



**UNIVERSITÀ DI PISA**

**DIPARTIMENTO DI FARMACIA**

*Corso di Laurea Specialistica in Chimica e Tecnologia Farmaceutiche*

Tesi di Laurea:

**STUDIO COMPUTAZIONALE DI VIRTUAL SCREENING PER LO  
SVILUPPO DI NUOVI INBITORI DELLA CHINASI BTK**

**Relatori:** Dr. Tiziano Tuccinardi  
Dr. Giulio Poli

**Candidato:** Piccinno Maria Eleonora

Settore Scientifico Disciplinare: **CHIM-08**

ANNO ACCADEMICO 2013-2014

## *INDICE*

CAPITOLO 1 : INTRODUZIONE .....	pag.1
1.1 Struttura .....	pag.1
1.2 Attivazione .....	pag.3
1.3 La BTK e l'artrite reumatoide .....	pag.4
1.4 BTK e leucemia .....	pag.5
1.5 Approcci per lo sviluppo di nuovi inibitori .....	pag.7
1.6 Ligandi .....	pag.10
1.7 Altri ligandi .....	pag.12
CAPITOLO 2 : METODI .....	pag.18
2.1 OMEGA .....	pag.18
2.2 ROCS .....	pag.19
2.2.1 Teoria .....	pag.20
2.2.2 Preparazione dei file di input .....	pag. 22
2.2.3 Interfaccia della Command Line .....	pag. 23
2.2.4 Parametri di input e output .....	pag. 24
2.2.5 Report file .....	pag. 25
2.3 FLAP .....	pag.27
2.3.1 FLAP(FINGERPRINTS FOR LIGANDS AND PROTEINS).pag.27	
2.3.2 Campi di Grid (GRID-MIFs) .....	pag.28
2.3.3 Punti e tetraedri farmacoforici .....	pag.30
2.3.4 Fingerprints e matching .....	pag.31
2.3.5 Receptor-Based Virtual Screening (RBVS) .....	pag.32

2.4 STUDI DI DOCKING CON GOLD 5.0 .....	pag.36
2.5 STUDI DI DINAMICA MOLECOLARE .....	pag.39
2.5.1 Preparazione dei file di input per la dinamica molecolare .....	pag.39
2.5.2 Minimizzazione e simulazione di dinamica .....	pag.40
CAPITOLO 3: PARTE SPERIMENTALE .....	pag.42
3.1 Analisi strutture cristallografiche .....	pag.43
3.2 Screening database Allvender .....	pag.45
3.3 Studi Structure-Based con FLAP.....	pag.46
3.3.1 Screening e risultati .....	pag.47
3.4 Studi di docking .....	pag.48
3.4.1 GOLD .....	pag.48
3.4.2 Docking preliminare .....	pag.49
3.4.3 Docking e risultati finali.....	pag.50
3.5 Studi di dinamica molecolare .....	pag.50
3.5.1 Studi preliminari .....	pag.51
3.5.2 Dinamica e risultati finali.....	pag.52
CAPITOLO 4: CONCLUSIONI .....	pag.54
BIBLIOGRAFIA .....	pag.57

# CAPITOLO 1

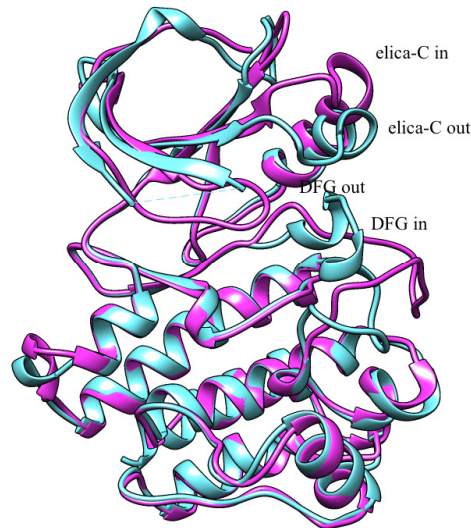
## INTRODUZIONE

La Bruton's Tyrosine Kinase (BTK) è un membro della famiglia delle tirosine chinasi TEC [1]. La BTK è espressa in molte cellule ematopoietiche come nei linfociti B, nei mastociti, e nei macrofagi, ma non è espressa nei linfociti T, nelle cellule natural killer (NK) e nelle cellule del plasma [2]. Una mutazione del gene umano che codifica per la BTK causa un'alterazione di essa, ciò provoca la mancanza dei linfociti B periferici e livelli bassi di immunoglobuline che portano all'insorgenza della malattia ereditaria XLA (X-linked agammaglobulinemia)[3]. Nella XLA il deficit immunitario primario è specifico dei linfociti B e data l'importanza di essi nella risposta immunitaria cellulo-mediata è avanzato l'interesse da parte delle compagnie farmaceutiche nella ricerca di molecole inibitrici della BTK come possibili farmaci per la cura di patologie autoimmunitarie, come l'artrite reumatoide, lupus erimatoso e sclerosi multipla [4]. Analogamente, vi è anche interesse per lo sviluppo di inibitori della BTK per il trattamento di neoplasie ematologiche, poichè l'attivazione aberrante della BTK è stata implicata nella patogenesi di linfomi e leucemie [5].

### 1.1 STRUTTURA

La BTK è una proteina multi-dominio composta da: un dominio pleckstrin homology (PH), un dominio TEC homology (TH), un dominio SH2, un dominio SH3 e un dominio catalitico C-terminale (SH1). Studi strutturali hanno rilevato che la BTK può adattarsi a due conformazioni: uno stato auto-inibitorio della proteina e uno stato attivo. Importante nell'osservazione della struttura è l'orientazione dell'elicaC (aminoacidi 439-452) che può trovarsi in una posizione distale (elica C "out"); questo stato è associato ad uno stato inattivo della proteina, infatti, come si può notare dall'immagine sotto riportata, questa conformazione porta all'occlusione dell'accesso del sito di legame da parte del loop di attivazione. Quando invece l'elica-C è in posizione

prossimale (elica-C "in") consente la formazione di un ponte tra gli amminoacidi K430 e E445 che indica lo stato attivo della chinasi. (Figura 1)



**Figura 1:** Stato attivo (in viola) e inattivo (in azzurro) della proteina

Il loop di attivazione è una sequenza aminoacidica (aminoacidi 542-559) nella quale è contenuto il residuo Y551 (importante per l'attivazione della proteina). Tale loop è estremamente mobile e modifica il sito di legame per l'ATP, quindi anche il sito di legame degli inibitori. Uno stato attivo della proteina si può osservare quando il loop di attivazione è orientato verso l'esterno, e uno stato inattivo quando è orientato verso l'interno (Figura 2). La posizione del loop non solo cambia le dimensioni e la forma della tasca di legame dell'ATP, ma crea anche due nuovi sub-pocket non presenti nelle strutture cristalline di altre chinasi. Uno dei due sub-pocket, di natura acida, è situato più in profondità ed è chiuso dalla catena laterale dei residui del loop attivato L547 e E550; l'altro sub-pocket, potenzialmente basico e presente più in superficie, è circoscritto dai residui M596, R525 e dal residuo del loop attivato K558. Entrambi i sub-pocket potrebbero essere esaminati ed utilizzati per il disegno di inibitori selettivi verso la BTK.



**Figura 2:** Loop di attivazione della BTK

Utile è anche l'osservazione nella struttura della sequenza aminoacidica DFG (aminoacidi 539-541) che può trovarsi in una posizione "in" e in una posizione "out". Quando è in posizione "out" ostacola il legame dell'ATP con la proteina che assume uno stato inattivo; quando invece la sequenza aminoacidica DFG si trova in posizione "in" lo stato dell'enzima risulta essere attivo [6-7]. La posizione della sequenza DFG è infatti determinata da quella del loop di attivazione: quando il loop è orientato verso l'interno la sequenza DFG si trova in posizione "out" e viceversa.

## 1.2 ATTIVAZIONE

L'attivazione del recettore di membrana (BCR) porta a sua volta all'attivazione della proteina chinasi Lyn, appartenente alla famiglia delle Src. La Lyn lega la proteina fosfatidilinositolo-3-chinasi che è coinvolta nella conversione del fosfatidilinositolo-4,5-difosfato in 3,4,5- fosfatidilinositolo. A questo punto la BTK è reclutata sulla membrana plasmatica attraverso l'interazione tra il suo dominio PH (dominio pleckstrin homology) e il 3,4,5- fosfatidilinositolo.

La proteina Lyn fosforila la BTK a livello della tirosina 551 (Y551) presente nel loop di attivazione e ciò porta all'autofosforilazione della tirosina 223 (Y223) presente nel dominio SH3 della BTK. La BTK così attivata fosforila e attiva la fosfolipasi C $\gamma$ , innescando una serie di eventi tra cui il trasporto di ioni calcio, l'attivazione di NF- $\kappa$ B, la proliferazione e la produzione di anticorpi [8].

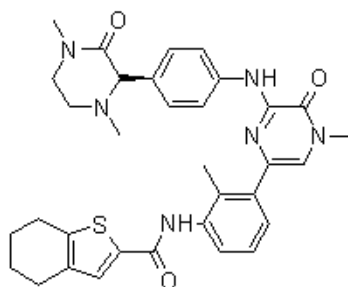
### 1.3 LA BTK E L'ARTRITE REUMATOIDE

L'artrite reumatoide (AR) è una malattia autoimmune che causa l'infiammazione cronica delle articolazioni, dei tessuti che le circondano, e di altri organi del corpo. Circa l'1% della popolazione mondiale è afflitta da AR, e l'insorgenza della malattia è 2-3 volte più frequente nelle donne rispetto agli uomini; la malattia può manifestarsi a qualsiasi età, ma l'esordio è più frequente nelle donne di età superiore ai 65 anni. La caratteristica principale della malattia è una poliartrite simmetrica che coinvolge principalmente le piccole articolazioni delle mani e dei piedi; è stato stimato che circa il 60% dei pazienti con AR soffre di una malattia progressiva, con conseguente distruzione articolare e disabilità.

Sebbene molti aspetti dell'eziologia e patogenesi dell'AR rimangano sconosciute, si pensa che l'insorgenza della malattia sia dovuta ad una risposta immunitaria verso uno o più antigeni presenti nel compartimento sinoviale. Molte cellule, tra cui i linfociti B, macrofagi, linfociti T, cellule endoteliali e fibroblasti, partecipano al processo infiammatorio, suggerendo la presenza di molteplici obiettivi cellulari per l'immunoterapia dell'artrite reumatoide. La terapia diretta contro linfociti B, come rituximab, un anticorpo monoclonale chimerico che elimina le cellule B legandosi al CD20 (antigene sulla superficie cellulare), ha reso evidente il ruolo centrale delle cellule B nella patogenesi dell'AR e di altre malattie infiammatorie.

La BTK svolge un ruolo nei linfociti B di elaborazione e presentazione dell'antigene [9]. Modelli animali sono stati utilizzati per studiare la patogenesi della malattia e testare nuove potenziali terapie per l'uso clinico. I tre modelli usati per la sperimentazione di potenziali agenti terapeutici per la AR sono: l'artrite indotta da adiuvante nel ratto e l'artrite collagene-indotta (CIA) in ratti e topi [11] [12]

Il GDC-0834 (Figura 3) è un inibitore altamente selettivo, ATP-competitivo della BTK, ed è stato sviluppato come un potenziale farmaco terapeutico per l'artrite reumatoide. Il farmaco inibisce la fosforilazione della BTK nel test in vitro e in vivo nei roditori, riuscendo a sopprimere l'artrite nel modello di topo CIA in maniera dose-dipendente ed inibendo significativamente l'infiammazione articolare e i danni della cartilagine.



**Figura 3:** Struttura del GDC-0834

#### 1.4 BTK E LEUCEMIA

La leucemia linfatica cronica (LLC) è una malattia del midollo osseo, un cancro dei linfociti B, ed è causata da una sovrapproduzione di linfociti B difettosi. Essa colpisce in prevalenza persone adulte e anziane; l'invecchiamento della popolazione ha portato ad un aumento del numero di casi di LLC. È definita cronica perché ha molto spesso un decorso estremamente lento con una manifestazione asintomatica, tanto che i malati possono continuare a svolgere una vita normale per molti anni senza terapia antileucemica. I linfociti B affetti da questa patologia perdono la capacità di andare incontro ad apoptosi, o morte cellulare programmata, quel processo tramite il quale quando una cellula non funziona più bene si "suicida" per non danneggiare l'organismo. Essi quindi vivono più a lungo del normale e si accumulano nel sangue periferico e in vari organi, come nel midollo osseo, nella milza e nel fegato. Con il passare del tempo viene compromessa la produzione delle cellule normali del sangue, con riduzione della produzione dei globuli rossi, delle piastrine e dei linfociti B normali, che limita le difese immunitarie.



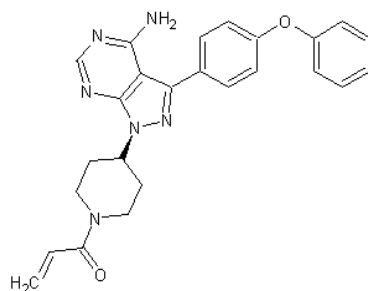
La chemioterapia e la radioterapia cercano soprattutto di eliminare le cellule “malate” agendo sulla loro continua proliferazione, ma mancano di una reale specificità d’azione e inoltre provocano ancora consistenti effetti collaterali. Le terapie più innovative puntano invece a sfruttare il sistema immunitario stesso. I tumori sono generati da “cellule difettose“ del nostro organismo, e il fatto che possano svilupparsi dimostra che il nostro sistema immunitario non le riconosce come estranee e non realizza un’efficiente risposta. I meccanismi effettori del sistema immunitario sono quindi potenzialmente in grado di distruggere in maniera selettiva le cellule tumorali, causando in questo modo pochissimi effetti collaterali, ma necessitano di qualcosa che induca e indirizzi la loro azione.

Il PCI-32765 (Figura 4) è un farmaco detto inibitore del recettore dei linfociti B; la molecola presente sulla superficie dei linfociti essenziale per l’attivazione e la funzione dei linfociti stessi, esso ha dimostrato un potenziale davvero impressionante non solo per quanto riguarda la sua efficacia, ma anche per il basso livello di tossicità. Il meccanismo d’azione del PCI-32765 si basa sulla selettiva e irreversibile inibizione della BTK. Da un punto di vista molecolare, il PCI-32765 provoca la morte programmata delle cellule e ostacola la migrazione e la successiva adesione in linfociti B maligni.

Durante la sperimentazione clinica di fase I/II, si è osservata una remissione duratura della malattia in pazienti con recidiva o resistenti alle terapie standard.

Attualmente la leucemia linfatica cronica è trattata con una combinazione di chemioterapie che però possono causare mielosoppressione, ovvero inibizione della funzionalità del midollo osseo che a sua volta comporta la diminuzione della produzione degli ematociti, che sono le cellule del sangue che generano globuli bianchi, globuli rossi e piastrine. La più grave conseguenza a cui questi trattamenti portano, è un’aumentata suscettibilità alle infezioni. Il vantaggio che offre il PCI-32765 è che non sopprime la produzione di ematociti e l’unico effetto collaterale è solitamente la diarrea. Questo farmaco è attualmente in fase di sviluppo in quanto si è finora dimostrato un composto promettente e un potente inibitore selettivo.

Pertanto, esso rappresenta un approccio terapeutico unico per il trattamento della LLC, poiché è un inibitore molto selettivo, facilmente somministrabile per via orale, che produce una risposta rapida ed efficace nella riduzione dei linfonodi, non è mielosoppressivo e ha una tossicità molto bassa. Una svolta nella cura di questa malattia [13].



**Figura 4** : Struttura chimica di PCI-32765

## 1.5 APPROCCI PER LO SVILUPPO DI NUOVI INIBITORI

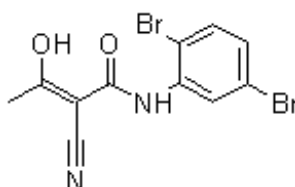
Sono stati utilizzati tre approcci generali per la progettazione di inibitori selettivi della BTK. Uno si è basato sul mirare il residuo aminoacidico presente nella proteina definito come residuo gatekeeper, esso delimita una piccola tasca idrofobica laterale chiamata back pocket, l'altro approccio si basa sulla possibilità di interagire con la catena laterale della cisteina C481, e come terzo approccio si è posto quello di poter occupare una pocket di specificità situata lontano dalla zona cerniera.[14]

### *RESIDUO GATEKEEPER T474*

La tasca posteriore della proteina chinasi è una regione lipofila separata dal sito di legame dell'ATP dal cosiddetto residuo "gatekeeper". Nella BTK questo è rappresentato dalla T474 che essendo un residuo relativamente piccolo permette agli inibitori ATP-competitivi di occupare la tasca.

Circa il 20% di tutte le proteine chinasi umane condividono la treonina come residuo gatekeeper, fornendo un punto di partenza per la progettazione razionale di inibitori selettivi della BTK. A nostra conoscenza, il primo tentativo di progettare inibitori-specifici della BTK che interagiscono con l'amminoacido T474 risale al 1999, quando i

ricercatori della Parker Hughes Cancer Center utilizzarono l'homology modeling e il docking per la progettazione razionale del metabolita leflunomide 10 ( LFM - A13, Figura 5)[15]. Sulla base del legame con la BTK ipotizzato per LFM-A13, è stato suggerito che le interazioni con l'amminoacido T474 potrebbero fornire selettività per altre chinasi.



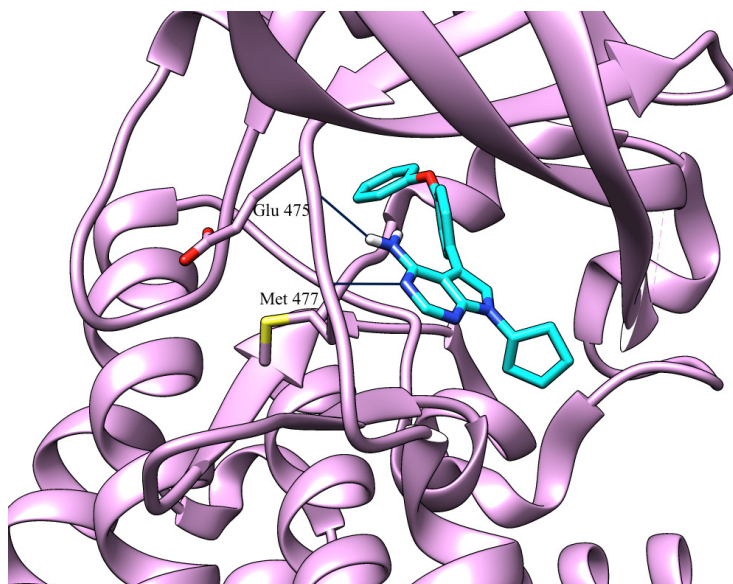
**Figura 5:** Struttura chimica di LFM-A13

L'ipotesi di migliorare la selettività verso la BTK di piccole molecole inibitrici che interagiscono con il residuo gatekeeper T474, è stata ulteriormente rafforzata dai dati di selettività ottenuti dall'inibitore irreversibile PCI-32765 [16]. Dallo screening di questo composti su oltre 100 chinasi è stato osservato che l'attività inibitoria più elevata era ottenuta su più di 10 chinasi che contenevano tutte treonina come residuo gatekeeper.è [16,17]. Questi dati indicano fortemente che l'interazione con il residuo gatekeeper T474 può contribuire in modo rilevante per la selettività di inibitori BTK.

#### *C481*

Il secondo approccio volto a cercare inibitori specifici per la BTK è quello di interagire con la catena laterale dell'aminoacido C481. L'allineamento di 491 chinasi umane dimostra che la cisteina è presente in questa posizione in sole nove chinasi: BLK, BMX, EGFR, ERBB2, ERBB4, ITK, JAK3, TEC, e TXK [16]. Tutte le nove chinasi hanno come gatekeeper un residuo di treonina nella posizione 474, tranne JAK3 e ITK, che presentano invece rispettivamente metionina e fenilalanina, come residui di gatekeeper. La struttura cristallina della BTK in complesso con il ligando B43 (Figura 6) mostra che l'anello ciclopentilico del ligando è a 4 Å dalla catena laterale dell'aminoacido C481

[18]. Aminoacidi più grandi della cisteina o della serina in questa posizione, sono presenti in circa il 70 % di tutte le chinasi umane, dove rischiano di scontrarsi con il ligando determinandone una significativa perdita di potenza inibitoria. Per questo motivo una favorevole interazione con il residuo C481 costituisce un buon elemento di selettività per gli inibitori della BTK.



**Figura 6 :** interazioni ad idrogeno del composto B43 con la *zona cerniera*

### *POCKET DI SPECIFICITA'*

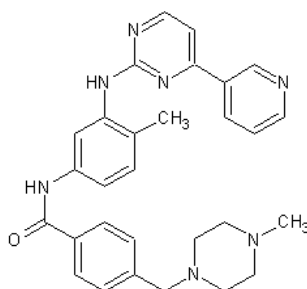
Un terzo approccio di grande successo indirizzato a progettare inibitori-specifici della BTK, è stato quello di riempire una "pocket di specificità", indicata anche come la "tasca H3", circa 15 Å lontano dalla regione cerniera.[19]

Il composto CGI1746 è un potente inibitore della BTK ATP-competitivo (IC<sub>50</sub> = 1.9 nM). L'alta selettività di questo inibitore e dei relativi analoghi, deriva dalla sua porzione tert-butilfenilica che occupa questa pocket con grande complementarità, essa è formata in gran parte dai residui L542, S543, V546 e Y551 del loop di attivazione. È interessante notare che la pocket di specificità è presente anche nella struttura cristallina della BTK complessata con B43 (Figura 6), anche se questo inibitore non occupa questa porzione.

Le catene laterali dei residui S543, V546, e Y551 sono presumibilmente responsabili dell'alta selettività per la pocket di specificità della BTK, inoltre le catene laterali di due aminoacidi in particolare contribuiscono alla formazione di essa nella BTK: l'aminoacido Y551 defosforilato e l'aminoacido D521 che formano un legame idrogeno intermolecolare. Dopo fosforilazione di Y551, si ha repulsione elettrostatica tra queste due catene laterali che dunque si allontanano impedendo la formazione della tasca H3. Legandosi e stabilizzando la pocket di specificità della BTK, il CGI1746 blocca il ciclo di fosforilazione e attivazione della BTK e, in misura molto minore, inibisce l'attività della BTK fosforilata. [19]

## 1.6 LIGANDI

### *IMATINIB*



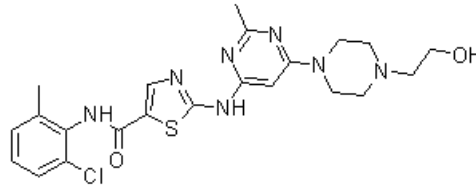
**Figura 7 :** Struttura chimica dell'imatinub

Una piccola molecola inibitrice di chinasi, l'Imatinib (Gleevec, Figura 7) si lega al sito attivo del dominio catalitico della chinasi Bcr-Abl e inibisce la sua attività tirosin-chinasica [2,3]. L'Imatinib rappresenta un caso paradigmatico per la terapia del cancro mirata ed è ora il trattamento di prima linea per la leucemia mieloide cronica (LMC) [4]. Esso è un inibitore di Bcr-Abl, con un  $IC_{50}$  nel range del nanomolare, ma con notevole selettività [6,7].

Un grave inconveniente della terapia con l'Imatinib, deriva dal fatto che molti pazienti sviluppano resistenza secondaria, che porta alla conseguente ricaduta del paziente e alla progressione della malattia. La resistenza è prevalentemente causata da mutazioni puntiformi nel dominio della chinasi Bcr-Abl [9,11]. Pertanto, sono stati sviluppati gli inibitori resistenti delle chinasi. Di questi, solo il Dasatinib (Sprycel, BMS-354825,

Bristol-Myers Squibb) è stato finora approvato per il trattamento di adulti con LMC e Bcr-Abl positivi alla leucemia linfocitica acuta con resistenza a Imatinib o intolleranza alla precedente terapia.[10]

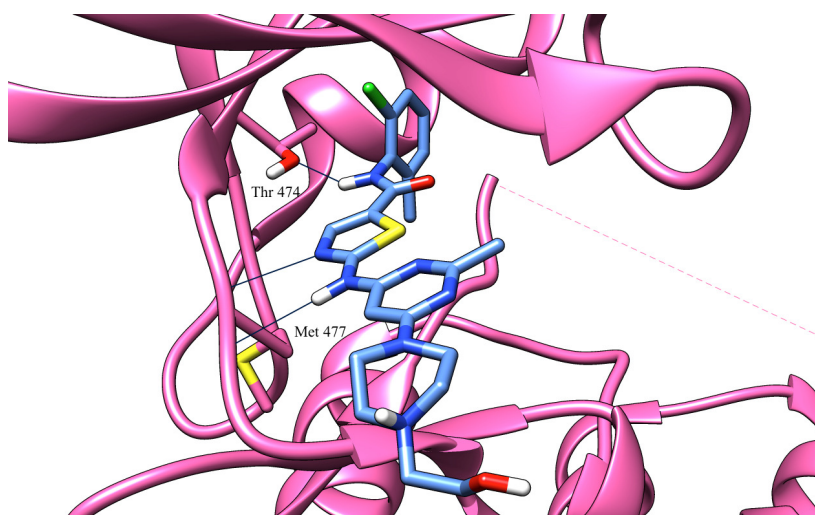
### *DASATINIB*



**Figura 8:** Struttura chimica del dasatinib

Il Dasatinib (Figura 8) inibisce l'attività della chinasi Bcr-Abl, comprese tutte le forme resistenti all'Imatinib clinicamente rilevanti, con l'eccezione della T315I, una mutazione comune [20,21]. Il Dasatinib è stato sviluppato come un inibitore delle chinasi Src e Abl ma, in seguito ha mostrato attività su un'ampia gamma di chinasi [10]. E' utilizzato per il trattamento della leucemia mieloide cronica (LMC). L'analisi dei mastociti derivati dai topi BTK-deficient ha suggerito che l'inibizione della BTK da parte del Dasatinib può essere responsabile della riduzione osservata nel rilascio di istamina in seguito al trattamento con esso; inoltre, il Dasatinib ha dimostrato di inibire il rilascio di istamina nei basofili umani primari e la secrezione di citochine pro-infiammatorie nelle cellule immunitarie.

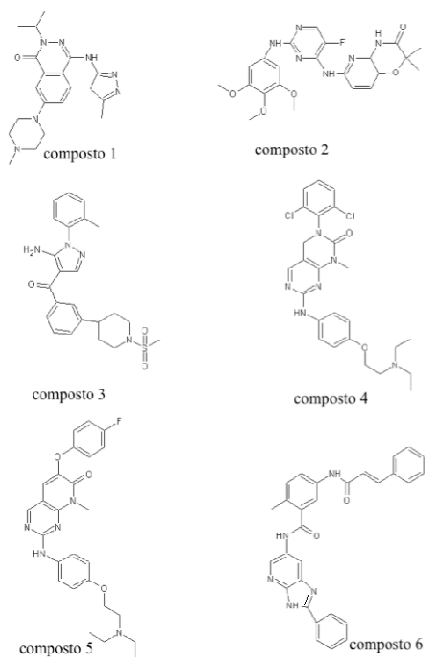
Il Dasatinib si lega al sito catalitico della BTK (figura 9) formando legami a idrogeno con i residui della regione cerniera e con il residuo gatekeeper T474; occupando anche la back pocket composta prevalentemente da residui idrofobici come Met449, Val458, Leu460 e Ile472, ma anche idrofilici come Lys430, Glu445 e Ser538 (Figura 8). L'anello o-clorometilfenilico del Dasatinib è infatti fortemente inclinato rispetto al piano della molecola per poter essere inserito nella tasca idrofobica.



**Figura 9** : Interazioni ad idrogeno del DASATINIB con la regione cerniera

## 1.7 ALTRI LIGANDI

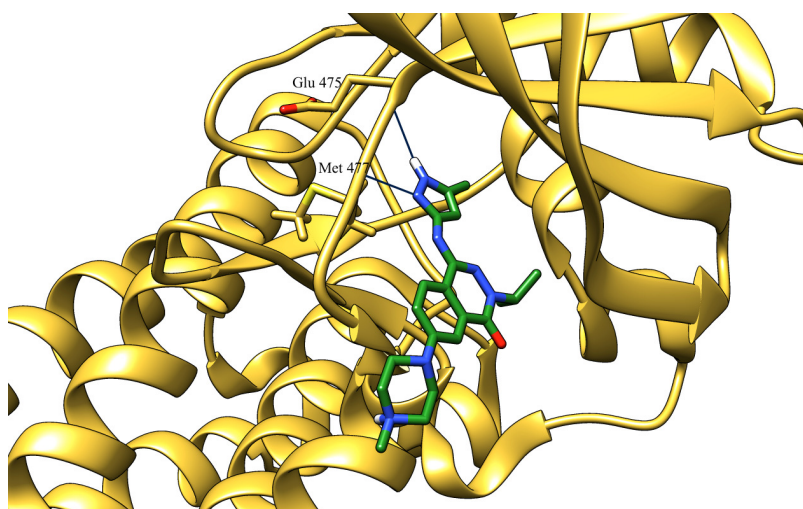
Dall'articolo Kuglstatte A.; Protein Sci 2011 [7] andremo ad esaminare alcuni dei più conosciuti inibitori della BTK, andandoci a soffermare anche sul loro modo di legarsi alla proteina. (Figura 10)



**Figura 10** : Struttura chimica dei composti presenti nell'artico Kuglstatte A. [7]

Il composto 1 induce la chiusura del loop di attivazione nella tasca di legame per l' ATP (Figura 11).

Esso forma tre legami idrogeno tramite la sua porzione amminopirazolica nella regione cerniera della BTK. Il nucleo ftalazinonico del composto 1 è posizionato tra l'amminoacido V427 del lobo N-terminale della BTK e l'amminoacido G480 del lobo C-terminale mentre il gruppo metilpiperazinico è esposto al solvente.



**Figura 11:** Interazioni ad idrogeno del composto 1 con la *zona cerniera*

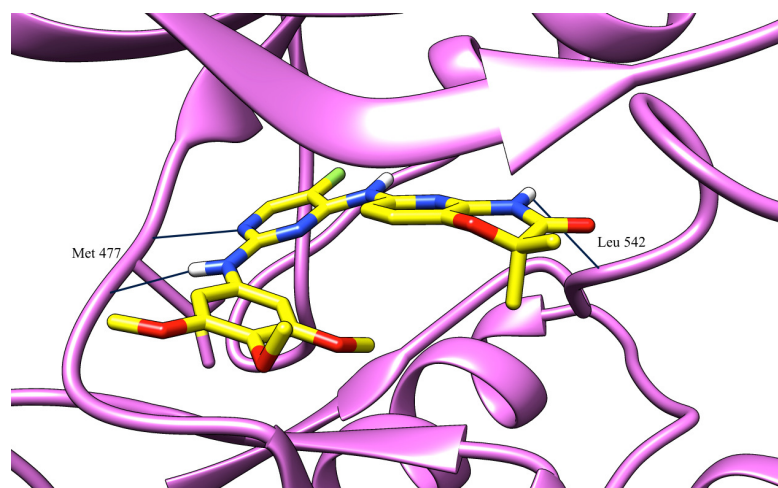
Il composto 2 (R406) stabilizza una conformazione atipica della sequenza DFG (Figura 12).

E' stato riportato che il composto 2 inibisce anche 15 altre chinasi con una selettività minore di circa 10 volte [22]. Esso forma due legami idrogeno nella regione cerniera della BTK tramite la sua porzione diammino-pirimidinica, che è inserita tra le catene laterali dell'amminoacido A428 nel lobo N -terminale e l'amminoacido L528 nel lobo C-terminale. La porzione trimetossi-fenilica del composto 2 è parzialmente esposta al solvente, posizionata tra le catene laterali di L408 e G480, il suo gruppo NH forma un legame idrogeno con il carbonile del backbone della L542 mentre la catena laterale dell'amminoacido L547 si posiziona tra il composto 2 e il lobo C –terminale.

Il tripeptide DFG deve adottare una conformazione atipica. La catena laterale del residuo D539 occupa lo spazio riempito nella conformazione DFG-in dalla catena laterale di F540. Nella conformazione DFG atipica stabilizzata dal composto 2, la catena laterale F540 è posizionata tra le catene laterali di F517 e quelle dei residui

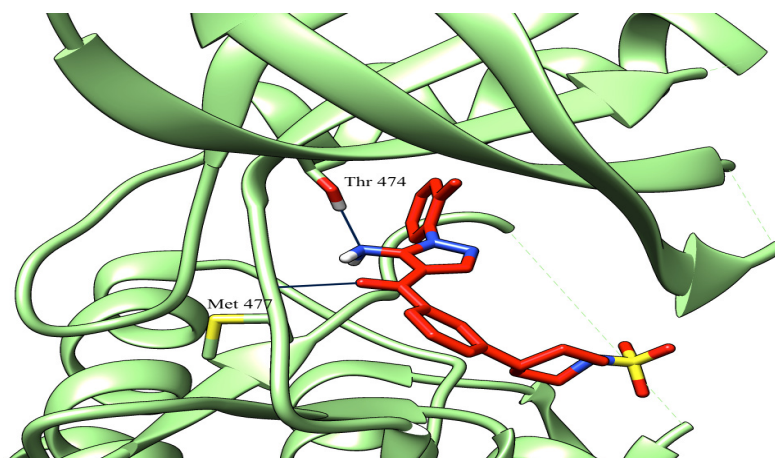


dell'elica-C E445, V448 e M449. La conformazione del loop di attivazione stabilizzato dal composto 2 rappresenta ancora un altro stato inattivo della BTK. Anche in questo caso, il ponte K430-E445 non si forma: la catena laterale K430 si impegna in un legame a idrogeno con il carbonile dell'aminoacido R544, e il residuo E445 dell'elica-C forma un ponte con la catena laterale del R554.



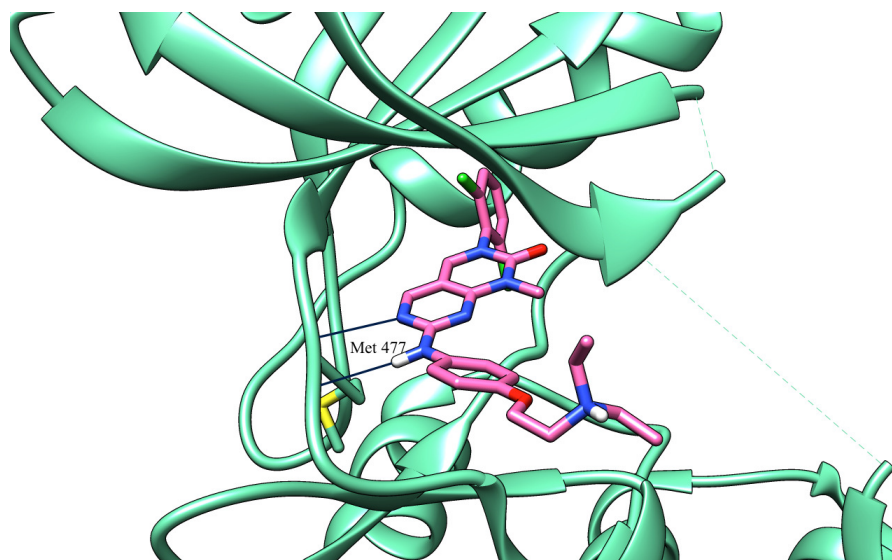
**Figura 12 :** Interazioni ad idrogeno del composto 2 con la *zona cerniera*

Il composto 3 si lega alla conformazione attiva della BTK (Figura 13) con modesta potenza ( $IC^{50} = 3,5 \mu M$ ). Esso forma due legami idrogeno con la *zona cerniera* della BTK e uno con la catena laterale del residuo gatekeeper T474. Il fenile del composto 3 è posizionato nella tasca posteriore tra T474 e il ponte K430 - E445. Il gruppo N-sulfonilpiperidinico è esposto al solvente



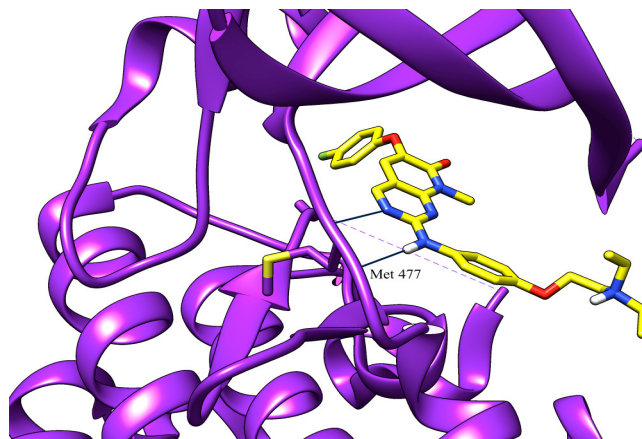
**Figura 13 :** Interazioni ad idrogeno del composto 3 con la *zona cerniera*

Il composto 4 come il composto 3 si lega alla conformazione attiva della BTK (Figura 14), esso inibisce molteplici proteine chinasi. È il più potente inibitore della BTK ( $IC_{50} = 48 \text{ nM}$ ) descritto in questo articolo. Il composto 4 si lega alla *zona cerniera* della BTK attraverso la sua porzione aminopirimidinica. Il gruppo di diclorofenilico del ligando occupa il backpocket e il gruppo fenossico è inserito tra l'amminoacido L408 e il G480. La porzione dietilamminica è altamente esposta al solvente.



**Figura 14 :** Interazioni ad idrogeno del composto 4 con la *zona cerniera*

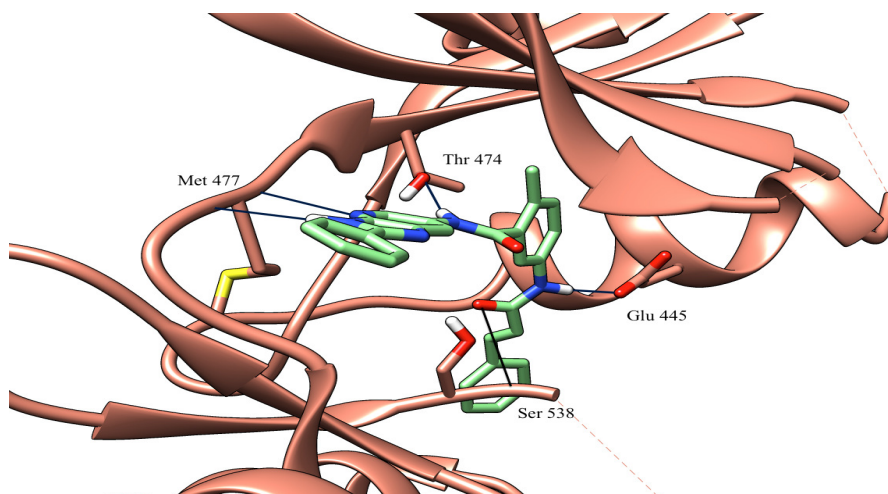
Il composto 5 induce la conformazione "out" dell'elica-C della BTK (Figura 15). Esso si lega alla regione cerniera della BTK nello stesso modo del composto 4, ad esso strutturalmente correlato, ma è circa 100 volte più potente. Il composto 5 contiene un atomo di ossigeno che connette il nucleo biciclico ed il sostituito clorofenilico. Nella BTK, la catena laterale clorofenilica del composto 5 non induce un cambiamento nella posizione del residuo gatekeeper T474, e quindi non può entrare nella tasca posteriore. Essa sporge invece nella regione DFG, dove provoca un cambiamento della posizione dell'amminoacido S538 e induce la conformazione "out" dell'elica-C. Questo infatti impedisce alla catena laterale E445 dell'elica-C di formare il ponte salino con K430 come osservato per il composto 4, a causa di un impedimento sterico. Interessante è quindi notare che, l'aggiunta di un solo atomo di ossigeno tra la porzione centrale e il sostituito laterale del ligando può spostare la conformazione della BTK da uno stato attivo ad uno stato inattivo.



**Figura 15 :** Interazioni ad idrogeno del composto 5 con la *zona cerniera*

Il composto 6 lega una conformazione della BTK DFG-out (Figura 16).

E' stato ipotizzato che cosiddetti inibitori chinasi di tipo II che si legano alla conformazione DFG-out possano fornire un punto di partenza per la progettazione di inibitori altamente selettivi rispetto agli inibitori di tipo I che legano la conformazione DFG-in [7,8]. Il composto 6, inibisce la BTK con un  $IC_{50} = 5,6 \mu M$ ; esso si lega alla *zona cerniera* tramite il suo nucleo imidazopirimidinico. L'anello fenilico legato direttamente al composto 6 è inserito tra L408 e G480. È interessante notare che questa caratteristica è condivisa da tutte le strutture descritte in quest'articolo, nonostante la vasta gamma di legami chimici vincolanti nella *zona cerniera*. La porzione metilfenilica del composto 6 è posizionata tra le catene laterali del residuo gatekeeper T474 e K430. L'adiacente gruppo ammidico forma legami idrogeno con le catene laterali dei residui E445 e S538. La porzione stirenica terminale del composto 6 induce la conformazione DFG-out spostando la catena laterale del residuo F540.



**Figura 16 :** Interazioni ad idrogeno del composto 6 nella *zona cerniera*

Quindi, in conclusione possiamo dire che le strutture cristalline del dominio chinasi della BTK complessate con diversi inibitori ATP-competitivi descritte qui possono dimostrare che piccole molecole legate all'enzima possono stabilizzarne le diverse conformazioni. La stabilizzazione delle conformazioni può verificarsi indipendentemente dallo stato di fosforilazione e dalla presenza di domini regolatori. Consideriamo questa una proprietà delle chinasi in generale, e non un fenomeno specifico della BTK.

.Le strutture BTK-ligando qui riportate, in combinazione con strutture precedentemente note, [5,8] illustrano che la BTK può essere stabilizzata in almeno quattro distinte conformazioni proteiche tutte raffiguranti stati cataliticamente inattivi. Quali di queste conformazioni sono fisiologicamente rilevanti per la regolazione dell'attività della BTK in vivo rimane da stabilire. Le informazioni strutturali a disposizione possono servire come punto di partenza per successive indagini. In generale, le diverse conformazioni della BTK stabilizzate con diversi inibitori dimostrano che la co-cristallizzazione con ligandi di diverse forme, dimensioni e legami chimici, può essere un ottimo strumento per esplorare la flessibilità conformazionale di una proteina.

La disponibilità di un insieme eterogeneo di determinate strutture proteiche della BTK, aumenta in modo rilevante la possibilità di ipotizzare la modalità di legame di piccoli inibitori per i quali non sono state depositati complessi cristallografici con la proteina, come per esempio il metabolita leflunomide LFM- A13.[16]

## CAPITOLO 2

### METODI

#### 2.1 OMEGA

La generazione di conformeri per piccole molecole è un grande problema. Omega è stato progettato principalmente per supportare gli esercizi di virtual screening ad alto rendimento. La riproduzione della conformazione bioattiva, era in precedenza il focus del programma con l'esclusione di tutte le altre possibili conformazioni. Una serie di modifiche sono state introdotte nella versione di Omega 2.0, che lo rendono un programma di generazione di conformeri con fini più generici. Una maggiore qualità del force field (MMFF e derivati) sostituisce la Dreiding force field. La costruzione del modello 3D è stata riprogettata per acquisire anche geometrie di angoli di legame non ideali, pur mantenendo un livello di simmetria appropriato per la ricerca di legami rotabili. La generazione di conformeri di anelli è effettuata in maniera esaustiva nei limiti del force field e della capacità di ottimizzare e identificare i minimi locali.

Per l'inizio dell'analisi conformazionale

- Aprire OMEGA
- Digitare omega2
- Digitare: 

```
omega2 -in nomefile.sdf -out nomefile.oeb.gz -
ewindow 25 -maxconfs 250 -maxtime 600 -rms 0.2 -
strictatomtyping false
```

Dove:

- in = input file
- out = output file
- ewindow = imposta la finestra energetica utilizzata per accettare o rifiutare i criteri per la generazione di conformeri. Qualsiasi conformero con un'energia inferiore alla somma dell'energia della finestra energetica con l'energia minima globale del conformero, sarà accettato. Conformeri con energia superiore saranno respinti. [valore di default = 10.0 Kcal/mol]

- maxconfs = imposta il numero massimo di conformeri che deve essere generato. [default = 400]
- maxtime = limite di tempo impiegato (in secondi) per generare i conformeri per ogni molecola. [default = 120.0]
- rms = imposta il minimo scarto quadratico medio (Root Mean Square) della distanza cartesiana, al di sotto della quale due conformeri sono considerati come duplicati.
- Strictatomtyping false = utilizza lo strict atom typing per l'MMFF94.

Impostando "true", si avrà esito negativo per qualsiasi molecola che contiene qualsiasi tipo di atomo che non dispone di parametri specificati.

Impostando "false", saranno presi in considerazione gli stessi tipi di atomi o simili che sono stati specificati nel file di parametro . [default = true]

## 2.2 ROCS

ROCS, acronimo di Rapid Overlay of Chemical Structures, è un software creato per effettuare ricerche su grandi database di molecole 3D tramite l'utilizzo di un metodo di sovrapposizione che permette di scovare composti non intuitivamente simili, che possono successivamente essere analizzati e sottoposti alle fasi posteriori di un processo di drug discovery.

Metodi simili di sovrapposizione hanno avuto solitamente un impatto marginale nello screening di grandi database, vista la loro scarsa velocità di sovrapposizione (circa 1-2 molecole per secondo); al contrario, ROCS è in grado di sovrapporre dai 600 fino agli 800 conformeri per secondo. A questo si aggiunge la possibilità di dividere un solito calcolo su più processori, traendo un vantaggio massimo dal lavoro contemporaneo di più macchine e rendendo possibile lo screening di database di decine di milioni di composti a tempi ragionevoli.

Per le sue ricerche, ROCS utilizza un metodo di sovrapposizione di tipo shape-based: le molecole sono allineate attraverso un processo di ottimizzazione del corpo rigido che massimizza il volume condiviso: nel fare ciò, il programma considera solamente gli heavy atoms dei ligandi e ignora gli idrogeni. Dal momento che in questo contesto forma e volume sono due entità correlate tra loro, la suddetta procedura di

massimizzazione del volume condiviso si rivela un utile strumento per raggiungere informazioni riguardo la somiglianza delle forme.

Sebbene il programma consista principalmente in un metodo shape-based, l'utente può includere nel processo di sovrapposizione e di analisi della somiglianza una sua personale considerazione della chimica, in modo da facilitare l'identificazione di quei composti che sono simili sia dal punto di vista della forma che dal punto di vista elettrostatico. La possibilità di unire uno screening riguardante la forma a uno specifico per la chimica, il tutto frazionabile su più macchine, rende ROCS un software incredibilmente potente per la ricerca di composti all'interno di grandi database.

Pur essendo storicamente uno strumento utilizzabile esclusivamente da command line, la release 3.0 ha incluso un visualizzatore grafico e un editore di query, denominato vROCS.

### 2.2.1 Teoria

Parlando di ROCS, e quindi del suo metodo di lavoro, deve essere ben chiaro il significato della parola "forma": in questo contesto si afferma che due entità hanno la stessa forma se il loro volume corrisponde perfettamente. Al contrario, più la forma differisce e più il volume occupato dalle stesse si allontana dal corrispondere. Il volume è invece definito come un campo scalare, ovvero una funzione che ha un singolo valore numerico in ogni punto dello spazio:

$$V(\text{volume}) = \int f(x, y, z) \, dv$$

La funzione del volume viene inoltre definita come *funzione caratteristica*. Dobbiamo aggiungere inoltre che due oggetti non possono avere la stessa forma se i loro volumi sono differenti; contemporaneamente, possono però avere lo stesso volume ma forma diversa. A questo punto possiamo scrivere una precisa definizione matematica di similitudine della forma considerando il seguente integrale:

$$S_1 = \int |f(x, y, z) - g(x, y, z)| \, dV$$

dove  $f$  e  $g$  sono due diverse funzioni caratteristiche.

Quando l'integrale assume un valore pari a zero allora  $f$  e  $g$  rappresentano esattamente la stessa funzione e corrispondono quindi alla stessa forma. Al contrario, più il valore dell'integrale si allontana da zero, più differenti sono le forme rappresentate da  $f$  e  $g$ .

L'integrale rappresenta quindi una quantità *metrica* tra le due grandezze  $f$  e  $g$ : il termine *metrico* viene usato raramente per riferirsi alla forma, ma qui assume un preciso significato matematico, cioè una distanza che:

- è sempre positiva;
- assume valore zero solo quando due entità sono identiche;
- rispetta la disuguaglianza triangolare, la quale afferma che se un'entità  $A$  dista  $x$  da un'entità  $B$ , e  $B$  dista  $y$  da  $C$ , allora la distanza tra  $A$  e  $C$  è compresa tra  $|x-y|$  e  $|x+y|$ .

Il tipo di metrica descritto da  $S_1$  viene indicato come *metrica*  $L_1$ .

Un altro tipo di metrica è la  $S_2$ :

$$S_2 = \sqrt{\int [f(x, y, z) - g(x, y, z)]^2 dv}$$

Questo integrale rappresenta quello standard utilizzato per definire l'uguaglianza di forma. Elevando ambo i membri dell'uguaglianza alla seconda otteniamo:

$$S_2^2 = \int f(x, y, z)^2 dV + \int g(x, y, z)^2 dV - 2 \int f(x, y, z) g(x, y, z) dV$$

Questa può essere definita come l'equazione fondamentale per il confronto della forma.

Essa può essere riscritta come:

$$S_{f,g} = I_f + I_g - 2O_{f,g}$$

I termini  $I$  rappresentano la sovrapposizione self-volume delle entità (molecole, in questo caso), il termine  $O$  invece è la sovrapposizione tra le due funzioni: mentre i due termini  $I$  sono indipendenti dall'orientamento, il termine  $O$  varia al variare di quest'ultimo. Trovare l'orientamento che massimizza il valore di  $O$  e che allo stesso tempo minimizzi  $S_{f,g}$  equivale a trovare la miglior sovrapposizione tra le due entità.

Il valore noto come *coefficiente di Tanimoto* può essere ottenuto ricombinando i termini



$I$  e  $O$  secondo l'equazione:

$$Tanimoto_{f,g} = \frac{O_{f,g}}{I_f + I_g - O_{f,g}}$$

In maniera simile è possibile definire un'alternativa misura della forma, utilizzando il *coefficiente di Tversky*, la cui equazione base è la seguente:

$$Tversky_{f,g} = \frac{O_{f,g}}{\alpha I_f + \beta I_g}$$

Normalmente,  $\alpha + \beta = 1$ . Il calcolo secondo Tversky dipende da quale self-volume acquisisce per primo il fattore  $\alpha$ : ROCS effettua due calcoli, il primo in cui è la query ad acquisire  $\alpha$ , e il secondo dove invece lo stesso pre-fattore viene assegnato alla molecola del database. Il valore del *coefficiente di Tversky* può assumere valori maggiori di 1, dal momento che il volume condiviso  $O_{f,g}$  può essere maggiore del self volume  $I_f$  (o  $I_g$ ) di una molecola.

In aggiunta allo shape-based, ROCS può come opzione considerare anche la "chimica" (*Color*) durante la sovrapposizione delle molecole, in maniera tale da poter identificare durante lo screening di un database tutte quelle molecole simili sia nella forma che nelle caratteristiche chimiche. I *color atoms* sono descritti come gaussiane e visibili in vROCS come sfere colorate.

Il programma è fornito di due *color force fields*, *Implicit Mills Dean* (di default) e *Explicit Mills Dean*, i quali sono descritti in color force field files (\*.cff) localizzati all'interno della ROCS data directory: questi sono utilizzati per identificare le similitudini chimiche tra la query e le molecole del database, e quindi per rifinire la sovrapposizione basata esclusivamente sullo shape.

### 2.2.2 Preparazione dei file di input

#### *File del Database*

L'utilizzo più comune di ROCS consiste nel sovrapporre una grande raccolta di composti (dbase file) su una molecola di riferimento, denominata *query*. Il formato più

comune per il database è il multi-conformer file .OEBinary, creato dal programma OpenEye OMEGA; alternativamente è possibile utilizzare SDF, MOL2 e PDB come formati alternativi.

Un altro tipo di file consentito è il dbase file: un file il cui nome termina per *.list* o *.lst* viene letto come una lista di file di molecole, una per linea.

### *File Query*

Il secondo file richiesto, necessario affinché una corsa di ROCS possa essere effettuata, è il file query, il quale può contenere una o più molecole su ciascuna delle quali verrà allineata volta volta una molecola del database. Aggiungendo il comando *-mcquery* allo script si sceglie di trattare il file query come una libreria di più conformeri di un solito composto: in questo caso il software allineerà ciascuna molecola del database ad ogni conformero del file query, riconsegnando comunque un unico file di output, contenente il miglior allineamento scaturito da tutte le sovrapposizioni possibili.

Tipologie alternative sono lo *shape query*, che comprende diversi elementi di forma, incluse molecole, griglie e feature di tipo color, e le griglie, utilizzabili anch'esse come file di riferimento.

### **2.2.3 Interfaccia della Command Line**

Una descrizione dell'interfaccia della Command Line può essere ottenuta eseguendo ROCS specificando l'opzione *help*:

```
prompt> ROCS -help
```

In questo modo genera il seguente output:

#### ***Help functions***

```
ROCS --help simple      : get a list of simple parameters  
ROCS --help all       : get a complete list of parameters
```

*ROCS --help <parameter>* : get a detailed help on parameter  
*ROCS --help html* : create an html help file for this program

Il calcolo di ROCS può quindi essere lanciato da shell, digitando il comando desiderato, come nell'esempio:

*rocs -query <queryfile> -dbase <dbasefile>*

Dove: -query= indica il file di riferimento sul quale effettuare la sovrapposizione;  
-dbase= specifica il nome del file da utilizzare come database.

#### **2.2.4 Parametri di input e output**

Oltre ai già noti files di query e database, richiesti per eseguire un calcolo di ROCS, sono disponibili ulteriori parametri per rifinire e migliorare il risultato del programma, tutti visibili eseguendo ROCS e aggiungendo l'opzione *help* come visto in precedenza.

#InputOptions :

- -param *<paramfile>* : definisce il file dei parametri. Esso contiene una vera e propria lista di tutti i parametri che possono essere utilizzati, invece che riportarli ciascuno sulla Command Line. In aggiunta, il file dei parametri scritto in seguito a un qualsiasi calcolo di ROCS può essere utilizzato in successive esecuzioni dello stesso programma, tramite il comando:

*rocs -param <paramFile>*

- -mcquery : tratta la query come una libreria di conformeri di un solito composto. Di default, questo parametro è posto come falso.
- -scdbase : non tratta due conformeri contigui del file del database come fossero generati da una solita molecola.

#OutputOptions :

- -prefix *<name>* : prefisso utilizzato per nominare tutti i file di output; di default è utilizzato il prefisso ROCS.

- -besthits <N> : numero massimo di hit che vengono riportati nel file di output. Ovviamente tutte le molecole presenti nella *hitlist* dovranno aver superato il valore di *cutoff*, se questo è stato specificato, e ivi saranno tenute o meno a seconda dello score specificato alla voce *rankby*. Di default, N è stabilito 500.
- -cutoff <F> : parametro che stabilisce se una particolare sovrapposizione è buona abbastanza da essere inclusa nella *hitlist*. F= -1.0 di default.
- -rankby <score> : definisce lo score da utilizzare per classificare gli hit. Le varie possibilità includono: TanimotoCombo; ShapeTanimoto (tanimoto); ColorTanimoto; ComboScore (combo); ScaledColor; RefTverskyCombo; RefTversky; RefColorTversky; FitTverskyCombo; FitTversky; FitColorTversky; Overlap. Di default gli hit sono classificati tramite la TanimotoCombo.
- -maxconfs <N>: numero massimo di sovrapposizioni da ottenere per ogni confronto tra query e molecole del database.

#HitsOutputOptions :

- -outputquery : posiziona la query come prima struttura all'interno del file di output. True di default.
- -ofomat <extension> : specifica il formato del file di output. Di default, viene mantenuto il formato SDF.

### 2.2.5 Report file

Oltre al file hits.sdf, contenente gli hit del database che meglio si sovrappongono alla query opportunamente classificati, un calcolo eseguito con ROCS fornisce come file di output un file report (\*.rpt), simile ad una tabella con su specificati i diversi campi.

*Name* è il nome del database di molecole. Se il database contiene più conformeri per una solita molecola, allora l'indice dello specifico conformero segue il nome della molecola.

*ShapeQuery* è il nome della query, cui i corrispondenti hit sono stati allineati. Anche in questo caso, se il riferimento è costituito da più conformeri l'indice dello specifico conformero segue il nome della molecola.

*Rank* è la classificazione numerica della hitlist basata sullo score scelto di applicare. Se non sono state effettuate modifiche, gli hit sono classificati in base al punteggio di Tanimoto.

*TanimotoCombo* fornisce un punteggio che tiene conto sia dello shape che del color, entrambi "secondo Tanimoto". Può assumere tutti i valori compresi tra 0 e 2.

*ShapeTanimoto* fornisce il punteggio di Shape Tanimoto, compreso tra 0 e 1.

*ColorTanimoto* fornisce il punteggio di Color Tanimoto, anch'esso compreso tra 0 e 1.

*FitTverskyCombo* fornisce un punteggio che tiene conto sia dello shape che del color (entrambi secondo Tversky), unendo la *FitTversky* con la *FitColorTversky*. Può assumere tutti i valori compresi tra 0 e 2.

*FitTversky* fornisce il valore di shape secondo l'equazione di Tversky (vedi 2.2.1 Teoria), assegnando a  $\beta$  un valore di 0.95. Precedentemente questo score era noto come Tversky(d).

*FitColorTversky* calcola il valore di color sempre secondo l'equazione di Tversky, assumendo  $\beta=0.95$ .

*RefTverskyCombo* fornisce un punteggio che tiene conto sia dello shape che del colore (entrambi secondo Tversky), unendo la *RefTversky* con la *RefColorTversky*. Può assumere tutti i valori compresi tra 0 e 2.

*RefTversky* fornisce il valore di shape secondo l'equazione di Tversky (vedi 2.2.1 Teoria), assegnando, questa volta, a  $\alpha$  un valore di 0.95. Questo punteggio era precedentemente chiamato Tversky(q).

*RefColorTversky* calcola il valore di color sempre secondo l'equazione di Tversky, assumendo  $\alpha=0.95$ .

*ScaledColor* calcola un valore graduato del colore, considerando il punteggio reale ottenuto da un hit e dividendolo per il punteggio di color ottenuto dalla query contro se stessa.

*ComboScore* unisce la ShapeTanimoto con la ScaledColor, fornendo un valore che tiene conto sia dello shape che del color, e che può assumere tutti i valori compresi tra 0 e 2.

*SubTan* è definita prendendo in considerazione il file di riferimento e la molecola del database alla fine della sovrapposizione e rimuovendo tutti gli atomi della molecola del database che distano maggiormente di 1.5 Å da qualsiasi atomo della query. A questo punto si calcola lo ShapeTanimoto per queste due strutture e il punteggio ottenuto è salvato proprio come SubTan.

*Overlap* consiste nel valore assoluto del volume sovrapposto tra la query e la molecola del database. Questo parametro è utilizzato quando è una griglia ad essere utilizzata come query.

## **2.3 FLAP**

### ***2.3.1 FLAP (FINGERPRINTS FOR LIGANDS AND PROTEINS)***

Il software FLAP è stato sviluppato per sfruttare le numerose informazioni fornite dalle strutture cristallografiche, dalle molecole presenti in letteratura e/o dai loro complessi. Esso rappresenta una procedura computazionale innovativa capace di esplorare lo spazio farmacoforico tridimensionale di ligandi e proteine, e fornire informazioni qualitative e quantitative relative alla complementarietà delle loro interazioni, utilizzando delle strutture di riferimento che permettano di effettuare delle sovrapposizioni ligando-ligando, proteina-proteina o proteina-ligando. Nei paragrafi seguenti verranno elucidati gli elementi sui quali sono basati i calcoli effettuati dal software e le metodologie di calcolo..

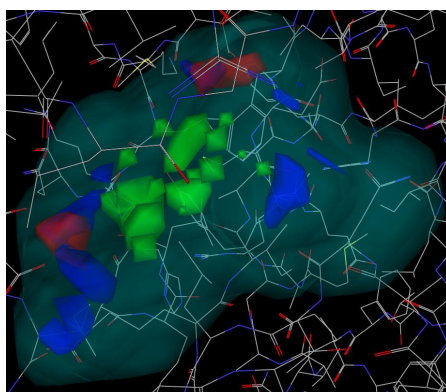
### **2.3.2 CAMPI DI GRID (GRID-MIFs)**

Le opportunità che il programma offre sono molteplici, spaziando da screening basati sulle informazioni recettoriali a modelli basati sulle caratteristiche dei ligandi. Ad ogni modo, alla base di qualsiasi tipo di analisi vi è l'esplorazione delle caratteristiche chimico-fisico-strutturali dei ligandi e/o delle cavità delle proteine forniti a FLAP, attraverso l'analisi dei campi di GRID (GRID-MIFs). I MIFs (molecular interaction fields, letteralmente "campi di forza molecolari") descrivono la variazione spaziale dell'energia di interazione tra una (macro)molecola target di riferimento e un determinato probe (letteralmente una "sonda"). La molecola target può essere una macromolecola (come una proteina), o un ligando. Un probe può essere costituito da una piccola molecola, un frammento di molecola o un singolo atomo: tale entità virtuale ha effettivamente il compito di ispezionare lo spazio circostante la molecola target (o quello definito da un sito o una tasca proteica) e rilevare le zone in cui un'interazione di una certa entità può essere stabilita tra il probe stesso ed il target, definendo così una mappa volumetrica relativa al dato tipo di interazione. I MIFs possono essere rilevati sulla base di varie soglie di energia, utilizzando superfici che delimitano le zone all'interno delle quali un dato tipo di interazione supera un certo valore energetico impostato. Esistono vari programmi in grado di calcolare i MIFs ma sicuramente il più efficace ed il più utilizzato è GRID che è presente all'interno di FLAP e costituisce le fondamenta delle sue tecniche di analisi.

A seconda della natura del probe utilizzato cambia il tipo di interazione che può essere stabilita tra questo ed il target, per cui probes diversi rilevano MIFs di natura ed entità diversa. Dalla combinazione di MIFs di natura diversa è possibile realizzare un quadro più o meno dettagliato delle varie possibilità di interazione che un determinato target offre, in relazione allo spazio ad esso circostante ed analizzare, ad esempio, le caratteristiche chimico-fisico-strutturali richieste da una determinata tasca recettoriale che un ligando deve soddisfare per potervi interagire o quelle che una generica molecola deve presentare come presupposto per poter dare le stesse interazioni di un ligando attivo di riferimento. Per quanto sia possibile i probes dovrebbero dunque essere scelti in modo tale da rappresentare al meglio tutte le possibili interazioni stabilibili da un potenziale ligando o dagli amminoacidi che definiscono una cavità recettoriale. Poiché 60 diversi GRID probes sono disponibili in FLAP teoricamente non ci sono limiti alla selezione delle loro combinazioni. In realtà, allo stato attuale delle tecniche

computazionali, per molte ragioni è necessario evitare l'utilizzo di più di sei probes, che devono dunque essere scelti accuratamente per poter rappresentare al meglio le interazioni principali dei target da analizzare. Ad ogni modo, normalmente vengono utilizzati 4 probes, che sono più che sufficienti per ottenere un'accurata descrizione delle principali possibilità interattive di una determinata molecola. Tali probes sono infatti proposti da FLAP come default e sono H, DRY, N1 e O, rispettivamente un probe volumetrico (per l'analisi del volume e della forma del target in esame), uno idrofobico (per l'analisi di possibili interazioni di Wan der Waals,  $\pi$ -stacking, forze di dispersione ecc.), un donatore ed un accettore di legami a idrogeno (per l'analisi di possibili interazioni dipolari con accettori e donatori di legami a idrogeno rispettivamente o con gruppi carichi). Si riporta il risultato del calcolo dei campi di GRID con tali probes per un target recettoriale (Figura1).

Poiché i MIFs calcolati per un target recettoriale sono utilizzati per identificare le posizioni più favorevoli per lo sviluppo di interazioni con il target stesso che devono essere occupate dagli atomi di una data molecola, è necessario che gli atomi dei potenziali ligandi forniti a FLAP come input siano classificati sulla base dei probes scelti: ogni atomo di ciascun ligando viene dunque associato al probe che meglio approssima le sue caratteristiche e le sue possibilità di interazione. FLAP normalmente fa tutto questo in automatico con le molecole fornitegli, ma la procedura può anche essere personalizzata. Questa analisi è fondamentale sia per i calcoli structure-based che per quelli ligand-based, in quanto questa "scomposizione" dei ligandi nei probes che meglio li rappresentano è utilizzata per guidarne la reciproca sovrapposizione così come la sovrapposizione con i MIFs dei target recettoriali.



**Fig.1** : Rappresentazione dei campi dei GRID calcolati per un sito recettoriale dai probes H (in celeste), DRY (in verde), N1 (in blu) e O (in rosso).



### **2.3.3 PUNTI E TETRAEDRI FARMACOFORICI**

Una volta che per ogni ligando e/o sito proteico sono stati calcolati i MIFs, le informazioni in essi contenute vengono condensate in pochi punti farmacoforici attraverso una funzione che li seleziona sia dal punto di vista energetico che spaziale: tali punti costituiscono le zone chiave di interazione all'interno dei MIFs ad essi associati. Questa selezione, che viene fatta automaticamente da FLAP può anche essere eseguita manualmente e rappresenta un punto critico dell'intero sistema. L'importanza e la delicatezza di questa operazione sono accentuate dal fatto che non possono essere utilizzati troppi punti farmacoforici (normalmente non più di 25-30 per probe) allo scopo di evitare problemi combinatoriali, ma allo stesso tempo la loro combinazione deve essere tale da offrire una descrizione sufficientemente accurata e dettagliata delle possibili interazioni sviluppabili da un ligando o dai residui di un sito proteico (i punti farmacoforici devono essere in grado di rappresentare adeguatamente l'insieme dei MIFs relativi ai corrispondenti elementi).

Alla fine di questa procedura di selezione viene realizzato, per ogni elemento analizzato, un totale di circa 100-200 punti (a seconda di quanto impostato) ciascuno dei quali è caratterizzato da tre diverse proprietà: il tipo di interazione energetica rappresentata (relativa alla natura del probe e del MIF a cui si riferisce), l'entità di tale interazione (ad esempio il valore in Kcal/mol dell'energia di interazione sviluppabile) e le coordinate cartesiane che ne identificano la posizione all'interno di una tasca recettoriale o nello spazio circostante un ligando. A questo punto FLAP procede automaticamente con la creazione di un modello matematico composto da tutte le possibili combinazioni di quartetti di punti farmacoforici, dando luogo a molteplici tetraedri farmacoforici che costituiscono gli elementi chiave del metodo di calcolo di FLAP: ogni ligando o sito recettoriale che FLAP utilizza come input viene infatti descritto dalla serie di quartetti o tetraedri farmacoforici che in esso sono stati individuati e su di essi si basa la sovrapposizione ligando-ligando e ligando-proteina. Per quanto riguarda i ligandi FLAP costruisce anche dei "tetraedri atomici", ovvero costituiti da quartetti di punti che rappresentano gli atomi pesanti delle molecole analizzate, precedentemente classificati sulla base dei probes utilizzati per il calcolo dei MIFs: anche questi tetraedri sono ovviamente essenziali sia per i calcoli ligand-based che receptor-based.

### ***2.3.4 FINGERPRINTS E MATCHING***

Il numero di tetraedri necessari per descrivere i diversi elementi utilizzati da FLAP varia a seconda della loro natura. Nel caso dei ligandi è necessario considerare più conformazioni, che sono create, selezionate e conservate in automatico da FLAP. Ad ogni conformazione saranno associati MIFs diversi e dunque punti farmacoforici diversi (che avranno quantomeno coordinate spaziali e reciproche distanze diverse); la stessa cosa vale per i tetraedri costruiti sugli atomi pesanti della molecola, le cui distanze e coordinate variano in base alla conformazione assunta dalla molecola stessa. Ogni conformazione sarà dunque descritta da tutta una serie di tetraedri ricavati dai punti farmacoforici ad essa associati e dalla particolare disposizione degli atomi pesanti.

La completa descrizione di una molecola viene dunque ottenuta considerando tutti i diversi tetraedri ricavabili da tutte le diverse conformazioni associate alla molecola stessa: il loro insieme costituisce una specie di mappa identificativa della molecola che, opportunamente codificata, dà luogo ad una grande matrice chiamata “fingerprint” (letteralmente “impronta digitale”). Quando viene creato un database di composti tutte queste informazioni (GRID-MIFs, punti farmacoforici, tetraedri farmacoforici, tetraedri atomici e infine le fingerprints ricavate) vengono memorizzate insieme alle relative strutture ed utilizzate per i vari calcoli di screening.

Nel caso dei siti recettoriali che sono definiti da una struttura cristallografica e dunque caratterizzati da residui amminoacidici assumenti una ben precisa conformazione, viene calcolata una singola serie di MIFs e dunque di punti e tetraedri farmacoforici, che vanno a descrivere la determinata tasca proteica. Inoltre non viene effettuata la mappatura degli atomi pesanti dei residui amminoacidici che delimitano il sito e dunque non vengono creati tetraedri atomici, dato che i calcoli receptor-based non utilizzano questo tipo di dato. Ad una tasca recettoriale è dunque associata una quantità di dati significativamente minore rispetto ad un ligando, in quanto la sua fingerprint è costituita dalla somma codificata dei soli tetraedri farmacoforici calcolati in base alla data posizione che i suoi residui assumono nel cristallo utilizzato.

Il metodo di calcolo che FLAP utilizza, indipendentemente dal tipo di screening che sta effettuando si basa sull’opportuna sovrapposizione dei tetraedri farmacoforici e atomici

di ligandi e siti proteici da analizzare, precedentemente realizzati come spiegato. La sovrapposizione avviene previa codifica dei tetraedri in bitstrings, che sono i veri elementi che vanno a costituire i fingerprints. Un bitstring è una serie di dati in grado di rappresentare univocamente, in un formato matematico, un determinato tetraedro. Di questi valori sei rappresentano le reciproche distanze individuabili congiungendo tra loro i quattro punti che costituiscono un dato tetraedro, altri quattro sono necessari per identificare a quale tipo appartiene ognuno dei quattro punti, cioè con quale probe si identifica (se fa parte di un tetraedro atomico) o a quale MIF si riferisce (se fa parte di un tetraedro farmacoforico) ed un ultimo valore è proporzionale alla somma delle energie associate ai quattro punti. In questo modo ciascun bitstring identifica un ben preciso quartetto di punti ed un tetraedro di un determinato volume associato ai punti stessi. Il volume del tetraedro è calcolato sulla base delle coordinate dei punti che ne rappresentano i vertici e delle loro distanze: si ottiene così un valore corredato di segno che fornisce anche indicazioni sulla chiralità del tetraedro. Dall'unione di tutti i bitstrings associati ad una molecola o un recettore, ovvero di ogni singola serie di dati descrivente un determinato tetraedro, calcolato per il determinato elemento, si ottiene la complessa matrice di dati che costituisce il fingerprint dell'elemento in questione. Quando FLAP effettua un qualsiasi screening inizialmente confronta tra loro tutti i vari bitstrings delle fingerprints di ogni molecola da analizzare con quelli del relativo template (l'elemento di riferimento, sia questo una molecola, una tasca recettoriale o un modello farmacoforico), ovvero sovrappone tutti i vari tetraedri delle molecole test con quelli del template, successivamente raffina la ricerca sulla base di altri elementi che variano a seconda del tipo di calcolo effettuato.

### **2.3.5 Receptor-Based Virtual Screening (RBVS)**

L'intero processo è basato sull'assunzione che l'affinità di una molecola per un sito recettoriale, e dunque l'entità delle interazioni che possono essere tra loro stabilite, sia una funzione additiva delle singole affinità/interazioni che possono essere individuate tra i singoli atomi e/o gruppi funzionali presenti nel ligando e nel sito stessi. Queste possibili interazioni vengono rilevate attraverso la sovrapposizione dei tetraedri atomici delle molecole da analizzare con i tetraedri costituiti dai punti farmacoforici relativi ai

MIFs calcolati per la tasca recettoriale. Quando un potenziale ligando, in una conformazione energeticamente favorevole, possiede quattro atomi che presentano una disposizione spaziale tale da permetterne la sovrapposizione con un ugual numero di punti farmacoforici recettoriali dello stesso tipo (che infatti esprimono la richiesta di atomi con quelle determinate caratteristiche in quelle precise posizioni), il ligando in questione può essere potenzialmente in grado di stabilire un'interazione di legame efficace con il recettore in esame.

Ovviamente nella pratica è molto difficile riscontrare una perfetta sovrapposizione tra gli atomi di un ligando ed i corrispondenti punti farmacoforici di un recettore, dunque quando se ne valuta la posizione reciproca nello spazio cartesiano è necessario accettare un certo grado di approssimazione. Ad esempio il risultato della sovrapposizione dei due tetraedri atom-based e pharmacophore-based deve essere considerato soddisfacente se ciascun atomo è a meno di 1Å di distanza dal corrispondente punto farmacoforico, molto buono o ottimo se le relative distanze si riducono a 0,5-0,25 Å, mentre si considerano coincidenti due tetraedri i cui punti corrispondenti si discostano tra loro per meno di 0,25 Å.

Quando viene effettuato un receptor-based virtual screening FLAP sovrappone tutti i tetraedri atomici di tutti i diversi conformeri di ciascuna molecola analizzata con tutti i tetraedri farmacoforici del recettore target. Quando ha luogo una sovrapposizione soddisfacente la ricerca viene comunque continuata in modo esaustivo, poiché più sovrapposizioni corrette possono essere riscontrate tra i diversi tetraedri di un ligando e di un target, anche perché sia in una molecola che in un recettore possono essere presenti tetraedri equivalenti ma generati da punti con coordinate diverse.

Il corretto matching tra due o più tetraedri non è condizione necessaria e sufficiente per determinare la possibilità di un effettivo binding tra il recettore ed il ligando, poiché anche la complementarità delle loro forme e dei loro volumi è un elemento fondamentale che non può essere trascurato e che costituisce un criterio di valutazione dei risultati ottenuti dalle varie sovrapposizioni. Una delle caratteristiche chiave di FLAP è infatti la sua capacità di vagliare e filtrare i risultati delle sovrapposizioni dei tetraedri di recettore e ligando sulla base della loro forma. Inizialmente infatti il numero totale delle sovrapposizioni ritenute corrette e momentaneamente memorizzate dal programma, quando si analizza un database di molecole con un RBVS, è estremamente elevato, ma molte di queste soluzioni possono essere rapidamente eliminate

considerando la posizione nello spazio di tutti gli atomi pesanti (non solo di quelli che descrivono i tetraedri di volta in volta considerati) dei ligandi esaminati. Se alcuni di questi si trovano all'esterno del volume escluso del recettore il tetraedro viene scartato: in questo modo il numero delle soluzioni fornite dalle sovrapposizioni viene sensibilmente ridotto. Alla fine di questo step FLAP valuta, per le soluzioni rimaste, la presenza di eventuali ingombri sterici tra ligandi e recettore e la loro importanza: le soluzioni presentanti contatti sterici ritenuti inaccettabili vengono scartate. Le soluzioni accettate (le varie pose di ogni molecola test) vengono poi ordinate sulla base di un punteggio che esprime l'accuratezza con cui gli atomi pesanti della molecola test sono inseriti all'interno dei MIFs calcolati per il sito recettoriale, ovvero che valuta quanto accuratamente ogni MIF presente nel sito è in grado di alloggiare la o le porzioni della molecola test che ne soddisfano le richieste e lasciar fuori quelle che presentano proprietà incompatibili. La pose, tra quelle accettate di ogni molecola test, che presenta il punteggio più alto viene utilizzata come rappresentativa del risultato della data molecola, ma anche tutte le altre pose vengono salvate tra le soluzioni relative al dato ligando.

In sostanza FLAP, per ogni ligando da analizzare, effettua prima un rapido calcolo per rilevare la presenza di tetraedri atom-based complementari a tetraedri pharmacophore-based del recettore e capire così se il ligando possiede le caratteristiche necessarie per poter interagire con esso, poi effettua la sovrapposizione di tali tetraedri per posizionare i conformeri del ligando ad essi associati all'interno della tasca proteica ed infine valuta la correttezza di ogni pose del ligando sulla base della complementarità di forma tra il ligando in quella data conformazione ed il recettore. Le soluzioni rimanenti sono dunque fortemente influenzate dalla forma della proteina e le coordinate delle conformazioni e dei tetraedri che le rappresentano vengono salvate.

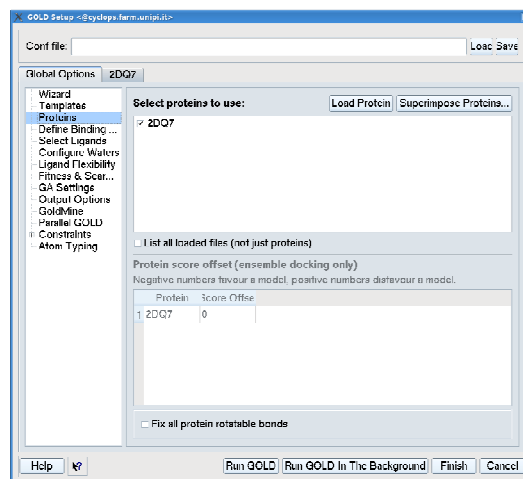
In tutto questo può sembrare che FLAP, in effetti, sia in grado di effettuare una specie di docking, anche se attraverso più passaggi, ma la realtà è diversa. Mentre un programma di docking calcola la totale energia di interazione tra un ligando e una proteina, FLAP considera esclusivamente pochi contributi alla volta (le quattro interazioni associate a una coppia di tetraedri sovrapposti) considerando come condizione necessaria e sufficiente per un buon risultato la corretta sovrapposizione di almeno due tetraedri combinata con l'assenza di repulsioni steriche tra ligando e recettore. Questo metodo produce delle soluzioni che possono poi essere ottimizzate

attraverso un docking vero e proprio all'interno del sito proteico, ma che non possono essere considerate un effettivo risultato di docking. L'altra fondamentale differenza è che FLAP opera in uno spazio assolutamente discreto, utilizzando esclusivamente i punti essenziali per descrivere i GRID-MIFs e/o la struttura delle piccole molecole, ricercandone una complementarità. Questa caratteristica permette a FLAP di effettuare un matching proteina-ligando basato sulle forze di legame stabilibili tra i due elementi, più che sulla complementarità di forma, che è invece valutata nello step successivo, al contrario di quanto avviene invece nel docking che è guidato inizialmente dalla complementarità di forma e in cui solo in un secondo momento vengono calcolate e quantificate tutte le possibili interazioni relative alla potenziale soluzione.

## **2.4 STUDI DI DOCKING CON GOLD 5.0**

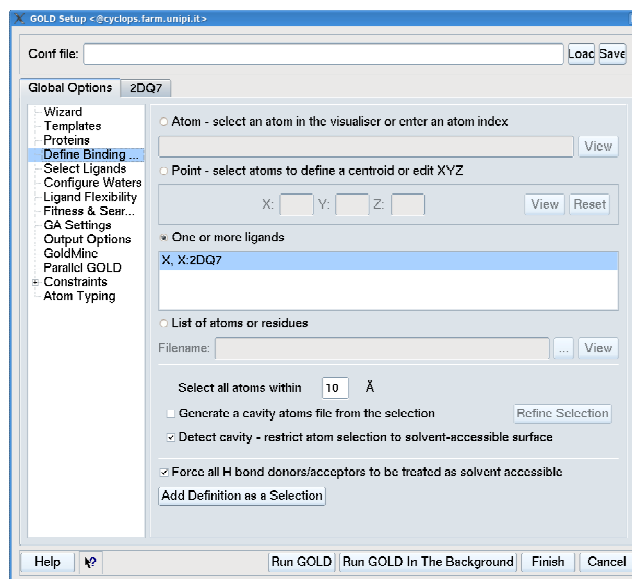
Gli studi di docking condotti sono stati eseguiti con il software GOLD 5.0. Il programma è in grado di effettuare il docking flessibile ed automatico di un set di composti all'interno di una struttura proteica. Le molecole analizzate vengono "posizionate" nel sito recettoriale, rintracciato all'interno della proteina fornita come input. Le varie soluzioni ottenute per ogni molecola vengono valutate sulla base di una funzione di scoring selezionata tra quelle disponibili e viene loro assegnato un punteggio: le "n" migliori soluzioni (secondo quanto impostato dall'utente) vengono restituite come output dal software.

La struttura proteica utilizzata per il calcolo è quella contenuta nel cristallo 3K54. Fornendo come input al programma il complesso cristallografico, questo permette di eliminare l'acqua, di aggiungere gli idrogeni alla proteina e al ligando e di estrarre quest'ultimo dal sito recettoriale per poi utilizzarlo (eventualmente) come riferimento per delimitare sito di legame e per analizzare i risultati del docking (Figura 2).



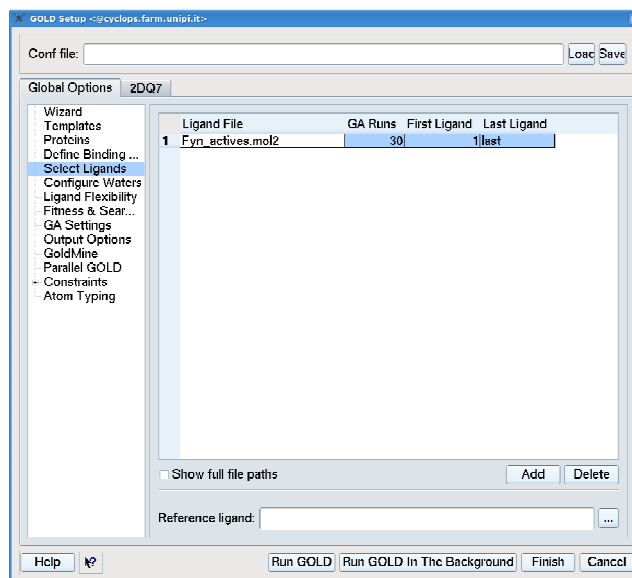
**Fig. 2:** Finestra di dialogo relativa al caricamento e all'elaborazione della proteina.

La tasca recettoriale all'interno della quale effettuare il docking è stata definita in modo tale da comprendere tutti gli amminoacidi contenuti in un raggio di 10 Å dalla proteina complessata, restringendo però la selezione dei contorni della tasca alla superficie accessibile al solvente (Figura 3).



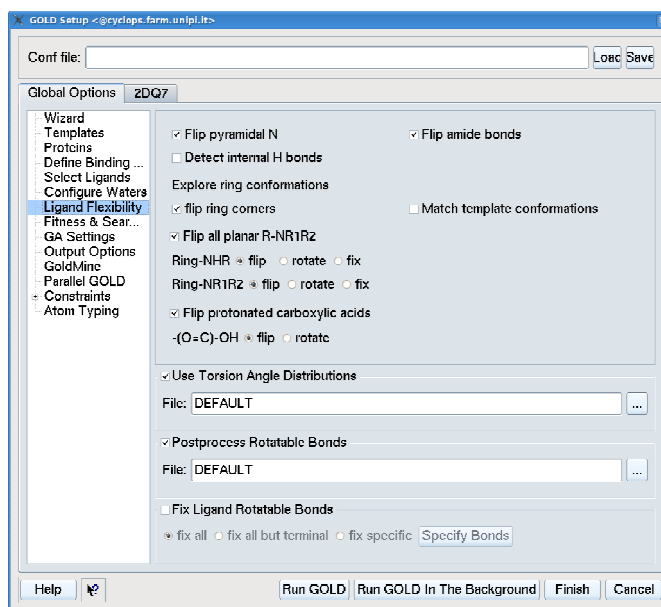
**Fig. 3:** Finestra di dialogo relativa alla definizione del sito di legame.

Ciascun ligando analizzato è stato sottoposto a 30 cicli di algoritmo genetico per il calcolo delle soluzioni di docking (Fig. 4).



**Fig. 4** :Finestra di dialogo relativa alla selezione dei ligandi su cui effettuare il docking.

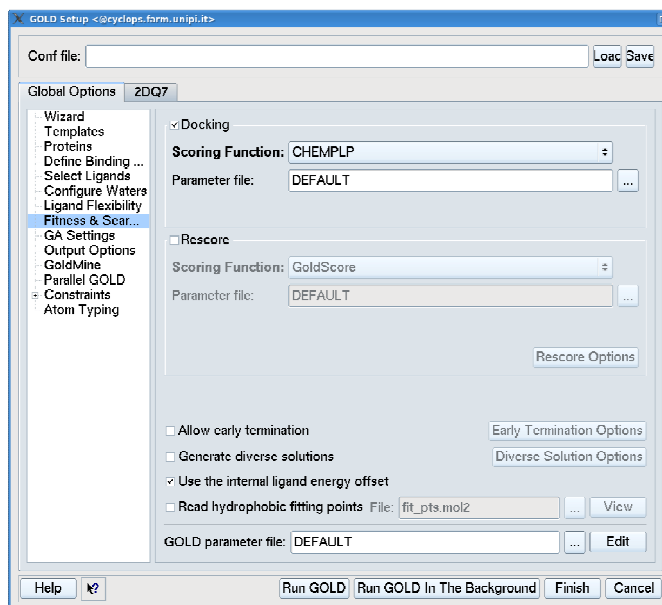
È stata attivata, per i ligandi analizzati, la possibilità di inversione conformazionale dei legami ammidici e degli anelli alifatici e la possibilità di inversione piramidale degli atomi di azoto piramidali, lasciando tutti gli altri parametri relativi alla flessibilità dei ligandi nell'impostazione di default (Fig. 5).



**Fig. 5** :Finestra di dialogo relativa alla definizione della flessibilità dei ligandi nel docking.

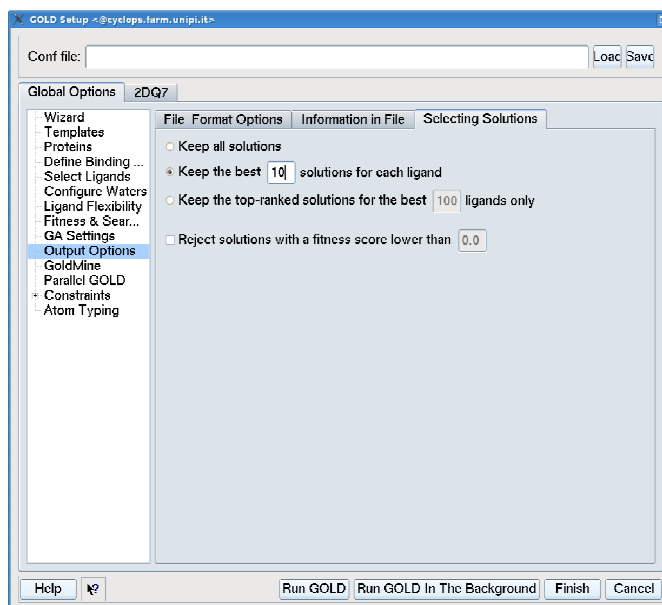
Come scoring function è stata scelta ChemPLP, disattivando la funzione “Allow early termination” (Fig. 6).





**Fig 6** :Finestra di dialogo relativa alla scelta delle funzioni di scoring.

È stata infine impostata la restituzione delle 10 migliori soluzioni per ogni ligando, salvate in un unico file in formato sdf (Fig. 7).



**Fig. 7**: Finestra di dialogo relativa alla selezione delle soluzioni di docking da restituire.

Per gli altri calcoli di docking è stato modificato solo quest'ultimo parametro: è stata infatti impostata la restituzione della migliore soluzione per ogni ligando "Keep the best 1 solutions for each ligand".

## 2.5 STUDI DI DINAMICA MOLECOLARE

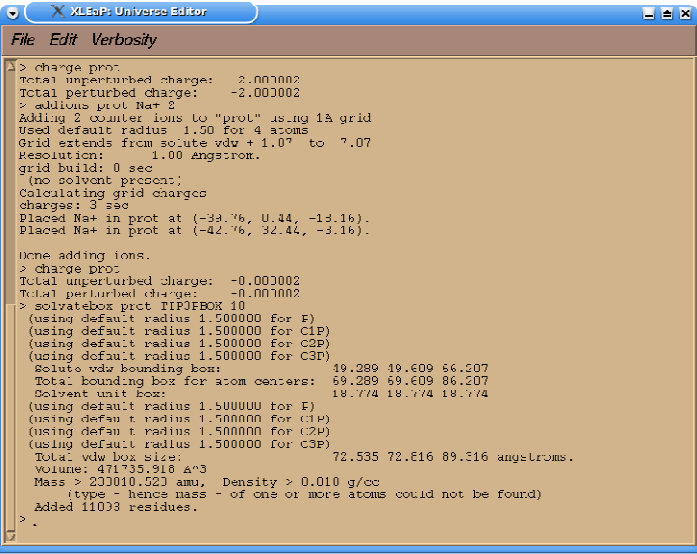
Gli studi di dinamica molecolare sono stati condotti utilizzando il software AMBER11: questo è in realtà costituito da una suite di programmi che insieme permettono di effettuare simulazioni di dinamica molecolare, specialmente su biomolecole. In realtà nessuno di tali programmi porta questo nome: ad ogni modo, per semplicità, è possibile considerare AMBER come un complesso software che racchiude in se tutti i vari programmi necessari per la realizzazione di uno studio di dinamica molecolare.

### 2.5.1 Preparazione dei file di input per la dinamica molecolare

Per effettuare uno studio di dinamica molecolare è necessario innanzitutto creare dei file di input, opportunamente elaborati, per tutti gli elementi sottoposti alla simulazione e dei file contenenti i parametri ad essi correlati. Se la simulazione interessa un complesso ligando-recettore, come in questo caso, è necessario effettuare il calcolo della carica totale del ligando utilizzando il programma antechamber, fornendogli come input il file del solo ligando estratto dal complesso. A questo punto, con il programma parmchk è possibile creare il file contenente i parametri di simulazione relativi al ligando, a partire dall'output di antechamber. Questi due programmi fanno parte della suite di AMBER ed è possibile accedervi dalla modalità non grafica.

Una volta creati i parametri relativi al ligando è possibile effettuare l'elaborazione del complesso proteina-ligando attraverso l'interfaccia grafica di AMBER, xleap. Dall'interfaccia devono essere innanzitutto caricati i force fields su cui saranno basati i calcoli della dinamica, il file dei parametri del ligando e quello contenente le coordinate della sola proteina estratta dal complesso (Fig. 8). A questo punto deve essere calcolata la carica totale della proteina in modo tale da poterla neutralizzare con l'aggiunta di un numero opportuno di ioni  $\text{Na}^+$  o  $\text{Cl}^-$  a seconda del valore e del segno della carica rilevata. Successivamente è necessario creare un box di acqua che circondi la proteina (normalmente di raggio 10 Å dalla proteina stessa). Come risultato dell'elaborazione del complesso proteina-ligando vengono infine creati un file contenente le coordinate dell'intero complesso (file.inpcrd) ed uno contenente tutti i relativi parametri

(file.prmtop): questi file sono fondamentali per la realizzazione di tutte le simulazioni di dinamica molecolare.



```
XLEAP: Universe Editor
File Edit Verbosity
> charge prot
Total unperturbed charge: 2.000002
Total perturbed charge: -2.000002
> addions prot Na+ 2
Adding 2 counter ions to "prot" using 1A grid
Used default radius 1.50 for 4 atoms
Grid extends from solute vdw + 1.07 to 7.07
Resolution: 1.00 Angstrom.
grid build: 0 sec
(no solvent present)
Calculating grid charges
charges: 3 sec
Placed Na+ in prot at (-33.76, 0.44, -13.16).
Placed Na+ in prot at (-42.76, 32.44, -3.16).

Done adding ions.
> charge prot
Total unperturbed charge: -0.000002
Total perturbed charge: -0.000002
> solvatebox prot TIP3PBOX 10
(using default radius 1.500000 for F)
(using default radius 1.500000 for ClP)
(using default radius 1.500000 for ClD)
(using default radius 1.500000 for ClE)
Solute vdw bounding box: 49.289 49.609 86.207
Total bounding box for atom centers: 59.289 59.609 86.207
Solvent unit box: 18.774 18.774 18.774
(using default radius 1.500000 for F)
(using default radius 1.500000 for ClP)
(using default radius 1.500000 for ClD)
(using default radius 1.500000 for ClE)
Total vdw box size: 72.535 72.616 89.516 angstroms.
Volume: 471735.918 A^3
Mass > 200010.520 amu, Density > 0.010 g/cc
(type - hence mass - of one or more atoms could not be found)
Added 11003 residues.
> .
```

Fig. 8. Elaborazione del complesso proteina-ligando dall'interfaccia grafica xleap.

## 2.5.2 Minimizzazione e simulazioni di dinamica

Per effettuare una minimizzazione e/o una simulazione di dinamica molecolare con AMBER sono necessari i file .inpcrd e .prmtop di input ed un file contenente i parametri relativi al calcolo.

I parametri di calcolo relativi ad una minimizzazione sono i seguenti:

- ❖ Imin: permette di scegliere se effettuare una minimizzazione o una simulazione di dinamica.
- ❖ Maxcyc: numero complessivo di cicli di minimizzazione
- ❖ Ncyc: numero di cicli di minimizzazione preliminari (compiuti con un forcefield meno severo)
- ❖ Ntb: permette di scegliere se eseguire la minimizzazione a volume o a pressione costante
- ❖ Ntr: da attivare se si vogliono impostare dei constraints
- ❖ Cut: permette di impostare la zona entro cui considerare gli atomi non legati

I parametri relativi ad una simulazione di dinamica molecolare sono più numerosi e complessi. La durata di una simulazione viene impostata specificando il numero di step di dinamica da effettuare, della durata di 0.002 ps ciascuno. Come nelle minimizzazioni anche nelle dinamiche è possibile impostare dei constraints che permettono di mantenere fisse le coordinate di alcuni atomi degli elementi analizzati.

È possibile effettuare la simulazione mantenendo costanti il volume del sistema o la pressione; la temperatura può essere fatta variare progressivamente oppure mantenuta ad un valore stabilito. In base alle condizioni scelte è necessario impostare parametri diversi con informazioni compatibili con la scelta effettuata. Normalmente si mira a realizzare una dinamica molecolare a temperatura e pressione costanti. I primi step di simulazione sono però condotti a basse temperature, in quanto è necessario innanzitutto fornire calore al sistema, e in queste condizioni il calcolo della pressione del sistema è abbastanza inaccurato: la scelta di effettuare questa prima fase di riscaldamento a pressione costante può quindi comportare dei problemi. Per questo motivo la fase iniziale della simulazione, in cui la temperatura del sistema viene fatta passare da 0°K a 300°K, viene condotta a volume costante. Una volta che il sistema si è equilibrato ed ha raggiunto una temperatura stabile intorno a 300°K è possibile passare a fasi successive di simulazione da condurre a pressione e temperatura costanti.

## CAPITOLO 3

### PARTE SPERIMENTALE

La BTK negli ultimi anni ha ispirato grande interesse nella comunità scientifica, pertanto lo scopo di questa tesi è stato quello di cercare nuovi inibitori di questo enzima attraverso lo studio di virtual screening di database di composti esistenti in commercio.

La metodologia di ricerca di tali molecole ha fatto uso di diversi programmi, via via sempre più selettivi, che ci hanno permesso partendo da un database commerciale di quasi centinaia di migliaia di composti, di selezionare un ristretto numero di molecole con buona probabilità di essere inibitori della BTK da acquistare e sottoporre infine ai saggi enzimatici per verificarne l'attività.

Attraverso l'utilizzo di composti noti in letteratura e analizzando le relative strutture cristallografiche, si è deciso di prendere in considerazione solo 3 delle 9 strutture cristallografiche note, in quanto la sovrapposizione di esse era quasi perfetta, e i relativi inibitori interagivano con la conformazione attiva dell'enzima.

Utilizzando il database Allvondor, all'interno del quale sono presenti circa 780.000 composti, si è eseguito un primo screening con il programma ROCS e un secondo screening col programma FLAP, impiegando approcci differenti, sia ligand-based che receptor-based. I risultati sono stati poi sottoposti al docking utilizzando come programma GOLD 5.0 e infine le molecole selezionate sono state poste all'analisi di dinamica molecolare (MD).

L'utilizzo della dinamica molecolare come ultimo passaggio è dettato dai lunghi tempi di calcolo che essa adotta, quindi ponendola come ultimo step si cerca di assegnare un numero ridotto di composti.

Le molecole risultanti dalla MD sono state poi visualizzate, selezionate e infine acquistate per essere saggate attraverso saggi enzimatici.

### 3.1 ANALISI STRUTTURE CRISTALLOGRAFICHE

L'analisi delle strutture cristallografiche è stata eseguita con il programma di visualizzazione Chimera. Dalla sovrapposizione dei nove cristalli, si è evidenziata una quasi totale sovrapposizione dei complessi riguardanti la conformazione attiva dell'enzima e quindi un elevato grado di similarità.

Tali strutture hanno codice PDB 3K54, 3PIZ e 3PJ1. L'unica differenza tra loro è la presenza della tirosina 551 fosforilata nella struttura 3K54, mentre gli altri due complessi sono stati ottenuti con l'enzima non fosforilato.

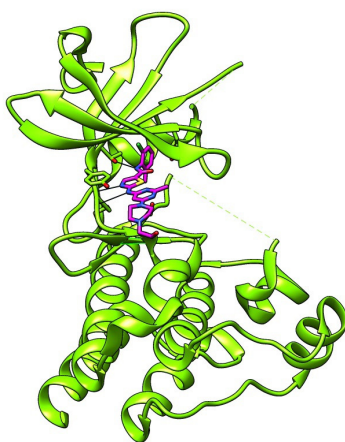
Tutte e tre le strutture sono caratterizzate dall'elica C in posizione "in", dalla sequenza aminoacidica DFG anch'essa in posizione "in", e dal loop di attivazione orientato verso l'esterno.

Tutti i complessi leganti la conformazione inattiva dell'enzima non sono stati presi in considerazione per lo sviluppo di questa tesi.

La struttura cristallografica 3K54 contiene un noto antileucemico, il Dasatinib ( $IC_{50} = 5\text{nm}$ , Figura 1).

Lo scheletro della proteina del Dasatinib presenta la definizione di un loop, posizionato in modo tale da chiudere e definire meglio il sito di legame dell'ATP.

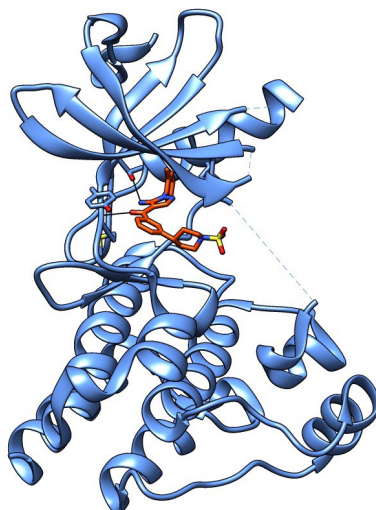
Il Dasatinib fa due legami idrogeno nella zona cerniera interagendo con il carbonile e con l'ammide e occupa una tasca idrofobica dietro il residuo Thr474 gatekeeper con il quale forma un ulteriore legame a idrogeno.



**Figura 1** : complesso cristallografico della BTK con il Dasatinib.

La struttura cristallografica appartenente al codice PDB 3PIZ, contiene il ligando 15Q\_3 (composto numero 3 dell'articolo Kuglstatter del 2010) ( $IC_{50} = 3,5 \mu M$ , Figura 2).

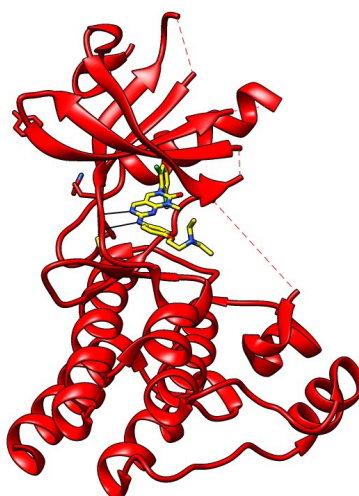
Forma due legami idrogeno nella *zona cerniera* della BTK e uno con la catena laterale del residuo gatekeeper T474.



**Figura 2** : complesso cristallografico della Btk con il 15Q\_3.

La struttura cristallografica con codice PDB 3PJ1, contiene il ligando 15Q\_4 (composto 4 dell'articolo Kuglstatter del 2010) ( $IC_{50} = 48 \text{ nM}$ , Figura 3).

Il composto 4 si lega alla *zona cerniera* della BTK attraverso la sua porzione aminopirimidinica formando 2 legami ad idrogeno.



**Figura 3** : complesso cristallografico della Btk con il ligando 15Q\_4.

Il punto di partenza per lo sviluppo dell'elaborato è stato quello di analizzare i relativi legami che questi tre ligandi formano nella *zona cerniera* dell'enzima, la scelta nell'utilizzare questi tre ligandi è stata dettata innanzitutto dal fatto che legavano la conformazione attiva della BTK ma anche tenendo in considerazione che due di esse formano un legame col residuo gatekeeper T474, il ché, come spiegato, poteva costituire un elemento di selettività.

Tutti e tre i ligandi formano due legami a idrogeno con l'enzima a livello della hinge region, essi inoltre presentano tutti una porzione idrofila che è esposta al solvente e una porzione lipofila che si posiziona nella back-pocket. Nei risultati finali quindi, ci si dovrebbe aspettare di ottenere molecole che replicano questo modo di interagire con la proteina.

### **3.2 SCREENING DATABASE ALLVENDOR**

Il database Allvendor contiene circa 780.000 molecole. Tale database è stato sottoposto all'analisi conformazionale con OMEGA2, impostando una generazione massima di 250 conformeri per composto e un limite di tempo di 600 secondi per generarli.

Le strutture ottenute sono state sottoposte ad uno screening ligand-based con ROCS, secondo il metodo della sovrapposizione shape-based dei tre ligandi presi in considerazione e contenuti nelle tre strutture cristallografiche precedentemente descritte. ROCS utilizza un metodo di allineamento delle strutture attraverso un processo di ottimizzazione di esse, quindi attraverso la somiglianza delle forme; inoltre tiene in considerazione anche composti che sono simili a livello elettrostatico.

Per ognuno dei tre composti di riferimento è stato effettuato uno screening del database Allvendor mirato all'identificazione di molecole ad essi strutturalmente simili, che potessero presentare le caratteristiche necessarie per poter interagire con il sito catalitico della BTK.

Come valore decisivo nella selezione delle strutture da portare nei successivi step, è stato impostato quello relativo al Tanimoto Combo, esso è il parametro che tiene in considerazione sia il livello di sovrapposizione della forma delle molecole analizzate



rispetto al composto di riferimento, sia la loro somiglianza dal punto di vista elettrostatico, (Cap. 2 pag. 27) e può assumere un valore minimo di 0 e un valore massimo di 2; con 0 si indica una cattiva sovrapposizione mentre con 2 si indica una sovrapposizione ottimale.

Si sono selezionate le strutture con valore maggiore o uguale a uno, andando quindi ad escludere quelle strutture la cui sovrapposizione e quindi similarità nella forma e nelle caratteristiche elettrostatiche fosse lontana da quella dei tre ligandi noti in letteratura.

I composti selezionati attraverso i tre diversi screening sono stati poi confrontati tra loro in modo tale da identificare i duplicati, che sono poi stati eliminati.

Questo studio ha portato come risultato a 1143 molecole per lo screening col Dasatinib, 3159 molecole per lo screening col 15Q\_4 e 857 strutture per quello col 15Q\_3, per un totale di 5159 strutture complessive.

### **3.3 STUDI STRUCTURE-BASED CON FLAP**

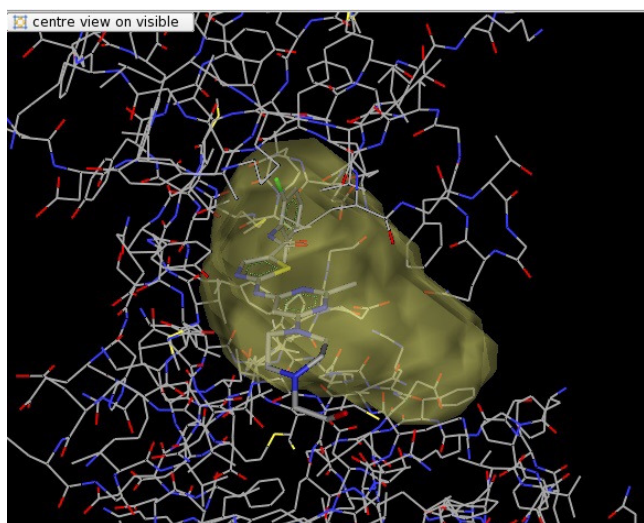
Avendo comunque un elevato numero di molecole ottenute dallo screening precedente, si è pensato di procedere con un'ulteriore selezione, utilizzando un pre-filtro receptor-based col programma FLAP, quindi si è cercato di filtrare ulteriormente le strutture attraverso una migliore possibilità nell'interazione con la proteina .

L'intero processo è basato sull'affinità di una molecola per un sito recettoriale, e dunque sulla natura delle interazioni che possono essere tra loro stabilite.

Il sito recettoriale preso in considerazione per questo screening è quello corrispondente al sito catalitico della BTK e come riferimento è stata utilizzata la struttura cristallografica in complesso con il Dasatinib, in quanto durante gli studi di validazione è risultata quella che dava dei migliori risultati nel posizionamento iniziale dei tre ligandi di riferimento, che mantenevano appunto le interazioni fondamentali.

### 3.3.1 SCREENING E RISULTATI

Dopo validazione si è proseguito con la creazione del database contenente le 5159 molecole derivanti dallo screening precedente e indicandole come decoys (molecole non attive), più le nostre 3 molecole di riferimento (Dasatinib,15Q\_3 e 15Q\_4) indicandole come attive, quindi il database complessivo era formato da 5162 molecole. Abbiamo avviato il relativo calcolo receptor-base utilizzando come pocket di riferimento la tasca occupata dal volume del Dasatinib . (Figura 4)

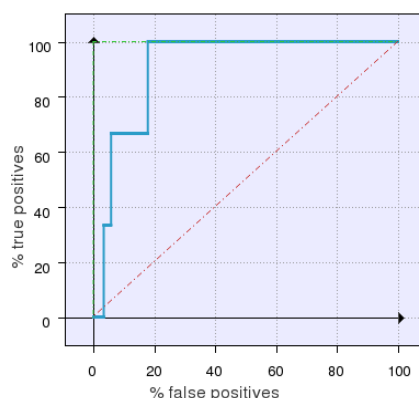


**Figura 4:** Pocket di riferimento utilizzata per lo studio del receptor-based .

Alla fine del calcolo, si sono dovuti analizzare i vari enrichment plot per decidere quale scoring function poteva essere giudicata migliore sulla base della AUC (area sotto la curva) ,che è indice della performance dello screening, e dei MIFs rispetto ai quali viene calcolato il valore di score stesso.

N1\*DRY è risultata la migliore tra le scoring function, poiché tiene in considerazione sia l'analisi delle possibili interazioni di Van der Waals,  $\pi$ -stacking e forze di repulsione (DRY), sia dei donatori di legami ad idrogeno presenti nelle strutture (N1).

Selezionando secondo questa scoring function, le molecole si sono così ridotte a 923, considerando conformi tutte le strutture con valore di score superiore a quello dell'ultimo ligando attivo presente nel database. (Figura 5 )



**Figura 5** :Enrichment plot del virtual screening . ScoringFunction N1\*DRY

Da questa immagine si può notare come l'ultimo ligando noto in letteratura e quindi definito come attivo, sia presente al di sotto del 20 % del numero totale di molecole che sono state soggette allo screening. Importante osservazione questa poichè ci ha permesso di sottoporre allo step successivo una ristretta percentuale di molecole con elevata affinità per il sito recettoriale.

### 3.4 STUDI DI DOCKING

#### 3.4.1 GOLD

Le 923 strutture ottenute dallo screening receptor-base di FLAP, sono state sottoposte al docking col programma GOLD, utilizzando ancora una volta come proteina, quella complesso BTK-Dasatinib.

GOLD ci fornirà una probabile conformazione attiva per ciascuna struttura risultante dall'analisi receptor-based con FLAP. Questi due programmi seguono due approcci completamente diversi nella disposizione del ligando all'interno del sito recettoriale.

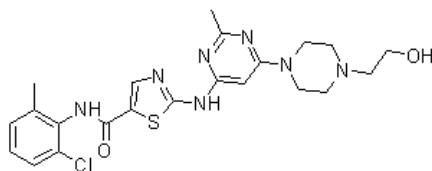
Il metodo di calcolo che utilizza FLAP si basa sull'opportuna sovrapposizione dei tetraedri farmacoforici e atomici di ligandi e sui siti proteici da analizzare, GOLD invece analizza il complesso proteina-ligando in termini energetici.

È importante dire che GOLD restituisce sempre un possibile posizionamento per ogni molecola all'interno del sito di legame, anche se la posizione della struttura è poco precisa, questo sta a significare che il numero di file di output sarà uguale a quello di input.

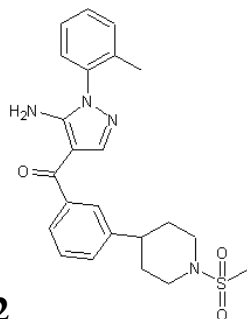
### 3.4.2 DOCKING PRELIMINARE

In una prima fase dovevamo capire se il programma GOLD era in grado di riprodurre la conformazione attiva delle molecole all'interno del sito recettoriale. Abbiamo quindi deciso di processare solo i tre ligandi attivi iniziali.

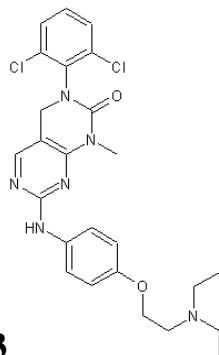
Riporto per semplicità le loro strutture nella Figura 6 .



**1**



**2**



**3**

**Figura 6 :** Struttura chimica del Dasatinib (1), 15Q\_3 (2), 15Q\_4(3)

Per fare questa prima analisi preliminare abbiamo utilizzato l'interfaccia grafica di GOLD, impostando come funzione di scoring la PLP. È stato dimostrato, in seguito a studi condotti nel nostro stesso laboratorio, come quest'ultima sia la funzione migliore per quanto riguarda le protein chinasi. I risultati sono stati visualizzati mediante il programma Chimera. In effetti, essi replicavano molto bene la posizione di legame dei composti cristallizzati e mantenevano tutte le interazioni ad idrogeno ritenute fondamentali. Si è effettuato un cross-docking dei tre ligandi e delle tre proteine.

Ciascun ligando è stato sottoposto al docking nelle proteine degli altri due ligandi, in maniera tale da verificare innanzitutto se il programma restituiva la pose effettiva del ligando all'interno del complesso e qualora questo fosse stata buona, poter processare le 923 strutture in tutti e tre i complessi contenenti i tre ligandi noti.

Dalla visualizzazione e dallo studio attento dei risultati di questo cross-docking è stato subito visibile come i tre ligandi fossero stati processati bene solo nella proteina complessata con il Dasatinib, ed è per questo che si è utilizzata essa nel docking con le 923 strutture.

### 3.4.3 DOCKING E RISULTATI FINALI

Le molecole ottenute mediante screening receptor-based con FLAP, sono state poi processate con GOLD. Il calcolo è stato lanciato da interfaccia non grafica utilizzando il file gold.conf generato durante il docking preliminare. I risultati così ottenuti sono stati in seguito visualizzati e analizzati singolarmente con Chimera, ponendo come parametro discriminante, le strutture in grado di stabilire almeno due dei tre legami ad idrogeno nella *zona cerniera*.

L'analisi ha portato all'ottenimento di 58 strutture, un numero ragionevole di composti da essere sottoposti allo studio di dinamica molecolare.

## 3.5 STUDI DI DINAMICA MOLECOLARE

Lo studio di dinamica molecolare consente di verificare la stabilità del complesso fornendogli energia. A differenza del docking, dove si considera il recettore come un'entità completamente immobile, nella dinamica molecolare anche l'enzima ha un certo grado di libertà conformazionale, che possiamo impostare noi variando determinati parametri. È un calcolo molto accurato che richiede tempi macchina lunghi, per questo motivo viene utilizzato in una delle ultime fasi del protocollo, in maniera da poter fornirgli un numero adeguato di molecole. Inoltre ha bisogno di una posizione di

legame predefinita per ciascun ligando. Viene quindi eseguita, quasi sempre, in seguito alla procedura di docking. Il processo di simulazione si divide fondamentalmente in due parti: la prima, costituita da due step di minimizzazione, la seconda, costituita da uno step di dinamica molecolare vera e propria. Ciascuno di questi step è caratterizzato da particolari restrizioni applicate sulla struttura molecolare e da una durata ben precisa:

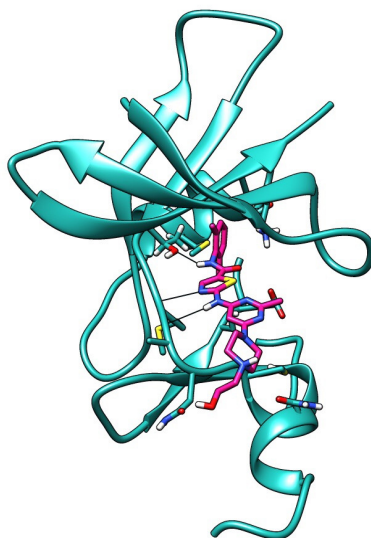
- Primo step di minimizzazione, in cui viene permesso solo il movimento delle molecole d'acqua del solvente (tutti gli atomi del complesso vengono mantenuti fissi);
- Secondo step di minimizzazione, in cui vengono mantenuti fissi solo i carboni  $\alpha$  della proteina (in modo da minimizzare il ligando e le catene laterali);
- Primo step di dinamica vero e proprio, della durata di 0,3 ns, a volume costante, in cui viene fornita energia al sistema mantenendo fissi i carboni  $\alpha$  della proteina (la temperatura passa progressivamente da 0°K a 300° K)
- Secondo step di dinamica di 0,7 ns a pressione e temperatura costanti ( $P=1$  atm e  $T=300^\circ\text{K}$ ). I carboni  $\alpha$  rimangono fissi.
- Terzo ed ultimo step di dinamica, della durata di 2 ns. I carboni  $\alpha$  vengono mantenuti fissi.

### 3.5.1 STUDI PRELIMINARI

Prima di procedere alla dinamica delle molecole provenienti da database commerciali, è necessario capire se il nostro complesso iniziale, di cui possediamo la struttura cristallografica, rimanga stabile in seguito ai vari passaggi di simulazione. Delle tre strutture di partenza si è scelto il complesso della BTK con il dasatinib. Lo abbiamo sottoposto ai due step di minimizzazione e quindi allo step di dinamica. Analizzando il risultato con il programma Chimera, è risultato evidente come il dasatinib mantenesse la sua posizione di legame praticamente invariata all'interno del sito recettoriale, senza perdere nemmeno una delle tre interazioni ad idrogeno iniziali.

Per alleggerire il calcolo, la catena amminoacidica del nostro enzima è stata ridotta mediante l'eliminazione di alcune sequenze che non erano ritenute importanti per

l'interazione con i ligandi. Il nuovo complesso è stato a sua volta sottoposto a minimizzazione e a cinque ns di dinamica. Ancora una volta il sistema è risultato essere stabile (Figura 7).



**Figura 7 :** Complesso BTK-Dasatinib al termine della simulazione di dinamica di prova.

### 3.5.2 DINAMICA E RISULTATI FINALI

La dinamica molecolare è stata condotta su tutte le molecole selezionate in seguito all'analisi dei risultati di docking, per un totale di 58 composti. Abbiamo utilizzato la struttura enzimatica "alleggerita" (da cui erano state eliminate alcune sequenze amminoacidiche non importanti nell'interazione coi ligandi). La creazione dei parametri e dei file di input necessari ad AMBER per effettuare i calcoli è stata realizzata attraverso una procedura automatizzata. Inoltre, al fine di velocizzare ulteriormente il calcolo, per ciascun ligando è stato condotto un solo step di minimizzazione (il secondo) e lo step di dinamica di 2 ns.

I risultati sono stati analizzati sulla base dell'RMSD medio tra la pose del ligando durante l'intera dinamica e la sua pose iniziale, ovvero quella proposta dal docking. Valori di  $\text{RMSD} \leq 2\text{\AA}$  indicheranno come il ligando non si sia allontanato dal sito

recettoriale e abbia mantenuto abbastanza fedelmente il binding mode di partenza. Dopo i due nanosecondi di simulazione sono risultati sufficientemente stabili 41 complessi. Abbiamo analizzato i risultati con il programma di visualizzazione Chimera. Sono state prese in considerazione tutte le molecole che mantenevano almeno 2 interazioni ad idrogeno fondamentali con la zona d'interesse della proteina per un tempo maggiore dell'85% dell'intera simulazione. Le molecole risultanti da questo studio risultano essere 10.

Dall'analisi di esse a livello strutturale, cinque possono essere considerate possibili inibitori della BTK, tra loro quattro presentano anche il legame con il residuo gatekeeper e questa ulteriore interazione può rappresentare un punto di partenza per una maggiore selettività nella ricerca degli inibitori per la BTK. Tali composti verranno in seguito acquistate per essere sottoposte a saggi enzimatici per valutarne sperimentalmente l'attività inibitoria nei confronti della BTK.



## CAPITOLO 4

### CONCLUSIONI

La BTK è una tirosin chinasi appartenente alla famiglia delle TEC chinasi, della quale fanno parte anche la BMX, la ITK, la TXK e la stessa TEC. È coinvolta in una serie di vie di trasduzione intracellulari che modulano la differenziazione e la crescita dei linfociti B. Mutazioni a livello della sequenza amminoacidica di questo enzima portano alla comparsa di una grave malattia ereditaria, la XLA (X-linked agammaglobulinemia), caratterizzata da bassi livelli di cellule B ed immunoglobuline. Queste osservazioni hanno fatto crescere un grande interesse nella comunità scientifica riguardo lo sviluppo di inibitori selettivi della BTK come possibili farmaci per la cura di malattie autoimmunitarie caratterizzate da una attività incontrollata dei linfociti B, prima fra tutte l'artrite reumatoide. È stato dimostrato come la BTK sia espressa anche in cellule della linea mieloidea (granulociti e monociti) implicate nella fase di infiammazione cronica dell'artrite reumatoide. Sembra inoltre che l'attivazione di questa chinasi abbia degli effetti antiapoptotici, i quali possono essere alla base dell'insorgenza di leucemie e linfomi. Proprio per questo motivo, la BTK si è rivelata un bersaglio molto interessante nello sviluppo di farmaci per la cura di tali patologie.

La sequenza amminoacidica della BTK presenta delle caratteristiche peculiari che sono molto importanti nella progettazione di ligandi altamente selettivi. Notiamo in particolare la cisteina 481 (C481), che ricorre in questa posizione in sole altre 9 chinasi e la treonina 474 (T474), che funziona da residuo "gatekeeper" ed è presente in circa il 20% di tutte le chinasi umane. Un terzo elemento per la selettività è lo specificity pocket, o H3-pocket, una tasca idrofobica caratteristica di una particolare conformazione inattiva della BTK.

Il nostro studio si è posto l'obiettivo di ricercare nuovi inibitori attivi sulla BTK applicando una procedura di virtual screening su database commerciali.

In letteratura sono riportati una serie di nove complessi cristallografici che rappresentano la BTK in uno stato attivo o inattivo. Lo stato attivo o inattivo dell'enzima è dettato dalla disposizione di un loop detto loop di attivazione e dalla posizione di un'elica detta elica C.

Attraverso il programma di visualizzazione molecolare Chimera è stato possibile allineare e confrontare le sequenze amminoacidiche di tali cristalli. Risulta evidente come i complessi con codice PDB 3K54, 3PJ1 e 3PIZ, i quali rappresentano la conformazione attiva della BTK, si sovrappongono in maniera quasi perfetta. I ligandi dei tre complessi leganti la forma attiva dell'enzima sono stati scelti per gli step successivi.

Lo studio è stato mirato ad analizzare i relativi legami ad idrogeno che questi ligandi formano con la *zona cerniera* dell'enzima, si è ritenuto importante tenere in considerazione che due ligandi su tre interagivano con la treonina T474 del residuo gatekeeper, ciò poteva essere considerato un' ulteriore elemento di selettività nella scelta delle molecole finali. Una delle strutture prese in considerazione è il Dasatinib, un noto antileucemico. Rilevante è stato lo studio del suo complesso (codice PDB 3K54) poiché si è notato da subito avere una definizione completa di un piccolo loop che circonda meglio la tasca di legame dell'ATP.

Dopo validazione, si è sottoposto il database commerciale Allvondor (780.000 strutture) ad uno screening ligand-based col programma ROCS, utilizzando come strutture di riferimento quelle dei tre ligandi leganti la conformazione attiva dell'enzima. Questo step ci ha portato alla selezione di 5159 composti, strutturalmente simili a quelli dei tre ligandi noti poiché il ligand-based è un metodo che si basa sulla sovrapposizione della forma e sulla similarità delle interazioni elettrostatiche.

I risultati ottenuti da questo screening sono stati sottoposti ad un'ulteriore analisi da parte di FLAP, questa volta però utilizzando un metodo receptor-based. L'intero processo è basato sull'affinità di una molecola per un sito recettoriale, e dunque la natura delle interazioni che possono essere tra loro stabilite. In questo studio è stata utilizzata la struttura cristallografica della BTK in complesso con il Dasatinib. L'analisi con FLAP ha portato ad un'ulteriore riduzione del numero di molecole che da 5159 sono passate a 923.

I risultati selezionati sono stati sottoposti ad una procedura di docking con GOLD. L'obiettivo è quello di arrivare a questo step con un maggior numero di molecole più affini per la pocket di legame. Questi due programmi seguono due approcci completamente diversi nella disposizione del ligando all'interno del sito recettoriale. Il metodo di calcolo di FLAP si basa sull'opportuna sovrapposizione dei tetraedri farmacoforici e atomici di ligandi e siti proteici da analizzare, mentre GOLD analizza il

complesso proteina-ligando in termini energetici.

I dati ottenuti dal docking sono stati visualizzati e analizzati tenendo in considerazione solo le molecole che formavano almeno due legami ad idrogeno nella *zona cerniera* dell'enzima e di conseguenza sono state scartate tutte le altre.

L'osservazione ha portato all'ottenimento di 58 strutture , un numero idoneo da sottoporre allo studio di dinamica molecolare, andando così a verificare la stabilità delle molecole all'interno del sito recettoriale. Questo ultimo step è quello che richiede più tempo di calcolo, quindi si cerca di giungere a questo punto con meno strutture possibili.

Dall'analisi dei risultati della dinamica molecolare, i ligandi selezionati si dovranno dimostrare stabili, mantenendo i legami ad idrogeno iniziali formati con la regione cerniera dell'enzima per un tempo superiore all'85% dell'intera simulazione.

Dopo un'accurata analisi strutturale il numero di molecole si è ridotto a cinque. E' importante far notare come quattro di esse hanno come interazione di legame quello con la treonina 474 del residuo gatekeeper.

Le quattro molecole così trovate saranno acquistate e saggiate attraverso saggi enzimatici al fine di valutarne l'attività inibitoria nei confronti della proteina chinasi.

## BIBLIOGRAFIA

[1] Vetrie, D.; Vorechovsky, I.; Sideras, P.; Holland, J.; Davies, A.; Flinter, F.; Hammarstrom, L.; Kinnon, C.; Levinsky, R.; Bobrow, M.; Smith, C. I. E.; Bentley, D. R. The gene involved in X-linked agammaglobulinaemia is a member of the src family of proteintyrosinekinases. *Nature* 1993, 361, 226–233.

[2] Smith, C. I.; Baskin, B.; Humire-Greiff, P.; Zhou, J. N.; Olsson, P. G.; Maniar, H. S.; Kjellen, P.; Lambris, J. D.; Christensson, B.; Hammarstrom, L.; Bentley, D.; Vetrie, D.; Islam, K. B.; Vorechovsky, I.; Sideras, P. Expression of Bruton's agammaglobulinemia tyrosinekinase gene, BTK, is selectively down-regulated in T lymphocytes and plasma cells. *J. Immunol.* 1994, 152, 557–565

[3] Conley, M. E.; Dobbs, A. K.; Farmer, D. M.; Kilic, S.; Paris, K.; Grigoriadou, S.; Coustan-Smith, E.; Howard, V.; Campana, D. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu. Rev. Immunol.* 2009, 27, 199–227.

[4] Chang, B. Y.; Huang, M. M.; Francesco, M.; Chen, J.; Sokolove, J.; Magadala, P.; Robinson, W.H.; Buggy, J. J. The Bruton tyrosine kinase inhibitor PCI-32765 ameliorates autoimmune arthritis by inhibition of multiple effector cells. *Arthritis Res Ther.* 2011 Jul 13;13(4),R115.

[5] Davis, R. E.; Ngo, V. N.; Lenz, G.; Tolar, P.; Young, R. M.; Romesser, P. B.; Kohlhammer, H.; Lamy, L.; Zhao, H.; Yang, Y.; Xu, W.; Shaffer, A. L.; Wright, G.; Xiao, W.; Powell, J.; Jiang, J. K.; Thomas, C. J.; Rosenwald, A.; Ott, G.; Muller-Hermelink, H. K.; Gascoyne, R. D.; Connors, J. M.; Johnson, N. A.; Rimsza, L. M.; Campo, E.; Jaffe, E. S.; Wilson, W. H.; Delabie, J.; Smeland, E. B.; Fisher, R. I.; Braziel, R. M.; Tubbs, R. R.; Cook, J. R.; Weisenburger, D. D.; Chan, W. C.; Pierce, S. K.; Staudt, L. M. Chronic active B-cellreceptor signalling in diffuse large B-cell lymphoma. *Nature* 2010, 463, 88–92.

[6] Marcotte, D. J.; Liu, Y. T.; Arduini, R. M.; Hession, C. A.; Miatkowski, K.; Wildes, C. P.; Cullen, P. F.; Hong, V.; Hopkins, B. T.; Mertsching, E.; Jenkins, T. J.; Romanowski, M. J.; Baker, D. P.; Silvian, L. F. Structures of human Bruton's tyrosine kinase in active and inactive conformations suggest a mechanism of activation for TEC family kinases. *Protein Sci.* 2010, 19, 429–439.

[7] Kuglstatter, A.; Wong, A.; Tsing, S.; Lee, S. W.; Lou, Y.; Villasenor, A. G.; Bradshaw, J. M.; Shaw, D.; Barnett, J. W.; Browner, M. F. Insights into the conformational flexibility of Bruton's tyrosine kinase from multiple ligand complex structures. *Protein Sci.* 2011, 20, 428–436.

- [8] Lin, L.; Czerwinski, R.; Kelleher, K.; Siegel, M. M.; Wu, P.; Kriz, R.; Aulabaugh, A.; Stahl, M.; Activation Loop Phosphorylation Modulates Bruton's Tyrosine Kinase (Btk) Kinase Domain Activity. *Biochemistry* 2009, 48, 2021–2032.
- [9] Shruti Sharma, Gregory Orłowski and Wenxia Song ; Btk Regulates B Cell Receptor-Mediated Antigen Processing and Presentation by Controlling Actin Cytoskeleton Dynamics in B Cells; *J Immunol* 2009; 182:329-339;
- [10] Di Paolo, J. A.; Huang, T.; Balazs, M.; Barbosa, J.; Barck, K. H.; Bravo, B. J.; Carano, R. A.; Darrow, J.; Davies, D. R.; DeForge, L. E.; Diehl, L.; Ferrando, R.; Gallion, S. L.; Giannetti, A. M.; Gribling, P.; Hurez, V.; Hymowitz, S. G.; Jones, R.; Kropf, J. E.; Lee, W. P.; Maciejewski, P. M.; Mitchell, S. A.; Rong, H.; Staker, B. L.; Whitney, J. A.; Yeh, S.; Young, W. B.; Yu, C.; Zhang, J.; Reif, K.; Currie, K. S. Specific Btk inhibition suppresses B cell- and myeloid cell-mediated arthritis. *Nat. Chem. Biol.* 2011, 7, 41–50.
- [11] A.M. Bendele ; BolderPATH, Inc., University of Colorado, MCD Biology, Colorado, USA ; Animal models of osteoarthritis ; *J Musculoskel Neuron Interact* 2001.
- [12] Hegen M; Keith JC Jr; Collins M; Nickerson-Nutter CL; Utility of animal models for identification of potential therapeutics for rheumatoid arthritis. *Ann Rheum Dis.* 2008 Nov;
- [13] [medicinamolecolare.wordpress.com/tag/leucemia-linfatica-cronica-llc](http://medicinamolecolare.wordpress.com/tag/leucemia-linfatica-cronica-llc)
- [14] Wyllie, A. H.; Kerr, J. F.; Currie, A. R. Cell death: the significance of apoptosis. *Int. Rev. Cytol.* 68, 251–306, 1980.
- [15] Korsmeyer, S. J. Regulators of cell death. *Trends Genet.* 11, 101– 105, 1995.
- [16] Cohen, J. J.; Duke, R. C.; Fadok, V. A.; Sellins, K. S. Apoptosis and programmed cell death in immunity. *Annu. Rev. Immunol.*, 10, 267–293, 1992.
- [17] Linette, G. P.; Korsmeyer, S. J. Differentiation and cell death: lesions from immune system. *Curr. Opin. Cell Biol.*, 6, 809–815, 1994.
- [18] D'Aura Swanson, C.; Paniagua, R. T.; Lindstrom, T. M.; Robinson, W. H. Tyrosine kinases as targets for the treatment of rheumatoid arthritis. *Nat. Rev.*

Rheumatol. 2009, 5, 317-324.

[19] Tsukada, S.; Saffran, D. C.; Rawlings, D. J.; Parolini, O.; Allen, R.C.; Klisak, I.; Sparkes, R. S.; Kubagawa, H.; Mohandas, T.; Quan, S. et al: Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell*. 1993, 72, 279-290.

[20] Lou, Y.; Owens, .T .D.; Kuglstatler, A.; Kondru, R. H.; Goldstein, D. M. Bruton's Tyrosine Kinase Inhibitors: Approaches to Potent and Selective Inhibition, Preclinical and Clinical Evaluation for Inflammatory Diseases and B Cell Malignancies *J Med Chem*. 2012, 55, 4539-4550.

[21] Lindstrom, T. M.; Robinson, W. H. A multitude of kinases - which are the best targets in treating rheumatoid arthritis? *Rheum Dis Clin North Am*. 2010, 36, 367-383.

[22] Firestein, G. S. Evolving concepts of rheumatoid arthritis. *Nature* 2003, 423, 356-361.