



Thèse présentée pour obtenir le titre de

Docteur de l'Université Louis Pasteur Strasbourg 1

Discipline : Sciences du Vivant
Spécialité : Biologie Structurale

par

Yann BRELIVET

Etude de voies de communication Application aux Récepteurs Nucléaires

Soutenue publiquement le 10 mars 2008 devant le jury :

Directeur de thèse
Rapporteur interne
Rapporteur externe
Rapporteur externe
Examineur
Examineur

Dino MORAS, Illkirch
Hélène DOLLFUS, Strasbourg
Vincent LAUDET, Lyon
William BOURGUET, Montpellier
Etienne BAULIEU, Paris
Olivier POCH, Illkirch

A mes parents,

Remerciements

Avant tout, je tiens à exprimer ma profonde reconnaissance aux Professeurs Hélène Dollfus, Etienne Baulieu, Vincent Laudet et William Bourguet, Directeur de Recherche, pour l'honneur qu'ils me font de juger cette thèse.

Cette thèse ne serait rien sans remercier, particulièrement, trois Hommes qui m'ont accompagné dans ce travail. Vous m'avez fait, tous les trois, avancer dans la science mais aussi dans la vie. Mon seul regret est de ne pas vous avoir rencontré plus tôt.

Mon directeur de thèse, Dino Moras, pour la formation scientifique de qualité qu'il m'a généreusement transmise, pour sa grande confiance et la liberté qu'il m'a accordé dans mes recherches et pour ses qualités humaines qu'il est encore rare de trouver aujourd'hui,

Mon non officiel « co-directeur de thèse », Olivier Poch. Cette passion qui t'anime, cet entrain, cette énergie font de toi ce que tu es ! Je me rappelle mes débuts, nos soirées à aligner des séquences... Tu as su me guider, me motiver, trouver les mots justes au moment opportun... tout simplement : *Merci ! « La Séquence ne ment jamais ! »* Cette thèse le confirme bien, non ?

Jean-Claude Thierry qui s'est toujours tenu informé de l'avancement de ma thèse, des conseils judicieux qu'il m'a prodigués et qui a su trouver les mots justes pour me rassurer dans mes « quelques » moments de doute en cette fin de thèse

Que de monde à remercier ! Allez, je me lance !!!

Remerciements à tous les membres du Laboratoire de Bio-informatique et Génomique Intégratives, en commençant par les piliers (non non, pas de bar) Raymond, Julie, Odile, Fred et Luc. Raymond, ta gentillesse, toujours disponible, le dur apprentissage du Tcl (pas convaincu ?) toujours partant pour une petite escapade au Mont Blanc ?... Notre Julie Nationale, que dis-je Internationale. Cette thèse est-elle « big-or-no ? » Ou bien parle-t-on de mollusques ??? Odile, ton humour et ta sympathie, a quand le prochain barbecue ? Fred, ton calme, la zen attitude en toute circonstance et Luc, ton humour et ta joie de vivre.

Et puis, il y a tous les autres...

Les résidents du mythique couloir Laetitia (membre du club des croqueurs de chokobons) et Manu (nos discussions ludiques et autres) ; Laurent, Sophie, Véro (toujours la pêche et avec le sourire svp, on se fait une soirée crêpes quand tu veux !), Wolfgang, Hoan (comme dirait Anne le meilleur standardiste vietnamien de

l'Institut), Ravi (jeune Docteur toujours zen), Nicolas (toujours une petite anecdote qui donne le sourire et fait du bien dans les moments de moins bien), Fryar (les Locusts n'ont qu'à bien se tenir !), et bien sur Anne, ma compagne de rédaction (Oh ça vaaa ! On est passé par des moments, hein ?), Claudine et nos aventures au synchrotron, les p'tis jeunes qui règnent en maîtres dans leur module Guillaume, Radwen, Laurent-P (humm la théorie de la co-évolution), David et Nicolas (et les RNs chez Alvinella ???), Yannick et pour finir les deux p'tits nouveaux Florence et Francisco : bienvenue à vous deux en Bioinfo !!!

Et puis, il y aussi les expatrié(e)s : Aurélie et Fred, Adeline (j'attends toujours ta prestation sur scène !) et Jean.

Merci à Serge pour la gestion des serveurs (prêt pour la montée du mont Ste Odile ?) et Alain (l'électronique n'a plus de secrets pour toi), à Christiane notre jeune retraitée, Anne (ta gentillesse, ta disponibilité, toujours à l'écoute) et Laetitia (merci pour les encouragements et pour les petites attentions culinaires ;-)).

Mes remerciements vont également à tous les membres du Département de Biologie et de Génomique Structurales et plus particulièrement :

Marc Ruff qui m'a initié à la cristallographie aux rayons X, Sylvia pour sa rigueur, son organisation et l'apprentissage des techniques de purification, Florence pour sa gentillesse, sa bonne humeur (même le lundi ?), JMG *alias* Jean-Marie (allez je te mets dans le groupe), Geo Trouvetou *alias* Benwa, Carole et Virginie, toujours disponibles pour répondre à mes questions techniques en fin de rédaction, Natacha R., Annick D. (merci, car c'est aussi grâce à toi que j'ai pu faire cette thèse), Chris, Yannick B., Serena (merci pour ton sourire, courage pour la fin de thèse), Isabelle H., Alexandra C., Tiphaine, Catherine, Bénédicte D-B., Isabelle B., Pierre A., Ferdinand, Judit, Anne-Catherine et tous les autres...

Je tiens aussi à remercier les membres de la plate-forme de Biologie et de Génomique Structurales : Loubna (toujours accueillante), Edouard (et ses remèdes naturels et les bons gâteaux), Matthieu (Oh cong !), Pierre P. (Maître *ès* robotique) et bien sûr Didier (pour toutes les discussions scientifiques et autres).

Merci également aux différents services communs qui nous facilitent quand même bien la vie au labo, y'a pas à dire ! Donc :

Serve Vicaire pour le séquençage d'ADN, Pascal Eberling pour la synthèse de peptides, l'équipe de Betty Heller pour la culture cellulaire, Manuela Argentini pour la protéomique, Hélène et Maïté pour leur gentillesse et leur disponibilité. Mais non, Armelle je ne t'ai pas oubliée !

J'aimerais également remercier Norbert, Natacha, Charlotte, Eva, Anne-So qui ont pris « un peu » de leur temps pour m'expliquer certaines manips.

Merci à tous les membres de l'OPPF (*Oxford Protein Production Facility*) que j'ai pu rencontrer lors d'un *Workshop* très riche en enseignements. L'accueil et la qualité scientifique étaient des plus remarquables.

Merci à la team Téléthon : Josiane, Mims, Anne-Sophie, Marie, Chantal... C'est reparti pour un tour cette année ?

Mille mercis à mes petites fées nourricières : Laetitia, Anne F., Anne N., Florence.

Merci à toute ma famille de Bretagne et d'Alsace et à tous mes amis pour leurs encouragements. Ils vont enfin pouvoir comprendre ce que j'ai fait pendant tant d'années. Les « Bretons » : Mémé, Annie, Robert, Jacques, Céline, Ronan, Gaëlle, Roger, Annette, Lionel. Les « Alsaciens » : René, Suzanne, Sandrine, Fabien, Matthieu, Louise. Mes ami(e)s : Sophie H. (c'est toujours un vrai plaisir pour moi de te revoir), la famille Dromard avec Laurence, Patrick et Lucas qui va bientôt être grand frère d'ailleurs, Romu, Berengère, Agathe, Sophie R. (j'ai droit à ma ratatouille, non ?), Sylvia, Dom, Karine, Cyril, Patrick T. ...

Un énorme MERCI de tout coeur à mes parents qui ont permis que tout cela se réalise et qui ont toujours été là pour m'encourager.

Enfin j'aimerais dire merci à mes compagnons de route anonymes rencontrés lors de mes sorties cyclistes. Si vous, lecteurs êtes tentés par l'aventure de Paris-Brest-Paris, faites moi signe !!!

Table des matières

Introduction	1 -
Chapitre 1 - La super-famille des récepteurs nucléaires	6 -
1.1. Fonctions biologiques et importance thérapeutique	6 -
1.2. Organisation modulaire des récepteurs nucléaires.....	8 -
1.2.1. Le domaine A/B	9 -
1.2.1.1. Structure et fonction	9 -
1.2.1.2. Interaction du domaine A/B avec des coactivateurs	10 -
1.2.1.3. Modifications post-traductionnelles	11 -
1.2.2. Le domaine C	12 -
1.2.2.1. Structure et fonction	12 -
1.2.2.2. Eléments structuraux impliqués dans la liaison à l'ADN et la dimérisation..	13 -
1.2.2.3. Les éléments de réponse.....	14 -
1.2.3. Le domaine D	19 -
1.2.4. Le domaine E	20 -
1.2.4.1. Structure et fonction	20 -
1.2.4.2. Interface de dimérisation	21 -
1.2.4.3. La poche de fixation du ligand.....	22 -
1.2.4.4. Le mécanisme du piège à souris.....	25 -
1.2.5. Le domaine F.....	26 -
1.2.5.1. Structure	26 -
1.2.5.2. Fonction.....	27 -
1.3. Evolution des récepteurs nucléaires.....	28 -
1.3.1. Un peu d'histoire.....	28 -
1.3.1.1. Précambrien : 4600-540 millions d'années	29 -
1.3.1.2. Paléozoïque : 540-250 millions d'années.....	29 -
1.3.1.3. Mésozoïque : 250-65 millions d'années.....	29 -
1.3.1.4. Cénozoïque : 65 millions d'années - Aujourd'hui.....	30 -
1.3.2.... Et un peu de systématique	30 -
1.3.3. Evolution et répartition phylogénétique des RNs	36 -
1.3.3.1. Origines de la diversification de la super-famille des RNs	38 -
1.3.3.2. Acquisition de la capacité à fixer un ligand	39 -
1.4. Mécanisme de la transactivation des RNs et sa régulation.....	42 -
1.4.1. Mécanisme général.....	42 -
1.4.2. Les cofacteurs.....	43 -
1.4.2.1. Les coactivateurs	44 -
1.4.2.1.1. SRC/p160	44 -
1.4.2.1.2. CBP/p300	44 -
1.4.2.2. Les cointégrateurs	46 -
1.4.2.3. Les corépresseurs	46 -
1.4.3. La transactivation	47 -
1.5. Les bases moléculaires de l'antagonisme	52 -
1.6. Modifications post-traductionnelles des RNs	54 -
1.7. Communication croisée avec d'autres facteurs de transcription	56 -
1.8. Localisation cellulaire	56 -
1.8.1. Localisation cytoplasmique.....	56 -
1.8.2. Localisation nucléaire	59 -
1.9. Objectifs de la thèse	60 -
Matériel et méthodes	61 -

Chapitre 2 - Bioinformatique	63 -
2.1. Ressources informatiques et bioinformatiques	63 -
2.1.1. Équipement et ressources informatiques	64 -
2.1.2. Savoir-faire bioinformatique du laboratoire	64 -
2.1.2.1. GScope : l'ossature bioinformatique du laboratoire	64 -
2.1.2.2. Ecriture de programmes : le langage Tcl/Tk	66 -
2.1.3. Les bases de données biologiques : bases de données ou données de base ?	66 -
2.1.3.1. Bases de données généralistes	66 -
2.1.3.1.1. Uniprot	67 -
2.1.3.1.2. PDB	68 -
2.1.3.2. Bases de données spécialisées	69 -
2.1.4. L'interrogation des banques	70 -
2.1.4.1. SRS	70 -
2.1.4.2. Recherche de similarité dans les bases de données	71 -
2.1.5. PipeAlign : de la protéine à l'alignement multiple	72 -
2.1.5.1. Ballast : traitement des résultats des recherches BlastP	73 -
2.1.5.2. DbClustal : construction de MACS	74 -
2.1.5.3. RASCAL : correction des MACS	74 -
2.1.5.4. LEON : évaluation de l'alignement multiple basée sur l'homologie	75 -
2.1.5.5. NorMD : évaluation de la qualité d'un MACS	75 -
2.1.5.6. Secator et DPC : classification des séquences au sein du MACS	76 -
2.1.5.7. Les autres outils informatiques utilisés	77 -
2.1.5.7.1. MACSIMS : gestion de l'information au sein des MACS	77 -
2.1.5.7.2. OrdAlie : analyse hiérarchisée d'alignements multiples	78 -
2.1.6. Les suites de programmes d'analyse de séquences	79 -
2.1.6.1. Le package GCG	79 -
2.1.6.2. EMBOSS	79 -
2.1.7. Analyse structurale des protéines	80 -
2.1.7.1. Le logiciel O	80 -
2.1.7.2. PyMOL	80 -
2.1.7.3. Modeller : construction de modèles par homologie	81 -
Chapitre 3 - Biochimie, biologie moléculaire et cellulaire	83 -
3.1. Techniques de biologie moléculaire et cellulaire	83 -
3.1.1. Les souches d' <i>Escherichia coli</i> utilisées	83 -
3.1.2. Vecteurs de clonage et d'expression	83 -
3.1.3. Antibiotiques et solutions pour milieux de culture	84 -
3.1.4. Tampons et milieux	84 -
3.1.4.1. Le milieu SOB	84 -
3.1.4.2. Le milieu TB	84 -
3.1.5. Electrophorèse en gel d'agarose	84 -
3.1.5.1. Tampon de dépôt	84 -
3.1.5.2. Tampon TBE 10X	85 -
3.1.5.3. Tampon TAE 50X	85 -
3.1.6. Préparation des bactéries chimio-compétentes et transformation	85 -
3.1.6.1. Méthode Inoue : bactéries chimio-compétentes DH5 α	85 -
3.1.6.2. Méthode au CaCl ₂ : bactérie chimio-compétente BL21	86 -
3.1.6.3. Transformation chimique de bactéries chimio-compétentes	86 -
3.1.7. Purification d'ADN plasmidique	87 -
3.1.7.1. Maxi-préparation	87 -
3.1.7.2. Mini-préparation	90 -

3.1.8. Stratégies de clonage	- 92 -
3.1.8.1. Le clonage Gateway®	- 92 -
3.1.8.1.1. Principe.....	- 92 -
3.1.8.1.2. La recombinaison	- 92 -
3.1.8.1.3. Les enzymes recombinantes	- 92 -
3.1.8.1.4. Les réactions de recombinaison	- 93 -
3.1.8.1.5. Protocole.....	- 94 -
3.1.8.2. Le clonage par enzymes de restriction	- 99 -
3.1.8.2.1. Digestion de l'ADN par des enzymes de restriction	- 99 -
3.1.8.2.2. Déphosphorylation des extrémités 5' phosphate du vecteur	- 99 -
3.1.8.2.3. Déprotéinisation et précipitation de l'ADN	- 100 -
3.1.8.2.4. Ligation	- 100 -
3.1.8.2.5. Transformation à partir du mélange de ligation	- 101 -
3.1.8.2.6. Obtention du vecteur de clonage et vérification de la séquence	- 101 -
3.1.9. Mutagenèse ponctuelle	- 101 -
3.2. Techniques biochimiques	- 103 -
3.2.1. Culture et production dans <i>E. coli</i>	- 103 -
3.2.1.1. Préculture	- 104 -
3.2.1.2. Culture	- 104 -
3.2.1.3. Récolte des bactéries	- 105 -
3.2.2. Lyse cellulaire	- 105 -
3.2.3. Ultracentrifugation	- 106 -
3.2.4. Chromatographies sur colonnes	- 107 -
3.2.4.1. Chromatographie d'affinité sur résine de chélation de cations métalliques.-	107 -
3.2.4.1.1. Colonnes HiTrap Chelating HP de 5ml.....	- 107 -
3.2.4.1.2. Colonne avec la résine Talon	- 109 -
3.2.4.2. Chromatographie d'exclusion moléculaire: Superdex 200 HR 16/60	- 111 -
3.2.5. Incubation avec l'acide rétinoïque	- 112 -
3.2.6. Protéolyse ménagée par la thrombine	- 112 -
3.3. Cristallisation	- 113 -
3.3.1. La plate-forme de cristallisation.....	- 113 -
3.3.2. Généralités sur la cristallisation	- 113 -
3.4. Caractérisation des échantillons.....	- 115 -
3.4.1. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes.....	- 115 -
3.4.1.1. Tampons et solutions.....	- 115 -
3.4.1.1.1. Gel de séparation à 12%	- 115 -
3.4.1.1.2. Pré-gel de concentration à 5%	- 115 -
3.4.1.1.3. Tampon dissociant et de charge concentré 5X.....	- 115 -
3.4.1.1.4. Tampon de migration 5X	- 116 -
3.4.1.1.5. Révélation au Bleu de Coomassie.....	- 116 -
3.4.2. Dosage des acides nucléiques et des protéines.....	- 116 -
3.4.3. La spectrométrie de masse	- 117 -
3.5. Techniques de caractérisation fonctionnelle.....	- 118 -
3.5.1. Expérience de retard sur gel	- 118 -
3.5.1.1. Principe.....	- 118 -
3.5.1.2. Marquage radioactif d'un oligonucléotide au [γ - ³² P] ATP	- 118 -
3.5.1.3. Transcription-Traduction in vitro	- 119 -
3.5.1.4. Retard sur gel	- 119 -
3.5.2. Test de transactivation en cellules HEK 293 EBNA.....	- 120 -
3.5.2.1. Culture	- 120 -

3.5.2.2. Transfection.....	- 120 -
3.5.2.2.1. Ensemencement en plaque de culture 24 puits.....	- 120 -
3.5.2.2.2. Protocole de transfection.....	- 121 -
3.5.2.2.3. Récolte des cellules.....	- 122 -
3.5.2.3. Test β -galactosidase.....	- 122 -
3.5.2.4. Test luciférase.....	- 123 -
Résultats.....	- 125 -
Chapitre 4 - Contexte bioinformatique.....	- 127 -
4.1. Analyse de protéines - Le rôle central des alignements multiples de séquences.....	- 127 -
4.1.1. Introduction.....	- 127 -
4.1.2. Définition de l'Alignement Multiple de Séquences Complètes.....	- 128 -
4.1.3. Applications du MACS.....	- 131 -
4.1.3.1. La phylogénie et la génomique comparative.....	- 131 -
4.1.3.2. Caractérisation fonctionnelle.....	- 133 -
4.1.3.3. Amélioration de la prédiction des structures secondaires et tertiaires.....	- 133 -
4.1.3.4. Prédiction des structures secondaires.....	- 134 -
4.1.3.5. Prédiction des structures tertiaires.....	- 135 -
4.1.3.6. Conservation des résidus.....	- 136 -
Chapitre 5 - Analyse du MACS de la super-famille des RNs.....	- 137 -
Chapitre 6 - Et les autres domaines dans tout ça ?.....	- 151 -
6.1. Le domaine A/B.....	- 151 -
6.1.1. Rappel.....	- 151 -
6.1.2. Le désordre.....	- 151 -
6.1.2.1. Caractérisation et fonction.....	- 151 -
6.1.2.2. KoAnno.....	- 154 -
6.1.2.2.1. Programmes de prédiction du désordre.....	- 155 -
6.1.2.2.1.1. RONN.....	- 155 -
6.1.2.2.1.2. GlobPlot2.....	- 155 -
6.1.2.2.1.3. DisEMBL1.4.....	- 156 -
6.1.2.2.1.4. FoldIndex.....	- 157 -
6.1.2.2.2. Le serveur web KoAnno.....	- 157 -
6.1.2.3. Applications.....	- 159 -
6.1.2.3.1. Les 48 RNs humains.....	- 159 -
6.1.2.3.2. Le RN des oestrogènes ER α	- 161 -
6.2. Le DBD.....	- 164 -
Chapitre 7 - Caractérisation des voies de communication.....	- 185 -
7.1. Genèse des mutants.....	- 187 -
7.2. Les différentes études menées.....	- 191 -
Chapitre 8 - Etude du système hERR γ	- 193 -
8.1. Les différents mutants testés.....	- 193 -
8.2. Etude fonctionnelle des mutants de hERR γ	- 194 -
8.3. Etude structurale des mutants hERR γ	- 194 -
8.3.1. Expression.....	- 194 -
8.3.2. Lyse cellulaire et ultracentrifugation.....	- 195 -
8.3.3. Purification des différents mutants.....	- 195 -
8.3.3.1. Le mutant E1a.....	- 196 -
8.3.3.1.1. Chromatographie d'affinité sur résine chélatant les ions nickel.....	- 197 -
8.3.3.1.2. Filtration sur gel.....	- 197 -
8.3.3.1.3. Protéolyse ménagée par la thrombine.....	- 197 -
8.3.3.1.4. Chromatographie d'affinité sur résine chélatant les ions nickel.....	- 198 -

8.3.3.1.5. Filtration sur gel	- 198 -
8.3.3.1.6. Rendement de purification	- 199 -
8.3.3.1.7. Pureté et homogénéité du mutant E1a.....	- 199 -
8.3.3.1.7.1. Gel natif.....	- 200 -
8.3.3.1.7.2. Spectrométrie de masse.....	- 200 -
8.3.3.2. Expression et purification du mutant E2a	- 202 -
8.3.3.3. Expression et purification des mutants E3a, E4a et E5	- 203 -
8.3.3.4. Bilan des purifications.....	- 204 -
8.3.4. Cristallisation du mutant E1a	- 205 -
8.3.5. Du cristal à la structure du mutant E1a	- 207 -
8.3.5.1. Collecte des données	- 207 -
8.3.5.2. Résolution de la structure et affinement.....	- 208 -
8.3.6. Etude de la structure tridimensionnelle du mutant E1a.....	- 211 -
8.3.6.1. Structure générale.....	- 211 -
8.3.6.2. Interface de dimérisation.....	- 212 -
8.3.6.3. Conformation de l'arginine 379	- 213 -
8.3.6.4. La poche de fixation du ligand	- 214 -
8.3.7. Bilan	- 218 -
Chapitre 9 - Etude du système hRAR α	- 221 -
9.1. Etude structurale	- 221 -
9.1.1. Expression de X0 et du mutant R1d.....	- 221 -
9.1.2. Lyse cellulaire et ultracentrifugation.....	- 222 -
9.1.3. Purification du complexe R1d-X0.....	- 223 -
9.1.3.1. Chromatographie d'affinité sur résine chélatant les ions nickel	- 224 -
9.1.3.2. Chromatographie de filtration sur gel	- 225 -
9.1.3.3. Rendement de purification	- 225 -
9.1.4. Cristallisation de l'hétérodimère R1d-X0	- 226 -
9.2. Etude de l'activité transcriptionnelle et de la dimérisation.....	- 226 -
9.3. Etude indirecte de la voie par phosphorylation	- 227 -
Chapitre 10 - Etude du système hXRRA	- 239 -
10.1. Les mutants disponibles	- 239 -
10.2. Etude structurale du mutant X1c	- 240 -
10.2.1. Expression du mutant X1c	- 240 -
10.2.2. Lyse cellulaire et ultracentrifugation.....	- 241 -
10.2.3. Purification du mutant X1c en présence de 9- <i>cis</i> RA	- 241 -
10.2.3.1. Chromatographie d'affinité sur résine chélatant les ions cobalt	- 243 -
10.2.3.2. Filtration sur gel	- 243 -
10.2.3.3. Rendement de purification	- 244 -
10.2.4. Cristallisation du mutant X1c.....	- 244 -
10.3. Etude fonctionnelle	- 245 -
10.3.1. Importance du pont connectant H8 à la région H9-10 pour l'oligomérisation	- 248 -
10.3.2. La LBP et le pont connectant H1 à H8 sont essentiels pour l'activité transcriptionnelle de hXRRA.....	- 249 -
Discussion - Perspectives	- 251 -
Annexes	- 265 -
Références Bibliographiques.....	- 275 -

Tables des Figures

Figure 1 Les deux voies majeures de transduction du signal	- 5 -
Figure 2 Organisation structurale et fonctionnelle de la super-famille des RNs	- 8 -
Figure 3 Illustration schématique du DBD de hRXR α	- 14 -
Figure 4 Types d'éléments de réponse des RNs.....	- 16 -
Figure 5 Vue stéréoscopique de différents modes d'interaction DBD-ADN.....	- 18 -
Figure 6 Structure du LBD du RN à la Vitamine D	- 21 -
Figure 7 Structures cristallographiques vues selon l'axe de symétrie d'ordre 2.....	- 22 -
Figure 8 Ligands et RNs associés	- 24 -
Figure 9 Effets de la fixation du ligand sur la conformation du LBD de hRXR α	- 26 -
Figure 10 Echelle des temps géologiques et biologiques.....	- 28 -
Figure 11 Arbre phylogénétique simplifié des eucaryotes.....	- 32 -
Figure 12 Morphologie de la larve d'un Urochordé.....	- 35 -
Figure 13 Anatomie de l'amphioxus.....	- 35 -
Figure 14 Les six sous-familles de la super-famille des RNs.....	- 37 -
Figure 15 Schéma illustrant l'acquisition de la capacité de fixation du ligand	- 41 -
Figure 16 Surfaces de reconnaissance des CoR et des CoA.....	- 48 -
Figure 17 La « pince électrostatique ».....	- 49 -
Figure 18 Mécanisme séquentiel de l'action des RNs	- 51 -
Figure 19 Effet des ligands agonistes et antagonistes	- 53 -
Figure 20 Les différentes étapes de « chaperonnage »	- 57 -
Figure 21 Transport du récepteur stéroïdien vers le noyau.....	- 58 -
Figure 22 Interface graphique de GScope.....	- 65 -
Figure 23 Evolution du nombre d'entrées de la banque Swiss-Prot.....	- 67 -
Figure 24 Evolution du nombre d'entrées de la banque TrEMBL.....	- 68 -
Figure 25 Evolution du nombre d'entrées de la banque PDB.....	- 69 -
Figure 26 Capture d'écran du serveur web SRS de l'IGBMC.....	- 71 -
Figure 27 Aperçu de la cascade de programmes constituant PipeAlign.....	- 73 -
Figure 28 Capture d'écran du serveur web de PipeAlign.....	- 76 -
Figure 29 Schéma de représentation des conservations adoptée dans OrdAlie.....	- 78 -
Figure 30 Capture d'écran du programme SeqLab.....	- 79 -
Figure 31 Protocole d'une maxi-préparation d'ADN plasmidique.....	- 89 -
Figure 32 Protocole d'une mini-préparation d'ADN plasmidique.....	- 91 -
Figure 33 Réactions catalysées par les enzymes recombinantes.....	- 93 -
Figure 34 Réaction BP	- 93 -
Figure 35 Réaction LR.....	- 93 -
Figure 36 Schéma expliquant la technique de mutagenèse dirigée utilisée.....	- 102 -
Figure 37 Système d'expression utilisé dans <i>E. coli</i>	- 104 -
Figure 38 Robots disponibles sur la plate-forme de cristallisation.....	- 113 -
Figure 39 Diagramme des phases	- 114 -
Figure 40 Principe de fonctionnement du spectromètre électrospray	- 117 -
Figure 41 Protocole de transfection.....	- 122 -
Figure 42 Diagramme de Venn de la classification des acides aminés.....	- 129 -
Figure 43 Vue schématique d'un MACS.....	- 130 -
Figure 44 Applications des alignements multiples de séquences protéiques.....	- 131 -
Figure 45 Les différentes relations d'homologie.....	- 132 -

Figure 46 MACS des RNs avec les différents domaines	138 -
Figure 47 Exemple de transition désordre - ordre	153 -
Figure 48 Echelle de Russel/Linding et Deleage/Roux.....	156 -
Figure 49 Echelle de propension des résidus	157 -
Figure 50 Interface de KoAnno.....	158 -
Figure 51 Capture d'écran présentant les résultats de prédiction du désordre	158 -
Figure 52 Bilan des prédictions du désordre des différents programmes	160 -
Figure 53 Bilan des prédictions du désordre des différents programmes	162 -
Figure 54 Région prédite ordonnée dans le domaine A/B de ER α	163 -
Figure 55 Les trois RNs tests.....	186 -
Figure 56 Protocole de clonage et de mutagenèse	188 -
Figure 57 Les différents mutants hERR γ	193 -
Figure 58 Expression des mutants E1a, E2a, E3a, E4a et E5.	195 -
Figure 59 Schéma de purification du mutant E1a.....	196 -
Figure 60 Gel natif (8-25%) de hERR γ	200 -
Figure 61 Spectre de masse en conditions dénaturantes du mutant E1a.	201 -
Figure 62 Spectre de masse en mode natif effectué sur le mutant E1a.	202 -
Figure 63 Purification des mutants E3a, E4a et E5 par filtration sur gel	204 -
Figure 64 Bilan de purification des différents mutants.....	204 -
Figure 65 Cristaux du mutant E1a en complexe avec le peptide SRC-1.....	206 -
Figure 66 Diagramme de Ramachandran du mutant E1a	210 -
Figure 67 Caractéristique du mutant E1a.....	211 -
Figure 68 Structure de l'homodimère cristallographique du mutant E1a.....	211 -
Figure 69 Superposition des structures cristallographiques des homodimères.....	212 -
Figure 70 Stéréovue de la conformation de la boucle H8-H9.....	213 -
Figure 71 La poche de fixation du ligand du mutant E1a.	214 -
Figure 72 Stéréovue de la densité électronique observée dans la LBP	215 -
Figure 73 Structure cristallographique du complexe ER α /œstradiol.....	216 -
Figure 74 Stéréovue de la superposition de l'œstradiol E2 de hER α	217 -
Figure 75 Exemples de ligands	218 -
Figure 76 Bilan de caractérisation des différents mutants de hERR γ	218 -
Figure 77 Comparaison de R0 et du mutant R1d.....	221 -
Figure 78 Schéma de purification de l'hétérodimère R1d-X0	224 -
Figure 79 Bilan des expériences de <i>pull down</i> et de transactivation	227 -
Figure 80 Site de phosphorylation (S369) de la PKA chez hRAR α	229 -
Figure 81 Les différents mutants hXR α	239 -
Figure 82 Schéma de purification du mutant X1c en présence de 9- <i>cis</i> RA	242 -
Figure 83 Cristal du mutant X1c.....	244 -
Figure 84 Vérification de la transcription-traduction <i>in vitro</i>	246 -
Figure 85 Les mutants exploitables des voies de communication.....	259 -

Table des Tableaux

Tableau 1 Les 48 gènes des récepteurs nucléaires humains	- 7 -
Tableau 2 Structures de domaines de liaison à l'ADN référencées dans la PDB	- 13 -
Tableau 3 Séquences des oligonucléotides GWS et GWR.	- 97 -
Tableau 4 Séquences des oligonucléotides de séquençage A1 et A2.	- 98 -
Tableau 5 Séquences des oligonucléotides UP1 et GWR.....	- 99 -
Tableau 6 Composition des différents tampon de lyse.....	- 106 -
Tableau 7 Composition des tampons A et B pour la purification des mutants E1a à E5-	108 -
Tableau 8 Composition des tampons A et B pour la purification du mutant R1d.....	- 109 -
Tableau 9 Composition des tampons A et B pour la purification du mutant X1c.	- 110 -
Tableau 10 Composition des différents tampons de filtration sur gel.....	- 111 -
Tableau 11 Liste des mutants pour l'étude structurale.....	- 192 -
Tableau 12 Rendement de purification du mutant E1a	- 199 -
Tableau 13 Rendement de purification du mutant E2a	- 202 -
Tableau 14 Statistiques de collecte des données du cristal du mutant E1a.....	- 208 -
Tableau 15 Statistiques d'affinement.....	- 209 -
Tableau 16 Tableau schématique présentant les différentes combinaisons de purification du mutant R1d.....	- 222 -
Tableau 17 Rendement de purification du complexe R1d-X0.....	- 225 -
Tableau 18 Tableau schématique présentant les différentes combinaisons de purification jusqu'à l'obtention de cristaux du mutant X1c.....	- 241 -
Tableau 19 Rendement de purification du mutant X1c	- 244 -
Tableau 20 Séquences des éléments de réponse DR1 et DR5.....	- 246 -
Tableau 21 Bilan des expériences de gel retard et de transactivation pour les différents mutants de hRXR α	- 247 -

Abréviations

9- <i>cis</i> RA	acide rétinoïque 9- <i>cis</i>
Å	Angström
ADN	Acide DésoxyriboNucléique
AF	<i>Activation Function</i>
BIPS	Plate-forme de Bioinformatique de Strasbourg
CBP	<i>CREB Binding Protein</i>
CoA	coactivateur
CoR	corépresseur
CREB	<i>cAMP-Response Element Binding</i>
CTE	<i>Carboxy Terminal Extension</i>
C-ter	C-terminale
Da	Dalton
DBD	<i>DNA Binding Domain</i>
DPC	<i>Density of Points Clustering</i>
DR	<i>Direct Repeat</i>
EMBOSS	<i>European Molecular Biology Open Software Suite</i>
EMSA	<i>ElectroMobility Shift Assay</i>
ERR	<i>Estrogen Related Receptor</i>
ESI-MS	<i>Electron Spray Ionisation-Mass Spectrometer</i>
FPLC	<i>Fast Performance Liquid Chromatography</i>
GCG	<i>Genetics Computer Group</i>
GF	filtration sur el
HAT	<i>Histone Acétyl Transferase</i>
HDAC	<i>Histone DeACetyl-transferase</i>
HMG	<i>High Mobility Group</i>
HMT	<i>Histone Methyl Transferase</i>
HRE	<i>Hormone Response Element</i>
HSP	<i>Heat Shock Protein</i>
IGBMC	Institut de Génétique et de Biologie Moléculaire et Cellulaire
IMAC	<i>Immobilized Metal Affinity Chromatography</i>
IPTG	IsoPropyl β -D-1-ThioGalactopyranoside
IR	<i>Inverted Repeat</i>
LBD	<i>Ligand Binding Domain</i>
LBGS	Laboratoire de Biologie et de Génomique Structurales
LBP	<i>Ligand Binding Pocket</i>
LEON	<i>multiple aLignment Evaluation Of Neighbours</i>
LMS	<i>Local Maximum Segment</i>
MACS	<i>Multiple Alignment of Complete Sequences</i>
MACSIMS	<i>Multiple Alignment of Complete Sequences Information Management System</i>
MAPK	<i>Mitogen-Activated Protein Kinase</i>
MBP	<i>Maltose Binding Protein</i>
mg	milligramme
min	minute
ml	millilitre
mM	millimolaire

ng	nanogramme
N-CoR	<i>Nuclear CoRepressor</i>
NLS	<i>Nuclear Localization Signal</i>
nm	nanomètre
NMR	<i>Nuclear Magnetic Resonance</i>
NorMD	<i>Normalized Mean Distance</i>
N-ter	N-terminale
NR-box	<i>Nuclear Receptor-box</i>
OrdAlie	<i>Ordered Alignment information explore</i>
PDB	<i>Protein Data Bank</i>
PCR	<i>Polymerase Chain Reaction</i>
PKA	<i>Protein Kinase A</i>
PLB	<i>Passiv Lysis Buffer</i>
pmol	picomole
RAR	<i>Retinoid acid Receptor</i>
RASCAL	<i>RApid Scanning and Correction of ALignment errors</i>
RCA	<i>Rolling Circle Amplification</i>
RM	Récepteur Membranaire
RN	Récepteur Nucléaire
RXR	<i>Retinoid X Receptor</i>
sec	seconde
SMRT	<i>Silencing Mediator for Retinoic acid and Thyroid hormone receptors</i>
SRS	<i>Sequence Retrieval System</i>
SUMO	<i>Small Ubiquitine-related MOdifier</i>
TBP	<i>TATAbox Binding Protein</i>
USP	<i>ultraspiracle</i>
µl	microlitre
U	Unité
vol	volume

Introduction

Unité fondamentale du vivant, la cellule est étroitement dépendante de son milieu. La quantité de nourriture, le niveau d'oxygène, le niveau de lumière ou encore la présence d'autres cellules ou organismes sont autant d'informations nécessaires à sa survie. Cette dépendance a amené les cellules à développer des systèmes de détection du signal et de communication remarquables, tant par leur diversité que par leur efficacité. Tous ces systèmes sont organisés selon un schéma similaire : un récepteur cellulaire détecte un signal ; il transmet l'information à des effecteurs ; ces derniers permettent alors la propagation et l'intégration du signal ; l'intégration du signal aboutit au stockage de l'information à plus ou moins long terme, ainsi qu'à l'adaptation du comportement cellulaire, fréquemment au travers de la transcription de programmes génétiques adaptés. Chez l'homme, plus d'un cinquième des gènes est directement ou indirectement impliqué dans de tels mécanismes de communication (Copley *et al.*, 2002).

Quels sont les mécanismes biologiques permettant à la cellule de relever un tel défi ? Comment la transmission exacte d'une information d'un type donné est-elle assurée, au même titre que des centaines d'autres différentes, et ce de manière simultanée ? Par quelle « magie » la cellule parvient-elle à atteindre un niveau de précision adapté dans un environnement aussi complexe et variable ?

L'une des réponses développée au cours des quelques trois milliards et demi d'années de développement et d'évolution du vivant qui nous précèdent, réside dans la régulation de l'association des molécules assurant la communication cellulaire (Bray and Duke, 2004). La notion de communication cellulaire revêt l'idée d'un transfert d'information. Le rôle de chacune des molécules participant d'une cascade de signalisation va donc être de conduire cette information.

Au niveau moléculaire, ce transfert d'information a été particulièrement étudié dans le cadre de l'allostérie. Ce phénomène se caractérise notamment par la propagation de réarrangements correspondant à des changements conformationnels locaux ou plus vastes de la protéine (Luque *et al.*, 2002). Cette cascade de changements conformationnels d'une région d'une protéine vers une autre définit la notion d'allostérie (Kern and Zuiderweg, 2003). Au niveau d'un récepteur, la reconnaissance et la liaison d'un ligand ne reposent généralement que sur un nombre restreint de résidus distribués autour de la zone de fixation

du ligand (Suel *et al.*, 2003). De même, la propagation du signal de liaison vers une nouvelle région du récepteur repose sur une chaîne d'interactions coopératives impliquant quelques résidus souvent distribués de façon hétérogène au sein des protéines et qui participent au phénomène d'allostérie fonctionnelle. En déterminant le squelette nécessaire à une allostérie fonctionnelle, ces résidus sont souvent impliqués dans le cœur architectural de la protéine et dans sa stabilité. L'importance de la stabilité structurale des protéines dans le phénomène d'allostérie pourrait ainsi expliquer le lien étroit existant entre propagation de changements conformationnels, oligomérisation et signalisation (Bray and Duke, 2004; Goodsell and Olson, 2000; Monod *et al.*, 1965; Sourjik, 2004).

Pour faire face à la multiplicité des signaux extérieurs, différents systèmes de reconnaissance et de transmission primaire du signal existent que l'on peut réunir autour de deux grands types qui aboutissent fréquemment à des facteurs de transcription régulant des programmes génétiques spécifiques : les Récepteurs Membranaires (RMs) et les Récepteurs Nucléaires (RNs) (Figure 1).

Dans le cas des RMs, médiateurs d'un message de prolifération cellulaire, la « molécule signal » ne pénètre pas directement dans la cellule. Il peut s'agir d'un petit ligand (histamine, adrénaline), d'un peptide (vasopressine, glucagon) ou d'une protéine (hormone de croissance). Après activation par leur ligand, les RMs transmettent un signal à l'intérieur de la cellule vers des facteurs de transcription nucléaires, comme NF- κ B et/ou le complexe AP-1, par le biais de messagers secondaires tels que l'AMPc, le GMPc, les phosphoinositides, le calcium et/ou *via* une cascade de phosphorylations faisant appel à des protéines kinases. Ces phosphorylations aboutiront *in fine* à la modification d'un facteur de transcription et à la régulation de la transcription des gènes placés sous son contrôle.

La signalisation par les RNs est une des voies majeures de transduction du signal inventée par les Métazoaires pour réguler directement la transcription des gènes. La super-famille des RNs, médiateurs de la signalisation intracellulaire, dont l'action se situe comme leur nom l'indique dans le noyau, se comportent eux-mêmes comme des facteurs transcriptionnels. La plupart des RNs sont capables de lier spécifiquement des petites molécules hydrophobes (hormones stéroïdiennes, acide rétinoïque, vitamine D3, hormones thyroïdiennes, acides gras, prostaglandines), qui pénètrent ou sont libérées au sein de la

cellule cible. Ces ligands jouent le rôle de signaux régulateurs qui modifient l'activité transcriptionnelle des RNs.

A la différence des autres facteurs de transcription de la voie des RMs, les RNs sont schématiquement capables d'assumer toutes les fonctions assurées par les divers protagonistes impliqués dans la voie des RMs, et ce en une seule entité protéique : se fixer à l'ADN, fixer un ligand, posséder plusieurs domaines activateurs de la transcription et réguler la transcription des gènes cibles au travers de l'action coordonnée des RNs et d'une variété de cofacteurs (coactivateurs et corépresseurs) présentant des activités enzymatiques variées (phosphorylation, etc.). Pour permettre l'intégration de ces signaux extracellulaires, ces différents facteurs de transcription interagissent entre eux et modulent leurs activités transcriptionnelles réciproques.

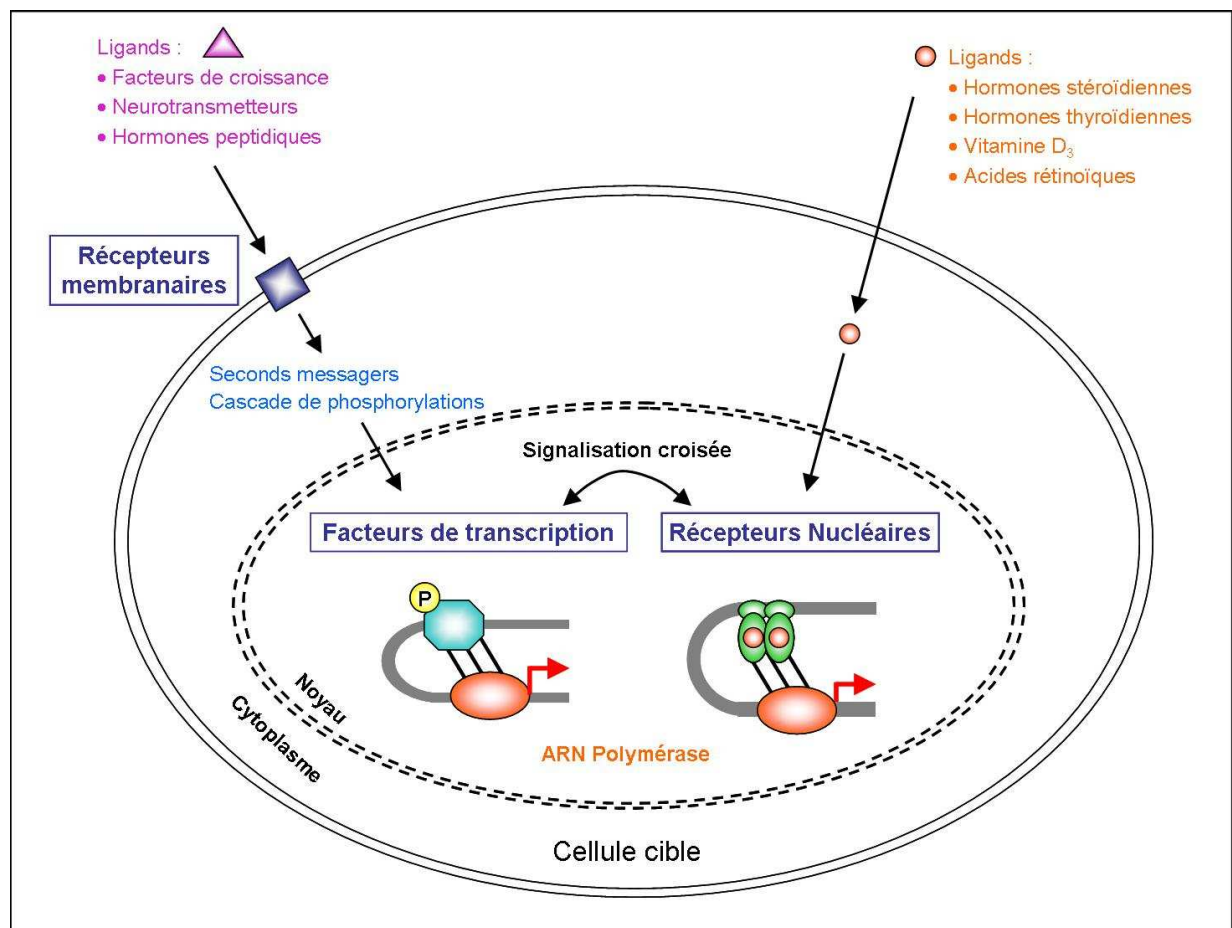


Figure 1 | Les deux voies majeures de transduction du signal impliquent les récepteurs membranaires et les récepteurs nucléaires. Les deux voies de signalisation peuvent interférer au niveau de la régulation transcriptionnelle de gènes cibles (signalisation croisée), ce qui permet d'établir un réseau fonctionnel d'intégration de l'information.

Chapitre 1 - La super-famille des récepteurs nucléaires

1.1. Fonctions biologiques et importance thérapeutique

Les RNs forment une super-famille spécifique de facteurs de transcription présents chez tous les Métazoaires. Cette super-famille comprend 48 gènes identifiés dans le génome humain (Tableau 1), 21 dans celui de *Drosophila melanogaster* et plus de 250 chez *Caenorhabditis elegans* (Robinson-Rechavi *et al.*, 2001). En régulant la transcription de nombreux gènes cibles, les RNs sont impliqués dans une multitude d'évènements physiologiques tels que l'embryogenèse, l'homéostasie, la croissance cellulaire, la reproduction, le développement, le métabolisme, la différenciation ou encore l'apoptose (Ribeiro *et al.*, 1995).

Les RNs ont la capacité de se fixer à l'ADN et la plupart d'entre eux sont capables de lier spécifiquement des petites molécules hydrophobes (hormones stéroïdiennes, acide rétinoïque, vitamine D3, hormones thyroïdiennes, acides gras, prostaglandines), qui pénètrent ou sont libérées au sein de la cellule cible. Ces ligands jouent le rôle de signaux régulateurs qui modifient l'activité transcriptionnelle des RNs. Bien que pour plusieurs de ces récepteurs une interaction à un ligand soit nécessaire à leur activation, il existe un ensemble de récepteurs, définis comme orphelins, pour lesquels aucun ligand n'a été identifié. L'identification de ligands naturels ou synthétiques a permis d'«adopter» certains de ces orphelins, comme c'est le cas pour RXR (*Retinoid X receptor*) et PPAR (*Peroxisome proliferator activated receptor*). Cependant, certains RNs comme Nurr1 (*Nur-related factor 1*) (Wang *et al.*, 2003), DHR38 (*Drosophila Hormone Receptor 38*) (Baker *et al.*, 2003), COUP-TF (*Chicken Ovalbumin Upstream Protein Transcription Factor*) et NGFI-B (*Nerve Growth Factor Induce*) semblent être de véritables orphelins (Mangelsdorf and Evans, 1995). Ces RNs orphelins se retrouvent dans chacune des six grandes sous-familles de RNs (Laudet, 1997) qui seront présentées dans la section 1.3.3.

Par cette propriété de fixer une hormone modifiant leur activité transcriptionnelle, les RNs constituent d'importantes cibles pharmacologiques, associées à des pathologies majeures (cancer, diabète, obésité, syndrome de résistance hormonale, etc.).

Nom		Isotype	Nomenclature	Ligand
TR	<i>Thyroid hormone receptor</i>	a	NR1A1	Hormone thyroïdienne (T3)
		β	NR1A2	Hormone thyroïdienne (T3)
RAR	<i>Retinoic acid receptor</i>	a	NR1B1	Acide rétinoïque
		β	NR1B2	Acide rétinoïque
		?	NR1B3	Acide rétinoïque
PPAR	<i>Peroxisome proliferator activated receptor</i>	a	NR1C1	Acide gras, leukotriène B4
		β	NR1C2	Acides gras
		?	NR1C3	Acides gras, prostaglandine J2
Rev-erb	<i>Reverse ErbA</i>	a	NR1D1	Orphelin
		β	NR1D2	Orphelin
ROR	<i>RAR-related orphan receptor</i>	a	NR1F1	Orphelin
		β	NR1F2	Orphelin
		?	NR1F3	Orphelin
LXR	<i>Liver X receptor</i>	a	NR1H3	Oxysterol
		β	NR1H2	Oxysterol
FXR	<i>Farnesoid receptor</i>	a	NR1H4	Acides biliaires, feraxamine
VDR	<i>Vitamin D receptor</i>		NR1I1	Vitamine D
PXR	<i>Pregnane X receptor</i>		NR1I2	Xenobiotiques
CAR	<i>Constitutive androstan receptor</i>		NR1I3	Xenobiotiques, phenobarbital
HNF4	<i>Human nuclear factor 4</i>	a	NR2A1	Orphelin
		β	NR2A3	Orphelin
		?	NR2A2	Orphelin
RXR	<i>Retinoid X receptor</i>	a	NR2B1	Acide rétinoïque 9-cis
		β	NR2B2	Acide rétinoïque 9-cis
		?	NR2B3	Acide rétinoïque 9-cis
TR2	<i>Testis receptor</i>		NR2C1	Orphelin
TR4			NR2C2	Orphelin
TLL	<i>Talless</i>		NR2E2	Orphelin
PNR	<i>Photoreceptor-specific nuclear receptor</i>		NR2E3	Orphelin
COUP-TF	<i>Chicken ovalbumin upstream promoter-transcription factor</i>	I	NR2F1	Orphelin
		II	NR2F2	Orphelin
EAR2	<i>ErbA2-related gene-2</i>		NR2F6	Orphelin
ER	<i>Estrogen receptor</i>	a	NR3A1	Oestradiol-17 β, tamoxifène, raloxifène
		β	NR3A2	Oestradiol-17 β, composés synthétiques
ERR	<i>Estrogen receptor-related</i>	a	NR3B1	Orphelin
		β	NR3B2	Orphelin
		?	NR3B3	Orphelin
GR	<i>Glucocorticoid receptor</i>		NR3C1	Cortisol, dexaméthasone, RU486
MR	<i>Mineralocorticoid receptor</i>		NR3C2	Aldosterone, spiro lactone
PR	<i>Progesterone receptor</i>		NR3C3	Progesterone, médroxyprogesterone acétate, RU486
AR	<i>Androgen receptor</i>		NR3C4	Testostérone, flutamide
NGFI-B	<i>NGF-induced factor B</i>		NR4A1	Orphelin
NURR1	<i>Nur-related factor 1</i>		NR4A2	Orphelin
NOR1	<i>Neuron-derived orphan receptor 1</i>		NR4A3	Orphelin
SF1	<i>Steroidogenic factor 1</i>		NR5A1	Orphelin
LRH1	<i>Liver receptor homologous protein 1</i>		NR5A2	Orphelin
GCNF	<i>Germ cell nuclear factor</i>		NR6A1	Orphelin
DAX-1	<i>DSS-AHC critical region on the chromosom, gene 1</i>		NR0B1	Orphelin
SHP	<i>Short heterodimeric partner</i>		NR0B2	Orphelin

Tableau 1 | Les 48 gènes des récepteurs nucléaires humains. Le code couleur correspond à celui six grandes sous-familles de RNs présentées dans la section 1.3.3.

L'activation ou la non-activation des RNs représente l'un des mécanismes central d'action de nombreuses molécules actives actuelles (œstrogènes, glucocorticoïdes, androgènes, stéroïdes, PPARs). Plus de 10% des 100 médicaments les plus vendus dans le monde ciblent ces RNs (la pilule contraceptive, la majorité des anti-inflammatoires, de nombreux médicaments contre le diabète, etc.).

1.2. Organisation modulaire des récepteurs nucléaires

Les RNs sont organisés selon une architecture modulaire commune en 5 ou 6 domaines (Krust *et al.*, 1986) représentée sur la Figure 2.

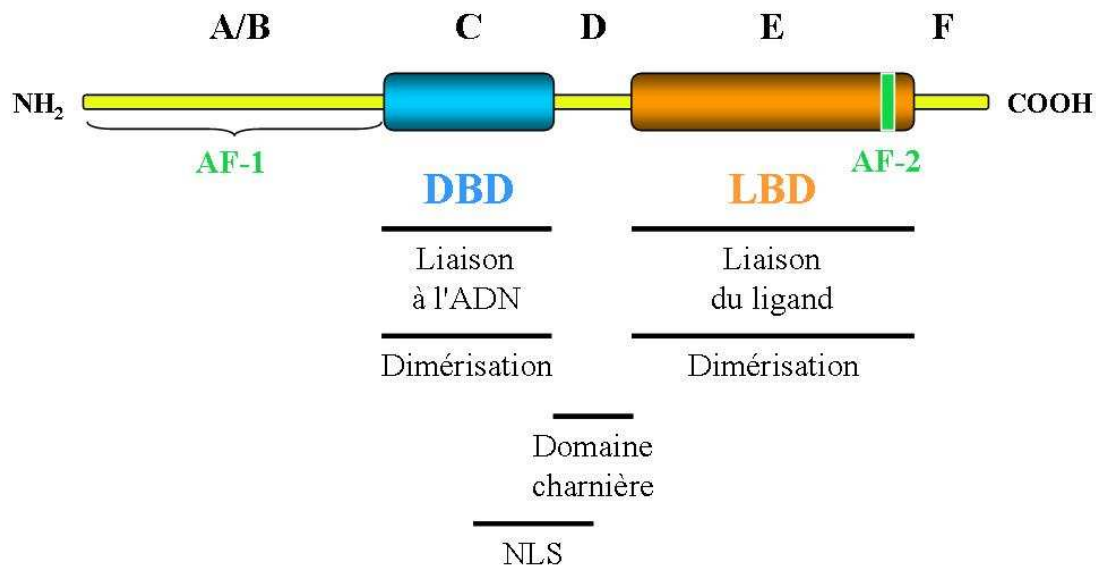


Figure 2 | Organisation structurale et fonctionnelle de la super-famille des récepteurs nucléaires. Les différents domaines (A/B, C, D, E et F) sont représentés des extrémités N- à C-terminales (N-ter et C-ter). Les régions C (DBD en bleu) et E (LBD en orange), respectivement impliquées dans la liaison à l'ADN et celle du ligand, sont conservées dans la super-famille. Les régions A/B, D et F, représentées en jaune, sont variables en séquence et en taille. La plupart des RNs possèdent deux fonctions de transactivation : AF-1 dans la région N-ter qui est ligand-indépendante ; AF-2, dans la partie C-ter du domaine E, qui est régulée par le ligand.

Le domaine amino-terminal A/B, variable en séquence et en longueur, contient une fonction d'activation de la transcription indépendante de la fixation du ligand AF-1 (*Activation Function-1*). Suit le domaine C de liaison à l'ADN (DBD, *DNA Binding Domain*), très conservé. Ce domaine comprend deux structures en doigt de zinc établissant des contacts avec des

séquences nucléotidiques spécifiques appelées éléments de réponse. Le DBD est, de plus, impliqué dans la dimérisation. Le domaine D est une région charnière reliant le DBD au domaine E de liaison du ligand (LBD, *Ligand Binding Domain*). Le LBD est lui aussi impliqué dans la dimérisation et contient une fonction d'activation de la transcription dépendante de la fixation du ligand AF-2 (*Activation Function-2*). Certains RNs possèdent un domaine F additionnel à l'extrémité C-ter.

1.2.1. Le domaine A/B

1.2.1.1. Structure et fonction

Le domaine A/B présente fréquemment une composition en acides aminés très biaisée (richesse en résidus de type sérine, thréonine, proline, asparagine, acide glutamique, glutamine, etc.) et représente une des, si ce n'est la région la moins conservée des RNs, à la fois en séquence et en taille, de 23 résidus pour le RN à la vitamine D (VDR) à 602 résidus pour le RN aux minéralocorticoïdes (MR). Historiquement, Krust et coll. (Krust *et al.*, 1986) se sont basés sur l'analyse des récepteurs aux œstrogènes de l'homme (hER) et du poulet (cER) pour définir les domaines A et B. Le domaine A correspond aux 37 premiers résidus de hER et le domaine B aux 142 résidus suivants, jusqu'au début du DBD. Cependant, la distinction des domaines A et B est généralement impossible.

Le domaine A/B, qui est une cible connue de phosphorylations, contient une fonction de transactivation indépendante du ligand, dont la capacité d'activation varie d'un RN à l'autre (Bastien *et al.*, 2000; Chen *et al.*, 2000; Warnmark *et al.*, 2003). Alors que la phosphorylation de ER α par TFIIH est ligand-dépendante (Chen *et al.*, 2000), celle de ER β par la protéine-kinase MAPK (*Mitogen-Activated Protein Kinase*) est ligand-indépendante et permet le recrutement du coactivateur (CoA) SRC-1 (Tremblay *et al.*, 1999). Ce dernier cas illustre bien la communication croisée entre la voie des RNs et celle des RMs (par le biais de MAPK).

La plupart des isoformes des RNs se distinguent dans cette région car elles sont issues soit d'un épissage alternatif, soit de l'utilisation de promoteurs différents (Giguere *et al.*, 1994).

Aucune donnée cristallographique concernant le domaine A/B d'un RN n'est disponible à ce jour. En l'absence de partenaire, il semblerait que le domaine A/B, dont AF-1, existe sous la forme d'un état désordonné, donc peu ou pas structuré (Dahlman-Wright and McEwan, 1996; McEwan *et al.*, 2007; Warnmark *et al.*, 2001). Cette propriété est vraie pour un grand

nombre de domaines d'activation des facteurs de transcription eucaryotiques qui présentent une faible propension intrinsèque à former des structures secondaires, comme c'est le cas, par exemple, pour NF- κ B (Schmitz *et al.*, 1994) ou p53 (Kussie *et al.*, 1996).

L'utilisation de la résonance magnétique nucléaire (NMR, *Nuclear Magnetic Resonance*) a montré que la région AF-1 de GR et les régions N-ter des récepteurs ER α et β sont désordonnées en solution. Il en est de même pour PR, d'après des études de protéolyse ménagée de la région N-ter seule. Cependant, en présence du DBD, cette région se structurerait, suggérant des interactions intramoléculaires ou des effets allostériques dans la protéine (Bain *et al.*, 2000; Bain *et al.*, 2001).

Un des dogmes centraux de la biochimie est que les protéines, pour fonctionner, doivent être structurées. La formation de structures secondaires serait donc cruciale à l'activation par la fonction AF-1. Il a été mis en évidence que la formation d'une hélice α est une étape importante dans l'activation des gènes par l'AF-1 (Dahlman-Wright and McEwan, 1996). Dans le cas de GR, des mutations ponctuelles qui agrandissent la région hydrophobe et permettent la formation de l'hélice α ont pour conséquence une augmentation de l'activité transcriptionnelle par rapport à la protéine sauvage. Dans leur ensemble, ces études convergent vers un modèle dans lequel le domaine AF-1 consiste en plusieurs modules pouvant former des conformations de type hélice α (Kumar and Thompson, 2003).

1.2.1.2. Interaction du domaine A/B avec des coactivateurs

Différents CoA, sans proximité de séquence ou de structure, interagissent avec la fonction AF-1 des RNs, mais les bases structurales de ces interactions ne sont pas encore élucidées.

Comme nous l'avons évoqué précédemment, tous les domaines A/B étudiés semblent être structurellement désordonnés, mais la formation de structures secondaires est un pré-requis pour l'activation par AF-1. Ainsi, un changement conformationnel du domaine A/B pourrait être induit par la liaison d'un partenaire. Ce phénomène a été observé pour ER α , où la formation d'un complexe entre le domaine N-ter seul et la TBP (*TATAbox Binding Protein*) s'accompagne d'un changement conformationnel (Warnmark *et al.*, 2001). L'explication la plus probable de cet effet est que la TBP, très structurée, stabilise ou induit une forme ordonnée correctement repliée du domaine N-ter de ce récepteur. L'étude de cette

interaction ER α /TBP a suggéré un mécanisme en deux étapes : une phase initiale d'association rapide entre le domaine N-ter désordonné de ER α et la surface de la TBP, suivie d'une phase de liaison lente durant laquelle une adaptation conformationnelle pourrait survenir.

1.2.1.3. Modifications post-traductionnelles

Dans ce contexte, il est particulièrement intéressant de noter que les domaines A/B, et par là même la fonction AF-1, des RNs sont les cibles de choix de nombreuses phosphorylations régulatrices. Ces modifications peuvent moduler le recrutement de CoA par l'AF-1 et permettre parfois l'activation des récepteurs en absence de ligand. Ainsi, les sérines 104, 106, 118 et 167 sont les principaux sites de phosphorylation de ER α . En particulier, la sérine 118 est une cible pour les kinases activées par les signaux mitogènes (MAPK, *Mitogen Activated Protein Kinases*). La résultante des mutations sérine en alanine de ces résidus est une diminution générale de l'activité transcriptionnelle de ER α (Glaros *et al.*, 2006). Il faut également souligner que cette sérine 118 peut également être phosphorylée par la cdk7 (*cyclin-dependent kinase 7*), une kinase appartenant au facteur général de transcription TFIIH (Chen *et al.*, 2000; Krstic *et al.*, 1997; Trowbridge *et al.*, 1997). De manière tout à fait intéressante, la kinase cdk7 phosphoryle également la région AF-1 de RAR α et RAR γ sur des sérines conservées (S77, S79 et S68 respectivement pour RAR α 1, RAR γ 1 et RAR γ 2 (Bastien *et al.*, 2000; Rochette-Egly *et al.*, 1997)).

Il est probable que ces phosphorylations influencent le repliement du domaine AF-1 ou du récepteur entier en stabilisant le complexe CoA/AF-1 par la création de nouveaux points d'interactions spécifiques entre le récepteur et le CoA.

Ainsi, la phosphorylation représente un niveau très important de la régulation de l'activité AF-1, et donc de l'activité globale des RNs. Leurs conséquences ne sont pas encore élucidées mais peuvent aller jusqu'à entraîner l'activation du RN en absence de ligand, illustrant la puissance de telles modifications régulatrices.

1.2.2. Le domaine C

1.2.2.1. Structure et fonction

Ce domaine (70 à 80 acides aminés, ~10kDa) confère aux RNs les propriétés de reconnaissance spécifique et de fixation d'une séquence d'ADN cible. Ce domaine, plus communément appelé DBD pour *DNA Binding Domain*, est le plus conservé dans la super-famille des RNs, sur une région d'environ 66 résidus. A ce jour, de nombreuses structures de DBDs de différents RNs, complexés ou non à l'ADN, obtenues par cristallographie aux rayons X ou RMN, sont disponibles (Tableau 2).

Le DBD est principalement composé de 2 motifs en doigts de zinc de type C2-C2 (classe II). Le premier [CI] présente un motif du type $CX_2CX_{13}CX_2C$ et est situé en N-ter de [CII], du type $CX_5CX_9CX_2C$ (C représente une cystéine et X un acide aminé quelconque). Chaque motif contient 4 cystéines chélatant un ion de zinc ainsi que 2 hélices α amphipatiques, orientées perpendiculairement l'une par rapport à l'autre, qui interagissent spécifiquement avec les séquences d'ADN des éléments de réponse hormonale (Freedman *et al.*, 1988a; Freedman *et al.*, 1988b; Luisi *et al.*, 1991). Cette structure ressemble aux doigts de zinc classiques du facteur de transcription TFIIIA de *Xenopus laevis* (Brown *et al.*, 1985; Miller *et al.*, 1985), ainsi qu'à un grand nombre de protéines liant les acides nucléiques (Klug and Rhodes, 1987). Contrairement à TFIIIA, dans lequel les doigts de zinc lient chacun l'ADN de façon indépendante, un seul motif de liaison au zinc est impliqué dans la liaison à l'ADN dans les RNs, l'autre permettant la dimérisation au contact de l'ADN.

Récepteur(s)	Oligomérisation	PDB	Méthode	Référence
En absence d'ADN				
rGR	Monomère	1gdc	RMN	(Baumann <i>et al.</i> , 1993)
	Monomère	2gda		
hRXR α	Monomère	1rxr	RMN	(Holmbeck <i>et al.</i> , 1998)
En complexe avec l'ADN				
hER α	Homodimère sur IR3	1hcq	RX	(Schwabe <i>et al.</i> , 1993)
hRev-ERB α	Monomère	1ga5	RX	(Sierk <i>et al.</i> , 2001)
hVDR	Homodimère sur DR3	1kb4	RX	(Shaffer and Gewirth, 2002)
hPR	Momodimère	2c7a	RX	(Roemer <i>et al.</i> , 2006)
hRAR α -hRXR α	Hétérodimère sur DR1	1dsz	RX	(Rastinejad <i>et al.</i> , 2000)
rGR	Homodimère sur IR4	1glu	RX	(Luisi <i>et al.</i> , 1991)
hRXR α	Homodimère sur DR1	1by4	RX	(Zhao <i>et al.</i> , 2000)
dECR-hRXR α	Hétérodimère sur IR1	1r0n	RX	(Devarakonda <i>et al.</i> , 2003)
dECR-dUSP		1r0o		
hERR2	Monomère	1lo1	RMN	(Gearhart <i>et al.</i> , 2003)
hRev-ERB α	Homodimère sur DR2	1a6y	RX	(Zhao <i>et al.</i> , 1998)
		1hlz		(Sierk <i>et al.</i> , 2001)
dECR-dUSP	Hétérodimère sur 20E	2han	RX	(Jakob <i>et al.</i> , 2007)
rNGFI-B	Monomère	1cit	RX	(Meinke and Sigler, 1999)
mSF1	Monomère	2ff0	RMN	(Little <i>et al.</i> , 2006)
hRXR α -hVDR	Hétérodimère sur DR3	1ynw	RX	(Shaffer and Gewirth, 2004)
hLRH1	Monomère	2a66	RX	(Solomon <i>et al.</i> , 2005)
hRXR α -hTR β	Hétérodimère sur DR4	2nll	RX	(Rastinejad <i>et al.</i> , 1995)

Tableau 2 | Structures de domaines de liaison à l'ADN référencées dans la PDB (*Protein Data Bank*). Les structures ont été résolues par cristallographie aux rayons X ou par RMN, en absence ou en complexe avec l'ADN. r pour *Rattus rattus*, h pour *Homo sapiens*, d pour *Drosophila melanogaster*, m pour *Mus musculus*, DR pour *Direct Repeat*, IR pour *Inverted Repeat* (palindrome), suivi d'un chiffre qui correspond à l'espacement entre les deux demi-sites (voir 1.2.2.3). 20E pour élément de réponse naturel pseudo palindromique du promoteur du gène *hsp27* de *D. melanogaster*. Les codes PDB correspondent aux codes d'accès des structures de DBD correspondantes dans la PDB (<http://www.rcsb.org>).

1.2.2.2. Éléments structuraux impliqués dans la liaison à l'ADN et la dimérisation

Les structures des complexes DBD-ADN ont permis de définir les déterminants de la spécificité de reconnaissance des récepteurs envers leur ADN. Plusieurs éléments ont été caractérisés (boîtes P, D, A et T) définissant ou contribuant à la spécificité de l'élément de réponse, à l'interface de dimérisation et aux contacts formés avec l'ADN. L'organisation

schématique de ces différents éléments structuraux est représentée sur la Figure 3 pour le DBD de hRXR α .

La boîte P (boîte Proximale) se situe au niveau de l'hélice H1 du premier motif à doigt de zinc et fait partie de l'hélice de reconnaissance (du demi-site) interagissant avec le grand sillon de l'ADN. La boîte D (boîte Distale) est, quant à elle, impliquée dans l'interface de dimérisation (notamment dans le cas des récepteurs aux stéroïdes) et dans la discrimination des éléments de réponse. Les boîtes T et A dont les rôles ont notamment été définis dans le cas du TR et du RXR interviennent respectivement dans la dimérisation et la reconnaissance de l'ADN au niveau du petit sillon. Ces deux boîtes correspondent à l'extension C-ter du DBD, ou CTE (*Carboxy Terminal Extension*) (Giguere, 1999).

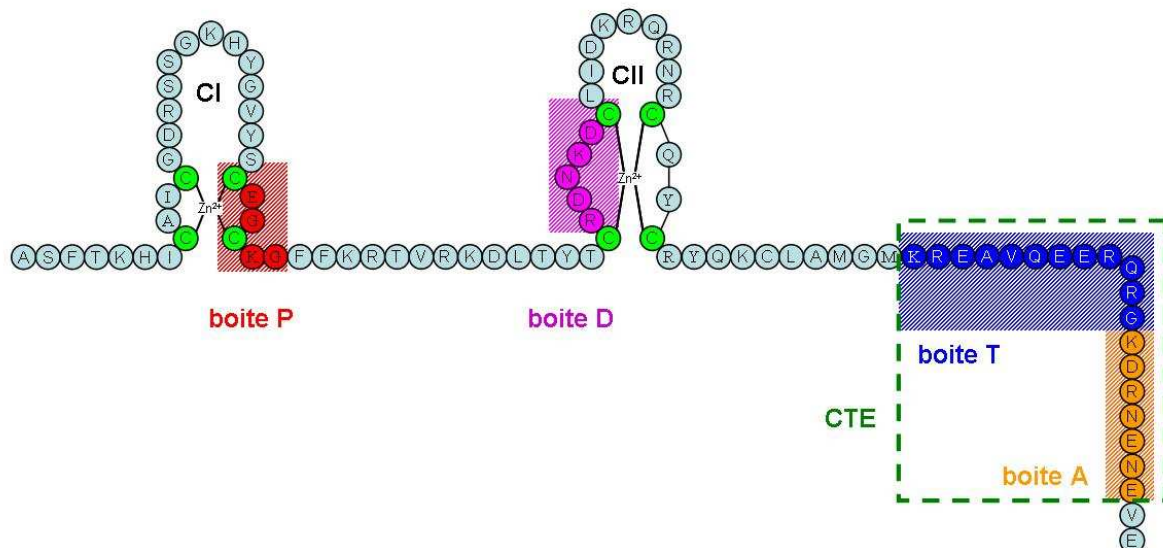


Figure 3 | Illustration schématique du DBD de hRXR α . Les cystéines sont colorées en vert, les boîtes P, D, T et A en rouge, mauve, bleu et orange respectivement. La région encadrée en pointillés correspond à la CTE.

1.2.2.3. Les éléments de réponse

Tous les RNs reconnaissent des éléments de réponse, ou demi-sites, dérivés d'une séquence hexanucléotidique canonique 5'-PuGGTCA-3' (Pu = A ou G) (Gronemeyer and Laudet, 1995; Gronemeyer and Moras, 1995). Des événements de mutation et de duplication ont généré toute une série de répétitions de ces séquences, caractérisées par des orientations et des polarités différentes. Il s'agit de répétitions palindromiques caractéristiques des récepteurs de la sous-famille NR3 (stéroïdiens) et espacées de trois paires de bases (aussi

appelées IR3 pour *Inverted Repeat*) (Bain *et al.*, 2007; Schwabe *et al.*, 1995); de répétitions directes (*DR, Direct Repeat*) pour les hétérodimères des sous-familles NR1 et NR4 avec RXR ; d'une séquence simple (un demi-site unique) dans le cas des RNs orphelins monomériques comme NGFI-B (Glass, 1994). Les répétitions directes (*DR, direct repeats*), notées DR1 à DR5, sont particulièrement intéressantes car elles présentent une polarité et sont responsables de la liaison sous forme homodimérique de RNs non stéroïdiens comme HNF4 et RXR et surtout, hétérodimérique avec RXR et des récepteurs tels que les récepteurs aux rétinoïdes (RAR), aux hormones thyroïdiennes (TR), à la vitamine D (VDR) et aux éicosanoïdes (PPAR). La spécificité de liaison à l'ADN dépend donc de la séquence des demi-sites, de l'orientation des répétitions de ces demi-sites et de l'espacement entre les séquences. L'organisation de ces séquences et la nature des associations entre récepteurs et éléments de réponse sont décrites sur les Figure 4 et Figure 5 (Mangelsdorf and Evans, 1995; Mangelsdorf *et al.*, 1995).

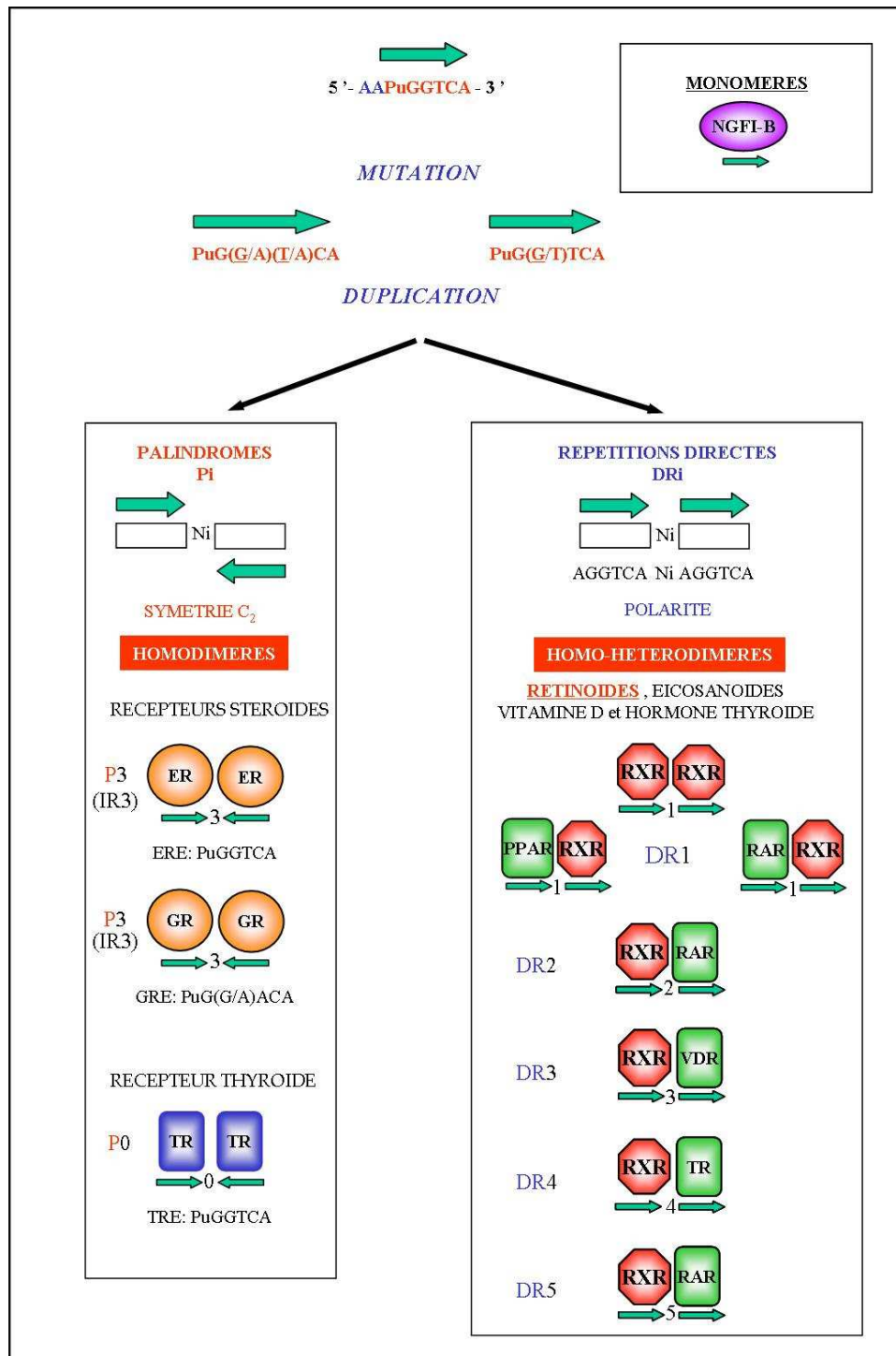


Figure 4 | Types d'éléments de réponse des RNs.

Le récepteur RXR est capable de former des homodimères sur des séquences DR1, mais il présente aussi la propriété essentielle de former des hétérodimères avec tous les autres membres de la sous-famille NR1 (Gronemeyer and Laudet, 1995). Il est donc un partenaire central et joue un rôle pivot dans toutes les voies de signalisation impliquant les hétérodimères RXR-autre récepteur (Kastner *et al.*, 1995). Chaque hétérodimère est spécifique

d'un DR donné. A titre d'exemple, un DR3 sera le plus souvent reconnu par un hétérodimère RXR-VDR, un DR4 par un hétérodimère RXR-TR et un DR5 par un hétérodimère RXR-RAR. (Glass, 1994; Kurokawa *et al.*, 1993; Mader *et al.*, 1993; Perlmann *et al.*, 1993; Zechel *et al.*, 1994). Contrairement au cas symétrique de la séquence palindromique reconnue par l'homodimère ER (Figure 4 et Figure 5B), l'asymétrie de l'élément de réponse DR se traduit par l'asymétrie dans la disposition de l'hétérodimère RXR-TR (Figure 5C). La structure du complexe RXR-TR-DR4 est représentative de l'organisation sur les DR (Rastinejad *et al.*, 1995).

De plus, la polarité des hétérodimères est stricte. Ainsi, la sous-unité RXR est systématiquement en position 5' dans les hétérodimères RXR-TR, RXR-VDR, alors qu'elle occupe la position 3' dans l'hétérodimère PPAR-RXR. Le cas des hétérodimères RXR-RAR et RAR-RXR, qui se fixent respectivement sur les éléments DR5 et DR1, est remarquable car les deux hétérodimères ont des propriétés très différentes au niveau de la transcription. La complexité et la diversité du répertoire de séquences DR associées à la combinatoire des interactions entre différents récepteurs sous forme d'homodimères ou d'hétérodimères sont à l'origine de régulations fines au niveau des voies de signalisation impliquant ces récepteurs.

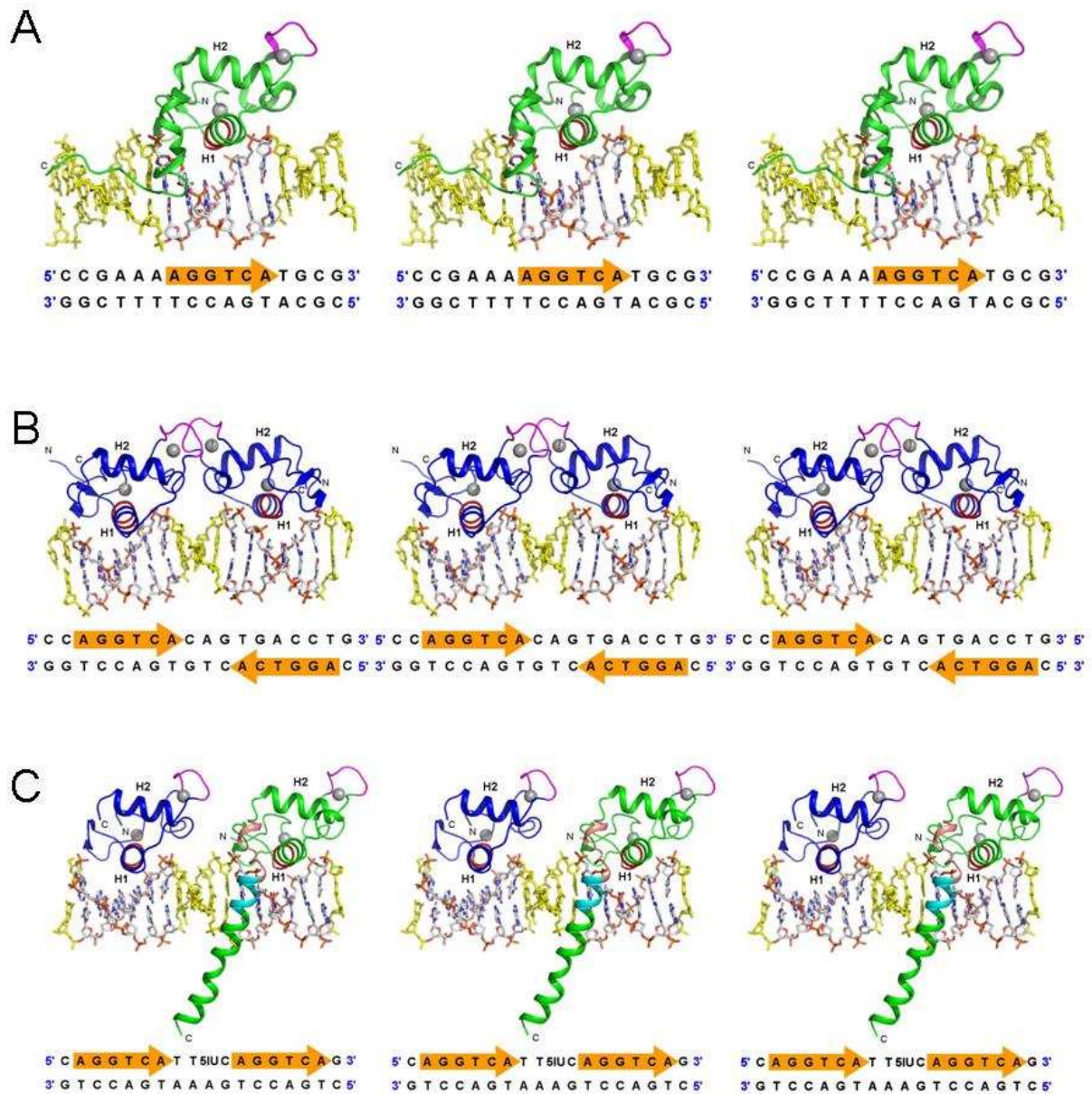


Figure 5 | Vue stéréoscopique¹ de différents modes d'interaction DBD-ADN. La boîte P est colorée en rouge, la boîte D en mauve, les atomes de zinc sont schématisés par des sphères grises. Le demi-site est représenté par une flèche orange. (A), structure cristallographique du monomère NGFI-B du rat (1CIT, (Meinke and Sigler, 1999)). (B), structure cristallographique de l'homodimère de hER α en complexe avec son élément de réponse IR3 (1HCQ, (Schwabe *et al.*, 1993). Le sens des hélices H2 de chaque monomère étant opposé, on dit que les deux monomères sont disposés « en tête à tête ». (C), structure cristallographique de l'hétérodimère hRXR α (bleu) et hTR β (vert) en complexe avec un élément de réponse DR4 (2NLL, (Rastinejad *et al.*, 1995). Le sens des hélices H2 de chaque monomère étant le même, on dit que les deux monomères sont disposés « en tête à queue ».

¹ Les figures peuvent être vues en trois dimensions à l'oeil nu, soit par la technique dite « parallèle » en utilisant les deux figures de gauche, soit par la technique dite « croisée » en utilisant les deux figures de droite.

1.2.3. Le domaine D

Le domaine D est, avec le domaine A/B, une des régions la moins conservée des RNs à la fois en séquence et en taille. Initialement, on pensait que la principale fonction du domaine D était de jouer le rôle d'une charnière flexible reliant le DBD (domaine C) au LBD (domaine E), permettant ainsi le libre positionnement du DBD par rapport au LBD. En effet, comme nous l'avons vu précédemment (Figure 5B et C) le DBD effectue une rotation de 180° et présente des interfaces de dimérisation différentes selon l'élément de réponse situé sur l'ADN (c'est-à-dire en fonction de l'orientation des demi-sites et de l'espacement entre les séquences), contrairement au LBD qui, comme nous le verrons par la suite, présente des interfaces dimériques sensiblement identiques quelque soit le type de dimérisation. Dans ce contexte, la flexibilité du domaine D devait permettre à certains RNs de se fixer comme dimères à la fois sur des éléments de réponse directs (DR) et inversés (IR) (Glass, 1994).

Cependant, des travaux récents ont montré que le domaine D joue un rôle actif au sein des RNs, en intégrant de nombreux signaux de régulation de leur localisation et de leur activité en général. Ainsi, dans le cas du récepteur de l'androgène (AR), cette région charnière est apparue comme un domaine multifonctionnel impliqué à la fois dans la fixation à l'ADN et dans la translocation nucléaire (Haelens *et al.*, 2007). Le domaine D de AR peut être acétylé, acétylation qui peut être impliquée dans l'activité transcriptionnelle (Fu *et al.*, 2000; Wang *et al.*, 2001) et dans le repliement de AR (Thomas *et al.*, 2004). Ce récepteur possède un signal de localisation nucléaire (NLS, *Nuclear Localization Signal*) bi-partite (617-RKCYEAGMTLGARARKLKK-633) constitué de deux groupes de résidus basiques (soulignés dans la séquence) et localisé en partie sur le DBD et la région charnière. Deux mutations (R629Q et K630T, en gras dans la séquence) situées dans la deuxième moitié du NLS et trouvées chez deux patients atteints de cancer de la prostate induisent une activité transcriptionnelle élevée, mais sont aussi à l'origine d'une translocation nucléaire incomplète. Toutefois, cette activité élevée est indépendante de la translocation nucléaire incomplète.

D'autres exemples concernant une translocation incomplète liée à des modifications dans le domaine D existent dans la littérature pour VDR (Michigami *et al.*, 1999) ou encore DAX-1 (Bernard *et al.*, 2006). Chez ER α , la sumoylation de la région D entraîne un effet positif de

l'activité transcriptionnelle (Sentis *et al.*, 2005). De même, la phosphorylation de résidus du domaine D de LRH-1 stimule la transcription.

De plus, des études ont montré que le domaine D pouvait servir de point d'ancrage pour des corépresseurs (CoR) (Horlein *et al.*, 1995) ainsi que des CoA, tels que PGC-1 (*PPAR Gamma Coactivator-1*) qui interagit à la fois avec une partie du DBD et de la région D de PPAR γ (Puigserver *et al.*, 1998) et de ER α (Tcherepanova *et al.*, 2000).

1.2.4. Le domaine E

1.2.4.1. Structure et fonction

Le second domaine conservé, le domaine E, correspond au domaine de fixation du ligand (LBD, *Ligand Binding Domain*). C'est un domaine multifonctionnel assurant la fixation du ligand dans une poche (LBP, *Ligand Binding Pocket*), la fonction de transactivation de la transcription dépendante du ligand (AF-2) sur l'hélice H12 (Gronemeyer and Laudet, 1995) et l'interface de dimérisation impliquée dans l'homo- ou l'hétérodimérisation. De plus, ce domaine interagit avec des cofacteurs et des protéines de choc thermique (HSP, *Heat Shock Protein*).

De nombreuses structures de LBDs (plus de 200) couvrant l'ensemble de la super-famille des RNs, résolues par cristallographie aux rayons X sont disponibles. Toutes les structures de LBD disponibles révèlent un domaine globulaire qui adopte un repliement similaire, composé de 12 hélices α (H1 à H12) anti-parallèles disposées en trois couches (Moras and Gronemeyer, 1998; Wurtz *et al.*, 1996) formant un sandwich d'hélices α anti-parallèles et d'un petit feuillet β de deux brins antiparallèles, situé entre les hélices H5 et H6. La couche centrale formée des hélices H4, H5, H6, H8 et H9 est prise en sandwich entre la couche formée des hélices H1 à H3 et la couche constituée des hélices H7 et H10 (Figure 6).

Des variations existent autour de cette organisation canonique, notamment au niveau de la zone de connexion entre les hélices H1 et H3, où l'on a pu observer l'absence de l'hélice H2 par exemple chez RAR γ (Renaud *et al.*, 1995) ou la présence d'une hélice H2' supplémentaire chez PPAR γ (Shiau *et al.*, 1998). De même, on observe d'autres variations telles que l'absence ou l'augmentation du nombre de brins du feuillet β , comme dans le cas de PPAR, une large insertion d'une soixantaine de résidus en amont de l'hélice H3 chez VDR, ou encore une

courte extension de 12 résidus à l'extrémité C-ter de H12 qui adopte une conformation de brin β chez PR.

L'analyse des séquences protéiques des LBDs des RNs a mis en évidence un motif signature (Wurtz 1998) d'une vingtaine d'acides aminés [(F,W)AKX₄FX₂LX₃DQX₂LL], situé entre l'extrémité C-ter de l'hélice H3 et le milieu de l'hélice H5. Ce motif contient la plupart des résidus conservés au sein de la super-famille des RNs nécessaires à la stabilisation du cœur du repliement canonique du LBD.

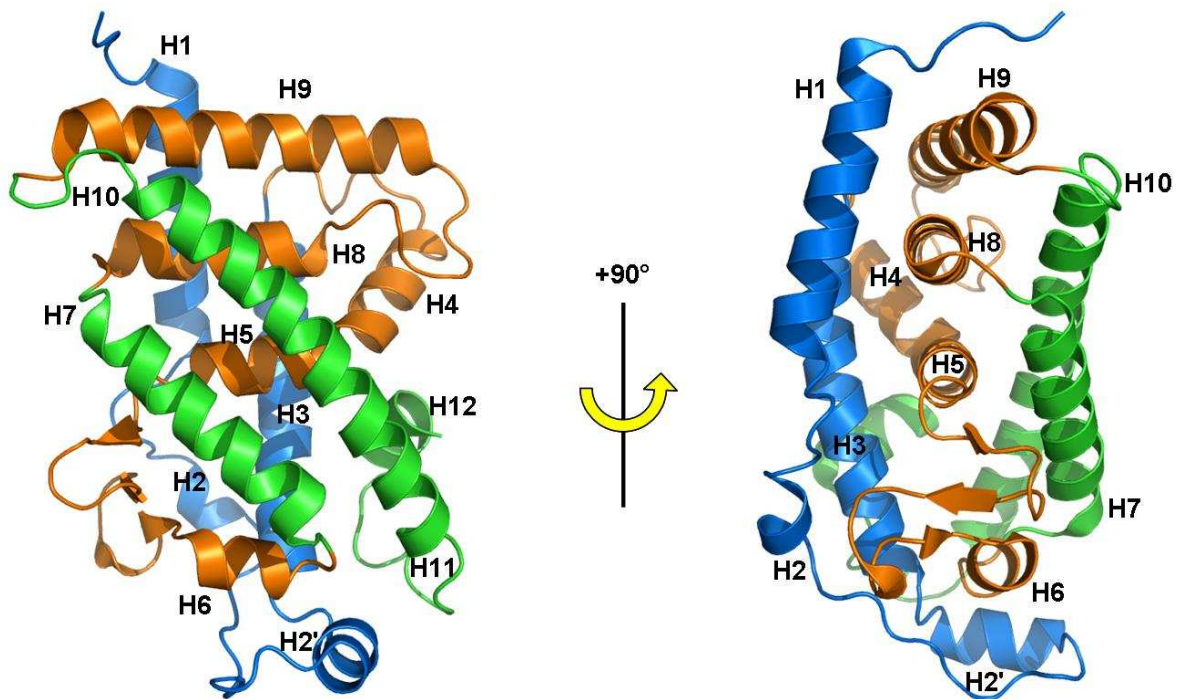


Figure 6 | Structure du LBD du RN à la Vitamine D (code PDB : 1S19 (Tocchini-Valentini *et al.*, 2004). Cette structure est constituée de 12 hélices α anti-parallèles et un feuillet de deux brins β . Les hélices α sont organisées en un sandwich d'hélices α avec en première couche les hélices H1, H2 et H3 (bleu), la couche centrale formée des hélices H4, H5, H6, H8 et H9 (orange) et la dernière couche des hélices H7 et H10 (vert). On peut noter l'existence d'une hélice H2' supplémentaire pour ce récepteur.

1.2.4.2. Interface de dimérisation

L'interface de dimérisation met principalement en jeu l'hélice H10 de chaque partenaire. Bien que les homodimères et les hétérodimères exhibent la même structure canonique avec les hélices H10 formant le squelette de l'interface (Figure 7), cette interface est stabilisée de plus par des contacts existant d'un côté entre l'hélice H7 d'un monomère et la boucle L8-9 (L pour *loop*) du partenaire et, d'un autre côté, entre l'hélice H9 et l'extrémité N-ter de l'hélice H10 (Bourguet *et al.*, 1995; Bourguet *et al.*, 2000; Egea *et al.*, 2002; Gampe *et*

al., 2000; Svensson *et al.*, 2003). La surface totale de dimérisation accessible pour les homodimères (3000 Å² pour ER) est plus grande que pour les hétérodimères (2400 Å² pour RXR-RAR) (Brzozowski *et al.*, 1997). En effet, les structures cristallographiques montrent que les hétérodimères dévient significativement de l'axe de symétrie d'ordre 2 observé chez les homodimères. De plus, l'association asymétrique d'un hétérodimère est plus stable que celle symétrique d'un homodimère RXR (Bourguet *et al.*, 2000).

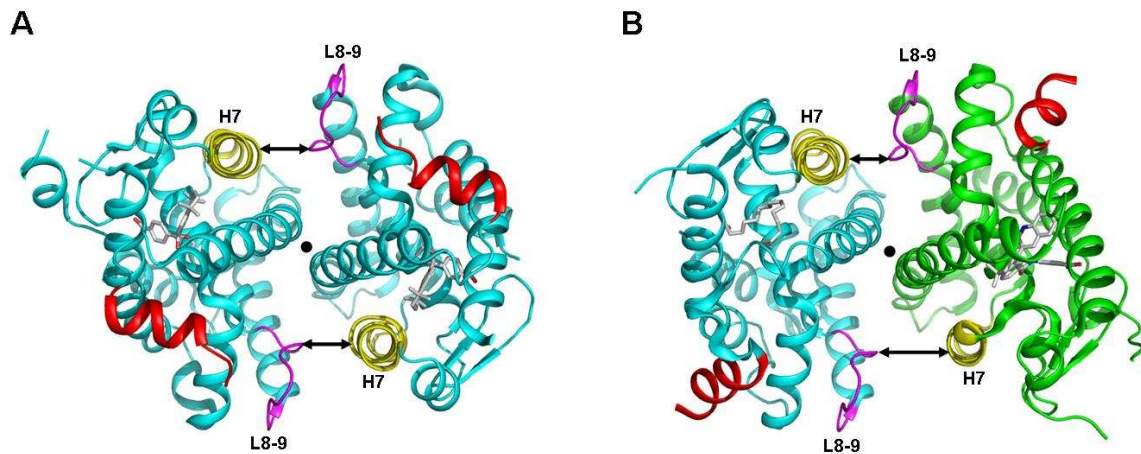


Figure 7 | Structures cristallographiques vues selon l'axe de symétrie d'ordre 2. (A), structure cristallographique de l'homodimère symétrique du LBD de hRXRα (cyan) complexé à son ligand BMS649 (code PDB : 1MZN). (B), structure cristallographique de l'hétérodimère asymétrique du LBD de hRXRα (cyan) et de hRARα (vert) liés à l'acide oléique et au BMS614, respectivement (code PDB : 1DKF). Les flèches indiquent les distances entre l'hélice H7 (jaune) et la boucle L8-9 (magenta). L'hélice H12 est colorée en rouge.

1.2.4.3. La poche de fixation du ligand

La poche de fixation du ligand, ou LBP, majoritairement hydrophobe, est localisée à l'intérieur même de la structure du LBD, encadrée par l'hélice H3 d'un côté et les hélices H7 et H10 de l'autre, ainsi que par le feuillet β . Au-delà des aspects liés à la reconnaissance et à la fixation du ligand par les résidus tapissant la LBP, se pose la question du mécanisme d'entrée et de sortie du ligand. Le mécanisme du piège à souris (Chapitre 1.2.4.4) est une solution élégante à ce problème. Il est basé sur un réarrangement structural majeur du LBD, mettant en jeu principalement la flexibilité de l'hélice H12, élément crucial qui endosse le rôle de couvercle de la LBP (Blondel *et al.*, 1999; Egea *et al.*, 2000; Moras and Gronemeyer, 1998).

Bien que le repliement des LBDs soit conservé au sein de la super-famille des RNs, les LBPs présentent une grande variété de taille, de 30Å³ pour le récepteur DHR38 à 1400 Å³ pour le récepteur PPAR (Li *et al.*, 2003). La grande taille de la poche, caractéristique des RNs orphelins comme PXR et PPAR, permet de fixer de nombreux ligands de tailles différentes au détriment de l'affinité de liaison pour ceux-ci. Au contraire, la LBP de petite taille des RNs classiques (non orphelins) limite ces récepteurs à ne reconnaître qu'un ligand spécifique, mais avec une affinité élevée. La spécificité de fixation du ligand est aussi définie par la forme de la LBP qui varie grandement d'un récepteur à l'autre pour accompagner une pléthore de fonctions contrôlées par ces récepteurs.

En plus de la taille et de la forme, la nature hydrophobe ou hydrophile de la surface de la LBP joue un rôle déterminant dans la spécificité de fixation du ligand. Affinité et sélectivité de reconnaissance du ligand sont assurées par les nombreux contacts hydrophobes et les points d'ancrage hydrophiles dans la poche. Dans le cas de l'acide rétinoïque *9-cis*, le ligand est capable de s'adapter à la forme des poches de RXR et RAR et d'optimiser les contacts hydrophobes. Par contre, pour les récepteurs stéroïdiens, les ligands sont plus rigides. La sélectivité ne serait pas uniquement due aux contacts de Van der Waals, qui ne suffiraient pas à distinguer les petites différences structurales des ligands (Figure 8).

	Ligand	Récepteur	Sous-famille	
	Structure	Nom		
Thyroïde		3, 3', 5 Triiodothyronine (T3)	TR	NR1
Isoprenoïdes et Eicosanoïdes		Prostaglandine J_2	PPAR	NR1
		Acide Retinoïque 9- <i>cis</i>	RXR RAR	NR2 NR1
		Acide Retinoïque tout- <i>trans</i>	RAR	NR1
Steroïdes		Vitamine D ₃	VDR	NR1
		Aldosterone	MR	NR3
		Cortisol	GR	NR3
		Progesterone	PR	NR3
		RU486 (mifepristone)		
		Testosterone	AR	NR3
		Oestradiol-17 β	ER	NR3

Figure 8 | Ligands et RNs associés. Le code couleur correspond à celui six grandes sous-familles de RNs présentées dans la section 1.3.3.

1.2.4.4. Le mécanisme du piège à souris

La comparaison des structures cristallographiques du LBD sous forme *apo* (absence de ligand) (Bourguet *et al.*, 1995) et *holo* (présence de ligand) (Egea *et al.*, 2000; Renaud *et al.*, 1995) a révélé un mécanisme intéressant grâce auquel les RNs deviennent transcriptionnellement actifs après fixation du ligand (Figure 9). Ce mécanisme, dénommé « modèle du piège à souris » (Moras and Gronemeyer, 1998; Wurtz *et al.*, 1996), induit une importante modification conformationnelle dans la structure du LBD et conduit à une compaction du LBD (de 39110 à 35870 Å³ chez RXR).

Les changements conformationnels majeurs impliquent les hélices H3, H11 et H12 et dans une moindre mesure l'hélice H6 et le feuillet de brins β . Les autres hélices ne sont pas affectées. La forme *apo* du LBD hRXR α comporte une hélice supplémentaire H2 connectée à l'hélice H3 par la boucle Ω . Cette même région correspond à une boucle entre les hélices H1 et H3 dans la forme *holo*. En absence de ligand, H11 stabilise la poche de fixation du ligand, cavité hydrophobe béante, en la remplissant de ses résidus apolaires.

Lors de la fixation du ligand, sous l'effet d'un changement de conformation de cette boucle Ω , H3 est déplacée d'une dizaine d'Angströms vers la poche de fixation du ligand. Ce mouvement accompagne le repositionnement de H11 dans la continuité de H10 et induit le basculement et le repositionnement de H12 qui était complètement exposée au solvant dans la forme *apo*, vers le cœur du LBD. Cette interaction est stabilisée par des contacts avec des résidus de H3 et H11.

Le résultat de ce réarrangement conformationnel, induit par la fixation du ligand, est la fermeture de la poche de fixation du ligand par l'hélice H12, qui agit comme un « couvercle » mobile et qui contient des résidus cruciaux pour la fonction de AF-2. L'orientation de l'hélice H12 est une conséquence des effets allostériques induits par la nature chimique particulière du ligand qui s'est fixé dans la LBP. Cette modification conformationnelle génère, de plus, de nouvelles surfaces d'interactions.

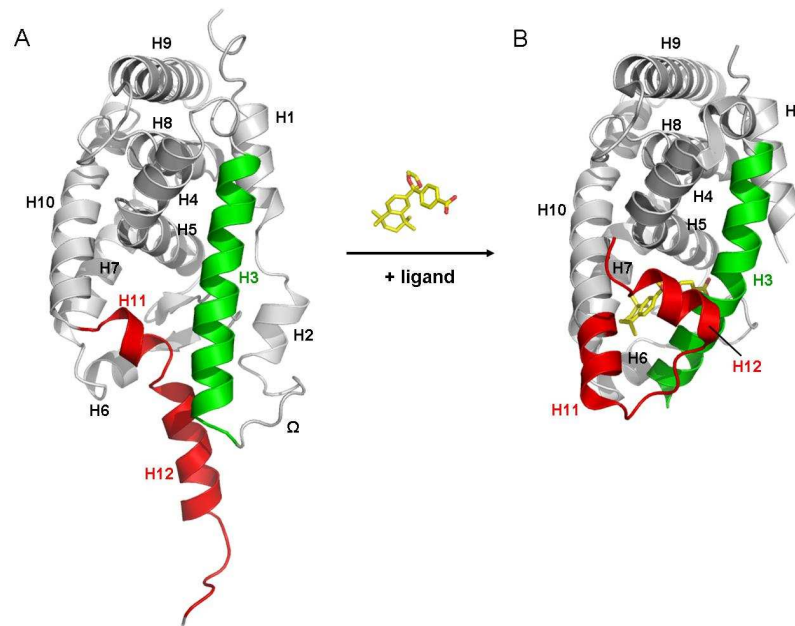


Figure 9 | Effets de la fixation du ligand sur la conformation du LBD de hRXR α . (A), Forme *apo* (absence de ligand) du LBD de hRXR α (Bourguet *et al.*, 1995), code PDB : 1LBD) représentée avec les éléments de structure secondaire subissant les réarrangements conformationnels majeurs lors de la fixation du ligand colorés en vert (H3) et en rouge (H11 et H12). (B), Forme *holo* (présence de ligand), lors de la fixation du ligand, dans ce cas BMS649 (Egea *et al.*, 2000), code PDB : 1MVC) de hRXR α , les modifications structurales majeures permettent à l'hélice H12 d'être stabilisée dans une position dite active.

1.2.5. Le domaine F

Certains RNs, comme HNF4, ER, PR, MR ou GR possèdent un domaine supplémentaire situé à leur extrémité C-ter, le domaine F.

1.2.5.1. Structure

A l'heure actuelle, seules les structures de ce domaine des récepteurs aux glucocorticoïdes (Bledsoe *et al.*, 2002), aux minéralocorticoïdes (Bledsoe *et al.*, 2005), à la progestérone (Williams and Sigler, 1998) et à l'androgène (Sack *et al.*, 2001) ont été résolues. La structure du domaine E-F du récepteur à la progestérone (PR) (Williams and Sigler, 1998) a révélé que le repliement de cette région en feuillet β antiparallèle sur le LBD engendre une gêne stérique qui entraîne la formation d'un homodimère non canonique. Cependant, en raison du nombre restreint de structures résolues de LBDs possédant un domaine F, il est difficile de juger de la validité de ce type de dimérisation.

1.2.5.2. Fonction

La fonction du domaine F a été principalement étudiée par délétion. La délétion du domaine F peut avoir des conséquences différentes sur l'affinité et la fixation du ligand. En effet, la délétion du domaine F de ER α augmente son affinité pour l'œstradiol contrairement à PR pour lequel la délétion entraîne une réduction de l'affinité et une diminution considérable de la fixation de la progestérone mais affecte peu l'affinité de l'antagoniste RU486 (voir 1.4.1) (Couette *et al.*, 1998; Tahiri *et al.*, 2001).

De plus, chez ER α , la délétion du domaine F diminue l'activité agoniste des anti-œstrogènes. Le domaine F possède une fonction de régulation de la réponse transcriptionnelle aux œstrogènes et anti-œstrogènes et les effets sont différents selon le type cellulaire (Koide *et al.*, 2007; Montano *et al.*, 1995; Nichols *et al.*, 1998; Schwartz *et al.*, 2002).

Chez HNF4 α 1 de rat, des études ont montré que la perte du domaine F diminuait la capacité du CoR SMRT (*Silencing Mediator for Retinoic acid and Thyroid hormone receptors*) (Chen and Evans, 1995) à interagir avec HNF4 α 1 *in vitro* et réprimait son activité *in vivo*. Le domaine F exerce donc un rôle de régulateur clé pour HNF4 α 1 qui l'aide dans la discrimination entre les CoA et les CoR. De plus, ce domaine module la capacité de SMRT à réprimer l'activité de HNF4 α 1 (Ruse *et al.*, 2002).

1.3. Evolution des récepteurs nucléaires

1.3.1. Un peu d'histoire...

Depuis les débuts de la vie sur Terre, une multitude de changements climatiques et géologiques se sont succédés et ont certainement influencé l'évolution des grands groupes d'animaux. L'évolution est divisée en quatre ères : le Précambrien, le Paléozoïque, le Mésozoïque et le Cénozoïque (Figure 10). Voici un petit voyage rapide à travers le temps...

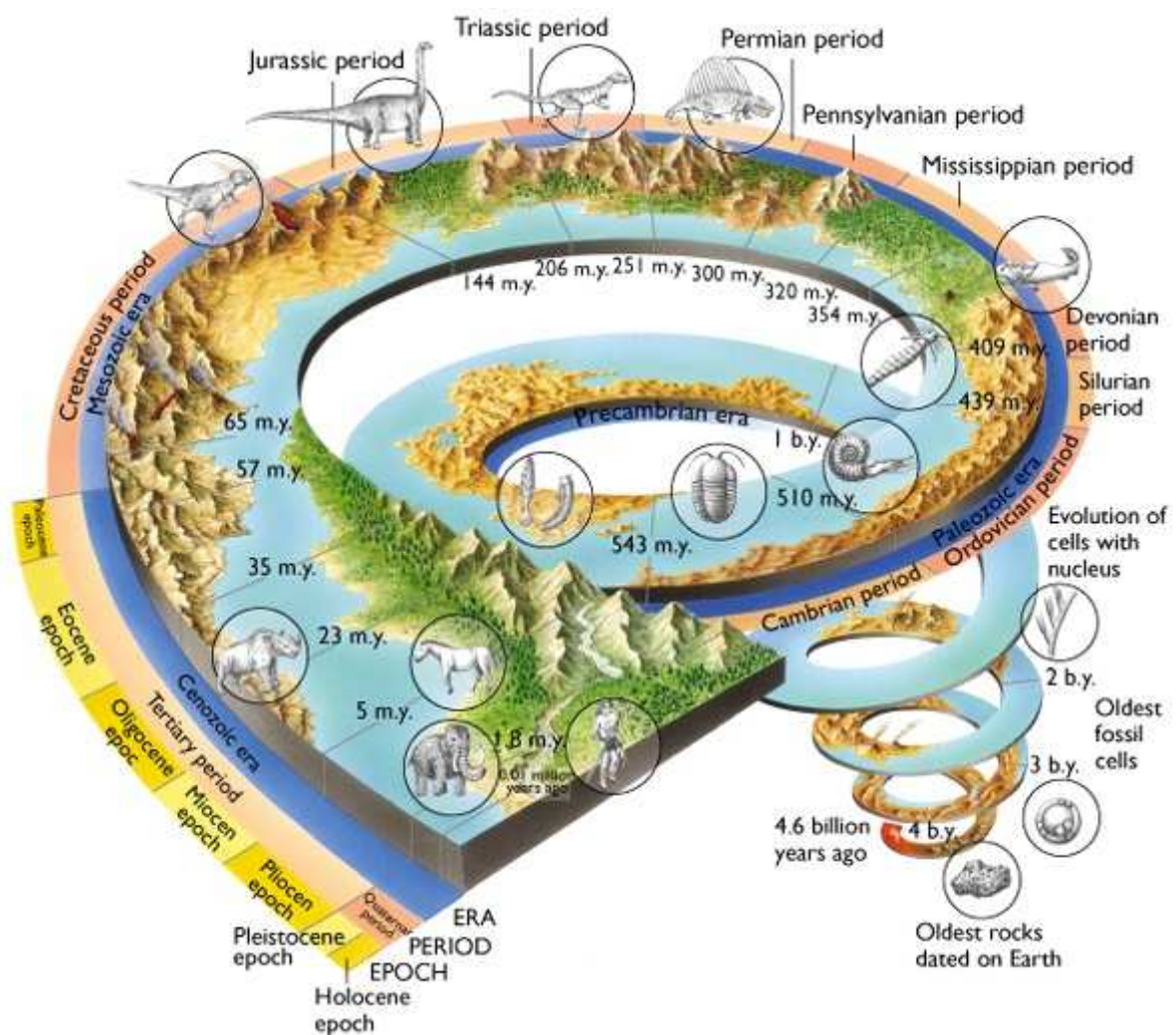


Figure 10 | Echelle des temps géologiques et biologiques (<http://pubs.usgs.gov/gip/geotime/time.html>).

1.3.1.1. Précambrien : 4600-540 millions d'années

Il s'agit de la plus longue période sur l'échelle des temps géologiques, puisqu'elle s'étend de la formation de la terre à l'émergence d'une abondante faune d'animaux. Peu de fossiles des premiers animaux sont disponibles. La plupart se fossilisent très mal et les roches qui pourraient contenir des fossiles des premiers animaux sont très vieilles et souvent endommagées. Dans les océans, les premières algues bleues, apparues vers 3,7 milliards d'années, fixent le carbonate de calcium, construisent des champignons de calcaire (les stromatolithes, du grec *strôma*, tapis, et *lithos*, pierre) et émettent les premières bulles d'oxygène. L'époque la plus ancienne pour laquelle de nombreux fossiles sont connus est la fin du Précambrien. La composition des dépôts de fossiles de cette époque a incité les paléontologistes à appeler le Précambrien le stade ver-méduse de l'évolution. A la fin du Précambrien, la plupart des embranchements actuels étaient déjà apparus, mais la faune était dominée par les méduses et les vers.

1.3.1.2. Paléozoïque : 540-250 millions d'années

Le Paléozoïque est une ère géologique divisée en six périodes, qui s'étend de 540 à 250 millions d'années. Elle est au centre de deux événements majeurs. Le premier qui s'est tenu lors de la période Cambrienne (540-505 millions d'années) et dont la résultante est une formidable explosion de la diversité des animaux avec l'apparition des poissons, des plantes à fleurs, des insectes, des Amphibiens et des Reptiles. Les cinq continents étaient encore fusionnés en un seul pour former la Pangée. Le deuxième événement majeur a eu lieu à la fin du Paléozoïque, lors de la période Permienne (290-248 millions d'années) avec la plus grande extinction en masse de l'histoire et la disparition d'environ 95% de toutes les espèces marines et de 70% des espèces vivant sur les continents induisant la diminution, voire la disparition de nombreux groupes végétaux et animaux.

1.3.1.3. Mésozoïque : 250-65 millions d'années

Le Mésozoïque, divisé en trois périodes (Trias, Jurassique, Crétacé) est surnommé l'âge des Reptiles. Pendant cette ère, un intense volcanisme est responsable de la fracturation de la Pangée : les continents commencent à dériver et le climat est chaud. Le Mésozoïque est connu sous le nom plus familier d'âge des Dinosauriens. Il est aussi marqué par l'apparition

des Oiseaux (*archaeopterix*), des Reptiles (*ichtyosaure, iguanodon, ptérodactyle*) et des Mammifères représentés par l'ordre des protothériens (Mammifères pourvus d'une poche marsupiale). Toutes les formes de vie modernes existent, bien que dans certains cas et en particulier celui des Mammifères, il s'agisse de formes primitives. A la fin de cette ère (Crétacé), il y a 65 millions d'années environ, est survenue l'extinction du Crétacé qui désigne l'extinction massive des dinosaures, à l'exception des oiseaux.

1.3.1.4. Cénozoïque : 65 millions d'années - Aujourd'hui

Le Cénozoïque est l'ère géologique actuelle. Elle représente la plus récente des trois subdivisions majeures de l'histoire des animaux, de la fin du Crétacé avec la disparition des Dinosaures jusqu'à la diversification des Oiseaux et des Mammifères, et, tout récemment à l'apparition de l'Homo. Le Cénozoïque est souvent appelé l'âge des Mammifères qui, à partir de quelques formes simples, ont divergé durant cette ère en divers embranchements d'animaux terrestres, marins et aériens. Un refroidissement graduel du climat a eu lieu au cours de cette ère. D'un point de vue géologique, les continents ont continué à dériver jusqu'à occuper leurs emplacements actuels.

1.3.2. ... Et un peu de systématique

La systématique est la branche de la biologie qui traite de la classification des organismes. Le principe central de cette discipline est de regrouper les espèces qui partagent certaines similitudes en s'appuyant sur des données phénotypiques (anatomie ou développement) ou génotypiques (classiquement ARN ribosomique) et qui proviennent d'une même lignée évolutive. La systématique est une branche importante de la biologie. Non seulement la standardisation de la nomenclature binomiale (genre et espèce) permet d'appeler les animaux par le même nom, mais le regroupement d'organismes apparentés nous permet d'inférer des caractéristiques sur un organisme à partir de ce qui est connu de son ancêtre ou de ses cousins. Cependant, la classification est en continuel changement. Au gré des nouvelles recherches et découvertes, le regroupement des organismes est modifié pour mieux refléter les différentes lignées évolutives. Comme il y a encore beaucoup d'éléments manquants, il existe souvent des contradictions qui sont autant d'hypothèses.

Il y a également une résistance naturelle au changement qui retarde l'adoption de certains regroupements par l'ensemble des zoologistes.

La Figure 11 présente l'arbre phylogénétique des Eucaryotes (du grec *eu*, vrai et *karuon*, noyau). Chez les Eucaryotes, les organismes sont généralement regroupés en quatre règnes : les Protistes (du grec *protos*, premier), les Végétaux, les Champignons et les Animaux. Le règne animal est composé d'organismes multicellulaires et hétérotrophes (qui utilisent une source de carbone organique).

La première branche du règne Animal, celle des Parazoaires (du grec *parazo*, à côté et *zoon*, animal), comprend des organismes eucaryotes pluricellulaires qui n'ont pas de véritables tissus comme les Eponges, contrairement au reste des animaux, les Métazoaires (du grec *Metazo*, après et *zoon*, animal). Les éponges sont des organismes aquatiques filtreurs, sessiles et sans plan de symétrie bien défini (asymétriques).

Les Métazoaires ont une répartition mondiale et occupent tous les milieux, y compris par exemple les fumeurs noirs localisés sur les dorsales océaniques. Il faut noter que la phylogénie des Métazoaires qui suit est, par ailleurs, encore pleine d'incertitudes.

Les Métazoaires se divisent ensuite selon le plan de symétrie du corps en animaux à symétrie radiale, ou diploblastiques, dont la paroi corporelle est formée de deux couches de tissus, l'ectoderme et l'endoderme, séparés par la mésoglée (ou mésenchyme) et en animaux à symétrie bilatérale, ou triploblastiques. Ces derniers possèdent trois véritables feuillettes cellulaires : l'ectoderme, le mésoderme, et l'endoderme.

Les Cnidaires (du grec *knide*, ortie et du latin *aria*, comme) forment un groupe charnière entre les premiers organismes multicellulaires et ceux qui sont apparus après. Les Cnidaires sont des animaux diploblastiques (anémones de mer, méduses, hydres) et représentent les premiers organismes au niveau d'organisation cellules-tissus.

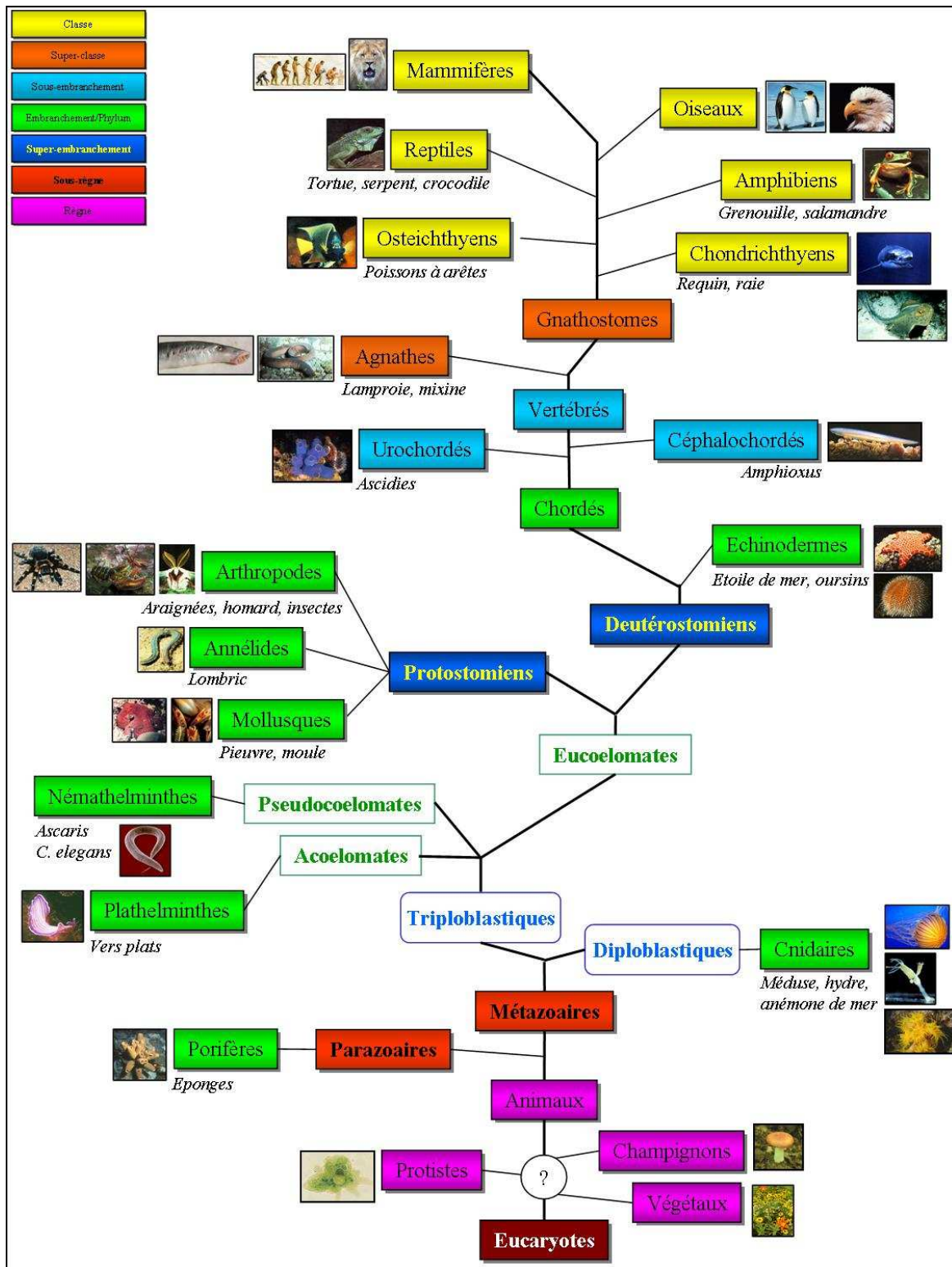


Figure 11 | Arbre phylogénétique simplifié des Eucaryotes (adaptée de <http://simulium.bio.uottawa.ca/bio2525/Notes/bio2525.htm>). La longueur des branches est arbitraire.

La branche suivante, celle des triploblastiques, est sujette aux plus grandes controverses. Classiquement, les animaux à symétrie bilatérale sont divisés en trois groupes, en fonction de la présence d'une cavité interne, le coelome (*koilos*, cavité). On a donc une branche d'animaux sans coelome, les Acoelomates, comme les vers plats de l'embranchement des Plathelminthes (du grec *platus*, plat et *helmins*, ver), des animaux avec une cavité interne qui n'est pas un « vrai » coelome, les Pseudocoelomates, comme l'ascaris de l'embranchement des Némathelminthes, et des animaux avec un « vrai » coelome, les Eucoelomates, comme le ver de terre de l'embranchement des Annélides et tous les animaux ayant émergé plus tardivement dans l'arbre phylogénétique.

La division suivante de l'arbre phylogénétique est basée sur des détails du développement embryonnaire. On divise les Eucoelomates en Protostomiens et Deutérostomiens (du grec *proto*, premier, *deutero*, second et *stomum*, bouche). Le nom de ces deux groupes reflète l'origine de la bouche lors du développement embryonnaire. Le super-embranchement des protostomiens se divise en trois embranchements, celui des Mollusques, des Annélides et des Arthropodes.

L'embranchement des Mollusques (du latin *mollis*, mou) est le deuxième embranchement le plus diversifié et comprend plus de 100 000 espèces vivantes. Ses représentants (huître, limace et pieuvre) occupent principalement le milieu marin, mais certains groupes ont colonisé le milieu terrestre. Contrairement aux Annélides, aux Arthropodes et aux Vertébrés, les Mollusques ne présentent pas une organisation métamérique. Ils pourraient dériver d'animaux bilatéraux acoelomates avec un épiderme cilié comme les Plathelminthes, qui auraient développé un coelome mais pas de métamères.

Les Annélides (du latin *anellus*, petit anneau) modifient l'architecture eucoelomate décrite chez les Mollusques. Au lieu d'avoir un coelome et toutes les structures qui lui sont associées, ils ont une série d'unités répétées qui contiennent chacune une cavité interne remplie de fluide. Cet arrangement d'unités répétées donne une apparence segmentée aux Annélides. Cette transformation de l'architecture est appelée métamérisation et chaque segment produit est un métamère. Dans la forme ancestrale, tous les métamères sont semblables et les structures dérivées de l'ectoderme et du mésoderme sont répétées dans chaque segment.

L'embranchement des Arthropodes (du grec *arthron*, articulation et *podos*, pied) est celui qui a le plus de succès sur notre planète. Ils sont présents dans tous les habitats, des montagnes enneigées aux fosses abyssales, des déserts aux forêts tropicales. Le critère principal de ce groupe est la présence d'appendices articulés et d'un exosquelette formé d'une cuticule qui recouvre entièrement l'extérieur de l'animal.

Les Echinodermes (du grec *echinos*, épine et *derma*, peau) constituent, avec par exemple les étoiles de mer et les oursins, le premier embranchement des Deutérostomiens. Les Echinodermes ont une symétrie radiale et si ce seul trait était utilisé pour regrouper les animaux, ils pourraient être apparentés aux Cnidaires. Cependant, ce n'est pas le cas. Si on observe le cycle de vie des Echinodermes, on réalise que la larve possède une symétrie bilatérale et que seul le stade adulte a une symétrie radiale. Les Echinodermes actuels sont, pour la plupart, mobiles. Ils se déplacent lentement à l'aide de pieds ambulacraires (tubes creux remplis de liquide), la bouche orientée vers le bas.

Les embranchements vus jusqu'à présent représentent plus de 96% des espèces animales vivantes. Le dernier embranchement est celui des Chordés (du latin *chorda*, corde), animaux bilatéraux et métamérisés. L'embranchement des Chordés est l'un des rares embranchements à avoir colonisé avec succès les environnements aquatiques et terrestres. L'origine des Chordés remonte à plus de 500 millions d'années, probablement au Cambrien, et est divisé en trois sous-embranchements : les Urochordés (du grec *ura*, queue et du latin *chorda*, corde), les Céphalochordés (du grec *cephalo*, tête et du latin *chorda*, corde) et les Vertébrés (du latin *vertebra*, relatif aux vertèbres). Ces trois sous-embranchements partagent les quatre structures communes aux Chordés : une notocorde qui assure le support longitudinal et qui est réduite chez les Vertébrés, où son rôle principal est rempli par la colonne vertébrale ; un cordon nerveux tubulaire dorsal (tube neural), la moelle épinière chez les Vertébrés ; des fentes pharyngiennes qui servent à la filtration chez les Chordés ; une queue post-anale bien développée (Figure 12).

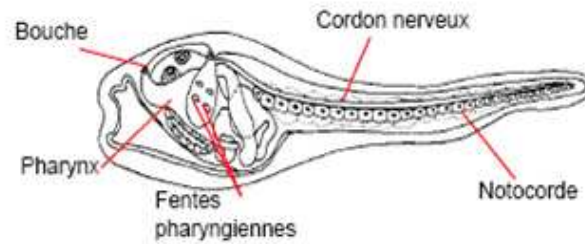


Figure 12 | Morphologie de la larve d'un Urochordé (adaptée de <http://simulium.bio.uottawa.ca/>).

Le sous-embranchement des Chordés regroupe toute une série d'organismes issus d'une radiation adaptative importante. La radiation adaptative concerne l'évolution d'une variété d'espèces à partir d'un ancêtre commun. Chaque espèce est adaptée à une niche donnée et les descendants peuvent être d'aspects très différents par rapport à leur ancêtre ainsi que les uns par rapport aux autres. Les poissons, grenouilles, serpents, dinosaures, oiseaux et même l'espèce humaine ont le même ancêtre. Dans le sous-embranchement des Urochordés (ascidie), il existe un petit groupe d'animaux qui ont un cycle vital dimorphe avec un stade adulte sessile (immobile) et une larve nectonique (nageant librement).

Dans le second sous-embranchement, celui des Céphalochordés (amphioxus), l'architecture ancestrale a été modifiée pour produire un organisme ressemblant à un poisson élancé comme *Branchiostoma lanceolatum* ou amphioxus (Figure 13). C'est dans ce sous-embranchement que les caractéristiques ancestrales de l'embranchement sont les plus évidentes.

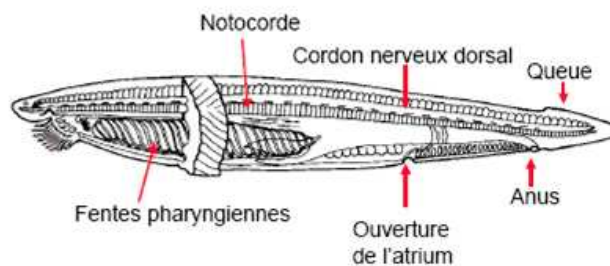


Figure 13 | Anatomie de l'amphioxus (adaptée de <http://simulium.bio.uottawa.ca/>).

Le troisième et plus important sous-embranchement est celui des Vertébrés. Le rôle de la notocorde comme élément squelettique est réduit et sa fonction ancestrale remplie par des os et cartilages. Le nom du sous-embranchement est tiré de celui des unités de base du nouveau squelette axial : les vertèbres. Dans la forme ancestrale, les ouvertures dans la paroi

du pharynx étaient supportées par des arches branchiales cartilagineuses, empêchant ainsi le pharynx de s'écraser sur lui même. Ces structures sont toujours présentes chez les Vertébrés sans mâchoires, les Agnathes (du grec *a*, sans et *gnathos*, mâchoire), comme la lamproie et la mixine.

Les Vertébrés qui ont une mâchoire, les Gnathostomes (du grec *gnathos*, mâchoire), se divisent en six classes : les Chondrichthyens (du grec *chondros*, cartilage et *ichthys*, poisson) avec le requin et la raie, les Osteichthyens (du latin *os*, os et du grec *ichthys*, poisson) avec les poissons à arêtes, les Amphibiens (du grec *amphi*, double et *bios*, vie) avec la grenouille et la salamandre, les Reptiles (du grec *herpeton*) avec la tortue, le serpent et le crocodile, les Oiseaux (du latin *avis*) et les Mammifères (du latin *mammalia*).

Comme mentionné précédemment, la position de certaines branches fait débat, en particulier au sein des Protostomiens. En s'appuyant sur des données moléculaires, certains spécialistes (Aguinaldo *et al.*, 1997; Mallatt and Winchell, 2002) proposent une division des Protostomiens en deux groupes majeurs : les Ecdysozoaires, regroupant par exemple les Nématelminthes et les Arthropodes, et les Lophotrochozoaires rassemblant notamment les Plathelminthes, les Annélides et les Mollusques. La discussion reste entièrement ouverte !

1.3.3. Evolution et répartition phylogénétique des RNs

Les RNs, absents des Végétaux, Protistes et Champignons, représentent une des plus abondantes classes de facteurs de transcription présents chez les Métazoaires (Escriva *et al.*, 1998). Par ailleurs, ils sont également absents des Parazoaires, qui incluent les éponges (Escriva *et al.*, 1997). Apparus à un stade précoce de l'évolution des Métazoaires, les RNs sont retrouvés dans les organismes les plus primitifs que sont les Cnidaires (méduse, hydre, anémone de mer), mais aussi chez tous les Triploblastiques, tant Acoelomates que Eucoelomates. Les RNs sont, en particulier, bien caractérisés dans le phylum des Chordés (Poissons, Amphibiens, Reptiles, Oiseaux et Mammifères), les Nématelminthes (*Caenorhabditis elegans*) et les Arthropodes (Insectes tels que *Drosophila melanogaster*).

L'analyse phylogénétique basée sur les alignements de séquences des DBDs et des LBDs des RNs a permis de diviser la super-famille des RNs en six sous-familles (Figure 14) (Laudet, 1997).

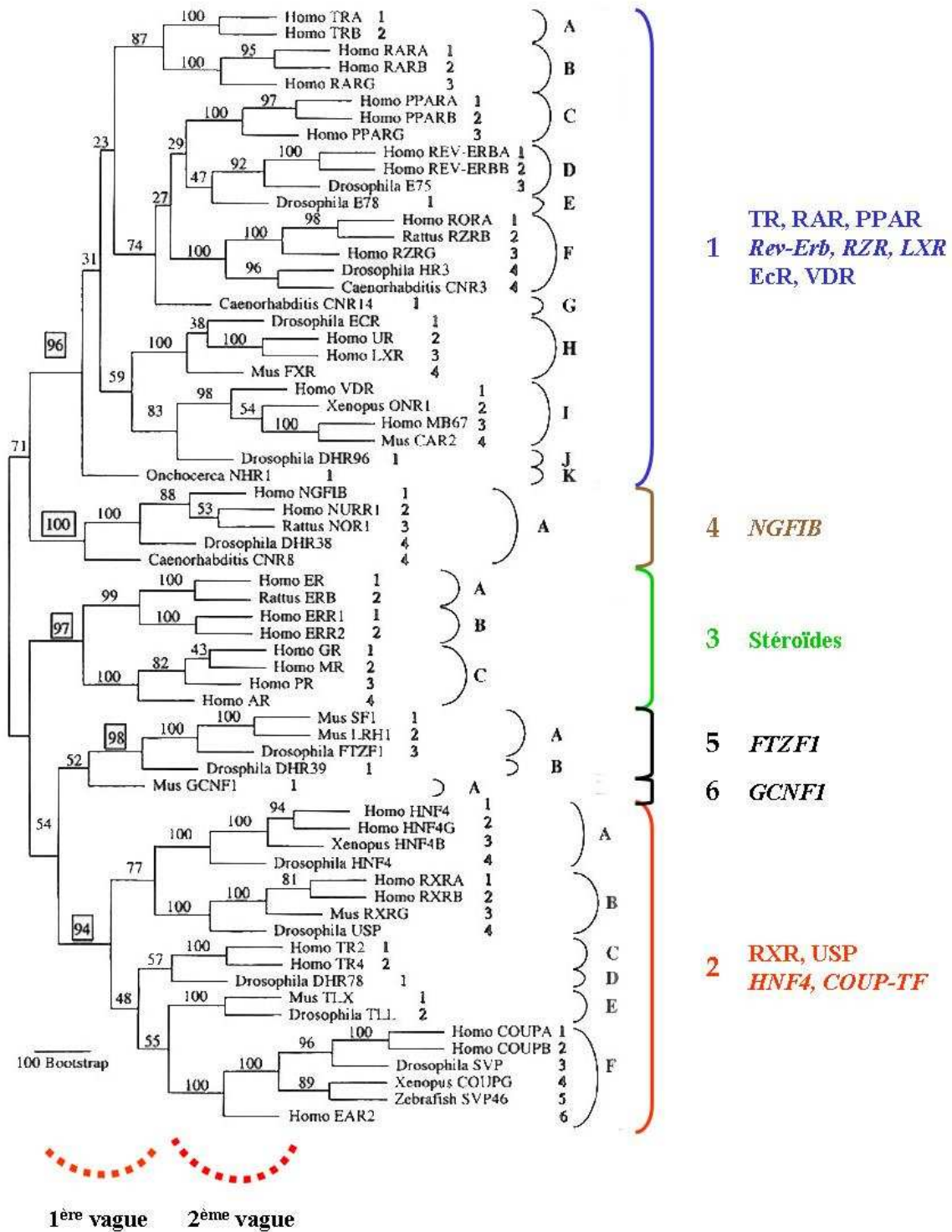


Figure 14 | Les six sous-familles de la super-famille des RNs basées sur l'étude phylogénétique de 65 gènes codant pour les RNs des Vertébrés, Arthropodes et Nématodes (adaptée de (Laudet, 1997)).

Une nouvelle nomenclature des RNs, inspirée du système élaboré pour les cytochromes P450, a été établie afin de réduire la confusion découlant du grand nombre de noms triviaux donnés aux récepteurs (Nebert *et al.*, 1987). Cette nomenclature est

principalement basée sur les travaux de Vincent Laudet (Laudet, 1997). Ainsi, le nom indique respectivement, pour chaque RN, la sous-famille, le groupe et la forme paralogue (The Nuclear Receptors Committee (1999))(Figure 14 et Tableau 1). Par exemple, NR2B1 est le gène qui code pour le récepteur nucléaire *NR* (*Nuclear Receptor*) de la classe 2, du gène B et de la forme paralogue 1, c'est-à-dire RXR α .

1.3.3.1. Origines de la diversification de la super-famille des RNs

Suite à ces résultats, un modèle concernant la diversification de la super-famille des RNs a été proposé. La diversité des RNs observée actuellement (Escriva *et al.*, 2000; Laudet, 1997) (Figure 14) serait principalement due à la succession de deux périodes de duplication de gènes. L'épisode le plus marquant dans l'évolution des RNs serait survenu très tôt lors de l'émergence des Métazoaires, probablement avant la divergence des Cnidaires, avec l'apparition des principales sous-familles actuelles de RNs. Cette première période a été gouvernée par la duplication de gènes, comme observé dans d'autres familles de protéines, par exemple les familles des protéines tyrosines-phosphatases, des protéines G et des tyrosines-kinases.

Ensuite, après la séparation Arthropodes/Vertébrés, la seconde période de duplications propre aux Vertébrés aurait conduit à l'émergence des divers groupes paralogues à l'intérieur de chaque sous-famille (TR α , TR β , RAR α , RAR β , RAR γ , etc.). Cependant, à ces duplications se joignent aussi des événements de gain et de perte de gènes qui jouent également un rôle important dans l'évolution des différents RNs chez les animaux à symétrie bilatérale (Bertrand *et al.*, 2004). Ces événements, estimés à plus de 75, ne sont pas distribués de manière régulière dans l'arbre phylogénétique. En effet, il existe des branches correspondantes à des périodes de duplication de gènes, d'autres à des périodes de perte et un petit nombre de branches où les deux événements se recouvrent. De plus, il semble que lorsque la duplication des gènes de RNs est fréquente, la perte de gènes dans la super-famille des RNs est, quant à elle, rare.

Il est intéressant de noter qu'aucun homologue de la sous-famille NR3 (stéroïdiens), n'a été mis en évidence chez *C. elegans* et *D. melanogaster*. Cette observation suggère que cette sous-famille s'est formée après la division Arthropodes/Vertébrés et est par voie de conséquence spécifique aux Deutérostomiens. Cependant, l'existence des récepteurs

stéroïdiens n'est pas l'apanage des Chordés. En effet, un récepteur stéroïdien (ER) d'un mollusque, *Aplysia californica*, a pu être récemment cloné (Thornton *et al.*, 2003). Il semble donc plus prudent de dire que ces récepteurs sont absents de certains organismes parce qu'ils ont été perdus au cours de l'évolution, mais qu'ils étaient bien présents dans le génome ancestral des animaux à symétrie bilatérale. De plus, le statut précis du ER de ce mollusque reste encore à établir par une caractérisation fonctionnelle détaillée, que ce soit au niveau de la fixation du ligand ou de l'activité transcriptionnelle. En effet, les expériences préliminaires menées par Thornton *et coll.* ont montré que ce récepteur ER n'était pas régulé par l'œstrogène (Thornton *et al.*, 2003).

Seuls les membres des sous-familles NR2 et NR5 (COUP-TF, RXR, FTZ-F1) existent chez tous les Métazoaires. Ils ont été mis en évidence chez les Cnidaires et seraient parmi les premiers récepteurs apparus, conduisant à l'hypothèse que le premier RN appartiendrait à l'une de ces deux sous-familles.

Outre la multitude de récepteurs et de ligands, une source supplémentaire de complexité et de diversité provient également du fait que, pour certains récepteurs, on distingue plusieurs isotypes (par exemple, RAR α , RAR β , RAR γ), et pour un isotype donné plusieurs isoformes (RAR α 1 et RAR α 2). Cette diversité a une double origine : les isotypes correspondent à des gènes différents (généralement localisés sur des chromosomes distincts) ; les isoformes sont générées par l'utilisation différentielle de promoteurs distincts pour un même gène combinée à des mécanismes d'épissage alternatif.

1.3.3.2. Acquisition de la capacité à fixer un ligand

L'examen de l'arbre évolutif montre que les RNs orphelins ne forment pas un groupe homogène distinct des RNs classiques, qui possèdent un ligand, mais sont répartis dans l'ensemble de l'arbre phylogénétique présenté Figure 14. De plus, les RNs fixant des ligands ne sont pas regroupés entre eux en fonction de la nature de ce ligand. En effet, les récepteurs NR1 (TR, RAR, VDR, PPAR) sont très proches alors que leurs ligands proviennent de voies de biosynthèse diverses et sont donc complètement différents. Au contraire, RAR et RXR (NR2), membres de sous-familles très différentes, fixent des ligands similaires (respectivement l'acide rétinoïque *all-trans* et *9-cis*). Ces résultats font ainsi apparaître une indépendance entre l'histoire des RNs et leur capacité à reconnaître des ligands.

Contrairement aux récepteurs orphelins, et ce dans la plupart des cas, les RNs des Vertébrés dont les ligands sont connus ne possèdent pas d'homologues chez les Arthropodes. Cette observation a amené plusieurs auteurs (Detera-Wadleigh and Fanning, 1994; Escriva *et al.*, 2000; Laudet, 1997) à proposer le modèle suivant : les premiers membres de la famille des RNs auraient été des orphelins, « primitifs », alors que les RNs à ligands seraient plus « modernes » (plus récents) et dériveraient probablement des précédents.

Le cas du récepteur RXR est particulièrement intéressant. En effet, il fixe l'acide rétinoïque *9-cis* depuis l'apparition des Cnidaires jusqu'aux Vertébrés, sauf chez les Arthropodes pour lesquels son homologue USP, produit du locus génétique *ultraspiracle* chez *Drosophila melanogaster* (Oro *et al.*, 1990), est incapable de fixer ce ligand. Une des explications possibles est la suivante : un événement unique serait survenu très tôt dans l'évolution des Métazoaires, avant la division Cnidaires/animaux triploblastiques, caractérisé par l'acquisition de la propriété de fixer l'acide rétinoïque *9-cis*, suivi d'un deuxième événement définit lui, par une perte spécifique de cette capacité à fixer ce ligand chez les Arthropodes. Les récepteurs EcR, VDR et PXR auraient acquis leur capacité de fixation avant la division Arthropodes/Vertébrés. Les RNs de la sous-famille NR3 (stéroïdiens) sont uniquement présents chez les Chordés, suggérant que l'acquisition de fixation du ligand dans cette sous-famille s'est déroulée très tôt durant l'évolution des Chordés. Par contre, les membres des groupes H et I de la sous-famille NR1 sont présents à la fois chez les Arthropodes et les Vertébrés et fixent des ligands similaires aux stéroïdes, suggérant un gain précoce de la capacité à fixer un ligand durant l'évolution des Métazoaires, probablement l'ancêtre commun des organismes bilatéraux (Figure 15). La plupart des autres RNs à ligands des sous-familles NR1 (TR, PPAR, RAR) et NR3 (ER) auraient donc acquis leur capacité à lier un ligand spécifiquement au niveau des Vertébrés.

Il est intéressant de noter qu'une fois la capacité de fixation d'une classe particulière de ligands acquise, celle-ci n'a pas été (ou très peu) modifiée durant l'évolution.

Toutes ces observations tendent donc à supposer que la super-famille des RNs aurait évolué à partir d'un RN orphelin très ancien apparu avec l'émergence des Métazoaires les plus primitifs (Cnidaires). Les descendants auraient secondairement acquis plusieurs fois la capacité à fixer un ligand, et ce indépendamment dans chaque sous-famille. Cette hypothèse

pourrait expliquer la structure conservée des récepteurs nucléaires en contraste avec la grande diversité de leurs ligands (Escriva *et al.*, 1997; Laudet, 1997).

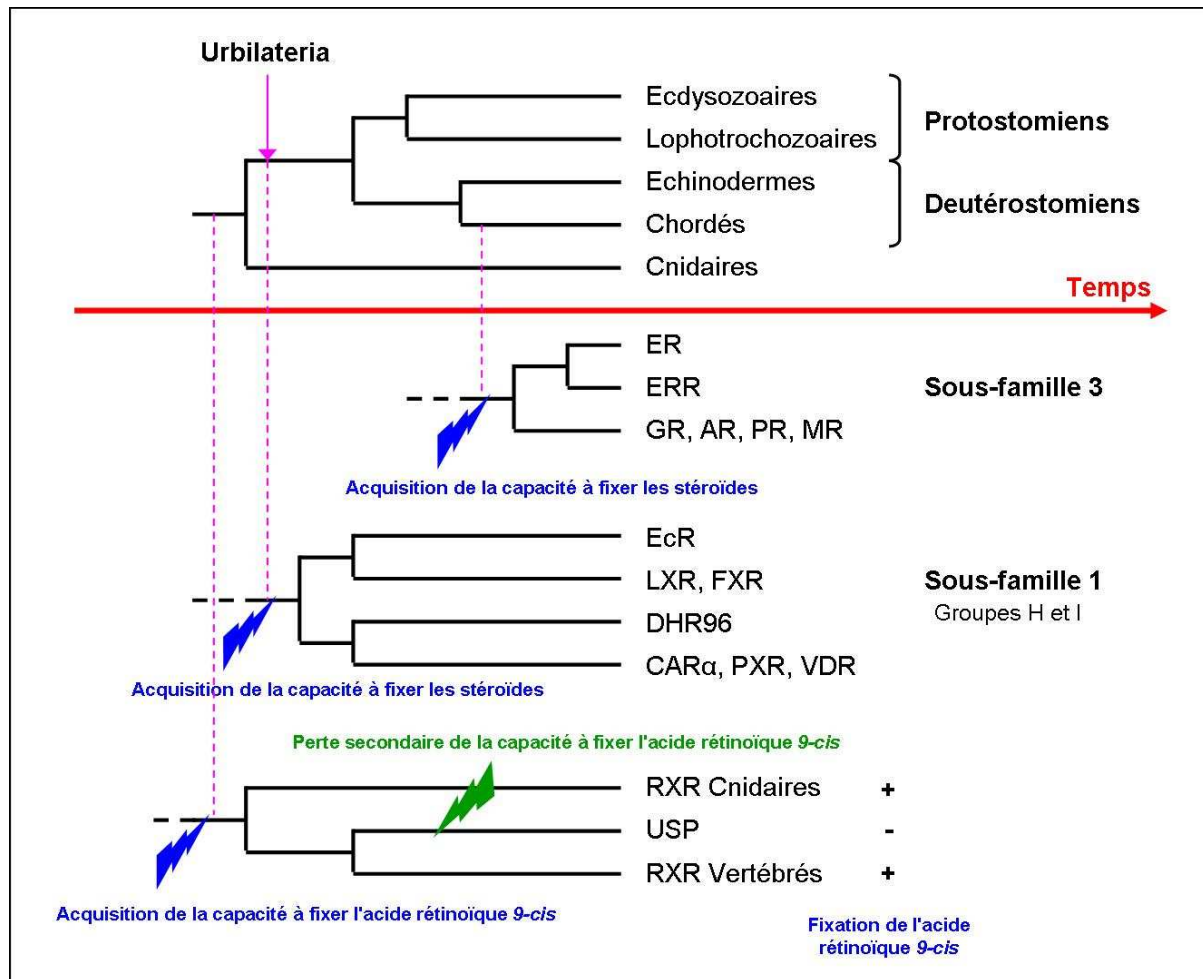


Figure 15 | Schéma illustrant l'acquisition de la capacité de fixation du ligand (adapté de (Escriva *et al.*, 1997)). Arbre phylogénétique schématique des Métazoaires (en haut). Les différentes périodes durant lesquelles la fixation du ligand a été acquise au cours de l'évolution sont indiquées pour les récepteurs NR3, NR1H et NR1I, ainsi que la perte secondaire de la capacité à fixer l'acide rétinoïque 9-cis propre aux arthropodes.

Cependant, qu'en était-il du ligand ? En effet, cette théorie suppose que les deux protagonistes, le ligand et le récepteur, préexistaient avant l'évènement qui a conduit à la capacité de fixer un ligand. Il semble évident que les hormones existaient bien longtemps avant leurs récepteurs. En effet, les stéroïdes sont retrouvés non seulement chez les Métazoaires, mais aussi chez les plantes et les champignons.

Si la théorie « le RN ancestral est un récepteur orphelin » est vraie, comment l'activité transcriptionnelle de ce récepteur était-elle régulée en absence de ligand ? Il est maintenant

admis que les RNs dont les ligands sont connus régulent la transcription des gènes par l'intermédiaire de cofacteurs comme des CoA ou des CoR (Horwitz *et al.*, 1996). La fixation du ligand induit une modification structurale du LBD générant des surfaces d'interaction pour les cofacteurs. Ainsi, pour ces récepteurs, le ligand est nécessaire pour induire le changement conformationnel du LBD et, par là même, pour induire l'activité transcriptionnelle du récepteur.

Les structures cristallographiques des RNs orphelins montrent une structure 3D homologue à celle des récepteurs dits classiques. Cette conservation structurale suggère un mode d'activation similaire à celui des RNs classiques. De plus, on sait aussi que les RNs, comme le récepteur des œstrogènes, peuvent être activés par d'autres processus que la fixation du ligand, notamment par des phosphorylations ou à la suite d'interactions entre protéines. Cela permet d'imaginer que ces premiers récepteurs pouvaient être des facteurs de transcription classiques dont l'activité était régulée par un changement de conformation, induit par phosphorylation par exemple. Sur ce mode ancestral, la capacité à fixer des ligands serait venue s'ajouter plus tardivement de façon à affiner les modes de régulation de ces molécules.

1.4. Mécanisme de la transactivation des RNs et sa régulation

1.4.1. Mécanisme général

Le modèle d'action d'un RN est généralement représenté par un récepteur cytoplasmique inactif en complexe avec des protéines de choc thermique (HSP, *Heat Shock Protein*) (Pratt and Toft, 1997) qui, une fois activé par la fixation de son ligand, se déplace vers le noyau afin d'activer la transcription des gènes cibles. Si ce modèle est vrai pour certains récepteurs stéroïdiens, la plupart des RNs sont localisés dans le noyau et sont souvent liés à l'ADN en absence de leur ligand. Il est également connu que des RNs, en absence de ligand ou en réponse à des ligands antagonistes² ou agonistes, vont réprimer ou

² Un **agoniste** présente des analogies de structure avec un ligand physiologique auquel il peut se substituer et interagit avec un récepteur spécifique en déclenchant les événements biochimiques et cellulaires observés. Un **antagoniste** présente également des analogies de structure avec un ligand physiologique. Il interagit avec un récepteur spécifique et bloque le déclenchement des événements biochimiques et cellulaires.

activer la transcription *via* le recrutement de gros complexes multiprotéiques au niveau du promoteur de leurs gènes cibles.

Les RNs régulent donc l'expression des gènes cibles en réponse à la liaison de l'hormone par de multiples moyens impliquant le changement de compartiment cellulaire, l'interaction avec l'ADN et des cofacteurs, les modifications post-traductionnelles, la communication avec d'autres facteurs de transcription, etc.

1.4.2. Les cofacteurs

Le terme de cofacteur regroupe de nombreuses protéines, généralement sous forme de complexes multiprotéiques, qui interagissent avec les RNs fixés au niveau des régions promotrices des gènes cibles en affectant leur taux de transcription (Glass and Rosenfeld, 2000; McKenna *et al.*, 1999a; McKenna *et al.*, 1999b). Ces cofacteurs sont impliqués à la fois dans la répression et dans l'activation des gènes cibles.

Il existe différentes familles de cofacteurs ayant différentes fonctions : les CoA qui activent la transcription tels que ceux de la famille p160 dont SRC-1, la famille CBP/p300 ; les CoR qui répriment la transcription comme NCOR et SMRT ; les cointégrateurs (TRAP/DRIP) qui facilitent l'interaction des RNs avec la machinerie transcriptionnelle.

La régulation de la transcription est aussi régulée par le remodelage de la chromatine par des cofacteurs à activité ATPase (SWI/SNF). Une chromatine ouverte (euchromatine) facilite l'assemblage de la machinerie transcriptionnelle basale, contrairement à une chromatine condensée fortement ordonnée (hétérochromatine) qui empêche la fixation de la TBP (*TATA Box Binding protein*) et réprime la transcription. La structure de la chromatine peut également être modifiée par la modification covalente des histones du nucléosome par acétylation. Ce phénomène réduit le nombre d'interactions entre l'ADN et le nucléosome en diminuant la charge nette positive des histones. C'est un phénomène réversible qui implique une activité enzymatique de déacétylation. Les deux activités d'histone acétyl-transférase (HAT) et d'histone déacétyl-transférase (HDAC) sont respectivement associées aux CoA et aux CoR.

1.4.2.1. Les coactivateurs

1.4.2.1.1. SRC/p160

La première famille de CoA identifiée est la famille SRC/p160 (Aranda and Pascual, 2001). D'un poids moléculaire d'environ 160kDa, cette famille inclue les CoA SRC-1/NCoA-1, SRC-2/TIF-2/GRIP-1/NCoA2 et SRC-3/pCIP/ACTR/A1B1/TRAM1/RAC3. Ces trois membres montrent une similarité de séquence d'environ 40%, suggérant une certaine redondance fonctionnelle. Dans certains cas, une forte spécificité est observée, à l'exemple de SRC-3 qui augmenterait sélectivement l'activité transcriptionnelle de ER α et non celle de ER β (Suen *et al.*, 1998).

1.4.2.1.2. CBP/p300

Cette famille joue essentiellement le rôle de CoA pour divers facteurs de transcription, coordonne l'assemblage de complexes protéiques additionnels (aspect essentiel à sa fonction de CoA) et acétyle divers substrats (McKenna *et al.*, 1999a; McKenna *et al.*, 1999b; Torchia *et al.*, 1997). CBP (*CREB Binding Protein*) a été initialement caractérisé comme un CoA requis pour l'activation efficace des promoteurs régulés par l'AMPc par CREB (*cAMP-Response Element Binding*). CBP sert également de CoA à CREB en particulier par le recrutement du complexe ARN polymérase II et interagit avec une grande variété de facteurs de transcription incluant AP-1 impliqué dans la prolifération cellulaire (Fronsdal *et al.*, 1998), NF- κ B dans les réponses anti-inflammatoires (Perkins *et al.*, 1997), p53 dans l'apoptose (Avantaggiati *et al.*, 1997; Chen *et al.*, 1997) mais aussi avec les RNs (Chakravarti *et al.*, 1996; Fronsdal *et al.*, 1998; Kamei *et al.*, 1996).

Ces CoA interagissent avec les RNs *via* un motif caractérisé par sa courte séquence signature LXXLL, où L représente une leucine et X un acide aminé quelconque, également appelé NR-box (*Nuclear Receptor-box*). La séquence contenant ce motif adopte une structure en hélice α amphipatique (deux tours d'hélice) qui va venir s'amarrer par l'intermédiaire des leucines dans le sillon hydrophobe créé à la surface du LBD. Des mutations dans ce motif abolissent l'interaction avec les RNs (Heery *et al.*, 1997). Par conséquent, ces motifs sont nécessaires et suffisants pour l'interaction directe avec les RNs *in vivo*.

Il est intéressant de noter que certains CoA possèdent plusieurs motifs LXXLL (7 chez SRC-1), dont la plupart semblent être fonctionnels *in vitro*, en terme de capacité de liaison aux RNs. Cette multiplicité reflète soit la redondance du système, soit la spécificité de l'interaction (Ding 1998). Il existe des arguments en faveur de cette dernière. En effet, des expériences de mutagenèse ont montré l'importance des résidus jouxtant le motif LXXLL, du côté N-ter ou C-ter. Par conséquent, même les CoA homologues tels que SRC-1/SRC-2/SRC-3 ont des préférences pour certains RNs. Ainsi, les récepteurs TR et ER se lient préférentiellement au motif LXXLL II, GR et AR au motif LXXLL III de p160 (Darimont *et al.*, 1998; Ding *et al.*, 1998; Kalkhoven *et al.*, 1998; Voegel *et al.*, 1998). En utilisant des motifs chimères pour l'étude de l'interaction avec ER, il a été montré que les résidus basiques du côté N-ter du motif II étaient nécessaires pour la liaison préférentielle de ce motif (Mak *et al.*, 1999).

Parmi ces CoA se trouvent des complexes qui vont modifier les histones. Certains possèdent une activité histone acétyltransférase (HAT, *Histone Acetyl Transferase*) et vont acétyler les histones, tels que CBP/p300, P/CAF (p300/CBP Associated Factor) et la famille de CoA p160 (Aranda and Pascual, 2001; Glass and Rosenfeld, 2000). D'autres possèdent une activité histone méthyltransférase (HMT, *Histone Methyl Transferase*) et vont méthyler les histones, comme par exemple CARM-1 (*Coactivator-associated aRginine Methyl Transferase*). Ces complexes, en modifiant les histones, desserrent la structure des nucléosomes et rendent les gènes plus accessibles pour la transcription. D'autres complexes possédant des activités de remodelage de la chromatine dépendantes de l'ATP (SWI/SNF) entrent en jeu. Ils induisent une ouverture de la chromatine fortement condensée et produisent un environnement transcriptionnel permissif au niveau du promoteur qui est nécessaire à l'activation de la transcription.

L'activité histone acétyl-transférase intrinsèque de CBP et p300 a été démontrée initialement en utilisant des histones comme substrats. *In vitro*, CBP et p300 acétylent non seulement les histones libres, mais aussi les histones assemblés dans des complexes nucléosomaux suggérant que les nucléosomes peuvent être modifiés de manière similaire *in vivo* (Glass and Rosenfeld, 2000).

Certains CoA se lient à l'AF-1, d'autres à l'AF-2 et d'autres encore aux deux régions AF-1 et 2 (Warnmark *et al.*, 2003). En plus des protéines décrites ci-dessus, fonctionnant comme HAT, HMT, remodeleurs et médiateurs, de nombreux autres CoA avec diverses fonctions ont été décrits (McKenna *et al.*, 1999b). Par exemple, les protéines HMG-1 et 2 (*High Mobility Group*), qui lient l'ADN en induisant une courbure et qui accroissent la liaison à l'ADN, ainsi que l'activité transcriptionnelle des récepteurs aux stéroïdes (Boonyaratanakornkit 1998). Un autre CoA très surprenant a été identifié, SRA (*Steroid Receptor RNA Activator*), qui agit sélectivement comme CoA de l'AF-1 des récepteurs stéroïdiens et qui diffère des autres CoA dans sa nature car c'est un transcrit d'ARN et non une protéine (Lanz *et al.*, 1999).

1.4.2.2. Les cointégrateurs

Le complexe multiprotéique TRAP/DRIP, qui interagit respectivement avec TR et VDR d'une manière dépendante du ligand, est composé d'environ 16 protéines dont la taille est comprise entre 70 et 230 kDa. Ce complexe est recruté par la fonction AF-2 en réponse à la liaison du ligand agoniste, par le biais d'une seule sous-unité (TRAP220/DRIP205) possédant un motif identique à celui trouvé chez les CoA (Aranda and Pascual, 2001). Différentes sous-unités de ce complexe sont homologues aux protéines du Médiateur (Gu *et al.*, 1999; Rachez *et al.*, 1998), un complexe qui s'associe, en compagnie de protéines SRB (*Suppressor of carboxyterminal Repeat domain B*), à l'ARN polymérase II. Il semblerait donc que le complexe TRAP/DRIP soit capable de recruter l'ARN polymérase II. Le complexe TRAP/DRIP pourrait fonctionner indépendamment ou en séquence avec le système p160 et CBP.

1.4.2.3. Les corépresseurs

Plusieurs CoR ont été mis en évidence, notamment SMRT (*Silencing Mediator for Retinoic acid and Thyroid hormone receptors*) (Chen and Evans, 1995) et N-CoR (*Nuclear CoRepressor*) (Horlein *et al.*, 1995). Ils interagissent en absence de ligand avec RAR et TR associés à RXR sur l'ADN et répriment la transcription basale. N-CoR et SMRT servent de CoR à de nombreux RNs tels Rev-erba, Rev-erbβ, COUP-TF, et ER et PR liés à des antagonistes.

Ils exercent leur activité répressive sur des RNs sous forme *apo via* le recrutement d'un complexe à activité histone déacétylase (HDAC) qui conduit à la condensation de la chromatine, phénomène inhibant le recrutement de facteurs de transcription et du complexe ARN polymérase II (Kadonaga, 1998) et ainsi l'expression des gènes cibles. Les différents CoR et les différentes HDACs semblent former plusieurs complexes, dont certains possèdent aussi des activités de remodelage de la chromatine dépendant de l'ATP.

Ces CoR se fixent aux RNs sous forme *apo* par le biais de motifs conservés de type LXX(I/H)XXX(I/L) (où L représente la leucine, I l'Isoleucine, H l'Histidine et X un acide aminé quelconque) qui ressemblent au motif LXXLL retrouvé dans les CoA. Cependant, la séquence contenant le motif d'interaction des CoR forment trois tours d'hélice α et est donc plus étendue (un tour d'hélice supplémentaire) que celle des CoA (Hu and Lazar, 1999; Nagy *et al.*, 1999; Perissi *et al.*, 1999).

D'autres protéines différentes de SMRT et N-CoR, qui fonctionnent comme CoR des RNs, ont été identifiées : Alien (Dressel *et al.*, 1999) interagit avec TR, COUP-TF-1 de la drosophile, EcR et DAX-1, mais pas avec RAR ou RXR, ni avec USP de la drosophile ; RIP140 (Cavailles *et al.*, 1995) ; REA (*Estrogen Receptor Activity*) (Montano *et al.*, 1999) ; SUN-CoR (*Small Ubiquitous Nuclear CoRepressor*) (Zamir *et al.*, 1997) ; Zip 67 (*Zinc finger protein 67kDa*) (Borud *et al.*, 2003) ; ainsi que deux CoR, Hr (Hairless) et LCoR (*Ligand dependent CoRepressor*), qui présentent un mécanisme d'action particulier. En effet, Hr a été identifié comme CoR du récepteur orphelin constitutivement actif ROR α (*Retinoic acid receptor related Orphan Receptor α*). Hr présente un mode de répression particulier de ROR α , puisqu'il interagit avec ROR α par le biais de deux motifs LXXLL qui sont considérés généralement comme motifs d'interaction des CoA. De plus, l'hélice H12 de ROR α semble importante pour cette interaction avec Hr (Moraitis *et al.*, 2002). Quant à LCoR, c'est un CoR qui recrute les récepteurs liés à des agonistes à travers un motif LXXLL (Fernandes *et al.*, 2003).

1.4.3. La transactivation

Schématiquement, la transactivation est un processus séquentiel en trois étapes. En absence de ligand, le LBD de certains RNs (à l'exception des récepteurs stéroïdiens) est lié à des CoR tels que N-CoR ou SMRT qui recrutent des complexes transcriptionnels exerçant une activité de déacétylation des histones (HDAC). Cette déacétylation conduit à une

condensation de la structure de la chromatine dont la résultante est la répression de l'expression du gène cible (Figure 18A). Les CoR s'arriment à un sillon hydrophobe présent à la surface du LBD généré par les hélices H3 et H4 (Figure 16) *via* leur motif LXX(I/H)XXX(I/L) similaire au motif LXXLL des CoA (Jepsen and Rosenfeld, 2002). La séquence contenant le motif des CoR se fixe sur un sillon hydrophobe formé à la surface du LBD qui est partiellement chevauchant avec celui des CoA (Xu *et al.*, 2002).

La fixation d'un ligand agoniste induit des effets allostériques complexes, décrits Figure 9, amenant l'hélice H12 à adopter une position radicalement différente, sur le coeur du LBD, qui désorganise le sillon hydrophobe et conduit à la dissociation des CoR. Ce nouveau positionnement de l'hélice H12 du récepteur *holo* génère de nouvelles surfaces d'interaction et permet aux CoA à activité acétylase (HAT) d'interagir (Figure 18B), par leur court motif LXXLL, au niveau d'une crevasse hydrophobe formée par les hélices H3, H4 et H12 (Figure 16). L'échange dynamique, dépendant du ligand, des CoR pour les CoA sert d'interrupteur et permet de passer de la répression à l'activation de la transcription.

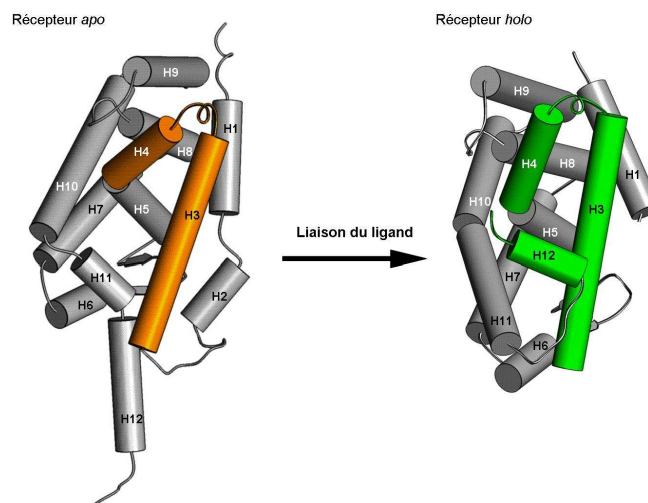


Figure 16 | Surfaces de reconnaissance des CoR et des CoA. Récepteur *apo* : en absence de ligand, les CoR sont liés au LBD d'un RN *apo* (code PDB : 1LBD) au niveau des hélices H3 et H4 (orange). Récepteur *holo* : la liaison du ligand agoniste induit un changement conformationnel de l'hélice H12 qui la positionne sur le coeur du LBD. Cette nouvelle orientation de H12 chasse les CoR et permet le recrutement des CoA au niveau des hélices H3, H4 et H12 (vert) du RN *holo* (code PDB : 1MVC).

La séquence contenant le motif LXXLL des CoA adopte une structure en courte hélice α amphipatique qui va venir s'amarrer au niveau du sillon hydrophobe précédemment cité. Une réelle complémentarité de surface existe entre la structure du peptide et celle du sillon. Le peptide est, de plus, maintenu par des interactions électrostatiques formant une « pince électrostatique ». Cette « pince » est constituée de deux résidus chargés, strictement conservés dans la super-famille des RNs, le premier localisé sur l'hélice H3 (K284 pour hRXR α) et le deuxième sur l'hélice H12 (E453 pour hRXR α) (Figure 17). Cette crevasse hydrophobe est similaire, mais pas identique, à celle impliquée dans l'interaction avec les CoR et joue un rôle décisif dans l'interaction avec les CoA.

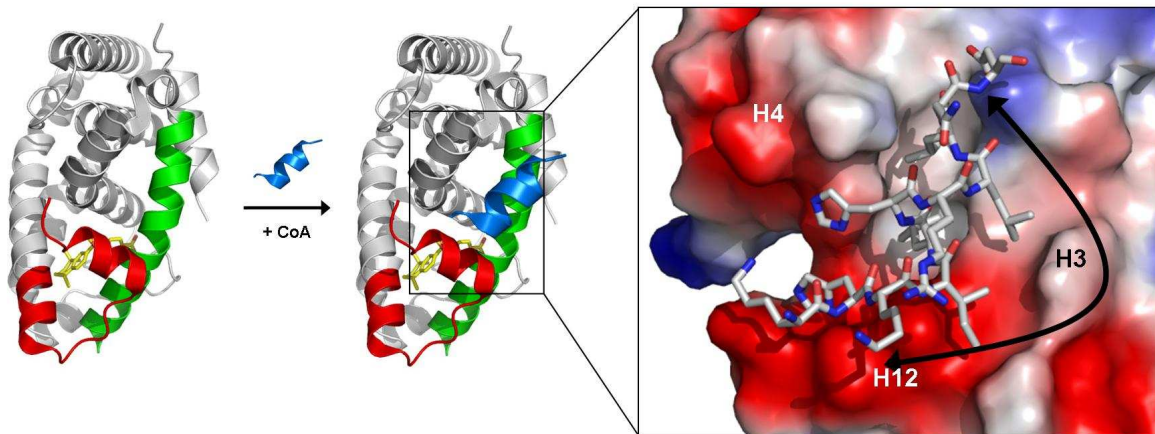


Figure 17 | La « pince électrostatique ». Cette « pince » est constituée de deux résidus chargés qui interagissent avec les deux extrémités du peptide CoA en augmentant ainsi l'interaction du CoA avec le LBD. Les deux extrémités de la « pince » sont indiquées par la flèche noire. Les surfaces correspondantes aux hélices H3, H4 et H12 sont indiquées.

Des complexes remodelant la chromatine vont générer une structure ouverte de la chromatine au niveau du promoteur. Une fois la chromatine décondensée, les RNs recrutent l'ARN polymérase II *via* son association avec le complexe TRAP220/DRIP205, possédant aussi un motif LXXLL, et par dissociation du complexe HAT suite à son acétylation. Le Médiateur peut alors se fixer et recruter la machinerie de transcription basale qui permet l'initiation de la transcription (Figure 18C).

Le changement conformationnel du LBD, induit par la fixation d'un agoniste, est donc une manifestation structurale de la fonction de transactivation d'un RN. Il faut noter que l'effet allostérique du ligand sur la conformation du LBD peut aussi moduler l'activité de la fonction AF-1 du domaine A/B, par exemple, par une communication croisée

intramoléculaire entre les domaines N et C-ter. Les mécanismes structuraux de ce processus restent une énigme, puisque la structure d'un RN entier n'a pas encore été résolue, mais dans le cas du récepteur à l'androgène (AR), des résultats concernant une interaction directe entre les régions N et C-ter ont été obtenus (He and Wilson, 2002).

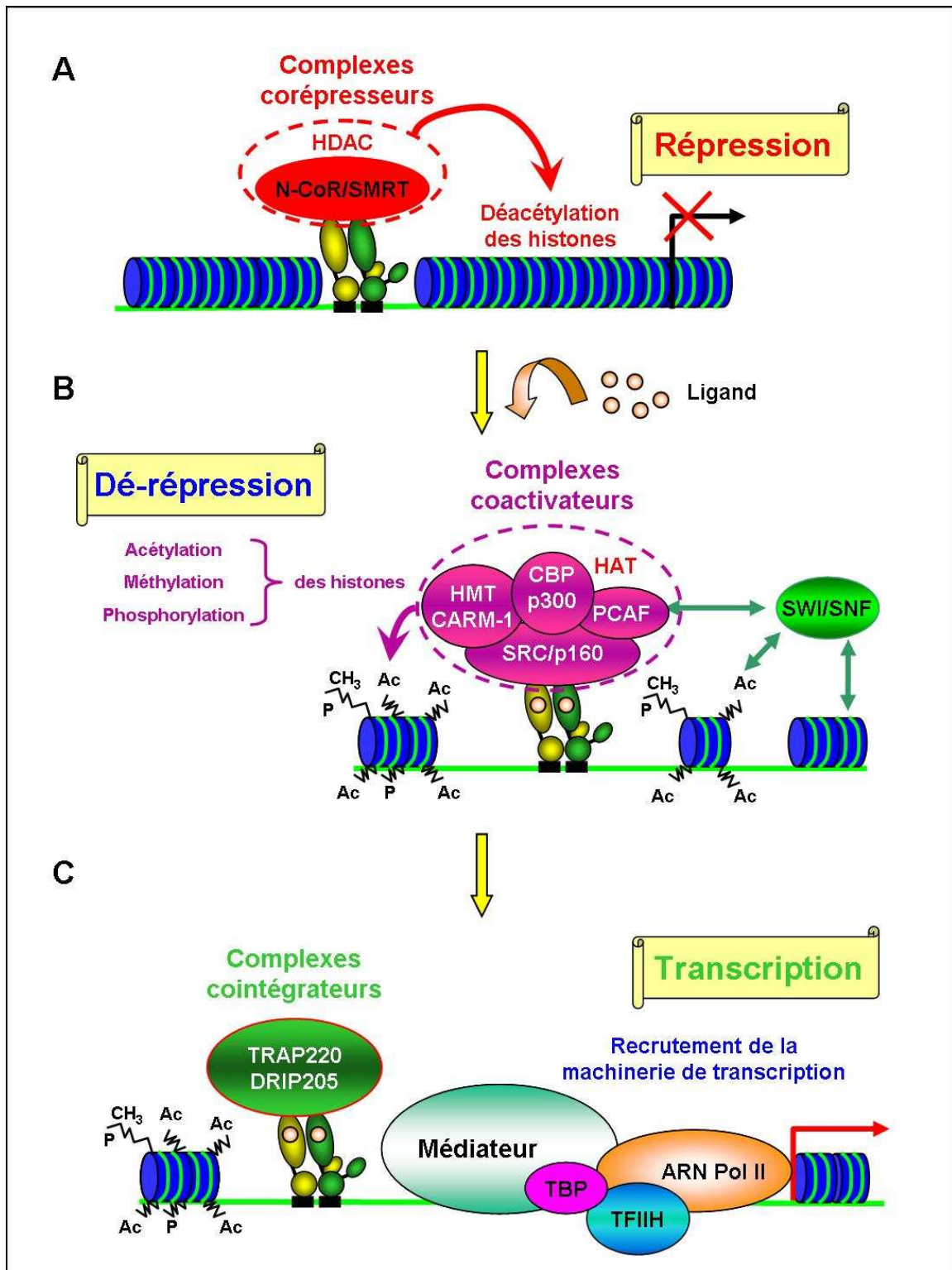


Figure 18 | Mécanisme séquentiel de l'action des RNs (adaptée de (Bastien and Rochette-Egly, 2004)). La répression active (A) implique le recrutement d'un complexe CoR exerçant une activité déacétylase (HDAC). Suite à la fixation du ligand (B), ce complexe CoR est dissocié et remplacé par un complexe CoA à activité HAT, HMT ou ATPase, dont la conséquence est la décondensation de la chromatine (dé-répression). Dans la troisième étape (C), le CoA se dissocie suite à son acétylation qui diminue sa capacité à interagir avec les RNs ou suite à sa dégradation par le protéasome. Le complexe TRAP220/DRIP205 le remplace, permettant le recrutement de la machinerie transcriptionnelle et ainsi la transcription.

1.5. Les bases moléculaires de l'antagonisme

Les RNs sont impliqués dans la régulation de fonctions physiologiques essentielles et sont des cibles pharmacologiques de choix pour la conception de molécules modifiant leurs activités. Ainsi, les récepteurs aux éicosanoïdes (PPAR) sont activement étudiés pour la mise au point de molécules régulant les métabolismes lipidiques et glucidiques. Des applications médicales concernent le traitement de l'obésité ou du diabète. C'est également le cas des récepteurs aux stéroïdes (ER et PR) étudiés pour la mise au point de composés régulant les réponses physiologiques sous le contrôle des hormones œstrogènes et progestatives. Des applications médicales concernent notamment la contraception et les traitements de certains cancers (cancer du sein). Les rétinoïdes sont aussi utilisés dans le traitement de certaines affections de la peau (psoriasis ou acné) et certains cancers.

Des composés agonistes et antagonistes sont disponibles pour divers récepteurs et la connaissance des structures des LBDs permet de rationaliser la conception de nouveaux composés, en analysant les mécanismes d'action des agonistes et des antagonistes. Par agoniste, on définit un composé agissant sur un récepteur donné et provoquant une réponse similaire (d'une amplitude plus ou moins importante) à celle occasionnée par un ligand naturel. Au contraire, un antagoniste désigne un composé agissant sur le même récepteur en s'opposant à la réponse occasionnée (soit en la diminuant, soit en l'annulant) par le ligand naturel ou un agoniste.

La résolution des structures du LBD d'un même récepteur lié à des composés agonistes et antagonistes a permis de comprendre les bases moléculaires de l'agonisme et de l'antagonisme dans le cas des RNs. Historiquement, les structures qui ont permis d'élucider un tel mécanisme sont : la structure de ER α lié à l'œstradiol (le ligand agoniste naturel) ou au raloxifène (un antagoniste) (Brzozowski *et al.*, 1997); la structure de ER α lié au Di-Ethyl-Stilbestrol (un agoniste) ou au 4-hydroxy tamoxifène (un antagoniste) (Shiau *et al.*, 1998) et la structure de ER α lié à la génistéine (un agoniste partiel).

La Figure 19 présente les structures *holo* du LBD de hER α lié à un agoniste, le 2-(but-1-enyl)-17 β -oestradiol (Code PDB : 2G5O) ou à un antagoniste, le 4-hydroxy-tamoxifène (Shiau *et al.*, 1998) (Code PDB : 3ERT) avec, en comparaison, la structure du récepteur hRXR α sous forme *apo* (Bourguet *et al.*, 1995) (Code PDB : 1LBD). Il apparaît qu'agonistes et antagonistes se lient dans la même poche, mais selon des modalités

différentes. Les principales différences observées concernent la position occupée par l'hélice H12 portant le domaine de transactivation AF-2.

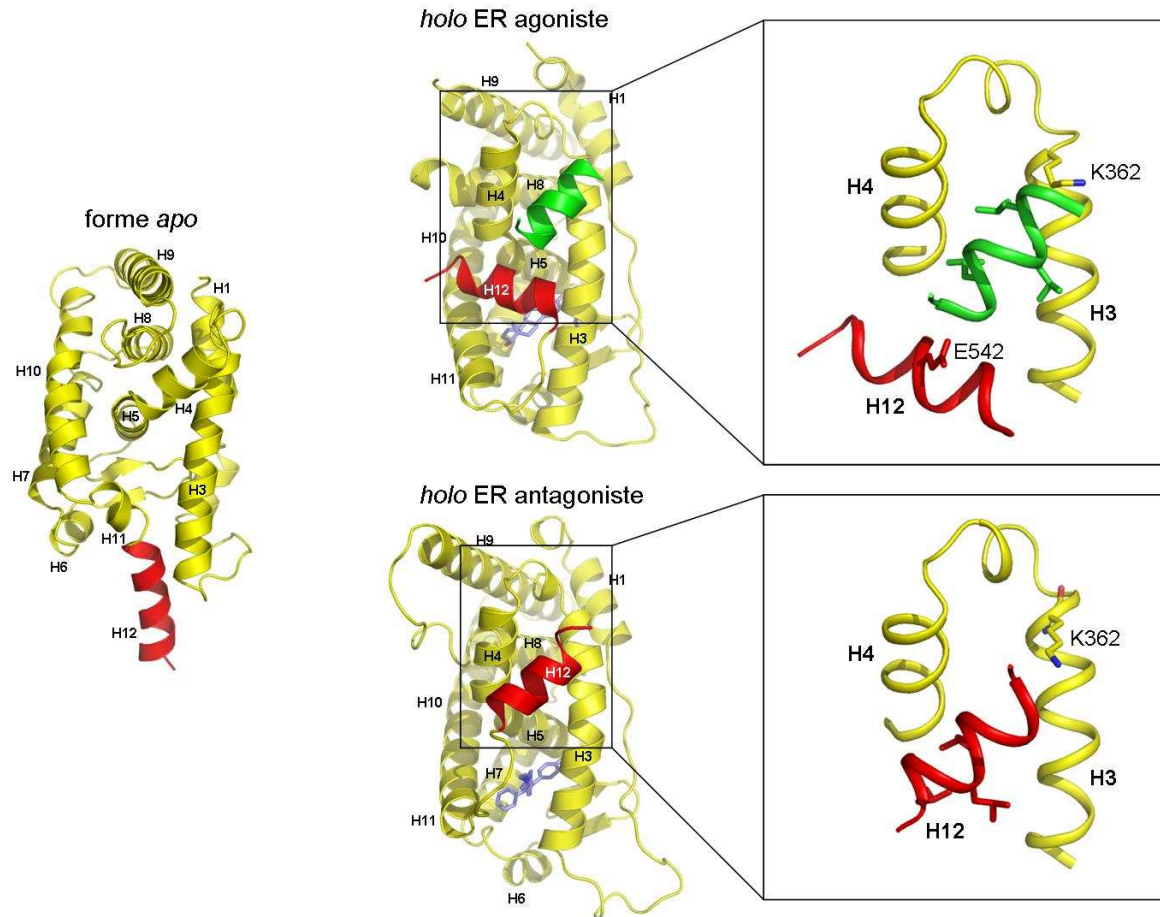


Figure 19 | Effet des ligands agonistes et antagonistes sur le positionnement de l'hélice H12. Comparaison des structures tridimensionnelles des LBDs de : RN sous forme *apo* ; de ER α complexé à un ligand agoniste, l'œstradiol, et le peptide (vert) portant le motif LXXLL du CoA GRIP1 (Code PDB : 2G5O) ; de ER α complexé à un ligand antagoniste, le 4-hydroxytamoxifène (Code PDB : 3ERT). Il est intéressant de noter que l'hélice H12 dans la structure antagoniste occupe la même position que celle occupée par le motif d'interaction LXXLL du CoA dans la structure agoniste.

La conformation occupée par H12 est identique que l'on considère les complexes avec le ligand naturel ou les divers agonistes. H12 occupe alors une surface délimitée par les hélices H3, H5 et H10/H11. La fixation de ligands antagonistes déplace l'hélice H12 de sa position dite « agoniste » à sa position dite « antagoniste » en générant une contrainte stérique. En effet, ces ligands portent des chaînes latérales encombrantes, qui débordent de la poche de fixation du ligand, et pointent vers l'hélice H12. Celle-ci occupe alors une autre surface délimitée par les hélices H3 et H4.

Ces structures confirment l'existence du récepteur sous trois conformations distinctes : une conformation *apo* (avec l'hélice H12 en conformation étendue) et deux conformations *holo*, où l'hélice H12 adopte deux conformations distinctes selon qu'un agoniste ou un antagoniste soit fixé. La conformation agoniste est définie par une seule et unique position de H12 et est indispensable à l'établissement d'interactions ultérieures productives entre le récepteur et un CoA.

1.6. Modifications post-traductionnelles des RNs

Le niveau d'expression, ainsi que la fonction des protéines, doivent être constamment ajustés et sont donc finement régulés. Concernant l'expression des protéines, l'initiation de la transcription, de même que le traitement de l'ARN, son transport puis la traduction et le repliement des protéines sont soumis à une stricte régulation. L'activité des protéines est ensuite contrôlée par la régulation de leur localisation et de leur demi-vie. De plus, la fonction des protéines est aussi modulée rapidement par des modifications post-traductionnelles, qui consistent en l'attachement aux protéines de divers groupements : méthyle, phosphoryle, acétyle, etc. Les RNs, mais aussi leurs corégulateurs et les histones, peuvent subir diverses modifications post-traductionnelles qui vont modifier leur activité : ubiquitination, sumoylation, acétylation, méthylation et phosphorylation (Baek and Rosenfeld, 2004).

La phosphorylation par différentes kinases de résidus dans les domaines A/B, C et E a été démontrée. En règle générale, la (ou les) phosphorylation(s) coopère(nt) avec l'hormone pour atteindre le niveau maximal d'activation transcriptionnelle. Cependant, il a été montré qu'elle pouvait atténuer, et même terminer, la réponse au ligand en excluant les récepteurs du complexe de transcription, en réduisant leur affinité de liaison au ligand ou en induisant leur dégradation par la voie ubiquitine-protéasome (Rochette-Egly, 2003).

L'ubiquitination est la conjugaison à un substrat de l'ubiquitine, un polypeptide de 76 résidus. La mono-ubiquitination est impliquée dans la régulation transcriptionnelle, alors que la poly-ubiquitination est impliquée dans les voies de dégradation des protéines. L'ubiquitination de plusieurs RNs a été reportée, et notamment ER, AR, PPAR α , GR, RAR γ , RXR α et TR sont régulés par le système ubiquitine-protéasome (Baek and Rosenfeld, 2004).

La sumoylation est une modification identifiée récemment qui consiste en la conjugaison d'un petit polypeptide appelé SUMO (*Small Ubiquitine-related MOdifier*) à des résidus lysines spécifiques des protéines cibles et qui produit divers effets. Le récepteur aux androgènes est sumoylé et cette sumoylation est importante pour son activité de répression (Nishida and Yasuda, 2002). Au contraire, la sumoylation du récepteur aux glucocorticoïdes augmente son activité transcriptionnelle et régule sa stabilité (Le Drean *et al.*, 2002).

L'acétylation de plusieurs RNs tels que AR et ER α a été reportée. L'acétylation de ER α contribue à l'atténuation de l'activité transcriptionnelle (Wang *et al.*, 2001), alors que dans le cas de AR elle est importante pour la transactivation dépendante du ligand (Fu *et al.*, 2003).

Comme on vient de le voir, l'activité des RNs peut être finement régulée directement, mais aussi indirectement *via* des modifications de leurs corégulateurs et de leur niveau d'expression. De plus, elle intègre aussi les informations provenant d'autres voies de signalisation. Il a été suggéré que ces modifications influencent la localisation cellulaire, l'activité enzymatique et la stabilité des protéines cibles et pouvaient donc être importante dans la régulation temporelle du recrutement séquentiel de ces différents corégulateurs.

L'activation de la transcription d'un gène implique le recrutement orchestré des composants de la machinerie transcriptionnelle basale et de facteurs intermédiaires, concomitant avec une altération locale de la structure de la chromatine par des modifications post-traductionnelles de la queue des histones et du remodelage du nucléosome.

De récents travaux (Hager *et al.*, 2006) ont apporté, grâce au développement de nouvelles techniques d'études destinées à la caractérisation de facteurs de transcription en cellules vivantes et en temps réel, une nouvelle vision concernant cette cinétique. Il est apparu que, contrairement à ce que l'on pensait, le récepteur effectuait des « allers-retours » rapides (environ 1 min en contact avec l'ADN) et continus (toutes les 5 min) sur la chromatine. Ce comportement dynamique offre des possibilités intéressantes dans le cadre de communications croisées entre les récepteurs, constamment disponibles, et certaines voies de signalisation qui peuvent rapidement modifier la réponse d'un récepteur donné.

1.7. Communication croisée avec d'autres facteurs de transcription

En plus de leur mode d'action classique par liaison à des séquences spécifiques d'ADN (HRE, *Hormone Response Element*), les RNs peuvent également réguler l'expression de gènes ne possédant pas dans leur région régulatrice de séquences HRE mais un site de fixation pour un autre facteur de transcription tel que le site AP-1 où se lient les facteurs de transcription Fos et Jun. Dans ce cas, le récepteur, lié à l'hormone, ne se lie pas directement à l'ADN mais module la transcription en interagissant avec d'autres protéines par exemple un autre facteur de transcription.

1.8. Localisation cellulaire

La localisation cellulaire du récepteur en absence de ligand est un point important à souligner dans le mode de fonctionnement de ces régulateurs transcriptionnels activés par la fixation d'un ligand.

1.8.1. Localisation cytoplasmique

En absence de leurs ligands, certains récepteurs stéroïdiens, tels GR, sont séquestrés dans le cytoplasme (Pratt and Toft, 1997); (Picard, 2006). Ils sont complexés à des protéines chaperonnes, notamment des protéines de choc thermique (HSP, *heat shock protein*), en particulier HSP70 et HSP90 pour GR (Dittmar *et al.*, 1997) et PR (Kosano *et al.*, 1998). HSP70 et HSP90 sont des protéines à activité ATPase. Leurs activités sont régulées par diverses protéines co-chaperonnes : HSP40, p23, Hop (*HSP70/HSP90 organizing protein*) et Hip (*HSP70 interacting protein*). HSP40 est une protéine co-chaperonne de HSP70 qui stimule son activité ATPase et facilite la fixation du prochain composant du complexe, Hop. Hop permet de mettre en contact HSP70 et HSP90. La protéine p23 se lie à HSP90 et stabilise le complexe ATP-HSP90-GR. Des immunophilines comme FKBP51 et FKBP52 sont aussi impliquées, en particulier FKBP52, qui intervient dans l'import nucléaire de GR (Grad and Picard, 2007) (Figure 20).

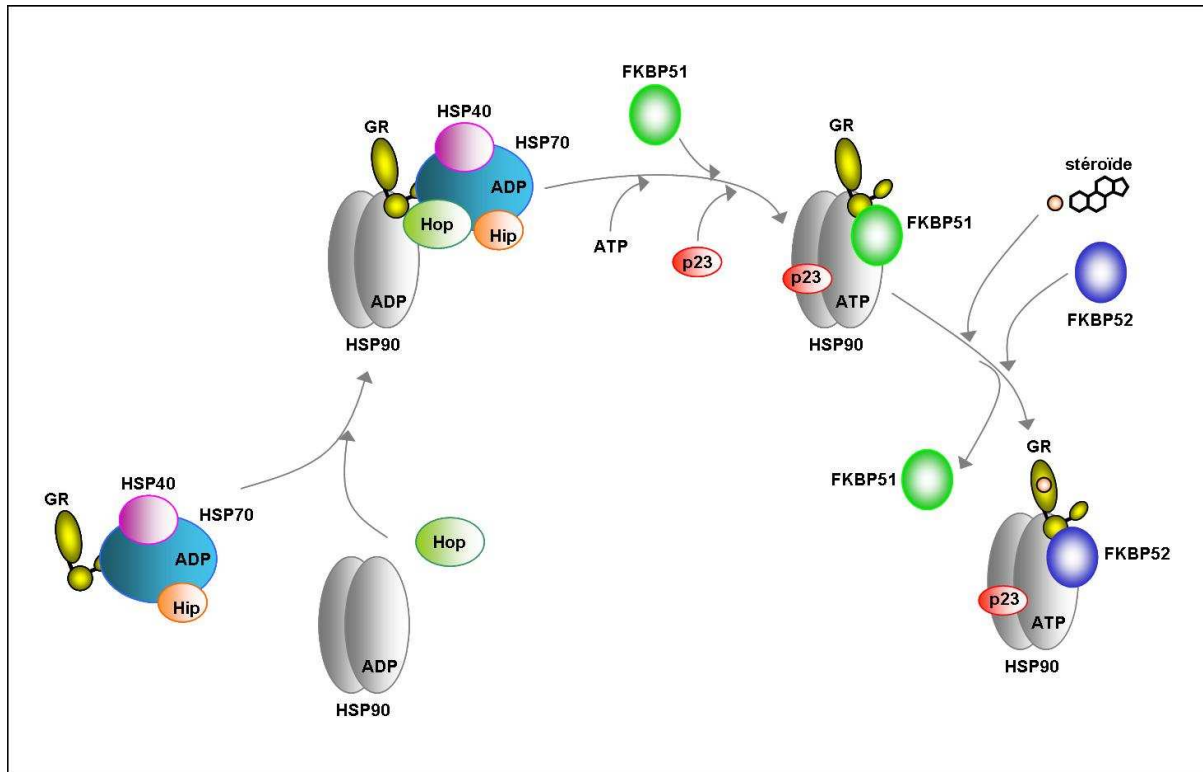


Figure 20 | Les différentes étapes de « chaperonnage » des récepteurs stéroïdiens (adaptée de (Grad and Picard, 2007)). Le récepteur stéroïdien *apo* s'associe en premier lieu avec HSP70 et HSP40. L'interaction de ces deux protéines maintient un repliement correct du récepteur. Cette étape dépend de l'activité ATPase d'HSP70 et conduit au recrutement de Hip et de Hop au niveau d'HSP70. Hop lie HSP90 de manière indépendante et permet à HSP90 de contacter le récepteur. En présence d'ATP, p23 stabilise HSP90 au niveau du LBD, maintenant la conformation dans un état de haute affinité pour le ligand. Puis Hop est déplacée par une immunophiline. En absence de ligand, le complexe GR-HSP90 interagit préférentiellement avec l'immunophiline FKBP51. Après fixation de l'hormone, elle est remplacée par FKBP52.

Les HSP facilitent le repliement du LBD du RN (Nathan and Lindquist, 1995; Pratt and Toft, 1997) et interviennent à différents niveaux : dans la capacité de fixation du ligand (Picard *et al.*, 1990); (Pratt and Toft, 1997) ; dans la régulation de la liaison à l'ADN où HSP90 est capable de dissocier les récepteurs ER (Sabbah *et al.*, 1996) et GR (Kang *et al.*, 1999) liés à l'ADN ; dans la translocation nucléaire où les chaperonnes influencent le transport entre le cytoplasme et le noyau des récepteurs stéroïdiens (DeFranco, 1999) ; dans la dégradation et le recyclage où le traitement à la geldanamycine, un inhibiteur de l'activité ATP-asique d'HSP90 qui provoque la perte de la capacité de liaison de GR et de PR et qui pourrait être attribuée à un mécanisme de dégradation impliquant le protéasome 20S (Smith *et al.*, 1995; Whitesell and Cook, 1996).

L'interaction RN-chaperonnes stabilise le récepteur en absence de ligand, en le maintenant sous une forme monomérique latente stable. La fixation du ligand au récepteur permet la translocation vers le noyau en particulier grâce à FKBP52, qui interagit avec la dynéine et qui lie GR aux microtubules (Czar *et al.*, 1994). Alternativement, GR, seul ou complexé avec FKBP52, pourrait se lier à des importines pour migrer vers le noyau (Figure 21). L'*holo*-récepteur dimérique active la transcription des gènes cibles, après s'être fixé sur les séquences d'ADN spécifiques situés dans le promoteur et recruté des CoA (Grad and Picard, 2007; Picard, 2006; Pratt and Toft, 1997).

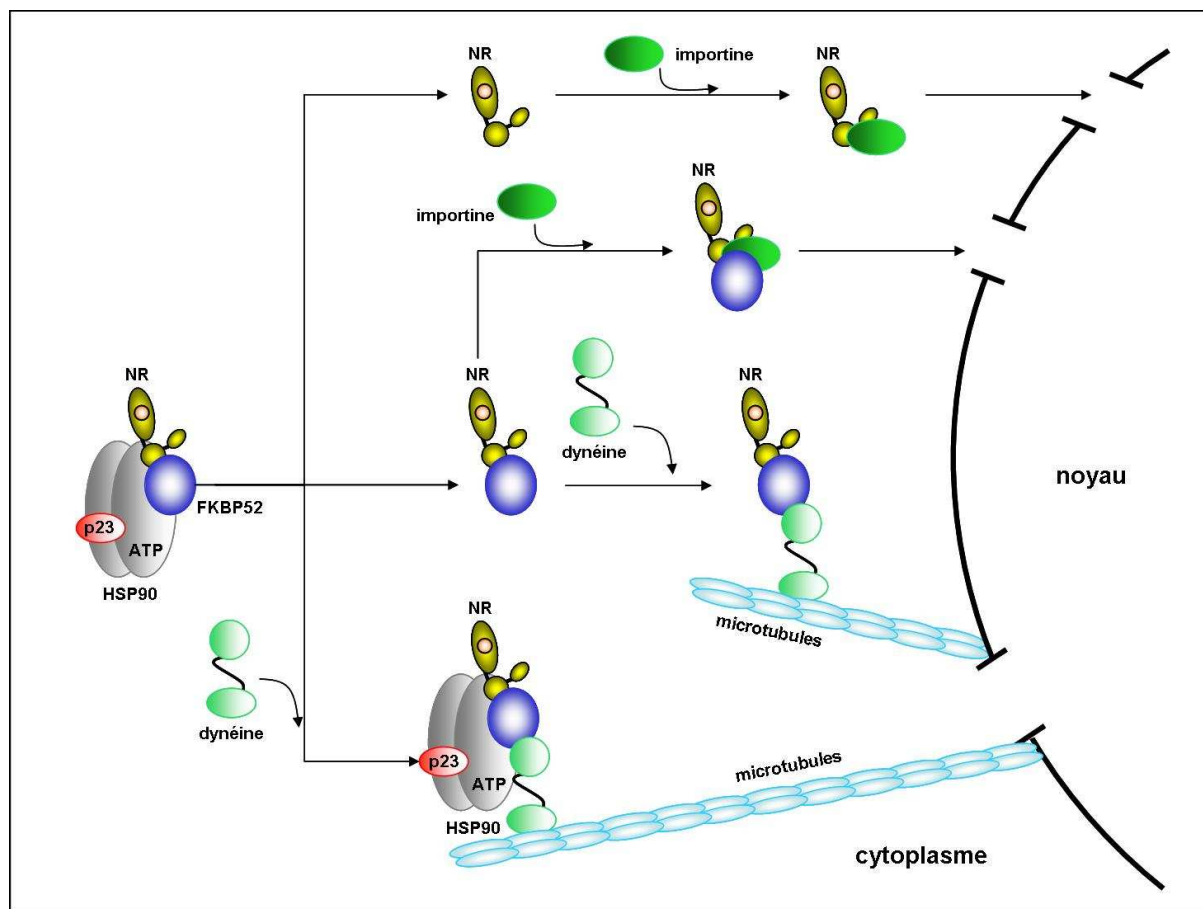


Figure 21 | Transport du récepteur stéroïdien vers le noyau (adaptée de (Grad and Picard, 2007)). Après fixation du ligand, l'immunophiline FKBP51 est remplacée par FKBP52. FKBP52 s'associe avec la dynéine qui permet au complexe GR-HSP90 d'interagir avec les microtubules pour le transport vers le noyau. De manière alternative, GR, seul ou en complexe avec FKBP52, pourrait se lier à des importines permettant ainsi le passage du pore nucléaire.

1.8.2. Localisation nucléaire

A la différence des récepteurs stéroïdiens, la localisation des récepteurs des rétinoïdes (RXR et RAR) ou de l'hormone thyroïde (TR) est nucléaire en absence de ligand. En effet, contrairement à celui des récepteurs aux hormones stéroïdes, le LBD des récepteurs de l'acide rétinoïque, ou celui des TR et VDR, n'interagit pas avec les complexes *HSP*, en particulier HSP90 (Dalman *et al.*, 1991), responsables de la séquestration des récepteurs stéroïdiens dans le cytoplasme en absence de ligand. Ils sont déjà fixés sur les séquences cibles d'ADN, aussi bien sous forme d'homodimères que d'hétérodimères sous forme *apo* en complexe avec des CoR tels que SMRT et NCoR réprimant l'activité transcriptionnelle. En absence de ligand, ils répriment la transcription des gènes cibles. La fixation du ligand (agoniste) modifie la structure des récepteurs et active alors la transcription des gènes cibles en recrutant des CoA et en relarguant les CoR (Bhattacharyya *et al.*, 1997), (Zhu *et al.*, 1998). De plus, le fait qu'ils puissent agir en tant que répresseurs de l'activité transcriptionnelle en absence de ligand, implique qu'ils soient présents dans le noyau indépendamment de la présence d'acide rétinoïque. La répression des gènes par les récepteurs des rétinoïdes en absence de ligand peut dicter le destin cellulaire. Elle est notamment essentielle lors de la formation de la tête chez le xénope et du développement du squelette chez la souris (Weston *et al.*, 2002).

1.9. Objectifs de la thèse

La plupart des RNs fonctionnent comme dimères, soit sous la forme d'homodimères soit sous la forme d'hétérodimères. Du point de vue fonctionnel, le LBD est le domaine le plus important. Outre son rôle dans la fixation du ligand, il est également impliqué dans la dimérisation (homodimère ou hétérodimère avec RXR) des récepteurs ainsi que dans la régulation de la transcription dépendante du ligand. Le LBD est le contributeur majeur à la stabilité du dimère par des contacts inter-protéines. Grâce aux nombreuses structures de LBDs disponibles, ces points de contact sont connus. Le LBD possède, en plus, des surfaces d'interaction pour des cofacteurs, soit des CoA soit des CoR. Cependant, les mécanismes moléculaires qui contrôlent cette association fonctionnellement importante ne sont pas encore élucidés.

Le sujet de ma thèse a consisté à établir les bases moléculaires de la discrimination entre ces deux états, homo- et hétérodimère en se basant sur une approche multidisciplinaire combinant la bioinformatique, la biologie moléculaire, structurale et cellulaire sur trois systèmes tests qui ont été choisis pour leur comportement oligomérique différent : $ERR\gamma$ qui est un récepteur homodimérique constitutivement actif (sans ligand), $RAR\alpha$ qui hétérodimérise avec $RXR\alpha$ et $RXR\alpha$ qui peut lui même homodimériser.

Matériel et méthodes

Chapitre 2 - Bioinformatique

2.1. Ressources informatiques et bioinformatiques

A l'ère post génomique, et dans un contexte où les études à haut débit se généralisent, la communauté scientifique est confrontée à une production importante de données dont la diversité, en termes de types et de propriétés en complexifie l'analyse et la compréhension. La solution vient peut être des Champollion du 21^{ème} siècle : les bioinformaticiens. Fouiller les bases de données, aligner des séquences, prédire les structures tridimensionnelles des protéines, analyser les expressions du génome, etc. autant de tâches qui requièrent des connaissances et des compétences non seulement en biologie, en biochimie, en mathématique, en statistique, voire en physique et en linguistique, mais aussi en informatique. Dans ce contexte, l'analyse bioinformatique doit faire appel à des banques de données régulièrement mises à jour, à l'automatisation de certains traitements ainsi qu'au stockage organisé des résultats obtenus. Ces contraintes résument les trois défis bioinformatiques dans l'analyse des séquences protéiques : gérer, traiter et visualiser. En d'autres termes : pérenniser, valoriser et exploiter.

Au cours de cette thèse, nous avons bénéficié de l'infrastructure existante au Laboratoire de Biologie et de Génomique Structurales (LBGS) et des services fournis par la *Plate-forme de Bioinformatique de Strasbourg* (BIPS) (<http://bips.u-strasbg.fr/>). BIPS est une plate-forme à haut débit localisée à l'IGBMC et dédiée à l'analyse de séquence ainsi qu'à la génomique comparative et structurale. Elle a été labellisée *Plate-forme Nationale RIO* en 2003 et constitue la *Plate-forme de Bioinformatique du Génopôle Grand-Est « du Gène au Médicament »*.

Dans ce chapitre, les équipements informatiques utilisés, ainsi que le savoir-faire bioinformatique du laboratoire, seront décrits succinctement. Nous présenterons ensuite les principales banques de données sur lesquelles nos études se sont appuyées, ainsi que les moyens mis à notre disposition pour les interroger. Nous terminerons ce chapitre en nous intéressant plus précisément aux outils que nous avons utilisés dans le cadre des études relatives aux alignements multiples de séquences protéiques d'une part et aux structures tridimensionnelles des protéines d'autre part.

2.1.1. Équipement et ressources informatiques

L'équipement informatique et son bon fonctionnement tiennent une part importante dans le déroulement de travaux de recherche et plus particulièrement lors d'études bioinformatiques à haut débit. Les paragraphes suivants décrivent l'environnement informatique dans lequel l'ensemble des travaux décrits dans ce manuscrit ont été réalisés. Le nom usuel des machines est indiqué en italique.

A l'IGBMC, trois serveurs centraux sont dédiés à la bioinformatique et aux calculs intensifs. *Titus* est le support du serveur *web* ainsi que des services interactifs fournis par BIPS. Il s'agit d'un serveur SUN E450 quadri-processeurs UltraSPARC II, cadencé à 400 MHz, de 1 Go de mémoire vive. Son architecture est de type 64 bits et il dispose d'une installation Solaris 9. Les deux autres serveurs sont des serveurs de calculs. *Beaufort* est un cluster de six machines Compaq ES40 quadri-processeurs Alpha ev67 cadencés à 667 MHz disposant de 16 Go (machine *ouragan*) ou de 4 Go (machines *blizzard*, *cyclone*, *tempeste*, *tornado* et *trombe*) de mémoire vive reliées par Memory Channel. L'architecture est de type 64 bits et dispose d'une part, d'une installation Tru64 UNIX en version 5.1. Le serveur *Star* est, quant à lui, composé d'un cluster de six machines SUN V40Z quadri-processeurs Opteron cadencés à 2,6 GHz, disposant de 32 Go (*star 1-2*) ou 16 Go (*star 4-6*) de mémoire vive et, d'autre part de deux machines SUN M2200 bi-processeurs bi-coeur Opteron cadencés à 2,6 GHz, disposant de 4 Go de mémoire vive (*star 7-8*). L'architecture est de type 64 bits et dispose d'une installation Solaris 10 (*star 1-2*) ou Linux (*star 3-8*). Le serveur de disque est basé sur un équipement Sun V480 (Solaris 9) qui dispose de 8 Téraoctets (To) d'espace disque en RAID5, partagé avec les autres serveurs *via* NFS (*Network File System*) ainsi que sur les serveurs *star 1-2* disposant de 9 To d'espace disque en RAID5.

2.1.2. Savoir-faire bioinformatique du laboratoire

2.1.2.1. GScope : l'ossature bioinformatique du laboratoire

GScope (Figure 22) (Ripp, manuscrit en préparation) est une plate-forme de génomique dédiée à l'étude massive de données, qu'il s'agisse de génomes complets ou d'un ensemble de gènes ou de protéines. Elle est développée au laboratoire en grande partie par Raymond Ripp, qui maintient la cohérence générale de cet outil et est enrichie par tous les

programmes individuels de chaque utilisateur ou développeur de GScope. La plate-forme est écrite en Tcl/Tk et gère des fichiers au format ascii. Une interface web est également disponible.

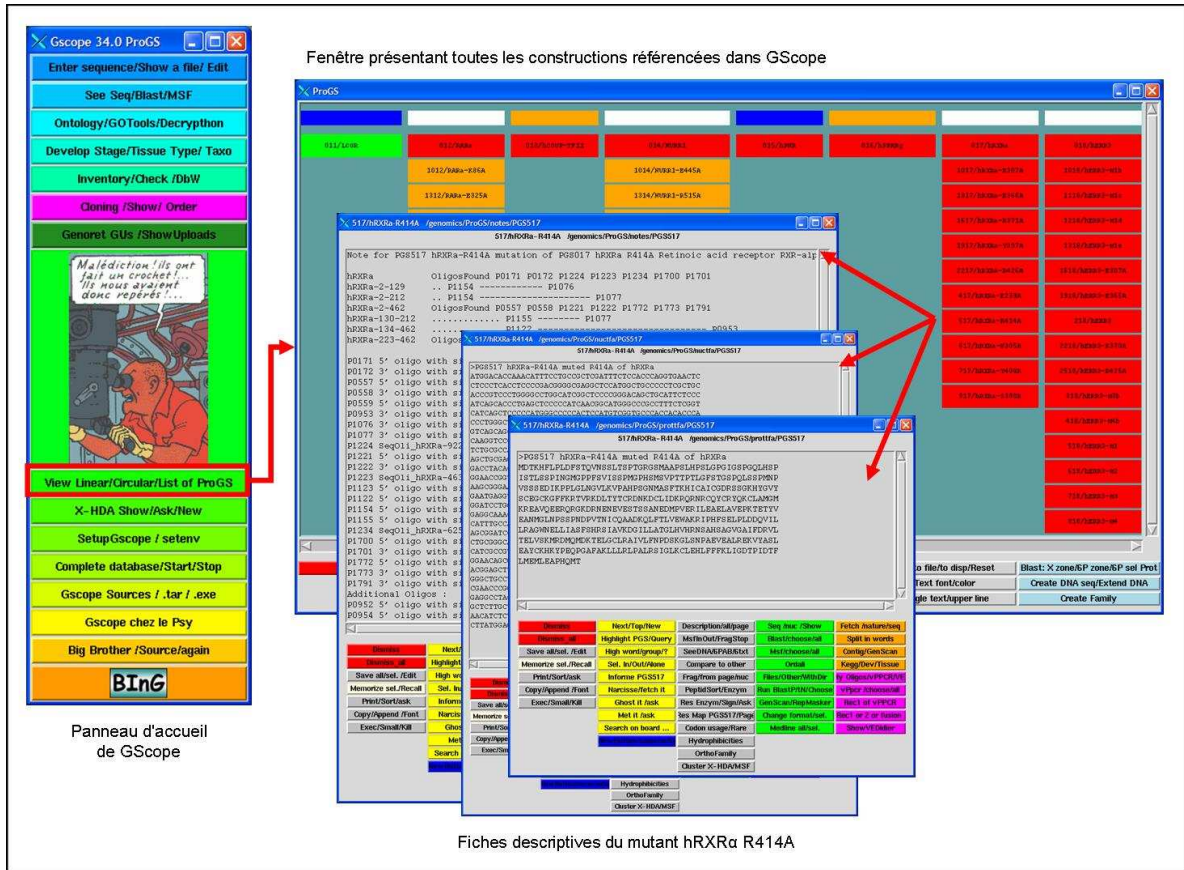


Figure 22 | Interface graphique de GScope.

Au cours de ce travail, GScope a été utilisé pour réaliser les recherches de séquences protéiques et de structures dans les bases de données ainsi que pour construire les différents alignements multiples de séquences protéiques. De plus, une composante de GScope, dédiée au clonage, nous a permis de créer, étudier, stocker et gérer toutes les informations nucléotidiques ou peptidiques relatives à nos gènes d'intérêt en réalisant, de manière automatique, le dessin des oligonucléotides nécessaires au clonage ou à la mutagenèse, la commande de ces oligonucléotides, la création virtuelle des produits de PCR et des vecteurs d'expression, la vérification du séquençage, etc.

2.1.2.2. Ecriture de programmes : le langage Tcl/Tk

Afin de rester cohérent avec les développements réalisés au sein de la plate-forme GScope, le langage de programmation utilisé au sein du laboratoire est principalement le Tcl/Tk (<http://www.tcl.tk>).

Le Tcl/Tk a l'avantage d'être un langage de script interprété, qui ne nécessite par conséquent aucune compilation. Il permet de plus le déploiement rapide de nouvelles applications, qui seront parfaitement « portables » sur n'importe quelle plate-forme informatique (Windows, MacOS et tout système Unix ou Linux).

Le Tcl/Tk est composé de deux parties : le Tcl (*Tool Command Language*), un langage de commande et le Tk (*ToolKit*), extension graphique de haut niveau. Le Tcl/Tk est tout à fait adapté à la lecture et la manipulation de fichiers, toutes les variables étant considérées comme des chaînes de caractères. Complété par Tk, le Tcl est un langage de programmation qui permet de créer aisément des interfaces graphiques très complètes et indispensables pour le biologiste.

2.1.3. Les bases de données biologiques : bases de données ou données de base ?

Sur Internet, la jungle des bases de données s'étend chaque jour un peu plus. L'afflux de séquences impose une gestion rigoureuse des bases de données et un accès aisé, conditions indispensables au bon déroulement d'analyses bioinformatiques. Un certain nombre de bases de données générales et spécialisées sont régulièrement installées et mises à jour automatiquement sur les serveurs de BIPS. Ces bases de données sont disponibles au format GCG (Butler, 1998) et peuvent être interrogées en utilisant le programme SRS (*Sequence Retrieval Software*) (Etzold and Argos, 1993).

2.1.3.1. Bases de données généralistes

Les principales bases de données publiques de séquences et de structures sont installées sur les serveurs de l'IGBMC.

2.1.3.1.1. Uniprot

La banque de séquences protéiques Uniprot (Wu *et al.*, 2006) est la réunion de deux banques, SwissProt et TrEMBL (Boeckmann *et al.*, 2003). Les séquences de la banque TrEMBL sont issues de la traduction automatique des séquences codantes de la banque généraliste de séquences nucléiques EMBL. Après validation et annotation (à titre d'exemples : la (ou les) fonction(s) de la protéine, les domaines structuraux, les modifications post-traductionnelles, les variants, les références bibliographiques, les mutations connues associées à d'éventuelles pathologies, etc.) par des experts, ces séquences protéiques sont intégrées dans la banque SwissProt. Cette banque se veut la référence en matière de banques de données de séquences protéiques annotées et représente une véritable exception au sein des banques de séquences généralistes. En effet, elle constitue une banque protéique non redondante qui privilégie la qualité et la richesse des annotations à l'exhaustivité de sa collection de séquences. Cependant, l'arrivée continue de nouvelles séquences issues de projets de séquençage de génomes pénalise la banque SwissProt, dont la richesse des annotations ne peut suivre une telle explosion de données. Ces éléments se reflètent parfaitement dans la croissance quasi linéaire du nombre d'entrées dans la banque SwissProt (Figure 23), contrairement à celle de la banque TrEMBL qui est exponentielle (Figure 24).

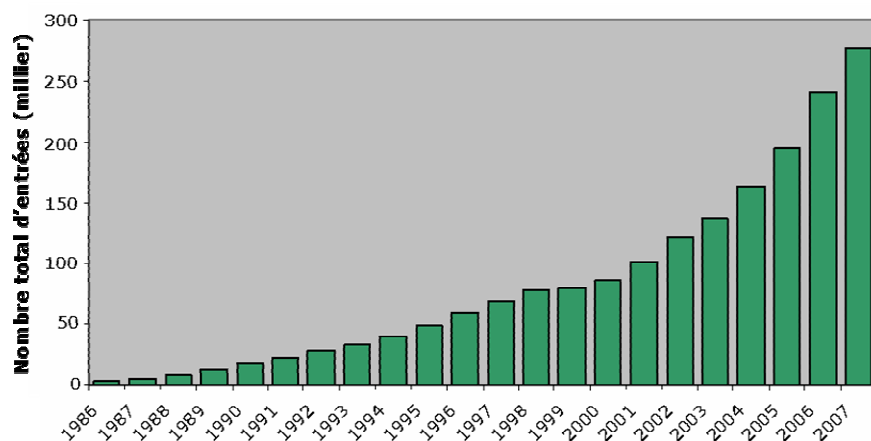


Figure 23 | Evolution du nombre d'entrées de la banque Swiss-Prot depuis sa création en 1986. Les données sont extraites du site <http://www.expasy.org/sprot/relnotes/#SPstat>.

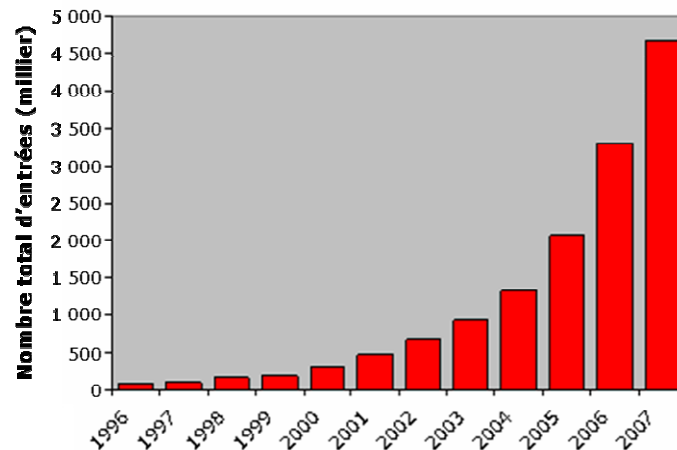


Figure 24 | Evolution du nombre d'entrées de la banque TrEMBL depuis sa création en 1996. Les données sont extraites du site http://www.ebi.ac.uk/trembl/Documents/old_trembl_rel_notes.html.

2.1.3.1.2. PDB

La PDB (*Protein Data Bank*) (Kouranov *et al.*, 2006) est la principale banque internationale de structures tridimensionnelles de protéines déterminées par cristallographie aux rayons X ou par Résonance Magnétique Nucléaire (RMN). Cette banque, fondée par le *Brookhaven National Laboratory* (BNL), contenait sept structures en 1971 ; au 14 août 2007 elle en comptait 45213. Les entrées de la banque comprennent des informations sur les structures primaires et secondaires des molécules considérées, les coordonnées atomiques, souvent les détails des expériences (conditions de cristallisation, empilement cristallin, statistiques d'affinement, etc.) ainsi que des références bibliographiques. Bien que le nombre de structures de macromolécules biologiques connues à l'heure actuelle soit très inférieur à celui des séquences, en raison des nombreuses difficultés que soulèvent les techniques de détermination de structures, celui-ci croît actuellement à une vitesse comparable à celle observée pour les séquences protéiques il y a quelques années (Figure 25), grâce aux programmes de génomiques structurales.

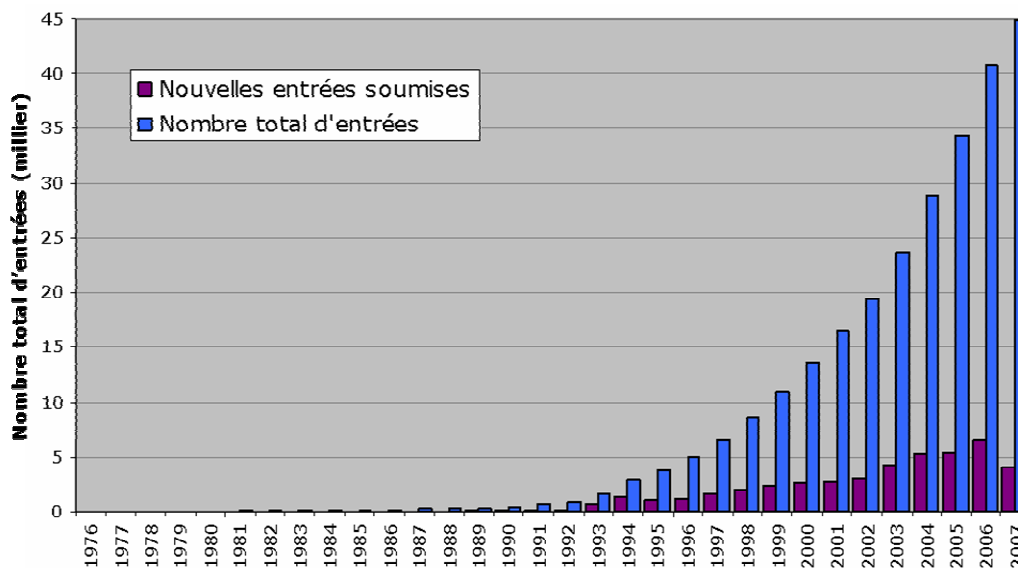


Figure 25 | Evolution du nombre d'entrées de la banque PDB depuis 1976 (adaptée de <http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>)

On peut cependant noter une redondance importante dans la PDB, plusieurs structures tridimensionnelles pouvant correspondre à la même séquence, selon les conditions d'obtention de la structure, la finesse de sa résolution, la présence de ligands différents ou de mutations.

2.1.3.2. Bases de données spécialisées

InterPro (Mulder *et al.*, 2007) est une banque intégrée qui regroupe des informations relatives aux familles de protéines, aux domaines protéiques ainsi qu'aux sites fonctionnels des protéines. Elle est distribuée depuis 1999 par l'European Bioinformatics Institute (EBI). La base de données InterPro représente une collaboration entre un certain nombre de banques de données de signatures de protéines, incluant les banques de domaines Pfam (Bateman *et al.*, 2004), Prodom (Bateman *et al.*, 2004), Smart (Letunic *et al.*, 2006) et les banques de motifs Prints (Attwood, 2002) et Prosite (Hulo *et al.*, 2006). Les signatures des protéines sont intégrées manuellement au sein des entrées d'InterPro et sont vérifiées afin de disposer d'informations biologiques et fonctionnelles fiables. InterPro fournit également des liens vers d'autres banques spécialisées, comme Gene Ontology (Ashburner *et al.*, 2000).

2.1.4. L'interrogation des banques

2.1.4.1. SRS

Le système SRS (*Sequence Retrieval System*) (Etzold and Argos, 1993) permet d'interroger, à l'aide d'une même interface, pratiquement n'importe quelle collection de séquences disponible sous la forme de fichiers texte. A l'heure actuelle, plus de 1300 banques de données différentes (séquences d'acides nucléiques et protéiques, structures, domaines protéiques ou encore voies métaboliques) sont interrogeables par SRS. La liste des serveurs SRS dans le monde est disponible à l'adresse suivante <http://downloads.lionbio.co.uk/publicsrs.html> et on peut citer en particulier ceux de l'EBI, de l'Institut Pasteur, du *Sanger Institute* et bien entendu de l'IGBMC.

Les concepteurs de SRS ont développé leur propre langage de programmation, ICARUS (*Interpreter of Commands and RecUrsive Syntax*), qui permet d'indexer toute collection structurée. Les fichiers plats des bases de données, organisés en champs structurés, sont parcourus par des interpréteurs syntaxiques et les champs contenant les données sont alors indexés. Cette indexation permet un accès rapide aux différentes entrées grâce à des requêtes complexes, mais aussi par des requêtes croisées qui exploitent les liens existant entre les différentes bases. Actuellement, la version 7 de SRS est installée au BIPS.

SRS peut être utilisé en ligne de commande sous UNIX ou par l'intermédiaire de son interface *web* sur le serveur de l'IGBMC <http://bips.u-strasbg.fr/srs/>. C'est cette dernière interface (Figure 26) qui est la plus couramment utilisée par les biologistes du laboratoire pour accéder aux informations relatives à une entrée.

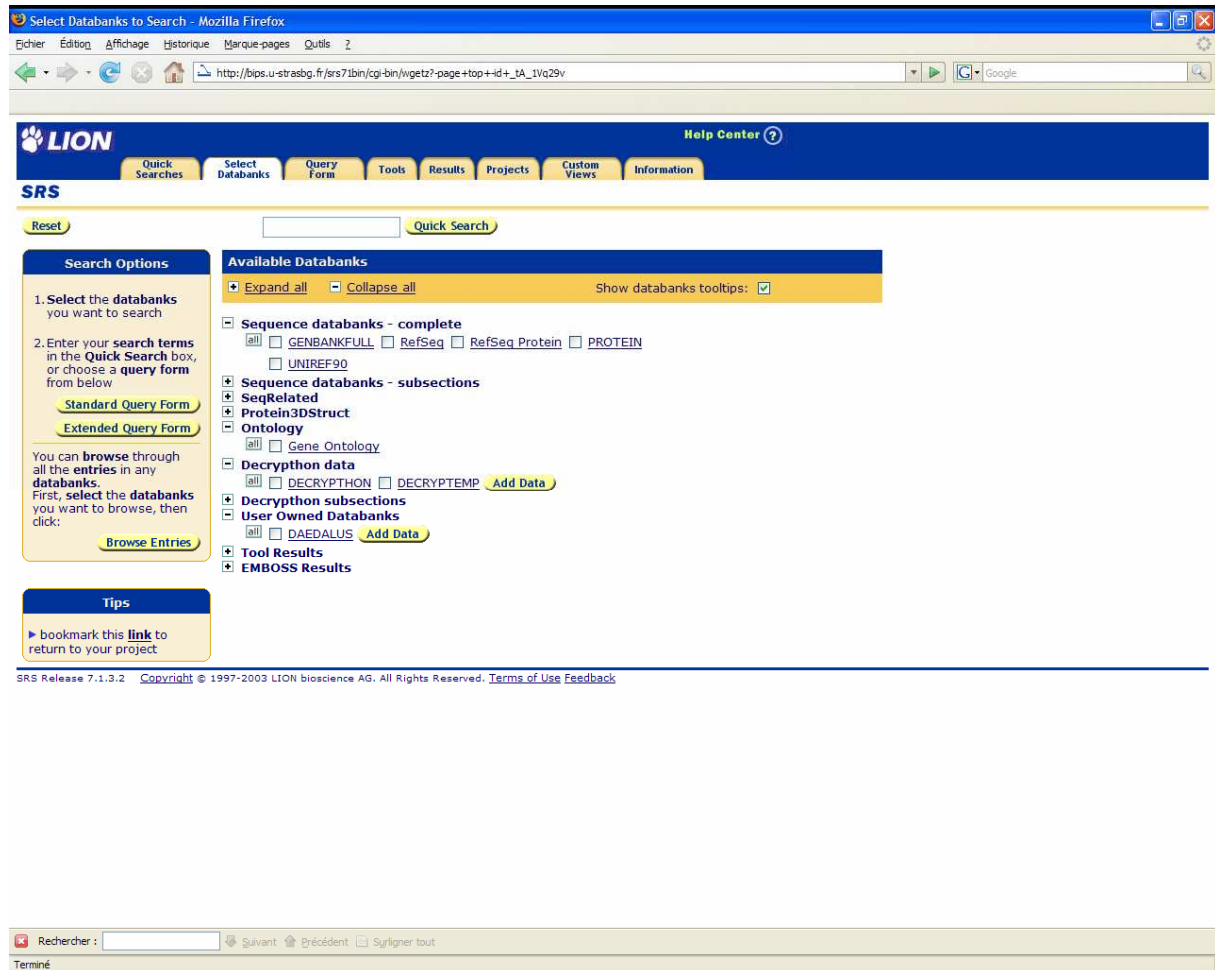


Figure 26 | Capture d'écran du serveur web SRS de l'IGBMC.

2.1.4.2. Recherche de similarité dans les bases de données

La recherche de similarité entre une séquence protéique inconnue et toutes les séquences protéiques d'une base de données est une des premières étapes à laquelle un biologiste est confronté s'il veut identifier ou mieux caractériser sa protéine d'intérêt. Au cours de ce processus, les séquences sont comparées afin d'obtenir le plus de correspondances possibles.

Nous avons, au cours de nos travaux, effectué les recherches de similarité dans les banques de séquences avec la série des programmes Blast (Altschul *et al.*, 1997), en utilisant les paramètres proposés par défaut. Nous avons en particulier utilisé le programme BlastP, dédié à la recherche de similarité dans les banques de séquences protéiques à partir d'une séquence protéique. BlastP fournit un alignement local des régions les mieux conservées entre la protéine d'intérêt et chacune des protéines détectées au sein de la banque. Pour chaque alignement, un score et une estimation statistique de sa signification biologique, la

valeur d'« expect » (ou *E-value*) est présentée. Cette dernière prend en compte plusieurs facteurs, comme la longueur des séquences, le score de similarité entre les deux séquences ainsi que la taille de la banque.

Nous considérons de manière générale que les séquences ayant une *E-value* inférieure ou égale à 0,005, présentent une similarité significative avec la séquence d'intérêt.

2.1.5. PipeAlign : de la protéine à l'alignement multiple

La première partie des études que nous avons menées est basée sur l'analyse d'alignements multiples de séquences protéiques complètes (MACS, *Multiple Alignment of Complete Sequences*) (Lecompte *et al.*, 2001). La comparaison de séquences protéiques est l'une des tâches les plus courantes et complexes de la bioinformatique et pourtant souvent négligée. La réalisation d'alignements multiples est une condition préalable essentielle à de nombreuses analyses de protéines, telles que la modélisation, la reconstruction phylogénétique ou la description de sites conservés ou variables au sein d'une même famille. PipeAlign (Plewniak *et al.*, 2003), outil d'analyse de famille de protéines développé au Laboratoire de Bioinformatique et de Génomique Intégratives (LBGI), permet la construction automatique d'un MACS de qualité au sein duquel les séquences sont classées en sous-familles. Il effectue une recherche initiale dans les bases de données UniProt et PDB par BlastP à partir d'une séquence protéique unique ou d'un jeu de séquences protéiques. Cette recherche dans les bases de données est suivie d'une cascade de six programmes couplés qui peuvent également être utilisés de manière indépendante. Ces derniers sont décrits ci-dessous (Figure 27).

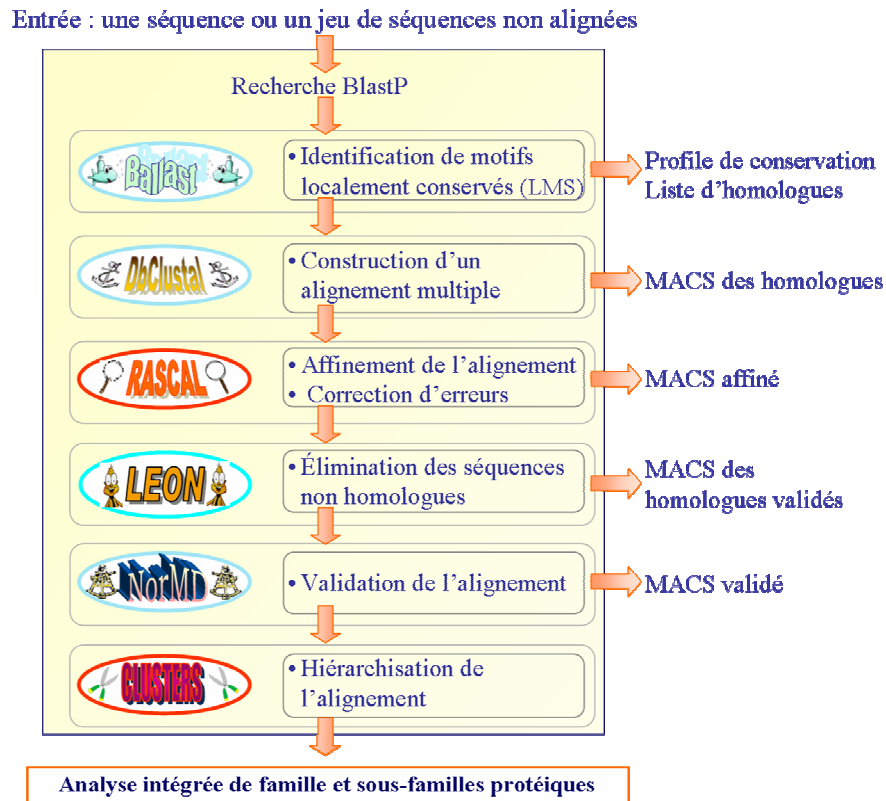


Figure 27 | Aperçu de la cascade de programmes constituant PipeAlign (adaptée de (Plewniak *et al.*, 2003).

2.1.5.1. Ballast : traitement des résultats des recherches BlastP

Ballast (Plewniak *et al.*, 2000) construit un profil de conservation à partir des séquences détectées (ou *hits*) par BlastP dans les bases de données. La contribution de chaque *hit* dans le profil est proportionnelle à sa significativité, c'est-à-dire à son *E-value*. Le profil de conservation est ensuite lissé et des pics y sont détectés en utilisant la dérivée seconde du profil lissé. Ces pics définissent les segments de conservation maximale ou LMSs (*Local Maximum Segments*), qui correspondent aux segments de séquences les mieux conservés entre la séquence initiale et les *hits* détectés par BlastP. Les positions des LMSs dans chaque séquence sont identifiées et conservées dans un fichier comme une liste d'ancres locales utilisées par DbClustal.

2.1.5.2. DbClustal : construction de MACS

DbClustal (Thompson *et al.*, 2000) est un programme d'alignement multiple de séquences complètes qui conjugue les avantages de l'alignement global et de l'alignement local.

En effet, ClustalW (Thompson *et al.*, 1994), programme d'alignement multiple basé sur l'algorithme d'alignement global développé par Needleman et Wunsch (Needleman and Wunsch, 1970), a longtemps été utilisé au laboratoire pour la construction des MACS. Les inconvénients d'une méthode basée uniquement sur l'alignement global ont cependant été mis en évidence (Thompson *et al.*, 1999), notamment dans le cas d'alignements de séquences contenant des insertions ou des extensions N-terminales ou C-terminales. DbClustal a été développé pour palier ces insuffisances. Ce programme reste basé sur l'algorithme de ClustalW, mais intègre en plus les informations de conservation locale mises en évidence par Ballast, en se servant des LMS comme points d'ancrage pour la construction de l'alignement multiple global.

2.1.5.3. RASCAL : correction des MACS

DbClustal est basé sur un algorithme utilisant des approximations. Il est donc possible que des erreurs soient introduites au sein de l'alignement multiple. Le programme RASCAL (*R*apid *S*canning and *C*orrection of *A*lignment errors) (Thompson *et al.*, 2000; Thompson *et al.*, 2003) a été développé pour détecter ces erreurs et les corriger. L'alignement multiple obtenu en sortie du programme DbClustal est divisé horizontalement (blocs de forte conservation) et verticalement (présence de sous-familles) pour former un « quadrillage » au sein duquel les régions bien alignées, et donc fiables, peuvent être identifiées et différenciées selon leur degré de conservation ou leur appartenance à une sous-famille. Les erreurs potentielles d'alignement sont détectées en comparant les profils des régions fiables. RASCAL réaligne chaque région mal alignée en utilisant un algorithme proche de celui implémenté dans ClustalW. La correction de l'alignement est restreinte aux régions les moins fiables, permettant une stratégie de ré-affinement plus performante.

2.1.5.4. LEON : évaluation de l'alignement multiple basée sur l'homologie

Un alignement multiple n'ayant de sens que si les séquences protéiques alignées sont homologues, l'étape suivante de PipeAlign a été mise en place pour détecter, au sein du MACS, les séquences n'appartenant pas à la famille d'intérêt. Le programme LEON (*multiple aLignment Evaluation Of Neighbours*) (Thompson *et al.*, 2004) se base sur les régions fiables, encore appelées *core blocks*, déterminées par RASCAL. LEON profite de la nature transitive des relations d'homologie : l'information des séquences intermédiaires est mise à profit pour mettre en évidence les régions conservées des séquences les plus divergentes. Les blocs de conservation de chaque sous-famille du MACS sont ensuite reliés, afin de former des régions contiguës de conservation, considérées comme homologues à la séquence initiale. La composition en acides aminés des séquences de l'alignement est également prise en compte par l'incorporation d'un certain nombre d'algorithmes de détection de segments, dont la composition est biaisée. Finalement, les séquences qui ne contiennent que des blocs disjoints et aucune région homologue sont retirées du MACS.

En sortie de LEON, nous disposons donc d'un MACS de bonne qualité qui ne contient que des séquences partageant au moins une région homologue à la séquence initiale.

2.1.5.5. NorMD : évaluation de la qualité d'un MACS

NorMD (*Normalized Mean Distance*) (Thompson *et al.*, 2001), est une fonction objective qui est utilisée pour évaluer la qualité d'un MACS produit par les quatre étapes précédentes. NorMD combine les avantages des techniques basées sur les scores de colonnes avec la sensibilité des méthodes introduisant des scores de similarité de résidus.

Le score assigné à l'alignement est indépendant du nombre de séquences que contient le MACS, de leur pourcentage d'identité, de leur longueur, etc. Ceci permet de comparer des scores NorMD entre alignements indépendants. Le score assigné par NorMD sera, de manière générale, compris entre 0 et 1. Plus le score est proche de 1, plus la qualité de l'alignement peut être considérée comme satisfaisante. Après analyse du comportement de NorMD sur de nombreux cas référencés, un seuil de 0,3 a été fixé, en dessous duquel la qualité du MACS est considérée comme insatisfaisante.

2.1.5.6. Secator et DPC : classification des séquences au sein du MACS

La classification des séquences au sein de l'alignement est la dernière étape intégrée à PipeAlign. Les programmes Secator (Wicker *et al.*, 2001) et DPC (*Density of Points Clustering*) (Wicker *et al.*, 2002) permettent de classer, ou hiérarchiser, des séquences d'un MACS dans de potentielles sous-familles de séquences fonctionnelles, le nombre de sous-familles créées étant déterminé de façon automatique par ces programmes.

La sortie finale de la suite de programmes PipeAlign est un MACS validé de haute qualité, dans lequel les séquences sont classées en sous-familles.

PipeAlign est mis à disposition de la communauté scientifique par l'intermédiaire du site web (Figure 28) de la plate-forme de bioinformatique de Strasbourg (<http://bips.u-strasbg.fr/PipeAlign/>). Nous disposons également de la version locale de chacun des programmes qui constitue la cascade de PipeAlign, ce qui nous permet de concevoir des « PipeAlign » à façon, ajustés au mieux à nos besoins, en fonction des études menées.

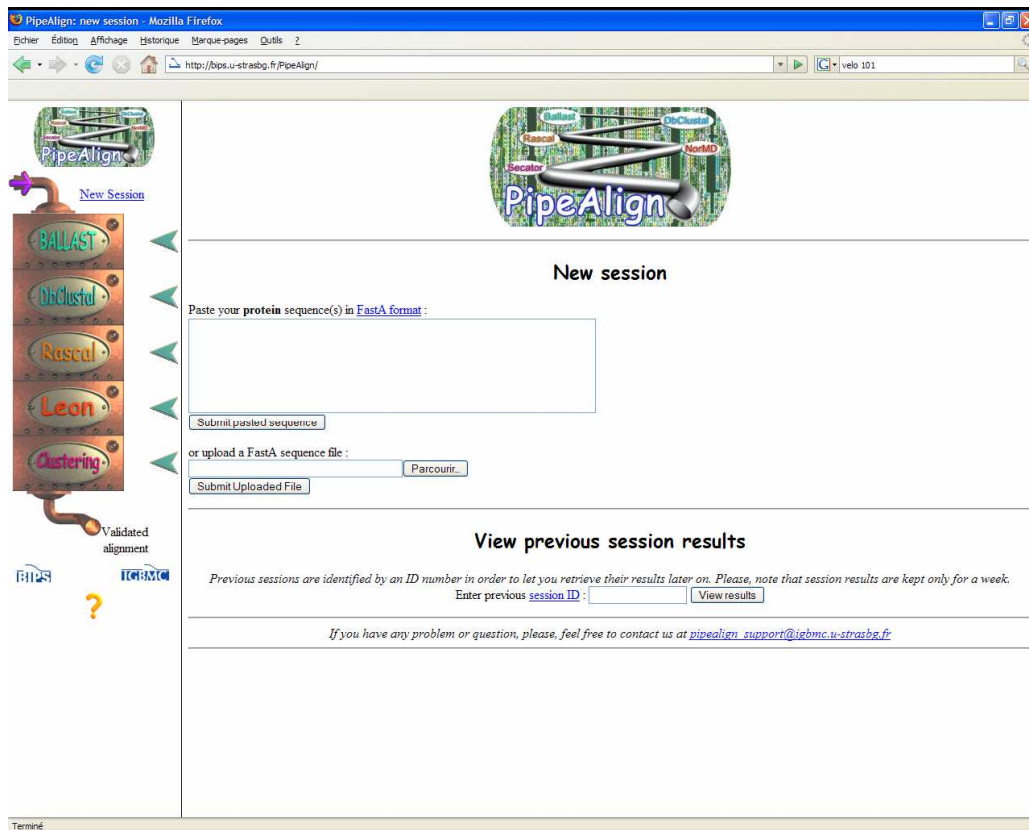


Figure 28 | Capture d'écran du serveur web de PipeAlign.

2.1.5.7. Les autres outils informatiques utilisés

2.1.5.7.1. MACSIMS : gestion de l'information au sein des MACS

MACSIMS (*Multiple Alignment of Complete Sequences Information Management System*) (Thompson *et al.*, 2006) est un système de gestion de l'information basé sur l'ontologie des alignements multiples, MAO (*Multiple Alignment Ontology*) (Thompson *et al.*, 2005), répertorié sur OBO (*Open Biological Ontologies*, <http://obofoundry.org/>), le site de référence des ontologies biologiques.

MACSIMS permet une annotation automatique des MACS par l'intégration et l'organisation de différents types de données dans le cadre de l'alignement multiple. MACSIMS combine des méthodes exploitant l'analyse des bases de connaissances et la prédiction de séquences *ab initio*. En s'appuyant sur les données de conservation au sein de l'alignement multiple hiérarchisé et la pertinence des informations générées, une étape de validation croisée des informations recueillies permet la mise en évidence des données avérées. L'information validée des séquences connues est alors propagée aux séquences inconnues, les caractérisant ainsi par des annotations fiables et détaillées.

Les informations relatives aux séquences de l'alignement sont collectées au sein de diverses banques de données comme Uniprot, PDB et InterPro et sont de plusieurs natures : des données taxonomiques, des données relatives aux résidus des sites actifs ou de liaison, des domaines structuraux et fonctionnels, des annotations de l'ontologie des gènes GO (*Gene Ontology*), etc. Les segments transmembranaires, ceux de faible complexité et les régions dites *coiled coil* sont prédits à partir de la séquence primaire des protéines de l'alignement. Les informations collectées ou générées par MACSIMS sont disponibles au format XML, un format structuré permettant une exploitation informatique automatique à haut débit et sont aussi accessibles au biologiste pour l'analyse visuelle à travers une interface web simple et conviviale : <http://bips.u-strasbg.fr/MACSIMS/>. MACSIMS facilite ainsi la collecte automatique d'informations ainsi que l'extraction de connaissances et fournit un outil interactif d'interrogation et de visualisation des résultats.

2.1.5.7.2. OrdAlie : analyse hiérarchisée d'alignements multiples

OrdAlie pour *Ordered Alignment information explorer* (Moulinier, manuscrit en préparation) est un outil d'analyse et d'édition des alignements. OrdAlie est à même de gérer aussi bien les formats traditionnels des alignements multiples comme MSF et FASTA (Annexes 1 et 2) que les alignements, au format XML, annotés par MACSIMS. OrdAlie permet une analyse des MACS et l'exploration des séquences, structures et fonctions des protéines, ainsi que des relations évolutives entre ces dernières. Les séquences de l'alignement peuvent être automatiquement ou manuellement classées en sous-familles, ce qui permet une analyse hiérarchique de la conservation des résidus au niveau de la famille entière ou des sous-familles. La représentation des conservations fournie par OrdAlie (Figure 29) permet d'évaluer aisément la pertinence des groupes constitués et met en évidence aussi bien les régions de forte conservation caractéristiques de la famille que les régions ou résidus discriminants, c'est-à-dire conservés à l'intérieur d'une sous-famille et strictement absents dans les autres groupes. Ces oppositions sont révélatrices des spécificités de chaque sous-famille à l'intérieur de la famille.

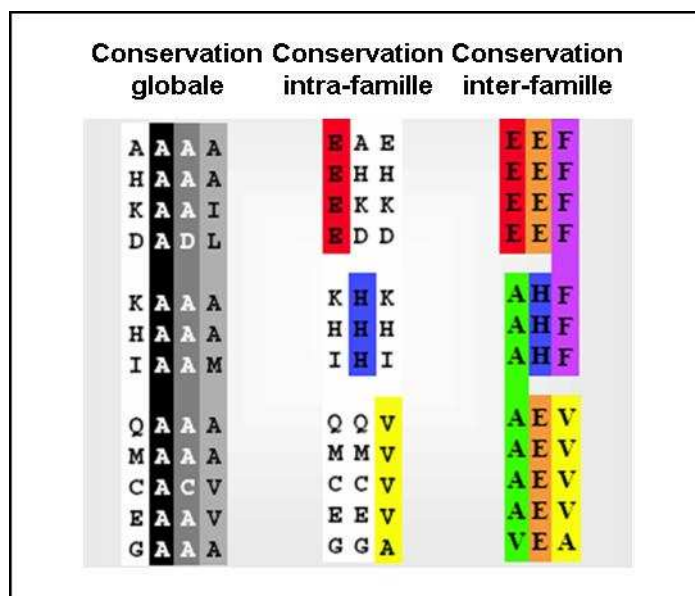


Figure 29 | Schéma de représentation des conservations adoptée dans OrdAlie. La conservation globale concerne tous les membres de la famille (noir : conservation stricte, gris foncé : au moins 80% de résidus identiques, gris clair : au moins 60% de résidus partageant les mêmes propriétés physico-chimiques) ; la conservation intra-famille : plus de 80% d'identité dans une sous-famille (rouge pour la sous-famille 1, bleu pour la sous-famille 2, jaune pour la sous-famille 3) ; conservation inter-famille : plus de 80% d'identité entre deux sous-familles (violet pour les sous-familles 1 et 2, vert pour les sous-familles 2 et 3, orange pour les sous-familles 1 et 3).

2.1.6. Les suites de programmes d'analyse de séquences

2.1.6.1. Le package GCG

Le package GCG (*Wisconsin Package Version 10.2, Genetics Computer Group Madison Wisconsin*) est une suite de plus de 150 logiciels recouvrant la quasi-totalité des programmes dédiés à l'analyse de séquences. La version 10 permet de manipuler, de visualiser, d'analyser, d'éditer et de comparer des séquences des bases de données au format GCG installées localement. Parmi divers logiciels d'analyse, l'éditeur d'alignements multiples SeqLab a été couramment utilisé pour la visualisation et la correction manuelle de plusieurs alignements multiples réalisés au cours de ma thèse. SeqLab est doté d'une interface graphique dotée dont les nombreuses fonctions permettent l'étude des alignements multiples de manière conviviale (Figure 30).

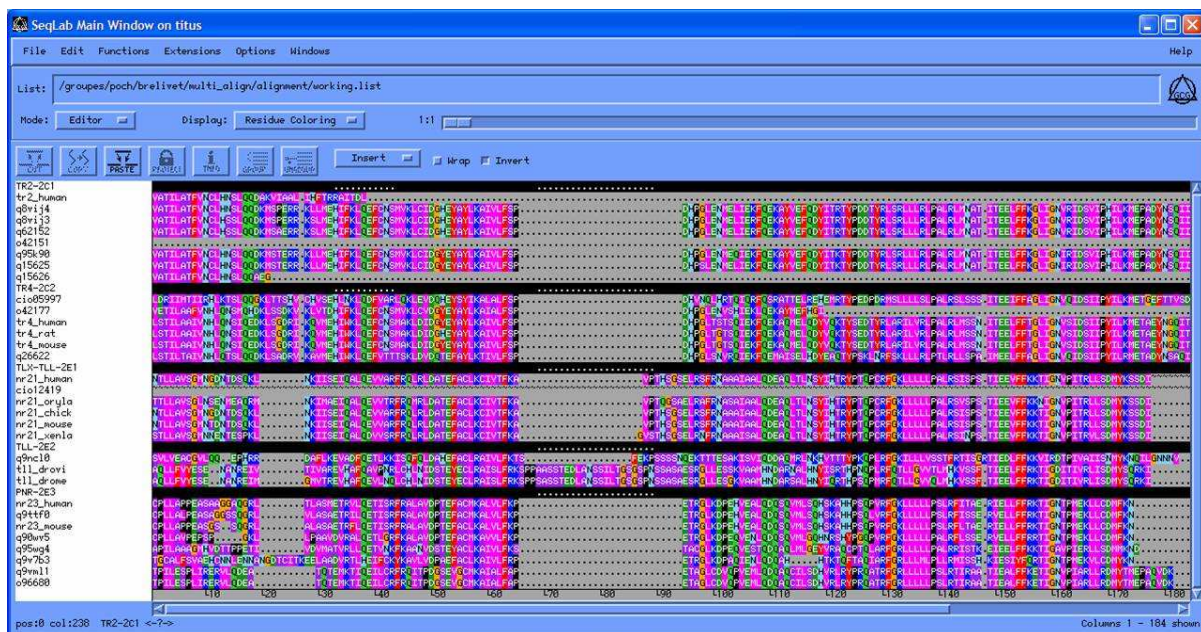


Figure 30 | Capture d'écran du programme SeqLab.

2.1.6.2. EMBOSS

EMBOSS (*European Molecular Biology Open Software Suite*) est une suite de programmes dédiée à l'analyse de séquences (<http://emboss.sourceforge.net>) qui regroupe les outils majeurs d'analyse de séquences. Ce logiciel est compatible avec différents systèmes

(UNIX et Windows) et une interface web y est intégrée. EMBOSS peut être couplé à SRS et accède ainsi directement aux banques de séquences indexées par ce dernier.

2.1.7. Analyse structurale des protéines

2.1.7.1. Le logiciel O

Les structures tridimensionnelles de protéines ont été superposées en utilisant le logiciel O. O est un ensemble de programmes dédiés à la modélisation des macromolécules. Ce logiciel d'analyse graphique est construit sous la forme d'une base de données. Toutes les données utilisées peuvent être stockées dans une librairie locale. O permet de visualiser les molécules, soit sous forme de tracé du squelette des carbones α , soit sous forme de protéines entières avec les chaînes latérales. Il permet non seulement une analyse très fine grâce à des calculs de distances inter-atomiques mais également de superposer des structures, de manipuler des modèles dans des cartes de densité, de construire et d'appliquer des opérations de symétrie sur des molécules ou encore d'améliorer des modèles. Il présente une interface graphique très complète avec un menu que l'utilisateur peut définir lui-même selon ses besoins. Il fonctionne simultanément avec une fenêtre de ligne de commande et une fenêtre de visualisation.

2.1.7.2. PyMOL

PyMOL (<http://pymol.sourceforge.net>) est un logiciel de visualisation et d'analyse de la structure tridimensionnelle des molécules. Des fonctions permettent de visualiser et d'explorer les différentes parties ou propriétés d'une molécule (comme les charges, les résidus hydrophobes et polaires, etc.). Il fonctionne sur toutes les plates-formes, incluant UNIX, Windows et Macintosh. Cet outil est capable d'importer tous les formats de modèles moléculaires disponibles et peut être contrôlé au moyen de scripts (liste de commandes). La qualité et la diversité des représentations disponibles, ainsi que les possibilités offertes à l'utilisateur, rendent cet outil indispensable à l'analyse structurale.

2.1.7.3. Modeller : construction de modèles par homologie

Modeller (Sali *et al.*, 1995), version 9v1, est installé en local sur le serveur *Star*. Il existe une version fonctionnant sur le système d'exploitation Windows. Modeller permet de construire des modèles protéiques tridimensionnels grâce à une méthode de minimisation des contraintes spatiales, qu'il définit en terme de « fonction de probabilités de densité ». Modeller est souvent utilisé pour la modélisation comparative. L'utilisateur peut fournir un alignement comprenant la séquence qu'il veut modéliser (séquence cible) et une ou plusieurs structures de référence. Le processus de calcul automatique du modèle tridimensionnel comporte trois étapes. Dans la première étape, Modeller extrait les contraintes spatiales (contraintes de distance des liaisons, des angles, des angles dièdres et des contacts entre les atomes non liés) de la séquence cible, à partir des informations trouvées dans l'alignement structural. Dans la seconde étape, il estime ces contraintes, soit à partir de la structure de référence (pour les résidus conservés dans l'alignement), soit à partir de celles calculées d'après une base de données constituée de 105 alignements contenant 416 protéines de structures connues (pour les résidus non conservés). Ainsi les « relations » extraites, exprimées en terme de « densité de probabilités conditionnelles », sont utilisées comme des contraintes spatiales. Par exemple, les probabilités pour les différentes valeurs des angles dièdres de la chaîne principale sont calculées en fonction du type de résidu considéré, de la conformation du résidu équivalent et de l'identité de séquence entre les 2 protéines. Dans la dernière étape, Modeller combine ces contraintes spatiales avec les termes du champ de force CHARMM dans une fonction qu'il optimise pour construire le « modèle tridimensionnel tout atome ». Pour chaque étude de modélisation, nous avons généré cinq modèles et conservé celui présentant l'énergie la plus basse.

Chapitre 3 - Biochimie, biologie moléculaire et cellulaire

3.1. Techniques de biologie moléculaire et cellulaire

3.1.1. Les souches d'*Escherichia coli* utilisées

Deux souches bactériennes, dont les génotypes sont présentés en Annexe 3, ont été utilisées : la souche DH5 α au cours des étapes de clonage, d'amplification de vecteur et lors du sous-clonage dans le vecteur pSG5 pour les études fonctionnelles ; la souche BL21(DE3) dans le cadre des étapes d'expression. Le génome de la souche BL21(DE3) a intégré le lysogène du bactériophage λ DE3 et comporte notamment le gène codant l'ARN polymérase du bactériophage T7 sous contrôle du promoteur lacUV5 inductible à l'IPTG. De plus, la souche *E. coli* BL21(DE3) utilisée pour l'expression est déficiente en certaines protéases qui pourraient dégrader la protéine recombinante et diminuer le rendement de purification (Studier and Moffatt, 1986; Studier *et al.*, 1990).

3.1.2. Vecteurs de clonage et d'expression

Le plasmide pDONR207 (*genta*^R, *Invitrogen*) a été utilisé dans le cadre du clonage basé sur la technologie Gateway[®]. Les vecteurs d'expression p0GWA, pHGWA, pHMGWA (Busso *et al.*, 2005) (*amp*^R) (LBGS) (Busso *et al.*, 2005), basés sur le vecteur pET22b (*Novagen*) permettent d'exprimer les protéines, fusionnées ou non à différentes étiquettes à leur extrémité N-terminale (N-ter). Le vecteur p0GWA ne possède pas d'étiquette à son extrémité N-ter. Il est utilisé dans le cadre de co-purification avec un autre vecteur possédant une étiquette en N-ter. Le vecteur pHGWA permet de produire une protéine fusionnée à un polypeptide hexahistidine. Le vecteur pHMGWA possède, quant à lui, en plus du polypeptide hexahistidine, une séquence d'ADN codant pour la protéine MBP (*Maltose Binding Protein*) connue pour augmenter la solubilité de la protéine à laquelle elle est fusionnée. Ces polypeptides de fusion peuvent être éliminés par protéolyse spécifique, en utilisant la thrombine. Les vecteurs plasmidiques pSG5 (*amp*^R, *Stratagene*) et pG4M (*amp*^R) sont des vecteurs d'expression eucaryotes utilisés lors des études fonctionnelles.

3.1.3. Antibiotiques et solutions pour milieux de culture

Toutes ces solutions sont filtrées en conditions stériles sur membrane de porosité 0,22 μm et stockées à -20°C . La concentration d'ampicilline est à 100 mg/ml et est conditionnée dans l'eau ; celle de la gentamycine est à 7 mg/ml dans l'eau ; celle de l'IPTG à 0,5 M dans l'eau.

3.1.4. Tampons et milieux

3.1.4.1. Le milieu SOB

Ce milieu est composé de 2% (poids/volume) de bacto-tryptone, de 0,5% (poids/volume) d'extrait de levure, de NaCl 10 mM, de KCl 2,5 mM, de MgCl_2 10 mM et de MgSO_4 10 mM. Le bacto-tryptone et l'extrait de levure sont pesés pour le volume de milieu approprié, le volume est ensuite ajusté, puis le milieu est autoclavé. Les différents sels sont ajoutés dans des conditions stériles. Le pH du milieu est ajusté avec NaOH pour se situer entre 6,7 et 7,0.

3.1.4.2. Le milieu TB

Le milieu TB est constitué de Pipes 10 mM pH=7, de CaCl_2 15 mM, de MnCl_2 55 mM et de KCl 250 mM. Le pH est ajusté à pH=7,0 avec du KOH 5 N avant d'ajouter le MnCl_2 . Le milieu est filtré sur membrane de porosité 0,22 μm dans des conditions stériles.

3.1.5. Electrophorèse en gel d'agarose

3.1.5.1. Tampon de dépôt

Le tampon de dépôt sur gel d'agarose est constitué de 10 mM Tris-HCl pH=8,0, 1 mM EDTA, de bleu de bromophénol 0,25% (poids/vol), de bleu de xylène cyanol FF 0,25% (poids/vol) et de glycérol 40% (vol/vol).

3.1.5.2. Tampon TBE 10X

Le tampon TBE 10X contient, pour un l l, 108 g de Tris base, 55 g d'acide borique, 40 ml d'EDTA 0,5 M pH=8,0.

3.1.5.3. Tampon TAE 50X

Le tampon TAE 50X se compose, pour 1 l, de : 242 g de Tris base, 57,1 ml d'acide acétique glacial, 100 ml d'EDTA 0,5M pH=8,0.

La masse adéquate d'agarose (*Appligène*) (1% poids/vol) est dissoute par chauffage au micro-onde dans le tampon TAE ou TBE 1X. Pour visualiser l'ADN par fluorescence sous ultraviolets, du bromure d'éthidium (BET) (stock 10 mg/ml) est ajouté dans le gel à 10 µg/ml. Le gel se solidifie par refroidissement.

3.1.6. Préparation des bactéries chimio-compétentes et transformation

Toutes les étapes sont réalisées dans des conditions stériles. Des tests ont été effectués pour vérifier la compétence des différentes souches en fonction des méthodes de préparation et nous avons constaté une meilleure efficacité de transformation pour les bactéries DH5α préparées selon la méthode Inoue (Inoue *et al.*, 1990) et pour les bactéries BL21(DE3) selon la méthode au CaCl₂ (Dagert and Ehrlich, 1979) (David Rosé, communication personnelle).

3.1.6.1. Méthode Inoue : bactéries chimio-compétentes DH5α

Une pré-culture de 100 ml de milieu LB est inoculée à partir d'un stock glycérol et incubée à 37°C sous agitation constante pendant la nuit. La pré-culture est refroidie pendant 8 h à 4°C. Une culture de 250 ml de milieu SOB estensemencée par 4 ml de cette pré-culture (pour avoir une densité optique DO_{600nm} de départ de 0,1) puis incubée à 18°C sous agitation constante (190 rpm) jusqu'à atteindre une DO_{600nm} de 0,4. La culture est ensuite refroidie 10 min sur la glace, puis centrifugée pendant 15 min à 2500 g à 4°C. Le culot est re-suspendu dans 40 ml de milieu TB froid. Ce mélange est incubé 10 min sur glace puis re-centrifugé pendant 15 min à 2500 g à 4°C. Le culot est à nouveau re-suspendu dans 10 ml de tampon TB froid. Du DMSO est rajouté goutte à goutte à une concentration finale de 7% (vol/vol). Ce

mélange est laissé 10 min sur la glace, puis dispensé en fractions aliquotes de 50 et 100 μ l. Les fractions aliquotes sont finalement congelées dans l'azote liquide et conservées à -80°C.

3.1.6.2. Méthode au CaCl₂ : bactérie chimio-compétente BL21

Une pré-culture de 100 ml de milieu LB est inoculée à partir d'un stock glycérol et incubée à 37°C sous agitation constante pendant la nuit. Le lendemain, une culture de 1 l de milieu LB estensemencée avec 20 à 25 ml de cette pré-culture (pour avoir une DO_{600nm} de départ comprise entre 0,05 et 0,1), incubée à 37°C sous agitation constante (190 rpm) jusqu'à atteindre une DO_{600nm} de 0,4. La culture est ensuite refroidie 20 min sur la glace puis centrifugée pendant 15 min à 2500 g à 4°C. Le culot est re-suspendu dans 250 ml de NaCl 10 mM. Ce mélange est incubé 10 min sur glace puis re-centrifugé pendant 15 min à 2500 g à 4°C. Le culot est à nouveau repris dans 250 ml de CaCl₂ 75 mM et incubé sur glace pendant 35 min. Le mélange est à nouveau centrifugé 15 min à 2500 g à 4°C, le culot est repris dans 15 ml de CaCl₂ 75 mM. Après rajout de 3 ml de glycérol 87%, la solution est homogénéisée, incubée 20 min dans la glace puis dispensée en fractions aliquotes de 50 et 100 μ l. Les fractions aliquotes sont finalement congelées à -80°C.

3.1.6.3. Transformation chimique de bactéries chimio-compétentes

Une fraction aliquote de bactéries chimio-compétentes (50 μ l) est mélangée à l'ADN plasmidique (50 ng) et le mélange est laissé 30 min sur la glace. Un choc thermique est alors réalisé à 42°C pendant 1 min. Le mélange est ensuite placé sur la glace pendant 2 min avant d'y rajouter 150 μ l de milieu LB. La réaction de transformation est placée à 37°C sous agitation constante pendant 1 h. Différents volumes de mélange de transformation (entre 20 et 150 μ l au maximum) sont étalés sur milieu LB agar supplémenté de l'antibiotique adéquat.

3.1.7. Purification d'ADN plasmidique

Les méthodes de maxi- et de mini-préparations d'ADN plasmidique sont basées sur la méthode de Birnboim et *coll.* (Birnboim and Doly, 1979).

3.1.7.1. Maxi-préparation

Le kit « NucleoBond® Xtra Maxi » (*Macherey-Nagel*) est utilisé pour réaliser une purification d'ADN plasmidique en grande quantité. Ce kit, basé sur la chromatographie échangeuse d'anions, permet d'obtenir environ 1mg d'ADN plasmidique pur à partir d'une culture de 400ml.

Un milieu de culture LB de 300 ml est inoculé avec une colonie isolée prélevée sur une boîte LB agar après une transformation. Le milieu de culture LB est supplémenté de 300 µl de l'antibiotique adéquat pour éviter la perte du plasmide. Cette culture est incubée à 37°C sous agitation constante (190 rpm) entre 12 et 16 h (toute une nuit).

Les étapes sont décrites sur la Figure 31. Les cellules sont récoltées par centrifugation à 4000 rpm pendant 15 min à 4°C. Le surnageant est éliminé et le culot cellulaire est re-suspendu dans 12 ml de tampon RES additionné de RNase A en pipetant les cellules par aspiration-refoulement dans un tube Falcon™ de 50 ml. Les cellules sont ensuite lysées par l'addition de 12 ml de tampon LYS (NaOH/SDS). Cette suspension est homogénéisée avec précaution en inversant le tube 5 fois, puis incubée à température ambiante pendant 4 min. Il ne faut surtout pas utiliser de vortex, sinon l'ADN chromosomique se fractionnerait, se détacherait des débris cellulaires et contaminerait alors la suspension. Dans ces conditions, les protéines et l'ADN chromosomique sont dénaturés. L'ARN est dégradé par la RNase A. Ensuite, 12 ml de tampon de neutralisation NEU (contenant de l'acétate de potassium) sont ajoutés à la suspension, causant la précipitation du SDS en KDS (potassium dodecyl sulfate), des protéines, de l'ADN chromosomique et des débris cellulaires. L'ADN plasmidique retrouve sa forme super-enroulée et reste en solution. Le lysat est immédiatement mélangé en inversant avec précaution le tube. Ce mélange est centrifugé à 4000 rpm pendant 15 min pour culotter les précipités. La colonne, avec le filtre, est équilibrée avec 25 ml de tampon EQU. Après la centrifugation, le surnageant est simultanément clarifié et chargé sur la colonne, où l'ADN se fixe sur la résine. La colonne se vide par gravité. Le filtre et la colonne

sont lavés avec 15 ml de tampon d'équilibration EQU. Le filtre est alors retiré de la colonne et la colonne lavée avec 25 ml de tampon de lavage WASH. Enfin, l'ADN plasmidique est élué avec 15 ml de tampon d'éluion ELU dans un tube Falcon™ de 50 ml. L'éluion est réalisée dans des concentrations de sel élevées et par modification du pH de 6,5 à 9. Dans ces conditions alcalines, la charge positive de la résine échangeuse d'anions est neutralisée et l'ADN plasmidique relâché.

Le kit « NucleoBond® Finalizer » est utilisé à la suite de cette purification pour concentrer et dessaler rapidement l'ADN plasmidique.

L'ADN est précipité avec 0,7 volume d'isopropanol à température ambiante, soit 10,5 ml. Ce mélange est vigoureusement agité puis laissé reposer pendant 2 min. Le précipité est chargé sur la membrane d'un *Finalizer* à l'aide d'une seringue et le filtrat est éliminé. Le précipité est ensuite lavé avec 2 ml d'éthanol à 70% puis la membrane est séchée en y faisant passer de l'air, afin d'éliminer toute trace d'éthanol. Finalement, l'ADN plasmidique purifié est élué avec 600 µl de tampon TE (10 mM Tris-HCl pH=7,5 ; 1 mM EDTA). Pour un rendement maximal, il est préconisé de réaliser l'éluion 2 fois. Le premier éluat est à nouveau chargé sur la membrane du *Finalizer* et élué une seconde fois, dans le même tube collecteur, pour permettre la récupération d'un maximum du plasmide. La concentration et la qualité du plasmide sont respectivement mesurées et vérifiées par spectrométrie UV à 260nm en utilisant le spectrophotomètre ND-1000 NanoDrop® (*NanoDrop*). La qualité du plasmide est également vérifiée par électrophorèse sur gel d'agarose 1% (poids/vol de TAE 1X) additionné de 1µl de bromure d'éthidium (BET) (10 mg/ml).












1. Récolte des cellules	 20min, 4°C, 3500g Elimination du surnageant
2. Equilibrage de la colonne	 Tampon EQU 12ml
3. Lyse cellulaire alcaline	 Tampon RES 12ml Tampon LYS 12ml 5min, TA Tampon NEU 12ml
4. Clarification et charge du lysat	 Clarification du lysat  20min, 4°C, 3500g Charge du lysat
5. Lavage 1 avec le filtre	 Tampon EQU 15ml Elimination du filtre
6. Lavage 2 sans le filtre	 Tampon WASH 25ml
7. Elution de l'ADN	Tampon ELU 15ml
8. Précipitation de l'ADN	Isopropanol 10,5ml
9. Charge du précipité sur la membrane	 NucleoBond Finalizer
10. Lavage du précipité	 Ethanol 70% 5ml
11. Séchage de la membrane	 10 x air
12. Elution de l'ADN	 Elution en 2 fois Elution 1 Tampon TE 600µl Elution 2 Avec l'éluat de la 1 ^{ère} élution

Figure 31 | Protocole d'une maxi-préparation d'ADN plasmidique avec le kit NucleoBond® Xtra Maxi (Macherey-Nagel).

3.1.7.2. Mini-préparation

Le kit NucleoSpin® Plasmid (*Macherey-Nagel*) est utilisé pour réaliser une purification d'ADN plasmidique. Il permet d'obtenir jusqu'à 25 µg d'ADN plasmidique purifié à partir d'une culture de 4 ml (Figure 32). La composition des tampons du kit n'est pas donnée par le fabricant.

Un milieu de culture LB de 4 ml, supplémenté de 4 µl de l'antibiotique adéquat, est inoculé avec une colonie isolée prélevée sur une boîte LB agar après une transformation. Cette culture est incubée à 37°C à 190 rpm durant la nuit (12 à 16 h).

Les cellules sont ensuite récoltées par centrifugation à 2500 g pendant 10 min à 4°C. Le surnageant est éliminé et le culot bactérien est repris dans 250 µl de tampon A1 supplémenté de RNase A. Les cellules sont ensuite lysées par l'addition de 250 µl de tampon A2 (NaOH/SDS). Cette suspension est homogénéisée avec précaution en inversant les tubes 6 à 8 fois, puis incubée à température ambiante durant 5 min. Le tampon A3 est ajouté (300 µl) et la suspension est à nouveau homogénéisée par inversion des tubes. Ce mélange est centrifugé pendant 10 min à 11000 g à température ambiante. Le surnageant est déposé sur une colonne NucleoSpin® Plasmid, où l'ADN se fixe sur la membrane en silice. Les colonnes sont centrifugées 1 min à 11000 g. Le filtrat est éliminé. La membrane est lavée une première fois par 500 µl de tampon AW et par centrifugation à 11000 g pendant 1 min, puis une deuxième fois par 600 µl de tampon A4 (éthanol). Deux centrifugations permettent ensuite de sécher correctement la membrane. Enfin, l'ADN plasmidique est élué par 30 µl de tampon AE (5 mM Tris-Cl pH=8,5) par centrifugation d'1 min à 11000 g. La concentration et la qualité du plasmide sont respectivement mesurée et vérifiée par spectrométrie UV à 260nm en utilisant le spectrophotomètre ND-1000 NanoDrop® (*NanoDrop*). La qualité du plasmide est également vérifiée par électrophorèse sur gel d'agarose 1% (poids/vol de TAE 1X) additionné de 1 µl de BET (10 mg/ml).








<p>1. Récolte des cellules</p>	 <p>10min, 4°C, 2500g</p>
<p>2. Lyse cellulaire</p>	 <p>Tampon A1 250µl Tampon A2 250µl 5min T° ambiante (TA) Tampon A3 300µl</p>
<p>3. Clarification du lysat</p>	 <p>10min, TA, 11000g</p>
<p>4. Fixation de l'ADN sur la membrane en silice</p>	 <p>Charge du surnageant 1min, TA, 11000g</p>
<p>5. Lavage de la membrane en silice</p>	 <p>Tampon AW 500µl 1min, TA, 11000g Tampon A4 600µl 1min, TA, 11000g</p>
<p>6. Séchage de la membrane en silice</p>	 <p>2min, TA, 11000g</p>
<p>7. Elution de l'ADN</p>	 <p>Tampon AE 30µl 1min, TA 1min, TA, 11000g</p>

Figure 32 | Protocole d'une mini-préparation d'ADN plasmidique avec le kit NucleoSpin® Plasmid (Macherey-Nagel).

3.1.8. Stratégies de clonage

3.1.8.1. Le clonage Gateway®

3.1.8.1.1. Principe

Dans notre laboratoire, la technologie Gateway®, technique de clonage basée sur les propriétés de recombinaison du bactériophage λ (Bushman *et al.*, 1985; Cassuto *et al.*, 1971; Landy, 1989), est couramment employée. Cette technologie permet un transfert rapide de séquences d'ADN d'intérêt depuis un plasmide navette vers n'importe quel plasmide d'expression compatible permettant d'utiliser de multiples systèmes pour l'analyse de l'expression protéique.

3.1.8.1.2. La recombinaison

Le système de recombinaison du bactériophage λ est un système de recombinaison conservative (ni gain ni perte de nucléotides) site spécifique, qui lui permet de s'intégrer dans le génome de *E. coli* et permet le passage du cycle lytique au cycle lysogénique du bactériophage. Cette recombinaison se fait par l'intermédiaire de séquences dites d'attachements spécifiques (sites *att*): *attB* sur le chromosome d'*E. coli* et *attP* sur le bactériophage λ . Ces sites *att* servent de sites d'interaction pour les protéines de recombinaison. Dès l'intégration du bactériophage λ , la recombinaison se fait entre les sites *attB* et *attP* pour donner les sites *attL* et *attR*. L'ADN flanquant les sites de recombinaison est échangé et, après recombinaison, les sites *att* sont inclus dans la séquence hybride. La recombinaison est conservative et ne nécessite aucune synthèse d'ADN. Elle peut être utilisée avec n'importe quel type d'ADN (super-enroulé, linéaire ou relaxé).

3.1.8.1.3. Les enzymes recombinantes

La recombinaison du bactériophage λ est catalysée par une variété d'enzymes qui se lient à des sites spécifiques d'attachement. Ces enzymes diffèrent selon le cycle du bactériophage λ . Le cycle lysogénique est catalysé par l'intégrase (Int) du bactériophage λ et les protéines du facteur d'intégration d'*E. coli* (IHF). Le cycle lytique du bactériophage λ est, quant à lui, catalysé par l'intégrase (Int) du bactériophage λ , l'excisionase (Xis) et les protéines du facteur d'intégration d'*E. coli* (IHF). Pour le clonage Gateway®, ces enzymes

sont produites artificiellement et les composés nécessaires à la recombinaison ont été modifiés et adaptés pour augmenter la spécificité et l'efficacité de la réaction (Hartley *et al.*, 2000). Ces enzymes constituent les mélanges BP Clonase® et LR Clonase® (Figure 33).

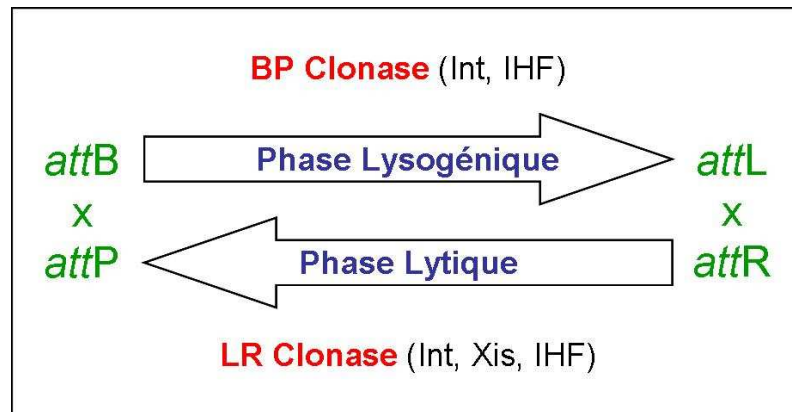


Figure 33 | Réactions catalysées par les enzymes recombinantes.

3.1.8.1.4. Les réactions de recombinaison

Deux réactions de recombinaison constituent la base de la technologie Gateway®. La réaction BP (Figure 34) permet la recombinaison entre un substrat *attB* (produit PCR flanqué des sites *attB* ou clone d'expression flanqué des sites *attB* linéarisé) et un substrat *attP* (un vecteur donneur, pour créer un clone d'entrée (pENTR) contenant les sites *attL*). Cette réaction est catalysée par l'enzyme BP Clonase®.



Figure 34 | Réaction BP (adapté de Invitrogen).

La réaction LR (Figure 35) réalise la recombinaison entre un substrat *attL* (vecteur d'entrée) et un substrat *attR* (vecteur de destination ou pDEST) pour créer un clone d'expression contenant les sites *attB*. Cette réaction est catalysée par l'enzyme LR Clonase®.

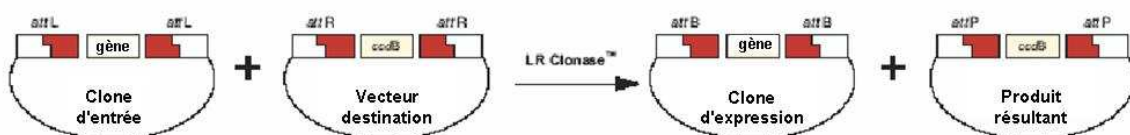


Figure 35 | Réaction LR (adapté de Invitrogen).

Les vecteurs donneur et destination possèdent une cassette contenant le gène *ccdB* pour la sélection négative. Après la réaction BP ou LR, cette cassette est remplacée par le gène d'intérêt. La présence du gène *ccdB* permet la sélection négative des vecteurs donneur et destination après la recombinaison et la transformation. En effet, la protéine CcdB interfère avec l'ADN gyrase d'*E. coli* (Bernard and Couturier, 1992) et inhibe ainsi la croissance de la plupart des souches d'*E. coli* couramment utilisées pour le clonage.

3.1.8.1.5. Protocole

Dessin des oligonucléotides contenant les séquences attB pour amplifier le gène d'intérêt

Le clonage Gateway® requiert le dessin d'un oligonucléotide « sens » comportant le site *attB1* et d'un oligonucléotide « anti-sens » comportant le site *attB2*. Un oligonucléotide court « sens » est utilisé pour une première amplification puis un oligonucléotide générique pour une deuxième amplification. L'oligonucléotide court comporte la séquence codante pour un site de clivage à une protéase, éventuellement un site de restriction et 18 à 25 nucléotides spécifiques de la cible. L'oligonucléotide générique contient la séquence *attB1* suivie de la séquence codante pour un site de clivage à une protéase. Nous avons choisi de placer la séquence codante pour un site de clivage à cet endroit afin de pouvoir facilement éliminer de la protéine d'intérêt non seulement d'éventuelle(s) fusion(s) N-ter, mais également la séquence protéique codée par la séquence *attB1* (TSLYKKAG). L'oligonucléotide « anti-sens » est un oligonucléotide long comportant, dans l'orientation 5' vers 3', la séquence *attB2*, d'éventuels signaux et 18 à 25 nucléotides spécifiques de la séquence.

Amplification des domaines par PCR

Cette méthode enzymatique très fréquemment utilisée est efficace pour amplifier des fragments d'ADN *in vitro*. En général, une PCR réalisée dans un volume final de 20 µl suffit à obtenir suffisamment de matériel nécessaire à la réaction BP. L'amplification est réalisée en par l'intermédiaire de deux PCR consécutives en utilisant principalement la Phusion ADN polymérase (2 U/µl) (*Finnzymes*) selon le protocole décrit ci-dessous.

Une première PCR est réalisée dans un volume de 20 µl avec l'oligonucléotide court spécifique « sens » et l'oligonucléotide long « anti-sens ». A 1 µl d'ADN matriciel (10 ng) sont

ajoutés 4 μl de tampon HF 5X, 0,25 μl de chaque oligonucléotide à 10 μM , 2 μl de dNTP (2 mM chacun), 0,4 U de Phusion ADN polymérase et de l'eau ultra-pure.

Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 30 sec à 98°C, une amplification de 20 cycles comprenant une étape de dénaturation thermique à 98°C pendant 7 sec, suivie d'une étape d'hybridation des amorces à 55°C pendant 20 sec et une étape d'élongation des amorces à 72°C dont la durée dépend de la taille du fragment à amplifier (15 sec/kpb). Une élongation finale à 72°C pendant 7 min termine la PCR.

La 2^{ème} PCR se déroule dans les mêmes conditions en utilisant l'oligonucléotide générique « sens », l'oligonucléotide long « anti-sens » et 1 μl d'ADN matriciel provenant de la première PCR. Les conditions de la PCR sont identiques à celle de la première. Seul le nombre de cycles de l'amplification change : il s'agit dans ce cas de 35 cycles. La réaction PCR est vérifiée en déposant 2 μl de la PCR (+1 μl de bleu de charge) sur gel d'agarose 1% (poids/vol) contenant 1 μl de BET (10 mg/ml) pour 100 ml de tampon TAE.

Si l'amplification avec la Phusion ADN polymérase ne fonctionne pas, nous avons obtenu de bons résultats en utilisant l'Expand High Fidelity ADN polymérase (3,5 U/ μl) de *Roche Diagnostics*. La PCR est relancée dans un volume réactionnel de 20 μl selon le protocole suivant. A 1 μl d'ADN matriciel (50 ng) sont ajoutés 2 μl de tampon Expand HiFi 10X, 0,7 μl d'oligonucléotide court spécifique « sens » (10 μM), 0,7 μl d'oligonucléotide long « anti-sens » (10 μM), 2 μl de dNTP (2 mM chacun), 0,5 μl d'Expand High Fidelity ADN polymérase et de l'eau ultra-pure. En option, 2 à 5 % final de DMSO (Dimethyl sulfoxide) peuvent être ajoutés. Le DMSO inhibe les hybridations non spécifiques et déstabilise les structures secondaires de la matrice ou des oligonucléotides.

Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 4 min à 94°C. Une première amplification de 10 cycles comprenant une étape de dénaturation thermique à 94°C pendant 1 min, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C pendant 1 min. Une élongation finale à 72°C pendant 7 min termine la PCR. Une deuxième amplification de 25 cycles comprenant une étape de dénaturation thermique à 94°C pendant 1 min est

réalisée, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C pendant 1 min. A chaque cycle, le temps d'élongation est augmenté de 5 sec. Une élongation finale à 72°C pendant 7 min termine la PCR.

Purification des produits PCR

Les produits PCR sont purifiés, notamment dans le but d'éliminer les oligonucléotides *attB* et les dimères d'oligonucléotides *attB*. Les dimères d'oligonucléotides peuvent recombiner efficacement avec le vecteur donneur lors de la réaction BP et ainsi augmenter le bruit de fond après la transformation dans *E. coli*. Le kit NucleoSpin® ExtractII (*Macherey-Nagel*) est utilisé directement à partir de la réaction PCR. Avec la méthode NucleoSpin ExtractII, l'ADN (1 vol) est fixé sur une membrane en silice d'une colonne NucleoSpin ExtractII en présence de sels chaotropiques (guanidine) par centrifugation à 11000 g pendant 1 min. Les contaminants sont éliminés par une simple étape de lavage avec 600 µl de tampon NT3 contenant de l'éthanol et par centrifugation à 11000 g pendant 1 min. Une deuxième centrifugation à 11000 g pendant 2 min permet de bien sécher la membrane et d'éliminer toute trace d'éthanol. Finalement, l'ADN purifié est élué dans des conditions de force ionique faible avec 15 à 50 µl de tampon NE (Tris-HCl 5 mM pH=8,5) par une centrifugation à 11000 g pendant 1 min. Les produits PCR purifiés sont quantifiés en utilisant le spectrophotomètre ND-1000 NanoDrop® (*NanoDrop*).

Réaction BP entre le produit PCR-attB et un vecteur donneur (pDONR207) pour générer un clone d'entrée (pENTR)

La réaction BP se déroule dans un volume final de 5 µl en présence de 10 à 30 fmol de produit PCR, 100 ng de pDONR207, 1 µl de tampon BP 5X, 1 µl de BP Clonase mix et de l'eau ultra-pure.

Pour favoriser la réaction de recombinaison, le mélange est incubé toute la nuit à température ambiante. La réaction est stoppée par l'ajout de 1,2 µg de Protéinase K qui dégrade la BP Clonase en 10 min à 37°C. A ce stade, la réaction BP peut se conserver à -20°C avant transformation.

Transformation

Toutes les étapes suivantes sont effectuées en conditions stériles. La totalité de la réaction BP est ajoutée à 50µl de bactéries DH5α chimio-compétentes préalablement décongelées sur la glace. Le mélange est incubé durant 30min sur glace. Les bactéries sont transformées par choc thermique à 42°C pendant 1min. Après 2min d'incubation sur glace, 150µl de milieu LB sont ajoutés et les bactéries sont mises à incuber à 37°C pendant 1h sous agitation. Toute la réaction de transformation est étalée sur boîte de Pétri standard contenant du LB agar complétement de gentamycine (7 µg/ml final). La boîte est placée à 37°C toute la nuit.

Vérification de la présence du gène d'intérêt dans le pENTR par PCR en utilisant les oligonucléotides GWS et GWR

Deux colonies sont repiquées dans 3 ml de milieu LB complétement de gentamycine (7 µg/ml final). Après incubation à 37°C pendant la nuit sous agitation constante, ces mini-cultures pourront être utilisées pour les mini-préparations d'ADN.

La PCR se déroule selon le protocole suivant en utilisant la Taq ADN polymérase (*NEB*) (5 U/µl). La réaction se déroule dans un volume final de 20 µl en présence de 1,5 µl de la mini-culture, 2 µl de tampon Thermo Pol 10X, 2 µl de dNTP (2 mM chacun), 0,5 µl de Taq ADN polymérase, 1 µl de chacun des oligonucléotides GWS et GWR à 10 µM (Tableau 3) et de l'eau ultra-pure.

Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 4 min à 94°C. Une amplification de 20 cycles comprenant une étape de dénaturation thermique à 94°C pendant 1 min, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C dont la durée dépend de la taille du fragment à amplifier (1 min/kpb). Une élongation finale à 72°C pendant 5 min termine la PCR.

Oligonucléotide	Séquence
GWS (« sens »)	5'-GTACAAAAAAGCAGGCTTC-3'
GWR (« anti-sens »)	5'-GTACAAGAAAGCTGGGTC-3'

Tableau 3 | Séquences des oligonucléotides GWS et GWR.

La réaction PCR est vérifiée en déposant 8 µl de la PCR (+2 µl de bleu de charge) sur gel d'agarose 1% (poids/volume) contenant 1 µl de BET (10 mg/ml) pour 100 ml de tampon TAE. Si la présence de l'insert est confirmée, la mini-culture est utilisée pour une mini-préparation d'ADN suivant le protocole présenté auparavant.

Séquençage du pENTR

Les bornes des produits PCR sous-clonés, au niveau des sites de recombinaison, ainsi que le début et la fin de la séquence du gène d'intérêt, sont vérifiés en utilisant les oligonucléotides A1 et A2 (Tableau 4). Pour une réaction de séquençage, 20 µl d'ADN plasmidique à une concentration de 100 ng/µl et 10 µl de l'oligonucléotide à 2 pmol/µl sont nécessaires.

Oligonucléotide	Séquence
A1 (« sens »)	5'-TCGCGTTAACGCTAGCATGGATCTC-3'
A2 (« anti-sens »)	5'-GTAACATCAGAGATTTTGAGACAC-3'

Tableau 4 | Séquences des oligonucléotides de séquençage A1 et A2.

Réaction LR entre le clone d'entrée (pENTR) et les vecteurs destination pDEST

La réaction LR se déroule dans un volume final de 5µl et requiert 100ng de pDEST, 100ng de pENTR, 1µl de tampon LR 5X, 1µl de LR Clonase mix et du tampon TE pH=8. Le mélange est incubé au moins 1h à température ambiante. Puis la réaction est stoppée par l'ajout de 1,2µg de Protéinase K qui dégrade la LR Clonase. Ce mélange est incubé à 37°C pendant 10min. A ce stade, la réaction LR peut se conserver à -20°C pendant une semaine avant transformation.

Transformation

Le protocole est exactement le même que celui précédemment décrit pour la réaction BP, mis à part qu'il s'agit d'ajouter 1 µl de la réaction LR à 50 µl de bactéries DH5α chimio-compétentes. Si peu, ou aucune colonie, n'est obtenue, il est conseillé de transformer le reste de la réaction LR.

Vérification de la présence de la fusion en N-ter et du gène d'intérêt dans le pDEST par PCR en utilisant les oligonucléotides « sens » UP1 et « anti-sens » GWR

Deux colonies sont repiquées dans 3 ml de milieu LB complémenté de l'antibiotique adéquat et incubées à 37°C pendant la nuit sous agitation constante. Ces mini-cultures pourront être utilisées pour les mini-préparations d'ADN.

La PCR se déroule selon le même protocole que précédemment en utilisant la Taq ADN polymérase (NEB) (5 U/ μ l) et les oligonucléotides UP1 « sens » et GWR « anti-sens » (Tableau 5).

La réaction PCR est vérifiée en déposant 8 μ l de la PCR (+2 μ l de bleu de charge) sur gel d'agarose 1% (poids/volume) contenant 1 μ l de BET (10 mg/ml) pour 100 ml de tampon TAE.

Oligonucléotide	Séquence
UP1 (« sens »)	5'-GATGCGTCCGGCGTAG-3'
GWR (« anti-sens »)	5'-GTACAAGAAAGCTGGGTC-3'

Tableau 5 | Séquences des oligonucléotides UP1 et GWR.

3.1.8.2. Le clonage par enzymes de restriction

3.1.8.2.1. Digestion de l'ADN par des enzymes de restriction

Les plasmides ou les fragments PCR sont digérés par les enzymes de restriction appropriées en les incubant 4h en présence de 10 U/ μ g d'ADN à la température optimale de l'enzyme (le plus souvent 37°C) et dans le tampon adéquat proposé par le fournisseur.

3.1.8.2.2. Déphosphorylation des extrémités 5' phosphate du vecteur

Les phosphatases dites alcalines, car actives à pH alcalin, catalysent l'élimination d'un groupement phosphate en 5' d'une chaîne d'ADN. La phosphatase alcaline d'origine bovine (CIP : *calf intestinal phosphatase*) a été utilisée.

Après linéarisation du vecteur plasmidique, ce dernier est déphosphorylé par l'enzyme CIP (Quatum Appligène©) afin d'éviter une fermeture, ou auto-ligation, du plasmide sur lui-même. Lors de l'insertion du fragment d'ADN double brin à étudier, celui-ci apporte ses deux groupements phosphate en 5' et participe à deux des quatre liaisons esters qui seraient nécessaires à la formation d'un ADN double brin recombinant circulaire clos.

Le plasmide, linéarisé par une ou plusieurs enzymes de restriction, est traité à la phosphatase alcaline (10 U/ μ g d'ADN) dans un tampon 10 mM Tris-HCl pH=7,9, 1 mM EDTA et incubé à 37°C. La phosphatase est inactivée par une déprotéinisation et une précipitation à l'éthanol. Il est extrêmement important d'inactiver efficacement l'enzyme pour que celle-ci n'interfère pas avec la suite des étapes de clonage. En effet, des traces de CIP peuvent déphosphoryler le fragment PCR et empêcher toute ligation.

3.1.8.2.3. Déprotéinisation et précipitation de l'ADN

Cette opération permet l'inactivation de toutes les enzymes résiduelles. La solution contenant l'ADN est déprotéinisée en présence d'un volume de phénol/chloroforme (1/1). Les protéines précipitent à l'interface, le surnageant renfermant l'ADN ainsi purifié est prélevé. L'ADN est précipité en présence d'acétate de sodium 0,3 M pH=5 par addition de 2,5 vol d'éthanol absolu à froid (-20°C) pendant la nuit. Après centrifugation (14000 g, 30 min, 4°C), le culot est séché et repris dans un volume approprié de tampon TE.

3.1.8.2.4. Ligation

La ligation entre un fragment d'ADN et un vecteur plasmidique, tous deux digérés par les mêmes enzymes de restriction, se fait dans un rapport molaire vecteur/fragment de 1 à 5. La quantité d'ADN est mesurée par spectrométrie UV à 260nm en utilisant le spectrophotomètre ND-1000 NanoDrop® (*NanoDrop*).

La ligation est réalisée en présence d'1 U d'ADN ligase T4 dans le tampon [Tris-HCl 50 mM pH=7,6, MgCl₂ 10 mM, DTT 1 mM], pendant 4 h à 4°C. Pour 10 μ l de volume réactionnel, 100 ng de vecteur déphosphorylé sont utilisés.

Cette étape est la plus délicate du clonage. La stoechiométrie relative entre l'insert et le vecteur est déterminante, c'est pourquoi plusieurs concentrations en insert (concentration constante de vecteur) sont systématiquement testées. Parallèlement, un témoin de ligation, le vecteur linéarisé déphosphorylé, est réalisé.

3.1.8.2.5. Transformation à partir du mélange de ligation

Le milieu de ligation est utilisé pour transformer directement les bactéries DH5 α par méthode chimique. Après transformation, des fractions aliquotes sont étalées sur milieu LB solide supplémenté de l'antibiotique adéquat.

3.1.8.2.6. Obtention du vecteur de clonage et vérification de la séquence

Quelques colonies issues de la transformation sont prélevées à l'aide de pics stériles et amplifiées en milieu liquide supplémenté de l'antibiotique adéquat. L'ADN plasmidique est purifié par mini-préparation d'ADN en utilisant le kit Nucleospin® Plasmid (*Macherey-Nagel*). La présence et la taille de l'insert sont vérifiées par électrophorèse sur gel d'agarose à partir d'une digestion par des enzymes de restriction ou d'une PCR.

Afin de vérifier que l'ADN polymérase n'a pas introduit de mutation(s), la séquence de la zone amplifiée est contrôlée par séquençage des deux brins. Le séquençage a été réalisé par le service commun de séquençage de l'IGBMC, dirigé par Serge Vicaire.

3.1.9. Mutagenèse ponctuelle

La mutation souhaitée est introduite par la technique RCA (*Rolling Circle Amplification*) ou PCR en cercle tournant (Figure 36). Cette technique nécessite deux oligonucléotides complémentaires de 33 bases portant la (les) base(s) modifiée(s) correspondant à la mutation désirée en position centrale, ainsi qu'une ADN polymérase thermostable et processive comme la Pfu ADN polymérase.

La réaction se déroule dans un volume final de 50 μ l en présence de 40 ng de matrice, auquel sont ajoutés : 1 μ l de Pfu ADN polymérase, le tampon adéquat, 10 μ l de dNTP à 2 mM, 20 pmol de chaque oligonucléotide portant la mutation et de l'eau ultra-pure. En option, 2 à 10% final de DMSO peuvent être ajoutés.

Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 3 min à 94°C, suivie d'une amplification de 15 cycles comprenant une étape de dénaturation thermique à 94°C pendant 30 sec, une étape d'hybridation des amorces à 55°C pendant 1 min et une étape d'élongation des amorces à 68°C dont la durée dépend de la taille du fragment à amplifier (2 min/kpb). Un témoin négatif est réalisé sans oligonucléotide. La réaction PCR est vérifiée en déposant 10 μ l de la réaction (+5 μ l de bleu

de charge) sur gel d'agarose 1% (poids/vol) contenant 1 µl de BET (10 mg/ml) pour 100 ml de tampon TAE.

La particularité de cette technique réside dans la capacité que présente l'endonucléase *Dpn I* à discriminer l'ADN méthylé de l'ADN non méthylé. En effet, l'ADN du plasmide parental non muté, qui a servi de matrice pour la PCR, est méthylé ce dernier étant produit dans une souche comportant le gène *dam*. Ce plasmide parental est éliminé lors de la digestion par l'endonucléase *DpnI* qui ne reconnaît son site de restriction 5'-GATC-3' que s'il est méthylé au niveau de l'adénine (en position N6), ce qui n'est bien sûr pas le cas de l'ADN néo-synthétisé. Les 40 µl restants de la réaction PCR précédente, ainsi que le témoin négatif, sont donc digérés par 20 U d'enzyme *DpnI* pendant 2 h à 37°C. La digestion est vérifiée en déposant 5 µl de la réaction (+1 µl de bleu de charge) sur gel d'agarose 1% (poids/vol) contenant 1 µl de BET (10 mg/ml) pour 100 ml de tampon TAE.

La dernière étape de ce protocole consiste à transformer des bactéries DH5α par 10 µl de plasmide muté. Lors de la transformation, les cassures simple brin du plasmide muté sont réparées par la cellule.

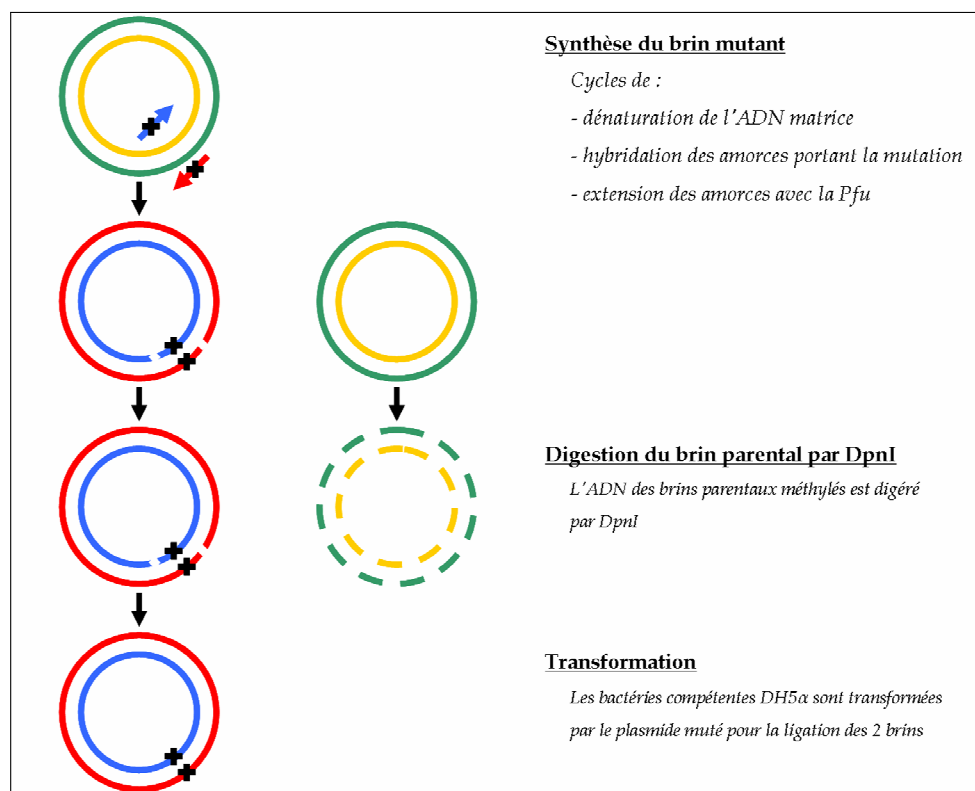


Figure 36 | Schéma expliquant la technique de mutagenèse dirigée utilisée (méthode RCA).

3.2. Techniques biochimiques

3.2.1. Culture et production dans *E. coli*

Le système d'expression utilisé est le système pET (*Novagen*) basé sur l'ARN polymérase du bactériophage T7, inductible à l'IPTG, décrit dans la Figure 37. L'IPTG (IsoPropyl β -D-1-ThioGalactopyranoside) est un analogue du lactose fréquemment utilisé comme inducteur de l'opéron lactose. L'IPTG n'étant pas métabolisé par *E. coli*, sa concentration reste constante et le taux d'expression des gènes lac p et lac o ne varie pas.

Ce système est très puissant pour exprimer les protéines recombinantes dans *E. coli*. Le gène cible est cloné dans un plasmide sous le contrôle d'un promoteur transcriptionnel du bactériophage T7. La transcription du gène cible est réalisée uniquement par l'ARN polymérase de T7, dont le gène est inséré dans le génome de la bactérie hôte par l'intermédiaire d'un dérivé lysogène du bactériophage λ , le bactériophage DE3. Le gène codant l'ARN polymérase T7 est sous la dépendance du promoteur *lacUV5* inductible à l'IPTG. De plus, le gène cible et le gène codant l'ARN polymérase T7 présentent un site Lac opérateur pour un contrôle optimal de la transcription.

En absence d'inducteur, le répresseur produit par le gène *lacI* reconnaît spécifiquement le site opérateur Lac et s'y fixe. L'encombrement de ce complexe (un tétramère) est tel que l'ARN polymérase est incapable de se fixer au site promoteur Lac, la transcription/traduction de l'ensemble des gènes étant par conséquent réprimées.

La présence d'inducteur (IPTG) induit un changement conformationnel du répresseur LacI. Ainsi modifié, le répresseur LacI est incapable de se fixer sur le site opérateur Lac, permettant la transcription, puis la traduction, de l'ARN polymérase T7. L'ARN polymérase T7 se fixe alors sur le promoteur T7 présent sur le vecteur et le gène cible est exprimé.

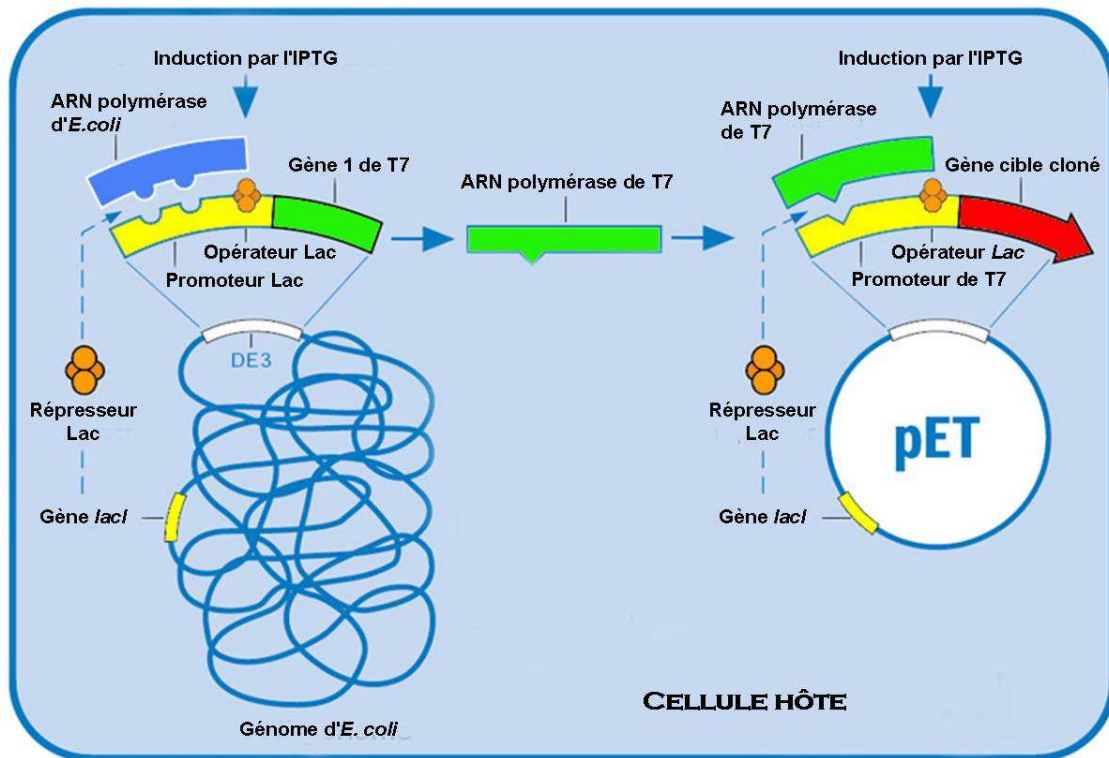


Figure 37 | Système d'expression utilisé dans *E. coli*.

3.2.1.1. Préculture

Une pré-culture de 200 ml de milieu LB, supplémenté de 200 μ l d'ampicilline, est directement inoculée avec une colonie isolée d'une boîte LB Agar obtenue après étalement d'une transformation de bactéries BL21(DE3) par le plasmide souhaité. Cette pré-culture est cultivée à 37°C sous agitation constante durant la nuit.

3.2.1.2. Culture

La pré-culture est utilisée pourensemencer 6 l de milieu LB, supplémenté de 1 ml d'ampicilline, répartis en 6 fois 1 l dans des Erlenmeyers de contenance totale égale à 5 l. En général, 25ml de pré-culture/l de milieu sont utilisés de façon à avoir une DO_{600nm} de 0,1 au moment de l'inoculation. La première phase de culture a lieu à 37°C sous agitation constante (190 rpm). La progression de la culture est suivie par lecture de la DO_{600nm} .

Pour les protéines sauvages X0 et R0, l'induction de l'expression de la protéine se fait à 37°C, pour les différents mutants elle se fait à 25°C. Pour cela, à une DO_{600nm} comprise entre 0,4 et 0,6, la température est abaissée et, une fois la température atteinte, l'IPTG est ajouté à la concentration finale de 0,5 mM.

3.2.1.3. Récolte des bactéries

Après 3 h d'induction, les bactéries sont récoltées par centrifugation (20 min à 4000 g à 4°C) et soigneusement lavées avec 50 ml de milieu LB frais. Les culots bactériens secs sont ensuite pesés, congelés et conservés à -20°C. Les rendements masse/vol moyens sont de l'ordre de 4 à 5 g de cellules/l de culture.

3.2.2. Lyse cellulaire

Les conditions de lyse sont à adapter selon les protéines étudiées. Les culots bactériens sont re-suspendus dans le volume adéquat de tampon de lyse. Les différentes conditions de lyse et de volumes sont présentées dans le tableau ci-dessous. La lyse s'effectue principalement par sonication, à l'aide d'un sonicateur de type Vibracell 72412 équipé d'une sonde de 13 mm de diamètre, durant 10 min de temps effectif par des impulsions d'ultra-sons toutes les 2 sec à 40% d'amplitude à 4°C, en limitant la température à 10°C. L'agent réducteur, ainsi que tous les inhibiteurs de protéases, sont ajoutés au mélange juste au début du processus de lyse. Lorsqu'un ligand est utilisé (*9-cisRA*), la sonication est réalisée dans l'obscurité. Une fraction aliquote correspondant à l'extrait brut est prélevée et déposée sur un gel d'acrylamide en conditions dénaturantes.

	mutants E1a...E5	mutant R1d hétérodimère R1d-X0	mutant X1c hétérodimère X1c-R0
Volume de culture	1 l	R1d : 3 l R1d : 3 l ; X0 : 2 l	X1c : 2 l X1c : 2 l ; R0 : 3 l
Volume tampon de lyse	50 ml	150 ml	150 ml
NaCl	500 mM	500 mM	150 mM
KCl			150 mM
NH₄Cl			100 mM
TrisHCl pH=8	20 mM	20 mM	20 mM
CHAPS		4 mM	
β-mercaptoéthanol	5 mM	5 mM	
Imidazole	5 mM		
DIFP	0,1 mM	0,1 mM	0,1 mM
glycérol		10%	10%
C12M		0,1%	
PIC (<i>Protease Inhibitor Cocktail</i>)			au 1/100 ^{ème}
9-cisRA (10⁻²M)		500 µl	500 µl

Tableau 6 | Composition des différents tampon de lyse.

3.2.3. Ultracentrifugation

L'extrait brut, issu de la sonication, est ensuite clarifié par ultracentrifugation pendant 2 h à 45000 g à 4°C. Le surnageant correspondant à la fraction soluble est maintenu à 4°C et immédiatement utilisé. Après ultracentrifugation, le culot obtenu est re-suspendu dans un volume équivalent pour obtenir la fraction culot. Deux fractions aliquotes, correspondant à l'extrait soluble et au culot, sont prélevées et sont déposées sur un gel d'acrylamide en conditions dénaturantes.

3.2.4. Chromatographies sur colonnes

Les étapes de purification par chromatographie ont été réalisées sur un système de chromatographie FPLC (*Fast Performance Liquid Chromatography*) BioLogic Workstation (BioRad).

3.2.4.1. Chromatographie d'affinité sur résine de chélation de cations métalliques

Deux types de colonnes ont été utilisés pour cette chromatographie d'affinité : la HiTrap Chelating (GE Healthcare) et la résine Talon (Clontech).

Le principe de cette chromatographie, dite IMAC (*Immobilized Metal Affinity Chromatography*), est basé sur l'emploi de cations métalliques divalents (Co^{2+} , Zn^{2+} , Ni^{2+} , Cu^{2+}) chélatés par des groupements de type imido-acétates, greffés sur une matrice de type sépharose pour séparer des protéines portant des étiquettes polyhistidines. Ces cations immobilisés vont interagir avec les polyhistidines de la protéine produite et permettre de séparer rapidement et efficacement la protéine de fusion du reste des protéines bactériennes endogènes. Après une étape de lavage, la protéine d'intérêt peut être séparée du support en utilisant l'imidazole comme agent d'élution.

3.2.4.1.1. Colonnes HiTrap Chelating HP de 5ml

Ces colonnes sont en polypropylène, composé biocompatible et inerte pour les protéines à purifier. Elles sont remplies de Sépharose Haute Performance. Cette résine consiste en des billes d'agarose fortement réticulées auxquelles l'acide imido-diacétique a été couplé par des groupements éther stables, *via* un bras espaceur de sept atomes. Le gel est stable dans une large gamme de pH (de 2 à 14) et supporte tous les tampons couramment utilisés. Le débit de charge et d'élution est de 5 ml/min.

Les colonnes sont fournies dépourvues d'ions métalliques et doivent être chargées avec l'ion désiré avant utilisation, ici le nickel. Après la charge de l'extrait soluble et le lavage, l'élution des protéines recombinantes retenues sur le gel peut être réalisée par différentes méthodes : modification du pH, agent chélateur ou molécule compétitrice. La protéine est éluée soit par un gradient d'imidazole dans le cas des différents mutants de hERRY, soit par paliers dans le cas du mutant R1d.

Mode opératoire pour la purification des mutants E1a à E5

Après dépôt de l'extrait soluble, on procède comme suit :

- la colonne est lavée sur 10 volumes de colonne (CV) avec 100% de tampon A (Tableau 7),
- puis un deuxième lavage sur 10 CV avec 98% tampon A/2% tampon B (15 mM Imidazole) afin d'éliminer les protéines fixées de manière non spécifique,
- un gradient linéaire de 15 mM à 500 mM d'Imidazole sur 30 CV est appliqué,

	Tampon A	Tampon B
NaCl	500 mM	500 mM
TrisHCl pH=8	20 mM	20 mM
Imidazole	5 mM	500 mM
β-mercaptoéthanol	5 mM	5 mM
DIFP	0,1 mM	0,1 mM

Tableau 7 | Composition des tampons A et B pour la purification des mutants E1a à E5.

La colonne est régénérée dans une solution d'EDTA (EDTA 50 mM et NaCl 500 mM) sur 10 CV.

Mode opératoire pour la purification du mutant R1d

Après dépôt de l'extrait soluble, on procède comme suit :

- la colonne est lavée sur 20 CV avec 95% tampon A/5% tampon B (50 mM Imidazole) (Tableau 8) afin d'éliminer les protéines fixées de manière non spécifique,
- l'élution de la protéine n'est pas réalisée par un gradient d'Imidazole mais par paliers. Un premier palier est réalisé sur 14 CV avec 85% tampon A/15% tampon B (150 mM Imidazole), puis un deuxième sur 10 CV avec 70% tampon A/30% tampon B (300 mM Imidazole).

	Tampon A	Tampon B
NaCl	500 mM	500 mM
TrisHCl pH=8	20 mM	20 mM
Imidazole		1 M
CHAPS	4 mM	4 mM
C12M	0,1%	0,1%
glycérol	10%	10%
β-mercaptoéthanol	5 mM	5 mM
DIFP	0,1 mM	0,1 mM

Tableau 8 | Composition des tampons A et B pour la purification du mutant R1d.

3.2.4.1.2. Colonne avec la résine Talon

La résine Talon (*Clontech*) est une résine IMAC chargée avec des cations Co^{2+} . L'avantage de cette résine réside dans la nature du cation fixé et dans le type de groupement chélatant. Le cobalt interagit moins fortement avec les protéines à polyhistidines, la fixation est moins forte, mais plus spécifique. Il y a donc moins de fixation aspécifique et certains contaminants sont mieux séparés. De plus, l'élution se fait à des concentrations plus faibles en imidazole, ce qui est plutôt favorable, en particulier dans notre cas. En effet il a été remarqué que l'imidazole pouvait, dans certains cas, entraîner l'agrégation et la précipitation des protéines. Enfin, le type de groupement chélatant l'ion métallique réduit le risque de co-élution accidentelle de celui-ci avec la protéine. Ce dernier point est très important car ce phénomène a été observé pour d'autres résines chélatant les métaux et avec les autres types de cations métalliques (cobalt y compris). Ces cations sont une source potentielle d'hétérogénéité de la protéine (en particulier si le peptide polyhistidine n'est pas éliminé). Ils peuvent induire une agrégation (cas du zinc), voire des modifications non covalentes ou covalentes (cas du nickel, du cuivre et du cobalt).

Le cobalt est particulièrement sensible à l'oxydation et ce type de résine ne tolère pas des concentrations élevées en agents réducteurs. Le DTT est à éviter (pas plus de 3 mM) et l'utilisation du β -mercaptoéthanol à faible concentration (5 mM) est préférable. Des concentrations trop élevées en agents réducteurs provoquent généralement un brunissement de la résine par réduction du cation métallique en métal, et une élution prématurée de la

protéine en présence de traces de cations plus ou moins fortement associés à la protéine (phénomène d'autant plus marqué lorsque l'étiquette polyhistidine est encore présente). Le cation métallique est alors généralement très difficile à éliminer même en utilisant des agents chélateurs à haute concentration (EDTA ou citrate).

Mode opératoire pour la purification du mutant X1c

Cette résine ne demande pas d'être préparée : elle est simplement lavée dans l'eau ultra-pure pour être prête à l'emploi. On utilise une colonne de volume effectif de 2,5 à 5 ml. Le débit de charge et d'élution est de 5 ml/min. Après dépôt de l'extrait soluble, on procède comme suit :

- la colonne est lavée avec 95% tampon A/5% tampon B (50 mM imidazole) (Tableau 9), ce qui permet d'éliminer les protéines liées de manière non spécifique.
- la protéine fusionnée ou non à l'étiquette polyhistidine est éluée sélectivement avec 85% tampon A/15% tampon B (soit 150 mM imidazole) puis avec 70% tampon A/30% tampon B (soit 300 mM imidazole). Il ne s'agit pas d'une élution en gradient, mais par paliers

La colonne est nettoyée avec du tampon B, puis du tampon A. La colonne est réutilisable 2 à 3 fois.

	Tampon A	Tampon B
NaCl	150 mM	150 mM
TrisHCl pH=8	20 mM	20 mM
KCl	150 mM	150 mM
NH₄Cl	100 mM	100 mM
Imidazole	5 mM	1 M
glycérol	10%	10%
DIFP	0,1 mM	0,1 mM

Tableau 9 | Composition des tampons A et B pour la purification du mutant X1c.

3.2.4.2. Chromatographie d'exclusion moléculaire: Superdex 200 HR 16/60 (*Pharmacia*)

Cette technique, plus communément appelée filtration sur gel, permet de séparer les protéines selon leur taille et leur forme, par des granules de gel poreux. Les grosses protéines (dont le diamètre est supérieur à celui des pores) sont exclues et par conséquent éluées les premières, au niveau du volume mort (ou volume d'exclusion). Les petites et moyennes protéines voient leur migration ralentie car elles progressent à travers les billes et seront éluées plus tardivement. Les protéines sont donc éluées dans l'ordre inverse des masses moléculaires.

Il existe une relation linéaire entre le volume d'éluion et le logarithme de la masse moléculaire de la protéine. La fonction reliant le poids moléculaire de la protéine et le volume d'éluion sur une colonne d'exclusion moléculaire est la suivante :

$$\text{Log (PM)} = f (K_{AV})$$

avec K_{AV} , le coefficient de partage entre la phase liquide et la phase gel

$$K_{AV} = (V_e - V_m) / (V_t - V_m)$$

avec V_e , le volume d'éluion ; V_t , le volume total et V_m , le volume mort.

Pour éviter tout risque de contamination, certaines colonnes sont réservées aux purifications en absence de ligand, d'autres en présence de ligand. Les conditions de filtration sur gel, différentes selon la protéine, sont indiquées dans le Tableau 10 ci-dessous.

	mutants E1a ... E5	mutant R1d	mutant X1c
NaCl	500 mM	150 mM	100 mM
TrisHCl pH=8	50 mM à pH=8	10 mM à pH=7,5	10 mM à pH=8
CHAPS		2 mM	
β-mercaptoéthanol			28 mM
C12M		0,1%	
glycérol		2%	
DTT		5 mM	

Tableau 10 | Composition des différents tampons de filtration sur gel.

Avant d'être injecté sur la colonne de filtration sur gel, l'échantillon est concentré à l'aide d'une unité *Centriprep* jusqu'à un volume inférieur à 3 ml.

3.2.5. Incubation avec l'acide rétinoïque

Le *9-cisRA* est un composé se présentant sous la forme de paillettes cristallines jaunes. C'est un composé dangereux car potentiellement cancérigène et tératogène. De plus, il est extrêmement sensible à la lumière ultraviolette et à l'oxydation. Toutes les étapes de purification (et de cristallisation) effectuées en présence d'acide rétinoïque ont donc été réalisées dans l'obscurité. Ce dérivé de la vitamine A étant très hydrophobe et très peu soluble dans l'eau, il ne se présente sous forme concentrée qu'en solution saturée dans l'éthanol. La concentration molaire moyenne à saturation est de l'ordre de 25 mM. Ceci pose des problèmes lors de l'étape de saturation de la protéine en acide rétinoïque, l'alcool induisant toujours une légère agrégation ainsi qu'une précipitation.

Un volume de solution éthanolique de *9-cisRA* n'excédant pas 2% du volume total de la solution de protéine est progressivement ajouté à la protéine afin de saturer le récepteur en ligand sans trop provoquer de dénaturation liée à l'éthanol. Ce mélange est incubé à 4°C dans l'obscurité pendant 1 à 2 h.

3.2.6. Protéolyse ménagée par la thrombine

L'élimination, par protéolyse ménagée à la thrombine, de l'étiquette polyhistidine (ou de la fusion située à l'extrémité N-ter) des protéines surproduites et purifiées est réalisée en utilisant de la thrombine bovine plasmatique (*Sigma Biochemicals*). L'enzyme est conditionnée dans un tampon Tris-HCl 20 mM pH=8 et glycérol 30% (vol/vol) à la concentration de 0,25 U/ μ l et conservée à -20°C.

Cette protéase à sérine requiert la présence de cations calcium divalents ($[Ca^{2+}] = 5$ mM) et un pH compris entre 8,0 et 8,5. Elle est susceptible d'être inhibée par les agents chélatants (tels que l'EDTA) ou complexants (tels que l'imidazole) ainsi que par les inhibiteurs classiques de protéases à sérine (DIFP). Il s'est avéré que cette enzyme est plus résistante aux inhibiteurs de protéases à sérine qu'il n'y paraît : l'ajout de calcium n'est pas nécessaire et des concentrations en imidazole de l'ordre de 50 mM ne semblent pas inhiber l'enzyme. Dans notre cas, nous n'effectuons pas de traitement particulier pour inactiver l'enzyme.

3.3. Cristallisation

3.3.1. La plate-forme de cristallisation

La plate-forme de cristallisation du DBGS dispose de plusieurs outils d'aide à la cristallisation. Les tests de cristallisation du mutant E1a ont été réalisés à l'aide d'un robot de cristallisation *Tecan Workstation 150*. Ce robot permet de réaliser des gouttes (1 μ l de protéine + 1 μ l de la solution de criblage) en boîtes 96 conditions en une vingtaine de minutes. Depuis, le DBGS a fait l'acquisition d'un nouveau robot de cristallisation, *Cartesian Honeybee 8+1*, permettant de réaliser des nanogouttes de l'ordre de 100 à 500 nl de protéine par condition. Avec ce nouveau robot, 18 μ l de protéine sont suffisant pour réaliser une boîte de 96 conditions avec des gouttes de 100 nl. Une tête 96 supports de cônes a été rajoutée à l'ancien robot (*Tecan Workstation 150*) permettant ainsi le remplissage des réservoirs d'une boîte 96 conditions en moins d'1 min. Nous disposons aussi d'une station de pipetage (*Tecan Miniprep-75*) permettant de préparer 96 conditions de cristallisation en moins de deux heures (Figure 38).

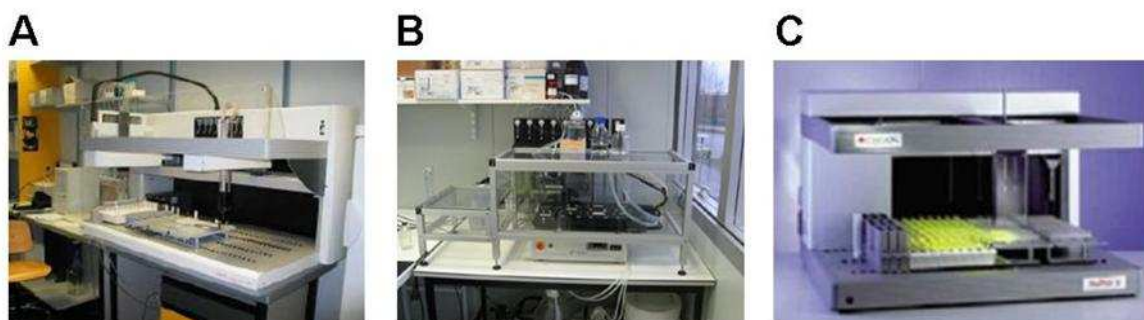


Figure 38 | Robots disponibles sur la plate-forme de cristallisation. (A), *Tecan Workstation 150*. (B), *Cartesian Honeybee 8+1*. (C), *Tecan Miniprep-75*.

3.3.2. Généralités sur la cristallisation

La solubilité d'une macromolécule est fonction de nombreux paramètres tels que la concentration de la macromolécule, le pH, la température, la force ionique, l'effet d'additifs etc. La cristallisation est un compromis entre les facteurs thermodynamiques (solubilité) et cinétiques (nucléation et croissance cristalline). Lors de la cristallisation, la solubilité de la macromolécule diminue jusqu'à atteindre l'état de nucléation favorable à l'obtention de cristaux, sans toutefois tomber dans la zone de précipitation (Figure 39).

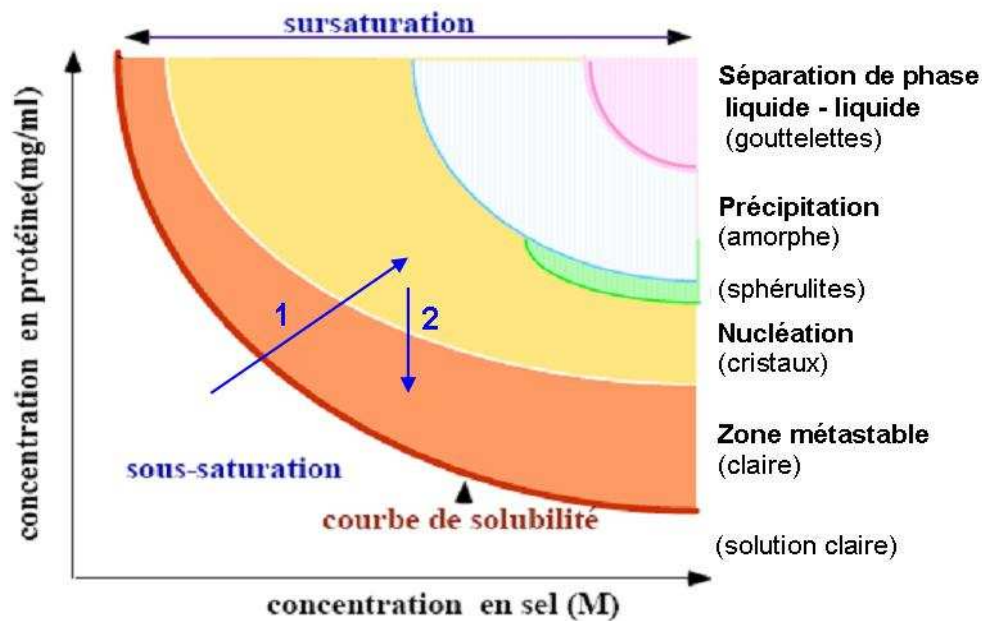


Figure 39 | Diagramme des phases. Les trois étapes de la cristallisation : la sursaturation et la nucléation en 1 et la croissance cristalline en 2.

La méthode utilisée pour les cristallisations se base sur la technique de diffusion de vapeur en goutte assise. L'équilibre s'établit entre la goutte et le réservoir par diffusion des espèces volatiles jusqu'à ce que la tension de vapeur soit la même dans la goutte et dans le réservoir. Dans la plupart des cas, la goutte est plus diluée que le réservoir ce qui entraîne une concentration de la goutte. La protéine qui ne diffuse pas reste dans la goutte et se trouve elle aussi concentrée. L'état de sursaturation s'établit donc progressivement en augmentant simultanément la concentration de la protéine et de l'agent précipitant, ce qui peut conduire à l'obtention de cristaux.

La recherche des conditions de cristallisation s'est effectuée en utilisant plusieurs kits (*screens*) de criblage commerciaux distribués par *Hampton Research* et *Qiagen*. Ces kits contiennent différentes conditions de cristallisation permettant de tester très rapidement un grand nombre de conditions différentes, augmentant ainsi les chances d'obtention de cristaux. Les kits de *Hampton Research* utilisés sont les suivants : *Index™*, *SaltRX™*, *Natrix™* et *Membfac™*, ceux de *Qiagen* sont *The Anions*, *The Cations*, *The Ammonium Sulfates*, *The Sparse Matrix 1*, *The Classics*, *The JCSG+ screen* et *The ProComplex screen*. D'autres kits développés par notre laboratoire, sont aussi disponibles, tels que *The PEG-Ion pH*, *The PEG-Ion Classic* et deux kits dédiés spécialement aux RNs : *NR-LBD 48* et *NR-LBD 96* (Zeyer et coll.). Ces différentes conditions de criblage pourront donner, soit des cristaux, soit

des indications sur le comportement de la protéine vis-à-vis du type d'agent précipitant utilisé, du pH optimal, de la concentration de la protéine et de la température optimale. Toutes ces informations seront utilisées pour optimiser les conditions de cristallisation dans le but d'améliorer les microcristaux obtenus ou d'obtenir des cristaux.

Les essais de cristallisation ont été effectués, soit en boîte *Low Profile (Greiner)* 96 puits avec une protéine, soit en boîte *Corning 3.1* avec la même protéine à trois concentrations différentes. La cinétique de la diffusion dépend des différences de concentration dans la goutte et dans le réservoir mais également de la géométrie du système, du volume de la goutte et de sa distance au réservoir (DeTitta & Luft, 1995). Les gouttes sont formées du mélange de la protéine avec la solution du réservoir dans les rapports 1:1.

3.4. Caractérisation des échantillons

3.4.1. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes

3.4.1.1. Tampons et solutions

3.4.1.1.1. Gel de séparation à 12%

Pour 1 gel, la composition est la suivante : 1,05 ml de Tris-HCl 1,5 M pH=8,8 et SDS 0,4% (poids/vol), 1,25 ml d'Acrylamide/Bis-Acrylamide 40/0,8 (poids/vol), 42,35 µl de persulfate d'ammonium (APS) 10% (poids/vol), 4,15 µl de TEMED et 1,8 ml d'eau ultra-pure.

3.4.1.1.2. Pré-gel de concentration à 5%

Pour 1 gel, la composition est la suivante : 530 µl de Tris-HCl 0,5 M pH=6,8 et SDS 0,4% (poids/vol), 250 µl d'Acrylamide/Bis-Acrylamide 40/0,8 (poids/vol), 20 µl de persulfate d'ammonium (APS) 10% (poids/volume), 2 µl de TEMED et 1,28 ml d'eau ultra-pure.

3.4.1.1.3. Tampon dissociant et de charge concentré 5X

Ce tampon est composé de 250 mM Tris-HCl pH=6,8, glycérol 50% (vol/vol), SDS 10% (poids/vol), 140 mM β-mercaptoéthanol, 100 mM DTT, bleu de bromophénol 0,05%

(poids/vol) et d'eau ultra-pure. Les échantillons mélangés dans le tampon dissociant sont dénaturés par chauffage dans un bloc chauffant 5 min à 100°C.

3.4.1.1.4. Tampon de migration 5X

Un litre de tampon de migration 5X contient 30 g de Tris-Base, 188 g de glycine et 10 g de SDS. La migration s'effectue à une intensité constante de 250 V et 80 mA.

3.4.1.1.5. Révélation au Bleu de Coomassie

Le gel est démoulé de son support, mis à fixer et colorer par immersion dans une solution de coloration au bleu de Coomassie pendant 15 min à température ambiante et sous agitation. La solution stock de bleu de Coomassie concentrée 2X contient pour 200 ml : 4 g de bleu de Coomassie, 120 ml d'éthanol et 80 ml d'eau ultra-pure. La solution de fixation et de coloration au bleu de Coomassie contient pour 100 ml : 50 ml de la solution stock de bleu de Coomassie concentrée 2X, 10 ml d'acide acétique et 40 ml d'eau ultra-pure. Le gel est ensuite décoloré dans la solution de décoloration pour révéler les protéines précédemment fixées et colorées. La solution de décoloration contient 10% d'acide acétique, 30% d'éthanol (vol/vol) et de l'eau ultra-pure.

3.4.2. Dosage des acides nucléiques et des protéines

Les acides nucléiques sont dosés par l'utilisation du spectrophotomètre NanoDrop® ND-1000 (*NanoDrop*). L'avantage de cet appareil réside dans l'utilisation d'un petit volume d'échantillon (de l'ordre du microlitre) pour mesurer la concentration en acides nucléiques (ADN/ARN), en sachant qu'à une unité de DO_{260nm} correspond une concentration d'ADN double brin de 50 ng/ μ l. Une lecture de l'absorbance à 260nm et 280nm permet de déterminer la pureté de l'ADN. Si l'ADN est pur, le rapport des $DO_{260nm/280nm}$ doit être compris entre 1,8 et 2. Plus élevé, il indiquerait une contamination par des ARNs. S'il est plus faible, il indiquerait une contamination par les protéines (280nm) ou par du phénol (270nm). Les protéines sont dosées soit avec le spectrophotomètre NanoDrop® ND-1000, en utilisant les coefficients d'extinction molaire linéique calculés à partir de la composition en acides aminés, soit par la méthode de Bradford en dosage spectrophotométrique à 595nm en utilisant la sérum albumine bovine pour établir la courbe d'étalonnage.

3.4.3. La spectrométrie de masse

Les macromolécules biologiques à analyser sont placées dans un tampon volatil (acétate d'ammonium 50 mM) et le principe de l'analyse est le suivant : l'échantillon est ionisé en solution par bombardement électronique avant de passer en phase gazeuse. Une série de lentilles va ensuite focaliser le faisceau, puis les ions vont évoluer en fonction de leur rapport masse/charge au sein du TOF (*Time Of Flight*) où règne le vide poussé. Les ions de masse élevée auront un temps de vol inférieur aux ions de masse faible, ce qui permet leur séparation. Un traitement informatique des mesures du rapport masse/charge permet de calculer la masse de la macromolécule qui a généré ces ions par fragmentation. Nous effectuons la plupart de nos caractérisations sur un ESI-MS (*Electron Spray Ionisation Mass Spectrometer*), un électrospray à temps de vol dont le fonctionnement est schématiquement représenté sur la Figure 40.

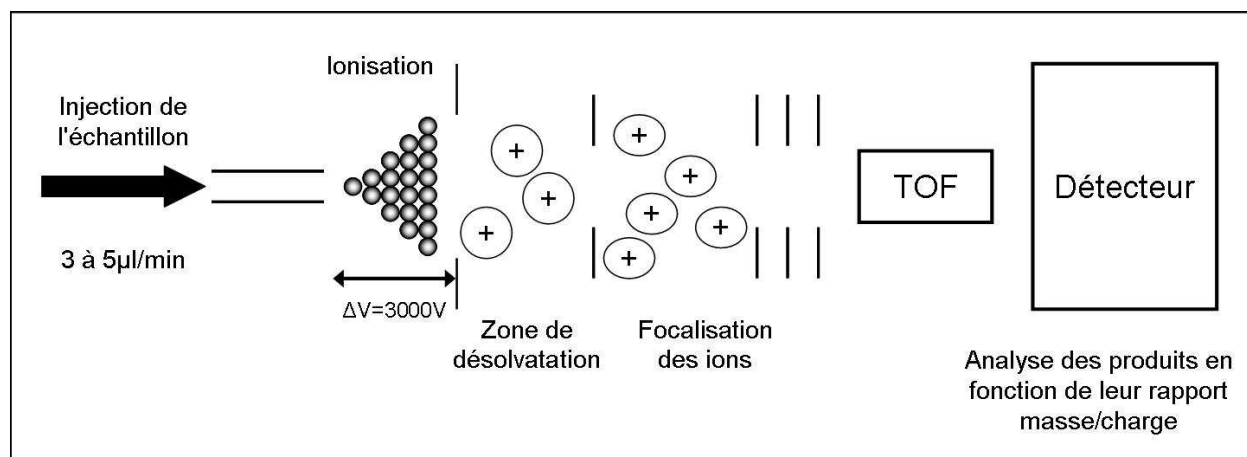


Figure 40 | Principe de fonctionnement du spectromètre électrospray à temps de vol.

L'analyse par spectrométrie de masse peut être effectuée en conditions dénaturantes (pour des protéines concentrées à 5 µM) ou natives (pour des protéines concentrées à 20 µM). En conditions dénaturantes, l'ensemble des liaisons non covalentes est dénaturé. En conditions natives, les liaisons ioniques spécifiques sont conservées ce qui permet d'étudier l'état d'oligomérisation des macromolécules ainsi que leur interaction potentielle avec des substrats, tels que des ligands ou des peptides. Les mesures de spectrométrie de masse ont été réalisées à l'IGBMC par Hélène Nierengarten.

3.5. Techniques de caractérisation fonctionnelle

3.5.1. Expérience de retard sur gel

3.5.1.1. Principe

Plus communément appelé EMSA (*ElectroMobility Shift Assay*), le retard sur gel est la technique la plus simple pour étudier une interaction entre un acide nucléique marqué et des protéines. Elle repose sur le fait qu'un complexe ADN-protéine migrera différemment dans un gel non-dénaturant que chaque entité séparée (ADN ou protéine). Ce retard de migration permet de déterminer rapidement si une séquence particulière d'ADN a été reconnue et liée par une protéine. La spécificité de l'interaction ADN-protéine peut être vérifiée en incubant le(s) complexe(s) formé(s) avec un excès de sonde froide, de l'ADN non-spécifique et un anticorps reconnaissant la protéine. Le complexe ternaire ADN-protéine-anticorps aura une migration encore plus retardée qualifiée de *supershift*. Il est cependant possible qu'un anticorps qui fonctionne très bien en Western Blot soit inefficace en *supershift*, en raison d'une moins bonne accessibilité de l'épitope lorsque la protéine est fixée sur un acide nucléique.

3.5.1.2. Marquage radioactif d'un oligonucléotide au [γ - 32 P] ATP

La T4 polynucléotide kinase catalyse le transfert du phosphate de [γ - 32 P] ATP sur l'extrémité 5' de l'oligonucléotide synthétique (les oligonucléotides sont synthétisés sans phosphate à leur extrémité 5'). Le protocole suivant permet le marquage de 10 pmoles d'oligonucléotide synthétique. La réaction de transfert du phosphate radioactif se réalise en mélangeant 2 μ l de l'oligonucléotide (10 pmoles/ μ l), 3 μ l de tampon 10X, 4 μ l de [γ - 32 P] ATP (Activité Spécifique : 3000 Ci/mmoles) soit 5 pmoles/ μ l, 11 μ l d'eau ultra-pure et 1 μ l de T4 polynucléotide kinase (10 U/ μ l). Le mélange est incubé 90 min à 37°C. Par la suite, sont rajoutés 3 μ l de NaCl 1 M, puis 120 μ l de NaCl 100 mM. La sonde radioactive est ensuite purifiée pour éliminer la radioactivité non incorporée par chromatographie. La purification est réalisée sur une résine d'exclusion Sephadex G50. Cette purification permet d'éliminer environ 90% de la radioactivité non incorporée. La radioactivité de l'oligonucléotide est ensuite mesurée sur un compteur de scintillation LS 6500 (*Beckman*). Le produit final est utilisé immédiatement ou conservé à -20°C.

3.5.1.3. Transcription-Traduction *in vitro*

Le système de Transcription-Traduction TnT® T7 (*Promega*) est utilisé pour transcrire et traduire *in vitro* les gènes clonés en aval du promoteur T7 de l'ARN polymérase dans le vecteur pSG5 (Annexe 4). Les plasmides suivants sont utilisés comme matrice : pSG5-X0, pSG5-X1a, pSG5-X1b, pSG5-X1c, pSG5-X3, pSG5-X4a, pSG5-X5, pSG5-R0. A 40 µl de mélange réactionnel (*TnT® T7 Quick Master Mix*) sont ajoutés 1 µl de méthionine 1 mM, 1 µg d'ADN plasmidique matrice et de l'eau ultra-pure pour un volume final de 50 µl. Le milieu de réaction est incubé à 30°C pendant 1 h 30 min. Deux témoins sont réalisés : le premier contient uniquement le *TnT® T7 Quick Master Mix* et de l'eau ultra-pure alors que le deuxième contient le *TnT® T7 Quick Master Mix*, le plasmide pSG5 et de l'eau ultra-pure. Puis, 50 µl de KCl 1 M sont ajoutés pour obtenir 100 µl de protéine à une concentration finale de KCl de 500 mM. L'intégrité des protéines est vérifiée en réalisant la même réaction en présence de méthionine-³⁵S, sans rajouter de KCl en fin de réaction. Dans ce cas, une fois la transcription-traduction *in vitro* terminée, une fraction aliquote de 5 µl de chaque réaction est ajoutée à 20 µl de tampon dissociant 5X. Ces échantillons sont dénaturés par chauffage durant 5 min à 90°C. Une fraction aliquote de 15 µl des échantillons dénaturés est chargée sur un gel d'acrylamide en conditions dénaturantes 12,5%. Les protéines du gel sont ensuite transférées par Western Blot en seulement 7 min en utilisant le système iBlot (*Invitrogen*), sur une membrane de nitrocellulose, qui sera autoradiographiée.

3.5.1.4. Retard sur gel

Les RNs, en quantité équimolaire (2 µl de chaque hétérodimère ou 4 µl pour un homodimère) à une concentration finale en KCl de 100 mM, sont incubés pendant 15 min dans la glace en présence de 13 µl de tampon de fixation [50 mM Tris-HCl pH=8,0, 0,5 mM EDTA, 2 mM DTT, 25% glycérol] contenant 2 µg de polydIdC.

Les compétitions sont réalisées en présence d'un excès d'oligonucléotide double brin non marqué de 1000 fois (10 pmol). Les *supershifts* sont obtenus en présence de 1 µl d'anticorps monoclonal de souris dirigé contre la protéine. Dans le cadre de l'étude de l'effet du ligand sur la dimérisation, 1 µl d'acide rétinoïque *9cis* à 10⁻⁴ M a été rajouté aux protéines. Les compétitions, les *supershifts* et le ligand sont incubés pendant 15 min supplémentaires dans la glace. Puis, 10 fmoles (3 µl) d'oligonucléotide double brin marqué au [³²P] ATP

(75000 cpm) sont ajoutés aux différentes réactions et incubés 15 min à température ambiante. Finalement, à ces différentes réactions, sont rajoutés 2 µl de bleu de dépôt (bleu de bromophénol, 60% glycérol) en faisant bien attention à ne pas perturber le mélange réactionnel.

L'électrophorèse des complexes ADN-protéines est effectuée en gel de polyacrylamide acrylamide/bis-acrylamide 30/0,8 (poids/vol) dans un tampon TBE 0,5X, à voltage constant (200 V, 20 mA) pendant 2 h avec une pré-migration du gel de 2 h à 130 V, 25 mA et 100 W. Le gel est alors transféré sur papier filtre (*Whatman*), séché sous vide à 75°C pendant 1 h puis autoradiographié en présence de film *Kodak BioMax MR* à une température de -80°C pendant la nuit avant le développement du film.

3.5.2. Test de transactivation en cellules HEK 293 EBNA

3.5.2.1. Culture

Les cellules adhérentes HEK 293 EBNA sont cultivées en milieu Dulbecco (DMEM additionné de 1 g/L de glucose) avec 10% de sérum de veau foetal, gentamycine et 1 mg/ml de G418 (Généticine) dans des flasques de 75 cm² à 37°C et 5% CO₂ en atmosphère humide. Les cellules sont régulièrement passées afin qu'elles n'atteignent jamais une confluence de 100%. Pour ce faire, le milieu est aspiré et les cellules sont rincées avec 10 ml de tampon PBS (*Phosphate Buffer Saline*) stérile 1X. Le PBS est aspiré avant d'ajouter 2 ml de trypsine diluée, en recouvrant l'ensemble des cellules. On laisse agir la trypsine durant 1 min, après en avoir aspiré l'excès. La trypsine, en détruisant les protéines de liaison, brise les liens existant entre les cellules et permet de dissocier le tapis cellulaire. Les cellules sont décollées par quelques chocs sur les côtés de la flasque, puis reprises dans 20 ml de milieu de culture. Cet ajout en excès du milieu de culture inhibe l'action de la trypsine. Les cellules sont finalement diluées au 1/5^{ème} pour obtenir une confluence de 90% deux jours plus tard.

3.5.2.2. Transfection

3.5.2.2.1. Ensemencement en plaque de culture 24 puits

Les cellules d'une flasque de 75 cm² arrivées à confluence sont reprises dans 10 ml de milieu, après rinçage au PBS 1X et traitement à la trypsine. Les cellules sont ensuite comptées sur une cellule Neubauer afin d'inoculer les puits d'une plaque de culture avec 500 µl à une

concentration de 200 000 cellules/ml en veillant à obtenir une bonne répartition des cellules sur toute la surface des puits. Pour compter les cellules, une fraction aliquote de 20 μ l de la suspension cellulaire est diluée avec 20 μ l d'une solution de bleu de trypan. Le bleu de trypan, a tendance à pénétrer dans les cellules qu'il rencontre. Une fois dans la cellule, un mécanisme d'exclusion va éjecter cette molécule dans le milieu extérieur. Ce mécanisme nécessitant de l'énergie, seules les cellules possédant une source d'ATP peuvent le mettre en place. Ainsi, les cellules viables expulseront la molécule et resteront réfringentes, petites et rondes au microscope. Au contraire, les cellules mortes n'auront pas les moyens de la rejeter et resteront bleues, gonflées et de plus en plus étalées. Le nombre total de cellules viables/ml peut donc être déterminé par la formule $N \times B \times 10^4$ où N représente le nombre moyen de cellules non colorées comptées par carré et B le facteur de dilution en bleu de trypan. La plaque de culture est incubée 24 h à 37°C pour que les cellules soient entre 50 et 60% de confluence le jour de la transfection.

3.5.2.2.2. Protocole de transfection

La transfection (Figure 41) est réalisée en utilisant un agent transfectant, le jetPIE™ (*Qbiogen*). Il s'agit d'un cation qui compacte l'ADN et forme des complexes chargés positivement qui sont capables d'interagir avec les protéoglycanes anioniques de la surface cellulaire et d'entrer dans la cellule par endocytose. Une fois dans l'endosome le réactif retient les protons, ce qui tamponne le pH endosomal et évite la dégradation de l'ADN. L'influx de protons induit ensuite la rupture de l'endosome et l'ADN est libéré dans le cytoplasme.

Pour une transfection, 0,5 μ l à 500 ng/ μ l de plasmide rapporteur (par exemple UAS-Luciférase) (soit 250 ng), 1 μ l à 25ng/ μ l de plasmide d'expression DBD-Gal4-LBD (soit 25 ng), 0,5 μ l à 100 ng/ μ l de plasmide exprimant LacZ (PCH110) (soit 50 ng), complété à 1 μ g d'ADN par du plasmide *carrier* (pBluescript) sont dilués dans un volume final de 50 μ l de NaCl 150 mM.

A ce mélange d'ADN, sont ajoutés 2 μ l de jetPIE™ dilué dans 50 μ l de NaCl 150 mM. La solution est vortexée 10 sec, brièvement centrifugée et incubée 30 min à température ambiante. La totalité de cette solution ADN/jetPIE™ (100 μ l) est transférée goutte-à-goutte dans un puits contenant le milieu de culture et les cellules à une

confluence de 50 à 60%. La plaque de culture est incubée à 37°C et 5% CO₂ en atmosphère humide. Les cellules sont rincées avec 500 µl de tampon PBS stérile 1X, 10 h après la transfection, puis 500 µl de milieu frais sont ajoutés, avec ou sans ligand (10⁻⁶ M finale). La plaque de culture est remise à incuber à 37°C. Les cellules sont récoltées après 24 h.

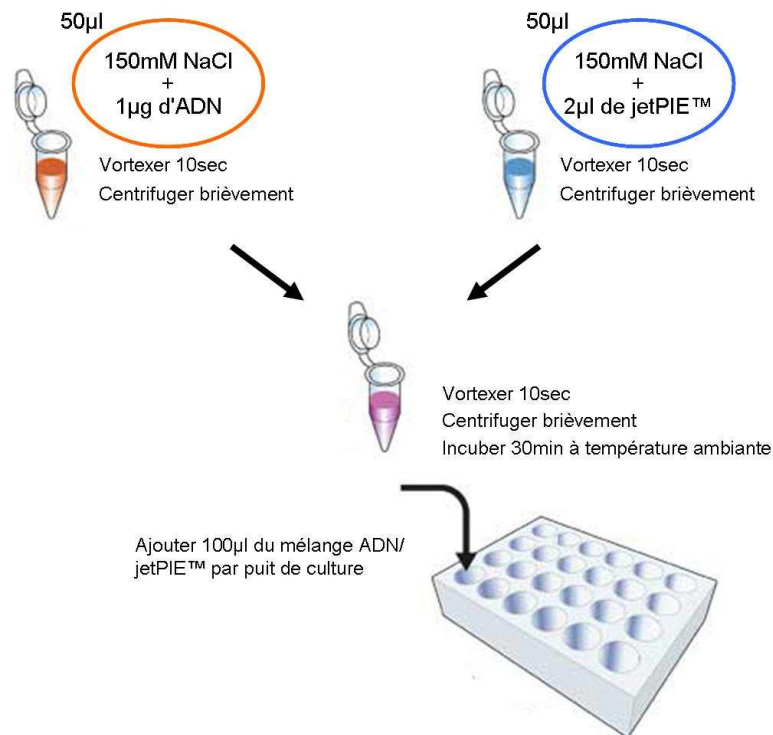


Figure 41 | Protocole de transfection (adaptée de *Polyplus-transfection*).

3.5.2.2.3. Récolte des cellules

Le milieu de culture est aspiré, les cellules sont rincées avec 10 ml de tampon PBS 1X stérile, puis lysées avec 100 µl par puit de tampon de lyse *Passiv Lysis Buffer* (PLB) 1X de *Promega* durant 15 min. Les lysats sont transférés dans une plaque 96 puits et centrifugés 10 min à 1500 rpm. Ces lysats clarifiés sont transférés dans une nouvelle plaque 96 puits et immédiatement utilisés ou conservés à -20°C.

3.5.2.3. Test β-galactosidase

On pipette 20 µl de lysat dans une plaque 96 puits transparente auxquels sont ajoutés 180 µl de solution ONPG (0,2 g ONPG pour 50 ml d'eau ultra-pure/tampon β-galactosidase dans un rapport 1/5). Le tampon β-galactosidase se compose pour 500 ml, de : 60 mM de

Na_2HPO_4 , 40 mM de $\text{Na}_2\text{H}_2\text{PO}_4$, 10 mM KCl, 50 mM de β -mercaptoéthanol, 1 mM de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ et d'eau ultra-pure.

La plaque est incubée à 37°C jusqu'à obtention d'une coloration jaune suffisante (30 à 40 min). La réaction est stoppée par 75 μl de Na_2CO_3 1 M. On mesure ensuite la $\text{DO}_{415\text{nm}}$ (comprise entre 0,1 et 1), puis les unités Gal sont calculées suivant la formule :

$$U_{\text{gal}} = (100 \times V_{\text{tot}} \times \text{DO}_{415\text{nm}}) / (V_{\text{test}} \times \text{Tps})$$

où V_{tot} représente le volume de tampon de lyse total (ici 100 μl), V_{test} le volume de lysat utilisé pour l'essai (ici 20 μl) et Tps le temps en heure passé à 37°C.

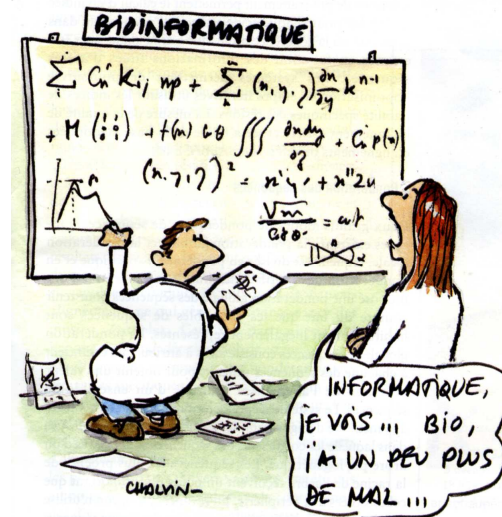
3.5.2.4. Test luciférase

On pipette 10 μl de lysat dans une plaque 96 puits auxquels sont ajoutés 50 μL de réactif luciférine. Le réactif luciférine, préparé extemporanément, est composé de 530 μM d'ATP, 470 μM de luciférine, 270 μM de Coenzyme A et un demi-volume de tampon luciférase 2X. Le tampon luciférine 2X contient 40 mM de Tris-Phosphate pH7,8, 2,14 mM de MgCl_2 , 5,4 mM de MgSO_4 , 0,2 mM de EDTA, 66,6 mM de DTT et de l'eau ultra-pure.

Les mesures des tests de l'activité luciférase ont été réalisées sur un luminomètre *Berthold Technologies LB96P*, avec 10 sec de temps de lecture.

Résultats

Chapitre 4 - Contexte bioinformatique



4.1. Analyse de protéines - Le rôle central des alignements multiples de séquences

4.1.1. Introduction

Au cours de l'évolution, des événements de mutations aléatoires sont survenus, lesquelles ont modifié les séquences des gènes codant pour l'ARN et les protéines. Les mutations, n'entraînant pas de remaniement chromosomique majeur, peuvent être i) spontanées si elles apparaissent naturellement lors de la réplication des gènes ou ii) induites par des facteurs extérieurs, ou agents mutagènes, ces derniers pouvant être des facteurs physiques (rayons ultraviolets, X, gamma) ou chimiques tels que les pesticides et les solvants. Généralement les mutations sont corrigées immédiatement par des mécanismes complexes et efficaces de réparation de l'ADN.

Quelque soit son origine, la mutation est susceptible d'engendrer une cascade de modifications pouvant entraîner des altérations importantes.

Plusieurs types de mutations existent, notamment, les mutations ponctuelles définies par la substitution d'un nucléotide de l'ADN par un autre (mutations faux-sens, non sens,

silencieuses), ou par l'insertion/délétion d'un à plusieurs centaines de nucléotides. D'autres mécanismes entrent en jeu entraînant des réarrangements chromosomiques où les brins d'ADN subissent une double cassure qui peuvent engendrer par exemple une anomalie du nombre de chromosomes ou l'apparition de nouvelles combinaisons de gènes. La plupart des mutations sont dites neutres. Elles ne modifient pas le fonctionnement de la protéine et n'ont pas de conséquences phénotypiques macroscopiques. Cependant, certaines mutations, par exemple la mutation de résidus du site actif d'une enzyme ou la mutation d'un résidu empêchant une protéine d'adopter sa structure spatiale correcte qui conditionne son fonctionnement pourraient être à l'origine de conséquences pour le moins dramatiques en engendrant des protéines non fonctionnelles. Si ces protéines sont impliquées dans des processus biologiques nécessaires à la survie de la cellule (organisme), la mutation peut être responsable de la mort de la cellule.

Ainsi, au cours de l'évolution, les résidus fonctionnellement et structuralement importants pour la protéine subissent une forte pression de sélection et doivent être conservés. En comparant les séquences proches les unes des autres et en regardant les résidus qui restent les mêmes (conservés) dans tous les membres de la famille, on peut ainsi déduire quels sont les résidus déterminants au bon fonctionnement de la protéine.

Aussi, la comparaison de plusieurs séquences, ou alignement multiple, est devenue un outil incontournable dans le domaine de la biologie moléculaire moderne, essentiel à de nombreuses analyses de protéines telles que la modélisation, la reconstruction phylogénétique ou la description de sites conservés ou variables dans une même famille de protéines.

4.1.2. Définition de l'Alignement Multiple de Séquences Complètes

Dans la plupart des représentations, un alignement multiple représente une liste de séquences utilisant un code à une lettre pour chaque nucléotide ou acide aminé. Chaque ligne horizontale dans l'alignement correspond à une séquence simple et les résidus équivalents selon leurs propriétés physico-chimiques et équivalents d'un point de vue évolutif, structural ou fonctionnel sont alignés verticalement. Le diagramme de Venn (Figure 42) ayant trait aux propriétés des acides aminés donne une idée de la compatibilité des résidus en fonction des caractéristiques physico-chimiques à maintenir.

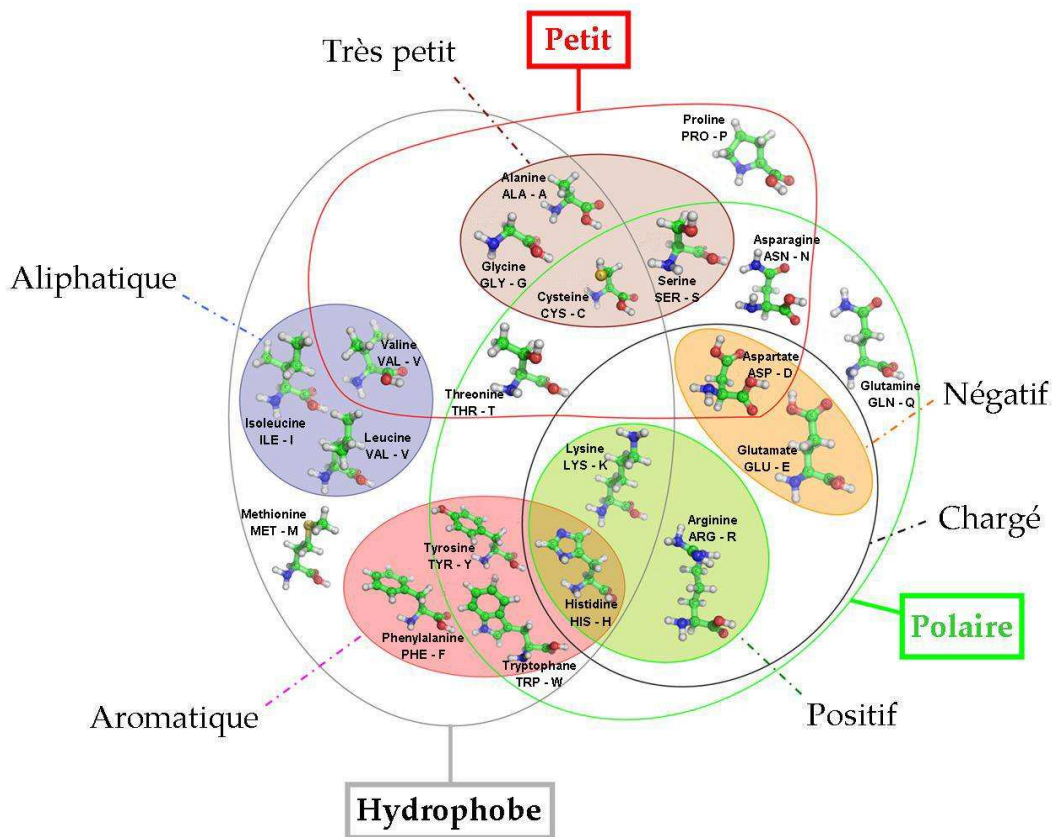


Figure 42 | Diagramme de Venn de la classification des acides aminés selon leurs propriétés physico-chimiques (Taylor, 1986).

L'alignement multiple représente un outil extrêmement fascinant pour l'analyse des familles de protéines, qui plus est lorsqu'il s'agit d'un alignement multiple de séquences complètes ou MACS, acronyme de *Multiple Alignment of Complete Sequences*, la quintessence des outils du bioinformaticien. Le MACS (Figure 43) consiste à aligner plusieurs séquences homologues dans leur intégralité (alignement global) et permet d'intégrer pour chaque séquence à la fois une analyse horizontale en tenant compte de sa longueur totale et une analyse verticale en l'intégrant dans son contexte évolutif. Lorsque les séquences sont de longueurs différentes ou si elles présentent des variations de composition de séquence, il est nécessaire d'introduire des insertions dans un groupe de séquences pour « recaler » le reste de la séquence ; ces insertions correspondant à des lacunes ou *gaps* dans les autres groupes.

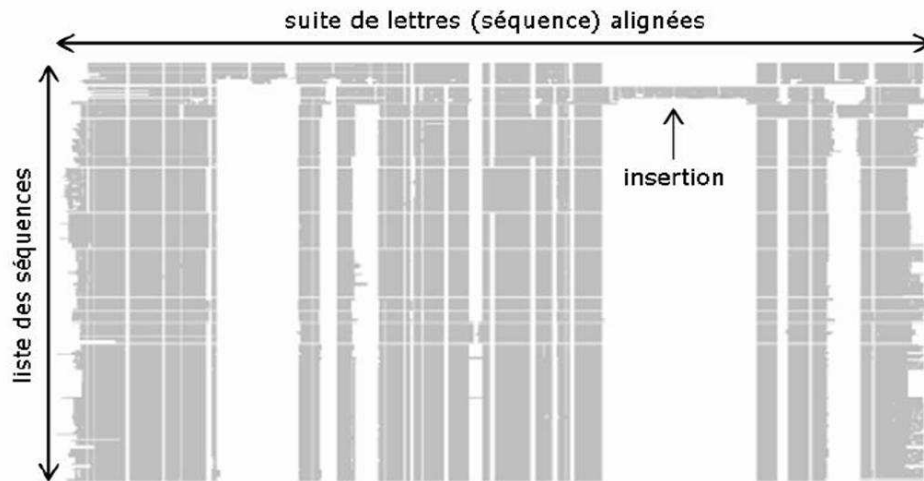


Figure 43 | Vue schématique d'un MACS. Le MACS contient des séquences d'une famille de protéines alignées sur toute leur longueur. Ces protéines peuvent être issues de plusieurs organismes. L'alignement peut nécessiter l'introduction d'insertion dans un groupe de séquences correspondant à des lacunes ou *gaps* dans d'autres.

De nombreux programmes d'alignement multiple de séquences ont été développés au cours des 20 dernières années (Edgar and Batzoglou, 2006), le plus connu et couramment utilisé étant sans doute ClustalW (Thompson *et al.*, 1994). Beaucoup de programmes utilisent de la couleur pour différencier l'information. Pour les ADN ou ARN, l'utilisation de couleur permet de différencier les nucléotides. Pour les alignements de protéines, elle permet d'indiquer les propriétés des acides aminés, ce qui aide à conclure sur la conservation du rôle d'un acide aminé substitué (Figure 42).

Dans la compréhension du fonctionnement de la vie, les protéines jouent un rôle essentiel. On part donc de l'hypothèse que des protéines comportant des séquences similaires risquent fort de posséder des propriétés physico-chimiques identiques et des fonctions similaires. À partir de l'identification de similarités entre la séquence d'une première protéine dont on connaît le mécanisme d'action et celle d'une deuxième protéine dont on ne connaît pas le mécanisme de fonctionnement, nous pouvons ainsi inférer des similitudes structurelles ou fonctionnelles sur la séquence non connue.

4.1.3. Applications du MACS

La comparaison des séquences au sein d'un MACS est devenue un outil incontournable dans plusieurs domaines de la biologie moléculaire moderne, des études de phylogénie jusqu'à la prédiction de structures 2D et 3D, en passant par la prédiction des fonctions d'une protéine, des études de réseaux d'interaction, d'annotation fonctionnelle, de génomique comparative, etc (Figure 44).

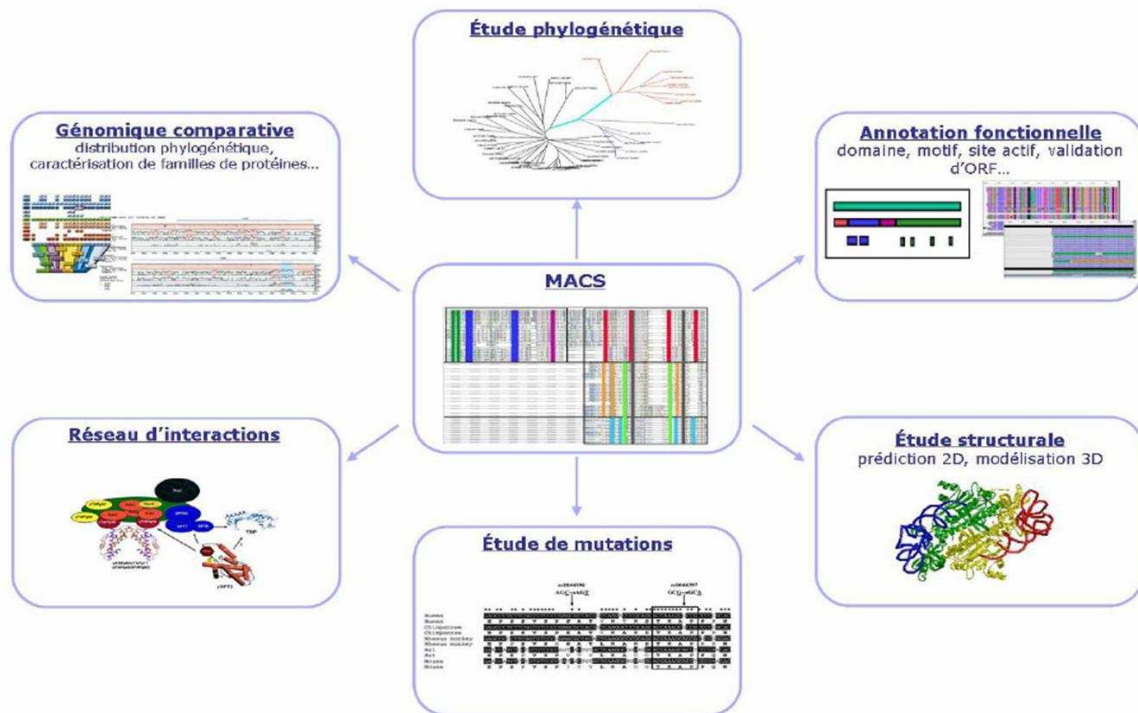


Figure 44 | Applications des alignements multiples de séquences protéiques en biologie moderne.

L'alignement multiple permet de questionner un nombre important de domaines biologiques, dont en voici quelques exemples.

4.1.3.1. La phylogénie et la génomique comparative

La phylogénie est la discipline ayant pour objectif la reconstruction de l'histoire évolutive des espèces en vue d'établir leur parenté, et dans le cas de la phylogénie moléculaire, elle est basée sur la comparaison des séquences de leurs gènes ou de leurs protéines. A titre d'exemple, l'arbre de la vie actuel, dans lequel le monde vivant est divisé en trois domaines (bactérie, archaebactérie et eucaryotes) a été construit d'après des études de comparaison de séquences d'ARN ribosomal. L'étude d'une séquence protéique dans son contexte évolutif (Phillips 2000) constitue une étape cruciale dans sa caractérisation et est

basée sur l'analyse de sa famille. De plus, étudier les phénomènes évolutifs afin de retracer la phylogénie des espèces et de comprendre les mécanismes sous-jacents constitue un des enjeux majeurs de la génomique comparative. Elle tente en effet, d'interpréter la multitude de données disponibles afin de proposer des modèles d'évolution, de comprendre l'organisation des gènes et de retracer les processus responsables de la conservation/divergence entre les différents génomes de différentes espèces.

La qualité de l'alignement (choix des séquences et des espèces) constitue la base dans la reconstruction phylogénétique, un mauvais alignement pouvant ruiner la reconstruction. Une étape fondamentale dans l'analyse est d'établir la correspondance entre les séquences, c'est-à-dire trouver les séquences homologues où l'homologie désigne une relation d'évolution entre des gènes. Deux gènes sont homologues s'ils sont le résultat d'un processus d'évolution divergent à partir d'un ancêtre commun (Fitch, 1970).

Il existe plusieurs types de relations d'homologie comme illustrés sur la Figure 45 :

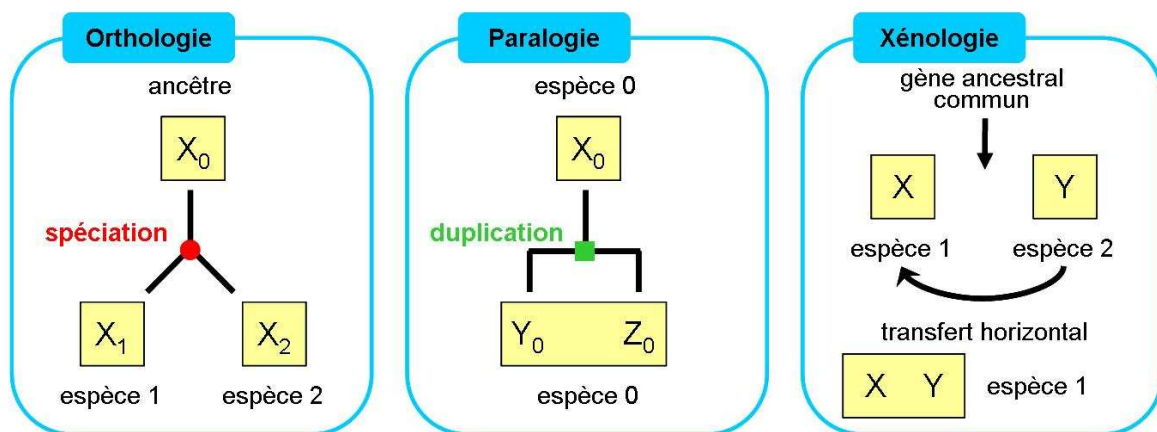


Figure 45 | Les différentes relations d'homologie.

- les gènes orthologues sont des gènes homologues qui ont évolué à partir d'un gène ancestral suite à un événement de spéciation,
- les gènes paralogues sont des gènes homologues qui ont évolué à partir d'un gène ancestral suite à un événement de duplication,
- les gènes xénologues sont des gènes homologues présents au sein d'un même organisme, l'un ayant été hérité « verticalement » et le second par transfert horizontal, c'est-à-dire lors de l'acquisition du matériel génétique d'une autre espèce.

Les familles des gènes paralogues présentent souvent des fonctions similaires ou proches, mais pas nécessairement identiques, tandis que les gènes orthologues ont généralement conservé la même fonction. La distinction entre gènes paralogues et orthologues constitue donc un critère important dans le cadre d'études évolutives (Pellegrini *et al.*, 1999).

4.1.3.2. Caractérisation fonctionnelle

Les alignements multiples sont couramment utilisés dans le cadre de la caractérisation fonctionnelle de protéines, principalement pour la validation de séquences protéiques, en permettant de détecter les erreurs de séquençage, les décalages du cadre de lecture, les erreurs de prédiction de la structure du gène (codon initiateur, site d'épissage intron/exon) (Bianchetti *et al.*, 2005), ou pour la prédiction de la fonction par la propagation d'informations des séquences annotées vers les séquences non annotées.

4.1.3.3. Amélioration de la prédiction des structures secondaires et tertiaires

Dans un contexte structural, l'utilisation de l'alignement multiple permet d'améliorer la prédiction des structures secondaires (Lee *et al.*, 2006) et tertiaires (Moult, 2005; Yang and Honig, 2000).

Différentes méthodes expérimentales permettent de résoudre la structure tertiaire des protéines comme la cristallographie aux rayons X, la spectroscopie par résonance magnétique nucléaire ou encore la cryomicroscopie. Cependant, ces méthodes sont difficiles et coûteuses et la détermination de la structure d'une protéine reste un processus lent. En effet, on peut noter que la principale banque de données de structures tridimensionnelles, la PDB, ne contient que 48 161 entrées au 1^{er} janvier 2008, alors que la banque de séquences protéiques UniProt (Consortium 12.6) contient 5 362 532 entrées au 4 décembre 2007.

Afin de pallier à ce manque de données expérimentales, des méthodes automatiques *in silico* de prédiction des structures secondaires et tertiaires des protéines ont été développées. Il se dégage deux types de méthodes, l'une basée sur la prédiction locale de la structure des protéines et l'autre basée sur la prédiction globale. Néanmoins, ces deux approches sont étroitement liées dans la mesure où la structure tertiaire elle-même impose la structure secondaire locale.

4.1.3.4. Prédiction des structures secondaires

La séquence en acides aminés d'une protéine code pour sa structure 3D et il est courant de prédire un certain nombre de caractéristiques structurales à partir de cet enchaînement de résidus. Les principaux éléments structuraux pouvant être prédits à partir de la structure primaire d'une protéine sont :

- les éléments de structures secondaires (hélice α et feuillet β), dont la prédiction constitue un vaste domaine d'étude depuis les années 70 (revue dans (Rost, 2001)). Les premières méthodes publiées utilisaient le fait que, dans les structures tridimensionnelles, certains acides aminés sont retrouvés préférentiellement dans certaines classes de structures secondaires ; au courant des années 80, l'environnement local des résidus dans la séquence a été pris en compte. Les méthodes actuelles se sont diversifiées et utilisent couramment l'information de séquences homologues (Heringa, 2000). Elles permettent d'obtenir un taux de bonne prédiction supérieur à 75%. Des approches différentes étant utilisées selon les programmes, il est courant d'obtenir des résultats contradictoires, c'est pourquoi il est conseillé de comparer plusieurs méthodes afin d'obtenir un consensus. Plusieurs logiciels sont disponibles, notamment, PREDATOR (Frishman and Argos, 1995), NNSSP (Salamov and Solovyev, 1997), PHD (Rost, 1996), PSIPred (Jones, 1999) ou encore ANTHEPROT (Deleage *et al.*, 2001),
- les régions *coiled-coils*, répétitions d'heptapeptides, qui sont des hélices oligomérisées enroulées les unes autour des autres pour former des super-hélices. Plusieurs programmes de prédiction des *coiled-coils* existent, citons COILS (Lupas *et al.*, 1991), PAIRCOIL (Berger *et al.*, 1995) et MULTICOIL (Wolf *et al.*, 1997), par exemple,
- les régions transmembranaires, ou plus précisément les hélices α transmembranaires. Tous ces programmes de prédiction reposent sur les profils d'hydrophobicité de la chaîne polypeptidique impliquant 16 à 25 résidus et supposent que leurs structures secondaires seront en hélice α (revue dans (Chen *et al.*, 2002)),
- les régions désordonnées, qui n'adoptent pas de structure 3D bien définie et qui peuvent être impliquées dans de nombreux processus biochimiques importants

(reconnaissance moléculaire, transduction de signal) (Wright and Dyson, 1999). Différents paramètres, comme la composition en acides aminés, la complexité de la séquence, l'hydrophobicité ou la charge nette, sont pris en compte par les programmes de prédiction disponibles, comme DISOPRED2 (Ward *et al.*, 2004a), GlobPlot2 (Linding *et al.*, 2003b), FoldIndex (Prilusky *et al.*, 2005) ou encore RONN (Thomson *et al.*, 2003).

Dans ce contexte, on peut noter qu'en biologie et génomique structurales, le désordre intrinsèque d'une protéine est souvent source de problèmes survenant lors des étapes ultérieures d'expression, de purification ou de cristallisation. Pour minimiser ces problèmes, il est nécessaire de délimiter au mieux les régions potentiellement désordonnées de la protéine étudiée.

Un certain nombre d'outils de prédiction du désordre, construits sur des algorithmes différents et ayant chacun leur propre spécificité, sont actuellement disponibles (Vucetic *et al.*, 2005) (<http://www.disprot.org/predictors.php>). Une des stratégies d'amélioration de l'efficacité de prédiction est d'utiliser plusieurs méthodes performantes et de combiner les différents résultats obtenus. Dans cet esprit, nous avons développé un serveur web **KoAnno** (**Chaos Annotation**) regroupant quatre programmes de prédiction du désordre et disponible à l'adresse suivante <http://www-bio3d-igbmc.u-strasbg.fr/~brelivet/KoAnno/>. Ce projet, décrit au chapitre 6.1.2.2 s'intègre dans le cadre des recherches consacrées à l'étude des RNs au sein du Département de Biologie et Génomique Structurales et du Laboratoire de Bioinformatique et Génomique Intégratives. De plus, cette application est facilement accessible à tout utilisateur et permet la visualisation simultanée de l'ensemble des résultats de prédiction d'une protéine ou d'une famille de protéines.

4.1.3.5. Prédiction des structures tertiaires

La modélisation par homologie est la méthode *in silico* la plus couramment utilisée pour obtenir un modèle 3D d'une protéine lorsque sa structure n'a pas été déterminée expérimentalement. C'est un moyen de construire un modèle 3D d'une protéine d'intérêt sur la base d'une autre protéine « support » de structure expérimentale connue (Sali et Blundell 1993). Son principe repose sur le fait que des protéines homologues, avec des séquences

d'acides aminés proches, présentent des structures 3D similaires. La modélisation par homologie se base sur un alignement de qualité entre les séquences. Les coordonnées atomiques de la protéine support sont affectées à la séquence d'intérêt et les chaînes latérales sont positionnées de manière à minimiser l'énergie du système. Cette technique est efficace si le degré d'identité des séquences est suffisamment élevé, généralement de 30% ou plus. Il faut savoir, qu'en théorie, plus le pourcentage d'identité entre les deux séquences est élevé, meilleur sera le modèle. MODELLER (Sali *et al.*, 1995), Geno3D (Combet *et al.*, 2002), SWISS-MODEL (Schwede *et al.*, 2003), 3D-JIGSAW (Bates *et al.*, 2001) sont des exemples d'outils performants de modélisation par homologie.

4.1.3.6. Conservation des résidus

Une des applications majeures de l'alignement multiple réside dans l'analyse des conservations au sein d'une famille de protéines, notamment les résidus strictement conservés dans toutes les séquences (résidus ou motifs d'importance structurale et/ou fonctionnelle) ou encore les résidus conservés spécifiquement dans un sous-groupe de séquences (résidus discriminants de ce sous-groupe) (del Sol Mesa 2003).

« *La Séquence ne ment jamais !* »

Olivier Poch

Chapitre 5 - Analyse du MACS de la super-famille des RNs

Résumé de l'article

Les RNs sont des facteurs de transcription qui contrôlent un grand nombre d'évènements physiologiques au travers de la régulation de la transcription de gènes cibles. La plupart des RNs fonctionnent comme dimères, soit en homodimères soit en hétérodimères avec RXR/USP, le LBD représentant le contributeur majeur à la stabilité du dimère par de nombreux contacts protéine-protéine. Cependant, les mécanismes moléculaires qui gouvernent cette association fonctionnellement importante ne sont pas encore élucidés. Pour établir les bases moléculaires de cette discrimination entre les deux états, homo- ou hétérodimère, nous avons réalisé, puis analysé, un alignement multiple de séquences complètes (MACS), basé sur la structure, de tous les RNs des Métazoaires.

L'origine de ce travail a donc débuté par la réalisation du MACS des RNs couvrant l'ensemble des domaines, du domaine A/B au domaine F. Dans ce but, nous avons recherché de manière exhaustive l'ensemble des séquences similaires dans les banques de protéines à partir des 48 RNs humains. Nous avons détecté 1432 séquences protéiques, les fragments de longueur inférieure à 20% de la longueur totale de la protéine ont été éliminés soit 147 séquences, ainsi que 203 séquences contenant au moins une insertion/délétion dans un motif. Deux MACS sont disponibles. Le premier est composé de toutes les séquences protéiques (706) de RNs sans les 260 RNs de *C. elegans* qui pourraient apporter un biais non négligeable dans l'analyse. Ces séquences de *C. elegans* sont analysées en utilisant un second MACS qui lui est composé des génomes complets, disponibles lors de la réalisation de ce MACS, et qui a été utilisé pour des analyses de génomique comparative.

Même avec les programmes d'alignements disponibles, la réalisation du MACS des RNs a nécessité un travail long et complexe impliquant de nombreux processus cycliques

d'ajustements manuels et d'analyse des structures secondaires, pour obtenir un MACS de haute qualité (Figure 46).

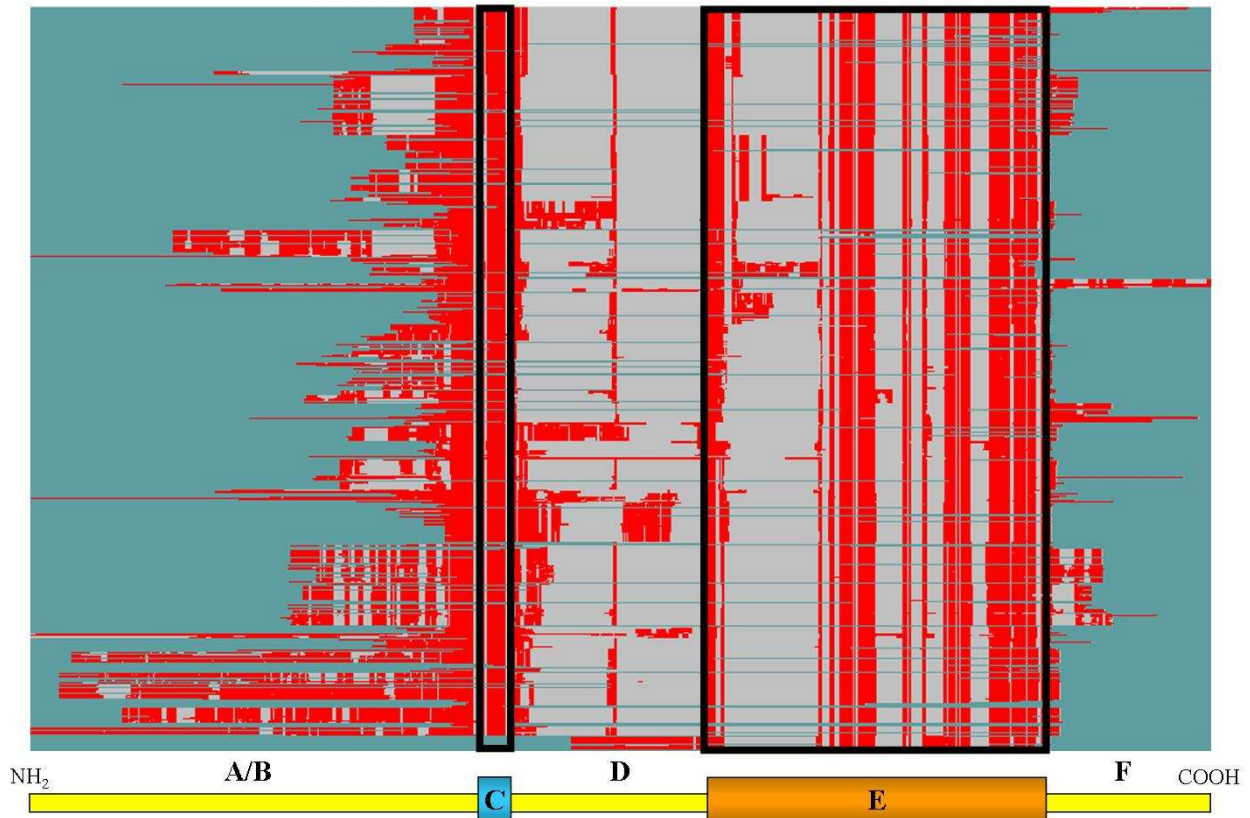


Figure 46 | MACS des RNs avec les différents domaines : le domaine A/B, C (DBD, coloré en bleu), D, E (LBD, coloré en orange) et le domaine F.

La Figure 46 représente une vue générale du MACS des RNs. Les régions conservées correspondantes aux deux domaines conservés des RNs, le DBD et le LBD, sont aisément identifiables comme l'atteste la superposition de la représentation schématique d'un RN. Le LBD, composé d'environ 250 acides aminés, présente des petites régions conservées qui correspondent aux différents éléments de structures secondaires, les douze hélices α et le feuillet β . Nous pouvons également observer les régions moins conservées, comme le domaine A/B, la région charnière (D) reliant le DBD au LBD et le domaine F présent dans quelques RNs. Par sa complexité, ce MACS validé manuellement et structuralement est un excellent « cas test » pour tous les programmes développés dans notre laboratoire. Le cercle vertueux liant recherche de base et applications joue à plein.

Le sujet de cette étude étant « la dimérisation du LBD », nous nous sommes donc focalisés sur l'alignement des LBDs. L'analyse a révélé la présence de deux jeux de résidus conservés et différentiellement conservés, qui pourraient être fonctionnellement et/ou structurellement importants et qui ont permis de diviser la super-famille des RNs en deux classes distinctes en relation avec leur comportement oligomérique. Ces résultats ont été examinés au niveau structural, pour chaque classe, en utilisant les structures tridimensionnelles des LBDs disponibles. Il en est ressortit que tous les homodimères appartiennent à la classe I, dans laquelle, les résidus identifiés contribuent à une voie de communication constituée de deux ponts salins reliant l'interface de fixation des cofacteurs à l'interface de dimérisation. La classe II est définie par la présence d'un unique résidu différentiellement conservé situé (Arginine) dans la boucle H8-H9. Cette Arginine est impliquée dans un pont salin, qui fait de lui la signature des hétérodimères. La voie de communication de la classe II diffère de celle de la classe I par les interfaces qu'elle connecte. Pour la classe II, il s'agirait d'une voie de communication liant la poche de fixation du ligand à l'interface de dimérisation. Il est, de plus, intéressant de noter, que tous ces résidus ne sont pas localisés à la surface du LBD.

Au cours de l'évolution, les résidus fonctionnellement et structurellement importants subissent une forte pression de sélection. L'analyse de l'alignement multiple de *C. elegans* a montré que les motifs canoniques $[(I,L)lx(A,I)Exxh]$, $[hxcAHxxT]$ et $[(F,W)AKX_4FX_2LX_3DQX_2LL]$ (I, Ile, T, Thr, L, Leu, A, Ala, E, Glu, F, Phe, W, Trp, K, Lys, D, Asp, Q, Gln, h, acide aminé hydrophobe, l, acide aminé à longue chaîne latérale, c, acide aminé chargé et x un acide aminé quelconque) (Wurtz *et al.*, 1996) pour le LBD des RNs sont peu ou pas conservés. En dépit de cette variabilité, notre analyse a montré la présence de l'Arginine conservée spécifique de la classe I en N-ter de H10 et l'absence stricte de l'Arginine différentiellement conservée caractéristique de la classe II. Les RNs de *C. elegans* appartiendraient donc à la classe I. Ces observations laissent supposer que les RNs ancestraux seraient des facteurs de transcription homodimériques.

Publication N°1

Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level

Yann Brelivet, Sabrina Kammerer, Natacha Rochel, Olivier Poch, Dino Moras

EMBO reports, 2004, Vol. 5, N° 4, Pages 423-429

Pages 143-149 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1038/sj.embor.7400119>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

Chapitre 6 - Et les autres domaines dans tout ça ?

6.1. Le domaine A/B

« *Le chaos est souvent source de vie, alors que l'ordre génère des habitudes.* »

Henry Adams

6.1.1. Rappel

Le domaine A/B des RNs représente une des régions la moins conservée des RNs à la fois en séquence et en taille, de 23 résidus pour VDR à 602 résidus pour MR. Ce domaine possède fréquemment une composition en acides aminés très biaisée (richesse en résidus de type Ser, Thr, Pro, Asn, Glu, Gln, etc.). Malgré de nombreuses tentatives, aucune structure cristallographique du domaine A/B seul n'a encore été résolue à ce jour. En absence de partenaire, il semblerait que ce domaine A/B existe sous la forme d'un état désordonné donc peu ou pas structuré (Dahlman-Wright and McEwan, 1996; McEwan *et al.*, 2007; Warnmark *et al.*, 2001). Cette propriété est commune à un grand nombre de domaines d'activation des facteurs de transcription eucaryotiques qui présentent une faible propension intrinsèque à former des structures secondaires comme c'est le cas, par exemple, pour NF- κ B (Schmitz *et al.*, 1994) ou p53 (Kussie *et al.*, 1996).

6.1.2. Le désordre

6.1.2.1. Caractérisation et fonction

Selon le paradigme classique séquence-structure-fonction de la biologie structurale, la structure 3D d'une protéine, déterminée en grande partie par sa séquence primaire, est au cœur de sa fonction biologique (Anfinsen, 1973). En 1999, Wright et Dyson (Wright and Dyson, 1999) ont appelé à une réévaluation majeure de cette vision, suggérant qu'un grand nombre de régions protéiques et de protéines pouvaient exercer leur fonction en absence d'un repliement bien défini. Il a été montré, par exemple, que des régions désordonnées pouvaient jouer un rôle essentiel dans des voies de signalisation cellulaire (Frankel and Kim,

1991; Weinreb *et al.*, 1996). Les protéines ou régions protéiques n'adoptant pas de structure 3D bien définie ont été nommées « intrinsèquement désordonnées ».

Ces dernières années, différents travaux ont permis de mieux caractériser les régions désordonnées et ont abouti à la détermination de différents paramètres souvent utilisés pour prédire *in silico* le désordre intrinsèque dans les protéines : la composition en acides aminés, la complexité de la séquence, l'hydrophobicité, la charge nette, autant de paramètres fortement interconnectés (Dunker and Obradovic, 2001; Romero *et al.*, 1997; Romero *et al.*, 1998; Uversky, 2002).

Les régions et protéines intrinsèquement désordonnées sont généralement caractérisées par une faible hydrophobicité et une charge nette élevée. L'hydrophobicité moyenne est définie par la somme de l'hydrophobicité de tous les résidus divisée par le nombre total de résidus de la séquence protéique. La charge nette moyenne se définit quant à elle comme la charge nette à pH=7, c'est-à-dire le nombre total de résidus Asp et Glu chargés négativement et de résidus Arg et Lys chargés positivement divisé par le nombre total de résidus de la protéine (Uversky, 2002; Uversky *et al.*, 2000). De plus, les régions désordonnées présentent une faible complexité, c'est-à-dire une très faible variété de résidus (par exemple, une région riche en Pro), et une composition en acides aminés biaisée. Ces régions contiennent des taux élevés en certains acides aminés tels que Glu, Lys, Arg, Asp, Gly, Gln, Ser et Pro et de faibles taux en Ile, Leu, Val, Trp, Phe, Tyr, Cys et Asn (Romero *et al.*, 2001). La présence de résidus aromatiques et leur capacité d'interaction ont été associées à des forces de structuration pouvant inhiber le désordre (Burley and Petsko, 1985). D'autre part, les acides aminés chargés (Glu, Asp, Arg et Lys) se retrouvent souvent en grand nombre dans des régions désordonnées puisqu'un déséquilibre de charge tend à favoriser le désordre. Cette différence de composition explique la faible hydrophobicité et la charge nette élevée caractérisant les régions désordonnées.

D'un point de vue expérimental, le désordre des protéines est identifié par diverses méthodes telles que la cristallographie aux rayons X, la RMN ou le dichroïsme circulaire. Chacune de ces méthodes détecte différemment le désordre (Tompa, 2002). Au mieux, les régions désordonnées sont invisibles par ces techniques ; au pire, elles peuvent perturber entièrement l'expérience en affectant la solubilité et/ou la cristallisation (Oldfield *et al.*, 2005).

La caractérisation des régions désordonnées est d'autant plus ardue que certaines régions impliquées dans des mécanismes de reconnaissance moléculaire subissent parfois des transitions désordre - ordre lors de la liaison du partenaire physiologique (Figure 47) (Weinreb *et al.*, 1996). De telles interactions moléculaires peuvent être de type enzyme - substrat, récepteur - ligand, protéine - protéine, protéine - ARN ou protéine - ADN (Huber, 1979).

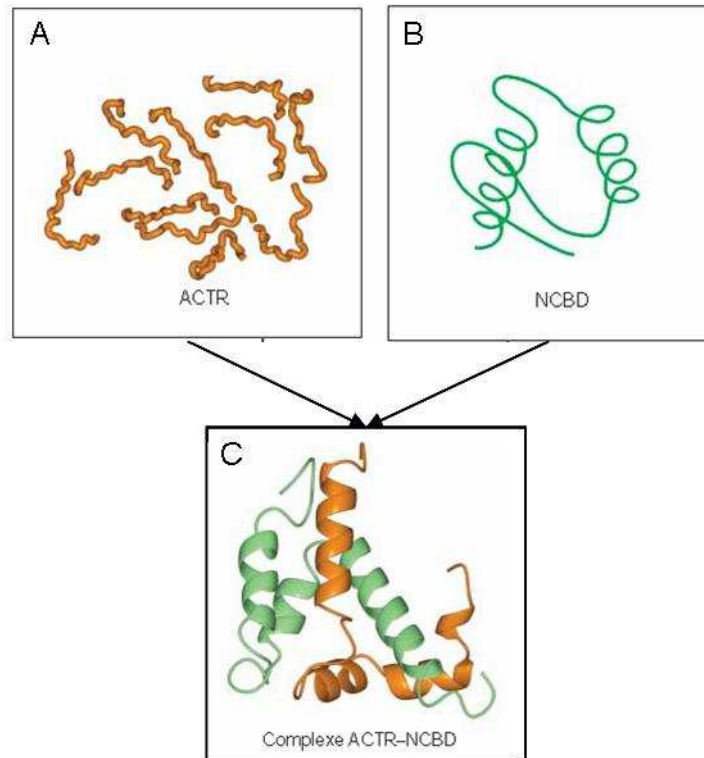


Figure 47 | Exemple de transition désordre - ordre. (A), Domaine d'interaction désordonné ACTR (*ACTivator for Thyroid hormone and Retinoid receptors*) du CoA p160 coloré en orange. (B), Domaine d'interaction désordonné NCBD (*Nuclear-receptor Co-activator Binding Domain*) de CBP (*cAMP-responsive (CRE)-binding (CREB) protein*) coloré en vert. (C), Structure correctement repliée du complexe ACTR-NCBD (code PDB : 1KBH).

La transition désordre-ordre lors de la fixation à un partenaire peut aboutir à une interaction combinant à la fois une forte spécificité et une faible affinité de liaison avec le partenaire évitant ainsi une fixation irréversible souvent inappropriée pour des processus biologiques.

Des exemples de domaines protéiques intrinsèquement désordonnés et fonctionnels sont constamment rapportés. Les fonctions ou les mécanismes auxquels sont reliés les

régions désordonnées incluent la régulation de la transcription et de la traduction, la transduction du signal cellulaire, la phosphorylation des protéines, la régulation de l'auto-assemblage de grands complexes multiprotéiques comme le flagelle des bactéries ou le ribosome. Certaines régions désordonnées jouent le rôle de chaperonnes pour d'autres protéines et pour l'ARN et fonctionnent comme éléments de reconnaissance (Tompa and Csermely, 2004). Compte tenu des multiples implications, il n'est pas surprenant qu'une forte proportion (30-60%) de protéines eucaryotiques soit prédite comme contenant de longues régions désordonnées (≥ 40 résidus consécutifs). Ces grands segments non structurés seraient fréquents dans les protéines régulatrices comme les protéines de signalisation et les protéines associées à certains types de cancers, les régulateurs transcriptionnels et les protéines de liaison à l'ARN, mais ils seraient relativement rares dans les protéines ayant une fonction catalytique (Iakoucheva *et al.*, 2002; Vucetic *et al.*, 2003; Ward *et al.*, 2004b).

6.1.2.2. KoAnno

En biologie et génomique structurales, de nombreux problèmes lors de l'expression, de la purification et/ou cristallisation d'une protéine sont directement ou indirectement reliés à la présence d'une zone, ou d'un grand segment, présentant un désordre intrinsèque. Pour minimiser ces problèmes, il est nécessaire de prédire et de délimiter au mieux les régions potentiellement désordonnées. De nombreux logiciels de prédiction des régions désordonnées dans les protéines, s'appuyant sur des algorithmes différents et ayant chacun leur propre définition du désordre, sont actuellement disponibles (<http://www.disprot.org/predictors.php>) (Vucetic *et al.*, 2005). Une des stratégies d'amélioration de l'efficacité de prédiction est d'utiliser plusieurs méthodes performantes et de combiner les différents résultats obtenus. C'est dans ce contexte que nous avons développé KoAnno (pour *Annotation du Chaos*), un serveur web regroupant quatre programmes de prédiction des régions désordonnées dans les protéines.

6.1.2.2.1. Programmes de prédiction du désordre

Parmi l'ensemble des programmes disponibles, certains n'ont pas été implémentés dans notre serveur *web* KoAnno pour des questions de gestion et de traitement des résultats, notamment ceux dont les résultats sont envoyés par e-mail (DISOPRED2 (Ward *et al.*, 2004a) et NORSp (Liu and Rost, 2003)) ou encore ceux qui ne sont pas en accès libre (PONDR(Li *et al.*, 1999)). Nous avons finalement arrêté notre choix sur quatre programmes de prédiction du désordre : RONN (Yang *et al.*, 2005), DisEMBL1.4 (Linding *et al.*, 2003a), GlobPlot2 (Linding *et al.*, 2003b) et FoldIndex (Prilusky *et al.*, 2005). Ce choix a également été guidé par l'étude réalisée par Yang *et coll.* (Yang *et al.*, 2005) qui ont comparé les résultats des programmes RONN, PONDR, FoldIndex, DisEMBL1.4, GlobPlot2, DISOPRED2 et Prelink et ont montré que les quatre programmes choisis sont parmi les plus performants.

6.1.2.2.1.1. RONN

RONN est une extension du réseau de neurones BBFNR (*Bio-Basis Functional Neural Networks*) (Thomson *et al.*, 2003) destiné à l'origine à la caractérisation de sites de clivage protéolytique et maintenant adapté à la détection de régions désordonnées.

Des bibliothèques d'alignements de séquences de repliements connus (ordonnés, désordonnés ou un mélange des deux) ont été réalisées manuellement d'après des données structurales. Si l'on veut connaître les zones ordonnées ou désordonnées d'une séquence quelconque, une méthode consiste à réaliser l'alignement de cette séquence par rapport aux séquences de ces bibliothèques. Suivant le degré de similitude, les zones ordonnées ou désordonnées peuvent être prédites. Cette prédiction est réalisée par un réseau de neurones entraîné sur ces bibliothèques de séquences connues.

6.1.2.2.1.2. GlobPlot2

GlobPlot2, outil d'identification de régions globulaires et désordonnées dans les protéines, utilise une approche simple basée sur une somme de propensions des acides aminés à se trouver dans un état ordonné ou désordonné en utilisant l'échelle de Russel/Linding (Figure 48) (Linding *et al.*, 2003b).

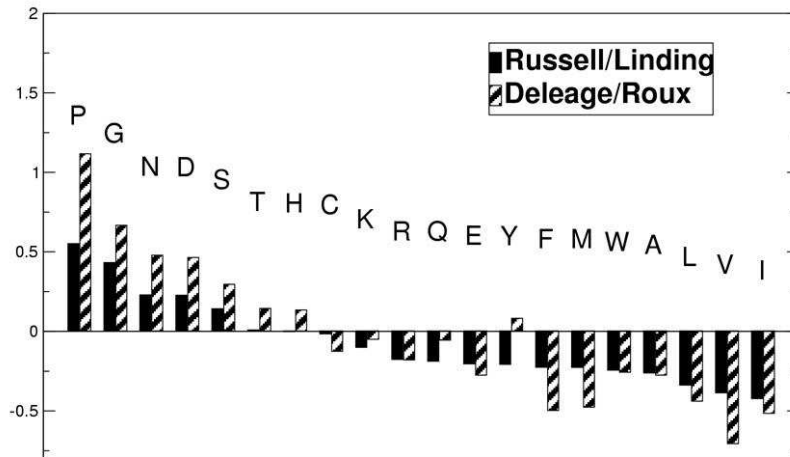


Figure 48 | Echelle de Russel/Linding et Deleage/Roux.

Le terme propension a été introduit en 1987 par Deleage et *coll.* (Deleage and Roux, 1987) et se base sur les travaux de Chou et Fasman réalisés en 1974 (Chou and Fasman, 1974) traitant des paramètres conformationnels des acides aminés dans les structures secondaires et les structures de type *Random Coil* (RC). Après avoir utilisé et constaté que l'échelle de Deleage/Roux ne s'adaptait pas ou peu à la détection des régions désordonnées, les auteurs de GlobPlot2 ont recalculé les propensions des acides aminés à se trouver soit dans des structures secondaires régulières (hélices α ou feuillet de brins β) définies par DSSP (Kabsch and Sander, 1983) soit dans d'autres structures de type RC (boucles, *turns*, etc.).

6.1.2.2.1.3. DisEMBL1.4

DisEMBL1.4, un autre réseau de neurones, se base sur trois critères différents pour estimer le désordre. Le premier critère *LOOPS/COILS* est défini par DSSP (Kabsch and Sander, 1983). Les résidus présents dans les hélices α (H), les hélices 3_{10} (G) ou les feuillet β (E) sont considérés comme ordonnés. Les résidus présents dans les états de type T (*hydrogen bonded turn*) ; S (*bend*), région présentant une forte courbure d'au moins 70° ; I (hélice π) sont considérés comme des *LOOPS* ou *COILS*. Les *LOOPS/COILS* n'étant pas nécessairement désordonnées, une prédiction du désordre basée uniquement sur cette définition pourrait se révéler imparfaite. Le deuxième critère *HOTLOOPS* représente des boucles avec un degré de mobilité élevé (facteur B). De nombreuses tentatives de prédiction du désordre basées sur les facteurs B ont été réalisées, mais il s'agit d'une stratégie délicate car les facteurs B peuvent être fortement modifiés dans une structure en fonction de l'environnement structural et de l'empilement local (Brooks and Karplus, 1985; Smith *et al.*, 2003). Pour caractériser le

désordre dans les protéines, DisEMBL1.4 utilise une technique basée sur une échelle de propension de mobilité des résidus en fonction des facteurs B (Smith *et al.*, 2003). Les coordonnées absentes dans une structure résolue par cristallographie aux rayons X définissent le troisième critère de définition du désordre par le terme *REM465*. En effet, les densités électroniques non assignées reflètent souvent un désordre intrinsèque et ont été utilisées très tôt pour la prédiction du désordre (Li *et al.*, 2000).

Des réseaux de neurones ont été entraînés sur des jeux de données correspondant aux trois définitions du désordre (*LOOPS*, *HOTLOOPS* et *REM46*). L'échelle de propension du désordre des acides aminés (Figure 49) pour chaque définition reflète le contenu des différents jeux de données utilisés pour l'entraînement des réseaux de neurones.

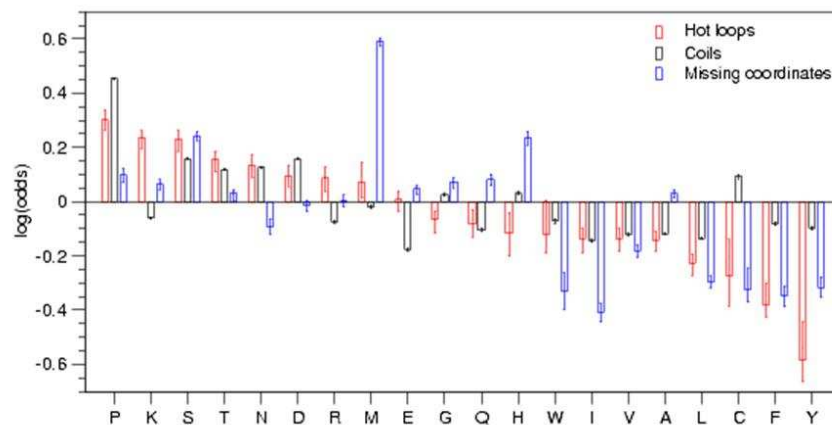


Figure 49 | Echelle de propension des résidus à se trouver dans un état désordonné.

6.1.2.2.1.4. FoldIndex

FoldIndex (Uversky *et al.*, 2000) prédit si une séquence protéique donnée est intrinsèquement désordonnée en implémentant l'algorithme défini par Uversky et *coll.* basé sur la moyenne des résidus hydrophobes de la protéine et sur la valeur absolue de sa charge nette.

6.1.2.2.2. Le serveur web KoAnno

Sans entrer dans les détails d'implémentation et de fonctionnement, KoAnno est un serveur web CGI (*Common Gateway Interface*) disponible à l'adresse suivante : <http://www-bio3d-igbmc.u-strasbg.fr/~brelivet/KoAnno/>. L'interface de KoAnno (Figure 50) offre à l'utilisateur le choix d'exécuter les prédictions du désordre soit pour une séquence

protéique unique (*Run KoAnno for an Accession Number*) soit pour un alignement multiple (*Run KoAnno for all sequences of a Multiple Alignment*). La seule restriction concerne le nom de la séquence protéique qui doit absolument être un *Accession number* (AC) du type P12345 (UniProt).

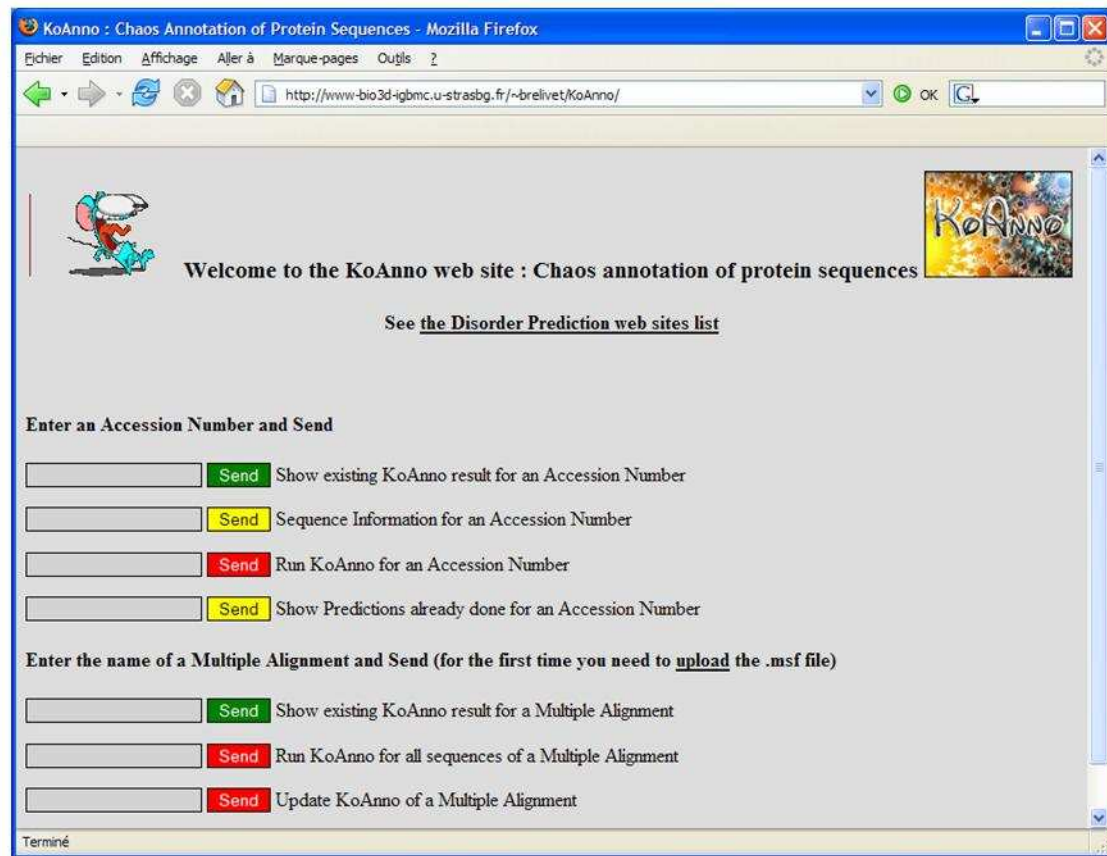


Figure 50 | Interface de KoAnno.

Le résultat de l'analyse d'une protéine, par exemple le RN aux glucocorticoïdes de l'homme (hGR) dont l'*Accession number* est P04150 s'affiche sur une page HTML où toutes les prédictions du désordre sont représentées en orange (Figure 51).

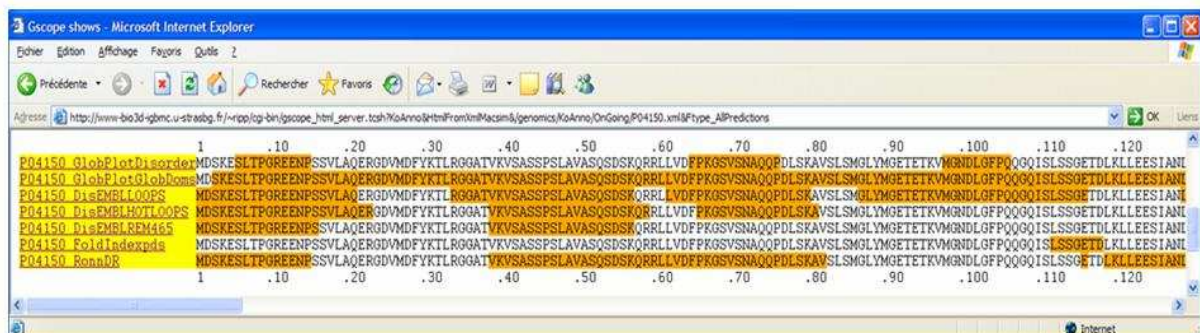


Figure 51 | Capture d'écran présentant les résultats de prédiction du désordre (orange) des différents programmes pour les 120 premiers acides aminés de l'AC P04150 correspondant à hGR.

6.1.2.3. Applications

6.1.2.3.1. Les 48 RNs humains

Le MACS des 48 RNs humains a été extrait de celui comprenant tous les RNs des Métazoaires. Pour faciliter la comparaison des différentes prédictions obtenues, les résultats sont représentés par des courbes indiquant le pourcentage de résidus prédits désordonnés par un programme pour chaque position du MACS (Figure 52).

Si nous analysons les prédictions obtenues pour les deux domaines structurés et bien caractérisés, le DBD et le LBD, seuls RONN et DisEMBL_REM465 prédisent correctement le DBD comme non désordonné tandis que le LBD (à l'exception de quelques boucles) est globalement prédit comme structuré par l'ensemble des programmes. Nous pouvons noter que DisEMBL_LOOPS sur-prédit la présence de boucles, alors que DisEMBL_HOTLOOPS semble les sous-prédire. Concernant les piètres résultats obtenus par FoldIndex dans le domaine DBD, ils pourraient être liés à son algorithme basé sur l'hydrophobicité et sur la présence de charge nette des résidus. Or les DBDs des RNs sont classiquement composés de nombreux acides aminés basiques, tandis que les LBDs (bien prédits) sont plutôt riches en acides aminés hydrophobes.

De même, concernant le LBD, nous pouvons noter que les différentes structures secondaires (hélices et feuillet) sont généralement bien prédites. La région située entre H1 et H3 apparaît relativement désordonnée. C'est une prédiction correcte justifiée par l'absence de densité dans de nombreuses cartes de densité électronique déterminées lors de la résolution de structures de LBDs par cristallographie aux rayons X. Cette région correspond à un domaine désordonné de type plus ou moins transitoire. En effet, la forme *apo* (en absence de ligand) de certains LBDs présente une hélice additionnelle (H2) dans la région connectant les hélices H1 et H3. Dans la forme *holo* (présence d'un ligand), l'hélice H2 se déplie permettant à la boucle H1-H3 d'interagir plus fortement avec le reste de la protéine. Cette conformation différente de la boucle H1-H3 illustre probablement la dynamique de cette région accompagnant le mouvement de H3 qui subit un changement conformationnel lors de la liaison d'un ligand agoniste ou antagoniste.

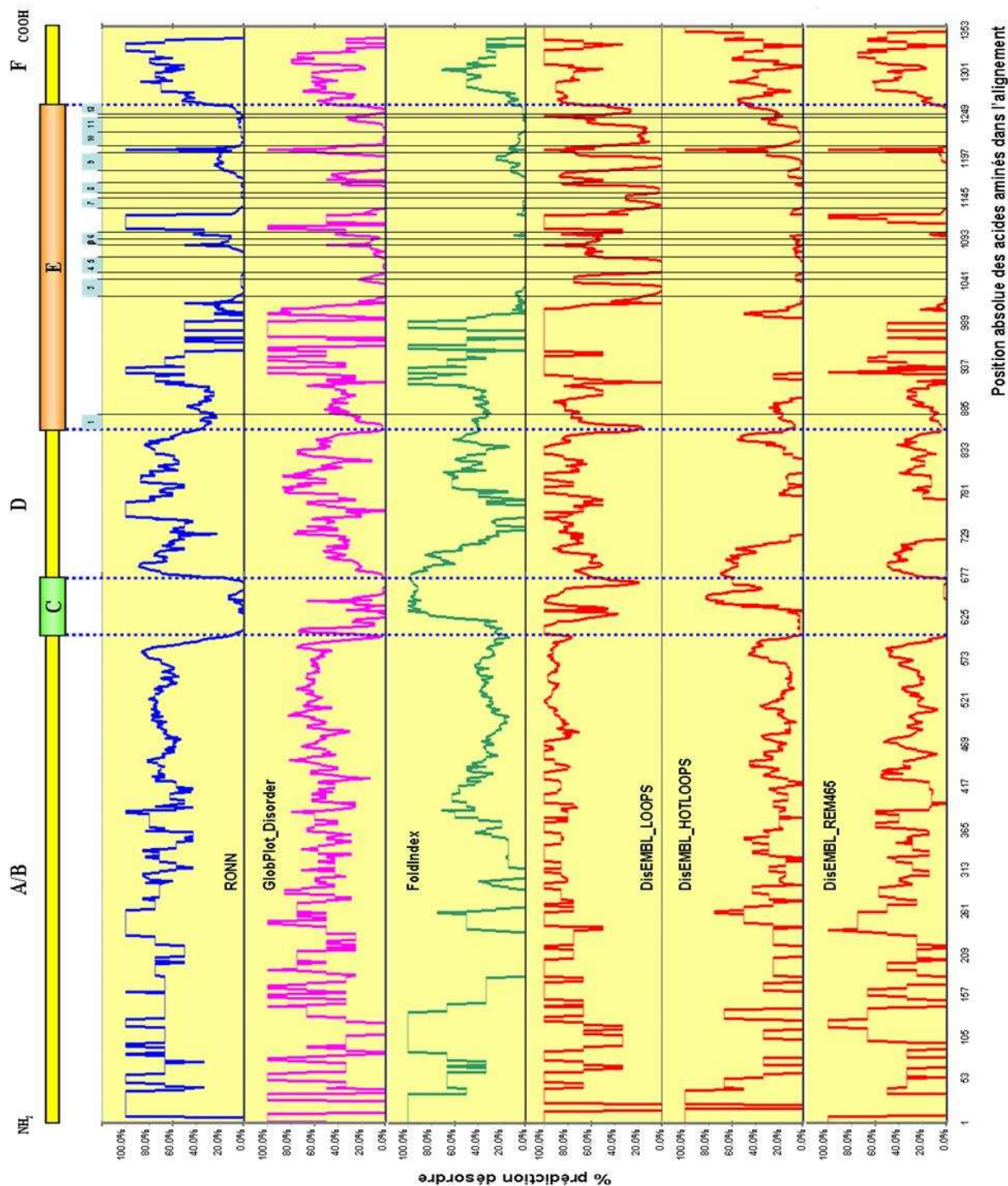


Figure 52 | Bilan des prédictions du désordre des différents programmes concernant les 48 RNs humains. Chaque courbe représente le pourcentage de prédiction du désordre d'un programme en fonction de la position absolue des résidus dans le MACS des 48 RNs humains. Chaque point, d'une position absolue du MACS, correspond au nombre de résidus prédits désordonnés par un programme, divisé par le nombre de séquences ayant un acide aminé à cette position. La valeur finale est ramenée à un pourcentage. Le résultat du programme RONN est coloré en bleu ; GlobPlot_Disorder en mauve ; FoldIndex en vert ; DisEMBL_LOOPS, DisEMBL_HOTLOOPS et DisEMBL_REM465 en rouge. L'organisation structurale (A/B, C, D, E et F) d'un RN est superposée aux courbes afin d'en faciliter l'analyse en délimitant les différents domaines. Les structures secondaires (hélices α et feuillet de brins β) du LBD sont représentées en bleu clair.

Pour les domaines A/B, D et F non caractérisés structuralement, les résultats sont beaucoup plus divergents. En raison de la très grande variabilité de longueur et de séquence du domaine A/B, les acides aminés, souvent caractéristiques d'une sous-famille de RNs, ne sont pas parfaitement alignés entre les différentes sous-familles, entraînant une difficulté d'analyse par rapport à la position absolue des résidus. Cependant, les résultats de prédiction du domaine A/B confirment les grandes tendances et plus particulièrement la sur-prédiction du désordre par DisEMBL_LOOPS.

6.1.2.3.2. Le RN des oestrogènes ER α

La prédiction du désordre a été testée sur le MACS de la sous-famille des RNs des oestrogènes (ER) et plus particulièrement l'isoforme ER α (Figure 53). Les grandes tendances observées précédemment lors de l'analyse des 48 RNs humains sont retrouvées : la sur-prédiction générale de boucles par DisEMBL_LOOPS et la sous-prédiction par DisEMBL_HOTLOOPS ; le DBD correctement prédit non désordonné par RONN et DisEMBL_REM465 ; les structures secondaires du LBD sont généralement bien prédites par les différents programmes de prédiction.

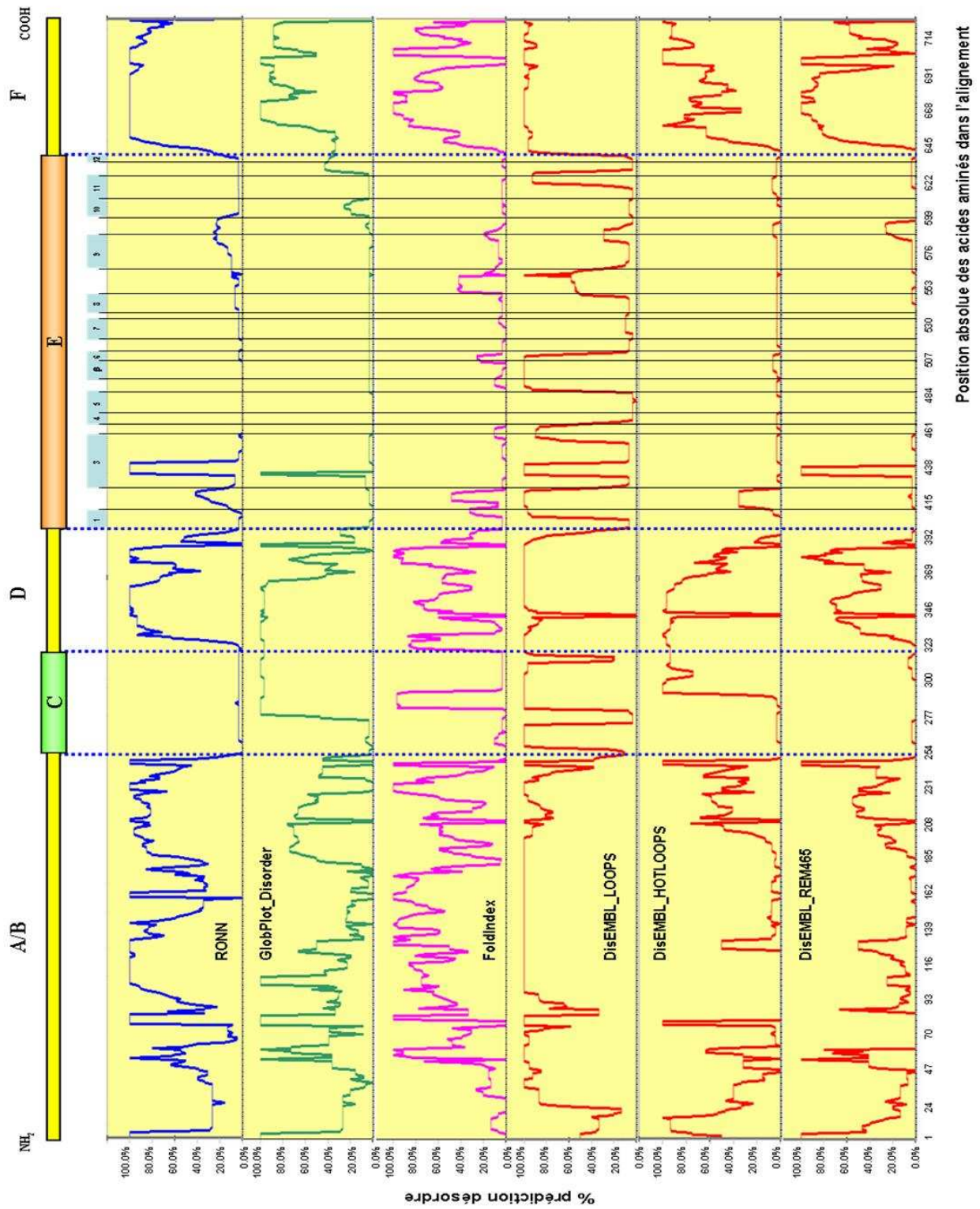


Figure 53 | Bilan des prédictions du désordre des différents programmes concernant ERα. Les conventions sont les mêmes que celles utilisées dans la Figure 52.

Dans le cas d'ER α , il nous est possible d'analyser de façon fiable les prédictions de désordre des domaines A/B qui sont correctement alignés entre eux. Il semble qu'une petite région autour de la position 70, prédite comme ordonnée, se détache de l'ensemble du domaine A/B. Il est intéressant de noter qu'à cette région correspond, dans le MACS de cette sous-famille, un motif conservé de type $\phi x \phi$ avec ϕ un acide aminé aromatique et x un acide aminé quelconque (Figure 54).

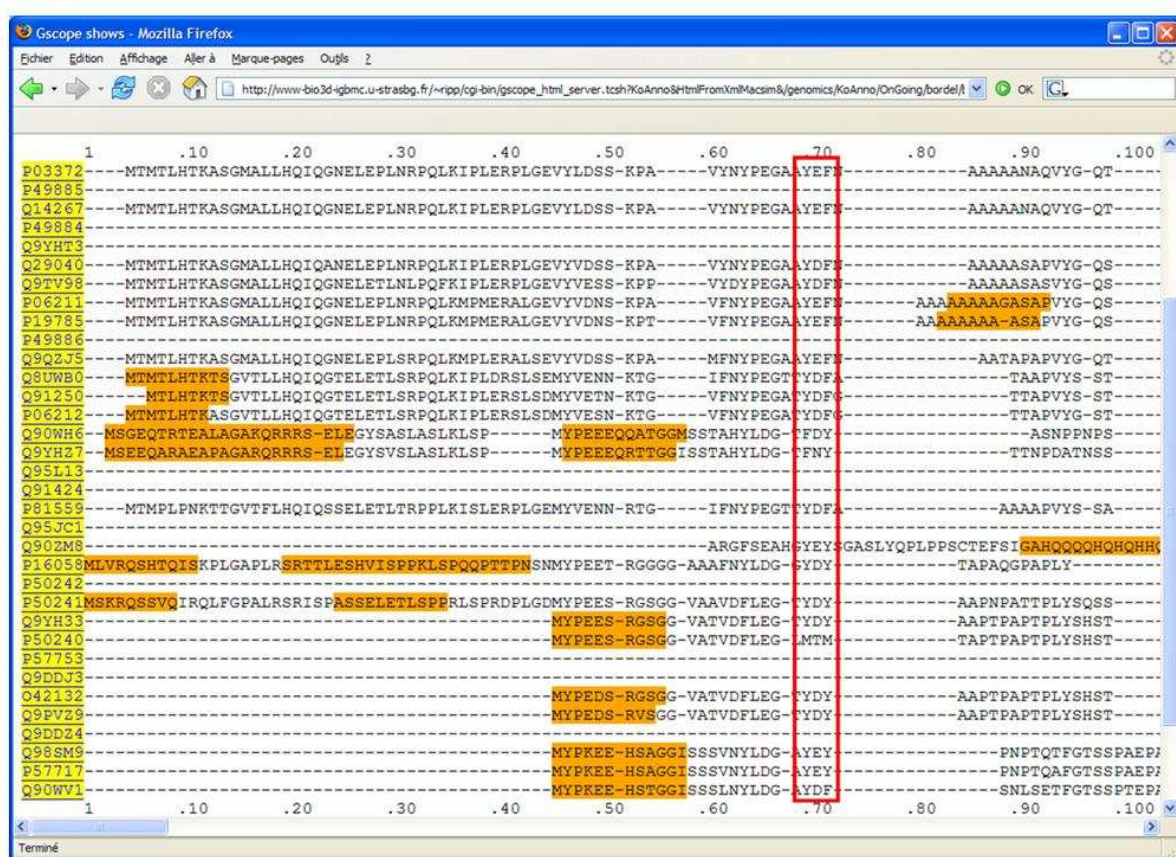


Figure 54 | Région prédite ordonnée dans le domaine A/B de ER α .

Il est intéressant de noter que ce motif est présent chez tous les récepteurs stéroïdiens de la sous-famille NR3 à l'exception de ERR. Cette observation pourra servir de piste pour la définition de nouveaux domaines pour de prochains tests de cristallographie.

6.2. Le DBD

Résumé de l'article

Les RNs forment une super-famille de facteurs de transcription qui régulent divers processus biologiques comme la croissance, le développement et l'homéostasie. Il est maintenant établi que les fonctions génomiques des RNs, dans le noyau, sont activées suite à la fixation de ligands lipophiles. Une fois en dehors du noyau, les RNs peuvent assurer des fonctions non génomiques additionnelles rapides et indépendantes de la transcription des gènes. Cependant, peu d'information est disponible quant à l'activation ou la répression de ces fonctions non génomiques. Un mécanisme évident est de modifier la localisation nucléaire de ces récepteurs. La phosphorylation régule divers aspects de la fonction des RNs, indépendamment de la fixation du ligand. Les RNs sont phosphorylés dans leurs domaines de transactivation (domaine A/B et LBD) et dans le DBD. La phosphorylation du domaine A/B et du LBD par différentes kinases peut affecter la dimérisation, le recrutement de cofacteurs et l'activité transcriptionnelle. La phosphorylation du DBD a aussi montré une diminution dans la dimérisation et dans la fixation à l'ADN de plusieurs RNs. Parce que de nombreux signaux de localisation nucléaire (NLS, *Nuclear Localization Signal*) ou d'export nucléaire (NES, *Nuclear Export Signal*) sont localisés dans le DBD, il est possible que la phosphorylation du DBD affecte la localisation nucléaire des RNs.

Nous avons montré, chez le RN orphelin HNF4 α , que la protéine kinase C (PKC) phosphoryle une Sérine (S78) très conservée située entre les deux doigts de zinc du DBD. Cette S78 est adjacente à plusieurs résidus chargés positivement (Arginine ou Lysine) qui sont impliqués dans la localisation nucléaire de HNF4 α . Un mutant mimant la phosphorylation, S78D, diminue la fixation à l'ADN, la capacité transcriptionnelle, la stabilité de HNF4 α et modifie également la localisation nucléaire du récepteur. Les résultats ont également montré que la PKC phosphoryle le DBD d'autres RNs et que les mutants équivalents, mimant la phosphorylation, induisent une localisation cytoplasmique des récepteurs. De plus, la modélisation de la phosphorylation de cette S78 a apporté une aide à la compréhension du mécanisme responsable de la diminution de la fixation à l'ADN.

Cette S78 étant conservée (Ser/Thr) chez tous les RNs non stéroïdiens, nous proposons que cette phosphorylation pourrait être un mécanisme commun de régulation de la fonction des RNs.

Publication N°2

Phosphorylation of a Conserved Serine in the Deoxyribonucleic Acid Binding Domain of Nuclear Receptors Alters Intracellular Localization

Kai Sun, Vedrana Montana, Karthikeyani Chellappa, Yann Brelivet, Dino Moras, Yutaka Maeda, Vladimir Parpura, Bryce M. Paschal, Frances M. Sladek

Molecular Endocrinology, 2007, Vol. 21, N° 6, Pages 1297–1311

Pages 169-183 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1210/me.2006-0300>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

Chapitre 7 - Caractérisation des voies de communication

Le but fondamental dans la signalisation est de comprendre la communication, ce qui, dans le cadre d'une protéine, peut se traduire par le processus par lequel un signal émis d'un site fonctionnel d'une protéine se propage de manière fiable à un autre site fonctionnel distant.

L'analyse bioinformatique du MACS des LBDs des RNs a révélé une surprenante architecture d'interactions entre des acides aminés conservés et différenciellement conservés au sein des sous-familles. En effet, un petit nombre de résidus forme des réseaux physiquement connectés qui relient des sites fonctionnels distants dans la structure tertiaire des RNs : pour la classe I, de l'interface de fixation des cofacteurs, laquelle fixe des complexes protéiques de régulation qui modulent l'activité transcriptionnelle, à l'interface de dimérisation qui permet une interaction avec un partenaire ; pour la classe II, de la LBP, qui interagit avec de petites molécules lipophiles, à l'interface de dimérisation.

A ce stade, nous étions intellectuellement confrontés à un problème majeur : de quelle manière caractériser au mieux ces deux voies de communication qui gouvernent l'oligomérisation (homo- ou hétérodimérisation) des RNs ?

Une stratégie ambitieuse a été élaborée qui a nécessité la construction de nombreux mutants afin de remplacer progressivement au sein de RNs tests les résidus de la classe I par ceux de la classe II et *vice versa*. Trois RNs humains tests ont été choisis pour leur comportement oligomérique différent : hERR3 qui est un récepteur homodimérique constitutivement actif (sans ligand), hRAR α qui hétérodimérise avec hRXR α et hRXR α qui peut lui même homodimériser (Figure 55). Concernant la genèse des multiples mutants des trois systèmes tests, deux possibilités s'offraient à nous. La première était d'obtenir la totalité des mutants à un temps t et ensuite de les étudier. La deuxième possibilité, que nous avons choisie, était de caractériser un, ou une série de, mutant dès qu'il était construit et de continuer le clonage des autres mutants en parallèle.

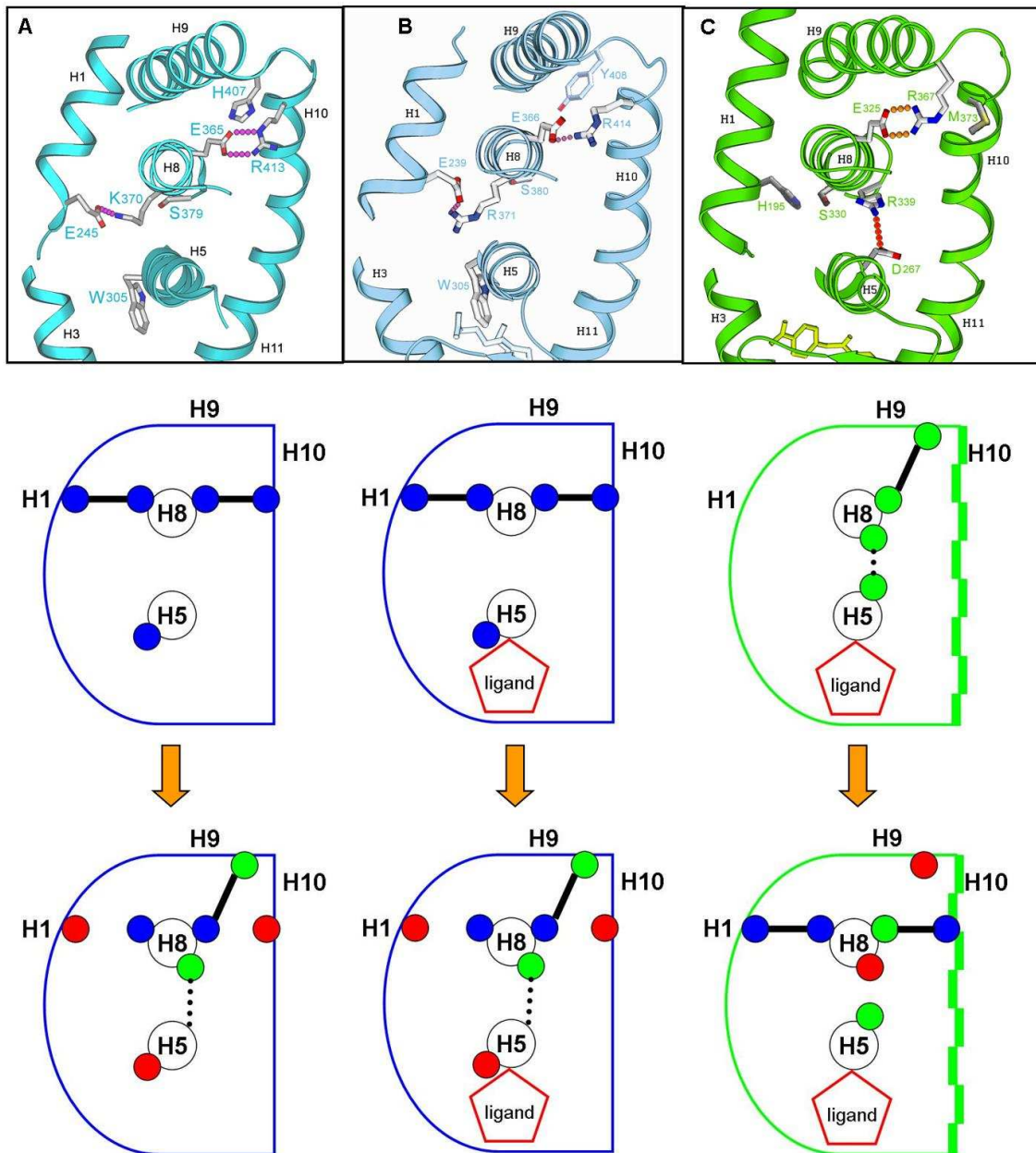


Figure 55 | Les trois RNs tests. En haut, voies de communication chez (A), hERR γ , (B) hXR α et (C) hRAR α . Au milieu, la schématisation des voies de communication dans les différents RNs. En bas, les quintuples mutants correspondant pour chaque système à l'échange de classe. Les points bleus correspondent aux résidus spécifiques de la classe I de hERR γ et hXR α , les points verts aux résidus spécifiques de la classe II de hRAR α , les points rouges à la mutation d'un résidu d'une classe en alanine. Le trait en pointillés représente une connexion située spatialement en arrière de celle en trait plein.

7.1. Genèse des mutants

L'étude d'un système complet nécessite la construction de 32 mutants par RN test, soit 96 mutants pour l'ensemble des trois RNs tests hERR γ , hRAR α et hRXR α . Dans ce but, notre stratégie a permis d'optimiser au maximum l'ensemble du processus. Si tout fonctionne correctement, une semaine est nécessaire pour l'ajout d'une mutation une fois le vecteur de référence obtenu. Dans le cadre de la stratégie mise en place, l'obtention validée d'un mutant implique l'utilisation de deux oligonucléotides de mutation, quatre oligonucléotides de séquençage (deux pour les sites de recombinaison et deux pour la mutation) et donc quatre réactions de séquençage, deux ou trois réactions LR (voir section 3.1.8.1.4) et suivant le cas deux sous-clonages dans les vecteurs d'expression eucaryote.

La méthode employée pour la construction des différents mutants combine la technologie du clonage Gateway® (voir 3.1.8.1) et la technique de mutagenèse dirigée RCA (voir 3.1.9) (Figure 56). Le choix des limites des domaines a été basé sur les structures résolues par Bourguet et *coll.* (Bourguet *et al.*, 2000) soit pour hRAR α des résidus 175 à 462 et pour hRXR α des résidus 223 à 462 et par Greschick et *coll.* (Greschick *et al.*, 2002) soit pour hERR γ des résidus 229 à 458.

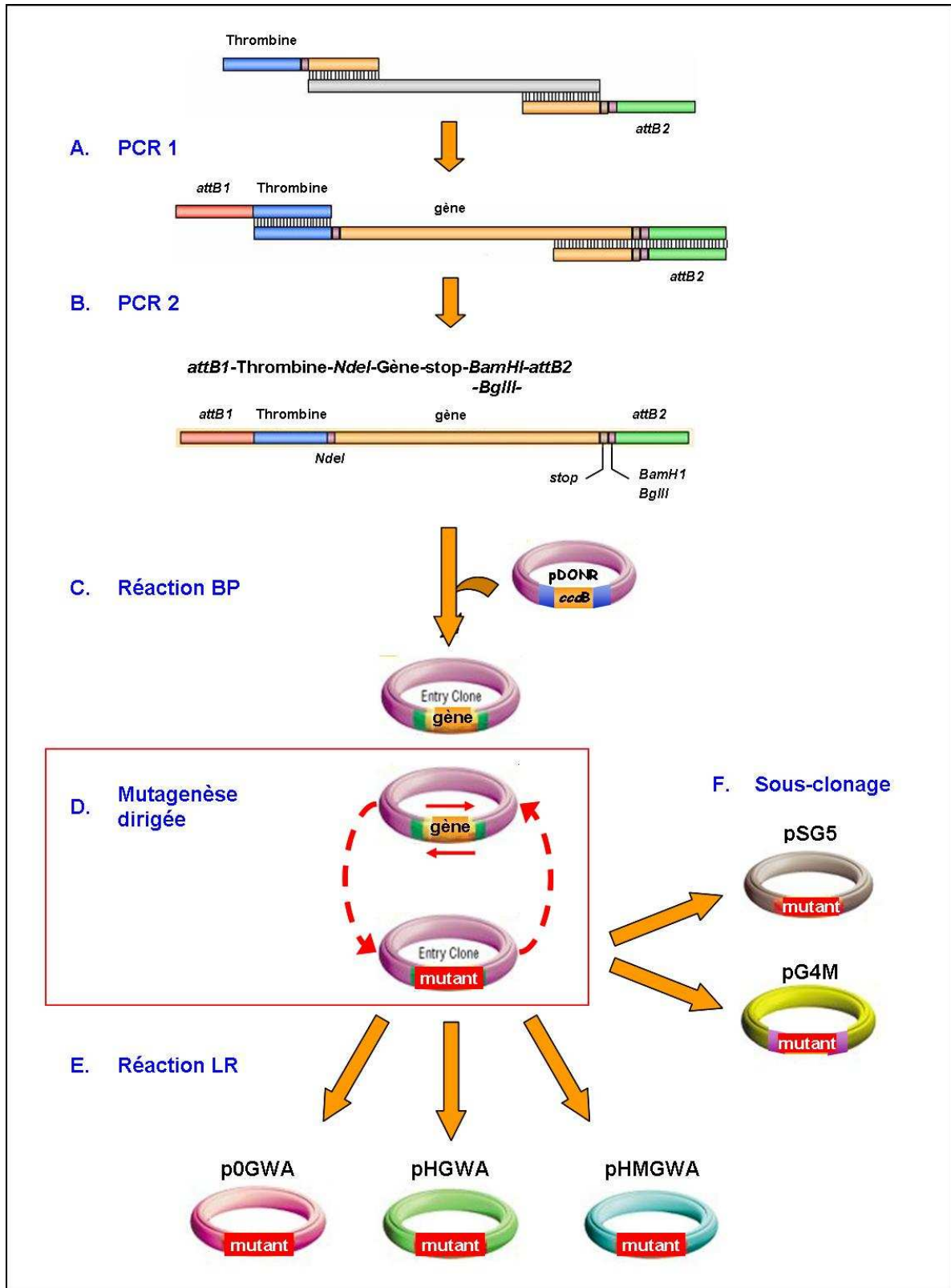


Figure 56 | Protocole de clonage et de mutagenèse.

Parmi les nombreuses et différentes techniques de clonage existantes, le système Gateway apparaît comme un choix adapté dans le cadre d'une politique de clonage à haut débit. La technologie Gateway permet, en effet, un transfert rapide de séquence d'ADN depuis un plasmide navette (pENTR) vers n'importe quel plasmide d'expression (pDEST) compatible. Contrairement à d'autres techniques comme le clonage par restriction-ligation, par LIC (*Ligation Independent Cloning*) ou par RF (*Restriction Free*), la technologie Gateway permet de s'affranchir de l'utilisation d'enzymes spécifiques pour préparer le vecteur et l'insérer comme les enzymes de restriction, la ligase ou encore la T4 ADN polymérase. De nombreux avantages, tels un large catalogue de vecteurs compatibles respectant le cadre ouvert de lecture, le séquençage unique du pENTR, un protocole aisé, sont autant d'atouts à cette technique installée et développée au sein du DBGS. Toutefois ce système nécessite la synthèse de longs oligonucléotides pouvant être à l'origine de problèmes lors de l'amplification par PCR du gène d'intérêt à insérer. Dans la plupart des cas, pour remédier à ce problème, nous utilisons un premier oligonucléotide « sens » qui comporte la séquence codante pour le site de clivage à la thrombine et un site de clivage à la protéase *NdeI* pour les trois RNs, un oligonucléotide « sens » générique qui contient la séquence *attB1* suivie de la séquence codante pour le site de clivage à la thrombine et un troisième oligonucléotide « anti-sens » qui comporte, dans l'orientation 5' vers 3', la séquence *attB2*, un site de clivage à une protéase, *BglII* pour hRXR α et hRAR α ou *BamHI* pour hERR γ et un codon stop (Figure 56A et B et section 3.1.8.1.5). Les produits PCR, correspondants aux trois RNs sauvages, sous-clonés sont vérifiés par séquençage au niveau des sites de recombinaison ainsi qu'au niveau de la séquence du gène d'intérêt.

Le fragment d'ADN est cloné, par la réaction BP, dans le vecteur pDONR207 et génère ainsi un pENTR, notre vecteur de référence pour l'étape de mutagenèse suivante (Figure 56C). On pourra noter une efficacité de clonage de l'ordre de 80%.

C'est à ce stade qu'intervient le processus de mutagenèse dirigée (Figure 56D), mais c'est aussi à ce stade que nous nous sommes heurtés à de nombreuses complications dans l'obtention des différents mutants. Les différentes mutations ponctuelles sont réalisées selon la technique RCA sur les trois pENTR générés précédemment et requièrent l'utilisation de

deux oligonucléotides complémentaires de 33 bases portant la (les) base(s) modifiée(s) correspondant à la mutation désirée en position centrale. Les problèmes rencontrés sont probablement dus à ces oligonucléotides de mutations qui présentent en général un pourcentage élevé en GC de l'ordre de 70%, et une T_m comprise entre 70 et 80°C. L'utilisation d'adjuvant, comme le DMSO, à des concentrations finales variant de 5 à 10% (poids/vol) ou des gradients de température, nous ont en général permis de résoudre ces problèmes. Le séquençage permet de confirmer la mutation et de vérifier si l'ADN polymérase n'a pas introduit de mutation(s) dans les sites de recombinaison. Pour obtenir un double mutant, il est nécessaire que le simple mutant soit validé et ainsi de suite.

La dernière étape du protocole consiste à cloner les différents mutants dans les divers vecteurs d'expression, soit par réaction LR, dans la suite du clonage Gateway, d'un clone d'entrée pENTR avec un vecteur de destination pDEST pour générer un clone d'expression (Figure 56E), soit par un sous-clonage, en utilisant les enzymes de restrictions, dans les vecteurs pSG5 et pG4M (Figure 56F) dans le cadre des études fonctionnelles. Trois pDEST différents ont été utilisés pour des études structurales : p0GWA, pHGWA et pHMGWA. Les cartes des différents vecteurs sont présentées en Annexe 4.

L'expérience du clonage accumulée dans le laboratoire, nous a amené à choisir de placer la séquence codant pour le site de clivage à la thrombine avant le site de restriction *NdeI* afin de pouvoir facilement éliminer de la protéine d'intérêt non seulement d'éventuelle(s) fusion(s) N-ter comme la MBP mais également la séquence protéique codée par la séquence *attB1* (TSLYKKAG) (Annexe 5). Après clivage à la thrombine, il ne subsiste que quatre acides aminés (RGS~~H~~) étrangers à la séquence insérée dans le vecteur.

Afin de réduire la confusion découlant de la complexité et du nombre important de mutants, nous avons été amenés à établir un système simple de codification des mutants qui se base sur un code à trois caractères : une lettre, suivie d'un chiffre, puis une lettre. La première lettre, E, R ou X désigne respectivement le RN hERR γ , hRAR α ou hRXR α . Le chiffre représente le nombre de mutations, 0 pour la protéine sauvage, 1 pour une simple mutation, etc. La dernière lettre permet de différencier chaque mutant possédant le même

nombre de mutation(s). Par exemple le mutant X1c correspond au mutant simple de hXR α W305A (Tableau 11). A ce jour, un grand nombre de mutants (38), répertoriés dans le Tableau 11, sont disponibles pour les systèmes hERR γ , hXR α et hRAR α .

7.2. Les différentes études menées

L'impact sur le comportement des RNs d'une ou de plusieurs mutations jusqu'au changement de classe a été examiné au travers des quatre propriétés : la stabilité, la dimérisation en présence d'un fragment d'ADN contenant l'élément de réponse du RN en absence/présence de 9-*cis*RA, l'activité transcriptionnelle et la structure cristallographique.

En raison de son statut particulier au sein des RNs, nous nous sommes principalement focalisés sur le système hXR α , c'est-à-dire les mutants X1a, X1b, X1c, X3a, X4a, X5 (Tableau 11). Le récepteur hXR α est, en effet, capable d'homodimériser ainsi que d'hétérodimériser avec certains RNs de la classe II. Les études fonctionnelles ont nécessité le sous-clonage des différentes constructions, à partir des différents pENTR correspondants dans les vecteurs d'expression eucaryotes, soit pSG5 dans le cadre de l'étude de la dimérisation, soit pG4M dans le cadre de l'étude de la transactivation.

Protéine	Limites	Mutation	Label	Vecteurs d'expression
hRXR α	223-462	sauvage	X0	pOGWA, pHGWA
hRXR α	223-462	E239A	X1d	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	W305A	X1c	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	S380R	X1e	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	Y408R	X1a	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	R414A	X1b	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/R414A	X2a	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A	X2b	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/S380R	X2c	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/S380R/R414A	X3a	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A/S380R	X3b	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A/S380R	X3c	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A/R414A	X3d	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A/S380R/R414A	X4b	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/S380R/Y408R/R414A	X4a	pOGWA, pHGWA, pHMGWA
hRXR α	223-462	E239A/W305A/S380R/Y408R/R414A	X5	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	sauvage	R0	pOGWA, pHGWA
hRAR α	175-462	H195E	R1a	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	S332K	R1e	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	R339A	R1d	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	R367A	R1b	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	M373R	R1c	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	H195E/R339A	R2a	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	H195E/R367A	R2b	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	R339A/R367A	R2c	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	R367A/M373R	R2d	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	R339A/R367A/M373R	R3a	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	H195E/S332K/R341A	R3b	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	H195E/S332K/R341A/M373R	R4	pOGWA, pHGWA, pHMGWA
hRAR α	175-462	H195E/S332K/R341A/ R367/M373R	R5	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	sauvage	E0	pOGWA, pHGWA
hERR γ	229-458	E245A	E1b	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	S379R	E1a	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	E245A/S379R	E2a	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	S379R/R413A	E2b	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	E245A/W305A	E2c	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	W305A/S379R	E2d	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	E245A/S379R/R413A	E3b	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	S379R/H407R/R413A	E3a	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	E245A/S379R/H407R/R413A	E4	pOGWA, pHGWA, pHMGWA
hERR γ	229-458	E245A/W305A/S379R/H407R/R413A	E5	pOGWA, pHGWA, pHMGWA

Tableau 11 | Liste des mutants pour l'étude structurale.

Chapitre 8 - Etude du système hERR γ

8.1. Les différents mutants testés

Les premiers mutants de hERR γ disponibles sont ceux présentés ci-dessous (Figure 57).

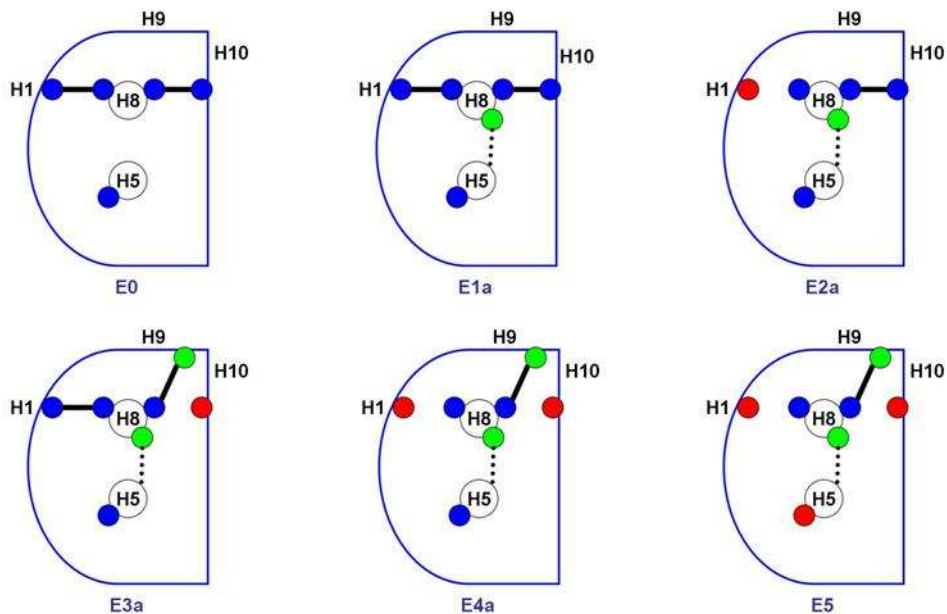


Figure 57 | Les différents mutants hERR γ . Les points bleus correspondent aux résidus spécifiques de la classe I de hERR γ , les points verts représentent la mutation d'un résidu de hERR γ en acide aminé spécifique de la classe II et les points rouges correspondent à la mutation d'un résidu de hERR γ en alanine.

Dans le cas du mutant E1a, il s'agit d'essayer de recréer le pont salin spécifique de la classe II (hétérodimères avec RXR) faisant intervenir l'arginine différenciellement conservée de la boucle H8-H9 avec le résidu acide universellement conservé situé au niveau des hélices 4 et 5. De cette série de constructions, E1a est le seul mutant à ne pas avoir de mutation impliquant un des résidus spécifique de la voie de communication de la classe I. Il s'agit dans ce cas, de l'ajout d'une connexion, qui plus est, interne à la structure, donc à priori sans risque de conséquence majeur sur la stabilité globale du LBD. Dans le double mutant E2a, une deuxième mutation a été introduite en plus de celle de E1a. Dans ce cas, le pont salin joignant les hélices H1 et H8 est brisé. Le triple mutant E3a est caractérisé par la mutation de

E1a, la mutation en alanine de l'arginine différentielle conservée en N-ter de H10 spécifique de la classe I ayant pour effet de déconnecter le deuxième pont salin entre H8 et H10, la mutation d'une histidine en arginine située dans H9 pour tenter de recréer le deuxième pont salin caractéristique de la classe II. Dans le quadruple mutant E4a, les ponts salins de la classe I ont été remplacés par ceux de la classe II. Le quintuple mutant E5 correspond à un récepteur « équivalent » de la classe II. Il s'agit, par ces multiples et diverses mutations, d'étudier le comportement oligomérique de ce récepteur.

8.2. Etude fonctionnelle des mutants de hERR γ

Il est à noter que dans le cas du système hERR γ , nous n'avons pu réaliser l'ensemble des tests fonctionnels. D'une part, n'ayant pas les éléments de réponse de hERR γ , les expériences de retard sur gel n'ont pu être réalisées. D'autre part, pour une raison encore non déterminée (problème de transfection ou autre), les tests de transactivation n'ont pas fonctionné. Ces deux tests sont bien entendu importants dans le cadre de notre étude, mais pour l'instant restent en suspend.

8.3. Etude structurale des mutants hERR γ

Dans cette partie sont décrites les purifications des cinq mutants présentés Figure 57. Le protocole de purification utilisé pour le mutant E1a ci-après est identique pour les suivants. Le LBD (229-458) des différents mutants a été cloné dans un vecteur d'expression pHMGWA (*amp*^R) où le gène codant le mutant se trouve en fusion avec une étiquette polyhistidine en extrémité N-ter suivie d'une fusion avec la protéine MBP (*Maltose Binding Protein*). La MBP est censée augmenter la solubilité et permettre une production plus importante de protéine.

8.3.1. Expression

Les conditions d'expression dans *E. coli* sont détaillées dans la section 3.2.1 La Figure 58 illustre l'expression similaire des différents mutants de hERR γ .

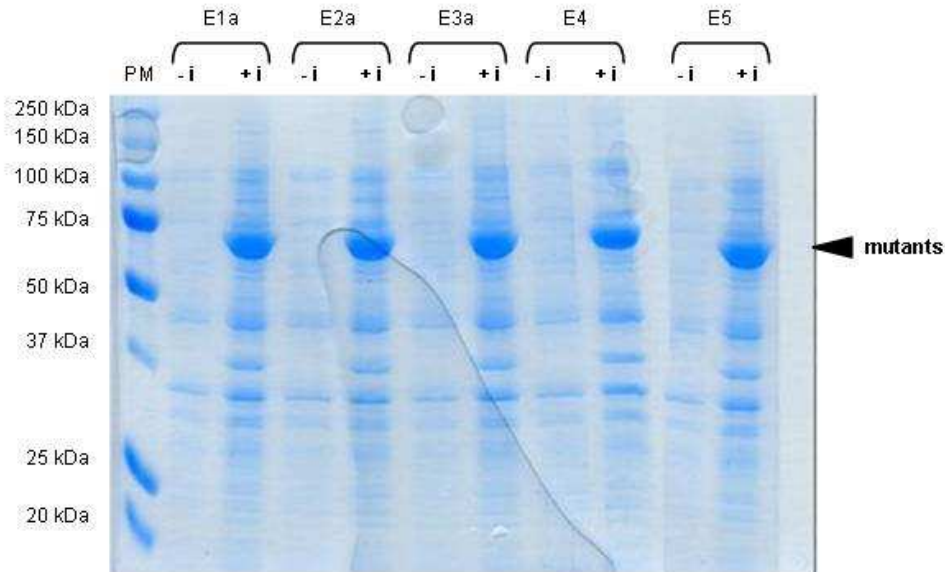


Figure 58 | Expression des mutants E1a, E2a, E3a, E4a et E5. « - i » correspond à une fraction aliquote avant induction, « + i » après induction à l'IPTG.

8.3.2. Lyse cellulaire et ultracentrifugation

La procédure consiste à lyser, principalement par sonication, les cellules des culots provenant des cultures précédentes des différents mutants. Les culots sont re-suspendus dans un volume adéquat de tampon puis lysés. L'extrait brut, issu de la sonication, est ensuite clarifié par ultracentrifugation (2 h à 45000 g à 4°C) (voir sections 3.2.2 et 3.2.3).

8.3.3. Purification des différents mutants

Les conditions expérimentales sont décrites dans la section 3.2.4. Le lysat cellulaire clarifié est purifié selon un protocole en cinq étapes illustré Figure 59 : une chromatographie d'affinité au nickel, une filtration sur gel, une protéolyse par la thrombine pour éliminer l'étiquette polyhistidine et la fusion MBP, une deuxième chromatographie d'affinité au nickel pour ne conserver que le mutant E1a sans fusion et finalement une deuxième filtration sur gel. Le choix de ce protocole est tiré des enseignements de notre laboratoire.

8.3.3.1. Le mutant E1a

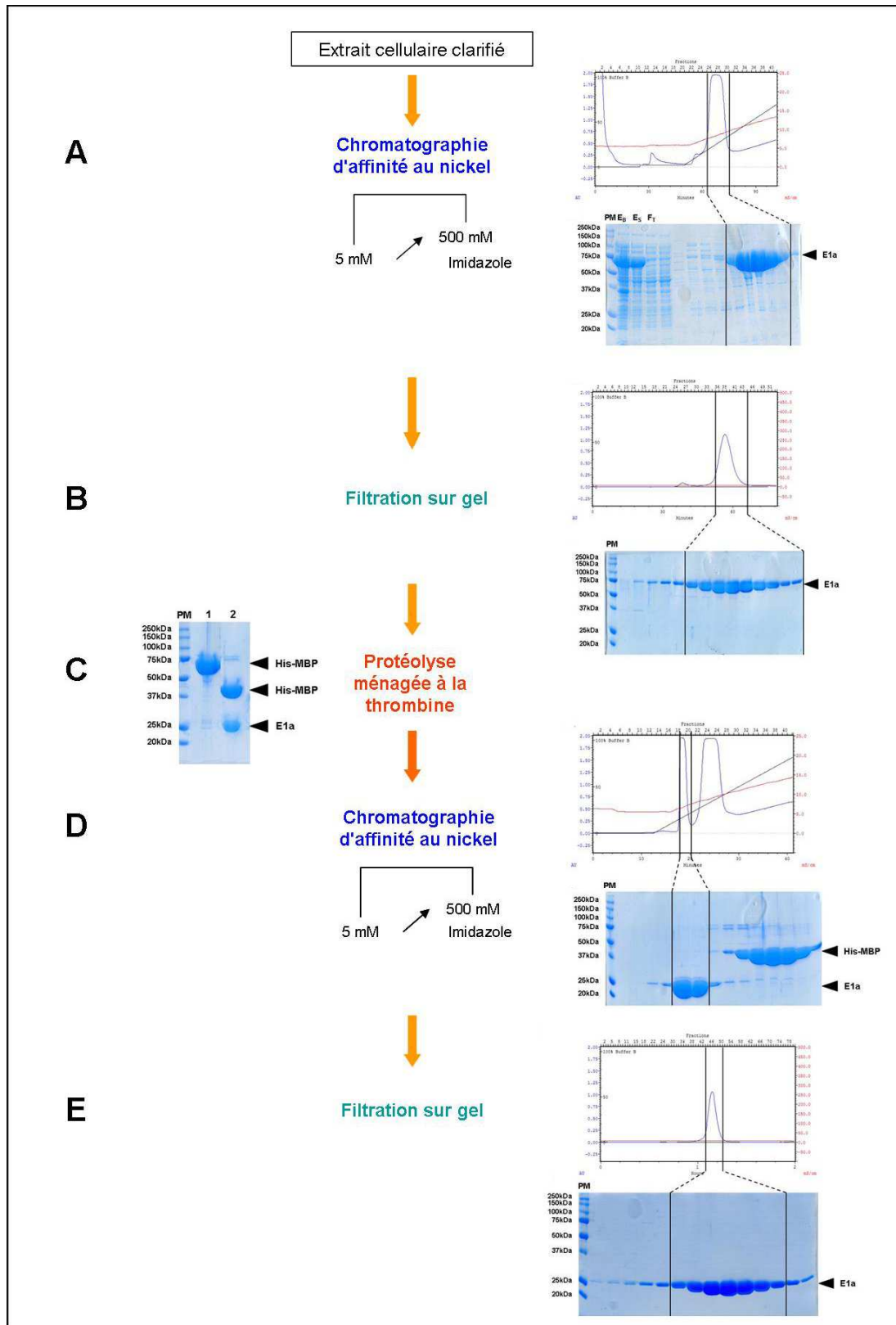


Figure 59 | Schéma de purification du mutant E1a. Chromatogramme : le tracé bleu correspond à la DO_{280nm} , le rouge à la conductivité en mS/cm et le gris au gradient d'Imidazole. Gel SDS-PAGE 12% : PM, marqueurs de poids moléculaire ; E_B, Extrait Brut après sonication ; E_S, Extrait Soluble, F_T, Flow Through.

8.3.3.1.1. Chromatographie d'affinité sur résine chélatant les ions nickel

Le principe de la chromatographie d'affinité est basé sur l'emploi de cations métallique divalents (Zn^{2+} , Ni^{2+} , Cu^{2+} , Co^{2+}) chélatés par des groupements de type iminodiacétate pour séparer des protéines portant des polyhistidines. Le zinc permet l'utilisation de fortes concentrations en agent réducteur, contrairement au nickel et au cuivre, plus sensibles à la réduction du cation métallique en métal. Ces cations immobilisés vont interagir avec la fusion polyhistidine de la protéine et permettre de purifier rapidement et efficacement la protéine. Dans notre situation, aucun agent réducteur n'a été utilisé, notre choix de cations s'est porté sur le nickel. L'élution est réalisée dans un gradient d'imidazole. Lors de cette première chromatographie d'affinité, le mutant E1a est éluée avec 105 mM d'Imidazole. Les fractions de colonne sont analysées sur gel SDS-PAGE 12% (Figure 59A). La masse apparente du mutant E1a est en adéquation à celle théorique de 73 kDa correspondant à l'étiquette polyhistidine (2 kDa), la fusion MBP (45 kDa) et le LBD du mutant E1a (26 kDa). On peut noter la présence de quelques contaminants, l'un d'une masse d'environ 37 kDa et un autre d'environ 40 kDa, probablement des protéines d'*E. coli*, mais en proportion négligeable.

8.3.3.1.2. Filtration sur gel

La filtration sur gel permet de séparer les protéines en fonction de leur masse moléculaire apparente. Cette étape chromatographique permet de caractériser l'échantillon ainsi que d'éliminer toute trace éventuelle d'agrégats ou d'oligomères (Figure 59B). Nous observons deux pics : le premier correspond aux agrégats du mutant E1a sur lequel s'est adsorbé le contaminant d'*E. coli* de 37 kDa que cette étape chromatographique a permis d'éliminer comme l'atteste le gel SDS-PAGE 12% : le pic majoritaire correspond au mutant E1a sous forme homodimérique de masse moléculaire apparente de 140 kDa, sans trace d'agrégat ni d'autres espèces oligomériques.

8.3.3.1.3. Protéolyse ménagée par la thrombine

A ce stade, la protéine présente déjà un excellent degré de pureté, suffisant pour permettre l'obtention de cristaux utilisables dans des expériences de diffraction aux rayons X. Cependant, la grande fusion His-MBP placée en amont de la protéine d'intérêt peut être à

l'origine de sérieux problèmes lors des tests de cristallisation. En effet, la MPB est connue pour former des dimères, tétramères, etc. et ainsi pourrait influencer jusqu'à altérer la propre dimérisation du mutant E1a lui-même. Pour remédier à ce problème, une protéolyse ménagée est réalisée sur la protéine de fusion en utilisant la thrombine bovine plasmatique. Celle-ci est susceptible d'être inhibée par les agents complexants (Imidazole) et en principe les inhibiteurs classiques de protéases à sérine (DIFP). Le fait d'avoir réalisé une filtration sur gel a aussi permis de conditionner le mutant E1a dans un tampon adapté à cette protéolyse. L'étape de protéolyse de la fusion His-MBP a été réalisée en ajoutant 1U de thrombine par mg de protéine. La réaction est incubée à 4°C durant la nuit. Dans ces conditions, la protéolyse est considérée comme totale et seule une faible fraction (3%) correspondant à la protéine non coupée est observée (Figure 59C). De plus, aucune réaction de protéolyse secondaire n'est décelée. Aucune perte de protéine n'est à constater suite à cette coupure (Tableau 12) puisque pour une mole de His-MBP-E1a on récupère une mole de protéine de fusion His-MBP et une mole du mutant E1a. Nous pouvons considérer le rendement de cette étape à 100%.

8.3.3.1.4. Chromatographie d'affinité sur résine chélatant les ions nickel

Pour éliminer la fusion après clivage, le produit de la digestion est chargé sur une colonne pour une deuxième chromatographie d'affinité au nickel qui retient spécifiquement la fusion polyhistidine-MBP, ainsi que la protéine non clivée, alors que la protéine d'intérêt se retrouve dans l'effluent de la colonne. Cependant, il semble que le mutant E1a clivé interagisse encore, certes faiblement, avec la colonne, probablement en raison des résidus His présent à la surface du LBD. En effet, le chromatogramme montre deux pics qui correspondent respectivement au mutant E1a (26 kDa) élué en début de gradient à environ 10 mM d'Imidazole et à la fusion polyhistidine-MBP (47 kDa) (Figure 59D).

8.3.3.1.5. Filtration sur gel

Afin de vérifier si l'élimination de la protéine de fusion MBP n'a pas eu de conséquence dramatique sur le repliement et la solubilité de la protéine, qu'elle ne soit pas agrégée, une deuxième filtration sur gel a été réalisée (Figure 59E). Un avantage supplémentaire de cette étape chromatographique est de conditionner le mutant E1a en modifiant le tampon en vue des tests de cristallisation. Un seul pic est observé,

correspondant au mutant E1a sous forme homodimérique (52 kDa) sans trace d'agrégat ni d'autres espèces oligomériques, et présentant une excellente pureté (supérieure à 98%).

A ce stade, le mutant E1a est stable dans le tampon de la filtration sur gel (500 mM NaCl, 50 mM TrisHCL pH=8) sans présence de la protéine de fusion polyhistidine-MBP et la mutation ne semble pas avoir déstabilisé la structure du LBD.

8.3.3.1.6. Rendement de purification

Le rendement de purification est de l'ordre de 13 mg du mutant E1a pour 1 l de culture, quantité suffisante pour des tests de cristallisation (Tableau 12).

E1a	Volume (ml)	Concentration (mg/ml)	Quantité (mg)	Rendement (%)
Chromatographie d'affinité 1 au nickel	25	3,5	87,5	
Concentration 1	3	24,2	72,6	83
Filtration sur gel 1	13,5	5,2	70,1	83
Clivage à la Thrombine				
Chromatographie d'affinité 2 au nickel	4	6,3	25,0	100
Concentration 2	1,1	12,5	13,8	55
Filtration sur gel 2	10,5	1,2	12,6	92

Tableau 12 | Rendement de purification du mutant E1a.

8.3.3.1.7. Pureté et homogénéité du mutant E1a

La deuxième filtration sur gel a montré un échantillon d'une très grande pureté (Figure 59E). Après purification, la protéine a été caractérisée par électrophorèse en condition native (Figure 60) et par spectrométrie de masse en conditions dénaturantes (Figure 61) et natives (Figure 62).

8.3.3.1.7.1. *Gel natif*



Figure 60 | Gel natif (8-25%) de hERR γ . (1), hERR γ sauvage ; (2), E1a.

Le gel natif montre une migration équivalente entre hERR γ sauvage et le mutant E1a indiquant donc que ce dernier est homogène.

8.3.3.1.7.2. *Spectrométrie de masse*

La spectrométrie de masse a été utilisée pour mesurer la masse des macromolécules purifiées et la comparer à la masse théorique attendue de façon à être certain qu'aucune dégradation n'ait eu lieu au cours de leur purification. Le résultat obtenu est un spectre de masse représentant les rapports m/z (masse/charge) des ions détectés selon l'axe des abscisses et l'abondance relative de ces ions selon l'axe des ordonnées. Le mutant E1a a été caractérisé par spectrométrie de masse en conditions dénaturantes et natives.

La spectrométrie de masse en conditions dénaturantes montre l'existence d'une protéine de masse 26535,21 Da qui correspond à la masse attendue de 26533,96 Da (Figure 61).

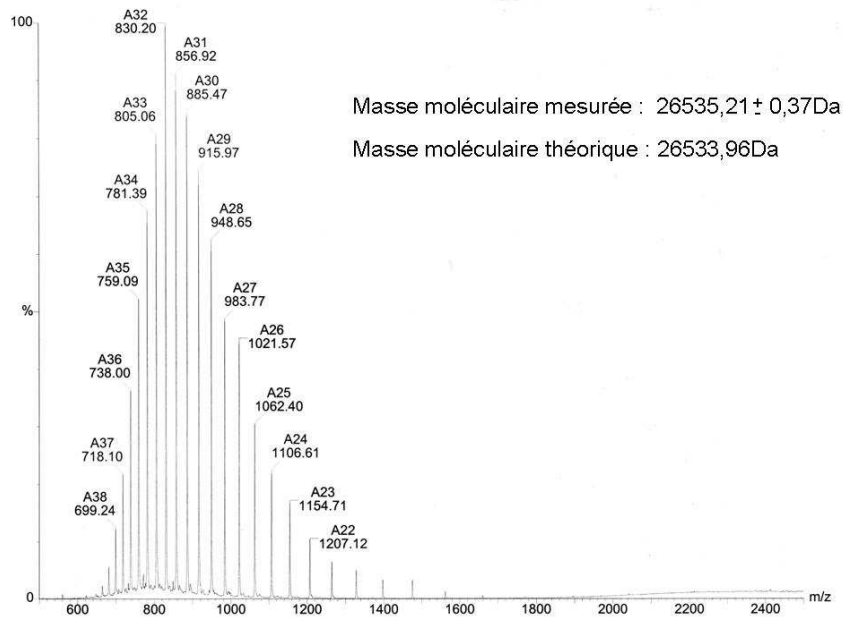


Figure 61 | Spectre de masse en conditions dénaturantes du mutant E1a.

En conditions natives, deux résultats inattendus ont été révélés (Figure 62). Dans le cas de ERR γ sauvage, une seule espèce a été détectée : la forme homodimérique (Greschik *et al.*, 2002). Contrairement à ERR γ sauvage, l'analyse du mutant a montré l'existence de deux espèces : une forme homodimérique mais surtout la présence d'une forme monomérique. Cette analyse n'étant pas quantitative, nous ne pouvons pas savoir laquelle des deux espèces est majoritaire ou s'il existe un équilibre. Cette mutation dont le rôle semblait anodin au niveau de la dimérisation apparaît en réalité intervenir dans ce processus. Qui plus est, cette arginine, impliquée dans le pont salin caractéristique de la classe II, ne se situe pas au niveau de l'interface de dimérisation.

Plus surprenant encore, cette spectrométrie de masse en conditions natives a également révélé la présence d'un ligand d'une masse de 182 Da. Il s'agit encore ici, d'une différence notable de ERR γ sauvage qui est constitutionnellement actif (Greschik *et al.*, 2002).

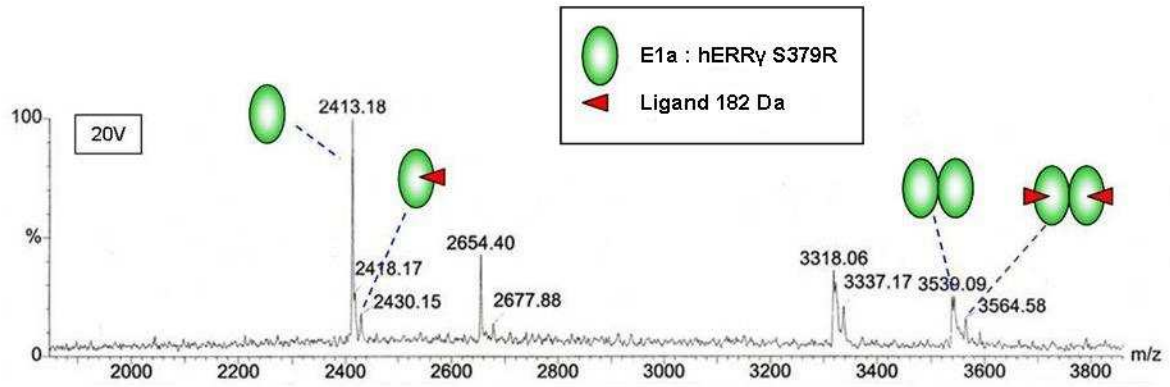


Figure 62 | Spectre de masse en mode natif (ESI-TOF) effectué sur le mutant E1a.

Ces analyses nous permettent, par ailleurs, de confirmer que nous avons bien purifié le mutant E1a dans son intégralité, sans qu'aucune protéolyse ne se soit produite au cours de la production ou de la purification, qu'il n'y a pas eu de modifications post-traductionnelles et qu'il ne reste pas de contaminant.

8.3.3.2. Expression et purification du mutant E2a

La masse de cellules obtenue, après récolte, est d'environ 2,2 g/l de culture. Cependant, le rendement de purification (Tableau 13) est de l'ordre de 339 μ g de protéine pure du mutant E2a pour 1 l de culture.

E2a	Volume (ml)	Concentration (mg/ml)	Quantité (mg)	Rendement (%)
Chromatographie d'affinité 1 au nickel	20	1,2	24,6	
Concentration 1	3	6	18,2	74
Filtration sur gel 1	7,5	1,6	11,6	64
Clivage à la Thrombine				
Chromatographie d'affinité 2 au nickel	1,5	0,9	1,41	30
Filtration sur gel 2	6	0,06	0,34	24

Tableau 13 | Rendement de purification du mutant E2a.

Il est intéressant de noter, que pour un volume équivalent de culture du mutant E1a, environ quatre fois moins de protéine soluble ont été obtenus à la fin de la première

chromatographie d'affinité au nickel. La quantité finale de protéine du mutant E2a soluble pure s'en ressent bien évidemment aussi. De plus, lors du clivage à la thrombine, plus de 70% de la protéine a été perdue. Pour rappel, dans le cas du mutant E2a, contrairement au mutant E1a, un résidu de la voie de communication, à savoir le résidu acide de l'hélice H1 a été muté en alanine. Il semble donc que ce mutant ne soit pas aussi stable que le premier. A ce stade, l'expression étant similaire à celle du mutant E1a (Figure 58), la purification devra être optimisée, de manière à obtenir une plus grande quantité de protéine mais surtout beaucoup plus stable. Dans cette optique, différents paramètres pourront être modifiés :

- la concentration et la nature du sel : NaCl, KCl, NH₄Cl
- l'ajout de détergents destinés à améliorer la solubilité : CHAPS
- le tampon utilisé
- le volume de tampon de lyse

8.3.3.3. Expression et purification des mutants E3a, E4a et E5

La masse de cellules obtenue pour chacun de ces trois mutants, après récolte, est de l'ordre de 3,7 g/l de culture. Le comportement de ces mutants est identique lors de leur purification qui s'est limitée à deux étapes, une chromatographie d'affinité au nickel suivie d'une filtration sur gel. A l'issue de cette filtration sur gel, les mutants sont élués dans le volume mort de la colonne et sembleraient donc être tous sous une forme agrégée (Figure 63). Cependant, nous avons bon espoir d'obtenir de la protéine soluble. En effet, il apparaît une « traîne » derrière le pic qui correspond au mutant non agrégé, bien sur encore en fusion avec la MBP. A ce stade, il faudrait reprendre en considération les conditions d'expression, bien que le profil d'expression soit identique pour tous les mutants, et bien sur aussi les conditions de purification. D'autres bactéries pourront être testées, ainsi que la température et la durée d'induction. Différentes protéines de fusion pourront être utilisées comme la thioredoxine qui n'a pas tendance à dimériser.

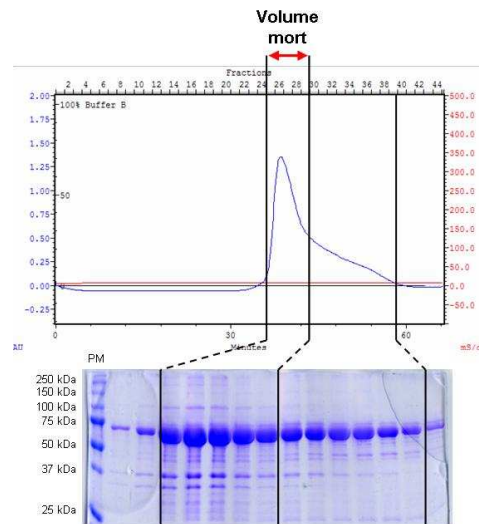


Figure 63 | Purification des mutants E3a, E4a et E5 par filtration sur gel.

8.3.3.4. Bilan des purifications

A ce stade de l'expérience nous pouvons déjà retirer un enseignement concernant l'impact des différents mutants sur la stabilité de la structure (Figure 64).

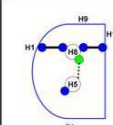
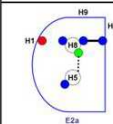
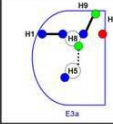
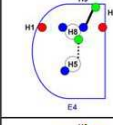

	Expression	Purification
 <p>E1a</p>	✓	✓
 <p>E2a</p>	✓	✓
 <p>E3a</p>	✓	∅
 <p>E4</p>	✓	∅
 <p>E5</p>	✓	∅

Figure 64 | Bilan de purification des différents mutants.

En effet, plus il y a de mutations, moins le mutant est soluble. Mais surtout, la mutation d'un résidu spécifique de la voie de communication de la classe I engendre une instabilité structurale et par voie de conséquence un problème de solubilité.

8.3.4. Cristallisation du mutant E1a

Dans le cadre de l'étude structurale du mutant E1a, nous avons pu bénéficier des équipements disponibles sur la plate-forme de Biologie et de Génomique Structurales de Strasbourg. Ces facilités, issues du développement de la génomique structurale à haut débit, nous permettent de réaliser à la fois un gain de temps et une économie de matériel biologique non négligeables.

Le mutant E1a, purifié à plus de 98% a été concentré à 5 mg/ml, puis co-cristallisé en présence d'un excès de peptide contenant le second motif LXXLL du domaine d'interaction de SRC-1 (686-RHKILHRLQLQEGSPS-700) selon la technique de diffusion de vapeur en goutte assise (1 μ l de la solution E1a/SRC-1 + 1 μ l de solution de cristallisation contre un réservoir de 80 μ l). Dans le cadre de la cristallisation de ERR γ sauvage, la présence du peptide SRC-1 s'est révélée être indispensable dans l'obtention de cristaux (communication personnelle, Carole Peluso-Iltis) en permettant de plus le positionnement correct et le maintien de l'hélice H12 dans une conformation agoniste stable. L'utilisation des kits *The Ammonium Sulfates*, *The Classics*, *The JCSG+ screen*, *The PEG-Ion pH* et *NR-LBD 96* s'est montrée efficace en nous permettant d'obtenir des cristaux en quelques jours à 4°C et 24°C (Figure 65). Les conditions de cristallisation du mutant E1a sont différentes de celle qui a permis d'obtenir la structure tridimensionnelle de ERR γ sauvage (100 mM TrisHCl pH8, 1,5 M ammonium sulfate et 15% glycérol).

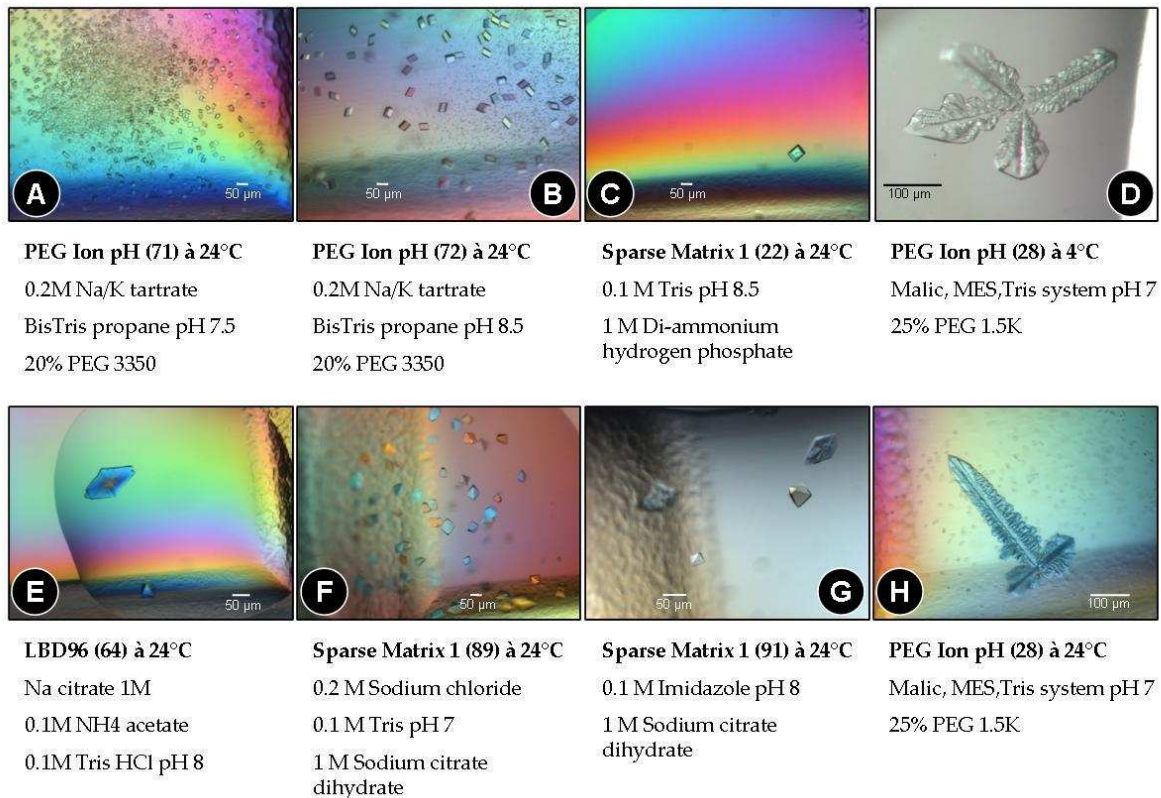


Figure 65 | Cristaux du mutant E1a en complexe avec le peptide SRC-1. Le nombre entre parenthèses correspond à la condition du *screen*.

Ces nombreux cristaux diffèrent selon la taille, inférieure à 50 μm (Figure 65A, B, C et F) jusqu'à 300 μm (Figure 65D et H) et la forme : parallélépipède (Figure 65A et B), bipyramide (Figure 65C, E, F et G) et « style fougère » (Figure 65D et H). La majorité des cristaux ont été obtenus à une température de croissance de 24°C. Seule une condition nous a permis d'obtenir des cristaux à 4°C et 24°C, cependant ces cristaux n'étant pas mono cristallins, ils n'ont pas été utilisés. Il est intéressant de noter que l'augmentation du pH, de 7,5 à 8,5, dans le cas des conditions A et B a permis d'augmenter la taille et d'améliorer la forme des cristaux. Cependant, ces microcristaux, de petite taille (moins de 50 μm) n'étaient pas utilisables pour une expérience de diffraction aux rayons X. La condition E nous paraissait être de bonne facture, qui plus est, le cristal, obtenu dans le *screen* NR-LBD96 spécialement dédié aux RNs, est de taille correcte de l'ordre de 75 x 75 x 150 μm^3 .

8.3.5. Du cristal à la structure du mutant E1a

Les données cristallographiques ont été traitées avec la suite logicielle HKL2000 (Otwinowski and Minor, 1997) et la plupart des calculs cristallographiques ont été effectués par la chaîne de programmes CCP4 (Collaborative Computational Project, 1994).

8.3.5.1. Collecte des données

Le cristal de la condition 64 du *screen NR-LBD 96* (Figure 65E) a directement été utilisé au synchrotron de l'ESRF à Grenoble (www.esrf.fr/) en utilisant de l'huile (Fomblin 3900) comme agent cryoprotectant. La collecte de données a montré que ce cristal appartenait au groupe d'espace $P4_12_12$ ($P4_32_12$ pour ERR γ sauvage (Greschik *et al.*, 2002)) avec les paramètres de maille suivants : $a=b=63,94$ Å, $c=136,86$ Å et $\alpha=\beta=\gamma=90^\circ$. Les données de diffraction à une résolution de 1,4 Å ont été collectées sur la ligne ID14-EH1. A ce jour, 12 structures de ERR γ sauvage ont été résolues à une résolution allant de 1,6 à 3,2 Å, dont 4 dans des conformations constitutivement actives (code PDB : 1KV6, 1TFC, 2GP0 et 2GP7), 6 en présence de ligands antagonistes tels que le diéthylstilbestrol (1S9P) ou le tamoxifène (1VBJ, 1S9Q, 2EWP, 2GPU et 2GPV) et 2 en présence de ligands synthétiques, le GSK4716 (2GPP) de poids moléculaire de 282,32 Da et le bisphénolA (2E2R) de poids moléculaire de 228,28 Da. Les statistiques relatives à ces données sont résumées dans le Tableau 14 ci-dessous.

E1a	
Collecte de données	ID14-EH1
Groupe d'espace	P4 ₁ 2 ₁ 2
Nombre de réflexions	56 796
Paramètres de maille	a=b=63,94 Å c=136,86 Å $\alpha=\beta=\gamma=90^\circ$
Résolution	57,9-1,4
R _{sym} * (%)	6 (34)
Complétude (%)	97,5 (98)
Multiplicité	20,4 (16,5)

Tableau 14 | Statistiques de collecte des données du cristal du mutant E1a.
*haute résolution : 1,4-1,44. Les valeurs entre parenthèses correspondent à la dernière tranche de résolution. $R_{sym} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$ où I est l'intensité mesurée pour chaque réflexion et $\langle I \rangle$ l'intensité moyenne de réflexions équivalentes.

8.3.5.2. Résolution de la structure et affinement

La structure a été résolue par remplacement moléculaire (AMoRe) en utilisant la structure cristallographique de l'homodimère de hERR γ sauvage (code PDB : 1KV6) complexé à un peptide contenant le second motif LXXLL du domaine d'interaction de SRC-1 (686-RHKILHRLLEQEGSPS-700) comme modèle (Greschik *et al.*, 2002).

Ce modèle initial a été affiné à l'aide du programme REFMAC5 (Murshudov *et al.*, 1997) qui a ensuite été utilisé comme source de phases initiales pour un processus cyclique automatique d'affinements/reconstruction d'un modèle à l'aide du programme ARP/wARP (Perrakis *et al.*, 2001; Perrakis *et al.*, 1999) en utilisant toutes les données jusqu'à 1,4 Å. La carte finale, de type $2mF_{obs}-DF_{calc}$ s'est révélée d'une très bonne qualité. Un processus cyclique de constructions manuelles à l'aide du programme O (Jones *et al.*, 1991) et d'affinement cristallographique classique à l'aide du programme REFMAC5 a permis d'aboutir au modèle actuel correspondant à un facteur cristallographique de 0,2 (R_{libre} 0,224) avec une très bonne géométrie (Tableau 15).

hERR γ LBD S379R	
Affinement	
Résolution	19,3-1,4
Nombre de réflexions	52515
Complétude	97,5%
R _{factor} * (%)	20% (24)
R _{free} (%)	22,4 (27)
Facteurs B moyens (Å ²)	20,8
RMS bond lengths (Å)	0,007
RMS bond angles (°)	1,132

Tableau 15 | Statistiques d'affinement. *haute résolution : 1,4-1,44.

Ce modèle contient :

- le LBD du mutant E1a des résidus 229 à 454
- le peptide contenant le second motif LXXLL du domaine d'interaction de SRC-1 (686-RHKILHRLQLQEGSPS-700)
- 483 molécules d'eau

Il est à noter que tous les résidus se trouvent dans une conformation correcte comme l'atteste le diagramme de Ramachandran présenté Figure 66.

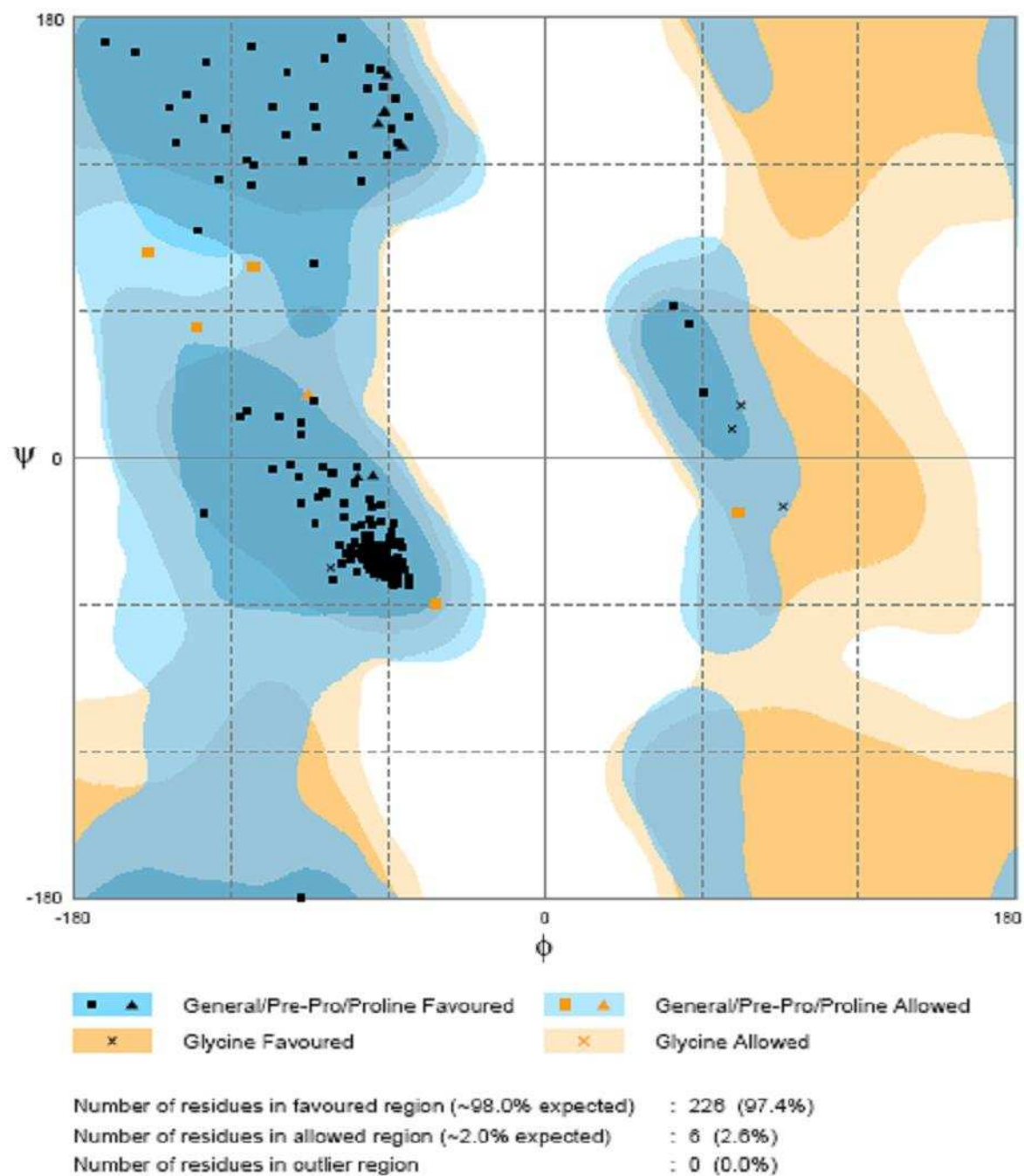


Figure 66 | Diagramme de Ramachandran du mutant E1a. Les résidus en conformation favorable sont représentés en noir, ceux en conformation permise en orange.

8.3.6. Etude de la structure tridimensionnelle du mutant E1a

8.3.6.1. Structure générale

Pour rappel, il s'agissait, par cette mutation, de recréer le pont salin spécifique de la classe II (hétérodimères avec RXR) faisant intervenir l'arginine différentiellement conservée de la boucle H8-H9 avec l'acide glutamique (E307) situé au niveau des hélices H4-5 (Figure 67), sans perturber la voie de communication spécifique de la classe I.

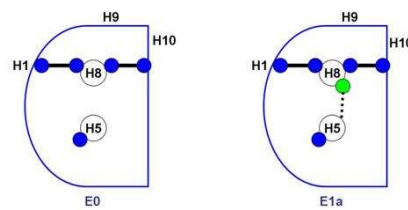


Figure 67 | Caractéristique du mutant E1a.

Nous présentons ici la structure de l'homodimère cristallographique du mutant E1a (hERR γ LBD S379R) des résidus 229 à 454 (Figure 68). Ce mutant adopte la structure canonique en sandwich d'hélices α et présente une conformation identique à celle de hERR γ sauvage, c'est-à-dire une conformation transcriptionnellement active typique des RNs liés à un agoniste.

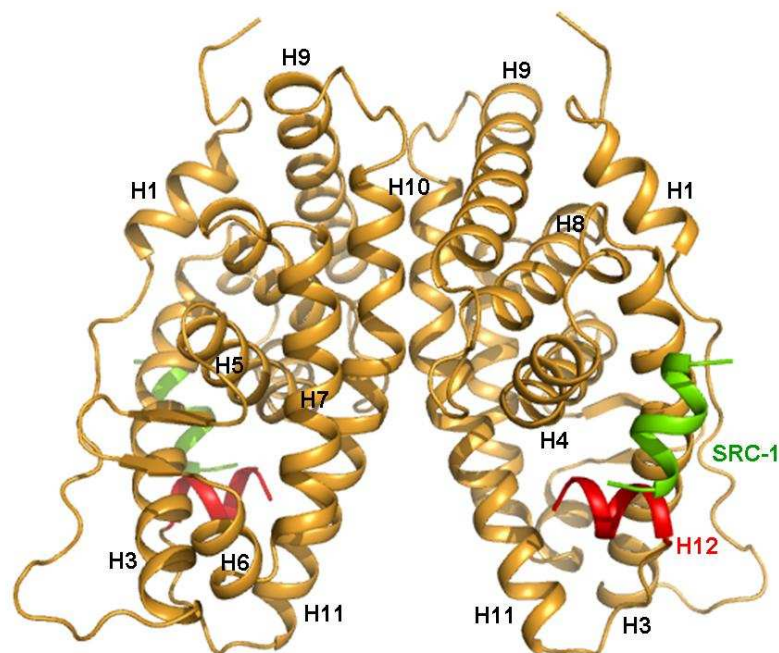


Figure 68 | Structure de l'homodimère cristallographique du mutant E1a (hERR γ LBD S379R). L'hélice H12 est colorée en rouge et le peptide SRC-1 en vert.

8.3.6.2. Interface de dimérisation

La superposition d'un monomère de hERR γ sauvage et du mutant E1a montre une similarité importante, illustrée par un *rmsd* (*root mean square déviation* ou déviation moyenne) faible : 0,52 sur 225 atomes de Ca. On peut noter un léger changement de conformation de chaque monomère qui peut être mis en évidence par la superposition sur un monomère, du mutant E1a sur hERR γ sauvage, des deux homodimères. Le deuxième monomère du mutant E1a présente alors une rotation de 6° et une translation de 0,07 Å par rapport au deuxième monomère de hERR γ sauvage, comme l'atteste la Figure 69.

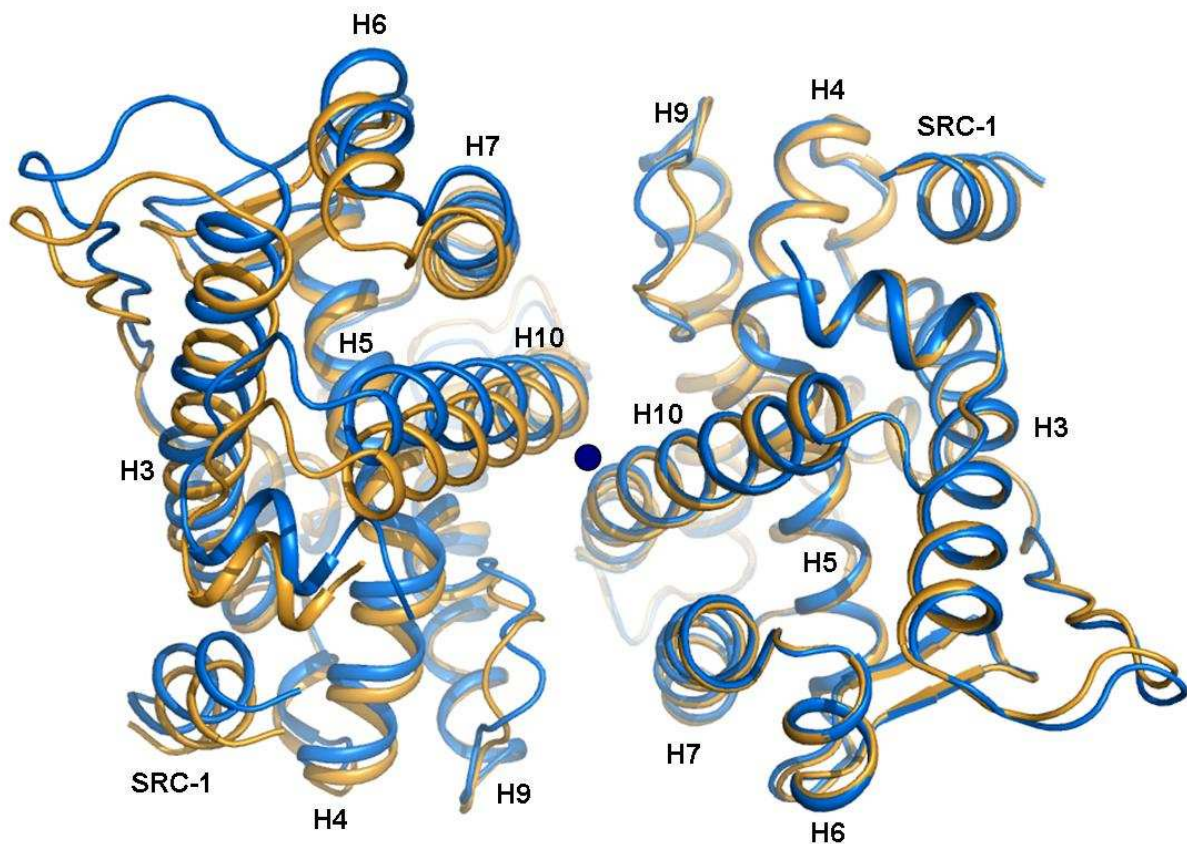


Figure 69 | Superposition des structures cristallographiques des homodimères hERR γ sauvage (bleu) et du mutant E1a (orange) vue selon l'axe de symétrie d'ordre 2 (point bleu).

8.3.6.3. Conformation de l'arginine 379

La Figure 70 présente le pont salin (pointillés rouges), caractéristique de la classe II, existant entre les résidus D267 (H4-5) et R339 (L8-9) de hRAR α .

La différence majeure entre les deux structures hERR γ sauvage et E1a réside essentiellement au niveau de la boucle H8-H9 et plus précisément au niveau de la mutation de la sérine 379 en arginine. Il apparaît que cette mutation engendre une augmentation locale des facteurs B en apportant une flexibilité supplémentaire à la boucle H8-H9 (colorée en orange) visible sur la Figure 70. En effet, le facteur d'agitation moyen pour l'ensemble de la structure est de l'ordre de 20 Å². Or, dans cette boucle, des résidus 378 à 386, le facteur d'agitation augmente jusqu'à 40 Å² avec un maximum à 57 Å² au niveau de la mutation. Cette flexibilité n'est apparemment pas suffisante pour permettre à l'arginine S379R du mutant E1a d'établir le pont salin équivalent de la classe II avec l'acide glutamique E307 localisé au niveau des hélices H4-H5.

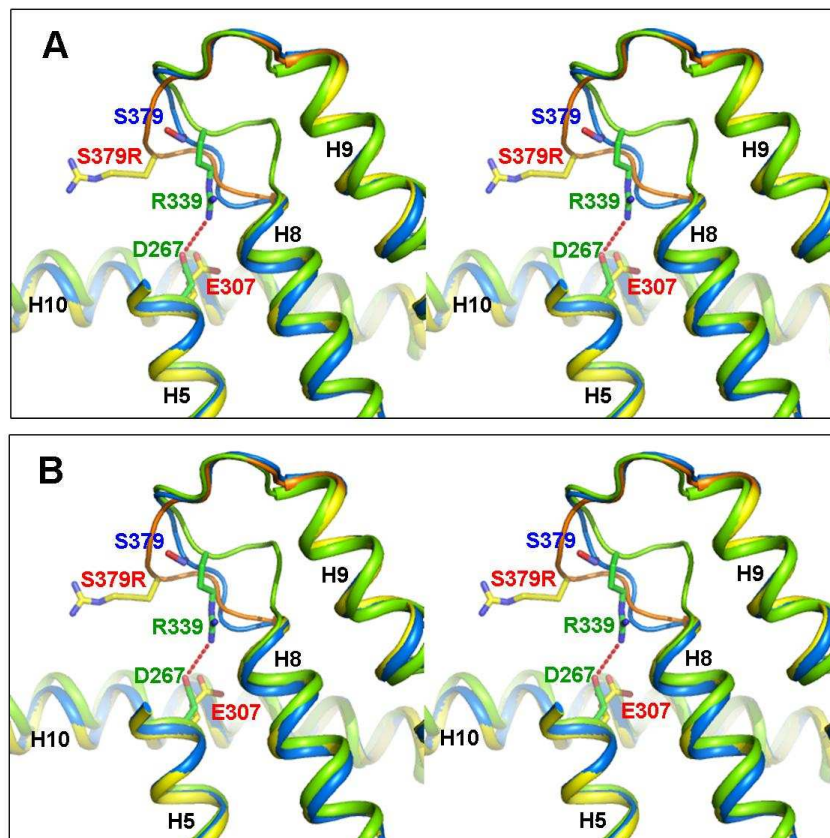


Figure 70 | Stéréovue de la conformation de la boucle H8-H9 de hRAR α (vert), hERR γ (bleu) et E1a (jaune). Le pont salin entre les résidus D267 et R339 de hRAR α caractéristique de la classe II est indiqué par les pointillés rouges et la boucle H8-H9 est coloriée en orange (A), stéréovue selon la technique dite parallèle, (B), stéréovue selon la technique dite croisée.

En raison des facteurs B importants au niveau de cette région, il est évident que cette mutation déstabilise la boucle H8-H9. Toutefois, l'analyse de la structure n'a pas montré d'interaction entre cette S379R avec un résidu de l'interface de dimérisation ou un autre acide aminé du même monomère.

8.3.6.4. La poche de fixation du ligand

La LBP de hERR γ et celle du mutant E1a sont identiques. Elle est formée de douze résidus (Figure 71), principalement hydrophobes (L268, A272, L309, I310, V313, L345, H434, F435, F450, L454) et chargés (E275, R316) et son volume (220Å³) est le plus petit observé chez les RNs. Les deux résidus chargés sont conservés à travers tous les membres de la famille ERR et ER.

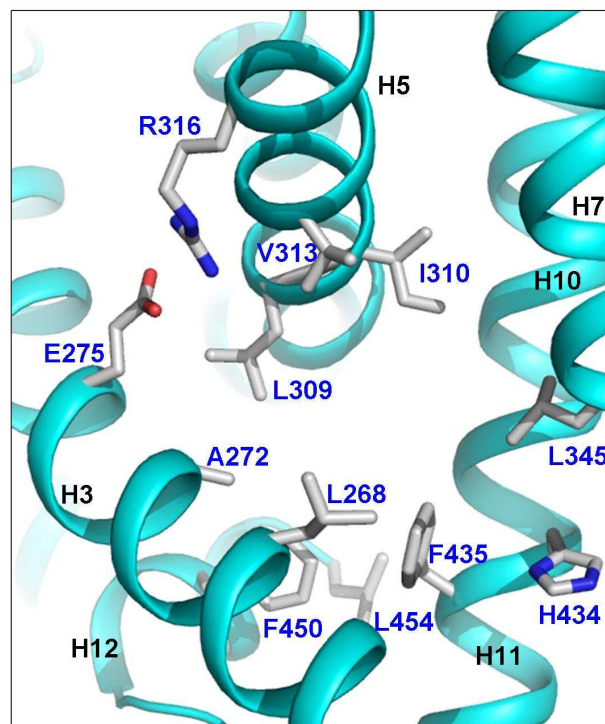


Figure 71 | La poche de fixation du ligand du mutant E1a.

Le LBD de hERR γ a été la première structure cristallographique dans laquelle une conformation agoniste du LBD en présence d'un peptide CoA était adoptée en absence de ligand. Cependant, dans la LBP du mutant E1a se trouve un ligand comme l'atteste la présence significative de densité électronique (Figure 72). De plus, l'existence d'un ligand, de 182 Da, est par ailleurs confirmée par les résultats obtenus par la spectrométrie de masse.

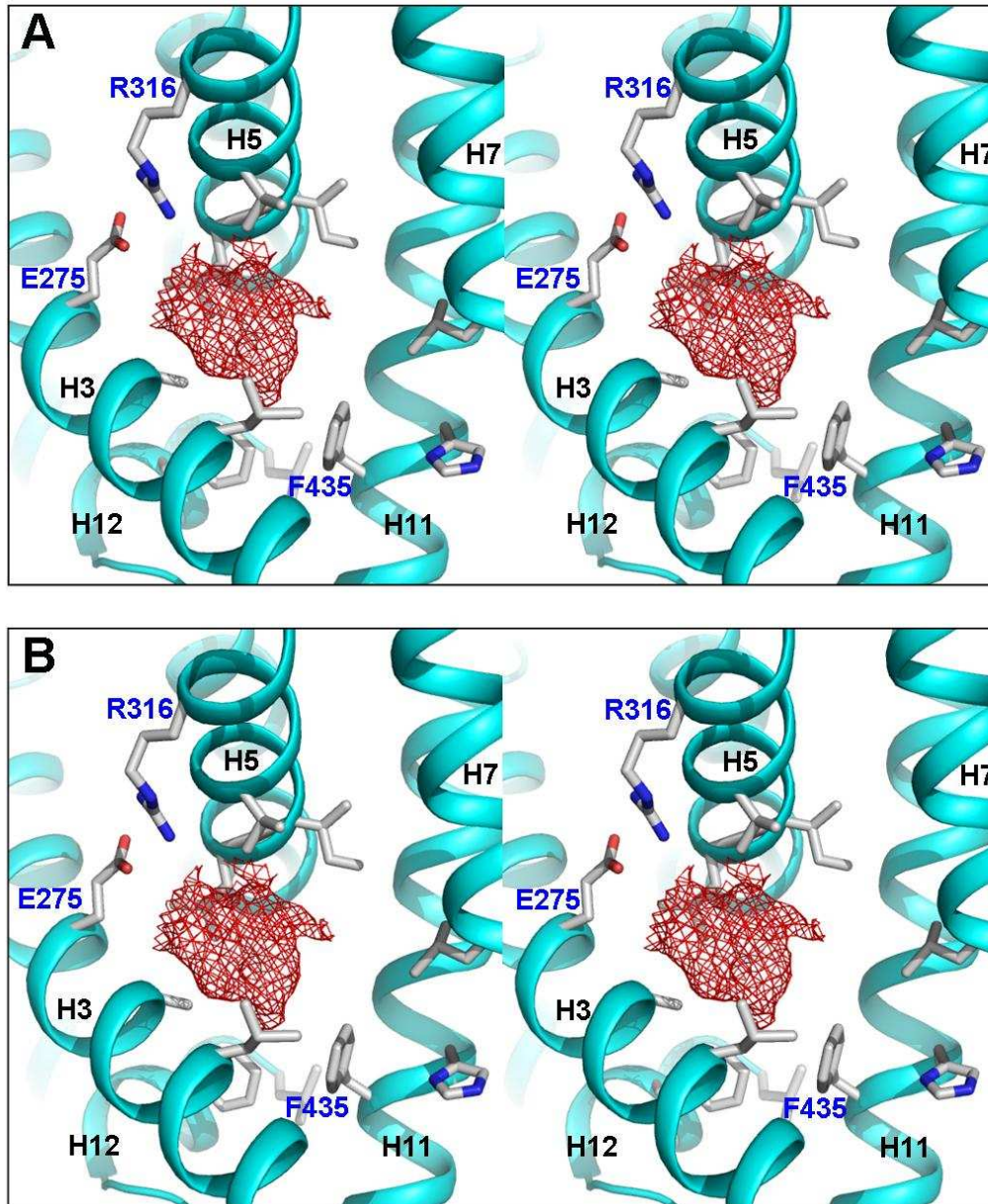


Figure 72 | Stéréovue de la densité électronique observée dans la LBP du mutant E1a. La densité électronique est représentée en rouge. (A), stéréovue selon la technique dite parallèle, (B), stéréovue selon la technique dite croisée.

Les RNs ERR γ et ER α présentent 36,6% d'identité. La résolution de la structure cristallographique de ER α complexé à l'œstradiol (E2) (code PDB : 1QKT) (Gangloff *et al.*, 2001) a montré que deux résidus chargés de la LBP de ER α , E353 et R394, établissaient des liaisons hydrogènes avec le groupe 3-OH de l'anneau A de E2, tandis que le résidu H524 connectait le groupe 17 β -OH de l'anneau D.

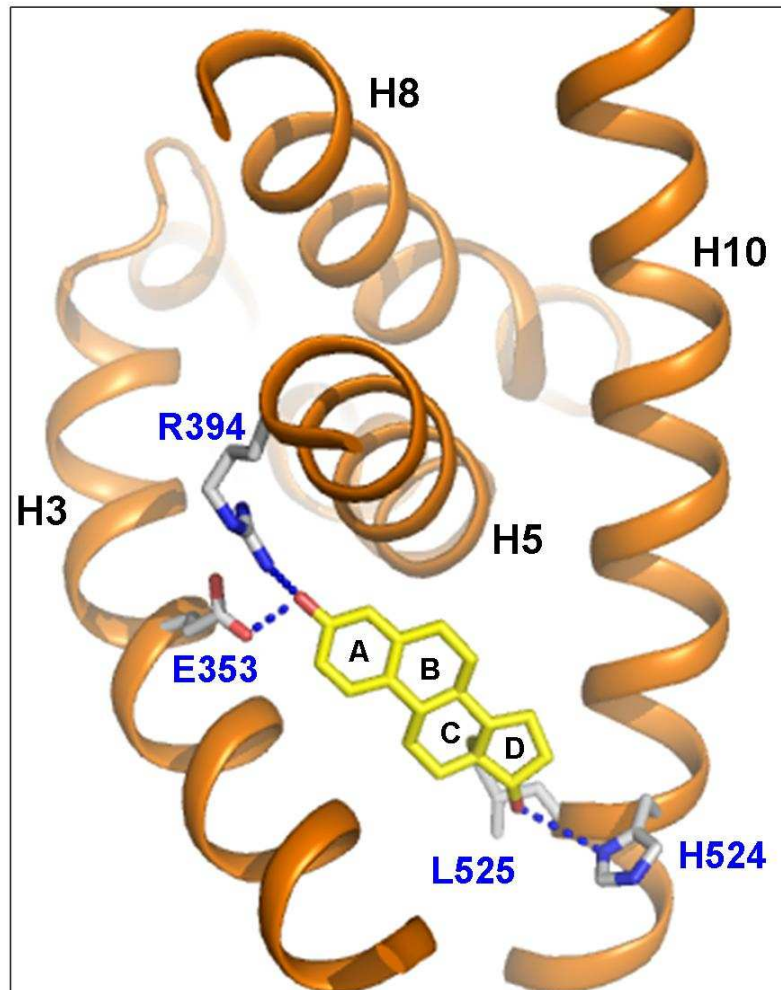


Figure 73 | Structure cristallographique du complexe ER α /œstradiol (E2) (code PDB : 1QKT) (Gangloff *et al.*, 2001). Le récepteur ER α est représenté en orange, le ligand E2 en jaune.

Les deux résidus connectant le groupe 3-OH de E2 sont également présents et dans la même conformation aussi bien dans le récepteur ERR γ sauvage que dans le mutant E1a (E275 et R316). Une différence essentielle concerne la L525 ER α , où le résidu équivalent correspond à la F435 chez ERR γ et E1a. Dans le but de déterminer de possibles agonistes pour ERR γ , Greschick *et coll.* (Greschik *et al.*, 2002) ont superposé le complexe ER α /E2 avec ERR γ et modélisé E2 dans la LBP. Cette modélisation a montré que l'anneau A de E2 se plaçait correctement dans la LBP de ERR γ . Cependant, une gêne stérique entre l'anneau D de E2 et le résidu F435 (L525 pour ERR γ sauvage) excluait la fixation de E2 dans la LBP de ERR γ . Ces observations sont également vérifiées pour le mutant E1a (Figure 74).

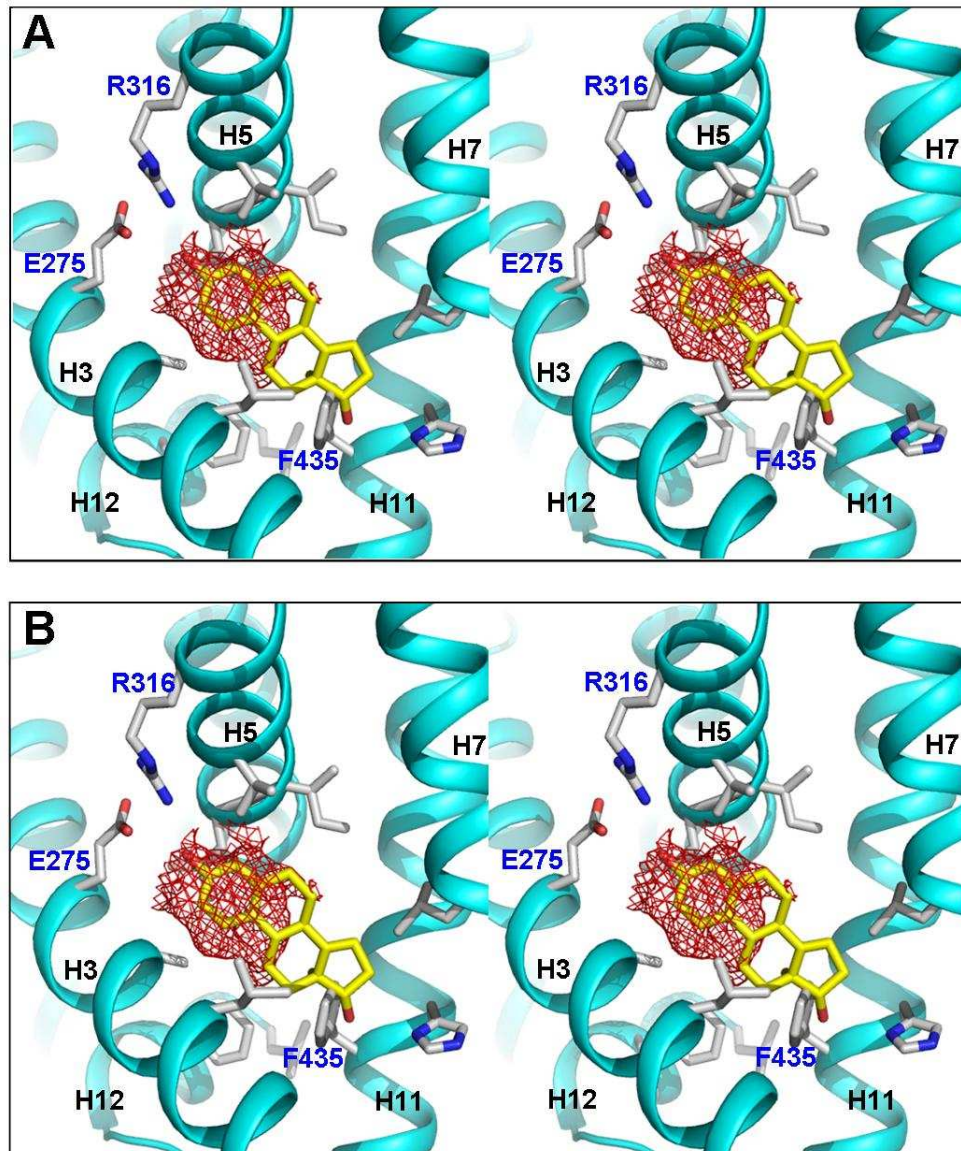


Figure 74 | Stéréovue de la superposition de l'oestradiol E2 de hER α dans la LBP du mutant E1a. La densité électronique observée dans la LBP du mutant est représentée en rouge, le ligand E2 en jaune.(A), stéréovue selon la technique dite parallèle, (B), stéréovue selon la technique dite croisée.

En raison de la contrainte imposée par ces résidus, il avait alors été suggéré que la taille d'un hypothétique agoniste de ERR γ devrait mesurer moins de la moitié de la taille de E2, c'est-à-dire ne comprendre que les anneaux A et B. La superposition du mutant E1a avec le complexe ER α /E2 confirme cette hypothèse comme l'illustre la Figure 74. En effet, la densité électronique correspondant au ligand de 182 Da, observée dans la LBP du mutant E1a, englobe les anneaux A et B de E2 (Figure 74).

A ce stade, nous disposons de peu d'information concernant la structure du ligand présent dans la LBP du mutant E1a. Les recherches devront s'orienter dans la direction d'un composé chimique basé sur deux cycles et d'un poids proche de 182 Da. Une recherche rapide a permis de définir quelques exemples de molécules chimiques pouvant servir de pistes (Figure 75).

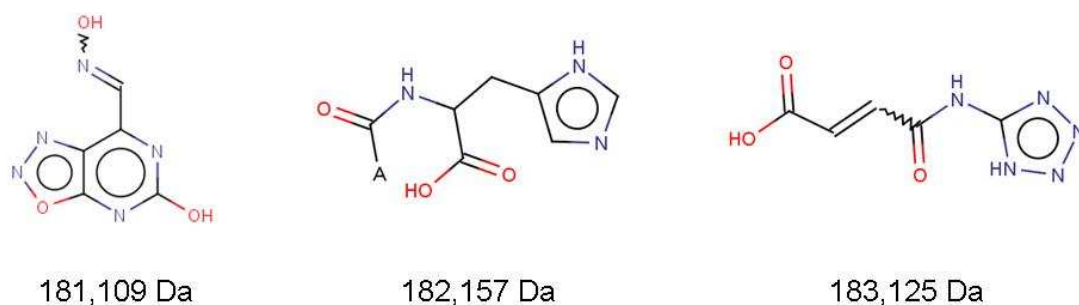


Figure 75 | Exemples de ligands (<http://cdb.ics.uci.edu/>).

8.3.7. Bilan

A ce jour, nous n'avons pu obtenir la structure cristallographique que d'un seul mutant sur les cinq testés (Figure 76).

	Expression	Purification	Cristallisation	Cristaux	Structure
E1a	✓	✓	✓	✓	✓
E2a	✓	✓	∅		
E3a	✓	∅			
E4	✓	∅			
E5	✓	∅			

Figure 76 | Bilan de caractérisation des différents mutants de hERR γ .

Une question reste en suspend. En effet, si notre hypothèse est fondée, le quintuple mutant E5, qui correspond à un récepteur « équivalent » de la classe II (un récepteur qui hétérodimérise avec RXR), devrait lui aussi pouvoir interagir avec RXR. Des études ultérieures sont déjà prévues pour tester cette hypothèse.

Chapitre 9 - Etude du système hRAR α

9.1. Etude structurale

Au moment d'initier l'étude structurale du système hRAR α , nous ne disposions que de mutants simples, notamment le mutant R1d qui correspond à l'unique résidu différenciellement conservé de la classe II, à savoir l'arginine située dans la boucle H8-9 (Figure 77). L'étude précédente du système hERR γ nous a montré que la mutation d'un résidu impliqué dans la connexion de l'hélice H8 vers la boucle H8-H9 engendrait des problèmes au niveau de la solubilité du mutant. Ainsi, dans le cadre de cette étude, nous nous sommes focalisé sur le mutant R1d, dont l'arginine n'est pas directement impliquée dans l'interface de dimérisation.

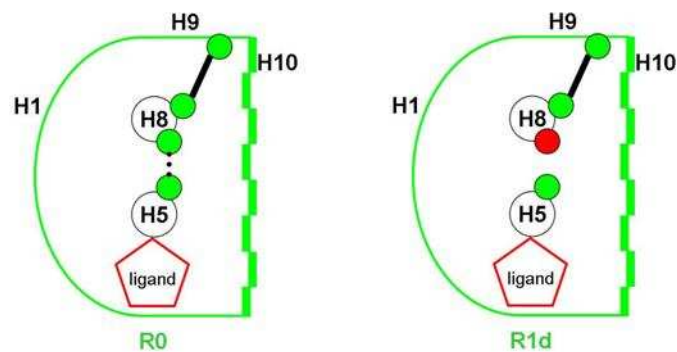


Figure 77 | Comparaison de R0 et du mutant R1d.

Il s'agissait, par cette mutation, d'étudier le rôle de l'arginine différenciellement conservée impliquée dans le pont salin caractéristique de la classe II sur le comportement général de ce mutant et l'effet sur l'hétérodimérisation avec X0.

9.1.1. Expression de X0 et du mutant R1d

Le mutant R1d a été cloné dans un vecteur d'expression pHGWA (*amp^R*) où le gène codant le mutant se trouve en fusion avec une étiquette polyhistidine en extrémité N-ter, alors que X0, cloné dans un vecteur d'expression p0GWA (*amp^R*), sera produit sous sa forme non fusionnée avec une étiquette polyhistidine. La masse de cellules obtenue, après récolte,

est d'environ 4,5 g/l de culture pour le mutant R1d et 3 g/l de culture pour X0. Les conditions d'expression dans *E. coli* sont détaillées dans la section 3.2.1.

Différentes combinaisons de purifications, illustrées dans le Tableau 16, ont été testées. La purification, en deux étapes, de l'hétérodimère R1d-X0, surlignée en jaune, est décrite ci-dessous.

	Expression	Purification				Cristallisation	Cristaux
		Lyse cellulaire	Chromatographie d'affinité	Clivage Thrombine	Filtration sur gel		
		- Lig + Lig					
R1d-pHGWA	✓	✓	✓		✓	∅	
R1d-pHGWA	✓	✓	✓		✓	∅	
Co-purification avec X0-p0GWA							
R1d-pHGWA	✓	✓	✓		✓	∅	
R1d-pHGWA	✓	✓	✓		✓	∅	

Tableau 16 | Tableau schématique présentant les différentes combinaisons de purification du mutant R1d.

La purification d'un hétérodimère pose le problème de la stoechiométrie entre les deux sous unités. La stratégie de purification est basée sur les connaissances acquises dans notre laboratoire concernant ce sujet ainsi que sur les travaux de Bourguet et *coll.* (Bourguet *et al.*, 2000).

9.1.2. Lyse cellulaire et ultracentrifugation

La procédure consiste à lyser ensemble, par sonication, les cellules des culots provenant de 3 l de culture du mutant R1d et de 2 l de X0. Les culots sont re-suspendus dans un volume adéquat de tampon puis lysés (voir 3.2.2 et 3.2.3). Cette opération est effectuée à l'obscurité car une solution éthanolique saturée d'acide rétinoïque 9-*cis*RA est ajoutée dans l'extrait brut. Ce dernier, issu de la sonication, est ensuite clarifié par ultracentrifugation (2 h à 45 000 g à 4°C).

9.1.3. Purification du complexe R1d-X0

Les deux sous-unités X0 et R1d, produites séparément dans *E. coli*, sont co-purifiées. Ceci présente l'avantage de favoriser la formation de l'espèce stable que constitue l'hétérodimère surtout en présence de 9-*cis*RA. D'autre part le mutant R1d étant limitant (moins soluble sous forme monomérique), il nous est facile de travailler en excès de X0 et de se débarrasser de cet excès à la première étape de purification. X0, non complexé, ne sera pas retenu sur la colonne d'affinité interagissant avec le polypeptide hexahistidine uniquement présent sur le mutant R1d. L'adjonction de ligand se fait dès la première étape de préparation afin d'obtenir une saturation effective de l'hétérodimère en 9-*cis*RA.

La purification de ce mutant s'effectue en deux étapes : une chromatographie d'affinité au nickel suivie d'une filtration sur gel (Figure 78).

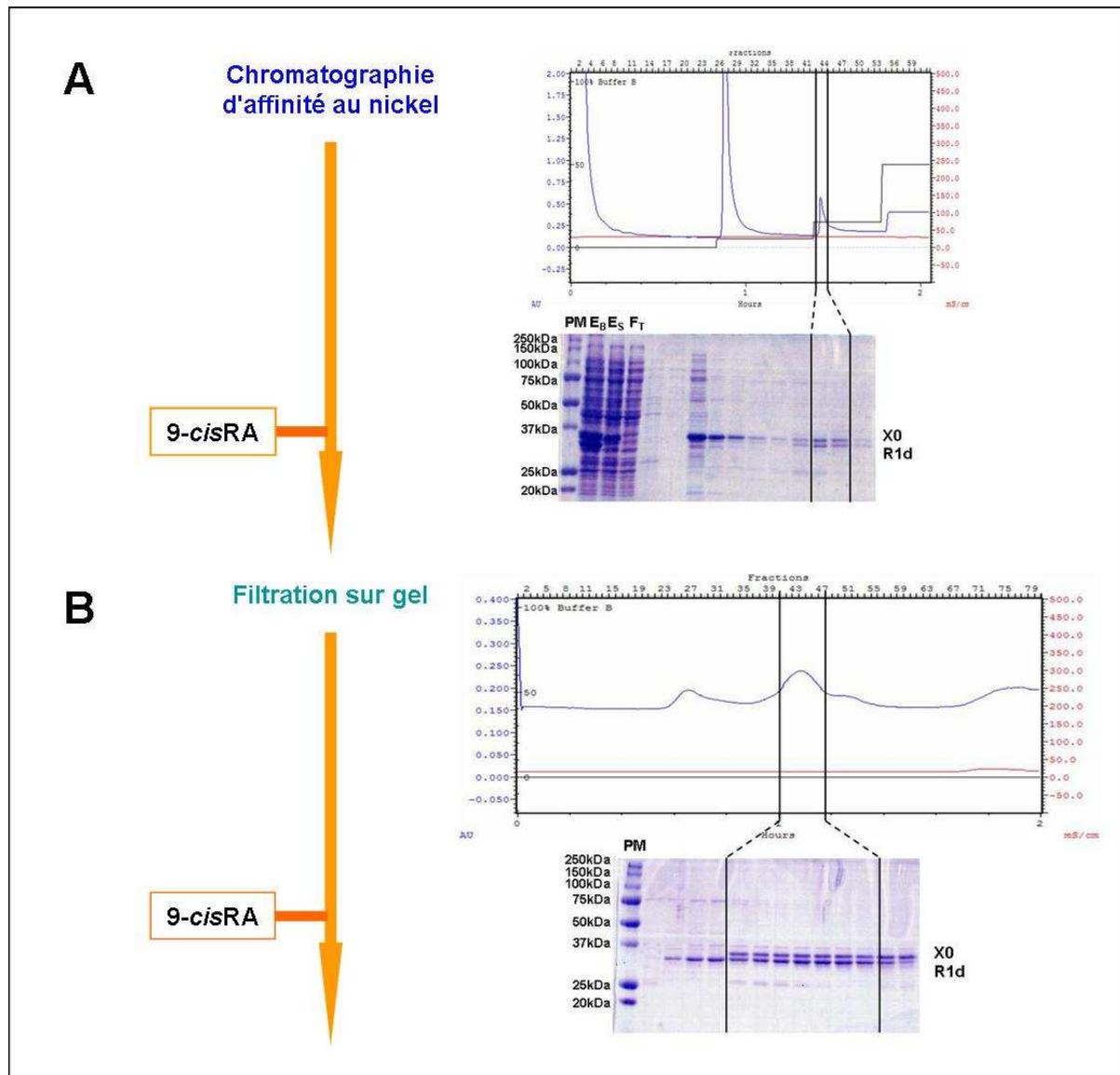


Figure 78 | Schéma de purification de l'hétérodimère R1d-X0 en présence de 9-cisRA. Chromatogramme : le tracé bleu correspond à la DO_{280nm}, le rouge à la conductivité en mS/cm et le gris au gradient d'Imidazole. Gel SDS-PAGE 12% : PM, marqueurs de poids moléculaire ; E_B, Extrait Brut après sonication ; E_S, Extrait Soluble, F_T, Flow Through.

9.1.3.1. Chromatographie d'affinité sur résine chélatant les ions nickel

Le principe et les conditions expérimentales sont décrits dans la section 3.2.4. Le lysat clarifié est chargé sur une colonne d'affinité par chélation de cations métalliques Ni²⁺ (Figure 78A). A la suite de cette chromatographie, nous observons deux pics : le premier correspond au lavage permettant d'éliminer les protéines fixées de manière non spécifique ainsi que l'excès de X0 non retenu sur la colonne, le second correspond à l'hétérodimère R1d-X0. Nous pouvons également noter, d'après le gel SDS-PAGE 12%, la migration atypique de X0. En effet, le poids moléculaire calculé de X0 est de 29 407 Da, alors que celui du mutant R1d est

de 35 161 Da. En conditions dénaturantes, X0 aurait du migrer plus loin et se situer sous le mutant R1d. Cette migration atypique de X0 est connu et a été confirmée par la spectrométrie de masse MALDI-TOF, réalisée par Manuela Argentini (plate-forme de Protéomique). Les fractions correspondantes à l'hétérodimère sont rassemblées. Une solution éthanolique saturée de 9-*cis*RA est ajoutée au complexe et incubée 2 h à 4°C à l'obscurité.

9.1.3.2. Chromatographie de filtration sur gel

La filtration sur gel permet de séparer les protéines en fonction de leur masse moléculaire apparente, d'éliminer toute trace éventuelle d'agrégats ou d'oligomères et de conditionner l'hétérodimère R1d-X0 en vue des tests de cristallisation. L'échantillon est concentré, puis injecté sur la colonne de filtration sur gel (Figure 78B). Nous pouvons observer deux pics aplatis : le premier correspond à des agrégats du mutant R1d et le deuxième à l'hétérodimère R1d-X0. Les fractions sont analysées sur gel SDS-PAGE 12% et rassemblées. A nouveau, une solution éthanolique saturée de 9-*cis*RA est ajoutée à l'hétérodimère R1d-X0. La stœchiométrie définitive est apparemment idéale.

9.1.3.3. Rendement de purification

Le rendement de purification est de l'ordre de 3,4 mg de l'hétérodimère R1d-X0 (Tableau 17).

R1d : hRARα R339A	Volume (ml)	Concentration (mg/ml)	Quantité (mg)	Rendement (%)
Chromatographie d'affinité au nickel	10	0,64	6,4	
Concentration	3	1,3	3,9	61
Filtration sur gel	10,5	0,3	3,4	87

Tableau 17 | Rendement de purification du complexe R1d-X0.

9.1.4. Cristallisation de l'hétérodimère R1d-X0

Une solution éthanolique saturée de 9-*cis*RA est ajoutée à l'hétérodimère R1d-X0 et laissée incuber durant la nuit à 4°C à l'obscurité. Le complexe R1d-X0 est ensuite concentré à 4 mg/ml, puis cristallisé selon la technique de diffusion de vapeur en goutte assise (0,7 μ l de la protéine + 0,7 μ l de solution de cristallisation contre un réservoir de 75 μ l) à 24°C en utilisant les kits : *The Sparse Matrix 1*, *The Classics*, *The JCSG+*, *The PEG-Ion pH* et *NR-LBD 96*. A ce jour, nous n'avons encore obtenu aucun cristal que ce soit pour la purification décrite ci-dessus que pour les autres présentées Tableau 16.

Le principal paramètre différenciant l'hétérodimère R1d-X0 de R0-X0, mise à part la mutation, réside dans le type de ligand. En effet, dans le cas de l'hétérodimère R0-X0, il s'agit d'un ligand antagoniste BMS614. Peut être que l'utilisation de ce ligand nous permettra, dans une étude ultérieure, d'obtenir des cristaux. D'autres paramètres pourraient avoir une influence dans l'obtention de cristaux : augmenter la concentration, l'ajout d'un peptide CoA, éliminer l'étiquette polyhistidine, etc.

9.2. Etude de l'activité transcriptionnelle et de la dimérisation

Pour rappel, les résultats fonctionnels obtenus suite à notre étude préliminaire des résultats issus de l'analyse bioinformatique des LBDs de la super-famille des RNs (Breliet *et al.*, 2004) sont présentés ci-dessous (Figure 79). Les deux résidus, l'arginine R339 et R367 de hRAR α , impliqués dans les deux ponts salins caractéristiques de la classe II ont été mutés en alanine. L'effet de ces deux mutations, R367A (mutant R1b) et R339A (mutant R1d), a été examiné au niveau de l'hétérodimérisation avec hRXR α par des expériences de *pull down*, ainsi que par leur capacité à activer la transcription.

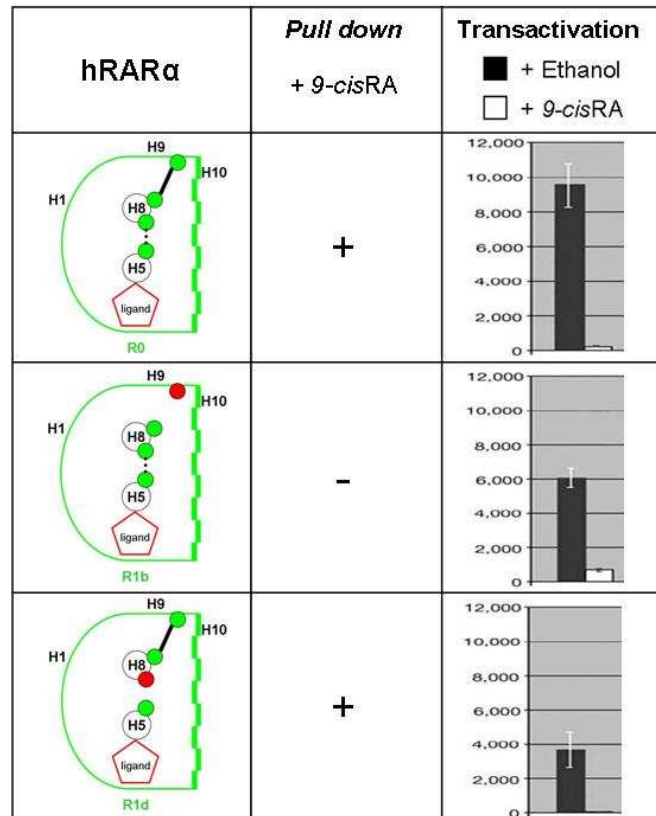


Figure 79 | Bilan des expériences de *pull down* et de transactivation pour les mutants R1b et R1d.

Ces résultats nous montrent que le mutant R1b a un impact évident sur la stabilité du dimère contrairement à R1d. Alors que le mutant R1b présente un faible effet sur l'activité transcriptionnelle, celle du mutant R1d est réduite de plus de moitié. Il avait été alors suggéré que la stabilité de l'hétérodimère n'était pas le rôle principal de l'arginine différenciellement conservée de la boucle H8-H9.

9.3. Etude indirecte de la voie par phosphorylation

Résumé de l'article

L'importance de la voie de communication de la classe II a été étudiée au niveau du système RAR α par l'étude du rôle de la phosphorylation de résidus impliqués dans les régions spécifiques de la classe II. Ce projet visait à déchiffrer les bases structurales de la communication entre les RNs et les composants de la machinerie transcriptionnelle afin de comprendre comment leurs fonctions étaient modulées par la phosphorylation.

Les RNs communiquent avec les différentes composantes de la machinerie de transcription comme l'ARN polymérase II, TFIIB, TBP ou TFIIH pour moduler l'expression de gènes spécifiques (Aranda and Pascual, 2001; Bastien and Rochette-Egly, 2004). Cette communication moléculaire peut être accomplie soit par des interactions protéine-protéine directes entre les RNs et la machinerie de transcription basale, soit par des interactions indirectes médiées par des corégulateurs. La transcription d'un gène cible par les RNs n'est pas seulement régulée par le ligand mais aussi au travers de réseaux coordonnés faisant intervenir divers événements de modifications post-traductionnelles comme la phosphorylation (Bastien and Rochette-Egly, 2004). Dans la plupart des cas, les phosphorylations participent à la dynamique et au recrutement séquentiel de CoA, de complexes remodelant la chromatine et/ou de composants de la machinerie de transcription (Rochette-Egly, 2003).

Les RNs sont principalement phosphorylés dans le domaine A/B, mais aussi dans le LBD par plusieurs kinases. Par exemple, RAR α est phosphorylé par la PKA (*Protein Kinase A*) (Rochette-Egly *et al.*, 1995) au niveau de la boucle H9-H10 sur la sérine 369. Il est intéressant de noter que ce site de phosphorylation se situe deux résidus après l'arginine 367 impliquée dans le pont salin caractéristique de la classe II, connectant l'hélice H8 à l'hélice H9 et qui, en concert avec le deuxième pont salin connectant la boucle H8-H9 aux hélices H4-5, maintient la conformation du LBD. Les données expérimentales ont dans un premier temps, montré que la cycline H (composante de TFIIH) se fixe au niveau de la boucle H8-H9 de RAR α . Dans un deuxième temps, la phosphorylation de la sérine 369 engendre une augmentation de l'interaction de la cycline H avec RAR α .

Ce résultat majeur obtenu sur un récepteur sauvage permet de relier la voie de communication avec les mécanismes de régulation de l'activité transcriptionnelle (Gaillard *et al.*, 2006) (Figure 80). En effet, la cycline H appartient au complexe TFIIH qui est directement impliquée dans la régulation de l'activité transcriptionnelle de l'ARN polymérase II.

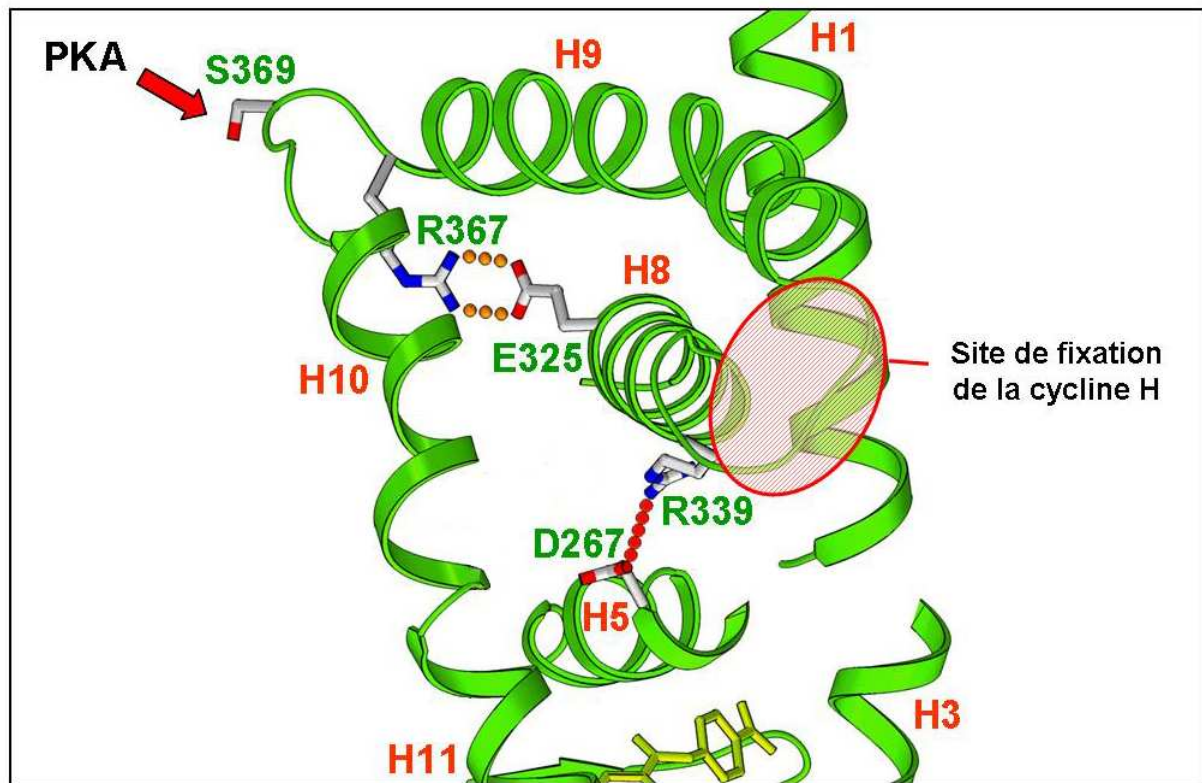


Figure 80 | Site de phosphorylation (S369) de la PKA chez hRAR α et site de fixation de la cycline H.

Publication N°3

Phosphorylation by PKA potentiates retinoic acid receptor α activity by means of increasing interaction with and phosphorylation by cyclin H/cdk7

Emilie Gaillard, Nathalie Bruck, Yann Brelivet, Gaétan Bour, Sébastien Lalevée, Annie Bauer, Olivier Poch, Dino Moras, Cécile Rochette-Egly

Proceedings of the National Academy of Sciences of the United States of America, 2006, Vol.103, N° 25, Pages 9548–9553

Pages 233-238 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1073/pnas.0509717103>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>

Chapitre 10 - Etude du système hRXR α

10.1. Les mutants disponibles

Concernant l'étude du système hRXR α , nous disposons d'une nouvelle collection de mutants, de la simple mutation jusqu'au changement de classe avec un quintuple mutant, ainsi qu'avec différents mutants intermédiaires (Figure 81).

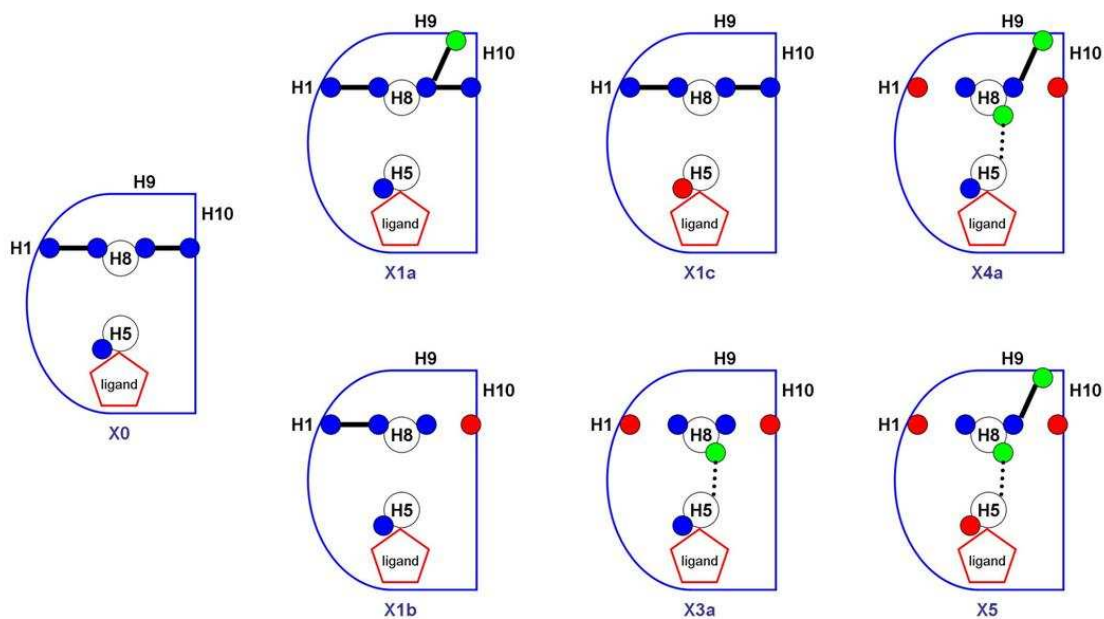


Figure 81 | Les différents mutants hRXR α . Les points bleus correspondent aux résidus spécifiques de la classe I de hRXR α , les points verts représentent la mutation d'un résidu de hRXR α en acide aminé spécifique de la classe II et les points rouges correspondent à la mutation d'un résidu de hRXR α en alanine.

Dans le cas du mutant X1a, il s'agissait d'essayer de recréer le deuxième pont salin spécifique de la classe II (hétérodimères avec RXR) en mutant une tyrosine, située dans H9, en arginine. Le mutant X1b est défini par la perte du deuxième pont salin caractéristique de la classe I connectant H8 à H10. Le tryptophane pointant dans la poche de fixation du ligand et caractéristique de la classe I a été muté en alanine dans le mutant X1c. De cette série de constructions, X1c est le seul mutant à ne pas avoir de mutation impliquant un des résidus spécifiques de la voie de communication de la classe I. Le triple mutant X3a se caractérise par la perte des deux ponts salins de la classe I connectant l'interface de fixation des cofacteurs à l'interface de dimérisation et la mutation d'une sérine en arginine pour tenter de recréer le

pont salin spécifique de la classe II faisant intervenir l'arginine différenciellement conservée de la boucle H8-H9 avec le résidu acide situé au niveau des hélices 4 et 5. Dans le quadruple mutant X4a, les ponts salins de la classe I ont été remplacés par ceux de la classe II. Le quintuple mutant X5 correspond à un récepteur « équivalent » de la classe II.

10.2. Etude structurale du mutant X1c

L'étude structurale des différents mutants du système hERR γ nous a enseigné qu'une simple mutation d'un résidu de la voie de communication avait un impact majeur sur la solubilité. Seule la structure cristallographique du mutant E1a a pu être résolue, pour lequel la mutation ne concerne aucun résidu impliqué directement dans la voie. Dans le même esprit, nous nous sommes donc, dans un premier temps, focalisés sur l'étude structurale du mutant X1c. Il s'agissait, par cette mutation, d'étudier le rôle du tryptophane différenciellement conservé de la classe I sur la fixation du ligand et l'impact de cette mutation sur le comportement oligomérique de ce récepteur, en sachant que ce tryptophane n'intervenait pas directement dans la voie de communication.

Le problème posé par la purification de X0 réside principalement dans l'existence d'un équilibre complexe, notamment dans le cas du X0 non lié à son ligand. En effet, le X0 non lié se présente sous deux formes interconvertibles caractérisées comme un homodimère et un tétramère. Cependant, la présence du ligand 9-*cis*RA déplace l'équilibre en faveur du dimère. C'est pour cette raison que la purification se fait toujours en présence d'un excès de ligand afin de saturer le récepteur et d'avoir une préparation aussi homogène que possible.

10.2.1. Expression du mutant X1c

Le domaine de fixation de l'hormone du mutant X1c (223-462) a été cloné dans un vecteur d'expression pHGWA (*amp^R*) où le gène codant le mutant se trouve en fusion avec une étiquette polyhistidine en extrémité N-ter. La masse de cellules obtenue, après récolte, est d'environ 5 g/l de culture et l'expression est correcte. Les conditions d'expression dans *E. coli* sont détaillées dans la section 3.2.1.

10.2.2. Lyse cellulaire et ultracentrifugation

La procédure consiste à lyser, principalement par sonication, les cellules des culots provenant de des cultures du mutant X1c. Les culots sont re-suspendus dans un volume adéquat de tampon puis lysés. Cette opération est effectuée à l'obscurité car une solution éthanolique saturée de 9-*cis*RA est ajoutée dans l'extrait brut. Ce dernier, issu de la sonication, est ensuite clarifié par ultracentrifugation (2 h à 45 000 g à 4°C) (voir 3.2.2 et 3.2.3).

10.2.3. Purification du mutant X1c en présence de 9-*cis*RA

Différentes combinaisons de purifications, illustrées dans le Tableau 18, ont été testées. Seule une, surlignée en jaune, nous a permis d'obtenir une forme cristalline de ce mutant et est décrite ci-dessous.

	Expression	Purification				Cristallisation	Cristaux
		Lyse cellulaire	Chromatographie d'affinité	Clivage Thrombine	Filtration sur gel		
		- Lig + Lig					
X1c-pHGWA	✓	✓	✓	✓	✓	✓	
X1c-pHGWA	✓	✓	✓	✓	✓	∅	
Co-purification avec R0-pHGWA							
X1c-p0GWA	✓	✓	✓		✓	∅	
X1c-p0GWA	✓	✓	✓		✓	∅	
X1c-p0GWA	✓	✓	✓	✓	✓	∅	

Tableau 18 | Tableau schématique présentant les différentes combinaisons de purification jusqu'à l'obtention de cristaux du mutant X1c.

Les conditions expérimentales de purification sont décrites dans la section 3.2.4. Le lysat cellulaire clarifié est purifié selon un protocole en deux étapes : une chromatographie d'affinité au cobalt suivie d'une filtration sur gel (Figure 82). Le choix de ce protocole est basé sur les enseignements de notre laboratoire et d'après les travaux d'Egea et coll. (Egea *et al.*, 2000).

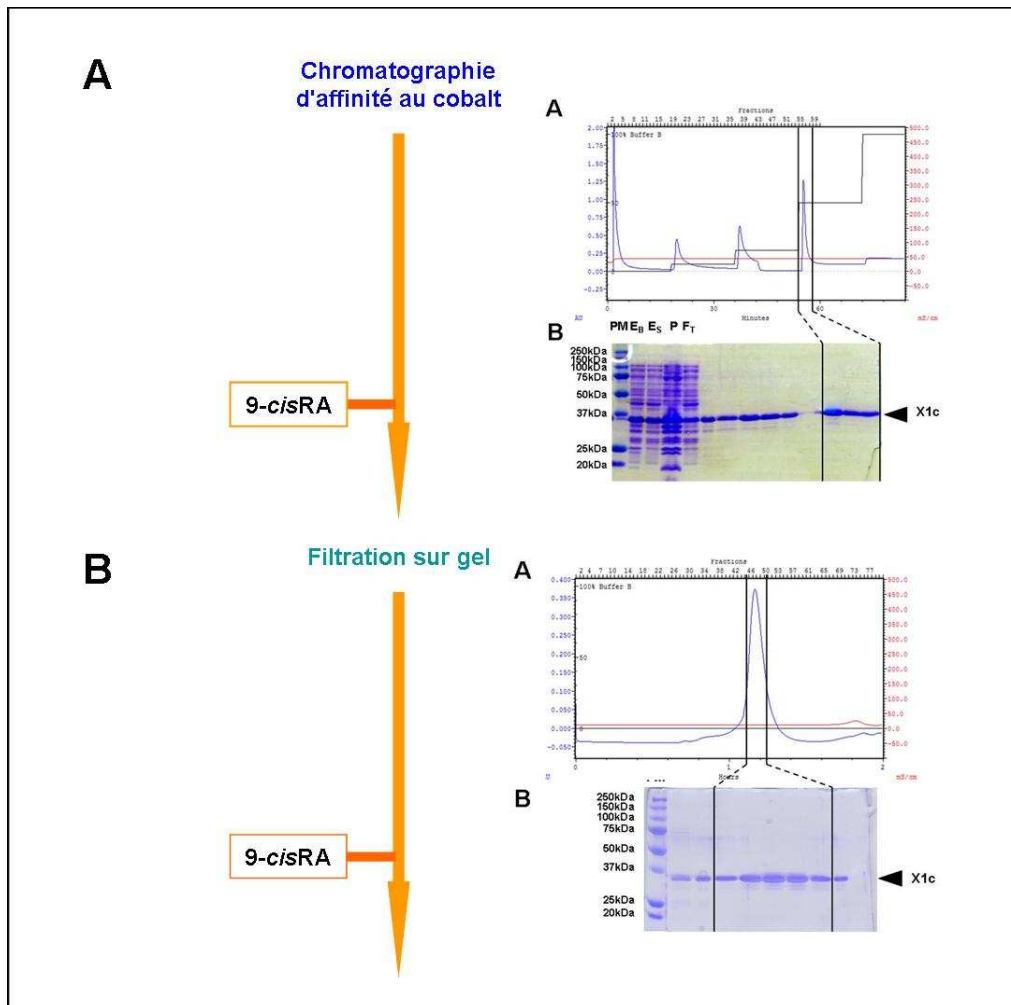


Figure 82 | Schéma de purification du mutant X1c en présence de 9-cisRA. Chromatogramme : le tracé bleu correspond à la DO_{280nm} , le rouge à la conductivité en mS/cm et le gris au gradient d'Imidazole. Gel SDS-PAGE 12% : PM, marqueurs de poids moléculaire ; EB, Extrait Brut après sonication ; Es, Extrait Soluble, P, culot, F_T, Flow Through.

10.2.3.1. Chromatographie d'affinité sur résine chélatant les ions cobalt

Le lysat clarifié est chargé sur une colonne d'affinité par chélation de cations métalliques Co²⁺. A la suite de cette chromatographie, nous observons trois pics (Figure 82A). Nous pouvons également noter, d'après le gel SDS-PAGE 12%, la migration atypique de X1c déjà observée pour X0 lors de la purification de l'hétérodimère R1d-X0 (voir section 9.1.3.1). En effet, le poids moléculaire calculé de X1c est de 30 346 Da alors que d'après le gel, il apparaît plus proche de 37 kDa. Les fractions correspondantes au troisième pic, contenant une plus faible proportion de contaminants, ont été rassemblées. Une solution éthanolique saturée de *9-cisRA* est rajoutée.

10.2.3.2. Filtration sur gel

La filtration sur gel permet de séparer les protéines en fonction de leur masse moléculaire apparente, d'éliminer toute trace éventuelle d'agrégats ou d'oligomères (Figure 82B). L'échantillon est concentré, puis injecté sur la colonne de filtration sur gel. Un seul pic est observé, correspondant au mutant X1c de masse moléculaire apparente d'environ 70 kDa, sans trace d'agrégat ni d'autres espèces oligomériques et présentant une excellente pureté (supérieure à 98%). Si nous n'avions pas utilisé de ligand *9-cisRA*, nous aurions observé un deuxième pic correspondant à la forme tétramérique du mutant, ce qui n'est pas le cas ici. On peut noter la présence d'un contaminant situé sous le mutant X1c mais en quantité négligeable. Ce mutant X1c existerait donc sous la forme d'un dimère, mais cette hypothèse reste à confirmer. Cette filtration sur gel permet également d'éliminer l'excès de *9-cisRA* et de conditionner le mutant X1c en vue des tests de cristallisation.

A ce stade, le mutant X1c est stable dans le tampon de la filtration sur gel (100 mM NaCl, 10 mM TrisHCL pH8, 0,1 mM DIFP et 28 mM β -mercaptoéthanol) en présence de la fusion polyhistidine.

10.2.3.3. Rendement de purification

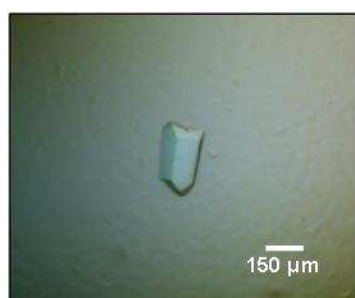
Le rendement de purification est de l'ordre de 7 mg de mutant X1c pour 3 l de culture (Tableau 19).

X1c : hXR α W305A	Volume (ml)	Concentration (mg/ml)	Quantité (mg)	Rendement (%)
Chromatographie d'affinité au cobalt	25	0,74	18,5	
Concentration	3	3,8	11,4	60
Filtration sur gel	7,5	0,9	6,7	59

Tableau 19 | Rendement de purification du mutant X1c.

10.2.4. Cristallisation du mutant X1c

Le mutant X1c, purifié à plus de 98% a été concentré à 7 mg/ml, puis cristallisé selon la technique de diffusion de vapeur en goutte assise puis cristallisé selon la technique de diffusion de vapeur en goutte assise (100 nl de la protéine + 100 nl de solution de cristallisation contre un réservoir de 70 μ l) en utilisant les kits *Index*TM, *SaltRX*TM, *The Ammonium Sulfates*, *The Sparse Matrix 1*, *The Classics*, *The JCSG+*, *The ProComplex*, *The PEG-Ion pH*, *PEG-Ion « maison »* et *NR-LBD96*. Ce criblage ne nous a permis d'obtenir sur 912 conditions testées, en trois semaines à 24°C, qu'une seule forme cristalline dans la condition 94 du *screen ProComplex* (Figure 83).



ProComplex (94) à 24°C

50 mM Magnesium chloride

5% PEG 4000

0,1 M MES pH=6,5

10% 2-propanol

Figure 83 | Cristal du mutant X1c. Le nombre entre parenthèses correspond à la condition du *screen*.

Cet unique cristal a été reproduit puis testé. Cette forme cristalline s'est révélée être très fragile au montage dans une *cryoloop*, probablement en raison d'un pourcentage de solvant élevé. Différents agents cryoprotectants ont été utilisés : différents types d'huile

(Fomblin), le glycérol et l'éthylène glycol. Malheureusement, ces cristaux de protéine ne diffractent pas.

Ce résultat a été obtenu dans les derniers mois de cette thèse. Une étude complémentaire reste à réaliser sur ce mutant très intéressant, à savoir :

- la spectrométrie de masse en conditions natives nous permettra de vérifier si le mutant X1c est capable ou non de fixer le ligand *9-cisRA*
- un gel natif et l'ultracentrifugation analytique nous donneront des informations concernant l'état d'oligomérisation de X1c
- la co-cristallisation avec un peptide CoA (Tif-2). Le CoA permettra probablement de stabiliser H12 qui en absence de ce tryptophane risque d'être instable et engendrer une instabilité de l'empilement cristallin à l'origine de cette fragilité observée lors du montage dans les *cryoloop*.

Dans les expériences ultérieures, nous poursuivrons également la purification des mutants de la voie de communication en sachant que nous serons sans doute confrontés aux mêmes problèmes rencontrés lors de l'étude structurale du système hERR γ .

10.3. Etude fonctionnelle

Pour déterminer si les différentes mutations affectent la fonction de hRXR α , nous avons testé l'activité transcriptionnelle des différents mutants en utilisant un gène rapporteur, la luciférase, sous le contrôle d'un élément de réponse synthétique, un 17-mère, de GAL4 placé en amont du promoteur de la thymidine kinase du virus *Herpes simplex* (17m-tk-LUC). Nous avons ensuite testé la capacité des différents mutants à homodimériser sur un élément de réponse DR1 et à hétérodimériser avec hRAR α sauvage sur un élément de réponse DR5 (Tableau 20) par des expériences de gel retard en absence/présence du ligand *9-cisRA*.

Elément de réponse	Séquence
DR1	TGTAGAGGTCAGAGGTCCTC ACATCTCCAGTCTCCAGTGAG
DR5	AGGGTTCACCGAAAGTTCCTCGCA TCCCAAGTGGCTTTCAAGTGAGCGT

Tableau 20 | Séquences des éléments de réponse DR1 et DR5.

Une quantité équivalente de protéine sauvage et de mutants a été utilisée dans les expériences de gel retard comme l'atteste la Figure 84.

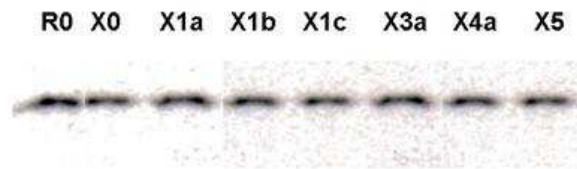


Figure 84 | Vérification de la transcription-traduction *in vitro* des protéines sauvages de hRXR α (X0) et de hRAR α (R0) et des différents mutants de hRXR α (X1a, X1b, X1c, X3a, X4a et X5).

Pour en faciliter l'interprétation, l'ensemble des résultats de ces deux expériences est rassemblé dans le Tableau 21 ci-dessous. Les gels des expériences de retard sur gel sont présentés en Annexe 6.

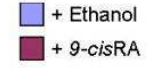
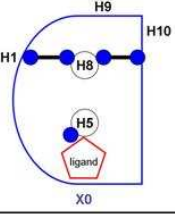
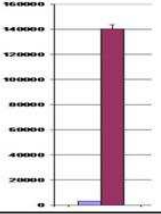
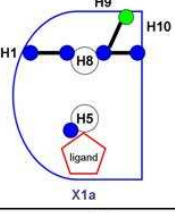
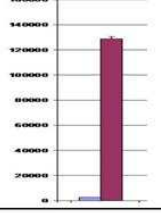
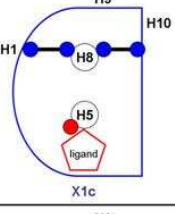

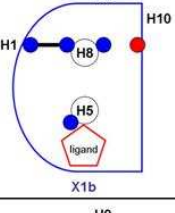
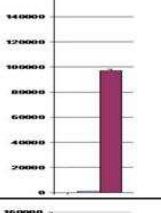
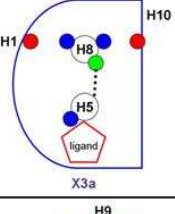

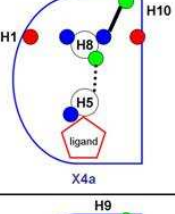
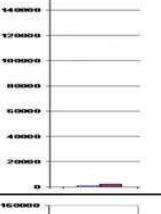
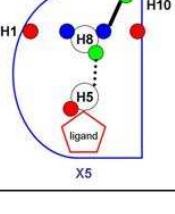
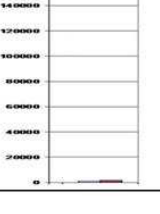
hRXR α	Homodimérisation (DR1)		Hétérodimérisation (DR5) avec R0		Transactivation 
	- AR 9-cis	+ AR 9-cis	- AR 9-cis	+ AR 9-cis	
 X0	+	++	+	++	
 X1a	++	+++	+	++	
 X1c	++	+++	-	++	
 X1b	-	-	-	-	
 X3a	-	-	-	-	
 X4a	-	-	-	++	
 X5	-	+++	-	++	

Tableau 21 | Bilan des expériences de gel retard et de transactivation pour les différents mutants de hRXR α .

10.3.1. Importance du pont connectant H8 à la région H9-10 pour l'oligomérisation

Pour investiguer les différentes contributions des résidus de la voie de communication spécifique de la classe I sur le comportement oligomérique de hRXR α , la capacité de différents mutants à homo- et/ou à hétérodimériser a été analysée par des expériences de retard sur gel, au travers de l'efficacité du dimère à se fixer sur des éléments de réponse. Noter que l'expression différentielle ou la stabilité des différents mutants de hRXR α ne sont pas à prendre en compte dans ces résultats, puisque tous les mutants sont exprimés de manière similaire (Figure 84).

Remarquablement, la mutation de l'arginine différentiellement conservée R414A en N-ter de H10, chez les mutants X1b et X3a, a un impact clair sur la dimérisation en abolissant l'homodimérisation et l'hétérodimérisation avec hRAR α . Cependant, le rétablissement de la connexion entre l'hélice H8 et H9 des mutants X4a et X5, par la mutation de l'histidine H408 en arginine équivalente de la classe II, restaure l'hétérodimérisation avec hRAR α . Dans le cas du mutant X5, l'« équivalent » d'un récepteur de la classe II, l'homodimérisation peut être, elle-même, restaurée en présence de *9-cisRA*. Il semblerait donc que le tryptophane pointant dans la LBP ait un rôle dans la dimérisation, puisque la caractéristique principale qui différencie les mutants X4a et X5 correspond justement à ce résidu. Cette observation concorde avec le résultat obtenu pour le mutant X1c. En effet, ce dernier homo- et hétérodimérise en présence de *9-cisRA*.

Il ressort de ces études que le plus important est de stabiliser la boucle H9-H10, puisque la conversion d'un pont en un autre permet le rétablissement d'une forme de dimérisation. Tous ces résultats confirment donc l'importance cruciale du pont connectant l'hélice H8 à la région H9-10, c'est-à-dire vers l'interface de dimérisation, pour l'oligomérisation tout en sachant que les résidus de la voie ne sont pas directement impliqués dans cette interface.

10.3.2. La LBP et le pont connectant H1 à H8 sont essentiels pour l'activité transcriptionnelle de hXR α

La voie de communication spécifique de la classe I suggère un « échange d'informations » entre la surface d'interaction des cofacteurs, adjacente à l'hélice H1, et l'interface de dimérisation. L'extrémité C-ter de H1 est très proche de la crevasse de fixation du motif LXXLL commun à tous les cofacteurs (Heery *et al.*, 1997; Shiau *et al.*, 1998). En effet, des mutations localisées dans cette région empêchent la fixation de NCoR chez le récepteur TR (Horlein *et al.*, 1995). Or, l'activité transcriptionnelle d'un récepteur dépend du positionnement correct de l'hélice H12 dans une conformation agoniste, ce qui inclue bien entendu les surfaces d'interaction pour les CoR et CoA. Pour évaluer l'importance des résidus de la voie de communication de la classe I sur l'activité transcriptionnelle du récepteur, la capacité des différents mutants de hXR α à activer la transcription d'un gène rapporteur en présence du ligand 9-*cis*RA a été examinée. Remarquablement, la mutation du tryptophane W305A, pointant dans la LBP, du mutant X1c diminue radicalement l'activité transcriptionnelle de hXR α . De manière similaire, lorsque le pont connectant H1 à H8 des mutants X3a et X4a est brisé, l'activité transcriptionnelle de hXR α est abolie. Noter que dans le cas du mutant X1a, mutation correspondant au résidu équivalent impliqué dans le pont spécifique de la classe II, l'activité transcriptionnelle est sensiblement équivalente à celle de hXR α sauvage. Il est intéressant de noter que, d'après les résultats obtenus, le mutant X1b n'homodimérise ni hétérodimérise avec R0 et cependant son activité transcriptionnelle reste significative. Une hypothèse possible serait que ce mutant existerait sous une forme monomérique.

Tous ces résultats démontrent clairement que le tryptophane et le pont connectant H1 à H8 sont essentiels pour maintenir la capacité de hXR α à activer la transcription du gène rapporteur en réponse au ligand.

De nombreux enseignements ont d'ores et déjà pu être tirés de ces différentes études. Cependant, pour mieux appréhender la compréhension de cette voie de communication, d'autres mutants complémentaires (certains sont déjà disponibles) sont nécessaires.

Discussion - Perspectives

Cette thèse est dédiée à l'étude de l'oligomérisation des Récepteurs Nucléaires (RNs), facteurs de transcription impliqués dans de nombreux processus biologiques. Dans ce cadre, deux approches à haut débit complémentaires, caractéristiques de l'ère post-génomique, ont été utilisées : la bioinformatique et la génomique structurale couplées à des expériences de biologie moléculaire. De ce point de vue, au-delà des résultats obtenus qui seront discutés dans la seconde partie de ce chapitre, ces différents travaux ont également contribué à mettre en évidence les forces et les faiblesses de chacune de ces deux approches.

L'étude des RNs a montré l'importance que peut avoir la bioinformatique dans des études appliquées à la recherche biologique. L'illustration la plus flagrante de cette importance est sans aucun doute la mise en évidence des deux classes de RNs par l'analyse couplée des séquences et des structures. Cependant, nous avons également pu constater, au cours de nos travaux, que le biologiste se retrouve fréquemment confronté à un problème récurrent, nouveau pour ce domaine scientifique, à savoir l'abondance de données. Cette abondance est d'autant plus problématique que la qualité des données initiales à traiter est tout à fait variable. Cette forte hétérogénéité de la qualité des données est sans conteste liée à la production même des données qui se base couramment sur différentes technologies et analyses à haut débit couplant, à chaque étape, approches expérimentales et informatiques. Ce couplage intime se vérifie aussi bien au niveau de la genèse de séquences (ADNc ou génome) et ce à toutes les étapes de séquençage, de prédiction des gènes ou d'annotation structurale ou fonctionnelle qu'au niveau des études de génomique fonctionnelle, les « omiques », qui s'appuient toutes sur de nombreuses étapes de simplifications et d'approximations inhérentes aux multiples approches statistiques disponibles. Bien qu'indispensable aux approches à haut débit, l'émergence de cette « biologie statistique » pose le problème des nombreuses approximations et/ou erreurs générées qui restent souvent méconnues et inaccessibles au biologiste. Plus grave encore, ces erreurs peuvent s'accumuler, jusqu'à masquer le message biologique recherché. De ce point de vue, le système biologique que représentent les RNs n'a pas fait exception à la règle puisque nous avons été confrontés, dès le début de notre étude, à de nombreuses erreurs de prédiction de gènes et de protéines. Ces erreurs ont abouti, tous organismes confondus, à l'élimination d'environ 25% (350) des 1432 protéines ou fragments homologues à un RN. Ceci illustre bien l'importance du

phénomène, même dans le cas d'un système fortement étudié comme les RNs, et ce dès le premier stade de l'analyse, à savoir la séquence primaire.

Pour étudier en profondeur cette famille de protéines, nous nous sommes largement appuyés sur la puissance intégrative de l'alignement multiple de séquences complètes que ce soit pour l'étude globale par domaine ou pour l'étude fine des conservations au sein de cette famille, notamment les conservations différentielles au sein du LBD. Cependant, le rôle de cet alignement validé manuellement et structuralement ne s'arrête pas là et illustre bien les avantages d'une synergie entre analyse de séquences et développement d'outils bioinformatiques. En effet, d'ores et déjà, cet alignement, qui intègre des protéines présentant des compositions, longueurs, patrons de conservation et organisations en domaines complexes, constitue un système de référence utilisé au laboratoire pour mieux estimer les avantages et inconvénients des nouveaux algorithmes de construction d'alignement multiple. Par ailleurs, l'alignement des RNs nous a amené à développer un nouveau prototype d'analyse des régions désordonnées au sein des séquences, KoAnno, qui a non seulement permis de révéler des motifs potentiellement intéressants dans les domaines A/B des RNs, mais qui a surtout permis d'approfondir notre réflexion sur le rôle du chaos au sein des protéines. De ce point de vue, il est clair qu'à l'avenir nous envisageons d'améliorer les outils de prédiction de ces régions, notamment en s'appuyant sur les alignements multiples, afin d'intégrer systématiquement ces informations aussi bien dans les stratégies de clonage et d'expression de protéines que dans l'analyse des résultats.

L'étude fine des conservations différentielles au sein des LBDs a permis de révéler l'existence de deux grandes classes de RNs réunissant certaines des six sous-familles déjà définies essentiellement sur la base d'alignements de DBDs (Laudet, 1997) : NR2, 3, 5 et 6 pour la classe I et NR1 et 4 pour la classe II. Ces deux classes se distinguent par un nombre très restreint de résidus (trois pour la classe I et deux pour la classe II) qui reflètent bien la variabilité des LBDs lorsque l'on considère un large éventail d'organismes (de *C. elegans* à l'homme). Au niveau de l'analyse, l'apport du génome de *C. elegans* a été essentiel car il a révélé que les motifs canoniques $[(I,L)lx(A,I)Exxh]$, $[hxcAHxxT]$ et $[(F,W)AKX_4FX_2LX_3DQX_2LL]$ n'étaient pas conservés universellement. Dès lors, il est apparu

que les résidus constitutifs des classes étaient non seulement peu nombreux (un résidu acide universel et quelques résidus communs à des sous-familles distinctes) mais surtout isolés et dispersés dans différentes zones de l'alignement. Lors de l'analyse structurale, ces résidus épars, qui ne représentaient pas un motif classique au sens de la bioinformatique (plusieurs résidus colinéaires conservés), ont révélé qu'ils constituaient en fait un véritable « motif spatial » préservant des ponts salins qui traversent l'architecture des LBDs. De toute évidence, en réunissant un grand nombre de séquences et de structures à haute résolution, la famille des RNs représente un système privilégié pour mettre en évidence un tel « motif spatial ». Cependant, il est clair qu'à l'avenir, grâce aux efforts de la génomique et de la génomique structurale, le nombre de systèmes ouverts à l'étude de « motifs spatiaux » va se multiplier, entraînant l'émergence d'une nouvelle approche bioinformatique couplant intimement les mondes de l'analyse des structures primaires et quaternaires.

L'introduction de la notion de classes et de ponts salins différenciellement conservés au sein de la famille des RNs s'est révélée particulièrement fructueuse en permettant de connecter une surface commune (l'interface de dimérisation) à des surfaces différentes, fonctionnellement actives : l'interface de reconnaissance des cofacteurs (classe I) et la poche de fixation du ligand (classe II). Ce résultat, établi sur la base de l'analyse de l'ensemble des structures de LBDs, a permis, qui plus est, de relier directement les classes avec le comportement oligomérique des RNs : la classe I comprend tous les homodimères et la classe II, tous les RNs qui forment des hétérodimères avec RXR (vertébrés) ou USP (invertébrés). Ceci nous a amené à proposer l'existence de deux voies de communication connectant les surfaces actives des RNs. Cette hypothèse s'intègre parfaitement dans les connaissances accumulées sur les RNs qui pointent fréquemment l'importance, pour l'activation de la transcription, d'interactions allostériques complexes entre les surfaces fonctionnelles du LBD et ses nombreux partenaires (Germain *et al.*, 2006). Ainsi, la fixation d'un partenaire tel que le ligand induit un changement conformationnel majeur du site de fixation d'autres partenaires (les cofacteurs) et de l'hélice H12 entraînant l'échange entre corépresseur et coactivateur et l'activation de la transcription. Dans le même esprit, de nombreux résultats ont révélé un lien étroit entre modifications post-traductionnelles du LBD et modifications des interactions avec différents acteurs de la transcription tels que : l'ADN et les éléments de réponse

génomiques, les ligands, les corégulateurs, d'autres voies majeures de transcription, le protéasome et, bien évidemment, la machinerie de transcription elle-même. Ces quelques exemples illustrent bien la vision actuelle des RNs qui apparaissent de plus en plus au cœur d'un réseau d'interactions complexes détectant les signaux externes et internes à la cellule, les interprétant et les transmettant à des effecteurs afin d'adapter le comportement cellulaire au travers de la transcription de programmes génétiques adaptés. Dans ce contexte, il est intéressant de rappeler que, par delà leur type d'interactions, homodimère et hétérodimère diffèrent fondamentalement dans leur capacité à intégrer des signaux complexes : les homodimères reconnaissant théoriquement un seul type d'élément de réponse et un seul ligand tandis que les RNs, en hétérodimérisant avec RXR ou USP, sont à même d'introduire une importante combinatoire basée sur des éléments de réponse et des ligands différents. De ce point de vue, on peut noter que nos résultats ont révélé que *C. elegans*, Métazoaire relativement simple, ne posséderait que des RNs homodimères de classe I. Ce résultat s'inscrit bien dans les hypothèses évolutives actuelles qui suggèrent que l'apparition des homodimères serait antérieure à celle des hétérodimères et aux vagues de duplications et événements de gains et pertes qui ont jalonné l'histoire des RNs.

Pour tester expérimentalement notre hypothèse concernant les voies de communication, nous avons décidé d'élaborer une stratégie basée sur une série de mutations progressives des résidus spécifiques des classes I et II, afin de transformer un RN de classe I (homodimère) en RN de classe II (hétérodimère) et *vice versa*. Cette stratégie ambitieuse a semblé la seule susceptible de vraiment valider conjointement les deux voies de communication. De plus, compte tenu des nombreux signaux et partenaires associés aux RNs, il nous est très vite apparu qu'il serait très difficile de distinguer les effets spécifiques des mutations dans un contexte de mécanismes fortement interconnectés (reconnaissance de l'ADN, reconnaissance du ligand, dimérisation, transactivation, etc.).

Pour atteindre notre objectif, la stratégie élaborée s'est appuyée sur :

- trois systèmes tests choisis pour leur comportement oligomérique différent impliquant ERR γ (classe I) qui est un récepteur homodimérique orphelin constitutivement actif ; RAR α (classe II) qui hétérodimérise avec RXR α ; RXR α (classe

I), le récepteur pivot, qui hétérodimérise avec l'ensemble des RNs de classe II et qui, de plus, homodimérise,

- la construction de pas moins de 38 mutants tous systèmes confondus, allant du simple au quintuple mutant,

- quatre approches expérimentales complémentaires pour analyser l'impact d'une ou plusieurs mutations incluant : la spectrométrie de masse en conditions natives et dénaturantes afin d'étudier le comportement d'un LBD mutant vis-à-vis de sa capacité à fixer un ligand et à dimériser ; des expériences de retard sur gel pour analyser la dimérisation de protéine DBD-*hinge*-LBD mutante en présence d'un fragment d'ADN contenant son élément de réponse ; des tests d'activité transcriptionnelle, en présence et absence de ligand, pour analyser la transactivation et tester l'intégrité du mécanisme dit du « piège à souris » et, enfin, la diffraction des rayons X afin d'obtenir, au niveau atomique, la compréhension structurale des conséquences des mutants.

Il est clair qu'un tel projet à moyen/haut débit n'est concevable que dans une logique de génomique structurale qui implique la disponibilité d'une plate-forme, telle que la plate-forme de Biologie Structurale de Strasbourg, fortement robotisée et qui propose des protocoles standardisés et rôdés. Cependant, cela nécessite également la mise en place d'un véritable module expérimental afin de systématiser les étapes d'obtention, de validation, d'expression puis de purification des différents mutants. Dans l'absolu, ce module expérimental, en s'appliquant successivement à chaque mutant intermédiaire, devait permettre l'obtention de l'ensemble des mutants multiples, du simple au quintuple et ce pour tous les systèmes étudiés. Dans la pratique, hélas, certains mutants ayant fait défaut, les trois systèmes de RNs n'ont pas pu être investigués avec le même degré de précision. De même, l'ensemble des quatre approches expérimentales envisagées n'a pu être appliqué à tous les mutants pour des raisons multiples allant de simples problèmes techniques aux problèmes inhérents au système étudié, tel ERR γ pour lequel, à ce jour, aucun ligand n'est disponible. Dès lors, si l'on fait un bilan rapide, on peut clairement pointer le système RXR α sur lequel nous nous sommes concentrés et pour lequel six mutants (trois simples, un triple, un quadruple et un quintuple), tous solubles et exploitables, ont été obtenus tandis que seuls

deux simples mutants étaient testés dans le système RAR α et un seul pour ERR γ malgré l'obtention de cinq mutants (un mutant par étape du simple au quintuple).

Malgré ces données manquantes, les résultats obtenus ont déjà permis de tirer quelques enseignements sur l'importance et le rôle des voies de communication au sein des RNs.

Pour rappel, schématiquement, les deux voies de communication qui traversent les RNs par leur cœur structural (les hélices H8 et H5) permettent de connecter au travers de deux « motifs spatiaux » distincts, d'une part, une région commune, correspondant au début (classe II) ou à la fin (classe I) de la boucle H9-H10 impliquée dans l'interface de dimérisation et, d'autre part, deux surfaces fonctionnellement actives, à savoir, l'interface de reconnaissance des cofacteurs (classe I) et la poche de fixation du ligand (classe II). Bien que spatialement éloignées, ces dernières surfaces participent toutes deux à la transactivation au travers du mécanisme dit du « piège à souris ». Ce mécanisme, suite à l'adjonction du ligand, implique un mouvement important de l'hélice H12 en position d'activation de la transcription (position agoniste) entraînant ainsi l'obturation de la poche de fixation du ligand en synergie avec une présence de coactivateurs appropriés à l'interface de reconnaissance des cofacteurs. Dans ce contexte, l'essentiel de notre effort a porté sur la compréhension précise du lien entre reconnaissance du ligand, oligomérisation et transactivation au travers de l'étude coordonnée de multiples mutants.

Toutefois, les résultats obtenus sur les protéines non mutées de RAR α par l'équipe de Cécile Egly (Gaillard *et al.*, 2006) ont suggéré l'implication des voies de communication dans une intégration encore plus globale des signaux et partenaires associés aux RNs. Ces expériences, qui ont porté sur l'étude du rôle de la phosphorylation par la PKA de la Sérine S369 localisée dans la boucle H9-H10, ont révélé que la phosphorylation de ce résidu, situé à l'interface de dimérisation, augmente l'efficacité du recrutement de la cycline H du côté opposé, à savoir l'interface d'interaction avec les cofacteurs. Par delà la confirmation de l'importance des voies de communication au sein des RNs, cette étude permet de les impliquer dans les mécanismes de régulation de la transactivation liés au cycle cellulaire. En

effet, la cycline H appartient au complexe TFIIH qui est directement impliquée dans la régulation de la transactivation par l'intermédiaire de la phosphorylation du domaine C-terminal de l'ARN polymérase II.

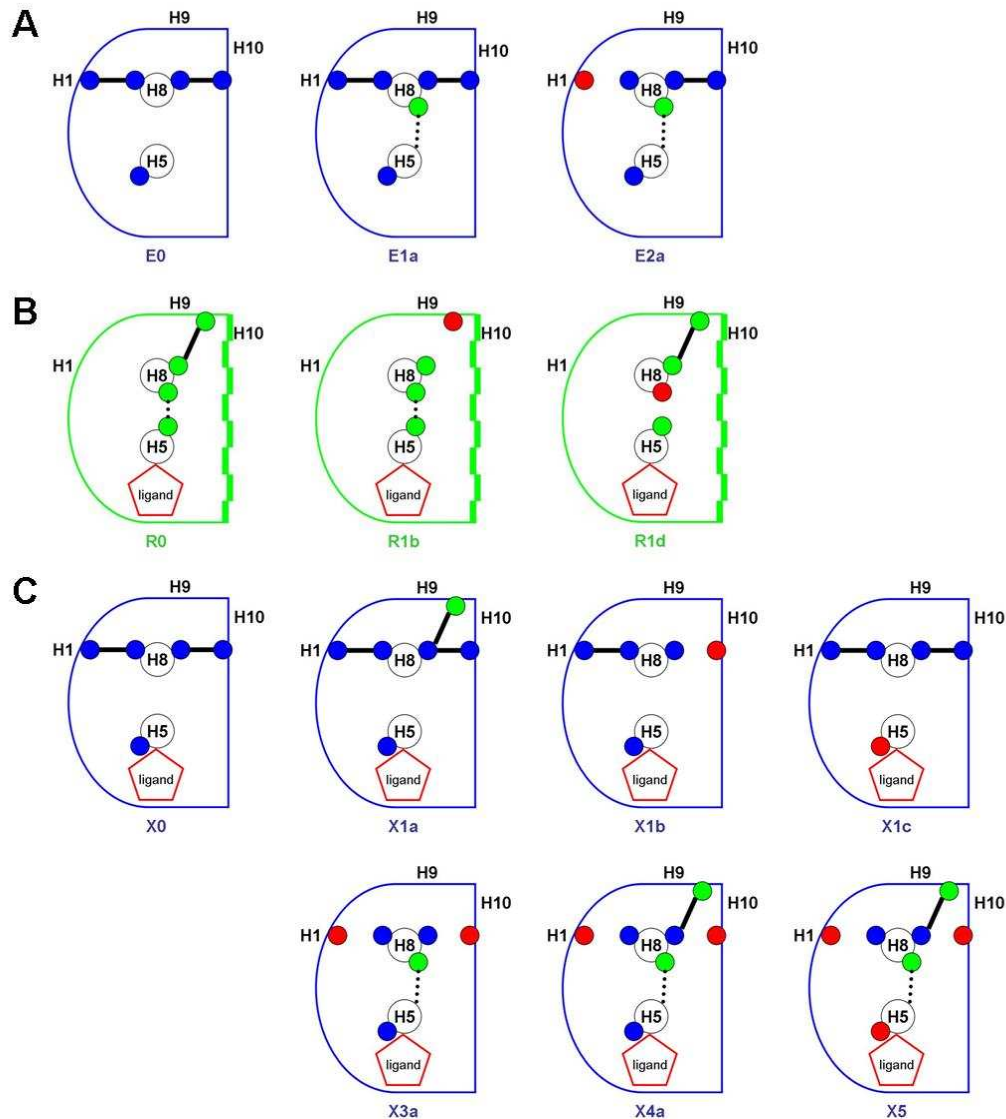


Figure 85 | Les mutants exploitables des voies de communication. (A), le système hERR γ . (B), le système hRAR α . (C), le système hRXR α .

En ce qui concerne l'analyse des mutants exploitables des voies de communication (Figure 85), de nombreux éléments indiquent que, tous systèmes confondus, le pont reliant la boucle H9-H10 au cœur structural des RNs (hélices H8 et H5) joue un rôle essentiel dans le contrôle de la dimérisation. Cette conclusion découle non seulement de toutes les mutations qui, en impliquant directement un résidu de ce pont, entraînent l'interruption du lien entre la

boucle H9-H10 et l'hélice H8 (tels les mutants R1b, X1b, X3a) et la perte totale de la dimérisation que des mutations qui maintiennent ou rétablissent le lien et une forme de dimérisation (tels les mutants E1a, R1d, X1a, X4a, X5). Dans ce dernier lot, il est capital de distinguer les simples mutants E1a et X1a, qui correspondent à l'ajout d'un résidu de classe II, à savoir, pour E1a, le résidu reliant la poche de fixation du ligand à l'hélice H8 et, pour X1a, le résidu reliant le début de la boucle H9-H10 à l'hélice H8. Ces mutants possèdent une voie de communication de classe I intacte. Cependant, dans les deux cas, la simple introduction d'un résidu de classe II, qui pouvait sembler mineure, entraîne des perturbations dans l'état d'oligomérisation allant d'un renforcement de l'homodimérisation pour X1a à la perturbation de l'homodimérisation, comme cela a été observé pour le mutant E1a. L'étude approfondie du mutant E1a a été particulièrement fructueuse, puisqu'au travers des expériences de spectrométrie de masse en conditions natives, nous avons pu montrer que l'ajout du résidu de classe II aboutit à l'existence de deux populations en échange dynamique correspondant aux formes monomérique et homodimérique. De plus, la résolution de la structure de ce mutant a permis d'établir que l'homodimère lui-même présente des perturbations dans son mode de reconnaissance du partenaire, entraînant une rotation de près de 6° entre les deux monomères mutés. Par ailleurs, la résolution à 1.4 Å de la structure du mutant E1a a révélé la présence d'un ligand d'un poids moléculaire de 182 Da, confirmée par la spectrométrie de masse. Ce résultat inattendu ouvre la voie de la recherche du ligand fonctionnel de ce RN jusqu'alors considéré comme orphelin.

L'obtention de nombreux mutants perturbant la dimérisation est également à considérer au regard des nombreuses études de mutagenèse impliquant des résidus directement localisés dans les interfaces de dimérisation (Aggelidou *et al.*, 2006; Bogan *et al.*, 2000; Chan and Wells, 2007; Lee *et al.*, 1998; Vivat-Hannah *et al.*, 2003; Vivat *et al.*, 1997). Dans leur ensemble, ces études, comme celle menée par Vivat *et coll.* (Vivat-Hannah *et al.*, 2003), ont abouti à la conclusion qu'il était très difficile d'identifier des résidus essentiels à la dimérisation au sein des grandes surfaces d'interaction des RNs qui peuvent varier de 1300 Å² pour RXR jusqu'à 1700 Å² pour ER. Ainsi, dans le cadre de l'étude de Vivat *et coll.*, il a été possible de remplacer par une alanine un grand nombre de résidus (jusqu'à sept) sans entraîner de perturbation majeure de la dimérisation. *A contrario*, dans notre étude, toute

mutation d'un résidu de voie reliant la boucle H9-H10 au cœur structural, voire le rajout d'un résidu d'une classe sur une voie intacte, ont entraîné des perturbations dans l'oligomérisation. Clairement, ces résultats suggèrent que le contrôle de l'état d'oligomérisation et surtout, le positionnement efficace de l'interface de dimérisation est indissociable d'une voie de communication fonctionnelle et renforce l'importance de ces voies dans les phénomènes allostériques intégrant reconnaissance du ligand, dimérisation et transactivation.

Les expériences d'activation de la transcription réalisées sur des RNs mutés dans les résidus classe-spécifiques impliquant d'une part, l'interface de reconnaissance des cofacteurs (pour la classe I) et, d'autre part, la poche de reconnaissance du ligand (pour la classe II) ont tous abouti à la perte totale de la capacité de transactivation. Ce résultat s'est vérifié aussi bien pour les mutants de classe I, tels X1c, X3a, X4a et X5, que pour l'unique mutant de classe II exploitable dans le cadre de tests de transactivation, à savoir R1d. Nos résultats sont en accord avec les nombreux travaux (Chen *et al.*, 2005; He *et al.*, 2006; Iordanidou *et al.*, 2005; Marimuthu *et al.*, 2002; Solomon *et al.*, 2001) qui ont montré l'importance d'une coordination de ces deux surfaces actives dans le mécanisme dit du « piège à souris ». En effet, ces deux surfaces sont essentielles au positionnement de la forme agoniste de l'hélice H12 et il semble que les remplacements des résidus de classe I par les résidus de classe II, comme c'est le cas des mutants X3a, X4a et X5, ne suffisent pas à rétablir la capacité de transactivation d'un RN. Dans ce contexte, afin de comprendre le positionnement de l'hélice H12 et de déterminer l'état (agoniste, antagoniste) des récepteurs mutés, un de nos objectifs majeurs est de réaliser des études structurales additionnelles des mutants précédemment cités qui sont stables en solution et que nous avons pu produire en grande quantité.

Pour finir, on peut noter qu'au regard de notre objectif premier de transformer un RN de classe I en classe II et *vice versa*, un grand pas a été franchi dans le système RXR α de classe I. Ce résultat majeur est illustré par les quadruple (X4a) et quintuple (X5) mutants de RXR α qui possèdent l'ensemble des résidus de classe II. Dans les deux cas, nos résultats ont montré le rétablissement de diverses formes d'oligomérisation qui plus est, dépendantes du ligand. En effet, comme nous l'avons vu (Tableau 21), en absence de ligand, aucun phénomène

d'homodimérisation ou d'hétérodimérisation n'est observé, tandis qu'en présence du ligand, X4a et X5 hétérodimérisent avec RAR α . Cependant, ces deux mutants ne sont pas équivalents puisque X5 a la capacité d'homodimériser en présence du ligand. Ce dernier résultat est particulièrement surprenant si l'on considère que la différence majeure entre ces deux mutants réside dans la mutation, chez X5, du résidu classe I spécifique qui pointe dans la poche du ligand. Dans l'absolu, cette mutation aurait du transformer le mutant X5 en récepteur purement de classe II, donc incapable d'homodimériser. Ce résultat paradoxal montre clairement que des expériences complémentaires vont être nécessaires pour comprendre vraiment le comportement de ces deux mutants, notamment au travers de l'obtention des structures qui seront les seules susceptibles de nous éclairer sur le mode de reconnaissance et le positionnement exact des interfaces de dimérisation. Cependant, pour paradoxal qu'il soit, ce résultat indique également que la voie reliant la poche du ligand et l'interface de dimérisation existe dans ces mutants puisque la mutation, chez X4a, du résidu pointant dans la poche entraîne un comportement d'homodimérisation différent entre X4a et X5.

Tous ces résultats nous confortent dans l'idée que pour intégrer en une seule protéine les nombreux mécanismes de signalisation, les RNs ont développé des voies de communication intra- et inter-moléculaires afin d'interconnecter l'ensemble des surfaces actives et partenaires impliqués dans le système et « d'informer » d'éventuelles modifications du milieu. Sans aucun doute, la réalité moléculaire sera plus riche et plus complexe que cette vision très anthropomorphique et l'étude et la compréhension fine de ce type de voies n'en est encore qu'à ses débuts. En effet, il reste encore de nombreuses expériences à réaliser et à imaginer pour révéler les nombreux mécanismes allostériques à l'œuvre pour réaliser des intégrations impliquant de multiples partenaires nucléiques, protéiques et moléculaires, l'édification, la modification et la dégradation de complexes dynamiques et surtout l'établissement d'une machinerie à même d'assurer la transcription de programmes génétiques idoines face aux changements des conditions externes ou aux différentes étapes de développement. Dans cette longue marche, le grand nombre de mutants et d'approches expérimentales déjà disponibles nous permet de disposer d'un véritable système modèle qui pourra non seulement éclairer le rôle et le mode d'action des

voies au sein des RNs, mais aussi, nous l'espérons, fournir des enseignements utiles pour de futures études abordant l'intégration des signaux cellulaires au sein des complexes macromoléculaires.

Annexes

Annexe 1 | Exemple d'alignement multiple au format FASTA

```
>RARa
LDILILRICTRYTPEQDAETGLLSAICLICGDRQDLEQPDRVDMLQEPLLEALKVYVRKR
RPSRPHMFPKMLMKITDLRSISAK
>RARb
LDILILRICTRYTPEQDTETGLLSAICLICGDRQDLEEPTKVVDKLQEPLLEALKIYIRKR
RPSKPHMFPKILMKITDLRSISAK
>RARG
LDILMLRICTRYTPEQDTETGLLSAICLICGDRMDLEEPEKVDKLQEPLLEALRLYARRR
RPSQPYMFPRLMKITDLRGISTK
>PPARa
YEAIFAMLSVNMNKD-DSDISLFAAIIICGDRPGLLNVDGHIKMQEGIVHVLRLHLQSN
HPDDIFLFPKLLQKMADLRQLVTE
>PPARb
HEAIFAMLASIVNKD-DSDLALFIAAIIICGDRPGLMNVPRVEAIQDTILRALEFHLQAN
HPDAQYLFPKLLQKMADLRQLVTE
>PPARG
HEIIYTMLASLMNKD-DSDLAIFIAVILSGDRPGLLNVDKPIEDIQDNLLQALELQKLN
HPSSQLFAKLLQKMTDLRQIVTE
>TRa
MEIMSLRAAVRYDPESDTEVALLQAVLLMSTDRSGLLCVVDKIEKSQEAYLLAFEHYVNR
KHNIPHFWPKLLMKEREVQSSILY
>TRb
MEIMSLRAAVRYDPESDTEVALLQAVLLMSSDRPGLACVERIEKYQDSFLLAFEHYINYR
KHHVTHFWPKLLMKVTDLRMIGAC
>VDR
IEVIMLSNESFTMDDEEEHVLLMAICIVSPDRPGVQDAALIEAIQDRLSNTLQTYIRCR
HPP-HLLYAKMIQKLADLRSLNEE
>RORa
LEVVFIRMCRAFDSQ-EDEIALFSAFVLMASDRSWLQEKVKIEKLQQKIQLALQHVLOKN
HRED-GILTKLICKVSTLRALCGR
```

Annexe 2 | Exemple d'alignement multiple au format MSF (*Multiple Sequence Format*)

MSF: 84 Type: P February 1, 2008 19:33 Check: 2642 ..

Name: RARa	Len: 84	Check: 4420	Weight: 0.10
Name: RARb	Len: 84	Check: 3010	Weight: 0.10
Name: RARg	Len: 84	Check: 2955	Weight: 0.10
Name: PPARa	Len: 84	Check: 2556	Weight: 1.00
Name: PPARb	Len: 84	Check: 1347	Weight: 1.00
Name: PPARg	Len: 84	Check: 2575	Weight: 1.00
Name: TRa	Len: 84	Check: 5144	Weight: 1.00
Name: TRb	Len: 84	Check: 3295	Weight: 1.00
Name: VDR	Len: 84	Check: 3934	Weight: 1.00
Name: RORa	Len: 84	Check: 3406	Weight: 1.00

//

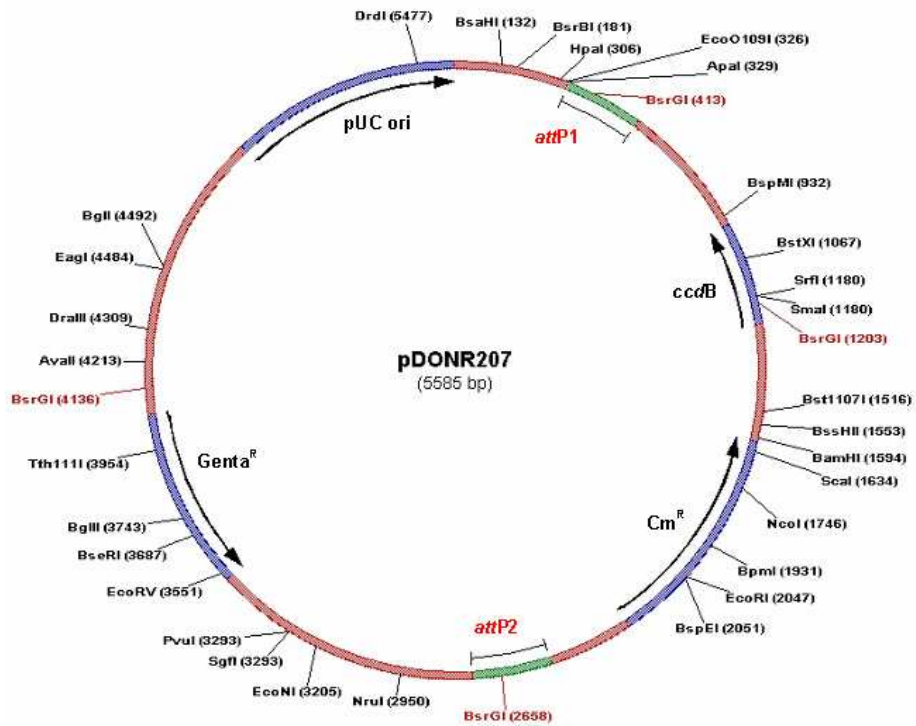
	1				50
RARa	LDILILRICT	RYTPEQDAET	GLLSAICLIC	GDRQDLEQPD	RVDMLQEPLL
RARb	LDILILRICT	RYTPEQDTET	GLLSAICLIC	GDRQDLEEPT	KVDKLQEPLL
RARg	LDILMLRICT	RYTPEQDTET	GLLSAICLIC	GDRMDLEEPE	KVDKLQEPLL
PPARa	YEAIFAMLSS	VMNKD.DSDI	SLFVAAIICC	GDRPGLLNVG	HIEKMQEGIV
PPARb	HEAIFAMLAS	IVNKD.DSDL	ALFIAAIILC	GDRPGLMNVP	RVEAIQDTIL
PPARg	HEIIYTMLAS	LMNKD.DSDL	AIFIAVIIIS	GDRPGLLNVK	PIEDIQDNLL
TRa	MEIMSLRAAV	RYDPESDTEV	ALLQAVLLMS	TDRSGLLCVD	KIEKSQEAYL
TRb	MEIMSLRAAV	RYDPESDTEV	ALLQAVLLMS	SDRPGLACVE	RIEKYQDSFL
VDR	IEVIMLSRNE	SFTMDDEEEH	VLLMAICIVS	PDRPGVQDAA	LIEAIQDRLS
RORa	LEVVFIRMCR	AFDSQ.EDEI	ALFSAFVLMS	ADRSWLQEKV	KIEKLQKIQ
	51			84	
RARa	EALKVYVRKR	RPSRPHMFPK	MLMKITDLRS	ISAK	
RARb	EALKIYIRKR	RPSKPHMFPK	ILMKITDLRS	ISAK	
RARg	EALRLYARRR	RPSQPYMFPR	MLMKITDLRG	ISTK	
PPARa	HVLRLHLQSN	HPDDIFLFPK	LLQKMADLRQ	LVTE	
PPARb	RALEFHLLQAN	HPDAQYLFPK	LLQKMADLRQ	LVTE	
PPARg	QALELQLKLN	HPESQLFAK	LLQKMTDLRQ	IVTE	
TRa	LAFEHYVNRH	KHNIPHFWPK	LLMKEREVQS	SILY	
TRb	LAFEHYINYR	KHHVTHFWPK	LLMKVTDLRM	IGAC	
VDR	NTLQTYIRCR	HPP.HLLYAK	MIQKLADLRS	LNEE	
RORa	LALQHVLLQKN	HRED.GILTK	LICKVSTLRA	LCGR	

Annexe 3 | Géotypes des souches BL21(DE3) et DH5 α

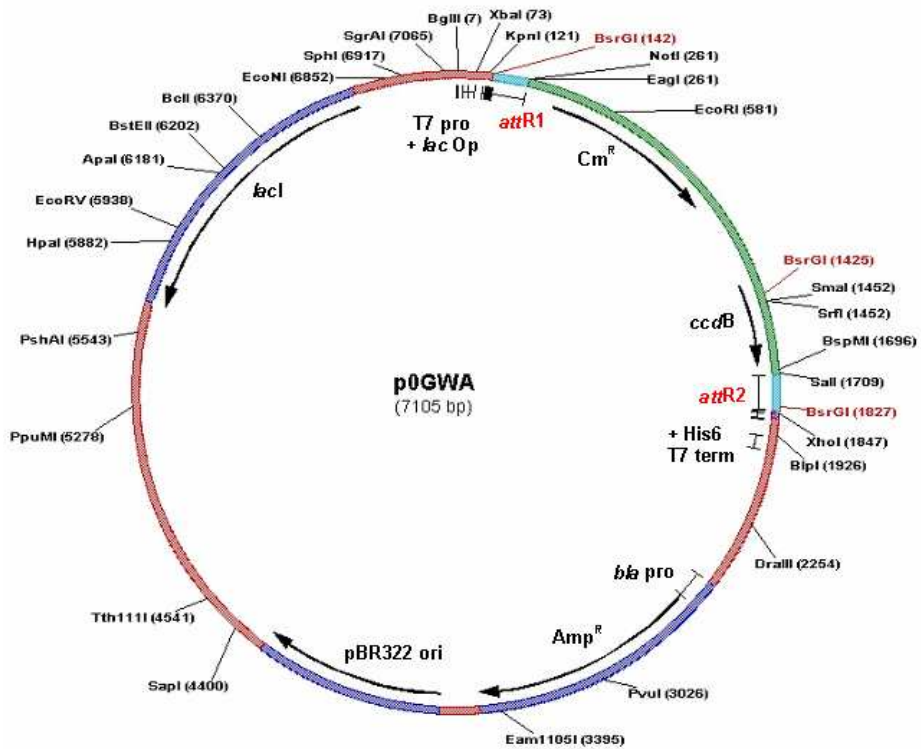
BL21(DE3) : F⁻ *ompT hsdS_B(r_B⁻ m_B⁻) gal dcm* (DE3)

DH5 α : F⁻ ϕ 80*dlacZ* Δ M15 Δ (*lacZYA-argF*) U169 *recA1 endA1 hsdR17(r_K⁻, m_K⁺) phoA supE44*
 λ *thi-1 gyrA96 relA1*

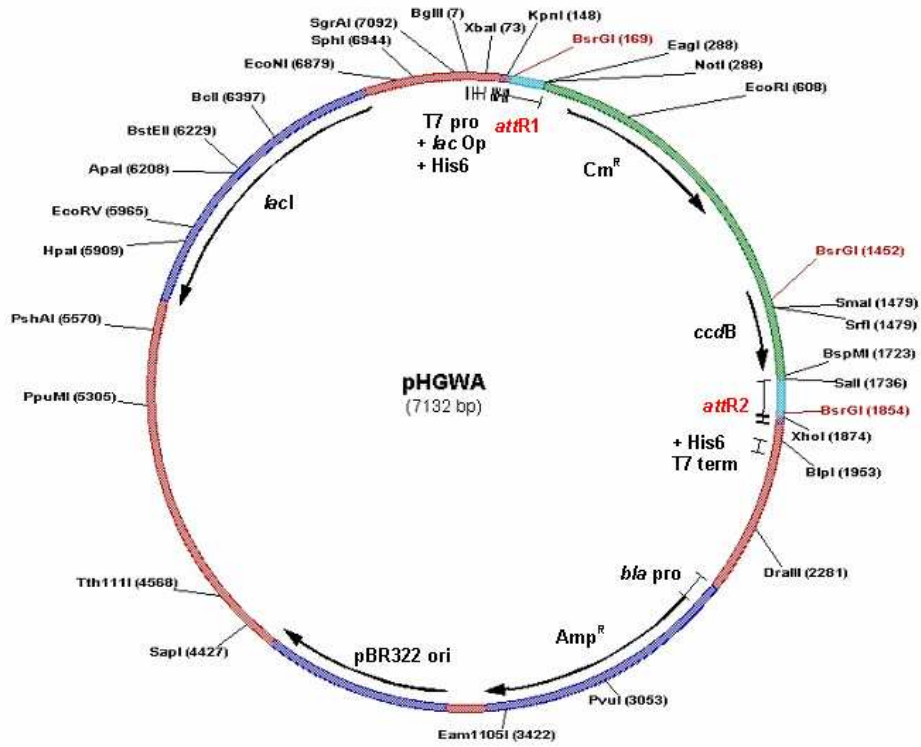
Annexe 4 | Carte de restriction du vecteur pDONR207



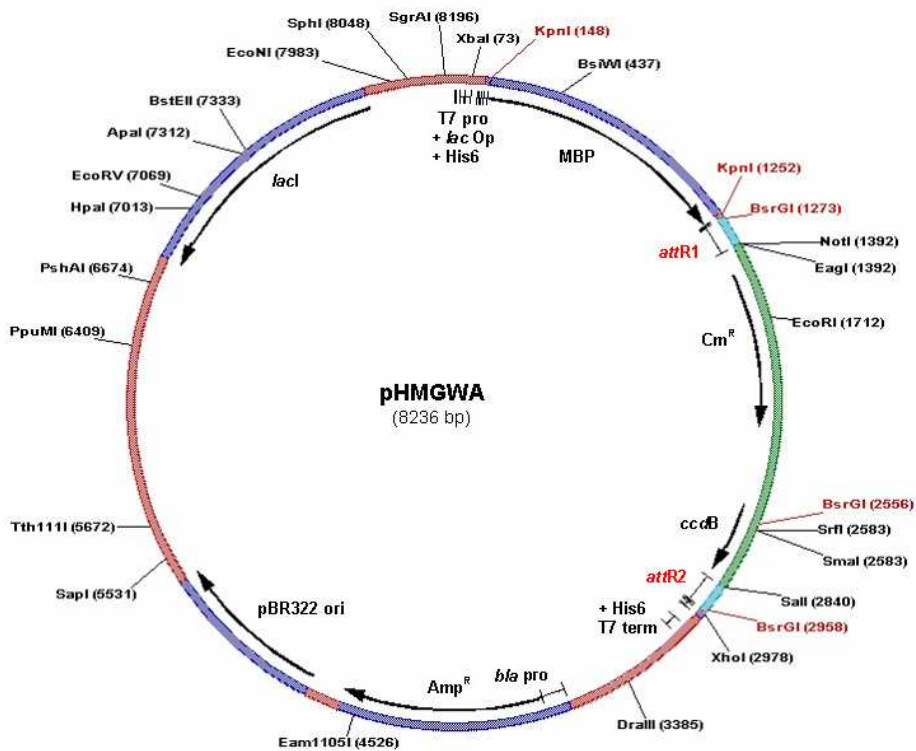
Carte de restriction du vecteur p0GWA



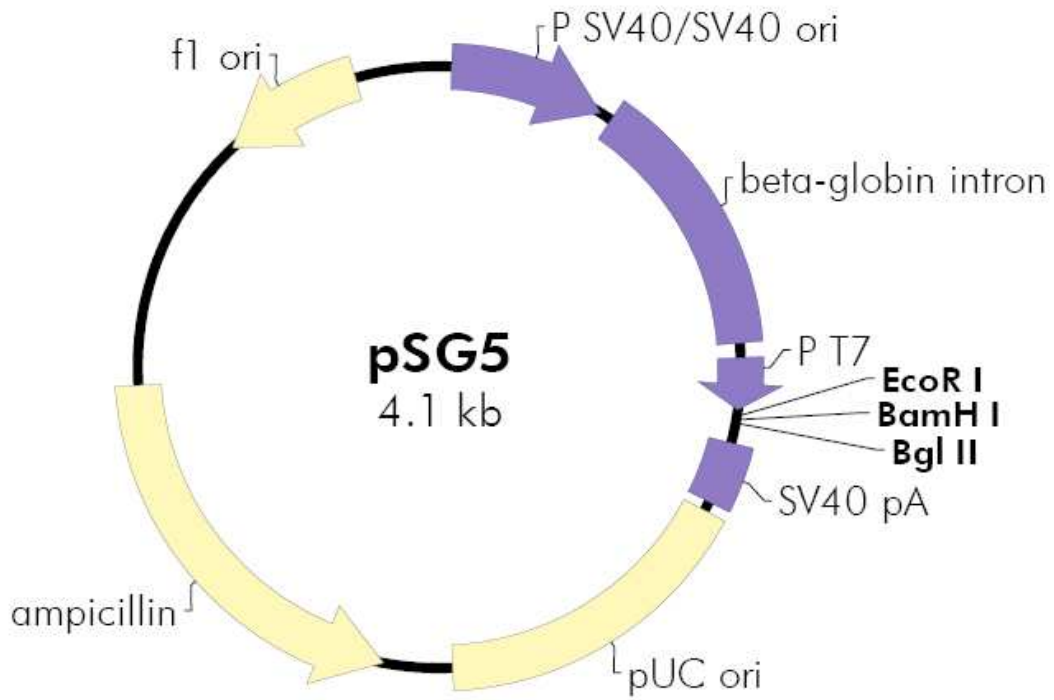
Carte de restriction du vecteur pHGWA



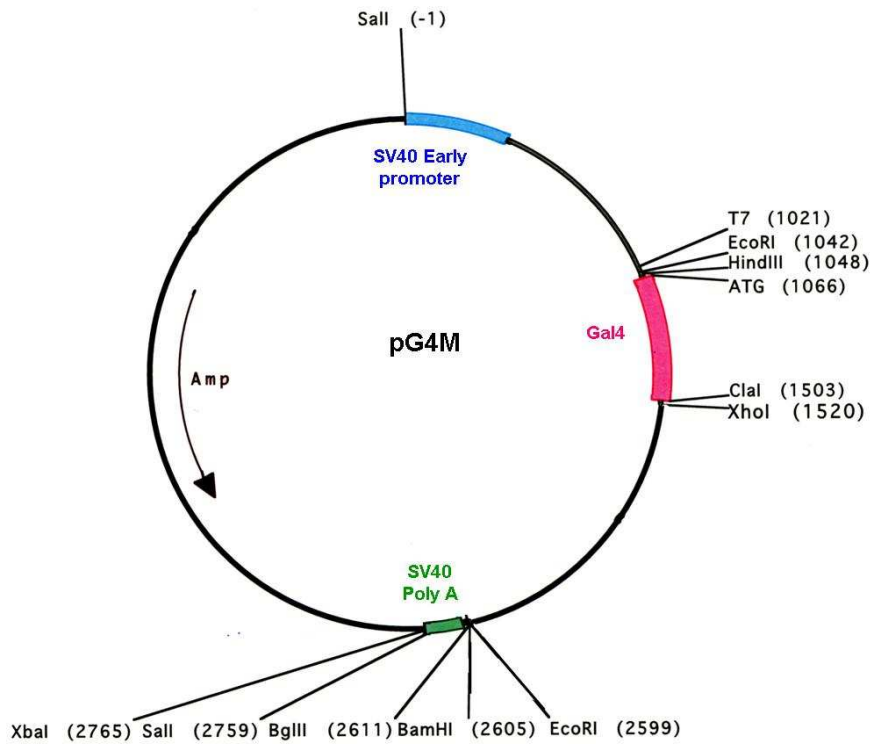
Carte de restriction du vecteur pHMGWA



Carte de restriction du vecteur pSG5



Carte de restriction du vecteur pG4M



Annexe 5 | Séquences en nucléotides et en acides aminés des différentes fusions des vecteurs p0GWA, pHGWA et pHMGWA.

p0GWA

ATGGGTACCGGATCTTACATCACAAGTTTGTACAAAAAGCAGGCTTCCTG
M G T G S Y I T S L Y K K A G F L
GTGCCACGCGGTTCTCATATG- gène -TGA-BglIII ou BamHI
V P R G S H M -protéine-stop

pHGWA

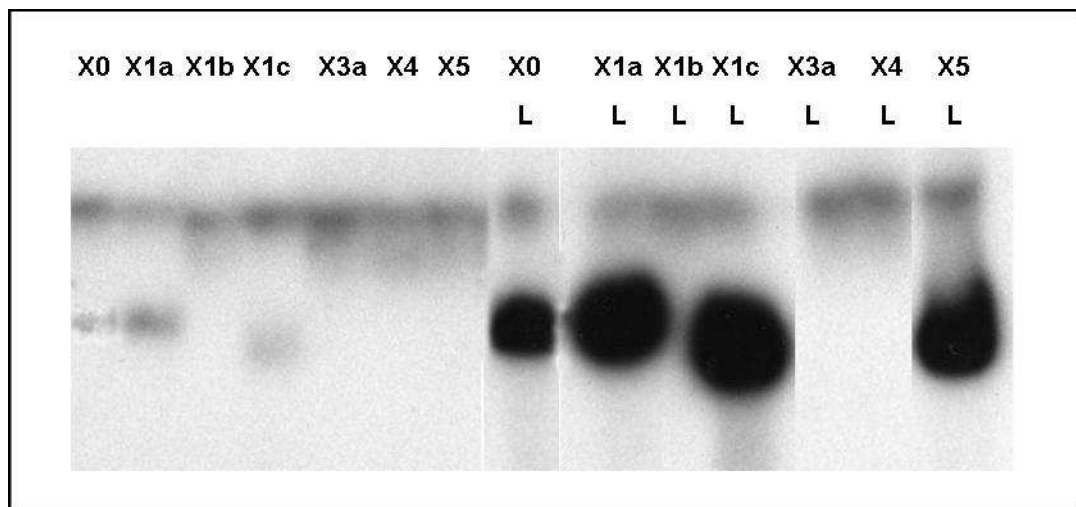
ATGGGCAGCAGCCATCATCATCATCATCACGGTACCGGATCTTACATCACA
M G S S H H H H H H G T G S Y I T
AGTTTGTACAAAAAGCAGGCTTCCTGGTGCCACGCGGTTCTCATATG- gène -TGA-BglIII ou BamHI
S L Y K K A G F L V P R G S H M -protéine-stop

pHMGWA

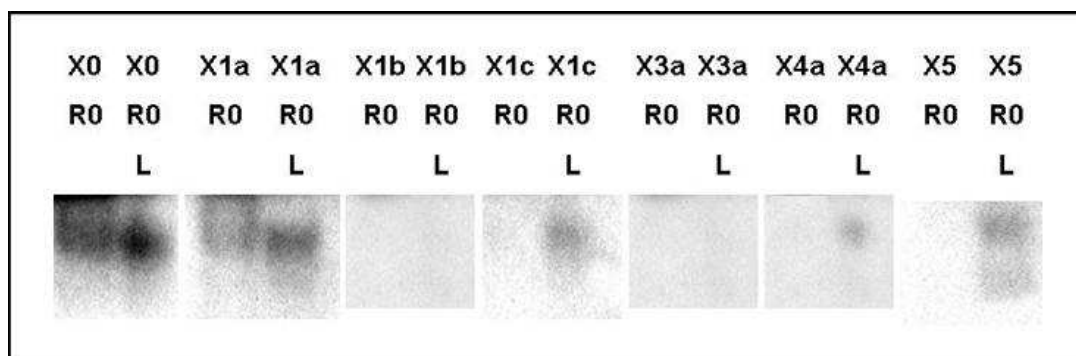
ATGGGCAGCAGCCATCATCATCATCATCACGGTACCAAACTGAAGAAGGTAACTGGTA
M G S S H H H H H H H G T K T E E G K L V
ATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCCGGTAAGAAATTCGAG
I W I N G D K G Y N G L A E V G K K F E
AAAGATACCGGAATTAAGTACCCGTTGAGCATCCGGATAAACTGGAAGAGAAATCCCA
K D T G I K V T V E H P D K L E E K F P
CAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGT
Q V A A T G D G P D I I F W A H D R F G
GGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTCCAGGACAAG
G Y A Q S G L L A E I T P D K A F Q D K
CTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATC
L Y P F T W D A V R Y N G K L I A Y P I
GCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACC
A V E A L S L I Y N K D L L P N P P K T
TGGGAAGAGATCCCGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATG
W E E I P A L D K E L K A K G K S A L M
TTCAACCTCAAGAACCCTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGTTATGCG
F N L Q E P Y F T W P L I A A D G G Y A
TTCAAGTATGAAAACGGCAAGTACGACATTAAGACGCTGGGCGTGGATAACGCTGGCGCG
F K Y E N G K Y D I K D V G V D N A G A
AAAGCGGCTGACCTTCCTGGTTGACCTGATTAAAAACAACACATGAATGCAGACACC
K A G L T F L V D L I K N K H M N A D T
GATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGC
D Y S I A E A A F N K G E T A M T I N G
CCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAAACGGTACTGCCG
P W A W S N I D T S K V N Y G V T V L P
ACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCC
T F K G Q P S K P F V G V L S A G I N A
GCCAGTCCGAACAAGAGCTGGCAAAAGAGTTCCCTCGAAAACCTATCTGCTGACTGATGAA
A S P N K E L A K E F L E N Y L L T D E
GGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAG
G L E A V N K D K P L G A V A L K S Y E
GAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCAGAAAGGTGAA
E E L A K D P R I A A T M E N A Q K G E
ATCATGCCGAACATCCCGCAGATGTCGCTTCTGGTATGCCGTGCGTACTGCGGTGATC
I M P N I P Q M S A F W Y A V R T A V I
AACGCCGCCAGCGGTCGTCAGACTGTGATGAAGCCCTGAAAGACGCGCAGACTGGTACC
N A A S G R Q T V D E A L K D A Q T G T
GGATCTTACATCACAAGTTTGTACAAAAAGCAGGCTTCCTGGTGCCACGCGGTTCTCAT
G S Y I T S L Y K K A G F L V P R G S H
ATG- gène -TGA-BglIII ou BamHI
M -protéine-stop

Étiquette Histidine
MBP
Thrombine
NdeI

Annexe 6 | Gels des expériences de retard sur gel



Test de la capacité des différents mutants du système hRXR α à homodimériser sur un élément de réponse DR1 en absence/présence du ligand *9-cisRA*.



Test de la capacité des différents mutants du système hRXR α à hétérodimériser avec R0 sur un élément de réponse DR5 en absence/présence du ligand *9-cisRA*.

Références Bibliographiques

1999. A unified nomenclature system for the nuclear receptor superfamily. *Cell* 97(2):161-3.
- Aggelidou E, Iordanidou P, Demetriades C, Piltsi O, Hadzopoulou-Cladaras M. 2006. Functional characterization of hepatocyte nuclear factor-4 alpha dimerization interface mutants. *Febs J* 273(9):1948-58.
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387(6632):489-93.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-402.
- Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181(96):223-30.
- Aranda A, Pascual A. 2001. Nuclear hormone receptors and gene expression. *Physiol Rev* 81(3):1269-304.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT and others. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-9.
- Attwood TK. 2002. The PRINTS database: a resource for identification of protein families. *Brief Bioinform* 3(3):252-63.
- Avantaggiati ML, Ogryzko V, Gardner K, Giordano A, Levine AS, Kelly K. 1997. Recruitment of p300/CBP in p53-dependent signal pathways. *Cell* 89(7):1175-84.
- Baek SH, Rosenfeld MG. 2004. Nuclear receptor coregulators: their modification codes and regulatory mechanism by translocation. *Biochem Biophys Res Commun* 319(3):707-14.
- Bain DL, Franden MA, McManaman JL, Takimoto GS, Horwitz KB. 2000. The N-terminal region of the human progesterone A-receptor. Structural analysis and the influence of the DNA binding domain. *J Biol Chem* 275(10):7313-20.
- Bain DL, Franden MA, McManaman JL, Takimoto GS, Horwitz KB. 2001. The N-terminal region of human progesterone B-receptors: biophysical and biochemical comparison to A-receptors. *J Biol Chem* 276(26):23825-31.
- Bain DL, Heneghan AF, Connaghan-Jones KD, Miura MT. 2007. Nuclear receptor structure: implications for function. *Annu Rev Physiol* 69:201-20.
- Baker KD, Shewchuk LM, Kozlova T, Makishima M, Hassell A, Wisely B, Caravella JA, Lambert MH, Reinking JL, Krause H and others. 2003. The Drosophila orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell* 113(6):731-42.
- Bastien J, Adam-Stitah S, Riedl T, Egly JM, Chambon P, Rochette-Egly C. 2000. TFIIF interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7. *J Biol Chem* 275(29):21896-904.
- Bastien J, Rochette-Egly C. 2004. Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* 328:1-16.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL and others. 2004. The Pfam protein families database. *Nucleic Acids Res* 32(Database issue):D138-41.
- Bates PA, Kelley LA, MacCallum RM, Sternberg MJ. 2001. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins Suppl* 5:39-46.
- Baumann H, Paulsen K, Kovacs H, Berglund H, Wright AP, Gustafsson JA, Hard T. 1993. Refined solution structure of the glucocorticoid receptor DNA-binding domain. *Biochemistry* 32(49):13463-71.

- Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS. 1995. Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A* 92(18):8259-63.
- Bernard P, Couturier M. 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* 226(3):735-45.
- Bernard P, Ludbrook L, Queipo G, Dinulos MB, Kletter GB, Zhang YH, Phelan JK, McCabe ER, Harley VR, Vilain E. 2006. A familial missense mutation in the hinge region of DAX1 associated with late-onset AHC in a prepubertal female. *Mol Genet Metab* 88(3):272-9.
- Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M. 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. *Mol Biol Evol* 21(10):1923-37.
- Bhattacharyya N, Dey A, Minucci S, Zimmer A, John S, Hager G, Ozato K. 1997. Retinoid-induced chromatin structure alterations in the retinoic acid receptor beta2 promoter. *Mol Cell Biol* 17(11):6481-90.
- Bianchetti L, Thompson JD, Lecompte O, Plewniak F, Poch O. 2005. vALId: validation of protein sequence quality based on multiple alignment data. *J Bioinform Comput Biol* 3(4):929-47.
- Birnboim HC, Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6):1513-23.
- Bledsoe RK, Madauss KP, Holt JA, Apolito CJ, Lambert MH, Pearce KH, Stanley TB, Stewart EL, Trump RP, Willson TM and others. 2005. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *J Biol Chem* 280(35):31283-93.
- Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, McKee DD, Consler TG, Parks DJ, Stewart EL, Willson TM and others. 2002. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* 110(1):93-105.
- Blondel A, Renaud JP, Fischer S, Moras D, Karplus M. 1999. Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. *J Mol Biol* 291(1):101-15.
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I and others. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31(1):365-70.
- Bogan AA, Dallas-Yang Q, Ruse MD, Jr., Maeda Y, Jiang G, Nepomuceno L, Scanlan TS, Cohen FE, Sladek FM. 2000. Analysis of protein dimerization and ligand binding of orphan receptor HNF4alpha. *J Mol Biol* 302(4):831-51.
- Borud B, Mellgren G, Lund J, Bakke M. 2003. Cloning and characterization of a novel zinc finger protein that modulates the transcriptional activity of nuclear receptors. *Mol Endocrinol* 17(11):2303-19.
- Bourguet W, Ruff M, Chambon P, Gronemeyer H, Moras D. 1995. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha. *Nature* 375(6530):377-82.
- Bourguet W, Vivat V, Wurtz JM, Chambon P, Gronemeyer H, Moras D. 2000. Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol Cell* 5(2):289-98.
- Bray D, Duke T. 2004. Conformational spread: the propagation of allosteric states in large multiprotein complexes. *Annu Rev Biophys Biomol Struct* 33:53-73.

- Brelivet Y, Kammerer S, Rochel N, Poch O, Moras D. 2004. Signature of the oligomeric behaviour of nuclear receptors at the sequence and structural level. *EMBO Rep* 5(4):423-9.
- Brooks B, Karplus M. 1985. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci U S A* 82(15):4995-9.
- Brown RS, Sander C, Argos P. 1985. The primary structure of transcription factor TFIID has 12 consecutive repeats. *FEBS Lett* 186(2):271-4.
- Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M. 1997. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389(6652):753-8.
- Burley SK, Petsko GA. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 229(4708):23-8.
- Bushman W, Thompson JF, Vargas L, Landy A. 1985. Control of directionality in lambda site specific recombination. *Science* 230(4728):906-11.
- Busso D, Delagoutte-Busso B, Moras D. 2005. Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal Biochem* 343(2):313-21.
- Butler BA. 1998. Sequence analysis using GCG. *Methods Biochem Anal* 39:74-97.
- Cassuto E, Lash T, Sriprakash KS, Radding CM. 1971. Role of exonuclease and protein of phage lambda in genetic recombination. V. Recombination of lambda DNA in vitro. *Proc Natl Acad Sci U S A* 68(7):1639-43.
- Cavailles V, Dauvois S, L'Horset F, Lopez G, Hoare S, Kushner PJ, Parker MG. 1995. Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *Embo J* 14(15):3741-51.
- Chakravarti D, LaMorte VJ, Nelson MC, Nakajima T, Schulman IG, Juguilon H, Montminy M, Evans RM. 1996. Role of CBP/P300 in nuclear receptor signalling. *Nature* 383(6595):99-103.
- Chan LS, Wells RA. 2007. Manipulation of reciprocal salt bridges at the heterodimerization interface alters the dimerization properties of mouse RXRalpha and PPARgamma1. *Biochem Biophys Res Commun* 358(4):1080-5.
- Chen CP, Kernytsky A, Rost B. 2002. Transmembrane helix predictions revisited. *Protein Sci* 11(12):2774-91.
- Chen D, Riedl T, Washbrook E, Pace PE, Coombes RC, Egly JM, Ali S. 2000. Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIID and participation of CDK7. *Mol Cell* 6(1):127-37.
- Chen H, Lin RJ, Schiltz RL, Chakravarti D, Nash A, Nagy L, Privalsky ML, Nakatani Y, Evans RM. 1997. Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell* 90(3):569-80.
- Chen J, Kim J, Dalton JT. 2005. Discovery and therapeutic promise of selective androgen receptor modulators. *Mol Interv* 5(3):173-88.
- Chen JD, Evans RM. 1995. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* 377(6548):454-7.
- Chou PY, Fasman GD. 1974. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* 13(2):211-22.
- Combet C, Jambon M, Deleage G, Geourjon C. 2002. Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* 18(1):213-4.
- Copley RR, Letunic I, Bork P. 2002. Genome and protein evolution in eukaryotes. *Curr Opin Chem Biol* 6(1):39-45.

- Couette B, Jalaguier S, Hellal-Levy C, Lupo B, Fagart J, Auzou G, Rafestin-Oblin ME. 1998. Folding requirements of the ligand-binding domain of the human mineralocorticoid receptor. *Mol Endocrinol* 12(6):855-63.
- Czar MJ, Owens-Grillo JK, Yem AW, Leach KL, Deibel MR, Jr., Welsh MJ, Pratt WB. 1994. The hsp56 immunophilin component of untransformed steroid receptor complexes is localized both to microtubules in the cytoplasm and to the same nonrandom regions within the nucleus as the steroid receptor. *Mol Endocrinol* 8(12):1731-41.
- Dagert M, Ehrlich SD. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* 6(1):23-8.
- Dahlman-Wright K, McEwan IJ. 1996. Structural studies of mutant glucocorticoid receptor transactivation domains establish a link between transactivation activity in vivo and alpha-helix-forming potential in vitro. *Biochemistry* 35(4):1323-7.
- Dalman FC, Scherrer LC, Taylor LP, Akil H, Pratt WB. 1991. Localization of the 90-kDa heat shock protein-binding site within the hormone-binding domain of the glucocorticoid receptor by peptide competition. *J Biol Chem* 266(6):3482-90.
- Darimont BD, Wagner RL, Apriletti JW, Stallcup MR, Kushner PJ, Baxter JD, Fletterick RJ, Yamamoto KR. 1998. Structure and specificity of nuclear receptor-coactivator interactions. *Genes Dev* 12(21):3343-56.
- DeFranco DB. 1999. Regulation of steroid receptor subcellular trafficking. *Cell Biochem Biophys* 30(1):1-24.
- Deleage G, Combet C, Blanchet C, Geourjon C. 2001. ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities. *Comput Biol Med* 31(4):259-67.
- Deleage G, Roux B. 1987. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng* 1(4):289-94.
- Detera-Wadleigh SD, Fanning TG. 1994. Phylogeny of the steroid receptor superfamily. *Mol Phylogenet Evol* 3(3):192-205.
- Devarakonda S, Harp JM, Kim Y, Ozyhar A, Rastinejad F. 2003. Structure of the heterodimeric ecdysone receptor DNA-binding complex. *Embo J* 22(21):5827-40.
- Ding XF, Anderson CM, Ma H, Hong H, Uht RM, Kushner PJ, Stallcup MR. 1998. Nuclear receptor-binding sites of coactivators glucocorticoid receptor interacting protein 1 (GRIP1) and steroid receptor coactivator 1 (SRC-1): multiple motifs with different binding specificities. *Mol Endocrinol* 12(2):302-13.
- Dittmar KD, Demady DR, Stancato LF, Krishna P, Pratt WB. 1997. Folding of the glucocorticoid receptor by the heat shock protein (hsp) 90-based chaperone machinery. The role of p23 is to stabilize receptor.hsp90 heterocomplexes formed by hsp90.p60.hsp70. *J Biol Chem* 272(34):21213-20.
- Dressel U, Thormeyer D, Altincicek B, Paululat A, Eggert M, Schneider S, Tenbaum SP, Renkawitz R, Baniahmad A. 1999. Alien, a highly conserved protein with characteristics of a corepressor for members of the nuclear hormone receptor superfamily. *Mol Cell Biol* 19(5):3383-94.
- Dunker AK, Obradovic Z. 2001. The protein trinity--linking function and disorder. *Nat Biotechnol* 19(9):805-6.
- Edgar RC, Batzoglou S. 2006. Multiple sequence alignment. *Curr Opin Struct Biol* 16(3):368-73.
- Egea PF, Mitschler A, Moras D. 2002. Molecular recognition of agonist ligands by RXRs. *Mol Endocrinol* 16(5):987-97.
- Egea PF, Mitschler A, Rochel N, Ruff M, Chambon P, Moras D. 2000. Crystal structure of the human RXRalpha ligand-binding domain bound to its natural ligand: 9-cis retinoic acid. *Embo J* 19(11):2592-601.

- Escriva H, Delaunay F, Laudet V. 2000. Ligand binding and nuclear receptor evolution. *Bioessays* 22(8):717-27.
- Escriva H, Langlois MC, Mendonca RL, Pierce R, Laudet V. 1998. Evolution and diversification of the nuclear receptor superfamily. *Ann N Y Acad Sci* 839:143-6.
- Escriva H, Safi R, Hanni C, Langlois MC, Saumitou-Laprade P, Stehelin D, Capron A, Pierce R, Laudet V. 1997. Ligand binding was acquired during evolution of nuclear receptors. *Proc Natl Acad Sci U S A* 94(13):6803-8.
- Etzold T, Argos P. 1993. SRS--an indexing and retrieval tool for flat file data libraries. *Comput Appl Biosci* 9(1):49-57.
- Fernandes I, Bastien Y, Wai T, Nygard K, Lin R, Cormier O, Lee HS, Eng F, Bertos NR, Pelletier N and others. 2003. Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. *Mol Cell* 11(1):139-50.
- Fitch WM. 1970. Distinguishing homologous from analogous proteins. *Syst Zool* 19(2):99-113.
- Frankel AD, Kim PS. 1991. Modular structure of transcription factors: implications for gene regulation. *Cell* 65(5):717-9.
- Freedman LP, Luisi BF, Korszun ZR, Basavappa R, Sigler PB, Yamamoto KR. 1988a. The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. *Nature* 334(6182):543-6.
- Freedman LP, Yamamoto KR, Luisi BF, Sigler PB. 1988b. More fingers in hand. *Cell* 54(4):444.
- Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. *Proteins* 23(4):566-79.
- Fronsdal K, Engedal N, Slagsvold T, Saatcioglu F. 1998. CREB binding protein is a coactivator for the androgen receptor and mediates cross-talk with AP-1. *J Biol Chem* 273(48):31853-9.
- Fu M, Rao M, Wang C, Sakamaki T, Wang J, Di Vizio D, Zhang X, Albanese C, Balk S, Chang C and others. 2003. Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Mol Cell Biol* 23(23):8563-75.
- Fu M, Wang C, Reutens AT, Wang J, Angeletti RH, Siconolfi-Baez L, Ogryzko V, Avantiaggiati ML, Pestell RG. 2000. p300 and p300/cAMP-response element-binding protein-associated factor acetylate the androgen receptor at sites governing hormone-dependent transactivation. *J Biol Chem* 275(27):20853-60.
- Gaillard E, Bruck N, Brelivet Y, Bour G, Lalevee S, Bauer A, Poch O, Moras D, Rochette-Egly C. 2006. Phosphorylation by PKA potentiates retinoic acid receptor alpha activity by means of increasing interaction with and phosphorylation by cyclin H/cdk7. *Proc Natl Acad Sci U S A* 103(25):9548-53.
- Gampe RT, Jr., Montana VG, Lambert MH, Miller AB, Bledsoe RK, Milburn MV, Kliewer SA, Willson TM, Xu HE. 2000. Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol Cell* 5(3):545-55.
- Gangloff M, Ruff M, Eiler S, Duclaud S, Wurtz JM, Moras D. 2001. Crystal structure of a mutant hERalpha ligand-binding domain reveals key structural features for the mechanism of partial agonism. *J Biol Chem* 276(18):15059-65.
- Gearhart MD, Holmbeck SM, Evans RM, Dyson HJ, Wright PE. 2003. Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J Mol Biol* 327(4):819-32.
- Germain P, Staels B, Dacquet C, Spedding M, Laudet V. 2006. Overview of nomenclature of nuclear receptors. *Pharmacol Rev* 58(4):685-704.

- Giguere V. 1999. Orphan nuclear receptors: from gene to function. *Endocr Rev* 20(5):689-725.
- Giguere V, Tini M, Flock G, Ong E, Evans RM, Otulakowski G. 1994. Isoform-specific amino-terminal domains dictate DNA-binding properties of ROR alpha, a novel family of orphan hormone nuclear receptors. *Genes Dev* 8(5):538-53.
- Glaros S, Atanaskova N, Zhao C, Skafar DF, Reddy KB. 2006. Activation function-1 domain of estrogen receptor regulates the agonistic and antagonistic actions of tamoxifen. *Mol Endocrinol* 20(5):996-1008.
- Glass CK. 1994. Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr Rev* 15(3):391-407.
- Glass CK, Rosenfeld MG. 2000. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* 14(2):121-41.
- Goodsell DS, Olson AJ. 2000. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29:105-53.
- Grad I, Picard D. 2007. The glucocorticoid responses are shaped by molecular chaperones. *Mol Cell Endocrinol* 275(1-2):2-12.
- Greschik H, Wurtz JM, Sanglier S, Bourguet W, van Dorsselaer A, Moras D, Renaud JP. 2002. Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3. *Mol Cell* 9(2):303-13.
- Gronemeyer H, Laudet V. 1995. Transcription factors 3: nuclear receptors. *Protein Profile* 2(11):1173-308.
- Gronemeyer H, Moras D. 1995. Nuclear receptors. How to finger DNA. *Nature* 375(6528):190-1.
- Gu W, Malik S, Ito M, Yuan CX, Fondell JD, Zhang X, Martinez E, Qin J, Roeder RG. 1999. A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol Cell* 3(1):97-108.
- Haelens A, Tanner T, Denayer S, Callewaert L, Claessens F. 2007. The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor. *Cancer Res* 67(9):4514-23.
- Hager GL, Elbi C, Johnson TA, Voss T, Nagaich AK, Schiltz RL, Qiu Y, John S. 2006. Chromatin dynamics and the evolution of alternate promoter states. *Chromosome Res* 14(1):107-16.
- Hartley JL, Temple GF, Brasch MA. 2000. DNA cloning using in vitro site-specific recombination. *Genome Res* 10(11):1788-95.
- He B, Gampe RT, Jr., Hnat AT, Faggart JL, Minges JT, French FS, Wilson EM. 2006. Probing the functional link between androgen receptor coactivator and ligand-binding sites in prostate cancer and androgen insensitivity. *J Biol Chem* 281(10):6648-63.
- He B, Wilson EM. 2002. The NH(2)-terminal and carboxyl-terminal interaction in the human androgen receptor. *Mol Genet Metab* 75(4):293-8.
- Heery DM, Kalkhoven E, Hoare S, Parker MG. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387(6634):733-6.
- Heringa J. 2000. Computational methods for protein secondary structure prediction using multiple sequence alignments. *Curr Protein Pept Sci* 1(3):273-301.
- Holmbeck SM, Foster MP, Casimiro DR, Sem DS, Dyson HJ, Wright PE. 1998. High-resolution solution structure of the retinoid X receptor DNA-binding domain. *J Mol Biol* 281(2):271-84.
- Horlein AJ, Naar AM, Heinzl T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamei Y, Soderstrom M, Glass CK and others. 1995. Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377(6548):397-404.

- Horwitz KB, Jackson TA, Bain DL, Richer JK, Takimoto GS, Tung L. 1996. Nuclear receptor coactivators and corepressors. *Mol Endocrinol* 10(10):1167-77.
- Hu X, Lazar MA. 1999. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* 402(6757):93-6.
- Huber R. 1979. Conformational flexibility in protein molecules. *Nature* 280(5723):538-9.
- Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ. 2006. The PROSITE database. *Nucleic Acids Res* 34(Database issue):D227-30.
- Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323(3):573-84.
- Inoue H, Nojima H, Okayama H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96(1):23-8.
- Iordanidou P, Aggelidou E, Demetriades C, Hadzopoulou-Cladaras M. 2005. Distinct amino acid residues may be involved in coactivator and ligand interactions in hepatocyte nuclear factor-4alpha. *J Biol Chem* 280(23):21810-9.
- Jakob M, Kolodziejczyk R, Orlowski M, Krzywda S, Kowalska A, Dutko-Gwozdz J, Gwozdz T, Kochman M, Jaskolski M, Ozyhar A. 2007. Novel DNA-binding element within the C-terminal extension of the nuclear receptor DNA-binding domain. *Nucleic Acids Res* 35(8):2705-18.
- Jepsen K, Rosenfeld MG. 2002. Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci* 115(Pt 4):689-98.
- Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292(2):195-202.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47 (Pt 2):110-9.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577-637.
- Kadonaga JT. 1998. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92(3):307-13.
- Kalkhoven E, Valentine JE, Heery DM, Parker MG. 1998. Isoforms of steroid receptor co-activator 1 differ in their ability to potentiate transcription by the oestrogen receptor. *Embo J* 17(1):232-43.
- Kamei Y, Xu L, Heinzel T, Torchia J, Kurokawa R, Glass B, Lin SC, Heyman RA, Rose DW, Glass CK and others. 1996. A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* 85(3):403-14.
- Kang KI, Meng X, Devin-Leclerc J, Bouhouche I, Chadli A, Cadepond F, Baulieu EE, Catelli MG. 1999. The molecular chaperone Hsp90 can negatively regulate the activity of a glucocorticosteroid-dependent promoter. *Proc Natl Acad Sci U S A* 96(4):1439-44.
- Kastner P, Mark M, Chambon P. 1995. Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life? *Cell* 83(6):859-69.
- Kern D, Zuiderweg ER. 2003. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13(6):748-57.
- Klug A, Rhodes D. 1987. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb Symp Quant Biol* 52:473-82.
- Koide A, Zhao C, Naganuma M, Abrams J, Deighton-Collins S, Skafar DF, Koide S. 2007. Identification of regions within the F domain of the human estrogen receptor alpha that are important for modulating transactivation and protein-protein interactions. *Mol Endocrinol* 21(4):829-42.

- Kosano H, Stensgard B, Charlesworth MC, McMahon N, Toft D. 1998. The assembly of progesterone receptor-hsp90 complexes using purified proteins. *J Biol Chem* 273(49):32973-9.
- Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM. 2006. The RCSB PDB information portal for structural genomics. *Nucleic Acids Res* 34(Database issue):D302-5.
- Krstic MD, Rogatsky I, Yamamoto KR, Garabedian MJ. 1997. Mitogen-activated and cyclin-dependent protein kinases selectively and differentially modulate transcriptional enhancement by the glucocorticoid receptor. *Mol Cell Biol* 17(7):3947-54.
- Krust A, Green S, Argos P, Kumar V, Walter P, Bornert JM, Chambon P. 1986. The chicken oestrogen receptor sequence: homology with v-erbA and the human oestrogen and glucocorticoid receptors. *Embo J* 5(5):891-7.
- Kumar R, Thompson EB. 2003. Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Mol Endocrinol* 17(1):1-10.
- Kurokawa R, Yu VC, Naar A, Kyakumoto S, Han Z, Silverman S, Rosenfeld MG, Glass CK. 1993. Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. *Genes Dev* 7(7B):1423-35.
- Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274(5289):948-53.
- Landy A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* 58:913-49.
- Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97(1):17-27.
- Laudet V. 1997. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J Mol Endocrinol* 19(3):207-26.
- Le Drean Y, Mincheneau N, Le Goff P, Michel D. 2002. Potentiation of glucocorticoid receptor transcriptional activity by sumoylation. *Endocrinology* 143(9):3482-9.
- Lecompte O, Thompson JD, Plewniak F, Thierry J, Poch O. 2001. Multiple alignment of complete sequences (MACS) in the post-genomic era. *Gene* 270(1-2):17-30.
- Lee S, Lee BC, Kim D. 2006. Prediction of protein secondary structure content using amino acid composition and evolutionary information. *Proteins* 62(4):1107-14.
- Lee SK, Na SY, Kim HJ, Soh J, Choi HS, Lee JW. 1998. Identification of critical residues for heterodimerization within the ligand-binding domain of retinoid X receptor. *Mol Endocrinol* 12(3):325-32.
- Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. 2006. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 34(Database issue):D257-60.
- Li X, Obradovic Z, Brown CJ, Garner EC, Dunker AK. 2000. Comparing predictors of disordered protein. *Genome Inform Ser Workshop Genome Inform* 11:172-84.
- Li X, Romero P, Rani M, Dunker AK, Obradovic Z. 1999. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* 10:30-40.
- Li Y, Lambert MH, Xu HE. 2003. Activation of nuclear receptors: a perspective from structural genomics. *Structure* 11(7):741-6.
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. 2003a. Protein disorder prediction: implications for structural proteomics. *Structure (Camb)* 11(11):1453-9.
- Linding R, Russell RB, Neduva V, Gibson TJ. 2003b. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 31(13):3701-8.

- Little TH, Zhang Y, Matulis CK, Weck J, Zhang Z, Ramachandran A, Mayo KE, Radhakrishnan I. 2006. Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol Endocrinol* 20(4):831-43.
- Liu J, Rost B. 2003. NORSp: Predictions of long regions without regular secondary structure. *Nucleic Acids Res* 31(13):3833-5.
- Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352(6335):497-505.
- Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* 252(5009):1162-1164.
- Luque I, Leavitt SA, Freire E. 2002. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu Rev Biophys Biomol Struct* 31:235-56.
- Mader S, Chen JY, Chen Z, White J, Chambon P, Gronemeyer H. 1993. The patterns of binding of RAR, RXR and TR homo- and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains. *Embo J* 12(13):5029-41.
- Mak HY, Hoare S, Henttu PM, Parker MG. 1999. Molecular determinants of the estrogen receptor-coactivator interface. *Mol Cell Biol* 19(5):3895-903.
- Mallatt J, Winchell CJ. 2002. Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. *Mol Biol Evol* 19(3):289-301.
- Mangelsdorf DJ, Evans RM. 1995. The RXR heterodimers and orphan receptors. *Cell* 83(6):841-50.
- Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P and others. 1995. The nuclear receptor superfamily: the second decade. *Cell* 83(6):835-9.
- Marimuthu A, Feng W, Tagami T, Nguyen H, Jameson JL, Fletterick RJ, Baxter JD, West BL. 2002. TR surfaces and conformations required to bind nuclear receptor corepressor. *Mol Endocrinol* 16(2):271-86.
- McEwan IJ, Lavery D, Fischer K, Watt K. 2007. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl Recept Signal* 5:e001.
- McKenna NJ, Lanz RB, O'Malley BW. 1999a. Nuclear receptor coregulators: cellular and molecular biology. *Endocr Rev* 20(3):321-44.
- McKenna NJ, Xu J, Nawaz Z, Tsai SY, Tsai MJ, O'Malley BW. 1999b. Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. *J Steroid Biochem Mol Biol* 69(1-6):3-12.
- Meinke G, Sigler PB. 1999. DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B. *Nat Struct Biol* 6(5):471-7.
- Michigami T, Suga A, Yamazaki M, Shimizu C, Cai G, Okada S, Ozono K. 1999. Identification of amino acid sequence in the hinge region of human vitamin D receptor that transfers a cytosolic protein to the nucleus. *J Biol Chem* 274(47):33531-8.
- Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *Embo J* 4(6):1609-14.
- Monod J, Wyman J, Changeux JP. 1965. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* 12:88-118.
- Montano MM, Ekena K, Delage-Mourroux R, Chang W, Martini P, Katzenellenbogen BS. 1999. An estrogen receptor-selective coregulator that potentiates the effectiveness of

- antiestrogens and represses the activity of estrogens. *Proc Natl Acad Sci U S A* 96(12):6947-52.
- Montano MM, Muller V, Trobaugh A, Katzenellenbogen BS. 1995. The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. *Mol Endocrinol* 9(7):814-25.
- Moraitis AN, Giguere V, Thompson CC. 2002. Novel mechanism of nuclear receptor corepressor interaction dictated by activation function 2 helix determinants. *Mol Cell Biol* 22(19):6831-41.
- Moras D, Gronemeyer H. 1998. The nuclear receptor ligand-binding domain: structure and function. *Curr Opin Cell Biol* 10(3):384-91.
- Moult J. 2005. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* 15(3):285-9.
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R and others. 2007. New developments in the InterPro database. *Nucleic Acids Res* 35(Database issue):D224-8.
- Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-55.
- Nagy L, Kao HY, Love JD, Li C, Banayo E, Gooch JT, Krishna V, Chatterjee K, Evans RM, Schwabe JW. 1999. Mechanism of corepressor binding and release from nuclear hormone receptors. *Genes Dev* 13(24):3209-16.
- Nathan DF, Lindquist S. 1995. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol Cell Biol* 15(7):3917-25.
- Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W and others. 1987. The P450 gene superfamily: recommended nomenclature. *DNA* 6(1):1-11.
- Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48(3):443-53.
- Nichols M, Rientjes JM, Stewart AF. 1998. Different positioning of the ligand-binding domain helix 12 and the F domain of the estrogen receptor accounts for functional differences between agonists and antagonists. *Embo J* 17(3):765-73.
- Nishida T, Yasuda H. 2002. PIAS1 and PIASx α function as SUMO-E3 ligases toward androgen receptor and repress androgen receptor-dependent transcription. *J Biol Chem* 277(44):41311-7.
- Oldfield CJ, Ulrich EL, Cheng Y, Dunker AK, Markley JL. 2005. Addressing the intrinsic disorder bottleneck in structural proteomics. *Proteins* 59(3):444-53.
- Oro AE, McKeown M, Evans RM. 1990. Relationship between the product of the *Drosophila* ultraspiracle locus and the vertebrate retinoid X receptor. *Nature* 347(6290):298-301.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96(8):4285-8.
- Perissi V, Staszewski LM, McInerney EM, Kurokawa R, Kronen A, Rose DW, Lambert MH, Milburn MV, Glass CK, Rosenfeld MG. 1999. Molecular determinants of nuclear receptor-corepressor interaction. *Genes Dev* 13(24):3198-208.
- Perkins ND, Felzien LK, Betts JC, Leung K, Beach DH, Nabel GJ. 1997. Regulation of NF- κ B by cyclin-dependent kinases associated with the p300 coactivator. *Science* 275(5299):523-7.
- Perlmann T, Rangarajan PN, Umesono K, Evans RM. 1993. Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes Dev* 7(7B):1411-22.

- Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. 2001. ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr* 57(Pt 10):1445-50.
- Perrakis A, Morris R, Lamzin VS. 1999. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6(5):458-63.
- Picard D. 2006. Chaperoning steroid hormone action. *Trends Endocrinol Metab* 17(6):229-35.
- Picard D, Khursheed B, Garabedian MJ, Fortin MG, Lindquist S, Yamamoto KR. 1990. Reduced levels of hsp90 compromise steroid receptor action in vivo. *Nature* 348(6297):166-8.
- Plewniak F, Bianchetti L, Brelivet Y, Carles A, Chalmel F, Lecompte O, Mochel T, Moulinier L, Muller A, Muller J and others. 2003. PipeAlign: A new toolkit for protein family analysis. *Nucleic Acids Res* 31(13):3829-32.
- Plewniak F, Thompson JD, Poch O. 2000. Ballast: blast post-processing based on locally conserved segments. *Bioinformatics* 16(9):750-9.
- Pratt WB, Toft DO. 1997. Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev* 18(3):306-60.
- Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann JS, Silman I, Sussman JL. 2005. FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*.
- Puigserver P, Wu Z, Park CW, Graves R, Wright M, Spiegelman BM. 1998. A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* 92(6):829-39.
- Rachez C, Suldan Z, Ward J, Chang CP, Burakov D, Erdjument-Bromage H, Tempst P, Freedman LP. 1998. A novel protein complex that interacts with the vitamin D3 receptor in a ligand-dependent manner and enhances VDR transactivation in a cell-free system. *Genes Dev* 12(12):1787-800.
- Rastinejad F, Perlmann T, Evans RM, Sigler PB. 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* 375(6528):203-11.
- Rastinejad F, Wagner T, Zhao Q, Khorasanizadeh S. 2000. Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1. *Embo J* 19(5):1045-54.
- Renaud JP, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H, Moras D. 1995. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* 378(6558):681-9.
- Ribeiro RC, Kushner PJ, Baxter JD. 1995. The nuclear hormone receptor gene superfamily. *Annu Rev Med* 46:443-53.
- Robinson-Rechavi M, Carpentier AS, Duffraisse M, Laudet V. 2001. How many nuclear hormone receptors are there in the human genome? *Trends Genet* 17(10):554-6.
- Rochette-Egly C. 2003. Nuclear receptors: integration of multiple signalling pathways through phosphorylation. *Cell Signal* 15(4):355-66.
- Rochette-Egly C, Adam S, Rossignol M, Egly JM, Chambon P. 1997. Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIID and phosphorylation by CDK7. *Cell* 90(1):97-107.
- Rochette-Egly C, Oulad-Abdelghani M, Staub A, Pfister V, Scheuer I, Chambon P, Gaub MP. 1995. Phosphorylation of the retinoic acid receptor-alpha by protein kinase A. *Mol Endocrinol* 9(7):860-71.
- Roemer SC, Donham DC, Sherman L, Pon VH, Edwards DP, Churchill ME. 2006. Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. *Mol Endocrinol* 20(12):3042-52.
- Romero, Obradovic, Dunker K. 1997. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Inform Ser Workshop Genome Inform* 8:110-124.

- Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S, Dunker AK. 1998. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput*:437-48.
- Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. 2001. Sequence complexity of disordered protein. *Proteins* 42(1):38-48.
- Rost B. 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* 266:525-39.
- Rost B. 2001. Review: protein secondary structure prediction continues to rise. *J Struct Biol* 134(2-3):204-18.
- Ruse MD, Jr., Privalsky ML, Sladek FM. 2002. Competitive cofactor recruitment by orphan receptor hepatocyte nuclear factor 4 α 1: modulation by the F domain. *Mol Cell Biol* 22(6):1626-38.
- Sabbah M, Radanyi C, Redeuilh G, Baulieu EE. 1996. The 90 kDa heat-shock protein (hsp90) modulates the binding of the oestrogen receptor to its cognate DNA. *Biochem J* 314 (Pt 1):205-13.
- Sack JS, Kish KF, Wang C, Attar RM, Kiefer SE, An Y, Wu GY, Scheffler JE, Salvati ME, Krystek SR, Jr. and others. 2001. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci U S A* 98(9):4904-9.
- Salamov AA, Solovyev VV. 1997. Protein secondary structure prediction using local alignments. *J Mol Biol* 268(1):31-6.
- Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins* 23(3):318-26.
- Schmitz ML, dos Santos Silva MA, Altmann H, Czisch M, Holak TA, Baeuerle PA. 1994. Structural and functional analysis of the NF-kappa B p50 C terminus. An acidic and modular transactivation domain with the potential to adopt an alpha-helical conformation. *J Biol Chem* 269(41):25613-20.
- Schwabe JW, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75(3):567-78.
- Schwabe JW, Chapman L, Rhodes D. 1995. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure* 3(2):201-13.
- Schwartz Z, Sylvia VL, Guinee T, Dean DD, Boyan BD. 2002. Tamoxifen elicits its anti-estrogen effects in growth plate chondrocytes by inhibiting protein kinase C. *J Steroid Biochem Mol Biol* 80(4-5):401-10.
- Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31(13):3381-5.
- Sentis S, Le Romancer M, Bianchin C, Rostan MC, Corbo L. 2005. Sumoylation of the estrogen receptor alpha hinge region regulates its transcriptional activity. *Mol Endocrinol* 19(11):2671-84.
- Shaffer PL, Gewirth DT. 2002. Structural basis of VDR-DNA interactions on direct repeat response elements. *Embo J* 21(9):2242-52.
- Shaffer PL, Gewirth DT. 2004. Structural analysis of RXR-VDR interactions on DR3 DNA. *J Steroid Biochem Mol Biol* 89-90(1-5):215-9.
- Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL. 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95(7):927-37.
- Sierk ML, Zhao Q, Rastinejad F. 2001. DNA deformability as a recognition feature in the reverb response element. *Biochemistry* 40(43):12833-43.

- Smith DF, Whitesell L, Nair SC, Chen S, Prapapanich V, Rimerman RA. 1995. Progesterone receptor structure and function altered by geldanamycin, an hsp90-binding agent. *Mol Cell Biol* 15(12):6804-12.
- Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G. 2003. Improved amino acid flexibility parameters. *Protein Sci* 12(5):1060-72.
- Solomon C, Macoritto M, Gao XL, White JH, Kremer R. 2001. The unique tryptophan residue of the vitamin D receptor is critical for ligand binding and transcriptional activation. *J Bone Miner Res* 16(1):39-45.
- Solomon IH, Hager JM, Safi R, McDonnell DP, Redinbo MR, Ortlund EA. 2005. Crystal structure of the human LRH-1 DBD-DNA complex reveals Ftz-F1 domain positioning is required for receptor activity. *J Mol Biol* 354(5):1091-102.
- Sourjik V. 2004. Receptor clustering and signal processing in *E. coli* chemotaxis. *Trends Microbiol* 12(12):569-76.
- Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189(1):113-30.
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60-89.
- Suel GM, Lockless SW, Wall MA, Ranganathan R. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10(1):59-69.
- Suen CS, Berrodin TJ, Mastroeni R, Cheskis BJ, Lyttle CR, Frail DE. 1998. A transcriptional coactivator, steroid receptor coactivator-3, selectively augments steroid receptor transcriptional activity. *J Biol Chem* 273(42):27645-53.
- Svensson S, Ostberg T, Jacobsson M, Norstrom C, Stefansson K, Hallen D, Johansson IC, Zachrisson K, Ogg D, Jendeberg L. 2003. Crystal structure of the heterodimeric complex of LXRalpha and RXRbeta ligand-binding domains in a fully agonistic conformation. *Embo J* 22(18):4625-33.
- Tahiri B, Auzou G, Nicolas JC, Sultan C, Lupo B. 2001. Participation of critical residues from the extreme C-terminal end of the human androgen receptor in the ligand binding function. *Biochemistry* 40(29):8431-7.
- Taylor WR. 1986. The classification of amino acid conservation. *J Theor Biol* 119(2):205-18.
- Tcherepanova I, Puigserver P, Norris JD, Spiegelman BM, McDonnell DP. 2000. Modulation of estrogen receptor-alpha transcriptional activity by the coactivator PGC-1. *J Biol Chem* 275(21):16302-8.
- Thomas M, Dadgar N, Aphale A, Harrell JM, Kunkel R, Pratt WB, Lieberman AP. 2004. Androgen receptor acetylation site mutations cause trafficking defects, misfolding, and aggregation similar to expanded glutamine tracts. *J Biol Chem* 279(9):8389-95.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-80.
- Thompson JD, Holbrook SR, Katoh K, Koehl P, Moras D, Westhof E, Poch O. 2005. MAO: a Multiple Alignment Ontology for nucleic acid and protein sequences. *Nucleic Acids Res* 33(13):4164-71.
- Thompson JD, Muller A, Waterhouse A, Procter J, Barton GJ, Plewniak F, Poch O. 2006. MACSIMS: multiple alignment of complete sequences information management system. *BMC Bioinformatics* 7:318.
- Thompson JD, Plewniak F, Poch O. 1999. A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res* 27(13):2682-90.
- Thompson JD, Plewniak F, Ripp R, Thierry JC, Poch O. 2001. Towards a reliable objective function for multiple sequence alignments. *J Mol Biol* 314(4):937-51.

- Thompson JD, Plewniak F, Thierry J, Poch O. 2000. DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res* 28(15):2919-26.
- Thompson JD, Prigent V, Poch O. 2004. LEON: multiple aLignment Evaluation Of Neighbours. *Nucleic Acids Res* 32(4):1298-307.
- Thompson JD, Thierry JC, Poch O. 2003. RASCAL: rapid scanning and correction of multiple sequence alignments. *Bioinformatics* 19(9):1155-61.
- Thomson R, Hodgman TC, Yang ZR, Doyle AK. 2003. Characterizing proteolytic cleavage site activity using bio-basis function neural networks. *Bioinformatics* 19(14):1741-7.
- Thornton JW, Need E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 301(5640):1714-7.
- Tocchini-Valentini G, Rochel N, Wurtz JM, Moras D. 2004. Crystal structures of the vitamin D nuclear receptor liganded with the vitamin D side chain analogues calcipotriol and seocalcitol, receptor agonists of clinical importance. Insights into a structural basis for the switching of calcipotriol to a receptor antagonist by further side chain modification. *J Med Chem* 47(8):1956-61.
- Tompa P. 2002. Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527-33.
- Tompa P, Csermely P. 2004. The role of structural disorder in the function of RNA and protein chaperones. *Faseb J* 18(11):1169-75.
- Torchia J, Rose DW, Inostroza J, Kamei Y, Westin S, Glass CK, Rosenfeld MG. 1997. The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* 387(6634):677-84.
- Tremblay A, Tremblay GB, Labrie F, Giguere V. 1999. Ligand-independent recruitment of SRC-1 to estrogen receptor beta through phosphorylation of activation function AF-1. *Mol Cell* 3(4):513-9.
- Trowbridge JM, Rogatsky I, Garabedian MJ. 1997. Regulation of estrogen receptor transcriptional enhancement by the cyclin A/Cdk2 complex. *Proc Natl Acad Sci U S A* 94(19):10132-7.
- Uversky VN. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11(4):739-56.
- Uversky VN, Gillespie JR, Fink AL. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41(3):415-27.
- Vivat-Hannah V, Bourguet W, Gottardis M, Gronemeyer H. 2003. Separation of retinoid X receptor homo- and heterodimerization functions. *Mol Cell Biol* 23(21):7678-88.
- Vivat V, Zechel C, Wurtz JM, Bourguet W, Kagechika H, Umemiya H, Shudo K, Moras D, Gronemeyer H, Chambon P. 1997. A mutation mimicking ligand-induced conformational change yields a constitutive RXR that senses allosteric effects in heterodimers. *Embo J* 16(18):5697-709.
- Voegel JJ, Heine MJ, Tini M, Vivat V, Chambon P, Gronemeyer H. 1998. The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *Embo J* 17(2):507-19.
- Vucetic S, Brown CJ, Dunker AK, Obradovic Z. 2003. Flavors of protein disorder. *Proteins* 52(4):573-84.
- Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, Cortese MS, Lawson JD, Brown CJ, Sikes JG and others. 2005. DisProt: a database of protein disorder. *Bioinformatics* 21(1):137-40.
- Wang C, Fu M, Angeletti RH, Siconolfi-Baez L, Reutens AT, Albanese C, Lisanti MP, Katzenellenbogen BS, Kato S, Hopp T and others. 2001. Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity. *J Biol Chem* 276(21):18375-83.

- Wang Z, Benoit G, Liu J, Prasad S, Aarnisalo P, Liu X, Xu H, Walker NP, Perlmann T. 2003. Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* 423(6939):555-60.
- Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. 2004a. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20(13):2138-9.
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004b. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337(3):635-45.
- Warnmark A, Treuter E, Wright AP, Gustafsson JA. 2003. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. *Mol Endocrinol* 17(10):1901-9.
- Warnmark A, Wikstrom A, Wright AP, Gustafsson JA, Hard T. 2001. The N-terminal regions of estrogen receptor alpha and beta are unstructured in vitro and show different TBP binding properties. *J Biol Chem* 276(49):45939-44.
- Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT, Jr. 1996. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35(43):13709-15.
- Weston AD, Chandraratna RA, Torchia J, Underhill TM. 2002. Requirement for RAR-mediated gene repression in skeletal progenitor differentiation. *J Cell Biol* 158(1):39-51.
- Whitesell L, Cook P. 1996. Stable and specific binding of heat shock protein 90 by geldanamycin disrupts glucocorticoid receptor function in intact cells. *Mol Endocrinol* 10(6):705-12.
- Wicker N, Dembele D, Raffelsberger W, Poch O. 2002. Density of points clustering, application to transcriptomic data analysis. *Nucleic Acids Res* 30(18):3992-4000.
- Wicker N, Perrin GR, Thierry JC, Poch O. 2001. Secator: a program for inferring protein subfamilies from phylogenetic trees. *Mol Biol Evol* 18(8):1435-41.
- Williams SP, Sigler PB. 1998. Atomic structure of progesterone complexed with its receptor. *Nature* 393(6683):392-6.
- Wolf E, Kim PS, Berger B. 1997. MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci* 6(6):1179-89.
- Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321-31.
- Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R and others. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* 34(Database issue):D187-91.
- Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, Gronemeyer H. 1996. A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* 3(2):206.
- Xu HE, Stanley TB, Montana VG, Lambert MH, Shearer BG, Cobb JE, McKee DD, Galardi CM, Plunket KD, Nolte RT and others. 2002. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* 415(6873):813-7.
- Yang AS, Honig B. 2000. An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J Mol Biol* 301(3):691-711.
- Yang ZR, Thomson R, McNeil P, Esnouf RM. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*.

- Zamir I, Dawson J, Lavinsky RM, Glass CK, Rosenfeld MG, Lazar MA. 1997. Cloning and characterization of a corepressor and potential component of the nuclear hormone receptor repression complex. *Proc Natl Acad Sci U S A* 94(26):14400-5.
- Zechel C, Shen XQ, Chen JY, Chen ZP, Chambon P, Gronemeyer H. 1994. The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *Embo J* 13(6):1425-33.
- Zhao Q, Chasse SA, Devarakonda S, Sierk ML, Ahvazi B, Rastinejad F. 2000. Structural basis of RXR-DNA interactions. *J Mol Biol* 296(2):509-20.
- Zhao Q, Khorasanizadeh S, Miyoshi Y, Lazar MA, Rastinejad F. 1998. Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell* 1(6):849-61.
- Zhu XG, Hanover JA, Hager GL, Cheng SY. 1998. Hormone-induced translocation of thyroid hormone receptors in living cells visualized using a receptor green fluorescent protein chimera. *J Biol Chem* 273(42):27058-63.

Les Récepteurs Nucléaires (RN) sont des facteurs de transcription ligand-dépendants contrôlant un grand nombre de processus biologiques au travers de la régulation de la transcription des gènes. Organisés en multidomains, la plupart des RN fonctionnent soit en homodimères soit en hétérodimères, le domaine de fixation du ligand (LBD) étant le contributeur majeur de la stabilité des différents dimères. Le sujet de ma thèse a consisté à établir les bases moléculaires de la discrimination entre ces deux états par deux approches complémentaires : la bioinformatique génomique et structurale couplée à des expériences de biologie moléculaire. L'approche bioinformatique a permis de scinder la super-famille des RN en deux classes reliées au comportement oligomérique. Tous les homodimères appartiennent à la classe I, dans laquelle des résidus différenciellement conservés forment deux ponts salins connectant l'interface de dimérisation à la surface de fixation des cofacteurs tandis que les hétérodimères de classe II présentent d'autres résidus spécifiques connectant l'interface de dimérisation à la poche de fixation du ligand. Ces données nous ont permis de proposer l'existence de deux voies de communication connectant les surfaces actives des RNs par l'intermédiaire de deux jeux de ponts salins, véritables « motifs spatiaux » spécifiques des classes et différenciellement conservés de *C. elegans* jusqu'à l'homme.

L'étude expérimentale des deux voies de communication s'est basée sur une série de mutations progressives des résidus spécifiques de classe, afin de transformer un RN de classe I (homodimère) en classe II (hétérodimère) et *vice versa*. Trois RNs tests ont été choisis pour leur comportement oligomérique différent (classe I : ERR γ et RXR α ; classe II : RAR α) et l'impact des mutations a été examiné par quatre approches expérimentales analysant la stabilité et le type de dimérisation, la capacité de transactivation et la structure. Ces études ont montré que toute perturbation des voies impliquant l'interface de dimérisation entraîne des modifications du comportement oligomérique. Ce résultat suggère que le contrôle de l'état d'oligomérisation et le positionnement efficace de l'interface de dimérisation sont indissociables d'une voie de communication fonctionnelle et renforce l'importance de ces voies dans les phénomènes allostériques intégrant reconnaissance du ligand, du RN partenaire et des cofacteurs. De plus, nos résultats suggèrent l'implication directe des voies dans la coordination entre les surfaces impliquées dans ces reconnaissances et dans l'intégration des différents signaux cellulaires aboutissant à la transactivation. Par ailleurs, la résolution à 1.4 Å de la structure d'un mutant de ERR γ a fourni des pistes pour l'analyse, au niveau moléculaire, du lien entre voies et dimérisation, et pour la recherche d'un ligand fonctionnel de ce RN jusqu'alors considéré comme orphelin.

Ainsi il semble que, pour intégrer en une seule protéine, les nombreux mécanismes de signalisation impliqués dans l'adaptation de la transcription aux variations des milieux internes et externes, les RNs ont développé des voies de communication inter- et intramoléculaires qui « connectent et informent » l'ensemble des surfaces actives et partenaires impliqués dans le système. Finalement, au regard des enseignements biologiques et méthodologiques obtenus, les RNs apparaissent comme un système modèle de choix pour aborder l'intégration des signaux cellulaires au sein de complexes macromoléculaires impliquant de multiples partenaires nucléiques, protéiques et moléculaires, une dynamique complexe d'édification, de modification et de dégradation et surtout, l'établissement d'une machinerie en accord avec les cycles cellulaires, hormonaux et de développement.

Nuclear receptors (NRs) are ligand-dependent transcription factors that control a large number of physiological events through the regulation of gene transcription. Organized in several domains, NRs function either as homodimers or as heterodimers, where the ligand binding domain (LBD) plays the major role for the stability of the different dimers. The aim of my PhD consisted in establishing the molecular basis that discriminates the two states using two complementary approaches: genomic and structural bioinformatics combined with experimental data from molecular biology. Bioinformatic approaches have allowed us to partition the superfamily of NRs in two distinct classes based on their oligomeric behavior. All homodimers belong to class I, where differentially conserved residues form two salt bridges connecting the dimerization interface to the cofactor binding site, whereas class II heterodimers are characterized by other specific residues connecting the dimerization interface to the ligand binding pocket. Based on these data, we propose the existence of two communication pathways connecting active surfaces of NRs through two salt bridges representing "spatial motifs" specific for these classes and differentially conserved from *C. elegans* to human.

The experimental approach used to study the two communication pathways was based on a progressive series of mutations of class specific amino acids, with the aim of transforming a class I NR (homodimer) to a class II NR (heterodimer) and *vice versa*. Three NRs were chosen based on their different oligomeric behavior (class I: ERR γ and RXR α ; class II: RAR α) and the impact of the mutations was studied by four experimental approaches to analyse the stability and type of dimerization, the transactivation capacity and the structure. These studies have shown that any perturbation of pathways involved in the dimerization interface caused modifications in the oligomeric behavior. This result suggests that controlling the oligomeric status and the correct positioning of the dimerization interface cannot be dissociated from a functional communication pathway and thus reinforces the importance of such pathways in allosteric phenomena in the context of ligand-, NR partner and cofactor recognition. Furthermore, our results suggest a direct implication of these pathways in the coordination between the surfaces concerned in such recognition and in the integration of different cellular signals leading to transactivation. Furthermore, resolving the structure of mutant ERR γ , at 1.4 Å, will allow us to study, at the molecular level, the link between communication pathways and dimerization, and to search for a functional ligand of this NR that has so far been considered an orphan receptor.

Thus, it seems that, to integrate in a single protein entity, the numerous signaling mechanisms implicated in adaptation of transcription to the variations of internal and external environments, NRs have developed inter- and intramolecular communication pathways that interact and pass on information between the active surfaces and partners involved in the system. Taken together, NRs appear to be an ideal model to study the integration of cellular signals within macromolecular complexes, implicating binding of nucleic acids, proteins and other molecular partners, modification, degradation and the establishment of machinery inherently involved in cellular, hormonal and developmental cycles.

