Choix d'une cible à intérêt biologique:

Dans Le cadre de notre thématique de recherche qui porte sur le développement de nouveaux inhibiteurs de l'α-glucosidase, l'enzyme Maltase Glucoamylase (MGAM) intestinale de souche humaine (EC: 3.2.1.20) de la famille GH31[1] a été sélectionnée comme cible biologique pour realiser notre criblage qui vise le développement de nouveaux dérivés de xanthones inhibiteurs de l'α-glucosidase en tant que substances médicamenteuses candidates.

La structure tridimensionnelle de format Pdb de la protéine qui représente le domaine catalytique de MGAM (ntMGAM), co-cristalisée avec le Miglitol est disponible dans la Protein Data Bank sous le code PDB 3L4W avec une résolution de 2.0Å. Ce récepteur est un monomère de protéine avec une séquence d'une longueur de 875 acides aminés. Le miglitol identifié précédemment en tant qu'excellent inhibiteur de MGAM a été utilisé au cours de cette étude comme inhibiteur de référence afin d'identifier le site actif du récepteur.

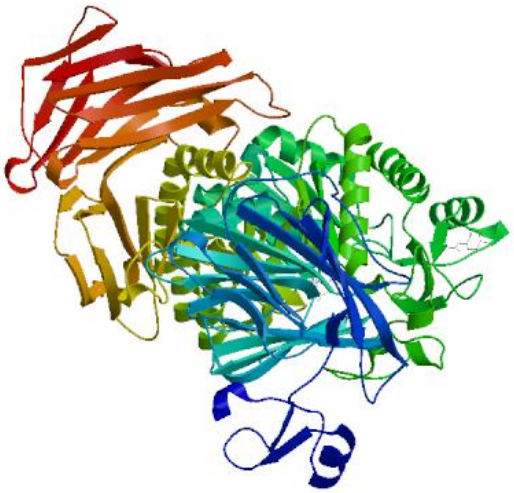

Figure I. 1. Structure de Maltase GlucoAmylase (PDB : 3L4W)

2. Chimiothéque de ligands : Création et préparation

Notre choix s'est porté sur une famille de ligands cibles ayant prouvé leur effet inhibiteur selectif vis-à-vis de l'α-glucosidase : les dérivés de xanhtones [2]. Une sélection de ces dérivés à partir de la littérature a mené à la construction d'une chimiothéque virtuelle à cribler comprenant 189 dérivés de xanthones qui n'ont pas encore montré une activité

inhibitrice vis-à-vis de l'α-glucosidase. Leurs structures moléculaires sont illustrées dans l'annexe (X_1 à X_{189}).

On peut noter que certains ligands ne sont pas décrits dans la littérature. En effet, ils ont été sélectionnés après avoir apporté quelques modifications structurales virtuelles, dans le but d'enrichir notre librairie et de générer une base de ligands divers en terme de structure, en variant les groupements fonctionnels portés par le squelette de base dibenzo-γ-pyrone.

Apres avoir créé la librairie, les structures des ligands ont été dessinées, transformées en 3D en format SDF et ensuite minimisées afin d'obtenir une conformation de basse énergie pour chaque ligand à l'aide du logiciel Hyprechem7 en utilisant la méthode semi-emperique PM3.

Le processus du docking moléculaire:

Tous les docking effectués lors de cette étude ont été assistés par Molegro Virtual Docker (MVD 2012.5.5.0, Molegro ApS). Notre choix de ce logiciel est justifié par sa meilleure performance pour étudier et prédire le mode d'interaction ligand-récepteur en comparaison avec d'autres logiciels fréquemment utilisés tels que Autodock, Superflex, Flex and Gold[3-5].

Pour une meilleure visualisation après les simulations du docking, les résultats obtenus ont été stockés et traités avec le logiciel PyMol[6].

3.1. Post-préparation du récepteur et des ligands par MDV:

Les fonctions de préparation automatique intégré dans MVD permettent d'importer le complexe cible et les ligands dans l'espace de travail «work space» dans lequel le récepteur est considéré comme un corps rigide en gardant les résidus fixes et les ligands sont flexibles avec l'exclusion des molécules d'eau. Les préparations sont principalement :

- ajout des hydrogènes explicites
- assignement des charges
- correction des types des liaisons (simple ou double).
- détection des torsions et les liaisons rotatives.

3.2. Détection des cavités et définition du site actif :

Molegro virtual Docker (MVD) possède un algorithme de détection de cavités moléculaires dans le volume tridimensionnel du récepteur afin de prédire les sites d'interactions possibles. Cet algorithme basé sur une grille « grid-based » fonctionne comme suit:

Une grille discrète d'une résolution de 0,8 Å couvrant tout le volume de la protéine est créée. A chaque point de cette grille, une sphère de rayon 1,4 Å est placée en vérifiant l'intersection de ces sphères avec celles de Van der Walls des atomes de protéine. Les points de grille qui chevauchent avec les atomes de protéine sont définis comme points inaccessibles et les autres sont considérés points accessibles.

Chaque point accessible de cette grille sera analysé afin de vérifier s'il fait partie de la cavité par l'insertion des lignes de directions aléatoires à partir de ce point jusqu'à franchir les limites de la grille. Si la majorité des directions franchissent d'autres points accessibles au cours de leurs chemin (à revoir défaut au moins 12 sur 16 directions différentes testés), ce point est consideré faisant partie d'une cavité. Ceci est répété un certain nombre de fois pour définir tous les points accessibles des cavités. La dernière étape consiste à déterminer les régions de cavités en connectant les points voisins. Les cavités trouvées sont ensuite classées en fonction de leur volume.

Par le biais de cet algorithme, nous avons détecté cinq cavités pour la MGAM (Figure2)

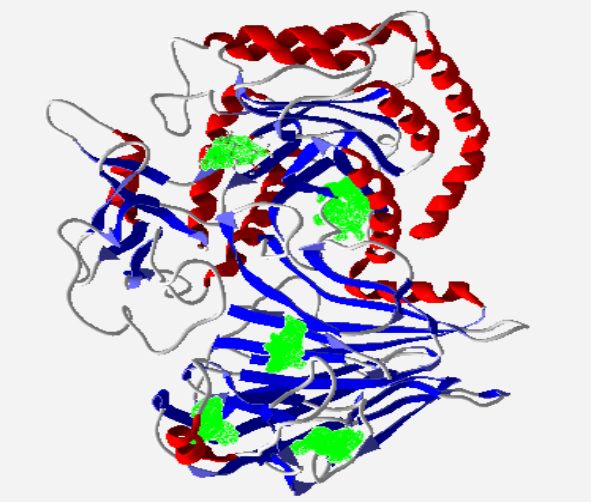

Figure I. 2. Illustration des cinq cavités du MGAMdétectées (en vert) par l'algorithme de détection de MVD

Les cavités sont classées selon leur volume (tableau 1).

Tableau I. 1: les cavités détectées avec l'algorithme de détection de MVD pour 3L4W

Cavité détectée	**Volume ($Å^3$)**	**Surface ($Å^2$)**	**Coordonnés (X,Y,Z)**
Cavité 1	86.02	312.32	(50.44, 111.11, 21.79)
Cavité 2	68.09	240.64	(55.10, 77.11, -7.6)
Cavité 3	62.46	179.2	(41.55, 76.92, -3.55)
Cavité 4	53.76	188.16	(54.08, 73.55, 8.67)
Cavité 5	51.2	147.2	(44.95, 91.39, 35.06)

Lors de notre investigation des cavités candidates pour un éventuel site d'interaction de notre étude docking, à l'aide de l'algorithme de prédiction de cavités, nous avons remarqué que le ligand de référence (miglitol) est co-cristallisé dans la cavité 5 et entouré par les résidus catalytiques. Nous avons donc décidé de conserver cette cavité pour la considérer comme centre de recherche (research center) pour le processus du docking. Cette sphère de recherche couvre tous les résidus du site actif avec un rayon de 15 Å.

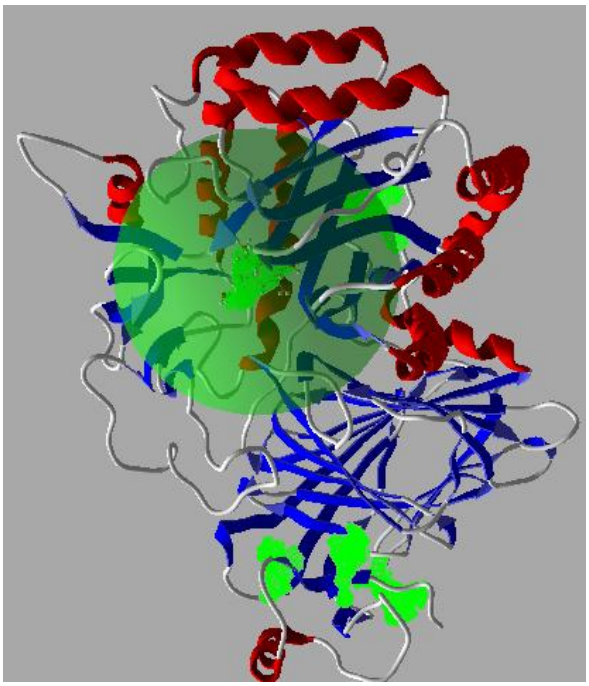

Figure I. 3. Illustration de l'espace de recherche

3.3. Protocole et validation :

3.3.1. Algorithme de recherche :

L'algorithme de recherche MolDock Optimizer utilisé dans MVD est un algorithme évolutionnaire basé sur une évolution différentielle guidée (Guided Differential Evolution). L'innovation dans cet algorithme introduite par Storn et Price[7], en comparaison avec d'autres algorithme évolutionnaires connus, consiste en une approche différente pour sélectionner et modifier les solutions de candidats (individus). La principale idée innovante dans le DE est la création de solutions dites "enfants" à partir d'une différence par rapport aux individus initiaux "parents" de la population.

-Représentation des individus de la population :

Seules les propriétés des ligands sont représentées pour les individus puisque la protéine reste rigide lors du docking. Ainsi, chaque solution est encodée en représentant sa position, son orientation et sa conformation avec des coordonnées cartésiennes, un vecteur de rotation, un angle de rotation et un angle de torsion pour chaque liaison flexible.

Ces paramètres sont attribués initialement au ligand de manière aléatoire pour définir les individus "parents" de la population. Pour chaque individu de la population initiale, une position aléatoire à l'intérieur de l'espace de recherche (déjà défini) est attribuée. L'initialisation de l'orientation se déroule en choisissant simplement l'axe d'orientation (entre -1,0 et +1,0, suivi par la normalisation des valeurs pour former un vecteur unitaire) et l'angle de rotation compris entre -180° et + 180°. Les angles flexibles de torsion (le cas échéant) se voient attribuer également un angle aléatoire entre -180°et + 180°.

-Principe de l'algorithme:

Une des raisons pour lesquelles l'algorithme évolutionnaire différentielle est attribué au MVD c'est sa capacité à nous permettre l'exploitation d'une diversité importante de la population: la progression des solutions parentales ,initialement générées de manière aléatoire avec une grande variation, entrainera la création de nouvelles solutions "enfants" largement différentes.

Cet algorithme converge vers les meilleures solutions évaluées à l'aide d'une fonction de score, ce qui a pour résultat la diminution de la taille de la population et la création de meilleurs solutions enfants. Ensuite, les solutions enfants remplacent les parents, si et

seulement si elles sont plus adaptées. Dans le cas contraire, le parent survit et est transmis à la prochaine génération (appelé itération de l'algorithme) (figure 4). Ce processus est exécuté aussi longtemps que la condition de fin de recherche n'est pas remplie.

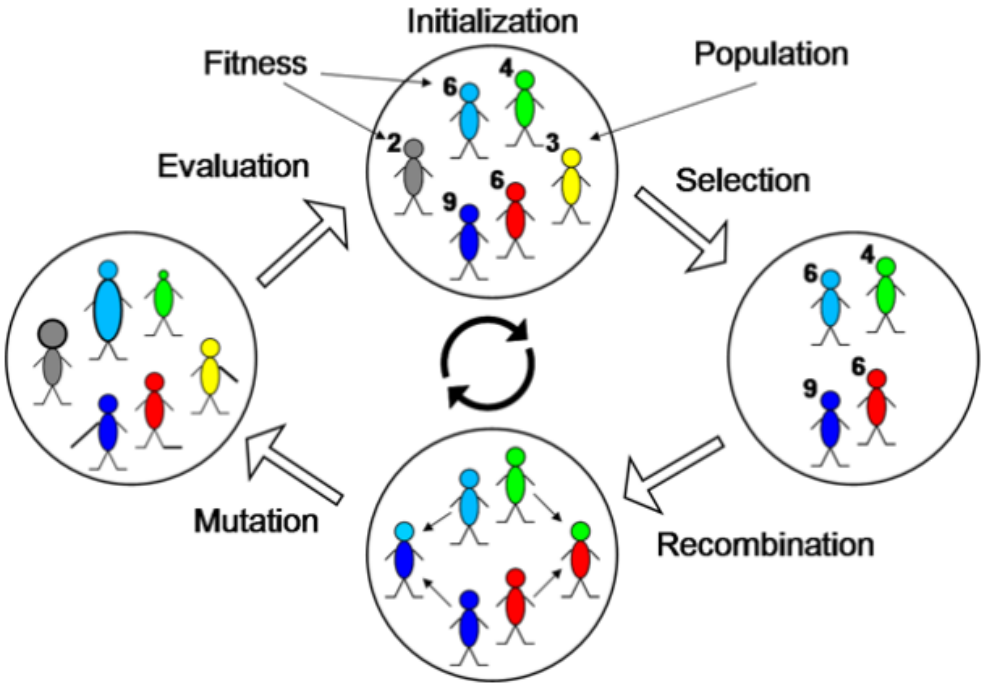

Figure I. 4. Principe de l'algorithme évolutionnaire différentiel de MDV

Nous avons utilisé Moldock optimizer comme algorithme de recherche avec le réglage algorithmique suivant:

- Nombre de tours (runs) : 100
- Taille de population : 150
- Itération maximale de l'algorithme: 2000
- Taux de croisement = 0,9
- Fin de recherche : basée sur la variation «variation-based»

3.3.2. Fonction de score:

Lors de la simulation du docking plusieurs poses candidates peuvent être obtenus pour chaque ligand, ici, nous sélectionnons les meilleures poses avec les énergies les plus basses par le biais de la fonction de score basée sur une grille «MolDock Score [Grid]» qui précalcule les valeurs d'énergies d'interaction sur une grille cubique régulièrement espacée

afin de réduire le temps de calcul. Cette fonction de score utilisée par MVD est un dérivé amélioré des fonctions de score PLP (piecewise linear potential) initialement proposées par Gehlhaar et al.[8, 9] et développées plus tard par Yang et al.[10] avec l'insertion de nouveaux termes de liaisons hydrogène et de charges.

La fonction de score, E_{Score}, est définie par les termes d'énergie suivants:

$$E_{score} = E_{inter} + E_{intra} \qquad eq.1$$

Avec : E_{inter} : énergie d'interaction ligand-protéine

E_{intra} : énergie d'interaction intramoléculaire du ligand

L'énergie d'interaction ligand-protéine E_{inter} est donnée par l'équation suivante:

$$E_{inter} = \sum_{i \in ligand} \sum_{j \in proteine} \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4 r_{ij}^2} \right] \qquad eq.2$$

Cette équation prend en compte tous les atomes lourds (i) du ligand et les atomes lourds (j) de la protéine ainsi que les atomes des cofacteurs et des molécules d'eau (le cas échéant).

- E_{PLP}: le potentiel linéaire PLP

- $332.0 \frac{q_i q_j}{4 r_{ij}^2}$: décrit les interactions électrostatiques entre les atomes chargés. La valeur numérique 332.0 est donnée pour fixer l'unité de l'énergie électrostatique en Kcal/mol.

- $D(r) = 4 r_{ij}$: constante diélectrique

- q_i et q_j: representent respectivement les charges des atomes du ligand et des atomes de la protéine.

Des exemples de charges des atomes sont donnés dans le tableau suivant :

Tableau I. 2: Templates de charges prisent en compte par MolDock Score [Grid].

Atomes de ligand	Atomes de protéines	Charges
Les atomes N dans $-C(NH_2)_2$	His (atomes N) Arg (atomes N)	0.5
Les atomes N dans $-N(CH_3)_2$ et $-(NH_3)$	Lys (atomes N)	1.0
Les atomes O dans -COO, $-SO_4$, $-PO_2$	Asp (atomes O) Glu (atomes O)	-0.5
Les atomes O dans $-PO_3$		-0.66
Les atomes O dans $-SO_3$		-0.33
Les atomes N dans $-SO_2NH$		-0.1
Ions métalliques tel que Na^+		+1.0
Ions métalliques tel que Ca^{+2}et Fe^{+2}		+2.0

Le potentiel E_{PLP} prend en considération principalement deux paramètres; le premier optimise les interactions stériques (Van der Waals) entre les atomes et l'autre optimise les potentiels des liaison hydrogènes.

Une liaison est considérée comme une liaison hydrogène si l'un des atomes peut donner un atome d'hydrogène et l'autre atome peut l'accepter. Les types d'atomes qui peuvent contribuer à la formation de liaisons hydrogène sont classés selon le tableau 3.

Tableau I. 3: Type d'atomes contribuant à la formation d'une liaison hydrogène.

Atomes	Types
N et O sans H liés	Accepteur
N et S avec un ou plusieurs H liés	Donneur
O avec H attachés	Accepteur et donneur

L'équation suivante représente l'énergie d'interaction intramoléculaire du ligand E_{intra} :

$$E_{intra} = \sum_{i \in ligand} \sum_{j \in ligand} E_{PLP}(r_{ij}) + \sum_{liaisons\ flexibles} A\,[1 - \cos(m.\theta - \theta_0)] + E_{clash} \qquad eq.3$$

Le premier terme de cette équation correspond à la somme d'énergies d'interaction entre toutes les paires d'atomes du ligand. Le second terme représente l'énergie de torsion, cette énergie dépend essentiellement de l'hybridation des atomes liés (tableau 4). θ est l'angle de torsion de la liaison, cet angle ne correspond pas nécessairement à une seule torsion, une moyenne est définie dans le cas où plusieurs torsions sont déterminées. Le dernier terme, E_{clash}, correspond à une constante de pénalité attribuée à l'équation afin d'éliminer les conformations infaisables du ligand; la valeur de cette constante est de 1000 si la distance entre deux atomes lourds est inférieure à 2,0 Å. Si un atome lourd du ligand est situé en dehors de la région du site d'interaction (l'espace de recherche) une constante de pénalité E_{clash} de 10000 est attribuée.

Tableau I. 4: Les paramètres correspondant à l'énergie de torsion.

Liaison	θ_0	m	A
sp^2-sp^3	0.0	6	1.5
sp^3-sp^3	π	3	3.0
sp^2-sp^2	0.0	2	3.0

A la fin du processus du docking, un ensemble de poses prometteuses est généré et les termes d'énergie calculés (tableau 5) sont stockés et regroupés dans un fichier de format «mvdresults» destiné à être inspecté dans le «work space» du MVD ou avec Molegro Data Modeller.

Tableau I. 5: Les termes d'énergie calculés avec la fonction de score MolDock Optimizer.

Terme d'énergie	Signification
$\mathbf{E_{Tolat}}$	Energie totale MolDock Score
$\mathbf{E_{Inter\ (protein\ -\ ligand)}}$	Energie d'interaction MolDock Score totale entre le ligand et la proteine).
$\mathbf{E_{Steric}}$	Energie d'interaction stérique entre la protéine et le ligand (calculé par PLP).
$\mathbf{H_{Bond}}$	Energie de liaison hydrogène entre la protéine et le ligand (calculé par PLP).
Electro	Energie d'interaction électrostatique protéine-ligand pour courte distance interatomique (r <4.5Å)
ElectroLong	Energie d'interaction électrostatique protéine-ligand pour longue distance interatomique (r >4.5Å)
$\mathbf{E_{Intra}}$	Energie total MolDock Score interne du ligand (interactions intramoléculaires).
$\mathbf{E_{Intra\ (steric)}}$	Energie d'interaction stérique du ligand (calculé par PLP).
$\mathbf{E_{Intra\ (hbond)}}$	Energie de liaison hydrogène du ligand (calculé par PLP).
$\mathbf{E_{Intra\ (elec)}}$	Energie d'interaction électrostatique du ligand
$\mathbf{E_{Intra\ (tors)}}$	Energie de torsion du ligand

3.3.3. Validation du protocole de docking:

Pour valider notre protocole du docking basé sur une combinaison de l'algorithme de recherche Moldock Optimizer et la fonction de score Moldock Score [Grid], nous avons évalué la précision de positionnement de la meilleure pose (top-scoring) du ligand de reference "Miglitol". Cette pose est obtenue après application du protocole du docking sur le complexe miglitol-site actif qui a généré plusieurs poses. La meilleure est choisie selon l'énergie totale d'interaction (MolDock Score) la plus basse d'une valeur de -81.747 Kcal /mol.

Cette énergie sera également utilisée comme énergie de référence avec les autres ligands de la chimiothéque à cribler. Cette validation nécessite deux étapes.

La première étape est le «Re-docking», considéré comme la méthode d'evaluation la plus populaire. Elle a pour but la détermination de la capacité du protocole du docking à générer une pose avec un positionnement dans le site d'interaction identique ou très proche à celui de la structure cristallographique originale. Cette méthode s'intéresse à l'écart quadratique moyen « RMSD » (Root Mean Square Deviation)[11], qui est une mesure géométrique de la distance entre les positions atomiques du ligand de référence co-cristallisé expérimentalement avec les cordonnés $(x_{cristal}, y_{cristal}, z_{cristal})$ et celles de la structure prédite ligand (pose) $(x_{pose}, y_{pose}, z_{pose})$ repositionnée après application du protocole du docking dans le site d'interaction. L'écart RMSD est donné par l'équation suivante:

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{cristal_i} - x_{pose_i})^2 + (y_{cristal_i} - y_{pose_i})^2 + (z_{cristal_i} - z_{pose_i})^2} \qquad eq.4$$

Avec :

- n : le nombre total d'atomes lourds.

- $(x_{cristal_i}, y_{cristal_i}, z_{cristal_i})$ et $(x_{pose_i}, y_{pose_i}, z_{pose_i})$ sont respectivement les cordonnées cartésiennes des atomes identiques de la structure expérimentale et de la structure prédite.

Il est considéré que les meilleurs protocoles du docking génèrent des valeurs RMSD inférieure à 2,0Å, cette valeur est un seuil au-dessus duquel la précision du positionnement est jugée mauvaise, entrainant l'invalidation du protocole[12].

Pour notre étude, le calcul de la valeur RMSD est effectué avec MVD en appliquant le protocole du docking décrit précédemment avec définition du miglitol co-cristallisé comme un ligand de référence. La simulation du re-docking a généré un RMSD équivalent à 0.31 Å (figure 5), largement inférieur au seuil, ce qui indique que notre protocole a réussi et qu'il est applicable pour le reste de cette étude.

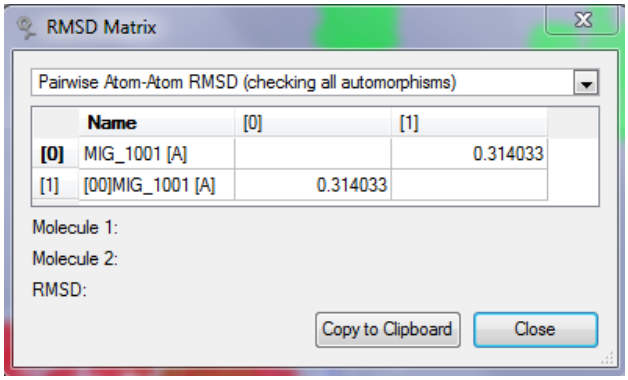

Figure I. 5. Matrice de RMSD générée par MDV après la simulation du re-docking

La superposition du miglitol et sa meilleure pose générée (figure 6), visualisée par Pymol, confirme la précision du positionnement et la génération d'une conformation quasi similaire à celle cristallographique originale.

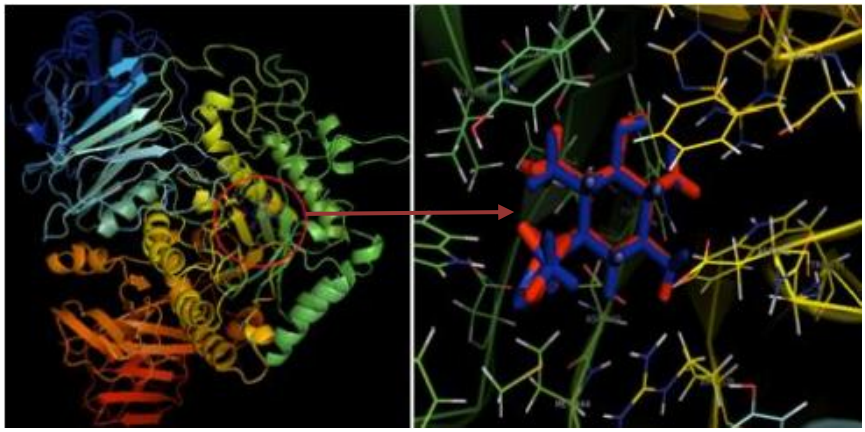

Figure I. 6. La superposition de la structure cristallographique du ligand (en bleu) et la pose obtenue après re-docking (en rouge), RMSD =0.31Å

La deuxième partie consiste en une classification de précision du positionnement basée sur les interactions «interactions-based accuracy classification (IBAC)» qui repose sur une inspection visuelle des structures co-cristallisées et prédites[13], le grand avantage de cette classification par rapport au RMSD est la prise en compte des interactions. Cette classification débute par une inspection visuelle du complexe co-cristallisé ligand-site d'interaction pour déterminer précisément l'ensemble des interactions, suivie par une autre inspection du complexe pose prédite-site d'interaction. Le positionnement de la pose prédite est dit correct si la conformation et son orientation ainsi que toutes les interactions clés sont similaires à celles de la structure originale du ligand de référence. Dans le cas contraire, le positionnement est déclaré incorrect.

En inspectant visuellement le site d'interaction du complexe cristallographique 3L4W, le miglitol montre une bonne affinité avec les résidus du site actif et notamment les résidus catalytiques. Cette affinité qui est à l'origine de son potentiel inhibiteur, se traduit par la formation de liaisons hydrogène intéressantes avec les résidus Asp 327, His 600, Arg526 et notamment avec les résidus catalytiques Asp 443 et Asp542.

En observant de plus près les résidus en interaction avec la pose (Figure 7), nous avons pu constater que cette pose interagit avec la même manière que le ligand cristallographique, grâce à des liaisons hydrogène de les même résidus: Asp 327, His 600, Arg526, le catalyseur nucléophile Asp 443 et le catalyseur acide / base Asp542.

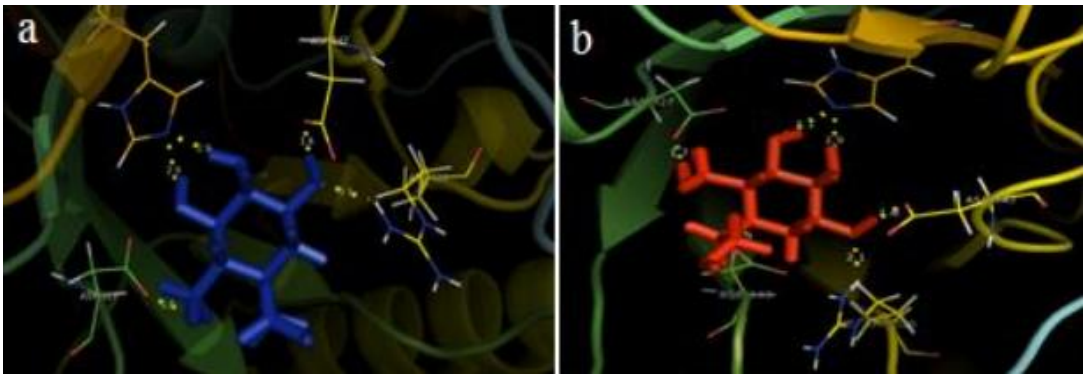

Figure I. 7. visualisation comparative entre le mode d'interaction du ligand cristallographique avec les résidus adjacents (a), et celui de la pose dans le même site d'interaction (b)

La conformation et l'orientation de la pose prédite quasi identiques à celles du ligand de référence prouvé avec un RMSD de 0.31Å et les modes d'interaction très similaires indiquent une excellente précision de positionnement validant ainsi notre protocole.

4. Criblage de la chimiothèque :

Le criblage débute par l'importation de tous les ligands de notre chimiothéque de 189 dérivés xanthone dans le «work space» du MVD, les ligands subissent une post-préparation avec l'outil de préparation automatique du logiciel afin de vérifier et de s'assurer que les propriétés structurales de chaque ligand sont prises en compte par le logiciel.

Une fois la chimiothéque importée, tous les ligands seront orientés vers l'espace de recherche déjà défini en appliquant le protocole précédemment décrit et validé. Le «scoring» des poses obtenues pour chaque ligand (après 100 tours de recherche) mène à un classement des meilleures poses basé sur leurs énergies d'interaction «Moldock score». Ainsi, pour chaque élément de la chimiothéque, une seule pose prédite est sélectionnée, et donc 189 poses