

Université Hassan II Casablanca

Faculté Des Sciences Ben M’sick

Centre d'Etudes Doctorales

**Sciences et Applications**



# THESE DE DOCTORAT

**Formation Doctorale : Biologie, Santé et Environnement (BSE)**

# **Laboratoire : Laboratoire Biologie et Santé**

### **Spécialité : Biologie moléculaire et Bioinformatique**

## Présentée par :

# El Khattabi Lamiae

# **Conception de la première chimiothèque marocaine de phytocomposés pour la découverte de médicaments et application des méthodes avancées de la modélisation moléculaire sur la résistin et PPAR $\gamma$ impliqués dans le diabète de type 2 et ses complications cardiovasculaires**

Soutenue le 26/10/ devant le jury composé de :

## Membre du jury


Les travaux présentés dans cette thèse ont été réalisés sous la direction du Pr. SAILE Rachid, directeur du Centre de Recherche en Biotechnologie et Santé, et chef du Laboratoire de Biologie et Santé à la Faculté des Sciences Ben M'Sik de l'Université Hassan II de Casablanca, et sous le co-encadrement du Dr. BARAKAT Abdelhamid, chef du Laboratoire de Génomique et Génétique Humaine de l'Institut Pasteur de Casablanca.



## DEDICACES

*Cette thèse représente ma modeste contribution à la science, avec l'espoir de laisser une empreinte significative directe ou indirecte pour les générations futures ...*

## REMERCIEMENTS

J'adresse mes vifs remerciements à Monsieur le Doyen de la Faculté des sciences Ben M'sik, Université Hassan II de Casablanca, **Pr. Mohammed TALBI**, qui nous a permis de réaliser ce travail dans les meilleures conditions, ainsi que pour le bon déroulement de notre cursus universitaire.

Je tiens à exprimer ma profonde gratitude envers mon directeur de thèse, **Professeur Rachid SAILE**, directeur du Laboratoire de Biologie et Santé et directeur du Centre de Recherche en Biotechnologie et Santé de la Faculté des sciences Ben M'sik. Son soutien infaillible, son engagement sans limites et sa vision toujours éclairante ont été des éléments cruciaux tout au long de cette thèse, nécessitant un temps et un effort considérables. Sa rigueur académique, sa créativité débordante et son ouverture d'esprit ont grandement contribué au succès de ce travail. Les connaissances étendues qu'il possède et sa passion contagieuse pour la recherche ont été une véritable source de motivation pour moi. Sa générosité intellectuelle, associée à une grande humanité, témoignent de son excellence académique et de ses réalisations remarquables en tant que chercheur de premier plan. Sans son soutien précieux et ses conseils avisés, cette thèse n'aurait certainement pas pu être réalisée. Je suis extrêmement reconnaissante envers Pr. **Rachid SAILE** pour son dévouement, son temps et ses efforts, qui ont joué un rôle essentiel dans cette réalisation académique majeure. Sans lui, cette thèse n'aurait certainement pas vu le jour.

J'aimerais également exprimer ma profonde gratitude envers plusieurs autres personnes qui ont contribué de manière significative à la réalisation de cette thèse. Tout d'abord, je tiens à remercier chaleureusement le **Professeur Anas KETTANI**, expert et référence en modélisation moléculaire. C'est grâce à son soutien et à ses conseils éclairés que j'ai pu me lancer dans ce domaine passionnant de recherche. Ses connaissances approfondies et son expertise ont été d'une valeur inestimable tout au long de mon parcours.

Un grand merci au **Docteur Abdelhamid BARAKAT** chef du laboratoire de Génomique et Génétique Humaine de l'Institut Pasteur du Maroc. Je tiens à lui exprimer ma reconnaissance pour m'avoir accueillie au sein de son laboratoire, offrant un environnement propice à la recherche et à l'apprentissage. Il a été pour moi un repère scientifique précieux et a contribué de manière significative à l'avancement de ce travail par sa bienveillance constante.

Je tiens à remercier chaleureusement **Pr. Halima LEBRAZI** et le **Professeur Hassan TAKI**, chercheurs et professeurs au Laboratoire de Biologie et Santé. Leur expertise et leurs précieux conseils ont grandement enrichi cette recherche et ont été d'une valeur inestimable.

Enfin, je tiens à remercier **Salma AATOURI**, **Youssra AALILOU**, **Salwa ZOUHDI** et **Fairouz MOUSTAID**, toutes étudiantes de master et doctorantes au Laboratoire de Biologie et Santé. Leur soutien et leur collaboration dans la vérification des données et dans les échanges constructifs ont grandement contribué à la qualité et à la fiabilité des résultats de cette thèse.

Je souhaite également exprimer ma reconnaissance envers **Mr Omar IJMOUAN**, ingénieur informaticien, pour son aide précieuse dans l'élaboration de la base de données MPDB (Moroccan Phytochemical Database). Son expertise en informatique et sa contribution à la gestion des données ont été essentielles pour le bon déroulement de cette recherche.

Je voudrais remercier spécialement ma famille qui m'a toujours soutenu. Ma douce mère **Mme Laila Alaoui BELGHITI** et mon père **EL KHATTABI Noureddine** qui ont toujours cru en moi et qui ont su m'encourager et me soutenir dans les moments les plus difficiles. Merci à Mes adorables Kaoutar, Malak, Zainab, Amine et Yahya pour leur amour et les encouragements continus prodigués pendant ces six années. Ils ont été d'un grand soutien.

Je suis profondément reconnaissante envers toutes ces personnes, des professeurs émérites aux doctorantes, pour leur engagement, leur expertise et leur soutien inestimable tout au long de cette thèse. Leur contribution collective a été d'une importance capitale, et je leur suis infiniment reconnaissante pour leur collaboration et leur impact positif sur cette recherche.

## PRÉAMBULE

La flore marocaine, riche et diverse, a toujours été au cœur des préoccupations scientifiques, notamment ses plantes aromatiques et médicinales. Dans un monde en quête de thérapies naturelles face aux enjeux médicaux contemporains, il est indiscutablement crucial d'explorer et de valoriser cette biodiversité, surtout face à des maladies touchant des millions de personnes. Ce travail de thèse se subdivise en trois volets majeurs :

Le premier volet, mené en collaboration avec le Laboratoire de Biologie et Santé de la Faculté des Sciences Ben M'Sik, Université Hassan II de Casablanca, et le Laboratoire de Génomique et Génétique Humaine de l'Institut Pasteur du Maroc, s'est consacré à l'analyse des variants du gène RETN liés au diabète de type 2 (DT2) et d'autres maladies métaboliques. L'objectif principal était d'approfondir notre compréhension des mécanismes moléculaires associés à la résistance à l'insuline. En se focalisant sur l'impact de certains variants génétiques sur la structure de la résistin, un protocole de recherche basé sur la modélisation moléculaire et la bioinformatique pour l'évaluation de l'impact des SNPs a été élaboré. Ce protocole a par la suite été adopté par plusieurs doctorants pour d'autres études.

La deuxième phase, en collaboration avec la CNRST et l'Agence Marocaine des Plantes Aromatiques et Médicinales, a porté sur la création de la "Moroccan Phyto-Chemical Database" (MPDB). Soutenue financièrement par la CNRST, cette base de données recense de nombreux phyto-composés marocains issus de la littérature scientifique. Pour faciliter l'accès à cette mine d'informations, la MPDB a été mise en ligne, devenant ainsi un outil inestimable pour la communauté scientifique internationale via le site [www.mpdb.org](http://www.mpdb.org).

Enfin, le troisième volet s'est articulé autour de l'utilisation de la MPDB pour le diabète de type 2 et les maladies cardiovasculaires. Cette partie de la recherche a évalué le potentiel thérapeutique des phyto-composés face aux cibles thérapeutiques associées à ces maladies, mettant en lumière l'importance de la flore marocaine comme réservoir potentiel de molécules thérapeutiques.

Cette thèse, tout en mettant en valeur le patrimoine botanique marocain, a révélé le potentiel thérapeutique des phyto-composés qu'il renferme. Elle esquisse des perspectives prometteuses pour l'élaboration de nouvelles thérapies basées sur les phyto-composés naturels.



## **LISTE DES ABREVIATIONS ET ACRONYMES**

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

DMAO : Découverte de médicaments assistée par ordinateur

COOH : Groupe carboxyle

HCEFLCD : Haut-Commissariat des Eaux et forêts et de la Lutte Contre la Désertification

NH<sub>2</sub> :Groupe amine

## LISTE DES TABLEAUX

<i>Tableau 1. Résultats attendus de précision des SNP de RETN prédis comme délétères dans PredictSNP et outils intégrés.</i>	<b>77</b>
<i>Tableau 2. Détails des nsSNP sélectionnés comme délétères parmi les SNP rapportés, leur analyse de conservation par ConSurf et la prédition des sites de post-traduction par ModPred.</i>	<b>77</b>
<i>Tableau 3. Prédiction du changement de stabilité protéique à l'aide d'I-MUTANT 2 et Mupro.</i>	<b>78</b>
<i>Tableau 4. Valeur RMSD et nombre total de liaisons hydrogène après minimisation de chaque modèle.</i>	<b>80</b>
<i>Tableau 5. Interactions observées entre les trois monomères de la protéine résistin dans les complexes natifs et les 9 mutants.</i>	<b>82</b>
<i>Tableau 6. Les SNPs (UTR mRNA) qui ont été prédis comme étant fonctionnellement significatifs par UTRscan.</i>	<b>82</b>
<i>Tableau 7. Résultats des principaux phyto-composés amarrés contre PPAR<math>\gamma</math>, avec leur affinité de liaison respective, les résidus de liaison hydrogène, et les noms des plantes marocaines sources.</i>	<b>90</b>
<i>Tableau 8. Résultats des meilleurs amarrés contre PPAR<math>\gamma</math>, avec leur affinité de liaison respective, les résidus interagissant par liaisons hydrogène et hydrophobes.</i>	<b>92</b>
<i>Tableau 9. Valeurs moyennes et écarts-types des énergies de liaison des complexes calculées par les méthodes g_mmpbsa.</i>	<b>97</b>
<i>Tableau 11. 100 ligands provenant de l'espèce cannabis sativa et 500 ligands sont des phytocomposés issus des plantes marocaines.</i>	<b>117</b>
<i>Tableau 10. résultats du criblage avec la molécule référentielle.</i>	<b>122</b>
<i>Tableau 12. Analyse d'énergie de liaison et RMSD.</i>	<b>124</b>
<i>Tableau 13. Les ligands du Cannabis sativa et leur analogues</i>	<b>126</b>
<i>Tableau 14. Types des liaisons avec les acides aminés ciblés dans le PCSK9</i>	<b>127</b>
<i>Tableau 15. Les phytocomposés de MPDB qui ont montré une haute potentiel de liaison</i>	<b>129</b>
<i>Tableau 16. Types des liaisons avec les acides aminés ciblés dans le PCSK9</i>	<b>129</b>
<i>Tableau 17. Le triage de ces 5 composés se base sur le respect de 3 critères de Règle de Lipinski au moins.</i>	<b>131</b>

# LISTE DES FIGURES

Figure 1. La formation de la liaison peptidique entre deux acides aminés au niveau de la séquence primaire protéique	6
Figure 2. Structures secondaires, les hélices alpha en haut et feuillets bêta en bas	7
Figure 3. Structure de l'enzyme triosephosphate isomérase	8
Figure 4. Structure 3D de la protéine PCSK9 visualisée en différente représentation par le logiciel pymol	9
Figure 5. Exploitation des structures tertiaires pour la conception des inhibiteurs ou modulateurs, visualisation par pymol, exemple d'enzyme Lanosterol synthase en jaune impliquée dans la synthèse de cholestérol.	10
Figure 6. Diagramme schématique montrant le processus de détermination de la structure 3D des protéines par cristallographie aux rayons X	11
Figure 7. Structure de la protéine ABL1a humaine résolu par la RMN, PDB ID : 6XR7	12
Figure 8. Schéma de la technique de cryo-microscopie électronique (Murata et Wolf 2018).	13
Figure 9. Les étapes de modélisation par homologie.	14
Figure 10. Superposition des deux structures protéiques verte (expérimentale et bleu prédite par alpha fold . (Jumper et al., 2021).	15
Figure 11. Angles dièdres décrivant l'orientation de deux liaisons peptidiques liées au même atome Ca	16
Figure 12. Diagramme de Ramachandran avec les régions les plus favorisées (vert foncé) et les régions supplémentaires autorisées (vert clair).	17
Figure 13. Nombre des structures 3D déterminés expérimentalement et disponible dans la base de données PDB <a href="https://www.rcsb.org">https://www.rcsb.org</a>	18
Figure 14. Exemple d'un fichier pdb, l'atome numéro 80 est un carbone (C) dans la lysine (LYS), résidu numéro 11 de la chaîne A. Ses coordonnées x, y, z sont respectivement 16.865, 23.437 et 0,587.	19
Figure 15. Stratégie globale de conception de médicaments assistée par ordinateur, partant des données génomique et protéique jusqu'à l' identification de lead.	23
Figure 16. Schéma de la classification des méthodes de la DMAO.	24
Figure 17. Les étapes de la modélisation des relations quantitatives structure-activité QSAR.	25
Figure 18. Représentation schématique de principe de docking semi flexible ou la protéine est considéré comme rigide alors que le ligand est flexible.	27
Figure 19. Représentation 2D des composés compromettent les structures qui sont largement signalées comme étant de faux positifs considérés comme PAINS, les Quinones à gauche ou les Rhodanines à droite .(Jasial, Hu, & Bajorath, 2017).	29
Figure 20. Liaison hydrogènes	31
Figure 21. Liaison hydrogènes	31
Figure 22. Forces de van der Waals	32
Figure 23. Schéma des critères de la règle de Lipinski : poids moléculaire (PM), surface polaire (PSA), Angles de rotation , accepteurs de liaisons hydrogène (HBA), donneurs de liaisons hydrogène (HBD) et log P.	34
Figure 24. Prévalence de diabète en 2021 dans le monde selon la fédération internationale de diabète (IDF).	
42	

<b>Figure 25. Classification des diabètes selon l'OMS.</b>	<b>43</b>
<b>Figure 26. Mode d'action des thiazolidinediones suite à son fixation sur le récepteur PPAR<math>\gamma</math>.</b>	<b>44</b>
<b>Figure 27. Liaison entre agoniste complet, rosiglitazone, Les éléments de la structure secondaire sont représentés sous forme de ruban, avec les acides aminés impliqués dans la liaison au ligand. Chaque atome est coloré séparément selon le même schéma de couleurs : vert clair pour le carbone ; rouge pour l'oxygène ; bleu pour l'azote ; jaune pour le soufre. (Nolte et al., 1998).</b>	<b>46</b>
<b>Figure 28. Superposition de complexes PPAR<math>\gamma</math> (gris) avec des agonistes partiels connus (codes pdb : 3D6D, 4PVU, 4PWL, 4JL4, 4JAZ, 4E4K, 2Q5P, 2Q6S, 2Q5S, 4E4Q, 5F9B) sur le complexe PPAR<math>\gamma</math>.</b>	<b>48</b>
<b>Figure 29. Baisse de l'expression de la resistine lors de l' activation de PPAR<math>\gamma</math> par ses ligands exogènes les thiazolidinediones (TZDs).</b>	<b>49</b>
<b>Figure 30. Implication de la résistine dans la survenue de diabète type 2 et d'athérosclérose. IL, interleukine; NF-<math>\kappa</math>B, facteur nucléaire-<math>\kappa</math>B; TNF<math>\alpha</math>, facteur de nécrose tumorale-<math>\alpha</math> (CodoñerFranch et Alonso-Iglesias 2015).</b>	<b>50</b>
<b>Figure 31. Le diabète type 2 (en millions d'individus) en 2000 et prévus pour 2030, adapté de (P Zimmet, Alberti, et Shaw 2001).</b>	<b>51</b>
<b>Figure 32. Rôle de HDL dans le transport inverse de cholestérol et la protection contre la maladie cardiovasculaire (Nayoung et al, 2016).</b>	<b>52</b>
<b>Figure 33. Régulation de taux des LDL dépend de la régulation des récepteurs LDL des cellules de mammifères , Modifié, (Goldstein &amp; Brown, 2009).</b>	<b>54</b>
<b>Figure 34. Schéma illustrant l'internalisation de LDL natives et modifiées par oxydation par les récepteurs des macrophages (Steinberg, 1997).</b>	<b>55</b>
<b>Figure 35. Schéma de la protéine PCSK9, montrant la localisation des domaines en homologie avec le prodomaine et le domaine catalytique de la famille subtilisine des proprotéines convertases.</b>	<b>57</b>
<b>Figure 36. Rôle de PCSK9 dans le recyclage des récepteurs LDL sur la surface cellulaires, figure modifiée, (Soutar &amp; Naoumova, 2007).</b>	<b>58</b>
<b>Figure 37. Relation pharmacocinétique de l'alirocumab, du PCSK9 et du cholestérol LDL (LDL-C) après administration sous-cutanée de l'alirocumab. modifié de (Farnier, 2015), L'alirocumab réduit la PCSK9 libre.</b>	<b>59</b>
<b>Figure 38. Exemple de composés naturels d'huiles essentielles et leurs activités biologiques rapportées avec des applications pharmaceutiques potentielles. (Nascimento, L et, al.2020).</b>	<b>61</b>
<b>Figure 39. Structures chimiques des quatre classes de polyphénoles; Acides phénoliques, flavonoïdes, stilbènes et lignanes (Pandey et Rizvi, 2009).</b>	<b>64</b>
<b>Figure 40. Le modèle tridimensionnel de la résistine humaine généré en utilisant la modélisation par homologie par le serveur web Swiss model. Le modèle est un trimère avec trois chaînes A (rouge), B (verte) et C (bleu).</b>	<b>76</b>
<b>Figure 41. Différence RMSD et liaisons hydrogène totales.</b>	<b>79</b>
<b>Figure 42. Réseau d'interaction protéine-protéine de la résistine en utilisant le serveur STRING.</b>	<b>82</b>
<b>Figure 43. Interaction 3D de la rosiglitazone et des phyto-composés identifiés dans le site actif de PPAR<math>\gamma</math>, liaison H représentée par des pointillés jaunes, (a) rosiglitazone (b) 3,4-DICQA en couleur orange (c) Acide chlorogénique en couleur verte</b>	<b>91</b>
<b>Figure 44. Structures 2D de la rosiglitazone de référence et des deux meilleurs phyto-composés dockés (a) Roziglitazone (b) Acide 3,4-dicaféoylquinique (c) Acide chlorogénique.</b>	<b>92</b>

<b>Figure 45. (a) Tracé RMSD des complexes PPAR<math>\gamma</math>, (b) Tracé du rayon de giration de différents complexes PPAR<math>\gamma</math> avec des panneaux codés en couleur, et (c) Tracé de la surface accessible au solvant (SASA) de différents complexes PPAR<math>\gamma</math> avec des panneaux codés en couleur.</b>	<b>95</b>
<b>Figure 46. Tracé des fluctuations moyennes quadratiques (RMSF) du PPAR<math>\gamma</math> complexé avec deux phyto-composés différents et la rosiglitazone, Apo PPARg en tant que référence pour la simulation de 100 ns. (a) Fluctuation de l'hélice 3 du PPAR<math>\gamma</math>, (b) Fluctuation de H12, et (c) Fluctuation de la feuille bêta.</b>	<b>96</b>
<b>Figure 47. Diagramme 2D des motifs de liaison hydrogène observés pour les deux complexes différents et la rosiglitazone en tant que référence du PPAR<math>\gamma</math> pendant la simulation de 100 ns.</b>	<b>97</b>
<b>Figure 48. Disposition simplifiée du schéma de la base de données MPDB. Les liens entre les tables représentent des relations de plusieurs à plusieurs.</b>	<b>105</b>
<b>Figure 49. Capture d'écran de la page du menu principal. La recherche de phyto-composés peut être effectuée avec différents mots clés comme illustré.</b>	<b>106</b>
<b>Figure 50. Capture d'écran de la page de résumé du composé. L'interface du site web MPDB est montrée, avec une capture d'écran de la partie de la page de résumé du composé pour l'entrée MNPDB00156.</b>	<b>106</b>
<b>Figure 51. Classification des phyto-composés de la MPDB. La classification a été obtenue à partir de ClassyFire.</b>	<b>108</b>
<b>Figure 52. Nombre de violation des propriétés physicochimiques des phyto-composés de la base de données, basées sur la "règle des cinq" de Lipinski.</b>	<b>109</b>
<b>Figure 53. Activités biologiques des phyto-composés de MPDB. Le diagramme circulaire montrant la distribution des différentes activités biologiques des phyto-composés dont l'utilisation est documentée dans leurs références dans MPDB.</b>	<b>110</b>
<b>Figure 54. Schéma de la méthodologie suivi dans le criblage virtuel des données de la MPDB.</b>	<b>113</b>
<b>Figure 55. : Le principe du UniProt et méthode de travail.</b>	<b>114</b>
<b>Figure 56. Des résidus manquants dans la séquence d'acides aminés de la PCSK9.</b>	<b>115</b>
<b>Figure 57. La structure de PCSK9 après avoir complété les acides aminé.</b>	<b>116</b>
<b>Figure 58. Interface de la MPDB de téléchargement des structures 3D des phyto-composés marocains.</b>	<b>117</b>
<b>Figure 59. Visualisation des acides aminés de site actif de la PCSK9 concernant C-Ter.</b>	<b>121</b>
<b>Figure 60. Interface d'interaction entre PCSK9 et LDLR, PCSK9 en doré et LDLR en bleu ciel.</b>	<b>121</b>
<b>Figure 61. Analyse 2D de mode de liaison entre l' inhibiteur identifié <i>in vitro</i> et PCSK9.</b>	<b>122</b>
<b>Figure 62. Diagramme 2D de l'interaction obtenu entre PCSK9 et l'inhibiteur identifié composé 16.</b>	<b>123</b>
<b>Figure 63. Visualisation 2D de l'interaction entre 6-Prenylapigenin et 6-prenylluteolin.</b>	<b>127</b>
<b>Figure 64. Visualisation 2D de l'interaction obtenu par docking moléculaire entre 6-Prenylapigenin et 6-prenylluteolin.</b>	<b>128</b>
<b>Figure 65. Réseau d'interaction entre PCSK9 et deux phyto-composés Excelside B a gauche et apigenin 7-o-rutinoside.</b>	<b>129</b>
<b>Figure 66. Réseau d'interaction entre PCSK9 et deux phyto-composés Excelside B et apigenin7-o-rutinoside.</b>	<b>130</b>



## LISTE DES PUBLICATIONS ET COMMUNICATIONS

**Elkhattabi, Lamiae**, Imane Morjane, Hicham Charoute, Soumaya Amghar, Hind Bouafi, Zouhair Elkarhat, Rachid Saile, Hassan Rouba, et Abdelhamid Barakat. 2019. « *In Silico Analysis of Coding/Noncoding SNPs of Human RETN Gene and Characterization of Their Impact on Resistin Stability and Structure* ». Édité par Jonathan M. Peterson. Journal of Diabetes Research 2019 (mai): 4951627. <https://doi.org/10.1155/2019/4951627>.

**Lamiae Elkhattabi**, Salwa Zouhdi, Fairouz, Abdelhamid Barakat, Rachid Saile. Molecular docking analysis of PPAR $\gamma$  with phytochemicals from Moroccan medicinal plants. Bioinformation 19(7): 795-805 (2023).

**Elkhattabi, L.**, Zouhdi, S., Mousstead, F., Mohtadi, K., Fougrache, H., Badri, W., Taki, H., Lebrazi, H., Kettani, A., & SAILE, R. (September, 2023). Development of Moroccan Phytochemicals DataBase for drug discovery and virtual screening. Soumis à Development of Moroccan Phytochemicals DataBase for drug discovery and virtual screening.

**Elkhattabi, L.**, Morjane, I., Charoute, H., Saile, R. & Barakat, A. Computational screening and analysis of the functional and structural impact of SNPs of the human RETN gene associated to type 2 diabetes. Atherosclerosis 315, e199–e200 (2020).

Morjane, I., Charoute, H., **Elkhattabi, L.**, Saile, R. & Barakat, A. Association of the C. 56C>G APOA5 gene polymorphism with coronary artery disease: Moroccan case-control study and an updated meta-analysis. Atherosclerosis 315, e189 (2020).

## Résumé

Dans un premier temps, l'étude du gène RETN a permis de mettre en évidence l'impact potentiel de SNP délétères sur la résistine, une protéine impliquée dans les troubles métaboliques. L'analyse des SNP non-sens dans le gène RETN a été approfondie, en portant une attention particulière aux SNP hautement conservés susceptibles d'influer sur la stabilité de la protéine. Ces découvertes suggèrent que certains de ces SNP pourraient jouer un rôle clé dans les voies métaboliques associées à la résistine. Dans une seconde partie, la recherche s'est orientée vers les composés extraits de plantes marocaines, en explorant leurs propriétés thérapeutiques potentielles, notamment leurs activités antimicrobiennes, anticancéreuses et antidiabétiques. Le 3,4-DICQA, l'acide chlorogénique et le lutéoloside ont émergé comme des candidats prometteurs pour améliorer la sensibilité à l'insuline via PPAR $\gamma$ , une cible majeure dans le diabète de type 2. Enfin, le projet a abouti à la création de la base de données des phytochimiques marocains (MPDB), fournissant des informations essentielles sur 710 phytochimiques, dont leurs structures 3D, leurs propriétés physico-chimiques, leurs activités biologiques, ainsi que leurs résultats de docking. Cette base de données offre une plateforme pour explorer les mécanismes d'action des phytochimiques marocains et facilite l'accès aux analogues commerciaux via Mcule. En conclusion, ce travail de thèse offre des perspectives passionnantes pour la compréhension des mécanismes pathologiques et la découverte de médicaments potentiels, tout en renforçant le potentiel des phytochimiques marocains dans le domaine de la santé.

# SOMMAIRE

<b>PRÉAMBULE</b>	<b>III</b>
<b>DEDICACES</b>	<b>V</b>
<b>REMERCIEMENTS</b>	<b>VI</b>
<b>LISTE DES ABREVIATIONS ET ACRONYMES</b>	<b>VIII</b>
<b>LISTE DES TABLEAUX</b>	<b>IX</b>
<b>LISTE DES FIGURES</b>	<b>X</b>
<b>LISTE DES PUBLICATIONS ET COMMUNICATIONS</b>	<b>XIII</b>
<b>SOMMAIRE</b>	<b>XIV</b>
<b>INTRODUCTION GENERALE</b>	<b>1</b>
<b>PARTIE I. REVUE BIBLIOGRAPHIQUE</b>	<b>4</b>
CHAPITRE 1. BIOINFORMATIQUE ET MODÉLISATION MOLÉCULAIRE	5
Introduction	5
Les protéines	5
L'importance de la structure tertiaire	9
Prédiction d'impact des variants sur la structure et la fonction protéiques	20
Découverte de médicament assistée par ordinateur	22
La règle de Lipinski	34
Simulation par dynamique moléculaire	35
CHAPITRE 2. LE DIABÈTE DE TYPE 2	41
1. Introduction	41

2. Le Diabète de type 2	41
3. PPAR- $\gamma$ : Le récepteur activé par les proliférateurs de peroxysomes gamma	45
La résistin	48
Complications cardiovasculaires de Diabète type 2	51
<b>CHAPITRE 4. LES PHYTO-COMPOSÉS DANS LE TRAITEMENT DE DT2 ET SES COMPLICATIONS ET BASE DE DONNÉES</b>	<b>60</b>
1. Phyto-composés	60
Phytothérapies	64
Les plantes aromatiques et médicinales du Maroc	65
Bases de données des phyto composés internationales	66
<b>OBJECTIFS DE LA THÈSE</b>	<b>68</b>
<b>PARTIE II : TRAVAUX DE RECHERCHE</b>	<b>69</b>
CHAPITRE 1. MODELISATION MOLECULAIRE DE LA RÉSISTINE HUMAINE IMPLIQUÉ DANS LE DIABÈTE DE TYPE II	71
1. Introduction	71
Matériel et Méthodes	72
Résultats	76
Discussion	83
Conclusion	84
CHAPITRE 2. PPAR ET CRIBLAGE VIRTUEL	86
1. Introduction	86
Méthodes	88
Résultats	90
Discussion	98
Conclusion	100
CHAPITRE 3. CONCEPTION DE LA PREMIÈRE BASE DE DONNÉES DES COMPOSÉS ISSUS DES PAMS	102
Introduction	102
Méthodes	103
Résultats	105
<i>Discussion</i>	110
<i>Conclusion</i>	112
CHAPITRE 4. EXPLOITATION DES DONNÉES DE LA MPDB VIS À VIS LA PCSK9	112
1. Introduction	112
1. Méthodes	114
<i>Résultats</i>	120
<i>Visualisation 3D de la protéine PCSK9</i>	120
Docking moléculaire de composé 16 l'inhibiteur connu DE PCSK9	121

La sélection de ces 5 composés repose sur la satisfaction d'au moins 3 critères de la règle de Lipinski.	131
Discussion	131
<b>CONCLUSION GÉNÉRALE</b>	<b>134</b>
<b>BIBLIOGRAPHIE GÉNÉRALE</b>	<b>136</b>
<b>INDEX DES AUTEURS</b>	<b>158</b>
<b>TABLE DE MATIÈRES</b>	<b>168</b>

## INTRODUCTION GENERALE

Le diabète de type 2 est une affection métabolique chronique caractérisée par une résistance à l'insuline et une insuffisance relative de la sécrétion d'insuline. Il est bien établi que les personnes atteintes de diabète de type 2 présentent un risque accru de maladies cardiovasculaires (MCV), notamment de coronaropathie, d'accident vasculaire cérébral et de maladie vasculaire périphérique. Selon l'American Heart Association (AHA), les adultes diabétiques sont deux à quatre fois plus susceptibles de mourir d'une maladie cardiaque que ceux qui ne le sont pas. De plus, l'Organisation mondiale de la santé (OMS) a identifié le diabète comme un facteur de risque majeur pour les MCV. La coexistence du diabète et des MCV aggrave le pronostic, rendant la prise en charge clinique complexe. La gravité de cette association est soulignée par le fait que la majorité des personnes atteintes de diabète de type 2 meurent de complications cardiovasculaires. Les défis actuels comprennent la nécessité d'une meilleure compréhension des mécanismes sous-jacents et du développement de stratégies thérapeutiques efficaces pour prévenir et traiter les MCV chez les patients diabétiques. Les études d'association pangénomique (GWAS) ont révélé plusieurs variants génétiques qui augmentent le risque de ces pathologies. De plus, le gène RETN, qui code pour la résistin, une hormone produite par les adipocytes, a également été étudié en relation avec le DT2 et les MCV. Au fil des recherches, une association solide a été établie entre des niveaux élevés de résistin et le risque accru de diabète de type 2 (DT2) (You et al., 2006; Silha et al., 2003). Ces découvertes ont ouvert des perspectives prometteuses pour renforcé la compréhension de lien entre les MCV et la physiopathologie du DT2 . Cependant, malgré ces avancées, un domaine demeure insuffisamment exploré. Bien que la relation entre la résistin et le DT2 soit bien établie, la littérature actuelle présente des lacunes concernant l'impact des variants génétiques de la résistin sur cette relation et la survenue des MCV. Il est donc impératif d'évaluer l'impact de ces variants génétiques sur la pathologie du DT2. Une telle évaluation pourrait non seulement enrichir la compréhension actuelle de la maladie, mais aussi ouvrir la voie à des stratégies thérapeutiques plus ciblées.

La modélisation moléculaire et la bioinformatique jouent un rôle crucial dans l'évaluation de l'impact des variants génétiques sur la structure tridimensionnelle des protéines et, par conséquent, sur leur

fonction. Les variants génétiques, tels que les mutations ponctuelles, peuvent entraîner des changements dans la conformation des protéines, affectant ainsi leur stabilité, leur dynamique et leur capacité à interagir avec d'autres molécules. Grâce à la modélisation moléculaire, il est possible de prédire comment une mutation spécifique peut influencer la structure tridimensionnelle d'une protéine. La bioinformatique offre des outils pour analyser et interpréter d'énormes ensembles de données génomiques, permettant d'identifier des variants potentiellement délétères et de prédire leur impact sur la fonction protéique. Ces techniques fournissent des informations précieuses pour comprendre les mécanismes moléculaires sous-jacents des maladies et pour concevoir des interventions thérapeutiques ciblées. Aujourd'hui, l'utilisation de la technologie moderne de chimio-informatique permet d'accélérer la recherche, offrant des économies de temps et d'argent pour de meilleurs résultats. La découverte de médicaments assistée par ordinateur (DMAO) demeure un moyen efficace pour identifier les principaux composés potentiels pour un large éventail de maladies. Un certain nombre d'approches informatiques sont utilisées pour identifier les nouveaux composés ayant un potentiel thérapeutique à partir d'énormes bibliothèques de composés. L'approche DMAO progresse quotidiennement. La tendance récente dans la conception de médicaments est de concevoir de manière rationnelle des thérapies puissantes avec des effets multi-ciblés, des efficacités plus élevées et moins d'effets secondaires, notamment en termes de toxicité.

Ces techniques informatiques permettent de simuler et d'analyser les interactions entre des molécules et des cibles biologiques spécifiques, telles que les protéines ou les récepteurs. Dans le contexte du diabète de type 2 et des MCV, où plusieurs protéines et récepteurs sont impliqués dans la pathogenèse, citant l'exemple de PPAR $\gamma$  (Peroxisome Proliferator-Activated Receptor gamma), récepteur nucléaire qui joue un rôle crucial dans la régulation du métabolisme des lipides et du glucose. Les thiazolidinediones (TZD) sont une classe de médicaments qui activent spécifiquement le PPAR $\gamma$  et sont utilisées dans le traitement du diabète de type 2 (DT2) pour améliorer la sensibilité à l'insuline. Cependant, l'utilisation des TZD est associée à plusieurs effets indésirables, tels que la prise de poids, la rétention d'eau, une augmentation du risque de fractures et potentiellement une insuffisance cardiaque. Face à ces défis, les composés naturels offrent une alternative prometteuse. Ces molécules, souvent présentes dans les aliments ou les plantes médicinales, peuvent activer le PPAR $\gamma$  avec une moindre incidence d'effets secondaires. L'avantage des composés naturels réside dans leur biodisponibilité, leur faible toxicité et leur capacité à cibler plusieurs voies métaboliques simultanément. Cependant, la recherche continue pour identifier des composés naturels efficaces et sûrs pour le traitement du DT2 via l'activation du PPAR $\gamma$ . Les plantes aromatiques et médicinales (PAM) ont été utilisées depuis longtemps pour traiter les maladies cardiovasculaires. À présent, les médicaments utilisés dans le traitement de divers cancers, maladies cardiovasculaires, diabète, etc., proviennent souvent de PAM ou de leurs dérivés, comme l'aspirine et la réserpine, qui représentent

encore des médicaments courants dans le monde. Les PAM sont devenues l'une des ressources les plus importantes pour développer de nouveaux composés thérapeutiques. Récemment, plusieurs articles scientifiques ont été publiés dans des revues à comité de lecture décrivant les effets positifs des PAM sur le processus de guérison de diverses maladies humaines et animales. En effet, entre 1981 et 2014, plus de 50% des médicaments nouvellement développés ont été élaborés à partir de produits naturels, principalement des plantes (Newman DJ et al., 2016). Le Maroc est caractérisé par une diversité florale spécifique et originale. D'après l'Agence nationale des plantes médicinales et aromatiques (ANPMA), environ 4 200 variétés ont été identifiées au niveau national, dont 600 plantes possèdent des propriétés médicinales et aromatiques. Cela a permis au Maroc d'occuper la deuxième place mondiale après la Turquie. La diversité florale marocaine est utilisée par des millions de personnes à des fins médicinales. Cependant, l'utilisation médicinale de la plupart de ces composés n'a pas encore été pleinement explorée. Bien que des bases de données de petites molécules librement accessibles, contenant des composés isolés de sources florales, aient été établies dans de nombreux pays, il n'existe actuellement aucune base de données en ligne de ce type au Maroc. Les découvertes de nouveaux composés thérapeutiques provenant de PAM marocaines à l'aide d'approches avancées de DMAO sont également limitées en raison de l'indisponibilité de structures 3D précises. D'où le besoin et la nécessité de la mise en place d'une base de données marocaine organisée intégrant toutes les informations pertinentes pour faciliter l'exploration de l'importance médicinale des composés issus de la biodiversité marocaine. Cette base permettrait de fournir aux communautés scientifiques nationales et mondiales les phyto-composés du Maroc avec des informations pertinentes sur les structures et les propriétés pour les études de DMAO, permettant ainsi l'exploration et l'exploitation des phyto-composés du Maroc. Ainsi, pour les études de DMAO accélère considérablement le processus de découverte de médicaments, réduisant le temps et les coûts associés à la recherche de nouveaux agents thérapeutiques pour traiter le diabète de type 2 et les MCV.

## PARTIE I. REVUE BIBLIOGRAPHIQUE

# CHAPITRE 1. BIOINFORMATIQUE ET MODÉLISATION MOLÉCULAIRE

## 1.1 Introduction

La Bioinformatique est une discipline scientifique qui applique l'informatique, les mathématiques et la statistique à l'étude de la biologie. Elle permet de stocker, traiter, analyser et interpréter de grandes quantités de données biologiques générées par des expériences ou des séquençages de l'ADN ou de l'ARN. Grâce à cette approche, on peut mieux comprendre les processus biologiques et faire de nouvelles découvertes dans les domaines de la biologie, de la médecine et de la pharmacologie. La Bioinformatique est donc devenue une discipline essentielle pour les chercheurs et les scientifiques, leur offrant des outils et des méthodes pour explorer le monde biologique à une échelle sans précédent grâce à des algorithmes sophistiqués et à des bases de données génomiques, la Bioinformatique permet de détecter et d'analyser les variations génétiques, ainsi que d'étudier leurs conséquences fonctionnelles (Koonin et al., 2013). Par ailleurs, la prédition de structures protéiques, notamment grâce à des méthodes de modélisation comparative et ab initio, facilite la compréhension des mécanismes moléculaires et des interactions protéine-protéine (Baker & Sali, 2001).

La découverte de médicaments assistée par ordinateur (DMAO) représente également un domaine d'application majeur pour la Bioinformatique et la modélisation moléculaire. Ces techniques permettent d'identifier des cibles thérapeutiques potentielles, de concevoir des molécules actives et de sélectionner des candidats médicaments en optimisant leur affinité, leur sélectivité et leur pharmacocinétique (Ferreira et al., 2015). La DMAO a contribué à l'accélération du processus de découverte de médicaments et à la réduction des coûts associés, tout en améliorant leur efficacité et leur sécurité (Paul et al., 2010). La Bioinformatique et la modélisation moléculaire sont des outils indispensables pour l'étude des mutations, la prédition des structures protéiques et la découverte de médicaments assistée par ordinateur. Leur impact sur la recherche fondamentale et appliquée a transformé notre approche de la biologie et ouvert la voie à de nouvelles avancées thérapeutiques.

## 1.2 Les protéines

Les protéines sont des biomolécules qui ont des fonctions diverses et variées. Elles peuvent être des enzymes qui catalysent des réactions biochimiques, des protéines structurelles qui maintiennent l'intégrité structurelle des membranes cellulaires, des protéines de motilité qui facilitent les mouvements cellulaires et de l'organisme, et des protéines membranaires qui agissent comme des récepteurs ou des canaux de transport d'ions. La variété des fonctions des protéines est due à leur capacité à former des structures tridimensionnelles grâce aux propriétés distinctes des vingt acides aminés naturels qui composent les polypeptides.

La compréhension des processus de repliement des protéines est essentielle pour comprendre leur structure et leur fonction, et cela peut avoir des implications importantes pour la biologie, la

médecine et la conception de médicaments et de thérapies. Les efforts de recherche dans ce domaine ont utilisé des méthodes expérimentales et informatiques pour étudier les caractéristiques des séquences protéiques qui les amènent à se replier correctement dans leur structure tridimensionnelle préférée. (Dill & MacCallum, 2012; Kar & Mishra, 2011; Rohl et al., 2004).

### 1.3 LES NIVEAUX D'ORGANISATION DES PROTÉINES

#### 1.3.1 STRUCTURE PRIMAIRE

La structure primaire est constituée par la séquence linéaire d'acides aminés qui compose une protéine, reliés par des liaisons peptidiques (Pauling et al., 1951). Une liaison peptidique se forme lorsque le groupe carboxyle ( $\text{COOH}$ ) d'un acide aminé réagit avec le groupe amine ( $\text{NH}_2$ ) d'un autre acide aminé, libérant une molécule d'eau et créant une liaison covalente entre les deux acides aminés. Cette liaison est stable et présente des propriétés géométriques particulières, telles que la coplanarité des atomes impliqués. La séquence unique d'acides aminés détermine la manière dont la protéine se replie pour former des structures secondaires, tertiaires et quaternaires, qui à leur tour, influencent la fonction biologique de la protéine.

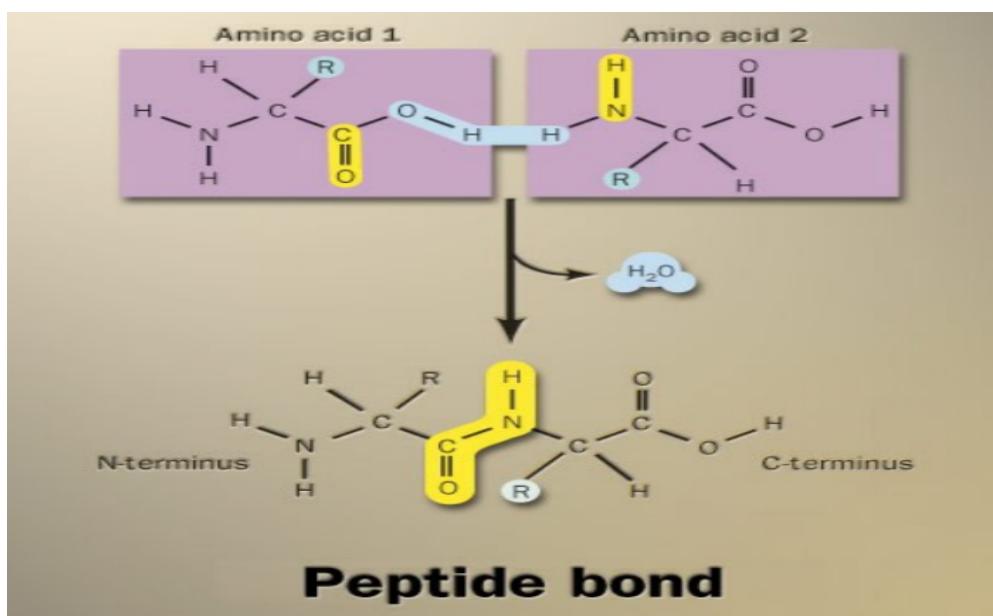


FIGURE 1. LA FORMATION DE LA LIAISON PEPTIDIQUE ENTRE DEUX ACIDES AMINÉS AU NIVEAU DE LA SÉQUENCE PRIMAIRES

### 1.3.2 STRUCTURE SECONDAIRE

La structure secondaire des protéines est l'organisation régulière d'une partie de la chaîne polypeptidique. Elle est maintenue par des liaisons hydrogène entre les acides aminés qui ne sont pas voisins dans la chaîne. Les deux types principaux de structures secondaires sont l'hélice  $\alpha$  et le feuillet  $\beta$ . L'hélice  $\alpha$  est une structure enroulée en forme de spirale, tandis que le feuillet  $\beta$  est une organisation de brins pliés (figure 1). Les coudes et les boucles constituent également des éléments structuraux importants dans les protéines, permettant de connecter les différentes parties de la chaîne. Ils sont souvent constitués de courbes et de torsions, qui peuvent être stabilisées par des liaisons hydrogène et d'autres interactions entre les acides aminés (Alberts et al., 2015).

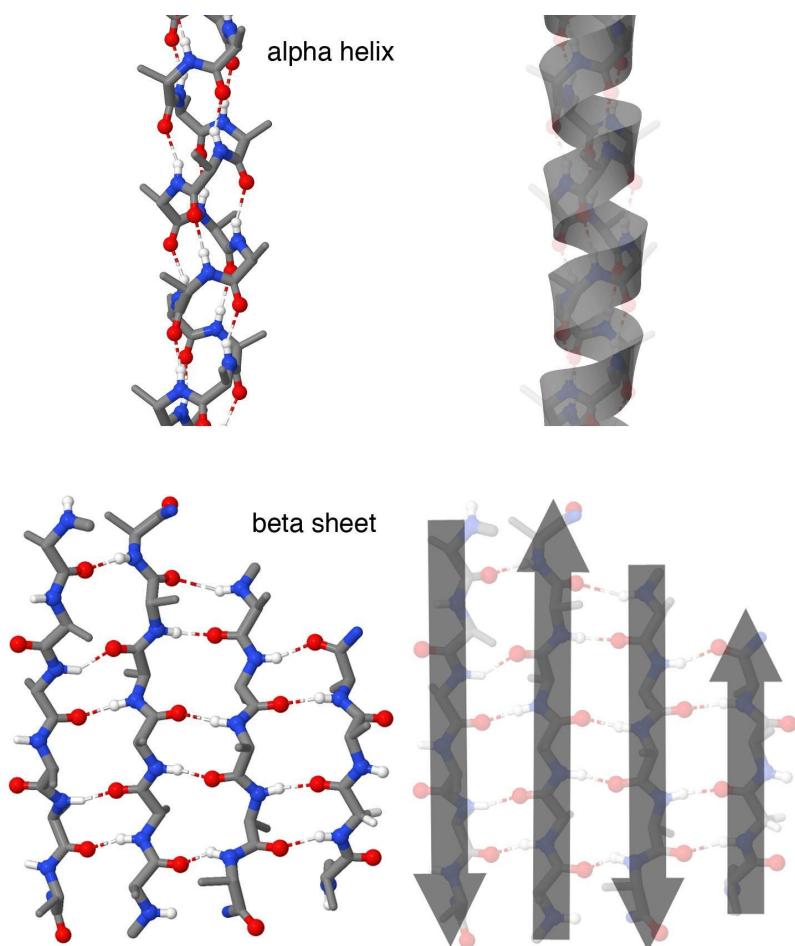


FIGURE 2. STRUCTURES SECONDAIRES, LES HÉLICES ALPHA EN HAUT ET FEUILLETS BÊTA EN BAS

Les boucles sont des éléments clés de la structure des protéines, car elles peuvent avoir une grande influence sur la fonction et la stabilité de la protéine. Elles peuvent également jouer un rôle dans la reconnaissance et la liaison à d'autres molécules. Les boucles peuvent être plus flexibles que les hélices  $\alpha$  et les feuillets  $\beta$ , et peuvent donc contribuer à la capacité de la protéine à se replier et à se déplier en réponse à des signaux environnementaux (Voet et al., 2016).

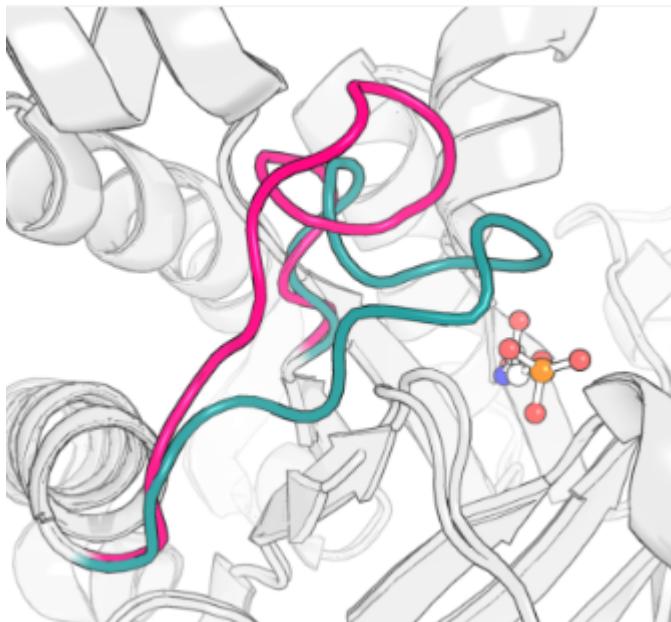


FIGURE 3. STRUCTURE DE L'ENZYME TRIOSEPHOSPHATE ISOMÉRASE

Lorsque le substrat se lie, une boucle passe d'une conformation "ouverte" (rose, entrée PDB 1TPD) à une conformation "fermée" (verte, 1TRD), ce qui empêche l'accès du solvant au site actif et stabilise le composé intermédiaire de la réaction.

#### 1.3.3 STRUCTURE TERTIAIRE

La structure tertiaire d'une protéine est définie comme étant la description de son repliement, c'est-à-dire de l'agencement tridimensionnel de ses boucles et structures secondaires régulières, ainsi que des liaisons covalentes post-traductionnelles telles que les ponts disulfures, ou de tout autre élément comme les ions. Ce repliement est essentiel pour la fonction biologique de la protéine. Les liaisons qui maintiennent la structure tertiaire des protéines peuvent être de différents types. Les liaisons covalentes telles que les ponts disulfures et les liaisons peptidiques sont responsables de la stabilité de la structure. Les interactions non-covalentes telles que les liaisons ioniques, les liaisons hydrogène, les forces de van der Waals et les interactions hydrophobes contribuent également à la formation de la structure tridimensionnelle des protéines. Les interactions hydrophobes sont particulièrement importantes car elles permettent aux acides aminés hydrophobes de se regrouper au centre de la protéine, ce qui minimise leur exposition à l'eau. Les liaisons ioniques et les liaisons hydrogène sont également importantes pour la stabilité de la structure tertiaire des protéines, car elles peuvent aider à maintenir les interactions entre les différents domaines de la protéine. La structure tertiaire peut être influencée par des facteurs environnementaux tels que le pH, la température et la présence d'autres molécules (Alberts B et al. 2015). Des modifications de la structure tertiaire peuvent avoir des conséquences fonctionnelles et sont associées à plusieurs maladies.

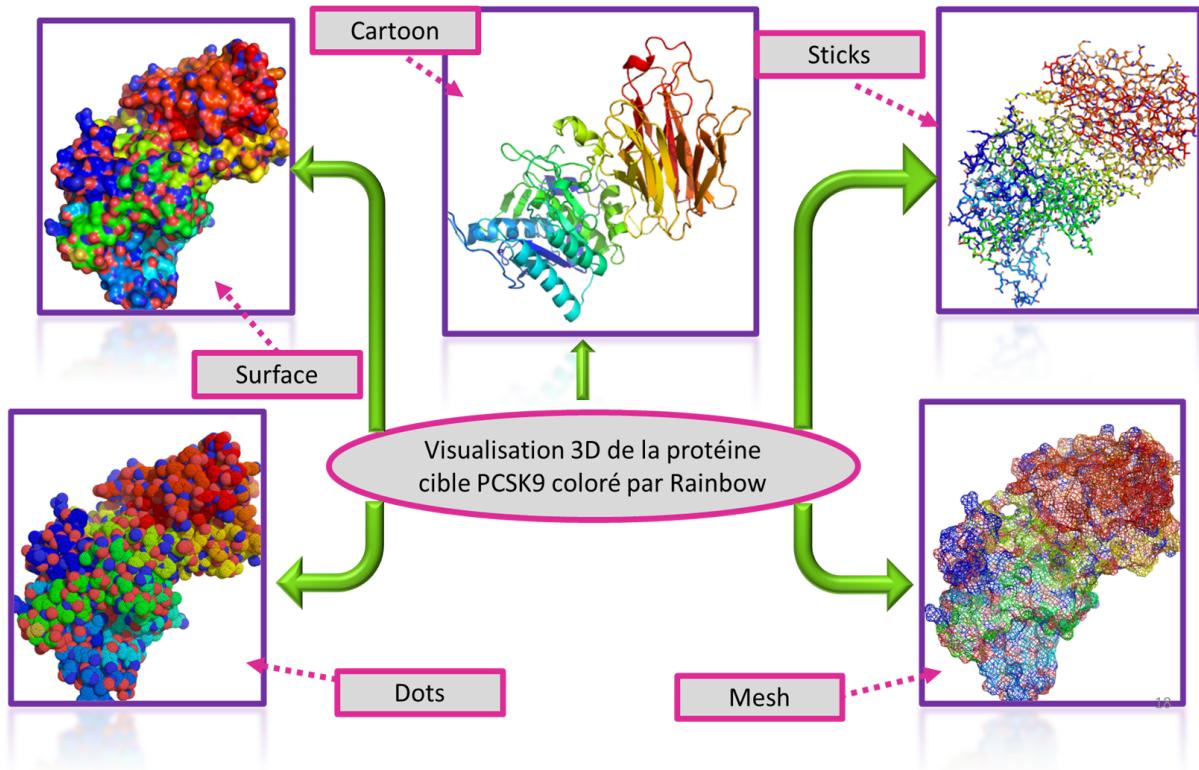
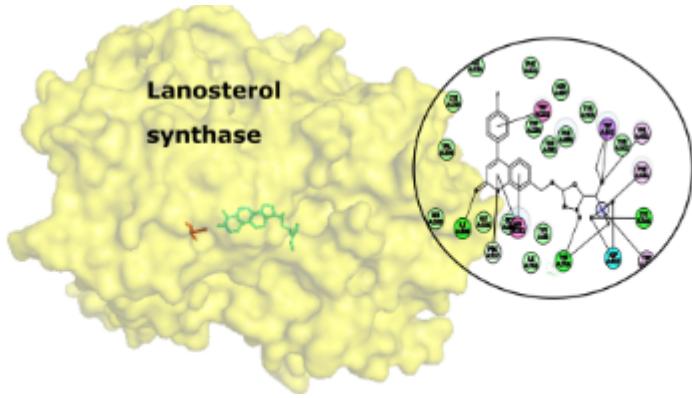


FIGURE 4. STRUCTURE 3D DE LA PROTÉINE PCSK9 VISUALISÉE EN DIFFÉRENTE REPRÉSENTATION PAR LE LOGICIEL PYMOL

#### 1.4 L'importance de la structure tertiaire

La structure tridimensionnelle des molécules biologiques, en particulier des protéines, joue un rôle crucial dans la compréhension des mécanismes biologiques et la découverte de médicaments. Les structures tridimensionnelles des protéines permettent de comprendre les mécanismes moléculaires et les interactions entre protéines et ligands, facilitant ainsi la conception de molécules thérapeutiques ciblées et efficaces (Harrison, 2016). Par exemple, la résolution de la structure tridimensionnelle de la protéase du VIH-1 a permis de concevoir des inhibiteurs de cette enzyme, tels que l'indinavir, un médicament essentiel dans le traitement du SIDA (Erickson et al., 1990). De même, la détermination de la structure de la kinase B-Raf, impliquée dans certains cancers, a conduit au développement du vemurafenib, un inhibiteur spécifique utilisé pour le traitement du mélanome métastatique (Bollag et al., 2010). Ces exemples illustrent comment la connaissance des structures tertiaires des protéines peut mener à des avancées thérapeutiques significatives.



**FIGURE 5. EXPLOITATION DES STRUCTURES TERTIAIRES POUR LA CONCEPTION DES INHIBITEURS OU MODULATEURS, VISUALISATION PAR PYMOL, EXEMPLE D'“ENZYME LANOSTEROL SYNTHASE EN JAUNE IMPLIQUÉ DANS LA SYNTHÈSE DE CHOLESTROL.**

## 1.5 MÉTHODE EXPÉRIMENTALE DE LA DÉTERMINATION DES STRUCTURES TERTIAIRES

### 1.5.1 X-RAY DIFFRACTION

La diffraction des rayons X est une méthode largement utilisée pour déterminer la structure tridimensionnelle des protéines. Le principe de cette technique repose sur l'interaction des rayons X avec la matière, et plus précisément avec les atomes d'une protéine cristallisée. Lorsque les rayons X traversent le cristal, ils sont diffractés par les atomes qui composent la structure, produisant un motif de diffraction complexe. En analysant ce motif, il est possible de reconstruire la structure tridimensionnelle de la protéine à l'échelle atomique (Drenth, 2007) figure 5 Parmi les avantages de la diffraction des rayons X, on peut citer la résolution élevée qu'elle permet d'atteindre, généralement inférieure à 3 Å, ce qui est suffisant pour déterminer la position des atomes dans la protéine et comprendre les mécanismes moléculaires en jeu (Rupp, 2009). De plus, la diffraction des rayons X peut être utilisée pour étudier une grande variété de protéines, y compris des complexes protéiques et des protéines liées à des ligands.

Cependant, la méthode présente également des inconvénients. Le principal défi de la diffraction des rayons X est la nécessité d'obtenir des cristaux de protéines de qualité, un processus souvent laborieux et difficile à maîtriser (McPherson, 2004). De plus, certaines protéines ne cristallisent pas facilement ou présentent des cristaux de mauvaise qualité, rendant la détermination de leur structure tridimensionnelle par diffraction des rayons X difficile, voire impossible.

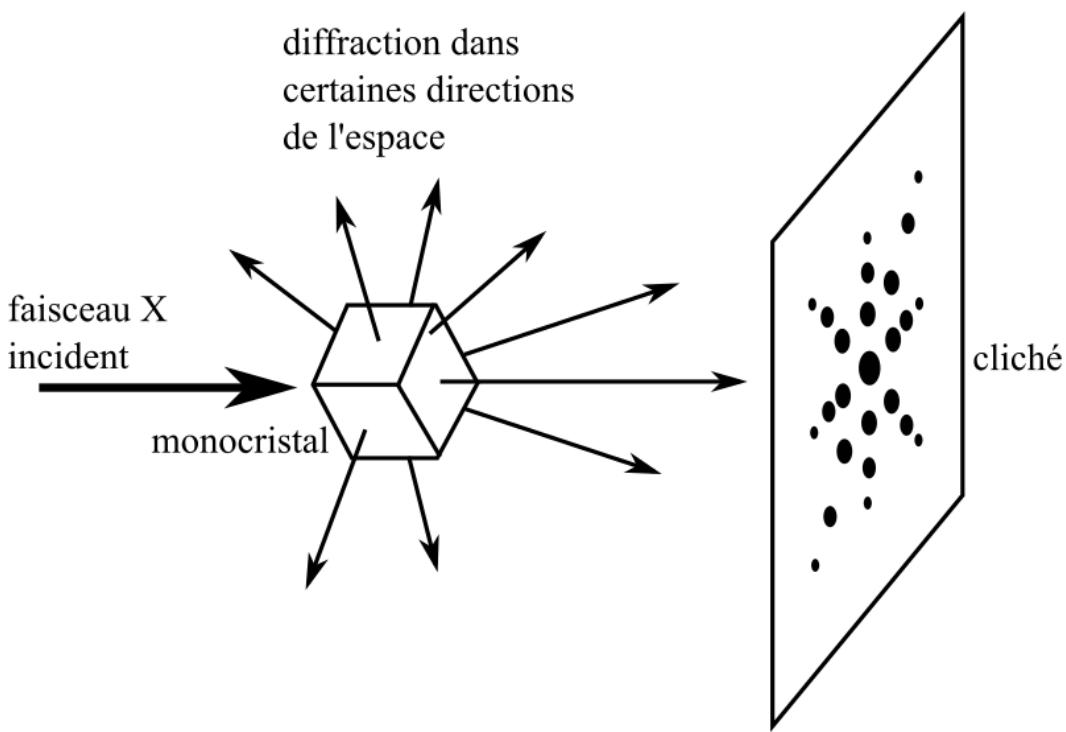


FIGURE 6. DIAGRAMME SCHÉMATIQUE MONTRANT LE PROCESSUS DE DÉTERMINATION DE LA STRUCTURE 3D DES PROTÉINES PAR CRISTALLOGRAPHIE AUX RAYONS X

### 1.5.2 RÉSONANCE MAGNÉTIQUE NUCLÉAIRE

La Résonance Magnétique Nucléaire (RMN) est une technique expérimentale qui permet d'obtenir un ensemble de contraintes de distances et d'angles sur les acides aminés constituant une protéine. En combinant ces contraintes spécifiques à la protéine avec des données génériques, il est possible d'obtenir un ensemble de modèles de la protéine satisfaisant ces contraintes. Le principe de la RMN est basé sur le comportement des noyaux atomiques en présence d'un champ magnétique intense, qui leur permet de s'orienter de manière spécifique. Les signaux émis par les noyaux atomiques lorsqu'ils sont soumis à un champ magnétique, permettent de déterminer leur position et leur orientation dans l'espace. L'avantage de cette technique expérimentale est qu'elle permet d'étudier la protéine en solution, ce qui permet d'observer sa structure dans un milieu dynamique, contrairement à la Radio-cristallographie qui fige la protéine dans un cristal. En effet, la protéine étant libre de ses mouvements en solution, la fluctuation de son repliement engendre un ensemble de contraintes reflétant cette mobilité. Ainsi, la RMN permet de fournir un ensemble de modèles pour la protéine, alors que la Radio-cristallographie propose un modèle unique. L'inconvénient de la RMN est que sa résolution est limitée par la taille de la protéine, et donc, elle n'est pas adaptée pour l'étude des protéines de grande taille. De plus, le temps de mesure pour la RMN est plus long que celui de la radio-cristallographie. (Sattler, M., et al, 1999).

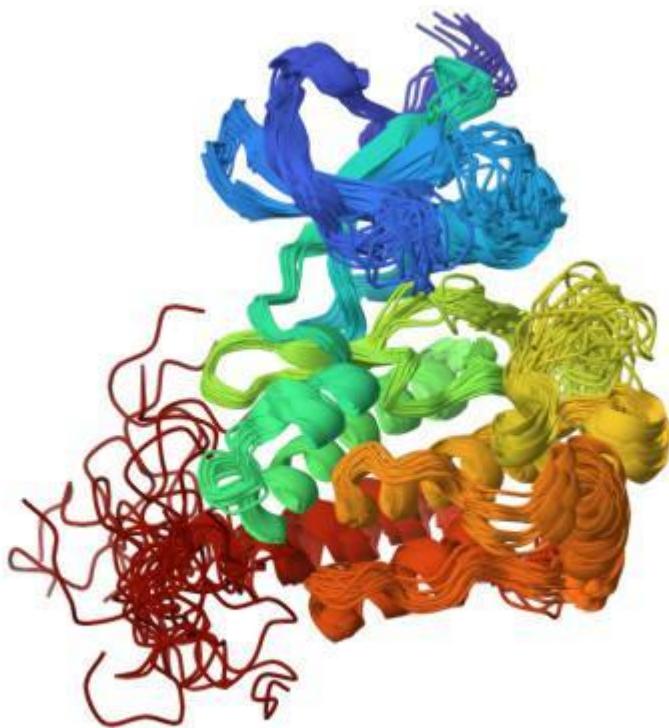


FIGURE 7. STRUCTURE DE LA PROTÉINE ABL1A HUMAINE RÉSOLU PAR LA RMN, PDB ID : 6XR7

### 1.5.3 MICROSCOPE ÉLECTRONIQUE

Le microscope électronique est une technique d'imagerie permettant d'étudier la structure des protéines à haute résolution (Bai et al., 2015). L'avantage principal de cette technique est sa haute résolution, qui peut atteindre une résolution sub-angstromique, permettant de visualiser les détails de la structure protéique, même au niveau des atomes individuels (Cheng, 2018). De plus, la préparation des échantillons pour le microscope électronique est plus rapide et moins complexe que pour la cristallographie aux rayons X, ce qui permet d'obtenir des résultats plus rapidement. Cependant, le microscope électronique présente également des inconvénients. Tout d'abord, la préparation des échantillons peut être délicate, car les protéines doivent être fixées sur une grille métallique et être suffisamment fines pour permettre la traversée des électrons (Cheng, 2018). De plus, l'imagerie par microscope électronique nécessite souvent l'utilisation de vitrification, ce qui peut provoquer des distorsions et des changements de conformation de la protéine (Liao et al., 2013). Enfin, la résolution atteinte par le microscope électronique peut être limitée par la taille de la protéine étudiée (Bai et al., 2015). Malgré ces inconvénients, le microscope électronique reste une technique d'imagerie puissante pour la détermination de la structure 3D des protéines. En combinaison avec d'autres techniques telles que la cristallographie aux rayons X, la RMN ou la spectroscopie de masse, le microscope électronique peut fournir des informations précieuses sur la structure et la fonction des protéines (Bai et al., 2015).

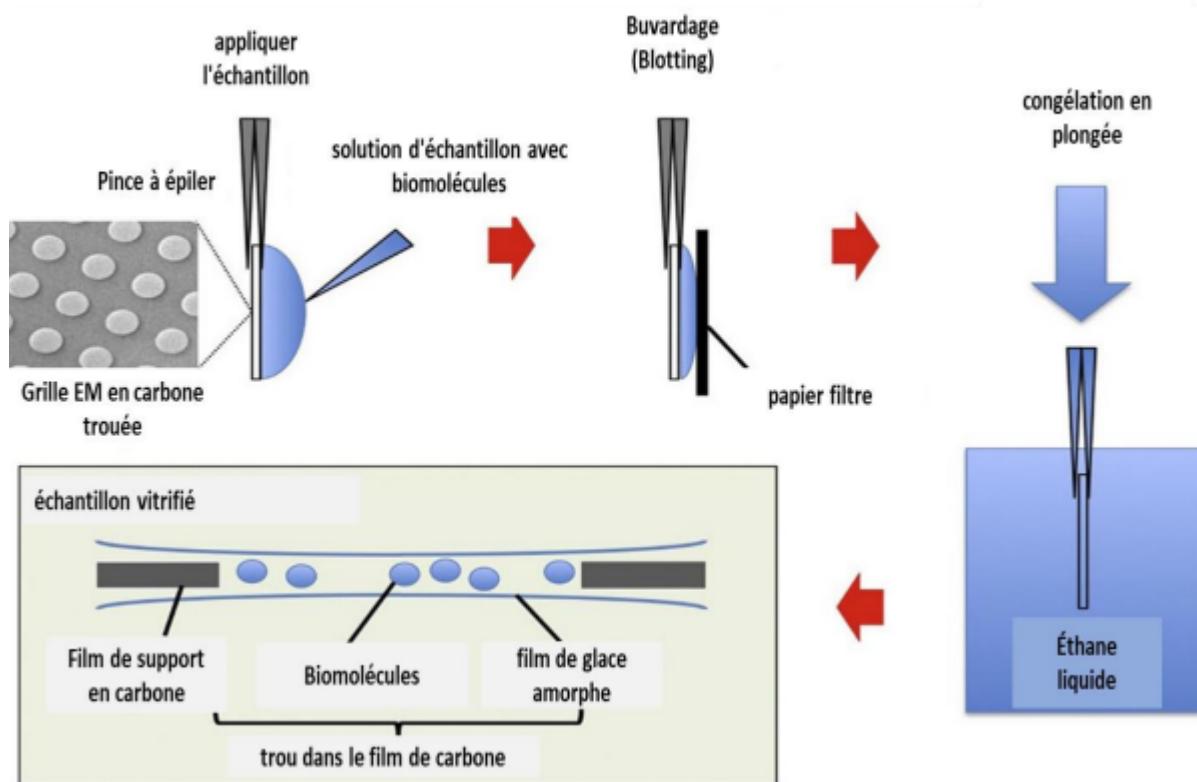


FIGURE 8. SCHÉMA DE LA TECHNIQUE DE CRYO-MICROSCOPIE ÉLECTRONIQUE (MURATA ET WOLF 2018).

### 1.6 PRÉDICTION DE LA STRUCTURE TERTIAIRE PAR LA BIOINFORMATIQUE

La structure 3D des protéines est essentielle pour comprendre leur fonction et leur rôle vital. Bien que les efforts aient permis de résoudre environ 100 000 structures (Jumper et al., 2021), cela ne représente qu'une petite partie des séquences protéiques existantes. Ainsi, les approches de prédiction de la structure des protéines, un enjeu majeur en bioinformatique, peuvent être divisées en quatre catégories principales que nous traiterons dans ce paragraphe.

#### 1.6.1 MÉTHODE DE MODÉLISATION PAR HOMOLOGIE

Le principe des méthodes basées sur les modèles, également appelées méthodes d'homologie ou de modélisation comparée, repose sur l'identification d'une protéine de structure connue ayant une séquence similaire à la protéine cible. L'hypothèse sous-jacente est que des séquences similaires adoptent généralement des structures tridimensionnelles similaires. La structure 3D de la protéine cible est prédictive en utilisant la structure de cette protéine homologue comme modèle et en ajustant les régions divergentes pour s'adapter à la séquence cible (Sali & Blundell, 1993).

Plusieurs logiciels utilisent ce principe pour prédire la structure 3D des protéines. En plus de MODELLER, d'autres exemples incluent SWISS-MODEL et Phyre2.

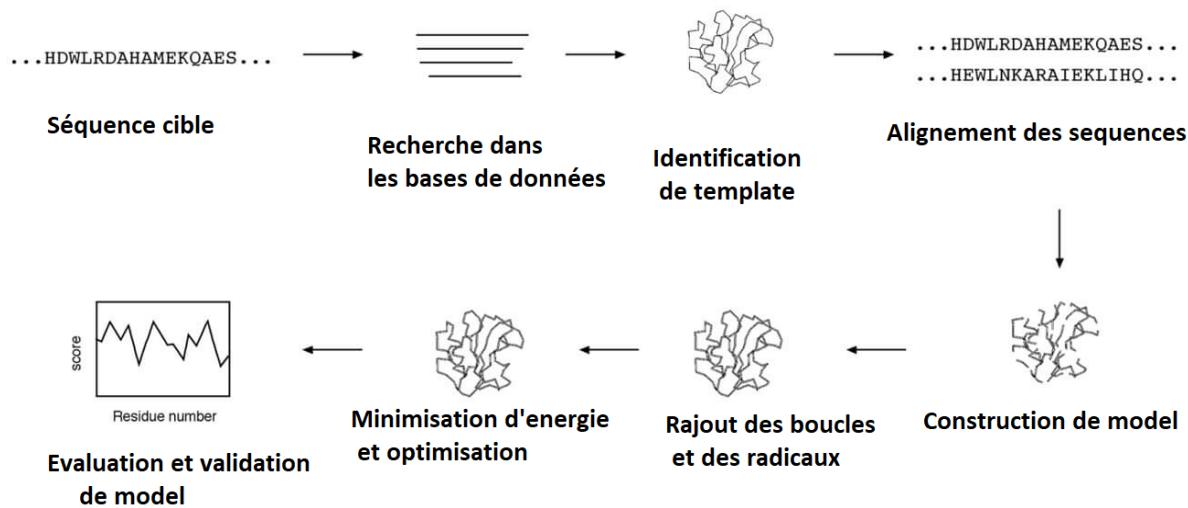


FIGURE 9. LES ÉTAPES DE MODÉLISATION PAR HOMOLOGIE.

#### 1.6.2 MÉTHODE AB INITIO

Ces approches qui tentent de prédire la structure d'une protéine uniquement à partir de sa séquence d'acides aminés, sans utiliser d'informations sur des structures connues. Le principe de ces méthodes repose sur la minimisation de l'énergie libre et l'exploration de l'espace conformationnel à l'aide de techniques de recherche, comme le recuit simulé ou les algorithmes de Monte Carlo. Deux exemples de logiciels qui utilisent le principe ab initio sont ROSETTA (Rohl et al., 2004) et QUARK (Xu & Zhang, 2012). Bien que ROSETTA soit principalement connu pour son approche basée sur les fragments, il comprend également un module ab initio pour la prédiction de la structure des protéines.

#### 1.6.3 MÉTHODES BASÉES SUR LES FRAGMENTS

Ces méthodes, basées sur les fragments, construisent la structure 3D de la protéine cible en assemblant des fragments de structures protéiques connues, en tenant compte des contraintes locales et globales de la séquence cible (Rohl et al., 2004). Un exemple de logiciel qui utilise cette approche est ROSETTA. Un autre logiciel basé sur la même approche est I-TASSER.

I-TASSER (Iterative Threading ASSEmbly Refinement) est un outil de prédiction de structure qui combine le threading de séquence, l'assemblage basé sur les fragments et la minimisation d'énergie pour générer des modèles 3D de protéines (Zhang, 2008).

#### 1.6.4 MÉTHODES BASÉES SUR L'APPRENTISSAGE PROFOND

Ces méthodes utilisent des réseaux de neurones profonds pour prédire directement la structure 3D d'une protéine à partir de sa séquence. Elles ont récemment montré des résultats prometteurs, surpassant les autres approches dans de nombreux cas (Jumper et al., 2021). Un exemple de logiciel utilisant cette approche est Alpha Fold, développé par DeepMind.

AlphaFold a été utilisé pour prédire avec succès la structure de la protéine humaine GCN4 leucine zipper (GCN4-LZ), qui est une région coiled-coil. La prédiction d'AlphaFold a montré une forte corrélation avec la structure expérimentale connue de cette protéine (Jumper et al., 2021).

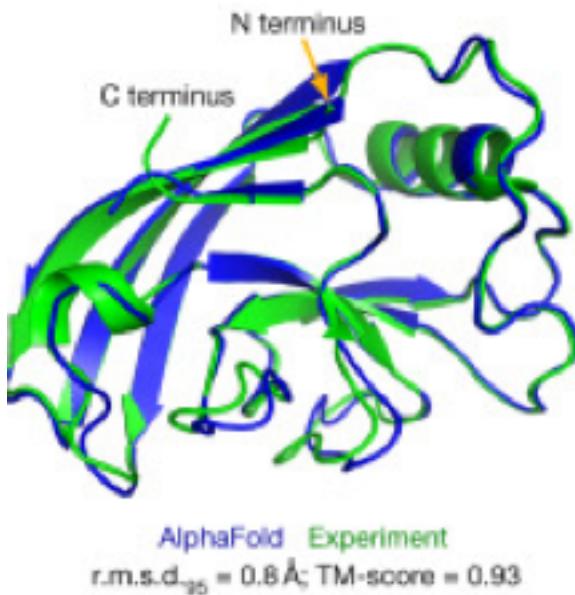


FIGURE 10. SUPERPOSITION DES DEUX STRUCTURES PROTÉIQUES VERTES (EXPÉRIMENTALE ET BLEU PRÉDITE PAR ALPHA FOLD . (JUMPER ET AL., 2021).

#### 1.6.5 VALIDATION DE LA STRUCTURE TERTIAIRE PRÉDITE PAR BIOINFORMATIQUE

La validation de la structure tertiaire d'une protéine est une étape cruciale pour vérifier si la structure prédictive est correcte et fiable. La structure tertiaire fait référence à la configuration tridimensionnelle globale adoptée par une protéine, résultant des interactions entre les acides aminés distants le long de la chaîne polypeptidique. Pour valider cette structure, plusieurs critères peuvent être évalués.

##### a. minimisation d'énergie

La minimisation d'énergie est une étape importante dans la modélisation moléculaire qui permet de prédire la structure la plus stable et la plus probable d'une molécule. Elle est utilisée pour calculer la conformation la plus stable de la molécule, en minimisant l'énergie potentielle de la molécule. Les interactions énergétiques entre les atomes de la molécule sont prises en compte dans le calcul de l'énergie potentielle. Les forces qui affectent chaque atome sont dérivées de cette fonction d'énergie potentielle, et la conformation de la molécule est ajustée en conséquence pour atteindre le minimum énergétique global. La minimisation d'énergie peut être réalisée à l'aide de différentes méthodes, telles que la méthode de gradient, qui calcule les dérivées partielles de la fonction d'énergie par rapport aux coordonnées des atomes, puis déplace la molécule dans la direction où la fonction d'énergie diminue le plus rapidement. Cette méthode est souvent utilisée pour optimiser les structures de protéines, de peptides et de petites molécules. La minimisation d'énergie est également

utile dans la préparation de systèmes pour d'autres méthodes de modélisation moléculaire, comme la dynamique moléculaire, qui simule les mouvements moléculaires dans le temps. En minimisant l'énergie potentielle avant l'utilisation de la dynamique moléculaire, on peut obtenir des structures de départ plus fiables, ce qui peut améliorer la qualité des simulations.

### b. ramachandran plot

Le graphique de Ramachandran est largement utilisé pour évaluer la qualité d'une structure protéique prédite ou expérimentale. Le graphique de Ramachandran, également appelé diagramme ou tracé de Ramachandran, est un outil graphique utilisé pour visualiser et évaluer les angles de torsion  $\phi$  (phi) et  $\psi$  (psi) du squelette polypeptidique des protéines. Il a été développé par G. N. Ramachandran, C. Ramakrishnan et V. Sasisekharan en 1963 (Ramachandran, Ramakrishnan, & Sasisekharan, 1963).

Le principe du graphique de Ramachandran repose sur la représentation des angles de torsion  $\phi$  et  $\psi$  pour chaque résidu d'acide aminé dans une protéine. Les angles  $\phi$  et  $\psi$  décrivent les rotations autour des liaisons peptidiques entre les acides aminés et sont essentiels pour déterminer la conformation tridimensionnelle d'une protéine. Le graphique de Ramachandran illustre les combinaisons d'angles  $\phi/\psi$  qui sont stéréo chimiquement possibles et énergétiquement favorables pour les acides aminés.

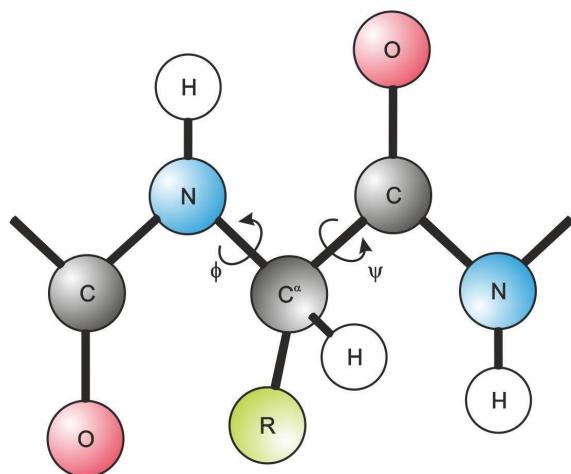


FIGURE 11. ANGLES DIÈDRES DÉCRIVANT L'ORIENTATION DE DEUX LIAISONS PEPTIDIQUES LIÉS AU MÊME ATOME CA

Les zones autorisées du graphique sont déterminées par les contraintes stéreo chimiques des liaisons et des angles, ainsi que par les interactions entre les atomes. Ces zones sont généralement classées en trois catégories : zones favorisées, zones autorisées et zones interdites. Les conformations favorisées correspondent aux angles de torsion les plus fréquemment observés dans les structures protéiques expérimentales. Les conformations autorisées sont moins fréquentes, mais encore possibles. Les conformations interdites correspondent aux angles de torsion qui entraînent des collisions stériques entre les atomes et sont donc très improbables.

Le graphique de Ramachandran est largement utilisé pour évaluer la qualité d'une structure protéique prédictive ou expérimentale. Une structure de haute qualité aura généralement la majorité de ses résidus dans les zones favorisées et autorisées, tandis que la présence de résidus dans les zones interdites peut indiquer des erreurs ou des problèmes locaux avec la structure.

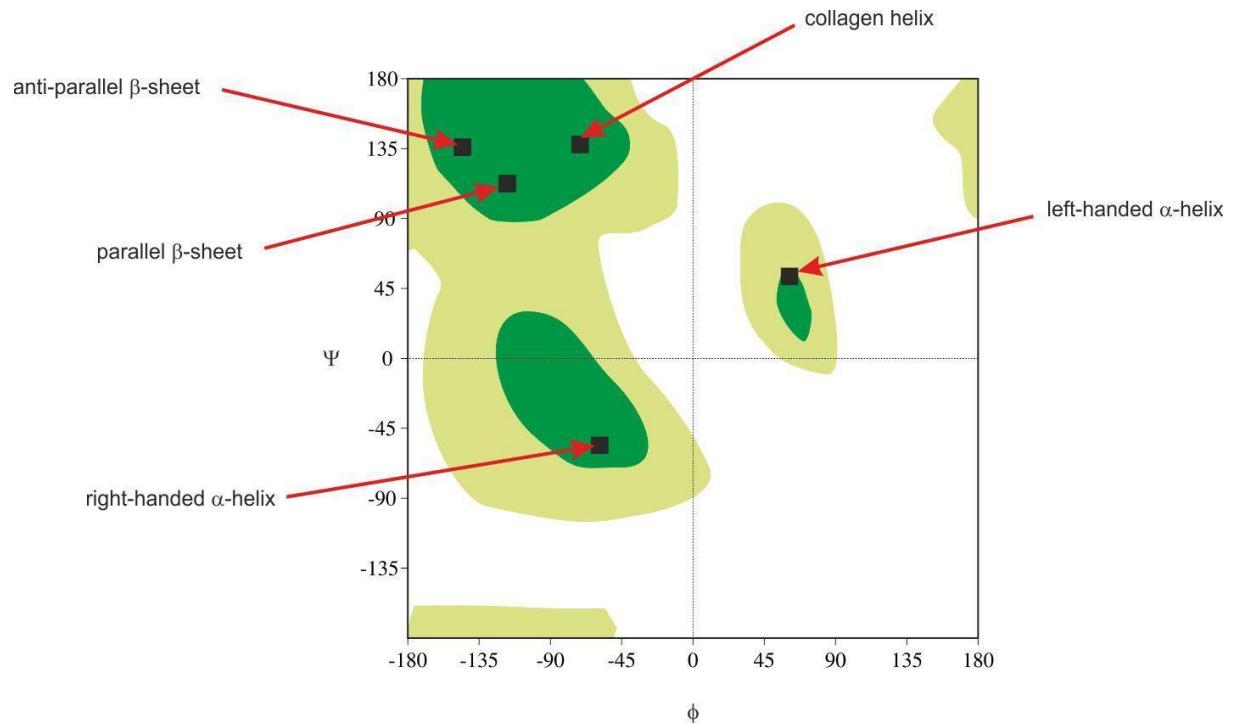


FIGURE 12. DIAGRAMME DE RAMACHANDRAN AVEC LES RÉGIONS LES PLUS FAVORISÉES (VERT FONCÉ) ET LES RÉGIONS SUPPLÉMENTAIRES (VERT CLAIR).

### C. PROCHECK

Le programme PROCHECK (Laskowski et al., 1993) est un outil informatique qui permet de valider les coordonnées de la structure protéique. Il calcule plusieurs paramètres stéréochimiques pour un modèle protéique donné, puis les compare avec les valeurs "idéales" obtenues à partir d'une base de données de structures protéiques à haute résolution bien raffinées dans la Protein Data Bank (PDB ; Bernstein et al., 1977). Les résultats de ces vérifications sont affichés sous forme de tracés colorés faciles à comprendre au format PostScript (Adobe Systems Inc., 1985). Cette méthode est utile pour le raffinement d'une structure protéique car elle permet de mettre en évidence les zones qui peuvent poser problème et nécessiter une attention particulière. En outre, PROCHECK ne nécessite que les coordonnées atomiques tridimensionnelles de la structure, ce qui lui permet de vérifier la "qualité" globale de tout modèle de structure protéique, qu'il soit dérivé expérimentalement par cristallographie ou RMN, ou construit par modélisation par homologie. Les résultats de PROCHECK peuvent aider à diriger la cristallographie vers les parties de la structure nécessitant une attention particulière, ce qui permet de gagner du temps et des efforts lors de l'analyse des données.

## 1.7 BASE DE DONNÉES DE RÉFÉRENCES DES PROTÉINES PDB

La base de données PDB (Protein Data Bank) est la principale base de données pour les structures de macromolécules biologiques résolues expérimentalement. Elle contient actuellement plus de 180 000 structures, dont plus de 170 000 sont des protéines et des peptides. Les autres structures comprennent des acides nucléiques, des complexes protéine-ADN, des complexes protéine-ARN et des petites molécules organiques. La PDB est régulièrement mise à jour avec de nouvelles structures qui sont déposées par des chercheurs du monde entier.

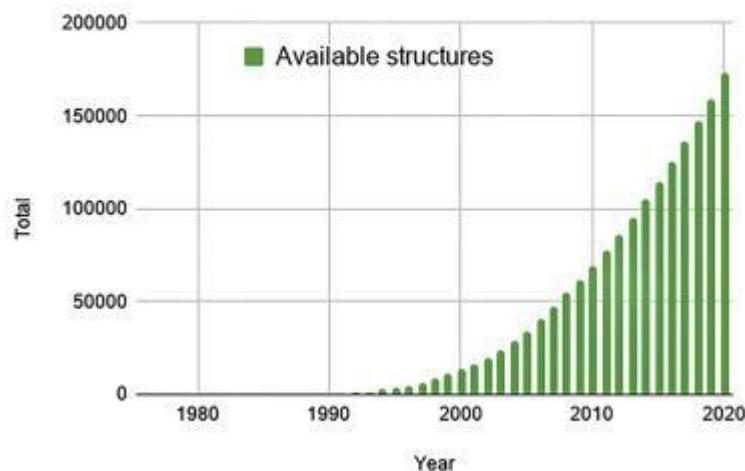


FIGURE 13. NOMBRE DES STRUCTURES 3D DÉTERMINÉS EXPÉRIMENTALEMENT ET DISPONIBLE DANS LA BASE DE DONNÉES PDB  
[HTTPS://WWW.RCSB.ORG](https://www.rcsb.org)

### 1.7.1 LE FICHIER PDB

Le fichier PDB (Protein Data Bank) contient des informations sur la structure tridimensionnelle d'une molécule, comme les coordonnées des atomes et les liaisons entre eux, ainsi que des informations sur les conditions expérimentales de résolution de la structure, telles que la technique de cristallisation et les références bibliographiques. Bien que d'autres formats aient été définis pour décrire les structures de macromolécules biologiques, les fichiers .pdb restent le moyen le plus courant pour communiquer des données sur les structures moléculaires de la PDB et sont également le format le plus couramment pris en charge par les logiciels de visualisation et de traitement de structures. L'un des avantages du format .pdb est sa facilité de lecture et d'interprétation pour l'utilisateur, même s'il n'est pas familier avec les données de structures moléculaires. Cependant, le format .pdb n'est pas assez descriptif pour permettre une exploitation facile et automatisée des données de structures, en particulier lorsqu'il s'agit d'exploiter les données expérimentales associées aux structures, ce qui peut rendre difficile l'automatisation du processus d'analyse, tout en maintenant un certain niveau de cohérence, de précision et de reproductibilité.

ATOM	80	CE	LYS	A	11	16.865	23.437	0.587	1.00	100.00	C
ATOM	81	NZ	LYS	A	11	16.300	24.743	1.031	1.00	50.42	N
ATOM	82	N	ALA	A	12	21.743	18.025	1.437	1.00	18.17	N
ATOM	83	CA	ALA	A	12	23.107	17.671	1.637	1.00	10.44	C
ATOM	84	C	ALA	A	12	23.555	16.291	1.157	1.00	15.53	C
ATOM	85	O	ALA	A	12	24.697	16.040	0.726	1.00	16.26	O

FIGURE 14. EXEMPLE D'UN FICHIER PDB, L'ATOME NUMÉRO 80 EST UN CARBONE (C) DANS LA LYSINE (LYS), RÉSIDU NUMÉRO 11 DE LA CHAÎNE A. SES COORDONNÉES X, Y, Z SONT RESPECTIVEMENT 16.865, 23.437 ET 0,587.

### 1.7.2 PDBsum

PDBSum est un outil en ligne qui fournit un résumé des caractéristiques importantes d'une structure de protéine dans la base de données PDB. Il offre une visualisation graphique des chaînes présentes dans la protéine, ainsi que des différents éléments de structures secondaires le long de chaque chaîne. En outre, il permet aux utilisateurs de faire des recherches à travers la base de données PDB, bien que cela soit limité aux informations textuelles des fichiers de structure tels que TITLE, HEADER, COMPND et AUTHOR. Cependant, dans Helix Explorer, un outil que nous détaillerons dans le prochain chapitre, nous avons pour objectif de permettre aux utilisateurs de faire des recherches directement dans la structure elle-même, plutôt que de se limiter aux métadonnées associées.

## 1.8 LES LOGICIELS DE VISUALISATIONS DE STRUCTURES 3D

Pour visualiser le fichier PDB de la structure d'une molécule, il est nécessaire d'utiliser un logiciel de visualisation moléculaire 3D. Ce logiciel convertit les coordonnées cartésiennes de la molécule en un modèle 3D qui peut être manipulé et observé. Les coordonnées 3D de la molécule sont inscrites dans des fichiers PDB qui fournissent un standard pour la représentation des données de structures macromoléculaires. Les logiciels de visualisation moléculaire permettent de mettre en évidence les particularités structurelles des molécules avec des jeux de couleurs ou de formes. Le schéma de couleur CPK est le plus couramment utilisé pour la représentation des molécules. Ces logiciels peuvent également être utilisés pour visualiser la trajectoire d'une simulation numérique, ce qui permet d'observer l'évolution structurale d'un complexe moléculaire au cours du temps ou de comprendre des phénomènes biologiques tels que le passage de molécules à travers une membrane ou le repliement d'une région protéique suite à son interaction avec un partenaire biologique.

### 1.8.1 PyMol

PyMol est un logiciel de visualisation moléculaire 3D très populaire dans les domaines de la biochimie, de la biologie structurale et de la chimie. Basé sur le langage de programmation Python. PyMol est capable de générer des images de haute qualité grâce à des techniques de lancer de rayons et est utilisé dans plus d'un quart des images de structures moléculaires présentes dans les articles scientifiques. Il dispose également de nombreuses fonctionnalités de rendu visuel pour une représentation optimale des molécules. PyMol est la base de plusieurs travaux d'équipe pour la

visualisation moléculaire dans différents environnements immersifs, ce qui montre son adaptabilité et sa capacité à satisfaire les besoins des chercheurs en matière de visualisation moléculaire. Le téléchargement de la version éducationnelle gratuite est disponible sur <https://pymol.org/ep>.

### 1.8.2 VMD

VMD est un outil de visualisation des structures 3D qui est spécialisé dans la visualisation et l'analyse des résultats de simulations de dynamique moléculaire. Cet outil est très apprécié en raison de ses nombreuses extensions et de sa capacité à être connecté à NAMD, un programme de dynamique moléculaire, permettant ainsi des simulations moléculaires interactives. Il est également bien adapté pour le rendu temps réel de grands modèles moléculaires grâce à des méthodes de rendu graphique directement effectuées et optimisées sur GPU.

## 2. LES VARIANTES ET LEURS IMPACT SUR LES PROTEINES

Les variants génétiques se réfèrent à des différences dans la séquence d'ADN entre individus ou populations. Ces variants peuvent être des modifications simples d'une seule base ou nucléotide (appelées polymorphismes mononucléotidiques ou SNPs), des insertions ou des délétions de plusieurs bases, ou des réarrangements plus complexes de la séquence d'ADN (Kidd et al., 2008)<sup>1</sup>. Ils peuvent se produire dans des régions codantes (exons) ou non codantes (introns et régions intergéniques) de l'ADN. La classification des variants génétiques est généralement basée sur leur impact potentiel sur la fonction des gènes et des protéines. Ils sont classés en variants bénins, possiblement pathogènes, et pathogènes. Les variants bénins sont considérés comme ayant peu ou pas d'impact sur la fonction de la protéine, tandis que les variants pathogènes sont associés à une maladie ou un phénotype abnormal (Richards et al., 2015).

### 2.1 LES VARIANTES NON CODANTS : DÉFINITION ET CLASSIFICATION

Les variants non codants se réfèrent aux modifications génétiques qui se produisent dans les régions non codantes de l'ADN, c'est-à-dire les régions de l'ADN qui ne sont pas traduites en protéines. Ces régions comprennent les introns (les parties non codantes d'un gène) et les régions intergéniques (les parties de l'ADN entre les gènes). Malgré qu'ils ne codent pas pour des protéines, ces variants peuvent avoir un impact significatif sur la régulation des gènes et peuvent contribuer à la maladie (Kellis et al., 2014). La classification des variants non codants est semblable à celle des variants génétiques en général, avec la distinction supplémentaire des variants régulateurs. Les variants régulateurs peuvent affecter l'expression d'un gène en modifiant la séquence d'ADN d'éléments régulateurs tels que les promoteurs et les enhancers (Maurano et al., 2012). La classification des variants non codants peut également prendre en compte l'impact potentiel sur l'épissage, où des modifications dans les séquences non codantes peuvent altérer la façon dont l'ARN est assemblé, produisant des protéines de formes différentes (Sterne-Weiler et al., 2011).

### **2.3 IMPACT DES VARIANTS CODANTS**

Les variants codants sont des modifications génétiques qui surviennent dans les exons, les parties de l'ADN qui sont transcrites et traduites en protéines. L'impact de ces variants dépend de leur nature. Les variants codants peuvent être synonymes, c'est-à-dire qu'ils ne modifient pas l'acide aminé produit, ou non synonymes, modifiant l'acide aminé produit et potentiellement la fonction de la protéine (Sauna et Kimchi-Sarfaty, 2011). Les variants non synonymes peuvent être des mutations missense, qui changent un acide aminé pour un autre, des mutations nonsense, qui introduisent un codon stop prématué, ou des mutations frameshift, qui décalent le cadre de lecture de la protéine (Mort et al., 2008). Ces dernières peuvent avoir des conséquences importantes sur la fonction de la protéine, menant potentiellement à la maladie. En plus de l'impact direct sur la séquence de la protéine, les variants codants peuvent également affecter la stabilité de la protéine, sa localisation cellulaire, et son interaction avec d'autres molécules (Yates and Sternberg, 2013). Les conséquences de ces modifications peuvent varier considérablement, allant d'un impact négligeable à une maladie grave.

## **3. PRÉDICTION DE L'IMPACT DES VARIANTS PAR LES OUTILS DE BIOINFORMATIQUE**

Il existe de nombreux outils bioinformatiques qui peuvent aider à prédire l'impact des variants génétiques sur la structure et la fonction des protéines. Parmi ceux-ci, PolyPhen-2 (Adzhubei et al., 2010) et SIFT (Kumar et al., 2009) utilisent des informations sur la séquence de la protéine et des informations structurelles pour prédire si un variant est probablement dommageable. D'autres outils, tels que DMAO (Kircher et al., 2014) et DANN (Quang et al., 2015), ont été développés pour prédire l'impact des variants non codants, en utilisant des données d'apprentissage automatique pour intégrer des informations sur une grande variété de caractéristiques génétiques et génomiques.

### **3.1 SIFT**

SIFT (Sorting Intolerant From Tolerant) est un outil de bioinformatique qui prédit l'impact d'un variant d'acide aminé sur la fonction d'une protéine basé sur le degré de conservation de l'acide aminé dans les séquences de protéines alignées (Kumar et al., 2009). SIFT utilise une approche statistique pour estimer si un variant particulier est toléré (c'est-à-dire sans effet) ou délétère (c'est-à-dire dommageable). Les variants qui sont rares ou absents dans les séquences alignées sont considérés comme délétères.

### **3.2 POLYPHEN**

PolyPhen-2 (Polymorphism Phenotyping v2) est un autre outil qui prédit l'impact fonctionnel des variants d'acides aminés. PolyPhen-2 utilise une combinaison de caractéristiques séquentielles, phylogénétiques et structurelles pour prédire si un variant est probablement bénin, probablement dommageable, ou probablement dommageable (Adzhubei et al., 2010). PolyPhen-2 prend en compte

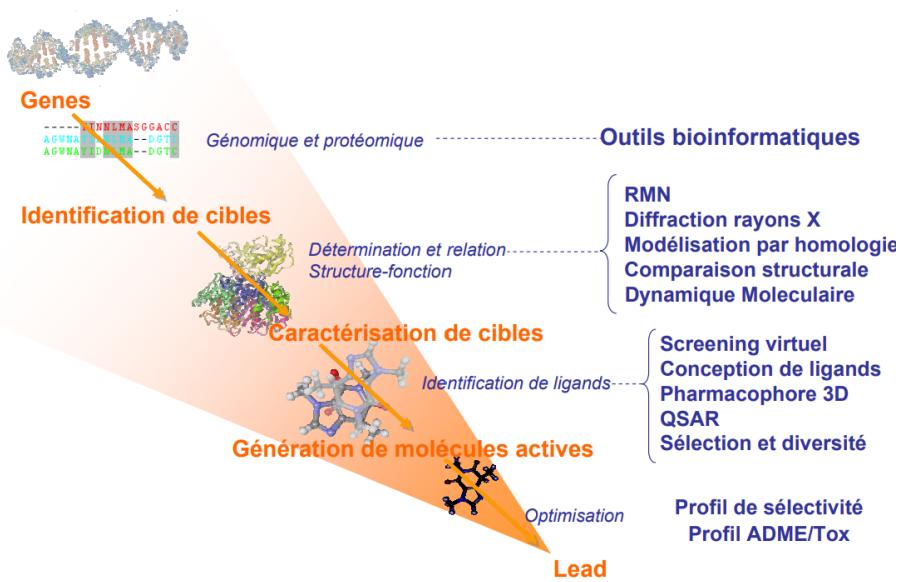
la conservation des acides aminés, la position du variant dans la protéine, et les caractéristiques structurales de la protéine pour faire sa prédition.

### ***3.3 ANALYSE DE CONSERVATION POUR L'ÉVALUATION DE L'IMPACT DES VARIANTS SUR LES PROTEINES***

L'analyse de la conservation est une méthode couramment utilisée pour évaluer l'impact des variants génétiques. Les sites de séquence hautement conservés au cours de l'évolution sont souvent fonctionnellement importants, et les modifications à ces sites sont plus susceptibles d'avoir un impact sur la fonction de la protéine (Siepel et al., 2005). De nombreux outils, y compris SIFT et PolyPhen-2, utilisent des informations sur la conservation des séquences pour aider à prédire l'impact des variants. D'autres outils, tels que PhyloP et GERP, fournissent des scores de conservation qui peuvent être utilisés pour évaluer l'impact potentiel des variants (Cooper et al., 2005; Pollard et al., 2010).

## **4. Découverte de médicament assistée par ordinateur (DMAO)**

La découverte de nouveaux médicaments est un processus complexe et coûteux qui peut prendre des années, voire des décennies. Les étapes impliquées comprennent la conception de composés chimiques, la synthèse, la caractérisation, la sélection, l'optimisation et finalement les essais cliniques. Tout au long de ce processus, il y a de nombreuses contraintes à prendre en compte, notamment le coût et la durée. La découverte de médicaments assistée par ordinateur est un domaine qui vise à réduire ces contraintes en utilisant des outils informatiques pour aider à la conception, la sélection et l'optimisation de composés chimiques potentiellement actifs. Les approches informatiques, telles que la modélisation moléculaire, la simulation de dynamique moléculaire, le criblage virtuel et la conception de médicaments rationnelle, peuvent aider à accélérer le processus de découverte de médicaments en permettant une exploration plus rapide et plus efficace de l'espace des composés. Cela peut entraîner des économies de temps et d'argent, tout en augmentant les chances de découvrir des médicaments efficaces. La DMAO a suscité une grande attention dès les années 1980 grâce aux avancées en biologie structurale (Ahlstrom et al., 1998). Les méthodologies de DMAO ont permis d'accélérer et d'optimiser le processus de recherche et développement de médicaments, répondant aux taux de réussite et aux coûts élevés de ce processus (Hajduk et al., 2008). Les progrès en cristallographie aux rayons X, spectroscopie RMN et puissance de calcul ont contribué à l'implémentation de la DMAO dans l'industrie pharmaceutique (Jorgensen, 2009). La DMAO a joué un rôle clé dans le développement de plusieurs médicaments approuvés, notamment pour le traitement de l'hypertension et du VIH (Hann et Keseru, 2012). La boîte à outils de la chimie computationnelle, de la simulation moléculaire et des approches chimio-informatiques offre aujourd'hui une variété d'outils informatiques pour aider à la conception de médicaments (Walters et Namchuk, 2013). Il est important d'utiliser ces outils avec une compréhension approfondie afin de produire des résultats valables dans un temps raisonnable (McGovern et Shoichet, 2003).



**FIGURE 15. STRATÉGIE GLOBALE DE CONCEPTION DE MÉDICAMENTS ASSISTÉE PAR ORDINATEUR, PARTANT DES DONNÉES GÉNOMIQUE ET PROTÉIQUE JUSQU'À L' IDENTIFICATION DE LEAD.**

Les deux principales catégories de la DMAO (découverte de médicaments assistée par ordinateur), le SBDD (structure-based drug design) et le LBDD (ligand-based drug design), ont considérablement contribué à la recherche et au développement de médicaments plus efficaces. Le SBDD utilise des informations sur la structure tridimensionnelle d'une cible thérapeutique, telle qu'une protéine ou un enzyme, pour concevoir des molécules qui interagissent spécifiquement avec elle. Cette approche permet de concevoir des molécules très spécifiques pour la cible et de réduire ainsi les effets secondaires sur d'autres parties du corps. De plus, elle est particulièrement utile pour les cibles thérapeutiques pour lesquelles il n'existe pas encore de ligand connu. Quant au LBDD, il utilise des informations sur les molécules connues qui se lient à une cible thérapeutique spécifique pour prédire les interactions moléculaires entre la cible et de nouvelles molécules. Cette approche permet de concevoir des molécules avec une structure similaire à celles qui se lient déjà à la cible et peut être utilisée pour concevoir des molécules plus efficaces et avec moins d'effets secondaires.

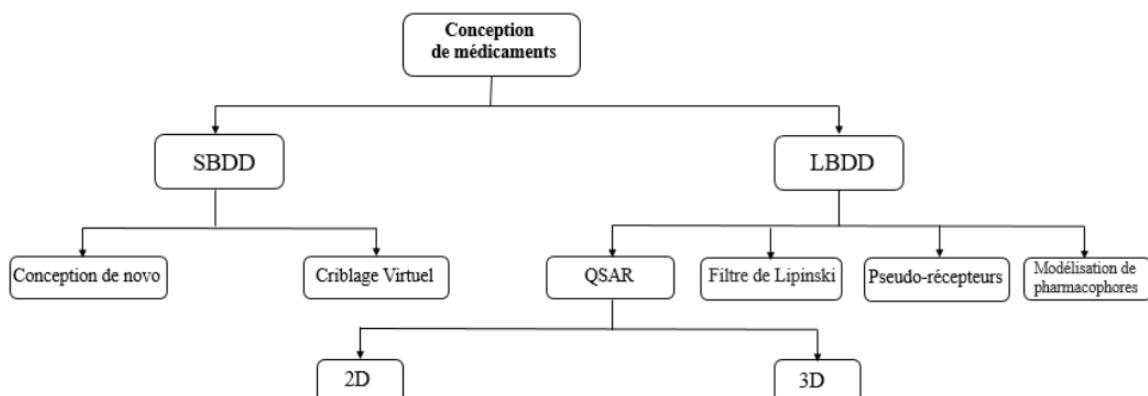


FIGURE 16. SCHÉMA DE LA CLASSIFICATION DES MÉTHODES DE LA DMAO.

#### 4.1 CONCEPTION DE MÉDICAMENTS À BASE DE LIGANDS

La LBDD (Ligand-Based Drug Design) est largement utilisée dans la DMAO en l'absence d'informations tridimensionnelles sur le récepteur et lorsque le site de liaison et la cible biologique ne sont pas précisément identifiés. Cette approche repose sur des molécules qui se lient à la cible biologique d'intérêt pour créer un modèle pharmacophore basé sur le ligand, qui définit les caractéristiques structurelles minimales qu'une molécule doit posséder pour se lier à la cible. Un modèle de la cible peut être développé sur la base des composés connus associés à la cible biologique, et utilisé pour trouver de nouvelles entités moléculaires présentant les mêmes caractéristiques et pouvant se fixer à la cible. La LBDD utilise la modélisation des relations quantitatives structure-activité (QSAR) pour établir une corrélation entre les propriétés des molécules calculées théoriquement et leur activité biologique obtenue expérimentalement. Cette corrélation obtenue permet de prédire l'activité de nouveaux analogues de ligands, contribuant ainsi à la conception de médicaments plus efficaces.

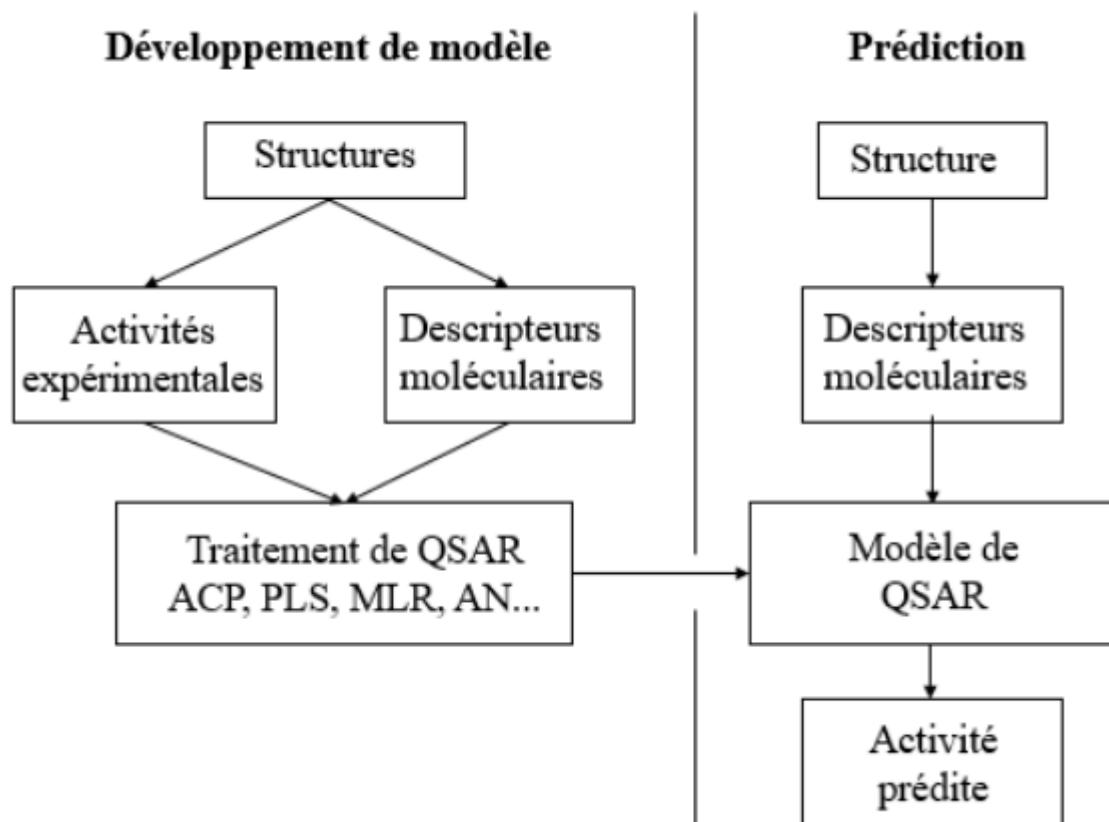


FIGURE 17. LES ÉTAPES DE LA MODÉLISATION DES RELATIONS QUANTITATIVES STRUCTURE-ACTIVITÉ QSAR.

## **4.2 CONCEPTION DE MÉDICAMENTS À BASE DE STRUCTURES**

La SBDD (Structure-Based Drug Design) est une méthode innovante de conception de médicaments qui repose sur les données structurelles tridimensionnelles de la cible médicamenteuse obtenues par des techniques de pointe telles que la cristallographie aux rayons X et la résonance magnétique nucléaire. Ces techniques fournissent des images hautement détaillées de la cible biologique qui permettent d'identifier les caractéristiques clés des interactions protéine-médicament à haute résolution.

Les méthodes de docking récepteur-ligand et de scoring sont souvent utilisées dans la SBDD pour positionner chaque ligand dans le site actif de liaison d'un récepteur et prédire le mode de liaison le plus favorable. Le docking est une technique de modélisation moléculaire qui permet de prédire l'orientation des ligands dans le site actif de la cible biologique. En utilisant les algorithmes de scoring, les chercheurs peuvent évaluer l'activité attendue de chaque composé en fonction de l'affinité de liaison approximative.

### **4.3 DOCKING MOLECULAIRE**

Le docking moléculaire est une méthode computationnelle clé en conception de médicaments assistée par ordinateur basée sur la structure de la cible. L'objectif de cette méthode est de prédire l'orientation d'une molécule (ligand) lorsqu'elle est liée à une autre (protéine, ADN, etc.), c'est-à-dire la "cible". Cette approche peut être utilisée pour la découverte de médicaments pour prédire la position et l'orientation d'une molécule de médicament dans le site de liaison d'une molécule de protéine cible.

Les principales étapes de cette approche sont :

#### **4.3.1 Principe d'algorithme de docking**

Ce processus implique la prédiction de la conformation du ligand, de l'orientation et de la position dans le site de liaison de la molécule cible. La molécule de ligand peut avoir divers états de rotation et de translation, ce qui entraîne un espace de recherche énorme. L'objectif est d'explorer les conformations et orientations possibles du ligand dans la poche de liaison de la protéine. De nombreux algorithmes sont utilisés à cette fin, tels que les méthodes de Monte Carlo, les algorithmes génétiques, et le recuit simulé, entre autres. Il existe différents types de méthodes de docking moléculaire, parmi lesquelles :

**a. Docking rigide (Rigid docking)** : Cette méthode suppose que les ligands et les récepteurs sont rigides, c'est-à-dire que leur conformation ne change pas pendant l'interaction. Cette méthode est plus rapide mais moins précise car dans la réalité, les molécules sont flexibles et changent de conformation pendant l'interaction.

**b. Docking flexible (Flexible docking) :** Cette méthode prend en compte la flexibilité des molécules. C'est une simulation plus réaliste, mais elle nécessite plus de puissance de calcul et est donc plus lente.

**c. Docking semi-flexible :** C'est une approche intermédiaire entre le docking rigide et le docking flexible. Dans le docking semi-flexible, la protéine cible est considérée comme rigide tandis que le ligand est considéré comme flexible. Cette approche tente de trouver un équilibre entre la précision du docking flexible et l'efficacité du docking rigide. Dans cette méthode, la protéine est généralement représentée par une grille d'énergies, calculée à partir de la structure 3D de la protéine. Les conformations possibles du ligand sont générées en faisant varier les angles de torsion des liaisons chimiques qui sont libres de tourner. Pour chaque conformation du ligand, une fonction d'énergie est calculée, qui prend en compte les interactions entre le ligand et la protéine, ainsi que l'énergie interne du ligand. La conformation qui minimise cette fonction d'énergie est considérée comme la meilleure prédition de la conformation liée.

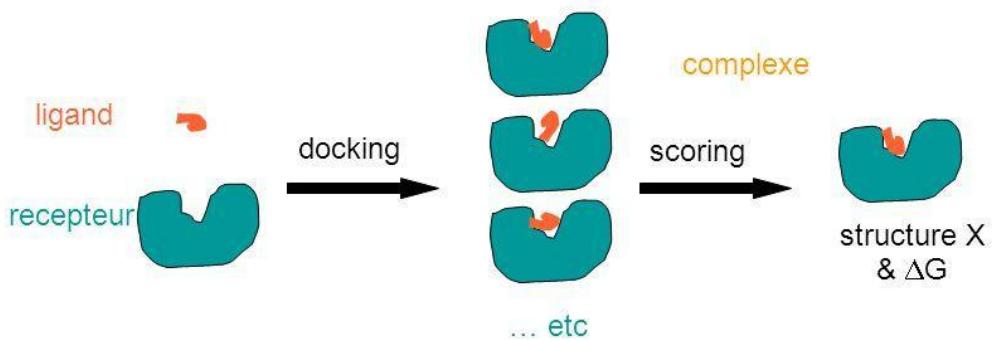


FIGURE 18. PRÉSENTATION SCHÉMATIQUE DE PRINCIPE DE DOCKING SEMI FLEXIBLE OU LA PROTÉINE EST CONSIDÉRÉE COMME RIGIDE ALORS QUE LE LIGAND EST FLEXIBLE.

#### 4.3.2 Énergie de liaison et fonctions de Notation (Scoring Functions)

L'énergie de liaison est calculée à partir de l'équation de force de van der Waals et des interactions électrostatiques entre le ligand et la cible biologique. Cette énergie est ensuite évaluée à l'aide de la fonction de scoring pour fournir une estimation de la force de l'interaction entre le ligand et la cible. L'équation de force de van der Waals se base sur les interactions entre les charges partielles positives et négatives des atomes du ligand et de la cible. Cette force est calculée en fonction de la distance entre les atomes, de leurs rayons de van der Waals et des termes de dispersion et de répulsion. Les interactions électrostatiques sont calculées à partir des charges partielles des atomes du ligand et de la cible, ainsi que de leur position relative.

L'énergie de liaison dans AutoDock est calculée à partir de l'équation de force de van der Waals et des interactions électrostatiques entre le ligand et la cible biologique, qui est exprimée comme suit :

$$E = E_{vdw} + E_{elec} + E_{desolv} + E_{torsion}$$

Où  $E_{vdw}$  est l'énergie de van der Waals,  $E_{elec}$  est l'énergie électrostatique,  $E_{desolv}$  est l'énergie de désolvatation et  $E_{torsion}$  est l'énergie de torsion. Chaque terme de cette équation est calculé en fonction de la distance entre les atomes, de leurs rayons de van der Waals, de leurs charges partielles et de leur position relative.

#### 4.3.3 VALIDATION DE DOCKING

La validation de docking est essentielle pour déterminer la fiabilité des prédictions et pour améliorer les algorithmes de docking et de scoring. Plusieurs techniques sont utilisées pour valider les résultats de docking, notamment :

##### a. Comparaison avec des structures expérimentales

Les structures tridimensionnelles de complexes ligand-récepteur obtenues par cristallographie aux rayons X, diffraction des neutrons ou résonance magnétique nucléaire (RMN) servent de référence pour évaluer la précision des prédictions de docking (Warren et al., 2006). Les prédictions sont comparées aux structures expérimentales en termes de position, d'orientation et de conformation du ligand.

##### b. Mesure de la performance des algorithmes de docking

Pour valider un algorithme de docking, on évalue généralement sa capacité à prédire correctement la pose du ligand dans le récepteur. Des mesures telles que le root-mean-square deviation (RMSD) entre les atomes du ligand prédit et ceux du ligand expérimental sont utilisées pour évaluer la précision du docking. Un RMSD faible indique une meilleure concordance entre la prédition et la référence.

##### c. Cross-docking et redocking

Le cross-docking consiste à prédire la pose d'un ligand dans plusieurs structures de récepteurs, tandis que le redocking consiste à prédire la pose d'un ligand dans la même structure de récepteur à partir de laquelle il a été extrait. Ces deux approches permettent d'évaluer la robustesse et la reproductibilité des prédictions de docking.

##### d. Composés interférentiels pan-essai

Les composés interférentiels pan-essai (PAINS) sont des substances chimiques qui sont souvent associées à des résultats faussement positifs dans les cibles à haut débit. Ces composés ont tendance à interagir de manière non spécifique avec de nombreuses cibles biologiques plutôt que d'affecter spécifiquement une cible souhaitée. De nombreux PAINS partagent un certain nombre de groupes fonctionnels perturbateurs. Bien que divers filtres aient été proposés et utilisés dans le cadre du dépistage virtuel et de la conception assistée par ordinateur de médicaments, la précision de ces filtres en termes de signalement ou de non-signalement des composés a été critiquée. Parmi les

PAINS courants, on trouve la toxoflavine, les isothiazolones, les hydroxyphényl hydrazones, la curcumine, les phénols-sulfonamides, les rhodanines, les étones, les quinones et les catéchols.

Un outil automatisé permettant de filtrer les composés pour déterminer les PAINS est disponible sur le lien suivant :

<https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=FAF-Drugs4#forms::FAF-Drugs4>

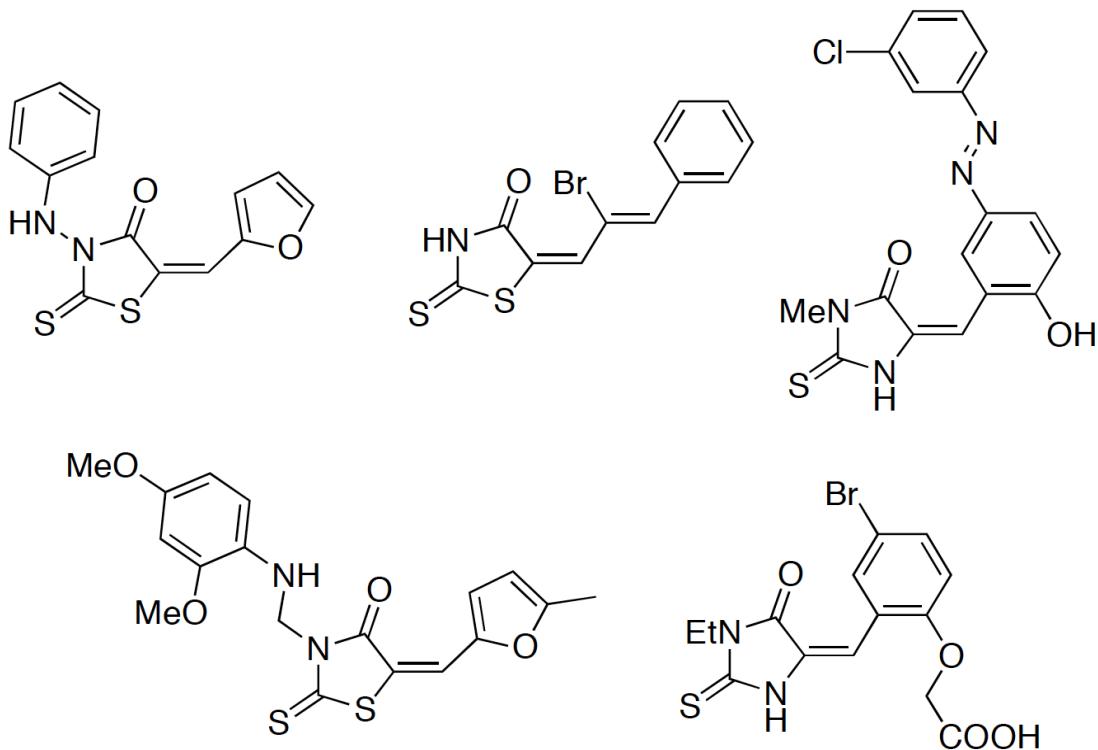


FIGURE 19. REPRÉSENTATION 2D DES COMPOSÉS COMPROMETTENT LES STRUCTURES QUI SONT LARGEMENT SIGNALÉES COMME ÉTANT DE FAUX POSITIFS CONSIDÉRÉS COMME PAINS, LES QUINONES À GAUCHE OU LES RHODANINES À DROITE .(JASIAL, HU, & BAJORATH, 2017).

#### 4.3.4 ANALYSE DE RÉSULTATS DE DOCKING

##### a. Analyse de RMSD et d'affinité de liaison

L'analyse du Root Mean Square Deviation (RMSD) constitue une étape essentielle dans l'évaluation des résultats de docking. Le RMSD est un indicateur statistique qui permet d'évaluer la différence entre deux superpositions moléculaires, généralement entre une structure de référence et une structure prédictive (Heifetz, 2012). Un RMSD faible signifie que la conformité prédictive est proche de la structure de référence, indiquant un docking réussi. À l'inverse, un RMSD élevé indique une grande différence entre les deux structures, ce qui peut être interprété comme un échec du docking.

L'analyse d'affinité, en revanche, est utilisée pour évaluer la force de l'interaction entre la molécule cible et le ligand. Les scores d'affinité sont calculés à l'aide de diverses méthodes, qui incluent souvent des termes énergétiques pour les forces de van der Waals, les forces électrostatiques, et la solvatation (Meng et al., 2011). En général, plus l'énergie de liaison est négative, plus l'affinité est

forte car l'énergie de liaison entre une molécule cible et un ligand est souvent exprimée en termes d'énergie libre de Gibbs. L'énergie libre de Gibbs est une mesure du potentiel maximal de travail que peut réaliser une réaction à pression et température constantes. Dans le contexte de l'interaction entre une molécule cible et un ligand, une énergie libre de liaison plus négative signifie que l'interaction est plus stable et que le ligand a une plus grande affinité pour la molécule cible. En effet, une énergie de liaison négative indique qu'il y a une libération d'énergie lors de la formation du complexe cible-ligand, ce qui est favorable sur le plan énergétique (Meng et al., 2011). C'est cette propriété qui fait que l'analyse des scores d'énergie de liaison est cruciale pour le docking moléculaire, car elle permet de prédire quelles molécules sont les plus susceptibles de se lier efficacement à la cible. Cela dit, il faut souligner que même si l'énergie de liaison est un indicateur important, d'autres facteurs, tels que la spécificité de liaison et la cinétique de liaison, sont également importants dans la prédiction des interactions ligand-cible réussies.

### b. Analyse d'interaction protéine-ligand

Les interactions ligand-protéine sont d'une importance capitale dans de nombreux aspects de la biologie et de la chimie. Ces interactions se produisent lorsque des ligands, tels que des petites molécules, des médicaments ou des hormones, se lient spécifiquement à des protéines cibles dans l'organisme. Différents types d'interactions ligand-protéine peuvent se produire, notamment les liaisons hydrogène, les interactions électrostatiques, les interactions hydrophobes et les interactions de Van der Waals. Ces interactions permettent la formation de complexes stables entre les ligands et les protéines, ce qui peut induire des changements conformationnels dans la protéine et moduler son activité biologique. L'importance de ces interactions réside dans leur capacité à réguler les processus biologiques, tels que la signalisation cellulaire, le transport des molécules, la reconnaissance antigène-anticorps et l'inhibition enzymatique. On distingue plusieurs types d'interaction, principalement :

- **Liaison hydrogen**

Une liaison hydrogène est une interaction spécifique entre un atome d'hydrogène et un autre atome électronégatif tel que l'oxygène, l'azote ou le fluor. Cette interaction est de nature dipôle-dipôle, ce qui signifie qu'elle implique la polarité des atomes impliqués. Dans une liaison hydrogène, l'atome d'hydrogène est partiellement positif en raison de la faible électronégativité de l'hydrogène, tandis que l'atome électronégatif (par exemple, l'oxygène) est partiellement négatif en raison de sa forte électronégativité. Cette différence de charge crée une attraction électrostatique entre les deux atomes, ce qui donne lieu à une liaison hydrogène.

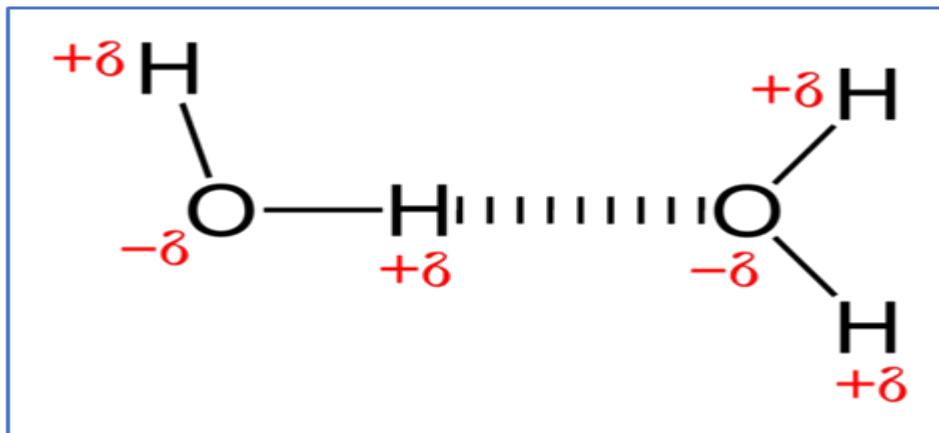


FIGURE 20. LIAISON HYDROGÈNES

En ce qui concerne la reconnaissance protéine/ligand, les liaisons hydrogène jouent un rôle clé dans la formation des complexes entre les protéines et leurs ligands. Les liaisons hydrogène entre les résidus d'acides aminés de la protéine et les atomes du ligand contribuent à la spécificité et à l'affinité de l'interaction protéine-ligand.

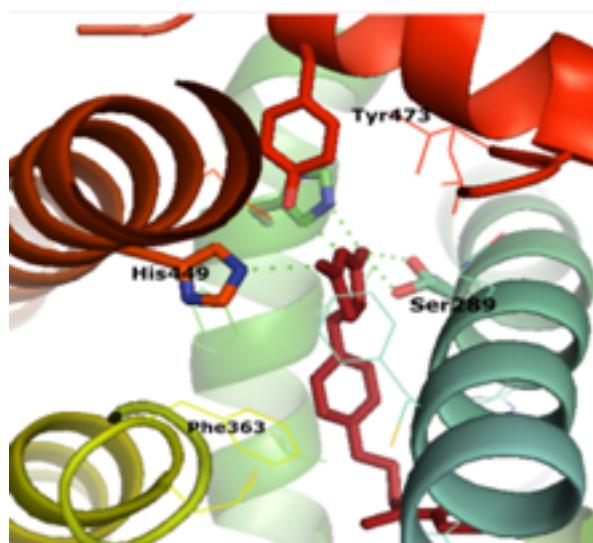


FIGURE 21. LIAISON HYDROGÈNES

- *Interaction de van Der Waals*

Les interactions de Van der Waals entre les protéines et les ligands jouent un rôle crucial dans la reconnaissance et la liaison spécifique des ligands aux sites actifs des protéines. Ces interactions se produisent lorsque les régions apolaires des ligands, telles que les groupes alkyl et les cycles aromatiques, interagissent avec les régions apolaires des résidus d'acides aminés dans les poches de liaison des protéines. Les forces de Van der Waals entre ces régions non polaires sont principalement des forces d'attraction induites par les fluctuations des charges électriques. Lorsque le ligand se rapproche de la protéine, les régions apolaires du ligand et de la protéine se rapprochent également, permettant aux nuages électroniques de ces régions de se chevaucher. Les fluctuations des charges

électriques dans les deux molécules créent alors des régions momentanément positives et négatives, ce qui génère une interaction attractive de Van der Waals. Cette interaction est renforcée par la complémentarité des formes et des tailles des régions apolaires du ligand et de la protéine, permettant un ajustement optimal et une interaction étroite.

Les forces de Van der Waals contribuent à la stabilité des complexes protéine-ligand en fournissant une interaction supplémentaire entre les deux molécules. Ces interactions sont essentielles pour la reconnaissance sélective des ligands, car elles contribuent à la formation de liaisons plus fortes entre le ligand et le site actif de la protéine. De plus, les forces de Van der Waals peuvent également influencer la conformation du ligand et favoriser la formation de conformations favorables pour l'interaction avec la protéine.

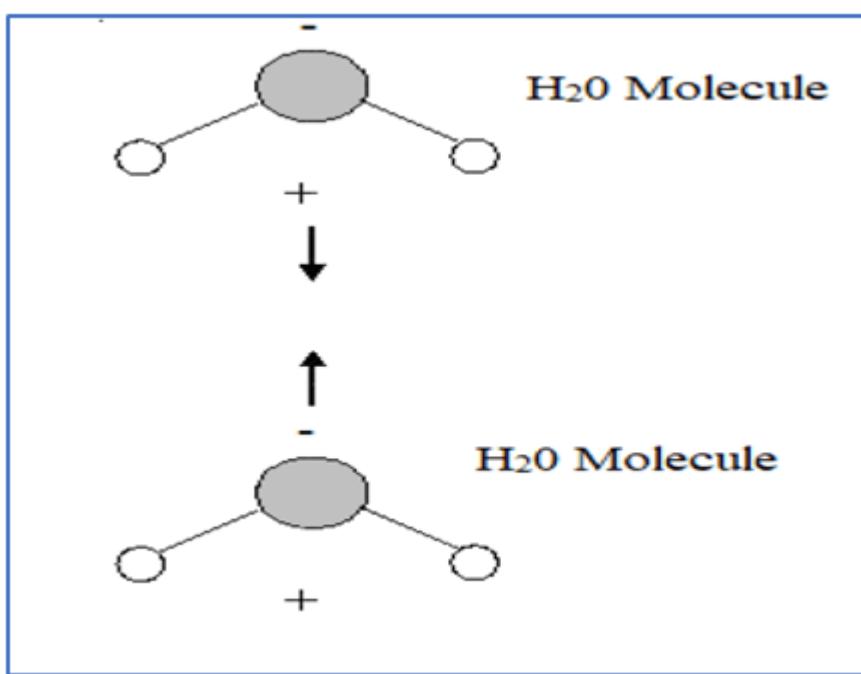


FIGURE 22. FORCES DE VAN DER WAALS

- *Interaction hydrophobic*

Les interactions hydrophobes entre les ligands et les protéines sont également d'une importance fondamentale dans la reconnaissance et la liaison spécifique des ligands aux sites actifs des protéines. Les régions hydrophobes des ligands et des poches de liaison des protéines ont tendance à se regrouper afin de minimiser leur contact avec l'eau environnante. Cette association hydrophobe entre le ligand et la protéine peut contribuer de manière significative à la stabilité et à la spécificité de l'interaction protéine-ligand. Lorsque le ligand pénètre dans la poche de liaison de la protéine, les résidus hydrophobes du ligand et de la protéine peuvent former des contacts étroits, permettant ainsi l'interaction hydrophobe. Ces interactions se produisent grâce à l'alignement des groupes hydrophobes des deux molécules, minimisant ainsi l'exposition à l'eau. L'association hydrophobe

entre le ligand et la protéine peut jouer un rôle clé dans la reconnaissance sélective des ligands, en contribuant à la formation de complexes stables et spécifiques. De plus, les interactions hydrophobes peuvent influencer la conformation du ligand et favoriser la formation de conformations adaptées à l'interaction avec la protéine. L'ajustement optimal des régions hydrophobes du ligand dans la poche de liaison peut contribuer à la stabilisation de la structure du complexe protéine-ligand.

Les interactions hydrophobes fournissent une contribution significative à la reconnaissance protéine-ligand, en particulier dans les régions apolaires où d'autres interactions sont moins prédominantes.

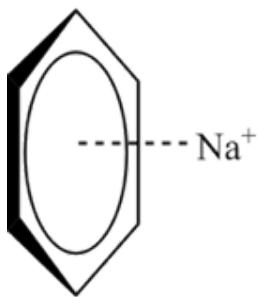
- *Interactions ioniques*

Les forces ioniques, aussi appelées interactions électrostatiques, sont des interactions qui se produisent entre des particules chargées de manière opposée. Elles jouent un rôle crucial dans la fonction et la stabilité des protéines. Les acides aminés qui composent les protéines peuvent avoir des charges positives ou négatives, et ces charges peuvent interagir avec les ions présents dans les ligands. Ces forces ioniques sont essentielles pour maintenir la structure tridimensionnelle de la protéine et permettent également une interaction spécifique et forte entre la protéine et le ligand. Elles contribuent à la spécificité de la liaison, en attirant des ligands de charge opposée vers le site actif de la protéine, et peuvent également jouer un rôle dans la catalyse des réactions chimiques par les enzymes. Il est cependant important de souligner que les forces ioniques ne sont qu'une parmi de nombreuses interactions potentielles entre une protéine et un ligand, et peuvent agir en concert avec d'autres types d'interactions, comme les liaisons hydrogène, les interactions hydrophobes et les forces de van der Waals.

- *Interaction Pi-cation*

Les interactions pi-cation sont des interactions non-covalentes qui peuvent jouer un rôle significatif dans la liaison ligand-protéine. Un cation est un atome ou une molécule qui porte une charge positive, tandis qu'un système pi est un système d'atomes où les électrons sont délocalisés, généralement trouvés dans les cycles aromatiques.

L'interaction pi-cation survient lorsque le cation interagit avec la densité électronique du système pi. Un exemple courant d'interaction pi-cation est celle qui se produit entre un cycle aromatique, comme le benzène ou le phényle, et un ion métallique ou un acide aminé chargé positivement, comme l'arginine ou la lysine.



Dans la liaison ligand-protéine, ces interactions peuvent être cruciales. Par exemple, un ligand qui contient un cycle aromatique pourrait interagir par interaction pi-cation avec un résidu d'acide aminé de la protéine qui est chargé positivement. Cette interaction peut aider à stabiliser la liaison du ligand à la protéine, augmentant ainsi l'efficacité et la spécificité de la liaison.

Un exemple notable est la liaison de nombreux médicaments aux protéines. Certains médicaments contiennent des cycles aromatiques qui peuvent interagir par interaction pi-cation avec des résidus d'acides aminés chargés positivement sur la protéine cible. Ces interactions peuvent augmenter l'efficacité du médicament en améliorant sa liaison à la protéine cible.

#### 4.4 La règle de Lipinski : prédiction de la ressemblance des ligands aux médicaments

La règle de Lipinski, également connue sous le nom de "règle des cinq", est une règle empirique utilisée en pharmacologie pour prédire la solubilité et la biodisponibilité d'un composé chimique (Lipinski, 1997). Cette règle stipule que pour qu'un composé soit efficace en tant que médicament, il doit posséder certaines propriétés physico-chimiques, notamment un poids moléculaire inférieur à 500 daltons, une logP (coefficient de partage octanol/eau) inférieure à 5, pas plus de 5 liaisons hydrogène donneur, et pas plus de 10 liaisons hydrogène accepteur. Cette règle est basée sur l'observation que les composés qui respectent ces critères ont tendance à être plus solubles et à avoir une biodisponibilité plus élevée, ce qui les rend plus efficaces en tant que médicaments. La règle de Lipinski est largement utilisée dans la conception de médicaments pour filtrer les composés potentiels en début de recherche et pour maximiser les chances de succès dans les essais cliniques.

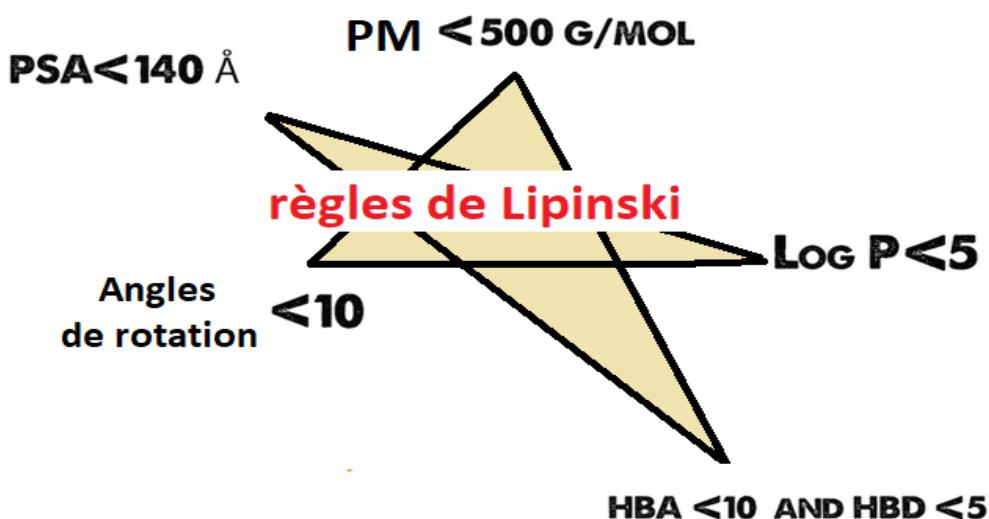


FIGURE 23. SCHÉMA DES CRITÈRES DE LA RÈGLE DE LIPINSKI : POIDS MOLÉCULAIRE (PM), SURFACE POLAIRE (PSA), ANGLES DE ROTATION , ACCEPTEURS DE LIAISONS HYDROGÈNE (HBA), DONNEURS DE LIAISONS HYDROGÈNE (HBD) ET LOG P.

#### 4.5 Simulation par dynamique moléculaire

La dynamique moléculaire (DM) est une technique de simulation largement utilisée pour étudier les macromolécules biologiques, telles que les protéines (Karplus & McCammon, 2002). Les simulations de DM permettent d'examiner le comportement dynamique des protéines à différentes échelles de temps, des mouvements internes rapides aux changements conformationnels lents et au repliement des protéines (Freddolino & Schulten, 2009). La DM est également utilisée pour rationaliser les propriétés expérimentales mesurées au niveau moléculaire et affiner les structures déterminées par des méthodes de rayons X ou de RMN (Lindahl & Hess, 2008).

L'interaction entre les techniques informatiques et expérimentales dans le domaine des simulations de DM est ancienne et mutuellement bénéfique (Leach, 2001). Les méthodes théoriques aident à comprendre et analyser les données expérimentales, tandis que les données expérimentales sont essentielles pour valider et améliorer les techniques et protocoles de calcul (Orozco & Luque, 2000).

Avec les progrès récents, il est désormais possible de réaliser des simulations de systèmes plus réalistes, incluant des molécules d'eau explicites, des contre-ions et un environnement complet de type membrane (Kutzner et al., 2011). Cela permet d'étudier de nouvelles propriétés en temps réel et d'approfondir notre compréhension des systèmes biomoléculaires complexes (Shaw et al., 2010).

Au cours des dernières années, des progrès significatifs ont été réalisés dans la représentation des systèmes de dynamique moléculaire (DM), notamment en ce qui concerne les meilleurs champs de force, le traitement des interactions électrostatiques à longue portée et les conditions limites du système, ainsi que les algorithmes améliorés pour contrôler la température et la pression (Piana et al., 2015; Darden et al., 1998). En théorie, les simulations DM dans un solvant explicite pourraient

modéliser les mouvements concertés importants en biologie, tels que les changements conformationnels, la formation de structures secondaires et tertiaires et le repliement des protéines (Karplus, M et al., 2002). Cependant, les simulations DM sont coûteuses en calcul et dépendent du développement de champs de force améliorés (Van Gunsteren & Berendsen, 1990). Les codes de simulation MD modernes optimisés pour les architectures GPU permettent d'exécuter rapidement des simulations de complexes récepteur-ligand en tenant compte de la flexibilité du récepteur et du ligand (Stone et al., 2010). Plusieurs progiciels sont disponibles pour la modélisation des biomolécules, tels que GROMACS (Abraham et al., 2015), AMBER (Case et al., 2005), NAMD (Phillips et al., 2005), YASARA (Krieger et al., 2002) et CHARMM (Brooks et al., 2009). Ces programmes de simulation avancés permettent différents types de simulations et fournissent des informations utiles sur les interactions protéine-ligand et les fluctuations structurelles des protéines (Henzler-Wildman & Kern, 2007). La DM complète bien les techniques basées sur les rayons X et la RMN pour étudier les complexes difficiles (Rieping et al., 2007). Toutefois, les données structurelles expérimentales sont nécessaires pour évaluer et améliorer la précision des modèles d'ensemble générés par DM (Lindorff-Larsen et al., 2012). La DM est utile pour étudier les interactions faibles entre un ligand et une protéine intrinsèquement désordonnée, ainsi que pour générer des ensembles conformationnels et estimer les énergies libres relatives de liaison (Krieger et al., 2006).

#### 4.5.1 LES ÉQUATIONS DE MOUVEMENT

Les équations de mouvement sont un ensemble d'équations différentielles qui décrivent la dynamique d'un système en fonction des forces qui lui sont appliquées. Dans le contexte des simulations de dynamique moléculaire (DM), les équations de mouvement sont utilisées pour décrire la trajectoire des atomes au cours du temps en fonction des forces interatomiques déterminées par le champ de force. Les équations de mouvement sont généralement basées sur la deuxième loi de Newton, qui stipule que la force agissant sur un atome est égale à la masse de cet atome multipliée par son accélération ( $F = ma$ ). Pour chaque atome dans le système, cette équation différentielle relie la position, la vitesse et l'accélération en fonction du temps, et les forces sont dérivées de l'énergie potentielle définie par le champ de force.

Pour résoudre ces équations de mouvement, un intégrateur numérique est utilisé pour simuler la trajectoire atomique au cours du temps avec une discrétisation temporelle, appelée pas de temps. L'algorithme de Verlet et ses variantes, tels que le Verlet à vitesse et le Verlet de Liapunov, sont couramment utilisés pour intégrer les équations de mouvement dans les simulations de DM, car ils offrent une bonne conservation de l'énergie et une stabilité numérique.

#### 4.5.2 INTERVALLE DE TEMPS

Dans les simulations de dynamique moléculaire (DM), les équations de mouvement sont basées sur la deuxième loi de Newton ( $F = ma$ ), qui relie la force ( $F$ ), la masse ( $m$ ) et l'accélération ( $a$ ) d'un atome.

Pour résoudre ces équations de mouvement, un pas de temps discret ( $\Delta t$ ) est utilisé pour discréteriser la trajectoire des atomes au cours du temps.

L'intervalle de temps, ou pas de temps, est un paramètre clé dans les simulations de DM, car il détermine la résolution temporelle et la stabilité numérique de la simulation. Un pas de temps trop grand peut entraîner des erreurs numériques et une mauvaise conservation de l'énergie, tandis qu'un pas de temps trop petit peut rendre la simulation très coûteuse en temps de calcul sans apporter d'amélioration significative de la précision.

Pour les simulations de biomolécules, un pas de temps typique se situe entre 1 et 5 femtosecondes (1 fs = 10 à la puissance -15 s). Ce choix d'intervalle de temps permet de capturer avec précision les mouvements rapides des atomes d'hydrogène, qui ont généralement les vibrations les plus rapides dans les biomolécules. Des techniques telles que l'intégration multiple de pas de temps peuvent être utilisées pour augmenter l'efficacité des simulations en utilisant des pas de temps plus longs pour les interactions à plus longue portée et plus lentes, tout en conservant des pas de temps plus courts pour les interactions rapides.

#### 4.5.3 CHAMP DE FORCE

Les champs de force jouent un rôle crucial dans les simulations de dynamique moléculaire (DM) et la modélisation des biomolécules. Ils représentent l'ensemble des paramètres et des fonctions mathématiques utilisés pour décrire les interactions entre les atomes d'un système moléculaire. Ces interactions comprennent les liaisons covalentes, les angles de valence, les termes diédriques, les interactions non covalentes de van der Waals et les forces électrostatiques.

Plusieurs champs de force ont été développés au fil des ans pour traiter différents types de systèmes moléculaires, notamment les protéines, les acides nucléiques, les lipides et les petites molécules. Parmi les champs de force les plus couramment utilisés pour les biomolécules, on trouve :

##### a. AMBER (ASSISTED MODEL BUILDING WITH ENERGY REFINEMENT)

Ce champ de force est largement utilisé pour les protéines, les acides nucléiques et les lipides. Il a été développé à l'origine par Peter Kollman et ses collègues (Cornell et al., 1995).

##### b. CHARMM (CHEMISTRY AT HARVARD MOLECULAR MECHANICS)

Un autre champ de force populaire pour les simulations de biomolécules, développé à l'origine par Martin Karplus et ses collègues (MacKerell et al., 1998).

##### c. GROMOS (GRONINGEN MOLECULAR SIMULATION)

Ce champ de force a été développé principalement pour les protéines et les lipides, avec une mise à jour régulière des paramètres (Oostenbrink et al., 2004).

#### *d. OPLS-AA (OPTIMIZED POTENTIALS FOR LIQUID SIMULATIONS-ALL ATOM)*

Un champ de force développé par William L. Jorgensen et ses collègues, destiné à la modélisation des protéines et d'autres biomolécules (Jorgensen et al., 1996).

#### 4.5.4 MINIMISATION D'ÉNERGIE

La minimisation d'énergie est une étape cruciale dans la simulation de dynamique moléculaire qui vise à obtenir une configuration d'énergie minimale pour le système étudié. Cette étape utilise des méthodes d'optimisation numérique et des modèles de potentiel d'interaction pour ajuster les positions atomiques et réduire l'énergie potentielle du système. une configuration d'énergie minimale pour le système étudié. Cette étape est essentielle pour éliminer les contraintes et les interactions indésirables initiales, afin de préparer le système pour la simulation proprement dite. L'objectif de la minimisation d'énergie est de trouver les positions des atomes qui minimisent l'énergie potentielle du système, en ajustant les coordonnées atomiques de manière itérative. Pour cela, différentes méthodes d'optimisation numérique sont utilisées, telles que la méthode du gradient, la méthode de conjugaison de gradients ou encore la méthode de Newton-Raphson. Lors de la minimisation d'énergie, les forces interatomiques sont calculées en utilisant des modèles de potentiel d'interaction, tels que les potentiels de Lennard-Jones ou les potentiels de champ de force. Ces potentiels décrivent les interactions entre les atomes, prenant en compte les effets de la liaison chimique, des interactions à longue portée et des interactions électrostatiques. Il existe également des algorithmes spécifiques pour la minimisation d'énergie, tels que l'algorithme de descente de gradient ou l'algorithme de gradient conjugué, qui permettent d'itérer sur les positions atomiques jusqu'à atteindre une convergence énergétique satisfaisante. La convergence est généralement considérée comme atteinte lorsque la variation de l'énergie potentielle devient négligeable ou lorsque les forces interatomiques deviennent inférieures à un certain seuil. La minimisation d'énergie peut être réalisée en utilisant des logiciels de simulation de dynamique moléculaire largement utilisés, tels que GROMACS (Abraham et al., 2015) ou AMBER (Salomon-Ferrer et al., 2013), qui fournissent des outils et des algorithmes avancés pour cette étape.

#### 4.5.5 CONTRÔLE DE LA TEMPÉRATURE ET DE LA PRESSION

Contrôle de la température et de la pression : Des thermostats et barostats sont utilisés pour maintenir la température et la pression constantes lors des simulations de dynamique moléculaire, afin d'assurer des conditions réalistes et de reproduire les conditions expérimentales.

Le thermostat de Langevin est un exemple de thermostat qui contrôle la température en ajoutant un terme de friction et une force aléatoire aux équations du mouvement (Grest & Kremer, 1986). Ce thermostat est particulièrement utile dans les simulations de systèmes biologiques, car il permet de modéliser l'interaction des biomolécules avec leur environnement de manière plus réaliste. Le barostat de Berendsen est un exemple de barostat qui ajuste la pression du système en modifiant les

dimensions de la boîte de simulation de manière isotrope ou anisotrope (Berendsen et al., 1984). Ce barostat est souvent utilisé pour maintenir la pression constante dans les simulations de dynamique moléculaire, en ajustant la taille du système pour compenser les variations de pression. D'autres méthodes de contrôle de la température et de la pression existent, telles que le thermostat de Nosé-Hoover (Nosé, 1984; Hoover, 1985) et le barostat de Parrinello-Rahman (Parrinello & Rahman, 1981), offrant des alternatives en fonction des besoins spécifiques des simulations.

#### 4.5.6 LA PRODUCTION DE LA DYNAMIQUE

Pendant cette étape, le système est soumis à une évolution temporelle réaliste en utilisant les équations de mouvement newtoniennes pour chaque atome. La simulation de production permet de générer une trajectoire moléculaire qui fournit des informations détaillées sur les propriétés thermodynamiques, cinétiques et structurales du système. Elle permet de comprendre les changements conformationnels, les réactions chimiques, les mouvements collectifs des molécules et bien plus encore. Pour réaliser la simulation de production, des méthodes numériques avancées et des algorithmes d'intégration temporelle sont utilisés. Différentes approximations et schémas d'intégration, tels que l'algorithme de Verlet ou la méthode de saut de pas (leapfrog), sont employés pour résoudre les équations de mouvement à chaque pas de temps. Il existe de nombreux logiciels de simulation de dynamique moléculaire qui offrent des fonctionnalités pour la simulation de production, tels que NAMD (Phillips et al., 2005), GROMACS (Abraham et al., 2015) et AMBER (Salomon-Ferrer et al., 2013). Ces logiciels sont couramment utilisés pour simuler des systèmes biologiques et chimiques complexes et ont démontré leur efficacité dans diverses études de recherche. La durée de la simulation de production dépend des objectifs de l'étude et des caractéristiques du système étudié. Les simulations peuvent varier de quelques picosecondes à plusieurs microsecondes, voire plus longtemps pour des systèmes plus grands et plus complexes.

#### 4.5.7 DONNÉES OBTENUS PAR LES SIMULATIONS DE DYNAMIQUE MOLÉCULAIRE

Les simulations de dynamique moléculaire (MD) sont devenues un outil crucial pour comprendre les dynamiques complexes et les propriétés structurales des systèmes biologiques et chimiques au niveau atomique. Ces simulations génèrent d'importantes quantités de données, mais leur interprétation peut souvent être un défi, nécessitant une solide compréhension à la fois de la méthodologie computationnelle et des bases théoriques des phénomènes observés. Les simulations de dynamique moléculaire fournissent des données sur la position et la vitesse de chaque atome dans le système tout au long de la simulation. L'un des principaux objectifs des simulations de MD est de comprendre le mouvement des atomes ou des molécules, à partir duquel on peut déduire les propriétés dynamiques du système telles que les coefficients de diffusion, les vitesses de réaction et les fréquences vibratoires (Leach, 2001). L'interprétation la plus directe des données consiste à examiner la trajectoire du système, qui comprend les positions de tous les atomes à chaque instant.

En visualisant ces trajectoires, nous pouvons comprendre les mouvements physiques et les interactions des molécules dans le système, tels que les changements conformationnels, les événements de liaison ou les corrélations dynamiques entre différentes régions d'une molécule (Karplus & Petsko, 1990). Un concept critique dans l'interprétation des simulations de MD est le paysage d'énergie potentielle, qui décrit la relation entre les configurations atomiques d'un système et leurs énergies potentielles correspondantes. En explorant ce paysage, les chercheurs peuvent identifier les états stables d'un système (c'est-à-dire les minima d'énergie) et les états de transition (c'est-à-dire les maxima d'énergie) qui les séparent, ce qui est essentiel pour comprendre la cinétique et la thermodynamique du système (Wales, 2003). De plus, les moyennes sur l'ensemble sont souvent calculées à partir des trajectoires de MD pour obtenir des propriétés thermodynamiques telles que l'énergie moyenne, la température et la pression du système. Ces quantités peuvent être liées aux propriétés macroscopiques via la mécanique statistique, établissant un lien entre la dynamique microscopique et le comportement macroscopique du système (Frenkel & Smit, 2002).

#### *A. ROOT MEAN SQUARE DEVIATION (RMSD)*

Le RMSD est une mesure de la distance moyenne entre les atomes (généralement les atomes du squelette) de protéines superposées. Dans le contexte des simulations de MD, le RMSD est souvent utilisé pour évaluer les différences structurelles et comparer les changements conformationnels d'une protéine par rapport à une structure de référence au fil du temps. Cela permet d'évaluer la stabilité globale et l'étendue de l'écart par rapport à la structure initiale pendant la période de simulation (Grant et al., 2006).

#### *B. ROOT MEAN SQUARE FLUCTUATION (RMSF)*

Alors que le RMSD mesure l'écart d'une structure protéique, le RMSF donne une indication de la variabilité ou de la flexibilité des résidus protéiques individuels au cours d'une simulation. Des valeurs élevées de RMSF indiquent généralement des régions très flexibles de la protéine, telles que des boucles ou les extrémités des hélices alpha et des feuillets bêta, tandis que des valeurs faibles de RMSF peuvent indiquer une région plus rigide ou structurée (Karplus & McCammon, 2002).

#### *C. RAYON DE GIRETION (R<sub>G</sub>)*

Le rayon de giration est une mesure de la distance quadratique moyenne pondérée par la masse d'un ensemble d'atomes par rapport à leur centre de masse commun. Dans le contexte de la MD, R<sub>G</sub> est un descripteur utile de la compacité d'une protéine ou d'une autre macromolécule. Les variations de R<sub>G</sub> au cours d'une simulation peuvent indiquer des changements conformationnels tels que le repliement ou le dépliement d'une protéine (Bhattacharyya & Bhat, 2020).

#### *D. SURFACE ACCESSIBLE AU SOLVANT (SASA)*

Le SASA mesure la surface d'une biomolécule accessible à un solvant. Dans les simulations de MD, la surveillance du SASA peut fournir des informations sur les changements d'exposition de la surface d'une protéine au solvant, ce qui peut être essentiel dans des processus tels que les interactions protéine-protéine, le repliement des protéines et la liaison des ligands. Une diminution du SASA au fil du temps correspond souvent à une molécule devenant plus compacte, tandis qu'une augmentation peut indiquer un dépliement ou une expansion (Shrake & Rupley, 1973).

## CHAPITRE 2. LE DIABÈTE DE TYPE 2

### 1. Introduction

Le diabète de type 2 est une maladie métabolique chronique caractérisée par une résistance à l'insuline et une insuffisance de la production d'insuline, entraînant une hyperglycémie (American Diabetes Association, 2021). Cette affection est la forme la plus courante de diabète, représentant environ 90-95 % de tous les cas de diabète (CDC, 2021). Les causes du diabète de type 2 sont multifactorielles, impliquant des facteurs génétiques et environnementaux, tels que l'obésité, le manque d'exercice et une alimentation déséquilibrée (NIDDK, 2016). Des protéines clés dans la régulation du métabolisme du glucose et des lipides, telles que le PPARG (récepteur activé par les proliférateurs de peroxysomes gamma) et la résistin, ont été associées au développement du diabète de type 2 (Kadowaki et al., 2006; Steppan et al., 2001). Le PPARG est crucial pour la sensibilité à l'insuline et les agonistes du PPARG, tels que les thiazolidinediones, sont utilisés pour traiter le diabète de type 2 (Kahn et al., 2000). La résistin, une hormone adipocytaire, est également impliquée dans la résistance à l'insuline (Steppan et al., 2001). Les traitements actuels pour le diabète de type 2 comprennent des modifications du mode de vie, des médicaments oraux, tels que la metformine et les thiazolidinediones, et des injections d'insuline (American Diabetes Association, 2021). Malgré ces traitements, le diabète de type 2 demeure une cause majeure de morbidité et de mortalité, avec un taux de mortalité plus élevé chez les personnes atteintes de diabète par rapport à celles qui ne le sont pas (Saeedi et al., 2019). Les complications cardiovasculaires sont particulièrement préoccupantes pour les personnes atteintes de diabète de type 2. Elles présentent un risque accru de maladie coronarienne, d'accident vasculaire cérébral et d'autres maladies cardiovasculaires (Emerging Risk Factors Collaboration, 2010). La gestion efficace de la glycémie et la réduction des autres facteurs de risque, tels que l'hypertension et l'hyperlipidémie, sont cruciales pour minimiser ces complications et améliorer la qualité de vie des personnes atteintes de diabète de type 2.

Cependant, les thiazolidinediones, bien que efficaces dans le traitement du diabète de type 2, sont associées à divers effets indésirables, tels que la prise de poids, l'oedème et une augmentation du risque de fractures (Dormandy et al., 2005; Nissen & Wolski, 2007). Par conséquent, il est crucial de poursuivre la recherche afin de développer de nouveaux traitements plus sûrs et efficaces pour le diabète de type 2.

### 2. Le Diabète de type 2

#### 2.1 DÉFINITION

Le diabète de type 2 (DT2) est une maladie métabolique chronique caractérisée par une résistance à l'insuline et une altération de la sécrétion d'insuline par les cellules bêta du pancréas. Contrairement au diabète de type 1, où il y a une destruction auto-immune des cellules bêta du pancréas, le diabète

de type 2 résulte généralement d'un mode de vie sédentaire, d'une mauvaise alimentation, de l'obésité et de facteurs génétiques. Les personnes atteintes de diabète de type 2 présentent des taux élevés de glucose dans le sang (hyperglycémie), ce qui peut entraîner des complications à long terme, telles que des problèmes cardiovasculaires, rénaux, nerveux et oculaires. Le traitement du diabète de type 2 implique généralement des modifications du mode de vie, telles qu'une alimentation équilibrée et l'exercice, ainsi que des médicaments antidiabétiques pour aider à contrôler la glycémie (American Diabetes Association, 2021).

## 2.2 PRÉVALENCE

La prévalence du diabète de type 2 a considérablement augmenté au cours des dernières décennies à l'échelle mondiale. Selon le rapport mondial sur le diabète de l'Organisation mondiale de la santé (OMS) en 2021, environ 537 millions de personnes sont atteintes de diabète dans le monde, dont 90 % sont des cas de diabète de type 2 (Organisation mondiale de la santé, 2021). Au Maroc, le diabète de type 2 est également en hausse. Selon une étude réalisée en 2019 par l'International Diabetes Federation (IDF), le nombre de personnes atteintes de diabète au Maroc était d'environ 2,5 millions, avec une prévalence estimée à 12,4 % de la population adulte (International Diabetes Federation, 2019).

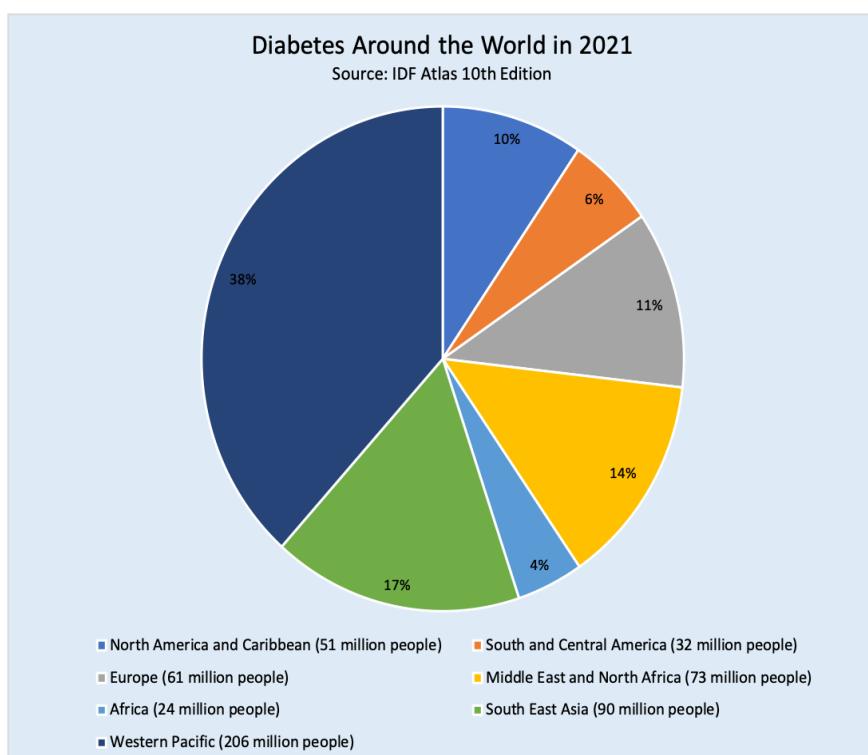


FIGURE 24. PRÉVALENCE DE DIABÈTE EN 2021 DANS LE MONDE SELON LA FÉDÉRATION INTERNATIONALE DE DIABÈTE (IDF).

## 2.3 PHYSIOPATHOLOGIE DE DIABÈTE DE TYPE 2

Le diabète de type 2 (T2DM) étant le plus répandu dans le monde selon l'OMS (figure 2), est caractérisé par une résistance à l'insuline, une augmentation de la production de glucose par le foie

et une réduction de la sécrétion d'insuline par les cellules  $\beta$  pancréatiques. La résistance à l'insuline, qui se produit principalement dans le muscle, le foie et le tissu adipeux, se développe à la suite d'interactions entre des facteurs environnementaux et génétiques, notamment l'obésité, l'inactivité physique, l'âge, le sexe, l'origine ethnique et les antécédents familiaux. La progression de la résistance à l'insuline au T2DM implique une réponse de défaillance des cellules  $\beta$  qui comprend des changements de la masse et de la fonction des cellules  $\beta$ , une augmentation de l'apoptose des cellules  $\beta$  et une altération de l'action des hormones incrétines. En plus de la résistance à l'insuline et de la dysfonction des cellules  $\beta$ , d'autres anomalies physiopathologiques sont impliquées dans le développement du T2DM, notamment le métabolisme anormal des lipides hépatiques, l'inflammation subclinique et l'activation du système rénine-angiotensine-aldostérone. Kahn, S. E., Cooper et al., 2014).

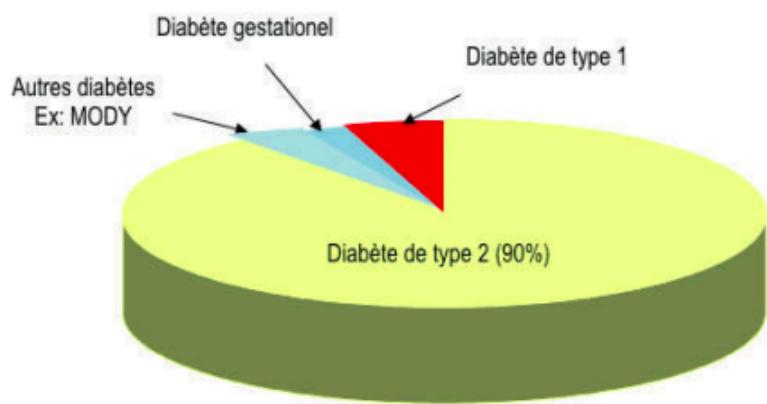


FIGURE 25. CLASSIFICATION DES DIABÈTES SELON L'OMS.

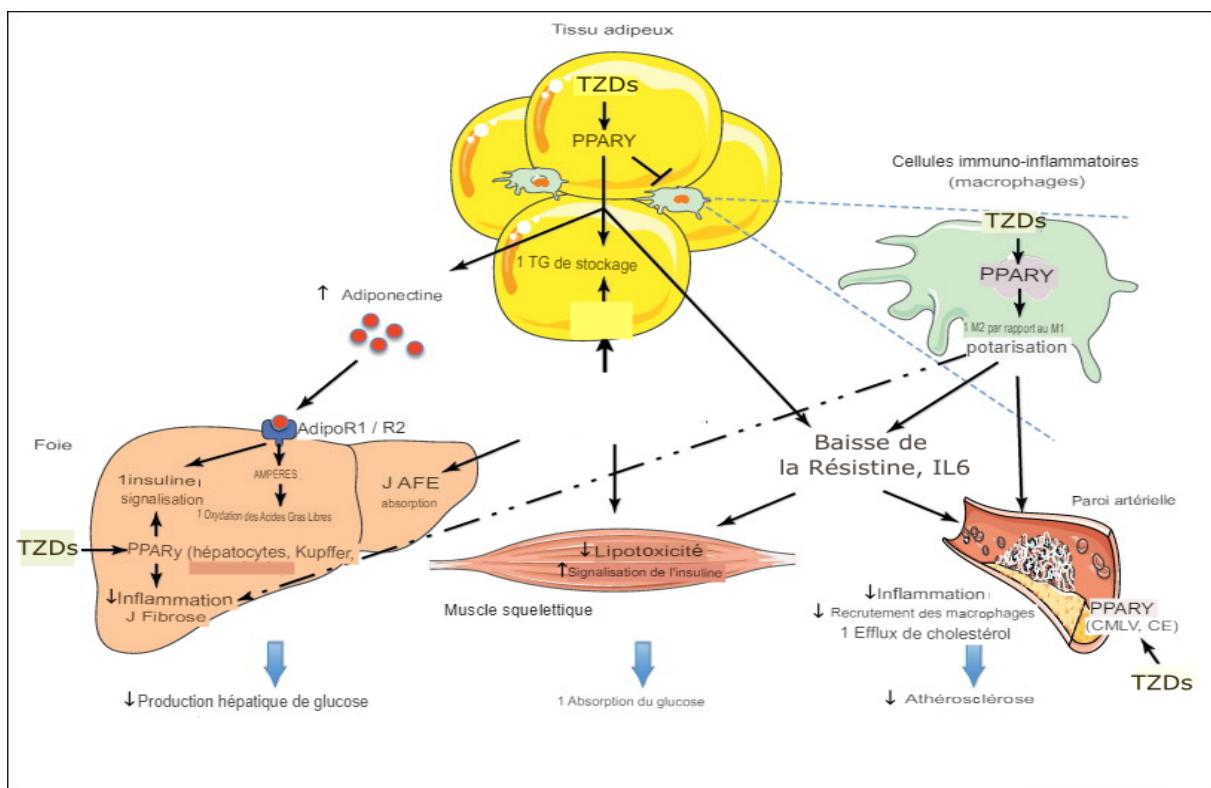
#### 2.4 TRAITEMENT DE DT2 PAR LES THIAZOLIDINEDIONES

Les thiazolidinediones (TZD) sont une classe de médicaments antidiabétiques qui agissent en augmentant la sensibilité des tissus à l'insuline. La rosiglitazone et la pioglitazone sont les deux principaux médicaments de cette classe. Les TZD ont été découvertes dans les années 1990, et la rosiglitazone a été approuvée par la FDA en 1999, suivie de la pioglitazone en 2000. Ces médicaments ont été largement utilisés pour traiter le diabète de type 2 en raison de leur efficacité à réduire la glycémie, de leur facilité d'utilisation et de leur profil d'innocuité relativement bon. Cependant, des inquiétudes ont été soulevées quant à leur sécurité cardiovasculaire, en particulier pour la rosiglitazone, qui a été associée à une augmentation du risque d'événements cardiovasculaires. En réponse à ces préoccupations, l'utilisation de la rosiglitazone a été fortement restreinte ou interdite dans de nombreux pays, et la pioglitazone a également fait l'objet d'un examen attentif. Les TZD sont une classe de médicaments antidiabétiques qui ont été découverts dans les années 1990. Ils ont été largement utilisés pour traiter le diabète de type 2, mais des inquiétudes ont été soulevées quant à

leur sécurité cardiovasculaire. Toutefois, des études récentes ont suggéré que la pioglitazone pourrait avoir des avantages cardiovasculaires (Yki-Järvinen, 2019).

#### 2.4.1 CIBLE BIOLOGIQUE DE THIAZOLIDINEDIONES

La découverte de la cible biologique des thiazolidinediones (TZD) a été le résultat de recherches menées dans les années 1990 pour comprendre les mécanismes sous-jacents à la résistance à l'insuline (Lehmann et al., 1995). Les études ont révélé que la résistance à l'insuline est due en partie à une diminution de l'expression des récepteurs de l'insuline dans les tissus cibles. Des recherches supplémentaires ont montré que les TZD agissent en activant un récepteur nucléaire appelé le récepteur activé par les proliférateurs de peroxyosomes gamma (PPAR- $\gamma$ ) (Lehmann et al., 1995). Ce récepteur est présent dans les adipocytes, les cellules musculaires et les cellules hépatiques, et régule l'expression de nombreux gènes impliqués dans la régulation du métabolisme des lipides et des glucides. Les TZD se lient spécifiquement au PPAR- $\gamma$  et activent sa transcription, ce qui conduit à une augmentation de la sensibilité des tissus à l'insuline et à une diminution de la production hépatique de glucose (Lehmann et al., 1995). Les effets des TZD sur la sensibilité à l'insuline sont donc indirects, résultant d'une modulation de l'expression génique dans les tissus cibles. Les recherches sur les TZD ont également montré que ces médicaments ont des effets anti-inflammatoires et anti-athérosclérotiques, qui peuvent contribuer à leurs effets bénéfiques sur les complications du diabète. Cependant, l'activation du PPAR- $\gamma$  peut également avoir des effets indésirables, tels que la prise de poids, l'œdème et l'augmentation du risque de fracture osseuse (Lehmann et al., 1995).



## FIGURE 26. MODE D'ACTION DES THIAZOLIDINEDIONES SUITE À SON FIXATION SUR LE RÉCEPTEUR PPAR $\gamma$ .

### 3. PPAR- $\gamma$ : Le récepteur activé par les proliférateurs de peroxysomes gamma

#### DÉCOUVERTE DE PPAR- $\gamma$

Dans les années 1980, les scientifiques ont commencé à étudier les peroxysomes, des organites cellulaires impliqués dans divers processus métaboliques, notamment la dégradation des acides gras. Ils ont observé que certains produits chimiques, tels que les hypolipidémiants de la famille des fibrates, avaient la capacité de stimuler la prolifération de peroxysomes dans les cellules hépatiques de rongeurs, mais leur mécanisme d'action restait mystérieux. En 1990, Ronald M. Evans et ses collègues ont identifié un nouveau type de récepteur nucléaire, qu'ils ont appelé "récepteur activé par les proliférateurs de peroxysomes" (Peroxisome Proliferator-Activated Receptor, PPAR) en raison de sa capacité à se lier à des produits chimiques qui stimulent la prolifération de peroxysomes. Ce premier PPAR a été appelé PPAR $\alpha$ . Peu après, deux autres types de PPAR ont été découverts : PPAR $\beta/\delta$  et PPAR $\gamma$ . Ce dernier a attiré l'attention en raison de son rôle crucial dans l'adipogenèse, c'est-à-dire la formation de nouvelles cellules adipeuses à partir de précurseurs. Parallèlement, un autre groupe de chercheurs étudiait les thiazolidinediones (TZD), une classe de médicaments utilisés pour le traitement du diabète de type 2. Ils ont découvert que ces médicaments agissent en se liant à un récepteur spécifique, qui s'est avéré être PPAR $\gamma$ . Cette découverte a permis de comprendre le mécanisme d'action des TZD et a ouvert la voie à de nouvelles recherches sur les rôles de PPAR $\gamma$  dans le métabolisme et la régulation de l'insuline.

#### 3.1 RÔLE DE PPAR- $\gamma$

Le récepteur activé par les proliférateurs de peroxysomes gamma (PPAR- $\gamma$ ) est un récepteur nucléaire impliqué dans la régulation de nombreux processus biologiques, y compris la différenciation cellulaire, la réponse inflammatoire et le métabolisme des lipides et des glucides (Grygiel-Górniak, 2014). Le PPAR- $\gamma$  est principalement exprimé dans les adipocytes, mais il est également présent dans d'autres tissus, tels que les cellules musculaires, hépatiques et intestinales. Le PPAR- $\gamma$  est activé par un certain nombre de ligands naturels, tels que les acides gras et les eicosanoïdes, ainsi que par des médicaments tels que les thiazolidinediones (TZD) (Parks et al., 1999). L'activation du PPAR- $\gamma$  conduit à la transcription de nombreux gènes impliqués dans le métabolisme des lipides et des glucides, y compris la régulation de la synthèse et du stockage des graisses. Des études ont également montré que le PPAR- $\gamma$  joue un rôle important dans la régulation de la réponse inflammatoire en réprimant l'expression de gènes pro-inflammatoires tels que l'interleukine-1 $\beta$  et le facteur de nécrose tumorale alpha (Li et al., 2011). En outre, l'activation du PPAR- $\gamma$  peut contribuer à la réduction de la résistance à l'insuline en augmentant la sensibilité des tissus à l'insuline (Grygiel-Górniak, 2014). Parmi les gènes régulé par le PPAR- $\gamma$  est le gène RETN qui code pour la protéine résistin (Maeda et al., 2001).

### 3.2 MÉCANISME D'ACTIVATION DE PPAR- $\gamma$

#### 3.2.1 ACTIVATION PAR UN AGONISTE COMPLET

L'activation du récepteur PPAR- $\gamma$  par un agoniste complet, comme la rosiglitazone, implique une série de changements conformationnels et d'interactions au niveau moléculaire. La rosiglitazone se lie dans la poche de liaison profonde du domaine de liaison au ligand (LBD) de PPAR- $\gamma$ . Le résidu crucial dans le LBD de PPAR- $\gamma$  pour la liaison de la rosiglitazone est le résidu Histidine 323 (His 323) sur l'hélice 5 ( $\alpha$ 5). Cette histidine forme une liaison hydrogène avec l'oxygène polaire de la rosiglitazone (Nolte et al., 1998). De plus, la rosiglitazone forme des liaisons hydrogène avec le résidu Serine 289 (Ser289) sur l'hélice 3 ( $\alpha$ 3) et le résidu Cysteine 285 (Cys285) (Bruning et al., 2007). Le résidu Tyr473, situé sur l'hélice 12 ( $\alpha$ 12), joue un rôle crucial dans la stabilisation de la conformation active de PPAR- $\gamma$ . Lorsqu'un agoniste complet comme la rosiglitazone se lie à PPAR- $\gamma$ , cela entraîne un changement conformationnel qui fait pivoter l'hélice 12 et place Tyr473 dans une position où il peut former une liaison hydrogène avec le ligand. Cette liaison hydrogène contribue à stabiliser la conformation active de PPAR- $\gamma$  (Nolte et al., 1998). De plus, Tyr473 forme également une liaison hydrogène avec le résidu His449, ce qui contribue à la stabilisation de l'interface entre les hélices 3 ( $\alpha$ 3), 4 ( $\alpha$ 4) et 12 ( $\alpha$ 12) (Xu et al., 1999). Lorsque la rosiglitazone se lie, elle provoque un changement de conformation dans l'hélice 12 ( $\alpha$ 12), également appelée boucle de fonction d'activation-2 (AF-2). Cette boucle peut ensuite recruter des co-activateurs, tels que le coactivateur de récepteur nucléaire (NCoR) et le coactivateur de récepteur nucléaire associé au silençage médiateur (SMRT), pour aider à augmenter l'activité de transcription de PPAR- $\gamma$  (Chandra et al., 2008).

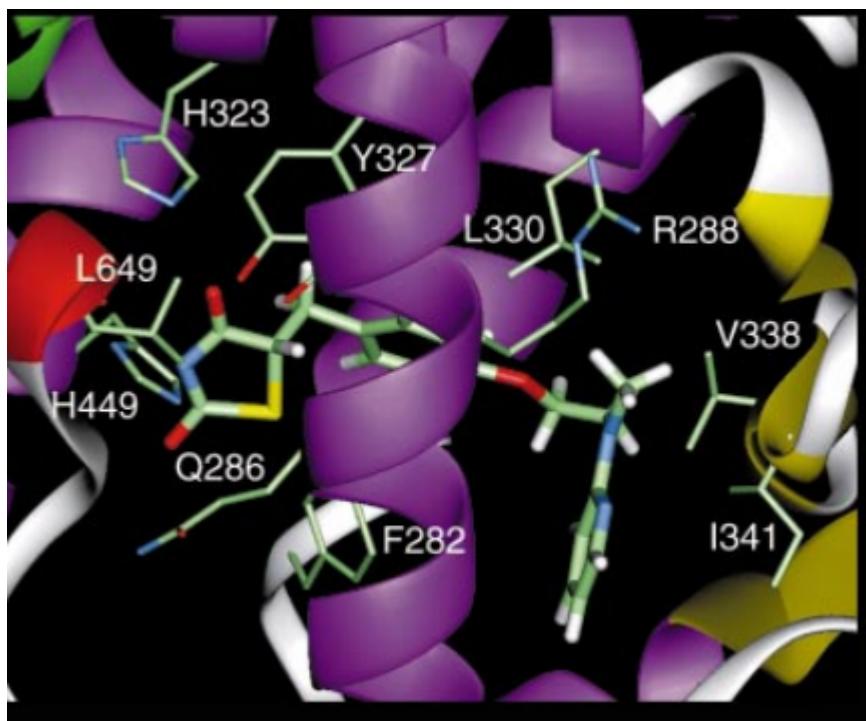


FIGURE 27. LIAISON ENTRE AGONISTE COMPLET, ROSIGLITAZONE, LES ÉLÉMENTS DE LA STRUCTURE SECONDAIRE SONT PRÉSENTÉS SOUS FORME DE RUBAN, AVEC LES ACIDES AMINÉS IMPLIQUÉS DANS LA LIAISON AU LIGAND. CHAQUE ATOME EST COLORÉ SÉPARÉMENT

SELON LE MÊME SCHÉMA DE COULEURS : VERT CLAIR POUR LE CARBONE ; ROUGE POUR L'OXYGÈNE ; BLEU POUR L'AZOTE ; JAUNE POUR LE SOUFRE. (NOLTE ET AL., 1998).

### 3.2.2 ACTIVATION DE PPAR- $\gamma$ PAR UN AGONISTE PARTIEL

Les agonistes partiels de PPAR- $\gamma$ , tels que le SR1664, activent le récepteur de manière différente par rapport aux agonistes complets. Ils sont conçus pour activer PPAR- $\gamma$  sans provoquer les effets secondaires indésirables associés aux agonistes complets, tels que la prise de poids. Lorsqu'un agoniste partiel se lie à PPAR- $\gamma$ , il ne cause pas de changement conformationnel majeur dans l'hélice 12 ( $\alpha$ 12), également connue sous le nom de boucle de fonction d'activation-2 (AF-2). Contrairement aux agonistes complets, qui stabilisent l'hélice 12 dans une position qui favorise le recrutement de co-activateurs, les agonistes partiels laissent l'hélice 12 dans une position relativement flexible (Bruning et al., 2007). Cette différence conformationnelle se traduit par une affinité réduite pour les co-activateurs, ce qui permet à PPAR- $\gamma$  d'éviter certaines des réponses transcriptionnelles associées aux agonistes complets. En même temps, les agonistes partiels conservent une certaine activité agoniste, ce qui leur permet d'induire une réponse transcriptionnelle, bien que à un niveau inférieur à celui des agonistes complets. Concernant les liaisons hydrogène, l'agoniste partiel SR1664 fait également des liaisons hydrogène avec His323 et Tyr473, comme les agonistes complets. Cependant, ces interactions ne sont pas suffisantes pour provoquer un changement conformationnel majeur dans l'hélice 12 (Marciano et al., 2015). Ainsi, l'activation de PPAR- $\gamma$  par des agonistes partiels implique un équilibre subtil entre l'activation et la répression, permettant une modulation plus précise de l'activité transcriptionnelle.

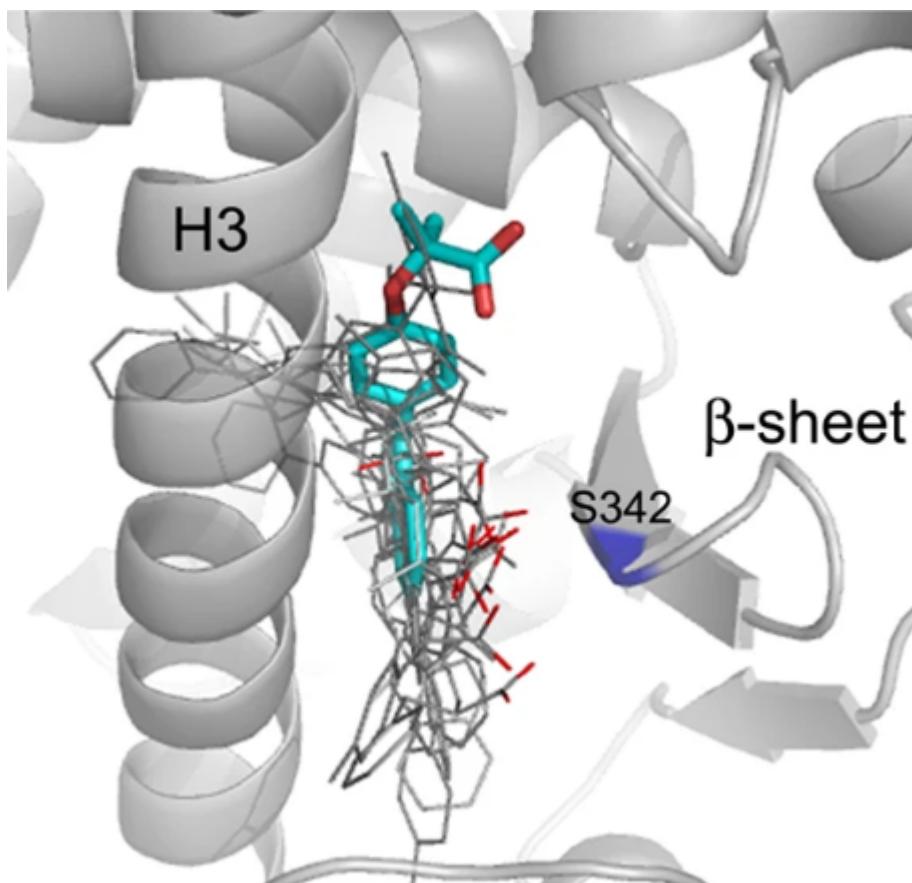


FIGURE 28. SUPERPOSITION DE COMPLEXES PPAR $\gamma$  (GRIS) AVEC DES AGONISTES PARTIELS CONNUS (CODES PDB : 3D6D, 4PVU, 4PWL, 4JL4, 4JAZ, 4E4K, 2Q5P, 2Q6S, 2Q5S, 4E4Q, 5F9B) SUR LE COMPLEXE PPAR $\gamma$ .

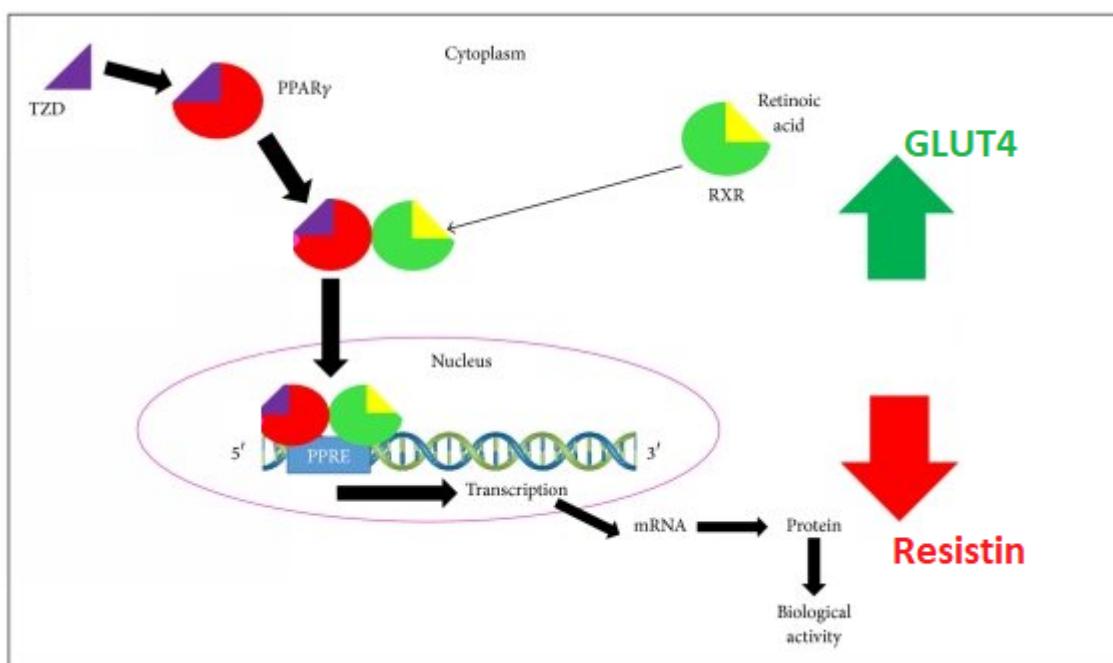
#### 4. La résistine

La résistine est une adipokine produite principalement par les adipocytes blancs et impliquée dans la régulation de l'homéostasie énergétique et de la sensibilité à l'insuline. Le gène RETN est situé sur le chromosome 19p13.2 et contient trois exons qui codent pour la résistine protéine de 108 acides aminés. Elle est également produite par d'autres tissus, tels que les macrophages et les cellules endothéliales, et peut jouer un rôle dans la réponse inflammatoire (Gao et al., 2017). Des études ont montré que les niveaux de résistine sont augmentés chez les souris obèses et les humains atteints de diabète de type 2 (Steppan et al., 2001). Des études ont également montré que la résistine peut jouer un rôle dans la résistance à l'insuline en inhibant la signalisation de l'insuline dans les tissus périphériques, tels que le foie et les muscles squelettiques (Fasshauer et Paschke, 2003). En outre, la résistine peut jouer un rôle dans la régulation de la lipogenèse et de la lipolyse, ce qui peut avoir un impact sur l'homéostasie énergétique (Gao et al., 2017).

##### 4.1 DÉCOUVERTE DE LA RÉSISTINE

L'article de Steppan et al. (2001) décrit la découverte de la résistine, une adipokine produite par les adipocytes et impliquée dans la régulation de l'homéostasie énergétique et de la sensibilité à l'insuline. Les auteurs ont montré que les niveaux de résistine sont augmentés chez les souris obèses

et les humains atteints de diabète de type 2. Ils ont également montré que la résistine est produite par les adipocytes et est exprimée à des niveaux élevés chez les souris obèses et les humains atteints de diabète de type 2. Le lien entre le PPAR- $\gamma$  et le gène RETN qui code pour la résistine a été établi par des études ultérieures. En effet, le PPAR- $\gamma$  a été identifié comme un régulateur de la transcription du gène RETN. Le PPAR- $\gamma$  se lie directement à la région promotrice du gène RETN et régule ainsi son expression. Des études ont également montré que les thiazolidinediones (TZD), des agonistes du PPAR- $\gamma$  utilisés dans le traitement du diabète de type 2, peuvent réduire les niveaux de résistine chez les souris obèses et les humains atteints de diabète de type 2 (Yamauchi et al., 2001). Cela suggère que l'activation du PPAR- $\gamma$  peut avoir un impact sur la régulation de la résistine et peut contribuer à améliorer la sensibilité à l'insuline chez les patients atteints de diabète de type 2.



**FIGURE 29. BAISSE DE L'EXPRESSION DE LA RESISTINE LORS DE L' ACTIVATION DE PPAR $\gamma$  PAR SES LIGANDS EXOGÈNES LES THIAZOLIDINEDIONES (TZDs).**

#### 4.2 LA RÉSISTINE : MEDIATEUR ENTRE DT2 ET MCV

Le diabète de type 2 est un trouble métabolique complexe qui est associé à une large gamme de complications, y compris les maladies cardiovasculaires (MCV). En effet, les MCV sont une cause majeure de morbidité et de mortalité chez les patients atteints de diabète de type 2, et le risque de MCV est deux à quatre fois plus élevé chez ces patients que dans la population générale. Un médiateur potentiel du lien entre le diabète de type 2 et les MCV est le facteur de transcription PPAR- $\gamma$ , qui joue un rôle crucial dans la différenciation des adipocytes, le métabolisme des lipides et la sensibilité à l'insuline. PPAR- $\gamma$  est fortement exprimé dans le tissu adipeux et est une cible de la classe de médicaments antidiabétiques les thiazolidinediones qui ont montré une amélioration de la sensibilité à l'insuline, réduire l'inflammation et abaisser la pression artérielle chez

les patients atteints de diabète de type 2, mais leur utilisation est également associée à un risque accru d'insuffisance cardiaque (Nesto et al., 2004). Un autre médiateur potentiel du lien entre le diabète de type 2 et les MCV est l'adipokine résistine. La résistine est produite par le tissu adipeux et a été montrée pour favoriser l'inflammation et la résistance à l'insuline dans divers tissus, y compris la paroi vasculaire. Plusieurs études ont lié des niveaux élevés de résistine à un risque accru de MCV et d'événements cardiovasculaires chez les patients atteints de diabète de type 2 (Gao et al., 2017). Des études récentes ont suggéré qu'il peut y avoir une interaction complexe entre PPAR- $\gamma$ , la résistine et les MCV chez les patients atteints de diabète de type 2. Par exemple, l'activation de PPAR- $\gamma$  a été montrée pour réduire l'expression de la résistine dans le tissu adipeux, ce qui peut à son tour diminuer le risque de MCV chez ces patients (Mandard et Müller, 2005).

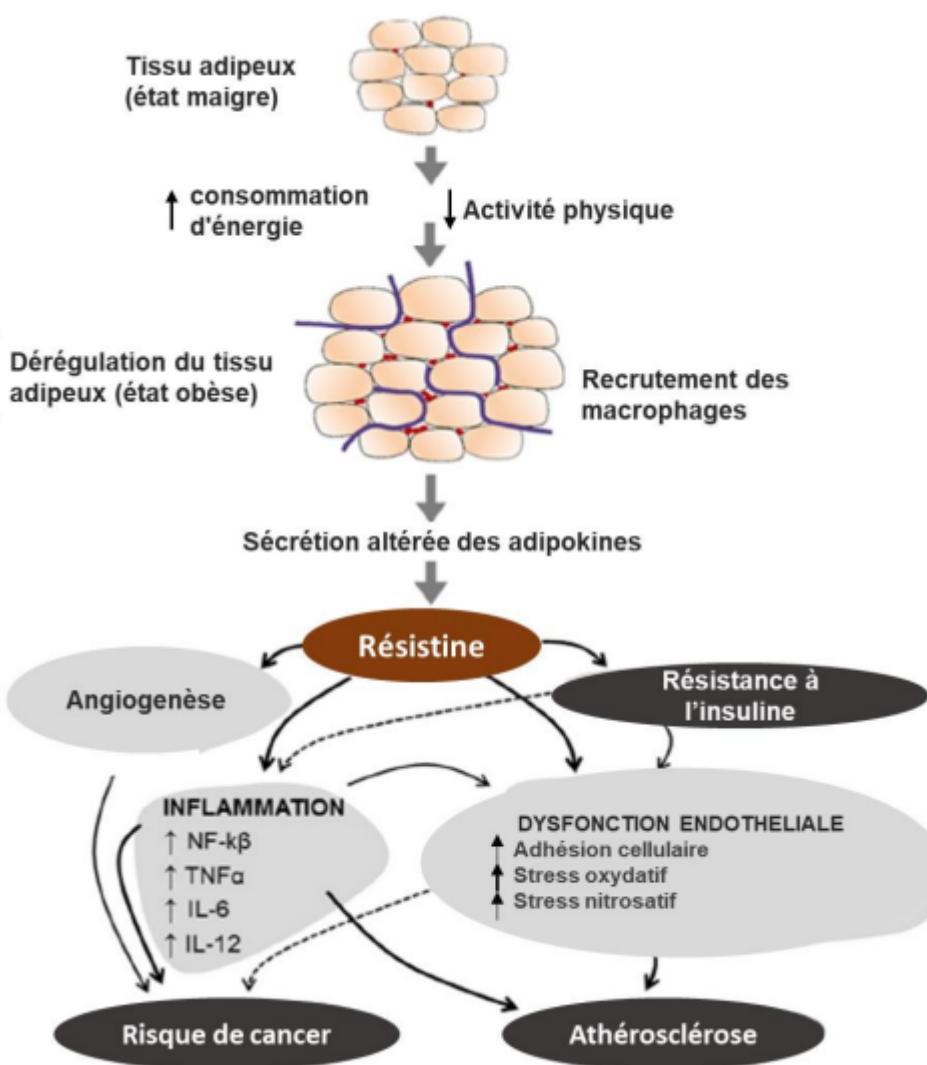


FIGURE 30. IMPLICATION DE LA RÉSISTINE DANS LA SURVENUE DE DIABÈTE TYPE 2 ET D'ATHÉROSCLÉROSE. IL, INTERLEUKINE; NF- $\kappa$ B, FACTEUR NUCLÉAIRE-KB; TNF $\alpha$ , FACTEUR DE NÉCROSE TUMORALE-A (CODOÑERFRANCH ET ALONSO-IGLESIAS 2015).

## 5. Complications cardiovasculaires de Diabète type 2

Le diabète de type 2 (DT2) est une affection chronique qui peut entraîner une multitude de complications à long terme, parmi lesquelles les complications cardiovasculaires sont particulièrement préoccupantes. Les patients atteints de DT2 ont un risque deux à quatre fois plus élevé de maladies cardiovasculaires (MCV) par rapport à ceux sans diabète (Emerging Risk Factors Collaboration, 2010). Les MCV, notamment l'infarctus du myocarde, l'insuffisance cardiaque et les accidents vasculaires cérébraux, sont la principale cause de morbidité et de mortalité chez les patients atteints de DT2 (Rawshani et al., 2017).

Plusieurs mécanismes pathophysiologiques sont impliqués dans l'augmentation du risque de MCV chez les patients atteints de DT2. Le DT2 est associé à une dyslipidémie caractérisée par des taux élevés de triglycérides, des taux faibles de lipoprotéines de haute densité (HDL) et une prédominance de particules de lipoprotéines de basse densité (LDL) de petite taille et denses, ce qui favorise l'athérosclérose (Taskinen & Borén, 2015). De plus, l'hyperglycémie chronique contribue au stress oxydatif, à la dysfonction endothéliale, à l'inflammation et à la thrombogenèse, qui sont tous des facteurs clés de la pathogenèse des MCV (figure 3) (Louis et al, 2021). Cependant, malgré le risque élevé, les interventions ciblant le contrôle glycémique, la gestion des lipides et la pression artérielle, ainsi que les facteurs de risque du mode de vie, peuvent réduire considérablement le risque de MCV chez les patients atteints de DT2 (Cosentino et al., 2020).

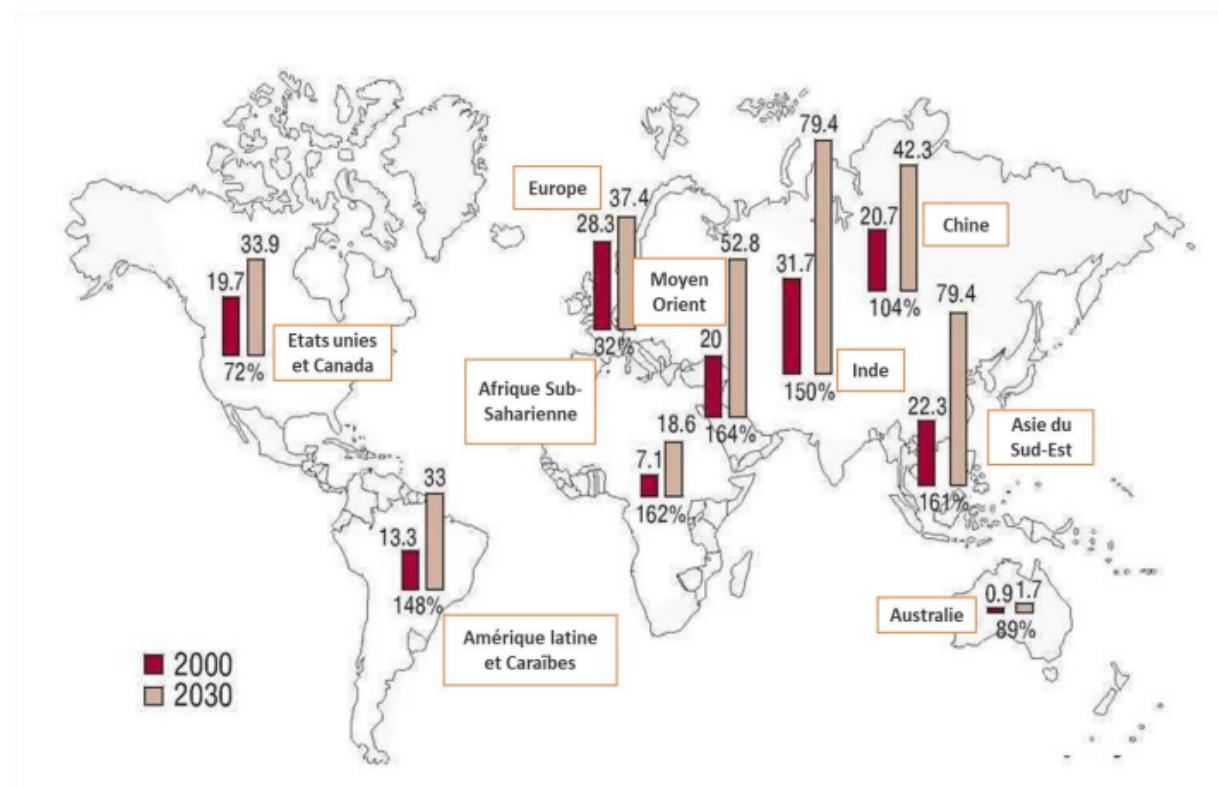


FIGURE 31. LE DIABÈTE TYPE 2 (EN MILLIONS D'INDIVIDUS) EN 2000 ET PRÉVUS POUR 2030, ADAPTÉ DE (P ZIMMET, ALBERTI, ET SHAW 2001).

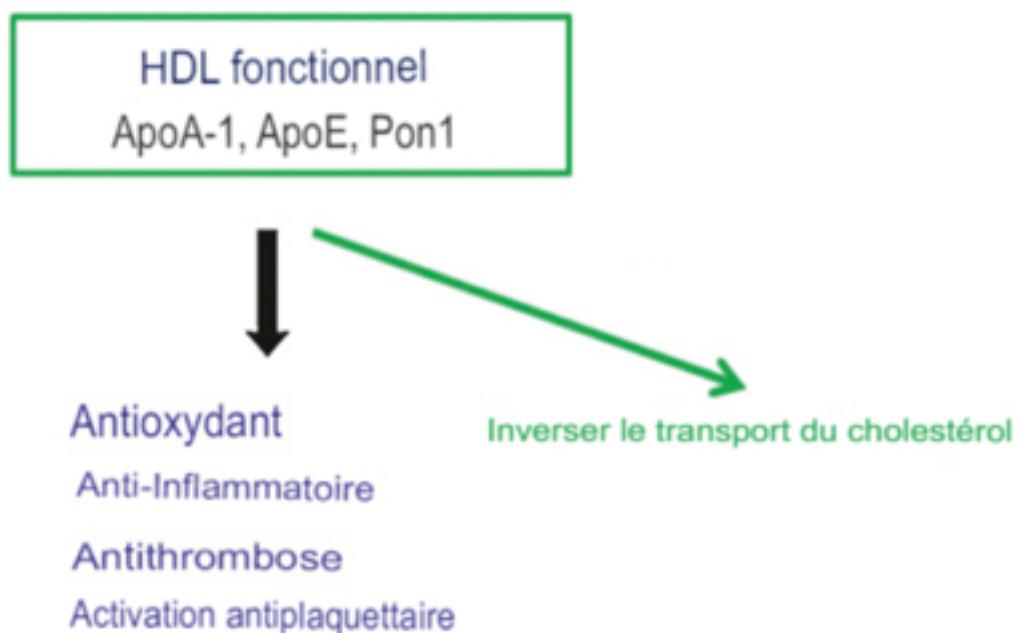
### 5.1 RAPPEL SUR LES LIPOPROTÉINES

Les lipoprotéines sont des particules complexes composées de lipides et de protéines qui jouent un rôle crucial dans le transport des lipides, y compris le cholestérol, dans le corps. Elles sont classées en différentes catégories en fonction de leur densité, notamment les lipoprotéines de haute densité (HDL), les lipoprotéines de basse densité (LDL) et les lipoprotéines de très basse densité (VLDL).

Les LDL, en revanche, sont souvent appelées "mauvaises" lipoprotéines, car un excès de LDL peut contribuer à l'accumulation de cholestérol dans les parois des artères, favorisant le développement de l'athérosclérose. Les LDL transportent le cholestérol du foie vers les tissus périphériques et déposent le cholestérol dans les cellules en se liant à des récepteurs spécifiques sur leur surface (Brown et Goldstein, 1986). L'oxydation des LDL est un facteur clé dans la formation des plaques d'athérosclérose, ce qui souligne l'importance de protéger les LDL contre l'oxydation (Steinberg, 1997).

#### 5.1.1 HDL

Les HDL, souvent appelées "bonnes" lipoprotéines, jouent un rôle clé dans le transport inverse du cholestérol. Elles sont responsables de la collecte de l'excès de cholestérol des tissus périphériques, y compris des cellules des parois des vaisseaux sanguins, et de son transport vers le foie pour être métabolisé et excrété (figure) (Nayoung et al, 2016). Les HDL agissent également comme des antioxydants et ont des propriétés anti-inflammatoires et anti-thrombotiques qui contribuent à la protection contre les maladies cardiovasculaires (Rader, 2016).



**FIGURE 32. RÔLE DE HDL DANS LE TRANSPORT INVERSE DE CHOLESTÉROL ET LA PROTECTION CONTRE LA MALADIE CARDIOVASCULAIRE (NAYOUNG ET AL, 2016).**

### 5.1.2 VLDL

Les VLDL sont des précurseurs des LDL et transportent principalement les triglycérides du foie vers les tissus périphériques. Ensuite, les VLDL sont métabolisées en LDL par des enzymes dans la circulation sanguine. Les niveaux élevés de VLDL peuvent être associés à une augmentation des taux de triglycérides sanguins, ce qui est un facteur de risque pour les maladies cardiovasculaires (Pirillo et Catapano, 2013).

### 5.1.3 LDL

Les lipoprotéines de basse densité, couramment abrégé en LDL, est souvent associé à des effets négatifs sur la santé, notamment en raison de son rôle dans l'établissement de l'athérosclérose (Kwiterovich, 2000). Les LDL sont des particules complexes qui facilitent le transport du cholestérol et d'autres lipides du reste du corps vers le foie. Le LDL est la fraction de cholestérol présente au sein de ces particules LDL. Lorsque le LDL est présent en quantités excessives, il peut s'agréger sur les parois des artères, formant ainsi des plaques athéromateuses (Sniderman et al., 2003). Une concentration élevée de LDL dans le sang est un facteur de risque significatif pour l'apparition de pathologies cardiovasculaires (Mach et al., 2020). L'excès de LDL dans les parois artérielles induit une réaction inflammatoire qui mène à l'athérosclérose, une affection caractérisée par l'accumulation de dépôts lipidiques, ou plaques, sur les parois artérielles (Lusis, 2000). Ces plaques peuvent provoquer le rétrécissement de l'artère ou se rompre, entraînant des complications graves comme un infarctus du myocarde ou un accident vasculaire cérébral (Libby, 2002). Par conséquent, le contrôle du taux de LDL est une stratégie essentielle pour la prévention des maladies cardiovasculaires.

#### a. Régulation de taux de LDL

La régulation du taux des lipoprotéines de basse densité (LDL) dans le sang est un processus complexe et multifactoriel. Au cœur de ce processus se trouvent les récepteurs LDL (LDLR), qui sont responsables de la liaison et de l'internalisation des LDL, facilitant ainsi l'élimination du LDL du flux sanguin (Goldstein & Brown, 2009). Des mutations dans le gène LDLR peuvent entraîner une hypercholestérolémie familiale, une condition caractérisée par des niveaux élevés de LDL et un risque accru de maladies cardiovasculaires (Soutar & Naoumova, 2007). La régulation du LDL est également influencée par la proprotein convertase subtilisin/kexin type 9 (PCSK9). Le PCSK9 se lie aux LDLR, augmentant ainsi leur dégradation et réduisant l'élimination du LDL du sang, ce qui entraîne une augmentation du taux de LDL (Seidah et al., 2003).

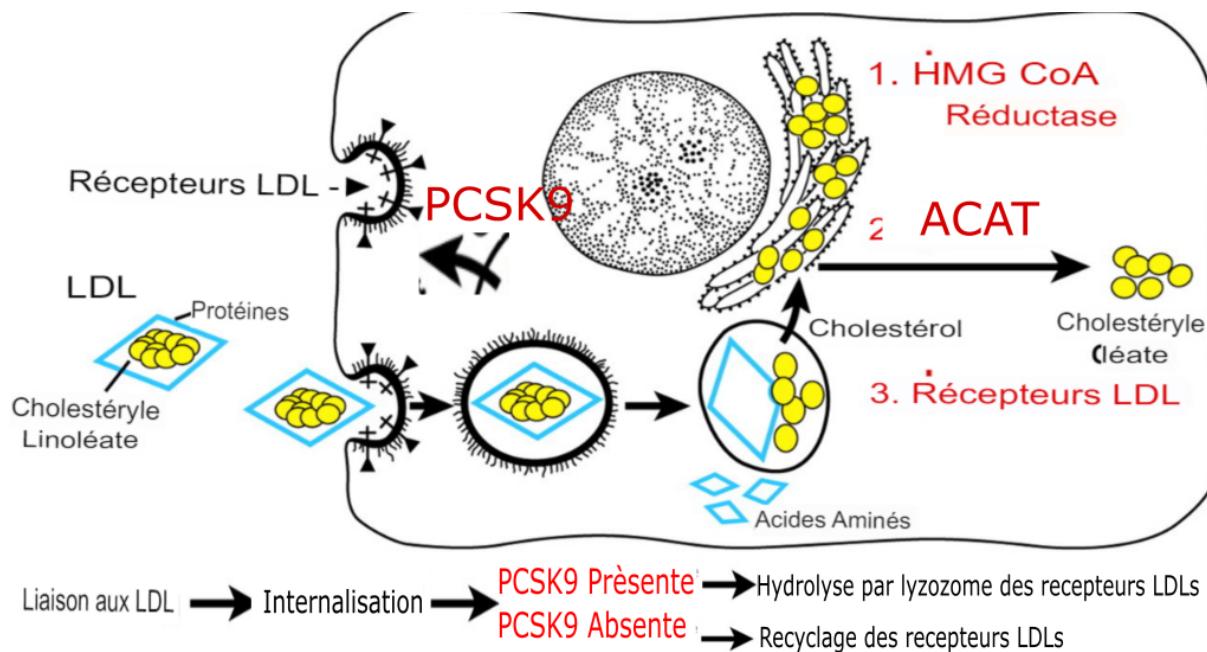


FIGURE 33. RÉGULATION DE TAUX DES LDL DÉPEND DE LA RÉGULATION DES RÉCEPTEURS LDL DES CELLULES DE MAMMIFÈRES , MODIFIÉ, (GOLDSTEIN & BROWN, 2009).

### b. LDL oxydée

L'oxydation des LDL peut être initiée par différents processus, notamment l'oxydation enzymatique et non enzymatique. L'oxydation enzymatique se produit lorsque les enzymes présentes dans les cellules inflammatoires, telles que les macrophages, génèrent des espèces réactives de l'oxygène (ROS) qui attaquent les LDL et les rendent oxydées (Parthasarathy et al., 1992). L'oxydation non enzymatique peut être provoquée par des radicaux libres, tels que les radicaux hydroxyle et les peroxydes lipidiques, qui sont générés lors de réactions chimiques complexes. Une fois oxydées, les LDL oxydées (oxLDL) deviennent pro-inflammatoires et sont reconnues par les récepteurs spécifiques sur les cellules endothéliales. Cela déclenche une réponse inflammatoire, entraînant l'adhésion des cellules inflammatoires, telles que les monocytes, aux parois des vaisseaux sanguins. Les monocytes se transforment ensuite en macrophages qui ingèrent les oxLDL et se transforment en cellules chargées de lipides, appelées cellules spumeuses (Steinberg, 1997). L'accumulation des cellules de la mousse et des lipides oxydés dans la paroi des vaisseaux sanguins forme des plaques d'athérosclérose, réduisant le calibre des vaisseaux et limitant le flux sanguin. De plus, les oxLDL stimulent la prolifération des cellules musculaires lisses de la paroi des vaisseaux, contribuant à la formation de la plaque. L'oxydation des LDL peut être favorisée par divers facteurs, notamment le tabagisme, l'hypercholestérolémie, l'hypertension artérielle, le diabète et l'inflammation chronique. Des antioxydants naturels présents dans l'alimentation, tels que les vitamines C et E, ainsi que d'autres composés bioactifs, peuvent réduire l'oxydation des LDL en neutralisant les radicaux libres et en protégeant les lipoprotéines contre les dommages (Aviram et al., 2000).

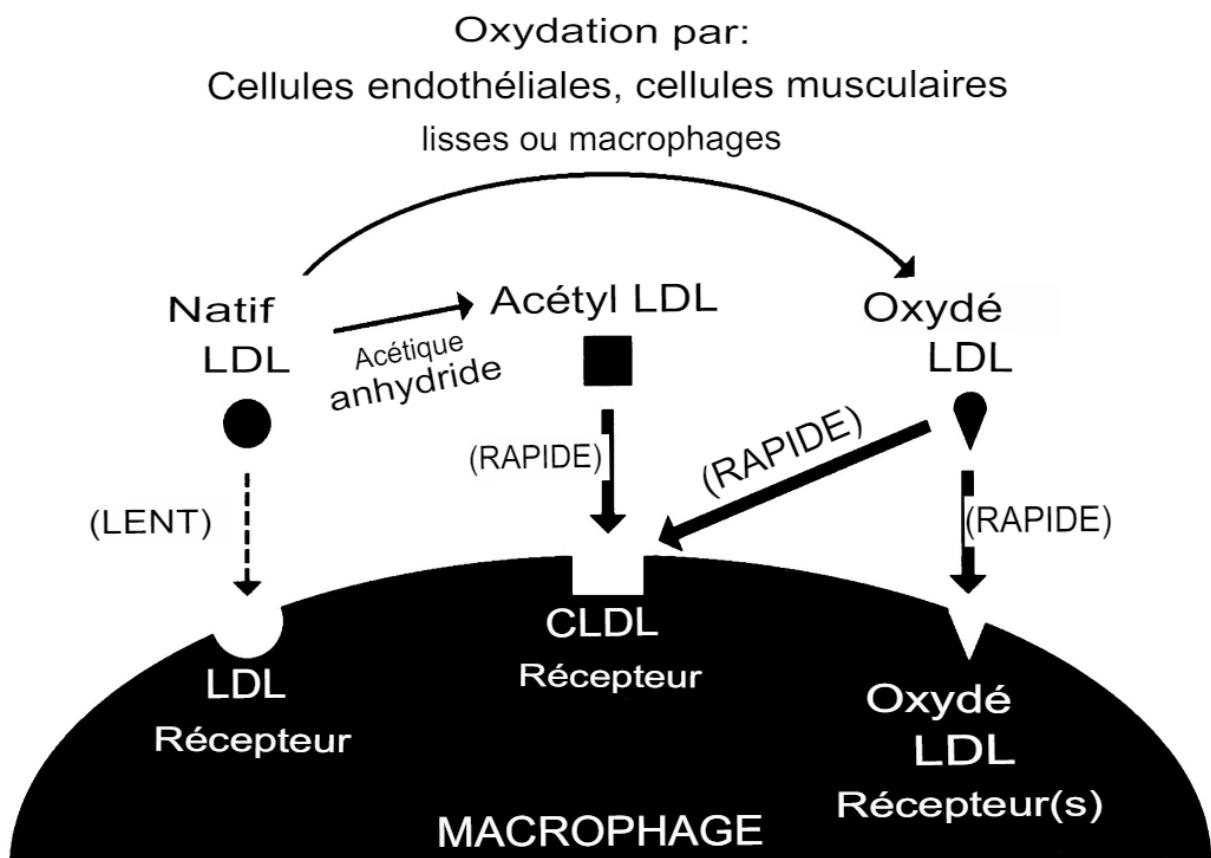


FIGURE 34. SCHÉMA ILLUSTRANT L'INTERNALISATION DE LDL NATIVES ET MODIFIÉES PAR OXYDATION PAR LES RÉCEPTEURS DES MACROPHAGES (STEINBERG, 1997).

### c. LDL et hypercholestérolémie

L'hypercholestérolémie, une condition caractérisée par des niveaux élevés de cholestérol dans le sang, est fortement corrélée à des taux élevés de cholestérol des lipoprotéines de basse densité (LDL). Le LDL est le principal contributeur à l'accumulation de plaques dans les artères, un processus connu sous le nom d'athérosclérose, qui peut mener à des maladies cardiaques et à des accidents vasculaires cérébraux (Libby, 2002). L'hypercholestérolémie peut être due à des facteurs génétiques, comme dans l'hypercholestérolémie familiale où des mutations dans le gène LDLR entravent l'élimination efficace du LDL du sang (Soutar & Naoumova, 2007). Les facteurs environnementaux, tels que le régime alimentaire et le mode de vie, jouent également un rôle dans la régulation du taux de LDL. Une alimentation riche en graisses saturées et en cholestérol peut augmenter le taux de LDL, tandis que l'activité physique peut aider à le réduire (Mannu et al., 2013). Plus récemment, le PCSK9 a été identifié comme un régulateur clé de l'hypercholestérolémie, en influençant la dégradation des récepteurs LDL et donc le taux de LDL dans le sang (Seidah et al., 2003). Les thérapies ciblant le PCSK9 se sont avérées efficaces pour réduire le taux de LDL chez les patients atteints d'hypercholestérolémie (Sabatine et al., 2017).

#### d. rôle de LDL dans l'Athérosclérose

Le LDL, la lipoprotéine de basse densité, joue un rôle déterminant dans le développement de l'athérosclérose, caractérisée par l'accumulation de plaques dans les parois artérielles (Libby, P., 2002). Les LDL peuvent s'accumuler sur les parois des artères, formant des plaques d'athérome qui réduisent le flux sanguin et peuvent conduire à des complications cardiovasculaires graves (Goldstein, J. L., & Brown, M. S., 2009). De plus, l'oxydation des LDL peut engendrer une réponse inflammatoire qui accélère le processus d'athérosclérose (MacRae et al., 2019). L'hypercholestérolémie, notamment l'hypercholestérolémie familiale, peut entraîner des taux élevés de LDL et donc augmenter le risque d'athérosclérose (Soutar, A. K., & Naoumova, R. P., 2007). En outre, des études ont montré que la réduction des niveaux de LDL, que ce soit par des modifications du mode de vie ou par des traitements médicamenteux, peut aider à réduire le risque d'athérosclérose et de maladies cardiovasculaires associées (Cholesterol Treatment Trialists' (CTT) Collaborators, 2010; Sabatine, M. S. et al., 2017).

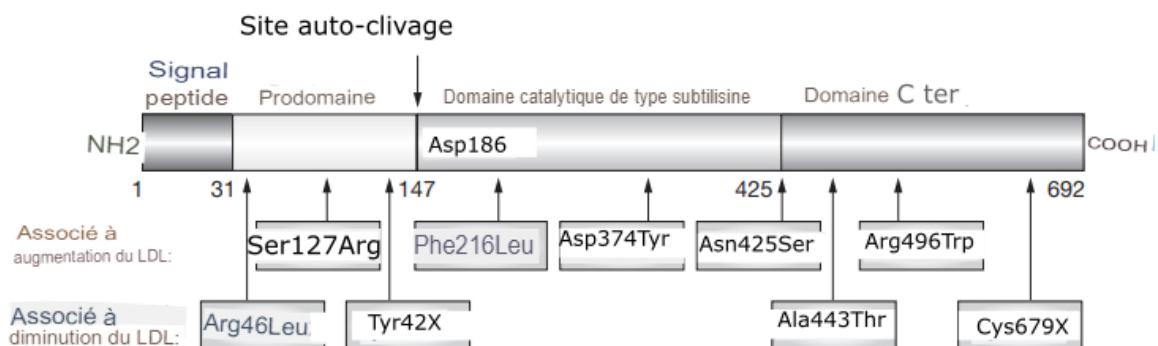
### 6. PCSK9 : NOUVELLE CIBLE THERAPEUTIQUE CARDIOVASCULAIRE

#### 6.1 DÉCOUVERTE DE PCSK9

PCSK9 est une protéine impliquée dans le métabolisme du cholestérol. Sa découverte a été faite en 2003 par deux groupes de recherche indépendants (Abifadel et al., 2003) (Cohen et al., 2006). Les études ont montré que PCSK9 régule la quantité de récepteurs de LDL (LDLR) à la surface des cellules hépatiques, ce qui a un impact sur le taux de cholestérol dans le sang (Horton et al., 2009). Les mutations du gène PCSK9 ont été identifiées comme étant liées à des variations du taux de cholestérol LDL dans le sang. Les mutations de gain de fonction du gène PCSK9 ont été étudiées en profondeur en raison de leur lien avec les maladies cardiovasculaires. Ces mutations entraînent une augmentation de l'activité de la protéine PCSK9, ce qui conduit à une diminution des récepteurs du LDL-cholestérol sur les cellules hépatiques et à une augmentation des taux de cholestérol LDL circulant dans le sang (Fan et al., 2007).

Les individus porteurs des mutations de gain de fonction de PCSK9 ont un risque accru de développer des maladies cardiovasculaires, notamment des maladies coronariennes et des accidents vasculaires cérébraux (Benn et al., 2010; Benn, Watts, Tybjærg-Hansen, & Nordestgaard, 2016). En raison de l'activité accrue de la protéine PCSK9, la dégradation des récepteurs du LDL-cholestérol est accélérée, entraînant une accumulation de cholestérol LDL dans les vaisseaux sanguins et favorisant ainsi la formation de plaques athéroscléreuses. Cependant les mutations perte-de-fonction du gène PCSK9 sont associées à une réduction significative du taux de cholestérol LDL chez les individus porteurs de ces mutations, cela se traduit par une augmentation du nombre de récepteurs du LDL-cholestérol à la surface des cellules hépatiques, qui sont responsables de l'élimination du cholestérol LDL du sang (Horton et al., 2009). Des études épidémiologiques ont démontré que les individus porteurs de

mutations perte-de-fonction de PCSK9 présentaient un risque réduit de développer des maladies cardiovasculaires telles que les accidents vasculaires cérébraux et les maladies coronariennes (Abifadel et al., 2003; Cohen et al., 2006). Cette observation suggère que la modulation de l'activité de PCSK9 peut être une stratégie thérapeutique prometteuse pour réduire le risque de ces maladies.

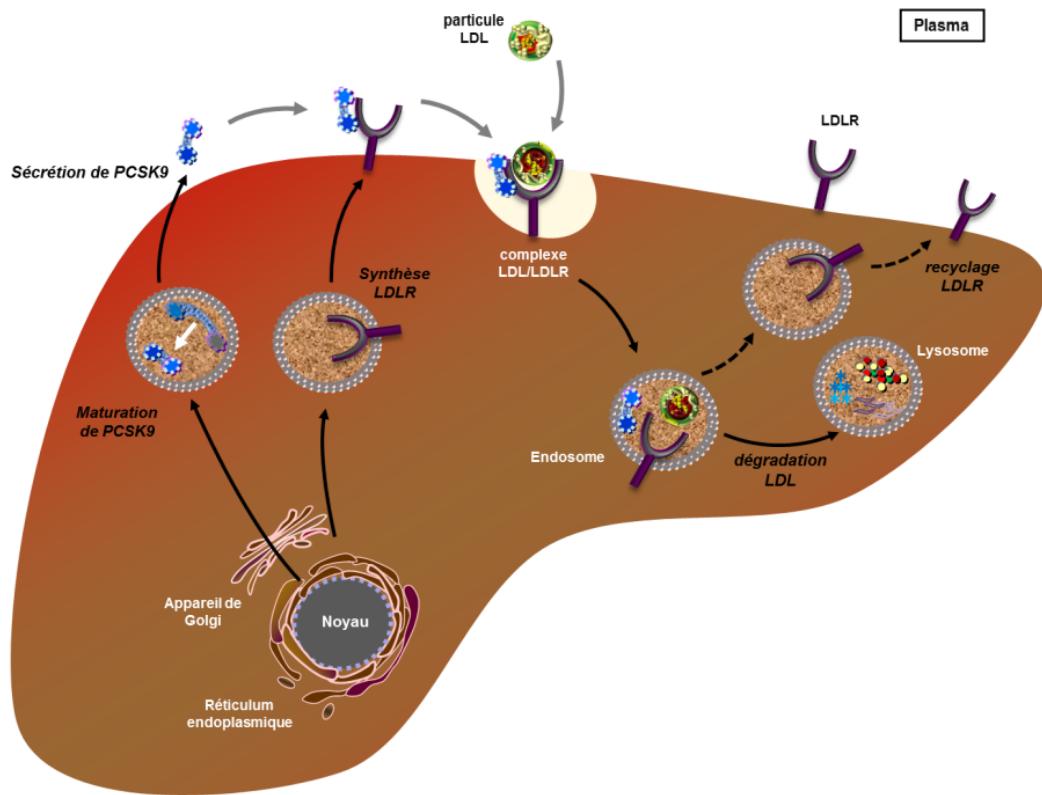


**FIGURE 35. SCHÉMA DE LA PROTÉINE PCSK9, MONTRANT LA LOCALISATION DES DOMAINES EN HOMOLOGIE AVEC LE PRODOMAINE ET LE DOMAINE CATALYTIQUE DE LA FAMILLE SUBTILISINE DES PROPROTÉINES CONVERTASES.**

**LES VARIANTS ASSOCIÉS À L'HYPERCHOLESTÉROLÉMIE FAMILIALE SONT INDIQUÉS DANS LA PREMIÈRE LIGNE , FIGURE MODIFIÉE, (SOUTAR & NAOUMOVA, 2007).**

#### 6.2 RÔLE DE PCSK9 DANS LA RÉGULATION DE TAUX DE LDL

PCSK9 est une protéine qui régule la quantité de récepteurs de LDL (LDL-R) à la surface des cellules hépatiques, ce qui a un impact sur le taux de cholestérol dans le sang. Les mutations induisant une amplification fonctionnelle de PCSK9 sont associées à une hypercholestérolémie autosomique dominante due à une croissance des concentrations en LDL. À l'inverse, des mutations perte de fonction de PCSK9 sont associées à une diminution de ces concentrations et à une réduction de 88% du risque d'apparition de maladies coronariennes. Dans le réticulum endoplasmique, l'auto clivage de Pro-PCSK9 est primordial à sa maturation et à sa sécrétion. PCSK9 s'associe avec EGF-A du récepteur aux LDL (LDLR) et engendrant sa dégradation lysosomale. A l'instar de LDLR, PCSK9 est positivement régulé par les statines via le Sterol Responsive element Binding Protein -2 (SREBP2). Les statines sont très utilisées en clinique dû à leurs effets hypcholestérolémiants, ajouter des inhibiteurs de PCSK9 pourrait améliorer leur efficacité (Le May C et al., 2009).



**Figure 36. Rôle de PCSK9 dans le recyclage des récepteurs LDL sur la surface cellulaires, (SABOURET, FARNIER, ET PUYMIRAT 2019).**

### 6.3 INHIBITEURS DE PCSK9

L'expression de PCSK9 est sous l'influence de la concentration intracellulaire de stérols. Cette interaction complexe est clairement démontrée dans la Figure 2. Lorsqu'un individu suit un traitement à base de statines, ce qui entraîne une réduction du taux de cholestérol à l'intérieur des cellules hépatiques, on observe simultanément une augmentation de l'expression des récepteurs LDL (LDL-R) ainsi que de PCSK9. Cette coordination entre l'expression de LDL-R et PCSK9, régulée par la concentration intracellulaire de cholestérol, peut potentiellement limiter l'effet pharmacologique de réduction du LDL-C, particulièrement lorsque le traitement à base de statines est administré à des doses élevées. Ce phénomène surprenant observé lors de l'administration de statines suggère qu'il pourrait être judicieux d'envisager l'inhibition de PCSK9 comme une stratégie complémentaire pour potentialiser l'efficacité du traitement à base de statines. En d'autres termes, en associant l'utilisation de statines avec des inhibiteurs de PCSK9, il serait possible d'optimiser le nombre de récepteurs LDL fonctionnels présents à la surface des hépatocytes. Cette approche synergique permettrait ainsi de maximiser l'élimination du LDL-C du sang, réduisant ainsi considérablement le risque de maladies cardiovasculaires. Les inhibiteurs de PCSK9 sont des médicaments développés pour réduire les taux de cholestérol LDL dans le sang (Sabouret, Farnier, et Puymirat 2019). Une étude menée par Michel Farnier et publiée en 2014 a démontré que les anticorps monoclonaux ciblant PCSK9 sont une

méthode efficace pour inhiber cette protéine. Ces anticorps ont été bien tolérés lors des essais de phases II et III, sans effets secondaires à court et à moyen terme. Ils permettent une diminution significative du LDL, atteignant jusqu'à 70 % avec les doses les plus efficaces. L'efficacité de ces inhibiteurs a été évaluée en combinaison avec des statines chez des patients atteints d'hypercholestérolémie familiale ou d'autres troubles métaboliques lipidiques, ainsi que chez des patients présentant une intolérance aux statines, et même en monothérapie. Actuellement, plusieurs études sont en cours pour évaluer l'efficacité de la prévention cardiovasculaire en réduisant les lipoprotéines athérogènes grâce à l'inhibition de PCSK9. L'objectif de ces essais est de déterminer les avantages cardiovasculaires potentiels et de clarifier la sécurité associée à des niveaux très bas de LDL (Farnier, 2015).

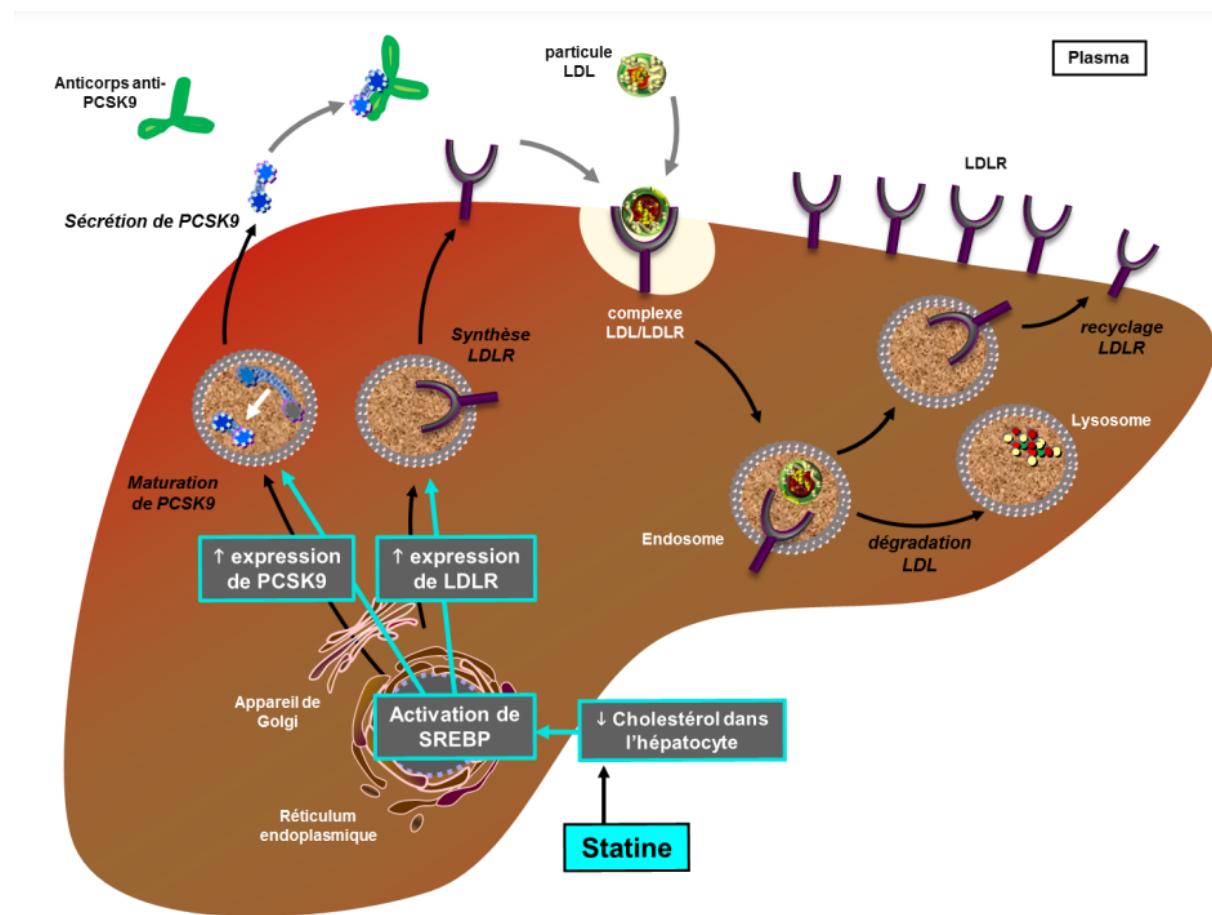


FIGURE 37. MÉCANISME D'ACTION COMPLÉMENTAIRE DES STATINES ET DES ANTICORPS MONOCLONAUX DIRIGÉS CONTRE PCSK9 (SABOURET, FARNIER, ET PUYMIRAT 2019).

## CHAPITRE 4. LES PHYTO-COMPOSÉS DANS LE TRAITEMENT DE DT2 ET SES COMPLICATIONS ET BASE DE DONNÉES

### 1. Phyto-composés

Les phyto-composés sont des composés biologiquement actifs présents dans les plantes. Ils ne sont pas essentiels à la vie humaine au même titre que les vitamines et les minéraux, mais ils ont des effets bénéfiques sur la santé humaine en raison de leurs propriétés antioxydantes, anti-inflammatoires et anticancéreuses (Liu, 2003). Les phyto-composés sont classés en différentes catégories, notamment les polyphénols, les caroténoïdes, les glucosinolates et les phytoestrogènes, chacune ayant des effets spécifiques sur la santé (Manach et al., 2004). Les polyphénols, par exemple, peuvent aider à prévenir les maladies cardiovasculaires, tandis que les caroténoïdes peuvent aider à prévenir certains types de cancer. Il est important de noter que la plupart des phyto-composés sont mieux absorbés par l'organisme lorsqu'ils sont consommés dans le cadre d'un régime alimentaire équilibré, riche en fruits, légumes, céréales complètes et légumineuses.

#### 1.1 CARACTÉRISTIQUES

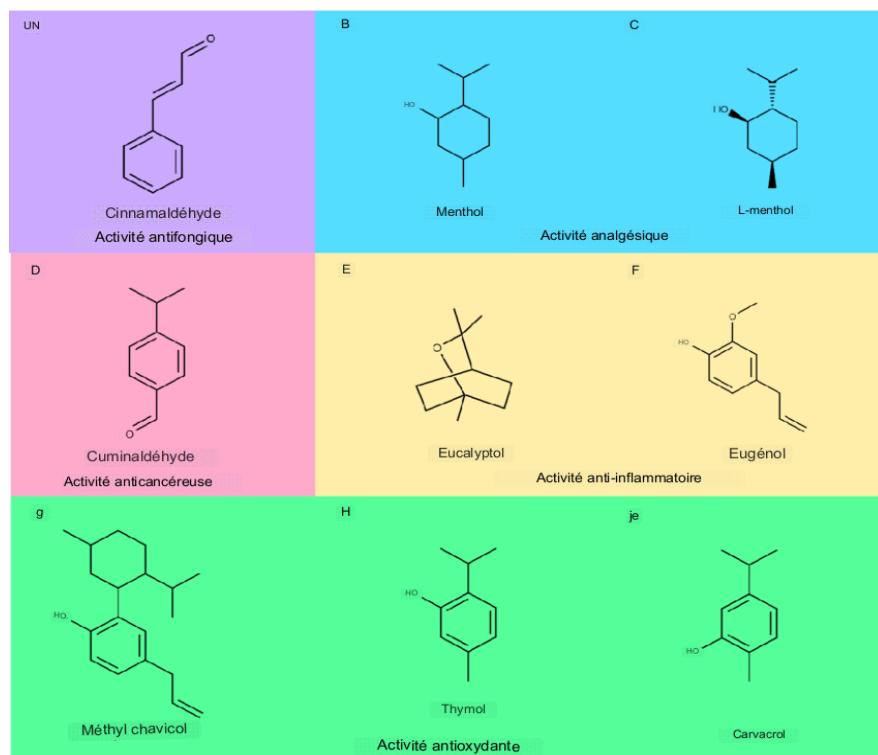
Les composés phyto-composés, non seulement se distinguent par leur diversité, mais aussi par leur rôle complexe et interactif dans la santé humaine. La présence de ces composés varie considérablement d'une plante à l'autre, et même au sein d'une même plante, en fonction des conditions de croissance et de la maturation (Liu, 2004). Les alcaloïdes, une autre classe de phyto-composés, sont connus pour leurs effets physiologiques puissants, notamment leur action comme stimulants du système nerveux central (Berkov et al., 2012). Les phyto-composés ont également montré une capacité à moduler l'activité enzymatique, influencer la production d'hormones, stimuler le système immunitaire, et même interférer avec la réPLICATION de l'ADN, ce qui pourrait avoir des implications pour la prévention du cancer (Liu, 2004; Mehta et al., 2004).

Les phyto-composés, bien que abondants dans notre alimentation, restent largement inexplorés en termes de leur potentiel thérapeutique. Les mécanismes précis par lesquels ils exercent leurs effets bénéfiques ne sont pas encore entièrement compris, en raison de leur diversité chimique et de leur interaction complexe avec les processus biologiques (Liu, 2004). Cependant, des études récentes suggèrent que les phyto-composés pourraient agir en synergie, les effets bénéfiques d'un composé étant améliorés par la présence d'autres composés (Liu, 2003). Cette synergie pourrait expliquer pourquoi les fruits et légumes entiers semblent avoir des effets protecteurs plus puissants que leurs composés individuels pris isolément (Liu, 2003). Les phyto-composés offrent donc un potentiel considérable pour le développement de nouvelles stratégies de prévention et de traitement des maladies, mais des recherches supplémentaires sont nécessaires pour exploiter pleinement ce potentiel (Higdon et al., 2007; Mehta et al., 2010).

## 1.2 ACTIVITÉS BIOLOGIQUES

L'activité biologique des phyto-composés est multifacette et complexe, englobant plusieurs aspects de la santé humaine. Leur rôle principal est souvent associé à leur capacité à agir comme antioxydants, qui neutralisent les radicaux libres qui peuvent causer des dommages oxydatifs aux cellules et contribuer à la progression de diverses maladies, y compris les maladies cardiovasculaires et le cancer (Pandey et Rizvi, 2009). Par exemple, les polyphénols et les caroténoïdes sont deux classes de phyto-composés connues pour leurs puissantes propriétés antioxydantes (Krinsky et Johnson, 2005; Manach et al., 2004).

En outre, de nombreux phyto-composés ont montré des activités anti-inflammatoires, capables de moduler la réponse immunitaire et de réduire l'inflammation, ce qui peut avoir des implications significatives pour la prévention et le traitement des maladies inflammatoires chroniques (González-Castejón et Rodriguez-Casado, 2011). Par exemple, les isothiocyanates, qui sont des produits de dégradation des glucosinolates trouvés dans les légumes crucifères, ont été associés à une réduction de l'inflammation et à une protection contre certaines formes de cancer (Higdon et al., 2007). Les phyto-composés peuvent également influencer le microbiote intestinal, avec des effets potentiellement bénéfiques sur la santé digestive et métabolique (Cardona et al., 2013).



**FIGURE 38. EXEMPLE DE COMPOSÉS NATURELS D'HUILES ESSENTIELLES ET LEURS ACTIVITÉS BIOLOGIQUES RAPPORTÉES AVEC DES APPLICATIONS PHARMACEUTIQUES POTENTIELLES. (NASCIMENTO, L ET AL. 2020).**

### **1.3 ACTIVITÉ ANTIBACTÉRIENNE**

Les composés provenant des plantes sont reconnus pour leur vaste gamme d'activités bénéfiques, notamment leur activité antibactérienne. De nombreux extraits de plantes ont démontré des propriétés antibactériennes contre diverses souches bactériennes pathogènes. Par exemple, l'extrait d'ail (*Allium sativum*) a montré une activité antibactérienne contre des pathogènes courants tels que *Staphylococcus aureus* et *Escherichia coli* (Bayan et al., 2014). De même, l'extrait de thé vert (*Camellia sinensis*) a été étudié pour son activité antibactérienne contre des bactéries pathogènes telles que *Helicobacter pylori*, impliquée dans les ulcères gastriques (Jalil et al., 2015). Un autre exemple est l'extrait de curcuma (*Curcuma longa*), qui a démontré une activité inhibitrice contre diverses souches de bactéries, y compris *Salmonella* et *Staphylococcus* (Bhawana et al., 2011). Ces exemples illustrent l'efficacité des composés d'origine végétale dans leur capacité à lutter contre les infections bactériennes.

### **1.4 ACTIVITÉ ANTI CANDIDOSE**

Les petits composés provenant des plantes ont également montré une activité anti-candidose prometteuse, ciblant spécifiquement les infections causées par les espèces de *Candida*. Par exemple, l'huile essentielle de thym (*Thymus vulgaris*) a démontré une activité anti-candidose en inhibant la croissance de *Candida albicans*, en réduisant la viabilité cellulaire et en perturbant la morphologie des cellules fongiques (Chaieb et al., 2007). De plus, le composé naturel appelé coumarine, présent dans certaines plantes, a montré une activité inhibitrice contre *Candida albicans* en interférant avec la croissance et la formation de biofilms fongiques (Nguefack et al., 2004). Ces résultats soulignent le potentiel des composés d'origine végétale dans la lutte contre les infections à *Candida* et leur utilisation potentielle en tant qu'agents antifongiques.

### **1.5 ACTIVITÉ ANTIFONGIQUE**

Les composés provenant des plantes ont également démontré une activité antifongique efficace contre diverses infections fongiques, y compris les infections à *Candida*. Par exemple, l'huile essentielle d'origan (*Origanum vulgare*) a montré une forte activité antifongique contre les espèces de *Candida*, notamment *Candida albicans*, *Candida glabrata* et *Candida krusei* (Pina-Vaz et al., 2004). De plus, le composé naturel safranal, présent dans le safran, a été étudié pour son activité antifongique contre *Candida albicans*. Des études ont montré que le safranal inhibe la croissance de *Candida albicans* en altérant la morphologie et la viabilité des cellules fongiques (Yazdanparast et al., 2008). Ces résultats soulignent le potentiel des composés d'origine végétale pour lutter contre les infections fongiques, en particulier celles causées par les espèces de *Candida*.

### **1.6 ACTIVITÉ ANTI INFLAMMATOIRE**

Les composés issus des plantes ont montré des propriétés anti-inflammatoires intéressantes, contribuant à la modulation de la réponse inflammatoire dans l'organisme. Par exemple, la

curcumine, un composé actif présent dans le curcuma (*Curcuma longa*), a été étudiée pour son activité anti-inflammatoire. Des études *in vitro* et *in vivo* ont démontré que la curcumine inhibe les voies de signalisation pro-inflammatoires, réduit la production de cytokines inflammatoires et régule l'expression de gènes liés à l'inflammation (Gupta et al., 2013). De plus, le gingembre (*Zingiber officinale*) est un autre exemple de plante connue pour son activité anti-inflammatoire. Les composés bioactifs du gingembre, tels que les gingérols et les shogaols, ont montré des effets inhibiteurs sur les médiateurs inflammatoires tels que les prostaglandines et les cytokines inflammatoires (Grzanna et al., 2005). Ces exemples mettent en évidence le potentiel des composés issus des plantes dans la modulation de l'inflammation et leur utilisation potentielle en tant qu'agents anti-inflammatoires naturels.

#### **1.7 ACTIVITÉ ANTIDIABÉTIQUE**

Les composés issus des plantes ont montré un potentiel intéressant dans le domaine de l'activité antidiabétique, contribuant à la régulation de la glycémie et à la prévention du diabète. Par exemple, la berbérine, un alcaloïde naturel présent dans diverses plantes médicinales, a démontré des effets antidiabétiques significatifs. Des études ont montré que la berbérine peut améliorer la sensibilité à l'insuline, réguler le métabolisme des glucides et réduire la glycémie chez les patients atteints de diabète de type 2 (Yin et al., 2008). De plus, le curcuma (*Curcuma longa*) a également été étudié pour son activité antidiabétique. La curcumine, un composé actif présent dans le curcuma, a montré des propriétés hypoglycémiantes en réduisant la résistance à l'insuline et en régulant le métabolisme des glucides et des lipides (Chuengsamarn et al., 2014). Ces exemples mettent en évidence le potentiel des composés issus des plantes dans la gestion du diabète et leur utilisation potentielle en tant qu'agents antidiabétiques.

#### **1.8 ACTIVITÉ ANTI-OXYDANTE**

Les composés issus des plantes sont connus pour leur activité antioxydante, aidant à neutraliser les radicaux libres et à prévenir les dommages oxydatifs dans le corps. Par exemple, les polyphénols, une classe de composés présents dans de nombreuses plantes, ont été largement étudiés pour leur puissante activité antioxydante. Des études ont montré que les polyphénols présents dans des aliments tels que les baies, le thé vert, le cacao et les légumes à feuilles vertes ont la capacité de piéger les radicaux libres et de protéger les cellules contre les dommages oxydatifs (Pandey et Rizvi, 2009).

De plus, la vitamine C, présente dans divers fruits et légumes, est un antioxydant bien connu qui agit en neutralisant les radicaux libres et en régénérant d'autres antioxydants dans le corps (Carr et Frei, 1999). Ces exemples soulignent l'importance des composés issus des plantes dans la promotion de la santé en fournissant une activité antioxydante protectrice contre le stress oxydatif.

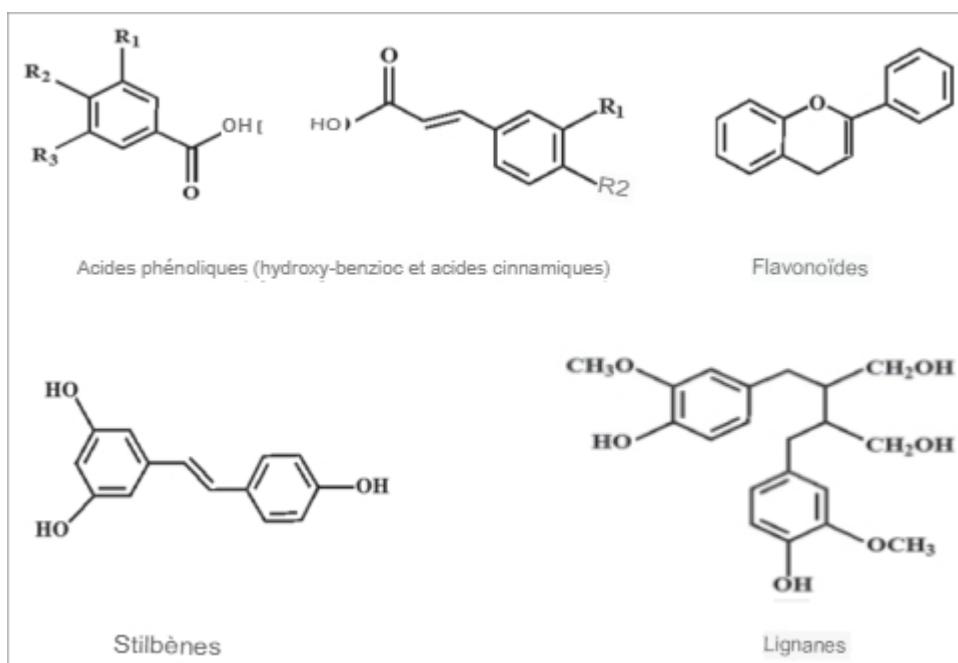


FIGURE 39. STRUCTURES CHIMIQUES DES QUATRE CLASSES DE POLYPHÉNOLES; ACIDES PHÉNOLIQUES, FLAVONOÏDES, STILBÈNES ET LIGNANES (PANDEY ET RIZVI, 2009).

### 1.9 ACTIVITÉ ANTI CARDIOVASCULAIRE

Les composés issus des plantes ont démontré des propriétés bénéfiques dans le domaine de l'activité anti cardiovasculaire, contribuant ainsi à la santé du système cardiovasculaire. Par exemple, les flavonoïdes, des composés naturels présents dans de nombreux fruits et légumes, ont été largement étudiés pour leurs effets cardioprotecteurs. Des études épidémiologiques ont montré une association inverse entre la consommation de flavonoïdes et le risque de maladies cardiovasculaires, tels que les accidents vasculaires cérébraux et les maladies coronariennes (Hollman et al., 2011). De plus, l'ail (*Allium sativum*) a également été étudié pour son activité cardioprotectrice. L'ail contient des composés soufrés, tels que l'allicine et le disulfure de diallyle, qui ont montré des effets bénéfiques sur la réduction de la pression artérielle, la diminution du taux de cholestérol LDL et l'amélioration de la fonction endothéliale (Ried et al., 2013). Ces exemples soulignent le potentiel des composés issus des plantes dans la promotion de la santé cardiovasculaire et leur utilisation potentielle dans la prévention des maladies cardiovasculaires.

## 2. Phytothérapies

### 2.1 PHYTOTHÉRAPIE DE DT2

L'importance croissante de la phytothérapie dans le traitement du diabète de type 2 est fondée sur l'efficacité potentiellement synergique de nombreux phyto-composés. Des études ont montré que ces composés peuvent interagir avec plusieurs voies métaboliques impliquées dans la pathogenèse du diabète, offrant une approche plus holistique pour traiter cette maladie complexe (Bnouham et al., 2006). Par exemple, certaines plantes médicinales, comme le ginseng et la berbérine, ont

démontré des effets hypoglycémiants, en partie grâce à la présence de phyto-composés qui peuvent augmenter la sensibilité à l'insuline et améliorer le métabolisme du glucose (Xie et al., 2011; Vuksan et al., 2000). En outre, de nombreux phyto-composés ont montré des activités anti-inflammatoires et antioxydantes, qui peuvent être particulièrement bénéfiques dans le contexte du diabète de type 2. Le stress oxydatif et l'inflammation sont tous deux impliqués dans la progression de cette maladie, et l'atténuation de ces processus pourrait aider à prévenir ou à retarder l'apparition des complications liées au diabète (Maritim et al., 2003; Higdon et al., 2007). Par exemple, la curcumine, un polyphénol trouvé dans le curcuma, a été étudiée pour son potentiel à réduire l'inflammation et le stress oxydatif chez les patients diabétiques (Chuengsamarn et al., 2012). Cependant, malgré l'abondance de recherches précliniques et cliniques sur l'efficacité des phyto-composés dans le traitement du diabète de type 2, il est important de noter que ces composés ne sont pas destinés à remplacer, mais plutôt à compléter, les traitements traditionnels. De plus, la variabilité interindividuelle dans la réponse aux phyto-composés, ainsi que la nécessité de dosages précis pour éviter les effets indésirables, soulignent l'importance d'une approche personnalisée en phytothérapie (Ekor, 2014; Williamson, 2013).

## 2.2 PHYTOTHÉRAPIE D'HYPERCHOLESTÉROLÉMIE

L'hypercholestérolémie, un facteur de risque majeur pour les maladies cardiovasculaires, est un autre domaine où la phytothérapie a démontré un potentiel prometteur. De nombreuses plantes médicinales et leurs phyto-composés ont été étudiées pour leur capacité à moduler le métabolisme du cholestérol et à réduire les niveaux de LDL. Par exemple, le psyllium, une source de fibres alimentaires solubles, a été démontré pour réduire les taux de LDL lorsqu'il est consommé en conjonction avec un régime alimentaire (Anderson et al., 2000). De même, la berbérine, un alcaloïde présent dans plusieurs plantes médicinales, a montré des effets hypcholestérolémiants dans diverses études (Kong et al., 2004). De plus, de nombreux phyto-composés ont démontré une activité antioxydante, qui peut jouer un rôle crucial dans la prévention de l'oxydation du LDL, un processus clé dans la pathogenèse de l'athérosclérose (Lobo et al., 2010). Par exemple, les polyphénols trouvés dans le thé vert ont été étudiés pour leur capacité à inhiber l'oxydation du LDL, ce qui pourrait potentiellement réduire le risque d'athérosclérose et de maladie cardiovasculaire (Legeay et al., 2015). Cependant, bien que la phytothérapie offre un potentiel considérable pour le traitement de l'hypercholestérolémie, il est important de noter que ces approches devraient être utilisées en complément des stratégies de style de vie et des médicaments conventionnels, et non pas comme un substitut. De plus, comme pour le diabète, la réponse aux phyto-composés peut varier considérablement d'une personne à l'autre, soulignant l'importance d'une approche personnalisée dans la phytothérapie (Ekor, 2014 ; Williamson, 2013).

### 2.3 Les plantes aromatiques et médicinales du Maroc

Le Maroc est un pays reconnu pour sa riche biodiversité de plantes aromatiques et médicinales (PAMs), qui jouent un rôle fondamental dans la santé publique et la médecine traditionnelle marocaine. Le Maroc possède une riche biodiversité, comprenant plus de 4200 espèces de plantes. Parmi celles-ci, environ 800 sont utilisées pour leurs propriétés médicinales selon l'agence nationale des plantes aromatiques et médicinales du Maroc (ANPAM). Le pays est particulièrement réputé pour des plantes comme le thym (*Thymus satureioides*), l'origan (*Origanum compactum*), la menthe (*Mentha suaveolens*) et la lavande (*Lavandula stoechas*), toutes reconnues pour leurs attributs aromatiques et médicinaux. En outre, l'arganier (*Argania spinosa*), plante endémique du Maroc, est largement utilisé pour ses propriétés médicinales et cosmétiques. En effet, l'huile d'argan, extraite des graines de l'arganier, est connue pour ses nombreux effets thérapeutiques (Charrouf & Guillaume, 2008). Sur le plan économique, la demande pour les plantes médicinales et les produits dérivés naturels a connu une croissance significative (Cunningham et al., 2020). Cependant, la surexploitation de ces ressources naturelles précieuses et la dégradation de leur habitat représentent des menaces importantes pour la biodiversité marocaine. Par exemple, l'armoise, une plante couramment utilisée dans la médecine traditionnelle marocaine, a montré des propriétés hypoglycémiantes et hypolipémiantes, suggérant un potentiel pour le traitement du diabète et de l'hypercholestérolémie (Eddouks et al., 2002 ; Maghrani et al., 2004). Outre les plantes traditionnellement utilisées pour leurs propriétés médicinales, le Maroc est également célèbre pour ses plantes aromatiques, comme le thym, la menthe et la sauge, qui sont largement utilisées pour leur saveur et leur parfum, mais qui possèdent également des propriétés médicinales potentielles. Par exemple, le thym marocain a démontré des propriétés antioxydantes et antimicrobiennes, qui peuvent être bénéfiques pour la santé cardiovasculaire et la prévention des infections (Fadli et al., 2012 ; Zekkour et al., 2015). Néanmoins, malgré l'abondance et l'utilisation généralisée des PAMs au Maroc, il reste encore beaucoup à découvrir sur leurs propriétés médicinales et leur potentiel thérapeutique. La recherche future devrait se concentrer sur la caractérisation plus détaillée des phyto-composés présents dans ces plantes, ainsi que sur l'évaluation de leurs effets sur la santé humaine dans des essais cliniques bien conçus (Alaoui et al., 2017; El-Hilaly et al., 2003).

### 3. Bases de données des phyto composés internationales

Dans le cadre de l'étude des plantes médicinales, comme celles du Maroc, l'accès à des informations précises sur les composés phyto-composés est crucial. Il existe plusieurs bases de données internationales qui se consacrent à la collecte et à la catalogation de ces informations. La base de données Dr. Duke's Phytochemical and Ethnobotanical Databases, par exemple, offre une compilation complète des composés bioactifs identifiés dans différentes espèces de plantes (Duke, 1992). De plus, la base de données PhytoHub offre des informations sur la composition en phytonutriments de

nombreux aliments (Rothwell et al., 2013). Ces ressources sont extrêmement utiles pour les chercheurs qui étudient l'activité biologique des phyto-composés. Par exemple, elles peuvent être utilisées pour identifier les composés bioactifs présents dans une plante donnée, ou pour comparer la composition des phyto-composés de différentes plantes. De plus, elles peuvent aider à identifier les plantes qui contiennent des composés ayant des activités biologiques spécifiques, comme les activités antioxydantes, anti-inflammatoires ou antimicrobiennes (Ulrich-Merzenich et al., 2009; Liu, 2013). Cependant, alors que ces bases de données offrent des informations précieuses, il est important de noter qu'elles ne sont pas exhaustives et que de nombreuses plantes et composés restent à découvrir et à cataloguer (Atanasov et al., 2015).

## **Objectifs de la thèse**

- 1- Analyser l'impact des variants génétiques du gène RETN sur le diabète de type 2 (DT2) et les maladies métaboliques à l'aide de la modélisation moléculaire et les outils de bioinformatique.
- 2- Évaluer les effets des phyto-composés sur le PPAR gamma, impliqué dans le DT2, par le biais du criblage virtuel et de simulations de dynamique moléculaire.
- 3- Créer la première base de données marocaine des composés issus des plantes marocaines, nommée MPDB (Moroccan Natural Phytochemicals Database).
- 4- Explorer le potentiel thérapeutique des composés présents dans la MPDB pour identifier un composé capable de se lier à la PCSK9, en vue du développement d'un nouvel inhibiteur d'origine naturelle pour un futur traitement des maladies cardiovasculaires.

## II. TRAVAUX DE RECHERCHE



## PARTIE 1. MODELISATION MOLECULAIRE DE LA RÉSISTINE HUMAINE IMPLIQUÉ DANS LE DIABÈTE DE TYPE II

## 1. Introduction

La résistine est une adipokine pro-inflammatoire qui appartient aux protéines du domaine C-terminal riches en cystéine appelées molécules de type résistine (RELMs) et est principalement sécrétée par les adipocytes chez les rongeurs et les macrophages chez l'homme (Steppan C. et al., 2001; Patel S., et al. 2004). Le gène codant pour la résistine (RETN) est situé sur le chromosome 19p13.2. Il a été démontré que la résistine est liée à plusieurs troubles inflammatoires, notamment l'obésité, le diabète de type 2, les maladies cardiovasculaires et l'asthme (Engert et al., 2002; Demirci et al., 2017). Cette protéine a des effets qui antagonisent l'action de l'insuline. Certaines études ont montré que la résistine affecte le transport du glucose et provoque la dégradation de l'insuline-stimulé du substrat du récepteur de l'insuline-1 (IRS-1) conduisant à l'induction de la résistance à l'insuline (Steppan C. et al., 2001; Rui L., et al., 2002). Les niveaux de résistine circulants ont été signalés comme étant significativement augmentés chez les souris obèses génétiquement et induites par l'alimentation et ont diminué avec l'administration du médicament antidiabétique Rosiglitazone (Steppan C. et al., 2001). De plus, une étude cas-témoin avec des patients atteints de diabète sucré de type 1 a montré que la combinaison d'insuline et de Rosiglitazone a diminué significativement les niveaux de résistine et de leptine (Guclu M., et al., 2015). La compréhension de la variation génomique est l'un des principaux défis de l'actuel domaine de recherche en génomique en raison du nombre énorme de variations génétiques sur le génome humain. Les polymorphismes mononucléotidiques (SNPs) représentent les variations génétiques les plus abondantes dans l'ensemble du génome humain, allant de 3 à 5 millions chez chaque individu (Dakal T. C., et al., 2017). La plupart des SNPs sont neutres, mais certains SNPs contribuent directement à la prédisposition aux maladies en modifiant la fonction des gènes ou en tant que marqueurs génétiques pour trouver les mutations causant des maladies à proximité par des études d'association et basées sur la famille (Wang X., et al., 2004). Les scientifiques pensent que ces variantes peuvent également influencer la réponse à certains médicaments (Giacomini K. M., et al., 2007). Les SNPs qui modifient les acides aminés codés sont appelés polymorphismes mononucléotidiques non synonymes (nsSNPs). Les nsSNPs, qui représentent environ la moitié de toutes les modifications génétiques liées aux maladies humaines, peuvent influencer la structure et/ou la fonction de la protéine résultant avec des effets neutres ou délétères (Cargill M., et al., 1999; Yue P., et al., 2016). De plus, Les régions non traduites 5' et 3' (UTR) sont d'un intérêt majeur car elles peuvent affecter l'expression des gènes, les activités post-transcriptionnelles et de traduction, et donc être d'une pertinence fonctionnelle (Mayr C. et al.; Chatterjee S. et al., 2009), l'étude de l'ADN non codant est également importante car elle contient la plupart des SNPs signalés dans le génome humain. Les polymorphismes dans les régions non codantes sont d'une grande importance car ils peuvent affecter l'expression des gènes, les activités post-transcriptionnelles et de traduction. Les variantes génétiques dans RETN ont été étudiées, décrivant l'association de ce gène avec les niveaux de résistine circulants. Beckers et al. ont identifié

la première mutation faux-sens C78S dans la résistine chez un patient obèse morbide et sa mère obèse, encourageant l'étude des mutations dans la région codante pour élucider leur implication dans la pathogenèse (Beckers S.,et al., 2011). Il a été estimé que les facteurs génétiques peuvent expliquer jusqu'à 70% de la variation des niveaux de résistine circulants (Menzaghi C.,et al., 2006). Cependant, les associations entre les SNPs de RETN et les variables anthropométriques et les modifications liées à l'obésité ont montré des résultats incohérents (Engert J. C.,et al., 2002; Norata G. D., et al., 2007; Mattevi V. S.,et al., 2004).

Se basant sur l'importance du gène RETN dans plusieurs maladies inflammatoires, en particulier les anomalies métaboliques, nous avons mené une analyse informatique en utilisant des prédicteurs de mutation comme SIFT, PolyPhen, PANTHER et PhD-SNP, Predict SNP, des outils de conservation et de stabilité, et enfin une analyse structurelle afin d'identifier les SNPs les plus délétères fonctionnellement dans les régions codantes et non traduites de ce gène.

## 2. Matériel et Méthodes

### 2.1 COLLECTE DE DONNÉES

Les informations sur les SNPs du gène RETN ont été collectées à partir de dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>). La séquence d'acides aminés de la résistine (accès NCBI: NP\_001180303) a été récupérée à partir de la base de données de protéines du NCBI (<http://www.ncbi.nlm.nih.gov/protein>). La structure théorique de la résistine (PDB ID: 1LV6) a été abandonnée car elle n'est pas en accord avec la structure cristalline disponible pour la résistine de souris actuellement.

### 2.2 PRÉDICTION DES NSSNPs DÉLÉTÈRES

PredictSNP1.0 (<http://loschmidt.chemi.muni.cz/predictsnp1/>) (Bendl J.,et al., 2014)[24] a été utilisé comme prédicteur de l'effet des mutations sur la fonction des protéines. Cette ressource est un classificateur de consensus qui permet d'accéder aux huit meilleurs outils de prédition: SIFT, Polyphen-1, Polyphen-2, MAPP, PhD-SNP, SNAP, PANTHER, PredictSNP, et nsSNPAnalyzer. SIFT (Sorting Intolerant from Tolerant) prédit si une substitution d'acide aminé affecte la fonction de la protéine en se basant sur l'homologie des séquences et les propriétés physiques des acides aminés (Ng P. C.,et al., 2003)[25]. SIFT prend une séquence de requête et utilise les informations d'alignement multiple pour prédire les substitutions tolérées et délétères pour chaque position de la séquence de requête.

PolyPhen-1 utilise un ensemble d'experts de règles empiriques pour prédire l'impact possible des substitutions d'acides aminés, PolyPhen-2 (Polymorphism Phenotyping v2) prédit l'influence potentielle d'une substitution d'acide aminé sur la structure et la fonction d'une protéine humaine en utilisant l'alignement de séquences multiples pour aligner les séquences en utilisant des informations structurales, MAPP (Multivariate Analysis of Protein Polymorphism) analyse la variation

physicochimique présente dans chaque colonne d'un alignement de séquences de protéines et prédit l'impact des substitutions d'acides aminés sur la fonction de la protéine (Stone E. A.,et al., 2005)[26].

PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms) est un prédicteur basé sur des machines à vecteurs de support (SVM) utilisé pour classer les nsSNPs en mutations causant des maladies génétiques humaines ou en mutations bénignes (Capriotti E.,et al., 2006)[27]. SNAP (screening for non-acceptable polymorphisms) est une méthode basée sur des réseaux de neurones pour prédire les effets fonctionnels des SNPs non-synonymes en utilisant des informations sur les protéines dérivées *in silico* (Bromberg Y.,et al., 2007)[28].

PANTHER (Protein Analysis Through Evolutionary Relationships) estime la probabilité qu'un nsSNP particulier cause un effet fonctionnel sur la protéine en utilisant la préservation évolutionnaire position-spécifique (Tang et Thomas 2016), nsSNPAnalyzer utilise une méthode d'apprentissage automatique appelée Random Forest pour prédire si le nsSNP a un effet phénotypique (Tang H.,et al 2016)[29] basé sur l'alignement de séquences multiples et l'information de la structure 3D. Enfin, PredictSNP1.0 affiche les scores de confiance générés par chaque outil et une prédiction de consensus sous forme de pourcentages en utilisant leurs valeurs d'exactitude observées pour faciliter les comparaisons (Bendl J.,et al., 2014).

### 2.3 ANALYSE DE CONSERVATION

Le serveur web ConSurf (<http://consurf.tau.ac.il/>) a été utilisé. Cet algorithme basé sur le web prédit les régions fonctionnelles cruciales d'une protéine en estimant le degré de conservation des acides aminés basé sur l'homologie. Les grades vont de 1 à 9 estimant l'étendue de la conservation de l'acide aminé tout au long de l'évolution. Par conséquent, le grade 9 représente le résidu le plus fortement conservé, et les nombres descendent jusqu'à 1, qui représente la région la moins conservée. Cet outil analyse la conservation au niveau des nucléotides et des acides aminés.

### 2.4 PRÉDICTION DES SITES DE MODIFICATION POST-TRADUCTIONNELLE

ModPred (<http://www.modpred.org/>) est un serveur web utilisé pour prédire les sites de modification post-traductionnelle (PTM) sur des séquences uniques et multiples, le serveur est constitué d'un ensemble de modèles de régression logistique bootstrap pour chaque type de PTM, formés à partir de 126 036 sites PTM non redondants vérifiés expérimentalement, à partir de la littérature et de la base de données (Bao L., et al., 2005)[30]. Les résultats sont donnés tels que résidu, modification, score, confiance et remarques. Dans cette étude, seuls les PTMs de confiance moyenne et élevée ont été pris en considération.

### 2.5 PRÉDICTION DU CHANGEMENT DE STABILITÉ DES PROTÉINES

Le changement de la stabilité des protéines dû aux nsSNPs a été prédit en utilisant I-Mutant2.0 (<http://folding.biifold.org/cgi-bin/i-mutant2.0>) qui est un outil web basé sur des machines à vecteurs

de support (SVM) utilisé pour la prédition automatique des changements de la stabilité des protéines dû au SNP. Il fournit la valeur prévue du changement d'énergie libre (DDG) et le signe de la prédition comme Augmentation ou Diminution. La valeur DDG est calculée à partir de la valeur d'énergie libre de Gibbs de dépliage de la protéine mutée moins la valeur d'énergie libre de Gibbs de dépliage du type sauvage en Kcal/mol. DDG > 0 signifie que la stabilité de la protéine a augmenté et DDG < 0 signifie que la stabilité de la protéine a diminué (Pejaver V.,et al., 2014).

La stabilité a également été vérifiée par l'outil Mupro (<http://mupro.proteomics.ics.uci.edu/>). Ce serveur est basé sur deux méthodes d'apprentissage automatique: les machines à vecteurs de support et les réseaux de neurones. Les deux ont été formés sur un large ensemble de données de mutations, montrent une performance comparable, mais ont des spécificités légèrement différentes. une précision supérieure à 84% et calculent un score entre -1 et 1 comme la confiance de la prédition. Le score de confiance <0 indique que la mutation diminue la stabilité de la protéine, tandis qu'un score de confiance >0 signifie que la mutation augmente la stabilité de la protéine (Capriotti E.,et al., 2005).

## 2.6 BALAYAGE DES SNPs UTR SUR LE SITE UTR

Les régions non traduites 5' et 3' (UTRs) ont des rôles cruciaux dans la dégradation, la traduction et la localisation des ARNm ainsi que la régulation de l'interaction protéine-protéine. Nous avons utilisé UTRscan <http://itbtools.ba.itb.cnr.it/utrscan> pour prédire les SNPs fonctionnels dans les UTRs 5' et 3'. L'outil UTRscan permet à l'utilisateur de rechercher des séquences soumises par l'utilisateur pour l'un des motifs présents dans UTRsite. UTRsite tire ses données de UTRdb, une base de données curée qui met à jour les ensembles de données UTR par le biais de l'extraction de données primaires et de la validation expérimentale (Cheng J.,et al., 2006; Chatterjee S.,et al., 2009)[7,33]. Pour ce faire, les données primaires au format FASTA ont été soumises et les résultats étaient sous la forme de noms de signal et de leurs positions dans le transcript.

## 2.7 ANALYSE STRUCTURALE

Modélisation de la structure native et des mutants : Le transcrit avec la séquence de référence NP\_001180303.1 a été utilisé pour l'analyse computationnelle. Nous avons sélectionné la structure cristalline de rayons X de la resistine de *Mus musculus* de la Banque de données sur les protéines (PDB) avec le code PDB 1RGX comme modèle pour générer une resistine humaine par modélisation par homologie en utilisant la plateforme Swiss Model (<https://swissmodel.expasy.org>) avec un QMEAN de -1.83 et une identité de séquence de 55.56% (Figure 6).

nnUCSF Chimera, qui est un programme hautement extensible développé par la Resource for Biocomputing, Visualization, and Informatics à l'Université de Californie, San Francisco pour la visualisation interactive et l'analyse des structures moléculaires et des données connexes, a été utilisé

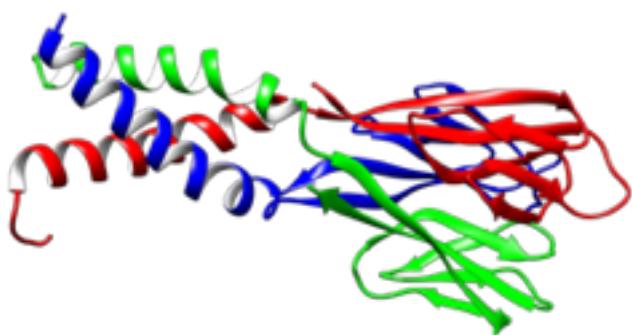
pour confirmer les positions correspondantes des mutations et pour construire les 15 modèles mutants (Pesole G.,et al., 2002).

L'énergie de minimisation pour les structures de type sauvage et mutantes a été réalisée par le serveur NOMAD-Ref basé sur Gromacs comme champ de force par défaut, nous avons utilisé la méthode du gradient conjugué pour l'optimisation des structures 3D (Pettersen E. F.,et al., 20).

Prédiction de RMSD et du total des liaisons hydrogène : UCSF Chimera, a servi à nouveau à vérifier l'analyse de déviation RMS en superposant à la fois les structures natives et mutantes et enfin, il a servi à calculer les valeurs totales de h-bonds pour chaque structure.

Analyse d'interaction : COCOMAPS (bioCOnplexes COnact MAPS) est une application web pour analyser et visualiser efficacement l'interface dans les complexes biologiques de protéine-protéine en utilisant des cartes de contact intermoléculaires. Le fichier d'entrée était le modèle d'homologie de la résistine en format PDB. Dans notre étude, nous utilisons COCOMAPS pour analyser l'interaction entre les trois monomères de la protéine de résistine (Lindahl E.,et al., 2006). Pour ce faire, nous avons téléchargé le fichier PDB du trimère de résistine (A, B et C comme identifiants de chaîne pour chaque monomère), nous avons ensuite comparé les interfaces d'interaction entre les deux chaînes A et B considérées comme la Molécule 1 interagissant avec la troisième chaîne C considérée comme la Molécule 2 (les interactions incluent les résidus de la chaîne A et de la chaîne B ensemble interagissant avec la chaîne C).

Prédiction des interactions protéine-protéine : STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, disponible à <http://string-db.org>) est une base de données d'interactions protéiques connues et prédites, couvrant actuellement 9643763 protéines de 2031 organismes. Cette base de données fournit une évaluation critique et une intégration des interactions protéine-protéine, y compris les associations directes (physiques) et indirectes (fonctionnelles) (Vangone A.,et al., 2011)[37].



**FIGURE 40.** LE MODÈLE TRIDIMENSIONNEL DE LA RÉSISTINE HUMAINE GÉNÉRÉ EN UTILISANT LA MODÉLISATION PAR HOMOLOGIE PAR LE SERVEUR WEB SWISS MODEL. LE MODÈLE EST UN TRIMÈRE AVEC TROIS CHAÎNES A (ROUGE), B (VERTE) ET C (BLEU).

### 3. Résultats

#### 3.1 SNP DATASET

Le gène RETN étudié dans ce travail a été extrait début octobre 2018 de la base de données dbSNP ([www.ncbi.nlm.nih.gov/snp/?term=RETN](http://www.ncbi.nlm.nih.gov/snp/?term=RETN)). Il contenait un total de 1075 SNPs, dont 78 étaient des nsSNPs, 35 étaient des SNPs synonymes codants, 339 étaient dans la région non codante, qui comprend 18 SNPs dans la région 5'UTR, 35 SNPs étaient dans la région 3'UTR et 287 étaient dans la région intronique.

#### 3.2 PRÉDICTION DES NSSNPs DÉLÉTÈRES

Un total de 78 nsSNPs a été sélectionné pour notre enquête. La collection de nsSNPs a été analysée avec divers outils de prédition in silico pour mesurer leurs effets sur la pathogénicité et pour trouver les SNPs associés à la maladie. Tous les SNPs codants non synonymes qui ont été obtenus à partir de la base de données SNP, ont été chargés dans PredictSNP1.0, et tous les outils intégrés disponibles ont été sélectionnés pour la prédition. PredictSNP1.0 a établi des prédictions pour les huit outils et une prédition de consensus sous forme de pourcentages d'exactitudes attendues ainsi que l'effet de la mutation sur la fonction de la protéine comme "neutre" ou "délétère" pour les substitutions d'acides aminés. 15 mutations ont été prédites comme délétères par tous les outils intégrés, à l'exception de nsSNPAnalyzer et PANTHER qui n'ont donné aucune prédition pour aucune mutation. Selon SNAP, un total de 38 nsSNPs sur 54 a été prédit comme délétère (70,37%), suivi par MAPP avec 37 nsSNPs délétères (68,51%), PolyPhen-2 avec 31 nsSNPs (57,40%), PolyPhen-1 avec 25 nsSNPs (46,29 %), SIFT avec 26 nsSNPs (48,15%) et PhD-SNP avec 18 nsSNPs (33,33%). Les nsSNPs prédis comme délétères sont répertoriés dans le tableau 1 avec l'exactitude attendue et ont été sélectionnés pour une analyse plus approfondie (tableau 1).

TABLEAU 1. RÉSULTATS ATTENDUS DE PRÉCISION DES SNP DE RETN PRÉDITS COMME DÉLÉTÈRES DANS PREDICTSNP ET OUTILS INTÉGRÉS.

nsSNPs	ID Variants	PredictSNP	PolyPhe n-1	PolyPhe n-2	SIFT	MAP P	PhD-SNP	SNAP
I32S	rs89432 927	(0.86)	(0.74)	(0.60)	(0.79)	(0.77)	(0.85)	(0.72)
C51Y	rs759129635	(0.86)	(0.74)	(0.81)	(0.79)	(0.77)	(0.81)	(0.86)
G58R	rs760260537	(0.86)	(0.74)	(0.81)	(0.79)	(0.87)	(0.73)	(0.88)
G58E	rs763916942	(0.86)	(0.74)	(0.81)	(0.79)	(0.91)	(0.58)	(0.88)
G66R	rs566843624	(0.86)	(0.74)	(0.56)	(0.42)	(0.62)	(0.81)	(0.86)
G71R	rs772946179	(0.86)	(0.74)	(0.59)	(0.79)	(0.87)	(0.57)	(0.86)
C74W	rs532089804	(0.86)	(0.74)	(0.81)	(0.79)	(0.91)	(0.77)	(0.75)
C78S	rs199834487	(0.86)	(0.74)	(0.81)	(0.79)	(0.75)	(0.67)	(0.86)
G79C	rs111331676	(0.86)	(0.74)	(0.81)	(0.79)	(0.85)	(0.73)	(0.86)
R84C	rs779335092	(0.86)	(0.74)	(0.81)	(0.79)	(0.85)	(0.73)	(0.84)
D97E	rs768223197	(0.86)	(0.59)	(0.81)	(0.79)	(0.77)	(0.58)	(0.84)
W98C	rs776285077	(0.86)	(0.74)	(0.81)	(0.79)	(0.85)	(0.81)	(0.86)
W98L	rs1035187378	(0.86)	(0.74)	(0.81)	(0.52)	(0.91)	(0.58)	(0.88)
C103G	rs991039386	(0.86)	(0.74)	(0.81)	(0.79)	(0.76)	(0.58)	(0.88)
C104Y	rs891939673	(0.86)	(0.74)	(0.81)	(0.79)	(0.77)	(0.81)	(0.88)

### 3.3 ANALYSE DE LA CONSERVATION

Les résultats de l'analyse Consurf ont montré que 13 SNPs faux-sens délétères sont situés dans des régions hautement conservées, avec des valeurs de conservation comprises entre 7 et 9, ce qui suggère que ces positions sont importantes pour l'intégrité de la résistine. Parmi ceux-ci, quatre résidus ont été prédits comme étant exposés et fonctionnels, cinq autres ont été prédits comme étant enfouis et structurels, trois résidus enfouis et un résidu exposé (tableau 2), à l'exception de la position 84 qui a été prédite comme modérément conservée et de la position 71 qui a été prédite comme résidu variable, et donc, n'ont pas été sélectionnés pour l'analyse structurale.

TABLEAU 2. DÉTAILS DES NSNP SÉLECTIONNÉS COMME DÉLÉTÈRES PARMI LES SNP RAPPORTÉS, LEUR ANALYSE DE CONSERVATION PAR CONSURF ET LA PRÉDICTION DES SITES DE POST-TRADUCTION PAR MODPRED.

Position	Score de Conservation	B/E	F / S	PTM Sites (wild type)	Variants	PTM Sites (mutants)
I32	7	b	-	-	I32S	-
C51	9	b	s	-	C51Y	Phosphorylation Amidation
G58	9	e	f	clivage protéolytique	G58R	clivage protéolytique
					G58E	clivage protéolytique
G66	7	e	-	clivage protéolytique	G66R	clivage protéolytique
G71	4	b	-	-	G71R	clivage protéolytique ADP-ribosylation

C74	9	b	s	liaison disulfure	C74W	-
C78	9	b	s	liaison disulfure	C78S	O-linked glycosylation
G79	9	e	f	-	G79C	liaison disulfure
R84	6	e	-	ADP-ribosylation clivage protéolytique	R84C	liaison disulfure
D97	9	e	f	-	D97E	liaison disulfure
W98	8	b	-	C-linked glycosylation clivage protéolytique Amidation	W98C	liaison disulfure
					W98L	-
C103	9	b	s	liaison disulfure	C103	clivage protéolytique
C104	9	b	s	liaison disulfure	C104	Amidation clivage protéolytique

**PTM** : sites de modification post-traductionnelle ; **B** : enfoui ; **E** : exposé ; **F** : fonctionnel ; **S** : structurel.

### 3.4 DES SITES DE MODIFICATION POST-TRADUCTIONNELLE

ModPred a été utilisé pour prédire les sites de modification post-traductionnelle présents dans la protéine de résistin humaine. Seuls les PTM avec une confiance élevée ou moyenne ont été discutés. Dans la protéine native, la position R84 a été prédite comme un site d'ADP-ribosylation, W98 comme un site de glycosylation C-liée ou de clivage protéolytique, C103 et C104 comme des sites de liaison disulfure. Après mutations, C51Y est apparu comme un site d'amidation avec le changement de Cys en Tyr, tandis que la position W98 est devenue un site de liaison disulfure. Pour la position C104, il a été prédit que le changement de Cys en Tyr conférait un site d'amidation avec une grande confiance. Les résultats de ModPred sont présentés dans le tableau 2.

### 3.5 L'IMPACT DES MUTATIONS DÉLÉTÈRES PRÉDITES SUR LA STABILITÉ DE LA PROTÉINE DE RÉSISTINE

Nous avons analysé les 13 substitutions faux-sens prédites comme délétères à partir des étapes précédentes avec I-Mutant2.0 et le serveur web MUpro. Les nsSNPs prédits pour diminuer la stabilité avec les deux outils ont été sélectionnés pour une analyse structurale plus approfondie (Tableau 3).

TABLEAU 3. PRÉDICTION DU CHANGEMENT DE STABILITÉ PROTÉIQUE À L'AIDE D'I-MUTANT 2 ET MUPRO.

Position	I MUTANT-2 DDGValue (kcal/mol)	MUPRO DDG value SVM

		NN (kcal/mol)		
I32S	-2.45 (Dim)	-1.490 (Diminution)	-0.80 (Diminution)	-0.96 (Diminution)
C51Y	-0.89 (Dim)	-0.234 (Diminution)	-0.20 (Diminution)	-0.69 (Diminution)
G58R	-0.15 (Diminution)	-0.95 (Diminution)	-0.22 (Diminution)	-0.73 (Diminution)
G58E	-1.43 (Diminution)	-0.913 (Diminution)	0.23 (Augmentation)	-0.67 (Diminution)
G66R	-1.74 (Diminution)	-0.84 (Diminution)	0.48 (Augmentation)	0.51 (Augmentation)
C74W	-0.76 (Diminution)	-0.69 (Diminution)	0.54 (Augmentation)	0.85 (Augmentation)
C78S	-0.63 (Diminution)	-1.513 (Diminution)	-0.09 (Diminution)	0.87 (Augmentation)
G79C	-0.89 (Diminution)	-0.842 (Diminution)	-0.84 (Diminution)	-0.38 (Diminution)
D97E	0.52 (Augmentation)	-0.77 (Diminution)	0.025(Augmentation )	-0.65 (Diminution)
W98C	-1.14 (Diminution)	-0.442 (Diminution)	-0.77 (Diminution)	-0.92 (Diminution)
W98L	-0.59 (Diminution)	0.16 (Augmentation)	-0.10 (Diminution)	0.691 (Augmentation)
C103G	-1.05 (Diminution)	-1.63 (Diminution)	-0.80 (Diminution)	0.547 (Augmentation)
C104Y	0.32 (Augmentation)	-1.03 (Diminution)	-0.36 (Diminution)	-0.80 (Diminution)

DDG : delta delta G ; SVM : Machine à vecteurs de support ; NN : Neuronal network

### 3.6 ANALYSE STRUCTURALE

#### 3.6.1 MODÉLISATION DE LA STRUCTURE DE LA RÉSISTINE HUMAINE

En utilisant la structure cristalline aux rayons X (1rgx) comme modèle, nous avons modélisé la structure 3D de la résistine humaine native en utilisant le serveur web Swiss model. La figure 7 montre le modèle généré sous la forme d'un trimère avec trois monomères (A, B et C). Ce trimère a été utilisé pour construire les 9 modèles mutants de résistine humaine.

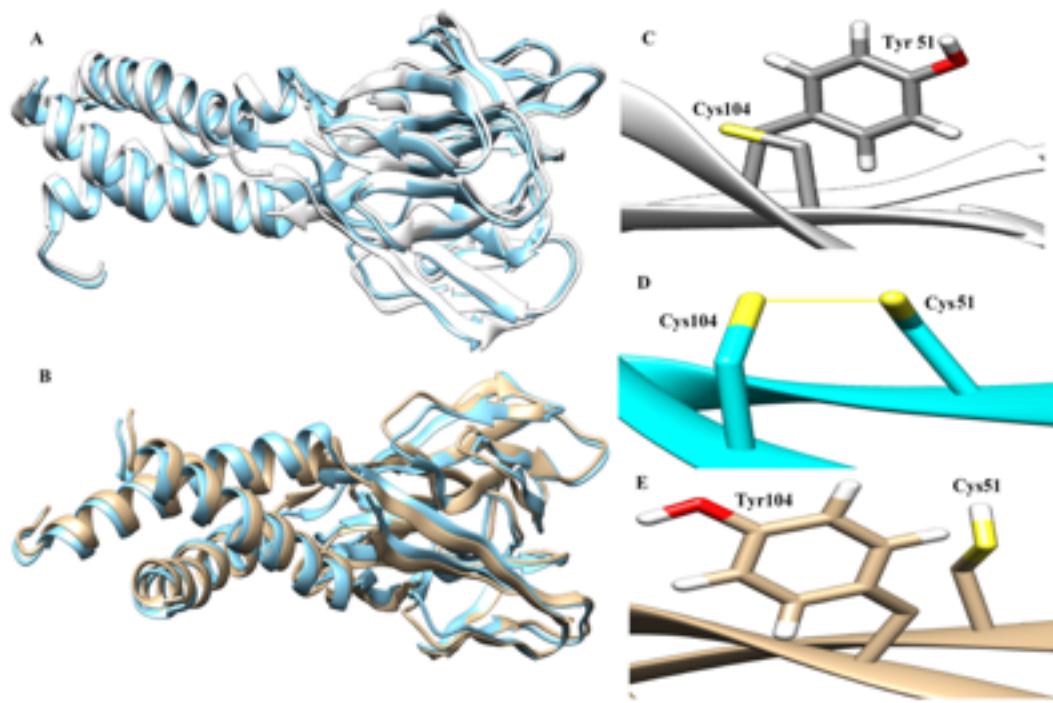


FIGURE 41. DIFFÉRENCE RMSD ET LIAISONS HYDROGÈNE TOTALES.

A. Superposition de la structure native de la résistine humaine (couleur cyan) sur la structure du mutant C51Y (couleur blanche) montrant un écart RMSD de 1.137Å

B. Superposition de la structure native de la résistine humaine (couleur cyan) sur la structure du mutant C104Y (couleur blanche) montrant un écart RMSD de 1.308Å

C. Rupture de la liaison disulfure dans le mutant résistine C51Y.

D. Liaison disulfure entre les deux résidus de cystéine 51 et 104 dans la résistine de type sauvage.

E. Rupture de la liaison disulfure dans le mutant résistine C104Y.

Les valeurs RMSD associées aux 9 mutants sont données dans le tableau 4. Lorsque la valeur RMSD augmente, l'écart entre les structures natives et mutantes sera plus grand et induira donc un changement dans l'activité protéique. Les C51Y et C104Y modifiés ont montré la plus grande RMSD, les résultats sont présentés dans la figure 7.B. De plus, le nombre total de liaisons h a été calculé pour évaluer leur contribution à la stabilité et au repliement de la protéine native. Toutes les structures mutées ont révélé un changement dans le nombre total de liaisons h par rapport à la résitine native, mais le mutant C104Y a montré une diminution remarquable, formant 254 liaisons h tandis que la structure native en formait 291. De plus, la visualisation de la structure native a montré que les résidus C51 et C104 forment une liaison disulfure entre eux (Figure 7.D), le changement de cystéine porté sur l'hélice alpha à ces positions induit la rupture du pont disulfure (Figure 7.C, 7.E) qui perturbe la structure de la protéine.

TABLEAU 4. VALEUR RMSD ET NOMBRE TOTAL DE LIAISONS HYDROGÈNE APRÈS MINIMISATION DE CHAQUE MODÈLE.

	RMSD (Å)	Total des Liaisons hydrogène
Natifs	0	291
I32S	0.023	297
C51Y	1.137	272
G58E	0.30	297
C78S	0.37	299
G58R	0.182	292
G79C	0.265	296
W98C	0.158	296
C10G	0.486	250
C10Y	1.308	254

### 3.7 ANALYSE DES INTERACTIONS

Les contacts à l'interface entre les acides aminés présents dans la protéine et les trois monomères de résistin ont été étudiés en utilisant CoCoMaps. Une variation du nombre de différents types d'interactions a été observée entre le complexe natif et les 9 mutants de résistin, les résultats sont présentés dans le tableau 5. En ce qui concerne le nombre d'interactions hydrophiles-hydrophiles, le complexe natif participait avec 262 interactions hydrophiles-hydrophiles. Les complexes mutants I32S, C51Y, G79C et C104Y, ont montré une augmentation significative du nombre d'interactions hydrophiles-hydrophiles avec respectivement 286, 266, 277 et 266 interactions, ce qui indique une réduction de l'hydrophobie de ces trimères mutants. De plus, le complexe mutant C103G a montré une augmentation significative du nombre d'interactions hydrophobes-hydrophobes, indiquant l'augmentation de l'hydrophobie de ce complexe mutant. De plus, le trimère mutant C51Y interagit avec seulement 75 résidus de la chaîne C formant le complexe trimère alors que la chaîne C dans le complexe natif interagit avec 78 résidus. Cette petite déviation peut perturber la formation du trimère de résistin.

### 3.8 PRÉDICTION DE L'EFFET DES SNPs SITUÉS DANS LA RÉGION UTR PAR LE SERVEUR UTRSCAN.

Le serveur UTRscan a été utilisé pour prédire l'effet des SNPs UTR sur le motif transcriptionnel. Six SNPs dans l'UTR 3', à savoir rs920569876, rs74176247, rs1447199134, rs943234785, rs76346269 et rs78048640 ont été prédits pour être dans les sites de polyadénylation et peuvent donc être responsables des phénotypes pathologiques. Les résultats sont présentés dans le tableau 6.

### 3.9 INTERACTIONS PROTÉINE-PROTÉINE À L'AIDE DE STRING :

La prédition des interactions protéine-protéine a indiqué que la résistin interagit avec 10 protéines dont l'insuline (ISN), le récepteur de l'insuline (INSR), la leptine (LEP), l'adiponectine (ADIPQ), le récepteur gamma activé par les proliférateurs de peroxysomes (PAPRG), le facteur de nécrose tumorale (TNF), l'interleukine 6 (TNF), la nicotinamide phosphoribosyltransferase (NAMPT), le

sous-groupe C, membre 1 de la sous-famille 3 des récepteurs nucléaires (NR3C1) et le prépropeptide ghrelin/obestatine (GHRL) (Figure 8).

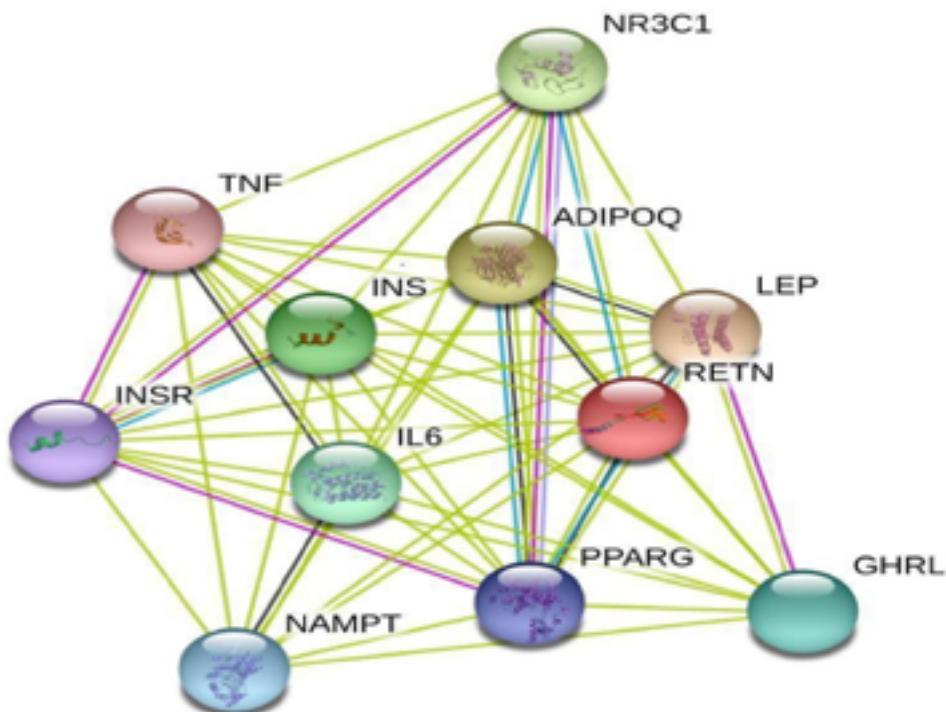


FIGURE 42. RÉSEAU D'INTERACTION PROTÉINE-PROTÉINE DE LA RÉSISTINE EN UTILISANT LE SERVEUR STRING.

TABLEAU 5. INTERACTIONS OBSERVÉES ENTRE LES TROIS MONOMÈRES DE LA PROTÉINE RÉSISTINE DANS LES COMPLEXES NATIFS ET LES 9 MUTANTS.

	Na tif	I32 S	C5 1Y	G5 8E	G5 8R	C7 8S	G 7 9 C	W 98 C	C1 03 G	C10 4Y
Nombre de résidus en interaction Molécule 1(Chaîne A Chaîne B)	29	129	130	127	129	129	128	130	130	129
Nombre de résidus en interaction Molécule 2 (Chaîne C)	8	78	75	77	78	78	78	78	78	75
Nombre d'interactions hydrophiles-hydrophobes	43	357	329	342	338	334	323	343	357	320
Nombre d'interactions hydrophiles-hydrophiles	62	286	266	257	258	261	277	251	259	266

Nombre d'interactions hydrophobes-hydrophobes	93	175	188	194	196	190	179	194	201	189
---	----	-----	-----	-----	-----	-----	-----	-----	-----	-----

TABLEAU 6. LES SNPs (UTR mRNA) QUI ONT ÉTÉ PRÉDITS COMME Étant FONCTIONNELLEMENT SIGNIFICATIFS PAR UTRSCAN.

SNP ID	Changt de nucléotide	Position UTR	Changt d'élément fonctionnel
rs920569876	A/G	3'UTR	Signale de la polyadénylation
rs74176247	A/G	3'UTR	Signale de la polyadénylation
rs1447199134	A/T	3'UTR	Signale de la polyadénylation
rs943234785	A/G	3'UTR	Signale de la polyadénylation
rs76346269	A/T	3'UTR	Signale de la polyadénylation
rs78048640	A/G	3'UTR	Signale de la polyadénylation

#### 4. Discussion

Avec la croissance du nombre de SNPs dans les bases de données, il devient difficile de sélectionner les SNPs à choisir pour contribuer au développement de la maladie. Ainsi, l'analyse computationnelle peut aider à sélectionner un nombre limité de NPs délétères prioritaires pour le dépistage des maladies génétiques simples et complexes (Szklarczyk D.,et al., 2015). Parmi les mutations affectant la fonction des protéines, les nsSNPs sont très fréquemment présentes et sont responsables de nombreuses maladies héréditaires (Sneha P.,et al., 2017). Les SNPs non synonymes ont été essentiellement décrits pour inhiber l'activité des protéines, la liaison ADN-protéine ou ADN-miARN. De nombreux gènes liés à des anomalies métaboliques ont été étudiés avec des approches computationnelles afin de prédire les mutations fonctionnelles telles que le facteur nucléaire 1 alpha des hépatocytes HNF1A, lapolipoprotéine AI APOAI, lapolipoprotéine E3 APOE3 et la lipoprotéine de haute densité HDL (Szklarczyk D.,et al., 2015; Desai M.,et al., 2018; Caulfield T. R., 2011). À ce jour, plus de 70 mutations faux-sens ont été rapportées dans le gène RETN humain. Cependant, les études d'association basées sur la population sur RETN restent insuffisantes. Par conséquent, la présente étude visait à étudier l'impact structurel et fonctionnel des SNPs présents dans le gène RETN en utilisant une analyse computationnelle approfondie. Un total de 78 nsSNPs dans le gène RETN a été extrait de dbSNP.

Les méthodes utilisées dans cette étude ont révélé l'importance d'utiliser divers algorithmes avec différentes capacités de prédiction comme un outil dynamique pour estimer l'effet des variations aux niveaux structurel et fonctionnel. L'utilisation d'un seul outil bioinformatique pour prédire les nsSNPs

potentiellement pathogènes peut ne pas être significative (Parra-Guillen Z. P., et al., 2012). Par conséquent, la présente étude était basée sur plusieurs outils computationnels, dont SIFT, PolyPhen1/2, MAAP, PhD-SNP, PredictSNP, SNAP, Consurf, ModPred, I-Mutant et MUpro, afin d'identifier les nsSNPs les plus délétères dans le gène RETN.

Quinze nsSNPs ont été prédits comme étant les SNPs les plus délétères par ces outils. De plus, en utilisant le serveur web ConSurf, 13 d'entre eux ont été prédits comme étant fortement conservés avec un score de conservation compris entre 7 et 9. De plus, la prédition des sites de PTM a montré que W98 est un site pour l'ADP ribosylation et la clivage protéolytique dans le type sauvage, tandis qu'après mutation (W98C) il est devenu un site pour la liaison disulfure avec l'acquisition d'un résidu Cys, cela peut induire des restrictions conformationnelles sur la protéine modifiant fortement son repliement, sa fonction et sa stabilité. Inversement, les résidus C103 et C104 ont été prédits comme des sites de PTM qui ont perdu des liaisons disulfure. Alors que la structure 3D de la liaison disulfure et les interactions sont très conservées dans la nature, leur perte peut affecter le repliement de la protéine. Neuf nsSNPs ont été significativement prédits pour affecter la stabilité. Ces résultats suggèrent que I32S, C51Y, G58E, G58R, C78S, G79C, W98C, C103G et C104Y pourraient être les SNPs les plus significatifs sur le plan structurel et fonctionnel dans la résistine humaine. Le variant C104Y a présenté la valeur RMS la plus élevée, soit 1.308 Å, suivi de C51Y avec un RMSD de 1.137 Å. De plus, ces variants ont eu une perte significative de liaisons hydrogène par rapport au type sauvage. De manière intéressante, les résidus C51, C78, C103 et C104 sont situés dans l'extrémité C-terminale riche en cystéine fortement conservée de la résistine, comme toutes les protéines de la famille des résistines qui partagent un domaine C-terminale commun avec un espacement invariant entre les positions des cystéines (1C-X11-2C-X8-3C-X-4C-X3-5C-X10-6C-X-7C-X-8C-X9-9C-10C) (Steppan C. M., et al., 2001; Gassoum A., et al., 2016; Holcomb I. N., et al., 2000). Les figures 7C et 2E montrent les mutants C51Y et C104Y qui ont perdu le pont disulfure après un changement de cystéine en résidus de tyrosine. Cela pourrait être susceptible de déstabiliser l'hélice portant des résidus qui pourraient être impliqués dans des sites de liaison avec d'autres protéines. En utilisant STRING, le réseau fonctionnel de la résistine a montré l'interaction avec dix protéines différentes impliquées unanimement dans les voies inflammatoires et métaboliques, suggérant la forte implication de la résistine dans les anomalies métaboliques.

## 5. Conclusion

Le gène RETN a été étudié dans ce travail en évaluant l'impact des SNPs délétères dans les régions codantes et non traduites grâce à une approche computationnelle. Sur un total de 78 SNPs non-sens dans le gène RETN, 15 ont été prédits comme les plus délétères en utilisant PredictSNP. Parmi ceux-ci, 9 nsSNPs ont été prédits comme étant hautement conservés et affectent la stabilité de la protéine. L'analyse structurale a révélé de hauts scores de RMSD, soit 1.308 Å et 1.137 Å pour les variants

C104Y et C51Y respectivement, avec une perte totale des liaisons hydrogène. Six SNPs UTR ont été prédits pour être dans des sites de polyadénylation. Par conséquent, nous concluons que 9 nsSNPs I32S, C51Y, G58E, G58R, C78S, G79C, W98C, C103G et C104Y et 6 substitutions dans 3'UTR, à savoir rs920569876, rs74176247, rs1447199134 rs943234785, rs76346269 et rs78048640 pourraient être d'importants candidats dans le processus pathologique de la résistiné, notamment dans les voies métaboliques.

## Partie 2. PPAR ET CRIBLAGE VIRTUEL

## 1. Introduction

Le syndrome métabolique est un ensemble de troubles qui ont un coût socio-économique élevé et qui est considéré comme une épidémie mondiale. Le syndrome métabolique fait référence à la présence simultanée d'obésité abdominale, de résistance à l'insuline, d'une augmentation des triglycérides (TG), d'une hypertension et d'une réduction du cholestérol des lipoprotéines de haute densité (HDL) (Alberti et Zimmet, 1998; Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001; Einhorn et al., 2003). Plusieurs définitions du syndrome métabolique existent actuellement (Alberti et al., 2005). Les facteurs physiologiques, biochimiques, cliniques et métaboliques interconnectés du syndrome métabolique augmentent directement le risque de maladie cardiovasculaire athérosclérotique (MCV), de diabète de type 2 (T2D) et de toutes les causes de mortalité (Wilson et al., 2005). Le syndrome métabolique confère un risque multiplié par cinq de développer un T2D et un risque multiplié par deux de développer une maladie cardiovasculaire (MCV) au cours des cinq prochaines années (Alberti et al., 2009). De plus, les patients atteints de syndrome métabolique présentent un risque multiplié par deux à quatre de faire un AVC, un risque multiplié par trois à quatre de faire un infarctus du myocarde (IM) et un risque multiplié par deux de décéder de tels événements par rapport à ceux sans syndrome (Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001), indépendamment d'une histoire préalable d'événements cardiovasculaires (Alberti et al., 2005). De plus, la présence à la fois du syndrome métabolique et de l'obésité provoque une inflammation tissulaire locale de faible intensité chronique et augmente la susceptibilité à d'autres affections telles que la stéatose hépatique, l'asthme, les troubles du sommeil, les calculs biliaires de cholestérol et certains types de cancer (Serafino-Agrusa et al., 2015; Caterson et al., 2004). La prise en charge du syndrome métabolique repose sur des modifications du mode de vie visant à rétablir l'équilibre énergétique, en plus des interventions pharmaceutiques. En ce qui concerne le traitement des patients atteints du syndrome métabolique, les médicaments utilisés ciblent différents aspects pertinents du syndrome métabolique tels que le poids corporel et la répartition des graisses, la résistance à l'insuline, l'hypertension, la dyslipidémie, l'hyperglycémie ou l'état prothrombotique et pro-inflammatoire établi (Rochlani et al., 2017). Le récepteur nucléaire activé par les proliférateurs des peroxysomes γ (PPAR $\gamma$ ) est l'un des facteurs de transcription activés par les ligands dans la superfamille des récepteurs hormonaux nucléaires et un régulateur essentiel de l'homéostasie du glucose et des lipides.

PPAR $\gamma$  est un régulateur essentiel de la sensibilité à l'insuline, de l'homéostasie lipidique, de l'inflammation et du métabolisme du glucose, et représente donc une cible pharmacologique importante pour la découverte de médicaments qui peuvent moduler simultanément plusieurs pathologies sous-jacentes du syndrome métabolique (Berger et al., 2005; Monsalve et al., 2013). Les activateurs de PPAR $\gamma$  se sont révélés puissants dans la lutte contre l'hypertension et le syndrome

métabolique (Stump et al., 2015; Mansour, 2014; Algandaby, 2020). Les thiazolidinediones (TZD) sont des agonistes de PPAR $\gamma$  et ont été approuvées en tant que première nouvelle classe de médicaments pour réduire la résistance à l'insuline chez les patients atteints de T2D (Stump et al., 2015; Fitzgerald et al., 2002; Laloyer et Staels, 2010). La classe des TZD agit en tant qu'agonistes complets de PPAR $\gamma$  via le mécanisme de verrouillage médié par la fonction d'activation 2 (AF 2). La troglitazone était le premier médicament de cette classe capable d'augmenter la sensibilité à l'insuline et la tolérance au glucose chez les sujets obèses (Plosker et Faulds, 1999). Ce médicament a été approuvé sur le marché américain en 1997 comme traitement du T2D. Il était disponible jusqu'à ce que la Food and Drug Administration (FDA) annonce son association avec le risque de survenue de l'hépatite fulminante. Malgré le retrait du médicament en raison d'une hépatotoxicité grave, il était clair que l'approche thérapeutique avec des agents augmentant la sensibilité à l'insuline comme la troglitazone avait un potentiel significatif pour le traitement des patients atteints de T2D. En conséquence, deux autres TZD, la rosiglitazone et la pioglitazone, ont été développées et approuvées pour la commercialisation. Les deux médicaments avaient des effets hypoglycémiants similaires à ceux de la troglitazone, mais avec un risque considérablement moindre d'effets secondaires hépatiques (Martens et al., 2002; Koshiyama et al., 2000). Les deux autres médicaments de la classe des TZD sont la rosiglitazone et la pioglitazone, qui sont tous deux utilisés cliniquement dans de nombreux pays pour le contrôle glycémique dans le traitement du T2D (Werner et Travaglini, 2001; Phillips et al., 2001). Cependant, la rosiglitazone a été retirée du marché européen. Aux États-Unis, l'utilisation de la rosiglitazone a été restreinte par la FDA en raison de certaines incertitudes scientifiques concernant la sécurité cardiovasculaire de son effet (Psaty et Furberg, 2007). Cette étude explore la découverte de médicaments par le biais de méthodes informatiques telles que le criblage virtuel, la modélisation basée sur la structure, les simulations de dynamique moléculaire et les calculs d'énergie libre. Elle vise à étudier les interactions entre la protéine PPAR $\gamma$  et les phytochimiques marocains, dans le but spécifique d'identifier des composés naturels présentant une grande stabilité et une forte affinité de liaison pour PPAR $\gamma$ . Les chercheurs comparent ensuite leurs résidus de liaison dans leur conformation de docking à ceux de la rosiglitazone décrite dans la littérature (Berger et Moller, 2002; Brust et al., 2018).

Cette recherche a des implications importantes pour le développement d'activateurs naturels alternatifs de PPAR $\gamma$ , car elle analyse les résidus de liaison et les groupes fonctionnels des composés identifiés, offrant ainsi un aperçu des mécanismes sous-jacents des interactions protéine-ligand de PPAR $\gamma$  (Genheden et Ryde, 2015). De plus, l'accent mis sur les phytochimiques marocains est particulièrement remarquable, car les composés naturels représentent une ressource précieuse et peu explorée pour la découverte de médicaments (Wang et al., 2022; Yan et al., 2020). L'identification de puissants composés naturels présentant une forte affinité de liaison et une grande stabilité pour PPAR $\gamma$  pourrait ouvrir la voie à de nouvelles thérapies potentielles pour des maladies telles que le

diabète de type 2 et le syndrome métabolique. Dans l'ensemble, cette étude fournit des informations précieuses pour la découverte de médicaments et met en évidence l'importance d'explorer les composés naturels dans la recherche de nouveaux traitements.

## 2.Méthodes

### 2.1 PRÉPARATION DE LA PROTÉINE

La structure cristalline du domaine de liaison du ligand de PPAR $\gamma$  en complexe avec la rosiglitazone (PDB ID 7awc) a été récupérée de la Protein Data Bank (PDB) avec une résolution de 1,74 Å (Berger et Moller, 2002). Ce complexe a été préparé pour le docking à l'aide des logiciels Pymol et PyRx (Trott et Olson, 2010). Le processus de préparation comprenait l'ajout de résidus et d'atomes manquants, l'élimination des métaux et des molécules d'eau indésirables, et l'ajout d'atomes d'hydrogène.

### 2.2 BIBLIOTHÈQUE DES PHYTO-COMPOSÉS MAROCAINS

Une bibliothèque de 600 phyto-composés marocains a été recueillie de diverses plantes aromatiques et médicinales marocaines. Cette bibliothèque n'a pas été créée sur la base de critères littéraires spécifiques liés au diabète ou à des maladies connexes, mais plutôt pour explorer la diversité des phyto-composés marocains. Ces composés ont été inclus dans la base de données des phyto-composés marocains (MPDB) à l'adresse <https://www.mpdb.org> (référence à ajouter lorsque disponible).

### 2.3 CRIBLAGE VIRTUEL

Nous avons effectué un Docking avec le programme open source de criblage virtuel PyRx, le site de liaison du récepteur a été défini en fonction du mode de liaison de l'agoniste total rosiglitazone (PDB ID 7awc), Pendant le processus de docking, la protéine était maintenue rigide et les ligands étaient flexibles avec toutes leurs liaisons de torsion libres de tourner. Une grille cubique de dimensions 21 Å×29Å× 24Å avec des points séparés de 0.375Å a été générée et a englobé tous les résidus du site actif de PPAR $\gamma$ .

### 2.4 INTÉGRATION DE L'ÉNERGIE DE LIAISON DANS LA BASE DE DONNÉES EN LIGNE

Le criblage virtuel a été effectué avec le programme open source PyRx (Trott et Olson, 2010). Le site de liaison du récepteur a été défini en fonction de la façon dont l'agoniste total rosiglitazone se lie à PPAR $\gamma$ . Durant le processus de docking, la protéine était maintenue rigide, alors que les ligands étaient flexibles.

Les valeurs d'énergie de liaison correspondant à la pose avec un RMSD de 0 Å ont été déterminées et incorporées dans la base de données en ligne MPDB. Cela a été fait dans le but de stimuler la recherche et l'industrie pour explorer et exploiter les propriétés distinctives des phyto-composés marocains (référence à ajouter lorsque disponible).

## 2.5 ANALYSE DE L'INTERACTION LIGAND-RÉCEPTEUR

L'analyse des interactions ligand-récepteur a été réalisée à l'aide de BIOVIA Discovery Studio Visualizer (Dassault Systèmes, 2020). Ce logiciel génère une visualisation 2D des liaisons hydrogène et des interactions hydrophobes, qui sont essentielles à l'affinité des composés pour le site actif de PPAR $\gamma$ . Pour évaluer les propriétés des trois composés d'intérêt similaires à celles d'un médicament, nous avons analysé leurs propriétés physico-chimiques conformément à la Règle des Cinq de Lipinski (Lipinski et al., 2001). Cette règle permet de prédire si un composé chimique a le potentiel de devenir un médicament oral efficace en fonction de ses propriétés d'absorption, de distribution, de métabolisme et d'excrétion (ADME).

## 2.6 SIMULATION DE DYNAMIQUE MOLÉCULAIRE

Nous avons également réalisé une simulation de dynamique moléculaire de 100 ns pour la protéine PPAR $\gamma$  et les deux complexes PPAR $\gamma$ -phytocomposé précédemment ciblés, en les comparant avec le comportement dynamique du complexe PPAR $\gamma$ -rosiglitazone (PDB ID: 7AWC) (Berger et Moller, 2002). Cette simulation a fourni des informations sur la stabilité et les changements structuraux des ligands, des protéines et des complexes protéine-ligand tout au long de différentes trajectoires. Les simulations ont été réalisées à l'aide du logiciel GROMACS 2018.2 (Abraham et al., 2015). Le processus de simulation a commencé avec les structures résultant de l'analyse de docking. Il comprenait généralement six étapes : génération de topologie, construction d'une boîte, solvatation, minimisation de l'énergie du système, équilibration du système et production de MD. Après quoi, des trajectoires ont été générées et les résultats analysés. Les paramètres de protonation et de minimisation ont été appliqués aux systèmes en utilisant le champ de force GROMOS96 43A1 (van Gunsteren et al., 1996). Les paramètres du fichier de topologie moléculaire pour les différents ligands ont été générés à l'aide du serveur PRODRG (Schuttelkopf et van Aalten, 2004). Le système a ensuite été neutralisé et la minimisation de l'énergie réalisée à travers 50 000 étapes de descente la plus abrupte. Après l'équilibration à 300 K et 1 bar pendant 100 ps, le système complet a été soumis à une simulation MD de 100 ns. Les coordonnées atomiques ont été stockées tous les 0,002 ps pendant la simulation pour une analyse ultérieure. Les trajectoires résultantes ont été analysées en détail en termes d'écart quadratique moyen (RMSD), de fluctuation quadratique moyenne (RMSF), de rayon de giration (Rg), de surface accessible au solvant (SASA) et du nombre de liaisons hydrogène. Ces analyses ont été effectuées à l'aide des fonctions intégrées de GROMACS et les graphiques correspondants ont été tracés avec le logiciel GRACE (Turner, 2005).

## 2.7 CALCUL DE L'ÉNERGIE DE LIAISON LIBRE

Les études de dynamique moléculaire ont été menées sur une période de 30 nanosecondes, et l'énergie libre de liaison des différents complexes a ensuite été calculée à l'aide du package "g\_mmpbsa". Cet outil logiciel a été développé en utilisant deux programmes open-source

populaires, GROMACS et APBS. La méthode MM-PBSA a été utilisée pour calculer les composants de l'énergie de liaison, à l'exception du terme entropique, et la contribution énergétique de chaque résidu à la liaison a été déterminée à l'aide d'un schéma de décomposition d'énergie (Genheden S, et al, 2015). Les résultats obtenus ont ensuite été utilisés comme entrées pour des scripts Python afin d'obtenir l'énergie de liaison finale.

### 3. Résultats

#### 3.1 CRIBLAGE DE LA COMPOSÉS MAROCAINES ET AMARRAGE MOLÉCULAIRE

La bibliothèque de ligands phyto-composés générée a été amarrée contre PPAR $\gamma$  et les composés amarrés ont été classés sur la base d'un filtre rigoureux qui inclut des facteurs tels que la force des liaisons hydrogène et un solide réseau de liaisons hydrophobes. Sur les 600 phyto-composés amarrées, les postes d'amarrage les mieux classées ont été sélectionnés. Après analyse de l'énergie de liaison, deux phyto-composés, le 3,4-DICQA et l'acide chlorogénique, ont été trouvés se lier avec une forte affinité dans le site actif du récepteur et forme un solide réseau d'interactions hydrogène et hydrophobes (Figure 9, Tableau 8). En particulier, le 3,4-DICQA avait un score de liaison de -8,2 Kcal/mol (Tableau 7) et a été trouvé se lier dans la même poche dans le site catalytique que l'agoniste total rosiglitazone. Il a formé des liaisons hydrogène avec un ensemble de résidus de la région AF1, y compris Tyr473 et His449, ainsi qu'avec Ser289, Tyr327, et Met364 du domaine de liaison au ligand (Figure 9b). De plus, Cys285 et Arg288 ont formé des interactions de soufre avec les cycles de benzène du 3,4-DICQA, impliquant de fortes interactions hydrophobes avec les résidus du domaine de liaison au ligand, y compris Ile341, Glu295, His323, Ser342, et Ile326 (Figure 10 b).

L'acide chlorogénique a été observé se lier par des liaisons hydrogène avec les résidus de la feuille bêta, y compris Ser342 et Ile341, ainsi qu'avec Ser289, Tyr327, Ile281, et Cys285, formant un réseau de liaisons hydrogène dans le site actif. Des interactions hydrophobes fortes et de van der Waals ont également été observées avec d'autres résidus de la région AF1 du domaine de liaison au ligand, tels que His323 et His449, comme illustré sur la Figure 10.

TABLEAU 7. RÉSULTATS DES PRINCIPAUX PHYTO-COMPOSÉS AMARRÉS CONTRE PPAR $\gamma$ , AVEC LEUR AFFINITÉ DE LIAISON RESPECTIVE, LES RÉSIDUS DE LIAISON HYDROGÈNE, ET LES NOMS DES PLANTES MAROCAINES SOURCES.

Nom et ID MPDB	Affinité de liaison	Source de plante marocaine	Interaction de liaisons h	Structure 2D
Rosiglitazone	-8.5 kcal /mol	Synthétique	Tyr473, Ser289, Cys285	

3,4-DICQA MNPDB00479	-8.6 kcal /mol	Cotula cinera	Tyr473, Met364, Ser289, Tyr327, His449	
Chlorogenic acid MNPDB00380	-8.0 kcal /mol	Anabasis aretioides Coss. & Moq Achyranthes aspera	Ser342, Ile341, Ser289, Tyr327, Ile281	

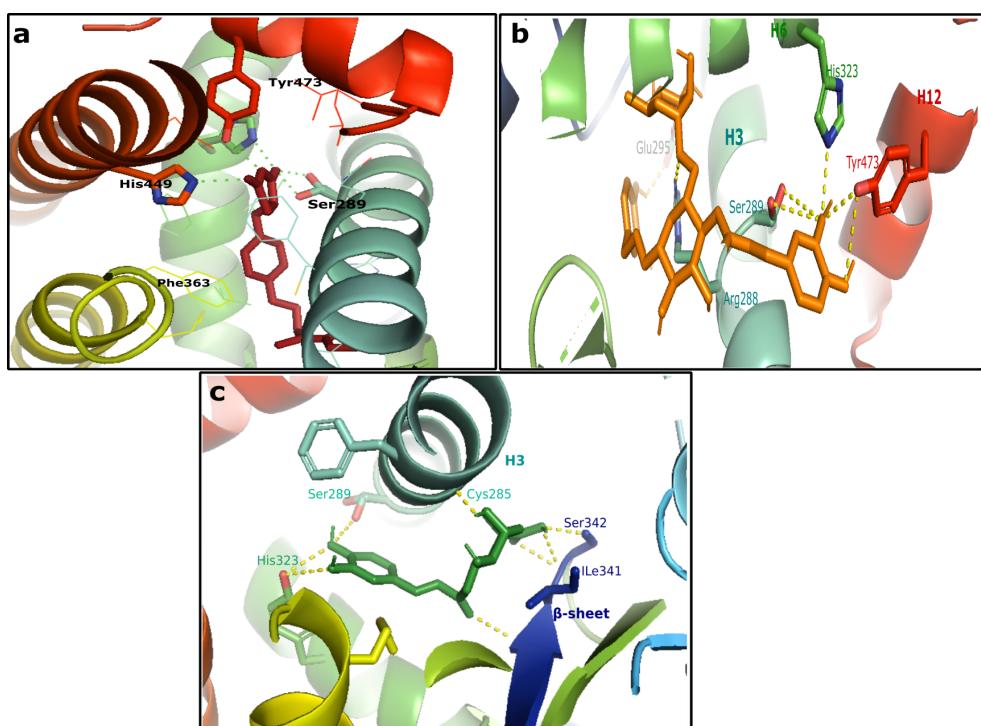


FIGURE 43. INTERACTION 3D DE LA ROSIGLITAZONE ET DES PHYTO-COMPOSÉS IDENTIFIÉS DANS LE SITE ACTIF DE PPAR $\gamma$ , LIAISON H  
REPRÉSENTÉE PAR DES POINTILLÉS JAUNES, (A) ROSIGLITAZONE (B) 3,4-DICQA EN COULEUR ORANGE (C) ACIDE CHLOROGÉNIQUE EN  
COULEUR VERT

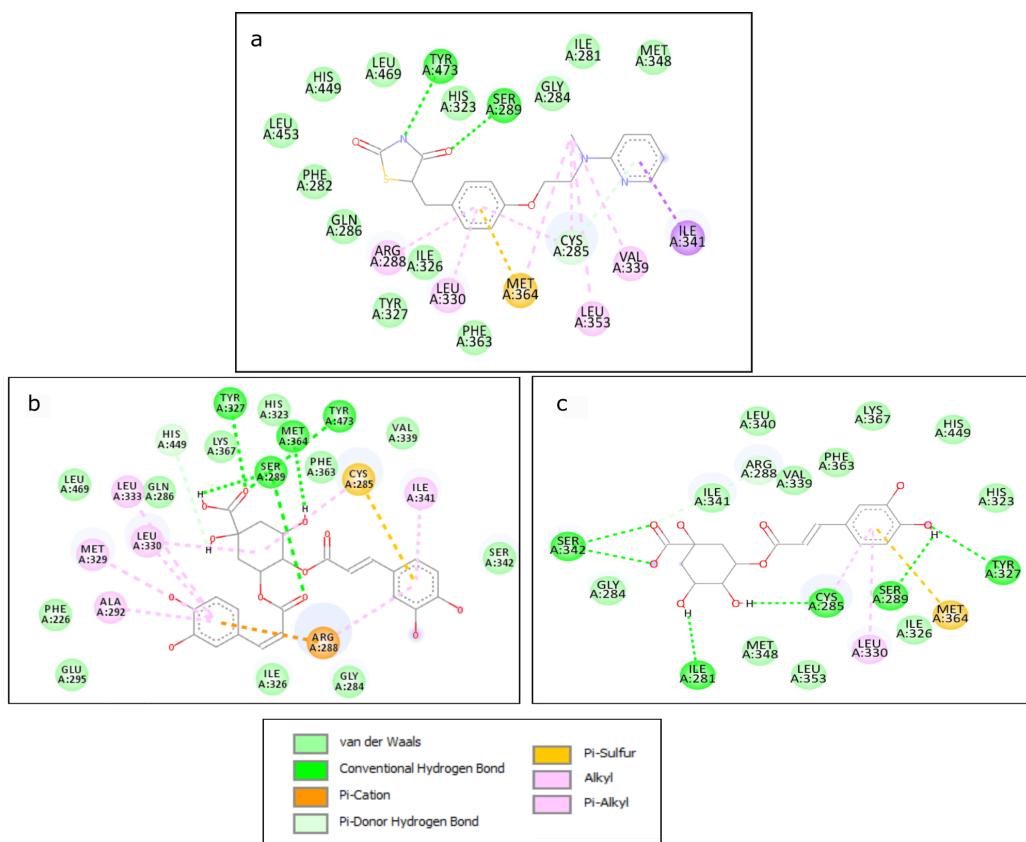


FIGURE 44. STRUCTURES 2D DE LA ROSIGLITAZONE DE RÉFÉRENCE ET DES DEUX MEILLEURS PHYTO-COMPOSÉS DOCKÉS (A) ROZIGLITAZONE (B) ACIDE 3,4-DICAFÉOYLQUINIQUE (C) ACIDE CHLOROGÉNIQUE.

### 3.2 PROPRIÉTÉS DE TYPE MÉDICAMENT DES RÉSULTATS IDENTIFIÉS

Le 3,4-DICQA et l'acide chlorogénique sont deux composés d'intérêt pour la découverte de médicaments. Lorsque évalués selon la règle des cinq de Lipinski, le 3,4-DICQA viole deux règles : le poids moléculaire et les donneurs de liaisons hydrogène, ce qui peut réduire sa probabilité d'être un médicament actif par voie orale réussi. En revanche, les deux acides chlorogéniques ne violent qu'une seule règle, les donneurs de liaisons hydrogène, tout en respectant les trois autres critères. Bien que cela suggère un potentiel plus élevé pour ces deux composés pour être des médicaments actifs par voie orale réussis, il est important de se rappeler que la règle des cinq de Lipinski est une ligne directrice avec des exceptions. Certains composés peuvent toujours être des médicaments réussis même s'ils ne respectent pas strictement les règles, tandis que d'autres qui répondent aux critères peuvent ne pas l'être. La règle des cinq sert d'outil précieux pour guider les premières étapes de la découverte de médicaments, mais elle ne doit pas être considérée comme un déterminant absolu du succès. (Tableau 9).

**TABLEAU 8. RÉSULTATS DES MEILLEURS AMARRÉS CONTRE PPAR $\gamma$ , AVEC LEUR AFFINITÉ DE LIAISON RESPECTIVE, LES RÉSIDUS INTERAGISSANT PAR LIAISONS HYDROGÈNE ET HYDROPHOBES.**

	MW	logP	HBA	HBD	Solubilité	Biodisponibilité orale	Résultat
3,4-DICQA MNPDB00479	<b>516.45</b>	<b>1.52</b>	<b>12</b>	<b>7</b>	<b>Bonne</b>	<b>Bonne</b>	<b>Accepté</b>
Chlorogenic acid MNPDB00380	<b>354.31</b>	<b>-0.42</b>	<b>9</b>	<b>6</b>	<b>Bonne</b>	<b>Bonne</b>	<b>Accepté</b>

### 3.3 SIMULATION MOLÉCULAIRE DYNAMIQUE

Comme démontré par l'étude de docking, ces composés ont montré une forte affinité et un solide réseau d'interaction avec PPAR $\gamma$ . Pour évaluer la stabilité des données de docking et pour vérifier la validité des résultats, les deux complexes (3,4-DICQA et acide chlorogénique) ont été simulés avec une dynamique moléculaire pendant 100ns à l'aide du logiciel GROMACS, et pour comparaison, la structure de référence de PPAR $\gamma$ -Rosiglitazone et PPAR $\gamma$ - seul sans ligand a également été soumise à une simulation MD de 100 ns. La dynamique moléculaire a été effectuée pendant 100 ns pour analyser la stabilité et la liaison H des cinq complexes. En conséquence, nous avons généré dans les Figures 3,4,5 et 6 les trajectoires MD pour trois complexes plus le PPAR $\gamma$  seul sans ligand pour révéler les changements se produisant en présence des modulateurs identifiés. Les trajectoires sont superposées les unes sur les autres, PPAR $\gamma$ -Rosiglitazone est représenté en rouge, PPAR $\gamma$ -3,4-DICQA en orange, PPAR $\gamma$ -Acide chlorogénique en vert, PPAR $\gamma$  seul en bleu. Pour le PPAR $\gamma$ -Rosiglitazone et le PPAR $\gamma$ -3,4-DICQA, la variation de Ca-RMSD des deux complexes varie entre 0 et 0,48 nm (Figure 11.a). En général, aucune variation considérable des fluctuations de RMSD entre les deux complexes n'a été observée, la variation de RMSD était similaire pendant 25ns-48ns et pendant 90ns jusqu'à la fin de la simulation.

Pour le PPAR $\gamma$ -Acide chlorogénique et le PPAR $\gamma$  seul, les deux complexes ont montré une grande stabilité par rapport à l'agoniste total rosiglitazone, le RMSD des deux complexes varie autour de 0,22 et 0,38 ns pendant les 100 ns de la simulation (Figure 11.a). Leur variation de RMSD était similaire dans les deux à l'exception de petites fluctuations pendant 33ns-39ns et entre 68ns-78ns où la variation de RMSD de PPAR $\gamma$  a augmenté par rapport à PPAR $\gamma$ -Acide chlorogénique (Figure 11.a).

Les dimensions globales des deux PPAR $\gamma$ -3,4-DICQA et PPAR $\gamma$ -Acide chlorogénique ont montré la même variation que le complexe de la rosiglitazone et tous les complexes restent stables autour de 1,9 nm pendant la simulation, à l'exception de PPAR $\gamma$  seul qui conserve les petites valeurs de Rg autour de 1,86 nm (Figure 11.b).

La surface accessible au solvant (SASA) a été calculée pour évaluer le maintien de l'empaquetage de la protéine dans le système (Figure 11.c). L'analyse SASA de tous les complexes a montré une modeste différence dans les propriétés de surface autour de 125-130 nm par rapport au complexe de

référence de rosiglitazone (entre 127-130nm<sup>2</sup>), à l'exception de PPAR $\gamma$ -Acide chlorogénique qui a montré des valeurs basses autour de 120 nm (Figure 11.c)

Le nombre de liaisons hydrogène impliquées dans toutes les interactions complexes varie entre 0 et 10 pendant les 100 ns de simulation (Figure 13). D'autre part, comme le montre la Figure 12, les valeurs de RMSF des différents complexes pendant 100 ns, en général, les résidus de PPAR $\gamma$  seul ont montré de fortes fluctuations par rapport à l'agoniste total et aux autres complexes. En effet, l'acide chlorogénique a montré des valeurs basses par rapport aux autres complexes (Figure 12.). De plus, les résidus de l'hélice 12 ont montré moins de valeurs de fluctuation avec la rosiglitazone (0,15nm), avec 3,4-DICQA et l'acide chlorogénique (0,18), tandis que le H12 dans PPAR $\gamma$  seul montre les valeurs élevées de RMSF arrivent à 0,54 nm suggérant que les interactions avec 3,4-DICQA et l'acide chlorogénique stabilisent le H12 de manière similaire à la rosiglitazone (Figure 12.b). L'observation de la valeur de flexibilité des résidus de feuille bêta (340-351) a montré que le récepteur seul montre les valeurs les moins élevées de RMSF (0,12-0,2nm) suivies de l'acide chlorogénique avec des valeurs entre 0,1nm et 0,27 nm, tandis que les résidus de feuille bêta avec le 3,4-DICQA et la rosiglitazone (Figure 4.b) ont montré les valeurs élevées de RMSF variant entre (0,8-0,33nm) et (0,1-0,27 nm) respectivement, ce qui laisse supposer que les résidus de feuille bêta stabilisés avec les liaisons hydrogène dans l'acide chlorogénique peuvent réaliser une interaction stable pendant le temps de simulation.

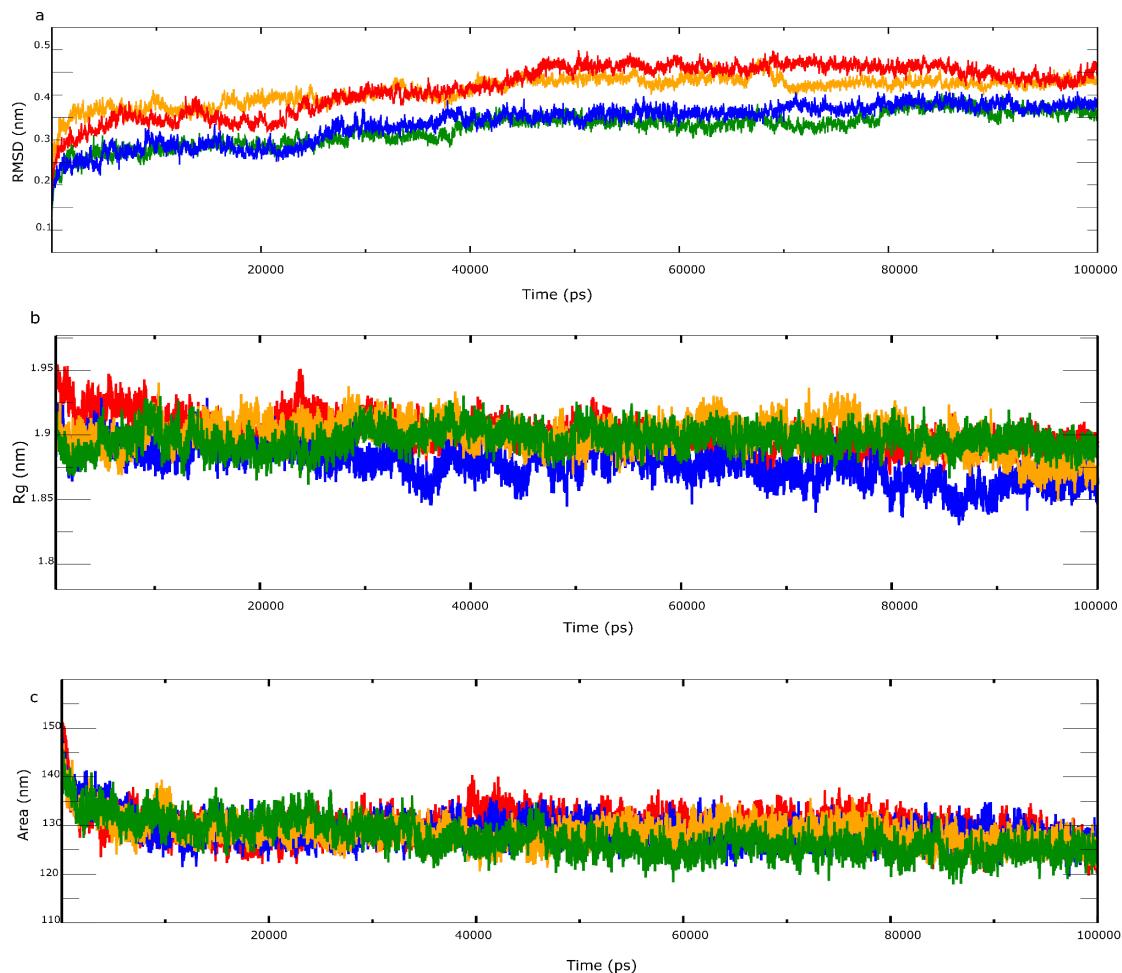


FIGURE 45. (A) TRACÉ RMSD DES COMPLEXES PPAR $\gamma$ , (B) TRACÉ DU RAYON DE GIRATION DE DIFFÉRENTS COMPLEXES PPAR $\gamma$  AVEC DES PANNEAUX CODÉS EN COULEUR, ET (C) TRACÉ DE LA SURFACE ACCESSIBLE AU SOLVANT (SASA) DE DIFFÉRENTS COMPLEXES PPAR $\gamma$  AVEC DES PANNEAUX CODÉS EN COULEUR.

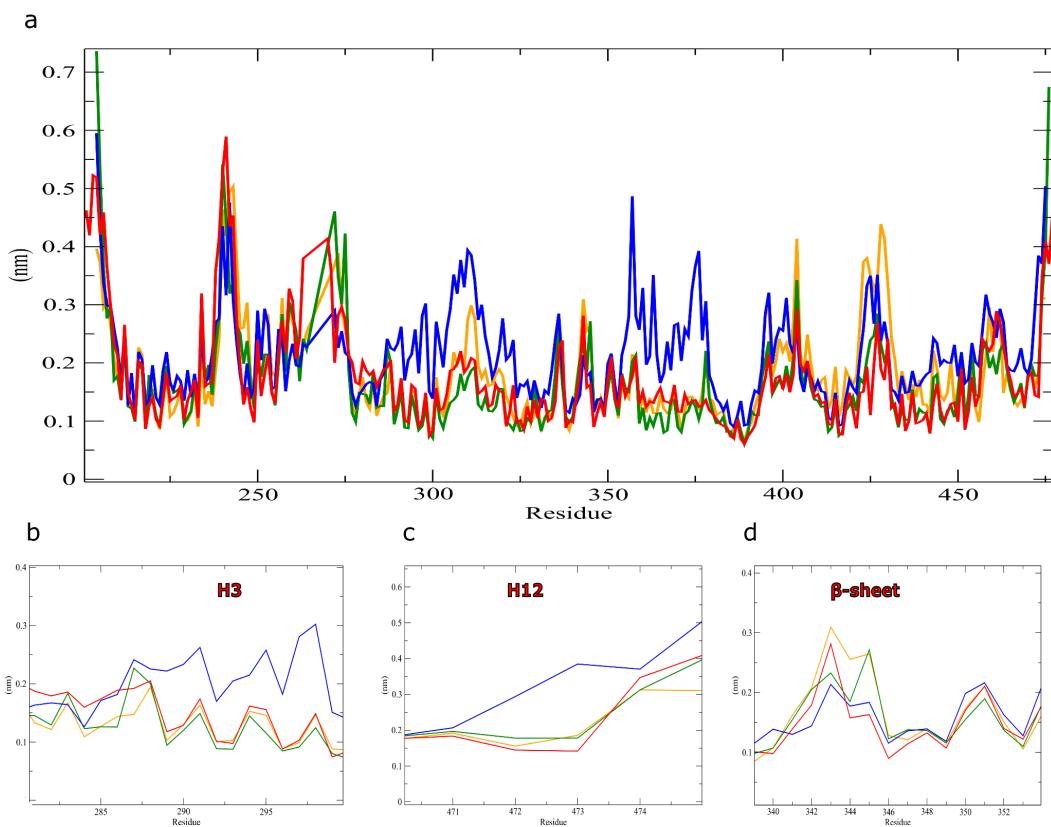


FIGURE 46. TRACÉ DES FLUCTUATIONS MOYENNES QUADRATIQUES (RMSF) DU PPAR $\gamma$  COMPLEXÉ AVEC DEUX PHYTO-COMPOSÉS DIFFÉRENTS ET LA ROSIGLITAZONE, APO PPAR $\gamma$  EN TANT QUE RÉFÉRENCE POUR LA SIMULATION DE 100 NS. (A) FLUCTUATION DE L'HÉLICE 3 DU PPAR $\gamma$ , (B) FLUCTUATION DE H12, ET (C) FLUCTUATION DE LA FEUILLE BÊTA.

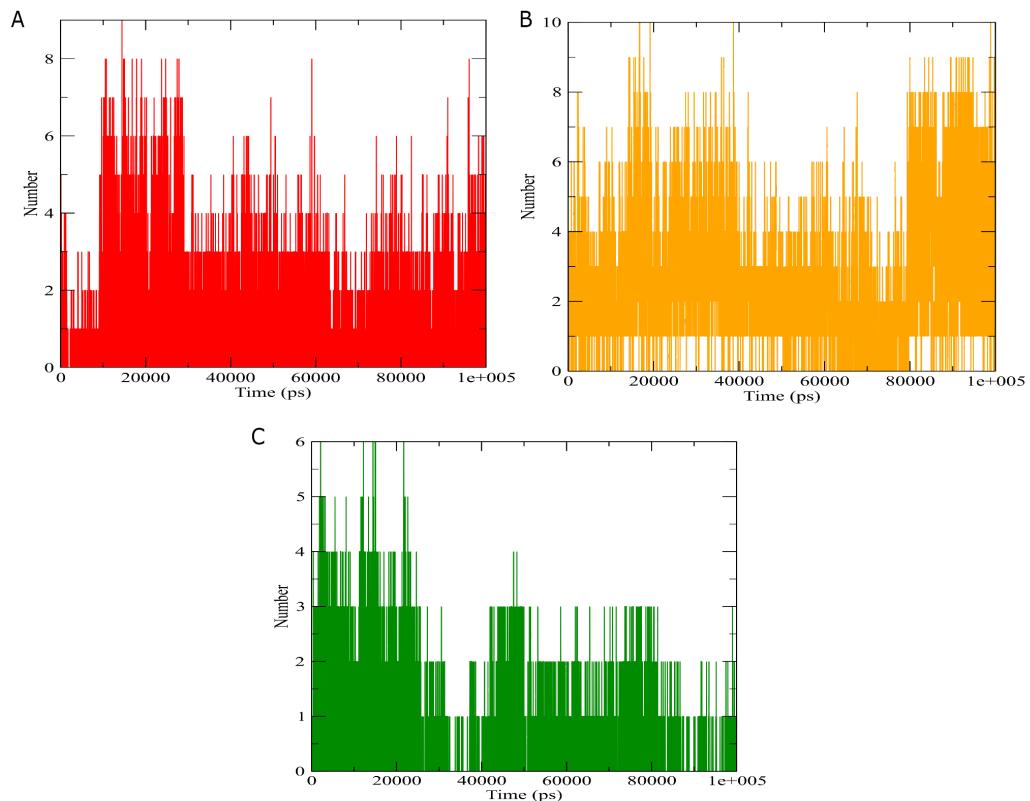


FIGURE 47. DIAGRAMME 2D DES MOTIFS DE LIAISON HYDROGÈNE OBSERVÉS POUR LES DEUX COMPLEXES DIFFÉRENTS ET LA ROSIGLITAZONE EN TANT QUE RÉFÉRENCE DU PPAR $\gamma$  PENDANT LA SIMULATION DE 100 NS.

### 3.4 ÉNERGIE DE LIAISON LIBRE

Une analyse comparative des paramètres énergétiques pour les phyto-composés interagissant avec le récepteur gamma activé par les proliférateurs de peroxysomes (PPAR gamma) a montré que l'acide chlorogénique (MNPDB00380) et le 3,4-DICQA (MNPDB00479) présentaient des affinités de liaison plus fortes par rapport à la rosiglitazone. L'acide chlorogénique avait l'affinité de liaison la plus forte avec une énergie de liaison de -139.442 kJ/mol, indiquant l'interaction la plus robuste avec le récepteur PPAR-gamma, tandis que le 3,4-DICQA suivait de près avec une énergie de liaison de -124.179179 kJ/mol. De plus, l'acide chlorogénique (MNPDB00380) a affiché l'énergie électrostatique la plus favorable (-58.928 kJ/mol), suggérant des interactions électrostatiques améliorées avec le récepteur, tandis que le 3,4-DICQA (MNPDB00479) a démontré des énergies électrostatiques moins favorables par rapport à la rosiglitazone. En ce qui concerne l'énergie de solvatation polaire, tous les composés ont présenté des valeurs dans une plage similaire, l'acide chlorogénique (MNPDB00380) présentant une valeur légèrement supérieure (46.069 kJ/mol) que les autres. En termes d'énergie de van der Waals, l'acide chlorogénique (MNPDB00380) a montré la valeur la plus favorable (-121.829 kJ/mol), impliquant des interactions non polaires fortes avec le récepteur. La rosiglitazone et le 3,4-DICQA (MNPDB00479) ont affiché des énergies de van der Waals moins favorables. En analysant l'énergie SASA, l'acide chlorogénique (MNPDB00380) est apparu comme le candidat le plus favorable (-4.754 kJ/mol), reflétant peut-être une exposition accrue au solvant ou des changements conformationnels lors de la liaison. À l'inverse, le 3,4-DICQA (MNPDB00479) a affiché une énergie SASA moins favorable par rapport à la rosiglitazone. Globalement, l'acide chlorogénique et le 3,4-DICQA sont recommandés comme candidats potentiels pour une enquête plus approfondie en tant qu'activateurs de PPAR gamma.

TABLEAU 9. VALEURS MOYENNES ET ÉCARTS-TYPES DES ÉNERGIES DE LIAISON DES COMPLEXES CALCULÉES PAR LES MÉTHODES G\_MMPBSA.

Nom du composé	Énergie de liaison	Énergie électrostatique	Énergie de solvatation polaire	Énergie de van der Waals	Énergie SASA
Rosiglitazone	-120.329 kJ/mol	-40.802 kJ/mol	36.747 kJ/mol	-106.221 kJ/mol	-10.053 kJ/mol
3,4-DICQA MNPDB00479	-124.179179 kJ/mol	-36.747 kJ/mol	36.822 kJ/mol	-112.053 kJ/mol	-12.201 kJ/mol

Chlorogenic acid MNPDB00380	-139.442 kJ/mol	-58.928 kJ/mol	46.069 kJ/mol	-121.829 kJ/mol	-4.754 kJ/mol
--------------------------------	-----------------	----------------	---------------	-----------------	---------------

#### 4. Discussion

Sur plusieurs siècles, les produits naturels ont été largement utilisés comme principale source de nombreux traitements de maladies. De nombreux constituants et/ou extraits dérivés de plantes ont été utilisés pour traiter des maladies complexes (Berger et al., 2005; Monsalve et al., 2013; Stump et al., 2015; Mansour, 2014; Algandaby, 2020). Des efforts de recherche importants ont été et continuent d'être entrepris afin d'explorer les structures naturelles prometteuses pour la découverte de médicaments (Fitzgerald et al., 2002). Comme l'incidence des troubles métaboliques et en particulier du diabète de type 2 (T2D) continue d'augmenter considérablement, la recherche de médicaments alternatifs et abordables semble atténuer la charge clinique de ces maladies dans le monde entier (Brust et al., 2018). Les médicaments contre le diabète ciblent leur genèse multifactorielle pour fournir un effet thérapeutique basé sur les sécrétagogues de l'insuline et les sensibilisateurs à l'insuline (Berger et al., 2002). Le récepteur activé par les proliférateurs de peroxysomes gamma ( $\text{PPARY}$ ) est un récepteur nucléaire qui joue un rôle crucial dans la régulation du métabolisme du glucose et des lipides, ce qui en fait un composant essentiel dans la gestion du diabète, du syndrome métabolique et des maladies cardiovasculaires (Berger et al., 2002; Monsalve et al., 2013). L'activation du  $\text{PPARY}$  a été montrée pour améliorer la sensibilité à l'insuline, favoriser la captation du glucose dans le tissu adipeux et le muscle squelettique et réduire l'inflammation, qui sont des facteurs critiques dans la pathogenèse du diabète de type 2 et du syndrome métabolique (Berger et al., 2002; Brust et al., 2018). De plus, l'activation du  $\text{PPARY}$  a été associée à des améliorations de la fonction endothéliale et à une réduction de l'athérosclérose, contribuant à une meilleure santé cardiovasculaire (Berger et al., 2002; Brust et al., 2018).

L'utilisation d'agonistes du  $\text{PPARY}$ , tels que les thiazolidinediones, s'est avérée efficace dans le traitement de ces conditions en modulant l'activité du  $\text{PPARY}$  et en ciblant ses effets en aval (Algandaby, 2020). Par conséquent, une compréhension plus approfondie du  $\text{PPARY}$  et de son activation est essentielle pour le développement de nouvelles stratégies thérapeutiques pour combattre le diabète, le syndrome métabolique et les maladies cardiovasculaires (Berger et al., 2005; Monsalve et al., 2013; Stump et al., 2015; Mansour, 2014). Les méthodes computationnelles ont été prouvées à travers de nombreux travaux, comme des méthodes puissantes pour guider la découverte de médicaments de molécules capables de se lier efficacement à des cibles biologiques, comme les protéines (Trott et Olson, 2010). Ces interactions peuvent être exploitées pour la découverte de modulateurs de protéines, ou de modulateurs dans le corps humain (Free Download: BIOVIA Discovery Studio Visualizer). Nos résultats sont en accord avec cette approche et suggèrent que les

molécules sélectionnées ont un fort potentiel d'affinité de liaison avec le PPAR $\gamma$  (Lagorce et al., 2017). Les résultats de docking ont montré que deux molécules sur 600 phytochimiques avaient les meilleurs scores de docking et montraient la force des liaisons hydrogène, et un solide réseau de liaisons hydrophobes avec des résidus fonctionnels, suggérant ainsi qu'elles pourraient activer le récepteur (Brust et al., 2018). Les mécanismes structurels sous-jacents à l'activation du PPAR $\gamma$  sont bien compris. Les agonistes stabilisent un état actif de la surface AF-2 en formant une interaction avec des résidus près de l'hélice 12 (Brust et al., 2018). Les agonistes complets du PPAR $\gamma$  forment une liaison hydrogène directe avec Tyr473 sur l'hélice 12, provoquant ainsi une activation transcriptionnelle (Brust et al., 2018). Les agonistes partiels ne forment généralement pas de liaison hydrogène avec Tyr473, mais ils stabilisent l'hélice 12 par des interactions avec d'autres régions de la poche de liaison du ligand (Brust et al., 2018). De plus, une étude récente indique que la liaison hydrogène du ligand à Arg288 pourrait être un médiateur critique de l'agonisme inverse sélectif du PPAR $\gamma$  qui semble prometteur pour améliorer l'indice thérapeutique associé aux ligands antidiabétiques du PPAR $\gamma$  (Brust et al., 2018). Les agonistes partiels sont délimités par H3, la feuille  $\beta$ , et la boucle  $\omega$ . Ici, l'acide chlorogénique modifie la structure de liaison des agonistes partiels au sein des phytocomposés identifiés. Le ligand occupe le domaine de liaison du ligand qui est principalement délimité par la feuille  $\beta$  et H3; en particulier, il forme une combinaison de liaisons H avec Arg288 de H3 et des liaisons H avec l'amide de la chaîne principale de Ser342, Glu343, ainsi que des interactions extensives de Van der Waals et hydrophobes avec Ile341 de la feuille  $\beta$  et Ala292 de H3. Cette feuille  $\beta$  n'a pas été trouvée en interaction avec l'agoniste complet, alors que nous suggérons que l'interaction de la feuille  $\beta$  pourrait être responsable de la compensation du manque de stabilisation de H12. En résumé, la molécule CGA forme des liaisons hydrogène avec Ser342, un acide aminé lié à la liaison de l'agoniste partiel. Il interagit également avec Ile341, Ser289, Tyr327 et Ile281. Sur la base de son interaction avec Ser342 et de l'absence d'interactions directes avec d'autres acides aminés associés à l'agoniste complet, l'acide chlorogénique est probablement un agoniste partiel (Peng et al., 2018). L'acide chlorogénique (CGA), appartient à la famille des acides hydroxycinnamiques et est formé par l'estérification de l'acide cafétique et de l'acide quinique (Wang et al., 2022), où il a été démontré que le CGA exerce des effets hypoglycémiants, hypolipidémiants, antibactériens, antioxydants et anti-inflammatoires (Yan et al., 2020; Li et al., 2009; Lou et al., 2011; Wu et al., 2007). Une étude récente a montré que l'acide chlorogénique et ses dérivés peuvent améliorer avec succès l'hyperglycémie et l'hyperinsulinémie et améliorer la fonction du pancréas, et pourraient être considérés comme un médicament prometteur pour le traitement du diabète en restaurant efficacement la fonction pancréatique (Li et al., 2009). De plus, le CGA a été démontré pour moduler le métabolisme du glucose et des lipides *in vivo*, à la fois chez les individus sains et chez ceux atteints de troubles métaboliques (Meng et al., 2013). La large distribution et les remarquables activités pharmacologiques de ces acides phénoliques naturels indiquent leur potentiel dans la

découverte et le développement de nouveaux médicaments naturels (Wang et al., 2022). Le CGA est produit dans les plantes par la voie de l'acide shikimique pendant la respiration aérobie. On peut le trouver dans la plante médicinale marocaine *l'Anabasis arctioides* et le *Coriandrum sativum L* (*Anabasis arctioides* Coss. & Moq.; Msaada et al., 2020; Wang et al., 2022). De plus, le CGA a été identifié comme un agoniste potentiel de PPAR $\gamma$ , similaire à la rosiglitazone, et a été démontré pour stimuler l'expression de PPAR $\gamma$ , en faisant un agent sensibilisant potentiel à l'insuline et un agent hypolipidémiant (Sanchez et al., 2017; Peng et al., 2018). Une étude menée en 2017 a suggéré que le CGA a des propriétés physicochimiques appropriées pour être considéré comme une molécule bioactive principale pour le développement de nouveaux agents avec des propriétés similaires (Sanchez et al., 2017). L'étude a également révélé que le CGA pourrait se lier à PPAR $\gamma$  et pourrait activer son expression (Sanchez et al., 2017). Le composé 3,4-DICQA forme des liaisons hydrogène avec Tyr473, Ser289 et His449, qui sont tous liés à la liaison de l'agoniste complet. De plus, il interagit avec Met364 et Tyr327. Compte tenu de ses interactions avec des acides aminés généralement associés à des agonistes complets, l'acide 3,4-dicafféoylquinique pourrait également être un agoniste complet (Brust et al., 2018).

Les simulations MD ont confirmé les études de docking puisque toutes les molécules testées ont tendance à rester liées à PPAR $\gamma$ . 3,4-DICQA, l'acide chlorogénique (CA) ont montré les meilleurs résultats dans les simulations MD et les calculs MM-PBSA, indiquant qu'ils ont une meilleure stabilité et pourraient être de puissants modulateurs contre PPAR $\gamma$  (Pronk et al., 2013; Wu et al., 2022). Cette étude a identifié trois phytochimiques à haut potentiel conducteur capables de se lier au site actif de PPAR $\gamma$ . L'ensemble des composés identifiés peut conduire à une solution thérapeutique contre le DT2 en ciblant efficacement le site de fonction de PPAR $\gamma$ . La rosiglitazone est un agoniste complet du récepteur activé par les proliférateurs de peroxysomes gamma (PPAR $\gamma$ ), et bien qu'elle soit efficace dans le traitement des troubles métaboliques, elle a été associée à un risque plus élevé d'effets indésirables par rapport aux agonistes partiels (Nissen et Wolski, 2007). De plus, des études ont suggéré que la rosiglitazone peut augmenter le risque de maladie cardiovasculaire (Nissen et Wolski, 2007; Psaty et al., 2007). Par conséquent, la découverte de nouveaux agonistes partiels du PPAR $\gamma$ , tels que l'acide chlorogénique et l'acide 3,4-dicafféoylquinique, pourrait offrir des alternatives thérapeutiques plus sûres et efficaces pour le traitement du diabète de type 2 et du syndrome métabolique.

## 5. Conclusion

Des extraits de plantes marocaines ont montré des activités antimicrobiennes, anticancéreuses et antidiabétiques. Cependant, les propriétés thérapeutiques de la plupart de ces composés n'ont pas encore été entièrement étudiées comme celles des composés indiens et chinois. Dans ce contexte, nous avons mené une recherche bibliographique pour extraire les composés marocains avec leurs

structures 3D afin de révéler leur impact moléculaire contre PPAR $\gamma$  en utilisant des approches avancées de découverte de médicaments assistées par ordinateur.

D'après les résultats du docking moléculaire et des simulations de dynamique moléculaire avec les calculs MM-PBSA, nous avons conclu que 3,4-DICQA interagit de manière similaire avec un agoniste complet et que l'acide chlorogénique, Lutéoloside se lie comme pourrait améliorer la sensibilité à l'insuline par l'intermédiaire de PPAR $\gamma$ . Par conséquent, nous suggérons que ces cinq phyto-composés pourraient être utilisés pour la modulation de PPAR $\gamma$  et pourraient être davantage modifiés et synthétisés pour développer des candidats médicamenteux potentiels contre le syndrome métabolique et surtout le T2DM.

## PARTIE 3. CONCEPTION DE LA PREMIÈRE BASE DE DONNÉES DES COMPOSÉS ISSUS DES PAMs

## 1. Introduction

L'extraordinaire diversité écologique et végétale du Maroc est principalement attribuée à sa situation géographique unique. Le pays abrite environ 4 500 espèces et sous-espèces de plantes vasculaires, dont un impressionnant 14,22% sont des espèces médicinales et 22% sont endémiques (Chaachouay et al., 2022)[1]. De plus, la diversité des écosystèmes du Maroc et le niveau élevé d'endémisme parmi ses espèces animales le distinguent des autres régions. En effet, le Maroc est considéré comme un point chaud mondial crucial pour la biodiversité, avec de nombreuses espèces exceptionnelles et irremplaçables trouvées uniquement dans cette région (Rankou et al., 2013). Les plantes médicinales ont historiquement été reconnues pour leur rôle significatif dans les soins de santé, offrant des remèdes pour divers maux tels que les maladies cardiovasculaires, l'inflammation et les systèmes immunitaires affaiblis (Nunes et al., 2020). Les phytochimiques ont démontré leur potentiel à réguler les activités biologiques, comme illustré par l'aspirine, reconnue pour ses propriétés thérapeutiques, y compris son potentiel antioxydant (Liu, 2003). La consommation régulière de régimes riches en phytochimiques, comprenant des fruits, des légumes et des céréales complètes, peut réduire le risque de maladie associé aux dommages oxydatifs (Liu, 2003). La large acceptation et la compatibilité des phytochimiques avec le corps humain, leurs caractéristiques structurelles uniques et leurs effets secondaires minimes contribuent à leur utilisation substantielle (Kumar et al., 2023). De plus, une proportion significative de médicaments pharmaceutiques est dérivée d'extraits de plantes (Newman & Cragg, 2020). Les structures diverses et les activités biologiques des phytochimiques captivent à la fois les scientifiques de laboratoire et computationnels, servant de modèles pour développer des petites molécules et des médicaments ; ils sont reconnus comme des structures privilégiées pour interagir avec les cibles protéiques des médicaments (Driggers et al., 2008). Leurs caractéristiques uniques et leur diversité structurelle continuent d'émerveiller les scientifiques dans le développement de médicaments inspirés de la nature (Driggers et al., 2008). Des méthodologies computationnelles ont été intégrées dans le pipeline de découverte de médicaments pour aider à l'identification et à l'élucidation de ces composés uniques, en se concentrant sur l'identification de motifs et de structures moléculaires pour guider la conception combinatoire et la sélectivité cible (Leelananda & Lindert, 2016). Au fil des ans, des disciplines telles que la chémoinformatique et la bioinformatique ont apporté d'importantes contributions à la découverte de médicaments à base de produits naturels (Arul Murugan et al., 2022). L'intégration de stratégies computationnelles, y compris le criblage virtuel et l'apprentissage automatique, a propulsé la recherche sur les phytochimiques, améliorant notre compréhension de ces composés. Par conséquent, ces avancées dans les approches computationnelles ont mis en évidence le potentiel des phytochimiques en tant que sources précieuses de médicaments modernes, ouvrant la voie au développement de nouveaux traitements efficaces (Sofowora et al., 2013). L'objectif de ce travail est de mettre en place une base

de données en ligne et gratuite dédiée aux plantes médicinales et aromatiques marocaines, une ressource complète pour les études de découverte de médicaments utilisant les phytochimiques marocains. Cette base de données permettra aux utilisateurs et aux chercheurs d'identifier les phyto-composés des plantes marocaines et d'accéder à leurs structures 3D, essentielles pour la modélisation moléculaire avancée et le criblage virtuel. En centralisant ces données, nous visons à optimiser l'exploration des phytochimiques marocains, favorisant la collaboration et mettant en évidence les bénéfices thérapeutiques potentiels de la biodiversité végétale du Maroc.

## 2. Méthodes

### 2.1 CONCEPTION DU SERVEUR ET DE LA BASE DE DONNÉES

MPDB est une plateforme de base de données gérée par le logiciel de gestion de base de données open source appelé MySQL (<http://www.mysql.com>). Le site web de la plateforme a été développé en utilisant PHP (<http://www.php.net/>) comme langage de programmation côté serveur, et JavaScript (<http://www.javascript.com/>) et Ajax comme langages de script côté client pour créer des interfaces conviviales. Pour garantir la sécurité de la plateforme, nous avons utilisé <https://www.openssl.org/> pour générer un certificat de sécurité qui crypte toutes les communications entre nos utilisateurs et le serveur de l'application.

### 2.2 COLLECTE DES COMPOSÉS

Les informations sur les composés ont été récupérées à partir d'articles de journaux et de thèses. Le nom de chaque plante et composé a été extrait, ainsi que toute information disponible sur l'activité du composé ou de la plante. Cependant, la notation SMILES pour certains composés n'était pas facilement disponible dans les publications. Pour remédier à cela, la base de données PubChem a été utilisée pour obtenir les SMILES et les structures 3D de chaque composé au format sdf. Le programme Open Babel (<http://openbabel.org>) a ensuite été utilisé pour générer divers fichiers 3D tels que pdbqt, mol2, et pdb. De plus, un ID PubChem a été ajouté aux données extraites si disponible. Pour créer des images de structure 2D de chaque composé, l'outil web Smi2Depict disponible à l'adresse <http://cdb.ics.uci.edu/cgi-bin/Smi2DepictWeb.py> a été utilisé.

### 2.3 RECHERCHE D'ANALOGUES

Pour identifier les analogues de chaque entrée dans la MPDB, nous avons utilisé la base de données Mcule (Kiss, Sandor, et Szalai 2012) et extrait un ensemble de composés au format SMILES, que nous avons téléchargé. Nous avons ensuite calculé les coefficients de Tanimoto ou scores de similarité en utilisant le logiciel OpenBabel (O'Boyle et al. 2011). Un seuil de coefficient de Tanimoto de 0,8 ou plus a été utilisé pour identifier les analogues. Nous avons intégré l'ID de l'anologue ainsi que leurs liens Mcule respectifs dans le MNPDB, comme illustré dans la Figure 3.

## 2.4 BASE DE DONNÉES RÉFÉRENCEÉ

Le site web de la MPDB fournit une liste complète de publications où les phyto-composés ont été identifiés, y compris le titre et le DOI de l'article de journal ou un lien direct vers la source. Étant donné que la base de données a été construite par extraction manuelle des informations à partir d'articles de revues, il a été jugé essentiel de maintenir un lien fort entre les références et les données obtenues à partir de celles-ci.

## 2.5 ANALYSE CHIMIO-INFORMATIQUE

La règle des cinq de Lipinski, un ensemble de directives largement accepté, s'est avérée être un outil efficace pour déterminer la ressemblance d'un composé avec un médicament. Cette règle prend en compte des paramètres physico-chimiques tels que le poids moléculaire, les donneurs et accepteurs de liaisons hydrogène, et le coefficient de partition octanol-eau calculé ( $c \log P$ ). Dans cette étude, nous avons évalué la conformité des composés dans la MPDB avec la règle de Lipinski.

## 2.6 VÉRIFICATION DES DONNÉES

Nous avons vérifié rigoureusement l'exactitude de notre ensemble de données de phyto composés à travers un processus manuel méticuleux. Nous avons comparé la formule chimique et la notation SMILES de chaque phytocomposé avec des sources fiables et comparé nos résultats avec la base de données PubChem. Nous avons minutieusement enquêté sur toute incohérence ou erreur dans notre ensemble de données et avons atteint un haut niveau de confiance en son exactitude et son intégrité. Ce processus de vérification méticuleux est essentiel pour la réussite des futures recherches dans le domaine des phyto-composés marocains, et notre ensemble de données est désormais une ressource fiable pour de telles recherches.

## 2.7 CLASSIFICATION PHYTO-COMPOSÉ

La classification des composés est utile pour évaluer leur diversité, et peut être réalisée selon différents schémas. Ici, ClassyFire, qui réalise une classification hiérarchique en utilisant des motifs structurels en royaumes, superclasses, classes et sous-classes, a été utilisé [Fantoukh OI, Dale OR, Parveen A et al (2019) Évaluation de la sécurité des phyto-composés dérivés du thé Rooibos sud-africain globalisé (*Aspalathus linearis*) à travers l'interaction avec CYP, PXR, et P-gp. *J Agric Food Chem* 67:4967–4975. <https://doi.org/10.1021/acs.jafc.9b00846>].

## 2.8 ACTIVITÉS DES COMPOSÉS

Dans cette étude, nous rapportons la distribution des activités biologiques pour un total de 490 phyto composés, tel que déterminé à partir de la référence source, où l'activité biologique significative des composés isolés a été déterminée, l'information a été enregistrée.

### 3. Résultats

#### 3.1 CONSTRUCTION ET CONTENU

Notre plateforme est accessible via le nom de domaine <https://mpdb.org/> et peut être consultée à l'aide de n'importe quel navigateur web. Le schéma de la base de données (comme illustré à la Figure 9) est principalement axé sur les tables des noms de plantes, des phyto-composés, et des références. Ces tables stockent des informations relatives aux composés, à leurs sources, et à leurs utilisations documentées. La table des sources fait référence aux organismes à partir desquels un composé donné a été isolé, tandis que la table des utilisations contient des informations sur toute activité spécifique qui a été testée pour un composé donné (comme l'activité antipaludique, antidiabétique, ou autres). Cela permet aux chercheurs et aux utilisateurs de rechercher et de filtrer facilement les données pour trouver l'information dont ils ont besoin.

Globalement, MPDB est une plateforme de base de données puissante et conviviale qui facilite l'accès et le travail avec les données chimiques. Son utilisation de logiciels libres et de cryptage assure la sécurité des données des utilisateurs, tandis que son schéma flexible et son interface en font un outil précieux pour les chercheurs et les professionnels de la chimie.

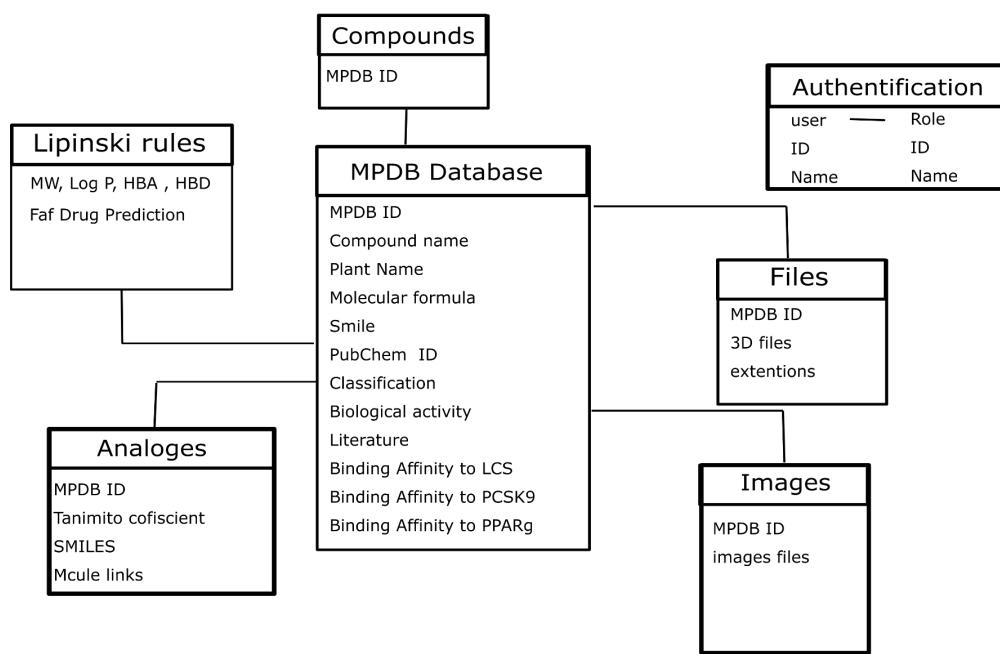


FIGURE 48. DISPOSITION SIMPLIFIÉE DU SCHÉMA DE LA BASE DE DONNÉES MPDB. LES LIENS ENTRE LES TABLES REPRÉSENTENT DES RELATIONS DE PLUSIEURS À PLUSIEURS.

The screenshot shows the MNPs Database homepage. At the top, there's a navigation bar with the MNPs Database logo, a Communication Center link, and a 'MNPs DATABASE' button. Below the navigation bar is a search bar with placeholder text: 'Explore Chemistry of Moroccan plants compounds : Quickly find chemical information and in silico results against cardiovascular targets'. A text input field contains the query 'Try Eugenol, Arbutus unedo, Antidiabetic activity, CC(=O)OC1CC2CCC1(C2(C)C), C15H24, Binding affinity to PPAR, Phenolic, PubChem ID'. Below the search bar is a green 'Filter' button. Underneath the search bar is a table with columns for 'ID MNPDB', 'Plant name', 'Compound name', and 'Actions'. A note says 'Please use the filter button to show results'.

**FIGURE 49. CAPTURE D'ÉCRAN DE LA PAGE DU MENU PRINCIPAL. LA RECHERCHE DE PHYTO-COMPOSÉS PEUT ÊTRE EFFECTUÉE AVEC DIFFÉRENTS MOTS CLÉS COMME ILLUSTRÉ.**

The screenshot shows the detailed view page for compound MNPDB00156. The top navigation bar includes the MNPs Database logo, a Communication Center link, and a 'Consultation des MNPs' button. The main content area is titled 'Dysokusone D'. It displays various details about the compound, including its ID (MNPDB00156), name (Dysokusone D), source (Thymus maroccanus), and chemical structure. The chemical structure is shown as a 2D diagram with atoms labeled CH<sub>3</sub>. Below the structure, there's a table with columns for MW, Log P, HBA, HBD, Lipinski violation, Solubility Forecast Index, Oral bioavailability, and FaF Dr predic. The values listed are 114.21, 2.16, 0, Good, 0, 0, Good, and Acceptable. To the right of the table is a '2D Structure' section with the chemical diagram. Other sections include 'Chemical and physical properties & FAF Drugs', 'Pubchem id', 'Source Link', and 'Literature'.

**FIGURE 50. CAPTURE D'ÉCRAN DE LA PAGE DE RÉSUMÉ DU COMPOSÉ. L'INTERFACE DU SITE WEB MPDB EST MONTRÉE, AVEC UNE CAPTURE D'ÉCRAN DE LA PARTIE DE LA PAGE DE RÉSUMÉ DU COMPOSÉ POUR L'ENTRÉE MNPDB00156.**

### 3.2 ANALOGUES DES PHYTO-COMPOSÉS

Ce processus de collection d'analogues nous a permis d'étendre la portée de la MNPDB en incorporant des analogues des composés marocains déjà présents dans la base de données. En incluant des analogues, nous avons pu identifier des pistes potentielles supplémentaires pour la découverte de médicaments et élargir davantage les applications potentielles de la MNPDB. Les coefficients de Tanimoto ont servi de mesure de la similarité entre les composés, nous permettant d'identifier ceux ayant le plus grand potentiel pour une enquête plus approfondie.

### 3.3 BASE DE DONNÉES RÉFÉRENCEÉE

Cette fonctionnalité permet au site web de fonctionner comme un outil de recherche précieux, spécifiquement dédié aux phyto-composés marocains. Si un utilisateur découvre un composé

pertinent, il peut naviguer de manière fluide vers la littérature disponible sur ce phyto-composé spécifique. Le site web offre également une opportunité à grande échelle d'explorer la littérature décrivant les phyto-composés marocains ayant une activité particulière. Cette ressource précieuse facilite non seulement l'accès à l'information, mais elle augmente aussi la visibilité des recherches menées par les chercheurs qui ont isolé ou travaillé avec ces composés. Un autre avantage de maintenir des informations entièrement référencées est la capacité de localiser des composés au sein d'une publication, ce qui peut parfois s'avérer difficile.

### 3.4 CLASSIFICATION DES PHYTO-COMPOSÉS

Les lipides prérol se distinguent par leur présence dominante, représentant environ 41,90% du total des composés phytochimiques répertoriés dans la base de données MPDB. Ces lipides prérol jouent un rôle indispensable dans un éventail de processus biologiques, agissant comme acteurs clés dans la signalisation cellulaire et orchestrant la régulation de diverses voies métaboliques. Dans l'ensemble de données MPDB, les flavonoïdes représentent environ 7,75%, démontrant leur prévalence en tant que pigments naturels au sein du règne végétal. Reconnus pour leurs puissantes propriétés antioxydantes et leurs attributs potentiels bénéfiques pour la santé, les flavonoïdes se présentent comme des composés bioactifs aux applications multiples.

Le domaine des alcaloïdes tropaniques, constituant environ 3,76%, émerge comme une autre catégorie fascinante de composés phytochimiques. Avec des effets pharmacologiques prononcés, leur présence dans la base de données MPDB pourrait captiver les chercheurs s'engageant dans le domaine complexe de l'exploitation des offrandes de la nature pour le développement de nouveaux médicaments. Les composés organo-oxygénés contribuent à environ 3,64%, tandis que les acyles gras représentent environ 3,29% du paysage phytochimique. Le benzène et ses dérivés substitués comptent pour environ 2,46%, et les composés sans classe spécifiée représentent 2,11%. Les acides carboxyliques et leurs dérivés contribuent à 1,88%, et les cannabinoïdes représentent 1,76%. Enfin, la catégorie "Autres", qui comprend l'anthracène, les lignanes d'aryltétraline, les benzopyranes, les acides cinnamiques et leurs dérivés, les coumarines et leurs dérivés, les lignanes de dibenzylbutane, le dihydropyrane, les époxydes, le lignand furanoïde, les furanes, les furofuranes, les indoles et leurs dérivés, les lactones, les oxanes, les peptidomimétiques, les phénanthrènes et leurs dérivés, les esters phénoliques, le phénol, les hydrocarbures polycycliques, les pyranes, les quinoléines et leurs dérivés, les hydrocarbures saturés, les stéroïdes et leurs dérivés, les tétrahydrofuranes, le tétraline et les hydrocarbures insaturés, représente 27,51% du total des composés phytochimiques répertoriés dans la base de données MPDB. Ces résultats indiquent une gamme diversifiée de composés ayant le potentiel pour diverses activités biologiques. La répartition des classes de composés est visible sur la Figure 51.

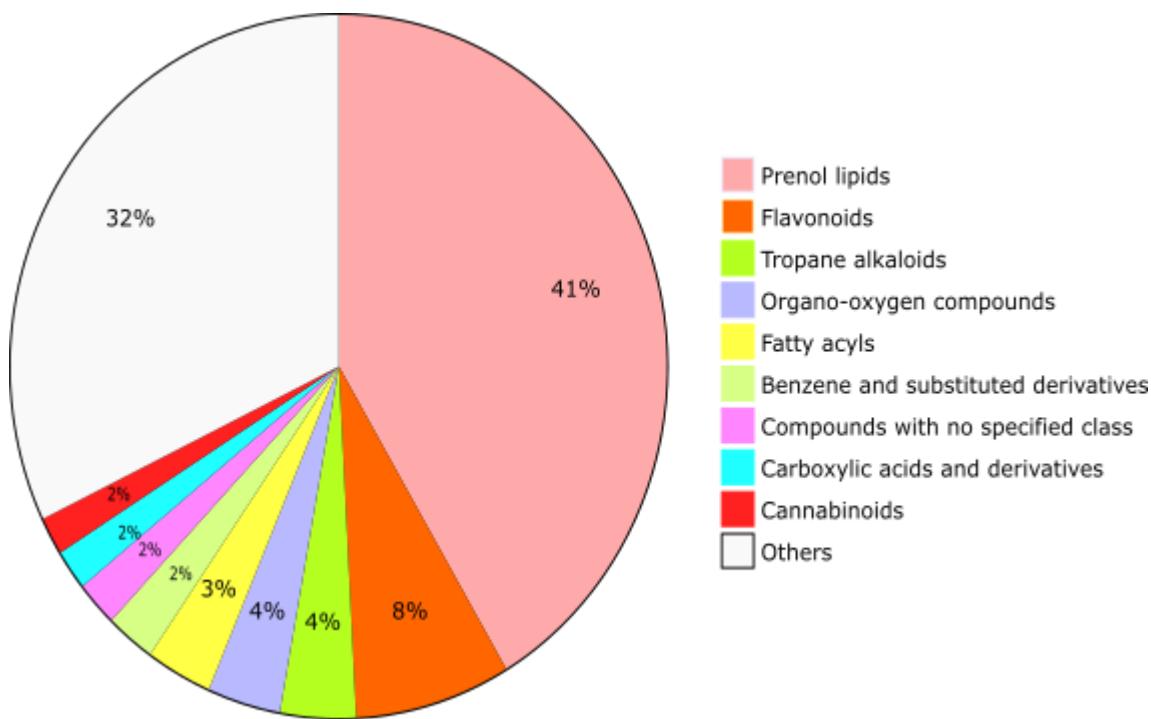


FIGURE 51. CLASSIFICATION DES PHYTO-COMPOSÉS DE LA MPDB. LA CLASSIFICATION A ÉTÉ OBTENUE À PARTIR DE CLASSYFIRE.

### 3.5 ANALYSE CHÉMOINFORMATIQUE

Comme illustré sur la Figure 52, notre analyse offre des perspectives sur la manière dont les composés de la MPDB sont en accord avec la Règle des Cinq de Lipinski. Un remarquable 64% des composés répondent entièrement aux critères, indiquant qu'ils possèdent probablement des propriétés favorables à une administration orale de médicaments. D'un autre côté, environ 27% des composés présentent une seule violation, suggérant un léger écart par rapport aux propriétés idéales, mais cela ne diminue pas automatiquement leur potentiel en tant que candidats médicamenteux. Une fraction plus petite, environ 3%, présente deux violations. De manière intéressante, bien que 5% des composés montrent trois violations, s'écartant des directives de Lipinski, ils présentent toujours un cadre structural qui pourrait inspirer les chimistes à concevoir de meilleurs médicaments. Ces données sont inestimables pour les initiatives de découverte de médicaments, guidant la priorisation des composés en fonction de leur potentiel en tant que médicaments administrés par voie orale. La base de données MPDB se distingue comme un atout précieux pour la découverte de médicaments, offrant un vaste éventail de composés potentiellement médicamenteux. Cependant, une exploration

plus approfondie du potentiel pharmacologique et thérapeutique de ces composés est justifiée dans les recherches ultérieures.

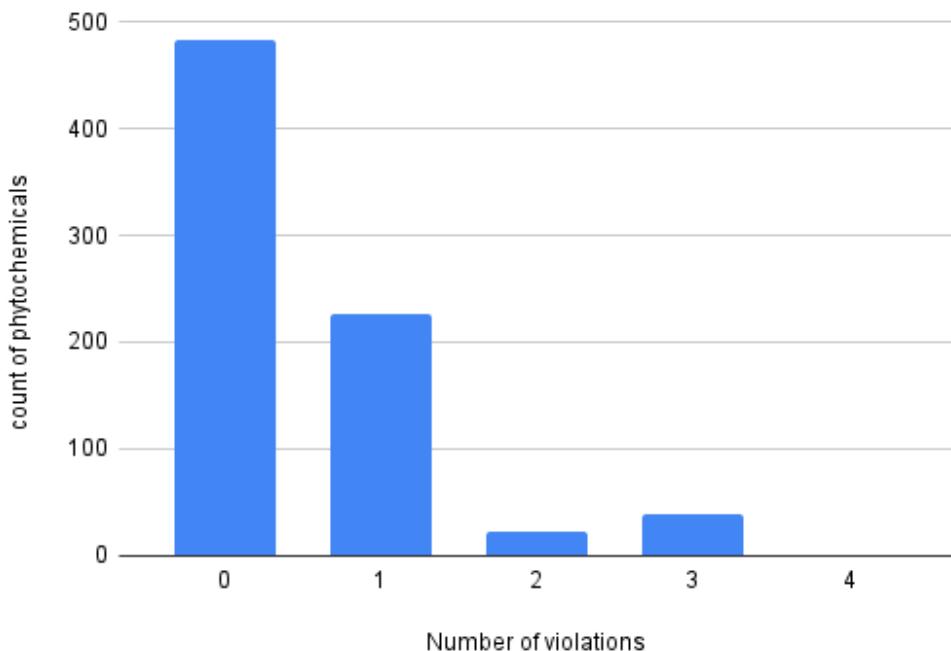


FIGURE 52. NOMBRE DE VIOLATION DES PROPRIÉTÉS PHYSICOCHIMIQUES DES PHYTO-COMPOSÉS DE LA BASE DE DONNÉES, BASÉES SUR LA "RÈGLE DES CINQ" DE LIPINSKI.

### 3.6 ACTIVITÉS DES COMPOSÉS

Dans cette étude, nous présentons la répartition des activités biologiques des phytocomposés de la MPDB, telles que déterminées à partir de la référence source. Nos résultats montrent que les activités biologiques les plus couramment observées sont antibactériennes (17,40%), antimicrobiennes (17,16%) et antioxydantes (16,18%). De plus, nous avons également observé des activités antidiabétiques (4,53%), anticandidales (2,08%), cytotoxiques (2,21%), anti-inflammatoires (5,64%) et induisant l'arrêt du cycle cellulaire et l'apoptose (3,80%). D'autres activités, représentant 4,18%, incluent les effets vasorelaxants, antifongiques, insecticides, antiviraux, analgésiques, antiprolifératifs, sédatifs et psychoactifs (Figure 52). Il est à noter que notre analyse révèle que 38,24% des enregistrements de composés n'ont pas d'informations sur leurs activités biologiques, soulignant la connaissance limitée actuellement disponible sur l'activité phytochimique de ces composés. Ces découvertes soulignent la nécessité d'approfondir les recherches sur les activités biologiques des

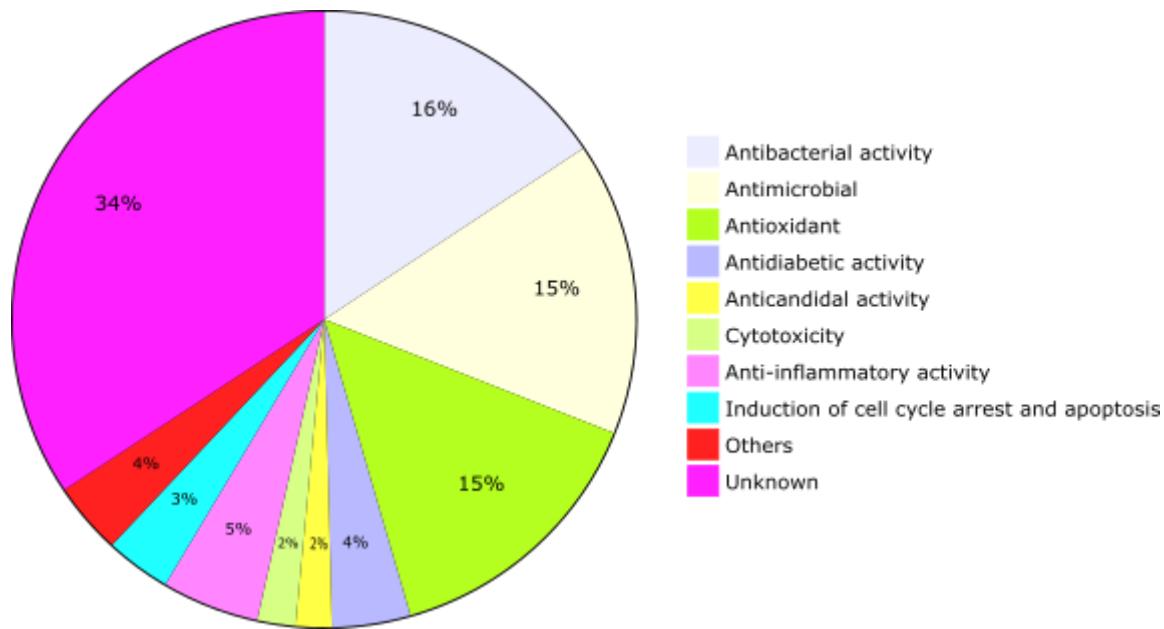


FIGURE 53. ACTIVITÉS BIOLOGIQUES DES PHYTO-COMPOSÉS DE MPDB. LE DIAGRAMME CIRCULAIRE MONTRANT LA DISTRIBUTION DES DIFFÉRENTES ACTIVITÉS BIOLOGIQUES DES PHYTO-COMPOSÉS DONT L'UTILISATION EST DOCUMENTÉE DANS LEURS RÉFÉRENCES DANS MPDB.

#### 4.Discussion

La Base de Données Phytochimiques Marocaine (MPDB) sert de ressource complète pour ceux qui s'intéressent au potentiel thérapeutique des phytochimiques des plantes marocaines. Des recherches ont montré que les plantes marocaines offrent un large éventail d'activités biologiques, notamment antimicrobiennes (El Abdali et al., 2023), anticancéreuses (Alami Merrouni & Elachouri, 2021), antidiabétiques (Bouyahya et al., 2021), et cardiovasculaires (Goldstein & Brown, 1990). La MPDB offre un accès à une collection soigneusement organisée de phyto-composés issus de ces plantes, ce qui en fait un outil essentiel pour les chercheurs en découverte de médicaments, médecine traditionnelle et phytochimie. La base de données contient des informations sur plus de 700 composés uniques, catégorisés par leurs motifs structuraux. Un segment notable de la base de données est dédié aux lipides prérol, reconnus pour leurs divers rôles biologiques et leur potentiel thérapeutique (Goldstein & Brown, 1990). Ces lipides jouent des rôles essentiels dans les processus cellulaires, tels que le maintien de l'intégrité de la membrane cellulaire, la modification des protéines et la transduction du signal. Un exemple notable de lipide prérol, le limonène, abondamment présent dans *Ammodaucus leucotrichus* (Muckensturm et al., 1997), le zeste de *Citrus limon* (Öntaş et al., 2016) et *Origanum grosii* (Bakha et al., 2019), est reconnu pour ses activités chémopréventives et chimiothérapeutiques contre divers cancers (Sun, 2007). De plus, le limonène présente des propriétés antimicrobiennes (Chee, Kim, & Lee, 2009) et possède des effets anti-inflammatoires (d'Alessio, Mirshahi, Bisson, & Bene, 2014). Les flavonoïdes sont également bien représentés, constituant 7,75%

de la base de données. Ces composés polyphénoliques, couramment trouvés dans les plantes, sont connus pour leurs bienfaits thérapeutiques (Harborne & Williams, 2000). La présence significative de plantes riches en flavonoïdes dans la MPDB souligne l'importance thérapeutique de ces composés. De nombreuses études ont souligné les propriétés antioxydantes des flavonoïdes (Chen et al., 2020), attribuant leurs bienfaits pour la santé à leur capacité à neutraliser les radicaux libres et à contrer le stress oxydatif. Leurs effets anticancéreux, anti-inflammatoires (Middleton, Kandaswami, & Theoharides, 2000) et antimicrobiens (Yuan et al., 2021) sont également bien établis. Les alcaloïdes tropaniques, représentant 3,76% de la base de données, se distinguent par leur squelette tropane unique, une structure à sept membres contenant de l'azote. Ces composés se trouvent principalement dans les plantes de la famille des Solanaceae, telles que Datura (*Datura stramonium*) et Henbane (*Hyoscyamus albus*). Parmi ceux-ci, l'atropine (McLendon & Preuss, 2023), la scopolamine (Riad & Hithe, 2023) et l'hyoscyamine (LiverTox, 2012) ont des avantages pharmacologiques significatifs. L'atropine est principalement utilisée pour traiter la bradycardie et comme agent pré-anesthésique, tout en servant également d'antidote contre l'empoisonnement cholinergique. La scopolamine est reconnue pour son efficacité contre le mal des transports et a des rôles potentiels dans le traitement des troubles du SNC. L'hyoscyamine offre un soulagement des troubles gastro-intestinaux et urinaires grâce à ses propriétés antispasmodiques. La base de données présente également une variété de composés. Les acyles gras, y compris les acides gras oméga-3 bénéfiques pour le système cardiovasculaire, représentent 3,29% (Harris, 2007). Les acides carboxyliques et leurs dérivés, cruciaux dans de nombreux processus biologiques, représentent 1,88% (Vane & Botting, 2003). Les cannabinoïdes, représentant 1,76%, ont récemment attiré une attention significative pour leurs rôles thérapeutiques potentiels, en particulier dans la gestion de la douleur et la neuroprotection (Russo, 2011). La décision récente du Maroc de légaliser le cannabis à des fins médicales, cosmétiques et industrielles reconnaît non seulement sa valeur thérapeutique, mais prépare également le terrain pour son utilisation mondiale élargie. Argania spinosa, ou l'arganier originaire du Maroc, est une mine de composés bioactifs (Charrouf & Guillaume, 2008). L'huile d'argan, extraite de ses graines, a été traditionnellement utilisée à des fins culinaires, cosmétiques et médicinales (Khallouki et al., 2003). Riche en acides gras insaturés comme l'acide oléique et linoléique, l'huile d'argan est connue pour ses propriétés anti-inflammatoires et protectrices du système cardiovasculaire (El Monfalouti, Guillaume, Denhez, & Charrouf, 2010). De plus, l'huile d'argan contient une teneur élevée en tocophérols, notamment le gamma-tocophérol, qui sont reconnus pour leurs propriétés antioxydantes qui protègent les cellules des dommages oxydatifs (Bennani et al., 2007). La teneur en polyphénols de l'huile amplifie davantage ses activités antioxydantes, anti-inflammatoires et antiprolifératives (Berrougui et al., 2003). La MPDB est méticuleusement conçue pour soutenir les techniques de découverte de médicaments assistées par ordinateur, telles que le docking moléculaire (Jones, Willett, Glen, Leach, & Taylor, 1997) et le criblage

virtuel (Kitchen, Decornez, Furr, & Bajorath, 2004). Les structures 3D précises de la base de données permettent aux chercheurs de prévoir les interactions thérapeutiques potentielles avec les protéines cibles (Walters, Stahl, & Murcko, 1998). Le criblage virtuel est devenu un pilier de la découverte de médicaments, facilitant l'identification de composés bioactifs à partir de bases de données vastes et accélérant ainsi la trajectoire de développement des médicaments.

## 5. Conclusion

La MPDB est une contribution significative au domaine de la découverte de médicaments, en particulier pour les chercheurs intéressés par le potentiel thérapeutique des produits chimiques végétaux marocains. La curation manuelle de la base de données à partir de la littérature scientifique garantit des données de haute qualité et fiables, ce qui en fait une ressource précieuse pour les efforts de découverte de médicaments.

## PARTIE 4. EXPLOITATION DES DONNÉES DE LA MPDB VIS À VIS LA PCSK9

## 1. Introduction

En tant que première étape pour explorer le potentiel des phyto-composés marocains de la MPDB, nous avons réalisé un criblage virtuel contre les cibles thérapeutiques des maladies cardiovasculaires. La maladie cardiovasculaire athérosclérotique est une condition bien connue caractérisée par une dysfonction endothéliale, qui est la première étape qui conduit à l'athérogenèse. L'hypercholestérolémie est l'une des causes les plus courantes de dysfonction endothéliale, induisant la constriction des gros vaisseaux sanguins dans le cœur. Bien que le cholestérol soit nécessaire à de nombreux processus physiologiques, en particulier dans la membrane cellulaire, une augmentation de la quantité de cholestérol, spécifiquement sous des conditions oxydantes, peut conduire à l'athérosclérose, entraînant une maladie cardiaque carotide, périphérique et coronaire.

PCSK9 est une protéase qui peut dégrader LDLR en interagissant avec le domaine extracellulaire de LDLR. La liaison de LDL à LDLR est essentielle pour diminuer l'accumulation de cholestérol et le risque de maladie coronarienne. Cependant, la liaison de PCSK9 à LDLR empêche la liaison de LDL à LDLR et favorise l'accumulation de cholestérol. Par conséquent, il est essentiel d'inhiber PCSK9 et son association avec LDLR pour prévenir la maladie coronarienne. Les thérapies ciblées contre PCSK9 sont basées sur deux principales stratégies : les thérapies biologiques avec des anticorps monoclonaux et les pharmacothérapies à petites molécules. Cependant, la découverte d'une pharmacothérapie à petites molécules pour l'inhibition de PCSK9 est encore en cours. Par conséquent, il est nécessaire d'explorer les phyto-composés et les composés naturels puisqu'ils sont reconnus comme des structures privilégiées pour interagir avec les cibles protéiques des médicaments. Leurs caractéristiques uniques et leur diversité structurelle continuent d'étonner les scientifiques dans le développement de médicaments inspirés des produits naturels.

Ici, nous présentons la première exploitation des données de MPDB contre PCSK9 pour trouver des hits qui se lient à la cible avec une haute affinité. Nous avons récupéré la structure cristalline de PCSK9 de la base de données PDB (PDB ID 6U26) et préparé le site actif. Nous avons ensuite réalisé un docking moléculaire en utilisant le programme de criblage en source ouverte PyRx. Les composés retenus ont été sélectionnés sur la base de la plus faible affinité de liaison et de la valeur de déviation quadratique moyenne (RMSD). Plus l'énergie est négative, plus l'affinité de liaison avec les phyto-composés cibles est forte. Enfin, l'énergie de liaison de chaque phyto-composé a été intégrée à la base de données. Ces résultats peuvent être davantage explorés par des études plus approfondies *in silico* et *in vitro*. La Figure 4 illustre la méthodologie globale suivie dans ce travail.

Ce chapitre explique et détaille la méthodologie suivie pour le criblage virtuel des phyto-composés marocains en vue d'identifier des composés qui se fixent à la PCSK9 avec une haute affinité.

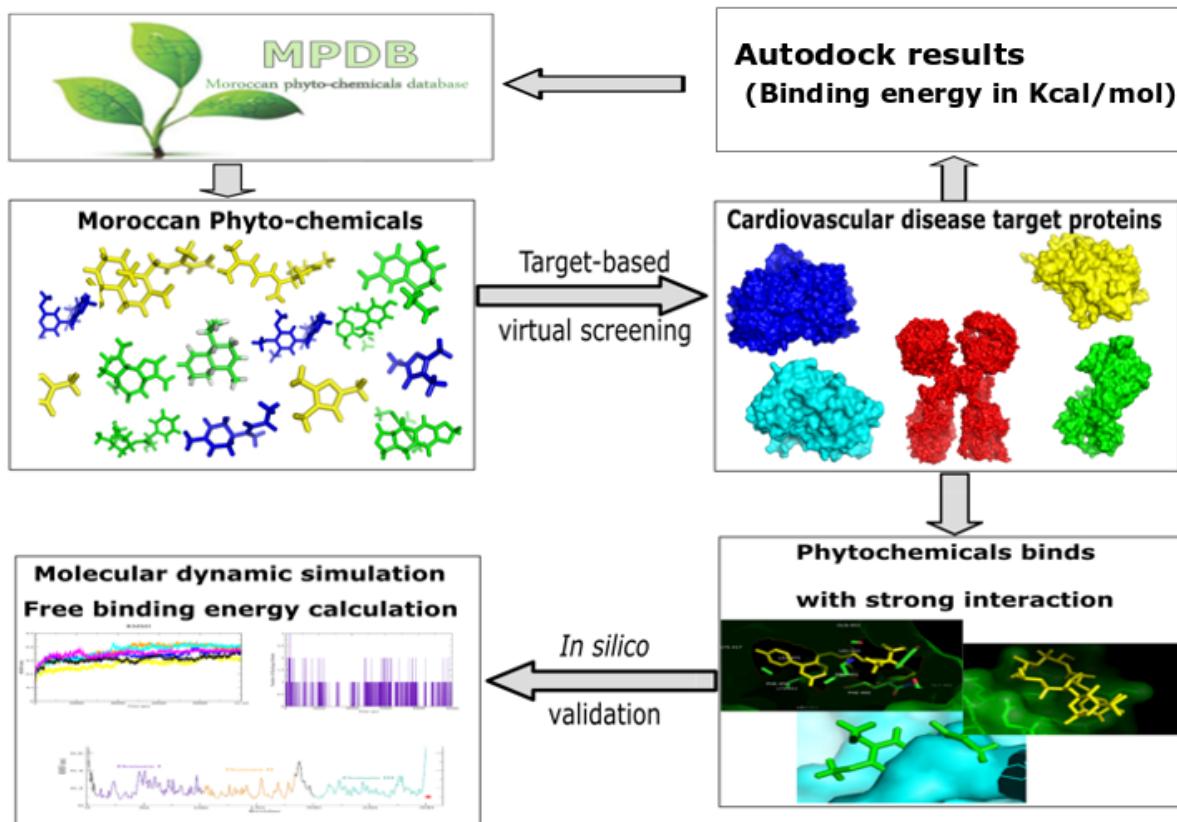


FIGURE 54. SCHÉMA DE LA MÉTHODOLOGIE SUIVI DANS LE CRIBLAGE VIRTUEL DES DONNÉES DE LA MPDB.

## 2. Méthodes

### 2.1 COLLECTE DES RENSEIGNEMENTS SUR LA CIBLE PCSK9

#### PDB

La PDB (Protein Data Base) est une base de données qui contient les structures tridimensionnelles de grandes molécules biologiques, comme les protéines et les acides nucléiques. Ces structures sont déterminées principalement par cristallographie aux rayons X, spectroscopie RMN ou microscopie électronique cryogénique, et sont fournies par des biologistes et des biochimistes du monde entier. La PDB est accessible gratuitement sur Internet à travers les sites web de ses membres (PDBe, PDBj, RCSB et BMRB). Dans cette étude, nous avons utilisé la PDB pour télécharger la structure 3D cristallisée de PCSK9.

#### UNIPROT

UniProt, ou "Universal Protein Resource", est une base de données centrale pour les informations sur les protéines. Elle est essentielle pour les chercheurs en biologie moléculaire et en bioinformatique, car elle offre un accès à une multitude de données annotées concernant la structure, la fonction, la localisation, et les interactions des protéines. Ces informations sont cruciales pour comprendre les fonctions biologiques et les mécanismes moléculaires sous-jacents. De plus, UniProt est régulièrement mis à jour, garantissant ainsi que les utilisateurs accèdent aux informations les plus récentes et les plus pertinentes. Dans le cadre de notre recherche, nous avons utilisé cette base de

données pour collecter toutes les informations concernant la PCSK9, nous permettant ainsi d'avoir une vue complète et détaillée sur la PCSK9.

#### DÉTERMINATION DU SITE ACTIF

La détermination de site actif est essentielle pour comprendre le mécanisme d'action et pour développer des inhibiteurs spécifiques. Le site actif d'une enzyme est la région où le substrat se lie et où a lieu la réaction catalytique. Pour la PCSK9, bien qu'elle n'ait pas de substrat traditionnel, son interaction avec le récepteur LDL (LDLR) est d'une importance primordiale. La région de la PCSK9 qui interagit avec le LDLR est souvent considérée comme son "site actif" fonctionnel. Des analyses de la structure 3D issus de l'expérimentale ont été utilisées pour déterminer précisément cette région de liaison. La structure cristalline de PCSK9 a permis d'identifier les sites de liaison de PCSK9 au domaine extracellulaire du LDLR.

N/A Fugure

#### FIGURE 55. : LE PRINCIPE DU UNIPROT ET MÉTHODE DE TRAVAIL.

##### PRÉPARATION DE LA CIBLE PCSK9

###### *PyMOL*

PyMOL est un outil puissant et largement utilisé dans le domaine de la modélisation moléculaire. Il offre aux chercheurs la capacité de visualiser, d'analyser et de manipuler des structures moléculaires en trois dimensions. L'une des principales forces de PyMOL réside dans sa capacité à produire des images de haute qualité des structures protéiques, ce qui est essentiel pour comprendre les interactions moléculaires, les sites actifs des enzymes et les domaines de liaison des ligands. De plus, il permet une analyse détaillée des interactions intermoléculaires, ce qui est crucial pour la conception de médicaments et la compréhension de l'interaction protéique. L'analyse structurale dans cette étude a été effectuée avec PyMOL afin d'identifier la surface d'interaction avec le LDLR .

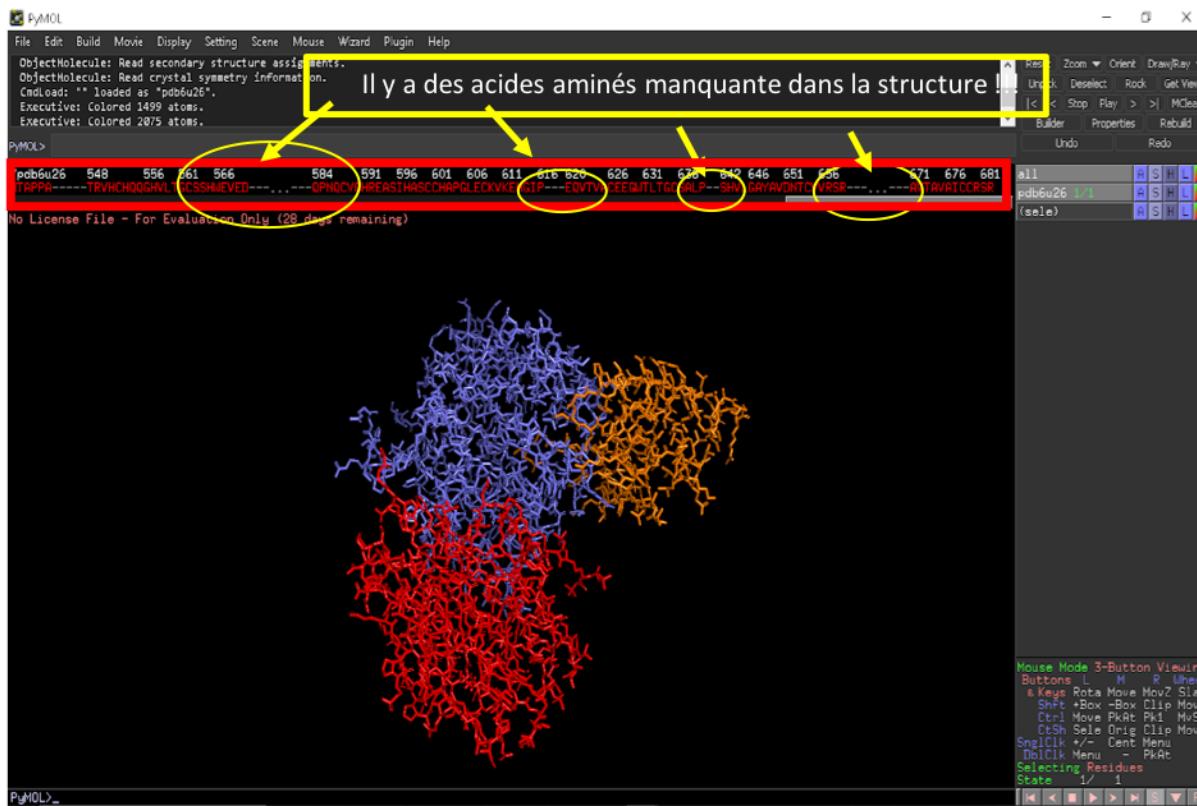


FIGURE 56. DES RÉSIDUS MANQUANTS DANS LA SÉQUENCE D'ACIDES AMINÉS DE LA PCSK9.

#### RECONSTRUCTION DES ACIDES AMINÉS MANQUANTS (6U26) DE PCSK9

Swiss model : est un serveur de modélisation homologique de structures protéiques entièrement automatisé, accessible via le serveur web Expasy, ou à partir du programme DeepView (Swiss Pdb-Viewer). Il y avait des acides aminés manquants durant l'expérimentation on va utiliser le logiciel Swiss model. L'objectif de ce serveur dans cette étude est non pas de construire la structure entière par homologie mais de compléter les résidus manquants de la structure 6U26 choisis en se basant sur la séquence primaire.

```
« MGTVSSRRSWWPLPLLLLLLGPAGARAQEQEDEDGYEELVLALRSEEDGLAEAPEHGTATFHRCAKDPWRLPGTYVVV
LKEETHLSQSERTARRLQAQAARRGYLTKLHVFHGLPGFLVKMSGDLLELALKLPHVDYIEEDSSVFAQSIPWNLERITPPRYR
ADEYQPPDGGSLEVYLLTSIQSDHREIEGRVMVTDFENVPEEDGTRFHRQASKCDSHGTHLAGVVSGRDAGVAKGASMR
SLRVLNCGKGTVSGTLIGLEFIRKSQQLVQPGPLVLLPLAGGYSRVLNAACQRLARAGVVLVTAAGNFRDDACLYSPASAPE
VITVGATNAQDQPVTLGLTNFGRCVDLFAPGEDIIGASSDCSTCFVSQSGTSQAAAHVAGIAAMMLSAEPELTLAELRQL
IHFSAKDVINEAWFPEDQRVLTPNLVAALPPSTHGAGWQLFCRTVWSAHSGPTRMATAVARCAPDEELLSCSSFSRSGKRRG
ERMEAQGGKLVCRAHNAFGGEGVYAIARCCLLPQANCSVHTAPPÆASMGTRVHCHQQGHVLTGCSSHWEVEDLGTHKP
PVLRPRGQPNCQCVGHREASIHASCCCHAPGLECKVKEHGIPAPQEQTVAEEGWTLTGCSALPGTSHVLGAYAVDNTCVRS
RDVTTGSTSEGAVTAICCRSRHLAQASQELQ ». 
```

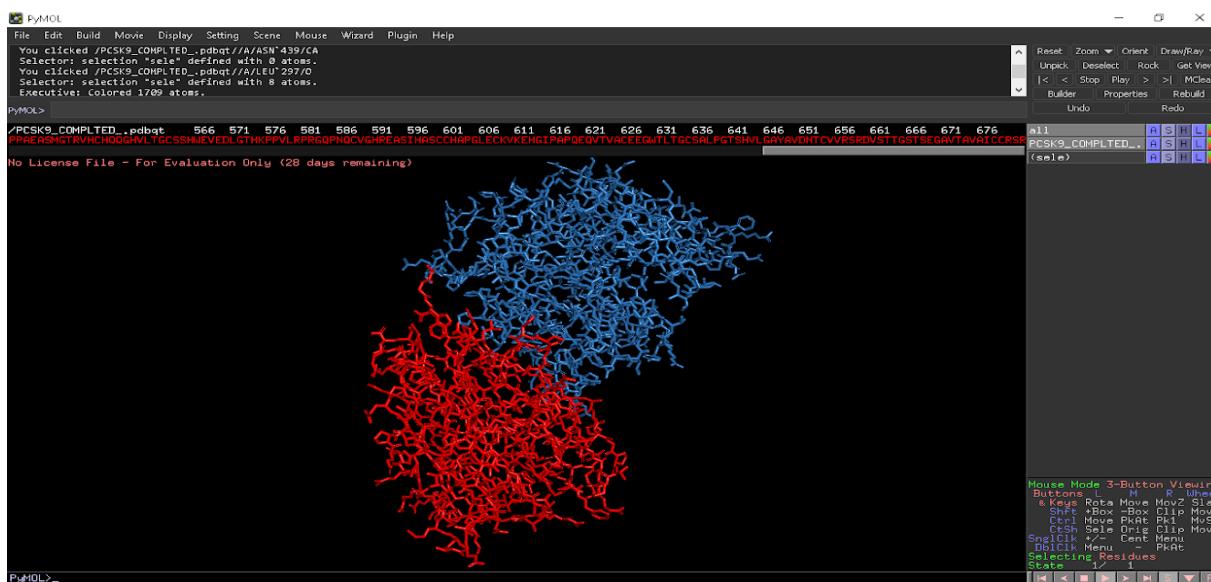


FIGURE 57. LA STRUCTURE DE PCSK9 APRÈS AVOIR COMPLÉTÉ LES ACIDES AMINÉ.

#### PRÉPARATION DE CHIMIOTHÈQUE

##### Collecte des structures 3D à partir de la base de recherche MPDB

##### *COLLECTE DES PHYTOCOMPOSÉS À PARTIR DE MPDB*

Dans cette partie, nous avons utilisé la MPDB pour télécharger les structures 3D des cannabinoïdes , La MPDB (Moroccan Phyto-Compounds DataBase).

The screenshot shows the MPDB interface with a search result for compound MNPDDB00199. The search bar at the top contains the identifier "MNPDDB00199". Below the search bar, there is a green button labeled "Filter". A red box highlights the text "CLIQUER sur details" (Click on details). A red arrow points from the top-left towards the search bar. Another red arrow points from the right towards the "Details" button next to the compound name "EXCELSIDE B". A third red box highlights the "Télécharger le fichier du phytocomposé format pdb" link at the bottom right of the detailed view.

FIGURE 58. INTERFACE DE LA MPDB DE TÉLÉCHARGEMENT DES STRUCTURES 3D DES PHYTO-COMPOSÉS MAROCAINS.

#### CHIMIOTHÈQUE

Cannabis est le nom d'une plante qui produit des substances psychoactives et médicinales, et parmi ces substances, il y a plus de 100 types différents de molécules organiques nommées cannabinoïdes, qui agissent sur le système nerveux et le cerveau.

**Ces composés sont classés en 11 types qui sont :**

TABLEAU 11. 100 LIGANDS PROVENANT DE L'ESPÈCE CANNABIS SATIVA ET 500 LIGANDS SONT DES PHYTOCOMPOSÉS ISSUS DES PLANTES MAROCAINES.

Composés chimiques	Nombre de composés
Type $\Delta^9$ -THC	18
Type $\Delta^8$ -THC	2

CBG	17
CBC	8
CBD	8
CBND	2
CBE	5
CBL	3
CBN	10
CBT	9
Type MISC	22
<b>TOTAL cannabinoïdes</b>	<b>94</b>
<b>TOTAL non cannabinoïdes</b>	<b>506</b>
<b>TOTAL</b>	<b>600</b>

Les phytocomposés sont des composés chimiques issus des plantes aromatiques marocaines naturelles, ils sont disposés dans la base de données MPDB.

Par suite on va utiliser **600 ligands** issus des espèces organiques en général, parmi ces composés on a **100 ligands provenant de l'espèce *Cannabis sativa*** et **500 ligands sont des phyto-composés issus des plantes marocaines**, Afin de révéler l'interaction entre la protéine PCSK9 et ces ligands naturels, on va réaliser une analyse de criblage.

#### *CRIBLAGE VIRTUEL BASÉ SUR LA STRUCTURE*

PyRx, également connu sous le nom d'AutoDock Tools, est un logiciel de conception assistée par ordinateur pour la modélisation moléculaire. Il est principalement utilisé pour la préparation et l'exécution de simulations de docking, une technique qui permet de prédire comment les petites molécules, comme les médicaments potentiels, se lient aux protéines et autres macromolécules biologiques. Cette capacité est essentielle pour comprendre les interactions moléculaires qui peuvent avoir des implications thérapeutiques. PyRx fournit une interface utilisateur graphique pour AutoDock Vina, facilitant ainsi la préparation des ligands et des protéines, la configuration des paramètres de docking et l'analyse des résultats. En tant qu'outil open source, PyRx est accessible à la communauté scientifique, permettant aux chercheurs de contribuer à son développement et de l'adapter à leurs besoins spécifiques.

#### *ANALYSE DE L'ÉNERGIE DE LIAISON*

PyRx est un logiciel de criblage virtuel qui analyse les interactions moléculaires en se basant sur différents profils d'interaction. Ces profils comprennent les interactions de van der Waals (vdW), les

interactions électrostatiques (Elec) et les liaisons hydrogène (HBond). Ces interactions jouent un rôle crucial dans la détermination de l'affinité entre une molécule et sa cible protéique. Après avoir effectué le criblage avec PyRx, les résultats sont généralement enregistrés sous la forme d'un fichier Excel (format csv). Pour analyser ces résultats, il est essentiel de filtrer les données en fonction de certaines affinités. Dans notre étude, nous avons choisi de retenir les molécules présentant une affinité inférieure à -9.1. Ce seuil a été déterminé en se basant sur l'affinité de la molécule référentielle après son criblage. Dans le fichier Excel, une étape de filtre est donc réalisée pour sélectionner uniquement les valeurs répondant à ce critère d'affinité, permettant ainsi d'identifier les molécules les plus prometteuses pour des études ultérieures.

#### ANALYSE DES INTERACTIONS LIGAND-PROTÉINE

L'analyse des interactions ligand-protéine est cruciale pour comprendre comment les molécules se lient et interagissent avec leurs cibles protéiques. Le logiciel Discover est un outil précieux dans ce domaine, permettant aux chercheurs de sonder en détail ces interactions. Avec Discover, il est possible de révéler les interactions de van der Waals, qui sont des forces attractives ou répulsives entre molécules qui ne sont pas dues à des liaisons covalentes ni à des interactions ioniques. Ces interactions jouent souvent un rôle essentiel dans la stabilisation de la liaison ligand-protéine. De plus, Discover permet également d'identifier les liaisons hydrogène, qui sont des interactions spécifiques entre un atome d'hydrogène et un atome électronégatif comme l'oxygène ou l'azote. Ces liaisons hydrogène sont cruciales pour la spécificité et l'affinité de la liaison, car elles peuvent fortement stabiliser l'interaction entre le ligand et la protéine. En utilisant Discover pour analyser les interactions ligand-protéine, les chercheurs peuvent obtenir une image détaillée de la manière dont un ligand se lie à sa cible protéique. Cela peut aider à la prédiction de leur efficacité et à la compréhension des mécanismes moléculaires sous-jacents de la liaison.

#### PREDICTION DE LA RESSEMBLANCE AUX MEDICAMENTS

FafDrugs est un outil en ligne essentiel pour les chercheurs en découverte de médicaments, permettant le filtrage et le formatage de chimiothèques pour le criblage virtuel. Il se base sur des critères pharmacocinétiques et toxicologiques pour évaluer le potentiel thérapeutique des molécules. En utilisant des règles bien établies, comme la règle des cinq de Lipinski, FafDrugs évalue la "drug-likeness" des molécules, c'est-à-dire leur probabilité d'être des médicaments potentiels. De plus, il identifie les motifs structurels associés à la toxicité, permettant d'éliminer précocelement les composés potentiellement nocifs. Outre ces fonctions de filtrage, FafDrugs offre également des capacités de formatage, rendant les chimiothèques compatibles avec divers logiciels de criblage virtuel. Ainsi, grâce à sa facilité d'accès et à ses fonctionnalités robustes, FafDrugs est devenu un allié incontournable pour optimiser les chimiothèques et accélérer le processus de découverte de médicaments.

## Résultats

### Visualisation 3D de la protéine PCSK9

La structure 3D de la protéine PCSK9 a été déterminée par Schroeder et al. en 2013 grâce à la diffraction des rayons X (Code PDB : 4NE9). La résolution obtenue était de 2,60 Å, avec une R-value de 0,211 et une R-free de 0,190, indiquant une bonne qualité de la structure cristalline.

#### IDENTIFICATION DE SITE ACTIF

Le site de liaison est principalement localisé sur la surface de PCSK9 est constitué des acides aminés allant de 367 à 381. En plus la triade catalytique, formée par les acides aminés Asp 186, His 226 et Ser 386, est essentielle à l'auto-activation de PCSK9. Par ailleurs, l'Asn 317 se situe dans la poche oxyanionique. La structure de PCSK9 comprend un domaine catalytique (acides aminés 153-474) et un domaine riche en cystéines et histidines. Dans le cadre de notre étude, nous avons ciblé à la fois la triade catalytique de PCSK9 dans le but d'inhiber son auto-activation et, la surface d'interaction à la LDLR. Fen Du et ses collaborateurs ont montré qu'une modification génétique de PCSK9 influence les LDLR. Ces chercheurs ont identifié une région de PCSK9, le domaine C-terminal (CD) constitué de 242 acides aminés et situé entre les positions 450 et 692, comme étant cruciale pour son auto-clivage, sa sécrétion et son activité.

#### POSITION DU SITE ACTIF

Dans la séquence protéique il est de 450 à 692 **C-TERMINAL** (ce qui est sélectionné en jaune) Fen Du., et al.

```
"MGTVSSRRSWPLLLLLLGPAGARAQEQEDEDGYEELVLALRSEEDGLAEAPEHGTATFHRCAKDPWRLPGTYVVVL  
KEETHLSQSERTARRLQAQAARRGYLTKILHVFHGLPGFLVKMSGDLLALKLPHVDYIEEDSSVFAQSIPWNLERITPPRYR  
ADEYQPPDGGSVLVEVYLLDTSIQSDHREIEGRVMVTDFENVPEEDGTRFHRQASKCDSHGTHLAGVVSGRDAGVAKGASMR  
SLRVLNCQGKGTVSGTLIGLEFIRKSQQLVQPVGPLVLLPLAGGYSRVLNAAQCQLARAGVVLVTAAGNFRDDACLYSPASAPE  
VITVGATNAQDQPVTGLGTNFGRCVDLFAPGEDIIGASSDCSTCFVSQSGTSQAAAHVAGIAAMMLSAEPELTIAELRQRL  
IHFSAKDVINEAWFPEDQRVLTPNLVAALPPSTHGAGWQLFCRTVWSAHSG PTRMATAVARCAPDEELLSCSSFSRGKRRG  
ERMEAQGGKLVCRAHNAFGGEVYAIARCCLLPQANCSVHTAPP AEASMGTRVHCHQQGHVLTGCSALPGTSHVLGAYAVDNTCVRS  
PVLRPRGQPQNQCVGHREASIHASCCCHAPGLECKVKEHGIPAPQEQTVAEEGWTLGCSALPGTSHVLGAYAVDNTCVRS  
RDVSTTGSTSEGAVTAVAICCRSRHLAQASQELQ "
```

#### La position :

La position du premier et le dernier acide aminé du site actif : 450 → 692, la figure 60 montre l'interface d'interaction entre la PCSK9 et LDLR avec les résidus participant dans l'interaction.

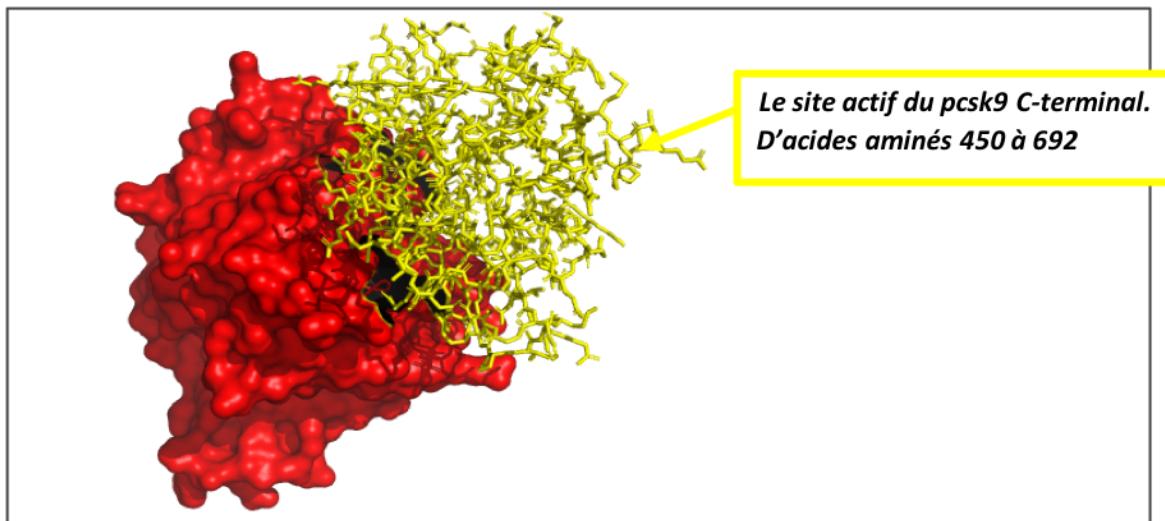


FIGURE 59. VISUALISATION DES ACIDES AMINÉS DE SITE ACTIF DE LA PCSK9 CONCERNANT C-TER.

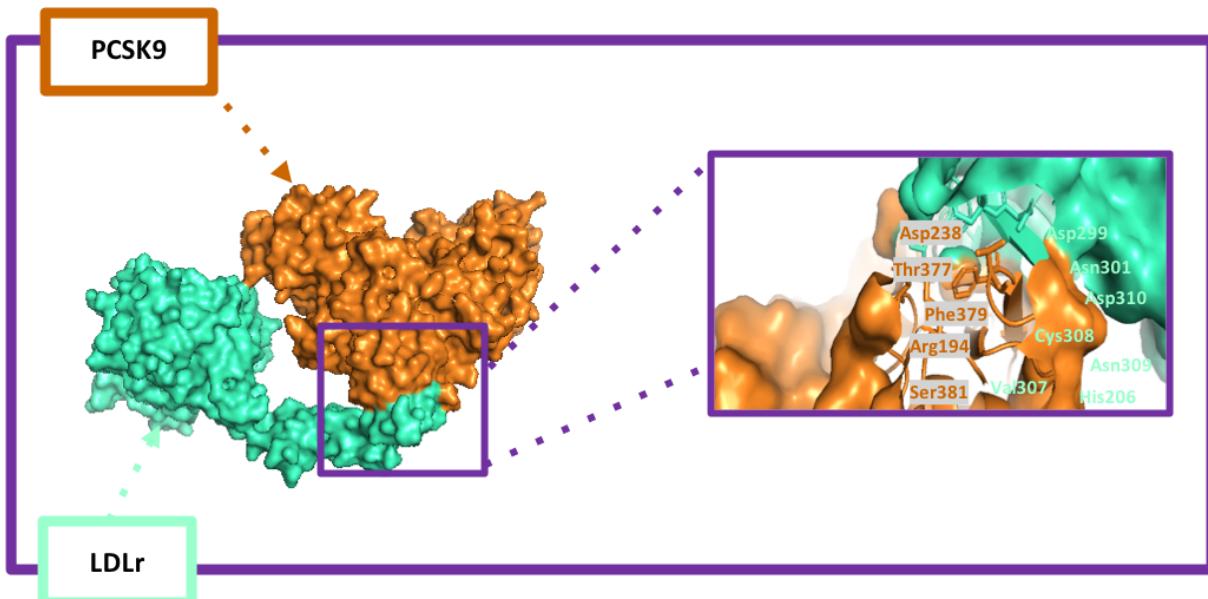


FIGURE 60. INTERFACE D'INTERACTION ENTRE PCSK9 ET LDLR, PCSK9 EN DORÉ ET LDLR EN BLEU CIEL.

#### Docking moléculaire de composé 16 l'inhibiteur connu DE PCSK9

D'après les recherches menées par Petrilli, W.L. et ses collaborateurs, le composé 16 a été mis en avant comme un inhibiteur potentiel de PCSK9. Dans le cadre de notre étude, nous avons réalisé un docking moléculaire de cette molécule pour la comparer à d'autres candidats potentiels. Les résultats du docking moléculaire sont présentés dans le tableau 10, incluant la nomenclature complète du composé 16. Les interactions 2D ont été visualisées et sont illustrées dans les figures 61 et 62.

TABLEAU 10. RÉSULTATS DU CRIBLAGE AVEC LA MOLÉCULE RÉFÉRENTIELLE.

PubChem	Nomenclature	Liaison d'affinité	RMSD
- New ligand (063)	<p>Acide</p> <p>4'-{[(1R)-6-{2-[2-({N^5~}-N^1-bis(tert-butoxycarbonyl)carbamimidoyl]-N^2-(tert-butoxycarbonyl)-L-ornithyl}amino)éthoxy]éthoxy}-1-méthyl-1-{2-oxo-2-[(1,3-thiazol-2-yl)amino]éthyl}-1,2,3,4-tétrahydroisoquinolin-7-yl]oxy}-2'-fluoro[1,1'-biphényl]-4-carboxylique</p> <p><b>Formule chimique : C<sub>53</sub>H<sub>69</sub>F N<sub>8</sub>O<sub>13</sub>S</b></p>	-8.9	0

LE DIAGRAMME 2D DE L'INTERACTION ENTRE L'INHIBITEUR CONNU ET LA PCSK9

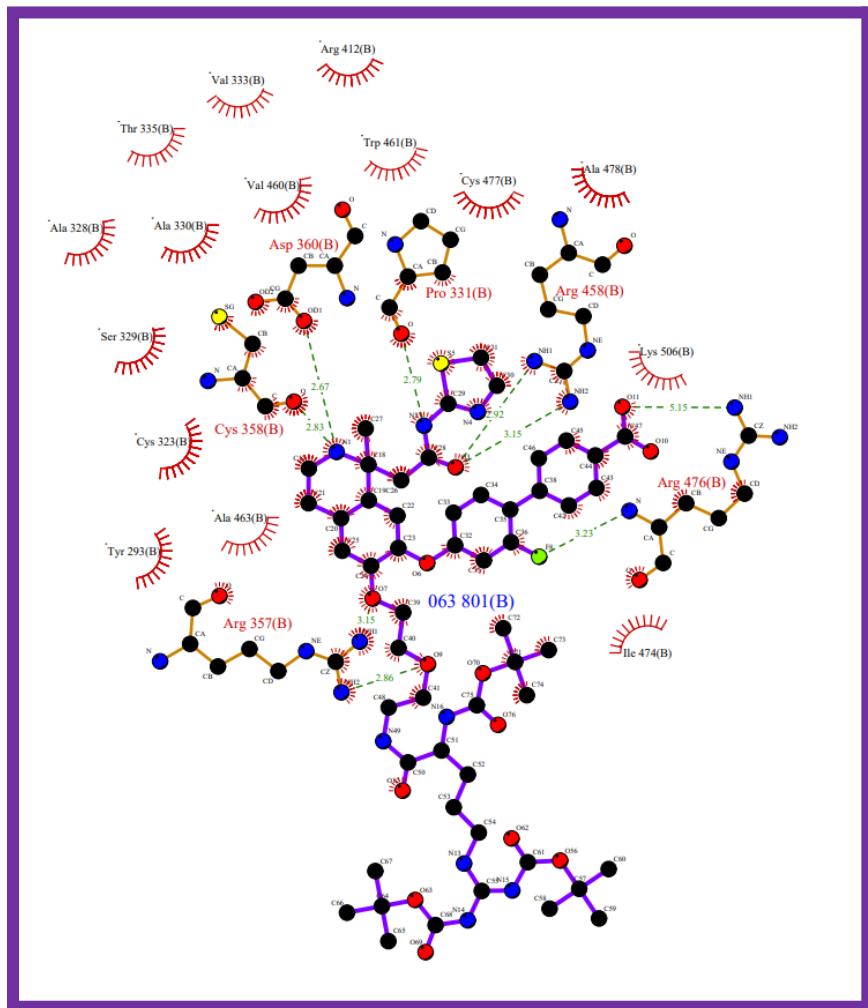


FIGURE 61. ANALYSE 2D DE MODE DE LIAISON ENTRE L'INHIBITEUR IDENTIFIÉ *IN VITRO* ET PCSK9.

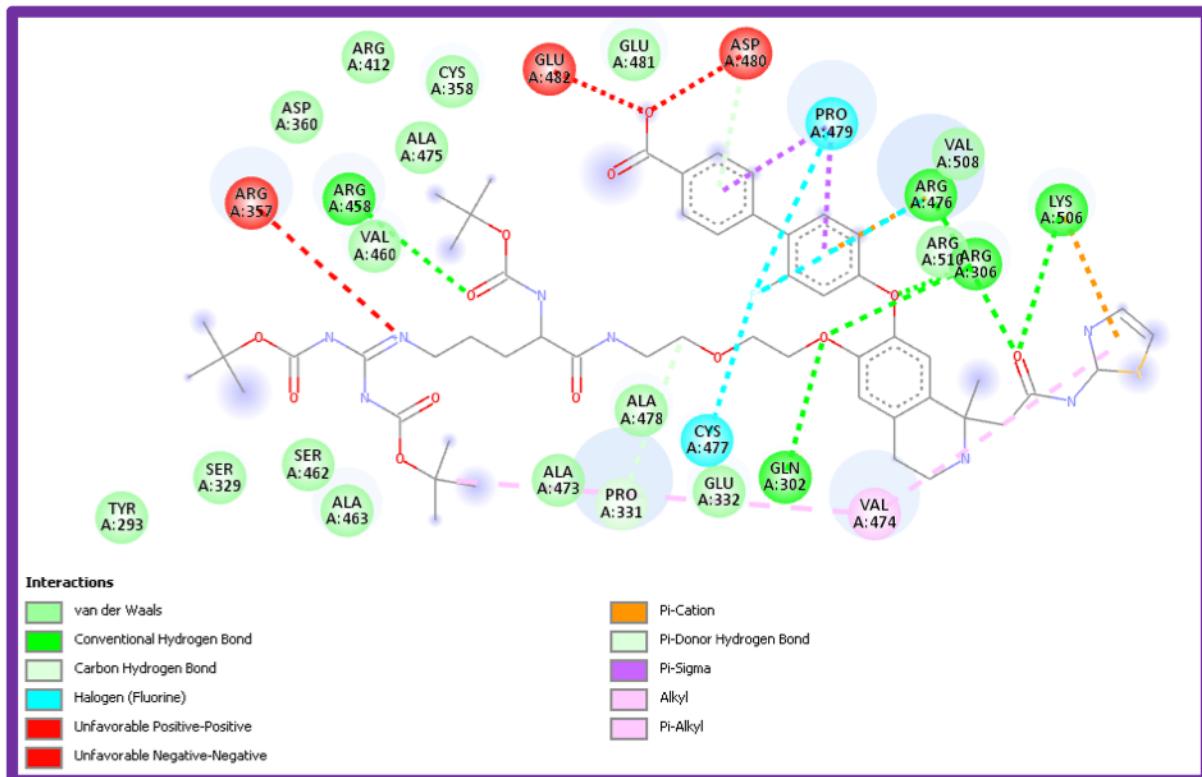


FIGURE 62. DIAGRAMME 2D DE L'INTERACTION OBTENU ENTRE PCSK9 ET L'INHIBITEUR IDENTIFIÉ COMPOSÉ 16.

L'ensemble des interactions moléculaires observées entre le composé 16 (molécule de référence) et la PCSK9 suite au docking moléculaire comprend : deux liaisons hydrogènes avec Arg458 et Arg476, une liaison de Van Der Waals avec Val460, ainsi que deux liaisons halogène/pi-alkyl avec Cys477 et Val477. Ces acides aminés font partie de ceux ciblés pour inhiber la protéine PCSK9.

#### RÉSULTATS DE CRIBLAGE VIRTUEL DES PHYTOCOMPOSÉS MAROCAINES

##### ANALYSE D'ÉNERGIE DE LIAISON ET RMSD

L'affinité et la Rmsd obtenues lors du criblage virtuel permettent d'évaluer la qualité des interactions entre les ligands et la protéine. Ainsi, il est possible d'identifier des composés organiques qui pourraient servir de futur inhibiteurs. Le tableau ci-après présente les composés ayant une affinité inférieure à -8.9 et une Rmsd égale à 0. Sur 600 ligands étudiés, issus de *Cannabis sativa* (100 composés) et de plantes marocaines (500 composés), le tableau met en évidence les ligands présentant une bonne affinité par rapport à la molécule de référence :

- 9 ligands provenant de *Cannabis sativa*
- 45 phytocomposés issus de plantes marocaines

TABLEAU 12. ANALYSE D'ÉNERGIE DE LIAISON ET RMSD.

ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	DU BINDING AFFINITY	Rmsd/u b
<b>LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES</b>					
10382485	---	CANNABIS SATIVA	6-Prenylapigenin	-9.3	0
59444416	---	CANNABIS SATIVA	THCVA	-9.2	0
93147	---	CANNABIS SATIVA	Tetrahydrocannabivarin	-9.4	0
293175	---		3,6,6,9-Tetramethyl-7,8 ,9,10-tetrahydro-6H-be nzo[c]chromen-1-ol	-9.9	0
134740	---	CANNABIS SATIVA	(+)-delta9-Tetrahydroca nnabinol	-9.5	0
14604081	MCULE-560452 5198-0	CANNABIS SATIVA	6-Prenylluteolin	-9.5	0
443123	---	CANNABIS SATIVA	6-prenylchrysin	-9.3	0
59444392	---	CANNABIS SATIVA	Cannabinol-C4	-9.7	0
10382701	---		1-Pentyl-(1naphthoyl) indole	-9.4	0
10891602	MNPDB00002	Balanophora tobiracola / Eichhornia crassipes	13-epi-manool	-9.3	0
5280681	MNPDB00025	Cheirolophus species	3-O-méthylquercétine	-9.4	0
10569	MNPDB00043	Palicourea c ondensata / Cedrus atlantica	abietic acid	-9.7	0
443474	MNPDB00045	Paliurus ramossissimus	Abietinol	-9.9	0
73170	MNPDB00052	Inula crithmoides	alpha-amyrin	-9.4	0
5280443	MNPDB00064	Artemisia mesatlantica	Apigenin	-9.4	0
71726407	MNPDB00092	Ceratonia siliqua(Penicillium citrinum)	citraquinochroman	-9.9	0
11694869	MNPDB00097	Juniperus species	dehydroabietinal	-9.7	0
442027	MNPDB00105	Tetraclinis articulata	Ferruginol	-10.4	0
5379096	MNPDB00116	Cheirolophus species	Jaceosidin	-10	0
5280863	MNPDB00117	Chenopodium murale	Kaempferol	-9.4	0
44627603	MNPDB00199	Fraxinus excelsior	EXCELSIDE B	-9.3	0
6474310	MNPDB00201	Inula viscosa	3,5-dicaffeoylquinic acid	-10.6	0

ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	BINDING AFFINITY	Rmsd/u b
<b>LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES</b>					
10382485	---	<i>CANNABIS SATIVA</i>	6-Prenylapigenin	-9.3	0
59444416	---	<i>CANNABIS SATIVA</i>	THCVA	-9.2	0
93147	---	<i>CANNABIS SATIVA</i>	Tetrahydrocannabivarin	-9.4	0
293175	---		3,6,6,9-Tetramethyl-7,8 ,9,10-tetrahydro-6H-be nzo[c]chromen-1-ol	-9.9	0
134740	---	<i>CANNABIS SATIVA</i>	(+)-delta9-Tetrahydroca nnabinol	-9.5	0
14604081	MCULE-560452 5198-0	<i>CANNABIS SATIVA</i>	6-Prenylluteolin	-9.5	0
443123	---	<i>CANNABIS SATIVA</i>	6-prenylchrysin	-9.3	0
59444392	---	<i>CANNABIS SATIVA</i>	Cannabinol-C4	-9.7	0
5281600	MNPDB00204	<i>Tetraclinis articulata</i>	amentoflavone	-9.8	0
5385553	MNPDB00205	<i>Adenocarpus anagyrifolius</i>	apigenin 7-O-glucoside	-9.3	0
9851181	MNPDB00206	<i>Adenocarpus anagyrifolius</i>	apigenin 7-O-rutinoside	-9.7	0
9064	MNPDB00212	<i>Tetraclinis articulata</i>	Catechin	-9.4	0
5281609	MNPDB00216	<i>Tetraclinis articulata</i>	cupressuflavone	-9.7	0
5319116	MNPDB00241	<i>Adenocarpus anagyrifolius</i>	luteolin 4'-Oglucoside	-10.4	0
5280637	MNPDB00242	<i>Adenocarpus anagyrifolius</i>	luteolin 7-O-glucoside,Cynarosi de, Luteoloside	-10	0
5281604	MNPDB00264	<i>Carya illinoiensis</i>	Azaleatin	-10.1	0
5489501	MNPDB00266	<i>Carya illinoiensis</i>	Caryatin	-9.8	0
5280343	MNPDB00268	<i>Carya illinoiensis</i>	Quercetin	-9.4	0
5319744	MNPDB00322	<i>Retama raetam</i>	3'-O-methylorobol	-9.6	0
-	MNPDB00323	<i>Cotula cinerea</i>	4,5- dicaffeoylquinic acid Malonyl	-9.3	0
5281697	MNPDB00328	<i>Retama raetam</i>	6-hydroxyapigenin	-9.2	0
5281642	MNPDB00330	<i>Lavandula dentata var. dentata</i>	6-hydroxyluteolin	-9.7	0
89644	MNPDB00343	<i>Mentha pulegium</i>	altersolanol A	-9.6	0

ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	BINDING AFFINITY	Rmsd/u b
<b>LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES</b>					
10382485	---	CANNABIS SATIVA	6-Prenylapigenin	-9.3	0
59444416	---	CANNABIS SATIVA	THCVA	-9.2	0
93147	---	CANNABIS SATIVA	Tetrahydrocannabivarin	-9.4	0
293175	---		3,6,6,9-Tetramethyl-7,8 ,9,10-tetrahydro-6H-be nzo[c]chromen-1-ol	-9.9	0
134740	---	CANNABIS SATIVA	(+)-delta9-Tetrahydroca nnabinol	-9.5	0
14604081	MCULE-560452 5198-0	CANNABIS SATIVA	6-Prenylluteolin	-9.5	0
443123	---	CANNABIS SATIVA	6-prenylchrysin	-9.3	0
59444392	---	CANNABIS SATIVA	Cannabinol-C4	-9.7	0
101637166	MNPDB00346	Fraxinus angustifolia	angustifolioside A	-9.5	0
44257813	MNPDB00349	Lavandula stoechas subsp. Atlantica	apigenin 7-allosyl(1?2)glucoside	-10.8	0
5319484	MNPDB00350	Lavandula stoechas subsp. Atlantica	apigenin 7-O-glucuronide	-10.9	0
5280704	MNPDB00351	Chrysanthemum viscidehirtum	Apigetrin	-9.8	0
6419835	MNPDB00378	Arbutus unedo	(-)catechin gallate	-9.9	0
1794427	MNPDB00380	Anabasis aretioides Coss. & Moq.	Chlorogenic acid	-9.5	0
162464	MNPDB00384	Artemisia herba-alba	Cirsilineol	-9.2	0
107971	MNPDB00397	Adenocarpus artemisifolius	Daidzin	-9.2	0
11541786	MNPDB00411	Globularia alypum	eriodictyol 7-O-sophoroside	-10.1	0
5280961	MNPDB00423	Coriandrum sativum L.	Genistein	-9.4	0
5281617	MNPDB00425	Mentha pulegium	Genkwanin	-9.3	0
445718	MNPDB00428	Fraxinus excelsior	3,8,9,10-TETRAHYDROX Y-7-HYDROXYMETHYL- 6-OXA-1,3-DIAZA-SPIR O[4.5]DECANE-2,4-DIO NE	-9.4	0
10621	MNPDB00443	Anabasis aretioides Coss. & Moq.	Hesperidin	-10	0

ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	DU BINDING AFFINITY	Rmsd/u b
<b>LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES</b>					
10382485	---	CANNABIS SATIVA	6-Prenylapigenin	-9.3	0
59444416	---	CANNABIS SATIVA	THCVA	-9.2	0
93147	---	CANNABIS SATIVA	Tetrahydrocannabivarin	-9.4	0
293175	---		3,6,6,9-Tetramethyl-7,8,9,10-tetrahydro-6H-benzo[c]chromen-1-ol	-9.9	0
134740	---	CANNABIS SATIVA	(+)-delta9-Tetrahydrocannabinol	-9.5	0
14604081	MCULE-560452 5198-0	CANNABIS SATIVA	6-Prenylluteolin	-9.5	0
443123	---	CANNABIS SATIVA	6-prenylchrysin	-9.3	0
59444392	---	CANNABIS SATIVA	Cannabinol-C4	-9.7	0
5281780	MNPDB00479	Cotula cinerea	3,4-dicaffeoylquinic acid	-9.3	0
14463052	MNPDB00492	Thymelaea lythroides	rutarensin	-10.1	0
5281699	MNPDB00494	Bupleurum spinosum	tamarixetin	-9.4	0
439246	MNPDB00704	Anabasis aretioides Coss. & Moq.	Naringenin	-9.2	0

#### VISUALISATION ET ANALYSE D'INTERACTION LIGAND-PROTEINE

Après l'analyse de ces 54 ligands qui ont un bon score, nous avons trouvé qu'il y a 5 composés qui ont une interaction aux acides aminés ciblée dans **notre sujet de recherche**, en général qu'on va les représentés comme suivant :

#### Le Cannabis Sativa

Pour **Cannabis sativa** nous avons trouvé qu'il y a **deux ligands** ont une interaction avec les acides aminés ciblées sont présenté dans le tableau suivant :

TABLEAU 13. LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES

LES LIGANDS DU CANNABIS SATIVA ET LEUR ANALOGUES					
ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	BINDING AFFINITY	Rmsd/ub
10382485	---	CANNABIS SATIVA	6-Prenylapigenin	-9.3 Kcal/mol	0

14604081	MCULE-56045 25198-0	CANNABIS SATIVA	6-Prenylluteolin	-9.5 Kcal/mol	0
----------	------------------------	-----------------	------------------	------------------	---

### Le Diagramme 2D

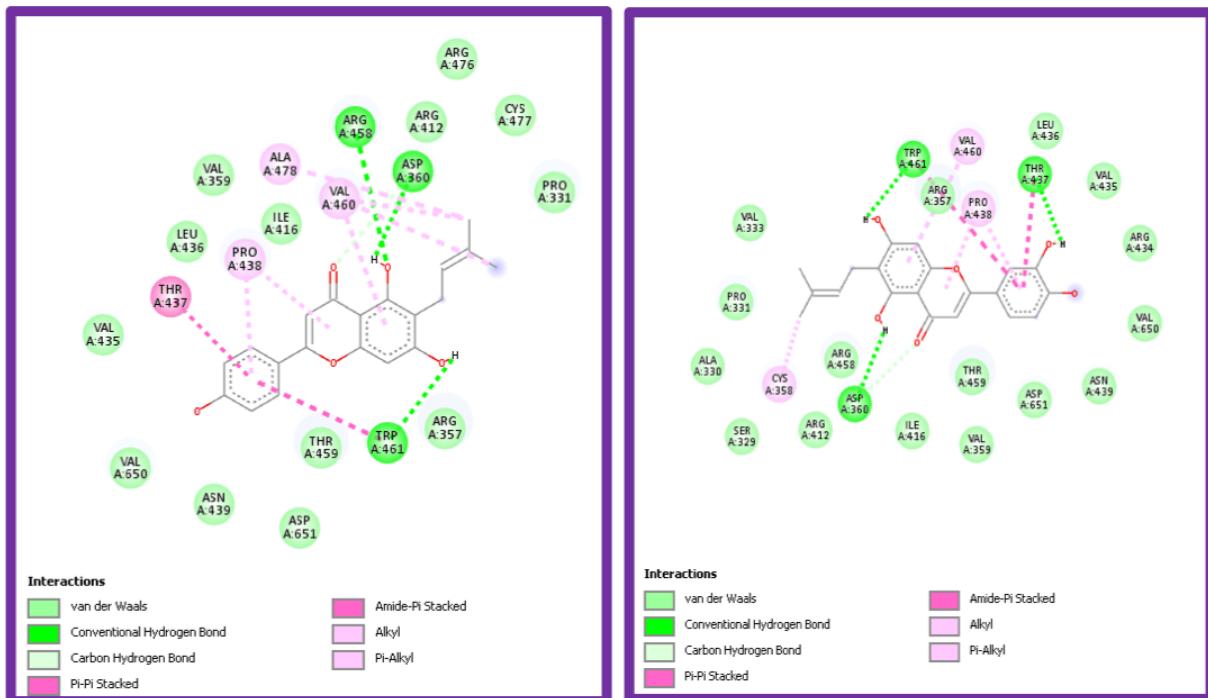


FIGURE 63. VISUALISATION 2D DE L'INTERACTION ENTRE 6-PRENYLAPIGENIN ET 6-PRENYLLUTEOLIN.

### Analyse des interactions ligand-protéine

Le tableau suivant montre les types des liaisons révélées entre les composés identifiés et la PCSK9 :

TABLEAU 14. TYPES DES LIAISONS AVEC LES ACIDES AMINÉS CIBLÉS DANS LE PCSK9

Nom du ligand	Liaisons hydrogènes	Autres types de liaisons chimiques		RMSD/ub
6-Prenylapigenin	Asp360, Arg458	VDW	HYDROPHOBES	0.0
		Cys477	Val460	
6-Prenylluteolin	Asp360	Ala330	Val460	0.0

### Visualisation 3D des interactions ligand-protéine

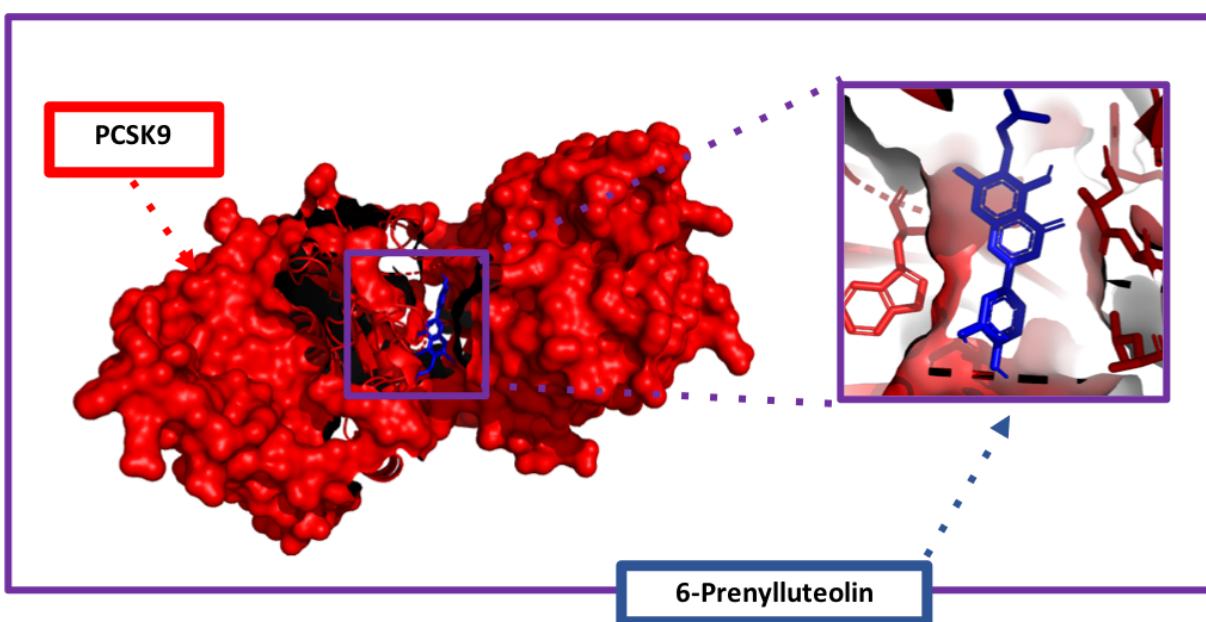
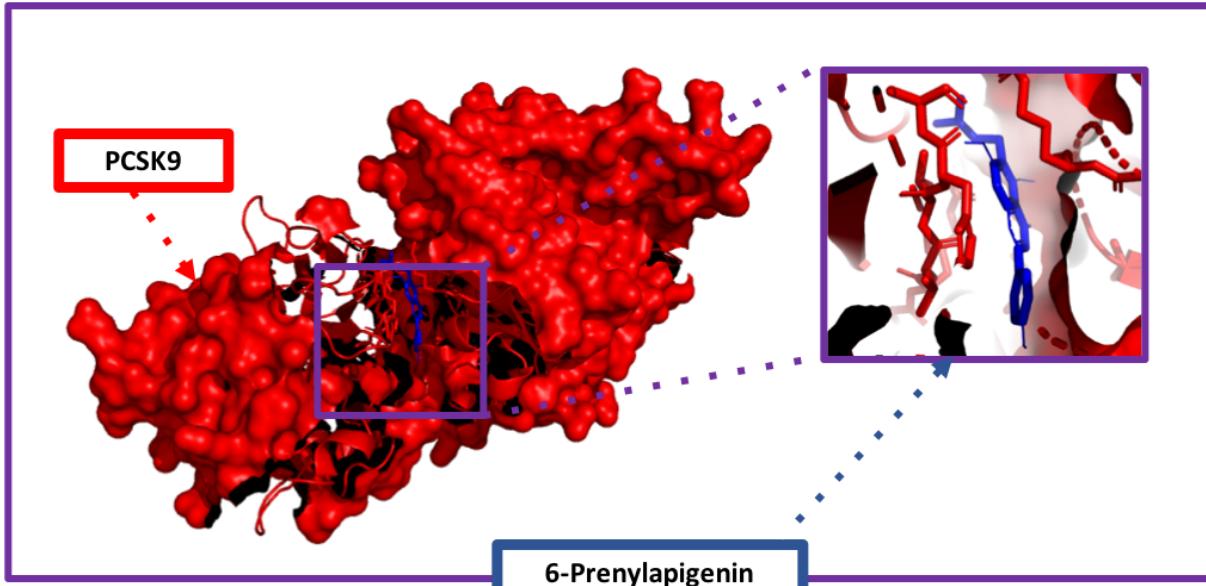


FIGURE 64. VISUALISATION 2D DE L'INTERACTION OBTENU PAR DOCKING MOLÉCULAIRE ENTRE 6-PRENYLAPIGENIN ET 6-PRENYLLUTEOLIN.

#### Les phyto-composés marocains

Pour les **phyto-composés** nous avons trouvé qu'il y a **trois composés** qui ont montré une interaction avec les acides aminés ciblées, le tableau 15 représente les résultats obtenus :

TABLEAU 15. LES PHYTOCOMPOSÉS DE MPDB QUI ONT MONtré UN HAUTE POTENTIELLE DE LIAISON

LES LIGANDS PHYTOCOMPOSÉS DE MNPDB					
ID PubChem	ID MNPDB	NOM DE LA PLANTE	NOMENCLATURE DU LIGAND	BINDING AFFINITY	Rmsd/ub
44627603	MNPDB00199	<i>Fraxinus excelsior</i>	EXCELSIDE B	-9.3	0
9851181	MNPDB00206	<i>Adenocarpus anagyrifolius</i>	apigenin 7-O-rutinoside	-9.7	0
----	MNPDB00323	<i>Cotula cinerea</i>	4,5-dicaffeoylquinic acid Malonyl	-9.3	0

### Le diagramme 2D

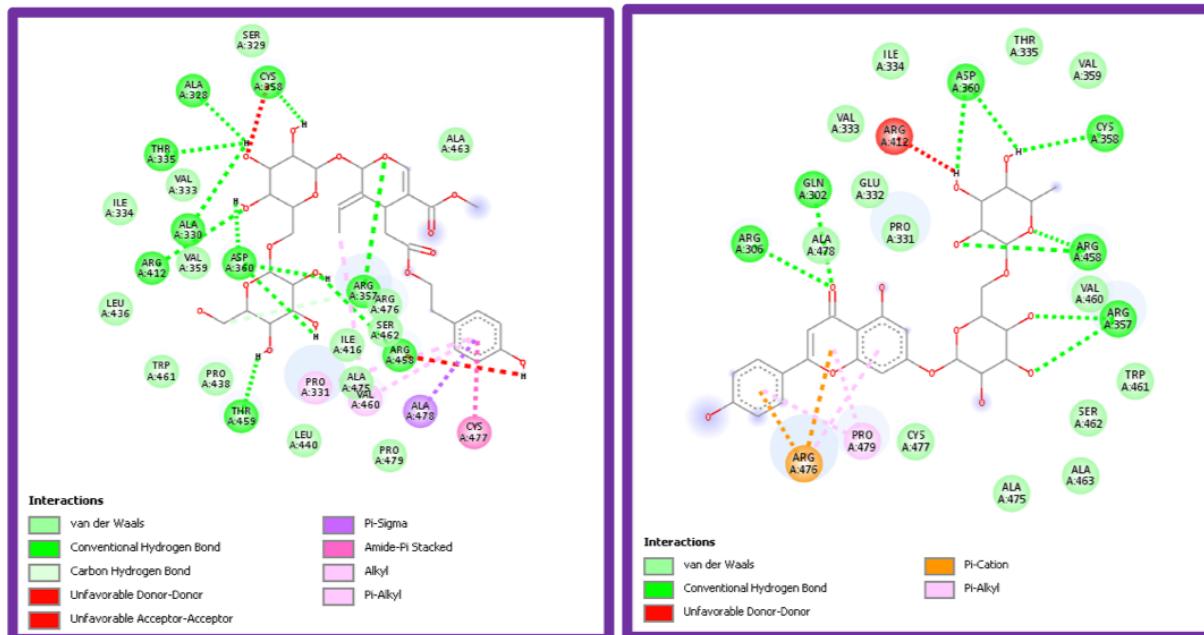


FIGURE 65. RÉSEAU D'INTERACTION ENTRE PCSK9 ET DEUX PHYTO-COMPOSÉS EXCELSIDE B A GAUCHE ET APIGENIN 7-O-RUTINOSIDE.

### Analyse des interactions ligand-protéine

Le tableau suivant montre types des liaisons avec les acides aminés ciblés dans le PCSK9 :

TABLEAU 16. TYPES DES LIAISONS AVEC LES ACIDES AMINÉS CIBLÉS DANS LE PCSK9

Nom du ligand	Liaisons hydrogènes	Autres types de liaisons chimiques		RMSD/ub
EXCELSIDE B	Asp360, Arg458, Arg357, Cys358	VDW	HYDROPHOBES	0.0
		--	Val460, Cys477	

apigenin 7-O-rutinoside	Asp360, Cys358, Arg357, Arg458,	Val460, Cys477	---	0.0
4,5- dicaffeoylquinic acid Malonyl	Asp360, Pro331, Arg476 Arg357	Ala330, Cys477	Val460	0.0

#### Visualisation 3D des interactions ligand-protéine

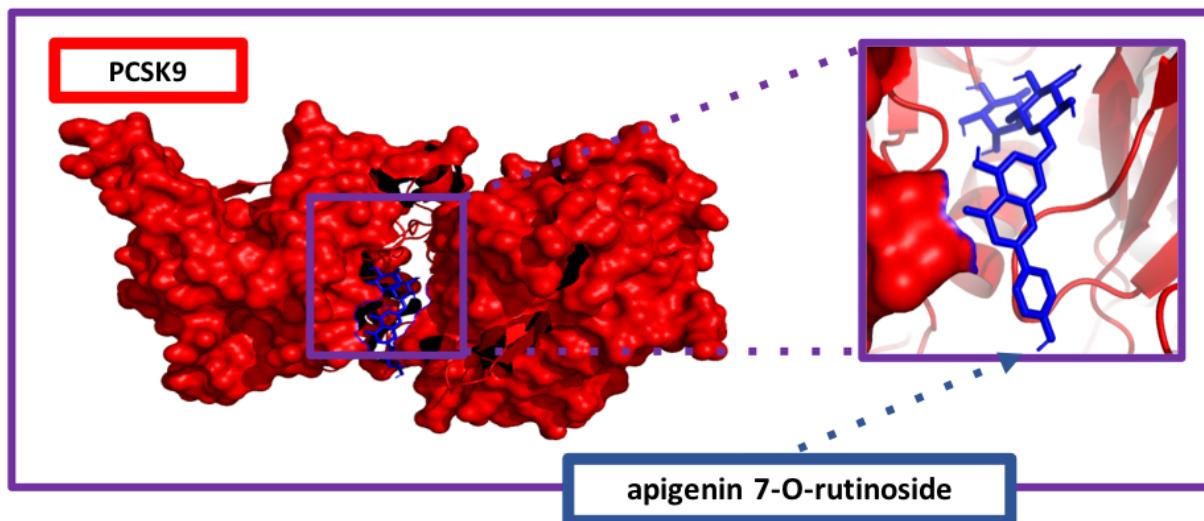
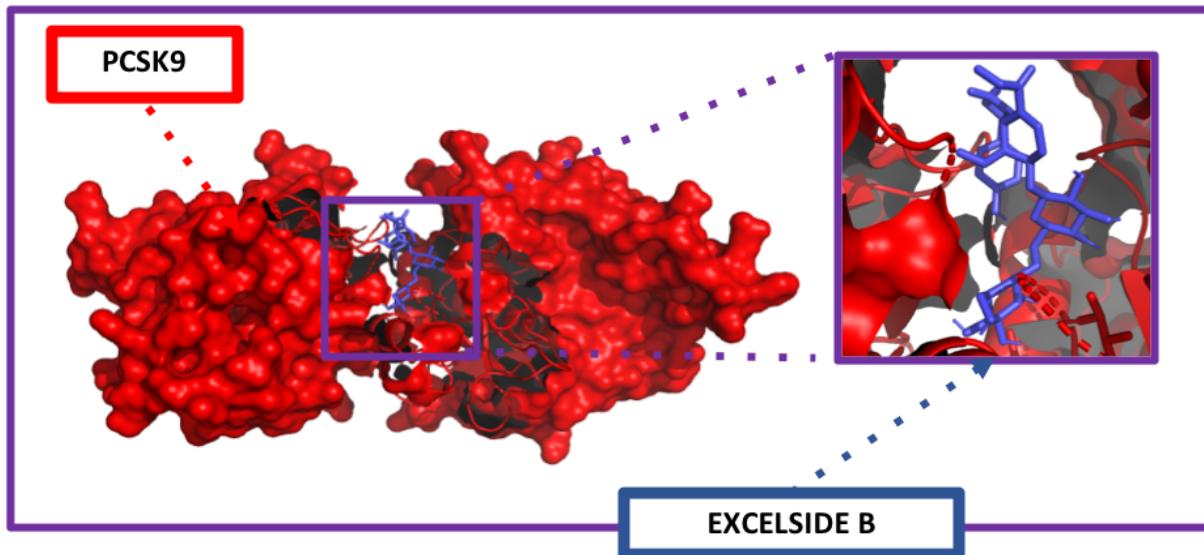


FIGURE 66. RÉSEAU D'INTERACTION ENTRE PCSK9 ET DEUX PHYTO-COMPOSÉS EXCELSIDE B ET APIGENIN 7-O-RUTINOSIDE.

#### PRÉDICTION DES PROPRIÉTÉS ADME-Tox

Suite aux résultats obtenus lors du criblage virtuel des composés naturels dérivés de *Cannabis sativa* et d'autres plantes marocaines, nous avons entrepris une évaluation de leurs propriétés physico-chimiques. Cette démarche avait pour objectif de prédire leurs pharmacocinétiques. Au total, cinq composés ont été retenus pour cette prédition ADME. Plus spécifiquement, parmi les composés de *Cannabis sativa*, deux ont montré une affinité pour la PCSK9 : il s'agit du 6-Prenylapigenin et du 6-Prenylluteolin. En ce qui concerne les autres plantes marocaines, trois phyto-composés ont été

identifiés : EXCELSIDE B, Apigenin-7-O-rutinoside et 4,5-dicaffeoylquinic acid Malonyl. L'étude de ces propriétés est fondamentale dans le processus de conception de médicaments. Pour être envisagés comme de potentiels candidats médicamenteux, ces composés doivent satisfaire aux critères établis par la règle de Lipinski. Les résultats sont présentés dans le tableau 17.

TABLEAU 17. LE TRIAGE DE CES 5 COMPOSÉS SE BASE SUR LE RESPECT DE 3 CRITÈRES DE RÈGLE DE LIPINSKI AU MOINS.

Règle de Lipinski								
	MW (g/mol)	logP	HBA	HBD	Lipinski violation	Solubility	Oral bioavailability	Result
<b>CANNABINOIDE SAXSTIVA</b>	6-Prenylapigenin	<b>338.4</b>	<b>3.7</b>	<b>5</b>	<b>3</b>	<b>0</b>	<b>good</b>	<b>good</b>
	6-Prenylluteolin	<b>354.4</b>	<b>4.2</b>	<b>6</b>	<b>4</b>	<b>0</b>	<b>good</b>	<b>good</b>
<b>Other Moroccan plants</b>	EXCELSIDE B	<b>686.7</b>	<b>-2.2</b>	<b>17</b>	<b>8</b>	<b>3</b>	<b>good</b>	<b>good</b>
	Apigenin-7-O-rutinoside	<b>578.5</b>	<b>-0.8</b>	<b>14</b>	<b>8</b>	<b>3</b>	<b>good</b>	<b>Good</b>
	4,5- dicaffeoylquinic acid Malonyl	<b>516.45</b>	<b>1.02</b>	<b>11</b>	<b>7</b>	<b>3</b>	<b>Good</b>	<b>Good</b>

La sélection de ces 5 composés repose sur la satisfaction d'au moins 3 critères de la règle de Lipinski.

### Discussion

L'inhibition de PCSK9 est une avancée majeure dans le domaine de la cardiologie et de la gestion des taux de cholestérol. PCSK9, ou proprotéine convertase subtilisine/kexine de type 9, est une enzyme produite dans le foie qui régule les récepteurs du LDL-cholestérol (le "mauvais" cholestérol) dans le sang. Lorsque PCSK9 est en excès, il dégrade ces récepteurs, entraînant une accumulation de cholestérol LDL dans le sang, ce qui peut augmenter le risque de maladies cardiovasculaires. Les médicaments inhibiteurs de PCSK9, tels que l'évolocumab et l'alirocumab, fonctionnent en bloquant l'action de cette enzyme, permettant ainsi aux récepteurs LDL de rester actifs et de réduire considérablement les niveaux de cholestérol LDL. Cette approche novatrice offre un moyen efficace de réduire le risque de maladies cardiaques chez les patients atteints d'hypercholestérolémie familiale ou d'autres troubles métaboliques, ou chez ceux qui ne parviennent pas à contrôler leur cholestérol avec d'autres traitements. Les anticorps monoclonaux, malgré leur efficacité indéniable dans la neutralisation de cibles spécifiques, présentent plusieurs inconvénients. Tout d'abord, leur taille et leur complexité moléculaire peuvent rendre leur production et leur purification coûteuses et complexes. De plus, leur administration nécessite souvent des injections intraveineuses ou sous-cutanées, ce qui peut entraîner des problèmes de tolérance et de conformité chez certains patients. En revanche, les petits composés inhibiteurs, tels que les inhibiteurs de PCSK9 sous forme de molécules de petite taille, offrent des avantages significatifs. Leur administration par voie orale est plus pratique pour les patients, favorisant ainsi une meilleure observance thérapeutique. De plus, ces petites molécules sont plus faciles à produire à grande échelle, ce qui peut potentiellement réduire les coûts de fabrication et rendre les traitements plus accessibles.

Dans cette étude, un criblage virtuel a été effectué à l'aide du logiciel PyRx sur les composés de la base de données MPDB. Nous avons sélectionné les composés présentant une forte affinité de liaison, avec un RMSD égal à 0. Les résultats ont identifié neuf composés issus de la plante marocaine Cannabis sativa, ainsi que 45 autres phyto-composés marocains parmi ceux présentant une affinité inférieure à -8.9 kcal/mol. Par la suite, nous avons procédé à une analyse structurale à l'aide de Discovery Studio afin d'étudier les interactions moléculaires entre chaque phyto-composé et la PCSK9. Cette analyse structurale a mis en évidence cinq composés interagissant avec les résidus fonctionnels de la PCSK9 grâce à des liaisons hydrogène, des forces de Van der Waals et des interactions hydrophobes.

Le composé identifié 10382485, issu de la plante Cannabis sativa, 6-Prenylapigenin, présente des liaisons hydrogènes robustes avec les résidus ciblés par notre recherche, à savoir (Asp360, Arg458). De plus, il établit une interaction de type Van der Waals avec le résidu (Cys477) et une liaison pi-alkyl avec l'acide aminé (Val460). Par ailleurs, le composé identifié 14604081, également issu de la plante Cannabis sativa et nommé 6-Prenylluteolin, affiche une liaison hydrogène forte avec l'un des résidus ciblés par notre recherche, à savoir (Asp360). De plus, il présente une interaction de type Van der Waals avec le résidu (Ala330) et une liaison pi-alkyl avec l'acide aminé (Val460). L'efficacité de l'interaction médicament-récepteur dépend fortement de la présence de liaisons hydrogène, qui renforcent la liaison. Par conséquent, ces composés pourraient potentiellement perturber la liaison entre PCSK9 et le récepteur LDLr, ce qui en ferait des candidats prometteurs en tant qu'inhibiteurs du mécanisme de PCSK9. Ensuite l'évaluation des propriétés ADME-Tox (absorption, distribution, métabolisme, excrétion et toxicité) de ces deux composés FAF-Drugs4 a été utilisé. Les résultats ont montré que ces deux composés présentent des caractéristiques physico-chimiques favorables et une bonne perméabilité et biodisponibilité. De plus, ils respectent les règles de Lipinski sans aucune violation, ce qui est un critère important pour la conception de médicaments potentiels. Les trois composés identifiés sont les suivants : MNPDB00199, issu de la plante Fraxinus Excelsior, nommé Excelside B, présente des liaisons hydrogènes fortes avec plusieurs résidus ciblés par notre étude, notamment Asp360, Arg458, Arg357 et Cys358, ainsi que des liaisons pi-alkyl/pi-sigma avec les acides aminés Val460 et Cys477. Le composé MNPDB00206, issu de la plante Adenocarpus anagyrifolius et appelé Apigenin-7-O-rutinoside, montre des liaisons hydrogènes solides avec les résidus ciblés par notre recherche, à savoir Asp360, Cys358, Arg357 et Arg458. De plus, il présente des interactions significatives de type Van Der Waals avec les résidus Val460 et Cys477, sans former d'autres types de liaisons avec les acides aminés ciblés.

En ce qui concerne le composé MNPDB00323, provenant de la plante Cotula cinerea et nommé 4,5-dicaffeoylquinic acid Malonyl, il établit des liaisons hydrogènes solides avec plusieurs résidus

ciblés par notre recherche, notamment Asp360, Pro331, Arg476 et Arg357. De plus, il présente des interactions significatives de type Van Der Waals avec les résidus Ala330 et Cys477, ainsi qu'une liaison carbone-hydrogène avec l'acide aminé Val460. Ces interactions moléculaires sont cruciales pour comprendre le potentiel de ces composés en tant qu'inhibiteurs de PCSK9. Leur capacité à former des liaisons hydrogènes et des interactions de type Van Der Waals avec les résidus clés de PCSK9 les positionne comme des candidats prometteurs pour perturber l'interaction entre PCSK9 et LDLr, ouvrant ainsi des perspectives passionnantes dans le développement de traitements potentiels pour les maladies cardiovasculaires. En bloquant la liaison entre la protéine PCSK9 et la protéine LDLr, ces interactions pourraient avoir un effet inhibiteur sur la PCSK9. Ces composés pourraient donc être des candidats potentiels pour le traitement du LDL sanguin en utilisant les phytocomposés comme agents thérapeutiques. La prédition ADME-Tox reste importante pour analyser les caractéristiques de ces phytocomposés on va utiliser le site web FAF-Drugs4, Ces phytocomposés présentent des violations des règles de Lipinski et ont donc des caractéristiques physico-chimiques défavorables.

La seule étude similaire à la nôtre a été menée par l'équipe Min et al., 2015 qui visait à réaliser un Criblage *in silico* de bibliothèques chimiques pour développer des inhibiteurs qui entravent l'interaction de PCSK9 avec le récepteur LDL en ciblant le site de liaison de PCSK9 avec le domaine EGF-A du LDLR.

## Conclusion générale

Le diabète de type 2 et les maladies cardiovasculaires représentent un défi majeur pour la santé mondiale. Les avancées technologiques, notamment dans les domaines de la bioinformatique et de la chimio-informatique, offrent des opportunités sans précédent pour aborder ces maladies sous un angle novateur. Les études pangénomiques ont mis en lumière des variants génétiques susceptibles d'augmenter le risque de DT2 et de MCV. En particulier, le gène RETN, code pour la résistine, a été identifié comme un acteur clé. Grâce à la modélisation moléculaire et à la bioinformatique, nous sommes désormais en mesure d'évaluer l'impact de ces variants sur la structure et la fonction des protéines, offrant ainsi des connaissances précieuses sur les mécanismes moléculaires sous-jacents. La découverte de médicaments assistée par ordinateur a également révolutionné notre approche de la recherche thérapeutique. En utilisant ces techniques, nous pouvons désormais explorer des bibliothèques de composés à une échelle sans précédent, identifiant rapidement des candidats prometteurs pour le traitement du DT2 et des MCV. Bien que le Maroc possède une riche biodiversité en terme de flore qui offre un réservoir inexploré de composés naturels potentiellement thérapeutiques, la mise en place d'une base de données marocaine dédiée, intégrant des informations détaillées sur ces composés, facilitera grandement leur exploration et leur exploitation vis-à-vis les cibles thérapeutiques de DT2 et MCV. Cette initiative, en plus de combler un vide dans le domaine de la recherche, positionne le Maroc comme un des leaders dans l'exploration de composés issus de plantes aromatiques et médicinales pour des problèmes de santé mondiaux. En combinant une compréhension approfondie de la biologie moléculaire avec les avancées technologiques modernes et la richesse de la biodiversité marocaine, nous sommes sur la voie de découvertes significatives qui pourraient transformer le traitement de ces maladies à l'avenir. La biologie moléculaire et la bioinformatique ont révolutionné notre compréhension des maladies et ont ouvert la voie à des approches thérapeutiques plus ciblées. Dans le cadre de cette recherche, nous avons exploré le potentiel des variants génétiques du gène humain RETN en relation avec le diabète de type 2 (DT2) et les maladies cardiovasculaires par l'aide des outils de bioinformatique. Le premier chapitre a mis en lumière l'impact des SNPs dans le gène RETN, suggérant leur rôle potentiel dans le changement conformationnel et fonctionnel, l'identification des variants du gène RETN a été une étape cruciale, en utilisant des données de séquençage à haut débit et en s'appuyant sur des bases de données reconnues telles que NCBI et ENSEMBL, nous avons pu mettre en évidence plusieurs variants et révélé que certaines mutations pourraient perturber la structure et la stabilité protéique,

suggérant un impact potentiel sur la fonction de la protéine et, par conséquent, sur la physiopathologie du DT2 et des maladies cardiovasculaires. Le deuxième chapitre a été particulièrement révélateur, démontrant le potentiel thérapeutique des composés marocains, notamment en ce qui concerne leur interaction avec PPAR $\gamma$ , un régulateur clé du métabolisme lipidique et glucidique. Nos découvertes ont montré que des composés tels que 3,4-DICQA interagissent de manière similaire avec PPAR $\gamma$  comme un agoniste complet, tandis que l'acide chlorogénique et le Lutéoloside pourraient améliorer la sensibilité à l'insuline via PPAR $\gamma$  similairement à des agonistes partiels. Le troisième chapitre a abouti à l'élaboration de la MPDB, une première base de données marocaines pour la découverte de médicaments basée sur les produits chimiques végétaux marocains avec les données nécessaires au criblage virtuel et à l'exploitation *in silico*. Enfin, le quatrième chapitre a capitalisé sur l'exploitation des données de la MPDB pour identifier des composés prometteurs contre la PCSK9, une nouvelle cible attractive contre les MCV. Les résultats obtenus au cours de cette thèse ouvrent plusieurs voies pour de futures recherches. Les prédictions computationnelles du premier chapitre nécessitent une validation expérimentale pour déterminer l'impact réel des SNPs sur la protéine. Les interactions des phyto-composés marocains avec PPAR $\gamma$ , mises en évidence dans le deuxième chapitre, méritent une exploration plus approfondie pour dévoiler leur potentiel thérapeutique réel et envisager des modifications chimiques pour optimiser leur efficacité. La MPDB, introduite dans le troisième chapitre, pourrait bénéficier d'une mise à jour régulière et de collaborations pour enrichir sa portée. Enfin, les composés identifiés pour la PCSK9 dans le quatrième chapitre méritent des études expérimentales pour confirmer leur efficacité et envisager leur développement en tant que médicaments potentiels contre les MCV. Ces perspectives, combinées, soulignent l'immense potentiel des ressources naturelles marocaines dans la conception de médicaments innovants. La recherche présentée dans cette thèse a jeté les bases d'une exploration plus approfondie des ressources phytochimiques, offrant une perspective riche pour le domaine de la recherche biomédicale.

## BIBLIOGRAPHIE GÉNÉRALE

Abifadel M, Varret M, Rabès JP, et al. *Mutations in PCSK9 cause autosomal dominant hypercholesterolemia*. Nat Genet. 2003;34(2):154-156.

Abraham, M. J., et al. (2015). GROMACS: *High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 1-2, 19-25.

Abraham, M. J., et al. (2015). GROMACS: *High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 1-2, 19-25.

Adobe Systems Inc. (1985). *PostScript language reference manual*. Addison-Wesley Professional.

ADVANCE Collaborative Group — *Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial) : a randomised controlled trial*. Lancet, 2007, 370 : 829-40.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. (2010). *A method and server for predicting damaging missense mutations*. Nature methods, 7(4), 248–249.

Ahlstrom, M. G., & Carlsson, J. (1998). *Computer-aided drug design: a historical perspective. Methods and principles in medicinal chemistry*, 11, 1-26.

Alaoui, K., Belabbes, R., & Cherrah, Y. (2017). *Use of medicinal plants in oral health care in the Marrakech region*. Phytothérapie, 15(4), 235-242.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2015). *Molecular biology of the cell* (6th ed.). Garland Science.

American Diabetes Association. (2021). *Standards of Medical Care in Diabetes—2021*. Diabetes Care, 44 (Supplement 1), S1-S232.

Anderson, J. W., Allgood, L. D., Lawrence, A., Altringer, L. A., Jerdack, G. R., Hengehold, D. A., & Morel, J. G. (2000). *Cholesterol-lowering effects of psyllium intake adjunctive to diet therapy in men and women with hypercholesterolemia: meta-analysis of 8 controlled trials*. The American Journal of Clinical Nutrition, 71(2), 472-479.

André VACHERON (2011). *COMMUNIQUÉ au nom de la Commission IV (Maladies cardiovasculaires): Complications cardiovasculaires du diabète*. Bull. Acad. Natle Méd., 2011, 195, no 1, 205-211, séance du 18 janvier 2011

Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., & Stuppner, H. (2015). *Discovery and resupply of pharmacologically active plant-derived natural products: A review*. *Biotechnology Advances*, 33(8), 1582-1614.

Aviram, M., et al. (2000). *Dietary antioxidants and paraoxonases against LDL oxidation and atherosclerosis development*. *Handbook of Experimental Pharmacology*, 146, 301-325.

Avogaro, A., Albiero, M., Menegazzo, L., de Kreutzenberg, S., & Fadini, G. P. (2011). *Endothelial dysfunction in diabetes: the role of reparatory mechanisms*. *Diabetes care*, 34 (Supplement 2), S285-S290.

Arul Murugan N, Ruba Priya G, Narahari Sastry G, Markidis S. Artificial intelligence in virtual screening: Models versus experiments. *Drug Discov Today*. 1 juill 2022;27(7):1913-23.

Bai, X.-C., McMullan, G., & Scheres, S. H. W. (2015). *How cryo-EM is revolutionizing structural biology*. *Trends in Biochemical Sciences*, 40(1), 49-57.

Baker, D., & Sali, A. (2001). *Protein structure prediction and structural genomics*. *Science*, 294(5540), 93-96.

Bayan, L., et al. (2014). *Garlic: A review of potential therapeutic effects*. *Avicenna Journal of Phytomedicine*, 4(1), 1-14.

Belmoukhtar, M., Bouyahya, A., & Abrini, J. (2021). *Moroccan Medicinal plants as inhibitors against SARS-CoV-2 main protease: in silico screening and molecular dynamic simulation*. *Computational Biology and Chemistry*, 92, 107431.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). *Molecular dynamics with coupling to an external bath*. *The Journal of Chemical Physics*, 81(8), 3684-3690.

Berkov, S., Zayed, R., & Doncheva, T. (2012). *Alkaloid patterns in some varieties of Datura stramonium*. *Fitoterapia*, 83(3), 619-625.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). *The protein data bank*. *Nucleic acids research*, 28(1), 235-242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., ... & Davies, G. J. (1977). *The protein data bank: a computer-based archival file for macromolecular structures*. *Journal of molecular biology*, 112(3), 535-542.

Bhattacharya, D., Nowotny, J., Cao, R., & Cheng, J. (2016). 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic acids research*, 44(W1), W406-W409.

Bhattacharyya, M., & Bhat, C. R. (2020). *A comprehensive study of the radius of gyration as an effective tool to study protein structure and stability*. *Biochemical and biophysical research communications*, 526(2), 485-491.

Bhawana, et al. (2011). *Curcumin nanoparticles: Preparation, characterization, and antimicrobial study*. *Journal of Agricultural and Food Chemistry*, 59(5), 2056-2061.

Bnouham, M., Ziyyat, A., Mekhfi, H., Tahri, A., & Legssyer, A. (2006). *Medicinal plants with potential antidiabetic activity-A review of ten years of herbal medicine research (1990-2000)*. *International Journal of Diabetes and Metabolism*, 14, 1-25.

Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P. N., Cho, H., ... & Zhang, C. (2010). *Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma*. *Nature*, 467(7315), 596-599.

Bouhaddouda, N., Benmansour, A., & Mekarnia, M. (2015). *Antioxidant properties and total phenolic content of three varieties of carob tree leaves from Algeria*. *Natural Product Research*, 29(24), 2305-2308.

Bousta, D., Soulaymani, A., & Saad, A. (2020). *Moroccan medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations*. *Journal of Biomolecular Structure and Dynamics*, 39(17), 6539-6547.

Brown, M. S., & Goldstein, J. L. (1986). *A receptor-mediated pathway for cholesterol homeostasis*. *Science*, 232(4746), 34-47.

Bruning, J. B., et al. (2007). *Partial agonists activate PPARgamma using a helix 12 independent mechanism*. *Structure*, 15(10), 1258-1271.

Bruning, J. B., et al. (2007). *Partial agonists activate PPARgamma using a helix 12 independent mechanism*. *Structure*, 15(10), 1258-1271.

Cardona, F., Andrés-Lacueva, C., Tulipani, S., Tinahones, F. J., & Queipo-Ortuño, M. I. (2013). *Benefits of polyphenols on gut microbiota and implications in human health*. *The Journal of Nutritional Biochemistry*, 24(8), 1415-1422.

Carr, A. C., & Frei, B. (1999). *Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans*. *The American Journal of Clinical Nutrition*, 69(6), 1086-1107.

CDC. (2021). National Diabetes Statistics Report. Centers for Disease Control and Prevention.

Chaieb, K., et al. (2007). *The chemical composition and biological activity of essential oil, Eugenia caryophyllata (Syzygium aromaticum L. Myrtaceae): A short review*. *Phytotherapy Research*, 21(6), 501-506.

Chandra, V., et al. (2008). Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. *Nature*, 456(7220), 350–356.

Charrouf, Z., & Guillaume, D. (2008). *Argan oil: Occurrence, composition and impact on human health*. *European Journal of Lipid Science and Technology*, 110(7), 632-636.

Cheng, Y. (2018). *Single-particle cryo-EM at crystallographic resolution*. *Cell*, 175(7), 1630-1638.

Cholesterol Treatment Trialists' (CTT) Collaborators. (2010). *Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials*. *The Lancet*, 376(9753), 1670-1681.

Chuengsamarn, S., et al. (2014). *Curcumin extract for prevention of type 2 diabetes*. *Diabetes Care*, 37(12), 3330-3332.

Chuengsamarn, S., Rattanamongkolgul, S., Luechapudiporn, R., Phisalaphong, C., & Jirawatnotai, S. (2012). *Curcumin extract for prevention of type 2 diabetes*. *Diabetes Care*, 35(11), 2121-2127.

Codoñer-Franch, P, et E. Alonso-Iglesias. 2015. « *Resistin: Insulin Resistance to Malignancy* ». Clinica Chimica Acta 438 (janvier): 46-54. <https://doi.org/10.1016/j.cca.2014.07.043>.

Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. *Sequence variations in PCSK9, low LDL, and protection against coronary heart disease*. N Engl J Med. 2006;354(12):1264-1272.

*Comprehensive Asymmetrical Docking Approach*. Chemical Biology & Drug Design, 80(3), 358–371.

Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A; NISC Comparative Sequencing Program. (2005). *Distribution and intensity of constraint in mammalian genomic sequence*. Genome research, 15(7), 901–913.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*. Journal of the American Chemical Society, 117(19), 5179-5197.

Cosentino, F., Grant, P. J., Aboyans,V., Bailey, C. J., Ceriello, A., Delgado, V., ... & Schnell, O. (2020). *2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD*. European Heart Journal, 41(2), 255-323.

Chaachouay N, Benkhnigue O, Zidane L. Ethnobotanical and Ethnomedicinal study of medicinal and aromatic plants used against dermatological diseases by the people of Rif, Morocco. J Herb Med. 1 mars 2022;32:100542.

Dauter, Z., Dauter, M., & Dodson, E. (1999). Jolly SAD. *Nature Structural Biology*, 6(11), 985-988.

Dill, K. A., & MacCallum, J. L. (2012). *The protein-folding problem, 50 years on*. Science, 338(6110), 1042-1046.

Dill, K. A., & MacCallum, J. L. (2012). *The protein-folding problem, 50 years on*. Science, 338(6110), 1042-1046.

Dobson, C. M. (2003). *Protein folding and misfolding*. Nature, 426(6968), 884-890.

Driggers EM, Hale SP, Lee J, Terrett NK. The exploration of macrocycles for drug discovery — an underexploited structural class. Nat Rev Drug Discov. 2008;7(7):608.

Dormandy, J. A., Charbonnel, B., Eckland, D. J., Erdmann, E., Massi-Benedetti, M., Moules, I. K., ... & PROactive investigators. (2005). *Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial*. Lancet, 366(9493), 1279-1289.

Drenth, J. (2007). *Principles of protein X-ray crystallography*. Springer Science & Business Media.

Duke, J. (1992). *Handbook of phytochemical constituents of GRAS herbs and other economic plants*. CRC Press.

Dunn, P. J. (2012). *A brief history of synthesis*. Pure and Applied Chemistry, 84(9), 1833-1847.

Eddouks, M., Maghrani, M., Lemhadri, A., Ouahidi, M. L., & Jouad, H. (2002). *Ethnopharmacological survey of medicinal plants used for the treatment of diabetes mellitus, hypertension and cardiac diseases in the south-east region of Morocco (Tafilalet)*. Journal of Ethnopharmacology, 82(2-3), 97-103.

Ekor, M. (2014). *The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety*. Frontiers in Pharmacology, 4, 177.

El-Hilaly, J., Hmammouchi, M., & Lyoussi, B. (2003). *Ethnobotanical studies and economic evaluation of medicinal plants in Taounate province (Northern Morocco)*. Journal of Ethnopharmacology, 86(2-3), 149-158.

Emerging Risk Factors Collaboration. (2010). *Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies*. Lancet, 375(9733), 2215-2222.

Emerging Risk Factors Collaboration. (2010). *Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies*. Lancet, 375(9733), 2215-2222.

Fadli, M., Saad, A., Sayadi, S., Chevalier, J., Mezrioui, N. E., Pagès, J. M., & Hassani, L. (2012). *Antibacterial activity of Thymus maroccanus and Thymus broussonetii essential oils against nosocomial infection – bacteria and their synergistic potential with antibiotics*. Phytomedicine, 19(5), 464-471.

Fasshauer, M., & Paschke, R. (2003). *Regulation of adipocytokines and insulin resistance*. Diabetologia, 46(12), 1594-1603. <https://doi.org/10.1007/s00125-003-1222-6>

*FDA approves Repatha to treat certain patients with high cholesterol*. FDA News Release. August 27, 2015.

Feinberg AP. (2007). *Phenotypic plasticity and the epigenetics of human disease*. Nature, 447(7143), 433–440.

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). *Molecular docking and structure-based drug design strategies*. Molecules, 20(7), 13384-13421.

Freddolino, P. L., & Schulten, K. (2009). *Common structural transitions in explicit-solvent simulations of villin headpiece folding*. Biophysical Journal, 97(9), 2338-2347.

Frenkel, D., & Smit, B. (2002). *Understanding molecular simulation: from algorithms to applications*. Academic press.

Gao, D., Trayhurn, P., & Bing, C. (2017). *Resistin and related cytokines*.

Goldstein, J. L., & Brown, M. S. (2009). *The LDL receptor. Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(4), 431-438.

González-Castejón, M., & Rodriguez-Casado, A. (2011). *Dietary phytochemicals and their potential effects on obesity: a review*. Pharmacological Research, 64(5), 438-455.

Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., & Caves, L. S. (2006). *Bio3d: an R package for the comparative analysis of protein structures*. Bioinformatics, 22(21), 2695-2696.

Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., & Caves, L. S. (2006). *Bio3d: an R package for the comparative analysis of protein structures*. Bioinformatics, 22(21), 2695-2696.

Grest, G. S., & Kremer, K. (1986). *Molecular dynamics simulation for polymers in the presence of a heat bath*. Physical Review A, 33(5), 3628-3631.

Gromiha, M. M., & Selvaraj, S. (2004). *Inter-residue interactions in protein folding and stability*. Progress in biophysics and molecular biology, 86(2), 235-277.

Grundy SM, Stone NJ, Bailey AL, et al. 2018

*AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.* J Am Coll Cardiol. 2019;73(24):e285-e350.

Grygiel-Górniak, B. (2014). Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications--a review. Nutrition Journal, 13, 17.  
<https://doi.org/10.1186/1475-2891-13-17>

Grzanna, R., et al. (2005). Ginger — An herbal medicinal product with broad anti-inflammatory actions. Journal of Medicinal Food, 8(2), 125-132.

Gupta, S. C., et al. (2013). Therapeutic roles of curcumin: Lessons learned from clinical trials. The AAPS Journal, 15(1), 195-218.

Hajduk, P. J., Huth, J. R., & Tse, C. (2008). Predicting protein druggability. Drug discovery today, 13(23-24), 1051-1058.

Hann, M. M., & Keseru, G. M. (2012). Finding the sweet spot: the role of nature and nurture in medicinal chemistry. Nature reviews Drug discovery, 11(5), 355-365.

Harrison, R. K. (2016). Phase II and phase III failures: 2013–2015. Nature Reviews Drug Discovery, 15(12), 817-818.

Head, M. S. (2006). A critical assessment of docking programs and scoring functions. Journal of Medicinal Chemistry, 49(20), 5912-5931.

Heifetz, A., et al. (2012). Understanding Molecular Recognition by G Protein-Coupled Receptors Using a

Higdon, J. V., Delage, B., Williams, D. E., & Dashwood, R. H. (2007). Cruciferous vegetables and human cancer risk: epidemiologic evidence and mechanistic basis. Pharmacological Research, 55(3), 224-236.

Hollman, P. C., et al. (2011). The biological relevance of direct antioxidant effects of polyphenols for cardiovascular health in humans is not established. The Journal of Nutrition, 141(5), 989S-1009S.

Hoover, W. G. (1985). *Canonical dynamics: Equilibrium phase-space distributions*. Physical Review A, 31(3), 1695-1697.

Horton JD, Cohen JC, Hobbs HH. *PCSK9: a convertase that coordinates LDL catabolism*. J Lipid Res. 2009;50 Suppl:S172-177.

Huang, S. Y., & Zou, X. (2007). *Advances and challenges in protein-ligand docking*. International Journal of Molecular Sciences, 11(8), 3016-3034.

International Diabetes Federation. (2019). *IDF Diabetes Atlas*, 9th Edition. Consulté le 1er avril 2023, à partir de <https://diabetesatlas.org/en/resources/>

Jalil, A. M. M., et al. (2015). *Effects of green tea (*Camellia sinensis*) on the pharmacokinetics of amoxicillin in rats*. Journal of Pharmacy and Pharmaceutical Sciences, 18(5), 704-713.

Jasial, S., Hu, Y., & Bajorath, J. (2017). *How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds*. Journal of Medicinal Chemistry, 60(9), 3879-3886. Published April 19, 2017. DOI: 10.1021/acs.jmedchem.7b00154.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*. Journal

Jorgensen, W. L. (2009). *The many roles of computation in drug discovery*. Science, 323(5918), 1793-1794.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). *Highly accurate protein structure prediction with AlphaFold*. Nature, 596(7873), 583–589.

Kadowaki, T., Yamauchi, T., Kubota, N., Hara, K., Ueki, K., & Tobe, K. (2006). *Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome*. The Journal of Clinical Investigation, 116(7), 1784-1792.

Kahn, B. B., Alquier, T., Carling, D., & Hardie, D. G. (2005). *AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism*. Cell Metabolism, 1(1), 15-25.

Kahn, S. E., Cooper, M. E., & Del Prato, S. (2014). *Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future*. The Lancet, 383(9922), 1068-1083. [https://doi.org/10.1016/S0140-6736\(13\)62154-6](https://doi.org/10.1016/S0140-6736(13)62154-6)

Kar, P., & Mishra, R. K. (2011). *Dynamics of protein folding: A review*. Journal of Biosciences, 36(4), 685-697.

Karplus, M., & McCammon, J. A. (2002). *Molecular dynamics simulations of biomolecules*. Nature Structural Biology, 9(9), 646-652.

Karplus, M., & Petsko, G. A. (1990). *Molecular dynamics simulations in biology*. Nature, 347(6294), 631-639.

Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. (2014). *Defining functional DNA elements in the human genome*. Proceedings of the National Academy of Sciences, 111(17), 6131–6138.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, et al. (2008) *Mapping and sequencing of structural variation from eight human genomes*. Nature 453: 56–64.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. (2014). *A general framework for estimating the relative pathogenicity of human genetic variants*. Nature genetics, 46(3), 310–315.

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). *Docking and scoring in virtual screening for drug discovery: Methods and applications*. Nature Reviews Drug Discovery, 3(11), 935-949.

Kong, W., Wei, J., Abidi, P., Lin, M., Inaba, S., Li, C., Wang, Y., Wang, Z., Si, S., Pan, H., Wang, S., Wu, J., Wang, Y., Li, Z., Liu, J., & Jiang, J. D. (2004). *Berberine is a novel cholesterol-lowering drug working through a unique mechanism distinct from statins*. Nature Medicine, 10(12), 1344-1351.

Krinsky, N. I., & Johnson, E. J. (2005). *Carotenoid actions and their relation to health and disease*. Molecular Aspects of Medicine, 26(6), 459-516.

Kumar P, Henikoff S, Ng PC. (2009). *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm*. Nature protocols, 4(7), 1073–1081.

Kwiterovich, P. O. (2000). *The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: a current review*. The American journal of cardiology, 86(12), 5-10.

Kumar A, P N, Kumar M, Jose A, Tomer V, Oz E, et al. Major Phytochemicals: Recent Advances in Health Benefits and Extraction Method. Molecules. 16 janv 2023;28(2):887.

Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 12 déc 2016;12(1):2694-718.

Liu RH. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. Am J Clin Nutr. sept 2003;78(3 Suppl):517S-520S

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). *PROCHECK: a program to check the stereochemical quality of protein structures*. Journal of applied crystallography, 26(2), 283-291.

Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR*. Journal of biomolecular NMR, 8(4), 477-486.

Leach, A. R. (2001). *Molecular modeling: principles and applications*. Prentice hall.

Legeay, S., Rodier, M., Fillon, L., Faure, S., & Clere, N. (2015). *Epigallocatechin Gallate: A Review of Its Beneficial Properties to Prevent Metabolic Syndrome*. Nutrients, 7(7), 5443-5468.

Lehmann, J. M., Moore, L. B., Smith-Oliver, T. A., Wilkison, W. O., Willson, T. M., & Kliewer, S. A. (1995). *An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma)*. Journal of Biological Chemistry, 270(22), 12953-12956. <https://doi.org/10.1074/jbc.270.22.12953>.

Le May C, Kourimate S, Langhi C, Chétiveaux M, Jarry A, Coméra C, Collet X, Kuipers F, Krempf M, Cariou B, Costet P. *Proprotein convertase subtilisin kexin type 9 null mice are protected from postprandial triglyceridemia*. Atherosclerosis. 2009;203(1):176-183.

Li, Y., Reddy, M. A., Miao, F., Shanmugam, N., Yee, J. K., Hawkins, D., Ren, B., & Natarajan, R. (2011). *Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent*

*inflammatory genes. Relevance to diabetes and inflammation.* Journal of Biological Chemistry, 286(32), 28562-28574. <https://doi.org/10.1074/jbc.M111.223925>

Liao, M., Cao, E., Julius, D., & Cheng, Y. (2013). *Structure of the TRPV1 ion channel determined by electron cryo-microscopy.* Nature, 504(7478), 107-112.

Libby, P. (2002). *Atherosclerosis: The New View.* Scientific American, 286(5), 46-55.

Libby, P. (2002). *Inflammation in atherosclerosis.* Nature, 420(6917), 868-874.

Lindahl, E., & Hess, B. (2008). *The power and limitations of NMR and X-ray crystallography in protein structure determination: a comparison of X-ray and NMR structures.* Journal of Chemical Education, 85(11), 1531.

Lipinski CA. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Journal of Pharmaceutical Sciences. 1997 May;86(5):483-488.

Liu, R. H. (2003). *Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals.* The American Journal of Clinical Nutrition, 78(3), 517S-520S.

Liu, R. H. (2004). *Potential synergy of phytochemicals in cancer prevention: mechanism of action.* The Journal of Nutrition, 134(12), 3479S-3485S.

Liu, Y. (2013). *In silico identification of anti-cancer compounds and plants from traditional Chinese medicine database.* Scientific reports, 3, 1021.

Lobo, V., Patil, A., Phatak, A., & Chandra, N. (2010). *Free radicals, antioxidants and functional foods: Impact on human health.* Pharmacognosy Reviews, 4(8), 118.

Lusis, A. J. (2000). *Atherosclerosis.* Nature, 407(6801), 233-241.

Mach, F., Baigent, C., Catapano, A. L., Koskinas, K. C., Casula, M., Badimon, L., ... & De Backer, G. (2020). *2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk.* European Heart Journal, 41(1), 111-188.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., & Karplus, M. (1998). *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*. The Journal of Physical Chemistry B, 102(18), 3586-3616.

Maghrani, M., Lemhadri, A., Jouad, H., Michel, J. B., & Eddouks, M. (2004). *Cholesterol and triglycerides lowering activities of caraway fruits in normal and streptozotocin diabetic rats*. Journal of Ethnopharmacology, 90(2-3), 323-329.

Manach, C., Scalbert, A., Morand, C., Rémesy, C., & Jiménez, L. (2004). *Polyphenols: food sources and bioavailability*. The American Journal of Clinical Nutrition, 79(5), 727-747.

Mandard, S., & Müller, M. (2005). *PPAR $\gamma$  and lipid metabolism in the gut*. Nutrients, 21(1), 12-22.

Mannu, G. S., Zaman, M. J., Gupta, A., Rehman, H. U., & Myint, P. K. (2013). *Evidence of lifestyle modification in the management of hypercholesterolemia*. Current Cardiology Reviews, 9(1), 2-14.

Mannu, G. S., Zaman, M. J., Gupta, A., Rehman, H. U., & Myint, P. K. (2013). *Lifestyle interventions in the management of hypercholesterolemia: Current evidence and future prospects*. Current Cardiology Reviews, 9(1), 2-14.

Marciano, D. P., et al. (2015). *Pharmacological repression of PPAR $\gamma$  promotes osteogenesis*. Nature Communications, 6, 7443.

Maritim, A. C., Sanders, R. A., & Watkins, J. B. (2003). *Diabetes, oxidative stress, and antioxidants: a review*. Journal of Biochemical and Molecular Toxicology, 17(1), 24-38.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al.

McGovern, S. L., & Shoichet, B. K. (2003). *Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes*. Journal of medicinal chemistry, 46(16), 2895-2907.

McPherson, A. (2004). *Introduction to protein crystallization*. Methods, 34(3), 254-265.

Mehta, R. G., Murillo, G., Naithani, R., & Peng, X. (2010). *Cancer chemoprevention by natural products: how far have we come?*. Pharmaceutical Research, 27(6), 950-961.

Meng, X. Y., et al. (2011). *Molecular docking: a powerful approach for structure-based drug discovery*. Current Computer-Aided Drug Design, 7(2), 146-157.

Meng, X. Y., et al. (2011). *Molecular docking: a powerful approach for structure-based drug discovery*. Current Computer-Aided Drug Design, 7(2), 146-157.

Michel Farnier, (2015). *PCSK9 : de la découverte aux perspectives thérapeutiques*. DOI : 10.1684/stv.2015.0884. 77-83 Volume 27, numéro 2, Mars-Avril 2015.

Michel Farnier. (2014). *Les inhibiteurs de PCSK9*

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. Journal of Computational Chemistry, 30(16), 2785-2791.

Mort M, Sterne-Weiler T, Li B, Ball EV, Cooper DN, Radivojac P, Sanford JR, Mooney SD. (2014). *MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing*. Genome Biology, 15(1), R19.

Moss SE, Klein R, Klein BE. *Cause-specific mortality in a population-based study of diabetes*. Am J Public Health. 1991;81 : 1158-62

Moult, J., Fidelis, K., Kryshtafovych, A., & Schwede, T. (2009). *A Critical Assessment of the Predictions in the CASP7 Confidence Interval for Automated Model Building and Model Refinement*. Proteins: Structure, Function, and Bioinformatics, 77(S9), 29-46.

Murata, K., et M. Wolf. 2018. « *Cryo-electron microscopy for structural analysis of dynamic biological macromolecules* ». Biochimica et Biophysica Acta (BBA) - General Subjects, Biophysical Exploration of Dynamical Ordering of Biomolecular Systems, 1862 (2): 324-34. <https://doi.org/10.1016/j.bbagen.2017.07.020>.

Nascimento, Lidiane & Moraes, Angelo & Santana, Kauê & Galúcio, João & Junior, Paulo & Costa, Cristiane & Cruz, Jorddy & Aguiar, Eloisa & Faria, Lênio. (2020). *Bioactive Natural Compounds and*

*Antioxidant Activity of Essential Oils from Spice Plants: New Findings and Potential Applications.*  
Biomolecules. 10. 988. 10.3390/biom10070988.

Nayoung Ahn Kijin Kim, 2016; *HDL-C in Cardiovascular Disease: Effect of Exercise*

Nesto, R. W., Bell, D., Bonow, R. O., Fonseca, V., Grundy, S. M., Horton, E. S., ... & Smith, S. C. (2004). *Thiazolidinedione use, fluid retention, and congestive heart failure: a consensus statement from the American Heart Association and American Diabetes Association.* Diabetes care, 27(1), 256-263.

Nguefack, J., et al. (2004). *Antimicrobial activity of the essential oils from Hymenocarate longiflorus and Nepeta cataria.* Phytotherapy Research, 18(8), 670-673.

Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. J Nat Prod. 27 mars 2020;83(3):770-803.

Nunes CDR, Barreto Arantes M, Menezes de Faria Pereira S, Leandro da Cruz L, de Souza Passos M, Pereira de Moraes L, et al. Plants as Sources of Anti-Inflammatory Agents. Mol Basel Switz. 15 août 2020;25(16):3726.

NIDDK. (2016). *Causes of Diabetes.* National Institute of Diabetes and Digestive and Kidney Diseases.

Nissen, S. E., & Wolski, K. (2007). *Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes.* New England Journal of Medicine, 356(24), 2457-2471.

Nolte, R. T., et al. (1998). *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.* Nature, 395(6698), 137–143.

Nolte, R. T., et al. (1998). *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.* Nature, 395(6698), 137–143.

Nosé, S. (1984). *A unified formulation of the constant temperature molecular dynamics methods.* The Journal of Chemical Physics, 81(1), 511-519.

Ntie-Kang, F., Zofou, D., Babiaka, S. B., Meudom, R., Scharfe, M., Lifongo, L. L., & Mbah, J. A. (2013). *AfroDb: a select highly potent and diverse natural product library from African medicinal plants.* PloS one, 8(10), e78085.

Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*. Journal of Computational Chemistry, 25(13), 1656-1676.

Organisation mondiale de la santé. (2021). Rapport mondial sur le diabète. Consulté le 1er avril 2023, à partir de <https://www.who.int/publications/i/item/9789240032248>

Orozco, M., & Luque, F. J. (2000). *Theoretical methods for the description of the solvent effect in biomolecular systems*. Chemical Reviews, 100(11), 4187-4226.

Pandey, K. B., & Rizvi, S. I. (2009). *Plant polyphenols as dietary antioxidants in human health and disease*. Oxidative Medicine and Cellular Longevity, 2(5), 270-278.

Pandey, K. B., & Rizvi, S. I. (2009). *Plant polyphenols as dietary antioxidants in human health and disease*. Oxidative Medicine and Cellular Longevity, 2(5), 270-278.

Parks, D. J., Blanchard, S. G., Bledsoe, R. K., Chandra, G., Consler, T. G., Kliewer, S. A., Stimmel, J. B., Willson, T. M., Zavacki, A. M., & Moore, D. D. (1999). *Bile acids: natural ligands for an orphan nuclear receptor*. Science, 284(5418), 1365-1368. <https://doi.org/10.1126/science.284.5418.1365>

Parrinello, M., & Rahman, A. (1981). *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 52(12), 7182-7190.

Parthasarathy, S., et al. (1992). *Oxidative modification of LDL: Its role in atherogenesis and the effects of antioxidants*. Free Radical Biology and Medicine, 13(3), 341-390.

Pauling, L., Corey, R. B., & Branson, H. R. (1951). *The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain*. Proceedings of the National Academy of Sciences, 37(4), 205-211.

Pellecchia, M., Sem, D. S., & Wüthrich, K. (2002). *NMR in drug discovery*. Nature Reviews Drug Discovery, 1(3), 211-219.

Phillips, J. C., et al. (2005). *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 26(16), 1781-1802.

Pina-Vaz, C., et al. (2004). *Antifungal activity of Thymus oils and their major compounds*. Journal of the European Academy of Dermatology and Venereology, 18(1), 73-78.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. (2010). *Detection of nonneutral substitution rates on mammalian phylogenies*. Genome research, 20(1), 110–121.

Prlic, A., Yates, A., Bliven, S.E., Rose, P.W., Jacobsen, J., Troshin, P.V., & Zardecki, C. (2010). *PDBsum new things*. Nucleic Acids Research, 38(Database issue), D453-D457. doi: 10.1093/nar/gkp973

Quang D, Chen Y, Xie X. (2015).

Rader, D. J. (2016). *High-density lipoproteins and reverse cholesterol transport: Moving beyond the cholesterol efflux capacity*. Arteriosclerosis, Thrombosis, and Vascular Biology, 36(4), 679-681.

Rawshani, A., Rawshani, A., Franzén, S., Sattar, N., Eliasson, B., Svensson, A. M., ... & Gudbjörnsdóttir, S. (2017). *Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes*. New England Journal of Medicine, 379(7), 633-644.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. (2015). *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology*. Genetics in Medicine, 17(5), 405–424.

Ried, K., et al. (2013). *The effect of garlic on blood pressure and cardiovascular risk factors: A systematic review and meta-analysis*. BMC Cardiovascular Disorders, 13, 49.

Robinson JG, Farnier M, Krempf M, et al. *Efficacy and safety of alirocumab in reducing lipids and cardiovascular events*. N Engl J Med. 2015;372(16):1489-1499.

Rohl, C. A., Strauss, C. E., Misura, K. M., & Baker, D. (2004). *Protein structure prediction using Rosetta*. Methods in Enzymology, 383, 66–93.

Rothwell, J.A., Perez-Jimenez, J., Neveu, V., Medina-Remón, A., M'Hiri, N., García-Lobato, P., Manach, C., Knox, C., Eisner, R., Wishart, D.S. and Scalbert, A., (2013). *Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content*. Database, 2013.

Rupp, B. (2009). *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science.

Rankou H, Culham A, Jury SL, Christenhusz MJM. The endemic flora of Morocco. *Phytotaxa*. 1 févr 2013;78(1):1-69.

Sofowora A, Ogunbodede E, Onayade A. The role and place of medicinal plants in the strategies for disease prevention. *Afr J Tradit Complement Altern Med AJTCAM*. 12 août 2013;10(5):210-29.

Sabatine, M. S., Giugliano, R. P., Wiviott, S. D., Raal, F. J., Blom, D. J., Robinson, J., ... & Somaratne, R. (2017). *Efficacy and safety of evolocumab in reducing lipids and cardiovascular events*. *New England Journal of Medicine*, 376(18), 1713-1722.

Sabatine, M. S., Giugliano, R. P., Wiviott, S. D., Raal, F. J., Blom, D. J., Robinson, J., ... & Somaratne, R. (2017). *Efficacy and safety of evolocumab in reducing lipids and cardiovascular events*. *New England Journal of Medicine*, 376(18), 1713-1722.

Sabatine, M. S., Giugliano, R. P., Wiviott, S. D., Raal, F. J., Blom, D. J., Robinson, J., ... & Somaratne, R. (2017). *Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease*. *New England Journal of Medicine*, 376(18), 1713-1722.

Sabatine MS, Giugliano RP, Keech AC, et al. *Evolocumab and clinical outcomes in patients with cardiovascular disease*. *N Engl J Med*. 2017;376(18):1713-1722.

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., ... & International Diabetes Federation Diabetes Atlas Committee. (2019). *Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas*, 9th edition. *Diabetes Research and Clinical Practice*, 157, 107843.

Sali, A., & Blundell, T. L. (1993). *Comparative Protein Modelling by Satisfaction of Spatial Restraints*. *Journal of Molecular Biology*, 234(3), 779–815.

Salomon-Ferrer, R., et al. (2013). *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald*. *Journal of Chemical Theory and Computation*, 9(9), 3878-3888.

Salomon-Ferrer, R., et al. (2013). *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald*. Journal of Chemical Theory and Computation, 9(9), 3878-3888.

Saltiel, A. R., & Olefsky, J. M. (1996). *Thiazolidinediones in the treatment of insulin resistance and type II diabetes*. Diabetes, 45(12), 1661-1669. <https://doi.org/10.2337/diab.45.12.1661>

Sattler, M., Schleucher, J., & Griesinger, C. (1999). *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulse field gradients*. Progress in Nuclear Magnetic Resonance Spectroscopy, 34(2), 93-158

Sauna ZE, Kimchi-Sarfaty C. (2011). *Understanding the contribution of synonymous mutations to human disease*. Nature Reviews Genetics, 12(10), 683–691.

Schneider, G., & Fechner, U. (2005). *Computer-based de novo design of drug-like molecules*. Nature Reviews Drug Discovery, 4(8), 649-663.

Seidah, N. G., Benjannet, S., Wickham, L., Marcinkiewicz, J., Jasmin, S. B., Stifani, S., ... & Prat, A. (2003). *The secretory proprotein convertase NARC-1: roles in liver regeneration and neuronal differentiation*. Proceedings of the National Academy of Sciences, 100(3), 928-933.

Seidah, N. G., Benjannet, S., Wickham, L., Marcinkiewicz, J., Jasmin, S. B., Stifani, S., ... & Prat, A. (2003). *The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation*. Proceedings of the National Academy of Sciences, 100(3), 928-933.

Shrake, A., & Rupley, J. A. (1973). *Environment and exposure to solvent of protein atoms. Lysozyme and insulin*. Journal of Molecular Biology, 79(2), 351-371.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. (2005). *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes*. Genome research, 15(8), 1034–1050.

Smith, J. C., Merzel, F., & Bondar, A. N. (2010). *Structure, dynamics and reactions of protein hydration water*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368(1914), 1169-1180.

Sniderman, A. D., Williams, K., Contois, J. H., Monroe, H. M., McQueen, M. J., de Graaf, J., & Furberg, C. D. (2003). *A meta-analysis of low-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol, and apolipoprotein B as markers of cardiovascular risk*. *Circulation: Cardiovascular Quality and Outcomes*, 4(3), 337-345.

Soutar, A. K., & Naoumova, R. P. (2007). *Genetic causes of familial hypercholesterolemia: Implications for diagnosis and treatment*. *Nature Clinical Practice Cardiovascular Medicine*, 4(4), 214-225.

Steinberg, D. (1997). *Low density lipoprotein oxidation and its pathobiological significance*. *Journal of Biological Chemistry*, 272(34), 20963-20966.

Steinberg, D. (2009). *The LDL modification hypothesis of atherogenesis: an update*. *Journal of Lipid Research*, 50(Supplement), S376-S381.

Stein EA, Raal F. *Reduction of low-density lipoprotein cholesterol by monoclonal antibody inhibition of PCSK9*. *Annu Rev Med*. 2014;65:417-431.

Steppan, C. M., Bailey, S. T., Bhat, S., Brown, E. J., Banerjee, R. R., Wright, C. M., ... & Lazar, M. A. (2001). *The hormone resistin links obesity to diabetes*. *Nature*, 409(6818), 307-312.

Steppan, C. M., Bailey, S. T., Bhat, S., Brown, E. J., Banerjee, R. R., Wright, C. M., Patel, H. R., Ahima, R. S., & Lazar, M. A. (2001). *The hormone resistin links obesity to diabetes*. *Nature*, 409(6818), 307-312.  
<https://doi.org/10.1038/35053000>

Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. (2011). *Loss of exon identity is a common mechanism of human inherited disease*. *Genome Res*, 21(10), 1563–1571.

Steward, A., Adhya, S., & Clarke, J. (2016). *Folding and unfolding: From simple models to complex systems*. *Current opinion in structural biology*, 36, 28-35.

Sabouret, Pierre, Michel Farnier, et Etienne Puymirat. 2019. « Inhibiteurs de PCSK9 : quelle place dans la prise en charge actuelle des dyslipidémies ? » *La Presse Médicale* 48 (3, Part 1): 227-37.  
<https://doi.org/10.1016/j.lpm.2019.01.009>.

Taskinen, M. R., & Borén, J. (2015). *New insights into the pathophysiology of dyslipidemia in type 2 diabetes*. Atherosclerosis, 239(2), 483-495.

Training; <http://dx.doi.org/doi:10.1016/j.imr.2016.07.001>.

Ulrich-Merzenich, G., Panek, D., Zeitler, H., Wagner, H., & Vetter, H. (2009). *New perspectives for synergy research with the "omic"-technologies*. Phytomedicine, 16(6-7), 495-508.

Vane, J. R., & Botting, R. M. (2003). *The mechanism of action of aspirin*. Thrombosis research, 110(5-6), 255-258.

Van Gunsteren, W. F., & Berendsen, H. J. (1990). *Computer simulation of molecular dynamics: Methodology*

Voet, D., Voet, J. G., & Pratt, C. W. (2016). *Fundamentals of biochemistry: Life at the molecular level (5th ed.)*. Wiley

Vuksan, V., Sievenpiper, J. L., Koo, V. Y., Francis, T., Beljan-Zdravkovic, U., Xu, Z., & Vidgen, E. (2000). *American ginseng (Panax quinquefolius L) reduces postprandial glycemia in nondiabetic subjects and subjects with type 2 diabetes mellitus*. Archives of Internal Medicine, 160(7), 1009-1013.

Wales, D. J. (2003). *Energy landscapes: calculating pathways and rates*. International Reviews in Physical Chemistry, 22(2), 237-282.

Walters, W. P., & Namchuk, M. (2013). *Designing screens: how to make your hits a hit*. Nature reviews Drug discovery, 12(4), 259-262.

Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., ... & Ziyyat, A., Legssyer, A., Mekhfi, H., Dassouli, A., Serhrouchni, M., & Benjelloun, W. (1997). *Phytotherapy of hypertension and diabetes in oriental Morocco*. Journal of Ethnopharmacology, 58(1), 45-54.

Williamson, E. M. (2013). *Synergy and other interactions in phytomedicines*. Phytomedicine, 20(10), 807-810.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., ... & Kent, S. B. (1991). *Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease*. Science, 245(4918), 616-621.

Xie, W., Zhao, Y., & Zhang, Y. (2011). *Traditional Chinese medicines in treatment of patients with type 2 diabetes mellitus*. Evidence-Based Complementary and Alternative Medicine, 2011.

Xu, D., & Zhang, Y. (2012). *Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field*. Proteins: Structure, Function, and Bioinformatics, 80(7), 1715–1735.

Xu, H. E., et al. (1999). *Molecular recognition of fatty acids by peroxisome proliferator-activated receptors*. Molecular Cell, 3(3), 397–403.

Yamauchi, T., Kamon, J., Waki, H., Terauchi, Y., Kubota, N., Hara, K., Mori, Y., Ide, T., Murakami, K., Tsuboyama-Kasaoka, N., Ezaki, O., Akanuma, Y., Gavrilova, O., Vinson, C., Reitman, M. L., Kagechika, H., Shudo, K., Yoda, M., Nakano, Y., ... Kadokawa, T. (2001). *The fat-derived hormone adiponectin reverses insulin resistance associated with both lipodystrophy and obesity*. Nature Medicine, 7(8), 941-946. <https://doi.org/10.1038/90984>

Yates CM, Sternberg MJ. (2013). *The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein–protein interactions*. Journal of Molecular Biology, 425(21), 3949–3963.

Yazdanparast, R., et al. (2008). *Safranal, a constituent of Crocus sativus (saffron), attenuated cerebral ischemia induced oxidative damage in rat hippocampus*. Journal of Pharmacy and Pharmaceutical Sciences, 11(4), 394-404.

Yin, J., et al. (2008). *Efficacy of berberine in patients with type 2 diabetes mellitus*. Metabolism, 57(5), 712-717.

Zekkour, N., Bouhaddouda, N., Benmansour, A., & Mekarnia, M. (2015). *Antioxidant properties and total phenolic content of three varieties of carob tree leaves from Algeria*. Natural Product Research, 29(24), 2305-2308.



# TABLE DE MATIÈRES

<b>PRÉAMBULE</b>	<b>III</b>
<b>DEDICACES</b>	<b>V</b>
<b>REMERCIEMENTS</b>	<b>VI</b>
<b>LISTE DES ABREVIATIONS ET ACRONYMES</b>	<b>VIII</b>
<b>LISTE DES TABLEAUX</b>	<b>IX</b>
<b>LISTE DES FIGURES</b>	<b>X</b>
<b>LISTE DES PUBLICATIONS ET COMMUNICATIONS</b>	<b>XIII</b>
<b>RÉSUMÉ</b>	<b>XIV</b>
<b>SOMMAIRE</b>	<b>XV</b>
<b>INTRODUCTION GENERALE</b>	<b>1</b>
<b>PARTIE I. REVUE BIBLIOGRAPHIQUE</b>	<b>4</b>
CHAPITRE 1. BIOINFORMATIQUE ET MODÉLISATION MOLÉCULAIRE	5
1 .1 Introduction	5
1.2 Les protéines	5
1.3 Les niveaux d'organisation des protéines	6
1.3.1 Structure primaire	6
1.3.2 Structure secondaire	7
1.3.3 Structure tertiaire	8
1.4 L'importance de la structure tertiaire	9
1.5 Méthode expérimentale de la détermination des structures tertiaires	10
1.5.1 X-Ray diffraction	10
1.5.2 Résonance Magnétique Nucléaire	11
1.5.3 Microscope électronique	12
1.6 Prédiction de la structure tertiaire par la Bioinformatique	13
1.6.1 Méthode de modélisation par homologie	13
1.6.2 Méthode ab initio	14
1.6.3 Méthodes basées sur les fragments	14
1.6.4 Méthodes basées sur l'apprentissage profond	14
1.6.5 Validation de la structure tertiaire prédictive par bioinformatique	15
a. minimisation d'énergie	15
b. ramachandran plot	16
C. PROCHECK	17
1.7 Base de données de références des protéines PDB	18
1.7.1 Le fichier PDB	18
1.7.2 PDBsum	19
1.8 Les logiciels de visualisations DE STRUCTURES 3D	19

1.8.1 PyMol	19
1.8.2 VMD	20
2. Les variants ET LEURS IMPACT SUR LES PROTEINES	20
2.1 Les variants non codants : définition et classification	20
2.3 Impact des variants codants	21
3. prédition de l'impact des variants PAR LES Outils de bioinformatique	21
3.1 SIFT	21
3.2 PolyPhen	21
3.3 Analyse de conservation pour l'évaluation de l'impact des variants SUR LES PROTEINES	22
4. Découverte de médicament assistée par ordinateur (DMAO)	22
4.1 Conception de médicaments à base de ligands	24
4.2 Coception de médicaments à base de structures	24
4.3 Docking moléculaire	25
4.3.1 Principe d'algorithme de docking	25
4.3.2 Énergie de liaison et fonctions de Notation (Scoring Functions)	26
4.3.3 Validation de docking	27
a. Comparaison avec des structures expérimentales	27
b. Mesure de la performance des algorithmes de docking	27
c. Cross-docking et redocking	27
d. Composés interférentiels pan-essai	27
4.3.4 Analyse de résultats de docking	28
a. Analyse de RMSD et d'affinité de liaison	28
b. Analyse d'interaction protéine-ligand	29
• Liaison hydrogen	29
• Interaction de van Der Walls	30
• Interaction hydrophobic	31
• Interactions ioniques	31
• Interaction PI-cation	32
4.4 La règle de Lipinski : prédition de la ressemblance des ligands aux médicaments	33
4.5 Simulation par dynamique moléculaire	33
4.5.1 Les équations de mouvement	34
4.5.2 Intervalle de temps	35
4.5.3 Champ de force	35
a. AMBER (Assisted Model Building with Energy Refinement)	36

b. CHARMM (Chemistry at HARvard Molecular Mechanics)	36
c. GROMOS (GROningen MOlecular Simulation)	36
d. OPLS-AA (Optimized Potentials for Liquid Simulations-All Atom)	36
<b>4.5.4 Minimisation d'énergie</b>	<b>36</b>
4.5.5 Contrôle de la température et de la pression	37
4.5.6 La production de la dynamique	37
4.5.7 Données obtenus par les simulations de dynamique moléculaire	38
a. Root Mean Square Deviation (RMSD)	38
b. Root Mean Square Fluctuation (RMSF)	39
c. Rayon de giration (Rg)	39
d. Surface accessible au solvant (SASA)	39
<b>CHAPITRE 2. LE DIABÈTE DE TYPE 2</b>	<b>40</b>
1. Introduction	40
2. Le Diabète de type 2	40
2.1 Définition	40
2.2 Prévalence	41
2.3 Physiopathologie de diabète de type 2	41
2.4 Traitement de DT2 par les thiazolidinedions	42
2.4.1 Cible biologique de Thiazolidenedions	43
3. PPAR-γ : Le récepteur activé par les proliférateurs de peroxysomes gamma	44
Découverte de PPAR-γ	44
3.1 Rôle de PPAR-γ	44
3.2 Mécanisme d'activation de PPAR-γ	45
3.2.1 Activation par un agoniste complet	45
3.2.2 Activation de PPAR-γ par un agoniste partiel	46
4. La résistine	47
4.1 Découverte de la résistine	47
4.2 La résistine : médiateur entre DT2 et MCV	48
5. Complications cardiovasculaires de Diabète type 2	50
5.1 Rappel sur les lipoprotéines	51
5.1.1 HDL	51
5.1.2 VLDL	52
5.1.3 LDL	52
a. Régulation de taux de LDL	52
b. LDL oxydée	53
c. LDL et hypercholestérolémie	54

d. rôle de LDL dans l'Athérosclérose	55
6. PCSK9 : Nouvelle cible thérapeutique cardiovasculaire	55
6.1 Découverte de PCSK9	55
6.2 Rôle de PCSK9 dans la régulation de taux de LDL	56
6.3 Inhibiteurs de PCSK9	57
<b>CHAPITRE 4. LES PHYTO-COMPOSÉS DANS LE TRAITEMENT DE DT2 ET SES COMPLICATIONS ET BASE DE DONNÉES</b>	<b>59</b>
1. Phyto-composés	59
1.1 Caractéristiques	59
1.2 Activités biologiques	60
1.3 Activité antibactérienne	61
1.4 Activité anti candidose	61
1.5 Activité antifongique	61
1.6 Activité anti inflammatoire	61
1.7 Activité antidiabétique	62
1.8 Activité anti-oxydante	62
1.9 Activité anti cardiovasculaire	63
2. Phytothérapies	63
2.1 Phytothérapie de DT2	63
2.2 Phytothérapie d'hypercholestérolémie	64
2.3 Les plantes aromatiques et médicinales du Maroc	64
3. Bases de données des phyto composés internationales	65
<b>OBJECTIFS DE LA THÈSE</b>	<b>67</b>
<b>II. TRAVAUX DE RECHERCHE</b>	<b>68</b>
PARTIE 1. MODELISATION MOLECULAIRE DE LA RÉSISTINE HUMAINE IMPLIQUÉ DANS LE DIABÈTE DE TYPE II	70
1. Introduction	71
2. Matériel et Méthodes	72
2.1 Collecte de données	72
2.2 Prédiction des nsSNPs délétères	72
2.3 ANALYSE DE CONSERVATION	73
2.4 Prédiction des sites de modification post-traductionnelle	73
2.5 Prédiction du changement de stabilité des protéines	73
2.6 Balayage des SNPs UTR sur le site UTR	74
2.7 Analyse structurale	74
3. Résultats	76
3.1 SNP dataset	76
3.2 Prédiction des nsSNPs délétères	76

3.3 Analyse de la conservation	77
3.4 des sites de modification post-traductionnelle	78
3.5 L'impact des mutations délétères prédictes sur la stabilité de la protéine de résistin	78
3.6 Analyse structurale	79
3.6.1 Modélisation de la structure de la résistin humaine	79
3.7 Analyse des interactions	80
3.8 Prédiction de l'effet des SNPs situés dans la région UTR par le serveur UTRscan.	81
3.9 Interactions protéine-protéine à l'aide de STRING :	81
4. Discussion	82
5. Conclusion	84
<b>PARTIE 2. PPAR ET CRIBLAGE VIRTUEL</b>	<b>85</b>
1. Introduction	86
2. Méthodes	88
2.1 Préparation de la protéine	88
2.2 Bibliothèque des phyto-composés marocains	88
2.3 Criblage virtuel	88
2.4 Intégration de l'énergie de liaison dans la base de données en ligne	88
2.5 Analyse de l'interaction ligand-récepteur	88
2.6 Simulation de dynamique moléculaire	89
2.7 Calcul de l'énergie de liaison libre	89
3. Résultats	90
3.1 Criblage de la COMPOSES MAROCAINES et amarrage moléculaire	90
3.2 Propriétés de type médicament des résultats identifiés	92
3.3 Simulation Moléculaire Dynamique	93
3.4 Énergie de liaison libre	97
4. Discussion	98
5. Conclusion	100
<b>PARTIE 3. CONCEPTION DE LA PREMIÈRE BASE DE DONNÉES DES COMPOSÉS ISSUS DES PAMs</b>	<b>102</b>
1. Introduction	103
2. Méthodes	104
2.1 Conception du serveur et de la base de données	104
2.2 Collecte des composés	104
2.3 Recherche d'analogues	104
2.4 Base de données référencée	105
2.5 Analyse chimio-informatique	105

2.6 Vérification des données	105
2.7 Classification phyto-composé	105
2.8 Activités des composés	105
<b>3. Résultats</b>	<b>106</b>
3.1 Construction et contenu	106
3.2 Analogues des phyto-composés	107
3.3 Base de données référencée	107
3.4 Classification DES phyto-composéS	108
3.5 Analyse chémoinformatique	109
3.6 Activités des composés	110
<i>4.Discussion</i>	<b>111</b>
<i>5.Conclusion</i>	<b>113</b>
<b>PARTIE 4. EXPLOITATION DES DONNÉES DE LA MPDB VIS À VIS LA PCSK9</b>	<b>114</b>
1. Introduction	115
2. Méthodes	116
2.1 Collecte des renseignements sur La cible PCSK9	116
PDB	116
UniProt	116
Détermination du site actif	117
Préparation de la cible PCSK9	117
PyMOL	117
Reconstruction des acides aminés manquants (6U26) de PCSK9	118
Préparation de chimiothèque	119
Collecte des structures 3D à partir de la base de recherche MPDB	119
Collecte des phytocomposés à partir de MPDB	119
Chimiothèque	120
Criblage virtuel basé sur la structure	121
Analyse de l'énergie de liaison	121
Analyse des interactions ligand-protéine	122
Prediction de la ressemblance aux médicaments	122
<b>Résultats</b>	<b>123</b>
<i>Visualisation 3D de la protéine PCSK9</i>	<b>123</b>
Identification de Site actif	123
Position du site actif	123
Docking moléculaire de composé 16 l'inhibiteur connu DE PCSK9	124
Le diagramme 2D DE L'INTERACTION ENTRE L'INHIBITEUR CONNU ET LA PCSK9	125

Résultats de Criblage virtuel des phytocomposés marocaines	126
Analyse d'énergie de liaison et RMSD	126
Visualisation et analyse d'interaction ligand-Protéine	129
Le <i>Cannabis Sativa</i>	129
Le Diagramme 2D	130
Analyse des interactions ligand-protéine	130
Visualisation 3D des interactions ligand-protéine	131
Les phyto-composés marocains	131
Le diagramme 2D	132
Analyse des interactions ligand-protéine	132
Visualisation 3D des interactions ligand-protéine	133
Prédiction des propriétés ADME-Tox	133
La sélection de ces 5 composés repose sur la satisfaction d'au moins 3 critères de la règle de Lipinski.	134
Discussion	134
<b>CONCLUSION GÉNÉRALE</b>	<b>137</b>
<b>BIBLIOGRAPHIE GÉNÉRALE</b>	<b>139</b>
<b>TABLE DE MATIÈRES</b>	<b>162</b>