Structural and functional investigation of the C-terminal intrinsically disordered fragment of ErbB2

Louise Pinet

▶ **To cite this version:**

Louise Pinet. Structural and functional investigation of the C-terminal intrinsically disordered fragment of ErbB2. Structural Biology [q-bio.BM]. Université Paris Saclay (COmUE), 2019. English. NNT : 2019SACLS375 . tel-02391272

HAL Id: tel-02391272
https://theses.hal.science/tel-02391272

Submitted on 3 Dec 2019

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 PARIS-SACLAY

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

NNT : 2019SACLS375

Thèse de doctorat

Structural and functional investigation of the C-terminal intrinsically disordered fragment of ErbB2

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

Ecole doctorale n°569 Innovation thérapeutique du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Biochimie et biologie structurale

Thèse présentée et soutenue à Gif-sur-Yvette, le 17 octobre 2019, par

LOUISE PINET

Composition du Jury :

Dominique Guianvarc'h Professeure des Universités, Université Paris-Sud (ICMMO)	Présidente
Olivier Lequin Professeur des universités, Université Pierre et Marie Curie (LBM)	Rapporteur
Malene Ringkøbing Jensen Directrice de recherche, CNRS (IBS)	Rapporteure
Mounira Amor-Guéret Directrice de recherche, CNRS (Institut Curie - Stress génotoxique et cancer)	Examinatrice
Ali Badache Directeur de recherche, INSERM (CRCM)	Examinateur
Françoise Guerlesquin Directrice de recherche, CNRS (LISM)	Examinatrice
Carine van Heijenoort Directrice de recherche, CNRS (ICSN)	Directrice de thèse
Nadine Assrir Chercheuse, CNRS (ICSN)	Co-encadrante de thèse

Contents

Acknowledgments **i**

Abbreviations **iv**

Résumé en français **viii**

Introduction **1**

Context of this work **4**

1 Intrinsically disordered proteins (IDPs) 4
1.1 Discovery and assessment of biological relevance of IDPs 4
1.1.1 Historical perspective 4
1.1.2 Current questions and challenges 5
1.2 Diversity of disordered states 6
1.3 The sequence-(dis)order relationship 7
1.3.1 Origin of disorder: composition bias in IDPs 7
1.3.2 The unique role of prolines in disorder 8
1.3.3 Predicting disorder 9
1.4 Functional specificities 9
1.4.1 Interaction modes of IDPs 10
1.4.2 IDPs in signaling 15
1.4.3 IDPs and diseases 15
1.5 Structural biology of IDPs 16
1.5.1 The goal: reconstruction of conformational ensembles 16
1.5.2 NMR, the choice atomic-resolution method to study flexible objects . . . 17
1.5.3 Complementary methods 19
2 The C-terminal tail of ErbB2 (CtErbB2) 21
2.1 The ErbB/EGFR/HER family of receptor tyrosine kinases 21
2.1.1 Cellular function (and disfunction) of ErbBs 22
2.1.2 Overall mechanism of signal transduction 23
2.1.3 The extracellular domain (ECD), ligand binding, and dimerization 24
2.1.4 The transmembrane helix (TMH) and N-terminal end of the juxtamembrane region (JMA) 26
2.1.5 The tyrosine kinase domain and its activation 27
2.1.6 The C-terminal tail (Ct), regulation and signal transduction 29
2.2 ErbB2/neu/HER2, a singular ErbB 33

2.2.1 ErbB2 and cancer . . . 34
2.2.2 ErbB2 conformation is ligand-independent and dimerization prone . . . 35
2.3 CtErbB2 . . . 37
2.3.1 Role of each phosphotyrosine (pY) in signal transduction . . . 37
2.3.2 Regulation of kinase activity . . . 40
2.3.3 Known structural features of CtErbB2 . . . 40
3 Grb2, a major adaptor protein in ErbB2-dependent pathways . . . 43
3.1 Grb2 in ErbB2 signaling . . . 43
3.2 General description of SH2 and SH3 domains . . . 44
3.2.1 SH2 domains and phosphotyrosines . . . 45
3.2.2 SH3 domains and polyproline motifs . . . 45
3.3 Previous structural descriptions of Grb2 . . . 47
3.3.1 Individual domains . . . 48
3.3.2 Domain organization in full-length Grb2 . . . 49
3.3.3 Oligomerization . . . 51
3.4 Known interaction modes of Grb2 . . . 53
3.4.1 SH2-phosphotyrosine (pY) interaction . . . 53
3.4.2 SH3 domains interactions . . . 54
3.4.3 Interdependence of binding of SH2 and SH3 domains . . . 57

Experimental methods **58**
4 Protein expression and purification . . . 58
4.1 Grb2 and CtErbB2 expressed in *E. coli* . . . 58
4.1.1 Constructs and mutants . . . 58
4.1.2 Expression in *E. coli* . . . 61
4.1.3 Purification . . . 62
4.2 Generation of viral stocks for protein expression in insects cells using the Bac-to-bac® expression system . . . 64
4.2.1 Principle of protein expression in Sf9 cells using the baculovirus expression system . . . 64
4.2.2 Generation of the viral stock . . . 66
4.2.3 Determination of the viral stock titer by flow cytometry . . . 66
5 Nuclear Magnetic Resonance (NMR) . . . 68
5.1 NMR samples, conditions, and assignments . . . 68
5.1.1 Principles of heteronuclear sequential assignment of proteins . . . 68
5.1.2 Backbone and side-chains assignment of CtErbB2 . . . 69
5.1.3 Backbone assignment of Grb2 . . . 70
5.2 Chemical shift analysis . . . 72
5.3 ^{15}N relaxation studies . . . 73
5.3.1 Dynamic information given by ^{15}N relaxation parameters. Expectations for globular proteins and IDPs. . . . 73
5.3.2 NMR experiments and analysis . . . 74
5.4 Scalar couplings . . . 75
5.5 Residual dipolar couplings (RDCs) . . . 76
5.5.1 (Residual) dipolar couplings . . . 76
5.5.2 Expected RDCs for an IDP . . . 76

5.5.3 Protocol . . . 77
5.6 Paramagnetic relaxation enhancements (PREs) . . . 78
5.7 NMR titrations . . . 80
6 Other biochemical, biophysical and bioinformatic techniques . . . 81
6.1 Circular dichroism (CD) . . . 81
6.1.1 CD of folded and unfolded proteins . . . 81
6.1.2 Protocol . . . 82
6.2 Small-angle X-ray scattering (SAXS) . . . 82
6.2.1 Principle of SAXS . . . 82
6.2.2 Data acquisition . . . 83
6.2.3 Data analysis . . . 85
6.3 Generation of conformational ensembles: Flexible Meccano . . . 87
6.4 Sequence alignments and distance trees of mammalian CtErbB2 . . . 87

Results **88**
7 Characterization of isolated, unphosphorylated CtErbB2 . . . 88
7.1 CtErbB2 is an IDP . . . 88
7.1.1 CtErbB2 is predicted to be disordered from its sequence . . . 88
7.1.2 Experimental evidence of overall absence of folding . . . 88
7.2 NMR assignment of backbone and side chains . . . 91
7.3 Characterization of proline conformation . . . 91
7.4 Characterization of transient local structure . . . 93
7.5 Dynamics of CtErbB2 . . . 95
7.6 Long-range contact(s) in CtErbB2 . . . 97
7.7 Conservation of CtErbB2 features amongst mammals . . . 98
7.8 Our study reveals new potential mechanisms for the regulation of signal transduction 99
8 Behavior of Grb2 in solution . . . 106
8.1 NMR backbone assignment of all constructs, instability of the NSH3 domain and effect of the Histidine tag . . . 106
8.1.1 Constructs and assignment . . . 106
8.1.2 Influence of the His-tag on the stability of the NSH3 and on the interaction between the NSH3 and SH2 domains . . . 109
8.2 Chemical shifts comparison of domains . . . 112
8.3 Dynamics of Grb2 . . . 114
8.4 Domain organization determined by SEC-SAXS . . . 115
8.4.1 The NSH3SH2 bidomain organization corresponds to the organization in the crystal structure of full-length Grb2 . . . 115
8.4.2 The SH2CSH3 bidomain adopts multiple conformations of different R_g . . . 116
8.4.3 FL Grb2 also adopts multiple conformations with different R_g . . . 117
8.5 New model for Grb2 domain organization and oligomeric state . . . 119
8.5.1 Stability of the NSH3 and His-tag . . . 119
8.5.2 Oligomeric state of apo-Grb2 in solution . . . 121
8.5.3 Domain organization in Grb2 monomer . . . 122
8.5.4 Importance for Grb2 interactions and biological function . . . 123
9 Interaction between CtErbB2 and Grb2 . . . 124
9.1 Grb2 and unphosphorylated CtErbB2 . . . 124

9.2 Need for tyrosine phosphorylation of CtErbB2 . . . 127
9.3 Grb2 and peptides of CtErbB2 . . . 128
9.3.1 Unphosphorylated Y-PxxP peptide P1 . . . 128
9.3.2 Phosphorylated Y-PxxP peptide P2 . . . 129
9.4 Possible consequences for Grb2-mediated signal transduction . . . 130
9.4.1 Selectivity of CtErbB2 polyproline motif for the NSH3 of Grb2 . . . 132
9.4.2 Importance of the organization of Grb2 domains for its interactions . . . 132
9.4.3 Revisiting the model of signal transduction by Grb2 . . . 133
10 Towards the characterization of functional states of CtErbB2: challenges of structural and functional studies of tyrosine phosphorylation . . . 135
10.1 Challenges and existing strategies . . . 136
10.1.1 Site-resolved detection and study of (tyrosine) phosphorylation . . . 136
10.1.2 Production of phosphorylated samples with multiple sites: site-directed synthetic approaches and biological relevance . . . 138
10.2 Three methods to produce phosphorylated CtErbB2 . . . 140
10.2.1 CtErbB2 co-expressed with a tyrosine kinase in *E. coli* . . . 140
10.2.2 *In vitro* phosphorylation followed by NMR . . . 142
10.2.3 Intracellular domain of ErbB2 expressed in Sf9 cells using the baculovirus expression system . . . 144
10.2.4 Overcoming the heterogeneity problem: using Tyr to Phe point mutants . . . 148

Conclusion and perspectives **150**

References **153**

Appendices **179**
I M9 medium composition . . . 180
II 1H, 13C and 15N assignments of the C-terminal intrinsically disordered cytosolic fragment of the receptor tyrosine kinase ErbB2 . . . 181
III The intrinsically disordered C-terminal tail of the ErbB2 receptor exhibits transiently structured elements likely to have functional relevance . . . 186
IV Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data . . . 234

Acknowledgments

Tout d'abord je tiens à remercier Malene Ringkjøbing Jensen et Olivier Lequin d'avoir accepté d'être rapporteur-e-s de cette thèse. J'espère qu'ils prendront plaisir à la lire et à en discuter. Merci aussi aux autres membres de ce grand jury : Mounira Amor-Guéret, Ali Badache, Françoise Guerlesquin et Dominique Guianvarc'h.

Un merci bien sûr tout particulier à ma directrice et à ma co-encadrante de thèse, Carine van Heijenoort et Nadine Assrir. Merci de m'avoir acceptée pour un stage de Master qui ne devait "pas mener à une thèse", et de m'avoir subtilement réorientée. Merci de m'avoir fait découvrir le monde de la recherche dans (presque) tous ses aspects. Merci pour les discussions (scientifiques mais pas que) passionnantes, souvent animées mais bienveillantes, et surtout pour votre confiance. Merci aussi de vous être adaptées aux oscillations régulières mais souvent imprévisibles de mon niveau d'assurance, et de m'avoir poussée quand j'en avais besoin. Nadine, merci de m'avoir guidée dans le monde obscur de la biologie, de ma première purification d'ErbB2 à la culture cellulaire et à mon premier Western Blot. Carine, merci de m'avoir initiée à la RMN, et de m'avoir confié cet aspect du projet avec une confiance qui m'a parfois intimidée. Merci pour les balades en forêt, dîners à Madrid ou simples pauses devant le 23B. Personne ne s'étonnera vraiment que je vous remercie aussi de montrer au quotidien que les femmes ne sont pas les dernières à avoir leur mot à dire, en recherche et ailleurs.

Un grand merci aux personnes de l'équipe de RMN avec qui j'ai travaillé directement dans le cadre de ce projet : YingHui, Alan et Ndèye, qui a été une très bonne étudiante pour ma première expérience d'encadrement. Merci à François B. qui a sacrifié des heures de sommeil pour aligner des séquences. Merci à Anaïs, impliquée presque malgré elle et dont la rigueur a permis de "sauver" une partie de ma thèse. Merci à tou-te-s celles-eux dont le travail, parfois dans l'ombre, m'a permis de travailler dans des condi-

tions exceptionnelles : Annie pour toute la partie Wetlab, gérée avec flexibilité et efficacité; François G., Jean-François, Nelly, Ewen et Christina pour leur aide à divers niveaux aux spectros, toujours dans la bonne humeur; Alda pour les aspects administratifs, auxquels j'arriverai peut-être un jour à comprendre quelque chose; Guillaume et David pour l'informatique, sans laquelle on serait bien démuni-e-s.

Merci aux personnes extérieures à l'ICSN avec qui j'ai eu le plaisir de travailler et d'échanger sur ce projet : Ali Badache et Lia N'Guyen dans le cadre de l'ANR qui a lancé le projet dans le laboratoire, et Françoise Guerlesquin, qui a continué bien après la fin de l'ANR à m'aider à garder le cap. Je remercie aussi Dominique Durand et Aurélien Thureau pour les expériences de SAXS et les discussions. Merci à Mickaël Bourge pour le FACS, Arnaud Poterszman et Simon Pichard pour les baculovirus. Merci également aux personnes avec qui j'ai collaboré sur d'autres projets, et qui ont jugé que j'avais quelque chose à y apporter : Maud Chan-Yao-Chong et Tâp Ha-Duong, Sophie Zinn-Justin et Philippe Cuniasse.

Un grand merci à toutes les personnes croisées (et souvent un peu plus) pendant ces trois ans et demi dans le labo, et qui ont fait de cette thèse une période incroyablement riche en partage.
Merci à tout le groupe de "jeunes" pour le soutien mutuel, les verres au Val Fleury, et autres voyages mouvementés dans des contrées lointaines et froides. Merci à Ludmilla d'avoir été la grande soeur de tout le groupe de thésard-e-s arrivé-e-s en 2016-2017. Merci à Camille d'avoir sacrifié une grande partie de ses premières minutes au labo le matin (et d'autres) pour un thé et un debrief. Merci à Gus pour son soutien et ses blagues, qui traversent maintenant l'Atlantique. Merci à Florian, qui, quoi que j'écrive, lira dans ces remerciements ce qu'il a envie d'y lire. Merci à Arthur, Hans, Christophe, Claire-Marie, Corentin, Anaïs et Luíza pour les pauses cafés et les discussions de couloir dont on se demandait parfois si elles auraient une fin.
Merci aux permanent-e-s qui apportent chacun-e de leur personnalité à ce laboratoire, et au bâtiment 23B : Alda, Guillaume et Véro pour leur franc-parler et leurs rires; Nelly pour sa gentillesse, son écoute et son caractère plus trempé qu'il n'y parait; François G. pour sa fausse discrétion, sa culture cinématographique et sa disponibilité; François B. pour sa culture scientifique, philosophique et générale qui parait sans limite et mène à des discussions passionnantes; Ewen pour sa curiosité scientifique et sa "tchatche"; Christina pour sa patience et son efficacité; Annie pour son immense gentillesse et son écoute; Nadine pour son écoute et sa sensibilité; Carine pour sa capacité à nous pousser à aller toujours plus loin, son ouverture, et ses conseils de lecture. Un immense merci à Jean-Nicolas, qui, malgré son départ de l'ICSN, continue d'être un repère,

et une source intarissable de confiance.

Merci à tou-te-s celles-eux qui m'ont soutenue en dehors du labo. Merci à Nicolas de m'avoir fait partager sa passion pour la RMN des protéines, et de m'avoir accompagnée au quotidien pendant le début de ma thèse. Merci à Diane, a.k.a Bibi, pour le soutien constant à tous les niveaux, les grandes réflexions existentielles, et les échanges au quotidien. Merci à Charline et Vincent pour les thés-potins, les muffins au chocolat quand ça allait moins bien, les Yakitoris et les discussions sur la recherche. Merci à "Spampom" et "Spampole" pour la danse, les rires, et l'empowerment. Merci à Pépé pour les soirées tisane, à Marius pour m'avoir accompagnée dans mes premiers trajets en RER B, et à Jack&Maillard et toute l'équipe retrouvée autour des films de Sharks. Merci aussi à Maud pour les déambulations au Louvre ou ailleurs, et à Maxence pour les dîners à Massy. Enfin, merci à mes parents, à mes frère et soeur et à mes grands-parents de ne pas raccrocher quand je parle pendant de longues minutes de protéine, de RMN et de labo, et du fait que "je suis faaaaatiguéééééee".

Merci.

Abbreviations

Chemicals, reagents, media and buffers

2xYT	2x Yeast extract-Tryptone
ATP	Adenosine triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
IAP	3-(2-Iodoacetamido)-PROXYL
LB	Lysogeny broth (sometimes referred to as Luria-Bertani medium)
MES	2-(N-morpholino)ethanesulfonic acid
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
PBS	phosphate buffered saline
PE	phycoerythrin
PIPES	piperazine-N,N'-bis(2-ethanesulfonic acid)
TCEP	tris(2-carboxyethyl)phosphine

Techniques, methods, experiments

AUC	Analytical ultracentrifugation
BEST	Band-selective excitation short-transient
CD	Circular dichroism
DLS	Dynamic light scattering
FSC	Forward-scattered light
HDX	Hydrogen-deuterium exchange
HSQC	Heteronuclear single quantum correlated spectroscopy
ITC	Isothermal Titration Calorimetry
MALDI-TOF	Matrix-assisted laser desorption ionization - Time of flight
MALS	Multi-angle light scattering
MS	Mass spectrometry
MST	Microscale thermophoresis
NOe	Nuclear Overhauser effect
NMR	Nuclear magnetic resonance
PRE	Paramagnetic relaxation enhancement
RDC	Residual dipolar coupling
SAXS	Small-angle X-ray scattering
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SSC	Side-scattered light
TROSY	Transverse relaxation-optimized spectroscopy

ErbB family-related abbreviations

CtErbB2	C-terminal tail of ErbB2
ECD	Extracellular domain
EGFR	Epidermal growth factor receptor
HER	Human epidermal growth factor receptor
ICD	Intracellular domain
JM	Juxtamembrane
MAPK	Mitogen-activated protein kinase(s)
RTK	Receptor tyrosine kinase
TKD	Tyrosine kinase domain
TMH	Transmembrane helix

Miscellaneous

AD	Alzheimer's disease
BMRB	Biological Magnetic Resonance Bank
CAP	Cancer-associated protein
δ	Chemical shift
FDA	United States Food and Drug Administration
FL	Full-length
IDP	Intrinsically disordered protein
MOI	Multiplicity of infection
MoRF	Molecular recognition feature
OD	Optical density
PD	Parkinson's disease
PDB	Protein Data Bank
PFU	Plaque-forming unit
PPII	Polyproline II
pS/pT/pY	Phosphoserine/phosphothreonine/phosphotyrosine
PTB domain	Phosphotyrosine binding domain
PTM	Post-translational modification
R_g	Radius of gyration
SH2/SH3 domain	Src homology 2/3 domain
SLiM	Short linear motif
SSP	Secondary structure propensity
TEV	Tobacco Etch Virus

Résumé en français

Introduction

Les protéines intrinsèquement désordonnées (IDPs), qui n'ont pas de structure tridimensionnelle stable, ont longtemps été considérées comme d'intérêt biologique limité. Depuis une vingtaine d'années, leur proportion élevée dans les protéomes, notamment eucaryotes (Ward et al., 2004), et leur rôle dans de nombreux processus physiologiques et pathologiques (Wright and Dyson, 1999) sont reconnus. En particulier, le désordre est commun dans les processus de signalisation (Iakoucheva et al., 2004). De récents développements méthodologiques, notamment en résonance magnétique nucléaire (RMN), permettent l'étude de ces protéines désordonnées, en parallèle de l'émergence de nouveaux concepts associés.

Dans les travaux présentés ici, nous nous intéressons à la région C-terminale désordonnée de ErbB2/HER2/neu (que nous appelons CtErbB2). ErbB2 fait partie de la famille de récepteurs tyrosine kinase des ErbBs, dont EGFR, ou ErbB1, est le membre le plus étudié et est utilisé comme modèle pour l'étude de leur mécanisme de transduction du signal (Kovacs et al., 2015b) : la liaison d'un ligand protéique au domaine extracellulaire du récepteur permet son homo- ou hétéro-dimérisation, l'activation de son domaine tyrosine kinase intracellulaire et la phosphorylation de sa région C-terminale. Cette région, par le biais de ses tyrosines phosphorylées, constitue une plateforme d'interaction avec de nombreuses protéines adaptatrices. Ces interactions activent les voies de signalisation correspondantes, contrôlant la division, la différentiation et la motilité cellulaires et la résistance à l'apoptose (Yarden and Sliwkowski, 2001). Dans ce modèle, deux aspects en particulier restent mal compris : les mécanismes spécifiques à chaque ErbB et les mécanismes d'interaction entre la queue C-terminale et les protéines adaptatrices.

Plus de 20% des cancers du sein sont associés à une surexpression de ErbB2, et cette forme dite HER2 + est particulièrement agressive (Slamon et al., 1987). Contrairement aux trois autres membres de la famille, l'activation d'ErbB2 ne nécessite pas l'association d'un ligand à son domaine extracellulaire (Cho et al.,

2003). La capacité intrinsèque d'ErbB2 à dimériser et l'activation de son domaine kinase n'étant donc pas contrôlées par un signal extracellulaire, la régulation de la phosphorylation et des interactions de sa queue C-terminale est d'autant plus cruciale. Plusieurs éléments suggèrent que cette région est globalement désordonnée (Keppel et al., 2017), mais aucune étude à haute résolution n'a été menée. Pourtant, son rôle dans la régulation de la transduction du signal en fait un modèle pour l'étude du lien entre désordre et fonction ainsi qu'une cible thérapeutique intéressante.

Mon travail de thèse a consisté à étudier dans un premier temps les caractéristiques structurales et dynamiques de CtErbB2 isolée et non phosphorylée. Dans un deuxième temps, je me suis intéressée à Grb2, un partenaire majeur de CtErbB2. Cette protéine adaptatrice contenant trois domaines (SH3-SH2-SH3) est connue notamment pour interagir avec une tyrosine phosphorylée de CtErbB2 (Dankort et al., 1997) et la connecter, en interagissant avec d'autres protéines par ses domaines SH3, à la voie de signalisation Ras/MAPK (Lowenstein et al., 1992). J'ai donc étudié l'organisation modulaire de Grb2 en solution, puis son interaction avec CtErbB2. Enfin, j'ai mis en place différentes méthodes permettant d'étudier CtErbB2 dans un état plus fonctionnel : dans son état phosphorylé, et dans le contexte de la région intracellulaire entière (avec le domaine kinase).

Contexte des travaux présentés

Les protéines intrinsèquement désordonnées (IDPs)

Le désordre dans les protéines Contrairement au premier dogme de la biologie structurale, et au concept de "clé-serrure" de Fischer (1894), toutes les protéines n'ont pas une structure tridimensionnelle unique déterminant leur fonction spécifique. Un grand nombre de protéines possèdent au moins 30 résidus consécutifs ne présentant pas de structure stable, et de telles régions sont dites intrinsèquement désordonnées. Chez l'homme, il est estimé que c'est le cas de 44% des protéines (Oates et al., 2013). La longueur des segments désordonnés, ainsi que leur composition en structures secondaires et leur compacité, sont fortement variables, et on parle donc de "continuum du désordre" (Uversky et al., 2005; Dyson and Wright, 2005).

Désordre et séquence Comme la structure d'une protéine globulaire, le niveau de désordre d'une IDP est largement dicté par sa séquence. Les IDPs sont globalement pauvres en résidus hydrophobes, enrichies en résidus chargés, polaires, en prolines et parfois en glycines et alanines (Habchi et al., 2014). Les séquences répétées sont aussi fréquentes dans les IDPs. Ces spécificités ont permis le développement de programmes

de prédiction de désordre à partir de la séquence d'une protéine.

Les prolines sont des acides aminés uniques, jouant un rôle particulier dans les IDPs (Mansiaux et al., 2011). Leur chaîne latérale cyclique entraine l'absence du proton amide du squelette, empêchant la formation de liaisons hydrogène nécessaires à la stabilisation d'hélices α ou de feuillets β. De plus, les contraintes générées par le cycle sont fortes. La proportion d'isomère *cis* de la liaison peptidique X-Pro (plusieurs %) est bien plus élevée que pour les autres acides aminés (moins de 1%) (Alderson et al., 2018). La cinétique d'interconversion avec la forme *trans* est très lente, souvent plusieurs minutes voire plusieurs heures (Reimer et al., 1998). De plus, les prolines favorisent un type particulier d'hélices auxquelles elles donnent leur nom : les hélices polyproline II (PPII). Ces hélices triangulaires de 3 résidus par tour, présentées en figure 3, sont étendues (3.1 Å par résidu) et bloquent les liaisons peptidiques en conformation *trans*. Elles sont fréquemment formées lors d'interactions de motifs riches en prolines avec des domaines SH3 ou WW (Macias et al., 2002).

Fonctions de IDPs Les spécificités physico-chimiques des IDPs sont liées à des spécificités fonctionnelles : elles sont sur-représentées lorsqu'une grande flexibilité est avantageuse (rôle de chaîne entropique), ou lorsque qu'une grande accessibilité et adaptabilité dans les interactions sont requises. Les complexes impliquant des IDPs peuvent eux-même présenter différents niveaux de désordre et différentes multiplicités de sites. Dans le cas le plus couramment décrit, la liaison est couplée au repliement local du partenaire désordonné au niveau d'un motif en interaction avec un site spécifique d'un partenaire ordonné. À l'autre extrême, un partenaire (ou les deux) peut conserver un grand niveau de désordre (on parle alors de "fuzzy complex"), et adopter de multiples conformations en échange entre plusieurs sites d'interaction. La notion d'allostérie, soit la régulation de l'activité d'un site par liaison d'un ligand à un autre site éloigné, liée originellement à l'observation de changements structuraux propagés d'un site structuré à l'autre, a aujourd'hui été étendue afin de prendre en compte l'allostérie liée à des changements de dynamique et à des transitions ordre/désordre. En lien avec ces caractéristiques de liaison des IDPs, leurs interactions sont souvent d'affinité modérée, tout en conservant une grande spécificité. L'association et la dissociation étant généralement rapides, les interactions impliquant des IDPs permettent une grande adaptabilité. Il n'est donc pas étonnant que les protéines désordonnées soient courantes dans les mécanismes de signalisation (Iakoucheva et al., 2004), et notamment dans les récepteurs comme ErbB2 (Kjaergaard and Kragelund, 2017).

Étude structurale du désordre La méthode expérimentale de choix pour étudier les IDPs avec une résolution atomique est la résonance magnétique nucléaire (RMN). Elle permet d'obtenir des informations à la fois structurales et dynamiques (de la centaine de ps à la s ou plus), ce qui revient à sonder l'ensemble conformationnel exploré par la protéine. Elle peut être combinée à des informations plus globales sur la forme et la taille moyenne de l'ensemble, obtenues par SAXS, et à des informations à l'échelle de la molécule unique (par FRET, spectroscopie de force ou nanopores uniques par exemple). Enfin, les données expérimentales peuvent être complétées par des études *in silico*, en générant (et sélectionnant) des ensembles conformationnels ou des trajectoires de dynamique moléculaire cohérents avec ces données, pour explorer des mécanismes possibles.

La queue C-terminale intrinsèquement désordonnée de ErbB2 (CtErbB2)

Les ErbBs Les quatre protéines ErbB (ErbB1/HER1/EGFR, ErbB2/HER2/neu, ErbB3/HER3 and ErbB4/HER4) sont des récepteurs tyrosine kinase (RTKs), présentés en figure 10. ErbB1 a été le premier RTK découvert, et sert classiquement de modèle pour comprendre le mécanisme de transduction du signal par les ErbBs. Les quatre ErbBs ont une structure globale commune : une région extracellulaire divisée en quatre domaines, une hélice transmembranaire, une région intracellulaire juxtamembranaire, un domaine tyrosine kinase et une queue C-terminale contenant des sites de phosphorylation tyrosines. Les voies de signalisation activées par les ErbBs comprennent notamment : la voie Ras/MAPK, la voie Akt/PI3K, la voie des Src kinases, la voie des STAT et la voie de MEMO (Yarden and Sliwkowski, 2001; Marone et al., 2004). Ces voies contrôlent la division, la différentiation, l'adhésion et la motilité cellulaire, ainsi que la résistance à l'apoptose (Pinkas-Kramarski et al., 1996; Jeon et al., 2015; Chausovsky et al., 2000; Zaoui et al., 2008; Giani et al., 1998). Les ErbBs sont toutes impliquées dans des processus physiologiques tels que le développement et la fonction de différents organes (Sebastian et al., 1998; Lee et al., 1995) et dans des processus pathologiques, en particulier l'apparition de cancers (Yarden and Sliwkowski, 2001; Yarden and Pines, 2012). Cependant, les ErbBs ont des spécificités fonctionnelles : leurs ligands, de la famille des facteurs de croissance épidermiques, diffèrent et ErbB2 n'a pas de ligand endogène identifié ; leur activité kinase varie et celle d'ErbB3 est très faible ; les processus physiologiques et pathologiques, notamment les types de cancers, dans lesquels elles sont impliquées sont différents (Yarden and Sliwkowski, 2001).

Le mécanisme général de signalisation par les ErbBs, déterminé principalement grâce aux études sur EGFR, se décompose en cinq étapes impliquant différentes parties des récepteurs (Kovacs et al., 2015b; Purba et al., 2017) et est récapitulé en figure 19 :

- La liaison d'un ligand à la région extracellulaire du récepteur entraine un changement structural dans l'organisation de ses quatre domaines (Ferguson et al., 2003). Alors que la forme libre inhibe la formation de dimères actifs pour la signalisation, la forme liée rend accessible la zone extracellulaire de dimérisation. Des dimères d'ErbBs sans ligands ont par ailleurs été observés (Tao and Maruyama, 2008), mais n'impliquent pas la dimérisation de la région extracellulaire et n'entrainent pas de signalisation.

- Dans le dimère actif ainsi formé, les hélices transmembranaires dimérisent dans une conformation où leurs extrémités intracellulaires sont éloignées et où l'hélice juxtamembranaire immédiatement adjacente (JMA) forme un dimère antiparallèle (Bragin et al., 2016).

- Le domaine tyrosine kinase est activé par contacts intermoléculaires dans un dimère asymétrique, où une kinase inactive dite "activatrice" active l'autre, appelée "réceptrice" (Zhang et al., 2006). Dans ce dimère asymétrique le segment juxtamembranaire le plus proche de la kinase (JMB) réceptrice stabilise l'interaction en interagissant avec la kinase activatrice (Red Brewer et al., 2009; Jura et al., 2009). Ce mécanisme est très différent de l'activation "classique" des RTKs par transphosphorylation du domaine kinase.

- Le domaine kinase ainsi activé phosphoryle la queue C-terminale en *trans* (Qian et al., 1994) et/ou en *cis*, par un mécanisme inconnu.

- Les tyrosines phosphorylées recrutent des protéines adaptatrices contenant majoritairement des domaines SH2 ou PTB, et ce recrutement active les voies de signalisation associées. La queue C-terminale et sa phosphorylation ont aussi été montrées comme étant impliquées dans la régulation de l'activité des domaines kinases (Gill et al., 2017). Cette observation est cohérente avec l'interaction observée entre ces deux régions dans les structures cristallographiques des domaines kinases, qui contiennent le début de la queue C-terminale (Lemmon et al., 2014). Une hélice, appelée AP-2 chez EGFR, située au début de la queue C-terminale, est notamment observée dans les structures de la kinase non activée. Elle interagit de façon intramoléculaire et intermoléculaire avec les domaines kinases, stabilisant un dimère symétrique inactif.

  Ces régions structurées dans les cristaux de domaines kinases sont les seules informations à haute résolution connues sur la queue C-terminale des récepteurs ErbBs. Le reste de la queue a été montré comme étant globalement désordonné (Keppel et al., 2017) et dynamique (Lee and Koland, 2005;

Lee et al., 2006), avec un comportement dépendant de l'état de phosphorylation (Lee et al., 2006; Bishayee et al., 1999).

ErbB2 et sa queue C-terminale (CtErbB2) ErbB2 a plusieurs particularités fonctionnelles dans la famille : elle n'a pas de ligand endogène connu et est capable d'activer des voies de signalisation de façon ligand-indépendante (Di Fiore et al., 1987; Brennan et al., 2000). Elle est surexprimée dans plus de 20% des cancers du sein, dits HER2 + (Slamon et al., 1987). Ces cancers sont particulièrement agressifs (Press et al., 1993). Ces particularités fonctionnelles sont associées à des particularités structurales, et notamment l'existence d'une unique conformation de sa région extracellulaire, celle permettant la dimérisation (Cho et al., 2003). Il manque donc à ErbB2 un mécanisme de régulation par rapport aux autres ErbBs, et la régulation de sa phosphorylation et de ses interactions est donc particulièrement importante.
La queue C-terminale d'ErbB2 contient 5 sites d'autophosphorylation connus (Hazan et al., 1990; Segatto et al., 1990), appelés Y_A, Y_B, Y_C, Y_D and Y_E, dont le rôle dans la transduction du signal a été étudié par Dankort et al. (1997, 2001a,b). La phosphorylation de Y_A a été montrée comme ayant un rôle inhibiteur de la transduction du signal, levé seulement par la phosphorylation de Y_D. La phosphorylation des quatre autres sites est activatrice de la signalisation, grâce à l'interaction avec différents partenaires, dont : Grb2 pour Y_B, Crk pour Y_C, Shc et MEMO pour Y_D, Dok-R pour Y_E.
En dehors de la possible formation de structures locales lors d'interactions, et de l'hélice AP-2 visible dans la structure cristallographique du domaine kinase de ErbB2 comme chez EGFR, on s'attend à ce que CtErbB2 soit globalement désordonnée, comme suggéré par l'étude de Keppel et al. (2017). Le but de notre étude est de caractériser à haute résolution la structure et la dynamique de CtErbB2, afin de comprendre comment sont régulées les interactions avec ses partenaires dans le cadre de l'activation des voies de signalisation. Nous avons en particulier étudié le partenaire Grb2.

Grb2, une protéine adaptatrice essentielle des voies de signalisation de ErbB2

Grb2 est une protéine modulaire purement adaptatrice, sans rôle catalytique, et dont la fonction repose exclusivement sur les interactions protéine-protéine. Elle a une organisation de type SH3-SH2-SH3. Elle lie différents RTKs (dont ErbB2) sur leurs phosphotyrosines par son domaine SH2 et fait le pont principalement avec la voie Ras/MAPK en recrutant d'autres protéines via ses domaines SH3 N- et C-terminaux (notés respectivement NSH3 et CSH3) (Lowenstein et al., 1992; Cheng et al., 1998). Grb2 est essentielle pour l'embryogenèse et est impliquée dans la tumorigenèse, notamment celle médiée par ErbB2 (Cheng

et al., 1998; Xie et al., 1995; Gril et al., 2007).

Les trois domaines de Grb2 Les domaines SH2 sont des domaines d'une centaine de résidus, ayant une structure canonique constituée d'un feuillet β central et de deux hélices, positionnées de chaque côté. Une telle structure est présentée en figure 25. Les domaines SH2 sont des domaines de liaison aux phosphotyrosines (pY, en contact avec une des faces du feuillet), et leur spécificité dépend des résidus positionnés en N-terminal de pY (en contact avec l'autre face du feuillet) (Liu et al., 2012). Le domaine SH2 de Grb2 a une structure globale canonique (Thornton et al., 1996), avec une spécificité pour les ligands de type pYxNx. La position de la chaîne latérale d'un tryptophane de Grb2 (W121) contraint la conformation de la chaîne de ses ligands en N-terminal de pY. Ce tryptophan force la formation d'un coude β (Marengere et al., 1994; Rahuel et al., 1996; Papaioannou et al., 2016) montré en figure 31, alors que les ligands des domaines SH2 adoptent classiquement une structure étendue.

Les domaines SH3 sont plus petits, généralement entre 50 et 75 résidus, et sont constitués de deux feuillets β antiparallèles (voir figure 26). Ils lient différents types de motifs, le plus couramment riches en prolines formant des hélices polyproline PPII, et notamment les motifs RxxPxxP (classe I) et PxxPxR (classe II) (Musacchio, 2002). Ces deux classes de peptides se lient avec des orientations opposées, permettant l'interaction des motifs xP (ou Px) avec des résidus aromatiques et une proline du SH3, et de l'arginine avec un résidu acide (E ou D) du SH3. Les interactions secondaires (en dehors de ces deux interactions principales) sont aussi courantes et participent à la spécificité de chaque SH3. Les deux SH3 de Grb2 ont une structure globale canonique (Goudreau et al., 1994; Kohda et al., 1994), présentée en figure 32c et 32d. Cependant, le domaine NSH3 a été montré comme échangeant entre sa forme repliée et une forme largement dépliée dans des conditions physiologiques (Goudreau et al., 1994). Les deux domaines SH3 ont une préférence pour les ligands de classe II plutôt que de classe I, mais le CSH3 peut aussi interagir avec des ligands de type RxxK contenant des prolines (Musacchio, 2002; Lewitzky et al., 2001; Harkiolaki et al., 2009). Sos, le facteur d'échange nucléotidique de Ras, peut se lier aux deux SH3, avec une préférence pour le NSH3 (McDonald et al., 2009). Les protéines Gab, elles, se lient au domaine CSH3 (Lewitzky et al., 2001; Harkiolaki et al., 2009).

Grb2 entière La seule structure de Grb2 entière est une structure cristallographique, sans ligand, donnée en figure 29 (Maignan et al., 1995). Dans cette structure, les deux domaines SH3 de Grb2 sont en interaction électrostatique, mais leurs sites d'interaction restent accessibles. La seule étude de Grb2 entière en solution a été effectuée avec des ligands peptidiques à la fois pour les domaines SH3 et le domaine SH2 (peptides extraits de Sos et d'EGFR respectivement) (Yuzawa et al., 2001). Cette étude montre une absence d'interaction notable entre les domaines, contrairement aux données cristallographiques. Aucune donnée n'est actuellement publiée sur le comportement de Grb2 en solution en l'absence de ligands.

L'état oligomérique de Grb2 est sujet à controverse. La structure cristallographique de Grb2 (Maignan et al., 1995) montre un dimère dans lequel les domaines SH2 et CSH3 interagissent (voir figure 30a), alors qu'en solution avec ligands peptidiques Grb2 est purement monomérique (Yuzawa et al., 2001). D'autres études conduisent à des constantes de dissociation du dimère de l'ordre du μM (McDonald et al., 2008b; Lin et al., 2012; Ahmed et al., 2015). Des études du domaine SH2 isolé ont aussi montré un dimère croisé métastable de ce domaine ("domain-swapping"), où l'hélice C-terminale est échangée entre deux monomères d'un dimère (Schiering et al., 2000; Nioche et al., 2002; Benfield et al., 2007; Papaioannou et al., 2016). Cependant, un tel échange dans la protéine entière ou *in vivo* n'a jamais été démontré.

Bien que les liaisons des domaines SH2 et SH3 à des peptides différents soient indépendantes les unes des autres (Cussac et al., 1994; Lemmon et al., 1994), la situation semble différente lorsque les ligands sont protéiques et non simplement peptidiques (Ravichandran et al., 1995). De plus, la liaison à un partenaire contenant à la fois un motif phosphotyrosine et un motif polyproline n'a jamais été étudiée structuralement, bien que la liaison de Grb2 à EGFR et ErbB2 *in cellulo* semble requérir à la fois les domaines SH2 et SH3 (Gill et al., 2017; Xie et al., 1995). De nombreuses informations sont donc nécessaires pour mieux comprendre l'interaction de Grb2 avec ErbB2 : l'arrangement des domaines en solution dans l'état libre, ainsi que la structure et la dynamique des complexes Grb2-ErbB2.

Méthodes expérimentales

Production et purification des protéines recombinantes étudiées

Expression dans *E. coli* et purification CtErbB2 (résidus 988-1255 de ErbB2) et Grb2 (domaines SH2, NSH3 et CSH3 respectivement isolés, bidomaines NSH3SH2 et SH2CSH3, et protéine entière) ont été produits de manière recombinante dans *E. coli*, selon des protocoles classiques. Selon les expériences, CtErbB2 était non marquée, marquée [U-^{15}N] ou marquée [U-^{15}N ; U-^{13}C], tout comme les constructions NSH3,

SH2, CSH3 et SH2CSH3 de Grb2. Le bidomaine NSH3SH2 ainsi que Grb2 entière ont été produits marqués [U-^{15}N ; U-^{13}C ; U-80% ^{2}H].

CtErbB2 a été exprimée sous la forme d'une protéine de fusion contenant en N-terminal une étiquette histidine pour faciliter la purification, suivie d'un domaine lipoyl permettant d'éviter les phénomènes d'agrégation et d'un site de clivage à la protéase TEV pour pouvoir obtenir la protéine sans étiquette. La protéine finale contenait seulement quatre résidus (GSHM) supplémentaires par rapport à la séquence native. Toutes les constructions de Grb2 ont été produites avec une étiquette histidine N-terminale suivie d'un site de clivage à la TEV. Après clivage, chaque construction contenait deux résidus (GA) supplémentaires par rapport à la séquence native des domaines correspondants.

Pour CtErbB2 comme Grb2, la purification a été effectuée par lyse à haute pression, chromatographie d'affinité sur colonne de nickel puis clivage à la TEV. La protéase utilisée contenait elle-même une étiquette histidine. Une seconde chromatographie d'affinité sur nickel a permis d'éliminer la TEV, les étiquettes clivées et les protéines non clivées. Une chromatographie d'exclusion stérique (SEC) finale a permis d'obtenir des protéines d'une pureté (estimée par SDS-PAGE) supérieure à 95 %, et d'intégrité vérifiée par spectrométrie de masse. Pour les constructions de Grb2 dans lesquelles l'étiquette histidine a été conservée, le protocole était le même en supprimant l'étape de clivage et la seconde chromatographie d'affinité.

Expression en cellules d'insectes La région intracellulaire entière d'ErbB2 (résidus 676-1255) a été produite en utilisant le système d'expression baculovirus Bac-to-bac®, en suivant les recommandations du protocole fourni par Invitrogen. Ce système d'expression est basé sur l'infection de cellules d'insectes (SF9) par un baculovirus, le gène d'intérêt ayant préalablement été inséré dans l'ADN viral par transposition depuis un plasmide bactérien. Le titre des stocks viraux obtenus a été déterminé par infection de cellules SF9, reconnaissance des cellules infectées par un anticorps anti-gp64 (protéine d'enveloppe virale) fluorescent, et détermination du pourcentage de cellules infectées par cytométrie de flux. La multiplicité d'infection (MOI, le ratio du nombre de virus sur le nombre de cellules), et donc le titre viral initial, a pu être retrouvée en considérant que le pourcentage de cellules effectivement infectées suit une loi de Poisson. Nous avons obtenu des titres viraux usuels de l'ordre de 10^8 PFU/mL.

Résonance magnétique nucléaire (RMN)

Les études par RMN de CtErbB2 ont été effectuées dans un tampon MES (40 mM MES pH 5.6, 200 mM NaCl, 2 mM TCEP) à 298 K, majoritairement à 950 MHz pour avoir la meilleure résolution possible. Les études de Grb2 ont été principalement effectuées dans un tampon PIPES (40 mM PIPES pH 7.2, 150 mM NaCl, 2 mM TCEP) à 308 K, majoritairement à 600 MHz, champ auquel le plus de pics sont observables à cause de phénomènes d'échange. Les études d'interaction par RMN ont été effectuées à 298 K dans le tampon PIPES.

Attributions L'attribution du squelette de CtErbB2 et Grb2 a été principalement effectuée à l'aide d'expériences tridimensionnelles d'attribution séquentielle basées sur les couples N-H amides. Les prolines n'ayant pas de proton amide, des expériences de type HCAN et CON (cette dernière en détection carbone), ainsi que le spectre HSQC du mutant Q161A, ont été utilisés pour l'attribution des régions riches en prolines de CtErbB2. Des expériences HADAMAC (Lescop et al., 2008) ont permis le transfert de l'attribution des HSQC ^{1}H-^{15}N de CtErbB2 entre différentes conditions (mutations, pH, température...). L'attribution des chaines latérales de CtErbB2 a été effectuée grâce à des expériences H(C)C(CO)NH, (H)CC(CO)NH, HBHANH et HBHA(CO)NH.

Études des déplacements chimiques Les déplacements chimiques sont informatifs de l'environnement local du noyau observé. Pour CtErbB2, ils ont été utilisés pour calculer des propensions de structures secondaires formées, grâce au programme SSP (Marsh et al., 2006). Pour Grb2, les structures secondaires ont été déterminées grâce à TALOS+ (Shen et al., 2009) et comparées aux structures secondaires attendues d'après les structures déjà publiées.

Relaxation des azotes amides du squelette Les paramètres de relaxation ^{15}N (vitesses de relaxation longitudinale R_1 et transverse R_2, et effet Overhauser hétéronucléaire $\{^{1}H\}$-^{15}N nOe) permettent d'étudier la dynamique des protéines dans une gamme allant de la centaine de ps ($\{^{1}H\}$-^{15}N nOes) à quelques ns (R_1 et R_2). Les valeurs de R_2 sont aussi sensibles aux phénomènes d'échange à l'échelle de la μs-ms. Ces paramètres ont été utilisés pour déterminer le comportement dynamique local des résidus le long de la chaîne de

CtErbB2, ainsi que la dynamique de chaque domaine dans Grb2 entière.

Couplages scalaires Les couplages scalaires (à travers les liaisons) entre protons amides et α ($^3J_{HNH\alpha}$) donnent des informations sur les angles dièdres dans une protéine. Ils ont ici été utilisés pour différencier des structures de type polyproline (PPII) et des structures de type brin β dans CtErbB2.

Couplages dipolaires résiduels (RDCs) Les couplages dipolaires, qui résultent en RMN d'une interaction anisotrope entre spins nucléaires à travers l'espace, donnent des informations d'orientation, notamment des liaisons N-H dans une protéine. Ils peuvent être utilisés pour obtenir des informations de structure secondaire ou d'orientation globale ou relative de domaines structurés. Ils sont moyennés à zéro en solution, mais peuvent être partiellement réintroduits (on parle alors de couplages dipolaires résiduels ou RDCs) par utilisation d'un milieu anisotrope. Ici, les RDCs de CtErbB2 ont été mesurés dans un gel d'acrylamide étiré (Chou et al., 2001), afin de compléter les données de structures secondaires données par les déplacements chimiques et les couplages scalaires.

Augmentation de relaxation par effet paramagnétique (PRE) Afin de compléter les données de structure locale par des informations sur les contacts à longue distance dans CtErbB2, le phénomène d'augmentation de relaxation par effet paramagnétique (PRE) a été exploité : une sonde paramagnétique (ici un nitroxide, le MTSL) est couplée à la protéine en une position choisie (une cystéine). Les signaux des atomes proches de cette sonde dans l'espace subissent une diminution d'intensité. Ainsi, en répétant l'expérience avec différentes cystéines (différents mutants, ici quatre permettant de coupler la sonde en position 45, 146, 227 et 248 respectivement), on peut sonder les contacts tout le long de la chaîne.

Autres méthodes biophysiques

Dichroïsme circulaire (CD) Le dichroïsme circulaire (CD) est basé sur la différence d'absorption de la lumière polarisée circulairement droite ou gauche par une molécule chirale. En particulier, dans l'UV lointain, les hélices α, feuillets β et les régions désordonnées ont des spectres CD très différents. Cela permet

donc l'analyse du contenu en structures secondaires d'une protéine, ici CtErbB2.

Diffusion des rayons X aux petits angles (SAXS) Le SAXS permet d'obtenir des informations à basse résolution en solution, typiquement sur la taille et la forme d'une macromolécule. Les paramètres qui peuvent en être extraits sont le rayon de giration (R_g), une estimation de la masse moléculaire, ainsi qu'une courbe de distribution des distances interatomiques. Ces paramètres, ainsi que la représentation de Kratky des données de SAXS, sont très différents selon le degré de globularité de la protéine étudiée, et sont donc un indicateur de repliement (Kikhney and Svergun, 2015; Receveur-Brechot and Durand, 2012). Les données expérimentales de SAXS peuvent être comparées à des courbes attendues pour un modèle donné (une structure déjà déterminée, un modèle de type polymère, etc). Nous avons utilisé le SAXS pour estimer le degré de structure de CtErbB2, et comparer son comportement à celui d'une chaîne dite "wormlike", flexible avec une longueur de persistance. Le SAXS a aussi été utilisé pour l'analyse de l'organisation des domaines de Grb2, et notamment sa comparaison avec la structure cristallographique. Nous avons enfin utilisé le programme EOM (Tria et al., 2015; Bernadó et al., 2007) pour affiner notre vision de l'organisation des domaines de Grb2 : nous avons généré des ensembles de structures à partir de la structure cristallographique dans laquelle les linkers ont été laissés flexibles, puis sélectionné des sous-ensembles récapitulant les données expérimentales. Nous avons ainsi pu obtenir une distribution des rayons de giration des structures dans ces sous-ensembles.

Résultats

Caractérisation de CtErbB2 isolée non phosphorylée

CtErbB2 présente un haut niveau de désordre CtErbB2 est très riche en prolines (44 prolines sur 268 résidus) et est prédite, de par sa séquence, comme très majoritairement désordonnée. Sa nature désordonnée est confirmée expérimentalement par l'absence de structure secondaire détectée par CD et sa migration ralentie sur gel SDS-PAGE par rapport à la migration attendue de par sa masse moléculaire. De plus, l'allure de sa courbe de Kratky et le bon ajustement de l'intensité de SAXS expérimentale par un modèle de type polymère "worm-like chain" avec une longueur de persistance d'environ 7 résidus confirment la nature hautement désordonnée de CtErbB2. Enfin, le spectre HSQC ^{1}H-^{15}N de CtErbB2 présente une très faible dispersion des signaux dans la dimension proton, confirmant l'absence de repliement. Le caractère désordonné de CtErbB2 a été observé par HDX-MS par Keppel et al. (2017) dans le contexte de la partie

intracellulaire entière de ErbB2, confirmant que ce n'est pas un artefact de notre construction. Il est aussi cohérent avec le rôle de la queue C-terminale dans la transduction du signal par ErbB2, requérant une haute accessibilité des sites de phosphorylation et d'interaction et la possibilité d'interactions multiples et adaptables selon le contexte.

Structures locales transitoires de CtErbB2 Malgré les difficultés liées au grand nombre de prolines dans CtErbB2 et à la faible dispersion des signaux, nous avons pu obtenir l'attribution de près de 99% des résonances du squelette, de 93% des azotes des prolines, et de 75 et 68% des protons et carbones des chaines latérales, respectivement. L'analyse des déplacements chimiques des carbones β et γ des prolines nous a permis de montrer qu'à part deux prolines (148 et 259), les populations d'isomères *cis* de chacune des prolines est inférieure à 10%. De nombreuses prolines étant précédées d'une sérine ou thréonine, une modification de l'état d'isomérisation selon la phosphorylation de ces résidus, ou la reconnaissance de motifs pS-P ou pT-P par des prolyl isomérases, sont envisageables. L'ensemble des déplacements chimiques le long de la séquence nous a aussi permis, grâce à l'algorithme SSP (Marsh et al., 2006) de détecter la présence de structures secondaires transitoires, dont une hélice α proche de l'extrémité N-terminale (résidus14-22) peuplée à environ 25% et des structures secondaires étendues de quelques acides aminés réparties tout le long de la séquence. Ces structures étendues, dont la position est corrélée à la densité de prolines le long de la séquence, ont été confirmées comme étant des hélices PPII (et non des brins β) grâce à l'analyse des couplages scalaires $^3J_{HNH\alpha}$. La conservation d'une haute densité de prolines dans ces régions de CtErbB2, que nous avons constatée chez les mammifères, renforce l'hypothèse de leur importance fonctionnelle. Les valeurs de RDCs mesurées sur CtErbB2 présentent des écarts significatifs par rapport à celles calculées pour un ensemble conformationnel de CtErbB2 qui se comporterait comme une chaîne statistique pure, généré par Flexible Meccano (Ozenne et al., 2012). Nous avons donc généré un ensemble conformationnel de CtErbB2 présentant des populations locales de structures secondaires issues de l'analyse des déplacements chimiques par SSP. Les RDCs calculés pour cet ensemble sont cohérents avec les valeurs expérimentales, confirmant la présence de ces structures secondaires.

Ces dernières sont probablement fonctionnellement importantes, notamment en tant que structures préformées pour des interactions intra- ou inter-moléculaires : l'hélice α correspond à l'hélice interagissant avec le domaine kinase dans les structures cristallographiques de ce dernier, et est possiblement impliquée dans la régulation de l'activité kinase selon l'état de phosphorylation de la protéine ; les hélices PPII sont des

structures connues pour interagir avec les domaines SH3 et WW. L'une d'elle contient notamment un motif de classe I (RxxPxxP, résidus 159-165) d'interaction avec des domaines SH3.

Dynamique locale de CtErbB2 Comme les propensions de structures secondaires, les paramètres de relaxation des azotes du squelette le long de la chaîne de CtErbB2 sont hétérogènes. Les parties les plus dynamiques à l'échelle de la ns sont les extrémités de la protéine, et les régions riches en glycines. Les régions riches en prolines, et partiellement structurées en hélices PPII, apparaissent légèrement moins flexibles, mais proches d'un comportement d'une chaîne statistique, ce qui confirme le caractère dynamique des courtes hélices PPII. Une région riche en histidines (résidus 56-61) montre elle la plus grande rigidité le long de la chaîne (visible sur les trois paramètres de relaxation). La région de l'hélice α présente un comportement dynamique complexe, qui ne peut pas être décrit par un temps de corrélation rotationnel unique, ce qui est cohérent avec une structuration transitoire. Enfin, deux autres régions présentent un comportement similaire mais ne présentent pas de structure secondaire notable : la région des résidus 23-49 et 225-244. L'origine de cette dynamique complexe pourrait être la présence de contacts transitoires à longue distance.

Il existe un contact à longue distance transitoire dans CtErbB2 Pour vérifier cette hypothèse, et déceler la présence de contacts longue distance dans CtErbB2, quatre expériences de PRE ont été effectuées. Elles montrent un effet significatif et réciproque entre les régions autour des résidus 30 et 230. La présence d'un contact transitoire est aussi suggérée par le rayon de giration mesuré par SAXS, qui est inférieur au rayon de giration prédit en tenant compte des structures secondaires observées. Ce contact longue distance dans CtErbB2 pourrait réguler l'accessibilité des sites de phosphorylation pour les domaines kinases du dimère actif, ou pour les protéines adaptatrices des différentes voies de signalisation.

La figure 50 récapitule les éléments structuraux caractérisés lors de notre étude, et leurs rôles potentiels dans les interactions intra- ou inter-moléculaires de ErbB2.

Comportement de Grb2 en solution

Stabilité du domaine NSH3 et influence de l'étiquette histidine Lors de l'attribution des spectres des différentes constructions de Grb2, des différences ont été constatées dans les constructions contenant le do-

maine NSH3 entre les formes avec étiquette histidine et celles où l'étiquette a été clivée à la TEV. Comme suggéré par de précédentes études, le domaine NSH3 seul est en équilibre entre une forme repliée et une forme dépliée. Nous avons observé que la forme dépliée est favorisée par la présence de l'étiquette histidine dans le domaine SH3 seul. Cependant, dans le bidomaine NSH3SH2 et Grb2 entière, la forme dépliée n'est pas visible lorsque l'étiquette histidine est présente, et apparait seulement lorsque l'étiquette est clivée. L'étiquette semble interagir à la fois avec le domaine NSH3 et le domaine SH2 dans ces constructions et stabilise la forme repliée du NSH3. De manière cohérente, les paramètres de relaxation azote de Grb2 entière indiquent que lorsque l'étiquette est présente, les domaines NSH3 et SH2 ont une dynamique très similaire, et distincte de celle du CSH3. Pour la suite de l'étude, la forme où l'étiquette histidine a été clivée a été utilisée.

Organisation des domaines et dynamique de Grb2 entière Le spectre HSQC ^{1}H-^{15}N de Grb2 entière est quasiment parfaitement superposable à la somme des spectres des trois domaines isolés, indiquant l'absence d'interaction notable entre les domaines dans la protéine entière. Ceci est en contradiction avec la structure cristallographique de Grb2, indiquant des contacts intramoléculaires entre les domaines NSH3 et CSH3, et intermoléculaires entre les domaines SH2 et CSH3 (dans le dimère). De plus, le comportement dynamique de Grb2, déterminé par les paramètres de relaxation ^{15}N, est différent de celui de la protéine avec étiquette, et n'indique pas de mouvement corrélé entre les différents domaines. De plus, ces paramètres de relaxation indiquent un monomère majoritaire en solution. Le comportement de Grb2 libre en solution est donc similaire à ce qui a précédemment été observé pour Grb2 en présence de ligands peptidiques d'EGFR et de Sos (Yuzawa et al., 2001).

Afin de confirmer cette interprétation, nous avons utilisé le SAXS, suivant directement une colonne SEC pour s'assurer de la monodispersité des échantillons, notamment en termes d'état oligomérique (Pérez and Vachette, 2017). Le bidomaine NSH3SH2 génère deux pics en SEC, indiquant la présence d'un dimère. L'analyse SAXS du pic correspondant au monomère indique que sa courbe SAXS est compatible avec la structure du bidomaine dans la structure cristallographique de Grb2 entière. À l'inverse, le bidomaine SH2CSH3 est très majoritairement monomérique, et la comparaison avec la courbe de SAXS prédite à partir de la structure cristallographique montre des divergences. En utilisant EOM et en définissant le linker SH2-CSH3 comme flexible, nous avons montré que les données de SAXS ne sont reproduites qu'en considérant plusieurs populations de R_g différents, certains plus petits et d'autres plus grands que celui de la structure cristallographique. La courbe SAXS de Grb2 entière n'est pas non plus compatible avec la struc-

ture cristallographique, et son ajustement nécessite de prendre en compte des populations de même R_g que la structure cristallographique, mais aussi des structures de plus grand R_g.
Nous avons donc défini un nouveau modèle, dans lequel Grb2 est monomérique, et ses trois domaines ne sont pas en interaction. Ce modèle est donné en figure 60. Une flexibilité plus grande du linker entre le domaine SH2 et le domaine CSH3 par rapport au linker entre les domaines NSH3 et SH2 est observée. Ce comportement asymétrique, tout en conservant une relative indépendance structurale et dynamique des trois domaines de Grb2, pourrait avoir des conséquences dans les interactions et la transduction du signal.

Interaction entre CtErbB2 et Grb2

Dans le modèle actuel de l'interaction ErbB2-Grb2, le domaine SH2 de Grb2 interagit avec la tyrosine Y_B de ErbB2 phosphorylée. Cette interaction n'a jamais été étudiée à haute résolution. De plus, nous avons montré que CtErbB2 contient des hélices PPII transitoirement formées et des motifs PxxP, avec lesquels les domaines SH3 de Grb2 pourraient interagir. Nous avons donc effectué l'étude de l'interaction entre Grb2 entière et CtErbB2 par RMN.

Interaction entre Grb2 et CtErbB2 non phosphorylée Nous avons d'abord étudié l'interaction entre Grb2 et CtErbB2 non phosphorylée. L'interaction est faible, mais les perturbations de déplacement chimique et d'intensité des signaux lors des expériences de titrage ont permis de mettre en évidence des zones d'interaction spécifiques. La plus affectée est celle contenant la tyrosine Y_B (résidu 152), ainsi qu'un segment riche en prolines ^{159}RPQPPSPR^{165}. Ce segment contient deux motifs PxxP susceptibles d'interagir avec des domaines SH3, et notamment un motif de classe I (RxxPxxP). Du côté de Grb2, le domaine NSH3 est le plus perturbé par l'interaction, au niveau du site connu de liaison à des peptides polyprolines, et quelques résidus du domaine SH2 sont impliqués. Ces résidus correspondent à la poche du SH2 liant non pas les phosphotyrosines, mais l'asparagine en position +2 conférant la spécificité pour Grb2. Cette asparagine est bien présente dans le motif contenant Y_B. Le domaine CSH3 n'est que peu perturbé par l'interaction.

Interaction entre Grb2 et des peptides de CtErbB2 non phosphorylés ou phosphorylés Pour aller plus loin dans la caractérisation de l'interaction, et en parallèle de l'optimisation de la phosphorylation des tyrosines dans CtErbB2 entière, nous nous sommes concentrées sur l'interaction de Grb2 avec des peptides

synthétiques de CtErbB2 contenant Y_B et le segment riche en prolines (résidus 141 à 176). L'interaction de Grb2 avec la forme non phosphorylée de ce peptide est similaire à celle observée avec CtErbB2 entière non phosphorylée. Elle implique majoritairement le site de liaison du domaine NSH3, avec une affinité faible (une constante de dissociation K_D de l'ordre de la centaine de micromolaire au millimolaire). L'échange entre la forme libre et liée de ce domaine est rapide à l'échelle des déplacements chimiques. Lorsque la tyrosine B du peptide de CtErbB2 est phosphorylée, la perturbation du domaine NSH3 de Grb2 est inchangée (en termes de vitesse d'échange et d'affinité). Cependant, la phosphorylation de la tyrosine permet le recrutement du domaine SH2 pour une interaction de plus grande affinité. L'échange de ce domaine entre la forme libre et liée est lent à l'échelle des déplacements chimiques, et avec un K_D plus faible que le NSH3, de l'ordre du micromolaire. Les résidus du SH2 impliqués dans l'interaction sont ceux du site de fixation de la phosphotyrosine et du site de reconnaissance de l'asparagine.

Conséquences fonctionnelles possibles L'implication de deux domaines de Grb2 dans l'interaction avec CtErbB2 change notre vision de la transduction du signal par cette protéine adaptatrice, comme présenté en figure 66. La spécificité du motif riche en proline de CtErbB2 pour le domaine NSH3 de Grb2 suggère une possible compétition pour la liaison de Grb2 avec Sos. L'interaction simultanée SH2-pY et NSH3-PxxP nécessiterait un rapprochement des domaines NSH3 et SH2 dans la forme liée en comparaison avec la forme libre, et l'organisation des domaines de Grb2 dans le complexe serait donc intéressante à étudier.
Du côté de ErbB2, mieux comprendre le mécanisme de liaison de CtErbB2 à Grb2 nécessite d'obtenir et d'étudier CtErbB2 phosphorylée. C'est un travail en cours, dont les enjeux, les défis et les méthodes ont été étudiés pendant ce travail de thèse.

Vers la caractérisation d'états fonctionnels de CtErbB2 : le défi des études structurales et fonctionnelles de la phosphorylation de tyrosines

La phosphorylation des tyrosines comme mécanisme de régulation dans la cellule a été découverte plus de vingt ans après celle des sérines et thréonines (Burnett and Kennedy, 1954; Eckhart et al., 1979). Les phosphotyrosines (pY) sont très largement moins abondantes que les phosphosérines (pS) et phosphothréonines (pT) à l'échelle d'un organisme (Hunter and Sefton, 1980). Les mécanismes moléculaires de phosphorylation des tyrosines sont souvent complexes, la majorité des tyrosine kinases étant des RTKs (Manning, 2002) requérant des phénomènes d'autophosphorylation. Leur étude est donc relativement récente et peu docu-

mentée.

Le défi de la détection des sites de phosphorylation La détection par des anticorps est longue, et ne permet pas un suivi en temps réel. La spectrométrie de masse est très employée, mais pose plusieurs problèmes : la faible sensibilité de détection des espèces phosphorylée (Thingholm et al., 2009), et la difficulté de résolution du site de phosphorylation, notamment lorsque la protéine est riche en prolines empêchant le clivage protéolytique complet, ou lorsque des séquences sont répétées, comme c'est le cas pour CtErbB2. La RMN est une technique adaptée, et déjà utilisée dans plusieurs études de phosphorylation de sérines et/ou théronines (Selenko et al., 2008; Liokatis et al., 2010; Theillet et al., 2012; Cordier et al., 2012, 2015; Landrieu et al., 2006; Sólyom et al., 2015). Elle permet un suivi cinétique et une résolution précise des sites de phosphorylation. Cependant, pour l'étude des tyrosines phosphorylées (particulièrement dans les protéines désordonnées), des problèmes de superposition des signaux se posent plus fréquemment que pour les sérines et thréonines. Plusieurs méthodes nécessitent donc d'être utilisées en parallèle.

Produire des échantillons de protéine phosphorylée pour des études structurales La production d'échantillons de protéines phosphorylées est aussi un défi majeur. Plusieurs techniques d'incorporation d'acides aminés phosphorylés à des positions choisies existent (phosphomimétiques, synthèse chimique, acides aminés non-naturels). Cependant, elles ne fournissent pas le(s) motif(s) de phosphorylation biologiquement pertinents. D'autres méthodes permettent de s'en approcher, en phosphorylant la protéine d'intérêt (*in vivo* ou *in vitro*) directement par une kinase (idéalement, la kinase endogène). Dans notre cas, le système le plus pertinent serait la reconstitution de dimères d'ErbBs, dans lesquels la phosphorylation a lieu *in vivo*.

Stratégies mises en place dans le cadre de ces travaux Ici nous avons choisi de développer trois approches, répondant à différentes questions sur la phosphorylation d'ErbB2 :

- La coexpression de CtErbB2 avec une tyrosine kinase dans *E. coli*, en utilisant une souche commerciale (TKB1, Agilent). Cette approche permet de produire, avec un protocole *a priori* relativement classique et permettant un marquage isotopique, des quantités de protéine phosphorylée suffisantes pour des études RMN complètes de l'état phosphorylé de CtErbB2. Cependant, le motif de phos-

phorylation engendré n'est pas contrôlé ni connu car la kinase coexprimée n'est pas indiquée. Le peu d'informations données par le fournisseur sur la souche rend aussi l'optimisation du protocole difficile. Pour l'instant, nous avons réussi à obtenir un échantillon phosphorylé non isotopiquement marqué et l'optimisation de la production en milieu marqué est en cours.

- La phosphorylation *in vitro* de CtErbB2 par la kinase ErbB2 (Life technologies). La phosphorylation peut ainsi être suivie en temps réel dans le tube RMN, avec une spécificité de séquence de la kinase respectée, mais sans la spécificité due à la topologie des dimères dans lesquels ont lieu la phosphorylation *in vivo*. Des problèmes ont été rencontrés, dus à la faible efficacité catalytique de la kinase, et sa contamination constatée par une sérine/thréonine kinase.

- La production en cellules d'insectes de la partie intracellulaire entière du récepteur ErbB2. Cette approche permettrait à la fois d'étudier le comportement de la queue C-terminale attachée au domaine kinase, et son auto-phosphorylation dans un contexte relativement proche du contexte biologique. En utilisant le système baculovirus Bac-to-Bac®, nous avons réussi à exprimer la portion intracellulaire d'ErbB2. La production à grande échelle pour la RMN et le marquage isotopique restent à réaliser.

Ces trois approches ont pour inconvénient de produire des protéines avec des motifs de phosphorylation *a priori* complexes et hétérogènes. Une approche serait de phosphoryler des sites uniques pour commencer par l'étude de l'effet des phosphorylations individuelles. Dans ce but, des mutants monotyrosines, dans lesquels toutes les tyrosines ont été mutées en phénylalanine sauf une, ont été produits. La phénylalanine a été choisie pour sa ressemblance structurale avec la tyrosine, combinée à l'impossibilité de sa phosphorylation due à l'absence du groupe hydroxyle. Nous avons effectué l'étude RMN du mutant où toutes les tyrosines ont été remplacées par des phénylalanines, et montré que ces mutations ne perturbent pas significativement le comportement de CtErbB2, validant cette approche. Elle nous permettra d'étudier notamment trois phosphorylations intéressantes, en s'appuyant sur l'étude de Dankort et al. (1997) : la phosphorylation de Y_A, qui a été montrée comme inhibant la signalisation, la phosphorylation de Y_B, cruciale pour l'interaction avec Grb2, et celle de Y_D, qui peut lever l'inhibition due à Y_A . La phosphorylation de Y_A et Y_D pourrait aussi modifier le contact longue distance entre ces deux régions que nous avons mis en évidence grâce aux expériences de PRE.

Conclusions et perspectives

Nous avons caractérisé en termes structuraux et dynamiques l'ensemble conformationnel de CtErbB2, et l'organisation des domaines de son partenaire Grb2. Nous avons aussi montré que l'interaction entre ces deux protéines met en jeu deux interactions locales : celle du domaine SH2 de Grb2 avec la tyrosine B phosphorylée de CtErbB2 et celle entre le domaine NSH3 de Grb2 et un motif polyproline de CtErbB2. La caractérisation structurale et dynamique complète des deux partenaires en complexe reste à effectuer, afin de déterminer notamment si et comment CtErbB2 se structure en interaction, et si l'organisation et la dynamique inter-domaine de Grb2 est modifiée. De plus, l'étude de l'effet du domaine kinase et de la phosphorylation sur CtErbB2 et ses interactions est fondamentale dans le but de comprendre ses mécanismes fonctionnels. Cette étude, avec notamment la production du domaine intracellulaire entier de ErbB2 pour la RMN, est en cours. Ces travaux sont un élément de plus vers la compréhension du comportement des protéines et régions intrinsèquement désordonnées et de leurs rôles.

INTRODUCTION

Intrinsically disordered proteins (IDPs) or regions (IDRs), that lack stable three-dimensional structure, have attracted much attention in the last twenty years. Previously often considered as artefacts or thought to have no physiological function, these proteins are now known to be widely spread in all organisms and especially eukaryotes (Ward et al., 2004), and to be involved in crucial processes in the cell. In particular, a high proportion of disorder is observed in signaling and regulatory pathways (Iakoucheva et al., 2002). Additionally, many post-translational modifications, and especially phosphorylation, occur preferably in disordered regions (Iakoucheva et al., 2004). Disordered regions are often found in receptors (Minezaki et al., 2007; Kjaergaard and Kragelund, 2017), and serve as linkers between folded domains and/or the membrane, regulatory domains, localization signals, hubs for signaling protein-protein interaction or post-translational modification platforms, or a combination of these roles (Kjaergaard and Kragelund, 2017). The C-terminal tail of the ErbB2, further noted CtErbB2 in the manuscript, is a good example of the role of such regions, and a typical illustration of how disordered regions have often been overlooked in mechanistic studies. With recognition of IDPs prevalence, the emergence of new associated concepts, and the development of customized experimental methods, especially in nuclear magnetic resonance (NMR), neglecting them is no longer conceivable.

ErbB2, also called HER2 (Human epidermal growth factor receptor 2) or neu, is part of the ErbB family of receptor tyrosine kinases that contains four members. The first member of this family to be discovered, the epidermal growth factor receptor (EGFR or ErbB1), is the most studied one and represents a model to understand how the ErbBs transduct signal from the extracellular space to signaling pathways. ErbB-dependent pathways are critical for cell division, motility and differentiation, and control of apoptosis. They include the Ras/MAPK pathway, the MEMO-RhoA-mDia1 pathway, the Akt/PI3K pathway, the STAT pathway, and the Src kinase pathway (Yarden and Sliwkowski, 2001; Marone et al., 2004). The studies carried out on EGFR led to the following model of the mode of action of ErbBs (Kovacs et al., 2015b): ligand binding triggers a conformational change in the extracellular domain, enabling homo- or hetero- dimerization of the receptors, and dimerization-induced activation of their intracellular tyrosine kinase domain (TKD). This domain

is then able to phosphorylate the C-terminal tail of the receptors, assumed to be a *trans*-phosphorylation. The tyrosine-phosphorylation of the C-terminal tails triggers interaction with adaptor proteins, activating the corresponding signaling pathways. This well-established model has two issues: it does not take into account the specificities of each ErbB, and lacks a description of the C-terminal tail and its mechanisms of interaction with the kinase domain(s) for phosphorylation and with the adaptor proteins for signal transduction.

In the case of ErbB2, no biologically relevant ligand has been identified for the extracellular domain, that has been found only in a dimerization-potent conformation (Cho et al., 2003). Physiologically, ErbB2 has been shown to be the preferred hetero-dimerization partner of other ErbBs (Graus-Porta et al., 1997). Pathologically, overexpression of ErbB2 is observed in more than 20% of breast cancers, and this HER2 + form of the disease is very aggressive (Slamon et al., 1987). The mechanism of kinase activation is believed to be, as for EGFR, only dimerization-dependent (Aertgeerts et al., 2011). Thus, no regulation mechanism other than receptor concentration seems to exist upstream of phosphorylation itself. Regulation of the phosphorylation and interactions of the C-terminal tail, which is the end of the chain before transducting signal to partners, is key. The absence of crystal structure for this region has led to the assumption of its lack of a stable structure, which is consistent with measurement of dynamics of this region, and observation of delayed electrophoretic migration. Very recently, its has finally been proven that the tail of EGFR and ErbB3 receptors is indeed disordered (Keppel et al., 2017). However, no residue-resolved information is accessible yet. At the beginning of this work, CtErbB2 was known to have several functions usually associated with structural disorder: it is a highly-phosphorylated region, and accessibility of these sites is crucial; it is an interaction hub for signaling, which requires plasticity and adaptability; it is believed to participate in regulation of the kinase domain, once again requiring adaptability. CtErbB2 is thus both an interesting model system to study the emerging "disorder-function" paradigm, and a potential target in the context of breast cancer.

To go further into this functional investigation, we chose to study the interaction between CtErbB2 and the adaptor protein Grb2. Grb2 is the link between ErbB receptors and the Ras/MAPK pathway that had been missing until the study by Lowenstein et al. (1992) and is crucial for both ErbB2 physiological and pathological roles. Grb2 is made of a central SH2 domain, and two terminal SH3 domains (named NSH3 at the N-terminus and CSH3 at the C-terminus). It has been shown to bind specifically to one tyrosine of CtErbB2 (numbered Y1139 in full-length ErbB2, Y152 in the tail, and often designated as Y_B for being the second auto-phosphorylation site in the tail) (Dankort et al., 1997). The sequence around this tyrosine matches the determined sequence specificity of the SH2 domain of Grb2 (Marengere et al., 1994), but no atomic model

of the binding exists. Besides, Grb2 SH3 domains could be involved in the CtErbB2-Grb2 interaction, since SH3 domains are known to interact with polyproline motifs, and that CtErbB2 is proline-rich. Grb2 is thus an interesting partner to study CtErbB2 binding modes.
Finally, the phosphorylation itself of CtErbB2 tyrosines is interesting to understand phosphorylation of IDPs. This system has several specificities that make the study of its phosphorylation both unique and challenging: it happens on tyrosines, and tyrosine phosphorylation remains poorly characterized compared to serines and threonines; there are at least 5 phosphorylation sites (Hazan et al., 1990; Segatto et al., 1990), distributed heterogeneously along CtErbB2 chain, leading to expected heterogeneous phosphorylation patterns; and the endogenous phosphorylation mechanism is not trivial, since it is autophosphorylation: phosphorylation happens in the context of ErbB dimers, with the ErbB kinases phosphorylating the tails, in a mechanism that has never really been elucidated and happens in *trans* (Qian et al., 1994), but potentially also in *cis*.

In this PhD work, I studied CtErbB2 both from the structural and functional points of view.
In the first chapter, I describe the context of this study. Section 1 is dedicated to understanding the specific questions arising from the study of intrinsically disordered proteins, from their sequence to their structural features and functional specificities, and ways to study them. In section 2, I present interest of studying the ErbB family of proteins, and why understanding the mechanism of signal transduction in their C-terminal tail, in particular in ErbB2, is necessary. Finally, section 3 is dedicated to CtErbB2 partner Grb2.
The second chapter presents the experimental approaches used to study this system: from production and purification of the studied proteins (section 4), to the NMR experiments (section 5) and complementary characterization techniques (section 6) used.
The third chapter presents the results of this study. In section 7, the unphosphorylated form of CtErbB2 is described in terms of structure and dynamics, at atomic resolution, and I show how the results can yield information relevant for ErbB2 function. In section 8, I present the study of Grb2 alone, and how the results allow a better understand of how it can perform its function as a modular protein. Section 9 is devoted to the analysis of the interaction between Grb2 and CtErbB2, showing SH2-tyrosine and SH3-polyproline components. Finally, section 10 presents methods that I implemented to produce and study phosphorylated samples of CtErbB2, and preliminary results.

CONTEXT OF THIS WORK

1 Intrinsically disordered proteins (IDPs)

1.1 Discovery and assessment of biological relevance of IDPs

Intrinsically disordered proteins (IDPs) are proteins that lack stable three-dimensional structure under physiological conditions. The idea that these proteins are functional and widely spread in life was slow to emerge, for it deeply challenged the structure-function paradigm.

1.1.1 Historical perspective

Since Emil Fisher and the concept of "lock and key" established in 1894 (Fischer, 1894), the idea that specificity of interactions (and therefore function) emerged from complementarity of well-defined, unique conformation of the partners had governed structural biology. It was reinforced while the number of elucidated structures grew from the one of myoglobin, solved by Kendrew et al. (1958), to nearly 11,000 structures at the end of the XX^{th} century in the Protein Data Bank (PDB). In the 50s, the importance of flexibility was acknowledged in a few studies, such as the study of "configurational adaptability" of bovine serum albumin by Karush (1950), and the one by Koshland and Brooks (1958) introducing the concept of "induced fit", but the importance of interdomain motion was not deeply investigated before the 60s and 70s (Monod et al., 1965; Koshland et al., 1966). The concept of intrinsic disorder, and the recognition of its functional importance, was only possible with the combination of experimental evidence, such as the observation of missing electron density in a large amount of crystal structures, or high sensitivity to proteases, and bioinformatics predicting high levels of disorder in protein sequences derived from high-throughput sequencing of genomes. This global shift in paradigm was really initiated at the end of the 90s, with the study by Romero et al. (1998) predicting that over 15,000 proteins of the SwissProt database have disordered segments of more than 40 residues, and the insightful review by Wright and Dyson (1999) asserting functionality of disorder: "The high proportion of these sequences in the genomes of all organisms studied to date argues for

important, as yet unknown functions, since there could be no other reason for their persistence throughout evolution.". This argument can be nuanced by the fact that statistically, protein sequences should be biased towards disorder rather than stable structure, and that in that sense we should rather wonder why structured proteins have been selected by evolution. But given the paradigm at that time, this was a bold statement. Since then, the importance of disorder in all kingdoms of life has been further assessed: in 2004, 2.0% of archaean, 4.2% of eubacterial and 33.0% of eukaryotic proteins were predicted to have disordered segments of more than 30 residues (Ward et al., 2004), and as of 2013, a proportion as high as 44% is predicted for human proteins (Oates et al., 2013).

1.1.2 Current questions and challenges

One of the remaining questions concerning the relevance of IDPs was to know to what extent structural disorder could persist in the crowded intracellular environment. The effect of molecular crowding on folding or unfolding is complex, difficult to predict, and crowding can have multiple outcomes (Fonin et al., 2018). Although the number of proteins which behavior has been studied *in cell* is still limited, molecular simulations, in-cell NMR, fluorescence studies and Fourier transform infrared spectroscopy have begun to unveil details of the life of IDPs in such environment, and showed that prokaryotic and eukaryotic cells can preserve high levels of disorder (Theillet et al., 2014, 2016).
The study of intrinsic disorder is still a developing field, with new systems investigated regularly and continuous methodological development. Latest reviews mainly focus on methodological aspects: NMR studies, with focus on in-cell studies (Milles et al., 2018; Sciolino et al., 2019), single-molecule fluorescence (Leblanc et al., 2018; Nasir et al., 2019), Small-angle X-ray scattering (Cordeiro et al., 2017) and simulations (Huang and MacKerell, 2018; Bhattacharya and Lin, 2019). There is also emerging interest into applying recently acquired knowledge on IDPs to drug development (Metallo, 2010; Uversky, 2010; Martinez et al., 2017; Kumar et al., 2017; Ruan et al., 2019).

With this paradigm shift, many systems can be looked at with new eyes. While disordered proteins that were ignored for a long time are now being deeply studied as model systems, such as α-synuclein or tau, interest is also growing for disordered parts of structured proteins that have been long disregarded, and for which atomic-scale description was not considered relevant for mechanistic studies. That is the case for the terminal tails of receptors such as ErbBs.

1.2 Diversity of disordered states

Proteins are not either ordered or disordered. A way of describing the diversity of their states is to classify functional proteins into four major thermodynamic states (Uversky, 2002), any of which possibly being native. In this Protein Quartet model shown in figure 1a, extended from the Protein Trinity model of Dunker et al. (2001), "function aris[es] from four specific conformations (ordered forms, molten globules, pre-molten globules, and random coils) and transitions between any two of the states". While random coils possess very few to no secondary or tertiary structures, pre-molten globules are a little more compact and contain residual secondary structures. Secondary structure content and compactness increases even further in molten globules, and are maximum in the ordered state.

The Protein Quartet is still a simplified view. While proteins that lack both secondary and tertiary structure all along their sequence represent one extreme case, globular proteins another, (pre-)molten globules are intermediate states on a spectrum that can be defined as the "continuum of disorder", with different lengths of intrinsically disordered regions (IDRs) and different levels of flexibility/residual secondary or tertiary structure in these regions (Uversky et al., 2005; Dyson and Wright, 2005). This continuum and different states along it are presented in figure 1b.

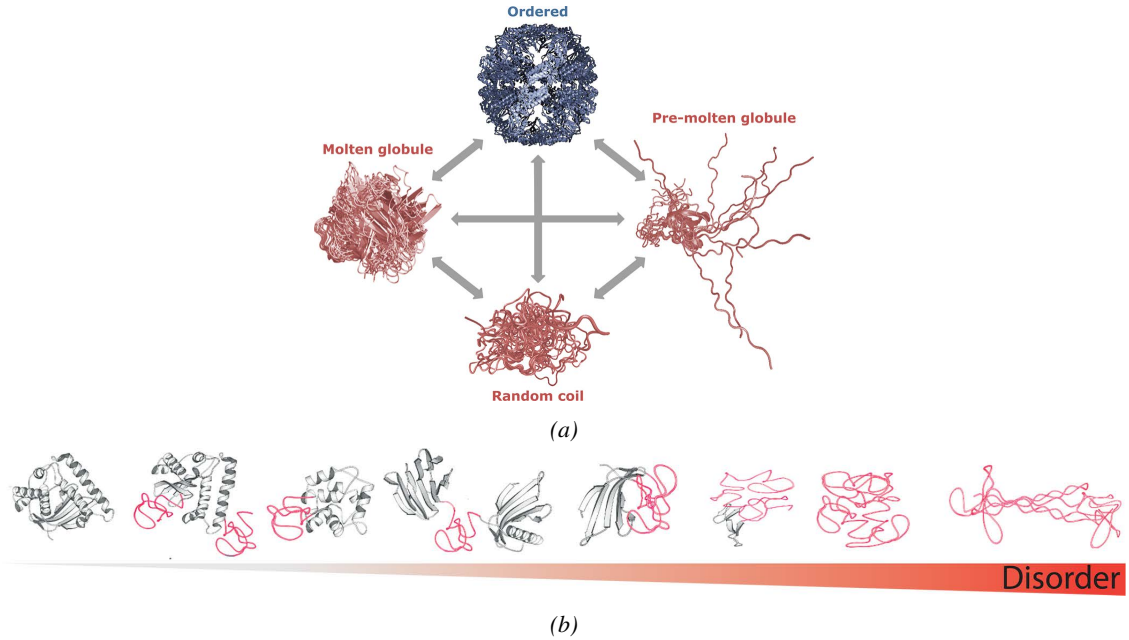
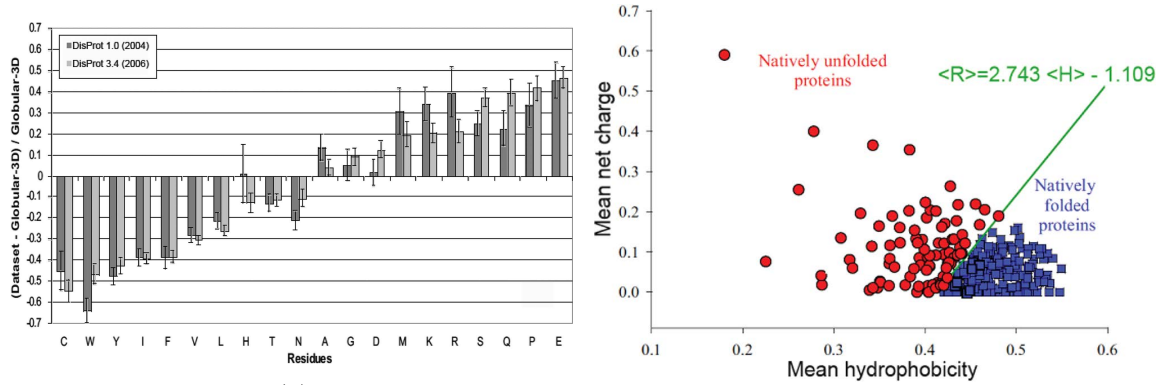


Figure 1 – Depiction of the diversity of disordered states. (a) Model of the Protein Quartet, described by Uversky (2002). Figure from Van Der Lee et al. (2014). (b) Continuum of disorder, from globular proteins to IDPs. From Habchi et al. (2014)

1.3 The sequence-(dis)order relationship

1.3.1 Origin of disorder: composition bias in IDPs

Anfinsen's dogma (Anfinsen, 1973), stating that the structure of a protein is determined by its amino acid sequence, can be easily extended to disordered proteins. Folding being mainly driven by the formation of a hydrophobic core, depletion in hydrophobic and aromatic residues is often the main characteristic of IDPs. They are instead enriched in polar and charged residues, as well as in secondary-structure disrupting residues (proline, and to a less extent glycine and alanine). The composition bias of IDPs is shown in figure 2, together with the resulting distinct area sampled by structured and disordered proteins in a Charge-Hydropathy plot.



(a) *(b)*

Figure 2 – Amino-acid composition features of intrinsically disordered proteins. From Habchi et al. (2014) (a) Bias of amino acid frequence in IDPs compared to globular proteins. (b) Mean net charge plotted against hydropathy for IDPs (red) and folded proteins (blue).

Consequently, IDPs are characterized by the absence of collapse (random coils), or by a non-specific, dynamic collapse (disordered globules).

Although repetitive sequences are not restricted to disordered proteins (for example collagen has Gly-X-X, and especially Gly-Pro-Hyp repeats, where Hyp is hydroxyproline), it is a characteristic often found in IDPs, which in addition to their overall composition bias, tend to have low sequence diversity.

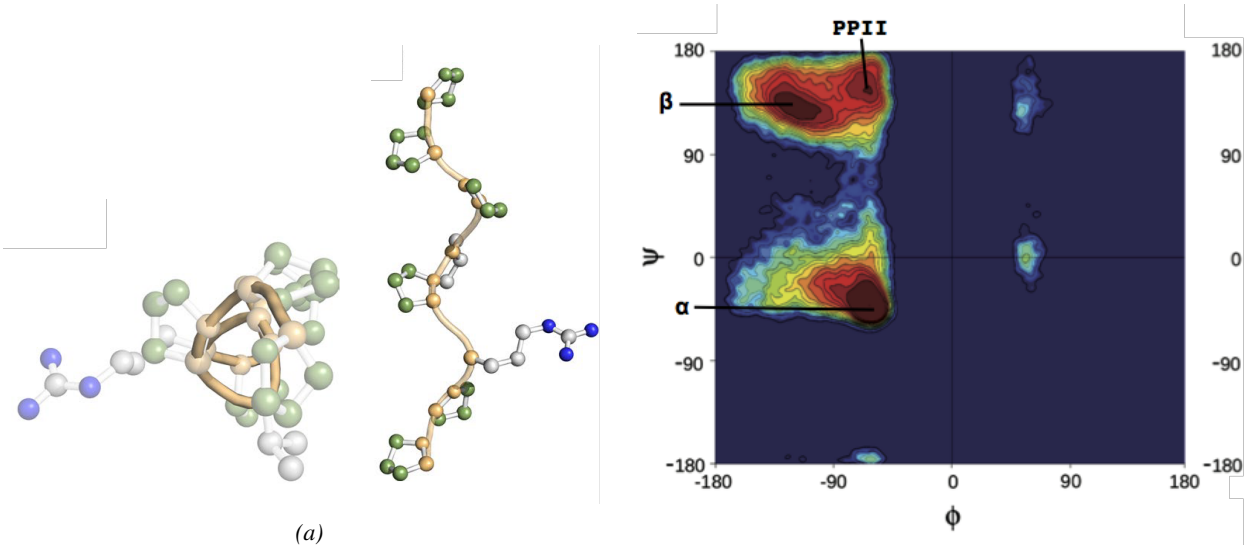
Let us have a closer look at the role of prolines, that have a particular role in protein order, disorder, and dynamics, and that are enriched in CtErbB2.

1.3.2 The unique role of prolines in disorder

This natural amino acid is unique, for its cyclic side-chain, covalently bound to the backbone nitrogen, restricts its dihedral ϕ angle, and because it has no amide proton to act as a H-bond donor. These features often exclude prolines from the core of canonical secondary structures, namely α helices and β sheets, and explains the high abundance of prolines in disordered proteins as seen in figure 2a.

Another interesting feature of prolines is the relatively high proportion of *cis* isomer of the X-Pro peptide bond: 5 to 80% of the *cis* isomer in short model peptides and up to 10% in unfolded polypeptides or proteins, exchanging with the *trans* isomer; in most folded proteins either the *cis* or *trans* isomer is populated due to constraints (Alderson et al., 2018). The exchange between the *cis* and *trans* isomers in short peptides is slow (hundreds to thousands of s) (Reimer et al., 1998). However, *in vivo* the process can be catalyzed by prolyl isomerases, leading to an acceleration by several orders of magnitude. An interesting thing to note is that the important Pin1 isomerase is specific to pSP and pTP motifs (Ranganathan et al., 1997), and that many Ser/Thr kinases recognize SP or TP motifs, showing an important interplay between prolyl isomerisation and phosphorylation (Lu et al., 2002).

Despite their enrichment in IDPs and IDRs, prolines have the ability to form specific kinds of structures, namely polyproline helices, PPI and PPII. While the right-handed PPI helices are only populated in some organic solvents (often aliphatic alcohols) and require all-*cis* peptide bonds, all *trans* PPII helices are quite abundant in aqueous solvents. PPII helices are left-handed, extended triangular structures of 3 residues per turn and 3.1 Å rise per residue (Mansiaux et al., 2011), as shown in figure 3a. About 5% of protein residues are estimated to adopt PPII conformation (not only prolines), but this number might be underestimated (Mansiaux et al., 2011). Indeed, PPII helices are not always taken into account in structural analyses, and their extended nature and dihedral angles (ϕ, ψ) of (-78°, -149°) can lead to confusion with β strands (see figure 3b). They are abundant in disordered proteins, especially (although not exclusively) in proline-rich ones (Mansiaux et al., 2011). No intra-helix hydrogen bonding is observed in PPII helices, consistent with their enrichment in prolines. This makes PPII helices good H-acceptors from the solvent or partners, and frequent interaction motifs, with SH3 or WW domains, or with DNA.



(a) (b)

Figure 3 – Polyproline II (PPII) helices. (a) PPII structure, from the top (left) or side (right). (b) Ramachandran plot showing the region defining PPII helices, close from β sheets. From Narwani et al. (2017).

1.3.3 Predicting disorder

These sequence specificities imply that disorder is, in most cases, rather easy to predict directly from sequence (with the Charge-Hydropathy plot introduced by Uversky et al. (2000) and the Hydrophobic Cluster Analysis developed by Gaboriaud et al. (1987) for example). However, most methods (PONDR®, found at `www.pondr.com/`, DisEMBL (Linding et al., 2003), found at `http://dis.embl.de/`, DISOPRED (Ward et al., 2004), found at `http://bioinf.cs.ucl.ac.uk/disopred`) are now indirect, and use machine learning, relying on experimental IDPs databases (the largest one being DisProt `http://www.disprot.org`, with 803 proteins and 2167 protein regions as of April 8th 2019). IUPRED (Dosztányi et al., 2005; Mészáros et al., 2018) (`https://iupred2a.elte.hu/`) is an combined method, using energy calculations based on contacts in the sequence predicted from databases of contacts in globular proteins. Each predictor performs differently depending on the characteristics of the considered protein (length, sequence composition...). The use of these predictors can help to design constructs of proteins, and to orient the choice towards the structural biology technique(s) that seem most powerful to study them (see section 1.5).

1.4 Functional specificities

IDPs perform a wide range of functions. Their adaptability often allows a single IDP to have multiple, independent functions; a phenomenon called *moonlighting*. A redefinition of the structure-function paradigm

has been proposed over time (Dunker et al., 2001; Uversky, 2017a) to include the disorder-function relationship and the role of exchange between order and disorder, and is given in figure 4. IDPs are enriched in some functional mechanisms: molecular assembly, molecular recognition, protein modification and entropic chain functions (Uversky, 2017a). Except when their role relies only on dynamics (entropic chains, flexible linkers), the functional specificities of IDPs mainly rely on the way they interact with partners, that is fundamentally different from the canonical lock and key model, and is closer to an intermolecular folding process.

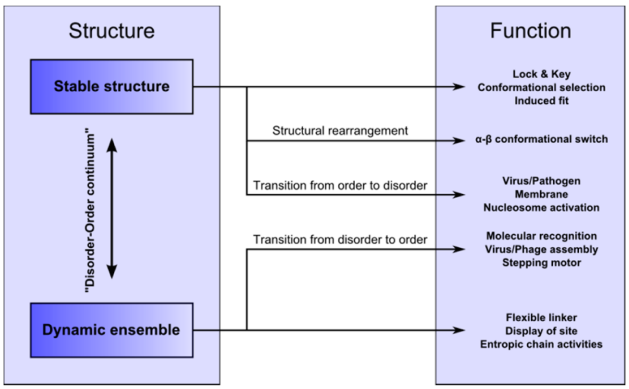


Figure 4 – The structure/disorder-function paradigm, developed by Dunker et al. (2001). Function can arise from order, disorder, or exchange processes between those states. Figure from Brucale et al. (2014).

1.4.1 Interaction modes of IDPs

Contrary to structured proteins, where the interaction site is usually defined by a surface, interaction of IDPs is mediated by short linear motifs (SLiMs) of around 3 to 10 residues, and/or by longer segments that adopt a secondary structure when bound and are called molecular recognition features (MoRFs) (Oldfield et al., 2005; Van Der Lee et al., 2014). SLiMs can be recognized for cleavage, post-translational modification, or as ligands to other, usually structured, proteins. They can be, and often are, embedded in MoRFs.
Hereafter is presented mainly the mechanism of binding of IDPs to structured partners. However, IDP/IDP interactions exist, for example mediated by strong electrostatic interactions (Borgia et al., 2017), or weak multiple binding that can lead to liquid-liquid phase transition (Uversky, 2017b).

Coupled folding and binding of IDPs Several types of complexes can be formed by the interaction of IDPs with their target: the best described one so far is when the interaction site of the IDP, then classified as a MoRF, folds upon binding (Wright and Dyson, 2009). MoRFs can form any type of secondary structure

(α-helices, β-sheets, polyproline helices, or even turns). The most commonly found one is the α-MoRF (Oldfield et al., 2005), since formation of α-helices only requires local interactions, as opposed to β-strands. MoRFs can either be recognized via conformational selection amongst different conformations preexisting in the free IDP, or "fold upon binding" (induced fit mechanism), as illustrated in figure 5. Examples of both processes have been found: while the binding of the IDP PUMA to Mcl-1 is independent on residual structure in the free state (Rogers et al., 2014), indicating an induced fit, the unbound state of PDE_γ exhibits transiently formed structure resembling its bound state to transducin (Song et al., 2008). However, for many IDPs, the actual mechanism is a complex mix between these extreme cases (Dogan et al., 2014).

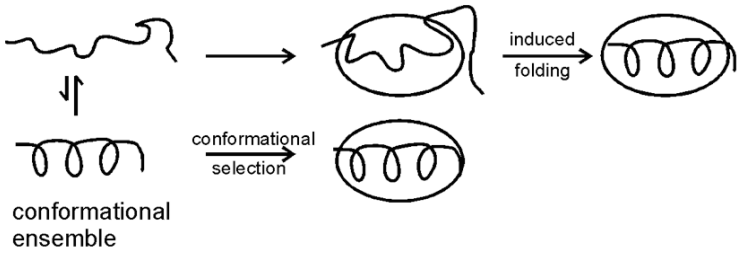


Figure 5 – Schematic representation of the two extreme coupled folding and binding mechanisms. From Wright and Dyson (2009).

Multiplicity of binding and flexibility in the complex IDP binding offers an additional level of complexity, in a much more common way than globular proteins: multiplicity of binding. Thanks to the relative independence of multiple interaction sites allowed by flexibility, and to the transient nature of most IDP complexes, IDPs can have a great number of different partners, either simultaneously on different sites, or successively on one particular site. For example, the CREB-binding protein CBP, can bind different proteins (p160 or IRF3) on the same site, adopting different structures (Waters et al., 2006; Qin et al., 2005). Conversely, a single structured partner can interact simultaneously with different sites of the IDP, creating a phenomenon of avidity, and different sites of an IDP can compete for a same interaction site of the partner, a phenomenon called allovalency that increases overall affinity (Olsen et al., 2017). In the extreme case, multiple sites on both the IDP and its partner, where the pairwise interactions are not highly specific, lead to a highly dynamic complex where the IDP remains widely disordered: this is called a "fuzzy complex" (Tompa and Fuxreiter, 2008). A database of fuzzy complexes can now be found at `http://protdyn-database.org/` (Miskei et al., 2017). The different possibilities of binding multiplicity between an IDP and a structured partner are schematically represented in figure 6.

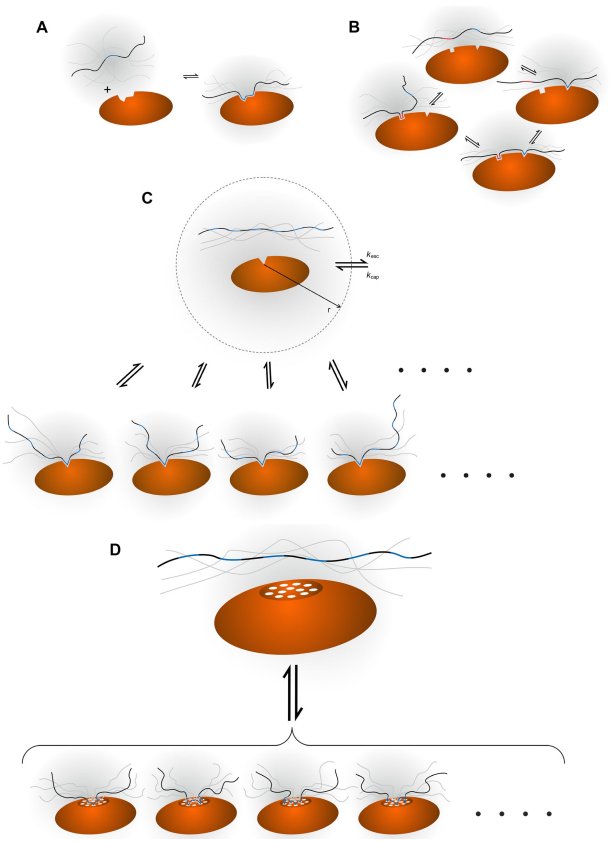


Figure 6 – Different binding mechanisms of one IDP to one globular partner. (a) Two-state model, involving one site on each partner. (b) Avidity, with two specific sites on each partner. (c) Allovalency, with multiple sites of the IDP competing for a single site of the globular partner. (d) Fuzzy complex, with multiple sites on both partners, without strong pairwise specificity. From Olsen et al. (2017).

Extension of the concept of allostery Allostery is defined as the regulation of activity at one site of a protein by binding of a ligand at another, remote, site. The concept of allostery was originally described by Monod, Jacob and Changeux in the context of a frozen structural description of proteins (Monod and Jacob, 1961; Changeux, 1961; Monod et al., 1965). It was thus restricted to structural changes, and the dynamics underlying the transition between states was not the focus. Recent works on redefinition of the concept of allostery, taking into account disorder and dynamics, have led to a more comprehensive view. Allostery can be defined as ligand-binding triggering changes in the conformational ensemble of the protein, including not only structural but also dynamic changes (Motlagh et al., 2014; Tompa, 2014; Berlow et al., 2018). New forms of allostery that are regrouped under the concepts of "ensemble allostery" (Motlagh et al., 2014) or "multistery" (Tompa, 2014) are presented in figure 7 and can include transitions between relatively rigid

states, modification of side-chain and/or backbone dynamics of folded domains, local folding or unfolding, or dynamic coupling between regions of IDPs.

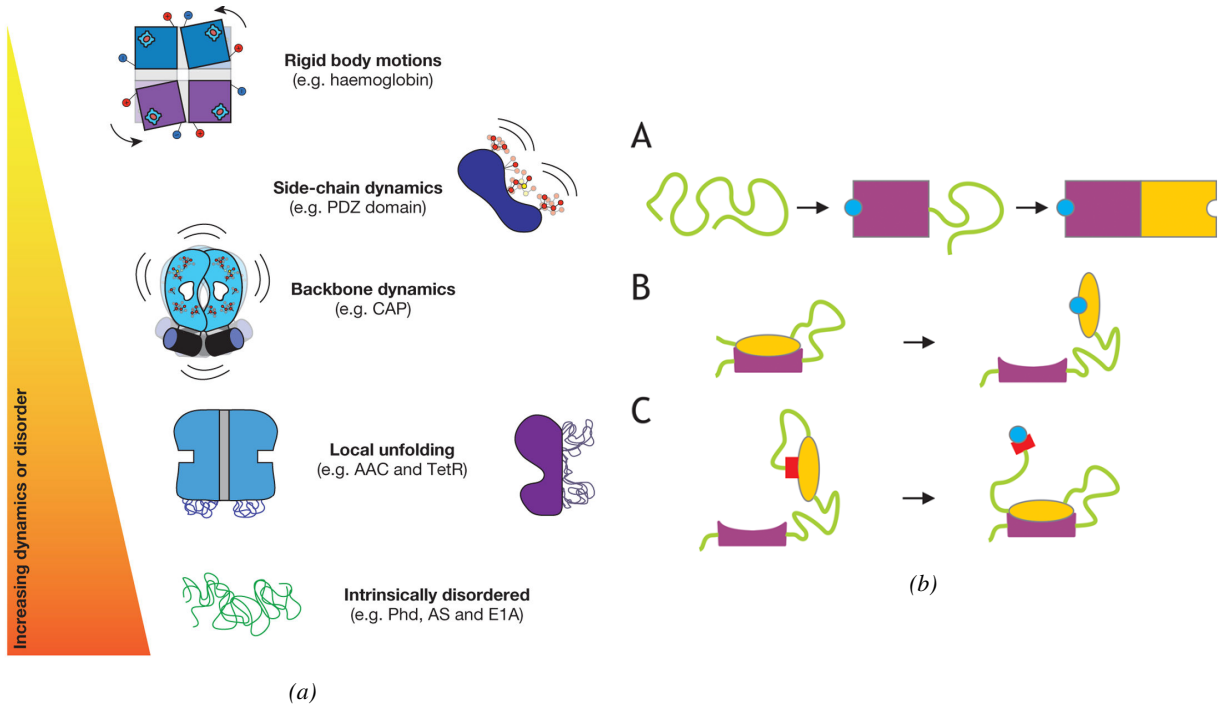


Figure 7 – Extension of the concept of allostery (a) taking into account different amounts of dynamics and disorder induced by ligand binding (Motlagh et al., 2014) and (b) taking into account different disorder-based mechanisms (Tompa, 2014).

Thermodynamics and kinetics of IDP binding In many cases, IDP binding to their partner is highly specific, but with a relatively low affinity (high dissociation constant K_D, in the order of μM or higher) (Zhou, 2012). Low affinity of IDPs is often attributed to the entropy cost of folding upon binding. Relatively high association rates are achieved through the so-called "fly-casting" mechanism, using the great capture radius allowed by flexibility, and thanks to adaptability of IDPs that reduces the need for "correct" orientation of the partners to bind. This leads to an increased number of "productive" encounter complexes, effectively leading to the formation of the final complex (Huang and Liu, 2009). Even in the case of coupled folding and binding, binding is often not drastically slowed down by folding, thanks to preformed structure in the free form.

For structured proteins, specificity is often associated to high affinity. How is IDP specificity achieved then? A first explanation is the usually extended interaction interface of IDPs, using short linear motifs; a second

feature is the existence, in many cases, of several secondary interaction sites (Didry et al., 2012; Delaforge et al., 2018). These characteristics facilitate a precise fit of the IDP to its target, and contribute to enhanced relative affinity, especially in the case of folding upon binding.

Several models exist to explain the peculiar behavior of IDPs regarding binding. Oldfield et al. (2005) explained the decoupling between affinity and specificity in the model of conformational selection, as shown in figure 8a. In such a model, the two-step mechanism, with folding first that only affects affinity, and then the binding step itself, on which depend both affinity and specificity, is the key to separately tune them. To explain the high association rates of IDPs to folded partners, a "dock-and-coalesce" model was described by Zhou et al. (2012) and is presented in figure 8b. The docking step is often the rate-limiting step, and is fast thanks to the low requirement for a specific orientation of the encounter.

There are of course counter-examples to this general model, and IDPs can bind with high affinity. A striking example is the recent description of the pM fuzzy complex between histone H1 and prothymosin-α, resulting from strong electrostatic interactions (Borgia et al., 2017).

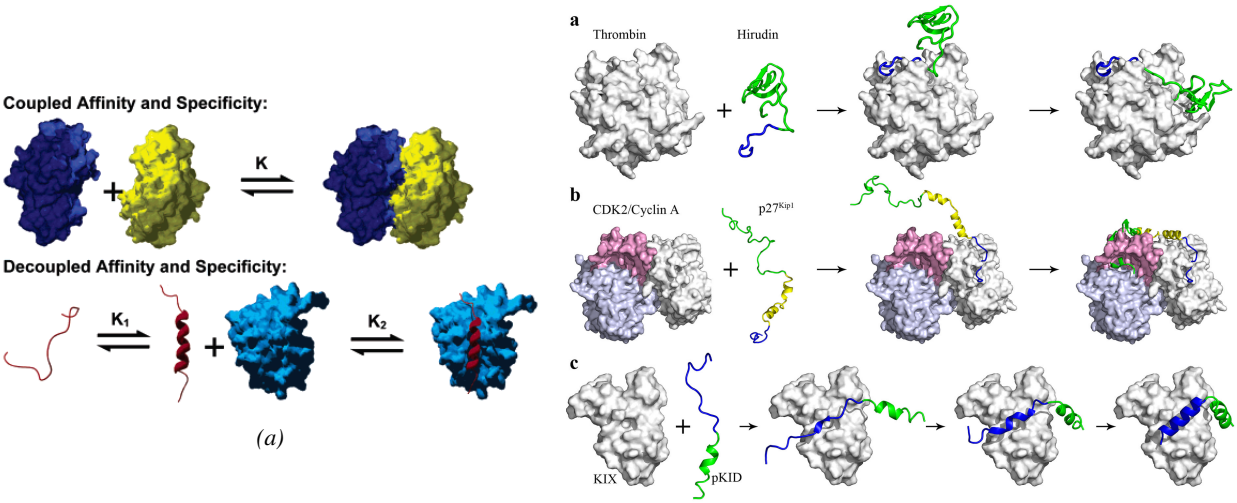


Figure 8 – Binding properties of IDPs (a) How disorder can decouple affinity and specificity, from the extreme case of conformational selection of an IDP (here Bad, binding to Bcl-xL). As opposed to proportionality between affinity and specificity expected for two rigid structured partners, with only one binding step, affinity can be independently tuned by the folding step of an IDP. From Oldfield et al. (2005). (b) The "dock-and-coalesce" mechanism proposed by Zhou et al. (2012), where the IDP docks to its partner by the interaction of a small segment, and then coalesces in a second step to form the complete binding interface.

1.4.2 IDPs in signaling

The low-affinity/high-specificity binding of IDPs is very useful in the context of signaling and regulation: relatively low affinities favor transient interactions, with high association and/or dissociation rates. More than globular proteins, IDPs are sensitive to various regulation mechanisms: they are prone to alternative splicing, experience high numbers of post-translational modifications, can modulate their accessibility to partners depending on conditions, and can undergo fast degradation and recycling. Consistently with these two characteristics, IDPs are overrepresented in signaling and regulatory processes: 66% of proteins in cell-signaling are predicted to contain disordered segments longer than 30 residues (Iakoucheva et al., 2002), allowing for rapid response to diverse inputs (Uversky et al., 2005).

Intrinsically disordered regions are found in many transmembrane proteins (Kjaergaard and Kragelund, 2017). As in the case of ErbB receptors C-terminal tail, the disordered regions is most often located in the cytoplasmic region of plasma membrane receptors (Minezaki et al., 2007). As reviewed by Kjaergaard and Kragelund (2017), the associated functions include formation of signaling complexes and regulation of the activity of a folded domain. For the first function, disordered segments can provide many short binding sites (SLiMs), that are adaptable and dynamically independent, in a relatively short sequence. Several examples are known, amongst which the N-methyl-D-aspartate (NMDA) disordered tail that is able to recruit many different proteins (CaM, CaMKII, PKC, PP2B, α-actinin), some of them binding to multiple sites (Strack and Colbran, 1998; Bayer et al., 2001; Leonard et al., 1999; Ehlers et al., 1996; Wyszynski et al., 1997). For the regulation of folded domain activity by disordered segments in transmembrane proteins, it has been reported in some channels (Goldin, 2003). In the case of receptor tyrosine kinases, and in particular ErbBs, the two roles of disordered regions described here can be relevant: the tail acts as a hub for adaptor proteins triggering signaling, and the tail is suspected to regulate kinase activity.

1.4.3 IDPs and diseases

Thanks to their chemical and structural adaptability and their binding promiscuity, IDPs are key to many cellular functions. As such, they are tightly regulated. This also implies that failure in such regulation can have dramatic consequences.

IDP aggregation and amyloidogenesis have been shown to be involved in many diseases, especially neurodegenerative diseases (Uversky, 2015). Amongst the most studied ones are α-synuclein, involved, *inter alia*, in Alzheimer's (AD) and Parkinson's (PD) diseases, tau, involved in AD , and Aβ, involved in AD

and Down's syndrome. Interestingly, the interactions of IDPs that lead to amyloidogenesis are high-affinity interactions, as opposed to usually observed low-affinity interactions of IDPs, with very stable supramolecular assembly. Disorder has also been shown to be overrepresented in cancer-associated proteins (CAP). 79% of human CAP have predicted disordered regions of 30 residues or more (Iakoucheva et al., 2002) (as opposed to an estimation of 44 % in all human proteins (Oates et al., 2013)). Such cancer-associated disorder-containing proteins include p53, BRCA1, PTEN and CBP amongst others. Other diseases often involving IDPs include cardiovascular diseases and diabetes. IDP-related diseases can originate from mutations, chromosomal translocation, altered levels of expression/degradation, aberrant alternative splicing or post-translational modifications (Uversky, 2014). Rational drug design approaches have to be adapted to the specificities of IDPs, and strategies to target them are beginning to emerge (Ruan et al., 2019).

1.5 Structural biology of IDPs

1.5.1 The goal: reconstruction of conformational ensembles

Having no stable structure, IDPs are better described as collections of rapidly interconverting conformers that we call conformational ensembles. The challenge for structural biology of such objects is therefore to gain insight into this ensemble from data that are usually an average of virtually an infinite number of states. This implies building model ensembles with enough structures to recapitulate experimental data. In these structures, the features that will be of particular interest will be the presence of partially formed secondary structures (potential MoRFs) and long-range contacts (intramolecular interactions) that can regulate accessibility of different regions of the protein and be critical for interactions and folding.

Even more than for structured proteins, IDP function is largely governed by dynamics. The second aspect of structural biology of IDPs is therefore to describe the different timescales of interconversion between these structures, and local dynamics of each of them, and to link these dynamic features to the timescale of biologically-relevant events. Dynamics will be particularly important to understand entropic chain functions, but also mechanisms of interaction.

Given that regulation mechanisms are numerous and particularly important for IDP function, the conformational ensemble of an IDP can be switched during its life cycle, for example by a post-translational modification or the binding of a partner. An extreme case of ensemble switching is the folding of the IDP 4E-BP2 into a four-stranded β domain upon multiple phosphorylation, described by Forman-Kay and col-

leagues (Bah et al., 2015). Comparing structural ensembles from the different potential states of the protein is thus crucial.

1.5.2 NMR, the choice atomic-resolution method to study flexible objects

Although disordered regions of proteins were historically (and sometimes still are) identified by crystallography (absence of crystal obtention or missing electron density), NMR has proven to be the best technique for atomic-resolution study of intrinsically disordered segments. It allows for both structural and dynamic description of such proteins, in conditions that are relatively close to physiological compared to other *in vitro* methods. Contrary to many other techniques, flexibility is rather favorable in NMR (if the dynamics if fast enough) as it enhances signal-to-noise ratio thanks to favorable nuclear spins relaxation properties for an IDP compared to a globular protein of the same size.

Challenges of IDPs NMR and specific requirements NMR for the study of IDPs suffer from the general lack of sensitivity of the technique, requiring high quantity of proteins of the order of the mg, even though it is less critical than for globular ones. High purity is important for structural biology, and even more for IDPs since they are more accessible to proteases and can be degraded rather fast.

On the point of view of spectroscopy itself, several problems are common for IDPs. The most obvious one is the poor spectral dispersion of proton signals (and especially amide protons), that all experience a similar environment, dominated by the interaction (and exchange) with the solvent. This phenomenon is increased by poor sequence diversity commonly encountered in disordered proteins, that can be overcome by specific labeling strategies (Urbanek et al., 2018). Therefore, (very) high fields are usually required for such studies. Accessibility to solvent also is an issue when the rate of exchange with water is in the intermediate/fast regime, and causes line broadening/extinction of the signal. That often prompts spectroscopists to choose low pH, low temperature conditions, to slow it down, with the risk to get further from physiological conditions. More generally, flexibility in the "wrong" timescales (that is, leading to an intermediate exchange regime) can lead to troublesome line broadening. That is a problem more commonly observed for (pre-) molten globules than for random coils. In these cases, adapting the magnetic field can be useful to fall back into the fast or slow exchange regime.

The last issue that is common to many IDPs for NMR studies is the high occurrence of prolines. Many routine protein-NMR experiments, starting with the ^{1}H-^{15}N HSQC, are backbone amide-proton based. Prolines lacking those amide protons lead to missing information along the protein sequence. This is partic-

ularly troublesome for sequential assignment, and "proton-less" NMR experiments (HCAN, CON... see section 5.1.2) have to be implemented to fill in the gaps and assign proline-rich regions.

Information accessible by NMR Here I describe the possibilities that NMR offers for the description of IDPs. The detailed description of each type of experiments are given in the experimental section 5.
A number of NMR parameters are rather easily accessible and give structural information. It is not to be forgotten that these parameters also depend on dynamics, since dynamics controls their averaging (or non-averaging). Local (secondary) structure of IDPs is mainly probed via chemical shifts, J couplings and residual dipolar couplings (RDCs). While chemical shifts can directly give access to rather precise estimates of secondary structure propensities (via different calculation methods such as δ2D (Camilloni et al., 2012) and SSP (Marsh et al., 2006), see section 5.2), quantitative interpretation of J couplings (see section 5.4) and RDCs (see section 5.5) is indirect and often requires comparison with simulated ensembles.
RDCs, and to a lesser extent chemical shifts, are also sensitive to transient long-range contacts. It is however difficult to decouple contribution from tertiary and secondary structure. Once again, comparison with back calculated values from simulated ensembles or between different conditions is necessary. With nOes being very weak (and often not detectable) for transient contacts in such flexible proteins, the method of choice to directly detect long-range intramolecular interactions in IDPs is to use paramagnetic relaxation enhancement (PRE) measurements, probing contacts in different regions of the protein (one by one, see section 5.6).
Dynamics can be probed by a number of different NMR experiments, depending on the timescale of the movements observed, as shown in figure 9. The most common type of experiments is ^{15}N relaxation measurements (R_1, R_2 and ^{1}H-^{15}N nOe, see section 5.3) that probe local dynamics, that is motions in the range of 100 ps-ns. To probe slower motions corresponding to either slow local motions or more global conformational changes, relaxation dispersion (μs to ms), CEST experiments (ms to s), or other magnetization transfer experiments (ms to dozens of s) can be used. For even slower motions, sequential NMR experiments (from simple 1D spectra to any type of longer experiments depending on the needed timepoints) can be recorded allowing monitoring of processes slower than the second range.
The NMR community for the study of disordered proteins is very active, and a lot of very good reviews exist on the topic, amongst which Kosol et al. (2013); Konrat (2014); Milles et al. (2018).

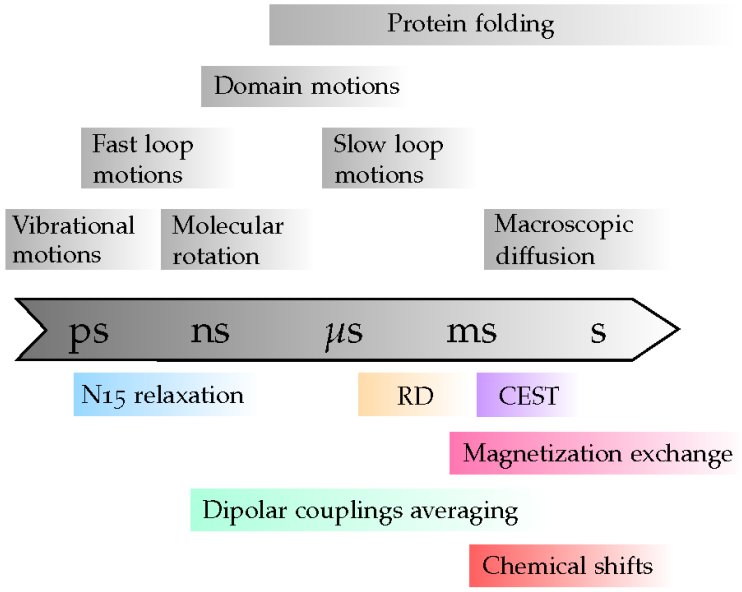


Figure 9 – Molecular motions in proteins (top) at different timescales, and NMR observables (bottom) allowing to probe them. From Deville (2015).

1.5.3 Complementary methods

NMR is very powerful to probe local structure and dynamics of flexible proteins, as well as intra- and intermolecular interactions. However, it is difficult to get information on the overall dimensions of the objects in solution by NMR. The dedicated technique to obtain this information is SAXS. By fitting the scattering curve with different models, size, shape, and compactness parameters can be calculated. More details are given in section 6.2. The drawback of both NMR and SAXS is that they only give information on the average over all the conformations in the ensemble that are exchanging fast enough compared to the experimental sampling. Two types of study of IDPs try to overcome this limit.

Experimentally, single-molecule analysis is rapidly emerging, with single-molecule fluorescence, single-molecule force spectroscopy, and single-nanopore studies. These techniques enable one to discriminate between ensemble and dynamic averaging (static vs. dynamic heterogeneity), and follow sequential events, for they have the best ensemble resolution possible (single particle), and a very broad range of timescales accessible. A method often used for IDPs is single-molecule FRET, that gives non-ensemble averaged distance information, contrary to NMR and SAXS. The application of single-molecule methods for the study of IDPs was recently reviewed by Brucale et al. (2014).

The other approach is to use simulations. Molecular dynamics allow, in combination with experimental techniques, to gain mechanistic insight at the atomic level. Major progresses have recently been made in the

development of specific force fields, adapted for disordered proteins, as opposed to force fields developed for globular proteins that often favor compaction. This implies adjustment of the potentials for backbone dihedral angles, and/or solvation. Efforts to optimize electrostatic potentials, often more critical for IDPs than for globular proteins, are also made. The last challenge is to ensure a sufficient sampling of conformational space. Given the high number of degrees of freedom, short simulations only explore a small portion of the energy landscape accessible to the IDP. Enhanced sampling methods are needed, and some have been developed, such as replica-exchange. These developments were very recently reviewed by Kasahara et al. (2019).

Some insight can also be gained by generating "static" conformational ensembles, to build models. Back calculation of experimental parameters from these models allows characterization of specific features compatible with experimental data. The first step is the generation of atomic models of the disordered protein, with the possibility to apply structural constraints to this ensemble. This can be done using Flexible Meccano, developed by Blackledge and colleagues (Ozenne et al., 2012). One can directly compare experimental data with the generated ensemble, or perform a selection step to obtain sub-ensembles fitting them. Two programs exist, the ENSEMBLE program developed by Forman-Kay and colleagues (Krzeminski et al., 2013), and the ASTEROIDS program developed by Blackledge and colleagues (Nodet et al., 2009). This type of approach, using NMR and SAXS data, has been successfully used on several systems, as reviewed by Sibille and Bernadó (2012).

The present work, studying the disordered tail of a receptor tyrosine kinase, is in line with ongoing research in the IDP field, and tackles many of the IDP-related issues: the role of a long disordered part in an overall structured protein; the influence of prolines on disordered states; the mechanism and influence of post-translational modifications; the role of disorder in protein-protein interactions and signaling; and the role of disorder in diseases. The main technique used was NMR, to investigate structure, dynamics, and interactions of CtErbB2, complemented by SAXS measurements, and generation of structural ensembles with Flexible Meccano.

2 The C-terminal tail of ErbB2 (CtErbB2)

2.1 The ErbB/EGFR/HER family of receptor tyrosine kinases

Receptor tyrosine kinases (RTKs) are key proteins for regulation of cell proliferation, cell motility, differentiation and apoptosis (Lemmon and Schlessinger, 2010). They are essential for development and several physiological processes, but are also involved in diseases, especially cancer (Du and Lovly, 2018). There are 58 RTKs in humans (Robinson et al., 2000), divided into 20 classes defined by phylogeny.

ErbB receptors were the first ones to be identified, and are therefore referred to as class I RTKs. ErbB1 was discovered by O'Keefe et al. (1974) as the epidermal growth factor receptor (EGFR). It was shown to have tyrosine kinase activity a few years later by Cohen et al. (1982). In 1984, sequencing unveiled its homology to a viral oncogene, v-erb-B (erythroblastic leukemia viral oncogene B) (Downward et al., 1984), and the ErbB family was named after it. It is also called the HER (Human Epidermal growth factor Receptor) or EGFR family, and contains four members: ErbB1/EGFR/HER1, ErbB2/neu/HER2, ErbB3/HER3, ErbB4/HER4, presented in figure 10.

As for all RTKs listed to date (Hubbard, 1999), ErbB receptors are made of a ligand-binding extracellular domain (ECD), a transmembrane helix (TMH), an intracellular juxtamembrane region (JM, composed of two parts JMA and JMB), a tyrosine kinase domain (TKD) and a C-terminal tail (Ct), also called regulatory domain. The whole receptors are about 140 kDa, and up to 185 kDa when they are glycosylated.

Despite this common domain organization of all four ErbBs, functional differences exist between them and are shown in figure 10. Most strikingly, no ligand has been found for ErbB2, and ErbB3 kinase has been shown to be very weakly active (about 1000 fold less active than EGFR) (Brignola et al., 2002a) . From the evolutionary point of view, it was determined that in invertebrates there is only one ErbB-like protein (called let-23 in *C. elegans*), and that the number of ligands is small (only one ligand identified in *C. elegans*, 4 in *Drosophila*). Following three steps of gene duplication (separating ErbB1/2 from ErbB3/4 first, with then another gene duplication of each), the four distinct ErbBs appeared in vertebrates, in which ligand diversity was also increased (Stein and Staros, 2006). Although EGFR and ErbB2 have different functional specificities, ErbB2 is therefore the closest homolog of EGFR on the family.

Hereafter I describe the physiological functions of ErbB proteins, the molecular origin of these functions, and their implication in diseases. Most of the conclusions were drawn from studies of EGFR, which is the model ErbB receptor. Their validity for ErbB2, the focus of this work, are discussed in section 2.2.

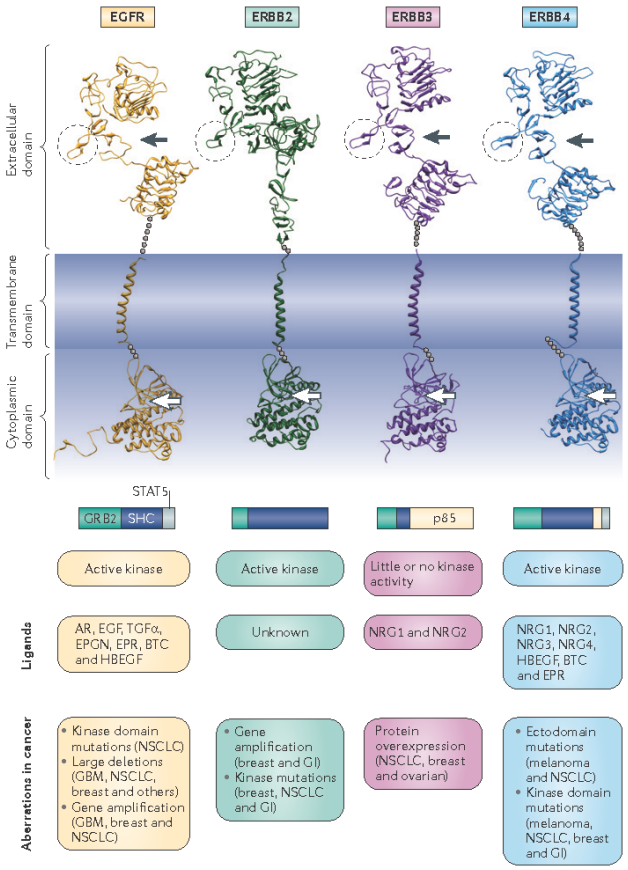


Figure 10 – The four ErbBs. (top) Crystal structures of the three main structured domains (each domain crystallized independently) of the four ErbB proteins. The extracellular domain (ECD) structures are in the tethered (ligand-free, inactive) state, except for ErbB2, which only exists in the dimerization-prone conformation (see section 2.1.3). Domains IV of the ECD for ErbB1/3/4 are not shown for clarity, but mask the dimerization loop. Black arrows: ligand-binding cleft. Dashed circle: dimerization loops. White arrow: ATP-binding cleft. The C-terminal tails are omitted in this figure, for no structure is known. Figure from Yarden and Pines (2012). (bottom) Main characteristics of each ErbB in terms of interaction partner, kinase activity, ligand binding and aberrations observed in cancer.

2.1.1 Cellular function (and disfunction) of ErbBs

ErbB receptors ensure a wide range of roles in development and adult organ function (Burden and Yarden, 1997; Chan et al., 2002), and null mutation of any of them is lethal at an early stage of embryogenesis. More specifically, ErbBs are involved in development of mammary glands (Sebastian et al., 1998), peripheral nervous system and Swann cells (Lee et al., 1995; Gassmann et al., 1995; Britsch et al., 1998; Riethmacher et al., 1997), and in cardiac development and function (Lee et al., 1995; Gassmann et al., 1995). At the cellular level, ErbBs are mitogenic (Pinkas-Kramarski et al., 1996; Vaskovsky et al., 2000), but are also involved in cell adhesion (Jeon et al., 2015), motility (Chausovsky et al., 2000; Marone et al., 2004), and

differentiation (Vaskovsky et al., 2000; Giani et al., 1998). They were shown to perform these functions by activating (mainly) the MAP-kinase, PI3K/Akt, Src kinase and STAT5 pathways (see figure 11), and more recently the MEMO pathway, described by Badache and colleagues (Marone et al., 2004), that controls cell motility.

Misregulation and mutations of ErbB proteins have been linked very early to cancer (Downward et al., 1984; Schechter et al., 1984). Different receptors, or receptor combinations, are involved in different types of cancer (Yarden and Sliwkowski, 2001; Yarden and Pines, 2012), as detailed in figure 10. Additionally, ErbBs are associated with neurodegeneration, and especially Alzheimer's disease (Chaudhury et al., 2003), cardiovascular diseases (Nakata et al., 1996) and psoriasis (Ben-Bassat and Klein, 2000). The nature of the ErbB-related diseases is consistent with their physiological tissue localization.

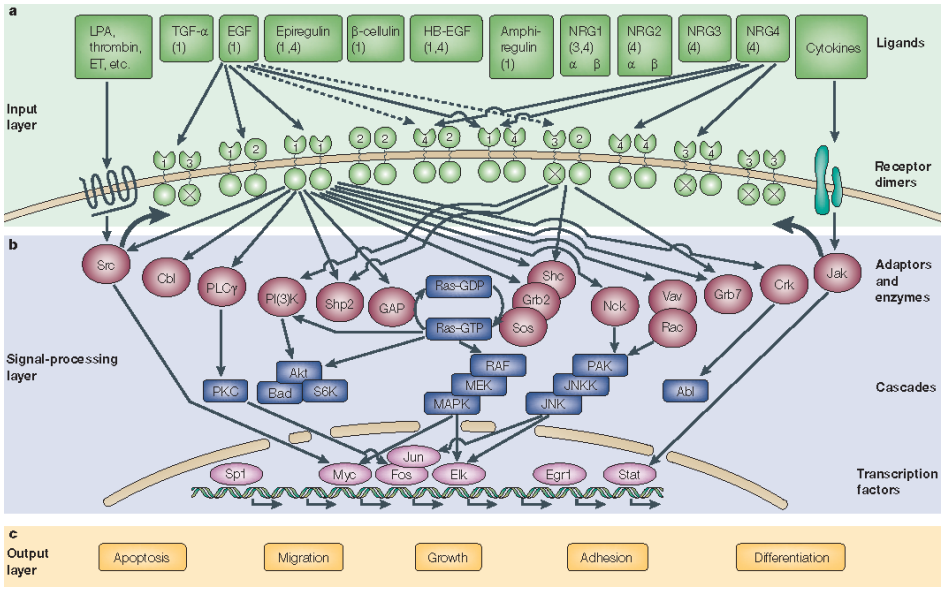


Figure 11 – Main ErbB signaling pathways, from the example of the EGFR/EGFR and ErbB2/ErbB3 dimers. Figure from Yarden and Sliwkowski (2001). The MEMO pathway, discovered more recently (Marone et al., 2004), is not shown here.

2.1.2 Overall mechanism of signal transduction

Given their implication in both important physiological functions and cancer, a lot of effort was put into understanding the mechanism of signal transduction by ErbBs. Structural studies were key to this under-

standing. Briefly, ligand binding to ErbB1, ErbB3 or ErbB4 ECD is followed by a conformational change that enables formation of active homo- or hetero-dimers of the receptors. ErbB2, on the contrary, does not bind any known ligand and is constitutively poised to dimerize. Dimerization enables the activation of the intracellular TKD of a least one monomer of the dimer and then leads to tyrosine phosphorylation of the C-terminal tail. Finally, the phosphorylated tails serve as anchors for interaction with adaptor proteins that trigger signaling. A very schematic view of the steps is given in figure 12, and the role of each domain is detailed in the following sections.

Understanding the general mechanism of signal transduction by ErbB receptors is crucial to understand the role of their C-terminal tails, their different states (signaling potent or not), their direct or indirect interactions with the other domains, and the relevance of the constructs to use for their study.

The numbering used in these sections is given in figure 12b for EGFR, the most studied ErbB, and ErbB2, the focus of our work.

2.1.3 The extracellular domain (ECD), ligand binding, and dimerization

ErbBs ECD is composed of about 630 residues divided into four domains (see figure 12b). Domains I and III are also called L1 and L2 domains, respectively, while domains II and IV are called cysteine-rich domains CR1 and CR2, respectively. The historical ligand EGF, that only binds ErbB1, has been joined by 10 other EGF-family ligands binding ErbB receptors (see figure 10).

Crystal structures of EGF-bound (Ogiso et al., 2002) and free (Ferguson et al., 2003) ECD of EGFR revealed the first step in ErbB dimerization (figure 13): in the free, autoinhibited state, the dimerization site in domain II is masked by interaction with domain IV. Upon ligand binding, mainly to domains I and III, a large conformational change frees domain II from intramolecular interaction with domain IV, and permits dimerization of the ECD. This conformational change for activation was confirmed *in vivo* using mutational studies and antibody-binding (Arndt-Jovin et al., 2014). Contrary to the dimerization mechanism of some other RTKs, the ligand itself does not dimerize. For each of the ErbB1/ErbB3/ErbB4 receptors, binding of different ligands leads to different types of dimers and different dimer association constants, with different signaling specificities (different pathways activated, different signaling intensities and signal lifetimes...) (Wilson et al., 2012; Macdonald-Obermann and Pike, 2014; Freed et al., 2017).

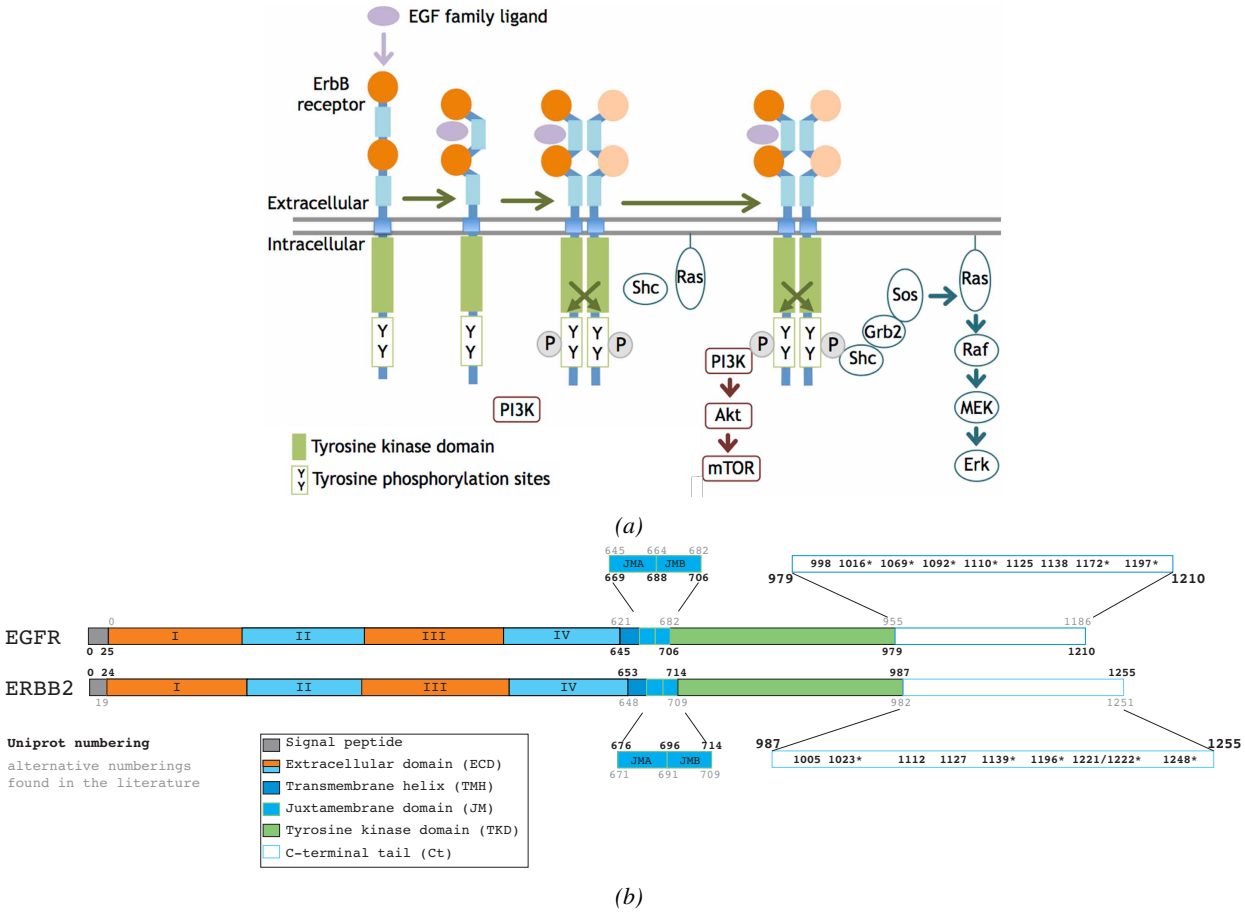


*Figure 12 – ErbBs domains function (a) Schematic representation of the role of each domain in ErbB signaling. Adapted from Iwakura and Nawa (2013). Only some of the signaling pathways are represented here. (b) Overall structure of EGFR and ErbB2 with numbering. The Uniprot numbering is used (unless otherwise stated) in this manuscript. For EGFR, the alternative numbering sometimes used excludes the signal peptide in the N-terminus. For ErbB2, the alternative numbering is that of rat ErbB2 (neu). In the C-terminal tail, all tyrosine positions are indicated, with * on the ones that are known autophosphorylation sites.*

In addition to this ligand-induced dimerization model, a ligand-induced activation of inactive dimers was also proposed. Indeed, preformed, ligand-free inactive dimers were shown to exist *in vitro* and *in vivo* (Moriki et al., 2001; Tao and Maruyama, 2008). In this model, the described ligand-triggered conformational change of the ECD happens in the preformed dimer. The subsequent conformational changes in the transmembrane and intracellular domain (described bellow) activate the kinase domain in the so-called "rotation model", reviewed by Purba et al. (2017). Between the two models, only the starting points really differ, but the activation mechanism itself is the same. Whether the monomers or inactive dimers predominate might actually depend on the nature of the receptor, level of expression, and localization in the membrane, since

membrane composition is not homogeneous (Arndt-Jovin et al., 2014).

Structural studies of the ECD were performed on monomers or homodimers. No structure of heterodimeric ECD of ErbB receptors has ever been described. It has been suggested that dimerization of the ECD only drives homo- but not hetero-dimer formation, that would be driven by other regions of the receptor (Ferguson, 2000).

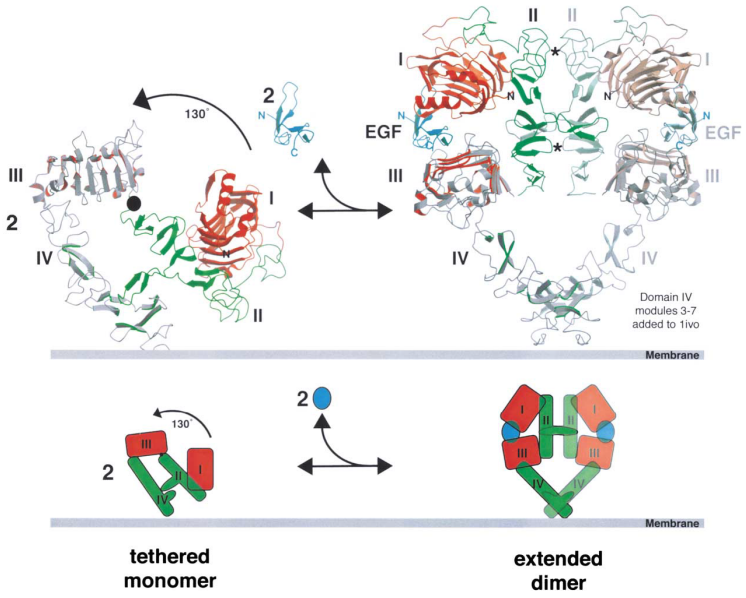


Figure 13 – EGFR auto-inhibition by its extracellular domain is released by binding of a ligand (here EGF, blue). Contacts in the dimer are made through the dimerization loop, and two other sites of domain II marked by asterisks. Figure from Burgess et al. (2003).

2.1.4 The transmembrane helix (TMH) and N-terminal end of the juxtamembrane region (JMA)

It has been shown that dimerization of the ECD itself is not sufficient for ErbB receptors signal transduction, but that other intermolecular interactions allowed by the more global conformational change are required (Yarden and Schlessinger, 1987). Consistently, truncated forms of ErbBs lacking the ECD are active (Downward et al., 1984), suggesting that the (ligand-free) ECD is auto-inhibitory. In particular, the TMH has been shown to enhance ligand-induced homodimerization of the ECD (Tzahar et al., 1996), and isolated TMH homo- and hetero-dimerize (Mendrola et al., 2002; Sharpe et al., 2002; Duneau et al., 2007). Furthermore, mutations in the TMH are involved in overactivation of ErbBs and diseases. For example, the Val to Glu mutation in position 659 (664 in another numbering in use, see figure 12b) of neu/ErbB2 is activating, and

is found in rat neuroglioblastoma (Bargmann and Weinberg, 1988).

The effect of this mutation, and the origin of activation of the receptor, has been suggested to be due to rotation of the TMH to form a hydrogen bond with the other TMH of the dimer (Bell et al., 2000). A recent NMR study by Bragin et al. (2016), using a ErbB2 construct containing the TMH and beginning of the juxtamembrane region (juxtamembrane region A, JMA), supports the importance of the rotation of the TMHs. This rotation, stabilized by a GxxxG motif, enables the C-termini of the helices to "open" in the activated state (see figure 14). Consistently with previous studies (Red Brewer et al., 2009; Endres et al., 2013; Zhang et al., 2006), in this "open" conformation JMA regions dimerize and are antiparallel. Conversely, the suggested model for the inactive state shows a membrane embedded, parallel, JMA dimer. Interaction of the intracellular part of ErbBs with the plasma membrane had already been suggested to be inhibitory (Arkhipov et al., 2013).

Even though the conformation and dimerization states of the transmembrane helix of ErbB receptors seem to link the effects of ligand binding to the ECD to conformational changes in the intracellular domains, the link has been suggested to be quite loose (Lu et al., 2010; Mi et al., 2011), with both active and inactive kinase dimers found in ligand-bound receptors. Reality is most probably way more complex than duality between all-active and all-inactive states.

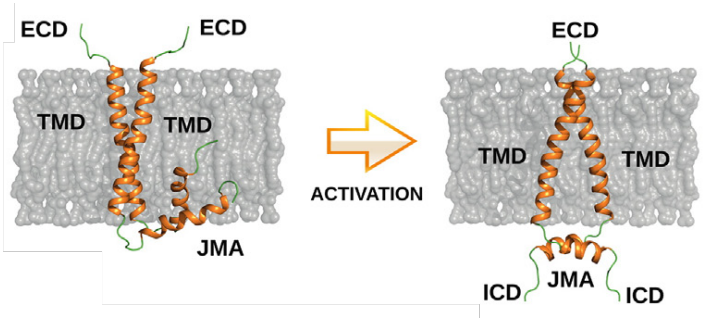


Figure 14 – Model of the conformation of the TMH and JMA of ErbB2 in the "inactive" and "active" states. From Bragin et al. (2016)

2.1.5 The tyrosine kinase domain and its activation

A non-canonical activation In many RTKs, monomeric kinase activity is low but non-zero. Dimerization of the receptors causes *trans*-tyrosine phosphorylation of the kinase or juxtamembrane domains, which dramatically increases kinase activity by changing the conformation of the activation loop (Hubbard, 2004). Such a mechanism was initially expected for ErbBs, and the expected *trans*-phosphorylation site in the ki-

nase is the conserved tyrosine Y869 in EGFR. However, Tyr to Phe mutagenesis studies (Gotoh et al., 1992) showed that optimal kinase activity in the dimers does not require phosphorylation of Y869. In accordance with that observation, the crystal structure of EGFR unphosphorylated kinase domain, solved by Stamos et al. (2002), showed a conformation typical of the active state of a tyrosine kinase. A structure resembling the inactive, unphosphorylated state of other kinases was described two years later, with a tyrosine kinase inhibitor (Lapatinib) bound (Wood et al., 2004).

Activation by asymmetric dimerization Without Tyr *trans*-phosphorylation as a mean to activate the kinase, the activation mechanism and necessity for dimerization was not understood. A closer look at the crystal lattice of the active EGFR from (Stamos et al., 2002) showed two main possible dimers: a symmetric one, and an asymmetric one, shown schematically in figure 15. Zhang et al. (2006) showed by mutagenesis of the interface residues and concentration-dependent activity assays that the active form is the asymmetric dimer. This dimer is similar to the CDK/cyclin complex, and one kinase monomer plays the role of an *activator*, promoting via intermolecular interactions the activation of the other, *receiver* kinase. The N-lobe of the *receiver* contacts the C-lobe of the *activator* with mainly hydrophobic interactions, and their relative position is characterized by a 3-fold screw rotation (120° rotation, 43 Å translation).
The mechanisms determining the relative position of each kinase (*receiver/acceptor*) is not clear. It has been proposed that it is dependent on the binding of ligand to each monomer (Macdonald-Obermann et al., 2012), but other studies suggested that it is intrinsic to the composition of the dimer, each receptor having a certain propensity to be the *receiver* (and EGFR being the "best *receiver*") (Ward and Leahy, 2015).

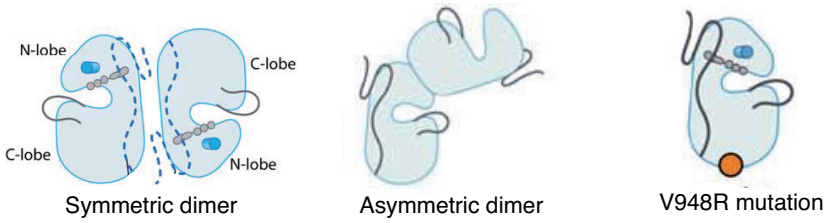


Figure 15 – Schematic representation of EGFR kinase symmetric dimer (left), asymmetric dimer (middle) and V948R mutation that allowed to determine that the asymmetric dimer is the functional one. From Zhang et al. (2006)

Role of the C-terminal juxtamembrane region (JMB) The asymmetric dimer supposes asymmetric positions of the juxtamembrane regions (JM), and especially their C-terminal moiety (JMB) which are not present in Zhang's construct. This JMB region was shown by Thiel and Carpenter (2007) to be important for kinase activity. Red Brewer et al. (2009) crystallized a construct containing JMA, JMB and the kinase domain. At the same time, Jura et al. (2009) compared the juxtamembrane-containing crystal structure of the kinase domain of ErbB4 (Wood et al., 2008) and the symmetric, inactive dimer structure of the V948R mutant to understand the role of this region. Both studies converged in saying that interaction of the JMB segment of the *receiver* with the C-lobe of the *activator* is necessary for activation, by acting as a "latch" thats locks the active dimeric conformation.

Model of activation in the whole transmembrane and intracellular domain A model recapitulating activation the mechanism of the kinase domain from the most recent studies is given in figure 16. From the observation that ErbB3 tail can be phosphorylated even though it is kinase dead, and other experiments with kinase-dead constructs (Qian et al., 1994), it is known that *trans*-autophosphorylation (phosphorylation of the tail of one ErbB monomer by the kinase of the other ErbB of the dimer) is efficient. The co-existence of *trans*-autophosphorylation with *cis*-autophosphorylation (phosphorylation by its very own kinase) has however never really been ruled out.

In the symmetric inactive dimer stabilized by the V948R mutation (see figure 15), described by Jura et al. (2009), the contact between the two kinases is mainly mediated by the beginning of the C-terminal tail of the receptor. This may explain regulation mechanism of the kinase activity of EGFR by the C-terminal tail, discussed in section 2.1.6 below.

2.1.6 The C-terminal tail (Ct), regulation and signal transduction

Tyrosine phosphorylation sites The C-terminal tail primary role in ErbB receptors is to ensure signal transduction via interactions of phosphorylated tyrosines with signaling proteins (via SH2 and PTB domains mainly). These tyrosines are called autophosphorylation sites when they are phosphorylated by ErbB kinases in the dimer, as opposed to tyrosines phosphorylated only by other tyrosine kinases in the cell. There

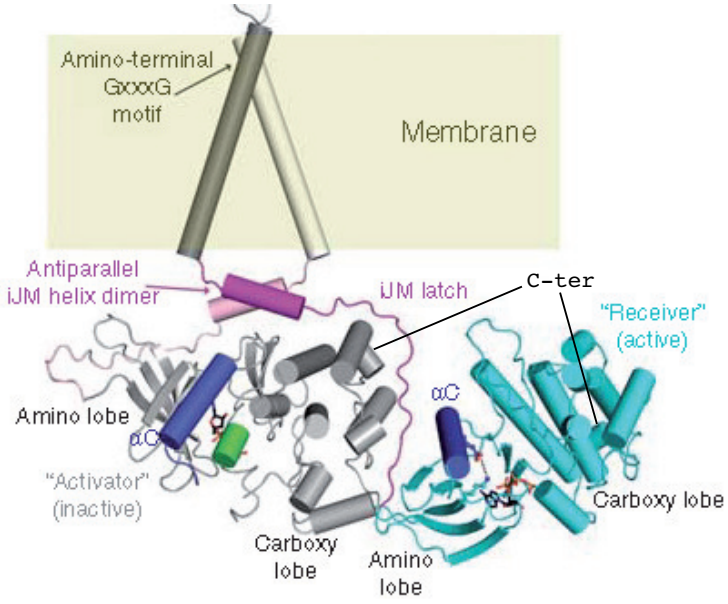


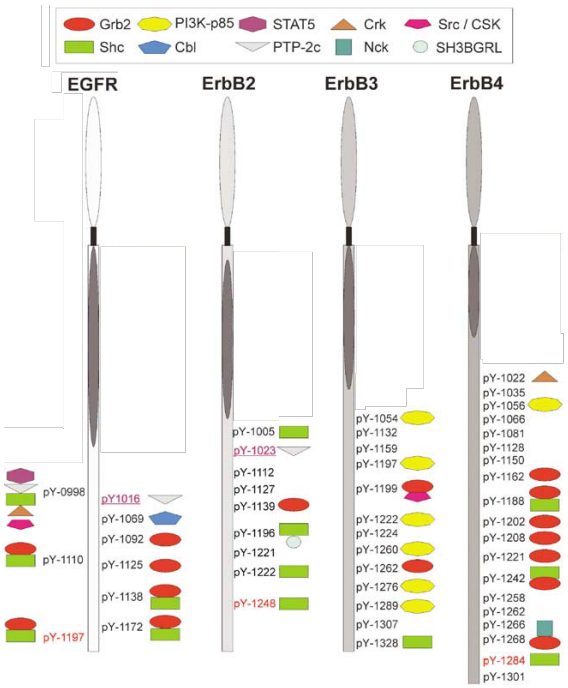
Figure 16 – Model of the active state of the transmembrane, juxtamembrane and kinase domains of EGFR, from Lemmon et al. (2014). The extreme C-terminal helix of each kinase, from where the C-terminal tail would start, is indicated.

are 9 tyrosines in EGFR C-terminal tail, amongst which six can be autophosphorylated. Four of them have been shown to be highly autophosphorylated sites (Y1092, Y1110, Y1172 and Y1197), and two have lower levels of phosphorylation (Y1016 and Y1069) (Bertics et al., 1985; Margolis et al., 1989; Walton et al., 1990). Phosphorylation studies of EGFR have been reviewed by Koland (2014), and a recapitulative table is shown in figure 17a. It was suggested that phosphorylation does not happen at many sites at once (Kim et al., 2016), but phosphorylation mechanism and patterns are widely unknown.

Partners of phosphorylated tyrosines of ErbBs Partners of ErbB phosphotyrosines include Grb2 and Shc (MAP kinase pathway), PI3K (Akt pathway), Src and STAT5 (Yarden and Sliwkowski, 2001; Schulze et al., 2005) (see figure 17b). Some partners are common between different ErbBs and some are not. We will focus of CtErbB2 partners in section 15. The C-terminal tail also includes motifs signaling for endocytosis and degradation/recycling, but these phenomena are not completely understood yet, and no direct interaction was shown (Sorkin and Goh, 2009).

P-site	*in vitro* studies	Studies in cultured cells
1016	+	-/+++
1069	++	+/++/+++
1092	+++	+/++/+++
1110	++/+++	++/+++
1125	ND	ND
1138	ND	+++
1172	+++	+/++/+++
1197	+++	+*/+++

(a)



(b)

*Figure 17 – Phosphorylated tyrosines of EGFR. (a) Review of different studies of the relative levels of phosphorylation of 8 (out of 9) tyrosines of EGFR C-terminal tail. *: greatly enhanced by ErbB2 expression. "/" indicates different results in different studies. Adapted from Koland (2014). (b) Phosphotyrosine interactome of the C-terminal tails of ErbB receptors. Adapted from Schulze et al. (2005).*

The C-terminal tail as a regulatory domain Additionally, the C-terminal tail is believed to participate in the regulation of the activity of the kinase domain, since it has been shown that deletion of some parts of the C-terminal tail modulates kinase activity. However, the results of the different studies (Walton et al., 1990; Sorkin et al., 1992; Alvarez et al., 1995; Cheng and Koland, 1996; Pines et al., 2010) diverge on the nature and precise regions of the tail involved in this regulation. This regulation could be phosphorylation-dependent, creating feedback loops. Indeed, C-terminal tyrosine to phenylalanine mutants of activated EGFR showed different kinase activities (Gill et al., 2017). Crystal structures (shown in figure 18) show the implication of the C-terminal tail in the kinase domain conformational state, supporting a regulatory function. In the inactive state, a helix is structured at the beginning of the C-terminal tail: this helix (995-1004) is sometimes called the AP-2 helix because it is responsible for binding to the AP-2 clathrin adaptor protein. It interacts with parts of the kinase N-lobe close to the N-lobe-C-lobe hinge region, which could alter activity. In the inactive, symmetric dimer of the V948R mutant (Jura et al., 2009), this helix also interacts with the N-lobe of the other monomer, and is followed by an extended part (1006-1014) interacting with the C-lobe

in a similar manner as the JMB in the active asymmetric dimer. This extended part instead forms another helix (1010-1018) in some other inactive and active structures (with varying length and position) interacting with the N-lobe. Molecular dynamics studies support the interplay between the C-terminal tail and JMB for kinase regulation (Mustafa et al., 2011). The C-terminal tail, at least its beginning, is even suspected to participate in the stabilization of the kinase core, since when the whole C-terminal tail removed, no crystals were obtained (Gajiwala, 2013).

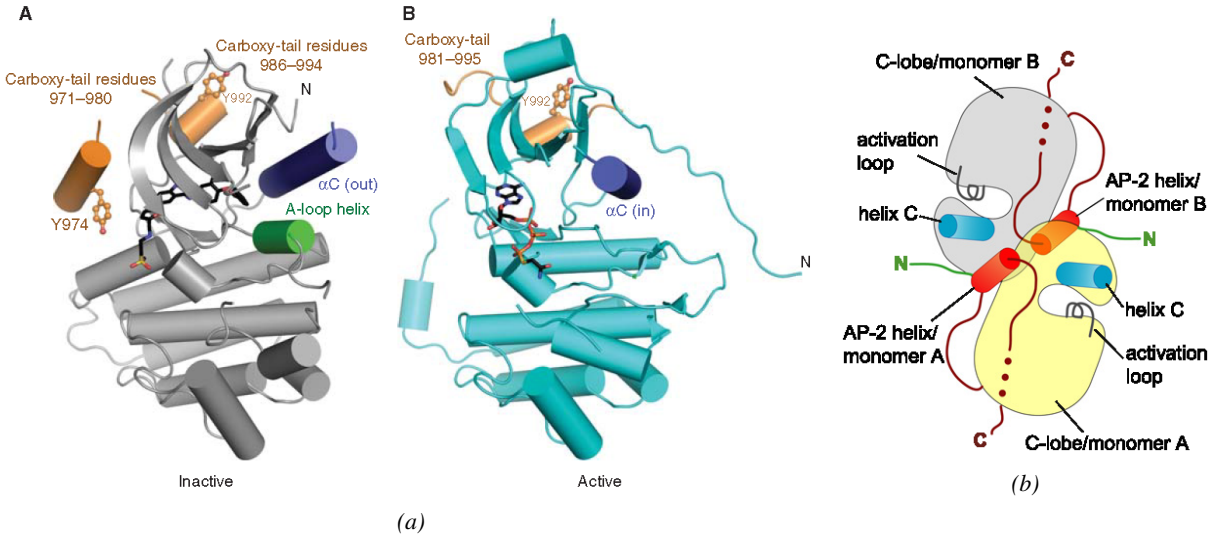


Figure 18 – Interaction of the C-terminal tail with the kinase domain of EGFR. (a) Inactive (Wood et al., 2004) and active (in the asymmetric dimer) (Zhang et al., 2006) kinase domain of EGFR showing interactions with the C-terminal tail. Side chains are shown for tyrosines (Y974, involved in endocytosis, and 992, autophosphorylation site) for which phosphorylation could disrupt interactions. Figure from Lemmon et al. (2014). (b) Inactive, symmetric dimer with the C-terminal tail preventing interaction of the JMB segment on the C-lobe. From Jura et al. (2009). Here the numbering is the alternative one (see figure 12b)

A model, as complete as possible today, of ErbB activation (in the "rotation model" supposing preformed inactive dimers) from ligand binding to phosphorylation of the C-terminal tail, is given in figure 19.

Order and disorder in ErbB C-terminal tails All the crystal structures only contain about 40 residues of the C-terminal tail of EGFR (out of 231). Flexibility, causing missing electron density even in the segments present in the crystals, is probably preventing crystallization of longer fragments. This is consistent with a

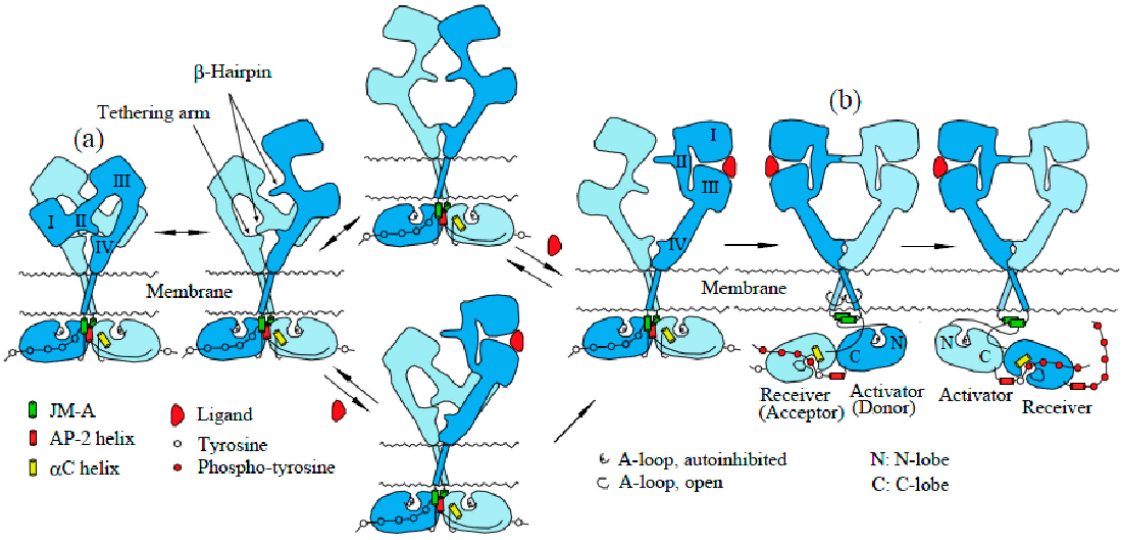


Figure 19 – Full model of ErbB activation in the "rotation model", supposing the existence of preformed, inactive dimers. From Purba et al. (2017).

study by Keppel et al. (2017) studying the C-terminal tails of ErbBs alone and in the context of the whole intracellular fragment, that uses a combination of biophysical techniques (HDX-MS, CD, SEC-MALS, DLS, AUC, SAXS) to show that they are disordered in EGFR and ErbB3, and at least solvent-accessible in ErbB2. Structural and dynamic features of the tails, and interaction with the kinase domain, are phosphorylation-dependent (Lee and Koland, 2005; Lee et al., 2006). However, the phosphorylation does not seem to cause any dramatic change in the secondary structure content (Lee et al., 2006), and what seems to be modified is the tertiary structure. For example, the observed contacts in the crystals could be modified by tyrosine phosphorylation (see figure 18). This is also what was suggested by a study by Bishayee et al. (1999): an antibody recognizes a conformation of the C-terminal tail of EGFR where two parts of the epitope, far in the sequence, are close in space. This recognition only happens when three tyrosines are phosphorylated (Y1016, Y1092 and Y1110) .

2.2 ErbB2/neu/HER2, a singular ErbB

ErbB2 was originally known as *neu*, since it was identified in rat neuroglioblastomas (Schechter et al., 1985), and its homology to EGFR/ErbB1 was found via their common similarity with the erythroblastic viral oncogene (Schechter et al., 1985). For historical reasons, EGFR has always been the most studied ErbB. However, interest for ErbB2 has been progressively growing, for it is involved in many different types of

aggressive cancer, and especially in breast cancer. Even though from an evolutionarily point of view ErbB2 is the ErbB closest to EGFR, lack of any known ligand, and peculiarities in its structure, have made it a good model of structure-function comparative studies.
Special emphasis is put in this section on features that seem specific to ErbB2 compared to mechanisms described in section 2.1.

2.2.1 ErbB2 and cancer

ErbB2 has been shown to be overexpressed (up to 100-fold, from about 10^4 to about 10^6 molecules/cell) in about 30% of breast cancers (Slamon et al., 1987). These cancers are called HER2-positive, and are associated with poor prognosis (higher relapse rate, and lower overall survival) (Press et al., 1993). ErbB2 is also overexpressed in some ovarian (Hellström et al., 2001), lung (Ugocsai et al., 2005), endometrial (Buza et al., 2014), or gastric (Abrahao-Machado and Scapulatempo-Neto, 2016) cancers. Mutations of ErbB2 have been observed in many cancer patients, especially in the kinase domain (Connell and Doherty, 2017). With these clinical data, knowledge that ErbB2-containing heterodimers are more potent than any homodimer (Graus-Porta et al., 1997; Lenferink et al., 1998), and that ErbB2 overexpression and/or mutation induces constitutively high ligand-independent activation, ErbB2 is a particularly attractive drug target.

Antibodies targeting ErbB2 Trastuzumab (Herceptin®, Genentech) was the first HER2-targeted drug on the market (approved by the FDA in 1998), aimed at breast cancer patients. This monoclonal humanized antibody binds the extracellular domain of ErbB2 (Carter et al., 1992; Cho et al., 2003), and acts by triggering antibody-dependent cellular cytotoxicity (ADCC), inhibiting dimerization due to steric hindrance, and preventing cleavage of the extracellular region. Trastuzumab was the 5th best-selling drug worldwide in 2018 (Nami et al., 2018). Although it greatly improves survival rates, resistance to trastuzumab is common and can be due to signaling via other receptors, or truncated forms of ErbB2 (Nahta and Esteva, 2007). Recently, an antibody-drug conjugate (ADC) where trastuzumab is coupled to the cytotoxic agent emtansine, T-DM1 (Kadcyla®, Lonza), has been developed, and FDA-approved in 2013. It allows cancer-cell targeted delivery of emtansine, coupled to the activity of trastuzumab.
Another antibody, pertuzumab or 2C4 (Agus et al., 2002) (Perjeta®, Genentech, FDA-approved in 2012) has been developped. Like trastuzumab, it binds to the extracellular domain of HER2 and triggers ADCC,

but prevents heterodimerization more efficiently (Agus et al., 2002). This is consistent with a binding epitope of pertuzumab on domain II (Franklin et al., 2004), contrary to trastuzumab that binds domain IV (Cho et al., 2003). Combined treatments with trastuzumab and chemotherapy is now the first-line treatment (Swain et al., 2015) for HER2-positive metastatic breast cancer.

Tyrosine kinase inhibitors (TKIs) Lapatinib (Tyverb ®, FDA-approved in 2007) is an inhibitor of the kinase domains of EGFR and ErbB2 (Nelson and Dolder, 2006) that binds to the ATP-binding pocket. Different TKIs are under development, that target EGFR, ErbB2 and ErbB3 (Neratinib), ErbB2 only (Tucatinib), or all ErbBs (Poziotinib, Pyrotinib).

A recent review of treatment of HER2-positive breast cancer was done by Escrivá-de Romaní et al. (2018). What comes out of those studies is that combined therapies are the best way to escape resistance. Targeting new domains of ErbB2 would therefore open new possibilities.

2.2.2 ErbB2 conformation is ligand-independent and dimerization prone

No biologically relevant high-affinity ligand for the extracellular domain of ErbB2 has been found, and its ligand-dependent activity has been linked to its heterodimerization with other, ligand-binding, ErbBs (Klapper et al., 1999). It has actually been shown to be their preferred heterodimerization partner, even in the presence of ligands for other ErbBs (Graus-Porta et al., 1997; Tzahar et al., 1996), and ErbB2/ErbB3 heterodimers have been shown to be the most potent ones, with the highest proliferation index (Holbro et al., 2003). ErbB2 has also been shown to be able to induce transformation (cells acquiring the properties of cancer) in a ligand-independent manner when overexpressed (Di Fiore et al., 1987; Brennan et al., 2000), suggesting signaling potential of homodimers.

Extracellular domain conformation These findings implied that the extracellular conformation of ErbB2 had to be different from the ligand-free, tethered conformation of EGFR (Ferguson et al., 2003), and closer to its bound, "open" conformation. This was confirmed with the crystal structure of the free rodent ErbB2, shown in figure 20, and of Herceptin-bound human ErbB2 (Cho et al., 2003). Domain II, which is responsible for dimerization of EGFR ECD, is exposed in ErbB2, as in ligand-bound EGFR, despite large differences

in the orientation of domains I/II relative to domains III/IV. This different orientation is due to hydrophobic interactions of domains I and III of ErbB2 that do not exist in EGFR, but instead are mediated by EGF. The domains II/IV interface of tethered EGFR do not exist in ErbB2. This explains why ErbB2 does not need a ligand to be locked in the "open" conformation. This structure was consistent with a quasi-simultaneously published structure of domains I-II-III of human ErbB2, which superimposes very well (Garrett et al., 2003). Without control of signaling by ligand-binding, ErbB2 lacks an important regulation mechanism of ErbBs. This observation makes it all-the-more interesting and important to understand how signal is transmitted downstream of the extracellular domain.

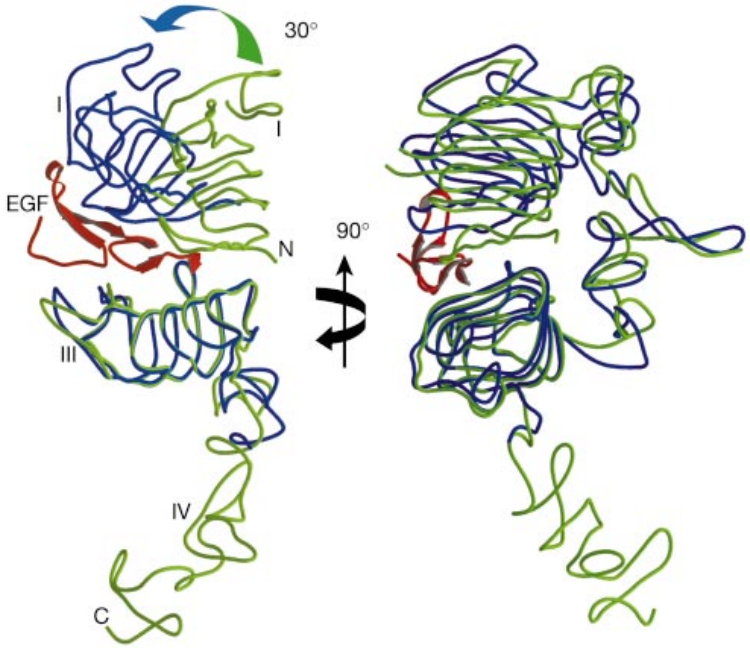


Figure 20 – Comparison of EGF (red)-bound EGFR (blue) domain I-II-III and ErbB2 (green, PDB 1N8Y) extracellular domains. On the left, domain II was omitted for clarity. Figure from Cho et al. (2003).

Homodimerization The existence of ErbB2 substancial ligand-independent activity and in cell microscopy studies revealed that ErbB2 homodimers are more frequently found than homodimers of other ErbBs (Arndt-Jovin et al., 2014). Surprisingly, the ErbB2 ECD did not crystallize as a dimer, it was shown to be monomeric by analytical ultracentrifugation (Horan et al., 1995), and mutagenesis suggested that ErbB2 dimerization did not involve the domain II residues key for EGFR dimerization (Franklin et al., 2004). Whether the ECD is involved in homodimerization of ErbB2 is unclear. However, other parts of ErbB2 are critical for homodimerization: the V659E (transmembrane) "activating" mutation (Bargmann et al., 1986; Akiyama et al.,

1991) increases constitutive homodimerization, tyrosine phosphorylation and transforming potency, while ^{987}VVI^{989}/AAA mutation (at the interface between the kinase domain and the C-terminal tail) abolishes homodimerization and phosphorylation of wild-type and V659E activated mutant (Penuel et al., 2002). The ^{987}VVI^{989} sequence was shown not to be as critical for heterodimerization and ligand-dependent signaling as for homodimerization and ligand-independent signaling (Penuel et al., 2002).

Kinase domain structure and activity ErbB2 kinase has been show to have relatively low kinase activity (a few fold less than EGFR and ErbB4) (Brignola et al., 2002b). Mutation studies linked this low activity to poor ATP-binding and turnover and to glycine residues located between the αC helix and the β4 of the kinase domain (Fan et al., 2008). This was confirmed when the crystal structure of ErbB2 kinase domain with a small kinase inhibitor was solved in 2011 (Aertgeerts et al., 2011) (PDB 3PP0). Although the asymmetric activated dimer is very similar to the one solved for EGFR, increased flexibility due to the previously identified glycines seems to destabilize the active site.

2.3 CtErbB2

2.3.1 Role of each phosphotyrosine (pY) in signal transduction

Five autophosphorylation sites have been identified in CtErbB2: Y1023 (called Y_A), Y1139 (Y_B), Y1196 (called Y_C), Y1221/1222 (Y_{D1} and Y_{D2}) and Y1248 (Y_E) (Hazan et al., 1990; Segatto et al., 1990). Sequence analysis showed that site B has a consensus sequence to bind Grb2 SH2 domain (pYxN) where pY is a phosphotyrosine, while sites C and E have consensus sequences for binding to PTB domains (NPxpY), and site D2 a sequence close to that consensus (NLxpY) (Dankort et al., 2001a).

Major studies of CtErbB2 interactome were conducted by Dankort, Muller and colleagues (Dankort et al., 1997, 2001a,b), who tried to understand the role of each tyrosine in transformation and signal transduction by the Ras pathway. The authors used tyrosine to phenylalanine mutations to abolish phosphorylation at specific sites. These mutants were derived from the V659E mutant (noted "NT") that induces transformation without need of elevated overexpression. Using this mutant enables to easily assess the status of ErbB2 signaling using focus formation assays. The authors made constructs with only one autophosphorylation site mutated to F (called for example "NT-A" when Y_A is mutated to F) or with all but one tyrosine mutated to F (called for example when only site A remains as a tyrosine "NT-YA").

Phosphorylation of Y_A inhibits signal transduction, contrary to all other autophosphorylation sites Main results of focus formations assays of the mutants studied by Dankort et al. (1997) are given in figure 21. They showed that the phosphorylation of tyrosine A alone inhibited focus formation even more than when all tyrosines are mutated to phenylalanines. Conversely, all other autophosphorylation sites (B to D) were individually sufficient for transformation. Besides, mutation of individual tyrosine autophosphorylation sites to F only slightly reduced foci formation, except for single mutation of Y_A to F, that enhanced transformation. Additional phosphorylation of Y_B, Y_C, or Y_E did not reverse Y_A inhibition, while Y_D did, suggesting some connection between those two sites.

Grb2 binds ErbB2 directly via phosphorylated Y_B, and indirectly via Shc on phosphorylated Y_D The authors showed in following studies (Dankort et al., 2001a,b) that signaling by Y_B and Y_D correlated with Grb2 binding to ErbB2 directly (Y_B, consistently with Grb2 sequence specificity) or indirectly (Y_D). Shc PTB domain was shown to bind to Y_D and Grb2 to bind to Shc phosphorylated SH2 domain. These interactions with Grb2 and Shc were shown to couple ErbB2 to the Ras/MAPK pathway, and to be responsible for the transformation induced by phosphorylation at these sites.
An interaction between the site B phosphotyrosine and the protein Grb7 SH2 domain has also been characterized by co-immunoprecipitation (Stein et al., 1994) and by NMR (Ivancic et al., 2003). The role of this interaction *in vivo* is less characterized than the one between Grb2 and ErbB2, but is believed to be involved in breast cancer (Stein et al., 1994).

Several PTB- and SH2-containing proteins bind phosphosites Y_C and Y_E Although Shc PTB domain was mainly shown to bind to site D, it was also shown to bind to site C (Marone et al., 2004). Crk adaptor protein, as well as PLCγ were also shown to bind to this site, although they contain SH2 domains and not PTB domains (Dankort et al., 2001b; Marone et al., 2004). Site E was shown to bind the PTB domain of Dok-R (Dankort et al., 2001a). Although the role of these interactions have not been investigated in details, all those proteins are involved in the MAPK pathway.

The main SH2- and PTB-containing proteins binding to each site of CtErbB2 are recapitulated in figure 22.

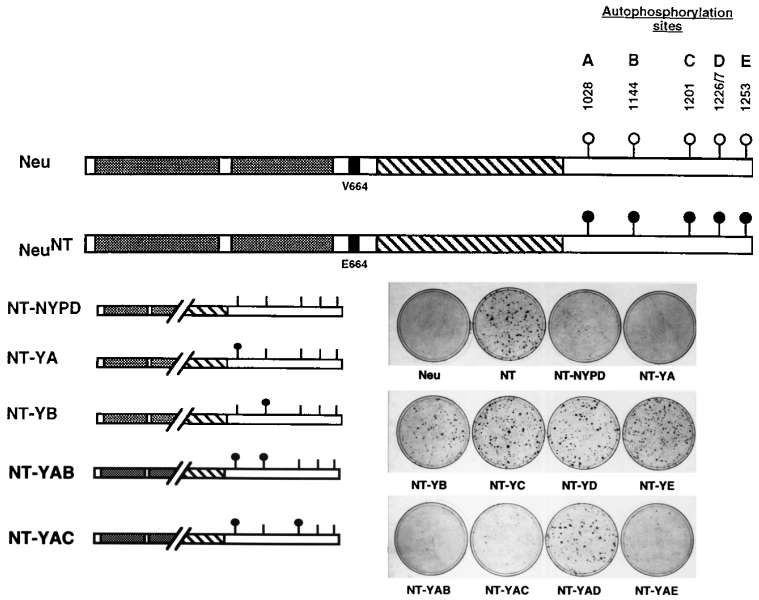


Figure 21 – Role of the different autophosphorylation sites of ErbB2. Transfection with different mutants of activated (NT) ErbB2/neu, of which schematic examples of constructs are given, and observation of foci formation (transformation). Restoration of any of four autophosphorylation sites (Y_B to Y_E) compared to an all-phenylalanine mutant (NYPD) restores significant transformation, while tyrosine A phosphorylation inhibits transformation. Inhibition by Y_A is lost only when Y_D is phosphorylated. Tyrosine phosphorylation in all but the NYPD mutant was verified. From Dankort et al. (1997). Here the numbering is the alternative (rat neu) one (see figure 12b)

Phosphorylated Y_D binds MEMO to control cell motility A strategy similar to that of Dankort et al. (1997) was used in the group of Ali Badache (Marone et al., 2004) to identify proteins interacting with autophosphorylation sites Y_D. They identified a new protein that they called MEMO (mediator of ErbB2-driven cell motility), binding on the second tyrosine of this site (Y1222, also numbered Y1227 or Y_{D2}). This protein is structurally related to bacterial dioxygenases and does not contain any SH2 or PTB domain, so it was first thought to interact indirectly with ErbB2. However, it was shown by pull-down experiments that the interaction was direct (Qiu et al., 2008), and the interaction between MEMO and a Y_D-derived peptide was described by NMR (Feracci et al., 2011). In the docking model, the phosphotyrosine binds in the large cleft of MEMO via three histidines and a cysteine making hydrogen bonds with the phosphate group. The minimal sequence for binding was shown to be ^{1217}FDNLYpYWDQD^{1226}. MEMO has been shown to control cell motility by regulating microtubule dynamics, formation of adhesion sites, and formation of actin-binding complexes (Zaoui et al., 2008).

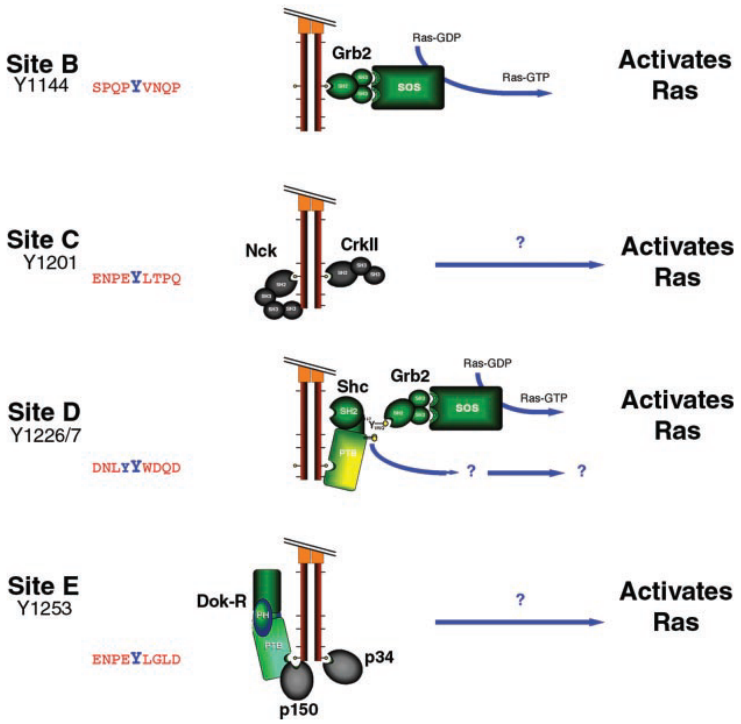


Figure 22 – Main SH2 and PTB domains interacting with CtErbB2 autophosphorylation sites. Figure from Dankort et al. (2001a). Here the numbering is the alternative (rat neu) one (see figure 12b)

2.3.2 Regulation of kinase activity

The effect of the C-terminal tail, and of its phosphorylation, on the activity of the kinase is unclear. Mechanisms similar to EGFR may be at stake but were not really investigated. From ErbB2 kinase domain structures (bound to inhibitors), both published in 2011 (Aertgeerts et al., 2011; Ishikawa et al., 2011), the AP-2 helix observed in EGFR is positioned in a similar fashion in ErbB2, next to the hinge between the kinase N and C lobes (see figure 23). This position could suggest a similar role in kinase activity regulation.

2.3.3 Known structural features of CtErbB2

In the free state CtErbB2 is expected to be highly disordered Except for the helix (Pro999-Leu1009) seen in the crystal structure of the kinase domain, there is no atomic-resolution data on the structure and dynamics of free CtErbB2. Although Keppel et al. (2017) focused on EGFR and ErbB3, they also showed by HDX-MS that CtErbB2 is highly solvent-accessible and probably disordered in the context of the whole intracellular region. In this study, the helix seen in the crystal structure also seems more protected than the

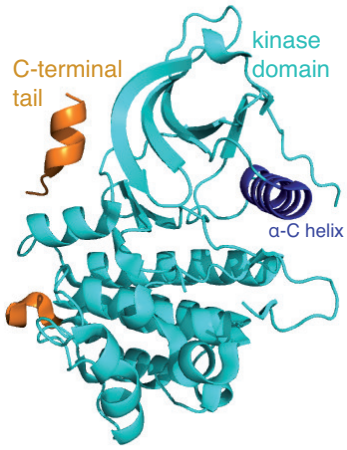


Figure 23 – Structure of the ErbB2 kinase domain, in complex with the SYR127063 inhibitor not shown (PDB 3PP0) (Aertgeerts et al., 2011).

rest of the tail, which suggests that there is more structuration in this region in solution, consistently with the kinase domain crystal structure.

Expected conformation of the phosphotyrosine binding sites in the bound state A phosphopeptide (from the receptor TrkA) bound to Shc PTB domain was shown to adopt a complex structure, completing a β-strand of the protein and forming a β-turn N-terminal to the pY (Zhou et al., 1995). Whether a similar structure is formed when CtErbB2 binds Shc PTB domain is unknown. In the MEMO-Y_D structure, the peptide adopts a helical structure in the region FDNLYpYW (Feracci et al., 2011). Phosphotyrosine peptides usually adopt an extended conformation when binding to SH2 domains (Liu et al., 2012). However, peptides binding to Grb2 SH2 have been shown to adopt a β-turn involving the phosphotyrosine and the residue in position +3 (Rahuel et al., 1996), and a peptide from ErbB2 Y_B site complexed with the Grb7 SH2 domain exhibits such a turn (Ivancic et al., 2003). Different kinds of local structure can thus be expected in different bound forms of CtErbB2.

Aims and strategies of this work CtErbB2 is expected to interact with many different partners, in an intramolecular manner (with its kinase domain for regulation), in the ErbB dimers (for potential *trans* phosphorylation), and with signaling proteins. To understand how these different interactions are regulated, we chose to adopt two complementary approaches:

- Study the structure and dynamics of free CtErbB2. The expected disordered nature of this tail indicates that both the existence of local prestructured motifs (or MoRFs), and the overall structure and dynamics of the region might allow regulation of multiple binding events, as described in the first section of this manuscript for IDPs.

- Study of individual interaction events. In parallel to the first approach, it is crucial to understand at the atomic level how CtErbB2 interacts with its partners. Each interaction will then be replaced in the general context of signaling. In this work, we chose to focus on one particular partner, Grb2, presented below.

3 Grb2, a major adaptor protein in ErbB2-dependent pathways

Grb2 (for Growth factor receptor-bound 2) was discovered in human as the protein ensuring signal transduction between RTKs and the Ras/MAPK pathway (Lowenstein et al., 1992). It is known as Sem-5 (Sex muscle abnormal protein 5, for its role in vulval development) in *C. elegans* and Drk (downstream of receptor kinase) in *Drosphila*. It is a typical example of adaptor protein, a protein that links different parts of signaling pathways, with no catalytic activity. It is entirely made of two protein-binding modules: a central SH2 domain flanked by two SH3 domains (called NSH3 at the N-terminus and CSH3 at the C-terminus).

3.1 Grb2 in ErbB2 signaling

Grb2 in signaling pathways Grb2 was originally shown to bind to phosphorylated RTKs via its SH2 domain, and to the Son of sevenless guanine nucleotide exchange factor (Sos) via its SH3 domains, triggering Ras/MAPK pathway signaling (Lowenstein et al., 1992; Gale et al., 1993; Buday and Downward, 1993). Grb2 has been shown to bind both directly and indirectly to ErbB2: it binds directly to phosphorylated Y_B and indirectly to phosphorylated Y_D via Shc (Dankort et al., 1997, 2001b). Shc binds Y_D via its PTB domain and phosphorylated Shc is recognized by Grb2 SH2 domain. Sos, by interacting with RTK-bound Grb2, localizes at the plasma membrane, getting closer to the Ras GTPase. Sos then exchanges the nucleotide bound to Ras from GDP to GTP.

Although the MAP kinase pathway is the RTK/Grb2-related pathway that is best described (see figure 24), Grb2 also interacts with other proteins determining the fate of receptor tyrosine kinase signaling: via binding to Gab1 (for Grb2-associated-binding protein 1), it links RTKs to the PI3K pathway controlling apoptosis and to Shp2 phosphatase reinforcing Ras-dependent signaling (Gu and Neel, 2003); binding to ubiquitin ligase Cbl controls ubiquitination of EGFR (Ravid et al., 2004); and binding to dynamin1 controls EGFR endocytosis (Yamazaki et al., 2002).

Physiological and pathological roles of Grb2 Grb2 role as an adaptor has been shown to be essential for early steps of embryogenesis, especially differentiation. This role has been directly correlated to the role of Grb2 in Ras pathway, connecting RTKs by its SH2 domain to Sos. Indeed, replacing Sos proline-rich motifs

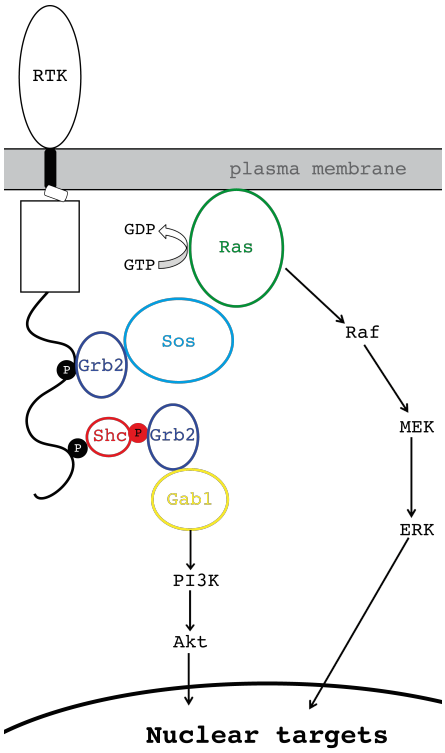


Figure 24 – Simplified scheme of the two main Grb2-dependent signaling pathways: PI3K/Akt and Ras/MAPK.

with Grb2 SH2 domain, thus allowing direct connection between RTKs and Sos, rescued differentiation in the Grb2-null cells (Cheng et al., 1998). The same study showed that Grb2 is also rate-limiting for mammary carcinoma induction in mice. Moreover, it has been shown to be upregulated in some breast cancers (Daly et al., 1994). Grb2 has been shown to be essential for ErbB2-induced transformation (Xie et al., 1995), and the efficacy of targeting Grb2 in HER-2 positive cancer cells has been proven (Gril et al., 2007).

3.2 General description of SH2 and SH3 domains

Src Homology 2 and 3 (SH2 and SH3) domains were discovered in proteins related to the cytoplasmic tyrosine kinase Src. Since then, they have been shown to be highly abundant in signaling pathways, and to often work together. Many proteins contain tandem SH2 and SH3 domains (Fyn, Grb2, Abl, Lck, GAP, PI3K, phospholipase Cγ...).

3.2.1 SH2 domains and phosphotyrosines

SH2 domains were discovered in 1986 in Pawson's lab (Sadowski et al., 1986), and are domains of about 100 residues that specifically recognize phosphorylated tyrosines. It was the first protein interaction domain to be identified. 111 human genes code for at least one SH2 domain, for a total of 121 human SH2 domains described so far (Liu et al., 2011). SH2 domains can be found in enzymes or non-catalytic proteins, and often participate in signaling pathways of receptor tyrosine kinases.

SH2 binding to a phosphopeptide do not trigger large conformational changes. The phosphotyrosine binds in a conserved pocket (see figure 25), with a conserved arginine residue in a FLVR motif. Specificity of SH2 domains for their substrate is given mainly by another pocket, formed by loops on the other face of the central β sheet. That pocket accommodates residues C-terminal compared to the phosphotyrosine (up to position +6). Residues N-terminal to the phosphotyrosine, or further in the sequence, can also participate in the interaction in some cases. The location of the two pockets requires, in most cases, the ligand peptide to be in an extended conformation perpendicular to the central β sheet. The specific, secondary interactions allow for dissociation constants that can be as low as nM (Felder et al., 1993), but are more usually in the μM range. SH2 domains can, in some cases, bind non-phosphorylated peptides but with significantly higher K_D, except for some SH2 domains such as SAP (Poy et al., 1999) and Shc (Charest et al., 1996) that can also quite specifically recognize non-phosphorylated peptides.

A website gathering comprehensive information about SH2 domains is accessible at `www.sh2domain.org`.

3.2.2 SH3 domains and polyproline motifs

SH3 domains are small domains of about 50 to 75 residues. Even though sequence conservation of SH3 domains is not very high (about 27% (Larson et al., 2000)), they have a conserved fold. They are made of two perpendicular antiparallel β sheets, with their termini close to each other and located on the side opposite to the binding site, as seen in figure 26a. SH3 domain-containing proteins are involved in recognition and assembly of other proteins (for example with recognition of dynamin by amphiphysin (Hinshaw, 2000)), membrane localization (of Sos thanks to binding to Grb2 (Buday and Downward, 1993)), and kinase regulation (in Crk (Sriram and Birge, 2012)).

SH3 domains mainly bind peptides containing the PxxP motif in a PPII conformation, with affinities usually in the μM range (Musacchio, 2002). Ligand-binding does not trigger any large conformational change. Two

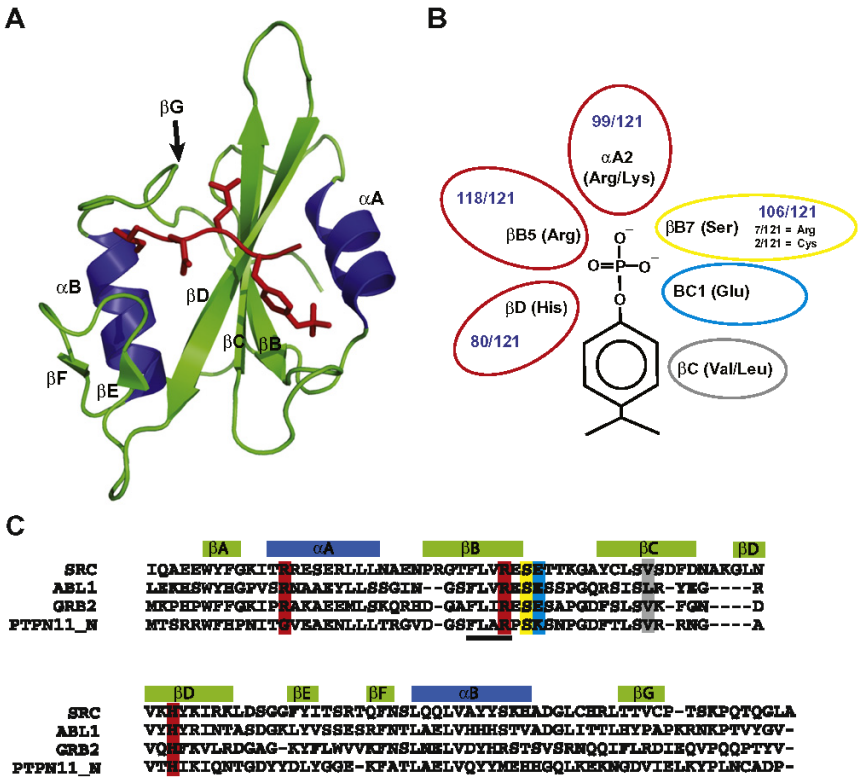


Figure 25 – Canonical structure of SH2 domains (A, here Src, PDB 1SPS) and binding to a phosphotyrosine peptide (red sticks). Here, the phosphotyrosine-binding pocket is on the right of the central β sheet, and the specificity pocket is on the left. Residues important in the phosphotyrosine-binding pocket are given in B, and alignment of different SH2 domains is given in C. From Liu et al. (2012)

main classes of ligands, that bind in opposite orientations, have been identified: class I ligands with consensus sequence RxxPxxP (said to bind in the *plus* orientation) and class II ligands with PxxPxR sequence (said to bind in the *minus* orientation) (Musacchio, 2002). Details of the two binding modes are given in figure 26b. In both cases, two pockets (named S1 and S2) bind one xP motif each, using aromatic residues and a proline. A third pocket (S3), located between the RT (β1-β2) and N-Src (β2-β3) loops, accommodates the positive charge of the arginine (sometimes a lysine) with a negatively charged residue, often a glutamate. The existence of two different binding orientations of polyproline motifs to SH3 domains is allowed by the symmetry of the PPII structure.

Many examples can be found of binding of SH3 domains to non-canonical motifs, with varying positions of the prolines or even absence of prolines in the ligand. Regions quite far from the canonical binding motif

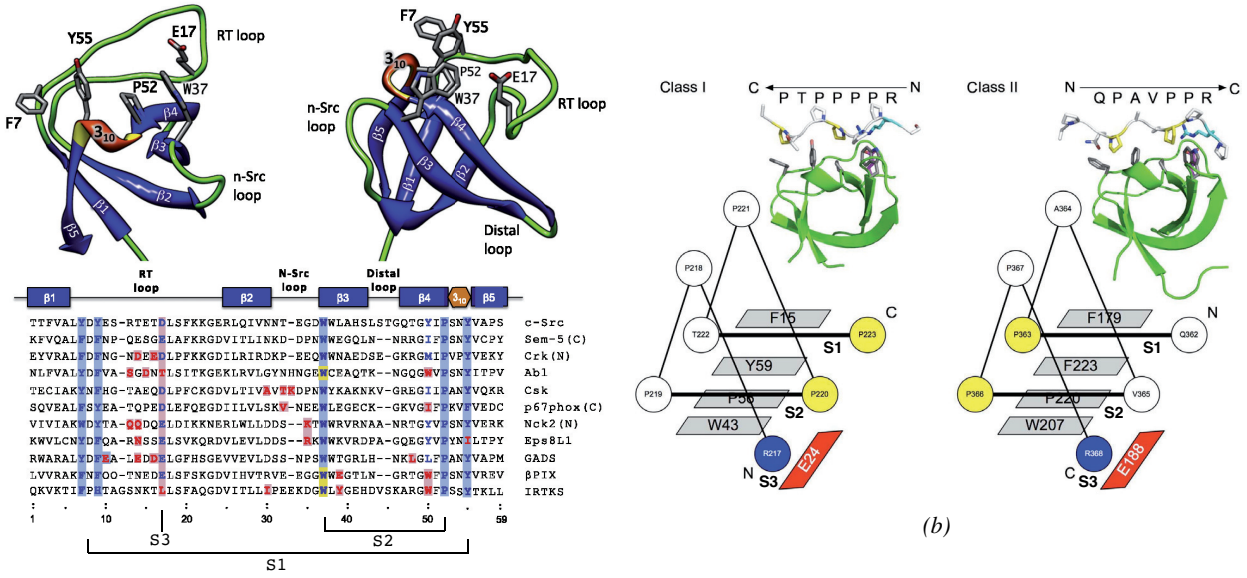


Figure 26 – Structure and binding of SH3 domains to polyproline motifs. (a) Structure of Sem-5 C-terminal SH3 domain and alignment of different SH3 domains. β1, β2 and β5 form one sheet, while β3 and β4 form the second one (which sometimes also contains the end of β2). Sticks of key residues for xP binding (blue background in the alignment) and the acidic pocket (red background) are shown. On the alignment, in blue are conserved residues, and in red are residues that confer binding specificity to each SH3. From Saksela and Permi (2012). (b) Binding of the canonical classes of polyproline ligands to SH3 domain. Left: β-PIX SH3/AIP-4 peptide complex as an exemple of class I ligand binding. Right: p40^{phox} SH3/p47^{phox} peptide complex as an example of class II ligand binding. From Kaneko (2008).

can also play an important role.

3.3 Previous structural descriptions of Grb2

Sequence analysis of Grb2 was sufficient to identify its SH2 and SH3 domains (Lowenstein et al., 1992). The first direct structural descriptions of Grb2 were the NMR solution structures of its NSH3 domain bound to Sos proline-rich peptides (PDB 1GBR)(Goudreau et al., 1994; Wittekind et al., 1994), and of its CSH3 domains (PDB 1GFC) (Kohda et al., 1994). These monodomain structure determinations were shortly followed by the crystal structure of the full-length, apo-Grb2 (PDB 1GRI) (Maignan et al., 1995), and then the isolated SH2 domain (PDB 1GHU) (Thornton et al., 1996). Although the folding of each individual domain is consistent with other domains of the same nature, interesting features appear in each of them and in their organization in the full-length adaptor.

3.3.1 Individual domains

SH3 domains There are only 3 structures of the NSH3 domain of Grb2 in the PDB. As suggested by its first structural study (by solution NMR) by Goudreau et al. (1994), the domain exhibits unfolding in several different conditions, even if its overall folding resembles that of other SH3 domains. It was said to be "degraded" and to dimerize (depending on the presence of DTT, as observed by mass spectrometry) at pH 6.5, and to be stabilized at pH 3.5. At this low pH, however, an equilibrium between a folded and an unfolded form was observed. The C32S mutant of the protein was not "degraded" and did not dimerize, but underwent the same folded/unfolded equilibrium at pH 3.5. Moreover, it had almost the same 3D structure as the wild type protein at pH 3.5. This indicates that cysteine oxidation only has an effect on dimerization, while pH controls folding stability. Later, the group of Mueller did not report stability issues working at pH 6.0 with 10mM DTT (Wittekind et al., 1994, 1997), consistent with this analysis.
Analysis of mutant structures also gives insight into this domain (un)stability. The C32S-P49L mutant was shown to be unfolded, although this proline is not essential in folding of other SH3 domain; The C32S-Y7V and C32S-E40T mutants exhibited slow folded/unfolded exchange with about 1:1 ratio, similar to the observed effect of low pH (Vidal et al., 1999). These particular residues do not seem to have a role more important than others for folding, and the unfolding induced by their mutation highlights the overall unstability of the domain. In other SH3 domains, pH, ionic strength, temperature, or even addition of proline-rich ligands, has been shown to have similar destabilizing effects. The control of folded/unfolded equilibrium depending on conditions may be a mechanism of regulation of interactions in the cell (Vidal et al., 1999).
No ligand-free structure of the NSH3 was deposited in the PDB. One of the two ligand-bound structure of the wild-type domain is presented in figure 27b.

The first structure of the CSH3 was solved by solution-state NMR by Kohda et al. (1994). It also has the canonical fold of an SH3 domain, with an additional small two-strand antiparallel β sheet in the RT loop compared to the canonical structure presented in figure 26. This sheet is found in many SH3 domains (Src, spectrin, PI3K, Fyn, PLCγ (Kohda et al., 1994)) but not in the NSH3. The cysteine that seems responsible for the NSH3 dimerization is deleted in the CSH3, even though a cysteine is also present at another position. For a comparison of NSH3 and CSH3 domains sequence and structure of Grb2, see figure 27.

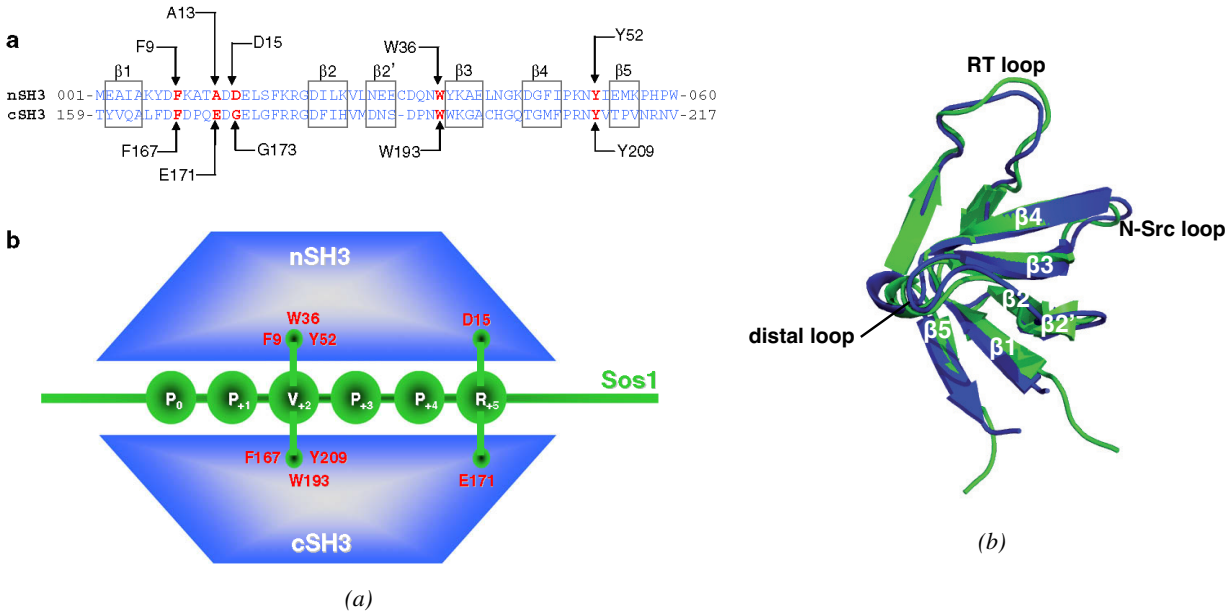


Figure 27 – Comparison of Grb2 NSH3 and CSH3 domains (a) Sequence alignment of Grb2 NSH3 and CSH3 and comparison of peptide binding. From McDonald et al. (2008a) adapted to fit strands numbering of figure 26. (b) Alignment of NSH3 (PDB 1GBQ, ligand bound, blue) and CSH3 (PDB 1GFC, free, green) domain structures of Grb2 in solution.

SH2 The first structure of Grb2 SH2 domain is the crystal structure of full-length, dimeric Grb2, at rather low resolution (3.1 Å) (Maignan et al., 1995) (see figure 29a). The SH2 domain shows canonical folding, but with a position of W121 that would prevent binding of extended peptides as seen in other SH2-phosphotyrosine peptide complexes. The solution structure of Grb2 SH2 domain by Thornton et al. (1996) shortly after confirmed the unusual pocket closed by W121, as shown in figure 28. Mutation of Src at the equivalent position to introduce tryptophan suggested that it is responsible for Grb2 selectivity for an asparagine in position +2 relative to the phosphotyrosine (Marengere et al., 1994).

Small differences between solution structures of the domain in phosphate compared to phosphate-free buffer, consistant with binding of phosphate in the phosphotyrosine-binding pocket, were found, especially in the loop around residues 87-94 (βB-βC) (Senior et al., 1998).

3.3.2 Domain organization in full-length Grb2

Only one structure of full-length Grb2 exists. This crystal structure solved by Maignan et al. (1995) is a dimer, with no ligand bound. Here we focus on the domain organization in each monomer (the two

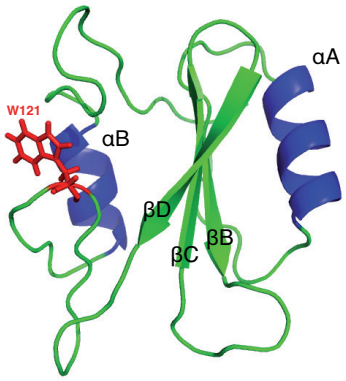


Figure 28 – Solution structure of Grb2 SH2 domain (Thornton et al., 1996) (PDB 1GHU). Compared to a canonical SH2 fold (figure 25), the position of W121 (red) would prevent an extended phosphotyrosine peptide to bind.

monomers are similar in the dimer), dimerization being discussed below in 3.3.3. Solution structures of all three individual domains of Grb2 superimpose well to this full-length structure (see figure 29a). The two SH3 domains are in contact but with a buried interface of moderate surface (1000 $Å^2$) and 5 hydrogen bonds. The authors suggest that this contact may not be present in solution, and that domains may be flexible relative to one another to facilitate binding to different partners. It is to be noted that even in this structure with inter-SH3 interactions, the binding sites for all three domains remain accessible and are not buried.

The hypothesis of flexible linkers between the domains was tested in solution by Yuzawa et al. (2001), but only in the ligand-bound state. Indeed, the authors studied full-length Grb2 with an EGFR peptide (EpYINSQV) and a Sos peptide (VPPPVPPRRR) at ratio Grb2:EGFR:Sos of 1:1.2:5. The results, presented in figure 29b, show that the ^{1}H-^{15}N HSQC spectrum of Grb2 is almost exactly the sum of the HSQC spectra of individual domains, and that each domain has independent dynamics in the full-length protein. Therefore, the three domains are expected to have independent behaviors when bound to different, non-connected peptides. However, it is to be noted that slightly higher chemical shifts differences between full-length Grb2 and the free NSH3 and SH2 are observed than for the CSH3, even though the values are still quite small. The CSH3 is also more freely rotating than the first two domains, and especially than the NSH3 even though it is very similar in size and structure. Weak interaction between the two first domains could be at stake. Whether these observations hold true for the ligand-free protein remains to be investigated.

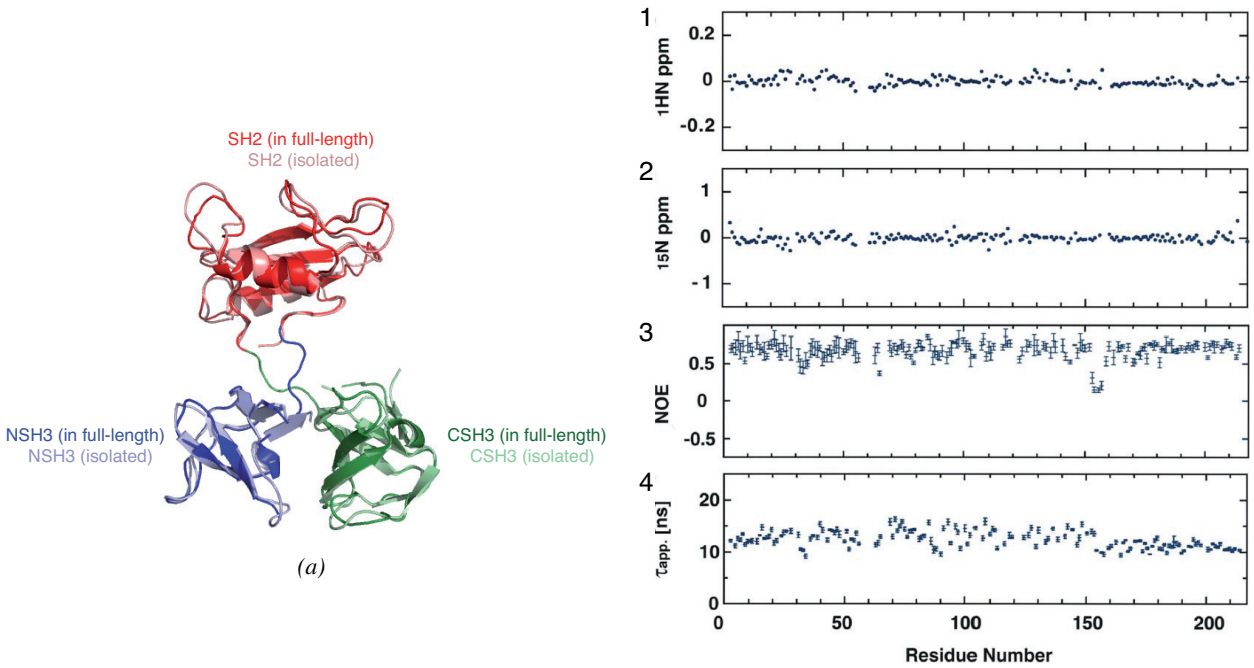


Figure 29 – Comparison of full-length Grb2 and individual domains. (a) Solution structures of individual domains of Grb2 (NSH3: light blue, PDB 1GBQ; SH2: pink, PDB 1GHU; CSH3: light green, PDB 1GFC) compared to the crystal structure of full-length Grb2 (NSH3: dark blue, SH2: red, CSH3: dark green, PDB 1GRI). (b) Comparison of proton (1) and nitrogen (2) chemical shifts of full-length Grb2 and individual domains, and ^{15}N relaxation of full-length Grb2 (3: heteronuclear nOe and 4: rotational correlation time calculated from T1 and T2 values). From Yuzawa et al. (2001).

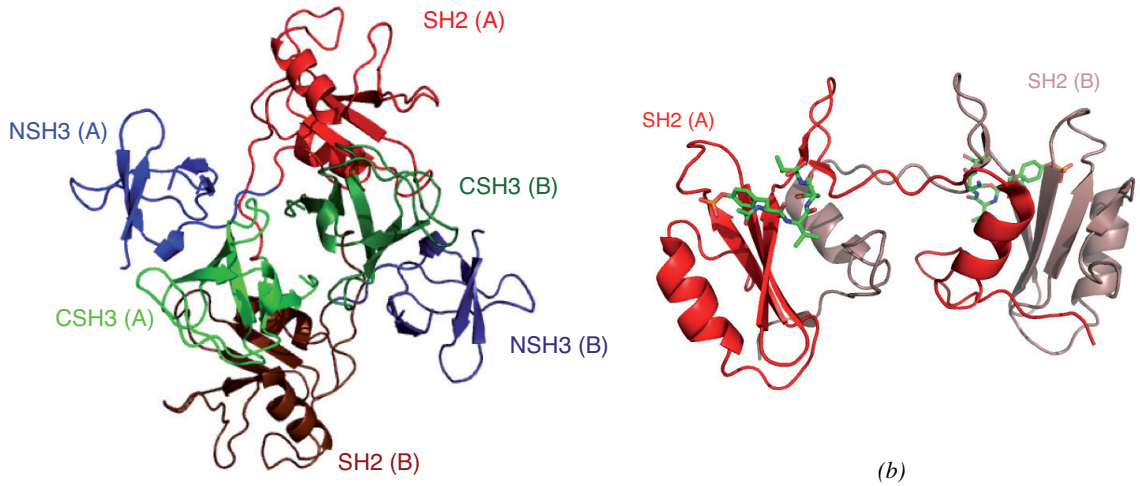
3.3.3 Oligomerization

Full-length Grb2 dimer In the literature, data about Grb2 oligomeric state seem contradictory. The crystal structure of the full-length is a dimer, with SH2 domain of one monomer contacting quite extensively the CSH3 of the other (as shown in figure 30a). This dimer interface is still leaving canonical SH2 and SH3 binding interfaces exposed for interaction.

Moreover, dimerization or higher-order oligomers have been observed during preparation of NSH3 (Goudreau et al., 1994) or SH2 (Thornton et al., 1996) samples. However, when trying to characterize those dimers or measure dimerization constants of Grb2 or its domains, no clear picture emerges. Different studies of Grb2 oligomeric state are recapitulated in figure 30c. Results diverge on the formation of non-covalent dimers, from almost all-monomeric Grb2 to low-μM dimer. The problem also resides in the absence of study of full-length Grb2 without any purification tag and without ligand. However, it is to be noted for all constructs that cysteins present in the SH3 domains (one in each) have to be reduced in order to prevent formation of

covalent dimers when working with constructs containing one or both of them.

Concerning the structure of the individual domains in the dimer, all studies suggest it is unchanged (as seen in the comparison of the individual domain structures to the crystallographic dimer of the full-length protein in figure 29a), except in the hypothesis of a domain-swapped SH2 domain, presented below. However, arrangement of domains relative to one another might be modified by dimerization (McDonald et al., 2008b).



(a) *(b)*

Ref	Construct	Ligand peptide	Dimeric state	Dimer interface	Affinity	Method
Guilloteau et al. (1996)	FL	None	unstable monomer	ND	ND	SEC, AUC, light scattering
Maignan et al. (1995)	FL	None	dimer	SH2-CSH3	ND	X-ray (PDB 1GRI)
Yuzawa et al. (2001)	FL	EGFR and Sos	monomer	None	>mM	NMR and SAXS
McDonald et al. (2008b)	Trx-FL-His	None	equilibrium	Assumed NSH3/CSH3	5 μM 20-30 μM	AUC ITC
Lin et al. (2012)	His-FL	None	equilibrium	ND	0.7 μM	MST
Ahmed et al. (2015)	His-FL	None	equilibrium	ND	0.7 μM	MST

(c)

Figure 30 – Grb2 dimerization. (a) Grb2 crystal dimer (PDB 1GRI). The two monomers are labeled A and B. (b) SH2 domain-swapped dimer structure (PDB 1FYR). (c) Literature data on Grb2 oligomeric state. SEC = Size Exclusion Chromatography; AUC = Analytical Ultracentrifugation; MST = Microscale Thermophoresis; ITC = Isothermal Titration Calorimetry. FL = full-length Grb2. Trx = Thioredoxin.

Domain-swapped SH2 dimer All studies until 2000 showed the SH2 domain of Grb2 to have canonical fold and to be monomeric. However, Schiering et al. (2000) crystallized a dimeric conformation of the SH2 in which the EF loop is extended, and the remaining C-terminal segment is swapped with another monomer (shown in figure 30b). Such domain-swapping had never been observed for a SH2 domain, and this structure is still unique today. The domain-swapped dimer reconstitutes two hybrid domains with overall canonical

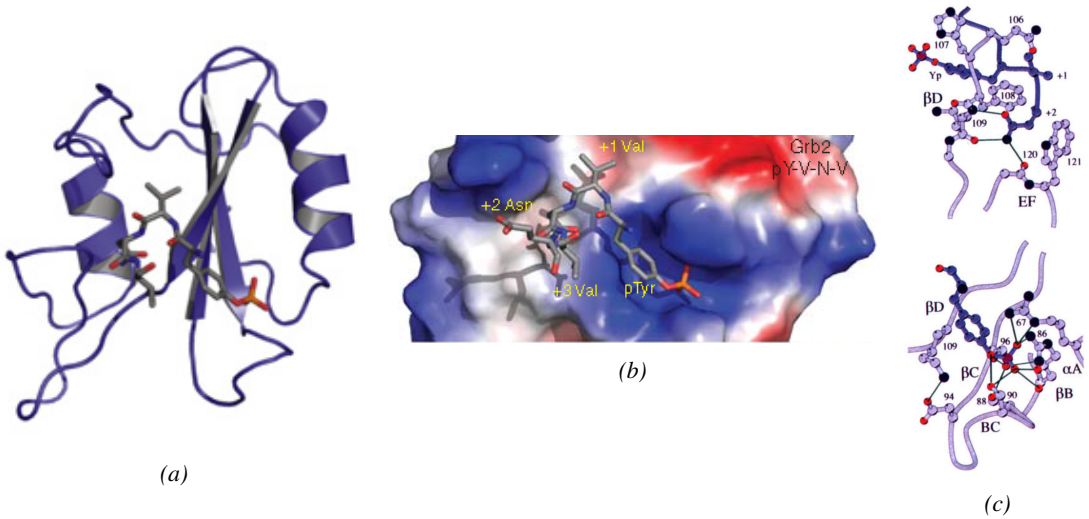
SH2 fold. The dimeric form could be separated from monomer by SEC (with a ratio 8:2), without return to an equilibrium of any of the two peaks in standard conditions. The dimeric form could be returned to monomer by heating, chemical denaturation and refolding, low pH, or addition of organic solvent, showing that the domain-swapped dimer is a metastable form. This domain-swapped dimer of the isolated SH2 domain has been observed by several groups since this study (Nioche et al., 2002; Benfield et al., 2007; Papaioannou et al., 2016). Domain-swapped dimer conformation has been shown to be independent on ligand binding (Nioche et al., 2002; Benfield et al., 2007; Papaioannou et al., 2016).
Presence of such swapping in full-length Grb2, and its biological relevance, have not been proven, even though Schiering et al. (2000) proposed a model of full-length, SH2-swapped Grb2 dimer.

3.4 Known interaction modes of Grb2

3.4.1 SH2-phosphotyrosine (pY) interaction

Grb2 is known to have specificity for pY-x-N-x substrates, with for example a K_D of about 0.3 μM for the SpYVNVQ peptide from Shc compared to 167.8 μM for the PQpYEEIPI peptide (typical substrate for Src SH2 domain), as measured by ITC (Papaioannou et al., 2016). This specificity has been shown to be associated with the presence of a tryptophan in the EF loop (W121) (Marengere et al., 1994), that would block the usually extended structure of the substrate peptide C-terminus of the phosphotyrosine (Maignan et al., 1995; Thornton et al., 1996). The peptide-bound structure of Grb2 SH2 (figure 31) showed that indeed the bound phosphotyrosine peptide (KPFpYVNV from BCR-Abl) adopts an unusual conformation with a β turn formed right after (C-terminus) the phosphotyrosine (Rahuel et al., 1996).
The role of W121 is mitigated in a more recent study (Papaioannou et al., 2016) that shows that Grb2 SH2 W121G substrates also bind in a β turn conformation, only slightly distorted. The mutation only lowered Grb2 specificity for N in position +2 but did not reverse it (the ratio of affinity for SpYVNVQ compared to PQpYEEIPI peptides as measured by ITC is still 10). The reverse T to W mutation in Src increased affinity for N+2 less drastically (42 fold increased affinity ratio) than previously measured by SPR (1200 fold increased affinity ratio) (Marengere et al., 1994) and led to comparable K_Ds for the two types of peptides. W121 is therefore not the only determinant for Grb2 SH2 specificity and the peculiar complex structure.
Peptides for which the structure of the complex with Grb2 SH2 have been solved include peptides derived from EGFR (EpYINQ, PDB 1ZFP), BCR-Abl (FpYVNV, PDB 1BMB), HGFR (TpYVNV, PDB 1FYR),

Shc (SpYVNV, PDB 1QG1) and CD28 (DpYMNM, PDB 3WA4). There is no structure of a complex between a peptide from ErbB2 (for which the known interaction site is EpYVNQ (tyrosine B) (Dankort et al., 1997)) and Grb2 SH2. However, there is a crystal structure of the Grb7 SH2 domain in complex with this ErbB2 peptide (PDB 1MW4), showing a similar β-turn conformation, even though this conformation is not due to a tryptophan in the EF loop but to a extension of this loop in Grb7.



(a) *(b)* *(c)*

Figure 31 – Binding of Grb2 SH2 domain to a pY-V-N-V peptide. (a) Cartoon representation (PDB 1BMB). Figure from Liu et al. (2012). (b) Surface representation (PDB 1BMB). Figure from Wagner et al. (2013) and (c) Sticks representation (PDB 1TZE) of the "specificity site" with W121 forcing the β turn conformation (top) and of the canonical phosphotyrosine binding site (bottom). Figure from Rahuel et al. (1996).

The effect of SH2 domain swapping on ligand binding is controversial, as affinities were the same between the monomeric and dimeric forms as measured by Schiering et al. (2000), and K_D increased 4 to 13-fold in the dimer as measured by Benfield et al. (2007). The latter observation is compatible with conformational change of the 121-123 region to an extended form in the swapped dimer, inducing reorientation of the W121 side chain and preventing stabilisation of the N+2 residue by W121 indole ring.

3.4.2 SH3 domains interactions

Musacchio (2002) made a review of peptides binding to both NSH3 and CSH3 domains. The result is reproduced in figure 32a. Even though several other peptides have been tested since, it gives an idea about the interaction motifs for these SH3 domains. They both bind preferentially to class II peptides (expectedly

in the *minus* orientation), as was observed by Kaneko (2008), with estimated preferences for class II vs class I of about 70% vs 30% for both domains.

Despite this common preference, the NSH3 and CSH3 domains do not have the same specificity, and do not bind the same proteins *in vivo*. It was shown that both NSH3 and CSH3 bind to Sos, although there is a preference of all Sos polyproline motifs for the NSH3 (Kohda et al., 1994; Wittekind et al., 1994; Vidal et al., 1999; McDonald et al., 2009). Conversely, Gab1 and Gab2 show specificity for the CSH3 domain, thanks to binding to non-PxxP motifs, RxxK motifs, which was not observed for the NSH3 (Lewitzky et al., 2001; Harkiolaki et al., 2009). In the structures of the Grb2-Gab2 peptides complexes (PDB 2VWF, 2W0Z), the peptides do not always adopt PPII structure, and make different contacts with Grb2 (residues Q170, E171 and E174). Additional hydrophobic contacts with prolines adjacent to the RxxK motif resemble the canonical binding of PxxP motifs. K_D is in the μM range, a common range of affinity for SH3 domains. For both SH3 domains, the presence of multiple PxxP binding sites on their target (Sos and Gab1 for the NSH3 and CSH3 respectively) was suggested to lead to multivalent binding increasing overall affinity, involving Grb2 dimerization (McDonald et al., 2012a,b). Moreover, non-PxxP regions may provide multiple secondary interactions also increasing affinity and specificity, since a Sos construct with scrambled sequence outside of the 4 PxxP motifs failed to tightly bind Grb2 (Bartelt et al., 2015). While it was suggested that different specificities of the two SH3 domains of Grb2 may lead to the formation of ternary Sos-Grb2-Gab1 complexes (McDonald et al., 2010), stoichiometry and steric hindrance prevent formation of such complex in the full-length Grb2 protein (McDonald et al., 2013).

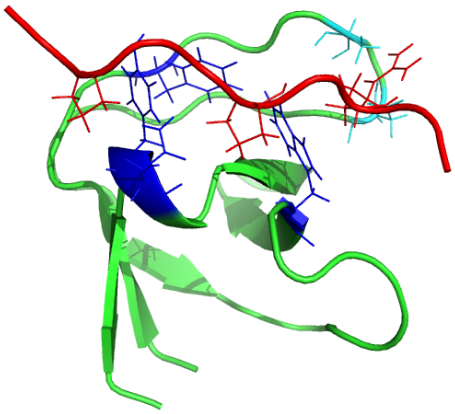
Despite common model of Grb2 interacting with receptors via a SH2-pY interaction, Lin et al. (2012) showed that Grb2 CSH3 was able to bind to a polyproline motif of the FGFR2 unphosphorylated C-terminal tail. This binding is believed to promote FGFR2 dimerization via Grb2 dimerization, promoting basal phosphorylation. In this model, Grb2 prevents extensive signaling by sterically blocking extensive phosphorylation and interaction with other proteins. Upon ligand binding and phosphorylation of Grb2 by FGFR2, the interaction is thought to be released for signaling via pY residues. Such possible mechanism was never investigated for ErbB receptors, but resemblance with FGFR receptors suggests it could exist.

SH3	Ligand	Consensus	Class	K_D (μM)
Grb2 N	Phage library consensus	+ZDX**PLPXLP**	I	
	Hs c-Cbl	PQ**RRPLPCTP**GD	I	
	Hs c-Cbl	WL**PRPIPKVP**VS	I	
	Hs Dynamin	GG**APPVPSRP**G	II	
	Hs Dynamin	GP**PPQVPSRP**N	II	
	Hs Dynamin	RA**PPGVPSRS**G	II	
	Hs Abl	LQ**APELPTKT**R	II	
	Hs Abl	AV**SPLLPRKE**R	II	
	Hs Abl	K**TAPTPPKRS**S	II	
	Hs c-Cbl	AS**LPPVPPRL**D	II	
	Hs SOS1 1)	P**VPPPVPPRRR**	II	
	Hs SOS1 (1)	P**VPPPVPPRRR**	II	3.5
	Hs SOS1 (1)	P**VPPPVPPRRR**	II	5.7
	Hs SOS1 (2)	DS**PPAIPPRQP**	II	
	Hs SOS1 (3)	ES**PPLLPPREP**	II	
	Hs SOS1 (4)	IA**GPPVPPRQS**	II	
	C3G	**PPPALPPKKR**	II	142.0
Grb2 C	Hs SOS1 (1)	P**VPPPVPPRRR**	II	39.0
	N-WASP	**PPPPPXR**	II	

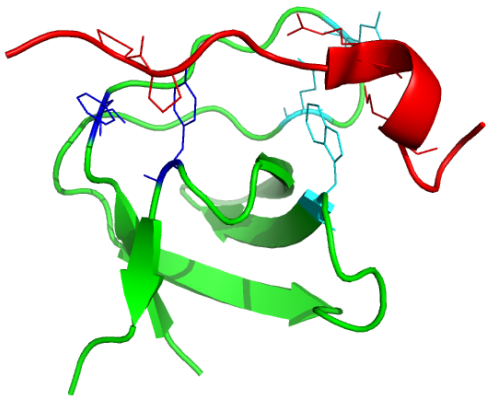
(a)

Domain	mutation(s)	PDB	ligand origin	ligand sequence	ligand class	ligand orientation
NSH3	None	1GBR	Sos	SPLLPKLPPKTYKRE	II	minus
		1GBQ		VPPPVPPRRR		
	C32S Y7V	1AZE				
CSH3	None	1IO6	designed	RHYRPLPPLP	I	plus
		2VWF	Gab2	IQPPPVNRNLKPDRK	non-canonical	minus
		2W0Z		APPPRPPKP		

(b)



(c)



(d)

Figure 32 – Ligand binding to Grb2 SH3 domains. (a) Characterized Grb2 SH3 ligands as of 2002 (Musacchio, 2002). (b) PDB structures of bound Grb2 SH3 domains. (c) NSH3 (green) bound to a Sos peptide (red), from PDB 1GBQ. Residues interacting with the PxxP motifs are in blue, and residues interacting with the arginine are in cyan. Sticks are shown for the prolines and arginine of the PxxPxR motif of the peptide. (d) CSH3 (green) bound to a RxxK peptide, from PDB 2VWF. Residues interacting with the R_0xxK_3 motif are in cyan, and residues interacting with prolines outside of the motif are in blue. Sticks are shown for R_0 and K_3, as well as for P_{-4} and P_{-5}.

3.4.3 Interdependence of binding of SH2 and SH3 domains

In a modular protein such as Grb2, one could wonder how binding of the SH2 and binding of the SH3 domains interfere.

It was shown by fluorescence and ITC that when the SH2 domain interacts with a phosphotyrosine peptide, the affinity of the SH3 domains for polyproline peptides is not changed, showing the absence of allosteric mechanism from the SH2 to the SH3 domains (Cussac et al., 1994; Lemmon et al., 1994). However, when using full-length proteins, interaction of Shc with the SH2 domain enhances binding of Sos to Grb2 SH3 domains *in vivo* (Ravichandran et al., 1995). Whether this is due to indirect effect via other proteins or to direct effects of complex conformation is not known.

When SH2-binding motifs and SH3-binding motifs coexist in the same protein, the problem is even more complex. Several examples exist where there seemed to be cooperativity of SH2 and SH3 domains to bind to a single partner *in vivo*. But once again, the direct or indirect nature of the effect is unknown. Even though RPTPα binding to Grb2 is known to occur via SH2-pY interaction, the CSH3 domain is required for the interaction *in vivo* (den Hertog and Hunter, 1996; Su et al., 1996). Gill et al. (2017) showed that both the SH2 and SH3 domains are required for Grb2 binding to EGFR single-tyrosine mutants. A similar phenomenon was observed for ErbB2 binding, for which binding to Grb2 is drastically reduced upon deletion of the NSH3 (Xie et al., 1995). Molecular details remain to be investigated.

Even though Grb2 is essential for ErbB2 signaling and is involved in ErbB2-related pathologies, a lot is still unknown about its behavior in solution, for it has mostly been studied domain-by-domain. The same observation can be made about Grb2 interactions: despite relatively abundant data on Grb2 individual domains interactions with peptides, very little is known about interactions in the context of the full-length protein. Thus, a lot of information is missing on how interaction of the different domains with different proteins allow signal transduction. Since we suspect Grb2 to interact with ErbB2 through both its SH2 and SH3 domains, this aspect is crucial.

EXPERIMENTAL METHODS

4 Protein expression and purification

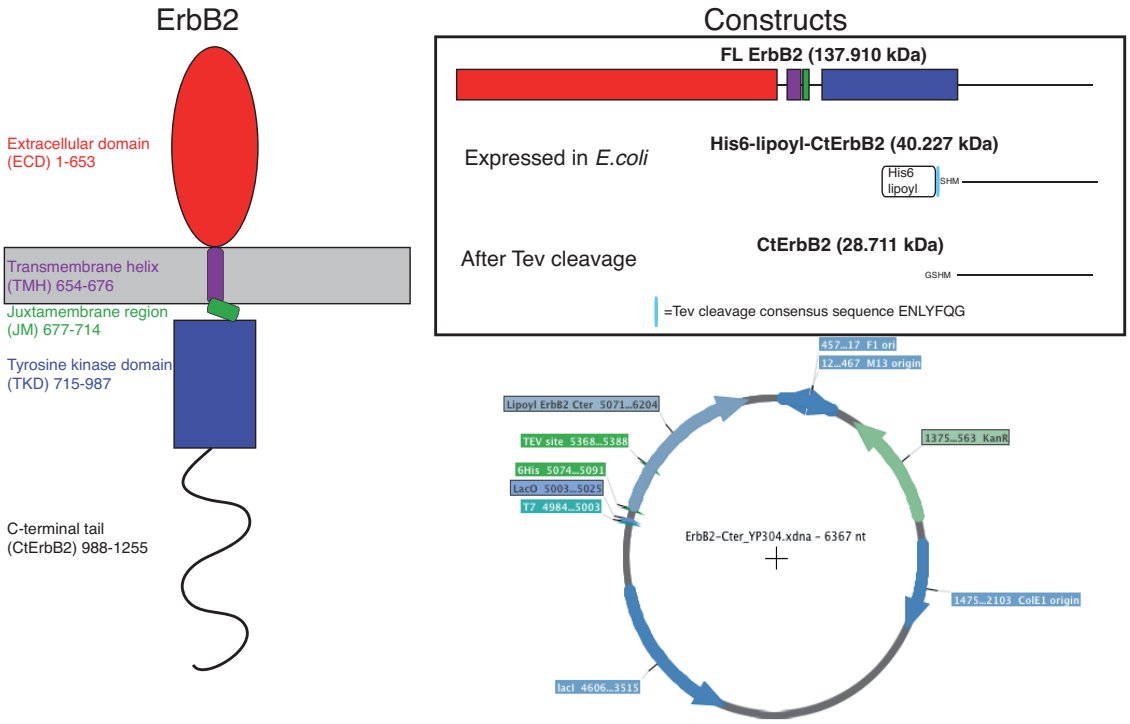
4.1 Grb2 and CtErbB2 expressed in *E. coli*

4.1.1 Constructs and mutants

CtErbB2 Constructs For the study of CtErbB2, the C-terminal tail was defined according to the literature as the portion of the protein following the kinase domain, and encompasses residues 988-1255, that we number 1 to 268 in this study. For purification purposes, a N-terminal His6 tag was added, as well as a lipoyl domain tag that increases solubility. Lipoyl is a 12 kDa domain from *B. stearothermophilus* E2p that is acidic (pI = 4.53), like CtErbB2, and prevents aggregation by electrostatic repulsion (Lebendiker and Danieli, 2014). The His6-lipoyl tag is followed by a TEV cleavage site, leaving 4 additional N-terminal residues (GSHM) when the tag is cleaved. The expressed construct and the final sequence after TEV cleavage are presented in figure 33. The gene coding for CtErbB2 was inserted in a YP304 vector also presented in figure 33, which also contains a gene conferring resistance to the antibiotic kanamycin.

CtErbB2 Mutants Different mutants of CtErbB2 were designed by site-directed mutagenesis:

- Q161A: for help in assignment.
- Four single-cysteine mutants for paramagnetic relaxation enhancement (PRE) experiments: C146S, C45S, (C45S, C146S, S227C) and (C45S, C146S, S248C). Each mutant was designed to have a single cysteine for coupling of the paramagnetic probe. How these mutants were used to site-specifically introduce a paramagnetic probe in CtErbB2 is described in section 5.6.
- Tyr-Phe mutants for selective phosphorylation: mutants "Y_A" (Y18F, Y125F, Y140F, Y152F, Y209F, Y234F, Y235F, Y261F), "Y_B" (Y18F, Y36F, Y125F, Y140F, Y209F, Y234F, Y235F, Y261F), "Y_C"



(a)

```
GSHMVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEPSEEEAPRSPLA
   1        10        20        30        40        50        60        70        80
PSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPAGAT
  90       100       110       120       130       140       150       160       170
LERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV
 180       190       200       210       220       230       240       250       260
```

(b)

Figure 33 – Construct of CtErbB2 expressed in E. coli and final CtErbB2 construct after His6-lipoyl tag cleavage by the TEV protease, compared to the full-length protein. (a) Constructs and vector. The molecular weights are given for non-isotopically labeled proteins. (b) Sequence of the CtErbB2 protein produced, after TEV cleavage, with the numbering used in this work.

(Y18F, Y36F, Y125F, Y140F, Y152F, Y234F, Y235F, Y261F), "Y_{D1}" (Y18F, Y36F, Y125F, Y140F, Y152F, Y209F, Y235F, Y261F), "Y_{D2}" (Y18F, Y36F, Y125F, Y140F, Y152F, Y209F, Y234F, Y261F) and "Y_E" (Y18F, Y36F, Y125F, Y140F, Y152F, Y209F, Y234F, Y235F). The reasons for designing such single-tyrosine mutants are given in section 10. The "0Y" mutant (Y18F, Y36F, Y125F, Y140F, Y152F, Y209F, Y234F, Y235F, Y261F), mutating all tyrosines to phenylalanines was also designed as a control.

All mutants had already been cloned in the lab before my thesis, except for the (C45S, C146S, S227C) mutant. It was made from the template (C45S, C146S) mutant with the following primers:

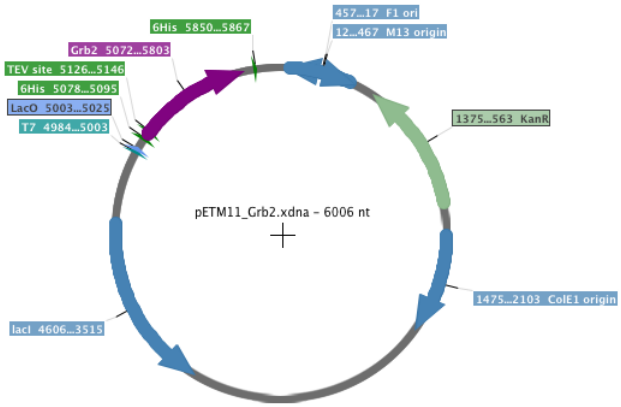
forward primer 5'-cgaaggctgggcagaaggcaggagg-3'

reverse primer: 5'-cctcctgccttctgcccagccttcg-3'

For cloning, the recommendations from the QuickChange II XL kit (Agilent Technologies) were followed. Subsequent amplification and purification of the plasmid were done following the recommendations from the NucleoBond ® Xtra Midi kit (Macherey-Nagel), and sequencing was performed by Eurofins Genomics.

Grb2 The domain organization of Grb2 is schematically presented in figure 34a. Each combination of domains of Grb2 was cloned into a pETM11 vector also containing the gene conferring resistance to kanamycin (shown in figure 34b for full-length Grb2): the three individual domains (NSH3, SH2 and CSH3), the two bidomains (NSH3SH2 and SH2CSH3) and the full-length protein (FL). Each construct contains a N-terminal His6 tag for purification purposes, followed by a TEV cleavage site. After cleavage, two additional N-terminal residues (GA) remain compared to the native sequence, as shown in figure 34c. The definition of all the constructs after TEV cleavage is given in figure 34d.

(a)



(b)

```
GAMEAIAKYDFKATADDELSFKRGDILKVLNEECDQNWYKAELNGKDGFIPKNYIEMKPHPWFFGKIPRAKAEEMLSKQRH
  1        10        20        30        40        50        60        70
DGAFLIRESESAPGDFSLSVKFGNDVQHFKVLRDGAGKYFLWVVKFNSLNELVDYHRSTSVSRNQQIFLRDIEQVPQQPT
80        90        100       110       120       130       140       150
YVQALFDFDPQEDGELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTPVNRNV
160       170       180       190       200       210
```

(c)

Construct	Numbering in the full-length protein (tag-cleaved protein)	MW (tag-cleaved, unlabeled protein)
Grb2 FL	(GA)1-217	25.334 kDa
NSH3	(GA)1-56	6.677 kDa
SH2	(GA)53-152	11.798 kDa
CSH3	(GA)153-217	7.653 kDa
NSH3-SH2	(GA)1-152	17.827 kDa
SH2-CSH3	(GA)53-217	19.306 kDa

(d)

Figure 34 – Grb2 constructs. (a) Domain structure of Grb2 (as defined in Uniprot). (b) pETM11 vector used for the expression of Grb2. (c) Sequence of the FL Grb2 construct produced after TEV cleavage, with the numbering used in this work. (d) Constructs of Grb2 studied in this work.

4.1.2 Expression in *E. coli*

Selective pressure was maintained on plates for transformation and in precultures and cultures using 30 $\mu g/mL$ of kanamycin.

Transformation The CtErbB2 or Grb2 plasmid was transformed into BL21 Star (DE3) *E. coli* cells with the following protocol: 50 μL of chemically competent cells were thawed on ice for 20-30 minutes. 1 μL of DNA (50 to 500 $ng/\mu L$) was added to the cells which were then gently swirled and kept on ice for another 20 to 30 minutes. The heat shock was performed in a water bath for 45 s at 42 °C. After cooling on ice for 2 min, 250 μL of SOC medium was added, and cells were incubated for 1 h at 37 °C with shaking (200 rpm). 50 μL of this culture was plated on LB/agar plates and incubated overnight at 37 °C.

Culture in unlabeled medium The next day, one colony from the plate was added to 50 mL of 2xYT medium and incubated overnight (37 °C, 200 rpm) as a preculture. This preculture was then transferred to 1L of 2xYT and grown until OD_{600nm} (optical density measured at 600 nm) reached 0.6 to 0.8. Protein expression was induced following the conditions given in table 1. Cells were then harvested by centrifugation at 6000 g for 20 min at 4 °C, and the pellet was flash frozen in liquid nitrogen for storage at -80 °C until protein purification.

Protein	Induction OD_{600nm}	Inducer	$T_{induction}$	$time_{induction}$
CtErbB2	0.6-0.8	2 mM IPTG	37°C	4-5h
Grb2	0.6-0.8	1 mM IPTG	25°C	overnight

Table 1 – Induction conditions for CtErbB2 and Grb2 in E. coli

Culture in ^{15}N or ^{15}N-^{13}C labeled medium The culture protocol in labeled medium was similar to culture in 2xYT, except that for isotopic labeling the 2xYT preculture was centrifuged at 1200 g for 20 min before resuspension in the M9 culture medium. M9 medium composition is given in table S1. Compared to the standard M9 medium composition for isotopic labeling, we added vitamins and trace elements to boost bacterial growth.

Culture for Grb2 80% deuteration The full-length construct of Grb2, and to a lesser extent the bidomain constructs, are quite large for folded proteins to be studied by NMR, due to slow tumbling and therefore high transverse relaxation rates. For the signal lifetime to be long enough to have good signal-to-noise ratio,

and especially to be able to record 3D spectra, a technique is to deuterate the protein. Thanks to the lower gyromagnetic ratio of deuterium compared to proton, its incorporation reduces the efficiency of the relaxation processes due to the strong dipolar couplings of protons. Since the purification and the experiments are performed in protonated buffers, the backbone deuteron amides have time to back-exchange to protons, allowing recording of regular ^{1}H-^{15}N-based NMR experiments. Besides, deuteration creates almost-isolated ^{1}H-^{15}N pairs, creating optimal conditions to record TROSY-based experiments at high magnetic fields (see section 5.1.3).

All full-length Grb2 samples were 80% deuterated. While the SH2CSH3 bidomain could be studied without deuteration, deuteration was necessary for the NSH3SH2 bidomain, especially for 3D spectra.

For production of deuterated Grb2, 8 mL of 2xYT preculture was transferred to 500 mL 2xYT and the culture was grown at 37 °C at 200 rpm until OD_{600nm} = 0.6 to 0.8. The culture was then centrifuged at 3500 g at 4 °C for 10 min, and the pellet was resuspended in 1 L unlabeled M9 medium and grown at 37 °C, 200 rpm until OD_{600nm} = 0.6 to 0.8 again. This step allowed for adaptation of bacteria to the minimal medium. The culture was centrifuged again at 3500 g at 4 °C for 10 min, the pellet was resuspended in 100 mL ^{15}N^{13}C deuterated M9 medium, and incubated for 1 h at 37 °C. 900 mL of ^{15}N^{13}C deuterated M9 was added and the culture was incubated for around 30 min before overnight induction as in the unlabeled protocol.

4.1.3 Purification

The composition of all buffers used is given in table 2. For both CtErbB2 and Grb2, that contain free cysteins, a reducing agent (TCEP) was included at all steps to prevent intra- and inter-molecular disulfide bond formation.

Buffer name	Used in step(s)	Composition	pH
Buffer A	HisTrap column and Ni-NTA resin	50 mM Tris-HCl, 300 mM NaCl, 2 mM TCEP	8
Lysis buffer	Lysis	buffer A + cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)	8
Buffer B	HisTrap column	buffer A + 500 mM imidazole	8
MES buffer (CtErbB2)	SEC/final buffer	40 mM MES, 200 mM NaCl, 2 mM TCEP	5.6
PIPES buffer (Grb2)	SEC/final buffer	40 mM PIPES, 150 mM NaCl, 2 mM TCEP	7.2
PBS buffer (Grb2)	SEC/final buffer	PBS, 2 mM TCEP	7.2

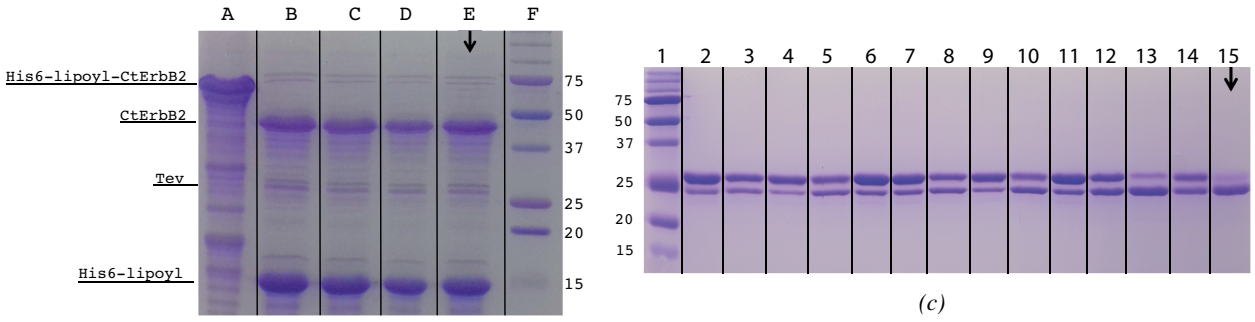
Table 2 – Buffers used for CtErbB2 and Grb2 purification.

For purification, the pellet was thawed on ice and resuspended in 50 mL of lysis buffer/L of culture. Lysis was achieved by cell disruption (Constant Systems Ltd.), the lysate was ultracentrifuged at 190,000 g for 1 h at 4 °C, and the filtered (0.22 μm) supernatant was loaded on a 5 mL HisTrap FF crude column (GE Healthcare Life Sciences). After a wash step, the protein was eluted with 150 mM imidazole (CtErbB2)

or 50 mM imidazole (Grb2). The purification tag was removed by TEV proteolytic cleavage, following conditions given in figure 35. Unsurprisingly, even at low temperature and with small amounts of TEV, the cleavage of CtErbB2 was complete, consistently with high accessibility due to disorder. On the contrary, cleavage of the full-length Grb2 tag was never complete, even with overnight incubation at 34 °C. The same observation was made for the NSH3 and NSH3SH2 constructs, suggesting that the His6 tag interacts with the NSH3 domain of Grb2. It is consistent with the fact that for these three constructs, the affinity of the tagged protein for the HisTrap column was low (the elution happened at low imidazole concentration, and partial elution was observed upon changes of flow rates even with no imidazole in the buffer).

Protein	Protein concentration	Tev:protein ratio	T	Duration	Gel well
CtErbB2	2.6 mg/mL	1:50	15 °C	overnight	B
		1:100			C
		1:50	4 °C		D
		1:100			E
Grb2	1 mg/mL	1:50	RT	2 h	2
	0.5 mg/mL		34 °C		3
		1:10	RT		4
			34 °C		5
	1 mg/mL	1:50	RT	4 h	7
	0.5 mg/mL		34 °C		8
		1:10	RT		9
			34 °C		10
	1 mg/mL	1:50	RT	overnight	6 and 11
				two overnight incubations	12
	0.5 mg/mL		34 °C	overnight	13
		1:10	RT		14
			34 °C		15

(a)



(b) *(c)*

Figure 35 – TEV cleavage conditions for CtErbB2 and Grb2. (a) Table of the different tests that were run. Chosen conditions are highlighted in gray. Each test corresponds to a lane of the 15% SDS-PAGE gels given for CtErbB2 in (b) and for Grb2 in (c). Chosen conditions are indicated with an arrow.

Since the TEV construct used contained a N-terminal His-tag, the excess of TEV protease and the cleaved tag were removed using a Ni-NTA resin (Thermo Scientific). The proteins were further purified by size exclusion chromatography using a Superdex-75 column (GE Healthcare Life Sciences) in MES buffer

(CtErbB2) or PIPES buffer (Grb2). In some early experiments, PBS buffer was used for Grb2, before it was changed for PIPES not to interfere with phosphotyrosine binding. The yield of [*U*-^{15}N; *U*-^{13}C]-CtErbB2 was about 6 mg for 1 L of M9 medium, while we obtained as much as 40 mg of [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-Grb2 for a 1 L culture. For all constructs (CtErbB2 and Grb2), purity of the final samples as assessed by SDS-PAGE was above 95%, and protein integrity was checked by mass spectrometry.

During all purification steps, great care was taken into keeping CtErbB2 samples at 4 °C, for its disordered nature makes it particularly prone to proteolytic degradation. Even with these precautions, proteolysis happened slowly over time, as seen from the appearance of numerous bands just below the one for CtErbB2 on SDS-PAGE gels before size exclusion chromatography (see figure 35b). Degradation was slowed down after the final purification step, but was still visible after typically 1 to 2 weeks in the spectrometer at 25 °C, even when antiprotease cocktails were used.

On the contrary, after the first affinity column, Grb2 was purified and kept at room temperature (except for long term storage, at -80 °C). At temperatures below about 10 °C, Grb2 samples looked turbid and viscous and the phenomenon was reversible. This was observed for the full-length protein, but also for the two bidomain constructs.

4.2 Generation of viral stocks for protein expression in insects cells using the Bac-to-bac® expression system

This expression system was used to produce the whole intracellular region of ErbB2, to study ErbB2 tyrosine phosphorylation, as exposed in section 10.

4.2.1 Principle of protein expression in Sf9 cells using the baculovirus expression system

The baculovirus expression system relies on the production of a recombinant protein via insertion of the gene of interest into a viral genome, transfection of cells to produce a viral stock, and finally infection of insect cells with this stock for protein expression. The time-consuming step of this expression system is the generation of a viral stock for each construct that needs to be expressed. Once it is done, the transfection is transient (there is no need to establish stable cell lines), fast and easy, and the protein can be harvested after only of few days. Infected cells are Sf9 cells, which are relatively easy to grow and passage, since they can be used both in adherent or suspension cultures.

The principle of the baculovirus Bac-to-bac® expression system (Invitrogen) is shown in figure 36. The

gene of interest is inserted in the viral genome in place of the gene coding for polyhedrin, a protein that is usually expressed at high levels by the virus to form occlusion bodies protecting it from the environment. Given that the viral plasmid (bacmid) is very large (around 130 kb), direct insertion of the gene of interest is difficult. In the Bac-to-bac® system, it is performed using transposition from a bacterial plasmid (pFastBac) to the bacmid in a specific strain of *E. coli* optimized for transposition. Once the bacmid contains the gene of interest, it is transfected into Sf9 cells for viral replication and creation of a viral stock, that can be further amplified by other cycles of infection/harvest of the virus. Protein production then simply consists of infection of Sf9 cells by addition of a controlled amount of the viral stock to the cell culture. The quantity of virus added is usually given in terms of multiplicity of infection (MOI), which is the quantity of virus (in plaque forming units, PFU) added per insect cell.

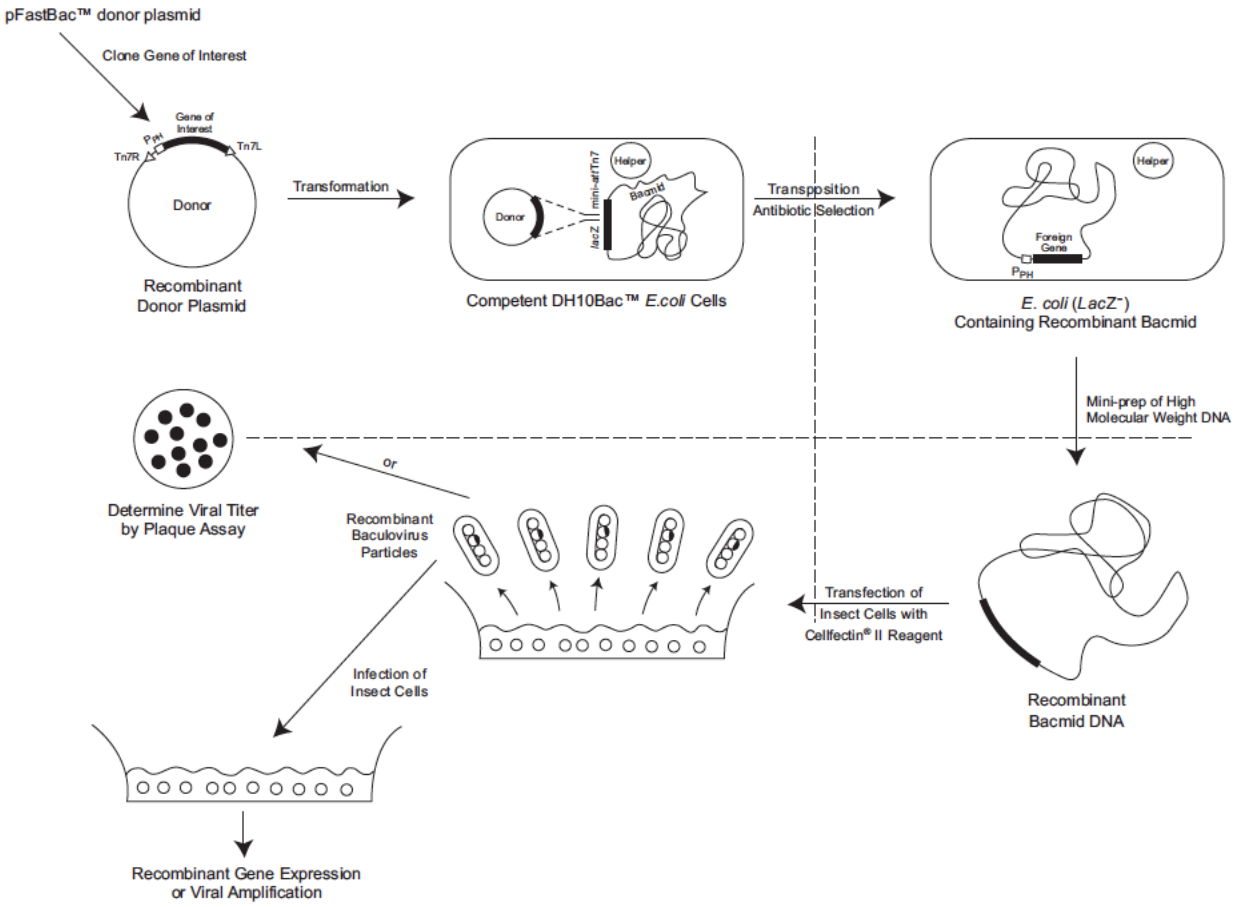


Figure 36 – Principle of protein production in insect cells using the baculovirus Bac-to-bac® expression system (Invitrogen)

4.2.2 Generation of the viral stock

We thank Dr. Arnaud Poterszman and Simon Pichard (baculovirus platform, IGBMC, Illkirch) for their advice and tests to guide us with this expression system. The bacmid was recombinantly generated in DH10Bac™ *E. coli* bacteria using the Bac-to-Bac® Baculovirus Expression System protocol (Invitrogen). The guidelines from the same protocol were followed for transfection and two cycles of virus amplification (resulting in a viral stock called P3).

Briefly, MAX Efficiency® DH10Bac™ competent cells were transformed with our pFastBac ™ construct by a 42 °C heat shock and plated on LB agar plus kanamycin/gentamicin/tetracyclin/X-gal/IPTG plates. After 48h at 37 °C, white colonies (for which both transformation and recombination of the gene of interest into the bacmid were successful) were picked and restreaked for confirmation. Colony PCR with pUC/M13 Forward and Reverse primers, followed by agarose gel electrophoresis, was performed for confirmation of successful recombination. Two clones with PCR products around 4 000 bp were selected and purified using the NucleoBond ® Xtra Midi EF (Macherey-Nagel). Transfection of the bacmid into Sf9 insect cells was performed in Grace's medium using Cellfectin® II reagent. Medium was collected 72 h after transfection and incubation (27 °C, 120 rpm), and after removing cells and debris by centrifugation and filtration (0.22 μm), the first viral stock (P1) + 2% fetal bovine serum was stored at 4°C in the dark. Two rounds of amplification were performed: the viral stock to amplify was added to log-phase Sf9 cells at a MOI of 0.05-0.1 and the medium was collected 3 to 7 days later to generate the new viral stock. This resulted in a final viral stock (P3) used for protein production.

4.2.3 Determination of the viral stock titer by flow cytometry

Usually, viral titers are obtained using plaque assays. However, these assays take at least one week. An alternative method was used using flow cytometry, and allowed the determination of viral titers in less than 24 hours. We thank Dr. Aurélie Albertini (I2BC, Gif-sur-Yvette) for the initial protocol. The method is based on recognition of infected cells, that display the baculovirus envelop protein gp64, with a phycoerythrin (PE)-conjugated anti-gp64 monoclonal antibody (AcV1, eBioscience). Determination of the proportion of infected cells (single or multiple infection) with flow cytometry allows for estimation of the MOI assuming a Poisson distribution.

2.10^6 Sf9 cells were infected with 100 μL of viral stock at different dilutions (from 1 to 1000-fold) in 6-well plates, and incubated at 27 °C for 18 hours with shaking (120 rpm). Cells were washed with 750 μL PBS

(by centrifugation at 450 g for 4 min and resuspension), centrifuged and resuspended in 100 μL of antibody solution (diluted 200-fold in PBS). Antibody binding was performed by incubation for 45 min at 4 °C on a rotative wheel. Cells were washed again with cold PBS, and final resuspension was performed in 300 μL PBS for flow cytometry.

Flow cytometry is a single-cell analysis method. Briefly, cells are discriminated thanks to the measure of forward-scattered light (FSC) that depends on their size, side-scattered light (SSC, at about 90 °) that depends on their granularity (or internal complexity), and fluorescence. Here, cells were analyzed by in a CytoFLEX S (Beckman Coulter) cytometer driven by the Cytexpert software. A first gating was made on FSC-SSC dot plot to focus on cells; doublets were discarded using SSC Area-SSC Height dot plot. PE fluorescence was excited by a 561-nm laser and collected through a 585/42 nm band pass filter. A total of 15000-20000 cells was acquired for each condition. The threshold of fluorescence for considering a cell as a positive event for PE (infected cell) was chosen as the highest level of autofluorescence observed for non-infected control cells. The present work has benefited from the CYTOMETRY facility of Imagerie-Gif, (`http://www.i2bc.paris-saclay.fr`), member of IBiSA (`http://www.ibisa.net`), supported by France BioImaging (ANR-10-INBS-04-01), and the Labex Saclay Plant Sciences (ANR-10-LABX-0040-SPS).

Once the percentage of infected (fluorescent) cells was determined, the MOI was calculated assuming that viral infection follows a Poisson distribution. That gives $MOI = -ln(P_{NI})$, P_{NI} being the percentage of non-infected cells. Since $MOI = \frac{N_{cells}}{N_{virus}}$ with N_{cells} the number of infected cells and N_{virus} the number of inoculated viruses (in PFU), the viral titer was easily determined knowing the initial number of cells. For the P3 viral stocks used, viral titer was between 1.10^8 and 5.10^8 PFU/mL, an expected value as provided by the manufacturer.

5 Nuclear Magnetic Resonance (NMR)

The typical samples for NMR studies were composed of 200-400 μM of protein, in 200 μL of buffer. We used 3 mm NMR tubes. Three different spectrometers, each equipped with a TCI cryoprobe (^{1}H, ^{13}C, ^{15}N, ^{2}H) and z-axis pulsed field gradients, were used: a Bruker Avance III 950 MHz spectrometer (22.3 T), a Bruker Avance III 800 MHz spectrometer (18.8 T), and a Bruker Avance I 600 MHz spectrometer (14.1 T)). Data were processed using different successive versions of Topspin (from 3.2 to 3.7, Bruker) and analyzed with CCPNMR (Vranken et al., 2005).

5.1 NMR samples, conditions, and assignments

5.1.1 Principles of heteronuclear sequential assignment of proteins

Thanks to advances in labeling and heteronuclear NMR pulse sequences and instrumentations, most NMR studies of proteins now begin with the ^{1}H, ^{15}N and ^{13}C backbone assignment of resonances. It relies on the ability to transfer magnetization between nuclei, thanks to the scalar couplings along the chain. Most commonly the excitation and observation are done on the proton, the most sensitive of the three nuclei ($\gamma_{1H} \approx 4\gamma_{13C} \approx -10\gamma_{15N}$). Using experiments transferring magnetization following a specific path, one is able to link ^{1}H, ^{15}N and ^{13}C resonances either in an intra-residue fashion, between sequential residues in the sequence, or both. The assignment process then relies on two processes. The first one is the determination of the nature of certain residues, using mainly $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts: typically, Gly, Ala , Ser and Thr have very recognizable patterns (see figure 37a). The second one is the identification of sequential links between them, as described below.

A typical example is the one of the HNCACB (given in figure 37b), which links the amide group of a residue i to both its $^{13}C_{\alpha}$ (green) and $^{13}C_{\beta}$ (blue) and the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of the preceding residue (i-1). The goal is then to find a HN strip (therefore belonging to another residue) that has its intra-residue $^{13}C_{\alpha}$-$^{13}C_{\beta}$ pattern matching the i-1 $^{13}C_{\alpha}$-$^{13}C_{\beta}$ pattern. This residue will therefore be assigned to i-1. When a sufficiently long stretch has been sequentially assigned with at least some of its residue types identified, it can be matched to sequences stretches expected from the known primary sequence of the protein.

This strategy works well in many cases, but others have to be implemented in certain cases, for example

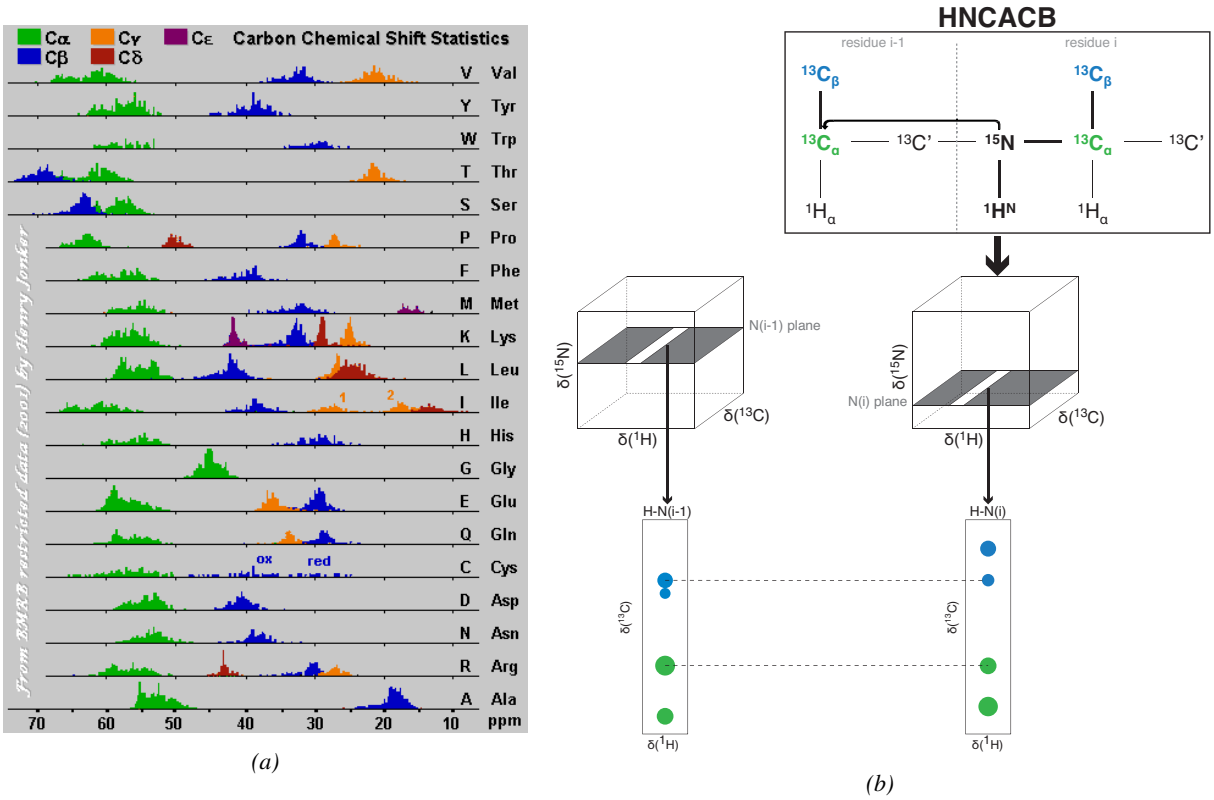


Figure 37 – Heteronuclear protein assignment. (a) Carbon chemical shift statistics for the 20 natural amino acid types in a protein, from the Biological Magnetic Resonance Bank (BMRB, `http://www.bmrb.wisc.edu/`). They allow unambiguous recognition of certain residue types (Gly, Ala, Ser and Thr mainly). (b) Principle of the backbone sequential assignment of a [U-^{15}N; U-^{13}C]-labeled protein, with an HNCACB experiment. The intra-residue carbon chemical shift pattern of a residue can be matched to the inter-residue carbon chemical shift pattern of another residue, identifying them as sequentially linked.

when carbon labeling is not feasible, when long experiments such as HNCACB (several hours to days) cannot be recorded due to limited stability of the sample, or when the protein is proline-rich, lacking many backbone amide proton and therefore creating too many interruptions in the sequential assignment. We had to use some of these strategies for CtErbB2, and they are described below.

5.1.2 Backbone and side-chains assignment of CtErbB2

Unless otherwise stated, CtErbB2 was studied in MES buffer pH 5.6 (see table 2) and spectra were recorded at 298 K. Consistently with reasons stated in 1.5.2, these acidic, low temperature conditions seemed optimal for this protein. Assignment experiments were performed at proton frequencies of 950 MHz or 800 MHz. High fields are absolutely necessary for a disordered protein of this size, for better resolution: 227 ^{1}H-^{15}N HSQC backbone cross peaks are expected (228 non-proline residues, minus the N-terminal glycine for

which unfavorable exchange is expected) in a proton dispersion range of about 1ppm only.
For backbone assignment, 3D HNCO, HN(CA)CO, HNCA, HNCACB, HN(CO)CA, HN(CO)CACB, HNCANH and HN(CO)CANH spectra were used. The last two experiments enable one to directly link backbone amides of sequential residues, and are particularly convenient and applicable to IDPs, since they require long signal lifetime for long magnetization transfer. All 3D experiments were recorded as BEST (band-selective excitation short-transient) experiments (Lescop et al., 2007). This technique allows reduction of the inter-scan delay (and thus great reduction of the experiment time) by optimizing longitudinal ^{1}H relaxation. The assignment of residue 161 (between prolines) was performed thanks to an HSQC spectrum of the Q161A mutant. For quick reassignment of the ^{1}H-^{15}N HSQC spectra upon slight spectral changes (temperature, pH, single point mutation...), HADAMAC experiments were used (Lescop et al., 2008). These amino acid-type edited experiments enable to distinguish 7 groups of residues, depending on side-chain C-C and C-H multiplicity, presence and chemical shift range of $^{13}C_{\beta}$, and number of carbonyl and aliphatic $^{13}C_{\gamma}$. The output spectrum is a pseudo-3D in which each ^{1}H-^{15}N HSQC plane contains the amide peaks of all residues following a particular amino acid type. For example, one plane will contain the amide peaks of all residues following a glycine. A superimposition of all the planes of an HADAMAC experiment used for assignment confirmation of the C146S mutant is shown in figure 38.
The side chains were assigned with 3D H(C)C(CO)NH, (H)CC(CO)NH, HBHANH and HBHA(CO)NH experiments.
To assign the particularly challenging proline-rich regions (CtErbB2 contains 44 prolines), we recorded:

- proton-detected HCAN, HCANi and HACA(CO)N 3D spectra, correlating $^{1}H_{\alpha}$ and $^{13}C_{\alpha}$ of residue i to ^{15}N of residue i and/or i+1
- a ^{13}C-detected 2D CON spectrum, correlating ^{15}N of residue i to ^{13}C' residues i and i-1

5.1.3 Backbone assignment of Grb2

Due to difficulties in removing the His6 purification tag, some His-tagged constructs were assigned and studied, in addition to the tag-cleaved constructs. Therefore the following constructs have been assigned: His-NSH3, NSH3, SH2, CSH3, His-NSH3SH2, NSH3SH2, SH2CSH3, His-Grb2 and Grb2 (Grb2 refers to the full-length construct). Unless otherwise stated, they were assigned and studied in PIPES buffer at pH 7.2 (see table 2). Assignment in PBS buffer was also initially performed, but left largely unused due to

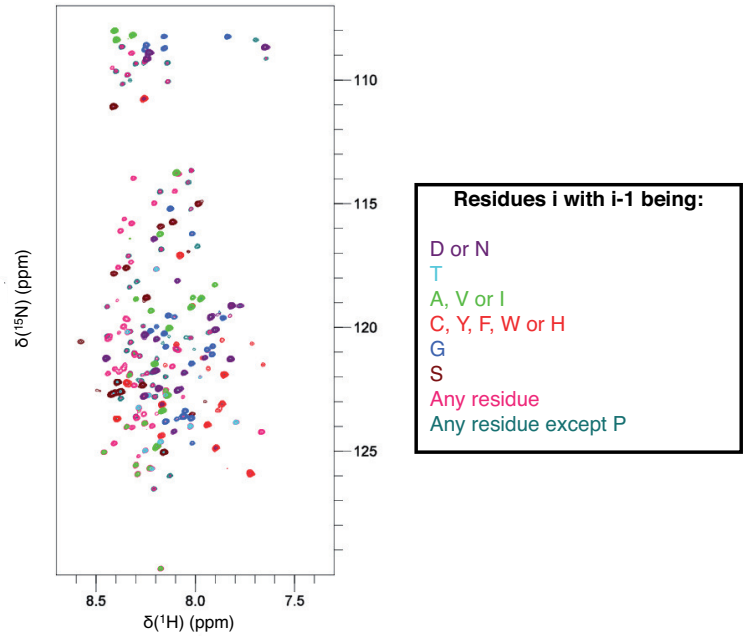


Figure 38 – HADAMAC experiment for quick confirmation of the assignment of the C146S mutant of CtErbB2. Here is shown the superposition of the 8 planes of the pseudo-3D.

observed interaction of the SH2 domain with phosphate that could interfere with the observation of SH2-phosphotyrosine interactions. The only construct for which the assignment was not transferred from PBS to PIPES buffer was the His-tagged NSH3 domain. Spectra were recorded at 308 K. Assignment experiments were mostly recorded at 600 MHz, for it gave the narrowest peaks due to the exchange processes at stake in the NSH3, but some experiments were also recorded at 800 or 950 MHz.

Backbone and $^{13}C_{\beta}$ assignments of all three individual domain constructs were performed on [*U*-^{15}N; *U*-^{13}C]-labeled samples, using different combinations of 3D experiments: HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and HNCANH in their BEST versions (Lescop et al., 2007). Backbone ^{1}H-^{15}N assignment of the SH2CSH3 bidomain was performed on a [*U*-^{15}N; *U*-^{13}C]-labeled sample mainly from transfer of the assignments of the SH2 and CSH3 monodomains thanks to good superimposition, and confirmed with HNCACB and HN(CO)CACB spectra. Assignments of the NSH3SH2 bidomain and of full-length Grb2 were performed on [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-labeled samples using the same types of experiments as for monodomains, occasionally transferring assignments from the isolated domains when necessary and possible. For the bidomains and FL Grb2, the 3D experiments were recorded in the BEST-TROSY version. TROSY (transverse relaxation-optimized spectroscopy)-based experiments (Per-

vushin et al., 1997) take advantage of the different widths of the components of a J-coupling multiplet. The component for which chemical shift anisotropy and dipole-dipole coupling effects on transverse relaxation cancel most efficiently is selected. This cancellation is best achieved at high fields. The technique gives narrower peaks than common experiments, increasing sensitivity and resolution, and is particularly interesting for large proteins with fast transverse relaxation.

5.2 Chemical shift analysis

Chemical shifts of protein residues depend on a variety of parameters, mainly amino acid type and neighbors, experimental conditions such as buffer composition, pH and temperature, but also, and most interestingly for us, on local structure. The nuclei most commonly used for prediction, because their chemical shifts are mostly influenced by secondary structure, are $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{13}C_{O}$. Amide proton and nitrogen chemical shifts are too strongly dependent on experimental conditions (pH, temperature), and are therefore more difficult to relate to secondary structure. It is to be noted that chemical shifts are also dependent on tertiary structure, but the effects are more subtle and less easy to predict.

Secondary structure calculations For Grb2, structural integrity of the individual domains (which structures are known) was verified using chemical shifts and TALOS + (`https://spin.niddk.nih.gov/bax/nmrserver/talos/`) (Shen et al., 2009). TALOS+ uses a set of 200 proteins which extensive chemical shift information have been deposited to the BMRB and for which a crystal structure is known. A neural network algorithm trained on this database then allows for prediction of backbone dihedral angles and secondary structure from chemical shifts and residue types.
For disordered proteins or regions, for which we look for structural propensities rather than stable structures, propensities are most often quantified using secondary chemical shifts (Wishart and Sykes, 1994) defined as

$$\Delta\delta = \delta - \delta_{RC}$$

where δ_{RC} is the expected chemical shift for a random coil. Different calculation methods exist, using different random coil chemical shifts databases. We used one of the most common ones, the SSP method (Marsh et al., 2006). This method takes into account several chemical shifts, weighted by the sensitivity of the corresponding nucleus to secondary structures, and returns a score that is expected to be a quantitative

estimate of the propensity for secondary structure. SSP scores are often simplified saying that a positive score indicates α helices and that negative scores indicate β-strands. However, this interpretation ignores the other types of structures, that can be important. More precisely, compact structures, which we can define as structures with a translation per residue typically below 2 Å, have a positive score. Most commonly, positive SSPs therefore indicate the presence of a α-helix, but can also point to 3_{10} helices. Although PPI helices, which are right-handed polyproline helices with only *cis* peptide bonds, are compact structures and could theoretically also be included in positive SSP types of structure, they have never been observed in water so far. Regions folding into extended structures, β-sheets but also PPII helices, with a typical translation per residue above 3 Å, have negative SSP scores. With default settings (that we used), the scores are averaged over 5 residues to smooth the profile. As recommended for disordered proteins, only $^{13}C_\alpha$, $^{13}C_\beta$ and $^{1}H_\alpha$ chemical shifts were used. The residues preceding prolines were removed from the calculation, since their chemical shifts are often extreme outliers.

Combined chemical shift perturbations (CCSP) When comparing spectra from different constructs or between different conditions (temperature, pH, presence of an interaction partner...) its is convenient to combine the chemical shift differences of several nuclei into a single value. To do so, the contributions from the different nuclei need to be weighted either by their gyromagnetic ratio or by the relative width of their chemical shift range. Therefore, to combine ^{1}H and ^{15}N chemical shift perturbation information, we used $CCSP = \sqrt{(\Delta\delta_{1H})^2 + (0.15\Delta\delta_{15N})^2}$.

5.3 ^{15}N relaxation studies

5.3.1 Dynamic information given by ^{15}N relaxation parameters. Expectations for globular proteins and IDPs.

Longitudinal (R_1) and transverse (R_2) ^{15}N relaxation rates, as well as steady-state heteronuclear {^{1}H}-^{15}N nuclear Overhauser effect values (hereafter simply referred to as {^{1}H}-^{15}N nOes) are routinely recorded to study protein dynamics. Each of these parameters is sensitive to different reorientation timescales of the N-H bond. While {^{1}H}-^{15}N nOes, sensitive to frequencies in the range of the proton resonance, account for dynamics in the sub-ns timescale, R_1 and R_2 rates are mainly sensitive to the contribution in the range of the nitrogen frequency (a few ns). The dependence of each parameters on the rotational correlation time τ_C of the NH bond, in a model of Brownian motions with a restricted libration ($S^2 = 0.8$, $\tau_i = 50ps$), is

given in figure 39. Additionally, R_2 is sensitive to slower dynamics, in the μs-ms timescale, and therefore to exchange processes commonly observed in proteins.

For a denatured or unfolded protein, a simple polymer model described by Schwalbe et al. (1997) suggests that the relaxation parameters should have bell shapes. The profile should be flat at the center of the protein, with increased flexibility (decreased R_1, R_2 and $\{^1H\}$-^{15}N nOes) at the extremities, governed by the length of the statistical element of the polymer. For a denatured protein, this length is empirically about 7 residues. Any difference from this bell-shape therefore indicates deviation from a random coil behavior. The use of NMR relaxation data for the investigation of intrinsically disordered protein is not trivial and is still being developed, as reviewed by Salvi et al. (2017).

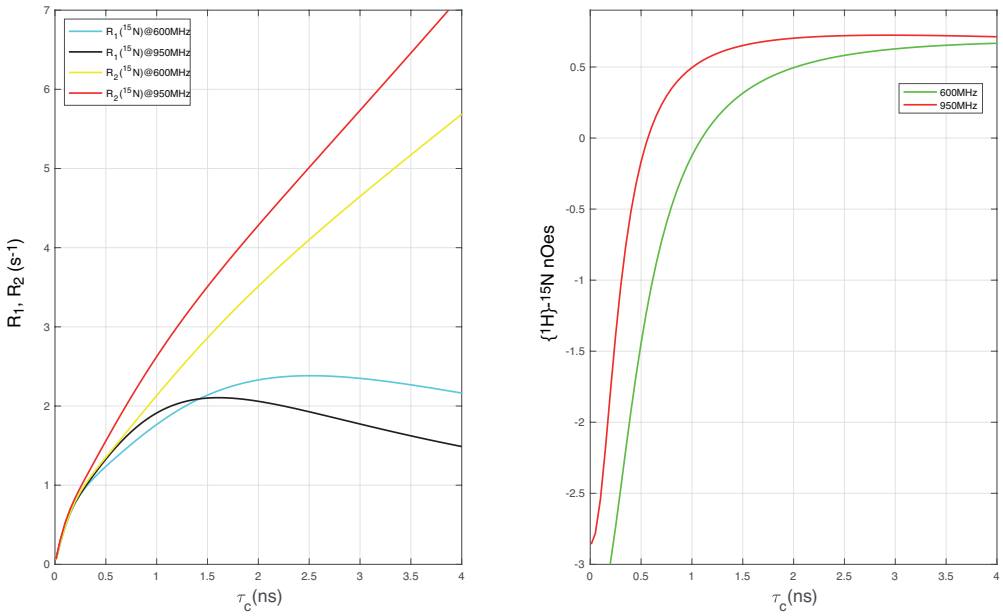


Figure 39 – Expected ^{15}N R_1, R_2 and $\{^1H\}$-^{15}N nOes for a chain with NH bonds experiencing a single rotational correlation time τ_C and libration with an order parameter $S^2 = 0.8$ and internal correlation time $\tau_i = 50ps$. The curves are drawn for two magnetic fields, 14.1 T (corresponding to a proton resonance frequency of 600 MHz) and 22.3 T (950 MHz proton frequency).

5.3.2 NMR experiments and analysis

For CtErbB2, the three parameters mentioned above were measured at proton frequencies of 600 and 950 MHz (nitrogen frequencies of about 60 and 95 MHz), on [U-^{15}N] labeled samples to prevent artefacts from ^{13}C coupling. The pulse-sequences used were the HSQC-based pulse sequence implemented by Bruker as interleaved pseudo-3D experiments. 10 to 14 delays were used: from 20 ms to 2 s for R_1 mea-

surements, and from 4 to 450 ms for R_2 measurements, with one delay being measured three times for error estimation. Relaxation rates were extracted fitting the curves to a mono-exponential decay. {^{1}H}-^{15}N nOes were measured with a TROSY-based Bruker pulse sequence using 120° pulses every 5 ms for 4 s for proton saturation. They were calculated as the intensity ratio of each peak in the experiment with or without proton saturation.

For Grb2, R_1, R_2 and {^{1}H}-^{15}N nOes were measured at 800 MHz, on [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H] samples, with carbon decoupling. Other than that, the pulse sequences were the same as for CtErbB2, with delays ranging from 17 to 305 ms for R_2 measurements, and from 20 ms to 1.8 s for R_1 measurements. From the R_1 and R_2 rates, in absence of conformational exchange contribution, it is possible to approximate the rotational correlation time τ_C using the equation $\tau_C \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{R_2}{R_1}-7}$ (Fushman et al., 1994). We compared this value with the value of τ_C derived from the Stokes-Einstein (SE) equation for an ideal spherical object. It was shown than τ_C (SE) found with the Stokes-Einstein equation was overestimated by a factor of two for globular proteins, giving after correction: $\tau_C \approx \frac{\tau_C(SE)}{2} = \frac{1}{2D_R} = \frac{4\pi\eta r^3}{k_B T}$ (Halle and Davidovic, 2003), where D_R is the rotational diffusion coefficient, η the viscosity of the buffer and r the radius of the sphere. Considering a globular, spherical protein with a partial specific volume of $0.73 cm^3.g^{-1}$ and with a viscosity of the buffer of 0.72 mPa.s at 308 K, it yields an estimated τ_C at this temperature $\tau_C \approx 0.615 * MW$ with τ_C in ns and MW the molecular weight in kDa.

5.4 Scalar couplings

Proton scalar couplings (or J_{HH} couplings) in a protein depend on dihedral angles, and thus give information about the local conformation of the polypeptide chain. An empirical formula described by Karplus links J values to the dihedral angle ϕ (Karplus, 1959): $^3J_{HH}(\phi) = Ccos(2\phi) + Bcos(\phi) + A$. However, it is sometimes difficult to find the most accurate Karplus parameters A, B and C, that depend greatly on the sequence of the protein and experimental conditions.

Here, homonuclear $^3J_{HN H\alpha}$ coupling constants were measured using the Bruker-implemented version of the 3D HNHA experiment of Vuister and Bax (Vuister and Bax, 1993). To analyze them more accurately, without the need to calculate dihedral angles, we compared them to $^3J_{HN H\alpha}$ calculated for our sequence supposing it is in a random coil conformation, using the server `https://spin.niddk.nih.gov/bax/nmrserver/rc_3Jhnha/` (Shen et al., 2018), that takes into account sample temperature and pressure. While $^3J_{HN H\alpha}$ higher than random coil indicate propensity for β-sheets, α helices have values lower than random coil, and PPII helices close or slightly lower than random coil (Smith et al., 1996).

5.5 Residual dipolar couplings (RDCs)

5.5.1 (Residual) dipolar couplings

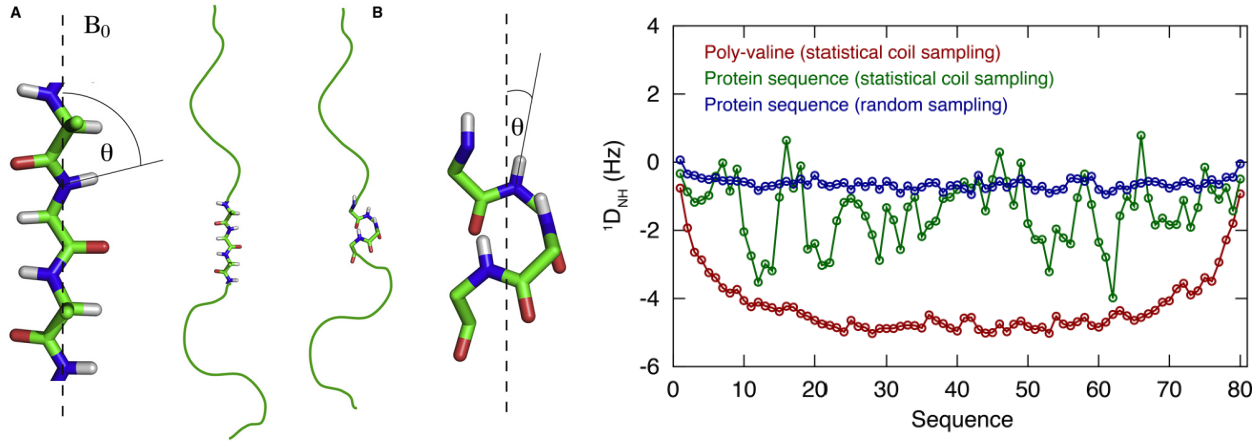
Dipolar coupling, in NMR, is an anisotropic interaction between two nuclear spins (that behave as magnetic dipoles), through space. The strength of this interaction between spin i and j is

$$D_{ij}(\theta, r_{ij}) = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2 r_{ij}^3} < \frac{3cos^2(\theta) - 1}{2} >$$

where θ is the angle between the internuclear ij vector and the static magnetic field B_0 (as shown in figure 40a for the N-H bond of a protein), and r_{ij} is the distance between the nuclei, supposed constant in the case of nuclei linked by a covalent bond. Even though the magnetic susceptibility tensor of proteins are not strictly isotropic, because of spatial and time averaging due to tumbling, in solution dipolar couplings average to 0. However, dipolar couplings can give valuable information on local orientation. It is possible to reintroduce them partially using slightly anisotropic media, and thus have access to this information in high resolution solution state NMR. They are then called residual dipolar couplings, or RDCs.

5.5.2 Expected RDCs for an IDP

Here we focus on backbone N-H RDCs, or $^1D_{NH}$, which are the easiest one to measure for a protein. Measurement of other types of RDCs often requires longer acquisition times, and increases the risk of protein degradation and/or dealignment of the medium. As shown in figure 40a, in the most common media that are aligned parallel to the magnetic field, α helices have their backbone N-H bonds roughly parallel to the direction of the chain and will give positive RDCs. Extended structures, such as β-sheets or PPII helices, with their N-H bond roughly perpendicular to the direction of the chain, will give negative RDCs. For a random coil, even when molecular tumbling is partially suppressed, absence of preferred orientation of the chain should intuitively result in spatial averaging and absence of measurable RDCs, as shown in blue in figure 40b. However, excluded volume effects locally constrain the chain, in a manner dependent on the type of amino-acid (Bernadó et al., 2005). This leads to expected RDC values mainly negative for disordered proteins (red and green in figure 40b). Deviation from the values determined for a statistical coil samplings indicates propensities for more extended (lower values) or α helical (higher values) structures.



(a) *(b)*

Figure 40 – Expected RDCs (a) θ angles of NH bonds relative to the magnetic field expected for extended (A) and α helical (B) structures. For extended structures, θ is close to 90°, while for α helices θ is close to 0. From Jensen et al. (2009). (b) Expected RDC values calculated with Flexible Meccano (see section 6.3) for a poly-valine chain sampling the statistical coil populations for this amino-acid (red) and for an arbitrary protein sequence, sampling statistical coil populations determined for each amino-acid type (green), or with completely random sampling of dihedral angles (blue). From Milles et al. (2018).

5.5.3 Protocol

Common aligning media include bicelles or bacteriophages, which due to their large anisotropic magnetic susceptibility align the medium parallel to the magnetic field. Acrylamide gels are also used, and can be compressed or stretched parallel or perpendicular to the NMR tube and therefore to the magnetic field. Alignment in bicelles or phages is stronger (dozens of Hz compared to a few Hz in acrylamide gels), resulting in more precise measurement of RDCs. However, the choice of the aligning medium must also take into account its (non) interaction with the protein of interest, that can be verified comparing the spectra in solution and in the alignment media. In our case, we found that PEG-hexanol mixtures interact with CtErbB2, resulting in heterogeneous peak broadening, and aberrant measurements of RDCs. We therefore chose the acrylamide gels, which was another technique that had been tested in the lab, and seemed less prone to yield specific interactions with proteins.

We used the method described by Chou et al. (2001). 6% acrylamide gels were polymerized in the MES buffer in a cylindrical chamber of 6mm diameter, cut to a length of 1cm, and then dialyzed in water overnight. The gels were then allowed to dry for several days at room temperature, and then stocked up to several months before use.

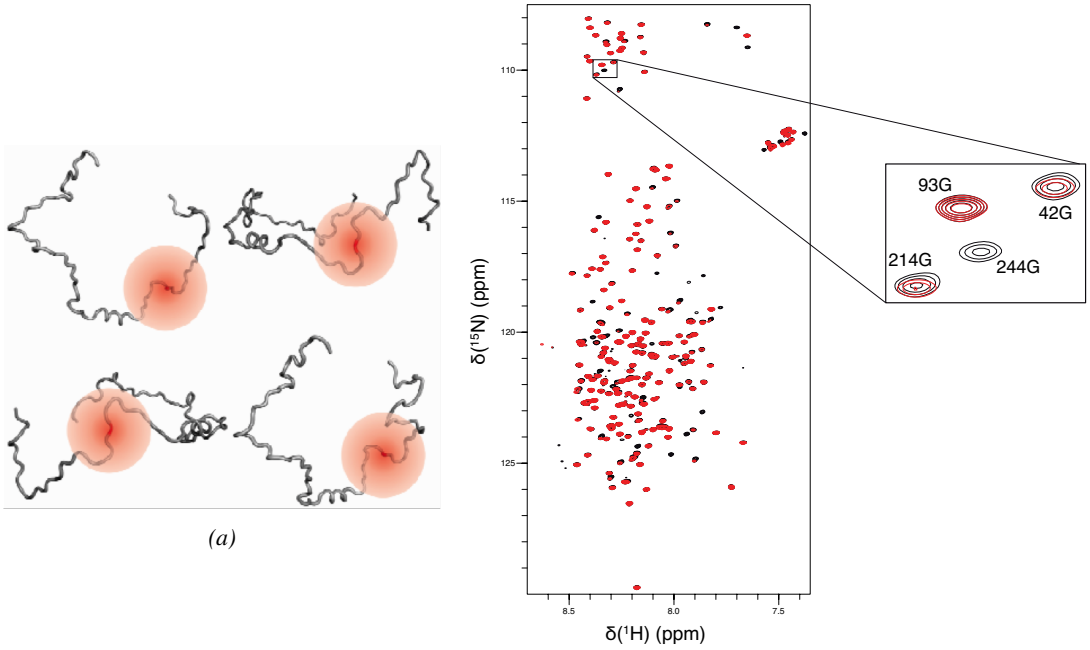
A BEST-type ^{1}H-^{15}N HSQC-IPAP spectrum of the protein was recorded in the standard buffer and condi-

tions with 2048 points in the indirect dimension, at 800 MHz. This spectrum allowed measurement of N-H scalar couplings. A gel was then rehydrated with 300 μL of this same protein solution, in the 6 mm chamber, for two days. It was then transferred carefully into a 5 mm, open-ended, Shigemi tube, which forced it to stretch, and left to settle overnight. The same BEST-type ^{1}H-^{15}N HSQC-IPAP was recorded, allowing to measure the sum of scalar and RDCs, and to calculate the RDCs. The whole experiment was repeated with a new sample, to check reproducibility and estimate errors. Each time, the obtained alignment was verified using a 1D deuterium spectrum, giving deuterium quadripolar splitting of respectively 4.98 and 5.66 Hz.

5.6 Paramagnetic relaxation enhancements (PREs)

Principle In NMR of globular proteins, distance information is routinely derived from nOe data. In most cases, the contacts in IDPs are too transient to give rise to a significant nuclear Overhauser effect. PRE experiments use the very high gyromagnetic ratio of unpaired electrons (more than 600 times higher than the one of proton) to generate stronger dipolar interactions. If a paramagnetic species is introduced at a specific position of the IDP sequence, nuclei close in space (as schematized by the red sphere in figure 41a) will relax faster and their peaks will have decreased intensity. Usually, the introduction of a paramagnetic species is done by coupling of a nitroxide compound on a cysteine via a disulfide bond. Typically, ^{1}H-^{15}N HSQC peaks of residues adjacent to the paramagnetic species become invisible, and the peaks of the residues that are brought close (even transiently) to the probe by long-range contacts will also show decreased intensity. An example is shown for CtErbB2 in figure 41b: introduction of the paramagnetic probe on position 227 induced decreased intensity of the peaks of residues relatively close in the sequence (e.g. residues 214 and 244), but also of residues that are far in the sequence (e.g. residue 42), revealing a long-range contact.

Protocol Four mutants of CtErbB2 were used to couple a nitroxide probe on four different cysteines in separate experiments: C146S, C45S, (C45S,C146S,S227C) and (C45S,C146S,S248C) (for respective paramagnetic tags on C45, C146, C227 and C248). These [*U*-^{15}N; *U*-^{13}C]-labeled mutants were reduced with 10 mM TCEP. TCEP was removed (using ZebaSpin Desalting columns, ThermoScientific), and the mutants were incubated at 15°C overnight with the paramagnetic probe (1:10 protein:probe molar ratio) in a 20 mM sodium borate buffer at pH 8.0. The buffer was then exchanged for the MES buffer pH 5.6, removing the excess of free paramagnetic probe. The ^{1}H-^{15}N HSQC spectrum of this paramagnetic form was recorded at



(a) *(b)*

Figure 41 – PRE experiments. (a) The paramagnetic probe, coupled on a cysteine residue of the protein, induces an increase in relaxation for any nucleus nearby (in a radius of about 20 Å represented here as a red sphere). The closer the nucleus, the faster its relaxation, the lower its resonance intensity. From Kosol et al. (2013). (b) Example of a PRE experiment on CtErbB2 (where the MTSL probe is coupled on residue 227). In black is the spectrum of the sample where the probe is reduced (diamagnetic, no effect on relaxation) and in red is the spectrum of the sample where the probe is in its oxidized state (paramagnetic).

950 MHz. The probe coupling was total, as assessed by MALDI-TOF mass spectrometry (data not shown). To obtain a reference spectrum of the non-paramagnetic protein, the sample was reduced with ascorbic acid (1:6 protein:ascorbic acid molar ratio) and a second HSQC spectrum was acquired. The assignment of this reference spectrum was verified by recording BEST-HNCA, BEST-HNCACB, BEST-HN(CO)CACB and HADAMAC experiments. Except for the peaks adjacent to the probe-coupled cysteine, no chemical shift difference was observed compared to the uncoupled protein, which ensured that the probe did not have any impact other than local on the structural ensemble of CtErbB2.

The probes that were used in this study are 3-(2-Iodoacetamido)-PROXYL (IAP) and S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL). We observed two sets of peaks for residues close to the IAP-coupled cysteine in the reduced form, indicating that the IAP probe was actually a racemic mixture. This prevented measurement of accurate intensities. This was not observed for MTSL, and therefore only the results with this second probe where kept. However, even without quantification, we could observe that the perturbed residues were the same with IAP than with MTSL.

5.7 NMR titrations

All NMR titrations between Grb2 and CtErbB2 (or CtErbB2 peptides) were recorded in the PIPES buffer (see table 2), at 298 K. The buffer was chosen for the pH to be closer to the physiological pH (compared to the MES buffer pH 5.6 used for the study of CtErbB2), and the temperature was chosen to be 298 K to avoid degradation of CtErbB2 over time (even though the optimal temperature for the stability and spectral quality of Grb2 is 308 K). Spectra were recorded at 950 MHz when studying the effect of binding on CtErbB2, and 600 MHz when studying the effect of binding on Grb2. For each point of a titration, ^{1}H-^{15}N HSQC spectra with identical parameters were recorded.

Titration of labeled CtErbB2 by unlabeled Grb2 Before titration, backbone assignment of CtErbB2 in PIPES buffer at 298 K was performed using HNCACB and HN(CO)CACB experiments. The titration was done on a 50 μM [*U*-^{15}N; *U*-^{13}C] unphosphorylated sample of CtErbB2. Addition of a sample of unlabeled FL Grb2 at 1.2 mM in the same buffer was done progressively to have titration points with CtErbB2:Grb2 ratios of 1:0, 1:1, 1:2, 1:5 and 1:10. Because the addition of Grb2 caused dilution, the final sample was concentrated back to its initial volume of 200 μL.

Titration of labeled Grb2 by unlabeled CtErbB2 Before titration, backbone assignment of Grb2 at 298 K was performed. The titration was done on a [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-labeled sample of Grb2 at 100 μM. Unlabeled, unphosphorylated CtErbB2 at 3.6 mM was added progressively to give Grb2:CtErbB2 ratios of 1:0, 1:1, 1:2, 1:3 and 1:5.

Titration of labeled Grb2 by unlabeled peptides of Grb2 The unphosphorylated P1 (sequence VAPLTCSPQPEYVNQPDVRPQPPSPREGPLPAARPA) and phosphorylated P2 (VAPLTCSPQPEpYVNQPDVRPQPPSPREGPLPAARPA) peptides were purchased from Schafer-N (Copenhagen), their pH was adjusted to 7 in water and they were lyophilized for titration experiments. Each titration was done on a [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-labeled sample of Grb2 at 100 μM. Peptides were added to final successive Grb2:peptide ratios of 1:0, 1:0.25, 1:0.5, 1:1, 1:2, 1:5, 1:10 and 1:20.

6 Other biochemical, biophysical and bioinformatic techniques

6.1 Circular dichroism (CD)

6.1.1 CD of folded and unfolded proteins

Circular dichroism is based on the principle that optically active solutions have different molar extinction coefficients ε at a given wavelength for left and right circularly polarized light. This difference is measured with the ellipticity, θ, which is defined as $\theta = tan^{-1}(b/a)$ where b and a are the minor and major axes of the ellipse formed by the resulting electric field vectors. The results are often normalized with molarity, using the molar ellipticity in $deg.cm^2.dmol^{-1}$ $[\theta] = \frac{100\theta}{Cl}$, where C is the molar concentration in M and l the path length of the cell in cm. For polymers (and especially proteins), it can be further normalized with the number of monomers. For proteins it gives the molar ellipticity per residue $MER = \frac{[\theta]}{n}$ where n is the number of residues in the protein.

Far-UV CD allows one to distinguish, and roughly quantify, the presence of different secondary structures in a protein. Spectra of different pure secondary structures are given in figure 42. Usually, PPII helices are not taken into account in CD studies, and are classified as "random coil". From denaturation studies of collagen, it has been suggested that the positive peak around 220 nm observed on the spectrum attributed to random coil actually arises from PPII structures (Lopes et al., 2014).

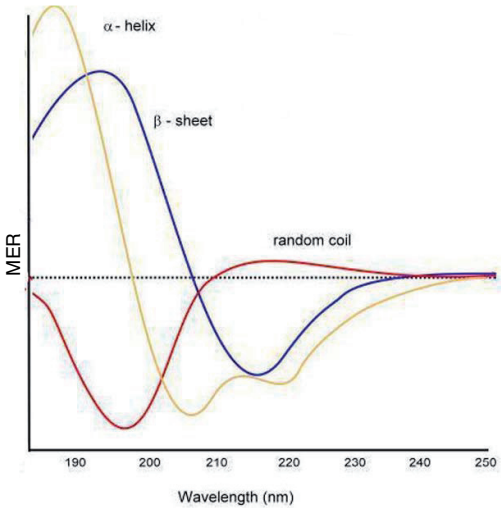


Figure 42 – Circular dichroism spectra of pure α-helix (yellow), pure β-sheet (blue), and pure random coil (red). From Dodero (2011).

6.1.2 Protocol

The sample for CD measurements was CtErbB2 at 198μM in the same buffer as the one for NMR. Measurements were performed on a Jasco J-810 spectropolarimeter, in a 0.01 mm path length quartz cell. The wavelength range was 190-260 nm, with a step of 0.1 nm. Temperature was controlled to 298 K using a sample cell temperature control unit (PFD 423S/L Peltier). The spectra consist of an average of 10 scans acquired at 50 nm/min, with buffer substraction and normalization of the signal with protein concentration and number of residues to yield results in molar ellipticity per residue (MER).

6.2 Small-angle X-ray scattering (SAXS)

6.2.1 Principle of SAXS

SAXS gives structural information at low resolution, such as the size and shape of a macromolecule. It is a method complementary to high-resolution techniques, and especially solution-state NMR, for the description of proteins in solution. The schematic representation of a SAXS experiment is given in figure 43a. A quartz capillary containing the solution to analyze is placed in front of a collimated monochromatic X-ray beam, and the intensity scattered by the sample is measured as a function of the momentum transfer q, also sometimes noted s, which is equal to $\frac{4\pi sin(\theta)}{\lambda}$, where 2θ is the scattering angle. The requirement in terms of sample quantity and buffer are roughly equivalent to NMR, but the sample has to be monodisperse (typically a single oligomeric state should be present). The sample should also be diluted enough so that intermolecular distances do not overlap with intramolecular distances information (the solution needs to be ideal).
The first parameters that can be obtained by SAXS are the radius of gyration R_g of the particle and its molecular weight. This is usually done using the Guinier approximation

$$ln(I(q)) = ln(I(0)) - \frac{q^2 R_g^2}{3}$$

which is valid in the region of small angles (the limit is usually set at $qR_g < 1.0$). The linear fitting of the Guinier plot (ln(I(q)) as a function of q^2) in this region gives I(0) and R_g. From I(0), knowing the concentration of the protein and estimating its partial specific volume from its sequence, an estimated molecular weight can be extracted. For IDPs, for which some conformations can have a very high radius of gyration, the Guinier region may be too narrow to extract reliable parameters. The distance distribution function P(r) may then be used.

P(r) is the autocorrelation function of the scattering curve, and is calculated via an inverse Fourier transform:

$$P(r) = 2r^2 \int_0^\infty q^2 I(q) \frac{sin(qr)}{qr} dq$$

It gives an histogram of the interatomic distances of the protein, and thus information on the shape of the protein. It gives D_{max}, the maximum r for which $P(r) \neq 0$. R_g and I(0) can also be extracted from the P(r) distribution, allowing to get rid of the Guinier limitations:

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr}$$

$$I(0) = 4\pi \int_0^{D_{max}} P(r) dr$$

The Kratky plot is a useful representation of SAXS data. It gives graphical information about the shape of the protein, and especially about its ordered or disordered nature, as shown in the following section. The original Kratky plot is a plot of $q^2 I(q)$ as a function of q. A normalized version was proposed by Durand et al. (2010), plotting $(qR_g)^2 \frac{I(q)}{I(0)}$, and allowing to directly compare proteins.
Examples of simulated scattering curves, P(r) distributions Kratky plots and dimensionless Kratky plots are given in figure 43b. One can clearly distinguish a globular from a modular or disordered protein. In particular, the dimensionless Kratky plot has a maximum at coordinates $(\sqrt{3}, 1.1)$ for globular proteins, and a plateau or increase at high q for disordered proteins. The use of SAXS for disordered proteins has been reviewed by Kikhney and Svergun (2015) and Receveur-Brechot and Durand (2012).

6.2.2 Data acquisition

CtErbB2 The samples were similar than those for NMR, except that the protein was not isotopically labeled. The sample was extensively dialyzed in the MES buffer before SAXS measurements, and the dialysis buffer was used for SAXS buffer substraction. Data acquisition was performed by Dr. Dominique Durand. The sample was placed in a capillary cell of 1.8 mm diameter at 285 K. The curves were recorded on a Bruker Nanostar instrument equipped with a Microstar rotating anode (λ=1.54 Å). The Vantec bidimensional detector was placed 75 cm away from the sample. Data for the buffer and the sample were recorded in an interleaved manner (10 one-hour recordings for each), and the buffer curves were averaged and sub-

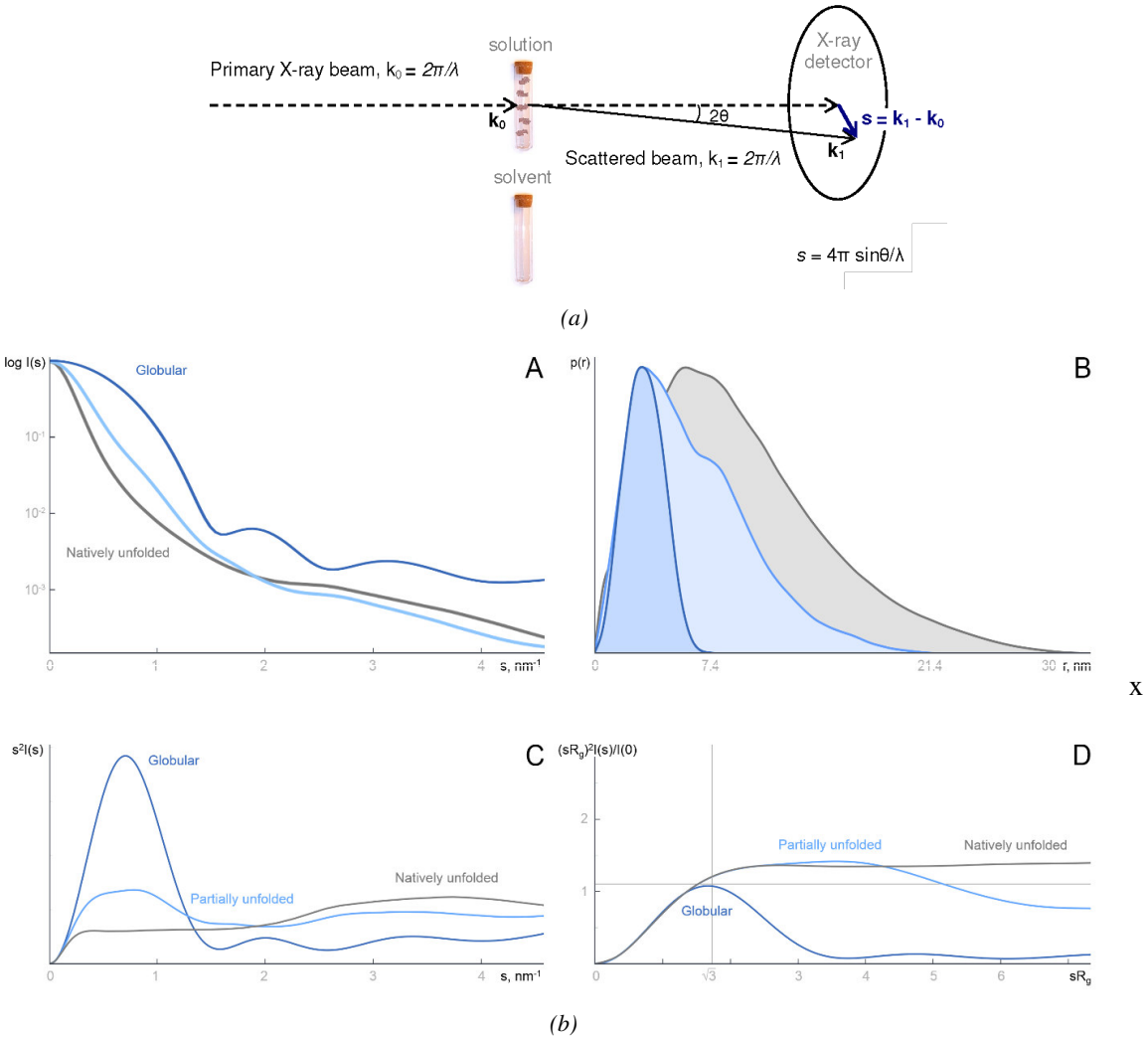


Figure 43 – SAXS of globular or (partially) disordered proteins (a) Schematic principle of a SAXS experiment. (b) Simulated SAXS scattering curve (A), P(r) distribution (B), Kratky plot (C) and dimensionless Kratky plot (D) for a 60 kDa protein, globular (dark blue), 50% unfolded (light blue) or completely unfolded (grey). From Kikhney and Svergun (2015).

stracted from the sample curves. The curves were normalized to the intensity of pure water. Absence of radiation damage was checked comparing curves over time. Several concentrations ranging from 50 to 200 μM were tested, yielding no difference in scattering, indicating the absence of oligomerization effects. The curves with more signal-to-noise ratio (at 200 μM) were therefore chosen for analysis.

Grb2 The SAXS measurements were done at the SWING beamline at the SOLEIL synchrotron (Saint Aubin, France) with Dr. Aurélien Thureau. The samples were injected at a flux of 0.3 mL/min into an Agilent Bio SEC-3-300 size exclusion column pre-equilibrated in the PIPES buffer (see table 2) and mounted

on an HPLC system, prior to its analysis in the X-ray beam (David and Pérez, 2009). 60 μL at 4.1 mg/mL, 50 μL at 11.4 mg/mL and 80 μL at 2.1 mg/mL for respectively full-length Grb2, the NSH3SH2 bidomain and the SH2CSH3 bidomain were injected. The X-ray energy was 12 keV ($\lambda = 1.033$ Å). The Eiger 4M pixel detector was used, at a distance of 2 m. Images were measured with 990ms exposure and 10 ms dead time, before void volume for the buffer (180 images) and after void volume for the sample (810 images).

6.2.3 Data analysis

CtErbB2 Data analysis was performed with Dr. Dominique Durand (Biochemistry, Biophysics and Structural Biology Department, I2BC, Orsay). The PRIMUS program (Konarev et al., 2003) was used for the analysis of scattering curves. The P(r) distance distribution was calculated using the GNOM program (Svergun, 1992). The dimensionless Kratky plot calculation was made according to Durand et al. (2010).
Given that CtErbB2 is highly disordered, we tried to fit the scattering curve to a polymer model, the wormlike chain model, using the Sharp-Bloomfield equation (Sharp and Bloomfield, 1968; Pedersen and Schurtenberger, 1996). This model from polymer physics assumes continuous flexibility along the chain, with excluded volume effects. This kind of model had been previously successfully used to analyze the behavior of intrinsically disordered proteins (Receveur-Brechot and Durand, 2012; Obrien et al., 2015). This model assumes:

$$\frac{I(q)}{I(0)} = I_{SB}(q)e^{-q^2R_c^2/2}$$

where

- $I_{SB}(q) = g_D(x) + \frac{b}{L}\left(\frac{4}{15} + \frac{7}{15x} - (\frac{11}{15} + \frac{7}{15x})e^{-x}\right)$
- $g_D(x) = 2\frac{e^{-x}+x-1}{x^2}$ is the Debye function (Pérez et al., 2001)
- $x = q^2Lb/6$
- L is the contour length of the chain
- b is the length of the statistical element
- R_c is the radius of gyration of cross section

The parameters obtained from the SAXS curves and from this model could then be compared to expected values for random coils. The theoretical value of the contour length for a protein is calculated as $L = Nl_0f$

where N is the number of residues, l_0 is the distance between two sequential C_α, and f is a geometrical factor taking into account dihedral angles. For an unstructured protein we took l_0 = 3.78 Å and f = 0.95 as in (Receveur-Brechot and Durand, 2012). The radius of gyration R_g can be estimated using Flory theory giving $R_g = R_0 N^\nu$, where R_0 and ν depend on the behavior of the polymer in solution. While for globular proteins empirical parameters $R_0 = \sqrt{3/5} * 4.75$ and $\nu = 0.29$ have been found (Wilkins et al., 1999), the behavior of IDPs is quite different, giving $R_0 = 2.54$ and $\nu = 0.522$ (Bernadó and Blackledge, 2009).

Grb2 Data analysis was performed by Dr. Aurélien Thureau (SWING beamline, SOLEIL synchrotron, Saint Aubin). The curves were processed using the FOXTROT program (`https://www.synchrotron-soleil.fr/en/beamlines/swing`) developed at the SOLEIL synchrotron. The curves derived from buffer measurement were averaged and subtracted from the protein curves. Then the radius of gyration R_g was calculated using a Guinier plot on the most intense curve and extended to all scattering curves, thus producing a plot of R_g and scattered intensity, I_0, for all images. The curves with a stable R_g value were then averaged to produce the protein scattering curve of scattered intensity as a function of the momentum transfer $q = \frac{4\pi sin(\theta)}{\lambda}$, where 2θ is the diffusion angle and λ is the wavelength. Further analysis (Guinier, Radius of gyration R_g, intensity at origin I_0 calculation, distance distribution function P(r), and MW estimation) were performed with the ATSAS program suite (Franke et al., 2017). The dimensionless Kratky plots were calculated based on Durand et al. (2010). The generation of the theoretical SAXS curves and their comparison with the experimental ones were done using Crysol 2.8.2 (Svergun et al., 1995).
For constructs for which the SAXS curves derived from the crystal structure did not fit the experimental data, we wanted to have an idea of the type of rearrangement of the domains that would fit the data better. EOM is a program allowing the generation of statistical ensembles of proteins with flexible parts, and select sub-ensembles that fit the SAXS data. EOM 2.0 (Tria et al., 2015; Bernadó et al., 2007) was used to generate 10,000 structures based on the crystal structure (PDB 1GRI) as the template. The 2 N-terminal residues from the experimental constructs, and flexible parts missing in the crystal were added to the template. Three structural pools were generated for each combination (either full-length Grb2 or NSH3SH2, or SH2CSH3). The 2 linkers (residues 57-65, 154-160) and the C-terminal part (residues 215-219) were considered as flexible regions. Five runs of 100 independent cycles of the EOM2.0 Genetic Algorithm (Gajoe) were run on each structural pool to estimate the variability in the distribution of the R_g values.

6.3 Generation of conformational ensembles: Flexible Meccano

We used Flexible Meccano (Ozenne et al., 2012) to generate conformational ensembles of CtErbB2. Flexible Meccano generates multiple random conformations of a given protein chain using a minimization algorithm. It takes into account residue-specific dihedral angles distribution restraints (from a X-ray crystallography database) and volume exclusion effects. We used Flexible Meccano to generate two ensembles. The number of conformations (10 000 for each ensemble) was set after testing different sizes and verifying convergence.

- First ensemble: a random coil ensemble with no local or long-range constraints.
- Second ensemble: an ensemble with constrained secondary structure propensities derived from the SSP scores and slightly adjusted to best fit the RDC data. In this ensemble, the α-helix was implemented between residues 15 and 20, with a 20% propensity. PPII helices were constrained using the dihedral angles (-78°, 149°), with propensities of 20% (residues 76-93, 178-191, 216-226 and 238-261), 30% (residues 111-122, 141-151 and 168-175) or 40% (residues 156-165).

6.4 Sequence alignments and distance trees of mammalian CtErbB2

This part of the work was done with Dr. François Bontems (ICSN, Gif-sur-Yvette). 3 iterations of PSI BLAST (`blast.ncbi.nlm.nih.gov`) were used to collect 500 sequences related to human CtErbB2. After each iteration, non-ErbB2, low-quality or fragment sequences were removed. After the final iteration, only one isoform per organism was chosen if the isoforms were identical. A first distance tree was built based on pairwise alignment with BLAST. This tree was used to select mammalian sequences on which a multiple sequence alignment was performed using Clustal Omega (`www.ebi.ac.uk/Tools/msa/`). Finally, this alignment was used as input to built the final tree.

RESULTS

7 Characterization of isolated, unphosphorylated CtErbB2

7.1 CtErbB2 is an IDP

7.1.1 CtErbB2 is predicted to be disordered from its sequence

We first sought to confirm that, as suggested by other works, CtErbB2 is indeed disordered. If we look at CtErbB2 sequence, the amino acid composition (presented in figure 44a) is not very much depleted in hydrophobic residues. It is acidic, with a theoretical pI of 4.29 (as calculated by Protparam `https://web.expasy.org/protparam/`, pI of 4.33 in our construct with the 4 GSHM N-terminal residues). Such pI is not unusual, even for globular proteins. Consequently, CtErbB2 is not clearly displaced towards predicted IDPs in the charge/hydropathy plot (see figure 44b). However, its very high content in prolines (44 residues out of 268) is expected to prevent formation of stable structure. The overall disorder prediction, as given by IUPRED and shown in figure 44c, is biased towards disorder all along the sequence of CtErbB2, with order propensity approaching 50% only in three regions (roughly around residues 10 to 50, 190 to 210 and 225 to 235).

7.1.2 Experimental evidence of overall absence of folding

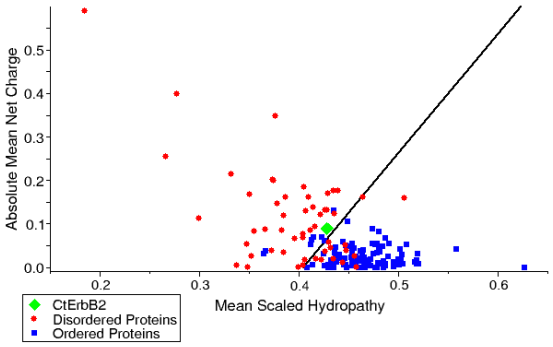
We combined different experimental approaches that were able to unambiguously show that CtErbB2 has low levels of secondary and tertiary structure.

During the purification process, one feature was readily notable: the SDS-PAGE gels of CtErbB2 showed a band at molecular weights of about 45 kDa (see figure 35b), whereas the expected mass of CtErbB2 is 28 kDa, and was confirmed by MALDI-TOF mass spectrometry (data not shown). This is a common feature of proteins containing high levels of flexibility, which slows down their electrophoretic migration.

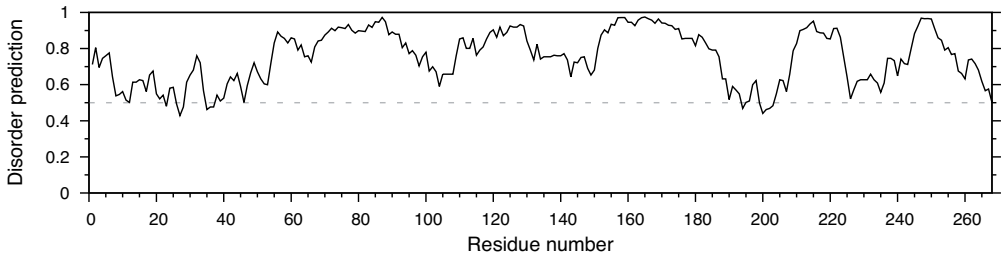
One of the first methods that was historically used to show the level of structure of a protein, and more precisely secondary structure, is circular dichroism. As seen in figure 45a, the CD spectrum of CtErbB2 has

Amino acid class	Average % (UniprotKB/Swiss-Prot)	% in CtErbB2
Hydrophobic (A, F, I, L, M, V, W, Y)	41.0	40.0
Negatively charged (D, E)	12.2	14.9
Positively charged (H, K, R)	13.6	7.8
Polar (N, Q, S, T)	19.9	18.7
Proline	4.7	16.4

(a)



(b)



(c)

Figure 44 – CtErbB2 (ErbB2 residues 988-1255) is predicted to be disordered from its sequence (a) Amino acid composition of ErbB2 compared to the average amino acid composition in UniprotKB/Swiss-Prot (from ProtScale `https://web.expasy.org/protscale/pscale/A.A.Swiss-Prot.html` *). (b) Charge/Hydropathy plot of CtErbB2 (calculted with PONDR* `http://www.pondr.com/` *). (c) IUPRED2a (*`https://iupred2a.elte.hu/`*) prediction for CtErbB2 (with the long disorder option)*

a strong negative band around 200 nm, and no sign of a negative band at 222 nm (indicative of α helices) or a positive band at 195 nm (indicative of β-sheets). This confirms the absence of significant levels of secondary folding of CtErbB2. SAXS gives more information on the global shape of the protein in solution, and the Kratky plot especially is a good visualization of the degree of globularity of a protein. In figure 45c are presented the SAXS scattering curve and the corresponding Kratky plot and distance distribution for CtErbB2. As expected for a non-globular, extended chain, the Kratky plot shows an increase at high qR_g values, and the P(r) is non symmetrical and extends towards high distances. We therefore tried to fit the scattering curve with a polymer model, the Sharp-Bloomfield equation, given in section 6.2.3. This model is a wormlike chain model that takes into account excluded volume effects (Sharp and Bloomfield, 1968). The fit is shown in red in figure 45c. The fit is very good, and we were able to extract several parameters: the length of the statistical element, about 20 Å, which is consistent with the value of 7 residues for a random coil found by Schwalbe et al. (1997); the radius of gyration of cross section, $R_C = 2.7$ Å, that gives the thickness of the chain; and the contour length L = 770 Å , which gives the length that we would obtain by

extending the chain without breaking secondary structure. This contour length is a little shorter than what would be expected for a random coil with no secondary structure (around 980 Å, see section 6.2 for details of the calculation), indicating that there is some (low) level of local structure in CtErbB2.

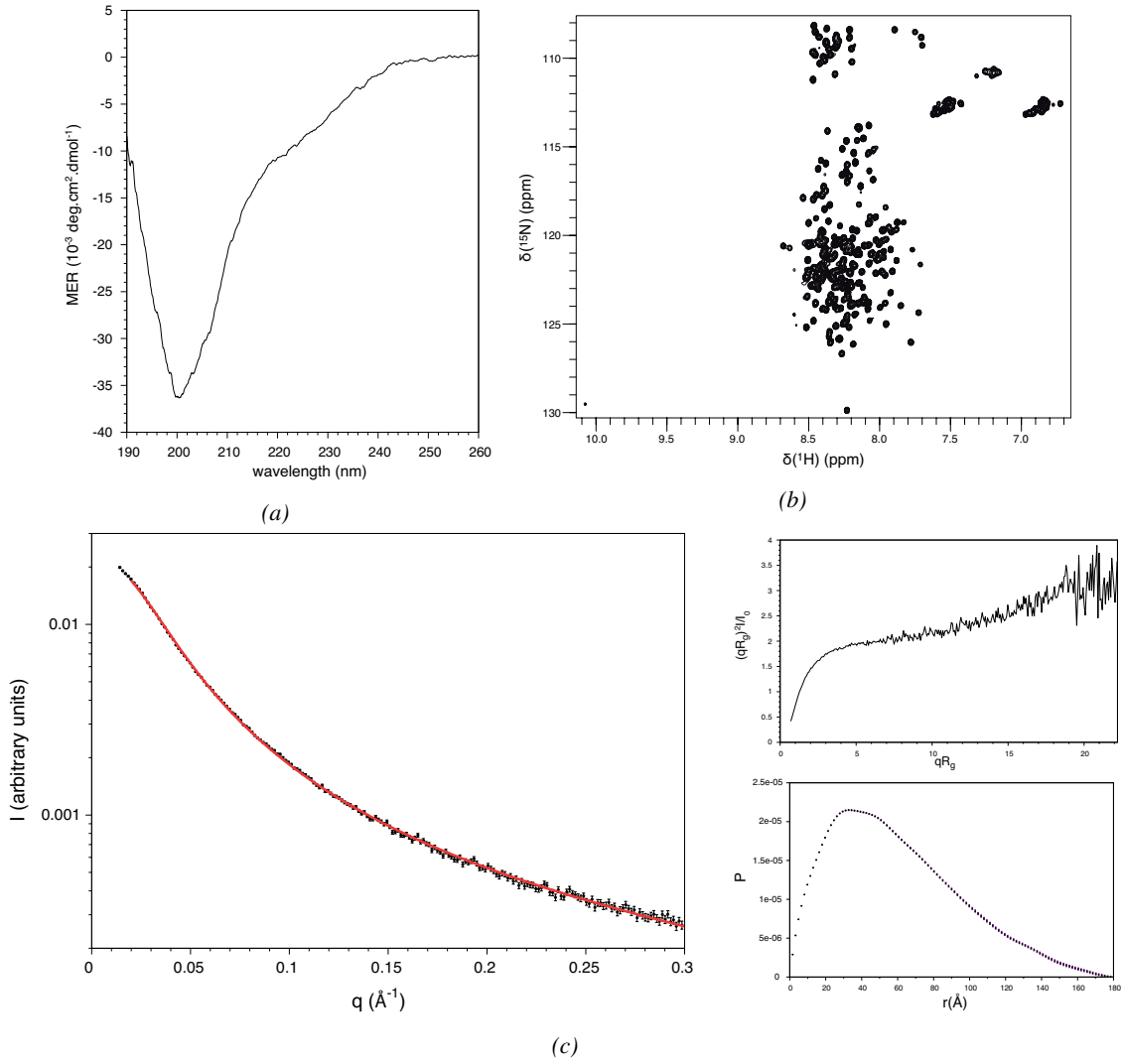


Figure 45 – CtErbB2 is an IDR. (a) Circular dichroism spectrum of CtErbB2. (b) HSQC spectrum of CtErbB2 at 950 MHz. (c) SAXS experimental (black) curve and model (red) using the Sharp-Bloomfield equation. Small panels: corresponding Kratky plot and distance distribution (P(r)) of CtErbB2.

These CD and SAXS data on CtErbB2 can be complemented by NMR studies at atomic resolution. Even without assignment, the ^{1}H-^{15}N HSQC spectrum of CtErbB2 (figure 45b) shows that all amide peaks are clustered in the central region of the proton dimension, with dispersion as low as 1 ppm. This indicates high levels of disorder. The only structure compatible with such spectra would be α helices, which also give

rise to peaks in this region, but stable helices were not seen on the CD spectrum.
Overall, these combined approaches prove that CtErbB2 is not stably structured in solution and is indeed an intrinsically disordered protein.

7.2 NMR assignment of backbone and side chains

We published the assignment of CtErbB2 in *Journal of Biomolecular NMR Assignment* (Wang et al., 2018), and the publication is reproduced in the appendix II. The corresponding BMRB entry is 26740. The details of the assignment strategy are given in the experimental section 5.1.2.
Almost 99% of backbone atoms of non proline residues were assigned: missing backbone ^{1}H-^{15}N are those of -3G, and -2S; all $^{13}C_{\alpha}$ and C' are assigned except for those of -3G; 96% of $^{1}H_{\alpha}$ were assigned. Among the 44 prolines, 93% had their ^{15}N assigned, as well as all but one $^{1}H_{\alpha}$ and all $^{13}C_{\alpha}$.
75% of side chain protons and 68% of side chain carbons (including $^{13}C_{\beta}$) of CtErbB2 were successfully assigned. High degree of overlap prevented more extensive assignment. More detailed information on the results of the assignment are given in table 3, and an assigned ^{1}H-^{15}N HSQC spectrum of CtErbB2 can be found in the assignment publication in appendix II.

Atom type	**Non-proline**		**Prolines**		**Total ratio assigned**
	not assigned	ratio assigned	not assigned	ratio assigned	
Backbone amide N	-3G, -2S	226/228	150P, 241P, 188P	41/44	267/272
Backbone amide H	-2S	226/227	X	X	226/227
$^{13}C_{\alpha}$	-3G	227/228	none	44/44	271/272
$^{1}H_{\alpha}$	-3G, 155Q	254/257	240P	43/44	297/301
C'	-3G	227/228	222P, 223P, 240P, 246P	40/44	267/272
$^{13}C_{\beta}$	none	199/199	162P, 222P, 223P, 240P, 246P	39/44	238/243

Table 3 – Details of the assignment of CtErbB2 backbone and $^{13}C_{\beta}$. In the side-chains (including $^{13}C_{\beta}$), 75% of protons and 68% of carbons were assigned.

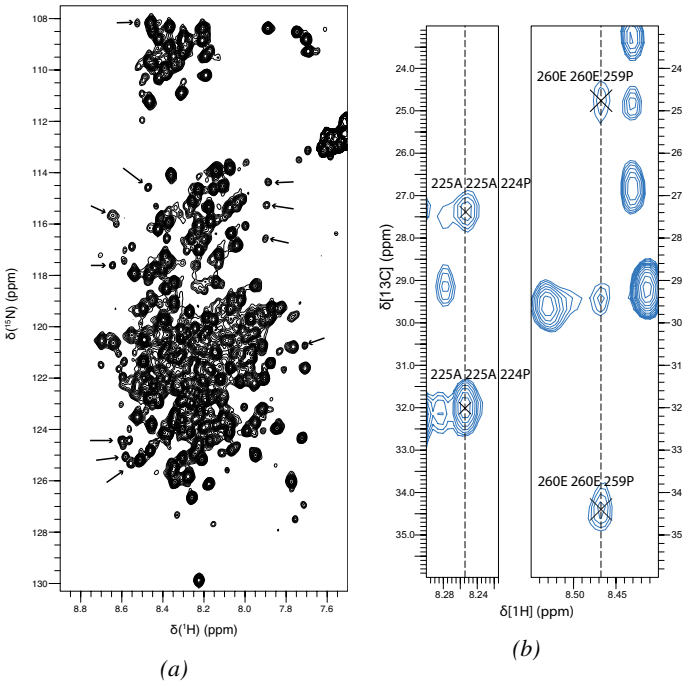
7.3 Characterization of proline conformation

In such a proline-rich protein as CtErbB2, one would expect that many forms of the protein coexist, each being characterized by a different set of proline isomers, exchanging in the slow regime. Here it could give $2^{44} = 1.7592186.10^{13}$ possibilities. However, the HSQC spectrum of CtErbB2 recorded in "usual" conditions (around 200 μM protein, 8 scans, 298 K, 950 MHz) shows a single visible form. Only by increasing the concentration samples (to about 400-500 μM) and/or the number of scans could we see minor peaks emerging from the noise level. Such a spectrum is shown in figure 46a. These minor peaks are in majority

too poorly populated to give enough signal in the 3D assignment spectra, and therefore were not assigned. Moreover, in such a disordered protein, the effect of proline isomerization is expected to impact the chemical shifts only for the neighboring residues, explaining that even if the dispersion of proline isomeric states is high, many peaks of different states will superimpose into a so-called "major" peak depending only on the isomeric state of very close prolines in the sequence.

We used the $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ chemical shifts (determined thanks to the (H)CC(CO)NH spectrum, shown in figure 46b) observed in the major form to see what is the predominant conformation of each proline of CtErbB2. Indeed, X-proline peptide bonds (X being any amino acid) in the *trans* conformation are expected to exhibit proline [$^{13}C_{\beta}$,$^{13}C_{\gamma}$] chemical shift couples in the range of [31.8 ± 1.0, 27.4 ± 0.9] ppm, while those of *cis* prolines should be around [33.8 ± 1.2, 24.4 ± 0.7] ppm (Shen and Bax, 2010). Results are given in figure 46c. For all but two of the assigned prolines, the X-proline bond is *trans*. For 148P and 159P, the form that was found and assigned is *cis*. The intensities of the corresponding peaks (the following residue in the HSQC and the carbon peaks in the 3D spectra) are lower than for the other prolines, but higher than the minor forms detected. This indicates that the detected *cis* form, even though significantly populated, may not be highly predominant. A *trans* form peak might also exist, but could not be found, probably due to peak overlapping. More specifically, we suspect (from spectra at different temperature or pH) that 149Q peak in the ^{1}H-^{15}N HSQC corresponding to the 148P *trans* form might be overlapped with 219Q, since the sequence around these residues is similar (PQP in both cases); similarly 260E could be superimposed with 208E or 151E (PEY sequence). However, this could not be confirmed.

Overall, our results are compatible with a recent publication stating that the *trans* conformation is expected to be populated at about 90% or more at each site in such an unfolded chain (Alderson et al., 2018). The question that we can then ask is why, in two defined positions in CtErbB2 sequence, the *cis* population is higher than expected. Asparagine and serine residues preceding prolines (as it it the case for these two prolines) do not seem to particularly promote the *cis* isomer in small peptides compared to other X amino acids in the XP sequence (Reimer et al., 1998). Aromatic residues preceding prolines do favor the *cis* isomer, but no such sequence is found in CtErbB2. The significant *cis* populations that we observe here are therefore probably due to other sequence or conformational effects in CtErbB2, consistent with the fact that other NP and SP sequences are present elsewhere and exhibit major *trans* conformations. We could not find any obvious determinant of this particular behavior.



(a) *(b)*

Proline number	$\delta^{13}C_\beta$ (ppm)	$\delta^{13}C_\gamma$ (ppm)	Deduced isomer
9	32.099 ± 0.054	27.091 ± 0.053	*trans*
12	32.094 ± 0.029	27.246 ± 0.008	*trans*
39	32.138 ± 0.007	27.354 ± 0.016	*trans*
46	32.144 ± 0.030	27.050 ± 0.020	*trans*
48	32.172 ± 0.077	27.148 ± 0.000	*trans*
50	31.891 ± 0.010	27.439 ± 0.000	*trans*
78	32.139 ± 0.022	27.413 ± 0.000	*trans*
84	32.053 ± 0.046	27.440 ± 0.000	*trans*
87	32.065 ± 0.045	27.316 ± 0.039	*trans*
90	32.111 ± 0.010	27.436 ± 0.000	*trans*
115	32.067 ± 0.013	27.435 ± 0.000	*trans*
119	32.116 ± 0.011	27.127 ± 0.000	*trans*
121	32.112 ± 0.022	27.240 ± 0.000	*trans*
129	32.141 ± 0.016	27.029 ± 0.000	*trans*
132	32.134 ± 0.029	27.336 ± 0.000	*trans*
134	32.008 ± 0.020	27.457 ± 0.000	*trans*
143	32.123 ± 0.005	27.344 ± 0.000	*trans*
148 (weak)	34.359 ± 0.030	ND	*cis*
150	32.125 ± 0.048	27.449 ± 0.000	*trans*
156	32.124 ± 0.005	27.276 ± 0.000	*trans*
160	32.096 ± 0.013	27.363 ± 0.000	*trans*
162	ND	ND	ND
163	32.047 ± 0.008	27.420 ± 0.000	*trans*
165	32.114 ± 0.020	27.352 ± 0.000	*trans*
169	32.189 ± 0.028	27.144 ± 0.000	*trans*
171	32.029 ± 0.025	27.419 ± 0.000	*trans*
175	32.106 ± 0.015	27.438 ± 0.000	*trans*
183	32.210 ± 0.029	27.433 ± 0.000	*trans*
188	32.186 ± 0.017	27.280 ± 0.000	*trans*
207	32.097 ± 0.006	27.163 ± 0.000	*trans*
212	32.045 ± 0.057	27.470 ± 0.000	*trans*
218	32.089 ± 0.015	27.342 ± 0.000	*trans*
220	32.157 ± 0.013	27.301 ± 0.000	*trans*
222	ND	ND	ND
223	ND	ND	ND
224	31.995 ± 0.028	27.365 ± 0.000	*trans*
228	32.001 ± 0.016	27.234 ± 0.000	*trans*
240	ND	ND	ND
241	31.969 ± 0.035	27.318 ± 0.000	*trans*
246	ND	ND	ND
247	32.100 ± 0.010	27.433 ± 0.000	*trans*
254	32.143 ± 0.008	27.447 ± 0.000	*trans*
259 (weak)	34.379 ± 0.017	24.764 ± 0.000	*cis*
267	31.973 ± 0.025	27.329 ± 0.000	*trans*

(c)

Figure 46 – Proline isomeric states in CtErbB2. (a) Proline $^{13}C_\beta$ and $^{13}C_\gamma$ chemical shifts of CtErbB2. Standard deviations calculated from the values in the different recorded spectra are given (0.000 when the chemical shift was only measured in one spectrum). For residues 148P and 259P, we suspect that the trans conformation also exists but is superimposed with another peak, preventing assignment. (b) Slices of the (H)CC(CO)NH spectrum showing prolines $^{13}C_\beta$ and $^{13}C_\gamma$ peaks for a trans and a cis isomer. (c) ^{1}H-^{15}N HSQC spectrum of CtErbB2 at 400µM, 8 scans, in the MES buffer pH 5.6. Arrows indicate some examples of minor peaks suspected to correspond to residues following minor isomeric states of prolines.

7.4 Characterization of transient local structure

Secondary structures are characterized by specific dihedral angles that can be probed using chemical shifts and scalar couplings. Moreover, different secondary structures have different orientations of their backbone NH bond, allowing to use NH RDCs as another probe. We used these three parameters to look for secondary structure elements in CtErbB2, and they are presented in figure 47. From the CD data, which do not show any strong band for α helices or β-sheets, we expect such secondary structures, if they exist, to be transient. Do to the absence of dispersed peaks in CtErbB2 HSQC, we do not expect to observe β-strands. MoRFs are

often made of preformed α helices (Fuxreiter et al., 2004), and proline-rich disordered segments are likely to get partially structured in polyproline (PPII) helices. We thus expect to find those two types of structures in CtErbB2.

We firstly used the SSP score by Forman-Kay's lab, that has the advantages of giving roughly quantitative secondary structure propensities (see figure 47). All along CtErbB2 sequence, no secondary structure more populated than about 30% can be detected. Overall, SSP scores alternate regularly between low positive (compact helices) and negative (extended structures) scores. Looking closely, the positive regions are N-capped by serines or aspartates (^{14}DSTFYRSLL^{22}, ^{71}DLTLGL^{75}, ^{100}DGDLGMG^{105}, ^{135}SETDGY^{140}, ^{231}DNLYY^{235}), which is consistent with the formation of short capped α helices. A similar behavior was already observed for the measles virus nucleoprotein N_{TAIL} disordered region (Jensen et al., 2011). The regions exhibiting negative SSP scores are less straightforward to interpret. They indicate extended regions, but chemical shifts do not allow one to discriminate β-strands from PPII helices. However, these extended regions correspond to regions of higher proline density (see blue vertical bars in figure 47), pointing to PPII helices. To confirm, and describe in more details, the most populated secondary structures, we used residual dipolar couplings, as well as scalar couplings, comparing them to values expected for random coils or more sophisticated conformational ensembles generated by Flexible Meccano (Ozenne et al., 2012).

- The longer and most populated α-helix in CtErbB2 is located in its N-terminus and has the sequence ^{14}DSTFYRSLL^{22}. It is estimated to be populated in about 25% of conformational states according to the SSP scores, and consistently correlates with a region of slightly positive $^{1}D_{NH}$ and low $^{3}J_{HNH\alpha}$ compared to expected random coil values (see figure 47).

- The extended stretches identified by the SSP scores, that are likely to be PPII helices, indeed correspond to $^{3}J_{HNH\alpha}$ values close to the random coil value. If they had been β-strands, we would have expected higher values (around 9 Hz, depending on the amino acid type). Using Flexible Meccano, we generated constrained conformational ensembles, in which secondary structures propensities were set to the SSP values. We then slightly adjusted these values to fit the RDC data. Only slight adjustment was necessary, showing that chemical shifts and RDCs are both consistent with PPII stretches defined with dihedral angles (-78°, 149°)) at propensities of 20% (residues 76-93, 178-191, 216-226 and 238-261), 30% (residues 111-122, 141-151 and 168-175) and 40% (residues 156-165). These transient PPII helices are compatible with the major *trans* conformation of prolines that we have determined.

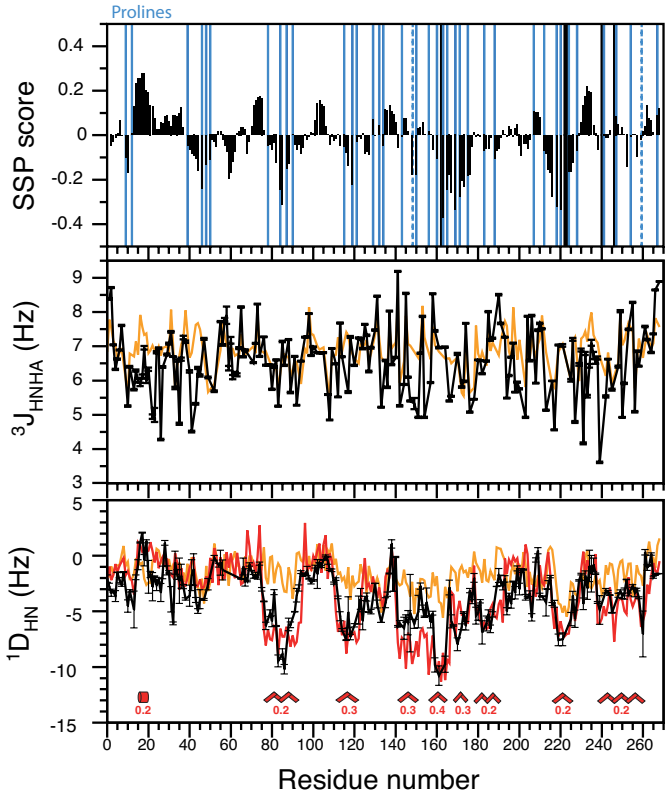


Figure 47 – Probing transient secondary structures of CtErbB2. (top) Secondary structure propensities calculated from combined chemical shifts (SSP score (Marsh et al., 2006)). (middle) Experimental (black) ^1H_N-$^1H_\alpha$ scalar couplings of CtErbB2 compared with scalar couplings expected from a random coil conformation of this sequence (Shen et al., 2018) (orange). (bottom) Backbone ^1H-^{15}N residual dipolar couplings (RDCs) of CtErbB2 measured in stretched acrylamide gels. The experimental RDCs (black) are compared with RDCs calculated for the random coil (orange) and the secondary-structure containing (red) Flexible Meccano ensembles. Corresponding secondary structures are indicated with a cylinder (α-helix) and with zig-zags (PPII helices), with the associated propensities.

7.5 Dynamics of CtErbB2

^{15}N R_1 and R_2 rates, as well as {^1H}-^{15}N nOes were measured for CtErbB2 at 14.1 and 22.3 T. It is noteworthy that obtaining reproducible relaxation data, and comparable data at different magnetic fields, is not an easy thing to do when working with IDPs: even more than for structured proteins, dynamics greatly depends on the conditions, and in particular very slight changes in pH or temperature can cause large variations. Great care was therefore taken in the control these parameters, to yield consistent data. The results are given in figure 48.

The first striking feature is that the relaxation parameters do not exhibit bell-shape profiles as expected from a random coil, and the dynamic behavior of the chain is highly heterogeneous. Four types of behavior

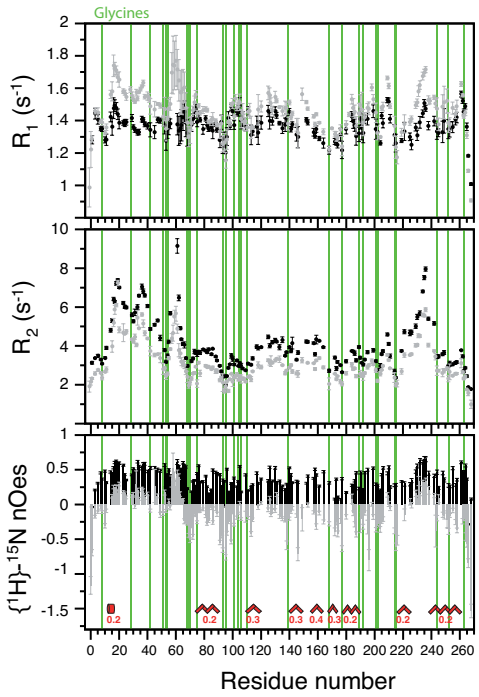


Figure 48 – ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and {^{1}H}-^{15}N nOes of CtErbB2, measured at 14.1 T (600 MHz proton resonance frequency, grey) and 22.3 T (950 MHz proton resonance frequency, black). The positions of glycines are indicated by green vertical bar.

can be distinguished. We analyzed them in the light of the expected relaxation parameters for ^{1}H-^{15}N bonds in segments exhibiting a single rotational correlation time τ_C, with small libration of the bond (see figure 39).

- A random-like, very dynamic behavior in the termini of the protein, and in the central, glycine-rich region (residues 65 to 120). With R_1 almost identical at the two magnetic fields, around 1.4 s^{-1}, and nOes of opposite signs (close to -0.5 at 14.1 T), the corresponding τ_C can be estimated to be around 0.7 ns.

- A rigid, histidine-rich region (residues 56-61). High R_2, R_1 higher at 14.1 than at 22.3 T, and nOes around 0.5 at the two fields point to a τ_C of about 4 ns, very high for a disordered protein. The histidine side-chains probably restrain motion in this region.

- A more complex dynamic behavior in the regions 15-49 and 225-244, which do not fit the single-τ_C model, indicating a mixture of flexible and restricted motions. The transient α-helix explains this behavior for region 15-22, but regions 23-49 and 225-244 do not exhibit significant secondary structure propensities. A possible explanation would be the presence of transient tertiary structure

(long-range contact).

- A close to random-coil behavior, slightly more restricted than the glycine-rich region, in the rest of the protein, corresponding to a τ_C around 1 ns. These regions contain the PPII helices, which do not exhibit sign of significantly restricted motion. This is consistent with the fact that PPII helices are more dynamic than "canonical", hydrogen-bonded secondary structures. Furthermore, they are not really longer than the length of the statistical elements of a random coil determined by Schwalbe et al. (1997), and are therefore not expected to change the segmental motion of the chain.

7.6 Long-range contact(s) in CtErbB2

A long disordered chain such as CtErbB2 statistically samples a wide conformational space. This is one of the advantages of IDPs, sometimes called the fly-casting mechanism, and increases encounter probabilities with partners.

We first used the SAXS data to see if we could detect any global compaction of the protein compared to a random coil. The radius of gyration determined via the P(r) distribution gave $R_g = 49.2$ Å. This is very different from the R_g expected for a globular protein of that size (18.7 Å) but consistent with a random coil (47.4 Å), both calculated using Flory theory (see section 6.2). However, when we compare the experimental value with the values calculated from our Flexible Meccano ensembles, we see that R_g should be even larger taking into account the secondary structures, and especially the PPII helices that extent the protein: indeed, the random coil ensemble generated by Flexible Meccano yields a R_g of 50.2 Å, while the ensemble containing secondary structure yields $R_g = 57.0$ Å. The expected, but not observed, increased R_g compared to a random coil due to the presence of the transient secondary structures may be compensated by long-range tertiary interactions.

We then tried to see whether certain regions experience specific contacts during this sampling in CtErbB2. To probe this at residue resolution, the dedicated technique is paramagnetic relaxation enhancement (PRE). The principle of this technique is exposed in section 5.6. We first coupled a paramagnetic MTSL probe at three positions in three distinct experiments to sample the extremities and the central part of CtErbB2 chain: at the two native cysteine positions (45 and 146) and at position 248. After preliminary experiments showing effects around residue 230, we added position 227. The results are presented in figure 49. A clear decrease in intensity is observed in the region around residue 230 when the probe is located on residue 45, and the reciprocal effect is observed when the probe is on residue 248, and even more strongly on residue

227. A relaxation enhancement is also observed around residue 230 when the probe is on position 146, but the effect is small and non-reciprocal, and therefore likely artifactual.

SAXS and PRE both indicate the presence of at least one long-range contact, located between the N to the C-terminus of protein. This contact is likely to be quite transient/lowly populated, since the decrease in R_g and the PRE are small.

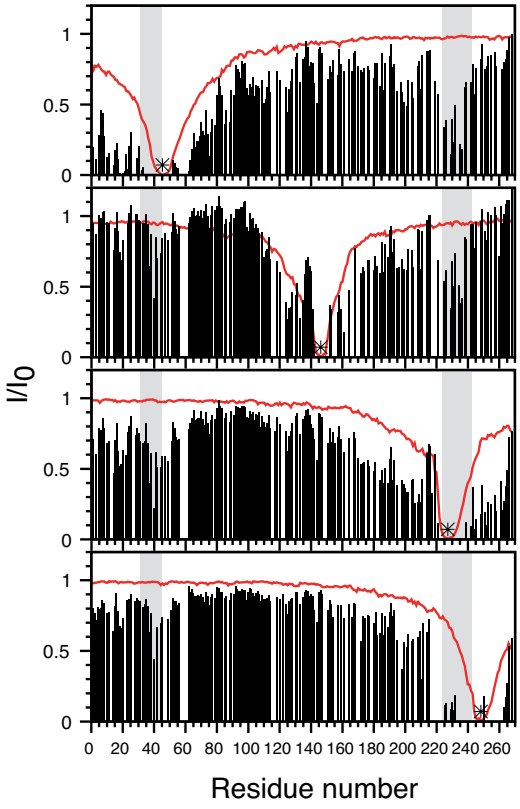


Figure 49 – Presence of a long-range contact in CtErbB2. Paramagnetic relaxation enhancement (PRE) profiles of CtErbB2 (black) compared to the profile expected from the Flexible Meccano ensemble including only secondary, but not tertiary, structures (red). Four experiments, each with MTSL coupled at a different cysteine position marked by an asterisk (45, 146, 227 and 248) have been performed. The two regions for which we observe significant and reciprocal effect are shown with grey background.

7.7 Conservation of CtErbB2 features amongst mammals

Usually, sequence conservation in intrinsically disordered proteins is much lower than for globular proteins, many constraints for structure conservation being released. However, we noticed a very high degree of conservation of CtErbB2 amongst different species. We therefore tried to look at sequences conservation of the structural and dynamic features observed here. We chose to focus on mammals, where all four ErbBs are

present as in humans. Since we do not expect strict sequence conservation for all regions, but for example region-specific conservation of proline density, we did not perform a global alignment but rather had a look at node-by-node conservation after building a tree of mammalian sequences with PSI-BLAST. The resulting alignments of the 21 nodes are presented in the Supplementary figures S2 of the publication submitted to JBC and given in appendix III, with the corresponding tree (figure S3) and node-by-node alignment (figure S4). The first striking feature of the alignment is the strict conservation of the 5 autophosphorylation sites (6 autophosphorylated tyrosines). But surprisingly, the conservation level remains very high even outside of these regions: compared to the sequence for *Homos sapiens*, one third of the sequence is strictly conserved and 15% more has a conservation score given by Clustal Omega higher than 0.5. This high conservation is also found in EGFR (Kovacs et al., 2015a). This indicates that the role of the other residues in the C-terminal tail of ErbB receptor is not random, and is not only driven by the conservation of flexibility. Interestingly, the first 50 residues of CtErbB2 are even more conserved than the rest of the tail: this encompasses the N-terminal α-helix, as well as the first region shown from PRE experiments to experience a transient long-range contact. The second region involved in this contact, around residue 230, is the second best conserved, with the stretch ^{228}PAFDNLYYW^{236} being strictly conserved. The conservation of the PPII helices is a little more complicated to analyze, since strict sequence conservation is not expected to be necessary. Yet, we can notice conservation of the overall proline density, and some mutations of prolines around PxxP motifs compensate each other thus leading to conservation of the motif (PxxP motifs 6 and 7 in figure S2 of appendix III).

7.8 Our study reveals new potential mechanisms for the regulation of signal transduction

The goal of our study was to understand the molecular bases for the ability of the C-terminal tail of the receptor ErbB2 to perform its function in signal transduction, starting from an inactive, non phosphorylated form. More specifically, we studied CtErbB2 structural and dynamic characteristics to determine its degree of accessibility and prefolding to interact with partners, both the ErbB kinases in the active dimers for phosphorylation, and adaptor proteins triggering signaling. Even though we here used a construct of the isolated tail, it is a first step towards understanding its crucial role in signal transduction by the full-length receptor. The secondary and tertiary structure elements that we found in CtErbB2 are recapitulated in figure 50. We managed to make a complete description of the local and long-range degree of structure of CtErbB2, by using an approach combining a wide range of NMR data with more global size and shape information given by SAXS. We also built a model conformational ensemble of 10 000 structures recapitulating the secondary

structures of the protein. Both its high degree of intrinsic disorder, its local folding into transient α- and PPII helices that we suspect to be MoRFs, and the presence of long-range contacts are central features determining how such a receptor tail can perform its function.

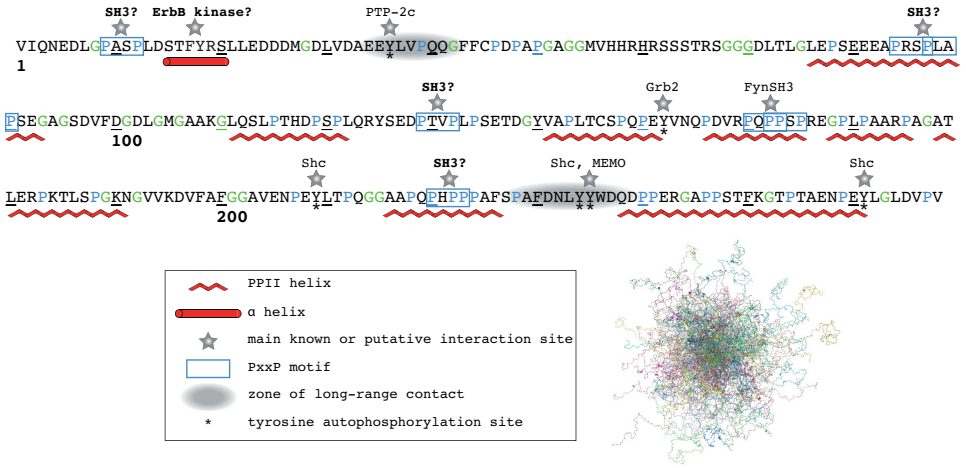


Figure 50 – Summary of our structural description of CtErbB2. In the sequence, glycines are in green and prolines are in blue. The stars indicate the interaction sites, either known (PTP-2c, Grb2, Fyn SH3, Shc, MEMO) (Dankort et al., 1997; Schulze et al., 2005; Marone et al., 2004) or putative (PxxP motifs potentially binding SH3 domains, indicated with a question mark). One hundred structures from the 10 000 structure Flexible Meccano ensemble with secondary structure are shown. The structures were aligned on residues -3 to 20.

Importance of disorder of CtErbB2 for multiple interactions Our study was conducted on a construct containing only the C-terminal tail of ErbB2. One could wonder whether the disorder observed here is still relevant in the context of the full-length receptor. Although the distance to the plasma membrane is too high for significant interaction with CtErbB2, the tail could still interact in an intramolecular fashion with the kinase domain, and one could argue that it could trigger folding. However, the recent study by Keppel et al. (2017) showed by HDX/MS that rates of solvent exchange of the C-terminal tail of ErbB2 are significantly higher than for the globular kinase core in a construct containing the whole ErbB2 cytoplasmic domain, and are consistent with a high degree of disorder. Therefore, observations made here are very likely to remain valid in the intact protein.

Additionally, it is expected that great flexibility of CtErbB2 is required for its function, which relies on accessibility of phosphorylation and interaction sites. The so-called fly-casting mechanism could be a great advantage of disorder in this instance. In the inactive state studied here, the tail needs to be able to sample a wide conformational space to encounter both ErbB kinases in the context of kinase-active dimers, and po-

tential other cytoplasmic partners (cytoplasmic kinases, regulation complexes...). Moreover, if this disorder is conserved in the phosphorylated, active, states, this feature could enable multiple interactions, high recruitment/dissociation rates to adaptor proteins for fast adaptability, all of these being crucial for an efficient signal transduction and regulation.

The N-terminal helix might be a prestructured motif for interaction with ErbB2 kinase and regulation CtErbB2 is likely to make transient contacts with either its own kinase domain, or the kinase of the other monomer of the dimer (in case of *trans*-phosphorylation), or both. The first hypothesis is supported by the two crystal structures of ErbB2 kinase (PDB 3PP0 and 3RCD): in both, the short segment of the C-terminal tail included in the construct used for crystallization retains long disordered segments but one helix interacts with the kinase core, and is located in our numbering ^{16}TFYRS^{20}. This corresponds to the center of the α-helix we have identified to be preformed in solution. Therefore, this helix can be considered to be a MoRF, and its transient folding in the free state could participate in kinase-binding affinity by reducing the binding entropic cost. This interaction could play a role in the regulation of the kinase activity of ErbB2, as it was suggested for EGFR (Walton et al., 1990; Sorkin et al., 1992; Alvarez et al., 1995; Cheng and Koland, 1996; Pines et al., 2010). Besides, it brings the whole C-terminal tail closer to the kinase domain, in a position where it would also be close to the kinase domain of the dimerization partner. The helix contains a tyrosine, which is not one of the 5 *trans*-autophosphorylation sites, but could also be phosphorylated by other kinases. It is easy to imagine that the introduction of a phosphate group could destabilize the helix and/or the interaction with the kinase domain.
The N-terminal helix could therefore be important for kinase activity regulation and phosphorylation, creating feedback mechanisms that are necessary for such important signaling pathways.

PPII helices carry SH3/WW interaction motifs Polyproline motifs, especially those structured in PPII helices, are known to be involved in protein-protein interactions with two types of folded domains: SH3 an WW domains, which are known to be abundant in receptor tyrosine kinase signaling pathways (Pawson, 1992). WW domains are sometimes considered as the phosphorylation-dependent equivalent of SH3 domains (Macias et al., 2002). The PPII helices found in CtErbB2 are short and transient. They are not especially rigid in the ns timescale, and are populated to a maximum of 40%. They therefore allow to keep a high degree of flexibility, while reducing the entropic cost of potential folding upon interaction.

While the consensus motif for interaction with SH3 domains is PxxP, with the two main ligand classes K/RxxPxxP (class I) and PxxPxR/K (class II), WW ligands include PPxY motifs (group I, with an unphosphorylated tyrosine), PPLP (group II), PPR or other proline-rich and K/R containing sequences (group III, potentially in competition for binding with SH3), and phospho-SP or phospho-TP (group IV) (Macias et al., 2002).

Three PPII stretches contain PxxP motifs: ^{81}EEA**P**RS**P**LA**P**SEGA^{94} (2 sequential motifs sharing one proline), ^{156}PDVR**P**Q**PP**S**P**REGP^{169} (two nested motifs, RxxPxxP and PxxPxxR), and ^{216}AAPQ**P**HP**P**PAFS^{227}. Two PxxP motifs are found outside of PPII regions: ^{5}EDLG**P**AS**P**LDST^{16} and ^{125}YSED**P**TV**P**LPSE^{136}. Amongst all these PxxP motifs, only one has a sequence recognized as canonical for binding to SH3 domains: the ^{159}RxxPxxP^{165} motif (class I motif). The ^{159}RPQPPSPRE^{167} peptide it is embedded in was shown in a previous study by Bornet et al. (2014) to fold in a PPII helix, and the RxxPxxP motif is responsible for interaction with the Src family kinase Fyn SH3 domain.

No consensus WW-binding PPxY, PPLP or PPR motifs are found in CtErbB2. However, seven SP motifs (residues 11-12, 86-87, 120-121, 147-148, 164-165, 187-188 and 227-228), and two TP motifs (residues 211-212 ans 253-254) exist, and could constitute, upon S/T phosphorylation, group III binding sites for WW domains. Although no direct interaction between a WW domain and ErbB2 has been observed, WW-containing proteins such as Pin1 and WWP1 have been shown to have a role in regulation of ErbB2-dependent pathways (Wulf et al., 2004; Chen et al., 2008).

In the current model of signal transduction initiation by ErbB receptors, the link between ErbBs and adaptor proteins is ensured by phosphotyrosines interacting with SH2/PTB/MEMO domains (Yarden and Sliwkowski, 2001; Schulze et al., 2005; Feracci et al., 2011). However, it is not excluded that other types of interactions involve ErbB2, for example polyproline/SH3 or polyproline/WW interactions. For example, Grb2 SH3 domains have been suggested to participate in the ErbB2/Grb2 interaction that is known to happen via pY/SH2 contacts (Xie et al., 1995). All in all, PPII-structured polyproline motifs expand the number of possible molecular recognition features of CtErbB2, with potential recognition by SH3 or WW domains.

Probable interplay between proline isomerization and S/T phosphorylation Out of the 44 prolines of CtErbB2, 7 are preceded by a serine, and 2 by a threonine. It is to be expected that S/T phosphorylation by proline-directed kinases, and isomerization by the prolyl isomerase Pin1, that recognizes pSP and pTP

sequences, could play a role in regulation of CtErbB2 conformation. Pin1 has been shown to be involved in ErbB2 regulation (Khanal et al., 2010; Lam et al., 2008), but the interaction has never been shown to be direct. A modification of the *cis*/*trans* ratio of prolines can have dramatic consequences in terms of overall conformation, changing the orientation of the chain, and local structure, since increased ratios of *cis* prolines can for example disrupt PPII helices, that require an all-*trans* conformation.

Long-range contacts as a mechanism for regulation of accessibility Our data show absence of a stable tertiary fold of CtErbB2, with the absence of long-range nOe cross-peaks on nOe-HSQC spectra (data not shown), moderate PRE effects, and overall shape parameters measured by SAXS compatible with an extended structure and very far from the behavior of a globular protein. Long-range contacts with long lifetime should also be detected on the RDC profile, as observed for example on α-synuclein (Bertoncini et al., 2005; Bernadó et al., 2005). However, the effect is difficult to differentiate from effects of local structure if the contact is very transient.

Even with no stable tertiary structure, a close comparison of the radius of gyration of the protein to the R_g expected from a random coil and from the ensemble containing the determined transient secondary structures reveals a small compaction. From our four PRE experiments, we observe a small but reproducible (also observed when using with a 3-(2-iodoamido) proxyl probe, data not shown) and reciprocal effect between the regions around residues 30-45 and 225-242. We have also observed that these two regions undergo partially restricted dynamics as shown by ^{15}N relaxation data. They are a little more hydrophobic than the rest of the protein, and contain some aromatic, possibly favoring π interactions. A small decrease in intensity is also observed in the region of residue 230 when the probe is on position 146, but the effect is not reciprocal (no clear effect around residue 146 when the probe is on position 227). It is possible that the probe "sticks" slightly on the region around residue 230.

The presence of a transient contact between the two ends of the tail of ErbB2 could have two roles:

- Bringing the tyrosine phosphorylation sites closer to the kinase domains in the active dimer. Even very transient contacts can be enough to trigger a first phosphorylation event and the signaling cascade. This effect would be added to the intramolecular interaction of the N-terminal α-helix with the kinase domain that also brings potential phosphorylation sites closer to the kinase.

- Regulating accessibility of the central interaction sites for adaptor protein partners, depending of the

phosphorylation status of the tail

In both cases, this would require that the long-range contacts be phosphorylation-dependent, which is very plausible: the main two sites for which contacts are observed encompass the tyrosine auto-transphosphorylation sites Y_A (Y36) and Y_D (Y235/236). We could imagine that additional charges at these positions could modify the transient contact. The potential functional role of this contact, and its dependence on phosphorylation, is supported by the study of Dankort et al. (1997) that shows that Y_A and Y_D have particular, interdependent role in ErbB2 signaling. When Y_A alone is phosphorylated (all other tyrosines being mutated to phenylalanines), it suppresses the potential of the V664E "activating" ErbB2 mutant to induce transformation of cells. But if Y_A and Y_D are simultaneously able to be phosphorylated, the transformation potency of the protein is recovered, contrary to any other combination between Y_A and another tyrosine phosphorylation site. A way to investigate individual and combined roles of these tyrosine phosphorylations would be particularly interesting, and is presented in section 10.2.4.

Perspectives: phosphorylation and interactions Our description of isolated, unphosphorylated CtErbB2 is a great step towards understanding the bases for the central role of this region of ErbB2 in signal transduction. There are three additional steps to complement its mechanistic characterization, in the context of our work:

- Characterize the effect of tyrosine phosphorylation on the conformational landscape of CtErbB2. Many of the features that we describe in this section are likely to be affected by phosphorylation. We therefore developed different approaches to obtain and study tyrosine phosphorylation, in the particular case of our system. However, this work, presented in section 10, aims at tackling more generally the problem of structural studies of tyrosine phosphorylation, which is still a challenge.

- Characterize the conformational states and phosphorylation process of CtErbB2 in the context of the full-length receptor, and eventually in the context of the homo- or heterodimers. This requires to transfer protein production in eukaryotic cells. Part of this challenge was undertaken, trying to produce the cytoplasmic portion of the receptor in insects cells for phosphorylation, and is also presented in section 10.

- Characterize the interactions of the tail with adaptor proteins, a process which is the heart of signal transduction by ErbB2. This work focuses on Grb2, a partner that is essential for ErbB2-initiated MAP kinase pathway signaling, and that contains domains able to interact with phosphotyrosines

and polyproline motifs. It is therefore particularly interesting both from a biological and mechanistic point of view. Influence of phosphorylation will also be a key aspect of the interaction study. The characterization of the Grb2 protein itself is presented in section 8, while the study of its interaction with CtErbB2 is presented in section 9.

Of course part of these questions, and many other questions, go past the aims of this thesis, or even of this project. S/T phosphorylation probably influences the conformation and interactions of the protein, as well as prolyl isomerization by isomerases; Many partners other than Grb2 could be studied, as well as the interdependence of these interactions...

8 Behavior of Grb2 in solution

Before being able to characterize the interaction between Grb2 and CtErbB2, we had to assign the spectra of free Grb2. We came across some surprises along the way, highlighting the necessity to study the behavior of this modular protein alone. From the literature data on Grb2, presented in section 3, two main questions remain that could affect the interaction with CtErbB2: What is the behavior of each domain relative to one another in the full-length protein in solution? And what is the oligomeric state of Grb2 in solution?

8.1 NMR backbone assignment of all constructs, instability of the NSH3 domain and effect of the Histidine tag

Unless otherwise stated (for the NSH3 domain), all assignments were performed and all the spectra shown were recorded at 308 K, on samples between about 300 and 600 μM in the PIPES buffer at pH 7.2 (40 mM PIPES, 150 mM NaCl, 2 mM TCEP). For the comparison between the His-tagged and tag-free NSH3 domain, the temperature was also 308 K but the buffer was PBS at pH 7.2 with 2 mM TCEP. The reason for this particular buffer for the NSH3 is that it was done by YingHui Wang, a previous post-doctoral researcher in the lab, and the experiments were not repeated in PIPES buffer.

8.1.1 Constructs and assignment

Due to difficult removal of the N-terminal purification histidine tag on the constructs containing the NSH3 domain (NSH3, NSH3SH2 and FL Grb2 constructs), assignments were done on both the tagged and non-tagged constructs. It gave very different spectra:

- For the NSH3: both with and without the tag, two set of peaks can be found and (at least partially) assigned. One exhibits well-dispersed peaks, and one has peaks gathered in the central region of the spectrum in the proton dimension. This SH3 domain was already shown by NMR to experience exchange between a folded and an unfolded form (Goudreau et al., 1994).

- Very surprisingly, an unfolded form of the NSH3 domain is also present in the spectra of the tag-free NSH3SH2 and FL Grb2 constructs, whereas only the fully folded form is visible in the spectra of the His-tagged NSH3SH2 and FL constructs. The His-tag, while destabilizing the NSH3 domain when in isolation, stabilizes it when at least the SH2 domain is present.

Construct	NH assigned	Special conditions
His-NSH3*	47/78	PBS buffer
NSH3*	48/56	(Also assigned in PBS buffer)
SH2	80/97	(Also assigned in PBS buffer)
CSH3	53/60	(Also assigned in PBS buffer)
His-NSH3SH2	118/170	(Also assigned in PBS buffer)
NSH3SH2*	123/148	
SH2CSH3	130/158	
His-Grb2	164/229	(Also assigned in PBS buffer)
Grb2*	173/207	

*Table 4 – Extent of assignment for each Grb2 construct. Our assignment is compared to the number of NH pairs that can be assigned, i.e. without counting the prolines and the N-terminal residue. For the constructs containing the NSH3 for which an unfolded form exists in equilibrium with a folded form (indicated with *), the extent of assignment is given for the folded form.*

The spectra are further analyzed in section 8.1.2, but these characteristics had to be taken into account for assignment. The effect of the His-tag was also investigated for the other constructs to see if it impacted the fold of the other two domains, but in the SH2, CSH3 and SH2CSH3 constructs the chemical shifts differences with and without the tag were restricted to very few N-terminal residues.
The extent of backbone N-H assignment for all the constructs, including the His-tagged ones when the NSH3 in present, is given in table 4. The $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ were assigned to a large extent, but the complete assignment is not finished for all constructs and is therefore not detailed here.

Integrity of the domains was verified for all constructs by comparison of the secondary structure calculated by TALOS+ (based on our chemical shifts) with the consensus fold recapitulated in Uniprot (entry P62993). The verification was done for the folded form of the NSH3 in the NSH3, NSH3SH2 and FL constructs. The extent of assignment for the unfolded forms is not sufficient for their complete characterization. The ^{1}H-^{15}N HSQC spectrum of the assignment of FL Grb2 is shown in figure 51b, 51c, and 51d. Residues for which peaks were not found, as shown in red in figure 51a, are mainly residues of the linkers between the NSH3 and SH2 and between the SH2 and CSH3 domains, and the region around C32 in the NSH3 domain. The linker residues are probably partially flexible, and exchange in the intermediate timescale might results in line broadening. The C32 region was already described as "poorly defined" by Goudreau et al. (1994). Its electron density is not defined in the crystal structure of the full-length Grb2 (PDB 1GRI) but is defined in the N-Src loop in the NMR peptide-bound structures of the isolated NSH3 domain (PDB 1GBR and 1GBQ), as shown if figure 29a. Once again, partial flexibility is probably at stake. Overall, spectra quality was better

at lower fields (600 MHz) compared to very higher fields (950 MHz), especially for peaks in the NSH3 domain.

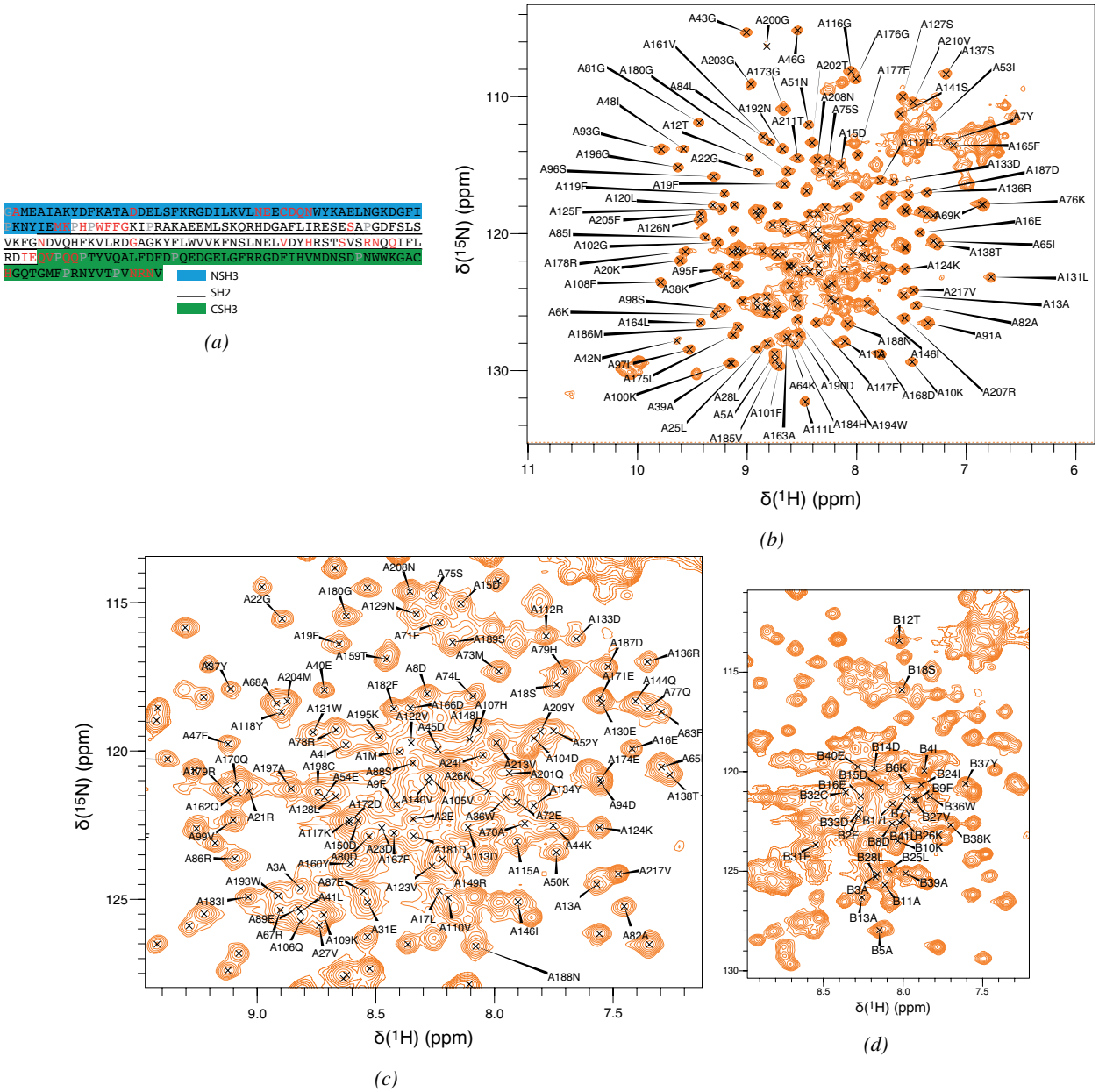


Figure 51 – Assignment of backbone H-N pairs of FL Grb2 (a) The missing assignments are shown in red along the sequence. The residues contained in each monodomain are indicated. The bidomains are simply the sum of the monodomains, with the exception of the additional GA residues in the N-terminus of each construct. (b, c, d) HSQC spectrum of [U-^{15}N; U-^{13}C; U-80% ^{2}H]-Grb2 with its assignment (308 K, PIPES buffer pH 7.2, 600 MHz). A indicates the folded form, and B the peaks corresponding to the unfolded form of the NSH3.

8.1.2 Influence of the His-tag on the stability of the NSH3 and on the interaction between the NSH3 and SH2 domains

Goudreau et al. (1994) showed that the stability of the NSH3 domain depends on pH. The charges in the His-tag could therefore be the reason for its interaction with the NSH3, which has a theoretical pI of 4.76. Given that cysteine oxidation was also shown to impact stability of the NSH3 domain via dimerization (Goudreau et al., 1994), it was verified that the folded/unfolded equilibrium did not depend on the quantity of the reducing agent in the buffer (TCEP 2 mM, tested up to 5 mM).

The His-tag destabilizes the isolated NSH3 The His-tag in our construct is 24 residues long. The HSQC spectra of the isolated NSH3 domain with the His-tag and with the His-tag cleaved are shown in figure 52a. The presence of the tag has two effects: it decreases the proportion of the folded form, and provokes chemical shift differences in the folded form, not only localized near the N-terminus. The combined ^{1}H-^{15}N chemical shift perturbations (CCSP) between the two constructs are shown in figure 52b. While the C32 region (around the N-Src loop) is missing in both construct, the region around residue 10 is only missing in the His-tagged constructs, suggesting broadening due to interaction with the tag or exchange between the folded and unfolded forms. The residues adjacent to the missing stretch around C32 are particularly affected by the presence of the His-tag, suggesting that this (negatively charged) region interacts with the tag.

The His-tag stabilizes the NSH3 in the NSH3SH2 bidomain and FL Grb2 by interacting with the SH2 domain In the NSH3SH2 domain, and the FL construct, the effect of the His-tag is reversed compared to the isolated NSH3: the tag stabilizes it. When the His-tag is not cleaved, only one set of peaks is observed. When the tag is cleaved, too many peaks can be seen on the HSQC spectrum, as shown on figure 53a. While only one form of the SH2 domain can be found, properly folded, two sets correspond to the NSH3 domain, one well dispersed and the other one corresponding to an unfolded form. The well-dispersed peak of this spectrum can be compared to the unique peaks of the His-tag construct (figure 53c). The presence of this His-tag strongly affects the whole bidomain, but some residues particularly stand out, additionally to the N-terminus where the tag is attached: the region around the N-Src loop and the residues 129N and 130E of the SH2 domain. These residues are highlighted in yellow in figure 53c. They are quite far apart when mapped on the crystal structure of full-length Grb2 (PDB 1GRI). The NSH3 and SH2 domains seem to

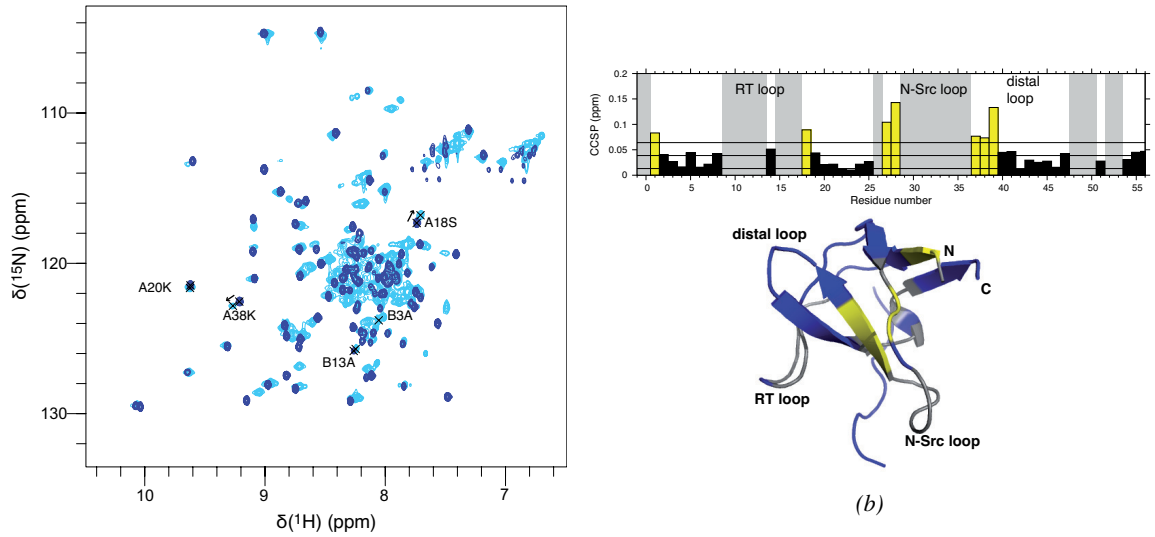


Figure 52 – Effect of the His-tag on the isolated NSH3 domain. (a) HSQC spectra were recorded in the PBS buffer, at 950 MHz. Cyan: His-NSH3; Dark blue: NSH3. Some example of peak assignments are shown (A: folded form; B: unfolded form). (b) Chemical shift perturbation caused by the His-tag. Grey: missing data due to lack of assignment in one or both of the constructs; black lines: mean+std (standard deviation), mean and mean-std values after outliers were removed; yellow: residues with CCSP > mean+std. The result is shown on the structure of the NSH3 domain alone (PDB 1GBQ).

interact, directly or indirectly via the His-tag. Given the nature of the residues involved, the interaction is probably electrostatic.

The influence of the His-tag is the same for full-length Grb2, and the regions of the protein affected by the His-tag are identical (data not shown). The CSH3 is not affected at all. The interpretation that the His-tag mediates an interaction between the NSH3 and SH2 domains is supported by ^{15}N relaxation data recorded on His-tagged full-length Grb2, presented in figure 54. The data clearly show that the dynamic behavior of His-Grb2 is separated in two parts: the NSH3SH2 that behaves as one dynamic entity, and the CSH3 that behaves as another. Mean R_1 and R_2 ($\pm$ std) values (after outliers were removed) are very similar for the NSH3 (0.77 ± 0.05 s^{-1} and 30.08 ± 2.63 s^{-1} respectively) and SH2 (0.74 ± 0.06 s^{-1} and 29.84 ± 3.63 s^{-1} respectively) domains, and significantly different from the rates for the CSH3 domain (1.13 ± 0.07 s^{-1} and 21.46 ± 1.79 s^{-1} respectively). If each domain was moving freely, the values for the NSH3 and CSH3 domain would be similar, since they have approximately the same size and shape, and would be different from the SH2. This structural and dynamic behavior of His-Grb2 is to be compared with the behavior of FL

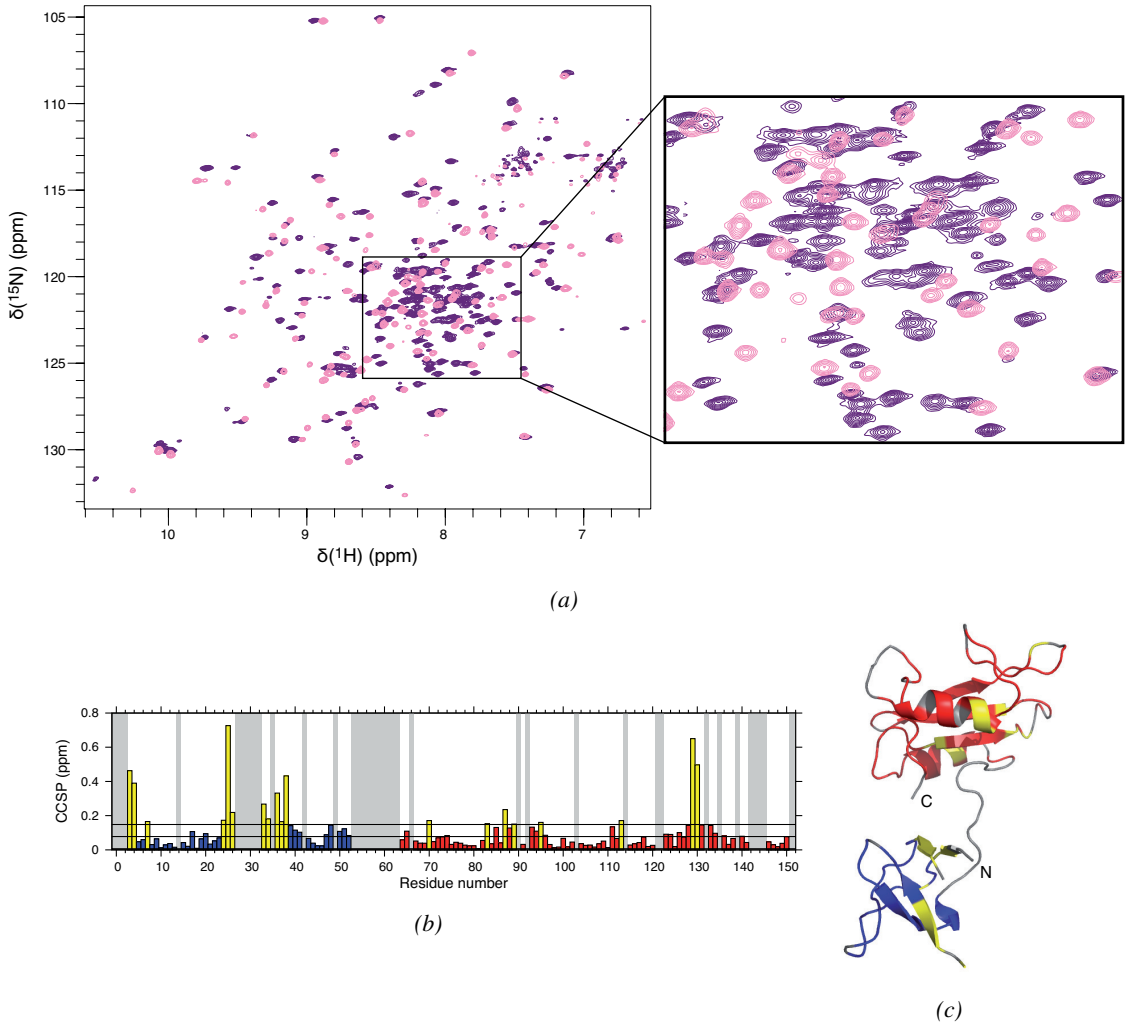


Figure 53 – Effect of the His-tag on the NSH3SH2 bidomain. (a) HSQC spectra of the His-NSH3SH2 (light pink, 800 MHz) and NSH3SH2 (purple, 950 MHz) constructs, with a zoom in the central region to show additional peaks in the NSH3SH2 spectrum. (b) Chemical shift perturbation caused by the His-tag in the NSH3SH2 bidomain. Grey: missing data due to lack of assignment in one or both of the constructs; black lines: mean+std (standard deviation) and mean values after outliers were removed; yellow: residues with CCSP > mean+std. (c) The result is mapped on the NSH3SH2 extracted from the crystal structure of FL Grb2 (PDB 1GRI).

Grb2 without the His-tag (sections 8.2 and 8.3).

Histidine tags are relatively small, and rather hydrophilic, and are therefore often considered not to affect protein structure and function. While that may be true for many proteins, several examples were found were a histidine tag influenced enzymatic activity (Majorek et al., 2014; Panek et al., 2013; Araújo et al., 2000), protein dynamics (Thielges et al., 2011), and/or protein stability in a pH and salt-dependent manner (Panek

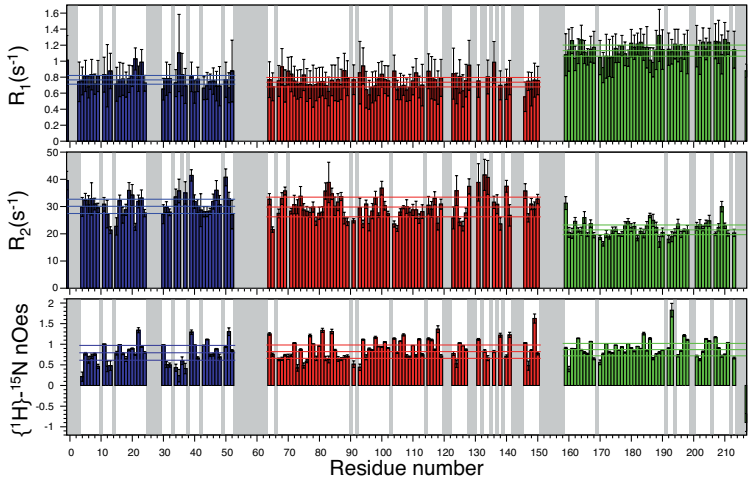


Figure 54 – ^{15}N relaxation parameters of His-Grb2 (full-length) measured at 800 MHz. The different colors show the delimitation of the domains. Blue: NSH3; red: SH2; green: CSH3. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed for each domain in the full-length protein.

et al., 2013; Booth et al., 2018). Given the large differences in structural and dynamic behavior noted between the tagged and tag-free forms in our case, the rest of the study was done on tag-cleaved proteins.

8.2 Chemical shifts comparison of domains

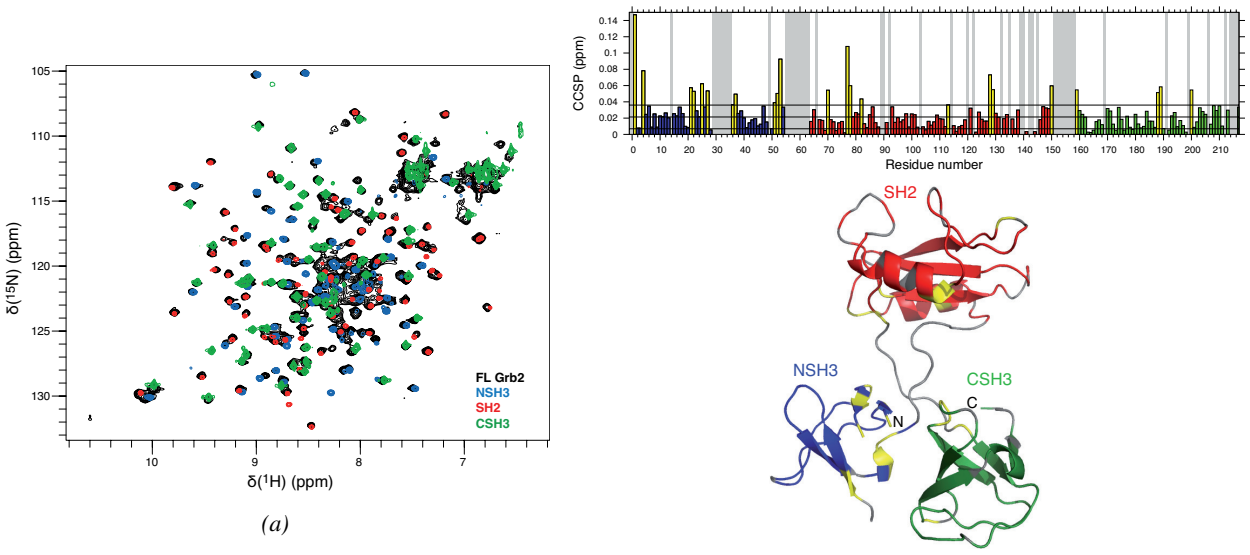
Grb2, as an adaptor protein, performs its biological function by binding to different partners. Regulation of these interactions, their order, and interdependence, are likely to have important consequences in Grb2 signaling, and notably in the Ras-MAPK pathway. This pathway has been shown to be the major pathway in ErbB2-dependent signaling (Dankort et al., 1997, 2001a,b). To regulate Grb2 interactions, the relative orientation and dynamics of its three domains, which can bind different partners, is key. While in the crystal structure of Grb2 (PDB 1GRI) an interaction between the two SH3 domains was observed (Maignan et al., 1995), in solution ligand-bound Grb2 was shown to behaves as three independent domains (Yuzawa et al., 2001). However, no study so far has investigated the solution structure and dynamics of apo-Grb2. Comparison with the bound forms of Grb2 with its different partners in the cell will then be a major step in understanding how they are regulated.

To try and understand what is the organization of each domain relative to one another in the full-length Grb2, the first approach that we chose was to compare the constructs of individual domains with their behavior in the full-length protein. In figure 55a are shown the overlaid ^{1}H-^{15}N HSQC spectra of full-length Grb2 and of the three individual domains (NSH3, SH2 and CSH3). The peaks of the individual domains superimpose

well with their corresponding peaks in FL Grb2, except for a few peaks corresponding to the edges of the domains. The additional peaks seen in FL Grb2 correspond to the unfolded form of the NSH3 domain. This demonstrates the absence of major structural changes or interaction of the domains in the full-length protein, in sharp contrast with what is seen in the crystal structure where the SH3 domains interact with each other. No dimer interface can be mapped either. The crystallographic dimer where the SH2 of one monomer contacts the CSH3 of another monomer is not detected in solution.

Still, some small local chemical shifts changes (mainly CCSP < 0.1 ppm) can be mapped in the sequence and on the crystal structure of Grb2 (figure 55b). The differences are all located on the sides of the domains that point to each other in the crystal structure, spatially close to the linkers. The same observations were made when comparing the individual domains with the bidomain constructs (NSH3SH2 and SH2CSH3), and the bidomain constructs with the FL protein.

Overall, our data suggest minimal interaction between the domains.



(a)

(b)

Figure 55 – Comparison of full-length Grb2 with individual domains (PIPES buffer, 308 K). (a) Overlay of the HSQC spectra of FL Grb2 (black, 600 MHz), the isolated NSH3 domain (blue, 600 MHz), the isolated SH2 domain (red, 800 MHz) and the isolated CSH3 domain (green, 950 MHz). (b) Combined ^{1}H-^{15}N chemical shift perturbations (CCSP) between the peaks of each residue in FL Grb2 and in the individual domains (top). Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed. For the NSH3, the peaks corresponding to the folded form where chosen for comparison between the isolated domain and the full-length protein. The CCSP for the assigned peaks of the unfolded form of the NSH3 were lower than for the folded form. The residues for which CCSP > mean+std are mapped on Grb2 crystal structure (bottom).

8.3 Dynamics of Grb2

We showed that structurally, the three domains of Grb2 appear independent, with only minor chemical shifts differences next to the delimitations of the domains. Depending of the length and stiffness of the linkers, they can still have various dynamics with respect to one another. As seen in figure 54, in the His-tagged protein, the NSH3 and SH2 domains behave as a single dynamic entity. However, in the tag-free protein, it is not the case, as shown in figure 56. The NSH3 shows highly heterogenous dynamics both in the sub-ns (as seen in the heteronuclear nOe values) and ns (as seen in the R_1 and R_2 rates) timescales, which is consistent with its heterogeneity, with folded and unfolded forms exchanging. The SH2 and CSH3 domains are more homogeneous along their respective sequences. Contrary to what was observed for the His-tagged protein, the mean R_1 and R_2 ($\pm$ std, after outliers were removed) relaxation rates are similar for the NSH3 ($1.15 \pm 0.22\ s^{-1}$ and $16.20 \pm 3.11\ s^{-1}$ respectively) and the CSH3 ($1.04 \pm 0.10\ s^{-1}$ and $15.56 \pm 1.41\ s^{-1}$ respectively) domains, consistently with the fact that they have similar sizes and are smaller than the SH2, which has significantly different relaxation rate constants ($0.80 \pm 0.07\ s^{-1}$ and $18.90 \pm 2.49\ s^{-1}$ respectively).

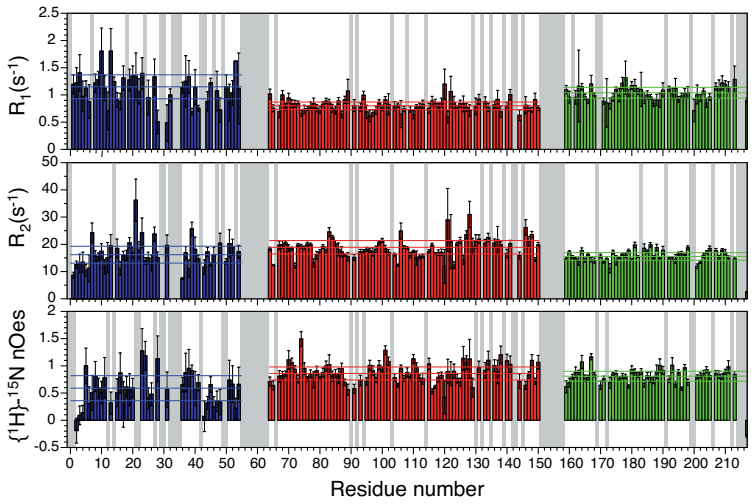


Figure 56 – ^{15}N relaxation parameters of FL Grb2 measured at 800 MHz. The different colors show the delimitation of the domains. Blue: NSH3; red: SH2; green: CSH3. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed for each domain in the full-length protein.

From the ratio of the relaxation rates averaged over the whole protein, we can also estimate a mean rotational correlation time of the NH bonds of Grb2. Details on the method of calculation are given in

section 5.3. With mean R_1 of $0.93 \pm 0.17\ s^{-1}$ and mean R_2 of $17.47 \pm 2.86\ s^{-1}$ (outliers being removed), this gives us $10.09 \pm 2.48\ ns$. If we compare this value with the one expected from the Stokes-Einstein model (described in section 5.3) at 308 K, which gives 15.58 ns, we can see that Grb2 tumbles more rapidly than an ideal globular protein of this size, suggesting some degree of flexibility. This is consistent with three domains that are not docked to one another. This also indicates that the protein is probably monomeric.
Even though the missing peaks in the linkers prevent us from directly getting information of their flexibility, the data that we have indicate that they are flexible enough to allow partially decorrelated dynamics of the three domains.

8.4 Domain organization determined by SEC-SAXS

To confirm our interpretation of the NMR data, get more information on the envelop of Grb2, and compare it with the crystal structure available, we recorded SAXS data on FL Grb2, as well as on the bidomains (NSH3SH2 and SH2CSH3). To make sure that the samples were homogeneous, which is essential for SAXS measurements, we used size-exclusion chromatography (SEC) coupled to SAXS (David and Pérez, 2009; Pérez and Vachette, 2017). This was all-the-more necessary since two peaks were observed by SEC when these constructs were purified, suggesting monomer/dimer equilibria.

8.4.1 The NSH3SH2 bidomain organization corresponds to the organization in the crystal structure of full-length Grb2

We first ran the SEC-SAXS experiment on the NSH3SH2 bidomain. The SEC elution profile, as well as the extrapolated intensity at q=0, I_0, and radius of gyration R_g along this profile, are shown in figure 57a. Firstly, we notice that this construct gives two size-exclusion peaks. From the corresponding SAXS data, the peaks correspond to proteins of respectively 33.1 and 17.8 kDa, while the expected mass of the construct is 17.8 kDa. A reasonable assumption is that the first peak corresponds to a dimeric form of the bidomain, and the second to the monomer, with about twice as much monomer compared to the dimer from the absorbance profiles at 280 nm. While the curves corresponding to the dimer have not been analyzed yet, the scattering intensity, dimensionless Kratky plot and distance distribution P(r) extracted for the monomer are given in figure 57b, 57c and 57d, respectively. The Kratky plot and P(r) are close from the profile expected for a globular protein, but with a shoulder in the P(r) distribution consistent with the presence of two distinct domains. Then, we compared the experimental SAXS curve with the one calculated by Crysol (Svergun

et al., 1995) for the NSH3SH2 domain structure extracted from the crystal structure of Grb2 (PDB 1GRI). The fit is very good, with a residual value $\chi = 1.15$ between the two curves. The NSH3SH2 bidomain therefore has an organization in solution fully compatible with the one seen in the crystal structure, with the two domains rather far from one another. No variation of the mean inter-domain distance can be detected by SAXS, with a measured R_g of 19.3 Å.

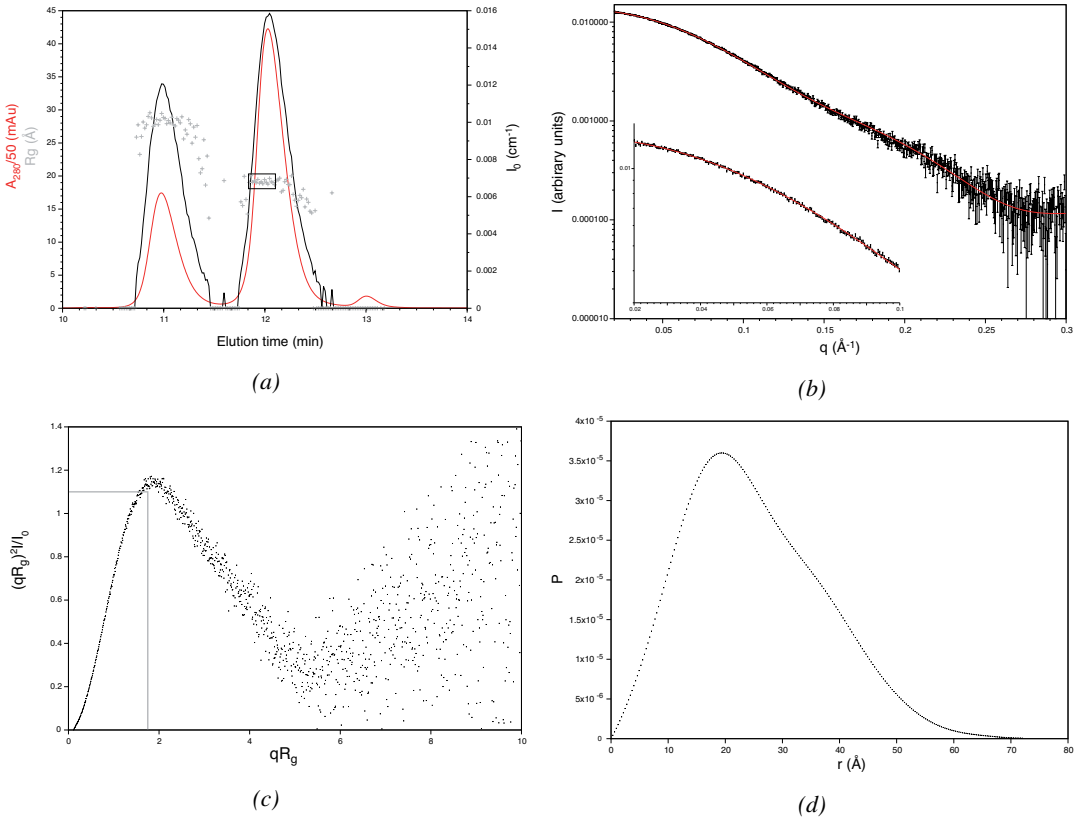


Figure 57 – SEC-SAXS data of Grb2 NSH3SH2 bidomain monomer. (a) SEC-SAXS profile. Along the elution profile are plotted the absorbance at 280 nm A_{280}, the radius of gyration R_g and the extrapolated scattered intensity at q=0 I_0. The chosen SAXS curves are taken in a zone of stable R_g (boxed). (b) Experimental scattering curve (black) and scattering curve from the NSH3SH2 bidomain extracted from the crystal structure (PDB 1GRI), calculated with crysol (red). $\chi = 1.15$. A zoom at small q is shown. (c) Dimensionless Kratky plot. In grey is the expected maximum for a globular protein. (d) Distance distribution between atoms, computed from the SAXS curve.

8.4.2 The SH2CSH3 bidomain adopts multiple conformations of different R_g

Then, we applied the same method to the SH2CSH3 bidomain. As shown in figure 58a, the proportion of dimer for this construct is minor, and no SAXS curve could be extracted from the corresponding peak. The monomer gave a molecular weight estimated by SAXS of 18.7 kDa, close to the expected value from the sequence (19.3 kDa). Contrary to the results for the NSH3SH2 bidomain, the Kratky plot given in fig-

ure 58c has a shifted maximum compared to the expected one for a globular protein, consistent with a more pronounced shoulder in the P(r) distribution shown in figure 58d. When we calculated the expected SAXS curve for the SH2CSH3 bidomain from the PDB structure of the full-length protein, the fit with our data was not as good as previously observed for the NSH3SH2 bidomain. The fit is shown in figure 58b, and gives $\chi = 1.42$. The domain organization of the SH2CSH3 bidomain is therefore not the same when isolated in solution than in the crystal structure of the full-length protein.

To have more information on the difference between the organization of our construct in solution and in the crystal, we used EOM (Tria et al., 2015; Bernadó et al., 2007). We defined as flexible the linker between the two domains (residues 154-160, slightly longer than the linker defined in Uniprot to make sure to have the best sampling). EOM generated 10,000 structures of the bidomain sampling the conformational space available with this flexible linker, starting from the crystal structure. The R_g distribution of the sub-ensembles that were selected by the program to fit the experimental data is shown in figure 58e. To fit the experimental curve, two populations with very different R_g are needed. They correspond to compact (low R_g) and extended (high R_g) conformers. The sampling of conformers with different R_g is not directly detected on the NMR spectra of this bidomain, in which we observe only one set of peaks. This suggests an exchange that is fast on the chemical shift timescale. Even in the most compact conformations selected by EOM, the domains are not very close and are not expected to interact very tightly with each other, consistently with the absence of significant chemical shift differences in the spectra of the individual domains and the bidomain. The presence of a higher degree of dynamics in the SH2CSH3 bidomain than in the NSH3SH2 bidomain is compatible with the absence of need for deuteration for the acquisition of 3D spectra of the SH2CSH3, contrary to the NSH3SH2.

8.4.3 FL Grb2 also adopts multiple conformations with different R_g

The analysis of SAXS data for the SH2CSH3 domain was finally repeated for the full-length Grb2 construct, as shown in figure 59. As for the SH2CSH3 bidomain, the proportion of dimer is small. FL Grb2 monomer (for which the molecular weight estimated by SAXS is 25.0 kDa, close to the expected value of 25.33 kDa) is overall more globular than the SH2CSH3 bidomain, as seen on the Kratky plot in figure 59c. This excludes the possibility of a completely extended form where all three domains are linearly arranged. The crystal structure, where the two SH3 domains are tightly packed (giving R_g = 21.36 Å), does not fit the solution data (R_g = 22.9 Å, figure 59b). Once again, in order to get a better idea of the domain organization, we used

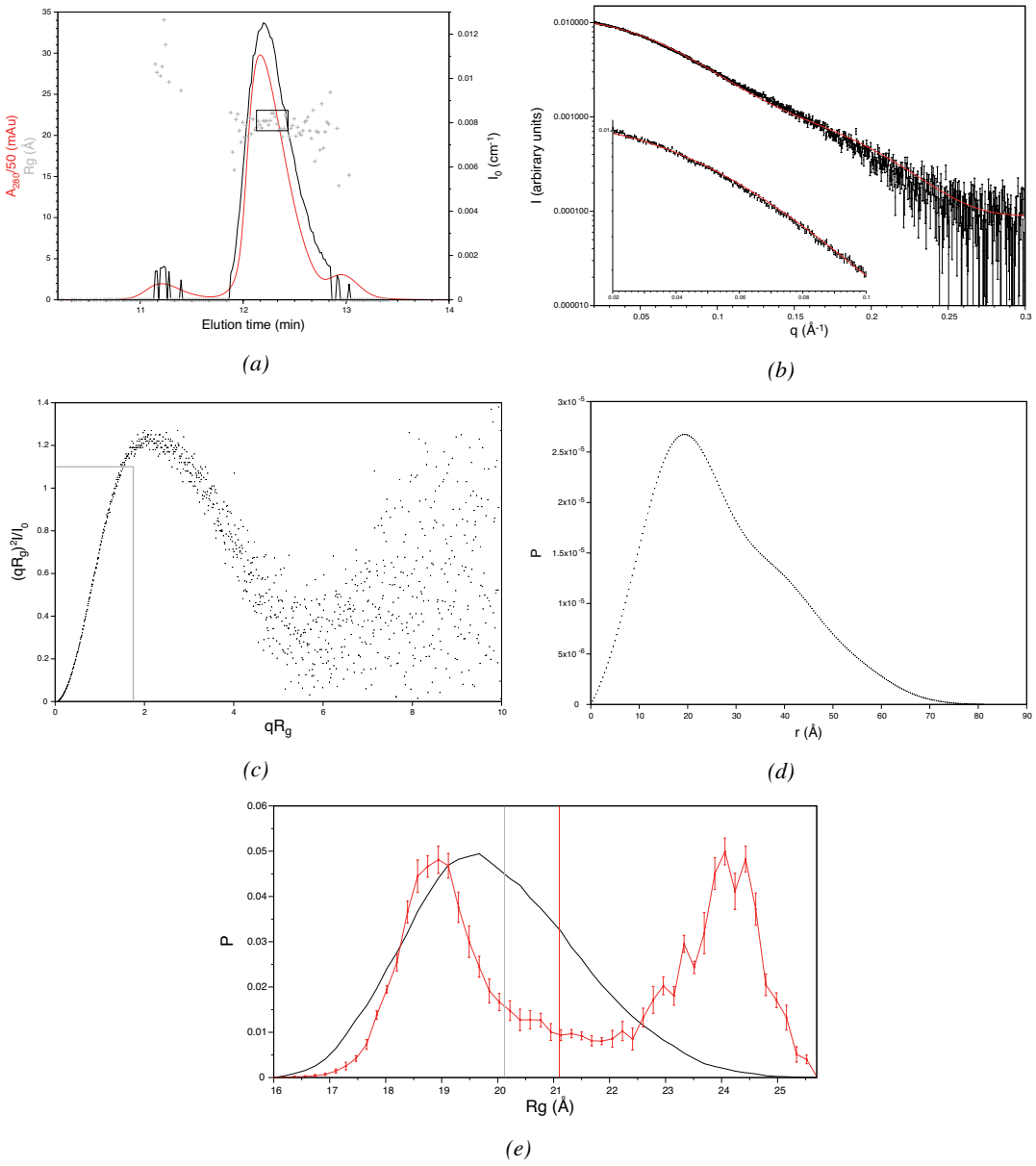


Figure 58 – SEC-SAXS data of Grb2 SH2CSH3 bidomain. (a) SEC-SAXS profile. Along the elution profile are plotted the absorbance at 280 nm A_{280}, the radius of gyration R_g and the extrapolated scattered intensity at q=0 I_0. The chosen SAXS curves are taken in a zone of stable R_g (boxed). (b) Experimental scattering curve (black) and scattering curve from the SH2CSH3 bidomain extracted from the crystal structure (PDB 1GRI), calculated with crysol (red). $\chi = 1.42$. A zoom at small q is shown. (c) Dimensionless Kratky plot. In grey is the expected maximum for a globular protein (d) Distance distribution between atoms, computed from the SAXS curve. (e) EOM R_g distribution for the SH2CSH3 bidomain monomer. Black: pool generated by Flexible Meccano; red: ensembles selected by EOM to fit the experimental SAXS data; grey bar: R_g of the bidomain from the crystal structure; red bar: experimental R_g.

EOM, defining as flexible the two linkers between the domains (residues 57-65 and 154-160). The results are shown in figure 59e. As for the SH2CSH3 bidomain, different types of conformations are needed to fit the experimental data, from a rather compact form, with a radius of gyration compatible with the crystal structure, to a more extended conformation. Structures of intermediate R_g are also populated. These data show that Grb2 samples conformations more extended than the crystal structure. From the observation made on the SH2CSH3 bidomain, the presence of extended conformations could come from extension of the SH2-CSH3 linker, but it could also indicate "breathing" of the protein, with the NSH3 and CSH3 experiencing different distance from one another.
Overall, flexibility of Grb2, with variations of the inter-domain distances, is consistent with its role as an adaptor protein.

8.5 New model for Grb2 domain organization and oligomeric state

Although Grb2 is an essential link between receptor tyrosine kinases and the Ras/MAPK signaling pathway (Lowenstein et al., 1992; Gale et al., 1993; Buday and Downward, 1993), the study of its solution structure without peptide ligands had never really been undertaken. Isolated domains had been characterized (Goudreau et al., 1994; Wittekind et al., 1994, 1997; Thornton et al., 1996; Kohda et al., 1994), as well as the full-length protein but either in a crystal (Maignan et al., 1995) or in solution with peptide ligands for all three domains (Yuzawa et al., 2001). Since the biological role of Grb2 relies on the ability of the different domains to interact with different partners, inter-domain flexibility is key and is not captured by crystallography. Binding might also change the dynamics, and therefore the description of the bound state only is not sufficient to understand Grb2 behavior. This is the reason why we characterized the solution state of apo-Grb2.

8.5.1 Stability of the NSH3 and His-tag

We showed that in our experimental conditions (40 mM PIPES pH 7.2, 150 mM NaCl, 2 mM TCEP, 308 K) the NSH3 domain of Grb2 is unstable and has a folded and unfolded form, in isolation but also in the NSH3SH2 bidomain and FL Grb2. This was already reported for this domain (Goudreau et al., 1994), but also for its *Drosophila* counterpart Drk (Zhang and Forman-Kay, 1995). Interestingly, we found that

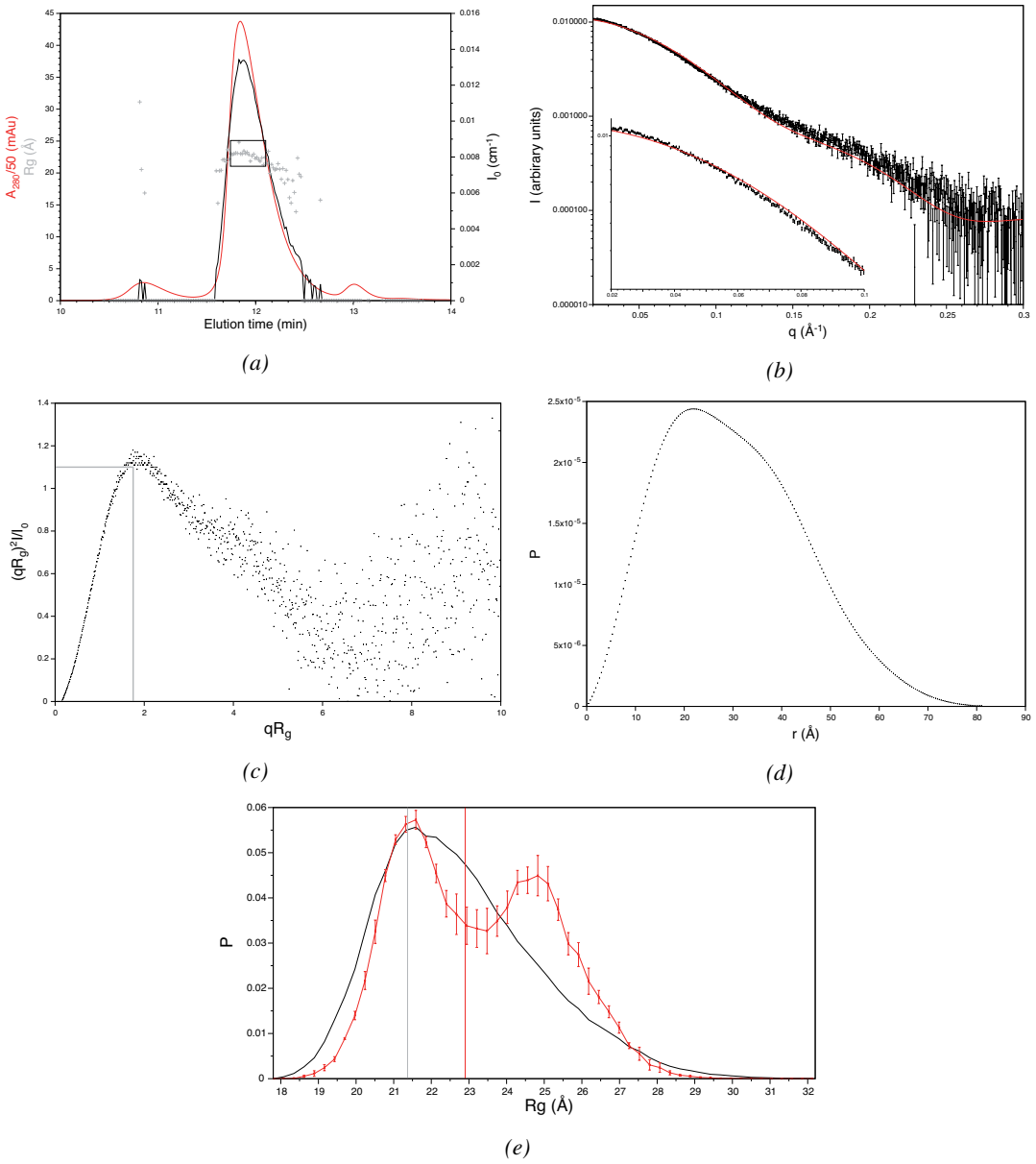


Figure 59 – SEC-SAXS data of full-length Grb2. (a) SEC-SAXS profile. Along the elution profile are plotted the absorbance at 280 nm A_{280}, the radius of gyration R_g and the extrapolated scattered intensity at q=0 I_0. The chosen SAXS curves are taken in a zone of stable R_g (boxed). (b) Experimental scattering curve (black) and scattering curve from the crystal structure (PDB 1GRI), calculated with crysol (red). $\chi = 1.84$. A zoom at small q is shown. (c) Dimensionless Kratky plot. In grey is the expected maximum for a globular protein (d) Distance distribution between atoms, computed from the SAXS curve. (e) EOM R_g distributions for FL Grb2 monomer. Black: pool generated by Flexible Meccano; red: ensembles selected by EOM to fit the experimental SAXS data; grey bar: R_g of the bidomain from the crystal structure; red bar: experimental R_g

a Histidine tag in the N-terminus destabilizes even more the isolated domain by interacting with acidic residues, but stabilizes it by favoring an interaction with the SH2 domain in the NSH3SH2 and FL Grb2.

This tag-mediated inter-domain interaction significantly changes the dynamics of the domains, docking the NSH3 and SH2 domain together. Besides, only one peak was observed by SEC for His-tagged Grb2 (data not shown), while two peaks are seen in the tag-free construct, suggesting an influence of the His-tag on dimerization. One should then be very careful about results obtained on His-tagged Grb2.

8.5.2 Oligomeric state of apo-Grb2 in solution

In the literature, there are very contradictory data about Grb2 oligomerization, ranging from a low-μM according to Lin et al. (2012) to high mM (Yuzawa et al., 2001; Guilloteau et al., 1996) dissociation constants of the dimer. The study by Lin et al. (2012) uses a His-tagged construct of the protein, and therefore one should be careful about the interepretation of the results, since we showed it changed Grb2 conformation and dynamics. For all constructs, NMR did not show any sign of the presence of a dimer (at concentrations of a few hundreds of μM), with only one set of peaks (except for the NSH3 folded/unfolded equilibrium), excluding non-symmetric dimers, and dynamics of the full-length giving estimated rotational correlation times incompatible with a dimer. The good superimposition between the spectra of the full-length protein and the mono- and bi-domains excludes formation of a dimeric interface, that would result in chemical shift differences. Therefore the crystallographic dimer, where the SH2 domain interacts with the CSH3 domain of the other monomer, is not seen in solution. It is not impossible that a minor dimeric form exists but can not be seen due to the combination of low quantity and faster transverse relaxation. The SEC-SAXS data confirm that the highly major form is a monomer for the SH2CSH3 bidomain as well as for the full-length protein. The NSH3SH2 presents a high proportion of dimer separated by size-exclusion chromatography, but was also injected at a higher concentration. The dimeric peak could not be analyzed by NMR or SAXS so far, by lack of time. A problem with SEC-SAXS is that the concentration of the protein which scattering curve is measured is unknown. A rough estimate of the dilution factor by the SEC column of 5 gives concentration of the FL Grb2, SH2CSH3 and NSH3SH2 constructs of about 30 μM, 130 μM and 20 μM respectively. If data showed that at similar concentrations, the NSH3SH2 bidomain does dimerize in a higher proportion than the other constructs, it would be compatible with a role of the C-terminal part of the SH2 domain in dimerization, as in the dimer-swapped SH2 dimer (Schiering et al., 2000; Nioche et al., 2002; Benfield et al., 2007; Papaioannou et al., 2016). This would have to be investigated further. Overall, our data are not compatible with a dimer dissociation constant below μM for FL Grb2. However, it does not exclude the possibility sub-mM dimer dissociation constant.

8.5.3 Domain organization in Grb2 monomer

In the crystal structure of Grb2 (PDB 1GRI), the two SH3 domains are far from the SH2 domain and interact with one another. In solution, with ligand peptides loaded, Yuzawa et al. (2001) showed that, on the contrary, the SH3 domains are not interacting. Here, we firstly showed by NMR that the domains in apo-Grb2 in solution are structurally and dynamically independent from one another, similarly to the peptide-loaded form. No interaction interface can be seen between the domains, and the three domains tumble with different rotational correlation times. From the analysis of SAXS data, we propose a model of apo-Grb2 in solution that has several populations of different radius of gyration, from a R_g similar to the crystal structure to a more extended population with significantly higher R_g. We used the bidomain constructs, NSH3SH2 and SH2CSH3 to understand the contribution of each inter-domain linker to this observed flexibility. It showed that while the NSH3SH2 bidomain is compatible with the organization in the crystal structure, the SH2CSH3 bidomain also has at least two populations of different R_g. The R_g of the compact form of the SH2CSH3 bodimain is smaller than the R_g of the crystal structure SH2CSH3 bidomain. Conversely, the R_g of the most compact form of the FL protein fits the R_g of the full-length crystal structure. The compact form of the SH2CSH3 bidomain is therefore influenced by the presence of the NSH3 domain. In the model we propose, even though the two interdomain linkers are flexible enough to decorrelate the dynamics of the three domains, the NSH3 and SH2 domains experience little to no movement compared to one another, while the CSH3 domain is able to sample a large conformational space between a compact and an extended form. This is consistent with the definition of the domains and of the linkers shown in figure 34a, where the NSH3-SH2 linker in shorter than the SH2-CSH3 linker. Moreover, the NSH3-SH2 linker contains a ^{57}PHP^{59} sequence that is expected to be particularly rigid, whereas the SH2-CSH3 also contains two prolines but in a ^{155}PQQP^{158} sequence expected to be less constrained. Variation of the distance between the two SH3 domains can not be excluded either. The resulting model is presented in figure 60.
Several complementary experiments could be run to confirm and complement this model:

- We could run EOM on the FL Grb2 data, with a fixed NSH3SH2 bidomain and a flexible linker between the SH2 and CSH3 domains, to confirm that the flexibility of the FL protein can arise solely from movement of the CSH3 relative to the NSH3 and SH2 domains.

- By NMR, we could record RDC and PRE experiments to have more information on the dynamics, orientations and distances between the domains.

- By NMR, relaxation dispersion experiments or CEST experiments would detect potential conformational exchange in the μs to ms timescale, relative to inter-domain movements or possible existence of a dimer. Analysis of Grb2 dynamics at several magnetic fields could also be informative.

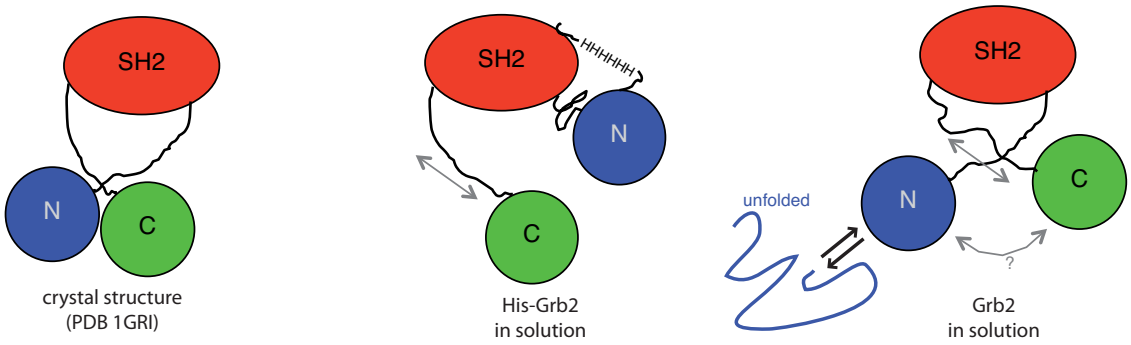


Figure 60 – Model of His-Grb2 and Grb2 (monomers) structure and dynamics in solution, compared to the crystallographic model.

8.5.4 Importance for Grb2 interactions and biological function

As noted by Yuzawa et al. (2001), in a modular adaptor protein such as Grb2, inter-domain dynamics is expected to be key, which makes the study of dynamics crucial to understand their biological function. In the case of the interaction with ErbB2, it would be essential to understand if, and how, Grb2 can interact with both the expected phosphotyrosine site via its SH2 domain and polyproline motifs via its SH3 domain(s). If the SH3 domains compete for several partners which are known to interact with polyproline motifs with moderate affinities, flexibility would also be a way to allow rapid binding and dissociation, in a manner rather similar to the binding modes of IDPs for that matter.

9 Interaction between CtErbB2 and Grb2

Even though Grb2 is known to interact with ErbB2 via the phosphorylated tyrosine B (Y_B, Y152 in our numbering, Y1139 in the full-length ErbB2 numbering) (Dankort et al., 1997), the interaction has never been characterized at the atomic level. The existing structures of Grb2 complexes only include single domains of Grb2, with peptides of partners. Besides, no peptide from ErbB2 has ever been studied in complex with Grb2. In the current model of signal transduction by Grb2, its SH2 domain interacts with RTKs while the SH3 domains interact with other signaling proteins: Sos, that is part of the Ras/MAPK pathway, interacts mainly with the NSH3 domain but can also interact with the CSH3 domain (McDonald et al., 2009); Gab1 and Gab2 large multi-site docking (LMD) proteins interact with the CSH3 domain, and are involved in the MAPK and PI3K pathways (Lewitzky et al., 2001; Harkiolaki et al., 2009).
Given the existence of several polyproline motifs in CtErbB2, as studied in section 7, and the observation that Grb2 binding to ErbB2 is impaired by deletion of its SH3 domains (Xie et al., 1995), it is likely that a SH3-polyproline interaction also participates in Grb2 binding to ErbB2. Several questions therefore remain regarding this interaction:

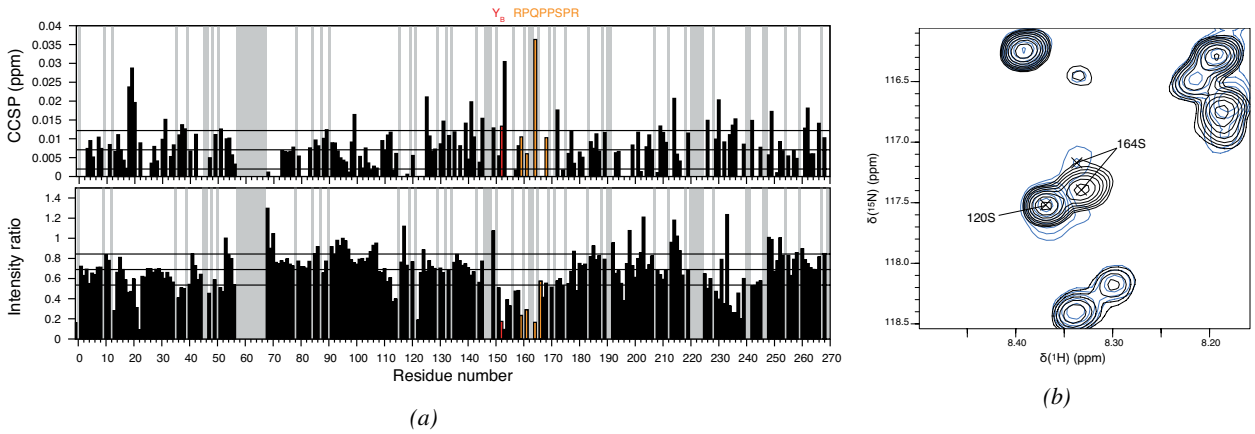
- Does Grb2 binding to ErbB2 involve SH3-polyproline interactions? Can these interactions interfere with binding to other partners of Grb2?
- How are the structural and dynamic features of CtErbB2 affected when bound to Grb2? Do the transiently formed PPII helices observed in free CtErbB2 fold upon binding to Grb2?
- How does the domain organization of Grb2 in solution that we determined, and in particular the inter-domain flexibility, influence binding?
- How does phosphorylation of ErbB2 tyrosines influence binding to Grb2?

First, we chose to investigate the binding of Grb2 to unphosphorylated CtErbB2. The binding is expected to be weak, since it was shown to require phosphorylation of Y_B *in vivo* (Dankort et al., 1997, 2001a,b). However, this could already allow detection of a potential SH3-polyproline interaction.

9.1 Grb2 and unphosphorylated CtErbB2

First we titrated 50 μM labeled CtErbB2 with unlabeled Grb2 in the PIPES buffer (pH 7.2, 298 K), following binding by NMR at 950 MHz. The effects are, as expected, weak. Some very small chemical shifts

changes can be observed, associated with some decreasing intensities, suggesting an intermediate exchange regime between the free and bound forms. Both chemical shifts and intensity effects at the final titration point (1:10 CtErbB2:Grb2 ratio) are plotted against CtErbB2 sequence in figure 61a.



(a) *(b)*

Figure 61 – Titration of CtErbB2 by Grb2. (a) ^{1}H-^{15}N HSQC comparison between CtErbB2 alone and with 10 equivalents of Grb2. ^{1}H-^{15}N CCSP (top) and intensity ratio (bottom) along CtErbB2 sequence. In red is Y_B (Y152) and in orange is the polyproline segment. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed. (b) S164 region of the HSQC spectrum of CtErbB2 alone (black) or with 10 equivalents of Grb2 (blue).

Three main regions are affected by the interaction, both in terms of intensity and chemical shifts: the N-terminal region around residue 20, the region between residues 150 and 170, and the region between residues 230 and 240. The middle region is the expected interaction site (Y152). Interestingly, it seems to interact with Grb2 (although weakly) even when not phosphorylated. The amide group of a nearby residue, S164, is also particularly affected, as shown on the spectra in figure 61b. It is in the center of a proline-rich segment ^{159}RPQPPSPR^{166}. This segment contains a canonical class I interaction motif for SH3 domains, RxxPxxP, and a PxxPxxR motif that is not canonical but is close to a class II PxxPxR motif (see section 3.2.2). This proline-rich segment is located in a transiently formed PPII helix of CtErbB2, with an estimated population of 40% (see section 7.4). These characteristics make it an ideal candidate for interaction with one or both of the Grb2 SH3 domains. The two other regions where amide groups are affected by the interaction also contain tyrosines: Y18 for the first one, Y234 and Y235 for the second one. None of them has the asparagine in position +2 that characterizes the interactions motifs for the SH2 domain of Grb2 (Marengere et al., 1994). The first one is not an autophosphorylation site, while Y234 and Y235 define the Y_D autophosphorylation site. Notably, the region of tyrosine 18 is in the transient α-helix, and the tyrosines 234 and 235 are in

the C-terminal region involved in the long-range contact that we have characterized in free CtErbB2. No potential SH3-binding polyproline motif can be found in these regions. The origin of the effect of Grb2 on these regions, direct or conformational, remains to be investigated. Grb2 binding could, for example, disrupt the long-range contact.

To investigate the binding on the side of Grb2, we did the reverse titration by following the effect of the addition of unlabeled CtErbB2 to a sample of 100 μM [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-Grb2, up to a Grb2:CtErbB2 ratio of 1:5, at 600 MHz. The conditions (buffer, temperature) were similar than for the first titration. The chemical shift and intensity perturbations at the final titration point are plotted along Grb2 sequence in figure 62. The main effect of CtErbB2 addition are small chemical shift perturbations in the N-terminal SH3 domain, indicating a fast exchange regime. The largest CCSP are seen for residues 12, 15, 37, 47, 50 and 51, consistent with the residues previously shown to be involved in binding of polyproline peptides (residues F9, A13, D15, W36 and Y52, see figure 27). A few residues experience significant chemical shift changes in the SH2, residues 72 and 140 in particular, while residues 107 to 109 experience small chemical shift changes combined with decreased intensity. Residues E72 and V140 are on the ligand binding face of the SH2 domain but are not known to be directly involved in binding of phosphopeptides. Lysine 109 is known to interact with both the ring of the ligand phosphotyrosine and with the asparagine in position +2 relative to the tyrosine. Perturbation of this lysine is consistent with binding of the tyrosine, even non phosphorylated. Residues of the SH2 domain that are known to interact with the phosphate group of phosphotyrosine ligands (R67, R86, S88, S90, see figure 25) are not affected by the binding here. The CSH3 domain is almost not affected by binding, except for a small chemical shift perturbation of residue 200, which is in the distal loop, far from the expected peptide binding site.

Overall, our data show binding of Grb2 via its N-terminal SH3 domain on at least one polyproline motif C-terminus of the tyrosine B of ErbB2. Additionally, weak but detectable binding of the SH2 domain to ErbB2 is detected, most probably on the tyrosine Y_B, which is the known interaction site of Grb2 on ErbB2 and the most perturbed tyrosine region in the reverse titration. Other regions of ErbB2 including tyrosines are affected by the binding, but do not show the consensus sequence for Grb2 binding. We could not measure reliable K_Ds, as saturation was not reached at the concentrations used, but we can estimate it to be

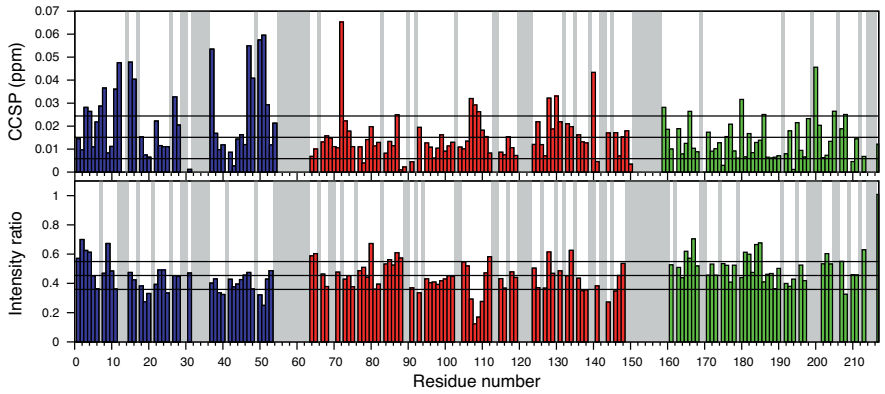


Figure 62 – ^{1}H-^{15}N HSQC comparison between Grb2 alone and with 5 equivalents of CtErbB2. ^{1}H-^{15}N CCSP (top) and intensity ratio (bottom) along Grb2 sequence. Different colors delimit the different domains: blue: NSH3; red: SH2; green: CSH3. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed.

in the mM range. Further structural characterization of the complex was done with peptides of ErbB2, as described below.

9.2 Need for tyrosine phosphorylation of CtErbB2

In the titration experiments presented so far, the binding of Grb2 to unphosphorylated CtErbB2 is too weak, and therefore the proportion of bound proteins is too low, to easily study the bound state. The results of most experiments would be dominated by the behavior of the free proteins in solution. This weak binding might not be biologically significant considering the expected range of the protein concentrations *in vivo*, which is consistent with the fact that CtErbB2-Grb2 binding has been shown to occur and be signaling-competent *in vivo* only when CtErbB2 is phosphorylated (Dankort et al., 1997). It is thus essential, to go further into the characterization of this interaction, to be able to phosphorylate CtErbB2 tyrosines, and especially the known interaction site for Grb2, Y_B. However, controlled phosphorylation of a multisite protein is not easy, and tyrosine phosphorylation is not as studied as serine/threonine phosphorylation. From our study with unphosphorylated CtErbB2, the known phosphotyrosine interaction site, Y_B, is close to the polyproline motif also interacting with Grb2. Peptides containing both these sites can thus be easily obtained by chemical synthesis in the unphosphorylated and phosphorylated form. In parallel to setting up methods to phosphorylate the whole CtErbB2, presented in section 10, we therefore chose to use a peptide approach to further study the CtErbB2-Grb2 interaction. To compare binding in the unphosphorylated and

phosphorylated state, we designed two peptides, P1 (unphosphorylated) and P2 (phosphorylated), containing both Y_B and the double PxxP/arginine motif RPQPPSPR. Their sequence is given in figure 63.

P1: ^{141}VAPLTCSPQPE**Y**VNQPDV**RPQPPSPR**EGPLPAARPA^{176}
P2: ^{141}VAPLTCSPQPE**pY**VNQPDV**RPQPPSPR**EGPLPAARPA^{176}

Figure 63 – ErbB2 P1 and P2 peptides designed for interaction with Grb2. The expected interaction sites, the tyrosine and the proline-rich segment, are in bold.

9.3 Grb2 and peptides of CtErbB2

Titrations of Grb2 with ErbB2 peptides were done in the same conditions as the titration of Grb2 with the whole CtErbB2 construct (on a [*U*-^{15}N; *U*-^{13}C; *U*-80% ^{2}H]-Grb2 sample at 100 μM in PIPES buffer at pH 7.2, at 298 K), and followed at 600 MHz. Titrations were performed up to a large quantity of peptide (Grb2:peptide ratio of 1:20, yielding a peptide concentration of 2 mM) to try to saturate Grb2.

9.3.1 Unphosphorylated Y-PxxP peptide P1

First we titrated Grb2 with the unphosphorylated P1 peptide, up to a Grb2:P1 ratio of 1:20. The overlaid HSQC spectra of the initial (1:0 ratio) and final (1:20 ratio) titrations points, with one intermediate titration point (1:5 ratio), are shown in figure 64a. The CCSP and intensity ratios between the initial and final points along the sequence are shown in figure 64b, and the largest shifts and intensity decreases are mapped onto Grb2 crystal structure in figure 64c. The NSH3 binding site exhibits exchange in the fast regime, while the region around K109 in the SH2 domain seems to experience intermediate exchange, with peaks of the free form broadening but no visible appearance of peaks corresponding to the bound form. The position of the affected peaks in the SH2, on top of the central β-sheet perpendicular to which the SH2-binding peptides are known to be located, is consistent with correct positioning of the peptide in the binding pockets.
The interaction we observe with both the NSH3 and SH2 domains are similar to what was observed with the whole C-terminal tail of ErbB2, except that the shifts that we observe here for the NSH3 are slightly larger, due to the higher ratio of peptide added. Still, we do not saturate 100 μM Grb2 with 2 mM peptide P1. The dissociation constant once again seems to be of the order of mM. This similarity of binding between the whole CtErbB2 and P1 indicates that the peptide we chose is a good model to study the influence of phosphorylation. We then performed the titration with the phosphorylated peptide P2, which is more biologically relevant.

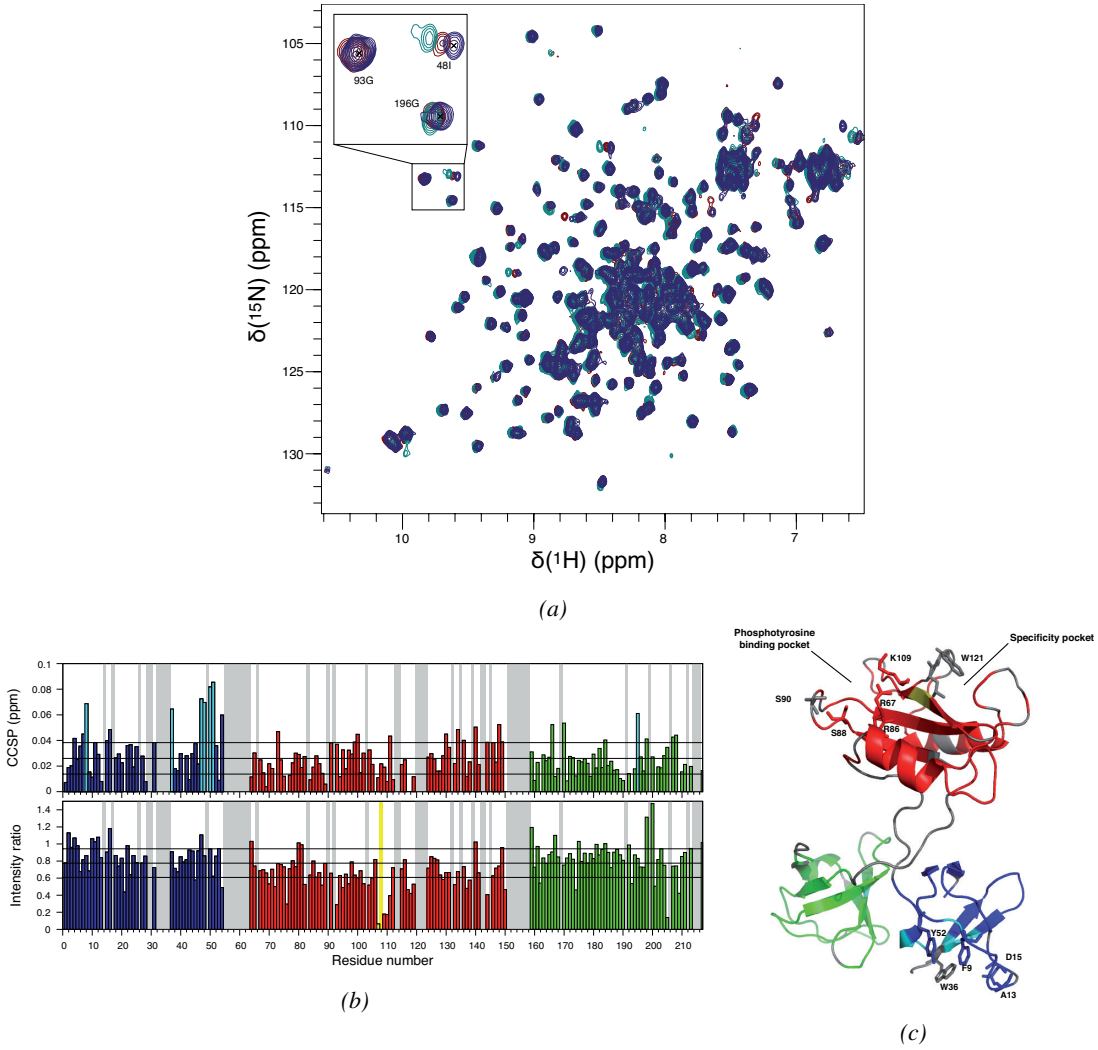


Figure 64 – Titration of Grb2 with unphosphorylated peptide P1. (a) ^{1}H-^{15}N HSQC of Grb2 alone (dark blue), with 5 equivalents of peptide P1 (maroon) and with 20 equivalents of P1 (green). A zoom shows peaks in the NSH3 (I48), SH2 (G93) and CSH3 (G196) domains. (b) Mapping of the differences between the initial and final titration points ^{1}H-^{15}N HSQC. Grey: missing residues; yellow: residues for which the intensity ratio (bottom) is < 0.1; cyan: residues for which the ^{1}H-^{15}N CCSP (top) is > 0.06 ppm. Yellow bands correspond to disappearing peaks. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed. (c) Color-coded mapping of the perturbations onto Grb2 crystal structure (PDB 1GRI). The residues known to be important for ligand binding in the NSH3 and SH2 domains are shown as sticks and named.

9.3.2 Phosphorylated Y-PxxP peptide P2

The titration of Grb2 with the phosphorylated peptide P2 was performed in the same conditions as the titration with the unphosphorylated peptide P1, up to a Grb2:P2 ratio of 1:20. The NSH3 domain is, as for P1, experiencing fast exchange. However, with this phosphorylated peptide, the SH2 is in slow exchange be-

tween the free and the bound form: the disappearance of the SH2 peaks are correlated with the appearance of peaks corresponding to the bound form. The overlaid HSQC spectra of the initial (1:0 ratio), final (1:20 ratio) and one intermediate (1:5 ratio) titration points are shown in figure 65a. The CCSP and intensity ratios of the free form (peaks that are disappearing) between the free form and the final titration point are shown in figure 65b. The mapping of the peaks in the fast and slow exchange regimes is shown in figure 65c.

The interaction of the NSH3 is similar to the binding to the unphosphorylated peptide P1: the same residues, corresponding to the expected binding site, are affected. Regrettably, uncertainties on the quantity of peptide added prevented reliable comparison of the estimated dissociation constant. We suspect that the quantity of added peptide P2 is lower than calculated, preventing the analysis of a potential change in K_D of the NSH3 between P1 and P2.

Contrary to what is observed for the NSH3, the kinetics of the SH2 binding dramatically changes when the peptide is phosphorylated. The exchange between the free and bound SH2 domain is slow when the peptide is phosphorylated, which is probably associated with a lower K_D compared to binding with the unphosphorylated P1 peptide. The binding sites involved in binding to P1 are also involved in binding to P2, but additional residues are taking part in the interaction, notably in the regions of R67, R86, S88 and S90, with are known to form the binding pocket for the phosphate group of the phosphorylated tyrosine. W121 is known to close the specificity pocket for the arginine 2 residues C-terminal of the tyrosine in the consensus Grb2-binding motif. Even though its peak is assigned, it unfortunately is too weak at our working concentration of 100 μM for the precise measure of its intensity or chemical shift perturbation, but it completely disappears upon binding. So far, the assignment of the bound form has not been done, but it will be a valuable source of information.

The precise analysis of the titration curves has not been done yet, but it will also tell us more about the simultaneity of the binding of the two domains.

9.4 Possible consequences for Grb2-mediated signal transduction

We obtained a first description at the atomic level of the binding between Grb2 and ErbB2. We showed that the SH2 domain of Grb2 interacts with the expected autophosphorylation site tyrosine B of ErbB2, even in the unphosphorylated state, with a high increase in affinity upon phosphorylation of the tyrosine. We also

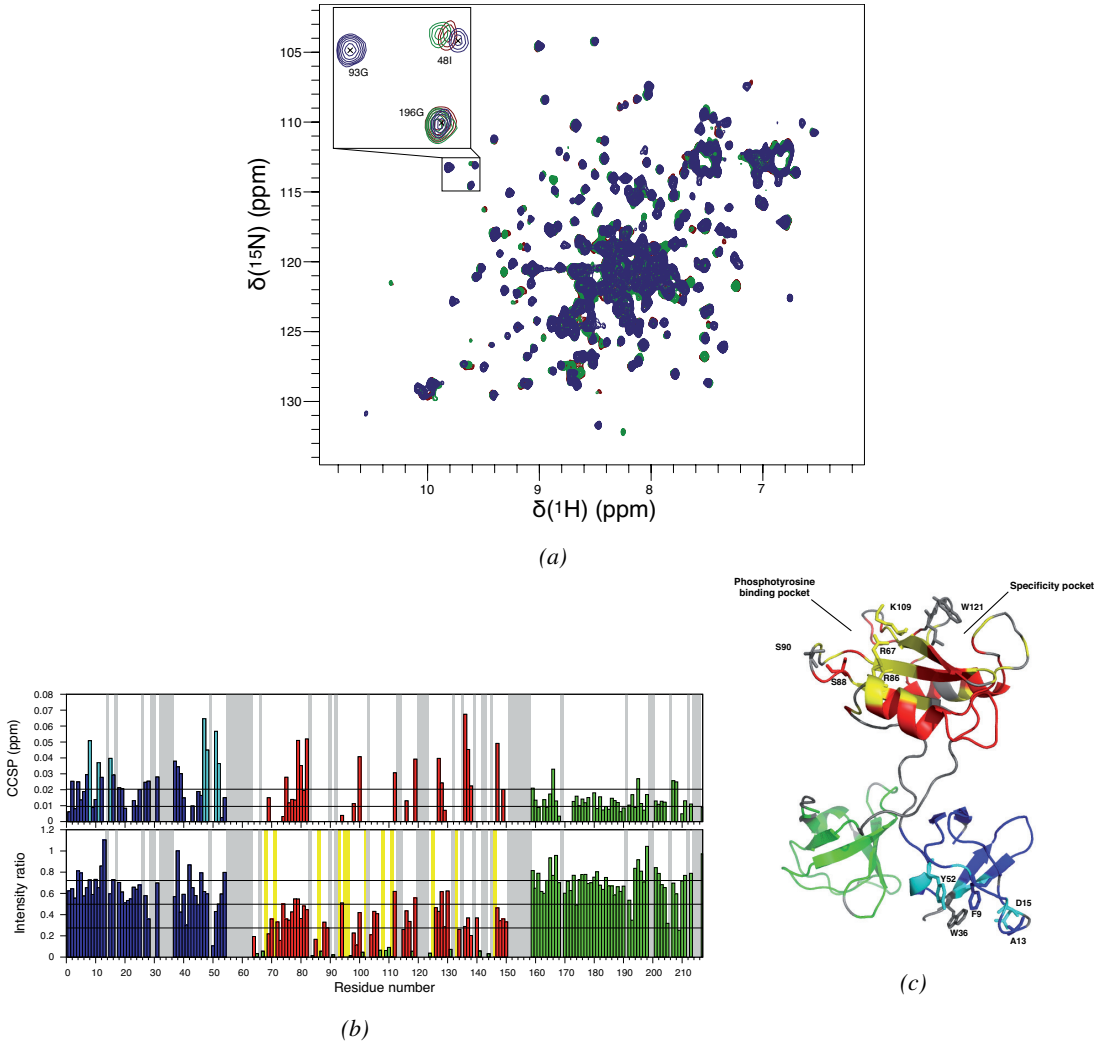


Figure 65 – Titration of Grb2 with phosphorylated peptide P2. (a) ^{1}H-^{15}N HSQC of Grb2 alone (dark blue), with 5 equivalents of peptide P2 (maroon) and with 20 equivalents of P2 (green). A zoom shows peaks in the NSH3 (I48), SH2 (G93) and CSH3 (G196) domains. (b) Mapping of the differences between the initial and final titration points ^{1}H-^{15}N HSQC. Grey: missing residues; yellow: residues for which the intensity ratio (bottom) is < 0.1; cyan: residues for which the ^{1}H-^{15}N CCSP (top) is > 0.035 ppm. Yellow bands correspond to disappearing peaks. Lines indicate mean+std (standard deviation), mean and mean-std values after outliers were removed. (c) Color-coded mapping of the perturbations onto Grb2 crystal structure (PDB 1GRI). The residues known to be important for ligand binding in the NSH3 and SH2 domains are shown as sticks and named.

showed that the N-terminal SH3 domain of Grb2 interacts with ErbB2 via a RPxPPxPR segment just C-terminal after the tyrosine. This secondary interaction could increase the overall affinity of Grb2 for ErbB2. This is consistent with the observation by Xie et al. (1995) using pull-down assays that the SH2 domain of Grb2 alone binds ErbB2 more weakly than full-length Grb2. The C-terminal SH3 domain of Grb2 does not seem to be affected by the interaction.

9.4.1 Selectivity of CtErbB2 polyproline motif for the NSH3 of Grb2

The RPxPPxPR segment of ErbB2 contains a RxxPxxP class I polyproline motif that has been shown by Guerlesquin and colleagues to be able to bind the SH3 domain of Fyn (Bornet et al., 2014). Even though many PxxP motifs are found in CtErbB2 sequence, as shown in figure 50, there is no canonical class II motif (PxxPxR) or other class I motif. Both NSH3 and CSH3 have been shown to be able to interact with class I motifs, even though for both domains the class II motifs have been more often characterized (see figure 32) (Musacchio, 2002).

Bornet et al. (2014) noticed that the RPxPPxPR segment does not bind the isolated CSH3 domain of Grb2, consistently with our results showing selectivity towards the NSH3 domain in FL Grb2. McDonald et al. (2009) observed that in Sos, the binding of the CSH3 domain required several arginines adjacent to the PxxP motif, which is not the case in the motif of CtErbB2 studied here. The only structure of the CSH3 bound to a class I motif (PDB 1IO6) involves a highly positively charged polyproline motif too (RHYRPLPPLP). Binding of the CSH3 to RxxK motifs has also been reported (Harkiolaki et al., 2009), and no such RxxK motif can be found in CtErbB2. The different selectivities of the NSH3 and CSH3 domains of Grb2 therefore seem to rely on electrostatic interactions.

9.4.2 Importance of the organization of Grb2 domains for its interactions

The Grb2 interaction sites described here are far from the linkers and from the regions of each domain that we identified to be affected by each other in FL Grb2 (see figure 55). Binding of each domain to ErbB2 is therefore not expected to be directly hindered by the presence of the other domains. However, given that the tyrosine and polyproline sites in CtErbB2 are very close from one another (only 6 residues between the tyrosine and the first arginine), the relative position of the NSH3 and SH2 domains is important to understand how binding occurs. The relative position of the NSH3 and SH2 domains in solution, as we determined by SAXS (see section 8.4), is compatible with their relative position in the crystal structure. In this structure, their respective binding pockets are too far from one another (as seen in figure 65c) for simultaneous binding to the polyproline and tyrosine motifs of the peptide. The domains have to be arranged differently in the bound form if their binding is simultaneous. Inter-domain dynamics could also be altered by binding.

9.4.3 Revisiting the model of signal transduction by Grb2

The current model of signal transduction by Grb2 is shown in figure 66: while the SH2 domain interacts with RTKs, and in particular ErbB2, the SH3 domains interact with downstream partners for signaling. Sos preferentially binds to the NSH3 (Chardin et al., 1993; McDonald et al., 2009), while Gab1 (Lewitzky et al., 2001) and Gab2 (Harkiolaki et al., 2009) bind to the CSH3. The interaction that we showed between the NSH3 of Grb2 and CtErbB2 polyproline motif has a lower affinity (mM range) compared to the NSH3-Sos interaction (μM range, measured by Ogura and Okamura (2013)). However, it could beneficiate from the increased local concentration due to the SH2-pY interaction, and compete with binding to the NSH3 in certain conditions, regulating Sos-dependent signaling.

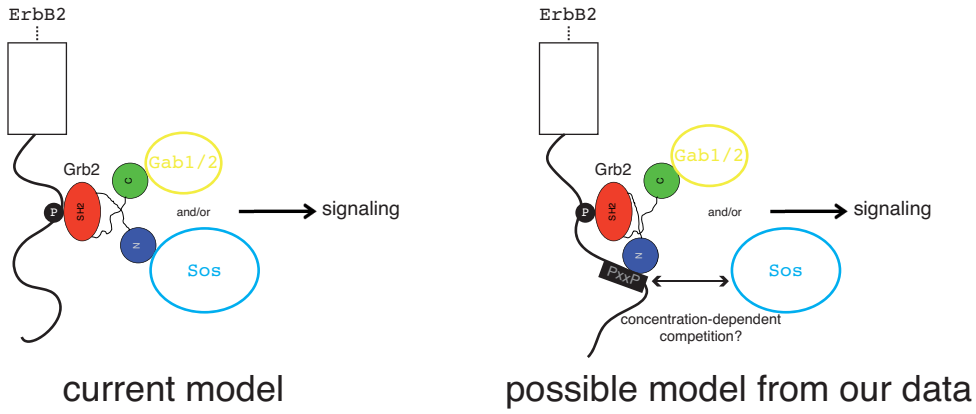


Figure 66 – Modification of the current model of Grb2 signal transduction. Only the intracellular region of ErbB2 is represented for clarity.

We did not study how the binding to Grb2 affects ErbB2 peptides conformation. While the PxxP motif is expected to be stabilized in a PPII helix upon interaction with the SH3 domain, the phosphotyrosine segment is expected to adopt a turn in its C-terminus. Whether and how these two expected local conformations are actually formed remains to be investigated. More globally, how the CtErbB2 conformational ensemble is affected by its binding to Grb2 is completely unknown, but could regulate other interactions, and impact signaling. Besides, we saw that sites outside of the Y_B/RPxPPxPR region are affected by the interaction with Grb2 in unphosphorylated CtErbB2. Whether it still holds when CtErbB2 is phosphorylated remains to be

investigated. For all these regions, we would need to obtain tyrosine phosphorylated samples of CtErbB2, and not just peptides. This is the focus of the next section.

10 Towards the characterization of functional states of CtErbB2: challenges of structural and functional studies of tyrosine phosphorylation

Post-translation modifications are a major regulation mechanism in the cell, and are particularly important in signaling. They allow context-dependent expansion of the protein readout beyond the common 20 amino acid. Amongst them, phosphorylation is the most commonly observed one, and the second one in terms of predicted sites (after N-glycosylation), from the Swiss-Prot database (Khoury et al., 2011). Phosphorylation can occur on the side-chains of serines, threonines and tyrosines (and much more rarely on histidines and aspartates). While serines and threonines share the same kinases, that were discovered in the 50s (Burnett and Kennedy, 1954), tyrosines are phosphorylated by specific kinases, namely tyrosine kinases, that were only discovered in 1979 with Src (Eckhart et al., 1979). Over the years, it has been shown that many tyrosine kinases are receptors, and to date 90 tyrosine kinases have been identified in the human genome (for a total of 525 kinases), amongst which 58 receptor tyrosine kinases (RTKs) (Manning, 2002). Most tyrosine kinases, and especially RTKs, are activated by transphosphorylation (of a tyrosine in their activation loop or juxtamembrane region) in oligomers (Lemmon and Schlessinger, 2010). In that regard, ErbB receptors are unique since dimerization-induced activation is not mediated by transphosphorylation of the kinase domains, but by inter-domain interaction (Stamos et al., 2002; Zhang et al., 2006).
Many essential processes, often linked to transmembrane signal transduction, depend on tyrosine phosphorylation (Hunter, 2009). Phosphotyrosines (pY) are less abundant than phosphoserines (pS) and phosphothreonines (pT): as of July 2019, from Swiss-Prot, there are 46973 predicted serine, 12599 predicted threonine, and 9282 predicted tyrosine phosphorylation sites in eukaryotic proteins (`http://selene.princeton.edu/PTMCuration/` (Khoury et al., 2011)). Additionally, phosphotyrosines have short lifetimes in the cell, due to high turnover of tyrosine phosphatases, leading to a contribution of pY to the total number of phosphosites in proteins of animal cells lower than 1% (Hunter and Sefton, 1980). All in all, tyrosine kinases and phosphorylation remain far less characterized than serine and threonine kinases and phosphorylation.

In CtErbB2, for interaction studies on a single phosphotyrosine motif, without taking into account the effect of the full-length protein, chemical synthesis of peptides was the way to go. However, to study the effect of phosphorylation on the conformation, and interactions, of the full-length C-terminal tail, other strategies had to be designed. This is very challenging due to the nature of the phosphorylation (on tyrosines), the size

of the protein and multiplicity of phosphosites, and the complexity of the native auto-transphosphorylation process. In this section, I present the associated specific challenges, and the three methods to overcome them that we have tried to apply in the lab, in parallel to our study of unphosphorylated CtErbB2.

10.1 Challenges and existing strategies

10.1.1 Site-resolved detection and study of (tyrosine) phosphorylation

One of the great challenges of the structural study of multiple phosphorylation is its detection at a resolution allowing information on each phosphorylation site. Here I present the most common and relevant methods for such detection, available in laboratories such as the ICSN, that could be used to study CtErbB2.

- Antibodies. Phosphospecific antibodies are the common technique to detect phosphorylation in cell biology studies. However, either they detect phosphorylation of an amino acid without sequence specificity (any pS or pT or pY) and do not give any information on the specific phosphorylation sites, or we need site-specific antibodies for all positions. This is cumbersome, does not really allow real-time monitoring, and very close phosphorylation sites (such as Y_{D1} and Y_{D2} in CtErbB2) can usually not be discriminated.

- Mass spectrometry. This is one of the most used technique to detect phosphorylation, since the addition of a phosphate group is expected to be easily detectable with an addition of 80 Da to the protein mass. However, site-resolved information is only accessible (routinely) with a proteolysis step. Depending on the protein sequence and structure, and protease(s) available, not all sites information can be obtained. Several problems combine for the study of CtErbB2 with this technique: proline-rich proteins are more difficult to proteolyse in enough peptides to cover the sequence; sequence repetitions, found in particular around CtErbB2 tyrosines, make the identification of cleaved peptides difficult; and adjacent sites such as D_1 and D_2 are virtually impossible to discriminate. An additional problem is that ionization of phosphorylated proteins is less efficient than ionization of their unphosphorylated counterpart, preventing accurate quantification. Enrichment is often needed (Thingholm et al., 2009). There is also a problem of isotopic labeling when working on NMR samples of large proteins, as in our case: isotopic heterogeneity can mask the phosphorylation information. Complex labeling strategies to identify and quantify phosphorylation exist (Przybylski et al., 2010), but require very specific expertise and are costly.

- NMR. It can give residue-specific, quantitative information about phosphorylation. NMR also has the major advantage to allow phosphorylation monitoring in real-time, in cases where the phosphorylation is slow enough to be followed by sequential experiments (in the order of 10 min for a "standard" HSQC, down to the min or even s timescale with fast experiments and concentrated samples). Given that phosphorylation sites are often located in disordered regions (Iakoucheva et al., 2004), NMR is all-the-more relevant for these studies. In the last few years, several groups have been studying phosphorylation by NMR, notably Selenko and colleagues (Selenko et al., 2008; Liokatis et al., 2010; Theillet et al., 2012; Mylona et al., 2016), Lippens and colleagues (Landrieu et al., 2006), Wolff and colleagues (Cordier et al., 2012, 2015), and Brutsher and colleagues (Sólyom et al., 2015), all focusing on serines and threonines. Few NMR studies exist on tyrosine phosphorylation (Deshmukh et al., 2011; Rosenlöw et al., 2014; Mayzel et al., 2014). Examples of how serine, threonine and tyrosine phosphorylation can be looked at by NMR is given in figure 67. There as some specific challenges associated with the NMR study of phosphorylation in our system:
  - By ^{1}H-^{15}N NMR: This is a very good way of observing serine or threonine phosphorylation, because their phosphorylated and unphosphorylated forms have very different backbone ^{1}H and ^{15}N chemical shifts, and the phosphorylated form has a peak outside of the crowded region of the spectrum (see figure 67 a and b). The case of tyrosine is unfavorable, since the phosphate group is quite far from the backbone and induces relatively small ^{1}H-^{15}N shifts (see figure 67 c and d). It has made detection of phosphorylation difficult to distinguish from subtle changes in conditions such as pH or temperature with standard HSQCs. Besides, phosphotyrosine peaks are located in already crowded regions of the spectrum.
  - By ^{13}C NMR: This is a way to get information close (CH_ε) to the actual addition of the phosphate group on tyrosines and have larger chemical shift changes upon phosphorylation (see figure 67d). Unluckily, the disordered nature of CtErbB2 causes superimposition of tyrosine side-chain carbons on ^{1}H-^{13}C HSQC, and requires to record time-consuming 3D spectra to have site-resolved information.
  - ^{31}P NMR: Being more sensitive than ^{13}C NMR, and giving signal only for phosphorylated residues in a protein sample (plus potential phosphate from the buffer), this leads to quite a sensitive detection of phosphorylation. However, the assignment of multi-phosphorylated proteins is not straightforward.

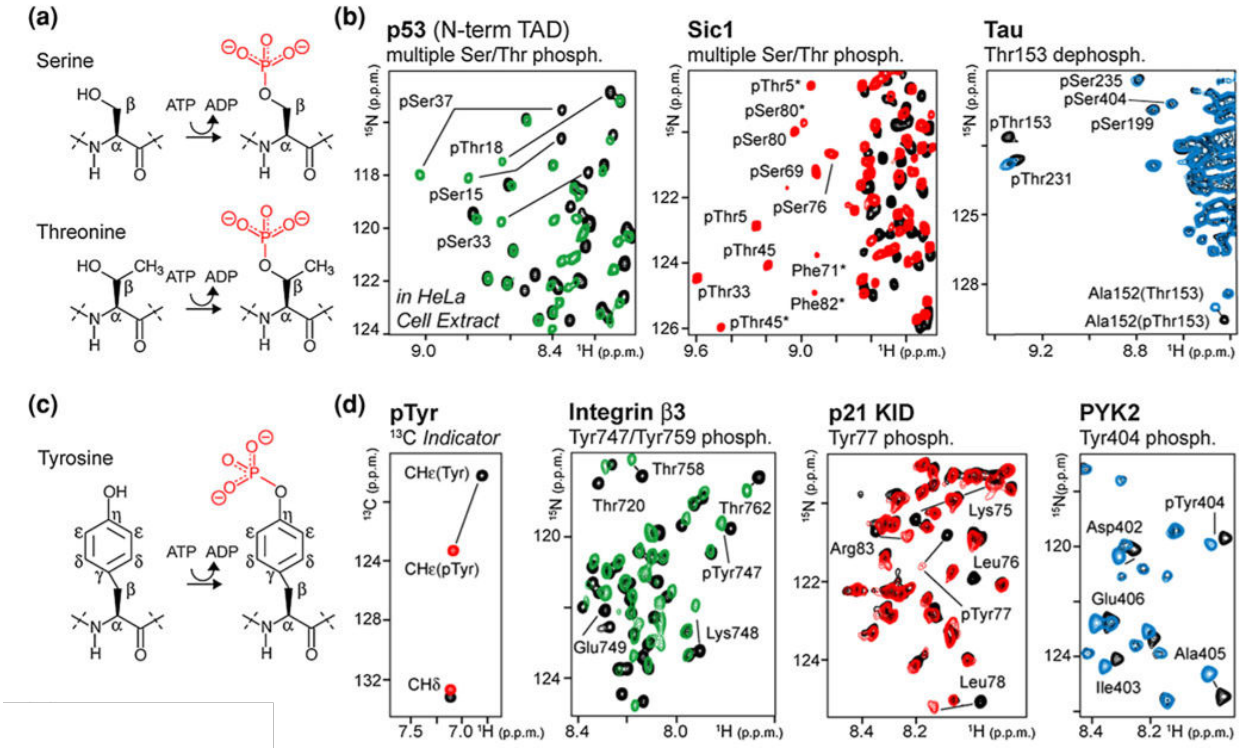


Figure 67 – Serine, threonine (a, b) and tyrosine (c,d) phosphorylation studied by NMR. Figure from Theillet et al. (2012).

Additionally to the problem of obtaining information on the phosphorylation state of each tyrosine site, there is also a major problem of heterogeneity of the sample that comes with many multiphosphorylated protein samples. This heterogeneity complicates the study of the effect of precise phosphorylation patterns.

10.1.2 Production of phosphorylated samples with multiple sites: site-directed synthetic approaches and biological relevance

We have seen that obtention of site-resolved information on a multiphosphorylated, heterogeneous sample is not easily done. Yet, biologically relevant samples are likely to be multiphosphorylated and heterogeneous. One way around would therefore be to generate samples with controlled phosphorylation patterns, to understand the role of each site, followed by the description of more complex and biologically relevant samples.

To produce site-directed phosphorylated proteins, three main approaches exist (Chen and Cole, 2015):

- The use of natural amino acid phosphomimetics introduced by site-directed mutagenesis. Phosphoserines and phosphothreonines have been mimicked by aspartates and glutamates since the end of the 1980s (Thorsness and Koshland, 1987). However, the validity of this approach is restricted by the structural and chemical difference between the phospho-amino acid and the mimic. For example, while Glu and Asp only have one possible negatively charged state, phosphate has two. Mimicking phosphorylation with Glu or Asp has been proven to be inaccurate in several instances, for example in α-synuclein (Paleologou et al., 2010) and Bcl-2 (Dai et al., 2013). Besides, these phosphomimetic residues are even more structurally different from pY than from pS and pT. We therefore dismissed this strategy. The reverse strategy, preventing phosphorylation by incorporation of a non-phosphorylatable analog, is more interesting for tyrosines: like alanine can be considered a non-phosphorylatable structural equivalent of serine, phenylalanine would be the non-phosphorylatable tyrosine. This approach has already been applied to RTKs from the PDGFR (Valius and Kazlauskas, 1993) and ErbB family (Dankort et al., 1997) in cell biology studies.

- Total chemical synthesis of proteins. It allows site-specific incorporation of any type of phosphorylated amino acid, but is generally limited to proteins of less than 60 residues. Ligation can be performed to then obtain longer proteins, but often requires peptides with N-terminal cysteines (Dawson et al., 1994). Furthermore, for long proteins with multiple phosphorylations, iterative ligations implies increased complexity, lower yields, and therefore higher costs. NMR studies would also require to isotopically label the synthesized peptide(s), further increasing the cost. We have thus restricted our use of this strategy to the small unlabeled peptides for titrations.

- The incorporation of phospho-residues as unnatural amino acids using a STOP codon. This technique has been tested for pS (Park et al., 2011), and more recently for pT (Zhang et al., 2017) and pY (Hoppmann et al., 2017). It is still a technique that requires specific expertise, that we do not have in the lab.

These approaches can be complemented by production of more biologically relevant phosphorylated form(s) of the protein of interest. When a protein is overexpressed recombinantly in the usual *E. coli* system, it is not phosphorylated. When the protein contains its own kinase domain, such as ErbB2, high-level expression of active proteins in *E. coli* is challenging, because of lack of solubility, misfolding and/or aggregation (Caspers et al., 1994; Rosano and Ceccarelli, 2014). Good levels of autophosphorylation are difficult to obtain (Shrestha et al., 2012; Bhoir et al., 2018). Other systems can be used, amongst which:

- *In vitro* phosphorylation of a recombinantly produced non-phosphorylated protein by either a recombinant, purified kinase, or by cell extracts. These two phosphorylation methods were monitored and compared by NMR by Wolff and colleagues (Cordier et al., 2012, 2015).

- For autophosphorylated proteins which contain their own kinase, overexpression in eukaryotic expression systems. Many kinases are produced this way, and in particular ErbB2 kinase is often overexpressed in insect cells using the baculovirus expression system (Brignola et al., 2002b; Fan et al., 2008; Monsey et al., 2010), including the commercial ones by Life technologies™ and PRECISIO™.

We are trying in the lab to combine all those approaches and apply them to the study of CtErbB2 phosphorylation.

10.2 Three methods to produce phosphorylated CtErbB2

We designed three methods to phosphorylate CtErbB2. All three have different challenges, advantages, drawbacks and costs.

10.2.1 CtErbB2 co-expressed with a tyrosine kinase in *E. coli*

To try and produce large amounts of tyrosine-phosphorylated CtErbB2, we used the Agilent TKB1 Competent Cells (Catalog #200134). This chemically competent strain, derived from the *E. coli* BL21(DE3) strain, allows for inducible expression of amounts of an unknown tyrosine kinase protein sufficient to phosphorylate a co-expressed tyrosine-containing protein, here our CtErbB2 construct. This method has the advantages of *E. coli* protein overexpression: it is fast, and allows for production of large quantities of proteins. Compared to a common *E. coli* production, the cost is higher, for a single transformation costs about 100 euros in bacteria. It also allows for the desired isotopic labeling in a manner overall similar to usual *E. coli* protocols, even though the protocol has to be adapted. The major drawback of this production technique is the total absence of control on the phosphorylation pattern, since no information is given by the supplier on the kinase nature and specificity. Moreover, the vector carrying this kinase is unknown, making it difficult to optimize the protocol.
The TKB1 strain allows for IPTG-induced production of a target protein, here CtErbB2, when transformed with the appropriate plasmid (since the strain carries the gene for T7 RNA polymerase under the control of the *lacUV5* promoter). Not much information is given on the induction mechanism of the tyrosine kinase, except "The TKB1 strain harbors a plasmid encoded inducible tyrosine kinase gene". The induction medium

for the kinase being tryptophan-free in the supplier's protocol, and containing indoleacrylic acid (see figure 68a), we think its expression is under control of the tryptophan operon. In addition to the antibiotic that is specific to our plasmid (kanamycin), the transformation and culture have to be done with tetracycline according to the Agilent protocol, suggesting it selects for bacteria containing the TK plasmid.
The optimization of the protocol has been done with Dr. Nadine Assrir, and I supervised Ndèye Marième Diouf, who was a bachelor student in the lab between June and August 2018, and Alan Even, a master student in the lab between March and July 2019, on this part of the project.

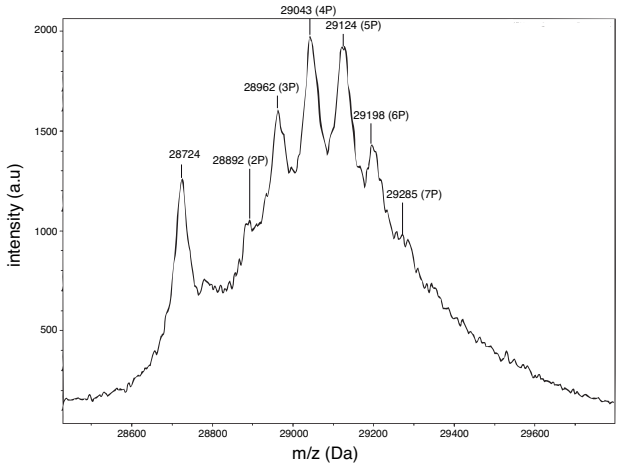
Optimization of the expression/purification protocol For expression of unlabeled phosphorylated CtErbB2, the Agilent manual of the TKB1 Competent Cells was followed. Briefly, the transformation protocol was similar to the one used for transformation of CtErbB2 in standard BL21 cells (see section 4.1.2), except β-mercaptoethanol was added to a concentration of 25 mM to the cells before transformation, and selection was maintained with both tetracycline and kanamycin. For expression of CtErbB2 and the tyrosine kinase, a single colony was transferred to 2 mL LB with both antibiotics and grown overnight at 37 °C with shaking. The next morning, the preculture was transferred to 50 mL LB with antibiotics until OD_{600nm} = 0.6. Expression of CtErbB2 was induced with 1 mM IPTG for 3 hours. The culture was then centrifuged at 2000 g and the pellet was resuspended in 100 mL TK induction medium (see figure 68a), and further incubated overnight at 37 °C with shaking. The protein was purified following the usual protocol for CtErbB2 (see section 4.1.3). It was then analyzed using MALDI-TOF mass spectrometry and showed several phosphorylation states, from unphosphorylated to phosphorylated 7 times (maybe more in small proportions), as shown in figure 68b. This number is intermediate between the total number of tyrosines (9) and the number of known autophosphorylation sites (5).

The production in a labeled medium is more complicated, since it requires to change the medium to a M9 medium even for the induction of CtErbB2 and not only for the induction of the kinase. This seems trivial, but not knowing the kinase plasmid makes it difficult to optimize. The M9 medium that we used was the TK induction medium given in figure 68a with NH_4Cl replaced by $^{15}NH_4Cl$ and removal of casamino acids. Depletion in tryptophan in such medium, if the kinase expression is indeed controlled by a tryptophan operon, may trigger expression of the kinase even in this absence of indoleacrylic acid. Therefore, we chose to induced the production of our protein and of the kinase simultaneously, changing the medium to the TK medium plus IPTG right after the OD_{600nm} reached 0.6. Although we tried a lot of different conditions, changing the induction OD_{600nm}, temperatures, times, and order of addition of IPTG and indoleacrylic acid,

TK Induction Medium
Modified 5× M9 Medium (per Liter)
30 g of dibasic sodium phosphate (Na_2HPO_4)
15 g of monobasic potassium phosphate (KH_2PO_4)
5 g of ammonium chloride (NH_4Cl)
2.5 g of NaCl
15 mg $CaCl_2$
Autoclave

1× TK Induction Medium (per Liter)
200 ml of modified 5× M9 medium
1 ml of 1 M $MgSO_4 \cdot 7H_2O$ (autoclave)
10 ml of 20% (w/v) glucose (filter sterilize)
5 ml of 20% casamino acids (filter sterilize)
0.1 ml of 0.5% thiamine-HCl (filter sterilize)
4 ml of 2.5 mg/ml indoleacrylic acid stock
1 ml of 30 mg/ml kanamycin [antibiotic selection for **target plasmid** (filter sterilize the stock solution and store at −20°C)
1 ml of 12.5 mg/ml tetracycline [mix tetracycline in 50% (v/v) ethanol and store at −20°C]
Dilute to 1 liter with sterile water

(a)



(b)

Figure 68 – Production of unlabeled phosphorylated CtErbB2 in the TKB1 strain. (a) TK induction medium composition for 1L. Reproduced and adapted from the Agilent TKB1 Competent Cells instruction manual (ampicillin has been changed to kanamycin, our selection antibiotic). (b) MALDI-TOF spectrum of unlabeled phosphorylated CtErbB2 expressed in the TKB1 strain and purified.

we could not reproduce our results with the unlabeled protocol. In some cases we had high levels of proteolysis (revealed by mass spectrometry, data not shown), and in no case were we able to unambiguously show presence of phosphorylation.

From the results we had on the unlabeled protein, we should be able to find a protocol to produce a labeled, phosphorylated protein. This is work in progress in the lab. In the meantime, the production of unlabeled protein is still interesting for interaction studies.

10.2.2 *In vitro* phosphorylation followed by NMR

To try and obtain a phosphorylated form of CtErbB2 closer to the native phosphorylation pattern, and be able to follow the kinetics of phosphorylation, we decided to use the ErbB2 kinase to phosphorylate our unphosphorylated construct of CtErbB2 expressed in *E. coli* in a intermolecular fashion. This approach respects the native phosphorylation specificity. However, since the topology is different from the topology in the activated dimers that trigger *trans*-autophosphorylation, we do not expect the pattern to be exactly similar to the native one.

The commercially available kinase construct (from Life technologies^{TM}, Catalog #PV3366) is produced in Sf9 insect cells with a baculovirus expression system, and contains the whole intracellular part of ErbB2,

including the juxtamembrane segment, kinase domain and C-terminal tail (residues 676-1255). It is quite expensive (468 euros/10 μg). We observed by mass spectrometry that this kinase is not catalytic: the amounts needed for significant phosphorylation levels are high (we managed to have significant levels of phosphorylation at a kinase:protein ratio of 1:2, with different levels of phosphorylation from 0 to +4, data not shown). This is consistent with literature reporting a relatively low activity of the ErbB2 kinase (Brignola et al., 2002b; Monsey et al., 2010). From this observation, we adapted the ratio of kinase compared to CtErbB2 for our NMR study, to have a relatively high kinase:CtErbB2 ratio, while keeping CtErbB2 concentration high enough for reasonable acquisition time for 2D experiments, and reasonable cost in commercial kinase. The protocol is given below.

Protocol The phosphorylation was done with 35 μM CtErbB2 in 200 μL of the following buffer: 20 mM Tris-HCl pH 7.2, 50 mM $MgCl_2$, 1 mM ATP, 2 mM TCEP, 0.01 % CHAPS, 5% D_2O and Halt^{TM} Protease and Phosphatase inhibitor Cocktail (Thermo Scientific). Manganese is usually used in phosphorylation buffers. However, it had to be replaced by magnesium because manganese is paramagnetic and causes severe broadening in NMR spectra. The kinase was added in a 1:50 ratio (10 μg in 25 μL of buffer composed of 50 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.01% CHAPS, 2 mM DTT, 50 % glycerol). Just after addition, a series of ^{1}H-^{15}N HSQC (2h) and ^{1}H-^{13}C HSQC (33 min) were recorded for about 24h at 298 K, at 950 MHz.

The commercial kinase is contaminated with S/T tyrosine kinase(s) The spectra that we obtained indeed showed what looks like phosphorylation but only on serines and/or threonines (see figure 69). The Ser/Thr region of the spectrum shows peaks disappearing and appearing, and in particular one peak appears in the region of pSer/pThr. This was observed on both experiments that we performed, with two different batches of the commercial kinase. Even with a relatively pure kinase (>70% according to the manufacturer), the high quantity that we had to use due to poor activity of the kinase increases the chance that much more active residual Ser/Thr kinases phosphorylate our protein.

A solution to overcome this problem would be to produce our own ErbB2 kinase, to be able to control its quality, and reduce cost.

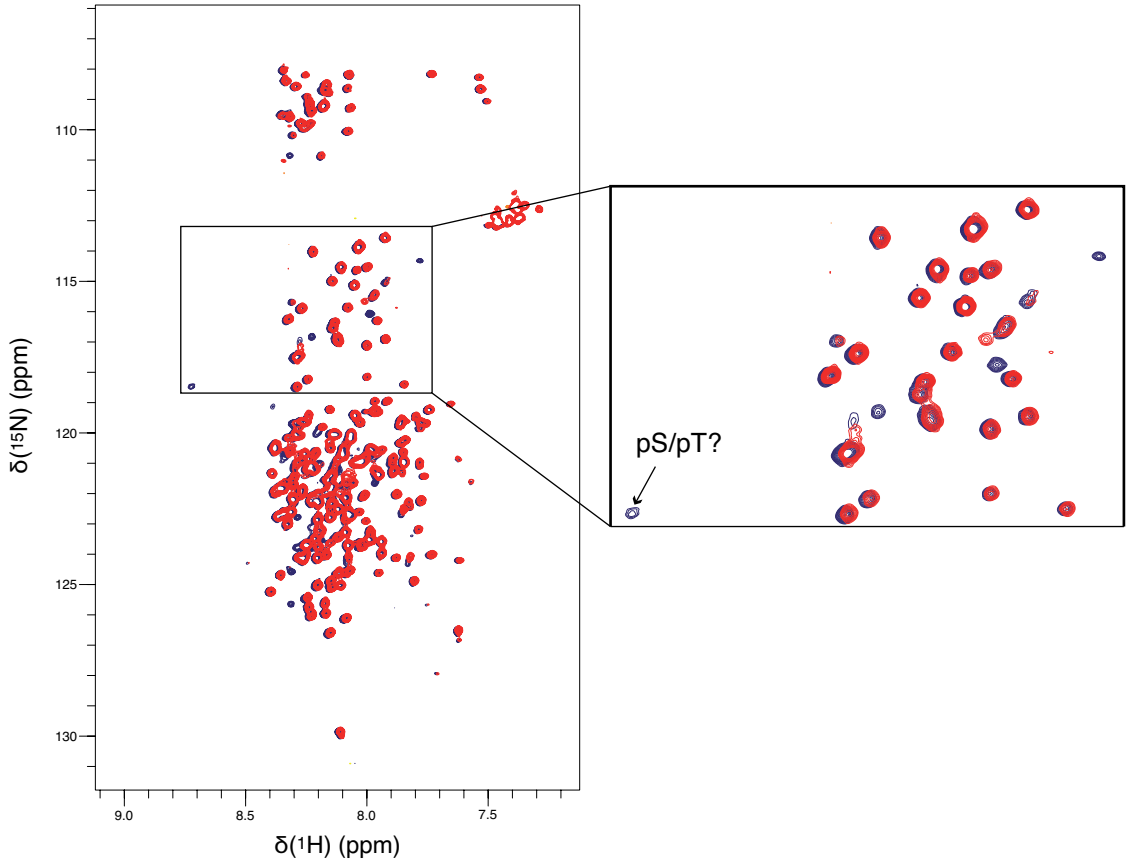


Figure 69 – First (red) and final (blue) ^{1}H-^{15}N HSQC spectra of the kinetic study of in vitro phosphorylation of wild-type CtErbB2 by the commercial ErbB2 kinase. Conditions are given in the protocol section.

10.2.3 Intracellular domain of ErbB2 expressed in Sf9 cells using the baculovirus expression system

The two approaches presented in section 10.2.1 and 10.2.2 are both ways to phosphorylate CtErbB2 in a way that is either non-specific (with the unknown kinase co-expressed in *E. coli*, section 10.2.1) or only partially specific (with the ErbB2 kinase lacking the specificity conferred by the topology of the ErbB dimer). It was shown that tyrosine kinase specificity is less dependent on the sequence of the substrate than Ser/Thr kinases specificity (Zhu et al., 2005), and is more dependent on the recruitment of the substrate (local concentration, accessibility). Reproducing ErbB2 autophosphorylation patterns thus necessitates to get as close as possible to the actual topology in which phosphorylation happens. Production of the whole ErbB dimers would be very challenging: the whole receptors are big (137.9 kDa for ErbB2) and are membrane proteins, which are not easy to solubilize for solution studies. They are not expected to be suitable for overexpression in *E. coli*. We therefore chose to try and produce the intracellular part of ErbB2, in insect cells using the baculovirus expression system. The construct (presented in figure 70a) was defined following the construct of the ki-

nase domain by Life technologies™, and contains the intracellular juxtamembrane region that is known to have an important role for dimerization of the kinase domains and activity. According to the tests by Life Technologies™, it should have an activity of about 0.1 nmol/min for 1 μg of kinase in 30 μL. Even though membrane anchoring increases ErbB2 kinase activity (by a factor of 3.5), it is not strictly necessary (Monsey et al., 2010). Being able to produce the intracellular ErbB2 would allow:

- The production of our own active kinase (unlabeled) to perform the *in vitro* phosphorylation described in section 10.2.2, but controlling the kinase purity and reducing the costs.
- The study of the behavior of autophosphorylated CtErbB2 in the context of the whole intracellular domain. This whole domain is big (around 65 kDa) for NMR. However, we hope that since the C-terminal tail was shown to be disordered even when attached to the kinase (Keppel et al., 2017), its favorable relaxation properties would yield good signal to noise ratio for this part. This kind of study of flexible parts of big proteins was already conducted on a viral nucleocapsid (Jensen et al., 2011), a DNA replication complex (De Biasio et al., 2015), and a nuclear receptor (Belorusova et al., 2016). This study would require that we isotopically label the protein. A protocol for labeling of protein produced in insect cells was developed in the lab a few years ago (Meola et al., 2014).
- The study of the kinase domain alone (residues 676-987) or the phosphorylated C-terminal tail alone (988-1255), made possible by insertion of a TEV cleavage site between the two domains in one of our constructs (that we call the TK-TEV-Ct construct, see figure 70a). We could compare the activity of the kinase domain with of without the tail, and compare the behavior of the phosphorylated tail with or without the kinase domain.

The principle of protein production in insect cells using the baculovirus expression system is given in section 4.2.1.

Constructs and vectors The intracellular domain (ICD) constructs correspond to the residues 676-1255 of the full-length ErbB2 receptor. We chose to put the purification His6 tag at the C-terminus as it is less likely to disrupt kinase activity that we want to preserve here.
The ICD insert for the pFastBac™, shown in figure 70b for the construct without the TEV cleavage site, and in figure 70c for the TK-TEV-Ct construct, comes from a pcDNA3 vector containing full-length ErbB2 provided by Dr. Ali Badache (CRCM, Marseille). The ribosome binding site was optimized for insect cells.

It was further modified to add the His6 tag, and to insert the TEV cleavage site for the TK-TEV-Ct construct. Cloning into the pFastBac^{TM} vector was done following guidelines of the Bac-to-bac® Baculovirus Expression System protocol (Invitrogen). The detailed protocols for the generation of the viral stock and its titration are given in sections 4.2.2 and 4.2.3. Viral stocks were generated for both constructs.

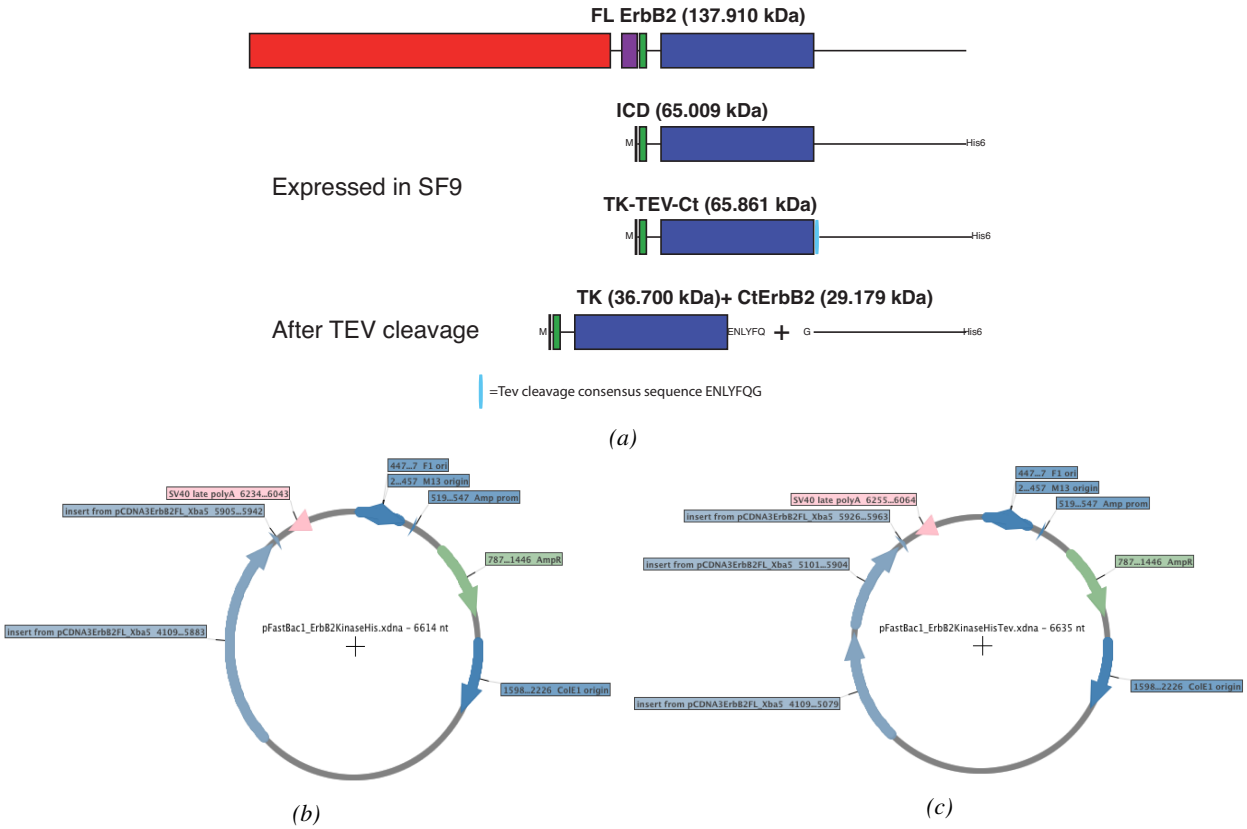
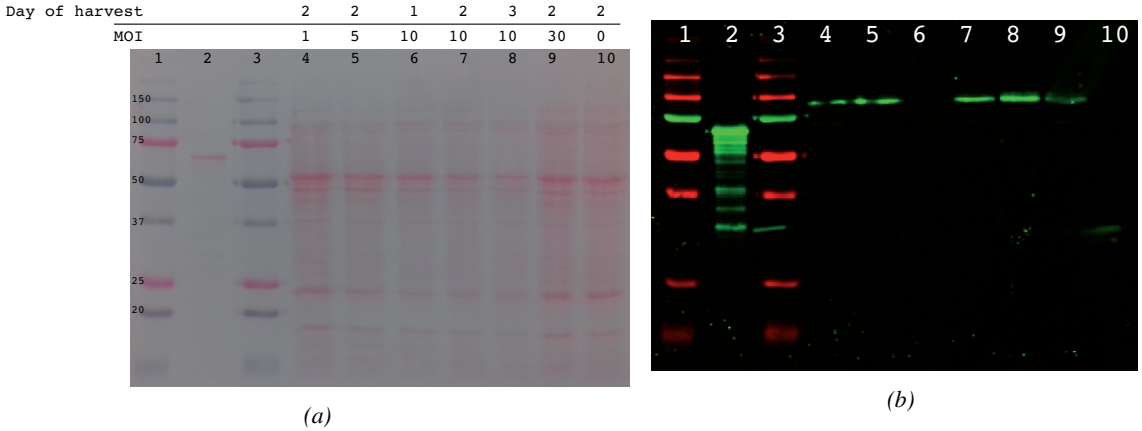


Figure 70 – ErbB2 constructs and pFastBac^{TM} vectors used for generation of Bacmids for expression in Sf9 cells. (a) The two protein constructs that we want to express in Sf9 cells, compared to the full-length ErbB2 receptor. The result of TEV cleavage is also given for the second construct (TK-TEV-Ct). (b) For the ICD construct. (c) For the TK-TEV-Ct construct.

Small-scale expression tests Once one has the viral stock at a known titer, the expression conditions can be optimized. In particular, two parameters can be tested: the MOI (multiplicity of infection), that is to say the virus to cell ratio for infection, and the expression time before harvest. The MOI has to be high enough for as many cells as possible to be infected, but without using too much of the viral stock for each production. The expression time, for non secreted proteins, must be long enough for a high quantity of proteins to be expressed, but with limited lysis so that the protein is not lost in the medium upon cell harvest. So far, we have done these tests only on the ICD construct.

3mL of Sf9 cells at 1.10^6 cells/mL were infected with the P3 viral stock corresponding to the ICD construct (3.10^8 PFU/mL). Different MOIs from 1 to 30, as well as different incubation times from 1 to 3 days before harvesting cells, were tested. The quantity of protein in the pellet was assessed by Western blotting (figure 71). Tests were not continued after day 3 because cell viability was going down dramatically. Before blotting, cells were lysed by sonication, the lysate was cleared by centrifugation, and enriched in His-tagged proteins by retention on Ni-NTA resin. Western blots were done on nitrocellulose membranes, using a primary anti-His tag mouse monoclonal antibody and a secondary IRDye® 800 CW Goat anti-mouse (IgG) antibody.



(a) (b)

Figure 71 – Expression tests for the ICD construct. Different MOIs were tested with harvest at day 2, and for MOI = 10, harvest at day 1, 2 and 3 was tested. (a) Ponceau staining of the membrane for the estimation of total protein loading and (b) Western blot revealing His-tagged proteins. In lane 2 is the positive control consisting of purified His-lipoyl-CtErbB2 (expressed in E. coli and purified as described in section 4.1, without cleavage of the His-lipoyl tag). Lanes 1 and 3 are molecular weight markers.

The anti-His Western blot shows a band between 75 and 100 kDa. This is consistent with the expected migration for our construct: its molecular weight is 65 kDa, but since the C-terminal tail is expected to remain very flexible, it is not surprising that it migrates at a higher apparent molecular weight. From the relative intensities of the bands (which should be normalized against total protein loading assessed from Ponceau staining shown in figure 71a), it seems that the amount of protein expressed does not dramatically increase when MOI is above 5. It also seems that the quantity of protein increases from day 2 to 3 even if lysis had begun to occur. The chosen conditions for expression of our protein would therefore be MOI = 5, and three days of expression before harvest.

We could not go further for now due to lack of time, but we have a proof of concept that we are able to

produce the ICD construct in Sf9 cells. Even though the construct does not recapitulate the native ErbB dimers, this would allow a leap forward in the understanding of the phosphorylation process by the ErbB kinase, and in the intramolecular interaction of the C-terminal tail with the kinase domain.

10.2.4 Overcoming the heterogeneity problem: using Tyr to Phe point mutants

We have three ways to produce phosphorylated CtErbB2, although optimization is still necessary for all three methods. Two problems persist: with our co-expression bacterial system (section 10.2.1), the phosphorylation pattern will not be specific of the ErbB kinases; and in all systems, heterogeneous phosphorylation on different sites will prevent easy characterization of the phosphorylated states.

We therefore chose to use tyrosine to phenylalanine mutations to selectively prevent phosphorylation on certain tyrosine sites, and select the sites that we want to keep accessible for phosphorylation. This approach is similar to the one that was used for several studies of ErbB2 phosphorylation (Dankort et al., 1997, 2001a,b; Marone et al., 2004). Our approach so far is to have mutants which can only be phosphorylated on one tyrosine (out of 6) autophosphorylation sites. These mutants with 8 point mutations (there are 9 tyrosines in total in CtErbB2) are called Y_A, Y_B, Y_C, Y_{D1}, Y_{D2} and Y_E and are described in section 4.1.1. These mutations are designed to selectively prevent phosphorylation without perturbing the behavior of the protein. To verify the latter, we assigned the 0Y mutant, where all the tyrosines are mutated to phenylalanines, by transferring the assignment from the wild-type and with the help of HNCACB, HN(CO)CACB and HNCO spectra. Then, we recorded ^{15}N T_1 and T_2 relaxation experiments to compare the dynamics of the mutant to that of the wild-type protein, at 950 MHz. The results are presented in figure 72. We noticed that the spectra recorded on the mutant had a small pH difference compared to the spectra recorded, as seen by a decreased intensity in the histidine-rich regions (around residue 50). These peaks becoming weak, we could not obtain reliable relaxation data on this region. Overall, the dynamic profile of the mutant is similar to the behavior of the wild-type protein. The same regions present higher R_2 and/or R_1 rates, even though the values are not exactly the same. These relaxation experiments being extremely sensitive to small changes in pH, temperature or small differences in the experiment parameters, we plan on recording new experiments, controlling these more closely, and also recording $\{^1$H$\}$-^{15}N nOes data. This would confirm the similarity of behavior of the 0Y mutant and the wild-type protein, validating our approach.

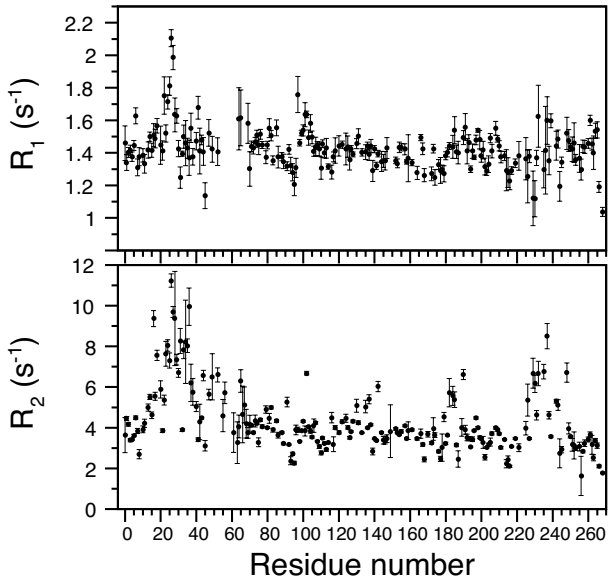


Figure 72 – ^{15}N R_1 and R_2 relaxation rates for the 0Y mutant of CtErbB2, measured at 22.3 T (950 MHz proton frequency).

The first mutants that we tried to use in our phosphorylation experiments described above, are the Y_A, Y_B and Y_D mutants. Although we did not manage to have any usable phosphorylated sample yet, these three mutants are particularly interesting. The phosphorylation of Y_A has been shown to inhibit the transformation potency of ErbB2 by Dankort et al. (1997). This could be an indirect effect, via recruitment of proteins involved in regulation mechanisms of ErbB2, or a more direct, conformational effect for example. The authors also showed that even though the negative effect of Y_A phosphorylation could not be rescued by phosphorylation of Y_B, Y_C or Y_E, simultaneous phosphorylation of Y_A and Y_D enabled to recover ErbB2 transformation potency. This suggests interdependence of the effect of phosphorylation on these two sites. This data, combined with our observation that the regions containing Y_A and Y_D experience a long-range contact, make these two sites particularly interesting. A Y_AY_D mutant, allowing simultaneous phosphorylation of the two tyrosines, would also be helpful. Our other mutant of particular interest, Y_B, would allow us to study the effect of phosphorylation on the interaction of Grb2 with Y_B, as done with our titrations with peptides, but in the context of the full-length protein. Since we showed that sites outside the primary phosphotyrosine motif are involved in the interaction, this could further improve our understanding of the interaction.

CONCLUSIONS AND PERSPECTIVES

In this work, we have characterized the structure and dynamics of the C-terminal tail of the receptor tyrosine kinase ErbB2 (CtErbB2). We have also undertaken the study of the two main phenomena that regulate its function: its phosphorylation and its phosphorylation-dependent interactions. We mainly used NMR to get atomic-resolution information, and we combined it with SAXS to obtain size and shape parameters.

First, we studied the conformational ensemble of isolated, unphosphorylated CtErbB2. No stable secondary or tertiary structure is present along the 268 residues of this tail, which makes it a particularly long and highly disordered region. A N-terminal α-helix is transiently formed, and is suspected to interact with the kinase domain in the full-length receptor. The numerous prolines in the tail participate in the transient folding of several segments into PPII helices. Two regions near the termini engage in a short-lived contact that could regulate CtErbB2 interactions.
We have shown that the structural heterogeneity of the tail is accompanied by high dynamic heterogeneity, as seen in ^{15}N relaxation data. However, we have not investigated the details of these dynamics heterogeneities. In particular, the regions involved in long-range contacts seem to undergo conformational exchange. Analysis of the relaxation data at several fields using different models, and complementary relaxation-dispersion or CEST experiments (depending on the exchange rate) could give some insight into the type of movement involved. The different dynamic behaviors experienced by glycine-rich and proline-rich regions could also be investigated in more detail.

Then, we studied a major interaction partner of ErbB2, the modular SH3-SH2-SH3 protein Grb2. Grb2 SH2 domain is known to interact with phosphotyrosine B of ErbB2 and its two SH3 domains link it to the Ras/MAPK pathway, a pathway critical for ErbB2 function and its role in cancer. Contrary to what is suggested by the crystal structure of Grb2, we showed by NMR that in its unliganded form, Grb2 domains do not interact with one another in solution. The NSH3 domain is exchanging between a folded and an unfolded form, and stays at a fix distance to the SH2 in the NSH3SH2 bidomain. Conversely, the fully

folded CSH3 domain experiences extension and shortening of its linker relative to the SH2 domain in the SH2CSH3 bidomain. Extension and compaction is also seen in the full-length protein. This behavior may influence binding to partners of Grb2.
Complementary RDC experiments and analysis of relaxation data would allow further characterization of the inter-domain orientations and dynamics. Additionally, Grb2 oligomeric state in controversial. In both the NMR and SAXS experiments that we performed, Grb2 in mainly monomeric. However, size exclusion chromatography shows peaks of dimers. We have not investigated the monomer/dimer equilibrium, that seems to have a dissociation constant higher than the μM K_D suggested by some studies. Once again, further relaxation-dispersion experiments could help detect the potential dimeric form, and SAXS analysis of the dimer peaks in SEC could give information about the organization of the dimer.

Having characterized both partners, we investigated the binding of Grb2 to unphosphorylated CtErbB2. Grb2 SH2 domain binds weakly to unphosphorylated CtErbB2 tyrosine B, and its NSH3 domain interacts with a proline-rich segment a few residues C-terminus of this tyrosine. Using ErbB2 peptides containing these two binding motifs, we showed that upon tyrosine phosphorylation the interaction with the SH2 domain of Grb2 is tighter, giving very different binding characteristics for the SH2-phosphotyrosine and the NSH3-polyproline interactions.
Peptides containing only the tyrosine or only the polyproline segment will be tested for interaction with Grb2 to investigate potential cooperativity between the SH2 and NSH3 binding. Peptides mutating the different prolines in the polyproline segment will also enable to determine if the canonical class I RxxPxxP motif is indeed the interaction motif, or if the other prolines and arginine are directly involved. Titrations of the whole CtErbB2 with Grb2 showed effects of binding on regions outside of the tyrosine B-polyproline segment. We will titrate Grb2 with peptides corresponding to these regions, to see if these effects are due to direct binding or indirect effects. The goal would eventually be to give up the peptide approach and study the interaction of Grb2 with the whole phosphorylated CtErbB2. In this context, multisteric effects as defined by Tompa (2014) as an extension of the concept of allostery could really be investigated.

Finally, an important perspective for this study is the obtention of phosphorylated CtErbB2. The different methods that we described would allow the study of the effect of phosphorylation on the structure and dynamics of CtErbB2 alone, and on its interactions. The study of the phosphorylation process itself, and more generally of the interaction of the C-terminal tail with the kinase domain of ErbB2 is another reason

for trying to produce constructs containing the kinase domain, using eukaryotic expression systems. We showed that we are able the express the intracellular region in insect cells, and the perspectives are therefore numerous.

All in all, our study is part of the ongoing global effort to understand how intrinsically disordered proteins perform their biological functions. In CtErbB2, the high level of disorder all along the sequence allows very high accessibility of all the tyrosine phosphorylation/interaction sites. We showed an even higher density of interaction sites along the tail that what was previously thought, with polyproline motifs adding up to (phospho)tyrosines. A folded region of this size could probably not fit as many interaction sites. We have also shown that in terms of transient secondary and tertiary structure, and in terms of dynamics, the tail is highly heterogeneous. Despite disorder, different sites have different behaviors, which suggests that there would be different binding modes of different regions of the tail, supporting the hypothesis of potentially different specificities of the multiple sites. In addition to this multiplicity, another layer of complexity and adaptability is added by phosphorylation, increasing the possibilities for regulation.
The interaction that we studied here is a model of the possibilities only offered by disorder: two motifs of very different nature (phosphotyrosine and polyproline), within only twenty residues, bind to different modules of a single protein, in a phosphorylation-dependent manner. In ErbB2, for which regulation by ligand-binding to a folded domain (the extracellular domain) is lacking, the regulation by this disordered region is particularly important. And the outcome of this regulation of interactions, determining the nature and intensity of signaling, is crucial for our understanding of physiological and pathological function, especially in cancer.

BIBLIOGRAPHY

Abrahao-Machado, L. F. and Scapulatempo-Neto, C. (2016). HER2 testing in gastric cancer: An update. *World J. Gastroenterol.*, 22(19):4619–25.

Aertgeerts, K., Skene, R., Yano, J., Sang, B. C., Zou, H., Snell, G., Jennings, A., Iwamoto, K., Habuka, N., Hirokawa, A., Ishikawa, T., Tanaka, T., Miki, H., Ohta, Y., and Sogabe, S. (2011). Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. *J. Biol. Chem.*, 286(21):18756–18765.

Agus, D. B., Akita, R. W., Fox, W. D., Lewis, G. D., Higgins, B., Pisacane, P. I., Lofgren, J. A., Tindell, C., Evans, D. P., Maiese, K., Scher, H. I., and Sliwkowski, M. X. (2002). Targeting ligand-activated ErbB2 signaling inhibits breast and prostate tumor growth. *Cancer Cell*, 2(2):127–37.

Ahmed, Z., Timsah, Z., Suen, K. M., Cook, N. P., Lee, G. R., Lin, C.-C., Gagea, M., Marti, A. a., and Ladbury, J. E. (2015). Grb2 monomer-dimer equilibrium determines normal versus oncogenic function. *Nat. Commun.*, 6:7354.

Akiyama, T., Matsuda, S., Namba, Y., Saito, T., Toyoshima, K., and Yamamoto, T. (1991). The transforming potential of the c-erbB-2 protein is regulated by its autophosphorylation at the carboxyl-terminal domain. *Mol. Cell. Biol.*, 11(2):833–42.

Alderson, T. R., Lee, J. H., Charlier, C., Ying, J., and Bax, A. (2018). Propensity for cis -Proline Formation in Unfolded Proteins. *ChemBioChem*, 19(1):37–42.

Alvarez, C. V., Shon, K.-J., Miloso, M., and Beguinot, L. (1995). Structural Requirements of the Epidermal Growth Factor Receptor for Tyrosine Phosphorylation of eps8 and eps15 , Substrates Lacking Src SH2 Homology Domains. *J. Biol. Chem.*, 270(27):16271–16276.

Anfinsen, C. B. (1973). Folding of Protein Chains. *Science (80-.).*, 181(4096):223–230.

Araújo, A., Oliva, G., Henrique-Silva, F., Garratt, R., Cáceres, O., and Beltramini, L. (2000). Influence of the Histidine Tail on the Structure and Activity of Recombinant Chlorocatechol 1,2-Dioxygenase. *Biochem. Biophys. Res. Commun.*, 272(2):480–484.

Arkhipov, A., Shan, Y., Das, R., Endres, N. F. N., Eastwood, M. P. M., Wemmer, D. E. D., Kuriyan, J., and Shaw, D. E. (2013). Architecture and membrane interactions of the EGF receptor. *Cell*, 152(3):557–569.

Arndt-Jovin, D. J., Botelho, M. G., and Jovin, T. M. (2014). Structure-Function Relationships of ErbB RTKs in the Plasma Membrane of Living Cells. *Cold Spring Harb. Perspect. Biol.*, 6(4):a008961–a008961.

Bah, A., Vernon, R. M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L. E., and Forman-Kay, J. D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*, 519(7541):106–9.

Bargmann, C. I., Hung, M.-C., and Weinberg, R. A. (1986). Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell*, 45(5):649–657.

Bargmann, C. I. and Weinberg, R. A. (1988). Increased tyrosine kinase activity associated with the protein encoded by the activated neu oncogene. *Proc. Natl. Acad. Sci.*, 85(15):5394–5398.

Bartelt, R. R., Light, J., Vacaflores, A., Butcher, A., Pandian, M., Nash, P., and Houtman, J. C. D. (2015). Regions outside of conserved PxxPxR motifs drive the high affinity interaction of GRB2 with SH3 domain ligands. *Biochim. Biophys. Acta - Mol. Cell Res.*, 1853(10):2560–2569.

Bayer, K.-U., De Koninck, P., Leonard, A. S., Hell, J. W., and Schulman, H. (2001). Interaction with the NMDA receptor locks CaMKII in an active conformation. *Nature*, 411(6839):801–805.

Bell, C. A., Tynan, J. A., Hart, K. C., Meyer, A. N., Robertson, S. C., and Donoghue, D. J. (2000). Rotational Coupling of the Transmembrane and Kinase Domains of the Neu Receptor Tyrosine Kinase. *Mol. Biol. Cell*, 11(10):3589–3599.

Belorusova, A. Y., Osz, J., Petoukhov, M. V., Peluso-Iltis, C., Kieffer, B., Svergun, D. I., and Rochel, N. (2016). Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor α in the Context of the Full-Length Protein. *Biochemistry*, 55(12):1741–1748.

Ben-Bassat, H. and Klein, B. Y. (2000). Inhibitors of tyrosine kinases in the treatment of psoriasis. *Curr. Pharm. Des.*, 6(9):933–42.

Benfield, A. P., Whiddon, B. B., Clements, J. H., and Martin, S. F. (2007). Structural and energetic aspects of Grb2-SH2 domain-swapping. *Arch. Biochem. Biophys.*, 462(1):47–53.

Berlow, R. B., Dyson, H. J., and Wright, P. E. (2018). Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *J. Mol. Biol.*, 430(16):2309–2320.

Bernadó, P., Bertoncini, C. W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005). Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.*, 127(51):17968–17969.

Bernadó, P. and Blackledge, M. (2009). A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys. J.*, 97(10):2839–2845.

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, 129(17):5656–5664.

Bertics, P. J., Weber, W., Cochet, C., and Gill, G. N. (1985). Regulation of the epidermal growth factor receptor by phosphorylation. *J. Cell. Biochem.*, 29(3):195–208.

Bertoncini, C. W., Jung, Y.-S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., and Zweckstetter, M. (2005). From The Cover: Release of long-range tertiary interactions potentiates aggregation of natively unstructured -synuclein. *Proc. Natl. Acad. Sci.*, 102(5):1430–1435.

Bhattacharya, S. and Lin, X. (2019). Recent advances in computational protocols addressing intrinsically disordered proteins. *Biomolecules*, 9(4).

Bhoir, S., Shaik, A., Thiruvenkatam, V., and Kirubakaran, S. (2018). High yield bacterial expression, purification and characterisation of bioactive Human Tousled-like Kinase 1B involved in cancer. *Sci. Rep.*, 8(1):4796.

Bishayee, A., Beguinot, L., and Bishayee, S. (1999). Phosphorylation of Tyrosine 992, 1068, and 1086 Is Required for Conformational Change of the Human Epidermal Growth Factor Receptor C-Terminal Tail. *Mol. Biol. Cell*, 10(3):525–536.

Booth, W. T., Schlachter, C. R., Pote, S., Ussin, N., Mank, N. J., Klapper, V., Offermann, L. R., Tang, C., Hurlburt, B. K., and Chruszcz, M. (2018). Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS Omega*, 3(1):760–768.

Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., Sottini, A., Soranno, A., Buholzer, K. J., Nettels, D., Kragelund, B. B., Best, R. B., Schuler, B., Borgia*, A., Borgia*, M. B., Bugge*, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., Sottini, A., Soranno, A., Buholzer, K. J., Nettels, D., Kragelund, B. B., Best, R. B., and Schuler, B. (2017). Extreme disorder in an ultra-high-affinity protein complex. *Nature*, 555(7694):61–66.

Bornet, O., Nouailler, M., Feracci, M., Sebban-Kreuzer, C., Byrne, D., Halimi, H., Morelli, X., Badache, A., and Guerlesquin, F. (2014). Identification of a Src kinase SH3 binding site in the C-terminal domain of the human ErbB2 receptor tyrosine kinase. *FEBS Lett.*, 588(12):2031–2036.

Bragin, P. E., Mineev, K. S., Bocharova, O. V., Volynsky, P. E., Bocharov, E. V., and Arseniev, A. S. (2016). HER2 Transmembrane Domain Dimerization Coupled with Self-Association of Membrane-Embedded Cytoplasmic Juxtamembrane Regions. *J. Mol. Biol.*, 428(1):52–61.

Brennan, P. J., Kumogai, T., Berezov, A., Murali, R., and Greene, M. I. (2000). HER2/neu: Mechanisms of dimerization/oligomerization. *Oncogene*, 19(53):6093–6101.

Brignola, P. S., Lackey, K., Kadwell, S. H., Hoffman, C., Horne, E., Carter, H. L., Stuart, J. D., Blackburn, K., Moyer, M. B., Alligood, K. J., Knight, W. B., and Wood, E. R. (2002a). Comparison of the Biochemical and Kinetic Properties of the Type 1 Receptor Tyrosine Kinase Intracellular Domains. *J. Biol. Chem.*, 277(2):1576–1585.

Brignola, P. S., Lackey, K., Kadwell, S. H., Hoffman, C., Horne, E., Carter, H. L., Stuart, J. D., Blackburn, K., Moyer, M. B., Alligood, K. J., Knight, W. B., and Wood, E. R. (2002b). Comparison of the Biochemical and Kinetic Properties of the Type 1 Receptor Tyrosine Kinase Intracellular Domains. *J. Biol. Chem.*, 277(2):1576–1585.

Britsch, S., Li, L., Kirchhoff, S., Theuring, F., Brinkmann, V., Birchmeier, C., and Riethmacher, D. (1998). The ErbB2 and ErbB3 receptors and their ligand, neuregulin-1, are essential for development of the sympathetic nervous system. *Genes Dev.*, 12(12):1825–36.

Brucale, M., Schuler, B., and Samorì, B. (2014). Single-Molecule Studies of Intrinsically Disordered Proteins. *Chem. Rev.*, 114(6):3281–3317.

Buday, L. and Downward, J. (1993). Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein, and Sos nucleotide exchange factor. *Cell*, 73(3):611–20.

Burden, S. and Yarden, Y. (1997). Neuregulins and their receptors: A versatile signaling module in organogenesis and oncogenesis. In *Neuron*, volume 18, pages 847–855.

Burgess, A. W., Cho, H.-S., Eigenbrot, C., Ferguson, K. M., Garrett, T. P. J., Leahy, D. J., Lemmon, M. A., Sliwkowski, M. X., Ward, C. W., and Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell*, 12(3):541–52.

Burnett, G. and Kennedy, E. P. (1954). The enzymatic phosphorylation of proteins. *J. Biol. Chem.*, 211(2):969–80.

Buza, N., Roque, D. M., and Santin, A. D. (2014). HER2/ neu in Endometrial Cancer: A Promising Therapeutic Target With Diagnostic Challenges. *Arch. Pathol. Lab. Med.*, 138(3):343–350.

Camilloni, C., De Simone, A., Vranken, W. F., and Vendruscolo, M. (2012). Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry*, 51(11):2224–2231.

Carter, P., Presta, L., Gorman, C. M., Ridgway, J. B., Henner, D., Wong, W. L., Rowland, a. M., Kotts, C., Carver, M. E., and Shepard, H. M. (1992). Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.*, 89(10):4285–4289.

Caspers, P., Stieger, M., and Burn, P. (1994). Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in Escherichia coli. *Cell. Mol. Biol. (Noisy-le-grand).*, 40(5):635–44.

Chan, R., Hardy, W. R., Laing, M. A., Hardy, S. E., and Muller, W. J. (2002). The Catalytic Activity of the ErbB-2 Receptor Tyrosine Kinase Is Essential for Embryonic Development. *Mol. Cell. Biol.*, 22(4):1073–1078.

Changeux, J. P. (1961). The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb. Symp. Quant. Biol.*, 26:313–8.

Chardin, P., Camonis, J., Gale, N., van Aelst, L., Schlessinger, J., Wigler, M., and Bar-Sagi, D. (1993). Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2. *Science (80-.).*, 260(5112):1338–1343.

Charest, A., Wagner, J., Jacob, S., McGlade, C. J., and Tremblay, M. L. (1996). Phosphotyrosine-independent Binding of SHC to the NPLH Sequence of Murine Protein-tyrosine Phosphatase-PEST. *J. Biol. Chem.*, 271(14):8424–8429.

Chaudhury, A. R., Gerecke, K. M., Wyss, J. M., Morgan, D. G., Gordon, M. N., and Carroll, S. L. (2003). Neuregulin-1 and ErbB4 Immunoreactivity Is Associated with Neuritic Plaques in Alzheimer Disease Brain and in a Transgenic Model of Alzheimer Disease. *J. Neuropathol. Exp. Neurol.*, 62(1):42–54.

Chausovsky, A., Waterman, H., Elbaum, M., Yarden, Y., Geiger, B., and Bershadsky, A. D. (2000). Molecular requirements for the effect of neuregulin on cell spreading, motility and colony organization. *Oncogene*, 19(7):878–888.

Chen, C., Zhou, Z., Liu, R., Li, Y., Azmi, P. B., and Seth, A. K. (2008). The WW domain containing E3 ubiquitin protein ligase 1 upregulates ErbB2 and EGFR through RING finger protein 11. *Oncogene*, 27(54):6845–6855.

Chen, Z. and Cole, P. A. (2015). Synthetic approaches to protein phosphorylation. *Curr. Opin. Chem. Biol.*, 28(3):115–22.

Cheng, A. M., Saxton, T. M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C. G., Cardiff, R. D., Cross, J. C., Muller, W. J., and Pawson, T. (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell*, 95(6):793–803.

Cheng, K. and Koland, J. G. (1996). Nucleotide Binding by the Epidermal Growth Factor Receptor Protein-tyrosine Kinase. *J. Biol. Chem.*, 271(1):311–318.

Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., and Leahy, D. J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature*, 421(6924):756–760.

Chou, J. J., Gaemers, S., Howder, B., Louis, J. M., and Bax, A. (2001). A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J. Biomol. NMR*, 21(4):377–382.

Cohen, S., Fava, R. A., and Sawyer, S. T. (1982). Purification and characterization of epidermal growth factor receptor/protein kinase from normal mouse liver. *Proc. Natl. Acad. Sci. U. S. A.*, 79(20):6237–41.

Connell, C. M. and Doherty, G. J. (2017). Activating HER2 mutations as emerging targets in multiple solid cancers. *ESMO Open*, 2(5):e000279.

Cordeiro, T. N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., and Bernadó, P. (2017). Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr. Opin. Struct. Biol.*, 42:15–23.

Cordier, F., Chaffotte, A., Terrien, E., Préhaud, C., Theillet, F. X., Delepierre, M., Lafon, M., Buc, H., and Wolff, N. (2012). Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.*, 134(50):20533–20543.

Cordier, F., Chaffotte, A., and Wolff, N. (2015). Quantitative and dynamic analysis of PTEN phosphorylation by NMR. *Methods*, 77-78:82–91.

Cussac, D., Frech, M., and Chardin, P. (1994). Binding of the Grb2 SH2 domain to phosphotyrosine motifs does not change the affinity of its SH3 domains for Sos proline-rich motifs. *EMBO J.*, 13(17):4011–4021.

Dai, H., Ding, H., Meng, X. W., Lee, S.-H., Schneider, P. A., and Kaufmann, S. H. (2013). Contribution of Bcl-2 Phosphorylation to Bak Binding and Drug Resistance. *Cancer Res.*, 73(23):6998–7008.

Daly, R. J., Binder, M. D., and Sutherland, R. L. (1994). Overexpression of the Grb2 gene in human breast cancer cell lines. *Oncogene*, 9(9):2723–7.

Dankort, D., Jeyabalan, N., Jones, N., Dumont, D. J., and Muller, W. J. (2001a). Multiple ErbB-2/Neu Phosphorylation Sites Mediate Transformation through Distinct Effector Proteins. *J. Biol. Chem.*, 276(42):38921–38928.

Dankort, D., Maslikowski, B., Warner, N., Kanno, N., Kim, H., Wang, Z., Moran, M. F., Oshima, R. G., Cardiff, R. D., and Muller, W. J. (2001b). Grb2 and Shc Adapter Proteins Play Distinct Roles in Neu (ErbB-2) - Induced Mammary Tumorigenesis : Implications for Human Breast Cancer. *Mol. Cell. Biol.*, 21(5):1540–1551.

Dankort, D. L., Wang, Z., Blackmore, V., Moran, M. F., and Muller, W. J. (1997). Distinct tyrosine autophosphorylation sites negatively and positively modulate neu-mediated transformation. *Mol. Cell. Biol.*, 17(9):5410–5425.

David, G. and Pérez, J. (2009). Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Crystallogr.*, 42(5):892–900.

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994). Synthesis of proteins by native chemical ligation. *Science*, 266(5186):776–9.

De Biasio, A., de Opakua, A. I., Mortuza, G. B., Molina, R., Cordeiro, T. N., Castillo, F., Villate, M., Merino, N., Delgado, S., Gil-Cartón, D., Luque, I., Diercks, T., Bernadó, P., Montoya, G., and Blanco, F. J. (2015). Structure of p15PAF-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nat. Commun.*, 6:6439.

Delaforge, E., Kragelj, J., Tengo, L., Palencia, A., Milles, S., Bouvignies, G., Salvi, N., Blackledge, M., and Jensen, M. R. (2018). Deciphering the Dynamic Interaction Profile of an Intrinsically Disordered Protein by NMR Exchange Spectroscopy. *J. Am. Chem. Soc.*, 140(3):jacs.7b12407.

den Hertog, J. and Hunter, T. (1996). Tight association of GRB2 with receptor protein-tyrosine phosphatase alpha is mediated by the SH2 and C-terminal SH3 domains. *EMBO J.*, 15(12):3016–27.

Deshmukh, L., Meller, N., Alder, N., Byzova, T., and Vinogradova, O. (2011). Tyrosine Phosphorylation as a Conformational Switch. *J. Biol. Chem.*, 286(47):40943–40953.

Deville, C. (2015). *Etude des états multiples des domaines WH2 en interaction avec l'actine par résonnance magnétique nucléaire*. Phd thesis, Université Pierre et Marie Curie.

Di Fiore, P., Pierce, J., Kraus, M., Segatto, O., King, C., and Aaronson, S. (1987). erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science (80-.).*, 237(4811):178–182.

Didry, D., Cantrelle, F.-X., Husson, C., Roblin, P., Moorthy, A. M. E., Perez, J., Le Clainche, C., Hertzog, M., Guittet, E., Carlier, M.-F., van Heijenoort, C., and Renault, L. (2012). How a single residue in individual β-thymosin/WH2 domains controls their functions in actin assembly. *EMBO J.*, 31(4):1000–1013.

Dodero, V. I. (2011). Biomolecular studies by circular dichroism. *Front. Biosci.*, 16(1):61.

Dogan, J., Gianni, S., and Jemth, P. (2014). The binding mechanisms of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.*, 16(14):6323–6331.

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16):3433–3434.

Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J., and Waterfield, M. D. (1984). Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature*, 307(5951):521–527.

Du, Z. and Lovly, C. M. (2018). Mechanisms of receptor tyrosine kinase activation in cancer. *Mol. Cancer*, 17(1):58.

Duneau, J.-P., Vegh, A. P., and Sturgis, J. N. (2007). A Dimerization Hierarchy in the Transmembrane Domains of the HER Receptor Family. *Biochemistry*, 46(7):2010–2019.

Dunker, A., Lawson, J., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.*, 19(1):26–59.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.*, 169(1):45–53.

Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, 6(3):197–208.

Eckhart, W., Hutchinson, M. A., and Hunter, T. (1979). An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. *Cell*, 18(4):925–933.

Ehlers, M. D., Zhang, S., Bernhardt, J. P., and Huganir, R. L. (1996). Inactivation of NMDA Receptors by Direct Interaction of Calmodulin with the NR1 Subunit. *Cell*, 84(5):745–755.

Endres, N. F., Das, R., Smith, A. W., Arkhipov, A., Kovacs, E., Huang, Y., Pelton, J. G., Shan, Y., Shaw, D. E., Wemmer, D. E., Groves, J. T., and Kuriyan, J. (2013). Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell*, 152(3):543–556.

Escrivá-de Romaní, S., Arumí, M., Bellet, M., and Saura, C. (2018). HER2-positive breast cancer: Current and new therapeutic strategies. *The Breast*, 39:80–88.

Fan, Y.-X., Wong, L., Ding, J., Spiridonov, N. A., Johnson, R. C., and Johnson, G. R. (2008). Mutational Activation of ErbB2 Reveals a New Protein Kinase Autoinhibition Mechanism. *J. Biol. Chem.*, 283(3):1588–1596.

Felder, S., Zhou, M., Hu, P., Ureña, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993). SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol. Cell. Biol.*, 13(3):1449–1455.

Feracci, M., Pimentel, C., Bornet, O., Roche, P., Salaun, D., Badache, A., and Guerlesquin, F. (2011). MEMO associated with an ErbB2 receptor phosphopeptide reveals a new phosphotyrosine motif. *FEBS Lett.*, 585(17):2688–92.

Ferguson, K. M. (2000). Extracellular domains drive homo- but not hetero-dimerization of erbB receptors. *EMBO J.*, 19(17):4632–4643.

Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H.-s., Leahy, D. J., and Lemmon, M. A. (2003). EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. *Mol. Cell*, 11(2):507–517.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der Dtsch. Chem. Gesellschaft*, 27(3):2985–2993.

Fonin, A. V., Darling, A. L., Kuznetsova, I. M., Turoverov, K. K., and Uversky, V. N. (2018). Intrinsically disordered proteins in crowded milieu: when chaos prevails within the cellular gumbo. *Cell. Mol. Life Sci.*, 75(21):3907–3929.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017). ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, 50(4):1212–1225.

Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., De Vos, A. M., and Sliwkowski, M. X. (2004). Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell*, 5(4):317–328.

Freed, D. M., Bessman, N. J., Kiyatkin, A., Salazar-Cavazos, E., Byrne, P. O., Moore, J. O., Valley, C. C., Ferguson, K. M., Leahy, D. J., Lidke, D. S., and Lemmon, M. A. (2017). EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. *Cell*, 171(3):683–695.e18.

Fushman, D., Weisemann, R., Thüring, H., and Rüterjans, H. (1994). Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J. Biomol. NMR*, 4(1):61–78.

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.*, 338(5):1015–1026.

Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J. P. (1987). Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences. *FEBS Lett.*, 224(1):149–155.

Gajiwala, K. S. (2013). EGFR: Tale of the C-terminal tail. *Protein Sci.*, 22(7):995–999.

Gale, N. W., Kaplan, S., Lowenstein, E. J., Schlessinger, J., and Bar-Sagi, D. (1993). Grb2 mediates the EGF-dependent activation of guanine nucleotide exchange on Ras. *Nature*, 363(6424):88–92.

Garrett, T. P., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Kofler, M., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2003). The Crystal Structure of a Truncated ErbB2 Ectodomain Reveals an Active Conformation, Poised to Interact with Other ErbB Receptors. *Mol. Cell*, 11(2):495–505.

Gassmann, M., Casagranda, F., Orioli, D., Simon, H., Lai, C., Klein, R., and Lemke, G. (1995). Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature*, 378(6555):390–394.

Giani, C., Casalini, P., Pupa, S. M., De Vecchi, R., Ardini, E., Colnaghi, M. I., Giordano, A., and Ménard, S. (1998). Increased expression of c-erbB-2 in hormone-dependent breast cancer cells inhibits cell growth and induces differentiation. *Oncogene*, 17(4):425–432.

Gill, K., Macdonald-Obermann, J. L., and Pike, L. J. (2017). Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. *J. Biol. Chem.*, 292(50):20744–20755.

Goldin, A. L. (2003). Mechanisms of sodium channel inactivation. *Curr. Opin. Neurobiol.*, 13(3):284–290.

Gotoh, N., Tojo, A., Hino, M., Yazaki, Y., and Shibuya, M. (1992). A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor. *Biochem. Biophys. Res. Commun.*, 186(2):768–774.

Goudreau, N., Cornille, F., Duchesne, M., Parker, F., Tocque, B., Garbay, C., and Roques, B. P. (1994). NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. *Nat Struct Biol*, 1(12):898–907.

Graus-Porta, D., Beerli, R. R., Daly, J. M., and Hynes, N. E. (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J.*, 16:1647–1655.

Gril, B., Vidal, M., Assayag, F., Poupon, M.-F., Liu, W.-Q., and Garbay, C. (2007). Grb2-SH3 ligand inhibits the growth of HER2+ cancer cells and has antitumor effects in human cancer xenografts alone and in combination with docetaxel. *Int. J. Cancer*, 121(2):407–415.

Gu, H. and Neel, B. G. (2003). The "Gab" in signal transduction. *Trends Cell Biol.*, 13(3):122–130.

Guilloteau, J. P., Fromage, N., Ries-Kautt, M., Reboul, S., Bocquet, D., Dubois, H., Faucher, D., Colonna, C., Ducruix, A., and Becquart, J. (1996). Purification, stabilization, and crystallization of a modular protein: Grb2. *Proteins*, 25(1):112–9.

Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. (2014). Introducing protein intrinsic disorder.

Halle, B. and Davidovic, M. (2003). Biomolecular hydration: From water dynamics to hydrodynamics. *Proc. Natl. Acad. Sci.*, 100(21):12135–12140.

Harkiolaki, M., Tsirka, T., Lewitzky, M., Simister, P. C., Joshi, D., Bird, L. E., Jones, E. Y., O'Reilly, N., and Feller, S. M. (2009). Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the SH3C Domain of Grb2. *Structure*, 17(6):809–822.

Hazan, R., Margolis, B., Dombalagian, M., Ullrich, A., Zilberstein, A., and Schlessinger, J. (1990). Identification of autophosphorylation sites of HER2/neu. *Cell Growth Differ.*, 1(1):3–7.

Hellström, I., Goodman, G., Pullman, J., Yang, Y., and Hellström, K. E. (2001). Overexpression of HER-2 in ovarian carcinomas. *Cancer Res.*, 61(6):2420–3.

Hinshaw, J. E. (2000). Dynamin and its role in membrane fission. *Annu. Rev. Cell Dev. Biol.*, 16(January):483–519.

Holbro, T., Beerli, R. R., Maurer, F., Koziczak, M., Barbas, C. F., and Hynes, N. E. (2003). The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc. Natl. Acad. Sci.*, 100(15):8933–8938.

Hoppmann, C., Wong, A., Yang, B., Li, S., Hunter, T., Shokat, K. M., and Wang, L. (2017). Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.*, 13(8):842–844.

Horan, T., Wen, J., Arakawa, T., Liu, N., Brankow, D., Hu, S., Ratzkin, B., and Philo, J. S. (1995). Binding of Neu Differentiation Factor with the Extracellular Domain of Her2 and Her3. *J. Biol. Chem.*, 270(41):24604–24608.

Huang, J. and MacKerell, A. D. (2018). Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.*, 48:40–48.

Huang, Y. and Liu, Z. (2009). Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *J. Mol. Biol.*, 393(5):1143–1159.

Hubbard, S. R. (1999). Structural analysis of receptor tyrosine kinases. *Prog. Biophys. Mol. Biol.*, 71(3-4):343–358.

Hubbard, S. R. (2004). Juxtamembrane autoinhibition in receptor tyrosine kinases. *Nat. Rev. Mol. Cell Biol.*, 5(6):464–471.

Hunter, T. (2009). Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.*, 21(2):140–6.

Hunter, T. and Sefton, B. M. (1980). Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc. Natl. Acad. Sci. U. S. A.*, 77(3):1311–5.

Iakoucheva, L. M., Brown, C. J., Lawson, J., Obradović, Z., and Dunker, A. (2002). Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.*, 323(3):573–584.

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, 32(3):1037–1049.

Ishikawa, T., Seto, M., Banno, H., Kawakita, Y., Oorui, M., Taniguchi, T., Ohta, Y., Tamura, T., Nakayama, A., Miki, H., Kamiguchi, H., Tanaka, T., Habuka, N., Sogabe, S., Yano, J., Aertgeerts, K., and Kamiyama, K. (2011). Design and synthesis of novel human epidermal growth factor receptor 2 (HER2)/epidermal growth factor receptor (EGFR) dual inhibitors bearing a pyrrolo[3,2-d]pyrimidine scaffold. *J. Med. Chem.*, 54(23):8030–8050.

Ivancic, M., Daly, R. J., and Lyons, B. A. (2003). Solution structure of the human Grb7-SH2 domain/erbB2 peptide complex and structural basis for Grb7 binding to ErbB2. *J. Biomol. NMR*, 27(3):205–19.

Iwakura, Y. and Nawa, H. (2013). ErbB1-4-dependent EGF/neuregulin signals and their cross talk in the central nervous system: pathological implications in schizophrenia and Parkinson's disease. *Front. Cell. Neurosci.*, 7(February):4.

Jensen, M. R., Communie, G., Ribeiro, E. A. J., Martinez, N., Desfosses, A., Salmon, L., Mollica, L., Gabel, F., Jamin, M., Longhi, S., Ruigrok, R. W. H., and Blackledge, M. (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci.*, 108(24):9839–9844.

Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P., and Blackledge, M. (2009). Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure*, 17(9):1169–1185.

Jeon, M., Lee, J., Nam, S., Shin, I., Lee, J., and Kim, S. (2015). Induction of fibronectin by HER2 overexpression triggers adhesion and invasion of breast cancer cells. *Exp. Cell Res.*, 333(1):116–126.

Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., Wemmer, D. E., Zhang, X., and Kuriyan, J. (2009). Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. *Cell*, 137(7):1293–1307.

Kaneko, T. (2008). The SH3 domain- a family of versatile peptide- and protein-recognition module. *Front. Biosci.*, Volume(13):4938.

Karplus, M. (1959). Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.*, 30(1):11–15.

Karush, F. (1950). Heterogeneity of the Binding Sites of Bovine Serum Albumin. *J. Am. Chem. Soc.*, 72(6):2705–2713.

Kasahara, K., Terazawa, H., Takahashi, T., and Higo, J. (2019). Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review. *Comput. Struct. Biotechnol. J.*, 17:712–720.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

Keppel, T. R., Sarpong, K., Murray, E. M., Monsey, J., Zhu, J., and Bose, R. (2017). Biophysical evidence for intrinsic disorder in the C-terminal tails of the epidermal growth factor receptor (EGFR) and HER3 receptor tyrosine kinases. *J. Biol. Chem.*, 292(2):597–610.

Khanal, P., Namgoong, G. M., Kang, B. S., Woo, E.-R., and Choi, H. S. (2010). The prolyl isomerase Pin1 enhances HER-2 expression and cellular transformation via its interaction with mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 1. *Mol. Cancer Ther.*, 9(3):606–16.

Khoury, G. A., Baliban, R. C., and Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.*, 1(1):90.

Kikhney, A. G. and Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.*, 589(19):2570–2577.

Kim, K. L., Kim, D., Lee, S., Kim, S.-J., Noh, J. E., Kim, J.-H., Chae, Y. C., Lee, J.-B., and Ryu, S. H. (2016). Pairwise detection of site-specific receptor phosphorylations using single-molecule blotting. *Nat. Commun.*, 7(1):11107.

Kjaergaard, M. and Kragelund, B. B. (2017). Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.*, 74(17):3205–3224.

Klapper, L. N., Glathe, S., Vaisman, N., Hynes, N. E., Andrews, G. C., Sela, M., and Yarden, Y. (1999). The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. *Proc. Natl. Acad. Sci.*, 96(9):4995–5000.

Kohda, D., Terasawa, H., Ichikawa, S., Ogura, K., Hatanaka, H., Mandiyan, V., Ullrich, A., Schlessinger, J., and Inagaki, F. (1994). Solution structure and ligand-binding site of the carboxy-terminal SH3 domain of GRB2. *Structure*, 2(11):1029–1040.

Koland, J. G. (2014). Coarse-Grained Molecular Simulation of Epidermal Growth Factor Receptor Protein Tyrosine Kinase Multi-Site Self-Phosphorylation. *PLoS Comput. Biol.*, 10(1):e1003435.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003). PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, 36(5):1277–1282.

Konrat, R. (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.*, 241(1):74–85.

Koshland, D. E. and Brooks, C. L. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, 44(2):98–104.

Koshland, D. E., Némethy, G., and Filmer, D. (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits *. *Biochemistry*, 5(1):365–385.

Kosol, S., Contreras-Martos, S., Cedeño, C., and Tompa, P. (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*, 18(9):10802–10828.

Kovacs, E., Das, R., Wang, Q., Collier, T. S., Cantor, A., Huang, Y., Wong, K., Mirza, A., Barros, T., Grob, P., Jura, N., Bose, R., and Kuriyan, J. (2015a). Analysis of the Role of the C-Terminal Tail in the Regulation of the Epidermal Growth Factor Receptor. *Mol. Cell. Biol.*, 35(17):3083–3102.

Kovacs, E., Zorn, J. A., Huang, Y., Barros, T., and Kuriyan, J. (2015b). A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.*, 84(1):739–764.

Krzeminski, M., Marsh, J. A., Neale, C., Choy, W. Y., and Forman-Kay, J. D. (2013). Characterization of disordered proteins with ENSEMBLE. *Bioinformatics*, 29(3):398–399.

Kumar, D., Sharma, N., and Giri, R. (2017). Therapeutic Interventions of Cancers Using Intrinsically Disordered Proteins as Drug Targets: c-Myc as Model System. *Cancer Inform.*, 16:117693511769940.

Lam, P. B., Burga, L. N., Wu, B. P., Hofstatter, E. W., Lu, K. P., and Wulf, G. M. (2008). Prolyl isomerase Pin1 is highly expressed in Her2-positive breast cancer and regulates erbB2 protein stability. *Mol. Cancer*, 7:91.

Landrieu, I., Lacosse, L., Leroy, A., Wieruszeski, J. M., Trivelli, X., Sillen, A., Sibille, N., Schwalbe, H., Saxena, K., Langer, T., and Lippens, G. (2006). NMR analysis of a Tau phosphorylation pattern. *J. Am. Chem. Soc.*, 128(11):3575–3583.

Larson, S. M., Davidson, A. R., and Davidson, A. R. (2000). The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. *Protein Sci.*, 9(11):2170–2180.

Lebendiker, M. and Danieli, T. (2014). Production of prone-to-aggregate proteins. *FEBS Lett.*, 588(2):236–246.

Leblanc, S. J., Kulkarni, P., and Weninger, K. R. (2018). Single molecule FRET: A powerful tool to study intrinsically disordered proteins. *Biomolecules*, 8(4).

Lee, K.-f., Simon, H., Chen, H., Bates, B., Hung, M.-c., and Hauser, C. (1995). Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature*, 378(6555):394–398.

Lee, N. Y., Hazlett, T. L., and Koland, J. G. (2006). Structure and dynamics of the epidermal growth factor receptor C-terminal phosphorylation domain. *Protein Sci.*, 15(5):1142–1152.

Lee, N. Y. and Koland, J. G. (2005). Conformational changes accompany phosphorylation of the epidermal growth factor receptor C-terminal domain. *Protein Sci.*, 14(11):2793–803.

Lemmon, M. A., Ladbury, J. E., Mandiyan, V., Zhou, M., and Schlessinger, J. (1994). Independent binding of peptide ligands to the SH2 and SH3 domains of Grb2. *J. Biol. Chem.*, 269(50):31653–31658.

Lemmon, M. A. and Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell*, 141(7):1117–1134.

Lemmon, M. A., Schlessinger, J., and Ferguson, K. M. (2014). The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases. *Cold Spring Harb. Perspect. Biol.*, 6(4):a020768–a020768.

Lenferink, A. E., Pinkas-Kramarski, R., van de Poll, M. L., van Vugt, M. J., Klapper, L. N., Tzahar, E., Waterman, H., Sela, M., van Zoelen, E. J., and Yarden, Y. (1998). Differential endocytic routing of homo- and hetero-dimeric ErbB tyrosine kinases confers signaling superiority to receptor heterodimers. *EMBO J.*, 17(12):3385–97.

Leonard, A. S., Lim, I. A., Hemsworth, D. E., Horne, M. C., and Hell, J. W. (1999). Calcium/calmodulin-dependent protein kinase II is associated with the N-methyl-D-aspartate receptor. *Proc. Natl. Acad. Sci.*, 96(6):3239–3244.

Lescop, E., Rasia, R., and Brutscher, B. (2008). Hadamard Amino-Acid-Type Edited NMR Experiment for Fast Protein Resonance Assignment. *J. Am. Chem. Soc.*, 130(15):5014–5015.

Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.*, 187(1):163–169.

Lewitzky, M., Kardinal, C., Gehring, N. H., Schmidt, E. K., Konkol, B., Eulitz, M., Birchmeier, W., Schaeper, U., and Feller, S. M. (2001). The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif. *Oncogene*, 20(9):1052–1062.

Lin, C. C. C.-C., Melo, F. A., Ghosh, R., Suen, K. M., Stagg, L. J., Kirkpatrick, J., Arold, S. T., Ahmed, Z., and Ladbury, J. E. (2012). Inhibition of basal FGF receptor signaling by dimeric Grb2. *Cell*, 149(7):1514–1524.

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003). Protein disorder prediction: Implications for structural proteomics. *Structure*, 11(11):1453–1459.

Liokatis, S., Dose, A., Schwarzer, D., and Selenko, P. (2010). Simultaneous Detection of Protein Phosphorylation and Acetylation by High-Resolution NMR Spectroscopy. *J. Am. Chem. Soc.*, 132(42):14704–14705.

Liu, B. A., Engelmann, B. W., and Nash, P. D. (2012). The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. *FEBS Lett.*, 586(17):2597–2605.

Liu, B. A., Shah, E., Jablonowski, K., Stergachis, A., Engelmann, B., and Nash, P. D. (2011). The SH2 Domain-Containing Proteins in 21 Species Establish the Provenance and Scope of Phosphotyrosine Signaling in Eukaryotes. *Sci. Signal.*, 4(202):ra83–ra83.

Lopes, J. L. S., Miles, A. J., Whitmore, L., and Wallace, B. A. (2014). Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein Sci.*, 23(12):1765–72.

Lowenstein, E. J., Daly, R. J., Batzer, A. G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E. Y., Bar-Sagi, D., and Schlessinger, J. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell*, 70(3):431–442.

Lu, C., Mi, L.-Z., Grey, M. J., Zhu, J., Graef, E., Yokoyama, S., and Springer, T. A. (2010). Structural Evidence for Loose Linkage between Ligand Binding and Kinase Activation in the Epidermal Growth Factor Receptor. *Mol. Cell. Biol.*, 30(22):5432–5443.

Lu, K. P., Liou, Y.-c., and Zhou, X. Z. (2002). Pinning down proline-directed phosphorylation signaling. *Trends Cell Biol.*, 12(4):164–172.

Macdonald-Obermann, J. L. and Pike, L. J. (2014). Different Epidermal Growth Factor (EGF) Receptor Ligands Show Distinct Kinetics and Biased or Partial Agonism for Homodimer and Heterodimer Formation. *J. Biol. Chem.*, 289(38):26178–26188.

Macdonald-Obermann, J. L., Piwnica-Worms, D., and Pike, L. J. (2012). Mechanics of EGF Receptor/ErbB2 kinase activation revealed by luciferase fragment complementation imaging. *Proc. Natl. Acad. Sci.*, 109(1):137–142.

Macias, M. J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.*, 513(1):30–37.

Maignan, S., Guilloteau, J. P., Fromage, N., Arnoux, B., Becquart, J., and Ducruix, A. (1995). Crystal structure of the mammalian Grb2 adaptor. *Science (80-.).*, 268(5208):291–293.

Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F., and Minor, W. (2014). Double trouble-Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.*, 23(10):1359–1368.

Manning, G. (2002). The Protein Kinase Complement of the Human Genome. *Science (80-.).*, 298(5600):1912–1934.

Mansiaux, Y., Joseph, A. P., Gelly, J. C., and de Brevern, A. G. (2011). Assignment of polyproline ii conformation and analysis of sequence - structure relationship. *PLoS One*, 6(3):1–15.

Marengere, L. E., Songyang, Z., Gish, G. D., Schaller, M. D., Parsons, J. T., Stern, M. J., Cantley, L. C., and Pawson, T. (1994). SH2 domain specificity and activity modified by a single residue. *Nature*, 369(6480):502–5.

Margolis, B. L., Lax, I., Kris, R., Dombalagian, M., Honegger, A. M., Howk, R., Givol, D., Ullrich, A., and Schlessinger, J. (1989). All autophosphorylation sites of epidermal growth factor (EGF) receptor and HER2/neu are located in their carboxyl-terminal tails. Identification of a novel site in EGF receptor. *J. Biol. Chem.*, 264(18):10667–71.

Marone, R., Hess, D., Dankort, D., Muller, W. J., Hynes, N. E., and Badache, A. (2004). Memo mediates ErbB2-driven cell motility. *Nat. Cell Biol.*, 6(6):515–522.

Marsh, J. a., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.*, 15(12):2795–2804.

Martinez, M. S., Nonell-Canals, A., Sanchez-Martinez, M., Martinez, M. S., Nonell-Canals, A., and Sanchez-Martinez, M. (2017). Intrinsically Disordered Proteins as Drug Targets. *MOJ Proteomics Bioinforma.*, 5(3):24–26.

Mayzel, M., Rosenlöw, J., Isaksson, L., and Orekhov, V. Y. (2014). Time-resolved multidimensional NMR with non-uniform sampling. *J. Biomol. NMR*, 58(2):129–139.

McDonald, C. B., Balke, J. E., Bhat, V., Mikles, D. C., Deegan, B. J., Seldeen, K. L., and Farooq, A. (2012a). Multivalent Binding and Facilitated Diffusion Account for the Formation of the Grb2-Sos1 Signaling Complex in a Cooperative Manner. *Biochemistry*, 51(10):2122–2135.

McDonald, C. B., Bhat, V., Mikles, D. C., Deegan, B. J., and Sel (2012b). Bivalent Binding Drives the Formation of Grb2-Gab1 Signaling Complex in a Non-Cooperative Manner. *FEBS Lett.*, 279(12):2156–2173.

McDonald, C. B., El Hokayem, J., Zafar, N., Balke, J. E., Bhat, V., Mikles, D. C., Deegan, B. J., Seldeen, K. L., and Farooq, A. (2013). Allostery mediates ligand binding to Grb2 adaptor in a mutually exclusive manner. *J. Mol. Recognit.*, 26(2):92–103.

McDonald, C. B., Seldeen, K. L., Deegan, B. J., Bhat, V., and Farooq, A. (2010). Assembly of the Sos1-Grb2-Gab1 ternary signaling complex is under allosteric control. *Arch. Biochem. Biophys.*, 494(2):216–25.

McDonald, C. B., Seldeen, K. L., Deegan, B. J., and Farooq, A. (2008a). Structural basis of the differential binding of the SH3 domains of Grb2 adaptor to the guanine nucleotide exchange factor Sos1. *Arch. Biochem. Biophys.*, 479(1):52–62.

McDonald, C. B., Seldeen, K. L., Deegan, B. J., and Farooq, A. (2009). SH3 Domains of Grb2 Adaptor Bind to PXψPXR Motifs Within the Sos1 Nucleotide Exchange Factor in a Discriminate Manner. *Biochemistry*, 48(19):4074–4085.

McDonald, C. B., Seldeen, K. L., Deegan, B. J., Lewis, M. S., and Farooq, A. (2008b). Grb2 adaptor undergoes conformational change upon dimerization. *Arch. Biochem. Biophys.*, 475(1):25–35.

Mendrola, J. M., Berger, M. B., King, M. C., and Lemmon, M. A. (2002). The Single Transmembrane Domains of ErbB Receptors Self-associate in Cell Membranes. *J. Biol. Chem.*, 277(7):4704–4712.

Meola, A., Deville, C., Jeffers, S. A., Guardado-Calvo, P., Vasiliauskaite, I., Sizun, C., Girard-Blanc, C., Malosse, C., van Heijenoort, C., Chamot-Rooke, J., and Others (2014). Robust and low cost uniform< sup> 15</sup> N-labeling of proteins expressed in< i> Drosophila</i> S2 cells and< i> Spodoptera frugiperda</i> Sf9 cells for NMR applications. *J. Struct. Biol.*, 188(1):71–78.

Mészáros, B., Erdös, G., and Dosztányi, Z. (2018). IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, 46(W1):W329–W337.

Metallo, S. J. S. (2010). Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.*, 14(4):481–488.

Mi, L.-Z., Lu, C., Li, Z., Nishida, N., Walz, T., and Springer, T. A. (2011). Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. *Nat. Struct. Mol. Biol.*, 18(9):984–989.

Milles, S., Salvi, N., Blackledge, M., and Jensen, M. R. (2018). Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. *Prog. Nucl. Magn. Reson. Spectrosc.*, 109:79–100.

Minezaki, Y., Homma, K., and Nishikawa, K. (2007). Intrinsically Disordered Regions of Human Plasma Membrane Proteins Preferentially Occur in the Cytoplasmic Segment. *J. Mol. Biol.*, 368(3):902–913.

Miskei, M., Antal, C., and Fuxreiter, M. (2017). FuzDB: Database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.*, 45(D1):D228–D235.

Monod, J. and Jacob, F. (1961). Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.*, 26:389–401.

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.*, 12(1):88–118.

Monsey, J., Shen, W., Schlesinger, P., and Bose, R. (2010). Her4 and Her2/neu tyrosine kinase domains dimerize and activate in a reconstituted in vitro system. *J. Biol. Chem.*, 285(10):7035–7044.

Moriki, T., Maruyama, H., and Maruyama, I. N. (2001). Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *J. Mol. Biol.*, 311(5):1011–1026.

Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature*, 508(7496):331–339.

Musacchio, A. (2002). How SH3 domains recognize proline. *Adv. Protein Chem.*, 61:211–268.

Mustafa, M., Mirza, A., and Kannan, N. (2011). Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: A molecular dynamics study. *Proteins Struct. Funct. Bioinforma.*, 79(1):99–114.

Mylona, A., Theillet, F.-X., Foster, C., Cheng, T. M., Miralles, F., Bates, P. A., Selenko, P., and Treisman, R. (2016). Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation. *Science*, 354(6309):233–237.

Nahta, R. and Esteva, F. J. (2007). Trastuzumab: triumphs and tribulations. *Oncogene*, 26(25):3637–3643.

Nakata, A., Miyagawa, J. I., Yamashita, S., Nishida, M., Tamura, R., Yamamori, K., Nakamura, T., Nozaki, S., Kameda-Takemura, K., Kawata, S., Taniguchi, N., Higashiyama, S., and Matsuzawa, Y. (1996). Localization of heparin-binding epidermal growth factor-like growth factor in human coronary arteries: Possible roles of HB-EGF in the formation of coronary atherosclerosis. *Circulation*, 94(11):2778–2786.

Nami, B., Maadi, H., and Wang, Z. (2018). Mechanisms Underlying the Action and Synergism of Trastuzumab and Pertuzumab in Targeting HER2-Positive Breast Cancer. *Cancers (Basel).*, 10(10):342.

Narwani, T. J., Santuz, H., Shinada, N., Melarkode Vattekatte, A., Ghouzam, Y., Srinivasan, N., Gelly, J. C., and de Brevern, A. G. (2017). Recent advances on polyproline II. *Amino Acids*, 49(4):705–713.

Nasir, I., Onuchic, P. L., Labra, S. R., and Deniz, A. A. (2019). Single-molecule fluorescence studies of intrinsically disordered proteins and liquid phase separation. *Biochim. Biophys. Acta - Proteins Proteomics*, 1867(10):980–987.

Nelson, M. H. and Dolder, C. R. (2006). Lapatinib: a novel dual tyrosine kinase inhibitor with activity in solid tumors. *Ann. Pharmacother.*, 40(2):261–9.

Nioche, P., Liu, W.-Q., Broutin, I., Charbonnier, F., Latreille, M.-T., Vidal, M., Roques, B., Garbay, C., and Ducruix, A. (2002). Crystal structures of the SH2 domain of grb2: highlight on the binding of a new high-affinity inhibitor. *J. Mol. Biol.*, 315(5):1167–1177.

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., and Blackledge, M. (2009). Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.*, 131(49):17908–17918.

Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztányi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., and Gough, J. (2013). D2P2: database of disordered protein predictions. *Nucleic Acids Res.*, 41(1).

Obrien, D. P., Hernandez, B., Durand, D., Hourdel, V., Sotomayor-Pérez, A. C., Vachette, P., Ghomi, M., Chamot-Rooke, J., Ladant, D., Brier, S., and Chenal, A. (2015). Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. *Sci. Rep.*, 5(August):1–11.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. *Cell*, 110(6):775–787.

Ogura, K. and Okamura, H. (2013). Conformational change of Sos-derived proline-rich peptide upon binding Grb2 N-terminal SH3 domain probed by NMR. *Sci. Rep.*, 3(1):2913.

O'Keefe, E., Hollenberg, M. D., and Cuatrecasas, P. (1974). Epidermal growth factor. Characteristics of specific binding in membranes from liver, placenta, and other target tissues. *Arch. Biochem. Biophys.*, 164(2):518–526.

Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005). Coupled Folding and Binding with α-Helix-Forming Molecular Recognition Elements. *Biochemistry*, 44(37):12454–12470.

Olsen, J. G., Teilum, K., and Kragelund, B. B. (2017). Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.*, 74(17):3175–3183.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-r., Jensen, M. R., Segard, S., Bernado, P., Charavay, C., and Blackledge, M. (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, 28(11):1463–1470.

Paleologou, K. E., Oueslati, A., Shakked, G., Rospigliosi, C. C., Kim, H.-Y., Lamberto, G. R., Fernandez, C. O., Schmid, A., Chegini, F., Gai, W. P., Chiappe, D., Moniatte, M., Schneider, B. L., Aebischer, P., Eliezer, D., Zweckstetter, M., Masliah, E., and Lashuel, H. A. (2010). Phosphorylation at S87 Is Enhanced in Synucleinopathies, Inhibits -Synuclein Oligomerization, and Influences Synuclein-Membrane Interactions. *J. Neurosci.*, 30(9):3184–3198.

Panek, A., Pietrow, O., Filipkowski, P., and Synowiecki, J. (2013). Effects of the polyhistidine tag on kinetics and other properties of trehalose synthase from Deinococcus geothermalis. *Acta Biochim. Pol.*, 60(2):163–6.

Papaioannou, D., Geibel, S., Kunze, M. B., Kay, C. W., and Waksman, G. (2016). Structural and biophysical investigation of the interaction of a mutant Grb2 SH2 domain (W121G) with its cognate phosphopeptide. *Protein Sci.*, 25(3):627–637.

Park, H.-S., Hohn, M. J., Umehara, T., Guo, L.-T., Osborne, E. M., Benner, J., Noren, C. J., Rinehart, J., and Soll, D. (2011). Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science (80-.).*, 333(6046):1151–1154.

Pawson, T. (1992). SH2 and SH3 domains. *Curr. Opin. Struct. Biol.*, 2(3):432–437.

Pedersen, J. S. and Schurtenberger, P. (1996). Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules*, 29(23):7602–7612.

Penuel, E., Akita, R. W., and Sliwkowski, M. X. (2002). Identification of a Region within the ErbB2/HER2 Intracellular Domain That Is Necessary for Ligand-independent Association. *J. Biol. Chem.*, 277(32):28468–28473.

Pérez, J. and Vachette, P. (2017). A Successful Combination: Coupling SE-HPLC with SAXS. *Adv. Exp. Med. Biol.*, 1009:183–199.

Pérez, J., Vachette, P., Russo, D., Desmadril, M., and Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-β protein investigated by small-angle X-ray scattering 1 1Edited by M. F. Moody. *J. Mol. Biol.*, 308(4):721–743.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.*, 94(23):12366–71.

Pines, G., Huang, P. H., Zwang, Y., White, F. M., and Yarden, Y. (2010). EGFRvIV: A previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism. *Oncogene*, 29(43):5850–5860.

Pinkas-Kramarski, R., Soussan, L., Waterman, H., Levkowitz, G., Alroy, I., Klapper, L., Lavi, S., Seger, R., Ratzkin, B. J., Sela, M., and Yarden, Y. (1996). Diversification of Neu differentiation factor and epidermal growth factor signaling by combinatorial receptor interactions. *EMBO J.*, 15(10):2452–2467.

Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J., Cantley, L. C., Terhorst, C., and Eck, M. J. (1999). Crystal Structures of the XLP Protein SAP Reveal a Class of SH2 Domains with Extended, Phosphotyrosine-Independent Sequence Recognition. *Mol. Cell*, 4(4):555–561.

Press, M. F., Pike, M. C., Chazin, V. R., Hung, G., Udove, J. A., Markowicz, M., Danyluk, J., Godolphin, W., Sliwkowski, M., Akita, R., Paterson, M. C., and Slamon, D. J. (1993). Her-2/neu Expression in Node-negative Breast Cancer: Direct Tissue Quantitation by Computerized Image Analysis and Association of Overexpression with Increased Risk of Recurrent Disease. *Cancer Res.*, 53(20):4960–4970.

Przybylski, C., Jünger, M. A., Aubertin, J., Radvanyi, F., Aebersold, R., and Pflieger, D. (2010). Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-orbitrap instrument. *J. Proteome Res.*, 9(10):5118–5132.

Purba, E., Saita, E.-i., and Maruyama, I. (2017). Activation of the EGF Receptor by Ligand Binding and Oncogenic Mutations: The "Rotation Model". *Cells*, 6(2):13.

Qian, X., LeVea, C. M., Freeman, J. K., Dougall, W. C., and Greene, M. I. (1994). Heterodimerization of epidermal growth factor receptor and wild-type or kinase-deficient Neu: a mechanism of interreceptor kinase activation and transphosphorylation. *Proc. Natl. Acad. Sci.*, 91(4):1500–1504.

Qin, B. Y., Liu, C., Srinath, H., Lam, S. S., Correia, J. J., Derynck, R., and Lin, K. (2005). Crystal structure of IRF-3 in complex with CBP. *Structure*, 13(9):1269–77.

Qiu, C., Lienhard, S., Hynes, N. E., Badache, A., and Leahy, D. J. (2008). Memo Is Homologous to Nonheme Iron Dioxygenases and Binds an ErbB2-derived Phosphopeptide in Its Vestigial Active Site. *J. Biol. Chem.*, 283(5):2734–2740.

Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J., and Grutter, M. G. (1996). Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode. *Nat. Struct Biol*, 3(7):586–589.

Ranganathan, R., Lu, K. P., Hunter, T., and Noel, J. P. (1997). Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell*, 89(6):875–86.

Ravichandran, K. S., Lorenz, U., Shoelson, S. E., and Burakoff, S. J. (1995). Interaction of Shc with Grb2 regulates association of Grb2 with mSOS. *Mol. Cell. Biol.*, 15(2):593–600.

Ravid, T., Heidinger, J. M., Gee, P., Khan, E. M., and Goldkorn, T. (2004). c-Cbl-mediated Ubiquitinylation Is Required for Epidermal Growth Factor Receptor Exit from the Early Endosomes. *J. Biol. Chem.*, 279(35):37153–37162.

Receveur-Brechot, V. and Durand, D. (2012). How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sci.*, 13(1):55–75.

Red Brewer, M., Choi, S. H., Alvarado, D., Moravcevic, K., Pozzi, A., Lemmon, M. A., and Carpenter, G. (2009). The Juxtamembrane Region of the EGF Receptor Functions as an Activation Domain. *Mol. Cell*, 34(6):641–651.

Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998). Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.*, 279(2):449–460.

Riethmacher, D., Sonnenberg-Riethmacher, E., Brinkmann, V., Yamaai, T., Lewin, G. R., and Birchmeier, C. (1997). Severe neuropathies in mice with targeted mutations in the ErbB3 receptor. *Nature*, 389(6652):725–730.

Robinson, D. R., Wu, Y.-M., and Lin, S.-F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene*, 19(49):5548–5557.

Rogers, J. M., Wong, C. T., and Clarke, J. (2014). Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure. *J. Am. Chem. Soc.*, 136(14):5197–5200.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.*, pages 437–48.

Rosano, G. L. and Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.*, 5(APR):1–17.

Rosenlöw, J., Isaksson, L., Mayzel, M., Lengqvist, J., and Orekhov, V. Y. (2014). Tyrosine Phosphorylation within the Intrinsically Disordered Cytosolic Domains of the B-Cell Receptor: An NMR-Based Structural Analysis. *PLoS One*, 9(4):e96199.

Ruan, H., Sun, Q., Zhang, W., Liu, Y., and Lai, L. (2019). Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discov. Today*, 24(1):217–227.

Sadowski, I., Stone, J. C., and Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.*, 6(12):4396–408.

Saksela, K. and Permi, P. (2012). SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Lett.*, 586(17):2609–2614.

Salvi, N., Abyzov, A., and Blackledge, M. (2017). Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.*, 102-103:43–60.

Schechter, A., Hung, M., Vaidyanathan, L., Weinberg, R., Yang-Feng, T., Francke, U., Ullrich, A., and Coussens, L. (1985). The neu gene: an erbB-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science (80-.).*, 229(4717):976–978.

Schechter, A. L., Stern, D. F., Vaidyanathan, L., Decker, S. J., Drebin, J. A., Greene, M. I., and Weinberg, R. A. (1984). The neu oncogene: An erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature*, 312(5994):513–516.

Schiering, N., Casale, E., Caccia, P., Giordano, P., and Battistini, C. (2000). Dimer Formation through Domain Swapping in the Crystal Structure of the Grb2-SH2-Ac-pYVNV Complex. *Biochemistry*, 39(44):13376–13382.

Schulze, W. X., Deng, L., and Mann, M. (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.*, 1(1):E1–E13.

Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997). Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. *Biochemistry*, 36(29):8977–8991.

Sciolino, N., Burz, D. S., and Shekhtman, A. (2019). In-Cell NMR Spectroscopy of Intrinsically Disordered Proteins. *Proteomics*, 19(6):1800055.

Sebastian, J., Richards, R. G., Walker, M. P., Wiesen, J. F., Werb, Z., Derynck, R., Hom, Y. K., Cunha, G. R., and DiAugustine, R. P. (1998). Activation and function of the epidermal growth factor receptor and erbB-2 during mammary gland morphogenesis. *Cell Growth Differ.*, 9(9):777–85.

Segatto, O., Lonardo, F., Pierce, J. H., Bottaro, D. P., and Di Fiore, P. P. (1990). The role of autophosphorylation in modulation of erbB-2 transforming function. *New Biol.*, 2(2):187–95.

Selenko, P., Frueh, D. P., Elsaesser, S. J., Haas, W., Gygi, S. P., and Wagner, G. (2008). In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat. Struct. Mol. Biol.*, 15(3):321–329.

Senior, M. M., Frederick, A. F., Black, S., Murgolo, N. J., Perkins, L. M., Wilson, O., Snow, M. E., and Wang, Y. S. (1998). The three-dimensional solution structure of the Src homology domain-2 of the growth factor receptor-bound protein-2. *J. Biomol. NMR*, 11(2):153–164.

Sharp, P. and Bloomfield, V. A. (1968). Light scattering from wormlike chains with excluded volume effects. *Biopolymers*, 6(8):1201–1211.

Sharpe, S., Barber, K. R., and Grant, C. W. M. (2002). Evidence of a Tendency to Self-Association of the Transmembrane Domain of ErbB-2 in Fluid Phospholipid Bilayers. *Biochemistry*, 41(7):2341–2352.

Shen, Y. and Bax, A. (2010). Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J. Biomol. NMR*, 46(3):199–204.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, 44(4):213–23.

Shen, Y., Roche, J., Grishaev, A., and Bax, A. (2018). Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. *Protein Sci.*, 27(1):146–158.

Shrestha, A., Hamilton, G., O'Neill, E., Knapp, S., and Elkins, J. M. (2012). Analysis of conditions affecting auto-phosphorylation of human kinases during expression in bacteria. *Protein Expr. Purif.*, 81(1):136–143.

Sibille, N. and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.*, 40(5):955–962.

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235(4785):177–82.

Smith, L. J., Bolin, K. a., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996). Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.*, 255(3):494–506.

Sólyom, Z., Ma, P., Schwarten, M., Bosco, M., Polidori, A., Durand, G., Willbold, D., and Brutscher, B. (2015). The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophys. J.*, 109(7):1483–1496.

Song, J., Guo, L.-W., Muradov, H., Artemyev, N. O., Ruoho, A. E., and Markley, J. L. (2008). Intrinsically disordered -subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure. *Proc. Natl. Acad. Sci.*, 105(5):1505–1510.

Sorkin, A. and Goh, L. K. (2009). Endocytosis and intracellular trafficking of ErbBs. *Exp. Cell Res.*, 315(4):683–696.

Sorkin, A., Helin, K., Waters, C. M., Carpenter, G., and Beguinot, L. (1992). Multiple autophosphorylation sites of the epidermal growth factor receptor are essential for receptor kinase activity and internalization. Contrasting significance of tyrosine 992 in the native and truncated receptors. *J. Biol. Chem.*, 267(12):8672–8.

Sriram, G. and Birge, R. B. (2012). Commentary: The carboxyl-terminal Crk SH3 domain: Regulatory strategies and new perspectives. *FEBS Lett.*, 586(17):2615–2618.

Stamos, J., Sliwkowski, M. X., and Eigenbrot, C. (2002). Structure of the Epidermal Growth Factor Receptor Kinase Domain Alone and in Complex with a 4-Anilinoquinazoline Inhibitor. *J. Biol. Chem.*, 277(48):46265–46272.

Stein, D., Wu, J., Fuqua, S., Roonprapunt, C., Yajnik, V., D'Eustachio, P., Moskow, J., Buchberg, A., Osborne, C., and Margolis, B. (1994). The SH2 domain protein GRB-7 is co-amplified, overexpressed and in a tight complex with HER2 in breast cancer. *EMBO J.*, 13(6):1331–1340.

Stein, R. A. and Staros, J. V. (2006). Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol. Biol.*, 6:1–17.

Strack, S. and Colbran, R. J. (1998). Autophosphorylation-dependent Targeting of Calcium/ Calmodulin-dependent Protein Kinase II by the NR2B Subunit of the N -Methyl- d-aspartate Receptor. *J. Biol. Chem.*, 273(33):20689–20692.

Su, J., Yang, L.-T., and Sap, J. (1996). Association between Receptor Protein-Tyrosine Phosphatase RPTPa and the Grb2 Adaptor. *J. Biol. Chem.*, 271(45):28086–28096.

Svergun, D., Barberato, C., and Koch, M. H. (1995). CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, 28(6):768–773.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, 25(4):495–503.

Swain, S. M., Baselga, J., Kim, S.-B., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J.-M., Schneeweiss, A., Heeson, S., Clark, E., Ross, G., Benyunes, M. C., and Cortés, J. (2015). Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. *N. Engl. J. Med.*, 372(8):724–734.

Tao, R.-H. and Maruyama, I. N. (2008). All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. *J. Cell Sci.*, 121(19):3207–3217.

Theillet, F.-X., Binolfi, A., Bekei, B., Martorana, A., Rose, H. M., Stuiver, M., Verzini, S., Lorenz, D., van Rossum, M., Goldfarb, D., and Selenko, P. (2016). Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature*, 530(7588):45–50.

Theillet, F.-X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., Li, C., Crowley, P. B., Gierasch, L., Pielak, G. J., Elcock, A. H., Gershenson, A., and Selenko, P. (2014). Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.*, 114(13):6661–6714.

Theillet, F.-X., Smet-Nocca, C., Liokatis, S., Thongwichian, R., Kosten, J., Yoon, M.-K., Kriwacki, R. W., Landrieu, I., Lippens, G., and Selenko, P. (2012). Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR*, 54(3):217–236.

Thiel, K. W. and Carpenter, G. (2007). Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proc. Natl. Acad. Sci.*, 104(49):19238–19243.

Thielges, M. C., Chung, J. K., Axup, J. Y., and Fayer, M. D. (2011). Influence of Histidine Tag Attachment on Picosecond Protein Dynamics. *Biochemistry*, 50(25):5799–5805.

Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009). Analytical strategies for phosphoproteomics. *Proteomics*, 9(6):1451–1468.

Thornton, K. H., Mueller, W. T., McConnell, P., Zhu, G., Saltiel, A. R., and Thanabal, V. (1996). Nuclear magnetic resonance solution structure of the growth factor receptor-bound protein 2 Src homology 2 domain. *Biochemistry*, 35(36):11852–11864.

Thorsness, P. E. and Koshland, D. E. (1987). Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. *J. Biol. Chem.*, 262(22):10422–5.

Tompa, P. (2014). Multisteric Regulation by Structural Disorder in Modular Signaling Proteins: An Extension of the Concept of Allostery. *Chem. Rev.*, 114(13):6715–6732.

Tompa, P. and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.*, 33(1):2–8.

Tria, G., Mertens, H. D. T., Kachala, M., and Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, 2(2):207–217.

Tzahar, E., Waterman, H., Chen, X., Levkowitz, G., Karunagaran, D., Lavi, S., Ratzkin, B. J., and Yarden, Y. (1996). A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol. Cell. Biol.*, 16(10):5276–5287.

Ugocsai, K., Mándoky, L., Tiszlavicz, L., and Molnár, J. (2005). Investigation of HER2 overexpression in non-small cell lung cancer. *Anticancer Res.*, 25(4):3061–6.

Urbanek, A., Morató, A., Allemand, F., Delaforge, E., Fournet, A., Popovic, M., Delbecq, S., Sibille, N., and Bernadó, P. (2018). A General Strategy to Access Structural Information at Atomic Resolution in Polyglutamine Homorepeats. *Angew. Chemie - Int. Ed.*, 57(14):3598–3601.

Uversky, V. N. (2002). Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.*, 11(4):739–756.

Uversky, V. N. (2010). Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D 2 concept. *Expert Rev. Proteomics*, 7(4):543–564.

Uversky, V. N. (2014). Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. *Front. Mol. Biosci.*, 1(July):1–24.

Uversky, V. N. (2015). Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front. Aging Neurosci.*, 7(March):1–6.

Uversky, V. N. (2017a). *Intrinsic Disorder, Protein-Protein Interactions, and Disease*. Elsevier Inc., 1 edition.

Uversky, V. N. (2017b). Protein intrinsic disorder-based liquid-liquid phase transitions in biological systems: Complex coacervates and membrane-less organelles. *Adv. Colloid Interface Sci.*, 239:97–114.

Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.*, 41(3):415–427.

Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005). Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling.

Valius, M. and Kazlauskas, A. (1993). Phospholipase C-γ1 and phosphatidylinositol 3 kinase are the downstream mediators of the PDGF receptor's mitogenic signal. *Cell*, 73(2):321–334.

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, 114(13):6589–6631.

Vaskovsky, A., Lupowitz, Z., Erlich, S., and Pinkas-Kramarski, R. (2000). ErbB-4 Activation Promotes Neurite Outgrowth in PC12 Cells. *J. Neurochem.*, 74(3):979–987.

Vidal, M., Goudreau, N., Cornille, F., Cussac, D., Gincel, E., and Garbay, C. (1999). Molecular and cellular analysis of Grb2 SH3 domain mutants: Interaction with Sos and dynamin. *J. Mol. Biol.*, 290(3):717–730.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*, 59(4):687–696.

Vuister, G. W. and Bax, A. (1993). Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha)) Coupling-Constants in N-15-Enriched Proteins. *J. Am. Chem. Soc.*, 115(17):7772–7777.

Wagner, M. J., Stacey, M. M., Liu, B. A., and Pawson, T. (2013). Molecular mechanisms of SH2- and PTB-Domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harb. Perspect. Biol.*, 5(12).

Walton, G. M., Chen, W. S., Rosenfeld, M. G., and Gill, G. N. (1990). Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced in vivo tyrosine phosphorylation of cell substrates. *J. Biol. Chem.*, 265(3):1750–4.

Wang, Y., Pinet, L., Assrir, N., Elantak, L., Guerlesquin, F., Badache, A., Lescop, E., and van Heijenoort, C. (2018). 1H, 13C and 15N assignments of the C-terminal intrinsically disordered cytosolic fragment of the receptor tyrosine kinase ErbB2. *Biomol. NMR Assign.*, 12(1):23–26.

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.*, 337(3):635–645.

Ward, M. D. and Leahy, D. J. (2015). Kinase Activator-Receiver Preference in ErbB Heterodimers Is Determined by Intracellular Regions and Is Not Coupled to Extracellular Asymmetry. *J. Biol. Chem.*, 290(3):1570–1579.

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M., and Heery, D. M. (2006). Structural diversity in p160/CREB-binding protein coactivator complexes. *J. Biol. Chem.*, 281(21):14787–95.

Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999). Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry*, 38(50):16424–16431.

Wilson, K. J., Mill, C., Lambert, S., Buchman, J., Wilson, T. R., Hernandez-Gordillo, V., Gallo, R. M., Ades, L. M., Settleman, J., and Riese, D. J. (2012). EGFR ligands exhibit functional differences in models of paracrine and autocrine signaling. *Growth Factors*, 30(2):107–116.

Wishart, D. S. and Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.*, 239(1982):363–92.

Wittekind, M., Mapelli, C., Farmer, B. T., Suen, K.-L., Goldfarb, V., Tsao, J., Lavoie, T., Barbacid, M., Meyers, C. A., and Mueller, L. (1994). Orientation of Peptide Fragments from Sos Proteins Bound to the N-Terminal SH3 Domain of Grb2 Determined by NMR Spectroscopy. *Biochemistry*, 33(46):13531–13539.

Wittekind, M., Mapelli, C., Lee, V., Goldfarb, V., Friedrichs, M. S., Meyers, C. A., and Mueller, L. (1997). Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: Direct refinement against NOEs, J-couplings and 1 H and 13 C chemical shifts. *J. Mol. Biol.*, 267(4):933–952.

Wood, E. R., Shewchuk, L. M., Ellis, B., Brignola, P., Brashear, R. L., Caferro, T. R., Dickerson, S. H., Dickson, H. D., Donaldson, K. H., Gaul, M., Griffin, R. J., Hassell, A. M., Keith, B., Mullin, R., Petrov, K. G., Reno, M. J., Rusnak, D. W., Tadepalli, S. M., Ulrich, J. C., Wagner, C. D., Vanderwall, D. E., Waterson, A. G., Williams, J. D., White, W. L., and Uehling, D. E. (2008). 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases. *Proc. Natl. Acad. Sci.*, 105(8):2773–2778.

Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., Alligood, K. J., Rusnak, D. W., Gilmer, T. M., and Shewchuk, L. (2004). A Unique Structure for Epidermal Growth Factor Receptor Bound to GW572016 (Lapatinib). *Cancer Res.*, 64(18):6652–6659.

Wright, P. E. and Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, 293(2):321–331.

Wright, P. E. and Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.*, 19(1):31–38.

Wulf, G., Garg, P., Liou, Y.-C., Iglehart, D., and Lu, K. P. (2004). Modeling breast cancer in vivo and ex vivo reveals an essential role of Pin1 in tumorigenesis. *EMBO J.*, 23(16):3397–3407.

Wyszynski, M., Lin, J., Rao, A., Nigh, E., Beggs, A. H., Craig, A. M., and Sheng, M. (1997). Competitive binding of α-actinin and calmodulin to the NMDA receptor. *Nature*, 385(6615):439–442.

Xie, Y., Pendergast, A. M., and Hung, M. C. (1995). Dominant-negative mutants of Grb2 induced reversal of the transformed phenotypes caused by the point mutation-activated rat HER-2/Neu. *J. Biol. Chem.*, 270(51):30717–24.

Yamazaki, T., Zaal, K., Hailey, D., Presley, J., Lippincott-Schwartz, J., and Samelson, L. E. (2002). Role of Grb2 in EGF-stimulated EGFR internalization. *J. Cell Sci.*, 115(Pt 9):1791–802.

Yarden, Y. and Pines, G. (2012). The ERBB network: at last, cancer therapy meets systems biology. *Nat. Rev. Cancer*, 12(8):553–563.

Yarden, Y. and Schlessinger, J. (1987). Self-phosphorylation of epidermal growth factor receptor: evidence for a model of intermolecular allosteric activation. *Biochemistry*, 26(5):1434–1442.

Yarden, Y. and Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.*, 2(2):127–137.

Yuzawa, S., Yokochi, M., Hatanaka, H., Ogura, K., Kataoka, M., Miura, K., Mandiyan, V., Schlessinger, J., and Inagaki, F. (2001). Solution structure of Grb2 reveals extensive flexibility necessary for target recognition. *J. Mol. Biol.*, 306(3):527–37.

Zaoui, K., Honoré, S., Isnardon, D., Braguer, D., and Badache, A. (2008). Memo-RhoA-mDia1 signaling controls microtubules, the actin network, and adhesion site formation in migrating cells. *J. Cell Biol.*, 183(3):401–408.

Zhang, O. and Forman-Kay, J. D. (1995). Structural Characterization of Folded and Unfolded States of an SH3 Domain in Equilibrium in Aqueous Buffer. *Biochemistry*, 34(20):6784–6794.

Zhang, X., Gureasko, J., Shen, K., Cole, P. A., and Kuriyan, J. (2006). An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell*, 125(6):1137–1149.

Zhang, Y., Ptacin, J. L., Fischer, E. C., Aerni, H. R., Caffaro, C. E., San Jose, K., Feldman, A. W., Turner, C. R., and Romesberg, F. E. (2017). A semi-synthetic organism that stores and retrieves increased genetic information. *Nature*, 551(7682):644–647.

Zhou, H. X. (2012). Intrinsic disorder: Signaling via highly specific but short-lived association. *Trends Biochem. Sci.*, 37(2):43–48.

Zhou, H. X., Pang, X., and Lu, C. (2012). Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys. Chem. Chem. Phys.*, 14(30):10466–10476.

Zhou, M. M., Ravichandran, K. S., Olejniczak, E. F., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J., and Fesik, S. W. (1995). Structure and ligand recognition of the phosphotyrosine binding domain of Shc. *Nature*, 378(6557):584–92.

Zhu, G., Liu, Y., and Shaw, S. (2005). Protein Kinase Specificity: A Strategic Collaboration between Kinase Peptide Specificity and Substrate Recruitment. *Cell Cycle*, 4(1):52–56.

Appendices

SUPPLEMENTARY INFORMATION

I M9 medium composition

Component	Concentration	
	^{15}N M9	^{15}N^{13}C M9
Na_2HPO_4	45 mM (6.4 g/L)	
KH_2PO_4	22 mM (3 g/L)	
NaCl	8.5 mM (0.5 g/L)	
$^{15}NH_4Cl$	18.5 mM (1 g/L)	
$MgSO_4$	2 mM	
$CaCl_2$	10 μM	
Vitamin mix	Gibco^{TM} MEM Vitamin Solution	
Trace elements	see (b)	
Glucose	4g/L	X
^{13}C Glucose	X	3g/L

(a) General M9 composition

Compound	Concentration (mM)
EDTA	17.1
$CaCl_2$	40.8
$CuSO_4$	17.6
$MnCl_2$	6
H_3BO_3	0.323
$ZnSO_4$	2.345
$FeSO_4$	21.6
Ascorbic acid	1.1

(b) 100X stock solution of trace elements

Table S1 – M9 medium composition (for [U-^{15}N] or [U-^{15}N; U-^{13}C] labeling). For deuteration, the same M9 is used with D_2O as solvent instead of H_2O.

PUBLICATIONS

II 1H, 13C and 15N assignments of the C-terminal intrinsically disordered cytosolic fragment of the receptor tyrosine kinase ErbB2

III The intrinsically disordered C-terminal tail of the ErbB2 receptor exhibits transiently structured elements likely to have functional relevance

This publication has been submitted to the *Journal of Biological Chemistry*.

IV Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data

Titre: Exploration structurale et fonctionnelle de la partie C-terminale intrinsèquement désordonnée de ErbB2

Mots clés: Résonance magnétique nucléaire (RMN), protéine intrinsèquement désordonnée (IDP), récepteur tyrosine kinase ErbB2, protéine adaptatrice Grb2, phosphorylation de tyrosines, interaction protéine-protéine

Résumé: ErbB2/HER2 est un récepteur tyrosine kinase de la famille d'EGFR (ErbB1) surexprimé dans plus de 20% des cancers du sein et associé à une forme particulièrement agressive de la maladie. Les récepteurs ErbBs sont actifs seulement sous forme de dimères, permettant la phosphorylation de leur queue C-terminale par leur domaine tyrosine kinase. La phosphorylation entraine l'interaction avec des protéines adaptatrices et l'activation de voies de signalisation, Ras/MAPK et PI3K/Akt principalement. Ces voies contrôlent la prolifération, la motilité cellulaire et la résistance à l'apoptose. Contrairement à ErbB1/3/4, ErbB2 dimérise en l'absence de ligand. Comprendre les autres mécanismes de régulation de la phosphorylation de ses tyrosines et de ses interactions est donc particulièrement intéressant.
ErbB2 a fait l'objet de nombreuses études structurales et fonctionnelles. Elles ont permis la mise au point de traitements ciblés efficaces mais sujets à l'apparition de résistance, dont l'anticorps Trastuzumab, ciblant sa partie extracellulaire. La queue C-terminale d'ErbB2 (CtErbB2) a été très souvent ignorée dans ces études. Cette partie étant intrinsèquement désordonnée, il a fallu attendre ces dernières années pour que les concepts et les outils permettant de l'étudier émergent.
Dans cette thèse, j'ai d'abord effectué la caractérisation structurale et dynamique de CtErbB2. J'ai montré que bien qu'étant dépourvue de toute structure stable, cette région riche en prolines possède plusieurs structures secondaires transitoires et un contact longue-distance participant très probablement à la régulation de ses interactions intra- et inter-moléculaires. Dans une deuxième partie je me suis intéressée à la caractérisation de la protéine adaptatrice Grb2, partenaire essentiel de ErbB2 pour l'activation de la voie des MAP kinases. L'organisation en solution des domaines de cette protéine modulaire dans sa forme libre était jusque là inconnue. J'ai ensuite étudié l'interaction entre Grb2 et CtErbB2, et montré que CtErbB2 interagit non seulement avec le domaine SH2 de Grb2 (par l'intermédiaire d'une phosphotyrosine), mais aussi avec son domaine SH3 N-terminal (grâce à un motif polyproline). Enfin, j'ai mis en place plusieurs stratégies de phosphorylation des tyrosines de CtErbB2, dans le but d'étudier plus largement l'effet des phosphorylations sur l'ensemble de cette région.

Title: Structural and functional investigation of the C-terminal intrinsically disordered fragment of ErbB2

Keywords: Nuclear magnetic resonance (NMR), intrinsically disordered protein (IDP), receptor tyrosine kinase ErbB2, adaptor protein Grb2, tyrosine phosphorylation, protein-protein interaction

Abstract: ErbB2/HER2 is a receptor tyrosine kinase of the EGFR (ErbB1) family overexpressed in 20% of breast cancers and associated to a particularly aggressive form of the disease. ErbB receptors are only active upon dimerization that enables phosphorylation of their C-terminal tail by their tyrosine kinase domain. Phosphorylation then triggers interaction with adaptor proteins and activation of signaling pathways, mainly Ras/MAPK and Akt/PI3K. Those pathways control cell proliferation, motility and resistance to apoptosis. Contrary to ErbB1/3/4, ErbB2 can dimerize without any ligand. Understanding other mechanisms of regulation of its tyrosine phosphorylation and of its interactions is thus particularly interesting.
ErbB2 structure and function have been extensively studied. This has led to the development of several FDA-approved targeted drugs, that are effective but to which resistance occurs, amongst which the Trastuzumab antibody that targets ErbB2 extracellular domain. The C-terminal tail of ErbB2 (CtErbB2) has been widely ignored in these studies. Since it is intrinsically disordered, the concepts and tools to study it have only emerged in the last few years.
In the present work, I have performed the structural and dynamic study of CtErbB2. I showed that despite its lack of any stable structure, this proline-rich region exhibits several transient secondary structures and a long-range contact that might participate in the regulation of its intra- and inter-molecular interactions. Then, I characterized the adaptor protein Grb2, which is a partner of ErbB2 that is essential for the activation of the MAPK pathway. The solution organization of the domains of this modular protein in its apo-form was unknown so far. I also studied the interaction between Grb2 and CtErbB2, showing that in addition to the known SH2-phosphotyrosine interaction, a polyproline motif of CtErbB2 binds to the N-terminal SH3 domain of Grb2. Finally, I implemented several strategies to phosphorylate CtErbB2 tyrosines, to study more extensively the effect of phosphorylation on the whole tail.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France