

République Algérienne Démocratique et Populaire
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université MOULOUD MAMMÈRI TIZI OUZOU
Faculté des sciences Département chimie

Mémoire

Présenté en vue de l'obtention du grade de

Master

Filière : Chimie

Spécialité : Chimie Pharmaceutique

Par

SALAH DEHBIA

Thème

Modélisation de la structure d'une protéine impliquée dans le cancer et de ses interactions avec quelques ligands

Soutenue le 30/09/2014 devant le jury composé de :

Mr	Saal Amar	Maitre de conférences-UMMTO	Président
Mme	Bouaziz-Terrachet Souhila	Maitre de conférences-UMMTO	Rapporteuse
Mme	Guechtouli Nabila	Maitre de conférences-UMMTO	Examinatrice

Remerciement

Mes remerciements les plus profonds vont à Madame **SOUHILA BOUAZIZ-TERRACHET**, Maitre de conférences à l'UMMTO, pour m'avoir proposé un sujet aussi passionnant.

Je lui suis vraiment reconnaissante pour son soutien constant, son expertise en modélisation moléculaire, la confiance qu'elle m'a accordée et pour avoir suivi et orienté l'avancement de ce travail.

Mes remerciements vont tout d'abord à Monsieur **AMAR SAAL**, Maitre de conférences à l'UMMTO d'avoir accepté de présider ce jury.

J'exprime mes remerciements chaleureux et respectueux à Madame **Nabila GUECHTOULI**, Maitre de conférences à l'UMMTO, pour l'honneur qu'elle me fait en acceptant de juger ce travail.

J'adresse un remerciement particulier à toutes les personnes du département de maths pour m'avoir permis d'accéder à des ressources informatiques nécessaires à la réalisation de certains calculs présentés dans ce manuscrit.

Je ne saurais oublier dans ces remerciements la Direction de la Faculté de Chimie, Monsieur le Doyen, le personnel technique et administratif ainsi que tous ceux qui, de près ou de loin ont contribué à la réalisation de ce travail.

Dédicaces

Du fond de mon cœur et avec l'intensité de mes émotions, Je dédie ce travail :

*A mes parents qui m'ont enseigné la persévérance dans mes études,
qui m'ont toujours été d'un grand secours par leur soutien
et leur encouragement pendant les moments difficiles, qu'ils trouvent ici le fruit de leur labeur
et le témoignage de mon grand amour*

Que le dieu me les protège.

A mes très chers frères Mohand Ou L'Hocine, Medjber

Ainsi à mes sœurs Ghania, Luiza, Houria

Pour leurs amours et leurs encouragements

Je formule les plus sincères vœux de bonheur et d'un avenir radieux

A mes amis les plus fidèles Ouiza, Chafiaa.

*A tous mes enseignants ainsi que tous mes collègues de promotion de
Master et spécialement Ouiza, Sarah, Hayet.*

Sommaire

I.	<i>Introduction générale</i>	1
	<i>Chapitre I : Rappels bibliographiques sur la protéine ADP-Ribosylation Factor 1 (ARF1) : cible potentielle dans le traitement du cancer du sein</i>	
I.	<i>Introduction</i>	4
II.	<i>Les protéines G</i>	4
III.	<i>Les familles de la protéine G</i>	5
III. 1	<i>La famille des protéines RAS</i>	6
III. 2	<i>La famille des protéines RHO</i>	7
III. 3	<i>La famille des protéines RAB</i>	7
III. 4	<i>La famille des protéines RAN</i>	7
III. 5	<i>La famille des protéines ARF</i>	8
	<i>III.5.1 Les protéines Arfs</i>	8
	<i>III.5.2 La protéine Arf1</i>	10
	<i>a. La Structure de la protéine Arf1</i>	10
	<i>b. Les inhibiteurs de la protéine Arf1</i>	11
	<i>Références bibliographiques</i>	12
	<i>Chapitre II : Les méthodes de la chimie théorique et du drug design</i>	
I.	<i>Introduction</i>	15
II.	<i>La mécanique quantique</i>	15
II.1	<i>Equation de Schrödinger</i>	15
	<i>I.3.1 Approximation de Born-Oppenheimer</i>	16
	<i>I.3.2 Approximation orbitale</i>	17
	<i>I.3.3 Approximation LCAO</i>	17
II.2	<i>Les différentes méthodes quantiques</i>	18
	<i>II.2.1 Méthode Ab-initio</i>	18
	<i>a. Formalisme Hartree-Fock</i>	18
	<i>1. Energie moyenne associée à un déterminant de Slater</i>	18
	<i>2. Méthodes de champ auto-cohérent et équations d'HF</i>	19
	<i>3. Equations de Roothann et Hall</i>	20

	<i>b. Calcul ab initio et les bases d'orbitales atomiques utilisées</i>	20
<i>II.2.2</i>	<i>Méthodes semi-empiriques</i>	22
	<i>a. Principe</i>	22
	<i>b. Versions basées sur l'approximation NDDO</i>	23
<i>III.</i>	<i>La mécanique moléculaire</i>	24
	<i>III.1. Bases et principes</i>	24
	<i>III.2 Champs de forces</i>	24
<i>IV.</i>	<i>Techniques du drug design</i>	25
	<i>IV.1. Approche 'receptor-based'</i>	25
	<i>IV.2. Alignement de structures tridimensionnelles de protéines</i>	26
<i>V.</i>	<i>Interactions intermoléculaires</i>	26
	<i>V.1 Interactions hydrophobes</i>	26
	<i>V.2 Interactions de Van Der Waals</i>	26
	<i>V.3 Interactions électrostatiques</i>	26
	<i>V.4 Interactions ioniques ou pont salin</i>	26
	<i>V.5 Liaisons hydrogènes</i>	27
	<i>V.6 Pont disulfure ou cystine</i>	27
	<i>Références bibliographiques</i>	28
	<i>Chapitre III : Modélisation de la structure de la protéine Arf1et de ses interactions avec une série de flavonoïdes</i>	
	<i>Introduction</i>	29
	<i>Partie A : Modélisation par alignement multiple de la structure de l'Arf1</i>	30
<i>I.</i>	<i>Objectif</i>	30
<i>II.</i>	<i>Méthodologie</i>	30
	<i>Partie B : Etude par docking moléculaire d'une série de flavonoïdes dans le site actif de l'Arf1</i>	35
<i>I.</i>	<i>Objectif</i>	35
<i>II.</i>	<i>Méthodologie</i>	35
	<i>II.1 Structure de départ</i>	35
	<i>II.1.1 Le récepteur</i>	35
	<i>II.2.2 Les molécules étudiées</i>	36
	<i>II.2 Méthodologie utilisée</i>	37
	<i>II.2.1 Préparation de la structure du récepteur</i>	37

<i>II.2.2</i>	<i>Localisation du site actif</i>	<i>37</i>
<i>II.2.3</i>	<i>Préparation des molécules étudiées</i>	<i>37</i>
<i>II.2.4</i>	<i>Protocole de docking</i>	<i>38</i>
	<i>a. Paramètres de la grille de calculs</i>	<i>38</i>
	<i>b. Paramètres de docking</i>	<i>38</i>
<i>III.3</i>	<i>Résultats et discussion</i>	<i>38</i>
<i>III.3.1</i>	<i>Validation des paramètres de docking</i>	<i>38</i>
<i>III.3.2</i>	<i>Sélection de la conformation dite bioactives</i>	<i>39</i>
<i>III.3.3</i>	<i>Mode de liaison des composes étudiés</i>	<i>40</i>
<i>III.3.4</i>	<i>Interactions mises en jeu</i>	<i>43</i>
<i>III.4</i>	<i>Conclusion</i>	<i>48</i>
	<i>Références bibliographiques</i>	<i>50</i>
	<i>Conclusion générale et Perspectives</i>	<i>51</i>

Introduction générale

La recherche et la synthèse de nouveaux composés chimiques et biochimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire, qui est une technique permettant de construire, visualiser, manipuler et d'analyser la structure et la réactivité d'un système de particules. L'évolution des outils informatiques ont permis le développement important de nouveaux outils spécialement dédiés à la modélisation moléculaire. Grâce à cette évolution et aux progrès en biologie moléculaire, les chercheurs parviennent à simuler l'action de substances thérapeutiques, d'où la naissance des techniques du «drug design».

Les techniques dites de drug design, ayant pour objectif la conception de molécules actives et utilisant les méthodes de la modélisation moléculaire, comprend deux types d'approches. Celle du « ligand-based » et celle du « receptor-based ». L'approche « ligand-based » utilise des techniques de calculs basées sur la connaissance du ligand, comme le QSAR (Quantitative Structure Activity Relationships) qui consiste à disposer d'une liste de molécules présentant une activité sur une cible donnée et ensuite à élaborer des modèles statistiques permettant de prédire l'activité d'une nouvelle molécule. L'autre approche : « receptor-based », basée sur la connaissance de la cible biologique, consiste à identifier quelles molécules peuvent se lier sur le site actif d'une protéine dans le but d'inhiber son action.

Les objectifs principaux de nos travaux consistent en une modélisation de la structure de la protéine Arf1 afin de localiser les différents motifs biologiques et le site actif de cette protéine puis l'étude du mode de reconnaissance d'une série de flavonoïdes avec les résidus du site actif de l'Arf1. Ces études ont été essentiellement réalisées au moyen des méthodes de la chimie théorique et du docking moléculaire.

La protéine ADP-Ribosylation Factor 1 (Arf1), une petite protéine G, existe sous six isoformes, Arf1-Arf6, répartis en trois classes. L'Arf1 et l'Arf6 sont les plus étudiées à cause de leurs rôles biologiques. Contrairement à l'Arf6 qui joue un rôle important dans la cellule, l'Arf1 est surexprimée dans les cellules invasives du cancer du sein et elle est impliquée dans la migration et la prolifération des cellules cancéreuses [1]. Ainsi, Arf1 constitue une cible potentielle dans le traitement anticancer du sein.

Après une introduction générale où nous présentons le but du travail, ce manuscrit s'articule autour de trois chapitres suivis d'une conclusion générale.

Dans le premier chapitre, nous reportons les résultats de notre revue bibliographique sur la protéine ADP-Ribosylation Factor 1 (Arf1) et de ses inhibiteurs.

Le second chapitre est réservé à la présentation des méthodes de la modélisation moléculaire que nous avons utilisée dans ce travail. Dans un premier temps nous avons définis: la mécanique quantique et la mécanique moléculaire. Ensuite, nous avons présenté un outil de drug design qui nous a permis d'étudier les différents modes de reconnaissance des ligands et la protéine Arf1 en l'occurrence le docking moléculaire.

Dans le troisième chapitre, nous avons présenté d'une part l'étude de modélisation par alignement multiple de trois structures cristallographiques de l'Arf1. L'objectif de cette simulation est de définir les différents motifs biologiques composant la structure de cette protéine. D'autre part, nous avons présenté les simulations de docking moléculaire, pour déterminer les différents modes de liaison d'une série de vingt flavonoïdes dans le site actif de l'Arf1.

La conclusion générale porte sur les principaux apports de nos travaux. Nous présentons également les perspectives que nos résultats laissent entrevoir afin d'améliorer les modes de liaisons obtenus.

Références bibliographiques

[1] Boulay PL, Cotton M, Melançon P, and Claing A, ARF1 controls the activation of the PI3K pathway to regulates EGF dependent growth and migration of breast cancer cells, *The Journal of Biological Chemistry*, 2008, 283(52) : 36425-36434.

CHAPITRE I

Rappels bibliographiques sur la protéine ADP-Ribosylation Factor 1 (ARF1) : cible potentielle dans le traitement du cancer du sein

I. Introduction

Le cancer est une maladie chronique de cellule qui peut être mortelle et qui apparait lorsque l'ADN des cellules est endommagé par des agents mutagènes. Quand la cellule devient cancéreuse, elle perd ses capacités de réparation mais elle ne meurt pas. Elle continue de se développer et finira par former une masse qu'on appelle " tumeur" [1]. Une tumeur peut être bénigne (non cancéreuse) ou maligne (cancéreuse).

Le cancer se subdivise en plusieurs classes selon les organes touchés : Les poumons, pancréas, côlon, les seins. Ce dernier est une tumeur maligne qui se développe au niveau du sein, c'est la forme la plus rencontrées chez la femme [2].

Dans le cadre des objectifs visés par les travaux présentés dans ce mémoire, nous résumons dans ce chapitre l'état actuel des connaissances concernant les structures des protéines impliquées dans le processus du cancer du sein.

II. Les protéines G

Les protéines G appartiennent à la famille des GTPase. Ces protéines ont la masse moléculaire généralement comprise entre 20 000 et 30 000 daltons. Elles peuvent être liées à la membrane, comme c'est le cas des grandes protéines hétérotrimériques *G*, qui sont constituées de sous-unités alpha (α), beta (β) et gamma (γ). Elles peuvent aussi être des monomères non liés à la membrane, comme c'est le cas des petites protéines *G*. Cette appellation revient au fait qu'elles sont composées d'une seule sous-unité très petite (environ 20 kDa).

Les petites protéines *G* peuvent se trouver sous deux conformations actives et inactives. Elles sont dites actives si elles sont liées au guanosine triphosphate (GTP) ou inactives lorsqu'elles sont liées au guanosine diphosphate (GDP) dont la structure chimique est représentée sur la figure 1. Elles passent d'une conformation à une autre suivant un cycle GDP/GTP comme c'est illustré dans la figure 2. Néanmoins, ce processus est extrêmement lent, c'est pourquoi il y'a intervention des facteurs d'échanges nucléotidiques (en anglais *guanine nucleotide exchange factor* GEF). Ces derniers stimulent la dissociation du GDP liée à la protéine *G*, donc la forme inactive, pour former la GTP liée à la protéine *G*, la forme active [3]. Le cas inverse c'est-à-dire le passage de la forme active (protéine *G* liée au GTP) à la forme inactive (protéine *G* liée au GDP), se fait par hydrolyse du GTP en GDP à l'aide d'une protéine nommée *GTPase activating protein* (GAP) [4]. Le passage d'une forme à une autre permet aux petites protéines *G* le déclenchement ou l'inhibition des réactions biochimiques dans

la cellule [5]. Les petites protéines G permettent le transfert d'informations à l'intérieure de la cellule, de ce fait elles sont impliquées dans la transduction de signaux.

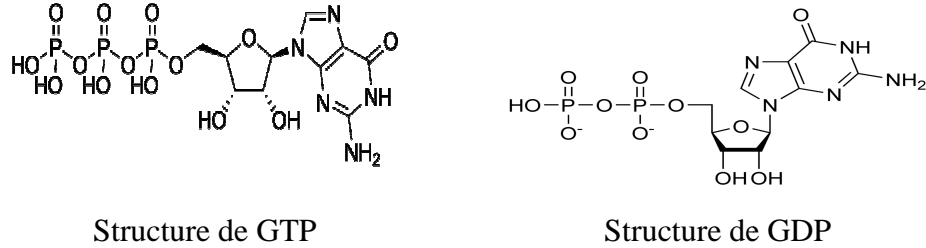


Figure 1 : structures chimiques des deux composés GTP et GDP

Nous présentons dans la figure suivante le cycle GDP/GTP des petites protéines G (PPG), passant d'une forme inactive à la forme active.

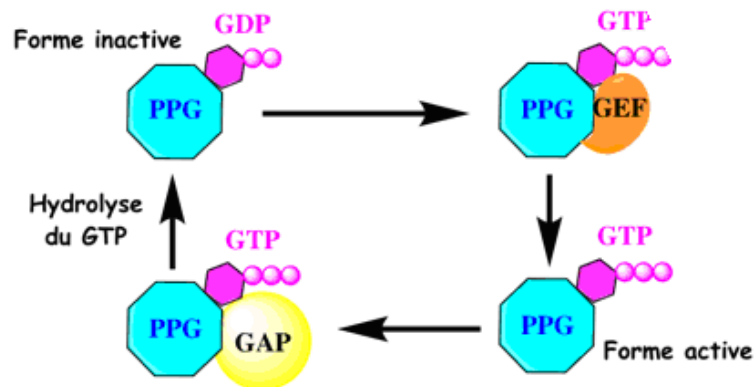


Figure 2 :cycle GDP/GTP des petites protéines G

III. Les familles de la protéine G

Les protéines G sont divisées en cinq familles selon leur homologie de séquence en acides aminés et leurs rôles primaires dans la cellule. Il s'agit donc des 5 familles suivantes : RHO, RAS, RAB, RAN et ARFs.

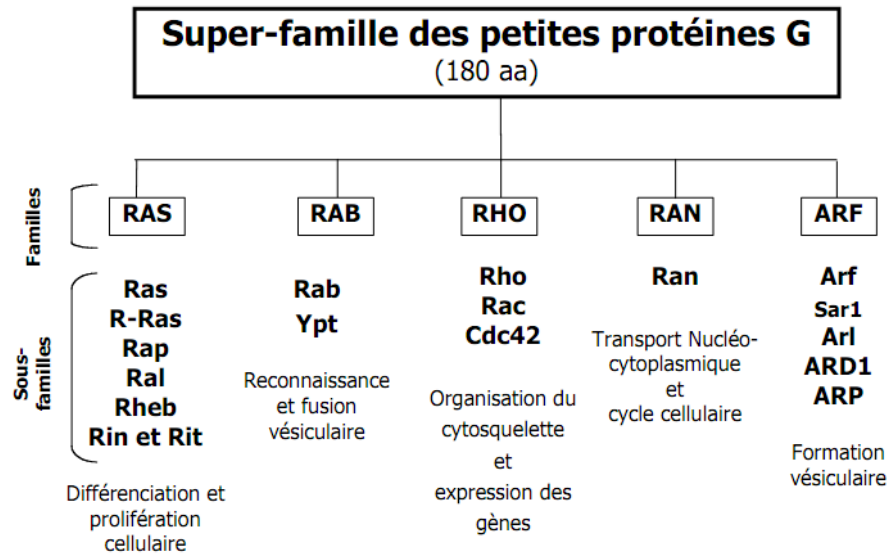


Figure 3 : Différents membres de la famille des petites protéines G

III.1. La famille des protéines RAS

La famille RAS est composée de plusieurs sous-familles, parmi les quelles on peut citer Ras, R-Ras, Rap, Ral, Rheb, Rin et Rit. Dans la sous-famille Ras ont trouve trois principales isoformes qui ont plus de 85% d'homologie de séquence en acides aminés : Harvey-Ras (H-Ras), Kirsten-Ras (K-Ras) et Neuroblastoma-Ras (N-Ras). Ces trois isoformes sont exprimées de manière ubiquitaire. Depuis la résolution des premières structures de H-Ras en 1988 [6-8], plusieurs structures de petites protéines G sont publiées. Ces structures sont composées d'une chaîne polypeptidique simple de 189 acides aminés, liée à la membrane par une queue lipidique constituée par des isoprénoïdes et des acides gras saturés.

Des études récentes ont montré que H-Ras et N-Ras n'interviennent pas dans le développement embryonnaire alors que K-Ras est essentiel [9]. Ces petites protéines sont très importantes dans de nombreux processus biologiques tels que la différenciation cellulaire et la prolifération. L'expression anormale de ces petites protéines conduit au développement de cancer du système lymphatique [10].

III.2. La famille des protéines RHO

La famille de GTPases **Rho** comprend six branches : Cdc42, Rac, Rho, RhoBT, Rnd et Miro. Comme dans le cas des protéines Ras, les protéines Rho sont exprimées de manière ubiquitaire et sont retrouvées dans de nombreuses espèces animales [11]. Ces protéines jouent un rôle dans de nombreux processus cellulaires tels que la réorganisation du cytosquelette, la progression du cycle cellulaire, l'expression génique, la polarité cellulaire, la migration et la transformation cellulaire [12]. En plus à toutes ces propriétés, les protéines Cdc42 sont impliquées dans l'inhibition de la migration et d'invasion dans les cancers [13].

III.3. La famille des protéines RAB

La famille Rab est la plus grande et la plus variée de la superfamille des petites protéines G Ras avec plus de 60 membres, parmi lesquels on peut citer La protéine neuronale humaine Rab6B, Rab et Ypt. Elles sont identifiées chez la levure, puis chez le rat. Dans ces protéines l'hydrolyse du GTP en GDP, c'est-à-dire le passage de la forme active à la forme inactive, est couplée à l'association et la dissociation des membranes. Ces protéines interviennent dans la régulation du transport vésiculaire [14].

III.4. La famille des protéines RAN

Les petites protéines G Ran, sont localisées dans le noyau et le cytoplasme des cellules et sont associées aux membranes. La forme inactive de cette protéine, c'est-à-dire Ran liée au GDP, se trouve en grande concentration dans le cytoplasme du fait de la présence de la protéine GAP qui aide à l'hydrolyse comme c'est déjà expliqué en haut. Ainsi, tous les Ran-GTP qui arrivent dans le cytoplasme vont être transformés en Ran-GDP. En revanche, la forme active, Ran-GTP, se trouve en grande concentration dans le noyau. Tous les Ran-GDP qui arrivent dans le noyau seront transformés en Ran-GTP du fait de la présence d'un facteur d'échange (GEF) qui aide à la transformation du Ran-GDP en Ran-GTP. Ces protéines interviennent dans le contrôle des phénomènes de fusion membranaire, le trafic des vésicules intracellulaires et la régulation du cycle cellulaire comme elles peuvent être impliquées dans le mécanisme de condensation de la chromatine [15].

III.5. La famille des protéines ARF

Cette sous-famille des GTPases a été découverte, en 1984, pour son rôle de co-facteur de l'activité de la toxine du choléra, responsable en grande parties des fortes diarrhées caractéristiques des symptômes du choléra [16]. Ce co-facteur a été nommé ADP-Ribosylation factor (ARF). Elle est composée des membres Arf, Sar1, Arl, ARD1 et ARP. Cette sous famille est caractérisée par la présence d'une extension N-terminale plus ou moins longue et divergente en séquence et en longueur. Les protéines ARF ont en commun un certain nombre de résidus comme leucine 34, tyrosine 81, l'aspartate 93, l'arginine 99 et le tryptophane 78. Il est bien de noter que ces résidus ne sont pas conservés chez les autres familles de petites protéines G.

III.5.1. Les protéines Arfs

Les protéines ADP-Ribosylation Factor (ARF_s) présentent 54% d'homologie de séquence avec les protéines SAR1 [4]. On les trouve chez toutes les cellules eucaryotes (homme, souris, rat, vache...) et sont absentes chez les bactéries. Elles forment une sous-famille de six isoformes soient d'Arf1 à Arf6, où Arf2 est absent chez l'humain. Ces protéines sont classées, chez les mammifères, selon leur taille et leur homologie de séquence en trois groupes distincts, soit la classe I (Arf1-Arf3), la classe II (Arf4 et Arf5) et la classe III (Arf6), où Arf6 est la plus divergente en termes de séquence en acides aminés comme le montre la figure 4.

Les isoformes Arf1-2 et 3 sont exprimés dans la levure, les tissus bovins, de rat et de souris, et n'ont pas été retrouvées chez d'autres mammifères. Chacune de ces classes assure des fonctions distinctes dans le transport cellulaire.

Classe I: Cette classe, regroupant les isoformes Arf1, Arf2 et Arf3, est distribuées majoritairement dans l'appareil de Golgi. Les trois isoformes de cette classe sont impliquées dans le contrôle de la formation de vésicules lors du transport intracellulaire. Les gènes Arf1 et Arf2 sont homologues à 96% et ils présentent seulement 54% d'homologie de séquence avec le gène Arf3.

Classe II : les isoformes regroupant cette classe soient Arf4 et Arf5, sont distribuées majoritairement dans l'appareil de Golgi. Ces deux isoformes assurent le transport cellulaire.

Classe III : Dans cette classe on trouve uniquement l'isoforme Arf6, l'une des protéines Arfs les plus étudiée. Elle est localisée dans la membrane plasmique et dans les membranes endosmales où elle assure le contrôle de la synthèse de lipides. Cette isoforme est impliquée dans l'endocytose des

récepteurs membranaire et dans l'organisation du cytosquelette d'actine [17]. Récemment, il a été démontré qu'Arf6 contrôle le processus d'invasion et de migration cellulaire dans le cancer du sein [18].

Nous présentons dans la figure suivante les séquences en acides aminés des six isoformes de Arf soient Arf1 à Arf6. En vert sont colorées les régions conservées et en jaune celles qui ne le sont pas.

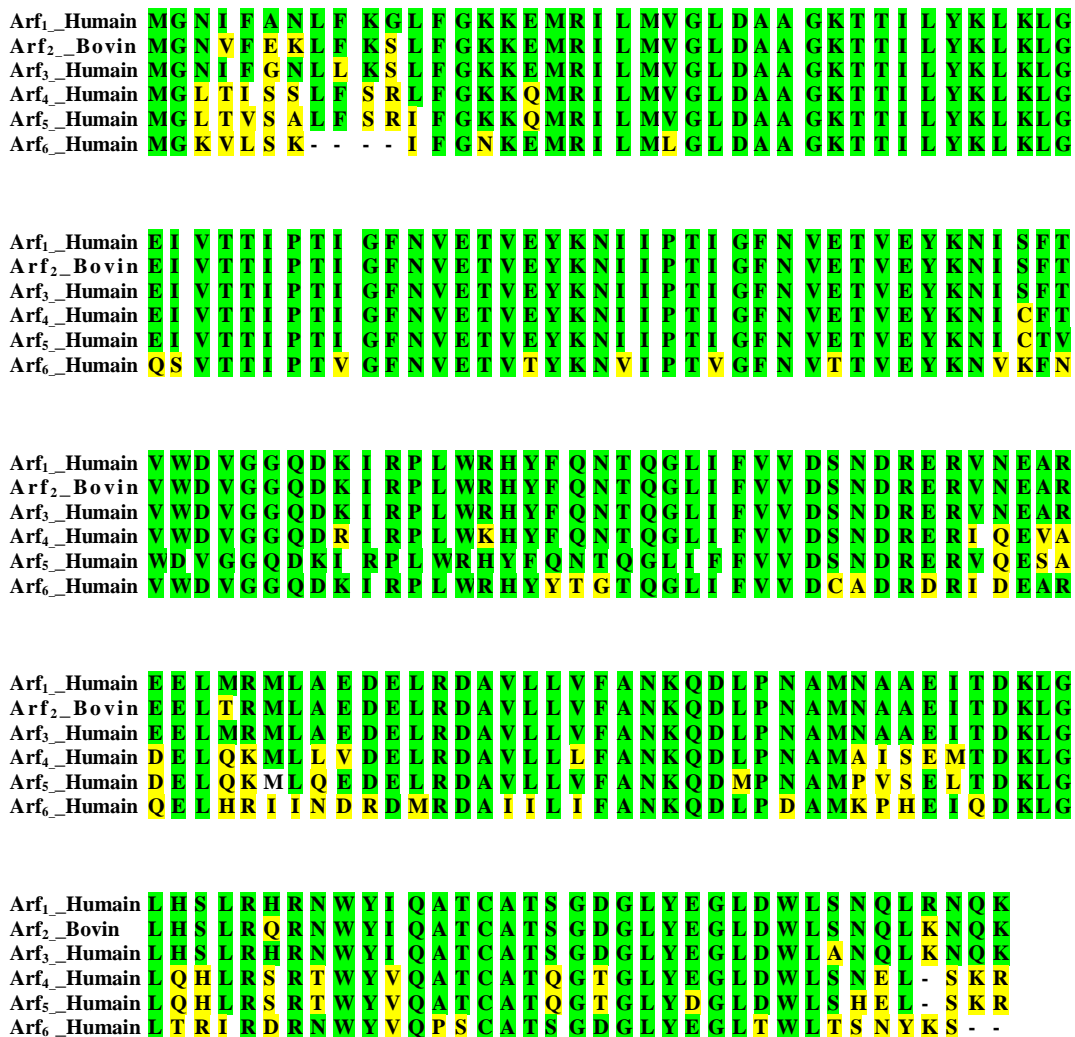


Figure 4 :séquences en acides aminés des six isoformes de l'ARF. Les parties colorées en vert représentent les régions conservées, les parties différentes en acides aminés sont colorées en jaune [19].

III.5.2. La protéine Arf1

La protéine Arf1 est l'isoforme la plus abondante de la famille Arf. Elle représente 90% des protéines de cette sous famille présentes dans une cellule. Elle a été identifiée au niveau de l'appareil de Golgi où elle exerce plusieurs fonctions comme le transport vésiculaire, la synthèse des lipides membranaires, la réorganisation du cytosquelette d'actine et la ségrégation des chromosomes, un événement clé de la division cellulaire [4]. Au cours des dernières années, il a été démontré que l'Arf1 est surexprimée dans les cellules invasives du cancer du sein et elle est impliquée dans la migration et la prolifération des cellules cancéreuses [20]. Ainsi, Arf1 constitue une cible potentielle dans le traitement anti-cancer du sein.

a. La Structure de la protéine Arf1

La petite protéine G Arf1, constituée de 181 acides aminés, présente 67% d'identité de séquence avec l'Arf6, c'est la plus différente de celle-ci en la comparant aux autres protéines Arfs (figure 4). Contrairement aux autres petites protéines G, la protéine Arf1 possède une glycine en position 2. Ces protéines possèdent deux régions importantes pour leur activité biologique, switch I et switch II. La région Switch I est située entre les acides aminés 45 et 54, la région switch II se situe entre les acides aminés 70 et 80. La partie entre ces deux régions, c'est-à-dire les acides aminés de 55 à 69, s'appelle interswitch. Le passage de la forme inactive à la forme active c'est-à-dire l'échange du GDP par le GTP, catalysé par un facteur d'échange (Guanine exchange factor; GEF), induit des changements conformationnels au niveau des régions switch I et switch II (voir la figure 5). ARNO est l'un de ces facteurs d'échange (GEF) capable d'activer Arf1 [21].

La figure suivante montre le changement conformationnel, l'interswitch et l'hélice N-terminal, de la protéine Arf1 lors de son passage de la forme inactive à la forme active.

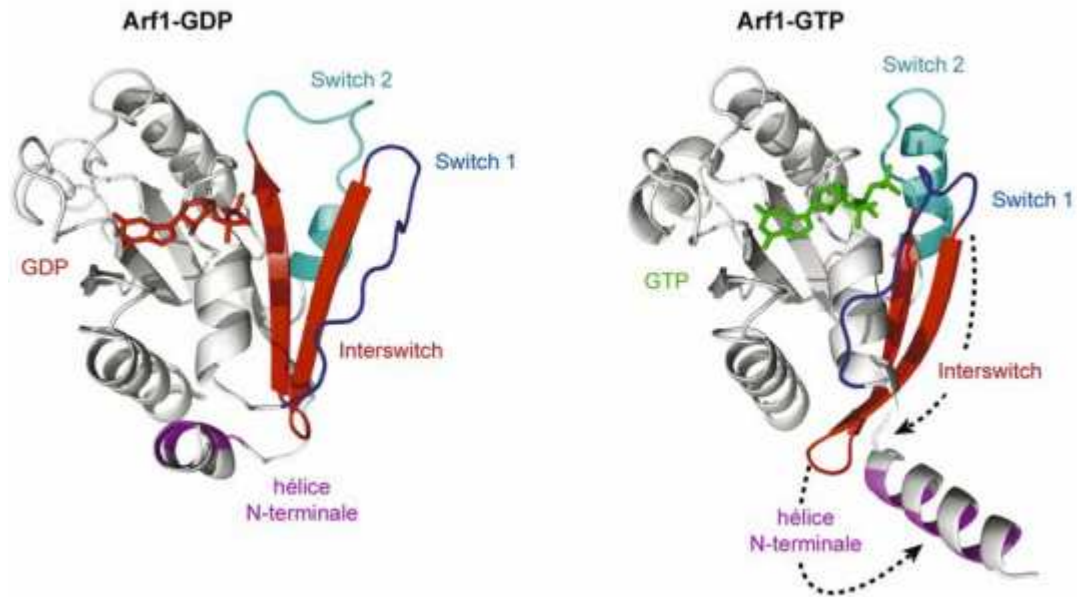


Figure 5 : Structures de la protéine Arf1 liée au GDP et au GTP [22]. La liaison du GTP provoque le glissement de l’interswitch qui déplace l’hélice N-terminale de la poche hydrophobe qu’elle occupe lorsqu’Arf1 est lié au GDP [23].

b. Les inhibiteurs de la protéine Arf1

A ce jour, trois inhibiteurs d’Arf1 sont connus. Le premier, la bréfeldine A (BFA) dont la structure chimique est représentée sur la figure 6a, a été identifiée il y’a plus de 65 ans, mais son rôle d’inhibiteur pour l’Arf n’a été élucidé que 40 ans plus tard [24]. La BFA, une toxine naturelle, cible spécifiquement la forme Arf1 et pas Arf6 liée au GDP et le domaine sec7 [25]. Cet inhibiteur n’empêche pas l’interaction d’Arf1 avec son GEF, mais elle les piège dans une interaction non fonctionnelle [26]. Récemment, la découverte par criblage in silico puis la synthèse d’un nouveau composé inhibiteur de l’activation des Arfs, le LM11 (voir la figure 6b), permet de bloquer l’activité catalytique des Arfs GEF sensibles et résistantes à la BFA [27]. Des résultats ont démontré que le LM11 inhibe Arf1-5 mais pas Arf6. Ces dernières années un chercheur japonais et ses collaborateurs développent un troisième inhibiteur l’AMF-26, dont la structure chimique est représentée sur la figure 6c, par des analyses de *pull-down* [28]. La même équipe montre, par des études de modélisation moléculaire, que cet inhibiteur occupe le même site actif que la BFA.

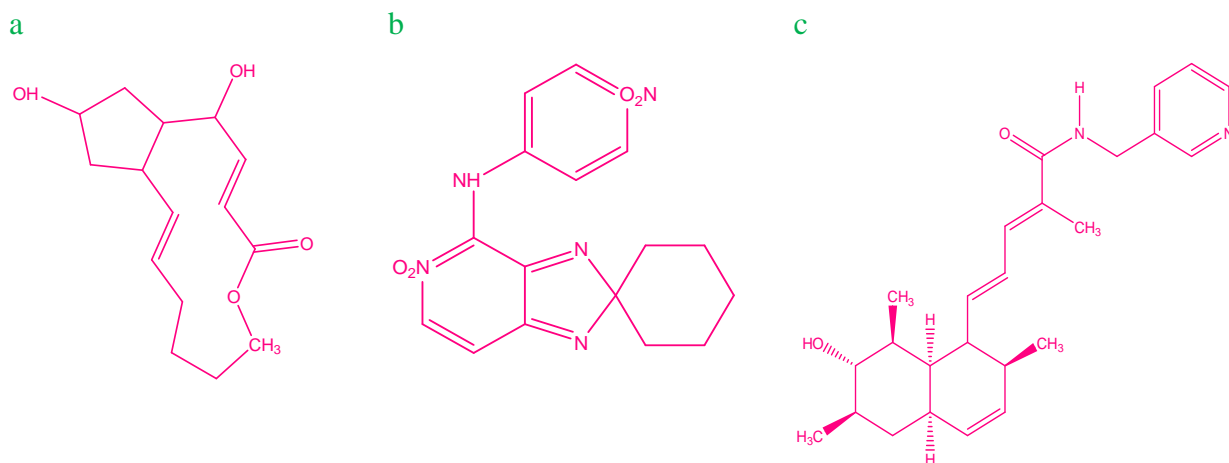


Figure 6 :structures chimiques des trois inhibiteurs d'Arf 1, a-BFA, b-LM11 et c-AMF26

Références bibliographiques

- [1] Lyonnet S, Prédilections génétiques aux cancers : Actualités et Perspectives en 2005. *Medicine Science*. 2005, 21(11) :962-968.
- [2] Pierre-Luc B, L'étude du rôle d'ARF1 dans la migration et la prolifération des cellules du cancer du sein. Ph Doct : Pharmacologie : Université de Montréal., 2011.
- [3] Joubert A, Etude biochimique et structurales des grands facteurs d'échange des protéines Arf_s. M : Sciences de la vie et de la terre : Ecole pratique des hautes études., 2007.
- [4] Ménétrey J, Etude structurale des petites protéines G : Rap 2A dans un complexe non catalytique avec le GTP et Arf₆ en complexe avec le GDP. Doct : Biophysique moléculaire : L'université Paris6., 2000.
- [5] Eric K, Schwartz J, Jessel T, Principe de la science neural. 4^{ème} ED. McGraw-Hill, New York., 2000.
- [6] Devos AM, Tong L, Milburn MV, Matias PM, Jancarik J, Noguchi S, Nishimura S, Miura K, Ohtsuka E, Kim SH, Three-dimensional structure of an oncogene protein: catalytic domain of human c-H-ras. *Science*. 1988, 239(4842):888-893.
- [7] Milburn MV, Tong L, deVos AM, Brünger A, Yamaizumi Z, Nishimura S, Kim SH, Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*. 1990, 247(4945):939-945.
- [8] Pai EF, Krengel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A, Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J*. 1990, 9(8):2351-2359.
- [9] Johnson L, Greenbaum D, Cichowski K, Mercer K, Murphy E, Schmitt E, Bronson RT, Umanoff H, Edelman W, Kucherlapati R, Jacks T, K-ras is an essential gene in the mouse with partial functional overlap with N-ras. *Genes Dev*. 1997 Oct 1, 11(19):2468-2481.
- [10] Downward J, Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer*. 2003 Jan, 3(1):11-22.
- [11] Goffinet M, Conception et obtention d'un anti corps spécifique des formes activées des RHO GTPase, Doct : Science de la vie et de la terre : Université Toulouse III-Paul-Sabatier., 2007.
- [12] Lewis-Saravalli S, ARF1 Contrôle la migration des cellules hautement invasives du cancer du sein via Rac1. M : Pharmacologie : Université de Montréal., 2012.
- [13] Pierre-Luc B, L'étude du rôle d'ARF1 dans la migration et la prolifération des cellules du cancer du sein. Ph Doct : Pharmacologie : Université de Montréal., 2011.
- [14] Magman R, Mécanisme de l'internalisation du récepteur CCK2 bases pharmaceutiques et structurales. Doct : Innovation pharmacologique : université Toulouse III-Paul Sabatier., 2011.
- [15] Bischoff FR, Postingsl H, Mitotic regulator protein RCC1 is complexed with anuclear-ras-related polypeptide. *Proc Natl Acad Sci USA*. 1999, 88:10830-10834.
- [16] Enomoto K and Gill DM, Cholera toxin activation of adenylate cyclase. Roles of nucleoside triphosphates and a macromolecular factor in the ADP ribosylation of the GTP-dependent regulatory component. *J Biol Chem*. 1980, 255(4):1252-1258.
- [17] Alazheq MA, Analyse du rôle cytohémine - 1 chez le neutrophile humain : Aspects signalétiques et fonctionnels. Doct : Microbiologie – Immunologie : Université Laval, 2011.
- [18] Hashimoto S, Onodera Y, Hashimoto A, Tanaka M, Hamaguchi M, Yamada A, Sabe H, Requirement for Arf6 in breast cancer invasive activities. *Proc Natl Acad Sci U S A*. 2004, 101(17): 6647-6652
- [19] Tsuchiya M, Price SR, Tsai SC, Moss J, Vaughan M, Molecular identification of ADP-ribosylation factor mRNAs and their expression in mammalian cells. *J Biol Chem*. 1991, 266:2772-2777.
- [20] Boulay PL, Cotton M, Melançon P, and Claign A, ARF1 controls the activation of the PI3K pathway to regulate EGF dependent growth and migration of breast cancer cells. *The Journal of Biological Chemistry*. 2008, 283(52): 36425-36434
- [21] Pasqualato S, Renault L, and Cherfils J, Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep*. 2002, 3(11): 1035-1041.
- [22] Amor JC, Harrison DH, Kahn RA, Ringe D, Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature*. 1994, 372:704-708.
- [23] Gillingham AK, Munro S, The small G proteins of the Arf family and their regulators. *Annu Rev Cell Dev Biol*. 2007, 23: 579-611.
- [24] Singleton VL, Bohonos N, and Ullstrup AJ, Decumbin, a new compound from a species of *Penicillium*. *Nature*. 1958, 181(4615): 1072-1073.
- [25] Peyroche A, Antonny B, Robineau S, Acker J, Cherfils J, Jackson CL, Brefeldin A acts to stabilize an abortive ARF-GDP-Sec7 domain protein complex: involvement of specific residues of the Sec7 domain. *Mol Cell*. 1999, 3(3): 275-285.
- [26] Peyroche A, Antonny B, Robineau S, et al. Brefeldin A acts to stabilize an abortive ARF-GDP-Sec7 domain protein complex: involvement of specific residues of the Sec7 domain. *Mol Cell*. 1999, 3: 275-285.

CHAPITRE I :Rappels bibliographiques sur la protéine ADP-Ribosylation Factor 1 (ARF1) : cible potentielle dans le traitement du cancer du sein

[27] Viaud J, Zeghouf M, Barelli H, Zeeh JC, Padilla A, Guibert B, Chardin P, Royer CA, Cherfils J, and Chavanieu A, Structure-based discovery of an inhibitor of Arf activation by Sec7 domains through targeting of protein–protein complexes.PNAS.2007, 104:10370-10375.

[28] Ohashi Y, Iijima H, Yamaotsu N, Yamazaki K, Sato S, Okamura M, Sugimoto K, Dan S, Hirono S, Yamori T, AMF-26, a novel inhibitor of the Golgi system, targeting ADP-ribosylation factor 1 (Arf1) with potential for cancer therapy.J Biol Chem. 2012 Feb 3, 287(6):3885-3897.

CHAPITRE II

Les méthodes de la chimie théorique et du drug design

I. Introduction

L'élaboration de modèles moléculaires s'effectue actuellement au moyen de deux types d'approches théoriques de base radicalement différentes et qui sont :

La mécanique quantique et la mécanique moléculaire

Dans ce chapitre, nous présentons d'une part les méthodes théoriques utilisées, d'autre part, les techniques du drug design qui nous ont servi dans l'élaboration des modèles mathématiques de prédiction des énergies de liaisons des molécules étudiées avec les résidus du site actif de l'Arf1.

II. La mécanique quantique

Les méthodes issues de la mécanique quantique sont basées sur la recherche de la meilleure répartition des électrons dans le champ engendré par les noyaux du système et cela en résolvant l'équation de Schrödinger.

II.1. Equation de Schrödinger

Les propriétés chimiques d'un système quantique sont déterminées par la fonction d'onde Ψ . Cette dernière dépend des variables d'espace de chaque particule et elle est la solution de l'équation de Schrödinger[1], qui s'écrit pour les états stationnaires :

$$\hat{H}\Psi = E\Psi \quad (1)$$

\hat{H} est l'opérateur hamiltonien moléculaire associé à l'énergie totale E du système étudié. Cethamiltonien ne prend pas en considération les interactions entre les électrons et les champs extérieurs au système. Il s'écrit pour un système polyélectronique constitué de n électrons et de N noyaux :

$$\hat{H} = \sum_{j=1}^N -\frac{\hbar^2}{2M_j} \Delta_j + \sum_{i=1}^n -\frac{\hbar^2}{2m_e} \Delta_i + \sum_{j=1}^N \sum_{j' > 1}^N Z_j Z_{j'} \frac{e^2}{r_{jj'}} - \sum_{j=1}^N \sum_{i=1}^n Z_j \frac{e^2}{r_{ij}} + \sum_{i=1}^n \sum_{i' > 1}^n \frac{e^2}{r_{ii'}} \quad (2)$$

Énergie cinétique
des noyaux

Énergie cinétique
des électrons

Énergie
derépulsion entre
les noyaux

Énergie
d'attraction
noyau-électron

Énergie de
répulsion entre les
électrons

Δ_i et Δ_j sont les *laplaciens* de l'électron i et du noyau j . Δ_i est donné par la relation :

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

m_e est la masse de l'électron, M_j, Z_j et $Z_{j'}$ sont respectivement la masse du noyau j , les numéros atomiques des atomes j et j' .

$r_{ii'}$ est la distance entre deux électrons i et i' , $r_{jj'}$ est la distance entre deux noyaux j et j' .

r_{ij} la distance entre un électron i et un noyau j et e la charge de l'électron.

$\hbar = \frac{h}{2\pi}$ avec h : la constante de planck.

Cependant, l'équation (1) ne peut être résolue analytiquement que pour des systèmes monoélectroniques, d'où la nécessité d'introduire des approximations et de résoudre le système numériquement [2].

II.1.1. Approximation de Born-Oppenheimer

Cette approximation permet de séparer le mouvement des électrons de celui des noyaux en négligeant l'énergie cinétique de ces derniers [3]. Elle se base sur le fait que les électrons sont beaucoup plus légers et ils beaucoup plus rapidement que les noyaux. Cela revient à considérer le troisième terme de l'équation 2 constant pour une géométrie donnée.

Dans le cadre de cette approximation, l'hamiltonien moléculaire peut se diviser en une partie nucléaire H_N et une partie électronique H_e .

La partie électronique est donnée par la relation :

$$\hat{H}_e = \sum_{i=1}^n h^c(i) + \sum_{i=1}^n \sum_{i'=1}^n \frac{e^2}{r_{ii'}} \quad (4)$$

Avec :

$$h^c(i) = -\frac{\hbar^2}{2m_e} \Delta_i - \sum_{j=1}^N Z_j \frac{e^2}{r_{ij}} \quad (5)$$

$h^c(i)$ est l'hamiltonien de cœur relatif à l'électron i , c'est un opérateur monoélectronique qui décrit le mouvement d'un électron dans le champ des N noyaux. Le dernier terme de l'équation (4) empêche la séparation de l'équation de Schrödinger en n équations monoélectroniques d'où la nécessité d'introduire une autre approximation sur l'expression de la fonction d'onde.

II.1.2. Approximation orbitale

La fonction d'onde électronique doit tenir compte du spin de l'électron, donc en plus des trois coordonnées spatiales définies par le vecteur espace, une quatrième coordonnée est nécessaire pour définir l'état du moment de spin [4].

$$\chi(i) = \Psi(i)\alpha(i) \quad (6)$$

$$\bar{\chi}(i) = \Psi(i)\beta(i) \quad (7)$$

$\alpha(i)$ et $\beta(i)$ sont les fonctions de spin correspondantes aux valeurs de spin $+1/2$ et $-1/2$ respectivement, $\chi(i)$ et $\bar{\chi}(i)$ sont des spin-orbitales de l'électron i . Ainsi, la fonction d'onde totale sera un simple produit de ces spin-orbitales

$$\Psi = |\chi_1(1)\chi_2(2)\dots\chi_m(n)| \quad (8)$$

Cette fonction d'onde ne satisfait pas au *principe d'exclusion de Pauli* [5], car elle ne prend pas en compte *l'indiscernabilité* des électrons. Ce problème est alors résolu en écrivant la fonction d'onde du système sous forme d'un déterminant de Slater [6,7].

Ce dernier est construit sur une base orthogonale de m fonctions d'ondes mono électroniques, tenant compte de l'état de spin et qu'on appelle spin-orbitales. La permutation de deux électrons implique la permutation de deux lignes conduisant à un changement de signe. Le problème réside alors dans l'obtention des meilleures spin-orbitales pour obtenir la fonction d'onde du système à n électrons.

$$\Psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(1) & \bar{\chi}_1(1) & \dots & \chi_m(1) & \bar{\chi}_m(1) \\ \chi_1(2) & \bar{\chi}_1(2) & \dots & \chi_m(2) & \bar{\chi}_m(2) \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot \\ \chi_1(n) & \bar{\chi}_1(n) & \dots & \chi_m(n) & \bar{\chi}_m(n) \end{vmatrix} \quad (9)$$

Le facteur de normalisation pour la fonction est $\frac{1}{\sqrt{n!}}$, n étant le nombre d'électrons.

II.1.3. Approximation LCAO

L'approximation LCAO [8] (*Linear Combination of Atomic Orbitals*) est basée sur le fait que la liaison chimique se forme suite au recouvrement de fonctions d'onde appartenant à 2 atomes suffisamment rapprochés. Ainsi, la fonction d'onde monoélectronique d'une

molécule peut être considérée comme une combinaison linéaire des fonctions d'onde atomiques. Ces dernières constituent les *fonctions de base* choisies pour la description du système.

$$\Phi_i = \sum_{r=1}^n C_{ir} \varphi_r \quad (10)$$

φ_r est une orbitale atomique et les coefficients C_{ir} sont inconnus. La recherche de la fonction d'onde revient à déterminer les coefficients C_{ir} .

II.2. Les différentes méthodes quantiques

La chimie quantique regroupe actuellement plusieurs méthodes de calculs, toutes basées sur la résolution de l'équation de Schrödinger. Parmi ces méthodes, il y a celles qui sont basées sur le traitement des intégrales de répulsion inter électronique (*ab initio* et semi empiriques) et celle qui est basée sur la densité électronique comme la théorie de la fonctionnelle de la densité (DFT).

II.2.1. Méthode *Ab-initio*

a. Formalisme *Hartree-Fock*[9]

Dans ce formalisme qui utilise les approximations de Born-Oppenheimer et orbitalaire, on peut se placer soit dans un cadre avec contrainte de spin (RHF : Restricted Hartree-Fock) ou dans celui sans contrainte de spin (UHF : Unrestricted Hartree-Fock). Le premier concerne les couches fermées où les OM sont doublement occupées donc pas d'électron célibataire. Le deuxième (UHF) concerne les couches ouvertes où les OM ne sont plus doublement occupées mais chaque spin-orbitale est remplacée par deux orbitales distinctes : la première décrit le spin α et l'autre le spin β . Comme les composés étudiés dans le cadre de cette thèse ne sont pas des radicaux, nous nous limiterons ici à décrire le formalisme RHF.

1. Energie moyenne associée à un déterminant de Slater

L'énergie électronique moyenne associée à la fonction d'onde Φ est donnée par la relation de Dirac suivante :

$$\langle \Phi | \hat{H} | \Phi \rangle = \bar{E}_e \quad (11)$$

Où Φ est la fonction d'onde de Slater et E_e l'énergie moyenne du système.

La résolution de cette équation est très complexe car il y a $n!$ Possibilités de répartir les électrons suivant les $m=n/2$ OM occupées, mais comme les spin-orbitales sont orthonormées le calcul se simplifie à cause de l'annulation d'un grand nombre de termes.

Sans donner les détails de calcul, l'intégration de l'équation (11) sur les variables d'espace conduit à l'énergie électronique suivante :

$$\bar{E}_e = \sum_{i=1}^m 2h_{ii}^c + \sum_{i=1}^m \sum_{j \neq i}^m (2J_{ij} - K_{ij}) \quad (12)$$

h_{ii}^c est l'intégrale monoélectronique de cœur, J_{ij} est l'intégrale bi-électronique coulombien et K_{ij} est l'intégrale bi-électronique d'échange.

$$h_{ii}^c = \langle \chi_i(1) | h^c(1) | \chi_i(1) \rangle \quad (13)$$

$$J_{ij} = \langle \chi_i(1)\chi_i(1) \left| \frac{e^2}{r_{12}} \right| \chi_j(2)\chi_j(2) \rangle \quad (14)$$

$$K_{ij} = \langle \chi_i(1)\chi_j(1) \left| \frac{e^2}{r_{12}} \right| \chi_i(2)\chi_j(2) \rangle \quad (15)$$

2. Méthodes de champ auto-cohérent [10] et équations d'HF

Le problème posé précédemment, concernant l'obtention des meilleures spin-orbitales pour obtenir la fonction d'onde du système à n électrons (équ. 9), est résolu ici en utilisant le théorème des variations [11]. Ce théorème implique que pour une variation $\partial\chi_i$ de la spin-orbitale χ_i appartenant à un ensemble orthonormal, on doit avoir une variation d'énergie ∂E nulle :

$$\int \partial \chi_i(1) \left[h^c(1) + \sum_{j \neq i}^m (2J_j(1) - K_j(1)) \right] \chi_i(1) d\tau_1 = 0$$

Ce qui conduit aux équations de Hartree-Fock suivantes :

$$F \chi_i = e_i \chi_i$$

e_i est la valeur propre de l'opérateur de Fock, elle correspond à l'énergie d'un électron décrit par la spin-orbitale χ_i dans le système à N atomes.

$$(18) \quad F(1) = h^c(1) + \sum_{j \neq i}^m [2J_j(1) - K_j(1)]$$

Les équations (18) aux valeurs propres qu'on appelle équations canoniques sont résolues de manière auto cohérente. Cette méthode consiste à choisir un jeu d'orbitales pour construire F et ensuite un nouveau jeu d'orbitales est généré. Ces nouvelles orbitales vont à leur tour servir à définir de nouveaux opérateurs J_j et K_j et donc un nouvel ensemble de fonctions propres de F . Cette procédure est répétée jusqu'à ce que la cohérence interne soit atteinte, d'où le nom de méthode du champ auto-cohérent (Self Consistent Field) [14], abrégé en utilisant les initiales anglaises SCF. Ce formalisme utilisant l'approximation LCAO conduit aux équations de Roothaan et Hall.

3. Equations de Roothaan et Hall

Dans ce cadre la fonction d'onde moléculaire s'écrit sous forme d'une combinaison linéaire des OA (équation 10) [12,13]. Plus le nombre de ces orbitaux atomiques est grand plus le résultat est proche de la réalité mais plus les calculs sont complexes et onéreux.

La résolution de ces équations, consistant à déterminer les coefficients des OM, se fait aussi de manière auto-cohérente comme dans le cas des équations d'HF. La difficulté dans la résolution des équations de Roothaan et Hall provient du grand nombre d'intégrales à calculer.

b. Calcul ab initio et les Bases d'orbitales atomiques utilisées

Le calcul *ab initio* repose sur la résolution des équations de Hartree-Fock Roothaan. Tous les électrons mis en jeu par chaque atome sont considérés. La précision des résultats obtenus dépend de la base d'OA utilisées [10]. Il faut noter que l'énergie est d'autant plus

stable que la base est plus étendue. Plus la base d'OA utilisée est étendue plus les calculs sont longs et les résultats sont précis. On peut dire que le choix des bases est encore aujourd'hui un problème qui n'est pas parfaitement résolu.

Les bases d'OA utilisées sont généralement des combinaisons linéaires fixées de gaussiennes que l'on appelle contractions. Elles ont été développées par l'équipe de John Pople dès la fin des années 1970. Nous pouvons citer : 3-21G [14], 6-31G [15],.... où 3 et 6 représentent le nombre de gaussiennes contractées et décrivant les couches internes, quant aux couches externes elles sont respectivement décrites par une contraction de 2 ou 3 gaussiennes et 1 gaussienne isolée. La notation : 6-31G* signifie l'ajout d'une fonction type d à une base comportant des orbitales s et p . La notation : 6-31G ** signifie l'ajout d'une orbitale $2p$ à l'atome d'hydrogène. L'ajout d'une orbitale diffuse sur un atome lourd est représenté par un signe + devant G : 3-21+G. Rappelons qu'une orbitale diffuse a le nombre quantique principal n supérieur à celui de valence. L'ajout de deux orbitales diffuses, une sur l'atome lourd et l'autre sur l'atome d'hydrogène, est représenté par un signe ++ devant G : 3-21++G.

Toutefois, si les méthodes *ab initio* donnent des résultats assez précis, elles ont pour inconvénient de nécessiter des temps de calcul très longs ; ce qui limite actuellement leur application à des systèmes comportant une quarantaine d'atomes.

Il faut noter que le calcul *ab initio*, ne tenant pas compte de la corrélation électronique, conduit à une énergie toujours supérieure à la valeur exacte. Pour cela, des méthodes dites 'post Hartree Fock' ont été développées. Parmi ces méthodes, nous pouvons citer les méthodes d'interaction de configuration et celles de Møller Plesset.

Le calcul *ab initio* est très gourmand en temps CPU (Central Processing Unit) à cause du nombre d'intégrales à déterminer et qui augmente avec la puissance quatrième du nombre d'OA utilisé [10]. De ce fait, des théories moins coûteuses sont parfois nécessaires, même au risque de perdre une partie de l'information. Parmi celles-ci, nous pouvons citer la théorie de la fonctionnelle de la densité (DFT) [16] et les méthodes semi-empiriques.

La DFT utilise les théorèmes fondamentaux de Hohenberg et Kohn [17] qui ont établi une relation fonctionnelle entre l'énergie de l'état fondamental et la densité électronique. Elle a pour objectif de décrire un système en considérant la densité comme variable de base. La précision des résultats obtenus avec la DFT présente un excellent compromis entre les résultats précis mais lents des méthodes *ab-initio* les plus sophistiqués et les résultats rapides mais grossiers des méthodes semi-empiriques que nous décrivons ci-dessous.

II.2.2. Méthodes semi-empiriques

a. Principe

Les méthodes semi empiriques (SE) ont pour but de réduire de façon significative le temps de calcul que nécessitent les méthodes *ab-initio*. Dans une méthode SE, une partie des calculs HF est remplacée par des paramètres ajustés sur des valeurs issues de calcul *ab initio* ou expérimentales. Elle ne considère que les électrons de valence qui sont décrits par les orbitales atomiques de type Slater. Cette approximation est basée sur le fait que les électrons impliqués dans la liaison chimique sont ceux de la couche de valence. Ces méthodes cherchent d'une part, à diminuer considérablement le nombre d'intégrales moléculaires à calculer et d'autre part, à éliminer complètement le calcul des intégrales biélectroniques, à trois et quatre centres. Toutes les méthodes SE négligent les intégrales de recouvrement dans la résolution des équations SCF [18] : c'est ce qu'on appelle approximation du RDN « Recouvrement Différentiel Nul » ou ZDO « Zero Differential Overlap ».

L'application du RDN simplifie les intégrales biélectroniques comme suit :

$$(rs|tu) = (rr|tt)\delta_{rs}\delta_{ts} \quad (19)$$

Avec
$$\delta_{rs} = \begin{cases} 0 & r \neq s \\ 1 & r = s \end{cases} \quad (\text{symbole de kronecker})$$

Les différentes méthodes SE qui existent actuellement résultent de la manière dont le RDN est appliqué sur les intégrales biélectroniques. Nous résumons, dans le tableau ci-dessous, les versions successives des méthodes SE générées par le degré d'application de cette approximation.

Tableau 1 : Différentes méthodes SE utilisées.

Degré d'application de RDN ou ZDO	Méthode	Versions
$\varphi_r \neq \varphi_s$	CNDO [19] (Complete Neglect of Differential Overlap)	CNDO/1
		CNDO/2 [20]
		CNDO/s [21]
$\varphi_r \neq \varphi_s$ sauf pour les intégrales mono centres (rs/rs)	INDO [20,21] (Intermediate Neglect of Differential Overlap) M : Modified Z : Zerner S : Symmetric	INDO
		ZINDO [22]
		MINDO/1 [23]
		MINDO/2 [23]
		MINDO/3 [23]
		INDO/s [24]
R≠S	NDDO [26,27] (Neglect of Differential Diatomic Overlap)	MNDO [28,29]
		AM1
		PM3

En général, toutes ces versions sont précises pour des familles de composés voisines de celles utilisées pour la paramétrisation. Cependant, celles basées sur l'approximation NDDO et que nous présentons ci-après, sont les plus utilisées actuellement.

b. Versions basées sur l'approximation NDDO

L'approximation NDDO [26,27] consiste à négliger les recouvrements diatomiques différentiels où interviennent les produits de deux orbitales centrées sur deux atomes différents. Les améliorations successives de la paramétrisation ont conduit au cadre MNDO et ses versions successives [30-34]. Cependant, MNDO décrit mal la liaison hydrogène [30]; ce qui a conduit l'équipe de Dewar à développer AM1 (Austin Model 1) [31]. Cette dernière ne diffère de MNDO que par le procédé empirique utilisé pour évaluer les répulsions entre cœurs [32]. En effet, selon Dewar et al, les formules utilisées dans MNDO conviennent mal aux distances de van der Waals. Dans le même ordre d'idées J. J. P. Stewart a proposé en 1989, sous la version révisée PM3 (parametric Method 3) [33], une reparamétrisation complète de MNDO. Celle-ci s'est effectuée sur des séries de molécules test au moyen d'un procédé automatique [33]. Il faut noter que PM3 présente le défaut de décrire mal les composés comportant des liaisons amides.

III. Mécanique moléculaire

III.1. Bases et principes

Contrairement à la mécanique quantique, la mécanique moléculaire (MM) ne tient pas compte explicitement des électrons. Elle traite les atomes et les électrons qui leur sont associés comme des masses ponctuelles chargées, obéissant aux lois de la mécanique classique.

L'énergie de la molécule est exprimée sous la forme d'une somme de contributions associées aux écarts de la structure par rapport à des paramètres structuraux de référence :

$$E = E_r + E_\theta + E_\Phi + E_\omega + E_{\text{van der Waals}} + E_{\text{électrostatique}} + E_{\text{liaisons hydrogène}} + E_{r\theta} \quad (20)$$

Les variables du calcul sont les coordonnées internes du système : longueur de liaison r , angles de valence θ , angles dièdres Φ , angles dièdres impropres ω , ainsi que les distances entre atomes non liés dont les interactions sont représentées par un potentiel de van der Waals le plus souvent de type Lennard-Jones et un potentiel électrostatique le plus souvent de type coulombien.

III.2. Champs de forces

La Mécanique Moléculaire (MM) est une méthode empirique qui consiste à dégager, grâce à des résultats de la mécanique classique, les propriétés des atomes et des molécules et à déduire des modèles mathématiques simples appelés *Champs de forces* [35-36]. Le champ de forces permet de décrire les forces agissant lorsque la géométrie d'équilibre est perturbée. Sous ce terme sont en fait regroupés deux éléments : d'une part l'expression des différentes fonctions contribuant au calcul énergétique et d'autre part les valeurs des différentes constantes paramétrant ces fonctions. De nombreux champs de forces sont disponibles dans différents logiciels de modélisation moléculaire. Chacun d'eux est limité à une classe de molécules ayant suffisamment de ressemblance structurelle et fonctionnelle. On peut ainsi trouver des champs de force destinés à la modélisation de petites molécules organiques [37], de macromolécules [38], de nucléotides [39] ou encore de complexes organométalliques [39]. Il est donc dans un premier temps important de vérifier si le champ de force choisi est adapté au système à étudier. Dans ce travail, nous avons utilisé la mécanique moléculaire pour

minimiser les structures des molécules dessinées et à étudier. Pour cela nous avons choisi le champ de force MM2 qui est bien adapté à la modélisation des petites molécules organiques [37].

V. Techniques du Drug design

Les techniques du *drug design* représentent l'ensemble des processus nécessaires pour découvrir de nouveaux médicaments. Elles ont pour objectif la conception rationnelle de médicaments. Elles sont basées soit sur l'étude des relations entre l'activité d'une molécule et sa structure soit sur l'étude du docking moléculaire qui nécessite la connaissance de la cible biologique. La première utilise des outils tels que le QSAR (*Quantitative Structure-Activity Relationships*) ou la CoMFA (*Comparative Molecular Field Approach*) et ne nécessite pas la connaissance de la cible biologique. Quant à la deuxième, dite 'receptor-based', elle est appliquée lorsqu'une structure tridimensionnelle (3D) fiable du récepteur est disponible.

V.1. Approche 'receptor-based'

Cette approche basée sur la cible, appelée le docking moléculaire ou encore l'arrimage moléculaire, est une technique qui permet de placer une petite molécule, dite ligand, dans le site actif d'une macromolécule, dite récepteur. L'objectif du docking moléculaire consiste à prédire à la fois la conformation active (ou viable) et l'orientation relative d'une molécule au sein du récepteur. Le plus souvent, la simulation se fait dans un champ de force simplifié permettant de prédire l'énergie d'interaction entre le ligand et la protéine, dont le but est de classer les différents résultats obtenus. Il faut noter que dans une simulation de docking, le récepteur et le ligand peuvent être considérés rigides ou flexibles. Dans le deuxième cas, le ligand est considéré flexible mais la flexibilité permise au récepteur est limitée exclusivement aux chaînes latérales des résidus du site actif ; on parle alors d'un docking semi-flexible, réservé au docking ligand-protéine. Plusieurs programmes utilisent ce type de docking parmi les quels on peut citer AutoDock[40].

V.2. Alignement de structures tridimensionnelles de protéines

L'alignement de structures tridimensionnelles de protéines est une manière de représenter deux ou plusieurs séquences de macromolécules biologiques (exemple protéines) les unes sous les autres, de manière à en faire ressortir les régions homologues ou similaires. L'objectif de l'alignement, réalisé par des programmes informatiques, est de disposer les acides aminés pour identifier les zones de concordance. Ceci nécessite en général

l'introduction de "trous" à certaines positions dans les séquences, de manière à aligner les caractères communs sur des colonnes successives. Ces trous correspondent à des insertions ou des délétions (appelés indel) d'acides aminés dans les séquences biologiques.

VI. Interactions intermoléculaires

VI.1. Interactions hydrophobes

Les acides aminés dont le radical est hydrophobe et apolaire ont la propriété d'empêcher la formation des liaisons hydrogène entre les molécules d'eau. Ils forment dans la structure des protéines des zones hydrophobes où les molécules d'eau ne peuvent échanger aucune liaison avec les radicaux d'acides aminés.

VI.2. Interactions de Van Der Waals

Ce sont des interactions entre atomes, molécules, ou entre une molécule et un cristal. Ces interactions Van Der Waals s'appliquent à très courte distance et ne concernent donc que les atomes de surface [41]. Elles sont nombreuses et contribuent essentiellement dans la recherche de la concordance stérique entre le ligand et la protéine réceptrice [42].

VI.3. Interactions électrostatiques

Certains acides aminés peuvent être chargés positivement ou négativement, ce qui donne lieu à des interactions électrostatiques. Ces interactions électrostatiques agissent à "longues" distances (autour de 10 Å), entre des groupes d'atomes électriquement chargés [43].

VI.4. Interactions ioniques ou pont salin

Ces liaisons résultent des forces électrostatiques dans une protéine entre les résidus chargés positivement et les résidus chargés négativement. Elles ont une importance certaine dans les liaisons entre la protéine et d'autres molécules [44].

VI.5. Liaisons hydrogène

Les liaisons hydrogènes sont des interactions intermoléculaires de faible énergie entre 20 à 30 KJ/mol qui agissent dans les systèmes biologiques, à très courte distance (0,8 à 2,8 Å) entre un élément très électronégatif portant au moins un doublet non liant et un atome d'hydrogène lié à un élément très électronégatif.

VI .6. Pont disulfure ou cystine

Pont disulfure est une liaison covalente qui se forme quand la protéine se replie (structure tertiaire) ou lors de la dissociation de sous-unités protéiques (structure quaternaire) par l'oxydation dans la protéine. On trouve deux des 20 acides aminés ont des radicaux contenant un atome de soufre c'est le cas de cystéine et la molécule résultante de la liaison de deux cystéines est la cystine.

Le pont disulfure ne se forme pas dans le cytoplasme qui est un environnement réducteur, mais lorsque les protéines sont secrétés ou exposées à la surface [45].

Références bibliographiques

- [1] Schrodinger E, Ann.,Physik. 1930, 61: 126.
- [2] Burrau.Klg O, DanskeVidenskab., Selskab. 1927, 7: 1.
- [3] Born M, Oppenheimer JR, Ann., Physik. 1927, 84: 457.
- [4] Slater JC, Phys. Rev. 1929, 34: 1293.
- [5] Pauli W, Physik Z. 1925, 31: 765.
- [6] Hariharan PC, Pople JA, Mol. Phys. 1974, 27: 209.
- [7] Slater JC, Phys. Rev. 1930, 35: 509; *ibid*, 1929, 34: 1293.
- [8] Pople JA, Beveridge DL, approximate molecular orbital theory.McGraw-Hill, New York. 1970, 330.
- [9] Slater JC, Phys. Rev. 1930, 35: 210.
- [10] Hartree DR, Proc. Cambridge Phil.Soc. 1928, 24: 9.
- [11] Rivail JL, "Eléments de chimie quantique à l'usage des chimistes".2^{ème} ED. Interditions et CNRS Editions., 1994, 313.
- [12] Roothaan CCJ, Rev. Mod. Phys. 1951, 23: 69.
- [13] Hall GG, Proc. Roy. Soc. London. 1951, 205: 541.
- [14] Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre W J, Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements". J. Am. Chem. Soc.1982, 104(10): 2797-2803.
- [15] Kohn W, Becke AD, Parr RG, Density Functional Theory of Electronic Structure".J. Phys. Chem.1996, 100(31): 12974-12980.
- [16] Del Bene J, Jaffe HH, J. Chem. Phys. 1968,48: 1807.
- [17] Parr R. J, J. Chem. Phys.1952,20:239.
- [18] (a) Pople JA, Beveridge DL, Approximate molecular Orbital Théory.Mac Graw-Hill. 1970; (b) Pople JA, Segal GA, J. Chem. Phys. 1965, 43: 136-149; 1966,44: 3289.
- [19] Pople JA, Beveridge DL, Dobosh PA, J. Chem. Phys. 1967, 47: 2026-2033.
- [20] Del Bene J, Jaffe HH, J. Chem. Phys. 1968, 48: 1807.
- [21] Leach AR, Molecular Modelling.,British Library. 1996, 05.
- [22] Bingham RL, Dewar MJS, J. Amer. Chem. Soc. 1975, 97:1285.
- [23] Chen SY, Hedges RM, Theor. Chim. Acta. 1973, 31: 275.
- [24] Coffey P, Jug K, J. Amer. Chem. Soc. 1973, 95: 7800.
- [25] Dewar MJS, Zoebisch EG, Healy EF, Stewart JJ, J. Amer. Chem. Soc. 1985, 107: 3902.
- [26] Stewart JP, Programme MOPAC, Quantum Chemistry Programm Exchange. N°455, Indiana University, looming, U. S. A.
- [27] Dewar MJ, Thiel W, J. Am. Chem. Soc. 1977, 99: 4899-4907.
- [28] Dewar MJS, Merz KM, Organometallics.1986, 5: 1494.
- [29] Stewart JJP, Rev. Comp. Chem., Lipkowitz KB, Boyd ed DB, VCH Publishers, New York.1990, 45.
- [30] Dewar MJS, Zoebisch E, Healy EF, Stewart JJP, J. Am. Chem. Soc.1985, 107 :3902-3909.
- [31] Rivail JL, "Eléments de chimie quantique à l'usage des chimistes".Interditions et Editions du CNRS, Paris.1989, 318.
- [32] Stewart JJP, J. Comput. Chem.1989, 10: 209-264.
- [33] van Gunsteren WF, Berendsen HJC, Angew. Chem. Int. Ed. Engl.1990, 29: 992.
- [34] van Gunsteren WF, in Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications.2:3-36, van Gunsteren WF, Weiner PK, Wilkinson eds AJ,Escom Science Publishers.1993.
- [35] Brooks CL, Karplus M, Pettitt BM, Proteins: a Theoretical Perspective of Dynamics, Structure and Thermodynamics.Wiley, New York.1988.
- [36] Biosym/MSI, Discover, user guide, ed. Biosym/MSI. San Diego., 1995.
- [37] Weiner SJ, Kollman PA, Nguyen DT, Case DA, An all-atom force field for simulations of proteins and nucleic acids.J. Comp. Chem.1986,7:230.
- [38] Allinger NLJ, Am. Chem. Soc. 1977, 99: 8127.
- [39] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK and Olson AJ, Journal of Computational Chemistry.1998, 19: 1639-1662.
- [40] Bouyer G, Rendu multimodal en Réalité Virtuelle : Supervision des interactions au service de la tâche. Doct : Université Paris XI, 2007.
- [41] Teniou S, Conception in silico de nouveaux inhibiteurs de la topoisomérase I de cellules cancéreuses.Magt : Technologie des Explorations Biochimiques : Université Mentouri Constantine, 2012.
- [42] Andrusier N, « FireDock: fast interaction refinement in molecular docking. ». Proteins. 2007, 69(1):139-59.
- [43] Salpin F, « laine et colorants : fixation, quantification et vieillissement. Étude par spectrométrie Roman ».Doct : Université Pierre et Marie Curie, 2008.
- [44] Martin TW, Derewenda ZS, « The name is bond -H bond ».Nature Structural Biology. 1999, 6(5):403-06.

CHAPITRE III

**Modélisation de la structure de la protéine Arf1 et de
ses interactions avec une série de Flavonoïdes**

I. Introduction

Dans le but de développer de nouveaux inhibiteurs de la protéine Arf1, nous avons entrepris dans ce chapitre une étude détaillée de la modélisation de la structure de cette protéine ainsi que de son mécanisme d'inhibition. Pour effectuer une approche assez réaliste, nous nous sommes laissés guidés par des informations de la littérature disponibles et présentées dans le chapitre I. Cette démarche nous a incités à structurer nos travaux en deux étapes successives que nous résumons brièvement comme suit :

- La première étape porte sur une étude de modélisation par alignement multiple de la structure de l'Arf1, la protéine impliquée dans le cancer du sein. Cette étude réalisée sur trois structures de départ de la protéine Arf1 nous a permis de mettre en évidence la structure optimale à utiliser dans la partie II.
- Ensuite, à l'aide de la technique du docking moléculaire, nous avons étudié le mode d'association le plus probable d'une série de vingt flavonoïdes dans la cavité catalytique de cette protéine. Cette étude totalement guidée par des informations disponibles dans la littérature, a permis de mettre en évidence les interactions mises en jeu entre ces composés et les acides aminés du site actif et de prédire l'énergie de liaison entre chaque composé et la protéine Arf1.

Nous présentons de manière détaillée ces étapes successives dans les 2 parties que nous notons A et B.

Partie A : Modélisation par alignement multiple de la structure de l'Arf1

I. Objectif

Comme nous l'avons mentionné dans le chapitre I, il existe seulement trois structures cristallographiques correspondant au complexe Arf1 /BFA et sec7 domaine dans la *Protein Data Bank* (PDB) [1]. Il est important de rappeler que le ligand cristallographique dans ces trois structures adopte la même conformation. Le problème est donc de rechercher la conformation à utiliser dans la deuxième partie de ce travail.

II. Méthodologie

Dans le but de choisir la structure cristallographique que nous allons utiliser dans notre travail, nous avons entrepris une étude de comparaison entre toutes les structures 3D d'Arf1 liées au domaine sec7 et BFA (BREFELDINE A) disponibles dans la *Protein Data Bank* (PDB) [1]. Un total de trois structures cristallographiques ont été extraites de la PDB (codes : 1R8Q de résolution = 1.86 [2], 1RE0 de résolution = 2.40 [2] et 1S9D de résolution = 1.80 [3]).

Nous portons dans le tableau suivant une description de chaque structure cristallographique.

Tableau 1 : Présentation des trois structures cristallographiques d'Arf1 extraites de la PDB ; les résidus colorés en bleu sont conservés dans les trois structures.

Structures-cristallographiques d'ARF1	Nombre de chaînes	Nombre d'acides aminés de la chaîne A	Acides aminés du site actif (AA < 7Å de la chaîne A)
1R8Q GDP	4 chaînes (A, F, B, E)	179 résidus	ILE20, LEU21, MET22, VAL23, GLY24, LYS30, LEU88, ILE33, THR64, GLY87, ILE89, PHE90, VAL91, LEU122, VAL123, PHE124, AFB403.
1RE0 GDP	2 chaînes (A, B)	155 résidus	ILE20, LEU21, MET22, VAL23, GLY24, LEU88, LYS30, THR64, VAL65, TRP66, ILE89, PHE90, VAL91, LEU121, LEU122, PHE124, HOH738, AFB702.
1S9D GDP	2 chaînes (A, E)	162 résidus	LEU21, MET22, VAL23, GLY24, LYS30, LEU88, ILE89, PHE90, VAL91, VAL92, LEU121, LEU122, VAL123, PHE124, AFB403.

Les structures cristallographiques portées dans le tableau 1 sont composées de plusieurs chaînes d'acides aminés (AA) (02 ou 04 chaînes). Elles comportent plus de 100 AA (179 AA dans la structure 1R8Q, 155 AA dans la structure 1RE0 et 162 AA dans la structure 1S9D). Les résidus du site actif ne sont pas conservés dans les trois structures, ce qui explique le fait que les protéines dans l'organisme ne sont pas figées.

Tableau 2 : Présentation des différents éléments composant les trois structures cristallographiques d'Arf1 extraites de la PDB.

Structures-cristallographiques d'ARF1	Hélice alpha	Feuillets beta	Coudes	Boucles
1R8Q GDP	*ILE4-LEU8 *LYS30-LEU37 *TRP78-PHE82 *ARG99-LEU111 *ALA136-LEU143 *LEU166-ASN179	*MET18-VAL23 *GLU54-GLU57 *ILE61-VAL65 *GLY87-ASP93 *ALA119-ASN126 *TRP153-ALA157	*LEU8-PHE9 *ASP26-ALA27 *LYS38-GLY40 *GLN83-THR85 *SER94-ASP96 *LEU116-ASP118 *GLN128-ASP129 *PRO131-ASN132 *LEU143-GLY144 *ALA160-GLY163	*LYS10-GLY11 *LEU12-PHE13 *GLY14-LYS15 *ILE49-PHE51 *LYS59-ASN60 *GLY70-GLN71 *LEU130-PRO131 *HIS150-ARG151
1RE0 GDP	*LYS30-LYS38 *ARG79-PHE82 *VAL100-ALA112 *ALA136-LYS142 *LEU166-ASN179	*ARG19-GLY24 *THR55-TYR58 *ILE61-VAL65 *THR85-SER94 *VAL120-ASN126 *TRP153-ALA157	*ASP26-ALA28 *LYS59-ILE61 *GLN83-ASN84 *SER94-ASN95 *ARG97-ARG99 *ARG117-ASP118 *GLN128-ASP129 *PRO131-ALA133 *LYS142-LEU145 *ALA160-GLY163	*GLY24-LEU25 *ALA27-GLY29 *GLY40-GLU41 *THR45-ILE46 *ILE49-PHE51 *GLY70-GLN71 *ARG97-GLU98 *LEU130-PRO131 *LEU148-HIS150
1S9D GDP	*LYS30-LEU37 *ARG99-ALA112 *ALA136-ALA137 *GLU138-LEU143 *LEU166-LEU177	*ILE20-VAL23 *THR55-TYR58 *ILE61-VAL65 *GLY87-ASP93 *VAL120-ASN126 *TRP153-ALA157	*ASP26-ALA27 *LYS59-ASN60 *GLN83-THR85 *SER94-ASP96 *ARG117- ALA119 *GLN128-ASP129 *PRO131-ASN132 *LYS142-GLY144 *ARG149- ARG151	*ALA28-GLY29 *ILE49-GTY50 *GLY70-GLN71 *LEU130-PRO131 *ARG151-ASN152

Il est bien de noter que les trois structures étudiées diffèrent aussi par le nombre et la localisation des différents motifs biologiques (Hélice alpha, Feuillets beta, coudes et les boucles) constituant les structures secondaires correspondantes, voir le tableau 2 et la figure 1.

Contrairement à la structure 1R8Q qui contient six hélices alpha, le nombre d'hélices alpha est le même dans les deux structure 1RE0 et 1S9D (elles sont composées de cinq hélices alpha). Les séquences de ces hélices sont faiblement conservées. Il est bien de noter aussi que la différence entre les structures étudiées réside dans le nombre et les séquences des boucles. Par ailleurs, les trois structures sont caractérisées par une forte conservation dans les séquences des feuilletts beta et des coudes.

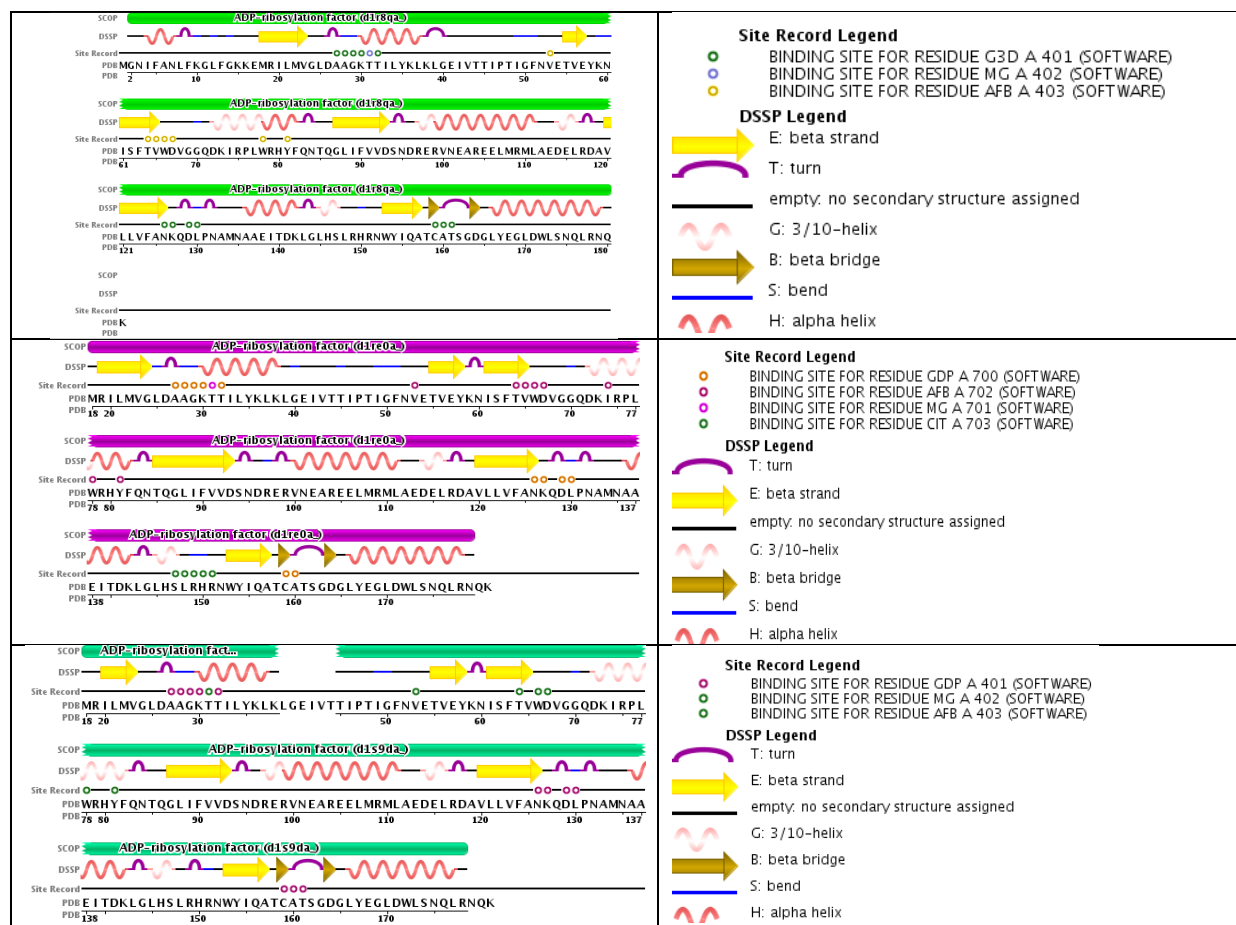


Figure 1: Séquence en acides aminés des trois structures étudiées 1R8Q, 1RE0, 1S9D[4].

Afin d'identifier les régions de similarité entre les trois structures cristallographiques d'Arf1 extraites de la PDB, nous les avons aligné. Nous portons le résultat de cette simulation sur les deux figures suivantes.

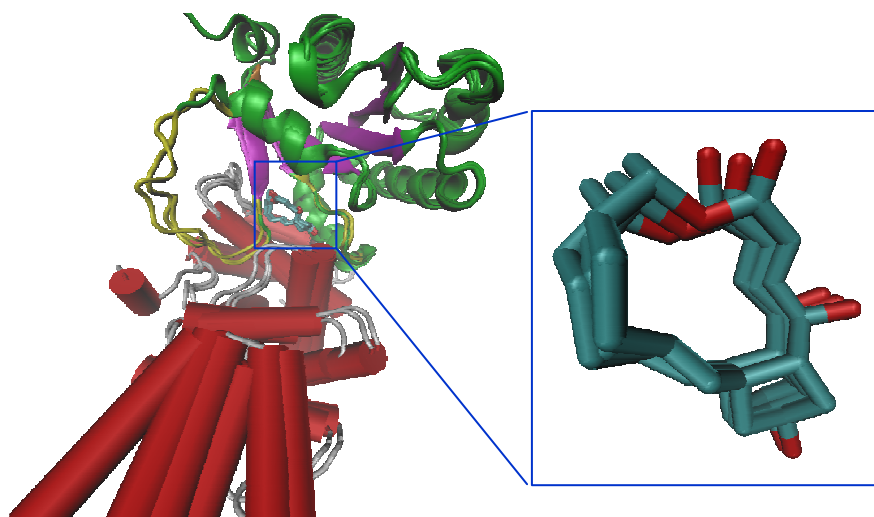


Figure 2 : Vue 3D de la superposition des trois structures cristallographiques de l'Arf1 (représenté en mode 'New Cartoon') et le domaine sec7 (représenté en mode 'cartoon') et BFA en tiges.

La figure 2 montre que la conformation du ligand cristallographique adopté dans les trois structures étudiées est pratiquement la même tandis que les trois structure de la protéine Arf1 ne sont pas parfaitement superposables.

```

1R8Q_A  MGNIFANLFKGLFGKKEMRILMVGLDAAGKTTILYKLLGEIVTTIPTIGFNVETVEYKNI
1RE0_A  M-----MRILMVGLDAAGKTTILYKLLGEIVTTIPTIGFNVETVEYKNI
1S9D_A  M-----MRILMVGLDAAGKTTILYKLL-----TIPTIGFNVETVEYKNI

1R8Q_A  SFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSDNRERVNEAREELMRMLAEDELRAVL
1RE0_A SFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSDNRERVNEAREELMRMLAEDELRAVL
1S9D_A SFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSDNRERVNEAREELMRMLAEDELRAVL

1R8Q_A  LVFANKQDLPNAMNAAEITDKLGLHSLRHRNWYIQATCATSGDGLYEGLDWLSNQLRNQM
1RE0_A  LVFANKQDLPNAMNAAEITDKLGLHSLRHRNWYIQATCATSGDGLYEGLDWLSNQLRN-M
1S9D_A  LVFANKQDLPNAMNAAEITDKLGLHSLRHRNWYIQATCATSGDGLYEGLDWLSNQLR--M
    
```

Figure 3 : Alignement des séquences en acides aminés des trois isoformes, en bleu sont représentées les parties conservées et en rouge celles qui ne le sont pas.

Les tirets enregistrés dans les structures 1RE0 et 1S9D (---) représentent des GAPs ou inDel. Ils signifient qu'à ces points les deux structures ont subi des mutations par DELétion au cours de laquelle les acides aminés **GNIFANLFKGLFGKKE** sont perdus. Ce manque d'acides aminés dans les deux structures cristallographiques (codes : 1RE0 et 1S9D), nous incite à choisir la structure de l'Arf1 dont le code PDB 1R8Q comme structure métabolique dans la suite de notre travail.

Partie B : Etude par docking moléculaire d'une série de Flavonoïdes dans le site actif de l'Arf1

I. Objectif

Dans la perspective de l'étude du mécanisme d'inhibition de la BFA au sein de l'Arf1, une étude par le docking moléculaire de ce ligand s'avère nécessaire pour parvenir à des résultats rationnels voire proches de la réalité. Cette étude cruciale consiste en fait, à rechercher le mode d'association ligand/récepteur le plus probable, c'est-à-dire respectant les informations expérimentales disponibles à ce propos.

II. Méthodologie

II.1. Structures de départ

II.1.1. Le récepteur

Les coordonnées de la protéine Arf1 sont obtenues à partir d'une structure cristallographique disponible dans la banque de données des protéines (PDB : *Protein Data Bank*) [1]. Nous avons choisi la structure codée 1R8Q, de résolution égale à 1.86 Å et correspondant au complexe Bréfeldine A/Arf1. La structure chimique du ligand BFA (Brefeldine A), l'inhibiteur de la protéine Arf1, est donnée dans la Figure 4.

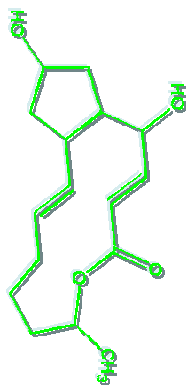


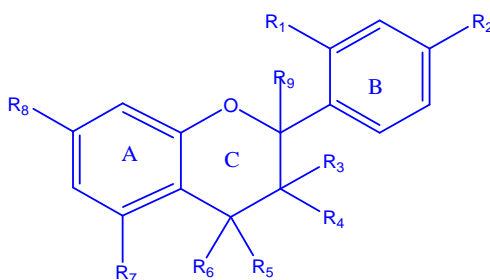
Figure 4 :Structure chimique de BFA

II.1.2. Les molécules étudiées

Les flavonoïdes, appelés également bio flavonoïdes, polyphénols, proanthocyanidines, catéchines ou flavonols, sont des substances présentes dans les plantes. Ils constituent un groupe de plus de 6 000 composés naturels. Ces composés sont à l'origine des teintes brunes, rouges et bleues des fleurs, des fruits et des légumes. Des études récentes ont démontré les diverses propriétés biologiques de ces composés, notamment les propriétés anti-oxydantes, vasculoprotectrices, anti hépatotoxiques, antiallergiques, anti-inflammatoires, anti-ulcéreuses et anti-tumorales [5].

Une base, contenant 20 molécules dérivées des flavonoïdes [6] et ayant des phényles et des groupements polaires, est présentée dans leTableau 3.

Tableau 3 : Structures chimiques des molécules étudiées.



molécule	R1	R2	R3	R4	R5	R6	R7	R8	R9
Flavo1	H	OH	H	OH	H	OH	OH	OH	H
Flavo2	H	OH	H	OH	H	H	OH	OH	H
Flavo3	H	OH	H	OH	H	OH	H	OH	H
Flavo4	H	OH	H	OH	H	H	H	OH	H
Flavo5	H	OH	H	OH	H	H	H	H	H
Flavo6	H	CH ₃	H	OH	H	H	H	H	H
Flavo7	H	CH ₃	H	H	H	OH	H	H	H
Flavo8	H	H	OH	OH	H	OH	OH	OH	H
Flavo9	H	H	H	OH	H	OH	OH	OH	H
Flavo10	H	H	H	H	H	OH	OH	H	H
Flavo11	H	H	H	OH	H	OH	H	H	H
Flavo12	H	H	CH ₃	OH	H	OH	H	H	H<
Flavo13	H	H	H	OH	H	H	H	H	H
Flavo14	H	H	H	H	C ₅ H ₁₁	OH	H	H	H
Flavo15	CH ₂ OH	H	H	OH	H	H	H	H	H
Flavo16	OH	H	H	OH	H	H	H	H	OH
Flavo17	H	H	H	OH	H	H	H	H	OH
Flavo18	H	H	H	OH	H	H	H	H	C ₆ H ₅
Flavo19	H	H	H	H	H	OH	H	H	OH
Flavo20	H	H	H	H	H	H	H	H	CH ₂ OH

II.2. Méthodologie utilisée

A l'aide de l'interface graphique AutodockTools (ADT), nous avons préparé et déterminé les paramètres *optima* que nous allons utiliser dans le cadre de ce travail

II.2.1. Préparation de la structure du récepteur

Une seule chaîne de la structure 1R8Q.pdb, dépourvue des molécules d'eau et du ligand, est sélectionnée pour la préparation de la structure de la cible. Une fois le ligand retiré, les hydrogènes manquants sont ajoutés et les charges de Gasteiger sont assignées [7].

II.2.2. Localisation du site actif

Avant de retirer le ligand du complexe, nous avons pris le soin de relever tous les résidus qui sont à une distance inférieure à 7 Å, du centre du ligand. Les résidus sélectionnés sont donnés dans le tableau 4.

Tableau 4 : résidus du site actif de l'Arf1

chaîne	Résidus								
A	TYR81	TRP78	ASP67	TRP66	VAL65	THR64	PHE51	ASN52	VAL53
E	MET194	THR197							

Ces acides aminés sont considérés dans notre étude comme des résidus englobant le site actif. Sachant que le programme Autodock4 [8] utilise un docking semi-flexible, nous avons séparé le système en une partie flexible, contenant les résidus du site actif, et une partie rigide, contenant le reste de la protéine.

II.2.3. Préparation des molécules étudiées

Nous avons modélisé les structures préliminaires des vingt molécules à l'aide de la chaîne de programmes de ChemOffice [9]. Une préoptimisation de la géométrie, dans le cadre de la mécanique moléculaire, en utilisant le champ de force MM2 [10]. Cette stratégie a pour but d'accélérer la convergence de l'optimisation de la géométrie réalisée ensuite dans le cadre de la mécanique quantique en utilisant la méthode semi-empirique PM3.

Les charges partielles de Gasteiger ont été assignées aux atomes à l'aide du logiciel Auto Dock Tools [11]. Un "centre de masse" pour les mouvements rigides externes qui sont

les rotations et les translations du ligand et également les liaisons flexibles permettant les déformations internes du ligand ont été définis.

II.2.4. Protocole de docking

a- Paramètres de la grille de calculs

Le programme AutoDock utilise dans les calculs de docking moléculaire une grille de potentiels, permettant d'augmenter la vitesse d'évaluation de l'énergie d'interaction ligand/récepteur. Le site actif ou la zone à explorer du récepteur est englobée dans une grille tridimensionnelle dont les paramètres sont définis de manière à permettre la libre orientation de chaque composé étudié. Ainsi, chaque ligand est docké dans cet espace délimitant le site de liaison dans l'Arf1. Cette grille, ayant pour centre (5.228, 6.841, 33.346), présente un volume égal à $42 \times 46 \times 50 \text{ \AA}^3$. L'espacement entre chaque point de la grille est de 0.375 \AA . Cette valeur, prise par défaut par AutoDock4, correspond au quart de la longueur d'une simple liaison C-C. Il faut noter que ce paramètre peut varier de 0.2 à 1.0 \AA , suivant la précision désirée de la grille.

b-Paramètres de docking

Le protocole de docking moléculaire adopté, dans ce travail, consiste à docker chaque ligand, considéré flexible, dans le site actif de l'Arf1. Les paramètres d'arrêt ont été repoussés à 2500.000 évaluations d'énergie et à 27.000 générations. Pour avoir un échantillon de solutions statistiquement intéressant, nous avons retenu 50 conformations pour chaque ligand. Chaque expérience de docking a consisté en une série de 50 cycles, utilisant l'algorithme génétique lamarckien (LGA). Quant aux autres paramètres de simulations, nous avons pris les paramètres par défaut d'Autodock.

II.3. Résultats et discussion

II.3.1. Validation des paramètres de docking

Nous avons effectué cette validation en comparant les deux conformations de BFA, à savoir, celle issue du docking utilisant ces paramètres et la conformation expérimentale issue du complexe 1R8Q. Nous présentons, dans la figure 5, la superposition des deux conformations RX et issue du docking. La figure montre que les deux conformations se superposent bien; ce qui conforte l'utilisation des paramètres établis dans la suite de notre étude structurale.

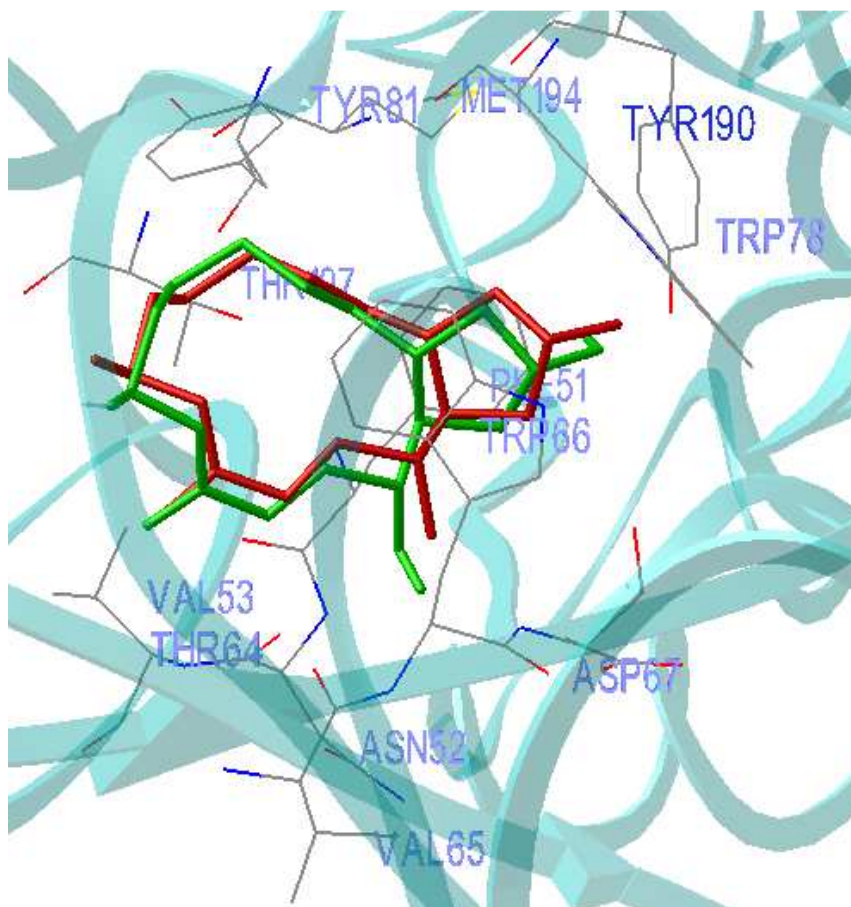


Figure 5 : superposition des deux conformations de BFA dans le site actif de l'Arf1 : la conformation issue du docking en vert et la conformation cristallographique en rouge.

II.3.2. Sélection de la conformation dite bioactive

Pour chaque ligand, l'opération de docking a généré plusieurs complexes regroupés en familles (clusters). Ces clusters classés par ordre d'énergie croissante présentent des populations composées de conformations prédites viables. Nous avons analysé en détails tous les clusters. La sélection de la conformation bioactive a été faite selon les critères suivants :

- 1) Le mode de liaison le plus proche de celui adopté par le ligand expérimental BFA dans la structure RX 1R8Q.
- 2) La plus petite énergie de liaison.
- 3) Le cluster le plus peuplé.
- 4) Le nombre de liaisons hydrogène.

Nous donnons dans le tableau 5 les énergies de liaison, ligand/récepteur, notées $\Delta G_{\text{binding}}$ obtenues par AutoDock pour les vingt composés étudiés.

Tableau 5 : énergie de liaison, $\Delta G_{\text{binding}}$ (kcal/mol), prédites par AutoDock des composés étudiés.

molécule	Energie de liaison
Flavo1	-24.09
Flavo2	-23.07
Flavo3	-23.05
Flavo4	-20.83
Flavo5	-23.73
Flavo6	-22.82
Flavo7	-24.0
Flavo8	-23.54
Flavo9	-20.16
Flavo10	-22.27
Flavo11	-22.6
Flavo12	-23.04
Flavo13	-23.68
Flavo14	-15.58
Flavo15	-22.14
Flavo16	-22.53
Flavo17	-23.15
Flavo18	-22.84
Flavo19	-23.71
Flavo20	-22.08

II.3.3. Mode de liaison des composés étudiés

L'analyse des résultats, obtenus par les simulations de docking, montre que les flavonoïdes s'insèrent bien dans le site actif de la protéine Arf1 présent dans le complexe cristallographique 1R8Q (Figure 6). La visualisation des conformations prédites bioactives montre les acides aminés qui constituent le site actif de la protéine Arf1.

Dans le but de prendre en considération le maximum d'interactions existantes entre le ligand et le récepteur, nous avons relevé toutes les interactions inférieures ou égales à 6 Å à partir du centre du ligand. Nous résumant dans le tableau 6 les interactions les plus significatives en termes de liaisons hydrogènes et d'interactions hydrophobes ainsi que les énergies de liaisons entre les composés étudiés et la protéine Arf1.

Tableau 6 : Mode de liaison des composés Flavo1-Flavo7
N° : Numéro de la conformation

Molécule	N°	Liaison Hydrogène			Résidus à proximité du ligand (< 4 Å)
		Atome du ligand	Résidu (groupement)	Distance (Å)	
Flavo 1	1	*OH (R ₂) *OH (R ₄) *OH (R ₈) *OH (R ₈) *OH (R ₈)	*ASP67 (OH de la chaîne latérale) *ASP198 (OH de la chaîne latérale) *THR64 (OH de la chaîne latérale) *THR55 (OH de la chaîne latérale) *TYR81	*2.795 (liaison H) *2.671 (liaison H) *2.743 (liaison H) *2.922 *4.714	ARG19, THR55, TYR81, ASP67, TRP 66, THR 64, PHE 51, VAL 53, ASP 198, VAL 204, THR 197.
Flavo 2	36	*OH (R ₂) *OH (R ₈) *OH (R ₈) *OH (R ₇)	*TYR190 *THR64 *ARG19 (NH du backbone) *VAL204	*2.985(liaison hydrogène) *2.909(liaison hydrogène) *2.915(liaison hydrogène) *4.772	GLU54, TYR81, ASP67, TRP66, VAL65, THR64, ASN52, VAL53, ASP198, VAL204, MET194, THR197.
Flavo 3	27	*OH (R ₈) *OH (R ₄) *OH (R ₆)	*ASP67 *ASP198 *TYR81	*2.784(liaison hydrogène) *3.514 *3.191	ARG19, THR55, TYR81, ASP67, TRP66, THR64, PHE51, ASN52, VAL53, VAL204, MET194, THR197.
Flavo 4	12	*OH (R ₈)	*ASP67	*2.820(liaison hydrogène)	ARG19, THR55, TYR81, ASP67, TRP66, THR64, PHE51, VAL53, ASP198, VAL204, THR197.
Flavo 5	23	*OH (R ₄) *OH (R ₄) *OH (R ₂)	*ASP67 *ASN52 *TYR190	*2.806(liaison hydrogène) *3.720(liaison hydrogène) *3.448	ARG19, LEU21, THR55, TYR81, TRP66, THR64, PHE51, ASN203, VAL204, MET194.
Flavo 6	30	*OH (R ₄) *OH (R ₄)	*ASP198 *TYR81	*3.041 *3.385	THR55, TYR81, ASP67, TRP66, THR64, PHE51, VAL53, TYR190, ASP198, VAL204, MET194.
Flavo 7	20	*OH (R ₆)	*THR197	*5.908	ARG19, TYR81, TRP66, THR64, PHE51, VAL53, ASP198, VAL204, THR197.

II.3.4. Interactions mises en jeu

Les composés étudiés, tableau 1, sont caractérisés par la présence d'un groupement hydroxyles (OH) sur plusieurs positions (R₁-R₉) des trois cycles A, B et C. Ces composés diffèrent par la nature chimique des substituants (R₁-R₉) (hydroxyle, n-pentyles, méthyle et cyclo-hexyl). Cette variation dans la structure chimique de ces substituants engendre des différences en termes d'interaction avec les résidus du site actif. L'analyse des résultats montre que le composé prédit le plus actif possède une énergie de liaison E = -24.09kcal/mol et que nous avons nommé Flavo1 dans le tableau 1, présente cinq groupements hydroxyles (OH) sur les positions R2, R4 et R6-R8. Il forme trois liaisons hydrogènes avec les résidus du site actif qui sont ASP67, ASP198 et THR64 (R2 / ASP67, d=2.795 Å ; R3 / ASP198, d=2.671 Å ; R8 / THR64, d=2.743 Å), comme le montre la figure 8. Une autre interaction de type liaison hydrogène faible a été constatée entre le groupement R8 et le résidu THR55 (R8 / THR55, d = 2.922 Å).

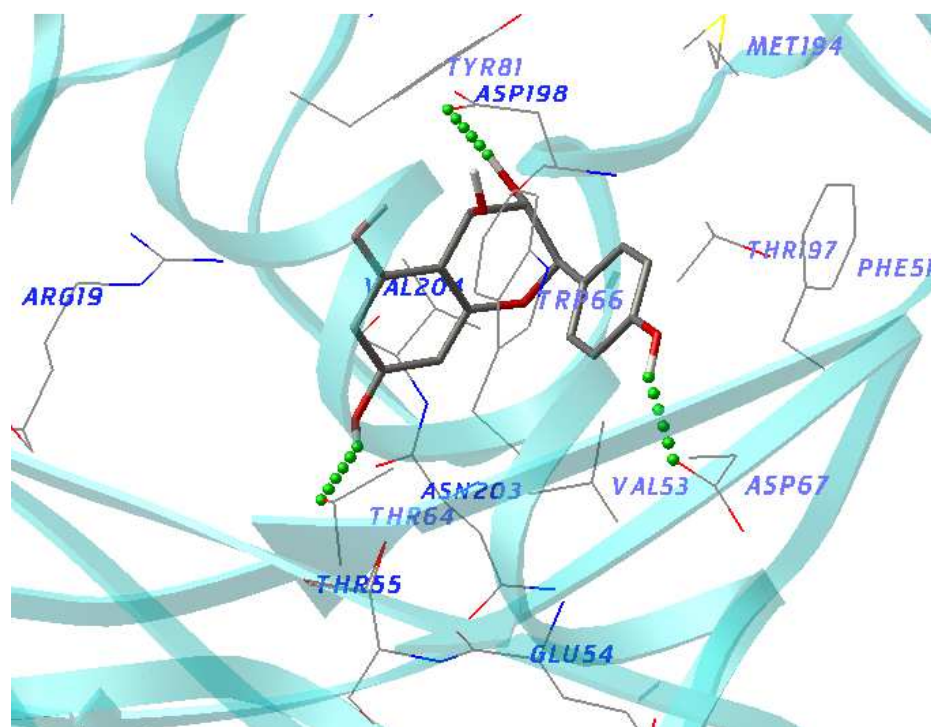


Figure 8 : Le positionnement possible du ligand **flavo1** dans le site actif de l'Arf1. Les liaisons hydrogènes sont représentées en pointillés. Seule la chaîne peptidique du récepteur est représentée en bleu.

La substitution d'un hydroxyle OH en positions R6 ou R7 par un hydrogène dans les composés Flavo2 ou Flavo3 respectivement, entraîne une augmentation dans les énergies des complexes Flavo2/1R8Q E = -23.07 kcal/mol et Flavo3/1R8Q E = -23.05 kcal/mol, voir le tableau 5. Ceci se traduit par une diminution de l'affinité entre ces deux composés et la cible biologique. Cette diminution s'explique par la perte de l'interaction électrostatique entre les deux groupements (R6, R7) et les résidus TYR81 (d = 3.714 Å) et ARG19 (d = 3.327 Å) respectivement (Figure 9).

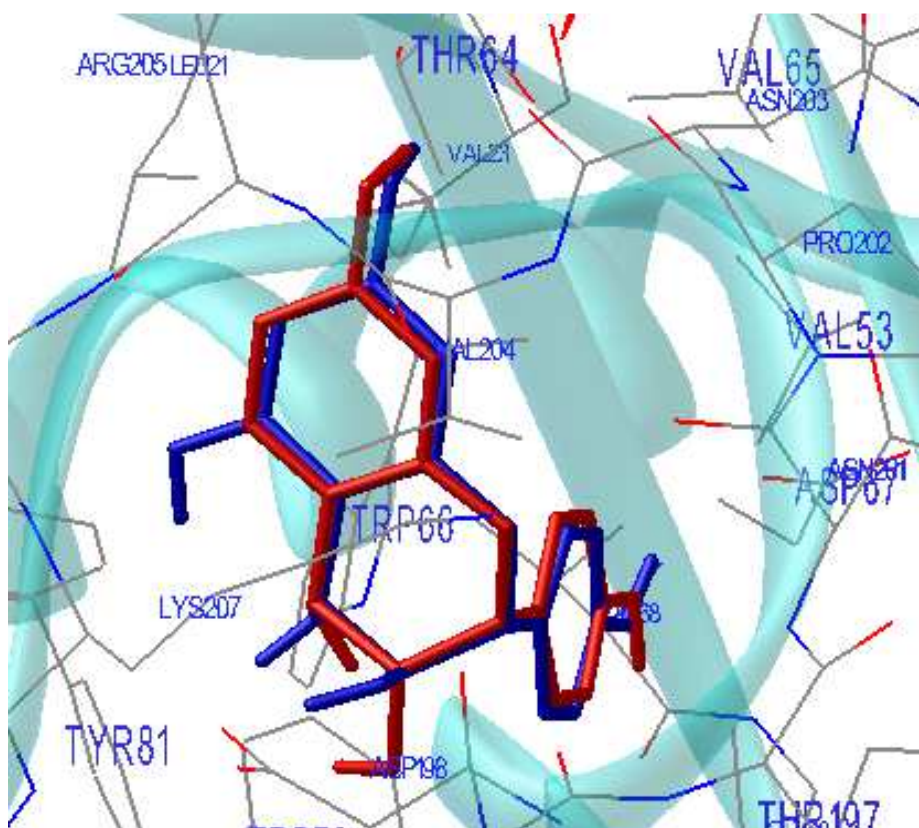


Figure 9 : superposition des meilleurs positionnements des composés: flavo1 (bleu) et flavo3 (rouge)

La suppression des deux groupements hydroxyles en position R6 et R7 donnant le composé Flavo4 entraîne une diminution de l'affinité de ce composé pour la protéine Arf1. La visualisation du complexe prédit viable de ce composé avec la protéine Arf1 montre l'absence de l'interaction électrostatique entre le groupement hydroxyle (OH) en position R7 du composé Flavo1 avec le résidu ARG19, voir la figure 10. La suppression des groupements

L'affinité de ce composé pour la cible Arf1 a augmenté dans ce cas, autrement dit : l'énergie du complexe ligand / récepteur est passée de $E = -23.73$ kcal/mol à $E = -24.0$ kcal/mol, voir le tableau 5. Cette différence dans l'énergie s'explique par la présence d'une liaison hydrogène, dans le composé Flavo7, entre le groupement R6 et le résidu TYR81 ($d = 3.041$ Å), voir la figure 11.

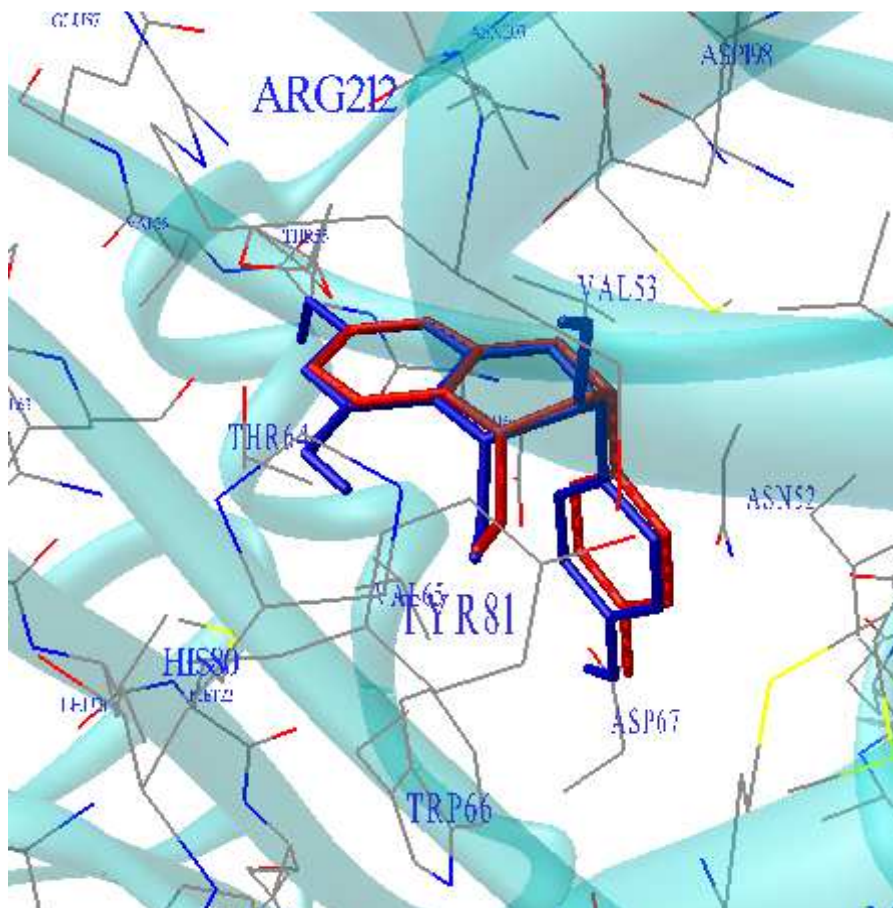


Figure 11 : superposition des meilleurs positionnements des composés: Flavo1 (bleu) et Flavo7 (rouge)

Tableau 7 : Mode de liaison des composés Flavo8-Flavo12
N° : Numéro de la conformation

Molécule	N°	Liaison Hydrogène			Résidus à proximité du ligand (< 4 Å)
		Atome du ligand	Résidu (groupement)	Distance (Å)	
Flavo 8	49	*OH(R ₈) *OH(R ₆)	*TRP78 *ASP67	*3.027(liaison hydrogène) *2.680(liaison hydrogène)	THR55, TYR81, TRP66, THR64, PHE51, VAL53, ASP198, MET194, THR197.
Flavo 9	50	*OH(R ₄) *OH(R ₇) *OH(R ₈) *OH(R ₈)	*THR197 *ASP198 *THR64 *THR55	*3.675 *2.932 *3.342 *3.402	TYR81, TRP78, TRP66, THR64, PHE51, VAL53, MET194, THR197, LEU21, THR55, ILE74, LEU77, TYR190, ASP198, VAL204, MET208.
Flavo 10	15	*OH (R ₇)	*TYR81	*3.023(liaison hydrogène)	TYR81, ASP67, THR64, PHE51, ASN52, VAL53, ASP198, VAL204, MET194.
Flavo 11	41	*OH(R ₄) *OH(R ₆)	*TYR81 *THR64	*4.729 *4.692	TYR81, TRP78, ASP67, TRP66, PHE51, VAL53, MET194, THR197, LEU77, HIS80, MET216.
Flavo 12	34	*OH(R ₄) *OH(R ₆)	*THR197 *TYR81	*5.833 *4.890	TYR81, TRP78, ASP67, TRP66, PHE51, VAL53, MET194, THR197, LEU77, TYR190.

La permutation entre les deux groupements R2 et R3 dans le composé Flavo1 donne le composé Flavo8. Ce changement dans la structure du composé Flavo1 entraîne une diminution de l'affinité pour la protéine Arf1. La suppression de ce groupement (composés Flavo9, Flavo11, Flavo10 et Flavo13) entraîne une importante diminution de l'affinité de ces composés pour la protéine Arf1. L'énergie est passée de E = -24.09 kcal/mol à E = -20.16 kcal/mol en passant du composé Flavo1 à Flavo9. Ces deux manipulations montrent l'importance de la nature du groupement en position R2. La visualisation des complexes ligand/récepteur pour ces deux composés, prédits viables montre que cette diminution d'énergie est due à la perte de l'interaction avec le résidu ASP67. L'insertion d'un groupement méthyle en position R3 du composé Flavo11, composé Flavo12, entraîne des interactions hydrophobes favorables avec les résidus TRP66, TYR81, voir le tableau 7.

Tableau 8 : Mode de liaison des composés Flavo13-Flavo15
N° : Numéro de la conformation

Molécule	N°	Liaison Hydrogène			Résidus à proximité du ligand (< 4 Å)
		Atome du ligand	Résidu (groupement)	Distance (Å)	
Flavo 13	8	*OH (R ₄)	*ASP198 (OH de la chaîne latérale)	*2.975(liaison hydrogène)	THR 55, TYR 81, TRP 66, THR 64, PHE 51, VAL 53, ASP 198, ASN 203, VAL 204, THR 197.
Flavo 14	50	*OH (R ₆)	*ASP198 (OH de la chaîne latérale)	*2.846	TYR81, ASP67, TRP66, VAL65, THR64, VAL53, MET194, ARG19, GLU54, ASP198, VAL204.
Flavo 15	33	*CH ₂ OH (R ₁) *OH(R ₄)	*ASP67 *ASP198	*2.911 *5.274	TRP78, TRP66, PHE51, MET194, THR197, ILE74, ALA157, TYR190, ILE193.

L'introduction d'un groupement alkyle en position R5, composé Flavo14, conduit à une énergie de liaison de $E = -15.58$ kcal/mol. Ce groupement occupe la poche de liaison constituée de résidus TYR81, HIS80, SER191, LEU195, MET208, MET194 et MET216. La faible affinité enregistrée entre ce composé et la protéine Arf1 s'explique par la présence de gênes stériques entre le groupement alkyle en position R5 et les résidus MET194 et MET208. Les distances entre ce groupement alkyle et ces résidus sont seulement ($d = 2.832$ Å) et ($d = 2.574$ Å) respectivement, voir le tableau 8.

Par ailleurs, l'introduction d'un hydroxyméthyle (CH₂OH) en position R1 du composé Flavo13, composé Flavo15 dans le tableau 1, entraîne une légère augmentation de l'affinité de ce composé pour la protéine Arf1. La visualisation du complexe prédit viable par nos calculs de docking montre des interactions électrostatiques entre ce groupement CH₂OH et les résidus du site actif.

Tableau 9 :Mode de liaison des composés Flavo16-Flavo20
N° : Numéro de la conformation

Molécule	N°	Liaison Hydrogène			Résidus à proximité du ligand (< 4 Å)
		Atome du ligand	Résidu (groupement)	Distance (Å)	
Flavo 16	40	*OH(R ₄) *OH(R ₉) *OH(R ₄)	*THR197 *THR197 *TYR81	*3.414 *3.438 *4.541	TYR81, ASP67, TRP66,THR64, PHE51, VAL53, MET194, ASP198, VAL204.
Flavo 17	25	*OH(R ₄)	*TYR81	*5.360	TYR81, ASP67, TRP66, THR64, PHE51, ASN52, VAL53, THR197, LEU21, ASP198.
Flavo 18	17	*OH (R ₄)	*THR64	*3.646	ARG19, PHE82 TYR81, ASP67, TRP66, THR64, PHE51, VAL53, VAL204, MET194, THR197.
Flavo 19	5	*OH (R ₉) *OH (R ₆) *OH (R ₉)	*THR197 (OH de la chaîne latérale) *ASP198 (2OH de la chaîne latérale) *ASP198 (NH du backbone)	*2.799 *2.754 *3.173	THR 55, TYR 81, THR 64, VAL 53, ASP 198, VAL 204, MET 194, THR 197.
Flavo 20	29	*OH (R ₉)	*TYR81	*5.003	THR55, TYR81, TRP66, THR64, PHE51, VAL53, ASN203, VAL204, MET194.

L'introduction de groupements polaire en position R9, composés Flavo16, Flavo17, Flavo18, Flavo19, Flavo20, et l'absence de substituant en R2, R7 et R8 fait baisser l'affinité de ces composés à la protéine Arf1 par comparaison au composé prédit le plus actif Flavo1. Notons que l'introduction d'un hydroxyle en position R6, composé Flavo19, entraine une amélioration de l'affinité pour la protéine concernée voir les tableaux 5 et 9.

II.4. Conclusion

Le travail présenté dans ce chapitre comprend deux parties. Dans la première, nous avons effectué une étude de Modélisation par alignement multiple de la structure de l'Arf1, la protéine impliquée dans le cancer du sein. Cette étude montre que les séquences en acides aminés des structures cristallographiques 1RE0 et 1S9D présentent des manques en acides aminés dans quelques régions en les comparants à la structure RX de code pdb 1R8Q. Les résultats de cette étude nous ont guidé dans le choix de la structure métabolique, de code PDB 1R8Q, pour la suite de notre recherche.

Dans la deuxième partie, nous avons effectué une étude menée au moyen du docking moléculaire semi-flexible et est consacrée à la compréhension du mode de reconnaissance par la protéine Arf1, d'une série de 20 Flavonoïdes issue de la littérature. Le docking du ligand issu de la structure cristallographique BFA a été effectué pour générer les paramètres *optima* du docking entrepris sur l'ensemble des molécules étudiées. Cette étude a mis en évidence un certain nombre d'interactions entre cette famille de composés et la protéine Arf1 ainsi que le meilleur positionnement pour chaque composé dans le site actif. L'analyse de nos résultats montre qu'AutoDock4 positionne les composés étudiés dans le site actif de l'Arf1 occupé par le ligand cristallographique. Notre étude a révélé l'importance de quelques résidus 'tapissant' le site actif, en particulier l'ASP67, ASP198, THR64, THR55 dans la liaison du ligand. Par ailleurs, les groupements hydroxyles des résidus ASP67, ASP198 et THR64 sont prédits, par nos simulations, impliqués dans des interactions par liaison hydrogène avec le ligand. Certains résidus aromatiques, en particulier la TRP66 et TYR81, peuvent être essentiels dans la reconnaissance du ligand moyennant des interactions hydrophobes.

En conclusion, nous pouvons dire que les résidus conservés, en particulier l'ASP67, ASP198, THR64, THR55, TRP66 et TYR81, peuvent être spécifiquement ciblés dans le but de concevoir les structures de nouveau anti-cancer du sein plus puissant. Cette étude structurale par le docking constitue une voie dans le développement de nouveaux inhibiteurs l'Arf1.

Références bibliographiques

- [1] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov PE, The protein Data Bank Nucleic Acids Research. Bourne 2000,28:235-242.
- [2] Renault L, Guibert B, Cherfils J, Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. 2003 Nature, 426:525-530.
- [3] Mossessova E, Corpina RA, Goldberg J, Crystal structure of ARF1*Sec 7 Complex with Brefeldine A and its implication for the guanine nucleotide exchange mechanism. MOL. Cell. 2003, 12: 1403-1411.
- [4] Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE, Structure of a B-DNA dodecamer: confirmation and dynamics. Proc. Natl. Acad. Sci. Usa. 1981, 78: 2179-2218.
- [5] Havsteen BH, The biochemistry and medical significance of the flavonoids. Pharmacol Therap. 2002, 96: 67-202.
- [6] Bolton E, Wang Y, Thiessen PA, Bryant SH, Pub Chem: Integrated Platform of Small Molecules and Biological Activities. Chapter 12 IN Annual Reports in Computational Chemistry. Elsevier: Oxford, UK. 2008, 217-240.
- [7] Gasteiger J, Marsali M, Tetrahedron. 1980, 36: 3219.
- [8] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ, Journal of Computational Chemistry. 1998, 19: 1639-1662.
- [9] Cambridge Scientific Computing (2003) ChemOffice Ultra. Cambridge Scientific Computing, Cambridge.
- [10] Allinger NL, J Am Chem Soc. 1977, 99: 8127-8134
- [11] Michel F. Sanner. Python: A Programming Language for Software Integration and Development. J. Mol. Graphics Mod. 1999, Vol 17, February. pp57-61.

Conclusion générale et perspectives

Les travaux présentés dans ce manuscrit sont orientés vers la recherche d'une voie thérapeutique dans le domaine de la pathologie du cancer du sein, au moyen de simulations bioinformatiques. La première porte sur la modélisation par alignement multiple de quelques structures protéiques, la deuxième consiste en une étude des interactions entre une série de flavonoïdes et la protéine impliquée dans le cancer du sein à savoir l'ADP-Ribosylation Factor 1 (Arf1).

La protéine ADP-Ribosylation Factor 1 (Arf1), une petite protéine G, existe sous six isoformes, Arf1-Arf6, répartis en trois classes. L'Arf1 et l'Arf6 sont les plus étudiées à cause leurs rôles biologiques. Contrairement à l'Arf6 qui joue un rôle important dans l'homéostasie de l'organisme, l'Arf1 est surexprimée dans les cellules invasives du cancer du sein et elle est impliquée dans la migration et la prolifération des cellules cancéreuses. Les liens de l'activité de ces iso-formes, aux effets homéostatiques et indésirables suscitent le développement d'inhibiteurs sélectifs pour la protéine Arf1.

Après avoir résumé les résultats de notre recherche bibliographique concernant l'état actuel des connaissances sur les protéines ARFs, nous avons présenté les méthodes de la modélisation moléculaire.

Notre travail de recherche est divisé en deux grandes parties. La première, consiste à modéliser la structure de la protéine Arf1 au moyen d'une méthode bioinformatique à savoir l'alignement multiple. Cette partie nous a permis d'identifier les différents motifs biologiques qui constituent la protéine Arf1 et qui ne sont pas conservés dans les trois structures étudiées ce qui nous a permis de constater la grande flexibilité de la protéine. Elle nous a permis aussi de choisir, parmi les structures cristallines de l'Arf1 disponibles dans la PDB, la structure métabolique que nous allons utiliser dans la suite de notre travail de recherche.

La deuxième partie de notre travail de recherche, consiste à étudier les interactions entre une série de vingt flavonoïdes et les résidus du site actif de l'Arf1. Pour cela, nous avons effectué des simulations de docking moléculaire pour déterminer les différents modes de liaison de cette famille de composés. La mise en évidence de résidus pertinents dans le site de liaison, nous a permis de discuter les différences d'énergies de liaisons prédites par notre étude pour ces composés. En résumé, nous pouvons dire que dans cette partie de notre travail, deux objectifs sont atteints. Le premier est la mise en évidence des interactions prédominantes ainsi que les acides aminés les plus importants pour la liaison d'un flavonoïde avec la protéine

Arf1 qui sont l'ASP67, ASP198, THR64, THR55, TRP66 et TYR81. Le second objectif est de donner une interprétation de la variation des énergies de liaisons prédites par AutoDock en fonction des variations structurelle des composés étudiés.

En conclusion, les recherches, réalisées au cours de la simulation du docking des flavonoïdes, ont permis la proposition de modèles de prédiction de la conformation bioactive. Cependant, le logiciel AutoDock4, utilisé dans ce travail ne permet pas une flexibilité totale à la protéine Arf1. En revanche, du fait de la grande flexibilité du site actif de la protéine, ces modèles sont insuffisants pour guider la conception directe de nouveaux inhibiteurs de l'Arf1. Afin de remédier à ces insuffisances et d'aller plus loin dans nos investigations, nous comptons poursuivre ce travail en utilisant des techniques plus poussé permettant d'appréhender la flexibilité de la protéine, comme la dynamique moléculaire. Il est bien de noter que les énergies de liaisons présentées dans ce mémoire sont prédites théoriquement par notre étude. Cependant, une étude expérimentale complémentaire soit in vitro ou in vivo est nécessaire pour vérifier la validité de ces résultats.