

UNIVERSITÀ DEGLI STUDI DI FIRENZE  
Dipartimento di Scienze Farmaceutiche

SCUOLA DI DOTTORATO IN SCIENZE  
CORSO DI DOTTORATO IN  
CHIMICA E TECNOLOGIA DEL FARMACO

XXI CICLO (2006-2008)  
*Settore scientifico disciplinare CHIM/08*

**Strumenti della modellistica molecolare applicati allo  
studio e ottimizzazione strutturale di composti di  
interesse farmaceutico**

**Candidato**

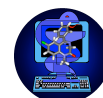
*Dr. Jacopo Sgrignani*

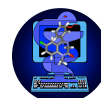
**Docente supervisore**

*Prof. Paola Gratteri*

**Coordinatore del corso di dottorato**

*Prof. Fulvio Gualtieri*





## Sommario

	Pag
<b>Introduzione</b>	3
<b>1. Descrizione dei target</b>	6
<b>1.1 Recettore nicotinico neuronale</b>	7
<b>1.2 Glicosilidrolasi</b>	13
1.2.1 <i>Beta-glicosidasi</i>	16
1.2.2 <i>Trealasi</i>	19
1.2.3 <i>Maltasi-glucoamilasi</i>	21
<b>2. Strumenti e Metodi</b>	23
<b>2.1 Meccanica Molecolare</b>	24
<b>2.2 Metodi basati sulla funzione d'onda</b>	30
2.2.1 <i>Metodo di Hartree Fock</i>	30
2.2.2 <i>Density Functional Theory (DFT)</i>	32
<b>2.3 Metodi QM/MM</b>	34
<b>2.4 Dinamica Molecolare (MD)</b>	36
2.4.1 <i>Note generali</i>	36
2.4.2 <i>Impostare un esperimento: aspetti pratici collegati alla realizzazione di un esperimento di MD</i>	39
<b>2.5 Docking</b>	43
2.5.1 <i>Uno fra i molti software di docking molecolare Glide Schördinger®</i>	47
<b>2.6 Metodi per il calcolo della free energy</b>	51
2.6.1 <i>Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA)</i>	51
2.6.2 <i>Linear Interaction Energy (LIE) o Linear Response Method (LRM) o Linear Interaction approximation (LIA).</i>	53
2.6.3 <i>Thermodynamic Integration</i>	53
<b>3. Sezione sperimentale</b>	57
<b>3.1 Recettore nicotinico neuronale <math>\alpha 4\beta 2</math></b>	58
<b>3.2 Glicosilidrolasi</b>	61

<b>4. Risultati e discussione</b>	66
<b>4.1 Recettore nicotinico neuronale <math>\alpha 4\beta 2</math></b>	67
<b>4.2 Glicosilidrolasi</b>	90
4.2.1 <i>Beta-glicosidasi</i>	92
4.2.2 <i>Trealasi</i>	98
4.2.3 <i>Maltasi-glucoamilasi</i>	105
<b>Bibliografia</b>	112

*L'ipotesi più importante di tutta la biologia, per esempio, è che tutto ciò che gli animali fanno, lo fanno gli atomi.*

*In altre parole, non vi è nulla che gli esseri viventi facciano che non possa essere inteso partendo dal punto di vista che essi sono composti di atomi i quali agiscono secondo le leggi della fisica.*

*Richard Phillips Feynman (1918 – 1988)*

## Introduzione

Lo studio della modalità di interazione fra ligandi e target macromolecolari rappresenta uno degli aspetti fondamentali del lavoro del moderno chimico farmaceutico.

Questo lavoro assume al giorno d'oggi, complice anche la sempre maggiore disponibilità di informazioni strutturali ottenute sperimentalmente tramite diffrazione a raggi X e spettroscopia NMR, i connotati di un progetto di "architettura" molecolare, nel quale si cerca di ottenere la massima complementarità fra i due partner, ossia il ligande e il target.

Gli approcci computazionali utilizzabili nel disegno e l'ottimizzazione di un nuovo ligande possono essere suddivisi in due classi: quelli che utilizzano la conoscenza della struttura tridimensionale (3D) del target biologico (*Structure-Based Drug Design, SBDD*) e quelli che sfruttano la possibilità di definire un sito di interazione virtuale (*Virtual Receptor Site, VRS*) a partire dall'informazione derivata dall'analisi di un insieme di molecole, congeneriche e non, interagenti tutte con lo stesso profilo farmacologico su un dato recettore (*Ligand-Based Drug Design, LBDD*).

In questo lavoro di tesi vengono presentate due applicazioni di SBDD.

Nella prima è stato affrontato il calcolo dell'energia libera di binding fra alcuni ligandi e il recettore nicotinico neuronale  $\alpha 4\beta 2$ , per mezzo dell'uso congiunto di due metodiche computazionali, la Dinamica Molecolare (MD) e il docking QM/MM.

La seconda applicazione ha riguardato sia la messa a punto di strategie computazionali per il calcolo dell'energia libera di binding di inibitori di enzimi appartenenti alla classe delle glicosil idrolasi e la comprensione dei determinanti molecolari che regolano l'inibizione.

Questo secondo progetto ha visto l'applicazione sia di metodi rigorosi, come la thermodynamic integration (TI), sia di metodi più rapidi come l' MM-PBSA

(Molecular Mechanics-Poisson Boltzman Surface Area) e la Linear Interaction Energy (LIE).

## *Descrizione dei Target*

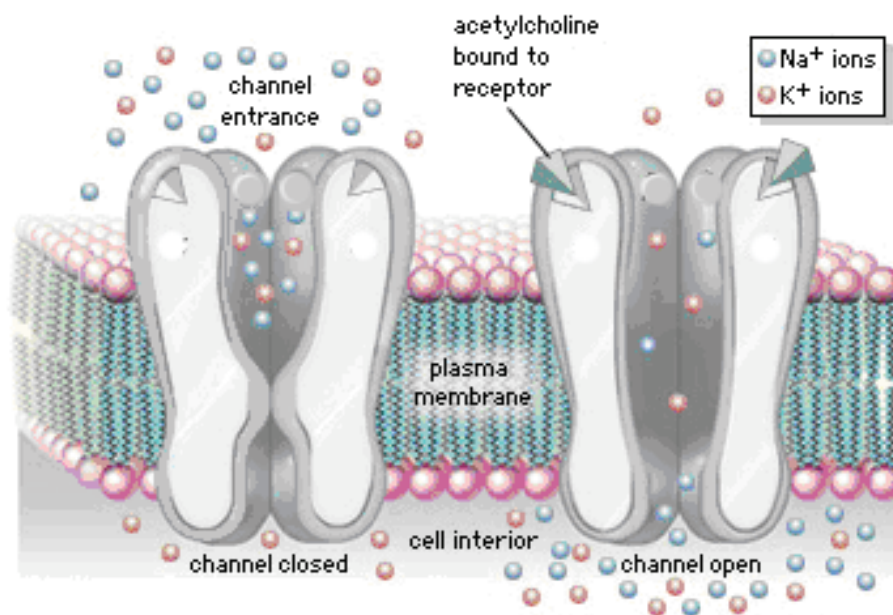
*Ecco perché ancora oggi io vo d'intorno investigando e ricercando...se ci sia alcuno...che io possa ritenere sapiente; e poiché sembrami che non ci sia nessuno, io vengo così in aiuto al dio dimostrando che sapiente non esiste nessuno.*

*da Platone, Apologia di Socrate.*



## Recettore nicotinic neuronale

I recettori nicotinici nAChRs sono recettori ionotropi (ligand-gated ion channel, LGIC) appartenenti alla superfamiglia dei Cys-loop receptor. Sono recettori di membrana formati da cinque subunità di tipo sia omomero (della sola tipologia  $\alpha$ ) che eteromero (di tipo  $\alpha$  e  $\beta$  nel caso dei recettori neuronali; di tipo  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$  nel caso dei recettori della placca muscolare) (Figura 1).



**Figura 1.** Figura schematica della struttura del nAChR.

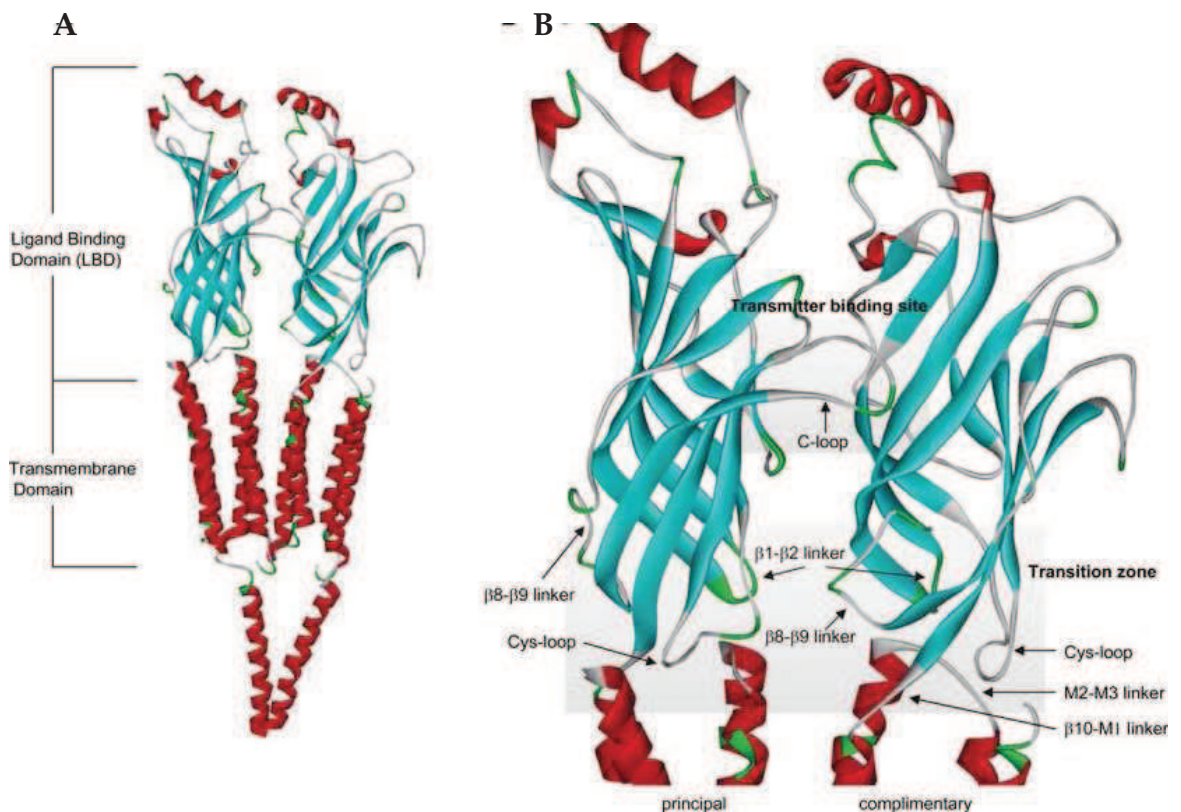
Un notevole impulso alla comprensione della funzione e struttura del nAChR si è avuto nel 2001, quando è stata resa disponibile nel Protein Data Bank (PDB)<sup>1</sup> la struttura, risolta mediante diffrazione ai raggi X, di una proteina solubile, ottenuta da *Lymnea stagnalis*, omologa della porzione N-terminale extracellulare del recettore nicotinic, l'Acetylcholine binding protein (AChBP).<sup>2</sup> Un ulteriore passo nella direzione della comprensione delle caratteristiche di struttura e funzione di questo recettore è stato segnato nel 2005 da Unwin che ha ottenuto la risoluzione mediante microscopia elettronica (risoluzione 4 Å) della struttura completa dell'nAChR di *Torpedo californica*.<sup>3</sup>

Da un punto di vista strutturale i nAChRs si compongono di tre parti (**Figura 2**):

- una porzione extracellulare N-terminale (NTD, N-terminal domain) che ospita il sito di binding e come tale anche detta ligand binding domain LBD;
- una porzione transmembrana (TM) che con i suoi 4 domini (M1, M2, M3, M4) forma il canale vero e proprio (ICD, Ion channel domain);
- un dominio C-terminale intracellulare.

Il LBD è formato da due insiemi di  $\beta$ -strand affacciati, ciascuno caratterizzato dalla presenza di loop di collegamento. A livello di LBD si distinguono 2 regioni, il transmitter binding site e la transition zone. Il transmitter binding site, all'interfaccia fra due subunità di cui una obbligatoriamente di tipo  $\alpha$ , è formato da un gruppo di aminoacidi a caratteristiche aromatiche ed idrofobiche ed è delimitato verso la porzione che guarda la membrana da un loop, detto loop C, sede del ponte disolfuro tipico dei recettori della Cys-loop receptor family. La zona di transizione è formata da loop comprendenti, oltre il già citato loop-C, il  $\beta$ 1- $\beta$ 2 linker, il  $\beta$ 8- $\beta$ 9 linker, il  $\beta$ 10-M1 linker ed il M2-M3 linker coinvolti nelle variazioni strutturali che si realizzano a partire dal transmitter binding site verso la porzione transmembrana (dominio M2) e responsabili dell'apertura/chiusura del canale ionico (**Figura 2**).

Il recettore nicotino può esistere in vari stati: lo stato chiuso caratterizzato dall'assenza dell'agonista, lo stato aperto e lo stato desensitizzato nel quale è presente l'agonista ma il canale risulta non attivato.



**Figura 2** A) Recettore nicotinicco da *Torpedo californica*. B) Particolare del LBD.

Globalmente AChBP è caratterizzata da una bassa omologia strutturale con il recettore nicotinicco, ma condivide con esso sia la stessa funzione, ossia lega l'acetilcolina, sia una un'elevata omologia a livello del binding site il che rende l'acetylcholine binding protein un buon template per la costruzione di modelli di omologia del recettore nicotinicco neuronale.<sup>4-8</sup>

A livello di SNC le subunità più abbondanti sono  $\alpha 4$ ,  $\beta 2$  e  $\alpha 7$ ; le subunità  $\alpha 7$  si assemblano a formare recettori omomerici mentre le  $\alpha 4$  e  $\beta 2$  recettori di tipo eteromericco a stechiometria prevalente  $(\alpha 4)_2(\beta 2)_3$ . In questo lavoro di tesi l'attenzione si è focalizzata su quest'ultimo tipo di recettore.

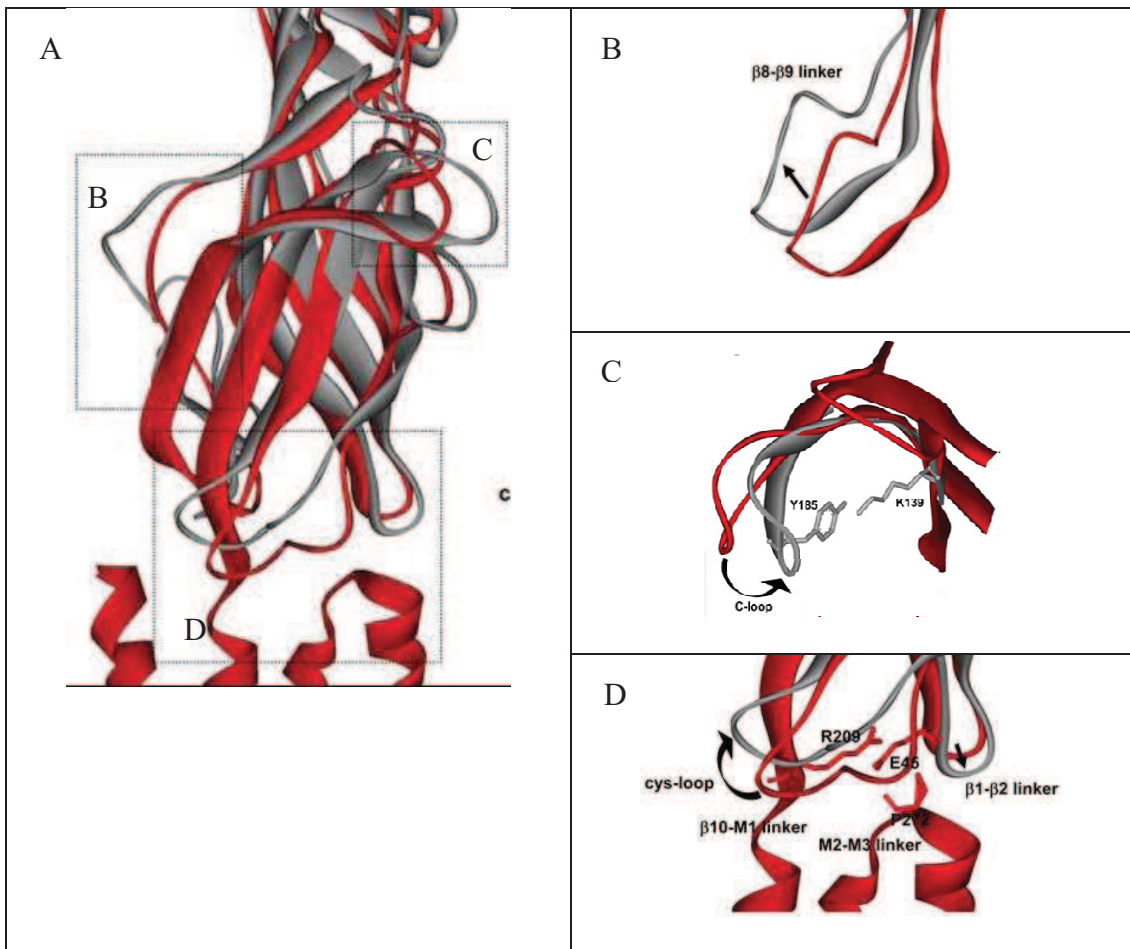
La sovrapposizione strutturale di complessi tanto con agonisti che antagonisti di AChBP presenti nel PDB (**Tabella 1**) ha permesso di distinguere due conformazioni principali del recettore nAChR.

**Tabella 1.** Strutture di AChBP ad oggi disponibili nel PDB.

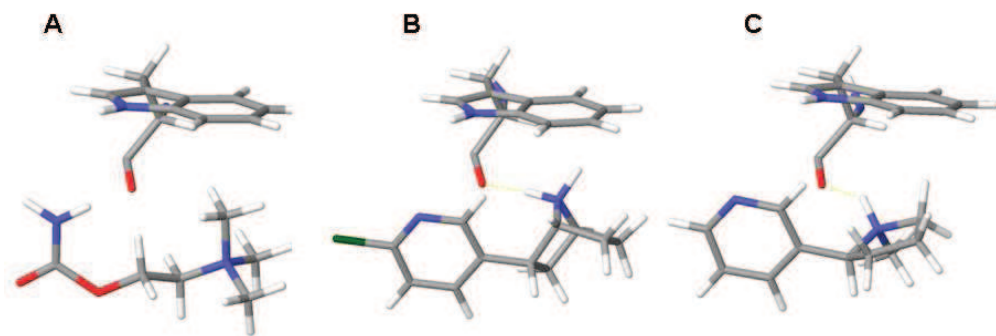
<b>PDB CODE</b>	<b>Titolo</b>	<b>Risoluzione (Å)</b>	<b>Specie</b>
1I9B <sup>2</sup>	AChBP in complesso con HEPS	2,70	<i>Lymnaea stagnalis</i>
1UV6 <sup>9</sup>	AChBP in complesso con Carbamilcolina	2,50	<i>Lymnaea stagnalis</i>
1UW6 <sup>9</sup>	AChBP in complesso con Nicotina	2,20	<i>Lymnaea stagnalis</i>
1UX2 <sup>9</sup>	AChBP in complesso con HEPES	2,20	<i>Lymnaea stagnalis</i>
2BJ0 <sup>10</sup>	AChBP forma apo	2,00	<i>Bulinus Truncatus</i>
2BR7 <sup>11</sup>	AChBP in complesso con HEPES	3,00	<i>Aplysia californica</i>
2BR8 <sup>11</sup>	AChBP in complesso con $\alpha$ -conotossina	2,40	<i>Aplysia californica</i>
2BYN <sup>12</sup>	AChBP forma apo	2,02	<i>Aplysia californica</i>
2BYP <sup>12</sup>	AChBP in complesso con $\alpha$ -conotossina	2,07	<i>Aplysia californica</i>
2BYQ <sup>13</sup>	AChBP in complesso con epibatidina	3,40	<i>Aplysia californica</i>
2BYR <sup>12</sup>	AChBP in complesso con metil-aconitina	2,45	<i>Aplysia californica</i>
2BYS <sup>12</sup>	AChBP in complesso con lobelina	2,05	<i>Aplysia californica</i>
2C9T <sup>14</sup>	AChBP in complesso con $\alpha$ -conotossina	2,25	<i>Aplysia californica</i>
2PGZ <sup>15</sup>	AChBP in complesso con cocaina	1,76	<i>Aplysia californica</i>
2UZ6 <sup>16</sup>	AChBP in complesso con $\alpha$ -conotossina	2,40	<i>Aplysia californica</i>
2ZJU <sup>17</sup>	AChBP in complesso con imidacloprid	2,58	<i>Lymnaea stagnalis</i>
2ZJV <sup>17</sup>	AChBP in complesso con clotianidina	2,70	<i>Lymnaea stagnalis</i>
3C79 <sup>18</sup>	AChBP in complesso con imidacloprid	2,48	<i>Aplysia californica</i>
3C84 <sup>18</sup>	AChBP in complesso con tiacloprid	1,94	<i>Aplysia californica</i>

Nella prima, caratteristica dei complessi con antagonisti, il loop C assume una conformazione aperta corrispondente allo stato chiuso (o resting state) del canale. La seconda configurazione è assunta in presenza di agonisti e vede il loop C in conformazione chiusa, corrispondente con tutta probabilità allo stato aperto o desensitizzato del recettore (**Figura 3**).

L'analisi strutturale dei complessi con AChBP/ligande, insieme ad alcuni studi di site directed mutagenesis<sup>19</sup> hanno consentito di studiare il ruolo svolto da alcuni aminoacidi nell'interazione del nAChR con i ligandi. È stato così possibile evidenziare il coinvolgimento di un residuo di triptofano (Trp143, numerazione secondo 1UV6) nell'interazione con gruppi ammoniacali carichi positivamente, di ligandi come la nicotina, l'epibatidina e l'acetilcolina, con i quali il residuo aminoacidico dà luogo sia ad un'interazione di tipo  $\pi$ -catione sia ad un legame ad idrogeno (**Figura 4**).



**Figura 3** A) nAChR nello stato chiuso in rosso sovrapposto a AChBP nello stato attivo o desensitizzato (in grigio). B) particolare del  $\beta 8$ - $\beta 9$  linker; C) particolare del movimento del loop C a seguito del binding di agonisti; D) particolare dei  $\beta 1$ - $\beta 2$  linker e  $\beta 10$ -M1 linker.



**Figura 4.** Interazioni fra Trp143 e A) Acetilcolina, B) Epibatidina e C) Nicotina.

Il recettore nicotinico regola molte funzioni sia nel sistema nervoso centrale che periferico, come dimostrano i molteplici effetti causati dalla nicotina: alterazioni cognitive, effetti sull'attenzione, sull'apprendimento, sulla

memoria, sull'attività locomotoria, sulla percezione del dolore e sulla dipendenza.<sup>20, 21</sup> Data la varietà di effetti in cui questo recettore è coinvolto, non sorprende che negli anni siano stati intrapresi numerosi studi di molecular modeling<sup>4-7,19,22,23</sup> allo scopo di individuare i determinanti strutturali necessari ad ottenere nuovi e potenti ligandi per questo recettore.

La tipologia di approccio di molecular modeling applicato allo studio del recettore nicotinico ha seguito di pari passo i progressi compiuti dalle tecniche di risoluzione ai raggi X nella determinazione strutturale di questo target. In particolare, è possibile riconoscere un'epoca precedente la risoluzione della struttura di AChBP, durante la quale il disegno e la progettazione di nuovi ligandi era affidata a metodiche di ligand based drug design e allo sviluppo di modelli 3D-QSAR. Questi studi hanno portato a definire alcune proprietà che i ligandi devono possedere per interagire con il recettore. In particolare è necessaria la presenza di: i) un azoto protonato in grado di stabilire un'interazione  $\pi$ -catione e/o un legame ad idrogeno con la controparte recettoriale; ii) un accettore di legame ad idrogeno e/o porzione ricca di elettroni  $\pi$ ; iii) un sistema ricco di elettroni  $\pi$  in grado di interazioni  $\pi$ - $\pi$ ; iv) interazioni steriche, coinvolgenti una porzione a caratteristiche idrofobiche della molecola.

Successivamente alla risoluzione di AChBP la progettazione di ligandi al nAChR si è avvalsa di numerosi modelli costruiti per omologia dei vari sottotipi di nAChR. Questi modelli hanno rappresentato il punto di partenza di numerosi studi di tipo structure-based, attraverso cui è stato possibile investigare le modalità di interazione target ligande. Hanno inoltre consentito di sperimentare l'applicazione di metodologie computazionali per la valutazione dell'energia libera di binding dei ligandi.

## *Glicosilidrolasi*

E' universalmente riconosciuto che i carboidrati rivestono un ruolo molto importante sia nell'immagazzinamento dell'energia, sia nel riconoscimento cellulare.<sup>24</sup> Durante gli ultimi anni, grazie agli sforzi di numerosi gruppi di ricerca, è stato evidenziato il coinvolgimento che questo tipo di molecole rivestono in molti meccanismi biologici e la loro importanza nella cura di malattie, quali il diabete, le infezioni virali, e sperimentalmente nel cancro.<sup>25</sup>

I maggiori successi nell'uso terapeutico dei carboidrati sono stati ottenuti nel campo del diabete di tipo II, nella terapia del quale vengono utilizzate molecole quali l'acarbiosio (Prandase®, Precose®, Glucobay®), e il miglitol (Glyset®), e nella terapia antivirale per la quale sono oggi in commercio molecole quali lo zanamivir (Relenza®) e l'oseltamivir (Tamiflu®), che agiscono come inibitori delle neuroaminidasi.

Inoltre l'interesse sempre crescente per i biocarburanti, sta aumentando l'attenzione per le reazioni di degradazione dei polisaccaridi.<sup>26</sup>

Un ulteriore ambito di ricerca molto attivo è quello che si concentra sullo studio degli enzimi coinvolti nella degradazione degli zuccheri, le glicosidasi o glicosilidrolasi (GH).

Tipicamente una percentuale compresa fra l'1-2 % dell'intero genoma di un organismo codifica per le glicosilidrolasi, ad oggi sono stati annotati migliaia di geni che codificano per queste proteine.

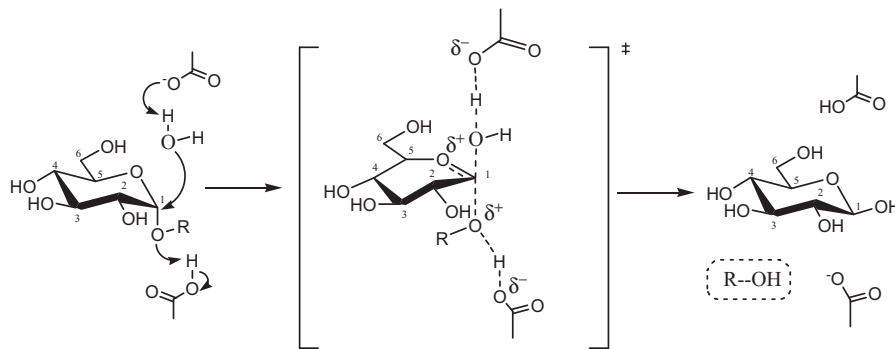
Le glicosilidrolasi sono coinvolte nel meccanismo di degradazione degli zuccheri, catalizzando la reazione di glicolisi attraverso ritenzione o inversione della configurazione sterica al centro anomero.

Nella reazione di idrolisi acido/base sono coinvolti due residui essenziali, che agiscono uno da donatore di protone e l'altro da nucleofilo.

In particolare, gli enzimi che scindono il legame glicosidico con un'inversione di configurazione al centro anomero, agiscono attraverso una singola sostituzione nucleofila (**Figura 5**) in condizioni di catalisi acido/base. Gli enzimi che ritengono la configurazione del centro anomero operano

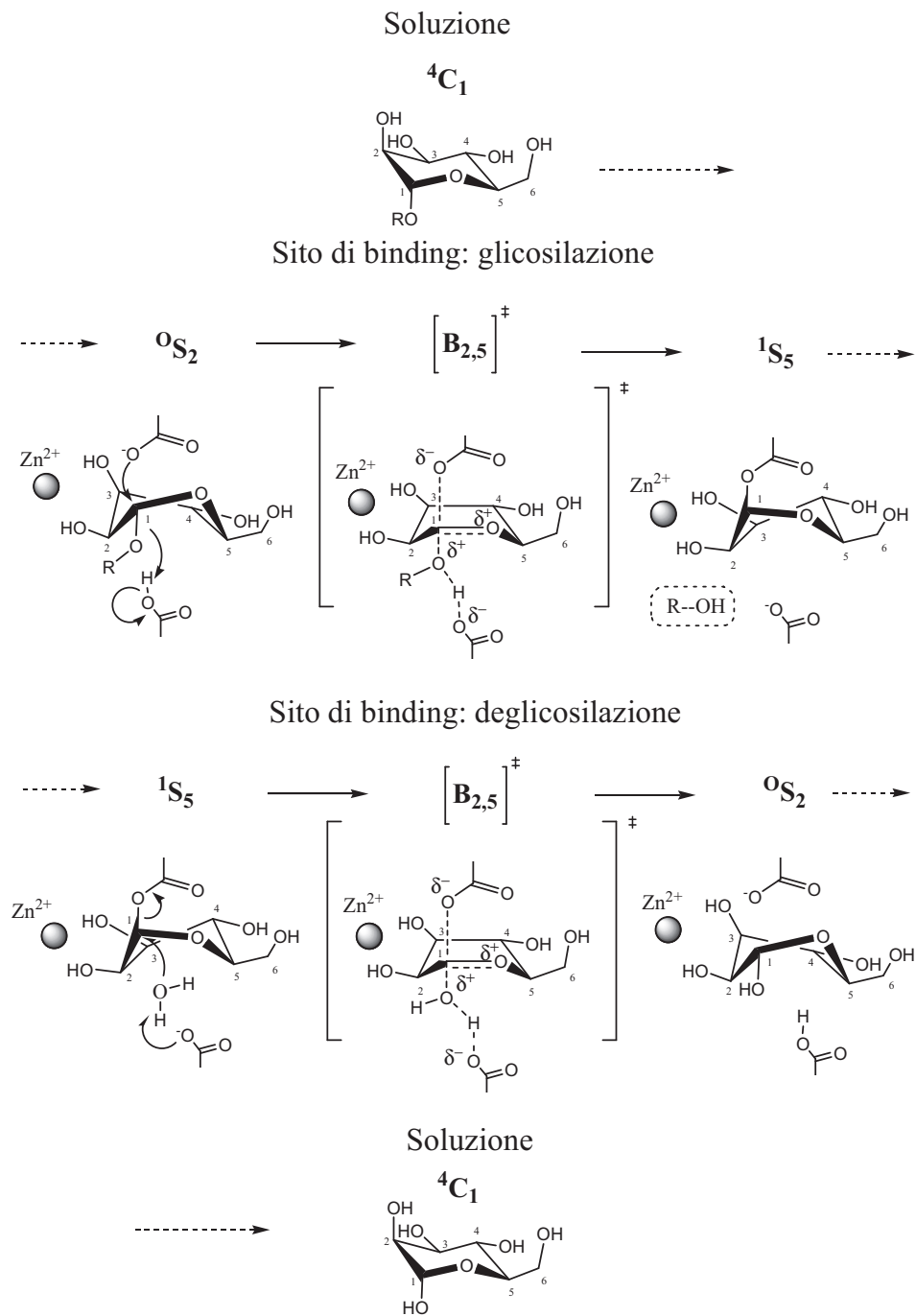
invece una reazione a due stadi, con formazione e successiva scissione di un intermedio covalente (**Figura 6**).

In questo lavoro di tesi, basandosi su approcci computazionale di diverso tipo ed avendo come obiettivo la progettazione di nuovi e più potenti inibitori, abbiamo studiato tre diversi enzimi appartenenti alla famiglia delle glicosilidrolasi: la beta-glicosidasi da *Thermotoga maritima* (famiglia GH1, classificazione del database CAZy), trealasi da *Escherichia coli* (famiglia GH37), la maltasi-glucoamilasi umana (famiglia GH31).



**Figura 5.** Meccanismo di reazione proposto per enzimi che operano con inversione di configurazione. Immagine adattata da 27.





**Figura 6.** Meccanismo di reazione proposto per enzimi che operano con ritenzione di configurazione. Immagine adattata da 27.

**Note:** I simboli che indicano le varie conformazioni d'anello sono assegnate in accordo con le regole di Cremer e Pople; secondo questa nomenclatura le conformazioni a sedia, mezza sedia, barca e barca distorta sono indicate con i simboli **C**, **H**, **B**, **S** e gli atomi che sono sopra o sotto il piano principale dell'anello sono indicati con indici in alto a sinistra o in basso a destra del simbolo, rispettivamente. Così, ad esempio  ${}^{2,5}B$  indica una barca che ha entrambe gli atomi numerati come 2 e 5 sopra al piano dell'anello.

A partire dagli anni '60, periodo nel quale è stato isolato, la noijrimicina (**Figura 7**), il primo inibitore conosciuto delle glicosidasi, sono state sintetizzate e isolate molte molecole con attività inibitoria verso questi enzimi; la maggior parte di queste molecole hanno carattere glicomimetico e sono riconducibili strutturalmente alla classe degli imminozuccheri (**Figura 8**).<sup>28</sup>

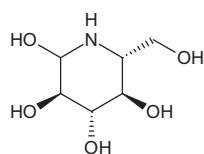
*Beta-Glucosidasi da *Thermotoga maritima* (famiglia GH1)*

La beta-glucosidasi da *Thermotoga maritima* appartiene alla famiglia delle glicosilidrolasi 1 (*TmGH1*) e, come tutti i membri appartenenti a questa famiglia, catalizza la reazione di glicolisi con ritenzione della configurazione al centro anomero. La reazione di glicolisi procede mediante la formazione di un intermedio covalente che viene poi idrolizzato attraverso un intermedio oxacarbenico (**Figura 6**). Nella catalisi sono coinvolti due residui chiave. Il primo (Glu166) fornisce assistenza protonica al gruppo uscente, ossia la catena polisaccaridica, comportandosi poi da base Brønsted nel successivo attacco nucleofilo mediato da una molecola d'acqua per la scissione dell'intermedio covalente. Il secondo residuo (Glu351) è il nucleofilo enzimatico responsabile della formazione dell'intermedio covalente.

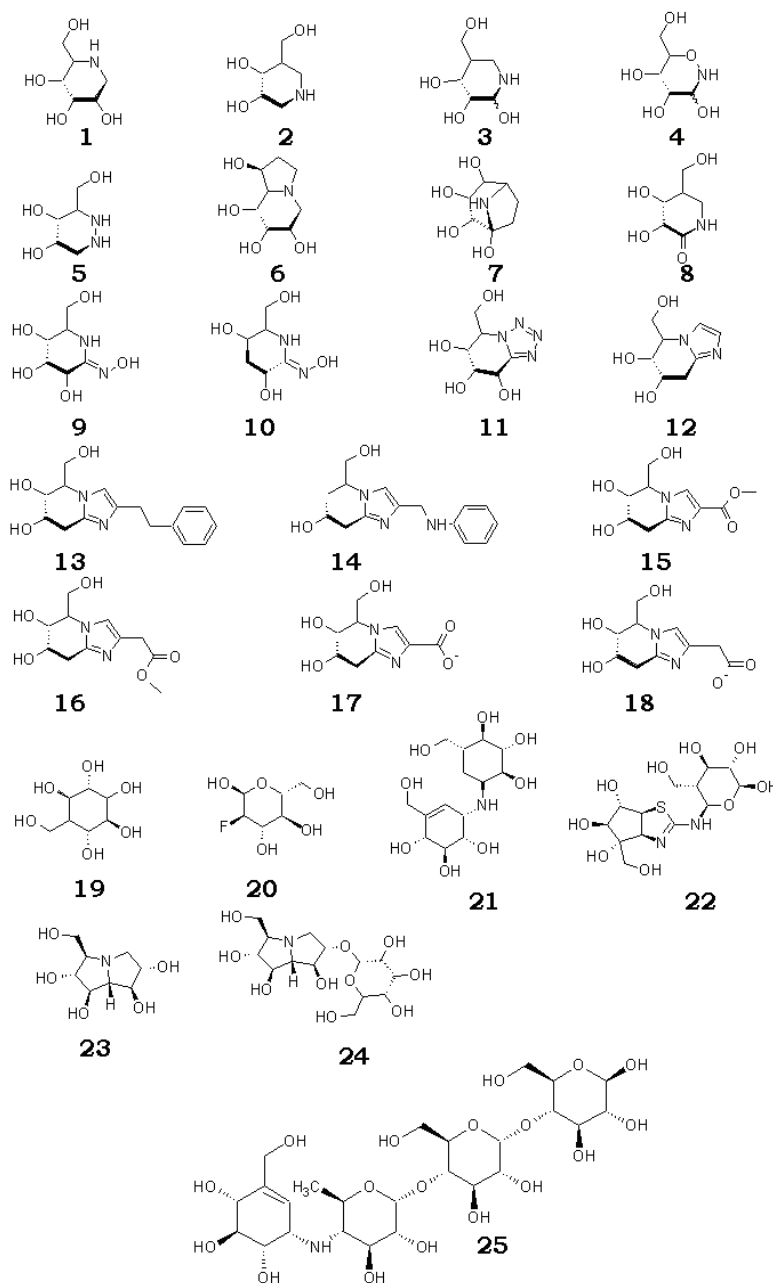
L'efficacia degli inibitori delle glicosidasi è stata associata al fatto che si comportino come analoghi dello stato di transizione oxacarbenico, anche se non sono state ancora chiarite le caratteristiche molecolari che determinano la loro azione.

Dal punto di vista strutturale la beta-glucosidasi è caratterizzata da una sequenza di 446 aminoacidi e da un folding ( $\beta/\alpha$ )<sub>8</sub> (**Figura 9**).

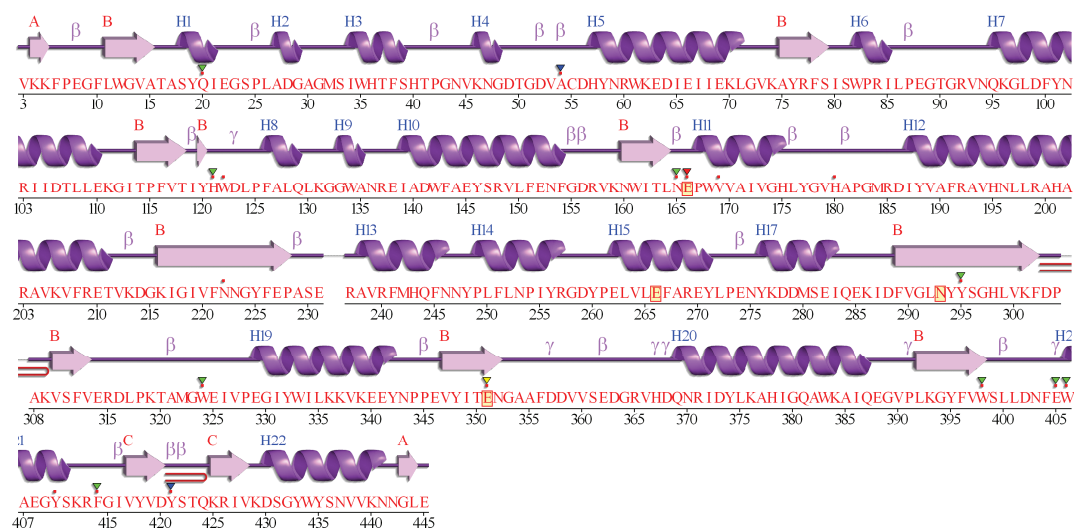
All'interno del PDB sono presenti circa una ventina di complessi tra la beta-glicosidasi e vari inibitori (**Tabella 2**), molti dei quali sono stati oggetto di studi sperimentali volti ad approfondire le forze che governano l'interazione fra i ligandi e l'enzima. Il più dettagliato fra questi studi è senza dubbio quello effettuato da Gloster et al.<sup>29</sup> dove vengono riportati dati strutturali e termodinamici relativi a 18 complessi con inibitori.




**Figura 7** noijrimicina primo inibitore delle glicosidasi isolato negli anni '60.



**Figura 8.** Inibitori delle GH utilizzati in questa tesi: **1** deoxynojirimicina, **2** isofagomina, **3** neuromicina, **4** tetraisossazina, **5** azafagomina, **6** castanospermina, **7** calistegina B<sub>2</sub>, **8** isofagomina lattame, **9** gluco-idroximolattame, **10** galatto-idroximolattame, **11** glucotetrazolo, **12** glucoimidazolo, **13** glucoimidazolo fenil sostituito, **14** glucoimidazolo fenilaminometil sostituito, **15** glucoimidazolo metossicarbonil sostituito, **16** glucoimidazolo metossicarbonilmetil sostituito, **17** glucoimidazolo carbossilato, **18** glucoimidazolo carbossimetil sostituito, **19** ciclofellitolo, **20** deossosfluoroglucosio, **21** validoxylamina A, **22** tiatreazolina, **23** casuarina, **24** casuarina 6- $\alpha$ -glucopiranosio, **25** acarbosio.



Sec. Struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...

HelixStrand

Motifs:  $\beta$  beta turn  $\gamma$  gamma turn  beta hairpin

CSA annotation:  catalytic residue

Residue contacts:  to ligand

Figura 9. Sequenza e struttura secondaria della beta-glucosidasi.

Tabella 2. Complessi inibitori/beta-glicosidasi da *Thermotoga maritima* ad oggi disponibili nel PDB e utilizzati in questo lavoro.

Molecola (a)	PDB CODE	Kd (b)	Risoluzione (Å)	Molecola (a)	PDB CODE	Kd (b)	Risoluzione (Å)
1	2CET <sup>30</sup>	12.9 $\mu$ M	1.97	11	1UZ1 <sup>31</sup>	240 nM	2.00
2	2J7C <sup>29</sup>	51 nM	2.09	12	2J78 <sup>29</sup>	56 nM	1.65
3	2J7E <sup>29</sup>	37 nM	2.19	13	2J7F <sup>29</sup>	9.6 nM	2.28
4	1OIF <sup>32</sup>	541 nM	2.12	14	1W3J <sup>33</sup>	10.8 nM	2.00
5	2CES <sup>30</sup>	65 nM	2.15	15	2J79 <sup>29</sup>	74 nM	1.94
6	2J7H <sup>29</sup>	2.1 $\mu$ M	1.95	16	2CBU <sup>34</sup>	48 nM	1.85
7	2J7D <sup>29</sup>	3.3 $\mu$ M	2.24	17	2CBV <sup>34</sup>	445 nM	1.95
8	2J7G <sup>29</sup>	290 nM	1.91	18	1OIM <sup>32</sup>	100 nM	2.15
					2J77 <sup>32</sup>		2.10
9	2J75 <sup>29</sup>	384 nM	1.85	19	2JAL <sup>35</sup>	ND	1.90
10	2J7B <sup>29</sup>	1.1 $\mu$ M	1.87	20	1OIN <sup>32</sup>	ND	2.15

(a): La numerazione dei composti è in accordo a quella riportata in Figura 8.

(b): Valori di Kd presenti in 29.

ND= dato non disponibile.

Grazie a questa imponente mole di dati sperimentali, la beta-glucosidasi, pur non avendo specifiche implicazioni pratico-terapeutiche, rappresenta un ottimo sistema per testare le performance delle metodologie di molecular modeling.

#### *Trealasi (famiglia GH37)*

A differenza della famiglia GH1, gli enzimi che appartengono alla famiglia GH37 presentano un meccanismo gli colitico con inversione di configurazione. All'interno di questa famiglia, l'unica proteina di cui è nota la struttura 3D è la trealasi, recentemente risolta da Gibson et al.<sup>36</sup>

Questa proteina è costituita da 565 aminoacidi con un folding ( $\alpha/\alpha$ )<sub>6</sub> barrel e da un punto di vista funzionale opera la scissione del trealosio (**Figura 10**).

Il trealosio è usato dagli insetti come riserva energetica per volare, data la facilità della sua trasformazione in glucosio ad opera della trealasi; molecole in grado di inibire quest'enzima hanno, dunque, un ottimo potenziale come insetticidi.

Per quanto riguarda il meccanismo catalitico, è stato ipotizzato che ad agire da base durante la sostituzione nucleofila sia il residuo Glu496, mentre il residuo Asp312 agisce come acido (**Figura 5**).



*Maltasi-Glucoamilasi umana (famiglia GH31)*

La maltasi-glucoamilasi (MGAM), insieme alla  $\alpha$ -glucoamilasi, è uno dei due enzimi responsabili della catalisi dell'ultimo step del processo di rilascio del glucosio nella digestione dell'amido.

MGMA è una proteina di membrana presente nelle cellule intestinali e possiede due domini catalitici omologhi, un N-terminale (NtMGAM) localizzato in prossimità della membrana, e uno C-terminale (CtMGAM) rivolto dalla parte del lumen. Entrambe le subunità operano con un meccanismo catalitico che prevede la ritenzione di configurazione (**Figura 6**) a livello del centro anomero ma l'esiguo numero di studi effettuati sui due domini catalitici, non ha consentito ancora di stabilire se questi posseggano una specificità per diversi substrati.

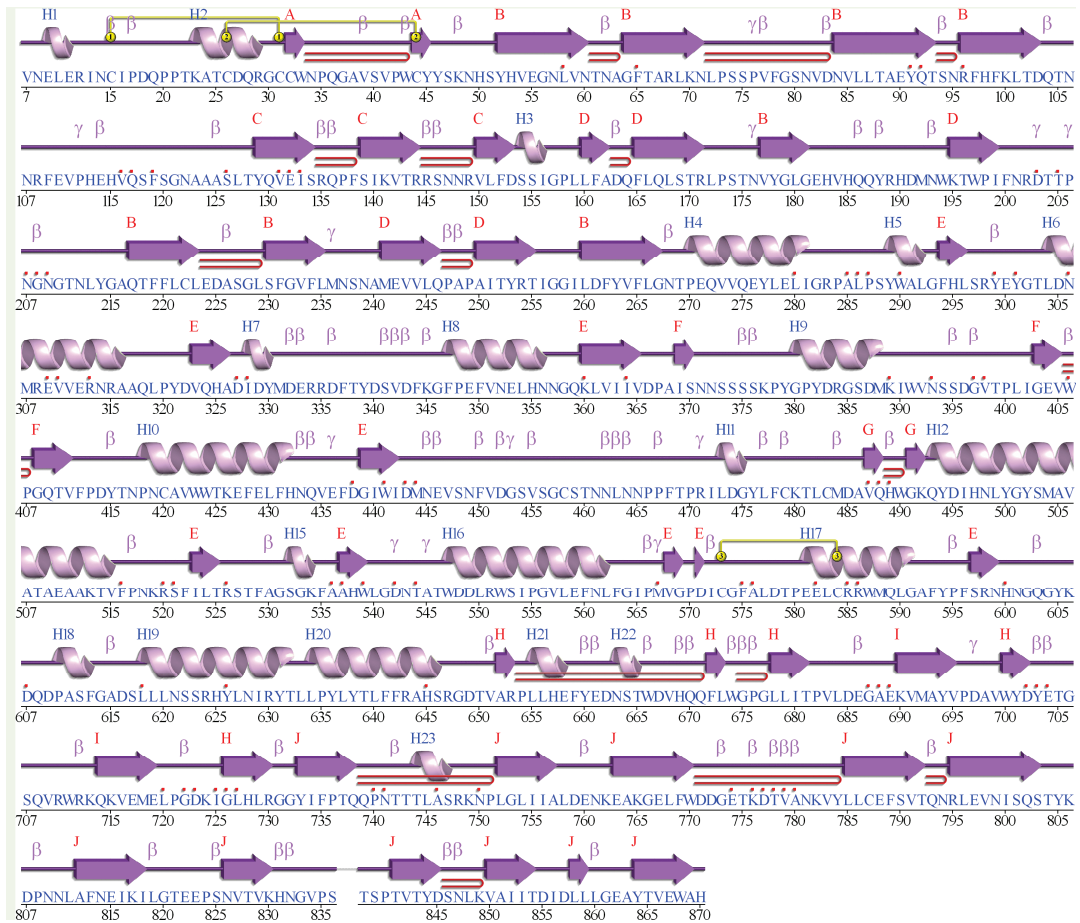
Ad oggi è nota soltanto la struttura del dominio N-terminale, recentemente risolta ai raggi X da Sim e Rose<sup>38</sup>, che ha evidenziato un folding ( $\beta/\alpha$ )<sub>8</sub> (**Figura 11**).

L'interesse allo studio di questo enzima risiede nella possibilità di comprendere meglio il meccanismo digestivo dell'amido, e attraverso questo progettare composti più efficaci per la cura del diabete di tipo II.

In particolare, nel PDB sono state depositate le strutture della forma apo (pdb code 2QLY<sup>38</sup>), e del complesso con l'inibitore l'acarbosio (Figura 8, composto 24) (pdb code 2QMJ<sup>38</sup>), composto già presente nelle preparazioni commerciali (Prandase®, Precose®, Glucobay®).

La struttura del complesso MGAM/casuarina (pdb code 2CTT) non è ancora disponibile nel PDB ma ci è stata fornita nell'ambito di una collaborazione con il prof. Rose dell' Ontario Cancer Research Institute.

Obiettivo comune nello studio dei sopracitati enzimi idrolitici implicati nella degradazione degli zuccheri è la delucidazione dei determinanti molecolari che guidano l'interazione con gli inibitori e l'analisi dell'influenza della






Sec. struc:  Helices labelled H1, H2, ... and strands by their sheets A, B, ...  
 HelixStrand  
 Motifs:  $\beta$  beta turn  $\gamma$  gamma turn  $\beta$  beta hairpin  
 Disulphides:  disulphide bond  
 Residue contacts:  to ligand

Figura 11. Sequenza e struttura secondaria di MGAM.

Tabella 4. Complessi inibitori e MGAM utilizzati in questo lavoro.

Molecola (a)	PDB CODE	Kd(nM)	Risoluzione (Å)	Molecola (a)	PDB CODE	Kd(nM)	Risoluzione (Å)
23	3CTT	0.45		25	2QMJ <sup>38</sup>	6200	1.90
24	ND	280		APO	2QLY <sup>38</sup>		2.00

(a): La numerazione dei composti è in accordo a quella riportata in Figura 8.

ND= dato non disponibile.

flessibilità del target macromolecolare nel processo di riconoscimento dei ligandi.



## *Strumenti e metodi.*

$$S = k \log W$$

Ludwig Eduard Boltzmann (1844 – 1906)

All'interno di questo capitolo verranno riassunte le basi teoriche e metodologiche dei vari approcci computazionali utilizzati in questo lavoro di tesi.

La prima parte del capitolo sarà dedicata a una breve descrizione delle basi della meccanica molecolare (MM), della meccanica quantistica (QM) e dei metodi misti QM/MM. Nella seconda parte verrà dato spazio ad una breve introduzione sulle basi teoriche delle simulazioni di dinamica molecolare (MD) e delle metodologie di docking. Infine, l'ultima parte del capitolo sarà dedicata alla discussione dei metodi utilizzati per la misura dell'affinità ligando/proteina.

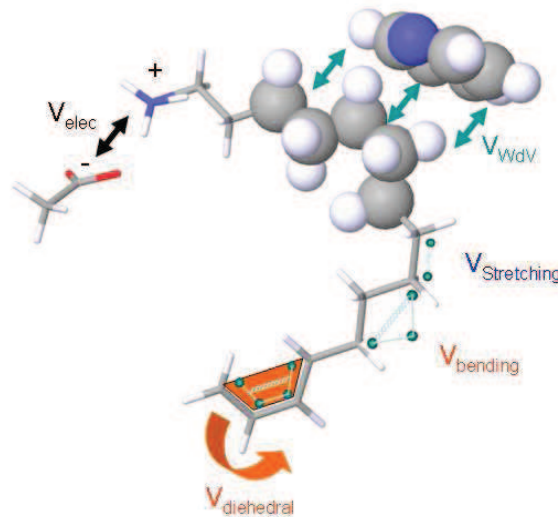
### **Meccanica molecolare**

Con il termine di meccanica molecolare vengono indicate tutte quelle metodologie che utilizzano le leggi della fisica classica (nella formulazione di Newton o di Lagrange) per descrivere il comportamento di piccole molecole e macromolecole biologiche.

I metodi di meccanica molecolare si basano sull'approssimazione di Born-Oppenheimer la quale consente di esprimere l'energia del sistema molecolare come funzione delle coordinate dei soli nuclei. Questi metodi si basano su modelli di interazione fra sistemi molecolari piuttosto semplici che consistono nel trattare le molecole come composti da masse (atomi) e molle (legami) descritte da movimenti consistenti in variazioni delle lunghezze di legame (stretching), apertura e chiusura degli angoli (bending), rotazioni e torsioni di legami.

Parte fondamentale dei calcoli di meccanica molecolare è il force field (FF). Il FF identifica sia l'equazione matematica che descrive le caratteristiche energetiche della molecola e l'insieme dei parametri utili a descrivere i vari tipi di atomi (atom type) costituenti la molecola, i legami che uniscono gli atomi e gli angoli formati dagli stessi.

La forma funzionale FF assume che l'energia potenziale di una molecola, anche detta energia sterica totale, possa essere espressa come somma di contributi dovuti ad interazioni di tipo intra e inter molecolare nel contesto del sistema di interesse (**Figura 1**).



**Figura 1.** Rappresentazione grafica dei termini del potenziale presenti in vari force field

Da un punto di vista generale la forma funzionale di un FF può essere espressa dall'**equazione 1** fermo restando che alcuni FF possono includere termini addizionali necessari, ad esempio, per modellare esplicitamente legami ad idrogeno.

$$\begin{aligned}
 V(r_1, r_2, \dots, r_N) = & \sum_{\text{legami}} \frac{1}{2} K_b [b - b_0]^2 + \sum_{\text{angoli}} \frac{1}{2} K_\theta [\theta - \theta_0]^2 + \sum \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \\
 & \sum_{\substack{\text{diedri} \\ \text{impropri}}} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2 + \sum_{\text{coppie}(i,j)} [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 + q_i q_j / (4\pi\epsilon_0 \epsilon_r r_{ij})]
 \end{aligned}
 \tag{eq. 1}$$

- $V(r_1, r_2, \dots, r_N)$  rappresenta l'energia potenziale del sistema. È funzione solo della posizione ( $r$ ) degli atomi ( $N$ ) che compongono il sistema; non tiene conto dei moti elettronici.
- $\sum_{\text{legami}} \frac{1}{2} K_b [b - b_0]^2$  modella le interazioni fra coppie di atomi legati e descrive la variazione dell'energia in funzione della deviazione della

lunghezza del legame rispetto al suo valore di riferimento  $b_0$ . L'espressione è una funzione di tipo armonico riconducibile alla legge di Hook.  $K_b$  rappresenta il parametro che determina la costante di deformazione lineare delle molla che simula il legame interatomico. Tutti i FF possiedono per ciascun tipo di legame specifici valori di  $K_b$  e  $b_0$  a seconda della natura del legame e degli atomi che lo formano.

- $\sum_{\text{angoli}} \frac{1}{2} K_\theta [\theta - \theta_0]^2$  descrive la deviazione degli angoli di legame dal loro valore di equilibrio  $\theta_0$  ed è espressa in maniera analoga a quanto visto per i legami. Anche in questo caso la variazione dell'energia associata alle deviazioni degli angoli di legame  $\theta$  rispetto al loro valore di equilibrio viene descritta tramite la legge di Hook o legge del potenziale armonico.
- $\sum \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)]$  è un termine che si riferisce alla variazione dell'energia associata alle torsioni. Questa variazione è descritta per mezzo di una funzione periodica nella quale  $n$  è il parametro che determina la periodicità ed è legato alla simmetria della molecola. Definisce il numero di minimi dell'energia potenziale  $V$  nella rotazione fra  $0$  e  $2\pi$ .  $\varphi$  rappresenta l'angolo di rotazione e  $\gamma$  il fattore di fase.
- $\sum_{\substack{\text{diedri} \\ \text{impropri}}} \frac{1}{2} K_\zeta [\zeta - \zeta_0]^2$  è un termine addizionale che rende conto delle interazioni fuori dal piano dei così detti termini di secondo ordine, e che quindi viene usato per mantenere la planarità di un sistema.

Questi primi quattro termini dell'espressione funzionale del FF si riferiscono al contributo dato dalle interazioni di legame all'energia potenziale. Queste interazioni che dipendono dalla connettività.

La maggior parte delle variazioni nella struttura e nell'energia dei sistemi molecolari sono dovute ai termini torsionali e di non legame, dovuti cioè ad

interazioni che non dipendono dalla connettività fra gli atomi interagenti. Sono interazioni che si realizzano “attraverso lo spazio” e sono di solito modellate come funzioni di una qualche potenza inversa della le interazioni carica-carica variano con  $1/r$ , carica-dipolo= con  $1/r^2$ , dipolo-dipolo con  $1/r^3$ , dipolo-quadrupolo con  $1/r^4$ , carica-dipolo indotto con  $1/r^4$ , dipolo-dipolo indotto con  $1/r^6$ .

I termini di non legame di un FF vengono ricondotti a due gruppi principali: le interazioni elettrostatiche e le interazioni di van der Waals (vdW), rappresentate nell' **equazione 1** dal termine:

$$\sum_{\text{coppie}(i,j)} \left[ C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 + q_i q_j / (4\pi\epsilon_0 \epsilon_r r_{ij}) \right]$$

LJ 6-12

L'energia di non legame è ottenuta come somma di potenziali fra coppie di atomi  $i$  e  $j$  e dipende in maniera inversa dalla distanza  $r_{ij}$  che separa gli atomi.

La componente di vdW delle interazioni di non legame è generalmente descritta con una funzione di Lennard Jones 6-12 (LJ 6-12). Si compone di due termini  $C_6(i,j)/r_{ij}^6$  rappresenta il contributo attrattivo dovuto alle forze di dispersione di London originate da dipoli istantanei che si formano durante le fluttuazioni della nube elettronica. Questa componente è prevalente a grandi distanze e decresce con  $r^6$ .

$C_{12}(i,j)/r_{ij}^{12}$  descrive la componente di tipo repulsivo dovuta a forze di scambio che si instaurano a breve distanza fra i nuclei e decade secondo la dodicesima potenza della distanza  $r$ .

Il terzo termine si riferisce alla componente elettrostatica dell'interazione che viene trattata con la legge di Coulomb, assegnando cariche parziali centrate sui nuclei, ed assumendo che la densità di carica intorno agli atomi abbia simmetria sferica.

Nel caso del legame ad idrogeno alcuni FF usano al posto del potenziale di LJ 6-12 una funzione di LJ 10-12 al fine di migliorare l'accuratezza con cui viene predetta la geometria del legame ad H. Altri force field più sofisticati, prendono in considerazione la distanza dell'idrogeno dall'accettore, l'angolo formato da donatore, idrogeno e accettore (DHA) e la deviazione del legame dalla direzione del più vicino lone pair dell'accettore.

Tutti i FF sono empirici e sono progettati per riprodurre le proprietà strutturali di set di composti usati per svilupparli.

Un'importante proprietà che un FF dovrebbe possedere è la trasferibilità della forma funzionale e dei parametri, intendendo per trasferibilità la possibilità di applicare con affidabilità e precisione i parametri derivati da un set su un più ampia generalità di molecole.

Esiste un'ampia varietà di FF che si differenzia per la forma funzionale e/o per i parametri usati. La tabella 1 riporta alcuni fra i più comuni FF ad oggi usati.

**Tabella 1.** Alcuni fra i force field più utilizzati.

<i>Force Field</i>	<i>Descrizione</i>
AMBER (Assisted Model Building and Energy Refinement)	Largamente usato per proteine e DNA
CHARMM (Chemistry at HARvard Molecular Mechanics).	Largamente usato per piccole molecole e macromolecole biologiche.
CVFF	Usato per piccole molecole e macromolecole biologiche
Gromos	Force field per nucleotidi, proteine e zuccheri
Gaff (General Amber Force Field).	Insieme di parametri per piccole molecole organiche.
OPLS (Optimized Potential for Liquid Simulations).	Largamente usato per piccole molecole e macromolecole biologiche, e sviluppato per riprodurre proprietà di sostanze organiche liquide.
CFF	Famiglia di force field per molecole organiche.
MMFF	Sviluppato nei laboratori merk per lo studio di molecole organiche.

I FF utilizzati in questo lavoro di tesi sono stati:

- per gli studi di docking, OPLS2001<sup>39</sup> che rappresenta una versione ottimizzata del FF OPLS<sup>40</sup>;
- per ciò che riguarda gli studi di dinamica molecolare (MD) Amber03, evoluzione del più noto FF Amber<sup>41</sup>, e in particolare per la porzione proteica del sistema in studio. Per ciò che riguarda i ligandi sono stati invece utilizzati i parametri del GeneralAmberForceField<sup>42</sup> (GAFF) per la valutazione di tutte le componenti del potenziale, ad eccezione delle cariche atomiche attribuite mediante la metodologia RESP (Restrained Electrostatic Potential), basata sull'uso del potenziale elettrostatico ( $\epsilon_p$ );

Il potenziale elettrostatico rappresenta l'energia di interazione fra una carica positiva puntiforme, posizionata in un punto dello spazio e i nuclei e gli elettroni di una molecola.

L'attribuzione delle cariche viene realizzata attraverso una sequenza di step:

1. posizionamento della molecola, considerata in termini QM come somma di nuclei ed elettroni, in una griglia tridimensionale;
2. determinazione del potenziale elettrostatico in ogni punto della griglia, risolvendo le funzioni d'onda esatte della meccanica quantistica, per mezzo di un metodo QM consistente con quello utilizzato per le altre cariche contenute nel FF;
3. sostituzione della rappresentazione della molecola in termini di orbitali molecolari (nuclei e densità elettronica) con la descrizione in termini di coordinate nucleari. Si associa quindi in maniera iterativa una carica ad ognuno di questi atomi e si calcola un potenziale elettrostatico approssimato  $\epsilon_p^{approx}$  per ogni nodo della griglia. Applicando il metodo dei minimi quadrati si individua quel set di cariche atomiche per le quali è minima la somma dei

quadrati delle differenze fra il potenziale approssimato e quello calcolato con il metodo di Hartree Fock (HF);

4. assimilazione a 0 del valore di carica per gli atomi di carbonio più interni

I problemi associati alla determinazione delle cariche attraverso il metodo RESP riguardano l'approssimazione insita nel calcolo, l'influenza della conformazione sulla scelta delle cariche, la scarsa riproducibilità delle cariche assegnate.

### Metodi basati sulla funzione d'onda

Fanno parte dei metodi basati sulla funzione d'onda tutte quelle metodologie che utilizzano una descrizione quanto meccanica (QM) delle molecole. Queste metodologie dette *ab-initio*, consentono di ottenere la risoluzione dell'equazione di Schrodinger.

I metodi QM *ab-initio* dal latino *dall'inizio*, non usano parametri empirici e permettono un'accurata descrizione delle molecole; diversamente dalla meccanica molecolare, forniscono informazioni su struttura elettronica della molecola, distribuzione di carica, potenziale elettrostatico, polarizzabilità e reattività molecolare.

#### Metodo di Hartree Fock

Il più noto metodo QM è il metodo di Hartree Fock (HF). È un metodo variazionale la cui applicazione porta quindi ad energie sempre maggiori o uguali a quella effettivamente posseduta dal sistema.

Il metodo HF può essere considerato un'approssimazione a singolo elettrone applicata a sistemi multielettronici, nella quale l'Hamiltoniano molecolare è suddiviso in Hamiltoniani che descrivono l'energia di un singolo elettrone.

Considerando un sistema molecolare di N elettroni, ognuno caratterizzato dal grado di libertà  $r_i$ , la funzione d'onda è data dal prodotto di Hartree

$$\psi_h(r_1, r_2, \dots, r_n) = \phi(r_1) \cdot \phi(r_2) \cdot \dots \cdot \phi(r_n) \quad (\text{eq. 2})$$

che consente di scrivere l'Hamiltoniano per questo tipo di sistema come:



$$H_e = T_e + V_{ne} + V_{ee} + V_{nn} \quad (\text{eq. 3})$$

dove:

- $T_e = \sum_{i=1}^n \frac{-\nabla_i^2}{2}$  (energia cinetica elettronica)
- $V_{ne} = \sum_i^n \sum_A^N \frac{-Z_A}{r_{iA}}$  (interazione elettroni nuclei)
- $V_{ee} = \sum_i^n \sum_j^n \frac{1}{r_{ij}}$  (interazione elettrone elettrone)
- $V_{nn} = \sum_A^N \sum_{B>A}^N \frac{Z_A Z_B}{R_{AB}}$  (interazione nucleo nucleo)

A e B rappresentano i nuclei,  $i$  e  $j$  gli elettroni,  $Z$  la carica nucleare,  $T$  l'operatore energia cinetica e  $V$  l'operatore energia potenziale.

L'Hamiltoniano può quindi essere riscritto come combinazione di questi termini:

$$H = \sum_{i=1}^n \frac{-\nabla_i^2}{2} + \sum_i^n \sum_A^N \frac{-Z_A}{r_{iA}} + \sum_i^n \sum_j^n \frac{1}{r_{ij}} + \sum_A^N \sum_{B>A}^N \frac{Z_A Z_B}{R_{AB}} \quad (\text{eq. 4})$$

dove  $V_{nn}$ , nell'ambito dell'approssimazione di Born-Oppenheimer, è indipendente dalle coordinate elettroniche,  $T_e$  e  $V_{ee}$  dipendono dalle coordinate elettroniche mentre il termine  $V_{ee}$  dipende da  $n(n-1)/2$  coordinate elettroniche.

Caratteristica fondamentale del metodo di HF, è quella di ridurre la complessità del calcolo di questo Hamiltoniano assumendo che i termini  $V_{ee}$  e  $V_{en}$  siano dovuti all'interazione di ogni singolo elettrone con una nube media formata dagli altri elettroni. In questo modo un complesso problema multidimensionale viene scomposto in tanti più piccoli problemi monodimensionali.

Questo tipo di approssimazione, sebbene sia alla base di uno dei più utilizzati metodi computazionali, può condurre ad errori significativi nella stima dell'energia di una molecola.

*Density Functional Theory (DFT)*

Un altro metodo QM che sta acquistando sempre più importanza è la Density Functional Theory (DFT), metodo rigoroso e computazionalmente molto efficiente per lo studio dello stato fondamentale delle molecole.

L'idea fondamentale alla base dei metodi DFT, che è valsa il premio nobel per la chimica 1998 a Walter Kohn e John A. Pople, è che lo stato fondamentale di un sistema quantico di  $n$  particelle può essere descritto in funzione della sua densità elettronica  $\rho(\mathbf{r})$ .

La prima vantaggiosa implicazione di questa impostazione è che la densità è un osservabile e come tale può essere misurata e visualizzata. Inoltre, essendo la densità funzione dello spazio, la dimensionalità del problema è ridotta da  $3N$  a 3. Quest'ultima considerazione fa sì che si possano ottenere, in tempi contenuti, risultati di affidabilità pari a quelli ottenibili con metodi computazionalmente più dispendiosi.

La densità elettronica è funzione dell'equazione d'onda e come tale si ottiene dal suo quadrato  $|\psi|^2$ ; dal momento che dal punto di vista matematico una funzione di funzione  $F(f)$  viene anche denominata funzionale, e che la teoria di Pople e Kohn esprime l'energia di una molecola in funzione della sua densità elettronica, a sua volta funzione dell'equazione d'onda, tale teoria viene anche indicata con Density Functional Theory (DFT). La veridicità di questo funzionale è contenuta nel primo teorema di Hohenberg e Kohn (1964) che dimostra, utilizzando una *reductio ad absurdum*, che una densità elettronica specifica univocamente uno ed un solo stato fondamentale di una molecola.

Purtroppo ad oggi non è stato ancora trovata una formulazione che espliciti efficacemente la funzione che correla energia e densità elettronica.

Nel 1965 Kohn e Sham proposero un metodo per minimizzare l'energia in funzione della densità elettronica combinando quest'ultima con la funzione d'onda.

Come punto di partenza essi ripartirono il funzionale energia secondo la seguente equazione:

$$E[\rho] = T_0[\rho] + \int [V_{ext}(r) + U_{el}(r)] \rho(r) dr + E_{ex}[\rho] \quad (\text{eq. 5})$$

dove  $T_0$  è il termine che descrive l'energia cinetica degli elettroni in un sistema che ha la stessa densità del sistema reale ma nel quale non sono presenti interazioni fra elettroni.

$U_{el}$  è il termine descritto dalla legge di Coulomb per l'interazione fra elettroni. Esso include esplicitamente l'interazione elettrone-elettrone visto che la corrispondente energia è:

$$U_{el}[\rho] = \int \int \frac{\rho(r')\rho(r)}{|r'-r|} dr dr' \quad (\text{eq. 6})$$

$V_{ext}$  è il potenziale esterno, ossia il potenziale esercitato dai nuclei:

$$V_{ext} = \sum_a \frac{-Z_a}{|R_a - r|}$$

L'ultimo termine,  $E_{ex}[\rho]$ , denominato exchange-correlation energy, tiene conto di tutti i contributi non considerati nei precedenti termini, ossia dello scambio elettronico, della correlazione elettronica, della correzione all'energia cinetica necessaria per avere l'energia cinetica in un sistema reale, ed infine delle correzioni per self-interaction.

Nel corso degli anni sono state elaborate molte diverse formulazioni di questo funzionale. Infatti non essendo nota la sua forma esatta è stato necessario ricorrere ad approssimazioni riguardo all'energia cinetica, all'energia di scambio e all'energia di correlazione. In una delle più semplici, la local density approximation (LDA), il funzionale dipende solamente dalla densità elettronica ai valori delle coordinate in cui il funzionale viene valutato.

La local spin density approximation (LSDA) è un altro tipo di approssimazione, generalizzazione diretta della LDA con incluso lo spin elettronico.

Le approssimazioni di gradiente generalizzato (generalized gradient approximations, GGA) sono ancora approssimazioni di tipo locale e tengono anche conto del gradiente della densità elettronica nelle medesime coordinate. Fa parte di quest'ultima classe di approssimazioni il funzionale B3LYP, ad oggi il più utilizzato per lo studio di piccole molecole organiche.

La valutazione delle performance di un funzionale è sostanzialmente empirica, motivo per cui molti autori tendono a classificare la DFT fra i metodi semiempirici.

Comunque assumendo che questo funzionale descriva sufficientemente bene l'energia del sistema, è possibile derivare l'equazione (eq. 5) di Kohn e Sham usando il principio variazionale, ed usare poi questa equazione per calcolare l'energia del sistema mediante un sistema iterativo.

#### *Metodi QM/MM*

I metodi misti quantomeccanica/meccanica molecolare (QM/MM) accoppiano metodologie di meccanica quantistica (QM) e meccanica molecolare (MM) nello studio di un sistema di interesse.

L'idea di base delle metodologie QM/MM consiste nel trattare solo una ristretta porzione del sistema a livello di teoria QM, mentre tutto il resto del sistema è descritto a livello MM.<sup>43</sup>

Per un sistema studiato con metodologia QM/MM sarà quindi possibile scrivere un Hamiltoniano del tipo:

$$H_{eff} = H_{MM} + H_{QM} + H_{QM-MM} \quad (\text{eq. 7})$$

dove  $H_{MM}$  esprime l'energia della porzione MM,  $H_{QM}$  esprime l'energia della porzione QM mentre  $H_{QM/MM}$  descrive l'energia relativa alle porzioni QM e MM a più stretto contatto ed interagenti fra loro.

Il punto più delicato delle metodologie QM/MM sta nella trattazione della così detta boundary region, ossia la regione di confine fra le aree del sistema

studiate a livello QM e quelle trattate a livello MM, e nel calcolo delle interazioni fra queste due porzioni.

I vari schemi per il calcolo QM/MM possono essere ricondotti a due gruppi principali, dipendentemente dal tipo di trattamento dell'interazione elettrostatica tra le due porzioni del sistema

- accoppiamento meccanico (mechanical embedding)
- accoppiamento elettrostatico (electrostatic embedding)

Nell'accoppiamento meccanico il calcolo della porzione QM viene eseguito ignorando la porzione MM mentre le interazioni fra le due porzioni vengono trattate a livello MM includendo cioè interazioni di legame (stretching), angolari (bending) e di non legame, (elettrostatiche e di van der Waals).

Nel caso dell'accoppiamento elettrostatico il calcolo della porzione QM è effettuato in presenza della porzione MM.

Visto che molti force field disponibili in letteratura come comprendono un grande set di cariche parziali centrate sugli atomi, è solitamente conveniente considerare le interazioni fra la parte QM e quella MM come un'interazione di carica all'interno dell'Hamiltoniano QM, mentre gli altri tipi di interazione vengono trattati a livello MM.

La differenza fra questi due approcci risiede principalmente nel trattamento della componente elettrostatica dell'interazione.

L'accoppiamento meccanico è senza dubbio più semplice di quello elettrostatico, ma soffre di alcuni inconvenienti. È innanzitutto necessario disporre di parametri di tipo MM per ambedue le porzioni QM ed MM, del sistema, si assume inoltre che la distribuzione di carica nella porzione QM non dipenda da quella nella porzione MM, il che può essere fonte di errore.

Diversamente dall'accoppiamento meccanico, l'accoppiamento elettrostatico non richiede parametri MM per la porzione QM, visto che l'interazione elettrostatica fra le due parti è inclusa come termine dell'Hamiltoniano QM.

In entrambi gli approcci le interazioni di legame e le interazioni di van der Waals fra le due porzioni sono calcolate a livello MM.

Nel caso in cui le due parti del sistema non interagiscano covalentemente l'implementazione di questa metodologia è abbastanza semplice; infatti, l'energia di interazione fra le due porzioni QM e MM viene calcolata classicamente, ossia utilizzando il potenziale di Lennard-Jones per le interazioni di vdW e la legge di Coulomb per le interazioni elettrostatiche.

Qualora le due porzioni QM e MM siano legate covalentemente, è necessario utilizzare accorgimenti che permettano di saturare la valenza dei due atomi coinvolti nel legame. Questa eventualità rappresenta una delle maggiori sfide per gli sviluppatori di sistemi QM/MM i quali durante gli anni hanno proposto numerose alternative fra le quali alcune basate sull'utilizzo di link atom e altre su local orbital.

Nei metodi del primo gruppo, ossia quelli basati sui link atom, un atomo "fittizio" viene utilizzato per saturare la valenza degli atomi coinvolti in legami che è necessario tagliare perché a ponte fra le due porzioni QM e MM del sistema; di solito a questo scopo vengono utilizzati atomi di idrogeno o dummy atom specificamente parametrizzati.

I metodi appartenenti al gruppo degli orbitali localizzati, descrivono il legame a ponte usando orbitali strettamente localizzati, derivati da calcoli su piccoli composti modello che, come accade per i parametri di un force field, si assume essere trasferibili.<sup>43</sup>

### *Dinamica Molecolare (MD)-Note generali*

La dinamica molecolare (MD) è una metodologia computazionale che permette di studiare, tramite la risoluzione delle equazioni del moto, l'evoluzione temporale di un sistema poliatomico.

Il risultato degli esperimenti di MD è una traiettoria, cioè una serie di conformazioni assunte in successione temporale dal sistema, che specifica la variazione della posizione e della velocità di tutte le particelle che lo compongono.

Gli enormi sviluppi avvenuti al giorno d'oggi nel campo delle strumentazioni hardware hanno consentito di simulare, per tempi sufficientemente estesi, i movimenti a cui vanno incontro i grandi sistemi biologici, così da ottenere dati confrontabili con le quantità misurabili sperimentalmente.

L'intervallo temporale di accadimento dei fenomeni biofisici, ossia la loro scala dei tempi, varia dai femto ai microsecondi (**Tabella 2**). E' perciò fondamentale poter impostare simulazioni di durata sufficiente a coprire la scala dei tempi in cui avviene il fenomeno a cui siamo interessati.

Alternativamente possono essere utilizzati metodi come la metadinamica<sup>44</sup>, il replica exchange e l'umbrella sampling in grado di accelerare fenomeni rari e renderli investigabili con simulazioni di minor durata.<sup>45</sup>

**Tabella 2.** Scala dei tempi associate ai principali processi biofisici.

<b>Tipologie di movimenti</b>	<b>Esempi funzionali</b>	<b>Tempi; scala delle ampiezze</b>
<p><i><b>Movimenti locali</b></i></p> <p>Fluttuazione degli atomi</p> <p>Movimenti delle catene laterali</p>	<p>Flessibilità del docking di ligandi;</p> <p>movimenti di diffusione nel tempo.</p>	<p>Da fs a ps (<math>10^{-15}</math> - <math>10^{-12}</math> s);</p> <p>meno di 1 Å.</p>
<p><i><b>Movimenti su media scala</b></i></p> <p>Movimenti dei loop</p> <p>Movimenti delle porzioni terminali</p> <p>Movimento a corpo rigido</p>	<p>Adattamento conformazionale del sito attivo;</p> <p>specificità di binding.</p>	<p>Da ns a <math>\mu</math>s (<math>10^{-9}</math> - <math>10^{-6}</math> s);</p> <p>1-5 Å.</p>
<p><i><b>Movimenti su larga scala</b></i></p> <p>Movimenti dei domini</p> <p>Movimenti delle subunità</p>	<p>Movimenti intra-dominio;</p> <p>transizioni allosteriche.</p>	<p>Da <math>\mu</math>s a ms (<math>10^{-6}</math> - <math>10^{-3}</math> s);</p> <p>5-10 Å.</p>
<p><i><b>Movimenti globali</b></i></p> <p>Transizione elica-coil</p> <p>Folding/unfolding</p> <p>Associazione di subunità</p>	<p>Attivazione ormonale;</p> <p>funzionalità proteica.</p>	<p>Da ms a h (<math>10^{-3}</math> - <math>10^{-4}</math> s);</p> <p>più di 10 Å.</p>

I fenomeni biofisici interessanti nel processo di Drug Design, ossia le variazioni conformazionali del ligande e del sito di binding, si collocano in una scala temporale che oscilla fra i femto e i nanosecondi, ed è quindi in questa scala che si collocano tutte le simulazioni di dinamica molecolare effettuate in questo lavoro di tesi.

Nelle simulazioni di dinamica molecolare, il moto delle particelle che costituiscono il sistema segue le leggi della meccanica classica. Questo tipo di approccio, come tutti quelli basati sulla meccanica molecolare, parte dall'approssimazione di Born-Oppenheimer (B-O), secondo la quale è possibile separare il moto dei nuclei da quello degli elettroni, in considerazione del fatto che la massa degli elettroni è molto più piccola di quella dei nuclei. In quest'ottica, dunque, gli elettroni possono rapidamente adattarsi ad ogni cambiamento della posizione dei nuclei ed è quindi possibile calcolare l'energia di una molecola, o macromolecola, come funzione delle sole coordinate nucleari, trascurando il moto degli elettroni, considerati nel loro stato fondamentale.

In MD, l'evoluzione temporale del sistema viene rappresentata mediante una traiettoria, cioè una serie di singole conformazioni campionate in stretti intervalli di tempo (time step).

La traiettoria viene ottenuta risolvendo la seconda legge di Newton ( $F = ma$ ). Questa ha carattere vettoriale e può essere scomposta applicandola a ognuna delle particelle componenti il sistema esprimendo l'accelerazione come derivata seconda della posizione rispetto al tempo.

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad (\text{eq. 8})$$

Secondo questa equazione una particella di massa  $i$  si muove lungo la coordinata  $x_i$  sotto l'azione della forza  $F_{x_i}$ , che è il risultato dell'interazione dell'atomo  $i$  con tutte le altre particelle del sistema. Questa forza può essere espressa come la derivata negativa di una funzione potenziale  $V(r_1, r_2, r_3, \dots)$



$r_n$ ), espressione dell' energia potenziale del sistema espressa in funzione delle posizioni dei vari atomi:

$$F_i = -\frac{\partial V}{\partial r_i} \quad (\text{eq. 9})$$

Le coordinate e le velocità di tutti gli atomi (traiettoria) vengono ottenute per integrazione dell'equazione di Newton, risolta simultaneamente per tutti gli atomi, in un piccolo intervallo di tempo  $dt$  (time step).

Il potenziale  $V$  viene derivato dalle equazioni dei FF di MM.

*Impostare un esperimento: aspetti pratici collegati alla realizzazione di un esperimento di MD.*

Le fasi in cui si sviluppa un esperimento di MD possono essere schematizzate come segue:

- punto di partenza per la realizzazione di un buon esperimento di dinamica molecolare, è la disponibilità di una struttura tridimensionale di buona qualità della macromolecola di interesse. Le strutture 3D delle macromolecole possono essere scaricate da apposite banche dati. Il più importante archivio di strutture macromolecolari 3D è il Protein Data Bank (PDB)<sup>1</sup> che raccoglie strutture risolte sperimentalmente (54559 al 1 dicembre 2008), mediante diffrazione a raggi X (46656 strutture), spettroscopia NMR (7598 strutture) o microscopia elettronica (204 strutture), o ottenute con metodi teorici, ad esempio mediante modellazione per omologia.

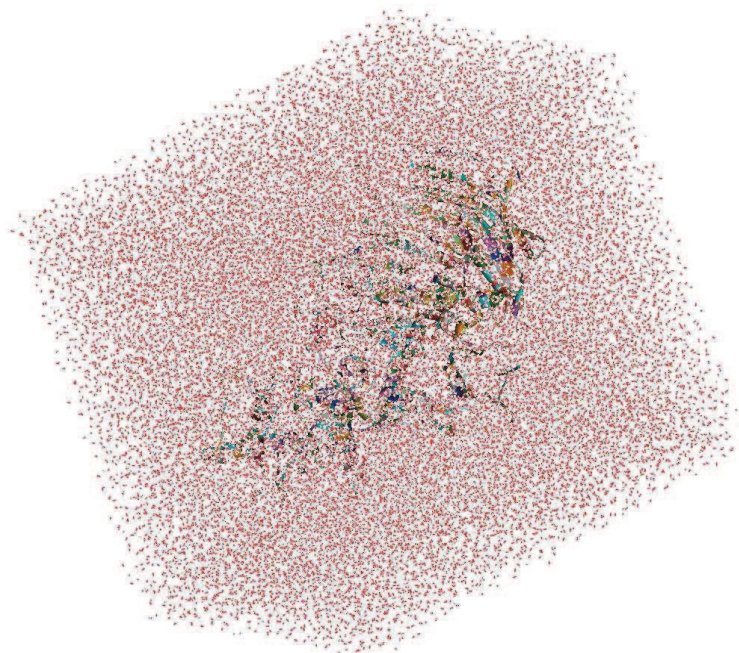
- Le simulazioni di MD devono essere effettuate in un ambiente il più simile possibile a quello effettivo delle macromolecole.

La simulazione delle condizioni di solvatazione sperimentale di macromolecole è oggi realizzata per mezzo delle Periodic Boundary Conditions (PBC).

In questo tipo di strategia computazionale la macromolecola viene posta all'interno di un contenitore (box), a forma di esaedro, di molecole d'acqua (**Figura 2**), replicato poi per simmetria su tutti i suoi lati; con questo

stratagemma si evita che le molecole d'acqua si trovino a contatto con il vuoto, dando luogo a fenomeni di tensione superficiale in cui le molecole di solvente tendono ad aggregarsi in determinati punti dello spazio.

La metodologia PBC sebbene sia ritenuta fra le più valide, porta con sé alcuni inconvenienti; infatti, a causa della replicazione per simmetria dei box di solvente, è necessario inserire nei parametri di calcolo un cut-off delle forze elettrostatiche che eviti l'interazione fra una molecola di solvente e la sua immagine riprodotta per simmetria.



**Figura 2.** Proteina solvatata in un contenitore (box) di forma esaedrica.

- Una volta ottenuta la struttura dalla banca dati, preparato il sistema utilizzando un appropriato modello di solvente, e assegnati i vari parametri necessari per risolvere le equazioni viste in precedenza, i vari programmi di MD procedono all'assegnazione delle velocità ai vari atomi. Questo è un passaggio molto delicato, infatti non è disponibile alcun dato sperimentale riguardo alla velocità degli atomi, le quali vengono poi attribuite in base alla

distribuzione di Maxwell standard alla temperatura  $T$  a cui si effettua la simulazione.

Questo tipo di assegnazione casuale può provocare alcuni problemi, fra cui la creazione di zone più calde all'interno del sistema simulato. A questo proposito è importante precisare che qualunque aggiustamento della temperatura, effettuato durante un esperimento MD, viene eseguito regolando le velocità associate ai singoli atomi. Per limitare al minimo questo tipo di fenomeni, solitamente, si procede ad un lento riscaldamento del sistema, in modo che la temperatura abbia la possibilità di uniformarsi.

La regolazione delle velocità degli atomi, e quindi la regolazione della temperatura, può essere effettuata utilizzando diversi algoritmi. Il più diffuso è il *Berendsen Coupling Algorithm*; questo algoritmo simula un bagno termico che, associato al sistema, permette di mantenere la temperatura costante. È importante notare che l'algoritmo di Berendsen non fissa la temperatura ma porta a un rilassamento esponenziale della temperatura istantanea a quella target e che inoltre non esiste una prova certa che l'algoritmo di Berendsen sia in grado di riprodurre la distribuzione canonica.

Allo scopo di ovviare a questo problema, sono stati messi a punto metodi alternativi per la termostattizzazione di un sistema durante un esperimento di dinamica molecolare. Uno dei più importanti è, senza dubbio, il termostato di Nosè-Hoover. L'idea fondamentale alla base di questo algoritmo è quella per la quale il flusso di calore fra il sistema ed un bagno termico ad esso accoppiato possa essere regolato aggiungendo un ulteriore grado di libertà alla lagrangiana descrivente il sistema. Il concetto di lagrangiana fu introdotto in origine in una riformulazione della meccanica classica nota come meccanica lagrangiana, e quindi rappresenta solo una formulazione matematicamente differente delle equazioni del moto di Newton.

L'uso dell'algoritmo di Nosè-Hoover permette di riprodurre al meglio la distribuzione canonica di un sistema.

Uno dei fattori più importanti in una simulazione di dinamica molecolare è l'intervallo di tempo o time step in cui vengono calcolate le forze agenti sui

vari atomi. Il valore di time step dovrebbe essere inferiore alla velocità del più rapido movimento del sistema. Nel caso di biomolecole il movimento più veloce è quello associato alla stretching dell'atomo di H, fenomeno che ha una frequenza di circa 10 fs; quindi un adeguato time step in questo tipo di simulazione dovrebbe essere di circa 0.5 fs.

Un time step molto piccolo ha però lo svantaggio di allungare in modo consistente i tempi necessari per effettuare una simulazione. Al fine di allungare i tempi di time step, e quindi effettuare simulazioni più lunghe, sono state messe a punto alcune tecniche, tra le quali la più famosa è quella cosiddetta SHAKE. Il protocollo SHAKE è correntemente utilizzato per aggiustare la posizione degli atomi di H dopo ogni step di dinamica molecole in modo che, anche utilizzando un time step più lungo, non si abbiano problemi di eccessiva fluttuazione dei valori di energia.

- Una volta effettuata la simulazione, è necessario procedere all'analisi dei dati ottenuti.

Perché il confronto fra dati sperimentali e calcolati sia significativo è necessario che le simulazioni avvengano nelle stesse condizioni in cui sono stati ottenuti i dati sperimentali.

Uno dei parametri più importanti utili alla valutazione dei risultati degli esperimenti di MD è l'RMSE (Root Mean Square Deviation) cioè la radice della deviazione quadratica media fra le coordinate degli atomi prima e dopo l'esperimento di MD.

Altri parametri ottenibili dal calcolo sono ad esempio il B-factor (fattore di temperatura) e il calore specifico per i quali è possibile il confronto con dati sperimentali.

In questo lavoro di tesi, le simulazioni sono state eseguite a pressione e temperatura costante (NPT), dal momento che questo è l'insieme termodinamico tramite cui vengono ottenuti i dati sperimentali.

L'analisi delle simulazioni è necessaria anche per arrivare ad individuare porzioni della traiettoria in cui alcune proprietà, come l'RMSD globale o l'energia, si siano sufficientemente stabilizzate.

La MD, sebbene rappresenti una grande possibilità per lo studio di macromolecole biologiche e più in generale di sistemi complessi, presenta anche alcuni inconvenienti. In primo luogo il fatto di utilizzare la meccanica classica per descrivere il moto degli atomi, anche se adeguato per molte situazioni, presenta dei limiti, come ad esempio nel trattamento degli idrogeni; il moto dei protoni, come denota la possibilità di avere un effetto tunnel, ha essenzialmente un "comportamento" quanto meccanico.

Inoltre, basandosi sull'approssimazione di Born-Openheimer, la MD non consente di trattare processi come il trasferimento elettronico o il comportamento di un sistema in uno stato eccitato.

Infine, non è secondario ricordare che nonostante le strumentazioni hardware diventino sempre più potenti, i tempi di simulazione non vanno oltre i nanosecondi, tempi che non permettono di analizzare fenomeni molto importanti, come ad esempio l'apertura e chiusura di un canale e il folding proteico, che avvengono su scale dei tempi più lunghe.

### *Docking*

Il molecular docking è una metodologia computazionale tramite cui si cerca di predire la struttura del complesso risultante dall'interazione fra due specie, di solito una piccola molecola (ligande) e una macromolecola target (proteina o acido nucleico).<sup>46, 47</sup>

Alla formazione del complesso sono dunque associati due aspetti fondamentali:

- la generazione della posa (pose generation), ovvero il posizionamento reciproco delle entità interagenti l'una rispetto all'altra.
- la valutazione delle pose (pose evaluation o scoring), ovvero la valutazione dell'energie di binding associate alla formazione del complesso ligande/target.

I software di docking molecolare disponibili oggi in commercio affrontano entrambi gli aspetti: dapprima effettuano un campionamento delle conformazioni dei ligandi all'interno del binding site, successivamente attribuiscono alle pose un punteggio (score) espressione della stima dell'affinità del ligande verso il target.

La maggior parte dei software di docking oggi disponibili per la ricerca dell'orientazione preferenziale di piccole molecole all'interno del proprio target biomacromolecolare, tratta la struttura recettoriale come un'entità rigida, permettendo solo alla piccola molecola (ligande) di adattarsi conformazionalmente al recettore.

L'approssimazione di questi metodi rappresenta un compromesso fra il contenimento della complessità computazionale altrimenti necessaria per poter affrontare e gestire un maggior numero di gradi di libertà conformazionale (flessibilità del recettore) e la possibilità di intraprendere studi di lead finding and optimization, attraverso la riproduzione *in silico* dell'adattamento del ligande all'interno del binding site.

Sono tuttavia sempre più frequenti moduli nei software di docking che affrontano il problema delle variazioni conformazionali mutuamente indotte da ligande e dal target al momento dello stabilirsi del legame (flexible ligand/flexible receptor).<sup>48, 49</sup> In quest'ottica si possono seguire diversi approcci:

- uso di un insieme (ensemble) di conformazioni recettoriali ottenute mediante simulazioni di dinamica molecolare o metodi montecarlo (MC), o ottenute sperimentalmente (Xray-time solved, NMR).<sup>50</sup>
- induced-fit: il mutuo adattamento indotto dal binding ligande/target può essere simulato attraverso:
  - soft docking: rappresenta una procedura che attribuisce una certa adattabilità (plasticità) al sito di binding permettendo che si verifichino alcuni contatti sterici fra il ligande e il target attraverso una modulazione (riduzione) dell'effetto del termine

repulsivo della funzione di valutazione energetica (scoring function).

- side chain flexibility: è una procedura che, applicata ad alcune pose risultanti dal docking con target rigido (flexible ligand/rigid receptor), sostituisce i residui del binding site con residui di alanina, effettua il docking con questa configurazione del recettore, e ricostruisce quindi tutte le pose risultanti le catene laterali dei residui effettivamente presenti all'interno del binding site.<sup>51</sup>

Qualunque sia la natura del software di docking, la sua affidabilità viene valutata determinando i due parametri di accuratezza di docking e scoring (docking e scoring accuracy).

La docking accuracy rappresenta la capacità del software di docking di riprodurre dati strutturali ottenuti sperimentalmente; è valutata misurando la radice delle deviazioni quadratiche medie fra le coordinate atomiche della posa del ligande (RMSD) ottenuta con l'algoritmo di docking, rispetto alle coordinate atomiche del ligande sperimentalmente.

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2} \quad (\text{eq. 10})$$

$\delta_i^2$  = distanza fra coppie di atomi equivalenti.

La scoring accuracy, cioè la capacità del software di predire l'energia libera di binding fra il ligande e il target, è valutata confrontando il punteggio attribuito alla posa dal software di docking con il dato biologico ottenuto sperimentalmente per i ligandi analizzati.

I software stimano l'affinità di binding per mezzo di funzioni di scoring,<sup>46, 52, 53</sup> le quali mancando di considerare appieno i contributi entropici e di

solvatazione, rendono piuttosto conto della complementarità ligande recettore che certamente è un prerequisito per il binding del ligande.

Le “scoring function” possono essere classificate in tre categorie principali:

Force field based: sono funzioni che utilizzano i contributi elettrostatico e di van der Waals, calcolati secondo le leggi della meccanica molecolare, allo scopo di ottenere una valutazione energetica dell’interazione ligande/proteina. Alcune di queste funzioni considerano anche la penalizzazione energetica conseguente all’assunzione della conformazione legata del ligande al recettore (energia di strain).

Funzioni empiriche: nelle funzioni di docking empiriche l’energia libera di binding è suddivisa in una somma di contributi ciascuno opportunamente pesato secondo un coefficiente determinato applicando una regressione lineare multipla su un set di complessi ligando/proteina la cui attività è nota e ottenuta sperimentalmente. I coefficienti sono tali da minimizzare la differenza fra il  $\Delta G$  sperimentale e quello ricavato con la funzione di scoring sul set di composti usati per svilupparla. Appartiene a questa classe la funzione ChemScore<sup>54</sup>, da cui deriva GlideScore, funzione parametro di scoring prevalentemente utilizzata in questo lavoro di tesi.

Knowledge-based: queste funzioni sfruttano le informazioni contenute nei database sperimentali quali il Cambridge Structural Database o Protein Data Bank (PDB); si basano sulla formulazione inversa della legge di Boltzmann la quale trasforma la frequenza con la quale l’atomo A del ligande si trova alla distanza  $r$  dall’atomo B del recettore in un’energia di interazione fra i due atomi espressa in funzione di  $r$ .

Ciascun software di docking implementa algoritmi propri per la fase di posing e per quella di scoring.

In questo lavoro di tesi il software Glide Schördinger® è stato prevalentemente usato.



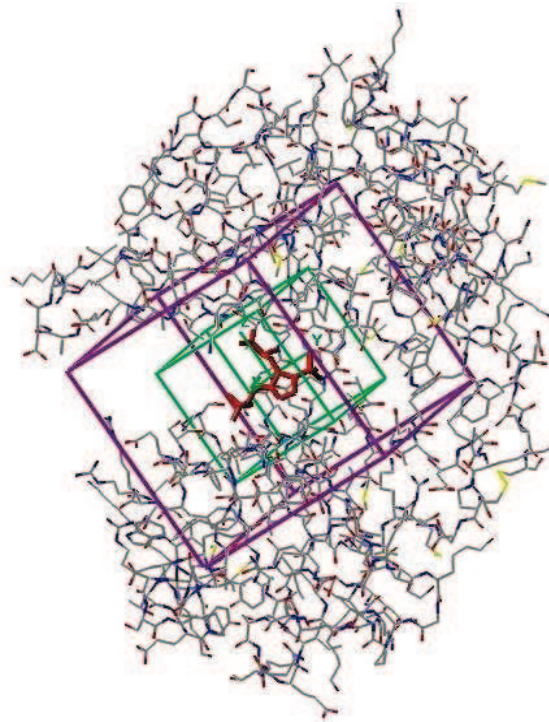
Uno fra i molti software di docking molecolare Glide Schördinger®

Glide<sup>55, 56</sup> (Grid-Based Ligand Docking with Energetics) è una procedura di docking molecolare, sviluppata da Schördinger® per la ricerca delle possibili posizioni assunte dai ligandi all'interno del proprio sito di binding.

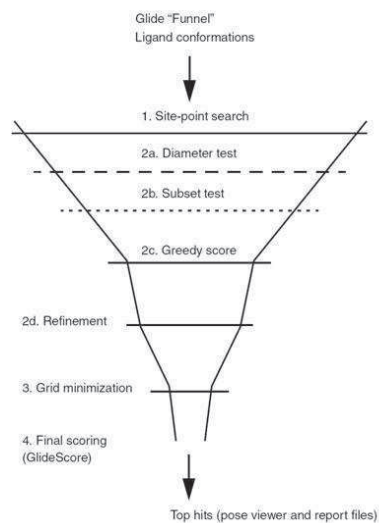
Glide è articolato in step successivi. Dapprima opera un controllo sulla struttura del recettore nella fase denominata *preparazione e rifinitura della proteina*. Quindi procede alla costruzione di una griglia (*grid set up*) utile a definire la forma e le proprietà del sito di binding attraverso il calcolo di campi elettrostatici e di van der Waals ottenuti con il force field (FF) OPLS-AA.<sup>39</sup> In questo modo si individuano sulla griglia dei cosiddetti site points. Operativamente, il *grid set up* prevede la definizione di due griglie denominate bounding box o ligand center box ed enclosing box (**Figura 3**), centrate in genere sul centroide del ligande. La bounding box (verde) racchiude l'area entro cui il centroide del ligande può muoversi durante il docking; l'enclosing box (viola), che contiene la bounding, rappresenta sia l'area per la quale sono calcolati i valori della griglia sia lo spazio fisico entro cui gli atomi del ligande vengono a trovarsi durante il docking. I residui all'esterno dell'enclosing box sono "congelati" e non sono considerati durante i calcoli.

La definizione di questi due spazi è molto importante, sia per la buona riuscita del docking, sia per la durata del calcolo. In particolare, la bounding box deve essere abbastanza grande da contenere tutti i ligandi che devono essere sottoposti al docking, mentre l'enclosing box deve essere di una grandezza tale da contenere tutti i residui che possono influenzare l'interazione con il ligande, senza però allungare eccessivamente i tempi di calcolo. Questa caratteristica si dimostra molto importante nell'analisi di grandi librerie di molecole.

La terza fase di Glide, consistente nel *docking* vero e proprio, si realizza attraverso l'applicazione di una serie di filtri gerarchici (**Figura 4**).



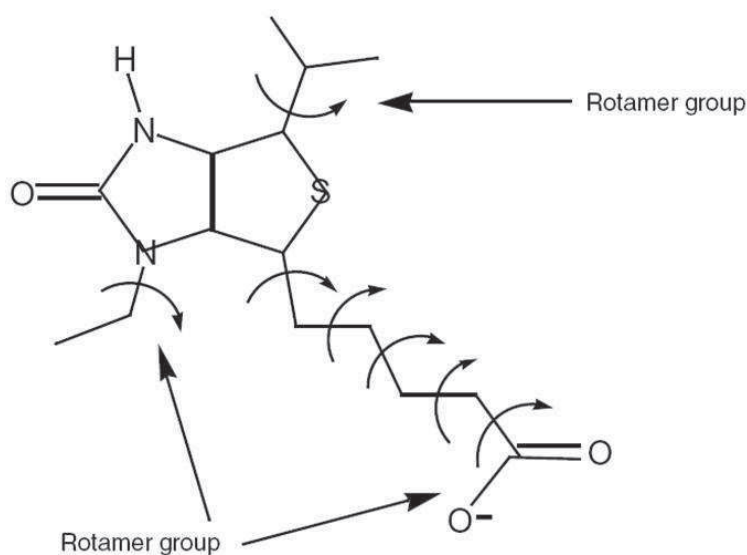
**Figura 3.** rappresentazione grafica della bounding box (in verde) e della enclosing box (in viola). Al centro della bounding box (in rosso) è mostrato il ligande S-ATPO cocristallizzato con la subunità GluR2.



**Figura 4.** Rappresentazione schematica della procedura di docking effettuata da Glide.

In un primo momento, il protocollo di docking esegue una ricerca conformazionale completa all'interno del sito di binding. A questo passaggio è associato un controllo delle conformazioni generate con eliminazione di quelle a più alta energia ritenute non idonee al binding con il recettore (ad esempio conformazioni che possiedono legami a idrogeno interni a lungo raggio).

In questa fase, ciascun ligande è considerato come composto da una parte centrale (core) e da una serie di gruppi rotameri, legati al core da un legame rotabile, ma non contenenti essi stessi legami rotabili. Il core rappresenta perciò quello che rimane nella struttura molecolare tolti i gruppi rotameri appena descritti (Figura 5).



**Figura 5.** Divisione di una molecola in "core" e parti rotabili.

Il numero delle possibili conformazioni assunte dal core dipende dal numero di legami rotabili, dal numero degli anelli a 5 o 6 membri conformazionalmente flessibili e dal numero di atomi di azoto a geometria trigonale asimmetrica (come, ad esempio, l'azoto nelle solfonamidi).

Di solito il numero di queste conformazioni non eccede 500, anche per sistemi con elevata libertà conformazionale, ed è minore ovviamente per sistemi rigidi. Tutte le possibili combinazioni delle conformazioni del core e dei vari gruppi rotameri vengono poi ricostruite e ciascuna sottoposta agli step successivi della procedura di docking nel binding site (docking flessibile), benché Glide sia anche capace di dockare conformazioni già precalcolate (docking rigido).

Su ciascuna conformazione del ligande trovata nella fase precedente, vengono poi calcolate le distanze fra il centro del ligande e la sua superficie. Queste distanze sono confrontate con quelle fra i site point della griglia e la superficie del recettore; il site point è selezionato se si osserva una buona coincidenza fra le distanze del recettore e quelle del ligande.

Tutte le conformazioni generate e ritenute accettabili sono sottoposte alla successiva procedura di docking.

Glide, a questo punto, procede effettuando una prima analisi sterica della cavità di binding, conformazioni che presentano un'elevato numero di contatti o "clash" con il recettore sono scartate.

Questo screening viene eseguito valutando, per le varie conformazioni accettate, la distanza degli atomi della conformazione dal diametro (inteso come la linea che congiunge i due atomi più distanti) del ligande. Vengono scartate quelle conformazioni per le quali si verificano un'eccessiva quantità di contatti fra gli atomi del ligande e quelli del recettore (Figura 3, fase 2a).

Successivamente, le pose non scartate sono mosse intorno al diametro della molecola (Figura 3, fase 2b) e valutate sulla base di uno score ad hoc (greedy score), basato sui seguenti termini:

- legami ad idrogeno instaurati dal ligande
- eventuali legami del ligande con metalli presenti nel sito di binding
- penalizzazione di contatti di tipo sterico (Figura 2, fase 2c).

Si passa a questo punto ad una rivalutazione (rescoring) delle migliori (di solito 5000) pose del greedy score attraverso una procedura di affinamento (Figura 3, fase 2d) che consente al ligande di muoversi rigidamente di  $\pm 1$  Å nelle direzioni X, Y e Z dello spazio. Solo un numero limitato, di solito 400, delle migliori pose rifinite è poi minimizzato all'interno della griglia che descrive il recettore (Figura 3, fase 3).

La scelta della struttura meglio dockata dalla procedura è guidata dal valore dell'Energy score (E<sub>model</sub>) che combina l'energy-grid score, l'affinità di binding predetta dal Glide score e l'energia conformazionale interna (internal energy strain) risultante dall'algoritmo di ricerca conformazionale applicato.

### Metodi per il calcolo della free energy

I “Free energy method” sono quei metodi che consentono di stimare la differenza di energia libera fra due differenti stati rappresentati, nell’ambito del drug design, dal ligande e dalla proteina nel loro stato libero e dal complesso ligande/proteina.

All’interno di questa ampia categoria di metodi teorici possiamo trovare metodi teoricamente molto complessi e rigorosi come la Free Energy Perturbation (FEP)<sup>57</sup>, la Thermodynamic Integration (TI)<sup>58</sup> e la Metadinamica,<sup>44, 59</sup> così come metodi meno rigorosi ma più rapidi come la teoria Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA)<sup>60, 61</sup> e la Linear Interaction Energy (LIE).<sup>60</sup>

#### *Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA)*

La metodologia MM-PBSA,<sup>60, 61</sup> è classificata fra cosiddetti end point method ed è utilizzabile per un rapido e accurato calcolo della differenza di energia libera legata alla formazione del complesso ligando-proteina.

In questo tipo di approccio è necessario produrre una serie di conformazioni delle specie molecolari coinvolte nell’equilibrio studiato (recettore, ligando e complesso ligando-recettore).

Queste conformazioni possono essere prodotte sia utilizzando metodi Monte Carlo (MC) sia simulazioni di MD, come è avvenuto in questo lavoro. Di solito, allo scopo di ridurre le risorse di calcolo necessarie, viene usata la sola traiettoria relativa al complesso ligando-proteina.

Per ogni snapshot della traiettoria considerato viene calcolata l’energia libera di Gibbs (G) per ognuna delle specie molecolari. La differenza di energia libera misurabile a seguito della formazione del complesso è espressa dalla equazione:

$$\Delta G = G_{\text{complex}} - (G_{\text{receptor}} - G_{\text{ligand}}) \quad (\text{eq. 11})$$

L'energia libera di binding  $\Delta$  dell'equazione 12 può essere espressa come somma di un contributo entalpico ( $\Delta H$ ) ed di un contributo entropico ( $\Delta S$ ).

$$\Delta G = \Delta H - T\Delta S \quad (\text{eq. 12})$$

L'entalpia di binding  $\Delta H$  può essere decomposta nei due contributi:

- $\Delta G_{MM}$  fornisce la variazione dell'energia conformazionale conseguente al binding. È calcolata tramite la meccanica molecolare.
- $\Delta G_{solv}$  rappresenta il contributo all'energia libera dell'energia di solvatazione.

$\Delta G_{MM}$  può essere ulteriormente decomposto nella somma dei contributi  $\Delta G_{vdW} + \Delta G_{elc} + \Delta G_{int}$  (energia di van der Waals, elettrostatica e conformazionale).

Il contributo all'energia libera di binding dato dall'energia di desolvatazione si compone di un termine  $\Delta G_{solv-pol}$  dovuto al contributo elettrostatico alla solvatazione e di un contributo delle forze non polari alla solvatazione  $\Delta G_{solv-nopol}$ .  $\Delta G_{solv-pol}$  viene calcolato mediante la risoluzione dell'equazione di Poisson-Boltzmann, mentre  $\Delta G_{solv-nopol}$  è correlato con la superficie accessibile al solvente (SASA, solvent accessible surface area) dalla seguente equazione:

$$\Delta G_{solv-nopol} = \gamma SASA + \beta \quad (\text{eq. 13})$$

Dove  $\gamma$  è la tensione superficiale e  $\beta$  è un coefficiente empirico.

Il contributo entropico al binding può essere determinato qualitativamente mediante normal mode analysis (NMA).

*Linear Interaction Energy (LIE) o Linear Response Method (LRM) o Linear Interaction approximation (LIA).*

Come già l'MM-PBSA, anche questo tipo di approccio al calcolo dell'energia libera è riconducibile ai così detti end point method, nei quali vengono simulati solo lo stato finale ed iniziale del processo di binding.

Il metodo è in grado di stimare la differenza di energia libera di binding fra due molecole utilizzando la seguente equazione.

$$\Delta G = \alpha \left\langle V_{L_{prot}}^{vdw} - V_{L_{solv}}^{vdw} \right\rangle + \beta \left\langle V_{L_{prot}}^{elec} - V_{L_{solv}}^{elec} \right\rangle + \gamma \left\langle V_{L_{prot}}^{cav} - V_{L_{solv}}^{cav} \right\rangle \quad (\text{eq. 14})$$

dove  $V_L^{vdw}$ ,  $V_L^{elec}$  e  $V_L^{cav}$  sono i contributi di van der Walls, elettrostatico e di cavità all'energia potenziale, stimati per il solo ligande all'interno della proteina e del solvente.  $\langle \rangle$  rappresenta la media all'interno di un insieme termodinamico (ensemble average).

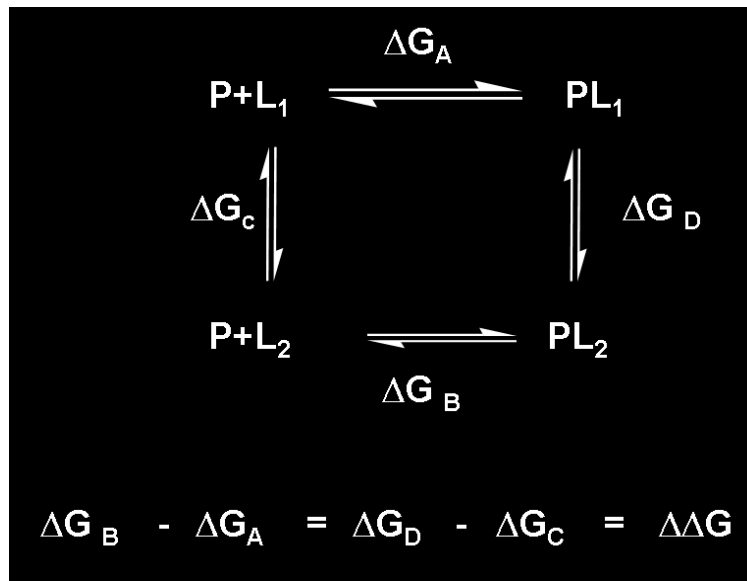
Queste misurazioni vengono effettuate in un insieme termodinamico con pressione, temperatura e numero di particelle costante (NPT), riprodotto mediante dinamica molecolare.

$\alpha$ ,  $\beta$  e  $\gamma$  sono coefficienti empirici che possono essere ottimizzati per ogni sistema e che sono stati oggetto di numerosi studi.<sup>62-64</sup> Per quanto riguarda il valore di  $\beta$ , numerosi autori, raccomandano l'uso del valore di 0.5.

### *Thermodynamic Integration*

Questo tipo di metodologia permette di calcolare in modo molto accurato le differenze in termini di  $\Delta G$  che si hanno in seguito al binding di due o più ligandi.

Dal punto di vista pratico, i metodi di TI calcolano la differenza fra il  $\Delta G$  di binding per un ligando  $L_1$  ( $\Delta G_A$ ) e il  $\Delta G$  di binding per un ligando  $L_2$  ( $\Delta G_B$ ), simulando la trasformazione di  $L_1$  e  $L_2$  attraverso step non fisici in un ciclo termodinamico del tipo:



**Figura 6.** Ciclo termodinamico con cui viene calcolata la differenza di energia libera fra due ligandi.

Il calcolo diretto di  $\Delta G_A$  e  $\Delta G_B$ , sebbene possibile dal punto di vista teorico, è solitamente impraticabile a causa di problemi di convergenza. È però possibile simulare i due processi caratterizzati dall'energia libera  $\Delta G_C$  e  $\Delta G_D$ , e da questi stimare il valore di  $\Delta\Delta G$  ossia la differenza fra le differenze di energia libera.

La trasformazione viene realizzata simulando, mediante dinamica molecolare, una serie di stati intermedi fra  $L_1$  e  $L_2$ , per i quali viene specificato uno specifico valore di un parametro di accoppiamento indicato con  $\lambda$ , che specifica la percentuale di trasformazione di  $L_1$  in  $L_2$ . Per  $\lambda = 0$  avremo una simulazione per il ligande  $L_1$  mentre per  $\lambda = 1$  avremo una simulazione per il ligande  $L_2$ .

Alla fine delle simulazioni sarà possibile ottenere il valore di  $\Delta\Delta G$  fra i due ligandi integrando il valore di  $dG/d\lambda$  per i vari step simulati secondo quanto descritto dall'equazione:

$$\Delta G = \int_0^1 \left\langle \frac{dH(\lambda)}{d\lambda} \right\rangle d\lambda \quad (\text{eq. 15})$$



dove  $H$  indica l'Hamiltoniano del sistema.

Durante la simulazione, per valori di  $\lambda$  diversi da 0 e 1, i parametri di non legame di  $L_1$  e  $L_2$  possono venire interpolati linearmente o attraverso soft-core potential (ossia potenziali di vdW modificati), mentre i parametri di legame e le masse atomiche vengono lasciate invariate. Il tipo di interpolazione può variare.

Nel caso dell'interpolazione di tipo lineare, è necessario effettuare una minimizzazione delle strutture di partenza per le varie simulazioni, così come monitorare i punti molto vicini a 0 e 1, dove le simulazioni tendono a mostrare una forte instabilità numerica.

Dal punto di vista pratico la trasformazione dei ligandi può essere affrontata secondo gli approcci single-topology e double-topology.

Nel primo approccio alcuni atomi della molecola di partenza vengono semplicemente trasformati in altri, mentre quelli conservati non vengono trattati in alcun modo; questo tipo di approccio è quello utilizzato quando si studino modifiche di sostituenti su un determinato lead.

Il secondo metodo invece, prevede la costruzione di due diverse topologie per due diverse molecole. In questo tipo di simulazioni ambedue le molecole  $L_1$  e  $L_2$  sono "geometricamente" presenti già all'inizio ma, mentre per una molecola sono "accesi" tutti i parametri di non legame, per l'altra saranno "spenti"; durante la trasformazione si passerà quindi attraverso degli stati intermedi che porteranno ad uno stadio dove saranno "accese" solo le interazioni della molecola che inizialmente era spenta.

Questo metodo è molto utile quando si valutano variazioni di  $\Delta G$  per molecole fra loro strutturalmente molto diverse o con un diverso binding mode. Tuttavia, allo scopo di evitare interazioni fittizie fra le due molecole, è necessario utilizzare una lista di esclusione che permetta di eliminare il contributo all'energia potenziale delle interazioni fra le due molecole.

Il numero di step con cui  $L_1$  è trasformato in  $L_2$ , come pure la lunghezza delle simulazioni, sono parametri molto delicati, e devono essere decisi per ogni sistema attraverso un'attenta analisi dei risultati intermedi.

La via scelta per la trasformazione, cioè come  $L_1$  viene trasformato in  $L_2$ , essendo l'energia libera una funzione di stato, non influenza i risultati ottenuti.

## *Sezione sperimentale*

*"... Non vogliate negar l'esperienza  
di retro al sol, del mondo senza gente.  
Considerate la vostra semenza  
fatti non foste a viver come bruti  
ma per seguir virtute e canoscenza"*

*(Dante Alighieri, Divina Commedia, Inferno  
canto 5VI, 116-120)*



*Dinamica molecolare*

Il modello del recettore  $\alpha 4\beta 2$  in complesso con due molecole di epibatidina (una per ogni binding site), è stato preparato utilizzando la procedura contenuta nel software di docking Glide.

Le cariche atomiche dell'epibatidina sono state attribuite utilizzando la procedura RESP, fittando un potenziale elettrostatico calcolato al livello di teoria HF/6-31G\*. Il ligando è stato simulato utilizzando i parametri del General Amber Force Field (GAFF)<sup>42</sup>, mentre per la proteina è stato utilizzato Amber03.<sup>68</sup>

L'intero sistema è stato solvatato con un box d'acqua con una distanza minima dalla proteina di 8 Å, e la neutralità del sistema è stata raggiunta aggiungendo 27 ioni Na<sup>+</sup>.

Prima di iniziare la simulazione di MD l'intero sistema è stato minimizzato con un criterio di convergenza di 0.01 kcal mol<sup>-1</sup> Å<sup>-1</sup>.

L'acqua e i controioni sono stati equilibrati, simulando il riscaldamento del sistema per 40 ps fino alla temperatura di 300 K. Successivamente è stata effettuata una simulazione di dinamica molecolare per 15 ns, usando un time step di 0.2 ps.

Durante questa fase il sistema è stato simulato in periodic boundary conditions (PBC); le interazioni di vdW e quelle elettrostatiche a corto raggio sono state stimate con un cut-off di 8 Å, mentre quelle a lungo raggio sono state stimate con il metodo Particle Mesh Ewald (PME).

Gli algoritmi di Berendsen<sup>69</sup> sono stati utilizzati per mantenere il valore della pressione a 1 atm e la temperatura a 300K.

Conformazioni del recettore utili a rappresentare la flessibilità nei successivi studi di docking (flexible ligand/flexible receptor) sono state selezionate per mezzo della procedura implementata in ACIAP.<sup>70, 71</sup> La cluster analysis è stata eseguita su un insieme di residui che, nel primo snapshot, si trovano nel raggio di 7 Å dal ligando. Dopo i primi 1.5 ns, necessari per raggiungere l'equilibrio strutturale sono state estratte 7 conformazioni rappresentative del

recettore (tre per il sito A e quattro per il sito B, Figura 2). Le conformazioni sono state minimizzate utilizzando un criterio di convergenza di 0.01 kcal mol<sup>-1</sup> Å<sup>-1</sup>.

Tutti i calcoli sono stati eseguiti su una work station LINUX con un processore Intel Xeon Quad 5320 e 4GB di RAM, con un tempo di 0.33ns/giorno per un totale di 45 giorni di calcolo.

### *Docking*

Tutti gli esperimenti di docking sono stati effettuati utilizzando il programma Glide 4.0.<sup>55</sup>

Le strutture tridimensionali di AChBP e  $\alpha 4\beta 2$  nAChR sono state preparate utilizzando l'apposito modulo della procedura Glide Schrodinger®.

Le griglie, costruite considerando uno dei cinque siti di binding equivalenti per ciò che riguarda AChBP ed entrambi i siti di binding all'interfaccia delle catene  $\alpha$  e  $\beta$  per ciò che riguarda  $\alpha 4\beta 2$ -nAChR, sono state centrate sul centroide del ligande complessato.

Il docking dei ligandi è stato condotto nella modalità standard e QM-polarized descritta di seguito, utilizzando i settaggi di default (nessun fattore di scala per gli atomi non polari della proteina, fattore di scala 0.8 per gli atomi polari). Le pose, tre per ciascun ligande, risultanti dal docking sono state selezionate utilizzando il valore di Emodel, parametro che combina Glide score, energia di interazione coulombiana, energia di vdW ed energia interna del ligande.

Il protocollo di docking QM-polarized descritto da Cho et al.<sup>72</sup>, è stato applicato sia ai complessi di AChBP sia agli snapshot estratti dai calcoli di MD sul modello del recettore  $\alpha 4\beta 2$ . In questo protocollo ad un primo esperimento di docking utilizzando le cariche attribuite dal FF, fa seguito una seconda fase di redocking, sulle tre pose provenienti dalla prima fase. A

queste pose sono assegnate cariche ESP calcolate per ciascun ligande all'interno della proteina.

Più in dettaglio alla porzione QM del sistema (ligande) viene assegnato un livello di teoria B3LYP/6-31G\*/LACVP\*, mentre la porzione MM (proteina) viene simulata usando il FF OPLS2001. Si ottiene così un potenziale elettrostatico da cui ricavare cariche ESP del ligando, da utilizzare nella fase di redocking. La posa fra le tre così trattate, che possiede il miglior valore di Emodel, risulta essere quella selezionata.

L'applicazione di questa procedura ai vari snapshot estratti dalle simulazioni di MD, fornisce per ciascun ligande tante pose quanti sono gli snapshot considerati. Per ciascun ligande viene quindi selezionata fra queste pose quella con il minor valore di Gscore.

La stima dei parametri di scoring (Emodel e Gscore) del protocollo QM-polarized avviene in modalità SP (standard precision) nella prima fase di docking ed in modalità SP o XP (extra precision) nella seconda fase di redocking. Le due funzioni di scoring SP e XP differiscono per il trattamento delle interazioni idrofobiche, la descrizione del legame ad idrogeno e il valore dei coefficienti con i quali vengono pesate le interazioni elettrostatica e di vdW.<sup>55, 56</sup>

Tutti i calcoli sono stati eseguiti su una work station LINUX con un processore Intel Xeon Quad 5320 e 4GB di RAM, impiegando 22.5 minuti per molecola.

### **Studi sulle Glicosilidrolasi**

#### *Docking su trealasi e MGAM*

Le strutture target/ligande di partenza non disponibili nel PDB e necessarie agli esperimenti di MD, sono state ottenute attraverso docking molecolare. In particolare sono stati preparati i complessi trealasi/casuarina e

MGAM/casuarina-6-O-alfa-glucoside a partire dai complessi 2JGO (trealasi/tiatetrazolina) e 3CTT (MGAM/casuarina).

Le griglie sono state centrate sul centroide del ligando complessato.

Durante gli esperimenti di docking sono stati utilizzati i settaggi di default (nessun fattore di scala per gli atomi non polari della proteina, fattore di scala 0.8 per gli atomi polari).

#### *Dinamica molecolare*

Le strutture cristallografiche dei complessi studiati (pdb code: 2JGO, 3CTT, 2CET, 2J7C, 2J7E, 2J7H, 2J7G, 2J7F), così come quelle dei complessi trealasi/casuarina e MGAM/6-O-alfa-glucoside ottenute mediante docking, sono state preparate utilizzando il modulo leap implementato in Amber.

Le cariche atomiche dei ligandi sono state attribuite utilizzando la procedura RESP, fittando un potenziale elettrostatico calcolato al livello di teoria HF/6-31G\*.

Il ligando è stato simulato utilizzando i parametri del General Amber Force Field (GAFF)<sup>42</sup>, mentre per la proteina è stato utilizzato Amber03.<sup>68</sup>

L'intero sistema è stato solvatato con un box d'acqua con una distanza minima dalla proteina di 10 Å per trealasi e maltasi intestinale umana e 12 Å per la beta-glucosidasi da *Thermotoga maritima*, rispettivamente.

La neutralità del sistema è stata raggiunta aggiungendo ioni Na<sup>+</sup>.

In questo caso tutte le simulazioni sono state effettuate utilizzando il software Gromacs 3.3.3.<sup>73</sup>

Prima di iniziare le simulazioni, l'intero sistema è stato minimizzato in due fasi, nella prima è stato minimizzato solamente il solvente, mentre sulla proteina e sul ligando sono stati applicati dei restrain di posizione; nella seconda fase l'intero sistema è stato lasciato libero di muoversi; in ambedue i casi è stato usato un criterio di convergenza di 0.01 kJ mol<sup>-1</sup> Å<sup>-1</sup>.

Anche l'equilibratura è stata effettuata in due diverse fasi: nella prima sono stati equilibrati solamente l'acqua e i controioni (tenendo ferme proteina e



ligande), simulando il riscaldamento del sistema fino alla temperatura di 298.15 K per 40 ps; nella seconda fase è stato simulato lo stesso processo di riscaldamento lasciando tutto il sistema libero di muoversi.

Successivamente è stata condotta una dinamica molecolare di 6 ns per la beta-glucosidasi, 4.5 ns per la trealasi e per 10.5 ns per MGAM, con un time step di 0.15 ps.

Durante questa fase il sistema è stato simulato in periodic boundary conditions (PBC); le interazioni di vdW e quelle elettrostatiche a corto raggio sono state stimate con un cut-off di 10 Å, mentre quelle a lungo raggio sono state stimate con il metodo Particole Mesh Ewald (PME).

L'algoritmo di Nosè-Hoover è stato utilizzato per mantenere il valore della temperatura a 298.15K, mentre la pressione è stata fissata al valore di 1 atm con l'algoritmo di Parrinello-Rahaman.<sup>74</sup>

Tutti i calcoli sono stati eseguiti su una work station LINUX con un processore Intel Xeon Quad 5320 e 4GB di RAM, con un tempo di 0.39 ns/giorno (totale 11.5 giorni) per la trealasi, 1.39 ns/giorno per la Beta-glucosidasi (totale 4.31 giorni), 0.38 ns/giorno (totale 27.6 giorni) per MGAM.

#### *Linear Interaction Energy*

Le energie necessarie per elaborare il valore di  $\Delta G$  secondo la teoria della Linear Interaction Energy sono state calcolate utilizzando le simulazioni MD del ligando all'interno della proteina descritte in precedenza, e le simulazioni del ligando in un box d'acqua con una distanza minima dal ligando di 12 Å, utilizzando il software Gromacs 3.3.3 e gli stessi parametri descritti per il complesso. I valori di  $\Delta G$  sono stati elaborati utilizzando il valore dell'interazione elettrostatica e di quella di vdW pesati con i parametri

proposti da diversi autori:  $\alpha = 0.181$  e  $\beta = 0.50$  quando il ligando è carico, o  $0.33$  quando il ligando è una specie neutra (proposti da Åqvist e collaboratori,<sup>75</sup>)  $\alpha = 0.476$  e  $\beta = 0.165$  (proposti da Jorgensen e collaboratori<sup>76</sup>) e  $\alpha = 0.418$  e  $\beta = 0.087$  (proposti da Wall e collaboratori<sup>77</sup>).

#### *Calcolo dell'energia libera di binding con MM-PBSA o MM-GBSA*

Tutti i calcoli di questo tipo sono stati effettuati estraendo i dati relativi alle tre specie molecolari coinvolte da una singola traiettoria prodotta utilizzando MD.

Il valore di  $G$  (energia libera di Gibbs) è stato calcolato utilizzando la seguente formula:

$$G = E_{MM} + G_{polar} + G_{nopolar} - TS_{MM} \quad (\text{eq. 1})$$

$E_{MM}$  è stato calcolato utilizzando il modulo sander di AMBER10,<sup>78</sup>  $G_{polar}$  è stato determinato risolvendo l'equazione di Poisson lineare (LPBE),  $G_{nopolar}$  è stato determinato utilizzando l'equazione:

$$G_{nopolar} = SURFTEN \times SASA + SURFOFF \quad (\text{eq. 2})$$

dove SASA è la superficie accessibile al solvente, mentre SURFTEN e SURFOFF sono stati fissati rispettivamente a  $0.0072$  e  $0.00$ .

Questi calcoli sono stati effettuati con il modulo pbsa di AMBER10, per quanto riguarda i calcoli di PBSA, mentre i calcoli GBSA sono stati effettuati utilizzando il modello di Onufriev-Bashford-Case<sup>79</sup> con raggi di Born modificati.

Dopo alcuni tentativi i migliori risultati sono stati ottenuti fissando la costante dielettrica esterna a  $80$  e quella interna  $1$ .

Il valore dalla concentrazione salina non è sembrato influenzare i valori calcolati, ragione per cui è stato settato a 0.

La valutazione delle interazioni di non legame è stata effettuata utilizzando alternativamente i raggi Parse o di vdW, in combinazione con le cariche amber/RESP.

Gli esperimenti di mutagenesi computazionale sono stati effettuati utilizzando i parametri sopra descritti per i calcoli di MM-GBSA, e le mutazioni di tutti gli atomi della proteina sono state effettuate utilizzando l'apposito script disponibile in AMBER9.

Tutti i calcoli sopra descritti sono stati effettuati su alcuni snapshot estratti dopo i primi 2ns delle simulazioni di dinamica molecolare.

Nel caso della beta-glucosidasi il contributo entropico è stato calcolato, mediante normal mode analysis, su sei snapshot estratti dagli ultimi 6 ns di dinamica.

Tutti i calcoli sono stati eseguiti su una work station LINUX con un processore Intel Xeon Quad 5320 e 4GB di RAM, impiegando circa 24 ore per la valutazione del contributo entalpico e 96 ore per quello entropico

## *Risultati e Discussione*

*Se i dati e la teoria non concordano, cambia i dati.*

*(A. Einstein 1879 – 1955, premio Nobel per la Fisica 1921)*

## Recettore nicotinicco $\alpha 4\beta 2$



Il coinvolgimento dei nAChRs in molte funzioni biologiche, rende questo recettore un target particolarmente interessante per il chimico farmaceutico impegnato nella progettazione e sviluppo di nuovi ligandi.

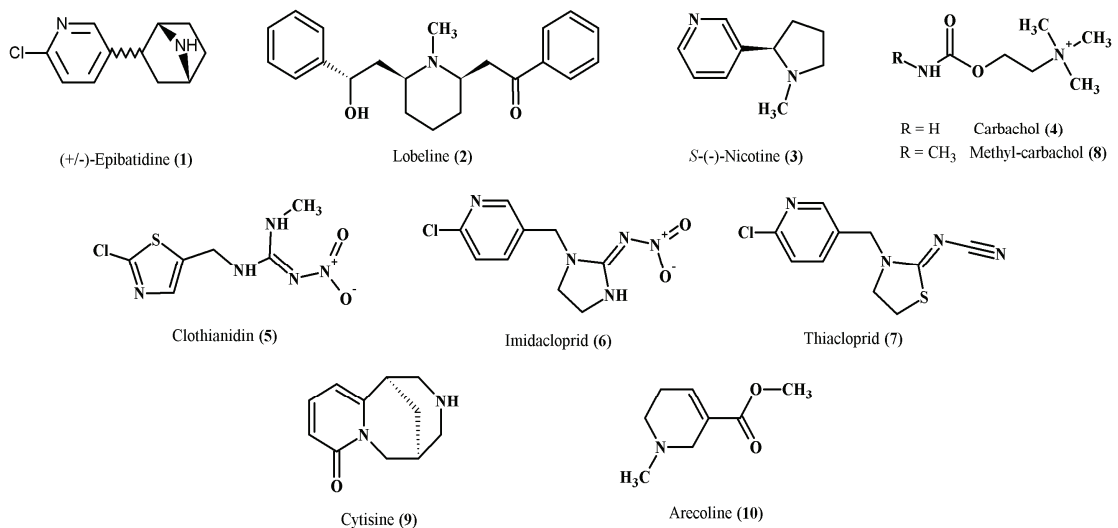
In questo lavoro di tesi è stata affrontata la messa a punto di un protocollo per la determinazione dell'energia libera di binding ( $\Delta G$ ) di complessi ligandi/recettore nicotinicco  $\alpha 4\beta 2$ , per mezzo di metodologie QM/MM accoppiate a studi di MD, si è inteso in questo modo prendere in considerazione fenomeni quali la flessibilità recettoriale e gli effetti di polarizzazione che hanno luogo durante il binding. Questi fenomeni assumono un ruolo importante in generale ma in modo particolare per questo specifico target.

### *Messa a punto e convalida del protocollo di lavoro*

La fase di messa a punto e convalida della procedura di docking QM/MM è stata condotta utilizzando dati strutturali, disponibili nel PDB, relativi all'Achetylcoline Binding Protein (AChBP) complessata con ligandi (**Figura 1**, composti 1, 2, 3, 5, 6 e 7) per i quali si disponeva del valore di  $K_d$  del complesso.

In particolare sono state selezionati i complessi:

1. 2BYQ<sup>13</sup> AChBP in complesso con Epibatidina (*Aplysia californica*) (Kd 14 nM<sup>80</sup>, risoluzione: 3.40 Å)
2. 1UV6<sup>9</sup> AChBP in complesso con Carbamilcolina (*Limnea stagnalis*) (Kd 5600 nM<sup>80</sup>, risoluzione: 2.20 Å)
3. 2BYS<sup>13</sup> AChBP in complesso con Lobelina (*Aplysia californica*) (Kd 0.3 nM<sup>12</sup>, risoluzione: 2.05 Å )
4. 1UW6<sup>9</sup> AChBP in complesso con Nicotina (*Limnea stagnalis*) (Kd 86 nM<sup>81</sup>, risoluzione: 2.20 Å )
5. 2ZJU<sup>17</sup> AChBP in complesso con Clotianidina (*Limnea stagnalis*) (Kd 1570 nM<sup>17</sup>, risoluzione: 2.58 Å )
6. 2ZJV<sup>17</sup> AChBP in complesso con Imidacloprid (*Limnea stagnalis*) (Kd 7260 nM<sup>17</sup>, risoluzione: 2.70 Å )
7. 3C79<sup>18</sup> AChBP in complesso con Imidacloprid (*Aplysia californica*) (Kd 63 nM<sup>18</sup>, risoluzione: 2.48 Å )
8. 3C84<sup>18</sup> AchBP in complesso con Tiacloprid (*Aplysia californica*) (Kd 14 nM<sup>18</sup>, risoluzione: 2.94 Å ).



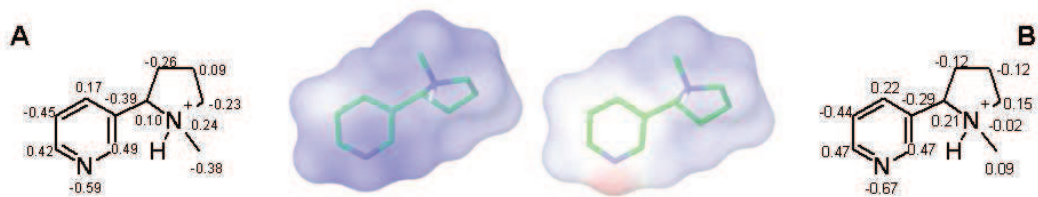
**Figura 1.** Agonisti nicotinici utilizzati in questo lavoro.

La fase di validazione è consistita nella verifica delle performance della procedura di docking QM-polarized eseguita in modalità SP nella prima fase di docking e in modalità sia SP che XP nella seconda fase di redocking.

Il protocollo QM-polarized prevede infatti un primo docking del ligande con cariche da FF (OPLS2001), seguito da un redocking con cariche ESP assegnate trattando il ligande a livello QM nell'ambiente della proteina trattata a livello MM.

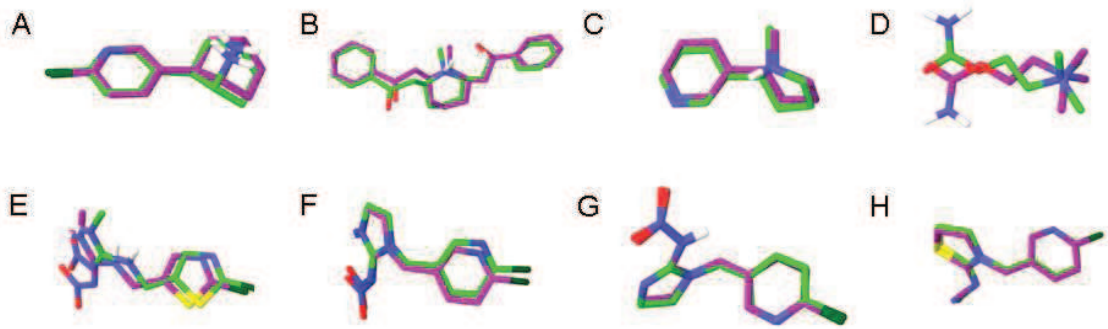
Questo tipo di procedura permette un miglior trattamento dell'interazione elettrostatica, che si riflette in un aumento dell'accuratezza del docking.

La Figura 2 mette in evidenza le differenze fra la superficie del potenziale elettrostatico ottenuta usando le cariche ESP e quelle da FF.



**Figura 2.** Potenziale elettrostatico plottato sulla superficie di Connolly usando **A)** cariche atomiche ESP derivate da un calcolo single point a B3LYP/6-31\* all'interno della proteina **B)** cariche derivate dal FF OPLS2001. I colori rosso e blu rappresentano rispettivamente potenziale negativo e positivo.

La procedura di docking sugli otto complessi di AChBP/ligandi elencati in precedenza è stata inoltre eseguita in modalità standard (non QM-polarized). I risultati riportati in Tabella 1 e la rappresentazione grafica di **Figura 3**, mostrano come tutti i protocolli di docking siano in grado di trovare l'orientazione del ligande nativo, con un valori di RMSD rispetto alla posa cristallografica quasi sempre inferiori ai 2 Å. Le due sole eccezioni sono le rappresentate dalla Clotianidina (modalità XP e SP con cariche da FF) e dall'Imidacloprid (modalità XP con cariche da FF).



**Figura 3.** Risultati del docking QM-polarized SP su AChBP, **A** (-)-Epibatidina (pdb code 2BYQ), **B** Lobelina (pdb code 2BYX), **C** (S)-Nicotina (pdb code 1UW6), Carbacolo (pdb code 1UV6), Clotianidina (pdb code 2ZJV), Imidacloprid (pdb code 2ZJU), Imidalcoprid (pdb code 3C79), Tiacloprid (pdb code 3C84). In verde gli atomi di carbonio della cristallografia, in magenta gli atomi di carbonio delle pose risultanti dal docking. Gli atomi di idrogeno riportati solo quelli polari.

Nel caso del ligande neonicotinoide Clotianidina è necessario osservare come all'errore associato al suo posizionamento in AChBP non corrisponda un analogo errore nella valutazione dell'affinità di binding.

Nel caso dell'Imidacloprid l'errore è con probabilità da imputarsi ad un fallimento del protocollo di docking standard.

E' tuttavia interessante notare che l'applicazione del protocollo QM-polarized consente di individuare per tutti i ligandi, compresi, Clotianidina e Imidacloprid, soluzioni con valori di RMSD inferiori ai 2 Å, e che i risultati più soddisfacenti si ottengono con la modalità di rescoring SP.

Infatti, il valore medio dell'errore nella predizione dell'energia libera di binding dei risultati QM-polarized SP è inferiore alle 2 kcal/mol ( $\Delta$  medio), che corrisponde a circa 3 unità di  $pK_d$ . Da un punto di vista chimico-farmaceutico questo risultato è certamente molto significativo dal momento che contribuisce alla riduzione del numero di falsi positivi che potrebbero risultare in un processo di virtual screening.



Tabella 1. Risultati per la valutazione del protocollo di docking su AChBP.

	QM-polarized docking						Standard docking					
	Modalità di scoring XP			Modalità di scoring SP			Modalità di scoring XP			Modalità di scoring SP		
Codice PDB (Ligande)	RMSD (Å)	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)	RMSD (Å)	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)	RMSD (Å)	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)	RMSD (Å)	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)
		Differenza ( $\Delta$ ) <sup>a</sup>			Differenza ( $\Delta$ ) <sup>a</sup>			Differenza ( $\Delta$ ) <sup>a</sup>			Differenza ( $\Delta$ ) <sup>a</sup>	
EBYQ (Epibatidina)	1.3	-9.3	-10.7	0.3	-9.1	-10.7	1.3	-9.5	-10.7	0.1	-9.1	-10.7
		$\Delta = -1.4$			$\Delta = -1.6$			$\Delta = -1.2$			$\Delta = -1.6$	
2BYS (Lobelina)	0.6	-8.4	-13.0	0.5	-11.3	-13.0	0.5	-14.7	-13.0	0.6	-10.9	-13.0
		$\Delta = -4.7$			$\Delta = -1.7$			$\Delta = 1.7$			$\Delta = -2.1$	
1UW6 (Nicotina)	0.3	-7.8	-10.1	0.4	-8.2	-10.1	0.3	-7.5	-10.1	0.3	-7.8	-10.1
		$\Delta = -2.3$			$\Delta = -1.9$			$\Delta = -2.6$			$\Delta = -2.3$	
1UV6 (Carbacolo)	1.4	-5.6	-7.2	1.4	-6.7	-7.2	1.3	-5.5	-7.2	1.5	-6.0	-7.2
		$\Delta = -1.6$			$\Delta = -0.5$			$\Delta = -2.0$			$\Delta = -1.2$	
2ZJV (Clotianidina)	0.9	-6.6	-7.0	1.3	-6.3	-7.0	2.3	-6.8	-7.0	6.3	-5.7	-7.0
		$\Delta = -0.4$			$\Delta = -0.7$			$\Delta = -0.2$			$\Delta = -1.3$	
2ZJU (Imidacloprid)	0.4	-7.3	-7.8	0.4	-6.2	-7.8	0.4	-7.3	-7.8	0.4	-6.0	-7.8
		$\Delta = -0.5$			$\Delta = -1.6$			$\Delta = -0.5$			$\Delta = -1.8$	
3C79 (Imidacloprid)	0.3	-6.6	-9.8	0.2	-6.4	-9.8	6.1	-5.3	-9.8	0.3	-6.2	-9.8
		$\Delta = -3.2$			$\Delta = -3.4$			$\Delta = -4.5$			$\Delta = -3.6$	
3C84 (Tiacloprid)	0.4	-5.0	-10.7	0.2	-6.9	-10.7	0.3	-5.0	-10.7	0.2	-6.3	-10.7
		$\Delta = -5.7$			$\Delta = -3.8$			$\Delta = -5.7$			$\Delta = -4.4$	
$\Delta$ medio		2.5 kcal/mol			1.9 kcal/mol			2.3 kcal/mol			2.3 kcal/mol	

$$^a \Delta = \Delta G_{\text{exp}} - \text{Gscore}$$

I risultati descritti testimoniano l'importanza di considerare gli effetti di polarizzazione durante un esperimento di docking, specie nel caso di composti, come quelli che si legano al recettore nicotino, per i quali è riconosciuta l'importanza degli effetti di polarizzazione sulla stabilizzazione del binding al target. E' noto infatti come l'interazione di tipo  $\pi$ -catione giochi un ruolo di primo piano nel legame di ligandi al recettore  $\alpha 4\beta 2$ -nAChR. Una valutazione quantitativa di questo tipo di interazione può essere fatta solo tenendo conto esplicitamente di effetti di polarizzazione<sup>82</sup>, trasferimento di carica e nube elettronica<sup>83-86</sup> ed il protocollo QM-polarized, utilizzato in questo lavoro, ha il vantaggio di considerare questi effetti almeno parzialmente.

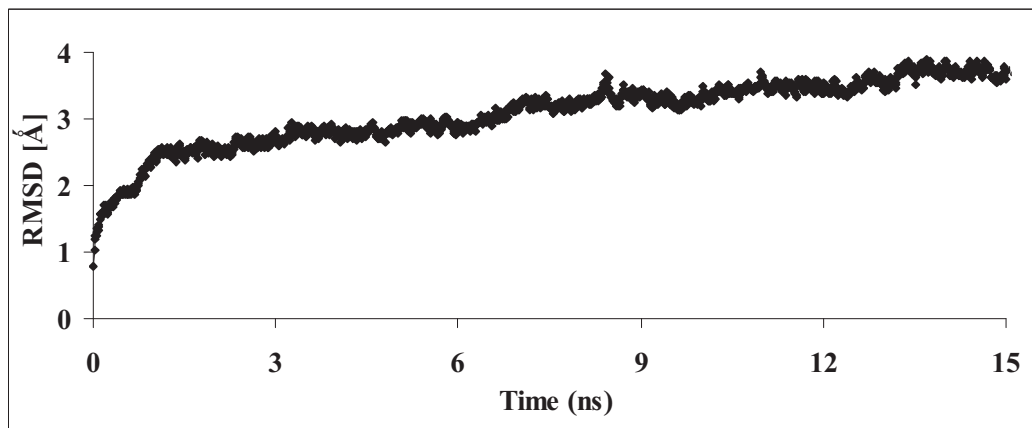
*Analisi della flessibilità del recettore: studi di dinamica molecolare MD sul modello di omologia  $\alpha 4\beta 2$ -nAChR*

Il PDB non fornisce alcuna struttura 3D risolta sperimentalmente di  $\alpha 4\beta 2$ -nAChR. Si è quindi resa necessaria la costruzione di un modello teorico del recettore  $\alpha 4\beta 2$  attraverso la tecnica della modellazione per omologia, usando come template la struttura di AChBP in complesso con nicotina (pdb code 1UW6<sup>9</sup>). Questa struttura è caratterizzata da un'omologia globale abbastanza bassa con  $\alpha 4\beta 2$ -nAChR (circa il 30%), anche se a livello del sito di binding quasi tutti i residui coinvolti nell'interazione risultano conservati.

Le procedure di docking impiegate di routine in ambito industriale ed accademico sono di tipo ligande flessibile/recettore rigido. Tener conto durante il docking della flessibilità del target biologico (docking flessibile/recettore flessibile) è infatti un processo dispendioso dal punto di vista computazionale, ed in genere limitato al solo movimento delle catene laterali della proteina. L'uso di procedure di dinamica molecolare può aiutare a superare questo tipo di limitazione consentendo, grazie anche alla messa a punto di codici completamente parallelizzati, lo studio della

variabilità conformazionale del target macromolecolare in un ambiente di solvente esplicito.

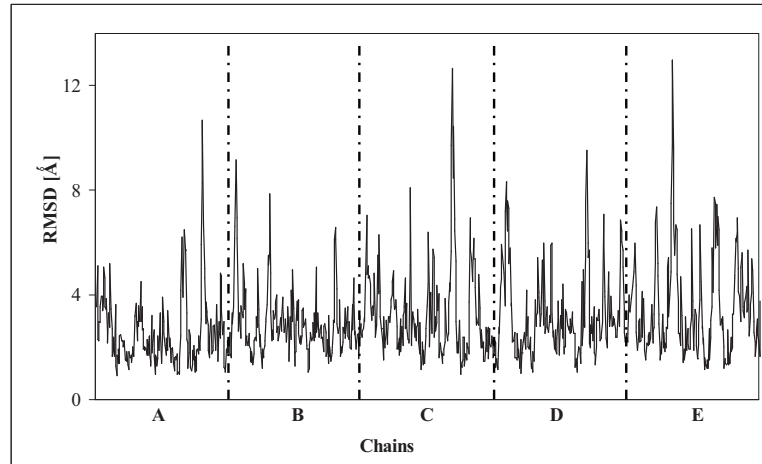
Gli studi di dinamica cominciano con una preventiva fase di ottimizzazione del modello del target ottenuto per omologia. Questa fase ha lo scopo di risolvere distorsioni che si possono verificare nella proteina e di consentire alle molecole del solvente (acqua) di rilassarsi intorno alla struttura molecolare. Il raggiungimento dell'equilibrio strutturale conseguente a questa operazione, avvenuta in presenza dell'agonista nicotino epibatidina nel binding site, è avvenuto dopo 1.5 ns dall'avvio della simulazione, protrattasi complessivamente per 15 ns (**Figura 4**).



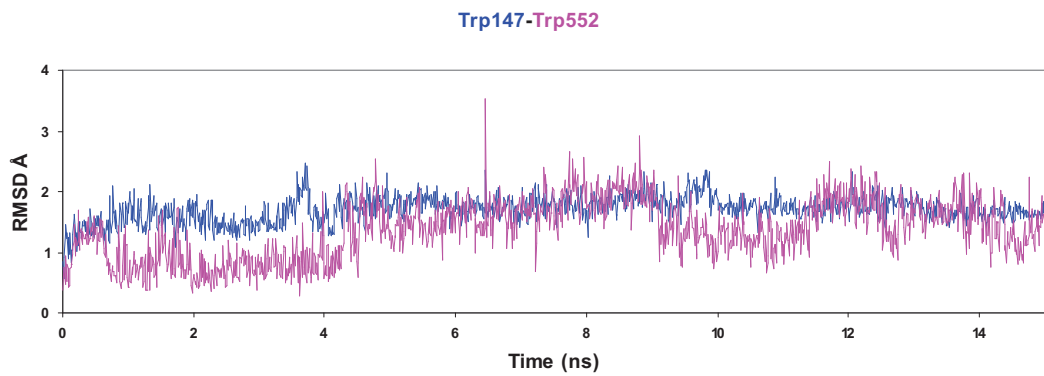
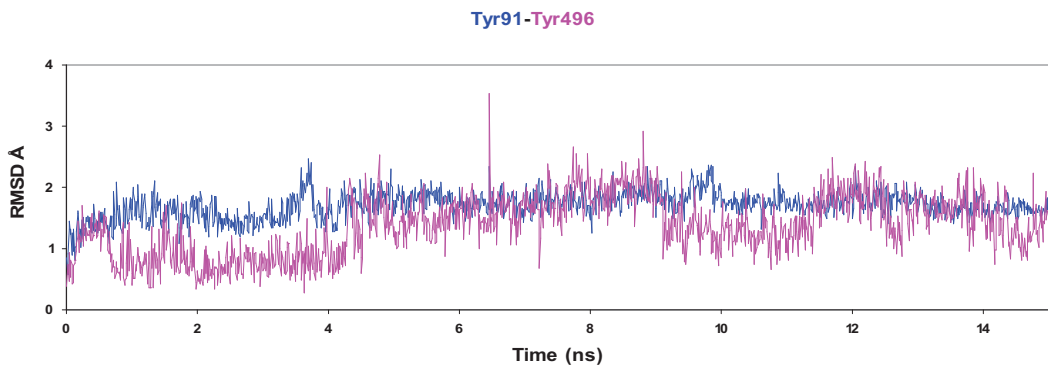
**Figura 4.** Valore di RMSD in Å per tutti i carboni alfa del complesso fra il modello del recettore in complesso con l'epibatidina *vs* tempo di simulazione.

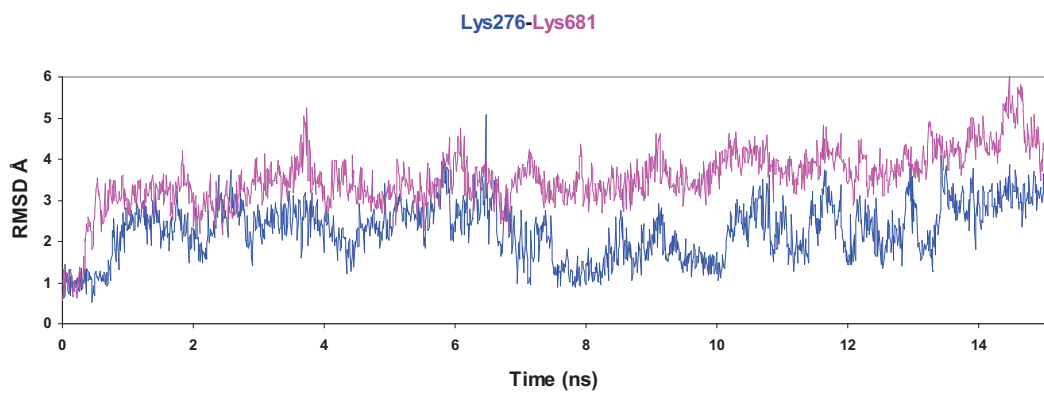
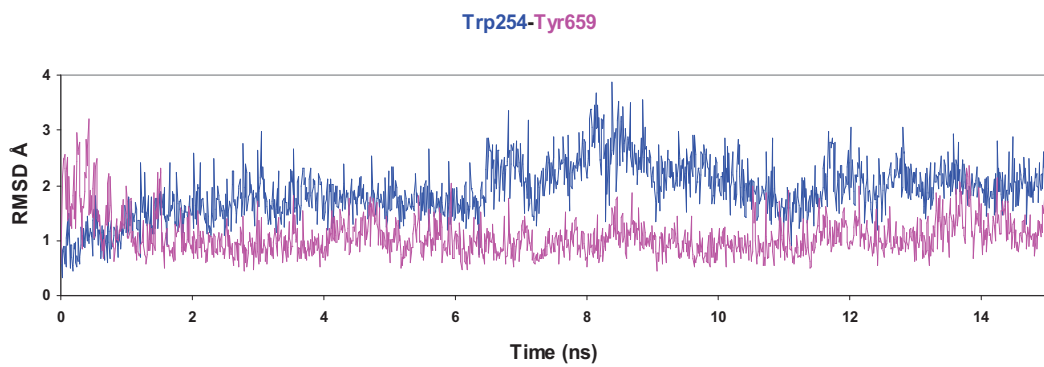
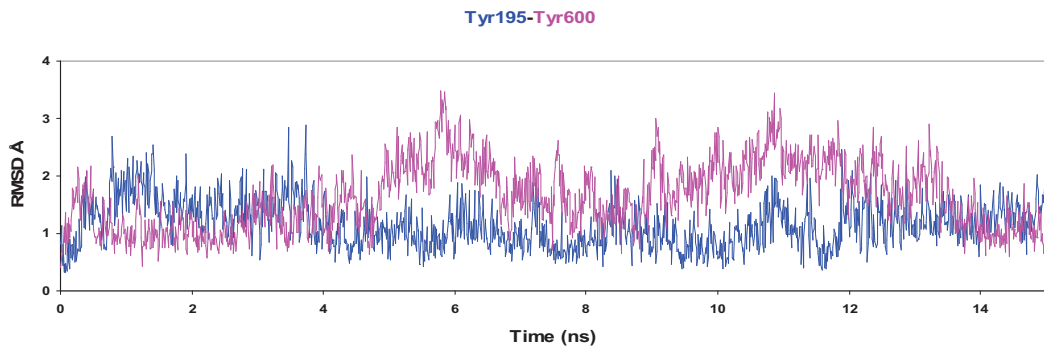
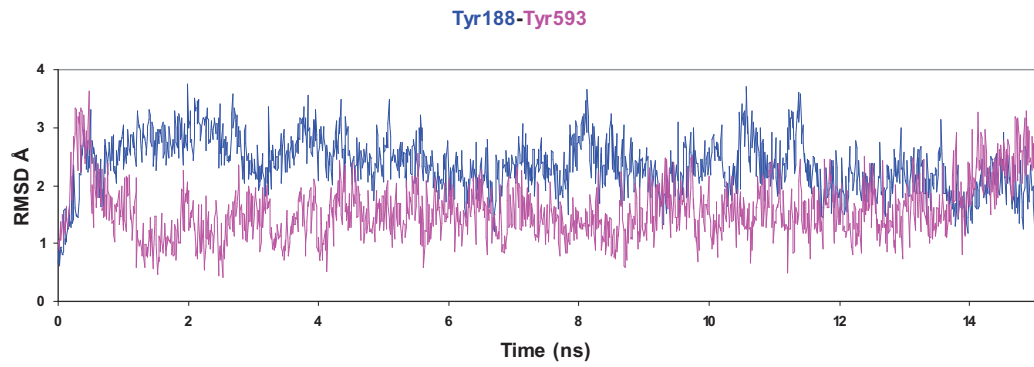
L'entità delle trasformazioni subite dal modello durante i 15 ns di MD sono state monitorate calcolando il valore dell' RMSD medio per ogni singolo residuo del modello (**Figura 5**). L'analisi del grafico mette in evidenza come i movimenti più significativi (RMSD minore di 5 Å) si abbiano a livello degli estremi di ciascuna delle 5 catene che formano la porzione extracellulare N-terminale del recettore, regioni che si trovano al confine con la porzione transmembrana non modellata, e quindi non inclusa nello studio di nAChR-

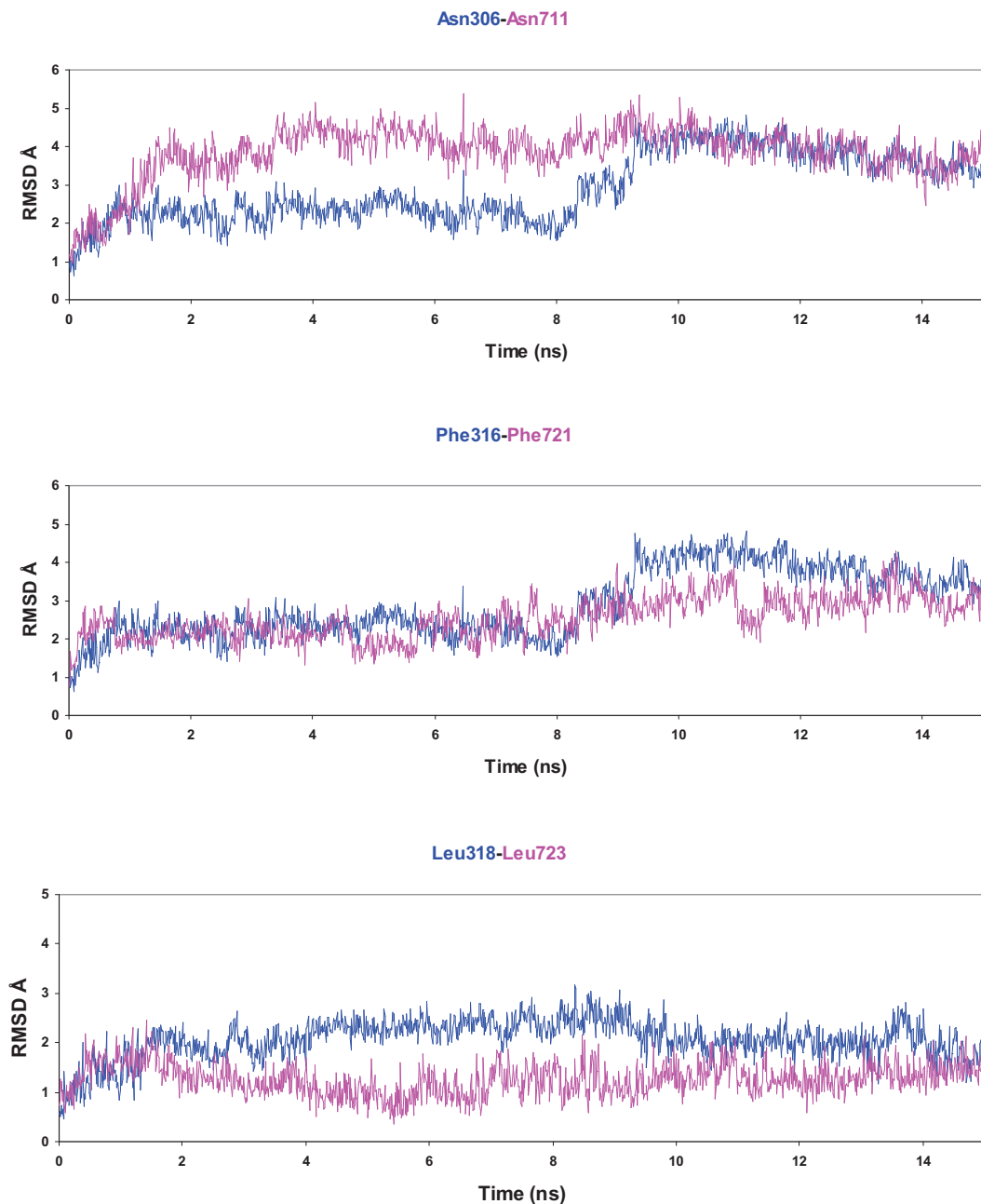
$\alpha 4\beta 2$ . Il probabile coinvolgimento di questi residui nel meccanismo di apertura/chiusura del canale ionico<sup>87</sup>, rende questo risultato meno sorprendente.



**Figura 5.** RMSD medio per ogni residuo in Å. Le lettere A-E indicano le 5 catene che compongono il modello.







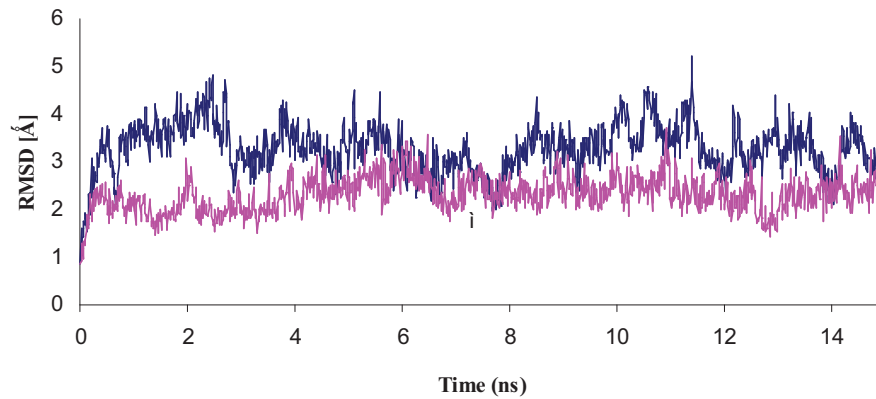
**Figura 6.** Analisi dell'andamento temporale del RMSD per i residui del sito binding.

Ulteriore verifica dei movimenti subiti dal modello è stata condotta determinando i valori di RMSD per alcuni residui del binding site rispetto alla loro posizione di partenza. Tale verifica è stata su nove residui (sito A Y91, W147, Y188, Y195, W254, K276, N306, F316, L318; per il sito B Y496, W552, Y593, Y600, W659, K681, N711, F721, L723) di entrambi i binding site, uno per ciascuna delle interfacce  $\alpha 4/\beta 2$  presenti in  $\alpha 4\beta 2$ -nAChR.

La Figura 6 riporta l'andamento del valore di RMSD di Figura 6 di questi residui durante l'intera simulazione MD.

L'osservazione degli andamenti dell' RMSD riportati nei grafici indica come la struttura del binding site si preservi sostanzialmente durante la dinamica e come questo comportamento si applichi ai residui di entrambi i binding site.

Un particolare riguardo, nello studio delle trasformazioni subite dal modello durante la dinamica, è stato rivolto ai residui che compongono il loop C, ovvero quella porzione recettoriale che con i suoi movimenti regola il legame di agonisti al binding site e di conseguenza blocca il passaggio degli ioni dal canale. Il monitoraggio dell'andamento durante la dinamica dei residui del loop C risulta particolarmente importante visto il collegamento diretto di questo loop alla porzione transmembrana del recettore nicotinic, non modellata per omologia. Sarebbe di conseguenza plausibile attendersi, data anche la mancanza di constrain imposti ai residui della porzione extracellulare, il verificarsi di consistenti variazioni conformazionali a carico dei residui di questo loop.<sup>88</sup> Tali variazioni tuttavia non si sono verificate (**Figura 7**) confortandoci sulla tenuta del modello della porzione extracellulare del recettore nicotinic durante le simulazioni di MD e sull'uso di snapshot diversi estratti dalla traiettoria di dinamica molecolare quali a rappresentanti della variabilità conformazionale del recettore (e quindi del fenomeno dell'adattamento indotto) nel successivo studio di docking di una serie di ligandi agonisti del recettore nicotinic  $\alpha 4\beta 2$  (**Figura 1**, composti 1, 3, 4, 8, 9 e 10)



**Figura 7.** Variazione nel tempo del valore di RMSD per i residui del loop C, Sito A (187-194) blu, Sito B (592-599) magenta.

*Applicazione del protocollo validato di docking QM-Polarized alla determinazione del  $\Delta G$  di complessi  $\alpha_4\beta_2$ -nAChR/ligandi.*

Il protocollo di docking QM-polarized validato su AChBP è stato applicato allo studio dell'interazione fra i ligandi nicotinici della (Figura 1, composti 1, 3, 4, 8, 9 e 10) ed un certo numero di snapshot derivati da simulazioni di dinamica molecolare.

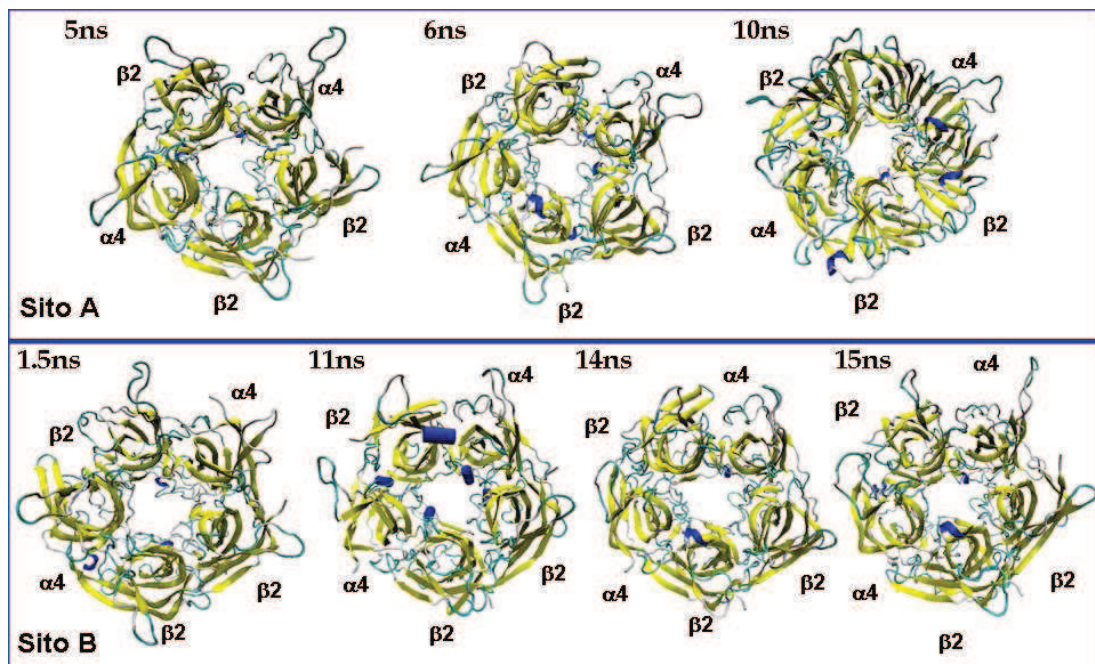
L'uso di questi snapshot quale strumento per la valutazione della flessibilità conformazionale del target recettoriale in studi di interazioni con ligandi, rappresenta un approccio ancora non troppo diffuso, seppure in alcuni casi usato,<sup>89,90</sup> per considerare l'adattamento indotto ligande/recettore che si verifica, nel docking, in seguito allo stabilirsi dell'interazione fra queste due entità.

Le condizioni sperimentali con le quali è stata condotta la simulazione di dinamica molecolare prevedevano la registrazione di uno snapshot ogni 10 ps per un totale di 1500 snapshot nei complessivi 15 ns di durata della dinamica. Su questi frame e per ciascuno dei due siti di binding di  $\alpha_4\beta_2$ -nAChR (siti A e B), è stata condotta un'analisi di cluster sui residui entro 7 Å dall'epibatidina per mezzo del software ACIAP<sup>70</sup> che ha condotto all'individuazione di 76 cluster per il sito A e 84 cluster per il sito B. La



selezione dei frame che, sulla base del criterio di Chauvenet di ACIAP, sono risultati rappresentativi dei cluster individuati ha condotto all'individuazione di 7 snapshot complessivi, tre per il sito A (a circa 5, 6 e 10 ns) e quattro per il sito B (a circa 1.5, 11.5, 14 e 15 ns).

Individuati i 7 snapshot rappresentativi della flessibilità conformazionale del recettore (**Figura 8**), su di essi è stato applicato il protocollo di docking QM-polarized per i ligandi: ( $\pm$ )-epibatidina (Figura 1, composto 1), S-(-)-nicotina (Figura 1, composto 3), carbacolo (Figura 1, composto 4), metilcarbacolo (Figura 1, composto 8), citisina (Figura 1, composto 9), R-arecolina e S-arecolina (Figura 1, composto 10) dei quali si dispone del valore di  $K_d$  determinato sperimentalmente.<sup>91</sup>



**Figura 8.** Rappresentazione tube dei sette snapshot estratti dalla MD.

La Tabella 2 riporta i risultati ottenuti dall'applicazione della procedura di docking QM-polarized in modalità SP e XP (cfr sezione sperimentale), selezione della posa con miglior Emodel fra le tre ottenute in seguito al docking nei singoli snapshot, e successiva selezione, fra le sette pose individuate, di quella a minor Gscore per ciascun ligande. In Tabella 2 è inoltre presente la colonna QM-SP<sub>rigido</sub> con gli score ottenuti sul primo

snapshot da MD ed il valore dell'errore medio di predizione di  $\Delta G$  ottenuto nelle tre modalità di docking.

**Tabella 2.** Risultati del docking QM polarized sugli snapshot del modello di  $\alpha_4\beta_2$ -nAChR.

Ligande	QM-XP flessibile			QM-SP flessibile			QM-SP rigido	
	Tempo	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)	Tempo	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)	Gscore (kcal/mol)	$\Delta G_{\text{exp}}$ (kcal/mol)
		<i>Differenza (<math>\Delta</math>)<sup>a</sup></i>			<i>Differenza (<math>\Delta</math>)<sup>a</sup></i>		<i>Differenza (<math>\Delta</math>)<sup>a</sup></i>	
(-)-Epibatidina	5 ns	-13.12	-12.1	6.3 ns	-8.3	-12.1	-7.5	-12.1
		$\Delta = 1.02$			$\Delta = -3.8$		$\Delta = -4.6$	
(+)-Epibatidina	5 ns	-13.22	-12.1	5 ns	-7.8	-12.1	-7.6	-12.1
		$\Delta = 1.12$			$\Delta = -4.3$		$\Delta = -4.5$	
S-(-)-Nicotina	6.3 ns	-9.0	-10.4	6.3 ns	-7.2	-10.4	-7.1	-10.4
		$\Delta = -1.4$			$\Delta = -3.2$		$\Delta = -3.3$	
Carbaccolo	15 ns	0.41	-8.3	5 ns	-6.7	-8.3	-4.2	-8.3
		$\Delta = -8.71$			$\Delta = -1.6$		$\Delta = -4.1$	
Metilcarbaccolo	13.8 ns	-6.7	-9.8	13.8 ns	-5.9	-9.8	-5.8	-9.8
		$\Delta = -3.1$			$\Delta = -3.9$		$\Delta = -4.0$	
Citisina	1.5 ns	-11.2	-11.8	1.5 ns	-8.0	-11.8	-6.7	-11.8
		$\Delta = -0.6$			$\Delta = -3.8$		$\Delta = -5.1$	
R-Arecolina	5 ns	-9.4	-8.2	11.5 ns	-6.5	-8.2	-6.2	-8.2
		$\Delta = 1.2$			$\Delta = -1.7$		$\Delta = -2$	
S-Arecolina	15 ns	-8.4	-8.2	5 ns	-6.3	-8.2	-6.3	-8.2
		$\Delta = 0.2$			$\Delta = -1.9$		$\Delta = -1.9$	
$\Delta$ medio		2.2 kcal/mol			3.0 kcal/mol		3.7 kcal/mol	

$$^a\Delta = \Delta G_{\text{exp}} - \text{Gscore}$$

È interessante notare come nella modalità XP flessibile, se non si considerano le due molecole non drug like carbacolo e metilcarbacolo, l'errore medio nella predizione del  $\Delta G$  diminuisce sensibilmente (inferiore a 1 kcal/mol). Le leggermente migliori prestazioni dei risultati in modalità XP rispetto a quelli in modalità SP, sono probabilmente da imputare al fatto che questa scoring function contiene contributi specifici per la valutazione dell'interazione  $\pi$ -catione e di stacking, di particolare rilievo per nAChR.

Nell'ottica di verificare se considerare i movimenti strutturali potesse effettivamente aumentare l'accuratezza delle predizioni, abbiamo comparato i risultati ottenuti usando gli snapshot estratti da MD con quelli ottenuti utilizzando un docking QM-polarized sulla prima conformazione del modello di  $\alpha 4\beta 2$ -nAChR (**Tabella 2**, ultima colonna).

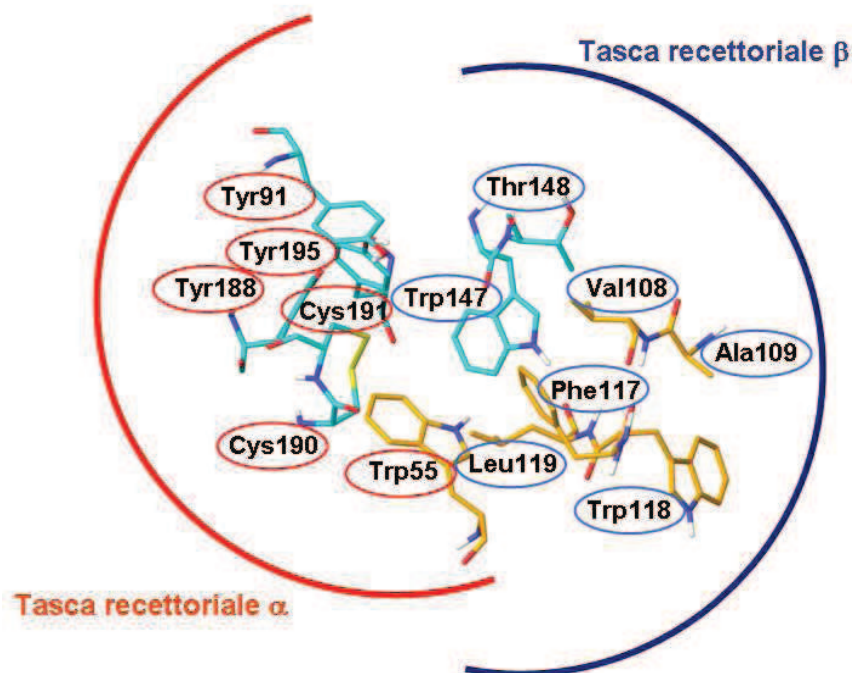
I risultati ottenuti mostrano una sensibile riduzione dell'accuratezza delle predizioni con un errore medio di 3.7 kcal/mol.

Considerati globalmente, questi risultati confermano che tener conto della flessibilità molecolare e gli effetti di polarizzazione guidati dalla carica migliora sensibilmente l'accuratezza dei dati ottenuti.

*Analisi del binding mode*

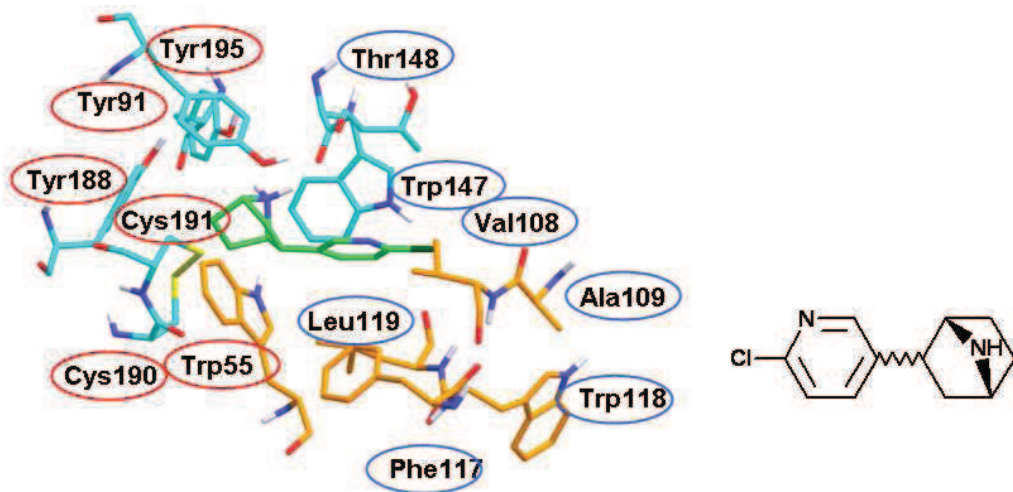
I risultati del docking QM-polarized mettono in evidenza due zone d'interazione all'interno del binding site. La prima (di seguito indicata come tasca recettoriale  $\alpha$ ) è formata da una serie di residui aromatici ( $\alpha$ 4Tyr91,  $\alpha$ 4Trp147,  $\alpha$ 4Tyr188,  $\alpha$ 4Tyr195,  $\beta$ 2Trp55) ed è coinvolta nella stabilizzazione della testa cationica dei ligandi attraverso interazioni elettrostatiche e  $\pi$ -catione, rinforzate dallo stabilirsi di uno o due legami ad idrogeno.

La seconda regione (di seguito indicata come tasca recettoriale  $\beta$ ) , è delimitata dai residui  $\alpha$ 4Thr148,  $\alpha$ 4Cys190,  $\alpha$ 4Cys191,  $\beta$ 2Ala108,  $\beta$ 2Val109,  $\beta$ 2Phe117,  $\beta$ 2Trp118 e  $\beta$ 2Leu119, e costituisce una porzione idrofobica nella quale si posiziona la porzione non carica del ligande.



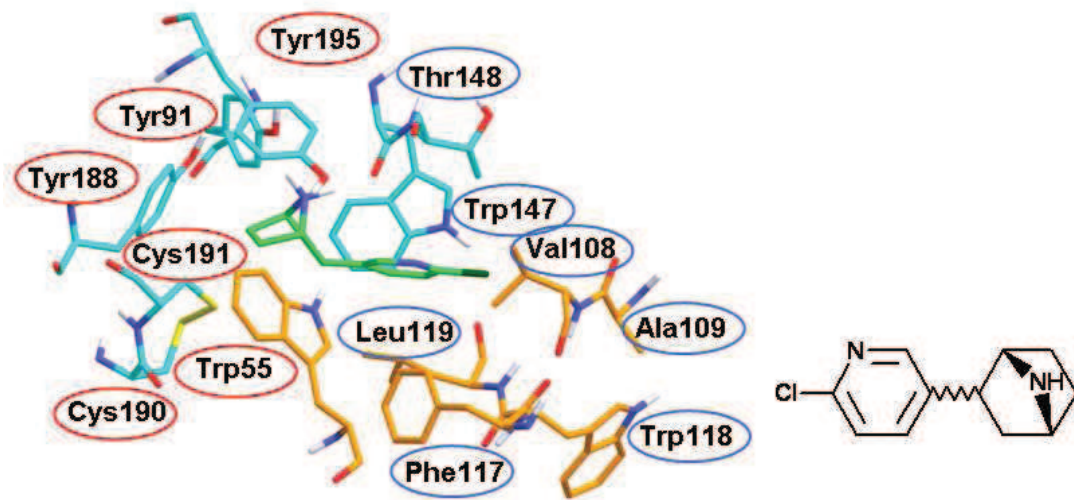
**Figura 9.** Sito di binding per il recettore nicotinic neuronale  $\alpha$ 4 $\beta$ 2. In rosso e blu le parentesi con indicazione delle tasche recettoriali  $\alpha$  e  $\beta$ , rispettivamente. La subunità  $\alpha$ 4 è colorata in blu, mentre la subunità  $\beta$ 2 in arancione. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

Nel complesso che il target recettoriale forma con la (-) epibatidina (Figura 1, composto 1), il gruppo amminico carico positivamente interagisce, tramite un legame ad idrogeno con l'ossigeno carbonilico del residuo  $\alpha$ 4Trp147 (distanza N-H...O, 1.82 Å (XP) o 1.79 Å (SP)) e con il gruppo idrossilico di  $\alpha$ 4Tyr91 (distanza N-H...O, 3.2 Å (XP) o 2.66 (SP)). La testa cationica del ligande è anche coinvolta in interazioni di tipo  $\pi$ -catione con gli anelli aromatici della tasca recettoriale  $\alpha$  che si trovano entro un raggio 4 Å. Infine il gruppo 6-cloro-piridile trova spazio nella tasca recettoriale  $\beta$ , anche se non la riempie completamente.



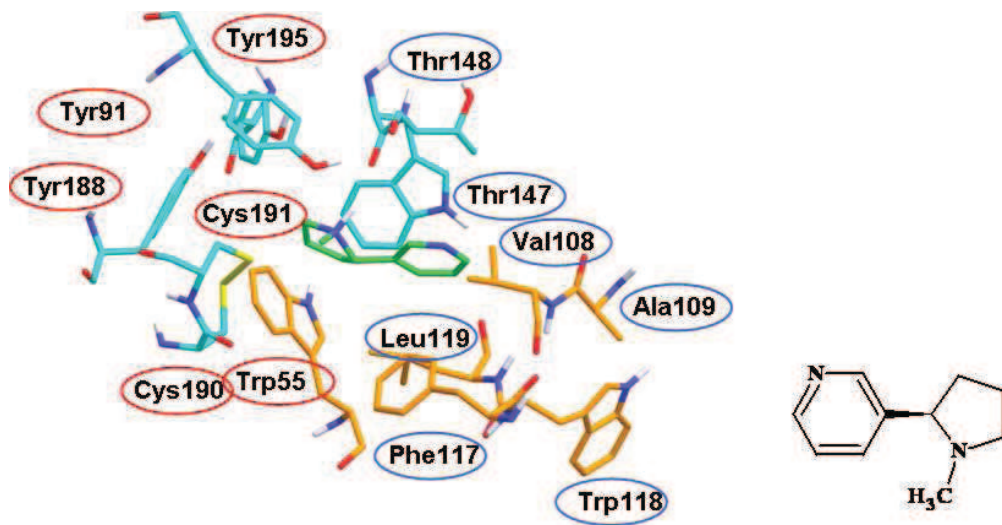
**Figura 10.** Risultati del docking QM-SP per (-)-epibatidina. La catena  $\alpha$ 4 è colorata in azzurro, mentre la catena  $\beta$ 2 in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

In accordo al simile dato di attività mostrato dai due enantiomeri, anche la (+)-epibatidina (Figura 1, composto 1) dà luogo ad interazioni analoghe con il recettore, posizionando la testa cationica all'interno della tasca recettoriale  $\alpha$ , formando un legame ad idrogeno con il gruppo carbonilico di  $\alpha$ 4Trp147 (distanza N-H...O, 1.84 Å (XP) o 1.67 Å (SP)). La molecola stabilisce interazioni di tipo NH... $\pi$  con  $\alpha$ 4Tyr195 (distanza NH... $\pi$  3.13 Å (XP) 2.75 Å (SP)). Analogamente a quanto avviene per la (-) epibatidina la porzione cloro-piridilica è posizionata all'interno della tasca recettoriale  $\beta$ .



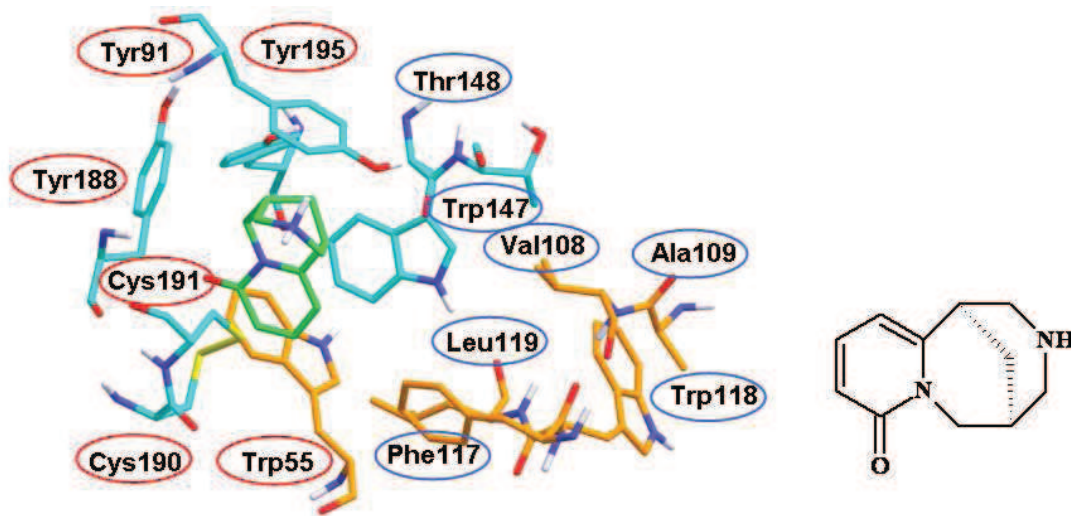
**Figura 11.** Risultati del docking QM-SP per (+)-epibatidina. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

La testa cationica della (S)-nicotina (Figura 1, composto 3) non interagisce con l'anello indolico di  $\alpha$ 4Trp147 (distanza fra la testa cationica e l'anello 5.3 Å (XP) o 5.05 Å (SP)), ma stabilisce un'interazione di legame ad idrogeno con il gruppo carbonilico di questo amminoacido (distanza N-H...O 2.26 Å (XP) o 2.04 Å (SP)). Questa osservazione è in accordo con i dati di mutagenesi sito diretta pubblicati da Cashin et al.,<sup>19</sup> che hanno evidenziato l'importanza della formazione di un legame ad idrogeno con questo residuo per il binding dei ligandi epibatidina e nicotina. Inoltre, anche la differente distanza fra la testa cationica di (3) e l'anello indolico di  $\alpha$ 4Trp147 (4.35 Å (XP) o 4.02 Å (SP)) e dell'interazione fra la testa cationica di (1) e l'anello indolico di  $\alpha$ 4Trp147 (5.29 Å (XP) o Å 5.05 (SP)) è coerente con la differente forza dell'interazione  $\pi$ -catione descritta da Cashin et al.<sup>19</sup>



**Figura 12.** Risultati del docking QM-SP per S-(-)-nicotina. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

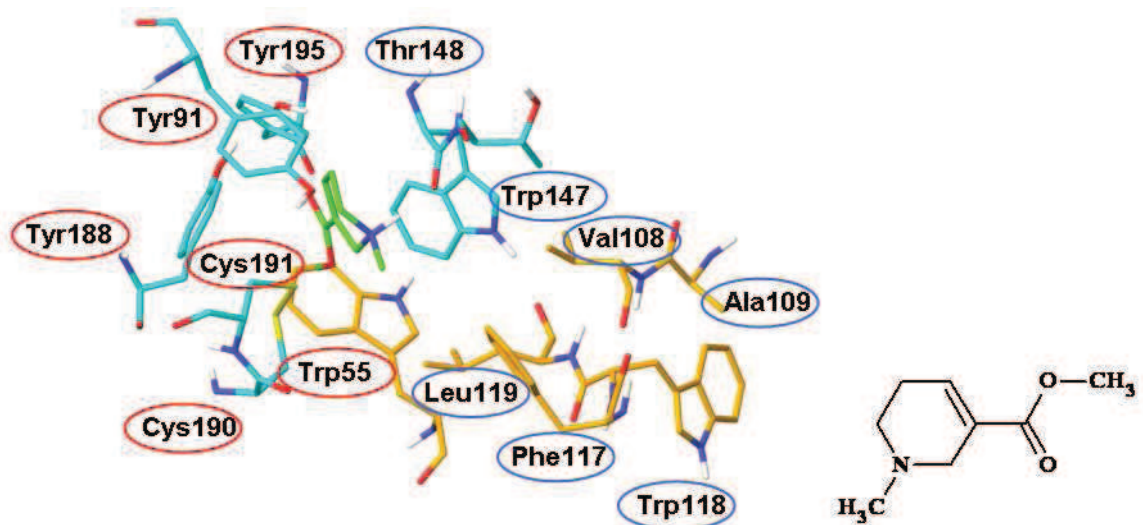
Il binding mode della citisina (Figura 1, composto 9) risulta abbastanza caratteristico. La molecola stabilisce infatti un'interazione di  $\pi$ - $\pi$  stacking con  $\alpha$ 4Trp147 (distanza  $\pi$ - $\pi$  4.37 Å (XP) 4.42 (SP), rinforzando l'interazione attraverso un legame  $\text{NH}\cdots\pi$  con  $\alpha$ 4Tyr195 (distanza  $\text{NH}\cdots\pi$  1.84 Å (XP) 2.64 Å (SP)). Diversamente da epibatidina e nicotina la porzione ciclica saturata della citisina non si posiziona nella tasca recettoriale  $\beta$ , bensì all'interno tasca recettoriale  $\alpha$ .



**Figura 13.** Risultati del docking QM-SP per la citisina. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

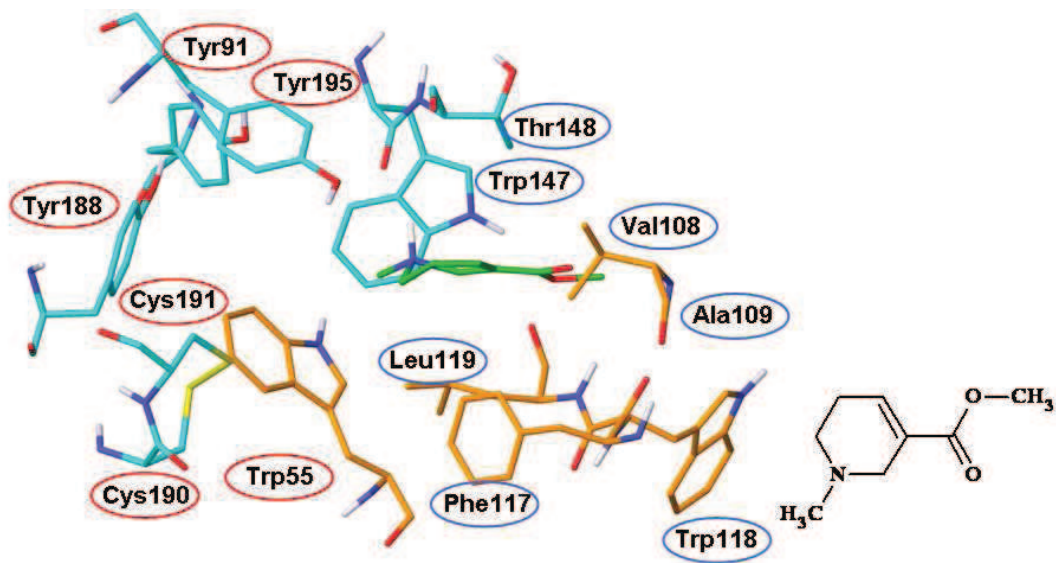
Per quanto riguarda l'arecolina (Figura 1, composto 10), sono state dockate ambedue le forme pseudo enantiomeriche che la molecola può assumere in soluzione, cioè quelle che originano dalla protonazione dell'atomo di N dell'anello tetraidropiridinico.

I due pseudo enantiomeri interagiscono in modo analogo localizzando la testa cationica in profondità all'interno del binding site e stabilendo un'interazione  $\pi$ -catione con  $\alpha$ 4Trp147 (distanza fra la testa cationica e l'anello aromatico 4.17 Å (XP) e 5.93 Å (SP) e 4.79 Å (XP) o 3.97 Å (SP) rispettivamente per gli pseudo enantiomeri R e S). Le diverse soluzioni, risultanti dal docking conducono a diverse orientazioni all'interno del binding site del metil-estere che tuttavia non si ripercuotono sullo score assegnato alle pose.



**Figura 14.** Risultati del docking QM-SP per R-arecolina. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.





Fi

**Figura 15.** Risultati del docking QM-SP per S-arecolina. La catena  $\alpha$  è colorata in azzurro, mentre la catena  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

Il carbacolo e il suo metil derivato (Figura 1, composti 4 e 8) essendo derivati ammoniaci quaternari, e come tali privi di un donatore di legame ad idrogeno a livello dell'atomo di N quaternario pur localizzando la testa cationica a livello della tasca recettoriale  $\alpha$ , non sono in grado di rinforzare l'interazione  $\pi$ -catione mediante la formazione di un legame ad idrogeno.

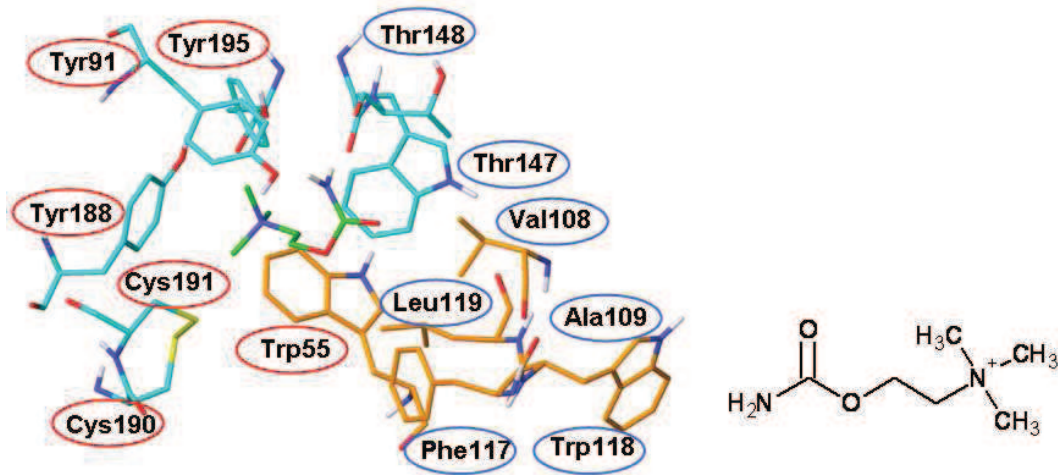
Nel caso del carbacolo le soluzioni risultanti dal docking in modalità XP e SP non sono convergenti. Infatti, la posa XP si stabilizza in una conformazione estesa che permette la formazione di un legame ad idrogeno con il gruppo idrossilico di  $\beta$ 2Ser235 (distanza N-H $\cdots$ O, 1.45 Å) della tasca recettoriale  $\beta$ .

Per contro nella soluzione SP è presente un legame ad idrogeno che si stabilisce con il gruppo carbonilico di  $\alpha$ 4Trp147 (distanza N-H $\cdots$ O, 1.32 Å) e il gruppo idrossilico di  $\alpha$ 4Tyr195 (distanza N-H $\cdots$ O, 1.66 Å).

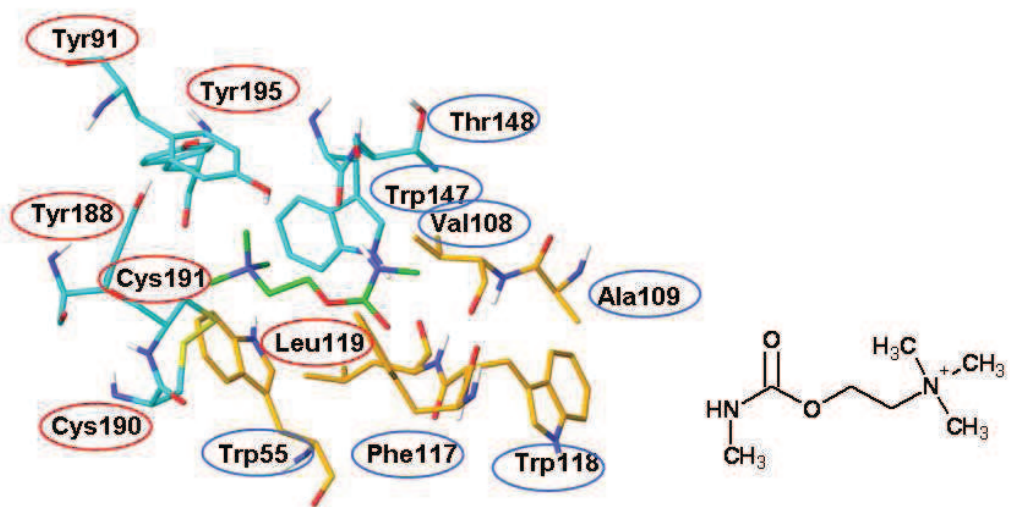
Tuttavia l'esclusione della presenza di legami ad idrogeno risultante dagli studi di mutagenesi sito diretta pubblicati da Dougherty e collaboratori<sup>19</sup> conferma la difficoltà dei software di docking, Glide incluso, nel proporre soluzioni convincenti per il docking di molecole non drug-like.

Il metil-carbacolo assume una conformazione simile a quella assunta dal carbacolo in modalità XP. La soluzione trovata è stabilizzata da un legame ad

idrogeno con l'ossigeno carbonilico di  $\alpha$ 4Trp147 (distanza N-H...O, 1.80 Å sia in modalità XP che SP).



**Figura 16.** Risultati del docking QM-SP per il carbacolo. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.



**Figura 17.** Risultati del docking QM-SP per il metilcarbacolo. La subunità  $\alpha$  è colorata in azzurro, mentre la subunità  $\beta$  in arancio. I residui della tasca recettoriale  $\alpha$  sono cerchiati in rosso quelli della  $\beta$  in blu.

In conclusione, i dati presentati dimostrano l'importanza, nel contesto dei recettori nicotinici, di combinare studi QM/MM e simulazioni MD. Tale accoppiamento non solo permette di considerare gli effetti di polarizzazione

del ligande nel contesto proteico, ma anche di catturare il comportamento dinamico del recettore.

Inoltre è importante notare che il protocollo proposto richiede un impegno computazionale non eccessivo (circa 22.5 minuti per ligande), tale da consentirne l'applicazione in un un processo di drug design.

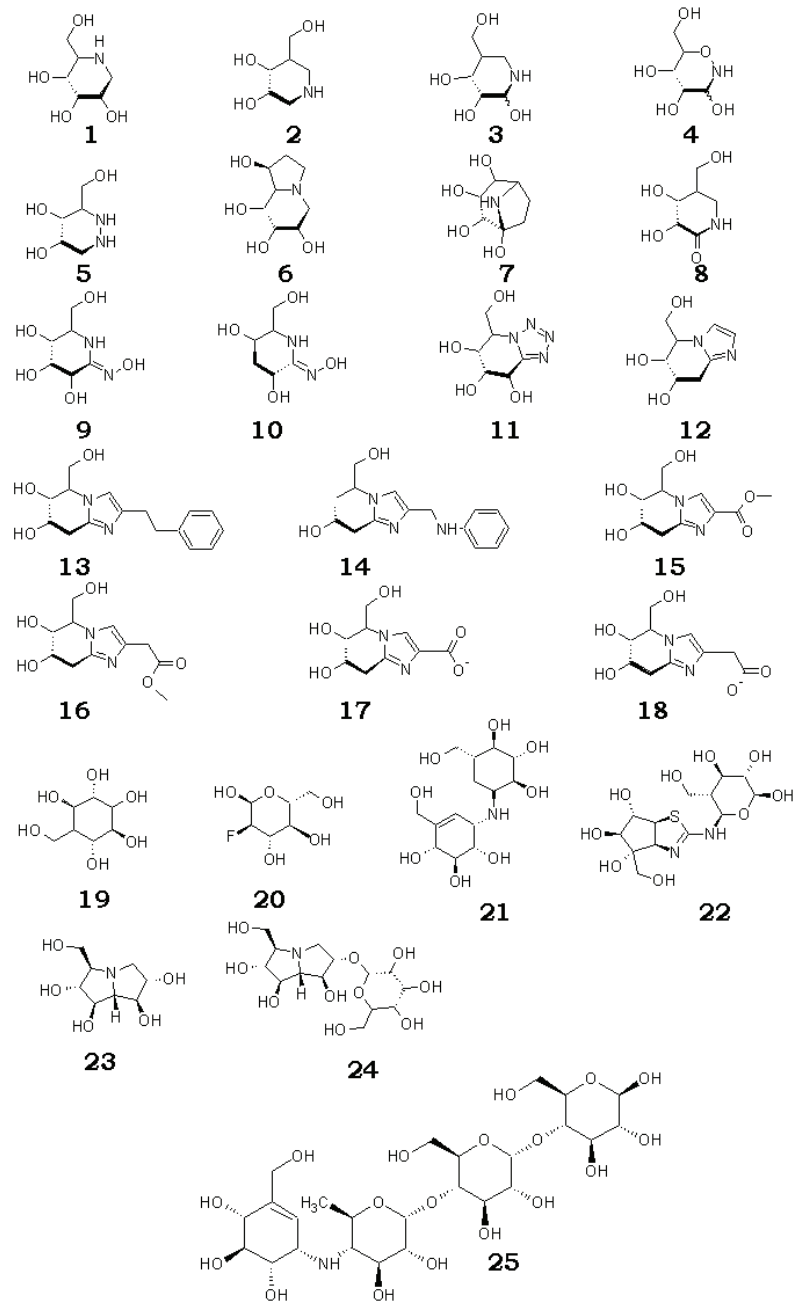
## Studi sulle glicosilidrolasi

L'idea di intraprendere lo studio di enzimi appartenenti alla classe delle glicosilidrolasi è originata da un soggiorno di studio della durata di sette mesi (giugno 2007-gennaio 2008) trascorso presso la Scuola Internazionale Superiore di Studi Avanzati (SISSA), settore di fisica statistica e biologica, sotto la guida del prof P. Carloni.

La disponibilità di dati termodinamici relativi a complessi unita alla presenza nel PDB delle strutture 3D risolte sperimentalmente rende questi enzimi dei target ideali per l'applicazione di procedure computazionali avanzate quali la Thermodynamic Integration (TI), la Linear Interaction Energy (LIE) e l' MM-PBSA/GBSA oggetto di ricerca durante lo stage presso la SISSA.

Fra i numerosi enzimi appartenenti a questa classe, abbiamo focalizzato l'attenzione su la beta-glucosidasi da *Thermotoga maritima*, la trealasi da *Escherichia coli* e la maltasi-glucoamilasi umana; tutti i ligandi oggetto dello studio appartengono alla classe degli iminozuccheri e, concordemente con i dati teorici<sup>92</sup> e sperimentali presenti in letteratura, sono stati modellati come molecole cariche positivamente.

Nello studio della beta-glucosidasi, le tecniche computazionali applicate al calcolo dell'energia libera di binding, sono state la Thermodynamic Integration (TI), l' MM-PBSA (Poisson Boltzmann Surface Area) e la Linear Interaction Energy.



**Figura 18 .** Inibitori delle GH utilizzati in questa tesi: **1** deoxynojirimicina, **2** isofagomina, **3** neuromicina, **4** tetraisossazina, **5** azafagomina, **6** castanospermina, **7** calistegina B<sub>2</sub>, **8** isofagomina lattame, **9** gluco-idroximolattame, **10** galatto-idroximolattame, **11** glucotetrazolo, **12** glucoimidazolo, **13** glucoimidazolo fenil sostituito, **14** glucoimidazolo fenilaminometil sostituito, **15** glucoimidazolo metossicarbonil sostituito, **16** glucoimidazolo metossicarbonilmetil sostituito, **17** glucoimidazolo carbossilato, **18** glucoimidazolo carbossimetil sostituito, **19** ciclofellitolo, **20** deossosfluoroglucosio, **21** validoxylamina A, **22** tiatreazolina, **23** casuarina, **24** casuarina 6- $\alpha$ -glucopiranosio, **25** acarbosio.

Per quanto riguarda gli enzimi trealasi e maltasi-glucoamilasi, l'approccio computazionale usato ha previsto:

- l'ottenimento, mediante docking molecolare delle strutture tridimensionali dei complessi ligando/proteina, nel caso in cui non fossero disponibili sul PDB;
- lo studio della stabilità del complesso mediante simulazioni di dinamica molecolare (MD);
- la determinazione dell'energia libera di binding mediante la Linear Interaction Energy;
- la determinazione mediante alanine scanning del contributo all'energia di binding degli aminoacidi della proteina.

#### *Studio della beta-glicosidasi*

La beta-glicosidasi è uno degli enzimi più studiati all'interno della classe delle glicosilidrolasi. Infatti, come già riportato nella parte introduttiva di questa tesi, esistono circa 18 strutture per questo enzima in complesso con diversi inibitori, risolte mediante diffrazione a raggi X, e delle quali è disponibile un'analisi termodinamica delle energie di binding.<sup>29</sup>

Questa disponibilità rende la beta-glicosidasi un ideale sistema di test per la valutazione e la messa a punto di metodi per il calcolo dell'energia libera di binding che possano poi essere utilizzati nell'ambito della progettazione e ottimizzazione di nuovi ligandi.

Gli approcci computazionali applicati a questo target sono stati la Thermodynamic Integration (TI), la Linear Interaction Energy (LIE) e l'MM-PBSA.

*Thermodynamic Integration (TI)*

La TI, sebbene possa essere utilizzata per il calcolo del valore assoluto di  $\Delta G$ , è solitamente impiegata per calcolare la differenza fra i valori di  $\Delta G$  ( $\Delta\Delta G$ ) per due diversi ligandi. Questo calcolo, oltre a fornire un'indicazione sul ligando più affine all'interno della coppia oggetto di studio, è anche notevolmente più efficiente dal punto di vista computazionale rispetto al calcolo di un  $\Delta G$  assoluto.

Dato l'elevato costo computazionale di questa metodica, preliminare all'applicazione della TI alla beta-glicosidasi è stata la simulazione del processo di trasformazione da deoxynojirimicina a isofagomina (Figura 18, composti 1, 2), due inibitori dell'enzima che, a fronte di una notevole similarità strutturale presentano un'elevata differenza di affinità.

Il calcolo del  $\Delta\Delta G$  è stato affrontato utilizzando sia un approccio double topology sia un approccio single topology.

L'approccio double topology, descritto nel capitolo 2, sia dal punto di vista operativo che da quello dei risultati è solitamente un'opzione meno conveniente del single topology

In questo approccio, infatti, la presenza di restrain di distanza necessari ad evitare l'allontanamento dal binding site del ligando rappresentato da dummy atom (cioè atomi privi di raggio e carica atomica), potrebbe apportare un contributo esterno al  $\Delta G$ .

Nel caso in studio il risultato ha evidenziato alcune differenze fra single e double topology, essendo il valore di  $\Delta\Delta G$  risultato pari a in 6.7 kcal/mol nel primo caso e 1.20 kcal/mol nel secondo.

Paragonato al valore di  $\Delta G$  sperimentale (-3.20 kcal/mol), l'errore commesso con l'approccio double topology non supera le 4 kcal/mol, in valore assoluto. I risultati ottenuti con entrambi gli approcci sono tuttavia di segno opposto al valore sperimentale e quindi non utili in un processo di drug design.

Una delle cause principali di predizione errata nei calcoli di TI è la mancata convergenza delle simulazioni. Allo scopo di controllare quest'aspetto le simulazioni double topology sono state prolungate fino a 3 ns per ogni

singolo valore del parametro di accoppiamento  $\lambda$  senza ottenere sostanziali variazioni del valore misurato; inoltre in ogni simulazione, ai diversi valori di  $\lambda$ , la convergenza dei valori di  $\delta G/\delta\lambda$  è stata controllata calcolando la media a blocchi.

Allo stato attuale del lavoro la causa più probabile di errore è riconducibile a una carente descrizione del ligando data dal force field Gaff. Inoltre è importante sottolineare, come ben evidenziato da Woo e Roux,<sup>93</sup> che un'ulteriore causa di errore potrebbe risiedere nella difficile stima dell'energia di idratazione di molecole cariche.

#### *Linear Interaction Energy (LIE)*

Viste le difficoltà incontrate nell'utilizzo della TI, abbiamo pensato di esplorare le opportunità offerte dai metodi end point per il calcolo dell'energia libera di binding.

Nello specifico, il metodo della LIE è stato applicato a complessi fra la beta-glicosidasi e sei diversi inibitori: deoxynojirimicina, isofagomina, neuromicina, castanospermina, isofagomina lattame, glucoimidazolo fenil sostituito (Figura 18, composti 1, 2, 3, 6, 8, 13).

Dal punto di vista computazionale l'impegno delle risorse di calcolo richiesto da questi metodi è abbastanza modesto, richiedendo due traiettorie di dinamica molecolare, una per il complesso ligando proteina e una per il ligando, immersi in un box di solvente esplicito.

Nel caso della beta-glicosidasi, visto che l'analisi delle traiettorie evidenziava una sostanziale convergenza delle strutture dei complessi già dopo il primo nano secondo, le simulazioni sono state arrestate a 6 ns.

Inizialmente, allo scopo di testare le performance di questo approccio computazionale abbiamo utilizzato i parametri proposti da Åqvist e collaboratori<sup>75</sup> ( $\alpha = 0.181$  e  $\beta = 0.50$  quando il ligando è carico o 0.33 nel caso di ligandi non carichi); successivamente, allo scopo di verificare se le prestazioni del metodo potessero essere migliorate, abbiamo utilizzato anche



i parametri proposti da Jorgensen e collaboratori<sup>76</sup> ( $\alpha = 0.476$  e  $\beta = 0.165$ ), e da Wall e collaboratori ( $\alpha = 0.418$  e  $\beta = 0.087$ ).<sup>77</sup>

**Tabella 3.** Predizioni del valore di  $\Delta G$  per i vari complessi con la beta-glicosidasi, la numerazione delle molecole si riferisce a quella della Figura 18.

<b>Molecola</b>	<b><math>\Delta G_{\text{sperimentale}}</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math> (Aqvist)</b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math> (Jorgensen)</b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math> (Wall)</b> kcal/mol
<b>1</b>	<b>-6.67</b>	-15.37	-8.61	-5.96
<b>2</b>	<b>-9.98</b>	-14.07	-7.21	-6.31
<b>3</b>	<b>-8.19</b>	-23.68	-10.64	-6.74
<b>6</b>	<b>-7.74</b>	-21.35	-10.65	-7.06
<b>8</b>	<b>-8.92</b>	-3.79	-5.07	-4.41
<b>13</b>	<b>-10.93</b>	-20.97	-13.28	-9.55
<b>Errore</b>		9.51	2.71	2.07

Come è possibile osservare dai dati riportati in Tabella 3, le predizioni ottenute nelle diverse condizioni non sono particolarmente incoraggianti.

Infatti in tutti i casi il coefficiente di correlazione ( $R^2$ ) fra il dato elaborato e quello sperimentale è inferiore a 0.25.

L'uso dei coefficienti elaborati di Wall et al fornisce il dato più promettente. Si riesce infatti a discriminare il ligando più attivo da quello meno attivo anche se nessuno dei set di parametri utilizzati permette di individuare il corretto rank di affinità. Ciò può dipendere dal ristretto intervallo di affinità dei ligandi, range dello stesso ordine dell'errore attribuibile al Force Field in simulazioni di dinamica molecolare.

#### *MM-PBSA/GBSA*

L'ultimo approccio che abbiamo utilizzato per calcolare l'energia libera di binding relativa alla formazione dei complessi con la beta-glicosidasi, è stato l' MM-PBSA/GBSA.

Questo approccio utilizza alcuni snapshot, ottenuti da un'unica traiettoria del complesso ligando-enzima, per calcolare il contributo entalpico ed entropico all'energia libera di binding.

La metodologia è stata applicata a tutti i complessi utilizzati per la Linear Interaction Energy, ad eccezione del complesso con la neuromicina (Figura 17, composto 3).

L'ottimizzazione delle condizioni sperimentali di analisi è avvenuta testando sia i modelli di solvente che utilizzano l'equazione di Poisson-Boltzmann (PBSA) per determinare la componente elettrostatica all'energia di solvatazione, sia quelli che utilizzano l'equazione Generalized Born (GBSA). Inoltre sono stati condotti test utilizzando i raggi atomici sia del FF Amber03, che del FF Parse (FF ottimizzato per l'uso di solventi impliciti).

Le costanti dielettriche del soluto e del solvente sono state settate rispettivamente a 1 e 80, dopo aver verificato che valori diversi portavano sempre a un decremento delle predizioni.

**Tabella 4.** Predizioni del valore di  $\Delta G$  per i vari complessi con la beta-glicosidasi utilizzando il FF Amber03 e il modello di solvente PBSA. La numerazione delle molecole si riferisce a quella della Figura 18.

<b>Molecola</b>	<b><math>\Delta G_{\text{sperimentale}}</math></b> kcal/mol	<b><math>\Delta H_{\text{calc}}</math></b> kcal/mol	<b><math>T\Delta S_{\text{calc}}</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math></b> kcal/mol
<b>1</b>	<b>-6.67</b>	-29.99	-20.06	-9.93
<b>2</b>	<b>-9.98</b>	-40.54	-15.34	-25.15
<b>6</b>	<b>-7.74</b>	-47.46	-18.96	-28.5
<b>8</b>	<b>-8.92</b>	26.72	-16.02	42.74
<b>13</b>	<b>-10.93</b>	-42.51	-20.78	-21.73

$R^2=0.007$

**Tabella 5.** Predizioni del valore di  $\Delta G$  per i vari complessi con la beta-glicosidasi utilizzando il FF Amber03 e il modello di solvente GBSA. La numerazione delle molecole si riferisce a quella della Figura 18.

<b>Molecola</b>	<b><math>\Delta G_{\text{sperimentale}}</math></b> kcal/mol	<b><math>\Delta H_{\text{calc}}</math></b> kcal/mol	<b><math>T\Delta S_{\text{calc}}</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math></b> kcal/mol
<b>1</b>	<b>-6.67</b>	-27.04	-20.06	-6.98
<b>2</b>	<b>-9.98</b>	-38.28	-15.34	-22.24
<b>6</b>	<b>-7.74</b>	-48.97	-18.96	-30.01
<b>8</b>	<b>-8.92</b>	-17.40	-16.02	-1.38
<b>13</b>	<b>-10.93</b>	-54.86	-20.78	-34.08

$$R^2=0.24$$

**Tabella 6.** Predizioni del valore di  $\Delta G$  per i vari complessi con la beta-glicosidasi utilizzando il FF Amber03 e raggi PARSE. La numerazione delle molecole si riferisce a quella della Figura 18.

<b>Molecola</b>	<b><math>\Delta G_{\text{sperimentale}}</math></b> kcal/mol	<b><math>\Delta H_{\text{calc}}</math></b> kcal/mol	<b><math>T\Delta S_{\text{calc}}</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}}</math></b> kcal/mol
<b>1</b>	<b>-6.67</b>	-12.06	-20.06	-7.99
<b>2</b>	<b>-9.98</b>	-27.66	-15.34	-12.32
<b>6</b>	<b>-7.74</b>	-32.79	-18.96	-13.83
<b>8</b>	<b>-8.92</b>	-0.51	-16.02	15.51
<b>13</b>	<b>-10.93</b>	-26.63	-20.78	-6.63

$$R^2=0.005$$

I risultati ottenuti usando le suindicate condizioni sperimentali (Tabelle 4, 5, 6) mostrano una scarsa correlazione tra i dati calcolati e quelli sperimentali.

Le migliori prestazioni si ottengono utilizzando i raggi Parse che quindi possono rappresentare un buon punto di partenza per future indagini tese a migliorare le performance del metodo nel predire l'affinità di iminozuccheri inibitori delle glicosilidrolasi.

### *Studio della trealasi*

La trealasi, la cui risoluzione strutturale è stata ottenuta da Gibson et al.<sup>36</sup> mediante diffrazione a raggi X, scinde i legami glucosidici del trealosio, polisaccaride implicato nel metabolismo degli insetti. Inibitori di questo enzima hanno dunque un potenziale come insetticidi.

In quest'ottica la comprensione dei determinanti molecolari che regolano l'interazione con questo enzima e la valutazione dell'energia libera di binding di inibitori enzimatici, assume un significato di rilievo.

I complessi inibitore/trealasi usati in questo studio sono stati:

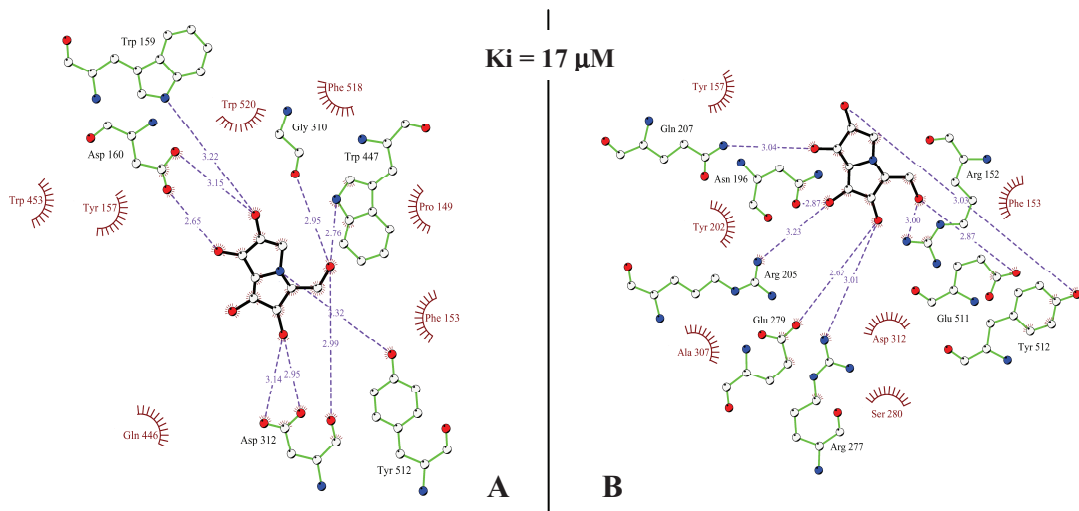
- 2JF4<sup>36</sup>, trealasi in complesso con validoxylamina A, ( $K_d$  10 nM, risoluzione 2.20 Å);
- 2JG0<sup>36</sup>, trealasi in complesso con tiatreazolina, ( $K_d$  9 nM, risoluzione 1.50 Å);
- Trealasi/casuarina-6-O-alfa-glucoside, complesso disponibile grazie ad una collaborazione con il gruppo del prof. G. Davies dell'università di York (UK).

Fra gli inibitori di questo enzima è stata inoltre considerata anche la molecola della casuarina, composto lead del gruppo del prof. A. Goti del dipartimento di Chimica Organica dell'Università di Firenze con il quale è da tempo attiva una collaborazione.

Tutti i complessi trealasi/ligande studiati sono stati sottoposti a simulazioni di dinamica molecolare a partire dalle coordinate atomiche dei complessi depositati nel PDB. Le coordinate del complesso trealasi/casuarina non risolto sperimentalmente, sono state ottenute tramite docking molecolare.

In particolare le soluzioni proposte dall'algoritmo di docking si clusterizzano sostanzialmente in due pose alternative (**Figura 19**).

La prima vede il ligando disporsi nella porzione più interna del sito di binding, detta anche -1, mentre la seconda vede il ligando occupare la porzione più esterna, sito +1.



**Figura 19.** Ligplot delle orientazioni risultanti dal docking della casuarina nei siti +1 (A) e -1 (B).

#### *Analisi delle simulazioni di MD*

Le simulazioni MD a cui sono stati sottoposti tutti i complessi trealasi/ligande hanno evidenziato una sostanziale stabilità sia del posizionamento del ligande che della struttura proteica.

L'analisi delle traiettorie di 2JF4 ha messo in luce la capacità della validoxylamina di formare mediamente 14 legami ad idrogeno con la proteina, l'assenza di molecole d'acqua in grado di coordinarsi e interagire stabilmente con il ligando ed il sostanziale mantenimento posizione iniziale e dei legami ad idrogeno formati dal ligande e descritti da Davies e collaboratori.<sup>58</sup>

Anche la tiatreazolina (pdb code 2JG0) mostra un comportamento analogo alla validoxylamina mantendo durante la traiettoria la conformazione risolta tramite Xray, formando un numero medio di 12 legami ad idrogeno e non coordinando in modo stabile alcuna molecola d'acqua.

Anche nel caso della casuarina-6-O-alfa-glucoside il complesso non sembra subire eccessive modificazioni strutturali durante le simulazioni di dinamica

molecolare riuscendo a mantenere tutte le interazioni che si osservano nel complesso cristallografico.

In questo caso il ligando è in grado di stabilire circa 12 legami ad idrogeno e non coordina permanentemente nessuna molecola d'acqua del solvente.

Per tutte le simulazioni effettuate, sfruttando la routine `g_energy` implementata in Gromacs è stato calcolato il contributo delle forze elettrostatiche e di vdW all'energia di binding (**Tabella 7**).

**Tabella 7.** Contributo elettrostatico e di vdW al  $\Delta G$  di binding per i vari complessi con la trealasi, la numerazione delle molecole si riferisce a quella della Figura 18.

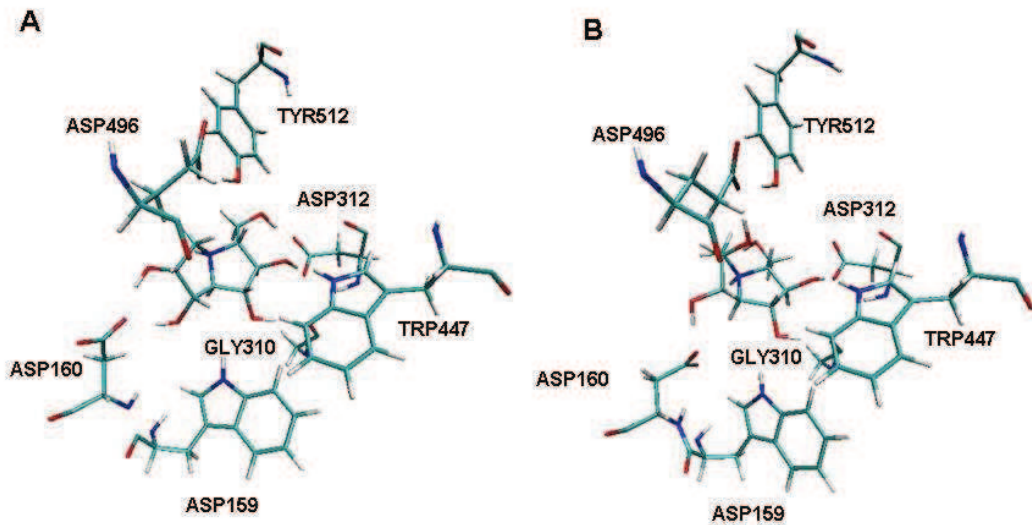
<b>Molecola</b>	<b><math>E_{\text{coul(SR)}}</math> Kcal/mol</b>	<b><math>E_{\text{vdW}}</math> Kcal/mol</b>
21	-135,53	-34.33
22	-134.01	-41.31
23 <sub>prima-posa</sub>	-90.71	-18.87
23 <sub>seconda-posa</sub>	-93.56	-17.53
24	-137,90	-33.02

Per quanto riguarda la casuarina ambedue le orientazioni prodotte dal docking sono state sottoposte a simulazioni di MD della durata 10.5 ns.

La posa che posiziona il ligando nel sito -1 (**Figura 20A**) si mantiene stabile per tutta la durata della simulazione, forma un totale di 8 legami ad idrogeno 6 con la proteina e 2 con molecole di solvente.

La posa che posiziona la casuarina nel sito +1 (**Figura 20B**), mostra invece un notevole riarrangiamento rispetto alla posizione che si ottiene dal docking, che la porta, alla fine della simulazione, a coincidere con la prima.

Anche in questo caso si formano 8 legami ad idrogeno, di cui 6 con la proteina e 2 con il solvente.



**Figura 20.** Struttura minimizzata dell'ultimo snapshot delle simulazione di MD condotte per il complesso trealasi-casuarina condotte partendo dalla prima posa (A) e dalla seconda posa (B).

Allo scopo di chiarire se il binding dei ligandi potesse provocare modificazioni conformazionali della proteina, mancando una struttura dell'enzima in forma apo, è stata eseguita una simulazione utilizzando la struttura cristallografica del complesso con la tiatetrazolina (pdb code 2JG0) privato del ligande.

L'ispezione visiva dei risultati di questa simulazione, oltre a non individuare importanti movimenti strutturali a carico dell catene laterali, ha evidenziato il fatto che circa 12 molecole d'acqua, di cui 6 localizzate nella parte più interna dell'enzima, in prossimità del residuo Asp160, possano essere spiazzate a seguito del legame dei ligandi con la proteina. Questo dato mette in rilievo il fatto che, ligandi come casuarina-6-O-alfa-glucoside, tiatetrazolina o la validoxylamina, formando un numero medio di circa 12 legami ad idrogeno, saturano in un modo quasi ottimale il network di possibili legami ad idrogeno instaurabili all'interno del binding site.

#### *Calcolo dell'energia libera utilizzando la LIE*

Le notevoli dimensioni dei complessi trealasi ligandi, che completamente solvatati contano più di 70000 atomi, suggeriscono l'uso di metodologie

semplici e veloci per la valutazione dell'affinità quali la Linear Interaction Energy.

La scarsità di dati biologici per inibitori ha reso impossibile, per questo enzima, procedere alla messa a punto di nuovi parametri ragione per cui sono stati utilizzati i set di parametri già testati sulla beta-glicosidasi. L'applicazione dei parametri di Åqvist ha permesso di calcolare la differenza di energia libera di binding ( $\Delta G$ ) con un errore di circa 1kcal/mol per tutte le molecole studiate (**Tabella 8**). Fa eccezione la validoxylamina (Figura 18, composto 21), per la quale l'errore commesso è circa 10 kcal/mol.

I migliori risultati si sono ottenuti con i coefficienti di Wall et al: quasi tutte le predizioni riproducono il risultato sperimentale con un errore inferiore alle 2 kcal/mol. Fa eccezione per la tiatreazolina (Figura 18, composto 22) il cui valore di energia libera di binding viene riprodotto con un errore di circa 3,65 kcal/mol.

**Tabella 8.** Predizioni del valore di  $\Delta G$  per i vari complessi con la trealasi, la numerazione delle molecole si riferisce a quella figura 18.

<b>Molecola</b>	<b><math>\Delta G_{\text{sperimentale}}</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}} (\text{Åqvist})</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}} (\text{Jorgensen})</math></b> kcal/mol	<b><math>\Delta G_{\text{calc}} (\text{Wall})</math></b> kcal/mol
<b>21</b>	<b>-11.1</b>	-21.51	-16.87	-12.80
<b>22</b>	<b>-10.6</b>	-9.26	-16.98	-14.75
<b>23</b> <sub>prima posa</sub>	<b>-6.50</b>	-5.38	-7.80	-6.48
<b>23</b> <sub>seconda posa</sub>	<b>-6.50</b>	-5.28	-8.30	-6.47
<b>24</b>	<b>-10.80</b>	-11.71	-13.67	-11.31
<b>Errore medio*</b>		3.67	4.08	1.34

\*l'errore medio è calcolato utilizzando il valore della prima posa della casuarina.

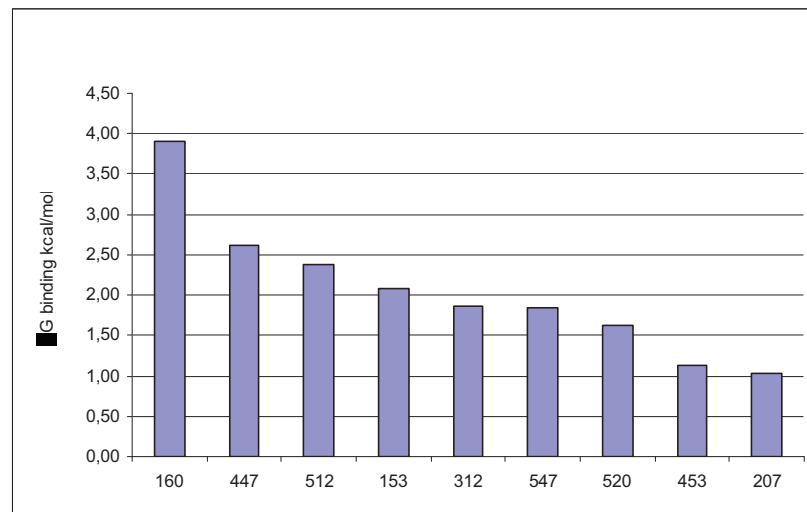
#### *Mutagenesi computazionale (Computational Alanine Scanning)*

La mutagenesi computazionale rappresenta una procedura tramite cui si calcola la differenza dell'energia libera di binding target/ligande che si origina per mutazione dei residui aminoacidici del target in residui di alanina. L'applicazione di questa procedura contribuisce a chiarire le interazioni importanti per il binding di ligandi al target e fornisce una stima

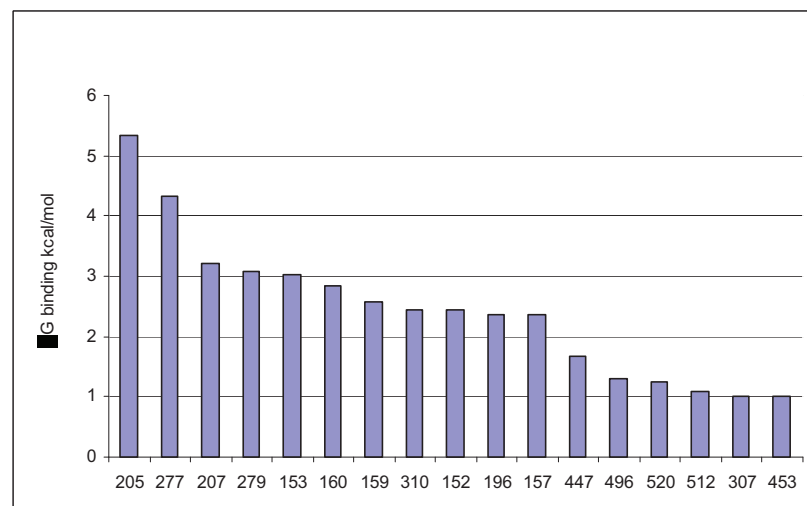


delle interazioni fra ligandi e proteina, così da poter identificare le driving force del binding.

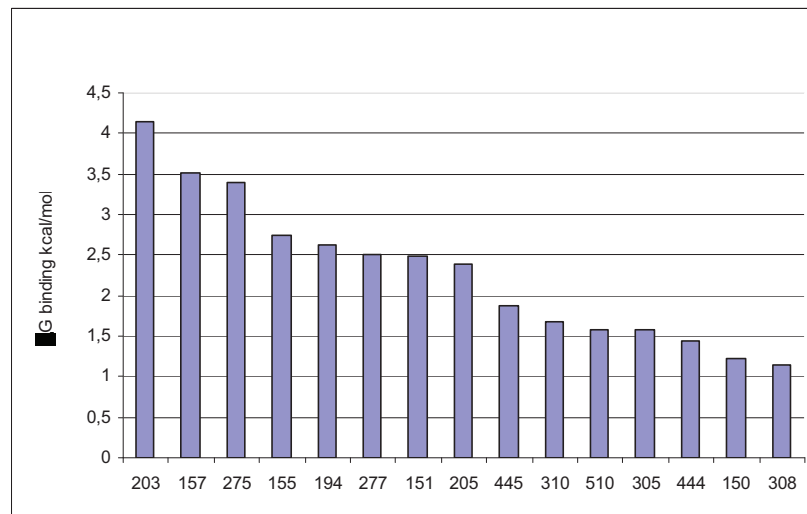
Le figure 21-24 mostrano il contributo al  $\Delta G$  di binding per ogni singolo aminoacido: le barre di maggior altezza indicano residui che contribuiscono di più all'energia di binding, è importante sottolineare l'assenza di valori negativi indice che tutte le mutazioni effettuate sono sfavorevoli al binding.



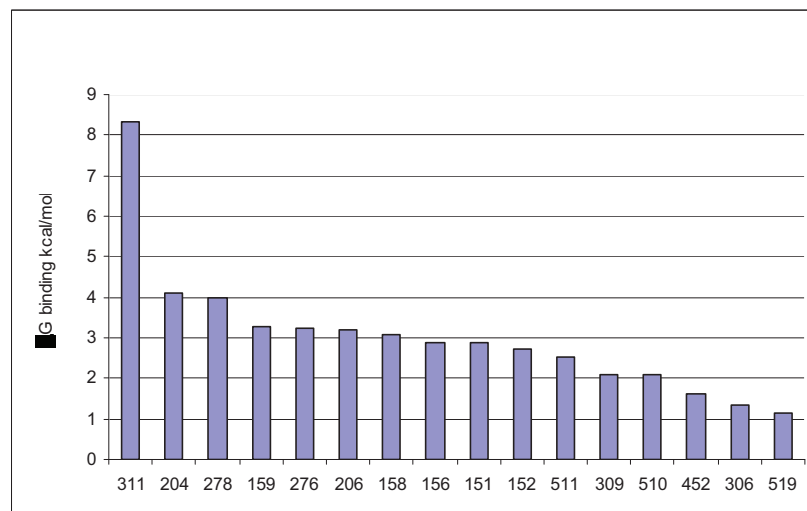
**Figura 21.** Contributo all'energia di binding per i vari aminoacidi nel primo complesso trealasi-casuarina. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.



**Figura 22.** Contributo all' energia di binding per i vari aminoacidi nel complesso trealasi-casuarina-6-O-alfa-glucoside. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.



**Figura 23.** Contributo all'energia di binding per i vari aminoacidi nel complesso trealasi-validoxylamina. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.



**Figura 24.** Contributo all'energia di binding per i vari aminoacidi nel complesso trealasi-tiatreazolina. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.

L'alanine scanning condotto sulla casuarina è stato eseguito sulla posa del sito -1 dato il riarrangiamento subito dalla posa del sito +1 messo in evidenza dagli esperimenti MD. Dall'osservazione della Figura 21 si evince come il maggior contributo al binding della casuarina è fornito dal residuo Asp160 e come anche i residui Trp447, Tyr512 e Phe153 giochino un ruolo molto importante.

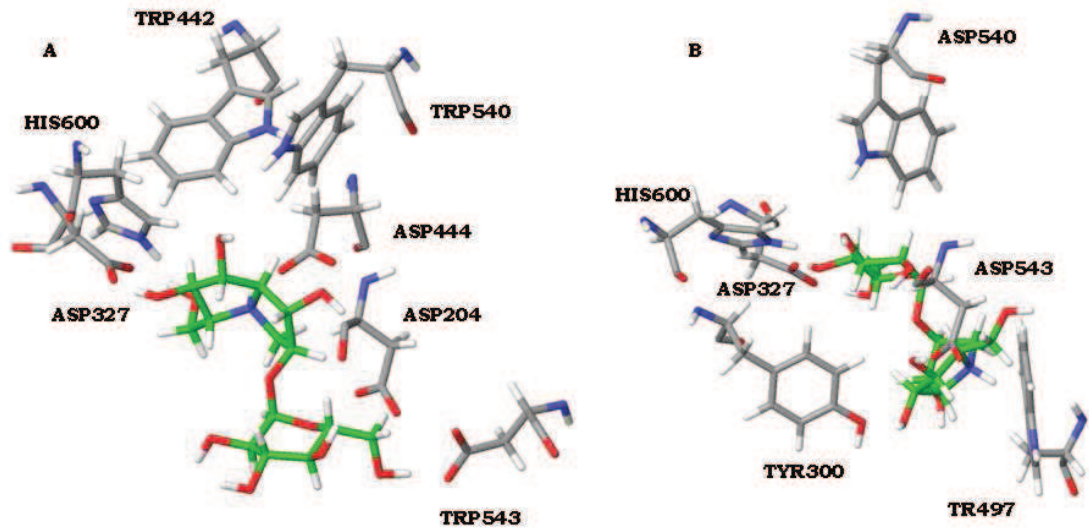
Tre dei primi quattro residui più importanti per il binding della casuarina-6-O-alfa-glucoside, Arg205 e Arg277 e Glu279, (**Figura 22**) sono localizzati nel sito -1 e interagiscono con la porzione glicosidica. Questo dato sottolinea l'importanza del sito -1 per il binding di inibitori.

L'alto valore che si registra per il residuo Asp311 nel complesso trealasi/tiatreazolina (**Figura 24**), questo è probabilmente dovuto alla formazione di un ponte salino con il ligande; infatti il valore misurato non si discosta molto da quello stimato da Gruia et al<sup>94</sup> per un ponte salino fra due aminoacidi che si trovino sulla superficie di una proteina.

#### *Studio della maltasi-glucoamilasi*

La maltasi-glucoamilasi è un enzima coinvolto nella degradazione dell'amido la cui struttura è stata recentemente risolta e depositata nel PDB (pdb code 2QMJ). L'importante ruolo fisiologico di questo enzima lo rende un interessante target per la cura del diabete. Anche per questo target lo studio delle interazioni fra inibitori e l'enzima è stato affrontato utilizzando simulazioni di dinamica molecolare. La scarsità dei dati strutturali relativi al target ha limitato le simulazioni ai complessi maltasi-glucoamilasi/casuarina e maltasi-glucoamilasi/casuarina-6-O-alfa-glucoside, partendo per quest'ultimo complesso dalle coordinate atomiche ottenute mediante docking.

Le soluzioni proposte dall'algoritmo di docking si raggruppano sostanzialmente in due pose. Nella prima la porzione agliconica della molecola si dispone all'interno del sito catalitico (**Figura 25A**), mentre nella seconda è la porzione casuarinica a disporsi nella parte più esterna dell'enzima (**Figura 25 B**).

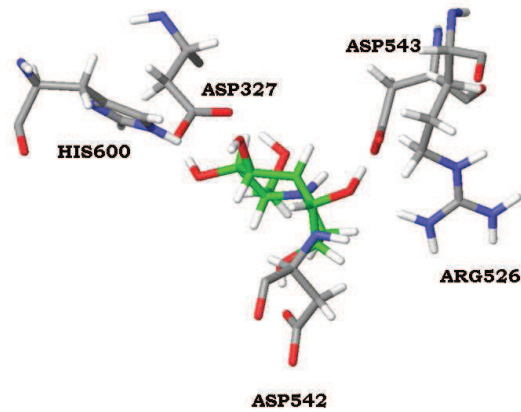


**Figura 25.** Struttura minimizzata dell'ultimo snapshot delle simulazione di MD condotte per il complesso MGAM/casuarina-6-O-alfa-glucoside, condotte partendo dalla prima posa (A) e dalla seconda posa (B).

#### *Analisi delle simulazioni di MD*

Le simulazioni di MD condotte a partire dalla struttura 3D del complesso maltasi-glucoamilasi/casuarina non hanno evidenziato alcun riarrangiamento strutturale del complesso, mantenendo la molecola di casuarina stabilmente posizionata nel sito catalitico (**Figura 26**).

L'analisi delle interazioni indica che la casuarina è in grado di formare circa 8 legami ad idrogeno di cui circa 6 con la proteina e 2 con molecole d'acqua del solvente.



**Figura 26.** Struttura minimizzata dell'ultimo snapshot delle simulazione di MD condotte per il complesso MGAM-casuarina.

Le simulazioni del complesso maltasi-glucoamilasi/casuarina-6-O-alfa-glucoside sono state condotte a partire dai due risultati del docking.

Le traiettorie relative alla prima posa, hanno confermato la sostanziale stabilità della struttura del complesso; il ligando è in grado di stabilire circa 12 legami ad idrogeno la maggior parte dei quali, circa 8, con molecole di solvente.

Anche le simulazioni effettuate sulla seconda posa hanno mostrato una sostanziale stabilità del complesso anche se rispetto alle coordinate iniziali la molecola di inibitore tende a ripiegarsi parzialmente durante la simulazione.

Il ligando stabilisce 14 legami ad idrogeno, di cui 11 con le molecole di solvente.

L'analisi dei contributi energetici al binding (**Tabella 9**), evidenzia come la maggior differenza fra la prima e la seconda posa della casuarina-6-O-alfa-glucoside sia dovuta al contributo elettrostatico.

**Tabella 9.** Contributo elettrostatico e di vdW al  $\Delta G$  di binding per i vari complessi con MGAM, la numerazione delle molecole si riferisce a quella della Figura 18.

Molecola	$E_{\text{coul(SR)}}$ Kcal/mol	$E_{\text{vdW}}$ Kcal/mol
23	-104.22	-14.01
24 <sub>prima-posa</sub>	-149.15	-20.58
24 <sub>seconda-posa</sub>	-140.22	-18.71

*Calcolo dell'energia libera utilizzando la LIE*

Come già nel caso della trealasi, le ragguardevoli dimensioni del target molecolare (125000 atomi) hanno indotto all'uso della metodica della LIE per il calcolo dell'energia libera di binding.

L'uso dei vari set di parametri suggeriti dalla letteratura (Aqvist, Jorgensen, Wall) ha condotto ai risultati riassunti in tabella 10.

**Tabella 10.** Predizioni del valore di  $\Delta G$  per i vari complessi con MGAM, la numerazione delle molecole si riferisca a quella della Figura 18.

Molecola	$\Delta G_{\text{sperimentale}}$ kcal/mol	$\Delta G_{\text{calc (Aqvist)}}$ kcal/mol	$\Delta G_{\text{calc (Jorgensen)}}$ kcal/mol	$\Delta G_{\text{calc (Wall)}}$ kcal/mol
23	<b>-8.65</b>	-8.65	-8.75	-6.21
24 <sub>prima-posa</sub>	<b>-4.64</b>	-7.16	-9.70	-7.26
24 <sub>seconda-posa</sub>	<b>-4.64</b>	-12.41	-7.33	-5.55

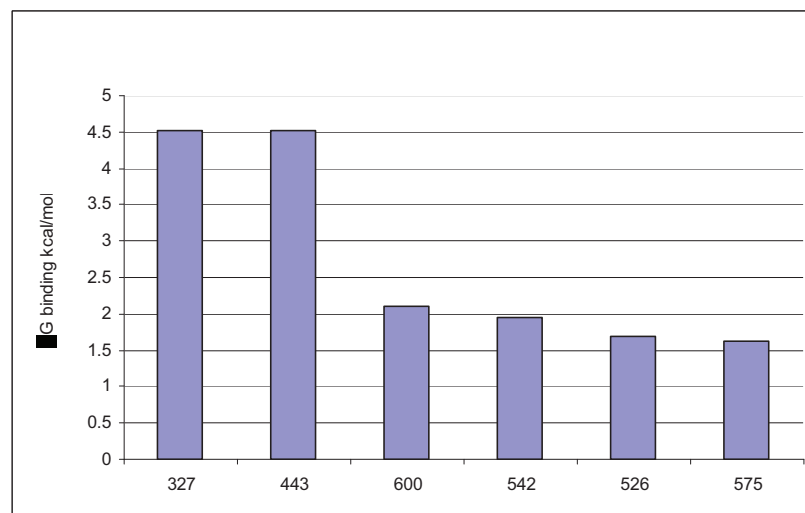
I coefficienti proposti da Wall et al<sup>77</sup> riproducono meglio degli altri i dati sperimentali. L'errore nella predizione dell'affinità della casuarina (Figura 18, composto 21) è infatti di circa 2.44 kcal/mol, mentre i risultati relativi alla casuarina-6-O-alfa-glucoside (Figura 18, composto 24), risultando abbastanza diversi nei valori di  $\Delta G$  le due pose, necessitano di alcune considerazioni. Il valore di  $\Delta G$  della prima posa (-7.26 kcal/mol), che posiziona la porzione agliconica nel sito più interno, è più negativo di quello sperimentale, e

addirittura più negativo di quello predetto per la casuarina (-6.21 kcal/mol). Il dato che si misura per la seconda posa (-5.55 kcal/mol) mantiene lo stesso trend osservato sperimentalmente rispetto alla casuarina ( $\Delta G_{\text{Wall}}$  -5.55 kcal/mol vs -6.21 kcal/mol ;  $\Delta G_{\text{exp}}$  -4.64 kcal/mol vs -8.65 kcal/mol) e si discosta da quello sperimentale per solo 1 kcal/mol.

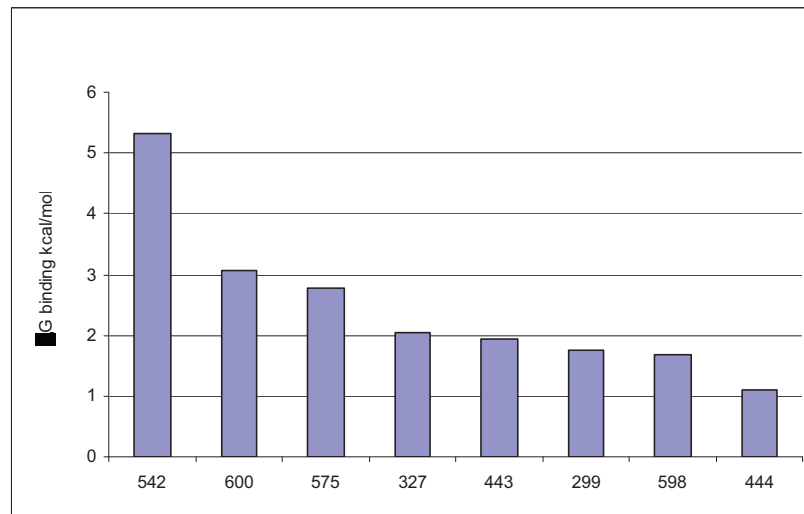
Ciò potrebbe ad ipotizzare che il glucoside interagisca con l'enzima nella seconda modalità, ossia con la porzione agliconica disposta nella parte più esterna.

#### *Computational Alanine Scanning*

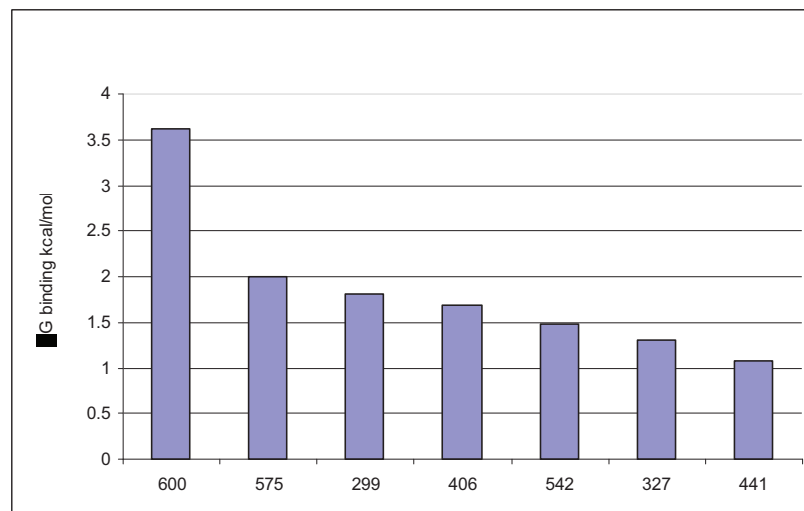
Le figure 27-29 mostrano il contributo all'energia libera di binding ed il grado di coinvolgimento dei residui aminoacidici della maltasi-glucoamilasi in complesso con la casuarina e la casuarina-6-O-alfa-glucoside.



**Figura 28.** Contributo all'energia di binding per i vari aminoacidi nel complesso maltasi-glucoamilasi/casuarina. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.



**Figura 29.** Contributo all' energia di binding per i vari aminoacidi della prima soluzione del docking maltasi-glucoamilasi/casuarina-6-O-alfa-glucoside. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.



**Figura 30.** Contributo all' energia di binding per i vari aminoacidi della seconda soluzione del docking maltasi-glucoamilasi/casuarina-6-O-alfa-glucoside. Sono riportati solo i residui la cui mutazione evidenzia un contributo maggiore di 1kcal/mol.

In entrambe le soluzioni provenienti dal docking della casuarina-6-O-alfa-glucoside i residui più coinvolti nel binding sono quelli del sito più interno, che rivestono un ruolo fondamentale anche per il binding delle casuarina His600, Asp542, Asp327.

Nel caso della casuarina, il maggior contributo al binding è fornito da due residui di aspartato, Asp327 e Asp443 e da un residuo di His600, che funge da donatore di legame ad idrogeno.



*Selettività trealasi/maltasi-glucoamilasi*

Il differente contributo all'energia di interazione dovuto alle forze di vdW, che emerge dall'analisi degli esperimenti di MD, potrebbe essere alla base dell'elevata selettività della casuarina-6-O-alfa glucoside per la trealasi.

Come è possibile osservare dai dati riportati in Tabella 7 e 9, la maggiore affinità della casuarina per la MGAM rispetto alla trealasi è da attribuirsi al più alto valore di energia dovuta all'interazione elettrostatica. Il contributo coulombiano in MGAM è maggiore di quello misurato nella trealasi di almeno 10 kcal/mol, mentre quello di vdW e il numero di legami ad idrogeno si mantengono di fatto paragonabili.

Gli esperimenti di mutagenesi computazionale dimostrano che nel caso della trealasi i siti +1 e -1 possiedono la stessa importanza, mentre in MGAM sembra esistere un solo sito di binding, osservazione confermata dal diverso contributo delle forze di vdW all'energia di interazione.

Queste osservazioni trovano conferma nel fatto che, sebbene negli esperimenti di MD le due pose della casuarina nella trealasi convergano ad un'unica soluzione, la maggiore affinità per l'enzima trealasi è mostrata da quei ligandi capaci di occupare entrambi i siti.

Una situazione opposta si verifica in MGAM infatti l'alanine scanning evidenzia il ruolo primario giocato dal sito -1. Inoltre la numerosità delle interazioni della casuarina-6-O-alfa-glucoside con le molecole di solvente, porta ad ipotizzare un contributo entropico sfavorevole al binding per quelle molecole in grado di occupare anche una porzione esterna dell'enzima.

Queste osservazioni potrebbero efficacemente spiegare la diminuita affinità per MGAM del glucoside rispetto alla casuarina.

**Bibliografia**

1. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-242.
2. Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van Der Oost, J.; Smit, A. B.; Sixma, T. K. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001**, *411*, 269-276.
3. Unwin, N. Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. *J. Mol. Biol.* **2005**, *346*, 967-989.
4. Bisson, W. H.; Scapozza, L.; Westera, G.; Mu, L.; Schubiger, P. A. Ligand selectivity for the acetylcholine binding site of the rat alpha4beta2 and alpha3beta4 nicotinic subtypes investigated by molecular docking. *J. Med. Chem.* **2005**, *48*, 5123-5130.
5. Grazioso, G.; Cavalli, A.; De Amici, M.; Recanatini, M.; De Micheli, C. Alpha7 nicotinic acetylcholine receptor agonists: prediction of their binding affinity through a molecular mechanics Poisson-Boltzmann surface area approach. *J. Comput. Chem.* **2008**, *29*, 2593-2602.
6. Huang, X.; Zheng, F.; Chen, X.; Crooks, P. A.; Dwoskin, L. P.; Zhan, C.-G. Modeling subtype-selective agonists binding with alpha4beta2 and alpha7 nicotinic acetylcholine receptors: effects of local binding and long-range electrostatic interactions. *J. Med. Chem.* **2006**, *49*, 7661-7674.

7. Huang, X.; Zheng, F.; Crooks, P. A.; Dwoskin, L. P.; Zhan, C.-G. Modeling multiple species of nicotine and deschloroepibatidine interacting with  $\alpha 4 \beta 2$  nicotinic acetylcholine receptor: from microscopic binding to phenomenological binding affinity. *J. Am. Chem. Soc.* **2005**, *127*, 14401-14414.
8. Le Novere, N.; Grutter, T.; Changeux, J.-P. Models of the extracellular domain of the nicotinic receptors and of agonist- and  $\text{Ca}^{2+}$ -binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 3210-3215.
9. Celie, P. H. N.; van Rossum-Fikkert, S. E.; van Dijk, W. J.; Brejc, K.; Smit, A. B.; Sixma, T. K. Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **2004**, *41*, 907-914.
10. Celie, P. H. N.; Klaassen, R. V.; van Rossum-Fikkert, S. E.; van Elk, R.; van Nierop, P.; Smit, A. B.; Sixma, T. K. Crystal structure of acetylcholine-binding protein from *Bulinus truncatus* reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors. *J. Biol. Chem.* **2005**, *280*, 26457-26466.
11. Celie, P. H. N.; Kasheverov, I. E.; Mordvintsev, D. Y.; Hogg, R. C.; van Nierop, P.; van Elk, R.; van Rossum-Fikkert, S. E.; Zhmak, M. N.; Bertrand, D.; Tsetlin, V.; Sixma, T. K.; Smit, A. B. Crystal structure of nicotinic acetylcholine receptor homolog AChBP in complex with an alpha-conotoxin PnIA variant. *Nat. Struct. Mol. Biol.* **2005**, *12*, 582-588.
12. Hansen, S. B.; Sulzenbacher, G.; Huxford, T.; Marchot, P.; Taylor, P.; Bourne, Y. Structures of *Aplysia* AChBP complexes with nicotinic agonists

and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **2005**, *24*, 3635-3646.

13. Bourne, Y.; Talley, T. T.; Hansen, S. B.; Taylor, P.; Marchot, P. Crystal structure of a Cbtx-AChBP complex reveals essential interactions between snake alpha-neurotoxins and nicotinic receptors. *EMBO J.* **2005**, *24*, 1512-1522.

14. Ulens, C.; Hogg, R. C.; Celie, P. H.; Bertrand, D.; Tsetlin, V.; Smit, A. B.; Sixma, T. K. Structural determinants of selective  $\hat{\pm}$ -conotoxin binding to a nicotinic acetylcholine receptor homolog AChBP. *Proc Natl Acad Sci U S A.* **2006**, *103*, 3615-3620.

15. Hansen, S. B.; Taylor, P. Galanthamine and non-competitive inhibitor binding to ACh-binding protein: evidence for a binding site on non-alpha-subunit interfaces of heteromeric neuronal nicotinic receptors. *J. Mol Biol.* **2007**, *369*, 895-901.

16. Dutertre, S.; Ulens, C.; Buttner, R.; Fish, A.; van Elk, R.; Kendel, Y.; Hopping, G.; Alewood, P. F.; Schroeder, C.; Nicke, A.; Smit, A. B.; Sixma, T. K.; Lewis, R. J. AChBP-targeted alpha-conotoxin correlates distinct binding orientations with nAChR subtype selectivity. *EMBO J.* **2007**, *26*, 3858-3867.

17. Ihara, M.; Okajima, T.; Yamashita, A.; Oda, T.; Hirata, K.; Nishiwaki, H.; Morimoto, T.; Akamatsu, M.; Ashikawa, Y.; Kuroda, S. i.; Mega, R.; Kuramitsu, S.; Sattelle, D. B.; Matsuda, K. Crystal structures of *Lymnaea stagnalis* AChBP in complex with neonicotinoid insecticides imidacloprid and clothianidin. *Invert. Neurosci.* **2008**, *8*, 71-81.

18. Talley, T. T.; Harel, M.; Hibbs, R. E.; Radic, Z.; Tomizawa, M.; Casida, J. E.; Taylor, P. Atomic interactions of neonicotinoid agonists with AChBP: molecular recognition of the distinctive electronegative pharmacophore. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7606-7611.
19. Cashin, A. L.; Petersson, E. J.; Lester, H. A.; Dougherty, D. A. Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **2005**, *127*, 350-356.
20. Gotti, C.; Riganti, L.; Vailati, S.; Clementi, F. Brain neuronal nicotinic receptors as new targets for drug discovery. *Curr. Pharm. Des.* **2006**, *12*, 407-428.
21. Romanelli, M. N.; Gratteri, P.; Guandalini, L.; Martini, E.; Bonaccini, C.; Gualtieri, F. Central Nicotinic Receptors: Structure, Function, Ligands, and Therapeutic Potential. *ChemMedChem* **2007**, *2*, 746-767.
22. Costa, V.; Nistri, A.; Cavalli, A.; Carloni, P. A structural model of agonist binding to the alpha3beta4 neuronal nicotinic receptor. *Br J Pharmacol.* **2003**, *140*, 921-931.
23. Glennon, R. A.; Dukat, M. alpha4beta2 nACh receptor pharmacophore models. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1841-1844.
24. Vocadlo, D. J.; Davies, G. J. Mechanistic insights into glycosidase chemistry. In press, *Curr. Opin. Chem. Biol.* **2008**.
25. Asano, N. Glycosidase inhibitors: update and perspectives on practical use. *Glycobiology* **2003**, *13*, 104.

26. Lopez-Casado, G.; Urbanowicz, B. R.; Damasceno, C. M. B.; Rose, J. K. C. Plant glycosyl hydrolases and biofuels: a natural marriage. *Curr. Opin. Plant. Biol.* **2008**, *11*, 329-337.
27. Chioccioli, M. Alla ricerca di nuovi inibitori per Glucoamilasi e Golgi alpha-Mannosidasi: messa a punto delle procedure di molecular modelling e progettazione di ligandi. Tesi di Laurea Specialistica in Chimica, Università di Firenze, 2007.
28. Lillelund, V. H.; Jensen, H. H.; Liang, X.; Bols, M. Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin. *Chem. Rev.* **2002**, *102*, 515-553.
29. Gloster, T. M.; Meloncelli, P.; Stick, R. V.; Zechel, D.; Vasella, A.; Davies, G. J. Glycosidase inhibition: an assessment of the binding of 18 putative transition-state mimics. *J. Am. Chem. Soc.* **2007**, *129*, 2345-2354.
30. Gloster, T. M.; Roberts, S.; Perugino, G.; Rossi, M.; Moracci, M.; Panday, N.; Terinek, M.; Vasella, A.; Davies, G. J. Structural, Kinetic, and Thermodynamic Analysis of Glucoimidazole-Derived Glycosidase Inhibitors. *Biochemistry* **2006**, *45*, 11879-11884.
31. Vincent, F.; Gloster, T. M.; Macdonald, J.; Morland, C.; Stick, R. V.; Dias, F. M. V.; Prates, J. A. M.; Fontes, C. M. G. A.; Gilbert, H. J.; Davies, G. J. Common inhibition of both beta-glucosidases and beta-mannosidases by isofagomine lactam reflects different conformational itineraries for pyranoside hydrolysis. *ChemBiochem* **2004**, *5*, 1596-1599.
32. Zechel, D. L.; Boraston, A. B.; Gloster, T.; Boraston, C. M.; Macdonald, J. M.; Tilbrook, D. M. G.; Stick, R. V.; Davies, G. J. Iminosugar

Glycosidase Inhibitors: Structural and Thermodynamic Dissection of the Binding of Isofagomine and 1-Deoxynojirimycin to beta-glucosidases. *J. Am. Chem. Soc.* **2003**, 125, 14313-14323.

33. Gloster, T. M.; Macdonald, J. M.; Tarling, C. A.; Stick, R. V.; Withers, S. G.; Davies, G. J. Structural, thermodynamic, and kinetic analyses of tetrahydrooxazine-derived inhibitors bound to beta-glucosidases. *J. Biol. Chem.* **2004**, 279, 49236-49242.

34. Gloster, T. M.; Madsen, R.; Davies, G. J. Dissection of conformationally restricted inhibitors binding to a beta-glucosidase. *Chembiochem* **2006**, 7, 738-742.

35. Gloster, T. M.; Madsen, R.; Davies, G. J. Structural basis for cyclophellitol inhibition of a beta-glucosidase. *Org. Biomol. Chem.* **2007**, 5, 444-446.

36. Gibson, R. P.; Gloster, T. M.; Roberts, S.; Warren, R. A. J.; Storch de Gracia, I.; Garcia, A.; Chiara, J. L.; Davies, G. J. Molecular basis for trehalase inhibition revealed by the structure of trehalase in complex with potent inhibitors. *Angew. Chem. Int. Ed. Engl.* **2007**, 46, 4115-4119.

37. F. Cardona, C. Parmegiani, E. Faggi, C. Bonaccini, P. Gratteri, L. Sym. T. M. Gloster, S. Roberts. G. J. Davies, D. R. Rose and A. Goti. Total syntheses of Casuarina and Its 6-O-alfa-Glucoside: Complementary Inhibition towards Glycoside Hydrolases of GH31 and GH37 Families. In press, *Chem. Eur. J.* **2009**.

38. Sim, L.; Quezada-Calvillo, R.; Sterchi, E. E.; Nichols, B. L.; Rose, D. R. Human intestinal maltase-glucoamylase: crystal structure of the N-

terminal catalytic subunit and basis of inhibition and substrate specificity. *J. Mol. Biol.* **2008**, 375, 782-792.

39. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.

40. Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, 10, 1657-1666.

41. Weiner, S. J.; Kollman, P. A.; Case, D. A., Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Jr., W., P.K. . A new force field for molecular mechanical simulation of nucleic acids and proteins *J. Am. Chem. Soc.* **1984**, 106, 765.

42. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-1174.

43. Lin, H.; Truhlar, D. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2007**, 117, 185-199.

44. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Nat.l Acad. Sci. U. S. A.* **2002**, 99, 12562-12566.



45. Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134-139.
46. Coupez, B.; Lewis, R. A. Docking and scoring--theoretically easy, practically impossible? *Curr. Med. Chem.* **2006**, *13*, 2995-3003.
47. Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56*, 235-249.
48. May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225-231.
49. Huang, S.-Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399-421.
50. Marco, E.; Gago, F. Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *ChemMedChem* **2007**, *2*, 1388-1401.
51. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534-553.
52. Kroemer, R. T. Structure-based drug design: docking and scoring. *Curr. Protein Pept. Sci.* **2007**, *8*, 312-328.

53. Jain, A. N. Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* **2006**, 7, 407-420.
54. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des.* **1997**, 11, 425-445.
55. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, 47, 1739-1749.
56. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, 49, 6177-6196.
57. Zwainzig, R. W. High-temperature equation of state by a perturbation method. I, Non-polar gases. *J. Chem. Phys.* **1954**, 22, 1420-1426.
58. Gilson, M. K.; Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, 36, 21-42.
59. Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, 127, 2600-2607.

60. Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, *13*, 3583-3608.
61. Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238-250.
62. Almlöf, M.; Brandsdal, B. O.; Aqvist, J. Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, *25*, 1242-1254.
63. Brandsdal, B. O.; Osterberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. Free energy calculations and ligand binding. *Adv. Protein Chem.* **2003**, *66*, 123-158.
64. Aqvist, J.; Marelus, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 613-626.
65. *Maestro v.7.5*, Schrödinger, L.L.C., New York, NY, 2005 (<http://www.schrodinger.com>).
66. *Macromodel v.9.0*, Schrödinger, L.L.C., New York, NY, 2005 (<http://www.schrodinger.com>).
67. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673-4680.

68. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999-2012.
69. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
70. Bottegoni, G.; Rocchia, W.; Recanatini, M.; Cavalli, A. ACIAP, Autonomous hierarchical agglomerative Cluster Analysis based protocol to partition conformational datasets. *Bioinformatics* **2006**, *22*, 58-65.
71. Bottegoni, G.; Cavalli, A.; Recanatini, M. A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs. *J. Chem. Inf. Model.* **2006**, *46*, 852-862.
72. Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915-931.
73. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306-317.
74. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182-7190.

75. Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided. Mol. Des.* **1998**, *12*, 27-35.
76. Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* **1997**, *40*, 1539-1549.
77. Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method. *J. Med. Chem.* **1999**, *42*, 5142-5152.
78. D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo,; M. Crowley, R. C. W., W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I.; Kolossváry, K. F. W., F. Paesani, J. Vanicek, X.Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke,; L. Yang, C. T., J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin,; Kollman, a. P. A. *AMBER10*, University of California: 2008.
79. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383-394.
80. Hansen, S. B.; Talley, T. T.; Radic, Z.; Taylor, P. Structural and ligand recognition characteristics of an acetylcholine-binding protein from *Aplysia californica*. *J. Biol. Chem.* **2004**, *279*, 24197-24202.
81. Hansen, S. B.; Radic, Z.; Talley, T. T.; Molles, B. E.; Deerinck, T.; Tsigelny, I.; Taylor, P. Tryptophan fluorescence reveals conformational

- changes in the acetylcholine binding protein. *J. Biol. Chem.* **2002**, *277*, 41299-41302.
82. Cubero, E.; Luque, F. J.; Orozco, M. Is polarization important in cation- $\pi$  interactions? *Proc. Natl. Acad. Sci. U. S. A.* **1998**; *95*, 5976-5980.
83. Cavalli, A.; Carloni, P.; Recanatini, M. Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* **2006**, *106*, 3497-3519.
84. Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. From ab initio quantum mechanics to molecular neurobiology: a cation- $\pi$  binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12088-12093.
85. Zacharias, N.; Dougherty, D. A. Cation- $\pi$  interactions in ligand recognition and catalysis. *Trends Pharmacol. Sci.* **2002**, *23*, 281-287.
86. Minoux, H.; Chipot, C. Cation- $\pi$  Interactions in Proteins: Can Simple Models Provide an Accurate Description? *J. Am. Chem. Soc.* **1999**, *121*, 10366-10372.
87. Gay, E. A.; Yakel, J. L. Gating of nicotinic ACh receptors; new insights into structural transitions triggered by agonist binding that induce channel opening. *J. Physiol.* **2007**, *584*, 727-733.
88. Cheng, X.; Wang, H.; Grant, B.; Sine, S. M.; McCammon, J. A. Targeted Molecular Dynamics Study of C-Loop Closure and Channel Gating in Nicotinic Receptors. *PLoS Computational Biology* **2006**, *2*, e134.

89. Kua, J.; Zhang, Y.; Eslami, A. C.; Butler, J. R.; McCammon, J. A. Studying the roles of W86, E202, and Y337 in binding of acetylcholine to acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *Protein Sci.* **2003**, *12*, 2675-2684.
90. Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693-705.
91. Borea, P. A.; Varani, K.; Gessi, S.; Merighi, S.; Dal Piaz, A.; Gilli, P.; Gilli, G. Receptor binding thermodynamics at the neuronal nicotinic receptor. *Curr. Top. Med. Chem.* **2004**, *4*, 361-368.
92. Zhou, J. M.; Zhou, J. H.; Meng, Y.; Chen, M. B. Molecular Dynamics Simulation of Iminosugar Inhibitor-Glycosidase Complex: Insight into the Binding Mechanism of 1-Deoxynojirimycin and Isofagomine toward beta-Glucosidase. *J. Chem. Theory Comput.* **2006**, *2*, 157-165.
93. Woo, H.-J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.
94. Gruia, A. D.; Fischer, S.; Smith, J. C. Kinetics of breaking a salt-bridge critical in protein unfolding. *Chemical Physics Letters* **2004**, *385*, 337-340.