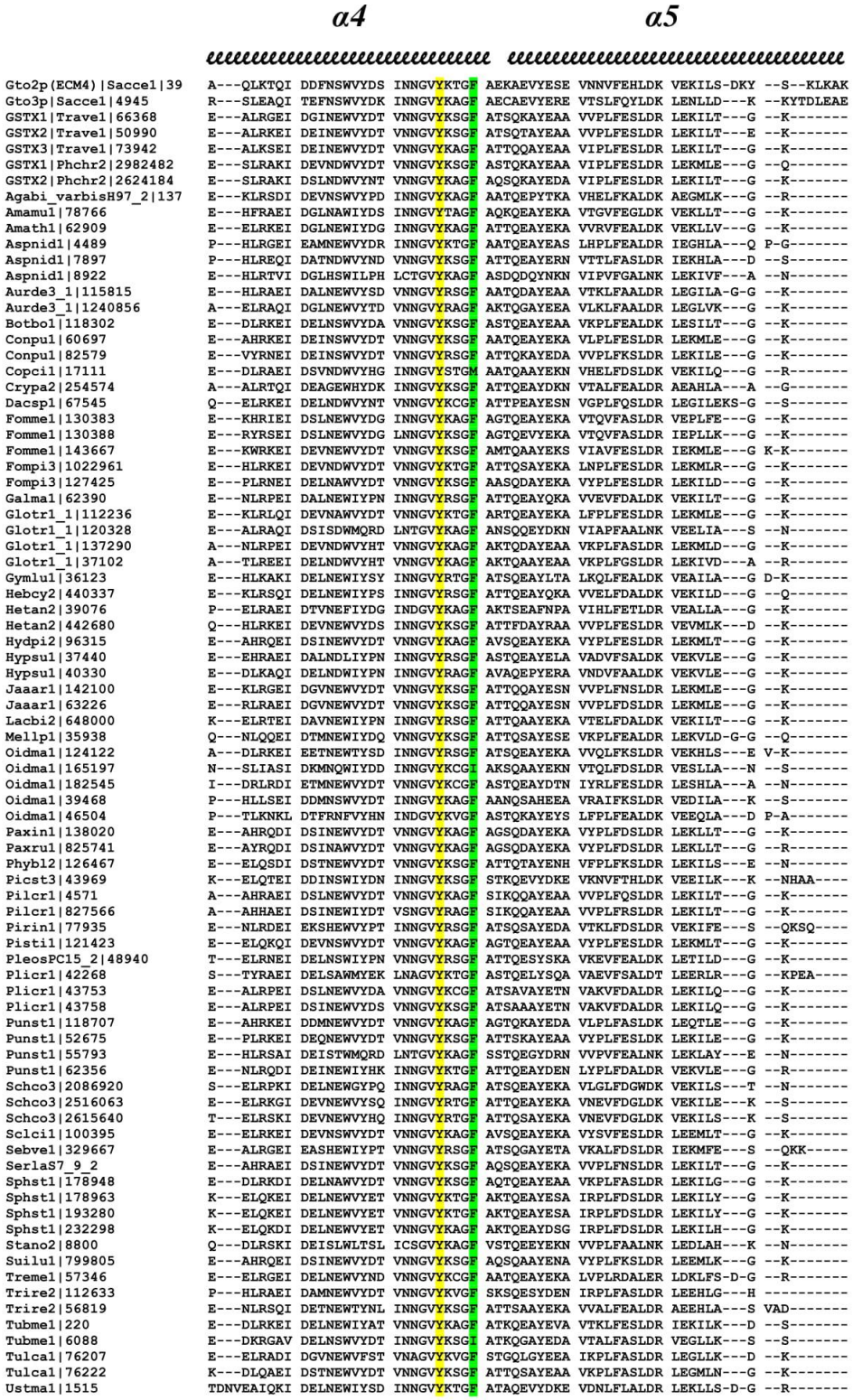
```
Gto2p(ECM4)|Sacce1|39   ---------- ---------- ---------- ---------- ---------- ---------M S---------
Gto3p|Sacce1|4945       ---------- ---------- ---------- ---------- ---------- ---------M S---------
GSTX1|Trave1|66368      M--------- ---------- ---------- ---------- ---------- ---------S S---------
GSTX2|Trave1|50990      ---------- ---------- ---------- ---------- ---------- ---------M S---------
GSTX3|Trave1|73942      ---------- ---------- ---------- ---------- ---------- ---------M A---------
GSTX1|Phchr2|2982482    MSFT------ ------TGST ------LQHQ --------EA QLQGAKDETV T---QLEQRA A---------
GSTX2|Phchr2|2624184    MTSLSASGSA IHITFLTNGL ------LDTL TRRFTSTPLR LLRPASYVQA PYISLRKANM ST--------
Agabi_varbisH97_2|137   M--------- ---------- ---------- ---------- ---------- ---------A A---------
Amamu1|78766            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Amath1|62909            ---------- ---------- ---------- ---------- ---------- ---------- M---------
Aspnid1|4489            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Aspnid1|7897            ---------- ---------- ---------- ---------- ---------- ---------M F---------
Aspnid1|8922            ---------- ---------- ---------- ---------- ---------- ---------- ----------
Aurde3_1|115815         ---------- ---------- ---------- ---------- ---------- ---------- ----------
Aurde3_1|1240856        ---------- ---------- ---------- ---------- ---------- ---------- ----------
Botbo1|118302           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Conpu1|60697            ---------- ---------- ---------- ---------- ---------- ---------- ----------
Conpu1|82579            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Copci1|17111            ---------- ---------- ---------- ---------- ---------- ---------- M---------
Crypa2|254574           ---------- ---------- ---------- ---------- ---------- ---------M K---------
Dacsp1|67545            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Fomme1|130383           ---------- ---------- ---------- ---------- ---------- ---------M A---------
Fomme1|130388           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Fomme1|143667           M--------- ---------- ---------- ---------- ---------- ---------S T---------
Fompi3|1022961          MTLV------ ------IRLL ------LASR T------RTS RLSATRFPSS LH-LHITSRT M---------
Fompi3|127425           ---------- ---------- ---------- ---------- ---------- ---------- M---------
Galma1|62390            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Glotr1_1|112236         ---------- ---------- ---------- ---------- ---------- ---------- ----------
Glotr1_1|120328         ---------- ---------- ---------- ---------- ---------- ---------M S---------
Glotr1_1|137290         ---------- ---------- ---------- ---------- ---------- ---------- ----------
Glotr1_1|37102          ---------- ---------- ---------- ---------- ---------- ---------- M---------
Gymlu1|36123            ---------- ---------- ---------- ---------- ---------- ---------- M---------
Hebcy2|440337           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Hetan2|39076            ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hetan2|442680           M--------- ---------- ---------- ---------- ---------- ---------A A---------
Hydpi2|96315            ---------- ---------- ---------- ---------- ---------- ---------M T---------
Hypsu1|37440            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Hypsu1|40330            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Jaaar1|142100           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Jaaar1|63226            ---------- ---------- ---------- ---------- ---------- ---------M P---------
Lacbi2|648000           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mellp1|35938            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Oidma1|124122           ---------- ---------- ---------- ---------- ---------- ---------M R---------
Oidma1|165197           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Oidma1|182545           ---------- ---------- ---------- ---------- ---------- ---------M E---------
Oidma1|39468            ---------- ---------- ---------- ---------- ---------- ---------M AP--------
Oidma1|46504            ---------- ---------- ---------- ---------- ---------- ---------M T---------
Paxin1|138020           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Paxru1|825741           MSADEQVKFL PGMATPIRQL IIRGCSLASQ RPL---IAPS SLPKHSHSLV E---RSSFRT S---------
Phybl2|126467           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Picst3|43969            ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pilcr1|4571             ---------- ---------- ---------- ---------- ---------- ---------M S---------
Pilcr1|827566           ---------- ---------- ---------- ---------- ---------- ---------M A---------
Pirin1|77935            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Pisti1|121423           ---------- ---------- ---------- ---------- ---------- ---------M S---------
PleosPC15_2|48940       ---------- ---------- ---------- ---------- ---------- ---------- ----------
Plicr1|42268            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Plicr1|43753            MS--------- ---------- ---------- ---------- ---------- --------TA A---------
Plicr1|43758            MS--------- ---------- ---------- ---------- ---------- ---------T A---------
Punst1|118707           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Punst1|52675            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Punst1|55793            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Punst1|62356            ---------- ---------- ---------- ---------- ---------- ---------M V---------
Schco3|2086920          ---------- ---------- ---------- ---------- ---------- ---------M T---------
Schco3|2516063          ---------- ---------- ---------- ---------- ---------- ---------M SS--------
Schco3|2615640          MSTTL----- -----FTRRW -----ALVGR TR----VPPD IIGTRA---- ---------M STYAR-----
Sclci1|100395           ---------- ---------- ---------- ---------- ---------- ---------M S---------
Sebve1|329667           ---------- ---------- ---------- ---------- ---------- ---------- ----------
SerlaS7_9_2             ---------- ---------- ---------- ---------- ---------- ---------M T---------
Sphst1|178948           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sphst1|178963           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sphst1|193280           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sphst1|232298           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Stano2|8800             ---------- ---------- ---------- ---------- ---------- ---------M T---------
Suilu1|799805           MAVAL----- ----IRSRLL ------MVSI K------PLT VVSSSPRSSL LY-HTRTMTS T---------
Treme1|57346            ---------- ---------- ---------- ---------- ---------- ---------M S---------
Trire2|112633           ---------- ---------- ---------- ---------- ---------- ---------M AS--------
Trire2|56819            ---------- ---------- ---------- ---------- ---------- ---------M LSASY-----
Tubme1|220              ---------- ---------- ---------- ---------- ---------- ---------M FFRSIFRQTF
Tubme1|6088             ---------- ---------- ---------- ---------- ---------- ---------M A---------
Tulca1|76207            ---------- ---------- ---------- ---------- ---------- ---------- M---------
Tulca1|76222            ---------- ---------- ---------- ---------- ---------- ---------M A---------
Ustma1|1515             ---------- ---------- ---------- ---------- ---------- ---------M V---------
```

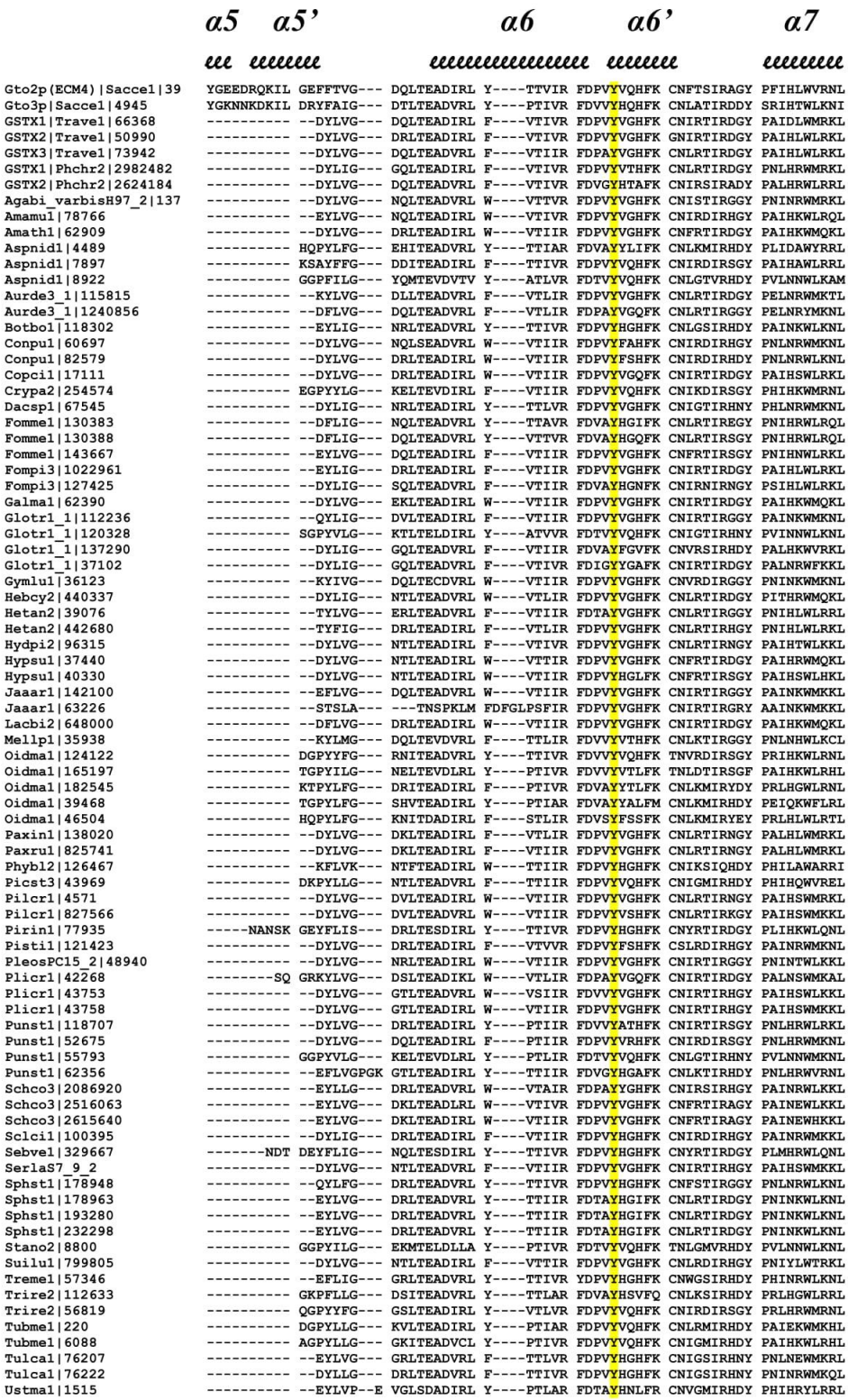
```
Gto2p(ECM4)|Sacce1|39   ---------- ---------- ---------- ---------- KQW---A-SG TNGAFK---- --R-QVSSFR
Gto3p|Sacce1|4945       ---------- ---------- ---------- ---------- EKS---A-SN NKAEFK---- --R-QSSPFR
GSTX1|Trave1|66368      ---------- ---------- --------TT RDVSHLSDI- SKS---I-SE PDGSFN---- --R-KPSVFR
GSTX2|Trave1|50990      ---------- ---------- --------ST RDVSHLTDI- TKR---P-LE PDGTFK---- --R-RPTSFR
GSTX3|Trave1|73942      ---------- ---------- --------AT RDVTHLTDI- SKG---L-AE ADGSFK---- --R-KASTFR
GSTX1|Phchr2|2982482    ---------- ---------- --------GG ASHELQSDI- SKM---K-TE DDGSFK---- --R-KAASFR
GSTX2|Phchr2|2624184    ---------G STLPHQHTAL QQANEQRLPA RD-PGQSDV- SKM---K-MQ ADGSFD---- --R-KPSTFR
Agabi_varbisH97_2|137   ---------- ---------- --------MT REVSSQSDI- GKM---K-TE PNGAYK---- --R-LEASFR
Amamu1|78766            ---------- ---------- --------ST KDISQFTDI- GKM---K-TE PDGTFK---- --R-QVSQFR
Amath1|62909            ---------- ---------- --------ST RDVTDQSDI- TKM---K-VE ADGSFK---- --R-LDASFR
Aspnid1|4489            ---------- ---------- ----------ENET- YKH-----AD KDGHFR---- --R-KDSVFR
Aspnid1|7897            ---------- ----LVLRR- --------AN QLTDISKKI- TDWV--NPGD KSGEFK---- --R-QASVFR
Aspnid1|8922            ---------- ---------- ------MPH- HEP-----ST AD-------- -------GWH
Aurde3_1|115815         ---------- ---------- --------MS RDVSSQVDI- SKG---KFHG ADGSFQ---- --R-APSSFR
Aurde3_1|1240856        ---------- ---------- --------MS RDTSSQSDI- TKL---P-TA ADGSFQ---- --R-APSSFR
Botbo1|118302           ---------- ---------- ----------MSQSDI- TKW-----HS DDGHFR---- --R-QQSTFR
Conpu1|60697            ---------- ---------- ---------- ---------- --M---A-TE KDGSFK---- --R-ADAAFR
Conpu1|82579            ---------- ---------- ---------- ---SKNADH- RIV---P-TD KDGSFK---- --R-PDSAFR
Copci1|17111            ---------- ---------- --------TS NDSSSQADV- YKA---P-RN EDGSIK---- --F-HTTSFR
Crypa2|254574           ---------- ---------- ---------- -QSTSPFKV- TNWV--DPKD KSGEFK---- --R-QVSSFR
Dacsp1|67545            ---------- ---------- ---------- ---AEKKDI- TSW-----AA KDGQFK---- --R-QASTFR
Fomme1|130383           ---------- ---------- --------SD AGTGDQTDL- TKL---E-RE KDGSFK---- --R-QASSFR
Fomme1|130388           ---------- ---------- --------SA PGTGDQTDLR NNL---S-RD KDGSFK---- --R-QVSTFR
Fomme1|143667           ---------- ---------- --------AT RDVSHQSDI- SKM---K-AE GDGEFK---- --R-KPSTFR
Fompi3|1022961          ---------- ---------- --------GT RDVTHVSDI- TKRV--Q-TA ADGSFL---- --R-PASQFR
Fompi3|127425           ---------- ---------- --------ST RDVTHLTDI- NKM---A-PE QDGSFN---- --R-RPSQFR
Galma1|62390            ---------- ---------- --------TT KDTTLQSDI- TKF---K-TE SDGSFK---- --R-ADASFR
Glotr1_1|112236         ---------- ---------- --------MS RNAEEQSDI- TRL---K-FE ADGSFK---- --R-LDSSFR
Glotr1_1|120328         ---------- ---------- ---------- ---SAATNI- HEL-----GT AD-------- -------GWH
Glotr1_1|137290         ---------- ---------- --------MS RDTRSQSDV- SKQ---L-IE KDGSFK---- --R-PDSVFR
Glotr1_1|37102          ---------- ---------- --------SG GDCQSQADV- SQQ---P-IG KDGSFK---- --R-PDSVFR
Gymlu1|36123            ---------- ---------- --------ST RDVSHQSDI- AKM---K-TE PDGTFK---- --R-LDSIFR
Hebcy2|440337           ---------- ---------- --------TT KDTTLQSDI- GKM---K-TE PDGSFK---- --R-LDASFR
Hetan2|39076            ---------- ---------- --------MS TTAVSQANP- AEK---K-GE PDGSFK---- --R-PASSFR
Hetan2|442680           ---------- ---------- --------AA RDVSHQSNI- AKM---R-TD GDGSFK---- --R-LPSTFR
Hydpi2|96315            ---------- ---------- --------SA RDSSHQSDI- GKM---K-LE ADGSFK---- --R-AASAFR
Hypsu1|37440            ---------- ---------- --------TA KDATLQSDI- TKM---K-RQ PDGSFN---- --R-APSSFR
Hypsu1|40330            ---------- ---------- --------TP KDSVSQSDV- GLM---K-RE PDGSFK---- --R-PDSSFR
Jaaar1|142100           ---------- ---------- --------MS KDVSLQADI- SKM---G-SQ PDGSFK---- --R-APASFR
Jaaar1|63226            ---------- ---------- --------AS KDVSLQADI- SKM---G-SQ PDGSFK---- --R-APASFR
Lacbi2|648000           ---------- ---------- ---------- ---------- --M---K-TE PDGSFK---- --R-LDASFR
Mellp1|35938            ---------- ---------- --NSNSKLDH SIQSSRKEI- TTW-----AS EDGHFR---- --R-QASSFR
Oidma1|124122           ---------- ----GQLL-- --------EL TSHSQATKI- TDWV--APND KSGEFK---- --R-GQSQFR
Oidma1|165197           ---------- ---------- --------ET KKTAAAESI- NSV---V-DA ATGEFK---- --R-PAAQFR
Oidma1|182545           ---------- ---------- ---------S KTEAESKDV- TEVVFKA-AE SYGHYQ---- --R-KASQFR
Oidma1|39468            ---------- ---------- --------ST ADNPEDKDV- YKF---N-FE KNGTFK---- --R-QVSSFR
Oidma1|46504            ---------- ---------- ---------- ----IKNDP- GSF-----AD DDGHFR---- --R-PEAKIR
Paxin1|138020           ---------- ---------- ---------M TSTSNQADI- GKM---K-LE SDGSFK---- --R-AASAFR
Paxru1|825741           ---------- ---------- --------SM TSTSNQADI- GKM---K-LE SDGSFK---- --R-AASAFR
Phybl2|126467           ---------- ---------- ---------- ----TKPAI- LNW-----AS KDGEFK---- --R-QQSVFR
Picst3|43969            ---------- ---------- ---------- ---------- LKF-----AA KDGAFH---- --R-RPSLFR
Pilcr1|4571             ---------- ---------- --------TK KDASLQSDI- TKF---R-HE PDGSFR---- --R-APSVFR
Pilcr1|827566           ---------- ---------- --------TK KDTSLQSDF- AKL---K-YE PDGSYK---- --R-IPSAFR
Pirin1|77935            ---------- ---------- --ENNSNIGA RDVSNMSDI- SKS---K-ME KDGSYN---- --R-RPSSFR
Pisti1|121423           ---------- ---------- --------SV RDTTLQSDI- GKM---K-LE NDGSFK---- --R-ASSAFR
PleosPC15_2|48940       ---------- ---------- ---------- ---------- --M---K-PE ADGSFK---- --R-ADASFR
Plicr1|42268            ---------- ---------- ---------- --TQVQTDP- TKR---E-TN PDGSFK---- --A-NVSTFR
Plicr1|43753            ---------- ---------- --------TQ RDTTLQSDI- TKM---K-TE PDGSFK---- --R-AASQFR
Plicr1|43758            ---------- ---------- --------TQ RDTTLQSDI- AKM---K-TE PDGSFK---- --R-ADAQFR
Punst1|118707           ---------- ---------- ---------- --STDLTDP- AKF---K-LQ ADGSLK---- --R-PQSAFR
Punst1|52675            ---------- ---------- --------AT RDVSKQSDI- TKM---T-TD PDGSFK---- --R-KASTFR
Punst1|55793            ---------- ---------- ---------- ---ATQAD-- --------GS DHKDFV---- --P-PDGSWH
Punst1|62356            ---------- ---------- ---------- ----LSRHL- RGFC--SPRS ASGSAKTGDT LLR-KVAVFT
Schco3|2086920          ---------- ---------- ---------- ---------- ------T-NA PDGSYQ---- --R-PPSQFR
Schco3|2516063          ---------- ---------- ---------- ---------- ------N-EK SDGSFK---- --R-QDSSFR
Schco3|2615640          -RLACPSHPQ PSLRYQPSK- --------LK RTMTTQLDIG TRV---Q-TD KDGSFV---- --R-PASSFR
Sclci1|100395           ---------- ---------- --------AS KDHTLQSDI- GKM---T-LE ADGSFK---- --R-AASSFR
Sebve1|329667           ---------- ---------- ----------MSDI- SKS---K-ME QDGSFN---- --R-KPSTFR
SerlaS7_9_2             ---------- ---------- --------AK RDTSHQSDI- AKL---K-TE ADGSFK---- --R-AASVFR
Sphst1|178948           ---------- ---------- ----------MSQSDI- TKS---NPID KDGSFK---- --R-AAASFR
Sphst1|178963           ---------- ---------- --------MS NHTSNQSVV- ------T-AN KQGAFV---- --R-PSSSFR
Sphst1|193280           ---------- ---------- --------MS NDTSNQSVV- ------T-AN KQGAFV---- --R-PSSSFR
Sphst1|232298           ---------- ---------- --------MS KDTSTPNPP- --------TN KQGAFV---- --R-TASSFR
Stano2|8800             ---------- ---------- ---------- ------TEI- HKY-----GT DD-------- -------GWH
Suilu1|799805           ---------- ---------- --------ST KDVSSQSDI- GKM---R-TE ADGSFK---- --R-AASSFR
Treme1|57346            ---------- ---------- ---------- ---TVTQKI- TDW---A-DK KDGSFK---- --R-QTSTFR
Trire2|112633           ---------- ---------- ---------- ---------- VQQ---T-YH DKGRFT---- --R-PDSTFR
Trire2|56819            ---------- ------LHCW -YPSDIPTSL QSSNQESKI- TDWV--KPGD KTGEFK---- --R-QVSSFR
Tubme1|220              SHLSSPLKAS PSLRYQLN-- --------YR AMSTSGQKI- TDWC--TPDS KRGEFN---- --R-KPSTFC
Tubme1|6088             ---------- ---------- ---------- -APIQKHKI- TEWV--AVDS KTGEFQ---- --R-KPSVFR
Tulca1|76207            ---------- ---------- --------SA PSNASQTPS- TDT-----AA AKGAWK---- --HPAPSSFR
Tulca1|76222            ---------- ---------- --------AT GNSDAQKNI- LTW---A--S NDGAFR---- --R-QASSFR
Ustma1|1515             ---------- ---------- ---------- -----EHSV- AVH-----AG KDGNFR---- --R-QASAFR
```

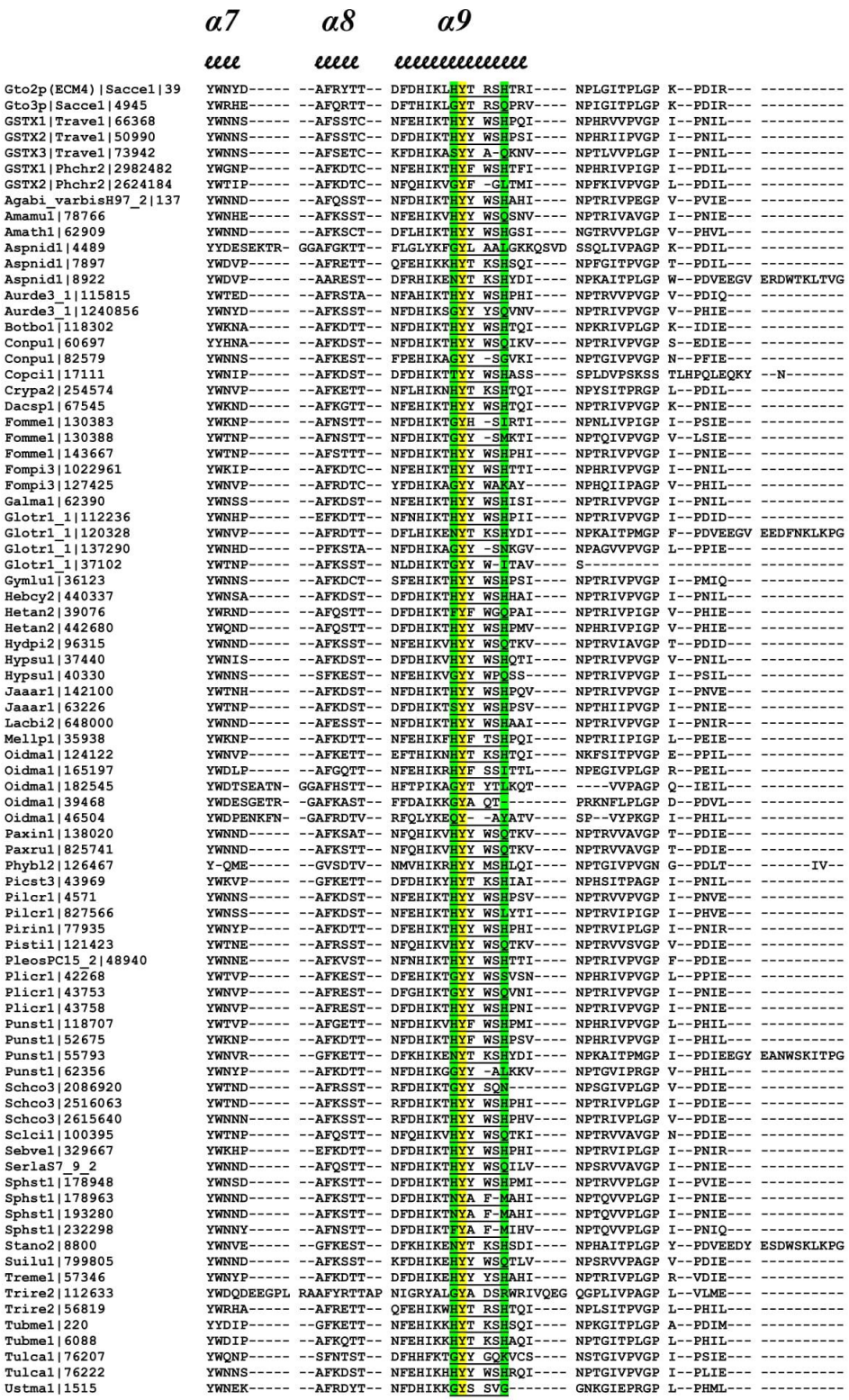

β1
α1
β2
Gto2p(ECM4)|Sacce1|39 ETIS--KQ-H PIYKPAKGRY WLYVSLACPW AHRTLITRAL KGLTSVIGCS VVHWH----- LDEKGWRFLD
Gto3p|Sacce1|4945 EIIS--AD-H PIYKPAKGRY WLYVALPCPW AQRTLITRAL KGLAPIIGCS VAHWH----- LDDKGWRFLE
GSTX1|Trave1|66368 NFVE--KG-G -AFAPEKGRY HLYVSYACPW ATRTLIVRKI KGLEEFIDVS VVSPR----- MGAHGWPFAN
GSTX2|Trave1|50990 DFIQ--KG-G -KFEPEKGRY HLYVSYACPW ATRALIVRKL KGLEDFIGVS VVSPQ----- MGAKGWPFAS
GSTX3|Trave1|73942 RFIE--KG-G -EFEPEKGRY HLYVAYSCPW ATRTLIVRKI KGLEEIVGVT IVSPL----- FSAHGWPFGD
GSTX1|Phchr2|2982482 NWIQ--PN-G -DFTPEKGRY HLYVSYACPW ATRTLIVRKL KGLEDFIGVT VVSPR----- MGSNGWPFAN
GSTX2|Phchr2|2624184 HWIS--QD-G -PFKPEKGRY HLYVSYGCPW ATRTLIVRQL KGLEDVISVS VVSPR----- MDANGWPFAN
Agabi_varbisH97_2|137 DVIA--AG-S -KFEPEADRY HLYICYACPW ATRTLIMRKL KGLESIISVT VVSPR----- MGSDGWAFAN
Amamu1|78766 NFIE--KG-G -RFEPEKDRY HLYVSYACPW ATRTLIMRKL KGLEDIIPVS VVSPH----- MGANGWPFAN
Amath1|62909 NWIE--KD-G -KFKPEKDRY HLYVSLACPW ATRTLIVRKL KGLEDIIPVT VVSPR----- MGTQGWPFAN
Aspnid1|4489 SWVSASPD-A -DFPAEKDRY VLYLNYGCPW AHRTNIVRSL KGLEDIIQLV VCDFE----- LGSQGWFFSE
Aspnid1|7897 NLISKEAG-A -EFPPEKDRY HLYVSYACPW AHRTLIVRKL KGLEDIISYT SVHWH----- LGEKGWRFAT
Aspnid1|8922 GSI---PD-S --FPAEPDRY HLYIGLFCPF AHRVNFVRHL KGLTELIDLS VVMPYPKGDS KGWPGWRFPA
Aurde3_1|115815 NTLT--PE-- ---QAQSGRY HLYVSYACPW ATRALIVRKL KGLDAHIGVT VVSPR----- MGEHGWPLGD
Aurde3_1|1240856 NTIA--PA-- ---DAQSGRY HLYVSLGCPW ATRTLIVRKL KGLEAHVGVT FVSPR----- MTADGWAFGD
Botbo1|118302 DFIT--KD-G -EFAPEKGRY HLYVSYACPW AHRTLIFREL KGLQDLIDVS VVSPH----- LGQDGWPFAS
Conpu1|60697 NWIE--KG-G -KFEPKKGRY HLYVSYACPW ATRTLITRKL KGLEDFIGVD VVSPH----- MGSDGWPFAN
Conpu1|82579 NWIE--KG-G -KFEPKKGRY HLYVAYSCPW ATRTLIVRKL KGLEDFISVD VVSPG----- ITPEGWVFAN
Copci1|17111 DTIE--RG-G -KFEPECDRY HLYASYACPW ATRTLIVRKL KGLEDII--- -LSPR----- LGANGWPFAS
Crypa2|254574 EWISSEPG-A -KFPPEKGRY HLYISYACPW ATRVIIARKL KGLEDFITVS AVHWS----- LGEKGWRFPT
Dacsp1|67545 DEIK--KG-G -QFEPEKGRY HLYVSYACPW AHRTLIVRAL KGLEDILPVS VVSPH----- MGQDGWPFAS
Fomme1|130383 DTIV--KG-G -KFGPESGRY HIYASFGCPW AHRTLIVRQV KGLGDIISVS IVSPR----- MDKDGWKFAN
Fomme1|130388 DTIV--KG-G -KFEPEPGRY HLYAALICPW AHRTLIVRKI KGLEDIISVT ITAPR----- MDKDGWPFAN
Fomme1|143667 NGIE--KG-G -RFEPERDRY HLYVSYACPW ACRTLITRKL KGLEEIIPFS VVSPR----- MGEHGWPFAN
Fompi3|1022961 DSIE--KG-G -KFEPEKGRY HLYVSYACPW ATRALIMRKL KGLEDFIDVS VVSPH----- MAAHGWPFAN
Fompi3|127425 EWIE--RG-G -KYAPEKDRY HLYVSYVCPW ATRTLIVRQL KGLEDIIPVT VVSPR----- MDEHGWPFAN
Galma1|62390 NIVE--KG-G -KFEPEQNRY HLYVSYACPW ATRTLIVRKL KGLDTIIPVT VVSPR----- MGPHGWPFAQ
Glotr1_1|112236 NQIQ--KG-G -KFEPEKDRY HLYVSYACPW ATRALIVRKL KGLENLIGVT VVSPR----- MGEHGWPFAN
Glotr1_1|120328 GVIS--PD-G -PFPPETGRY HLYIGLFCPF AHRANLVRHI KGLTDFIDVS IVKPYPKGDE KGWPGWRFPG
Glotr1_1|137290 NWVE--KG-G -KFAPEAGRY HLYVAYVCPW ATRTLIVRKL KGLHNIVSVT VASSR----- MDEHGWPFAN
Glotr1_1|37102 NWIE--KG-G -KFAPEADRY HLYVVYGCPW ATRTLIVRKL KGLEDIIPVT AASPR----- ADEHGWPFAK
Gymlu1|36123 NFIE--KG-G -QFEPEKNRY HLYVSLACPW ATRTLIVRKL KGLEDLISVT VVSPR----- MGTNGWPFAS
Hebcy2|440337 NTIE--KG-G -KFEPATDRY HLYVSYACPW ATRTLIVRKL KGLEPFIPVT AVSPR----- MGANGWPFAN
Hetan2|39076 SFVE--KG-G -QFPPENNRY HLYVAYGCPF ASRTLIVRKL KGLEDIISVT VCSPR----- FGPNGWPFAS
Hetan2|442680 SAVE--KG-G -QFPPEKDRY HLYVSYACPW ATRTLIMRKL KGLEDIVPVT VVSPH----- MGELGWPFAN
Hydpi2|96315 NWIQ--KD-G -EFTAEKDRY HLYVSYACPW ATRALITRKL KGLEDIIPVT VVSPH----- MGDNGWPFAS
Hypsu1|37440 NTIE--QG-G -KFAPETDRY HLYVSLACPW ATRTLIVRKL KGLDSLIPVT VVSPH----- MGAQGWPFAS
Hypsu1|40330 DVIE--KG-G -RFTPEEGRY HLYVAFSCPW ATRTLIVRKL KGLDSIIPVT IVSPR----- IGPNGWAFGS
Jaaar1|142100 NTIE--KG-G -KFEPEGDRY HLYVAYACPW ATRTLIVRKL KGLEALISVS VVSPH----- MGSNGWPFAS
Jaaar1|63226 NTIE--KG-G -KFEPEGDRY HLYVSYACPW ATRTLIVWKL KGLEALISMS VVSPH----- MGSSGWPFAS
Lacbi2|648000 NTIE--RG-G -RFEPERDRY YLYVSYACPW ATRTLIVRKL KRLDEFIPVT VVSPR----- MGTHGWPFAN
Mellp1|35938 DQIS--NE-G -KFKPEKGRY HIYVSLACPW AHRVLIFLQL KGLKDLVGVS VVHPF----- MGTEGWSWNA
Oidma1|124122 NFIK--KG-G -QFPPEKDRY HLYVSYACPW AHRTLIVRRL KGLEDIIPFT SVHWH----- MQELGWRFAT
Oidma1|165197 NHISSKPG-S -PFPPESGRY HLYVSYACPW AHRTLIMRNL KGLQSIIDFT SVHWYM---D LGGIGWRFAK
Oidma1|182545 NWVSADPE-A -EFPAESGRY VLYLNLGCPW AHRTNIVHKL KGLEDVIQVV VMGYV----- LTPEGWIYDG
Oidma1|39468 SWVSTEPG-A -RFPAEKDRY VLYMNRGCPW ASRANLVRTL KGLEDVIQMV VLDWE----- LFPEGWSFTG
Oidma1|46504 NFVSADPD-S -EFPAEANRY VLYVHYACPW AHRILIVRVL KALEDIIEVV ELD-K----- ADEIGWAFSG
Paxin1|138020 NSIQ--KD-G -EFTAEKDRY HLYVSYSCPW ATRTLITRKL KGLEDIIPVT AVSPH----- MGENGWPFAS
Paxru1|825741 NWIQ--KD-G -EFAAEKDRY HLYVSYACPW ATRTLITRKL KGLEDIIPVT VVSPH----- MGENGWPFAS
Phybl2|126467 NHIEADKN-A -TFAAEKDRY HLYVSWACPW AHRTAIVRKL KGLEDIISIS AVHYL----- LGEKGWKFAS
Picst3|43969 DFISNKPG-S -KFLAEANRY HLYVSFACPW AHRTLITRVL KGLSSVISVS VVHWH----- MDDKGWRFIN
Pilcr1|4571 SFIE--KG-G -QFEAEKDRY HLFVSYACPW ATRTLIVRKL KGLEDLIPVT IVSPH----- MGTEGWPFAN
Pilcr1|827566 SFIE--KE-G -QFEAERDRY HLYVSYACPW ATRTLIVRKL KGLEDLIPVT IVSPH----- MGTEGWPFAN
Pirin1|77935 NWIE--KG-G -KYEPEAGRY HLIVSYACPW AHRTLITRKL KGLEDIVSLS VVSPH----- MGSNGWPFKN
Pisti1|121423 NWIHP-KD-T -QFPAERDRY HLYVSYACPW ATRTLITRKL KGLEDIIPVT VVSPH----- MGENGWPFAS
PleosPC15_2|48940 NVIE--KG-G -RFEPEKDRY HLYVSYACPW ATRTLIVRKL KGLEDFIPVT VVSPR----- MGSQGWPFAN
Plicr1|42268 NFLE--KG-G -QFEPERGRY HLYVSYTCPY ATRTLIARKL KGLEDIIPIT VLSPK----- VGKNGFPFAN
Plicr1|43753 DFIA--KG-N -QFEPEKERY HLYVSYACPW ATRTLIVRKL KGLEDIIPVT IVSPR----- MGSLGWAFAN
Plicr1|43758 DFIA--KG-N -KFEPEKERY HLYVSYACPW ATRTLVVRKL KGLEDIIPVT VVSPR----- MGSLGWPFAN
Punst1|118707 DWIV--PGSG -KFEPEKGRY HLYVSLACPW AHRTLIVRKL KGLEEIIPIH VVSPH----- LGEHGWPFAS
Punst1|52675 DFIE--KG-G -KFEPEKGRY HLYVSYACPW ATRTIIVRKL KGLEDFIPVH VVSPR----- MGEHGWPFAN
Punst1|55793 SIIA--PD-S -EFPPEAGRY HLYIGLFCPF AHRANLVRHL KGLTSLIDVS VVKPYPKGDE KGWPGWCFPK
Punst1|62356 ARLPRIQG-P -HSSFVAGRY ILYVSYGCPW ASRTLIVRKL KGLEDVIDLA VVYYH----- LGENGWEFRP
Schco3|2086920 SIIE--KG-G -QFPPERDRY HLYVSYSCPW ATRTLIVRKL KGLEDIISFT PVVPR----- LGKQGWPFYN
Schco3|2516063 NFIE--KG-G -KFAPEHDRY HLYVSYACPW ATRALIMRKL KGLEDIIPFS AVSPR----- MGTNGWPFRR
Schco3|2615640 NFIE--KG-G -KFAPERDRY HLYVSYACPW ATRALIVRKL KGLEDIIPFT AVSPR----- MGKDGWPFRR
Sclci1|100395 NWIQS-KD-G -QFPAEKDRY HLYVSYACPW ATRTLITRKL KGLEDIIPVT IVSPR----- MGENGWPFAS
Sebve1|329667 DVIE--KG-G -RYEPEKGRY HLYVSYACPW AHRTLITRKI KGLEDYVDVT VVSPH----- MGNHGWPFAK
SerlaS7_9_2 HFIE--KG-G -KFEAERDRY HLYVSYSCPW ATRTLIVRKL KGLEDIIPVT VVSPH----- MGKDGWPFAS
Sphst1|178948 NHVT--KD-G -EFPAEAGRY HLYVSLACPW AHRTLILRKL KGLEDILPVS VVSPR----- MAENGWPFAN
Sphst1|178963 NFIE--KG-E -EFEPEAGRY HLYVSVACPW AHRALILRKL KGLEDIVCQS ---------- ----------
Sphst1|193280 NFIE--KG-G -EFEPEAGRY HLYVSVACPW AHRALILRKL KGLEDIVSLS IAAPR----- MHAKGWAFSN
Sphst1|232298 NFIE--KG-G -EFEPEAERY HLYVSIACPF AHRALILRKL KGLEDIISLS IAAPR----- MHAKGWAFSN
Stano2|8800 GVIK--EG-G -EFPPEKDRY HLYIGLFCPF AHRPNLVRHL KHLTSTLPIS VVRPYP---- KGEPGWRFD-
Suilu1|799805 NFIE--ND-G -KFTPEKGRY HLYVSYACPW ATRTLIARKL KGLEDFLAVS TVSPH----- MGAQGWPFKN
Treme1|57346 DVIG--PD-S -HFLPEKGRY HLYVSYACPW AHRTLIVRAL KGLEEYIDVS VVHPH----- MLEGGWHFVK
Trire2|112633 NFISKDPE-S -QFPAERGRY ALYLSPGCPW AHRTMIVRLL KGLDSIIDLY QVHFH----- MGSEGWYFSG
Trire2|56819 NWISRESG-A -QFPPEAGRY HLYVSYACPW AHRTLITRKL KGLDDIISFS VVHWH----- LGEKGWRFAT
Tubme1|220 DLVSREPN-A -KFPPERGRY HLYVSYACPW AHRTLITREL KGLQDIIGVS VVHWH----- LGENGWRFGT
Tubme1|6088 DAISKVPG-A -QFPPEKGRY HLYVSYACPW AHRTQIIREL KGLQDIISLS VVHWH----- MGSKGWRFAT
Tulca1|76207 SFIE--KG-G -QFPPEKDRY HLYVSKSCPW AGRTMIVRHL KGLESLIGLT ILTPY----- MTELGWAFKK
Tulca1|76222 NFIE--KG-G -EFPPEKGRY HLYISHACPW ANRTAIVRRM KGLESFIDVT VVSPY----- MGGLGWPFKK
Ustma1|1515 NTIS--KD-G -PHPPEKGRY ILYCALICPW ASRTLHTRAL KQLEDIIDVA IVHYS----- LTPNGWSFSV

```
Gto2p(ECM4)|Sacce1|39   MEKQLEDSED FLEHWHDVAG GIRTAKEDSS KSFAEIKNDS Q-----R--- -FMVDATNE- PH-YG---YK
Gto3p|Sacce1|4945       E----GDGKT NERHWFDIAG GISSVNLNTS TPVANIPNNA H-----R--- -LLVDGTDE- PH-YG---YK
GSTX1|Trave1|66368      V--------- ---------- ---------- --------D- A-----F--- ---PGADAD- PL-FG---AE
GSTX2|Trave1|50990      A--------- ---------- ---------- --------D- A-----Y--- ---PGADLD- PL-YD---VK
GSTX3|Trave1|73942      V--------- ---------- ---------- --------S- P-----F--- ---PGAEAD- PF-YN---AQ
GSTX1|Phchr2|2982482    V--------- ---------- ---------- --------D- P-----F--- ---PAADSD- PL-NN---AQ
GSTX2|Phchr2|2624184    V--------- ---------- ---------- --------D- P-----F--- ---PGADED- PL-HG---AR
Agabi_varbisH97_2|137   A--------- ---------- ---------- --------D- D-----Y--- ---PGAEVD- PI-HN---VK
Amamu1|78766            V--------- ---------- ---------- --------D- P-----F--- ---PSADVD- PL-YG---SE
Amath1|62909            I--------- ---------- ---------- --------D- N-----F--- ---PGAQVD- PV-RG---AN
Aspnid1|4489            S--------- ---------- ---------- --------G- R-----N--- ---GSAKKD- PL-YG---FT
Aspnid1|7897            P--------- ---------- ---------- --------D- E--D--I--- -PGANTTPD- PVHSS---YS
Aspnid1|8922            SN-------- ---------- ---------- --------D- E-----Y--- ---PGATVD- RL-FG---ED
Aurde3_1|115815         A--------- ---------- ---------- --------D- P-----F--- ---PGADRD- PH-EG---AQ
Aurde3_1|1240856        V--------- ---------- ---------- --------D- V-----F--- ---PGAETD- PN-EG---AR
Botbo1|118302           A--------- ---------- ---------- --------D- E-----Y--- ---PGAGTD- PL-YD---SK
Conpu1|60697            A--------- ---------- ---------- --------D- E-----Y--- ---PGATKD- SL-HGNE-VR
Conpu1|82579            V--------- ---------- ---------- --------H- D-----F--- ---PGATRD- SI-HGDE-VT
Copci1|17111            V--------- ---------- ---------- --------D- E-----F--- ---PGAETD- PL-YN---SQ
Crypa2|254574           A--------- ---------- ---------- --------E- DK-D--I--- -EGENVIPD- PLHED---FT
Dacsp1|67545            A--------- ---------- ---------- --------D- A-----F--- ---PGADSD- PN-EN---AK
Fomme1|130383           I--------- ---------- ---------- --------D- P-----F--- ---PGTDVD- TL-NH---AD
Fomme1|130388           I--------- ---------- ---------- --------D- P-----F--- ---PGAEVD- DL-NH---AD
Fomme1|143667           I--------- ---------- ---------- --------D- P-----F--- ---PAAGVD- PV-NH---AQ
Fompi3|1022961          V--------- ---------- ---------- --------D- A-----F--- ---PGADAD- PL-YG---AE
Fompi3|127425           I--------- ---------- ---------- --------D- P-----F--- ---PGADKD- PL-YG---AD
Galma1|62390            V--------- ---------- ---------- --------D- P-----F--- ---PAAGED- PI-YK---SE
Glotr1_1|112236         A--------- ---------- ---------- --------H- K-----F--- ---PDADVD- PL-YN---SD
Glotr1_1|120328         SA-------- ---------- ---------- --------D- E-----Y--- ---PRATVD- KL-FG---SE
Glotr1_1|137290         V--------- ---------- ---------- --------D- K-----L--- ---PGAEED- PF-YG---SE
Glotr1_1|37102          V--------- ---------- ---------- --------D- R-----F--- ---PDVEDD- PL-YG---SE
Gymlu1|36123            V--------- ---------- ---------- --------D- P-----F--- ---PGAEAD- PL-YN---SQ
Hebcy2|440337           V--------- ---------- ---------- --------D- P-----Y--- ---PAADED- QL-NH---SE
Hetan2|39076            V--------- ---------- ---------- --------E- P-----F--- ---PGAEVD- PL-YD---SK
Hetan2|442680           V--------- ---------- ---------- --------D- N-----F--- ---PGADVD- PL-YN---AQ
Hydpi2|96315            A--------- ---------- ---------- --------D- P-----Y--- ---PGADVD- PL-YN---SQ
Hypsu1|37440            V--------- ---------- ---------- --------D- K-----F--- ---SGADED- PL-YG---SA
Hypsu1|40330            V--------- ---------- ---------- --------D- R-----F--- ---EGAEED- PL-YQ---SA
Jaaar1|142100           V--------- ---------- ---------- --------D- Q-----F--- ---PCADID- PL-YN---SQ
Jaaar1|63226            V--------- ---------- ---------- --------D- Q-----F--- ---PCADID- PL-YN---SQ
Lacbi2|648000           A--------- ---------- ---------- --------D- P-----F--- ---PGADSD- PL-YQ---SE
Mellp1|35938            P--------- ---------- ---------- --------S- E-----F--N PSMTGVIKD- TL-YA---SE
Oidma1|124122           P--------- ---------- ---------- --------E- DR-D--V--- -PGENVEKD- PL-KGHEKFT
Oidma1|165197           A--------- ---------- ---------- --------D- EK----L--- -SGENVTPD- PV-NG---VE
Oidma1|182545           S--------- ---------- ---------- --------E- ---------- ---GSEPCD- PL-YG---FT
Oidma1|39468            R--------- ---------- ---------- --------D- ---------- ---GSDEID- PL-YG---FK
Oidma1|46504            T--------- ---------- ---------- ---------- ---------- ---TGPDKD- PR-YG---FT
Paxin1|138020           V--------- ---------- ---------- --------D- P-----Y--- ---PGADVD- PL-FN---SQ
Paxru1|825741           V--------- ---------- ---------- --------D- P-----Y--- ---PGADVD- PL-FN---SQ
Phybl2|126467           A--------- ---------- ---------- --------E- E-----C--- ---PGAIPD- TV-NN---AQ
Picst3|43969            D--------- ---------- ---------- --------E- ELKT--IDPK TDVSLGTVD- HL-YN---FK
Pilcr1|4571             V--------- ---------- ---------- --------D- E-----F--- ---PGAEVD- PL-YS---SE
Pilcr1|827566           V--------- ---------- ---------- --------D- E-----F--- ---PGAEVD- PL-YG---SE
Pirin1|77935            A--------- ---------- ---------- --------D- E-----F--- ---PGATDDG GI-MG---AQ
Pisti1|121423           V--------- ---------- ---------- --------D- P-----F--- ---PGADVD- PL-YN---SK
PleosPC15_2|48940       V--------- ---------- ---------- --------D- N-----F--- ---EASDED- PL-YK---SE
Plicr1|42268            A--------- ---------- ---------- --------D- EDSP--F--- ---PGAEDD- PL-YG---SK
Plicr1|43753            V--------- ---------- ---------- --------D- E-----F--- ---PGAEVD- PL-YS---SE
Plicr1|43758            V--------- ---------- ---------- --------D- E-----F--- ---PGANVD- PL-YN---SE
Punst1|118707           A--------- ---------- ---------- --------D- P-----Y--- ---PGADVD- PL-FD---AK
Punst1|52675            V--------- ---------- ---------- --------D- A-----Y--- ---PAADVD- PL-FN---AE
Punst1|55793            SE-------- ---------- ---------- --------D- E-----Y--- ---PGATPD- KL-FG---SE
Punst1|62356            E--------- ---------- ---------- ---------- ------V--- ---EGCTPD- PV-YG---AK
Schco3|2086920          EA-------- ---------- ---------- --------D- E-----Y--- TTHPGAEPD- PL-YG---SS
Schco3|2516063          N--------- ---------- ---------- --------D- NE-ESIF--- ---PAADPD- PL-HH---SA
Schco3|2615640          E--------- ---------- ---------- --------D- DE-ESIF--- ---PAADPD- PL-YQ---SA
Sclci1|100395           A--------- ---------- ---------- --------D- P-----F--- ---PGANVD- PL-YN---SQ
Sebve1|329667           A--------- ---------- ---------- --------D- A-----F--- ---PGAGDD- PV-IG---AE
SerlaS7_9_2             V--------- ---------- ---------- --------D- A-----Y--- ---PGADVD- PL-YN---AS
Sphst1|178948           V--------- ---------- ---------- --------D- E-----F--- ---PGATTD- PV-ND---SK
Sphst1|178963           ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sphst1|193280           V--------- ---------- ---------- --------D- N-----F--- ---PGATID- HV-NN---SE
Sphst1|232298           V--------- ---------- ---------- --------D- D-----F--- ---PGTTID- HV-NN---SE
Stano2|8800             ---------- ---------- ---------- --------E- T-----Y--- ---PNATPD- HL-FN---SR
Suilu1|799805           G--------- ---------- ---------- --------D- PS----W--- ---ADADDD- TE-HP-N-FQ
Treme1|57346            PSDFTSNPAP VSAHNF---- ---------- --------D- S-----F--- ---PGATSD- QL-NG---FH
Trire2|112633           E--------- ---------- ---------- --------G- ---------- ---GSLPED- PL-HG---FK
Trire2|56819            K--------- ---------- ---------- --------E- DK-D--V--- -PGENVIPD- PLHEG---FT
Tubme1|220              T--------- ---------- ---------- --------E- EAGE--Y--- -PEVAVGPE- AN-TG---AS
Tubme1|6088             E--------- ---------- ---------- --------E- EVAE--S--- -PEMTAAPE- PV-NG---AT
Tulca1|76207            A--------- ---------- ---------- --------D- ER-H--I--- ---DGVEED- PL-YG---SS
Tulca1|76222            A--------- ---------- ---------- --------A- DH-S--I--- ---EGAEDD- PL-YG---VS
Ustma1|1515             T--------- ---------- ---------- --------D- S-----V--- -----ETGD- PV-YG---FE
```


α2 β3 β4 α3 α4
Gto2p(ECM4)|Sacce1|39 RISDLYYKSD PQYSARFTVP VLWDLETQTI VNNESSEIIR ILNSSAFDEF VDDD--H-KK TD-----LVP
Gto3p|Sacce1|4945 RLSDFYFKTK PDYKGRFTVP VLWDLETCTI VNNESSDIIR IMNSAAFDEF VGEE--Y-RQ VR-----LVP
GSTX1|Trave1|66368 HVKDLYMRVN PDYEGRFTVP ILWDKKTSAI VNNESSEIIR MFNT-AFNDQ LPSD--K-AA ID-----LYP
GSTX2|Trave1|50990 HVSDLYLRAK PDYDGR--VP LLWDKKTATI MNNESSEIIR MFNS-AFDDQ LPAD--K-AA VD-----LYP
GSTX3|Trave1|73942 YVRDLYLRAD PKYEGRFTVP VLWDKKTETV VNNESSEIIR IFNT-AFNEF LPAD--K-AA IH-----LYP
GSTX1|Phchr2|2982482 HVKDLYLKVK PDYDGRFTVP VLWDKHTGTI VNNESSEIIR MFNT-AFNHL LPED--K-AK LD-----LYP
GSTX2|Phchr2|2624184 HLKDLYLRVE PEHAGRFTVP LLWDKQQDTI VNNESSDIIR MFNT-AFNGI LPQD--K-AA LD-----LYP
Agabi_varbisH97_2|137 HIKEIYLKIS PEYSGRFTVP LLYDKKLDTI VNNESSEIIR MFNS-AFNHL LPAD--K-AA LD-----FYP
Amamu1|78766 HIKDLYLRAD PNYEARFTVP VLWDKKTHTL VNNESSEIIR IFNS-AFNDL LPKE--K-AA LD-----FYP
Amath1|62909 HIKDLYLSVD PHYSGRFTVP VLWDKKLNTI VNTESSEIIR MFNT-AFNDL LPKD--K-AA LH-----YYP
Aspnid1|4489 TISQLYFKAD PEYKGRYTIP VLWDKKKETI VNNESSEIIR MFYT-EFDHL LPEE--R-RE VNQPGGGLYP
Aspnid1|7897 HLRDIYFANE PNYTGRFTVP VLYDKKTKRI VSNESSEIIR MFYY-EFDDL LPAQ--Y-QK VD-----LFP
Aspnid1|8922 YLHKIYFKAD SGYKGRYSVP VLWDRKTGTI VCNESLELLR WLPN-AFNDF IPSS--L-AS ID-----LYP
Aurde3_1|115815 HIKDLYLKAA PDYSGRFTVP VLWDTQDRTI VNNESSEIIR ILNT-AFNAL LPPA--H-AA LD-----LYP
Aurde3_1|1240856 YLKEIYFTAA PDYTGRFTVP VLWDKEKRTI VNNESSEIIR IFDT-AFNAL LPPD--R-AA LT-----LYP
Botbo1|118302 HVKDLYLRVS PNYTGRFTVP VLWDKKTQTI VNNESSEIIR IFNT-AFNDL IPAD--K-QK LD-----YYP
Conpu1|60697 HVKDLYFKVK PNYEGRFTVP VLWDKELHTI VSNESSEIIR MLNS-EFNEL LPKD--K-AA VD-----LYP
Conpu1|82579 HLRQLYFKAD PNYAGRYTVP TLWDKELETI VSNESADIIR MLNT-AFNDQ LPAD--K-AA ID-----IYP
Copci1|17111 YVKELYLRAN TAYTGRFSVP LLWDKKNQTI VNNESAEIIR IFNS-AFNEF LSED--K-AK LD-----LYP
Crypa2|254574 HLSEVYFESD PNYGGRFTVP ILYDKKQKII VNNESSEILR MFGT-VFDDQ LDEG--H-RN VV-----LYP
Dacsp1|67545 HVKDLYFVAQ PDYNMRFTVP VLWDKKLRTV VNNESSEIIR NFNT-AFNEF LPKE--K-AE LD-----YYP
Fomme1|130383 HIKDIYLKAD PEYNARYTVP ILWDKKLQTI VNNESSEIIR ILNA-SFNDL LPPE--K-SS ID-----LYP
Fomme1|130388 HVRDLYFKAQ PDFKDRFTVP VLWDKKLQTI VNNESSEIIR ILNS-SFNEL LPPE--K-AA ID-----LYP
Fomme1|143667 HIKDLYLKAD PNYNGRFTVP VLWDKKHQTI VNNESSEIIR IFNT-AFNEL VDAE--K-AK VD-----IYP
Fompi3|1022961 HIKDLYLRAA PNYEGRFSVP VLWDKKTHTV VNNESAEIIR MFNS-AFNHL LPAD--K-AK LD-----YYP
Fompi3|127425 HLKDVYLRVS PDYEGRFTVP LLFDKKTSSV VNNESSDIIR IFNS-SFNHL LPPE--K-AK LD-----LYP
Galma1|62390 HVKDLYIKAD PTYGGRFTVP VLWDKKLHTI VNNESSEIIR IFNT-AFNEF IPEE--K-AG LN-----FYP
Glotr1_1|112236 HVKDLYFRAE PEYGGRFTVP VLWDKKHETI VNNESSEIIR IFNT-AFNEF LPAD--K-AA ID-----IYP
Glotr1_1|120328 YLHEVYFRAD KEYKGRYSVP LLWDKKTGTI VNNESAELLR WLPT-AFNSL LDEE--H-AK ID-----LYP
Glotr1_1|137290 HIKDLYLRVD PDYTGRFTVP ILWDKKLETI VNNESSEIIR ILNN-AFNDL LPAD--K-AA LD-----LYP
Glotr1_1|37102 HIKDLYLKAD PDYSGRFTVP VLWDKKLETV VNNESSEIIR IFNS-AFNEI LPPDLEK-TK LD-----LHP
Gymlu1|36123 HIKDLYLSVN PDYSGRFTVP VLWDKKGKTI VNNESSEIIR ILNS-AFNHL LPEE--Y-AK VD-----LYP
Hebcy2|440337 HVKDLYLKVD PDYAGRFTVP ILWDKKHNTI VNNESSEIIR IFNT-AFNEF LPAD--K-AA LD-----YYP
Hetan2|39076 YVRDLYLRAD PNYAGKFTVP LLWDKKEHTI VNNESSEIIR ILNT-AFNAF LPEE--K-AA LD-----YYP
Hetan2|442680 HVKDLYLRAN PEYSGRFTVP VLWDKKQHTI VNNESSEIIR IFNS-AFNDT ISKE--K-AE LD-----FYP
Hydpi2|96315 HVRDLYYKAD PNYGGRFTVP VLWDKKKHTI VNNESSEIIR IFNS-AFNDM LPPE--K-AS VD-----IYP
Hypsu1|37440 YVKDLYFKAD PDYGGRFTVP VLWDKKLHTI VNNESSEIIR IFNT-AFNEF IPKE--K-AA LN-----YYP
Hypsu1|40330 HIKDLYFRAD PNYSGRFTVP VLWDKKLHTI VNNESSEIIR IFNT-AFNEF LPKE--K-AA LN-----YYP
Jaaar1|142100 HVKDLYLRCE PNYNGRFTVP VLWDKKHETI VNNESSEIIR IFNS-AFNHL IPKE--K-AE VD-----IYP
Jaaar1|63226 HVKDLYLRCE PNYNGRFTVP VLWDKKHETI VNNKSSEIIR IFNS-AFNHL IPKE--K-VE VD-----IYP
Lacbi2|648000 HVKDLYLKAE QNYGGRFTVP ILWDKKLQTI VNNESSEIIR MFNV-AFDEF IPAE--K-AA LD-----FYP
Mellp1|35938 KIKEIYFKAN HQYQGRFTVP VLWDKKEETI VNNESSEIIR MLNC-CFNEL IEEQ--Y-QK ID-----LYP
Oidma1|124122 HLRDIYFSAD PDYTGRFTVP TLYDTKTGQI VSNESSEIIR MFYS-EFDDL LPEK--Y-KN VD-----LFP
Oidma1|165197 TIRDLYLLAE PNYTGRFSVP VLWDKKTKTI VSNESSEIIR MFAT-EFNDL LDED--H-RS IN-----IIP
Oidma1|182545 KHSQLYHKAD PTYAGRFTVP ALWDRKRETI VNNESSEIIR MFYT-AFDAF IDPS--L-RE ISKP---LLP
Oidma1|39468 RLSELYFKAE PSFSGRYTVP VLWDKVHETI VSNESSEIIR MLYS-AFDAF LPAS--L-QE KNKPDGGLLP
Oidma1|46504 HLRGLYLKAF PGWNGRSSVP VLWDTKKETI VNNNASEILE MVYT-EFDAL IPEE--K-RE ATKGEAGFLP
Paxin1|138020 HVKDLYYKAD PNYAGRFTVP VLWDKKKNTV VNNESSEIIR MFNT-AFNDI LPPE--K-AS ID-----IYP
Paxru1|825741 HVKDLYYTAD PNYAGRFTVP VLWDKKKNTV VNNESSEIIR IFNT-AFNDI LPPG--K-AS ID-----IYP
Phybl2|126467 FIREIYFKAN PDYEGRFTVP VLWDKKHNTI VNNESSEIIR MFNT-AFNHF LPAE--K-AK LN-----YYP
Picst3|43969 RIRELYFKAE PDYVGRFTVP VLWDKKLETI VNNESSEIIR MLNL-EFNEL ATKE--G-AA ID-----IYP
Pilcr1|4571 HIKDLYFKAD PNYSGRFSVP VLWDKKNHTI VNNESSEIIR MFNT-AFNDL IPKD--K-AA VD-----IYP
Pilcr1|827566 HIKDLYLKAD PNYTGRFSVP VLWDKKNHTI VNNESSEIIR MFNA-AFNDL IPTD--K-AA VD-----IYP
Pirin1|77935 HVKDIYLKVQ PEYDGKFTVP VLFDTKSQSI VNNESSEILR MFNT-AFNDM IAPE--Q-AS LD-----LYP
Pisti1|121423 HVKDLYFKAD PNYGGRFTVP VLWDKKQNTV VNNESSEIIR MLNS-AFNDM LPPE--K-AT VD-----IYP
PleosPC15_2|48940 HVKDLYLKAD PDYGGRFTVP VLWDKKNHTI VNNESSEIIR IFNH-AFDEF IPKD--K-AE ID-----YYP
Plicr1|42268 YLRDLYLKAN PDYSGRPTIP VLWDKETHTI VNNDSAEIVR IFNS-AFNDA VAPE--A-AE LD-----LYP
Plicr1|43753 HVRDLYFKAD PDYGGRFTVP VLWDKKNHTI VNNESSEIIR MLND-GFNDD ISAE--H-AA LD-----LYP
Plicr1|43758 HVRDLYFKAD PDYGGRFTVP ILWDKKNHTI VNNESSEIIR MLNH-GFNEF IPAE--Q-AA LD-----FYP
Punst1|118707 HMKDLYLRVD PQYQARCSVP MLWDTKLSTI VNNESSEIIR MLYT-AFDAL LPES--V-AS LD-----LYP
Punst1|52675 HVKDIYLRVD PNYDGRFTVP ILFDTKHSTI VNNESSEIIR MFYT-AFDHL LPPE--Y-AN LD-----LYP
Punst1|55793 YLHEVYFRSN EQYKGRYSVP VLWDTKKGVI VNNESPEILR FLQT-AFDNL LPEP--Y-RS VS-----FYP
Punst1|62356 TIKEIYLRAD PEYALRLTVP VLFDKTLGTI VNNESSEIIR MLNT-DLDEF VTDE--H-KG VT-----YYP
Schco3|2086920 HVKDLYFRAQ PDYNGRFSVP ILWDKKQHTI VNNESSEIIR IFNT-AFNEL LAPE--Y-TS VD-----LYP
Schco3|2516063 HVKDLYFRAD PEYSGRFTVP VLWDKKLHTI VNNESSEIIR IFNE-GFNEL LPKE--K-AG LD-----FYP
Schco3|2615640 HVKDLYFRAD PNFSGRFTVP VLWDKKLHTI VNNESSEIIR IFNS-EFNDL LPPE--K-AA LD-----FYP
Sclci1|100395 HLKELYFKAD PDYSGRFTVP VLWDKKQHTI VNNESSEIIR IFNS-SFNDL LPSA--N-AA VD-----IYP
Sebve1|329667 HIKDLYLKVQ PDYDGKFTVP LLFDKKLNKA VNNESSEILR IFNT-AFNHL IPED--K-AA ID-----LYP
SerlaS7_9_2 HVKDLYFKAD PDYTGRFTVP VLWDKKNHTI VNNESSEIIR VFNT-AFNDI IPAD--K-AA ID-----IYP
Sphst1|178948 HVRDLYFKAD PNYGGRFTVP VLWDKKKNTV VNNESSEIIR IFNS-EFNHI LPSD--K-AQ ID-----LYP
Sphst1|178963 ---------- PVFR-KFTVP VLWDKKKETI VNNESSEIIR IFNS-GFNDI LPTE--K-AA LN-----FYP
Sphst1|193280 YIREVYFKAE PNYSGRFTVP VLWDKKKETI VNNESSEIIR IFNS-GFNDI LPTE--K-AA LN-----FYP
Sphst1|232298 YIREVYFKAE PNYSGRFTVP VLWDKKKETI VNNESAEIIR IFNS-GFNDI LPAE--K-AA VD-----LYP
Stano2|8800 FMHQLYFRDD PAYTGRYSVP LLWDKKTNRI VNNESAEMLK WLPH-AFDST IEDQ--KIRE ID-----FYP
Suilu1|799805 HLKELYLLAD PNYGGRFTVP VLWDKKLATI VSNESSEIIR MFNT-AFNHL LPAD--K-AA ID-----IYP
Treme1|57346 YLSEIYYVAS PEYKARFTVP VVWDTKEKTI VNNESSEVIR FFNT-GFNQS LSGK--E-KE LD-----LYP
Trire2|112633 KLKDLYLKAD PNYTGRYTVP VLWDKKADVL VNNESSEIIR MLYS-EFDEF LPEN--L-RE DNRPGKGLYP
Trire2|56819 HLRDVYFESD KEYAGRFTVP VLYDKKTRKI VSNESSEIIR MLGT-EFDDL VEEK--Y-RS VV-----LYP
Tubme1|220 FLRDIYFKAQ PDYDGRYTVP TFWDKKNNTI VSNESSEIIR FLYT-EFDDL VDES--K-KG IT-----YYP
Tubme1|6088 YLQDIYFKAD PQYGGRYTVP ILWDKKTSTI VNNESSEIIR LLYT-EFDDL VDEK--F-RG NL-----FYP
Tulca1|76207 HLRELYFKAN PSYEGRFTVP ALWDKKTETI VNNESSEIIR IFNN-AFNDL LPED--K-AT LD-----LYP
Tulca1|76222 HIRELYFKAN PEYSGRFTVP MLWDKKAETI VNNESAEIIR MFNT-AFNDQ LPAD--K-AA VD-----IYP
Ustma1|1515 NTKQLYLKAD PSYTARYTVP ILWDKKLGTI VNNESSEIIR ILYT-AFDDL LPDDSPA-KG KT-----YYP


α4
α5
Gto2p(ECM4)|Sacce1|39 A---QLKTQI DDFNSWVYDS INNGVYKTGF AEKAEVYESE VNNVFEHLDK VEKILS-DKY --S--KLKAK
Gto3p|Sacce1|4945 R---SLEAQI TEFNSWVYDK INNGVYKAGF AECAEVYERE VTSLFQYLDK LENLLD---K --KYTDLEAE
GSTX1|Trave1|66368 E---ALRGEI DGINEWVYDT VNNGVYKSGF ATSQKAYEAA VVPLFESLDR LEKILT---G --K-------
GSTX2|Trave1|50990 E---ALRKEI DEINEWVYDT VNNGVYKSGF ATSQTAYEAA VVPLFESLDR LEKILT---E --K-------
GSTX3|Trave1|73942 E---ALKSEI DEINEWVYDT VNNGVYKAGF ATTQQAYEAA VIPLFESLDR LEKILT---G --K-------
GSTX1|Phchr2|2982482 E---SLRAKI DEVNDWVYDT VNNGVYKSGF ASTQKAYEAA VIPLFESLDR LEKMLE---G --Q-------
GSTX2|Phchr2|2624184 E---SLRAKI DSLNDWVYNT VNNGVYKAGF AQSQKAYEDA VIPLFESLDR LEKILT---G --K-------
Agabi_varbisH97_2|137 E---KLRSDI DEVNSWVYPD INNGVYKAGF AATQEPYTKA VHELFKALDK AEGMLK---G --R-------
Amamu1|78766 E---HFRAEI DGLNAWIYDS INNGVYTAGF AQKQEAYEKA VTGVFEGLDK VEKLLT---G --K-------
Amath1|62909 E---ELRKEI DGLNEWIYDG INNGVYKAGF ATTQEAYEKA VVRVFEALDK VEKLLV---G --K-------
Aspnid1|4489 P---HLRGEI EAMNEWVYDR INNGVYKTGF AATQEAYEAS LHPLFEALDR IEGHLA---Q P-G-------
Aspnid1|7897 P---HLREQI DATNDWVYND VNNGVYKSGF ATTQEAYERN VTTLFASLDR IEKHLA---D --S-------
Aspnid1|8922 E---HLRTVI DGLHSWILPH LCTGVYKAGF ASDQDQYNKN VIPVFGALNK LEKIVF---A --N-------
Aurde3_1|115815 E---HLRAEI DALNEWVYSD VNNGVYRSGF AATQDAYEAA VTKLFAALDR LEGILA-G-G --K-------
Aurde3_1|1240856 A---ELRAQI DGLNEWVYTD VNNGVYRAGF AKTQGAYEEA VLKLFAALDR LEGLVK---G --K-------
Botbo1|118302 E---DLRKEI DELNSWVYDA VNNGVYKSGF ASTQEAYEAA VKPLFEALDK LESILT---G --K-------
Conpu1|60697 E---AHRKEI DEINSWVYDT VNNGVYKSGF AATQEAYEKA VLPLFESLDR LEKMLE---G --K-------
Conpu1|82579 E---VYRNEI DEINSWVYDT VNNGVYRSGF ATTQKAYEDA VVPLFKSLDR LEKILE---G --K-------
Copci1|17111 E---DLRAEI DSVNDWVYHG INNGVYSTGM AATQAAYEKN VHELFDSLDK VEKILQ---G --R-------
Crypa2|254574 A---ALRTQI DEAGEWHYDK INNGVYKSGF ATTQEAYDKN VTALFEALDR AEAHLA---A --G-------
Dacsp1|67545 Q---ELRKEI DELNDWVYNT VNNGVYKCGF ATTPEAYESN VGPLFQSLDR LEGILEKS-G --S-------
Fomme1|130383 E---KHRIEI DSLNEWVYDG INNGVYKAGF AGTQEAYEKA VTQVFASLDR VEPLFE---G --K-------
Fomme1|130388 E---RYRSEI DSLNEWVYDG LNNGVYKSGF AGTQEVYEKA VTQVFASLDR IEPLLK---G --K-------
Fomme1|143667 E---KWRKEI DEVNEWVYDT VNNGVYKSGF AMTQAAYEKS VIAVFESLDR IEKMLE---G K-K-------
Fompi3|1022961 E---HLRKEI DEVNDWVYDT VNNGVYKTGF ATTQSAYEKA LNPLFESLDR LEKMLR---G --K-------
Fompi3|127425 E---PLRNEI DELNAWVYDT VNNGVYKSGF AASQDAYEKA VYPLFESLDR LEKILD---G --K-------
Galma1|62390 E---NLRPEI DALNEWIYPN INNGVYRSGF ATTQEAYQKA VVEVFDALDK VEKILT---G --K-------
Glotr1_1|112236 E---KLRLQI DEVNAWVYDT VNNGVYKTGF ARTQEAYEKA LFPLFESLDR LEKMLE---G --K-------
Glotr1_1|120328 E---ALRAQI DSISDWMQRD LNTGVYKAGF ANSQQEYDKN VIAPFAALNK VEELIA---S --N-------
Glotr1_1|137290 A---NLRPEI DEVNDWVYHT VNNGVYKAGF AKTQDAYEAA VKPLFASLDR LEKMLD---G --K-------
Glotr1_1|37102 A---TLREEI DELNDWVYHT VNNGVYKAGF AKTQAAYEAA VKPLFGSLDR LEKIVD---A --R-------
Gymlu1|36123 E---HLKAKI DELNEWIYSY INNGVYRTGF ATSQEAYLTA LKQLFEALDK VEAILA---G D-K-------
Hebcy2|440337 E---KLRSQI DELNEWIYPS INNGVYRSGF ATTQEAYQKA VVELFDALDK VEKILD---G --Q-------
Hetan2|39076 P---ELRAEI DTVNEFIYDG INDGVYKAGF AKTSEAFNPA VIHLFETLDR VEALLA---G --K-------
Hetan2|442680 Q---HLRKEI DEVNEWVYDS INNGVYKSGF ATTFDAYRAA VVPLFESLDR VEVMLK---D --K-------
Hydpi2|96315 E---AHRQEI DSINEWVYDT VNNGVYKAGF AVSQEAYEKA VYPLFESLDR LEKMLT---G --K-------
Hypsu1|37440 E---EHRAEI DALNDLIYPN INNGVYRSGF ASTQEAYELA VADVFSALDK VEKVLE---G --K-------
Hypsu1|40330 E---DLKAQI DELNDWIYPN INNGVYRAGF AVAQEPYERA VNDVFAALDK VEKVLE---G --K-------
Jaaar1|142100 E---KLRGEI DGVNEWVYDT VNNGVYKSGF ATTQQAYESN VVPLFNSLDR LEKMLE---G --K-------
Jaaar1|63226 E---RLRAEI DGVNEWVYDT VNNGVYKSGF ATTQQAYESN VVPLFDSLDR LEKMLE---G --K-------
Lacbi2|648000 K---ELRTEI DAVNEWIYPN INNGVYRSGF ATTQAAYEKA VTELFDALDK VEKILT---G --K-------
Mellp1|35938 Q---NLQQEI DTMNEWIYDQ VNNGVYKSGF ATTQSAYESE VKPLFEALDR LEKVLD-G-G --Q-------
Oidma1|124122 A---DLRKEI EETNEWTYSD INNGVYRSGF ATSQEAYEKA VVQLFKSLDR VEKHLS---E V-K-------
Oidma1|165197 N---SLIASI DKMNQWIYDD INNGVYKCGI AKSQAAYEKN VTQLFDSLDR VESLLA---N --S-------
Oidma1|182545 I---DRLRDI ETMNEWVYDT VNNGVYKCGF ASTQEAYDTN IYRLFESLDR LESHLA---A --N-------
Oidma1|39468 P---HLLSEI DDMNSWVYDT INNGVYKAGF AANQSAHEEA VRAIFKSLDR VEDILA---K --S-------
Oidma1|46504 P---TLKNKL DTFRNFVYHN INDGVYKVGF ASTQKAYEYS LFPLFEALDK VEEQLA---D P-A-------
Paxin1|138020 E---AHRQDI DSINEWVYDT VNNGVYKAGF AGSQDAYEKA VYPLFDSLDR LEKLLT---G --K-------
Paxru1|825741 E---AYRQDI DSINAWVYDT VNNGVYKAGF AGSQDAYEKA VYPLFDSLDR LEKLLT---G --R-------
Phybl2|126467 E---ELQSDI DSTNEWVYDT VNNGVYKSGF ATTQTAYENH VFPLFKSLDR LEKILS---E --N-------
Picst3|43969 K---ELQTEI DDINSWIYDN INNGVYKSGF STKQEVYDKE VKNVFTHLDK VEEILK---K --NHAA----
Pilcr1|4571 A---AHRAEI DSLNEWVYDT VNNGVYKAGF SIKQQAYEAA VVPLFQSLDR LEKILT---G --K-------
Pilcr1|827566 A---AHHAEI DSINEWIYDT VSNGVYRAGF SIKQQAYEAA VVPLFRSLDR LEKILT---G --K-------
Pirin1|77935 E---NLRDEI EKSHEWVYPT INNGVYRSGF ATSQSAYEDA VTKLFDSLDR VEKIFE---S --QKSQ----
Pisti1|121423 E---ELQKQI DEVNSWVYDT VNNGVYKAGF AGTQEAYEAA VYPLFESLDR LEKMLT---G --K-------
PleosPC15_2|48940 T---ELRNEI DELNSWIYPN VNNGVYRSGF ATTQESYSKA VKEVFEALDK LETILD---G --K-------
Plicr1|42268 S---TYRAEI DELSAWMYEK LNAGVYKTGF ASTQELYSQA VAEVFSALDT LEERLR---G --KPEA----
Plicr1|43753 E---ALRPEI DSLNEWVYDA VNNGVYKCGF ATSAVAYETN VAKVFDALDR LEKILQ---G --K-------
Plicr1|43758 E---ALRPEI DSINEWVYDS VNNGVYKSGF ATSAAAYETN VAKVFDALDR LEKILQ---G --K-------
Punst1|118707 E---AHRKEI DDMNEWVYDT VNNGVYKAGF AGTQKAYEDA VLPLFASLDK LEQTLE---G --K-------
Punst1|52675 E---PLRKEI DEQNEWVYDT VNNGVYKSGF ATTSKAYEAA VYPLFESLDK LEKILE---G --K-------
Punst1|55793 E---HLRSAI DEISTWMQRD LNTGVYKAGF SSTQEGYDRN VVPVFEALNK LEKLAY---E --N-------
Punst1|62356 E---NLRQDI DEINEWIYHK INNGVYKTGF ATTQEAYDEN LYPLFDALDR VEKVLE---G --R-------
Schco3|2086920 S---ELRPKI DELNEWGYPQ INNGVYRAGF ATSQEAYEKA VLGLFDGWDK VEKILS---T --N-------
Schco3|2516063 E---ELRKGI DEVNEWVYSQ INNGVYRTGF ATTQEAYEKA VNEVFDGLDK VEKILE---K --S-------
Schco3|2615640 T---ELRSKI DEVNEWVYHQ INNGVYRTGF ATTQSAYEKA VNEVFDGLDK VEKILS---K --S-------
Sclci1|100395 E---ELRKEI DEVNSWVYDT VNNGVYKAGF AVSQEAYEKA VYSVFESLDR LEEMLT---G --K-------
Sebve1|329667 E---ALRGEI EASHEWIYPT VNNGVYRSGF ATSQGAYETA VKALFDSLDR IEKMFE---S --QKK-----
SerlaS7_9_2 E---AHRAEI DSINEWVYDT VNNGVYKSGF AQSQEAYEKA VVPLFNSLDR LEKILT---G --K-------
Sphst1|178948 E---DLRKDI DELNAWVYDT VNNGVYKSGF AQTQEAYEAA VKPLFASLDR LEKILG---G --K-------
Sphst1|178963 K---ELQKEI DELNEWVYET VNNGVYKTGF AKTQEAYESA IRPLFDSLDR LEKILY---G --K-------
Sphst1|193280 K---ELQKEI DELNEWVYET VNNGVYKTGF AKTQEAYESA IRPLFDSLDR LEKILY---G --K-------
Sphst1|232298 K---ELQKDI DELNEWVYET VNNGVYKAGF AKTQEAYDSG IRPLFDSLDR LEKILH---G --K-------
Stano2|8800 Q---DLRSKI DEISLWLTSL ICSGVYKAGF VSTQEEYEKN VVPLFAALNK LEDLAH---K --N-------
Suilu1|799805 E---AHRQEI DSINEWVYDT VNNGVYKSGF AQSQAAYENA VYPLFKSLDR LEEMLK---G --K-------
Treme1|57346 E---ELRGEI DELNEWVYND VNNGVYKCGF AATQEAYEKA LVPLRDALER LDKLFS-D-G --R-------
Trire2|112633 P---HLRAEI DAMNEWVYDT VNNGVYKVGF SKSQESYDEN IRPLFASLDR LEEHLG---H ----------
Trire2|56819 E---NLRSQI DETNEWTYNL INNGVYKSGF ATTSAAYEKA VVALFEALDR AEEHLA---S VAD-------
Tubme1|220 E---DLRKEI DELNEWIYAT VNNGVYKAGF ATKQEAYEVA VTKLFESLDR IEKILK---D --S-------
Tubme1|6088 E---DKRGAV DELNSWVYDT INNGVYKSGI ATKQGAYEDA VTALFASLDR VEGLLK---S --S-------
Tulca1|76207 E---ELRADI DGVNEWVFST VNAGVYKVGF STGQLGYEEA IKPLFASLDR LEGLLK---D --K-------
Tulca1|76222 K---DLQAEI DSTNEWVYDT VNNGVYKSGF ATTQSAYEAA VKPLFASLDR LEGMLN---G --K-------
Ustma1|1515 TDNVEAIQKI DELNEWIYSD INNGVYKTGF ATAQEVYDKE VDNLFLALDR LEKLLS-D-G --R-------


α5 α5' α6 α6' α7
Gto2p(ECM4)|Sacce1|39 YGEEDRQKIL GEFFTVG--- DQLTEADIRL Y----TTVIR FDPVYVQHFK CNFTSIRAGY PFIHLWVRNL
Gto3p|Sacce1|4945 YGKNNKDKIL DRYFAIG--- DTLTEADVRL Y----PTIVR FDVVYHQHFK CNLATIRDDY SRIHTWLKNI
GSTX1|Trave1|66368 ---------- --DYLVG--- DQLTEADIRL F----VTIVR FDPVYVGHFK CNIRTIRDGY PAIDLWMRKL
GSTX2|Trave1|50990 ---------- --DYLVG--- DRLTEADIRL F----VTIVR FDPVYVGHFK GNIRTIRDGY PAIHLWLRKL
GSTX3|Trave1|73942 ---------- --DYLVG--- DQLTEADVRL F----VTIIR FDPAYVGHFK CNLRTIRDGY PAIHLWLRKL
GSTX1|Phchr2|2982482 ---------- --DYLIG--- GQLTEADIRL F----VTIVR FDPVYVTHFK CNLRTIRDGY PNLHRWMRKL
GSTX2|Phchr2|2624184 ---------- --DYLVG--- DQLTEADVRL F----VTIVR FDVGYHTAFK CNIRSIRADY PALHRWLRRL
Agabi_varbisH97_2|137 ---------- --DYLVG--- NQLTEADVRL W----VTTVR FDPVYVGHFK CNISTIRGGY PNINRWMRKL
Amamu1|78766 ---------- --EYLVG--- NQLTEADIRL W----VTIVR FDPVYVGHFK CNIRDIRHGY PAIHKWLRQL
Amath1|62909 ---------- --DYLVG--- DRLTEADIRL W----VTIIR FDPVYVGHFK CNFRTIRDGY PAIHKWMQKL
Aspnid1|4489 ---------- HQPYLFG--- EHITEADVRL Y----TTIAR FDVAYYLIFK CNLKMIRHDY PLIDAWYRRL
Aspnid1|7897 ---------- KSAYFFG--- DDITEADIRL F----TTIVR FDPVYVQHFK CNIRDIRSGY PAIHAWLRRL
Aspnid1|8922 ---------- GGPFILG--- YQMTEVDVTV Y----ATLVR FDTVYVQHFK CNLGTVRHDY PVLNNWLKAM
Aurde3_1|115815 ---------- --KYLVG--- DLLTEADVRL F----VTLIR FDPVYVGHFK CNLRTIRDGY PELNRWMKTL
Aurde3_1|1240856 ---------- --DFLVG--- DQLTEADVRL F----VTLIR FDPAYVGQFK CNLRTIRGGY PELNRYMKNL
Botbo1|118302 ---------- --EYLIG--- NRLTEADVRL Y----TTIVR FDPVYHGHFK CNLGSIRHDY PAINKWLKNL
Conpu1|60697 ---------- --DYLVG--- NQLSEADVRL W----VTIIR FDPVYFAHFK CNIRDIRHGY PNLNRWMKNL
Conpu1|82579 ---------- --DYLVG--- DRLTEADIRL W----VTIIR FDPVYFSHFK CNIRDIRHDY PNLNRWLKNL
Copci1|17111 ---------- --DYLVG--- DRLTEADIRL W----VTIIR FDPVYVGQFK CNIRTIRDGY PAIHSWLRKL
Crypa2|254574 ---------- EGPYYLG--- KELTEVDIRL F----VTIIR FDPVYVQHFK CNIKDIRSGY PHIHKWMRNL
Dacsp1|67545 ---------- --DYLIG--- NRLTEADIRL Y----TTLVR FDPVYVGHFK CNIGTIRHNY PHLNRWMKNL
Fomme1|130383 ---------- --DFLIG--- NQLTEADVRL Y----TTAVR FDVAYHGIFK CNLRTIREGY PNIHRWLRQL
Fomme1|130388 ---------- --DFLIG--- DQLTEADVRL Y----VTTVR FDVAYHGQFK CNLRTIRSGY PNIHRWLRQL
Fomme1|143667 ---------- --EYLVG--- DQLTEADIRL F----VTIIR FDPVYVGHFK CNFRTIRSGY PNIHNWLRKL
Fompi3|1022961 ---------- --EYLIG--- DRLTEADIRL F----VTIIR FDPVYVGHFK CNIRTIRDGY PAIHLWLRKL
Fompi3|127425 ---------- --DYLIG--- SQLTEADVRL F----VTIIR FDVAYHGNFK CNIRNIRNGY PSIHLWLRKL
Galma1|62390 ---------- --DYLVG--- EKLTEADIRL W----VTIIR FDPVYVGHFK CNIRTIRDGY PAIHKWMQKL
Glotr1_1|112236 ---------- --QYLIG--- DVLTEADIRL F----VTIIR FDPVYVGHFK CNIRTIRGGY PAINKWMKNL
Glotr1_1|120328 ---------- SGPYVLG--- KTLTELDIRL Y----ATVVR FDTVYVQHFK CNIGTIRHNY PVINNWLKNL
Glotr1_1|137290 ---------- --DYLIG--- GQLTEADVRL F----VTIIR FDVAYFGVFK CNVRSIRHDY PALHKWVRKL
Glotr1_1|37102 ---------- --DYLIG--- GQLTEADVRL F----VTIVR FDIGYYGAFK CNIRTIRDGY PALNRWFKKL
Gymlu1|36123 ---------- --KYIVG--- DQLTECDVRL W----VTIIR FDPVYVGHFK CNVRDIRGGY PNINKWMKNL
Hebcy2|440337 ---------- --DYLIG--- NTLTEADVRL W----VTLIR FDPVYVGHFK CNLRTIRDGY PITHRWMQKL
Hetan2|39076 ---------- --TYLVG--- ERLTEADVRL F----VTIIR FDTAYVGHFK CNLRTIRGGY PNIHLWLRRL
Hetan2|442680 ---------- --TYFIG--- DRLTEADIRL F----VTLIR FDPVYVGHFK CNLRTIRHGY PNIHLWLRKL
Hydpi2|96315 ---------- --DYLVG--- NTLTEADIRL F----VTIVR FDPVYVGHFK CNLRTIRNGY PAIHTWLKKL
Hypsu1|37440 ---------- --DYLVG--- NTLTEADIRL W----VTTIR FDPVYVGHFK CNFRTIRDGY PAIHRWMQKL
Hypsu1|40330 ---------- --DYLVG--- NTLTEADIRL W----VTIIR FDPVYHGLFK CNFRTIRSGY PAIHSWLHKL
Jaaar1|142100 ---------- --EFLVG--- DQLTEADVRL W----VTIIR FDPVYVGHFK CNIRTIRGGY PAINKWMKKL
Jaaar1|63226 ---------- --STSLA--- ---TNSPKLM FDFGLPSFIR FDPVYVGHFK CNIRTIRGRY AAINKWMKKL
Lacbi2|648000 ---------- --DFLVG--- DRLTEADIRL W----VTIIR FDPVYVGHFK CNIRTIRDGY PAIHKWMQKL
Mellp1|35938 ---------- --KYLMG--- DQLTEVDVRL F----TTLIR FDVVYVTHFK CNLKTIRGGY PNLNHWLKCL
Oidma1|124122 ---------- DGPYYFG--- RNITEADVRL Y----TTIVR FDVVYVQHFK TNVRDIRSGY PRIHKWLRNL
Oidma1|165197 ---------- TGPYILG--- NELTEVDLRL Y----PTIVR FDVVYVTLFK TNLDTIRSGF PAIHKWLRHL
Oidma1|182545 ---------- KTPYLFG--- DRITEADIRL F----PTIVR FDVAYYTLFK CNLKMIRYDY PRLHGWLRNL
Oidma1|39468 ---------- TGPYLFG--- SHVTEADIRL Y----PTIAR FDVAYYALFM CNLKMIRHDY PEIQKWFLRL
Oidma1|46504 ---------- HQPYLFG--- KNITDADIRL F----STLIR FDVSYFSSFK CNLKMIRYEY PRLHLWLRTL
Paxin1|138020 ---------- --DYLVG--- DKLTEADIRL F----VTLIR FDPVYVGHFK CNLRTIRNGY PALHLWMRKL
Paxru1|825741 ---------- --DYLVG--- DKLTEADIRL F----VTIIR FDPVYVGHFK CNLRTIRNGY PALHLWMRKL
Phybl2|126467 ---------- --KFLVK--- NTFTEADIRL W----TTIIR FDPVYHGHFK CNIKSIQHDY PHILAWARRI
Picst3|43969 ---------- DKPYLLG--- NTLTEADVRL F----TTIIR FDPVYVQHFK CNIGMIRHDY PHIHQWVREL
Pilcr1|4571 ---------- --DYLVG--- DVLTEADIRL W----VTIIR FDPVYVGHFK CNLRTIRNGY PAIHSWMRKL
Pilcr1|827566 ---------- --DYLVG--- DVLTEADVRL W----VTIIR FDPVYVSHFK CNLRTIRKGY PAIHSWMKKL
Pirin1|77935 -----NANSK GEYFLIS--- DRLTESDIRL Y----TTIVR FDPVYHGHFK CNYRTIRDGY PLIHKWLQNL
Pisti1|121423 ---------- --DYLVG--- DRLTEADIRL F----VTVVR FDPVYFSHFK CSLRDIRHGY PAINRWMKNL
PleosPC15_2|48940 ---------- --DYLVG--- NRLTEADIRL W----VTIIR FDPVYVGHFK CNIRTIRGGY PNINTWLKKL
Plicr1|42268 --------SQ GRKYLVG--- DSLTEADIKL W----VTLIR FDPAYVGQFK CNIRTIRDGY PALNSWMKAL
Plicr1|43753 ---------- --DYLVG--- GTLTEADVRL W----VSIIR FDVVYVGHFK CNIRTIRHGY PAIHSWLKKL
Plicr1|43758 ---------- --DYLVG--- GTLTEADIRL W----VTIIR FDPVYVGHFK CNIRTIRHGY PAIHSWMKKL
Punst1|118707 ---------- --DYLVG--- DRLTEADIRL Y----PTIIR FDVVYATHFK CNIRTIRSGY PNLHRWLRKL
Punst1|52675 ---------- --DYLVG--- DQLTEADIRL F----PTIIR FDPVYVRHFK CNIRDIRSGY PNLHRWMKNL
Punst1|55793 ---------- GGPYVLG--- KELTEVDLRL Y----PTLIR FDTVYVQHFK CNLGTIRHNY PVLNNWMKNL
Punst1|62356 ---------- --EFLVGPGK GTLTEADIRL Y----TTIIR FDVGYHGAFK CNLKTIRHDY PNLHRWVRNL
Schco3|2086920 ---------- --EYLLG--- DRLTEADVRL W----VTAIR FDPAYYGHFK CNIRSIRHGY PAINRWLKKL
Schco3|2516063 ---------- --EYLVG--- DKLTEADLRL W----VTIVR FDPVYVGHFK CNFRTIRAGY PAINEWLKKL
Schco3|2615640 ---------- --EYLVG--- DKLTEADIRL W----VTIIR FDPVYVGHFK CNFRTIRAGY PAINEWHKKL
Sclci1|100395 ---------- --DYLIG--- DRLTEADIRL F----VTIIR FDPVYHGHFK CNIRDIRHGY PAINRWMKKL
Sebve1|329667 -------NDT DEYFLIG--- NQLTESDIRL Y----TTIVR FDPVYHGHFK CNYRTIRDGY PLMHRWLQNL
SerlaS7_9_2 ---------- --DYLVG--- NTLTEADVRL F----VTIIR FDPVYVGHFK CNIRTIRHGY PAIHSWMKKL
Sphst1|178948 ---------- --QYLFG--- DRLTEADVRL Y----TTIVR FDPVYHGHFK CNFSTIRGGY PNLNRWLKNL
Sphst1|178963 ---------- --EYLVG--- DRLTEADVRL Y----TTIIR FDTAYHGIFK CNLRTIRDGY PNINKWLKNL
Sphst1|193280 ---------- --EYLVG--- DRLTEADVRL Y----TTIIR FDTAYHGIFK CNLRTIRDGY PNINKWLKNL
Sphst1|232298 ---------- --EYLVG--- DRLTEADVRL Y----TTIIR FDTAYHGIFK CNLRTIRDGY PNINKWLKNL
Stano2|8800 ---------- GGPYILG--- EKMTELDLLA Y----PTIVR FDTVYVQHFK TNLGMVRHDY PVLNNWLKNL
Suilu1|799805 ---------- --DYLVG--- NTLTEADIRL F----VTTIR FDPVYVGHFK CNLRDIRHGY PNIYLWTRKL
Treme1|57346 ---------- --EFLIG--- GRLTEADVRL Y----TTIVR YDPVYHGHFK CNWGSIRHDY PHINRWLKNL
Trire2|112633 ---------- GKPFLLG--- DSITEADVRL Y----TTLAR FDVAYHSVFQ CNLKSIRHDY PRLHGWLRRL
Trire2|56819 ---------- QGPYYFG--- GSLTEADIRL Y----VTLVR FDPVYVQHFK CNIRDIRSGY PRLHRWMRNL
Tubme1|220 ---------- DGPYLLG--- KVLTEADIRL Y----PTIAR FDPVYVQHFK CNLRMIRHDY PAIEKWMKHL
Tubme1|6088 ---------- AGPYLLG--- GKITEADVCL Y----PTIVR FDPVYVQHFK CNIGMIRHDY PAIHKWLRHL
Tulca1|76207 ---------- --EYLVG--- GRLTEADVRL F----TTLVR FDPVYHGHFK CNIGSIRHNY PNLNEWMKRL
Tulca1|76222 ---------- --DYLLG--- DRLTEADVRL F----TTIVR FDPVYHGHFK CNIGSIRHNY PNINRWMKQL
Ustma1|1515 ---------- --EYLVP--E VGLSDADIRL Y----PTLAR FDTAYHNLFR CNVGMIRHDY PHIHRYLRRL




```
                                    α7          α8          α9
                                    εεεε        εεεεε       εεεεεεεεεεεεε
Gto2p(ECM4)|Sacce1|39   YWNYD----- --AFRYTT-- DFDHIKLHYT RSHTRI---- NPLGITPLGP K--PDIR--- ----------
Gto3p|Sacce1|4945       YWRHE----- --AFQRTT-- DFTHIKLGYT RSQPRV---- NPIGITPLGP K--PDIR--- ----------
GSTX1|Trave1|66368      YWNNS----- --AFSSTC-- NFEHIKTHYY WSHPQI---- NPHRVVPVGP I--PNIL--- ----------
GSTX2|Trave1|50990      YWNNS----- --AFSSTC-- DFDHIKTHYY WSHPSI---- NPHRIIPVGP I--PNIL--- ----------
GSTX3|Trave1|73942      YWNNS----- --AFSETC-- KFDHIKASYY A-QKNV---- NPTLVVPLGP I--PNIL--- ----------
GSTX1|Phchr2|2982482    YWGNP----- --AFKDTC-- NFEHIKTHYF WSHTFI---- NPHRIVPIGP I--PDIL--- ----------
GSTX2|Phchr2|2624184    YWTIP----- --AFKDTC-- NFQHIKVGYF -GLTMI---- NPFKIVPVGP L--PDIL--- ----------
Agabi_varbisH97_2|137   YWNND----- --AFQSST-- NFDHIKTHYY WSHAHI---- NPTRIVPEGP V--PVIE--- ----------
Amamu1|78766            YWNHE----- --AFKSST-- NFEHIKVHYY WSQSNV---- NPTRIVAVGP I--PNIE--- ----------
Amath1|62909            YWNND----- --AFKSCT-- DFLHIKTHYY WSHGSI---- NGTRVVPLGP V--PHVL--- ----------
Aspnid1|4489            YYDESEKTR- GGAFGKTT-- FLGLYKFGYL AALGKKQSVD SSQLIVPAGP K--PDIL--- ----------
Aspnid1|7897            YWDVP----- --AFRETT-- QFEHIKKHYT KSHSQI---- NPFGITPVGP T--PDIL--- ----------
Aspnid1|8922            YWDVP----- --AAREST-- DFRHIKENYT KSHYDI---- NPKAITPLGP W--PDVEEGV ERDWTKLTVG
Aurde3_1|115815         YWTED----- --AFRSTA-- NFAHIKTHYY WSHPHI---- NPTRVVPVGP V--PDIQ--- ----------
Aurde3_1|1240856        YWNYD----- --AFKSST-- NFDHIKSGYY YSQVNV---- NPTRIVPVGP V--PHIE--- ----------
Botbo1|118302           YWKNA----- --AFKDTT-- NFDHIKTHYY WSHTQI---- NPKRIVPLGP K--IDIE--- ----------
Conpu1|60697            YYHNA----- --AFKDST-- NFDHIKTHYY WSQIKV---- NPTRIVPVGP S--EDIE--- ----------
Conpu1|82579            YWNNS----- --AFKEST-- FPEHIKAGYY -SGVKI---- NPTGIVPVGP N--PFIE--- ----------
Copci1|17111            YWNIP----- --AFKDST-- DFDHIKTTYY WSHASS---- SPLDVPSKSS TLHPQLEQKY --N-------
Crypa2|254574           YWNVP----- --AFKETT-- NFLHIKNHYT KSHTQI---- NPYSITPRGP L--PDIL--- ----------
Dacsp1|67545            YWKND----- --AFKGTT-- NFEHIKTHYY WSHTQI---- NPTRIVPVGP K--PNIE--- ----------
Fomme1|130383           YWKNP----- --AFNSTT-- NFDHIKTGYH -SIRTI---- NPNLIVPIGP I--PSIE--- ----------
Fomme1|130388           YWTNP----- --AFNSTT-- NFDHIKTGYY -SMKTI---- NPTQIVPVGP V--LSIE--- ----------
Fomme1|143667           YWTNP----- --AFSTTT-- NFDHIKTHYY WSHPHI---- NPTRIVPVGP I--PNIE--- ----------
Fompi3|1022961          YWKIP----- --AFKDTC-- NFEHIKTHYY WSHTTI---- NPHRIVPVGP I--PNIL--- ----------
Fompi3|127425           YWNVP----- --AFRDTC-- YFDHIKAGYY WAKAY----- NPHQIIPAGP V--PHIL--- ----------
Galma1|62390            YWNSS----- --AFKDST-- NFEHIKTHYY WSHISI---- NPTRIVPVGP I--PNIL--- ----------
Glotr1_1|112236         YWNHP----- --EFKDTT-- NFNHIKTHYY WSHPII---- NPTRIVPVGP I--PDID--- ----------
Glotr1_1|120328         YWNVP----- --AFRDTT-- DFLHIKENYT KSHYDI---- NPKAITPMGP F--PDVEEGV EEDFNKLKPG
Glotr1_1|137290         YWNHD----- --PFKSTA-- NFDHIKAGYY -SNKGV---- NPAGVVPVGP L--PPIE--- ----------
Glotr1_1|37102          YWTNP----- --AFKSST-- NLDHIKTGYY W-ITAV---- S--------- ---------- ----------
Gymlu1|36123            YWNNS----- --AFKDCT-- SFEHIKTHYY WSHPSI---- NPTRIVPVGP I--PMIQ--- ----------
Hebcy2|440337           YWNSA----- --AFKDST-- DFDHIKTHYY WSHHAI---- NPTRIVPVGP I--PNIL--- ----------
Hetan2|39076            YWRND----- --AFQSTT-- DFDHIKTFYF WGQPAI---- NPTRIVPIGP V--PHIE--- ----------
Hetan2|442680           YWQND----- --AFQSTT-- DFDHIKTHYY WSHPMV---- NPHRIVPIGP V--PHIE--- ----------
Hydpi2|96315            YWNND----- --AFKSST-- NFEHIKVHYY WSQTKV---- NPTRVIAVGP T--PDID--- ----------
Hypsu1|37440            YWNIS----- --AFKDST-- DFDHIKVHYY WSHQTI---- NPTRIVPVGP V--PNIL--- ----------
Hypsu1|40330            YWNNS----- --SFKEST-- NFEHIKVGYY WPQSS----- NPTRIVPVGP I--PSIL--- ----------
Jaaar1|142100           YWTNH----- --AFKDST-- NFDHIKTHYY WSHPQV---- NPTRIVPVGP I--PNVE--- ----------
Jaaar1|63226            YWTNP----- --AFKDST-- DFDHIKTSYY WSHPSV---- NPTHIIPVGP I--PNIE--- ----------
Lacbi2|648000           YWNND----- --AFESST-- NFDHIKTHYY WSHAAI---- NPTRIVPVGP I--PNIR--- ----------
Mellp1|35938            YWKNP----- --AFKDTT-- NFEHIKFHYF TSHPQI---- NPTRIIPIGP L--PEIE--- ----------
Oidma1|124122           YWNVP----- --AFKETT-- EFTHIKNHYT KSHTQI---- NKFSITPVGP E--PPIL--- ----------
Oidma1|165197           YWDLP----- --AFGQTT-- NFEHIKRHYF SSITTL---- NPEGIVPLGP R--PEIL--- ----------
Oidma1|182545           YWDTSEATN- GGAFHSTT-- HFTPIKAGYT YTLKQT---- ----VVPAGP Q--IEIL--- ----------
Oidma1|39468            YWDESGETR- -GAFKAST-- FFDAIKKGYA QT-------- PRKNFLPLGP D--PDVL--- ----------
Oidma1|46504            YWDPENKFN- -GAFRDTV-- RFQLYKEQY- -AYATV---- SP--VYPKGP I--PHIL--- ----------
Paxin1|138020           YWNND----- --AFKSAT-- NFQHIKVHYY WSQTKV---- NPTRVVAVGP T--PDIE--- ----------
Paxru1|825741           YWNND----- --AFKSTT-- NFQHIKVHYY WSQTKV---- NPTRVVAVGP T--PDIE--- ----------
Phybl2|126467           Y-QME----- --GVSDTV-- NMVHIKRHYY MSHLQI---- NPTGIVPVGN G--PDLT--- ------IV--
Picst3|43969            YWKVP----- --GFKETT-- DFDHIKYHYT KSHIAI---- NPHSITPAGP I--PNIL--- ----------
Pilcr1|4571             YWNNS----- --AFKDST-- NFEHIKTHYY WSHPSV---- NPTRVVPVGP I--PNVE--- ----------
Pilcr1|827566           YWNSS----- --AFKDST-- NFEHIKTHYY WSLYTI---- NPTRVIPIGP I--PHVE--- ----------
Pirin1|77935            YWNYP----- --AFKDTT-- DFEHIKTHYY WSHPHI---- NPTRVIPLGP I--PNIR--- ----------
Pisti1|121423           YWTNE----- --AFRSST-- NFQHIKVHYY WSQTKV---- NPTRVVSVGP V--PDIE--- ----------
PleosPC15_2|48940       YWNNE----- --AFKVST-- NFNHIKTHYY WSHTTI---- NPTRIVPVGP F--PDIE--- ----------
Plicr1|42268            YWTVP----- --AFKEST-- DFEHIKTGYY WSSVSN---- NPHRIVPVGP L--PPIE--- ----------
Plicr1|43753            YWNVP----- --AFREST-- DFGHIKTGYY WSQVNI---- NPTRIVPVGP I--PNIE--- ----------
Plicr1|43758            YWNVP----- --AFREST-- DFDHIKTHYY WSHPNI---- NPTRIVPVGP I--PNIE--- ----------
Punst1|118707           YWTVP----- --AFGETT-- NFDHIKVHYF WSHPMI---- NPHRIVPVGP L--PHIL--- ----------
Punst1|52675            YWKNP----- --AFKDTT-- NFDHIKTHYF WSHPSV---- NPHRIVPVGP I--PHIL--- ----------
Punst1|55793            YWNVR----- --GFKETT-- DFKHIKENYT KSHYDI---- NPKAITPMGP I--PDIEEGY EANWSKITPG
Punst1|62356            YWNYP----- --AFKDTT-- NFDHIKGGYY -ALKKV---- NPTGVIPRGP V--PHIL--- ----------
Schco3|2086920          YWTND----- --AFRSST-- RFDHIKTGYY SQN------- NPSGIVPLGP V--PDIE--- ----------
Schco3|2516063          YWTND----- --AFKSST-- RFDHIKTHYY WSHPHI---- NPTRIVPLGP I--PDIE--- ----------
Schco3|2615640          YWNNN----- --AFKSST-- RFDHIKTHYY WSHPHV---- NPTRIVPLGP V--PDIE--- ----------
Sclci1|100395           YWTNP----- --AFQSTT-- NFQHIKVHYY WSQTKI---- NPTRVVAVGP N--PDIE--- ----------
Sebve1|329667           YWKHP----- --EFKDTT-- DFDHIKTHYY WSHPHI---- NPTRVIPLGP I--PNIR--- ----------
SerlaS7_9_2             YWNND----- --AFQSTT-- NFDHIKTHYY WSQILV---- NPSRVVAVGP I--PNIE--- ----------
Sphst1|178948           YWNSD----- --AFKSTT-- DFDHIKTHYY WSHPMI---- NPTRVVPLGP I--PVIE--- ----------
Sphst1|178963           YWNND----- --AFKSTT-- DFDHIKTNYA F-MAHI---- NPTQVVPLGP I--PNIE--- ----------
Sphst1|193280           YWNND----- --AFKSTT-- DFDHIKTNYA F-MAHI---- NPTQVVPLGP I--PNIE--- ----------
Sphst1|232298           YWNNY----- --AFNSTT-- DFDHIKTFYA F-MIHV---- NPTQVVPLGP I--PNIQ--- ----------
Stano2|8800             YWNVE----- --GFKEST-- DFKHIKENYT KSHSDI---- NPHAITPLGP Y--PDVEEDY ESDWSKLKPG
Suilu1|799805           YWNND----- --AFKSST-- KFDHIKEHYY WSQTLV---- NPSRVVPAGP V--PDIE--- ----------
Treme1|57346            YWNYP----- --AFKDTT-- DFDHIKEHYY YSHAHI---- NPTRIVPLGP R--VDIE--- ----------
Trire2|112633           YWDQDEEGPL RAAFYRTTAP NIGRYALGYA DSRWRIVQEG QGPLIVPAGP L--VLME--- ----------
Trire2|56819            YWRHA----- --AFRETT-- QFEHIKWHYT RSHTQI---- NPLSITPVGP L--PHIL--- ----------
Tubme1|220              YYDIP----- --GFKETT-- NFEHIKKHYT KSHSQI---- NPKGITPLGP A--PDIM--- ----------
Tubme1|6088             YWDIP----- --AFKQTT-- NFEHIKKHYT KSHAQI---- NPTGITPLGP L--PHIL--- ----------
Tulca1|76207            YWQNP----- --SFNTST-- DFHHFKTGYY GQKVCS---- NSTGIVPVGP I--PSIE--- ----------
Tulca1|76222            YWNNS----- --AFKDST-- NFEHIKHHYY WSHRQI---- NPTGIVPVGP I--PLIE--- ----------
Ustma1|1515             YWNEK----- --AFRDYT-- NFDHIKKGYS SVG------- GNKGIEPRGP L--PHML--- ----------
```

```
Gto2p(ECM4)|Sacce1|39   ----PL---- ---------- -------
Gto3p|Sacce1|4945       ----PP---- ---------- -------
GSTX1|Trave1|66368      ----PL---- ---------- -------
GSTX2|Trave1|50990      ----PL---- ---------- -------
GSTX3|Trave1|73942      ----PL---- ---------- -------
GSTX1|Phchr2|2982482    ----PLDA-- ---------- -------
GSTX2|Phchr2|2624184    ----PL---- ---------- -------
Agabi_varbisH97_2|137   ----PL---- ---------- -------
Amamu1|78766            ----PL---- ---------- -------
Amath1|62909            ----PL---- ---------- -------
Aspnid1|4489            ----PRDA-- ---------- -------
Aspnid1|7897            ----PKDKEV NAVKH----- -------
Aspnid1|8922            GVSHPAVLEF EKELNEE--- -------
Aurde3_1|115815         ----PL---- ---------- -------
Aurde3_1|1240856        ----PL---- ---------- -------
Botbo1|118302           ----PLN--- ---------- -------
Conpu1|60697            ----PL---- ---------- -------
Conpu1|82579            ----PL---- ---------- -------
Copci1|17111            --ANPVIDT- ---------- -------
Crypa2|254574           ----PLDKEV PAVEAAKA-- ----GKV
Dacsp1|67545            ----PL---- ---------- -------
Fomme1|130383           ----PL---- ---------- -------
Fomme1|130388           ----PL---- ---------- -------
Fomme1|143667           ----PLH--- ---------- -------
Fompi3|1022961          ----PFDD-- ---------- -------
Fompi3|127425           ----PLDA-- ---------- -------
Galma1|62390            ----PL---- ---------- -------
Glotr1_1|112236         ----PL---- ---------- -------
Glotr1_1|120328         GVKLPAVLDY ---------- -------
Glotr1_1|137290         ----PL---- ---------- -------
Glotr1_1|37102          ----ILLR-- ---------- -------
Gymlu1|36123            ----AL---- ---------- -------
Hebcy2|440337           ----PL---- ---------- -------
Hetan2|39076            ----AL---- ---------- -------
Hetan2|442680           ----PL---- ---------- -------
Hydpi2|96315            ----PL---- ---------- -------
Hypsu1|37440            ----PLKE-- ---------- -------
Hypsu1|40330            ----PLSL-- ---------- -------
Jaaar1|142100           ----PL---- ---------- -------
Jaaar1|63226            ----PL---- ---------- -------
Lacbi2|648000           ----PL---- ---------- -------
Mellp1|35938            ----ALDEDK ---------- -------
Oidma1|124122           ----REDEEV PAVAYALALR HSNTRRP
Oidma1|165197           ----GKQIEA VK-------- -------
Oidma1|182545           ----PLN--- ---------- -------
Oidma1|39468            ----PVAS-- ---------- -------
Oidma1|46504            ----PL---- ---------- -------
Paxin1|138020           ----PL---- ---------- -------
Paxru1|825741           ----PL---- ---------- -------
Phybl2|126467           ----PS---- ---------- -------
Picst3|43969            ----PL---- ---------- -------
Pilcr1|4571             ----DL---- ---------- -------
Pilcr1|827566           ----GL---- ---------- -------
Pirin1|77935            ----PLDS-- ---------- -------
Pisti1|121423           ----PL---- ---------- -------
PleosPC15_2|48940       ----PL---- ---------- -------
Plicr1|42268            ----PL---- ---------- -------
Plicr1|43753            ----PL---- ---------- -------
Plicr1|43758            ----PL---- ---------- -------
Punst1|118707           ----SL---- ---------- -------
Punst1|52675            ----PL---- ---------- -------
Punst1|55793            GVHLAA---- ---------- -------
Punst1|62356            ----PLDAK- ---------- -------
Schco3|2086920          ----PL---- ---------- -------
Schco3|2516063          ----PL---- ---------- -------
Schco3|2615640          ----PL---- ---------- -------
Sclci1|100395           ----PM---- ---------- -------
Sebve1|329667           ----PL---- ---------- -------
SerlaS7_9_2             ----PL---- ---------- -------
Sphst1|178948           ----PLDN-- ---------- -------
Sphst1|178963           ----PLTA-- ---------- -------
Sphst1|193280           ----PLTA-- ---------- -------
Sphst1|232298           ----PLTA-- ---------- -------
Stano2|8800             KVAHPRVLEA MSKL------ -------
Suilu1|799805           ----PL---- ---------- -------
Treme1|57346            ----PL---- ---------- -------
Trire2|112633           ----PL---- ---------- -------
Trire2|56819            ----PLEEEV RAVVA----- -----RV
Tubme1|220              ----PRDA-- ---------- -------
Tubme1|6088             ----PQDV-- ---------- -------
Tulca1|76207            ----PL---- ---------- -------
Tulca1|76222            ----SL---- ---------- -------
Ustma1|1515             ----PLDA-- ---------- -------
```

Figure S3. Multiple sequence alignment of putative fungal GSTX

All sequences were retrieved from MycoCosm on the JGI website. Sequences were aligned using MAFFT and manually annotated in order to highlight the secondary structure elements of *S. cerevisiae* ECM4 structure and specific residues (catalytic residues in yellow, G-site residues in blue, H-site residues in green). Motif that exhibits variations in helix α9 is underlined.

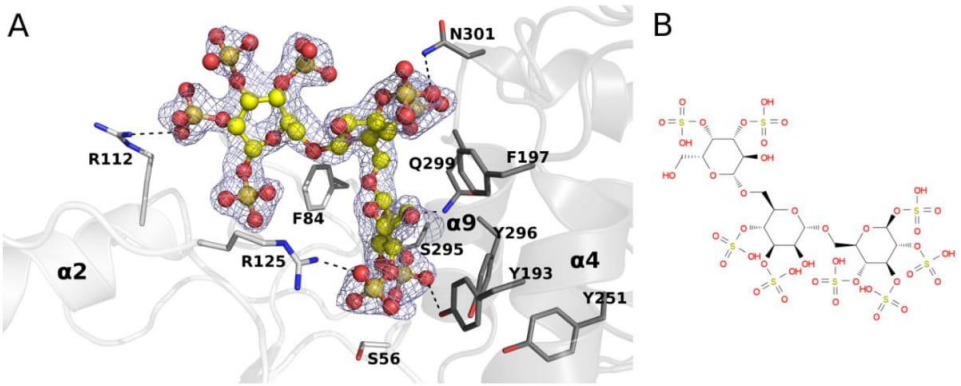


Figure S4. Dextran-sulfate complexed in TvGSTX3 C56S active site

A. 2Fo-Fc omit map is shown at 1.0 σ around dextran-sulfate (yellow sticks and spheres). Side-chains in the vicinity of the ligand are labeled and displayed as sticks and polar contacts are represented as dashed lines. **B.** Chemical structure of the ligand is shown (three units of the polymer).

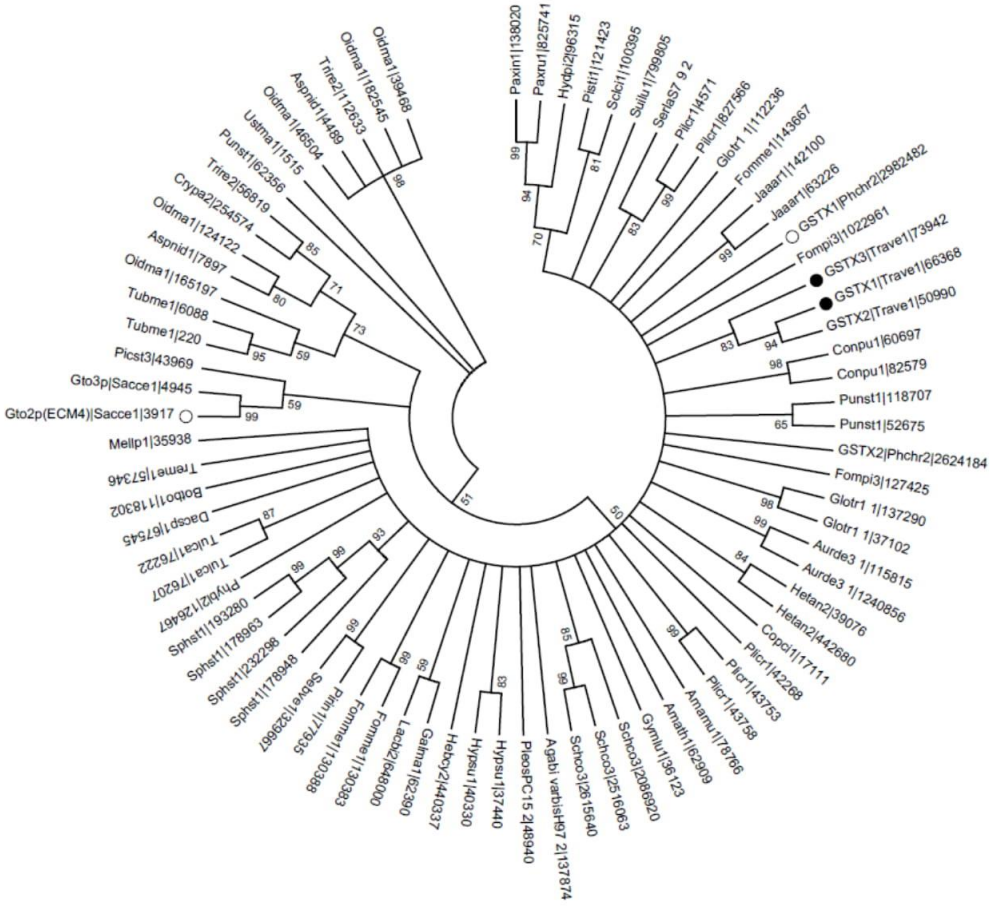


Figure S5. Neighbor-Joining analysis of putative GSTX sequences from various fungi

Sequences associated with existing solved structures are marked with a white circle (previous work) or a black circle (this study). The tree was condensed based on bootstrap values >50% (indicated below corresponding nodes). Abbreviations of organisms and protein accession numbers are provided on the JGI website (https://genome.jgi.doe.gov/fungi). Clustering of sequences in an organism-specific manner suggests recent diversification among the Xi class.

References

1 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).

2 Tickle, I. J. Experimental determination of optimal root-mean-square deviations of macromolecular bond lengths and angles from their restrained ideal values. *Acta crystallographica. Section D, Biological crystallography* **63**, 1274-1281; author reply 1282-1273, doi:10.1107/S0907444907050196 (2007).

3 Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615-619, doi:10.1093/nar/gkh398 (2004).

Discussion générale et Perspectives

Les organismes biologiques ont développé des systèmes biochimiques complexes leur permettant d'assurer leur survie face à la diversité des molécules potentiellement toxiques provenant de leur environnement. Ces systèmes incluent des enzymes de phase II telles que les glutathion transférases (GST), intervenant pour faciliter l'export des xénobiotiques en leur conjuguant un goupement glutathion. Les organismes fongiques dégradeurs de bois tels que *Phanerochaete chrysosporium* et *Trametes versicolor* présentent dans leur génome un nombre élevé de gènes codant pour des GST. *T. versicolor* possède au moins seize gènes codant pour des GST de classe Oméga, et trois gènes codant pour des GST de classe Xi. Le but de cette thèse était d'une part de caractériser structuralement différentes isoformes de TvGSTO et TvGSTX chez cet organisme modèle d'étude de la dégradation enzymatique du bois, et d'autre part de déterminer quelles molécules sont ciblées par les GSTO en utilisant une approche combinant des méthodes biochimiques et structurales. En parallèle de l'étude des TvGST, des études structurales et biochimiques ont également été entreprises sur la GST Xi ScECM4 retrouvée chez l'organisme modèle *Saccharomyces cerevisiae* ainsi qu'une étude de modélisation sur les isoformes ScGTO1 et ScGTO3.

Les lignes suivantes discutent des différents résultats obtenus et les replacent dans le contexte de la littérature. Dans un premier temps sont discutés les résultats structuraux et biochimiques des GST obtenus pour la classe Xi, puis pour la classe Oméga. Dans un second temps, la capacité des TvGSTO à reconnaître leurs molécules cibles, à savoir les polyphénols, est discutée. Les derniers paragraphes se veulent plus méthodologiques et évoquent les techniques qui ont été utilisées pour obtenir les complexes GST-ligand. L'utilisation de la cristallographie d'affinité y est notamment développée. Enfin, sont discutées les implications et les perspectives à long terme de ce travail de recherche.

- Les GST Xi

Les GST Xi sont des GST présentant une cystéine catalytique à l'instar des GST Oméga. En revanche, au départ de cette étude il était établi que ces deux classes présentaient des différences importantes tant dans leur structure que dans leur spécificité de substrat. Les travaux sur les GSTX mettaient en évidence une activité particulière « glutathionyl-(hydro)quinone réductase » (GHR, Xun et al, 2010) alors qu'une spécificité de substrat pour les glutathionyl-acétophénones avait été proposée pour les GSTO (Board et Anders, 2007 ; Meux et al, 2013). La résolution des structures cristallographiques d'une isoforme de la levure *S. cerevisiae* (ScECM4) et de deux isoformes du basidiomycète *Trametes versicolor*

(TvGSTX1 et TvGSTX3) confirme les particularités structurales de cette classe Xi et notamment le dimère atypique formé par l'interaction des deux monomères par leurs domaines tout hélice, initialement décrit dans la structure d'une isoforme de *P. chrysosporium* (PcGSTX1) (Meux et al, 2011). Néanmoins, nos résultats mettent en évidence des spécificités structurales pour ScECM4, notamment deux insertions (entre le brin β2 et l'hélice α2 dans le domaine thiorédoxine, et au niveau C-terminal de l'hélice α5 dans le domaine tout hélice). Ces insertions seraient également retrouvées chez les isoformes GTO1 et GTO3 de *S. cerevisiae* d'après les résultats issus d'une modélisation comparative. Toutefois aucune fonction particulière n'a été trouvée pour ces singularités qui seraient restreintes au clade *Saccharomycetacea* d'après notre analyse phylogénétique.

Concernant les isoformes de *Trametes versicolor*, nos études biochimique et structurale montrent que TvGSTX3 présente une dualité d'activité puisqu'elle peut réduire efficacement aussi bien la GS-ménadione (substrat des GSTX) que la GS-phenylacétophénone (substrat des GSTO) (Meux et al, 2011 ; Board et Anders, 2007). Les analyses structurales et les résultats de mutagenèse dirigée montrent que cette dualité d'activité serait liée à des résidus du site actif portés par l'hélice α9. Deux résidus histidines conservés dans la plupart des GSTX sont remplacés par une sérine et une glutamine dans TvGSTX3, aménageant une crevasse hydrophobe. Notre analyse phylogénétique suggère que cette hélice serait propice aux mutations car d'autres GSTX d'organismes fongiques présentent des variations dans cette région. Dans le cas des isoformes GTO1 et GTO3 de *S. cerevisiae*, nos résultats indiquent que les deux histidines sont également absentes. Ces différences pourraient suggérer de nouvelles activités catalytiques mais aussi la possibilité de réduire d'autres substrats que les GS-(hydro)quinones. En effet, il a été montré que GTO1 est retrouvée au niveau des peroxysomes avec un rôle potentiel dans le métabolisme des acides aminés soufrés: elle pourrait assurer la régulation par déglutathionylation d'une enzyme, la cystathionine β-lyase (Barreto et al, 2006). Cette étude réalisée avant la découverte de la fonction GHR est la seule à ce jour prêtant des substrats protéiques aux GSTX. Bien que n'ayant jamais été testés, des substrats de nature protéique ne sont pas à exclure puisque d'autres GST à cystéine, les GSTO, peuvent déglutathionyler des protéines (Kim et al, 2012 ; Menon et Board, 2013).

Le rôle biologique des GSTX chez les organismes fongiques n'est pas encore clairement élucidé. En 1997, il été suggéré que le gène codant pour ScECM4 serait lié à la biosynthèse de la paroi fongique (Lussier et al, 1997). De manière intéressante, nos résultats

cristallographiques obtenus pour le mutant TvGSTX3^{C56S} montrent que cette protéine est capable de lier un additif de cristallisation de type « polymère glycosidique », le dextran-sulfate. Compte-tenu de la présence de polymères glycosidiques dans la paroi fongique (Osumi, 1998), cette observation pourrait avoir une signification biologique vis-à-vis du rôle suggéré dans la biosynthèse de la paroi cellulaire. Chez les plantes et certaines bactéries du sol, d'autres rôles biologiques sont évoqués à propos des GSTX. A ce jour, le seul rôle avéré d'une GSTX est retrouvé dans la dégradation du pentachlorophénol pour PcpF de *Sphingobium chlorophenolicum* (Huang et al, 2008). Chez le peuplier, les deux isoformes PtGHR1 et PtGHR2 portent une séquence d'adressage et sont retrouvées au niveau du chloroplaste (Lallement et al, 2015). Plus généralement, il a été suggéré que les GSTX seraient retrouvées chez les organismes produisant par leur métabolisme des intermédiaires de type hydroquinone : chez les organismes photosynthétiques (plantes et cyanobactéries) et chez les organismes capables de dégrader la lignine (bactéries lignivores et champignons dégradeurs de bois) (Belchik et Xun, 2011). En effet, les intermédiaires de type hydroquinone sont capables de réagir spontanément avec le glutathion (Lam et al, 2012) et un moyen d'assurer la détoxication des produits formés serait de déglutathionyler ces espèces grâce à l'activité GHR des GSTX. Mais ces rôles restent des suggestions et des expériences supplémentaires seront nécessaires pour clarifier le rôle biologique des GST Xi en particulier chez les organismes photosynthétiques et lignivores.

- Structure des GST Oméga

Sur la base des séquences, les GSTO de *T. versicolor* peuvent être divisées en deux types phylogénétiquement distincts, précédemment décrits par Morel et collaborateurs (2009) dans une étude ciblant les GST de plusieurs champignons. Les GSTO de type II sont retrouvées au nombre de deux isoformes chez *T. versicolor* et présentent toutes deux une cystéine catalytique. Les GSTO de type III sont au nombre de quatorze isoformes : douze ont une sérine catalytique et deux présentent à la place une cystéine. Une extension du nombre de GSTO est observée chez d'autres basidiomycètes. De manière systématique, ce sont les GSTO de type III qui forment une extension de taille variable (pouvant atteindre quatorze isoformes chez *T. versicolor)*, alors que les GSTO de type II sont présentes au nombre quasi constant de deux à trois isoformes chez tous les champignons. Cela suggère des rôles biologiques distincts entre ces deux types, bien que peu de données soient disponibles pour le moment. Notre étude en collaboration avec l'équipe d'E. Gelhaye (laboratoire IAM) montre que les TvGSTO à sérine catalysent des réactions GSH-transférase alors que les TvGSTO à

cystéine catalysent des réactions thiol-transférase, malgré une forte conservation de structure. Les résultats structuraux montrent que le site de reconnaissance du glutathion est bien conservé entre les différentes isoformes, et que la différence majeure au niveau N-terminal se trouve dans la nature du résidu catalytique porté en N-terminal de l'hélice α1, soulignant le rôle crucial de ce résidu dans la catalyse assurée par les GST Oméga.

Au niveau structural, des différences notables ont pu être remarquées entre les structures des GSTO de *T. versicolor* résolues dans ce projet et celles d'autres organismes, malgré la conservation de caractéristiques communes. Le dimère particulièrement ouvert est une caractéristique également retrouvée dans la structure de HsGSTO1 humaine (Board et al, 2000). Cette particularité serait déterminante dans la capacité de HsGSTO1 à déglutathionyler des protéines (parmi lesquelles l'actine B, certaines protéines 'heat shock' et la protéine induite par la prolactine) avec un rôle dans la régulation de la signalisation cellulaire (Menon et Board, 2013). Par analogie, des substrats de nature protéique pourraient aussi être déglutathionylés par les TvGSTO. Précédemment, la présence de résidus catalytiques autres qu'une cystéine a également été relevée pour la structure de la GST Oméga 3 de *Bombyx mori* présentant un résidu asparagine et une absence d'activité GSH-transférase avec le substrat CDNB (Chen et al, 2011). Dans le cas des TvGSTO de type III, un résidu sérine est retrouvé dans la plupart des isoformes, conférant une activité GSH-transférase envers ce substrat.

Comparée aux GSTO humaines, les TvGSTO diffèrent essentiellement par l'absence d'une extension N-terminale rejoignant l'extrémité C-terminale, telle que celle observée chez HsGSTO1 (Board et al, 2000). Alors qu'aucune extension n'est présente chez les TvGSTO de type III, une extension différente se repliant en une hélice structurée supplémentaire est retrouvée pour TvGSTO3C. La deuxième différence notable est la présence chez les TvGSTO de type III d'une boucle d'insertion entre les brins β3 et β4 au niveau du domaine thiorédoxine. De manière intéressante, cette boucle de 15 résidus riche en proline recouvre le feuillet β, de manière similaire à l'extrémité N-terminale riche en proline de HsGSTO1 (Board et al, 2000). Cette boucle peut aussi être apparentée à l'insertion de localisation similaire retrouvée chez la GST Zeta du champignon pathogène *Coccidioides immitis,* de fonction actuellement inconnue (Edwards et al, 2011). Précédemment il a été montré la ressemblance structurale des GST de classe Oméga avec les GST de classe Tau (Thom et al, 2001). Ces deux classes partagent en effet la présence d'un dimère ouvert ainsi que d'une hélice α9 située à proximité du site actif. L'analyse structurale suggère que les plus proches homologues structuraux des TvGSTO sont les GSTU végétales, avant les GSTO d'*Homo*

sapiens et de *Bombyx mori*. Concernant les TvGSTOS de type III, cette ressemblance structurale est encore plus prononcée dans la mesure où l'extension N-terminale est absente et un résidu catalytique sérine peut être présent à l'instar des GSTU (Thom et al, 2001). Les implications biologiques de ces observations sont discutées dans la rubrique suivante.

- Potentiel des TvGSTO dans la reconnaissance des polyphénols

Au cours de ce projet, nous avons montré l'interaction des TvGSTO avec les molécules de la famille des polyphénols. Grâce aux expériences de TSA et d'inhibition cinétique, différentes molécules ont été identifiées parmi lesquelles des flavonoïdes, un stilbène et des hydroxybenzophénones. Deux sites d'interaction ont été mis en évidence grâce aux structures cristallographiques de complexes déterminées. Suivant les isoformes, ces interactions peuvent avoir lieu soit au niveau du site hydrophobe dans le site actif (site H), soit au niveau du site ligandine à l'interface du dimère (site L). Les spécificités structurales de chaque isoforme expliquent notamment le profil d'interaction spécifique de chacune d'elles. Cependant, aucune réaction catalytique avec les différentes molécules identifiées n'a été détectée. Les hydroxybenzophénones interagissent dans le site H de TvGSTO3S. Le métabolisme de ces molécules a été décrit chez *T. versicolor* et implique plutôt des glycosylations probablement par les glycosyl-transférases retrouvées dans la phase II (Gago-Ferrero et al, 2012). Le stilbène oxyresvératrol et certains flavonoïdes interagissent également au niveau du site H, notamment la dihydrowogonine dans l'isoforme TvGSTO3S. Pour certaines isoformes (TvGSTO6S et TvGSTO2C notamment) l'interaction avec des polyphénols (naringénine et *cis*-oxyresvératrol) a été démontrée au niveau du site L. D'après la littérature sur les GST, ce site L est le plus souvent retrouvé à l'interface du dimère (McTigue et al, 1995 ; Brock et al, 2013 ; Ahmad et al, 2017) et a été associé à un rôle de transport et/ou de piégeage de molécules hydrophobes (Hayes et al, 2005). Les flavonoïdes sont des molécules qui bien souvent présentent des activités anti-fongiques et sont retrouvées en faible quantité dans certains bois (Valette et al, 2017). Dans le cas des champignons dégradeurs de bois, la libération de ce type de molécules à l'interface bois-champignon a lieu lors de la dégradation des polymères lignine et cellulose et leur détoxication au niveau intracellulaire est probablement cruciale pour la survie du champignon. L'analyse de fructifications de *T. versicolor* par spectrométrie de masse a révélé la présence de nombreux flavonoïdes présents dans le champignon (Janjusevic et al, 2017). Ces molécules proviennent probablement du substrat sur lequel le champignon croît (sol, bois), puisqu'à ce jour aucune voie de synthèse des flavonoïdes par les champignons n'est communément admise par la

communauté scientifique (Janjusevic et al, 2017). D'autre part, aucune étude ne démontre que ces molécules peuvent être glutathionylées par les systèmes de détoxication fongiques. Chez le maïs, il a été démontré la glutathionylation de molécules de la famille des anthocyanes, dont la structure est proche de celles des flavonoïdes, par l'action de la GST Bz2 appartenant à la classe Tau (Marrs et al, 1995). Cette GST interviendrait ainsi dans le transport des dérivés glutathionylés d'anthocyanes à la vacuole. D'autres GST végétales auraient un rôle similaire de transport des flavonoïdes à la vacuole, parmi lesquelles des GST de classe Phi chez le pétunia (Mueller et al, 2000), avec une fonction de co-transport avec le glutathion sans réaction catalytique. Les TvGSTO pourraient ainsi présenter des rôles similaires de détoxication des polyphénols, soit :

(i) par conjugaison avec le glutathion, un rôle en cohérence avec la fixation de flavonoïdes au site H et à une distance cohérente du résidu catalytique, bien qu'aucun substrat de cette famille ne soit à ce jour décelé,

(ii) par transport sans réaction catalytique, au niveau du site H (n'excluant pas un co-transport avec le glutathion) ou du site L, le compartiment cellulaire d'adressage étant à ce jour indéterminé.

La similarité structurale des TvGSTO avec les GSTU végétales renforce l'hypothèse d'un rôle possible dans le transport, voire la détoxication des polyphénols. Il est tentant de penser que la multiplication du nombre d'isoformes de GSTO chez *T. versicolor* et chez d'autres champignons saprotrophes pourrait être liée à la diversité des polyphénols libérés lors de la dégradation du bois. A l'heure actuelle, peu de données biologiques sont disponibles sur les GST Oméga fongiques. Chez l'ascomycète pathogène de plante *Alternaria brassicicola*, des études transcriptomiques montrent la surexpression des transcrits correspondant aux GST, dont notamment une GST Oméga (Sellam et al, 2007) lors de l'exposition à l'allyl-isothiocyanate (une molécule antifongique synthétisée par les végétaux de la famille des Brassicacées). Concernant les dégradeurs de bois, des études transcriptomiques préliminaires chez *T. versicolor* suggèrent la surexpression de certains gènes de TvGSTO à cystéine en condition de carence (Eric Gelhaye, communication personnelle). Un rôle dans la réponse au stress est probablement retrouvé pour les TvGSTO. Des études supplémentaires seront nécessaires pour préciser la fonction biologique de ces GST chez les organismes fongiques.

- Méthodes d'obtention de structures de complexes GST-ligands

Dès le début du projet d'étude des glutathion transférases fongiques, nos efforts se sont portés sur la détermination de structures de complexes GST-ligands. Dans un premier temps, l'obtention des structures des complexes entre TvGSTO3S et les dérivés du glutathion par trempage des cristaux de la forme apo nous a confortés dans l'utilisation de cette méthode. Cependant celle-ci s'est révélée infructueuse dans l'obtention de structures de complexes impliquant des molécules plus hydrophobes (dissoutes dans du DMSO) telles que les hydroxybenzophénones. Aucun pic correspondant aux ligands potentiels n'a été visible dans les cartes de densité électronique. Précédemment, d'autres études avaient rapporté un échec dans le cas de trempage de cristaux alors que la co-cristallisation avec la même molécule avait conduit à l'obtention du complexe (Gelin et al, 2015). D'autre part, les tests d'obtention de complexes avec des ligands dont les interactions potentielles étaient inconnues (criblage « aveugle ») ont également fait l'objet d'un certain nombre de collectes se résumant à des échecs.

Deux éléments majeurs ont contribué à surmonter ces difficultés. D'une part, une stratégie de co-cristallisation a permis d'obtenir les structures des complexes de TvGSTO3S avec les HBPs, en utilisant les mêmes conditions de cristallisation que la forme apo. Pour les autres isoformes (TvGSTO2S, TvGSTO6S et TvGSTO2C) l'utilisation des conditions de cristallisation initiales, quel que soit le type de co-cristallisation (liquide ou sèche), n'a pas conduit à l'obtention de cristaux. Plutôt qu'une recherche de nouvelles conditions de cristallisation, une stratégie de « diffusion sèche » des ligands a été mise au point. Cette méthode repose sur la mise en contact de la liqueur mère contenant les cristaux de la forme apo avec le ligand potentiel séché au fond des puits des plaques de cristallisation : une alternative intéressante puisque le DMSO, solvant de la molécule d'intérêt ayant un effet néfaste sur les cristaux de protéines, est totalement évaporé comme c'est le cas dans la co-cristallisation sèche (Gelin et al, 2015). Une stratégie similaire est évoquée sur le site d'Hampton Research (https://www.hamptonresearch.com), dans la rubrique « astuces de cristallisation » : le ligand sous forme de poudre peut être ajouté à la liqueur mère contenant les cristaux. Les changements de conformation des TvGSTO lors de la liaison à leur ligand sont locaux uniquement et aucun réarrangement structural majeur n'a été observé. Cette méthode de diffusion sèche s'est révélée particulièrement utile et a permis de déterminer un nombre conséquent de structures de complexes en un temps relativement court.

Le deuxième élément crucial a été l'orientation des tests de co-cristallisation ou de trempage à partir des données de TSA obtenues pour les ligands potentiels. Cela nous a permis de sélectionner les meilleurs candidats et ainsi d'augmenter les chances d'obtention d'une structure d'un complexe. Cette approche a conduit à la détermination de neuf structures de complexes GST-polyphénol, et toutefois à autant de structures sans ligand malgré une valeur de déplacement de température de demi-dénaturation (ΔTm) significative. Une explication possible de ces échecs est donnée par la condition de cristallisation dont le pH peut différer du pH optimal de fixation du ligand et ainsi prévenir sa fixation à la GST par modification des propriétés électrostatiques de la protéine. L'utilisation des données biophysiques permet d'enrichir les résultats positifs issus d'une campagne de criblage de molécules (Schiebel et al, 2016). En revanche, cette stratégie implique une restriction des molécules testées qui peut conduire à manquer des résultats positifs potentiels (Schiebel et al, 2016). Dans l'ensemble, nos données illustrent l'utilité de la technique de TSA dans le cadre d'un projet de criblage cristallographique, comme cela a déjà été évoqué précédemment (Chung, 2007).

- Utilisation de la « cristallographie d'affinité »

Un de nos objectifs était de pouvoir identifier des ligands de GST à partir d'extraits naturels, en particulier des extraits de bois préparés par nos collaborateurs des laboratoires IAM et LERMAB. Des expériences d'inhibition avec une sonde fluorescente ont permis de pointer sur un extrait acétonique de duramen de merisier, stabilisant les isoformes de GSTO et contenant donc des ligands potentiels. Le fractionnement de cet extrait par HPLC, puis les expériences de RMN et de spectrométrie de masse suggéraient la présence de deux à trois molécules parmi lesquelles potentiellement un ou plusieurs flavonoïdes. Néanmoins, l'identification précise n'a pu être effectuée à partir de ces seules données. Une approche de cristallographie d'affinité à partir de l'isoforme TvGSTO3S a donc été développée : c'est-à-dire une co-cristallisation de la fraction d'intérêt avec la GST dans le but d'identifier le ligand ayant la meilleure affinité parmi le mélange. Cette méthode décrite dans la littérature a permis à Aguda et ses collaborateurs d'identifier un nouvel inhibiteur de cathepsine K présent en faible quantité dans un extrait d'actinomycète (Aguda et al, 2016). Nos données de diffraction à haute résolution (1,70 Å) acquises à partir du co-cristal nous ont permis d'identifier une densité électronique au niveau du site H de l'isoforme TvGSTO3S. Les informations de structure liées à la forme de cette densité étaient en accord avec un ligand de type flavonoïde et ont précisé la stéréochimie S de l'atome de carbone C2. Les positions des différents

substituants (groupements hydroxyles et méthoxyles) sur le noyau flavanone suggérées par la densité électronique, combinées à l'information structurale acquise par RMN et MS ont permis d'identifier sans ambigüité la dihydrowogonine. La présence de ce polyphénol dans le bois de merisier est documentée dans la littérature (McNulty et al, 2009).

La cristallographie d'affinité s'est révélée cruciale dans l'identification du polyphénol présent dans l'extrait de merisier interagissant avec l'isoforme TvGSTO3S. Il convient cependant de préciser que les données RMN et MS ont été indispensables à cette identification. Ces données permettent de confirmer la nature de la molécule suspectée, *a fortiori* dans le cas de molécules nouvellement identifiées. C'est dans ce cadre « multiméthodes » que la cristallographie d'affinité peut être réellement puissante. La cristallographie à elle-seule n'a pas la capacité réelle d'identifier de nouvelles molécules. Aguda et ses collaborateurs ont certes réussi à identifier un nouvel inhibiteur, le lichostatinal, dans le site actif de la cathepsine K sur la seule base de l'information structurale liée à la densité électronique (Aguda et al, 2016). L'obtention de la structure en complexe à partir de la molécule synthétique leur a permis d'assurer le résultat. La qualité, de même que la résolution des données de diffraction acquises, jouent évidemment un rôle déterminant dans l'identification. Par ailleurs, la stabilisation complète du ligand est nécessaire à son identification ainsi que son unicité. Dans le cas de deux ligands ayant des affinités proches pour la protéine, l'occupation simultanée au même site dans le cristal compromet en effet leur identification. Ces problèmes sont finalement similaires à ceux rencontrés lors de l'utilisation de cocktails de fragments lors d'une campagne de criblage cristallographique (Chilingaryan et al, 2012).

- <u>Perspectives appliquées à long terme</u>

L'utilisation des GST à des fins biotechnologiques est bien documentée. L'application la plus répandue est probablement retrouvée dans la purification de protéines fusionnées avec une étiquette GST sur colonne d'affinité de type sépharose-GSH (Lichty et al, 2005). L'affinité des GST pour le tripeptide est ainsi mise à profit pour faciliter la purification de protéines recombinantes. Plus généralement, la famille des GST est largement étudiée pour son potentiel en ingénierie protéique, c'est-à-dire dans l'amélioration ou la modification de leurs propriétés à des fins biotechnologiques ou industrielles. Ces applications sont notamment évoquées dans la revue de Perperopoulou et collaborateurs (2018). Brièvement, ces applications couvrent des domaines aussi variés que l'agriculture, l'environnement, la

médecine ou les nanobiotechnologies. A titre d'exemple, nous pouvons citer le développement de biosenseurs pour la détection de pesticides (Chronopoulou et al, 2012) ou d'herbicide (Oliveira et al, 2013) basée sur la capacité des GST à reconnaître ces molécules. En médecine, des inhibiteurs ciblant les GST humaines sont recherchés dans le cadre de certains cancers où ces protéines sont surexprimées et participent à la résistance aux chimiothérapies (Ramkumar et al, 2016 ; Karpusas et al, 2013). En nanotechnologie, des recherches portent sur l'ingénierie de ces protéines pour former des nanostructures ordonnées telles ques des nano-anneaux (Bai et al, 2013) ou des nano-câbles dans lesquels l'activité enzymatique des GST est conservée (Zhang et al, 2012).

A travers ce projet, nous avons exploré la possibilité d'utiliser les GST comme outils pour identifier des molécules aux propriétés intéressantes parmi des extraits de bois tropicaux. Notre hypothèse de départ était que, les systèmes de détoxication des champignons ayant évolué pour reconnaître les métabolites secondaires de plantes, ces enzymes pourraient potentiellement être utilisées pour reconnaître et identifier des molécules aux propriétés antimicrobiennes ou antioxydantes. L'utilisation d'une banque de six TvGSTO couplée à la technique de TSA sur une banque d'extraits d'arbres amazoniens a permis d'identifier l'oxyresvératrol. Ce polyphénol interagit avec une affinité de l'ordre du micromolaire avec le site H de l'isoforme TvGSTO2S, mais également à son site ligandine. L'oxyresvératrol est une molécule ayant des propriétés anti-inflammatoires (Chen et al, 2013), anti-bactériennes (Joung et al, 2016) et anti-oxydantes (Lorenz et al, 2003) et est notamment utilisée en industrie cosmétique. Notre approche confirme ainsi le potentiel des GSTO pour sonder la diversité chimique des extraits de bois et suggère que d'autres GST de micro-organismes pourraient également être utilisées dans le cadre d'une approche similaire.

Références Bibliographiques

I Ouvrages de référence

Campbell Biology, neuvième édition, Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB, 2011

Crystals, X-rays and Proteins, Sherwood D & Cooper J, 2010

Enzymatic Basis of Detoxication, Jakoby WB & Habig WH, 1980

Heterologous Gene Expression in *E.coli*, Xu MQ & Evans TC, 2011

The Fungi, troisième édition, Watkinson SC, Boddy L, Money NP, 2016

II Articles et revues

Adams, P. D., et al. (2010), 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', Acta Crystallogr D Biol Crystallogr, 66 (Pt 2), 213-21.

Aguda, A. H., et al. (2016), 'Affinity Crystallography: A New Approach to Extracting High-Affinity Enzyme Inhibitors from Natural Extracts', J Nat Prod, 79 (8), 1962-70.

Ahmad, L., et al. (2017), 'Structural evidence for Arabidopsis glutathione transferase AtGSTF2 functioning as a transporter of small organic ligands', FEBS Open Bio, 7 (2), 122-32.

Bai, Y., et al. (2013), 'Highly ordered protein nanorings designed by accurate control of glutathione S-transferase self-assembly', J Am Chem Soc, 135 (30), 10966-9.

Baldrian, Petr (2008), 'Wood-inhabiting ligninolytic basidiomycetes in soils: Ecology and constraints for applicability in bioremediation', Fungal Ecology, 1 (1), 4-12.

Barreto, L., et al. (2006), 'A peroxisomal glutathione transferase of Saccharomyces cerevisiae is functionally related to sulfur amino acid metabolism', Eukaryot Cell, 5 (10), 1748-59.

Belchik, S. M. and Xun, L. (2011), 'S-glutathionyl-(chloro)hydroquinone reductases: a new class of glutathione transferases functioning as oxidoreductases', Drug Metab Rev, 43 (2), 307-16.

Bergfors, T. (2003), 'Seeds to crystals', J Struct Biol, 142 (1), 66-76.

Board, P. G. and Anders, M. W. (2007), 'Glutathione transferase omega 1 catalyzes the reduction of S-(phenacyl)glutathiones to acetophenones', Chem Res Toxicol, 20 (1), 149-54.

Board, P. G. and Menon, D. (2016), 'Structure, function and disease relevance of Omega-class glutathione transferases', Arch Toxicol, 90 (5), 1049-67.

Board, P. G., et al. (2008), 'S-(4-Nitrophenacyl)glutathione is a specific substrate for glutathione transferase omega 1-1', Anal Biochem, 374 (1), 25-30.

Board, P. G., et al. (2000), 'Identification, characterization, and crystal structure of the Omega class glutathione transferases', J Biol Chem, 275 (32), 24798-806.

Bocedi, A., et al. (2016), 'Evolution of Negative Cooperativity in Glutathione Transferase Enabled Preservation of Enzyme Function', J Biol Chem, 291 (52), 26739-49.

Bocedi, A., et al. (2013), 'The impact of nitric oxide toxicity on the evolution of the glutathione transferase superfamily: a proposal for an evolutionary driving force', J Biol Chem, 288 (34), 24936-47.

Booth, J., Boyland, E., and Sims, P. (1961), 'An enzyme from Rat liver catalyzing conjugations with glutathione', Biochemical Journal, 79 (3), 516-24.

Bousset, L., et al. (2001a), 'Crystal structures of the yeast prion Ure2p functional region in complex with glutathione and related compounds', Biochemistry, 40 (45), 13564-73.

Bousset, L., et al. (2001b), 'Structure of the globular region of the prion protein Ure2 from the yeast Saccharomyces cerevisiae', Structure, 9 (1), 39-46.

Brock, J., G., Board P., and Oakley, A. J. (2013), 'Structural insights into omega-class glutathione transferases: a snapshot of enzyme reduction and identification of a non-catalytic ligandin site', Plos One, 8 (4).

Camarero, Susana, et al. (2007), 'Paper pulp delignification using laccase and natural mediators', Enzyme and Microbial Technology, 40 (5), 1264-71.

Carvalho, A. L., Trincao, J., and Romao, M. J. (2009), 'X-ray crystallography in drug discovery', Methods Mol Biol, 572, 31-56.

Chen, B. Y., et al. (2011), 'Structure-guided activity restoration of the silkworm glutathione transferase Omega GSTO3-3', J Mol Biol, 412 (2), 204-11.

Chen, Vincent B., et al. (2010), 'MolProbity: all-atom structure validation for macromolecular crystallography', Acta Crystallographica Section D: Biological Crystallography, 66 (Pt 1), 12-21.

Chen, X., et al. (2002), 'Structural identification of a bacterial quorum-sensing signal containing boron', Nature, 415 (6871), 545-49.

Chen, Yi-Ching, et al. (2013), 'Morus alba and active compound oxyresveratrol exert anti-inflammatory activity via inhibition of leukocyte migration involving MEK/ERK signaling', BMC Complementary and Alternative Medicine, 13 (1), 45.

Cheng, King-Fai and Leung, Ping-Chung (2008), 'General review of polysaccharopeptides (PSP) from C.versicolor : Pharmacological and clinical studies', Cancer Therapy, 6, 117-30.

Chilingaryan, Z., Yin, Z., and Oakley, A. J. (2012), 'Fragment-based screening by protein crystallography: successes and pitfalls', Int J Mol Sci, 13 (10), 12857-79.

Choi, J. H., Lou, W., and Vancura, A. (1998), 'A novel membrane-bound glutathione S-transferase functions in the stationary phase of the yeast Saccharomyces cerevisiae', Journal of Biological Chemistry, 273 (45), 29915-22.

Chronopoulou, Evangelia G., et al. (2012), 'Inhibition of human glutathione transferases by pesticides: Development of a simple analytical assay for the quantification of pesticides in water', Journal of Molecular Catalysis B: Enzymatic, 81, 43-51.

Chung, C. W. (2007), 'The use of biophysical methods increases success in obtaining liganded crystal structures', Acta Crystallogr D Biol Crystallogr, 63 (Pt 1), 62-71.

Cimmperman, Piotras, et al. (2008), 'A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands', Biophysical Journal, 95 (7), 3222-31.

Congreve, Miles, et al. (2003), 'A 'Rule of Three' for fragment-based lead discovery?', Drug Discovery Today, 8 (19), 876-77.

D'Arcy, A., Villard, F., and Marsh, M. (2007), 'An automated microseed matrix-screening method for protein crystallization', Acta Crystallogr D Biol Crystallogr, 63 (Pt 4), 550-4.

Danley, D. E. (2006), 'Crystallization to obtain protein-ligand complexes for structure-aided drug design', Acta Crystallogr D Biol Crystallogr, 62 (Pt 6), 569-75.

Davis, C., et al. (2011), 'The role of glutathione S-transferase GliG in gliotoxin biosynthesis in Aspergillus fumigatus', Chem Biol, 18 (4), 542-52.

Davis, I. W., et al. (2004), 'MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes', Nucleic Acids Res, 32 (Web Server issue), W615-9.

Dekant, W., Vamvakas, S., and Anders, M. W. (1990), 'BIOSYNTHESIS, BIOACTIVATION, AND MUTAGENICITY OF S-CONJUGATES', Toxicology Letters, 53 (1-2), 53-58.

Deroy, A., et al. (2015), 'The GSTome Reflects the Chemical Environment of White-Rot Fungi', PLoS One, 10 (10), e0137083.

Dixon, D. P. and Edwards, R. (2010), 'Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing', J Biol Chem, 285 (47), 36322-9.

Dixon, D. P., Cole, D. J., and Edwards, R. (2000), 'Characterisation of a zeta class glutathione transferase from Arabidopsis thaliana with a putative role in tyrosine catabolism', Arch Biochem Biophys, 384 (2), 407-12.

Eastwood, D. C., et al. (2011), 'The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi', Science, 333 (6043), 762-65.

Edwards, T. E., et al. (2011), 'Structures of a putative zeta-class glutathione S-transferase from the pathogenic fungus Coccidioides immitis', Acta Crystallogr Sect F Struct Biol Cryst Commun, 67 (Pt 9), 1038-43.

Emsley, P. and Cowtan, K. (2004), 'Coot: model-building tools for molecular graphics', Acta Crystallogr D Biol Crystallogr, 60 (Pt 12 Pt 1), 2126-32.

Erlanson, D. A., et al. (2016), 'Twenty years on: the impact of fragments on drug discovery', Nat Rev Drug Discov, 15 (9), 605-19.

Eswar, Narayanan, et al. (2006), 'Comparative Protein Structure Modeling Using Modeller', Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.], 0 5, Unit-5.6.

Evans, P. R. and Murshudov, G. N. (2013), 'How good are my data and what is the resolution?', Acta Crystallogr D Biol Crystallogr, 69 (Pt 7), 1204-14.

Fernandez-Canon, J. M. and Penalva, M. A. (1998), 'Characterization of a Fungal Maleylacetoacetate Isomerase Gene and Identification of Its Human Homologue', Journal of Biological Chemistry, 273, 329-37.

Fernandez-Canon, J. M., et al. (2002), 'Maleylacetoacetate isomerase (MAAI/GSTZ)-deficient mice reveal a glutathione-dependent nonenzymatic bypass in tyrosine catabolism', Mol Cell Biol, 22 (13), 4943-51.

Floudas, D., et al. (2012), 'The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes', Science, 336 (6089), 1715-9.

Gabadinho, Jose, et al. (2010), 'MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments', Journal of Synchrotron Radiation, 17 (5), 700-07.

Gago-Ferrero, P., et al. (2012), 'Evaluation of fungal- and photo-degradation as potential treatments for the removal of sunscreens BP3 and BP1', Sci Total Environ, 427-428, 355-63.

Garcera, A., et al. (2006), 'Saccharomyces cerevisiae cells have three Omega class glutathione S-transferases acting as 1-Cys thiol transferases', Biochem J, 398 (2), 187-96.

Gardiner, D. M., et al. (2004), 'The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus Leptosphaeria maculans', Mol Microbiol, 53 (5), 1307-18.

Gelin, M., et al. (2015), 'Combining 'dry' co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography', Acta Crystallogr D Biol Crystallogr, 71 (Pt 8), 1777-87.

Gomez, C., et al. (2011), 'In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST', Plant J, 67 (6), 960-70.

Green, A. R., et al. (2012), 'Structural understanding of the glutathione-dependent reduction mechanism of glutathionyl-hydroquinone reductases', J Biol Chem, 287 (43), 35838-48.

Hassell, A. M., et al. (2007), 'Crystallization of protein-ligand complexes', Acta Crystallogr D Biol Crystallogr, 63 (Pt 1), 72-9.

Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005), 'Glutathione transferases', Annu Rev Pharmacol Toxicol, 45 (1), 51-88.

Henrich, B., et al. (2009), 'PILATUS: A single photon counting pixel detector for X-ray applications', Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 607 (1), 247-49.

Hiscox, J., et al. (2010), 'Changes in oxidative enzyme activity during interspecific mycelial interactions involving the white-rot fungus Trametes versicolor', Fungal Genet Biol, 47 (6), 562-71.

Holm, P. J., et al. (2006), 'Structural basis for detoxification and oxidative stress protection in membranes', Journal of Molecular Biology, 360 (5), 934-45.

Huang, Y., et al. (2008), 'Maintenance role of a glutathionyl-hydroquinone lyase (PcpF) in pentachlorophenol degradation by Sphingobium chlorophenolicum ATCC 39723', J Bacteriol, 190 (23), 7595-600.

Janjusevic, L., et al. (2017), 'The lignicolous fungus Trametes versicolor (L.) Lloyd (1920): a promising natural source of antiradical and AChE inhibitory agents', J Enzyme Inhib Med Chem, 32 (1), 355-62.

Jeppesen, M. G., et al. (2003), 'The crystal structure of the glutathione S-transferase-like domain of elongation factor 1Bgamma from Saccharomyces cerevisiae', J Biol Chem, 278 (47), 47190-8.

Jhoti, H., et al. (2007), 'Fragment-based screening using X-ray crystallography and NMR spectroscopy', Curr Opin Chem Biol, 11 (5), 485-93.

Ji, Xinhua, et al. (1999), 'Structure and Function of Residue 104 and Water Molecules in the Xenobiotic Substrate-Binding Site in Human Glutathione S-Transferase P1-1', Biochemistry, 38 (32), 10231-38.

Johansson, A. S. and Mannervik, B. (2001), 'Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones', J Biol Chem, 276 (35), 33061-5.

Joung, Dae-Ki, et al. (2016), 'Antibacterial activity of oxyresveratrol against methicillin-resistant Staphylococcus aureus and its mechanism', Experimental and Therapeutic Medicine, 12 (3), 1579-84.

Kabsch, W. (2010), 'Xds', Acta Crystallogr D Biol Crystallogr, 66 (Pt 2), 125-32.

Kantardjieff, K. A. and Rupp, B. (2003), 'Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals', Protein Science, 12 (9), 1865-71.

Kantsadi, A. L., et al. (2014), 'Biochemical and biological assessment of the inhibitory potency of extracts from vinification byproducts of Vitis vinifera extracts against glycogen phosphorylase', Food Chem Toxicol, 67, 35-43.

Karathia, Hiren, et al. (2011), 'Saccharomyces cerevisiae as a Model Organism: A Comparative Study', PLoS ONE, 6 (2), e16015.

Karplus, P. A. and Diederichs, K. (2012), 'Linking crystallographic model and data quality', Science, 336 (6084), 1030-3.

Karpusas, Michael, et al. (2013), 'The Interaction of the Chemotherapeutic Drug Chlorambucil with Human Glutathione Transferase A1-1: Kinetic and Structural Analysis', PLOS ONE, 8 (2), e56337.

Katoh, Kazutaka and Standley, Daron M. (2013), 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', Molecular Biology and Evolution, 30 (4), 772-80.

Keegan, Ronan M. and Winn, Martyn D. (2008), 'MrBUMP: an automated pipeline for molecular replacement', Acta Crystallographica Section D: Biological Crystallography, 64 (Pt 1), 119-24.

Kim, K., et al. (2012), 'Glutathione s-transferase omega 1 activity is sufficient to suppress neurodegeneration in a Drosophila model of Parkinson disease', J Biol Chem, 287 (9), 6628-41.

Kolm, R. H., et al. (1995), 'ISOTHIOCYANATES AS SUBSTRATES FOR HUMAN GLUTATHIONE TRANSFERASES - STRUCTURE-ACTIVITY STUDIES', Biochemical Journal, 311, 453-59.

Kontur, W. S., et al. (2018), 'Novosphingobium aromaticivorans uses a Nu-class glutathione-S-transferase as a glutathione lyase in breaking the beta-aryl ether bond of lignin', J Biol Chem.

Kumari, V., et al. (2016), 'Irreversible Inhibition of Glutathione S-Transferase by Phenethyl Isothiocyanate (PEITC), a Dietary Cancer Chemopreventive Phytochemical', PLoS One, 11 (9), e0163821.

Ladner, J. E., et al. (2004), 'Parallel evolutionary pathways for glutathione transferases: Structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1', Biochemistry, 43 (2), 352-61.

Lallement, P. A., et al. (2014), 'The still mysterious roles of cysteine-containing glutathione transferases in plants', Front Pharmacol, 5, 192.

Lam, L. K., et al. (2012), 'Reduction of benzoquinones to hydroquinones via spontaneous reaction with glutathione and enzymatic reaction by S-glutathionyl-hydroquinone reductases', Biochemistry, 51 (25), 5014-21.

le Maire, A., et al. (2011), 'In-plate protein crystallization, in situ ligand soaking and X-ray diffraction', Acta Crystallogr D Biol Crystallogr, 67 (Pt 9), 747-55.

Lichty, J. J., et al. (2005), 'Comparison of affinity tags for protein purification', Protein Expr Purif, 41 (1), 98-105.

Lo Bello, M., et al. (2001), 'Human glutathione transferase P1-1 and nitric oxide carriers; a new role for an old enzyme', J Biol Chem, 276 (45), 42138-45.

Lorenz, P., et al. (2003), 'Oxyresveratrol and resveratrol are potent antioxidants and free radical scavengers: effect on nitrosative and oxidative stress derived from microglial cells', Nitric Oxide, 9 (2), 64-76.

Lussier, M., et al. (1997), 'Large scale identification of genes involved in cell surface biosynthesis and architecture in Saccharomyces cerevisiae', Genetics, 147 (2), 435-50.

Ma, X. X., et al. (2009), 'Structures of yeast glutathione-S-transferase Gtt2 reveal a new catalytic type of GST family', EMBO Rep, 10 (12), 1320-6.

Maafi, M. and Al-Qarni, M. A. (2018), 'Phi-order spectrophotokinetic characterisation and quantification of trans-cis oxyresveratrol reactivity, photodegradation and actinometry', Spectrochim Acta A Mol Biomol Spectrosc, 188, 64-71.

Mannervik, B. and Danielson, U. H. (1988), 'Glutathione transferases--structure and catalytic activity', CRC Crit Rev Biochem, 23 (3), 283-337.

Marinović, Mila, et al. (2018), 'Selective Cleavage of Lignin β-O-4 Aryl Ether Bond by β-Etherase of the White-Rot Fungus Dichomitus squalens', ACS Sustainable Chemistry & Engineering, 6 (3), 2878-82.

Marrs, K. A., et al. (1995), 'A GLUTATHIONE-S-TRANSFERASE INVOLVED IN VACUOLAR TRANSFER ENCODED BY THE MAIZE GENE BRONZE-2', Nature, 375 (6530), 397-400.

Marsh, M., et al. (2008), 'Structure of bacterial glutathione-S-transferase maleyl pyruvate isomerase and implications for mechanism of isomerisation', J Mol Biol, 384 (1), 165-77.

Masai, E., et al. (2003), 'Roles of the Enantioselective Glutathione S-Transferases in Cleavage of -Aryl Ether', Journal of Bacteriology, 185 (6), 1768-75.

Mashiyama, S. T., et al. (2014), 'Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere', PLoS Biol, 12 (4), e1001843.

Mathieu, Y., et al. (2013), 'Diversification of Fungal Specific Class A Glutathione Transferases in Saprotrophic Fungi', Plos One, 8 (11).

Mathieu, Y., et al. (2012), 'Characterization of a Phanerochaete chrysosporium glutathione transferase reveals a novel structural and functional class with ligandin properties', J Biol Chem, 287 (46), 39001-11.

Matthews, B. W. (1968), 'SOLVENT CONTENT OF PROTEIN CRYSTALS', Journal of Molecular Biology, 33 (2), 491-&.

McCoy, Airlie (2007), 'Solving structures of protein complexes by molecular replacement with Phaser', Acta Crystallographica Section D, 63 (1), 32-41.

McGoldrick, S., O'Sullivan, S. M., and Sheehan, D. (2005), 'Glutathione transferase-like proteins encoded in genomes of yeasts and fungi: insights into evolution of a multifunctional protein superfamily', FEMS Microbiol Lett, 242 (1), 1-12.

McNulty, J., et al. (2009), 'Isolation of flavonoids from the heartwood and resin of Prunus avium and some preliminary biological investigations', Phytochemistry, 70 (17-18), 2040-6.

McTigue, M. A., Williams, D. R., and Tainer, J. A. (1995), 'Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica

and its complex with the leading antischistosomal drug praziquantel', J Mol Biol, 246 (1), 21-7.

Menon, D. and Board, P. G. (2013), 'A role for glutathione transferase Omega 1 (GSTO1-1) in the glutathionylation cycle', J Biol Chem, 288 (36), 25769-79.

Meux, E., et al. (2013), 'New substrates and activity of Phanerochaete chrysosporium Omega glutathione transferases', Biochimie, 95 (2), 336-46.

Meux, E., et al. (2012), 'Sphingobium sp. SYK-6 LigG involved in lignin degradation is structurally and biochemically related to the glutathione transferase omega class', FEBS Lett, 586 (22), 3944-50.

Meux, E., et al. (2011), 'Glutathione transferases of Phanerochaete chrysosporium: S-glutathionyl-p-hydroquinone reductase belongs to a new structural class', J Biol Chem, 286 (11), 9162-73.

Molina, D. M., et al. (2007), 'Structural basis for synthesis of inflammatory mediators by human leukotriene C-4 synthase', Nature, 448 (7153), 613-U13.

Morel, F. and Aninat, C. (2011), 'The glutathione transferase kappa family', Drug Metab Rev, 43 (2), 281-91.

Morel, M., et al. (2009), 'The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete Phanerochaete chrysosporium', Cell Mol Life Sci, 66 (23), 3711-25.

Morel, M., et al. (2013), 'Xenomic networks variability and adaptation traits in wood decaying fungi', Microb Biotechnol, 6 (3), 248-63.

Morgenstern, R., Zhang, J., and Johansson, K. (2011), 'Microsomal glutathione transferase 1: mechanism and functional roles', Drug Metab Rev, 43 (2), 300-6.

Mosialou, E., et al. (1993), 'EVIDENCE THAT RAT-LIVER MICROSOMAL GLUTATHIONE TRANSFERASE IS RESPONSIBLE FOR GLUTATHIONE-DEPENDENT PROTECTION AGAINST LIPID-PEROXIDATION', Biochemical Pharmacology, 45 (8), 1645-51.

Mueller-Dieckmann, Christoph, et al. (2015), 'The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility', European Physical Journal Plus, 130 (4).

Mueller, Lukas A., et al. (2000), 'AN9, a Petunia Glutathione S-Transferase Required for Anthocyanin Sequestration, Is a Flavonoid-Binding Protein', Plant Physiology, 123 (4), 1561-70.

Muller, I. (2017), 'Guidelines for the successful generation of protein-ligand complex crystals', Acta Crystallogr D Struct Biol, 73 (Pt 2), 79-92.

Oakley, A. (2011), 'Glutathione transferases: a structural perspective', Drug Metab Rev, 43 (2), 138-51.

Oakley, A. J., et al. (1999), 'The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site)', J Mol Biol, 291 (4), 913-26.

Oliveira, T. I., et al. (2013), 'Molinate quantification in environmental water by a glutathione-S-transferase based biosensor', Talanta, 106, 249-54.

Osman, W. H. W., et al. (2018), 'Characterization of the glutathione S-transferases that belong to the GSTFuA class in Ceriporiopsis subvermispora: Implications in intracellular detoxification and metabolism of wood-derived compounds', Int J Biol Macromol.

Osumi, Masako (1998), 'The ultrastructure of yeast: Cell wall structure and formation', Micron, 29 (2), 207-33.

Patel, D., Bauman, J. D., and Arnold, E. (2014), 'Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery', Prog Biophys Mol Biol, 116 (2-3), 92-100.

Patskovsky, Y. V., Patskovska, L. N., and Listowsky, I. (2000), 'The enhanced affinity for thiolate anion and activation of enzyme-bound glutathione is governed by an arginine residue of human Mu class glutathione S-transferases', Journal of Biological Chemistry, 275 (5), 3296-304.

Pei, J. M. and Grishin, N. V. (2014), 'PROMALS3D: Multiple Protein Sequence Alignment Enhanced with Evolutionary and Three-Dimensional Structural Information', in D. J. Russell (ed.), Multiple Sequence Alignment Methods (Methods in Molecular Biology, 1079), 263-71.

Pellett, S. and Tracy, J. W. (2006), 'Mak16p is required for the maturation of 25S and 5.8S rRNAs in the yeast Saccharomyces cerevisiae', Yeast, 23 (7), 495-506.

Perperopoulou, F., Pouliou, F., and Labrou, N. E. (2018), 'Recent advances in protein engineering and biotechnological applications of glutathione transferases', Crit Rev Biotechnol, 38 (4), 511-28.

Petrussa, E., et al. (2013), 'Plant flavonoids--biosynthesis, transport and involvement in stress responses', Int J Mol Sci, 14 (7), 14950-73.

Pettersen, E. F., et al. (2004), 'UCSF Chimera--a visualization system for exploratory research and analysis', J Comput Chem, 25 (13), 1605-12.

Pflugrath, J. W. (2004), 'Macromolecular cryocrystallography--methods for cooling and mounting protein crystals at cryogenic temperatures', Methods, 34 (3), 415-23.

Pieper, Ursula, et al. (2004), 'MODBASE, a database of annotated comparative protein structure models, and associated resources', Nucleic Acids Research, 32 (suppl_1), D217-D22.

Polekhina, G., et al. (2001), 'Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity', Biochemistry, 40 (6), 1567-76.

Ramkumar, K., et al. (2016), 'Mechanistic evaluation and transcriptional signature of a glutathione S-transferase omega 1 inhibitor', Nat Commun, 7, 13084.

Reeves, C. D., et al. (2008), 'Genes for the biosynthesis of the fungal polyketides hypothemycin from Hypomyces subiculosus and radicicol from Pochonia chlamydosporia', Appl Environ Microbiol, 74 (16), 5121-9.

Reinemer, P, et al. (1991), 'The three-dimensional structure of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3 A resolution', EMBO Journal, 10 (8), 1997-2005.

Roret, T., et al. (2015), 'Evolutionary divergence of Ure2pA glutathione transferases in wood degrading fungi', Fungal Genet Biol, 83, 103-12.

Roth, M., et al. (2002), 'FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments', Acta Crystallographica Section D-Biological Crystallography, 58, 805-14.

Sanders, W. J., et al. (2004), 'Discovery of potent inhibitors of dihydroneopterin aldolase using CrystaLEAD high-throughput X-ray crystallographic screening and structure-directed lead optimization', J Med Chem, 47 (7), 1709-18.

Savitsky, P., et al. (2010), 'High-throughput production of human proteins for crystallization: the SGC experience', J Struct Biol, 172 (1), 3-13.

Scapin, Giovanna (2013), 'Molecular replacement then and now', Acta Crystallographica Section D: Biological Crystallography, 69 (Pt 11), 2266-75.

Schiebel, J., et al. (2016a), 'Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study', ACS Chem Biol, 11 (6), 1693-701.

Schiebel, J., et al. (2016b), 'High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits', Structure, 24 (8), 1398-409.

Schlosser, D., Grey, R., and Fritsche, W. (1997), 'Patterns of ligninolytic enzymes in Trametes versicolor . Distribution of extra- and intracellular enzyme activities during cultivation on glucose, wheat straw and beech wood', Applied Microbiology and Biotechnology, 47 (4), 412-18.

Sellam, A., et al. (2007), 'Transcriptional responses to exposure to the brassicaceous defence metabolites camalexin and allyl-isothiocyanate in the necrotrophic fungus Alternaria brassicicola', Mol Plant Pathol, 8 (2), 195-208.

Sheehan, D., et al. (2001), 'Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily', Biochem J, 360 (Pt 1), 1-16.

Shimizu, M., et al. (2005), 'Metabolic regulation at the tricarboxylic acid and glyoxylate cycles of the lignin-degrading basidiomycete Phanerochaete chrysosporium against exogenous addition of vanillin', Proteomics, 5 (15), 3919-31.

Shuker, S. B., et al. (1996), 'Discovering high-affinity ligands for proteins: SAR by NMR', Science, 274 (5292), 1531-4.

Stravodimos, G. A., et al. (2017), 'Affinity crystallography reveals the bioactive compounds of industrial juicing byproducts of Punica granatum for glycogen phosphorylase', Curr Drug Discov Technol.

Tamura, Koichiro, et al. (2013), 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', Molecular Biology and Evolution, 30 (12), 2725-29.

Terwilliger, T. C., et al. (2007), 'Ligand identification using electron-density map correlations', Acta Crystallogr D Biol Crystallogr, 63 (Pt 1), 101-7.

Thom, Russell, et al. (2001), 'The structure of a zeta class glutathione S-transferase from Arabidopsis thaliana: characterisation of a GST with novel active-site architecture and a putative role in tyrosine catabolism11Edited by I. A. Wilson', Journal of Molecular Biology, 308 (5), 949-62.

Tong, Z., Board, P. G., and Anders, M. W. (1998), 'Glutathione transferase zeta catalyses the oxygenation of the carcinogen dichloroacetic acid to glyoxylic acid', Biochem J, 331 (Pt 2), 371-4.

Umland, T. C., et al. (2001), 'The crystal structure of the nitrogen regulation fragment of the yeast prion protein Ure2p', Proc Natl Acad Sci U S A, 98 (4), 1459-64.

Vagin, Alexei and Teplyakov, Alexei (2010), 'Molecular replacement with MOLREP', Acta Crystallographica Section D, 66 (1), 22-25.

Valette, N., et al. (2017), 'Antifungal activities of wood extractives', Fungal Biology Reviews, 31 (3), 113-23.

van Beek, T. A., et al. (2007), 'Fungal bio-treatment of spruce wood with Trametes versicolor for pitch control: influence on extractive contents, pulping process parameters, paper quality and effluent toxicity', Bioresour Technol, 98 (2), 302-11.

Verdonk Marcel, et al. (2003), 'Improved protein–ligand docking using GOLD', Proteins: Structure, Function, and Bioinformatics, 52 (4), 609-23.

Wang, B., et al. (2011), 'Crystal structures and kinetic studies of human Kappa class glutathione transferase provide insights into the catalytic mechanism', Biochem J, 439 (2), 215-25.

Winn, M. D., et al. (2011), 'Overview of the CCP4 suite and current developments', Acta Crystallogr D Biol Crystallogr, 67 (Pt 4), 235-42.

Wlodawer, A., et al. (2013), 'Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination', FEBS J, 280 (22), 5705-36.

Xiao, G., et al. (1996), 'First-sphere and second-sphere electrostatic effects in the active site of a class mu gluthathione transferase', Biochemistry, 35 (15), 4753-65.

Xun, L., et al. (2010), 'S-Glutathionyl-(chloro)hydroquinone reductases: a novel class of glutathione transferases', Biochem J, 428 (3), 419-27.

Yamamoto, K., et al. (2013), 'Three-dimensional structure of a Bombyx mori Omega-class glutathione transferase', Biochem Biophys Res Commun, 438 (4), 588-93.

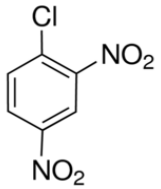
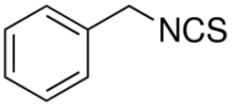
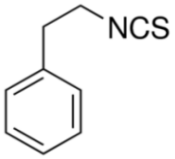
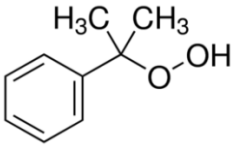
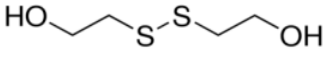
Yamamoto, T., et al. (2016), 'Oxidative trans to cis Isomerization of Olefins in Polyketide Biosynthesis', Angew Chem Int Ed Engl, 55 (21), 6207-10.

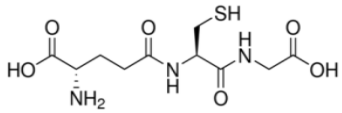
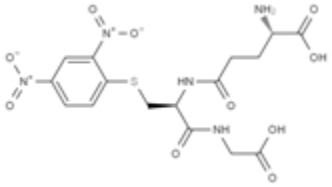
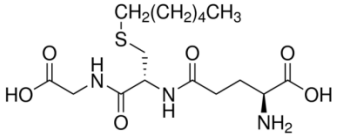
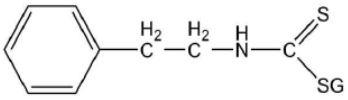
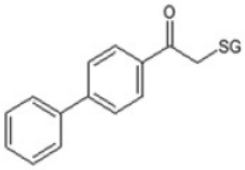
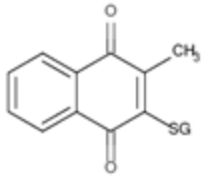
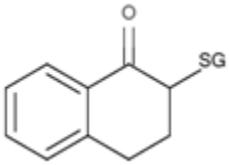
Zhang, Wei, et al. (2012), 'Self-assembly of glutathione S-transferase into nanowires', Nanoscale, 4 (19), 5847-51.

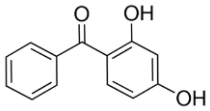
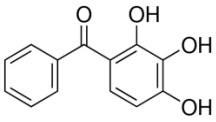
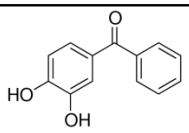
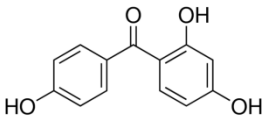
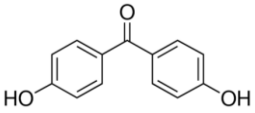
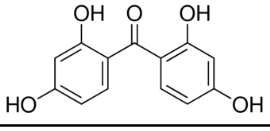
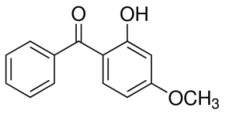
Zhou, H., et al. (2012), 'Structural insights into the dehydroascorbate reductase activity of human omega-class glutathione transferases', J Mol Biol, 420 (3), 190-203.

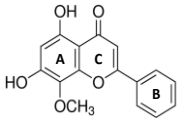
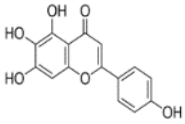
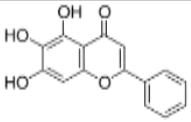
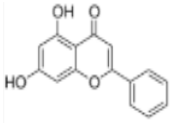
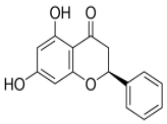
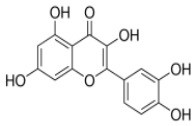
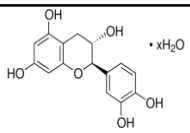
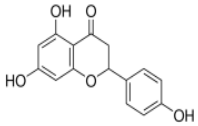
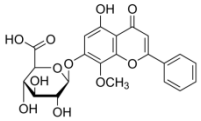
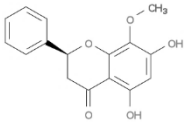
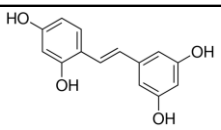
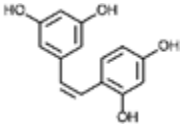
Annexes

I Structures topologiques des substrats et ligands utilisés au cours de ce projet

Substrats synthétiques	1-Chloro 2,4-dinitrobenzène	
	Benzyl-isothiocyanate	
	Phenethyl-isothiocyanate	
	Cumène hydroperoxyde	
	Hydroxy-ethyldisulfure	

Dérivés du glutathion	Glutathion (GSH)	
	Glutathionyl-dinitrobenzène (GS-DNB)	
	Glutathionyl-hexane (GS-hexane)	
	Glutathionyl-phenethyl-isothiocyanate (GS-PEITC)	
	Glutathionyl-phenylacetophenone (GS-PAP)	
	Glutathionyl-ménadione (GS-MEN)	
	Glutathionyl-tétralone	

Hydroxy-benzophénones (HBP)	2,4-HBP	
	2,3,4-HBP	
	3,4-HBP	
	2,4,4'-HBP	
	4,4'-HBP	
	2,2',4,4'-HBP	
	2-hydroxy 4-methoxy benzophénone	

Flavonoïdes	Wogonine	
	Scutellaréine	
	Baicaléine	
	Chrysine	
	Pinocembrine	
	Quercétine	
	Catéchine	
	Naringénine	
	Wogonoside	
	Dihydrowogonine	
Stilbènes	*Trans* oxyresvératrol	
	Cis oxyresvératrol	

Résumé : La détoxication est un processus biochimique qui est présent chez tous les organismes biologiques et qui leur permet d'assurer leur survie face aux xénobiotiques provenant de leur environnement. Les glutathion transférases (GST) représentent une large famille d'enzymes divisée en plusieurs classes qui participent à la phase II de détoxication en conjuguant le tripeptide glutathion au composé à éliminer. Par ailleurs, certaines GST ont un rôle non catalytique et assurent la séquestration ou le transport de molécules d'un compartiment cellulaire à un autre. Alors que l'activité catalytique des GST est étudiée depuis plusieurs décades, l'identification précise des molécules physiologiques ciblées par les GST reste un défi. Chez les organismes fongiques dégradeurs de bois, certaines classes de GST se sont étendues avec une multiplication du nombre d'isoformes codées au niveau génomique. Cette redondance serait le reflet de la diversité des molécules chimiques libérées lors de la dégradation enzymatique du bois. Au cours de cette thèse, des approches biochimiques et structurales ont été employées pour caractériser onze isoformes de GST du basidiomycète saprophyte *Trametes versicolor*. De plus, une approche utilisant des librairies de molécules a permis d'identifier une famille de ligands reconnus par ces GST, à savoir les polyphénols. Les modes d'interaction de ces ligands ont été décrits précisément à partir de la résolution de nombreuses structures cristallographiques. L'identification d'un flavonoïde à partir d'un extrait de bois de merisier (*Prunus avium*), arbre sur lequel croît *T. versicolor*, a notamment été permise par une approche de cristallographie d'affinité. Ces données suggèrent que les GST d'organismes fongiques saprotrophes pourraient prendre en charge les polyphénols libérés lors de la décomposition du bois.

Mots-clés : glutathion transférases, dégradation du bois, polyphénols, cristallographie aux rayons X, Trametes versicolor

Abstract : The ubiquitous biochemical process that enables each organism to cope with xenobiotics from its environment and thus ensures its survival is called detoxification. Glutathione transferases (GSTs) form a large family of enzymes divided into several classes. These enzymes participate in the detoxification phase II by conjugating the tripeptide glutathione to the molecule to be eliminated. Moreover, some GSTs are involved in non-catalytic processes such as sequestration or transport of molecules from one cellular compartment to another. Studies dedicated to the catalytic activity of GSTs have been ongoing for decades, yet precise identification of molecules targeted by GSTs remains challenging. In wood-decaying organisms, some of the GST classes have expanded with an increase of the number of isoforms encoded at the genomic level. This redundancy would reflect the diversity of the small molecules released upon wood enzymatic degradation. Through this thesis work, biochemical and structural approaches were used in order to characterize eleven GST isoforms from the saprotrophic fungus *Trametes versicolor*. In addition, the use of libraries of molecules helped in identifying polyphenols as a family of ligands that bind these GSTs. The molecular interaction modes were described precisely based on the resolution of numerous crystal structures. The identification of a flavonoid from an extract of the wild-cherry tree (*Prunus avium*) on which *T. versicolor* grows, was enabled by using an affinity crystallography approach. These data suggest that fungal GSTs could interact with plant polyphenols released during wood degradation.

Key-words : glutathione transferases, wood-degradation, polyphenols, X-ray crystallography, Trametes versicolor