



HAL
open science

VKORC1 et résistance aux antivitamines K : étude par modélisation moléculaire

Nolan Chatron

► **To cite this version:**

Nolan Chatron. VKORC1 et résistance aux antivitamines K : étude par modélisation moléculaire. Biologie moléculaire. Université Paris Saclay (COmUE), 2017. Français. NNT : 2017SACLN008 . tel-01540438

HAL Id: tel-01540438

<https://theses.hal.science/tel-01540438>

Submitted on 25 Sep 2017

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

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

NNT : 2017SACLN008



THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'ECOLE NORMALE SUPERIEURE DE CACHAN
(ECOLE NORMALE SUPERIEURE PARIS-SACLAY)

ECOLE DOCTORALE N° 573
Ecole Doctorale INTERFACES, Paris-Saclay

Spécialité de doctorat : Biologie

Par

Monsieur NOLAN CHATRON

**VKORC1 et Résistance aux Anti-vitamines K :
Etude par Modélisation Moléculaire**

Thèse présentée et soutenue à l'ENS Cachan, le 10 mars 2017

Composition du Jury :

M ROGNAN, Didier	DR CNRS, UNISTRA, Illkirch	Président du jury
M de BREVERN, Alexandre	DR INSERM, DSIMB, Paris	Rapporteur
M HINSEN, Konrad	CR CNRS, CBM, Orléans	Rapporteur
Mme RAHUEL-CLERMONT, Sophie	CR CNRS, IMoPA, Nancy	Examinatrice
M TROUVE, Alain	PR des Universités, ENS Cachan	Examineur
Mme TCHERTANOV, Luba	DR CNRS, ENS Cachan	Directrice de thèse
Mme LATTARD, Virginie	MC VetAgroSup, Lyon	Co-directrice de thèse
M PANSU, Robert	DR CNRS, ENS Cachan	Invité

Remerciements

Je remercie très chaleureusement Messieurs Alexandre de Brevern et Konrad Hinsen qui m'ont fait l'honneur d'être rapporteurs de ce manuscrit de thèse. Leurs conseils et suggestions ont été très précieux pour l'amélioration du manuscrit.

Je remercie très sincèrement Madame Sophie Rahuel-Clermont ainsi que Messieurs Didier Rognan et Alain Trouvé pour avoir accepté d'être examinateurs de cette thèse.

Je remercie bien évidemment l'équipe qui m'a encadré durant ces trois années de thèse. Je remercie Madame Luba Tchertanov pour son précieux encadrement, sa grande disponibilité, ses rectifications et conseils toujours efficaces et millimétrés tout au long de cette thèse, mais aussi pour toutes les discussions (scientifiques et beaucoup moins scientifiques) que nous avons pu avoir durant ces trois années. Je remercie Madame Virginie Lattard pour ses conseils précieux, pour la confiance qu'elle m'a accordée et par la même occasion pour sa confiance en la modélisation, et pour sa disponibilité. Je remercie également Monsieur Etienne Benoit (Professeur, VetAgro Sup Lyon) pour l'enthousiasme permanent dont il a fait preuve envers mes travaux de thèse, pour sa capacité à adopter un point de vue alternatif et ainsi toujours proposer de nouvelles pistes d'étude. Enfin, je vous remercie tous les trois, du fond du cœur, pour votre patience, votre sympathie toujours sincère à mon égard et pour toutes les discussions que nous avons eues au cours de mes trois années de thèse.

Je remercie les membres du CMLA, notamment Monsieur Nicolas Vayatis, qui m'a fait l'honneur de m'accueillir au sein du laboratoire. Je remercie Véronique Almadovar, Micheline Brunetti, Alina Muller et Virginie Pauchont, pour leur vitesse à résoudre les problèmes administratifs (qui m'ont toujours rebuté et, je pense, me rebuteront toujours autant), pour leur gentillesse, leur bonne humeur et leur humour. Je remercie Christophe et Sébastien pour toute l'aide qu'ils m'ont apportée en informatique, pour leur sympathie et leur humour. Je remercie Atman pour sa sympathie, sa simplicité, sa bonne humeur et ses nombreux mots d'encouragement, ainsi que pour m'avoir fait découvrir et permis de travailler sur Shiva.

Je remercie particulièrement Messieurs Bernard Chalmond et Alain Trouvé, pour toutes les discussions constructives, pour l'intérêt qu'ils ont porté à la partie « biologie » de ce projet de thèse, et pour le développement de la méthode ConfigScan, un grand et sincère merci. Je remercie également Nicolas Pajor, qui nous a malheureusement quittés, et qui était la toute première personne du CMLA que j'ai rencontré à l'ENS Cachan, toujours de bonne humeur, le sourire aux lèvres, à blaguer. Je profite de cette thèse pour lui rendre hommage et le remercier encore pour ses bons mots, et pour l'installation de TopDyn. Je remercie également Benjamin qui a pris la relève, et grâce à qui nous pouvons toujours calculer sur TopDyn aujourd'hui (et encore désolé pour toutes les fois où nous avons saturé la mémoire avec nos simulations !).

Je remercie toute l'équipe de VetAgro à Lyon, où j'ai toujours été accueilli dans la joie et les sourires. Merci à Abdé, Arnaud, Audrey, Claire, Jean-Valérie, Joffrey, Marlène, pour toute votre sympathie, toutes nos discussions et les bons moments (trop courts) passés ensemble lorsque je suis venu à Lyon. Je pense n'avoir oublié personne, mais sachez que je vous dis à tous un grand merci.

Je remercie également toute l'équipe BiMoDyM, les anciens comme les nouveaux, toutes ces personnes avec qui j'ai partagé le grand bureau du RdC Cournot pendant trois ans. Je remercie Florent et Priscila, ainsi que Yann et Nicolas, les quatre premières personnes avec qui j'ai travaillé à BiMoDyM. Pour les bons moments passés ensemble, les rires, les délires plus ou moins fins mais toujours drôles. Je remercie Sharad et Zoltan (inspecteur Palmai), avec qui j'ai partagé la plus grande partie de mon temps passé à BiMoDyM, merci pour tout les copains, pour les rires, pour les discussions plus sérieuses, pour le soutien, la solidarité dont vous avez fait preuve et qui, j'espère, continuera. Je remercie également Carole, Francesco, Samba, Guita et Martin, les stagiaires successifs passés dans l'équipe, j'espère que vous garderez tous un bon souvenir de votre séjour parmi nous. Bien évidemment, je remercie et salue tout spécialement François, François qui nous manque, disparu tragiquement, pour avoir illuminé pendant un an BiMoDyM avec ses rires, ses traits d'esprit, sa façon de toujours tout dédramatiser. Ce sont probablement les mots les plus douloureux de ce manuscrit, et du fond du cœur, je te remercie pour tout François, et te dédie l'ensemble de ces travaux de thèse. Nous ne t'oublierons jamais. Malgré la tristesse, je pense que tu aurais voulu que nous continuions dans la bonne humeur, et c'est ce que je vais tenter de faire en reprenant le fil de ces remerciements de façon plus légère.

Je remercie et salue joyeusement l'équipe de modélisation de l'Université de Nantes. Monsieur Vinh Tran, pour son précieux enseignement durant mon stage de M2, pour sa patience et son indulgence à mon égard lors de mes (nombreuses) petites boulettes durant ces 6 mois passés dans l'équipe, et pour son humour toujours fin et « l'air de rien ». Je remercie également Johann Hendrickx (alias le scripteur fou), qui m'a tout appris à mes débuts en programmation Python et m'a fait découvrir des fonctions insoupçonnées de PyMOL. Je remercie Lionel Hoffamn, qui m'a accompagné durant mes premiers pas en simulations de dynamique moléculaire. Je remercie enfin Messieurs Bernard Offman et Yves-Henri Sanéjouan pour leur accueil au sein de l'équipe et leur simplicité et gentillesse à mon égard durant ce stage de M2.

Je remercie et salue tout aussi joyeusement l'équipe de modélisation de l'Université de Reims. Monsieur Manuel Dauchez, qui m'a accueilli dans son équipe durant mon stage de M1. Monsieur Nicolas Belloy, qui m'a encadré durant ces 8 semaines de stage. Madame Stéphanie Baud, la première à m'avoir accueilli le premier jour de mon stage (où je suis brillamment arrivé en retard). Je les remercie tous les trois du plus profond de mon cœur, car c'est grâce à eux que j'en suis arrivé là aujourd'hui. Tous trois m'ont fait découvrir la modélisation moléculaire, avec patience, gentillesse, rigueur et surtout une grosse dose d'humour, chacun à leur façon mais tout aussi efficacement. Je remercie également Laurie Verzeaux, alias Mademoiselle Dieu (maintenant Docteur Mademoiselle Dieu), qui m'a toujours conseillé durant mon M2 et ma thèse grâce à son expérience (et oui ! être Mademoiselle Dieu implique une certaine ancienneté, le revers de la médaille !).

Je remercie également tous mes amis, depuis ma première année de licence jusqu'à la thèse, en passant par les collègues Ikea, Agathe (s'il te plaît, cesse de bavarder et termine ce BTS !), Claire (n'oublie pas la paille dans ta vodka-pomme), Laure, qui s'est toujours intéressée à mes travaux et m'a toujours soutenu, Ophélie, même si à Nîmes il pleut tout le temps, c'est sans rancune !, Henri, mon dernier pote de Nantes habitant à Paris, pour tout le mal qu'on a pu dire des gens en buvant des bières au Café d'Oz, Christophe, mon seul pote de la fac de Reims habitant à Paris, pour tout le bien qu'on a pu dire des gens en buvant des bières au Café d'Oz (encore !), Julien, alias Boss, devenu Docteur Boss avec quelques mois d'avance par rapport à moins, je te rejoins enfin copain ! Bastien, toujours au Congo, quand est-ce qu'on se fait une session surf copain !? Je remercie enfin Charline et Maud pour la garde de ma coloc' rousse moustachue lorsque je suis parti en congrès (ou surfer !).

Enfin, je tenais à remercier tout spécialement mes meilleurs amis de toujours, Geoffrey Caillaud et Pauline... Caillaud ! avec qui j'ai toujours partagé mes humeurs, mes joies, mes doutes, et qui m'ont toujours soutenu quoi qu'il arrive. Pendant que je réalisais cette thèse, vous fondiez un foyer, et ça... c'est la grande classe ! J'en profite également pour vous félicitez à nouveau pour le dernier arrivé, le petit Gabriel. Mon petit bonhomme, je ne te remercie pas car tu n'es pas encore en mesure de comprendre, mais qui sait, peut-être un jour vivras-tu toi aussi l'aventure d'une thèse ! Quoi qu'il arrive, je ne me fais aucun souci pour toi, avec les parents que tu as, tu iras... où tu voudras, et tu y arriveras à bon port. Je vous embrasse les amis.

Je terminerai ces remerciements par les trois personnes les plus importantes de ma vie, Papa, Maman et Tom. Merci pour tout, merci de m'avoir permis d'arriver jusqu'ici, car, quoi qu'on en dise, cette réussite n'est pas uniquement le fruit de mon travail. Papa, Maman, vous m'avez toujours fourni un environnement (social, de travail, financier) confortable, agréable et, sans tout ça, je n'aurais sûrement pas eu ce cursus universitaire. Mais bien au-delà des considérations matérielles, vous m'avez toujours soutenu, vous m'avez toujours encouragé sans jamais me forcer, vous avez su remettre un petit coup d'accélérateur quand cela était nécessaire, sans jamais me dégoûter de l'école ou des études. Je pense que c'est avant tout ce juste-milieu que vous avez su trouver qui m'a permis de devenir... hé bien tout simplement moi-même. Tom, tu n'es pas en reste, tu es sûrement celui de la famille avec qui j'ai le plus discuté de la thèse, celui qui a le plus compris ce que je faisais, tu m'as toujours écouté lorsque j'en parlais sans jamais montrer de lassitude (et pourtant, je parle !). J'en profite pour te souhaiter le meilleur à ton tour dans les études, tu es également fort bien parti, je n'ai vraiment aucune inquiétude pour toi.

A tous les trois, je vous dis un énorme merci, pour tout, et vous embrasse.

Avant propos

Cette thèse a été réalisée en codirection entre deux laboratoires académiques (ENS Cachan et VetAgro Sup Lyon). L'étude présentée vise à décrire au niveau structural les interactions entre une enzyme impliquée dans la coagulation, VKORC1, et ses inhibiteurs (les AVKs). La compréhension des mécanismes d'inhibition (et de résistance présentée par certains mutants de VKORC1) permettra à terme de concevoir de nouveaux AVKs surmontant ces résistances. L'intérêt des travaux présentés dans ce manuscrit est à la fois scientifique, thérapeutique et économique. En effet, les AVKs sont utilisés dans les thérapies anticoagulantes, la description structurale des interactions entre VKORC1 et AVKs permettra donc à terme de mieux comprendre les mécanismes de résistance et ainsi d'améliorer les traitements.

L'ensemble des travaux de modélisation présentés dans ce manuscrit ont été réalisés à l'ENS Cachan, dans l'équipe BiModyM du CMLA. Les calculs ont été réalisés sur le supercalculateur Turing de l'IDRIS et sur le cluster local TopDyn.

Au niveau des calculs présentés dans ce manuscrit, j'ai réalisé personnellement les étapes de modélisation de la structure 3D de VKORC1, de dynamique moléculaire, de docking et de calculs d'énergie libre de liaison. J'ai effectué toutes les analyses sous la supervision et en suivant les conseils et indications de Luba Tchertanov, en communiquant régulièrement avec Virginie Lattard pour m'assurer de la pertinence biologique de l'étude. La méthode *ConfigScan* a été développée par Bernard Chalmond et Alain Trouvé à partir des données de dynamique moléculaire que j'ai générées sur VKORC1.

Les travaux menés durant cette thèse ont permis la génération d'un modèle 3D de VKORC1, l'identification de l'état conformationnel actif de l'enzyme, la proposition du cycle catalytique détaillé de l'enzyme et la caractérisation de ses interactions avec son ligand naturel (la vitamine K) et trois inhibiteurs (AVKs) actuellement utilisés.

Ces travaux ont fait l'objet d'une publication commune avec VetAgro Sup Lyon et d'une publication entièrement consacrée à la modélisation et récemment soumise :

Human Vitamin K Epoxide Reductase: Retrieving of the metastable functional states from molecular dynamics simulations, N. Chatron et al.

Table des matières

Chapitre 1. INTRODUCTION	1
I. La coagulation sanguine et ses acteurs	1
1. Hémostase primaire	1
2. Hémostase secondaire ou coagulation sanguine	3
3. Hémostase tertiaire ou fibrinolyse	6
4. Découverte de la vitamine K et de ses fonctions biologiques	6
5. Mise en évidence du cycle de la vitamine K	9
II. La vitamine K époxyde réductase (VKORC1)	10
1. Identification du gène codant pour VKORC1	10
2. Mécanisme catalytique proposé pour VKORC1	11
3. VKORC1, activation directe ou indirecte ?	12
4. Modèles proposés pour le repliement du domaine membranaire de VKORC1	12
5. Inhibition de VKORC1 : les anticoagulants de type anti-vitamine K (AVK)	15
6. Problèmes liés à l'utilisation des AVKs en thérapie humaine	17
7. Problématiques liées à VKORC1	18
III. La modélisation moléculaire	21
1. Les acides aminés	21
2. Organisation structurale des protéines	22
3. Visualisation et représentations de la structure 3D des protéines	30
4. La dynamique moléculaire	32
Chapitre 2. METHODES	38
I. Banques de données	38
1. UniProt Knowledgebase	38
2. Protein Data Bank	39
II. Alignements de séquences	41
1. Comparaison de séquences avec Clustal Omega	41
2. Analyse de séquences avec BLAST	41
III. Prédictions de structures secondaires et topologie membranaire	43
1. Prédiction de structures secondaires	43
2. Prédiction de topologie membranaire	44

IV. Modélisation par homologie	45
V. Protocole de simulation de DM appliqué à VKORC1	49
1. Gromacs et ces fonctions	49
2. CHARMM36 – Champ de force utilisé pour les simulations de DM	50
3. Construction d'un modèle de membrane biologique	52
4. Solvatation du modèle de membrane et définition de la boîte de simulation	54
5. Minimisation de l'énergie potentielle du patch membranaire	55
6. Equilibration de la température du patch membranaire à volume constant	56
7. Equilibration de la pression du patch membranaire à température constante	57
8. Contrôle de la qualité de la bicouche lipidique : <i>GridMat</i>	58
9. Insertion du modèle 3D de VKORC1 dans la membrane biologique	60
10. Equilibration du système VKORC1 + membrane	63
11. Simulations de DM du système	64
VI. Analyse des données de simulations de DM par méthodes classiques	65
1. Inspection visuelle des données de simulations de DM (VMD/PyMOL)	65
2. Root Mean Square Deviations (RMSDs)	65
3. Root Mean Square Fluctuations (RMSFs)	66
4. Analyse en Composantes Principales (ACP)	66
5. Cross-correlations	67
6. <i>Clustering</i> des conformations	68
VIII. Détection de cavités à la surface de VKORC1 par <i>fpocket</i>	68
IX. Amarrage moléculaire	71
1. Algorithmes d'échantillonnage des conformations du ligand	71
2. Fonctions de notation des poses du ligand	74
3. Amarrage moléculaire : rigide ou flexible ?	75
4. Amarrage moléculaire des ligands à VKORC1 par AutoDock	76
X. Energie libre de liaison estimée par MM/PBSA	79
Chapitre 3. RESULTATS	81
I. Modélisation de la structure 3D de hVKORC1	81
1. Alignements de séquences de VKORs et recherche de protéines homologues	81
2. Prédiction des structures secondaires et de la topologie membranaire de hVKORC1 à partir de sa séquence peptidique	86
3. Modélisation par homologie de la structure 3D de VKORC1	90

4. Modélisation de la membrane du réticulum endoplasmique_____	93
5. Intégration du modèle structural de hVKORC1 dans la membrane du réticulum endoplasmique_____	94
II. Simulations de DM de hVKORC1 considérant les résidus de cystéine protonés_____	96
1. Evaluation de la stabilité structurale du modèle 3D de hVKORC1 par simulations de dynamique moléculaire de 100 ns_____	97
2. Recherche d'états métastables du modèle 3D de hVKORC1 par simulations de dynamique moléculaire d'1 μ s_____	107
3. Analyse des simulations de DM par des techniques mathématiques avancées____	113
III. Simulations de DM de hVKORC1 - Ponts disulfures prédits à partir des conformations métastables_____	122
1. Modélisation de structures 3D de hVKORC1 incluant les ponts disulfures prédits par simulations de DM_____	122
2. Détection de poches à la surface des quatre structures 3D prédites pour hVKORC1_____	128
IV. Sélectivité des modèles de hVKORC1 vis-à-vis de la vitamine K époxyde_____	133
1. Amarrage moléculaire de la vitamine K époxyde sur les quatre états enzymatiques prédits pour hVKORC1_____	133
2. Extension de la simulation de DM de l'état hVKORC1_C ₅₁ -C ₁₃₂ à 500 ns pour l'amarrage moléculaire de la vitamine K époxyde_____	139
V. Détermination par amarrage moléculaire du cycle enzymatique conduisant à la réduction de la vitK^E en vitK^Q par hVKORC1_____	144
VI. Etude des interactions au sein de complexes hVKORC1•Ligand_____	154
1. Etude des interactions entre hVKORC1 et la vitamine K époxyde_____	154
2. Etude des interactions entre hVKORC1 et trois anti-vitamines K (AVKs)_____	158
Chapitre 4. CONCLUSION ET PERSPECTIVES_____	165

Liste des figures

Figure 1.1. Mécanisme de l'hémostase primaire (phase plaquettaire)	2
Figure 1.2. Voies d'activation de la coagulation sanguine	4
Figure 1.3. Réaction de γ -carboxylation	4
Figure 1.4. Mécanisme de l'hémostase secondaire (coagulation sanguine)	5
Figure 1.5. Mécanisme de l'hémostase tertiaire (fibrinolyse)	7
Figure 1.6. Structures chimiques des différentes formes de la vitamine K	8
Figure 1.7. Cycle de la vitamine K	9
Figure 1.8. Mécanisme réactionnel de réduction de la vitamine K époxyde en vitamine K quinone par l'enzyme VKORC1	11
Figure 1.9. Modèles de repliement membranaire à 3 TMs et 4 TMs proposé pour VKORC1	14
Figure 1.10. Structure chimique des trois noyaux à partir desquels sont dérivés les différents AVKs	15
Figure 1.11. Mécanisme d'inhibition de VKORC1 par les AVKs	16
Figure 1.12. Propriétés physico-chimiques des acides aminés	23
Figure 1.13. Formation d'une liaison peptidiques, angles dièdres φ et ψ	24
Figure 1.14. Structure tridimensionnelle d'une hélice α	25
Figure 1.15. Structure tridimensionnelle d'un brin β et d'un feuillet β	25
Figure 1.16. Structure tertiaire d'une protéine	27
Figure 1.17. Structure d'une protéine	28
Figure 1.18. Relation entre les différents niveaux d'organisation structurale des Protéines	29
Figure 1.19. Principaux types de représentations 3D des protéines utilisées en modélisation moléculaire	31
Figure 1.20. Trajectoire de DM d'une protéine composés de N atomes	36
Figure 2.1. Evolution du nombre et complexité croissante des structures présentes dans la Protein Data Bank	40
Figure 2.2. Algorithme BLAST pour la recherche de régions similaires entre protéines	42
Figure 2.3. Protocole de modélisation par homologie	46
Figure 2.4. Modèle de membrane biologique construit par MemBuilder	53
Figure 2.5. Conditions périodiques aux limites appliquées à la boîte de simulation de DM	54

Figure 2.6. Aire par lipide et épaisseur de la bicouche lipidique d'une membrane biologique	58
Figure 2.7. Estimation de l'aire par lipide et l'épaisseur d'une bicouche lipide par GridMat	59
Figure 2.8. Algorithme d'insertion d'une protéine dans une membrane appliqué par la fonction <i>g_membed</i> de Gromacs	61
Figure 2.9. Algorithme de détection de poches à la surface de protéines par Fpocket	69
Figure 2.10. Algorithme de détection de poches à la surface de protéines au cours de simulations de DM par MDpocket	70
Figure 2.11. Principe de l'algorithme génétique pour l'échantillonnage des conformations d'un ligand	73
Figure 2.12. Principe de détermination de la grille de recherche des conformations d'un ligand par AutoDock	77
Figure 3.1. Séquence peptidique de hVKORC1 (forme canonique) choisie pour la modélisation	82
Figure 3.2. Séquences peptidiques des VKORs de différentes espèces alignées respectivement par rapport à hVKORC1	83
Figure 3.3. Alignement des séquences peptidiques de la VKORC1 humaine et d'une VKOR bactérienne (<i>Synechococcus sp.</i>)	84
Figure 3.4. Mécanisme enzymatique décrit pour bVKOR et proposé par analogie pour hVKORC1	86
Fig 3.5. Prédiction des structures secondaires et de la topologie membranaire de bVKOR à partir de sa séquence peptidique	87
Figure 3.6. Prédiction des structures secondaires et de la topologie membranaire de hVKORC1 à partir de sa séquence peptidique	88
Figure 3.7. Alignement des structures secondaires prédites pour hVKORC1 et des structures secondaires de son homologue bVKOR	89
Figure 3.8. Modélisation par homologie de la structure 3D de hVKORC1 à partir de la structure cristallographique de bVKOR	92
Figure 3.9. Composition lipidique générale de la membrane du RE mammalien	93
Figure 3.10. Modèle de la membrane du RE	94
Figure 3.11. Intégration du modèle 3D de hVKORC1 dans la membrane du RE Modélisée	95

Figure 3.12. RMSDs et interprétation par DSSP des structures secondaires du modèle 3D de hVKORC1	98
Figure 3.13. RMSFs des résidus du modèle 3D de hVKORC1 simulé et représentation de ces valeurs sur la structure du modèle	99
Figure 3.14. Localisation des résidus de leucine au niveau des hélices transmembranaires de hVKORC1	100
Figure 3.15. Analyse en composantes principales (ACP) de cinq simulations de DM de 100 ns de hVKORC1	102
Figure 3.16. Analyse des conformations similaires (<i>clustering</i>) de hVKORC1 obtenues au cours de cinq simulations de dynamique moléculaire de 100 ns	103
Figure 3.17. Analyse de la flexibilité de la boucle L de hVKORC1 au cours de cinq simulations de DM de 100 ns par application de métriques	104
Figure 3.18. Etude des distances S...S entre les atomes de soufre des quatre résidus C ₄₃ , C ₅₁ , C ₁₃₂ et C ₁₃₅ au cours de cinq simulations de DM de 100 ns	106
Figure 3.19. Caractérisation de la stabilité structurale globale de hVKORC1 au cours de simulations de DM d'1 μ s	108
Figure 3.20. Caractérisation de la déviation des hélices TMs et HH de hVKORC1 au cours de deux simulations de DM d'1 μ s	109
Figure 3.21. Etude des distances S...S entre les atomes de soufre des quatre résidus C ₄₃ , C ₅₁ , C ₁₃₂ et C ₁₃₅ au cours de deux simulations de DM de 1 μ s	111
Figure 3.22. Localisation des résidus hydrophobes au niveau de la boucle L de hVKORC1	112
Figure 3.23. Analyse des données de deux simulations de DM d'un μ s par ConfigScan	117
Figure 3.23.S1. Analyse spectrale de la trajectoire 3'' après l'ACP	118
Figure 3.23.S2. Analyse spectrale de la trajectoire 3'' sans réduction ACP	119
Figure 3.24. Structures 3D de hVKORC1_SH identifiés dans les minima d'énergie locaux durant deux simulations d'1 μ s	120
Figure 3.25. Positionnement de l'hélice HH par rapport aux quatre TMs au cours de deux simulations de DM d'1 μ s	121
Figure 3.26. Modélisation de structures de hVKORC1 incluant des ponts disulfures à partir des conformations (U min) identifiées par dynamique moléculaire	123
Figure 3.27. Caractérisation de la stabilité structurale globale des quatre modèles structuraux prédits pour hVKORC1	126
Figure 3.28. Modèles de hVKORC1 prédits par simulations de DM	127

Figure 3.29. Analyse des conformations similaires (<i>clustering</i>) de hVKORC1 pour chaque modèle prédit	129
Figure 3.30. Poches détectées à la surface des structures 3D pour chaque conformation la plus représentative de chaque modèle de hVKORC1	130
Figure 3.31. Volume des cavités détectées à la surface des structures 3D des conformations représentatives de chaque modèle de hVKORC1	132
Figure 3.32. Amarrage moléculaire de la vitK ^E (forme intégrale et forme tronquée) sur le modèle hVKORC1_C ₅₁ -C ₁₃₂	135
Figure 3.33. Amarrage moléculaire de la vitK ^E tronquée sur le modèle hVKORC1_C ₄₃ -C ₅₁ _C ₁₃₂ -C ₁₃₅	136
Figure 3.34. Amarrage moléculaire de la vitK ^E tronquée sur le modèle hVKORC1_C ₄₃ -C ₅₁	137
Figure 3.35. Amarrage moléculaire de la vitK ^E tronquée sur le modèle hVKORC1_C ₁₃₂ -C ₁₃₅	138
Figure 3.36. Caractérisation de la stabilité structurale globale du modèle hVKORC1_C ₅₁ -C ₁₃₂ au cours de la simulation de DM étendue à 500 ns	141
Figure 3.37. Détection de poches à la surface des structures 3D correspondant à la seconde conformation représentative de hVKORC1_C ₅₁ -C ₁₃₂ obtenue par extension de la simulation de DM à 500 ns	142
Figure 3.38. Amarrage moléculaire de la vitK ^E tronquée sur la seconde conformation représentative du modèle hVKORC1_C ₅₁ -C ₁₃₂	143
Figure 3.39. Mécanismes enzymatiques potentiels conduisant à la réduction de la vitK ^E en vitK ^Q	145
Figure 3.40. Amarrage moléculaire de la vitK ^H tronquée sur le modèle hVKORC1_C ₅₁ -C ₁₃₂ , considérant le résidu C ₁₃₅ à l'état réduit (S ⁻)	147
Figure 3.41. Amarrage moléculaire de la vitK ^H tronquée sur le modèle hVKORC1_C ₄₃ -C ₅₁ , considérant le résidu C ₁₃₅ à l'état réduit (S ⁻)	149
Figure 3.42. Amarrage moléculaire de la vitK ^Q tronquée sur le modèle hVKORC1_C ₄₃ -C ₅₁ , considérant les résidus C ₁₃₂ et C ₁₃₅ à l'état réduit (S ⁻)	150
Figure 3.43. Amarrage moléculaire de la vitK ^Q tronquée sur le modèle hVKORC1_C ₄₃ -C ₅₁ _C ₁₃₂ -C ₁₃₅	151
Figure 3.44. Cycle catalytique conduisant à la réduction de la vitK ^E en vitK ^Q par hVKORC1 postulé à partir des résultats de <i>docking</i> moléculaire	153

Figure 3.45. Reconstitution de la chaîne carbonée intégrale de la vitKE à partir de la pose de <i>docking</i> unique obtenue sur le modèle hVKORC1_C ₅₁ -C ₁₃₂ _____	155
Figure 3.46. Détermination de la contribution des résidus de hVKORC1_C ₅₁ -C ₁₃₂ dans les interactions de l'enzyme avec la vitK ^E au cours de deux simulations de DM de 100 ns concaténées_____	157
Figure 3.47. Complexes hVKORC1_C ₅₁ -C ₁₃₂ •AVK générés par <i>docking</i> moléculaire et pour lesquels une pose unique a été obtenue_____	158
Figure 3.48. Energies libres de liaison des complexes hVKORC1•vitKE et hVKORC1•AVK calculées au cours de simulations de DM de 100 ns concaténées_____	159
Figure 3.49. Comparaison entre les énergies libres de liaison calculées et les constantes d'inhibition mesurées expérimentalement pour trois hVKORC1•AVK_____	160
Figure 3.50. Détermination de la contribution des résidus de hVKORC1_C ₅₁ -C ₁₃₂ dans les interactions de l'enzyme avec l'acénocoumarol au cours de deux simulations de DM de 100 ns concaténées_____	161
Figure 3.51. Détermination de la contribution des résidus de hVKORC1_C ₅₁ -C ₁₃₂ dans les interactions de l'enzyme avec le difénacoum au cours de deux simulations de DM de 100 ns concaténées_____	162
Figure 3.52. Détermination de la contribution des résidus de hVKORC1_C ₅₁ -C ₁₃₂ dans les interactions de l'enzyme avec la phenprocoumone au cours de deux simulations de DM de 100 ns concaténées_____	163

Liste des tableaux

Tableau 1.1. Mutations de VKORC1 et phénotypes associés_____	18
Tableau 2.1. Fichiers générés par Gromacs au cours des simulations de DM_____	50
Tableau 3.1. Structures cristallographiques de bVKOR (<i>Synechococcus sp.</i>) disponibles dans la PDB_____	85

Liste des abréviations

ACP : analyse en composantes principales
AVK : anti-vitamine K
CHARMM : chemistry at Harvard macromolecular mechanics
CPU : central processing unit
DLPC : dilauroylphosphatidylcholine
DM : dynamique moléculaire
DPPC : dipalmitoylphosphatidylcholine
DSSP : dictionary of secondary structures of proteins
ELL : énergie libre de liaison
GG : gros grain
GGCX : gamma-glutamyl-carboxylase (γ -glutamyl-carboxylase)
GPU : graphics processing unit
HH : hélice horizontale
LINCS : linear constraint solver
MM : mécanique moléculaire
MM/PBSA : molecular mechanics / Poisson-Boltzman Surface Area
MQ : mécanique quantique
PME : particle mesh Ewald
PVKD : protéine vitamine K-dépendante
RE : réticulum endoplasmique
RMSD : root mean square deviation
RMSF : root mean square fluctuation
TM : hélice transmembranaire
TMX : thioredoxin-related transmembrane protein
vitK^E : vitamine K époxyde
vitK^H : vitamine K hydroxyquinone
vitK^Q : vitamine K quinone
VKOR : vitamine K époxyde réductase
VKORC1 : vitamin K epoxide reductase complex subunit 1
bVKOR : VKOR bactérienne
hVKORC1 : VKORC1 humaine

Chapitre 1. INTRODUCTION

I. La coagulation sanguine et ses acteurs

La **coagulation sanguine** est une propriété physiologique cruciale pour la survie d'un individu, consistant à stopper une hémorragie en cas de lésion d'un vaisseau sanguin. La coagulation participe donc à l'hémostase, son équilibre est assuré par des régulateurs enzymatiques dont les effets sont antagonistes. Ainsi, un déficit en facteur pro-coagulants augmente les risques d'hémorragie, tandis qu'un excès augmente les risques d'embolie et de thrombose [1, 2]. L'hémostase est hautement préservée chez les mammifères et démarre par l'hémostase primaire, suivie de l'hémostase secondaire (coagulation plasmatique), puis s'achève par l'hémostase tertiaire (fibrinolyse).

1. Hémostase primaire

L'hémostase primaire est subdivisée en deux étapes, la phase vasculaire puis la phase plaquettaire. La phase vasculaire correspond à une contraction du vaisseau sanguin lésé visant à limiter la perte de sang. Lorsque le sang atteint le sous-endothélium du vaisseau sanguin, la phase plaquettaire est déclenchée. Les plaquettes sanguines se fixent alors au collagène ou au facteur VIII présent au niveau du sous-endothélium (adhésion plaquettaire). Les plaquettes sanguines sont ensuite activées, exprimant à leur surface un récepteur au fibrinogène, le complexe glycoprotéique IIb/IIIa (également appelé intégrine $\alpha_{IIb}\beta_3$). Elles sécrètent également de l'ADP et du thromboxane A2, deux molécules permettant l'activation d'autres plaquettes sanguines (activation plaquettaire). Les plaquettes sanguines ainsi activées s'agrègent entre elles par fixation du fibrinogène sur les intégrines $\alpha_{IIb}\beta_3$ (agrégation plaquettaire). L'agrégat de plaquettes formé est appelé clou plaquettaire ou thrombus blanc et suffit à stopper le saignement lors d'une lésion peu importante d'un vaisseau sanguin. Ce clou plaquettaire sera ensuite consolidé par la mise en place de l'hémostase secondaire. En effet, lors de l'agrégation plaquettaire, l'orientation des phospholipides composant la membrane des plaquettes (notamment les phosphatidylsérines) est modifiée. Des groupements phosphate, non accessibles sur les plaquettes non agrégées, sont ainsi exposés à la surface du clou plaquettaire (**Figure 1.1**). Ils permettront la fixation de certains **facteurs de coagulation** activés lors de l'hémostase secondaire [3].

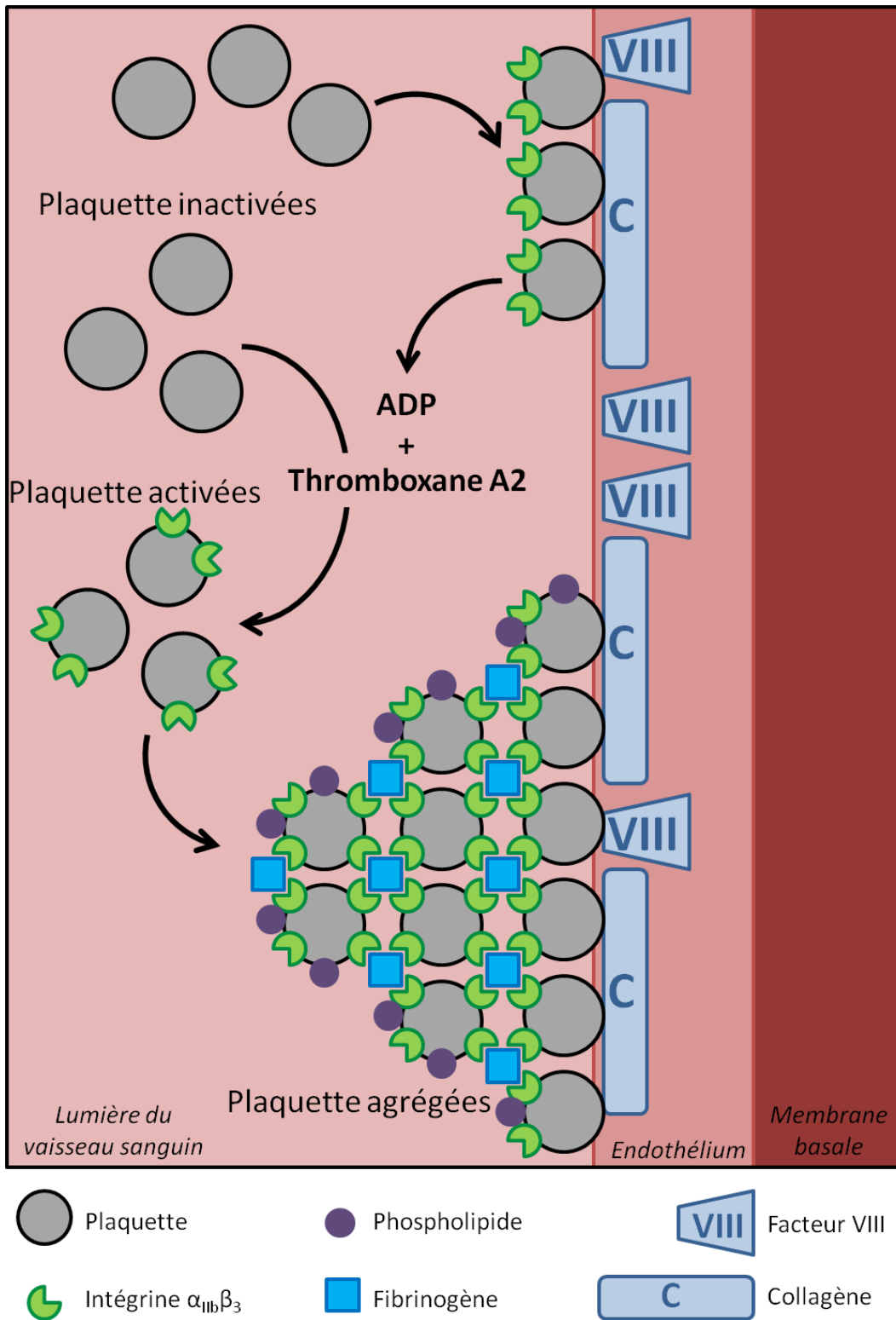


Figure 1.1. Mécanisme de l'hémostase primaire (phase plaquettaire). L'hémostase primaire provoque la formation d'un thrombus blanc par l'agrégation de plaquettes au niveau de la lésion vasculaire.

2. Hémostase secondaire ou coagulation sanguine

L'hémostase secondaire, ou coagulation, est une cascade de réactions enzymatiques conduisant à l'activation des facteurs de coagulation. Deux voies d'activation de la coagulation coexistent. La voie extrinsèque est activée par la libération d'activateurs tissulaires. La voie intrinsèque est déclenchée par des protéines plasmatiques, directement activées par l'endothélium du vaisseau sanguin lésé ou par les plaquettes activées durant l'hémostase primaire. Ces deux voies d'activation convergent ensuite en un mécanisme unique [1].

Les facteurs de coagulation impliqués dans l'hémostase secondaire, numérotés de I à XIII, existent tous à l'état de précurseurs inactifs. Les facteurs de coagulation II, VII, IX et X appartiennent à la famille des protéines vitamine K-dépendantes (PVKDs, **Figure 1.2**). En effet, leur activation nécessite une modification post-traductionnelle, la γ -carboxylation, catalysée par l'enzyme γ -glutamyl-carboxylase (GGCX) dont **la vitamine K est le cofacteur unique et indispensable**. La γ -carboxylation des protéines s'effectue dans les hépatocytes, les ostéoblastes et les cellules musculaires lisses. Cette réaction provoque une carboxylation des résidus d'acide glutamique au niveau de l'atome de carbone γ (**Figure 1.3**). Les résidus ainsi modifiés sont appelés acides γ -carboxy-glutamiques (Gla), dont la structure leur permet de former des liaisons très fortes avec des cations divalents, particulièrement les ions Ca^{2+} . Les facteurs de coagulation II, VII, IX et X activés sont donc capables de se fixer aux groupements phosphates exposés à la surface du clou plaquettaire formé en amont, par l'intermédiaire de liaisons calcium-dépendantes. Ces quatre facteurs se concentrent alors au niveau du thrombus blanc, interagissent les uns avec les autres et déclenchent la coagulation par production de filaments de fibrine insolubles à partir du fibrinogène. Ces filaments s'associent ensuite en réseau et entravent les cellules circulantes, notamment les globules rouges (**Figure 1.4**). L'amas insoluble ainsi formé constitue un caillot sanguin, également nommé thrombus rouge. La vitamine K assure donc une fonction primordiale, bien qu'indirecte, dans les deux voies d'activation de la coagulation. Elle est également impliquée dans l'activation des protéines C et S, inhibitrices de la coagulation [4, 5].

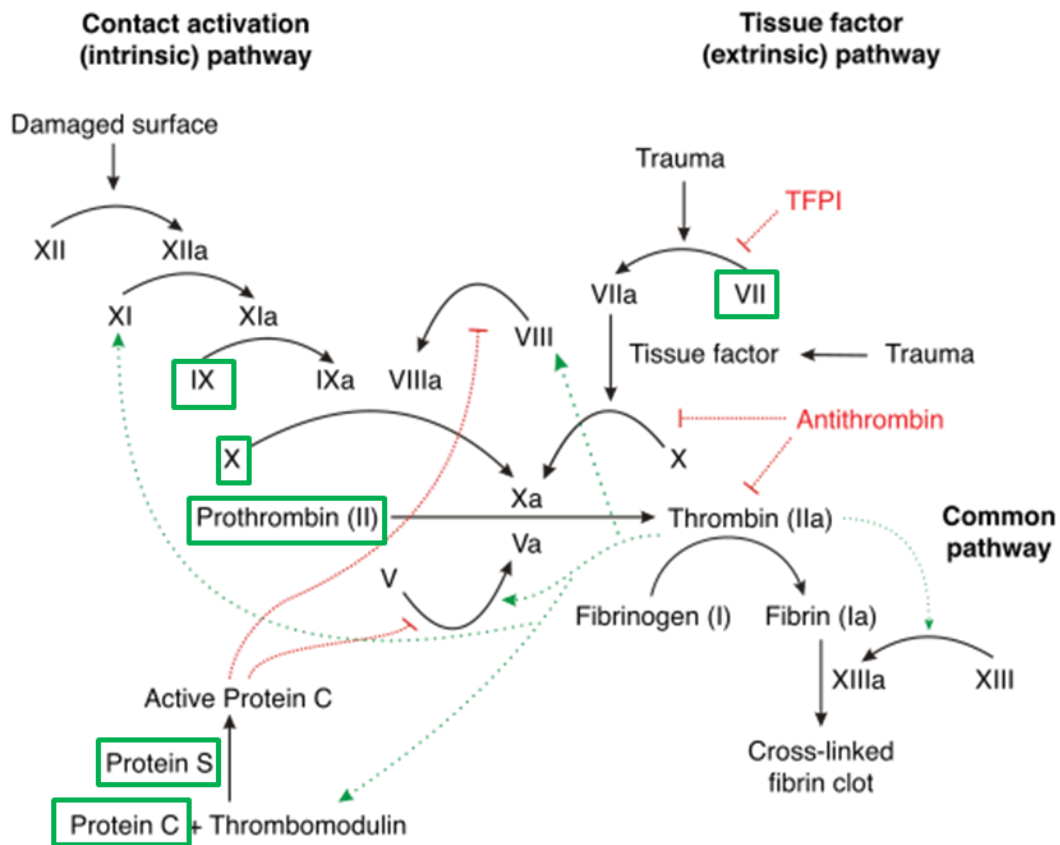


Figure 1.2. Voies d'activation de la coagulation sanguine[6]. Les voies activatrices sont représentées en noir, les voies d'inhibition en rouge et les boucles de rétrocontrôle par la thrombine en pointillés verts. Les protéines vitamine K-dépendantes sont encadrées en vert.

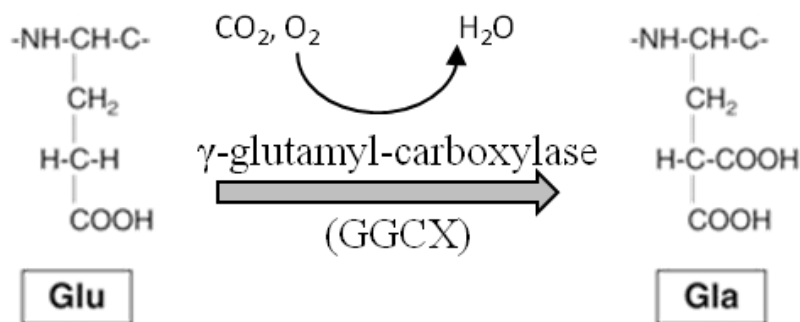


Figure 1.3. Réaction de γ -carboxylation. La γ -glutamyl-carboxylase fixe un groupement carboxyle au niveau de l'atome de carbone γ des résidus d'acide glutamique (Glu), les convertissant en résidus d'acide γ -carboxy-glutamiques (Gla). Chaque réaction de γ -carboxylation consomme une molécule de CO_2 et d' O_2 et produit une molécule d'eau.

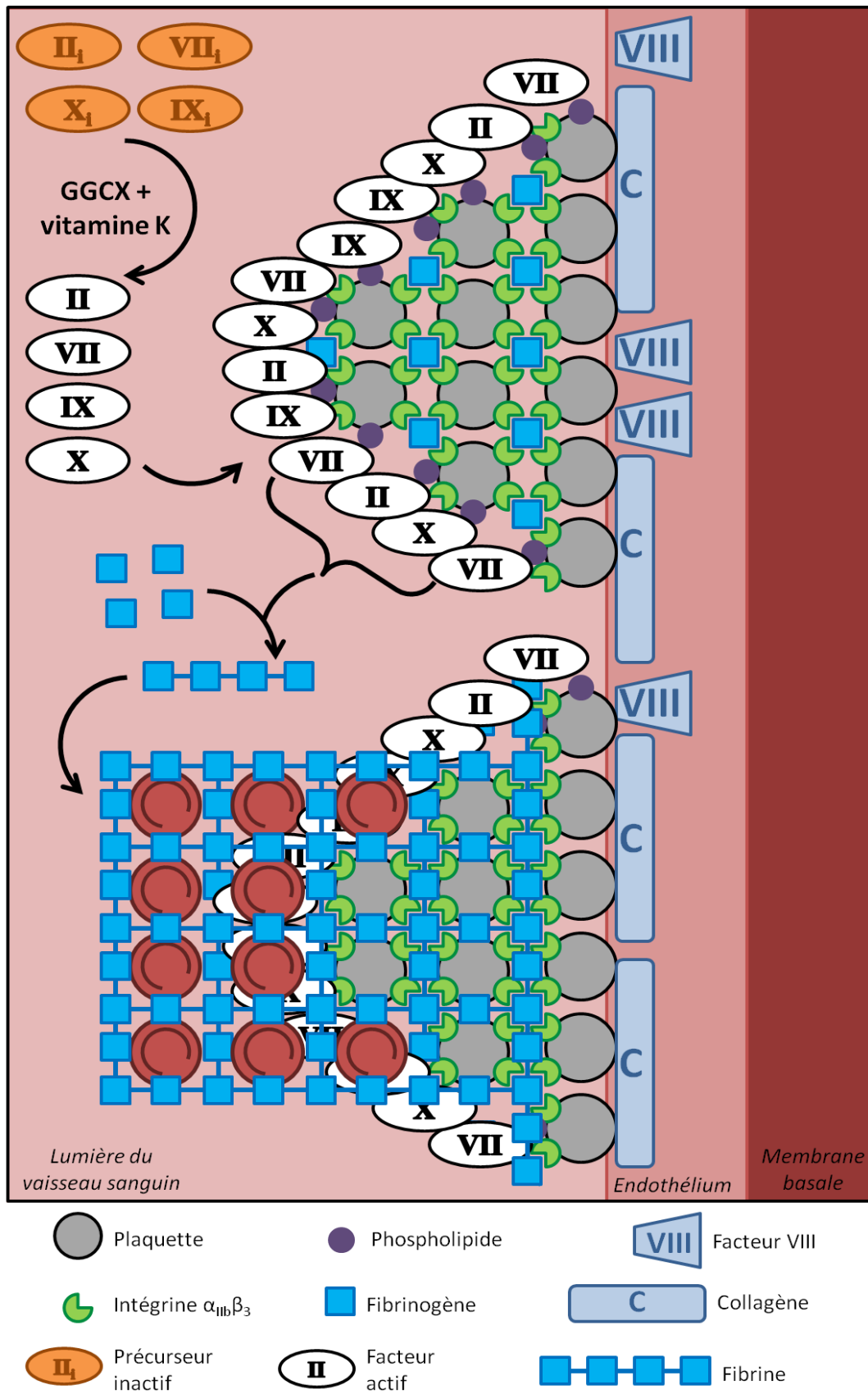


Figure 1.4. Mécanisme de l'hémostase secondaire (coagulation sanguine). L'hémostase secondaire provoque la formation d'un thrombus rouge par l'accumulation de globules rouges dans un réseau de fibrine.

3. Hémostase tertiaire ou fibrinolyse

L'hémostase tertiaire, ou fibrinolyse, vise à dissoudre le caillot sanguin formé précédemment pour éviter les complications thromboemboliques. Cette dernière étape de l'hémostase implique la production de plasmine à partir du plasminogène, son précurseur inactif dont la synthèse est hépatique. Deux activateurs du plasminogène, l'activateur tissulaire (tPA) et l'activateur de type urokinase (uPA), sont responsables de sa conversion en plasmine. Le tPA est produit par l'endothélium vasculaire ; il forme un complexe avec le plasminogène et la fibrine, aboutissant à la production de plasmine. L'uPA est synthétisé sous forme inactive (pro-urokinase) par les monocytes, les macrophages, les cellules épithéliales du glomérule rénal et les cellules endothéliales. Il est ensuite activé par la thrombine, puis transforme le plasminogène en plasmine indépendamment de la présence de fibrine. Le tPA et l'uPA sont inhibés par le *Plasminogen Activator Inhibitor* (PAI) en l'absence de coagulation. La plasmine activée se fixe finalement sur la fibrine et assure sa protéolyse. Les produits de dégradation de la fibrine (PDF) inhibent alors la formation de fibrine et l'agrégation plaquettaire (**Figure 1.5**).

4. Découverte de la vitamine K et de ses fonctions biologiques

La molécule de **vitamine K** a été découverte en 1935 par le biochimiste danois Henrik Carl Peter Dam, qui étudiait le rôle physiologique du cholestérol dans l'alimentation. Il a constaté que des poussins soumis à un régime appauvri en cholestérol mourraient d'un syndrome hémorragique après quelques semaines. La mort des animaux survenait y compris lors du rétablissement d'un taux normal de cholestérol dans leur alimentation. Ces observations lui suggèrent l'existence d'un facteur alimentaire inconnu mais indispensable à la coagulation, très probablement lié aux stérols. Ses recherches aboutissent à la découverte d'une molécule liposoluble, distincte du cholestérol, qu'il nomme alors vitamine K[7] (pour *koagulation*, en danois). Il parvient ensuite à extraire la vitamine K1 de la luzerne, en 1936. La synthèse chimique de cette molécule est réalisée trois ans plus tard par un chercheur américain, Edward Adelbert Doisy. Le prix Nobel de Physiologie (ou de Médecine) est décerné conjointement à Henrik Carl Peter Dam et Edward Adelbert Doisy en 1943, respectivement pour la découverte de la vitamine K et pour la détermination de sa structure chimique.

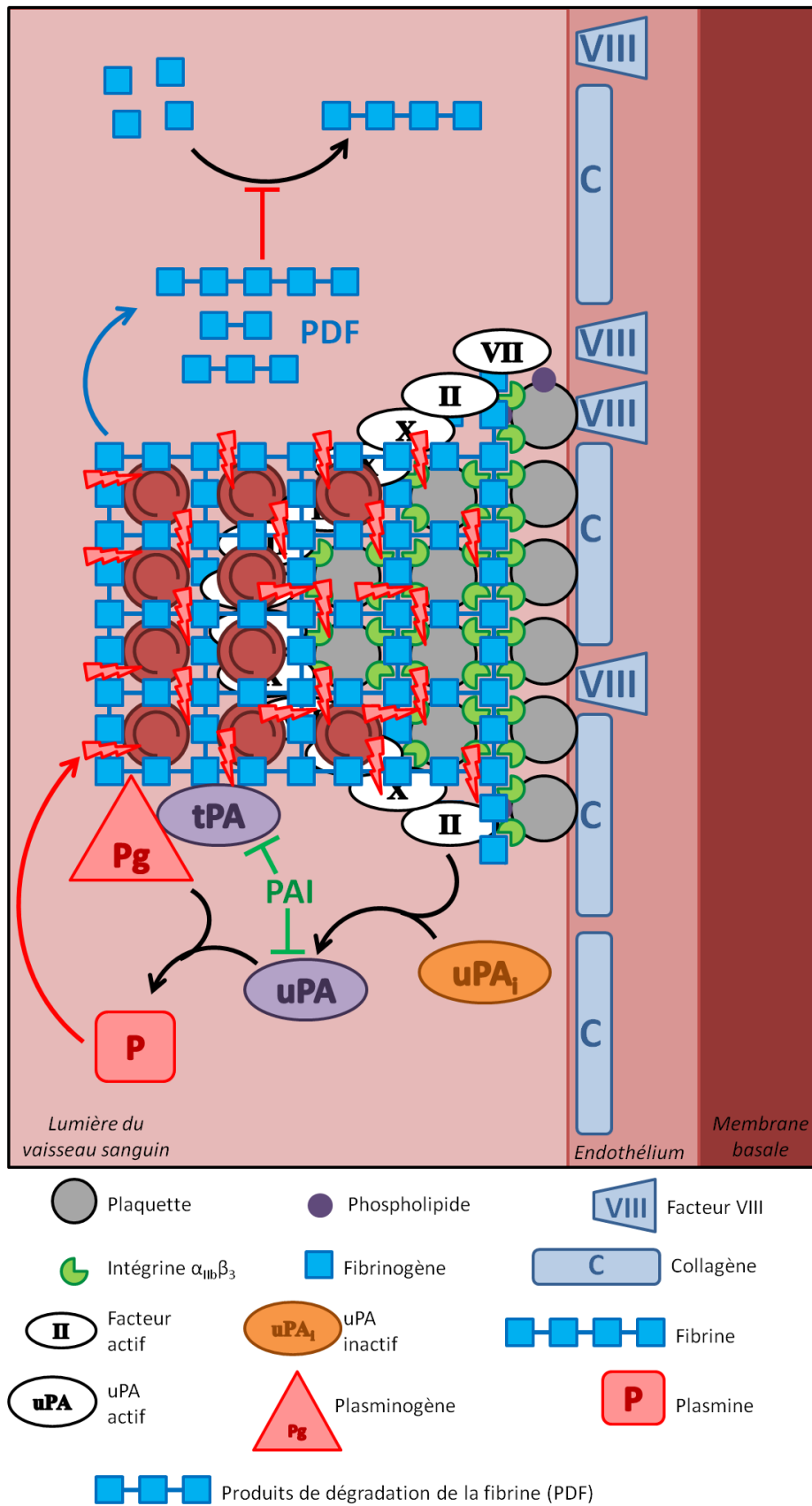


Figure 1.5. Mécanisme de l'hémostase tertiaire (fibrinolyse). L'hémostase tertiaire provoque la dissolution du thrombus rouge par la dégradation du réseau de fibrine.

Le terme vitamine K désigne en réalité une famille de molécules de type quinone. Trois variétés de vitamine K ont été identifiées. La vitamine K1, également nommée phylloquinone, est présente dans les légumes verts principalement. La vitamine K2, aussi appelée ménaquinone, est d'origine animale. Plusieurs sous-types de ménaquinones existent, de la ménaquinone-4 à la ménaquinone-14, selon le nombre de groupements isoprène composant leur chaîne hydrophobe. La vitamine K3, ou ménadione, est une forme synthétique de la vitamine K (**Figure 1.6**).

La vitamine K intervient dans de nombreuses fonctions biologiques, tels que le métabolisme osseux [8, 9], les processus cancéreux, la réponse inflammatoire [10], le stress oxydatif [11], la synthèse des sphingolipides [12] et l'activité exocrine du pancréas [13]. Le mécanisme d'action de la vitamine K dans la coagulation sanguine est cependant resté inconnu jusqu'à la mise en évidence de son rôle de cofacteur de la GGCX [14]

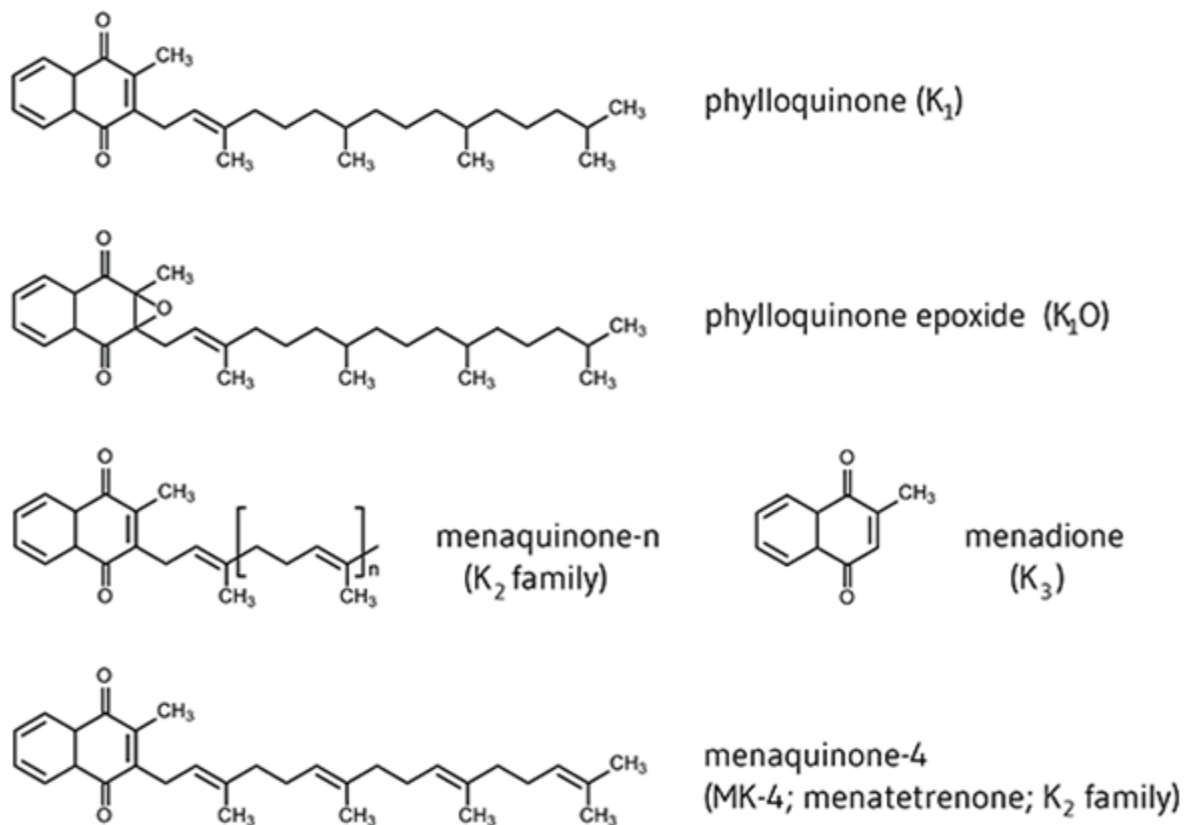


Figure 1.6. Structures chimiques des différentes formes de la vitamine K.

5. Mise en évidence du cycle de la vitamine K

La vitamine K est principalement stockée dans le foie, lieu de synthèse de la majorité des PVKDs. La vitamine K1 ne constitue qu'une faible part du stock de vitamine K hépatique mais est majoritairement utilisée comme cofacteur de la GGCX. Certaines vitamines K2, composant l'essentiel du stock de vitamine K hépatique, peuvent également remplir cette fonction [15]. La GGCX est activée en présence d'une forme réduite de la vitamine K, la vitamine K hydroquinone (vitK^H)[14]. Chaque réaction de γ -carboxylation provoque la transformation d'une molécule de vitK^H en vitamine K-2,3-époxyde (vitK^E) par la GGCX. Or, la vitK^E est incapable d'activer la GGCX, et l'apport exclusivement alimentaire de vitamine K est insuffisant pour assurer l'activité physiologique de la GGCX. Cette observation sous-entend l'existence d'un mécanisme biologique de recyclage de la vitK^E en vitK^H, permettant le maintien de l'activité de la GGCX. La réduction de la vitK^E en vitK^H s'effectue en deux étapes.

La vitK^E est dans un premier temps convertie en vitamine K quinone (vitK^Q, correspondant à la forme de la vitamine K telle qu'elle est ingérée), puis la vitK^Q est réduite en vitK^H, activant la GGCX. Ces deux réactions chimiques sont catalysées par une enzyme, la **vitamine K époxyde réductase (VKORC1, Figure 1.7)**[16].

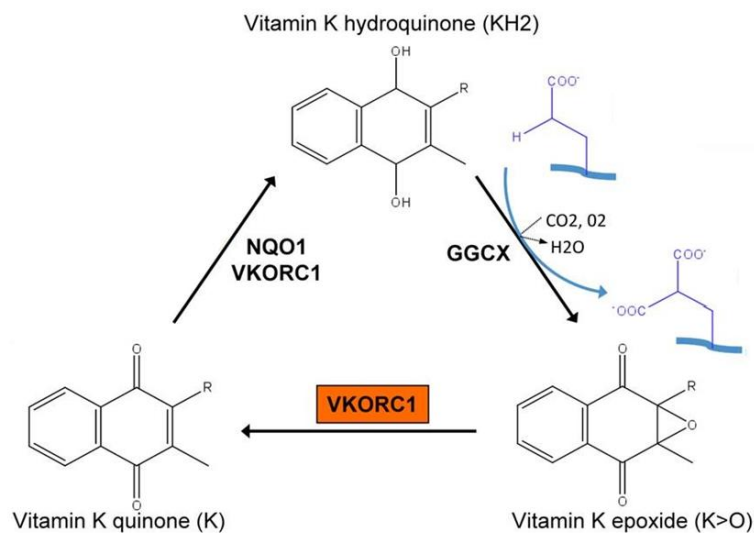


Figure 1.7. Cycle de la vitamine K [16]. La vitamine K est ingérée sous forme quinone puis réduite en hydroquinone, cofacteur de la GGCX. La vitamine K hydroquinone est ensuite transformée en vitamine K époxyde lors de la γ -carboxylation des résidus Glu par la GGCX. L'enzyme VKORC1 assure alors la réduction de la vitamine K époxyde en vitamine K quinone ; la molécule de vitamine K peut donc effectuer un nouveau cycle d'activation de la GGCX, assurant ainsi l'activation des facteurs de coagulation.

La réduction de la vitK^Q en vitK^H est également catalysée par une seconde enzyme, la NADPH quinone oxydoréductase 1 (NQO1). Une molécule de vitamine K peut ainsi être utilisée jusqu'à plusieurs centaines de fois par la GGCX. La VKORC1 est donc un acteur primordial de la coagulation, à travers sa fonction de recyclage de la vitamine K.

II. La vitamine K époxyde réductase (VKORC1)

1. Identification du gène codant pour VKORC1

L'activité vitamine K époxyde réductase (VKOR) a été décrite dès 1970 par Bell et Matschiner [17], qui ont mis en évidence l'inhibition de cette activité au niveau du foie par l'administration de coumaphène, une molécule anticoagulante développée en 1943. L'activité VKOR a d'abord été attribuée à un complexe multi-protéique, mais toutes les tentatives de purification des protéines intervenant dans ce mécanisme ont échoué. Cependant, la découverte conjointe d'un gène par deux équipes différentes en 2004 [18, 19] a remis en cause l'origine multi-protéique de l'activité VKOR. Ce gène de 5126 paires de bases, appelé *Vkorc1* (*vitamin K epoxide reductase complex subunit 1*), est localisé sur le chromosome 16 chez l'homme et code **une protéine de 163 acides aminés, VKORC1**, localisée dans la membrane du réticulum endoplasmique (RE). L'inhibition de ce gène par siARN diminue significativement l'activité VKOR [19], tandis qu'une surexpression de VKORC1 dans des cellules d'insecte provoque l'apparition d'une activité VKOR [18].

Le gène *Vkorc1* appartient à une famille de gènes homologues présents dans tous les domaines du vivant (vertébrés, insectes, végétaux, bactéries). Le séquençage de *Vkorc1* chez différentes espèces a révélé la conservation stricte de plusieurs acides aminés de la protéine codée, notamment **quatre résidus de cystéine** (C₄₃, C₅₁, C₁₃₂ et C₁₃₅ chez l'homme) [20, 21]. Les résidus C₁₃₂ et C₁₃₅ forment un **centre redox CXXC** retrouvé chez de nombreuses protéines de la famille des oxydoréductases. Ces deux résidus constituent **le site actif** de VKORC1 [22, 23].

2. Mécanisme catalytique proposé pour VKORC1

Un mécanisme de réduction de la vitK^E en vitK^Q, basé sur des études biochimiques, a été proposé par Silverman en 1981 (soit plus de vingt ans avant la découverte du gène *Vkorc1*) [24]. Le mécanisme suggéré implique les deux résidus de cystéine composant le site actif, ainsi qu'un acide aminé tiers assurant la fonction de catalyseur acide-base. Les deux résidus de cystéine formeraient un **pont disulfure** (S-S) dans la forme oxydée de l'enzyme. Ce pont serait d'abord réduit, permettant **une activation** de l'enzyme. La vitK^E se fixerait ensuite à l'enzyme réduite, l'atome d'oxygène du groupement époxyde étant protoné par le résidu acide mentionné précédemment. Le groupement thiol d'un des deux résidus de cystéine attaquerait dans le même temps le groupement époxyde, au niveau de l'atome de carbone en position 3. Le groupement hydroxyle ainsi formé serait à nouveau protoné, aboutissant à la réduction totale de la vitK^E en vitK^Q, à la libération d'une molécule d'eau et au retour de l'enzyme à l'état oxydé (**Figure 1.8**). Aucune étude expérimentale confirmant ce mécanisme réactionnel n'a été publiée depuis la découverte de VKORC1, mais ce modèle reste d'actualité.

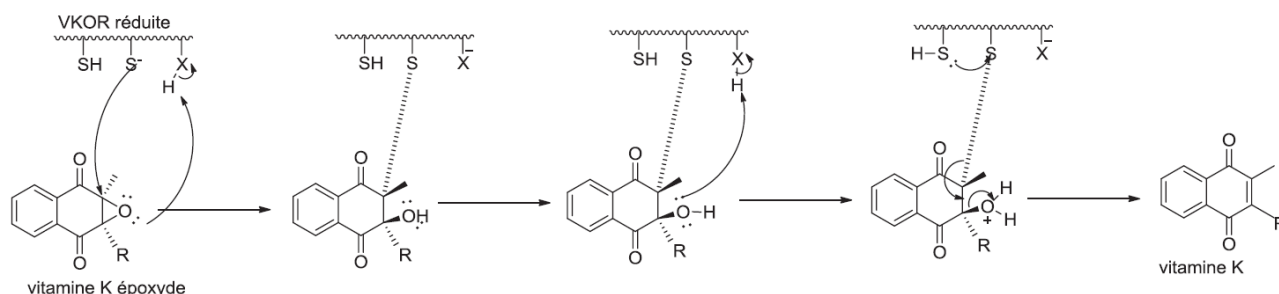


Figure 1.8. Mécanisme réactionnel de réduction de la vitamine K époxyde en vitamine K quinone par l'enzyme VKORC1 [24]. Le mécanisme proposé pour la réduction de la vitamine K époxyde en vitamine K quinone implique les deux résidus C132 et C135, ce dernier attaquant le groupement époxyde au niveau du carbone 2. Le groupement époxyde serait ensuite protoné par un troisième acide aminé, jusqu'à la réduction totale de la vitamine K époxyde en vitamine K quinone.

3. VKORC1, activation directe ou indirecte ?

Le motif redox CXXC constituant le site actif de VKORC1 nécessite une réaction de réduction pour être activé, l'intervention d'un réducteur physiologique est donc indispensable. Plusieurs molécules ont été suggérées pour effectuer ce rôle, tels que le glutathion, le système thiorédoxine/thiorédoxine réductase/NAPDH (capable de réduire le centre redox CXXC de VKORC1 [24, 25]) et la protéine disulfure isomérase (PDI [5]), qui a révélé une association étroite avec VKORC1 [26]. L'implication du glutathion a par la suite été exclue [27], tout comme celle du système thiorédoxine/thiorédoxine réductase/NAPDH [28]). La participation de la PDI a également été remise en cause [29]. Cependant, les études par co-immunoprécipitation effectuées ont suggéré l'intervention de trois autres protéines de la famille des *thioredoxin-related transmembrane proteins* (TMXs), la TMX1, la TMX4 et l'*endoplasmic resident protein 18* (Erp18). Le modèle initialement proposé pour la PDI serait donc valable en remplaçant cette protéine par l'une des trois TMXs.

La participation d'un partenaire protéique dans l'activation de VKORC1 est donc hautement probable, malgré les difficultés d'identification de la (ou des) protéine(s) impliquée(s). L'activation de VKORC1 peut s'effectuer par réduction directe du site actif C₁₃₂XXC₁₃₅ ou par l'intermédiaire des résidus C₄₃ et C₅₁ (hautement conservés chez toutes les VKORs), transférant le pouvoir du réducteur physiologique au site actif. Le rôle des résidus C₄₃ et C₅₁ reste néanmoins discuté, puisque la théorie de leur implication suggérée en 2005 par Rost [22] a été remise en question par les travaux publiés en 2007 par Jin [30]. La participation des résidus C₄₃ et C₅₁ à l'activation de VKORC1 est étroitement liée au repliement du domaine membranaire de VKORC1, qui est aussi largement débattu dans la littérature.

4. Modèles proposés pour le repliement du domaine membranaire de VKORC1

Deux modèles de **topologie du domaine membranaire**, basés sur des études expérimentales, ont été avancés pour l'enzyme VKORC1. Plusieurs analyses biochimiques et biophysiques ont conduit l'équipe de Tie à proposer dès 2005 un modèle de VKORC1 organisée en trois hélices transmembranaires (TMs) [31].

En revanche, d'autres données, notamment la résolution de la structure 3D d'un homologue bactérien de VKORC1 par cristallographie aux rayons X en 2010 [32], ont abouti à la suggestion d'un repliement de VKORC1 en 4 TMs par les équipes de Li et de Watzka [2, 32]. La structure 3D de l'ubiquinone réductase de *Synechococcus sp.* a révélé une région catalytique à 4 TMs et la présence d'un domaine thiorédoxine, absent chez l'enzyme mammalienne mais dont le repliement est similaire à celui des TMXs activant potentiellement VKORC1 [29, 32].

Le modèle de VKORC1 à 3 TMs implique une extrémité N-terminale localisée dans la lumière du RE, une longue boucle cytoplasmique reliant TM1 et TM2 et comprenant les résidus C₄₃, C₅₁, l'extrémité C-terminale étant cytoplasmique (**Figure 1.9**). Ce repliement est cohérent avec un mécanisme d'activation consistant en la réduction directe du site actif C₁₃₂XXC₁₃₅ situé à l'extrémité luminale de TM3, sans intervention des résidus C₄₃ et C₅₁ [31]. Le modèle de VKORC1 à 4 TMs suppose quant à lui une localisation luminale des extrémités N- et C-terminales ainsi que de la boucle reliant TM1 et TM2 et comprenant les résidus C₄₃ et C₅₁. Ce repliement de VKORC1 est en adéquation avec un mécanisme d'activation similaire à celui de la VKOR de *Synechococcus sp.*, soit le transfert du pouvoir réducteur entre une protéine de type thiorédoxine (située dans la lumière du RE) et le site actif (localisé à l'extrémité luminale de TM4) par l'intermédiaire des résidus C₄₃ et C₅₁.

Les résultats des études biochimiques et biophysiques menées sur VKORC1 pour déterminer son repliement membranaire restent à l'heure actuelle contradictoires [20, 22, 23, 30, 31, 33-38]. L'absence de structure 3D résolue expérimentalement pour une VKORC1 mammalienne, couplée à la difficulté d'identifier le nombre d'hélices transmembranaires et aux incertitudes quant au rôle des résidus C₄₃ et C₅₁, rendent impossible la description détaillée du mécanisme enzymatique de VKORC1.

En outre, le rôle biologique de VKORC1 fait de cette protéine une cible privilégiée dans les traitements anticoagulants. Plusieurs molécules inhibitrices de type anti-vitamine K (AVK) ont été développées afin de bloquer le recyclage de la vitamine K par VKORC1, et ainsi stopper le processus de coagulation. Ces inhibiteurs sont utilisés en thérapie humaine.

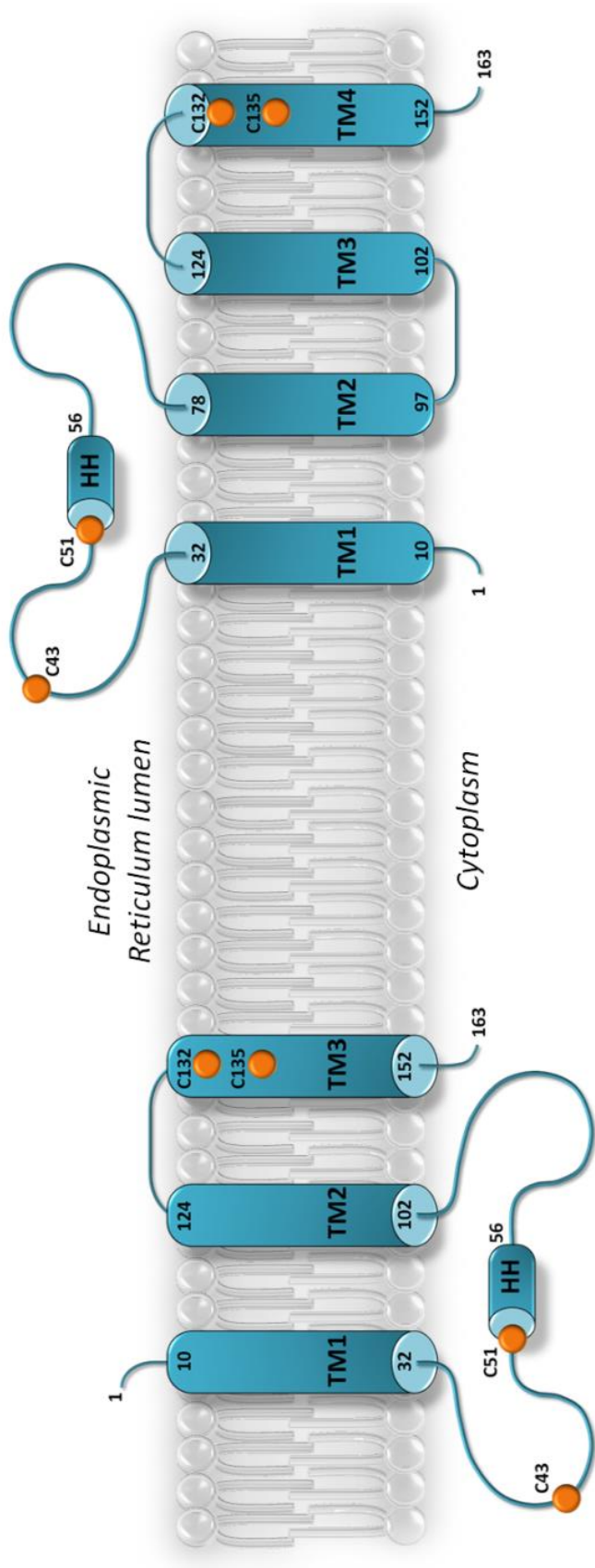


Figure 1.9. Modèles de repliement membranaire à 3 TMs et 4 TMs proposé pour VKORC1. Le modèle de repliement en 3 TMs proposé par Tie suggère une activation de VKORC1 par réduction directe du site actif C₁₃₂XXC₁₃₅, sans intervention des résidus C₄₃ et C₅₁. En revanche, le modèle de VKORC1 à 4 TMs avancé par les équipes de Li et de Watzka implique un transfert de pouvoir réducteur entre une protéine située dans la lumière du RE et le site actif C₁₃₂XXC₁₃₅ par l'intermédiaire des résidus C₄₃ et C₅₁.

5. Inhibition de VKORC1 : les anticoagulants de type anti-vitamine K (AVK)

Le premier **inhibiteur de type AVK** a été découvert en 1922 par un vétérinaire américain, Frank W. Schofield. Il a démontré que des épisodes hémorragiques survenant chez des bovins étaient dus à leur alimentation, composées de luzerne importée d'Europe [39]. La molécule responsable de ces hémorragies est identifiée en 1939 par Karl Paul Gerhard Link, un biochimiste américain, qui parvient à isoler sa structure cristalline. Cette molécule résulte de la dégradation de la luzerne, au cours de laquelle la coumarine est oxydée en 4-hydroxycoumarine, puis couplée au formaldéhyde pour enfin se dimériser, formant une molécule de dicoumarol. Link ne découvre la fonction anti-vitamine K du dicoumarol que vingt ans plus tard, en démontrant que l'administration de vitamine K contre les effets du dicoumarol [40].

Plusieurs autres inhibiteurs ont ensuite été développés à partir de trois noyaux principaux, la 4-hydroxycoumarine, la 4-hydroxy-thiocoumarine et l'indane-1,3-dione (**Figure 1.10**).

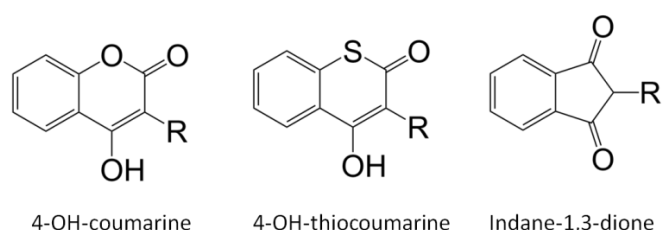


Figure 1.10. Structure chimique des trois noyaux à partir desquels sont dérivés les différents AVKs. La nature des groupements chimiques composant la chaîne R confère à chaque AVK ses spécificités physico-chimiques.

Les inhibiteurs de type AVK sont utilisés en médecine pour prévenir la survenue de phénomènes thromboemboliques, chez les patients portant des prothèses valvulaires cardiaques, souffrant de fibrillation auriculaire ou ayant subi un infarctus du myocarde. Les AVKs utilisés en thérapie sont dérivés des noyaux 4-hydroxycoumarine et indane-1,3-dione ; il s'agit du coumaphène, de l'acénocoumarol, de la fluindione (tous trois utilisés en France), ainsi que du tiocloumarol, de la phénindione et de la phenprocoumone.

Les AVKs sont des anticoagulants oraux empêchant le recyclage de la vitamine K par inhibition de VKORC1. Ils agissent au niveau de deux étapes du cycle de la vitamine, en bloquant la réduction de la vitK^E en vitK^Q d'une part, et la réduction de la vitK^Q en vitK^H d'autre part [18, 19]. Une inhibition potentielle de la NQO1 (**Figure 1.7**) par le dicoumarol a également été mise en évidence [41]. Le non renouvellement du stock de vitK^H résultant de cette inhibition provoque une diminution de l'activité de la GGCX. Les facteurs de coagulation restent donc à l'état inactif et incapables de se lier aux plaquettes agrégées lors de l'hémostase primaire. La formation des filaments de fibrine est finalement retardée et le temps de coagulation augmente. Un mécanisme d'inhibition détaillé a été proposé par Silverman en 1980, pour les AVKs dérivés du noyau 4-hydroxycoumarine. Il a suggéré une acylation par l'AVK d'un des deux résidus de cystéine du site actif de VKORC1, ou d'un acide aminé tiers nucléophile et spatialement proche du site actif. La formation d'un thioester à l'issue de cette réaction provoquerait ainsi une inhibition de nature irréversible (**Figure 1.11**), cohérente avec le mécanisme d'inhibition non compétitive, mis en évidence en 2005 [16].

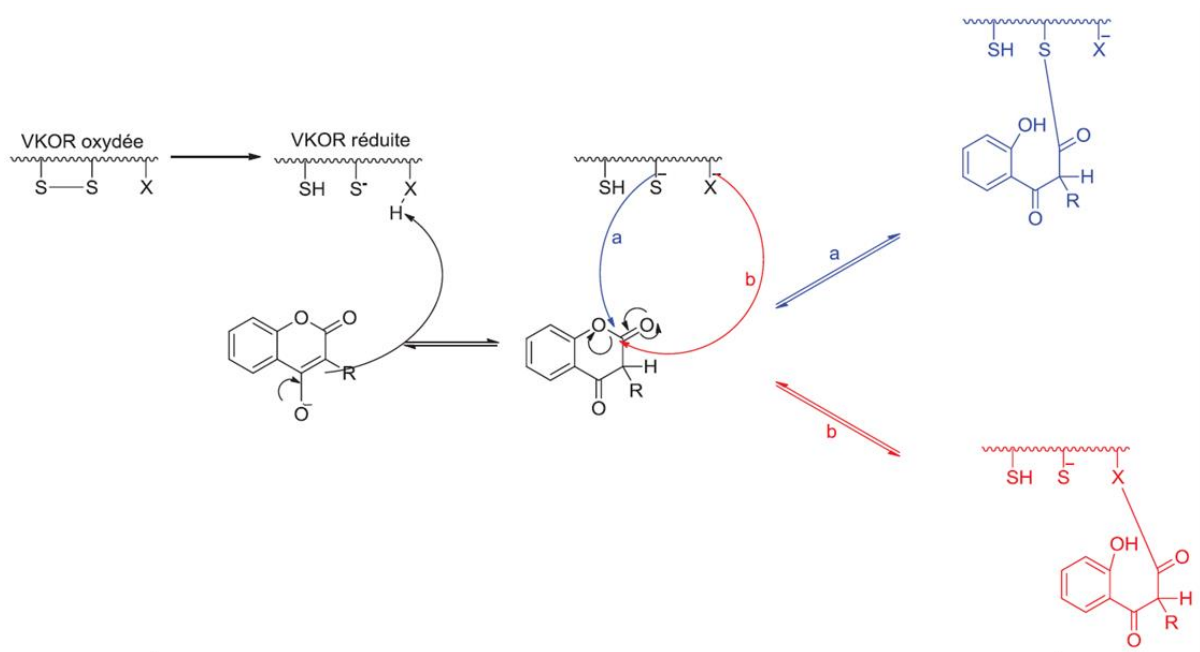


Figure 1.11. Mécanisme d'inhibition de VKORC1 par les AVKs. (a) Cas d'une acylation d'un des résidus de cystéine du site actif par l'AVK. (b) Cas d'une acylation d'un acide aminé nucléophile spatialement proche du site actif par l'AVK. Le thioester formé serait lié de façon covalente à l'enzyme, induisant le caractère irréversible de l'inhibition.

D'autres études, réalisées en 1983, suggèrent néanmoins une fixation du coumafène sur la forme oxydée de l'enzyme [42], contredisant la réaction décrite trois ans plus tôt par Silverman [24]. Le mécanisme concret de l'inhibition de VKORC1 par les AVKs reste donc incertain à l'heure actuelle.

6. Problèmes liés à l'utilisation des AVKs en thérapie humaine

Environ 900 000 personnes étaient sous traitement anticoagulant en France en 2007, soit plus d'1% de la population. Cependant, l'utilisation thérapeutique des anticoagulants de type AVK n'est pas sans risque. Une étude EMIR (Effets indésirables des Médicaments : Incidence et Risque) menée en 2007 a révélé que les accidents hémorragiques provoqués par les traitements aux AVKs constituaient la première cause d'hospitalisation due à des effets secondaires médicamenteux (environ 12% des cas, soit plus de 17 000 patients). La moitié de ces hospitalisations est provoquée par un mauvais usage des AVKs, mais l'autre moitié résulte d'une prise de traitement en conditions normales (étude EMIR 2007).

Plusieurs études ont donc été menées afin d'identifier les causes des variations de la réponse aux AVKs chez les patients sous traitement anticoagulant. De nombreuses analyses du gène *Vkorc1* ont mis en évidence le polymorphisme de ce gène. Certaines mutations ont pu être associées à un phénotype de résistance ou au contraire d'hypersensibilité aux AVKs, ainsi qu'à des modifications de l'activité VKOR catalysée par l'enzyme VKORC1 [16, 18, 43-48] (**Tableau 1.1**). D'autres mutations sont également soupçonnées de provoquer des modifications de l'activité VKOR ou de la sensibilité aux AVKs, mais leur effet n'a pu être établi avec certitude compte tenu des difficultés de caractériser *in vivo* l'effet des AVKs sur l'activité VKOR. Les mutations concernées sont L27T, V29L, D36T, D38T, V45A, S52L, S52W, W59R, W59L, V66M, V66G, G71A, N77S, N77T, L128R et R151Q.

Les mutations de VKORC1 identifiées chez l'homme induisent donc des effets très diversifiés sur la sensibilité de l'enzyme aux AVKs ainsi que sur son activité VKOR. Ces conséquences compliquent la mise en place de traitements anticoagulants chez certains patients, les réactions vis-à-vis des AVKs pouvant varier fortement d'un individu à l'autre. La compréhension des mécanismes induits par les mutations de VKORC1 sur la coagulation sanguine présente donc un intérêt majeur de santé publique.

Mutation	Activité VKOR	Sensibilité aux AVKs
A26P	--	R
A26T	+	R
H28Q	-	S
D36G	--	S
A41S	+	R
V54L	--	R
S56F	++	S
R58G	--	/
W59C	--	S
H68Y	+	R
I123N	+	R
Y139H	+	R

Tableau 1.1. Mutations de VKORC1 et phénotypes associés. Les effets induits par chaque mutation sur l'activité VKOR et sur la sensibilité de VKORC1 vis-à-vis des AVKs sont indiqués. Une activité VKOR fortement réduite (< 25% par rapport à la protéine sauvage) est indiquée par une coloration orange foncée et un signe "--" ; une activité VKOR légèrement réduite (entre 50 et 100% par rapport à la protéine sauvage) est indiquée par une coloration orange clair et un signe "-". Les mutations provoquant une résistance de VKORC1 aux AVKs sont indiquées par une coloration rouge et la lettre "R", tandis que les mutations provoquant une hypersensibilité de VKORC1 aux AVKs sont indiquées par une coloration verte et la lettre "S".

7. Problématiques liées à VKORC1

Suite à la description ci-dessus, on note quatre points essentiels.

1- l'enzyme VKORC1 joue un rôle indispensable dans le processus de coagulation sanguine, à travers sa fonction de recyclage de la vitamine K ;

2- l'inhibition de cette enzyme par des molécules de type AVK empêche le renouvellement du stock de vitamine K disponible dans l'organisme. L'administration orale d'AVKs constitue ainsi une stratégie privilégiée dans l'inhibition de la coagulation sanguine ;

3- le mécanisme de réduction de la vitK^E en vitK^Q par VKORC1 reste peu décrit, aucune étude n'ayant permis de confirmer définitivement le mécanisme réactionnel proposé par Silverman [24].

4- aucune structure 3D n'a été à ce jour résolue expérimentalement pour une VKORC1 mammalienne, ni sauvage ni mutée, compliquant d'autant plus la description structurale précise du fonctionnement de l'enzyme.

Au vu de l'état de l'art sur VKORC1, plusieurs questions restent ainsi sans réponse. L'enzyme VKORC1 adopte-t-elle un repliement tridimensionnel à trois ou quatre hélices transmembranaires chez les mammifères ?

Les résidus C₄₃ et C₅₁ sont-ils impliqués dans le mécanisme d'activation de VKORC1 ? Sur quelle conformation enzymatique de VKORC1 se fixe la vitK^E en conditions physiologiques ? Quels sont les effets structuraux des mutations modifiant l'activité VKOR et/ou la sensibilité de VKORC1 vis-à-vis des anticoagulants de type AVK ?

Quelques études par **modélisation moléculaire** ont été menées pour essayer de répondre à certaines de ces questions [2, 33, 35, 36]. Cependant, ces analyses restent peu approfondies. Les équipes de chercheurs ayant généré des modèles structuraux de VKORC1 ont construit des modèles dont le repliement membranaire correspond à l'organisation structurale suggérée par leurs propres résultats expérimentaux. Ces modèles impliquent des ponts disulfures formés *a priori*, par homologie directe avec l'enzyme bactérienne (bVKOR).

L'objectif des travaux présentés dans cette thèse est d'apporter des éléments de réponses aux questions posées ci-dessus, par des méthodes de modélisation moléculaire, tout en ayant la stratégie et l'approche les plus objectives possibles.

Nous avons débuté notre étude par une analyse de la séquence peptidique des VKORs, afin de chercher des protéines de séquence homologue disposant d'une structure 3D résolue expérimentalement. La structure 3D de la protéine identifiée a été utilisée comme '*template*' pour construire un modèle homologue de la VKORC1 humaine (hVKORC1). Nous avons postulé ce modèle à l'état entièrement protoné, les atomes de soufre des quatre résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ étant sous la forme SH, pour ne pas imposer la formation de ponts disulfures *a priori*.

Nous avons supposé que les simulations de dynamique moléculaire de ce modèle très flexible et sans contraintes produiraient un espace conformationnel très riche, dont les conformations seraient proche des états intermédiaires de l'enzyme. Nous avons utilisé des critères géométriques (distances S...S), probabilistes (populations des conformations) et d'énergie libre pour extraire ces conformations. En nous basant sur la prédiction de conformations-candidates, nous avons envisagé la construction des modèles 3D contenant des ponts S-S et leur exploration par simulations de DM. Une fois équilibrés, ces modèles peuvent servir de cibles pour étudier leur affinité avec la vitamine K.

La discrimination de leur sélectivité vis-à-vis de la vitamine K sous différentes formes permettra l'attribution du rôle fonctionnel de chaque modèle. La comparaison des affinités théoriques entre les modèles et la vitamine K ou des inhibiteurs (AVKs) avec les affinités mesurées expérimentalement permettra la validation des modèles et, par conséquent, du protocole proposé.

Selon ce protocole, différentes méthodes de modélisation moléculaire, simulation de dynamique moléculaire, docking, calcul d'énergie libre et de l'énergie de liaison ont été mises en applications et sont présentées dans la seconde partie de l'**INTRODUCTION**. Les algorithmes sous-jacents sont détaillés dans la partie **METHODES**. Les données générées par simulations DM ont été analysé par les méthodes statistiques classiques et par une nouvelle méthode développée au sein de notre laboratoire, le CMLA de l'ENS Cachan. Cette méthode avancée, basée sur l'analyse spectrale, a été développée au cours de l'analyse des données de simulation de DM de hVKORC1, en collaboration avec les mathématiciens, les Professeurs B. CHALMOND et A. TROUVE. L'ensemble des résultats de modélisation de hVKORC1 et des simulations de sa dynamique moléculaire sont expliqués dans la partie **RESULTATS**.

III. La modélisation moléculaire

La modélisation moléculaire consiste en la construction de **modèles 3D** à l'échelle microscopique (voire atomique) des molécules biologique. Elle permet ainsi d'analyser finement la structure des biomolécules *in silico*, à un niveau de description jusqu'à présent inaccessible ou limité pour les méthodes expérimentales (par exemple l'étude de la structure des protéines membranaires dans leur milieu naturel ou la caractérisation de complexes moléculaire). Les protéiques font partie des acteurs principaux des processus biologiques. Elles sont présentes chez tous les êtres vivants connus, quels que soient les domaines considérés (Archaeobactéries, Eubactéries et Eucaryotes). Les protéines, en dépit de la grande diversité de leurs rôles biologiques, sont toutes composées d'acides aminés, véritables briques élémentaires de la vie.

1. Les acides aminés

Au nombre de 20, les acides aminés sont composés d'un squelette d'atomes communs (peptide) et d'une chaîne latérale spécifique à chaque acide aminé, qui leur confère des propriétés physico-chimiques distinctes. Le fragment peptidique de chaque acide aminé est composé de 3 groupements : NH_2 - carbone alpha ($\text{C}\alpha$) - COOH . La chaîne latérale est toujours insérée sur l'atome de $\text{C}\alpha$ (**Figure 1.12**).

Le corps humain n'est capable de synthétiser que 11 acides aminés parmi les 20 ; les 9 autres sont dits "essentiels" et sont apportés par l'alimentation : phénylalanine, histidine, isoleucine, lysine, leucine, méthionine, thréonine, valine et tryptophane.

De plus, il existe deux acides aminés protéinogènes en plus des vingt cités ici : la pyrrolysine, présente uniquement chez certaines espèces d'Archaeobactéries, et la sélénocystéine, intervenant parmi quelques dizaines d'enzymes connues.

Le nombre d'acides aminés composant une protéine varie énormément (par exemple, 2 acides aminés pour l'aspartame, plus de 38 000 pour la titine). Chaque protéine présente plusieurs niveaux d'organisation structurale, de la structure primaire à la structure quaternaire.

2. Organisation structurale des protéines

La structure primaire d'une protéine correspond à la **séquence d'acides aminés** qui la composent (appelée chaîne peptidique). En effet, les atomes de la chaîne peptidique d'une protéine s'associent entre eux chimiquement. La liaison peptidique ainsi formée est une liaison covalente, très stable en conditions physiologiques et dont la rupture n'est effectuée que par des enzymes spécialisées. La formation de la liaison covalente peptidique est associée à une énergie importante, de l'ordre de 100 kcal/mol. La liaison peptidique s'effectue toujours entre le groupement COOH d'un acide aminé et le groupement NH₂ de l'acide aminé suivant. La liaison CONH formée est intrinsèquement plane en raison de sa nature chimique, mais deux rotations sont possibles autour des liaisons C α -CO de l'acide aminé i et NH-C α de l'acide aminé $i+1$. Deux angles dièdres peuvent ainsi être définis (**Figure 1.13**) :

- l'angle dièdre φ , dont le plan est défini par les atomes CO_(i-1)-NH_(i)-C α _(i)-CO_(i) ;
- l'angle dièdre ψ , dont le plan est défini par les atomes NH_(i)-C α _(i)-CO_(i)-NH_(i+1).

Certaines valeurs d'angles φ et ψ sont proscrites car elles induiraient des distances interatomiques trop courtes et seraient énergétiquement défavorables à la protéine.

Le deuxième niveau d'**organisation structurale des protéines** est représenté par les structures secondaires, pour lesquelles trois principaux types ont été définis : les hélices α , les brins β (qui peuvent s'associer entre eux pour former des feuilletts β) et les boucles.

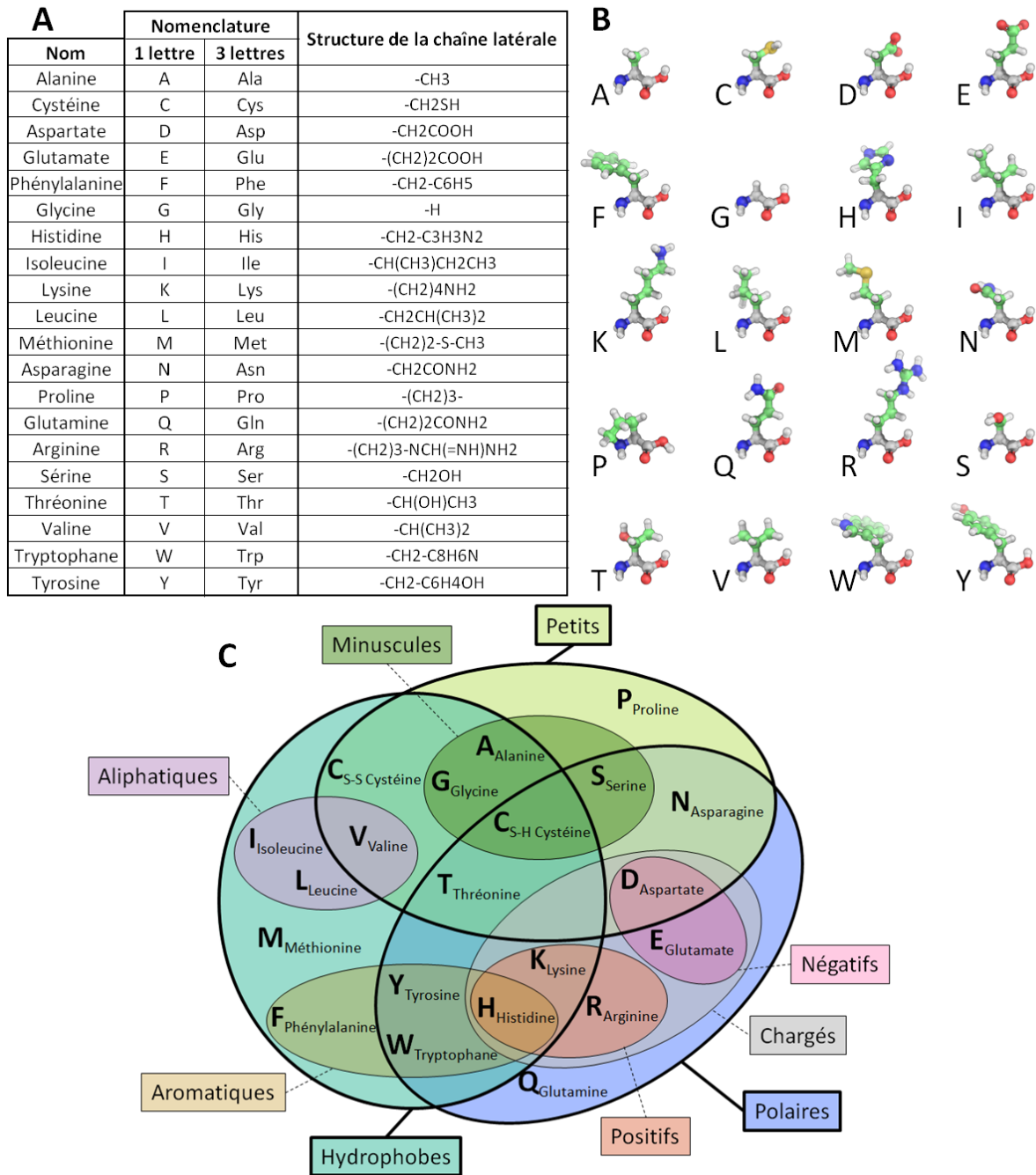


Figure 1.12. Propriétés physico-chimiques des acides aminés. (A) Tableau présentant les deux nomenclatures utilisées pour la notation des acides aminés, ainsi que la structure chimique des chaînes latérales. (B) Structure 3D des acides aminés. Les atomes sont représentés en sphères, les liaisons interatomiques sont représentées en *sticks* ; les atomes de carbone du squelette carboné et des chaînes latérales sont colorés en gris et en vert, respectivement. Les atomes d'azote, d'oxygène, de soufre et d'hydrogène sont colorés en bleu, rouge, jaune et blanc, respectivement. (C) Diagramme de Venn représentant les propriétés physico-chimiques de chaque acide aminé, induites par la diversité des chaînes latérales.

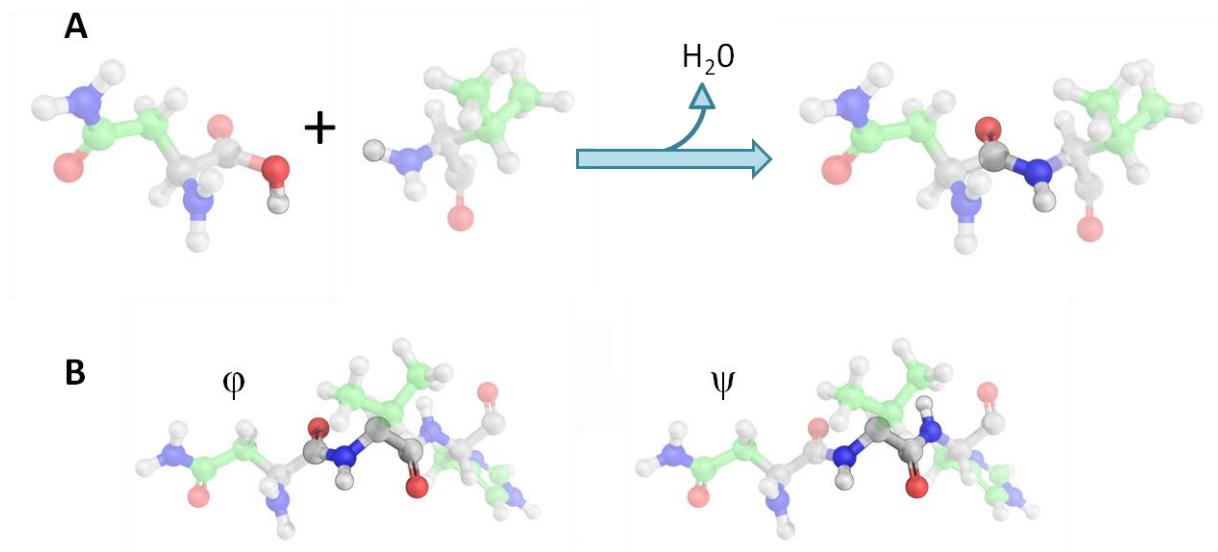


Figure 1.13. Formation d'une liaison peptidiques, angles dièdres φ et ψ . (A) La formation d'une liaison peptidique (CONH) entre deux acides aminés provoque la formation d'une molécule d'eau. (B) Les deux angles dièdres φ et ψ sont définis par les rotations possibles autour des liaisons C α -CO du résidu i et NH-C α du résidu $i+1$. Les atomes sont représentés en sphères, les liaisons interatomiques sont représentées en *sticks* ; les atomes de carbone du squelette carboné et des chaînes latérales sont colorés en gris et en vert, respectivement. Les atomes d'azote, d'oxygène et d'hydrogène sont colorés en bleu, rouge et blanc, respectivement. Seuls les atomes d'intérêt sont représentés en couleur opaque, les autres sont affichés en transparence.

Les hélices α correspondent à un enroulement de la chaîne peptidique sur elle-même, formant une spirale. Ce type de structure est maintenue par un réseau régulier de liaisons hydrogène (liaisons H) au niveau du squelette carboné de la protéine : le groupement CO du résidu i forme une liaison hydrogène avec le groupement NH₂ du résidu $i+4$. Le pas d'une hélice α est de 3,6 acides aminés (5,4 Å), toutes les chaînes latérales étant orientées vers l'extérieur de l'hélice (**Figure 1.14**). Les angles dièdres au sein des hélices α ont des valeurs moyennes de -60° pour φ et -45° pour ψ . Il existe d'autres types d'hélices, résultant d'une variation dans l'agencement du réseau de liaisons H. Ainsi, les hélices 3_{10} voient le groupement CO de l'acide aminé i former une liaison avec le NH₂ de l'acide aminé $i+3$, tandis que la formation d'une liaison H entre le groupement CO de l'acide aminé i et le groupement NH₂ de l'acide aminé $i+5$ aboutira à une structuration en hélice π . Les angles φ et ψ ont des valeurs moyennes de -50° et -25° dans les hélices 3_{10} et de -60° et -70° dans les hélices π , respectivement.

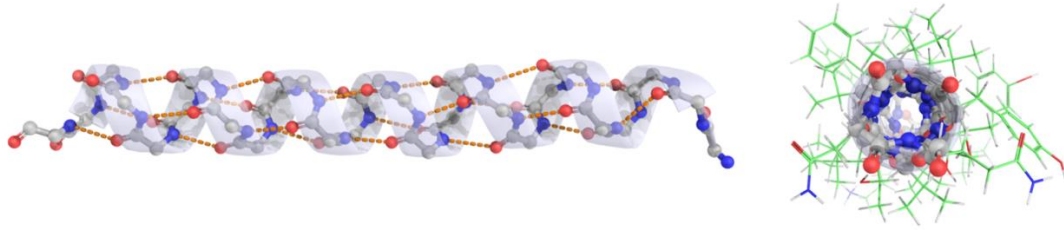


Figure 1.14. Structure tridimensionnelle d'une hélice α . Les liaisons hydrogène entre les atomes des résidus i et $i+4$ sont indiquées en pointillés oranges ; le squelette carboné est représenté en *cartoon* transparent (à gauche). L'orientation des chaînes latérales vers l'extérieur de l'hélice est mise en évidence (à droite). Les atomes du squelette carboné sont représentés en sphères, les liaisons interatomiques sont représentées en *sticks* ; les chaînes latérales sont représentées en ligne. Les atomes de carbone du squelette carboné et des chaînes latérales sont colorés en gris et en vert, respectivement. Les atomes d'azote, d'oxygène et d'hydrogène sont colorés en bleu, rouge et blanc, respectivement.

Les brins β sont également formés d'acides aminés contigus, dont les chaînes latérales sont distribuées alternativement de part et d'autre du squelette peptidique (les chaînes latérales des acides aminés i , $i+2$, $i+4$ etc. pointent dans une direction, celles des acides aminés $i+1$, $i+3$, $i+5$ etc. sont orientées à l'opposé). Plusieurs brins β peuvent ensuite s'associer pour former des feuillets β (parallèles ou antiparallèles), via un réseau de liaisons H toujours localisées au niveau du squelette peptidique (entre l'oxygène d'un acide aminé et l'azote d'un autre acide aminé situé sur le brin β voisin) ; les $C\alpha$ entre deux brins β ainsi associés sont donc adjacents (**Figure 1.15**). Les angles dièdres φ et ψ ont des valeurs moyennes de -120° et 115° au sein des feuillets β parallèles et de -140° et 135° dans les feuillets β antiparallèles, respectivement.

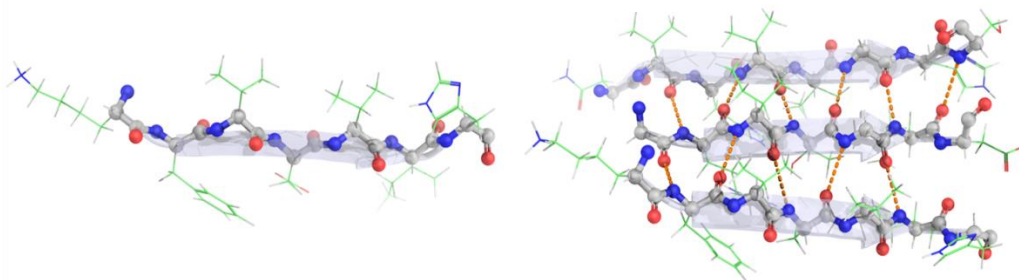


Figure 1.15. Structure tridimensionnelle d'un brin β et d'un feuillet β . Organisation des résidus dans un brin β (à gauche) et assemblage de plusieurs brins β en feuillet β (à droite). Les liaisons hydrogène entre les atomes du squelette carboné sont indiquées en pointillés oranges ; le squelette peptidique est représenté en *cartoon* transparent. L'orientation des chaînes latérales de part et d'autre du feuillet est mise en évidence. Les atomes du squelette peptidique sont représentés en sphères, les liaisons interatomiques sont représentées en *sticks* ; les chaînes latérales sont représentées en ligne. Les atomes de carbone du squelette peptidique et des chaînes latérales sont colorés en gris et en vert, respectivement. Les atomes d'azote, d'oxygène et d'hydrogène sont colorés en bleu, rouge et blanc, respectivement.

L'énergie associée à une liaison hydrogène seule est relativement faible, entre 0,5 et 3,5 kcal/mol. Cependant, les hélices α et les feuillets β sont stabilisés par de nombreuses liaisons H au niveau du squelette peptidique, dont la résultante énergétique est plus élevée. Les hélices α , brins β et feuillets β forment donc des régions semi-rigides au sein des protéines. En revanche, les boucles (*coil*), troisième catégorie de structure secondaire, montrent une absence de structuration particulière : aucun motif de liaisons H n'y est présent et les angles dièdres ne présentent pas de valeurs spécifiques.

Les boucles constituent ainsi des zones très flexibles et mobiles au sein des protéines, menant à des changements conformationnels locaux ou globaux

Certains acides aminés présentent une propension naturelle à former un type de structure secondaire défini [49]. Sur la base de ces observations, plusieurs méthodes ont été développées dans le but de prédire les structures secondaires d'une protéine à partir de sa structure primaire. Ces outils sont présentés dans la partie **METHODES**.

L'agencement des diverses structures secondaires les unes par rapport aux autres induit un troisième niveau d'organisation structurale des protéines, **la structure tertiaire**, extrêmement importante puisqu'elle confère aux protéines leur activité biologique. En effet, le repliement tridimensionnel de la chaîne peptidique crée un relief, une surface protéique définie ; certains acides aminés se retrouvent ainsi exposés à la surface de la protéine, d'autres sont enfouis, des cavités apparaissent, aboutissant à la formation de domaines protéiques (possédant chacun une fonction propre) et du ou des site(s) actif(s) (**Figure 1.16**).

La cohésion de la structure tertiaire des protéines est assurée par diverses liaisons. Les ponts disulfures, des liaisons covalentes entre atomes de soufre portés par deux résidus de cystéine, sont associés à une énergie d'environ 70 kcal/mol. Les interactions électrostatiques surviennent au niveau des particules élémentaires chargées (électrons et protons), provoquant une action attractive ou répulsive. Elles sont décrites par la loi de Coulomb, proposée au début du XVIII^e siècle par Charles-Augustin Coulomb. Cette loi permet d'estimer le gain (-) ou la perte (+) d'énergie lorsque l'on rapproche deux charges ponctuelles z_1e et z_2e de l' ∞ à la distance r . L'énergie $E = \pm \frac{z_1 z_2}{(4\pi\epsilon_0)\epsilon r} e^2$ associée à ce type d'interaction est faible, aux alentours de 2 kcal/mol par résidu.

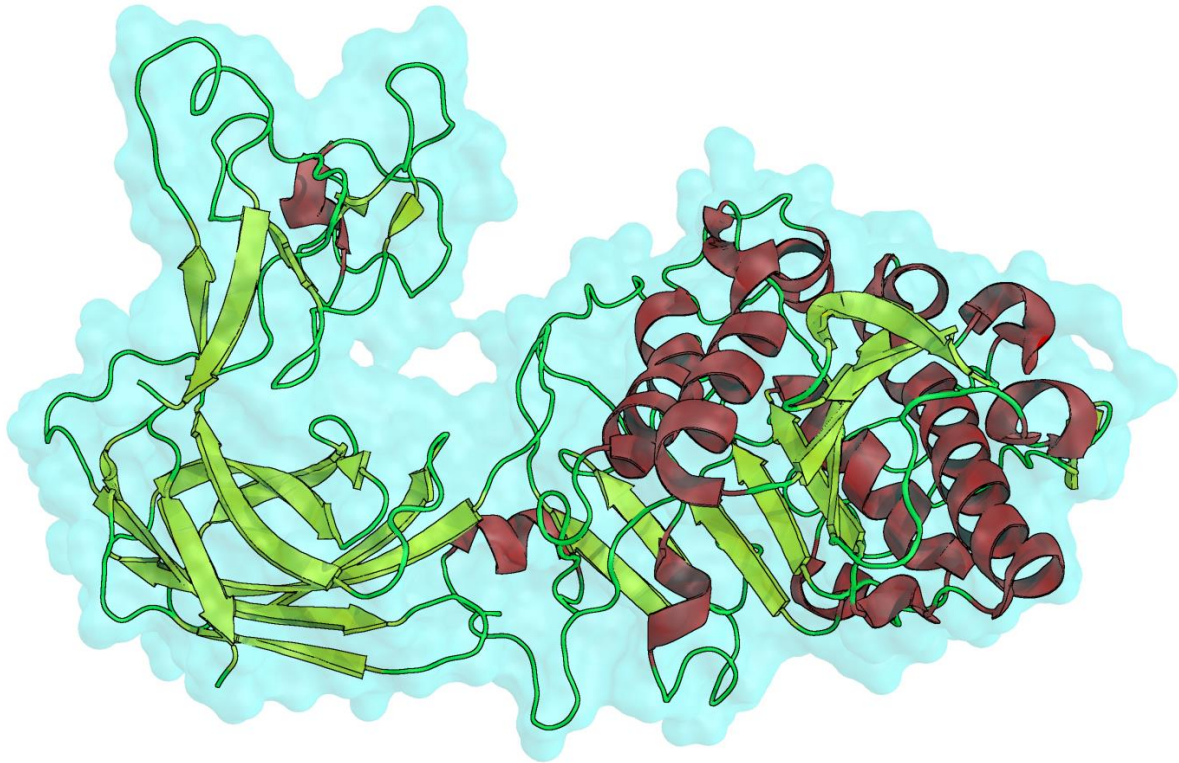


Figure 1.16. Structure tertiaire d'une protéine. La protéine est représentée en *cartoon*, mettant en évidence les différentes structures secondaires qui la composent. Les hélices α sont colorées en rouge, les feuillets β en jaune et les boucles en vert. La surface de la protéine est représentée en transparence est colorée en cyan.

Les interactions entre les électrons de deux atomes induisent des forces de van der Waals $E_{vdw} = -\frac{1}{r^6} \frac{Z\mu_1 Z\mu_2}{(4\pi\epsilon_0\epsilon)^2}$, dont l'énergie associée est également faible (environ 1 kcal/mol). Les forces de van der Waals ont trois origines : l'interaction électrostatique attractive ou répulsive entre deux multipôles permanents, selon leurs orientations (effets d'orientation ou forces de Keesom) ; l'interaction attractive entre un multipôle permanent et un multipôle induit (effets d'induction, forces de Debye) ; l'interaction électrostatique attractive entre deux multipôles induits (effets de dispersion, forces de London). Enfin, la nature hydrophobe ou hydrophile des acides aminés joue également un rôle important dans le repliement 3D des protéines. En effet, une protéine cytoplasmique possédera un cœur hydrophobe et une surface hydrophile, tandis qu'une protéine membranaire verra ses acides aminés hydrophiles enfouis et présentera une surface hydrophobe. L'énergie associée à une interaction hydrophobe entre deux résidus est inférieure à 1 kcal/mol.

Chacune des interactions non-covalentes participant à la structure tertiaire des protéines présente une énergie associée relativement faible. En revanche, ces interactions sont très nombreuses au sein des protéines, et la résultante de toutes ces forces constitue une énergie non négligeable, assurant le maintien de la structure 3D des protéines.

Enfin, certaines protéines disposent d'un niveau de repliement supplémentaire : la **structure quaternaire**. Il s'agit de protéines multimériques (également appelées oligomères), formées de plusieurs chaînes peptidiques (sous-unités ou monomères, **Figure 1.17**). Les liaisons assurant le maintien des différentes sous-unités entre elles sont du même type que celles intervenant dans les structures tertiaires, bien que les ponts disulfures interviennent plus rarement ; les interactions hydrophobes y sont prédominantes.

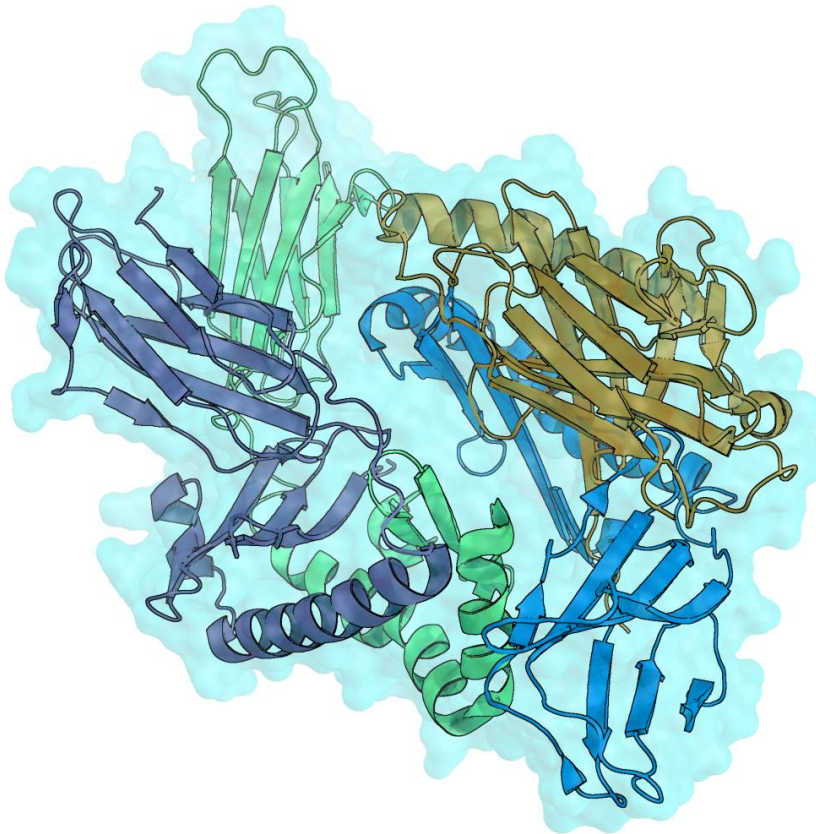


Figure 1.17. Structure d'une protéine. La protéine est représentée en *cartoon*, mettant en évidence les différentes structures secondaires qui la composent. Elle est composée de quatre sous-unités identiques, colorées respectivement en violet, vert, orange et bleu. Il s'agit donc d'un homotétramère. La surface de la protéine est représentée en transparence est colorée en cyan.

Le nombre de monomères formant une protéine oligomérique varie, de même que leur analogie ou leur différence. Par exemple, une protéine formée de deux monomères identiques sera nommée homodimère, tandis qu'une protéine composée de quatre sous-unités différentes sera qualifiée d'hétérotétramère.

Il existe donc un lien entre la séquence d'acides aminés et la structure tridimensionnelle d'une protéine, puisque la nature physico-chimique des résidus composant une protéine influence son organisation en structures secondaires, puis sa structure tertiaire, voire quaternaire le cas échéant (**Figure 1.18**). Cette relation séquence-structure présente un intérêt important, notamment pour l'étude de protéines dont la structure 3D est *a priori* inconnue (voir **METHODES**).

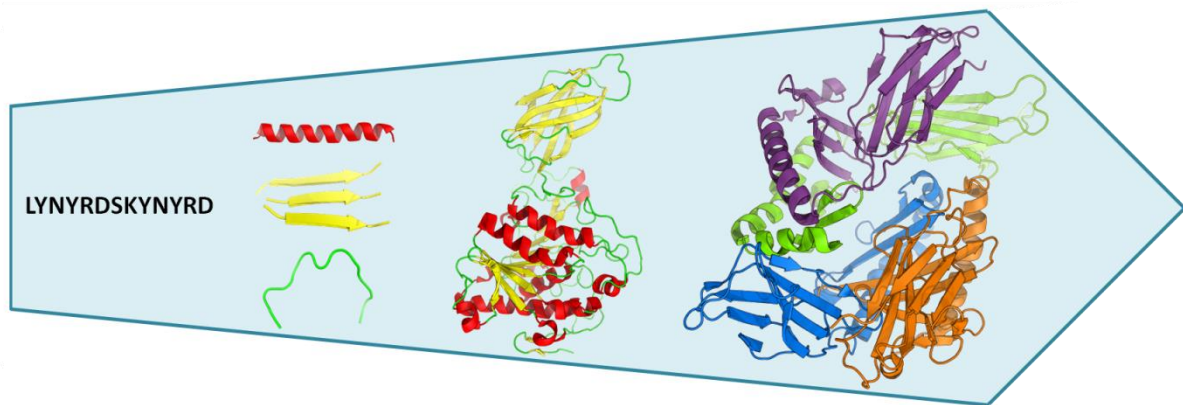


Figure 1.18. Relation entre les différents niveaux d'organisation structurale des protéines. La séquence peptidique (structure primaire) influence directement les structures secondaires, qui forment ensuite la structure tertiaire en s'agencant les unes par rapport aux autres. Certaines protéines, dites multimériques, sont composées de plusieurs chaînes peptidiques distinctes, associées entre elles pour former la structure quaternaire.

3. Visualisation et représentations de la structure 3D des protéines

Les **structures tridimensionnelles (3D)** de certaines protéines ont été résolues expérimentalement par cristallographie - diffraction aux rayons X, résonance magnétique nucléaire (RMN) ou microscopie électronique ; pour d'autres protéines, des modèles 3D sont construits à partir des structures 3D résolues pour des protéines semblables (voir **METHODES**). L'ensemble des structures 3D ainsi générées est collecté dans des bases de données structurales et accessibles via internet. Les fichiers des structures protéiques contiennent cependant une information limitée (séquence peptidique, coordonnées atomiques des atomes) et peu simple à interpréter directement en terme de représentation 3D. Des outils de visualisation tels que PyMOL [50], VMD [51], Maestro [52] et Chimera [53] ont donc été développés afin de faciliter la représentation et l'analyse visuelle en trois dimensions des structures protéiques. Ces logiciels permettent d'appréhender les structures 3D des protéines, d'identifier rapidement les structures secondaires présentes et le positionnement des domaines d'intérêt biologique, de mettre en évidence les résidus importants (d'après les données expérimentales) et d'afficher des informations supplémentaires (charges des résidus, potentiels électrostatiques). La visualisation 3D des protéines est indispensable à la compréhension de leur mode de fonctionnement ; par exemple, la comparaison de deux structures 3D d'une même protéine résolues en présence de deux ligands différents (ou de la protéine avec et sans ligand) permet d'observer rapidement les changements structuraux induits sur une protéine par un ligand donné.

Plusieurs types de représentations des structures 3D des protéines ont été mis en place, selon l'échelle à laquelle la protéine est étudiée (**Figure 1.19**). La représentation du squelette peptidique seul peut suffire pour une analyse de la structure globale d'une protéine (repliement), tandis que la visualisation des chaînes latérales des résidus est indispensable à l'étude des interactions entre une protéine et un ligand. Les observations ainsi effectuées permettent d'obtenir rapidement des informations pertinentes sur une protéine et peuvent orienter les travaux futurs, notamment dans le cas de projets exploratoires (par exemple dans le cas d'une protéine dont la structure 3D vient d'être résolue expérimentalement pour la première fois).

Toutefois, la structure d'une protéine obtenue par cristallographie ou par modélisation ne permet pas d'appréhender ses **changements conformationnels**, limitant le niveau d'information sur les relations « structure-fonction » de la cible. Des méthodes de **simulation de dynamique moléculaire** peuvent être utilisées dans ce contexte.

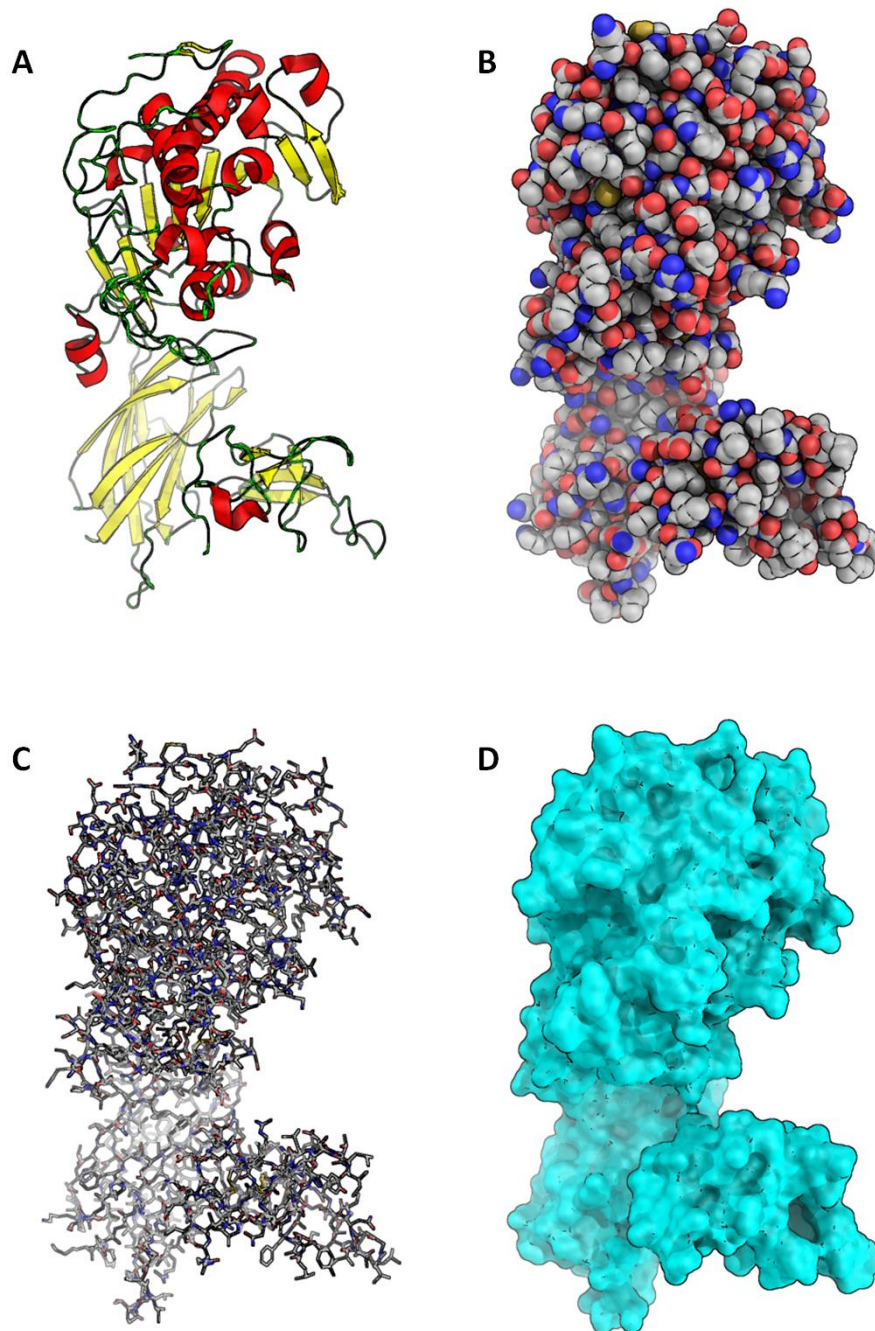


Figure 1.19. Principaux types de représentations 3D des protéines utilisées en modélisation moléculaire.

(A) Représentation en *cartoon* d'une protéine. Cette représentation n'affiche que le squelette carboné de la protéine et permet d'identifier rapidement les structures secondaires présentes (les hélices α sont colorées en rouge, les feuillets β en jaune et les boucles en vert). (B) Représentation en sphères d'une protéine. Cette représentation tout atome donne une idée du volume globale de la protéine, le rayon de chaque sphère correspondant au rayon de van der Waals de l'atome associé. (C) Représentation en *sticks* d'une protéine. Cette représentation tout atome permet d'observer toutes les liaisons interatomiques covalentes présentes dans la protéine. Les atomes de carbone, d'oxygène, d'azote et de soufre sont colorés en gris, rouge, bleu et jaune, respectivement. (D) Représentation en surface d'une protéine. Cette représentation permet d'identifier rapidement les cavités présentes au sein d'une structure protéique et pouvant potentiellement accueillir un ligand. La surface est colorée en cyan.

Elles offrent des moyens de caractériser les **propriétés dynamiques et énergétiques** des molécules et ouvrent l'accès à des échelles de temps et d'espace difficiles (ou impossibles) à sonder expérimentalement. Ces simulations numériques permettent notamment de décrire les propriétés structurales et dynamiques des molécules biologiques, en relation avec leurs fonctions établies et supposées. La dynamique moléculaire, combinée aux données expérimentales, contribue à une meilleure compréhension des mécanismes moléculaires des processus physiopathologiques. Elle est indispensable pour interpréter les données expérimentales et cliniques, formuler et examiner des hypothèses, stimuler le développement de nouvelles approches d'analyse, et joue un rôle important dans la recherche de nouveaux médicaments.

4. La dynamique moléculaire

Tous les acteurs du monde du vivant présentent une certaine flexibilité, chacun à leur échelle. Ainsi, un organisme vivant doit se mouvoir pour survivre (les plantes s'orientent par rapport au soleil, les animaux se déplacent pour se nourrir et se reproduire). De la même façon, les cellules composant un organisme peuvent se réarranger les unes par rapport aux autres (migration cellulaire dans le cas de la cicatrisation, phagocytose pour la réponse immunitaire). Enfin, les molécules présentes dans les cellules vivantes possèdent également des propriétés de déformation : association des filaments de myosine et d'actine puis basculement des têtes de myosine dans le cadre de la contraction musculaire, interaction des facteurs de transcription avec l'ADN pour la régulation génique, changements de conformation pour certaines réactions enzymatiques. Cette flexibilité des entités vivantes leur confère une capacité à s'adapter en permanence à leur environnement et constitue ainsi une des clés de la survie des individus.

Ainsi, une protéine présente une **plasticité conformationnelle**, lui permettant de s'associer à d'autres molécules (protéines, lipides, ADN, ARN, petites molécules) et dont dépend ainsi la fonction de cette protéine. Or, les structures 3D disponibles dans la *Protein Data Bank* (PDB) [54] ou modélisées *in silico* représentent un état statique d'une molécule, correspondant à une conformation moyenne (soit la plus forte probabilité d'observer chaque atome à un positionnement donné). Une méthode est donc indispensable pour reproduire les mouvements atomiques (et résiduels) et donc le comportement dynamique fonctionnel des molécules étudiées : la dynamique moléculaire (DM).

Les simulations de DM permettent d'accéder aux échelles spatiale et temporelle inhérentes aux biomolécules, présentant ainsi un apport considérable aux techniques *in vitro* et *in vivo* en visant à reproduire le comportement dynamique des biomolécules à leur propre échelle. La première simulation de DM a été réalisée en 1957 par Alder et Wainwright et consistait en la simulation des interactions entre les atomes composant un fluide. Chaque atome était considéré comme une unique sphère dure, ne subissant ni déformation ni perte d'énergie lors de collisions. Ces sphères n'exerçaient aucune force à distance les unes sur les autres ; lors d'un contact, une force de répulsion était induite, empêchant les volumes des deux sphères de se recouvrir. Rahman introduit en 1964 un potentiel continu dans les simulations, plus fin car tenant compte des nuages électroniques de atomes et de leurs contributions attractives et répulsives. La première simulation de DM (de 9,2 ps) d'une protéine (Bovine Pancreatic Trypsin Inhibitor, environ 500 atomes) dans le vide a été effectuée par A. MacCammon en 1976.

En quarante ans, les progrès techniques en informatique ont abouti à une croissance remarquable de la taille des systèmes étudiés et des temps de simulation. Outre la biomolécule d'intérêt, les simulations de DM actuelles prennent également son environnement en considération : molécules d'eau, ions, membrane lipidique si nécessaire, partenaire moléculaire éventuel (protéine, acide nucléique, petite molécule). Ainsi, des simulations de DM de systèmes de plusieurs centaines de milliers d'atomes, sur des durées de plusieurs centaines de nanosecondes (ns), sont de nos jours considérées comme classiques (le premier article relatant une simulation de DM de l'ordre de la microseconde (μ s) a été publié en 1998 par Duan et Kollman [55]). Par exemple, l'équipe de David Shaw effectue des simulations de l'ordre de la milliseconde (ms) grâce aux performances du supercalculateur Anton [56], accédant ainsi à l'échelle de temps correspondant aux transitions conformationnelles des protéines de grande taille. Cette évolution souligne l'amélioration constante des simulations de DM, en termes de vitesse de calculs, de temps de simulations et de taille et complexité des systèmes étudiés.

Trois méthodologies majeures ont été développées en DM : la mécanique moléculaire (MM), de la mécanique quantique (MQ) et des simulations gros grains (GG). Chacune de ces méthodes présente des avantages et des inconvénients spécifiques. Les simulations en MM permettent d'accéder à des échelles de temps de l'ordre de la centaine de nanoseconde (voire de la microseconde) mais ne rendent pas compte de la formation ou rupture des liaisons covalentes pouvant survenir durant les simulations de DM.

Les simulations en MQ reproduisent cette évolution des liaisons covalentes durant les simulations de DM, mais sont impossibles à mener à des échelles de temps de plusieurs nanosecondes, sur des systèmes contenant plusieurs milliers d'atomes.

Des méthodes hybrides combinant MQ et MM ont été mises en place afin de réunir les avantages des deux méthodes précédentes. Elles consistent en la description d'une petite partie du système (région d'intérêt, par exemple le site actif) en MQ, l'essentiel du système étant traité en MM. Cette méthode hybride permet d'effectuer des simulations à l'échelle de la nanoseconde, tout en intégrant la MQ à la zone d'intérêt choisie au départ. Enfin, les simulations de DM en gros grains consistent à représenter plusieurs atomes du système par une seule particule, nommée gros grain. Le nombre de particules composant le système est ainsi réduit, ces simulations en GG permettent ainsi d'atteindre des échelles de temps supérieures à la microseconde, voir à la milliseconde. Ce type de simulation provoque cependant la perte de la description tout atome du système simulé. Ces trois méthodes, bien que différentes, reposent toutes sur le même principe physique : les mouvements des atomes au cours des simulations de DM sont gouvernés par les forces interatomiques exercés dans le système étudié, décrites par un champ de force.

Les paramètres des champs de force sont semi-empiriques pour les simulations de MM, c'est-à-dire obtenus expérimentalement.

En revanche, les paramètres des champs de forces appliqués en simulation de MQ utilisent un potentiel énergétique issu d'études théoriques, considérant la structure électronique des atomes.

La méthode utilisée pour effectuer les simulations de DM présentées dans ces travaux est la DM dite "classique", soit la DM en MM. Tous les atomes composant le système à simuler sont considérés comme **un modèle mécanistique**. La dynamique d'un tel système peut, par approximation, être décrite par la résolution de la seconde loi de Newton :

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2}, i = 1 \dots N$$

où F_i représente l'ensemble des forces appliquées sur l'atome i , m_i est la masse de l'atome i , le vecteur $r_i = (x_i, y_i, z_i)$ représente les coordonnées de l'atome i , t le temps et N le nombre d'atomes composant le système étudié.

Les forces F_i appliquées sur un atome i à un instant t sont calculées à partir de l'équation suivante :

$$F_i(t) = - \frac{\partial V(r_1(t), r_2(t), \dots, r_N(t))}{\partial r_i(t)}$$

Une simulation de DM en approximation MM implique toujours la résolution simultanée de ces deux équations à chaque instant t de la simulation, par itération. Un schéma d'intégration doit donc être appliqué afin de calculer la trajectoire de chaque atome à intervalles de temps réguliers (discrétisation du temps). L'algorithme de Verlet [57] est l'un des plus utilisés en dynamique moléculaire, car il présente l'avantage de la simplicité et de l'efficacité des calculs (les vitesses ne sont pas conservées, seules les positions et forces appliquées aux atomes à chaque itération sont stockées et utilisées pour calculer les positions atomiques au temps suivant, ce qui induit une réversibilité dans le temps). Le pas d'intégration, noté δt , doit être inférieur à la période du mouvement de plus haute fréquence décrit par le système (soit la fréquence liée à l'élongation des liaisons covalentes C-H, N-H et O-H). δt sera donc de l'ordre de la femtoseconde (10^{-15} s).

Le mouvement concerté de tous les atomes décrit ainsi une trajectoire dans **un espace conformationnel**. Une simulation de dynamique moléculaire fournit un ensemble de "*snapshots*", qui apparaissent dans l'espace conformationnel comme un nuage de vecteurs conformationnels (**Figure 1.20**).

La fonction F_i représentant les forces appliquées à l'atome i est décrite par les équations et paramètres du champ de force, incluant les termes associés aux interactions liées (liaisons covalentes) et non liées (interactions électrostatiques et de van der Waals). L'équation globale d'un champ de force classique (MM en tout atome) est la suivante :

$$E_{\text{totale}} = E_{\text{liées}} + E_{\text{non liées}}$$

où $E_{\text{liées}} = E_{\text{élongation}} + E_{\text{angle}} + E_{\text{dièdre}}$

et $E_{\text{non liées}} = E_{\text{électrostatiques}} + E_{\text{van der Waals}}$

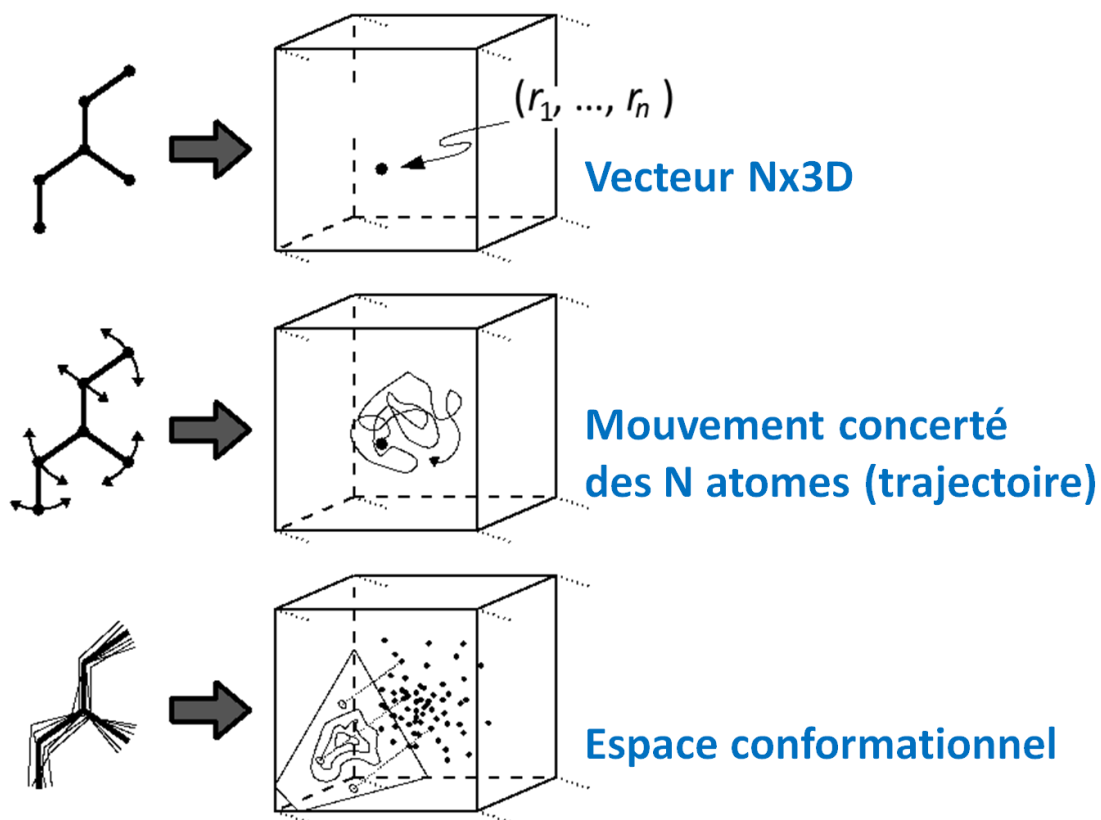


Figure 1.20. Trajectoire de DM d'une protéine composés de N atomes. Les mouvements atomiques décrits par la biomolécule étudiée entre l'instant initial ($t=0$) et l'instant $t+1$ sont obtenus grâce à un jeu de vitesses appelées vitesses initiales, généralement choisies aléatoirement. Ces vitesses initiales constituent la "pichenette" de départ, mettant tout le système en mouvement (la structure initiale de la biomolécule étant une conformation moyenne correspondant à une probabilité d'observation des positionnements de ses atomes).

Les champs de force peuvent contenir d'autres termes et paramètres afin d'affiner la description physique du système étudié, tels que les angles dièdres impropres (qui permettent la conservation du caractère planaire des cycles aromatiques) ou des termes de polarisation (dans le cas des champs de force polarisables, qui tiennent compte de l'influence des particules voisines dans le calcul d'une charge pour une particule donné). Ces termes supplémentaires distinguent les différents champs de force développés pour la DM, tout comme les paramètres associés aux différentes fonctions (par exemple l'énergie de déformation angulaire E_{angle} pour des atomes donnés). Le champ de force utilisé, les différents paramètres ainsi que les éléments spécifiques aux protéines membranaires et appliqués aux simulations de DM de VKORC1 sont présentés dans la partie **METHODES**.

Les simulations de DM en MM, malgré leur puissance, présentent des limites. En premier lieu, l'espace conformationnel d'une protéine exploré par DM classique est très restreint par rapport à la réalité biologique. En effet, les temps de simulation nécessaires à une exploration exhaustive de l'espace conformationnel d'une protéine sont inaccessibles par les moyens de calculs actuels. La génération de plusieurs trajectoires pour une même protéine, en faisant varier les vitesses initiales, permet de combler partiellement cette lacune en explorant différents sous-espaces conformationnels, donc différentes conformations représentatives de la protéine. Plusieurs alternatives ont été mises en place pour augmenter l'espace conformationnel exploré par DM. La méta-dynamique consiste à ajouter un potentiel limitant le retour du système à des états énergétiques déjà visités. Les simulations de DM dirigées guident l'évolution de la protéine de sa structure initiale jusqu'à une conformation finale. Cependant, les conformations adoptées par la protéine au cours de ce type de simulations ne sont pas toujours celles de plus faible énergie (les plus favorables structurellement).

Ensuite, les simulations de DM en approximation MM négligent les effets quantiques, tels les transferts de protons. Des contraintes sont en effet appliquées aux liaisons covalentes impliquant des atomes d'hydrogène (C-H, N-H et O-H) durant les simulations de DM classiques, afin de limiter leurs mouvements vibratoires. La définition des interactions non liées est également biaisée durant les simulations de DM classiques (elles sont par exemple tronquées par fixation d'une distance seuil) afin de réduire les temps de calcul.

Enfin, les conditions expérimentales appliquées aux structures à partir desquelles les paramètres composant les champs de force ont été déterminés varient. Les termes des champs de force peuvent ainsi provoquer des distorsions au niveau de la structure atomique du système simulé. Les différents champs de force sont en permanence rectifiés, améliorés, enrichis et réédités.

Chapitre 2. METHODES

I. Banques de données

L'avènement de l'informatique et le développement d'internet ont permis la création de bibliothèques recensant l'ensemble des connaissances générées (dans divers domaines, dont la bioinformatique) et leur mise à disposition (actuellement, hébergées sur des serveurs et disponibles en ligne ou installées localement). Les informations stockées dans ces banques de données sont vérifiées, validées, triées et organisées, assurant leur validité et leur accessibilité. Les logiciels et filtres interfacés facilitent l'accès à des éléments précis et permettent leur analyse (préliminaire ou détaillée). Les deux banques de données les plus utilisées concernant les protéines sont *UniProt*, contenant des informations sur leurs séquences peptidiques, et la *Protein Data Bank* qui recense leurs structures tridimensionnelles.

1. UniProt Knowledgebase

UniProt[58], abréviation de *Universal Protein Resource*, est une banque de données recensant les séquences peptidiques de milliers de protéines réparties dans tous les domaines du vivant. En juillet 2016, plus de 500 000 séquences avaient été manuellement annotées et examinées (vérification des informations dans la littérature afin d'assurer l'intégrité scientifique des données) et plus de 65 millions d'entrées étaient en attente de validation.

Parmi toutes les séquences protéiques disponibles, les données pour certaines espèces sont plus représentées que d'autres. *Homo sapiens* se classe par exemple à la onzième place, avec plus de 134 000 séquences peptidiques identifiées et mises à disposition. En comparaison, c'est le virus de l'immunodéficience humaine 1 (VIH-1) qui arrive en première position (plus de 700 000 entrées), malgré une différence de complexité de génome notable entre les deux espèces.

Outre les séquences d'acides aminés, *UniProtKB* indique, lorsque les données sont disponibles, les fonctions biologiques propres à chaque protéine, les pathologies dans lesquelles elles sont impliquées, les mutations (polymorphisme naturel, mutations induites par des facteurs environnementaux ou par les expérimentateurs) et leurs effets sur l'activité ou d'autres fonctions protéiques.

L'expression tissulaire de chaque protéine, sa localisation au sein de la cellule (membranaire, cytoplasmique, nucléaire...), les éventuelles modifications post-traductionnelles (acides aminés phosphorylés, glycosylés, ponts disulfures), les particularités topologiques (domaines transmembranaires, intracellulaires, extracellulaires) ainsi que les acides aminés composant le ou les site(s) actif(s) sont également répertoriés. Enfin, *UniProtKB* fournit plusieurs liens vers d'autres bases de données (structurales, protéomiques, interactions protéine-protéine, voies de signalisation...), les listes de publications traitant de chaque séquence et des groupes de séquences protéiques similaires (en terme d'identité de séquence).

2. Protein Data Bank

La *Protein Data Bank* (PDB) [54] est un répertoire mondial de dépôt d'informations sur les structures tridimensionnelles des protéines et des acides nucléiques, résolues expérimentalement par cristallographie, spectroscopie par résonance magnétique nucléaire ou microscopie électronique cryogénique (*cryo-EM*). A ce jour (16 janvier 2017, le nombre de structures tridimensionnelles disponibles dans la PDB s'élève à 125 795 (**Figure 2.1**). Parmi elles, environ 115 000 structures concernent des protéines, 3 000 des acides nucléiques et 6 000 des complexes protéine-acide nucléique. *Homo sapiens* est l'espèce présentant le plus de biomolécules dont la structure 3D est résolue (plus de 33 000 structures), suivie par *Escherichia coli* (aux alentours de 8 000 structures). La taille des molécules varie de quelques acides aminés (poids moléculaire de 300 kDa environ) pour les plus petites protéines jusqu'à des complexes moléculaires de très grande taille (poids moléculaire supérieur à 95 000 000).

La PDB est également interconnectée à divers outils de visualisation et d'analyse des données structurales, directement disponibles en ligne. Elle contient également des interfaces avec d'autres banques de données (banques de séquences ou structurales) ou d'autres outils d'analyse (tel que BLAST).

Le format des données structurales de la PDB (extension .pdb) de chaque protéine est standardisé et représente les coordonnées des atomes de chaque acide aminé dans l'espace cartésien. Les fichiers pdb contiennent également des informations biologiques sur la biomolécule concernée (structures secondaires, espèce à laquelle elle appartient, fonction biologique, localisation cellulaire) ainsi que sur les conditions expérimentales d'obtention de la structure tridimensionnelle (température, résolution, la symétrie...).

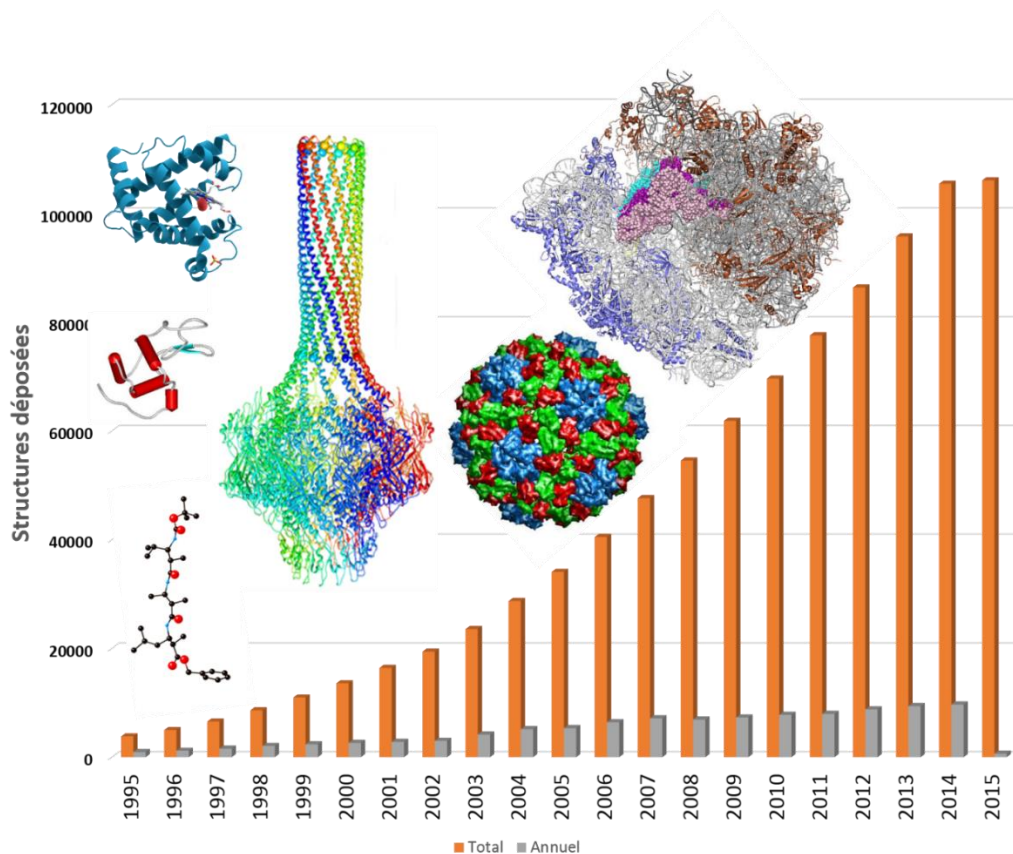


Figure 2.1. Evolution du nombre et complexité croissante des structures présentes dans la Protein Data Bank. Le nombre total par année et le nombre de structures déposées chaque année sont indiqués en orange et en bleu, respectivement. Les structures des protéines simples aux complexes macromoléculaires sont illustrées par un peptide, une globine, le lysozyme, un bactériophage, un virus et un ribosome.

Les paramètres de chaque atome de la biomolécule sont ensuite consignés dans l'ordre suivant (1 atome en 2 lignes) : numéro de l'atome, nom de l'atome, nom de l'acide aminé auquel il appartient (nomenclature 3 lettres), chaîne sur laquelle il est situé (dans le cas de complexes protéiques ou protéine-acide nucléique), numéro de l'acide aminé dans la séquence peptidique de la biomolécule, coordonnées orthonormées X, Y et Z de l'atome, facteur d'occupation de l'atome (indique la probabilité de localisation de l'atome à cette position), facteur de température isotrope ou anisotrope (indique la sphère ou l'ellipsoïde caractérisant le positionnement d'un atome), symbole de l'élément atomique, charge de l'atome.

II. Alignements de séquences

La séquence peptidique d'une protéine contient des informations primordiales, puisqu'elle influence directement sa structure (secondaire, tertiaire et éventuellement quaternaire). Différents outils d'analyse ont donc été développés afin de comparer les séquences de protéines distinctes, par alignement. Les alignements de séquences mettent en évidence les résidus conservés ou homologues entre différentes protéines. Ils permettent ainsi l'identification de sites fonctionnels possibles (sites actifs par exemple) pour des protéines dont la fonction est inconnue, ou encore de les apparenter à une famille de protéine définie.

1. Comparaison de séquences avec Clustal Omega

Clustal est un outil d'alignement multiple de séquence, la version *Omega*[59] étant sa version la plus récente et la plus performante. *Clustal Omega* remplace chaque séquence de l'alignement par un vecteur, chaque élément du vecteur représentant sa distance par rapport à une autre séquence de l'alignement. Ces vecteurs sont ensuite triés et regroupés par des méthodes classiques telles que la méthode statistique des K-means. Les alignements sont finalement effectués par comparaison de deux profils de modèles de Markov cachés. *Clustal Omega* permet également d'utiliser des modèles de Markov cachés issus d'alignements multiples déjà existants (profils d'alignements externes). Ainsi, dans le cas d'un alignement de séquences pour une famille de protéines donnée, si un alignement de séquences a déjà été effectué sur d'autres protéines de la même famille, *Clustal Omega* peut en extraire le modèle de Markov caché et le réinjecter dans le nouvel alignement afin d'améliorer sa précision. Il est également possible de procéder par itération : pour un alignement de séquences donné, le modèle de Markov caché obtenu est utilisé pour réaligner ces mêmes séquences et ainsi augmenter le niveau de précision de l'alignement.

2. Analyse de séquences avec BLAST

Une des méthodes les plus utilisées pour rechercher des régions similaires entre séquences nucléotidiques ou peptidiques est *Basic Local Alignment Search Tool (BLAST)*[60]. Il compare la séquence à analyser (séquence "cible") à toutes les séquences disponibles (séquences "modèles") dans différentes bases de données.

BLAST permet ainsi d'établir des liens évolutifs et fonctionnels entre séquences et d'identifier la famille de gènes ou de protéines à laquelle est potentiellement apparentée la séquence "cible" (Figure 2.2).

BLAST se décline en plusieurs versions selon la nature nucléotidique ou peptidique des séquences étudiées. *PSI-BLAST* (pour *Position-Specific Iterated BLAST*) est une extension consistant à exécuter *BLAST* par itérations, chacune déterminant une séquence consensus utilisée comme séquence "cible" pour l'itération suivante.

séquence requête (SR) : REEFANDSWELL

1- découpage de la SR en k-uplets (classiquement **quadruplets**) chevauchants :

REEFANDSWELL → REEF, EEFA, EFAN, FAND, ANDS, NDSW, DSWE, SWEL, WELL

2- construction du dictionnaire des **quadruplets homologues (QH)** :

- 20 acides aminés, donc $20^4 = 160\ 000$ quadruplets évalués (AAAA, AAAC, AAAD...)
- application d'une matrice de similitudes pour calculer les scores d'alignements

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	1	0	-2	-2	0	6													
D	-3	0	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	1	4						
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	2	2	4					
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	2	1	3	1	4			
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	3	7		
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

Matrice de similitudes BLOSUM62
(la plus couramment utilisée)

- les quadruplets présentant un score d'homologie supérieur au seuil défini par l'utilisateur sont conservés dans le dictionnaire final (REEF : KEEF, QEEF, RDEF, RQEF, RKEF...)

3- identification des quadruplet des **séquences modèles (SM)** présents dans le dictionnaire :

SM- RYTHIRDEF S N D T F E L Q G H

QH- RDEF

SR- REEF

4- extension de l'homologie en amont et en aval des quadruplets identifiés :

- le score d'alignement augmente d'autant plus que la séquence homologue est longue
- l'algorithme **stoppe** lorsque le score d'alignement cesse d'augmenter

SM- RYTHIRDEF S N D T F E L Q G H

RYTHI / RDEF S N D T F E L / Q G H

SR- ← REEF →

/ REEFANDSWEL / L

- les séquences modèles présentant un score d'alignement supérieur au seuil choisi sont conservées

5- calcul de l'espérance mathématique de chaque alignement :

- pour chaque alignement SM/SR, la valeur de l'espérance mathématique (E-value) est calculée
- plus l'E-value est faible ($< 10^{-10}$), moins l'alignement d'être dû au hasard

Figure 2.2. Algorithme BLAST pour la recherche de régions similaires entre protéines.

III. Prédiction de structures secondaires et topologie membranaire

1. Prédiction de structures secondaires

L'organisation d'une protéine en structures secondaires peut être caractérisée en considérant sa séquence peptidique seule, en s'appuyant sur la relation séquence-structure (voir **INTRODUCTION**). Différentes méthodes basées sur divers algorithmes ont ainsi été développés afin de prédire les structures secondaires d'une protéine à partir de sa séquence peptidique.

PREDATOR[61] est basé sur la probabilité de formation de liaisons hydrogène entre atomes de la chaîne principale de la séquence "cible". Ces probabilités sont estimées à partir de données structurale (PDB) pour chaque acide aminé, puis l'influence des résidus voisins est prise en compte. Un consensus par combinaison de ces deux prédictions est finalement établi pour chaque acide aminé, leur contribution à la formation de structures régulières (hélice α ou feuillet β) étant évaluée. *GOR IV*[62] utilise la théorie de l'information et les statistiques Bayésiennes pour estimer la probabilité de chaque acide aminé de la séquence 'cible' à former des structure secondaire spécifiques (hélice α ou feuillet β) prenons en compte les résidus voisins. *PSIPRED* s'appuie sur des matrices de score "position-spécifiques" (PSSM) [63] combinées à l'application de réseaux de neurones. La séquence "cible" est d'abord traitée par l'algorithme PSI-BLAST, trois itérations sont effectuées pour générer et affiner la *PSSM*. Cette dernière est ensuite découpée en sous-blocs de 15 résidus, chacun étant utilisé en entrée pour un premier réseau de neurones dont les résultats successifs sont filtrés par un second réseau. 10% des données obtenues par ce double traitement sont conservées comme jeu d'évaluation, les 90% restant servant à optimiser le réseau de neurones (pondération des données). L'apprentissage stoppe lorsque la précision du réseau de neurones appliqué au jeu d'évaluation commence à diminuer. *PROF*[64] utilise également les réseaux de neurones, mais l'algorithme démarre par l'application de l'outil GOR auquel des données évolutionnaires sont ajoutées. Les résultats prédits par GOR, combinés à l'utilisation de réseaux de neurones, sont utilisés pour l'apprentissage (supervisé et non supervisé), générant ainsi deux classificateurs. Les résultats sont ensuite analysés par discrimination linéaire ou réseaux de neurones, produisant deux nouveaux classificateurs employés là encore pour l'apprentissage.

Les résultats de ces deux derniers classificateurs sont finalement moyennés pour obtenir une prédiction unique pour chaque acide aminé de la séquence requête.

JPRED[65] s'appuie des données évolutives couplé à 6 algorithmes pour établir une prédiction consensus : *NNSSP* (prédiction des plus proches voisins), *PHD* (réseaux de neurones), *DSC* (discrimination linéaire), *MULPRED* (combinaison de méthodes consensus sur séquence seule), *ZPRED* (prédiction pondérée basée sur la conservation des résidus) et *PREDATOR* (propension à former des liaisons hydrogène).

2. Prédiction de topologie membranaire

La relation séquence-structure précédemment utilisée pour déterminer l'organisation en structures secondaires des protéines présente également un intérêt pour les protéines membranaires, notamment pour définir leurs domaines transmembranaires. Une série de méthodes bioinformatiques ont ainsi été développées pour définir les régions d'une protéine immergée dans la membrane, à partir de sa séquence peptidique uniquement.

L'algorithme de *HMMTOP*[66] est basé sur le concept suivant : la topologie des protéines transmembranaires peut être déterminée par une divergence maximale de composition en acides aminés de chaque segment de la séquence peptidique. Ainsi, des régions de topologies différentes (transmembranaires, extracellulaires ou intracellulaires) présentent de grandes différences dans leurs compositions respectives en acides aminés. Cette méthode identifie la localisation de certains segments utilisés, e.g., éléments structuraux (intracellulaire, extracellulaire, extrémité intracellulaire d'une hélice, extrémité extracellulaire d'une hélice, hélice transmembranaire). La précision de la méthode est augmentée par l'application d'un modèle de Markov caché, contrôlant la longueur des différents segments considérés. *OCTOPUS* [67] utilise aussi des modèles de Markov cachés couplés à des réseaux de neurones artificiels. Une recherche de séquences homologues à la séquence "cible" est tout d'abord effectuée via *BLAST*. Une séquence type est ainsi créée et utilisée comme entrée pour les réseaux de neurones, qui prédiront ensuite la localisation préférentielle de chaque résidu (transmembranaire, à l'interface, globulaire, dans une boucle, intra- ou extracellulaire). Ces prédictions servent alors d'entrées pour un modèle de Markov caché à double entrée, qui calcule la topologie la plus probable de chaque résidu. Les résultats de ces deux ensembles sont finalement combinés pour définir la localisation membranaire ou extra-membranaire de chaque acide aminé.

SPOCTOPUS[68] est basé sur le même algorithme qu'*OCTOPUS*, mais la méthode est renforcée par la prédiction du peptide signal de la séquence requête.

La séquence de ce dernier est déterminée par un modèle de Markov caché, en amont de l'application d'*OCTOPUS*. Cette étape améliore la précision de la prédiction de topologie membranaire. *PHOBIUS* et *POLYPHOBIUS*[69] s'appuient également sur un modèle de Markov caché, mais incluent des informations concernant des protéines homologues à la séquence "cible" (identifiées via *BLAST*). Les probabilités de chaque résidu de se trouver dans une hélice transmembranaire, en milieu extracellulaire ou intracellulaire, sont ensuite calculées pour la séquence "cible" ainsi que ses homologues. Une matrice moyenne des probabilités est finalement construite (par combinaison des séquences traitées) puis utilisée pour optimiser la précision de l'algorithme. *PHILIUS*[70] est similaire à *PHOBIUS* et *POLYPHOBIUS*, mais implique l'utilisation de réseaux Bayésiens dynamiques pour augmenter la précision des prédictions. *SCAMPI*[71] calcule la contribution de chaque acide aminé dans l'énergie libre d'insertion de la protéine dans la membrane. Sa performance est équivalente à celle des meilleures prédictions de topologie actuelles basées sur des statistiques.

IV. Modélisation par homologie

Les relations entre les différents niveaux d'organisation des protéines, de la structure primaire (séquence peptidique, 1D) à tertiaire (structure 3D), voire quaternaire (architecture 3D) le cas échéant (voir **INTRODUCTION**) permettent de combler partiellement le manque de données structurales expérimentales, notamment l'absence de structure tridimensionnelle pour une protéine "cible". En effet, deux protéines appartenant à la même famille possèdent des fonctions similaires, leur repliement tridimensionnel est donc sensiblement identique. Leurs séquences primaires sont également susceptibles de présenter une homologie élevée. En outre, les structures tridimensionnelles sont plus conservées entre protéines distinctes que leurs séquences peptidiques. Cette observation est d'autant plus vraie pour les protéines membranaires, chez lesquelles le repliement 3D est hautement conservé par rapport aux protéines cytoplasmiques, y compris pour des protéines dont l'identité de séquence est relativement basse (entre 20 et 35%)[72].

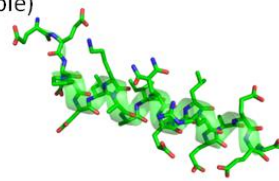
La **modélisation par homologie** se base sur cette relation afin d'estimer le repliement tridimensionnel d'une protéine de structure tertiaire inconnue.

Si la structure 3D d'une protéine A est résolue expérimentalement, et que sa séquence peptidique est proche de celle d'une protéine B dont on cherche à définir la structure tertiaire, il est possible d'estimer la structure 3D de la protéine B à partir de celle de la protéine A. La modélisation de la structure 3D d'une protéine par homologie est subdivisée en trois étapes principales (**Figure 2.3**).

Séquence peptidique de la **protéine cible (PC)** :
EDDPEL**K**GAL**Q**ETL**RDA**EG****

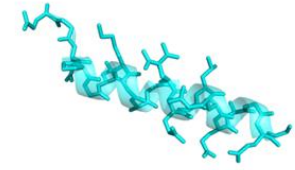
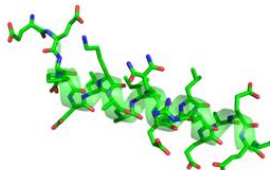
1- Identification d'une **protéine homologue (PH)** dont la structure 3D a été résolue expérimentalement (via BLAST par exemple)

EEDPDLKAAIQES**LRDA**EEA****



2- Alignement de séquences, construction du squelette carboné de la protéine cible et des chaînes latérales des **acides aminés identiques** entre les deux protéines

PC- EDDPEL**K**GAL**Q**ETL**RE**AD**EG**
PH- EEDPDL**KAAIQ**ES**LR**EA**EEA**



3- Construction des chaînes latérales des **acides aminés différents** entre la protéine cible et sa protéine homologue (via des bibliothèques recensant les conformères les plus probables)

PC- EDDPEL**K**GAL**Q**ETL**RE**AD**EG**
PH- EEDPDL**KAAIQ**ES**LR**EA**EEA**

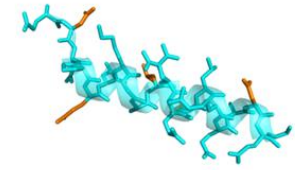
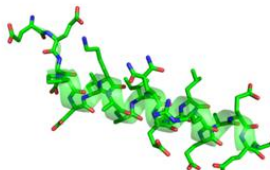


Figure 2.3. Protocole de modélisation par homologie.

Certaines protéines ou fragments protéiques ne présentent pas d'homologie avec les structures 3D caractérisées expérimentalement. De nouvelles méthodes *ab initio* (ou *de novo*) ont donc été développées pour pallier ce manque de données et générer leur structures. Elles reposent sur le concept que l'état natif d'une protéine correspond à son minimum global d'énergie libre. La modélisation *ab initio* consiste ainsi à rechercher ce minimum global en parcourant l'espace conformationnel de la protéine.

Chaque modèle 3D est obtenu par optimisation d'une fonction de densité de probabilité (probabilité d'observer chaque atome de la molécule à un positionnement donné dans l'espace cartésien). L'optimisation de cette fonction implique des méthodes de gradients conjugués et de dynamique moléculaire associée au recuit simulé afin que chaque atome de la molécule atteigne sa position la plus probable, donc la plus pertinente du point de vue structural. La connaissance d'une fonction d'énergie adéquate et une procédure de parcours de l'espace conformationnel efficace sont donc indispensables. Il est également nécessaire de limiter les temps de calculs, en réduisant par exemple le nombre d'atomes de la protéine considérés. Le compromis généralement adopté réside dans le choix adapté d'un sous-ensemble d'atomes du squelette du chaîne principale ou formé par les atomes C α et des chaînes latérale des acides aminés, combiné à une fonction d'énergie reflétant le plus fidèlement possible les effets du solvant et des atomes omis dans le sous-ensemble précédemment sélectionné.

La structure 3D de VKORC1 a été modélisée par homologie (le choix de la protéine homologue est détaillé dans la partie **RESULTATS**) via le logiciel *MODELLER* (version 9.2)[73]. *MODELLER* permet d'imposer un repliement tridimensionnel particulier à certains domaines de la protéine cible, par exemple des structures secondaires ou des ponts disulfures qui ont été mis en évidence expérimentalement en amont. Il est également possible de procéder à une modélisation itérative, en variant la protéine homologue utilisée au départ, ce qui améliore à chaque étape la précision du modèle 3D construit.

La modélisation par homologie d'une protéine cible suit deux étapes, chacune exécutée par un script codé en langage *Python* :

- alignement des séquences de la protéine cible et de la (ou des) protéines homologue(s) ;
- construction du (ou des) modèle(s) 3D de la protéine d'intérêt.

L'algorithme utilisé durant la première étape est cependant très simpliste, les alignements résultants ne sont donc pas toujours optimisés. Il est néanmoins possible de les rectifier manuellement, ou encore de fournir à *MODELLER* un alignement de séquences personnalisé, obtenu via *Clustal Omega* par exemple.

Construction des modèles pour les régions variables

Les régions variables sont des domaines (généralement des boucles) compris entre des régions conservées des protéines. La construction de telles boucles nécessite de trouver un fragment compatible avec le reste du modèle et dont la séquence est la plus proche possible de celle de la cible. D'une façon générale, le premier critère de sélection est la taille de la boucle. Toute protéine homologue possédant une boucle de taille semblable à celle de la cible fournit un fragment intéressant, dans la mesure où ce fragment est soumis à des contraintes d'environnement similaires dans l'homologue et dans le modèle. Si aucun fragment de même longueur n'est disponible mais que la boucle cible est supposée appartenir à une classe de boucles définie, la sélection d'un fragment appartenant cette classe est requise. Enfin, si la classe à laquelle appartient la boucle cible est inconnue, la recherche de fragments dont la géométrie est compatible avec celle des régions conservées qui bordent la boucle cible est nécessaire. La géométrie et la séquence des fragments candidats sont les seuls guides pour faire un choix cohérent. Cette méthode, utilisant des fragments de boucles connues pour construire les boucles d'un modèle, a été décrite par Jones et Thirup [74].

Lors du parcours d'une base de données pour la recherche de fragments de boucles, *MODELLER*[73] utilise comme critère de recherche les coordonnées des Ca des résidus du modèle bordant la boucle. Ces résidus constituent la boucle luminale chez VKORC1, composée de 46 résidus. Les coordonnées des résidus d'ancrage utilisés pour l'insertion des fragments de boucle dans le modèle proviennent du modèle ou de la base de données de fragments. Ces sources n'étant pas identiques, deux méthodes d'ajustement des coordonnées de ces zones de chevauchement ont été mises en place. La première consiste à calculer les coordonnées des résidus d'ancrage par une moyenne pondérée des coordonnées du fragment et du modèle. Le poids des coordonnées du fragment augmente lorsqu'on se rapproche de la boucle. La seconde, sans doute préférable, consiste à faire des ajustements des angles de torsion des résidus du fragment de la base pour superposer les résidus d'ancrage.

V. Protocole de simulation de DM appliqué à VKORC1

Les simulations de DM visent à simuler l'évolution d'un système de particules (une protéine) au cours du temps dans son microenvironnement (voir **INTRODUCTION**). Il s'agit de reproduire *in silico* une petite fraction du sous-espace biologique contenant la biomolécule étudiée. Les étapes de construction d'un système à simuler en DM sont sensiblement les mêmes, indépendamment du logiciel utilisé.

1. Gromacs et ces fonctions

Le logiciel libre *Gromacs* (version 4.6.5) [75] a été utilisé pour réaliser toutes les simulations de dynamique moléculaire présentées dans la partie **RESULTATS**. *Gromacs* dispose de nombreux avantages :

- il est composé de plusieurs sous-programmes exécutables en ligne de commande ; ainsi, chaque étape de la DM sera effectuée via quelques lignes de commande simples, ce qui permet de vérifier l'intégrité du système à simuler étape après étape ;

- parmi ces sous-programmes, de nombreuses fonctions d'analyse sont disponibles, permettant de traiter directement les simulations de DM réalisées sans problème de compatibilité de formats ;

- une forte communauté s'est développée autour de *Gromacs* et des fonctions additionnelles sont régulièrement mises à disposition par des équipes de modélisation ;

- *Gromacs* inclut plusieurs champs de force (voir la description dans la partie **V.2**) pour les simulations de DM (*AMBER*, *CHARMM*, *GROMOS*, *OPLS*) et ceux-ci sont directement modifiables par l'utilisateur manuellement ; les paramètres à appliquer aux lipides (dans le cadre de simulations de DM impliquant une protéine membranaire) ou à un ligand (pour les simulations de complexes protéine-ligand) peuvent par exemple être ajoutés ;

- *Gromacs* subdivise le système simulé en groupes d'atomes correspondant aux différentes entités le composant (protéine, carbones α , squelette carboné, eau, ions, lipides...), tous consignés dans un index ; il est possible de combiner ces différents groupes pour créer des indexes personnalisés, particulièrement utiles pour analyser des entités spécifiques ;

- l'exécution du code de *Gromacs* est parallélisable et une version pour réaliser les calculs sur GPU est également développée depuis *Gromacs* 4.6.

Il est donc aisément utilisable sur tout type de plateforme (stations de travail, clusters de calculs locaux, mésocentres, supercalculateurs), ce qui permet un gain de performance considérable.

Chaque simulation de DM effectuée par *Gromacs* génère un ensemble de fichiers standards (Tableau 2.1).

Extension	Type	Contenu
.gro	texte	coordonnées atomiques du système
.top	texte	paramètres topologiques des atomes du système (nature, charge, masse, liaisons, angles)
.mdp	texte	paramètres de simulation de DM à appliquer
.tpr	binaire	toutes les informations nécessaires à la simulation (combinaison des fichiers .gro, .top et .mdp)
.trr	binaire	coordonnées atomiques du système tout au long de la simulation (trajectoire)
.xtc	binaire	trajectoire compressée
.ndx	texte	groupes d'atomes principaux présents dans le système (protéine, molécules d'eau, squelette carboné, atomes de carbones α , ...)
.log	texte	différentes informations de la simulation de DM à chaque itération

Tableau 2.1. Fichiers générés par Gromacs au cours des simulations de DM.

2. CHARMM36 –Champ de force utilisé pour les simulations de DM

Le champ de force CHARMM36 [76] a été utilisé pour simuler les forces appliquées à chaque atome au cours des simulations de DM. CHARMM36 est un champ de force tout atome adapté aux simulations de DM de protéines, pour lequel nous disposons en outre des paramètres à appliquer aux lipides. L'équation regroupant les termes composant ce champ de force est la suivante :

$$E_{\text{totale}} = E_{\text{élongation}} + E_{\text{Urey Bradley}} + E_{\text{angle}} + E_{\text{dièdre}} + E_{\text{dièdre impropre}} + E_{\text{non liées}}$$

avec

$$E \text{ élongation} = \sum_{\text{élongation}} Kd(d - d_0)^2$$

où Kd est la constante de force de longueur de liaison, d la longueur de la liaison à un instant t et d_0 la longueur de la liaison de référence ;

$$E \text{ Urey Bradley} = \sum_{\text{Urey Bradley}} Kub(S - S_0)^2$$

où Kub est la constante de force d'Urey-Bradley, S la distance séparant deux atomes non liés mais séparés par deux liaisons covalentes à un instant t et S_0 la distance de référence entre ces deux atomes. Ce terme permet l'ajout d'une contrainte de distance entre deux atomes non liés;

$$E \text{ angle} = \sum_{\text{angle}} K\theta(\theta - \theta_0)^2$$

où $K\theta$ est la constante de force d'angle de liaison, θ l'angle de liaison entre trois atomes donnés à un temps t et θ_0 l'angle de liaison de référence entre ces trois atomes ;

$$E \text{ dièdre} = \sum_{\text{dièdre}} K\tau(1 + \cos(n\tau - \delta))$$

où $K\tau$ est la constante de force d'angle dièdre, n la périodicité de la rotation, τ l'angle dièdre et δ l'angle de phase ;

$$E \text{ dièdre impropre} = \sum_{\text{dièdre impropre}} K\phi(\phi - \phi_0)^2$$

où $K\phi$ est la constante de force d'angle dièdre impropre, ϕ l'angle dièdre impropre à un temps t et ϕ_0 l'angle dièdre impropre de référence. Contrairement aux angles dièdres "classiques", les liaisons entre quatre atomes formant un angle dièdre impropre ne sont pas linéaires, mais un atome est situé en position centrale et présente une liaison avec chacun des trois autres atomes;

$$E \text{ non liées} = \sum_{\text{non liées}} \left\{ \epsilon_{ij} \left[\left(\frac{R_{ij\text{min}}}{R_{ij}} \right)^{12} - \left(\frac{R_{ij\text{min}}}{R_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon l r_{ij}} \right\}$$

où ϵ_{ij} est la profondeur du puits de Lennard-Jones, $R_{ij\text{min}}$ la distance à laquelle le potentiel de Lennard-Jones est minimal, q_i et q_j les charges atomiques partielles respectives des atomes i et j , ϵl la constante diélectrique et R_{ij} la distance la distance entre les atomes i et j .

3. Construction d'un modèle de membrane biologique

L'étude des protéines membranaires (telle que VKORC1) implique la reconstitution d'une bicouche lipidique durant la préparation de simulations de DM. Il existe deux façons de procéder à cette étape :

- des patches membranaires de compositions lipidiques variées sont mis à disposition dans des banques de données ou sur les sites de différentes équipes de modélisation moléculaire ; les dimensions et la composition des patches sont donc limitées, mais ils présentent l'avantage d'être pré-équilibrées et sont ainsi prêts à l'emploi en DM ;

- des serveurs en ligne permettent de construire des patches lipidiques personnalisés (nombre et type de lipides à incorporer dans la fraction membranaire simulée) ; ces patches sur mesure nécessitent cependant une pré-équilibration complète avant l'application à la DM.

La membrane lipidique utilisée pour les simulations de DM de VKORC1 a été construite selon cette seconde méthode, grâce au serveur *Membrane Builder* (*MemBuilder*)[77]. Il propose de choisir :

- 1- le type de membrane à construire (bicouche lipidique, liposome ou micelle) ;
- 2- les dimensions du patch membranaire, de 64 à 400 lipides par monocouche ;
- 3- la composition lipidique précise de la membrane (proportion d'acide phosphatidique, phosphatidyl-choline, -éthanolamine, -glycérol, -sérine, cholestérol, sphingomyéline) ;
- 4- le champ de force utilisé (*AMBER*, *CHARMM*, *GROMOS*).

MemBuilder permet également de construire, dans le cas des bicouches et liposomes, des membranes asymétriques. Chaque monocouche présente alors une composition en lipides spécifique. Les membranes sont générées en formats .pdb et .gro (spécifique à *Gromacs* et donc bien adapté pour les simulations de DM réalisées dans ces travaux).

Une membrane composée de 200 molécules de DLPC (1,2-dilauroyl-sn-glycero-3-phosphocholine) a été construite en utilisant *MemBuilder*. Le choix du type de lipide est expliqué dans la partie **RESULTATS**. L'utilisation de molécules de DLPC présente trois avantages majeurs :

- 1- les paramètres à appliquer pour les simulations de DM ont été réajustés par rapport aux données expérimentales en 2012, puis implémentés dans le champ de force *CHARMM36* applicable via *Gromacs*[78] ; l'intégralité des simulations de DM présentées dans ce manuscrit ont été effectuées en utilisant ce champ de force personnalisé ;

2- ces nouveaux paramètres dispensent l'utilisateur de l'application d'une tension superficielle durant les simulations de DM ; celle-ci vise à empêcher les molécules d'eau de pénétrer dans la membrane durant les simulations (l'eau et les lipides étant non miscibles) et requiert habituellement l'ajout de contraintes durant chaque étape de la simulation ; or, les paramètres implémentés dans le champ de force *CHARMM36* contiennent un terme mimant la tension superficielle ;

3- la température de transition de phase du DLPC (271 K, mesurée expérimentalement) permet d'effectuer des simulations de DM à température physiologique (310 K) tout en conservant l'état liquide de la membrane ; d'autres types de lipides, comme le DPPC dont la température de fusion est de 314 K, imposent des températures de simulation élevées pour éviter le gel de la membrane et la perte de fluidité des lipides (suppression des processus comme la diffusion latérale).

L'agencement des lipides dans la bicouche modélisée par *MemBuilder* est très éloignée de leur l'organisation réelle dans les membranes biologiques (**Figure 2.4**). Outre l'étape de solvatisation du patch membranaire, le processus d'équilibration de la membrane modélisée sera primordial avant son utilisation pour les simulations de DM.

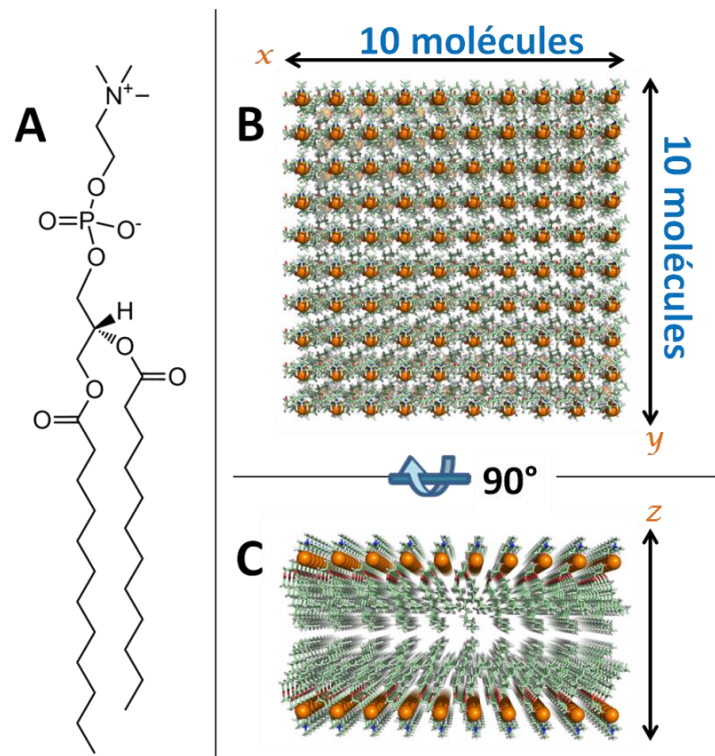


Figure 2.4. Modèle de membrane biologique construit par MemBuilder. (A) Structure chimique d'une molécule de DLPC. (B) Membrane lipidique modélisée par MemBuilder vue du dessus. (C) Membrane lipidique modélisée par MemBuilder vue de profil. Les lipides sont représentés en sticks, les atomes de phosphates en sphères. Les atomes de carbone, d'azote, d'oxygène, de phosphate et d'hydrogène sont colorés en vert, bleu, rouge, orange et blanc, respectivement.

4. Solvatation du modèle de membrane et définition de la boîte de simulation

L'étape de solvatation du patch modélisé a nécessité l'ajout 11 800 molécules d'eau de part et d'autre de la bicouche lipidique. Le modèle explicite TIP3P [79] a été utilisé pour représenter cet environnement aqueux. Ce modèle comprend trois sites d'interactions par molécule d'eau (correspondant chacun à un atome). Une charge ponctuelle est assignée à chaque atome et la géométrie de la molécule d'eau est représentée de façon rigide (longueurs et angle de liaisons fixes). Ce modèle simplifié permet de limiter les temps de calculs tout en restant efficace. Une simulation de DM impliquant une biomolécule et son environnement, la création d'une boîte est nécessaire pour contenir tous ces acteurs, ainsi que pour équilibrer convenablement le volume, la température et la pression du système. Cependant, les atomes du système sont susceptibles de dériver au cours de la simulation et risquent donc de "sortir" de cette boîte. Une boîte de simulation de DM doit donc répondre aux conditions périodiques aux limites : le système simulé est virtuellement répété à l'infini, de façon adjacente à chaque face de la boîte, à la manière d'un cristal infini parfait. Ainsi, lorsqu'un atome "sort" de la boîte de simulation par une face, un atome aux propriétés identiques est automatiquement recréé au niveau de la face opposée (**Figure 2.5**).

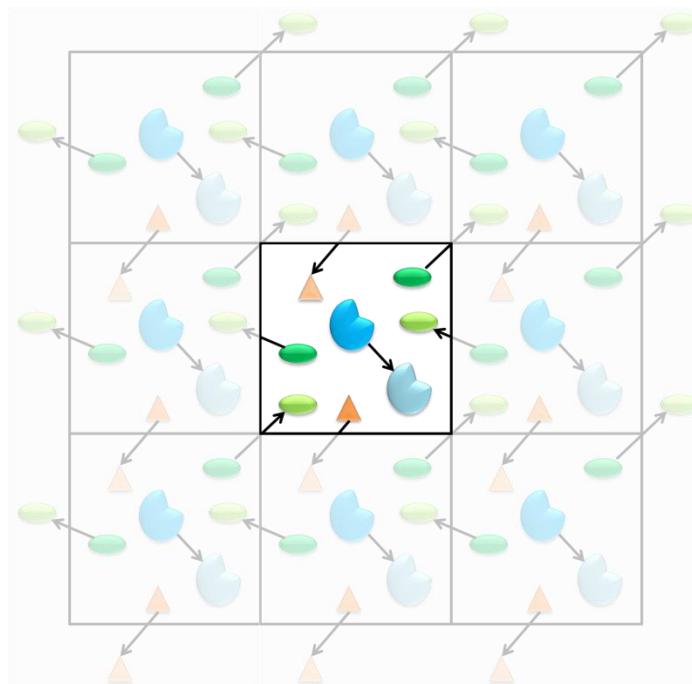


Figure 2.5. Conditions périodiques aux limites appliquées à la boîte de simulation de DM. La boîte est délimitée par un cadre noir. Les atomes sont représentés par des formes géométriques de couleurs différentes. Leur direction en cours de simulation est indiquée par une flèche noire. Les boîtes répétées à l'infini sur chaque face de la boîte de simulation sont représentées en transparence.

Deux contraintes s'imposent donc dans le choix de la boîte de simulation : seules certaines formes peuvent satisfaire aux conditions périodiques limites (cubique, triclinique, octaédrique, dodécaédrique), et la boîte doit être suffisamment volumineuse pour éviter que la biomolécule n'interagisse avec elle-même par périodicité. Les interactions à longue distance considérées en dynamique moléculaire s'étendent à 12 Å, une distance de 7 Å minimum doit donc séparer la surface de la biomolécule des limites de la boîte de simulation (ce qui évite toute interaction des atomes avec leur image périodique, alors à une distance de 14 Å). Le meilleur compromis réside en la définition d'une boîte de volume suffisamment important, combinée à la forme minimisant le nombre d'atomes de solvant nécessaires au remplissage de ladite boîte (influençant directement les temps de calcul de la simulation de DM). Dans notre cas, une boîte de simulation de forme cubique est la plus adéquate, considérant la forme du modèle de membrane généré.

5. Minimisation de l'énergie potentielle du patch membranaire

Les structure 3D des biomolécules, qu'elles proviennent de données expérimentales (résolution par cristallographie, RMN ou cryo-microscopie électronique) ou qu'elles soient modélisées (par homologie ou *de novo*), constituent toujours une représentation d'une densité (probabilité) de chaque atome à un positionnement donné (structure statique). Un nombre très limité des protéines caractérisés expérimentalement nécessite la construction de modèles théoriques.

Par conséquent, les structures 3D modélisées par homologie (ou *de novo*) présenteront les mêmes défauts que les structures expérimentales: longueurs et angles des liaisons covalentes parfois erronés, clashes stériques entre atomes, modifications structurales induites par d'éventuelles molécules limitant la flexibilité de la biomolécule. Une étape d'affinement, la minimisation d'énergie, s'avère donc nécessaire afin de rectifier toutes les imperfections structurales des molécules étudiées. Cette méthode utilise le gradient de surface d'énergie potentielle, directement lié aux forces de mécanique moléculaire, pour guider le système vers son minimum d'énergie le plus proche. Chaque atome est ainsi déplacé selon les forces appliquées par le reste du système, provoquant une réduction de l'énergie potentielle. La minimisation d'énergie est appliquée à tout le système (la protéine et les molécules de solvant étant ajoutées artificiellement) pour éviter les clashes stériques.

La position de chaque atome est finalement optimisée en fonction de l'énergie la plus favorable, rétablissant des distances interatomiques, longueurs et angles des liaisons covalentes physiquement adéquats.

La membrane modélisée a donc subi une minimisation d'énergie suivant l'algorithme du gradient (encore appelé algorithme de la plus forte pente, ou *steepest descent algorithm*, en anglais). Cette méthode applique à chaque itération k la relation suivante :

$$x(k) = x(k - 1) + \lambda(k)F(k)$$

où le vecteur x représente les $3N$ dimensions de configurations, $\lambda(k)$ le pas de chaque étape et $F(k)$ le vecteur de force, défini ainsi :

$$F(r) = -\frac{d}{dr}V(r) \text{ (} r \text{ étant le vecteur des coordonnées atomiques)}$$

Le pas de l'itération initiale est fixé à $0,1 \text{ \AA}$, puis réajusté après chaque itération selon l'évolution de l'énergie potentielle globale du système. Une augmentation de cette dernière signifie que le pas était trop élevé et que le minimum local le plus proche a été franchi ; le pas est donc réduit lors du passage à l'itération suivante. En revanche, si l'énergie potentielle globale diminue, le pas est accru durant la prochaine itération. Cet ajustement perpétuel du pas permet au système de s'orienter petit à petit vers le minimum local le plus proche. La minimisation est stoppée lorsque le gradient maximal devient inférieur à $1\ 000 \text{ kJ/mol/nm}$.

6. Equilibration de la température du patch membranaire à volume constant

La minimisation d'énergie effectuée selon l'approche décrite dans la partie **V.5** ne suffit pas seule à générer un système prêt à la simulation de DM. Il est également nécessaire de recréer les conditions de température les plus proches possibles du milieu biologique. Une étape de préparation du modèle effectuée via une courte simulation de DM permet d'établir une température typique du milieu biologique. La température de la membrane modélisée a donc été équilibrée à 310 K durant 1 ns , avec un pas d'intégration de 2 fs . Cette équilibration a été suivie par le contrôle de deux paramètres (l'aire par lipide et l'épaisseur de la bicouche lipidique, présentées dans la partie **V.8**) toutes les 100 ps .

L'algorithme de contrainte *LINCS* (*LI*Near *C*onstraint *S*olver) a été appliqué afin de fixer les longueurs et angles de liaison interatomiques résultant de la minimisation d'énergie potentielle globale. Le thermostat de couplage *vitesse-rescale* a été utilisé ; il s'agit d'une modification du thermostat de Berendsen [80] qui intègre un terme stochastique assurant la génération d'un ensemble canonique adéquat. Les molécules d'eau et de lipides sont considérées comme deux phases distinctes, avec une constante de temps de couplage égale à 100 fs pour chaque groupe de molécules. Les mouvements de translation du centre de masse de chacun de ces groupes de molécules (eau et lipides) sont supprimés afin d'empêcher un déplacement latéral de la bicouche lipidique par rapport aux molécules d'eau.

7. Equilibration de la pression du patch membranaire à température constante

A l'instar de la température, les conditions de pression pour la résolution expérimentale de structures 3D sont éloignées de celles propres aux milieux biologiques. Une dernière étape s'impose donc, similaire à l'équilibration de température mais concernant cette fois la pression du système. Une nouvelle simulation de DM est effectuée, conduisant le système vers une pression biologiquement pertinente. Les paramètres appliqués sont sensiblement identiques à ceux de l'équilibration de température (pas d'intégration de 2 fs, algorithme de contrainte *LINCS*, séparation des molécules d'eau et de lipide en deux phases avec suppression des translations des centres de masse). Cependant, l'ensemble de Nose-Hoover est cette fois utilisé comme thermostat, dans lequel la constante de temps (500 fs dans ce cas) ne concerne plus le couplage des molécules de chaque phase mais la période de fluctuation de la température (toujours 310 K), rendant ce thermostat plus précis. L'ensemble de Parrinello-Rahman [81] a été utilisé en association avec un couplage de pression semi-isotrope. Dans ces conditions, la pression est équilibrée dans le plan x - y d'une part et le long de l'axe z (séparant les deux phases aqueuse et lipidique) d'autre part, avec une période de fluctuation de pression égale à 5 ps. L'équilibration de pression a été prolongée sur une durée de 100 ns pour assurer une réorganisation correcte des lipides entre eux et vis-à-vis de la phase aqueuse, réorganisation contrôlée par la mesure de l'aire par lipide et de l'épaisseur de la bicouche lipidique toutes les 10 ns.

Les conditions de température et de pression simulées en DM correspondent à des ensembles, incluant tous les systèmes qui peuvent présenter des états microscopiques différents mais appartiennent à un unique état macroscopique ou thermodynamique. Les ensembles les plus couramment utilisés sont l'ensemble canonique NVT (collection de tous les systèmes dont l'état thermodynamique est caractérisé par un nombre d'atomes, un volume et une température fixes) et l'ensemble isobare-isotherme NPT (nombre d'atomes, température et pression fixes).

8. Contrôle de la qualité de la bicouche lipidique : *GridMat*

La vérification de la pertinence biologique du modèle membranaire est indispensable à chaque étape des simulations de DM. Les deux critères pouvant servir de points de contrôle sont l'aire par lipide et l'épaisseur de la bicouche lipidique (**Figure 2.6**).

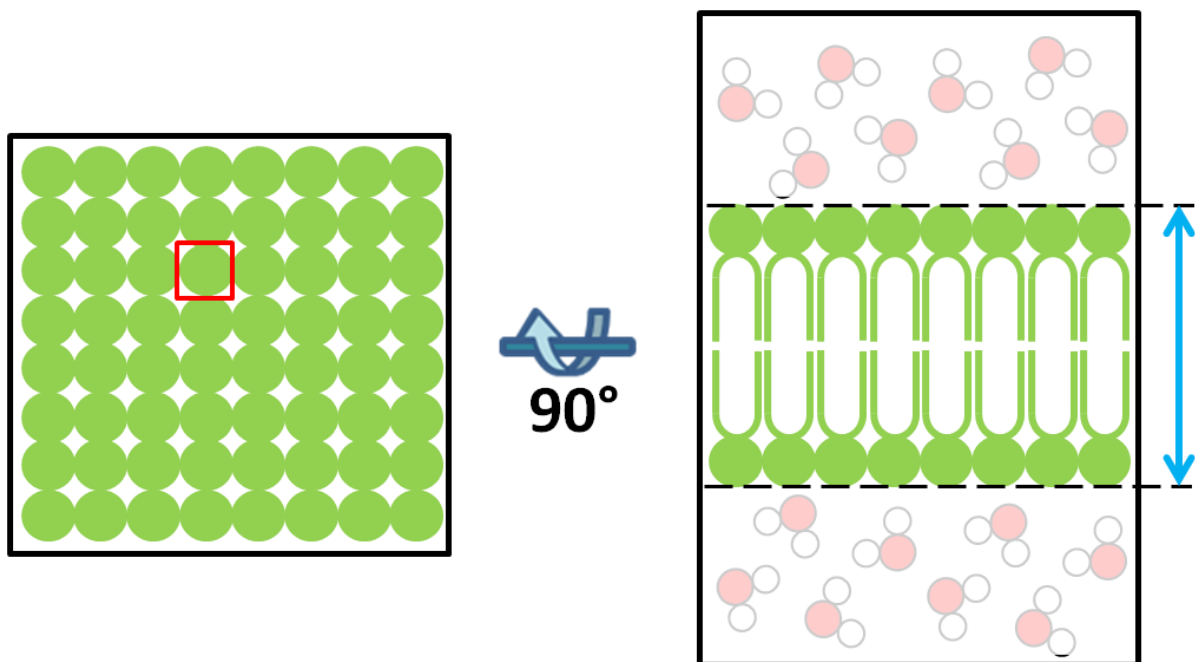


Figure 2.6. Aire par lipide et épaisseur de la bicouche lipidique d'une membrane biologique. L'aire par lipide correspond à la surface couverte par la tête phospholipidique dans le plan x-y, représentée par le cadre rouge sur le schéma de gauche. L'épaisseur de la bicouche est définie par la distance séparant les têtes phospholipidiques des lipides se faisant face dans le plan z ; elle est indiquée par la flèche bleue sur le schéma de droite.

Le logiciel *GridMAT-MD* (*Grid-Based Membrane Analysis Tool for Molecular Dynamics*) [82] calcule ces paramètres rapidement via l'exécution d'un script Perl (**Figure 2.7**). Une comparaison entre les données calculées par *GridMAT-MD* d'une part, et les épaisseurs de bicouches lipidiques et aires par lipides mesurées expérimentalement d'autre part, permet ensuite de vérifier l'adéquation du patch membranaire utilisé *in silico* avec le système biologique étudié.

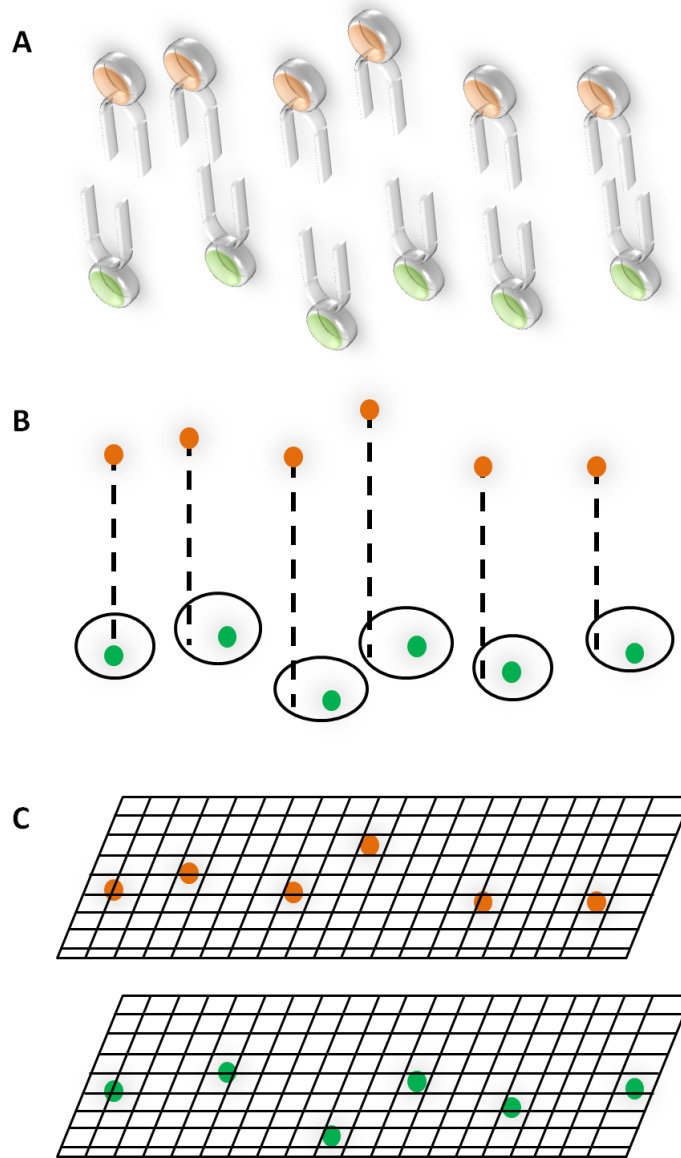


Figure 2.7. Estimation de l'aire par lipide et l'épaisseur d'une bicouche lipidique par GridMat. (A) Seules les têtes phospholipidiques des lipides membranaires sont considérées. (B) Pour chaque phospholipide d'un feuillet de la membrane, le phospholipide du feuillet opposé le plus proche du feuillet dans le plan x-y est recherché. Les distances entre chaque paire de phospholipides ainsi définies sont calculées puis moyennées pour obtenir l'épaisseur de la bicouche lipidique. (C) Une grille est définie dans le plan x-y pour chacun des deux feuillets. Chaque point de la grille est ensuite associé au phospholipide le plus proche dans le plan x-y. Le nombre de points associés à chaque phospholipide est ensuite divisé par le nombre de point total pour calculer l'aire par lipide de chaque phospholipide.

L'intégrité de la membrane modélisée a donc été vérifiée via *GridMAT-MD* après équilibration. L'aire par lipide calculée est de $62,81 \text{ \AA}^2$, celle mesurée expérimentalement est de $64,18 \text{ \AA}^2$ [78]. L'épaisseur de la bicouche calculée est de $30,7 \text{ \AA} \pm 2,7 \text{ \AA}$, celle mesurée expérimentalement est de $29,9 \text{ \AA} \pm 2,7 \text{ \AA}$. La membrane de DLPC modélisée et équilibrée est donc en bonne adéquation et prête à être utilisée pour la suite de l'étude.

9. Insertion du modèle 3D de VKORC1 dans la membrane biologique

Une fois la membrane lipidique correctement modélisée et équilibrée, l'insertion du modèle 3D de VKORC1 dans la membrane obtenue a été effectuée. Les étapes de génération du modèle structural de VKORC1 et le positionnement de ce modèle vis-à-vis de la membrane sont présentées dans la partie **RESULTATS**.

Lors de la préparation des simulations de DM du système VKORC1, le modèle 3D de l'enzyme a été intégré dans la bicouche lipidique via la fonction *g_membed* de *Gromacs*[83]. La protéine est d'abord positionnée manuellement dans la membrane, en utilisant des données expérimentales et/ou des outils bioinformatiques pour définir la région transmembranaire de la biomolécule. Les coordonnées de la protéine sont ensuite compressées dans le plan *x-y*. Le centre géométrique de la région transmembranaire de la protéine est utilisé comme point de référence. Les coordonnées atomiques *x* et *y* de la protéine sont alors réduites d'un facteur 100 par rapport à ce centre géométrique, générant une protéine filiforme le long de l'axe *z*. Toute molécule d'eau ou de lipide ayant au moins un atome dont les coordonnées recouvrent celles de la protéine (comprimée dans le plan *x-y*) est supprimée du système ; la fonction *g_membed* peut supprimer un nombre égal ou au contraire différent de lipides dans les deux monocouches de la membrane, afin d'adapter au mieux le trou formé à la forme réelle de la protéine étudiée. Une courte simulation de DM (1000 itérations) permet enfin la restauration des coordonnées initiales de la protéine. Chaque itération applique le principe classique de la DM, suivi d'une modification des coordonnées atomiques *x* et *y* de la protéine. Ces dernières sont multipliées chacune par un facteur 100, en prenant toujours le centre géométrique de la région transmembranaire comme point de référence. Toutes les interactions protéine-protéine (*y* compris covalentes) sont désactivées durant cette étape, afin de ne pas induire d'artefact lors du redimensionnement de la biomolécule (**Figure 2.8**).

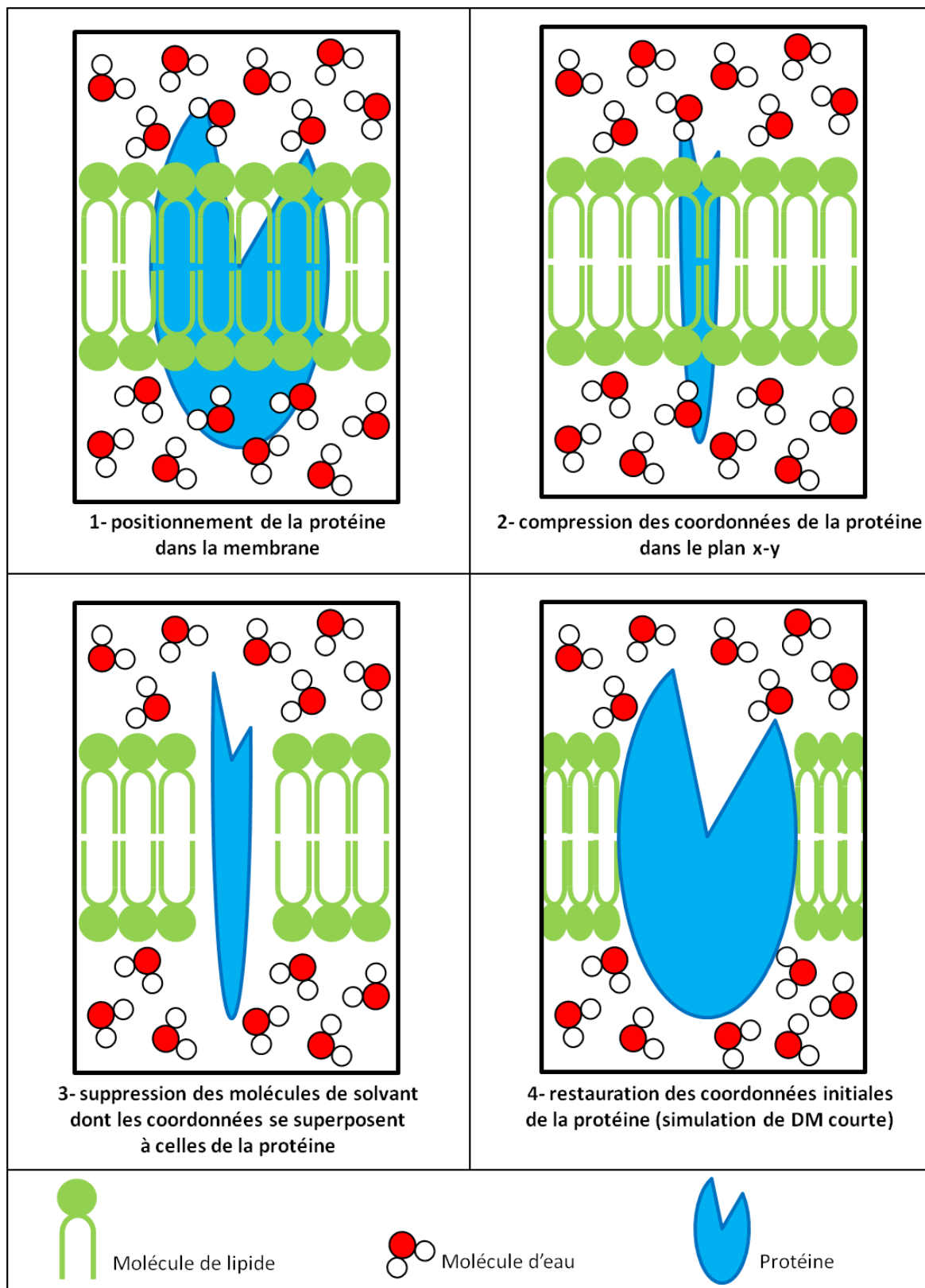


Figure 2.8. Algorithme d'insertion d'une protéine dans une membrane appliqué par la fonction *g_membed* de Gromacs.

Suite à l'application de cette méthode, la protéine retrouve sa conformation initiale tout en repoussant les lipides et molécules d'eau qui l'entourent. Cette spécificité de la méthode aboutit à un système proche de l'équilibre, puisque le réarrangement des lipides vis-à-vis de la protéine est optimal au cours de la dernière étape. L'étape d'équilibration de pression du système complet nécessitera néanmoins un temps de simulation compris entre 1 et 10 ns afin d'équilibrer la membrane dans son intégralité [83].

Les coordonnées atomiques du modèle VKORC1 ont donc d'abord été réduites d'un facteur 100 dans le plan x - y , puis 8 molécules de lipides et 52 molécules d'eau ont été supprimées du système (coordonnées atomiques superposées à celles de VKORC1). Ensuite, une simulation de DM de 2 ps (1 000 étapes, pas d'intégration de 2 fs) a été réalisée pour restaurer les coordonnées initiales du modèle VKORC1. Un seuil de 12 Å a été appliqué pour les interactions de van der Waals et de Coulomb, et l'algorithme *Particle Mesh Ewald* a été utilisé pour les interactions électrostatiques à longue distance; les ensembles de couplage *vélocité-rescale* et *Berendsen* (associé à un couplage de pression semi-isotrope) ont été utilisés pour la température et la pression, respectivement, en considérant la protéine et les lipides comme deux phases distinctes.

Deux autres protocoles existent pour l'intégration d'une protéine dans une membrane lipidique. Le premier implique l'utilisation de forces répulsives pour créer un trou dans la membrane, dans la zone d'insertion prévue de la protéine, avant l'introduction de cette dernière. Cependant, le réglage des paramètres nécessaires à la formation du trou le plus adapté à la forme de la protéine est très délicat. Le second est similaire à l'algorithme mis en œuvre par *g_membed*, mais les coordonnées atomiques de la protéine ne sont pas modifiées car toutes les molécules d'eau et de lipide dont les coordonnées se superposent à celles de la biomolécule sont directement supprimées. L'inconvénient majeur de cette méthode reste la durée conséquente de l'étape d'équilibration de pression (plusieurs dizaines, voire une centaine de nanosecondes), nécessaire à l'agglomération optimale des lipides restant autour de la biomolécule.

La fonction *g_membed* constitue donc la meilleure alternative à l'insertion du modèle 3D de VKORC1 dans la membrane biologique, en raison du temps d'équilibration de pression raisonnable (maximum 10 ns) et de sa simplicité d'utilisation (fonction implémentée dans Gromacs, donc utilisable en ligne de commande).

10. Equilibration du système VKORC1 + membrane

Une fois le modèle structural de VKORC1 intégré dans la membrane lipidique, les charges portées par la protéine (9 charges positives) ont été neutralisées via le remplacement de 9 molécules d'eau (choisies de façon aléatoire dans le système) par 9 ions Cl^- . L'énergie potentielle du système ainsi obtenu (modèle de VKORC1 intégré dans la membrane, entourée d'eau et de contre-ions) a ensuite été minimisée par l'algorithme du gradient, avec un pas d'itération initial fixé à 0,1 Å. La minimisation a été stoppée lorsque le gradient maximal est devenu inférieur à 1 000 kJ/mol/nm.

La température et la pression du système ont finalement été équilibrées en utilisant les mêmes paramètres que durant l'équilibration du patch membranaire seul, l'ensemble NVT *vélocité-rescale* durant l'équilibration de température, puis de Nose-Hoover pendant l'équilibration de la pression (en combinaison avec l'ensemble NPT de Parrinello-Rahman, lui-même associé à un couplage semi-isotrope). Le pas d'intégration était de 2 fs pour les deux étapes et l'algorithme de contrainte LINCS a été utilisé pour le maintien des longueurs et angles de liaison ; les coordonnées atomiques de la protéine ont encore été fixées par l'application de contraintes. Le système était subdivisé en trois phases pour les couplages de température et pression : le modèle de VKORC1, la membrane et les molécules d'eau groupées avec contre-ions. Les mouvements de translation des centres de masse des deux groupes *protéine+membrane* et *eau+ions* ont à nouveau été supprimés (pour éviter tout déplacement latéral d'une des phases vis-à-vis de l'autre). La température a été équilibrée à 310 K durant 100 ps, la pression à 1 bar durant 1 ns.

Le système généré à l'issue de ces étapes est solvato, minimisé et équilibré ; les contraintes appliquées aux coordonnées atomiques de la biomolécule sont alors levées et la simulation de DM dite "de production" peut être lancée. La préparation du système ne conduit donc à aucune exploration du paysage énergétique de la biomolécule concernée. Il s'agit uniquement de réarranger les atomes les uns par rapport aux autres et de recréer des conditions de simulations les plus fidèles possibles à la réalité biologique. Ces étapes mettent en évidence le double rôle de la dynamique moléculaire : affiner la structure 3D d'une biomolécule dans un premier temps (minimisation d'énergie, équilibrations de température et pression) puis explorer son espace conformationnel dans un second temps (dynamique de production).

11. Simulations de DM du système

Les paramètres attribués pour générer les données de production permettant l'exploration de l'espace conformationnel de la protéine sont les suivants : pas d'intégration de 2 fs, algorithme du saut de grenouille (*leap-frog algorithm* en anglais) [84] pour intégrer l'équation Newtonienne du mouvement, sauvegarde des coordonnées du système toutes les 5 ps, application des algorithmes *LINCS* (fixation des longueurs et angles des liaisons covalentes) et *Particle-Mesh Ewald (PME)*, interactions électrostatiques), fixation des seuils d'interactions électrostatiques et de van der Waals à 12 Å, subdivision du système en trois phases (protéine, membrane, eau+ions) et utilisation des ensembles de Nose-Hoover et Parrinello-Rahman pour maintenir une température et une pression (de type semi-isotrope) constantes (310 K et 1 bar, respectivement), suppression des mouvements de translation des groupes *protéine+membrane* et *eau+ions* pour éviter les déplacements latéraux de chaque phase vis-à-vis de l'autre. Les contraintes appliquées aux coordonnées atomiques de la protéine durant l'équilibration du système ont été levées afin de lui permettre une exploration de son espace conformationnel. Les paramètres appliqués aux simulations de DM considérant la présence d'un ligand (vitamine K ou AVK) sont identiques à ceux décrits précédemment. Les caractéristiques topologiques de chaque ligand (compatibles avec *Gromacs* et le champ de force *CHARMM36*) ont été obtenues via le serveur en ligne *SwissParam* développé par le *Swiss Institute of Bioinformatics* (SIB, <http://www.swissparam.ch/>). Le ligand a systématiquement été associé à la protéine durant chaque phase de préparation et de simulation : complexe "VKORC1•ligand" considéré comme une entité unique pour l'intégration dans la membrane, puis subdivision en trois groupes (protéine+ligand, membrane, eau+ions) pour les couplages de température et pression, et enfin regroupement *protéine+ligand+membrane* d'une part et *eau+ions* d'autre part pour la suppression des mouvements latéraux relatifs des phases. Les simulations de DM (données de production) ont été réalisées sur le supercalculateur Turing de l'IDRIS (mobilisant 256 CPUs pour une vitesse de simulation de 38 ns/j) (allocations 2014, 2015 et 2016-077291), ou sur le cluster hybride local du CMLA, TopDyn (CPU/GPU), en utilisant une combinaison hybride de 16 CPUs + 1 GPU (34 ns/j). En conséquence, l'algorithme de *Verlet*[85] a été utilisé pour la recherche des atomes voisins lors des simulations : seul algorithme proposé par *Gromacs* compatible avec l'accélération GPU, il a permis la conservation de paramètres de simulation identiques quel que soit le calculateur utilisé (Turing ou TopDyn), indispensable à la comparaison des trajectoires entre elles.

VI. Analyse des données de simulations de DM par méthodes classiques

1. Inspection visuelle des données de simulations de DM (VMD/PyMOL)

Les trajectoires générées par simulations de DM ont toutes été contrôlées visuellement en utilisant VMD (*Visual Molecular Dynamics*) [51], plus adapté que PyMOL [86] à la lecture des trajectoires de DM. Cette analyse visuelle des trajectoires a permis de contrôler l'intégrité du modèle de VKORC1 au cours des simulations de DM. En revanche, toutes les représentations de VKORC1 utilisées dans la partie **RESULTATS** pour analyser la structure de VKORC1 ont été réalisées avec PyMOL.

2. Root Mean Square Deviations (RMSDs)

La détection des déplacements des conformations décrits par une protéine constitue une première étape de validation de la stabilité de la biomolécule au cours d'une simulation de DM. Ainsi, une dérive significative de la protéine sera associée à l'observation d'un mouvement de grande amplitude lors de la visualisation de la trajectoire. La méthode standard d'analyse de ces déviations est le calcul des *Root Mean Square Deviations (RMSDs)*. Les valeurs de *RMSDs* de VKORC1 ont été calculées pour chaque conformation des trajectoires de DM, par rapport à la conformation initiale ($t = 0$ ns) de chaque trajectoire, selon la formule suivante :

$$RMSD (Tt, T0) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i |r_i(Tt) - r_i(T0)|^2}$$

Chaque conformation $r_i(Tt)$ de VKORC1 a été préalablement superposée à la conformation initiale $r_i(T0)$, en considérant les atomes de carbone α du domaine transmembranaire comme référence (résidus 9-28, 77-93, 102-123 et 132-151). Cette superposition évite la prise en compte des mouvements de rotation et translation décrits par la protéine durant les simulations de DM.

Les valeurs de RMSDs ont ensuite été calculées par la fonction *g_rms* de *Gromacs* sur les atomes de carbone α des résidus 9 à 151, afin de limiter le bruit induit par les fluctuations des extrémités N- et C-terminales de VKORC1.

3. Root Mean Square Fluctuations (RMSFs)

Le calcul des valeurs de RMSFs sur une protéine permet de mettre en évidence les régions dont les fluctuations atomiques sont les plus importantes (ou les plus faibles) au cours des simulations de DM. Les RMSFs ont été calculées sur VKORC1 par la fonction *g_rmsf* de *Gromacs*, selon la formule suivante :

$$RMSF(i) = \sqrt{\frac{1}{N} \sum_{t=1}^T |r_i(t) - \langle r_i \rangle|^2}$$

Les RMSFs ont été calculées sur les 163 atomes de carbone α de VKORC1, afin de limiter le bruit induit par les chaînes latérales, très flexibles en raison de leurs nombreuses liaisons pivotantes.

4. Analyse en Composantes Principales (ACP)

L'ACP est une méthode d'analyse de données permettant d'identifier des facteurs prédominants influençant un paramètre défini. En DM, l'ACP révèle les vecteurs propres (*eigenvectors*, directions 3D) privilégiés suivies par les atomes d'une protéine le long des trajectoires générées, par analyse de la covariance des fluctuations des coordonnées atomiques. La position moyenne de chaque atome au cours de la trajectoire étudiée est d'abord déterminée, puis la matrice de covariance C_{ij} est établie à partir des mouvements des atomes autour de leur position moyenne, selon la formule :

$$C_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle$$

où C_{ij} est l'élément de la matrice correspondant aux atomes i et j , r_i et r_j étant les coordonnées cartésiennes des atomes i et j , respectivement.

Seuls les atomes de carbone α des 163 résidus de VKORC1 ont été considérés. Chaque conformation de VKORC1 a préalablement été superposée à la conformation moyenne, en considérant les atomes de carbone α du domaine transmembranaire comme référence.

La matrice de covariance C a ensuite été diagonalisée selon la formule suivante :

$$AtCA = \omega$$

où A représente les vecteurs propres de la matrice de covariance, ω_i étant la valeur propre associée au i ème vecteur propre de la matrice A ; At est la transposée de A . Les valeurs propres définissent la fluctuation globale du système le long du vecteur propre associé.

Les vecteurs propres sont classés par ordre de fréquence croissant (donc par amplitude de mouvement décroissante). L'ACP a été réalisé par les fonctions `g_covar` et `g_anaeig` de *Gromacs*.

5. Cross-correlations

Les corrélations croisées (*cross-correlations*) entre atomes d'une protéine mettent en évidence les atomes dont les déplacements au cours des simulations de DM sont liés. La matrice de corrélations croisées est établie entre chaque paire d'atomes selon la formule suivante :

$$CC_{ij}^{ACP} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\langle \Delta r_i^2 \rangle^{1/2} \langle \Delta r_j^2 \rangle^{1/2}}$$

où i et j sont deux atomes de carbone α , Δr_i et Δr_j les déplacements des atomes i et j par rapport à leur position moyenne. L'élément de la matrice CC_{ij}^{ACP} est proche de 1 lorsque les mouvements des atomes i et j sont corrélés, et proche de -1 lorsqu'ils sont anti-corrélés. Il prend une valeur nulle lorsque que les mouvements ne présentent aucune corrélation particulière. Les matrices de *cross-correlations* ont été calculées par le package Bio3D [87] disponibles pour R, sur les atomes de carbone α des 163 résidus de VKORC1.

6. Clustering des conformations

La méthode de *clustering* permet d'identifier les conformations d'une protéine structurellement proches au cours d'une simulation de DM. Les *clusterings* effectués sur les simulations de DM de VKORC1 ont été réalisés par la fonction *g_cluster* de *Gromacs*, en utilisant l'algorithme Gromos. Chaque conformation de VKORC1 générée au cours d'une simulation de DM est préalablement été superposée à la conformation initiale, en considérant les atomes de carbone α du domaine transmembranaire comme référence. Les valeurs de RMSDs de toutes les conformations prises deux à deux sont ensuite calculées sur les atomes de carbone α des résidus 9-151 (les extrémités N- et C-terminales, très mobiles, sont négligées). Toutes les paires de conformations dont les valeurs de RMSDs sont inférieures à un seuil de RMSDs de 2 Å sont considérées comme "voisines". La conformation présentant le plus grand nombre de "voisines" et toutes ces conformations "voisines" sont regroupées en un *cluster* et retirées du groupe de conformations initial. Le même procédé est appliqué aux conformations restantes, jusqu'à ce que toutes les conformations soient associées à un *cluster*. Chaque conformation n'appartient ainsi qu'à un seul et unique *cluster*.

VIII. Détection de cavités à la surface de VKORC1 par *fpocket*

La structure tertiaire d'une protéine est directement liée à sa fonction. En particulier, sa surface est définie par des reliefs et des **cavités** (ou **poches**) dans lesquels sont susceptibles de venir se nicher des petites molécules ou des protéines-partenaires. La forme et la nature physico-chimique des poches définiront une spécificité vis-à-vis des molécules capables de se fixer et d'interagir avec la protéine. Des outils bioinformatiques pour l'étude *in silico* des cavités présentes au niveau des structures 3D des protéines sont disponibles. Le logiciel libre *fpocket* [88](téléchargeable ou utilisable en serveur web, <http://fpocket.sourceforge.net/>) a été utilisé pour détecter et analyser ces cavités.

Fpocket repose sur le concept de la sphère α , soit une sphère définie par 4 atomes en contact avec sa surface mais dont le volume interne ne contient aucun atome. Le volume des sphères α sera donc variable selon leur localisation au sein d'une protéine, les sphères les plus volumineuses se trouvant en surface, tandis que celles situées au cœur de la protéine présenteront un volume nettement plus petit (la densité atomique y étant plus élevée). Les cavités contiendront donc des sphères α de rayon intermédiaire.

Fpocket se base sur ce principe en filtrant toutes les sphères α définies pour une protéine en fonction de leur rayon. Une fois les sphères α de volume médian identifiées, celles dont les volumes se recouvrent ou dont les surfaces sont adjacentes sont regroupées pour définir une cavité (**Figure 2.9**).

Une extension de *fpocket*, *mdpocket*, a par la suite été mise en place pour analyser l'évolution d'une cavité au cours d'une simulation de DM. Il s'agit en réalité d'appliquer *fpocket* par itérations, pour plusieurs conformations (200 au minimum, idéalement) de la protéine décrivant la trajectoire à analyser.

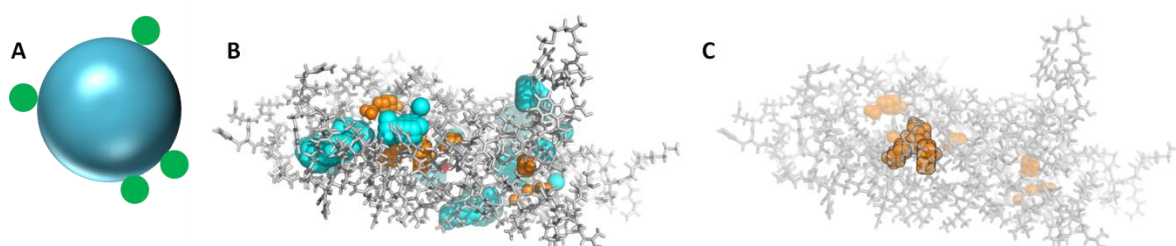


Figure 2.9. Algorithme de détection de poches à la surface de protéines par Fpocket. (A) Une sphère α (en bleu) est définie par un minimum de 4 atomes (en vert) en contact avec sa surface mais dont le volume est dépourvue d'atome. (B) Les sphères α situées à la surface de la protéine présentent un rayon important (sphères colorées en cyan), tandis que celles localisées au cœur de la protéine disposent d'un rayon faible (sphères colorées en orange). (C) Les sphères dont les volumes se recouvrent sont regroupées et forment ainsi des cavités (sphères opaques colorées en orange et entourées d'un trait noir). La protéine est représentée en *sticks* et colorée en gris.

Toutes ces conformations sont superposées les unes aux autres, puis *fpocket* détecte les cavités présentes pour chaque conformation. Les cavités (également superposées) sont ensuite extraites, sous forme de sphères α , puis une grille est construite sur l'ensemble des cavités. Enfin, la densité de sphères α est évaluée en chaque point de la grille, ce qui permet d'identifier les cavités persistantes (forte densité de sphères α) ou transitoires (faible densité de sphères α) au cours de la simulation de DM étudiée (**Figure 2.10**). Les cavités jugées intéressantes peuvent ensuite être analysées via *dpocket*, une autre extension de *fpocket* permettant d'obtenir un certain nombre de descripteurs de la poche étudiée – volume de la cavité, acides aminés composant la cavité, surfaces (totale, polaire et non-polaire) de la poche accessibles au solvant, charge globale portée par les acides aminés de la cavité.

Fpocket permet donc d'identifier la (ou les) cavité(s) d'une protéine à cibler comme site de liaison des ligands par l'amarrage moléculaire (présenté dans le paragraphe suivant) et fournit des caractéristiques physico-chimiques sur ces cavités potentielles, utilisables par la suite dans l'étude des interactions entre la protéine et divers ligands.

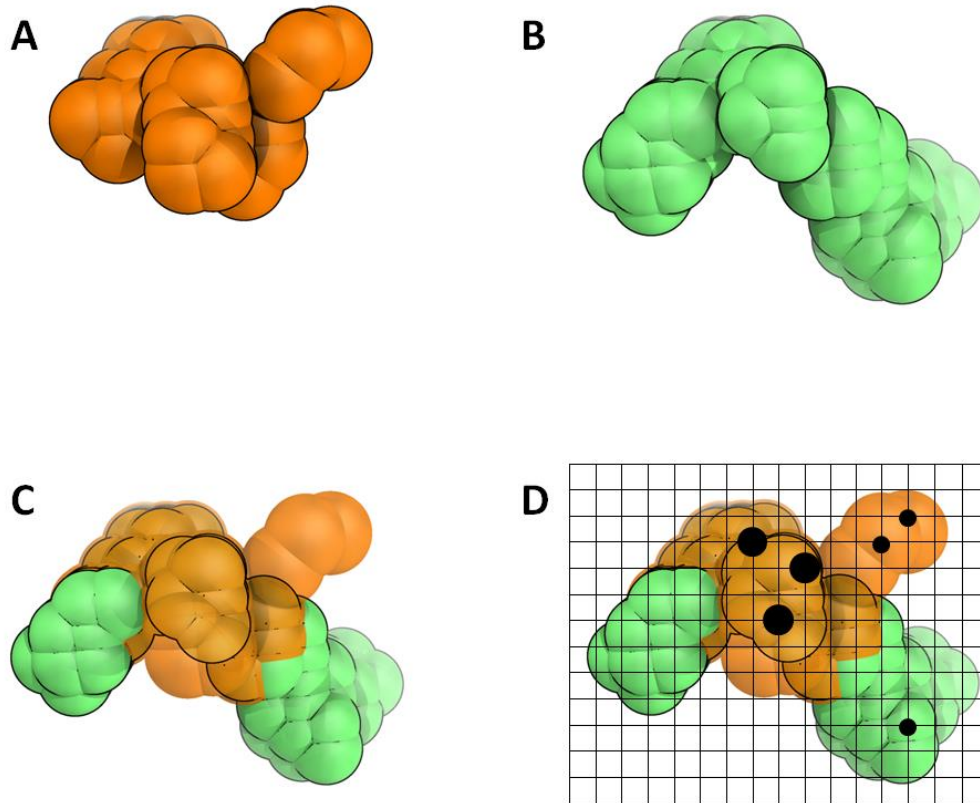


Figure 2.10. Algorithme de détection de poches à la surface de protéines au cours de simulations de DM par MDpocket. (A) et (B) Deux cavités identifiées sur deux conformations différentes d'une protéine générées au cours d'une simulation de DM. (C) Les cavités sont superposées. (D) Une grille est définie sur les cavités superposées. Pour chaque point de cette grille, la densité de sphères α est calculée, et représentée à titre d'exemple pour quelques points de la grille par un disque noir dont le diamètre est proportionnel à la densité de sphères α . Les cavités persistantes au cours d'une simulation de DM sont ainsi mises en évidence.

IX. Amarrage moléculaire

L'amarrage moléculaire (*molecular docking*) consiste à modéliser *in silico* les interactions entre un récepteur (une protéine) et un ligand (une petite molécule ou une autre protéine) à l'échelle atomique, visant la prédiction de l'orientation d'une molécule par rapport à une autre pour générer le complexe le plus stable possible. L'amarrage moléculaire peut s'appliquer à une protéine dont le site actif est inconnu (amarrage "aveugle"), expérimentalement caractérisé ou identifié *in silico* (via *fpocket* par exemple).

Cette méthode reposait à l'origine sur la théorie "clé-serrure" émise par Fischer en 1894 selon laquelle le ligand (clé) et le récepteur (serrure) présentent une complémentarité de forme conduisant à leur association.

Initialement, le ligand et le récepteur étaient tous les deux considérés comme des corps rigides. Koshland améliore ce postulat en 1963 avec la théorie de l'ajustement induit (*induced-fit*). Dans ce cas, le site actif de la protéine est sans cesse remodelé en fonction des interactions avec le ligand. Cette nouvelle perspective suggère la prise en compte de la flexibilité du récepteur et du ligand durant l'amarrage moléculaire, améliorant la précision des complexes récepteur-ligand prédits.

Indépendamment du niveau de considération de la flexibilité des partenaires, l'amarrage moléculaire se déroule toujours en deux étapes :

1- échantillonnage des conformations possibles du ligand dans le site actif de la protéine (position et orientation), chaque conformation prédite étant appelée *pose* ;

2- évaluation des affinités de liaison ligand-récepteur et classement des conformations prédites via une fonction de notation (*scoring*).

Plusieurs fonctions d'échantillonnage et de notation ont été mises en place et sont présentées dans les parties suivantes.

1. Algorithmes d'échantillonnage des conformations du ligand

La combinaison des six degrés de liberté de rotation-translation propres à une conformation donnée pour un ligand et de tous les degrés de liberté conformationnels du ligand et du récepteur génèrent un nombre très conséquent de modes de fixation entre ces deux molécules. Divers algorithmes ont donc été proposés pour échantillonner efficacement les conformations du ligand et ainsi limiter les temps de calcul.

Les algorithmes d'appariement sont basés sur une cartographie de la structure moléculaire du ligand et du site actif de la protéine. Ils impliquent des paramètres chimiques (tels que les atomes donneurs et accepteurs de liaisons hydrogène) et de complémentarité de forme entre le ligand et la poche du récepteur. Ces algorithmes sont très rapides et donc particulièrement adaptés au criblage de grandes bibliothèques de ligands.

Les méthodes de construction incrémentale considèrent le ligand sous forme de fragments. Le ligand est subdivisé en sous-groupes par rupture des liaisons pivotantes, puis un de ces fragments, souvent le plus volumineux ou celui portant le groupement actif du ligand, est amarré seul à la protéine.

Les autres fragments sont ensuite ajoutés par incréments. Différentes orientations sont testées pour chaque fragment afin qu'ils s'adaptent au mieux au site actif de la protéine, recréant ainsi la flexibilité du ligand. Ce principe est appliqué dans les logiciels de docking DOCK 4.0 [89], FLOG [90], LibDock [91] and SANDOCK[92].

Les algorithmes basés sur la méthode de Monte-Carlo génèrent diverses conformations pour un ligand, par rotation des liaisons et rotation ou translation de type corps rigide. Chaque conformation ainsi obtenue est évaluée par un critère de sélection basé sur une estimation de l'énergie libre de liaison (fonction de notation, pour laquelle différentes méthodes ont été mises en places et sont présentées dans la partie **IX.2**). Si ce critère est satisfait, la pose du ligand est sauvegardée et sera modifiée pour générer la conformation suivante à tester. L'avantage des méthodes de Monte-Carlo réside dans leur capacité à faire franchir au ligand des barrières d'énergie potentielle de surface, autorisant ainsi d'importants changements de conformation. Cette méthode a permis le développement des logiciels AutoDock [93], ICM [94], QXP [95] et Affinity.

Les algorithmes génétiques sont des algorithmes évolutionnistes, utilisant la théorie de la sélection naturelle émise par Darwin. Les degrés de liberté du ligand sont encodés en tant que chaînes binaires constituant les "gènes".

L'ensemble des gènes composent le chromosome (représentant en réalité une pose du ligand). Deux types d'opérateurs génétiques interviennent alors – les mutations, qui provoquent des changements aléatoires de gènes, et les *crossovers*, qui échangent les gènes entre deux chromosomes. Une nouvelle conformation du ligand résulte de chaque modification du gène par ces opérateurs génétiques.

Chacune des poses est évaluée par une fonction de notation, celles qui sont conservées (appelées "survivants" par analogie avec la génétique) servent de base à la génération de conformations suivante (**Figure 2.11**). Ce principe est appliqué dans les logiciels AutoDock [93], GOLD [96], DIVALI et DARWIN [97].

L'algorithme génétique lamarckien est une variante des algorithmes génétiques dits "darwinien", dans lesquels les modifications des gènes sont alors orientées, et non plus aléatoires. Ainsi, selon la théorie de Lamarck, les caractéristiques acquises par un individu durant sa vie peuvent devenir héréditaires. Ce principe est alors transféré aux conformations du ligand pour lequel les gènes sont modifiés un par un, régulièrement et de façon identique. Les modifications produisant une amélioration de l'énergie d'interaction récepteur-ligand (calculée par les fonctions de notation) sont conservées et transmises à la génération suivante. L'algorithme génétique lamarckien est disponible dans AutoDock [93].

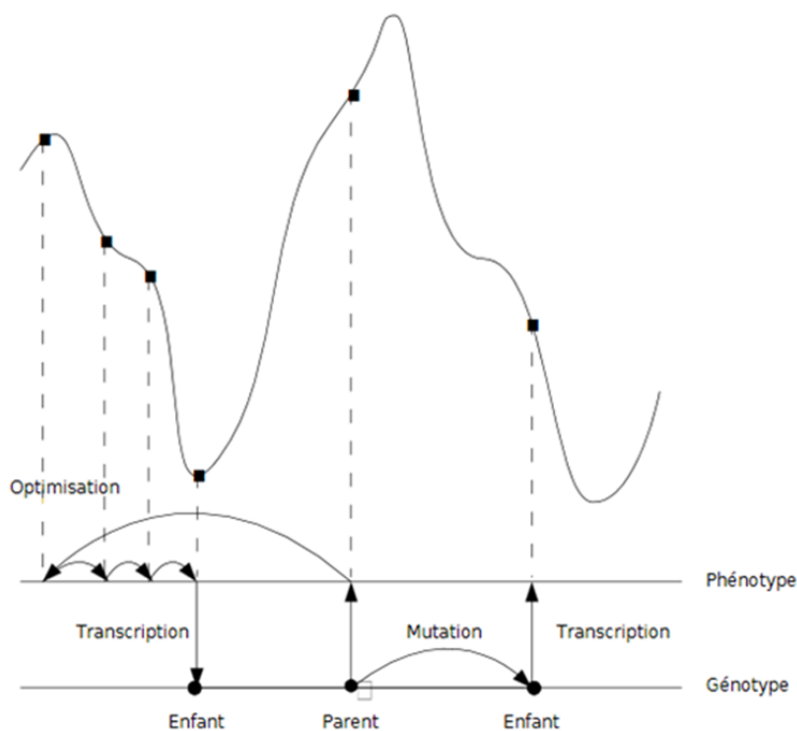


Figure 2.11. Principe de l'algorithme génétique pour l'échantillonnage des conformations d'un ligand.

2. Fonctions de notation des poses du ligand

Une évaluation des poses de ligand générées durant l'échantillonnage des conformations est ensuite nécessaire, afin d'identifier la (ou les) plus correcte(s). Ce rôle est assuré par les fonctions de notation qui estiment l'affinité de liaison entre la protéine et chaque pose du ligand (elles impliquent diverses simplifications et ne calculent donc pas la valeur absolue de l'affinité protéine-ligand). Plusieurs méthodes de notation ont été développées et peuvent être regroupées en trois groupes selon les calculs des fonctions : les fonctions basées sur un champ de force, les fonctions empiriques et les potentiels statistiques [98].

Les fonctions de notation basées sur un champ de force classique calculent la somme des interactions non covalentes (électrostatiques et van der Waals) afin d'évaluer l'énergie de liaison protéine-ligand. Les forces électrostatiques sont estimées par la loi de Coulomb, les interactions de van der Waals sont quant à elle décrites par un potentiel de Lennard-Jones. Ces fonctions de notation induisent des temps de calcul assez élevés, finalement limités par l'application d'un seuil de distance pour appréhender les interactions non covalentes.

Les effets à longue distance intervenant dans la fixation protéine-ligand sont néanmoins négligés par le recours à cette simplification. Des améliorations ont été apportées à ce type de fonction, considérant les liaisons hydrogène ainsi que les contributions de la solvation et de l'entropie dans l'affinité protéine-ligand [98]. Le recours à d'autres méthodes (perturbation de l'énergie libre ou énergie d'interaction linéaire) permet également d'affiner la prédiction de l'énergie de liaison.

Les fonctions de notations empiriques décomposent l'énergie de liaison en plusieurs composantes : liaisons hydrogène, interactions ioniques, effets hydrophobes et entropie de liaison. Chaque composante est multipliée par un coefficient puis toutes sont additionnées pour donner un score final. Les coefficients sont déterminés par analyse de régression sur un ensemble de complexes protéine-ligand dont les affinités de liaison sont connues.

L'avantage des fonctions empiriques réside dans la simplicité relative de l'estimation des composantes de l'énergie de liaison. Cependant, le nombre de ces termes et la façon dont ils sont estimés peuvent varier selon les logiciels d'amarrage utilisés.

Les potentiels statistiques reposent sur l'analyse de structures cristallographiques de complexes protéine-ligand, plus précisément sur le calcul des fréquences et/ou distances des contacts interatomiques protéine-ligand.

Le postulat de base est le suivant : plus un contact est favorable aux interactions protéine-ligand, plus sa fréquence d'apparition sera élevée. La distribution de ces fréquences est ainsi convertie en potentiel pour chaque couple d'atome observé. Le score final de chaque pose de ligand est donc calculé en favorisant les contacts préférentiels et en pénalisant les interactions répulsives entre chaque atome de la protéine et du ligand.

L'avantage des potentiels statistiques est leur simplicité et leur vitesse de calcul, rendant ces méthodes efficace pour le criblage de grandes bibliothèques de ligands. En revanche, les fréquences de certaines interactions interatomiques sont sous-estimées, le nombre de structures cristallographiques utilisées pour l'analyse statistique étant limité.

3. Amarrage moléculaire : rigide ou flexible ?

Outre la variété des méthodes d'échantillonnage et de notation des poses de ligand, trois grands types d'amarrage existent, selon la prise en compte de la flexibilité du récepteur et du ligand.

Ligand et récepteur rigides. Ces méthodes d'amarrage moléculaire reposent sur le modèle clé-serrure de Fischer et traitent le récepteur et le ligand comme deux corps rigides. L'espace de recherche des poses du ligand est donc très restreint, impliquant uniquement six degrés de liberté (trois pour les translations, trois pour les rotations). La flexibilité du ligand peut néanmoins être partiellement restituée via l'utilisation d'un ensemble de conformations distinctes préexistant. Ce type d'amarrage était le premier mis en application ; il peut encore s'avérer utile dans le cadre d'un complexe protéine-ligand pour lequel une interaction primordiale est préalablement connue.

Ligand et récepteur flexibles. Directement basée sur la théorie de l'ajustement induit de Koshland, cette méthode vise à recréer *in silico* la sélection de conformères et l'induction conformationnelle survenant lors de la fixation d'un ligand à son récepteur. Le processus de sélection de conformères se résume en la fixation sélective du ligand à une conformation spécifique de la protéine (parmi toutes les conformations possibles) ; l'induction conformationnelle décrit l'influence du ligand sur la protéine, l'orientant vers une conformation qu'elle n'adopterait pas spontanément à un état non lié. Cette induction peut être assimilée à une restructuration partielle de la protéine.

Deux méthodes principales sont disponibles pour restituer la flexibilité du récepteur :

- l'amarrage moléculaire souple diminue la valeur de l'énergie de répulsion de van der Waals dans la fonction de notation, autorisant un certain degré de chevauchement entre les atomes du ligand et ceux de la protéine ; la flexibilité ainsi reconstituée est artificielle, mais cette méthode présente l'avantage de la rapidité (les coordonnées du récepteur sont fixes, seuls les paramètres de van der Waals sont ajustés) ;

- l'utilisation de bibliothèques de rotamères, incluant un ensemble de conformations des chaînes latérales des acides aminés (issues d'analyses statistiques sur des données structurales expérimentales). Les avantages de cette méthode sont la vitesse d'échantillonnage des rotamères et le contournement des barrières énergétiques de minimisation.

Ligand flexible, récepteur rigide. Les deux méthodes précédentes décrivent la flexibilité des chaînes latérales des acides aminés. Cependant, le traitement de la flexibilité du squelette carboné des protéines reste un défi majeur de l'amarrage moléculaire, notamment par l'implication de ressources informatiques (puissance et temps de calcul) très importantes.

L'approche la plus répandue consiste donc à représenter la flexibilité du ligand mais à considérer le récepteur comme un corps rigide, aboutissant à un compromis entre la précision de la méthode et les temps de calcul nécessaires. La flexibilité du récepteur pourra être restituée via d'autres méthodes bioinformatiques, comme la dynamique moléculaire.

4. Amarrage moléculaire des ligands à VKORC1 par AutoDock

Tous les tests d'amarrage moléculaire présentés dans ces travaux ont été réalisés par le logiciel libre *AutoDock* (version 4.2) [93] ainsi que son interface graphique *AutoDockTools*. Ce programme inclut les algorithmes génétiques, génétiques lamarckiens et de Monte-Carlo pour l'échantillonnage des conformations du ligand considéré. La recherche de ces conformations s'effectue dans une grille prédéfinie par l'utilisateur (**Figure 2.12**). Dans notre cas, elle était localisée autour de la cavité identifiée par fpocket.

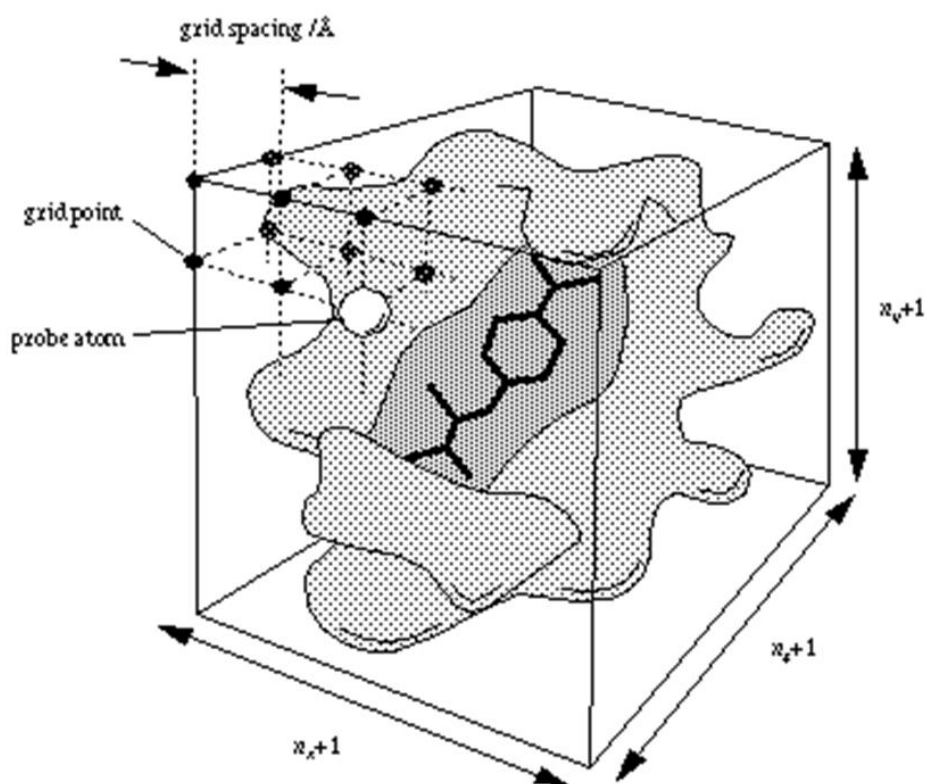


Figure 2.12. Principe de détermination de la grille de recherche des conformations d'un ligand par AutoDock. Le pas d'espacement de la grille (*grid spacing*) définit la résolution de la grille appliquée, correspondant au nombre de points utilisés (*grid point*). La recherche des conformations possible du ligand s'effectue ensuite en utilisant un atome sonde (*probe atom*) parcourant cette grille.

La fonction de notation des poses de ligand implémentée dans *AutoDock* est basée sur le champ de force *AMBER*, incluant les interactions électrostatiques et de van der Waals, les liaisons hydrogène et les contributions de la solvation et de l'entropie dans l'affinité récepteur-ligand. Chaque terme est pondéré par un facteur empirique calculé à partir de données expérimentales.

AutoDock permet de traiter la flexibilité du récepteur : l'utilisateur peut sélectionner les chaînes latérales de la protéine dont les différentes conformations seront échantillonnées et évaluées parallèlement à celles du ligand. Cette option reste cependant limitée par le nombre maximum de liaisons pivotantes que peut traiter *AutoDock* (34 rotations, incluant le ligand et le récepteur).

Tous les complexes récepteur-ligand présentés dans la partie RESULTATS ont suivi le protocole d'amarrage moléculaire suivant :

- la flexibilité du ligand est prise en compte, le récepteur est considéré comme rigide ; la flexibilité de la protéine est préalablement traitée par dynamique moléculaire, plusieurs de ses conformations ont été testées selon les complexes étudiés lors de l'amarrage moléculaire ;
- les charges partielles de Kollman ont été calculées sur le système (récepteur et ligand) ; ces charges sont dérivées d'un potentiel électrostatique issu de la chimie quantique ;
- les conformations du ligand sont échantillonnées via un algorithme génétique lamarckien ;
- les taux de mutations et de crossovers sont fixés à 2% et 80%, respectivement ;
- 150 conformations du ligand sont produites à chaque génération, 27 000 générations au maximum sont analysées, un seul survivant est autorisé à chaque génération ;
- les 100 poses du ligand présentant les scores les plus favorables sont finalement conservées.

La fonction de notation des poses d'*AutoDock* estime l'énergie de liaison entre le récepteur et le ligand ; le score obtenu pour chaque pose du ligand est d'autant plus négatif que l'interaction est favorable. Ces scores sont cependant estimatifs (pondération des termes énergétiques et omission de certains) et calculés sans intégration de l'effet d'induction provoqué par l'action réciproque entre le récepteur et le ligand.

Ainsi, la comparaison des scores permettra de différencier les poses les plus favorables pour un complexe récepteur-ligand donné, mais sera inefficace pour discriminer les affinités entre des complexes de natures différents (changement de ligand, de récepteur ou des deux). L'application d'autres méthodes est donc nécessaire pour définir les affinités respectives de différents ligands vis-à-vis d'un même récepteur, par exemple plusieurs inhibiteurs ciblant une même protéine (les AVKs inhibant l'enzyme VKORC1 dans notre cas).

X. Energie libre de liaison estimée par MM/PBSA

Le calcul de l'**énergie libre de liaison**(ELL) entre deux molécules (protéine-protéine ou protéine-ligand) permet d'estimer l'affinité entre ces deux entités. Plusieurs méthodes de calcul de l'énergie libre de liaison existent, parmi lesquelles la *Free Energy Perturbation* (FEP) et la *Thermodynamic Integration* (TI) sont les plus précises, mais aussi les plus complexes et nécessitant des temps de calcul importants. Une autre méthode, MM/PBSA (*Molecular Mechanics/Poisson Boltzmann Surface Area*), combine la mécanique moléculaire et des modèles de solvant explicites. La méthode MM/PBSA permet d'estimer l'ELL entre deux molécules le long d'une trajectoire de DM préalablement effectuée, en considérant d'une part l'énergie libre du complexe à laquelle sont soustraites les énergies libres de la protéine seule et du ligand (autre protéine ou petite molécule) seul :

$$\Delta E_{\text{liaison}} = E_{\text{complexe}} - (E_{\text{protéine}} + E_{\text{ligand}})$$

où E_{complexe} est l'ELL du complexe protéine-ligand, $E_{\text{protéine}}$ l'ELL de la protéine seule et E_{ligand} l'ELL du ligand seul. L'énergie libre E_x de chaque composante peut être décomposée de la façon suivante :

$$E_x = \langle E_{MM} \rangle - TS + \langle E_{\text{solvation}} \rangle$$

où E_{MM} est l'énergie potentielle moyenne de mécanique moléculaire dans le vide, TS représente la contribution de l'entropie à l'énergie libre dans le vide (T est la température, S le facteur d'entropie) et $E_{\text{solvation}}$ est l'énergie libre de solvation. Le facteur d'entropie n'est pas calculé durant l'estimation de l'ELL par MM/PBSA. Les ligands considérés dans cette étude étant semblables au niveau de l'entropie, le terme TS sera négligé, les ELL calculées seront donc des énergies libres relatives et non absolues. L'énergie libre E_x devient donc :

$$E_x = \langle E_{MM} \rangle + \langle E_{\text{solvation}} \rangle$$

avec $E_{MM} = E_{\text{covalentes}} + E_{\text{non covalentes}} = E_{\text{covalentes}} + (E_{\text{vdW}} + E_{\text{électrostatiques}})$ où $E_{\text{covalentes}}$ est nulle.

En effet, les calculs d'énergie libre sont effectués selon la méthode dite de la trajectoire unique, dans laquelle les énergies libres de la protéine et du ligand sont calculées à partir de la même trajectoire. Cette méthode implique une approximation selon laquelle les conformations de la protéine et du ligand sont considérées comme identiques entre les états libre et lié.

L'énergie libre associée aux liaisons covalentes est donc identique pour E_{complexe} et pour $E_{\text{protéine}} + E_{\text{ligand}}$.

$$E_{\text{solvation}} = E_{\text{polaire}} + E_{\text{non polaire}}$$

où E_{polaire} et $E_{\text{non polaire}}$ sont les contributions électrostatiques et non électrostatiques respectives à l'énergie libre de solvation. E_{polaire} est estimée par résolution de l'équation de Poisson-Boltzmann (PB) :

$$\nabla \cdot [\varepsilon(r)\nabla \cdot \varphi(r)] - \varepsilon(r)\tau(r)^2 \sinh[\varphi(r)] + \frac{4\pi\rho^f(r)}{kT} = 0$$

où $\varphi(r)$ est le potentiel électrostatique, $\varepsilon(r)$ la constante diélectrique, $\rho^f(r)$ la densité de charge et τ^2 dépend de la force ionique de la solution.

$$E_{\text{non polaire}} = E_{\text{cavité}} + E_{\text{vdW}}$$

où $E_{\text{cavité}}$ correspond au travail effectué par le solvant pour créer une cavité dans la molécule, et E_{vdW} est l'énergie d'attraction de van der Waals entre le solvant et la molécule.

Le modèle de *Solvent Accessible Surface Area* (SASA) a été utilisé pour calculer $E_{\text{non polaire}}$ contribuant aux ELL calculées dans les travaux présentés :

$$E_{\text{non polaire}} = \gamma A + b$$

où γ est un coefficient lié à la surface de tension du solvant, A est la SASA et b un paramètre d'ajustement.

Les ELL présentées dans ces travaux ont été calculées par la fonction `g_mmpbsa` de *Gromacs*[99]. Les ELL calculées sont donc des énergies relatives, elles nous ont permis de déterminer les affinités relatives des différents ligands testés avec l'enzyme VKORC1. Le nombre de conformations (*snapshots*) considérées et la constante diélectrique $\varepsilon(r)$ intervenant dans l'équation de Poisson-Boltzmann peuvent influencer les résultats [100]. Nous avons donc testé plusieurs combinaisons possibles, puis avons fixé le nombre de conformations à prendre en compte à 200 pour une simulation de DM de 100 ns, 400 pour une simulation de 200 ns (soit un pas de 500 ps). Un nombre plus important de conformations ne provoquait aucune modification de l'ELL calculée. La constante diélectrique $\varepsilon(r)$ a quant à elle été fixée à 4, cette valeur permettant d'obtenir la meilleure discrimination des ELL calculées pour les différents ligands étudiés.

Chapitre 3. RESULTATS

I. Modélisation de la structure 3D de hVKORC1

Une étape essentielle dans l'étude *in silico* de hVKORC1 est l'obtention d'un modèle structural de l'enzyme. Or, à notre connaissance, aucune structure 3D d'une VKOR mammalienne n'a été résolue expérimentalement au niveau atomique. De plus, les données biochimique et biophysiques concernant la topologie membranaire de hVKORC1 sont contradictoires.

Face à ces obstacles majeurs vers la description de la structure de hVKORC1, l'approche adoptée s'est déroulée en deux étapes. Une recherche de protéines de séquence homologue à hVKORC1 a d'abord été effectuée, afin d'identifier les résidus conservés entre ces protéines. Des méthodes bioinformatiques de prédictions de structures secondaires et de topologie membranaire ont été ensuite appliquées, nous permettant de prédire le repliement et la topologie membranaire de hVKORC1 à partir de sa séquence peptidique seule. Les résultats obtenus ont finalement été confrontés, afin de contrôler la convergence entre la topologie membranaire des protéines homologues, le repliement de hVKORC1 prédit à partir de sa séquence peptidique et les modèles topologiques (3 TMs ou 4 TMs) avancés par les expérimentateurs. L'ensemble de la démarche mise en place est présentée dans les paragraphes suivants.

1. Alignements de séquences de VKORs et recherche de protéines homologues

La séquence peptidique de l'enzyme hVKORC1 a été obtenue dans la banque de données *UniProt* (*Q9BQB6*). La séquence canonique a été considérée, parmi les 3 isoformes disponibles (**Figure 3.1**).

10 20 30 40 50 60 70 80
 MGSTWGSPPGW VRLALCLTGL VLSLYALHVK AARARDRDYR ALCDVGTAIS CSRVFSSRWG RGFGLVEHVL QDSILNQSN
 90 100 110 120 130 140 150 160
 SIFGCIFYTL QLLGCLRTR WASVIMLLSS LVSLAGSVYL AWILFFVLYD FCIVCITTYA INVSLMWLSF RKVQEPQGKA KRH

Figure 3.1. Séquence peptidique de hVKORC1 (forme canonique) choisie pour la modélisation. Les résidus sont numérotés par 10. Les résidus de cystéine du site actif (C₁₃₂ et C₁₃₅) ainsi que C₄₃ et C₅₁ sont surlignés en bleu.

Un alignement des séquences peptidiques des VKORs de 32 espèces a été réalisé en 2004 par Goodstadt et Ponting [21], afin d'identifier les résidus de VKOR conservés. Il a été démontré que les résidus de cystéine C₁₃₂ et C₁₃₅ (la numérotation correspond à la séquence de hVKORC1) étaient strictement conservés chez les VKORs des espèces étudiées (**Figure 3.3**). Ces deux résidus de cystéine, constituant le site actif de VKOR, forment un motif CXXC spécifique aux oxydoréductases (voir **INTRODUCTION.II.1**). L'appartenance de VKOR à cette famille d'enzymes a ainsi été établie. En outre, l'alignement de séquences a révélé une conservation stricte des résidus C₄₃ et C₅₁ chez les 32 espèces, apportant un premier argument en faveur de leur implication dans le mécanisme enzymatique de VKOR.

Une recherche de protéines homologues à hVKORC1 et dont la structure 3D est résolue expérimentalement a ensuite été effectuée, en utilisant l'outil pBLAST pour cribler les données structurales disponibles dans la PDB. La seule protéine homologue caractérisée par cristallographie est la VKOR de la cyanobactérie *Synechococcus sp.* (bVKOR). La séquence peptidique de bVKOR compte 283 résidus, soit 120 de plus que hVKORC1. Le fragment de séquence s'étendant des résidus 186 à 283 constitue un domaine thiorédoxine, absent chez hVKORC1. Les fragments de séquence les plus proches entre les enzymes humaine et bactérienne présentent une identité (21%) et une homologie (39%) faibles (**Figure 3.3**).

Quatre structures cristallographiques sont disponibles dans la PDB [32, 101] pour l'enzyme bactérienne, cristallisée en présence d'ubiquinone, son ligand naturel (**Tableau 3.1**). Le domaine de bVKOR homologue à hVKORC1 comprend 4 hélices transmembranaires (TM) (**Tableau 3.1**). Les données structurales et l'étude biochimique basée sur l'analyse de trois mutations simples, introduites au niveau des résidus cystéine [101] (**Tableau 3.1**), ont permis la description du mécanisme d'auto-activation de bVKOR. Le cycle enzymatique de la protéine bactérienne suggère un transfert du pouvoir réducteur du domaine thiorédoxine vers le site actif ¹³⁰CXXC₁₃₃ par l'intermédiaire des résidus C₅₀ et C₅₆ [101], schématisé dans la **Figure 3.4**. Par analogie, il a été admis que hVKORC1 était aussi structurée en 4 TMs, et que son mécanisme enzymatique était probablement similaire à celui de bVKOR [101].

	10	20	30	40	50	60	70	80	
hVKORC1	MGS T WGS P GW	VR L ALCL T GL	VL S L Y AL H V K	AA R AR D R D Y R	A LC D V G T A IS	C SR V F S SR R W G	R G F L VE H V L	G Q D S I L N Q S N	
bVKOR	--- T W L Q R H S	RL L LA I LA G L	G- S L L T A Y L T	Y T K L -TE Q PA	A FC T --GD G G	C DL V L S SR W A	E F L G I P T A AV	G LL G F L GV L -	
	13								
	90	100	110	120	130	140	150	160	
hVKORC1	S IF G CI F Y T L	Q LL G CI R ER	W AS V L M LS S	L V S LAG S V L	A W I L F F V L Y D	F CI V CI T TY A	I NV S L M W L S F	R K V Q E P O Q G KA	KRH
bVKOR	A LA--V L P--	D GL P L V K R R	W P A L F CL V SA	M T A FE-- M Y M	L Y L M V A V L R Q	F CM Y CI T AI I	L V A CL G -L V T	--V L GH R W L D	GGK (159-283) 158

Figure 3.3. Alignement des séquences peptidiques de la VKORC1 humaine et d'une VKOR bactérienne (*Synechococcus sp.*). Les résidus sont numérotés par 10. Les acides aminés conservés entre les deux espèces sont écrits en noir et surlignés en bleu ; les résidus similaires sont écrits en bleu. Les quatre résidus cystéine C₄₃, C₅₁, C₁₃₂ et C₁₃₅ (C₅₀, C₅₆, C₁₃₀ et C₁₃₃ chez bVKOR) sont encadrés en rouge. La séquence peptidique de bVKOR correspondant à hVKORC1 est numérotée en gris (seuls les résidus N- et C-terminaux sont indiqués). Les résidus 159 à 283 de bVKOR ne sont pas présentés dans un souci de clarté.

L'absence de domaine thiorédoxine chez hVKORC1 ne permet cependant pas de conclure quant au rôle des résidus C₄₃ et C₅₁ (homologues aux résidus C₅₀ et C₅₆ de bVKOR, respectivement). Une hypothèse avancée dans la littérature suggère l'intervention d'un partenaire protéique de fonction similaire au domaine thiorédoxine de bVKOR dans l'activation de hVKORC1 [29]. De plus, la faible homologie de séquence entre hVKORC1 et bVKOR impose d'autres recherches pour déterminer de façon plus probante l'organisation structurale de l'enzyme humaine. Des méthodes bioinformatiques de prédiction de structures secondaires et de topologie membranaire ont donc été utilisées.

3KP9 (chaîne A)	4NV2 (chaîne A)	4NV5 (chaîne A)	4NV5 (chaîne B)	4NV6 (chaîne A)
3.6 Å 16 – 279 53-55 91-92 C56S	3.6 Å 17 – 282 92-93 C50A	2.8 Å 17 – 282 49-53 92 155 C50A	2.8 Å 13 – 279 49-53 C50A	4.2 Å 17 – 282 92 155 C212A
C50 _{S...} C209 C56S C130 _{S...} C133 C212 _{SH}	C50A C56 _{S...} C130 C133 _{SH} C209 _{S...} C212	C56 _{S...} C130 C133 _{SH} C209 _{S...} C212	C56 _{S...} C130 C133 _{SH} C209 _{S...} C212	C50 _{S...} C209 C56 _{S...} C130 C133 _{SH} C212A

Tableau 3.1. Structures cristallographiques de bVKOR (*Synechococcus sp.*) disponibles dans la PDB. Chaque structure, identifiée par son code PDB (première ligne), est caractérisée par sa résolution, la séquence peptidique de la structure cristallisée (en bleu foncé), les résidus absents (en italique bleu clair) et les mutations (en vert). Les structures sont représentées en *cartoon* ; la région colorée en orange correspond au domaine homologue à hVKORC1, identifié d'après l'alignement de séquences (**Figure 3.3**) ; la région colorée en gris correspond à une hélice transmembranaire et au domaine thiorédoxine, absents chez hVKORC1. Le ligand (ubiquinone) est représenté en *sticks*, les atomes de carbone et d'oxygène sont colorés en rose et rouge clair, respectivement. La région des structures encadrée par des pointillés est présentée en détail. Les résidus de cystéine et le ligand sont représentés en *sticks* ; les atomes d'azote, d'oxygène et de soufre sont colorés en bleu, rouge et jaune, respectivement. Les résidus mutés sont représentés en *sticks* vert. Les ponts disulfures entre les résidus de cystéine sont codés par couleurs – noir pour les résidus appartenant au domaine homologue à hVKORC1, gris pour ceux du domaine thiorédoxine (absent chez hVKORC1) ; les résidus mutés sont en vert.

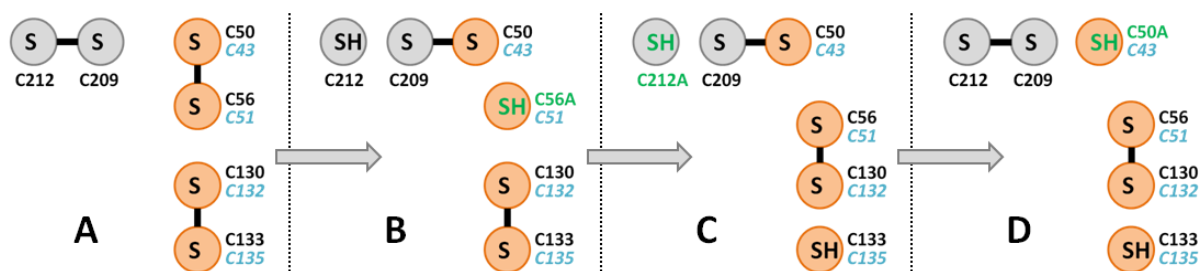


Figure 3.4. Mécanisme enzymatique décrit pour bVKOR et proposé par analogie pour hVKORC1 [101].

(A) Les résidus C₂₁₂ et C₂₀₉ du domaine thiorédoxine sont à l'état déprotoné (S) et forment un pont disulfure ; les résidus C₅₀, C₅₆, C₁₃₀ et C₁₃₂ du domaine VKOR sont également à l'état déprotoné et associés deux à deux pour former deux ponts disulfures. (B) La mutation C₅₆A empêche la formation d'un pont disulfure avec C₂₀₉ et force l'association entre les résidus C₂₀₉ et C₅₀ pour former un pont disulfure. Cette conformation mime un état protonné (SH) de C₅₆. (C) La mutation du résidu C₂₁₂A (alors considéré à l'état protonné) bloque sa réassociation avec C₂₀₉, alors forcé de former un pont disulfure avec C₅₀. Le résidu C₅₆ forme ainsi un pont disulfure avec C₁₃₀. C₁₃₃ est à l'état protonné. (D) La mutation C₅₆A, mimant l'état protonné de ce résidu, induit la formation de deux ponts disulfure, entre C₂₀₉ et C₂₁₂ d'une part et C₅₆ et C₁₃₀ d'autre part. Les deux conformations, présentées en (C) et (D), correspondent à un état actif de bVKOR, C₁₃₃ étant le résidu catalytique responsable de la dégradation de l'ubiquinone. Le mécanisme enzymatique ainsi postulé pour bVKOR implique un enchaînement de formations et ruptures de ponts disulfures. Les résidus localisés dans le domaine thiorédoxine (absent chez hVKORC1) sont colorés en gris, ceux appartenant à la région homologue à hVKORC1 sont colorés en orange. Les numéros des quatre résidus cystéine conservés entre les deux espèces sont indiqués en noir pour bVKOR et en italique bleu pour hVKORC1. Les résidus mutés sont annotés en vert.

2. Prédiction des structures secondaires et de la topologie membranaire de hVKORC1 à partir de sa séquence peptidique

La détermination du repliement structural de hVKORC1 à partir de sa séquence peptidique a été effectuée par l'application de méthodes de prédictions de structures secondaires et de topologie membranaire. Ces méthodes ont été développées sur des concepts différents et basées sur des algorithmes distincts (voir METHODES). Les consensus des structures secondaires et topologies membranaires prédites par ces méthodes ont ensuite été établis. Les logiciels utilisés ont été préalablement testés sur l'homologue bVKOR, puis les résultats obtenus ont été comparés aux structures cristallographiques. L'ensemble des méthodes utilisées a abouti à une prédiction fidèle des structures secondaires et des régions transmembranaires de bVKOR pour 96% de sa séquence peptidique.

Les cinq TM de la protéine bactérienne ainsi que le repliement du domaine thiorédoxine sont correctement prédits (Figure 3.5).

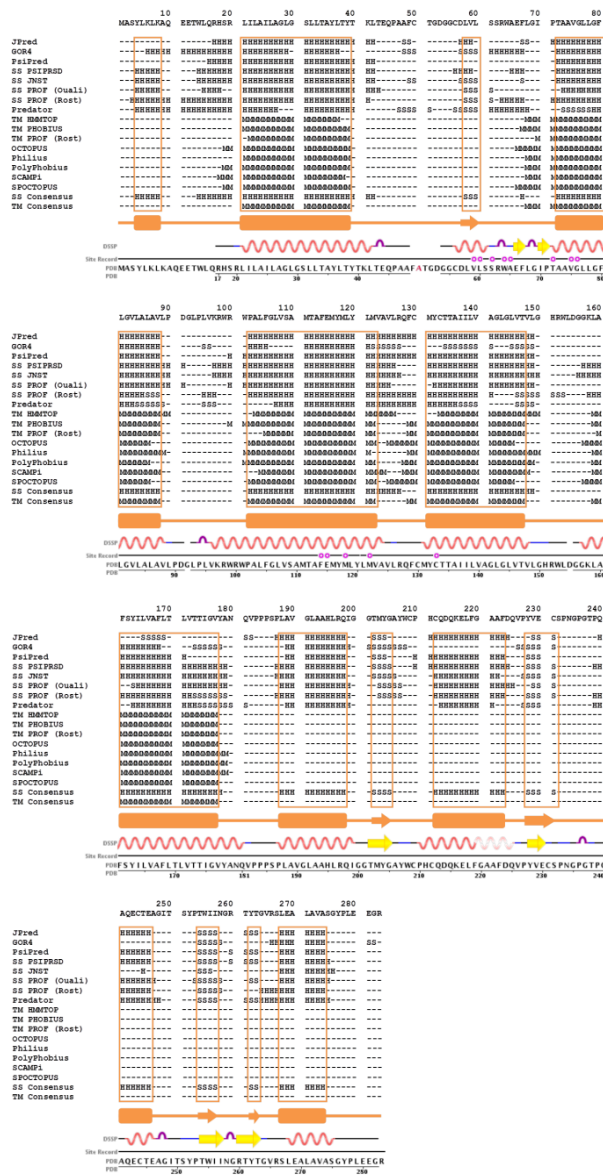


Figure 3.5. Prédiction des structures secondaires et de la topologie membranaire de bVKOR à partir de sa séquence peptidique. Les résidus de bVKOR sont numérotés par 10. Les méthodes de prédiction employées sont indiquées au début de chaque ligne. Les huit prédictions de structures secondaires sont présentées, suivies des huit prédictions de topologie membranaire. Les consensus établis sont présentés dans les deux dernières lignes. Les structures secondaires et localisations membranaires ainsi prédites sont délimitées par des cadres de couleur orange et schématisées sous les consensus, accompagnées de la structuration de bVKOR définie par DSSP. Les hélices α sont indiquées par des rectangles de couleur orange et des sinusoides colorés en rouge ; les feuilletts β sont représentés par des flèches, les boucles par une ligne. Pour chaque résidu, un repliement en hélice α est indiqué par un H, un brin β par un S et une absence de structuration régulière (boucle) par un tiret ; M indique la prédiction d'une localisation membranaire, un tiret signifie que le résidu concerné est extra-membranaire. Les résidus pour lesquels aucun consensus n'est obtenu sont considérés comme non structurés (boucle) pour les structures secondaires ou comme extra-membranaires pour la topologie membranaire.

Les seules prédictions erronées concernent deux brins β courts non prédits (résidus 66-67 et 70-71) et la petite hélice α horizontale (résidus 56-61, pour lesquels une structuration en brin β est prédite). Le protocole de prédiction ainsi validé a ensuite été appliqué à la séquence peptidique de hVKORC1, dont les résultats sont présentés dans la **Figure 3.6**.

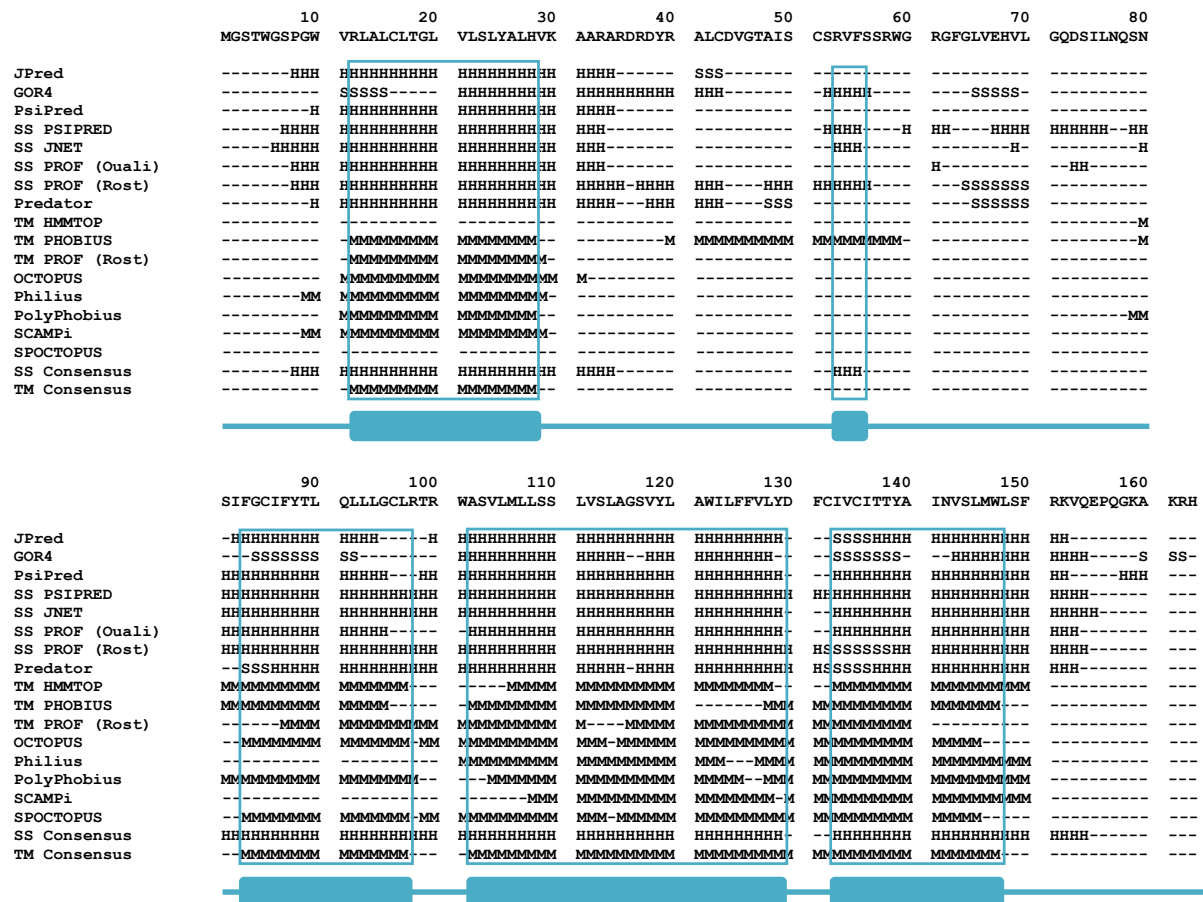


Figure 3.6. Prédiction des structures secondaires et de la topologie membranaire de hVKORC1 à partir de sa séquence peptidique. Les résidus de hVKORC1 sont numérotés par 10. Les méthodes de prédiction employées sont indiquées au début de chaque ligne. Les huit prédictions de structures secondaires sont présentées, suivies des huit prédictions de topologie membranaire. Les consensus établis sont présentés dans les deux dernières lignes. Les structures secondaires et localisations membranaires ainsi prédites sont délimitées par des cadres de couleur bleue et schématisées sous les consensus. Les hélices α sont indiquées par des rectangles de couleur bleue et les boucles par une ligne. Pour chaque résidu, un repliement en hélice α est indiqué par un H, une absence de structuration régulière (boucle) par un tiret ; M indique la prédiction d'une localisation membranaire, un tiret signifie que le résidu concerné est extra-membranaire. Les résidus pour lesquels aucun consensus n'est obtenu sont considérés comme non structurés (boucle) pour les structures secondaires ou comme extra-membranaires pour la topologie membranaire.

Le consensus obtenu à partir des résultats des prédictions indique **un repliement de hVKORC1 en quatre TMs**, s'étendant des résidus 12 à 28 (TM1), 83 à 97 (TM2), 101 à 129 (TM3) et 133 à 147 (TM4). TM1 et TM2 sont séparées par une longue boucle de 54 résidus (L), contenant une courte hélice α entre les résidus 53 et 55 (HH). Le repliement structural ainsi prédit pour hVKORC1 a été schématisé (**Figure 3.7**) et comparé au repliement de bVKOR (ID PDB : 4NV5) interprété par DSSP [102].



Figure 3.7. Alignement des structures secondaires prédites pour hVKORC1 et des structures secondaires de son homologue bVKOR. Les résidus sont numérotés par 10. La séquence de hVKORC1 est utilisée comme référence. Les structures secondaires prédites pour hVKORC1 sont schématisées en bleu au-dessus de l'alignement ; le repliement de bVKOR défini par DSSP et indiqué sous l'alignement. Les hélices α sont représentées par des rectangles bleus ou des sinusoides rouges, les boucles sont indiquées par une ligne simple. Les acides aminés conservés entre les deux espèces sont écrits en noir et surlignés en bleu ; les résidus similaires sont écrits en bleu. Les quatre résidus cystéine C₄₃, C₅₁, C₁₃₂ et C₁₃₅ (C₅₀, C₅₆, C₁₃₀ et C₁₃₃ chez bVKOR) sont encadrés en rouge.

Les trois TMs, TM1, TM3 et TM4, HH et la première moitié de la boucle L de hVKORC1 sont bien alignées vis-à-vis de leurs domaines homologues chez bVKOR (**Figure 3.7**). En revanche, TM2 (dont l'existence reste discutable dans la littérature [2, 16, 20, 21, 23, 31-38, 101]) est alignée sur la boucle reliant les deuxième et troisième hélices transmembranaires de bVKOR. Ce non-appariement résulte de la faible identité de séquence entre la TM2 de hVKORC1 et la région de bVKOR correspondant. Néanmoins, la prédiction d'un repliement à 4 TMs pour hVKORC1 et la topologie membranaire similaire de son homologue bVKOR convergent vers l'hypothèse d'une structuration de hVKORC1 en 4 TMs. Il été systématiquement observé que le repliement des protéines membranaires est fortement conservé en comparaison des protéines cytoplasmiques [72]. Cette observation est particulièrement vérifiée pour des protéines dont l'identité de séquence est comprise entre 20 et 35%. Par conséquent, nous avons choisi la structure cristallographique de bVKOR comme *template* pour générer un modèle 3D de hVKORC1 montrant une faible homologie de séquence (21%).

3. Modélisation par homologie de la structure 3D de VKORC1

La structure 3D de hVKORC1 a été modélisée à partir de la structure cristallographique de bVKOR présentant la meilleure résolution (ID PDB : 4NV5, 2.8 Å) [101]. Le modèle 3D de hVKORC1 a été généré par MODELLER (Méthodes), en utilisant l'alignement de séquences réalisé (**Figure 3.3**). Des contraintes de structures secondaires basées sur les prédictions de structures secondaires (**Figure 3.6**) ont été appliquées. Une structuration en hélice α a ainsi été imposée entre les résidus 12 et 28 (TM1), 53 et 55 (HH), 83 et 97 (TM2), 101 et 129 (TM3), 133 et 147 (TM4). Le modèle 3D ainsi généré pour hVKORC1 est présenté dans la **Figure 3.8**. Les résidus conservés ou homologues entre hVKORC1 et bVKOR sont majoritairement localisés dans les TMs et HH (**Figure 3.8.C**). La projection des prédictions de structures secondaires et topologie membranaire sur le modèle 3D (**Figure 3.8.D-E**) indique des taux de prédiction de 100% du cœur des 4 TMs. La localisation extra-membranaire de la boucle L et de l'hélice HH est également prédite avec une haute certitude (100%). Seules les extrémités de HH, de TM2 et de TM4 affichent des taux de prédiction de structures secondaires faibles (50%). Les taux de prédiction de topologie membranaire incertains (50%) concernent les extrémités de TM1 et de TM4.

Les précédents modèles structuraux de hVKORC1 générés par homologie avec bVKOR, publié dans la littérature, comprennent tous un pont disulfure entre les résidus C₅₁ et C₁₃₂, observé dans la structure *template*[101]. En effet, au cours de la réduction de la vitamine K, hVKORC1 adopte plusieurs états enzymatiques. Le pont disulfure liant les résidus de cystéine du site actif, C₁₃₂ et C₁₃₅, est "dynamique", les résidus de cystéine participant au cycle enzymatique sont donc alternativement à l'état protoné (SH) et déprotoné (S⁻). Un mécanisme d'activation similaire au cycle enzymatique de bVKOR a été proposé pour hVKORC1 (**Figure 3.4**), suggérant la formation d'un pont disulfure entre les résidus C₅₁ (localisé sur l'hélice HH, portée par la boucle L qui serait alors orientée dans la lumière du RE) et C₁₃₂. Cette association provoquerait la rupture du pont disulfure C₁₃₂-C₁₃₅, induisant un état protoné (SH) du résidu C₁₃₅, correspondant à la forme active de l'enzyme. Cependant, contrairement à bVKOR, hVKORC1 est dépourvue de domaine thiorédoxine (permettant l'auto-activation de l'enzyme bactérienne). L'implication des résidus C₄₃ et C₅₁ dans le mécanisme d'activation de hVKORC1 reste donc incertaine.

Pour éviter l'influence des données structurales de bVKOR (4NV5) sur la formation de ponts disulfures *a priori* chez hVKORC1, et pour rester objectifs vis-à-vis du rôle des résidus de cystéine dans le cycle enzymatique, nous avons considéré les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ de notre modèle structural à l'état protoné (SH) (**Figure 3.8.B**). L'absence de pont disulfure dans le modèle de hVKORC1 généré confère à la structure un maximum de flexibilité, sans contrainte préalable d'association entre résidus cystéine. Nous avons néanmoins présupposé que les simulations de DM de ce modèle pourraient générer des conformations montrant objectivement une prédisposition naturelle de certains résidus de cystéine à former des ponts disulfures S...S.

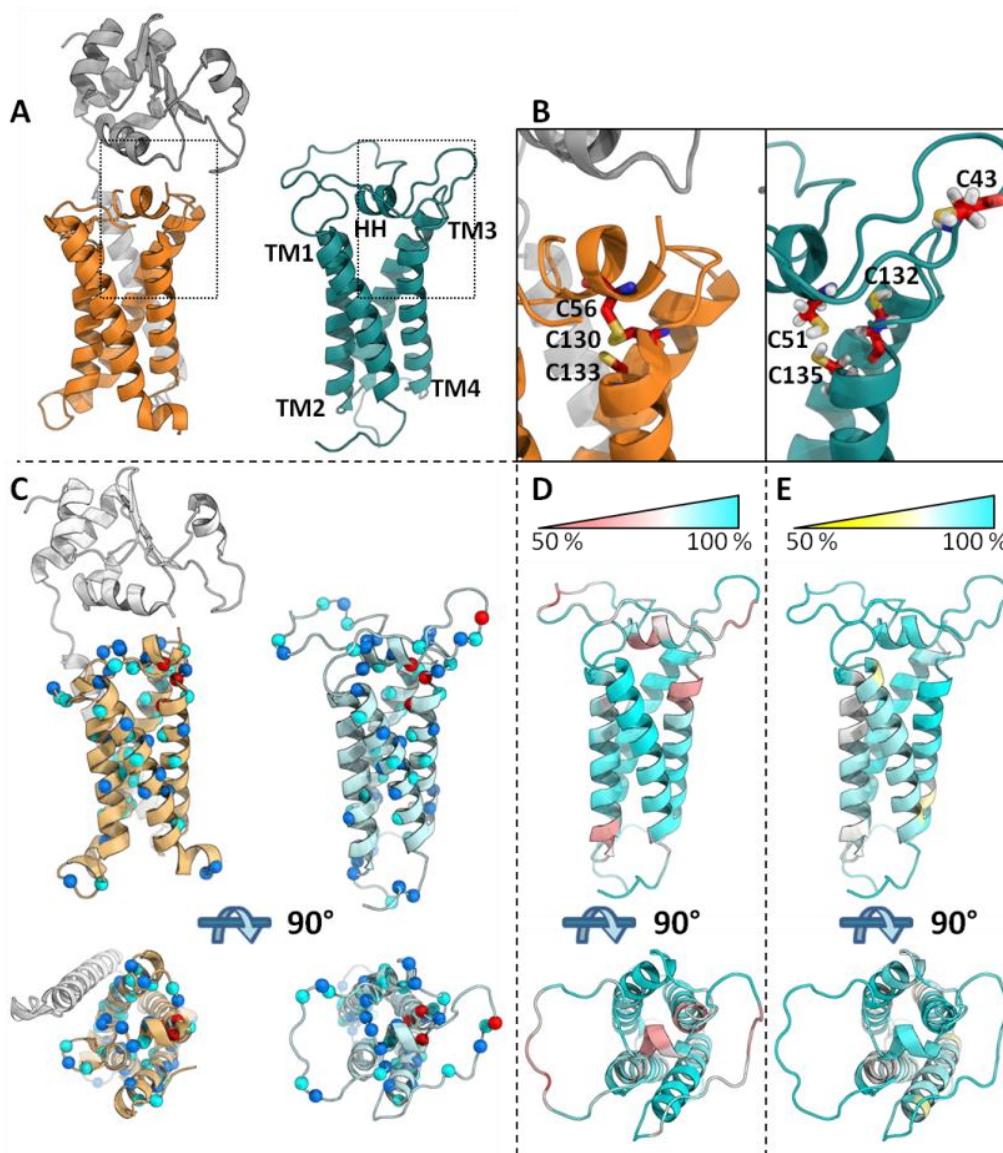


Figure 3.8. Modélisation par homologie de la structure 3D de hVKORC1 à partir de la structure cristallographique de bVKOR. (A) La structure cristallographique 4NV5 (en cartoon) de bVKOR est représentée à gauche ; la région homologue à hVKORC1 est colorée en orange, le reste de la protéine en gris. Le modèle structural généré pour hVKORC1 est représenté à droite (cartoon, en bleu). TM1 est composée des résidus 9 à 30, HH des résidus 51 à 56, TM2 des résidus 78 à 99, TM3 des résidus 102 à 125 et TM4 des résidus 132 à 152. Les régions encadrées par des pointillés sont représentées en détail dans le cadre B. (B) Les résidus de cystéine sont représentés en sticks, les atomes de carbone, d'azote et de soufre sont colorés en rouge, bleu et jaune, respectivement. Les résidus C₅₆ et C₁₃₀ (homologue à C₅₁ et C₁₃₂ chez hVKORC1) forment un pont disulfure chez bVKOR ; C₁₃₃ est à l'état protoné ; C₅₀ n'a pas été résolu. Les quatre résidus cystéine C₄₃, C₅₁, C₁₃₂ et C₁₃₅ du modèle 3D de hVKORC1 ont été modélisés à l'état protoné. (C) Les résidus conservés et homologues entre les deux protéines sont représentés par des sphères bleues et cyans, respectivement ; les résidus C₅₆, C₁₃₀ et C₁₃₃ (C₄₃, C₅₁, C₁₃₂ et C₁₃₅ chez hVKORC1) sont indiqués par des sphères rouges. (D) et (E) Les consensus des structures secondaires et de topologie membranaire prédites pour hVKORC1 sont encodés par des dégradés de couleurs, du rose au cyan et du jaune au cyan, respectivement ; les résidus colorés en cyan indiquent un taux de prédiction égal 100%, ceux colorés en jaune ou rouge un taux de 50%.

4. Modélisation de la membrane du réticulum endoplasmique

L'enzyme VKORC1 est localisée dans la membrane du réticulum endoplasmique (RE), la construction *in silico* d'un modèle de bicouche lipidique reflétant le plus fidèlement possible la membrane du RE (**Figure 3.9**) est souhaitable. Quatre types de lipides distincts composent la membrane du RE des mammifères, la phosphatidylcholine étant majoritaire (plus de la moitié des lipides). Un cinquième de la membrane du RE est constituée de phosphatidyléthanolamine, tandis que le phosphatidylinositol et la phosphatidylsérine composent à eux deux 14% de la membrane. Enfin, la sphingomyéline est très minoritaire puisqu'elle ne représente que 4% des lipides de la membrane du RE.

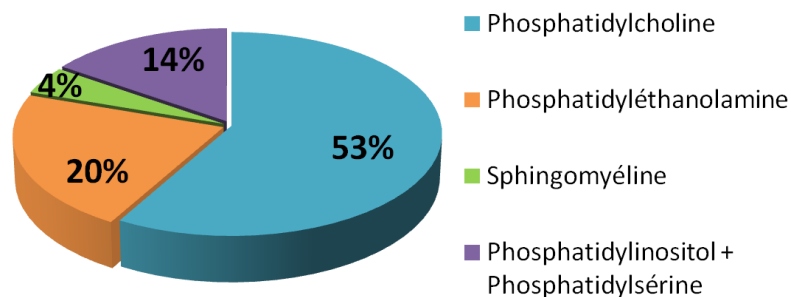


Figure 3.9. Composition lipidique générale de la membrane du RE mammalien.

Cependant, le site d'ancrage de VKORC1 à la membrane du RE n'est pas décrit dans la littérature à l'heure actuelle. De plus, les bicouches lipidiques modélisées *in silico* sont de faibles dimensions (entre 200 et 800 lipides, soit 70 à 140 Å de côté) [77]. La membrane modélisée ne pourra donc pas reproduire la composition réelle de la membrane du RE. En effet, les lipides minoritaires, telle que la sphingomyéline, seraient en effectif très faible (entre 2 et 4 par monocouche) et susceptibles d'être supprimés durant l'intégration du modèle 3D de VKORC1 dans la membrane. La présence d'un éventuel radeau lipidique (s'étendant sur plusieurs dizaines ou centaines de nanomètres) sera également impossible à représenter, compte tenu de la taille restreinte de la membrane modélisée.

Face à ces deux contraintes, le choix a été fait de modéliser une bicouche contenant uniquement le type de lipide majoritaire de la membrane du RE, la phosphatidylcholine (DLPC). Une bicouche lipidique composée de 200 molécules de DLPC a donc été modélisée à l'aide du serveur *MemBuilder* (**Figure 3.10**).

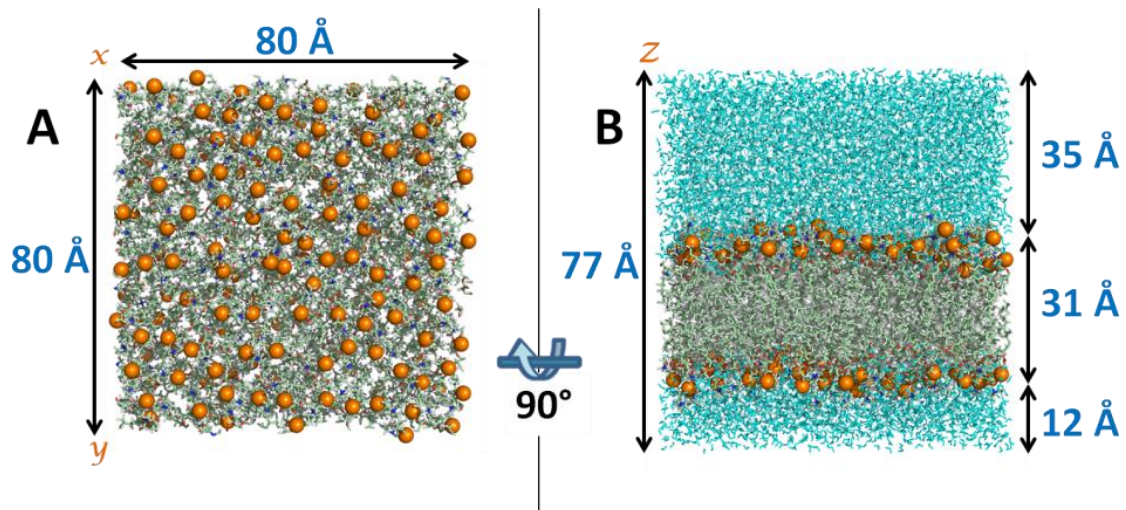


Figure 3.10. Modèle de la membrane du RE. La membrane modélisée intègre 200 molécules de 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC). Les dimensions de la membrane et l'épaisseur de la couche d'eau modélisées sont indiquées en bleu. (A) Membrane modélisée vue du dessus. Les molécules d'eau ne sont pas représentées pour plus de clarté. (B) Bicouche lipidique solvatée vue de côté. Les lipides sont représentés en sticks et colorés en vert, les atomes de phosphates sont indiqués par des sphères de couleur orange et les molécules d'eau sont représentées en *sticks* de couleur cyan.

5. Intégration du modèle structural de hVKORC1 dans la membrane du réticulum endoplasmique

La connaissance du positionnement de l'enzyme dans la bicouche lipidique est cruciale pour l'intégration correcte du modèle structural de hVKORC1 dans la membrane modélisée. Le domaine transmembranaire du modèle de hVKORC1 a dans un premier temps été défini grâce au serveur *Positioning of Proteins in Membrane* (PPM, voir METHODES). La région ainsi identifiée a été positionnée entre les deux plans définis par les positions moyennes des atomes de phosphate de chaque monocouche lipidique. Plusieurs propriétés des résidus composant hVKORC1 ont ensuite été considérées pour contrôler l'ancrage du modèle de hVKORC1 à la membrane (**Figure 3.11**), basées en grande partie sur les caractéristiques des protéines membranaires décrites par Nugent et Jones [103].

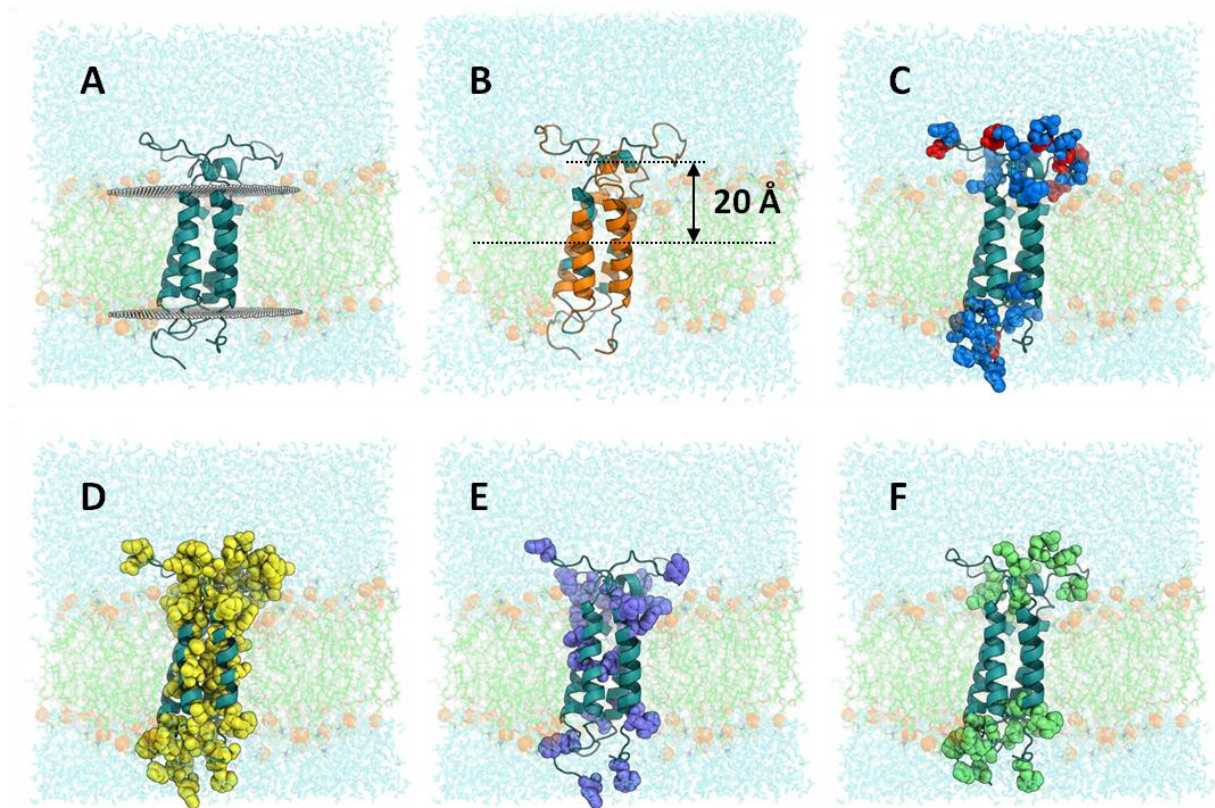


Figure 3.11. Intégration du modèle 3D de hVKORC1 dans la membrane du RE modélisée. (A) Les limites de la région membranaire de hVKORC1 définies par PPM sont indiquées par des sphères blanches. (B) Les résidus hydrophobes de hVKORC1 sont colorés en orange. L'axe de la petite hélice HH et le plan définissant le milieu de la membrane, distants de 20 Å, sont indiqués par des pointillés. (C) Les résidus de hVKORC1 chargés positivement et négativement sont représentés par des sphères bleues et rouges, respectivement. (D) Les résidus polaires de hVKORC1 sont représentés par des sphères jaunes. (E) Les résidus aromatiques de hVKORC1 sont représentés par des sphères mauves. (F) Les résidus R, K et W de hVKORC1 sont représentés par des sphères vertes. Le modèle 3D de hVKORC1 est représenté en cartoon bleu. Les lipides sont représentés en sticks verts, les atomes de phosphate par des sphères de couleur orange, les molécules d'eau en sticks de couleur cyan.

Cette étude a démontré que les hélices transmembranaires étaient riches en résidus hydrophobes, tandis que les résidus chargés sont généralement localisés au niveau des boucles cytoplasmiques, ou aux extrémités des TM (tout comme les résidus aromatiques). En revanche, les régions transmembranaires sont généralement pauvres en résidus polaires, lesquels se retrouvent alors enfouis vers l'intérieur de la protéine ou orientés vers les têtes phospholipidiques. Les résidus R, K et W sont quant à eux localisés à l'interface entre l'eau et la membrane. Enfin, toutes les boucles cytoplasmiques composées de 30 résidus ou plus et séparant deux hélices transmembranaires contiennent une hélice interfaciale, dont l'axe est parallèle au plan de la membrane.

Ces hélices sont composées de deux régions hydrophobe et hydrophile, orientées respectivement vers la membrane et la phase aqueuse. La distance séparant une hélice interfaciale du centre de la membrane est comprise entre 15 et 25 Å.

Dans notre modèle, la petite hélice HH présente toutes les caractéristiques d'une hélice interfaciale (**Figure 3.11.B**). Le positionnement du modèle 3D de hVKORC1 dans la membrane du RE reflète fidèlement ces propriétés spécifiques aux protéines membranaires. La fonction *g_membed* de GROMACS (**METHODES**) nous a ensuite permis d'intégrer la structure de hVKORC1 modélisée dans la membrane du RE.

Le système 'hVKORC1 + membrane' assemblé comporte, une fois le solvant ajouté, 2 599 atomes composant le modèle structural de hVKORC1, 194 molécules de DLPC (lipides), 11 771 molécules d'eau et 9 ions Cl⁻ ajoutés pour neutraliser les charges positives portées par hVKORC1.

II. Simulations de DM de hVKORC1 considérant les résidus de cystéine protonés

Le comportement dynamique du modèle 3D hVKORC1 a été étudié par simulations de dynamique moléculaire (DM). Ces simulations permettent en premier lieu d'affiner la structure générée par homologie, puis d'explorer une partie de l'espace conformationnel accessible au modèle 3D. L'absence de repliement régulier au niveau de la boucle L de hVKORC1 risque de conduire à une grande variabilité structurale de cette région. Cinq répliques (1-5) de 100 ns chacun ont donc été réalisés, en appliquant des vitesses initiales aléatoires pour étudier la diversité conformationnelle de hVKORC1. Deux simulations (1 et 2) ont ensuite été prolongées jusqu'à 1 µs (1→1' et 2→2') afin d'augmenter les probabilités d'atteindre un état métastable du modèle 3D de hVKORC1. Les résultats obtenus sont présentés dans les paragraphes suivants.

1. Evaluation de la stabilité structurale du modèle 3D de hVKORC1 par simulations de dynamique moléculaire de 100 ns

La détermination des structures secondaires par DSSP, le calcul des valeurs des RMSDs (**Figure 3.12**) et de RMSFs (**Figure 3.13**) ont permis de caractériser la stabilité structurale du modèle de hVKORC1 au cours des simulations. Les valeurs de RMSDs augmentent progressivement jusqu'à 30 ns de simulation, puis oscillent entre 3,5 et 7 Å selon les réplicas (**Figure 3.12.A**). L'amplitude des fluctuations reste cependant élevée jusqu'à 100 ns, la structure du modèle n'est donc que partiellement stabilisée. En revanche, le repliement des quatre TMs est très bien préservé durant les cinq simulations de DM, comme le montre l'assignement des structures secondaires par DSSP (**Figure 3.12.B**). La structure de l'hélice HH alterne entre hélice α et hélice_{3.10}. Certaines régions de la boucle L adoptent un repliement en coude (notamment les résidus délimitant les hélices TM1, TM2 et HH). Enfin, les valeurs de RMSFs (**Figure 3.13**) indiquent de faibles fluctuations atomiques des résidus localisés au niveau des TMs (environ 1 Å). Les résidus composant l'hélice HH et les boucles liant les TMs présentent des valeurs de RMSDs situées entre 1 et 3 Å. Les fluctuations les plus importantes concernent les extrémités N- et C-terminales ainsi que les résidus composant la boucle luminale L, pour lesquels les valeurs de RMSFs sont supérieures à 4 Å. La projection des valeurs de RMSFs sur le modèle 3D de hVKORC1 révèle que les régions de la boucle L les plus fluctuantes varient d'un réplica à l'autre (**Figure 3.13.B**).

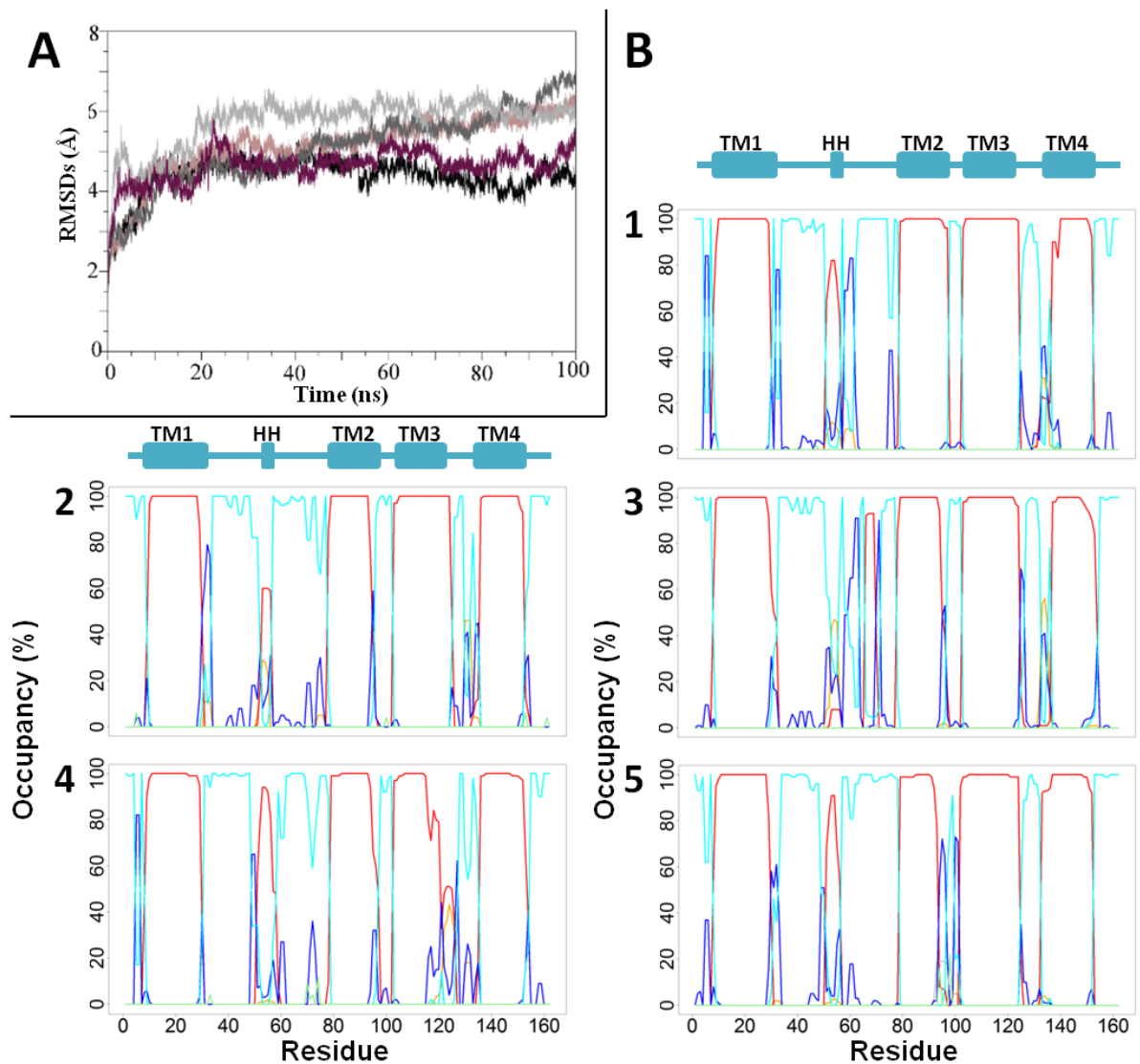


Figure 3.12. RMSDs et assignement par DSSP des structures secondaires du modèle 3D de hVKORC1. (A) Courbes des RMSDs du modèle 3D de hVKORC1 au cours de cinq simulations de DM de 100 ns chacune. Les RMSDs de chaque conformation de hVKORC1 ont été calculées sur les carbones α des résidus 9 à 151 (les résidus des extrémités N- et C-terminales ont été omis afin d'éliminer le bruit induit par leurs fluctuations), après alignement des carbones α des hélices transmembranaires sur les coordonnées de la conformation initiale ($t = 0$ ns). Les simulations de DM 1, 2, 3, 4 et 5 sont représentées par les courbes noire, marron foncé, gris clair, beige et gris foncé, respectivement (B) Interprétation par DSSP des structures secondaires de hVKORC1, indiquées en pourcentage de temps, au cours des cinq simulations de DM de 100 ns. Les repliements en hélice α , hélice $_{3,10}$, boucle, coude et brin β sont indiqués par les courbes colorées en rouge, orange, cyan, bleu foncé et vert clair, respectivement. Les structures secondaires de la conformation initiale de hVKORC1 ($t=0$ ns) sont indiquées par le schéma bleu, au-dessus des graphes.

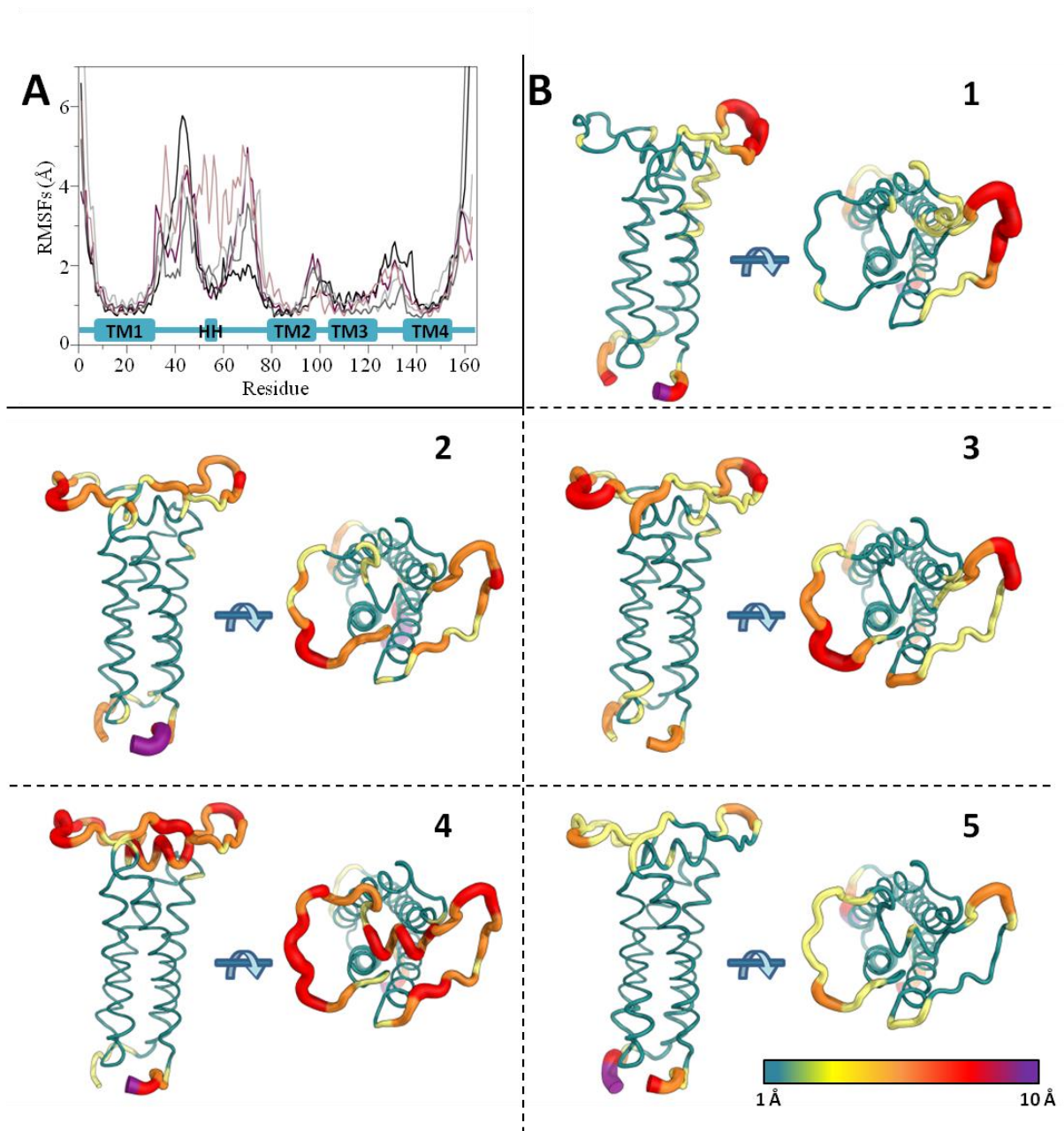


Figure 3.13. RMSFs des résidus du modèle 3D de hVKORC1 simulé et représentation de ces valeurs sur la structure du modèle. (A) Courbes des RMSFs du modèle 3D de hVKORC1 au cours de cinq simulations de DM de 100 ns chacune. Les RMSFs ont été calculées sur les 163 atomes du carbone α de la protéine, préalablement alignés sur les coordonnées de la conformation initiale ($t = 0$ ns). Les structures secondaires de la conformation initiale de hVKORC1 ($t = 0$ ns) sont indiquées par le schéma bleu situé sous la courbe des RMSFs. (B) Projection des valeurs de RMSFs au cours des cinq simulations de DM de 100 ns chacune sur la conformation initiale de hVKORC1, représentée en *tube*. Les valeurs de RMSFs croissantes sont codées par l'épaisseur de *tube* et par un dégradé de couleur, du bleu (1 Å) au violet (10 Å).

L'ensemble de ces résultats indique une bonne préservation du repliement global du modèle 3D hVKORC1_SH au cours des simulations de DM. Les hélices TMs sont très stables, l'architecture de ce domaine transmembranaire est stabilisée par des interactions hydrophobes, entre résidus de leucine notamment. Leur population est en effet très importante au niveau des TMs, particulièrement dans la partie inférieure où ils forment une "couronne" liant les quatre hélices entre elles (**Figure 3.14**).

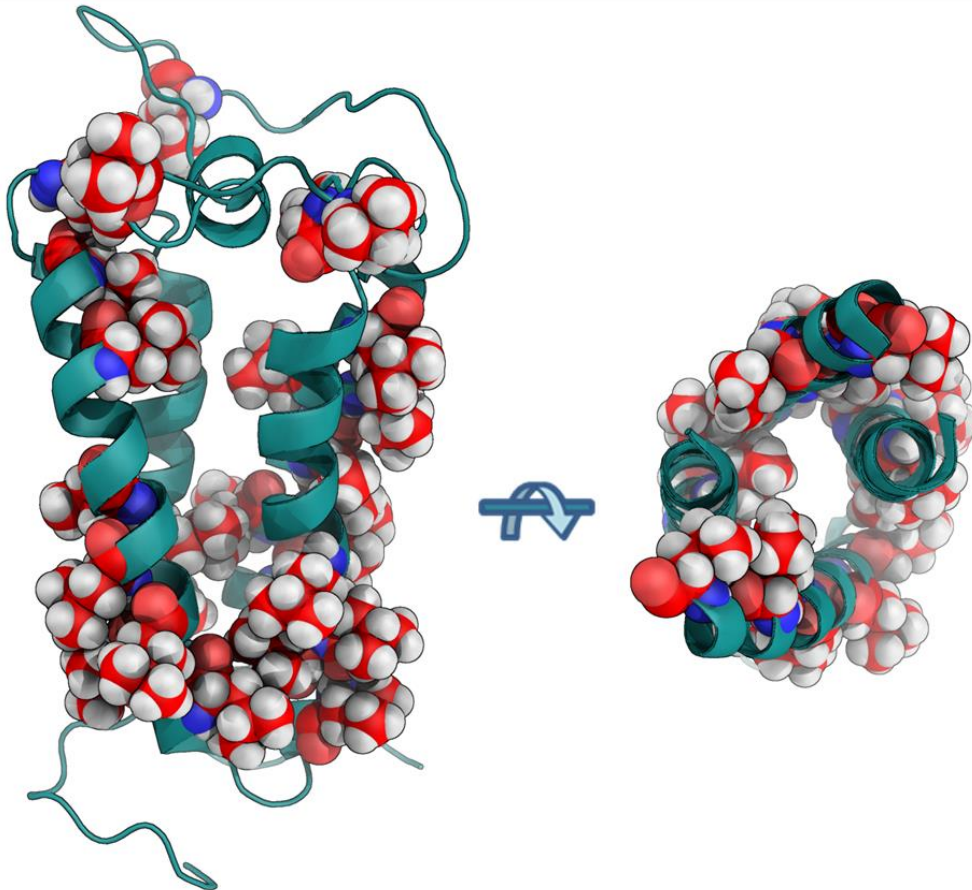


Figure 3.14. Localisation des résidus de leucine au niveau des hélices transmembranaires de hVKORC1. Les résidus de leucine sont représentés en sphères et colorés en rouge. Les atomes d'azote, d'oxygène et d'hydrogène sont représentés en bleu, rouge clair et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La boucle L n'est pas représentée sur la Figure de droite, pour plus de clarté.

La boucle L présente quant à elle des fluctuations importantes durant les simulations de DM. Ce domaine, en *coil*, semble être à l'origine de la variabilité conformationnelle du modèle de hVKORC1.

Nous avons donc étudié plus en détail le comportement dynamique de la boucle L, en utilisant l'analyse en composantes principales (ACP) et une méthode de *clustering* sur les conformations de hVKORC1 obtenues au cours des simulations de DM (**METHODES**).

Les dix premiers vecteurs propres de l'ACP décrivent entre 80 et 90% des mouvements collectifs (ou lents) du modèle de hVKORC1. Les atomes de hVKORC1 présentant ces mouvements de basse fréquence au cours des simulations de DM sont localisés au niveau de la boucle L (**Figure 3.15**). Les extrémités N- et C-terminales sont également très mobiles. La direction privilégiée suivie par les atomes le long des deux premiers vecteurs propres décrit un mouvement global de la boucle L permettant un rapprochement réciproque des deux demi-boucles. Cependant, la direction et l'amplitude de ce mouvement varient fortement d'un réplica à l'autre. Le regroupement des conformations similaires de hVKORC1 (*clustering*, basé sur les RMSDs calculées deux à deux entre chaque conformation d'une trajectoire, voir **METHODES**) révèle quant à lui la présence de conformations représentatives adoptées par le modèle de hVKORC1 au cours des simulations de DM (**Figure 3.16**). Le nombre élevé de *clusters* de conformations divergents et faiblement peuplés (entre 6 et 9 selon les réplicas, représentant chacun moins de 10% des conformations totales de chaque trajectoire) durant les 20 premières nanosecondes de simulation correspond au processus d'équilibration de la structure de hVKORC1. Les simulations 2, 3 et 5 montrent chacune la présence d'un seul *cluster* (représentant de 45% à 89% des conformations totales) entre 60 et 100 ns. Cette observation suggère la présence d'états métastables de hVKORC1 durant chacune des trois simulations. En revanche, trois *clusters* distincts sont identifiés entre 60 et 100 ns pour les réplicas 1 et 4.

Cette diversité de conformations met en évidence l'absence de convergence du modèle de hVKORC1 vers une conformation unique au cours des simulations de DM. De plus, l'application du *clustering* aux cinq simulations concaténées montre que les conformations représentatives identifiées sont spécifiques à chaque réplica, aucune n'est commune à plusieurs simulations (**Figure 3.16**). La flexibilité importante observée au niveau de la boucle L est donc à l'origine d'une grande variabilité de conformations. Les fluctuations atomiques des résidus concernés semblent aléatoires, nous avons donc cherché à décrire le mouvement de la boucle au cours des simulations de DM. Nous avons pour cela représenté la boucle par des points géométriques (*centroïdes*, **Figure 3.17**), puis avons analysé le motif défini par ces métriques au cours des simulations de DM.

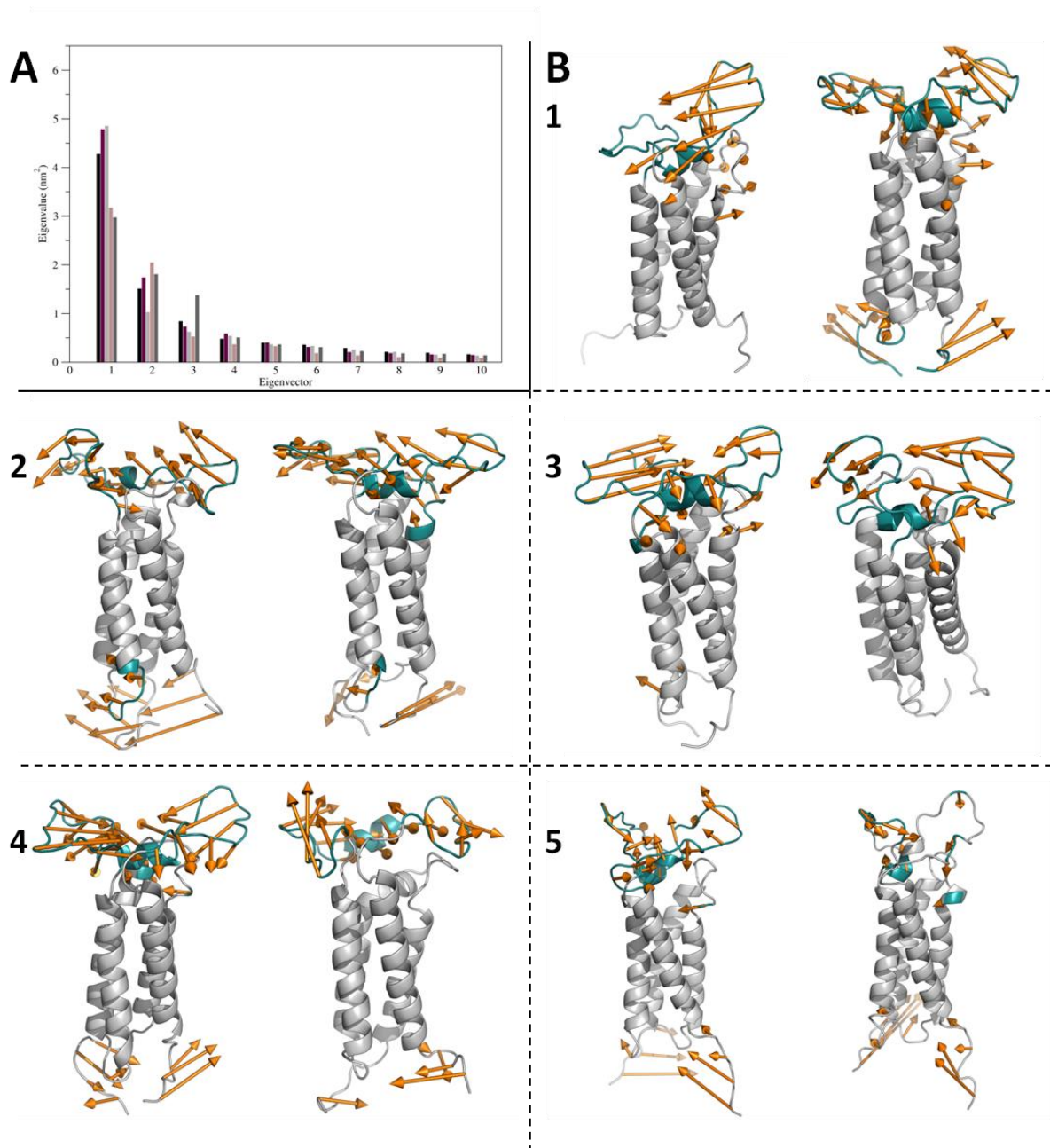


Figure 3.15. Analyse en composantes principales (ACP) de cinq simulations de DM de 100 ns de hVKORC1. L'ACP a été réalisé sur les 163 atomes de carbone α du modèle de hVKORC1. **(A)** Valeurs propres des dix premiers vecteurs propres de l'ACP. Les simulations de DM 1-5 sont représentées par les histogrammes colorés en noir, marron, gris clair, beige et gris foncé, respectivement. **(B)** Représentation des mouvements décrits par le modèle structural de hVKORC1 le long des deux premiers vecteurs propres de l'ACP, au cours de cinq simulations (1-5) de DM de 100 ns. Les flèches colorées en orange représentent la direction et l'amplitude des mouvements décrits. Le modèle 3D de hVKORC1 est représenté en *cartoon*, les régions les plus mobiles sont colorées en bleu.

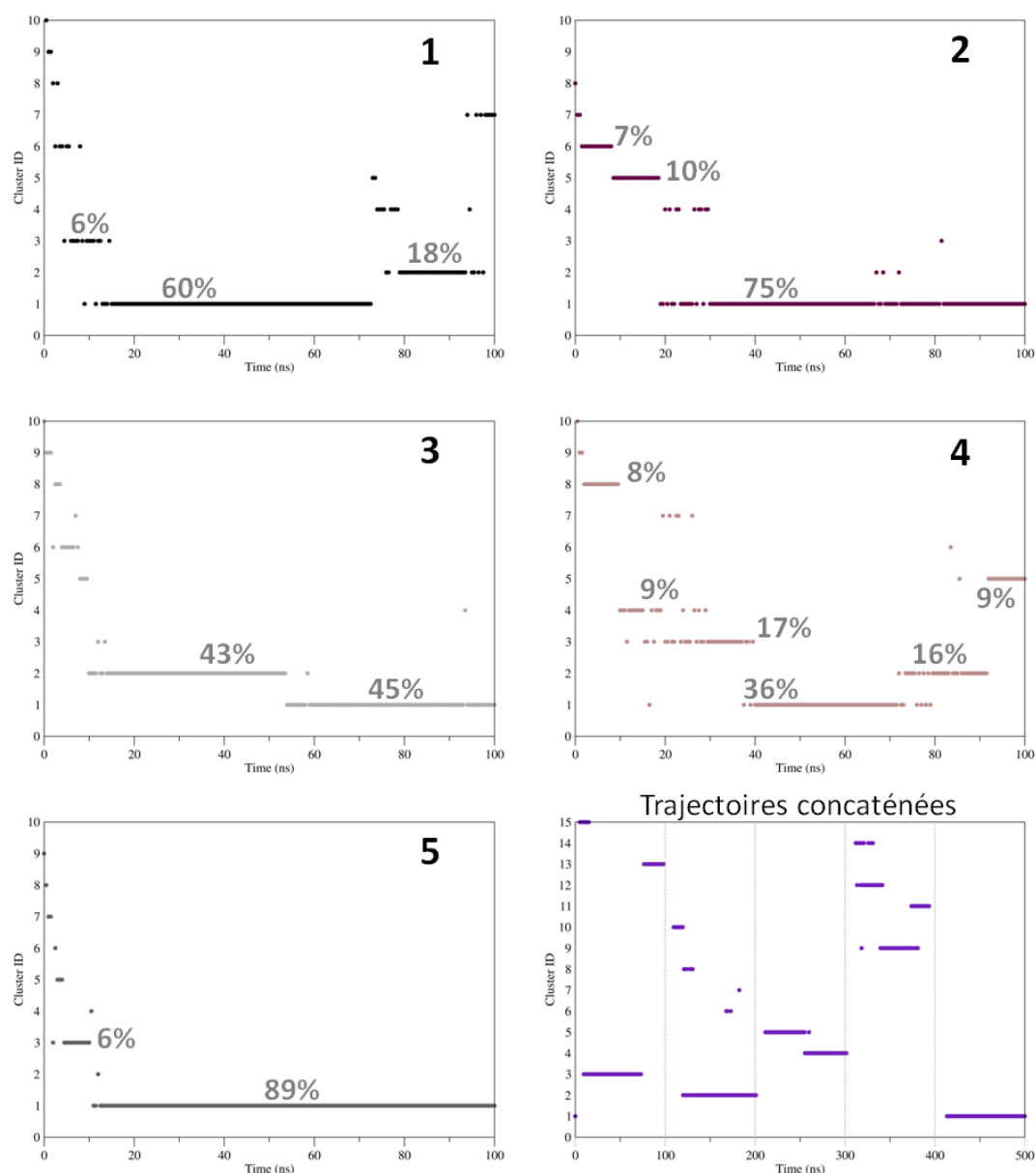


Figure 3.16. Analyse des conformations similaires (*clustering*) de hVKORC1 obtenues au cours de cinq simulations de dynamique moléculaire de 100 ns. Les conformations appartenant à un même *cluster* sont indiquées par des points, regroupés en lignes de niveau. La population de chaque *cluster* (proportion des conformations membres d'un même cluster par rapport au nombre de conformations total de la trajectoire) est indiquée en pourcentage, pour les *clusters* représentant plus de 5% des conformations totales. Les données de simulations 1, 2, 3, 4 et 5 sont représentées en couleur noire, marron, gris clair, beige et gris foncé, respectivement. Le graphique situé en bas à droite représente le *clustering* effectué sur les cinq trajectoires concaténées. Les lignes pointillées verticales délimitent chaque simulation de 100 ns. Le *clustering* est basé sur les RMSDs calculées deux à deux entre les atomes de carbone α des résidus 9-151 de chaque conformation d'une trajectoire. Deux conformations présentant une valeur de RMSD inférieure à 2 Å sont considérées comme membres d'un même *cluster*. Le *cluster* numéro 1 est toujours le plus peuplé, et chaque conformation appartient à un seul *cluster*.

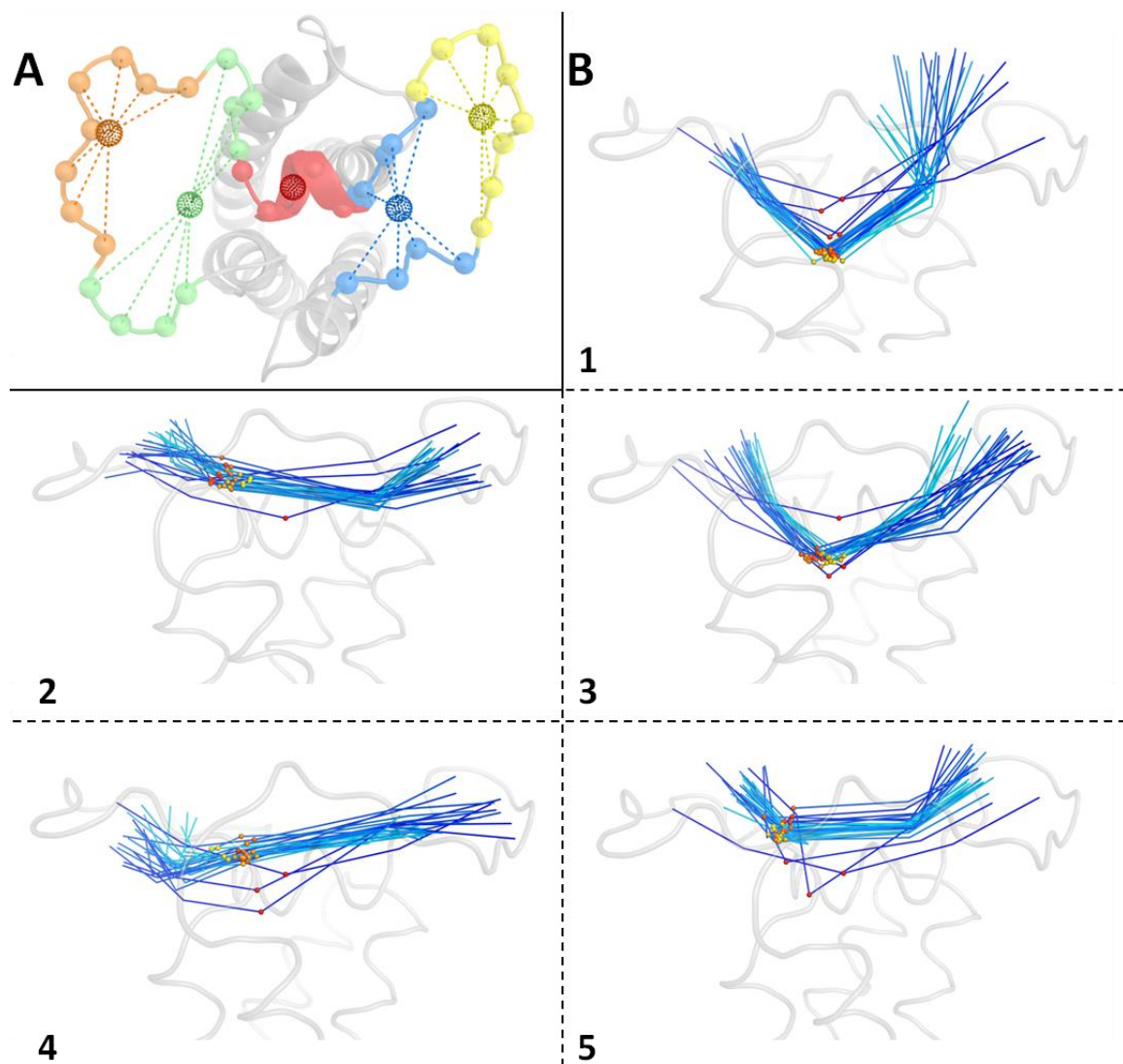


Figure 3.17. Analyse de la flexibilité de la boucle L de hVKORC1 au cours de cinq simulations de DM de 100 ns par application de métriques. (A) Les centres géométriques (*centroïdes*) sont définis à partir des coordonnées des atomes de carbone α des résidus concernés. Les sphères représentées en pointillés correspondent aux centres géométriques des résidus 62-69 (orange) ; des résidus 58-61 et 70-73 (verts); des résidus 52-57 de l'hélice HH (rouge) ; des résidus 36-39 et 48-51 (bleus) ; des résidus 40-47 (jaunes). Les atomes de carbone α des résidus composant la boucle L et l'hélice HH sont représentés par des sphères transparentes. Le modèle 3D de hVKORC1 est représenté en *cartoon*, les fragments de la boucle L sont colorés en fonction des centroïdes qu'ils définissent. (B) Projection dans le plan x-z des centroïdes définis sur la boucle L au cours de cinq simulations de DM de 100 ns. Les cinq centroïdes sont reliés par une ligne afin de représenter schématiquement les mouvements de la boucle. Le dégradé de couleurs correspond au temps de simulation, du bleu foncé ($t = 0$ ns) au cyan ($t = 100$ ns). Le centroïde de la petite hélice HH est indiqué par une sphère, colorée du rouge ($t = 0$ ns) au jaune ($t = 100$ ns). Le modèle 3D de hVKORC1 est représenté en *tube* gris.

Le suivi des métriques définies sur la boucle L au cours des simulations mettent en évidence deux comportements dynamiques distincts du modèle 3D de hVKORC1_SH. Au cours des simulations de DM 1 et 3, les extrémités distales de la boucle L se surélèvent par rapport à leur position initiale et se rapprochent l'une de l'autre. Dans cette conformation, l'hélice HH se rapproche des quatre TMs, et donc de la surface de la membrane. En revanche, les simulations 2, 4 et 5 révèlent un éloignement de l'hélice HH par rapport aux quatre TMs. Les extrémités distales de la boucle L fluctuent moins que lors des répliques 1 et 3, la boucle L conserve une conformation globale plus plane. Cette description a permis d'identifier deux conformations distinctes de la boucle L, en dépit des mouvements importantes de cette région de hVKORC1_SH.

La variation des inter-distances entre les atomes de soufre (S...S) des quatre résidus C₄₃, C₅₁ (tous deux localisés au niveau de la boucle L), C₁₃₂ et C₁₃₅ a ensuite été étudiée. Dans le modèle initial (t = 0 ns), les trois atomes de soufre des résidus C₅₁, C₁₃₂ et C₁₃₅ sont localisés proximité les un des autres (distances entre 4 et 7 Å), formant la base d'une pyramide. En revanche, l'atome de soufre du résidu C₄₃ est éloigné de cette pyramide d'une distance de 20 Å. Ces distances varient beaucoup au cours des simulations de DM, entre 4 et 24 Å. L'observation de ces inter-distances a mis en évidence des rapprochements privilégiés entre résidus de cystéines au cours des simulations de DM (**Figure 3.18**). En effet, certains résidus de cystéine s'apparient dans plusieurs simulations. C'est le cas des résidus C₅₁ et C₁₃₂ (S...S distants de 4 Å dans les simulations 1, 3 et 4), C₁₃₂ et C₁₃₅ (S...S distants de 6 Å dans les simulations 2, 3 et 5) et C₄₃ et C₅₁ (S...S distants de 7 Å dans les simulations 2 et 4). Les distances S...S des résidus C₅₁, C₁₃₂ et C₁₃₅ indiquent leur proximité réciproque durant la simulation 3. En revanche, certains rapprochements entre résidus de cystéine ne sont jamais observés, notamment pour les paires C₄₃-C₁₃₅ et C₄₃-C₁₃₂ (pour laquelle un rapprochement est observé durant la simulation 1, mais durant moins de 5 ns, donc considéré comme peu favorable). Trois paires de résidus de cystéine semblent s'associer préférentiellement, selon les distances S...S : C₄₃-C₅₁, C₅₁-C₁₃₂ et C₁₃₂-C₁₃₅, pouvant donc potentiellement former des ponts disulfures (S-S).

Cependant, le modèle de hVKORC1 ne semblait pas totalement équilibré durant deux simulations de DM parmi les 5 réalisées. De plus, la structure de la protéine VKORC1_SH adopte une grande variété de conformations. Nous avons donc choisi de prolonger les simulations 1 et 2 jusqu'à 1 µs, car elles présentent deux conformations différentes de la boucle L, reflétant une grande diversité conformationnelle de VKORC1_SH.

En effet, le modèle 3D de hVKORC1 semble avoir atteint un état d'équilibre au cours de la simulation 2, mais pas durant la simulation 1. La conformation de la boucle L ainsi que les résidus de cystéines appariés diffèrent également entre les deux répliques.

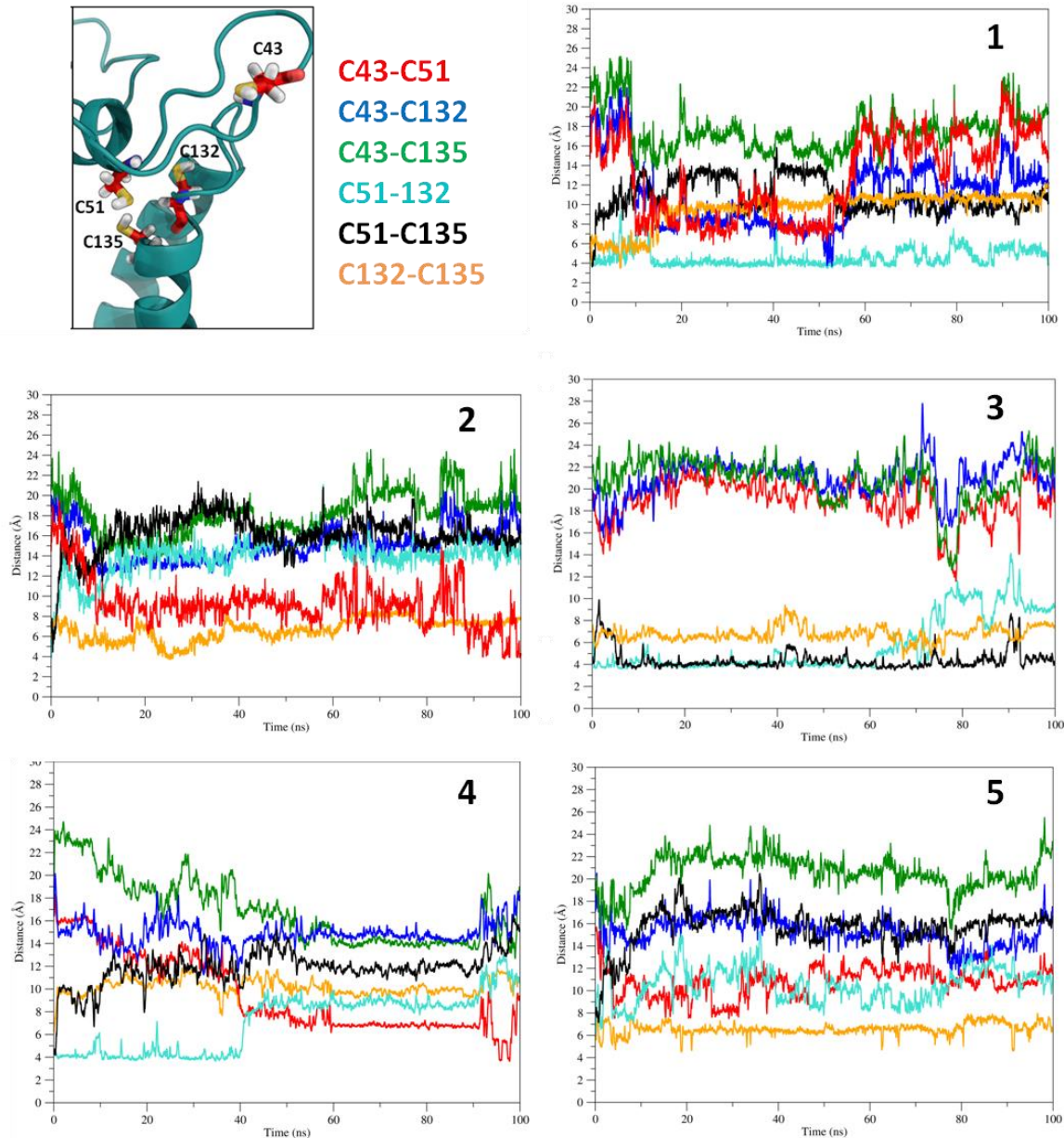


Figure 3.18. Etude des distances S...S entre les atomes de soufre des quatre résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ au cours de cinq simulations de DM de 100 ns. L'encart en haut à gauche présente la localisation des quatre résidus de cystéine (représentés en *sticks* rouges, les atomes de soufre, d'azote, d'oxygène et d'hydrogène étant colorés en jaune, bleu, rouge clair et blanc, respectivement) sur le modèle structural de hVKORC1 (représenté en *cartoon* bleu). Les distances S...S entre atomes de soufre des résidus de cystéine au cours des simulations de DM sont présentées dans les cinq graphes ; chaque paire de résidus de cystéine est représentée par une courbe dont la couleur est indiquée à côté de l'encart.

2. Recherche d'états métastables du modèle 3D de hVKORC1 par simulations de dynamique moléculaire d'1 μ s

La stabilité structurale du modèle 3D de hVKORC1 a été évaluée par le calcul des valeurs de RMSDs et RMSFs, ainsi que par l'interprétation des structures secondaires de la protéine par DSSP. Les courbes des RMSDs et RMSFs indiquent de faibles fluctuations atomiques des résidus localisés au niveau des hélices transmembranaires (RMSDs aux alentours de 2 Å, RMSFs comprises entre 1 et 2 Å), tandis que la boucle L présente les déviations les plus importantes (RMSDs entre 6 et 8 Å, RMSFs atteignant 4 Å pour les résidus les plus fluctuants) (**Figure 3.19.A-B**). Cette observation confirme la grande flexibilité de la boucle L par rapport au domaine membranaire de hVKORC1, déjà constatée durant les simulations de 100 ns. Néanmoins, les courbes des RMSDs montrent une stabilisation globale du modèle (y compris au niveau de la boucle L) après 500 ns, au cours des deux simulations de DM. Le repliement des hélices transmembranaires, à l'image des simulations de 100 ns, est hautement préservé au cours des simulations de DM d'1 μ s, comme le révèle l'assignement des structures secondaires par DSSP (**Figure 3.19.C**). Le repliement des résidus 51-56 formant l'hélice HH alterne principalement entre hélice α et hélice₃₁₀.

Nous avons défini de nouvelles coordonnées, les *centroïdes*, et avons suivi et leurs déplacements au cours des simulations de DM pour caractériser le comportement dynamique des hélices TMs et HH. Chaque hélice TM montre un déplacement significatif (5 Å) au cours des simulations (**Figure 3.20**). Ce déplacement n'est pas équivalent pour les deux extrémités de chaque hélice, en termes de valeur de déplacement (plus large pour le haut des hélices), et de direction pour certaines hélices (TM1, TM2 et TM4), produisant un effet d'enroulement du domaine membranaire. De plus, malgré la différence d'amplitude et de direction du déplacement des différentes hélices TMs, la tendance générale observée est une diminution des fluctuations des axes des *centroïdes* correspondant aux extrémités des quatre TMs (leur amplitude passe de 5 Å à 2 Å, **Figure 3.20.B**) au cours des simulations de DM. Une comparaison des deux simulations indique un déplacement des hélices TM différent. Les fluctuations de *centroïde* d'hélice HH passent de 8 Å à 2 Å durant les deux simulations de DM. Cependant, elle se stabilise dans une position plus éloignée des quatre TMs durant la simulation 2.

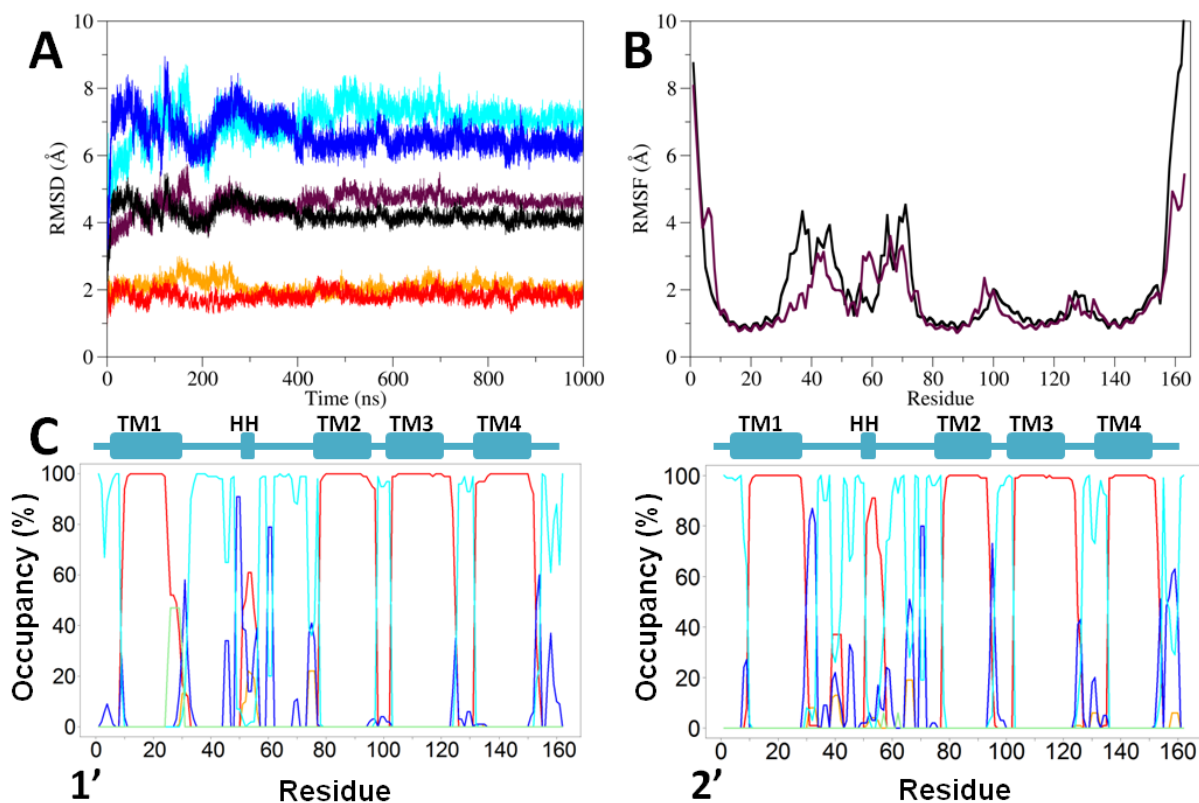


Figure 3.19. Caractérisation de la stabilité structurale globale de hVKORC1 au cours de simulations de DM d'1 μ s. (A) Courbes des RMSDs du modèle 3D de hVKORC1. Les courbes noire et marron correspondent aux RMSDs calculées sur les atomes de carbone α des résidus 9 à 151 (les résidus des extrémités N- et C-terminales ont été omis afin d'éliminer le bruit induit par leurs grandes fluctuations) au cours des deux simulations de DM. Les courbes en bleu et en cyan correspondent aux RMSDs calculées sur les atomes de carbone α des résidus composant la boucle L (résidus 29-76) au cours des deux simulations de DM. Les courbes en rouge et orange correspondent aux RMSDs calculées sur les atomes de carbone α des résidus composant les quatre TMs (résidus 9-28, 77-93, 102-123 et 132-151) au cours des deux simulations de DM. Les RMSDs ont été calculées après alignement des atomes de carbone α des hélices transmembranaires de chaque conformation sur les coordonnées de la conformation initiale ($t = 0$ ns). (B) Courbes des RMSFs du modèle 3D de hVKORC1 au cours de deux simulations de DM d'1 μ s. Les structures secondaires de la conformation initiale de hVKORC1 ($t = 0$ ns) sont indiquées sous le graphe. Les RMSFs des résidus de hVKORC1 ont été calculées sur les 163 atomes de carbone α de la protéine, préalablement alignés sur les coordonnées de la conformation initiale ($t = 0$ ns). (C) Assignement par DSSP des structures secondaires des conformations de hVKORC1 au cours de deux simulations de DM d'1 μ s (en pourcentage). Les repliements en hélice α , hélice $_{310}$, boucle, coude et brin β sont indiqués par les courbes en rouge, orange, cyan, bleu foncé et vert clair, respectivement. Les structures secondaires de la conformation initiale de hVKORC1 ($t = 0$ ns) sont indiquées au-dessus des graphes.

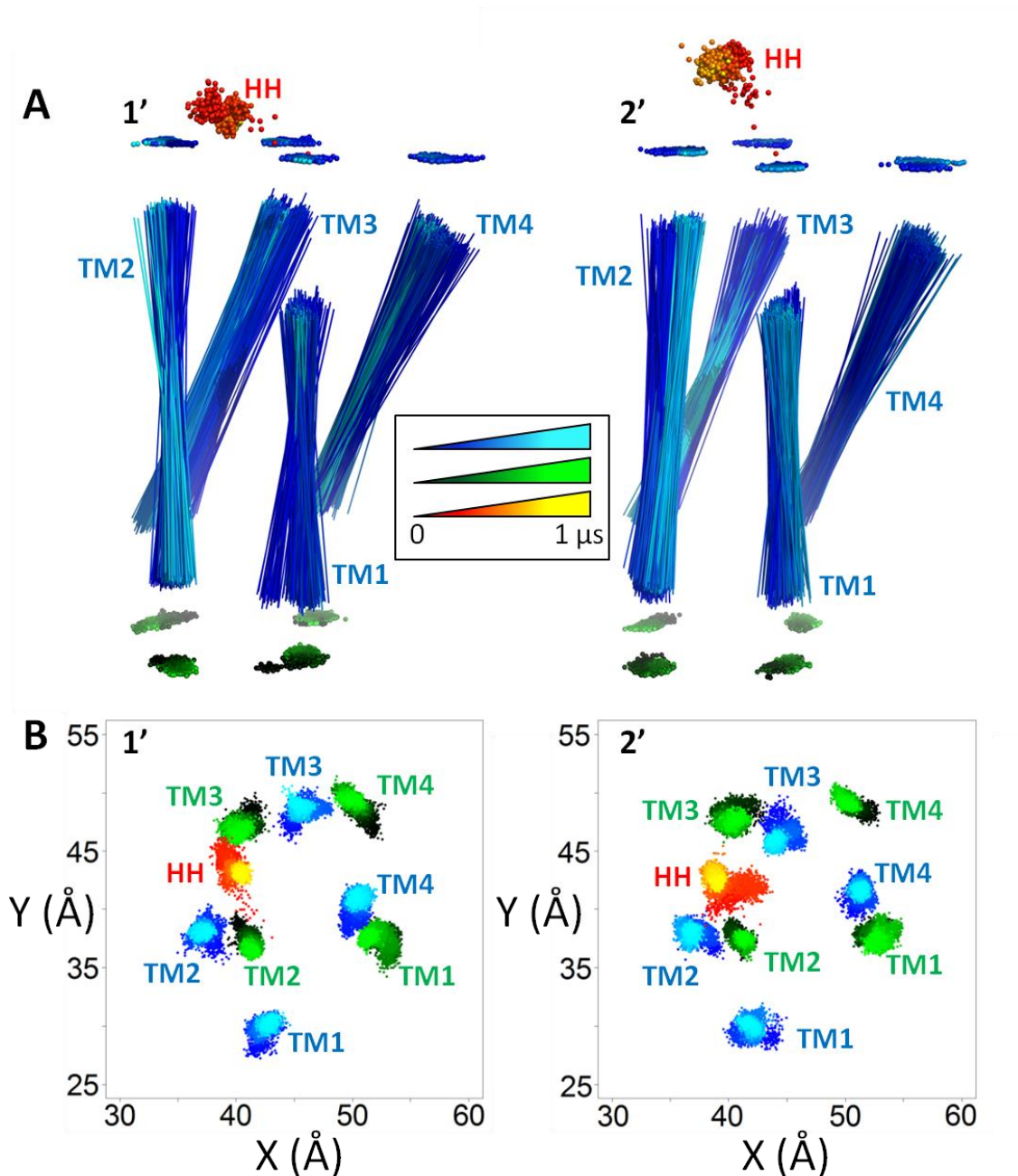


Figure 3.20. Caractérisation de la déviation des hélices TMs et HH de hVKORC1 au cours de deux simulations de DM d'1 μ s. (A) Les *centroïdes* de chaque extrémité des quatre TMs sont définis à partir des coordonnées des atomes de carbone α des quatre résidus formant le dernier tour de l'hélice (résidus 9-12 et 25-28 pour TM1, 77-80 et 90-93 pour TM2, 102-105 et 120-123 pour TM3, 132-135 et 148-151 pour TM4). Les deux *centroïdes* ainsi déterminés pour chaque TM sont reliés par une ligne représentant l'axe de la TM concernée. Un seul *centroïde* représente l'hélice HH, défini à partir des coordonnées des atomes de carbone α des résidus 52-57. Les axes des TMs et le *centroïde* de HH (lignes colorées du bleu foncé au cyan et sphères colorées du rouge au jaune, respectivement) sont représentés dans l'espace 3D au cours de deux simulations d'1 μ s chacune, à intervalles de temps d'1 ns. (B) Les coordonnées des *centroïdes* définissant HH (points colorés du rouge au jaune) et les extrémités supérieures (points colorés en bleu) et inférieures (points colorés en vert) des quatre TMs au cours des simulations de DM sont projetés dans le plan x-y. Le code couleur correspondant au temps de simulation MD est indiqué dans l'encart.

Nous avons ensuite étudié la diversité conformationnelle de la boucle L en appliquant les mêmes métriques (*centroïdes*) que celles utilisées pour analyser les simulations de 100 ns (**Figure 3.17**). Les données de chaque simulation sont groupées en un seul *cluster* à la fin des simulations, montrant une stabilisation de la boucle L en forme de plateau pour la simulation 1 et en forme de W pour la simulation 2 (**Figure 3.21.A**). Puis nous avons suivi la variation des distances S...S entre les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ au cours des deux simulations de DM d'1 μ s. Dans la structure initiale ($t = 0$ ns), les trois atomes de soufre des résidus C₅₁, C₁₃₂ et C₁₃₅ sont séparés par de courtes distances (4-6 Å). Le quatrième atome de soufre (résidu C₄₃) est éloigné de ce *cluster*, distant de 20 Å. Comme nous l'avons remarqué lors des simulations de 100 ns, le positionnement des atomes de soufre varie très fortement au cours d'une même simulation, produisant de grandes variations des distances suivies. Leurs valeurs sont réparties entre 4 et 24 Å pour les deux simulations, mais leur profil et leur comportement diffèrent largement entre les deux simulations.

Les résidus C₅₁ et C₁₃₂ sont les plus proches durant la simulation 1 (distance S...S comprise entre 4 et 8 Å), tandis que les résidus C₄₃ et C₅₁ d'une part, et C₁₃₂ et C₁₃₅ d'autre part, semblent s'associer préférentiellement au cours de la simulation 2 (distances S...S comprises entre 4 et 8 Å) (**Figure 3.21.B**). Les courtes distances S...S entre résidus de cystéine, ainsi que la stabilité de ces distances au cours des simulations de DM, suggèrent la stabilisation des conformations favorisant la formation de ponts disulfures entre résidus de cystéine.

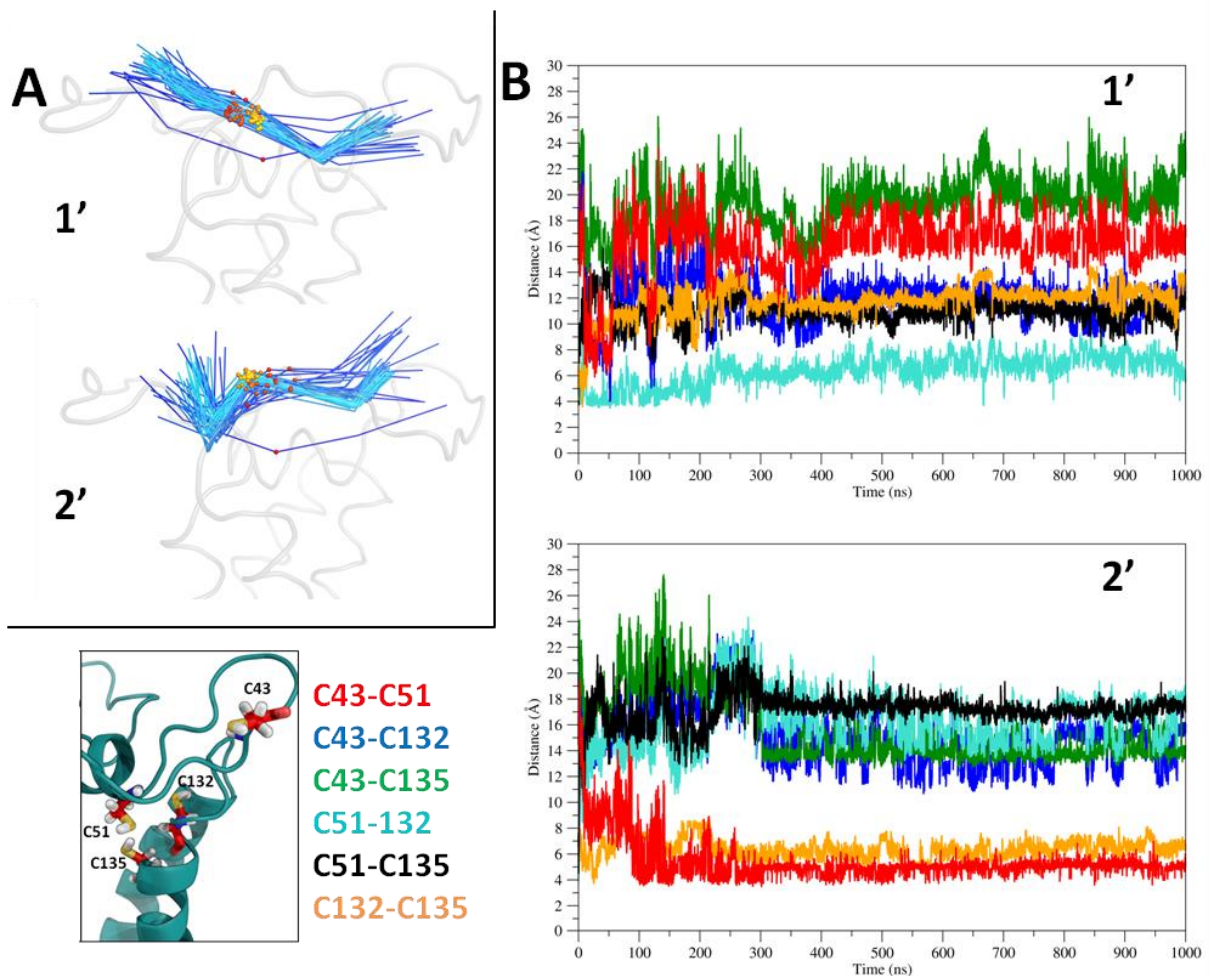


Figure 3.21. Etude des distances S...S entre les atomes de soufre des quatre résidus C43, C51, C132 et C135 au cours de deux simulations de DM de 1 μ s. (A) Les *centroïdes* définis sur la boucle L sont déterminés à partir des coordonnées des atomes de carbone α de différents fragments : résidus 62-69 ; résidus 58-61 et 70-73 ; résidus 36-39 et 48-51 ; résidus 40-47. Le *centroïde* représentant l'hélice HH est déterminé à partir des coordonnées des atomes de carbone α des résidus 52-57. Ces cinq *centroïdes* ont ensuite été projetés dans le plan x-z et reliés par des lignes. Le dégradé de couleurs correspond au temps de simulation, du bleu foncé (t = 0 ns) au cyan (t = 1 μ s). Le *centroïde* de la petite hélice HH est indiqué par une sphère colorée du rouge (t = 0 ns) au jaune (t = 1 μ s). Le modèle 3D de hVKORC1 est représenté en *tube* gris. (B) Variations des distances S...S entre atomes de soufre des résidus de cystéine au cours de deux simulations de DM d' 1 μ s. Chaque distance est distinguée par une courbe dont la couleur est indiqué à côté de l'encart, présentant la localisation des quatre résidus de cystéine (représentés en *sticks* rouges, les atomes de soufre, d'azote, d'oxygène et d'hydrogène étant colorés en jaune, bleu, rouge clair et blanc, respectivement) sur le modèle structural de hVKORC1 (représenté en *cartoon* bleu).

RESUME: les méthodes classiques d'analyse des données (RMSDs, RMSFs, ACP, *clustering* et paramètres structuraux) de dynamique moléculaire du modèle de VKORC1 ont mis en évidence des caractéristiques structurales et dynamique associables avec les fonctions de la protéine. En particulier, la capacité de la boucle luminale à effectuer des mouvements amples reflète sa plasticité conformationnelle, pouvant constituer un site de liaison pour un partenaire protéique (de type thioredoxine par exemple [29]). L'hypothèse d'un site de liaison est renforcée par la présence de nombreux résidus hydrophobes la boucle (60% des résidus qui la composent, **Figure 3.22**). Le mouvement corrélé des hélices transmembranaires et de la boucle luminale indique une adaptabilité conformationnelle globale de hVKORC1. Cette adaptabilité permet la formation ou la rupture de liaisons covalentes entre paires d'atomes de soufre, provoquant une réorganisation structurale et conformationnelle. L'ensemble de ces réarrangements aboutit à la formation d'états enzymatiques transitoires de hVKORC1. Les variations des distances S...S entre atomes de soufre de résidus de cystéine indiquent également une grande variabilité conformationnelle associée aux états multiples de l'enzyme. Les conformations dont les résidus de cystéine sont prédisposés à l'établissement de liaisons covalentes S...S sont considérées comme des conformations candidates des états intermédiaires de hVKORC1.



Figure 3.22. Localisation des résidus hydrophobes au niveau de la boucle L de hVKORC1. Les résidus hydrophobes de la boucle L sont colorés en rouge, les résidus non hydrophobes en bleu, sur la structure 3D de hVKORC1 et sur la séquence peptidique de la boucle L. La position de l'hélice HH est indiquée par le schéma situé au-dessus de la séquence peptidique. La structure 3D de hVKORC1 est représentée en *cartoon*, les hélices TMs sont colorées en gris clair.

3. Analyse des simulations de DM par des techniques mathématiques avancées

Les conformations de hVKORC1 montrent une grande variabilité et divergent entre les différentes simulations de MD, rendant difficile la capture des changements structuraux de la protéine au cours du temps. La qualification de certaines conformations comme réellement métastables et pertinentes physiquement et biologiquement nécessite une caractérisation plus fine, focalisée notamment sur la notion d'énergie libre. La détermination des états métastables nécessite de considérer dans sa globalité le caractère spatio-temporel des conformations obtenues au cours de simulations de DM, non restitué par les méthodes classiques comme celles basées sur les critères de RMSDs ou RMSFs. Les états métastables étant associés par définition à des minima d'énergie, un modèle de la dynamique contenant explicitement un terme d'énergie est nécessaire.

Ce modèle, ajusté à la trajectoire de DM, permettrait alors d'estimer le paysage d'énergie et ainsi de donner des indications pour définir les états métastables. En collaboration avec Alain TROUVE et Bernard CHALMOND (CMLA, ENS Cachan), nous avons développé une nouvelle méthode, *Configurational Scanning* (ConfigScan), basée sur les méthodes d'analyse spectrale.

Contexte théorique. Nous avons effectué une analyse statistique de chaque trajectoire de DM observée $\Xi = x_1, \dots, x_N$ en utilisant une estimation de la densité de conformations dans les espaces de faible dimension par le premier mode issu de l'ACP. La densité d'occupation locale $\pi(x)$ à une échelle donnée σ est estimée par un estimateur de densité de Parzen Window[104] avec un noyau gaussien :

$$\pi(x) \propto \sum_{\{i=1\}}^N \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

induisant le paysage d'énergie :

$$U(x) = -\log(\pi(x)) + C$$

Cette énergie ne doit pas être associée au potentiel (champ de force) qui anime la simulation de DM, mais considérée comme un proxy statistique pour analyser les régions de forte occupation liées aux états métastables potentiels visités durant la dynamique.

La présentation de la matrice de similarité entre conformations est de la forme suivante :

$$W_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

avec $(x_i) \propto W_{i,\cdot} \doteq \sum_{\{j=1\}}^N W_{i,j}$. Cependant, au-delà de la densité d'occupation, certains des aspects dynamiques peuvent être analysés empiriquement par l'estimation d'une marche aléatoire sur les données simulées, compatible avec la densité d'occupation. Considérons une marche aléatoire $n \rightarrow X_n$ sur les conformations MD telle que la probabilité de transition p_{ij} entre i et j soit donnée par :

$$P_{i,j} = P(X_{n+1} = j | X_n = i) \propto \frac{W_{i,j}^\alpha}{W_{i,\cdot}^\alpha W_{j,\cdot}^\alpha}$$

avec $\alpha = 1/2$. Cette marche aléatoire se rapproche, après le changement de temps $t_n = n\sigma^2$, d'une diffusion continue $t \rightarrow X_t$ suivant une diffusion de Langevin :

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t$$

donnant à l'équilibre une densité d'occupation $\pi(x)$ [105-107]. L'analyse spectrale du laplacien associé $L_{ij} = \delta_{ij} - P_{ij}$ donne des variables lentes intéressantes pour la dynamique estimée, fortement liée aux états métastables.

Traitement de données des simulations

Matrice de similarité. La matrice W , calculée sur les conformations brutes ou après réduction par ACP, produit un motif structuré en blocs, chacun d'eux étant constitué d'un ensemble de conformations proches (**Figure 3.23.A**). Les différents blocs sont vraisemblablement associés à des états métastables ou intermédiaires. Les blocs montrent des profils différents sur les trajectoires analysées qui s'agrègent (après 200-300 ns) au niveau de blocs dans la trajectoire 2' et s'arrangent sous forme de rosette en 1', indiquant potentiellement des conformations revisitées. La réduction de dimension par ACP ne dégrade pas le profil obtenu, mais produit un contenu plus contrasté. Cette réduction par ACP est très utile pour l'analyse des données, car elle réduit les calculs et améliore l'affichage. Chaque vecteur x_i est remplacé par un vecteur ξ_i de très petite dimension r réduite par ACP (typiquement, $r = 2$ ou $r = 3$).

La définition du paysage d'énergie est typiquement très grossière sur l'espace de conformations brut. La réduction par ACP permet d'extraire des composantes régulières en omettant celles de hautes fréquences, elle est par conséquent bénéfique pour l'analyse et la localisation des principaux bassins.

Profil d'énergie. U reflète la matrice de similarité, composée de blocs correspondant à des intervalles de temps où l'énergie est relativement stable. Cette propriété est soulignée par les profils des variables lentes, comme indiqué sur la **Figure 3.23.S1** et la **Figure 3.23.S2** pour la trajectoire 2'. Une variable lente est une fonction unidimensionnelle $q(x_i)$ liée à la dynamique de Langevin sous-jacente et donc à W , donnant une vue synthétique et simplifiée des blocs dans W , comme c'est le cas en regroupement spectral [106, 108].

Paysage d'énergie. Pour le profil d'énergie, les évolutions temporelles de l'énergie U (**Figure 3.23.B**) sont tracées sous forme de courbes $t_i \rightarrow U(x_i)$ ou $t_i \rightarrow U(\xi_i)$, selon que l'on considère la réduction de dimensions par ACP ou non. Cependant, il est intéressant de représenter le paysage d'énergie $U(x)$ comme une hypersurface paramétrée par la conformation x . Nous utilisons une réduction 2D pour afficher le paysage d'énergie comme une surface 2D en 3D comme $\xi \rightarrow U(\xi)$, et afficher l'évolution de la dynamique en tant que marche aléatoire sur ce paysage. Pour ce faire, sur la **Figure 3.23.C**, la trace de la trajectoire de la marche aléatoire sur le paysage d'énergie est tracée et révèle une version échantillonnée des principaux bassins visités. Le paysage d'énergie est utile pour définir les emplacements des minima d'énergie, comme indiqué sur la **Figure 3.23.D**. En effet, ces minima sont extraits d'une représentation lisse et continue de la surface échantillonnée, calculée à l'aide d'un modèle spline [109].

En résumé, la représentation stochastique de la trajectoire de DM est un moyen efficace d'analyser le paysage d'énergie et de localiser les bassins d'énergie. L'ensemble des descripteurs (matrice de similarité, variation d'énergie et surface d'énergie, couplées à l'analyse des variables lentes) sont les composantes de notre approche intitulée *Configuration Scanning* (ConfigScan). Appliqué à la trajectoire 2', le paysage d'énergie montre trois minima locaux bien définis et très peuplés, correspondant probablement aux états métastables de hVKORC1_SH. Ces minima sont liés entre eux par des cols décrivant la transition entre les différents états (**Figure 3.23.C**).

La dynamique conformationnelle observée de ce paysage à bassins multiples implique des déplacements de longue période dans un bassin, suivie d'une transition rapide vers un autre. Cet événement est décrit comme un «saut entre minima», déjà observé lors d'études de dynamique moléculaire par d'autres méthodes [110]. Le paysage d'énergie de la trajectoire 1', avec un minimum non significatif en début de simulation, peut être décrit comme une chute en feuille morte atteignant le minimum global.

Les conformations de plus faible énergie correspondant à chaque minimum local ont ensuite été identifiées (**Figure 3.23.D**). Il s'agit de la conformation $t = 963$ ns pour la simulation 1', et des conformations $t = 188$ ns, $t = 581$ ns et $t = 967$ ns pour la simulation 2'. Les structures 3D représentant ces conformations ont été extraites des simulations de DM (**Figure 3.24**). Leur analyse montre une proximité des résidus C_{51} et C_{132} (distance $S \cdots S$ égale à $6,4 \text{ \AA}$) pour l'état métastable atteint durant la simulation 1', suggérant une conformation favorable à la formation d'un pont disulfure entre les deux résidus de cystéine concernés. Un rapprochement des résidus C_{43} et C_{51} (distances $S \cdots S$ comprises entre $4,4 \text{ \AA}$ et $5,2 \text{ \AA}$) et C_{132} et C_{135} (distances $S \cdots S$ comprises entre $5,7 \text{ \AA}$ et $7,7 \text{ \AA}$) est observé durant la simulation 2', suggérant la formation potentielle de deux ponts disulfures dans la conformation obtenue.

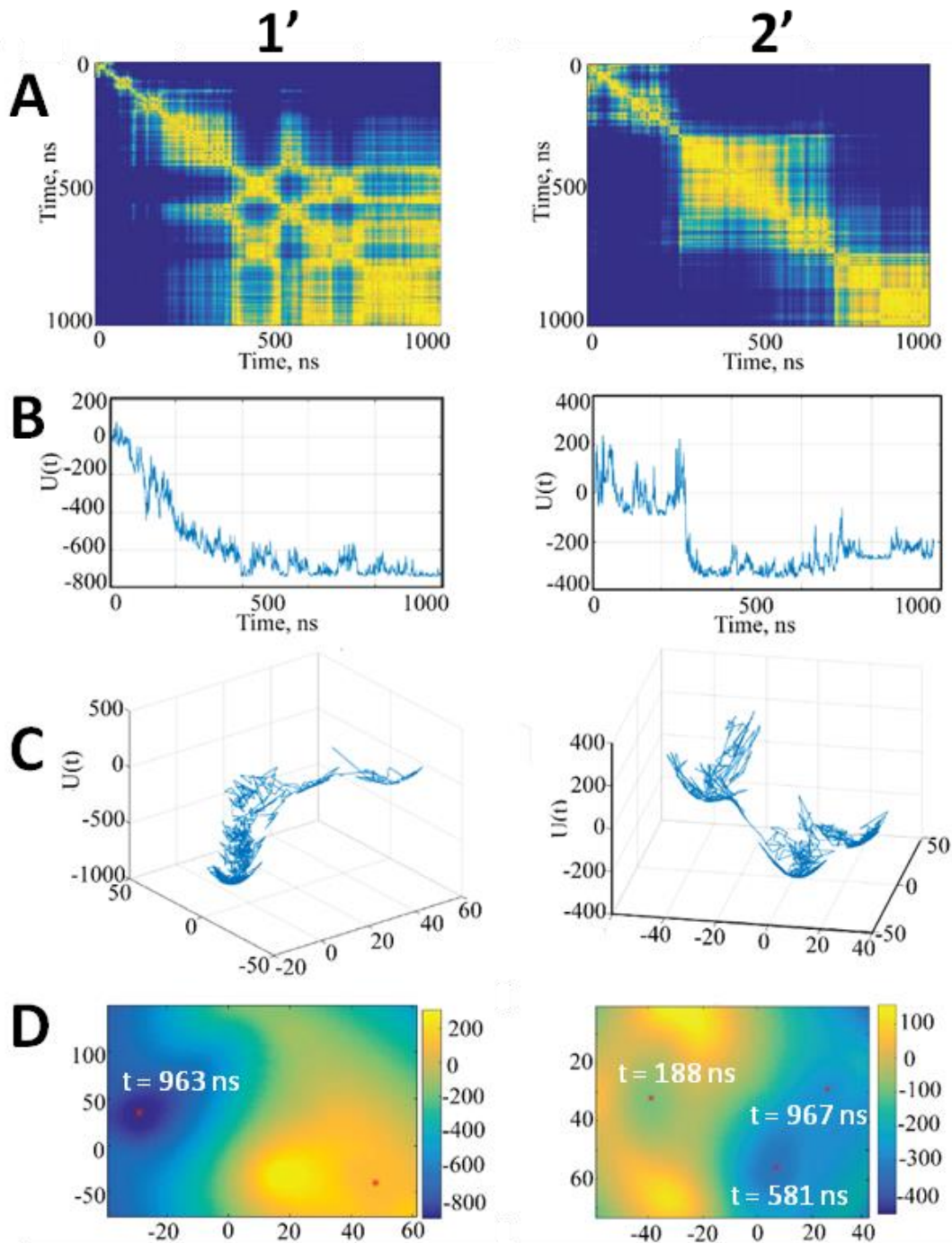


Figure 3.23. Analyse des données de deux simulations de DM d'1 μ s par ConfigScan (A) Matrices de similarités entre les conformations $\{x_i, i = 1, \dots, N\}$ de hVKORC1 obtenues au cours des deux simulations de DM, calculées sur les 163 atomes de carbone α de la protéine, après ACP des trajectoires 1' et 2'. Chaque vecteur x_i a été remplacé par le vecteur ξ_i de dimension 2. La distance entre chaque vecteur est codée par couleur, jaune pour les conformations similaires et bleu pour les éloignées. **(B)** Evolution temporelle de la variation $\delta U(\xi_i) \doteq U(\xi_i) - U(\xi_0)$ par rapport à l'énergie de la conformation initiale à $t = 0$ ns. **(C)** Trace des trajectoires des conformations après réduction 2D par ACP sur la surface d'énergie pour les trajectoires 1' et 2', présentées par rapport aux deux premières composantes de l'ACP. **(D)** Paysages d'énergie obtenus pour les deux trajectoires en représentation de surface spline.

Précisions sur la Méthode

Variables lentes. Une matrice de kernel Markov P est le graphe Laplacien $L = I - P$ normalisé, c'est-à-dire une généralisation du Laplacien standard sur une grille. Ses vecteurs propres ont une position clé dans le regroupement (*clustering*) spectral. Elles peuvent être approximées comme des variables indicatrices de clusters en Ξ , particulièrement lorsque les valeurs propres de P sont proches de 1. Dans ce cas, ces variables sont appelées variables lentes car leurs profils sont utiles pour identifier les bassins d'énergie, le passage de la marche aléatoire d'un bassin à un autre s'opérant à des échelles de temps longue. Les deux premières variables lentes pour la trajectoire 2', avec et sans réduction par ACP, sont représentées respectivement dans la **Figure 3.23.S1** et la **Figure 3.23.S2**. Puisque les variables lentes sont des variables indicatrices de clusters, nous pouvons utiliser cette propriété pour comparer les profils de chaque variable avant et après APC, en recherchant les changements de signe de ces profils.

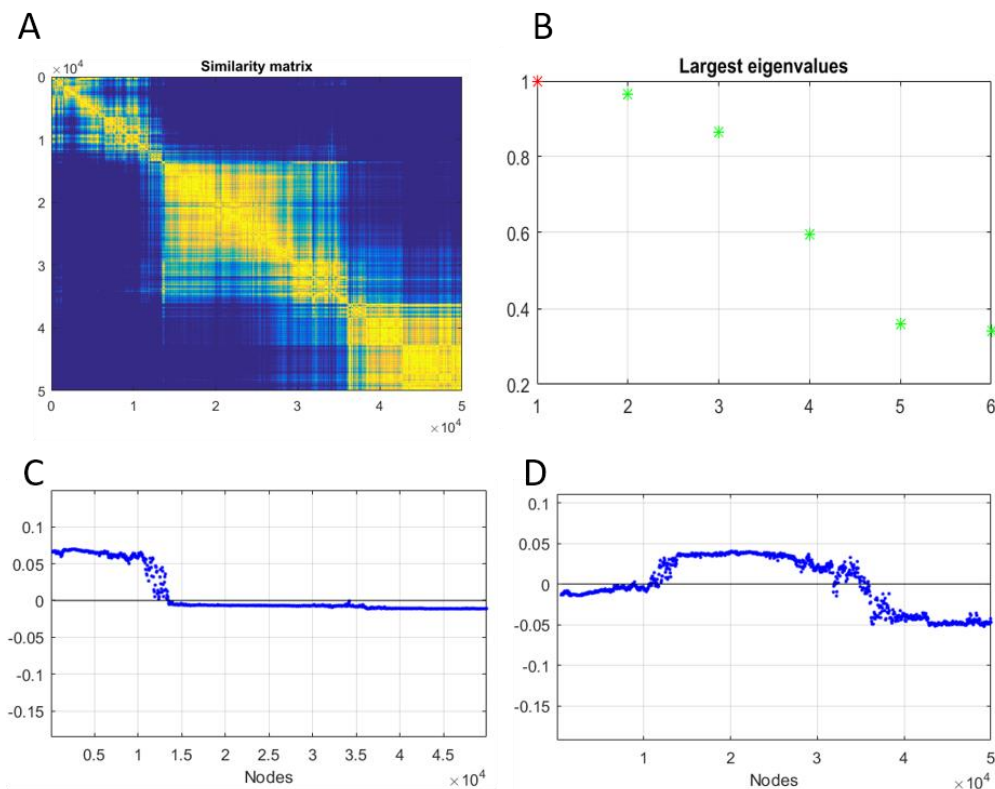


Figure 3.23.S1. Analyse spectrale de la trajectoire 2' après ACP. (A) Matrice de similarité W ; (B) Valeurs propres de A ; (C) Première variable lente; (D) Deuxième variable lente. En recherchant les changements de signe de cette variable lente, il apparaît que le bassin principal est approximativement situé entre le nœud 12500 ($t = 250\text{ns}$) et le nœud 30000 ($t = 600\text{ns}$), et le second bassin entre le nœud 37500 ($t = 750\text{ns}$) et le dernier nœud 50000 ($t = 1000\text{ns}$). Le troisième bloc entre le premier et le dernier nœud 10000 ($t = 200\text{ns}$) ne correspond pas à un bassin, mais à l'intervalle de temps pendant lequel la simulation MD s'est stabilisée.

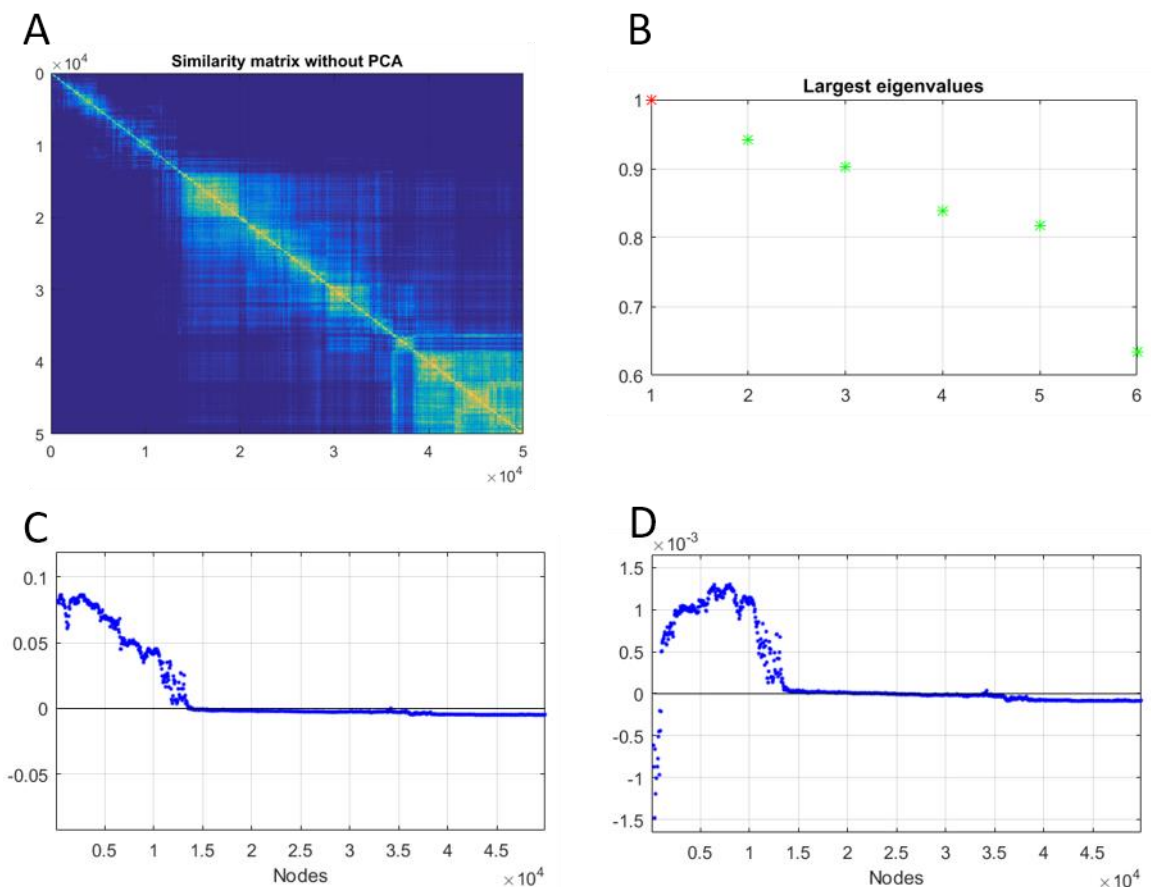


Figure 3.23.S2. Analyse spectrale de la trajectoire 2' sans réduction ACP. (A) Matrice de similarités W ; (B) Valeurs propres de A ; (C) Première variable lente; (D) Deuxième variable lente. L'emplacement des bassins est maintenant moins clair. Entre les nœuds 15000 et 30000, la pente du profil est très faible et il est difficile de déterminer l'emplacement exact du changement de signe.

Nous avons pu, avec cette nouvelle méthode, caractériser l'évolution des conformations de hVKORC1 en fonction du temps, puis avons représenté les conformations obtenues au cours de chaque trajectoire de DM comme une surface d'énergie temporellement dépendante. Cette estimation nous a permis d'identifier des minima locaux d'énergie bien définis et fortement peuplés, composés chacun de conformations reflétant différents états de hVKORC1. Deux conformations distinctes de l'enzyme ont été mises en évidence, la première (appelée conformation 1) présentant une distance $S \cdots S$ courte entre les atomes de soufre des résidus C_{51} - C_{132} , la seconde (conformation 2) présentant des distances courtes entre deux paires d'atomes de soufre des résidus C_{43} - C_{51} et C_{132} - C_{135} . Ces deux conformations suggèrent la formation de ponts disulfures au sein de VKORC1, et sont en conséquence des candidats prototypes pour la modélisation des états métastables de l'enzyme.

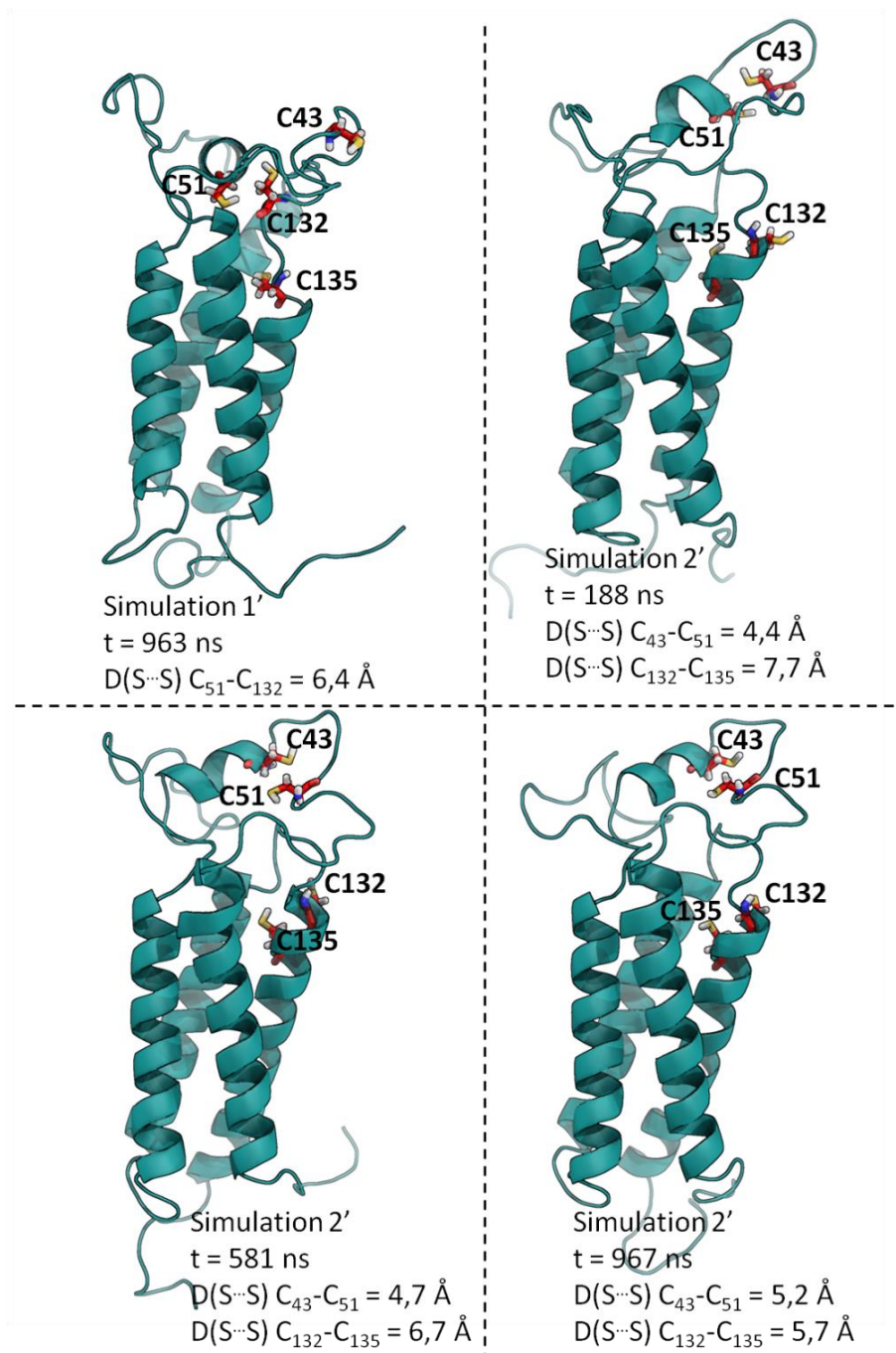


Figure 3.24. Structures 3D de hVKORC1_SH identifiés dans les minima d'énergie locaux durant deux simulations d'1 μ s. Le numéro et le temps de simulation à laquelle chaque conformation est observée sont indiqués sous les structures 3D. Les distances S··S entre les résidus de cystéine les plus proches sont également indiqués. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu ; les résidus de cystéines C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* rouges (les atomes de soufre, d'azote, d'oxygène et d'hydrogène sont colorés en jaune, bleu, rouge clair et blanc, respectivement).

La flexibilité de l'hélice HH joue probablement un rôle crucial dans l'association entre résidus de cystéine, puisqu'elle est stabilisée en position "basse" (proche des quatre TMs) dans la conformation 1 mais en position "haute" (éloignée des quatre TMs) dans la conformation 2. Nous avons étudié plus finement le positionnement de HH en analysant, au cours des simulations de DM, la distance entre le *centroïde* de HH et le plan moyen défini par les extrémités supérieures des quatre TMs. Cette distance, égale à 8 Å à $t = 0$ ns, se stabilise autour de 10 Å durant la simulation 1' (suggérant un pont disulfure C₅₁-C₁₃₂). Elle augmente jusqu'à 16 Å au cours de la simulation 2' (suggérant la présence de deux ponts disulfures, C₄₃-C₅₁ et C₁₃₂-C₁₃₅), puis se stabilise à 15 Å. Le résidu C₅₁ étant situé à l'extrémité de HH, nous pouvons donc postuler que cette petite hélice joue un rôle de "navette" entre les résidus C₁₃₂ du site actif (localisé sur TM4) et C₄₃ (localisé sur la boucle L), favorisant alternativement la formation d'un pont disulfure avec l'un ou l'autre de ces deux résidus (**Figure 3.25**).

Nos résultats suggèrent une implication des résidus C₄₃ et C₅₁ dans la formation de ponts disulfures. D'autres analyses sont cependant nécessaires pour confirmer notre hypothèse, notamment l'étude de modèles 3D incorporant les ponts disulfures identifiés. Ces investigations font l'objet de la partie suivante.

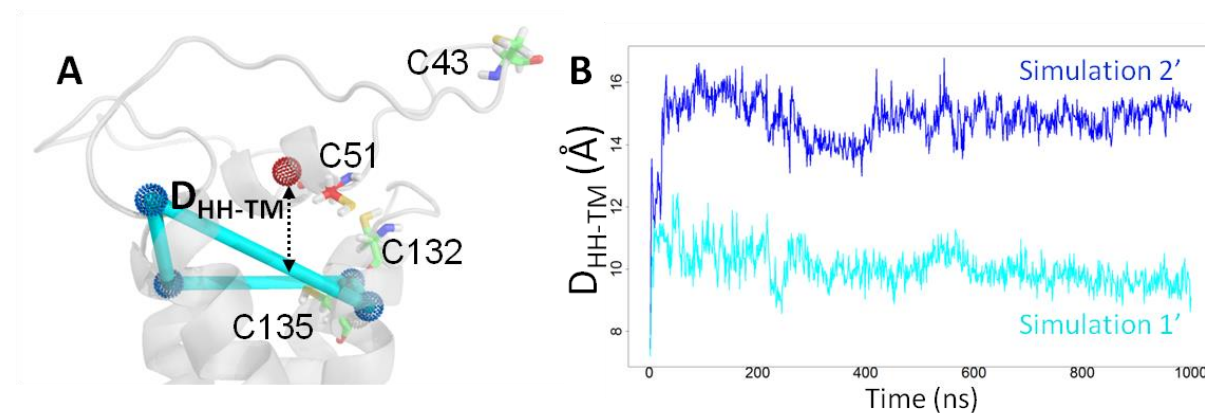


Figure 3.25. Positionnement de l'hélice HH par rapport aux quatre TMs au cours de deux simulations de DM d'1 µs. (A) Les *centroïdes* des extrémités supérieures des hélices TMs (sphères pointillées bleues) sont définis à partir des coordonnées des atomes de carbone α des résidus 25-28, 77-80, 120-123 et 132-135 ; le *centroïde* de l'hélice HH (sphère pointillée rouge) est défini à partir des coordonnées des atomes de carbone α des résidus 51-56. Le plan défini par les *centroïdes* des quatre TMs est indiqué par un tube cyan. Les résidus de cystéine C43, C51, C132 et C135 sont représentés en *sticks*. Le résidu C51 est coloré en rouge, les résidus C43, C132 et C135 en vert. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en gris clair. (B) Variations de la distance entre le *centroïde* de HH et le plan moyen des *centroïdes* des extrémités supérieures des quatre TMs au cours de deux simulations de DM d'1 µs.

III. Simulations de DM de hVKORC1 - Ponts disulfures prédits à partir des conformations métastables

1. Modélisation de structures 3D de hVKORC1 incluant les ponts disulfures prédits par simulations de DM

L'identification de conformations de hVKORC1_SH présentant des distances courtes entre atomes de soufre des résidus de cystéine nous a menés à postuler la formation de ponts disulfures entre ces résidus. Nous avons donc modélisé, à partir des conformations identifiées par DM de hVKORC1_SH, les structures 3D de hVKORC1 présentant les ponts disulfures identifiés. Les différents modèles ont été générés par MODELLER [73] en imposant pour chacun le ou les pont(s) disulfure(s) adéquat(s), par formation d'une liaison covalente entre les deux atomes de soufre concernés, dont la distance est alors fixée à 2,1 Å. Le premier modèle, basé sur la conformation 1 (distance S···S entre C₅₁ et C₁₃₂ égale à 6,4 Å), inclut un pont disulfure C₅₁-C₁₃₂ (**Figure 3.26.A**), les deux résidus C₄₃ et C₁₃₅ étant considérés à l'état protoné (SH). Un deuxième modèle, basé sur la conformation 2 (distances S···S entre les résidus C₄₃ et C₅₁ d'une part et C₁₃₂ et C₁₃₅ d'autre part égales à 4,7 Å et 6,7 Å, respectivement), inclut deux ponts disulfures C₄₃-C₅₁ et C₁₃₂-C₁₃₅ (**Figure 3.26.B**). Deux autres modèles ont également été dérivés à partir de la conformation 2, incluant chacun un seul pont disulfure, C₄₃-C₅₁ (résidus C₁₃₂ et C₁₃₅ à l'état protoné) ou C₁₃₂-C₁₃₅ (résidus C₄₃ et C₅₁ à l'état protoné, **Figure 3.26.C**). Ces deux modèles peuvent constituer des conformations intermédiaires entre les états représentés par les conformations 1 et 2. De plus, le choix de ces conformations a été renforcé par leur observation fréquente durant les simulations de 100 ns (simulations 3, 4 et 5).

Les quatre modèles ainsi générés sont notés respectivement hVKORC1_C₅₁-C₁₃₂, hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅, hVKORC1_C₄₃-C₅₁ et hVKORC1_C₁₃₂-C₁₃₅.

Chaque modèle optimisé a ensuite été inséré dans une membrane de DLPC, en utilisant la même méthode que celle appliquée au modèle de hVKORC1_SH dépourvu de pont disulfure (**Chapitre 3, partie I.5**). L'énergie potentielle de chaque système "hVKORC1 + membrane + eau + ions" a ensuite été minimisée, suivie par une équilibration de la température, puis de la pression. La comparaison entre chaque structure de hVKORC1 ainsi équilibrée et la conformation initiale ayant servi de "template" à sa modélisation montre des différences structurales, notamment au niveau de la boucle L (**Figure 3.26**).

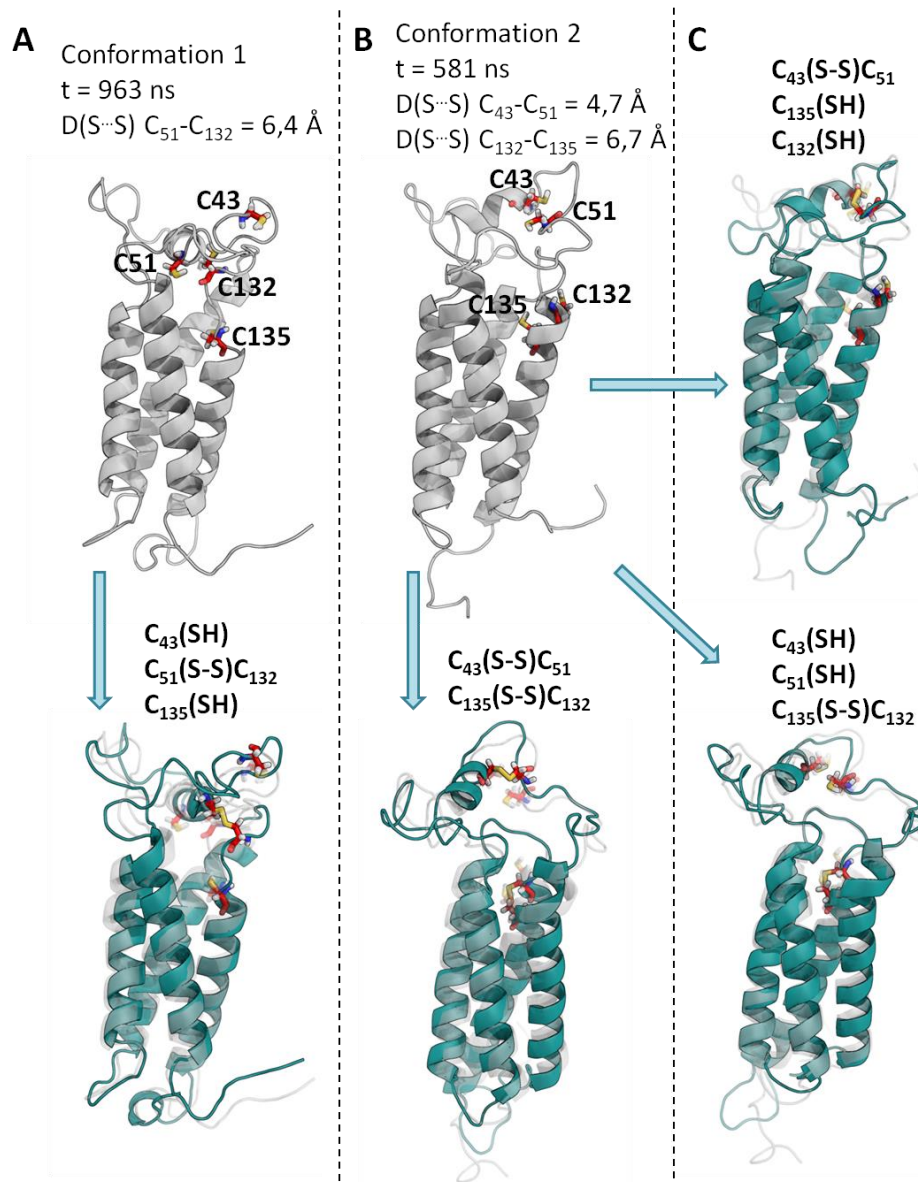


Figure 3.26. Modélisation de structures de hVKORC1 incluant des ponts disulfures à partir des conformations (U min) identifiées par dynamique moléculaire. (A) Modélisation d'une structure de hVKORC1 incluant un pont disulfure C₅₁-C₁₃₂ à partir de la conformation 1. (B) Modélisation d'une structure de hVKORC1 incluant deux ponts disulfures C₄₃-C₅₁ et C₁₃₂-C₁₃₅ à partir de la conformation 2. (C) Modélisation de deux structures de hVKORC1 incluant respectivement un pont disulfure C₄₃-C₅₁ et C₁₃₂-C₁₃₅ à partir de la conformation 2. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge ; les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon*. Les deux conformations de hVKORC1 identifiées par dynamique moléculaire sont colorées en gris ; le numéro de simulation, le temps auquel la conformation est observée ainsi que les résidus de cystéine dont les distances S...S sont les plus courtes sont indiquées au-dessus des conformations. Les modèles générés à partir de ces deux conformations et incluant des ponts disulfures sont colorés en bleu ; l'état protoné ou déprotoné de chaque résidu de cystéine est indiqué au-dessus des conformations. La conformation "template" utilisée pour chaque modèle est représentée en transparence afin de faciliter la comparaison entre les deux structures.

La formation de ponts disulfures modifiant la structure de hVKORC1, particulièrement le repliement de la boucle L, un affinement de la structure des quatre modèles générés est nécessaire. Nous avons donc effectué des simulations de DM d'une durée de 100 ns pour chacun des nouveaux modèles optimisés.

Les courbes des RMSDs (calculées sur les atomes de carbone α des résidus 9 à 151 de chaque conformation de hVKORC1 par rapport aux coordonnées de la conformation initiale) des simulations de DM montrent une équilibration des modèles après 20 ns (déviations comprises entre 2 et 5 Å). Le modèle hVKORC1_C132-C135 présente les déviations les plus importantes (stabilisation des valeurs de RMSDs autour de 4 Å, **Figure 3.27.A**).

Cette augmentation des déviations s'explique par la flexibilité de la boucle L, au niveau de laquelle les résidus C₄₃ et C₅₁ sont non-liés; la boucle L conserve ainsi un maximum de flexibilité et peut adopter une grande diversité de conformations. La comparaison avec les simulations de DM de hVKORC1_SH, pour lesquelles les valeurs de RMSDs se situent entre 3 et 7 Å (**Figure 3.12.A**), révèle l'effet stabilisateur des ponts disulfures sur l'ensemble de la structure de hVKORC1.

Les courbes des RMSFs (calculées sur les atomes de carbone α des 163 résidus de hVKORC1) indiquent des fluctuations atomiques faibles au niveau des TMs des quatre modèles (valeurs de RMSFs aux alentours d'1 Å), tandis que des différences apparaissent au niveau de la boucle L. En effet, le modèle hVKORC1_C43-C51_C132-C135 présente les fluctuations les plus faibles (environ 1 Å), seul un pic de fluctuations atteignant 3 Å est observé aux alentours du résidu 70, positionné sur la partie de la boucle L liant HH à TM2. Le modèle hVKORC1_C43-C51 montre des fluctuations équivalentes sur les deux moitiés de la boucle L, atteignant également 3 Å. Le modèle hVKORC1_C51-C132 présente un pic de fluctuations atteignant 5 Å autour du résidu 40 (proche du résidu C₄₃ à l'état SH), mais la seconde moitié de la boucle L reste peu fluctuante (valeurs de RMSFs inférieures à 2 Å). Le modèle hVKORC1_C132-C135 présente un profil similaire, les fluctuations les plus importantes (environ 4 Å) sont localisées autour du résidu C₄₃ (**Figure 3.27.B**). L'ensemble de ces observations démontre l'influence des ponts disulfures formés sur la flexibilité des modèles de hVKORC1, puisque chaque région de la boucle L contenant un résidu de cystéine impliqué dans un pont disulfure présente une réduction de ses fluctuations atomiques.

Le repliement des quatre TMs est hautement préservé pour chaque modèle de hVKORC1, comme le montre l'assignement des structures secondaires par DSSP (**Figure 3.27.C**). Le repliement de la boucle L et l'hélice HH est également très conservé pour le modèle hVKORC1_C51-C132, comme il l'était dans le modèle initial hVKORC1_SH.

Il présente cependant des différences dans les trois autres modèles, en particulier au niveau des résidus C51-S56 formant l'hélice HH dans hVKORC1_SH. Ces résidus adoptent un repliement en boucle dans le modèle hVKORC1_C43-C51_C132-C135, pour lequel le repliement régulier (en hélice α ou en hélice 3_{10}) est observé pour les résidus 40-45 et 55-65. Les résidus 40-45 du modèle hVKORC1_C43-C51 adoptent également un repliement en hélice α ; l'hélice HH est conservée mais se structure en hélice 3_{10} plus qu'en hélice α . Le modèle hVKORC1_C132-C135 montre une perte de structuration de l'hélice HH d'une part et un repliement en hélice α des résidus 40-45 et 55-65 d'autre part. Ces changements de structures secondaires dans les différents modèles de hVKORC1 confirment l'influence des ponts disulfures sur le repliement de la boucle L.

La formation d'un pont disulfure C₄₃-C₅₁ aboutit à un repliement en hélice α des résidus 40-45, alors qu'un pontage entre les résidus C₁₃₂ et C₁₃₅ provoque une perte de structuration de l'hélice HH et un repliement en hélice α des résidus 55-65. Les effets structuraux induits par la formation des ponts disulfures sont cumulatifs, puisque le modèle hVKORC1_C43-C51_C132-C135 présente une restructuration totale de la boucle L. L'hélice HH, outre son rôle de "navette" observé précédemment (**Figure 3.25**), montre également une grande variabilité structurale en fonction des ponts disulfures formés. Cette observation renforce notre théorie de l'influence de l'hélice HH dans l'association du résidu C₅₁ avec les résidus C₄₃ et C₁₃₂.

Nous avons caractérisé la variabilité structurale de la boucle L pour les quatre états conformationnels de hVKORC1 étudiés, et mis en évidence l'influence des ponts disulfures sur le repliement de cette boucle. Les quatre modèles structuraux décrits sont susceptibles de fixer la vitamine K. Le résidu C₁₃₅ est décrit dans la littérature comme étant le résidu catalytique, responsable de l'attaque de la vitamine K époxyde au niveau de l'atome de carbone 3, mais l'état enzymatique de hVKORC1 le plus favorable à la fixation de la vitamine K époxyde n'est pas clairement défini (**Figure 3.28**). Nous avons donc étudié l'affinité entre la vitamine K époxyde et les différents états conformationnels de hVKORC1.

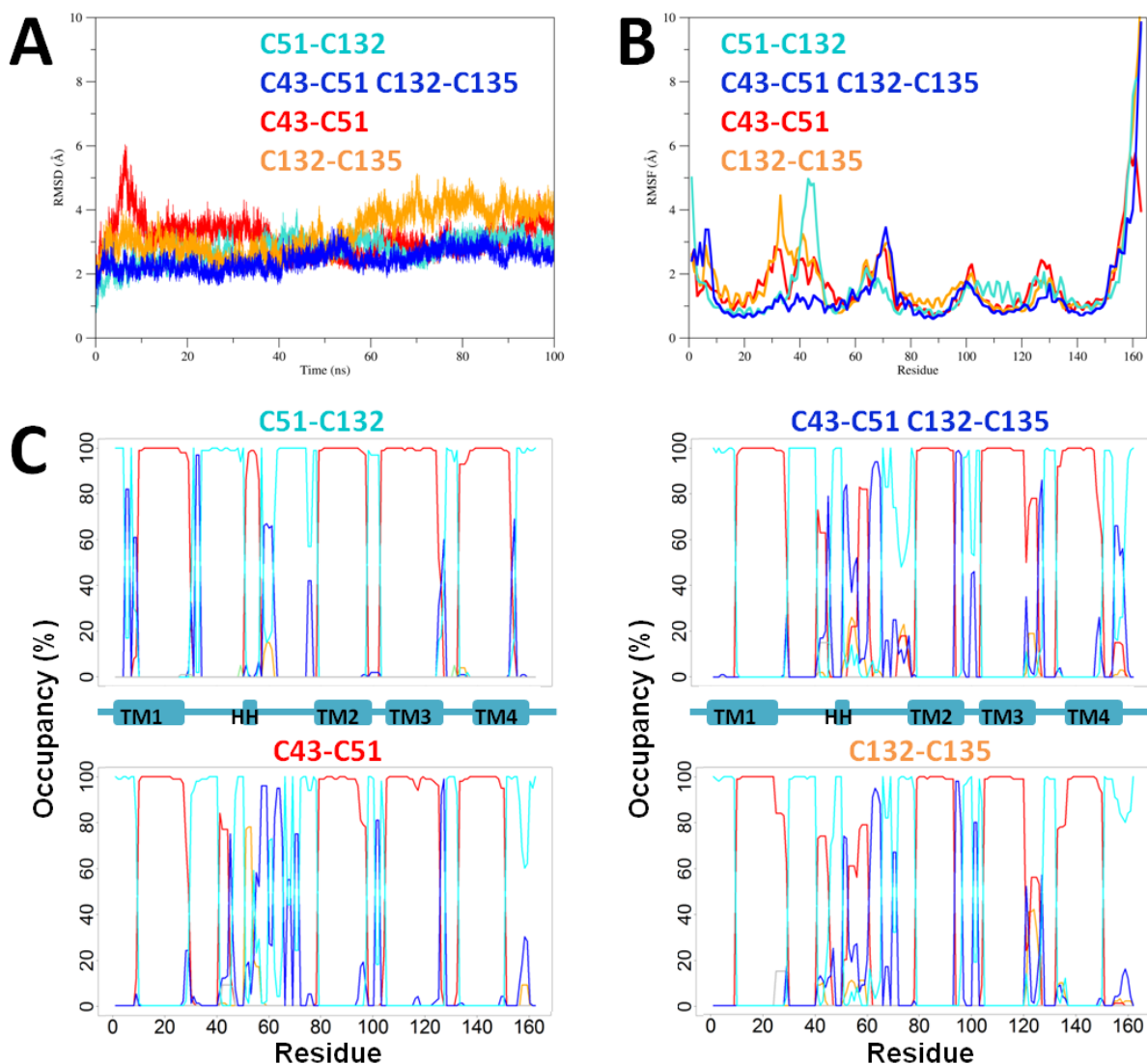


Figure 3.27. Caractérisation de la stabilité structurale globale des quatre modèles prédits pour hVKORC1. (A) Courbes des RMSDs des quatre modèles de hVKORC1 au cours de simulations de 100 ns. Les RMSDs ont été calculées sur les atomes de carbone α des résidus 9 à 151 (les résidus des extrémités N- et C-terminales ont été omis afin d'éliminer le bruit induit par leurs grandes fluctuations), après alignement des atomes de carbone α des hélices transmembranaires de chaque conformation sur les coordonnées de la conformation initiale ($t = 0$ ns). Les couleurs des états conformationnels associés à chaque courbe sont indiquées sur le graphe. (B) Courbes des RMSFs des quatre modèles de hVKORC1 au cours de simulations de DM de 100 ns. Les RMSFs ont été calculées sur les 163 atomes de carbone α de hVKORC1, préalablement alignés sur les coordonnées de la conformation initiale ($t = 0$ ns). Les couleurs des états conformationnels associés à chaque courbe sont indiquées sur le graphe. (C) Assignement par DSSP des structures secondaires des conformations des quatre modèles de hVKORC1 au cours de simulations de DM de 100 ns (en pourcentage). Les repliements en hélice α , hélice_{3,10}, boucle, coude et brin β sont indiqués par les courbes en rouge, orange, cyan, bleu foncé et vert clair, respectivement. Les structures secondaires de la conformation initiale ($t = 0$ ns) ainsi que les ponts disulfures présents dans les modèles de hVKORC1 sont indiqués au-dessus des graphes.

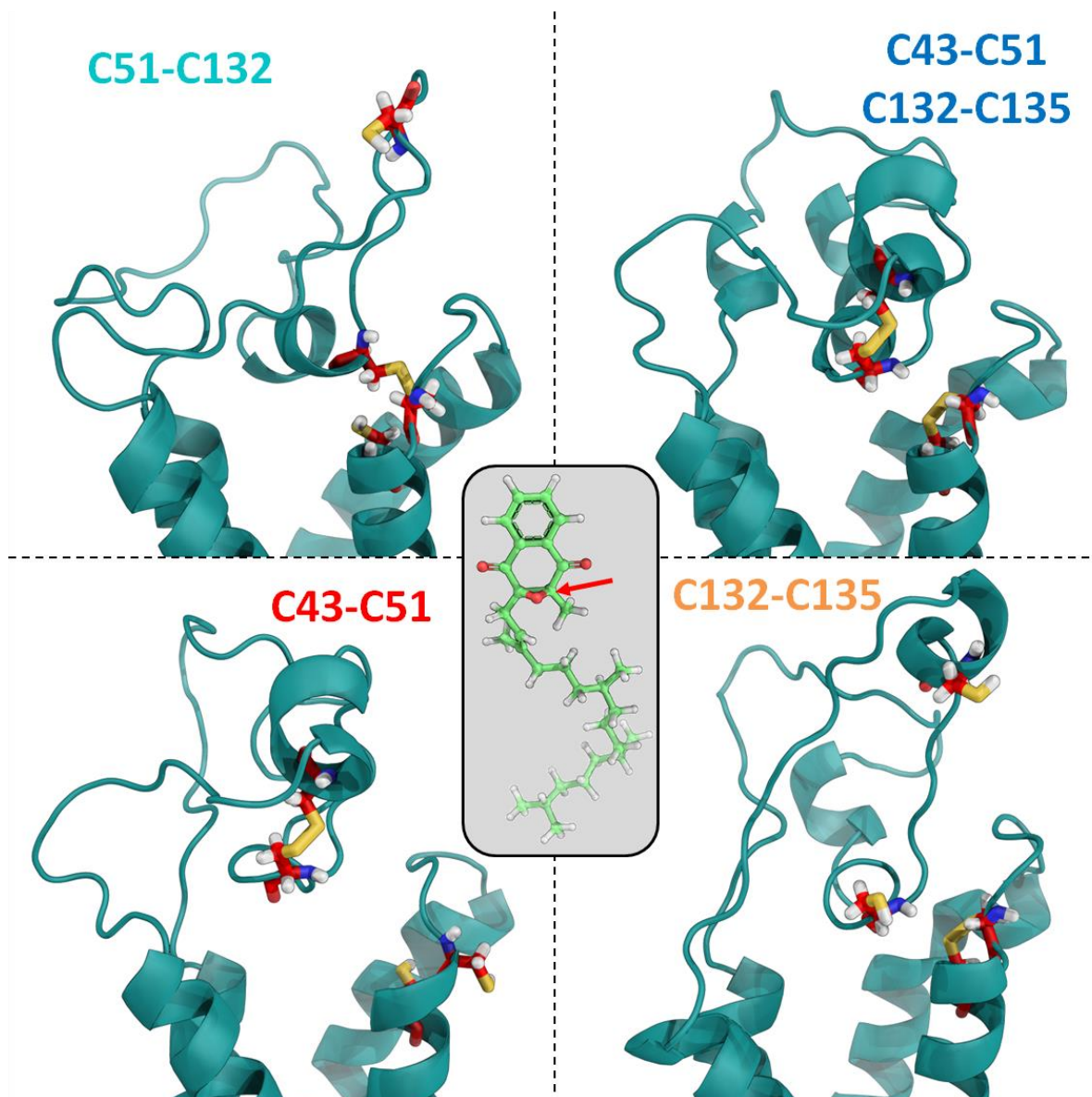


Figure 3.28. Modèles de hVKORC1 prédits par simulations de DM. Les quatre modèles de hVKORC1 sont représentés en *cartoon* et colorée en bleu, les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. La structure de la vitamine K époxyde est représentée en *sticks* et colorées en vert dans l'encart ; la flèche rouge indique l'atome de carbone 3.

2. Détection de poches à la surface des quatre structures 3D prédites pour hVKORC1

Nous avons appliqué une méthode de *clustering*, basé sur les RMSDs calculées deux à deux entre chaque conformation d'une trajectoire (voir **METHODES**), afin d'identifier les conformations représentatives de chaque modèle obtenues au cours des simulations de DM. Nos résultats ont mis en évidence un unique *cluster* représentatif pour chacun des modèles hVKORC1_C₅₁-C₁₃₂ (représentant 73% des conformations totales et regroupant les conformations obtenues de 30 à 100 ns de simulation), hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅ (représentant 76% des conformations totales et regroupant les conformations obtenues de 30 à 100 ns de simulation) et hVKORC1_C₄₃-C₅₁ (représentant 56% des conformations totales et regroupant les conformations obtenues de 40 à 100 ns de simulation). Le modèle hVKORC1_C₁₃₂-C₁₃₅ présente deux *clusters* principaux représentant chacun 37% et 24% des conformations totales, ainsi qu'un troisième *cluster* plus faiblement peuplé (représentant 18% des conformations totales, **Figure 3.29**). L'obtention de *clusters* multiples pour ce modèle s'explique par une grande flexibilité et donc une grande variabilité conformationnelle de la boucle L, due à l'absence de pont disulfure impliquant les résidus C₄₃ et C₅₁.

Nous avons ensuite recherché des poches à la surface des différents modèles de hVKORC1, en considérant les conformations représentatives de chaque *cluster* identifié. Nous avons utilisé l'outil MDpocket (voir **METHODES**) pour détecter les poches préservées parmi toutes les conformations composant chaque *cluster*. Nous avons ainsi identifié une poche pour chaque conformation représentative de chacun des quatre modèles de hVKORC1. Nous avons identifié une poche sur les conformations formant un unique *cluster* pour trois modèles. La poche détectée pour les modèles hVKORC1_C₁₃₂-C₁₃₅, est identique pour chacun des trois *clusters*. Malgré de la diversité de conformations propre à cet état de l'enzyme, la flexibilité de la boucle L ne semble pas affecter la géométrie de la poche.

La poche identifiée pour chacun des états conformationnels de hVKORC1 est localisée au niveau de la partie supérieure du domaine transmembranaire. Elle est délimitée par les quatre TMs et l'hélice HH. Sa géométrie diffère d'un état enzymatique à l'autre, tout comme son volume apparent, qui semble plus important pour le modèle hVKORC1_C₅₁-C₁₃₂ et moindre pour le modèle hVKORC1_C₁₃₂-C₁₃₅ (**Figure3.30.A**). De plus, la présence de lipides (chaînes hydrophobes) dans la poche détectée est systématiquement observée (**Figure3.30.B**).

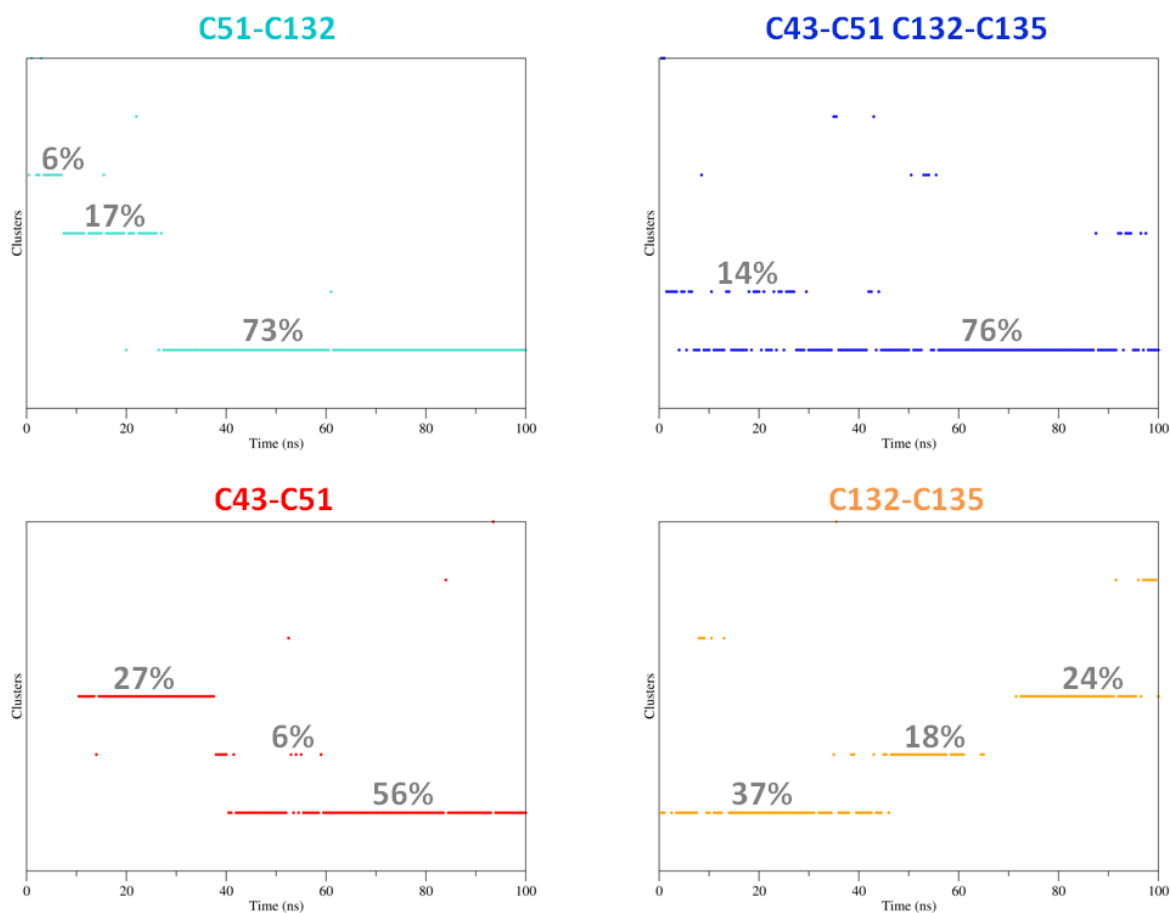


Figure 3.29. Analyse des conformations similaires (*clustering*) de hVKORC1 pour chaque modèle prédit.

Les conformations appartenant à un même *cluster* sont indiquées par des points, regroupés en lignes de niveau au cours de simulations de dynamique moléculaire de 100 ns. La population de chaque *cluster* (proportion des conformations membres d'un même cluster par rapport au nombre de conformations total de la trajectoire) est indiquée en pourcentage, pour les *clusters* représentant plus de 5% des conformations totales. Les ponts disulfures sont indiqués au-dessus de chaque graphe. Le *clustering* est basé sur les RMSDs calculées deux à deux entre les atomes de carbone α des résidus 9-151 de chaque conformation d'une trajectoire. Deux conformations présentant une valeur de RMSD inférieure à 2 Å sont considérées comme membres d'un même *cluster*. Le *cluster* numéro 1 est toujours le plus peuplé, et chaque conformation appartient à un seul *cluster*.

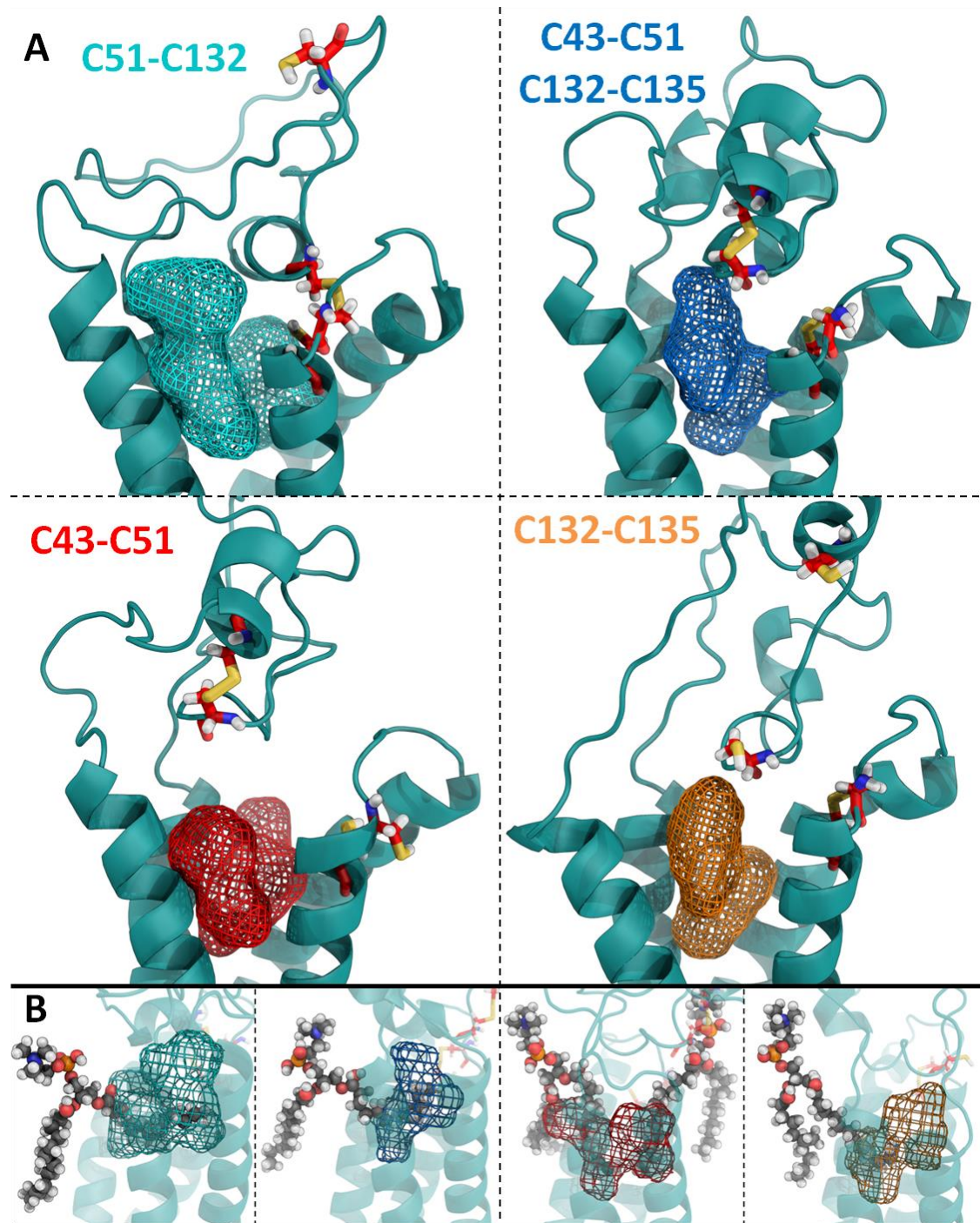


Figure 3.30. Poches détectées à la surface des structures 3D pour chaque conformation la plus représentative de chaque modèle de hVKORC1.(A) Les cavités détectées (MDpocket) à la surface des structures 3D de chaque état enzymatique de hVKORC1 sont représentées par un maillage (*mesh*). Les ponts disulfures présents dans chaque modèle de hVKORC1 sont indiqués au-dessus des structures 3D. (B) Les lipides dont la queue hydrophobe pénètre dans la cavité de VKORC1 durant les simulations de DM sont représentés en sphères, les atomes de carbones α sont colorés en gris foncés. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote, de phosphore et d'hydrogène sont colorés en jaune, rouge clair, bleu, orange et blanc, respectivement.

La cavité identifiée constitue un site de fixation potentiel pour la vitamine K époxyde (vitK^E). La pénétration des chaînes hydrophobes des lipides dans cette poche indique son caractère hydrophobe. La proportion de résidus hydrophobes constituant la cavité est comprise entre 27 et 30% selon les états conformationnels. Le volume de la poche occupé par les chaînes hydrophobes des lipides pourrait correspondre à la localisation de la chaîne hydrophobe de la vitK^E lors de sa fixation sur hVKORC1. Nous avons donc mesuré le volume de la poche pour chaque modèle de hVKORC1 sur la protéine seule (sans prendre en compte les lipides présent dans la cavité), en considérant les conformations composant les *clusters* hautement peuplés pour chaque modèle. Les résultats obtenus révèlent une tendance nette quant au volume des cavités propres à chaque modèle de hVKORC1.

La cavité la plus volumineuse est celle du modèle hVKORC1_C₅₁-C₁₃₂, dont le volume moyen est supérieur à 800 Å³ (**Figure 3.31**). Les cavités détectées pour les modèles hVKORC1_C₄₃-C₅₁ et hVKORC1_C₁₃₂-C₁₃₅ présentent un volume moyen similaire d'environ 600 Å³. La cavité du modèle hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅ est la plus étroite, avec un volume moyen de 500 Å³. Ces observations mettent en évidence l'effet des ponts disulfures sur la structure globale de hVKORC1, chaque état enzymatique possédant une poche dont le volume et la géométrie lui sont propres (le double pontage réduisant significativement le volume de la poche). Les différences entre les volumes de poche mesurés résultent de variations d'enfouissement ou d'exposition de certains résidus à la surface des cavités identifiées sur chaque état conformationnel de hVKORC1. Ainsi, cinq résidus (asparagine, histidine, isoleucine, tryptophane et valine) sont exposés à la surface de la cavité du modèle hVKORC1_C₅₁-C₁₃₂, mais sont enfouis pour les trois autres états conformationnels. Deux résidus (thréonine et valine) sont présents à la surface de la cavité du modèle hVKORC1_C₄₃-C₅₁, mais ne sont pas exposés pour les trois autres modèles. Enfin, deux résidus (arginine et asparagine) sont exposés à la surface de la cavité du modèle hVKORC1_C₁₃₂-C₁₃₅, mais enfouis pour les trois autres modèles. La localisation précise de chacun de ces résidus est présentée en détail dans la partie **VI (RESULTATS)**, lors de l'analyse de la contribution de chaque résidu à l'énergie libre de liaison entre VKORC1 et différents ligands. L'importance des écarts-types (de l'ordre de 100 Å³) est due aux chaînes latérales des résidus de hVKORC1, orientées alternativement vers l'intérieur ou l'extérieur de la poche au cours des simulations de DM.

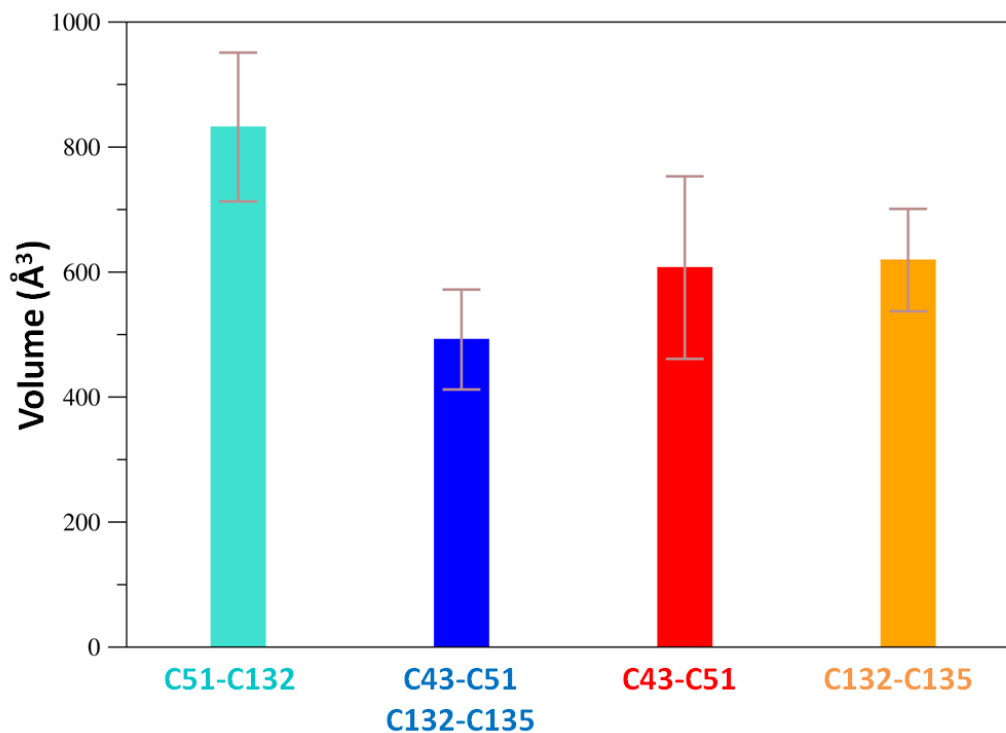


Figure 3.31. Volume des cavités détectées à la surface des structures 3D des conformations représentatives de chaque modèle de hVKORC1. Les écarts-types sont indiqués par une barre marron. Les modèles de hVKORC1 et les ponts disulfures présents sont indiqués sous le graphe.

Les volumes des cavités mesurés ainsi que les variations de résidus enfouis ou exposés à la surface des poches détectées constituent une première indication quant à la capacité de chaque état enzymatique à fixer la vitK^E. La localisation transmembranaire des cavités identifiées est en bonne adéquation avec le caractère hydrophobe de la vitK^E (conféré par la longue chaîne carbonée de la molécule). L'affinité de la molécule vis-à-vis de chaque modèle a ensuite été évaluée par amarrage moléculaire, dont les résultats sont présentés dans la partie suivante.

IV. Sélectivité des modèles de hVKORC1 vis-à-vis de la vitamine K époxyde

1. Amarrage moléculaire de la vitamine K époxyde sur les quatre états enzymatiques prédits pour hVKORC1

L'affinité de la vitK^E pour les quatre modèles de hVKORC1 a été estimée par amarrage (*docking*) moléculaire. Les poches identifiées précédemment sur les conformations représentatives de chaque modèle ont été définies comme site de fixation potentiel de la vitamine K. Pour chaque modèle, hVKORC1_C51-C132, hVKORC1_C43-C51_C132-C135, hVKORC1_C43-C51 et hVKORC1_C132-C135, 100 poses de *docking* de la vitK^E ont été effectuées. Le *docking* est effectué en considérant le ligand (vitK^E) flexible et la cible (hVKORC1) rigide. Les poses sont ensuite classées en fonction de leur score, correspondant à une estimation de l'énergie libre de liaison (ELL) entre la vitK^E et hVKORC1 (voir **METHODES**).

L'amarrage moléculaire de la vitK^E dans la cavité de hVKORC1_C51-C132 a généré 8 groupes de poses de la vitK^E, chaque groupe étant constitué de poses identiques. Le premier groupe contient 49% des poses totales de la vitK^E, et possède l'ELL la plus négative (-8,96 kcal/mol), donc la plus favorable à la fixation de la vitK^E sur hVKORC1 (**Figure 3.32.B**). Les 7 autres groupes de poses présentent des ELL et des populations décroissantes (le second groupe comprend 23% des poses totales et présente une ELL de -5,17 kcal/mol, le troisième groupe inclut 15% des poses totales et son ELL est de -4,71 kcal/mol).

Les 5 derniers groupes représentent une part non significative des poses, leurs populations étant inférieures à 10%, et sont donc considérées comme non représentatives. La longue chaîne carbonée de la vitK^E possède une grande flexibilité, pouvant perturber les résultats du *docking* effectué en adoptant une multitude de conformations. Cette chaîne hydrophobe n'intervenant pas directement dans la réaction enzymatique catalysée par hVKORC1 (le groupement époxyde de la vitamine K réduit par l'enzyme est localisé sur un cycle carboné), nous avons choisi de tronquer la vitK^E après le troisième atome de carbone de la chaîne hydrophobe (**Figure 3.32.A**), puis avons effectué un nouvel amarrage moléculaire en utilisant le même protocole paramètres.

Les nouveaux résultats montrent une réduction importante du nombre de groupes de poses (seulement 2, comparés aux 8 précédents). Le premier groupe est hautement majoritaire (88% des poses totales) et présente une ELL de -7,22 kcal/mol.

La pose de vitK^E correspondant est identique à la pose obtenue lors du *docking* de la vitK^E intégrale. Le second groupe ne contient que 12% des poses totales de vitK^E et son ELL (-7,08 kcal/mol) et très proche de celle du premier groupe (**Figure 3.32.C**).

Une comparaison des ELL pour deux poses identiques des vitK^E intégrale et tronquée indique une ELL moins favorable pour la forme tronquée de la molécule (-7,22 kcal/mol pour la forme tronquée, contre -8,96 kcal/mol pour la forme intégrale, **Figure 3.32.B-C**). Cette observation montre le rôle de la chaîne carbonée de la vitK^E dans l'interaction de la molécule avec hVKORC1. La coupure de la chaîne carbonée de la vitK^E permet cependant l'élimination des "poses parasites" lors de l'amarrage moléculaire, tout en conservant l'orientation et le positionnement de la tête cyclique de la vitK^E dans la cavité de hVKORC1 pour les poses majoritaires. La pose de *docking* la plus favorable pour l'interaction de la vitKE avec hVKORC1_C51-C132 correspond à un positionnement de la vitK^E à proximité des résidus du site actif C₁₃₂ et C₁₃₅, le groupement époxyde étant orienté vers ces résidus (**Figure 3.32.C**). Le même protocole de *docking* a été appliqué pour évaluer l'affinité entre la vitK^E (forme tronquée) et les trois autres modèles de hVKORC1.

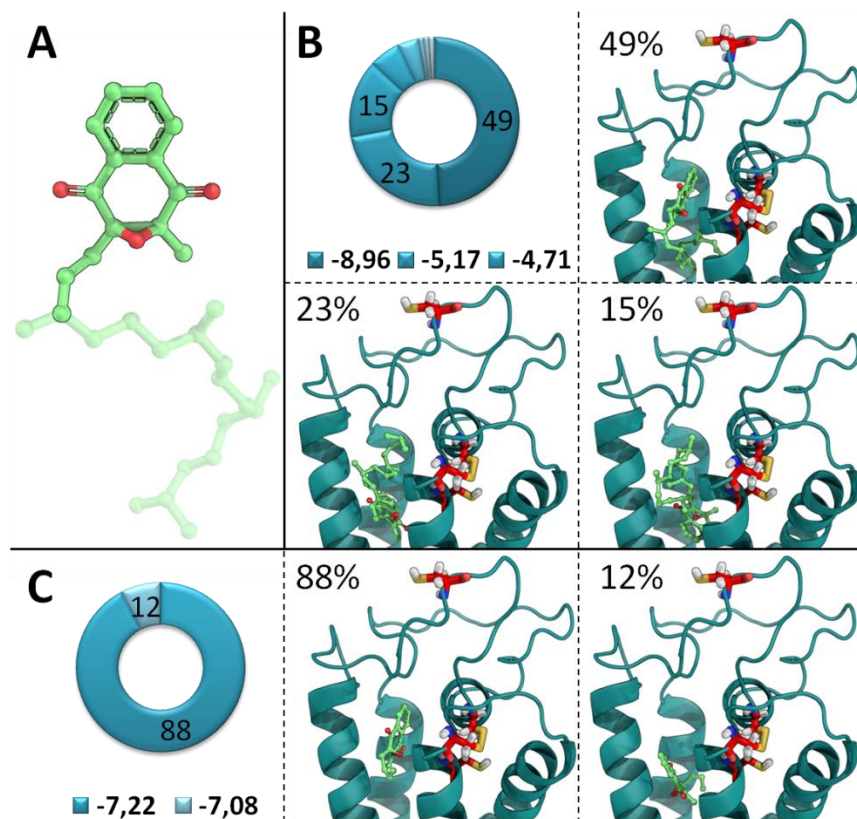


Figure 3.32. Amarrage moléculaire de la vitK^E (forme intégrale et forme tronquée) sur le modèle hVKORC1_C51-C132. (A) Représentation de la structure 3D de la vitamine K époxyde. La molécule est représentée en *sticks* verts ; les atomes sont représentés par des sphères ; les atomes d'oxygène sont colorés en rouge. La chaîne carbonée représentée en transparence correspond au fragment de la vitK^E tronqué avant de procéder au second *docking*. (B) Résultats du *docking* de la vitK^E intégrale sur le modèle hVKORC1_C51-C132. (C) Résultats du *docking* de la vitK^E tronquée sur le modèle hVKORC1_C51-C132. Les poses similaires obtenues au cours de chaque *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe est indiqué pour les groupes les plus représentatifs (incluant plus de 10% des poses totales). Les énergies libres de liaison (scores) associées à chaque groupe de poses sont indiquées sous le diagramme. Les structures 3D des complexes hVKORC1•vitK^E correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

Les résultats obtenus pour le modèle hVKORC1_C43-C51_C132-C135 indiquent deux groupes de poses, comprenant chacun 58% et 42% des poses totales et sont les ELL respectives sont de -7,03 kcal/mol et de -6,8 kcal/mol. La pose dont l'ELL est la plus négative correspond à un positionnement de la vitK^E à proximité du site actif de hVKORC1 (résidus C₁₃₂ et C₁₃₅), mais le groupement époxyde est orienté à l'opposé de ces deux résidus. La seconde pose montre un positionnement de la vitK^E plus éloigné du site actif de l'enzyme, le groupement époxyde étant également orienté à l'opposé des résidus C₁₃₂ et C₁₃₅ (**Figure 3.33**).

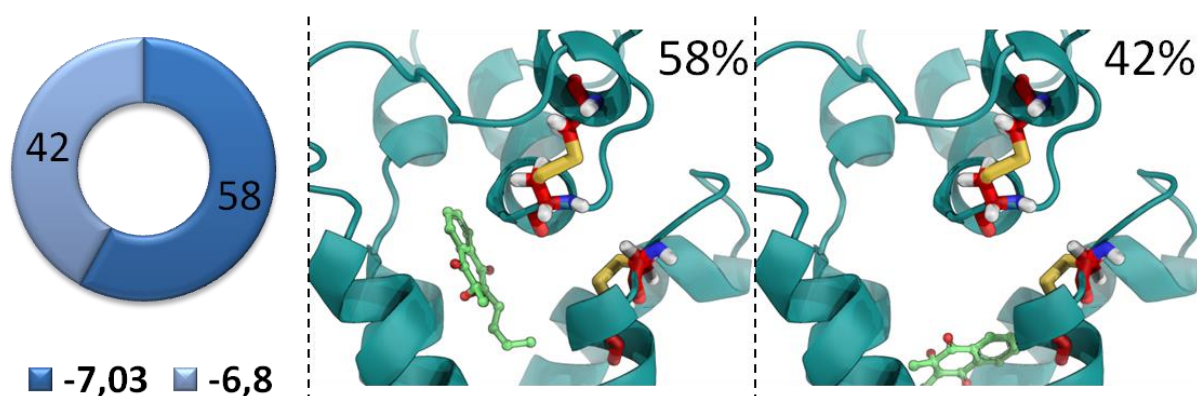


Figure 3.33. Amarrage moléculaire de la vitK^E tronquée sur le modèle hVKORC1_C43-C51_C132-C135. Les poses similaires obtenues au cours du *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe et les énergies libres de liaison associées à chaque groupe de poses sont indiqués sur le diagramme. Les structures 3D des complexes hVKORC1•vitK^E correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

Le *docking* de la vitKE sur le modèle hVKORC1_C43-C51 a généré 4 groupes de poses, regroupant chacun 74, 14, 10 et 2% des poses totales de vitKE (le dernier groupe est considéré comme non significatif étant donné sa faible population). Les 3 poses majoritaires, dont les ELL sont proches (-7,37, -7,17 et -7,15 kcal/mol, respectivement), correspondent toutes à un positionnement de la vitK^E éloigné du site actif de hVKORC1. Le groupement époxyde est orienté à l'opposé du site actif, à l'exception du groupe le moins représentatif (10% des poses totales), pour lequel le groupement époxyde est orienté vers TM4 (**Figure 3.34**).

Le *docking* de la vitK^E sur le modèle hVKORC1_C132-C135 a généré 5 groupes de conformations, incluant chacun 42, 25, 4 (poses considérées comme non représentatives), 16 et 13% des poses totales, et dont les ELL sont proches (-6,92, -6,86, -6,74 et -6,64 kcal/mol, respectivement). Les poses correspondant à chaque groupe indiquent toutes une vitK^E éloignée du site actif de hVKORC1. Le groupement époxyde est orienté à l'opposé des résidus C₁₃₂ et C₁₃₅ pour les deux premiers groupes, mais fait face au site actif pour les deux derniers groupes (**Figure 3.35**).

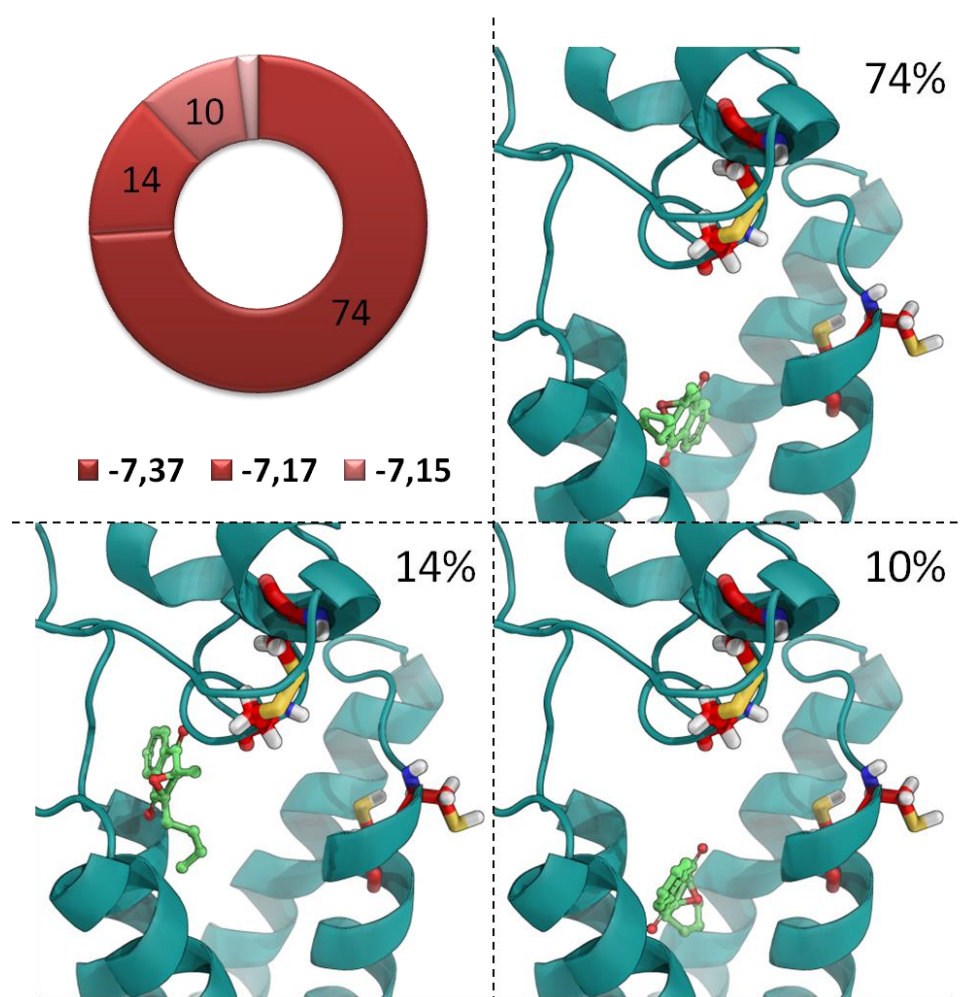


Figure 3.34. Amarrage moléculaire de la vitK^E tronquée sur le modèle hVKORC1_C43-C51. Les poses similaires obtenues au cours du *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe et les énergies libres de liaison associées à chaque groupe de poses sont indiqués sur le diagramme. Les structures 3D des complexes hVKORC1•vitK^E correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

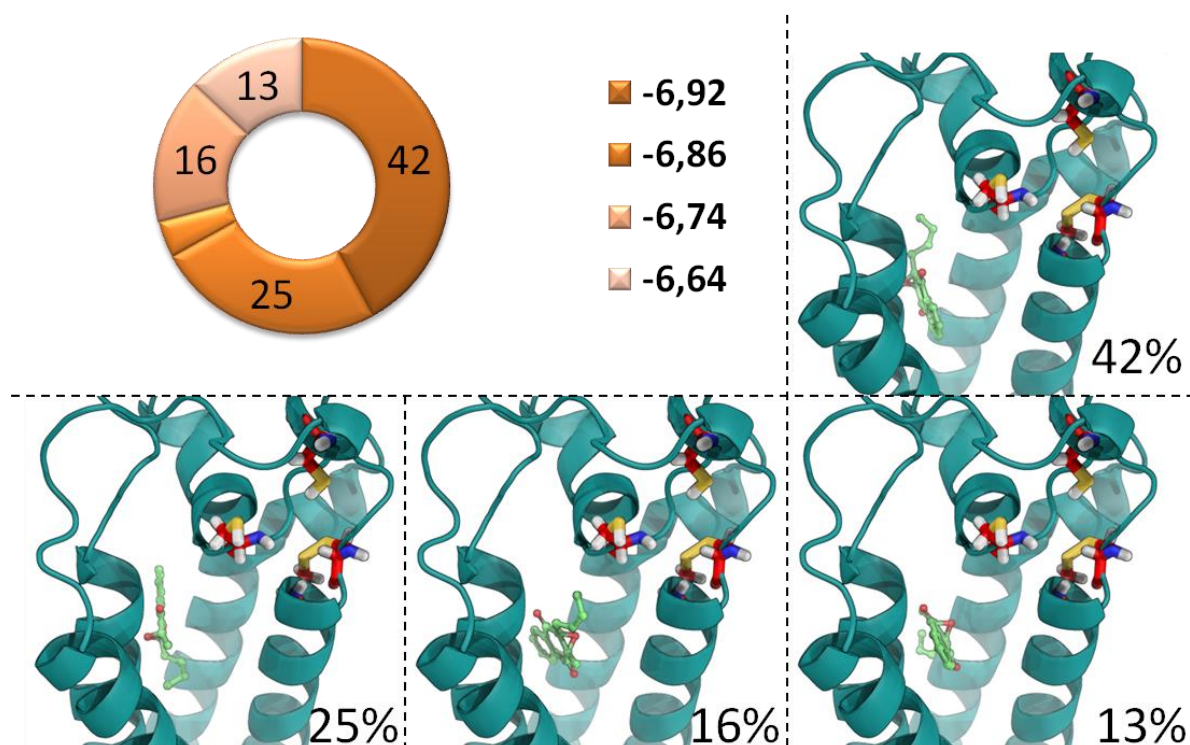


Figure 3.35. Amarrage moléculaire de la vitK^E tronquée sur le modèle hVKORC1_C132-C135. Les poses similaires obtenues au cours du *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe et les énergies libres de liaison associées à chaque groupe de poses sont indiqués sur le diagramme. Les structures 3D des complexes hVKORC1•vitK^E correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

Les poses de *docking* les plus significatives ont finalement été considérées. Les ELL estimées pour chaque complexe étant très proches (entre -7,37 kcal/mol et -6,80 kcal/mol), nous avons confronté les poses obtenues aux données expérimentales disponibles dans la littérature afin d'identifier le modèle de hVKORC1 le plus plausible pour la fixation de la vitK^E. Les données biologiques publiées suggèrent une attaque au niveau du carbone 3 de la vitK^E par l'atome de soufre du résidu C₁₃₅ de hVKORC1 lors de la réaction enzymatique [16]. La pose majoritaire obtenue lors de l'amarrage de la vitK^E sur le modèle hVKORC1_C51-C132 (88% des poses totales, ELL de -7,22 kcal/mol) est en bonne adéquation avec les données expérimentales, puisque la vitK^E est localisée à proximité du site actif de hVKORC1, son groupement époxyde étant orienté vers les résidus C₁₃₂ et C₁₃₅.

En revanche, le nombre important de poses distinctes de la vitK^E sur les autres modèles de hVKORC1 reflète leur faible sélectivité vis-à-vis de la molécule. De plus, les poses majoritaires obtenues après *docking* de la vitK^E sur ces autres modèles semblent moins pertinentes, puisqu'elles sont éloignées du site actif de hVKORC1, et le groupement époxyde est souvent orienté à l'opposé des résidus C₁₃₂ et C₁₃₅. Seule la pose majoritaire obtenue par amarrage de la vitK^E sur hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅ correspond à un positionnement de la vitK^E à proximité du site actif de hVKORC1. Cependant, l'orientation du groupement époxyde à l'opposé des résidus C₁₃₂ et C₁₃₅, ainsi que la présence d'un pont disulfure liant ces deux résidus, rendent hautement improbable la réduction de la vitK^E en vitamine K quinone (vitK^Q) par cet état conformationnel de hVKORC1.

L'état enzymatique identifié comme le plus favorable à la fixation de la vitK^E est donc hVKORC1_C₅₁-C₁₃₂. Néanmoins, l'orientation de la chaîne latérale du résidu C₁₃₅ à l'opposé de la pose de la vitK^E majoritaire (**Figure 3.32.C**) reste problématique. Nous avons choisi de prolonger la simulation de DM de 100 ns précédemment réalisée sur cet état conformationnel jusqu'à 500 ns, afin de laisser le modèle s'équilibrer plus longtemps. Nous espérons ainsi obtenir une conformation du modèle hVKORC1_C₅₁-C₁₃₂ dans laquelle la chaîne latérale du résidu C₁₃₅ est orientée dans la cavité de l'enzyme.

2. Extension de la simulation de DM de l'état hVKORC1_C₅₁-C₁₃₂ à 500 ns pour l'amarrage moléculaire de la vitamine K époxyde

La simulation de DM de l'état hVKORC1_C₅₁-C₁₃₂ précédemment réalisée a été étendue à 500 ns. La courbe des RMSDs (calculées sur les atomes de carbone α des résidus 9 à 151 de chaque conformation de hVKORC1 par rapport aux coordonnées de la conformation initiale) obtenues pour la simulation du modèle hVKORC1_C₅₁-C₁₃₂ étendue à 500 ns révèle la présence de deux plateaux. Le premier s'étend de 30 à 300 ns et correspond à la prolongation de l'état d'équilibre obtenu durant la première simulation de 100 ns ; les valeurs de RMSDs sont stabilisées aux alentours de 3 Å. Elles augmentent progressivement entre 300 et 350 ns, puis se stabilisent autour de 6 Å jusqu'à 500 ns, définissant un second plateau (**Figure 3.36.A**). Cette transition entre deux plateaux nettement définis est typique d'un changement de conformation d'une protéine au cours d'une simulation de DM. La courbe des RMSFs (calculées sur les atomes de carbone α des 163 résidus de hVKORC1) indique de faibles fluctuations atomiques au niveau des quatre TMs et de HH (aux alentours d'1 Å).

En revanche, les fluctuations atomiques de la boucle L augmentent considérablement par rapport à la simulation de 100 ns (**Figure 3.27.B**). Un premier pic de fluctuations dépassant 6 Å est observé pour les résidus 40 à 50 (leurs fluctuations restaient inférieures à 5 Å durant la simulation de 100 ns). Le second pic concerne les résidus 60 à 70, dont les fluctuations s'élèvent à 9 Å alors qu'elles ne dépassaient pas 2 Å durant la simulation de 100 ns (**Figure 3.36.B**). Cette augmentation significative des fluctuations atomiques de la boucle L reflète l'évolution conformationnelle de hVKORC1_C₅₁-C₁₃₂ au cours de la simulation prolongée jusqu'à 500 ns. Le repliement des quatre TMs et de HH de hVKORC1_C₅₁-C₁₃₂ est totalement préservé au cours de la simulation de 500 ns, comme le montre l'interprétation des structures secondaires par DSSP (**Figure 3.36.C**). Ce maintien du repliement de la protéine démontre que les augmentations de déviations et fluctuations atomiques mises en évidence par les courbes des RMSDs et RMSFs sont bien dues à un changement de conformation de l'enzyme. Le regroupement des conformations similaires de hVKORC1_C₅₁-C₁₃₂ par *clustering* (basé sur les RMSDs calculées deux à deux sur toutes les conformations de la protéine obtenue durant la simulation de DM, voir **METHODES**) montre trois *clusters* significatifs au cours de la simulation de 500 ns. Le premier *cluster* représentatif est présent de 30 à 250 ns de simulation et regroupe 39% des conformations totales de l'enzyme. Le second *cluster* s'étend de 350 à 500 ns de simulation et comprend 31% des conformations totales de hVKORC1.

Enfin, un troisième *cluster* intermédiaire est observé entre 250 et 350 ns de simulation, contenant 12% des conformations totales de la protéine (**Figure 3.36.D**). Les deux premiers *clusters* (39% et 31% des conformations totales de hVKORC1, respectivement) correspondent aux deux plateaux observés sur la courbe des RMSDs, tandis que le *cluster* intermédiaire (12% des conformations totales) reflète la transition entre ces deux plateaux (**Figure 3.36.A**).

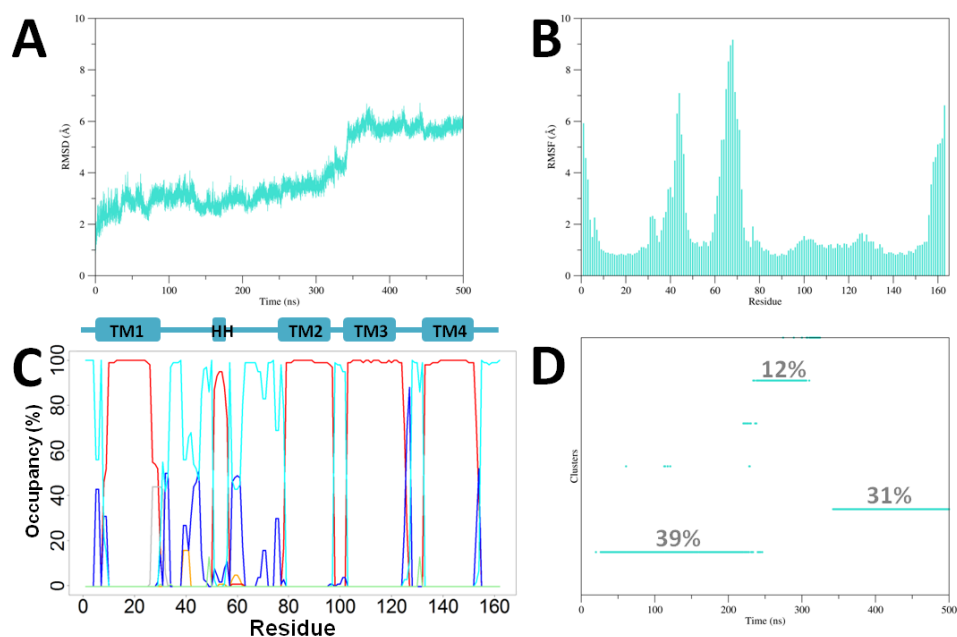


Figure 3.36. Caractérisation de la stabilité structurale globale du modèle hVKORC1_C51-C132 au cours de la simulation de DM étendue à 500 ns. (A) Courbes des RMSDs calculées sur les atomes de carbone α des résidus 9 à 151 de hVKORC1_C51-C132 (les résidus des extrémités N- et C-terminales ont été omis afin d'éliminer le bruit induit par leurs grandes fluctuations) au cours d'une simulation de DM de 500ns, après alignement des atomes de carbone α des hélices transmembranaires de chaque conformation sur les coordonnées de la conformation initiale ($t = 0$ ns). (B) Courbes des RMSFs de hVKORC1_C51-C132 au cours d'une simulation de DM de 500 ns. Les RMSFs ont été calculées sur les 163 atomes de carbone α de hVKORC1, préalablement alignés sur les coordonnées de la conformation initiale ($t = 0$ ns). (C) Interprétation par DSSP des structures secondaires des conformations de hVKORC1_C51-C132 au cours d'une simulation de DM de 500 ns (en pourcentage). Les repliements en hélice α , hélice_{3,10}, boucle, coude et brin β sont indiqués par les courbes en rouge, orange, cyan, bleu foncé et vert clair, respectivement. Les structures secondaires de la conformation initiale ($t = 0$ ns) sont indiquées au-dessus du graphes. (D) Analyse des conformations similaires (*clustering*) de hVKORC1_C51-C132 obtenues au cours d'une simulations de DM de 500 ns. Les conformations appartenant à un même *cluster* sont indiquées par des points, regroupés en lignes de niveau. La population de chaque *cluster* (proportion des conformations membres d'un même cluster par rapport au nombre de conformations total de la trajectoire) est indiquée en pourcentage, pour les *clusters* représentant plus de 5% des conformations totales. Le *clustering* est basé sur les RMSDs calculées deux à deux entre les atomes de carbone α des résidus 9-151 de chaque conformation d'une trajectoire. Deux conformations présentant une valeur de RMSD inférieure à 2 Å sont considérées comme membres d'un même *cluster*. Le *cluster* numéro 1 est toujours le plus peuplé, et chaque conformation appartient à un seul *cluster*.

La conformation représentative du premier *cluster* a déjà été identifiée durant la première simulation de 100 ns du modèle hVKORC1_C51-C132 et utilisée comme cible pour le *docking* de la vitK^E. Nous avons donc effectué une détection de poche à la surface de toutes les conformations de hVKORC1_C51-C132 composant le second *cluster* par MDpocket (voir **METHODES**). Un rapprochement entre TM1 et TM4 ainsi que TM1 et HH est observé pour cette seconde conformation représentative de hVKORC1_C51-C132, modifiant la géométrie de la cavité détectée. La poche identifiée présente un volume moyen inférieur à celui de la poche présente à la surface de la première conformation de hVKORC1_C51-C132 (environ 650 Å³, contre plus de 800 Å³ auparavant, **Figure 3.37**). Cette réduction du volume de la poche est susceptible de réduire l'espace des poses possibles de la vitK^E dans la cavité de hVKORC1_C51-C132 lors d'un amarrage moléculaire. Nous avons donc procédé à un nouveau *docking* moléculaire, en utilisant comme cible la seconde conformation représentative identifiée pour hVKORC1_C51-C132 et la cavité détectée à sa surface. 100 poses de la vitK^E sous forme tronquée ont été effectuées.

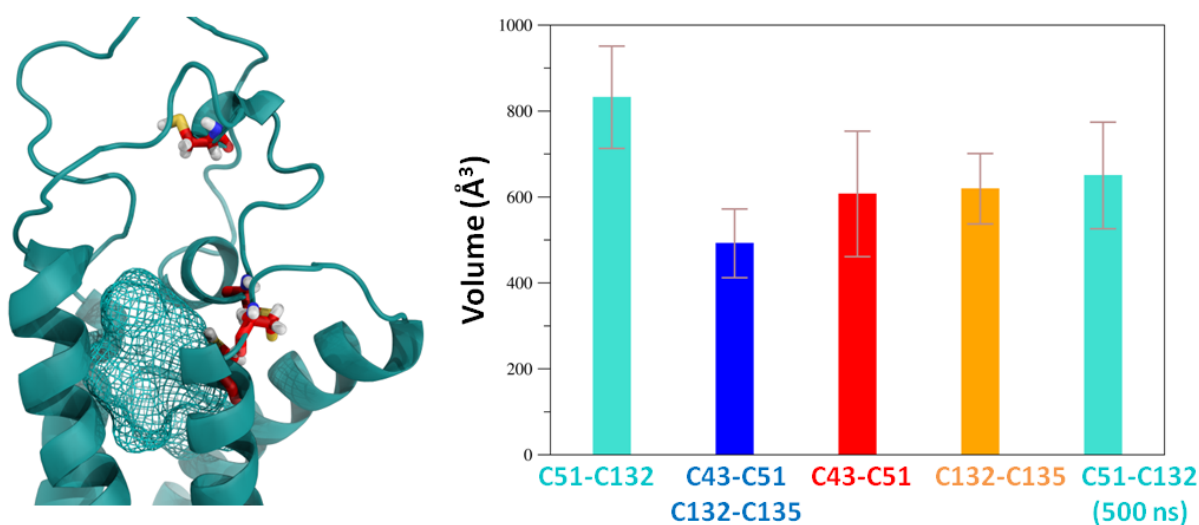


Figure 3.37. Détection de poches à la surface des structures 3D correspondant à la seconde conformation représentative de hVKORC1_C51-C132 obtenue par extension de la simulation de DM à 500 ns. La cavité détectée à la surface de la structure 3D de la seconde conformation représentative de hVKORC1_C51-C132 est représentée par un maillage (*mesh*) bleu clair. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge ; les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le volume moyen calculé pour la nouvelle cavité est comparé aux volumes moyens des cavités précédemment identifiées pour chaque état enzymatique de hVKORC1, représentés par un histogramme. Les écarts-types sont indiqués par une barre marron. L'état enzymatique de hVKORC1 associé à chaque cavité est indiqué sous le graphe.

Ce nouveau *docking* a généré 100 poses de la vitK^E totalement identiques, indiquant une seule conformation et un seul positionnement de la vitK^E dans la cavité de la nouvelle conformation de hVKORC1_C₅₁-C₁₃₂. La vitK^E est localisée à proximité du site actif de hVKORC1, son groupement époxyde est orienté vers le résidu C₁₃₅, lui-même orienté dans la cavité de l'enzyme. L'atome de carbone en position 3 de la vitK^E et l'atome de soufre du résidu C₁₃₅ sont ainsi distants de 4,1 Å. La pose obtenue est donc en parfaite adéquation avec les données expérimentales (**Figure 3.38**).

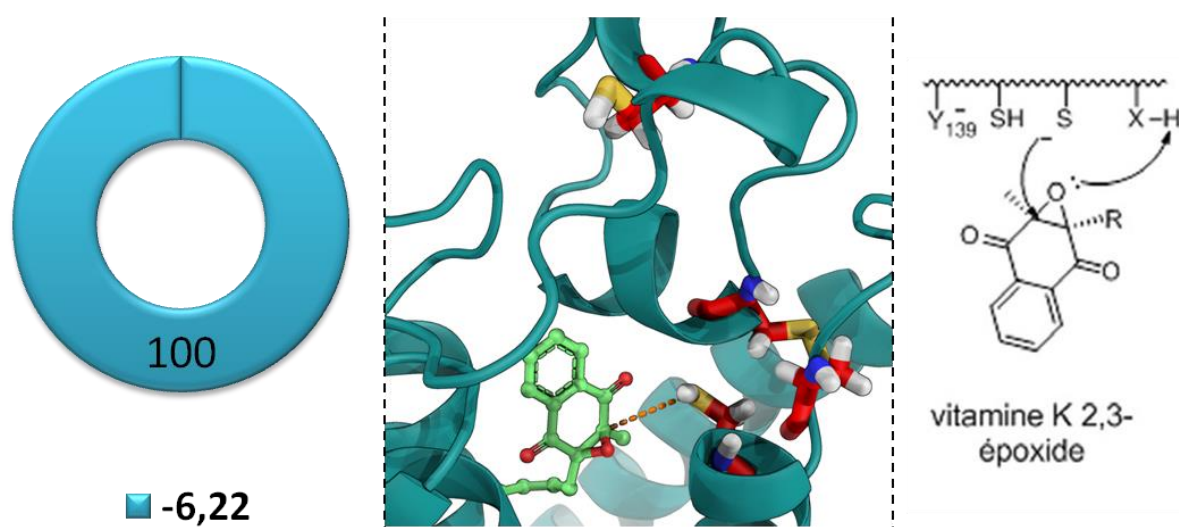


Figure 3.38. Amarrage moléculaire de la vitK^E tronquée sur la seconde conformation représentative du modèle hVKORC1_C₅₁-C₁₃₂. Les 100 poses identiques obtenues au cours du *docking* ont été regroupées en un seul groupe représenté par un diagramme. L'énergie libre de liaison associée à cette pose unique est indiquée sous le diagramme. La structure 3D du complexe hVKORC1•vitK^E obtenue est représentée. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D. Le schéma à droite correspond à l'association entre la vitK^E et hVKORC1 suggérée à partir des données biologiques [16].

Les cavités détectées à la surface de chacun des quatre états enzymatiques de hVKORC1 ont été utilisées comme cibles pour le *docking* de la vitK^E. Les résultats obtenus nous ont permis d'identifier le modèle hVKORC1_C₅₁-C₁₃₂ comme l'état enzymatique fixant sélectivement la vitamine K sous sa forme époxyde. Nous avons donc considéré cet état conformationnel comme point de départ du mécanisme enzymatique conduisant à la réduction de la vitK^E en vitK^Q.

V. Détermination par amarrage moléculaire du cycle enzymatique conduisant à la réduction de la vitK^E en vitK^Q par hVKORC1

L'état hVKORC1_C₅₁-C₁₃₂ a été précédemment identifié comme fixant la vitK^E et peut être considéré comme point de départ au cycle enzymatique provoquant une réduction de la vitK^E en vitK^Q. Cependant, deux possibilités existent concernant l'étape suivante de ce cycle catalytique.

Dans le premier cas, le résidu C₄₃ s'associerait avec le résidu C₅₁ pour former un pont disulfure, rompant le pont C₅₁-C₁₃₂ déjà existant ; le proton du résidu C₄₃ serait transféré au résidu C₁₃₂. L'enzyme hVKORC1 serait ainsi à l'état hVKORC1_C₄₃-C₅₁. Un *docking* de la vitK^E sur cet état enzymatique a déjà effectué (**Figure 3.34**), montrant l'existence de 4 poses distinctes pour lesquelles la vitK^E est toujours éloignée du site actif de hVKORC1. Les protons des résidus C₁₃₂ et C¹³⁵ seraient ensuite transférés un à un au groupement époxyde de la vitK^E, rompant successivement une des liaisons entre l'atome d'oxygène et les atomes de carbone en position 2 et 3.

Dans une seconde hypothèse, le proton du résidu C₁₃₅ serait d'abord transféré à l'atome d'oxygène du groupement époxyde, produisant une rupture de la liaison entre l'atome de carbone en position 2 ou 3 de la vitK^E et l'atome d'oxygène de son groupement époxyde. La vitamine K serait ainsi à sous forme hydroxyquinone (vitK^H). Le résidu C₄₃ formerait ensuite un pont disulfure avec C₅₁, rompant le pont C₅₁-C₁₃₂. Le proton du résidu C₄₃ serait transféré à C₁₃₂, puis au groupement hydroxyquinone de la vitK^H, achevant ainsi la réduction de la vitK^E en vitK^Q. La suite du cycle est commune aux deux hypothèses présentées ici. Une fois la vitK^E réduite en vitK^Q, les résidus C₁₃₂ et C₁₃₅ de hVKORC1 formeraient un pont disulfure, conduisant l'enzyme à adopter l'état hVKORC1_C₄₃-C₅₁-C₁₃₂-C₁₃₅. La vitK^Q peut être relâchée hors de hVKORC1 avant ou après la reformation du pont disulfure C₁₃₂-C₁₃₅ (**Figure 3.39**). Nous avons donc procédé à une nouvelle série de *dockings* afin d'identifier les étapes les plus probables du cycle enzymatique conduisant à la réduction de la vitK^E en vitK^Q et de définir l'ordre de rupture des deux liaisons du groupement époxyde. La sélectivité entre les différents modèles de hVKORC1 et les deux formes possibles de vitK^H a ainsi été évaluée. La notation vitK^{H2} correspond à une rupture de la liaison entre l'atome d'oxygène et l'atome de carbone en position 3, vitK^{H3} indique une rupture de la liaison entre l'atome d'oxygène et l'atome de carbone en position 2.

Nous avons donc effectué un nouveau *docking* de la vitK^{H2} ou vitK^{H3} sur hVKORC1_C51-C132, le résidu C₁₃₅ étant cette fois considéré à l'état déprotoné (S⁻). Cet état simule un transfert du proton du résidu C₁₃₅ au groupement époxyde de la vitamine K, juste après fixation de la vitK^E sur le modèle hVKORC1_C51-C132. Les résultats montrent une pose majoritaire de la vitK^{H2} (regroupant 64% des poses totales), identique à l'unique pose obtenue pour le *docking* de la vitK^E sur hVKORC1_C51-C132 (**Figure 3.38**). L'ELL correspondant à cette pose (-6,26 kcal/mol) est presque identique à celle de la seconde pose (-6,21 kcal/mol), qui regroupe 20% des poses totales et qui consiste en une rotation à 180° de la vitK^{H3} par rapport à la première pose, orientant le groupement hydroxyquinone à l'opposé du site actif. La distance entre l'atome de soufre du résidu C₁₃₅ et l'atome de carbone en position 3 de la vitK^{H2} est de 4,7 Å pour la pose majoritaire et de 4,9 Å pour la seconde pose. Les résultats concernant la vitK^{H3} montrent deux poses majoritaires, regroupant chacune 43 et 49% des poses totales, et dont les ELL sont très proches (-6,68 et -6,25 kcal/mol) ; il est donc difficile de distinguer la pose la plus pertinente (**Figure 3.40**). De plus, aucune des deux poses ne présente de positionnement de la vitK^{H3} similaire et celui de la vitK^E fixé à hVKORC1.

Les résultats de ce premier *docking* de la vitK^H sur ce modèle suggèrent donc un transfert du proton du résidu C₁₃₅ au groupement époxyde de la vitamine K immédiatement après fixation de la vitKE sur hVKORC1_C51-C132, rompant la liaison entre l'atome d'oxygène et l'atome de carbone en position 3 de la vitamine K.

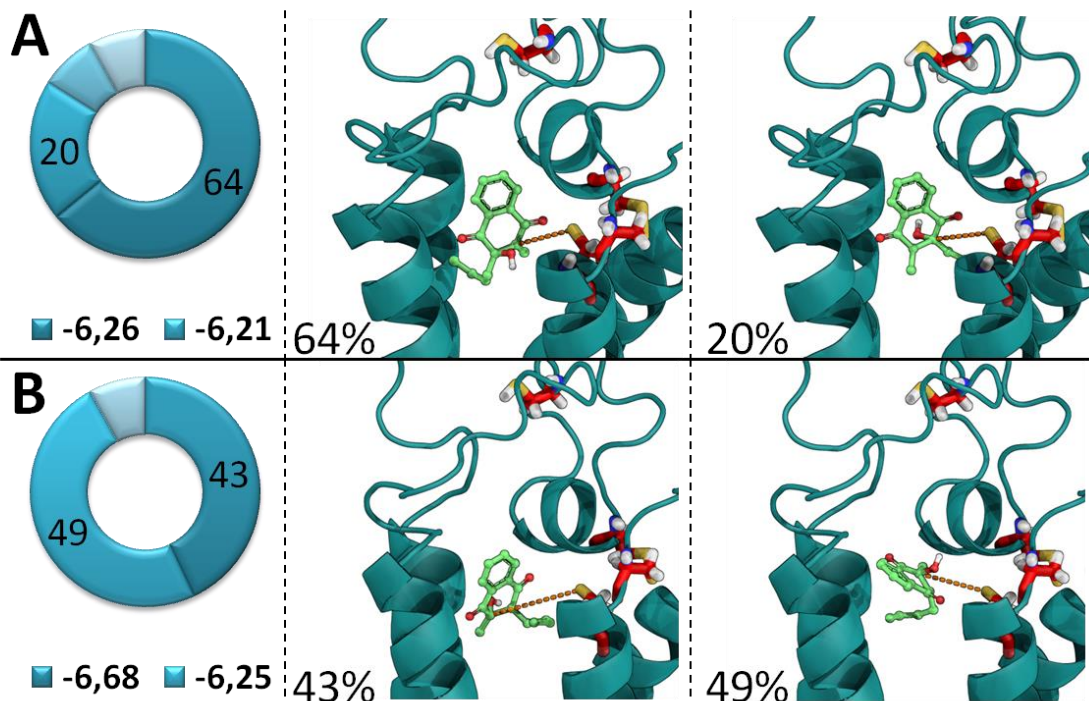


Figure 3.40. Amarrage moléculaire de la vitK^H tronquée sur le modèle hVKORC1_C51-C132, considérant le résidu C135 à l'état déprotoné (S⁻). (A) Résultats du *docking* de la vitK^{H2} tronquée sur hVKORC1_C51-C132_C135S⁻. (B) Résultats du *docking* de la vitK^{H3} tronquée sur hVKORC1_C51-C132_C135S⁻. Les poses similaires obtenues au cours de chaque *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe est indiqué pour les groupes les plus représentatifs (incluant plus de 10% des poses totales). Les énergies libres de liaison associées à chaque groupe de poses sont indiquées sous le diagramme. Les structures 3D des complexes hVKORC1•vitK^H correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^H est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

Un deuxième *docking* entre des vitK^{H2} ou vitK^{H3} sur hVKORC1_C43-C51 a été effectué, en considérant le résidu C135 à l'état déprotoné (S⁻). Ce nouvel amarrage moléculaire vise à simuler la formation d'un pont disulfure entre les résidus C₄₃ et C₅₁ et le transfert du proton de C₄₃ à C₁₃₂. Les résultats impliquant la vitK^{H2} montrent une pose majoritaire (76% des poses totales) et une seconde pose minoritaire (16% des poses totales) dont les ELL respectives sont de -6,09 et -5,64 kcal/mol. La distance entre l'atome de soufre du résidu C₁₃₅ et l'atome d'oxygène du groupement hydroxyquinone de la vitK^{H2} est de 13,1 Å pour la première pose et de 10,3 Å pour la seconde pose. Cette seconde pose, bien que minoritaire, pourrait correspondre à un pivotement de la vitK^{H2} par rapport à la position qu'elle occupait dans la cavité lors du *docking* précédent (**Figure 3.40**). Les résultats concernant la vitK^{H3} montrent trois poses constituant respectivement 39, 18 et 32% des poses totales et dont les ELL sont comprises entre -7,07 et -5,90 kcal/mol. Le positionnement de la vitK^{H3} est systématiquement éloigné du site actif de hVKORC1, la distance minimale entre l'atome de soufre du résidu C₁₃₅ et l'atome de carbone en position 3 de la vitK^{H3} est de 12,1 Å, pour la première pose (**Figure 3.41**).

Ces résultats renforcent l'hypothèse d'une rupture de la liaison entre les atomes de carbone en position 3 et d'oxygène du groupement époxyde de la vitamine K dans un premier temps.

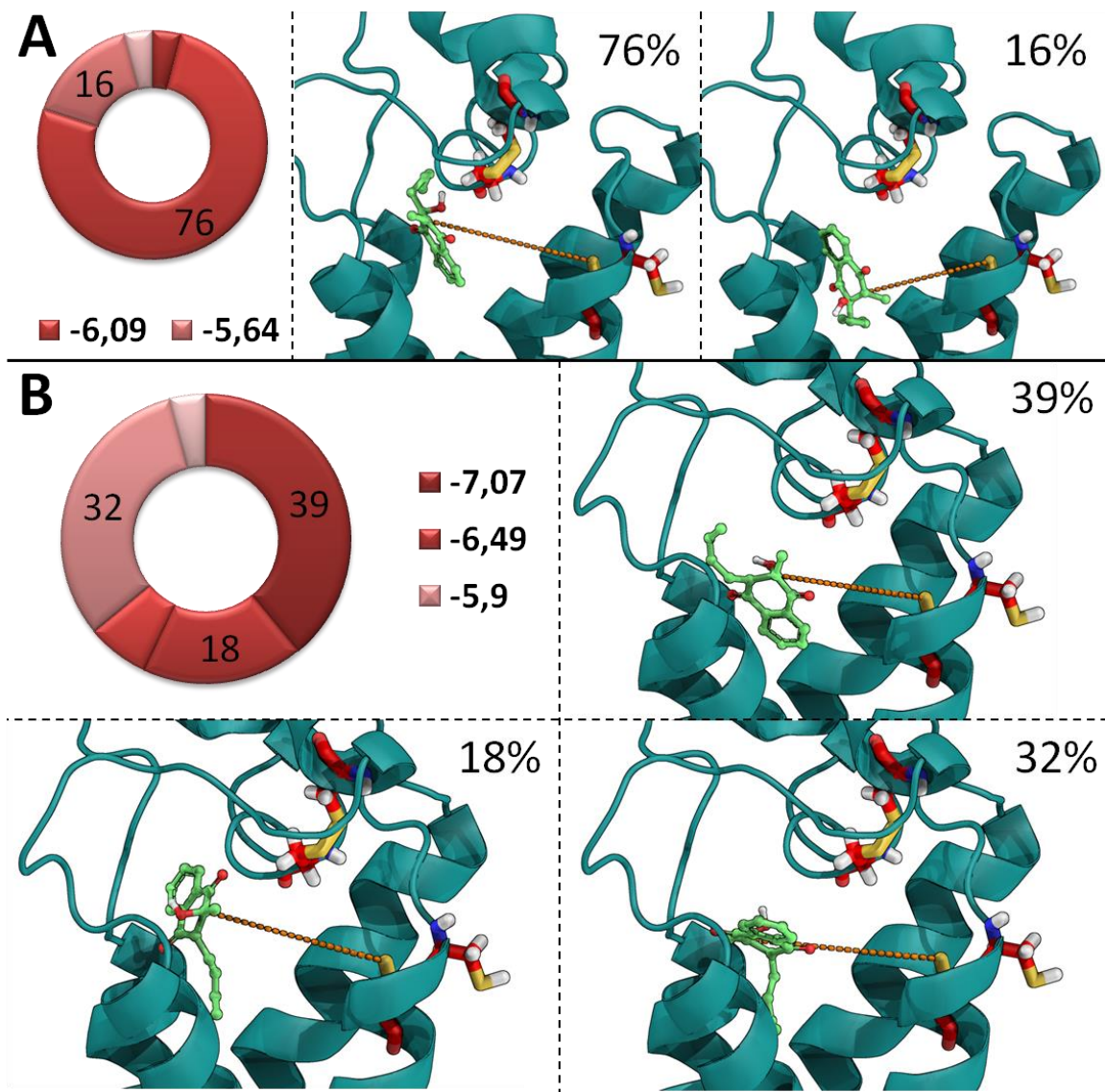
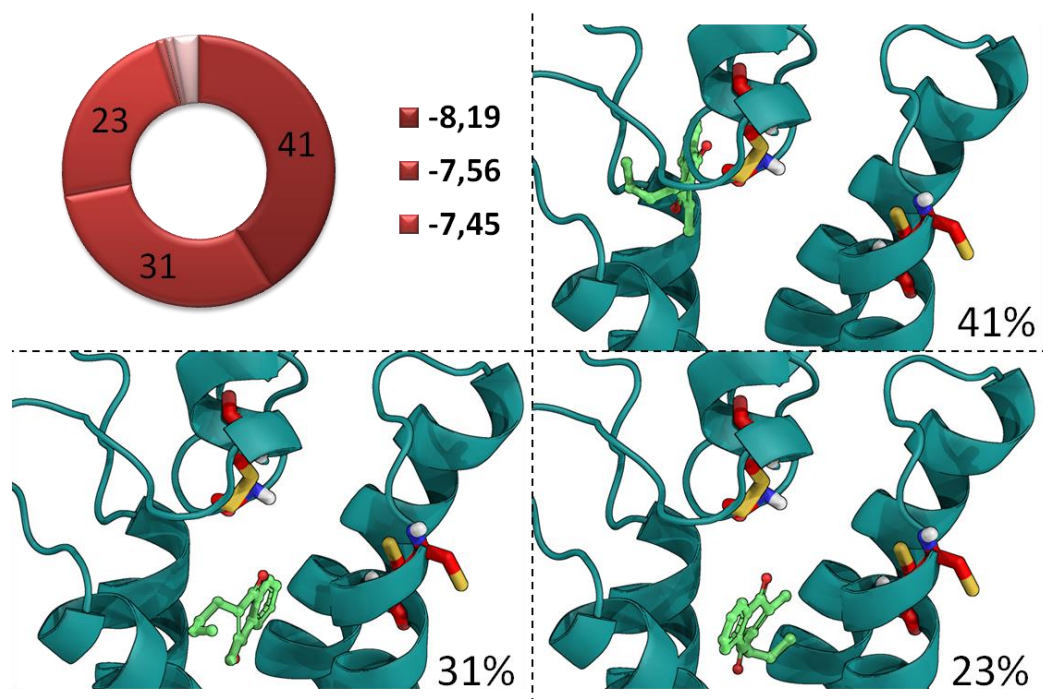


Figure 3.41. Amarrage moléculaire de la vitK^H tronquée sur le modèle hVKORC1_C43-C51, considérant le résidu C135 à l'état déprotoné (S⁻). (A) Résultats du *docking* de la vitK^{H2} tronquée sur hVKORC1_C43-C51_C135S⁻. (B) Résultats du *docking* de la vitK^{H3} tronquée sur hVKORC1_C43-C51_C135S⁻. Les poses similaires obtenues au cours de chaque *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe est indiqué pour les groupes les plus représentatifs (incluant plus de 10% des poses totales). Les énergies libres de liaison associées à chaque groupe de poses sont indiquées sous le diagramme. Les structures 3D des complexes hVKORC1•vitK^H correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^H est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

Nous avons ensuite étudié la sélectivité des états hVKORC1_C43-C51 (résidus C132 et C135 à l'état S⁻) et hVKORC1_C43-C51_C132-C135 vis-à-vis de la vitK^Q, afin de déterminer si la molécule pouvait se fixer sur ces états conformationnels de hVKORC1. Nous avons donc réalisé un *docking* de la vitK^Q sur hVKORC1_C43-C51, en considérant cette fois les deux résidus C132 et C135 à l'état déprotoné (S⁻). Cet amarrage moléculaire simule le transfert du proton du résidu C132 au groupement hydroxyquinone de vitK^H, donc sa réduction totale en vitK^Q et la formation d'une molécule d'eau (non prise en compte dans le *docking*). Les résultats indiquent trois poses majoritaires, regroupant respectivement 41%, 31% et 23% des poses totales et dont les ELL sont comprises entre -8,19 et -7,45 kcal/mol. La vitK^Q est systématiquement éloignée du site actif de hVKORC1 (**Figure 3.42**). Ces résultats suggèrent que la vitamine K serait relâchée hors de hVKORC1 dès sa réduction totale en vitK^Q.



En revanche, les résultats de l'amarrage moléculaire de la vitK^Q sur le modèle hVKORC1_C43-C51_C132-C135 révèlent l'existence d'une pose majoritaire (comprenant 60% des poses totales et dont l'ELL est de -7,04 kcal/mol). Deux autres poses sont observées, dont les ELL sont quasiment identiques (-7,01 et -6,93 kcal/mol) et qui constituent chacune 17% et 19% des poses totales. La vitK^Q est positionnée à proximité du site actif de hVKORC1 pour les deux premières poses (**Figure 3.43**), qui pourraient à nouveau correspondre à un pivotement de la molécule par rapport à son positionnement dans la cavité des modèles hVKORC1_C51-C132 et hVKORC1_C43-C51. Ces derniers résultats suggèrent donc l'impossibilité pour la vitK^Q de se fixer sur hVKORC1 une fois le pont disulfure entre les résidus C₁₃₂ et C₁₃₅ formé. De plus, la coexistence des deux résidus C₁₃₂ et C₁₃₅ à l'état déprotoné (S⁻) est très peu stable chimiquement. Leur association pour former un pont disulfure C₁₃₂-C₁₃₅ doit donc s'effectuer très rapidement. Nous avons ainsi considéré que la vitK^Q restait liée à hVKORC1 jusqu'à la formation du pont disulfure C₁₃₂-C₁₃₅, hypothèse renforcée par les résultats du dernier *docking*.

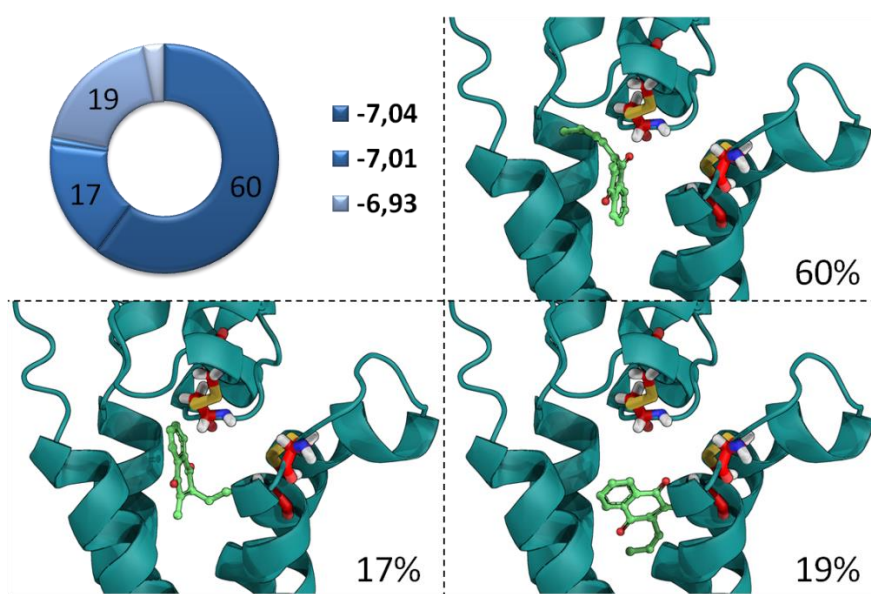


Figure 3.43. Amarrage moléculaire de la vitK^Q tronquée sur le modèle hVKORC1_C43-C51_C132-C135. Les poses similaires obtenues au cours de chaque *docking* sont regroupées et représentées par un diagramme. Le pourcentage de poses contenues dans chaque groupe est indiqué pour les groupes les plus représentatifs (incluant plus de 10% des poses totales). Les énergies libres de liaison associées à chaque groupe de poses sont indiquées sous le diagramme. Les structures 3D des complexes hVKORC1•vitK^Q correspondant aux poses les plus significatives sont représentées. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. La structure 3D de la vitK^Q est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. Le pourcentage de poses correspondant à chaque complexe est indiqué à côté des structures 3D.

L'ensemble des *dockings* effectués entre les différents états enzymatiques de hVKORC1 et les formes vitK^E, vitK^H et vitK^Q de la vitamine K nous a finalement permis de postuler le mécanisme enzymatique détaillé de la réduction de la vitK^E en vitK^Q. Nous avons identifié l'état hVKORC1_C₅₁-C₁₃₂ comme fixant la vitK^E. Le proton du résidu C₁₃₅ est ensuite transféré au groupement époxyde de la vitamine K, la réduisant en vitK^{H2} (les résultats de *docking* suggèrent en effet une rupture entre l'atome d'oxygène du groupement époxyde et l'atome de carbone en position 3 de la vitamine K). Puis, les résidus C₄₃ et C₅₁ s'associent pour former un pont disulfure C₄₃-C₅₁, rompant le pont C₅₁-C₁₃₂ existant. Le proton du résidu C₄₃ est transféré au résidu C₁₃₂, puis au groupement hydroxyquinone de la vitK^{H2}, la réduisant totalement en vitK^Q et provoquant la formation d'une molécule d'eau. Les résidus C₁₃₂ et C₁₃₅, à l'état déprotoné S⁻, s'associent alors pour former un pont disulfure C₁₃₂-C₁₃₅. La vitK^Q est finalement relâchée hors de hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅. L'apport de deux protons est ensuite nécessaire pour rompre le pont disulfure C₄₃-C₅₁, conduisant la protéine à l'état hVKORC1_C₁₃₂-C₁₃₅. L'association des résidus C₅₁ et C₁₃₂ permet à l'enzyme d'adopter à nouveau l'état actif hVKORC1_C₅₁-C₁₃₂ fixant la vitK^E. Le cycle enzymatique ainsi défini est présenté dans la **Figure 3.44**.

Les deux protons nécessaires à la réactivation de hVKORC1 une fois le cycle enzymatique effectué sont apportés par le milieu extérieur. Des données biologiques suggèrent l'intervention d'un partenaire protéique de la famille des thiorédoxines [29], fournissant à hVKORC1 les deux protons nécessaires à la rupture du pont disulfure C₄₃-C₅₁.

En revanche, un biais existe dans le *docking* réalisé. En effet, les conformations de hVKORC1_C₄₃-C₅₁ et hVKORC1_C₄₃-C₅₁_C₁₃₂-C₁₃₅ sur lesquelles sont arrimées la vitK^H et la vitK^Q sont extraites d'une simulation de DM effectuée sur l'enzyme seule. La cavité de hVKORC1 ne s'est donc pas ajustée à la vitamine K au cours de cette simulation de DM. Or, dans la réalité biologique, une fois la vitK^E arrimée à hVKORC1_C₅₁-C₁₃₂, la cavité de la protéine subit l'influence structurale de la vitamine K, même lors des changements d'états conformationnels.

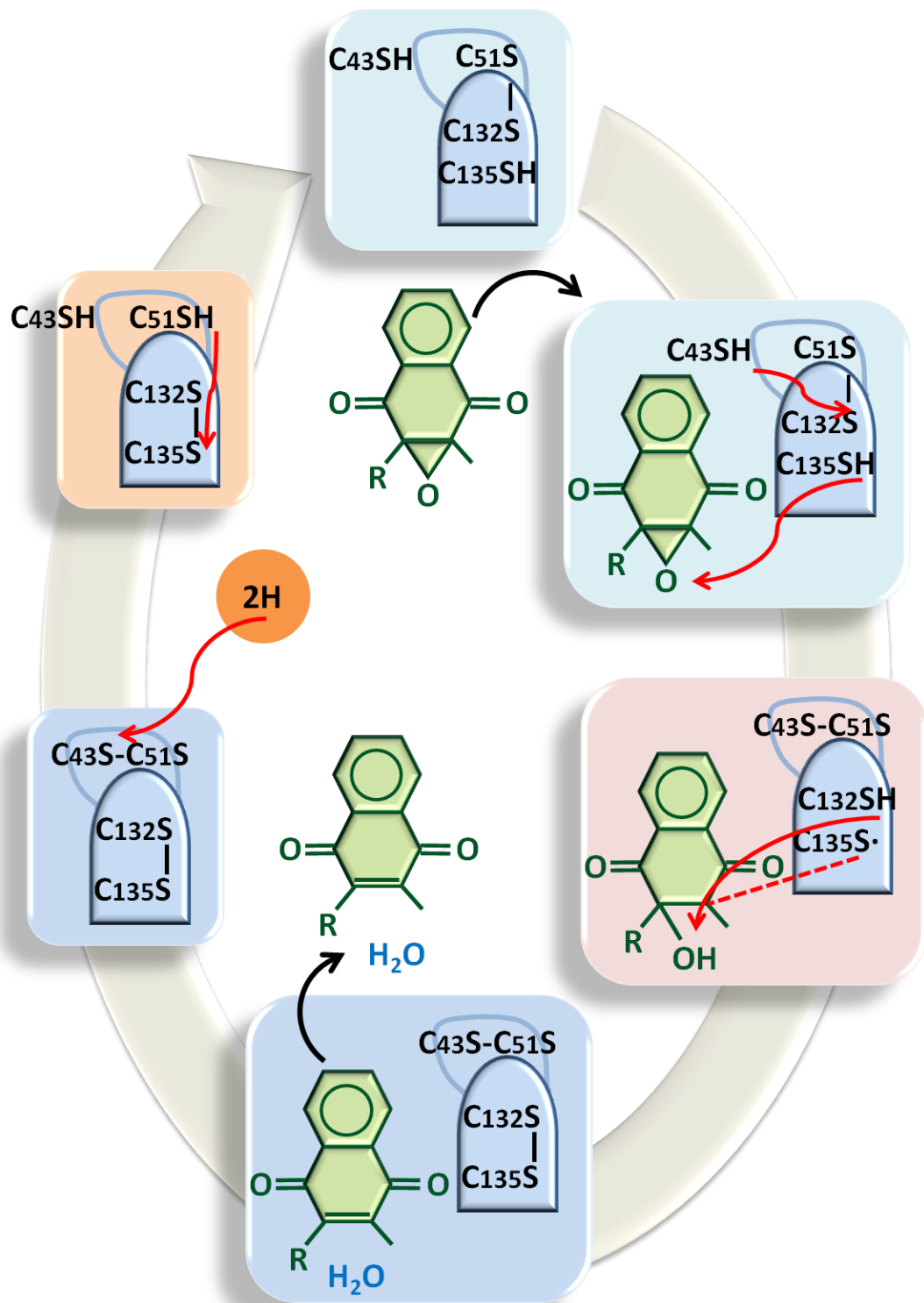


Figure 3.44. Cycle catalytique conduisant à la réduction de la vitK^E en vitK^Q par hVKORC1 postulé à partir des résultats de *docking* moléculaire. Les différentes étapes du cycle catalytique de hVKORC1, prédites à partir des résultats de *docking* moléculaire, sont indiquées. La structure de la vitamine K et l'état enzymatique de hVKORC1 associés à chaque étape sont représentés. La structure de la vitamine K est représentée en vert ; hVKORC1 est schématisée en bleu. Les flèches rouges indiquent les transferts de protons entre les résidus C₄₃, C₅₁, C₁₃₂, C₁₃₅ de hVKORC1 et le groupement époxyde ou hydroxyquinone de la vitamine K. Les flèches noires indiquent les étapes de fixation de la vitamine K sur hVKORC1 et de relargage de la molécule hors de l'enzyme.

Cette subtilité peut-être à l'origine de la variété de poses obtenues pour certains *dockings*, et du positionnement légèrement variable de la vitamine K d'un état enzymatique à l'autre. Néanmoins, l'obtention d'une pose unique de la vitK^E sur hVKORC1_C51-C132, en parfait accord avec les données biologiques [16], constitue une base solide pour la première étape du mécanisme de réduction de la vitK^E en vitK^Q. Nos résultats présentent en outre une tendance nette, nous permettant de postuler le cycle enzymatique de hVKORC1 malgré le biais mentionné précédemment.

VI. Etude des interactions au sein de complexes hVKORC1•Ligand

1. Etude des interactions entre hVKORC1 et la vitamine K époxyde

Nous avons précédemment obtenu un complexe entre la vitK^E et hVKORC1_C51-C132 par *docking* moléculaire. Cependant, nous avons considéré la vitK^E sous sa forme tronquée, et la différence d'ELL entre les poses obtenues pour les vitK^E intégrale et tronquée suggérait la participation de la chaîne hydrophobe à la stabilisation des interactions (**Figure 3.22**). De plus, le complexe résultant constitue une représentation figée des interactions entre la molécule et hVKORC1, insuffisant pour définir efficacement les résidus de hVKORC1 interagissant avec la vitK^E. Nous avons donc reconstitué l'intégralité de la chaîne hydrophobe de la vitK^E, en ajoutant successivement des groupements CH₂-CH₃ au niveau du dernier atome de carbone composant cette chaîne. L'énergie potentielle du complexe inséré dans une membrane de DLPC et solvaté a été minimisée à chaque étape, afin de réajuster convenablement la chaîne hydrophobe de la vitK^E dans la cavité de hVKORC1_C51-C132. La chaîne hydrophobe de la vitamine K intégralement reconstituée présente une conformation très étirée et traverse entièrement la cavité de hVKORC1 (**Figure 3.45**).

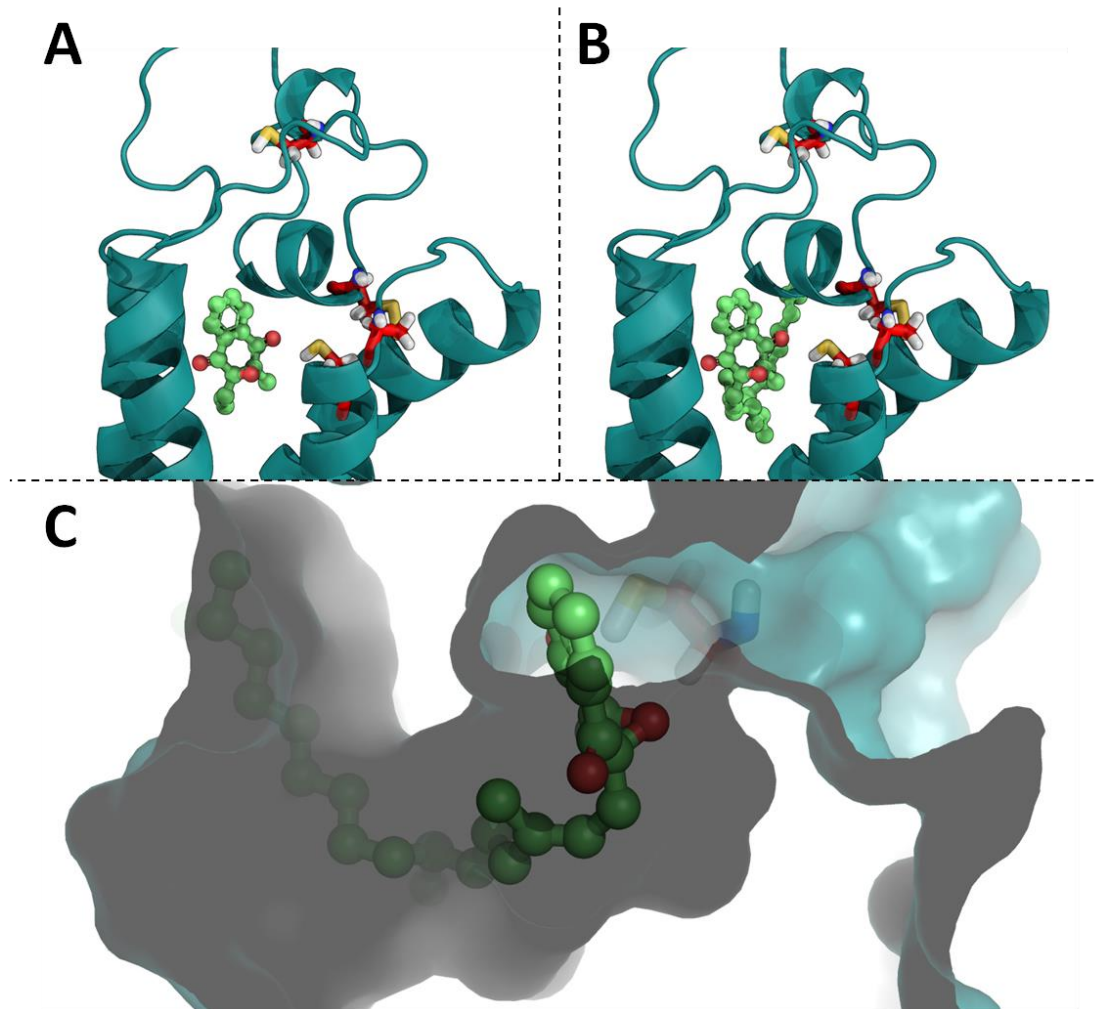


Figure 3.45. Reconstitution de la chaîne carbonée intégrale de la vitKE à partir de la pose de *docking* unique obtenue sur le modèle hVKORC1_C51-C132. (A) Représentation de la structure du complexe hVKORC1•vitK^E tronquée obtenue par *docking* moléculaire. (B) Représentation de la structure du complexe hVKORC1•vitK^E intégrale après reconstitution de la chaîne carbonée de la vitK^E. La structure 3D de hVKORC1_C51-C132 est représentée en *cartoon* et colorée en bleu ; les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* et colorés en rouge. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement. (C) Représentation en coupe du complexe hVKORC1•vitK^E intégrale. La surface de hVKORC1 est représentée en transparence et colorée en bleu. La structure 3D de la vitK^E est représentée en *sticks* verts ; les atomes de carbone et d'oxygène sont représentés par des sphères. Le résidu C₁₃₅ est représenté en *sticks* et coloré en rouge. Les atomes de soufre, d'oxygène, d'azote et d'hydrogène sont colorés en jaune, rouge clair, bleu et blanc, respectivement.

Deux simulations de 100 ns chacune du complexe formé par la vitK^E (intégrale) et hVKORC1_C₅₁-C₁₃₂ ont ensuite été effectuées. L'ELL entre les deux molécules a ensuite été estimé par la méthode MM/PBSA (*Molecular Mechanics/Poisson-Boltzmann Surface Area*, voir **METHODES**). Les ELL respectives calculées pour la première et la deuxième simulation sont de $-45,02 \pm 3,65$ kcal/mol et de $-40,22 \pm 4,03$ kcal/mol. La reproductibilité des résultats nous a permis de concaténer les deux trajectoires pour les traiter comme une seule simulation de 2x100 ns. L'ELL calculée sur les trajectoires concaténées est de $-45,03 \pm 3,65$ kcal/mol. La contribution de chaque résidu de hVKORC1 à l'ELL totale a ensuite été déterminée. Les résidus L₂₂, A₂₆, F₅₅, L₇₆, F₈₃, F₈₇, V₁₁₂, A₁₁₅, G₁₁₆, L₁₂₀ et V₁₃₄ favorisent le plus les interactions entre la vitK^E et hVKORC1 (**Figure 3.46**).

Le résidu F₅₅ présente la plus forte contribution à l'ELL ($-2,25$ kcal/mol) et stabilise le fragment cyclique de la vitK^E ; les autres résidus interagissent avec la chaîne hydrophobe. Les résidus N₈₀, T₁₃₈ et N₁₄₂ sont les moins favorables à la fixation de la vitK^E sur hVKORC1_C₅₁-C₁₃₂. Les interactions de type hydrophobe sont donc prédominantes dans la fixation de la vitamine K sur hVKORC1.

Le protocole utilisé a ainsi abouti à l'identification des interactions entre hVKORC1 et son ligand naturel, la vitK^E. Nous avons donc appliqué la même méthode à l'étude des interactions entre hVKORC1 et plusieurs inhibiteurs (AVKs). La conformation de hVKORC1_C₅₁-C₁₃₂ utilisée pour le docking de la vitK^E a été utilisée comme cible pour le *docking* des AVKs.

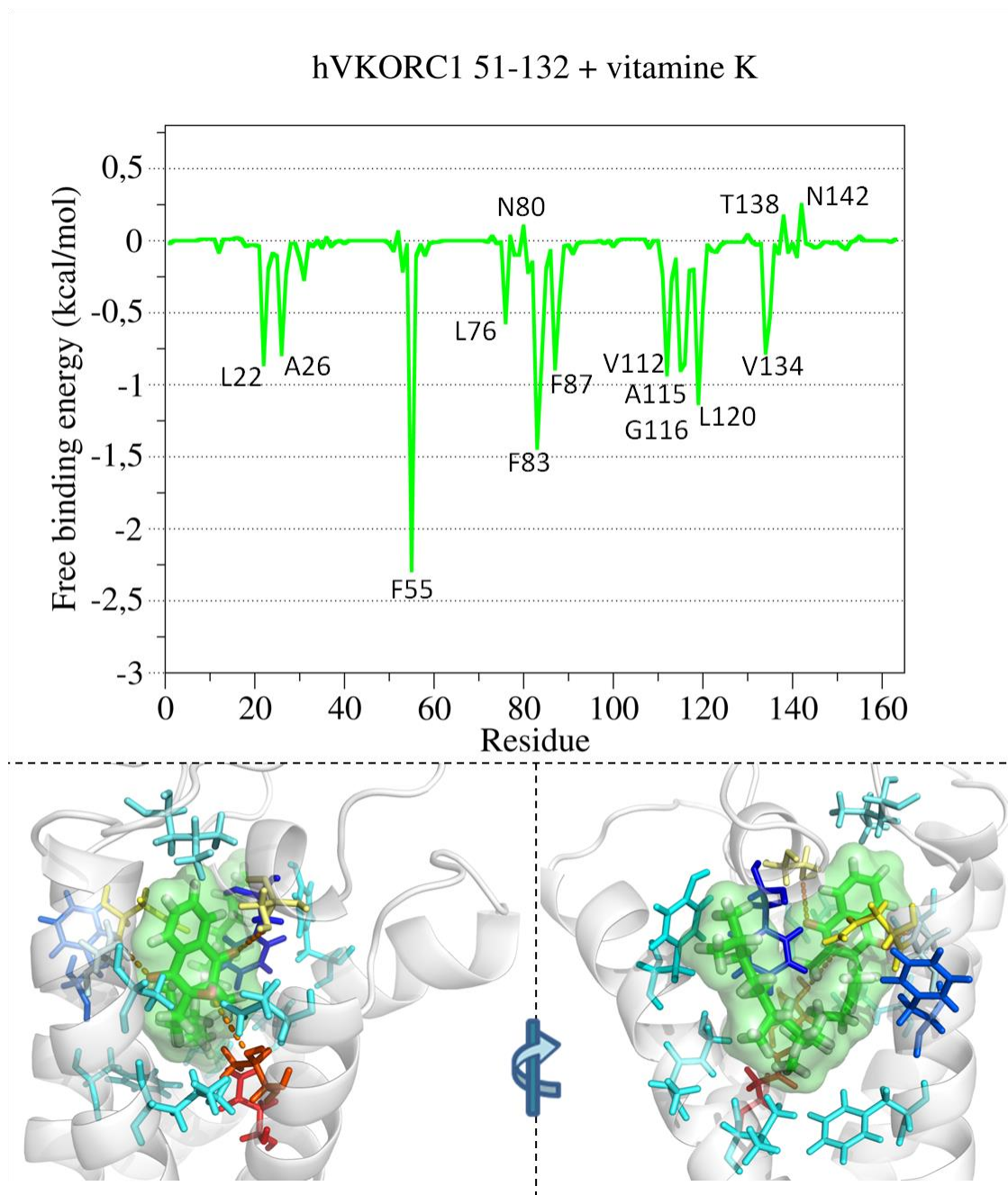


Figure 3.46. Détermination de la contribution des résidus de hVKORC1_C₅₁-C₁₃₂ dans les interactions de l'enzyme avec la vitK^E au cours de deux simulations de DM de 100 ns concaténées. Le graphe représente la contribution de chaque résidu de hVKORC1 à l'énergie libre de liaison totale entre l'enzyme et la vitK^E. La contribution calculée pour chaque résidu est ensuite reportée sur la structure 3D du complexe hVKORC1•vitK^E. Les résidus de hVKORC1 interagissant avec la vitK^E sont représentés en *sticks* et colorés selon leur contribution, du bleu foncé (résidu les plus favorables à la fixation de la vitK^E) au rouge (résidu les moins favorables à la fixation de la vitK^E). La structure 3D de la vitK^E est représentée en *sticks*, sa surface en vert transparent ; les atomes de carbone, d'oxygène et d'hydrogène sont colorés en vert, rouge et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* blanc transparent.

2. Etude des interactions entre hVKORC1 et trois anti-vitamine K (AVKs)

Le *docking* de trois inhibiteurs sur hVKORC1_C₅₁-C₁₃₂ a abouti à l'obtention d'une pose unique pour chacun des complexes générés. Les AVKs concernés sont l'acénocoumarol, la phenprocoumone et le difénacoum. Les poses obtenues pour chacun de ces AVKs indiquent un positionnement à proximité du site actif de hVKORC1, masquant ainsi le résidu C₁₃₅ et bloquant ainsi l'accès de la vitK^E (Figure 3.47).

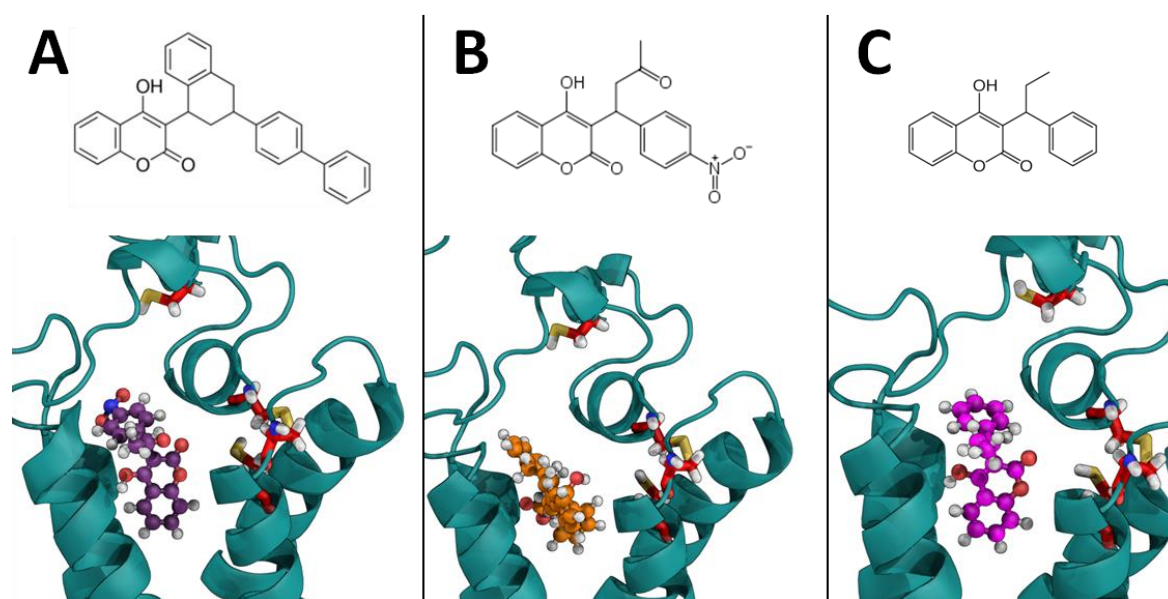


Figure 3.47. Complexes hVKORC1_C₅₁-C₁₃₂•AVK générés par *docking* moléculaire et pour lesquels une pose unique a été obtenue. (A) Complexe hVKORC1•acénocoumarol. (B) Complexe hVKORC1•difénacoum. (C) Complexe hVKORC1•phenprocoumone. La structure chimique de chaque AVK est indiquée en haut. La structure 3D des complexes obtenus par *docking* moléculaire est représentée en bas. La structure 3D de hVKORC1 est représentée en *cartoon* et colorée en bleu. Les résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ sont représentés en *sticks* rouges. La structure 3D de chaque AVK est représentée en *sticks* et en sphères. Les atomes d'oxygène, d'azote, de soufre et d'hydrogène sont colorés en rouge, bleu, jaune et blanc, respectivement. Les atomes de carbone de l'acénocoumarol, du difénacoum et de la phenprocoumone sont colorés en violet, orange et magenta, respectivement.

Des simulations de DM (deux répliques de 100 ns chacun) ont ensuite été effectuées sur chacun des trois complexes hVKORC1_C₅₁-C₁₃₂•AVK obtenus. Les deux trajectoires générées pour chaque complexe ont ensuite été concaténées et traitées comme une seule simulation de 2x100 ns afin de déterminer les ELL de hVKORC1 avec chaque AVK. L'acénocoumarol, le difénacoum et la phenprocoumone présentent des ELL respectives de $-25,50 \pm 3,30$, $-31,67 \pm 3,77$ et $-24,98 \pm 3,17$ kcal/mol (**Figure 3.48**). Ces résultats sont en bonne adéquation avec les données biologiques, puisqu'ils présentent un coefficient de corrélation de 0,66 avec les constantes d'inhibition mesurées par les biologistes de l'Ecole Vétérinaire de Lyon pour l'acénocoumarol, le difénacoum et la phenprocoumone ($0,33 \pm 0,18$, $0,12 \pm 0,03$ et $0,72 \pm 0,31$, respectivement, **Figure 3.49**). Cette adéquation entre nos prédictions et les données biologiques valident ainsi les complexes hVKORC1•AVK obtenus. Nous avons donc étudiée la contribution des résidus de hVKORC1 à la fixation des AVKs.

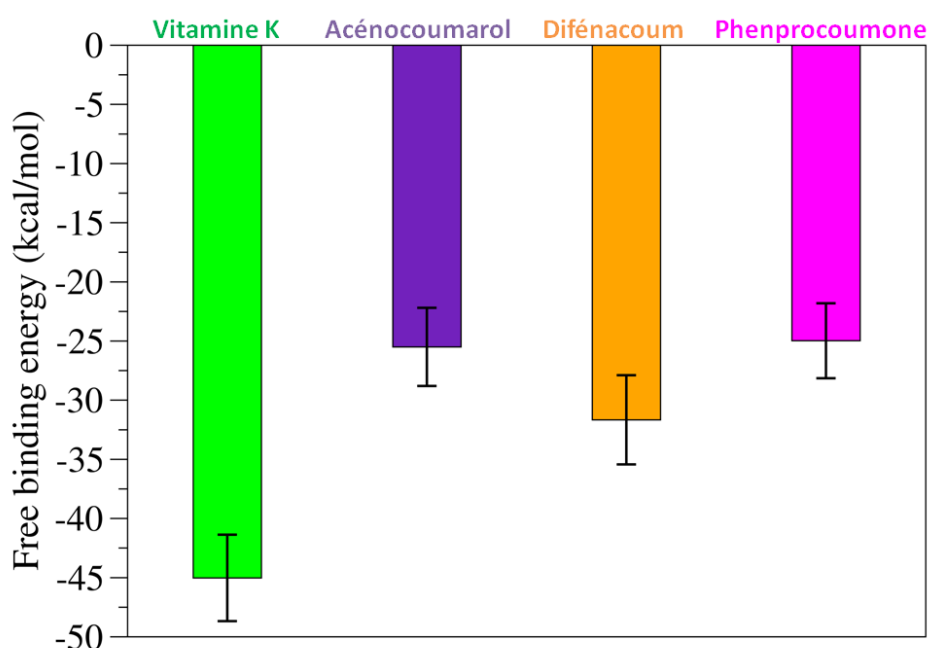


Figure 3.48. Energies libres de liaison des complexes hVKORC1•vitKE et hVKORC1•AVK calculées au cours de simulations de DM de 100 ns concaténées. L'énergie libre de liaison de quatre complexes hVKORC1•vitKE, hVKORC1•acénocoumarol, hVKORC1•difénacoum et hVKORC1•phenprocoumone a été calculée au cours de deux simulations de 100 ns concaténées, pour chaque complexe. Les résultats sont présentés sous forme d'histogramme, les écarts-types sont indiqués par une barre noire. Le ligand associé à hVKORC1 pour chaque complexe étudié est indiqué au-dessus du graphe.

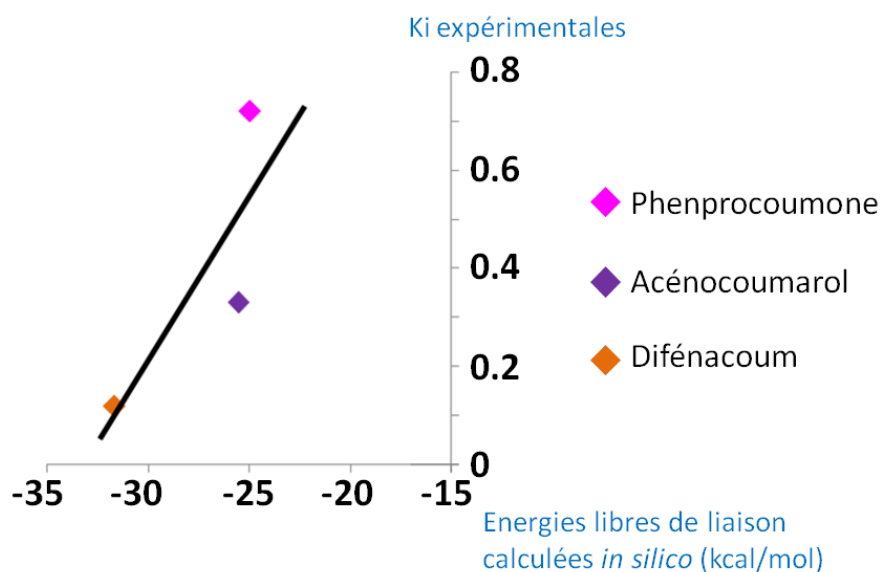


Figure 3.49. Comparaison entre les énergies libres de liaison calculées et les constantes d'inhibition mesurées expérimentalement pour trois hVKORC1•AVK.

Les résidus A₂₆, K₃₀, R₃₇, F₅₅, L₇₆, Q₇₈, L₁₂₀ et V₁₃₄ sont les plus favorables à la fixation de l'acénocoumarol sur hVKORC1. Les résidus F₅₅ et V₁₃₄ présentent les plus fortes contributions à l'ELL totale (-2,75 et -2,00 kcal/mol, respectivement). Les résidus D₃₆, D₃₈, D₇₃ et S₈₁ présentent les interactions avec cet inhibiteur les moins favorables (**Figure 3.50**).

Les résultats concernant le difénacoum indiquent une contribution favorable des résidus A₂₆, K₃₀, F₅₅, L₇₆ et V₁₃₄ à la fixation de l'AVK. Les résidus F₅₅ et V₁₃₄ présentent les interactions les plus favorables (ELL égale à -2,00 kcal/mol). Les résidus D₃₈ et N₇₇ sont les moins favorables à la fixation du difénacoum sur hVKORC1 (**Figure 3.51**).

Les résidus A₂₆, K₃₀, F₅₅, L₇₆ et V₁₃₄ sont les plus favorables à la fixation de la phenprocoumon sur hVKORC1. Le résidu F₅₅ présente la plus forte contribution à l'ELL totale (-2,50 kcal/mol). Les résidus S₂₃, N₇₇, S₈₁ et T₁₃₈ sont les moins favorables à la fixation de la phenprocoumon sur hVKORC1 (**Figure 3.52**).

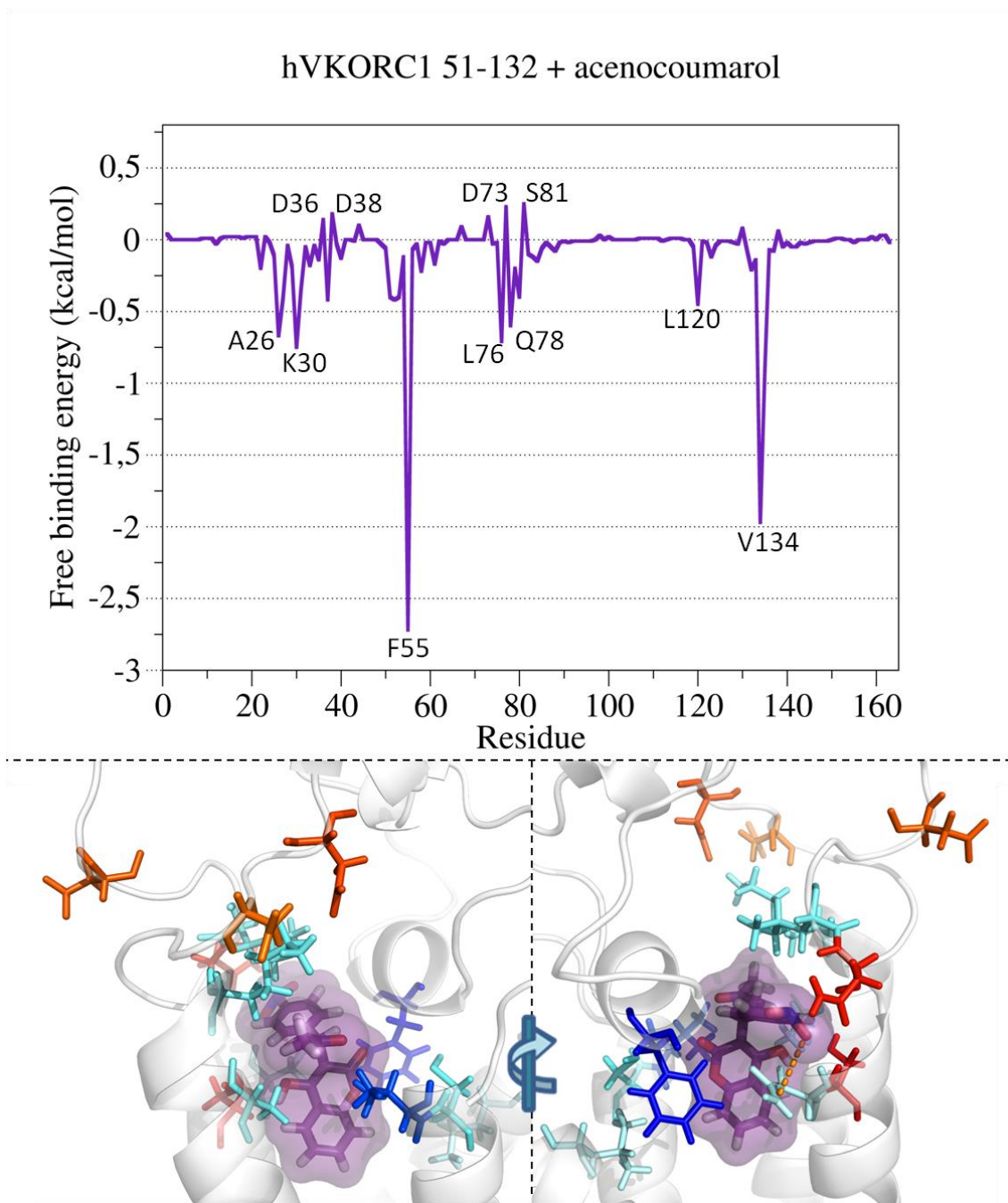


Figure 3.50. Détermination de la contribution des résidus de hVKORC1_{C51-C132} dans les interactions de l'enzyme avec l'acénocoumarol au cours de deux simulations de DM de 100 ns concaténées. Le graphe représente la contribution de chaque résidu de hVKORC1 à l'énergie libre de liaison totale entre l'enzyme et l'acénocoumarol. La contribution calculée pour chaque résidu est ensuite reportée sur la structure 3D du complexe hVKORC1•acénocoumarol. Les résidus de hVKORC1 interagissant avec l'acénocoumarol sont représentés en *sticks* et colorés selon leur contribution, du bleu foncé (résidu les plus favorables à la fixation de l'acénocoumarol) au rouge (résidu les moins favorables à la fixation de l'acénocoumarol). La structure 3D de l'acénocoumarol est représentée en *sticks*, sa surface en violet transparent ; les atomes de carbone, d'oxygène et d'hydrogène sont colorés en violet, rouge et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* blanc transparent.

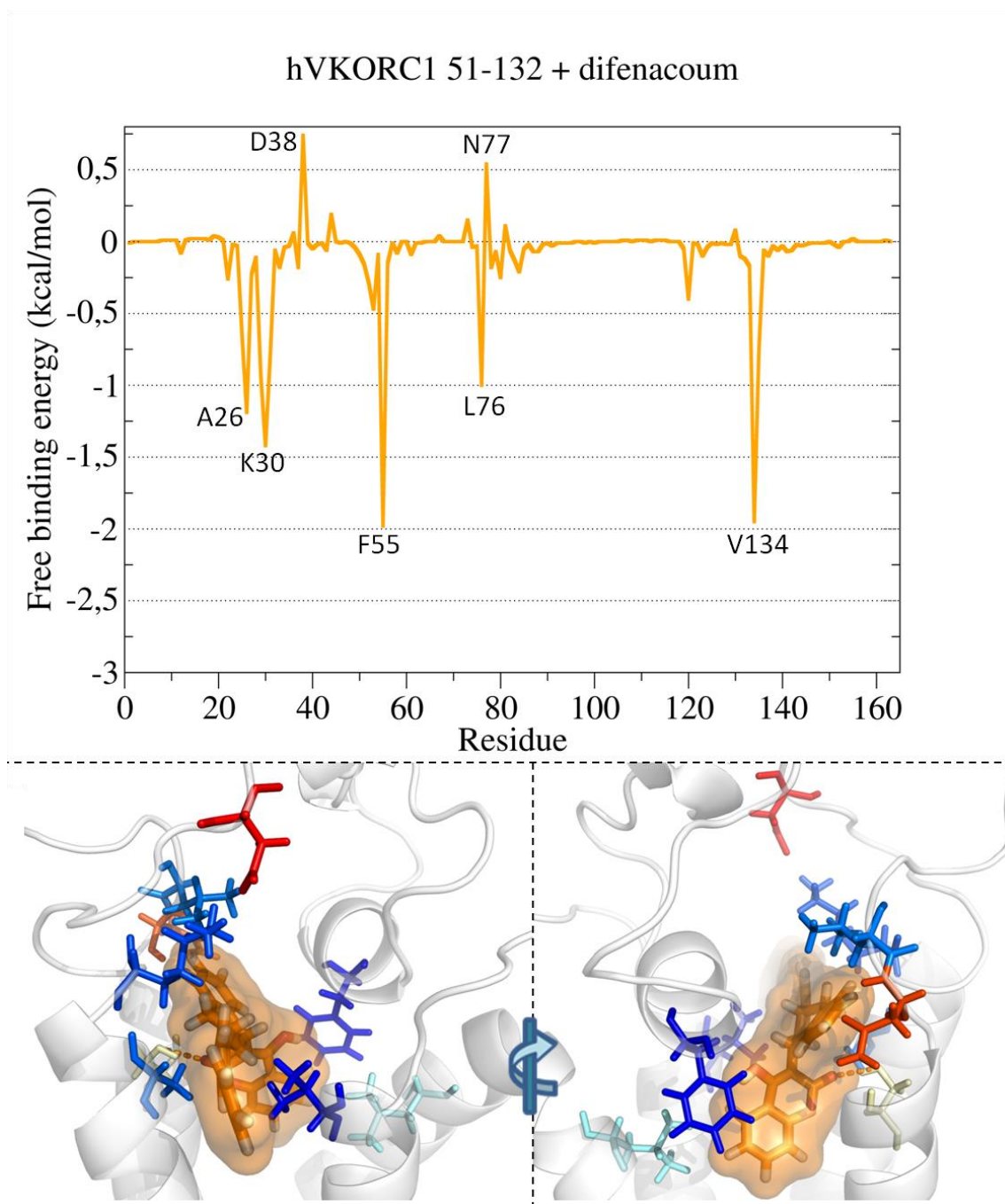


Figure 3.51. Détermination de la contribution des résidus de hVKORC1_{C51-C132} dans les interactions de l'enzyme avec le difénacoum au cours de deux simulations de DM de 100 ns concaténées. Le graphe représente la contribution de chaque résidu de hVKORC1 à l'énergie libre de liaison totale entre l'enzyme et le difénacoum. La contribution calculée pour chaque résidu est ensuite reportée sur la structure 3D du complexe hVKORC1•difénacoum. Les résidus de hVKORC1 interagissant avec le difénacoum sont représentés en *sticks* et colorés selon leur contribution, du bleu foncé (résidu les plus favorables à la fixation du difénacoum) au rouge (résidu les moins favorables à la fixation du difénacoum). La structure 3D du difénacoum est représentée en *sticks*, sa surface en orange transparent ; les atomes de carbone, d'oxygène et d'hydrogène sont colorés en orange, rouge et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* blanc transparent.

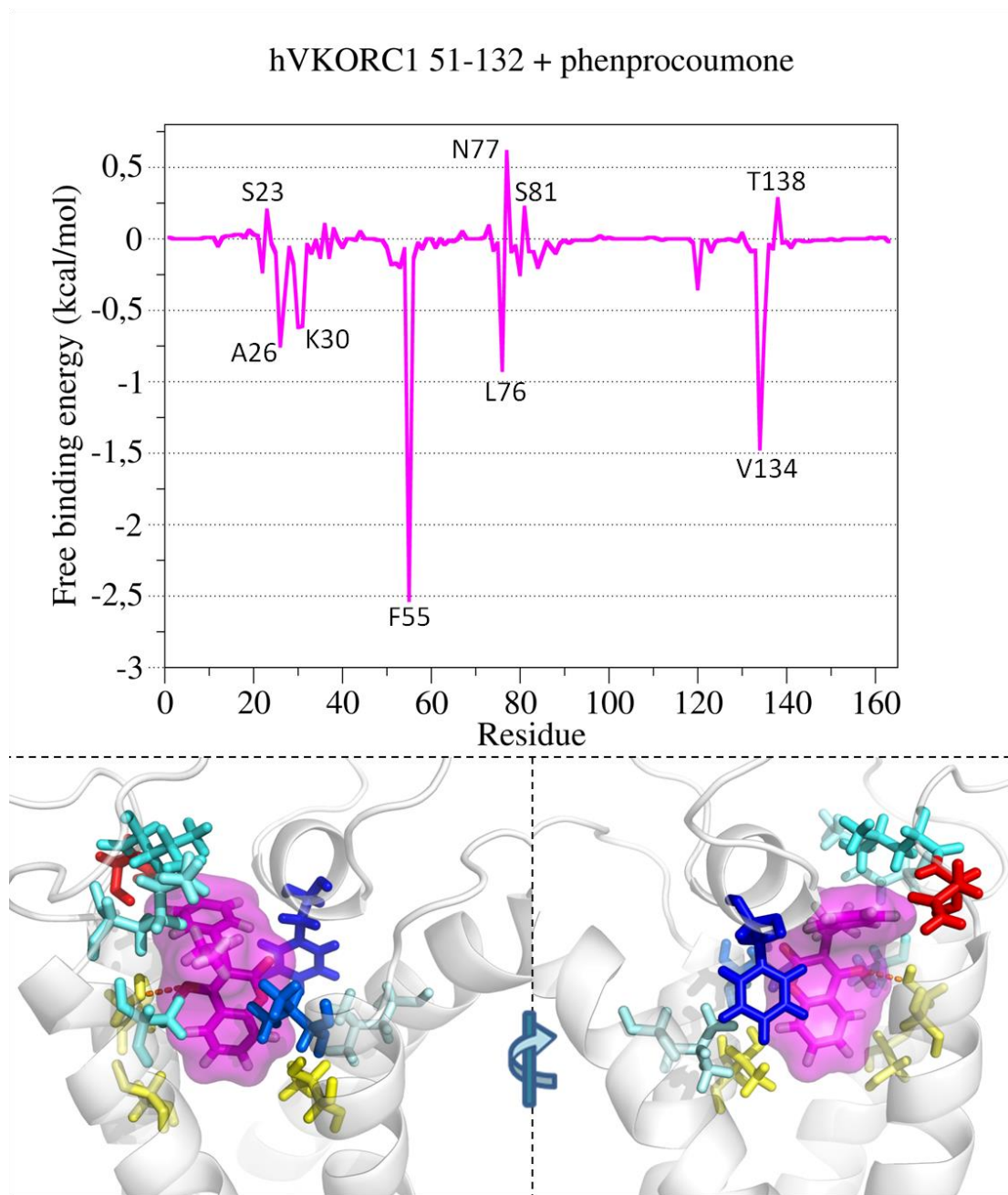


Figure 3.52. Détermination de la contribution des résidus de hVKORC1_{C51-C132} dans les interactions de l'enzyme avec la phenprocoumone au cours de deux simulations de DM de 100 ns concaténées. Le graphe représente la contribution de chaque résidu de hVKORC1 à l'énergie libre de liaison totale entre l'enzyme et la phenprocoumone. La contribution calculée pour chaque résidu est ensuite reportée sur la structure 3D du complexe hVKORC1•phenprocoumone. Les résidus de hVKORC1 interagissant avec la phenprocoumone sont représentés en *sticks* et colorés selon leur contribution, du bleu foncé (résidu les plus favorables à la fixation de la phenprocoumone) au rouge (résidu les moins favorables à la fixation de la phenprocoumone). La structure 3D de la phenprocoumone est représentée en *sticks*, sa surface en magenta transparent ; les atomes de carbone, d'oxygène et d'hydrogène sont colorés en magenta, rouge et blanc, respectivement. La structure 3D de hVKORC1 est représentée en *cartoon* blanc transparent.

L'ensemble des résultats générés nous a permis d'identifier les résidus impliqués dans les interactions de hVKORC1 avec la vitK^E et trois AVKs. Les interactions de type hydrophobe sont prédominantes dans la fixation de la vitK^E et des AVKs. Les quatre résidus A₂₆, F₅₅, L₇₆ et V₁₃₄ sont les plus favorables à la fixation de la vitK^E et des trois AVKs étudiés. Le résidu K₃₀ est uniquement impliqué dans la fixation des AVKs. Le résidu F₅₅ présente systématiquement la contribution la plus forte à l'ELL totale. Or, ce résidu a été identifié expérimentalement comme prépondérant dans la fixation de la warfarine (un autre AVK) sur VKORC1 [111]. Le résidu F₅₅ agit comme un couvercle maintenant la vitK^E à proximité du résidu C₁₃₅ du site actif. Seule la formation du pont disulfure entre les résidus C₅₁ et C₁₃₂ permet un positionnement adéquat du résidu F₅₅. En effet, comme nous l'avons démontré dans la partie **RESULTATS.II.3**, l'hélice HH (portant le résidu F₅₅) se trouve en position "basse", donc à proximité du site actif, lors du rapprochement des résidus C₁₃₂ et C₁₃₅ (**Figure 3.25**). La forme active de hVKORC1, correspondant au modèle hVKORC1_C₅₁-C₁₃₂, est donc liée à l'accessibilité du résidu C₁₃₅ à la vitK^E, ainsi qu'au rapprochement de l'hélice HH (donc du résidu F₅₅) du site actif, permettant le maintien de la vitK^E à proximité du résidu C₁₃₅.

Nous avons donc déterminé les modes d'interaction du complexe hVKORC1•vitK^E et de trois complexes hVKORC1•AVK. La compréhension au niveau structural du fonctionnement des AVKs actuellement utilisés en thérapie ouvre la voie à leur optimisation et la conception *in silico* d'une nouvelle génération d'AVKs.

Chapitre 4. CONCLUSION ET PERSPECTIVES

L'enzyme VKORC1 (*Vitamin K epOxide Reductase Complex subunit 1*), localisée dans la membrane du réticulum endoplasmique, effectue le recyclage de la vitamine K et joue ainsi un rôle essentiel dans le processus de coagulation sanguine. Par conséquent, la régulation de l'activité de VKORC1 constitue une stratégie privilégiée dans les traitements thérapeutiques anticoagulants. Des inhibiteurs anti-vitamine K (AVK) ont donc été développés, permettant d'empêcher le recyclage de la vitamine K par VKORC1, interrompant le processus de coagulation sanguine. Cependant, ni les mécanismes de réduction de la vitamine K époxyde (vitK^E) en vitamine K quinone (vitK^Q) par VKORC1, ni les mécanismes d'inhibition par les AVKs ne sont clairement établis. L'insuffisance des données expérimentales, liée aux difficultés de produire la protéine isolément puis de la purifier, est à l'origine de ce manque d'informations. De plus, les études biochimiques et biophysiques ont produit des résultats contradictoires sur le nombre d'hélices transmembranaires composant la protéine et sur le rôle des résidus C₄₃ et C₅₁ dans l'activation de l'enzyme. En outre, aucune structure tridimensionnelle n'a été résolue expérimentalement à ce jour pour une VKORC1 mammalienne, sauvage ou mutée, rendant inaccessible la description détaillée du fonctionnement de VKORC1 au niveau structural.

Nous avons entrepris l'étude de VKORC1 par des méthodes de modélisation moléculaire, avec pour objectif d'apporter des éléments de réponse aux différentes incertitudes portant sur l'enzyme. La modélisation moléculaire permet d'établir *in silico* des modèles structuraux à l'échelle atomique, à partir de protéines homologues dont la structure 3D a été résolue expérimentalement (par cristallographie aux rayons X, résonance magnétique nucléaire ou microscopie électronique). La flexibilité et la dynamique d'une protéine jouent un rôle aussi important pour son fonctionnement que sa structure, la description de la composante dynamique de la protéine est donc indispensable. Les simulations de dynamique moléculaire (DM) ouvrent l'accès au détail des mouvements atomiques. Les simulations de DM permettent ainsi d'explorer *in silico* l'espace conformationnel des protéines étudiées, actuellement inaccessible aux méthodes expérimentales. L'association de la modélisation moléculaire, des simulations de DM et de la biologie structurale aboutit à la description des relations entre la séquence, la structure, la dynamique et la fonction biologique des protéines, par des approches de biologie computationnelle.

Nous avons débuté notre étude par des prédictions de structures secondaires et de repliement membranaire à partir de la séquence primaire de hVKORC1. Nous avons postulé, à partir de ces résultats, un repliement tridimensionnel de hVKORC1 à quatre hélices transmembranaires (TMs). Nous avons également effectué une analyse de la séquence peptidique de la VKORC1 humaine (hVKORC1), afin de rechercher des protéines de séquence homologue dont la structure 3D avait été résolue expérimentalement.

Nous avons ensuite construit un modèle 3D de hVKORC1 en utilisant la structure d'un homologue bactérien (bVKOR) comme "*template*". Nous avons modélisé la structure 3D de VKORC1 en considérant les quatre résidus C₄₃, C₅₁, C₁₃₂ et C₁₃₅ à l'état protoné (SH), afin de mener notre étude le plus objectivement possible, indépendamment des données structurales de bVKOR. Aucun pont disulfure n'a donc été imposé lors de la modélisation de la structure de VKORC1, à la différence des études par modélisation menées précédemment par d'autres équipes. Nous avons suggéré une possibilité de prédire la formation de ponts disulfures par des simulations de DM de ce modèle dépourvu de contraintes.

Cinq simulations de DM de 100 ns chacune ont révélé une grande diversité conformationnelle de la boucle luminale reliant TM1 et TM2, permettant des rapprochements privilégiés entre certaines paires de résidus de cystéine. Deux simulations, la première montrant un rapprochement entre les atomes de soufre des résidus C₅₁ et C₁₃₂, la seconde un rapprochement entre les atomes de soufre des résidus C₄₃ et C₅₁ d'une part et C₁₃₂ et C₁₃₅ d'autre part, ont été prolongées jusqu'à 1 μ s. L'analyse de ces deux trajectoires prolongées a montré que le modèle de VKORC1 semblait avoir atteint un état d'équilibre. Cependant, la grande variabilité structurale de hVKORC1, induite par la flexibilité de la boucle L, a compliqué la caractérisation conformationnelle de l'enzyme. Nous avons donc développé une nouvelle méthode statistique avancée basée sur l'analyse spectrale, en collaboration avec les mathématiciens, les Professeurs B. CHALMOND et A. TROUVE (CMLA, ENS Cachan).

Cette méthode mathématique nous a permis d'estimer le paysage énergétique des conformations de hVKORC1 générées par DM. Les conformations associées aux minima d'énergie locaux ont été considérées comme des candidates probables pour la formation de ponts disulfures entre les paires de résidus C₄₃-C₅₁, C₅₁-C₁₃₂ et C₁₃₂-C₁₃₅.

Quatre modèles de hVKORC1, reflétant chacun une combinaison de ponts disulfures prédits, ont ensuite été construits puis étudiés par DM. L'affinité des différentes formes de la vitamine K (époxyde, hydroxyquinone et quinone) avec chacun de ces modèles a été évaluée. Les résultats nous ont permis d'identifier l'état conformationnel actif de hVKORC1 et de postuler le mécanisme enzymatique détaillé de l'enzyme.

Les résidus stabilisant la vitamine K dans le site actif de hVKORC1 et l'énergie d'interaction du complexe ont été caractérisés. Nous avons également étudié trois complexes hVKORC1•AVK (les trois AVKs étant l'acénocoumarol, le difénacoum et la phenprocoumone). Les affinités prédites entre chaque AVK étudié et hVKORC1 se sont avérées cohérentes avec les constantes d'inhibition mesurées expérimentalement. Cette corrélation des résultats prédits par modélisation moléculaire avec les données empiriques valide ainsi nos modèles théoriques et le protocole ayant généré ces résultats. De plus, une publication récente [112] rapporte la fixation d'un autre AVK, la coumaphène, sur un état conformationnel de hVKORC1 équivalent à celui prédit par nos travaux. Une autre publication a mis en évidence le rôle prépondérant du résidu F₅₅ dans la fixation de la coumaphène [122], correspondant à son rôle dans la reconnaissance de la vitamine K et des AVKs prédit par nos travaux. Ces résultats expérimentaux renforcent encore nos résultats de modélisation, confirmant la détermination correcte de la structure de VKORC1 ciblée par la vitamine K époxyde et trois AVKs.

En conséquence, le concept de modélisation et de simulation d'un modèle initialement dépourvu de contrainte structurale (sans pont disulfure imposé) pour prédire les conformations biologiquement pertinente de hVKORC1 est convenable, et la méthode d'analyse que nous avons développée est cohérente. La stratégie élaborée est applicable aux mutants de hVKORC1 pour décrire les mécanismes de modification du cycle enzymatique (suractivation ou inactivation) et pour développer de nouveaux inhibiteurs (AVKs) de VKORC1, modulant l'activité de l'enzyme et/ou surmontant les résistances actuelles induites par certaines mutations.

Les travaux présentés dans cette thèse et les résultats obtenus ouvrent la voie à une étude plus approfondie des mécanismes d'inhibition alternatifs de VKORC1. En effet, la flexibilité de la boucle luminale et les résidus hydrophobes qui la composent peuvent constituer un site d'interaction protéine-protéine putatif. En outre, l'implication des résidus C₄₃ et C₅₁ dans l'activation de VKORC1, révélée par nos résultats, est en adéquation avec l'activation de VKORC1 par une protéine partenaire, une des trois protéines thiorédoxines proposées dans la littérature. Une future étude par modélisation moléculaire des interactions entre ces trois protéines et les quatre états conformationnels prédits pour VKORC1 présente un intérêt majeur. Elle pourrait permettre de caractériser finement les interactions entre les thiorédoxines et VKORC1, et ainsi aboutir au développement de nouveaux inhibiteurs (non AVKs) empêchant l'activation de VKORC1 par sa ou ses protéine(s) partenaire(s).

Références

Reference List

1. C.Guyton, A. and J. E.Hall, *Précis de Physiologie Médicale*. Padoue, 2003.
2. Czogalla, K.J., et al., *Human VKORC1 mutations cause variable degrees of 4-hydroxycoumarin resistance and affect putative warfarin binding interfaces*. *Blood*, 2013. **122**(15): p. 2743-2750.
3. Baker, D.C. and J. Brassard, *Review of continuing education course on hemostasis*. *Toxicol. Pathol*, 2011. **39**(1): p. 281-288.
4. Berkner, K.L. and K.W. Runge, *The physiology of vitamin K nutrition and vitamin K-dependent protein function in atherosclerosis*. *J. Thromb. Haemost*, 2004. **2**(12): p. 2118-2132.
5. Stafford, D.W., *The vitamin K cycle*. *J. Thromb. Haemost*, 2005. **3**(8): p. 1873-1878.
6. Ilesanmi, O.O., *Pathological basis of symptoms and crises in sickle cell disorder: implications for counseling and psychotherapy*. *Hematol. Rep*, 2010. **2**(1): p. e2.
7. Dam, H., *The antihaemorrhagic vitamin of the chick*. *Biochem. J*, 1935. **29**(6): p. 1273-1285.
8. Koshihara, Y., K. Hoshi, and M. Shiraki, *Vitamin K2 (menatetrenone) inhibits prostaglandin synthesis in cultured human osteoblast-like periosteal cells by inhibiting prostaglandin H synthase activity*. *Biochem. Pharmacol*, 1993. **46**(8): p. 1355-1362.
9. Hara, K., et al., *The inhibitory effect of vitamin K2 (menatetrenone) on bone resorption may be related to its side chain*. *Bone*, 1995. **16**(2): p. 179-184.
10. Reddi, K., et al., *Interleukin 6 production by lipopolysaccharide-stimulated human fibroblasts is potently inhibited by naphthoquinone (vitamin K) compounds*. *Cytokine*, 1995. **7**(3): p. 287-290.
11. Li, J., et al., *Novel role of vitamin k in preventing oxidative injury to developing oligodendrocytes and neurons*. *J. Neurosci*, 2003. **23**(13): p. 5816-5826.
12. Lev, M. and A.F. Milford, *Effect of vitamin K depletion and restoration on sphingolipid metabolism in Bacteroides melaninogenicus*. *J. Lipid Res*, 1972. **13**(3): p. 364-370.
13. Thomas, M.W. and B.M. Gebara, *Difficulty breathing. Late onset vitamin K deficiency bleeding*. *Clin. Pediatr. (Phila)*, 2004. **43**(5): p. 499-502.
14. Oldenburg, J., et al., *The vitamin K cycle*. *Vitam. Horm*, 2008. **78**: p. 35-62.
15. Benton, M.E., P.A. Price, and J.W. Suttie, *Multi-site-specificity of the vitamin K-dependent carboxylase: in vitro carboxylation of des-gamma-carboxylated bone Gla protein and Des-gamma-carboxylated pro bone Gla protein*. *Biochemistry*, 1995. **34**(29): p. 9541-9551.
16. Tie, J.K. and D.W. Stafford, *Structure and function of vitamin K epoxide reductase*. *Vitam. Horm*, 2008. **78**: p. 103-130.
17. Bell, R.G. and J.T. Matschiner, *Vitamin K activity of phylloquinone oxide*. *Arch. Biochem. Biophys*, 1970. **141**(2): p. 473-476.
18. Rost, S., et al., *Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2*. *Nature*, 2004. **427**(6974): p. 537-541.
19. Li, T., et al., *Identification of the gene for vitamin K epoxide reductase*. *Nature*, 2004. **427**(6974): p. 541-544.
20. Oldenburg, J., et al., *Vitamin K epoxide reductase complex subunit 1 (VKORC1): the key protein of the vitamin K cycle*. *Antioxid. Redox. Signal*, 2006. **8**(3-4): p. 347-353.
21. Goodstadt, L. and C.P. Ponting, *Vitamin K epoxide reductase: homology, active site and catalytic mechanism*. *Trends Biochem. Sci*, 2004. **29**(6): p. 289-292.

22. Rost, S., et al., *Site-directed mutagenesis of coumarin-type anticoagulant-sensitive VKORC1: evidence that highly conserved amino acids define structural requirements for enzymatic activity and inhibition by warfarin*. *Thromb. Haemost.*, 2005. **94**(4): p. 780-786.
23. Rishavy, M.A., et al., *Novel insight into the mechanism of the vitamin K oxidoreductase (VKOR): electron relay through Cys43 and Cys51 reduces VKOR to allow vitamin K reduction and facilitation of vitamin K-dependent protein carboxylation*. *J. Biol. Chem.*, 2011. **286**(9): p. 7267-7278.
24. Silverman, R.B. and D.L. Nandi, *Reduced thioredoxin: a possible physiological cofactor for vitamin K epoxide reductase. Further support for an active site disulfide*. *Biochem. Biophys. Res. Commun.*, 1988. **155**(3): p. 1248-1254.
25. Johan, L., et al., *Vitamin K-dependent carboxylase. Possible role for thioredoxin in the reduction of vitamin K metabolites in liver*. *FEBS Lett.*, 1987. **222**(2): p. 353-357.
26. Wajih, N., S.M. Hutson, and R. Wallin, *Disulfide-dependent protein folding is linked to operation of the vitamin K cycle in the endoplasmic reticulum. A protein disulfide isomerase-VKORC1 redox enzyme complex appears to be responsible for vitamin K1 2,3-epoxide reduction*. *J. Biol. Chem.*, 2007. **282**(4): p. 2626-2635.
27. Whitlon, D.S., J.A. Sadowski, and J.W. Suttie, *Mechanism of coumarin action: significance of vitamin K epoxide reductase inhibition*. *Biochemistry*, 1978. **17**(8): p. 1371-1377.
28. Preusch, P.C., *Is thioredoxin the physiological vitamin K epoxide reducing agent?* *FEBS Lett.*, 1992. **305**(3): p. 257-259.
29. Schulman, S., et al., *Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners*. *Proc. Natl. Acad. Sci. U. S. A.*, 2010. **107**(34): p. 15027-15032.
30. Jin, D.Y., J.K. Tie, and D.W. Stafford, *The conversion of vitamin K epoxide to vitamin K quinone and vitamin K quinone to vitamin K hydroquinone uses the same active site cysteines*. *Biochemistry*, 2007. **46**(24): p. 7279-7283.
31. Tie, J.K., et al., *Membrane topology mapping of vitamin K epoxide reductase by in vitro translation/cotranslocation*. *J. Biol. Chem.*, 2005. **280**(16): p. 16410-16416.
32. Li, W., et al., *Structure of a bacterial homologue of vitamin K epoxide reductase*. *Nature*, 2010. **463**(7280): p. 507-512.
33. Tie, J.K. and D.W. Stafford, *Structural and functional insights into enzymes of the vitamin K cycle*. *J. Thromb. Haemost.*, 2016. **14**(2): p. 236-247.
34. Cao, Z., et al., *The membrane topology of vitamin K epoxide reductase is conserved between human isoforms and the bacterial enzyme*. *Biochem. J.*, 2016. **473**(7): p. 851-858.
35. Czogalla, K.J., M. Watzka, and J. Oldenburg, *Structural Modeling Insights into Human VKORC1 Phenotypes*. *Nutrients*, 2015. **7**(8): p. 6837-6851.
36. Wu, S., et al., *Membrane topology for human vitamin K epoxide reductase*. *J. Thromb. Haemost.*, 2014. **12**(1): p. 112-114.
37. Van Horn, W.D., *Structural and functional insights into human vitamin K epoxide reductase and vitamin K epoxide reductase-like1*. *Crit Rev. Biochem. Mol. Biol.*, 2013. **48**(4): p. 357-372.
38. Tie, J.K., D.Y. Jin, and D.W. Stafford, *Human vitamin K epoxide reductase and its bacterial homologue have different membrane topologies and reaction mechanisms*. *J. Biol. Chem.*, 2012. **287**(41): p. 33945-33955.
39. F.W.Schofield, *Canadian Veterinaty Record*, 1922.
40. LINK, K.P., *The discovery of dicumarol and its sequels*. *Circulation*, 1959. **19**(1): p. 97-107.

41. ERNSTER, L., *Transhydrogenase and TPNH oxidase activities of rat-liver mitochondria*. *Biochim. Biophys. Acta*, 1960. **38**: p. 170-173.
42. Fasco, M.J., et al., *Warfarin inhibition of vitamin K 2,3-epoxide reductase in rat liver microsomes*. *Biochemistry*, 1983. **22**(24): p. 5655-5660.
43. Hodroge, A., et al., *VKORC1 mutations detected in patients resistant to vitamin K antagonists are not all associated with a resistant VKOR activity*. *J. Thromb. Haemost*, 2012. **10**(12): p. 2535-2543.
44. Watzka, M., et al., *Thirteen novel VKORC1 mutations associated with oral anticoagulant resistance: insights into improved patient diagnosis and treatment*. *J. Thromb. Haemost*, 2011. **9**(1): p. 109-118.
45. Wallin, R., N. Wajih, and S.M. Hutson, *VKORC1: a warfarin-sensitive enzyme in vitamin K metabolism and biosynthesis of vitamin K-dependent blood coagulation factors*. *Vitam. Horm*, 2008. **78**: p. 227-246.
46. Weston, B.W. and P.E. Monahan, *Familial deficiency of vitamin K-dependent clotting factors*. *Haemophilia*, 2008. **14**(6): p. 1209-1213.
47. Bodin, L., et al., *Multiple genetic alterations in vitamin K epoxide reductase complex subunit 1 gene (VKORC1) can explain the high dose requirement during oral anticoagulation in humans*. *J. Thromb. Haemost*, 2008. **6**(8): p. 1436-1439.
48. Wilms, E.B., et al., *A new VKORC1 allelic variant (p.Trp59Arg) in a patient with partial resistance to acenocoumarol and phenprocoumon*. *J. Thromb. Haemost*, 2008. **6**(7): p. 1224-1226.
49. Munoz, V. and L. Serrano, *Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales*. *Proteins*, 1994. **20**(4): p. 301-311.
50. Yuan, S., et al., *PyMOL and Inkscape Bridge the Data and the Data Visualization*. *Structure*, 2016. **24**(12): p. 2041-2042.
51. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. *J. Mol. Graph*, 1996. **14**(1): p. 33-38.
52. Riliskis, L. and E. Osipov, *Maestro: an orchestration framework for large-scale WSN simulations*. *Sensors (Basel)*, 2014. **14**(3): p. 5392-5414.
53. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. *J. Comput. Chem*, 2004. **25**(13): p. 1605-1612.
54. Berman, H.M., et al., *The Protein Data Bank*. *Nucleic Acids Res*, 2000. **28**(1): p. 235-242.
55. Duan, Y. and P.A. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*. *Science*, 1998. **282**(5389): p. 740-744.
56. Dror, R.O., et al., *Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations*. *J. Gen. Physiol*, 2010. **135**(6): p. 555-562.
57. Omelyan, I.P., I.M. Mryglod, and R. Folk, *Optimized Verlet-like algorithms for molecular dynamics simulations*. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys*, 2002. **65**(5 Pt 2): p. 056706.
58. Magrane, M., *UniProt Knowledgebase: a hub of integrated protein data*. *Database (Oxford)*, 2011. **2011**: p. bar009.
59. Sievers, F. and D.G. Higgins, *Clustal Omega, accurate alignment of very large numbers of sequences*. *Methods Mol. Biol*, 2014. **1079**: p. 105-116.
60. Mount, D.W., *Using the Basic Local Alignment Search Tool (BLAST)*. *CSH. Protoc*, 2007. **2007**: p. db.
61. Frishman, D. and P. Argos, *Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence*. *Protein Eng*, 1996. **9**(2): p. 133-142.

62. Sen, T.Z., et al., *GOR V server for protein secondary structure prediction*. Bioinformatics, 2005. **21**(11): p. 2787-2788.
63. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices*. J. Mol. Biol, 1999. **292**(2): p. 195-202.
64. Ouali, M. and R.D. King, *Cascaded multiple classifiers for secondary structure prediction*. Protein Sci, 2000. **9**(6): p. 1162-1176.
65. Cuff, J.A., et al., *JPred: a consensus secondary structure prediction server*. Bioinformatics, 1998. **14**(10): p. 892-893.
66. Tusnady, G.E. and I. Simon, *Principles governing amino acid composition of integral membrane proteins: application to topology prediction*. J. Mol. Biol, 1998. **283**(2): p. 489-506.
67. Viklund, H. and A. Elofsson, *OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar*. Bioinformatics, 2008. **24**(15): p. 1662-1668.
68. Viklund, H., et al., *SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology*. Bioinformatics, 2008. **24**(24): p. 2928-2929.
69. Kall, L., A. Krogh, and E.L. Sonnhammer, *An HMM posterior decoder for sequence feature prediction that includes homology information*. Bioinformatics, 2005. **21 Suppl 1**: p. i251-i257.
70. Reynolds, S.M., et al., *Transmembrane topology and signal peptide prediction using dynamic bayesian networks*. PLoS. Comput. Biol, 2008. **4**(11): p. e1000213.
71. Bernsel, A., et al., *Prediction of membrane-protein topology from first principles*. Proc. Natl. Acad. Sci. U. S. A, 2008. **105**(20): p. 7177-7181.
72. Olivella, M., et al., *Relation between sequence and structure in membrane proteins*. Bioinformatics, 2013. **29**(13): p. 1589-1592.
73. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER*. Curr. Protoc. Protein Sci, 2016. **86**: p. 2-2.
74. Jones, T.A. and S. Thirup, *Using known substructures in protein model building and crystallography*. EMBO J, 1986. **5**(4): p. 819-822.
75. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free*. J. Comput. Chem, 2005. **26**(16): p. 1701-1718.
76. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles*. J. Chem. Theory. Comput, 2012. **8**(9): p. 3257-3273.
77. Ghahremanpour, M.M., et al., *MemBuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the GROMACS biomolecular simulation program*. Bioinformatics, 2014. **30**(3): p. 439-441.
78. Jambeck, J.P. and A.P. Lyubartsev, *Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids*. J. Phys. Chem. B, 2012. **116**(10): p. 3164-3179.
79. Price, D.J. and C.L. Brooks, III, *A modified TIP3P water potential for simulation with Ewald summation*. J. Chem. Phys, 2004. **121**(20): p. 10096-10103.
80. van Gunsteren, W.F. and H.J. Berendsen, *Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state*. J. Mol. Biol, 1984. **176**(4): p. 559-564.
81. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. J. Chem. Phys, 2007. **126**(1): p. 014101.
82. Allen, W.J., J.A. Lemkul, and D.R. Bevan, *GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics*. J. Comput. Chem, 2009. **30**(12): p. 1952-1958.

83. Wolf, M.G., et al., *g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation*. J. Comput. Chem, 2010. **31**(11): p. 2169-2174.
84. Hockney, R., S. Goel, and J. Eastwood, *Quiet high-resolution computer models of a plasma*. J. Comput. Phys., 1974. **14**: p. 148-158.
85. Verlet, L., *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*. Physical Review, 1967. **159**.
86. *The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC*.
87. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures*. Bioinformatics, 2006. **22**(21): p. 2695-2696.
88. Le, G., V. P. Schmidtke, and P. Tuffery, *Fpocket: an open source platform for ligand pocket detection*. BMC. Bioinformatics, 2009. **10**: p. 168.
89. Ewing, T.J., et al., *DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases*. J. Comput. Aided Mol. Des, 2001. **15**(5): p. 411-428.
90. Miller, M.D., et al., *FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure*. J. Comput. Aided Mol. Des, 1994. **8**(2): p. 153-174.
91. Rao, S.N., et al., *Validation studies of the site-directed docking program LibDock*. J. Chem. Inf. Model, 2007. **47**(6): p. 2159-2171.
92. Burkhard, P., P. Taylor, and M.D. Walkinshaw, *An example of a protein ligand found by database mining: description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex*. J Mol Biol, 1998. **277**(2): p. 449-66.
93. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. J. Comput. Chem, 2009. **30**(16): p. 2785-2791.
94. Neves, M.A., M. Totrov, and R. Abagyan, *Docking and scoring with ICM: the benchmarking results and strategies for improvement*. J. Comput. Aided Mol. Des, 2012. **26**(6): p. 675-686.
95. Alisaraie, L., L.A. Haller, and G. Fels, *A QXP-based multistep docking procedure for accurate prediction of protein-ligand complexes*. J. Chem. Inf. Model, 2006. **46**(3): p. 1174-1187.
96. Nurisso, A., et al., *Molecular docking using the molecular lipophilicity potential as hydrophobic descriptor: impact on GOLD docking performance*. J. Chem. Inf. Model, 2012. **52**(5): p. 1319-1327.
97. Taylor, J.S. and R.M. Burnett, *DARWIN: a program for docking flexible molecules*. Proteins, 2000. **41**(2): p. 173-191.
98. Meng, X.Y., et al., *Molecular docking: a powerful approach for structure-based drug discovery*. Curr. Comput. Aided Drug Des, 2011. **7**(2): p. 146-157.
99. Kumari, R., R. Kumar, and A. Lynn, *g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations*. J. Chem. Inf. Model, 2014. **54**(7): p. 1951-1962.
100. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. I. The accuracy of binding free energy calculations based on molecular dynamics simulations*. J. Chem. Inf. Model, 2011. **51**(1): p. 69-82.
101. Liu, S., et al., *Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer*. Nat. Commun, 2014. **5**: p. 3110.
102. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*. Biopolymers, 1983. **22**(12): p. 2577-2637.
103. Nugent, T. and D.T. Jones, *Membrane protein structural bioinformatics*. J Struct Biol, 2012. **179**(3): p. 327-37.
104. Parzen, E., *On the estimation of a probability density function and the mode*. Annals of Math. Stats., 1962. **33**: p. 1065-1076.

105. Coifman, R.R., et al., *Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems*. Multiscale Modeling and Simulation, 2008. **7**: p. 842-864.
106. Nadler, B., et al., *Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators*. Neural Information Processing Systems, 2005. **18**.
107. Nadler, B., et al., *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems*. Applied and Computational Harmonic Analysis, 2006. **21**: p. 113-127.
108. Von Luxbourg, U., *A tutorial on spectral Clustering*. Statistics and Computing, 2007. **17**(4): p. arXiv:0711.0189.
109. Chalmond, B., *Modeling and Inverse Problem in Image Analysis*. Applied Mathematical Sciences, 2003. **155**.
110. Harada, R. and A. Kitao, *Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to generate conformational transition pathway*. J. Chem. Phys, 2013. **139**(3): p. 035103.
111. Czogalla, K.J., et al., *Warfarin and vitamin K compete for binding to Phe55 in human VKOR*. Nat. Struct. Mol. Biol, 2017. **24**(1): p. 77-85.
112. Shen, G., et al., *Warfarin traps human vitamin K epoxide reductase in an intermediate state during electron transfer*. Nat. Struct. Mol. Biol, 2017. **24**(1): p. 69-76.

Titre : VKORC1 et résistance aux Anti-vitamines K : Etude par Modélisation Moléculaire

Mots clés : VKOR, vitamine K, modèle 3D, dynamique moléculaire, conformation, cible.

Résumé : VKORC1 est une enzyme membranaire du réticulum endoplasmique, responsable de la réduction de la vitamine K époxyde en vitamine K quinone, activant la synthèse des facteurs de coagulation. VKORC1 constitue ainsi une cible thérapeutique privilégiée des anticoagulants de type anti-vitamine K (AVKs). Cependant, certaines mutations de VKORC1 induisent une dérégulation physiologique et/ou une résistance aux AVKs. En l'absence de données structurales, plusieurs modèles topologiques de VKORC1 ont été proposés à partir des données biochimiques et biophysiques, fréquemment contradictoires. La topologie de VKORC1 ainsi que l'implication des résidus de cystéine (strictement conservés chez toutes les VKORs) dans le mécanisme enzymatique de la protéine restent incertaines. Nous avons construit, par des méthodes *in silico*, un modèle 3D de la VKORC1 humaine sauvage (hVKORC1^{WT}) à l'échelle atomique. Des simulations de dynamique moléculaire de la protéine, en considérant tous les résidus de cystéine sous leur forme oxydée (SH), nous ont permis d'identifier les résidus de cystéine les plus susceptibles de former des ponts disulfures. Nous avons par conséquent décrit les conformations métastables de hVKORC1^{WT} mimant les états fonctionnels de la protéine. L'étude de la recon-

naissance des formes époxyde et réduites de la vitamine K par les conformations prédites de hVKORC1^{WT} a mis en évidence leur sélectivité réciproque. Les conformations de hVKORC1^{WT} ciblées par chaque forme de la vitamine K et leur rôle dans le mécanisme de réduction de la molécule ont ainsi été caractérisés. Nous avons postulé à partir de ces résultats le mécanisme enzymatique détaillé de hVKORC1^{WT} et proposé la structure-cible des AVKs. Les interactions entre la cible prédite - hVKORC1^{WT} à l'état actif - et trois AVKs différents ont été décrites en termes d'énergie libre de liaison. Ces résultats ont été corrélés avec les constantes d'inhibition mesurées *in vivo*, validant nos prédictions théoriques. Notre protocole, développé pour hVKORC1^{WT}, est applicable aux formes mutées de l'enzyme. Il permettra la description des mécanismes modifiant l'activité réductase de hVKORC1 et/ou provoquant une résistance aux AVKs. Cette description ouvre la voie à la conception d'une nouvelle génération d'inhibiteurs surmontant les résistances aux AVKs. Notre concept et le protocole établi pour l'étude de hVKORC1 peuvent être étendus à l'analyse des VKORs mammaliennes et aux autres enzymes de la famille des oxydoréductases.

Title: VKORC1 and resistance to anti-vitaminK :Molecular Modeling Study

Keywords: VKOR, vitamine K, 3D model, molecular dynamics, conformations, target.

Abstract: VKORC1 is an endoplasmic reticulum membrane-resident enzyme, responsible for vitamin K epoxide reduction to vitamin K quinone that activates coagulation factors synthesis. VKORC1 is thus a prominent target of vitamin K agonists (VKAs) in anticoagulant therapies. However, some VKORC1 mutations lead to physiological dysregulation and/or VKAs resistance. No VKORC1 structural data is available, and postulated topological models are based on biochemical and biophysical experimental observations frequently contradicting. Topology of VKORC1 and involvement of cysteines residues (highly conserved in VKORs) in the protein enzymatic mechanism remain unclear. We built an *in silico* 3D model of the wild-type human VKORC1 (hVKORC1^{WT}) at the atomistic scale. Molecular dynamics simulations of the protein model, carrying all the cysteines residues in their oxidized form (SH), were used for identification of cysteines residues which may plausibly form disulfide bridges. We thus described hVKORC1^{WT} metastable conformations depicting the functionally relevant states of protein. Study of the vitamin K (in epoxide and in reduced forms) recognition by the predicted

conformations of hVKORC1^{WT} revealed their reciprocal selectivity. The hVKORC1^{WT} conformations targeted by each vitamin K form were established and their role in the reduction mechanism of this molecule was explained. Using our results, we postulated the comprehensive enzymatic mechanism of hVKORC1^{WT} and we proposed the 3D structure as the VKAs target. Interactions between the predicted target - hVKORC1^{WT} active state - and three different VKAs were characterized. The obtained affinities (free binding energy) were in good correlation with *in vivo* measured inhibition constants (Ki), thus validating our theoretical predictions. Our protocol developed for hVKORC1^{WT} is suitable for a study of its mutants. Description of the enzymatic mechanisms of mutated hVKORC1 will lead to understanding of the modified reductase activity or/and to explaining of its resistance to VKAs. Such data are crucial for the development of novel strategies in the design of a new generation of inhibitors overcoming VKAs resistance. Our concept and the established *in silico* protocol can be extended to analysis of mammalian VKORs and other oxidoreductases family proteins.

Université Paris-Saclay

Espace Technologique / Immeuble Discovery

Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

