



HAL
open science

Recherche de nouveaux antipaludiques par bioinformatique structurale et chémoinformatique : application à deux cibles : PfAMA1 et PfCCT

Émilie Pihan

► **To cite this version:**

Émilie Pihan. Recherche de nouveaux antipaludiques par bioinformatique structurale et chémoinformatique : application à deux cibles : PfAMA1 et PfCCT. Sciences agricoles. Université Nice Sophia Antipolis, 2013. Français. NNT : 2013NICE4039 . tel-01340854

HAL Id: tel-01340854

<https://theses.hal.science/tel-01340854v1>

Submitted on 2 Jul 2016

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.

Université de Nice-Sophia Antipolis



Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE NICE SOPHIA ANTIPOLIS

Discipline: Biologie, spécialité Interactions moléculaires et cellulaires

Ecole Doctorale: Sciences de la Vie et de la Santé (SVS)

**Recherche de nouveaux antipaludiques par
bioinformatique structurale et chémoinformatique.
Application à deux cibles : *PfAMA1* et *PfCCT*.**

présentée et soutenue publiquement

par

Emilie PIHAN

Le 2 Juillet 2013

Membres du jury

Pr. Serge Antonczak
Dr. Xavier Morelli
Dr. Mohammad Afshar
Dr. Rachel Cerdan
Dr. Henri Vial
Dr. Dominique Douquet

Président
Rapporteur
Rapporteur
Examinatrice
Examineur
Directrice de thèse

Remerciements

Ce travail de thèse a été effectué au sein de l'Institut de Pharmacologie Moléculaire et Cellulaire de Sophia Antipolis, sous la direction du Docteur Dominique Douguet. Il a été financé par une bourse ministérielle et a pris part à une ANR franço-hongroise nommée AddMal. Je tiens à remercier le docteur Gérard Lambeau pour m'avoir accueillie dans l'équipe "Physiopathologie Moléculaire des phospholipases A2 et de leurs médiateurs" ainsi que le Docteur Pascal Barbry pour m'avoir accueillie dans l'Institut.

Je tiens à remercier chaleureusement, en tout premier lieu, ma directrice de thèse Dominique Douguet qui m'a permise de m'épanouir en me laissant la liberté d'agir, en étant toujours disponible et en gardant un oeil bienveillant sur ma progression.

Je tiens également à remercier toutes les personnes ayant collaboré à mes recherches, ce travail ne serait pas ce qu'il est sans votre participation. Je pense bien évidemment à Marina, Alicia, Rachel, Henri, Marjorie, Maryse, Roberto, qui ont effectué les tests *in vitro*.

Je remercie vivement les Docteurs Xavier Morelli et Mohammad Afshar pour avoir accepté de rapporter mon manuscrit. Je suis très honorée de les compter parmi les membres de ce jury.

Je remercie le Professeur Serge Antonczak de me faire l'honneur de présider ma soutenance de thèse.

J'associe à ces remerciements les Docteurs Rachel Cerdan et Henri Vial pour avoir accepté de faire partie de mon jury de thèse. Qu'ils reçoivent l'expression de ma sincère gratitude.

Mes remerciements vont ensuite aux membres, passés et présents, de l'équipe du Docteur Gérard Lambeau. Un merci tout particulier à mes collègues de bureau. Julie, merci d'avoir toujours répondu présente pour nos pauses, sorties piscines reportées et autres déjeuners.

Je ne saurais conclure ces remerciements sans exprimer ma reconnaissance pour le soutien constant de ma famille. Un merci tout particulier à mes parents, mes soeurs, mes beaux-frères et leur ribambelle d'enfants!

Merci à mes amis d'école d'ingénieurs, qui ont rythmé mes années de doctorat par des voyages, des soirées, des rendez-vous au MK/GK, des sessions de beach volley et des déjeuners. Les lister serait un peu trop long, aussi soyez sûrs, que je pense à chacun d'entre vous. Merci à mes amies de Bretagne, toujours présentes lorsque je revenais aux sources ou lors de mes passages à Paris.

Et enfin merci à toi, Simon.

Table des matières

Abréviations.....	10
Introduction.....	13
I. Histoire du paludisme.....	15
II. Le parasite Plasmodium et le paludisme.....	17
1. Plasmodium, le parasite responsable du paludisme.....	17
2. Cycle de vie de Plasmodium falciparum.....	18
3. Situation actuelle mondiale.....	22
III. La pharmacopée du paludisme.....	23
1. Un peu d'histoire.....	23
1) La quinine.....	23
2) L'artémisinine.....	24
2. Médicaments antipaludiques approuvés.....	25
1) Utilisation de la base de médicaments e-Drug3D.....	25
2) La pharmacopée et les directives de l'OMS.....	25
a) les endoperoxydes : l'artémisinine et ses dérivés.....	27
b) les 4-aminoquinoléines : la quinine et ses dérivés.....	27
c) les 8-aminoquinoléines.....	29
d) les antifolates et les sulfonamides.....	29
e) les antibiotiques.....	30
f) les amino-alcools.....	30
g) autres.....	31
3) Classification ATC-DDD des antipaludiques par l'OMS.....	33
3. Médicaments en développement.....	35
1) Molécules.....	35
2) Vaccins.....	41
IV. Nouvelles cibles, nouveaux modes d'action.....	45
1. Inhibition du processus d'invasion du parasite et plus particulièrement, le complexe d'attache PfAMA1-PfRON2.....	46
1) Le processus d'invasion chez les apicomplexes.....	46
2) Les organelles de l'invasion	47
3) Les protéines de l'invasion.....	47
4) L'interaction protéine-protéine PfAMA1-PfRON2.....	48
2. Inhibition de la multiplication du parasite et plus particulièrement, une étape limitante de la biosynthèse de la phosphatidylcholine	51
1) La biogenèse membranaire chez Plasmodium.....	51
2) La voie de synthèse de novo Kennedy.....	52
a) Présentation	52
b) Molécules ciblant la voie de novo Kennedy.....	52
3) La PfCCT, enzyme limitante de la voie Kennedy.....	54

Chapitre 1 : Les chimiothèques et les outils de criblage.....57

I. Introduction.....	59
1. La chimoinformatique.....	59
2. Le criblage virtuel.....	59
1) Définition.....	59
2) Préparation des chimiothèques.....	60
3) Les filtres 'ligand-based'.....	61
4) Les filtres 'structure-based'.....	61
5) L'inspection visuelle ou la validation manuelle.....	62
3. Le protocole de criblage projet-dépendant.....	62
II. Les chimiothèques.....	64
1. Présentation.....	64
2. Préparation.....	65
1) Nettoyage des chimiothèques.....	65
2) Stéréoisomérisation.....	66
3) Génération de conformères.....	67
a) Définition.....	67
b) Omega.....	68
c) Corina.....	68
3. Préparation spécifique aux deux projets.....	69
1) Projet PfAMA1 : conformères des parties cycliques et acycliques.....	69
2) Projet PfCCT : filtre selon le volume, préparation des tautomères et des conformères de cycles.....	69
a) Filtre selon le volume.....	69
b) Génération des tautomères.....	70
c) Génération des conformères de cycles.....	71
III. Outils de criblage virtuel.....	72
1. Les approches 'ligand-based'.....	72
1) Méthode utilisant les structures 2D : empreintes moléculaires ECFP4....	72
2) Méthodes utilisant les structures 3D.....	74
a) approche basée sur la forme : ROCS (Grant et al., 1998).....	74
b) Pharmacophores.....	75
i. Qu'est-ce qu'un pharmacophore?.....	75
ii. Le programme Align-it, anciennement Pharaoh (Taminou et al., 2008).....	76
iii. Le programme PepMMsMIMIC (Floris et al., 2011).....	78
iv. Le programme Pharmer (Koes & Carlos J Camacho, 2011).....	79
2. Les approches 'structure-based'.....	80
1) Qu'est ce que l'amarrage moléculaire ('docking')?.....	80
2) Le programme Vina (Trott & Olson, 2010).....	81
3) Le programme PLANTS (Korb et al., 2006).....	83
4) Utilisation de Ligplot pour post-filtrer les solutions de docking.....	84
5) Visualisation des poses de docking : Pymol 3D.....	86

Chapitre 2 : Bloquage du processus d'invasion des érythrocytes par le parasite en ciblant le complexe d'attache AMA1-RON2 chez Plasmodium falciparum.....87

I. Contexte biologique.....89

1. Les interactions protéine-protéine (PPIs).....89
2. Problématique.....89

II. Données biologiques et structurales.....90

1. Présentation du complexe AMA1-RON2.....90
2. Résidus essentiels à l'interaction PfAMA1-PfRON2.....92
3. Analyse des structures cristallographiques.....94

III. Protocole de criblage virtuel96

1. Schéma général96
2. Premier filtre, basé sur le ligand (LBVS).....96
 - 1) Détermination du pharmacophore.....97
 - 2) PepMMsMIMIC98
 - 3) Pharmer.....99
 - 4) Align-it.....100
 - 5) Nombre de molécules retenues par chaque outil de pharmacophore....102
3. Deuxième filtre, basé sur la structure de la PfAMA1 (SBVS).....102
4. Troisième filtre visuel avec Pymol103
5. Liste des molécules achetées pour les tests in vitro.....103

IV. Validation expérimentale dans le cadre d'une collaboration. 106

1. Test in vitro de croissance parasitaire.....106
2. Test in vitro d'invasion parasitaire.....107
3. Tests d'affinité par SPR (Surface Plasmon Resonance).....108

V. Conclusion sur le criblage du complexe PfAMA1-PfRON2.....109

Chapitre 3 : Inhibition de la synthèse des phospholipides par le parasite lors du cycle érythrocytaire en ciblant la protéine CTP:phosphocholine CytidylylTransférase chez Plasmodium falciparum (PfCCT)..... 111

I. Problématique..... 113

II. Données sur la PfCCT..... 113

1. Présentation de l'enzyme PfCCT 113
2. Modélisation par homologie de la structure de la PfCCT 115
3. Analyse de l'interface du dimère..... 118
 - 1) Peptide antidimérisation..... 119
 - a) Modélisation informatique..... 119
 - b) Tests expérimentaux..... 122
 - i. Structure d'un peptide et dichroïsme circulaire..... 122
 - ii. Inhibition de l'activité catalytique en présence de peptide 122
 - 2) Mutations ponctuelles à l'interface..... 123
4. Le site actif du domaine N-terminal de la PfCCT..... 124
 - 1) Définition..... 124
 - 2) Description du site actif du domaine N-terminal de la PfCCT..... 124

III. Criblage virtuel 128

1. Schéma général 128
2. LBVS..... 129
 - 1) Criblage par similarité de forme..... 129
 - 2) Criblage par recherche de pharmacophore..... 130
 - 3) Criblage par similarité de structures..... 131
 - 4) Nombre de molécules retenues par chaque outil de LBVS..... 131
3. SBVS..... 132
 - 1) Préparation du monomère PfCCT-Nter pour le docking..... 132
 - 2) Choix des logiciels de docking..... 133
 - 3) Détermination du filtre basé sur la structure 134
 - 4) Application du filtre..... 135
4. Résultats du protocole de criblage virtuel..... 136
 - 1) Criblage de la chimiothèque e-Drug3D..... 137
 - 2) Molécules antipaludiques sans cible connue..... 137
 - a) Imidazolopiperazines..... 137
 - b) Criblage d'analogues de thymidine..... 138
 - 3) Visualisation des poses de docking avec Pymol 3D..... 139
 - 4) Molécules sélectionnées..... 140
 - 5) Synthèse d'analogues du Composé 12..... 145

IV. Validation expérimentale dans le cadre d'une collaboration.	146
1. Inhibition in vitro de la PfCCT recombinante et endogène.....	146
2. Type d'inhibition de la PfCCT recombinante.....	148
3. Test in vitro de croissance parasitaire.....	149
V. Conclusion sur le criblage de l'enzyme PfCCT.....	150
VI. La troisième enzyme de la voie de Kennedy, la PfCEPT.....	156
1. Constat de départ.....	156
2. Données récoltées dans les bases de données en ligne.....	157
1) Fonctions de la PfCEPT.....	157
2) Prédiction de la structure secondaire de la PfCEPT.....	161
3) Prédiction de la localisation des domaines transmembranaires.....	162
a) Résultats des prédictions par MEMSAT3 et MEMSAT-SVM.....	163
b) Résultats des prédictions par TMHMM.....	164
c) Résultats des prédictions par PHD.....	165
d) Résultats des prédictions par HMMTOP.....	165
e) Résultats des prédictions par TOPO2.....	166
f) Résultats des prédictions par PRED-TMR.....	166
g) Résultats des prédictions par TopPred sur Mobyle@Pasteur : prédiction des topologies des protéines membranaires.....	167
h) Résultats des prédictions par SOSUI.....	167
4) L'analyse des résultats des différentes méthodes de prédiction.....	168
a) 8 segments transmembranaires.....	168
b) 9 segments transmembranaires.....	168
3. I-Tasser, une plateforme pour la prédiction automatique de la structure par homologie et de la fonction d'une protéine (Roy et al., 2010).....	169
4. Modélisation ab initio avec l'outil Robetta (Chivian et al., 2003).....	170
Discussion.....	173
Annexes.....	179
Bibliographie.....	189

Abréviations

2,3D	2,3 Dimensions
AMA1	Apical Membran Antigen 1
AMB	Ambinter (chimiothèque)
CB	Express-Pick Chembridge (chimiothèque)
CCT	CTP:Choline-phosphate CytidylylTransférase
CEPT	Choline/Ethanolamine PhosphoTransférase
CK	Choline Kinase
CL	CombiLib Chembridge (chimiothèque)
CN	Chimiothèque Nationale (chimiothèque)
CTA	Combinaison Thérapeutique à base d'Artémisinine
CDP-choline	Cytidyl DiPhosphate-choline
CTP	Cytidyl TriPhosphate
DAG	DiAcyl Glycérol
DB	DrugBank (chimiothèque)
DC	Docking Consensus
DIMNP	Dynamique des Interactions Membranaires Normales et Pathologiques
DL	Drug-Like (chimiothèque)
ECFP	Extended-Connectivity FingerPrints
FAF-Drugs2	Free Admetox Filtering - Drugs2
FDA	Food and Drug Administration
GST	Glutathion S-Transférase
IC ₅₀	Half maximal Inhibitory Concentration
INN	International Nonproprietary Name
IRBC	Infected Red Blood Cells
LBVS	Ligand-Based Virtual Screening
MIC	Microneme protein
MMV	Medicines for Malaria Venture
MSP	Merozoite Surface Protein
MVI	Malaria Vaccine Initiative
NME	New Molecular Entity
OCT	Organic Cation Transporter
OMS	Organisation Mondiale de la Santé
PAINS	Pan-Assay Interference Compounds
PPI	Protein-Protein Interaction

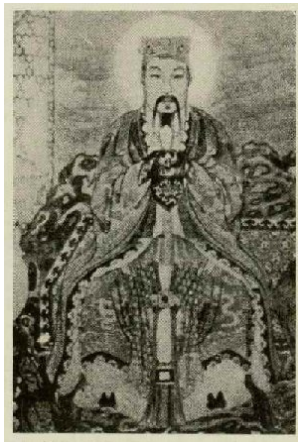
PC	Phosphatidylcholine
PDB	Protein Data Bank
PE	Phosphatidyléthanolamine
PL	Phospholipide
PLANTS	Protein-Ligand ANT System
QSAR	Quantitative Structure Activity Relationship
RMSD	Root Mean Square Deviation
ROCS	Rapide Overlay of Chemical Structures
RMN	Résonance Magnétique Nucléaire
RMSD	Root Mean Square Deviation
RON2	Rhoptry Neck Protein 2
ROP	Rhoptry Protein
SBVS	Structure-Based Virtual Screening
SDF	Structure Data File
SPR	Surface Resonance Plasmon
TPSA/PSA	Topological Polar Surface Area / Polar Surface Area
VD	Volume de Distribution
VMD	Visual Molecular Dynamics
WHO	World Health Organization

Introduction

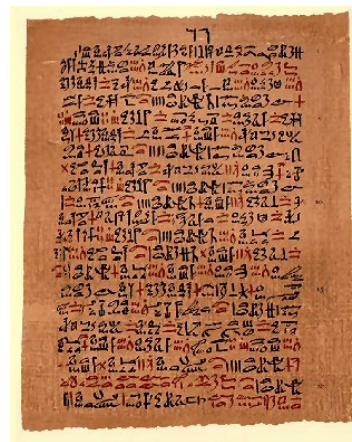
I. Histoire du paludisme

Le paludisme est une maladie parasitaire qui sévissait déjà avant l'apparition de l'homme. Ces très anciens parasites d'origine africaine auraient coévolué depuis plusieurs milliers d'années avec les anthropoïdes africains, puis avec les hommes. En effet, depuis que l'espèce humaine est entrée dans les temps historiques, toutes les grandes civilisations nous ont légué des traces écrites de leur confrontation avec le paludisme.

Les symptômes du paludisme ont été décrits pour la première fois dans Le Nei Ching, le Canon de Médecine chinoise, supposément écrit en 2700 avant J.C. par le mythique Empereur Huang Ti. On y associe la splénomégalie (augmentation du volume de la rate) à différents types de fièvres. Le paludisme y est décrit sous la forme de trois dragons symbolisant les trois phases de l'accès palustre : le premier d'entre eux tient un marteau (le mal de tête), le second porte un seau d'eau glacée (les frissons) et le troisième est ceint d'un diadème de fer porté au rouge (la fièvre).



Huang Ti



Page du papyrus d'Ebers

En 1600 avant J.C., le papyrus d'Ebers fait mention de ces mêmes symptômes et recommande « Un remède pour empêcher les moustiques de piquer : huile de moringa fraîche s'en enduire » (Ebers n° 846). Des hiéroglyphes du mur du temple de Dendara décrivent des fièvres récurrentes qui font suite à une crue du Nil excessive. Cette calamité, qui frappait annuellement les Égyptiens, a comme déterminatifs hiéroglyphiques un ciel de pluie, évoquant l'humidité et un homme à terre saignant de la tête, témoin de la gravité de cette affection. Ces descriptions semblent confirmées par l'analyse des rates de momies vieilles de 3000 ans dont certaines sont hypertrophiées. De même, le parasight TM F-test a permis le diagnostic rétrospectif et immunologique du paludisme à *Plasmodium falciparum* chez des momies africaines vieilles de 1500 à plus de 5000 ans (Miller et al., 1994).

En ce qui concerne le sous-continent indien, nous disposons d'une importante littérature médicale avec, notamment, le Charaka Samhita et le Susruta Samhita. Susruta expose une théorie de transmission des fièvres par le moustique (XVème siècle avant J.C.). La malaria y est décrite comme « la reine des maladies », attribuée généralement à la colère de Shiva. Les fièvres tierces et quartes sont clairement définies et ne laissent aucun doute quant au paludisme.

Au IVème siècle avant J.C., Hippocrate a réalisé ses premières descriptions cliniques des fièvres palustres. Au IIème siècle avant J.C., les Grecs et les Romains ont révélé une corrélation étiologique entre les fièvres intermittentes et la proximité de marécages. Le terme italien de "mal aria" (« mauvais air ») traduit bien la liaison faite entre les fièvres et les miasmes véhiculés dans l'air.

Aux XVIIIème et XIXème siècles, les premières études scientifiques sont menées en Europe. Il est d'ailleurs possible de retrouver des publications du Docteur Edwin Morris datant de 1843 sur PubMed. Pendant plus de trente ans, nombreux scientifiques ont cherché des liens entre les fièvres et l'environnement : présence ou non d'eau, de courants d'air...

La cause de la maladie sera découverte le 6 novembre 1880 à l'hôpital militaire de Constantine (Algérie) par un médecin de l'armée française, le Docteur Alphonse Laveran. Il observe le parasite dans les globules rouges de 44 personnes souffrant du paludisme. Il suggère que le paludisme est causé par ce protozoaire. Il publia le *Traité des fièvres palustres avec la description des microbes du paludisme* en 1884 et reçut le prix Nobel de physiologie ou médecine en 1907. Il introduisit aussi le terme francophone de paludisme, qui traduit la liaison "fièvres-marais" (terme latin « palus » = marais). C'est en 1897 que le médecin anglais Ronald Ross prouva que les moustiques anophèles sont les vecteurs de la malaria. Il reçut le Prix Nobel de Médecine en 1902. Il a montré comment le parasite entrait dans l'organisme, ce qui a permis d'autres découvertes importantes sur cette maladie et les moyens de la combattre.



Dr Alphonse Laveran



Dr Ronald Ross

Le paludisme affecte donc les humains depuis la nuit des temps et demeure un problème majeur de santé publique dans certaines régions du monde. Dans ce chapitre d'introduction, je détaillerai le cycle de vie du parasite *Plasmodium falciparum*, responsable du paludisme, et plus particulièrement le stade érythrocytaire, se déroulant dans l'hôte humain. En effet, des bases moléculaires et structurales de l'infection par le parasite sont nécessaires pour identifier des nouvelles cibles thérapeutiques et des nouvelles stratégies dans le traitement de la malaria. Ensuite, je présenterai les traitements historiques puis actuels en parcourant la base de données e-Drug3D (Pihan et al., 2012) ainsi que certains en étude clinique. Enfin, je vous montrerai l'intérêt des cibles thérapeutiques sur lesquelles j'ai travaillé : une protéine indispensable à l'invasion des parasites dans les globules rouges, Apical Membrane Antigen 1 chez *Plasmodium falciparum* (PfAMA1) et une enzyme essentielle à la biosynthèse des phospholipides, la CTP:phosphocholine cytidyltransferase (PfCCT). Le but de mes travaux de thèse est de collaborer à l'identification de nouvelles molécules antipaludiques par l'utilisation de la bioinformatique structurale et de la chémoinformatique.

II. Le parasite *Plasmodium* et le paludisme

1. *Plasmodium*, le parasite responsable du paludisme

Des cinq espèces pathogènes chez l'homme (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* et *P. knowlesi*), *Plasmodium falciparum* est la plus redoutable. Il est le responsable du plus grand nombre de cas et il est aussi le plus mortel et morbide. *Plasmodium falciparum* provoque les pathologies les plus sévères avec des complications pulmonaires, rénales, cérébrales et des anémies sévères.

Cet agent pathogène est un hématozoaire du genre *Plasmodium*, appartenant à l'embranchement des Apicomplexa/Sporozoa et à l'ordre des Hæmosporidae. Ce sont des parasites intracellulaires obligatoires des vertébrés transmis par des moustiques femelles. Ce parasite a un cycle de vie complexe qui requiert à la fois les humains et les moustiques pour hôtes. Le mérozoïte est le nom de la forme du parasite pendant le stade érythrocytaire. La plupart des données sur la structure des mérozoïtes a été obtenue chez *P. knowlesi*. Les mérozoïtes de cette espèce sont moins fragiles que les autres. La structure des mérozoïtes de *P. falciparum* a été décrite par Langreth en 1978. C'est une petite cellule polarisée de forme ovoïde de 1,5µm de long et 1µm de large. L'organisation structurale de Plasmodium est montrée sur la figure 1. La principale caractéristique morphologique est la présence d'un complexe apical à l'extrémité polarisée de la cellule, qui va servir lors de l'invasion. Le complexe apical consiste en deux vésicules sécrétoires, les rhoptries et les micronèmes. Ce sont les uniques organelles sécrétoires qui contiennent les produits nécessaires au mouvement, à l'adhésion aux cellules hôtes et à l'établissement de la membrane de la vacuole parasitophore qui entoure le parasite.

La phase érythrocytaire débute donc par le processus d'invasion des érythrocytes, qui met en jeu des protéines membranaires particulières, notamment AMA1 (provenant des micronèmes) et RON2 (provenant des rhoptries). Ces protéines sont nos cibles pharmacologiques et seront décrites au Chapitre 2.

Ensuite, la phase intra-érythrocytaire asexuée qui suit est responsable des symptômes cliniques du paludisme. Pendant cette étape, le parasite se multiplie et a un fort besoin en phospholipides (PL) pour construire les membranes. Les voies de biosynthèse des PL ont été étudiées. Certaines des enzymes impliquées sont prometteuses comme cibles thérapeutiques et feront l'objet du Chapitre 3.

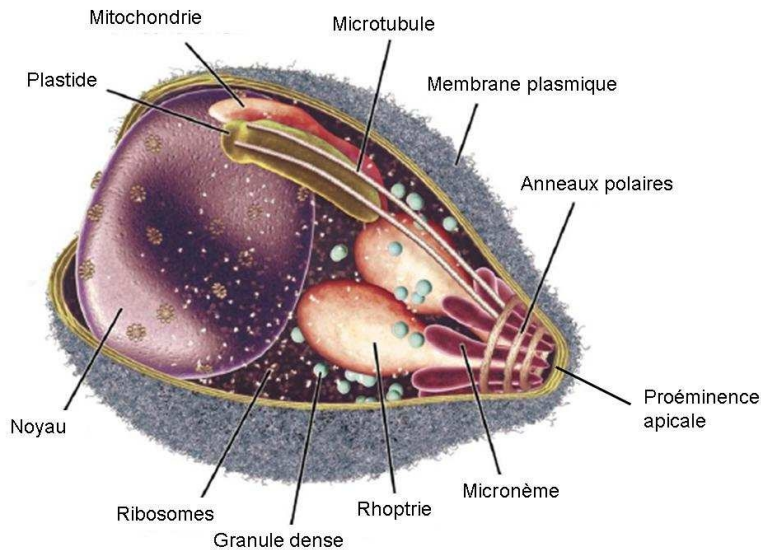


Figure 1 - Organisation structurale du mérozoïte de *P. falciparum* (Cowman & Crabb, 2006)

2. Cycle de vie de *Plasmodium falciparum*

Le cycle parasitaire chez l'anophèle commence quand un moustique femelle prend son repas sanguin sur un individu porteur de gamétocytes (Photo 1 de Hugh Sturrock, provenant du site <http://science.nationalgeographic.com/science/photos/malaria/>).

Alors que les formes sanguines asexuées sont détruites par les enzymes digestives du moustique, les gamétocytes mâles et femelles s'échappent rapidement de leur enveloppe érythrocytaire en se transformant en micro-gamètes (après exflagellation pour les mâles) ou macro-gamètes (pour les femelles) dans l'intestin moyen du moustique. La fécondation va aboutir à la formation d'un zygote dans



la demi-heure suivant la piqûre. Le zygote se transforme en ookinète allongé et mobile qui va migrer au travers de plusieurs cellules épithéliales du mésentère. Il va ensuite s'enkyster dans la paroi stomacale pour donner un oocyste. La sporogonie qui succède permet la formation de

sporozoïtes haploïdes. Au bout d'environ 10 jours, ces sporozoïtes passeront dans l'hémolymphe, certains vont être détruits par des mécanismes immunitaires du moustique tandis que d'autres vont migrer passivement jusqu'aux glandes salivaires de l'anophèle où ils se fixent. Ils s'agrègent dans les cavités situées dans la partie distale du système salivaire. Certains vont se détacher et migrer au niveau de la partie distale des canaux salivaires où ils seront transportés passivement vers la salive, prêts à être transmis au prochain hôte, lors du prochain repas sanguin. Entre 20 et 100 sporozoïtes seront inoculés à un hôte à l'occasion d'une nouvelle piqûre. Le temps de maturation chez l'anophèle est directement dépendant de la température extérieure moyenne; il s'effectue entre 6 et 20 jours (respectivement à 30 et à 20°C) avec une température minimale de 18°C nécessaire à la maturation.

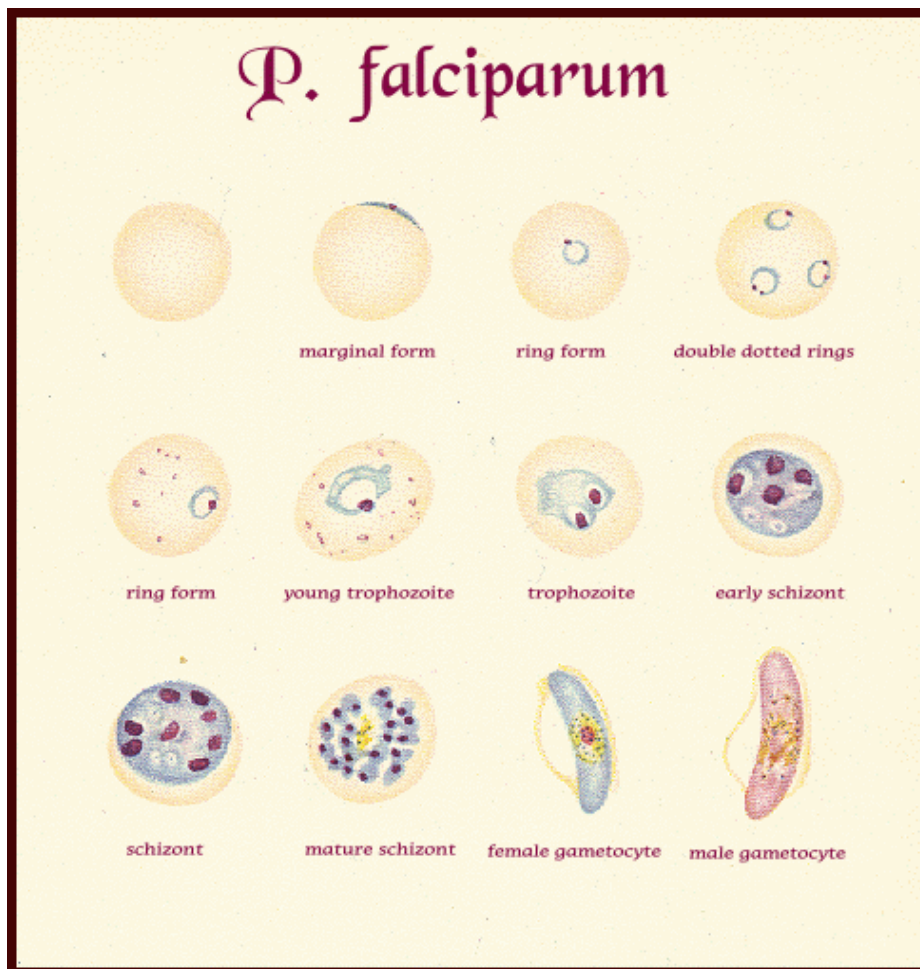


Figure 2 - Les différentes formes de *Plasmodium falciparum* au cours du cycle de vie. Cette diversité explique la difficulté de mettre au point un vaccin. (© Royal Perth Hospital, 2006)

Une fois injectés à l'homme, les sporozoïtes atteignent rapidement le foie où ils pénètrent dans les hépatocytes. Certains sporozoïtes vont cependant pénétrer dans des vaisseaux lymphatiques où ils seront dégradés par le système immunitaire. Dans l'hépatocyte, le sporozoïte entreprend un processus complexe de multiplication et de développement.

Rapidement, le sporozoïte se transforme en trophozoïte, entouré d'une membrane plasmique au sein d'une vacuole parasitophore qui va aboutir à la formation du mérozoïte. La cellule hépatique distendue et gonflée va éclater, déchargeant ainsi plusieurs milliers de mérozoïtes dans la circulation sanguine. Le mérozoïte est la deuxième forme invasive du parasite. Ces mérozoïtes ne peuvent cependant pas envahir de nouveaux hépatocytes. La durée de la schizogonie exo-érythrocytaire est caractéristique de l'espèce, avec une durée minimale de maturation de 5,5 jours chez *Plasmodium falciparum*.

Le cycle érythrocytaire se déroule à l'intérieur du globule rouge et comporte deux phases : une phase de multiplication asexuée, la schizogonie érythrocytaire suivie d'une phase de différenciation sexuée, la gamétocytogénèse (Figure 3). Le mérozoïte possède des organelles d'invasion caractéristiques des Apicomplexa (complexe apical, pellicule et «revêtement membranaire»), qui lui permettent de reconnaître et d'envahir l'érythrocyte. Le mérozoïte pénètre dans l'érythrocyte, il perd ses organelles et se différencie en une forme trophozoïte, arrondie, localisée dans la vacuole parasitophore à l'intérieur du cytoplasme érythrocytaire. Le jeune trophozoïte se transforme en schizonte, il va occuper tout le cytoplasme de l'érythrocyte. La durée totale de cette phase de développement intra-érythrocytaire est de 48 heures pour *P. falciparum*. Un schizonte mûr possède un nombre variable de mérozoïtes (16 à 32), qui est caractéristique de l'espèce *P. falciparum*. Les nouveaux mérozoïtes érythrocytaires ainsi formés sont libérés dans le système circulatoire sanguin à la fin de la schizogonie par l'éclatement du schizonte (Photo 2 de Albert Bonniers Forlag). Certains vont être détruits par le système immunitaire tandis que les autres vont de nouveau infecter des hématies afin de recommencer un cycle schizogonique.



Photo 2

La gamétocytogénèse correspond à la phase pendant laquelle les parasites libres dans la circulation sanguine peuvent se différencier en gamétocytes femelles et mâles. Le moustique prélèvera ces formes circulantes infectantes au cours d'un repas sanguin. La répétition du cycle érythrocytaire entraîne une augmentation progressive de la parasitémie qui peut atteindre et dépasser le seuil pyrogène et entraîner l'apparition de manifestations cliniques de l'accès palustre. Une caractéristique marquante du cycle érythrocytaire est le fait que les parasites ont tendance à se développer de façon synchrone, les parasites examinés sur un frottis sanguin sont fréquemment tous au même stade de développement et les symptômes cliniques, en particulier la fièvre, ont tendance à survenir à intervalles réguliers. Cette synchronicité du cycle biologique de *Plasmodium* serait en partie déterminée par le rythme circadien de l'hôte.

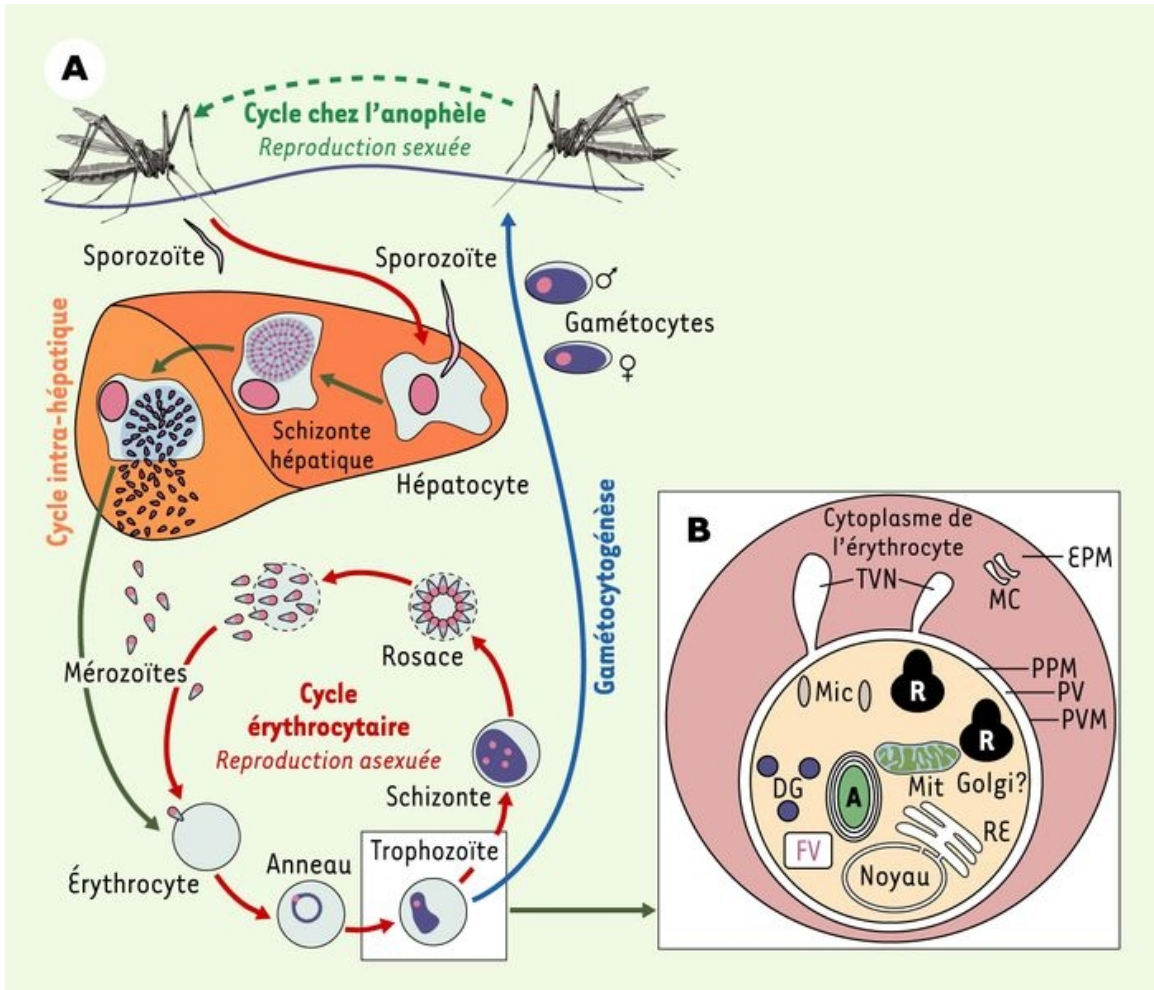


Figure 3 - Issue de (Dooren et al., 2000).

A) Cycle de vie de *Plasmodium falciparum*. Lors de son repas sanguin, le moustique du genre *Anopheles* injecte le parasite sous forme sporozoïte. Les parasites se multiplient dans les hépatocytes pendant une dizaine de jours puis sont libérés dans le sang sous forme de mérozoïtes qui envahissent les globules rouges. Cette prolifération intra-érythrocytaire qui passe par trois stades successifs (anneau, trophozoïte et schizonte) conduit à une destruction massive des hématies et constitue la phase symptomatique de la maladie. Au cours de cette phase, certains parasites se différencient en gamétocytes mâles ou femelles, qui seront ingérés par un moustique lors d'un repas sanguin sur un sujet impaludé. Après une multiplication sexuée et une migration vers les glandes salivaires, le moustique est prêt à infecter un nouvel hôte.

B) Structure du parasite *Plasmodium falciparum*. Bien que possédant sa propre membrane plasmique (PPM), le parasite se développe à l'intérieur d'une vacuole parasitophore (PV) dont la membrane (PVM) le sépare du cytoplasme érythrocytaire. Il possède les organites habituels d'un eucaryote (noyau, réticulum endoplasmique [RE] et appareil de Golgi rudimentaires, une mitochondrie [Mit]) ainsi que des organites spécifiques. Le complexe apical composé des micronèmes (Mic), des rhoptries (R) et des granules denses (DG) sert à l'invasion de la cellule hôte. La vacuole digestive (FV) permet au parasite de dégrader l'hémoglobine en hème. L'apicoplaste (A) est le vestige d'un plaste qui résulte d'une endosymbiose avec une algue rouge. Au cours de son développement, le parasite met en place dans le cytoplasme de l'érythrocyte un réseau membranaire (TVN : réseau tubulovésiculaire, MC : corpuscules de Maurer) impliqué dans le trafic moléculaire entre le parasite et la membrane de l'érythrocyte (EPM).

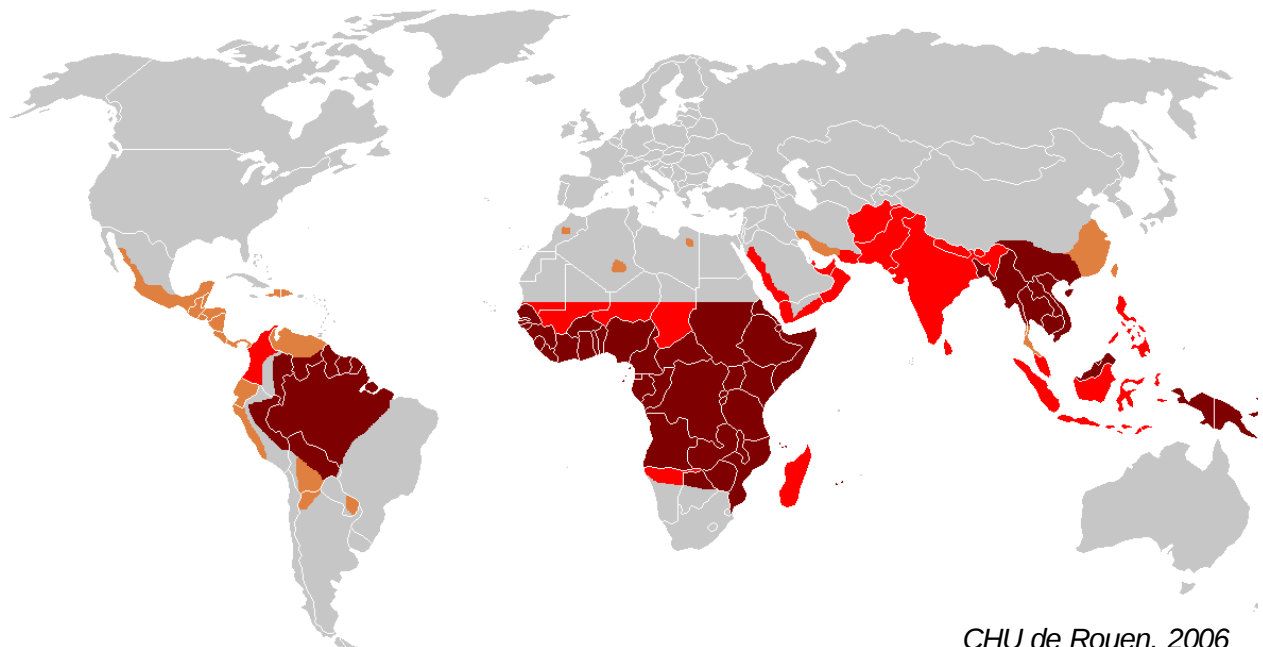
3. Situation actuelle mondiale

Malgré les efforts passés et récents, le paludisme est de loin la plus importante des maladies parasitaires tropicales dans le monde. Selon le 'World Malaria Report' 2011, il y a eu 216 millions de cas cliniques et 655 000 décès en 2010. Ce chiffre s'élevait aux alentours d'un million en l'an 2000. Cette diminution a été possible notamment par l'utilisation des moustiquaires imprégnées d'insecticides. On estime que près de la moitié de la population mondiale vit en zone d'endémie (Figure 4). La plupart des décès touche les enfants Africains de moins de 5 ans.

**VOILA UNE ARME DE DESTRUCTION MASSIVE
QUI N'INTÉRESSE PERSONNE.**



La prévention et le traitement du paludisme sont devenus très compliqués à cause de l'émergence des résistances aux médicaments développées par les parasites et celles aux insecticides développées par les moustiques. La situation est alarmante. Dans un nombre croissant de pays de l'Asie du Sud-Est, des résistances à l'artémisinine ont été reportées. Des résistances aux pyréthroides, les insecticides utilisés dans les moustiquaires et les sprays, ont aussi été reportées dans 27 pays d'Afrique et 41 pays dans le monde (World Health Organization, 2011).



CHU de Rouen, 2006

- ◆ Prévalence élevée de la chloriquino-résistance ou multi-résistance
- ◆ Chloriquino-résistance présente
- ◆ Pas de *Plasmodium falciparum* ou de chloriquino-résistance
- ◆ Pas de paludisme

Figure 4 - Répartition du paludisme et zones de résistances médicamenteuses.

III. La pharmacopée du paludisme

1. Un peu d'histoire

Deux antipaludiques extraits des plantes médicinales sont toujours parmi ceux les plus utilisés et les plus efficaces. Il s'agit de la **quinine**, qui demeure un antipaludique majeur préconisé dans le traitement des cas de paludisme grave ou d'urgence et l'**artémisinine** (qinghaosu) ainsi que ses dérivés (artéether, artésunate et artéméther) obtenus par hémisynthèse qui occupent actuellement une place primordiale lors de paludismes dits récessifs et neurologiques.

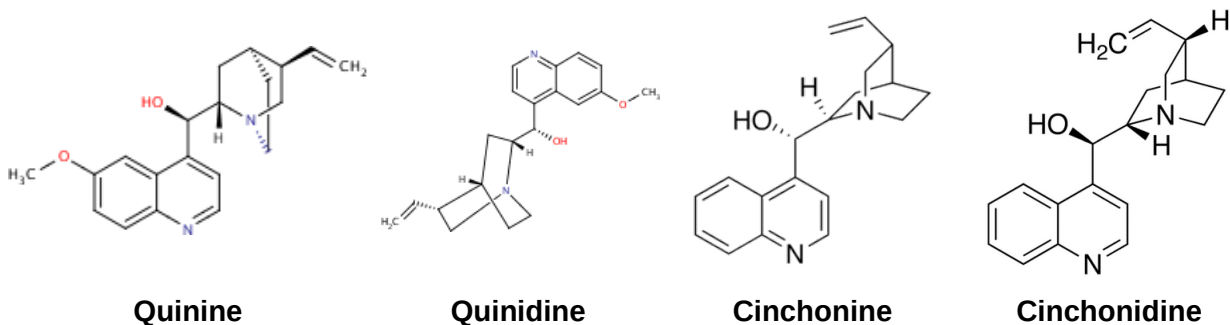
1) La quinine

Le quinquina (*Cinchona officinalis*) est un arbre de la famille des rubiacées, originaire d'Amérique du Sud. Il s'agit d'un arbre poussant sur les flancs montagneux des Andes. L'écorce de cette plante (Photo 3 de H. Zell) était utilisée par les indiens d'Amérique du Sud en médecine traditionnelle. Ce sont les jésuites espagnols qui en ont répandu l'usage au XVIème siècle. En 1630, au Pérou, Don Francisco Lopez avait observé l'utilisation par les Indiens de décoctions d'écorce de l'arbre pour leurs vertus curatives des accès fébriles. En 1820, les chimistes français Pierre Pelletier et Joseph Caventou extraient et identifient chimiquement une poudre amère de l'écorce du cinchona qu'ils appellent la **quinine**. Sa synthèse totale a été mise au point en 2001, mais économiquement peu intéressante, la quinine est encore aujourd'hui extraite de l'espèce *Cinchona* cultivée.



Photo 3

Une trentaine d'alcaloïdes ont été extraits de l'écorce de *Cinchona*, dont plusieurs sont actifs *in vitro* sur les souches de *P. falciparum*. Les quatre alcaloïdes les plus connus sont la quinine, la quinidine, la cinchonine et la cinchonidine. La combinaison des alcaloïdes (quinine, quinidine, cinchonine) est 2 à 10 fois plus efficace sur des souches quinorésistantes que chaque alcaloïde testé séparément.

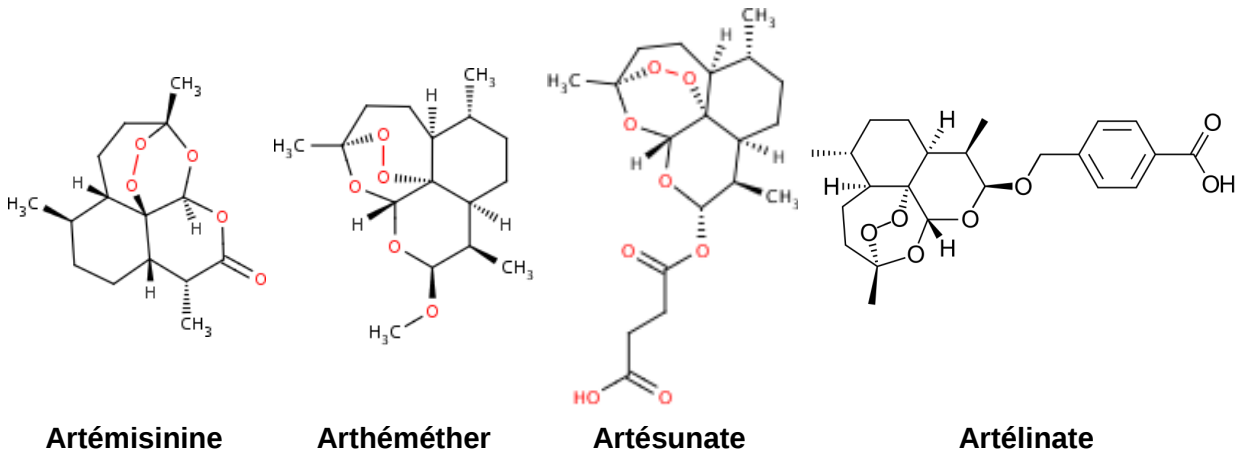


2) L'artémisinine



L'herbe verte qinghao (*Artemisia annua*, Photo 4 de Kristian Peters), connue de la pharmacopée traditionnelle chinoise et mentionnée pour la première fois en 168 av. J.-C., est redécouverte en Chine en 1971. Ils en isolent une substance cristalline appelée quighaosu, connue sous le nom d'artémisinine dont de nombreux dérivés ont été synthétisés afin de trouver des composés plus solubles. L'arthéméther est un dérivé

plus puissant que l'artémisinine, il est soluble dans les solutions huileuses et utilisé en injection intramusculaire. L'arténusate de sodium a une absorption per os très rapide, il peut également être administré par voie intraveineuse. L'artélate peut être administré par voie orale ou parentérale.



2. Médicaments antipaludiques approuvés

1) Utilisation de la base de médicaments e-Drug3D



1. Search 'malaria' in database retrieves 24 hits:

2. Search 'MALARIA' in **discarded compounds**:
FDA registration:

Figure 5 - Capture d'écran de la requête "malaria" sur la base de données e-Drug3D.

En utilisant la base de données en ligne e-Drug3D, j'ai pu lister les médicaments classés antipaludiques ayant été ou toujours commercialisés par la FDA (Food and Drug Administration). J'obtiens 24 hits qui correspondent à 19 molécules (Figure 5). En effet, une molécule peut exister avec plusieurs structures par l'énumération des différents énantiomères. Voici la liste des molécules : amodiaquine, artéméther, atovaquone, chloroquine, dapsonne, docetaxel, doxycycline, halofantrine, hydroxychloroquine, hydroxystilbamine, luméfántrine, méfloquine, primaquine, proguanil, pyriméthamine, quinidine, quinine, sulfadoxine et triméthoprim. Elles seront présentées plus en détails dans le paragraphe suivant.

Quatre molécules de cette liste (amodiaquine, halofantrine, hydroxystilbamine, sulfadoxine) ne sont plus commercialisées. Ils portent alors la mention "**DISCONTINUED**" (Annexe 1).

2) La pharmacopée et les directives de l'OMS

En 2012, Delves *et al.* ont classé une liste d'une cinquantaine de molécules approuvées ou en étude clinique (Figure 6). J'ai complété cette liste par 19 autres molécules, actuellement étudiées en clinique pour leurs propriétés antipaludiques. Les médicaments antipaludiques se classent en plusieurs groupes chimiques et il est utile d'avoir une certaine connaissance de leurs structures, de leurs propriétés chimiques et de leurs modes d'action. La classification des médicaments antipaludiques passés, actuels et futurs est difficile. Le but ici est de donner un aperçu des remèdes antipaludiques approuvés par la FDA ou préconisés par le World Health Organization (WHO=OMS), de leurs modes d'action et de leurs modes d'applications actuelles. Cependant, la liste n'est pas exhaustive. La figure 7 illustre le mode d'action des principaux antipaludiques au cours du cycle de vie du parasite.

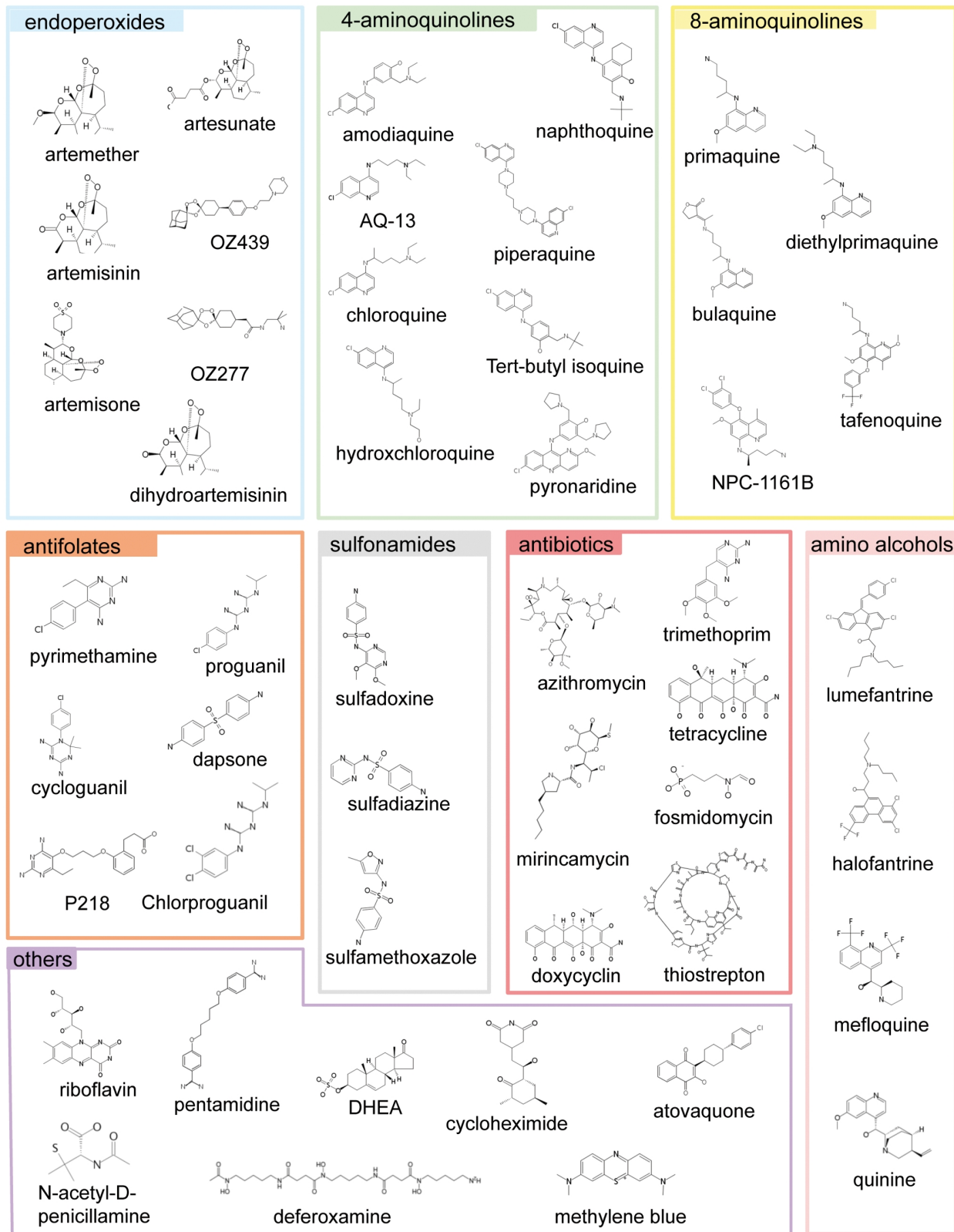
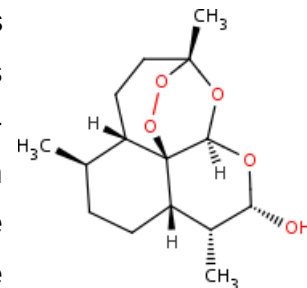


Figure 6 - Issue de (Delves et al., 2012). Les principales classes d'antipaludiques : les structures chimiques sont assemblées selon la classe chimique à laquelle ils appartiennent (endoperoxides, 4- et 8-aminoquinolines, amino-alcools) ou leur fonction (antifolates, antibiotiques) ou les deux (sulfonamides, une classe chimique d'antibiotiques utilisés dans les multi-thérapies antipaludiques).

a) les endoperoxides : l'artémisinine et ses dérivés

L'**artémisinine** est une molécule naturelle isolée d'une plante, l'*Artemisia annua*. Cette plante a été utilisée comme remède contre le paludisme depuis des milliers d'années en Chine. Elle contient d'autres molécules dérivées : l'artésunate et l'artéméther. Dans les directives pour le traitement du paludisme de l'OMS de 2007, il est bien stipulé que l'artémisinine et ses dérivés ne doivent pas être utilisés en monothérapie. En première intention, les dérivés d'artémisinine sont donc utilisés en combinaison avec d'autres molécules : la luméfantrine, l'amodiaquine, la méfloquine, la sulfadoxine-pyriméthamine. Les recommandations supplémentaires de l'OMS en 2010 stipulent que l'on doit utiliser la combinaison thérapeutique à base d'artémisinine (CTA) « **dihydroartémisinine**-pipéraquine » en première intention dans l'ensemble du monde. En deuxième intention, l'artésunate peut être couplée à des antibiotiques (tétracycline, doxycycline ou clindamycine). La CTA « artéméther-luméfantrine » peut être utilisée en prophylaxie.



Les dérivés de l'artémisinine (**artésunate**, **artéméther**) sont des schizonticides. Cette nouvelle classe d'antipaludiques de type peroxyde interfère aussi avec la digestion de l'hémoglobine, par libération de radicaux libres, toxiques pour le parasite. Les dérivés de l'artémisinine ont une action gamétocytocide, qui réduit la transmission et limite les risques de voir émerger des résistances (Woodrow et al., 2005).

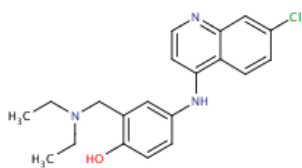
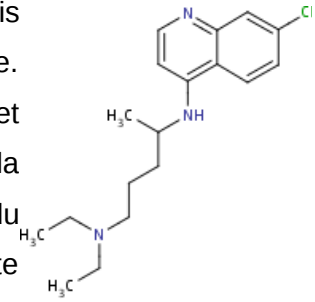
b) les 4-aminoquinoléines : la quinine et ses dérivés

La **quinine** est restée pendant longtemps le traitement de choix de l'accès palustre par sa rapidité d'action. Elle a été utilisée durant plus de trois siècles. Jusqu'aux années 30, c'était le seul médicament efficace pour le traitement du paludisme. C'est le seul médicament qui est resté largement et longtemps efficace pour le traitement des accès graves. Actuellement, elle est seulement utilisée dans le cadre du traitement du paludisme grave à *P. falciparum*, en partie à cause des effets secondaires indésirables. En effet, en Afrique, dans les années 30 et 40, les gens prenaient la quinine quand ils se croyaient atteints du paludisme, et la combinaison des infections répétées avec *P. falciparum* et le traitement inadapté avec la quinine, conduisaient, chez certains sujets, au développement d'hémolyse intravasculaire massive aiguë et d'hémoglobinurie, par exemple la fièvre bilieuse hémoglobinurique.

Plusieurs **aminoquinoléines** ont été synthétisées à partir du cycle de la quinine : **chloroquine**, **amodiaquine**, **méfloquine**. Lors du traitement par la quinine ou ses dérivés, ces produits se concentrent à l'intérieur des érythrocytes et tuent le parasite à son stade de développement schizonte. Ces molécules interfèrent avec la digestion de l'hémoglobine dans la

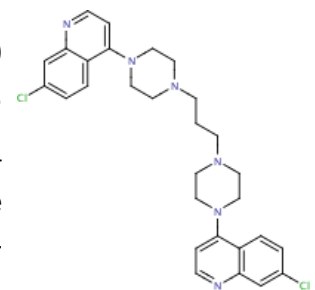
vacuole nutritive en inhibant la formation de l'hémozoïne. L'hémozoïne est un complexe formé par l'association des hèmes contenant le fer, toxique pour la parasite.

Parmi elles, la **chloroquine** a été utilisée pour la première fois dans les années 40, peu de temps après la Seconde Guerre mondiale. Elle est efficace pour la guérison de toutes les formes de paludisme et provoque peu d'effets secondaires quand elle est administrée selon la dose recommandée. Pour la campagne globale d'éradication du paludisme, ce fut la molécule la plus fréquemment utilisée. Cette campagne lancée en 1955 consistait en l'utilisation de la chloroquine en plus des mesures de contrôle du vecteur, c'est à dire les moustiques. Mais ces méthodes ont amené des résistances : les anophèles aux insecticides suite aux méthodes d'éradication de masse avec notamment l'utilisation du dichlorodiphényltrichloroéthane (DDT, 1955-1969) et la chimiorésistance de *P. falciparum* à la chloroquine. La lutte anti-vectorielle a échoué dans la plupart des pays tropicaux et est aujourd'hui pratiquement abandonnée. La résistance de *P. falciparum* à la chloroquine est apparue vers 1957 à la frontière entre la Thaïlande et le Cambodge et en Colombie (Payne, 1987) puis s'est étendue à l'Asie du Sud-Est, Centrale et Occidentale et à l'Amérique du Sud. A la fin des années 70, la chloroquino-résistance s'est observée en Afrique de l'Est et se propage depuis lors d'Est en Ouest à travers le continent (Wernsdorfer, 1991). La campagne d'éradication a donc été arrêtée pendant les années 70 et la chloroquine est maintenant utilisée en combinaison avec d'autres molécules. La plupart des souches de *P. falciparum* sont de nos jours résistantes à la chloroquine mais elle reste utilisée en association avec la **primaquine** pour traiter le paludisme à *P. vivax*.



L'efficacité de l'**amodiaquine** est meilleure que celle de la chloroquine, c'est pourquoi elle est utilisée en CTA avec l'artésunate dans les programmes de contrôle du paludisme en Afrique. Selon les recommandations de l'OMS de 2010, il est aussi préférable d'utiliser l'amodiaquine dans une CTA couplée à la sulfadoxine-pyriméthamine (trithérapie).

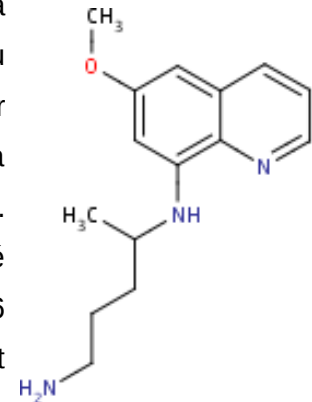
La **pipéraquline** est une bisquinoléine. Elle n'est utilisée qu'en association avec la dihydroartémisinine (DHA, aussi appelée arténimol) en première intention. Le mécanisme d'action exact de la pipéraquline n'est pas élucidé mais il est probablement identique à celui de la chloroquine, qui est un analogue structural proche. La structure bisquinolone volumineuse de la pipéraquline peut être importante pour l'activité sur les souches résistantes à la chloroquine.



c) les 8-aminoquinoléines

Le contrôle du paludisme est de plus en plus complexe à cause de l'émergence de souches de *Plasmodium falciparum* chimiorésistantes.

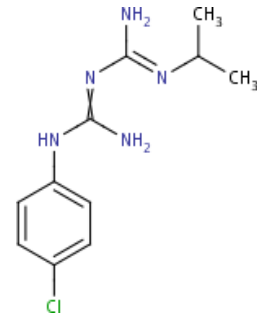
La **primaquine** est une 8-aminoquinoléine qui bloque la transmission du paludisme puisqu'elle agit sur les formes hépatiques du parasite ainsi que sur la gamétocytogénèse. Elle est recommandée par l'Organisation Mondiale de la Santé (OMS) pour contrôler la transmission du paludisme à *P.vivax* dans les zones endémiques. Cependant, elle nécessite un suivi médical à cause de la toxicité hémolytique de cette molécule chez les patients déficients en glucose-6 phosphate déhydrogénase (G6PD), particularité génétique qui prévaut dans les zones endémiques.



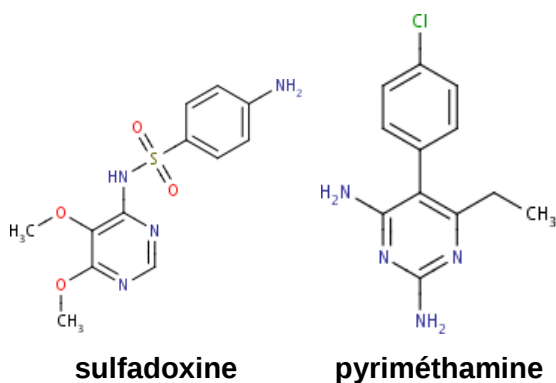
d) les antifolates et les sulfonamides

Les antifolates, comme le **proguanil** ou la **pyriméthamine**, agissent au niveau de la voie de synthèse des acides nucléiques du parasite (purine et thymine). Les antifolates inhibent la dihydrofolate réductase (DHFR), nécessaire à la synthèse de l'acide tétrahydrofolique (THF).

Le **proguanil** appartient aux antipaludiques de la classe des biguanides et il a été synthétisé pour la première fois en 1946. Il a une chaîne biguanide reliée à un noyau chlorophényl. Ce médicament est toujours utilisé en prophylaxie en association avec l'atovaquone.



Les sulfonamides, comme la sulfadoxine, sont aussi des antifolates. Les sulfonamides inhibent la dihydroptéroate synthétase (DHPS) qui produit l'acide dihydrofolique (DHF).



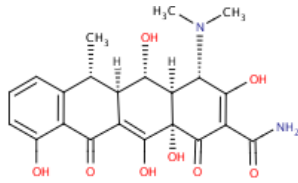
sulfadoxine

pyriméthamine

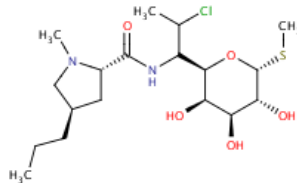
La plus ancienne des associations, la **sulfadoxine-pyriméthamine (Fansidar®)** n'est plus utilisée seule en raison des résistances. En effet, elle est capable de sélectionner rapidement des mutants résistants. Les nouveaux antipaludiques qui font l'objet de développements récents sont généralement associés, au moins en bithérapie. La CTA recommandée est artésunate couplée à l'association sulfadoxine-pyriméthamine.

e) les antibiotiques

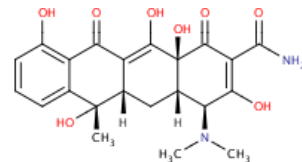
Cette classe contient entre autres la **doxycycline**, la **clindamycine** et la **tétracycline**. Ils inhibent la reproduction du parasite. Ces molécules sont utilisées en association à l'artésunate ou à la quinine en traitement de deuxième intention, selon les recommandations de l'OMS. Associées à la quinine, la doxycycline ou la clindamycine peuvent être utilisés en prophylaxie. Ces molécules empêchent le fonctionnement normal de l'apicoplaste de *Plasmodium falciparum* (Fichera & Roos, 1997).



doxycycline



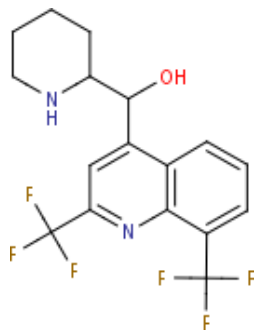
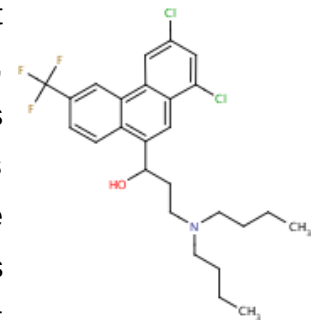
clindamycine



tétracycline

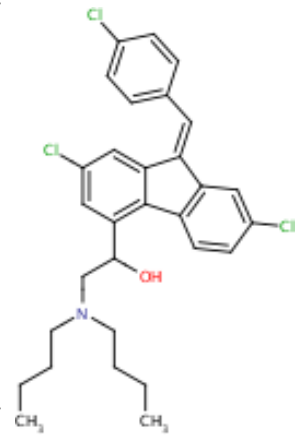
f) les amino-alcools

L'**halofantrine** appartient à la classe des composés phénanthrènes-méthanols. C'était un antipaludique efficace introduit dans les années 80, mais à cause de sa courte demi-vie de 1 à 2 jours, il n'était pas approprié à la prophylaxie. Malheureusement, des souches résistantes ont été de plus en plus signalées et il y avait quelques inquiétudes liées aux effets secondaires. L'halofantrine a donc été retirée du marché pour avoir été associée à des troubles neuropsychiatriques, à des douleurs abdominales, diarrhées, prurits et éruptions cutanées.



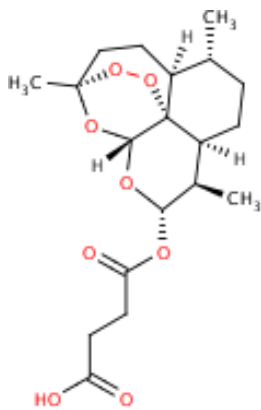
Introduite en 1971, la **méfloquine**, un dérivé de quinoline-méthanol, est structuralement proche de la quinine. Le médicament était très efficace contre le paludisme résistant aux autres traitements lorsqu'il fut introduit pour la première fois. En raison de sa longue demi-vie, la méfloquine était un bon médicament pour la prophylaxie antipaludique, mais l'apparition d'une résistance à grande échelle, en plus de divers effets secondaires indésirables, a réduit considérablement son utilisation. A cause de sa similarité avec la quinine, les deux médicaments ne doivent pas être administrés ensemble. Il a été rapporté divers effets secondaires indésirables. Elle est utilisée en CTA avec l'artésunate.

La **luméfantrine** est utilisée dans une CTA couplée à l'artéméter en prophylaxie, ou en traitement de première intention. Cette thérapie combinée cible les stades érythrocytaires de *Plasmodium*, même pour les souches chloriquino-résistantes. L'artéméter a une action rapide et est rapidement éliminée de l'organisme. La luméfantrine a un temps de demi-vie plus long et touche les parasites résiduels. Le mécanisme exact n'est pas connu, mais certains travaux laissent à penser qu'elle inhibe la formation de β -hématine en formant un complexe avec l'hémine. Elle inhiberait la synthèse des acides nucléiques et des protéines.



g) autres

L'**atovaquone** est une hydroxynaphtoquinone. C'est un inhibiteur puissant des fonctions mitochondriales en bloquant la chaîne de transfert d'électrons en se fixant sur une sous-unité du cytochrome b (CYTb). Elle a peu d'impact thérapeutique lorsqu'elle est utilisée seule. En combinaison avec un métabolite (proguanil), une intéressante synergie d'action s'observe grâce à une inhibition séquentielle de la synthèse des pyrimidines. Ce composé est utilisé en prophylaxie (Malarone®).



La **malarone** a été mise en circulation en Australie en 1998. L'atovaquone était déjà disponible en 1992 et était utilisée avec succès pour le traitement de *Pneumocystis carinii*. Cette combinaison a fait l'objet de quelques essais cliniques et a été trouvée efficace à 95%, même pour les formes résistantes de *P. falciparum*. Cette association est maintenant un générique.

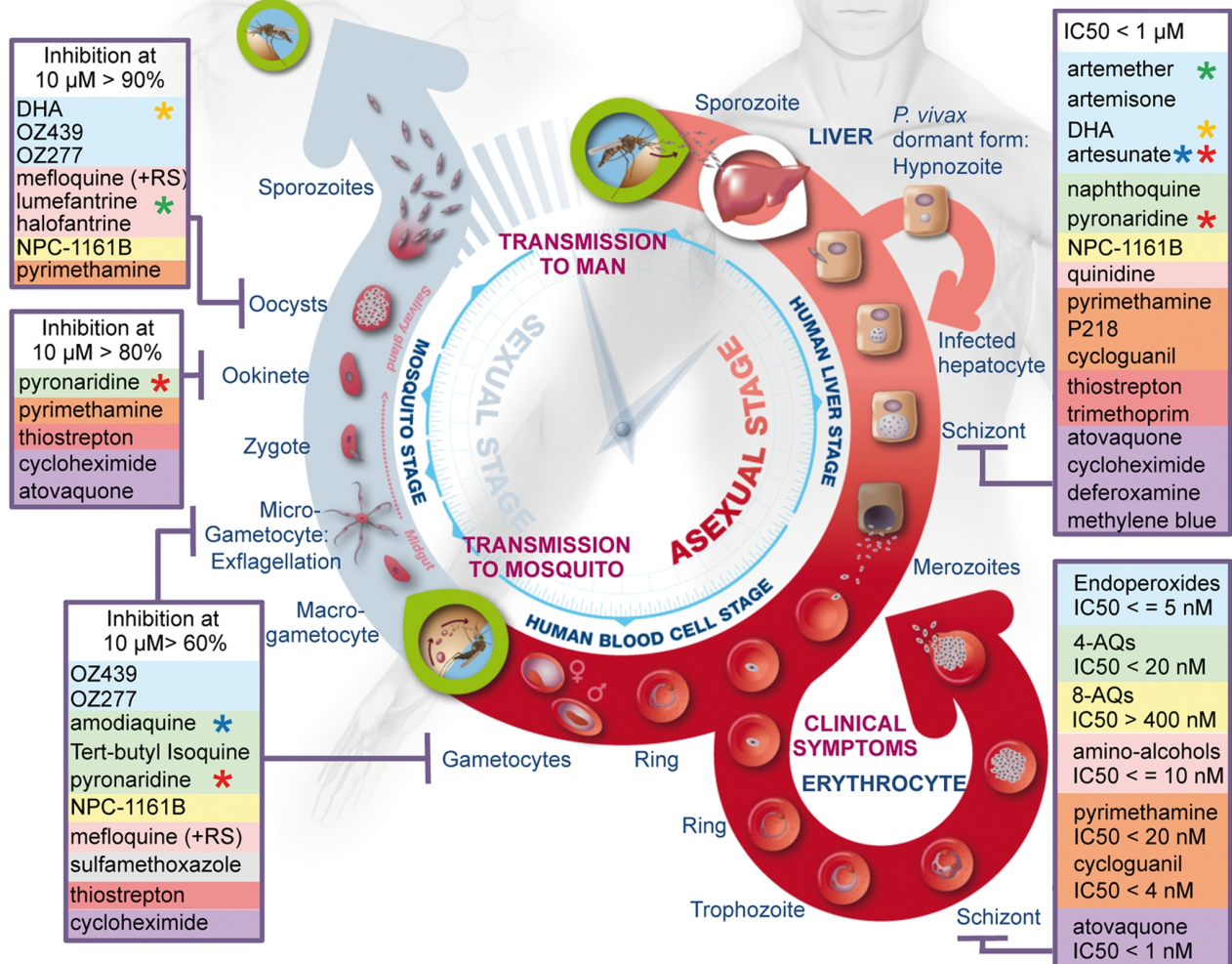


Figure 7 - Issue de (Delves et al., 2012). Résumé de l'activité sur le cycle de vie de *Plasmodium* des antipaludiques les plus utilisés. Les trois phases majeures (hépatique, érythrocytaire et vectorielle) du cycle de vie de *Plasmodium* sont présentées ici. Les deux points d'entrée menant à la transmission du parasite du vecteur à l'hôte et de l'hôte au vecteur sont indiqués par des cercles verts. Les formes du parasite, spécifiques de chaque phase sont précisées, et les molécules inhibitrices du développement de ces formes sont listées dans les encarts et colorées de la même façon que dans la figure 6. Les étoiles mettent en avant les composants des combinaisons basées sur l'artémisinine : vert (coartem), rouge (pyramax), orange (eurartesim), bleu (ASAQ).

3) Classification ATC-DDD des antipaludiques par l'OMS

ATC-DDD signifie Anatomique Thérapeutique Chimique-Defined Daily Dose. Dans cette classification, les médicaments ont un ou des identifiants. Ils sont répartis en plusieurs groupes selon l'organe ou le système sur lequel ils agissent et selon leurs propriétés chimiques, pharmacologiques et thérapeutiques. De manière générale, les médicaments sont classés en groupes dans cinq niveaux différents. Pour qu'une molécule apparaisse dans ce classement, il faut qu'une demande d'autorisation de mise sur le marché ait été soumise au moins dans un pays. Les compléments alimentaires et les produits de médecine traditionnelle ne sont, en général, pas inclus dans le système ATC. De plus, la priorité a été donnée aux préparations contenant des substances bien définies qui ont un nom INN (International Nonproprietary Name) et qui sont de nouvelles entités chimiques (NCE) ou biologiques autorisées à être commercialisées dans plusieurs pays, ou des substances existantes bien définies utilisées dans divers pays.

P ANTIPARASITIC PRODUCTS, INSECTICIDES AND REPELLENTS	
P01 ANTIPROTOZOALS	
P01B ANTIMALARIALS	
P01BA Aminoquinolines	
P01BA01	chloroquine
P01BA02	hydroxychloroquine
P01BA03	primaquine
P01BA06	amodiaquine
P01BB Biguanides	
P01BB01	proguanil
P01BB02	cycloguanil embonate
P01BB51	proguanil, combinaisons
P01BC Methanolquinolines	
P01BC01	quinine
P01BC02	mefloquine
P01BD Diaminopyrimidines	
P01BD01	pyrimethamine
P01BD51	pyrimethamine, combinaisons
P01BE Artemisinin and derivatives, plain	
P01BE01	artemisinin
P01BE02	athemether
P01BE03	artesunate
P01BE04	artemotil
P01BE05	artenimol
P01BF Artemisinin and derivatives, combinaisons	
P01BF01	artemether et lumefantrine
P01BF02	artesunate et méfloquine
P01BF03	artesunate et amodiaquine
P01BF04	artesunate, sulphamethopyrazine et pyrimethamine
P01BF05	artenimol et pipéraquline
P01BF06	artesunate et pyronaridine
P01BX Other antimalarials	
P01BX01	halofantrine

Le premier niveau qui nous concerne ici est P pour produits antiParasitaires, insecticides et révulsifs. Le deuxième niveau correspond au sous-groupe thérapeutique, qui est 01 pour les antiprotozoaires. Le troisième niveau est le sous-groupe pharmacologique B pour les antipaludiques. S'en suivent les divers groupes chimiques pour le quatrième niveau, au nombre de sept. Enfin, au cinquième niveau, on retrouve les différentes substances chimiques décrites précédemment.

Comme vous avez pu le voir dans la section précédente, il existe d'autres molécules utilisées dans la lutte contre le paludisme. Cependant, leur application première n'étant pas cette maladie, ces médicaments sont parfois classés dans d'autres groupes.

Par exemple, la tétracycline qui est un antibiotique utilisé en traitement de deuxième intention combinée à l'artésunate ou à la quinine, n'est pas classée dans les antipaludiques mais dans diverses autres classes : A01AB13, D06AA04, J01AA07, S01AA09, S02AA08, S03AA02.

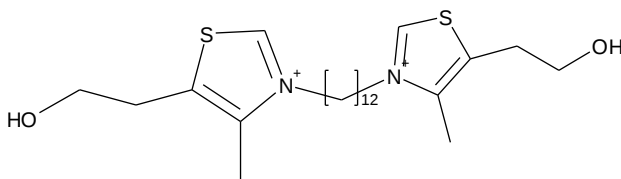
<p>A ALIMENTARY TRACT AND METABOLISM A01 STOMATOLOGICAL PREPARATIONS A01A STOMATOLOGICAL PREPARATIONS A01AB Antiinfectives and antiseptics for local oral treatment A01AB13 tetracycline</p>
<p>D DERMATOLOGICALS D06 ANTIBIOTICS AND CHEMOTHERAPEUTICS FOR DERMATOLOGICAL USE D06A ANTIBIOTICS FOR TOPICAL USE D06AA Tetracycline and derivatives D06AA04 tetracycline</p>
<p>J ANTIINFECTIVES FOR SYSTEMIC USE J01 ANTIBACTERIALS FOR SYSTEMIC USE J01A TETRACYCLINES J01AA Tetracyclines J01AA07 tetracycline</p>
<p>S SENSORY ORGANS S01 OPHTHALMOLOGICALS S01A ANTIINFECTIVES S01AA Antibiotics S01AA09 tetracycline</p>

3. Médicaments en développement

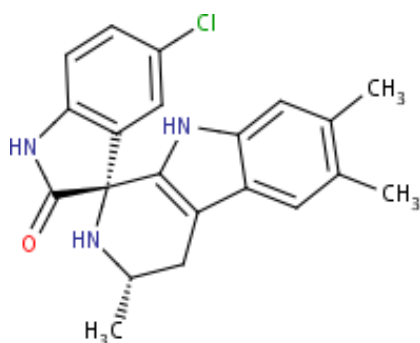
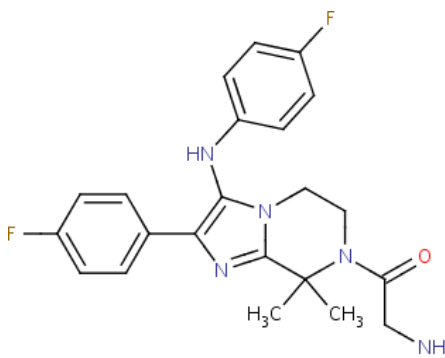
1) Molécules

Il y a d'ores et déjà des molécules présentant des nouvelles structures chimiques en développement clinique. Celles enregistrées en étude clinique enregistrées au service du Clinical Trials du NIH (<http://clinicaltrials.gov/>) en janvier 2013 sont présentées ici.

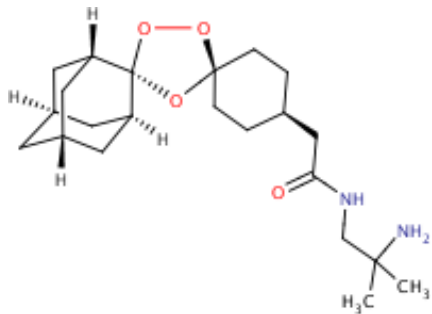
La molécule **SAR97276A** bithiazolinium (T3) est développée chez Sanofi Aventis. Elle cible le transporteur de cations *PfOCT* et la choline kinase *PfCK*. De cette manière, elle bloque la biosynthèse des phospholipides via la voie *de novo* Kennedy, essentielle au parasite. Elle a été découverte par l'équipe d'Henri Vial à Montpellier (Alberge et al., 2010). Cette molécule est entrée en essais cliniques de phase II.



Certaines **imidazolopiperazines** sont expérimentalement actives contre les stades sanguin, hépatique et gamétocyte de *Plasmodium falciparum*. Plus précisément, le **GNF156 (KAF156 Novartis)**, semble n'inhiber ni la biosynthèse protéique ni le cytochrome bc1 mais bloque le parasite dans ses stades précoces de développement. Son IC_{50} est de 10nM pour la souche 3D7 de *P.falciparum* et sa biodisponibilité orale est estimée à 67% chez les hommes. Cette molécule est entrée en essais cliniques de phase II.

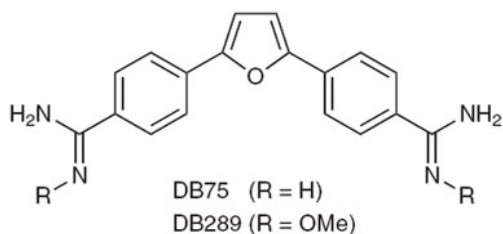
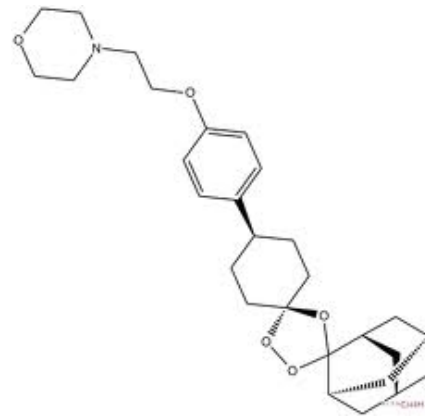


Le composé **NITD609 (KAE609, Novartis)**, appartient à la classe de spiroindolones. Il provient d'un criblage à haut débit de molécules naturelles (Rottmann et al., 2010). Il est actuellement en étude clinique de phase II. Ce composé bloque la synthèse protéique du parasite, en inhibant la protéine *PfATP4*. Il a de très bonnes propriétés pharmacocinétiques compatibles avec une seule prise orale par jour pour un traitement potentiel du paludisme à *P. falciparum* ou à *P. vivax*.



L'**artémolane (RBx11160 ou OZ277)** est un peroxyde synthétique. Il a été l'objet d'essais cliniques de phase II, afin d'identifier la bonne dose à administrer. Des doses quotidiennes de 100 et 200mg sont rapidement actives, efficaces et sûres sur tous les stades sanguins de *P.falciparum* sans effet sur les stades hépatiques. Cette molécule inhibe la PfATP6, une ATPase calcium sarcoplasmique du réticulum endoplasmique de *P.falciparum*. Elle s'accumule dans la vacuole digestive du parasite. Elle représente une alternative aux dérivés d'artémisinines dans les thérapies antipaludiques combinées (Valecha et al., 2010). Elle a été approuvée en Inde en 2011 combinée à la pipéraquline sous le nom commercial Synriam™.

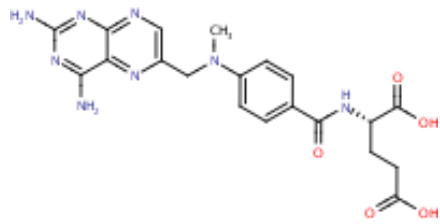
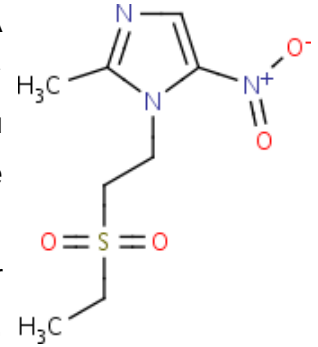
Le composé **OZ439**, développé par Medicines for Malaria Venture, est aussi un peroxyde synthétique, structuralement proche de l'artémolane. Il est en étude clinique de phase II sur *P.falciparum* et *P.vivax*. Il serait capable de guérir en une seule dose le paludisme chez l'animal et a été mis au point par une équipe internationale de chercheurs australiens, suisses, anglais et américains. Leur étude montre que l'antipaludique OZ439 est actuellement le produit le plus puissant pour détruire le parasite *Plasmodium falciparum* responsable de la maladie, et cela à tous les stades de son développement dans le sang. La phase I des essais cliniques a montré que ce composé était sans risque jusqu'à une dose unique de 1600mg. A cette dose, les concentrations plasmatiques sont telles que l'activité antiparasitaire dure 72 heures (Charman et al., 2011). Dans cette étude clinique, il s'agit aussi de trouver le meilleur partenaire pour une combinaison (pipéraquline, ferroquine, naphthoquine).



Le composé **DB289**, ou **pafuramide maléate**, est une diamidine entrée en essais cliniques pour une action prophylaxique contre le paludisme. C'est une prodrogue orale du métabolite actif connu, le DB75. Ce composé a été jusqu'en phase II où il a été arrêté car une parasitémie chez les patients a indiqué un échec

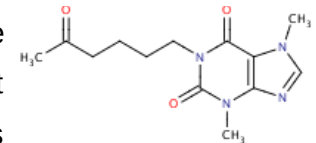
dans la prophylaxie. Le composé DB75 agirait en ciblant le noyau du parasite et la synthèse d'acides nucléiques (Purfield et al., 2009).

Le **tinidazole** est déjà une molécule approuvée par la FDA comme antibactérien pour le traitement des infections par *Trichomonas vaginalis*, *Entamoeba histolytica* and *Giardia lamblia*. En relarguant du nitrile, il crée des dommages à l'ADN. Il existe une unique étude publiée en Inde en 1988 montrant l'efficacité du tinidazole (2g par voie orale) contre *P.vivax* (SARMA, 1988). Un essai clinique en Thaïlande sur *P.vivax* a débuté en 2008, est allé jusqu'en phase II et s'est fini en 2012, mais aucun résultat n'est disponible.

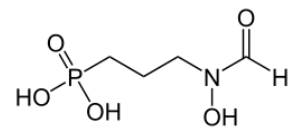


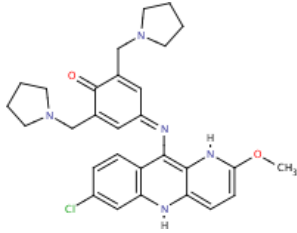
Le **méthotrexate** est déjà approuvé par la FDA comme anticancéreux. Cette molécule pourrait être utilisée contre le paludisme à des doses faibles, car c'est un inhibiteur de la dihydrofolate réductase. En effet, un essai clinique de phase I a montré qu'une faible dose (5mg par jour pendant 5 jours) de méthotrexate avait un profil pharmacocinétique sain sur des volontaires Kenyans. Cependant, les concentrations plasmatiques n'atteignent pas les 250-400nM requis pour éliminer l'infection paludique *in vivo*. D'autres études sont donc nécessaires pour confirmer que le méthotrexate pourrait servir dans le traitement contre le paludisme (Chilengi et al., 2011).

La **pentoxifylline** a été proposée comme thérapie complémentaire pour la malaria sévère. Jusqu'à aujourd'hui, elle a fait l'objet de plusieurs études cliniques sur les adultes, mais les résultats étaient partagés, le nombre de patients inclus était faible et la dose optimale n'avait pas été établie. Plus récemment, un essai clinique de phase II chez des enfants âgés de 9 à 96 mois atteints de neuropaludisme s'est déroulé en Afrique. Il s'agissait de déterminer la dose adéquate, c'est-à-dire la dose maximum tolérée, qui minimise les effets secondaires et permet une efficacité maximale. C'est cette dose qui sera utilisée dans une autre étude de phase II multicentres plus importante et une étude de phase III (Lell et al., 2010). Concernant son mode d'action, la pentoxifylline n'a pas d'action directe sur le parasite. En augmentant la déformabilité des érythrocytes, elle rend probablement plus difficile l'étape d'invasion par les parasites.



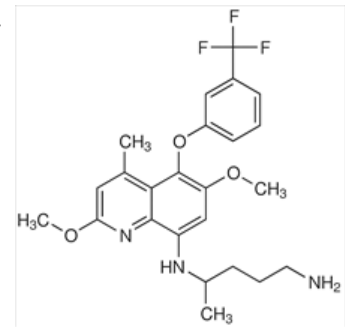
La **fosmidomycine** est un dérivé de l'acide phosphonique. Une étude clinique de phase II a montré qu'une combinaison fosmidomycine-clindamycine était un traitement sûr et efficace (Na-Bangchang et al., 2007). Cette combinaison est actuellement en étude clinique de phase III.



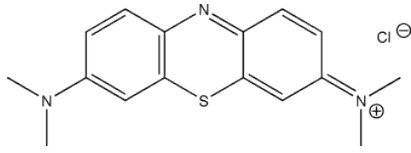


La **pyronaridine (ou malaridine)** est une 4-aminoquinoléine, qui inhibe la digestion de l'hème, toxique pour le parasite. Elle est actuellement en étude clinique de phase III en combinaison avec un ratio 3:1 avec l'artésunate, un dérivé d'artémisinine. Elle est efficace sur des souches résistantes à la chloroquine et à l'amodiaquine. Il n'y a pas de toxicité importante, et il existe une synergie avec les dérivés d'artémisinines (Croft et al., 2012). C'est un candidat idéal pour une nouvelle thérapie en combinaison avec l'artésunate.

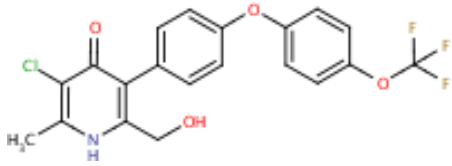
Le **tafénoquine** est une 8-aminoquinoléine. Elle a une activité avérée contre les formes hépatiques latentes de *P.vivax* (hypnozoïtes) *in vitro* et sur les patients. Ce composé pourrait permettre une durée de traitement plus courte (monodose) que les 14 jours requis par la primaquine. La première étude clinique entreprise par MMV et GSK porte sur la tolérance lors du dosage croissant chez des sujets présentant un déficit hétérozygote à la G6PD, en vue de déterminer la dose maximale à ne pas dépasser. Cette étude a démarré en 2009 et continue de recruter des sujets en 2012. L'efficacité et la tolérance du traitement du paludisme à *P. vivax* avec la tafénoquine sont également en cours d'évaluation dans le cadre d'une étude clinique de Phase IIb utilisant de la tafénoquine et de la chloroquine, présentées séparément. Cette étude, qui vient de démarrer au Pérou et en Thaïlande, se déroule en deux parties. Elle englobe également des sites et des patients d'autres pays d'Asie du sud-est et d'Amérique latine. Dans la première partie de l'étude, près de 320 patients seront recrutés afin d'identifier la dose adéquate de tafénoquine. La seconde partie permettra ensuite de déterminer la tolérance et l'efficacité de la dose sélectionnée sur un plus grand nombre de patients. L'équipe du projet est également en train d'évaluer les tests disponibles pour identifier les individus présentant un déficit enzymatique en G6PD qui pourraient être utilisés au point d'intervention avant le traitement avec une 8-aminoquinoline.



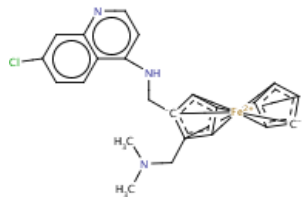
Actuellement, un ensemble d'**antirétroviraux** (antiretroviral therapy ART), lopinavir/ritonavir, emtricitabine/tenofovir disoproxil fumarate, efavirenz, navirapine, trimethoprim/sulfamethoxazole, font l'objet d'une étude clinique de phase II pour voir leur rôle dans la prévention et le traitement du paludisme. Une étude de phase III commencée en 2009 a déjà montré que l'ART lopinavir/ritonavir comparée à une autre ART réduit l'incidence du paludisme de 41% mais est accompagnée de sérieux effets secondaires (Achan et al., 2012).



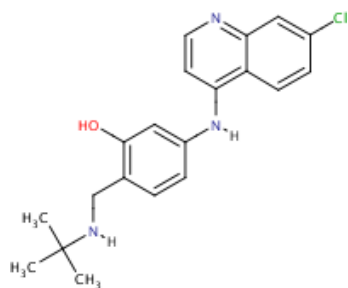
Le **bleu de méthylène** a été identifié en 1891 par Paul Ehrlich comme une molécule efficace contre le paludisme. Puis elle disparut à cause de ses effets secondaires réversibles : urine verte et yeux bleus. L'intérêt pour cette molécule est revenue récemment, surtout à cause de son prix bas. Les premières études de la combinaison bleu de méthylène/chloroquine sont décevantes. Mais, elle est rentrée en étude clinique de phase II en 2007 pour déterminer le profil d'efficacité d'une nouvelle formulation pédiatrique du bleu de méthylène avec l'amodiaquine ou l'artésunate et les résultats sont plus prometteurs.



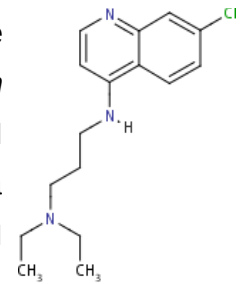
La **pyridinone (GSK932121)** inhibe la chaîne de transports des électrons dans la mitochondrie. Cette molécule a été développée en partenariat avec MMV (Medicines Malaria Venture). Elle est entrée pour la première fois en essais cliniques en 2009. De plus, des efforts sont en cours pour identifier un composé alternatif au cas où les résultats avec la pyridinone ne permettent pas d'aller plus loin dans les essais cliniques. Cela constituerait une nouvelle classe de composés avec le potentiel d'être très efficaces contre les souches résistantes de *P.falciparum* et *P.vivax*.



La **ferroquine** (ou ferrocène / SSR97193) est une 4-aminoquinoléine. Cette molécule est actuellement en développement chez Sanofi-Aventis, elle est en essais cliniques de phase II pour tester sa toxicité et son efficacité. C'est la première molécule organométallique. De plus, elle est active sur les souches résistantes à la chloroquine. Elle agit en ciblant les lipides, en inhibant la formation de l'hémozoïne et en générant des espèces réactives oxygénées.



AQ-13 est une 4-aminoquinoléine. Cette molécule est active sur les souches *P.falciparum* résistantes à la chloroquine. Elle a un profil pharmacocinétique similaire à celui de la chloroquine. Elle est en étude clinique de phase II depuis 2012, pour la comparer au Coartem, association artéméther-luméfántrine.



Le N-tert-butyl isoquine **GSK369796** est une 4-aminoquinoléine, développée en partenariat avec Liverpool, MMV et GSK. Elle agit en se liant à l'hématine, ce qui empêche sa dégradation par le parasite (O'Neill et al., 2009). Elle est actuellement en phase I d'essais cliniques pour évaluer sa pharmacocinétique.

Enfin, à travers le monde, de nombreuses initiatives visent à identifier de nouvelles molécules antipaludiques notamment parmi les plantes (Aguiar et al., 2012), (Campos et al., 2012). Pour exemple, l'**argémone mexicaine** (Photo 5 de B. Navez) est actuellement en essais cliniques de phase Ib, car c'est une plante médicinale traditionnelle connue comme ayant une activité antipaludique. Elle contient donc probablement des molécules intéressantes pour lutter contre le paludisme.



2) Vaccins

Actuellement, aucun vaccin n'est disponible pour combattre le paludisme. En 2011, 35 vaccins à visée antipaludique, correspondant à 154 études, étaient enregistrés par le WHO en étude préclinique ou clinique (Sauerwein et al., 2011). Une des difficultés majeures dans la mise au point d'un vaccin contre *Plasmodium* est, qu'au cours de sa vie, le parasite passe successivement par plusieurs stades avec des phases d'intense multiplication asexuée chez l'homme (dans les cellules du foie - phase hépatique - puis dans les globules rouges du sang - phase érythrocytaire) et une phase de reproduction sexuée suivie de multiplication, chez l'insecte. Chaque cycle se termine par la libération d'un parasite d'une forme différente, comme présenté dans la figure 2, donc porteur d'antigènes différents et induisant des réponses immunitaires différentes, ce qui complique d'autant la recherche d'un vaccin. Un des freins majeurs à la recherche d'un vaccin antipaludique reste l'absence de modèle animal : les rongeurs ne sont pas sensibles aux mêmes souches que l'homme, et les singes ne développent pas de pathologie cérébrale comme l'espèce humaine.

Le premier candidat contre le paludisme, le **SPf66**, a été formulé et testé en Colombie (Patarroyo et al., 1988) et plus tard produit aux Etats-Unis. C'est un peptide synthétique hybride contenant trois antigènes spécifiques des stades sanguins lié à la protéine circumsporozoïte du stade sporozoïte. De ce fait, c'est théoriquement un vaccin multistage. C'est un des premiers vaccins qui est entré en essais cliniques dans les régions endémiques. Dix essais cliniques ont été effectués en Afrique, Asie et Amérique du Sud. Les résultats étaient prometteurs initialement mais des études plus poussées ont montré un effet limité dans certains essais et aucun effet en Afrique. SPf66 n'était pas assez efficace pour être utilisé en routine dans la protection contre le paludisme (Graves & Gelband, 2006).

Actuellement, trois stades peuvent être ciblés pour essayer de mettre au point des vaccins. Les vaccins ciblant le stade exo-érythrocytaire visent à empêcher le sporozoïte de pénétrer ou de se développer dans les cellules hépatiques. Les vaccins ciblant le stade sanguin asexué empêchent les mérozoïtes de pénétrer ou de se développer dans les érythrocytes. Et enfin, les vaccins bloquant la transmission tentent d'empêcher le passage entre les deux hôtes, le moustique et l'homme.

RTS,S/AS01 est un vaccin ciblant le stade exo-érythrocytaire en phase 3 d'étude clinique. Il cible la protéine circumsporozoïte (CSP) lorsque le parasite infecte les cellules hépatiques. Sponsorisée par GSK et MVI (Malaria Vaccine Initiative), l'étude sur 4 ans a montré une protection à 43,4% chez les bébés de 5 à 17 mois la première année. Cette protection devient nulle après 3 ans surtout dans les zones à forte exposition à la malaria (Olotu et al., 2013).

Un autre vaccin, développé par la division d'immunologie et d'allergie du CHU Vaudois de Lausanne en suisse, s'appuie sur un peptide de synthèse **PfCS102** renforcé par un adjuvant, soit le Montanide ISA 720, soit le système adjuvant AS02A (GSK). Ce vaccin est basé sur un peptide chevauchant la région C-terminale 282-383 de la protéine circumsporozoïte (Roggero et al., 1995). Une première étude a montré que le peptide PfCS102 est sûr et hautement immunogénique (Audran et al., 2009). Cependant, une seconde étude clinique de phase I/IIa, PfCS102 ne montre aucune protection contre la paludisme avec la formulation utilisée (Genton et al., 2010).

En 2005, un vaccin a été développé à partir de protéines de surface du mérozoïtes et plus particulièrement à partir de la protéine **MSP3** (Chauhan et al., 2010). En étudiant les protéines parasitaires reconnues par le sérum de sujets « prémunis », vivant en zone d'endémie en Afrique et en Asie, les chercheurs ont pu mettre en évidence l'intérêt de ce nouvel antigène. Il existe en effet une très forte corrélation entre la présence d'anticorps dirigés contre cet antigène et la protection acquise par exposition naturelle à l'infection. Dans un essai clinique de phase I, les chercheurs de l'Institut Pasteur ont montré que MSP3 induisait chez l'homme la production d'anticorps capables d'éliminer le parasite. Les analyses ont montré que ces anticorps étaient aussi efficaces, ou plus efficaces, que ceux produits par des adultes africains « prémunis », pour éliminer les parasites, aussi bien *in vitro* en culture, que *in vivo* chez des souris humanisées. L'essai clinique a également révélé que ce candidat vaccin conférait des anticorps protecteurs de longue durée et qu'il était dénué d'effets secondaires chez l'homme. Fait notable, l'antigène MSP3 n'est pas polymorphe, c'est-à-dire qu'il ne présente pas de différence d'un parasite à l'autre qui permettrait à certains d'échapper aux anticorps induits par le vaccin. Un essai de phase Ib mené au Burkina Faso en 2007 visait initialement à examiner uniquement la tolérance et l'immunogénicité du vaccin chez 45 enfants âgés de 12 à 24 mois ayant reçu des doses de 15 ou 30 microgrammes en comparaison à un vaccin placebo. Néanmoins, les chercheurs n'ont pu s'empêcher de noter la protection conférée par la vaccination ainsi réalisée. Ces résultats prometteurs montrent l'intérêt de la démarche suivie par l'équipe de l'Institut Pasteur et soulèvent de grands espoirs quant au développement dans l'avenir d'un traitement vaccinal efficace contre ce fléau mondial qu'est le paludisme (Sirima et al., 2011).

AMA1 signifie antigène membranaire apical de type I. C'est une protéine essentielle à l'invasion et pourrait conduire à l'élaboration d'un vaccin antipaludique. Cependant, son polymorphisme important au niveau des boucles exposées au solvant pourrait limiter sa capacité à produire des anticorps protégeant les divers génotypes des parasites.

Une recherche dans la PDB sur la protéine AMA1 de *Plasmodium falciparum* retourne 8 structures dont 4 en complexe soit avec l'anticorps IgNAR, anticorps à chaîne lourde (2Z8V et 2Z8W) (Henderson et al., 2007) soit avec le fragment Fab (formé de la chaîne légère en entier et d'une partie de la chaîne lourde) de l'anticorps 1F9 (2Q8A et 2Q8B) (Coley et al., 2007). Une inspection visuelle de ces structures montrent que les deux anticorps cocrystallisés avec PfAMA1 se fixent au niveau du sillon hydrophobe exposé au solvant et plus précisément au niveau du site de fixation de l'épingle à cheveux de PfRON2 présenté dans le chapitre 2. Il a été montré que ces anticorps permettent d'inhiber l'invasion des mérozoïtes.

Une recherche dans la base de données des études cliniques retourne 25 résultats pour les vaccins relatifs à la protéine AMA1 :

- Cinq études concernent des vaccins contenant AdCh63 ou MVA, qui sont des virus inactivés modifiés qui ne peuvent se répliquer chez l'homme, et qui incluent du matériel génétique correspondant à des protéines spécifiques du paludisme (AMA1, MSP1, ME-TRAP, CS) exprimées pendant les stades hépatique et érythrocytaire. Ces vaccins ont pour but de stimuler une réponse immunitaire contre ces protéines et ainsi, fournir une protection contre le paludisme.
- Quatre essais cliniques concernant le vaccin **FMP2.1/AS02A** se déroulent au Mali et aux Etats-Unis. Le vaccin est composé d'une protéine recombinante exprimée chez *E.Coli* et basée sur AMA1 issu de *Plasmodium falciparum* 3D7 et renforcé par l'adjuvant AS02A. Il a été développé conjointement par l'Université du Maryland et l'Université de Bamako. L'essai clinique de phase I a montré que ce vaccin était sûr et immunogène (Thera et al., 2008). Un essai clinique de phase II est actuellement effectué en double aveugle. Cette étude inclut 400 enfants âgés de 1 à 6 ans, vivant à Bandiagara au Mali. Les enfants ont reçu 3 doses, par injection dans le bras. Les patients sont inclus dans l'étude pour 26 mois. Les premiers résultats montrent une efficacité de vaccination d'environ 20%. Normalement, un vaccin ne peut être mis sur le marché qu'avec un taux d'efficacité minimal aux alentours de 70-80%.

- Dix études concernent le vaccin **AMA1-C1**, une mixture de protéines recombinantes AMA1 provenant de deux souches de *Plasmodium falciparum* (FVO et 3D7), produites séparément dans *Pichia pastoris*. En effet, une immunisation avec une seule forme d'AMA1 ne protégerait pas contre les parasites exprimant différentes protéines AMA1. La combinaison de deux ou plusieurs gènes divergents d'AMA1 dans un seule formulation fournirait donc une protection plus large. Des études sont effectuées avec ces protéines complétées par d'autres substances comme Alhydrogel et/ou CPG 7909 pour évaluer l'augmentation de l'efficacité du vaccin (Mullen et al., 2008). Une étude en double aveugle a fait l'objet d'une phase II, elle concerne la formulation AMA1-C1 avec l'Alhydrogel. Aucune protection n'a été observée au premier abord (Sagara et al., 2009).
- Deux études cliniques en phase I concernent le vaccin **PfAMA-1-FVO[25-545]**. Il s'agit d'évaluer la toxicité de celui-ci en présence de divers adjuvants (hydroxyde d'aluminium, Montanide ISA 720 ou ASO2), mesurer la réponse humorale en quantifiant les anticorps, la prolifération des cellules T et la production de cytokines.
- Le vaccin PfCP2.9 a fait l'objet d'une étude clinique de phase I. Il est constitué d'une protéine de fusion contenant AMA1 et MSP1 avec l'adjuvant Montanide ISA720. Une étude préliminaire a montré que cette construction était sûre et immunogénique. Cependant, aucune activité fonctionnelle contre le parasite n'a été observée (Malkin et al., 2008).
- Le vaccin **DNA-Ad** fait l'objet d'une étude clinique de phase II. Il contient un antigène du stade hépatique (la protéine circumsporozoïte) et un antigène présent aux stades hépatique et sanguin (AMA1). Le but est d'éviter l'infection en tuant la majorité des parasites dans le foie.

Enfin en janvier 2010, l'université d'Orlando a annoncé un candidat-vaccin oral ou injectable à la fois efficace contre le paludisme et le choléra. Pour réaliser ce vaccin, la sous-unité B de la toxine responsable du choléra a été fusionnée à des antigènes du parasite responsable du paludisme : AMA1 ou MSP1 et ont été exprimés dans des chloroplastes de laitue et de tabac. Selon l'étude, ce vaccin injecté à des souris a été capable de les protéger durant 300 jours contre la pénétration du parasite dans les globules rouges et contre la toxine cholérique. La réponse immunitaire serait de type Tr1/Th2. Ce procédé breveté pourrait permettre de produire un vaccin peu cher et ne nécessitant pas de stockage réfrigéré en cas d'usage oral (Abdoreza et al., 2010).

IV. Nouvelles cibles, nouveaux modes d'action

Actuellement, divers efforts sont entrepris pour éradiquer le paludisme différemment que par le passé. Avant, la recherche se faisait essentiellement sur les molécules et les vaccins ciblant le stade sanguin de l'infection et l'espèce la plus pathogène et mortelle, *Plasmodium falciparum*.

Aujourd'hui, la recherche s'étend aux stades de transmission aux moustiques et par les moustiques car ils représentent des cibles plus vulnérables pour arrêter la progression du paludisme. Le scénario idéal serait en effet d'interrompre le paludisme aux deux points d'étranglement du cycle de vie du parasite, que vous trouverez figure 8. Le nombre de parasites passant de l'homme au moustique et inversement est très faible, de quelques dizaines à quelques centaines de parasites, comparé au nombre de parasites du stade sanguin. Cibler les stades sexuels du parasite avec des vaccins pourrait empêcher la transmission aux moustiques. Les vaccins développés autour de la protéine circumsporozoïte ciblent le stade hépatique. Il faudrait identifier d'autres antigènes du stade hépatique pour obtenir une protection complète.

Cependant, la recherche sur le stade sanguin est essentielle car c'est lui qui est symptomatique et mortel.

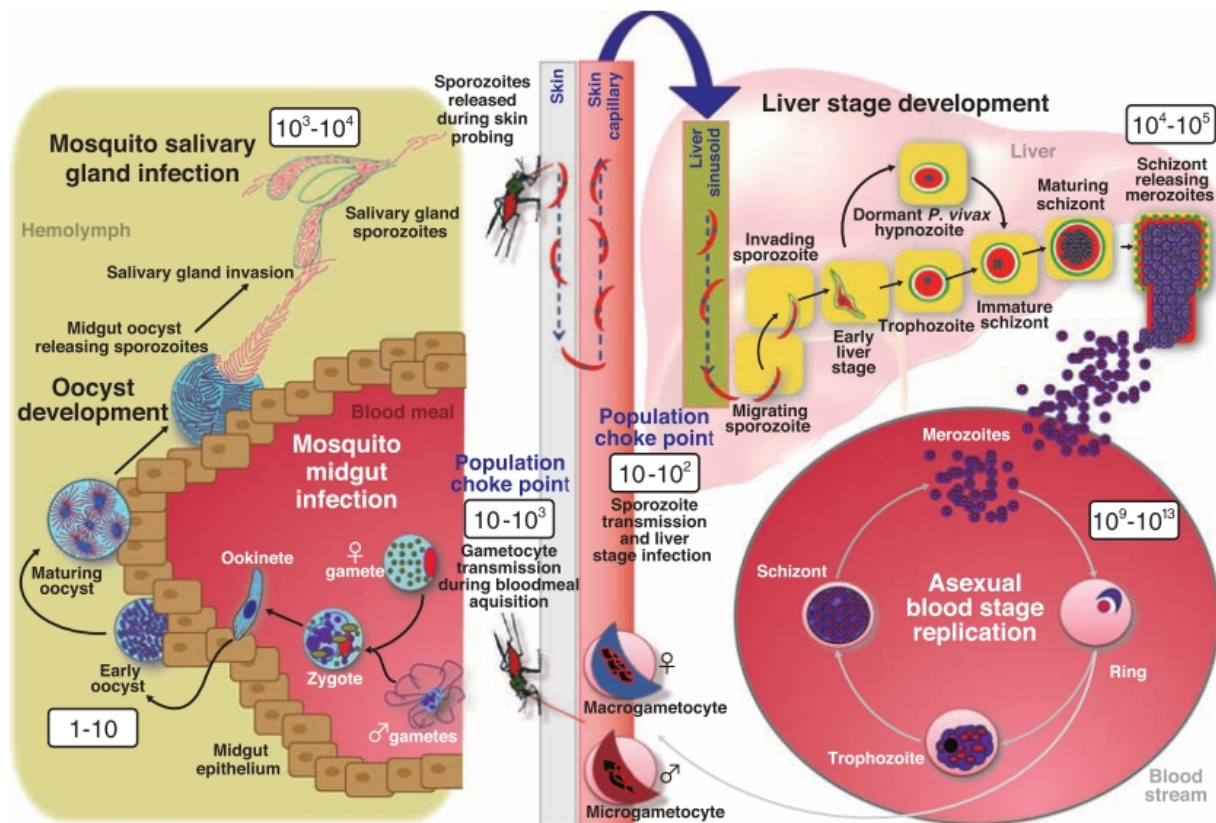


Figure 8 - Le cycle de vie de *Plasmodium* fournit divers sites d'intervention (Kappe et al., 2010).

Les manifestations cliniques du paludisme sont donc dues aux formes intra-érythrocytaires du parasite, les mérozoïtes, qui sont relargués par les schizontes matures au niveau du foie et qui constituent les formes invasives de *Plasmodium falciparum*. Mes travaux de thèse ont consisté à chercher des nouveaux inhibiteurs du stade sanguin, en ciblant soit *PfAMA1* soit *PfCCT*. Le but est, respectivement, d'empêcher l'invasion des érythrocytes par les parasites et de bloquer la biogénèse de la membrane du parasite pendant le cycle érythrocytaire.

En effet, chaque globule rouge envahi produit 16-32 nouveaux mérozoïtes. C'est pourquoi cibler l'invasion ou la réplication du parasite dans l'érythrocyte est intéressant dans le développement de nouvelles thérapies antipaludiques.

1. Inhibition du processus d'invasion du parasite et plus particulièrement, le complexe d'attache *PfAMA1-PfRON2*

1) Le processus d'invasion chez les apicomplexes

Le processus d'invasion de l'érythrocyte par le mérozoïte se déroule en plusieurs étapes. Tout d'abord, une reconnaissance longue distance des récepteurs de surface est impliquée (A) suivie par une réorientation où ces contacts de faible affinité sont maintenus. Une fois le pôle apical attaché à la membrane de l'érythrocyte (B), une jonction serrée est formée impliquant des interactions récepteur-ligand de forte affinité, par exemple l'interaction entre la protéine AMA1 et la protéine RON2 qui va nous intéresser plus particulièrement. Ensuite, la jonction serrée se déplace (et devient donc une jonction mobile) du pôle apical au pôle postérieur (C et D) grâce au moteur actine-myosine du parasite. Une fois le pôle postérieur atteint, les protéines d'adhésion de la jonction mobile sont protéolysées (E). Par ce processus, le parasite rentre dans l'érythrocyte en formant une vacuole parasitophore (Figure 9).

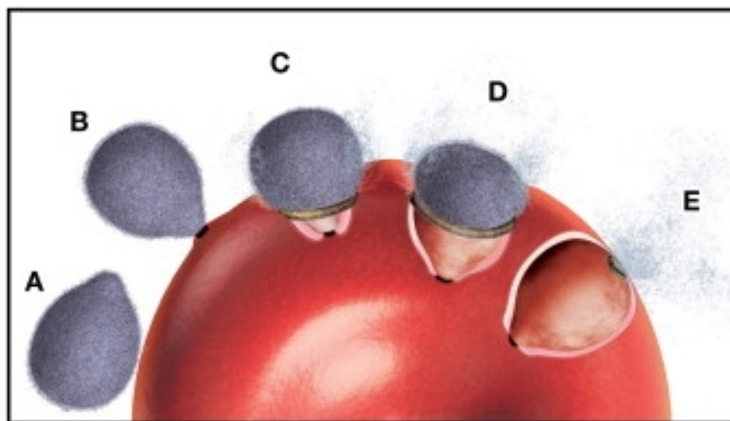


Figure 9 - Mécanisme global (Cowman & Crabb, 2006).

2) Les organelles de l'invasion

Les rhoptries et les micronèmes sont les organelles de l'invasion, spécifiques des apicomplexes, et situées à l'extrémité apicale de la cellule (Figures 1 et 10). Les rhoptries sont subdivisées en deux compartiments, le col effilé et le bulbe qui contiennent respectivement les protéines RONs et ROPs. Les micronèmes contiennent des protéines MICs, dont fait partie la protéine AMA1. Lors de l'invasion, le parasite sécrète de manière séquentielle le contenu des micronèmes puis des rhoptries.

L'équipe de Maryse Lebrun du DIMNP à Montpellier a montré que les protéines AMA1 et RON2, une fois sécrétées, étaient localisées au niveau de la jonction mobile. Ce sont des expériences de coimmunoprécipitation chez *Toxoplasma gondii* (un autre apicomplexe) et *Plasmodium falciparum* qui ont permis de déterminer les acteurs clés de cette jonction mobile (M. Lebrun et al., 2005), (Alexander et al., 2005), (Besteiro et al., 2009), (Mauld Lamarque et al., 2011).

3) Les protéines de l'invasion

Chaque étape de l'invasion des érythrocytes est gouvernée par des PPIs spécifiques. Certaines sont redondantes comme lors de la phase initiale d'attachement avec les protéines de la famille EBA. Ensuite, une de ces interactions spécifiques s'effectue entre deux protéines du parasite, AMA1 qui est transloquée à la surface du mérozoïte et RON2 qui est transférée à la membrane de l'érythrocyte pendant l'invasion. Il n'y a pas de voies alternatives. Le knock-out de ces protéines est impossible et le knock-down implique une diminution drastique de l'invasion. C'est un mécanisme très original qui est mis en jeu car le parasite fournit à la fois le ligand qu'il expose à sa surface et le récepteur qu'il insère dans la membrane de la cellule hôte. Il a été montré que l'interaction entre AMA1 et RON2 est essentielle pour la formation de la jonction mobile avec l'érythrocyte et une étape irréversible qui engage le parasite dans l'invasion.

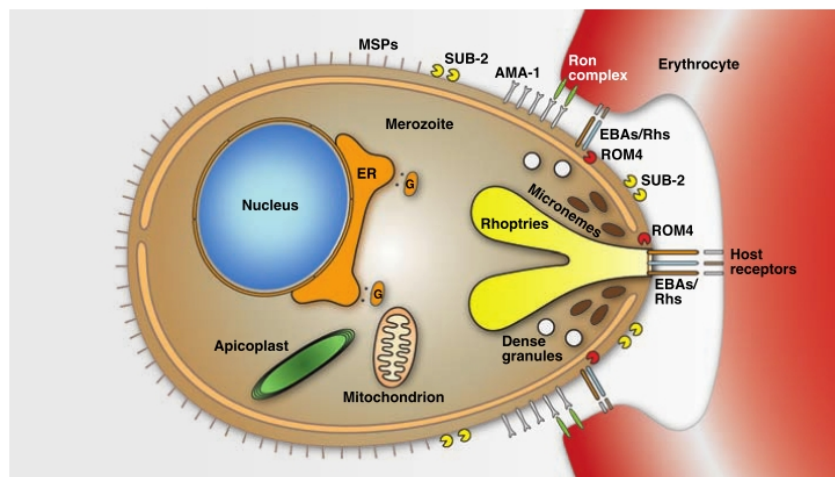


Figure 10 - Protéines impliquées dans l'invasion (Kappe et al., 2010).

4) L'interaction protéine-protéine PfAMA1-PfRON2

L'interaction entre les protéines RON2 et AMA1, qui a été montrée par coimmunoprécipitation chez *Tg*, est bien conservée chez *Pf*. L'anticorps de souris MAb F8.12.19 reconnaît le domaine III extracellulaire de *Pf*AMA1 qui est accessible avec ou sans perméabilisation des cellules. L'incubation de cellules de mammifères transfectées avec *Pf*AMA1 et avec une protéine recombinante *Pf*RON2-5 (peptide correspondant à la boucle extracellulaire, entre les deux domaines transmembranaires, des acides aminés 2020 à 2067 en fusion à la GST) permet de détecter leur liaison. La protéine recombinante GST seule (contrôle) ne présente pas d'affinité pour *Pf*AMA1 (Figure 11).

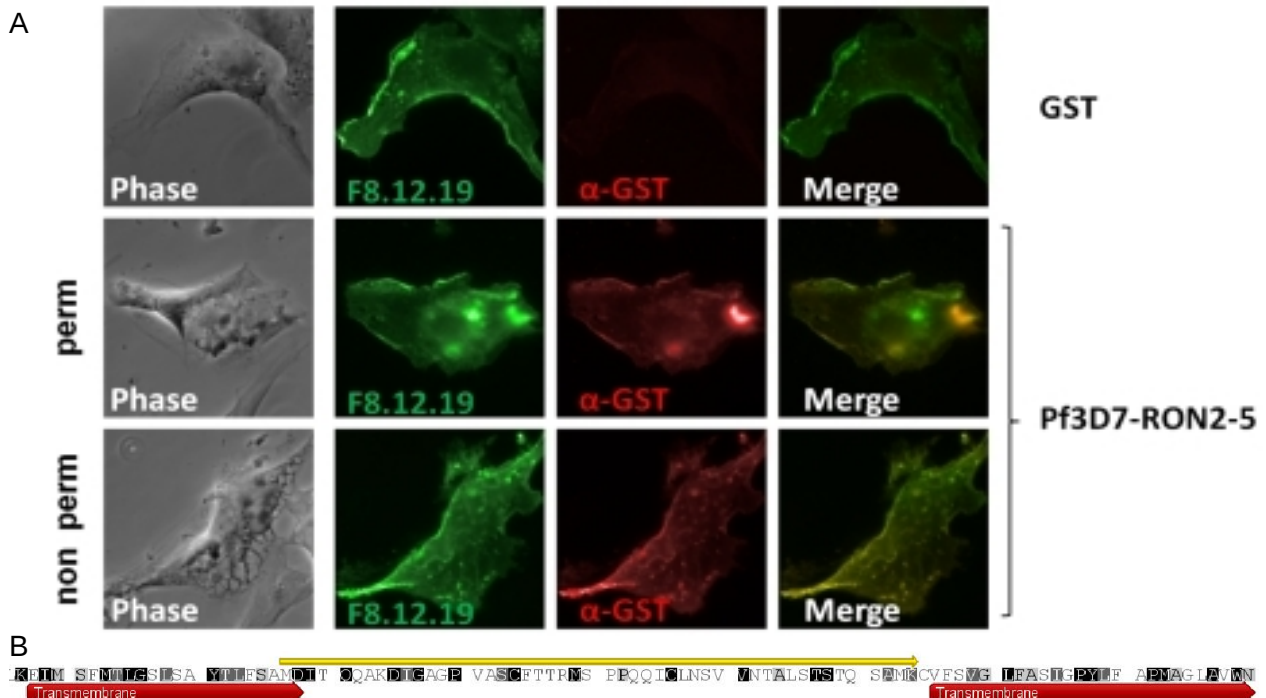


Figure 11 - Issue de (Lamarque et al., 2011). A) Liaison du peptide *Pf*RON2-5 à *Pf*AMA1.

B) Séquence du peptide *Pf*RON2-5 et position par rapport aux domaines transmembranaires.

Il est donc possible de schématiser ce complexe à la jonction mobile de la façon suivante :

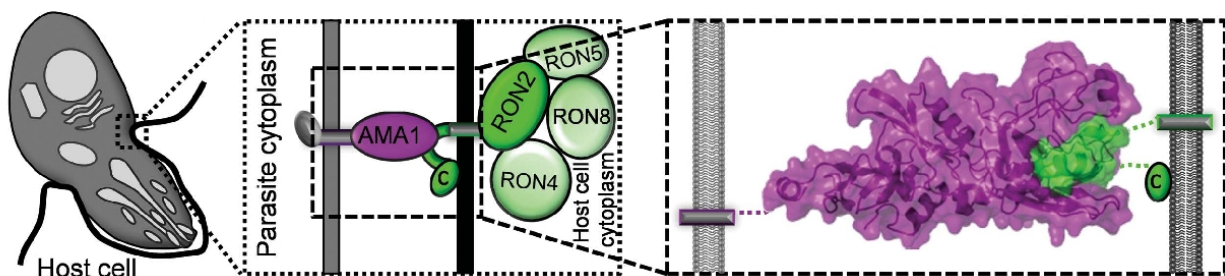


Figure 12 - Issue de (Tonkin et al., 2011). Schéma du complexe AMA1-RON2 à la jonction d'un apicomplexe.

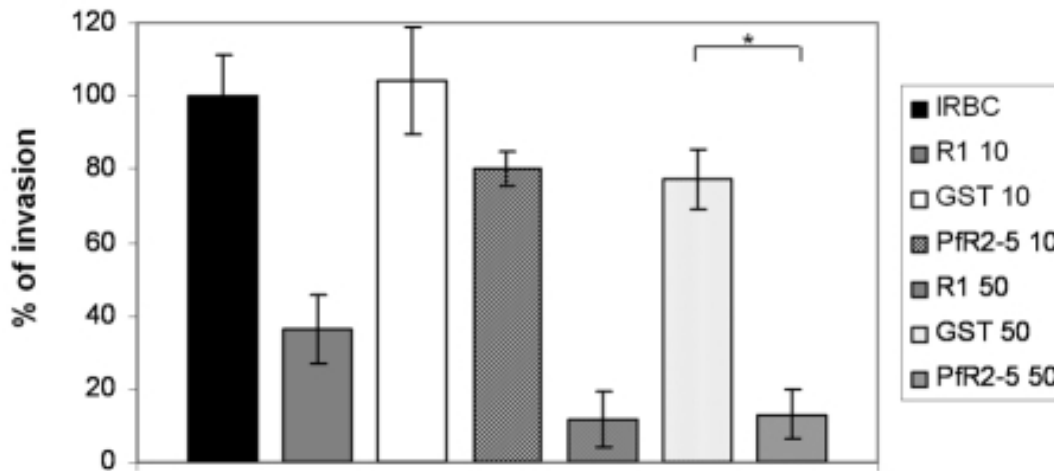


Figure 13 - Issue de (Lamarque et al., 2011). Les peptides R1*Pf*RON2-5 inhibent l'invasion des érythrocytes par *Plasmodium falciparum*.

*Pf*RON2-5, qui est une version tronquée de *Pf*RON2, correspond donc à la région de *Pf*RON2 qui se lie à AMA1. Il a été montré qu'elle inhibe l'invasion des globules rouges par les mérozoïtes de *P. falciparum*. Une expérience de phage display a aussi permis d'identifier un peptide, nommé R1 (Harris et al., 2005), et dont l'effet a aussi été testé ici. Le test d'inhibition de l'invasion a été effectué avec 2% de schizontes matures, qui ont été incubés avec 10 µg/ml (4.3 µM) or 50 µg/ml (21 µM) du peptide R1 ou équivalent en concentration de *Pf*RON2-5 ou de la protéine GST. La parasitémie des érythrocytes contrôles (IRBC : Infected Red Blood Cells) seize heures après invasion a été utilisée comme référence (100%). Le peptide R1 inhibe l'invasion du parasite, de la même manière que le fait *Pf*RON2-5 (t-test, * $p < 0.05$) (Figure 13).

Cette expérience montre que si l'on arrive à bloquer l'amarrage de RON2, on sera alors capable d'empêcher la jonction mobile de se former et ainsi de bloquer l'invasion des érythrocytes par les parasites.

Comme vous avez pu le constater, AMA1 est une des cibles majeures pour le développement de vaccins antipaludiques. C'est une protéine conservée chez *Plasmodium* mais également dans le phylum des Apicomplexes. Nos collaborateurs canadiens (Equipe de Martin Boulanger, Université de Victoria) ont cocrystallisé la structure d'AMA1 en complexe avec une partie de RON2 de *Plasmodium falciparum* contenant 39 acides aminés (peptide *Pf*RON2sp1, PDB 3ZWZ). *Pf*AMA1 a aussi été cocrystallisée avec le peptide R1 de séquence VFAEFLPLFSKFGSRMHILK (PDB 3SRJ) (Vulliez-Le Normand et al., 2012). Ces deux structures sont présentées dans la figure 14.

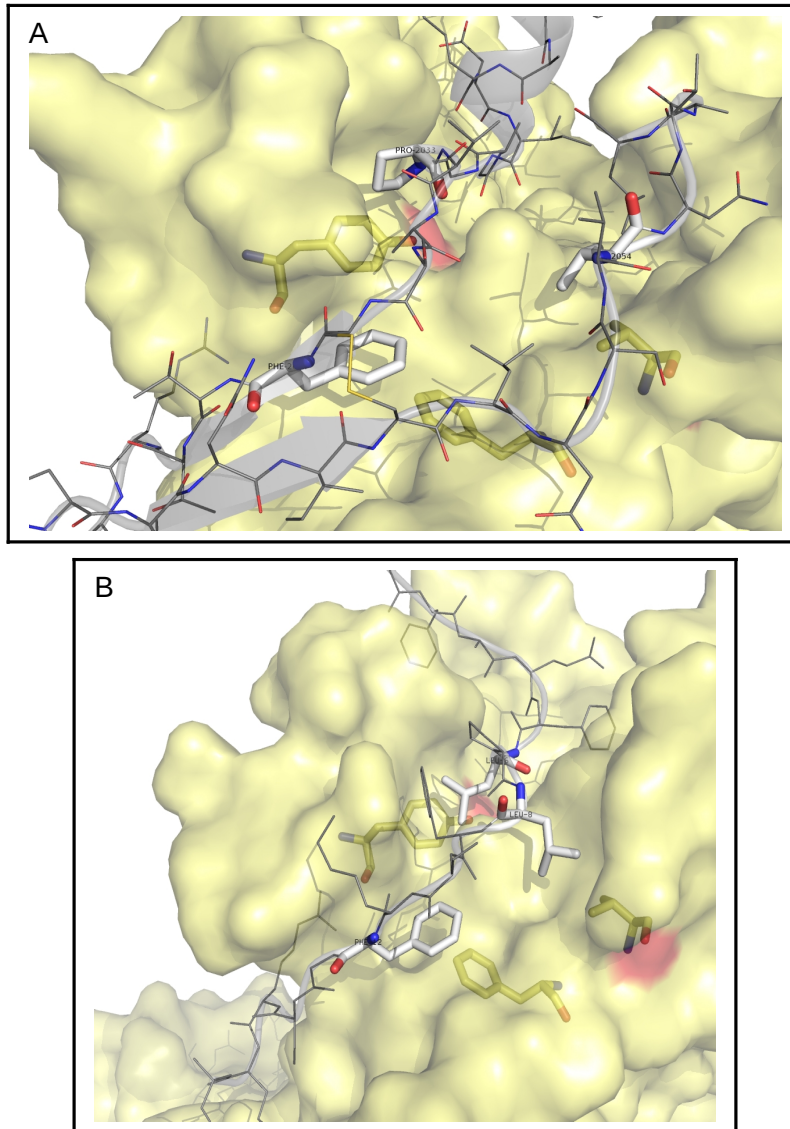


Figure 14 - Costructures de *Pf*AMA1 avec A) le peptide *Pf*RON2sp1 (PDB 3ZWX) et B) le peptide R1 (PDB 3SRJ) (Vulliez-Le Normand et al., 2012). La protéine AMA1 est représentée par une surface jaune et les peptides sont en blanc. Les acides aminés en bâton sont les résidus clés de l'interaction. Elles apportent des informations très utiles dans une démarche de recherche de nouveaux ligands par criblage virtuel.

De par les diverses études récentes et les informations qu'elles apportent, ce complexe *Pf*AMA1-*Pf*RON2 constitue une cible très intéressante pour développer de nouveaux médicaments contre le paludisme.

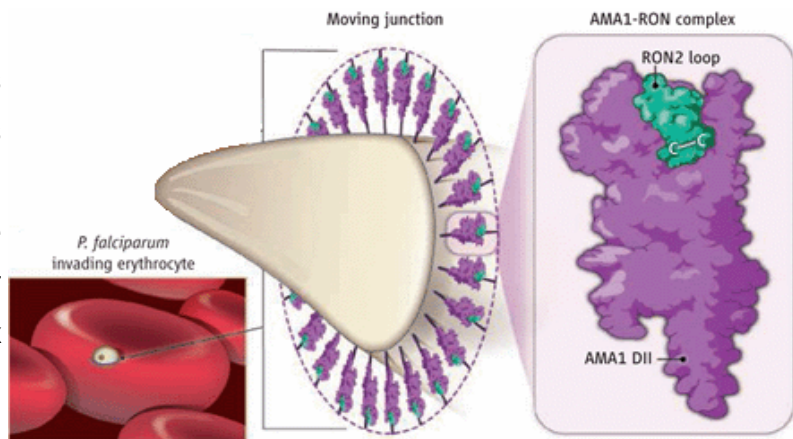
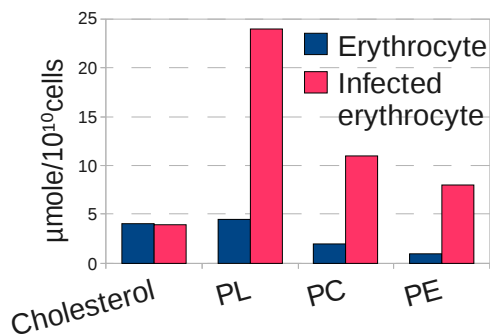


Image issue de (Baum & Cowman, 2011)

2. Inhibition de la multiplication du parasite et plus particulièrement, une étape limitante de la biosynthèse de la phosphatidylcholine

1) La biogénèse membranaire chez *Plasmodium*

La biogénèse membranaire de *Plasmodium* se déroule au cours de sa phase intra-érythrocytaire, suivant l'étape d'invasion. Le parasite prolifère à l'intérieur d'une vacuole parasitophore, dans l'érythrocyte mature, une cellule ayant perdu son noyau et la capacité de synthétiser protéines et lipides. Pour assurer sa prolifération (*P. falciparum* peut donner jusqu'à 32 nouveaux parasites au cours de son cycle de 40-48 heures), le parasite intracellulaire doit synthétiser de nombreuses membranes. Contrairement à la plupart des organismes eucaryotes, les membranes de *Plasmodium* sont dénuées de cholestérol et ont pour composante lipidique essentielle des phospholipides. Les deux phospholipides majoritaires sont la phosphatidylcholine (40-50%) et la phosphatidyléthanolamine (35-45%); le phosphatidylinositol



représente jusqu'à 11% tandis que la sphingomyéline y est quasi absente (Vial & Mamoun, 2005). Il a été montré qu'il y a six fois plus de phospholipides (PL) dans les érythrocytes infectés que dans les érythrocytes sains. Le phospholipide le plus abondant est la phosphatidylcholine (PC) (Déchamps et al., 2010).

Plasmodium doit posséder et met en place ses propres voies de biosynthèse pour effectuer une néo-synthèse d'une quantité considérable de phospholipides (jusqu'à 6 fois le contenu de la cellule hôte). L'approche pharmacologique consiste à cibler cette biosynthèse indispensable à la construction des structures membranaires du parasite et donc à sa survie. Les phospholipides sont synthétisés à partir de têtes polaires et d'acides gras puisés dans le plasma. Ce métabolisme paraît unique par son intensité et la multiplicité des voies métaboliques utilisées. Il combine des voies généralement présentes chez les procaryotes, les eucaryotes et, encore plus remarquable, les plantes. La principale voie de biosynthèse de phosphatidylcholine semble être la voie de synthèse *de novo*, dite voie de Kennedy, qui met en jeu trois enzymes. Il a aussi été montré que la sérine pouvait être décarboxylée en éthanolamine, suivi d'une formation de phosphoryléthanolamine. Les données génomiques ont montré qu'une activité phosphoryléthanolamine méthylase, jusque-là uniquement décrite chez les plantes, pouvait prendre le relais conduisant à de la phosphorylcholine, et ensuite à la phosphatidylcholine. Ces voies de synthèse n'apparaissent pas pouvoir se compenser et paraissent donc non redondantes.

2) La voie de synthèse *de novo* Kennedy

a) Présentation

La voie de synthèse *de novo* dite « voie de Kennedy » permet la synthèse de PC à partir de choline plasmatique et met en jeu trois enzymes : la choline kinase (CK), la CTP:phosphocholine cytidyltransférase (CCT) et la choline/éthanolamine phosphotransférase (CEPT) (Figure 15). Des knock-outs des enzymes de cette voie ont été tentés chez *Plasmodium berghei* mais les parasites y sont réfractaires pour les gènes *ck*, *cct*, *ect* et *cept* codant, respectivement pour CK, CCT, ECT et CEPT, les enzymes de la voie de biosynthèse des phospholipides (Déchamps et al., 2010). Les enzymes impliquées dans la voie *de novo* Kennedy sont donc très importantes pour la survie du parasite et sont donc des cibles pharmacologiques prometteuses.

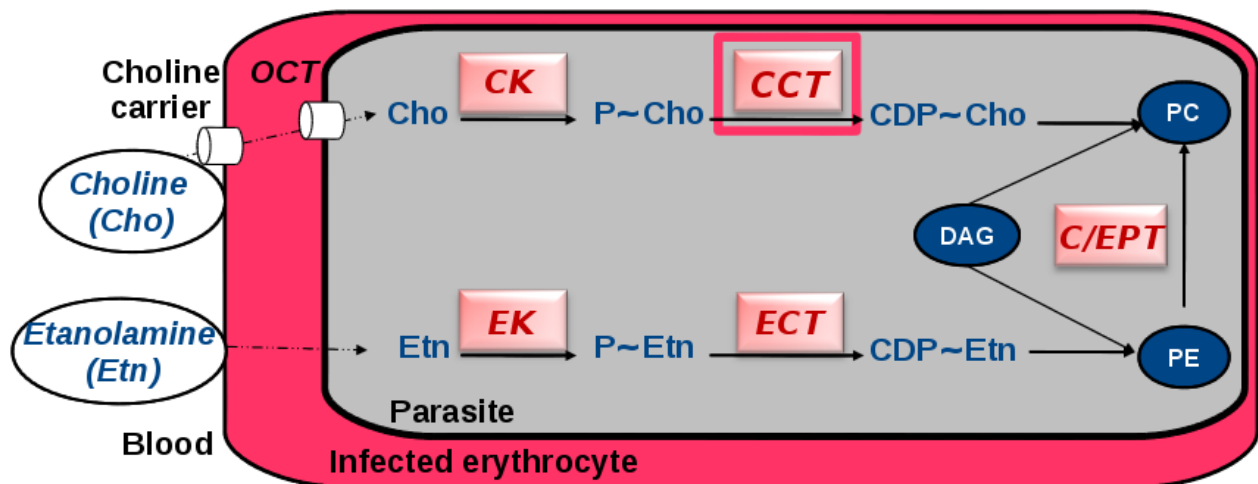
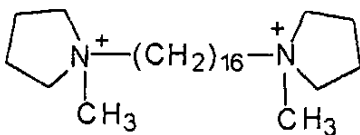


Figure 15 - Voie de synthèse *de novo* des phospholipides (Cho : choline, P~Cho : phosphocholine, CDP~Cho : CDP-choline, PC : phosphatidylcholine, Etn : éthanolamine, P~Etn : phosphoéthanolamine, CDP~Etn : CDP-éthanolamine, PE : phosphatidyléthanolamine).

b) Molécules ciblant la voie *de novo* Kennedy

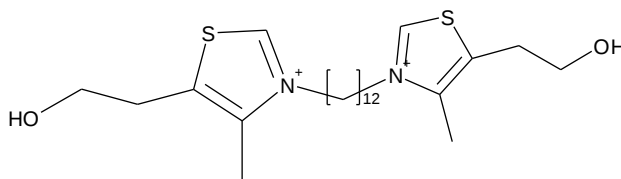
Des analogues de choline capables de bloquer l'import de choline et/ou d'altérer une ou plusieurs des enzymes de synthèse de phosphatidylcholine ont déjà été identifiés (Vial & Mamoun, 2005), (Roggero et al., 2004). Ces analogues ont été optimisés pour leur activité antipaludique. Une première génération de *bis*-ammonium quaternaire, dont le G25 est le



composé *leader*, a montré une très puissante activité antipaludique à la fois *in vitro* et *in vivo* (Calas et al., 2000), (Wengelnik et al., 2002), (Ancelin et al., 2003) et (Ancelin & Calas, 2003).

Cependant, la structure chargée de ces molécules ne leur permet pas un bon passage de la barrière gastro-intestinale, empêchant ainsi une administration par voie orale.

Il a donc été développé des pro-drogues neutres susceptibles de se convertir en composés *bis*-thiazolium ayant une activité antipaludique *in vitro* à moins de 10 nM (Nicolas et al., 2005). Ces composés sont actifs *in vivo* chez la souris à de très faibles doses par voie parentérale ou orale, et cela à faible ou haute parasitémie, et sont très faiblement toxiques (1 000 fois moins toxiques sur les cellules de mammifères que sur *Plasmodium*). Il a aussi été montré que les *bis*-thiazolium inhibent spécifiquement la biosynthèse de phosphatidylcholine, sans toucher la synthèse d'ADN ou des autres phospholipides (Vial et al., 2004). De plus, ils s'accumulent plus de 200 fois spécifiquement dans l'érythrocyte infecté. Ce ciblage spécifique de ces produits vers les cellules cibles est le rêve de tout pharmacologue et permet de travailler à faible dose et ainsi d'éviter les effets cholinergiques de ces produits.



Une molécule de cette famille des bithiazolium, nommée SAR97276A ou T3, est entrée en essais cliniques de phase II. Elle est développée chez Sanofi Aventis. Elle a été découverte par l'équipe d'Henri Vial à

Montpellier. Elle cible principalement l'entrée de choline dans la parasite via un ou deux transporteurs. Ce candidat est actif sur les souches résistantes de *P. falciparum* et est utilisé pour soigner les cas de malaria sévère. Cependant, aucune formulation orale n'a pu être réalisée, limitant de ce fait son utilisation. L'inhibition de la biosynthèse des PL est donc une stratégie thérapeutique validée.

Enfin, il a été récemment montré qu'une partie des composés était accumulée dans la vacuole digestive du parasite intracellulaire où ils peuvent interagir avec l'hème (métabolite de l'hémoglobine, toxique pour le parasite), interaction qui paraît contribuer à l'activité antiplasmodiale (Biagini et al., 2003). La puissance des composés s'exprime par une activité cytotoxique non réversible après un faible temps de contact.

Les *bis*-thiazolium sont donc des antipaludiques uniques par leur accumulation spécifique dans les érythrocytes infectés, leur double mécanisme d'action et leur faible toxicité. Le métabolisme de la phosphatidylcholine n'est la cible d'aucun antipaludique actuel, et apparaît comme une alternative réaliste aux antipaludiques pour lesquels une chimiorésistance existe.

3) La PfCCT, enzyme limitante de la voie Kennedy

La CCT est l'enzyme catalysant l'étape limitante de la synthèse *de novo* de la PC, PL majeur de la membrane. En effet, la CDP-choline intracellulaire reste à une concentration très faible comparée à la phosphocholine (Ancelin & Vial, 1989). Ses substrats naturels sont la phosphocholine et le CTP et la CDP-choline est le produit de la réaction. La CTP:Choline-phosphate Cytidylyltransférase catalyse donc la réaction présentée dans la figure 16.

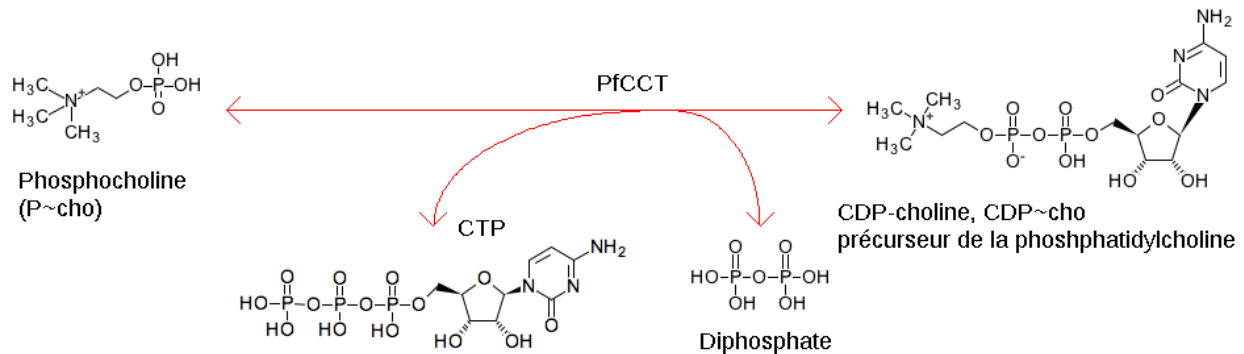


Figure 16 - Réaction catalysée par la PfCCT.

Les valeurs des paramètres cinétiques confortent l'idée que la CCT est l'enzyme limitante par sa faible affinité pour son substrat et sa faible vitesse de réaction.

Enzyme endogène	Substrat	Km (μM)	Vmax (pmol/min/10 ⁷ parasites)
PfCK	Choline	31,2 \pm 4,1	63,3 \pm 11,6
PfCCT	Phosphocholine	511 \pm 24.8	5,2 \pm 1.4
PfC/EPT	CDP-choline	13,2 \pm 0,7	17,0 \pm 1,5

Tableau 1 - Valeurs des paramètres cinétiques des enzymes de la voie de biosynthèse de la phosphatidylcholine (Wein et al., 2012).

La CCT est une protéine amphitropique, opérant une conversion entre une forme soluble inactive et une forme liée à la membrane active (Yeo et al., 1995). Elle contient 896 acides aminés, un domaine catalytique et un domaine de liaison à la membrane dupliqués et séparés par un long linker, dont la fonction est inconnue (Larvor et al., 2003).

La structure expérimentale de la *PfCCT* n'est pas encore connue mais il existe une structure de la CCT chez le rat en complexe avec son produit, la CDP-choline (Lee et al., 2009). La domaine catalytique de la *PfCCT* partage 51% d'identité avec la *rCCT* et 76% d'identité (92% similarité) si on ne considère que les résidus du site actif. Un modèle par homologie va pouvoir être construit avec ce support (Figure 17), ce qui donnera par ailleurs la possibilité de faire un criblage de petites molécules par amarrage (SBVS). De plus, la conformation de la CDP-choline telle qu'elle existe dans le site actif de la protéine *rCCT* va nous servir de référence pour tous les outils de criblage virtuel basés sur le ligand (LBVS : Ligand-Based Virtual Screening).

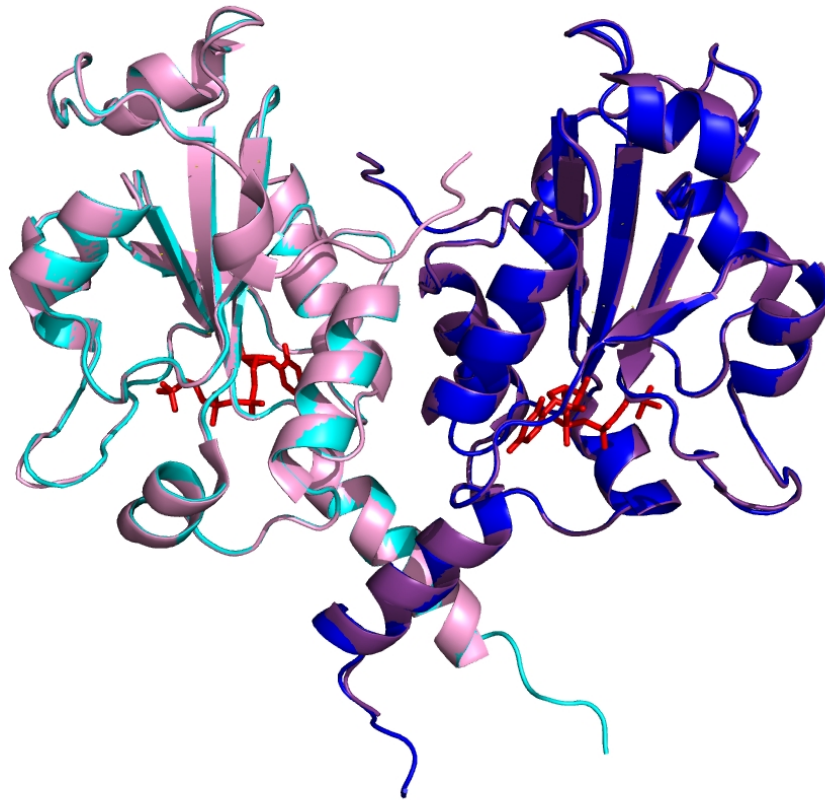
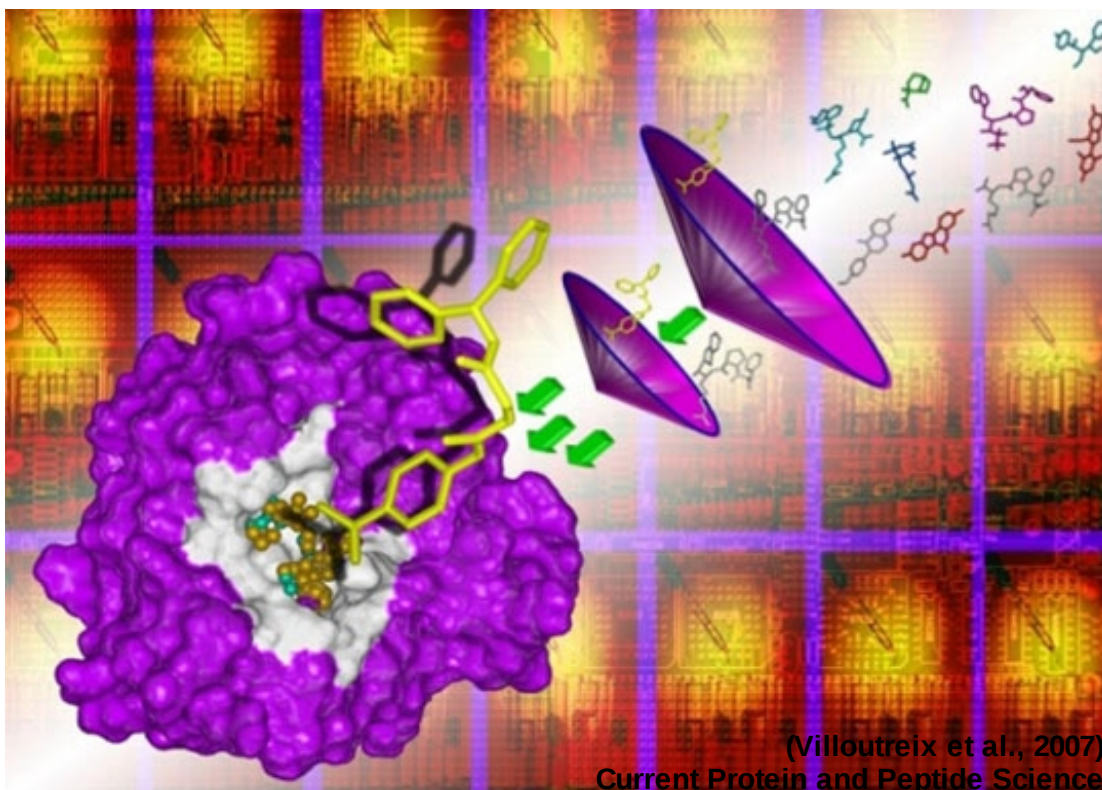


Figure 17 - Structure cristallographique et modèle de la CCT de *Rattus norvegicus* en rose et *Plasmodium falciparum* en bleu, respectivement.

Le but est de développer une nouvelle classe de médicaments contre le paludisme ciblant la biogénèse de la membrane du parasite durant le cycle érythrocytaire de *Plasmodium falciparum*.

Chapitre 1 : Les chimiothèques et les outils de criblage



I. Introduction

1. La chimoinformatique

La **chimoinformatique** a été clairement définie lors de la manifestation scientifique "Workshop Chemoinformatics in Europe : Research and Teaching" (29 mai - 1er juin 2006 à Obernai, France) : "La chimoinformatique est une discipline scientifique qui a évolué durant les 40 dernières années. Elle est à l'interface entre la chimie et l'informatique. On a réalisé que dans plusieurs domaines de la chimie, la quantité importante de données produites par la recherche en chimie ne peut être analysée que par des méthodes informatiques. De plus, beaucoup de problèmes en chimie sont si complexes que les nouvelles approches utilisant des solutions basées sur des méthodes informatiques sont nécessaires. Ainsi, des méthodes ont été développées pour construire des bases de données de molécules, des bases de données de réactions, pour prédire les caractéristiques physiques, chimiques et biologiques des molécules et des matériaux, pour la conception de médicaments, pour l'élucidation des structures, pour la prédiction des réactions chimiques et pour la conception des synthèses organiques. La recherche et le développement en chimoinformatique sont essentiels. Pour améliorer notre connaissance des phénomènes chimiques, pour que l'industrie reste compétitive dans une économie globale. Les méthodes chimoinformatiques peuvent être appliquées dans n'importe quel domaine de la chimie, de la chimie analytique à la chimie organique. C'est d'une importance particulière en conception et en développement de médicaments."

2. Le criblage virtuel

1) Définition

Le criblage virtuel est une des activités de la chimoinformatique, qui utilise des modèles informatiques pour sélectionner des composés pour les tests expérimentaux. On assigne un score à chaque molécule dans une base de données moléculaires, afin de les classer et une très faible fraction des molécules ayant les meilleurs scores est sélectionnée pour les tests expérimentaux. Les méthodes disponibles pour donner un score aux molécules peuvent être relativement simples ou complexes.

Le criblage expérimental à haut-débit est la principale méthode en chimie médicinale pour sélectionner de nouvelles molécules têtes de série ou candidats médicaments. Mais, en raison du coût élevé de la synthèse et du criblage d'une très grande quantité de composés, le criblage virtuel apparaît comme une étape préliminaire importante. Le criblage virtuel se définit comme un procédé *in silico* permettant de filtrer des chimiothèques en fonction de l'affinité prédite de chaque molécule pour une cible biologique prédéfinie. Il permet de focaliser l'approche expérimentale sur les molécules candidates les plus prometteuses, lorsque qu'une

approche de criblage expérimental nécessite une plate forme technologique performante. Le criblage virtuel nécessite des informations sur les propriétés physicochimiques et/ou géométriques des composés actifs et sur la structure 3D de la cible. Selon la disponibilité ou non de ces informations, différentes méthodes de criblage virtuel peuvent être utilisées. Les méthodes basées sur les ligands (ligand-based) sont utilisées lorsque les informations émanant des composés actifs ou des ligands naturels de la cible biologique sont disponibles. Les méthodes basées sur la structure (structure-based) sont utilisées lorsque les données sur la structure 3D de la cible biologique sont disponibles. La forte progression des structures tridimensionnelles expérimentales des macromolécules disponibles a provoqué un fort développement des méthodes basées sur la structure. Lorsque les informations aussi bien sur les composés actifs que celles sur les structures 3D de la cible biologique sont disponibles, il est possible d'utiliser les deux méthodes séparément afin de comparer leur efficacité ou bien de les utiliser conjointement afin d'introduire tout le rationnel connu sur le sujet.

2) Préparation des chimiothèques

Le nombre de molécules disponibles pour un criblage est important. On estime à 10 million le nombre de molécules disponibles à l'achat. C'est pourquoi il est important de supprimer les molécules qui ont des propriétés indésirables (Baell & Holloway, 2010). Ces propriétés peuvent varier d'un projet à l'autre, d'une cible protéique à l'autre, mais quelques règles générales sont appliquées spécifiquement en pharmacologie. Une courte liste de propriétés considérées comme indésirables pour des molécules devant passer en tests expérimentaux serait : la forte propension à se lier covalamment à la protéine cible, la présence de groupements toxiques, la solubilité, la présence de propriétés pouvant interférer dans les tests expérimentaux (ex : aggrégation).

Il est préférable de supprimer ce genre de molécules avant l'étape de criblage virtuel. On s'assure ainsi que les composés retenus par le criblage virtuel conviendront lors des tests expérimentaux. Il y a plein d'exemples de filtres bien connus correspondant aux quatre classes de filtres ci-dessus. Ils sont, par exemple, rassemblés dans FAF-Drugs2 (<http://bioserv.rpbs.univ-paris-diderot.fr/FAF-Drugs/>, (Lagorce et al., 2011)). Ceux permettant de prédire la biodisponibilité orale ont été largement explorés, le filtre le plus connu est la règle des 5 de Lipinski (Lipinski et al., 2001).

3) Les filtres 'ligand-based'

Les méthodes '**ligand-based**' utilisent l'observation que des molécules similaires à une molécule active auront plus de chance d'être actives qu'une molécule prise au hasard. Les approches de ce type considèrent la structure chimique en 2D ou en 3D, la forme, les propriétés physicochimiques ou la distribution des groupements chimiques (c'est-à-dire l'arrangement dans l'espace des groupes constituant un pharmacophore) pour évaluer la similarité avec la molécule active.

La quantité et la qualité de l'information requise pour pouvoir utiliser ces techniques varient. Les approches basées sur la similarité avec une molécule, ligand-based, peuvent ne requérir qu'une molécule active, qui peut provenir de la littérature, des brevets, ou des données expérimentales. Les approches QSAR (Quantitative Structure-Activity Relationship : relations quantitatives structure-activité), elles, requièrent un nombre de molécules actives plus grand couvrant une large gamme d'activité contre la protéine cible. La qualité du modèle QSAR dépend en très grande partie de la qualité des données d'activité, de telle sorte qu'un modèle fiable de QSAR est généralement construit à partir de données de liaison ou d'inhibition acquises avec précision et non pas via un criblage haut-débit imprécis. En effet, lors d'un criblage haut-débit, une seule concentration est testée et la diversité structurale des molécules est trop grande. Pour établir un bon modèle QSAR, il faut disposer de résultats concernant des molécules inactives et une série de molécules actives structurellement proches.

4) Les filtres 'structure-based'

Les méthodes '**structure-based**' utilisent la structure expérimentale tridimensionnelle de la protéine pour prédire quelles molécules vont s'y lier. Alors que ces méthodes semblent plus intuitives, la complexité des interactions protéine-ligand est difficilement numérisable et beaucoup d'approximations doivent être utilisées. En pratique, ces approximations peuvent limiter l'efficacité de ces méthodes.

Concernant les méthodes structure-based, l'amarrage de molécules (ou 'docking') requiert une structure expérimentale ou un modèle par homologie de la structure protéique cible. Les modèles de pharmacophore qui incluent l'information du récepteur nécessitent aussi une structure expérimentale du complexe entre une molécule active et sa protéine cible. De plus, le docking dans des structures protéiques ne contenant pas de ligand ou dans des modèles construits par homologie est moins précis.

5) L'inspection visuelle ou la validation manuelle

Alors que l'application d'un protocole de criblage virtuel à de larges bases de données peut permettre de supprimer automatiquement un grand nombre de composés indésirables, on ne peut pas substituer l'intervention humaine. Un chimoinformaticien expérimenté peut fournir un contrôle qualité de grande valeur aux résultats de criblage virtuel, s'assurant que les bons composés ont été sélectionnés pour les bonnes raisons. Un outil qui permet une visualisation, une manipulation et une évaluation simple de centaines ou de milliers de composés aide le chimoinformaticien à inspecter la liste des candidats provenant des criblages virtuels et à juger la qualité et la pertinence des résultats.

3. Le protocole de criblage projet-dépendant

En résumé, ces grandes étapes de préparation, de filtres automatiques et manuels se retrouvent dans les protocoles de criblage virtuel que j'ai développé pour les deux projets (Figure 18). Tous les logiciels utilisés vont être présentés plus en détail dans la suite du Chapitre 1, tandis que les paramètres spécifiques à chaque projet seront développés dans les Chapitres 2 et 3.

De manière succincte, une étape de filtre supprime les composés indésirables.

Puis un filtre "ligand-based" est appliqué, divers logiciels assignent un score à chaque molécule selon leur similarité de forme, de pharmacophore ou de familles structurales avec le ligand de référence. Les molécules sont alors rangées selon leurs scores de similarités.

Celles ayant les scores suffisants sont sélectionnées pour l'étape suivante, qui est un filtre "structure-based". On va amarrer les molécules dans le site actif de la protéine cible ou dans le domaine de liaison à une autre protéine (*PfCCT* ou *PfAMA1* respectivement), et on obtiendra un nouveau score représentant la "force" de la liaison. On pourra aussi utiliser le type d'interactions établies par ces molécules dans leurs modes de fixation pour évaluer leur pertinence.

La dernière étape est celle de la sélection manuelle, où les molécules répondant aux critères de scores et/ou d'interactions spécifiques au projet seront examinées et sélectionnées. D'autres critères de choix peuvent s'ajouter et inclure un bon visuel des poses de docking dans les sites actifs, la présence de fonctionnalités intéressantes, une chimie qui ferait d'une molécule un bon point de départ pour l'élaboration d'analogues plus actifs, l'existence d'un seul mode de liaison à la protéine cible ou la présence d'un type structural non couvert par un brevet.

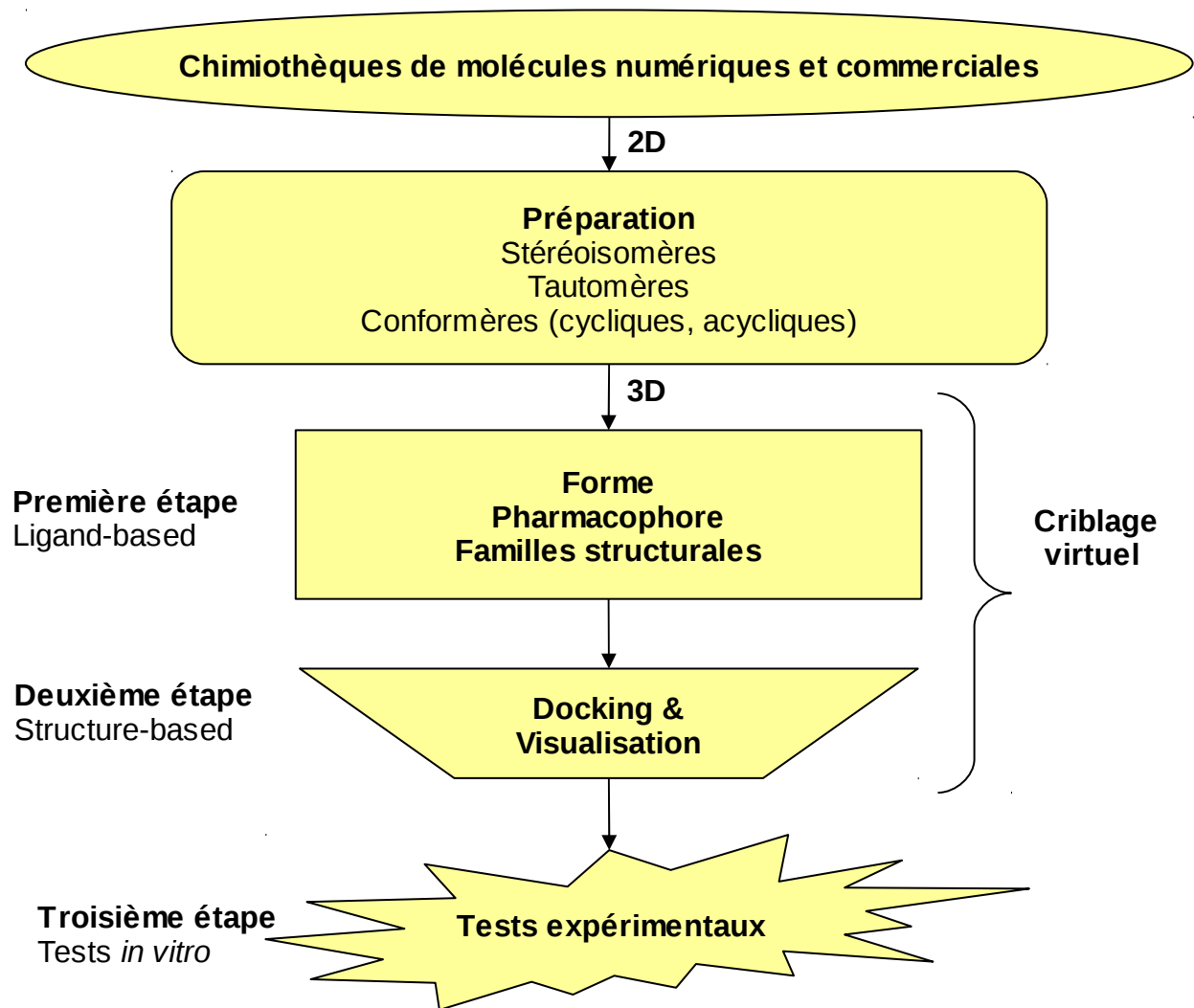


Figure 18 - Protocole de criblage virtuel global, mettant en jeu successivement une préparation des chimiothèques, des méthodes ligand-based et des méthodes structure-based.

II. Les chimiothèques

1. Présentation

Nom	Abréviation	Taille (en nombre de molécules)
e-Drug3D	e-Drug3D	1545
Elle a été développée au laboratoire et contient les médicaments passés et présents approuvés par la FDA. J'ai préparé les fichiers proposés au téléchargement contenant différents tautomères et/ou conformères de chaque molécule. (Pihan et al., 2012)		
DrugBank	DB	6631
Elle contient les petites molécules approuvées par la FDA (Food and Drug Administration), les molécules de type biotechnologique (protéine et peptide) approuvées par la FDA, les produits nutraceutiques et les molécules expérimentales (non approuvées).		
Chimiothèque Nationale	CN	43878
Elle regroupe les collections de produits de synthèse, de composés naturels et d'extraits naturels existants dans des laboratoires publics français.		
CHEMBL	CHEMBL	28379
Elle contient les molécules dont on sait qu'elles ont une action sur <i>Plasmodium falciparum</i> . (La MalariaBox est donc incluse dans cette base.)		
Express-Pick Chembridge	CB	485778
Elle contient des composés chimiques synthétisés par la société Chembridge, choisis pour leur nouveauté, leur diversité et leurs propriétés druglike.		
Ambinter	AMB	5053511
Elle contient des molécules pour le criblage haut-débit, des composés pour la chimie combinatoire, ainsi qu'une large sélection de fragments et de produits naturels. Ambinter est un revendeur de molécules organiques.		
Drug-Like	DL	2020026
Elle contient les molécules commercialement disponibles chez 29 vendeurs en 2008-2009 (Albany, Research ART-CHEM, Asinex, Asis Chem, ChemBridge, ChemDiv, ChemStar, Enamine, FluroChem, InterBioScreen, IVK Labs, Key Organics, Life Chemicals, Maybridge, Nanosyn, Oakwood Chemicals, Otava Chemicals, Peakdale, Pharmeks, Princeton Chemicals, SALOR, Specs, Spectrum, TimTec, TOSLab, Tripos, Ufark, UORSY, Vitas-M Lab), et qui répondent aux règles de Lipinski (Lipinski et al. 2001) et de Veber (Veber et al., 2002).		
CombiLib Chembridge	CL	336085
Elle contient des molécules produites à partir de 430 nouveaux scaffolds synthétisés par la société Chembridge. Ces composés sont différents de ceux de l'Express-Pick ou de n'importe quel vendeur afin de couvrir une partie différente de l'espace chimique.		
TOTAL		7974288

Tableau 2 -Présentation des diverses chimiothèques.

L'univers de l'espace chimique possible est très large : l'espace des petites molécules, c'est-à-dire les molécules d'un poids moléculaire inférieur à 500, est estimé à 10^{60} (Dobson, 2004). A contrario, l'espace chimique disponible à l'achat est bien plus restreint. Nous couvrons une fraction presque négligeable de l'espace chimique virtuel. Aujourd'hui, il existe un nombre important de bases de données de molécules chimiques, contenant soit des molécules de la chimie combinatoire, soit des substances naturelles, soit des molécules déjà connues comme intéressantes pour certaines cibles biologiques. Tout cela contribue à élargir l'espace chimique réel. En raison des problèmes de synthèse, on ne considère souvent que les molécules caractérisées expérimentalement. Pour exemple, j'ai téléchargé sept bases de données, contenant toutes des molécules en deux dimensions (2D), et majoritairement disponibles à l'achat. Elles regroupent seulement 8 millions de molécules, ce qui reste extrêmement faible à l'échelle de l'espace chimique possible.

2. Préparation

1) Nettoyage des chimiothèques

Comme évoqué précédemment, la première étape de n'importe quel protocole de criblage virtuel est de supprimer les composés indésirables des chimiothèques. Pour cela, nous avons choisi le programme **Filter** (OpenEye Scientific Software Inc.). Il permet de filtrer les composés selon divers critères :

- la présence de certains éléments ou de groupes fonctionnels
- les caractéristiques telles que le poids moléculaire ou la solubilité
- la règle des 5 de Lipinski (Lipinski et al., 2001)

Le programme Filter peut fixer un état de protonation cohérent avec un pH neutre pour toutes les molécules d'une chimiothèque. Cela permet de s'assurer que tous les groupes carboxyles soient déprotonés et toutes les alkylamines soient protonées. De manière plus générale, chaque groupe fonctionnel sera représenté d'une unique façon.

Le problème des duplicats dans les chimiothèques est récurrent, surtout lorsque les molécules ont été récupérées à partir de diverses sources. Filter peut être utilisé pour éliminer ces duplicats rapidement lors du processus de filtre. De cette façon, les scores ne seront pas calculés deux fois pour la même molécule plus tard dans le protocole.

Le protocole de nettoyage que j'ai développé inclut la suppression des sels, des métaux et des duplicats. Les composés ionisables ont été convertis dans leur forme la plus probable à pH=7,4. Les composés les plus gros, c'est-à-dire dont le poids moléculaire dépasse 1000 daltons, ont été supprimés. Pour la base de données Ambinter, seules les molécules entre 200 et 600 daltons ont été gardées. Pour ce faire, le logiciel filter avec les options -unique true (pour supprimer les duplicats) et -filter element_filter.txt (filtrant sur la composition en atomes et sur le poids moléculaire, Figure 19) ont été utilisés. La protonation à pH neutre est effectuée par défaut, il est donc inutile d'ajouter une option.

```
#!/*****  
#Copyright (C) 2000-2005, 2010 by OpenEye Scientific Software, Inc.  
#*****/  
#This file defines the rules for filtering multi-structure files based on  
#properties and substructure patterns.  
  
MIN_MOLWT 0 (200 pour Ambinter)  
MAX_MOLWT 1000 (600 pour Ambinter)  
  
ALLOWED_ELEMENTS H,C,N,O,F,P,S,Cl,Br,I
```

Figure 19 - Contenu du fichier element_filter.txt.

2) Stéréoisomérisation

La stéréoisomérisation n'est qu'une notion comparative entre deux molécules possédant la même formule semi-développée mais qui ont un arrangement différent dans l'espace (Figure 20). Deux stéréoisomères sont deux configurations possibles d'une molécule et peuvent avoir des comportements tout à fait différents, vis-à-vis d'une cible protéique par exemple. C'est pourquoi il est important de les énumérer dans le cas où celle-ci n'est pas spécifiée.

La distribution Omega d'OpenEye inclut **flipper**, un programme qui permet d'énumérer les stéréoisomères d'une molécule. Flipper prend en charge l'énumération de la stéréochimie R/S (incluant N et S) ainsi que la stéréochimie cis/trans. Flipper utilise des algorithmes de graphes pour déterminer quels atomes sont des stéréocentres. Si un stéréocentre a une stéréochimie déjà désignée, par défaut, flipper la laissera telle quelle. Par contre, si le stéréocentre n'a pas de stéréochimie particulière, flipper va énumérer les deux états possibles de stéréochimie au niveau du stéréocentre.

Flipper doit être utilisé en amont de la génération des coordonnées tridimensionnelles. C'est pourquoi les molécules en entrée doivent être au plus en deux dimensions. On utilise les valeurs par défaut : **flipper -in input.sdf -out output.sdf**. Ce programme n'est pas capable d'ajouter un numéro à chaque stéréoisomère. J'ai donc écrit un script python, addNBstereo, qui ajoute un numéro à chaque identifiant afin que chaque structure ait un nom unique.

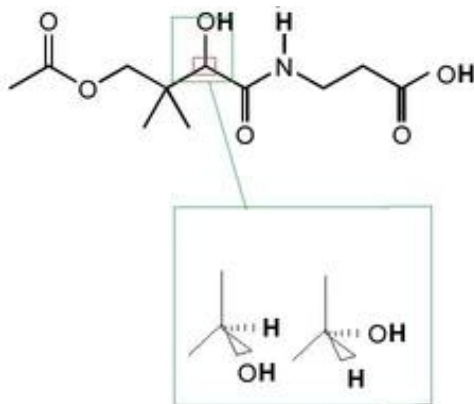


Figure 20 - Deux stéréoisomères possibles au niveau d'un stéréocentre carbone.

3) Génération de conformères

a) Définition

Les conformations d'une molécule sont les arrangements des atomes qui ne se différencient que par des rotations autour de liaisons simples. Ce terme a été introduit dans les années 1930 par N. Haworth lors de son étude des sucres. La représentation statique des molécules telles qu'on la pratique couramment n'est qu'une simplification graphique qui ne correspond pas à la réalité. En fait, les atomes sont animés de mouvements d'ampleur limitée les uns par rapport aux autres. Il existe des déformations de part et d'autre d'une position moyenne que l'on peut mettre en évidence par spectrométrie infra-rouge notamment : déformations angulaires qui affectent les angles de liaison, étirements et vibrations des liaisons qui affectent leur longueur, des rotations libres si l'orbitale de liaison possède une symétrie axiale et des déformations d'ampleur plus ou moins limitée si la molécule possède un degré de liberté moindre comme dans les cycles saturés à plus de cinq atomes. Tous ces mouvements se font sans qu'il y ait rupture de liaison et on passe librement d'une forme à une autre moyennant un très faible apport d'énergie : quelques kcal/mol (<15). Les différentes positions limites prises par la molécule lorsque ces déformations se produisent s'appellent les conformations. On dit aussi que la molécule se présente sous différents conformères.

Pour générer les conformations possibles de chaque stéréoisomère, j'ai choisi deux programmes : Omega de OpenEye et Corina. J'ai utilisé le premier pour générer les conformères acycliques et le second avec une option spécifique pour générer les conformères de cycle.

b) Omega

Omega est un programme utilisé pour convertir les structures en deux dimensions en trois dimensions. Omega utilise une méthode systématique basée sur des contraintes d'angle de torsion pour générer les conformations d'une molécule. Les règles sont déduites d'analyse des conformations de molécules trouvées dans les bases de données de structures expérimentales comme la CSD et la PDB. Omega s'assure que les conformères produits sont bas en énergie en utilisant le champ de force MMFF94. La combinaison d'une production de conformères systématique, basée sur des règles avec un seuil d'énergie approprié assure que les conformères générés par Omega sont à la fois divers et explorent les régions de basse énergie de l'espace conformationnel, incluant ceux proches de la conformation expérimentale.

L'option `-maxconfs` permet de donner le nombre maximum de conformations à générer. Par défaut, `-fromCT` a pour valeur `VRAI`, c'est-à-dire qu'Omega génère un ensemble initial de coordonnées en utilisant uniquement la table de connection de la molécule en entrée. Il y trouve en plus des connections établies entre les atomes, les informations concernant la stéréochimie mais aussi les états de charge de chaque atome.

c) Corina

Corina (Molecular Networks) est un programme développé par le groupe de recherche du Professeur Johann Gasteiger de l'Université d'Erlangen, en Allemagne. Corina génère automatiquement des structures en trois dimensions à partir de la constitution d'une molécule exprimée dans une table de connection ou un code linéaire (SMILES). Il utilise pour cela des angles de torsion référencés dans la Cambridge Structural Database : les "templates". Le programme peut également prendre en compte la variabilité de certains cycles. Ce programme ne calcule pas d'énergie et ne fait pas de minimisation, à la différence d'Omega. C'est donc un programme puissant et fiable pour convertir d'importantes chimiothèques contenant plusieurs milliers de structures.

Le script développé dans l'outil interne LEA3D, `sdfnoh`, a des options spécifiques de Corina : `-d rc,names,mc=NB,r2d,flapn,sc,3dst`, qui font que ce programme ne génère que des conformères de cycle (option `rc`), et considère de façon simultanée les différents cycles présents dans une molécule (option `rs`). Il est possible de donner un nombre maximum de conformères souhaité (option `mc=NB`). La numérotation des divers conformères est faite automatiquement sous la forme `cXXXX` (option `names`).

3. Préparation spécifique aux deux projets

A partir de cette étape, deux types de chimiothèque ont été préparées selon l'usage que l'on voulait en faire. Pour le projet *PfAMA1*, elle contient des conformères de cycle et des conformères acycliques. Pour le projet *PfCCT*, elle contient des tautomères et des conformères de cycle.

1) **Projet *PfAMA1* : conformères des parties cycliques et acycliques**

Dans ce projet, c'est un peptide, *PfRON2*, qui est notre ligand de référence en interaction avec la protéine *PfAMA1*. Il n'est donc pas possible d'utiliser certaines méthodes ligand-based, comme la similarité de forme ou les familles structurales. Notre choix s'est porté vers le pharmacophore car seuls divers points répartis dans l'espace et ayant une spécificité physicochimique est nécessaire pour le définir.

La recherche par pharmacophore, pour être efficace, implique d'avoir bien pris en compte la flexibilité des molécules. C'est pourquoi on va préparer les chimiothèques en terme de conformères pour les parties cycliques puis pour les parties acycliques des molécules.

J'ai d'abord utilisé Corina pour générer jusqu'à dix conformations de cycle (mc=10) puis j'ai utilisé Omega pour générer jusqu'à 10 conformations acycliques par conformère de cycle (- maxconfs 10).

2) **Projet *PfCCT* : filtre selon le volume, préparation des tautomères et des conformères de cycles**

a) **Filtre selon le volume**

Comme vous pouvez le voir dans la figure 21, la poche catalytique de la *PfCCT* est plutôt concave. Ainsi, seul un nombre limité de molécules peuvent potentiellement s'y loger. Le volume du site actif de la *PfCCT* a donc été estimé grâce à l'outil Voidoo. La poche aurait un volume d'environ 630 Angströms cube. J'ai alors choisi une valeur limite concernant le volume des molécules de 700 Angströms cube, et j'ai supprimé les molécules dont le volume était supérieur. (Kleywegt & Jones, 1994)

Pour cela, il était nécessaire que chaque molécule soit en trois dimensions. J'ai donc utilisé Omega avec -maxconfs=1. De cette sorte, une seule conformation est générée, celle de plus basse énergie. Ensuite, j'ai écrit un script, `filtre_volume.py`, qui prend un fichier sdf en entrée et renvoie un fichier contenant les molécules dont le volume est inférieur à 700 Angströms cube, et un fichier contenant les molécules trop grosses.

Ce script fait appel à un script perl (volume.pl), qui calcule le volume de chaque molécule par le programme NSC (Eisenhaber et al., 1995) avec pour référence les rayons de van der Waals tels qu'ils sont donnés par (Bondi, 1964).

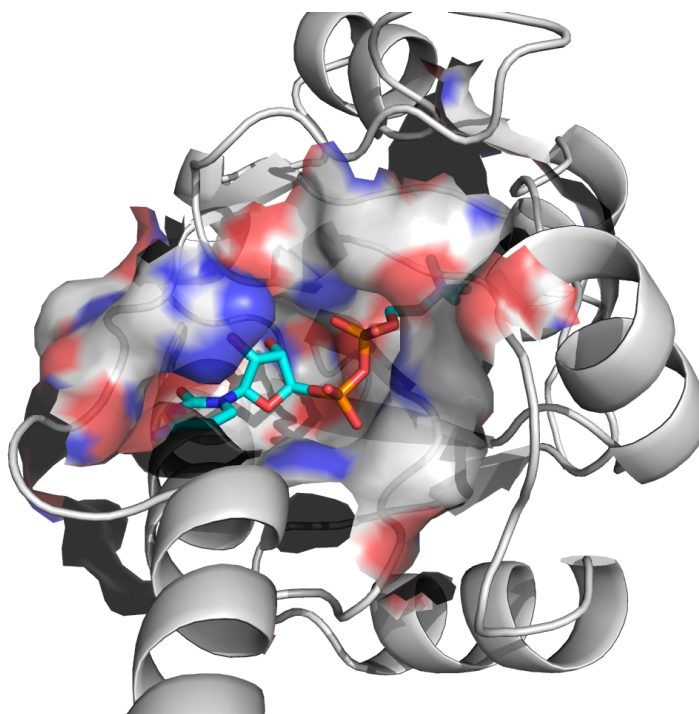


Figure 21 - CDP-choline dans la poche du monomère N-terminal de la *PfCCT*.

b) Génération des tautomères

Ensuite, afin de mieux représenter chaque molécule par ses différents états possibles, j'ai utilisé le programme **tautomers** inclu dans QuacPac d'OpenEye pour générer les formes tautomères les plus probables. Un composé présente du tautomérisme s'il peut être représenté par deux structures qui sont liées par un mouvement intramoléculaire d'un hydrogène d'un atome à un autre (Figure 22). Il est important de savoir que ces réarrangements tautomériques produisent des empreintes, des structures tridimensionnelles, des propriétés électrostatiques et des liaisons hydrogènes différentes. Les protéines se lient parfois préférentiellement à un tautomère présent en faible abondance. C'est pourquoi la génération des tautomères est un aspect à ne pas oublier lors de la préparation de chimiothèques destinées au criblage virtuel (Martin, 2009).

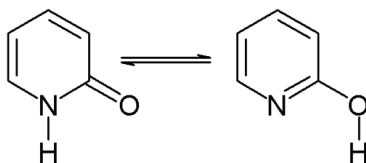


Figure 22 - Deux tautomères en équilibre de la 2-pyridone.

c) Génération des conformères de cycles

Une fois les tautomères énumérés, j'ai utilisé Corina pour générer les conformères de cycle (jusqu'à 20 ou 100 selon les chimiothèques). Je n'ai pas généré de conformères acycliques car les outils de docking utilisés lors du filtre "structure-based" simulent cette flexibilité.

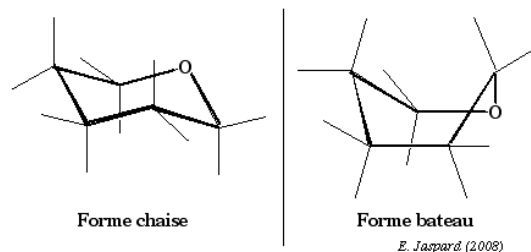


Figure 23 - Deux conformères du tétrahydropyrane.

Ces préparations suppriment quelques molécules ne répondant pas aux premiers critères de filter mais les étapes de génération de stéréoisomères, tautomères et conformères augmentent considérablement la taille des chimiothèques.

	Nombre de structures	Nombre de molécules
DB		6631 molécules
<i>PfAMA1</i>	237770	6018
<i>PfCCT</i>	254825	5935
CN		43 878 molécules
<i>PfAMA1</i>	3369554	40463
<i>PfCCT</i>	2493966	40243
CHEMBL		28 379 molécules
<i>PfAMA1</i>	1558659	26674
<i>PfCCT</i>	1167303	26449
CB		485 778 molécules
<i>PfAMA1</i>	5683251	481196
<i>PfCCT</i>	6060089	481421
DL		2 020 026 molécules
<i>PfAMA1</i>	53207674	1967476
<i>PfCCT</i>	32316253	1967820
AMB		5 053 511 molécules
<i>PfAMA1</i>	122921055	4895514
<i>PfCCT</i>	77313514	4899277
CL		336 085 molécules
<i>PfAMA1</i>	16746149	326863
<i>PfCCT</i>	10917375	326858

Tableau 3 - Les chimiothèques, le nombre de structures et de molécules après préparation pour chaque projet : *PfAMA1* ou *PfCCT* (Pour les différentes étapes, se référer à l'Annexe 2).

III. Outils de criblage virtuel

Je vais vous présenter les différents logiciels utilisés dans ces protocoles de criblages virtuels et les méthodes qui y sont développées. La combinaison spécifique de chacun d'entre eux pour chaque projet sera explicitée dans les chapitres 2 et 3, concernant respectivement le projet *PfAMA1* et le projet *PfCCT*.

1. Les approches 'ligand-based'

Cette démarche repose sur les connaissances acquises concernant d'autres molécules interagissant avec la cible d'intérêt biologique. Ces autres composés peuvent alors être utilisés comme références pour définir les caractéristiques que doit posséder une molécule afin d'interagir avec le système biologique d'intérêt. Au minimum, une technique ligand-based n'a besoin que d'une seule molécule active comme point de départ. Les outils présentés ci-après sont des logiciels qui permettent de trouver des similarités avec une référence de type ligand.

1) Méthode utilisant les structures 2D : empreintes moléculaires ECFP4

ECFP signifie Extended-Connectivity Fingerprints. Ce sont des empreintes topologiques circulaires développées pour la caractérisation moléculaire et la recherche de similarités. Elles sont très populaires parmi les outils de recherche de similarités dans la découverte de médicaments. (Rogers & Hahn, 2010)

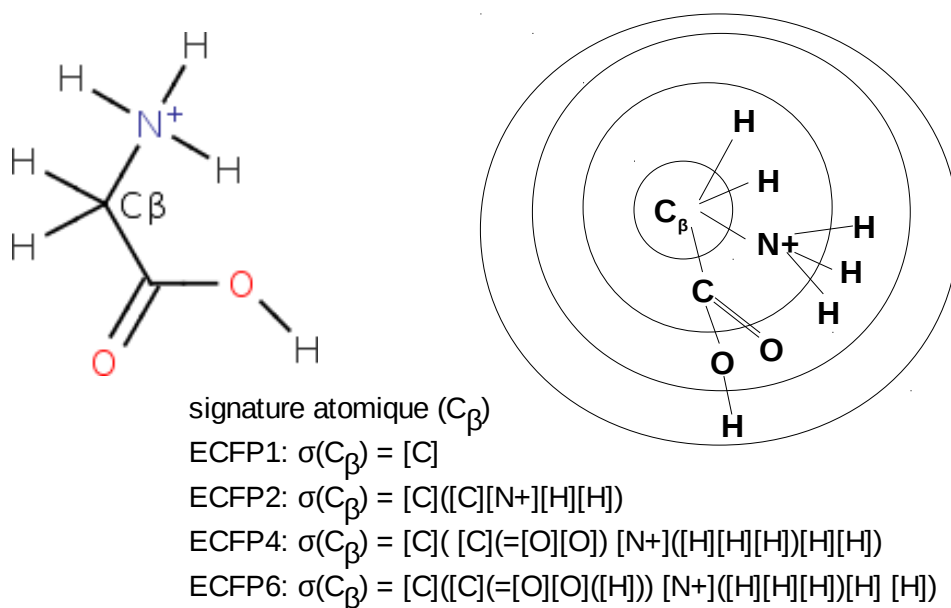
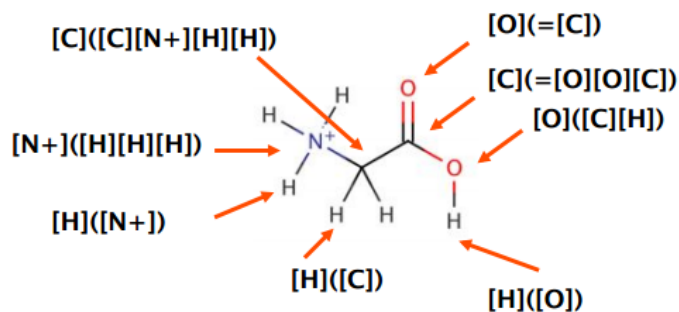


Figure 24 - Exemple de calcul d'ECFP de diamètre 1 à 6 pour l'atome C_{β} de la glycine.



$${}^1\sigma(\text{glycine}) = [\text{O}](=[\text{C}]) + [\text{O}]([\text{C}][\text{H}]) + [\text{N}+]([\text{H}][\text{H}][\text{H}]) + [\text{C}](=[\text{O}][\text{O}][\text{C}]) + [\text{C}]([\text{C}][\text{N}+][\text{H}][\text{H}]) + [\text{H}]([\text{O}]) + 3 [\text{H}]([\text{N}+]) + 2 [\text{H}]([\text{C}])$$

Figure 25 - Exemple de calcul d'ECFP1 pour la glycine entière.

Ces empreintes constituent ainsi une signature moléculaire. Pour cela, il s'agit de faire la somme des signatures de tous les atomes de la molécule. Ces empreintes représentent des structures moléculaires au moyen du voisinage circulaire d'atomes. Elles peuvent être très vite calculées. Elles ne sont pas prédéfinies et peuvent représenter un nombre très important de caractéristiques moléculaires différentes (incluant les informations stéréochimiques). Elles sont développées pour représenter à la fois la présence ou l'absence de fonctionnalité, car les deux sont cruciaux pour analyser l'activité moléculaire.

Ces empreintes circulaires apportent une méthode de criblage rapide et très efficace pour la recherche de similarités, en approximant mieux les attentes d'un chimiste médicinal que les empreintes linéaires (de type 'path-based'). Les empreintes linéaires sont utilisées pour chercher la présence de sous-structures particulières dans une molécule. Elles consistent à couper la molécule en fragments possibles de longueurs diverses, et de coder la présence ou l'absence de cette partie d'empreinte par 1 ou 0. Les empreintes circulaires sont plus appropriées que les empreintes linéaires pour les molécules contenant des ramifications ou des structures plus complexes, comme des cycles fusionnés (Bender et al., 2009)

2) Méthodes utilisant les structures 3D

Etant donné que les molécules se liant à leur protéine cible exercent sur elle un effet par leur structure en 3D, il y a un intérêt constant dans les techniques 3D dans le criblage ligand-based. La technique la plus connue est la recherche par pharmacophore, qui tente d'extraire des groupements fonctionnels d'une molécule active (ou d'un ensemble de molécules actives) importants dans la liaison à la protéine cible. Le criblage est alors conduit par l'identification de molécules qui pourraient potentiellement correspondre à cet ensemble de caractéristiques.

Une autre technique 3D tente de donner un score aux molécules basé sur leur similarité de forme globale par rapport à une molécule de référence, contrairement au pharmacophore qui cherche seulement leur capacité à correspondre à un set de caractéristiques.

a) approche basée sur la forme : ROCS (Grant et al., 1998)

Divers logiciels sont basés sur le concept que des molécules ayant des formes similaires vont pouvoir se loger dans le même site actif. Aujourd'hui, le logiciel standard pour cette méthode est ROCS, signifiant Rapid Overlay of Chemical Structures. Mais il en existe d'autres : Catalyst (Accelrys), Parashift (Livingstone et al., 2007), Hex, ShaEP (Vainio et al., 2009)...

ROCS est un programme de comparaison rapide de la forme, basé sur le fait que les molécules ont une forme similaire si leurs volumes se superposent bien. Elle utilise des fonctions Gaussiennes pour représenter le volume de chaque molécule. Cette représentation mathématique permet de minimiser la meilleure superposition globale aisément. C'est un outil puissant de criblage virtuel, qui peut rapidement identifier des composés avec une forme similaire à un composé de référence. La rapidité de ROCS, 1000 à 2000 superpositions par seconde soit 20 à 40 molécules par seconde, permet de cribler une base entière de multi-conformères en seulement un jour sur un processeur. C'est la méthode ligand-based la plus rapide. Des travaux récents ont montré que ROCS fait aussi bien, voire mieux, que des approches de type structure-based, en terme de performance globale mais aussi en terme de qualité sur certains cas d'étude (Hawkins et al., 2007),(Venhorst et al., 2008).

ROCS va au delà de l'approche abstraite des pharmacophores, en utilisant la totalité de la forme moléculaire de la référence, avec ses caractéristiques chimiques en plus. Les molécules des chimiothèques se voient attribuer un score relatif à la molécule de référence en utilisant une mesure physique rigoureuse de la similarité en trois dimensions, avec en parallèle une mesure du niveau de correspondance des groupements chimiques (donneur ou accepteur d'hydrogène, charge positive et négative, centre aromatique, groupement hydrophobe). Ainsi, ROCS est la combinaison d'une approche rigoureuse de superposition de forme et d'une estimation simple de la similarité chimique. Cette approche permet de donner des bons scores pas seulement aux molécules structurellement proches de la molécule de référence, mais aussi aux molécules de classes chimiques très différentes de celle de la référence ('scaffold hopping').

b) Pharmacophores

i. Qu'est-ce qu'un pharmacophore?

Selon la définition officielle de l'IUPAC (International Union of Pure and Applied Chemistry) en 1998, le pharmacophore est un ensemble de caractéristiques stériques et électroniques nécessaires à assurer les interactions supramoléculaires optimales avec une structure d'une cible biologique spécifique et d'en déclencher ou d'en diminuer la réponse biologique.

Un pharmacophore est donc un concept abstrait, c'est un ensemble des groupes fonctionnels importants (présence de charges, donneur/accepteur de liaisons hydrogènes, groupe hydrophobe...), dont la présence est indispensable pour que la molécule soit active. Le pharmacophore définit également la position relative dans l'espace de chacun de ces groupes les uns par rapport aux autres. La modélisation moléculaire et/ou le criblage basé sur les similarités de pharmacophore s'est révélé être une méthode importante et utile dans la découverte de nouvelles molécules. Un exemple de découverte grâce à ce principe fut le tirofiban en 1992. Elle mime un tripeptide, RGD, qui est la séquence minimale du fibrinogène se fixant au complexe de glycoprotéines GP IIb/IIIa. Ils ont cherché dans la collection Merck des composés qui possédaient une fonction amine et une fonction carboxylate séparées par 10 à 20 Angströms (approximation de la distance entre la guanidine et l'acide carboxylique de RGD). C'est alors devenu un anticoagulant, qui a été mis sur le marché en 1999 sous le nom d'Aggrastat.

Cette méthode est l'une des plus utilisées actuellement dans le criblage virtuel. En considérant les données sur les molécules actives, il sera possible de déduire des interactions clés et des groupements chimiques importants pour l'activité vis-à-vis de la cible. Ainsi, on pourra rechercher des analogues, répondant aux critères géométriques définis.

ii. Le programme Align-it, anciennement Pharao (Taminau et al., 2008)

Align-it est un outil pour aligner les molécules selon leurs pharmacophores. Les fonctionnalités de cet outil sont doubles : générer les pharmacophores à partir des structures moléculaires (Figure 26) et aligner ces pharmacophores à une molécule de référence donnée. Un score est alors calculé, qui correspond au volume commun résultant de l'alignement. Chaque point de pharmacophore est représenté par un volume gaussien en trois dimensions, défini par son centre et sa diffusion, et est caractérisé par un type fonctionnel. Ces types fonctionnels sont considérés comme étant importants dans la liaison sélective des molécules. Ils sont au nombre de neuf : aromatique, donneur de liaison hydrogène, accepteur de liaison hydrogène, lipophile, charge négative, charge positive, donneur et accepteur de liaison hydrogène, aromatique et lipophile, sphère d'exclusion.

L'alignement utilisé dans Align-it est rigide, c'est-à-dire qu'il n'y a pas de flexibilité apportée aux molécules et le programme ne fonctionne qu'avec une seule conformation fixe à la fois. C'est pourquoi il faut l'utiliser sur des chimiothèques préalablement préparées en terme de conformères.

Les coordonnées des molécules vont être modifiées car chacune des molécules va être alignée avec la molécule de référence. Align-it propose trois scores différents pour évaluer la similarité entre deux pharmacophores. V_o est le volume de chevauchement maximal entre les deux pharmacophores. V_{ref} est le volume du pharmacophore de référence, V_{db} est le volume du pharmacophore de la molécule qu'on cherche à évaluer. Le score TANIMOTO est une mesure globale et la mesure par défaut pour évaluer la similarité entre deux pharmacophores. La mesure TVERSKY_REF est destinée à identifier des composés dont le pharmacophore est un sur-ensemble du pharmacophore de référence. La mesure TVERSKY_DB est destinée à identifier des composés dont le pharmacophore est un sous-ensemble du pharmacophore de référence. Toutes ces mesures retournent des valeurs entre 0 et 1.

$$TANIMOTO = V_o / (V_{ref} + V_{db} - V_o)$$

$$TVERSKY_REF = V_o / V_{ref}$$

$$TVERSKY_DB = V_o / V_{db}$$

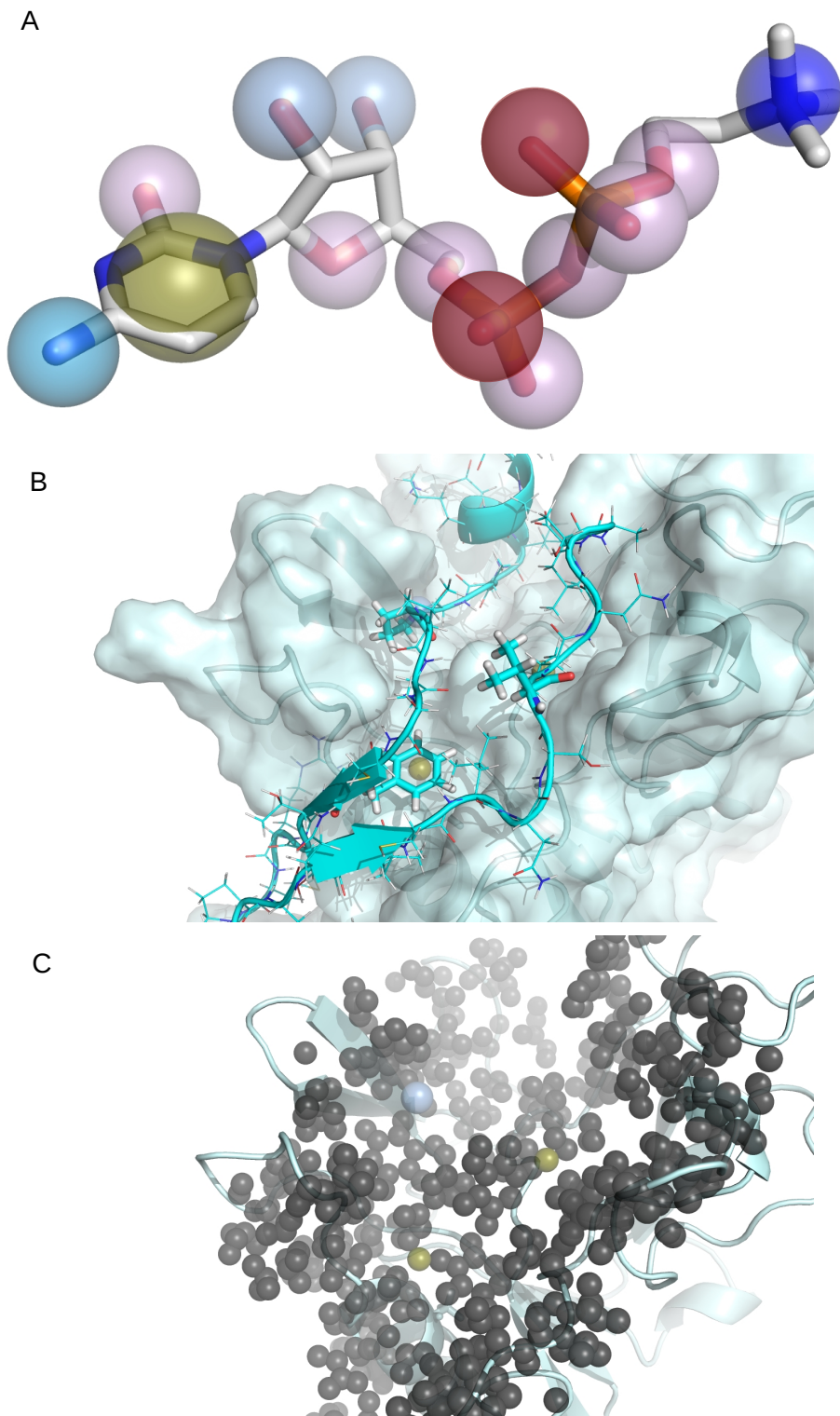


Figure 26 - Exemple d'un pharmacophore généré par Align-it à partir A) d'une molécule ou B) d'un peptide. Chaque couleur de sphère représente un type fonctionnel : jaune, aromatique et lipophile; bleu ciel, donneur de liaison hydrogène; rose, accepteur de liaison hydrogène; bleu clair, donneur et accepteur de liaison hydrogène; rouge, charge négative; bleu foncé, charge positive C) Ajout des sphères d'exclusion (en gris) comme contrainte.

iii. Le programme PepMMsMIMIC (Floris et al., 2011)

PepMMsMIMIC est un outil de criblage virtuel accessible uniquement en ligne (<http://mms.dsfarm.unipd.it/pepMMsMIMIC/>) (Figure 27)) nécessitant une interaction protéine-protéine/peptide comme point de départ. Son but est de proposer des composés chimiques contenant des éléments mimant un peptide naturel. Ainsi, on espère que les composés retenus garderont la capacité à interagir avec la cible protéique pour produire l'effet attendu. Plusieurs méthodes de scoring ont été développées. Celle qui permet de réduire le nombre de faux positifs est une fonction hybride prenant en compte le score de similarité de forme basé sur des méthodes USR (Ultrafast Shape Recognition) pour 40% du score total et le score du matching du pharmacophore (Pharmacophoric Fingerprint Similarity) pour 60% du score total. Cet outil ne donne pas les coordonnées de superposition (Floris et al., 2011).

A

1. Upload a valid PDB file

Choisissez un fichier Aucun fichier choisi Upload

or load an example peptide Load

[see DOI:10.2210/pdb1ycr/pdb, PDB ID: 1YCR, chain B]

B

2. Please pick at least 3 residues from the Jmol window

[VAL]2054:B

[PHE]2038:B	Side chain <input checked="" type="checkbox"/>	backbone CO <input type="checkbox"/>	backbone NH <input type="checkbox"/>	✗
[PRO]2033:B	Side chain <input checked="" type="checkbox"/>	backbone CO <input type="checkbox"/>	backbone NH <input type="checkbox"/>	✗
[VAL]2054:B	Side chain <input checked="" type="checkbox"/>	backbone CO <input type="checkbox"/>	backbone NH <input type="checkbox"/>	✗

3. Select the scoring method

hybrid search (60% pharmacoforic similarity, 40% shape)

Enter your e-mail address for job completion alert (optional)

pihan@ipmc.cnrs.fr

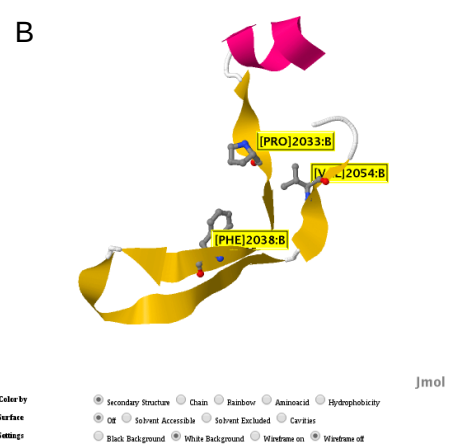


Figure 27 - Serveur PepMMsMIMIC en ligne : A) choix des paramètres de recherche et B) choix des résidus constituant le pharmacophore.

iv. Le programme Pharmer (Koes & Carlos J Camacho, 2011)

Pharmer est un outil de recherche et d'alignement de pharmacophore intéressant dans le sens où il stocke les pharmacophores des chimiothèques dans une base de données pour accélérer la recherche. Le pharmacophore d'une molécule est un ensemble de pharmacophores à trois points formant des triangles enregistrés avec des coordonnées exactes dans un index spatial. Une requête est alors décomposée en triangles, et les résultats de ces requêtes dans l'index spatial sont reconstruits dans un alignement de composés par rapport au pharmacophore de référence. Pour améliorer l'extensivité de la recherche, une approche de filtre de Bloom est utilisée pour représenter les empreintes moléculaires (Bloom, 1970). Un filtre de Bloom est une structure de données probabiliste compacte inventée par Burton Howard Bloom en 1970. Lors du test de la présence d'un triangle dans un pharmacophore entier, un filtre de Bloom permet de savoir avec certitude que le triangle est absent du pharmacophore entier (il ne peut pas y avoir de faux négatif) et avec une certaine probabilité que le triangle peut être présent dans le pharmacophore entier (il peut y avoir des faux positifs). La taille d'un filtre de Bloom est fixe et indépendante du nombre d'éléments contenus (la structure peut donc être extrêmement compacte). Ce genre de filtre permet d'éviter des appels inutiles à une très grande base de données en vérifiant tout de suite qu'un triangle n'est pas présent. Le filtre n'étant pas parfait, la recherche inutile aura toutefois lieu dans certains cas, mais une grande partie sera néanmoins évitée, multipliant ainsi le nombre de requêtes utiles possibles à matériel donné. Cet outil est donc extrêmement rapide une fois les chimiothèques préparées. En effet, il est capable de cribler deux millions de structures en moins d'une minute (Koes & Camacho, 2011).

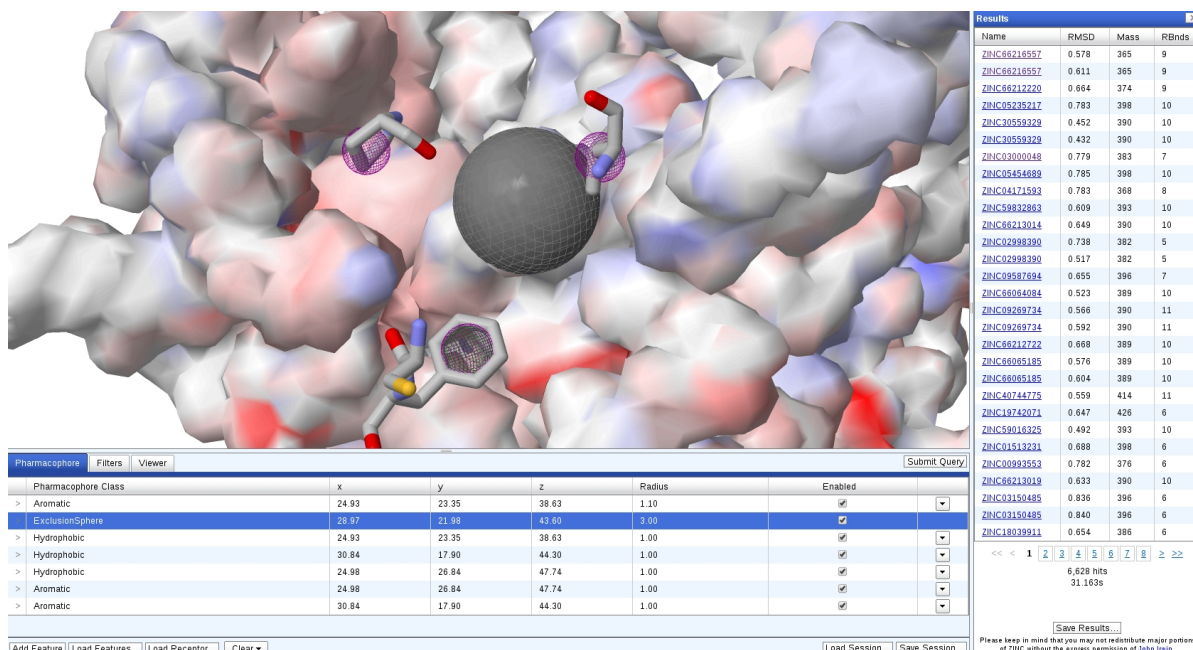


Figure 28 - Serveur ZINCPharmer en ligne avec la représentation du pharmacophore.

2. Les approches 'structure-based'

Cette technique repose sur les connaissances de la structure tridimensionnelle de la cible biologique. Ces connaissances sont obtenues par des méthodes expérimentales telles que la cristallographie par rayons X et la spectroscopie par résonance magnétique nucléaire (RMN). La base de données PDB est la base de référence des structures tridimensionnelles des grandes molécules biologiques en complexe, ou non, des protéines et des acides nucléiques. La PDB contient 89921 structures en date du 29 Avril 2013. Par contre, pour les structures non déterminées expérimentalement, il est parfois possible de réaliser un modèle, construit à l'aide de la modélisation par homologie. Ce modèle est généralement basé sur la structure expérimentale d'une protéine proche (homologue). Le criblage virtuel de modèles par homologie reste plus difficile à cause des approximations sur les coordonnées des atomes.

1) Qu'est ce que l'amarrage moléculaire ('docking')?

L'interaction entre une protéine et son substrat est la première étape de la plupart des réactions biologiques. Caractériser la liaison et définir quels sont les résidus mis en jeu, est donc primordial pour pouvoir expliquer les mécanismes qui influent sur l'affinité entre deux molécules. De même, la découverte de nouvelles drogues activant ou inhibant l'activité biologique d'une protéine peut se faire en prédisant leur affinité. C'est dans ce but que des techniques de modélisation moléculaire, regroupées sous le nom d'amarrage moléculaire ou "docking", ont été développées. Il est à noter que les substrats et les inhibiteurs sont définis par le terme générique de "ligand", tandis que le terme "récepteur" est utilisé pour la cible sur laquelle le ligand va se fixer, que ce soit une enzyme ou un récepteur protéique.

Il y a deux aspects principaux dans le docking. Tout d'abord, une fonction de **score** doit permettre de discriminer les bons complexes des mauvais, en évaluant l'affinité entre le ligand et le récepteur. Différentes fonctions de score existent : celles basées sur un champ de force (utilisant les énergies non-liantes), celles empiriques (dérivées de mesures expérimentales) et celles basées sur les connaissances (théoriques). Ensuite, l'**algorithme** de recherche doit trouver le meilleur complexe mesuré par la fonction de score, sachant que le nombre de complexes possibles est astronomique. Afin d'obtenir la meilleure pose entre le ligand et le récepteur, trois catégories d'algorithmes de recherche existent : la recherche stochastique (recuits simulés, calculs de Monte-Carlo, algorithmes génétiques (Autodock)), la construction incrémentale, et la génération d'un ensemble de conformères de faible énergie.

Aujourd'hui, le challenge dans le domaine du docking reste entier. Il faut pouvoir énumérer proprement les conformations du ligand (et du récepteur dans le cas du docking flexible). Ce problème peut être amélioré grâce à l'augmentation significative des ressources de calculs, et au développement de nouveaux algorithmes de recherche efficaces (conformations

précalculées, énumération ou non des tautomères). Le scoring quant à lui est un problème plus important. Le développement de nouvelles fonctions de score conceptuellement différentes va être nécessaire (consensus de fonctions de score, ajout d'un terme d'entropie, ajout d'un terme de solvation, fonctions de score spécifiques de la cible protéique, "machine-learning") (Yuriev & Ramsland, 2013)

Parmi tous les programmes d'amarrage moléculaire qui ont déjà prouvé leur efficacité, nous avons choisi d'utiliser Vina et PLANTS dans nos protocoles de criblage virtuel.

2) Le programme Vina (Trott & Olson, 2010)

Autodock Vina est un programme de docking moléculaire et de criblage virtuel, qui offre une bonne performance et une simplicité d'utilisation. Autodock Vina est une optimisation du programme AutoDock (Morris & Huey, 2009), qui permet d'imposer au système l'exploration de l'espace conformationnel en autorisant certains degrés de liberté au ligand, tout en maintenant fixe le récepteur. La fonction de score utilisée dans Autodock Vina est différente de celle utilisée dans Autodock (empirique, basée sur l'énergie libre de liaison), ce qui fait de Vina un nouveau logiciel. Ces programmes sont tous les deux conçus pour être capables de gérer les torsions des ligands. Une comparaison de ces deux outils a été publiée et Autodock Vina montre une plus grande efficacité, notamment dans le traitement des ligands avec un nombre de torsions important. De plus, il est plus rapide et permet de tirer avantage de multiples processeurs (Chang et al., 2010).

L'**algorithme** du programme de docking Vina est un algorithme de recherche d'optimisation globale stochastique de type Iterated Local Search. Au cours du docking, plusieurs calculs parallèles vont être effectués à partir de conformations aléatoires. Chacun de ces calculs est réalisé sur un certain nombre de pas, qui consistent en une perturbation aléatoire de la conformation suivie d'une étape d'optimisation locale réalisée selon la méthode Broyden-Fletcher-Goldfarb-Shanno (BFGS). Après chaque pas, la conformation obtenue est acceptée ou rejetée selon le critère de Metropolis, le nombre de pas étant établi en fonction de la complexité du problème étudié, c'est à dire de la taille de l'espace de recherche et du nombre de torsions autorisées sur le ligand. Les résultats de tous les calculs sont ensuite rassemblés et clusterisés afin de fournir une liste de solutions définitives classées en fonction de leur score. Le paramètre « exhaustiveness » (exhaustivité) régit le nombre de conformations aléatoires et donc de calculs menés sur le système. Ces calculs pouvant être réalisés en parallèle sur des processeurs différents, ce paramètre est directement lié au nombre de processeurs employés. C'est le seul élément qui permette d'influencer le résultat de docking, Autodock Vina étant conçu pour pouvoir être utilisé de façon très simple.

La fonction de **score** développée dans le programme de docking Vina est une combinaison d'une approche basée sur les connaissances et d'une approche empirique. Cette dernière permet de faire la somme des contributions énergétiques intramoléculaires et intermoléculaires du système d'une manière inspirée par le programme Xscore, qui utilise une fonction de score empirique. Elle s'appuie sur un training set issu de la PDBbind. Cependant, contrairement à son modèle, la fonction de score d'Autodock Vina n'exploite pas seulement les valeurs d'affinité des ligands, mais également les préférences conformationnelles des complexes du training set. Elle combine ainsi les avantages des fonctions de score empiriques et des fonctions de score basées sur les connaissances.

La fonction de score empirique comprend les termes suivants :

$$\Delta G_{\text{binding}} = \Delta G_{\text{gauss}} + \Delta G_{\text{repulsion}} + \Delta G_{\text{hbond}} + \Delta G_{\text{hydrophobic}} + \Delta G_{\text{tors}}$$

ΔG_{gauss} est constitué de deux fonctions gaussiennes et représente le terme stérique

$\Delta G_{\text{repulsion}}$ est égal au carré de la distance entre atomes est défavorable

ΔG_{hbond} correspond au terme des interactions hydrogènes

$\Delta G_{\text{hydrophobic}}$ correspond au terme des interactions hydrophobes

ΔG_{tors} est proportionnel au nombre des liaisons flexibles du ligand

Cette fonction de score a été calibrée avec 1300 complexes provenant de la base de données PDB-Bind. L'erreur standard est de 2.85 kcal/mol.

Pendant le docking, tous les résidus de la protéine sont gardés fixes tandis que les ligands sont considérés comme flexibles. Les torsions sur les ligands sont attribuées automatiquement lors de la conversion du format MOL2 au format interne PDBQT. Le site actif a été défini manuellement de manière à englober la totalité des résidus constituant le site actif. Le centre de la boîte se trouve au niveau du centre de gravité de la CDP-choline en complexe avec la PfCCT. Les dimensions de la boîte sont de 25×25×25 Å³.

Les paramètres par défaut de Vina utilisés pendant les calculs dans la structure de la PfCCT dont les résidus sont fixes (docking rigide), sont :

- exhaustiveness = 8
- nombre de poses générées = 9

Les fichiers de sortie de Vina en PDBQT ont été convertis en PDB à l'aide de scripts écrits dans le laboratoire.

3) Le programme PLANTS (Korb et al., 2006)

L'**algorithme** du programme de docking PLANTS, pour Protein-Ligand ANT System, est basé sur une classe d'algorithmes d'optimisation stochastique appelé "optimisation par colonies de fourmis" (Ant Colony Optimization). C'est la première fois que cette technique est utilisée pour résoudre le problème du docking protéine-ligand. L'optimisation par colonies de fourmis s'inspire du comportement des fourmis lorsque celles-ci sont à la recherche de nourriture. Il a ainsi été démontré qu'en plaçant une source de nourriture reliée au nid par une passerelle, formée d'une branche courte et d'une branche longue, les fourmis choisissaient toutes le chemin le plus court après un certain laps de temps. En effet, chaque fourmi se dirige en tenant compte des traces de phéromone déposées par les autres membres de la colonie qui la précèdent. Comme cette phéromone s'évapore, ce choix probabiliste évolue continuellement. Les fourmis marquent les meilleures solutions, et tiennent compte des marquages précédents pour optimiser leur recherche.

Dans le cas du docking protéine-ligand, cette colonie de fourmis artificielles est utilisée pour trouver la conformation d'énergie minimale du ligand dans le site actif. PLANTS considère le ligand comme flexible. Les fourmis sont utilisées pour marquer les conformations de basse énergie par des chemins de phéromones. Le déplacement du ligand et les angles de torsion sont des variables continues. Pour pouvoir utiliser l'optimisation par colonies de fourmis, une discrétisation de ces valeurs a été nécessaire. Pour les trois degrés de liberté de translation, une longueur d'intervalle de 0.1 Angström a été choisie. Pour tous les degrés de liberté de torsion, un intervalle de 1° a été pris. La piste de phéromones artificielle est modifiée à chaque itération pour générer des conformations de basse énergie avec une plus grande probabilité. (Korb et al., 2006).

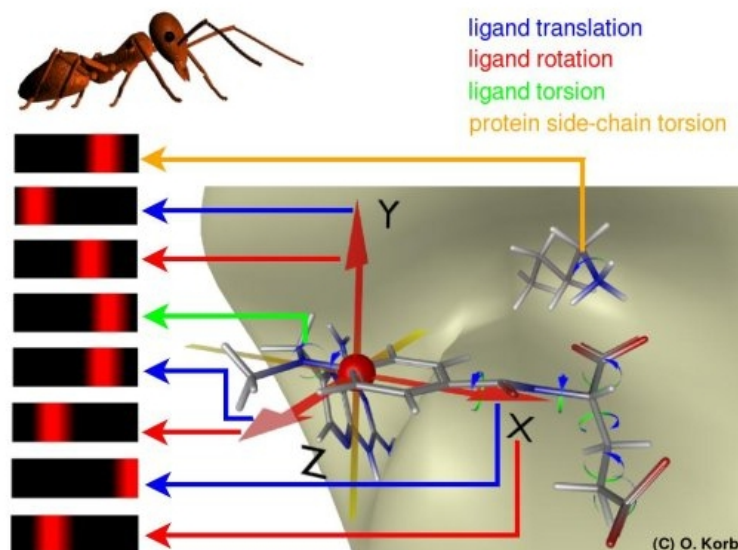


Figure 29 - Méthode de recherche du programme de docking PLANTS.

Dans cette méthode, la qualité de la fonction de score utilisée pour évaluer les complexes est d'une importance majeure dans le succès de la méthode. La fonction de **score** standard développée dans le programme de docking Plants est empirique et s'appelle $f_{\text{PLANTS}_{\text{CHEMPLP}}}$ (Korb & Stützle, 2009). Elle est dérivée de telle sorte à reproduire les affinités des complexes déterminées expérimentalement. Elle est basée en partie sur des fonctions de scores et des champs de force déjà publiés. La fonction de score PLP (f_{plp}), pour Piecewise Linear Potentiel, est utilisée pour modéliser la complémentarité stérique entre la protéine et le ligand. D'autres termes du Chemscore du programme de docking GOLD rendent compte de l'importance de l'angle dans les termes de liaison hydrogène (f_{hb} , $f_{\text{hb-ch}}$, $f_{\text{hb-CHO}}$) et de liaison aux métaux (f_{met} , $f_{\text{met-coord}}$, $f_{\text{met-ch}}$, $f_{\text{met-coord-ch}}$). Le potentiel de torsion du champ de force Tripos (f_{tors}) couplé à un terme de clash des atomes lourds (f_{clash}) est utilisé pour prendre en compte les interactions intraligand. Enfin, le site actif étant défini par une sphère dans PLANTS, la valeur c_{site} rend compte de la position par un potentiel quadratique si le ligand est hors de la sphère.

$$f_{\text{PLANTS}_{\text{CHEMPLP}}} = f_{\text{plp}} + f_{\text{hb}} + f_{\text{hb-ch}} + f_{\text{hb-CHO}} + f_{\text{met}} + f_{\text{met-coord}} + f_{\text{met-ch}} + f_{\text{met-coord-ch}} + f_{\text{clash}} + f_{\text{tors}} + c_{\text{site}}$$

Afin de paramétriser le score, 31 structures de complexes ont été utilisées, différentes fonctions ont été testées et celle retrouvant le maximum de positions natives avec un RMSD minimal a été retenue (Model M3 : 25/31 bonnes positions, RMSD moyen 1,45 Angströms).

4) Utilisation de Ligplot pour post-filtrer les solutions de docking

Suite au docking, il faut pouvoir classer les molécules. Le score calculé par chaque programme sera une première indication. Mais parfois, les interactions établies grâce à la position préférentielle trouvée par le logiciel de docking vont être aussi pertinentes. C'est pourquoi, grâce au logiciel Ligplot, j'ai ajouté cette composante aux méthodes 'structure-based'.

Le programme Ligplot génère automatiquement une représentation schématique en deux dimensions d'un complexe protéine-ligand ou protéine-protéine. Par défaut, c'est le programme HBPLUS (McDonald & Thornton, 1994) qui liste l'ensemble des liaisons hydrogènes et interactions hydrophobes au sein de la structure, respectivement dans *filename.hhb* et *filename.nnb*. Ligplot va alors lire ces fichiers de sortie en plus du fichier *filename.pdb* pour générer ses propres fichiers nécessaires à la construction des diagrammes d'interaction. On va donc pouvoir savoir quels résidus sont impliqués dans les interactions hydrogènes en cherchant dans *ligplot.hhb*.

Le fichier **ligplot.sum** est intéressant car il résume en un seul fichier les noms et les numéros des résidus impliqués dans n'importe quel type d'interaction (hydrogène ou hydrophobe). Ainsi, si l'on connaît les résidus avec lesquels on veut qu'une interaction s'établisse, de n'importe quel type, on peut aller chercher dans ce fichier si ce résidu est présent. De plus, la partie du résidu qui interagit est notifié par M (main-chain) pour la chaîne principale et S (side-chain) pour la chaîne secondaire. Ligplot a été développé pour permettre une inspection rapide d'un nombre important de complexes protéine-ligand, et devient dans ce sens un outil très utile dans un protocole de criblage virtuel.

Voici des exemples de fichiers ligplot.sum (Projet PfAMA1) et ligplot.hhb (Projet PfCCT).

```

Composé 3
60810067 de la CombiLib (Hit Pharmer)
ligplot.sum output :

```

Residues:	Interacting residues						Hydrogen bonds Non-bonded			
	M	M	S	M	S	M	S	M	S	contacts
[MOL 1 1] -> [TYR 251 A]	0	0	0	0	0	0	0	0	0	3
[MOL 1 1] -> [TYR 236 A]	0	0	0	0	0	0	0	0	0	2
[MOL 1 1] -> [TYR 234 A]	0	0	0	0	0	0	0	0	0	6
[MOL 1 1] -> [PRO 226 A]	0	0	0	0	0	0	0	0	0	2
[MOL 1 1] -> [MET 224 A]	0	0	0	0	0	0	0	0	0	1
[MOL 1 1] -> [PRO 184 A]	0	0	0	0	0	0	0	0	0	1
[MOL 1 1] -> [PHE 183 A]	0	0	0	0	0	0	0	0	0	2
[MOL 1 1] -> [THR 171 A]	0	0	0	0	0	0	0	0	0	1
[MOL 1 1] -> [VAL 169 A]	0	0	0	0	0	0	0	0	0	1
[MOL 1 1] -> [TYR 142 A]	0	0	0	0	0	0	0	0	0	2

```

Composé 12
Amb16782937_0@2_5 d'Ambinter (Hit ROCS)
ligplot.hhb output :

```

Donor	Acceptor	Distance
VAL A 40 N	lig d 1 024	3.08
TYR A 41 N	lig d 1 024	3.30
LYS A 78 NZ	lig d 1 023	2.88
HIS A 124 NE2	lig d 1 022	2.90
lig d 1 N19	TYR A 129 OH	3.27
THR A 171 N	lig d 1 021	3.06

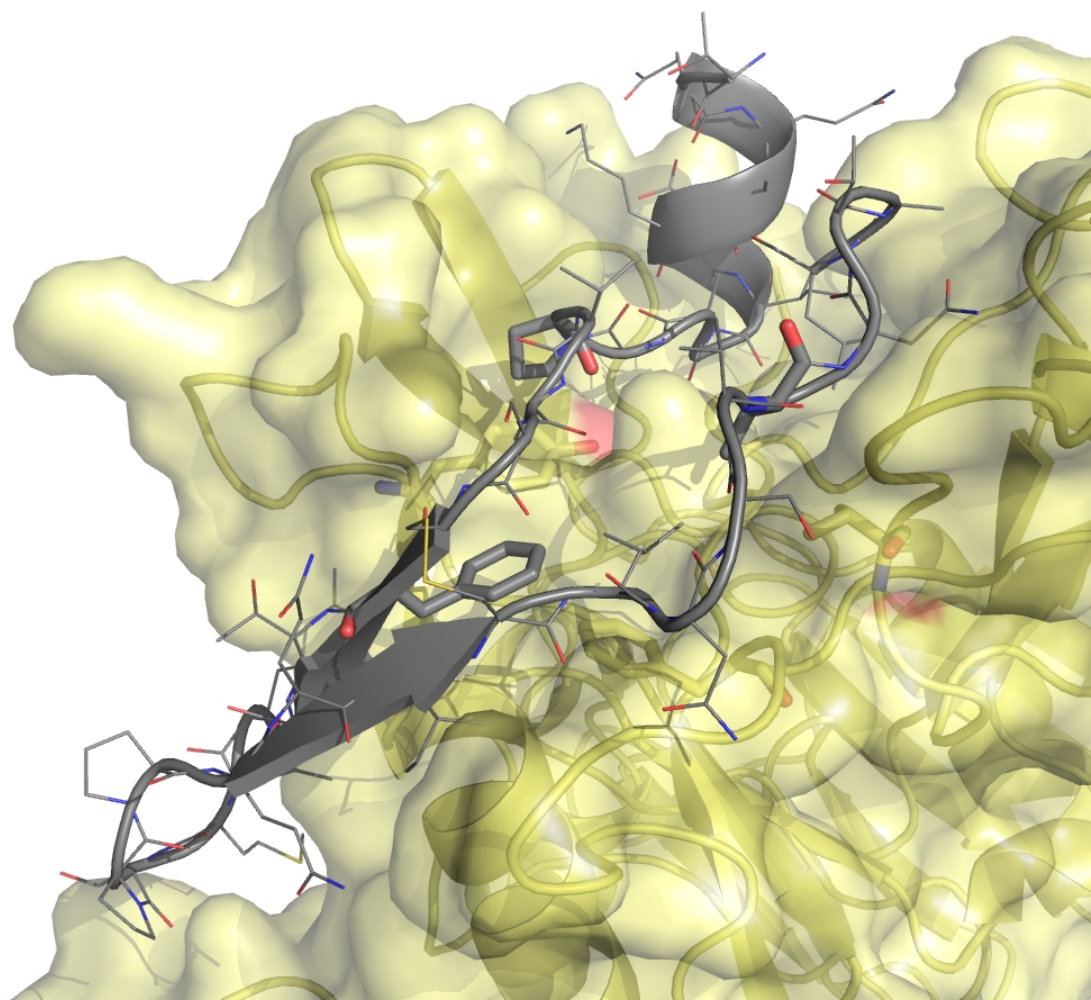
Figure 30 - Contenu des fichiers ligplot.sum et ligplot.hhb.

5) Visualisation des poses de docking : Pymol 3D

La sélection finale des molécules à commander pour l'évaluation expérimentale passe par une inspection manuelle et individuelle des interactions 3D de chaque touche virtuelle avec le récepteur. Pour cela, j'ai utilisé le logiciel de visualisation Pymol avec la 3D. Le choix se fera aussi selon la disponibilité des molécules chez leurs fournisseurs respectifs. En effet, selon le laps de temps séparant le téléchargement du catalogue électronique de la chimiothèque et la commande, le pourcentage de molécules devenues indisponibles augmente significativement.

En résumé de ce chapitre, le criblage virtuel de chimiothèques est devenu une méthode chémoinformatique utilisée en routine de manière à identifier des ligands pour des cibles d'intérêt thérapeutique. Il faut garder en mémoire que cette technologie est très dépendante de la précision des coordonnées 3D de la molécule de référence et de la cible et génère donc de nombreux faux négatifs. Les phases de préparation de la chimiothèque et de la fouille des résultats pour détecter de faux positifs sont aussi importantes que le criblage automatique lui-même. Il est plus intéressant de considérer le nombre de nouveaux chémotypes dans les ligands identifiés et validés par criblage que le taux de touches. De ce point de vue, l'outil qu'est le criblage virtuel est un complément naturel du chimiste médicinal afin de proposer des châssis moléculaires susceptibles de conduire rapidement à des chimiothèques focalisées à haute valeur ajoutée. Nous allons maintenant voir ce que peut apporter cette méthode informatique au travers de deux projets, que j'ai développés pendant ma thèse. Le but général est d'identifier de nouvelles molécules antipaludiques. Pour cela, deux cibles ont été identifiées au sein du laboratoire de Dynamique des Interactions Membranaires Normales et Pathologiques de Montpellier (DIMNP). Le chapitre 2 concerne l'identification des molécules inhibant une interaction protéine-protéine impliquée dans la jonction mobile : *PfAMA1/PfRON2*. Le chapitre 3, quant à lui, s'intéresse plus particulièrement à l'enzyme *PfCCT*, impliquée dans la voie de biosynthèse des phospholipides.

**Chapitre 2 : Bloquage du processus
d'invasion des érythrocytes par le
parasite en ciblant le complexe
d'attache AMA1 RON2 chez
*Plasmodium falciparum***



I. Contexte biologique

1. Les interactions protéine-protéine (PPIs)

Les interactions protéine-protéine jouent des rôles cruciaux dans de nombreux processus biologiques, comme certaines pathologies ou les interactions hôte-pathogène. Des petites régions de l'interface de ces PPIs, appelés "hot-spots", peuvent être responsables d'une extrême spécificité et sont essentielles à une interaction de haute affinité. Ainsi, ces sites orthostériques (reconnus par le ligand agoniste endogène de l'enzyme) sont des cibles intéressantes pour des inhibiteurs de PPIs car ils permettront d'interférer d'une manière très spécifique.

La Nutlin3 est une molécule ciblant une interaction protéine-protéine. Elle inhibe l'interaction du complexe protéique MDM2-p53, ce qui a pour conséquence de réactiver la fonction de la protéine suppresseur de tumeur p53 (LaRusch et al., 2007). Cette molécule est actuellement en étude clinique pour le traitement contre le cancer. D'autres petites molécules inhibitrices de PPIs, ciblant le VIH, la dengue ou le virus Ebola ont pu être identifiées. Ainsi, ces molécules inhibant les PPIs offrent des opportunités pour cibler les interactions hôte-pathogène et prévenir certaines maladies (Jubb et al., 2012).

2. Problématique

Cette première étude concerne la recherche de composés inhibiteurs de l'interaction AMA1-RON2 chez *Plasmodium falciparum*, qui a été présentée dans l'introduction. Ces travaux ont dans un premier temps nécessité une bonne compréhension des données biologiques et structurales, afin d'en dériver un pharmacophore efficace. Dans un deuxième temps, nous avons utilisé ce pharmacophore lors d'un protocole de criblage virtuel. Enfin, grâce à une collaboration, l'équipe de Maryse Lebrun, directrice de recherche à l'UMR5235 DIMNP Montpellier, a testé les molécules retenues suite au criblage virtuel dans un test de croissance parasitaire et dans un test *in vitro* d'invasion. Elles ont aussi été testées par SPR (Surface Plasmon Resonance) pour détecter une interaction avec la protéine PfAMA1.

II. Données biologiques et structurales

1. Présentation du complexe AMA1-RON2

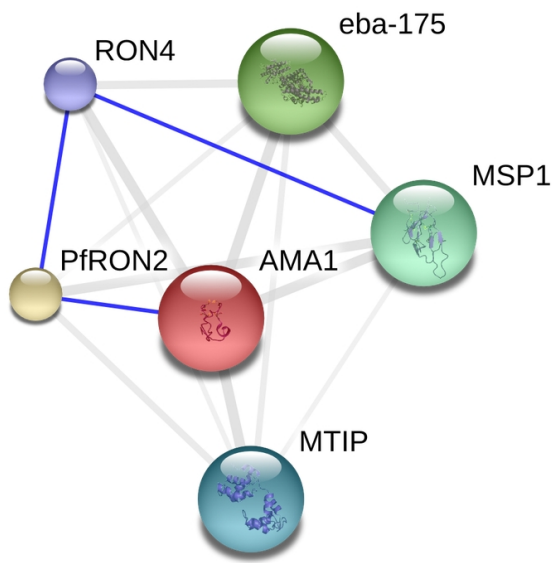
L'interrogation de la base de données UniprotKB permet de retrouver la protéine Apical Membrane Antigen 1 de *Plasmodium falciparum* (isolat 3D7). Son numéro d'accèsion est le Q7KQK5 (Q7KQK5_PLAF7). Cette protéine est codée par le gène *ama1* (numéro d'accès dans PlasmoDB : PF11_0344). La taxonomie auquel cette protéine appartient est la suivante : Eukaryota > Alveolata > Apicomplexa > Aconoidasida > Haemosporida > Plasmodium > Plasmodium (Laverania). Les annotations GO nous indiquent que cette protéine est impliquée dans l'entrée du parasite dans la cellule hôte (GO:0030260), ce qui explique sa pathogénicité (GO:0009405). En effet, elle fait partie du complexe apical (GO:0020007) et est présente à la membrane (GO:0016020).

Plusieurs structures cristallographiques de *PfAMA1* ont été résolues et sont disponibles dans la Protein Data Bank: 1Z40 (*PfAMA1* seule), 2Z8V et 2Z8W (*PfAMA1* en complexe avec un anticorps IgNAR), 2Q8A et 2Q8B (*PfAMA1* en complexe avec l'anticorps 1F9), 3SRJ (*PfAMA1* en complexe avec le peptide inhibiteur R1), 3SRI (*PfAMA1* en complexe avec le peptide de 29 acides aminés dérivé de *PfRON2sp2*) et 3ZWZ (*PfAMA1* en complexe avec le peptide de 39 acides aminés dérivé de *PfRON2sp1*). 3SRJ, 3SRI et 3ZWZ sont les structures qui vont nous intéresser pour établir le protocole de criblage virtuel. Ce seront les références sur lesquelles nous nous appuierons.

```
>tr|Q7KQK5|Q7KQK5_PLAF7  Apical  membrane  antigen  1,  AMA1
OS=Plasmodium falciparum (isolate 3D7) GN=AMA1 PE=1 SV=1
MRKLYCVLLLSAFEFTYMINFGRGQNYWEHPYQNSDVYRPINEHREHPKEYEYPLHQEHT  60
YQQEDSGEDENTLQHAYPIDHEGAEPAPQEQLFSSIEIVERSNYMGNPWTTEYMAKYDIE 120
EVHGSGIRVDLGEDAEVAGTQYRLPSGKCPVFGKGIIENSNTTFLTPVATGNQYLKDGG 180
FAFPPTEPLMSPMTLDEMRFYKDNKYVKNLDELTLCSRHAGNMIPDNDKNSNYKYPAVY 240
DDKDKKCHILYIAAQENNGPRYCNKDESKRNSMFCFRPAKDISFQNYTYLSKNVVDNWEK 300
VCPRKNLQNAKFGLWVDGNCEDIPHVNEFPAIDLFECNKLVFELSASDQPKQYEQHLTDY 360
EKIKEGFKNKNASMIKSAFLPTGAFKADRYKSHGKGYNWGNYNTETQKCEIFNVKPTCLI 420
NSSYIATTALSHPIEVENNFPCSLYKDEIMKEIERESKRIKLNDNDDEGNKKIIAPRIF 480
ISDDKSLKPCDPEMVSNSTCRFFVCKCVERRAEVTSNNEVVVKEEYKDEYADIPEHKP 540
TYDKMKIIIIASSAAVAVLATILMVYLYKRKGNAEKYDKMDEPQDYGKSNSRNDEMLDPEA 600
SFWGEEKRASHTTPVLMEKPYY 622
```

Figure 31 - Séquence complète de *PfAMA1*, contenant 622 acides aminés. Le domaine DI est représenté en bleu, le domaine DII en vert, le domaine DIII en rouge et la boucle DII surlignée en jaune.

La base de données des interactions protéine-protéine connues et prédites, STRING (Franceschini et al., 2013), propose les partenaires les plus probables.



Le trait bleu représente une interaction directe entre *PfAMA1* et *PfRON2*. *PfRON2* est le premier partenaire fonctionnel prédit avec un score de 0.999. La protéine Rhoptry Neck 2, *PfRON2*, d'identifiant Uniprot Q8IKV6 (Q8IKV6_PLAF7), est codée par le gène *PfRON2*, au niveau du locus PF14_0495. La taxonomie auquel cette protéine appartient est la suivante : Eukaryota > Alveolata > Apicomplexa > Aconoidasida > Haemosporida > Plasmodium > Plasmodium (Laverania).

Figure 32 - Réseau d'interactions de la protéine *PfAMA1* (<http://string-db.org/>)

PfRON2 est une protéine transmembranaire complète de 2189 acides aminés. Plusieurs structures cristallographiques contenant une partie de *PfRON2*, celle se liant directement au sillon hydrophobe de *PfAMA1*, ont été résolues et sont disponibles dans la Protein Data Bank : 3SRI (*PfAMA1* en complexe avec le peptide de 29 acides aminés dérivé de *PfRON2*sp2) et 3ZWZ (*PfAMA1* en complexe avec le peptide de 39 acides aminés dérivé de *PfRON2*sp1). La structure entière de *PfRON2* n'est pas connue.

La structure de la protéine AMA1 a d'abord été bien décrite chez *Plasmodium vivax* (Pizarro et al., 2005), puis chez *Plasmodium falciparum* (Bai et al., 2005), *Toxoplasma* (Crawford et al., 2010), *Babesia* et *Neospora* (Tonkin et al., 2013). La structure reste la même chez les autres espèces de *Plasmodium*. La protéine AMA1 est constituée d'une région ectoplasmique N-terminale, d'une région transmembranaire et d'un petit domaine cytoplasmique C-terminal. Le domaine ectoplasmique est découpé en trois domaines distincts : DI [108-301], DII [302-441] et DIII [442-662]. Seuls les domaines I et II ont été résolues par cristallographie. Chacun de ces deux domaines porte en son coeur un domaine PAN conservé, qui consiste en cinq feuillets beta, autour d'une hélice alpha de 12 résidus (Pizarro et al., 2005), (Bai et al., 2005). Les domaines PAN d'AMA1 sont relativement exceptionnels quand on les compare à ceux d'autres protéines. Par exemple, seul un pont cystéine et six résidus hydrophobes sont conservés. En fait, en raison de la divergence de ces deux domaines, ils n'ont pas été tout de suite identifiés comme des domaines PAN. De plus, les boucles qui connectent les feuillets beta 1 et 3 et 4 et 5

adoptent des conformations variables selon les souches, et les feuillets beta eux-mêmes sont variables en longueur et en position. Ces irrégularités dans les domaines PAN, particulièrement les boucles flexibles connectant les feuillets bêta résultent de l'évolution d'AMA1. Cela permet de détourner les anticorps des régions fonctionnellement critiques du domaine PAN, fournissant au parasite une voie pour échapper aux réponses humorales (Healer et al., 2004).

Le domaine I est important pour la fonction biologique d'AMA1, car il contient un sillon hydrophobe d'une surface de 1700 Å² dans lequel une boucle du domaine II (DII loop) peut venir se loger. De cette manière, le sillon hydrophobe du domaine I est masqué à *PfRON2* (Vulliez-Le Normand et al., 2012), la protéine partenaire de *PfAMA1* lors de l'invasion (Figure 33).

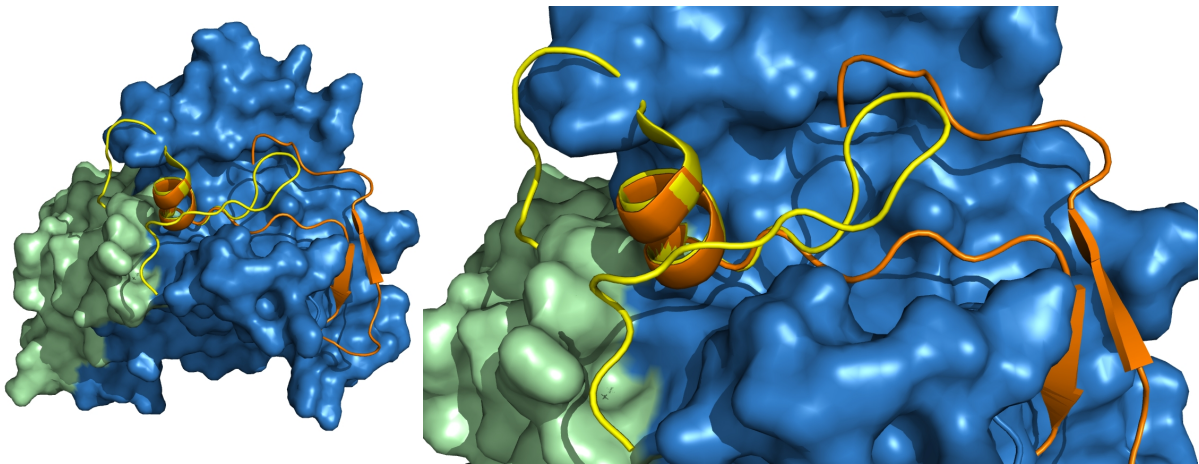


Figure 33 - Domaines I (bleu) et II (vert) de *PfAMA1* dans sa forme apo (où la boucle du domaine II (jaune) vient se loger dans le sillon hydrophobe) ou en interaction avec le peptide *PfRON2sp1* (orange), respectivement PDB 1Z40 (Bai et al., 2005) et 3ZWZ (Vulliez-Le Normand et al., 2012).

2. Résidus essentiels à l'interaction *PfAMA1-PfRON2*

Les membres du phylum des Apicomplexes, qui inclut le parasite du paludisme *Plasmodium*, partagent de nombreuses caractéristiques dans leurs mécanismes d'invasion malgré leurs spécificités vis à vis de leurs cellules hôtes et les caractéristiques de leurs cycles de vie. Une comparaison entre *Plasmodium* et *Toxoplasma* montre une insertion de deux acides aminés dans la boucle cystine de *PfRON2* par rapport à *TgRON2*, et la structure apo de *PfAMA1* montre un sillon apical étendu par rapport à celui de *Toxoplasma gondii*. Ceci est donc idéal pour compléter la boucle cystine étendue, qui en plus d'être importante pour l'interaction, doit avoir un rôle dans la régulation de la spécificité inter-espèce. (Tonkin et al., 2011)

Les résultats biologiques ont montré qu'il existe des "hot-spots" nécessaires à la formation spécifique du complexe *PfAMA1-PfRON2* (Vulliez-Le Normand et al., 2012). Des mutations de résidus clés chez *PfAMA1* ou *PfRON2* suppriment l'interaction entre les protéines

recombinantes (Tonkin et al., 2011). Les résultats sont visibles dans la figure 34. Les résidus importants de *Pf*RON2 pour l'interaction avec *Pf*AMA1 sont PRO2033 (cage hydrophobe), PRO2044, ARG2041 et PHE2038. Pour le partenaire AMA1, le résidu clé est PHE183. De plus, il a été montré que le peptide provenant de *Pf*RON2 inhibe l'invasion des érythrocytes par les mérozoïtes de *Plasmodium falciparum*.

Ces données permettent de localiser les zones clés de l'interaction. Elles sont importantes pour le design rationnel de petites molécules, de peptidomimétiques et des vaccins. Il va alors être envisageable de trouver des ligands qui répondent à ces critères physicochimiques spécifiques dans l'espace (pharmacophore).

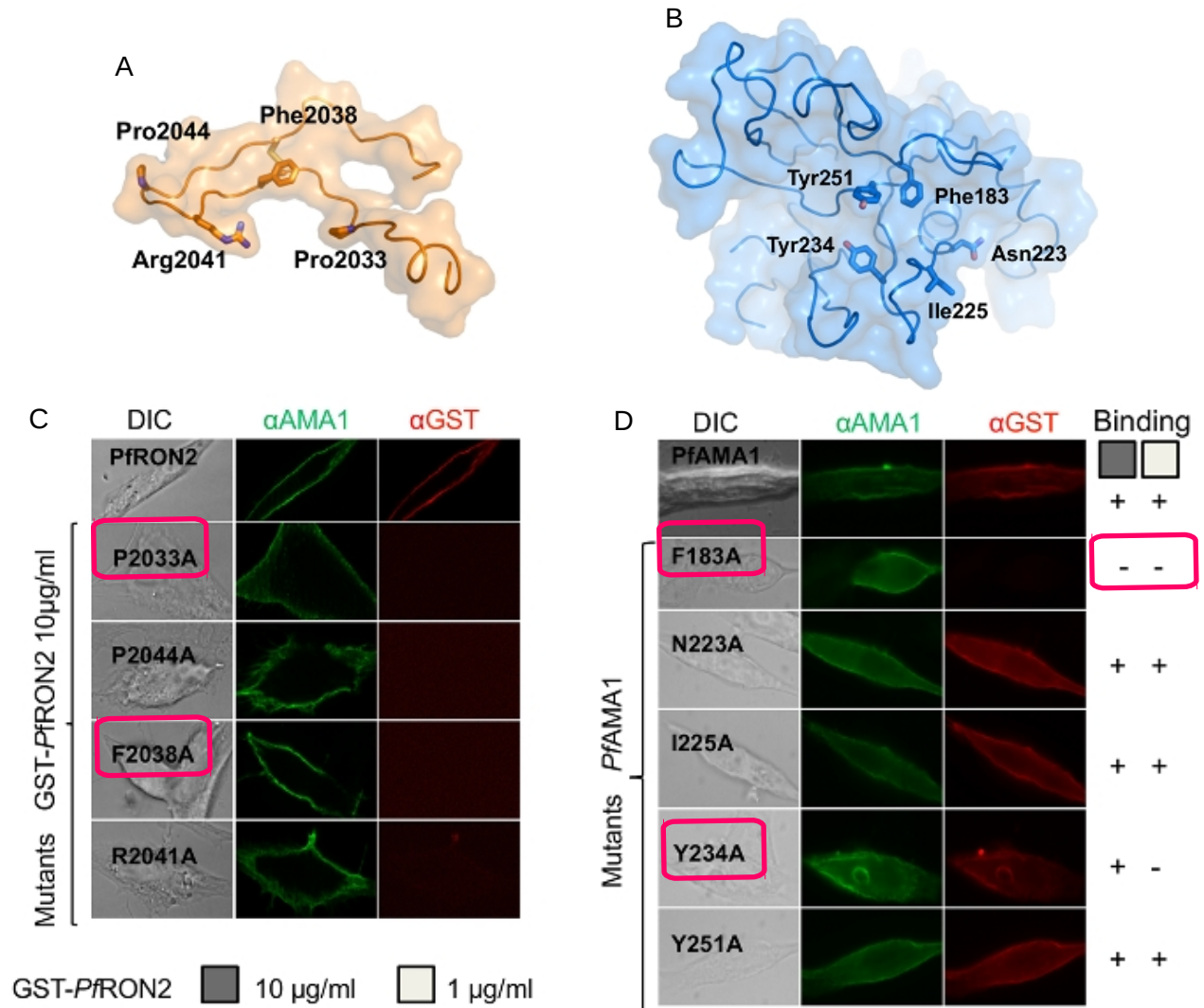


Figure 34 - Issue de (Tonkin et al., 2011). Des mutations effectuées sur *Pf*AMA1 et *Pf*RON2 révèlent des résidus essentiels à cette interaction. Position des résidus mutés dans la structure de A) *Pf*RON2 et de B) *Pf*AMA1. C) Les quatre résidus de *Pf*RON2 sont essentiels à l'interaction. En effet, leur mutation entraîne une perte de la colocalisation (colonne alpha-GST). D) Côté *Pf*AMA1, la mutation de la PHE183 abolit l'interaction et la mutation de la TYR234 la diminue.

3. Analyse des structures cristallographiques

La structure du complexe *PfAMA1* avec un peptide dérivé de la région extracellulaire de *PfRON2* a été résolue (PDB 3SRJ). Cette structure montre des spécificités claires concernant l'interaction *PfAMA1-PfRON2sp1*. Le site de liaison de *PfAMA1* comprend un sillon hydrophobe et une région qui devient exposée suite au déplacement de la 'DII loop'.

En parallèle, une autre structure a été déterminée : celle de *PfAMA1* en complexe avec le peptide R1. Ce dernier a été obtenu par une recherche de peptides inhibiteurs par la technique du phage display par le groupe de Foley. La méthode du phage display est une méthode qui permet la sélection d'un peptide grâce à sa présentation sur la surface de phages. Concrètement, *PfAMA1* a été immobilisée sur des plaques de 96 puits et exposée à une banque contenant plus de 5 fois 10^8 clones ayant à leur surface des peptides aléatoires de 20 acides aminés. Après lavage, les clones s'étant fixés sont élués, et c'est ainsi que le peptide R1 a été sélectionné comme le meilleur peptide (parmi deux) se fixant de manière spécifique à AMA1 (Harris et al., 2005).

Parmi les résidus clés de la poche hydrophobe de *PfAMA1*, il y a la phénylalanine 183, résidu nécessaire à la liaison à *PfRON2*. Cette phénylalanine est conservée au sein de tous les isolats de *Plasmodium falciparum*. Cela indique une contrainte fonctionnelle gouvernant une étape essentielle dans l'invasion et représente une cible nouvelle qui peut être potentiellement exploitée dans la thérapie antipaludique.

Cette structure a révélé que ce peptide R1 mimait le peptide *PfRON2* par sa position vis à vis de *PfAMA1*. Ces résultats ont permis de confirmer les résidus clés de l'interaction entre *PfAMA1* et *PfRON2*. La structure du ligand *PfRON2sp1* lié à AMA1 (PDB 3ZWZ) nous donne un point de départ idéal pour réaliser un criblage virtuel. Quand on compare les deux structures, on retrouve trois résidus : LEU6, LEU8 et PHE12 pour le peptide R1 et PRO2033, VAL2054 et PHE2038 pour *PfRON2sp1* en face respectivement de TYR234, VAL169 et PHE183 (Figure 35). Ceux-ci indiquent la position du pharmacophore consensus conforme aux trois poches décrites par (Tonkin et al., 2013).

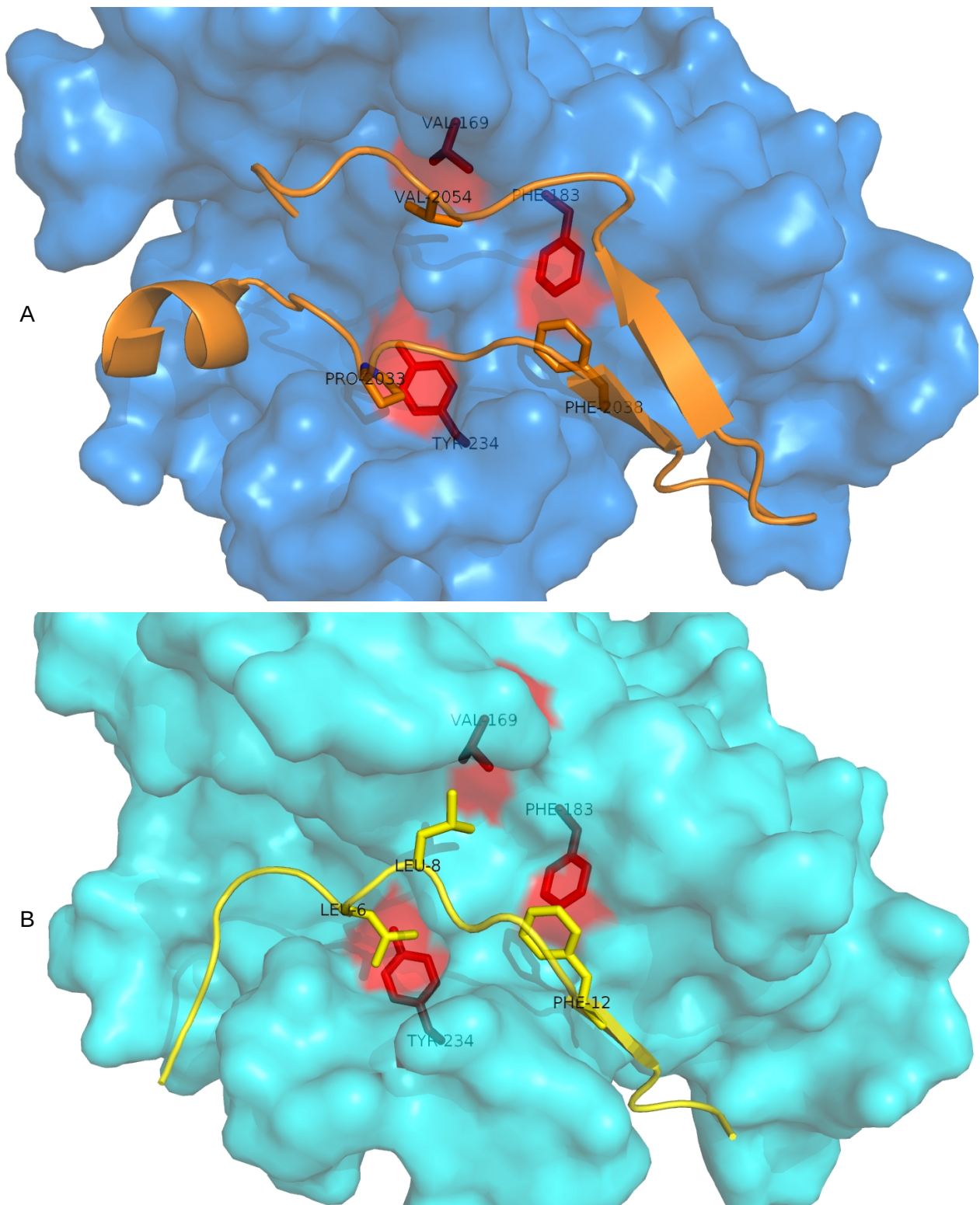


Figure 35 - La protéine AMA1 de *Plasmodium falciparum* en complexe avec
 A) PfrON2sp1 (PDB 3ZWZ)
 B) le peptide R1 (PDB 3SRJ)
 Les résidus clés de PfAMA1, impliqués dans l'interaction sont colorés en rouge.

III. Protocole de criblage virtuel

1. Schéma général

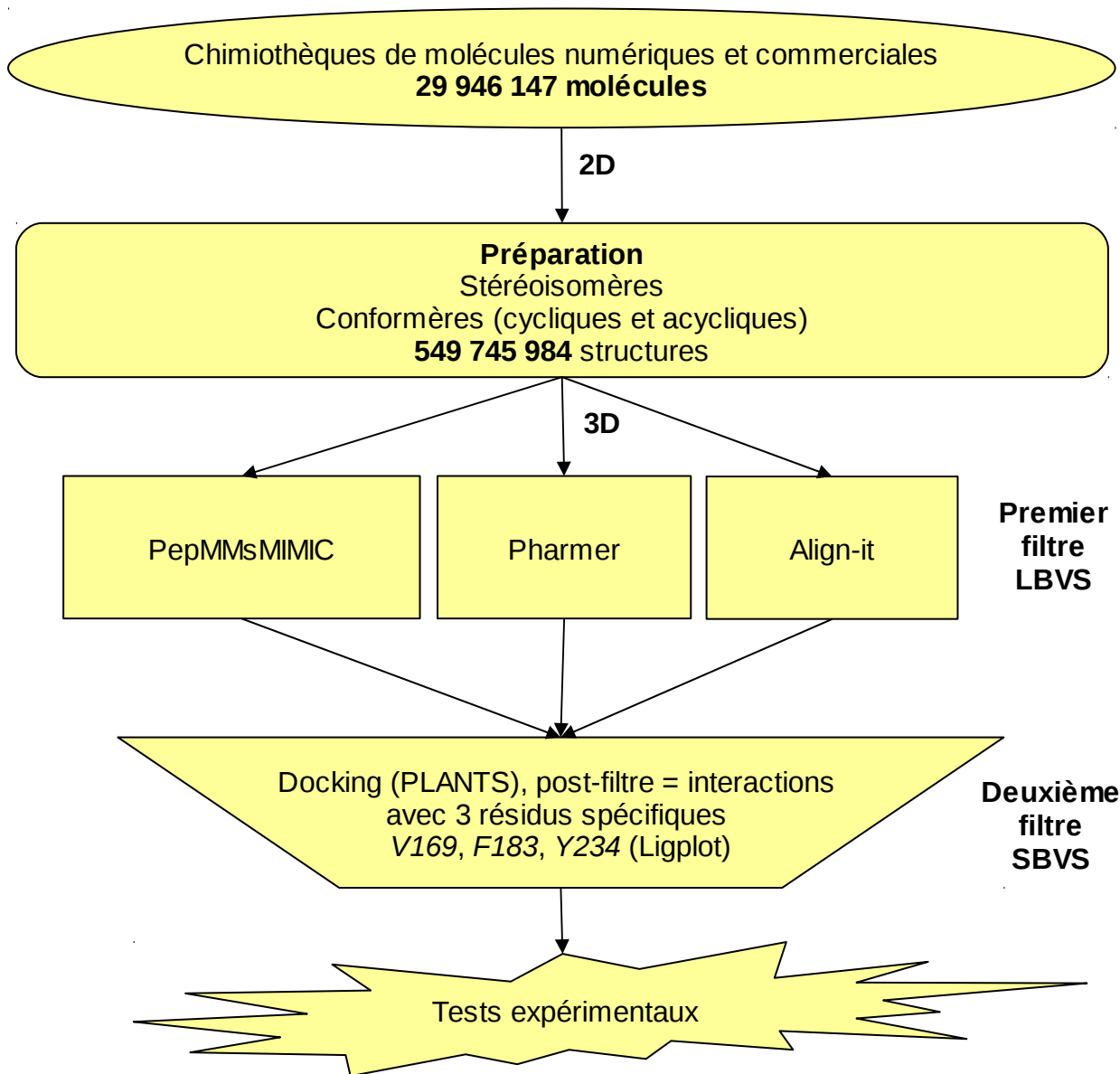


Figure 36 - Protocole de criblage virtuel développé pour déterminer des inhibiteurs de l'interaction PfAMA1-PfRON2.

2. Premier filtre, basé sur le ligand (LBVS)

Pour ce projet, un protocole basé sur le pharmacophore a été mis en place. Il utilise comme premier filtre trois outils qui se classent dans le domaine du LBVS : **PepMMsMIMIC**, **Pharmer** et **Align-It**. Toutes les bases n'ont pas été criblées avec les trois outils, entre autres parce que PepMMsMIMIC est un serveur en ligne criblant la base MMsINC et que Pharmer nécessite une préparation de chimiothèques qui peut prendre beaucoup d'espace mémoire.

1) Détermination du pharmacophore

Pour construire un pharmacophore, il y a certaines règles à suivre. Pour que l'utilisation de ce concept soit efficace, il faut utiliser entre 2 et 4 points. Il est aussi conseillé de restreindre l'espace en utilisant des volumes d'exclusion, qui vont permettre de rendre compte de la surface de la protéine.

A partir des données structurales et d'interaction, j'ai pu identifier quels résidus étaient essentiels. J'ai donc choisi parmi l'ensemble des résidus essentiels, 3 résidus qui n'étaient pas trop éloignés (*Pf*RON2sp1 : VAL2054, PHE2038 et PRO2033). Comme la TYR251 de *Pf*AMA1 se situe au centre de ces trois points, il était indispensable d'ajouter cette contrainte particulière. Il a donc été spécifié une sphère d'exclusion à l'endroit de la tyrosine (Figure 37).

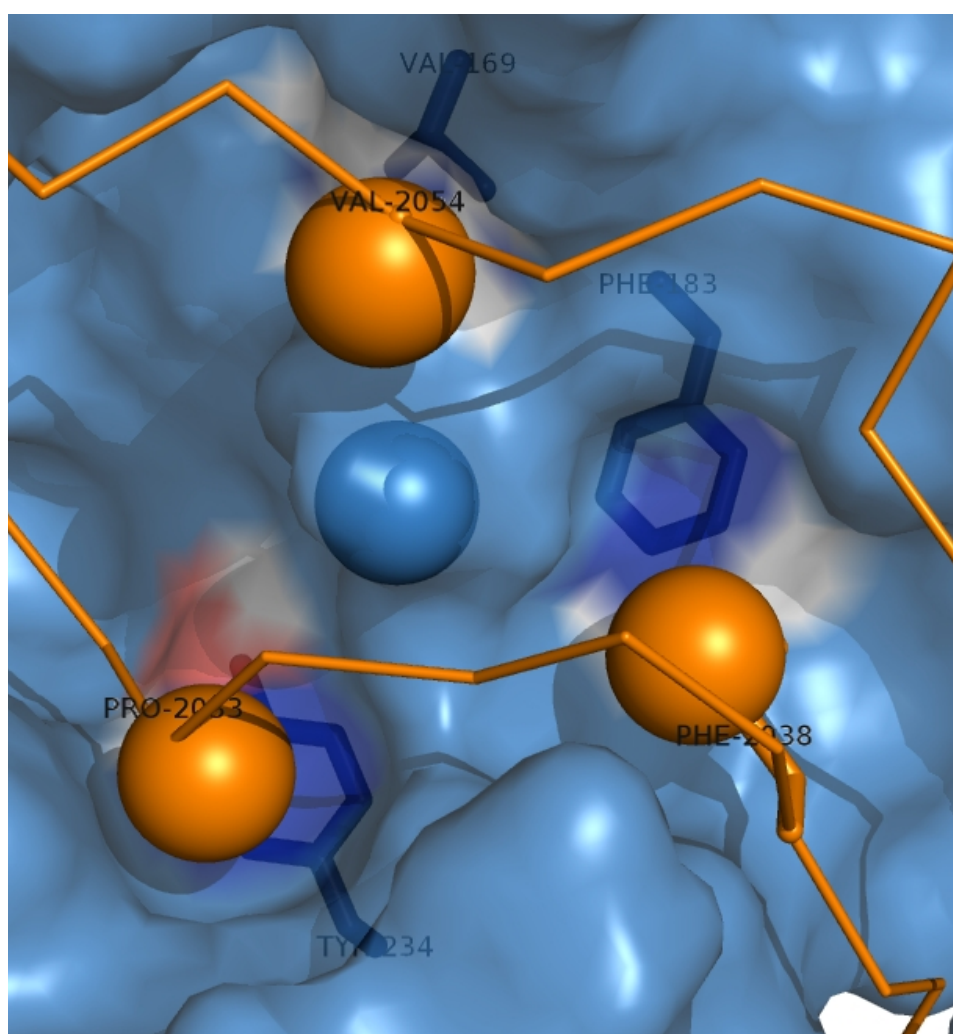


Figure 37 - Représentation du pharmacophore utilisé.

2) PepMMsMIMIC

PepMMsMIMIC est un serveur en ligne (<http://mms.dsfarm.unipd.it/pepMMsMIMIC/>), qui à partir du choix de 3 points minimum d'un peptide en trois dimensions, est capable de faire une recherche pharmacophorique tridimensionnelle. Seules deux chimiothèques ont été criblées par cet outil.

La première est MMsINC proposée par le serveur en ligne. Elle contient environ 17 million de conformères représentant 3,9 million de composés commerciaux. En résultat, les 200 meilleures molécules sont listées.

La deuxième est la collection Express-Pick de la Chembridge (CB), qui a pu être criblée localement après préparation préalable par le développeur, Matteo Floris de l'Université de Cagliari en Sardaigne (Italie). Ici, les 1000 meilleures molécules ont été retenues.

Le pharmacophore est déterminé par trois points qui sont les chaînes latérales de trois résidus clés : PRO2033 (sc : side chain) PHE2038 (sc) et VAL2054 (sc) de *Pf*RON2sp1. La recherche est une recherche hybride : 60% pharmacophorique (Pharmacophoric Fingerprint Similarity) et 40% basé sur la forme (Ultra Shape Recognition).

The screenshot displays the PepMMsMIMIC web interface. At the top, there are navigation links: [pepMMsMIMIC home](#), [Background](#), [Tutorial](#), [FAQ](#), and [MMsINC home](#). Below this is a "Wellcome to pep:MMs:MIMIC" section with a list of key information:

- **At a glance:** pepMMsMIMIC is a web-oriented tool that, given a peptide three-dimensional structure, is able to automate a multiconformers three-dimensional similarity search among 17 million of conformers calculated from 3.9 million of commercially available chemicals collected in the MMsINC database.
- **For a complete description of the workflow,** please read the [Background Section](#).
- **Who developed pep:MMs:MIMIC:** Matteo Floris and Stefano Moro.
- **How to cite:** Floris M., Masciocchi J., Fanton M., and Moro S., *Swimming into peptidomimetic chemical space using pepMMsMIMIC*, *Nucl. Acids Res.* (2011) first published online May 27, 2011, doi:10.1093/nar/gkr287 [[open access link](#)]

The main interface is divided into three steps:

- 1. Upload a valid PDB file:** A file upload area with a "Choisissez un fichier" button and an "Upload" button. Below it, a "Load" button for an example peptide is shown, with a link to "See DOI:10.2210/pdb/1y9p4b, PDB ID: 1YCP, chain B".
- 2. Please pick at least 3 residues from the Jmol window:** A dropdown menu shows "[VAL]2054:B". Below it, three residues are listed with their side chain (sc), backbone carbon (CO), and backbone hydrogen (HB) atoms selected: [PHE]2038:B, [PRO]2033:B, and [VAL]2054:B. Each row has a small "X" icon to the right.
- 3. Select the scoring method:** A dropdown menu is set to "hybrid search (60% pharmacoforic similarity, 40% shape)". Below it, there is a text input field for an optional e-mail address (containing "pihan@ipmc.cnrs.fr") and a "submit" button.

On the right side, a 3D molecular model is shown in a Jmol viewer. The protein backbone is yellow, and the side chains of the three selected residues are highlighted in pink. Labels for [PRO]2033:B, [PHE]2038:B, and [V]2054:B are visible. Below the model is a control panel with "Color by" (Secondary structure, chain, rainbow, aminoacid, hydrophobicity), "Surface" (On, Solvent Accessible, Solvent Excluded, Cerius), and "Settings" (Black Background, White Background, Windows on, Windows off) options, along with a "Reset Display" button.

Figure 38 - Capture d'écran de l'outil en ligne PepMMsMIMIC. Choix des résidus constituant le pharmacophore utilisé lors du criblage de la base de molécules MMsINC.

3) Pharmer

Pharmer est un outil de recherche par pharmacophore extrêmement rapide qui peut traiter des millions de structures chimiques en seulement quelques secondes. A la différence des autres logiciels développés, le temps d'exécution de Pharmer est proportionnelle à la complexité de la requête et non à la taille de la base de données que l'on crible.

Il existe un webserveur qui permet de cribler la ZINC : <http://zincpharmer.csb.pitt.edu/pharmer.html>. On peut aussi télécharger la version standalone : <http://sourceforge.net/projects/pharmer/>, ce qui m'a permis de cribler l'ensemble des chimiothèques que j'avais préparées. Le plus long dans ce logiciel est la création (option dbcreate) des chimiothèques de telle sorte qu'elles puissent être criblées (option dbsearch) à partir d'un fichier file.query préalablement préparé sur le webserveur et contenant les spécifications suivantes :

- TYR251 ExclusionSphere (3 Angströms, centre atome CZ, 28.968 21.977 43.600)
- PRO2033 Aromatique, Hydrophobe
- PHE2038 Aromatique, Hydrophobe
- VAL2054 Aromatique, Hydrophobe

The screenshot displays the ZINCPharmer web interface. On the left, a 3D molecular model shows a ligand (a benzene ring with a hydroxyl group) docked into a protein's pharmacophore, which is represented by a grey sphere and several colored regions (red, blue, white). Below the model is a table with columns for 'Pharmacophore Class', 'x', 'y', 'z', 'Radius', and 'Enabled'. The table lists various pharmacophore features and their coordinates and radii. On the right, a table titled 'Results' shows a list of ZINC IDs, RMSD values, Mass, and Filter counts. The interface includes navigation buttons like 'Add Feature', 'Load Features', 'Load Receptor', 'Clear', 'Load Session', and 'Save Session'.

Pharmacophore Class	x	y	z	Radius	Enabled
Aromatic	24.93	23.35	38.63	1.10	<input type="checkbox"/>
ExclusionSphere	28.97	21.98	43.60	3.00	<input checked="" type="checkbox"/>
Hydrophobic	24.93	23.35	38.63	1.00	<input checked="" type="checkbox"/>
Hydrophobic	30.84	17.90	44.30	1.00	<input checked="" type="checkbox"/>
Hydrophobic	24.98	26.84	47.74	1.00	<input checked="" type="checkbox"/>
Aromatic	24.98	26.84	47.74	1.00	<input checked="" type="checkbox"/>
Aromatic	30.84	17.90	44.30	1.00	<input checked="" type="checkbox"/>

Name	RMSD	Mass	Filters
ZINC66216052	0.578	365	9
ZINC66216057	0.611	365	9
ZINC66212220	0.664	374	9
ZINC05235217	0.783	388	10
ZINC30589329	0.452	390	10
ZINC30589320	0.432	390	10
ZINC30300048	0.779	383	7
ZINC05494889	0.785	398	10
ZINC04713593	0.783	388	8
ZINC02932863	0.609	393	10
ZINC66213014	0.648	390	10
ZINC02983390	0.738	382	5
ZINC02983390	0.517	382	5
ZINC09587694	0.655	396	7
ZINC66084084	0.523	389	10
ZINC09289734	0.566	390	11
ZINC09289734	0.592	390	11
ZINC66212722	0.668	389	10
ZINC66085185	0.576	389	10
ZINC66085185	0.604	389	10
ZINC02402725	0.558	414	11
ZINC18742071	0.647	426	6
ZINC59016325	0.482	393	10
ZINC01513221	0.688	398	6
ZINC00933553	0.782	376	6
ZINC66213018	0.633	390	10
ZINC0310485	0.836	396	6
ZINC0310485	0.840	396	6
ZINC18928911	0.654	386	6

Figure 39 - Capture d'écran de l'outil en ligne ZINCPharmer.

Il n'y a pas de seuil pour le score mais il est possible d'en mettre un pour le RMSD. Les molécules sont retenues seulement si elles répondent à tous les éléments requis avec une distance de tolérance. Dans notre cas, le pharmacophore contient 3 centres aromatiques hydrophobes et une sphère d'exclusion positionnés spécifiquement dans l'espace. Les caractéristiques des molécules doivent être dans ces sphères et ne pas passer par la sphère d'exclusion.

4) Align-it

Align-it est un outil de recherche par pharmacophore qui aligne les molécules selon leurs pharmacophores. Ceux-ci sont représentés par des volumes gaussiens en 3D. La nature permissive de ces fonctions continues a un effet bénéfique sur le problème d'optimisation introduit lors de l'alignement. En terme de vitesse de calcul, c'est le logiciel le plus couteux en temps parmi les trois. Cet outil n'est pas implémenté dans un webserveur. J'ai utilisé la version standalone.

Le pharmacophore est tel qu'il contient les trois résidus : PRO2033, PHE2038, VAL2054 et 577 sphères d'exclusion basées sur la structure de *PfAMA1*, générées grâce au plugin Pymol (create exclusion spheres). Cela a permis de générer le pharmacophore de référence : 3ptsExcl_Align-it.phar (Figure 40). L'option -cutOff 0.01 permet d'éliminer les molécules dont le score est 0. Le criblage se fait avec l'option -withExclusion true qui permet de prendre en compte la présence des sphères d'exclusion lors de la procédure d'optimisation.

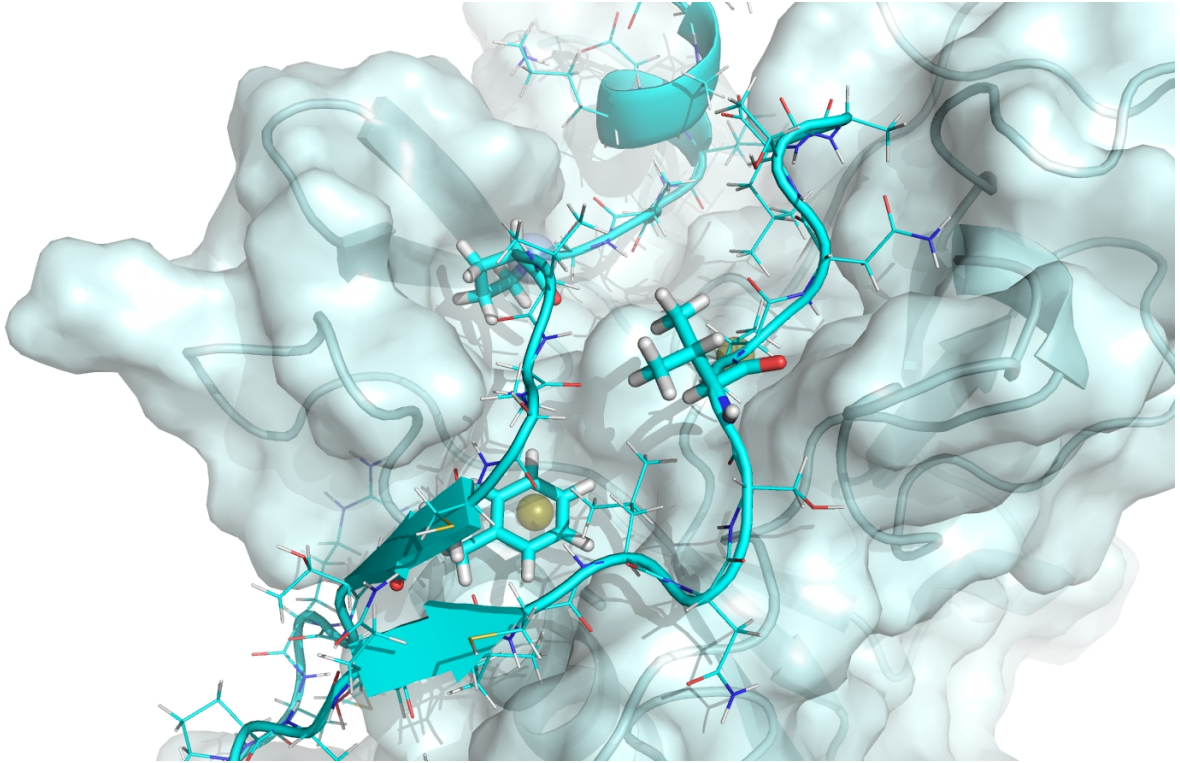
```
Generated_by_AlignIt_Pymol_Plugin
EXCL 38.6040000916 35.1360015869 50.6529998779 1.4 0 0.000 0.000 0.000
EXCL 33.51399993936.1850013733 50.4580001831 1.4 0 0.000 0.000 0.000
...
EXCL 29.8379993439 34.8829994202 58.4949989319 1.4 0 0.000 0.000 0.000
EXCL 29.5380001068 37.1279983521 58.1049995422 1.4 0 0.000 0.000 0.000
HYBL 25.1125 23.3021 38.7054 1.42857142857 0 0.000 0.000 0.000
HYBH 26.507 26.336 48.606 1.0 1 27.4039 26.2774 49.0443
HYBL 30.997 18.121 44.160 1.42857142857 0 0.000 0.000 0.000
$$$$
```

Figure 40 - Fichier 3PtsExcl_Align-it.phar contenant les contraintes géométriques du pharmacophore pour le programme Align-it.

Comme le calcul par Align-It est un peu long, j'ai pré-filtré les chimiothèques par le poids moléculaire (MW > 450 Daltons) et la solubilité (TPSA>80). J'ai choisi ces seuils en regardant les valeurs pour les molécules retenues par les autres outils de pharmacophore. Le nombre de liaisons libres doit aussi être inférieur ou égal à 10 pour la CN. Pour les chimiothèques CB et CL, la solubilité est évaluée par le logSw. Celui-ci doit être supérieur à -2.5 pour espérer que la molécule soit soluble dans l'eau.

Le score utilisé est le TVERSKY_REF, car il permet d'identifier des composés dont le pharmacophore contient au minimum le pharmacophore de référence. Le seuil a été fixé à 0,6.

A



B

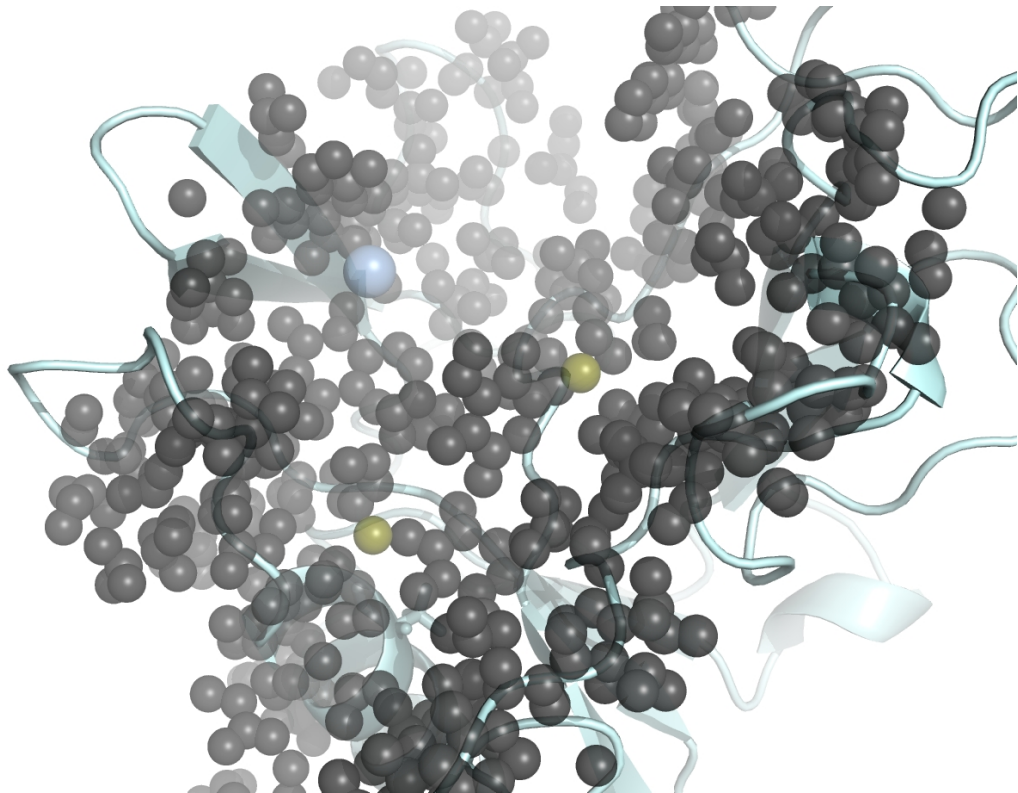


Figure 41 - Représentation A) du pharmacophore Align-it (sphères jaunes et bleues) et B) des sphères d'exclusion (sphères noires).

5) Nombre de molécules retenues par chaque outil de pharmacophore

Chimiothèque	Structures (Molécules)	PepMMsMIMIC	Pharmer	Align-it
DB	237 770 (6018)	-	49	4 805
CN	3 369 554 (40 463)	-	9223	96 265
CHEMBL	1 558 659 (26 674)	-	3047	32 829
CB	5 683 251 (481 196)	1000	599	50 263
DL	53 207 674 (1 967 476)	-	887	178 861
AMB	122 921 055 (4 895 514)	-	8276	2 145 861
CL	16 746 149 (326 863)	-	7463	8822
MB	6 797 (400)	-	55	946
MMsINC	170 000 000 (3 900 000)	200	-	-
ZINC	176 000 000 (18 300 000)	-	6628	-
e-Drug3D	15 075 (1 545)	-	63	910
Total	549 745 984 (29 946 147)	1 200	36 290	2 519 562

Tableau 4 - Nombre de molécules retenues dans chaque chimiothèque par les trois outils de pharmacophore.

3. Deuxième filtre, basé sur la structure de la *PfAMA1* (SBVS)

Le deuxième filtre suivant celui basé sur un pharmacophore appartient au SBVS. Il nécessite la structure de la protéine cible, *PfAMA1*. Pour cela, dans un premier temps, une minimisation et une dynamique moléculaire de 2ns avec Amber9 a été faite pour s'assurer que le complexe *PfAMA1-PfRON2sp1* était stable. La protéine *PfAMA1* est résolue des acides aminés 108 à 350 puis des acides aminés 388 à 440. Une liaison artificielle a donc été créé par tleap entre les résidus 350 et 388. La séquence du résidu 350 à 388 correspond à la DII loop non visible en cristallographie en raison de sa trop grande flexibilité. Le peptide *PfRON2sp* est résolu de 2023 à 2058, et contient donc 35 résidus. C'est une portion de la protéine *PfRON2* qui se lie à *PfAMA1*. La production de 2ns montre que le complexe est stable et a permis de préparer la protéine *PfAMA1* pour le docking.

Le docking se fait uniquement avec le logiciel PLANTS et le post-filtering est basé sur les interactions avec les résidus du sillon de *PfAMA1*. Pour être retenue, une molécule doit établir des interactions avec trois résidus spécifiques : **TYR234**, **PHE183** et **VAL169**. Suite à l'étape de docking et de post-filtre grâce à Ligplot, le nombre de molécules retenues suite à chaque outil de pharmacophore est listé dans le tableau 5. Cela permet de voir que cette étape est déterminante pour réduire considérablement le nombre de molécules à visualiser par la suite.

Chimiothèque	Structures (Molécules)	PepMMsMIMIC	Pharmer	Align-it
DB	237 770 (6018)	-	2	166
CN	3 369 554 (40 463)	-	1326	118
CHEMBL	1 558 659 (26 674)	-	190	1560
CB	5 683 251 (481 196)	8	24	336
DL	53 207 674 (1 967 476)	-	47	1591
AMB	122 921 055 (4 895 514)	-	574	237
CL	16 746 149 (326 863)	-	523	93
MB	6 797 (400)	-	9	11
MMsINC	170 000 000 (3 900 000)	1	-	-
ZINC	176 000 000 (18 300 000)	-	274	-
e-Drug3D	15 075 (1 545)	-	27	44
Total	549 745 984 (29 946 147)	9	2 996	4 156

Tableau 5 - Nombre de molécules retenues dans chaque chimiothèque suite au docking.

4. Troisième filtre visuel avec Pymol

Quand il y a trop de candidats à regarder, je supprime les molécules dont le nombre de liaisons libres est supérieur à 11, dont le poids moléculaire est inférieur à 450 Da et dont la TPSA est supérieure à 80, ou je recherche une sous-structure particulière.

Chimiothèque	Molécules	PepMMsMIMIC	Pharmer	Align-it
Retenues	6	0	4	2

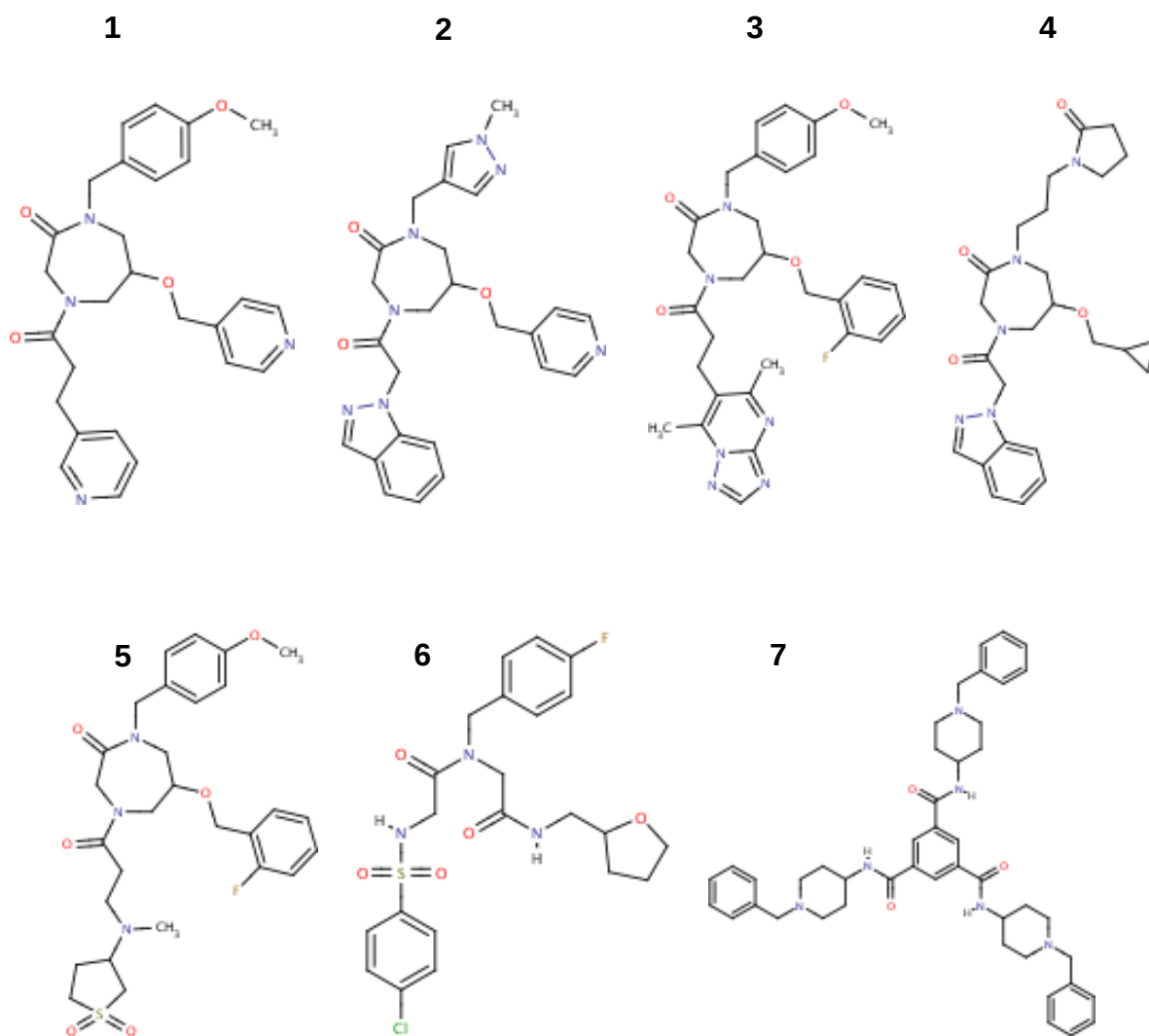
Tableau 6 - Nombre de molécules retenues à l'achat suite au criblage virtuel.

5. Liste des molécules achetées pour les tests *in vitro*

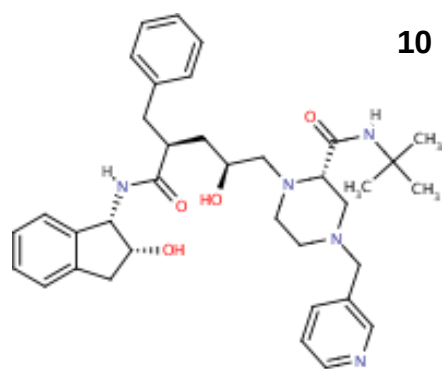
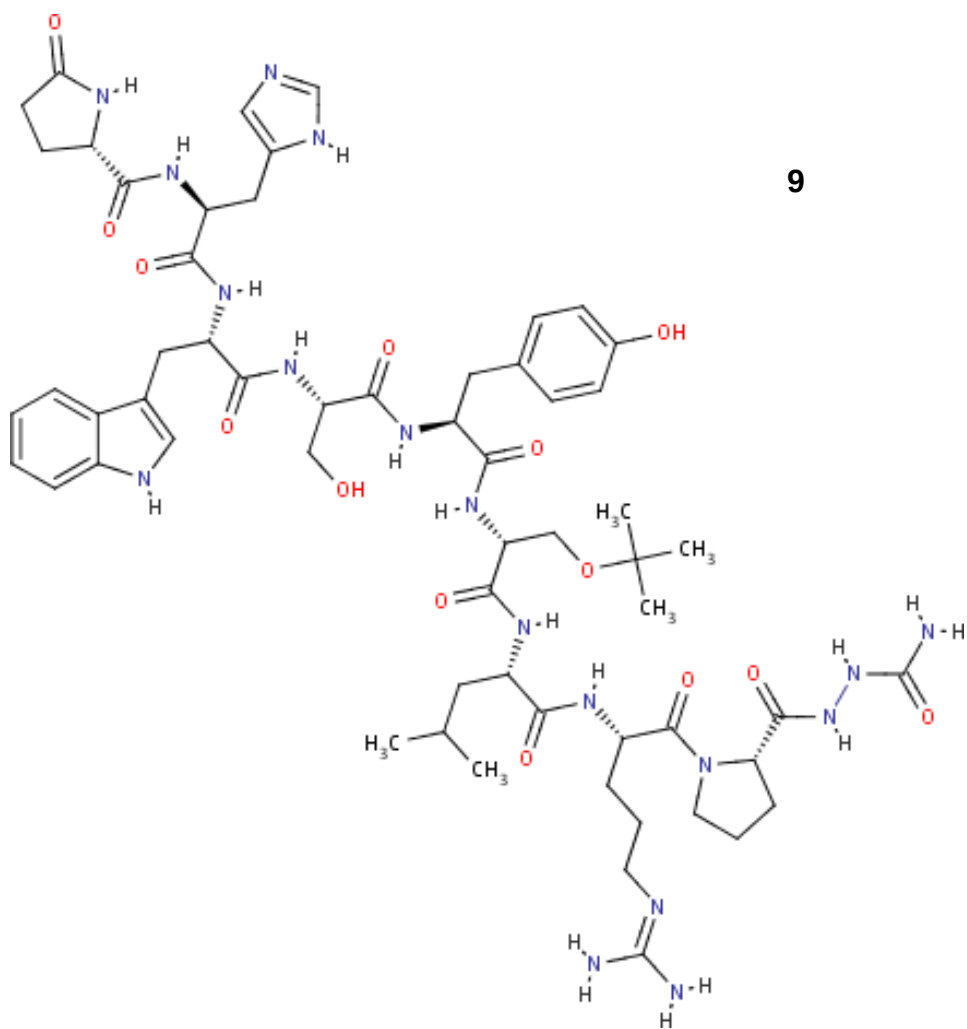
Le premier cycle d'achat de molécules contient sept molécules. Le deuxième cycle en contient deux. J'ai utilisé le programme PPI-HitProfiler et toutes les molécules du tableau 7 passent ce filtre. Elles possèdent donc bien les caractéristiques des inhibiteurs protéine-protéine (Sperandio et al., 2010). Les cinq premières molécules appartiennent à la même famille structurale. Cette famille est intéressante car les molécules sont embranchées, avec un cycle central assez flexible, ce qui permet aux substituants d'occuper les trois poches de PfAMA1.

Composé	ID	Provenance	Base de données	ChemSpiderID
1	19889551	Pharmer	CombiLib	23050414
2	81949477	Pharmer	ZINC	22412868
3	60810067	Pharmer	CombiLib	23059540
4	10144802	Align-It	CombiLib	22118875
5	66321848	Align-It	CombiLib	23065734
6	MolPort-000-075-803	Pharmer	ZINC	2458206
7	MolPort-001-534-338	Docking	MalariaBox	4229378
9	Goserelin	Pharmer	e-Drug3D	4470656
10	Indinavir	Align-It	e-Drug3D	4515036

Tableau 7 - Liste des molécules achetées.



Tour 1



Tour 2

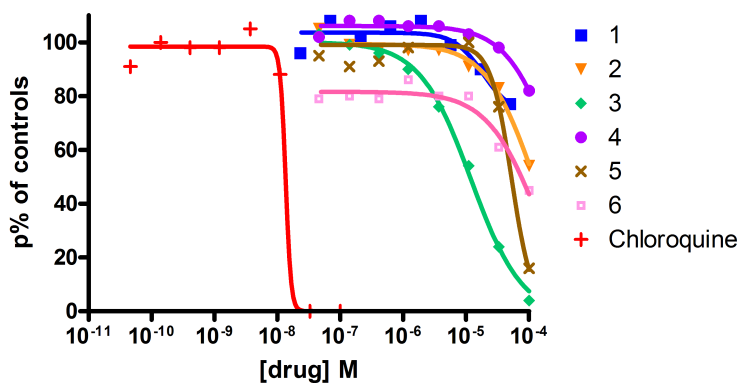
IV. Validation expérimentale dans le cadre d'une collaboration

Le groupe 'Invasion des cellules hôtes et Survie intracellulaire' du laboratoire DIMNP de Montpellier dirigée par Maryse Lebrun a développé des tests pour évaluer les composés. La protéine recombinante *PfAMA1* utilisée dans les différentes expériences a été produite à l'Université de Victoria au Canada par l'équipe de Martin Boulanger.

1. Test *in vitro* de croissance parasitaire

Marjorie Maynadier, du laboratoire DIMNP de Montpellier, a déterminé l'activité antipaludique *in vitro* (IC_{50}) contre *P.falciparum* (souche 3D7) des composés selon le microtest isotopique décrit par Desjardins (Desjardins et al., 1979). L' IC_{50} , **concentration inhibitrice médiane**, est une mesure de l'efficacité d'un composé. Cette mesure quantitative indique quelle quantité d'un composé est nécessaire pour inhiber à moitié un processus biologique donné.

Le composé n°1 est peu soluble à $2 \cdot 10^{-2} M$ en DMSO et le composé n°7 est insoluble à $4 \cdot 10^{-2} M$ en DMSO et n'a donc pas pu être testé. Les composés 3 et 5 montrent une activité et une IC_{50} a pu être définie. Ces composés appartiennent à la même famille structurale. Pour les autres composés, nous ne définissons pas d' IC_{50} car il n'y a pas d'inhibition complète aux concentrations testées (Figure 42).



Composé	IC_{50} (n=2)
3	12,5 μM
5	38,5 μM
T3	3,9 nM
CQ	12,5 nM

Figure 42 - Résultats du test *in vitro* de croissance parasitaire des composés 3 et 5 en comparaison avec le bithiazolinium (T3) et la chloroquine (CQ).

Expérience réalisée par Marjorie Maynadier du DIMNP.

(Les régressions non linéaires ont été effectuées par GraphPad Prism)

2. Test *in vitro* d'invasion parasitaire

Bien que les composés n'aient pas été choisis pour leur docking sur *Tg*AMA1, un test d'invasion développé au laboratoire DIMNP de Montpellier a été effectué sur *Toxoplasma gondii* par Maryse Lebrun tel qu'il est décrit dans (Lamarque et al., 2011). De manière préliminaire, on peut noter qu'une concentration de 100µM du composé 3 diminue l'invasion de 98% et une concentration de 100µM du composé 5 diminue l'invasion de 66%.

Nous avons positionné par docking le composé 3 dans la structure de *Tg*AMA1, qui est très proche de celle de *Pf*AMA1. La position de la Phe163 chez *Tg* comparée à celle de la Phe183 chez *Pf*, induit un mode de binding différent de celui observé avec *Pf*AMA1 (Figure 43) mais ce mode de binding est envisageable. Le composé 3 montre une bonne inhibition de l'invasion chez *Toxoplasma gondii* avec une IC_{50} d'environ 5µM (expérience de dose-réponse).

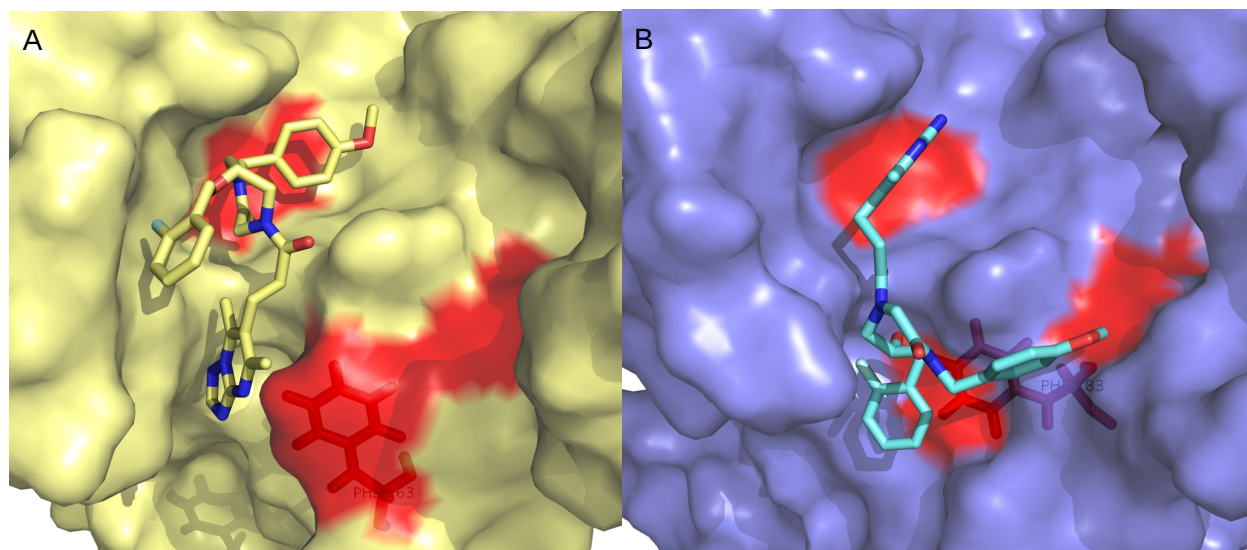


Figure 43 - Prédiction du mode de liaison du composé 3 par le logiciel de docking PLANTS dans A) le sillon hydrophobe de *Tg*AMA1 et B) dans celui de *Pf*AMA1. En rouge, position des résidus et équivalents chez *Tg* qui définissent le pharmacophore.

3. Tests d'affinité par SPR (Surface Plasmon Resonance).

Martine Pugnère, travaillant sur la plateforme "Protéomique Imagerie et Interactions Moléculaires" de Montpellier, a réalisé des expériences préliminaires par la technique du SPR. La résonance des plasmons de surface est un phénomène physique utilisé comme méthode de mesure de la liaison d'un ligand (dans le cas présent, les composés issus du criblage virtuel) sur un récepteur adsorbé à la surface d'une couche métallique (ici *PfAMA1*, protéine exprimée par l'équipe de Martin Boulanger, Université de Victoria). Un système de détection SPR mesure la variation de l'indice de réfraction au voisinage de l'interface quand le ligand se fixe aux récepteurs. La rifampicine, qui n'est pas sensée se lier à *PfAMA1*, est utilisée comme valeur de référence. Les composés 3 et 9 montrent une liaison faible à la protéine *PfAMA1* (Figure 44).

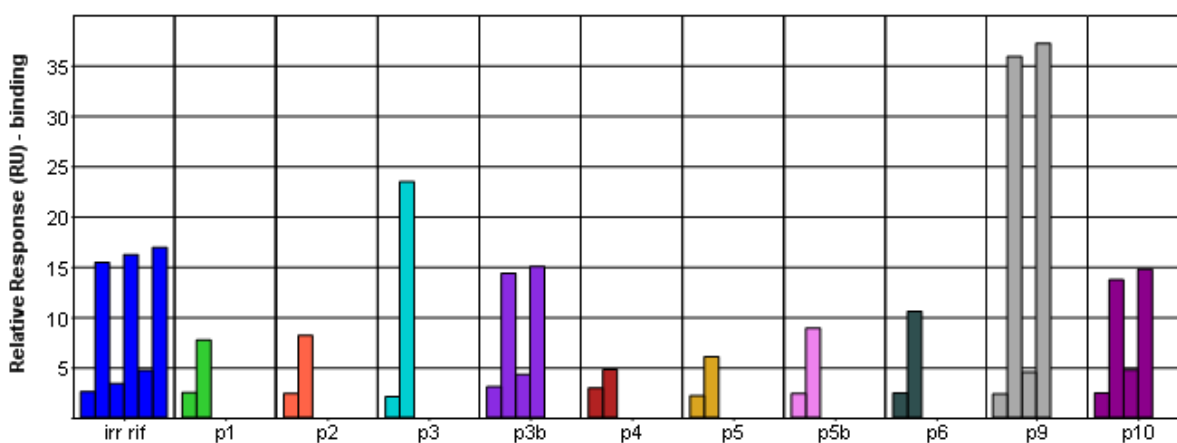


Figure 44 - Réponse relative des différents composés pour la protéine *PfAMA1*. Chaque mesure a été effectuée à 0 et 50 μ M de 1 à 3 fois. p3b et p5b correspondent à un second lot de composés 3 et 5. Expérience réalisée par Martine Pugnère.

En conclusion, les premiers résultats de SPR montrent que l'affinité de ces molécules pour la protéine *PfAMA1* est faible mais existante. La prochaine étape est de déterminer où se fixent les molécules 3 et 9 car la technique du SPR ne permet pas de savoir où se fixent exactement les composés. Au laboratoire, Roberto Delgadillo (post-doctorant) travaille à la mise au point d'un test de compétition utilisant la polarisation de fluorescence. Ce dernier devrait nous permettre de savoir si les composés se fixent dans le sillon hydrophobe de *PfAMA1*. Le test consiste à analyser des différences de signal en présence d'un inhibiteur. Ce test permettra aussi de cribler de petites chimiothèques réelles.

V. Conclusion sur le criblage du complexe *PfAMA1-PfRON2*

Plasmodium falciparum est un parasite intracellulaire obligatoire. L'invasion de la cellule hôte, l'érythrocyte, est un processus complexe et crucial pour la survie et la réplication du parasite. Il fait intervenir des organites dont ceux du complexe apical : les micronèmes et les rhoptries. Pendant l'invasion, le parasite forme une étroite jonction avec la cellule, la jonction mobile. Cette jonction met en jeu deux protéines clés, AMA1 (sécritée par les micronèmes) et RON2 (sécritée par les rhoptries).

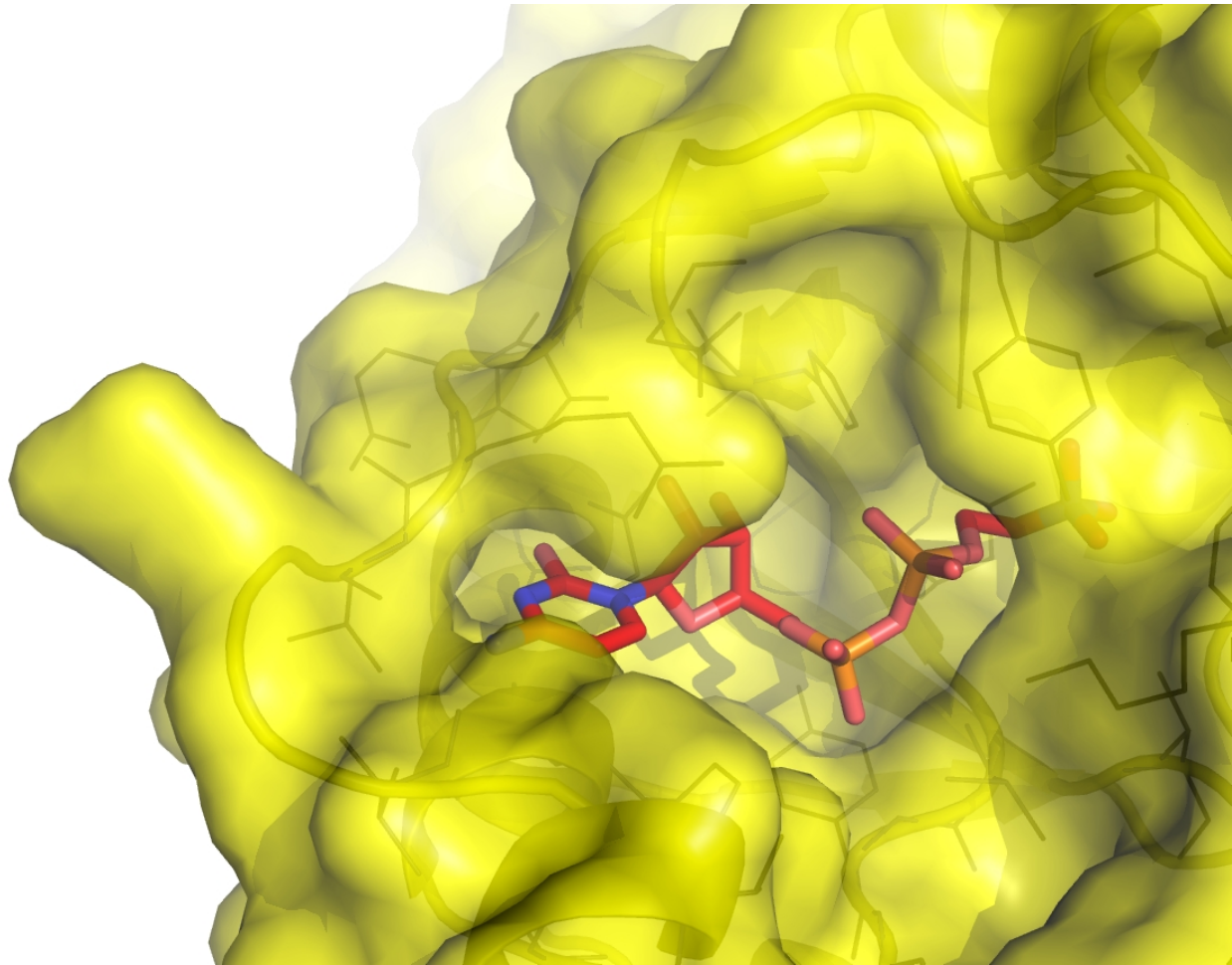
L'utilisation des données structurales disponibles nous a permis de dériver un pharmacophore de ce complexe protéine-protéine. J'ai mis au point un protocole de criblage virtuel qui intègre ce pharmacophore et la structure de la protéine *PfAMA1* lors de l'étape de docking afin de filtrer les différentes chimiothèques.

J'ai sélectionné neuf molécules, cinq appartiennent à la même famille chimique avec un scaffold 1,4 diazepam-2-one central identique (composés 1 à 5). Les quatre autres possèdent aussi une structure qui leur permettent d'atteindre les trois poches importantes au niveau du sillon hydrophobe de *PfAMA1*.

Les premiers résultats *in vitro* issus de nos collaborations identifient le composé 3 comme potentiellement ligand de *PfAMA1*. Le composé 5 de la même famille structurale, ne se différenciant que par un de ses trois substituants a aussi montré une inhibition de la croissance parasitaire. Il faut alors envisager de chercher d'autres analogues du composé 3.

Il faut maintenant s'assurer que la fixation de ce composé s'effectue effectivement au niveau du sillon hydrophobe, grâce au test de compétition utilisant la polarisation de fluorescence développée au laboratoire. Si tel est le cas, l'équipe de Martin Boulanger de l'Université de Victoria pourra tenter de cocrystalliser le composé avec sa cible, *PfAMA1*.

**Chapitre 3 : Inhibition de la synthèse
des phospholipides par le parasite
lors du cycle érythrocytaire en
ciblant la protéine
CTP:phosphocholine
CytidylylTransférase chez
Plasmodium falciparum (PfCCT)**



I. Problématique

L'importance de la biogenèse membranaire et de la voie *de novo* Kennedy lors de la phase intra-érythrocytaire de *Plasmodium falciparum* a été présentée dans l'introduction.

La deuxième partie de ma thèse concerne donc la recherche de composés inhibiteurs de l'enzyme *PfCCT*, deuxième enzyme de la voie *de novo* Kennedy. Ces composés pourront empêcher la multiplication des parasites dans les érythrocytes. En bloquant la synthèse de phosphatidylcholine, phospholipide majeur de la membrane parasitaire, ils permettront d'empêcher la biogenèse membranaire.

Ces travaux ont nécessité une bonne compréhension de la structure de cette enzyme, notamment grâce à la structure cristallographique disponible chez le rat. S'en est suivi un protocole de criblage virtuel en deux étapes : une étape basée sur un LBVS puis un docking dans le site actif du modèle de la *PfCCT*, construit par homologie. Ensuite, grâce à une collaboration avec le laboratoire DIMNP de Montpellier, nous avons pu tester les molécules sélectionnées sur l'enzyme recombinante et endogène ainsi que sur le parasite en culture.

Les tests biologiques actuels sur cette cible ne permettent pas de tester de nombreuses molécules (de l'ordre des dizaines). Nous sommes dans un contexte de criblage expérimental à bas débit. Nous devons donc sélectionner les composés à tester avec beaucoup d'attention.

II. Données sur la *PfCCT*

1. Présentation de l'enzyme *PfCCT*

L'interrogation de la base de données UniprotKB indique que cette choline-phosphate cytidyltransférase de *Plasmodium falciparum* (isolat 3D7) a pour numéro d'accèsion le Q8IEE9 (Q8IEE9_PLAF7). Cette protéine est codée par le gène *ctP*, au niveau du locus MAL13P1.86. La taxonomie à laquelle cette protéine appartient est la suivante : Eukaryota > Alveolata > Apicomplexa > Aconoidasida > Haemosporida > Plasmodium > Plasmodium (*Laverania*). Les annotations GO nous indiquent que cette protéine est impliquée dans un processus de biosynthèse (GO:0009058). En effet, elle fait partie de la voie métabolique de biosynthèse *de novo* des phospholipides (voie Kennedy). Elle est aussi responsable d'une activité catalytique (GO:0003824). Son identifiant de réaction catalytique est EC=2.7.7.15. La *PfCCT* catalyse la réaction (KEGG R01890) CTP + choline phosphate \rightleftharpoons diphosphate + CDP-choline (Figure 16).

La séquence de la *PfCCT* contient 896 acides aminés et contient 3 domaines dupliqués, dont le domaine catalytique. Les différents domaines sont deux domaines spécifiques des cytidyltransférases (Pfam : CTP_transf_2 PF01467 / InterPro : Cytidylyltransferase-like domain IPR004820), deux domaines de liaison à la membrane et deux domaines putatifs de phosphorylation.

```

MIIKVNSVQTQKSQDGNKVHNSQDDMDNEEKSIIRIYADGVYDLLHLGHMKQLEQAKHVDK 60
NVTLIVGVTGDNETRKFQIVQTLERTETLKHIRWVEIISPCPWITPEFMEEHKID 120
YVAHDDIPYANNQKKKKKKKSKGKSFDFEENEDIYAWLKRAGKFKATQRTEGVSTDLI 180
VRILKNYEDIERSLQRGIHPNELNIGVTKAQSIMKKNLIRWGEKVDELTKVTLTDKP 240
LGTDFDQGIDIIRDKVHDLFKLWRYHSHKLLKDFAKSFDPMFIIIRKRYRKDNL SAMYLS 300
DSNYFSSMKDELKKKKSLSTFNNIDEYYSADEEDTRDNSFYRVINKIANHTYYSEKE 360
NYETCFELEHSLRDKKKSLIFKDKNMKLSRDDLLHFTNPMNSKYDHNYYYYNNTYDKLSN 420
VTFNLNGIKKKSVDNESSYIKTAYHDIIDEDKNLLNKEKSFNKDEKSKKKSILETDQYKN 480
CEDQQFDNKYNHIIIGKKKSDSTYDNSLDEKSSNEKKSIDKQYSNMDSNMFHDCKTM 540
LSEHNESIENNDINGKQKEHIKKGENSENQDVPDTPNDAVPDDDDDDNSNDESEYES 600
SQMDSEKNKGSIKNSKNVVIYADGVYDMLHLGHMKQLEQAKKLFENTTLIVGVTSDNETK 660
LFGKQVVQTLERTETLKHIRWVEIISPCPWVITPEFLEKYKIDYVAHDDIPYANNQKK 720
KKKKKSKGKSFDFEENEDIYAWLKRAGKFKATQRTEGVSTDLIVRILKNYEDIERSL 780
QRGIHPNELNIGVTKAQSIMKKNLIRWGEKVDELTKVTLTDKPLGTDQGVENLQVK 840
FKELFKIWKNASNKLITDFTRKLEATSYLTSIQNIIDYEIENDDYASSNFDETSS 896

```

domaine catalytique (avec les sites conservés chez les cytidyltransférases et la boucle polylysine)
domaine de liaison à la membrane
domaine putatif de phosphorylation

Figure 45 - Séquence de la *PfCCT*, contenant 896 acides aminés.

Les domaines spécifiques des cytidyltransférases sont les domaines intéressants car ce sont les responsables de l'activité catalytique et contiennent donc chacun un site catalytique. Ces domaines adoptent un repliement particulier appelé Rossmann-fold (InterPro : Rossmann-like alpha/beta/alpha sandwich fold IPR014729). C'est un motif structural trouvé chez les protéines qui lient les nucléotides et ses dérivés. La structure contenant deux répétitions est composée de quatre feuillets bêta liés à deux paires d'hélices alpha dans l'ordre topologique beta-alpha-beta (x2). Le motif tient son nom de Michael Rossmann qui le premier a découvert ce motif récurrent chez les protéines liant les nucléotides. Nous allons rechercher des molécules occupant ces poches afin d'inhiber la réaction catalysée par la *PfCCT*. Cependant, la structure 3D de la *PfCCT* n'est pas encore résolue par cristallographie. C'est la raison pour laquelle j'ai construit un modèle par homologie.

2. Modélisation par homologie de la structure de la *PfCCT*

Concernant les données structurales disponibles sur la famille protéique des CCT, il existe des modèles RMN sur la partie membranaire isolée (PDB 1PEI et 1PEH) mais aussi une structure cristallographique de la CCT chez *Rattus norvegicus* (PDB 3HL4) (Lee et al., 2009). Il existe aussi deux structures cristallographiques des domaines catalytiques de la protéine CTP:glycérol-3-phosphate cytidyltransferase (GCT) de *Bacillus subtilis* (PDB 1COZ et 1N1D) et une structure cristallographique de la GCT de *Staphylococcus aureus* (PDB 2B7L). La structure cristallographique de l'ECT humaine (PDB 3ELB) a aussi été déterminée révélant un repliement du domaine catalytique en Rossmann-fold très similaire aux structures de CCT et GCT connues. Il n'y a aucune donnée concernant la (C/E)PT de *Plasmodium falciparum* qui est une protéine transmembranaire, catalysant la réaction suivant celle de la *PfCCT*. J'en ferai mention dans une autre partie de ce chapitre.

Quand une structure 3D d'un membre de la famille protéique est disponible, il est possible de s'en servir pour modéliser la structure des autres membres. Un bon alignement des séquences est important. La qualité de l'alignement dépend du taux d'identité mais pas seulement. Avec un taux d'identité supérieur à 40%, la modélisation est considérée facile. Entre 30 et 40%, il est souvent nécessaire d'optimiser l'alignement à la main. En dessous de 30%, la procédure de modélisation est plus complexe. De nombreuses méthodes existent et sont régulièrement évaluées (Runthala, 2012).

Dans le cas présent, pour modéliser la *PfCCT*, la structure de référence ('template') est la structure de la *rCCT* (PDB 3HL4). J'ai utilisé MUSCLE (**M**Ultiple **S**equences **C**omparison by **L**og-**E**xpectation) pour aligner dix séquences de familles protéiques très proches ce qui nous permet d'avoir un meilleur alignement. Trois familles peuvent être utilisées : la famille des CCT avec la séquence N-ter et C-ter de *Plasmodium falciparum*, la séquence de rat et celle humaine, la famille des ECT (*Plasmodium falciparum*, *Rattus norvegicus*, *Homo sapiens*) et la famille des GCT (2 séquences de *Bacillus subtilis* et une de *Staphylococcus aureus*). Les séquences au format fasta sont récupérées sur la base UniprotKB ou sur la PDB si une structure existe : *PfCCT_Nter* (UniProt Q8IEE9) , *PfCCT_Cter* (UniProt Q8IEE9), *rCCT* (UniProt P19836), *hCCT* (UniProt P49585), *rECT* (UniProt O88637), *hECT* (UniProt Q99447), *PfECT* (UniProt Q8IDM2), *bsGCT* (UniProt P27623), *saGCT* (UniProt O05155), *bsGCT* (UniProt Q8RKI6). Les identités de séquence sont de l'ordre de 13 à 33% sur le monomère entier (*PfCCT*-Nter ou *PfCCT*-Cter, Annexe 3) , mais l'identité de séquence est de 51% quand on aligne un domaine catalytique de la *PfCCT* à la partie de la *rCCT* résolue par cristallographie.

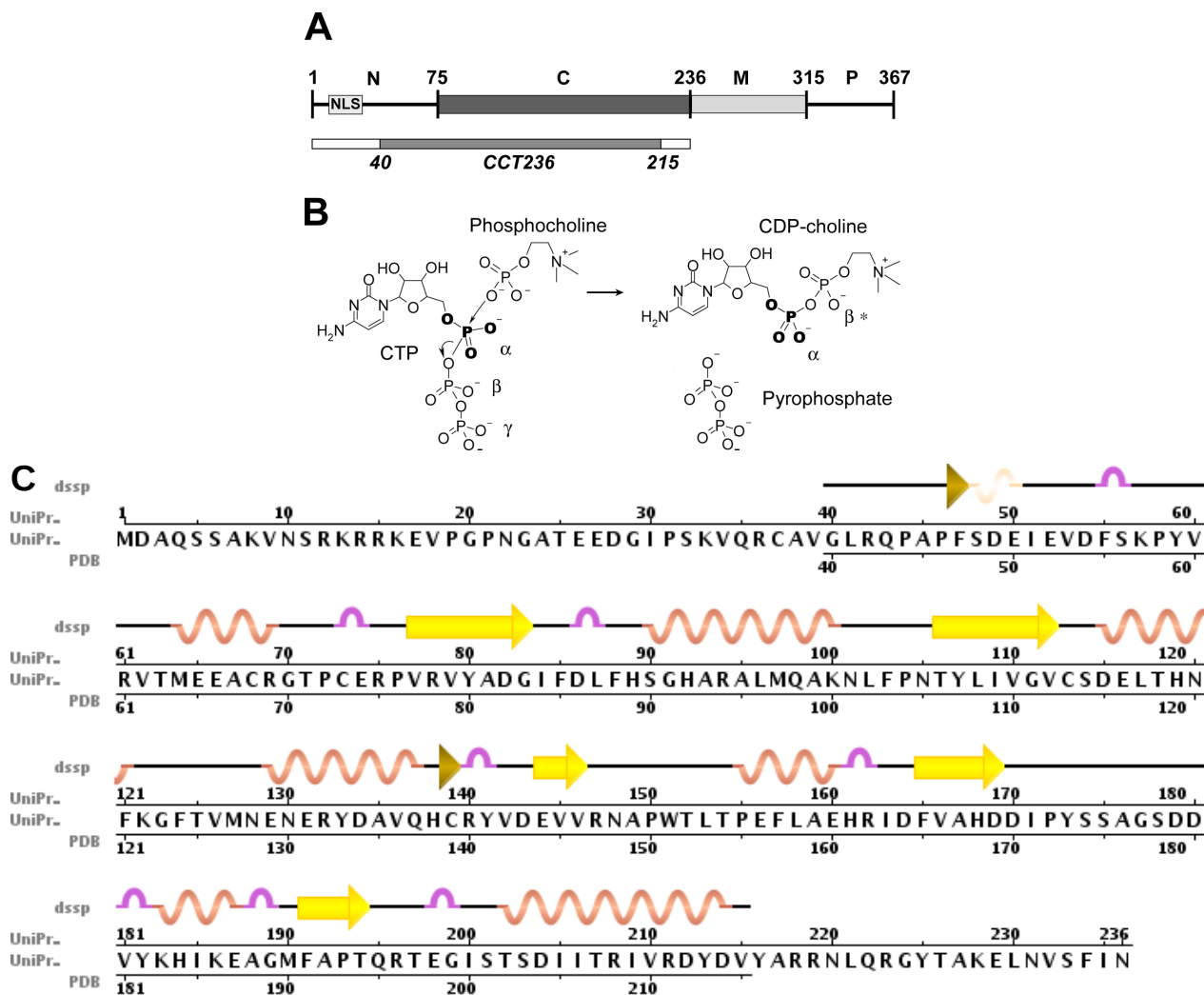


Figure 46 - Issue de (Lee et al., 2009). A) Organisation en domaines de la séquence de la CCT de *Rattus norvegicus*. B) Mécanisme catalytique de la rCCT. C) Structure secondaire de la structure résolue de la rCCT.

J'ai utilisé le programme Modeller 9v8 (Eswar et al., 2006) pour modéliser la PfCCT. Modeller est un logiciel qui construit des modèles par homologie de structures tridimensionnelles. Il est nécessaire d'établir un alignement préalable entre les séquences des protéines dont on veut un modèle et la séquence de la protéine dont on connaît déjà la structure. Modeller optimise alors automatiquement la construction de modèles.

Pour construire le modèle, j'ai utilisé le mode multi-chaines (dimère dans notre cas) et j'ai laissé les ligands CDP-choline cocrystalisés avec la CCT de *Rattus norvegicus*, protéine qui me sert de référence ou 'template'. J'ai supprimé dans la séquence de la PfCCT les boucles additionnelles polylysines n'existant pas dans la structure prise comme référence.

Model name	Modeller score	LGscore	MaxSub
<i>PfCCT</i> .B99990001.pdb	2221,00073	4,019	0,348
<i>PfCCT</i> .B99990002.pdb	2149,06934	3,582	0,342
<i>PfCCT</i> .B99990003.pdb	2063,43237	3,758	0,318
<i>PfCCT</i> .B99990004.pdb	2082,00952	3,966	0,346
<i>PfCCT</i> .B99990005.pdb	2085,56934	3,572	0,323
<i>PfCCT</i> .B99990006.pdb	1930,52283	3,696	0,331
<i>PfCCT</i> .B99990007.pdb	2087,33228	3,862	0,316
<i>PfCCT</i> .B99990008.pdb	2095,68750	3,958	0,350
<i>PfCCT</i> .B99990009.pdb	2075,49927	3,811	0,331
<i>PfCCT</i> .B99990010.pdb	2030,75525	3,720	0,337
Correct		> 1,5	> 0,1
Bon		> 3	> 0,5
Très bon		> 5	> 0,8

Tableau 8 - Valeurs des différents scores permettant d'évaluer les dix modèles de la *PfCCT*.

3. Analyse de l'interface du dimère

Les gènes de la CTP:phosphocholine cytidyltransférase chez *Plasmodium* sont annotés comme des copies uniques. De manière surprenante, les gènes de la CCT chez *Plasmodium* codent pour deux segments dupliqués, chacun contenant un domaine catalytique cytidyltransférase (C1 et C2), un domaine de liaison à la membrane (M1/M2) et un domaine de phosphorylation (P1/P2). Entre les deux segments, il y a un long linker partiellement conservé. Les autres CCT connues contiennent un seul domaine catalytique et aussi une copie du domaine de liaison à la membrane et du domaine de phosphorylation. Il a été reporté qu'elles formaient des dimères en solution (Taneva et al., 2008) et la structure cristallographique de la CCT de *Rattus norvegicus* présente une unité asymétrique composé d'un dimère. Ainsi, il est probable que les deux segments dupliqués de la protéine *PfCCT* s'associent pour former un « pseudo-dimère » intramoléculaire. D'autre part, la protéine recombinante constituée du domaine catalytique C-terminal seul sans la boucle polylysine (en marron dans la figure 45) du site catalytique (appelée *PfCCT* MΔK) est capable de s'homodimériser (Nagy et al., 2013).

La dimérisation (intra- ou intermoléculaire) semble donc être conservée au sein de la famille des cytidyltransférases. L'importance des interactions entre les deux domaines catalytiques nous a amené à tester l'hypothèse selon laquelle la perturbation de cette interface dimérique pourrait avoir une conséquence sur l'activité enzymatique.

1) Peptide antidimérisation

a) Modélisation informatique

A partir du modèle du dimère de la *PfCCT* provenant de Modeller, j'ai analysé l'interface de dimérisation afin d'imaginer un peptide capable d'empêcher la dimérisation. Pour cela, j'ai utilisé le plugin Intervor dans Pymol (Loriot & Cazals, 2010). Une interface entre deux partenaires est définie comme une collection de faces de Voronoï les séparant (en gris et jaune sur l'image). De manière équivalente, chaque face identifie deux atomes à l'interface. Pymol 1.3 me permet de visualiser cette interface et les résidus impliqués (Figure 48).

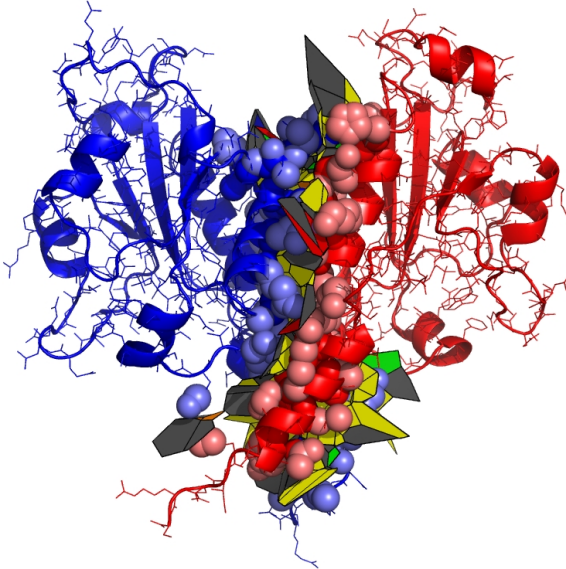


Figure 48 - Résultat d'Intervor sur le dimère *PfCCT*. (En bleu est représenté le monomère *PfCCT*-Cter et en rouge, le monomère *PfCCT*-Nter. Les sphères représentent les atomes se situant à l'interface des deux domaines catalytiques.)

A partir de la liste des résidus à l'interface, j'ai identifié les interactions hydrogènes entre les différents résidus présents à l'interface et je les ai listés ci-dessous.

Position dans <i>PfCCT</i> -Nter	<i>PfCCT</i> -Nter	<i>PfCCT</i> -Cter
Hélice A	K50	E675
Hélice C	Q80	Y772
	E87	R767
Hélice E	R182	E672
Boucle L3	H94	R681
	R96	D584
	R96	H679
	R96	I680
	R96	V683
	V98	R68

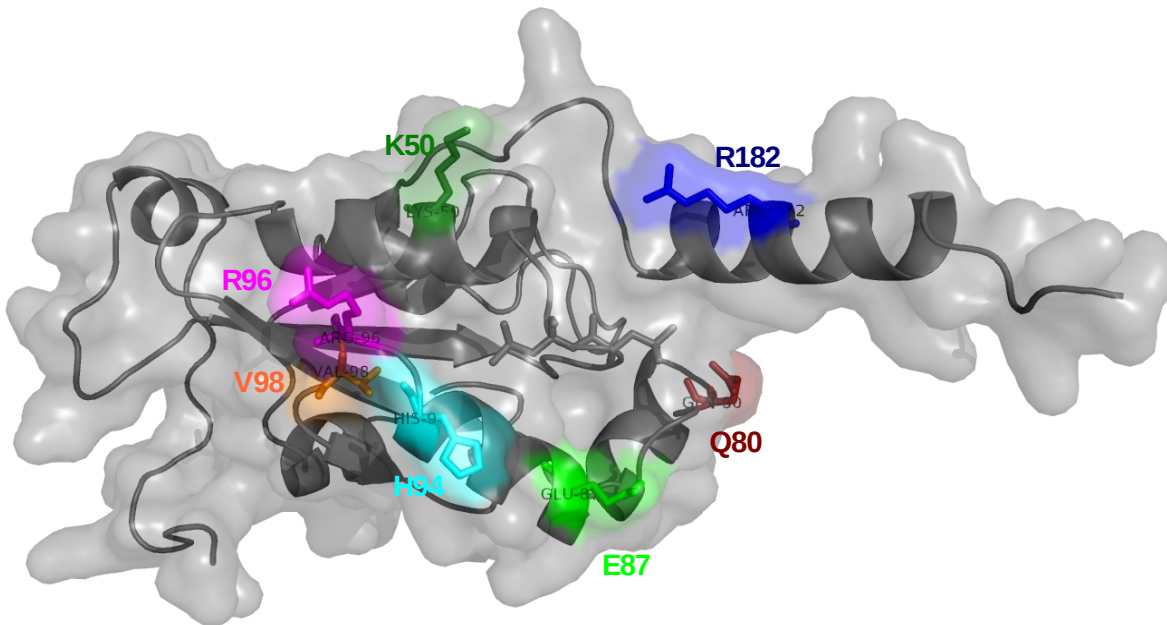


Figure 49 - Ensemble des contacts polaires entre le monomère N-terminal et le monomère C-terminal de *PfcCT*. Position des résidus impliqués dans les interactions hydrogènes au sein du monomère *PfcCT*-Nter.

Grâce à ces différents résultats, j'ai cherché une sous-séquence contenant ces résidus de l'interface et établissant des liaisons hydrogènes avec l'autre partie de l'interface. J'ai été capable de déterminer deux peptides de 20 acides aminés. Les peptides "anti-dimérisation" correspondent à l'hélice C dans la description donnée par (Lee et al., 2009) sur la structure de la CCT de *Rattus norvegicus*.

```

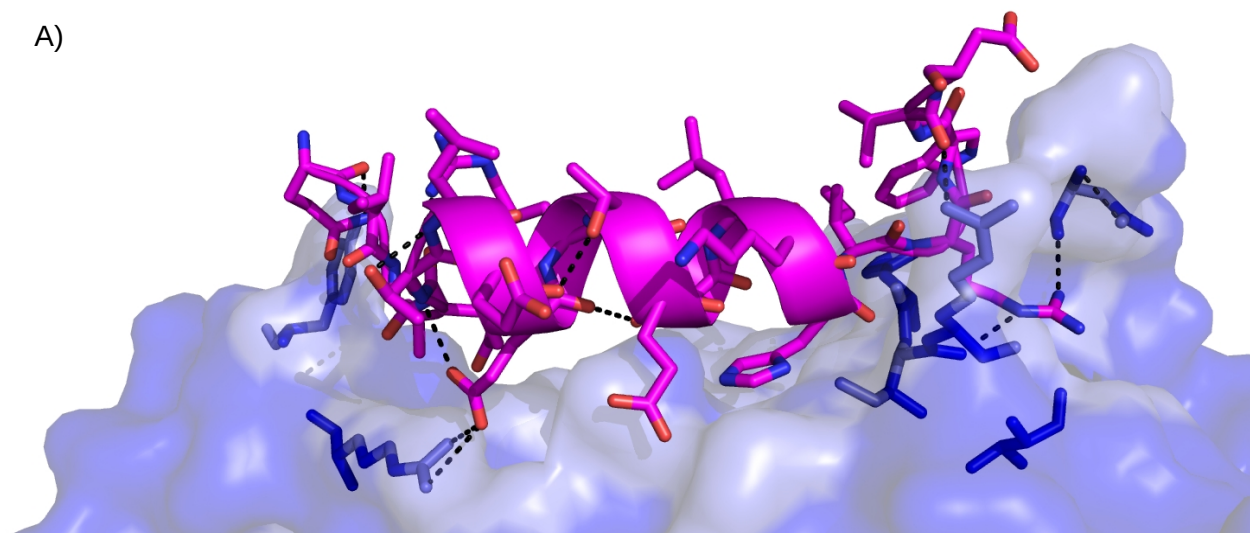
A) MIIKVNSVQTTQKSQDGNKVHNSQDDMDNEEKSIIRIYADGVYDLLHLGHMKQLEQAKHVDK 60
NVTLIVGVTGDNETRKFKGQIVQTLERTETLKHIRWVDEIISPCPWITPEFMEEHKID 120
YVAHDDIPYANNQKKKKKSKGKSFSFDEENEDIYAWLK RAGKFKATQRTEGVSTTDLI 180
VRILKNYEDYIERSLQRGIHPNELNIGVTKAQSIKMKKNLIRWGEKVTDELTKVTLTDKP 240
LGTDFDQGIDIIRDKVHDLFKLWRYHSKLLKDFAKSFDPMFIIIRKRYRKDNLSAM YLS 300
DSNYFSSMKDELKKKKSLSSTFNNIDEYYSADEEDTRDNSFYRVINKIANHTYYSEKE 360
NYETCFELEHSLRDKKKSLIFKDKNMKLSRDDLLHFTNPMNSKYDHNYYYNNTYDKLSN 420
VTFNLNGIKKKSVDNESSYIKTAYHDIIDEDKNLLNKEKSFNKDEKSKKKSILETDQYKN 480
CEDQQFDNKYNHIIIGKKS DSTYDNSLDKESSEKLLSIDKGQYSNMDSNYFHDCKTM 540
LSEHNESIESSNNDINGKQKEHIKKGENSENQDVPDTPDAVPDDDDDDNSNDESEYES 600
SQMDSEKNKGSIKNSKNVVIYADGVYDMLHLGHMKQLEQAKKLFENTTLIVGVTSDNETK 660
LFKQVVQTLERTETLKHIRWVDEIISPCPWVITPEFLEKYKIDYVAHDDIPYANNQKK 720
KKKKKSKGKSFSFDEENEDIYAWLK RAGKFKATQRTEGVSTTDLIVRILKNYEDYIERSL 780
Q RGIHPNELNIGVTKAQSIKMKKNLIRWGEKVTDELTKVTLTDKPLGTDFDQGVENLQVK 840
FKELFKIWKNASNKLITDFTRKLEATSYLTSIQNIIDYEIENDDYASSNFDDDETSS 896

```

B)

PfCCT_Nter (Pep1)	QIVQTL EERT TETL KHIRWVD (80-99)
PfCCT_Cter	QVVQTL EERT TETL KHIRWVD (665-684)

Figure 50 - Séquence de A) la PfCCT et B) des peptides proposés comme inhibiteurs potentiels de la dimérisation. En gras, les acides aminés communs au domaine Nter et Cter. En gras italique, les acides aminés qui sont à l'interface mais qui ne sont pas communs aux deux domaines. En couleurs, les contacts polaires entre les acides aminés à l'interface. En jaune, les acides aminés impliqués dans une hélice. En souligné, les peptides proposés.



B) PAT (Protein Analysis Toolkit)

UserSeq1.1d	QIVQTL EERT ETL KHIRWVD
consensus.2d	CHHHHHHHHH HHHHHHcccc
UserSeq1.2d.DSC	CCCCHHccc HHHHHHcccc
UserSeq1.2d.PSIPRED	CHHHHHHHHH HHHHHHcccc
UserSeq1.2d.SIMPA96	CHHHHHHHHH HHHHHHHHc
UserSeq1.ac.NETASA	EBBEBBEEEE EEBEEBBBE

Figure 51 - A) Le peptide "anti-dimérisation" N-ter en interaction avec le monomère PfCCT-Cter. B) Prédiction de la structure secondaire et de l'accessibilité au solvant du peptide "anti-dimérisation".

Le serveur en ligne Protein Analysis Toolkit (PAT) du CBS à Montpellier prédit un repliement en hélice de ce peptide (Gracy & Chiche, 2005). NETASA est un outil de prédiction d'accessibilité au solvant (B signifie enfoui (buried) et E exposé (exposed)) (Ahmad & Gromiha, 2002). De plus, certains serveurs de docking protéine-protéine (Zdock (Chen et al., 2003) et ClusPro (Comeau et al., 2003)) parviennent à replacer ce peptide dans sa position native vis-à-vis du monomère.

b) Tests expérimentaux

Le peptide Pep1 ainsi qu'un peptide scramble ont été testés. Un peptide scramble est un peptide qui a la même composition en acides aminés mais dans un ordre différent. La structuration du peptide sera alors différente. Ces expériences ont été réalisées par Alicia Contet, étudiante en thèse sous la direction de Rachel Cerdan et Henri Vial au laboratoire DIMNP de Montpellier.

i. Structure d'un peptide et dichroïsme circulaire

Une évaluation de la structuration du peptide Pep1 (séquence : QIVQTLERTETLKHIRWVD, 20aa, pI = 5,53, poids moléculaire = 2494 Da) par dichroïsme circulaire a été effectuée et le spectre de Pep1 se rapproche de celui obtenu pour un random-coil. Il n'y aurait donc pas de structuration de ce peptide en hélice dans l'eau.

ii. Inhibition de l'activité catalytique en présence de peptide

L'hypothèse est que la protéine a besoin d'être sous forme dimérique pour être active. La présence de Pep1 engendrerait une séparation du dimère et une inhibition de l'activité enzymatique. Pour vérifier cette hypothèse, une expérience d'inhibition de l'activité catalytique a été réalisée.

La PfcCT-Nter est inhibée totalement par 1mM de peptide. Cependant, la même expérience avec le peptide scramble (séquence : VREILDHTRQKTWTEQVELI, 20 aa, pI = 5,53 poids moléculaire = 2494 Da) a mené à la même inhibition. Cette expérience permet de conclure que l'inhibition n'est pas spécifique.

2) Mutations ponctuelles à l'interface

Alternativement, avec Alicia Contet du DIMNP, nous avons réfléchi aux types de mutations qui pourraient empêcher la dimérisation. Pour choisir les résidus de l'interface à muter, plusieurs stratégies ont été envisagées (les résidus potentiels sont ceux du monomère *PfCCT-Nter*) :

- muter des résidus générant des interactions hydrogènes ou ioniques (ex : E87, H94, R96, V98)
- muter des petits résidus (ex : M49 ou V98) en résidus plus gros pour créer une gêne stérique dans les zones où les deux monomères sont très proches
- muter des résidus impliqués dans des interactions hydrophobes enfouies en résidus polaires pour favoriser l'interaction avec le solvant (I183)

Les mutations qui ont été testées sont les suivantes, elles ont été réalisées simultanément : M49R / E87R / I183R (Figure 52). Après lyse et ultracentrifugation des bactéries exprimant la protéine, une bande correspondant au monomère est observée uniquement dans la fraction insoluble. Ces mutations induisent donc l'insolubilité de la protéine recombinante.

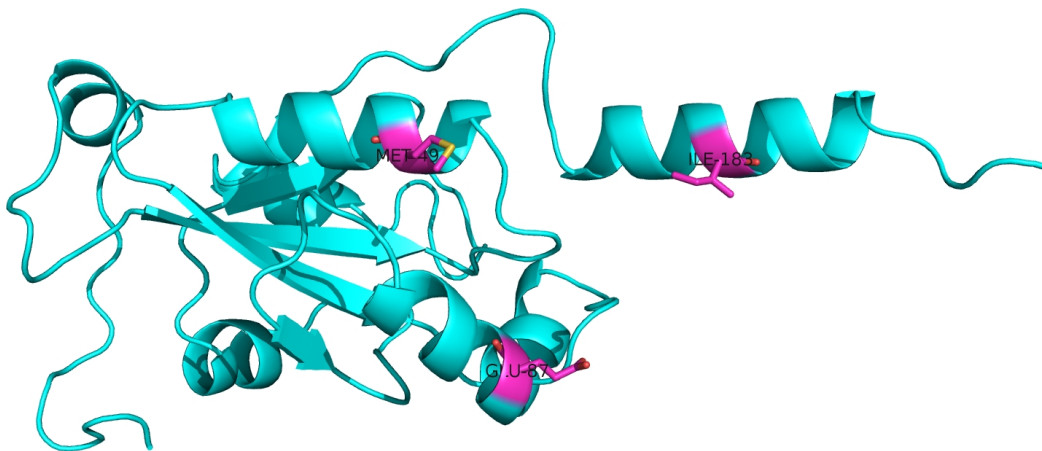


Figure 52 - Positions des résidus mutés en arginine.

Etant donné que les résidus mutés pointent vers l'extérieur de la protéine et ne sont pas enfouis, l'hypothèse est que ce monomère serait bien replié mais que l'interaction avec un autre monomère ne pouvant plus s'effectuer, une aggrégation des monomères se produit. De plus, il n'a jamais été observé de monomère sur gel natif pour la protéine non mutée. Nous pouvons ainsi nous demander si cette dimérisation n'est pas indispensable à la solubilité de cette protéine?

En conclusion, ces résultats préliminaires suggèrent l'importance de la dimérisation des domaines catalytiques de la *PfCCT* et la difficulté à perturber cette interface. Des études plus approfondies devront être mise en oeuvre afin d'interférer avec l'activité catalytique de l'enzyme, comme par exemple l'optimisation de la séquence du peptide antidimérisation.

4. Le site actif du domaine N-terminal de la *PfCCT*

1) Définition

La fonction des enzymes est liée à la présence dans leur structure d'un site particulier appelé le site actif qui a la forme d'une cavité ou d'un sillon. Les molécules sur lesquelles agit une enzyme sont définies comme les substrats de la réaction enzymatique. Chaque enzyme reconnaît spécifiquement une ou plusieurs molécules de substrat selon un principe de complémentarité, grâce à des sites de reconnaissance et de fixation situés à sa surface. Ce modèle statique a été affiné, par exemple en modèle dynamique, où l'enzyme n'est plus complémentaire de son substrat dans son état fondamental, mais dans son état actif : le substrat modifie légèrement la conformation de l'enzyme.

La fixation du substrat sur l'enzyme a pour conséquence la formation du complexe enzyme-substrat, ou complexe E/S, indispensable à la réaction enzymatique. Cette dernière fait intervenir les acides aminés d'un site catalytique. L'enzyme peut alors accélérer considérablement une des réactions biochimiques faisant intervenir ce substrat. À la fin de la réaction, l'enzyme est intacte et peut intervenir sur une autre molécule de substrat. Pendant la réaction, le substrat a pu être transformé en produit de la réaction.

Le site actif est constitué de deux parties :

- Le site de reconnaissance (ou site de liaison au substrat), permettant de fixer le substrat grâce à certains acides aminés.
- Le site catalytique (où a lieu la transformation du substrat), permettant de transformer le substrat grâce à des acides aminés qui interagissent avec le substrat.

2) Description du site actif du domaine N-terminal de la *PfCCT*

Pour lister et représenter les résidus constituant le site actif de la CCT de *Plasmodium falciparum*, j'ai étudié la structure de la CCT de *Rattus norvegicus*. Les alignements des séquences des CCT de plusieurs organismes mettent en évidence des motifs conservés dans la famille des cytidyltransférases : HXGH et RTEGVSTT (encadrés dans la figure 53 et surlignés en jaune dans l'annexe 3).

HXGH est la séquence spécifique permettant la liaison au CTP et la catalyse.

RTEGVSTT est la séquence signature spécifique de la superfamille des cytidyltransférases. Cette séquence permet la reconnaissance du motif cytidyl.

Ci-après sont donnés les détails des interactions entre la CDP-choline, produit de la réaction, et la protéine *PfCCT*-Nter. Trois des résidus des motifs conservés établissent des liaisons hydrogènes avec la CDP-choline (encadrés colorés dans la figure 53). Le tableau 9 permet une comparaison des acides aminés des différents sites actifs (*PfCCT*-Nter, *PfCCT*-Cter, *hCCT* et *rCCT*).

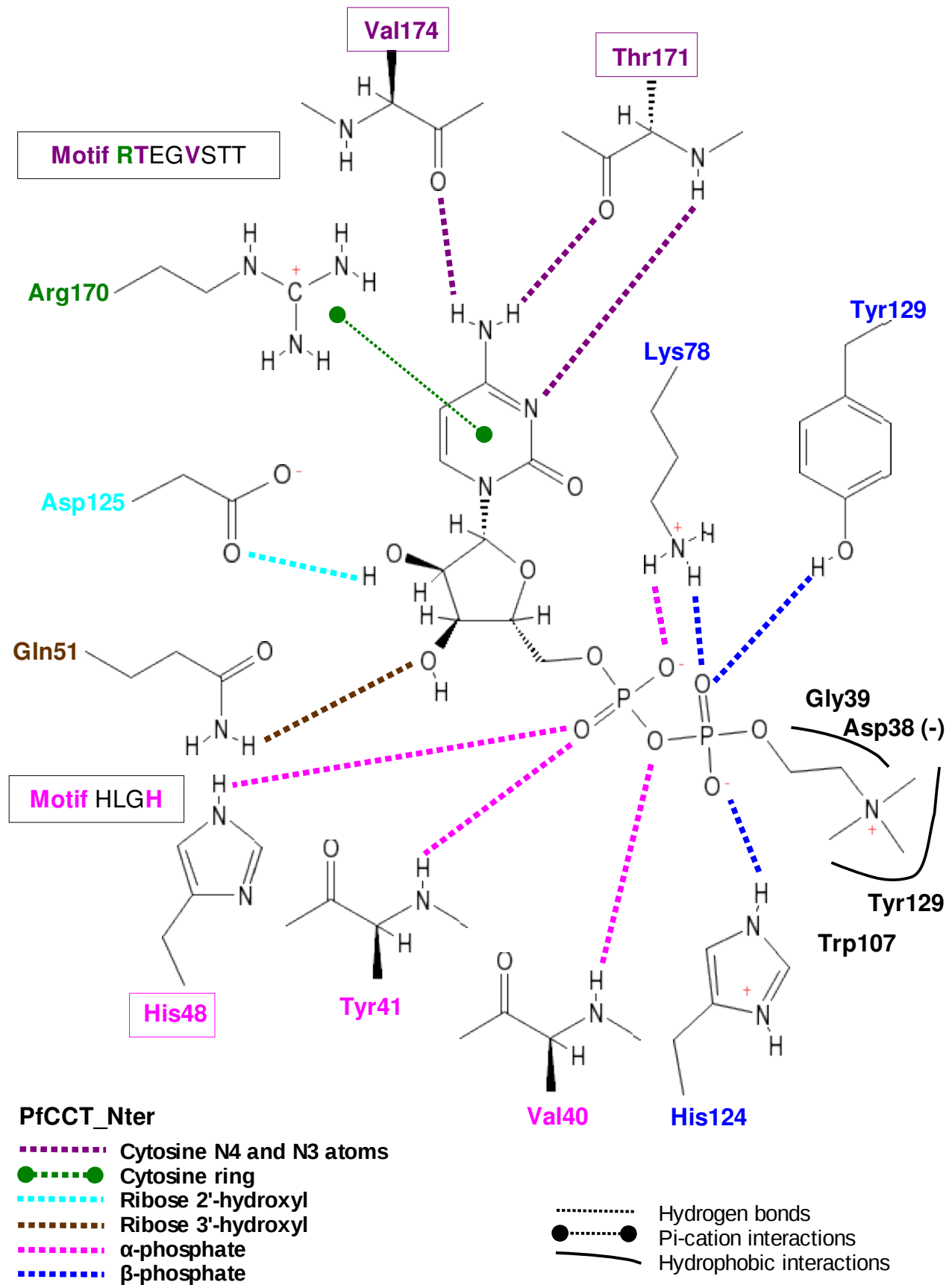


Figure 53 - Résidus du domaine N-terminal de la *PfCCT* constituant le site actif.

rCCT		hCCT	PfCCT Nter	PfCCT Cter	Atomes du ligand
CDP	D82	D82	D38	D623	-
choline	G83	G83	G39	G624	-
CDP	I84	I84	V40	V625	Premier groupe phosphate
CDP	F85	F85	Y41	Y626	Premier groupe phosphate
CDP	D86	D86	D42	D627	-
CDP	H89	H89	H45	H630	-
CDP	G91	G91	G47	G632	-
CDP	H92	H92	H48	H633	Premier groupe phosphate
-	A95	A95	Q51	Q636	SPECIFICITE
CDP	Q98	Q98	Q54	Q639	-
choline	C113	C113	T69	T638	-
CDP	K122	K122	K78	K663	Premier ou second groupe phosphate
choline	W151	W151	W107	W692	-
CDP	A167	A167	A123	A708	-
CDP	H168	H168	H124	H709	Premier ou second groupe phosphate
CDP	D169	D169	D125	D710	Premier ou second groupe phosphate
CDP	Y173	Y173	Y129	Y714	Second groupe phosphate
choline	V181	V181	I155	I740	-
choline	Y182	Y182	Y156	Y741	-
CDP	T194	T194	T168	T753	-
CDP	Q195	Q195	Q169	Q754	-
CDP	R196	R196	R170	R755	Cytosine
CDP	T197	T197	T171	T756	Azote et groupement NH2 de la cytosine
CDP	I200	I200	V174	V759	Groupement NH2 de la cytosine
CDP	T202	T202	T176	T761	-

Résidus des deux sites conservés dans la famille des cytidyltransférases (HXGH et R[T/S]XG[V/I]STT)

Résidus conservés entre les sites actifs de la CCT de rat, humain et de *Plasmodium falciparum*

Résidus établissant des liaisons hydrogènes avec le ligand

Résidus établissant des contacts hydrophobes avec le ligand

Résidus partageant des interactions Pi-Pi ou Pi-cation avec le ligand

Tableau 9 - Liste des résidus impliqués dans l'interaction avec la CDP-choline.

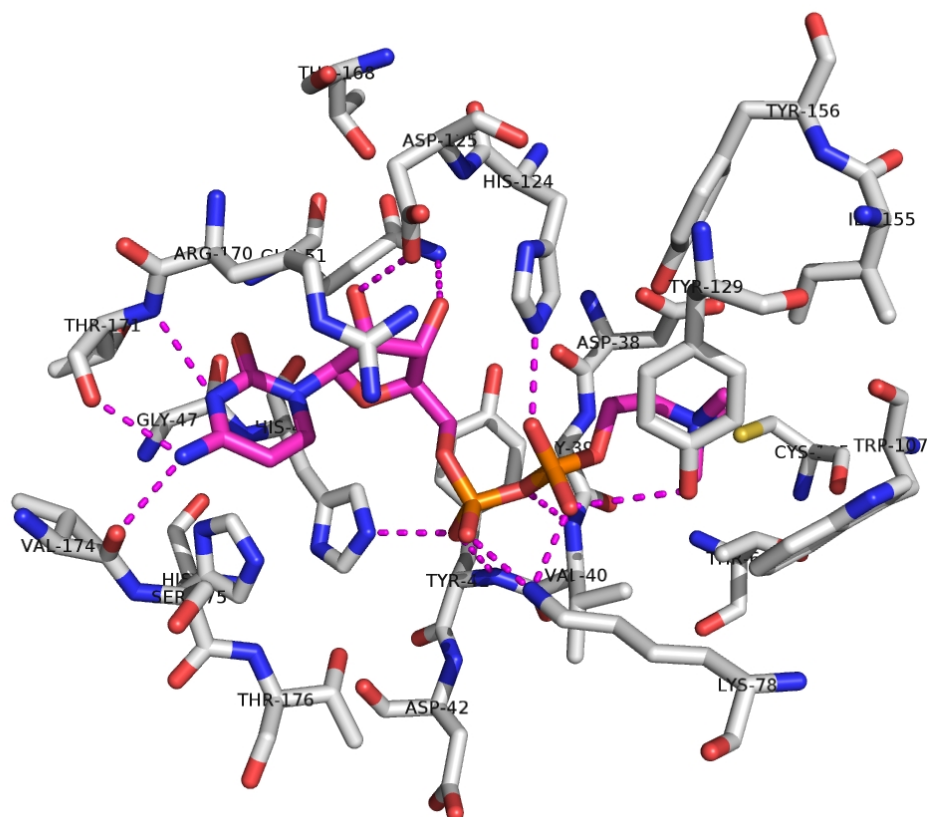


Figure 54 - Réseau des douze interactions hydrogènes de la CDP-choline établies par dix résidus des vingt-quatre constituant le site actif de la *PfCCT*-Nter (modèle avec position de la glutamine 51 optimisée).

Les sites actifs du domaine N-terminal et C-terminal de la *PfCCT* sont identiques.

Lors de la modélisation par homologie de la *PfCCT*, nous pouvons nous poser la question de la spécificité qu'auront les molécules sélectionnées pour *Plasmodium falciparum*. En effet, il existe une protéine homologue de la CCT chez *Homo sapiens*. La CTP:phosphocholine cytidyltransférase A est présente dans tous les tissus chez l'Homme, et pourrait donc être responsable d'effets secondaires. Lorsque nous alignons les deux séquences, nous observons une seule différence notable au niveau du site actif (ligne bleue du tableau 9) : A95 (chez *Homo sapiens*) ↔ Q51 (chez *Plasmodium falciparum*). C'est un point très important qu'il faudra garder à l'esprit lors de l'optimisation des molécules. Il faudra peut-être mettre à profit cette seule différence majeure du site actif pour essayer de développer une molécule spécifique de la *PfCCT*, qui n'a pas, ou peu, d'effet sur la *hCCT*.

Quand nous comparons le site actif de la CCT humaine et celui de la CCT de *Rattus norvegicus*, nous observons qu'ils sont identiques. Lors des tests *in vivo* sur le rat, il sera alors possible d'évaluer les effets secondaires des molécules retenues car la *rCCT* sera présente.

III. Criblage virtuel

1. Schéma général

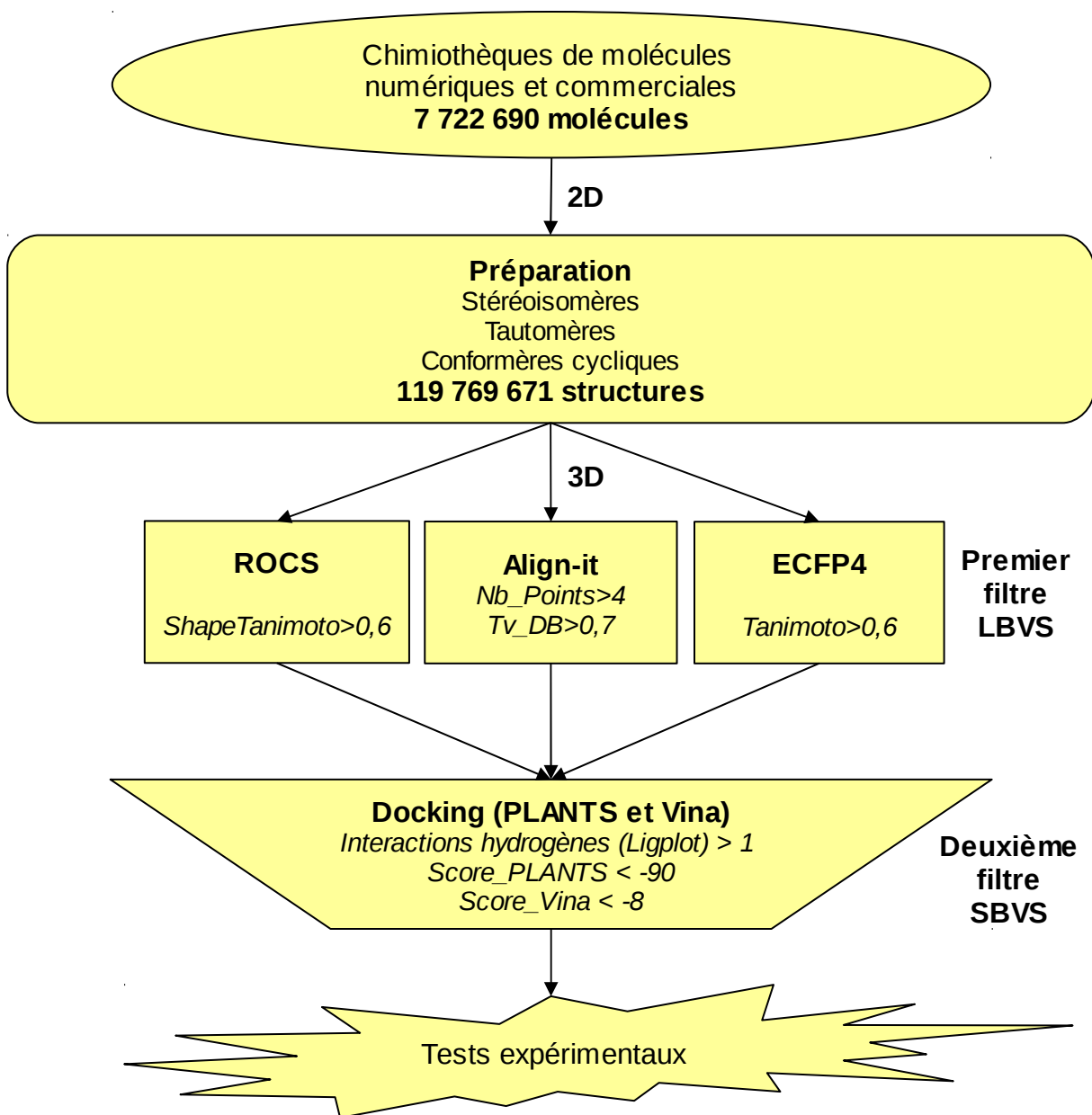
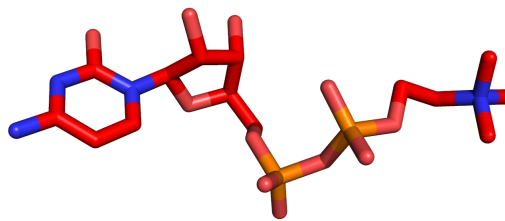


Figure 55 - Protocole de criblage virtuel développé pour déterminer des inhibiteurs de l'enzyme PfCC2.

2. LBVS

Dans ce premier filtre, trois logiciels se basant sur des concepts distincts ont été utilisés. Leur point commun est l'utilisation comme référence d'une molécule, la CDP-choline, dont la conformation est celle observée par cristallographie dans le site actif de la rCCT.



1) Criblage par similarité de forme

Le premier outil, **ROCS**, cherche des similarités de forme. ROCS est implémenté avec deux champs de force distincts : Implicit Mills Dean par défaut et Explicit Mills Dean. Le champ de force chimie-dépendant peut-être modifié avec l'option "-chemff ExplicitMillsDean".

Le champ de force Implicit Mills Dean inclut un modèle simple adoptant un pH=7. Les groupes fonctionnels se voient assigner une valeur prédéfinie, indépendamment de l'état de protonation de la molécule dans la chimiothèque. Par exemple, si une molécule contient un carboxylate, le champ de force Implicit Mills Dean va considérer un centre anionique, que ce groupe soit protoné ou déprotoné dans la chimiothèque.

Le champ de force Explicit Mills Dean quant à lui, interprète la protonation et l'état de charge de chaque molécule exactement comme elle est dans la chimiothèque. Ainsi, si un sulfate est protoné et neutre, il ne sera pas considéré comme un anion. Comme j'ai préparé les diverses chimiothèques à pH=7,4 et surtout que j'ai énuméré divers tautomères pour chaque molécule, qui diffèrent donc par leur protonation, j'ai choisi d'utiliser le champ de force Explicit Mills Dean.

Le champ de force chimie-dépendant est utilisé pour mesurer la similarité chimique entre la requête et la molécule de référence (score ColorTanimoto) et aussi pour optimiser les superpositions de forme. Le champ de force chimie-dépendant décrit les types d'atomes environnement-dépendant, les liens entre groupements fonctionnels et types d'atomes (à noter que seuls les atomes lourds sont pris en compte, les hydrogènes sont ignorés), le type d'interaction entre deux atomes (attractif ou répulsif). Les interactions entre deux atomes de même type sont toujours attractives. Le terme de poids décrit la force de l'interaction, relative aux gradients de forme. La gamme de poids affecte la portée de l'interaction.

- Donneur : Groupes fonctionnels qui peuvent agir comme des donneurs de liaisons hydrogènes, comme le groupement hydroxyle -OH
- Accepteur : Groupes fonctionnels qui peuvent agir comme des accepteurs de liaisons hydrogènes, comme le groupement carboxylate
- Anion : Groupes fonctionnels avec une charge négative localisée ou délocalisée
- Cation : Groupes fonctionnels avec une charge positive localisée ou délocalisée
- Hydrophobe : Groupes aromatiques ou aliphatiques
- Cycles : Cycles de taille définie, c'est-à-dire de 4 à 7 atomes

Tableau 10 - Groupes fonctionnels décrits dans les champs de force ROCS.

La valeur de score utilisée pour le filtre est uniquement le score de forme (ShapeTanimoto), et non pas le score chimie-dépendant (ColorTanimoto) même si l'alignement en tient compte. En effet, je voulais utiliser cet outil uniquement pour la superposition de forme afin de trouver de nouveaux chémotypes (différents des groupements choline et cytidyl). Sinon, la définition par coloration physicochimique rejoindrait le concept de pharmacophore. Pour ce dernier, j'ai préféré utiliser un autre logiciel, Align-it. Le seuil de score ShapeTanimoto choisi pour ce logiciel est de 0,6 et aucun nombre maximum de solutions n'a été spécifié.

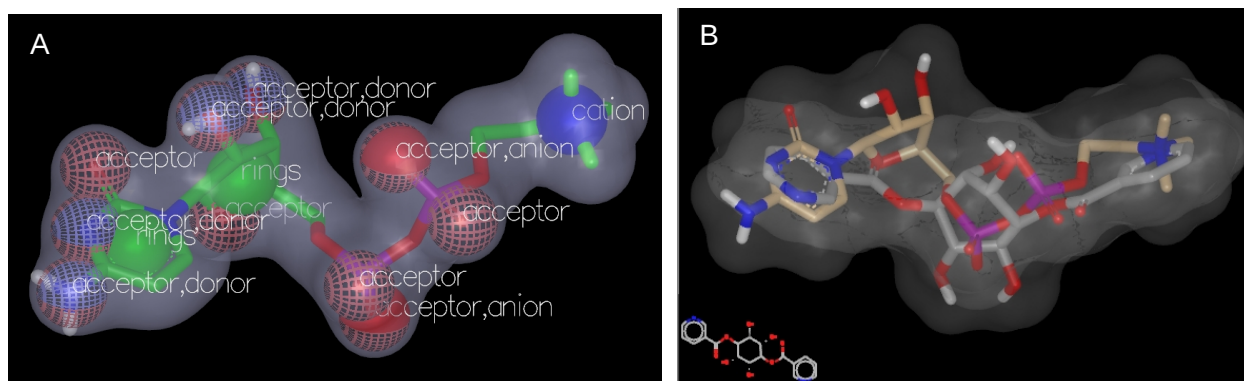


Figure 56 - A) Représentation de la forme et de la coloration chimique de la molécule de référence. B) Superposition de la CDP-choline avec un candidat en gris (ShapeTanimoto=0,6120).

2) Criblage par recherche de pharmacophore

Le second outil est **Align-it™** de Silicos. Il se base sur le concept du pharmacophore. La CDP-choline contient 16 points dans son pharmacophore. Les molécules des bases de données passent ce filtre si elles contiennent au moins 4 points de pharmacophore en commun avec le pharmacophore de référence et si leur score Tversky_DB est supérieur à 0,7. Le score Tversky_DB permet d'identifier des composés dont le pharmacophore est un sous-ensemble du pharmacophore de référence.

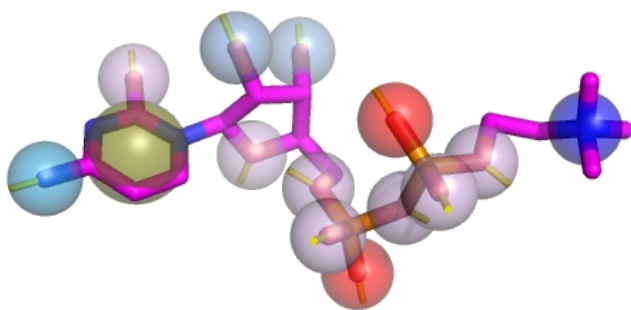


Figure 57 - Pharmacophore de référence, contenant 16 points et qui a été établi par **Align-it™** à partir de la molécule de référence, la CDP-choline. Les sphères rouges représentent une charge négative et masquent deux sphères roses représentant deux accepteurs de liaison hydrogène.

3) Criblage par similarité de structures

Le troisième et dernier outil, **ECFP4** (4 correspond au diamètre utilisé), utilise des empreintes circulaires spécifiques de familles structurales. Ces empreintes constituent la signature moléculaire. C'est la somme des signatures de tous les atomes d'une molécule.

Le seuil choisi pour le score Tanimoto est 0,6. Cet outil permet surtout de chercher des sous-structures de la molécule de référence, qui est la CDP-choline.

4) Nombre de molécules retenues par chaque outil de LBVS

Chimiothèque	Structures (Molécules)	ROCS	Align-it	ECFP4
DB	254 825 (5 935)	383 (84)	517 (218)	4 468 (130)
CN	2 493 966 (40 243)	499 (123)	5 667 (1 117)	620 (8)
CHEMBL	1 167 303 (26 449)	1 119 (611)	1 201 (292)	192 (4)
CB	6 060 089 (481 421)	16 052 (11 474)	12 422 (3 243)	0 (0)
DL	32 316 253 (1 967 820)	107 939 (74 711)	42 497 (10 576)	0 (0)
AMB	77 313 514 (4 899 277)	213 670 (13 675)	77 914 (861)	16 259 (21)
e-Drug3D	163 721 (1 545)	17 (15)	399(156)	7(1)
Total	119 769 671 (7 422 690)	339 679 (100 693)	140 617 (16 463)	21 546 (164)

Tableau 11 - Nombre de molécules retenues suite à chaque type de criblage.

3. SBVS

1) Préparation du monomère PfCCT-Nter pour le docking

Le criblage a été effectué sur le site actif du monomère N-terminal de la PfCCT, car c'est le monomère que le laboratoire de Montpellier exprimait principalement au début du projet. Cependant, il faut tout de même remarquer que le site actif du monomère C-terminal est absolument identique, comme vous pouvez le voir dans le tableau 9 comparant les résidus des sites actifs de différentes CCT (*Rattus norvegicus*, *Homo sapiens*, *Plasmodium falciparum*).

Comme la structure cristallographique de la PfCCT n'existe pas, j'ai construit un modèle par homologie à partir de la costructure de la CCT du rat avec le produit de la réaction, la CDP-choline. Cette modélisation vous a été présentée précédemment dans ce chapitre. Cependant, ce n'est pas cette structure que j'ai utilisé pour le criblage virtuel. En effet, je voulais garder le site actif le plus ressemblant possible à celui de la CCT de *Rattus norvegicus*. J'ai donc utilisé SCWRL4 (Krivov et al., 2009). SCWRL4 est un programme de positionnement des chaînes latérales basé sur une banque de rotamères possibles. L'alignement utilisé est le même que celui qui a servi à construire le modèle par homologie avec Modeller. Les résidus identiques entre *Plasmodium falciparum* et *Rattus norvegicus* sont gardés dans la même position tandis que les résidus différents voient leur position choisie dans une bibliothèque de rotamères. Nous utilisons l'option qui permet de ne pas ajouter les hydrogènes (-h). Pour ajouter les hydrogènes au modèle obtenu par SCWRL4, j'ai utilisé PrepDock (<http://swift.cmbi.ru.nl/servers/html/prepdock.html>), un serveur en ligne. Suite à cela, j'ai choisi l'état protoné (+1) pour l'HIS124 comme le suggère le réseau de liaisons déduit de la costructure.

La seule différence majeure de la CCT entre *Plasmodium falciparum* et *Rattus norvegicus* est une glutamine (Q51) à la place d'une alanine (A95). J'ai recherché la position de la GLN51 qui nous permettait de replacer correctement la CDP-choline avec le logiciel Autodock. Le but est de retrouver la conformation de la CDP-choline telle qu'elle est dans le site actif de rCCT. Ainsi, le criblage virtuel pourra s'effectuer sur une structure rigide. Pour cela, je me suis focalisée sur les poses proposées par Autodock d'énergie minimale, qui présentent une géométrie du pont hydrogène optimale. En effet, dans la structure de *Rattus norvegicus*, il y a une molécule d'eau (HOH401) en interaction avec la CDP-choline mimant cette glutamine (Figure 58). A partir de cette structure, j'ai relaxé tous les résidus à 6 Å de la CDP-choline (ligand CDC), en faisant trois minimisations à l'aide d'une dynamique moléculaire (Amber9). Tous les résidus à plus de 6 Å sont maintenus fixes. Suite à ces étapes de minimisations, seule l'arginine 170 voit sa position modifiée : elle s'écarte du motif cytidyl.

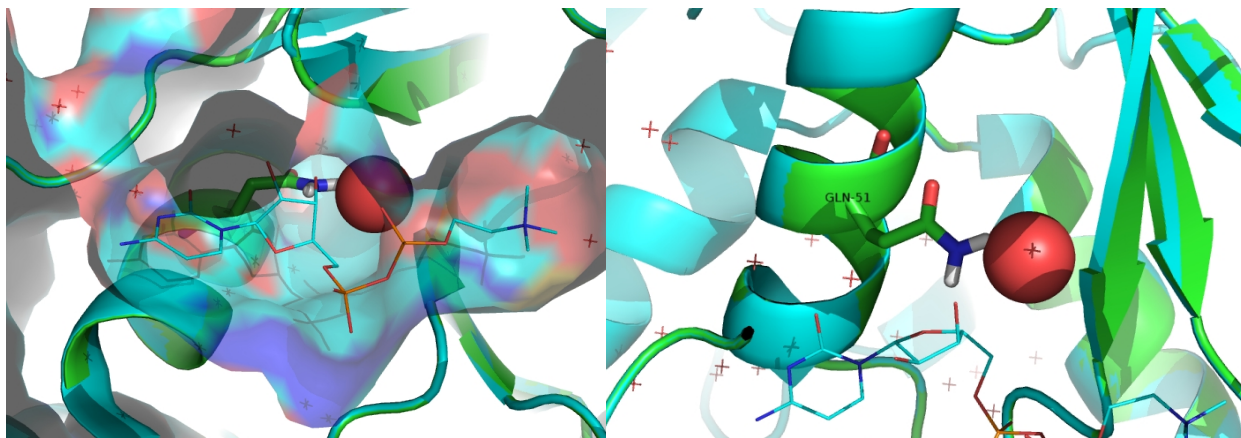


Figure 58 - Positions de la molécule d'eau 401 dans la structure 3HL4 (bleu) et de la GLN51 dans le modèle de *PfCCT-Nter* (vert).

2) Choix des logiciels de docking

Suite au criblage des bases de données DB et CN, j'ai voulu voir si les résultats de docking étaient corrélés. Pour cela, j'ai représenté les valeurs de score correspondantes à deux logiciels parmi Autodock, Vina, et Plants. Nous observons que Vina et PLANTS sont les plus corrélés en comparaison à Autodock (Figure 59). C'est la raison pour laquelle j'ai choisi d'utiliser ces deux logiciels.

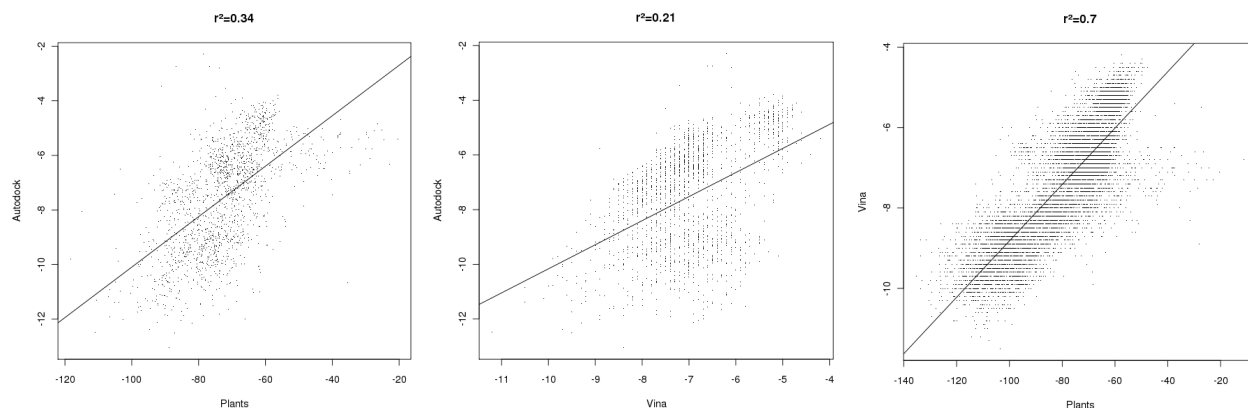


Figure 59 - Corrélations entre les scores de trois logiciels de docking : Autodock, Vina et PLANTS.

A la place d'Autodock, qui nous a permis de choisir une structure rigide de *PfCCT-Nter*, avec la glutamine 51 optimisée, j'ai donc choisi d'utiliser le logiciel de docking Vina qui a l'avantage d'être plus rapide et plus performant d'après la littérature. La préparation de la structure pour Vina a été réalisée grâce à AutodockTools. Lors des calculs, il nous faut spécifier le récepteur (*PfCCT-Nter*), le ligand (chaque molécule de la base), le centre du site actif (coordonnées x, y, z récupérées dans AutodockTools) et la taille de la boîte de recherche (25x25x25 Å³).

Pour le logiciel PLANTS, il faut d'abord générer le site actif en précisant le résidu souhaité comme centre : ici la CDP-choline CDC 165 A puis le rayon du site, soit 8 Å. Ensuite il

faut appeler le fichier contenant les paramètres choisis et spécifier le fichier au format sdf contenant les molécules à cribler.

3) Détermination du filtre basé sur la structure

J'ai choisi de coupler les résultats des deux logiciels de docking afin d'obtenir un consensus. Pour être retenue, une molécule de la chimiothèque doit présenter une pose de docking intéressante dans les deux logiciels : avoir un bon score et établir un minimum de liaisons hydrogènes. Pour déterminer quels seuils de score choisir lors du criblage virtuel, j'ai énuméré les différents tautomères et conformères de cycle de la CDP-choline. Ils sont au nombre de 3 tautomères x 7 conformères de cycle = 21 structures. Les résultats sont présentés dans le tableau 12.

Numéro de structure	Score Vina	Liaisons Hydrogènes Vina	Score PLANTS	Liaisons Hydrogènes PLANTS	Position identique à la costructure ?	
					Vina	PLANTS
CDP_choline@1_0	-9,6	8	-105,71	4	oui	non
CDP_choline@1_1	-9,2	7	-108,49	6	oui	non
CDP_choline@1_2	-8,8	5	-113,19	6	non	oui
CDP_choline@1_3	-9,4	7	-120,68	7	oui	oui
CDP_choline@1_4	-8,4	4	-111,10	5	non	non
CDP_choline@1_5	-8,4	5	-89,75	6	non	non
CDP_choline@1_6	-8,3	7	-105,98	3	non	non
CDP_choline@2_0	-9,2	8	-114,99	8	oui	oui
CDP_choline@2_1	-8,7	7	-107,27	8	non	oui
CDP_choline@2_2	-9	5	-105,92	5	non	non
CDP_choline@2_3	-9,3	7	-113,46	6	oui	non
CDP_choline@2_4	-8,7	3	-110,77	5	non	non
CDP_choline@2_5	-8,5	4	-96,97	3	non	non
CDP_choline@2_6	-8,9	6	-102,12	2	non	non
CDP_choline@3_0	-9,4	7	-116,22	7	oui	oui
CDP_choline@3_1	-8,2	4	-116,17	7	non	oui
CDP_choline@3_2	-8,8	4	-99,09	4	non	non
CDP_choline@3_3	-9,3	7	-106,84	7	oui	oui
CDP_choline@3_4	-8,7	5	-104,00	5	non	non
CDP_choline@3_5	-8,6	5	-91,84	4	non	non
CDP_choline@3_6	-8,9	5	-97,13	5	non	non
Moyenne	-8,87	5,71	-101,79	5,38		

Tableau 12 - Score et nombre de liaisons hydrogènes avec les résidus du site actif des 21 conformères de la CDP-choline par deux logiciels de docking, Vina et PLANTS. La dernière colonne indique si la pose de docking Vina ou PLANTS reproduit celle trouvée dans la costructure de la rCCT.

Suite à ces résultats, j'ai défini mon consensus de cette manière :

DOCKING CONSENSUS (DC)	{	Score Vina < -8
		Liaisons hydrogènes Vina > 1
		Score PLANTS < -90
		Liaisons hydrogènes > 1

Parfois, lorsque le nombre de molécules à visualiser était vraiment trop important, j'ai ajouté d'autres contraintes. J'ai augmenté le nombre minimal de liaisons hydrogènes de 2 à 4. Afin de s'assurer de la solubilité de la molécule, importante pour les tests *in vitro*, j'ai demandé un logP entre 0 et 3.

4) Application du filtre

<i>Chimiothèque</i>	<i>Structures (Molécules)</i>	<i>ROCS + DC</i>	<i>Align-it + DC</i>	<i>ECFP4 + DC</i>
DB	254 825 (5 935)	9 (9)	0 (0)	2 753 (97)
CN	2 493 966 (40 243)	10 (10)	3 (3)	8 (2)
CHEMBL	1 167 303 (26 449)	55 (39)	3 (2)	115 (2)
CB	6 060 089 (481 421)	620 (491)	81 (41)	0 (0)
DL	32 316 253 (1 967 820)	8 060 (6 795)	508 (243)	0 (0)
AMB	77 313 514 (4 899 277)	16 967 (13 675)	1 188 (641)	7 688 (22)
e-Drug3D	163 721 (1 545)	0 (0)	0 (0)	0 (0)
Total	119 769 671 (7 422 690)	25 721 (21 019)	1 783 (930)	10 564 (123)

Tableau 13 - Nombre de molécules retenues suite à chaque type de criblage et du consensus de docking.

4. Résultats du protocole de criblage virtuel

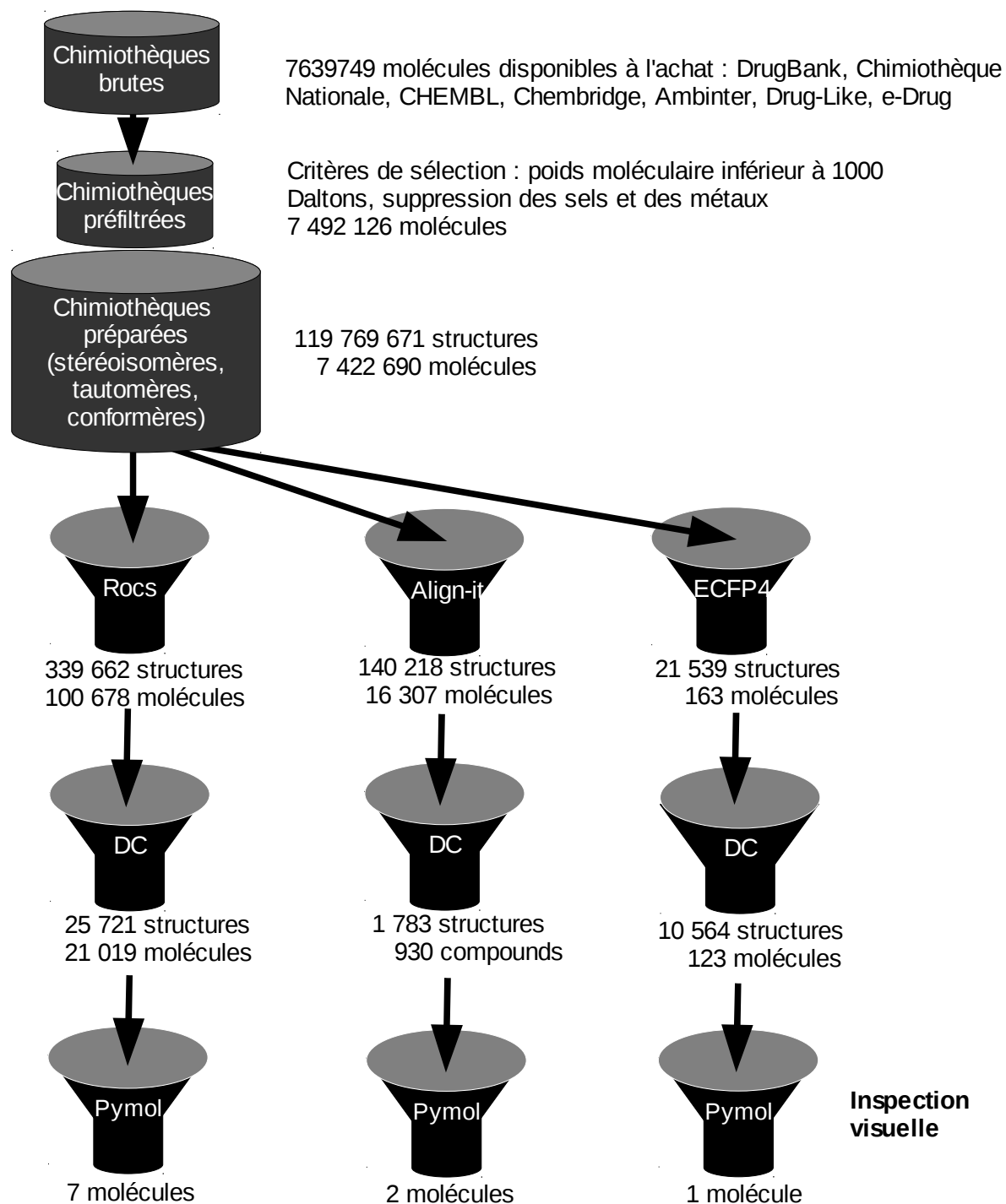


Figure 60 - Criblage virtuel utilisé pour identifier des inhibiteurs potentiels de la *PfCCT* avec le nombre de molécules et de structures qui passent chaque filtre. (DC : Docking consensus, c'est-à-dire Score_Plants <-90, Score_Vina <-8 et chaque position de docking partageant au moins deux interactions hydrogènes avec celles établies par la CDP-choline).

En complément, les paragraphes suivant mettent l'accent sur le criblage des molécules médicaments ou d'antipaludiques variés.

1) Criblage de la chimiothèque e-Drug3D

La base de données e-Drug3D, développée au laboratoire, contient les médicaments passés et présents approuvés par la FDA (Pihan et al., 2012). L'idée en criblant cette base est de trouver une nouvelle cible à un médicament pré-existant. Nous n'avons identifié qu'un seul hit par les empreintes circulaires ECFP4 : la cytarabine, médicament utilisé dans le traitement de la leucémie. Elle a un score Tanimoto supérieur à 0,6. C'est une sous-structure de la CDP-choline et un analogue de cytidyl. Cependant, les scores de docking n'était pas suffisant pour qu'elle passe le filtre consensus de docking.

Select all Unselect all		Pharmacology FDA Registration Chemistry Properties Similarities Most similar drugs Fragments					
Subselect		ININ	Class	Route (list)	Target	PDB	Links
1	CYTARABINE	CYTARABINE	ATC L01BC01 ANTIVIRAL ANTINEOPLASTIC ANTIMETABOLITES IMMUNOSUPPRESSIVE AGENTS	INJECTION	CYTIDINE DEAMINASE DNA POLYMERASE BETA SERUM ALBUMIN		DRUGBANK DB00987 ChEMBL

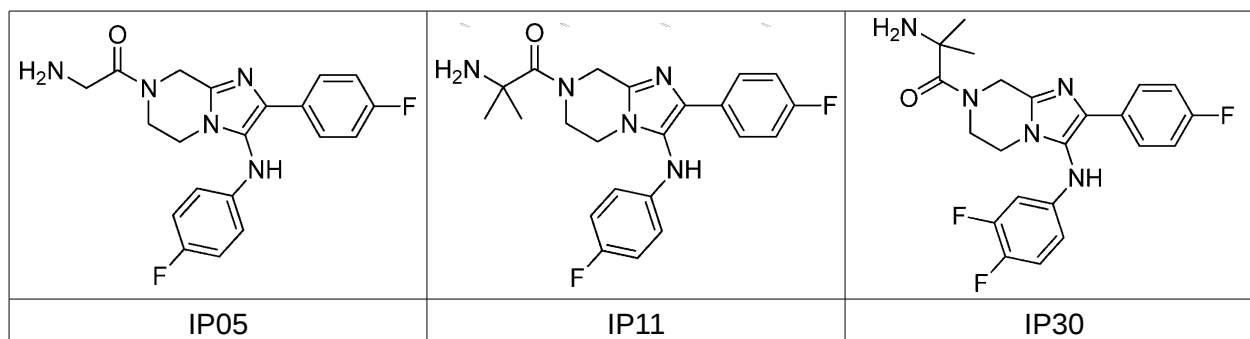
Figure 61 - Capture d'écran de l'entrée "CYTARABINE" dans la base de données e-Drug3D.

2) Molécules antipaludiques sans cible connue

Au cours de ma thèse, j'ai effectué une veille bibliographique des molécules antipaludiques publiées sans cible connue ou ciblant d'autres enzymes afin de récupérer des inhibiteurs potentiels de la *PfCCT*.

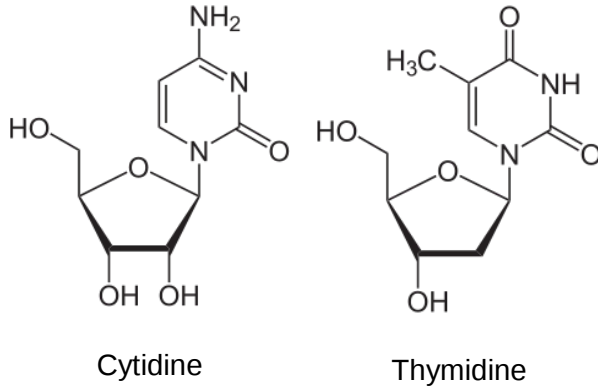
a) Imidazolopiperazines

A la lecture de la publication (Wu et al., 2011), j'ai sélectionné les trois meilleures imidazolopiperazines pour les cribler dans le site actif de la *PfCCT*. Cette famille de molécules est à l'origine du GNF156 (ou KAF156) développé par Novartis (page 26).



La visualisation des différentes poses de docking montre que ces molécules ne semblent pas capable de compéter avec la CDP-choline car elles ne la miment pas assez bien. Elles sont trop petites et se superposent mal à la molécule de référence.

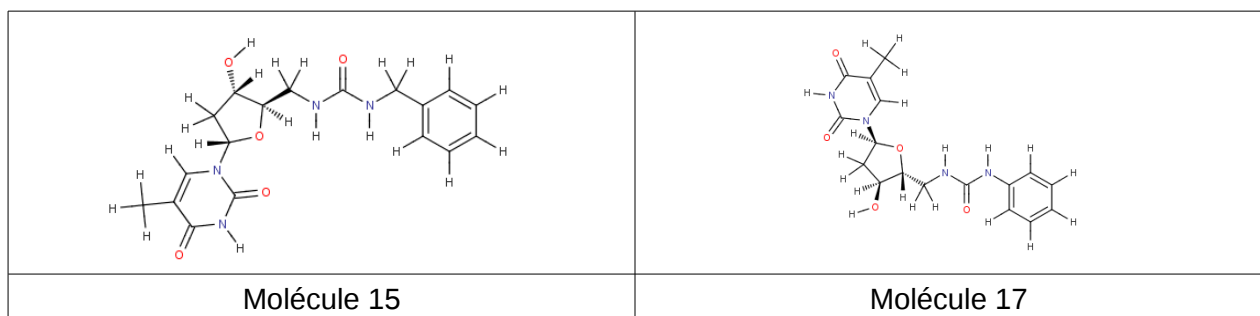
b) Criblage d'analogues de thymidine



→ Le CTP, substrat de la *PfCCT*, est un dérivé triphosphate d'un nucléoside, la cytidine. Nous nous sommes alors demandé si des dérivés d'un autre nucléoside, la thymidine, pourraient se lier au site actif de la *PfCCT*. Nous avons une chimiothèque développée pour cibler la TMPK de *Mycobacterium tuberculosis* à notre disposition, contenant 544 molécules.

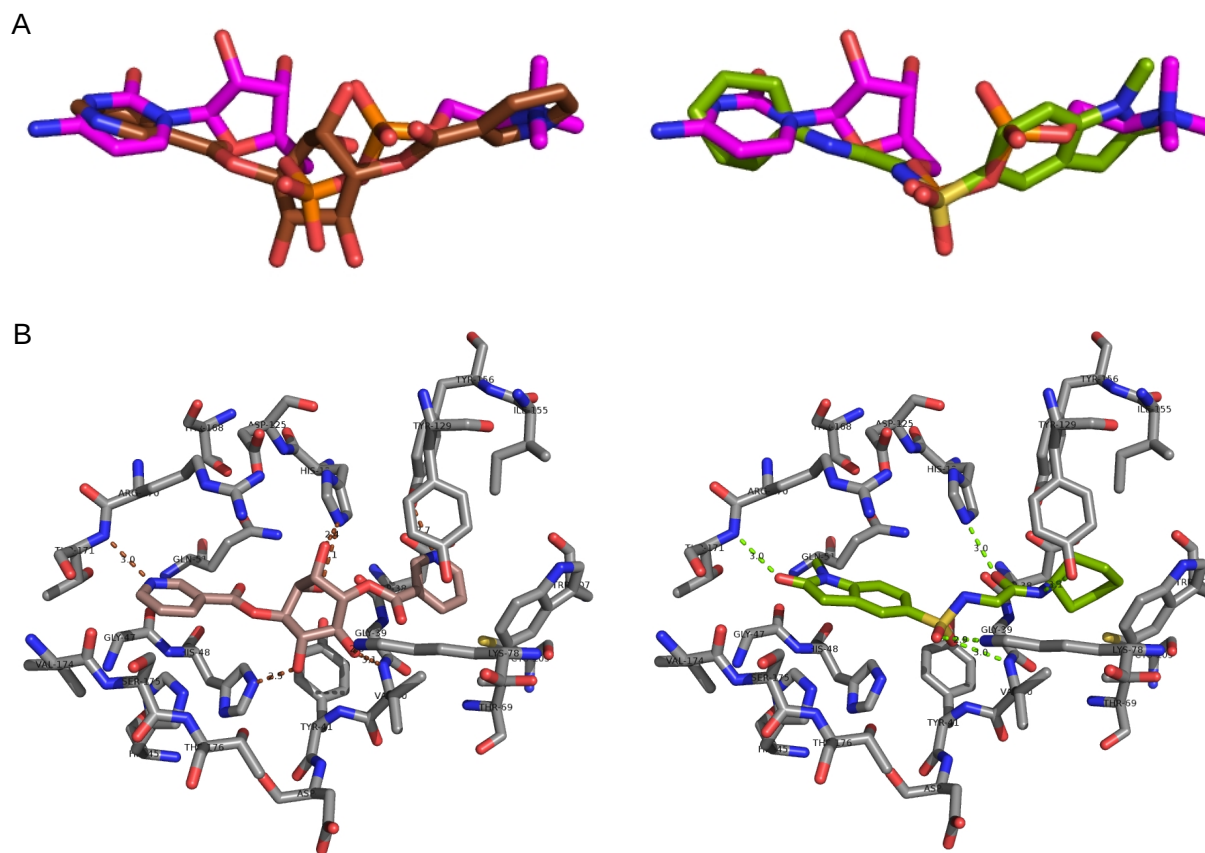
Nous avons choisi de ne garder que 221 molécules, dont le poids moléculaire se situait entre 300 et 600. Après préparation, ces molécules sont représentées par 2807 structures. 686 structures, représentant 84 molécules ont passé le filtre DC. Mais après visualisation, aucune molécule n'a été retenue.

→ D'autres analogues de thymidine ciblant la *PfTMPK* ont été publiés (Cui et al., 2012). Nous les avons également criblés. Nous pouvons observer que le K_i de la molécule 17 pour la *PfTMPK* est de 166 μ M alors que l' IC_{50} est de 23 μ M sur la croissance parasitaire de *Plasmodium falciparum*. D'autres voies d'inhibition sont peut-être mises en jeu. C'est pour cette raison que j'ai criblé les meilleurs analogues d'alpha-thymidine (les molécules 15 et 17) dans le site actif de notre cible, la *PfCCT*. Après inspection visuelle sous Pymol, il semble évident que ces deux molécules sont trop petites pour occuper le site actif de la *PfCCT*.



3) Visualisation des poses de docking avec Pymol 3D

J'ai visualisé les milliers de poses en les comparant à la CDP-choline. J'ai sélectionné les composés qui mimaient au mieux cette molécule, qui permettaient de garder des liaisons hydrogènes importantes et qui avaient une forme globale similaire à celle de la molécule de référence. Par exemple, dans la figure 62, les composés 1 et 12 répondent bien à ces divers critères.



4) Molécules sélectionnées

Dans un premier temps, j'ai criblé les petites bases de données comme la DB, la CN et la CB. Après visualisation des poses de docking, j'ai sélectionné des molécules afin de les acheter. Le premier tour contient 5 molécules (dont une de la chimiothèque DL car il était possible de l'acheter chez le vendeur Chembridge). Le deuxième tour contient 9 molécules, certaines ont été retenues dans les chimiothèques DL et AMB, d'autres sont des analogues de choline et de cytidyl. Puis le troisième et quatrième tour contiennent des analogues d'un hit identifié lors du deuxième tour. Le cinquième tour contient trois analogues de synthèse du composé 12.

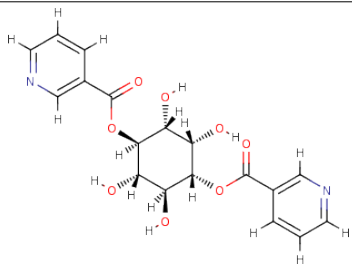
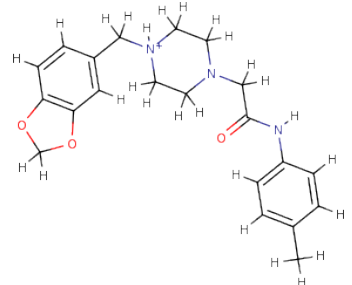
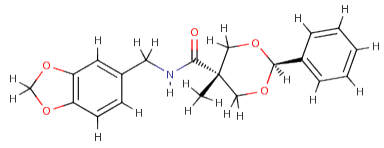
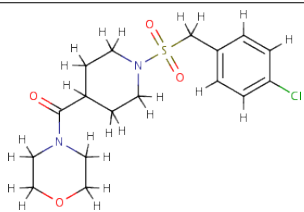
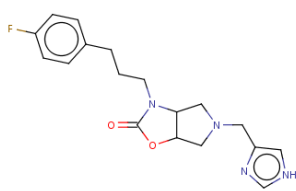
Fournisseur	Identifiant	Structure
Chembridge (provient de la chimiothèque CB via l'outil ROCS) Poids moléculaire : 391	5233902 Composé 1	
Chembridge (provient de la chimiothèque CB via l'outil ROCS) Poids moléculaire : 368	6509688 Composé 2	
Chembridge (provient de la chimiothèque CB via l'outil ROCS) Poids moléculaire : 355	7964114 Composé 3	
Chembridge (provient de la chimiothèque CB via l'outil Align-it) Poids moléculaire : 387	9191613 Composé 4	
Chembridge (provient de la chimiothèque DL via l'outil ROCS) Poids moléculaire : 344	15073438 Composé 5	

Tableau 14 - Molécules achetées lors du premier tour.

Fournisseur	Identifiant	Structure
Enamine (provient de la chimiothèque AMB via l'outil ROCS) Poids moléculaire : 415	Z395283652 Composé 6	
Chembridge (provient de la chimiothèque AMB, analogue de choline) Poids moléculaire : 331	5132269 Composé 7	
Chembridge (provient de la chimiothèque DL via l'outil ROCS) Poids moléculaire : 394	68041941 Composé 8	
Toronto Research Chemicals (provient de la chimiothèque AMB, analogue de choline) Poids moléculaire : 437	H941000 Composé 9	
Vitas-M Lab (provient de la chimiothèque AMB, analogue de choline) Poids moléculaire : 463	STK367919 Composé 10	
Otava (provient de la chimiothèque AMB, analogue de cytidyl) Poids moléculaire : 348	5060030060 Composé 11	
Otava (provient de la chimiothèque AMB via l'outil ROCS) Poids moléculaire : 359	1458183 Composé 12	
Otava (provient de la chimiothèque AMB, via l'outil Align-it) Poids moléculaire : 399	1645949 Composé 13	
SantaCruz Biotechnology (provient de la chimiothèque AMB via l'outil ECFP4) Poids moléculaire : 381	sc-210418 Composé 14	

Tableau 15 - Molécules achetées lors du deuxième tour.

Suite au deuxième tour, nous avons identifié une molécule pouvant inhiber la PfCCT avec une IC₅₀ de 826 ± 88 µM (Composé 12). J'ai donc cherché des analogues commerciaux de cette molécule en utilisant Chemspider (<http://www.chemspider.com/>). Pour cela, j'ai cherché les molécules similaires ayant un score tanimoto supérieur ou égal à 90% (Figure 63).

Simple **Structure** Advanced ▾ More searches...

1. Input your structure (choose a, b or c)

a. Upload a structure file (MOL, SDF, CDX) or image file (PNG, JPG, GIF).
Choisissez un fichier | Aucun fichier choisi

b. Convert to structure using a Name, SMILES, InChI or ChemSpider ID.
[] Convert

c. Click the image to draw out the structure yourself.

2. Edit molecule

Exact
 Substructure
 Similarity

Search Options
Tanimoto ▾ >=90% ▾ ⓘ

Powered By **GGA**
GGA Software Services LLC

Conditions for Using Bing:
GGA's Molecular Search Engine

► Options

Search Clear form Search Hits Limit: 100 ▾

Figure 63 - Capture d'écran de la requête effectuée sur Chemspider.

Il y a 178 molécules dont 165 qu'il est possible de télécharger. La préparation en stéréoisomères, tautomères et conformères génère 6332 structures. Suite au docking, 2999 structures représentant 124 molécules passent le filtre DC. Suite à une inspection visuelle, j'ai retenu 38 molécules. Parmi elles, j'ai éliminé celles qui avaient une TPSA inférieure à 80 Å². Il en restait alors 30. Parmi celles-ci, nous en avons retenues 4 à acheter (Composés 15 à 18).

Fournisseur	Identifiant	Structure
Otava (Analogue du composé 12) Poids moléculaire : 359	1456680 Composé 15	
ChemDiv (Analogue du composé 12) Poids moléculaire : 393	C464-1907 Composé 16	
ChemDiv (Analogue du composé 12) Poids moléculaire : 407	E197-0228 Composé 17	
ChemDiv (Analogue du composé 12) Poids moléculaire : 407	E570-0585 Composé 18	

Tableau 16 - Molécules achetées lors du troisième tour.

Les résultats du troisième tour ont montré que le groupement benzyl (Composé 15) se plaçant dans la poche choline n'augmentait pas l'affinité vis-à-vis de la *PfCCT*. Pour le tour 4, j'ai donc gardé le cycle cyclohexyl fixe et j'ai cherché des groupements analogues à l'indoline du composé 12. Nous avons choisi deux molécules dans la liste des 30 analogues du composé 12 pour en choisir deux autres, beaucoup plus similaires (Composés 19 et 20).

D'autre part, nous avons fait une recherche sur Chempider, de molécules contenant la sous-structure du linker du composé 12 (Figure 63). Nous avons récupéré 7195 candidats dont 6128 structures étaient disponibles au téléchargement.

Simple **Structure** Advanced ▾ More searches...

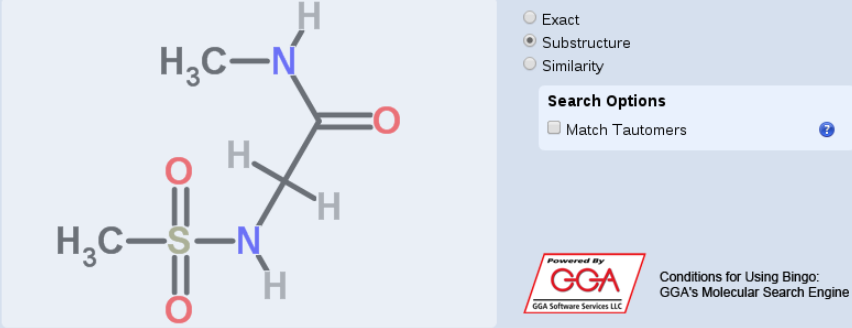
1. Input your structure (choose a, b or c)

a. Upload a structure file (MOL, SDF, CDX) or image file (PNG, JPG, GIF).
 Choisissez un fichier | Aucun fichier choisi

b. Convert to structure using a Name, SMILES, InChI or ChemSpider ID.


c. Click the image to draw out the structure yourself.

2. Edit molecule



Exact
 Substructure
 Similarity

Search Options
 Match Tautomers


 Powered By
 GGA Software Services LLC
 Conditions for Using Bingo:
 GGA's Molecular Search Engine

Options

Search Hits Limit: 10000 ▾

Figure 64 - Capture d'écran de la requête effectuée sur Chemspider.

Suite à la préparation, nous avons généré 52 472 structures représentant 6053 molécules. Nous avons diminué ce nombre en ne gardant que les molécules dont le TPSA était supérieur à 80, le nombre de liaisons libres inférieur à 10 et le poids moléculaire entre 300 et 500. De ces molécules, nous n'avons généré que le tautomère le plus probable. Le filtre DC retient alors 3906 structures représentant 1252 molécules. Comme nous voulions impérativement une interaction hydrogène avec les chaînes principales de la THR171 ou de la VAL174, nous avons post-filtré les résultats pour ne garder que 912 structures représentant 371 molécules. L'inspection visuelle sous Pymol nous a aiguillé vers deux molécules : les composés 21 et 22. Ils n'étaient pas présents dans la liste des molécules similaires à plus de 90% du composé 12. Ces composés établissent une liaison hydrogène avec la thréonine 171 mais pas avec la valine 174.

Dans ce quatrième tour, les modifications sont beaucoup plus petites : passage d'un cycle à 5 atomes à un cycle à 6 pour le bicyclic (Composé 19), suppression d'un méthyle (Composé 20), ajout d'un éthyle, et/ou remplacement d'un carbone par un oxygène (Composés 21 et 22).

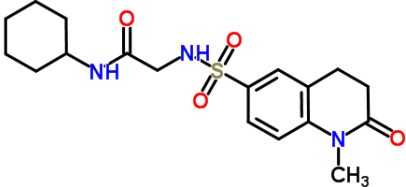
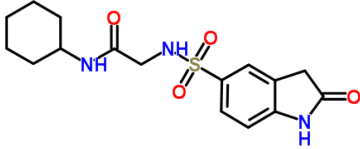
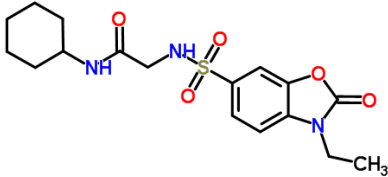
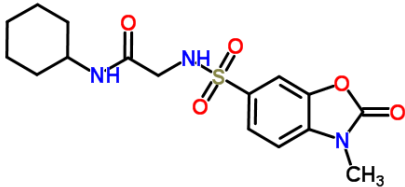
Fournisseur	Identifiant	Structure
Otava (Analogue du composé 12) Poids moléculaire : 379 CSID 25289726	3385854 Composé 19	
Otava (Analogue du composé 12) Poids moléculaire : 351 CSID 25372521	1458181 Composé 20	
Otava (Analogue du composé 12) Poids moléculaire : 381 CSID 25109204	1489864 Composé 21	
Otava (Analogue du composé 12) Poids moléculaire : 367 CSID 25372514	1458173 Composé 22	

Tableau 17 - Molécules achetées lors du quatrième tour.

5) Synthèse d'analogues du Composé 12

Suite aux premiers résultats expérimentaux, nous avons décidé de nous concentrer plus particulièrement sur le composé 12. En effet, ce composé est chimiquement attractif car la fonction sulfonamide est courante dans les médicaments approuvés et la synthèse est aisée. Afin d'améliorer l'affinité, nous avons cherché des analogues commerciaux (troisième et quatrième tours) mais nous avons aussi proposé nos propres analogues (cinquième tour, contenant 3 molécules). Pour la synthèse chimique, nous avons fait appel à Sylvie Pochet, chimiste de l'Institut Pasteur. Les tests expérimentaux de ces trois composés sont en cours. Les structures chimiques ne peuvent pas, pour l'instant, être dévoilées.

IV. Validation expérimentale dans le cadre d'une collaboration

Le groupe de la 'Biologie des Lipides' du laboratoire DIMNP Montpellier a développé les tests d'inhibition enzymatique sur protéine recombinante et endogène et ceux sur la croissance parasitaire de *Plasmodium falciparum*.

La *PfCCT* est l'enzyme limitante de la voie de biosynthèse de la phosphatidylcholine. Elle a été caractérisée au niveau enzymatique dans diverses études et dans le laboratoire DIMNP de Montpellier. Son affinité pour ses substrats est faible :

Publication	(Lee et al., 2009)	(Yeo et al., 1997)	(Nagy et al., 2013)	DIMNP
Enzyme	<i>rCCT</i> (1-236)	<i>PfCCT</i> (528-896)	<i>PfCCT</i> (528-795) Δ K	<i>PfCCT</i> (528-795)
K _m (phosphocholine)	0,33mM	0,49mM	1,8mM	3,6mM
K _m (CTP)	7,3mM	10,9mM	0,168mM	5,5mM

Tableau 18 - Valeurs d'affinité de constructions de la *PfCCT* pour ses substrats.

1. Inhibition *in vitro* de la *PfCCT* recombinante et endogène

L'activité de la protéine recombinante *PfCCT*(528-795), domaine C-terminal, a été évaluée par Marina Lavigne en présence des 25 composés issus du criblage virtuel. Le composé 12 présente la meilleure inhibition reproductible (n=3) de l'enzyme avec une IC₅₀ de 826 ± 88 μM.

Le meilleur composé, le numéro 12, a aussi été testé sur *PfCCT* endogène. On observe une IC₅₀ d'environ 1,3 mM. Cela reste dans l'ordre de grandeur de ce qui a été trouvé sur la *PfCCT* recombinante. Les autres molécules n'ont pas été testées sur la *PfCCT* endogène.

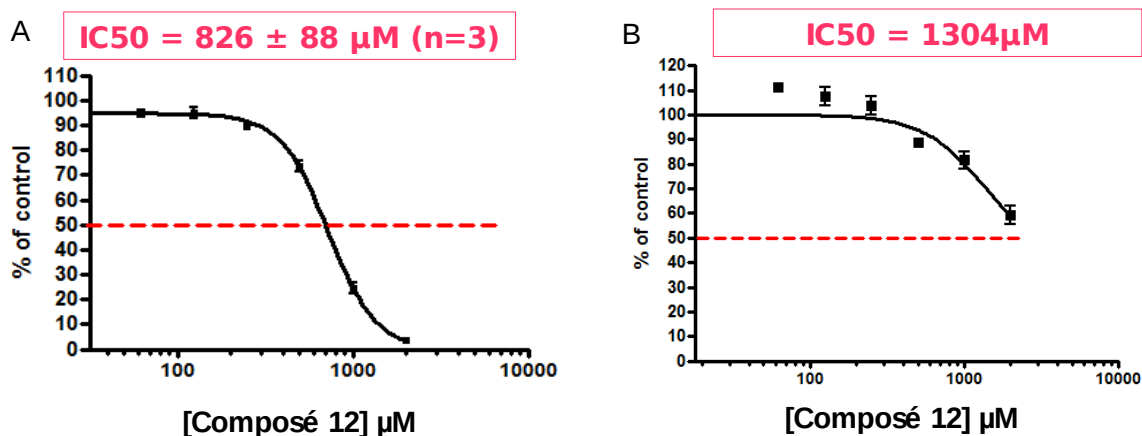
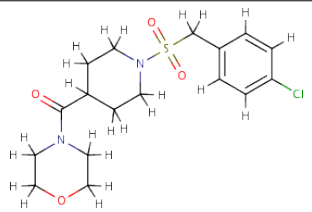
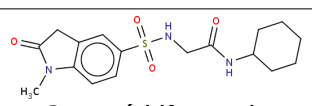


Figure 65 - Inhibition A) du domaine C-terminal de la *PfCCT* [598-795] et B) de la *PfCCT* endogène (souche 3D7) par le composé 12. Expériences réalisées par Marina Lavigne.

J'ai utilisé le serveur en ligne FAF-Drugs2 (Lagorce et al., 2011) sur les 25 composés afin de détecter les PAINS décrits précédemment par (Baell & Holloway, 2010). Les PAINS sont des sous-structures retrouvées dans des composés actifs sur différentes protéines dans plusieurs expériences de high throughput screening (HTS). Ils sont actifs de façon non spécifique. Le composé 9 est un PAINS par la présence d'un groupement azo.

Le serveur FAF-Drugs2 permet aussi de chercher les sous-structures indésirables qui induisent de la toxicité lors des test *in vitro* et qui sont présentées à l'adresse suivante : <http://bioserv.rpbs.univ-paris-diderot.fr/FAF-Drugs/groups.html>.

N°	Provenance	PfCCT recombinante	Remarques	FAF-Drugs2 Sous-structures indésirables
1	CB, ROCS		Instable dans l'eau	Aucune
2	CB, ROCS		ABANDON	Methylenedioxy
3	CB, Align-it		ABANDON	Methylenedioxy
4	CB, ROCS	500µM		Aryl halide Sulfonamide
5	DL, ROCS			Aryl halide
6	AMB, ROCS		Insoluble	Alkene
7	AMB, analogue de choline			Aucune
8	DL, ROCS			Aucune
9	AMB, analogue de choline			Azo1 Azo2 Phenol
10	AMB, analogue de choline		Insoluble	Alkene
11	AMB, analogue de cytidyl			Aucune
12	AMB, ROCS	826 µM ± 88 µM (n=3)	 Compétitif pour la phosphocholine	Sulfonamide
13	AMB, Align-it			Aryl halide
14	AMB, ECFP4			Rejetée : Sulfonate ester Arenesulfonyl
15	Analogue du composé 12		Précipité en DMSO	Sulfonamide
16	Analogue du composé 12		Peu soluble	Sulfonamide
17	Analogue du composé 12		Peu soluble	Sulfonamide

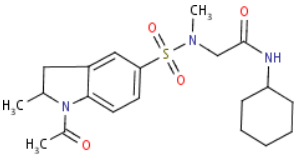
18	Analogue du composé 12	672 μ M	 Peu soluble	Sulfonamide
19	Analogue du composé 12			Sulfonamide
20	Analogue du composé 12			Sulfonamide
21	Analogue du composé 12			Sulfonamide
22	Analogue du composé 12			Sulfonamide
23	Analogue synthétisé du composé 12			Sulfonamide
24	Analogue synthétisé du composé 12			Sulfonamide
25	Analogue synthétisé du composé 12			Sulfonamide

Tableau 19 - Résultats des expériences d'inhibition *in vitro* sur la *PfCCT* recombinante. Les cellules en gris foncé représentent une IC_{50} pour la *PfCCT* supérieure à 1mM. Les cellules en gris clair représentent les expériences qui n'ont pas pu être réalisées.

2. Type d'inhibition de la *PfCCT* recombinante

Plusieurs types d'inhibition ont été bien caractérisés. Les enzymes peuvent être inhibées de façon compétitive (a), quand le substrat et l'inhibiteur sont en compétition pour se lier au même site actif. Elles peuvent être inhibées de façon non-compétitive (b) quand l'inhibiteur se lie autre part sur l'enzyme et diminue son activité. Enfin, elles peuvent être inhibées de façon incompétitive (c), quand l'inhibiteur se lie uniquement au complexe enzyme-substrat.

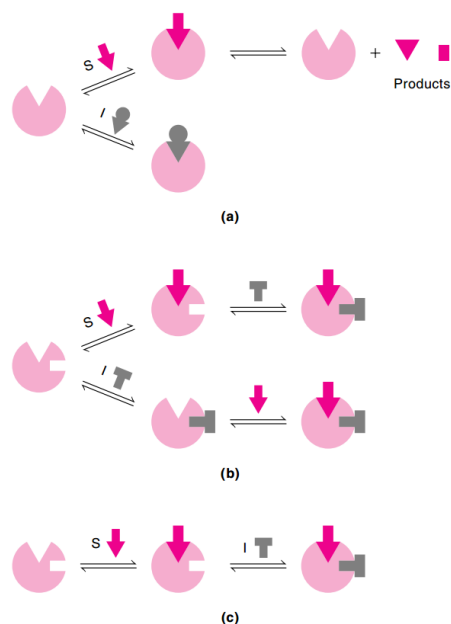


Figure 66 - Mécanismes d'inhibition (a) compétitive, (b) non-compétitive et (c) incompétitive.

Une expérience a donc été menée par Marina Lavigne pour déterminer le type d'inhibition du composé présentant la meilleure inhibition de la *PfCCT*, c'est-à-dire le composé 12. On a cherché son type d'inhibition vis-à-vis d'un des substrats, la phosphocholine. D'après les premiers résultats, il semble que le composé 12 inhibe la *PfCCT* de manière compétitive pour le substrat phosphocholine. Nous pouvons donc supposer que cette molécule va se fixer dans le site actif.

3. Test *in vitro* de croissance parasitaire

Les différents composés ayant montré une activité sur l'enzyme recombinante *PfCCT*(528-795) ont été évalué sur les parasites par Marina Lavigne. Le composé 12 présente la meilleure inhibition de la croissance parasitaire avec une IC_{50} de $37 \pm 19 \mu M$ ($n=3$).

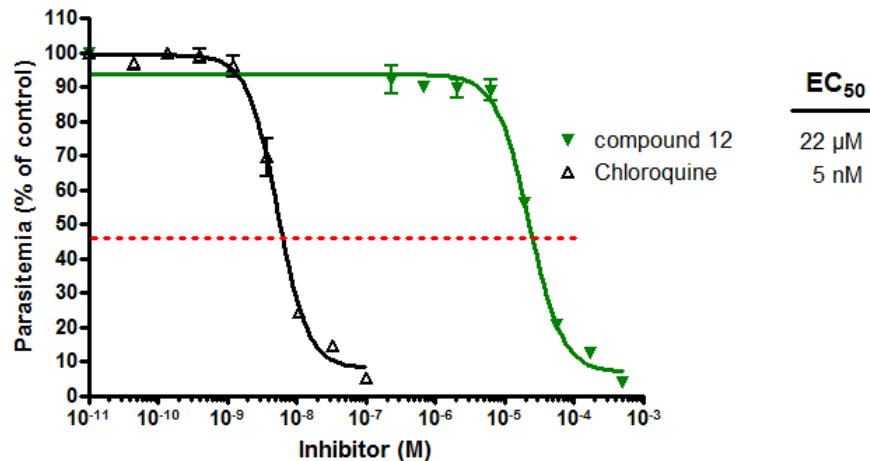


Figure 67 - Inhibition de la croissance parasitaire en présence du composé 12 ou de la chloroquine (référence). Expérience réalisée par Marina Lavigne.

Les composés 4 et 18 qui ont une IC_{50} inférieure à 1mM sur la *PfCCT* recombinante ont des valeurs respectives d' IC_{50} de $54 \pm 25 \mu M$ ($n=3$) et $130 \mu M$ sur la croissance parasitaire. Les autres composés montrent aussi des IC_{50} de l'ordre de dizaines de micromolaire mais à la vue de leur activité faible sur notre enzyme cible, la *PfCCT*, ils doivent agir par d'autres voies métaboliques. Ils n'apportent donc pas d'information du point de vue structure-fonction de notre cible.

V. Conclusion sur le criblage de l'enzyme PfCCT

Plasmodium falciparum est un parasite intracellulaire obligatoire. La multiplication des parasites dans la cellule hôte, l'érythrocyte, nécessite une quantité importante de phospholipides, et de façon majoritaire la phosphatidylcholine, constituant essentiel de la membrane des parasites. La biogenèse membranaire fait intervenir diverses voies de biosynthèse mais principalement la voie *de novo* Kennedy. Celle-ci se déroule en trois étapes catalysées par trois enzymes. L'enzyme limitante est la PfCCT.

La structure de la PfCCT n'ayant pas encore été résolue, j'ai d'abord réalisé un modèle à partir de la structure de la CCT de *Rattus norvegicus*, leurs domaines catalytiques étant proches (51% de similarité). J'ai mis au point un protocole de criblage utilisant les différentes données à notre disposition : la CDP-choline telle qu'elle est positionnée dans le site actif de la rCCT, le modèle de la PfCCT construit par homologie, et le réseau d'interactions hydrogènes entre les résidus du site actif de la PfCCT et la CDP-choline (Figure 53). Suite à une inspection visuelle, j'ai choisi 25 composés en plusieurs tours. Les positions d'amarrage vont être présentées dans les figures 68 à 75. Dans ces figures, la molécule de référence, la CDP-choline est en rose et les interactions qu'elle établit sont en pointillés rose (douze interactions hydrogènes de référence).

Le **composé 1** mime assez bien les deux phosphates de la CDP-choline grâce à ses quatre fonctions hydroxyles sur le cyclohexyle central. L'azote du groupement pyridyle permet une interaction avec la thréonine 171 (Figure 62B, gauche). Ce composé n'étant pas stable dans l'eau à cause de ses deux fonctions esters (rétro-estérification possible), il a été abandonné.

La partie morphol du **composé 4** établit une interaction hydrogène avec la thréonine 171. Le groupement sulfonamide au niveau de linker permet de mimer les phosphates de la CDP-choline et établit des interactions avec la chaîne principale du résidu valine 40 et avec la chaîne latérale de l'histidine 124. Côté choline, le groupement chlorobenzyle occupe bien la poche et peut créer des interactions de type pi-stacking avec la tyrosine 129.

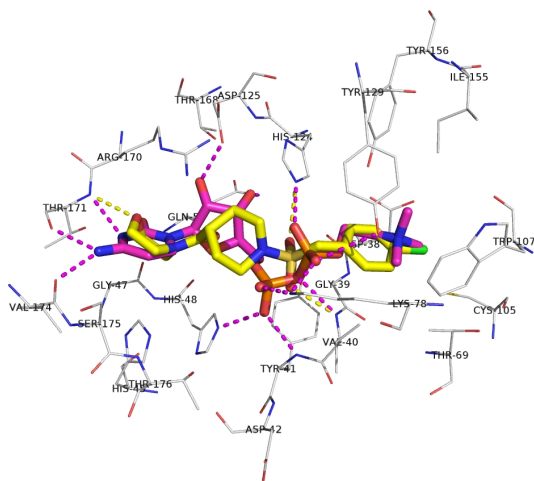


Figure 68 - Position du composé 4 (jaune) obtenue par amarrage. Les interactions hydrogènes que ce composé établit avec les résidus du site actif de la *PfCCT-Nter* sont représentées en pointillés jaune.

Les résultats expérimentaux de nos collaborateurs sur les deux premiers tours ont identifié un candidat : le composé 12. Ces données ont aiguillé notre recherche vers des analogues du composé 12 ou de son linker. Malheureusement, les analogues sélectionnés n'ont pas montré une meilleure activité que le composé 12 sur la croissance parasitaire. Nous pouvons tout de même essayer d'en déduire quelques relations structure-activité.

Le **composé 15** ne diffère du composé 12 que par son groupement benzyle situé dans la poche choline. Le benzyle est un groupement plus rigide que le cyclohexyle. Cette rigidité est transmise au groupement hydroxyle au niveau du linker qui perd l'interaction avec l'histidine 124 et au groupement sulfonamide qui mime alors moins bien le premier phosphate de la CDP-choline. Cette modification ne permet pas d'améliorer l'inhibition de la *PfCCT*. C'est pourquoi, les analogues suivants contiennent un groupement cyclohexyle dans la poche choline.

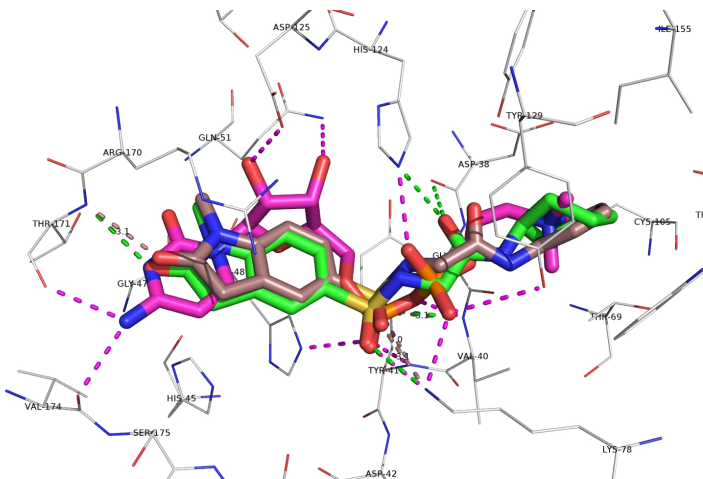


Figure 69 - Comparaison des positions du composé 15 (marron) et du composé 12 (vert) obtenues par amarrage.

Le **composé 16** quant à lui possède un cyclopentyle côté choline mais la présence d'un groupement différent sur le noyau dihydroindole par rapport au composé 12 ne permet pas de conclure quant à l'implication de chaque groupement dans la diminution de l'inhibition de la PfCCT. Il faudrait tester un analogue du composé 12 qui diffère uniquement côté choline, avec un cyclopentyle à la place du cyclohexyle. A la vue de la position d'amarrage, on suppose que le groupement sur le noyau dihydroindole crée une gêne stérique et génère une position non optimale du carbonyle vis-à-vis de la thréonine 171. De même, les groupements sulfonamide et hydroxyle du linker miment moins bien la CDP-choline que le composé 12.

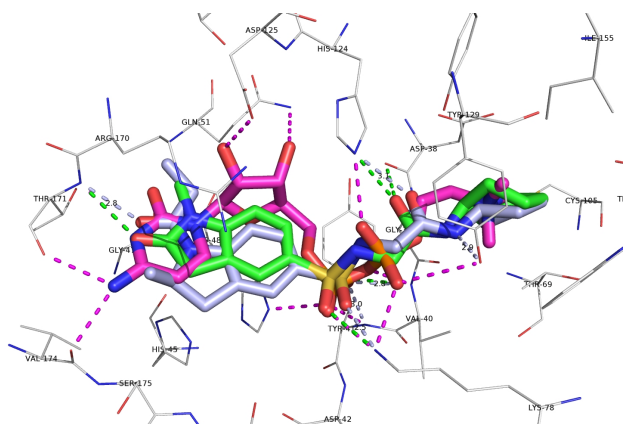


Figure 70 - Comparaison des positions du composé 16 (gris) et du composé 12 (vert) obtenues par amarrage.

Le **composé 17** possède deux méthyles sur le noyau indoline et un méthyle au niveau du linker. Ce dernier ne semble pas déranger la liaison car il est présent dans le composé 18, composé qui a montré une inhibition comparable à celle du composé 12. Ce sont donc les deux méthyles du bicyclic qui doivent gêner la liaison.

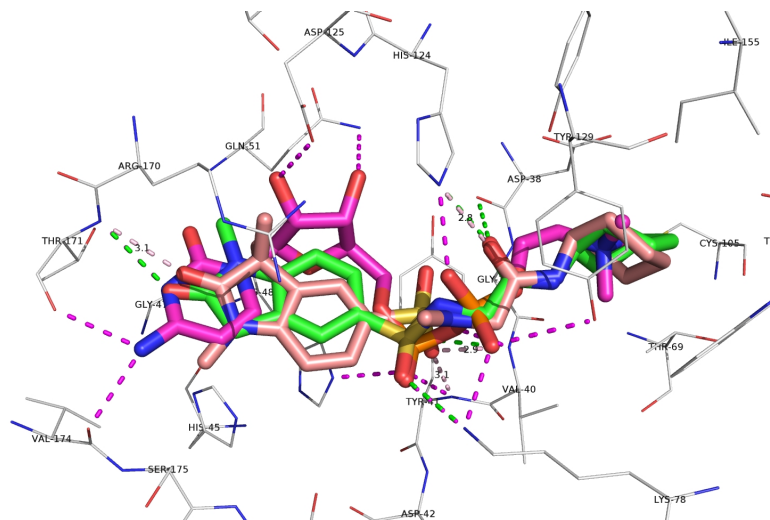


Figure 71 - Comparaison des positions du composé 17 (rose) et du composé 12 (vert) obtenues par amarrage.

Le **composé 19** diffère uniquement par un noyau tétrahydroquinoline à la place du noyau dihydroindole. Cette modification n'améliore pas l'affinité, elle modifie la position du groupement carbonyle, et l'angle par rapport à la thréonine 171. De la même manière que pour les composés 15, 16 et 17, le groupement sulfonamide mime moins bien le premier phosphate de la CDP-choline que le composé 12.

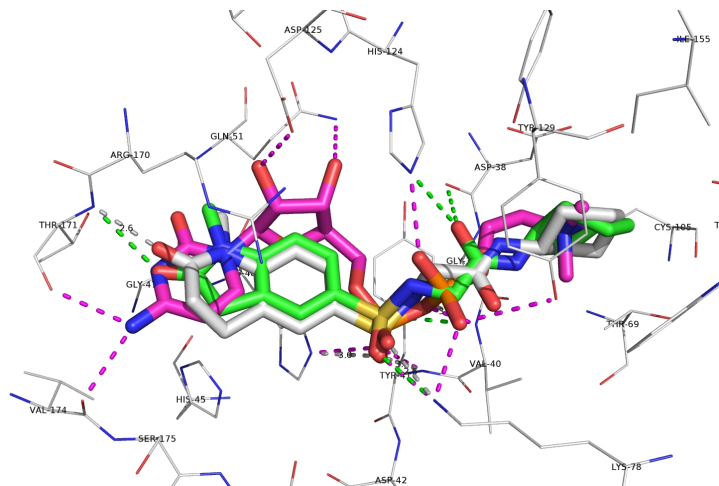


Figure 72 - Comparaison des positions du composé 19 (blanc) et du composé 12 (vert) obtenues par amarrage.

Le **composé 20** ne diffère du composé 12 que par la perte de son méthyle sur le noyau dihydroindole. Comme pour les composés précédents, cela engendre une modification de la position du noyau dihydroindole et donc de la position du groupement carbonyle, qui n'a alors plus le bon angle avec la thréonine 171 pour établir une liaison. Son absence diminue l'inhibition de la PfcCT ($IC_{50} > 2mM$).

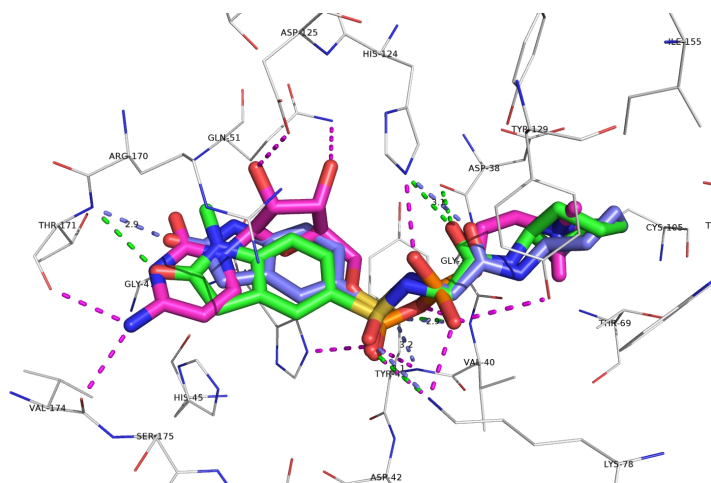


Figure 73 - Comparaison des positions du composé 20 (violet) et du composé 12 (vert) obtenues par amarrage.

Les **composés 21 et 22** possèdent un oxygène à la place d'un carbone dans le cycle à 5 atomes du noyau dihydroindole. Leurs positions d'amarrage sont quasiment identiques à celle du composé 12. La seule présence de l'oxygène diminue l'affinité vis-à-vis de la PfcCT ($IC_{50} > 1mM$).

Lorsque que l'on visualise plus précisément le mode de liaison du composé 12 ou du composé 18 (composés ayant montré une $IC_{50} < 1\text{mM}$ sur la *PfCCT* recombinante) par rapport à la position de la CDP-choline, on observe que le bicyclic mime assez mal le fragment cytidyl (Figures 69 à 75).

Pour le **composé 12**, seule l'interaction avec la chaîne principale de la thréonine 171 est conservée. Les interactions hydrogènes que les deux groupement hydroxyles du ribose établissent avec la glutamine 51 et l'asparagine 125 ne sont pas retrouvées. La sulfonamide et le groupement hydroxyle présents au niveau du linker du composé 12 permettent de mimer parfaitement les deux groupements phosphates de la CDP-choline, en interagissant avec la chaîne principale de la valine 40 et la lysine 78 (correspondant au premier groupement phosphate de la CDP-choline), et l'histidine 124 (correspondant au deuxième groupement phosphate de la CDP-choline). Au niveau de la poche choline, le cyclohexyle établit des liaisons de type van der Waals avec le tryptophane 107.

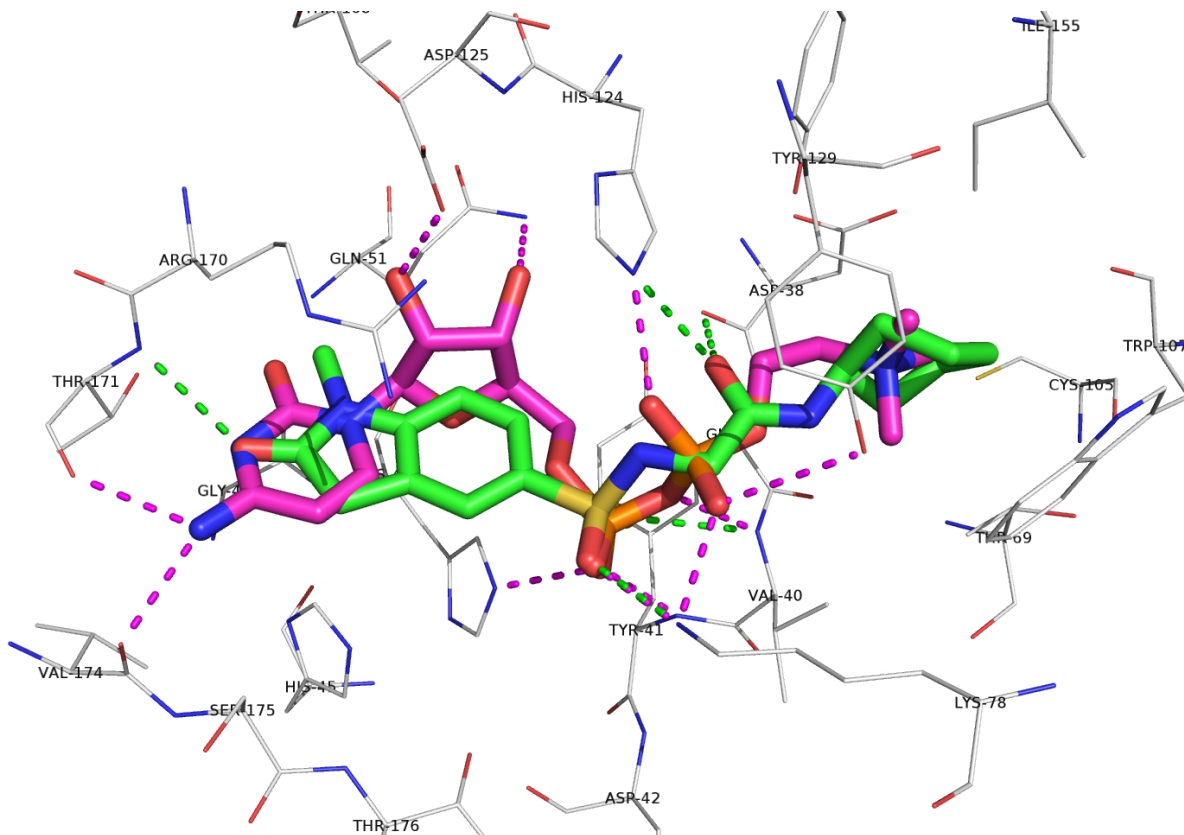


Figure 74 - Position du composé 12 (vert) obtenue par amarrage. Les interactions hydrogènes que ce composé établit avec les résidus du site actif de la *PfCCT*-Nter sont représentées en pointillés.

Pour le **composé 18**, la présence d'un groupement différent sur le noyau dihydroindole induit une pose de docking différente de celle du composé 12. Cependant, l'interaction avec la chaîne principale de la thréonine 171 est conservée. Au niveau du premier phosphate, on observe des interactions hydrogènes du groupement sulfonamide avec la lysine 78 et les chaînes principales de la valine 40 et la tyrosine 41. Puis au niveau du deuxième phosphate, le groupement hydroxyle interagit avec l'histidine 124.

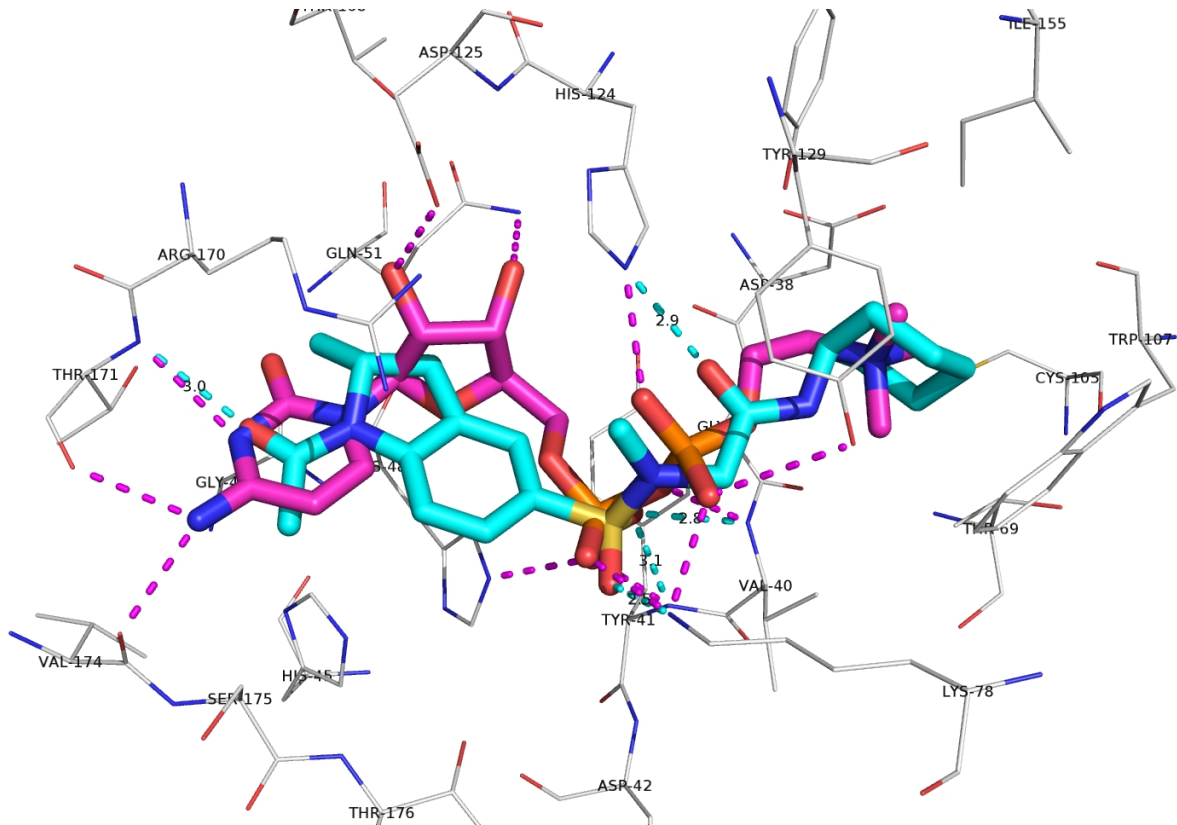


Figure 75 - Position du composé 18 (bleu) obtenue par amarrage. Les interactions hydrogènes que ce composé établit avec les résidus du site actif de la *PfCCT-Nter* sont représentées en pointillés. La CDP-choline est en rose.

Le point vers lequel il faut maintenant se tourner pour améliorer l'affinité est probablement le côté cytidyl pour essayer d'aller récupérer des interactions avec les résidus GLN51 et ASP125. Il faut aussi s'assurer que le linker soit capable de mimer au mieux les deux phosphates présents dans la CDP-choline.

VI. La troisième enzyme de la voie de Kennedy, la *Pf*CEPT

1. Constat de départ

Il est probable que les molécules sélectionnées par le criblage agissent sur la *Pf*CEPT. En effet, les composés sélectionnés pour les tests *in vitro* ont été choisis pour leur similarité avec le produit de la réaction de la *Pf*CCT, la CDP-choline. Mais la CDP-choline est aussi le substrat de la réaction enzymatique suivante dans la voie Kennedy, catalysée par la *Pf*CEPT.

De plus, l'observation de la forte différence d'activité des composés sur la *Pf*CCT recombinante ou endogène ($IC_{50}(\text{composé } 12) : 826 \mu\text{M} \pm 88 \mu\text{M}$ et $1304 \mu\text{M}$, figure 65) par rapport à l'activité sur le parasite entier ($IC_{50}(\text{composé } 12) : 37 \pm 19 \mu\text{M}$, figure 67) nous a conforté dans l'hypothèse de cette possible inhibition. La *Pf*CEPT est la troisième enzyme de la voie *de novo* Kennedy, la voie de synthèse *de novo* de la phosphatidylcholine PC, qui a pour deuxième substrat le diacylglycerol DAG. Elle produit la phosphatidylcholine (Figure 15).

Nos collaborateurs ont montré de façon préliminaire que la *Pf*CEPT est inhibée par le composé 12 avec une IC_{50} de $21 \mu\text{M}$, qui est du même ordre de grandeur que l' IC_{50} du composé 12 sur la croissance parasitaire ($37 \pm 19 \mu\text{M}$).

Pour compléter mon travail, j'ai fait une prédiction de la structure de cette protéine afin de modéliser le site actif. Pour cela, une méthodologie était à suivre et est expliquée ci-dessous dans la figure 76.

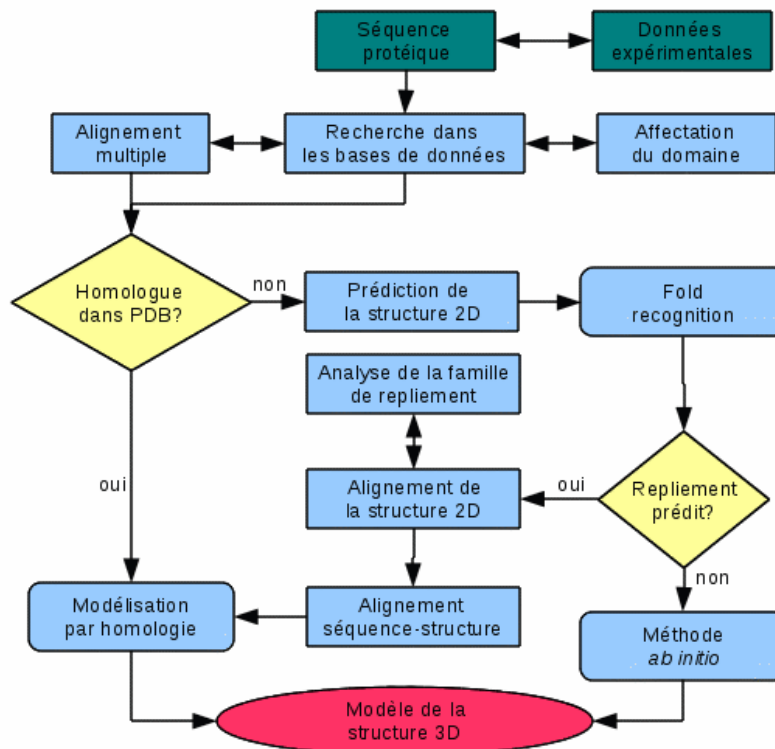


Figure 76 - Méthodologie pour prédire la structure tridimensionnelle d'une protéine.

2. Données récoltées dans les bases de données en ligne

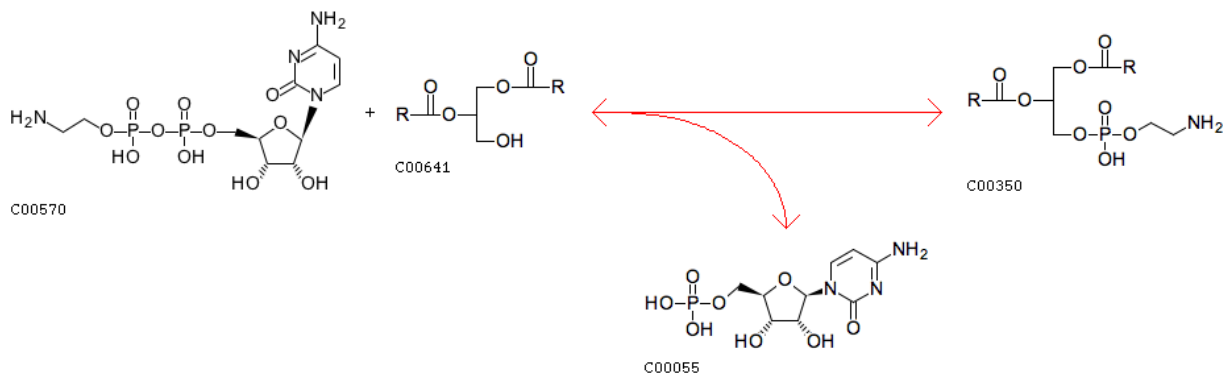
1) Fonctions de la *Pf*CEPT

Lorsque l'on recherche cette protéine dans la base de données Uniprot, nous la trouvons sous la dénomination "plasmodium falciparum ethanolaminophosphotransferase". Son identifiant Uniprot est le C6KTB9. La référence de cette séquence protéique dans la base NCBI est XP_966266.2. La référence de la séquence nucléique associée dans NCBI est XM_961173.2. Le nom du gène est PFF1375c. Elle est identifiée comme une potentielle éthanolaminophosphotransférase, catalysant la réaction EC 2.7.8.1 (CDP-ethanolamine + 1,2-diacylglycerol \rightleftharpoons CMP + phosphatidylethanolamine). L'annotation "Inferred by homology" nous indique que l'existence de cette protéine en tant que CEPT est probable car il existe des orthologues dans des espèces proches.

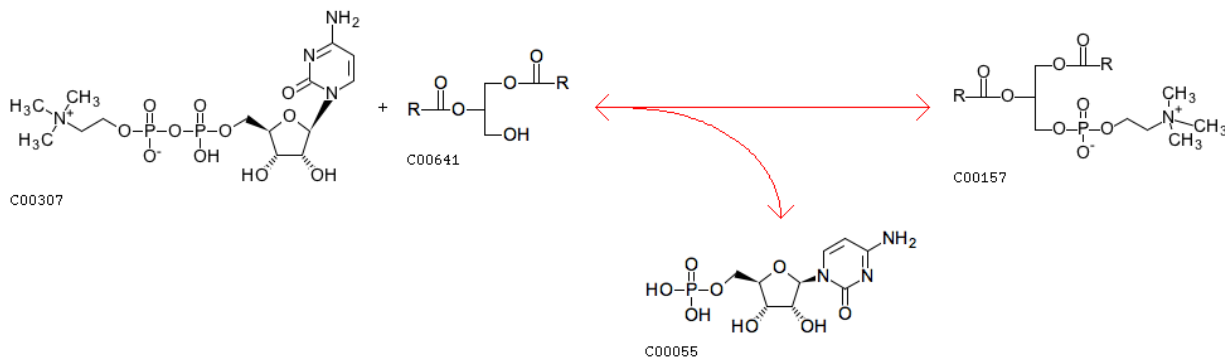
```
>tr|C6KTB9|C6KTB9_PLAF7 Ethanolaminophosphotransferase, putative
OS=Plasmodium falciparum (isolate 3D7) GN=PFF1375c PE=3 SV=1
MGILFKLNKSVYSNCKSYVYKSGGNSLLDNLDFAFWNICVKFVPKSITPNLLTLLGFLCS
TIAFVLTFVFDSQNKKYDYIYIYAGIFLFIYQTFDALDGKQARKTNTSSPLGQLFDHGCD
SITTSFFVFIACKAAGFPKSLIYYILLAIVQIQGYMFSWMEYHTKVFNTSVGKIGTTESH
VLVITMCILRGLKGAGVFQRTTLKDILPKSISSVLCKCVLSVTMNYVILIPVIFFIILSI
SRCTYIGMQNAKKKSEAIMQLAVVYIFSLMQYYFYTNVTPKTELICFSIALYSCFYN
LFINLSTILKIKMDLIPVPVIFYYSILLIFLKRNTKYKFLKHPLLKDSYILYYILFFGI
IYLLDYAHTIITNICKELNITFIFNKYK
```

Figure 77 - Séquence de la protéine *Pf*CEPT, contenant 391 acides aminés. Le domaine signature des phosphotransférases est mis en évidence.

Cette protéine a une activité enzymatique double (Vial et al., 1984). Sur la base de données PlasmoDB, son identifiant est PF3D7_0628300. Cette enzyme est identifiée comme une potentielle choline/ethanolaminophosphotransferase (CEPT). Elle a donc aussi une activité cholinephosphotransferase EC 2.7.8.2 (CDP-choline + 1,2-diacylglycerol \rightleftharpoons CMP + phosphatidylcholine). Ses valeurs caractéristiques pour la voie de la choline ont d'ailleurs été déterminées : $K_m(\text{CDP-choline})=13.2\pm 0.7\mu\text{M}$ et $V_{\text{max}}(\text{CDP-choline})=17.0\pm 1.5\text{pmol./min/107 cells}$ (Wein et al., 2012). Les réactions effectuées par la *Pf*CEPT sont présentées dans la figure 78.



R02057 EC 2.7.8.1



R01321 2.7.8.2

Figure 78 - Réactions catalysées par la *Pf*CEPT.

Selon Prosite, elle est prédite comme appartenant à la famille des CDP-alcool phosphatidyltransférase de classe I, car elle contient un domaine signature PS00379 du résidu **98 à 120** : **D-G-x(2)-A-R-x(7,8)-G-x(3)-D-x(3)-D** représenté en bleu dans la figure 77.

Un nombre important de phosphatidyltransférases partagent une séquence conservée. Ces enzymes sont impliquées dans la biosynthèse des phospholipides. Elles partagent la faculté de catalyser le déplacement d'un CMP à partir d'un CDP-alcool par un second alcool avec la formation d'une liaison phosphodiester et la rupture concomitante d'un lien anhydre phosphoride. Ces enzymes sont des protéines d'une taille variant de 200 à 400 acides aminés. La séquence conservée contient trois acides aspartiques (D98, D116 et D120) et est localisée dans la partie **N-terminale** de la protéine. L'identifiant InterPro de ce domaine est **IPR000462**. L'identifiant Pfam est **Pf01066**. Cette séquence peut être celle constituant le **site actif**.

La PfCEPT est reliée à trois termes de Gene Ontology :

- GO:0008654 "phospholipid biosynthetic process" : Les réactions chimiques et les voies métaboliques résultant en la formation de phospholipides, c'est-à-dire tout lipide contenant un acide phosphorique et un mono ou diester.

- GO:0016020 "membrane" : Double couche lipidique qui entoure les cellules, et les organelles chez les eucaryotes; la bicouche peut être simple ou double et contient des protéines associées.

- GO0004307 "ethanolaminephosphotransferase activity" : Catalyse de la réaction CDP-éthanolamine + 1,2-diacylglycérol → CMP + phosphatidyléthanolamine

En résumé, c'est une protéine impliquée dans la voie de biosynthèse des phospholipides (les voies de la phosphatidyléthanolamine et de la phosphatidylcholine), elle est ancrée à la membrane et a une activité de phosphotransférase.

Phosphatidylethanolamine and phosphatidylserine metabolism

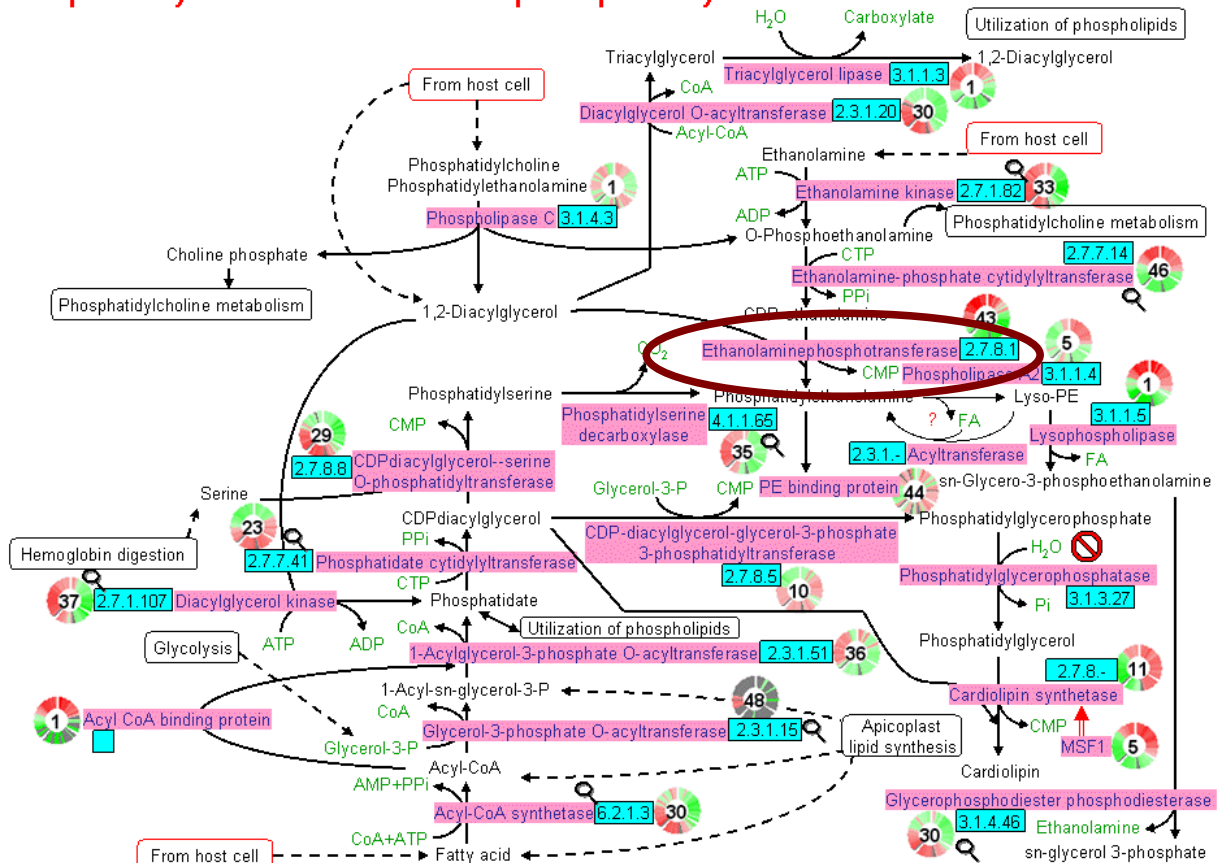


Figure 79 - Métabolisme de la phosphatidyléthanolamine.

(<http://priweb.cc.huji.ac.il/malaria/>)

Phosphatidylcholine metabolism

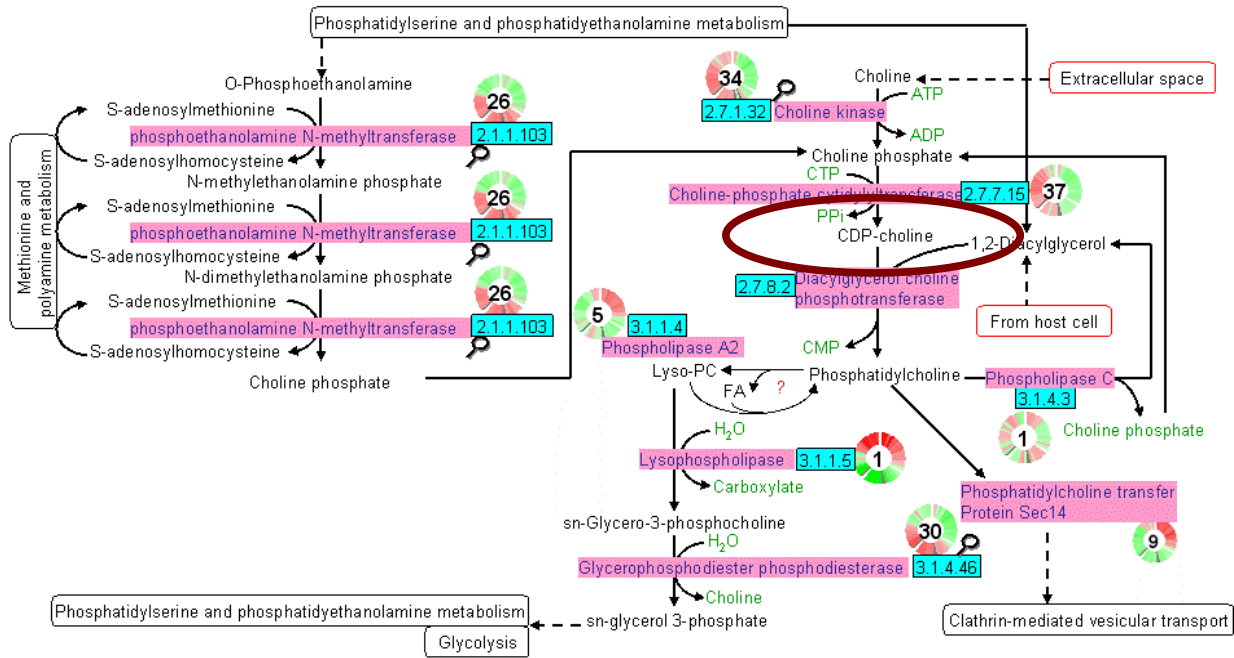


Figure 80 - Métabolisme de la phosphatidylcholine.

(<http://priweb.cc.huji.ac.il/malaria/>)

Il n'y a pas de structure actuellement disponible chez d'autres organismes. Cependant, à la vue des informations récoltées sur la page d'Uniprot (notamment le domaine signature des phosphatidyltransferase de classe I, en bleu dans la Figure 81, qui doit se trouver côté cytoplasmique pour avoir accès aux substrats CDP-choline et DAG) et en utilisant divers outils en ligne, nous avons obtenu des informations quant à son repliement. Nous allons donc pouvoir proposer une topologie membranaire pour cette protéine de *Plasmodium falciparum*.

2) Prédiction de la structure secondaire de la PfCEPT

Afin de prédire la structure secondaire de cette enzyme, j'ai utilisé un serveur en ligne : PSIPRED (Predict Secondary Structure) dans sa version 3.0. C'est une méthode de prédiction de structure secondaire basée sur deux réseaux de neurones sans rétroaction. Elle réalise une analyse sur le résultat d'un PSI-BLAST (Position Specific Iterated – BLAST). La PfCEPT est une protéine composée essentiellement d'hélices alpha.

```
# PSIPRED HFORMAT (PSIPRED V3.0)
Conf: 976202697565021256413479885045555999998741468777302667777899
Pred: CCCCCCCCHHHHCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHH
AA: MGILFKLNKSVYSNCKSYVYKSGGNSLLDNLFDAFWNICVKFVPKSITPNLLTLLGFLCS
      10      20      30      40      50      60
Conf: 99999998623788874179999999999987206401122112899973023321010
Pred: HHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCHHHHHHCCCCCCHHHHHHHCC
AA: TIAFVLTFVFD SQNKKYDYIYIYAGIFLFIYQTFDALDGKQARKTNTSSPLGQLFDHGCD
      70      80      90      100     110     120
Conf: 135689999999971999728999999999878999999996465676057766954899
Pred: HHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCEEEECCCCCHHHH
AA: SITTSFFVFIACKAAGFPKSLIYYILLAIVQIQGYMFSWMEYHTKVFNTSVGKIGTTESH
      130     140     150     160     170     180
Conf: 999999998763154333222112223321100122233201111345588788787541
Pred: HHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHC
AA: VLVITMCILRGLKGAGVFQRTTLKDILPKSISSVLCKCVLSVTMNYVILIPVIFFIILSI
      190     200     210     220     230     240
Conf: 211465443144541003467799999999998885057899840368789999999998
Pred: CEEEEEEEECCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHH
AA: SRCTYIGMQNAKKKSEAIMQLAVVYIFSLMQYFYFYTNTVTPKTELCFSIIALYSCFYN
      250     260     270     280     290     300
Conf: 999988665310258896247899999999987765320059884508999999999999
Pred: HHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHHHH
AA: LFINLSTILKIKMDLIPVPVIFYLSILLIFLKRNTKYKFLKHPLLKDSYILYYILFFGI
      310     320     330     340     350     360
Conf: 9988838899999821107222540698899
Pred: HHHHHHHHHHHHHHHHHHCCCEEECCCCC
AA: IYLLDYAHTIITNICKELNITFIFNKYK
      370     380     390
```

Figure 81 - Prédiction de la structure secondaire de la protéine PfCEPT par PSI-PRED. Jaune : hélice, rose : feuillet, bleu : domaine signature des CDP-alcool phosphatidyltransférases.

3) Prédiction de la localisation des domaines

transmembranaires

Selon l'annotation GO, cette protéine se trouve liée à la membrane. De plus, on observe un nombre important d'hélices alpha, qui pourraient venir s'ancrer à la membrane, et constituer des segments transmembranaires. Les hélices transmembranaires font habituellement autour de 20 acides aminés de long. L'identification de ces hélices transmembranaires peut se faire sur la base des échelles d'hydrophobicité. Comme l'intérieur de la couche lipidique est hydrophobe, il semble nécessaire que les acides aminés traversant la membrane soient aussi hydrophobes. Cependant, les pompes membranaires et les canaux ioniques contiennent aussi des résidues chargés et polaires au sein des segments transmembranaires généralement non-polaires.

L'utilisation de l'hydrophobicité pour prédire les hélices transmembranaires permet de proposer une topologie transmembranaire d'une protéine, c'est-à-dire de savoir quelles régions sont dans la cellule, quelles régions sont en dehors de la cellule et combien de fois la chaîne protéique traverse la membrane. Le profil d'hydrophobicité de la protéine *PfCEPT* montre une succession de zones hydrophobes et hydrophiles, ce qui peut déjà faire penser à des segments transmembranaires (Figure 82). Ces prédictions sont à prendre avec beaucoup de précautions. Cependant, couplé aux autres données récupérées, on est capable de proposer une topologie de la *PfCEPT*. Pour cela, j'ai utilisé neuf serveurs différents pour trouver un consensus sur la prédiction.

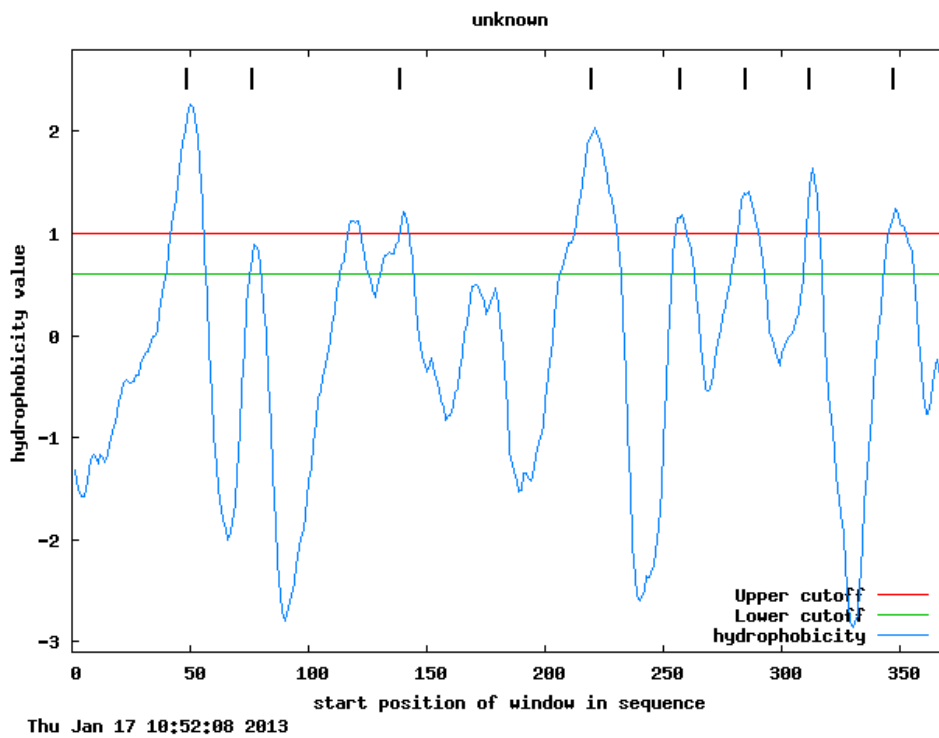
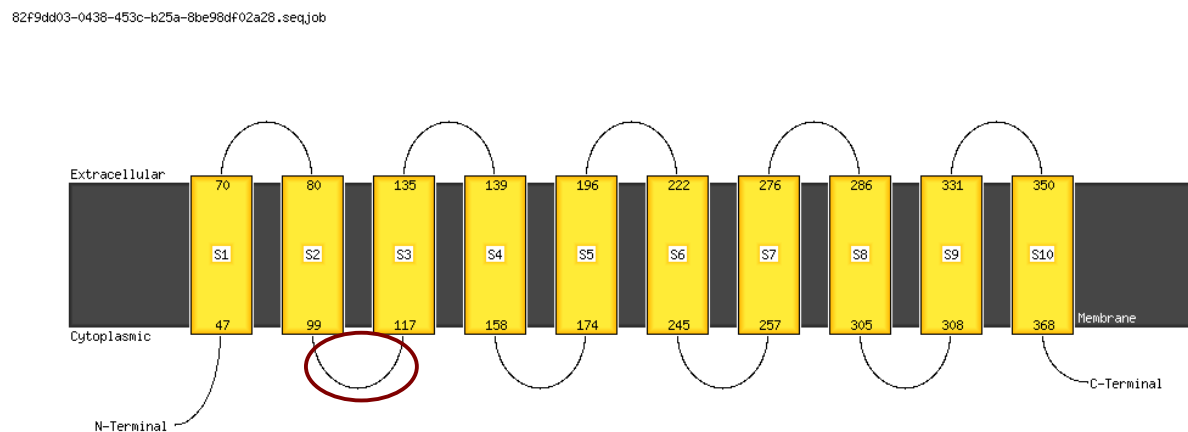
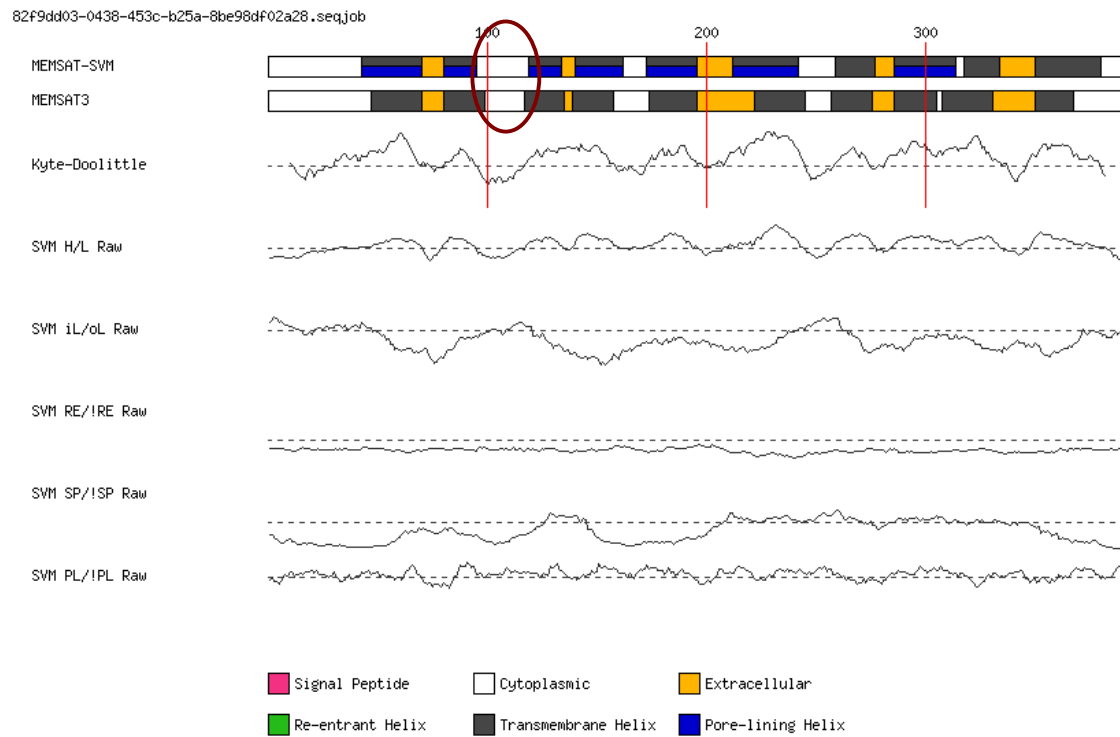


Figure 82 - Profil d'hydrophobicité de la protéine *PfCEPT* provenant du serveur TopPred (Von Heijne, 1992) <http://mobyli.pasteur.fr/cgi-bin/portal.py?#forms::toppred>

a) Résultats des prédictions par MEMSAT3 et MEMSAT-SVM

Sur le serveur PSI-PRED (<http://bioinf.cs.ucl.ac.uk/psipred/>, (Nugent & Jones, 2009)), il existe des outils spécifiques pour prédire les segments transmembranaires : MEMSAT3 (Transmembrane Topology Prediction) et MEMSAT-SVM (Transmembrane helix prediction).

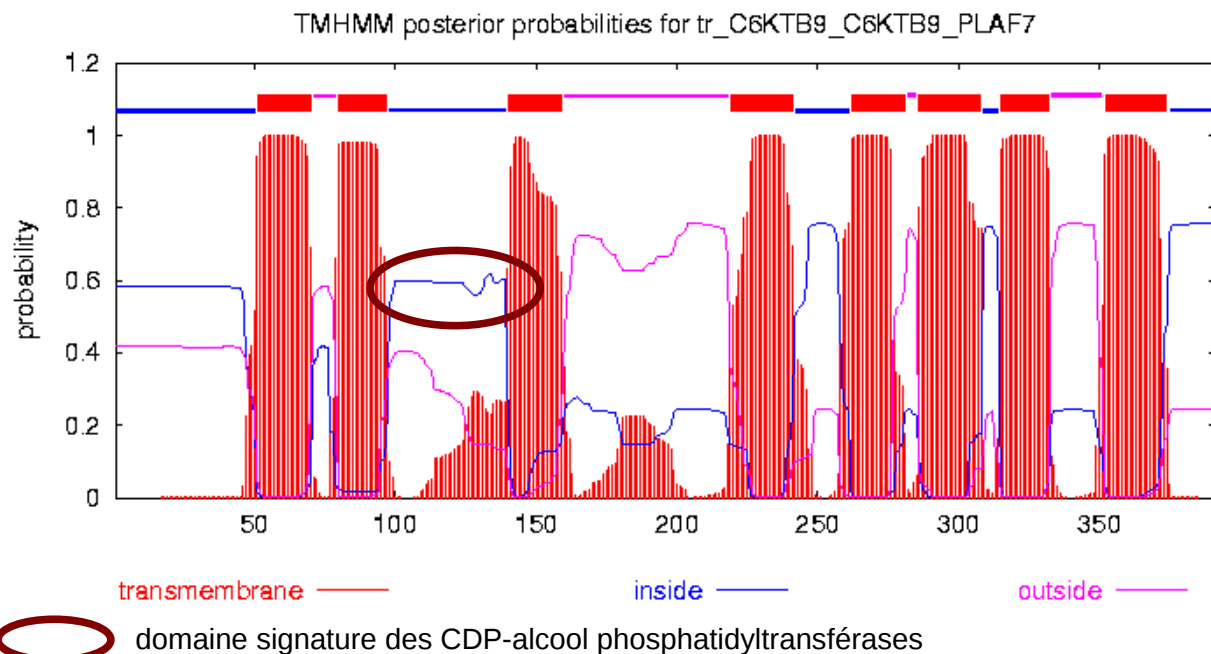


Le domaine signature, encerclé en noir, se situe du côté cytoplasmique. Ces serveurs prédisent 10 segments transmembranaires.

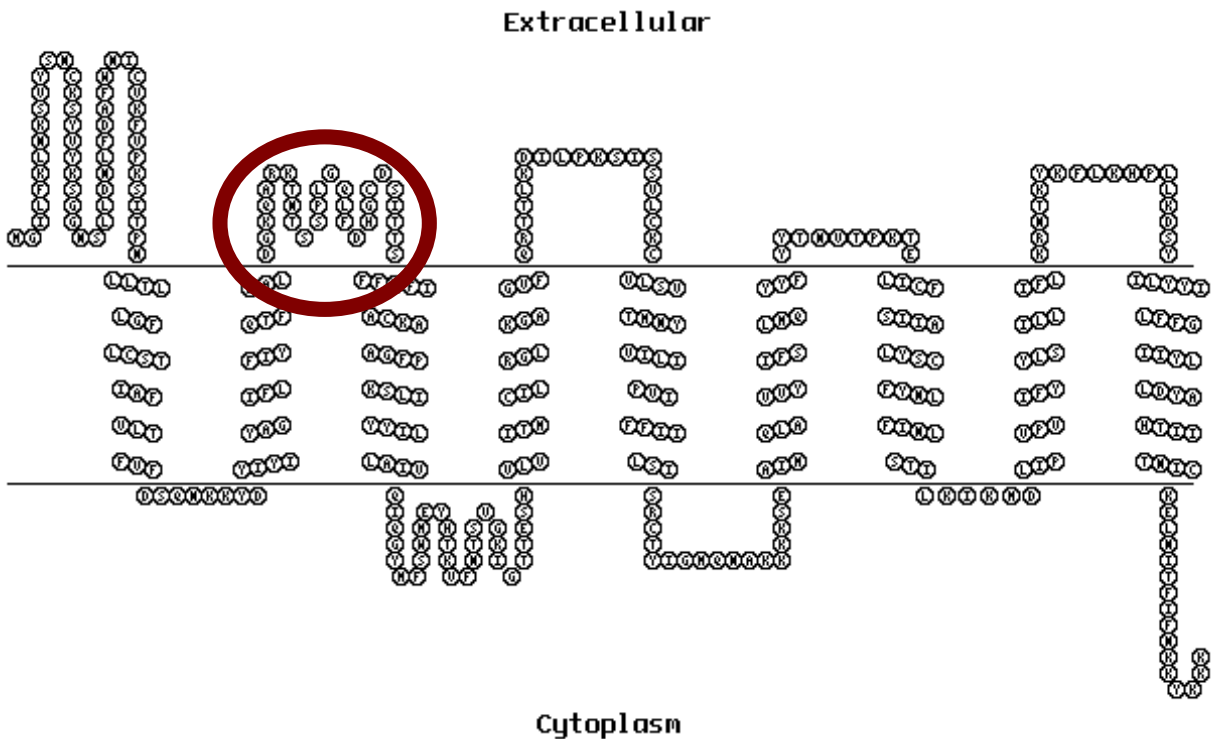
b) Résultats des prédictions par TMHMM

Le serveur TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) prédit 8 régions transmembranaires, aux endroits précisés dans la table suivante. La zone contenant la séquence conservée des phosphotransférases devrait plutôt se situer dans la partie interne de la cellule. La probabilité que cette zone, en rouge, soit à l'intérieur est plus importante (environ 60%).

tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	inside	1	50
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	51	70
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	outside	71	79
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	80	97
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	inside	98	139
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	140	159
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	outside	160	218
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	219	241
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	inside	242	261
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	262	281
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	outside	282	285
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	286	308
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	inside	309	314
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	315	332
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	outside	333	351
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	TMhelix	352	374
tr_C6KTB9_C6KTB9_PLAF7	TMHMM2.0	inside	375	391



e) Résultats des prédictions par TOPO2



Cette prédiction par le serveur TOPO2 (<http://www.sacs.ucsf.edu/TOPO2/>) est écartée car la séquence conservée doit se trouver côté cytoplasmique.

f) Résultats des prédictions par PRED-TMR

Ce serveur de prédiction des régions transmembranaires dans les protéines PRED-TMR (<http://athina.biol.uoa.gr/PRED-TMR/>) prédit 8 segments transmembranaires. Il n'y a pas d'informations vis-à-vis de la position intra ou extracellulaire des régions entre les hélices.

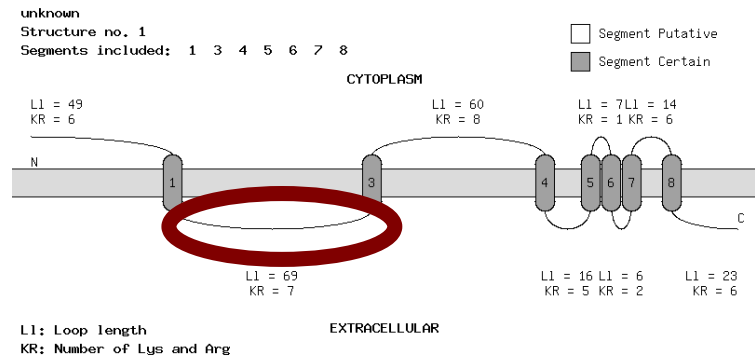
	1	2	3	4	5	6
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
	.		.		.	
0000	MGILFKLNKSVYSNCKSYVYKSGGNSLLDNLFDAFWNICVKFVPKSITPN	LLTLLGFLCS				
0060	TI A FVLT FV FDSQNKKYD YIYIYAGIFLFIYQTFD ALDGKQARKTNTSSPLGQLFDHGCD					
0120	SITTSFFVFIACKAAGFPK LIYYILLAI V QIQGYMFSW MEYHTKVFNTSVGKIGTTESH					
0180	VLVITMCILRGLKGAGVFQRTTLKDILPKSISVLCCK VLSVTMNYVILIPVIFFI ILSI					
0240	SRCTYIGMQNAKKKSE AIMQLAVVYIFSLMQYFY YTNVTPKTEL LICFSIALYSCFY N					
0300	LFINL STILKIKMD LIPVPVIFYLSILLIFL KRNTKYKFLKHPLLKDSY ILYYILFFGI					
0360	IYLLDYAHTII TNICKELNITFIFNKKYKKK					

(en bleu, régions prédites comme transmembranaires)

**g) Résultats des prédictions par TopPred sur Mobyle@Pasteur :
prédiction des topologies des protéines membranaires**

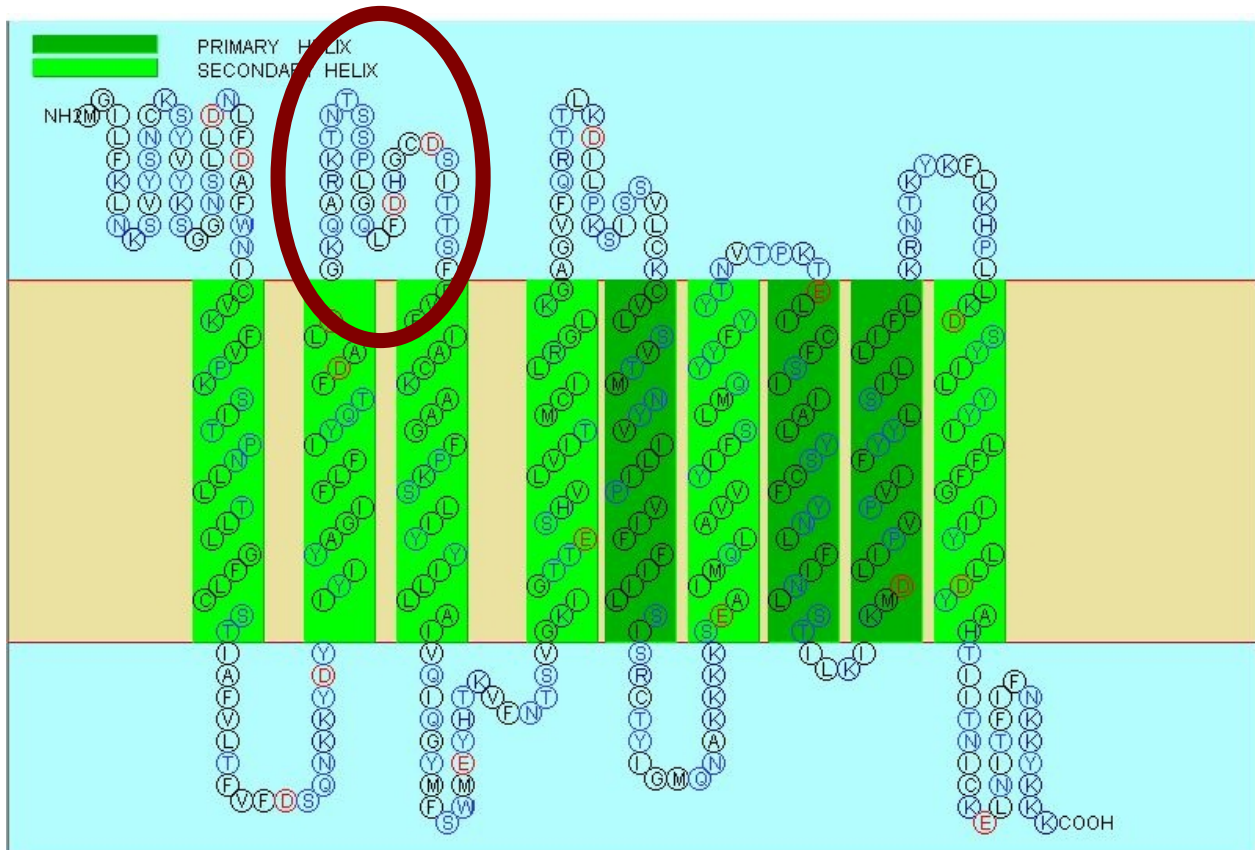
Deux possibilités sont proposées. La première est plus satisfaisante car elle positionne la séquence conservée du côté intracellulaire. Elle propose 8 hélices transmembranaires.

Helix	Begin - End	Score	Certainty
1	50 - 70	2.265	Certain
2	77 - 97	0.901	Putative
3	140 - 160	1.221	Certain
4	221 - 241	2.039	Certain
5	258 - 278	1.188	Certain
6	286 - 306	1.422	Certain
7	313 - 333	1.639	Certain
8	348 - 368	1.238	Certain



h) Résultats des prédictions par SOSUI

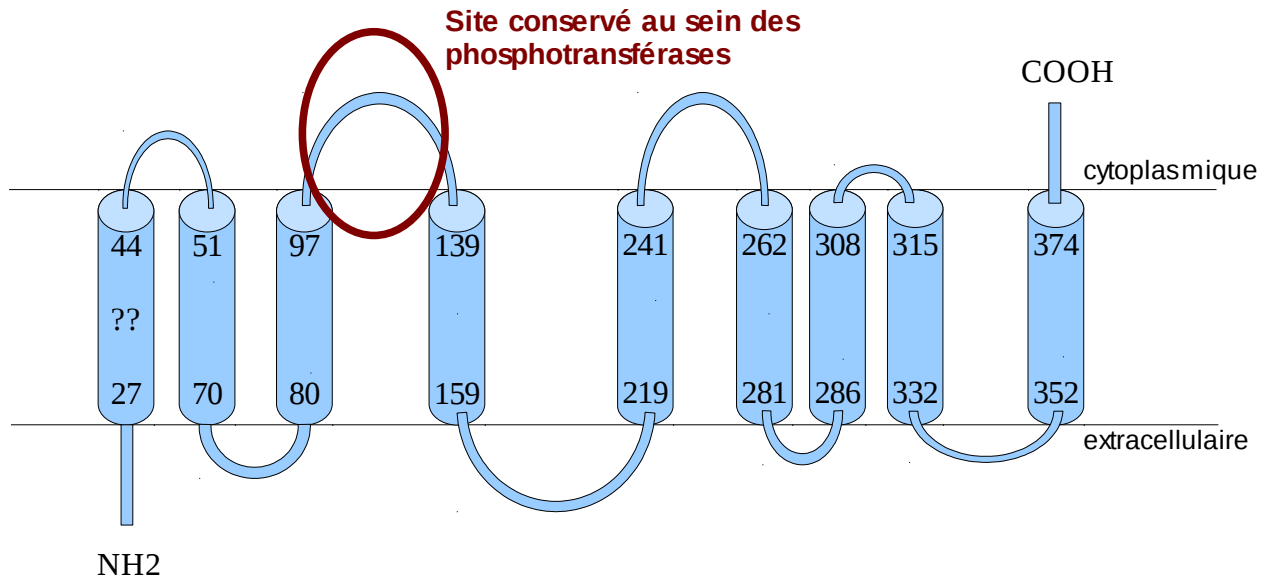
Cette protéine est prédite par ce serveur (<http://bp.nuap.nagoya-u.ac.jp/sosui/>) avec 9 hélices transmembranaires. L'orientation par rapport au cytoplasme n'est pas spécifié.



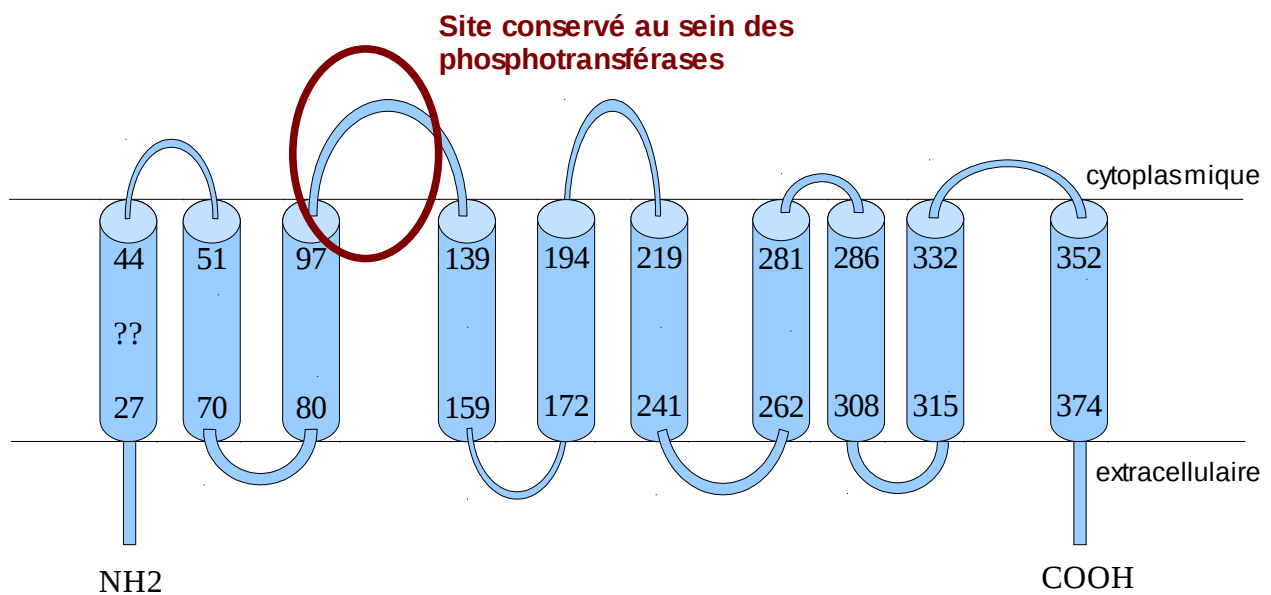
4) L'analyse des résultats des différentes méthodes de prédiction

Je propose deux topologies possibles pour la PfCEPT. La première contient 8 segments transmembranaires et la seconde 9. Un autre segment transmembranaire, le premier, est putatif. Il n'a été trouvé qu'une fois parmi les différents serveurs (HMMTOP). Le N-terminal serait donc plutôt du côté intracellulaire. Le segment supplémentaire [172-194] possible a pour effet d'inverser la localisation des boucles suivantes.

a) 8 segments transmembranaires



b) 9 segments transmembranaires



Nous souhaitons maintenant proposer un modèle 3D de la structure de cette protéine.

3. I-Tasser, une plateforme pour la prédiction automatique de la structure par homologie et de la fonction d'une protéine (Roy et al., 2010)

Ce serveur en ligne donne tout d'abord une prédiction de la structure secondaire de cette protéine à partir de sa séquence ainsi qu'une prédiction de l'accessibilité au solvant de chaque résidu. Les résultats se recourent avec ceux obtenus précédemment avec PsiPred.

<http://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S124655/>

Plusieurs modèles ont été construits.

Name	C-score	Exp.TM-Score	Exp.RMSD	No.of decoys	Cluster density
Model1:	-1,80	0.50+-0.15	11.0+-4.6	600	0,1835
Model2:	-1,80			600	0,1835
Model3:	-1,80			600	0,1835
Model4:	-1,80			600	0,1835
Model5:	-2,14			429	0,1312

Figure 83 - Valeurs de score des cinq modèles proposés par I-Tasser.

C-score est un score de confiance pour estimer la qualité des modèles proposés par I-TASSER, il est compris entre -5 et 2. En général, pour avoir un bon repliement, le modèle doit avoir un C-score inférieur à -1,5. Un TM-score supérieur à 0,5 indique que le modèle a une topologie correcte et un TM-score inférieur à 0,17 signifie que la similarité provient du hasard. Les 10 meilleures structures utilisées comme templates sont alignées à la séquence dont on souhaite la structure, c'est-à-dire la séquence de la PfCEPT. Les valeurs d'identité de séquence que ce soit sur la région alignée ou la séquence entière de la template varient entre 5% et 18%, ce qui est généralement trop faible pour utiliser la modélisation comparative.

Ce serveur propose aussi de prédire quel type d'enzyme est la protéine d'intérêt. On sait que l'on a affaire à une choline/éthanolaminephosphotransférase d'EC 2.7.8.2 ou 2.7.8.1. Aucun de ces nombres n'est proposé dans le top 5. Le C-scoreEC est entre 0.123 et 0.170, ce qui est trop faible.

La prédiction du site actif aussi est erronée. Le C-scoreLB est entre 0,02 et 0,06 alors que le Cscore varie entre 0 et 1 et est meilleur quand il est élevé. De plus, nous savons que la séquence signature de cette famille d'enzyme se situe en N-terminal entre le résidu 98 et 120.

Les modèles et les prédictions proposés par I-Tasser ne peuvent être retenus en raison de l'inexistence de structures expérimentales pour des protéines homologues connues. Lorsque l'on visualise plus précisément les 5 modèles, que l'on regarde le nombre d'hélices "transmembranaires" et que l'on localise la zone signature de la famille des phosphotransférases, nous observons que :

Dans le modèle 1, il y a trop d'hélices par rapport à la prédiction de la structure secondaire. Dans le modèle 2, les hélices semblent trop courtes pour traverser la membrane. Dans le modèle 3, la région conservée qui devrait se situer sur la face interne de la membrane pour avoir accès aux substrats et effectuer la réaction enzymatique se situe au sein de la protéine. Dans le modèle 4, les hélices semblent trop courtes pour traverser la membrane. Dans le modèle 5, la longueur des hélices alpha est plus appropriée et la région conservée est accessible, mais il y a un nombre trop important d'hélices alpha par rapport à ce qui a été prédit par les serveurs de prédiction de domaines transmembranaires.

En conclusion, l'utilisation d'une modélisation comparative est inadaptée dans le cas présent.

4. Modélisation *ab initio* avec l'outil Robetta (Chivian et al., 2003)

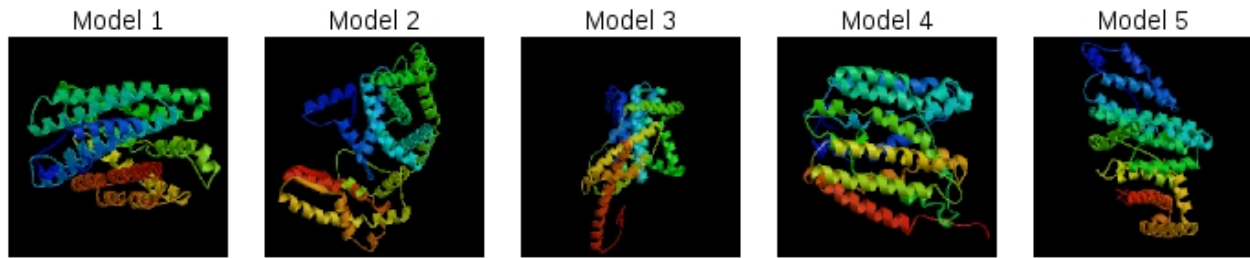
<http://robetta.bakerlab.org/results.jsp?id=35391>

Robetta combine les techniques *ab initio* et par homologie pour réaliser des modèles. Ce serveur utilise la méthode d'insertion de fragments de Rosetta pour construire les domaines pour lesquels on ne trouve pas d'homologues dans la PDB. Les modèles construits par homologie sont réalisés à partir de structures PDB retrouvées par les outils UW-PDB-BLAST ou HHSEARCH et alignées par diverses méthodes comme HHSEARCH, Compass et Promals. Les régions non structurées sont assemblées aux fragments structurés et optimisées pour s'aligner à la structure de la template. La procédure est complètement automatisée.

Deux prédictions ont été obtenues. La première est une prédiction par modélisation comparative avec pour template la protéine Yyip, transporteur de zinc (PDB 3H90) avec une valeur de confiance de 0.038. La deuxième est une prédiction *ab initio* avec deux domaines : le premier des résidus [1-267] a une valeur de confiance de 248. Le deuxième domaine [26 -391] a une valeur de confiance de 5.

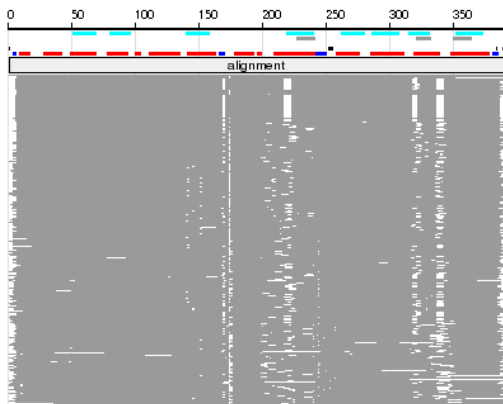
Parmi les modèles proposés, on recherche ceux qui peuvent s'ancrer à la membrane avec 8 ou 9 segments transmembranaires. Aucun ne répond à ces critères. A ce stade, nous n'avons donc pas de représentation 3D de la PfCEPT par manque de données expérimentales sur cette famille de protéines. Nous ne pouvons donc pas envisager de criblage virtuel de type 'structure-based' pour identifier des inhibiteurs ou modéliser le mode de fixation du composé 12.

Full Structure Predictions



Ginzu Domain Prediction 1 (RECOMMENDED: click on domain number for details: alignments, model energies, and MAMMOTH detections) ▲

Domain	Span	Source	Reference Parent	Parent Span	Confidence	Annotations
domain 1	1-391	alignment	3h90A_201	1-283	0.038800	TRANSPORT PROTEIN



Ginzu Domain Prediction 2 (RECOMMENDED: click on domain number for details: alignments, model energies, and MAMMOTH detections) ▲

Domain	Span	Source	Reference Parent	Parent Span	Confidence	Annotations
domain 1	1-260	msa	--	--	248.650997	--
domain 2	261-391	msa	--	--	5.295994	--

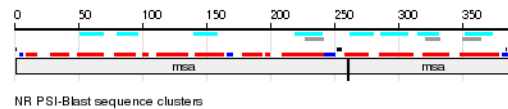
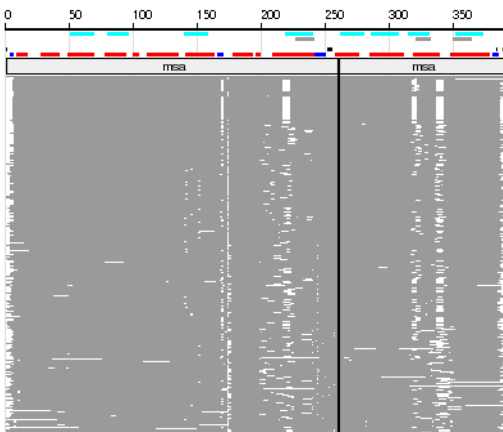


Figure 84 - Résultats de Robetta.

En conclusion, l'utilisation des différents serveurs a permis de proposer des topologies 2D réalistes, en concordance avec les données biologiques. Malheureusement, il faudra attendre la publication d'une structure d'un homologue ou de la protéine chez *Plasmodium falciparum* pour confirmer nos prédictions de la structure 3D de la PfCEPT.

Discussion

Le contrôle du paludisme à *Plasmodium falciparum* nécessite l'identification de nouvelles molécules candidates agissant sur des nouvelles cibles thérapeutiques. Deux équipes du laboratoire DIMNP de l'Université de Montpellier ont identifié des cibles au cours de leurs études sur les Apicomplexes. La première est un complexe protéique, *PfAMA1-PfRON2*, essentiel à l'invasion des érythrocytes par les mérozoïtes. La deuxième est l'enzyme *PfCCT* impliquée dans la biosynthèse de la phosphatidylcholine, phospholipide majoritaire de la membrane des parasites. Dans ces deux cas, c'est le stade érythrocytaire du parasite qui est ciblé.

Mon travail de thèse s'est donc porté sur ces deux nouvelles cibles que l'on considère thérapeutiques dans le traitement du paludisme à *Plasmodium falciparum*. Ces deux projets partagent la caractéristique d'être associés à des criblages expérimentaux à bas débit. Seules quelques dizaines de molécules pouvaient être testées au cours de ma thèse. Pour ce faire, j'ai intégré les données structurales disponibles dans des protocoles de criblage virtuel adaptés à chaque projet. Ces deux études ont été menées en collaboration avec les équipes du DIMNP de Montpellier effectuant les tests expérimentaux. Parmi les composés issus des criblages virtuels, il a été possible d'identifier des premiers candidats inhibiteurs, qu'il va maintenant falloir optimiser.

Le projet *PfAMA1* est un projet d'inhibition d'interface entre deux protéines. Ce type de projet est considéré comme ambitieux (Mullard, 2012) mais pas impossible comme le montre les résultats sur les protéines de la famille Bcl (Petros et al., 2010). *PfAMA1* possède un sillon hydrophobe suffisamment enfoui pour être considéré comme 'drug-able'. A partir de structures expérimentales et des expériences de liaison avec des peptides, j'ai pu dériver un pharmacophore (filtre LBVS). J'ai aussi utilisé la structure de la protéine lors de la deuxième étape de filtre (filtre SBVS). Le protocole de criblage virtuel m'a permis de sélectionner neuf molécules. Sept d'entre elles ont été testées sur la croissance parasitaire et sur l'invasion. Les neuf molécules ont aussi été soumises à une expérience par la technique du SPR. Le composé 3 a montré une activité dans ces trois expériences. Sa liaison au sillon hydrophobe de *PfAMA1* va être testée par une expérience de polarisation de fluorescence. Si le composé 3 se fixe effectivement, des expériences de cocristallisation pourront être menées par notre collaborateur canadien.

En résumé, à partir de la structure de la protéine *PfAMA1* en complexe avec son ligand naturel, la partie exposée de *PfRON2*, ou avec le peptide R1, il a été possible parmi près de 30 millions de molécules, d'identifier des molécules montrant une activité sur la croissance parasitaire de *Plasmodium falciparum*. Cela a été possible grâce à un protocole de criblage virtuel intégrant les données structurales disponibles.

Le projet *PfCCT* est un projet de 'drug discovery' plus classique puisqu'il s'agit d'une enzyme possédant un site actif enfoui. Ici, l'étape de filtre basé sur le ligand est une combinaison de diverses méthodes et utilise la CDP-choline comme référence (filtre LBVS). L'étape de filtre basée sur la structure (SBVS), l'amarrage moléculaire, a été effectuée sur un modèle de la *PfCCT* que j'ai réalisé. La position du seul résidu qui diffère entre *Rattus norvegicus* et *Plasmodium falciparum*, la glutamine 51, a été optimisée pour reproduire la conformation du produit de la réaction, la CDP-choline, cocrystallisée dans le site actif de la *rCCT*.

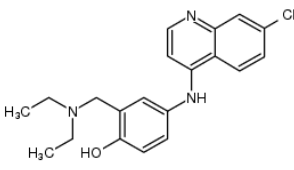
Les trois molécules qui ont montré une activité sur la *PfCCT* ($IC_{50} < 1\text{mM}$) miment le groupement cytidyl de la CDP-choline que de façon partielle. Il faudrait donc modifier ces molécules de façon à récupérer des interactions hydrogènes importantes avec la chaîne latérale de la valine 174, ainsi qu'avec les résidus asparagine 125 et glutamine 51 (résidu spécifique à *Plasmodium falciparum*). Au niveau du premier phosphate, il serait intéressant de pouvoir interagir avec l'histidine 48.

En parallèle, les molécules candidates sont actuellement testées sur deux lignées cellulaires humaines pour évaluer leur toxicité, ainsi que sur la *PfCEPT*, la troisième enzyme de la voie *de novo* Kennedy. On s'attend à ce que ces molécules se fixent plus ou moins fortement aux protéines *hCCT* et *PfCEPT*. En effet, la *CCT* humaine est très similaire à la *PfCCT* et la *PfCEPT* fixe le produit de la réaction de la *PfCCT*, que l'on mime avec nos inhibiteurs. D'un point de vue pharmacologique, la fixation des inhibiteurs à la *hCCT* sera liée au volume de distribution (VD) de la molécule, qui correspond au rapport de la quantité totale de médicament dans l'organisme sur la concentration plasmatique du médicament. Plus son VD sera faible, plus l'inhibiteur sera concentré dans le compartiment sanguin et moins il se diffusera dans les tissus de l'hôte. Ce type de paramètre pharmacocinétique n'est pas modélisable à ce stade du projet. L'inhibition simultanée de la *PfCEPT* peut-être par contre un avantage pour renforcer le blocage de la biosynthèse des phospholipides. D'autre part, dès que la structure 3D de la *PfCEPT* sera résolue, il sera possible d'utiliser ces données pour optimiser les composés ayant montré une activité sur cette protéine. Il sera aussi possible de mettre au point un nouveau protocole de criblage virtuel pour cibler la voie *de novo* Kennedy via cette enzyme.

En résumé, à partir du produit de la réaction (CDP-choline) et d'un modèle de la structure de la protéine cible (*PfCCT*), il a été possible parmi près de 8 millions de molécules, d'identifier des molécules montrant une activité sur la protéine recombinante et/ou sur la croissance parasitaire de *Plasmodium falciparum*. Cela a été possible grâce à un protocole de criblage virtuel intégrant les données disponibles, essentiellement la structure de la CCT de *Rattus norvegicus* en complexe avec la CDP-choline. Nos collaborateurs montpelliérains et hongrois travaillent à la résolution de la structure du domaine catalytique de la *PfCCT* par RMN et/ou diffraction aux rayons X.

Ces différentes études illustrent l'aide que peut apporter la bioinformatique structurale et la chémoinformatique. Outre les informations déduites des modèles, la sélection rationnelle de ligands assistée par ordinateur permet de suppléer aux limites du criblage expérimental, surtout lorsqu'il est à bas débit en donnant priorité à un nombre raisonnable de molécules. L'élément déterminant est l'intégration des données expérimentales et structurales permettant de définir les interactions essentielles entre la protéine et ses ligands. C'est cela qui donne de la valeur à la recherche *in silico* et qui permet de sélectionner des molécules actives. Les touches identifiées sont des points de départ importants dans ce type de projets car ils vont soit servir de sondes dans l'étude de la protéine dans son contexte biologique, soit ouvrir une piste pharmacologique et aboutir à un candidat médicament.

Annexes

<input type="button" value="Select all"/> <input type="button" value="Unselect all"/> <input type="button" value="Subselect"/>		<input type="button" value="Pharmacology"/> <input type="button" value="FDA Registration"/> <input type="button" value="Chemistry"/> <input type="button" value="Properties"/> <input type="button" value="Similarities"/> <input type="button" value="Most similar drugs"/> <input type="button" value="Fragments"/>				
1 <input checked="" type="checkbox"/>	INN	Class	Route (list)	Target	PDB	Links
AMODIAQUINE DISCONTINUED 	AMODIAQUINE HYDROCHLORIDE	ATC P01BA06 ANTIPROTOZOAL AMEBICIDES ANTI-INFLAMMATORY AGENTS, NON- STEROIDAL ANTIMALARIALS	ORAL	CYTOCHROME P450 2C8 (CYP2C8) CYTOCHROME P450 2D6 (CYP2D6) HEMOGLOBIN SUBUNIT ALPHA HISTAMINE N- METHYLTRANSFERASE		DRUGBANK DB00613 CHEMBL

Annexe 1 - Exemple de l'amodiaquine pour la mention DISCONTINUED.

DB	Nombre de structures	Nombre de molécules
DB_prep	6631	6631
DB_filtered	6100	6100
DB_stereo	20968	6100
DB_3D	20796	6032
DB_FilterVol	16579	5937
DB_taut	37125	5937
DB_taut3D (100conf)	254825	5935
DB_10confCorina	151100	6022
DB_10confCorina_10confOmega	237770	6018
CN	Nombre de structures	Nombre de molécules
CN_prep	43878	43878
CN_filtered	41577	41577
CN_stereo	233418	41577
CN_3D	203838	40518
CN_FilterVol	181610	40267
CN_taut	321034	40267
CN_taut3D (100conf)	2493966	40243
CN_10confCorina	1433722	40479
CN_10confCorina_10confOmega	3369554	40463
CHEMBL	Nombre de structures	Nombre de molécules
CHEMBL_prep	28379	28379
CHEMBL_filtered	28032	28032
CHEMBL_stereo	92435	28032
CHEMBL_3D	70763	26743
CHEMBL_FilterVol	60633	26493
CHEMBL_taut	188867	26493
CHEMBL_taut3D (100conf)	1167303	26449
CHEMBL_10confCorina	403924	26682
CHEMBL_10confCorina_10confOmega	1558659	26674
CB	Nombre de structures	Nombre de molécules
CB_prep	485778	485778

CB_filtered	485642	485642
CB_stereo	724225	485642
CB_3D	689960	481424
CB_FilterVol	689960	481424
CB_taut	1810229	481424
CB_taut3D (100conf)	6060089	481421
CB_10confCorina	2320106	481421
CB_10confCorina_10confOmega	5683251	481196
DL	Nombre de structures	Nombre de molécules
DL_prep	2020026	2020026
DL_filtered	1993321	1993321
DL_stereo	3328406	1993321
DL_3D	3209717	1967868
DL_FilterVol	3209717	1967868
DL_taut	6731905	1967868
DL_taut3D (20conf)	32316253	1967820
DL_10confCorina	15044524	1967820
DL_10confCorina_10confOmega	53207674	1967476
AMB	Nombre de structures	Nombre de molécules
AMB_prep	5053512	5053512
AMB_filtered (MW 200-600)	4935909	4935909
AMB_stereo	7604696	4935909
AMB_3D	7432605	4899728
AMB_taut	19017282	4899728
AMB_taut3D (20conf)	77313514	4899277
AMB_10confCorina	30933329	4899277
AMB_10confCorina_10confOmega	122921055	4895514
CL	Nombre de structures	Nombre de molécules
CL_prep	336085	336085
CL_filtered	336085	336085
CL_stereo	635272	336085
CL_3D	621593	328006
CL_taut	1246430	328006
CL_taut3D (100conf)	10917375	326858
CL_10confCorina	4215430	326863
CL_10confCorina_10confOmega	16746149	326863

Annexe 2 - Les chimiothèques criblées, le nombre de structures et de molécules après chaque étape de la préparation (Préparation commune aux deux projets → DB_prep : préparation des identifiants en tête de chaque molécule, DB_filtered : filtre avec le programme Filter, DB_stereo : génération des stéréoisomères avec le programme flipper, DB_3D: génération d'une conformation 3D avec le programme Omega) (Projet PfCCT → DB_taut : génération des tautomères avec le programme tautomers, DB_taut3D : génération des conformères de cycles avec le programme Corina) (Projet PfAMA1 → DB_10confCorina : génération de 10 conformères de cycle avec le programme Corina, DB_10confCorina_10confOmega: génération de 10 conformères acycles avec le programme Omega).

Identities computed with respect to: (10) PfcCT-Nter|Q8IEE9

Colored by: consensus/60.0% and property

			1 [.	.	.	.	:	.	60
1	saGCT 005155 2B7L	18.3%	-----							
2	bsGCT P27623 1COZ,1N1D	20.5%	-----							
3	bsGCT Q8RKI6	18.6%	-----							
4	PfECT Q8IDM2	13.2%	MSNQFLVDTIYNHGYMRKFLSILRSVKKNDKFQYIMKLCENTNIEEDEELYKVFINELHN							
5	rECT 088637	18.8%	-----M							
6	hECT Q99447 3ELB	20.8%	-----							
7	rCCT P19836 3HL4	33.1%	-----MDAQS							
8	hCCT P49585	22.8%	-----MDAQC							
9	PfCCT-Cter Q8IEE9	50.1%	-----LSEHNESIESSNNDING							
10	PfCCT-Nter Q8IEE9	100.0%	-----MIIKVN							
			61	:	.	120
1	saGCT 005155 2B7L	18.3%	-----							
2	bsGCT P27623 1COZ,1N1D	20.5%	-----							
3	bsGCT Q8RKI6	18.6%	-----							
4	PfECT Q8IDM2	13.2%	ISNTSSTRKKNDSSCNSTNNTGNNKENGSNNTHFEDMKDINSSISNN-SEEINEFEID							
5	rECT 088637	18.8%	IRNGHGAGGAAGLKPGG-----							
6	hECT Q99447 3ELB	20.8%	-----							
7	rCCT P19836 3HL4	33.1%	SAKVNSRKRREKVPNGATEEDGIPSKVQRCVGLRQPAPFSDEIEVDFSKPYVRVTE							
8	hCCT P49585	22.8%	SAKVNSRKRREKVPNGATEEDGIPSKVQRCVGLRQPAPFSDEIEVDFSKPYVRVTE							
9	PfCCT-Cter Q8IEE9	50.1%	KQKEHIKKGNSNQDVPDPTNPAVDDDD-----DDNSNDESEYESSQMDSE							
10	PfCCT-Nter Q8IEE9	100.0%	VQTQKSQDGNKVHNSQDDMDNE-----							
			121	:	.	180
1	saGCT 005155 2B7L	18.3%	-----MKRVITY--GTYD LLHYGH IELLRRAREM-- GDYLIV ALSTDEFNQIKHK-K							
2	bsGCT P27623 1COZ,1N1D	20.5%	-----MKKVITY--GTFD LLHWGH IKLLERAKQL-- GDYLVV AISTDEFNLQKQK-K							
3	bsGCT Q8RKI6	18.6%	-----MKKVITY--GTFD LFHYGH MKLLERAKNL-- GDYLIV GLSTDEFNLQKQK-K							
4	PfECT Q8IDM2	13.2%	SSTSTQEKTKETRIYVDGIFD LSHSGH FNAMRQAKKL-- GDIVV VGINSDEDALNSKGVK							
5	rECT 088637	18.8%	-----QRTVRVWCDGCYD MVHYGH SNQLRQARAM-- GDYLIV GVHTDEEIAKHKG-P							
6	hECT Q99447 3ELB	20.8%	----SMGRRARVWCDGCYD MVHYGH SNQLRQARAM-- GDYLIV GVHTDEEIAKHKG-P							
7	rCCT P19836 3HL4	33.1%	EACRGTPCERPVRVYADGIFD LFHSGH HARALMQAKNLPNTYLIVGVCSDELTHNFKG-F							
8	hCCT P49585	22.8%	EASRGTPCERPVRVYADGIFD LFHSGH HARALMQAKNLPNTYLIVGVCSDELTHNFKG-F							
9	PfCCT-Cter Q8IEE9	50.1%	KNKGSIKNSK NVVI YADGVYD MLHLGH MKQLE QA KKLFENTTLIVGVTS DN ETKLFK G -Q							
10	PfCCT-Nter Q8IEE9	100.0%	-----EK SIRI YADGVYD LLHLGH MKQLE QA KHV DK NVTLIVGV TG DN ET RK FK G-Q							
			181	:	.	240
1	saGCT 005155 2B7L	18.3%	SYDY EQ RKM LESIRY VDL VI PEKG WGQ KED VEK FDV DFV M GHD W-----							
2	bsGCT P27623 1COZ,1N1D	20.5%	AYHSY EH RK LIE TI RY V DE VI PE KN WE QK KQ DI ID H NI D V F M G DD W-----							
3	bsGCT Q8RKI6	18.6%	SHHSY EH RK FIE TI DL V NE VI PE KN WE QK IS DI Q K H D ID TF VI G DD W-----							
4	PfECT Q8IDM2	13.2%	PIYT Q ER G AL IAG CK W DE VI IGTKYN VD MD LL E K YN CD Y AA H G T L AYD-----							
5	rECT 088637	18.8%	PVFT Q ER Y K M V Q A IK W DE V V PA AP Y VT LE T L D K H NC D FC V H G ND IT LT-----							
6	hECT Q99447 3ELB	20.8%	PVFT Q ER Y K M V Q A IK W DE V V PA AP Y VT LE T L D K Y NC D FC V H G ND IT LT-----							
7	rCCT P19836 3HL4	33.1%	T VM N EN ER Y DA V Q H CR Y V DE V VR NA P WT L T PE FLA EH RI D F V AH- DD IP Y SS-----							
8	hCCT P49585	22.8%	T VM N EN ER Y DA V Q H CR Y V DE V VR NA P WT L T PE FLA EH RI D F V AH- DD IP Y SS-----							
9	PfCCT-Cter Q8IEE9	50.1%	V V Q T LE ER T ET L K H IR W DE I I SP C P W V V T PE FL E K Y K ID Y V AH- DD IP Y ANN Q KK KK KK							
10	PfCCT-Nter Q8IEE9	100.0%	I V Q T LE ER T ET L K H IR W DE I I SP C P W I T PE F ME EH K ID Y V AH- DD IP Y ANN Q KK KK KK							

	241	:	3	300
1 saGCT 005155 2B7L	18.3%	-----	-EGEFDFLKD	KCEVIYLR	RTEGISTT	KIKQELY	-----		
2 bsGCT P27623 1COZ, 1N1D	20.5%	-----	-EGKDFDLK	DQCEVVYLP	RTEGISTT	KIKKEIA	-----		
3 bsGCT Q8RKI6	18.6%	-----	-KGKDFDLK	EYCEVIYLR	RTDGIISTT	QIKKDMA	-----		
4 PfECT Q8IDM2	13.2%	-----	-KNGTCCYEE	VRKFNKLIK	RSYGISTT	TIINHL	QAVNNSNYSSNNN		
5 rECT 088637	18.8%	-----	-VDGRD	TYEEVKQ	AGRYRECKRTQGVSTT	DLVGRML	-----		
6 hECT Q99447 3ELB	20.8%	-----	-VDGRD	TYEEVKQ	AGRYRECKRTQGVSTT	DLVGRML	-----		
7 rCCT P19836 3HL4	33.1%	-----	-AGSD	DVYKHIKE	AGMFAPTORTEGISTT	DIITRIV	-----		
8 hCCT P49585	22.8%	-----	-AGSD	DVYKHIKE	AGMFAPTORTEGISTT	DIITRIV	-----		
9 PfCCT-Cter Q8IEE9	50.1%	KSKGKSF	SFDEENE	DIYAWLKR	AGKFKATQRTTEGVSTT	DLIVRIL	-----		
10 PfCCT-Nter Q8IEE9	100.0%	KSKGKSF	SFDEENE	DIYAWLKR	AGKFKATQRTTEGVSTT	DLIVRIL	-----		
	301	:	360
1 saGCT 005155 2B7L	18.3%	-----							
2 bsGCT P27623 1COZ, 1N1D	20.5%	-----							
3 bsGCT Q8RKI6	18.6%	-----							
4 PfECT Q8IDM2	13.2%	NNNNNNN	TLVNSNN	NNNNNDTNS	SVSTNEIS	DINNETNYV	NTNTNSEQL	DNFNKNDNP	
5 rECT 088637	18.8%	-----							
6 hECT Q99447 3ELB	20.8%	-----							
7 rCCT P19836 3HL4	33.1%	-----							
8 hCCT P49585	22.8%	-----							
9 PfCCT-Cter Q8IEE9	50.1%	-----							
10 PfCCT-Nter Q8IEE9	100.0%	-----							
	361	420
1 saGCT 005155 2B7L	18.3%	-----							
2 bsGCT P27623 1COZ, 1N1D	20.5%	-----							
3 bsGCT Q8RKI6	18.6%	-----							
4 PfECT Q8IDM2	13.2%	NILEITEE	QIYNSEL	GISDDN	KTQVSEQ	QHIDTL	PKNLLNR	NRCHITTSQIY	QFIDNNE
5 rECT 088637	18.8%	PHPTPAG	DTLSSEV	SSQCPG	GGQSPW	TGVSQFL	QTSQKII	-----	-QFASGKE
6 hECT Q99447 3ELB	20.8%	-----							
7 rCCT P19836 3HL4	33.1%	-----							
8 hCCT P49585	22.8%	-----							
9 PfCCT-Cter Q8IEE9	50.1%	-----							
10 PfCCT-Nter Q8IEE9	100.0%	-----							
	421	480
1 saGCT 005155 2B7L	18.3%	-----							
2 bsGCT P27623 1COZ, 1N1D	20.5%	-----							
3 bsGCT Q8RKI6	18.6%	-----							
4 PfECT Q8IDM2	13.2%	LIKKKK	KNKVVY	VDGSFD	IFHIGH	LRILENA	KKLGD--	YLLV	GMHSDEVVQ
5 rECT 088637	18.8%	P---	QPGETVI	YVAGAF	DLFHIGH	VDFLQEV	HKLAKR	PPYVIA	GLHFDQEV
6 hECT Q99447 3ELB	20.8%	P---	QPGETVI	YVAGAF	DLFHIGH	VDFLQEV	HKLAKR	PPYVIA	GLHFDQEV
7 rCCT P19836 3HL4	33.1%	-----							
8 hCCT P49585	22.8%	Y---	HLQERV	DKVKK	KKV-----	D-----	V---	EEKS	KEFVQKVEE
9 PfCCT-Cter Q8IEE9	50.1%	I---	KMKKNL	IRWGE	KVT-----	DELTKV	-TLTDKPL	---	GTDFDQ
10 PfCCT-Nter Q8IEE9	100.0%	I---	KMKKNL	IRWGE	KVT-----	DELTKV	-TLTDKPL	---	GTDFDQ

		481	.	5	540
1	saGCT 005155 2B7L	18.3%	-----						
2	bsGCT P27623 1COZ,1N1D	20.5%	-----						
3	bsGCT Q8RKI6	18.6%	-----						
4	PfECT Q8IDM2	13.2%	VSLLERTLNLVAMKVVDDVVIGAPWVI-----TESFIKRFH						
5	rECT 088637	18.8%	MNLHERTLNVLACRYVSEVVIGAPYSV-----TAELLNHFK						
6	hECT Q99447 3ELB	20.8%	MNLHERTLNVLACRYVSEVVIGAPYAV-----TAELLSHFH						
7	rCCT P19836 3HL4	33.1%	-----						
8	hCCT P49585	22.8%	IQKWEEKSREFIGSFLEM--FGPEGAL-----KHMLKEGKG						
9	PfCCT-Cter Q8IEE9	50.1%	FKIWKNASNKLITDF-----						
10	PfCCT-Nter Q8IEE9	100.0%	FKLWRYHKKLLKDFAKS--FDPMFIIIRKRYRKNLSAMYLSDSNYFSSMKDELKKKK						
		541	:	6 600
1	saGCT 005155 2B7L	18.3%	-----						
2	bsGCT P27623 1COZ,1N1D	20.5%	-----						
3	bsGCT Q8RKI6	18.6%	-----						
4	PfECT Q8IDM2	13.2%	IDVVVVRTIVDYIYSNNEIDPYDIPKLN-----YQELSSSNITTYEIIQRIEKN						
5	rECT 088637	18.8%	VDLVCHGK-TEIVPDRDGDSPYEEPKRRGI-----FCQIDSGSLTTLIVQRIIKN						
6	hECT Q99447 3ELB	20.8%	VDLVCHGK-TEIIPDRDGDSPYQEPKRRGI-----FRQIDSGSLTTLIVQRIITN						
7	rCCT P19836 3HL4	33.1%	-----						
8	hCCT P49585	22.8%	RMLQAISP-KQSPSSPTRERSPPSPFRWP-----FSGKTSPPCPANLSRHKAAY						
9	PfCCT-Cter Q8IEE9	50.1%	-----TRKLEATSYLTSIQNIIDYEIE						
10	PfCCT-Nter Q8IEE9	100.0%	SLSSTFNIDEYYSADDEEDTRDNSFYRVINKIANHTYSEKENYETCFELEHSLRDKKK						
		601] 643
1	saGCT 005155 2B7L	18.3%	-----						
2	bsGCT P27623 1COZ,1N1D	20.5%	-----						
3	bsGCT Q8RKI6	18.6%	-----						
4	PfECT Q8IDM2	13.2%	K-KYLMRNMSKRNKKEESIWETSNTYAINN-----						
5	rECT 088637	18.8%	RLEYEARNQKKEAKELAFLEALRQQAQPRGETD-----						
6	hECT Q99447 3ELB	20.8%	RLEY-----						
7	rCCT P19836 3HL4	33.1%	-----						
8	hCCT P49585	22.8%	DISEDEED-----						
9	PfCCT-Cter Q8IEE9	50.1%	NDDYASSNFDEETSS-----						
10	PfCCT-Nter Q8IEE9	100.0%	SLIFKDKNMKLSRDDLLHFTNPMNSKYDHNYYYYNNTYDKLS						

[MView](#) 1.49, Copyright © 1997-2006 Nigel P. Brown

Annexe 3 - Aligement des séquences de la famille des cytidyltransférases, utilisé pour construire le modèle comparatif de la *PfCCT*.

Annexe 4 - Publication <http://bioinformatics.oxfordjournals.org/content/28/11/1540.full>

Pihan, E., Colliandre, L., Guichou, J.-F. & Douguet, D., 2012. e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics*, 28(11), pp.1540–1.

e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design

Emilie Pihan¹, Lionel Colliandre², Jean-François Guichou² and Dominique Douguet^{2,*}

¹Institut de Pharmacologie Moléculaire et Cellulaire, CNRS UMR7275, Université Nice – Sophia Antipolis, 660, route des Lucioles 06560 Valbonne and ²Centre de Biochimie Structurale, INSERM U1054, CNRS UMR5048 – Université Montpellier 1 et 2, Montpellier, France

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: In the drug discovery field, new uses for old drugs, selective optimization of side activities and fragment-based drug design (FBDD) have proved to be successful alternatives to high-throughput screening. e-Drug3D is a database of 3D chemical structures of drugs that provides several collections of ready-to-screen SD files of drugs and commercial drug fragments. They are natural inputs in studies dedicated to drug repurposing and FBDD.

Availability: e-Drug3D collections are freely available at <http://chemoinfo.ipmc.cnrs.fr/e-drug3d.html> either for download or for direct *in silico* web-based screenings.

Contact: douguet@ipmc.cnrs.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on December 7, 2011; revised on March 30, 2012; accepted on April 5, 2012

1 INTRODUCTION

Approved drugs represent very attractive and valuable starting point for drug discovery. Studies of drugs entering the market show that most of their drug progenitors (or lead structures) were known drugs or clinical candidates (Proudfoot, 2002; Teague, 2011). The rationale is that a drug usually acts on more than one target and may exhibit previously unknown functions due to promiscuous off-target interactions explaining efficacy and/or side-effects (Campillos *et al.*, 2008). Examples of new therapeutic indications for old drugs include the anti-emetic and hypnotic Thalidomide originally withdrawn as a leprostatic agent (Teo *et al.*, 2005) or the anti-neoplastic Finasteride for the treatment of hair loss (approved in 1997). Over the years, an increasing number of pharmaceutical companies have been interested in 'drug repurposing' in order to reduce the risk and cost of the development of new chemical entities (Tobinick, 2009). The fragment-based drug design (FBDD) approach represents another validated alternative to high-throughput screening (Murray and Blundell, 2010). It consists of identifying low molecular weight molecules that are able to interact with a defined binding site. Hits are then combined or grown into high-affinity ligands. Over the past decade, FBDD has become established as an effective approach and has led to the discovery of high-affinity bioactive molecules, some of which are in clinical development. The 'rule of three' provides a useful guideline when designing the fragment

screening collection (Congreve *et al.*, 2003), but the selection criteria may also include 'privileged structures' that are commonly found in known drug molecules (Hartshorn *et al.*, 2005; Lepre, 2001). This strategy is based on the assumption that molecules resembling existing drugs are more likely to possess appropriate ADMET properties than random molecules. e-Drug3D has been purposely designed to provide free and ready-to-screen virtual collections of approved drugs and of their commercially available substructures (fragments). These collections of compounds are natural inputs for various cheminformatic and virtual screening applications.

2 e-DRUG3D COLLECTIONS

e-Drug3D is an annotated database based on the 'Drugs@FDA Data File' released by the US Food and Drug Administration (FDA). Over the years, several publicly available and drug-specific databases have emerged including DrugBank (Knox *et al.*, 2011), ChEMBL/Drugs (Gaulton *et al.*, 2012) and SuperDrug (Goede *et al.*, 2005). SuperDrug was the first free resource that provided 3D conformers of drug structures. In developing e-Drug3D, our objective was to further expand such computed drug structures by adding recent approved drugs and providing a collection of commercially available drug fragments. e-Drug3D currently contains 1519 annotated 3D structures of 1305 different FDA-approved drugs of molecular weight <2000 (last update: June 2011). Chiral centers are checked and enantiomers are differentiated. We also provide two supplementary collections resulting from (i) the calculation of the most probable tautomeric and ionic states at pH 7.4 and (ii) the generation of multiple conformations for ring systems. Finally, we produced a collection of commercial substructures of drugs. The current collection contains 1383 commercial fragments that are represented by a single conformation. Preparation, analyses and comparisons of the two collections are given in the Supplementary Material. The e-Drug3D database is accessible through our web interface for browsing. It can be queried by keyword search, structural criteria or similarity search. Tools are also provided to select subsets of drugs for comparisons. Drug and fragment collections are directly linked to the *in silico* screening web server e-LEA3D (Douguet, 2010).

3 A FBDD EXAMPLE

The FBDD approach was successfully used to identify novel inhibitors of human cyclophilins. These proteins are enzymes exhibiting a peptidyl-prolyl cis–trans isomerase activity and playing

*To whom correspondence should be addressed.

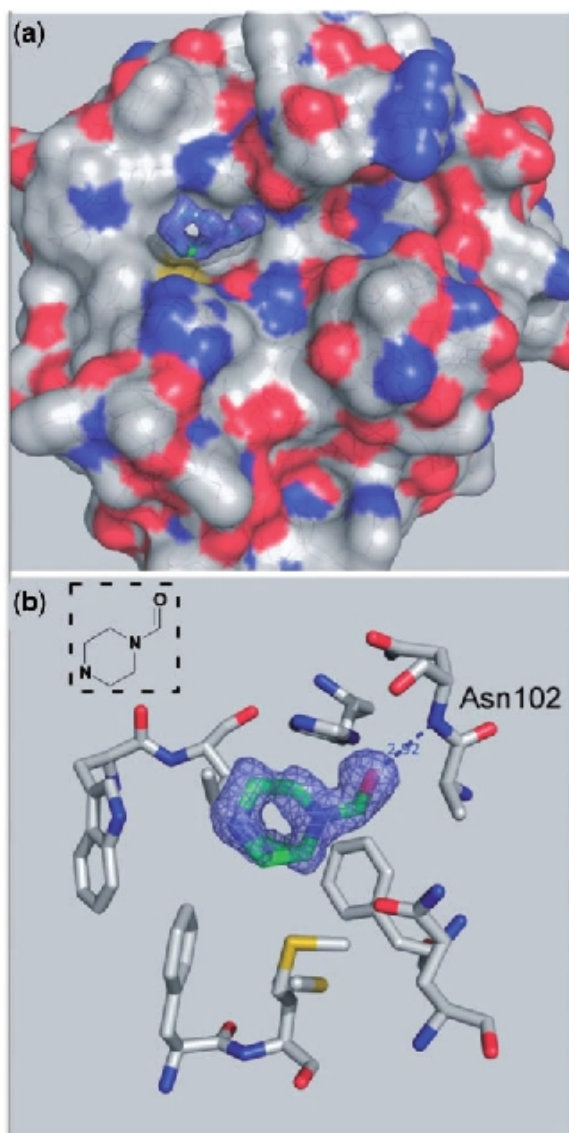


Fig. 1. (a) View of the 3D surface of the human cyclophilin D complexed with 1-formylpiperazine (PDB ID 3RCK). The fragment is represented by its electronic density in blue (map contour at 1.0σ). (b) Residues implicated in the binding are shown along with the hydrogen bond between the carbonyl moiety of 1-formylpiperazine and the backbone nitrogen of residue Asn102. Inset— 2D sketch of 1-formylpiperazine. Pictures were generated by using the program Pymol.

a role in immunity and viral infection, including Hepatitis C virus (HCV; Guichou *et al.*, 2011; Watashi, 2010). Several fragments

were selected by virtual screening and soaked with cyclophilin D crystals (Schlatter *et al.*, 2005). Among them, the compound 1-formylpiperazine corresponds to one fragment of our drug collection that is a substructure of the drugs Delavirdine, Doxazosin, Prazosin and Terazosin. Figure 1 shows the structure of the human cyclophilin D with this fragment (PDB ID 3RCK). Chemical optimization of several co-crystallized hit fragments led to the identification of submicromolar inhibitors of cyclophilin A, B and D (Guichou *et al.*, 2011). These molecules show similar inhibition in infectious models and no cytotoxicity at the effective concentrations. The discovery of non-peptidic cyclophilin inhibitors is a breakthrough in the ongoing search for new anti-HCV agents as well as a successful example of the application of FBDD to difficult targets.

ACKNOWLEDGEMENT

We thank D.Barbier for his help in managing the web server.

Funding: French National Research Agency [Grant FRAGSCREEN ANR-07-JC-JC-0046-01].

Conflict of Interest: none declared.

REFERENCES

- Campillos, M. *et al.* (2008) Drug target identification using side-effect similarity. *Science*, **321**, 263–266.
- Congreve, M. *et al.* (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today*, **8**, 876–877.
- Douguet, D. (2010) e-LEA3D: a computational-aided drug design web server. *Nucleic Acids Res.*, **38**, W615–W621.
- Gaulton, A. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
- Goede, A. *et al.* (2005) SuperDrug: a conformational drug database. *Bioinformatics*, **21**, 1751–1753.
- Guichou, J.-F. *et al.* (2011) New inhibitors of cyclophilins and uses thereof. Patent WO2011076784.
- Hartshorn, M.J. *et al.* (2005) Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.*, **48**, 403–413.
- Knox, C. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
- Lepre, C.A. (2001) Library design for NMR-based screening. *Drug Discov. Today*, **6**, 133–140.
- Murray, C.W. and Blundell, T.L. (2010) Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.*, **20**, 497–507.
- Proudfoot, J.R. (2002) Drugs, leads and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.*, **12**, 1647–1650.
- Schlatter, D. *et al.* (2005) Crystal engineering yields crystals of cyclophilin D diffracting to 1.7 Å resolution. *Acta Crystallogr. D*, **61**, 513–519.
- Teague, S.J. (2011) Learning lessons from drugs that have recently entered the market. *Drug Discov. Today*, **16**, 398–411.
- Teo, S.K. *et al.* (2005) Thalidomide as a novel therapeutic agent: new uses for an old product. *Drug Discov. Today*, **10**, 107–114.
- Tobinick, E.L. (2009) The value of drug repositioning in the current pharmaceutical market. *Drug News Perspect.*, **22**, 119–125.
- Watashi, K. (2010) Alisporivir, a cyclosporin derivative that selectively inhibits cyclophilin, for the treatment of HCV infection. *Curr. Opin. Invest. Drugs*, **11**, 213–224.

Bibliographie

- Abdoreza, D.-S., Melissa, S., Nallapali, S., Verma, D., Singh, N.D., Banks, R.K., Chakrabarti, D. & Daniell, H., 2010. Chloroplast-derived vaccine antigens confer dual immunity against cholera and malaria by oral or injectable delivery. *Plant Biotechnology Journal*, 8(2), pp.223–242.
- Achan, J., Kakuru, A., Ikilezi, G., Ruel, T., Clark, T.D., Nsanzabana, C., Charlebois, E., Aweeka, F., Dorsey, G., Rosenthal, P.J., Havlir, D. & Kanya, M.R., 2012. Antiretroviral agents and prevention of malaria in HIV-infected Ugandan children. *The New England journal of medicine*, 367(22), pp.2110–8.
- Aguiar, A.C.C., Rocha, E.M. Da, Souza, N.B. De, França, T.C. & Krettli, A.U., 2012. New approaches in antimalarial drug discovery and development: a review. *Memórias do Instituto Oswaldo Cruz*, 107(7), pp.831–45.
- Ahmad, S. & Gromiha, M.M., 2002. NETASA: neural network based prediction of solvent accessibility. *Bioinformatics (Oxford, England)*, 18(6), pp.819–24.
- Alberge, B., Gannoun-Zaki, Leila, Bascunana, C., Tran van Ba, C., Vial, Henri & Cerdan, Rachel, 2010. Comparison of the cellular and biochemical properties of Plasmodium falciparum choline and ethanolamine kinases. *The Biochemical journal*, 425(1), pp.149–58.
- Alexander, D.L., Mital, J., Ward, G.E., Bradley, Peter & Boothroyd, J.C., 2005. Identification of the moving junction complex of Toxoplasma gondii: a collaboration between distinct secretory organelles. *PLoS pathogens*, 1(2), p.e17.
- Ancelin, Marie L, Herbute, S., Ringwald, P. & Vial, H.J., 2003. Potent inhibitors of Plasmodium phospholipid metabolism with a broad spectrum of in vitro antimalarial activities. *Antimicrobial agents and Chemotherapy*, 47(8), pp.2590–2597.
- Ancelin, M. & Calas, M, 2003. In vivo antimalarial activities of mono-and bis quaternary ammonium salts interfering with Plasmodium phospholipid metabolism. *Antimicrobial agents and Chemotherapy*, 47(8), pp.2598–2605.
- Ancelin, M. & Vial, HJ, 1989. Regulation of phosphatidylcholine biosynthesis in Plasmodium-infected erythrocytes. *Biochimica et Biophysica Acta*, 1001, pp.82–89.
- Audran, R., Lurati-Ruiz, F., Genton, B., Blythman, H.E., Ofori-Anyinam, O., Reymond, C., Corradin, G. & Spertini, F., 2009. The synthetic Plasmodium falciparum circumsporozoite peptide PfCS102 as a malaria vaccine candidate: a randomized controlled phase I trial. *PLoS one*, 4(10), p.e7304.
- Baell, J.B. & Holloway, G. a, 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of medicinal chemistry*, 53(7), pp.2719–40.
- Bai, T., Becker, M., Gupta, A., Strike, P., Murphy, Vince J, Anders, R.F. & Batchelor, A.H., 2005. Structure of AMA1 from Plasmodium falciparum reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36), pp.12736–41.
- Baum, J. & Cowman, A.F., 2011. Revealing a parasite's invasive trick. *Science (New York, N.Y.)*, 333(6041), pp.410–411.

- Bender, A., Jenkins, J.L., Scheiber, J., Sukuru, S.C.K., Glick, M. & Davies, J.W., 2009. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *Journal of chemical information and modeling*, 49(1), pp.108–19.
- Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J.-F. & Lebrun, M., 2009. Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS pathogens*, 5(2), p.e1000309.
- Biagini, G. a., Richier, E., Bray, P., Calas, Michèle, Vial, Henri & Ward, S., 2003. Heme binding contributes to antimalarial activity of bis-quaternary ammoniums. *Antimicrobial agents and Chemotherapy*, 47(8), pp.2584–2589.
- Bloom, B.H., 1970. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7), pp.422–426.
- Bondi, A., 1964. van der Waals volumes and radii. *The Journal of physical chemistry*, 68(3), pp.441–451.
- Calas, Michèle, Ancelin, Marie L, Cordina, G., Portefaix, P., Piquet, G., Vidal-Sailhan, V. & Vial, Henri, 2000. Antimalarial activity of compounds interfering with *Plasmodium falciparum* phospholipid metabolism: comparison between mono-and bisquaternary ammonium salts. *Journal of medicinal chemistry*, pp.505–516.
- Chang, M.W., Ayeni, C., Breuer, S. & Torbett, B.E., 2010. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS one*, 5(8), p.e11955.
- Charman, S.A., Arbe-Banes, S., Bathurst, I.C., Brun, R., Campbell, M., Charman, W.N., Chiu, F.C.K., Chollet, J., Craft, J.C., Creek, D.J., Dong, Y., Matile, Hugues, Maurer, M., Morizzi, J., Nguyen, T., Papastogiannidis, P., Scheurer, C., Shackelford, D.M., Sriraghavan, K., Stingelin, L., Tang, Y., Urwyler, H., Wang, Xiaofang, White, K.L., Wittlin, S., Zhou, L. & Vennerstrom, J.L., 2011. Synthetic ozonide drug candidate OZ439 offers new hope for a single-dose cure of uncomplicated malaria. *Proceedings of the National Academy of Sciences of the United States of America*, 108(11), pp.4400–5.
- Chauhan, V.S., Yazdani, S.S. & Gaur, D., 2010. Malaria vaccine development based on merozoite surface proteins of. *Human Vaccines*, (September), pp.757–762.
- Chen, R., Li, L. & Weng, Z., 2003. ZDOCK: An initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Genetics*, 52(November 2002), pp.80–87.
- Chilengi, R., Juma, R., Abdallah, A.M., Bashraheil, M., Lodenyo, H., Nyakundi, P., Anabwani, E., Salim, A., Mwambingu, G., Wenwa, E., Jemutai, J., Kipkeu, C., Oyoo, G.O., Muchohi, S.N., Kokwaro, G., Niehues, T., Lang, T. & Nzila, A., 2011. A phase I trial to evaluate the safety and pharmacokinetics of low-dose methotrexate as an anti-malarial drug in Kenyan adult healthy volunteers. *Malaria journal*, 10(1), p.63.
- Chivian, D., Kim, D.E., Malmström, L., Bradley, Philip, Robertson, T., Murphy, P., Strauss, C.E.M., Bonneau, R., Rohl, C. a & Baker, D., 2003. Automated prediction of CASP-5 structures using the Robetta server. *Proteins*, 53 Suppl 6(February), pp.524–33.
- Coley, A.M., Gupta, A., Murphy, Vince J, Bai, T., Kim, H., Foley, Michael, Anders, R.F. & Batchelor, A.H., 2007. Structure of the malaria antigen AMA1 in complex with a growth-inhibitory antibody. *PLoS pathogens*, 3(9), pp.1308–19.

- Comeau, S.R., Gatchell, D.W., Vajda, S. & Camacho, C. J., 2003. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20(1), pp.45–50.
- Cowman, A.F. & Crabb, B.S., 2006. Invasion of red blood cells by malaria parasites. *Cell*, 124(4), pp.755–66.
- Crawford, J., Tonkin, M.L., Grujic, O. & Boulanger, M.J., 2010. Structural characterization of apical membrane antigen 1 (AMA1) from *Toxoplasma gondii*. *The Journal of biological chemistry*, 285(20), pp.15644–52.
- Croft, S.L., Duparc, S., Arbe-Barnes, S.J., Craft, J.C., Shin, C.-S., Fleckenstein, L., Borghini-Fuhrer, I. & Rim, H.-J., 2012. Review of pyronaridine anti-malarial properties and product characteristics. *Malaria journal*, 11(1), p.270.
- Cui, H., Carrero-Lérida, J., Silva, A.P.G., Whittingham, J.L., Brannigan, J. a, Ruiz-Pérez, L.M., Read, K.D., Wilson, K.S., González-Pacanowska, D. & Gilbert, I.H., 2012. Synthesis and Evaluation of α -Thymidine Analogues as Novel Antimalarials. *Journal of medicinal chemistry*, 55(24), pp.10948–57.
- Déchamps, S., Wengelnik, K., Berry-Sterkers, L., Cerdan, Rachel, Vial, H.J. & Gannoun-Zaki, Laila, 2010. The Kennedy phospholipid biosynthesis pathways are refractory to genetic disruption in *Plasmodium berghei* and therefore appear essential in blood stages. *Molecular and biochemical parasitology*, 173(2), pp.69–80.
- Delves, M., Plouffe, D., Scheurer, C., Meister, S., Wittlin, S., Winzeler, E. a, Sinden, R.E. & Leroy, D., 2012. The activities of current antimalarial drugs on the life cycle stages of *Plasmodium*: a comparative study with human and rodent parasites. *PLoS medicine*, 9(2), p.e1001169.
- Desjardins, R.E., Canfield, C.J., Haynes, J.D. & Chulay, J.D., 1979. Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrobial agents and Chemotherapy*, 16(6), pp.710–718.
- Dobson, C.M., 2004. Chemical space and biology. *Nature*, 432(7019), pp.824–8.
- Dooren, G. Van, Waller, R. & McFadden, G., 2000. Traffic Jams: Protein Transport in *Plasmodium falciparum*. *Parasitology Today*, 4758(00), pp.421–427.
- Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M., 1995. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3), pp.273–284.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U. & Sali, A., 2006. Comparative protein structure modeling using MODELLER. J. E. Coligan, B. M. Dunn, D. W. Speicher, & P. T. Wingfield, eds. *Current protocols in bioinformatics editorial board Andreas D Baxeavanis et al*, Chapter 2(November), p.Unit 2.9.
- Fichera, M.E. & Roos, D.S., 1997. A plastid organelle as a drug target in apicomplexan parasites. *Nature*, 390(6658), pp.407–9.

- Floris, M., Masciocchi, J., Fanton, M. & Moro, S., 2011. Swimming into peptidomimetic chemical space using pepMMsMIMIC. *Nucleic acids research*, 39(Web Server issue), pp.W261–W269.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C. & Jensen, L.J., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research*, 41(Database issue), pp.D808–15.
- Genton, B., D'Acromont, V., Lurati-Ruiz, F., Verhage, D., Audran, R., Hermsen, C., Wolters, L., Reymond, C., Spertini, F. & Sauerwein, R., 2010. Randomized double-blind controlled Phase I/IIa trial to assess the efficacy of malaria vaccine PfCS102 to protect against challenge with *P. falciparum*. *Vaccine*, 28(40), pp.6573–80.
- Gracy, J. & Chiche, L., 2005. PAT: a protein analysis toolkit for integrated biocomputing on the web. *Nucleic acids research*, 33(Web Server issue), pp.W65–71.
- Grant, J., Gallardo, M. & Pickup, B., 1998. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *Journal of computational chemistry*, 17(14), pp.1653–1666.
- Graves, P. & Gelband, H., 2006. Vaccines for preventing malaria (SPf66). *Cochrane database of systematic reviews (Online)*, (2), p.CD005966.
- Harris, K., Casey, J., Coley, A., Masciantonio, R., Sabo, J., Keizer, D., Lee, E., McMahon, A., Norton, R., Ander, R. & Foley, M., 2005. Binding hot spot for invasion inhibitory molecules on *Plasmodium falciparum* apical membrane antigen 1. *Infection and Immunity*, 73(10), pp.6981–6989.
- Hawkins, P.C.D., Skillman, a. G. & Nicholls, A., 2007. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry*, 50(1), pp.74–82.
- Healer, J., Murphy, V., Hodder, A.N., Masciantonio, Rosella, Gemmill, A.W., Anders, R.F., Cowman, A.F. & Batchelor, A., 2004. Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum*. *Molecular microbiology*, 52(1), pp.159–68.
- Von Heijne, G., 1992. Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology*, 225(2), pp.487–494.
- Henderson, K. a, Streltsov, V. a, Coley, A.M., Dolezal, O., Hudson, P.J., Batchelor, A.H., Gupta, A., Bai, T., Murphy, Vincent J, Anders, R.F., Foley, Michael & Nuttall, S.D., 2007. Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. *Structure (London, England : 1993)*, 15(11), pp.1452–66.
- Jubb, H., Higuieruelo, A.P., Winter, A. & Blundell, T.L., 2012. Structural biology and drug discovery for protein-protein interactions. *Trends in pharmacological sciences*, 33(5), pp.241–8.
- Kappe, S.H.I., Vaughan, A.M., Boddey, J. a & Cowman, A.F., 2010. That was then but this is now: malaria research in the time of an eradication agenda. *Science (New York, N.Y.)*, 328(5980), pp.862–6.

- Kleywegt, G.J. & Jones, T. a, 1994. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 2), pp.178–85.
- Koes, D.R. & Camacho, Carlos J, 2011. Pharmer: efficient and exact pharmacophore search. *Journal of chemical information and modeling*, 51(6), pp.1307–14.
- Korb, O. & Stützele, T., 2009. Empirical scoring functions for advanced protein– ligand docking with plants. *Journal of chemical information and Modeling*, 49, pp.84–96.
- Korb, O., Stützele, T. & Exner, T., 2006. PLANTS: Application of ant colony optimization to structure-based drug design. *Ant Colony Optimization and Swarm*, pp.247–258.
- Krivov, G.G., Shapovalov, M. V & Dunbrack, R.L., 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4), pp.778–95.
- Lagorce, D., Maupetit, J., Baell, J., Sperandio, O., Tufféry, P., Miteva, M. a, Galons, H. & Villoutreix, B.O., 2011. The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. *Bioinformatics (Oxford, England)*, 27(14), pp.2018–20.
- Lamarque, Mauld, Besteiro, S., Papoin, J., Roques, M., Vulliez-Le Normand, B., Morlon-Guyot, J., Dubremetz, J.-F., Fauquenoy, S., Tomavo, S., Faber, B.W., Hocken, C.H., Thomas, A.W., Boulanger, M.J., Bentley, G.A. & Lebrun, M., 2011. The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites. *PLoS pathogens*, 7(2), p.e1001276.
- LaRusch, G.A., Jackson, M.W., Dunbar, J.D., Warren, R.S., Donner, D.B. & Mayo, L.D., 2007. Nutlin3 blocks vascular endothelial growth factor induction by preventing the interaction between hypoxia inducible factor 1alpha and Hdm2. *Cancer Research*, 67(2), pp.450–454.
- Larvor, M.-P., Cerdan, Rachel, Gumila, C., Maurin, L., Seta, P. & Roustan, C., 2003. Characterization of the lipid-binding domain of the Plasmodium falciparum CTP: phosphocholine cytidylyltransferase through synthetic-peptide studies. *Biochemical*, 375, pp.653–661.
- Lebrun, M., Michelin, A., El Hajj, H., Poncet, J., Bradley, P.J., Vial, Henri & Dubremetz, J.F., 2005. The rhoptry neck protein RON4 re-localizes at the moving junction during Toxoplasma gondii invasion. *Cellular microbiology*, 7(12), pp.1823–33.
- Lee, J., Johnson, J., Ding, Z., Paetzel, M. & Cornell, R.B., 2009. Crystal Structure of a Mammalian CTP: Phosphocholine Cytidylyltransferase Catalytic Domain Reveals Novel Active Site Residues within a Highly Conserved Nucleotidyltransferase Fold. *Journal of Biological Chemistry*, 284(48), pp.33535–33548.
- Lell, B., Köhler, C., Wamola, B., Olola, C.H., Kivaya, E., Kokwaro, G., Wypij, D., Mithwani, S., Taylor, T.E., Kremsner, P.G. & Newton, C.R.J.C., 2010. Pentoxifylline as an adjunct therapy in children with cerebral malaria. *Malaria journal*, 9(1), p.368.
- Lipinski, C., Lombardo, F., Dominy, B. & Feeney, P., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 46(1-3), pp.3–26.

- Livingstone, D.J., Clark, T., Ford, M.G., Hudson, B.D. & Whitley, D.C., 2007. QSAR studies using the parashift system. *SAR and QSAR in environmental research*, 19(3-4), pp.285–302.
- Loriot, S. & Cazals, F., 2010. Modeling macro-molecular interfaces with Intervor. *Bioinformatics (Oxford, England)*, 26(7), pp.964–5.
- Malkin, E., Hu, J., Li, Z., Chen, Zhihui, Bi, X., Reed, Z., Dubovsky, F., Liu, J., Wang, Q., Pan, X., Chen, T., Giersing, B., Xu, Y., Kang, X., Gu, J., Shen, Q., Tucker, K., Tierney, E., Pan, W., Long, C. & Cao, Z., 2008. A phase 1 trial of PfCP2.9: an AMA1/MSP1 chimeric recombinant protein vaccine for Plasmodium falciparum malaria. *Vaccine*, 26(52), pp.6864–73.
- Martin, Y.C., 2009. Let's not forget tautomers. *Journal of computer-aided molecular design*, 23(10), pp.693–704.
- McDonald, I.K. & Thornton, J.M., 1994. Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology*, 238(5), pp.777–793.
- Miller, R.L., IKram, S., Armelagos, S., G.J., Walker, R., Harer, W.B., Shiff, C.J., Baggett, D. & Carrigan, M., 1994. Diagnosis of Plasmodium falciparum infections in mummies using the rapid manual ParaSight™-F test. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88(1), pp.31–32.
- Morris, G. & Huey, R., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), pp.2785–2791.
- Mullard, A., 2012. Protein-protein interaction inhibitors get into the groove. *Nature reviews. Drug discovery*, 11(3), pp.173–5.
- Mullen, G.E.D., Ellis, R.D., Miura, K., Malkin, E., Nolan, C., Hay, M., Fay, M.P., Saul, A., Zhu, D., Rausch, K., Moretz, S., Zhou, H., Long, C. a, Miller, L.H. & Treanor, J., 2008. Phase 1 trial of AMA1-C1/Alhydrogel plus CPG 7909: an asexual blood-stage vaccine for Plasmodium falciparum malaria. *PloS one*, 3(8), p.e2940.
- Na-Bangchang, K., Ruengweerayut, R., Karbwang, J., Chauemung, A. & Hutchinson, D., 2007. Pharmacokinetics and pharmacodynamics of fosmidomycin monotherapy and combination therapy with clindamycin in the treatment of multidrug resistant falciparum malaria. *Malaria Journal*, 6, p.70.
- Nagy, G.N., Marton, L., Kramos, B., Olah, J., Revesz, A., Vekey, K., Delsuc, F., Hunyadi-Gulyas, E., Medzihradzsky, K.F., Lavigne, M., Vial, Henri, Cerdan, Rachel & Vertessy, B.G., 2013. Evolutionary and mechanistic insights into substrate and product accommodation of CTP:phosphocholine cytidyltransferase from Plasmodium falciparum. *FEBS Journal*.
- Nicolas, O., Margout, D. & Taudon, N., 2005. Pharmacological properties of a new antimalarial bithiazolium salt, T3, and a corresponding prodrug, TE3. *Antimicrobial agents ...*, 49(9), pp.3631–3639.
- Nugent, T. & Jones, D.T., 2009. Transmembrane protein topology prediction using support vector machines. *BMC bioinformatics*, 10, p.159.

- Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Awuondo, K.O., Leach, A., Lievens, M., Lebouilleux, D., Njuguna, P., Peshu, N., Marsh, K. & Bejon, P., 2013. Four-Year Efficacy of RTS,S/AS01E and Its Interaction with Malaria Exposure. *New England Journal of Medicine*, 368(12), pp.1111–1120.
- O'Neill, P.M., Park, K.B., Shone, A.E., Maggs, J.L., Roberts, P., Stocks, P.A., Biagini, G.A., Bray, P.G., Gibbons, P., Berry, N., Winstanley, P.A., Mukhtar, A., Bonar-Law, R., Hindley, S., Bambal, R.B., Davis, C.B., Bates, M., Hart, T.K., Gresham, S.L., Lawrence, R.M., Brigandi, R.A., Gomez-delas-Heras, F.M., Gargallo, D. V & Ward, S.A., 2009. Candidate Selection and Preclinical Evaluation of N-tert-Butyl Isoquine (GSK369796), An Affordable and Effective 4-Aminoquinone Antimalarial for the 21st Century. *Journal of medicinal chemistry*, 52, pp.1408–1415.
- Patarroyo, M.E., Amador, R., Clavijo, P., Moreno, A., Guzman, F., Romero, P., Tascon, R., Franco, A., Murillo, L.A. & Ponton, G., 1988. A synthetic vaccine protects humans against challenge with asexual blood stages of Plasmodium falciparum malaria. *Nature*, 332(6160), pp.158–161.
- Payne, D., 1987. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitology Today*, 3(8), pp.241–246.
- Petros, A.M., Huth, J.R., Oost, T., Park, C.-M., Ding, H., Wang, Xilu, Zhang, H., Nimmer, P., Mendoza, R., Sun, C., Mack, J., Walter, K., Dorwin, S., Gramling, E., Lador, U., Rosenberg, S.H., Elmore, S.W., Fesik, S.W. & Hajduk, P.J., 2010. Discovery of a potent and selective Bcl-2 inhibitor using SAR by NMR. *Bioorganic & medicinal chemistry letters*, 20(22), pp.6587–91.
- Pihan, E., Colliandre, L., Guichou, J.-F. & Douguet, D., 2012. e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics*, 28(11), pp.1540–1.
- Pizarro, J.C., Vulliez-Le Normand, B., Chesne-Seck, M.-L., Collins, C.R., Withers-Martinez, C., Hackett, F., Blackman, M.J., Faber, B.W., Remarque, E.J., Kocken, C.H.M., Thomas, A.W. & Bentley, G. a, 2005. Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science (New York, N.Y.)*, 308(5720), pp.408–11.
- Purfield, A.E., Tidwell, R.R. & Meshnick, S.R., 2009. The diamidine DB75 targets the nucleus of Plasmodium falciparum. *Malaria journal*, 8, p.104.
- Rangel Campos, W., Catarino, P.P. & Hoppe, H.C., 2012. New frontiers in ethnomedicine: exploring the Angolan indigenous pharmacopeia for novel antimalarial drugs. *Drug discovery today*, 17(13-14), pp.651–3.
- Rogers, D. & Hahn, M., 2010. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5), pp.742–754.
- Roggero, M.A., Filippi, B., Church, P., Hoffman, S.L., Blum-Tirouvanziam, U., Lopez, J.A., Esposito, F., Matile, H, Reymond, C.D. & Fasel, N., 1995. Synthesis and immunological characterization of 104-mer and 102-mer peptides corresponding to the N- and C-terminal regions of the Plasmodium falciparum CS protein. *Molecular Immunology*, 32(17-18), pp.1301–1309.

- Roggero, R., Zufferey, R., Minca, M., Richier, E., Calas, Michèle, Vial, Henri & Ben Mamoun, C., 2004. Unraveling the mode of action of the antimalarial choline analog G25 in *Plasmodium falciparum* and *Saccharomyces cerevisiae*. *Antimicrobial agents and Chemotherapy*, 48(8), pp.2816–2824.
- Rottmann, M., McNamara, C., Yeung, B.K.S., Lee, M.C.S., Zou, B., Russell, B., Seitz, P., Plouffe, D.M., Dharia, N. V, Tan, J., Cohen, S.B., Spencer, K.R., González-Páez, G.E., Lakshminarayana, S.B., Goh, A., Suwanarusk, R., Jegla, T., Schmitt, E.K., Beck, H.-P., Brun, R., Nosten, F., Renia, L., Dartois, V., Keller, T.H., Fidock, D. a, Winzeler, E. a & Diagana, T.T., 2010. Spiroindolones, a potent compound class for the treatment of malaria. *Science (New York, N.Y.)*, 329(5996), pp.1175–80.
- Roy, A, Kucukural, A. & Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, 5(4), pp.725–738.
- Runthala, A., 2012. Protein structure prediction: challenging targets for CASP10. *Journal of Biomolecular Structure and Dynamics*, 30(5), pp.607–615.
- Sagara, I., Dicko, A., Ellis, R.D., Fay, M.P., Diawara, S.I., Assadou, M.H., Sissoko, M.S., Kone, M., Diallo, A.I., Saye, R., Guindo, M.A., Kante, O., Niamebele, M.B., Miura, K., Mullen, G.E.D., Pierce, M., Martin, L.B., Dolo, A., Diallo, D.A., Doumbo, O.K., Miller, L.H. & Saul, A., 2009. A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine*, 27(23), pp.3090–3098.
- SARMA, P.S.A., 1988. Tinidazole: a new drug in the treatment of vivax malaria. *Current therapeutic research*, 43(5), pp.954–956.
- Sauerwein, R.W., Roestenberg, M. & Moorthy, V.S., 2011. Experimental human challenge infections can accelerate clinical malaria vaccine development. *Nature Reviews Immunology*, 11(1), pp.57–64.
- Sirima, S., Cousens, S. & Druilhe, P., 2011. Protection against malaria by MSP3 candidate vaccine. *New England Journal of Medicine*, 365(11), pp.1062–1064.
- Sperandio, O., Reynès, C.H., Camproux, A.-C. & Villoutreix, B.O., 2010. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug discovery today*, 15(5-6), pp.220–9.
- Taminau, J., Thijs, G. & De Winter, H., 2008. Pharao: pharmacophore alignment and optimization. *Journal of molecular graphics & modelling*, 27(2), pp.161–9.
- Taneva, S., Dennis, M. & Ding, Ziwei, 2008. Contribution of each membrane binding domain of the CTP: phosphocholine cytidyltransferase-alpha dimer to its activation, membrane binding, and membrane cross-. *Journal of Biological Chemistry*, 283(42), pp.28137–28148.
- Thera, M. a et al., 2008. Safety and immunogenicity of an AMA-1 malaria vaccine in Malian adults: results of a phase 1 randomized controlled trial. *PLoS one*, 3(1), p.e1465.
- Tonkin, M.L., Crawford, J., Lebrun, M.L. & Boulanger, M.J., 2013. *Babesia divergens* and *Neospora caninum* apical membrane antigen 1 structures reveal selectivity and plasticity in apicomplexan parasite host cell invasion. *Protein science: a publication of the Protein Society*, 22(1), pp.114–27.

- Tonkin, M., Roques, M., Lamarque, M.H., Pugnière, M., Douguet, Dominique, Crawford, J., Lebrun, M. & Boulanger, M.J., 2011. Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. *Science*, 333(22 July 2011), pp.463–467.
- Trott, O. & Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), pp.455–461.
- Vainio, M.J., Puranen, S.J. & Johnson, M.S., 2009. ShaEP: Molecular Overlay Based on Shape and Electrostatic Potential. *Journal of Chemical Information and Modeling*, 49(2), pp.492–502.
- Valecha, N., Looareesuwan, S., Martensson, A., Abdulla, S.M., Krudsood, S., Tangpukdee, N., Mohanty, S., Mishra, S.K., Tyagi, P.K., Sharma, S.K., Moehrle, J., Gautam, A., Roy, Arjun, Paliwal, J.K., Kothari, M., Saha, N., Dash, A.P. & Björkman, A., 2010. Arterolane, a new synthetic trioxolane for treatment of uncomplicated *Plasmodium falciparum* malaria: a phase II, multicenter, randomized, dose-finding clinical trial. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 51(6), pp.684–91.
- Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W. & Kopple, K.D., 2002. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, 45(12), pp.2615–23.
- Venhorst, J., Núñez, S., Terpstra, J.W. & Kruse, C.G., 2008. Assessment of scaffold hopping efficiency by use of molecular interaction fingerprints. *Journal of medicinal chemistry*, 51(11), pp.3222–9.
- Vial, Henri & Ben Mamoun, C., 2005. Plasmodium lipids: metabolism and function. In I. W. Sherman, ed. *Molecular approaches to malaria*. pp. 327–352.
- Vial, Henri, Thuet, M.J. & Philippot, J.R., 1984. Cholinephosphotransferase and ethanolaminephosphotransferase activities in *Plasmodium knowlesi*-infected erythrocytes. Their use as parasite-specific markers. *Biochimica et Biophysica Acta*, 795(2), pp.372–83.
- Vial, H.J., Wein, Sharon, Farenc, C., Kocken, C., Nicolas, O., Ancelin, Marie Laure, Bressolle, F., Thomas, A. & Calas, Michèle, 2004. Prodrugs of bisthiazolium salts are orally potent antimalarials. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43), pp.15458–63.
- Vulliez-Le Normand, B., Tonkin, M.L., Lamarque, M.H., Langer, S., Hoos, S., Roques, M., Saul, F. a, Faber, B.W., Bentley, G. a, Boulanger, M.J. & Lebrun, M., 2012. Structural and functional insights into the malaria parasite moving junction complex. *PLoS pathogens*, 8(6), p.e1002755.
- Wein, S, Maynadier, M., Bordat, Y., Perez, J., Maheshwari, S., Bette-Bobillo, P., Tran Van Ba, C., Penarete-Vargas, D., Fraisse, L., Cerdan, R & Vial, H, 2012. Transport and pharmacodynamics of albitiazolium, an antimalarial drug candidate. *British journal of pharmacology*, 166(8), pp.2263–76.

- Wengelnik, K., Vidal, V., Ancelin, Marie L, Cathiard, A.-M., Morgat, J.L., Kocken, C.H., Calas, Michèle, Herrera, S., Thomas, A.W. & Vial, H.J., 2002. A class of potent antimalarials and their specific accumulation in infected erythrocytes. *Science (New York, N.Y.)*, 295(5558), pp.1311–4.
- Wernsdorfer, W.H., 1991. The development and spread of drug-resistant malaria. *Parasitology Today*, 7(11), pp.297–303.
- Woodrow, C.J., Haynes, R.K. & Krishna, S., 2005. Artemisinins. *Postgraduate medical journal*, 81(952), pp.71–78.
- World Health Organization, 2011. World Malaria Report, 2011.
- Wu, T. et al., 2011. Imidazolopiperazines: Hit to Lead Optimization of New Antimalarial Agents. *Journal of Medicinal Chemistry*, 54(14), pp.5116–5130.
- Yeo, H., Larvor, M., Ancelin, M. & Vial, H.J., 1997. Plasmodium falciparum CTP: phosphocholine cytidyltransferase expressed in Escherichia coli: purification, characterization and lipid regulation. *Biochemical Journal*, (324), pp.903–910.
- Yeo, H., Widada, J., Mercereau-Puijalon, O. & Vial, H.J., 1995. Molecular cloning of CTP: phosphocholine cytidyltransferase from Plasmodium falciparum. *European Journal of Biochemistry*, 233, pp.62–72.
- Yuriev, E. & Ramsland, P. a, 2013. Latest developments in molecular docking: 2010-2011 in review. *Journal of molecular recognition : JMR*, 26(5), pp.215–39.



Cette œuvre est mise à disposition sous licence Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France. Pour voir une copie de cette licence, visitez <http://creativecommons.org/licenses/by-nc-nd/3.0/fr/> ou écrivez à Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Résumé en français

Le paludisme est causé par cinq espèces du genre *Plasmodium*, *P. falciparum* étant le plus mortel. Des résistances de certaines souches du parasite ont été rapportées pour tous les médicaments mis sur le marché. Les moustiques vecteurs du parasite sont résistants aux insecticides et aucun vaccin n'est disponible. Cette maladie est un problème économique et de santé publique pour les pays en voie de développement. Mes travaux de thèses visent à identifier de nouveaux traitements contre le paludisme, en ciblant deux nouvelles protéines.

Les Apicomplexes ont développé un mécanisme unique d'invasion, impliquant une interaction forte entre la cellule hôte et la surface du parasite, appelée jonction mobile. La caractérisation structurale et fonctionnelle du complexe AMA1-RON2 a ouvert la voie à la découverte de petites molécules capables d'empêcher l'interaction AMA1-RON2 et de ce fait, l'invasion. Le parasite a aussi besoin de phospholipides pour construire sa membrane durant le cycle érythrocytaire. Il y a six fois plus de phospholipides dans les érythrocytes infectés que dans les érythrocytes sains. Notre stratégie est d'inhiber la voie de synthèse *de novo* Kennedy et plus précisément, son étape limitante catalysée par la *PfCCT*.

Des filtres basés sur le ligand (LBVS) et sur la structure (SBVS) ont été utilisés pour tester virtuellement les chimiothèques commerciales que j'ai préparées. Pour chaque projet, des molécules ont été sélectionnées pour leurs scores de docking et les interactions qu'elles établissent avec les résidus clés de la protéine. En combinant la bioinformatique structurale et la chémoinformatique, nous avons identifié des inhibiteurs potentiels des deux cibles protéiques.

Titre en anglais

Identification of new antimalarial molecules by structural bioinformatics and cheminformatics. Application to two targets: *PfAMA1* and *PfCCT*.

Résumé en anglais

Human malaria is caused by five parasitic species of the genus *Plasmodium*, *P. falciparum* being the most deadly. Drug resistance of some parasite strains has been reported for commercial drugs. Vector mosquitoes are resistant to perythroid insecticides and no successful vaccine is available. This disease is a public and economic health issue for developing countries. My PhD projects investigate new treatments for malaria, by targeting two new proteins.

Apicomplexa parasites have developed a unique invasion mechanism involving a tight interaction formed between the host cell and the parasite surfaces called Moving Junction. The structural and functional characterization of the AMA1-RON2 complex pave the way for the design of low molecular weight compounds capable of disrupting the AMA1-RON2 assembly and thereby invasion. The parasite also needs phospholipids to build its membrane during the erythrocytic cycle. There are six times more phospholipids in infected erythrocytes compared to healthy ones. Our strategy is to inhibit the *de novo* Kennedy pathway and more precisely its rate-limiting step catalysed by the enzyme *PfCCT*.

Filters were used for ligand-based (LBVS) and structure-based virtual screening (SBVS) of commercial chemical databases that I have prepared. For each project, molecules were selected in terms of their docking scores and their interactions with key active site residues. By combining structural bioinformatics and cheminformatics, we identified potential inhibitors of the two protein targets.

Mots-clés/Keywords

Paludisme, Criblage Virtuel, AMA1, CCT, chémoinformatique, bioinformatique structurale
Malaria, Virtual Screening, AMA1, CCT, cheminformatics, structural bioinformatics

Intitulé et adresse du laboratoire

UMR7275 CNRS IPMC (Institut de Pharmacologie Moléculaire et Cellulaire)
660 Route des Lucioles, 06560 Valbonne, France