



Ministero dell'Università e della
Ricerca Scientifica e Tecnologica



Università degli Studi di Palermo

Facolta' di Farmacia

Dottorato di Ricerca in "Scienze Farmaceutiche"
XVII CICLO Triennio 2002-2005

METODI CHEMIOMETRICI NELLO STUDIO DEGLI INIBITORI DI HIV-1: TECNICHE STATISTICHE MULTIVARIATE E MOLECULAR DOCKING

Dottorando:

Dott. Marco Tutone

Docente guida:

Chiar.mo Prof. Anna Maria Almerico

Coordinatore del Dottorato:

Chiar.mo Prof. Gaetano Dattolo

Dipartimento Farmacochimico, Tossicologico e Biologico

INDICE

1. INTRODUZIONE	Pag. 1
2. HIV-1	Pag. 2
3. TRASCRITTASI INVERSA (RT)	Pag. 4
3.1 – Organizzazione molecolare	Pag. 4
3.2 – Inibitori della RT	Pag. 5
3.3 – Mutazioni e resistenze	Pag. 5
4. INTEGRASI (IN)	Pag. 7
4.1 – Organizzazione Molecolare	Pag. 7
4.2 – Inibitori di HIV-1 Integrasi	Pag. 9
5. PROTEASI RETROVIRALE (PR)	Pag. 13
5.1 – Organizzazione molecolare	Pag. 13
5.2 - Inibitori della Proteasi (IP)	Pag. 15
6. SCOPO DEL LAVORO	Pag. 19
7. MATERIALI E METODI	Pag. 20
7.1 – Hardwares e softwares	Pag. 20
7.2 – Analisi QSAR	Pag. 20
7.3 – PCA (Principal Component Analysis)	Pag. 20
7.4 – DA (Discriminant Analysis)	Pag. 21
7.5 - Indice di Similarità	Pag. 21
7.6 – CA (Cluster Analysis)	Pag. 22
7.7 – Regressione Multilineare (BMLR)	Pag. 24
7.8 – Test di Validazione	Pag. 26
7.9 – Docking Molecolare - LigandFit®	Pag. 28
8. RISULTATI E DISCUSSIONE	Pag. 34
8.1 – Analisi multivariata applicata agli NNRTI	Pag. 34
8.2 – Analisi multivariata applicata agli Inibitori della Proteasi	Pag. 42
8.3 – Confronto dei dati di analisi multivariata e docking per NNRTI e inibitori di PR	Pag. 49
8.4 – Considerazioni finali all’analisi comparata tra metodi multivariati e docking	Pag. 55
8.5 – Analisi multivariata applicata agli Inibitori di IN	Pag. 56
9. CONCLUSIONI	Pag. 62
BIBLIOGRAFIA	Pag. 63
Appendice I – Inibitori della Trascrittasi	Pag. 67
Appendice II – Inibitori della Proteasi	Pag. 70
Appendice III – Attività degli inibitori di Integrasi inclusi nello studio	Pag. 75
Appendice IV – Cluster Analysis	Pag. 84
Appendice V – Derivati inclusi nell’analisi QSAR	Pag. 88

1. INTRODUZIONE

Una significativa riduzione di tempo e costi nella scoperta di nuovi farmaci è sempre stato uno dei principali obiettivi della ricerca farmaceutica.

E' noto che¹ è necessario sintetizzare in media circa 5000 composti per arrivare ad avere un solo derivato attivo sul mercato. Con l'avvento della chimica combinatoriale e delle reazioni in fase solida si è raggiunto un punto di svolta nella ricerca, rendendo disponibili per rapidi saggi, sia biologici sia computerizzati (High Throughput Screenings), milioni di composti, sintetizzati in pochi steps, generalmente automatizzati.

Nell'ambito di questo progresso tecnologico anche l'analisi QSAR ed il Docking Molecolare hanno ricoperto, e ricoprono tuttora, un ruolo di primaria importanza. Pur trattandosi di indagini che necessitano a priori della sintesi e della valutazione di molecole potenzialmente attive nei confronti di un bersaglio biologico, queste tecniche hanno permesso di effettuare enormi passi in avanti.

Infatti, si può tranquillamente sostenere che dagli inizi del novecento ai primi anni ottanta è stato scoperto un numero di molecole farmacologicamente attive pari a quello trovato dagli anni ottanta fino ad oggi, cioè da quando la QSAR ed il Docking hanno cominciato ad avere un'enorme influenza sulle fasi di scoperta e sviluppo dei farmaci.

In quest'ambito si inserisce il mio lavoro di tesi, che prevede l'utilizzo di tecniche di QSAR e di Docking Molecolare per comprendere le relazioni tra specifici targets biologici ed i loro potenziali inibitori; in particolare mi sono occupato dello studio di farmaci ad attività *anti-HIV-1*.

2. HIV-1

HIV-1 è sicuramente il più noto fra i retrovirus. Oggi il suo ciclo vitale (Figura 1) è conosciuto nei minimi dettagli, ed i possibili target farmacologici risultano pertanto numerosi. Si possono distinguere degli ENTRY STEPS, che coinvolgono le glicoproteine dell'envelope virale, e dei POST-ENTRY STEPS, che coinvolgono prodotti di geni accessori virali e le proteine cellulari con le quali essi interagiscono.

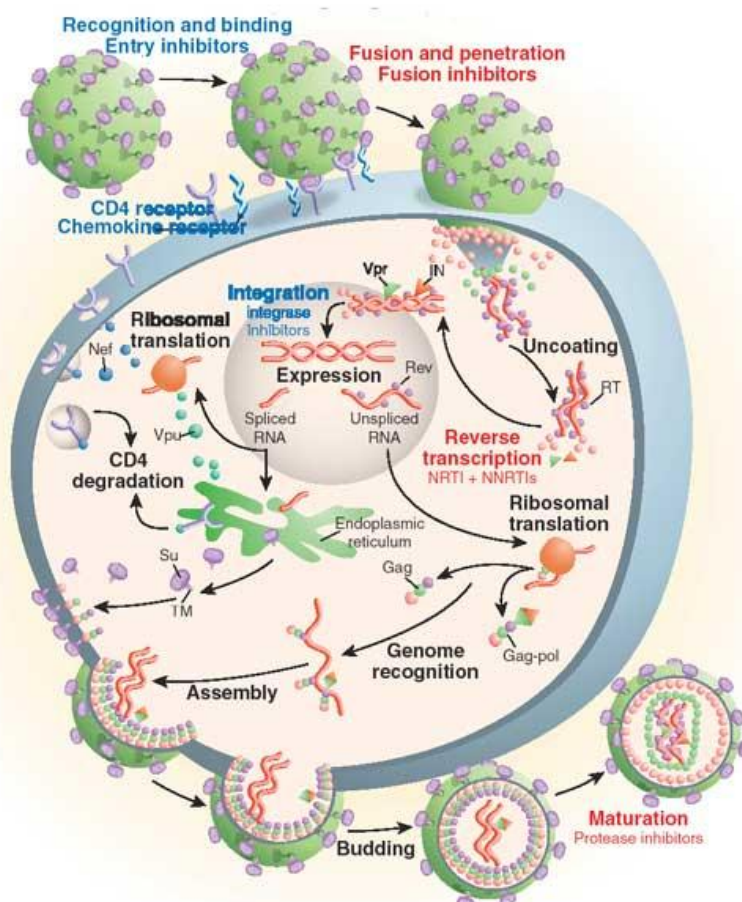


Figura 1. Ciclo vitale di HIV-1

Il passaggio limitante la velocità dell'infezione retrovirale è spesso l'iniziale adesione del virione alla cellula, che avviene prima che le glicoproteine dell'envelope virale interagiscano con specifici recettori della cellula ospite, innescando la fusione. Molti sono i meccanismi proposti; il più credibile di questi implica l'associazione di regioni cariche positivamente sulle glicoproteine virali dell'envelope con proteoglicani carichi negativamente (eparansolfato) presenti sulla superficie cellulare. L'envelope del virione di HIV (native) presenta sulla

superficie un complesso glicoproteico trimerico gp120-gp41, prodotti proteolitici di gp160, un precursore da 160 kDa. In questo complesso, il fattore di fusione gp41 è nascosto; il legame della glicoproteina gp120 al recettore CD4 dei linfociti T-Helper produce dei cambiamenti conformazionali che inducono l'esposizione del loop gp120V3, il quale di conseguenza interagisce con elementi strutturali di un appropriato corecettore, in genere una delle numerose proteine G a 7 domini transmembrana, che funzionano da recettore per le chemochine (CCR5 o CXCR4). (Figura 2)

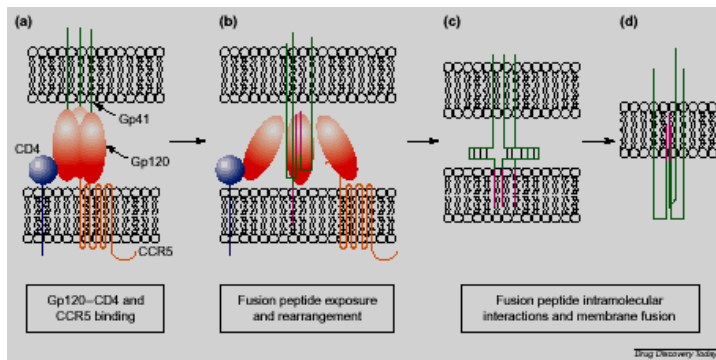


Figura 2. Fusione del virus con il linfocita CD4

Una volta che il virione si è fuso con la cellula ospite, avviene il poco conosciuto fenomeno dell'UNCOATING, cioè il rilascio degli acidi nucleici virali dal nucleo del capsido nel citoplasma della cellula ospite. Dopo questa fase, il virus procede alla trasformazione del proprio genoma da RNA a DNA attraverso un enzima DNA polimerasi-RNA dipendente [Trascrittasi Inversa (RT)] costitutivo del virus. Nel completamento della trascrizione, il cDNA virale (a struttura circolare) sintetizzato è traslocato nel nucleo della cellula, dov'è integrato nel DNA ospite per mezzo dell'enzima Integrasi (IN). Sfruttando poi l'apparato sintetico della cellula, il virus procede alla sintesi di polipeptidi precursori delle proteine strutturali, i quali sono poi processati in proteine funzionanti ad opera della Proteasi retrovirale (PR). L'assemblaggio delle particelle e la maturazione originano virioni pronti ad iniziare un nuovo ciclo di infezione.

3. TRASCRIPTASI INVERSA (RT)

3.1 – Organizzazione molecolare

La Trascrittasi Inversa (RT) è un'eterodimero costituito da due unità polipeptidiche, uno di 66 kDa (p66) e l'altra di 51 kDa (p51). L'unità polipeptidica p51 deriva dalla p66 per scissione proteolitica del dominio C-terminale.

La subunità p66 è costituita da due componenti (p15 e p51) ed è suddivisibile in 5 domini: il dominio della RNAsi-H e 4 sottodomini del dominio polimerasi (pol). La sua conformazione simile a quella di una mano destra, ha portato a definire tre sottodomini (dita, palmo, pollice); il quarto sottodominio pol, trovandosi interposto tra il resto del pol e RNAsi-H, viene chiamato "connection". (Figura 3)

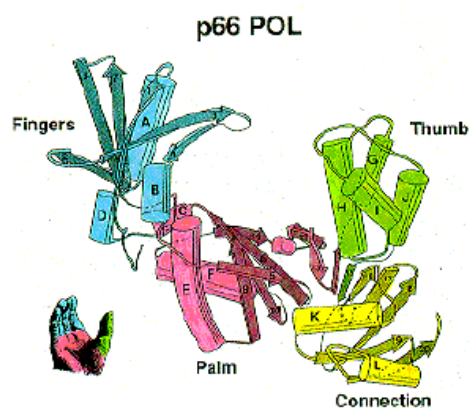


Figura 3. Struttura a "mano destra" di RT

Il ruolo dell'unità p51 non è del tutto chiara poiché solo il dominio pol della p66 catalizza la polimerizzazione. Le due unità polipeptidiche hanno la stessa sequenza aminoacidica, ma nella p51 manca la componente p15 corrispondente al dominio della RNAsi-H nella p66; ciò comporta una struttura secondaria differente e di conseguenza i peptidi equivalenti risultano non attivi. Una tasca di dimensioni sufficientemente grandi per poter accogliere l'ibrido RNA-DNA si trova tra RNAsi-H e il sito attivo pol della p66; alcuni inibitori come la Nevirapina si pongono in una tasca profonda del sito attivo pol senza sovrapporsi ad esso. La p51, diversamente dalla p66, non presenta alcuna cavità ed i residui che si pensa siano coinvolti nel processo catalitico (Asp185, Asp186, Asp110) sono essenzialmente sommersi, rendendo poco chiaro il ruolo di questa subunità. Le differenze più evidenti nella posizione relativa e nei contatti tra domini si riscontrano per il sottodominio "connection". Nella p66 esso è connesso con RNAsi-H con l'estremità del pollice e con il collegamento dell'unità p51, mentre nella

p51 è connesso a tutti gli altri tre sottodomini della p51 stessa così come al collegamento della p66.²

3.2 – Inibitori della RT

A causa dell'importanza di RT nella replicazione di HIV, gli inibitori di questo enzima sono potenziali agenti terapeutici nella battaglia contro il virus. Una classe di inibitori è costituita dagli analoghi nucleosidi come AZT, ddI, ddC, d4T. Questi didesossi composti mancano di un ossigeno in posizione 3' e causano la terminazione della catena del DNA quando vengono incorporati nel filamento crescente di DNA. Un'altra classe di composti che inibisce RT è quella degli inibitori non nucleosidi (NNRTI). E' stato dimostrato che questi inibitori si vanno a legare in una tasca che si trova tra due beta foglietti del palmo della p66, a circa 10 Å di distanza dal sito attivo della polimerasi. La superficie interna di questa tasca è prevalentemente idrofobica essendo primariamente costituita da leucina, valina, triptofano e residui di tirosina. Sebbene gli NNRTI siano composti chimicamente diversi, le strutture cristalline rivelano un modo comune di binding. Studiando il complesso RT-Nevirapina si è dedotto che il segmento da Tyr181 a Tyr188 assume una forma Ω -loop caratterizzata da due catene connesse da almeno due legami idrogeno. Gli NNRTI assumono una conformazione rigida "butterfly-like" che si sovrappone con una conformazione complementare a quella formata dai residui di Tyr181 e Tyr188 del loop.

Sono stati postulati tre possibili meccanismi per l'inibizione di RT da parte di queste classi di composti:

1. possibile alterazione della geometria del sito catalitico;
2. possibile chiusura dei sottodomini palmo e pollice, con riduzione della mobilità del sottodominio pollice;
3. probabile deformazione del punto di aggancio del relativo filamento del sito attivo polimerasico.

Inoltre le interazioni ligando-proteina possono variare da inibitore a inibitore e non sono da trascurare le interazioni con le molecole d'acqua di solvatazione anche se la risoluzione attuale della struttura dei complessi ligando-enzima non permette una precisa collocazione di queste.

3.3 – Mutazioni e resistenze

Inevitabilmente, la rapida replicazione di HIV e la sua innata variabilità genetica porta a generazioni di varianti del virus resistenti ai farmaci.

Anche gli NNRTI portano rapidamente all'emergere di ceppi mutanti di HIV-1, e quindi di resistenza al farmaco,^{3,4} poiché le mutazioni si verificano in porzioni dell'enzima vicine al presunto sito di legame degli NNRTI.^{5,6}

Le mutazioni a livello dei residui aminoacidici che circondano la tasca di binding per gli NNRTI possono dare origine a resistenza al farmaco attraverso meccanismi diversi:⁷

1. determinando la perdita della interazione ottimale tra l'inibitore e la proteina (per esempio in Tyr-181, la mutazione Y181C può determinare una diminuzione dell'interazione di tipo π con gli NNRTI);
2. creando impedimento sterico tra l'inibitore e la proteina (per esempio in Gly-190, le mutazioni possono determinare un'interazione sterica repulsiva con porzioni essenziali degli NNRTI, quale ad esempio l'anello ciclopropile della Nevirapina);
3. inducendo cambiamenti conformazionali importanti nella tasca di binding in seguito ad una qualsiasi mutazione.

Con l'eccezione della mutazione E138K, che si produce nella subunità p51, tutte le mutazioni che determinano resistenza agli NNRTI hanno luogo nella subunità p66 della RT di HIV-1.

4. INTEGRASI (IN)

4.1 – Organizzazione Molecolare

L'integrazione del DNA virale permette al virus di accedere al corredo genetico della cellula infettata; questo evento assicura pertanto che sia la cellula madre che le successive generazioni di cellule figlie esprimano i geni virali, mantenendo pertanto l'infezione allo stato cronico.⁸

L'enzima Integrase (IN) è il responsabile dell'inserimento del genoma virale all'interno del patrimonio genetico della cellula infettata. HIV-1 Integrase è una proteina di 32 kDa che presenta una sequenza altamente simile a quella dei suoi omologhi retrovirali, specie quella di ASV. E' presente *in vivo* sia come monomero, sia come un equilibrio dinamico di dimeri, trimeri ed oligomeri di ordine superiore. L'organizzazione molecolare del polipeptide comprende tre domini funzionali indipendenti. (Figura 4)

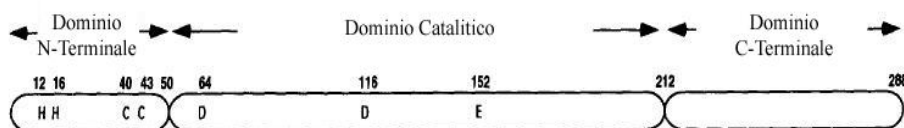


Figura 4. Struttura molecolare di HIV-1 Integrase

Il dominio N-terminale (residui 1-50) è caratterizzato dalla presenza di un motivo *Zinc Finger HHCC* (12,16-40,43), che ricorda molto da vicino il canonico *CCHH* presente in un gran numero di *DNA Binding Proteins (DBPs)*. È stato dimostrato che la regione *HHCC* lega ioni Zn^{2+} stabilizzando pertanto il dominio N-terminale e mediando le interazioni proteina-proteina. Comunque, a differenza della struttura α - β osservata nelle altre *DBPs*, il dominio legante lo zinco di HIV-1 IN presenta una struttura interamente ad α -elica.

Il dominio catalitico di HIV-1 IN si estende dal residuo 51 al 212, e presenta un motivo altamente conservato D,D(35)E, comprendente i residui di Asp64 ed Asp116 separati da Glu152 mediante 35 aminoacidi nella sequenza primaria; la sostituzione di uno di questi residui acidi comporta una perdita irreversibile dell'attività enzimatica, dal momento che molti studi indicano questo dominio come responsabile dell'attività di riconoscimento delle estremità di DNA virale.⁹

Il dominio C-terminale (residui 213-288) è quello che presenta meno analogie tra i vari retrovirus; tuttavia, come risulta dalla perdita di attività dovuta alla sua delezione, è essenziale nel catalizzare la reazione di integrazione. La sua struttura è composta da due β -foglietti antiparalleli, arrangiati in maniera simile alle regioni SH3 (*Src Homology Region 3*), che

mediano la dimerizzazione in soluzione. È stato inoltre dimostrato che il dominio C-terminale possiede capacità non-specifiche di legame con il DNA.

Lo sviluppo di saggi *in vitro* per la reazione di integrazione ha rivelato la necessità di un cofattore metallico, come Mn^{2+} o Mg^{2+} ; tuttavia, a causa della bassa concentrazione intracellulare di Mn^{2+} ($<10^{-7}$ M) rispetto a Mg^{2+} ($\sim 10^{-3}$ M), si tende generalmente ad assumere quest'ultimo come il vero cofattore *in vivo*. Il meccanismo attraverso il quale il cofattore metallico interviene nella reazione catalizzata dall'enzima non è ben definito, sebbene sia ormai assodato che a livello della regione D,D(35)E il cambio conformazionale causato dalla presenza dello ione metallico serve a mantenere il sito catalitico nella giusta configurazione attiva.

Durante l'infezione del linfocita T CD4 Integrasi entra a far parte del cosiddetto "Complesso di pre-integrazione" (PIC, *Pre-Integration Complex*). Questo è costituito da un insieme di proteine che assiste IN nell'integrazione, e che comprende DNA virale, IN, proteine della matrice virale (MA), trascrittasi inversa (RT), proteina del nucleocapside (NC), proteina virale R (Vpr) e proteina intracellulare HMG I(Y).

L'integrazione del DNA virale comprende due singoli passaggi, illustrati nella Figura 5.

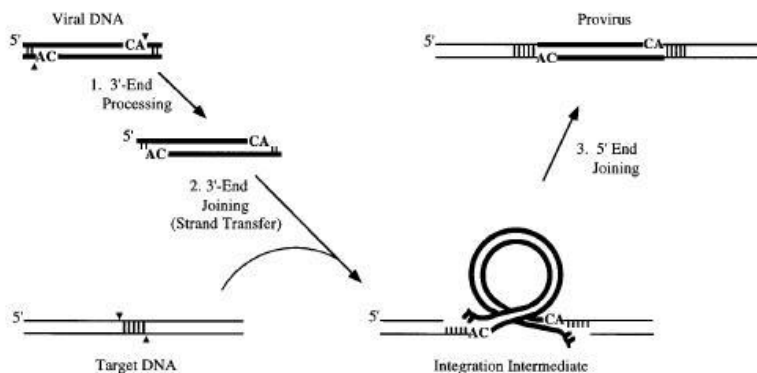


Figura 5. Meccanismo di integrazione del genoma virale

La preparazione del segmento di cDNA virale inizia nel citoplasma, dove IN catalizza il taglio endonucleotidico (*3'-Processing*) all'estremità 3'-terminale di una sequenza conservata 5'-CA-3', esponendo due gruppi -OH in 3'. Successivamente il DNA, con legate tutte le proteine del PIC, viene trasportato all'interno del nucleo, dove avviene la seconda reazione catalizzata da IN, detta di *Strand Transfer*. Questa consiste nell'inserzione, tramite attacco nucleofilo, dei due gruppi -OH liberati durante la reazione precedente, sul genoma della cellula ospite. La reazione avviene in maniera pressoché sequenza-indipendente, e la scelta del sito specifico di attacco sembra essere suggerita da caratteristiche del DNA, quali lo stato

di impaccamento della cromatina e la presenza di DNA Binding Proteins. Dal momento che in posizione 5' il DNA appena integrato risulta privo di due basi, l'apparato cellulare di riparazione del DNA provvede a riparare i gaps formati dall'integrazione (*5'-Gap repairing*).

Una terza attività catalitica di HIV-1 IN, chiamata *Disintegrazione*, è stata descritta *in vitro*; questa reazione è apparentemente opposta a quella di Strand Transfer, e conduce alla separazione del DNA virale precedentemente integrato e del DNA cellulare. Non vi è comunque evidenza di un'attività di disintegrazione catalizzata *in vivo* da Integrasi.

4.2 – Inibitori di HIV-1 Integrasi

L'importante ruolo di IN nel ciclo replicativo di HIV-1 rende la ricerca di inibitori di questo enzima una promettente strategia per combattere AIDS. Lo sviluppo estensivo di tali farmaci è iniziato piuttosto recentemente, e ad oggi sono conosciute relativamente poche classi di inibitori¹⁰, classificabili nelle seguenti categorie (Figura 6):

- Polianioni (I)
- Catecoli e Bis-catecoli (II)
- Derivati dell'acido aurintricarbossilico (III)
- Curcumine e derivati (IV)
- Tirfostine (V)
- Cumarine e derivati (VI)
- Derivati triazinici (VII)
- Derivati chinolinici e stirilchinolinici (VIII)
- Arilamidi (IX)
- Idrazidi e Salicilidrazidi (X)
- Derivati di acidi dichetonici (XI)
- Diarilsolfoni (XII)
- Indani e spirobiindani (XIII)
- Derivati dell'acido cicorico (XIV)
- Vari (XV)

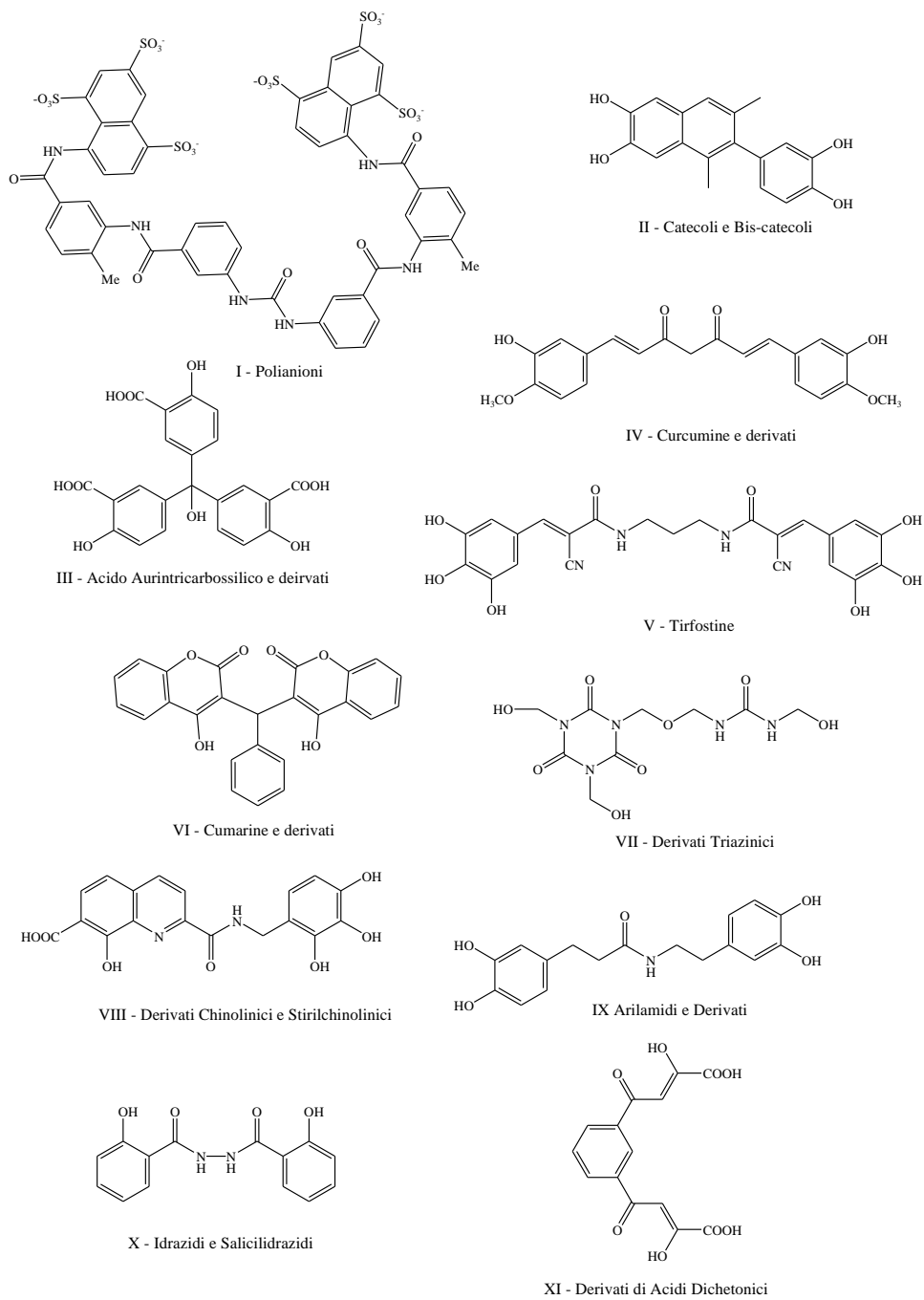
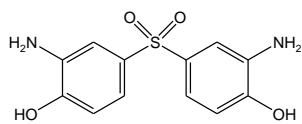
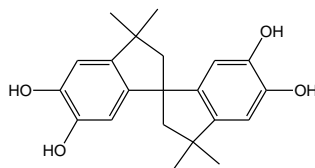


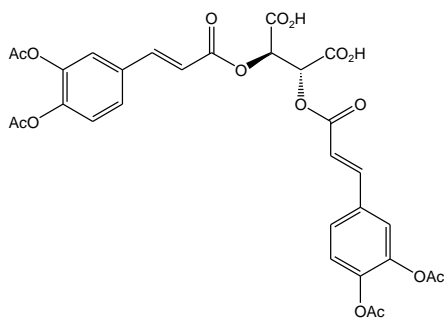
Figura 6. Classi di inibitori di HIV-1 IN



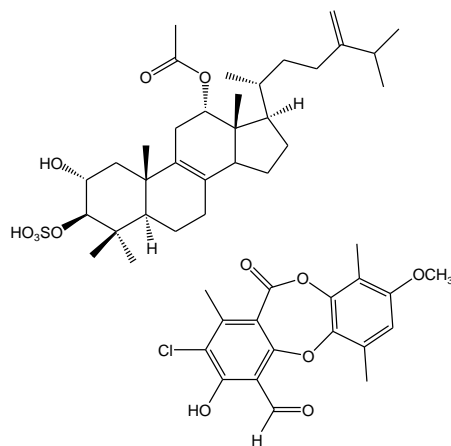
XII - Diarilsolfoni



XIII - Indani e Spirobiindani



XIV - Derivati dell'Acido Cicorico



XV - Vari

Figura 6 (continua)

Sebbene le attuali conoscenze circa la struttura e la funzione di IN siano alquanto limitate (non sono disponibili ad oggi né una struttura intatta della proteina con tutti e tre i suoi domini, né una struttura del complesso enzima/DNA) sono stati effettuati degli esperimenti di docking dimostranti che l'esame dell'architettura del sito attivo può rivelarsi un metodo utile per la progettazione di nuovi inibitori. Secondo i risultati di questi esperimenti¹¹ il sito di legame degli inibitori è formato da Asp64, Cys65, Thr66, His67, Glu92, Asp116, Gln148, Ile151, Glu152, Asn155, Lys156 e Lys159. (Figura 7)

Questo sito è chiaramente dominato da residui polari, e la sfida nella progettazione di nuove strutture consiste nel rendere il più forte possibile il legame degli inibitori sfruttando tutte le interazioni disponibili, specialmente i legami ad H, e minimizzando nel contempo i conflitti. Un sito di primaria importanza, che può servire da ancora per una forte interazione dei ligandi, è dato dalle due lisine e dai residui vicinali. Carbossilati, sulfonati e fosfati sono i gruppi d'elezione per l'interazione a questo livello, ma anche gruppi tetrazolici o catecolici possono legarsi su questo sito; il legame è virtualmente completato dalle interazioni con Thr66 ed Hys67.

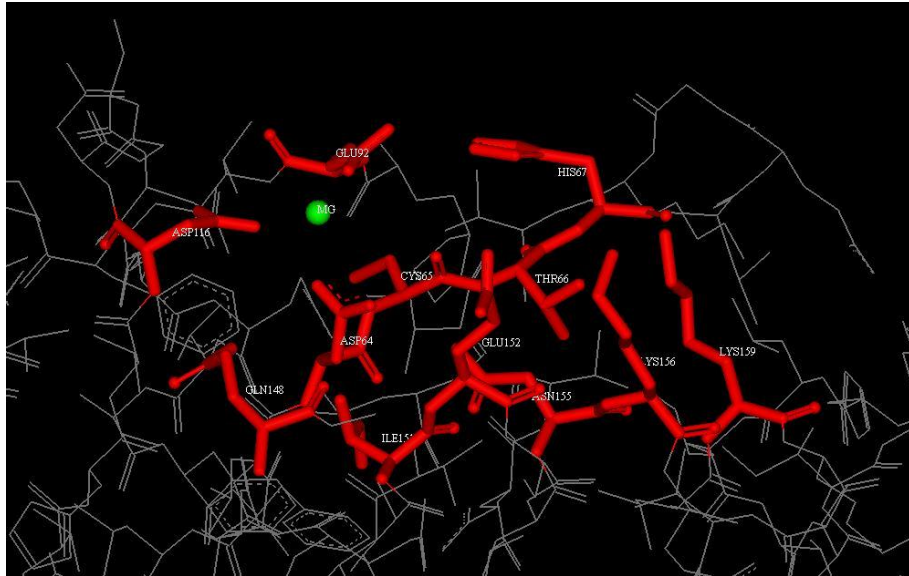


Figura 7. Sito attivo di HIV-1 IN

Cys65 è occasionalmente coinvolta in interazioni favorevoli, e merita una particolare attenzione quando si esplora la regione attorno a Thr66, His67 e Glu92, al fine di prevenire contatti repulsivi con i gruppi carbonilici del backbone proteico. Nella regione attorno allo ione Mg^{2+} , solamente Asp116 sembra essere disponibile, attraverso il suo gruppo carbossilato, ad accettare legami ad H nell'interazione con il ligando. Il gruppo carbossilato del terzo residuo catalitico, Glu152, sembra ottimo per l'accesso a legami, anche se con alcuni ligandi potrebbe rivelarsi troppo esposto, fuori dal sito attivo, per permettere delle interazioni stabili. Infine Gln148 sembra costituire la scelta ideale come secondo sito-ancora per stabilizzare le interazioni di un ligando al sito attivo.

5. PROTEASI RETROVIRALE (PR)

5.1 – Organizzazione molecolare

La prima indicazione che i retrovirus utilizzino una strategia di precursori proteici si è basata sull'osservazione che cellule infettate da AMV (Avian Myeloblastosis Virus) producono una proteina iniziale che è molto più grande di quelle trovate nel virus. Questo precursore è chiamato Gag-protein e include sempre le tre principali proteine del virione: matrice (MA), capsid (CA), nucleocapsid (NC).¹²

Molti retrovirus, compreso HIV-1, iniziano da un precursore Gag per generare proteine mature più piccole. L'esistenza di un precursore retrovirale implica l'esistenza di una proteasi capace di idrolizzare i legami peptidici (nello specifico una aspartato proteasi).

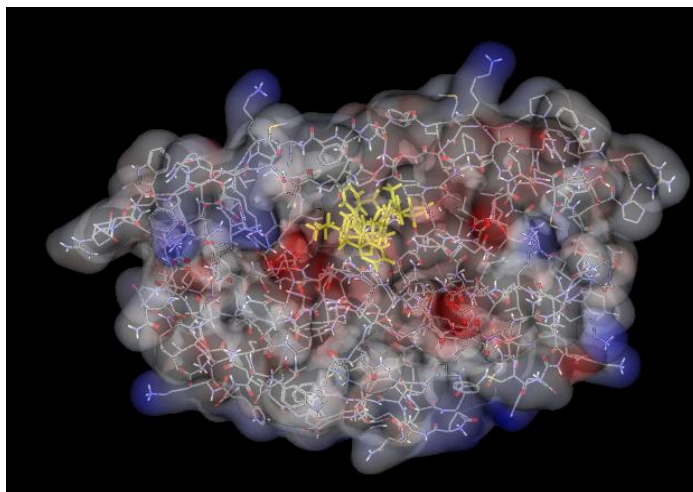


Figura 8. Proteasi Retrovirale

Le proteasi eucariotiche come renina, pepsina e catepsina D sono a singola catena con due motivi Asp-Ser/Thr-Gly nel sito attivo.

Le PR hanno un singolo motivo nella catena perciò per la formazione del sito attivo è necessaria la presenza di una seconda catena proteica. HIV-1 PR è quindi un enzima dimerico formato da due polipeptidi associati non covalentemente. Ogni monomero, costituito da 99 aminoacidi, contiene il gruppo aspartato necessario alla catalisi. Gli aminoacidi sono numerati da 1 a 99 e da 101 a 199, oppure con la notazione 1-1', dove 101 o 1' indicano il primo aminoacido della seconda catena identica alla prima. Studi cristallografici sulla proteasi hanno permesso di mappare con estrema precisione la struttura del sito attivo che presenta:

- Due residui di aspartato (25-25' oppure 25-125) che giacciono sul “pavimento di una crepa del sito attivo” e coordinano una molecola d’acqua utilizzata per idrolizzare il legame peptidico target.
- La crepa costituita da aminoacidi idrofobici che partecipano al riconoscimento del substrato.
- La parte superiore della crepa coperta da un lembo mobile.
- Una regione della proteina di ogni subunità che forma una β -struttura sopra la crepa. Questo lembo però non può muoversi per lasciare entrare il substrato. (Figura 8, 10)^{12, 13}

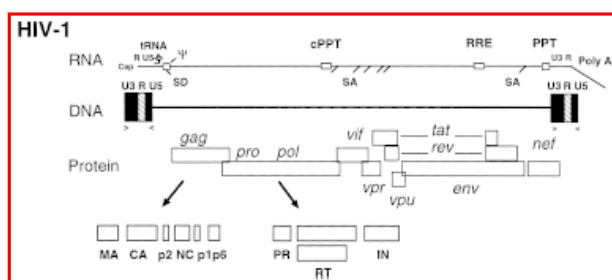


Figura 9. Mappa genica di HIV-1

La PR è codificata dal gene Pro del virus. Il gene Pro di differenti retrovirus si trova sempre tra il gene Gag, che codifica per le principali proteine strutturali del virus, ed il gene Pol, che codifica per la RT e la IN.

Nel caso di HIV-1, i geni Pro e Pol sono nello stesso reading-frame. I reading-frames di Gag e Pro-Pol si sovrappongono per circa 200 nucleotidi. (Figura 9)

Il precursore Gag-Pro-Pol è indirizzato ai siti di assemblaggio-particelle usando il Gag targeting signal. La dimerizzazione di questo precursore conduce all’attivazione di PR. E’ stato proposto che piccole quantità di PR dimerica attiva nel precursore Gag-Pro-Pol rilasciano monomeri di PR che poi dimerizzano e producono svariati cleavage.¹²

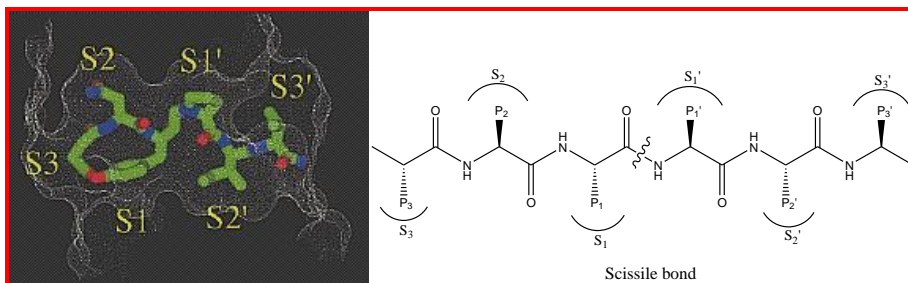


Figura 10. Subsiti di legame nel sito attivo e posizionamento del substrato

Per un efficiente cleavage il substrato della PR deve contenere almeno 7 aminoacidi e legarsi tramite interazioni ioniche, Van der Waals o legami a idrogeno.¹⁴ Analisi di strutture cristalline hanno mostrato che il substrato legato assume una β -struttura, che consente l'interazione con lo scheletro peptidico. Nelle β -strutture le catene laterali dell'inibitore si localizzano nei subsiti di legame presenti nella crepa, assumendo arrangiamento lineare.^{12, 14} (Figura 10)

Il meccanismo di azione della proteasi è piuttosto semplice (Figura 11) ed è coincidente in tutto e per tutto con quello delle altre proteine appartenenti alla famiglia delle proteasi. Una volta inseritosi nel sito recettoriale, il substrato subisce attacco nucleofilo da parte di una molecola d'acqua sul carbonile del legame amidico target mediato da uno dei residui di aspartato del sito catalitico. La protonazione dell'azoto amidico porta ad un intermedio tetraedrico zwitterionico che collassa nei prodotti di idrolisi.

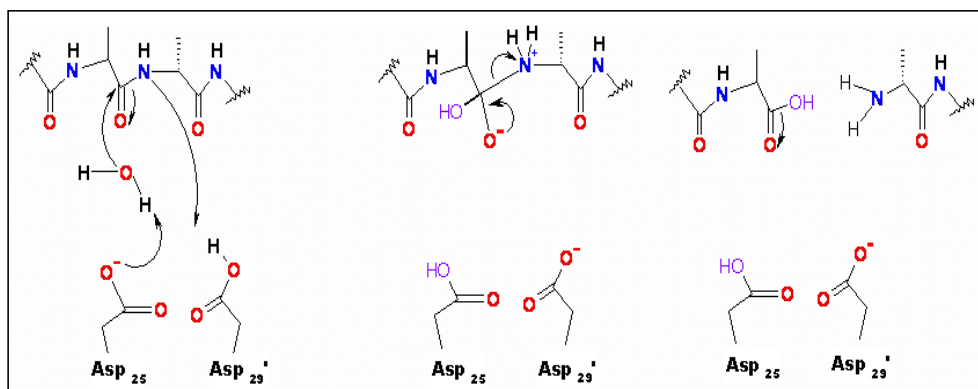


Figura 11. Meccanismo d'azione della proteasi

5.2 - Inibitori della Proteasi (IP)

Sono state studiate e sviluppate numerose classi di composti capaci di inibire la proteasi virale. La maggior parte è stata progettata in modo da presentare una struttura complementare al sito di binding ed è costituita da composti peptidomimetici.

Il punto di svolta delle ricerche per un efficace inibitore è stato però l'introduzione di un gruppo idrossilico nella posizione equivalente al legame amidico. Il gruppo idrossilico spiazza la molecola d'acqua dal sito attivo e coordina i residui di aspartato mimando lo stato di transizione (intermedio tetraedrico).¹² (Figura 12)

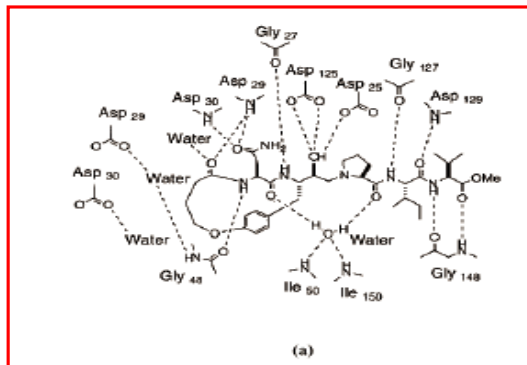


Figura 12. Stabilizzazione di un inibitore nel sito attivo

Tutti gli IP sviluppati sinora, utilizzano questa strategia. Il gruppo idrossilico può essere presente sotto forma di: idrossietilene, idrossietilamina, idrossietilaminosulfonamide. La presenza del sostituente –OH è indispensabile per il meccanismo di inibizione della Proteasi, ma non è sufficiente per la specificità necessaria per legare solo la PR.

Gli inibitori delle Proteasi approvati per uso clinico sono 6 (Figura 13) e presentano tutti le medesime caratteristiche:

1. Hanno grandi catene laterali idrofobiche vicino all'idrossile;
2. Sono più corti della lunghezza del peptide substrato;
3. Hanno caratteristiche di solubilità elevate in quanto uno dei principali problemi dei primi IP era la scarsa biodisponibilità dovuta all'eccessiva lipofilia delle catene laterali;
4. Si legano ad albumina e α -1 glicoproteina acida.¹²

Farmaci 1^ Generazione	Trade Name
Saquinavir	Invirase® / Fortovase®
Indinavir	Crixivan®
Nelfinavir	Viracept®
Amprenavir	Agenerase®
Ritonavir	Norvir®
Lopinavir / Ritonavir	Kaletra®

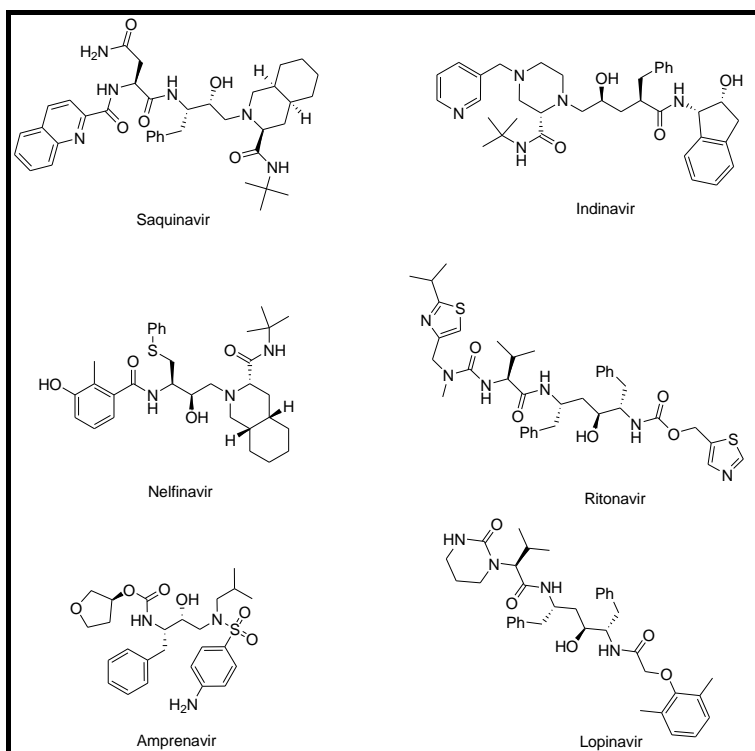


Figura 13. Farmaci approvati per uso clinico a giugno 2004

Uno dei problemi principali connesso con l'uso degli inibitori peptidomimetici è rappresentato da severi side-effects, soprattutto in terapie a lungo termine, a carico dell'apparato cardiovascolare. Ma il problema maggiore connesso all'utilizzo di questi farmaci è dovuto all'insorgere di mutazioni nella PR che conducono alla farmacoresistenza, così come avviene per gli inibitori della RT. Tali mutazioni, dovute all'alta velocità di riproduzione del virus (più di 10^{10} particelle virali prodotte ogni giorno)¹³ e alla totale assenza di attività riparatorie durante la sintesi del DNA-RNA dipendente, possono essere classificate come primarie e secondarie.

MUTAZIONI PRIMARIE: appaiono di solito rapidamente, sono specifiche per ogni farmaco e diminuiscono la sensibilità al farmaco localizzandosi nella tasca di legame riducendo la capacità catalitica dell'enzima;

MUTAZIONI SECONDARIE: solitamente si accumulano nel genoma virale che possiede già mutazioni primarie. Queste mutazioni conferiscono minore o nessuna riduzione nella sensibilità al farmaco, ripristinano le capacità di replicazione del virus che porta mutazioni primarie e possono anche incrementare il livello di resistenza di questi virus. Non compaiono necessariamente nel sito attivo della proteina. A oggi sono state osservate più di 20 mutazioni

(vedi Tabella 1) a seguito di terapia con IP. Per superare il problema la ricerca si è indirizzata verso nuovi inibitori non-peptidici, i quali si caratterizzano per uno spettro di azione più ampio, offrendo anche una migliore biodisponibilità e, probabilmente, minore resistenza.

INIBITORE PROTEASI	MUTAZIONI PRIMARIE	MUTAZIONI SECONDARIE
Indinavir	M46L V82A F T S	L10I/R/V, K20M/R, L24I, V32I, M36I, I54V, A71V/T, G73S/A, V77I, I84V, L90M
Ritonavir	V82A F T S	K20M/R, V32I, I32F, M36I, M46I/L, I54V/L, A71V/T, V77I, I84V, L90M
Saquinavir	G48V, L90M	L10I/R/V, I54V, A71V/T, G73S, V77I, V82A, I84V
Nelfinavir	D30N, L90M	L10F/I, M36I, M46I/L, A71V, V77I, V82A/F/T/S, I84V, N88D
Amprenavir	I50V, I84V	L10F/I/R/V, V32I, M46I, I47V, I54V

Tabella 1. Mutazioni associate con la resistenza agli inibitori della Proteasi utilizzati in terapia

6. SCOPO DEL LAVORO

Il progetto di ricerca, che ho svolto presso il Dipartimento Farmacochimico Tossicologico Biologico, è stato improntato sullo studio degli inibitori dei tre principali enzimi di HIV-1 ossia RT, PR e IN nel tentativo di esplorare le caratteristiche che possono modulare l'attività biologica tenendo anche conto dei fenomeni connessi all'insorgere della resistenza.

Lo studio effettuato in questi tre anni si è articolato essenzialmente in due parti. La prima ha riguardato RT e PR, la seconda IN. Tenendo conto del fatto che ormai da diversi anni sono in commercio inibitori di RT e PR e che, come precedentemente sottolineato, tali farmaci in terapie a lungo termine conducono a fenomeni di resistenza, mi sono proposto di sviluppare un metodo per valutare l'eventuale insorgere di questo fenomeno. Partendo da un database di molecole, descrittori molecolari, indici di similarità e dati di sensibilità e resistenza nei confronti dei più comuni ceppi mutanti (L100I, K103N, V106A, Y181C per RT; V82A, V82F, I84V per PR) costruito utilizzando dati di letteratura, ho applicato dapprima tecniche di analisi statistica multivariata, quali PCA e DA, per sviluppare un modello che potesse spiegare e prevedere l'eventuale resistenza o suscettibilità di molecole contro i ceppi mutanti. Successivamente si è effettuato il confronto dei risultati ottenuti dai metodi statistici multivariati con quelli provenienti da studi di molecular modeling nel tentativo di validare reciprocamente i dati.

Nella seconda parte del mio dottorato mi sono occupato dello studio degli inibitori dell'Integrasi. Considerando che in letteratura è presente un'enorme quantità di dati sugli inibitori di HIV-1 IN, ho cercato di selezionare una serie di strutture rappresentative delle classi di inibitori più efficaci e di costruire dei modelli che descrivano le attività di 3'-Processing e di Strand Transfer in funzione delle loro caratteristiche strutturali. Per IN non sono disponibili in commercio dei farmaci da utilizzare in terapia e di conseguenza non esistono dati riguardanti l'insorgere di mutazioni che conducano alla resistenza, per cui non è ancora possibile sviluppare dei modelli per la valutazione dell'insorgere di resistenza.

7. MATERIALI E METODI

7.1 – Hardwares e softwares

Tutti i calcoli di analisi statistica multivariata e docking sono stati effettuati su workstations della Silicon Graphics (modelli Indigo2 R4400 e Octane2) e su PC con OS WinXP Pro.

I softwares utilizzati sono CS ChemDraw Ultra 7.0, TSAR 3.2 (Tools for Structure Activity Relationships), ASP 3.21 (Automated Similarity Package), VAMP 7.0, CORINA 1.81, CODESSA PRO, Ligandfit®.

7.2 – Analisi QSAR

Lo studio delle relazioni che intercorrono tra la struttura molecolare di un composto e la sua attività si fonda su un approccio razionale basato sull'assunzione che esistano certe relazioni tra la struttura molecolare (S) e l'attività biologica (A) dei composti; si cerca cioè di determinare la funzione $f(S,A)$ che mette in relazione questi due parametri. L'obiettivo finale degli studi QSAR è quello di comprendere i meccanismi dell'azione farmacologica o tossicologica, suggerendo nuove prospettive per la sintesi di composti con attività biologica definita.

Per lo studio delle relazioni QSAR sono state proposte diverse metodologie che si distinguono tra loro sia per l'impostazione generale nell'approccio al problema, sia per le modalità con cui sono descritte le molecole (le diverse tipologie di descrittori), sia per i metodi matematici utilizzati per estrapolarne l'informazione necessaria. Comunque, tutte le strategie utilizzate accoppiano comunemente la rappresentazione delle strutture molecolari mediante opportuni descrittori, alla ricerca di relazioni quantitative tra descrittori ed attività biologica utilizzando principalmente l'analisi statistica multivariata.¹⁵

In particolare sono state utilizzate le tecniche PCA, DA, CA, BMLR

7.3 – PCA (Principal Component Analysis)

La PCA è un metodo che può essere usato per ridurre un grande numero di variabili, senza che ci sia una perdita di informazioni. E' particolarmente utile per:

- ridurre la dimensione di un set di dati, consentendo di visualizzare una grande quantità di informazioni;
- studiare le relazioni tra differenti descrittori;
- preparare i dati per ulteriori analisi.

Quando si esegue la PCA su una serie di dati, le variabili originarie sono trasformate in un nuovo set ortogonale di combinazioni lineari, generato per conservare la varianza dei dati originari.

$$PC_1 = a_{(1,1)}v_1 + a_{(1,2)}v_2 + \dots + a_{(1,n)}v_n$$

$$PC_2 = a_{(2,1)}v_1 + a_{(2,2)}v_2 + \dots + a_{(2,n)}v_n$$

$$PC_q = a_{(q,1)}v_1 + a_{(q,2)}v_2 + \dots + a_{(q,n)}v_n$$

Ogni componente principale (PC) è una combinazione di variabili iniziali (v), definita utilizzando dei loading coefficients (a). La maggior parte della varianza è contenuta nelle prime PC, la prima PC, infatti, di solito, mostra la varianza più grande. Ogni PC ha un eigenvalue (autovalore) associato, che mostra quanta varianza è esplicata da quella PC.

7.4 – DA (Discriminant Analysis)

La DA è una tecnica utilizzata per separare due o più classi di composti usando un numero di variabili esplicative. Le regole di classificazione derivate descrivono una superficie che separa le classi; può essere utilizzata per prevedere l'appartenenza ad una data classe.

TSAR usa il Mahalanobis distance discrimination algorithm.¹⁶ Per default, utilizza una procedura graduale per selezionare un subset di variabili esplicative ed ottimizzare le regole di classificazione. La variabile entrante, ad ogni stadio, è quella che dà il maggiore incremento nella distanza totale di Mahalanobis.

La selezione graduale delle variabili esplicative termina quando lo stopping criterion selezionato è stato raggiunto. In genere si sceglie lo stopping criterion **Automatic** che determina lo stop o quando tutti i punti sono correttamente classificati (nessuna variabile dovrebbe dare più del 5% di incremento nella distanza totale di Mahalanobis), oppure dopo che tutte le variabili sono state sommate.

7.5 - Indice di Similarità

La misura della similarità tra due molecole, basata su proprietà fisiche di lipofilia, forma e potenziale elettrostatico è ottenuta utilizzando un approccio analitico nel quale le funzioni Gaussiane sono sostituite nelle equazioni di calcolo. Le proprietà sono paragonate attraverso lo spazio utilizzando un incremento infinitesimale di griglia. La similarità per shape è stata calcolata utilizzando la modifica di Meyer all'equazione di Carbo. Meyer mostro che l'equazione di Carbo¹⁷ per le densità elettroniche può essere adattata alla misura della

similarità in termini di shape attraverso un semplice algoritmo modificato. Ogni molecola è rappresentata come un sistema di sfere interconnesse con raggio uguale al raggio di Van der Waals, centrate sul nucleo atomico. La coppia di molecole sovrapposte è piazzata in una griglia tridimensionale e una maglia di punti viene esplorata. Alla fine vengono contati:

O₁ numero dei punti della griglia che giacciono solo all'interno della molecola 1

O₂ numero dei punti della griglia che giacciono solo all'interno della molecola 2

B numero dei punti della griglia che giacciono all'interno di entrambe le molecole.

Il numero totale dei punti della griglia in ogni molecola è quindi dato da:

$$T_1 = O_1 + B$$

$$T_2 = O_2 + B$$

L'analogo dell'algoritmo dell'indice di Carbo è

$$S^c_{12} = \frac{B}{(T_1 T_2)^{1/2}}$$

7.6 – CA (Cluster Analysis)

I metodi di CA forniscono informazioni sulla presenza di raggruppamenti, detti *clusters*, utilizzando il concetto di similarità. Esso è la trasposizione matematica del concetto di analogia, e si configura come il complemento del concetto di dissimilarità; infatti per misurare quantitativamente quest'ultimo si può utilizzare il concetto di distanza, definendo come dissimili due elementi posti ad una certa distanza tra loro.

Si possono utilizzare molti tipi di distanze. In questo studio si è scelto di utilizzare la distanza Euclidea, definita come:

$$d_{\text{Euc}} = \sqrt{\sum_{j=1}^p (x_{sj} - x_{tj})^2}$$

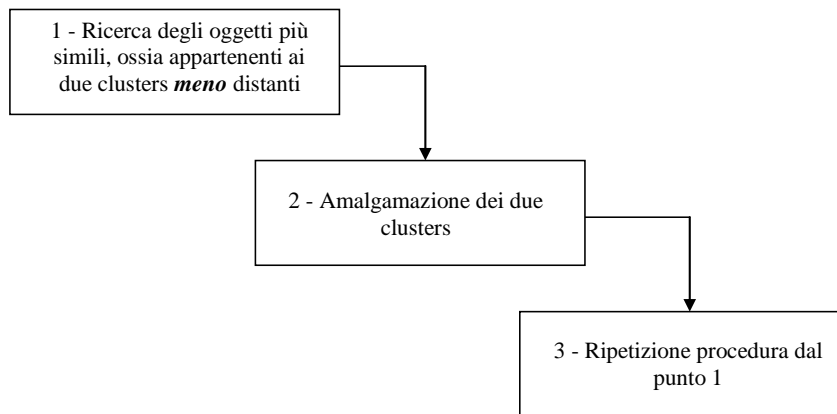
dove x_{sj} è l'elemento s j-esimo della matrice delle distanze.

La *matrice delle distanze* è definita come una matrice n x n (dove n è il numero di oggetti), in cui in ogni riga, rappresentante un oggetto, vi sono tutte le distanze degli altri oggetti da quello considerato (essa è ovviamente una matrice simmetrica, ed i suoi elementi diagonali hanno valore zero, poiché la distanza di ogni oggetto da se stesso è nulla).

Una volta definita la corrispondenza fra distanze e similarità l'algoritmo di clustering trasforma la matrice delle distanze in *matrice di similarità*, i cui elementi diagonali hanno valore uguale ad uno giacché la similarità di ogni oggetto con se stesso è massima.

I metodi di CA sono numerosi, e possono essere suddivisi in metodi *gerarchici*, e *non gerarchici*; i metodi gerarchici, a loro volta, possono essere distinti in *agglomerativi* e *non agglomerativi*. I cluster che ciascun metodo individua sono caratterizzati dalla loro posizione nello spazio p -dimensionale da un *centroide*, definito come il vettore delle medie delle variabili calcolate per gli oggetti assegnati al cluster, e da un *centrotipo*, definito come l'oggetto più rappresentativo di un cluster, in virtù della sua minore distanza dal centroide.

Nella formazione dei cluster, ogni oggetto è inizialmente considerato come facente parte di un cluster formato da un solo membro (se stesso); i cluster vengono in seguito considerati a gruppi, ed esaminata la loro distanza in termini di similarità, vengono infine fusi insieme. L'intero processo di clustering segue la procedura riportata nello Schema 1.¹⁸



Schema 1. Processo di clustering

I metodi di clustering utilizzati comprendono i seguenti:

1. **Ward's Reciprocal Nearest Neighbors (WRNN)** – utilizza l'algoritmo dei Reciprocal Nearest Neighbors (RNN), che comporta la minor perdita possibile di informazione;
2. **Single Linkage (SL)** – utilizza la minima distanza possibile tra due clusters;
3. **Group Average (GA)** – utilizza la distanza tra le medie pesate dei due clusters;
4. **Complete Linkage (CL)** – utilizza la massima distanza possibile tra i due clusters.

Indicando con f il cluster ottenuto dalla fusione del cluster s con il cluster t , di dimensioni rispettivamente n_s , n_t , ed $n_f = n_s + n_t$, la simmetria del cluster f con un altro cluster viene calcolata nei precedenti metodi di clustering secondo:

- **WRNN:** $s_{kf} = \frac{(n_s + n_k)s_{ks}}{n_f + n_k} + \frac{(n_t + n_k)s_{kt}}{n_f + n_k} - \frac{n_k s_{st}}{n_f + n_k}$
- **SL:** $s_{kf} = \min(s_{ks}, s_{kt})$

- **GA:** $S_{kf} = \frac{1}{2}(S_{ks} + S_{kt})$
- **CL:** $S_{kf} = \max(S_{ks}, S_{kt})$

Il risultato di questa procedura viene normalmente rappresentato mediante un grafico detto dendrogramma, che permette un'analisi altamente informativa della gerarchia delle similarità tra gli oggetti considerati. (Figura 14)

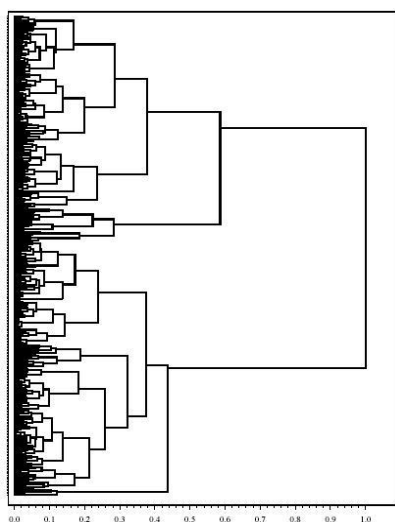


Figura 14. Esempio di Dendrogramma relativo ad una Cluster Analysis

7.7 – Regressione Multilineare (BMLR)

L'analisi QSAR prevede essenzialmente la soluzione di un problema di regressione multilineare. Questo può essere espresso in modo compatto e conveniente utilizzando la notazione di matrice.^{19,20,21} Supponendo che ci siano n proprietà in Y ed n valori calcolati per ogni descrittore molecolare k nelle colonne di X, allora Y_i , X_{ik} ed e_i possono rappresentare, rispettivamente, i valori i-esimi della variabile Y (proprietà), i valori i-esimi di ciascuno dei descrittori X, ed i valori residui i-esimi ignoti. Ponendo questi termini in una matrice si ha:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \dots \\ \dots \\ \dots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & \dots & \dots & X_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & \dots & \dots & \dots & X_{nk} \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \dots \\ \dots \\ \dots \\ e_n \end{bmatrix}$$

Il modello di regressione multipla in notazione matriciale può essere espresso allora come:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

dove \mathbf{b} è il vettore di colonna dei coefficienti (b_1 è per l'intercetta) e k è il numero di coefficienti di regressione ignoti per i descrittori. Ricordando che lo scopo della regressione multipla è minimizzare la somma dei quadrati dei residui:

$$\min_{\mathbf{b}} \|\mathbf{e}\|^2$$

I coefficienti di regressione che soddisfano questo criterio si trovano risolvendo il sistema di equazioni lineari (moltiplicando entrambi i lati per \mathbf{X}')

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\mathbf{b}$$

Quando le variabili X sono linearmente indipendenti (una matrice $\mathbf{X}'\mathbf{X}$ di ordine massimo) c'è una soluzione unica per il sistema di equazioni lineari. Uno dei modi per risolvere il suddetto sistema consiste nel moltiplicare entrambi i lati della formula di matrice per le equazioni normali della matrice inversa $\mathbf{X}'\mathbf{X}$ per dare:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Un altro modo è risolvere direttamente il sistema suddetto usando la fattorizzazione LS (indeterminato, $n < k$) o QR per il sovradeterminato ($n > k$); questo metodo è più generale, e non richiede l'inversione di matrice che necessita di lunghi tempi di calcolo. Possono essere usati anche i metodi di decomposizione del valore singolo, ma di solito tali metodi sono significativamente più lunghi e diventano vantaggiosi solo quando esiste una forte dipendenza lineare che diminuirebbe la qualità dei modelli. Il terzo modo per risolvere il problema di dipendenza lineare di variabili (quando il determinante della matrice $\mathbf{X}'\mathbf{X}$ è maggiore di zero) è l'inversione generale della matrice, ma questo è di solito fuori dalla sfera di QSAR.

Un principio fondamentale dei metodi dei minimi quadrati, della regressione lineare multipla in particolare, è che la varianza della variabile dipendente può essere partizionata secondo la fonte. Supponendo che una variabile dipendente (proprietà) sia regredita su uno o più descrittori e che, per convenzione, la variabile dipendente sia scalata, così che la sua media sia 0, si può calcolare in seguito un'identità base di minimi quadrati, in cui la somma totale dei quadrati dei valori della variabile dipendente sia uguale alla somma dei quadrati dei valori predetti più quella dei residui.

In generale:

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$

dove il termine di sinistra è la somma totale delle deviazioni quadratiche dei valori osservati della variabile dipendente dalla media della variabile dipendente, ed i termini sulla destra sono:

- La somma delle deviazioni quadratiche dei valori osservati della variabile dipendente dalla media della variabile dipendente;
- La somma delle deviazioni quadratiche dei valori osservati della variabile dipendente dai valori predetti, cioè la somma dei residui quadratici.

Oppure, altrimenti:

$$\mathbf{SS}_{Total} = \mathbf{SS}_{Model} + \mathbf{SS}_{Error}$$

Da notare che il valore di \mathbf{SS}_{Total} è sempre lo stesso per ogni particolare insieme di dati, ma \mathbf{SS}_{Model} ed \mathbf{SS}_{Error} variano con l'equazione di regressione. Assumendo di nuovo che la variabile dipendente sia scalata, così che la sua media sia zero, i valori di \mathbf{SS}_{Model} ed \mathbf{SS}_{Error} possono essere calcolati con:

$$\begin{aligned}\mathbf{SS}_{Model} &= \mathbf{b}'\mathbf{X}'\mathbf{Y}' \\ \mathbf{SS}_{Error} &= \mathbf{Y}\mathbf{Y}' - \mathbf{b}'\mathbf{X}'\mathbf{Y}'\end{aligned}$$

Assumendo che $\mathbf{X}'\mathbf{X}$ sia di ordine massimo:

$$\begin{aligned}r^2 &= 1 - \frac{\mathbf{SS}_{Error}}{\mathbf{SS}_{Total}} \\ s^2 &= \frac{\mathbf{SS}_{Error}}{n - k - 1} \\ F(k, n - k - 1) &= \frac{\mathbf{SS}_{Model}}{ks^2}\end{aligned}$$

Dove r^2 è il quadrato del coefficiente di correlazione, ossia la misura della bontà del modello, s^2 è una stima imparziale dei residui o della varianza dell'errore ed F è il criterio di Fisher ($k, n-1-1$) dei gradi di libertà.

Se $\mathbf{X}'\mathbf{X}$ non è di ordine massimo, si sostituisce k con l'ordine $(\mathbf{X}'\mathbf{X})+1$.

7.8 – Test di Validazione

Nell'analisi QSAR occasionalmente si ottengono modelli non totalmente affidabili.²² I problemi dell'affidabilità possono essere classificati come l'overfitting ed i modelli casuali. Il secondo problema può essere risolto solamente dal giudizio umano e soggettivo basato sulla giustificabilità di ogni accertamento dell'incertezza di una particolare predizione. Il problema di overfitting è uno dei problemi più comuni nello sviluppo di ogni tipo di modello, ed è solitamente risolto usando criteri di validazione obiettivi. Il metodo più comune in chemiometria è la "Cross Validation" (CV).²³

Il metodo più semplice, e più comunemente utilizzato di CV è il metodo *Leave-One-Out* (LOO, *Escludi-uno*). L'idea è di predire il valore di una proprietà di un composto dall'insieme di dati che a turno è predetto dall'equazione di regressione calcolata per ogni altra combinazione di dati. Il metodo tende ad includere componenti non necessarie nel modello,

ed è stato modulato²⁴ per essere asintoticamente incorretto. Inoltre il metodo non funziona bene per dati fortemente clusterizzati²³ e sottovaluta l'errore predittivo vero.²⁵

Un altro metodo di CV utilizzato è *Leave-Many-Out (LMO, Escludi-molti)*. A causa della complessità combinatoriale del calcolo, questo metodo dà luogo a bassa produttività, e quindi è suscettibile di ulteriori semplificazioni. La valutazione dei risultati di questo tipo di validazione può essere effettuata utilizzando l'approccio Monte Carlo.²³

I *test di Randomizzazione*²⁶ sono test statistici in cui i dati sono elaborati ripetutamente; viene calcolato un test statistico (come r^2 , criterio t, criterio F, etc...) per la permutazione di ciascun dato. La proporzione delle permutazioni dei dati, con valori di test statistici grandi quanto il valore dei risultati ottenuti, determina la significatività dei risultati. Per saggiare la bontà della correlazione multilineare, le permutazioni del vettore Y sono trattate attraverso procedure di regressione multilineare mantenendo fisse le colonne della matrice X.

A causa di un aumento fattoriale del tempo di calcolo, in dipendenza dalla misura del vettore Y, il metodo Monte Carlo è stato spesso utilizzato per produrre tests di randomizzazione.

Tali tests risultano molto validi quando applicati a metodi di regressione, ma nel caso della DA si è preferito utilizzare tests in grado di descrivere l'abilità predittiva del modello contro la probabilità di assegnazione dei singoli oggetti dovuta al caso. Sono stati utilizzati quindi i tests di Huberty [proportional chance (Cpro) and maximum chance (Cmax)].²⁷ Il criterio di proportional chance è calcolato dalla somma dei quadrati della porzione che ogni gruppo rappresenta del campione (equazione 1). Il criterio di maximum chance è la porzione di oggetti nel gruppo più grande (equazione 2).

$$C_{pro} = p^2 + (1 - p)^2 \quad (\text{eq. 1})$$

p = porzione di oggetti in un gruppo

(1-p) = porzione di oggetti nell'altro gruppo

$$C_{max} = (n_L / N_L) (100) \quad (\text{eq. 2})$$

n_L = numero di oggetti nel più grande dei due gruppi

N_L = numero totale di oggetti nei due gruppi combinati

L'accuratezza della classificazione del modello o minimum hit-rate (min. hit-rate) dovrebbe essere un quarto più grande di quella raggiunta dal caso (Cpro and Cmax).

L'altro test di validazione utilizzato per la DA è il chi-square di Pearson (Chi2).²⁸ Il chi-square test di significatività statistica è una serie di formule matematiche che paragonano la frequenza attualmente osservata di qualche fenomeno (nel nostro campione) con la frequenza attesa se non ci fosse alcuna relazione tra le due variabili nella popolazione più grande.

Quindi, [chi-square test](#) i risultati attuali contro l'ipotesi nulla e afferma se ci sono differenze sufficienti per superare una certa probabilità che sia dovuta ad un errore di campionamento.

7.9 – Docking Molecolare - LigandFit®

In passato sono stati proposti numerosi metodi per il docking di ligandi nei siti attivi di proteine a struttura nota. In generale, i metodi di docking *shape-based* implicano la caratterizzazione tridimensionale del sito attivo e la generazione di orientazioni dei ligandi complementari al volume del recettore. Sono noti diversi metodi per rappresentare il sito attivo di una proteina e per generare le orientazioni dei ligandi di forma compatibile con il sito. In uno dei primi pionieristici lavori in questo campo, Kuntz e collaboratori²⁹ rappresentavano la forma del sito attivo utilizzando un set di sfere sovrapposte di diverso raggio che toccavano la superficie del sito attivo; similmente, i ligandi erano anch'essi rappresentati da un set di sfere. La sovrapposizione del set di sfere del ligando e del sito attivo originavano il docking. Più recentemente sono stati sviluppati diversi tipi di algoritmi che consentono il docking proteina-ligando in cui il ligando è flessibile ma la proteina recettore è rigida, e l'interazione produce una modificazione del ligando. In questo modo si può predire il legame tra ligando e proteina e valutare l'energia totale di interazione nella configurazione ottimizzata di legame.

Nel nostro caso si è deciso di utilizzare LigandFit®, un modulo del Modeling Environment del software *Cerius*² della Accelerlys Inc.

LigandFit garantisce un rapido ed accurato protocollo per il docking di piccoli ligandi nei siti attivi di proteine considerando complementarità di forma tra ligando e sito attivo. Impiega un sistema di comparazione delle “shape” usato per primo da Oldfield per il fitting di molecole attraverso distribuzioni di densità elettronica ottenute dai raggi X.^{30, 31}

La procedura di docking di LigandFit si articola in due stadi:

1. Identificazione della cavità e selezione della regione della proteina come sito attivo per il docking;
2. Docking dei ligandi nel sito selezionato.

Griglie tridimensionali regolari di punti sono impiegate per l'identificazione del sito e anche per stimare l'energia di interazione del ligando con la proteina durante il docking. Il primo stadio (identificazione della cavità) implica l'uso di un algoritmo di “riempimento”. Mentre la procedura per l'identificazione dei vuoti nella proteina è piuttosto diretta, la determinazione dei limiti della cavità e della zona di solvatazione risulta più problematica.

Il secondo stadio del metodo (il docking) impiega un protocollo che comporta:

1. Ricerca conformazionale di ligandi flessibili usando un campionamento stocastico per selezionare i valori degli angoli di torsione variabili;
2. Selezione di una orientazione basata sulla comparazione della forma del conformero del ligando con quella del sito;
3. Stima della qualità del docking utilizzando un calcolo di energia grid-based per valutare l'energia di interazione ligando-proteina.

L'algorithmo per l'identificazione della cavità inizia costruendo una griglia rettangolare con 0.5 Å di spacing. La dimensione della griglia è determinata calcolando l'estensione della proteina lungo i tre assi cartesiani, ed aggiungendo un limite a questa estensione. Ogni punto della griglia è poi classificato come punto libero od occupato. I punti occupati sono quelli che giacciono all'interno della zona di contatto con il più vicino atomo della proteina. La distanza di contatto è uguale al raggio dell'atomo della proteina (2.5 Å per gli atomi pesanti, 2.0 Å per gli idrogeni). I grid point esterni a questa distanza di contatto sono considerati free (non-occupati). Lo step successivo consiste nel dividere i free points in due tipologie: quelli che costituiranno i siti attivi e quelli che si trovano fuori dai siti attivi. Questo implica l'eliminazione delle connessioni tra i vari siti, che avviene tramite i punti liberi che giacciono fuori dalla proteina (Figura 15). I siti attivi distinti potranno essere individuati rimuovendo questi punti liberi utilizzando un "eraser" (MODALITA' PROTEIN SHAPE).

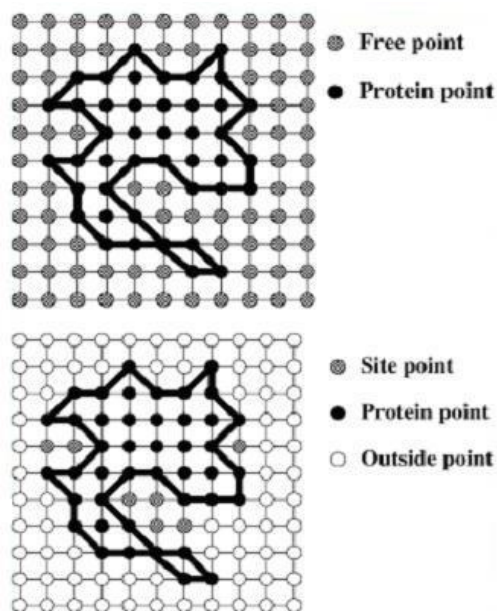


Figura 15. Procedura di identificazione delle cavità nella proteina

In taluni casi, è disponibile la struttura sperimentale tridimensionale di un ligando già dockato nel sito attivo; è allora possibile costruire il sito attivo sulla base della orientazione nota del ligando. Invece di utilizzare l'eraser, sono considerati tutti i free points che giacciono dentro il raggio di qualche atomo dal ligando. In genere il raggio degli atomi pesanti del ligando è convenzionalmente posto uguale a 2.5 Å, mentre il raggio degli atomi di idrogeno è posto a 2.0 Å (**MODALITA' DOCKED LIGAND**).

Una volta identificato il sito attivo si procede al docking dei ligandi. Per un dato ligando, il metodo prevede una procedura iterativa nella quale sono generate conformazioni random del ligando per uno specificato numero di volte, $N_{MaxTrials}$. La procedura mantiene una "Save List" nella quale sono conservate le strutture con il miglior docking trovate dall'algoritmo. Si può specificare il numero delle strutture, N_{Save} , da conservare nella Save List. Se la Save List è piena ed il candidato ha una similarità di shape peggiore di quelli già salvati, questo viene scartato, altrimenti il candidato è selezionato per il docking. Nella ricerca conformazionale del ligando si utilizza il metodo Monte Carlo. Durante questa ricerca, la lunghezza dei legami e gli angoli di legame non vengono modificati, ma sono randomizzati solo gli angoli torsionali (ad eccezione di quelli negli anelli), perciò le molecole dovrebbero essere minimizzate per assicurare la corretta lunghezza dei legami e degli angoli di legame. Il limite superiore della perturbazione diedrale random è di 180° per tutte le torsioni, mentre il limite inferiore dipende dal numero di atomi rotabili: maggiore è il numero di atomi attaccati al legame, minore è la dimensione della perturbazione diedrale.

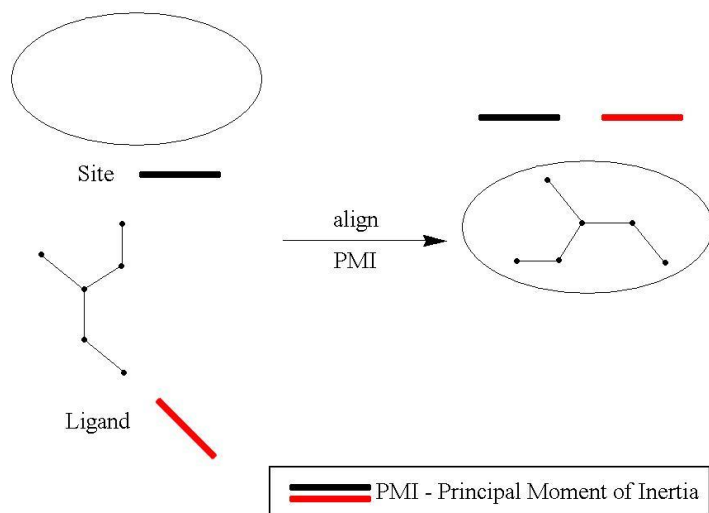


Figura 16. Steps di fitting

Generato il nuovo conformero, il fitting è effettuato in due steps. Dapprima si paragona il momento di inerzia principale (PMI) non mass-weighted del sito attivo (Figura 16) con il PMI non mass-weighted del ligando secondo le seguenti equazioni:

PMI: Principal Moment of Inerzia (P_x, P_y, P_z)

$$\text{Ratio}_{xy} = \frac{P_x}{P_y}$$

$$\text{Ratio}_{xz} = \frac{P_x}{P_z}$$

$$\text{Ratio}_{yz} = \frac{P_y}{P_z}$$

$$\text{Fit}_{\text{PMI}} = \sqrt{\Delta \text{ratio}_{xy}^2 + \Delta \text{ratio}_{xz}^2 + \Delta \text{ratio}_{yz}^2}$$

Se il valore (Fit_{PMI}) è superiore al valore soglia o non è migliore dei fittings precedentemente effettuati, non viene effettuato alcun processo di docking e si prosegue esaminando un altro ligando od un altro conformero dello stesso ligando. Se però Fit_{PMI} è migliore dei risultati salvati precedentemente, il ligando è allora posizionato nel *binding site* secondo il valore di PMI. Poiché PMI è uno scalare, ci sono quattro possibili posizioni (“orientazioni”) (Figura 17) per orientare il ligando nel sito attivo; per ogni posa viene calcolato il corrispondente Docking Score (D) secondo:

$$D = - (\text{energia di interazione Ligando-Proteina} + \text{energia interna del Ligando})$$

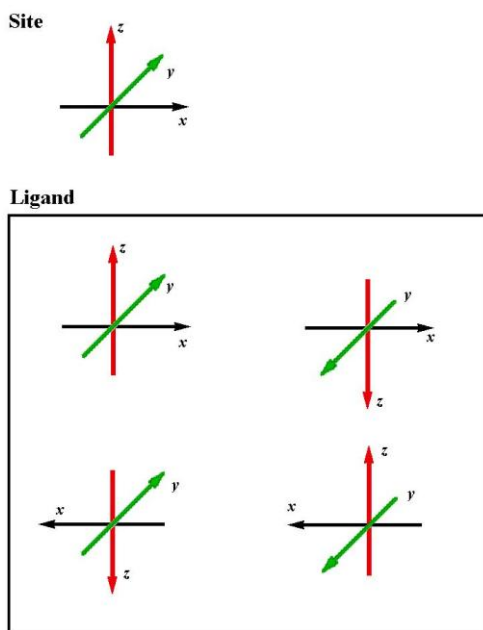


Figura 17. Procedura di docking

L'energia di interazione è calcolata dalla somma dell'energia di Van der Waals e dell'energia elettrostatica. Per ridurre i tempi di calcolo dell'energia di interazione, che altrimenti sarebbero troppo lunghi, viene utilizzato un sistema di interpolazione tri-lineare. L'energia interna del ligando è calcolata come somma delle energie di Van der Waals ed elettrostatiche. L'energia di Van der Waals è computata usando la stessa funzione utilizzata per quella di interazione, sommando tutte le coppie di atomi di non legame interni del ligando, separate da almeno tre legami; stessa valutazione viene effettuata per l'energia elettrostatica.

La procedura di docking fin qui descritta tenta di produrre orientazioni di ligandi che hanno energia di interazione favorevole con la proteina. Mentre le funzioni di energia impiegate per questo scopo tendono a favorire orientazioni ragionevoli dei ligandi, non sono però derivate per predire affinità di legame o per selezionare un ligando piuttosto che un altro. Quindi, bisogna utilizzare delle funzioni di scoring empiriche per prioritizzare i ligandi dockati (LIGSCORE1, LIGSCORE2, PLP1, PLP2, PMF, LUDI).

LIGSCORE, funzione di scoring rapida e semplice, è fittata con GFA (*Genetic Function Algorithm*) e fornisce l'affinità di binding, in termini di pK_i . I descrittori utilizzati sono tre:

$$\mathbf{LIGSCORE1} = pK_i = \mathbf{VdW} + (\mathbf{C} + \mathbf{pol}) + \mathbf{Totpol}^2$$

VdW è espressa in Kcal/mol, C+pol è espressa in \AA^2 , Totpol² è espressa in \AA^4 ; questi ultimi due rappresentano descrittori di superficie, ma mentre C+pol esprime le interazioni attrattive tra ligando e proteina, Totpol² esprime anche interazioni repulsive. Equazioni simili sono usate per LIGSCORE a seconda del descrittore VdW utilizzato (LIGSCORE1, LIGSCORE2) o del force field (CFE o Dreiding).³²

PLP (Piecewise Linear Potential) è anch'essa una funzione di scoring rapida e semplice, e viene espressa in unità arbitrarie di energia; la relazione con l'affinità di binding è del tipo:

$$pK_i = -\mathbf{PLP}$$

Essa tiene conto dei donatori e degli accettori di H, nonché delle interazioni non polari.

Esistono due forme della funzione PLP (1 e 2): la differenza consiste nel diverso tipo di interazioni che vengono considerate quando si calcola $E_{\text{Ligando-Proteina}}$. In PLP1 si considerano E_{Sterica} ed E_{Hbond} , mentre in PLP2 il contributo è dato da $E_{\text{Repulsive}}$ o $E_{\text{Dispersive}}$ che provengono da interazioni ligando-donatore o accettore-accettore ed in genere da tutti gli altri tipi di interazioni.^{33, 34}

PMF (Potential of Mean Force) è una funzione definita come la somma delle interazioni di energia libera fra tutte le coppie interatomiche del complesso ligando-proteina. Gli scores vengono riportati in unità di energia arbitraria; maggiore è il valore di $-PMF$, maggiore è l'affinità del ligando. In genere la relazione di energia è data da:

$$\Delta G_{\text{Bind}} = \frac{\text{PMF_score}}{\epsilon}$$

dove ϵ è uno scaling factor.³⁵

LUDI è una funzione di scoring proposta da Bohm. E' data dalla somma di cinque contributi:

1. contributo proveniente dai legami H ideali;
2. contributo proveniente dalle interazioni ioniche perturbate;
3. contributo proveniente dalle interazioni lipofile;
4. contributo dovuto alla perdita di entropia rotazionale e traslazionale del ligando;
5. contributo dovuto al freezing dei gradi di libertà interni.

La relazione di energia viene espressa in kcal/mol ed è data da:

$$\text{LUDI_score} = k \Delta G_{\text{bind}}$$

dove k assume il valore di -77.33 mol/kcal.²⁹

Tali funzioni empiriche utilizzano modelli di regressione [Multi Linear Regression (MLR) o (GFA)] per fittare i coefficienti su un set di termini fisicamente opportuni al fine di riprodurre le affinità di binding sperimentali di un training set di complessi ligando-proteina noti.

8. RISULTATI E DISCUSSIONE

8.1 – Analisi multivariata applicata agli NNRTI

Più di 30 differenti classi di NNRTI sono state descritte fino ad oggi. In questo studio sono stati utilizzati cinquantacinque derivati NNRTI, le cui strutture sono riportate in Appendice I. La selezione include la maggior parte dei derivati presenti correntemente nel database del National Institute of Allergy and Infectious Diseases (NIAID),³⁶ e derivati descritti in letteratura.³⁷ Sono stati presi in considerazione i composti più attivi di ogni classe. La Tabella 2 mostra per i derivati oggetto di studio i dati di sensibilità/resistenza contro i più comuni ceppi mutanti di RT. I derivati con resistant ratio >15 (Attività sul ceppo mutante /Attività sul wt) sono considerati come resistenti (R); i derivati con resistant ratio <15 sono considerati come suscettibili (S).³⁸

Formattato: Italiano (Italia), Non Evidenziato

I composti sono stati inizialmente disegnati come rappresentazioni 2D utilizzando ChemDraw Ultra 7.0 e le strutture 3D sono state ottimizzate con metodi semiempirici (MNDO o PM3) utilizzando i moduli CORINA e COSMIC di TSAR 3.2. Per i derivati chirali per i quali la stereochimica della forma attiva non è nota sono stati presi in considerazione entrambi gli enantiomeri (cfr derivati 5/6; 27/28; 33/34; 43/44; 48/49).

È stato inizialmente calcolato un pool di più di 80 descrittori, tra i quali sono stati selezionati 11 descrittori da includere nel database che meglio identificano le caratteristiche molecolari che possono essere correlate con l'attività biologica (volume ellissoidale, $\log \log P$, donatori di legami H, accettori di legami H, lipolo totale, area superficiale, calore di formazione, E_{LUMO} , E_{HOMO} , momento di dipolo totale, energia totale in acqua).³⁹ (Tabella 3) La selezione di variabili strutturali è stata fatta considerando le numerose strutture cristallografiche di HIV-1 RT complessato con NNRTI, le quali supportano il ben noto binding mode delle differenti classi di NNRTI. In particolare il

~~Since resultant correlation matrix gave coefficients <0.7 , meaning that no significant 0.9 correlation was found the whole set of descriptors could be used.~~

Formattato: Italiano (Italia), Non Evidenziato

Formattato: Italiano (Italia), Non Evidenziato

volume ellissoidale, che è definito dal momento di inerzia, e l'area superficiale accessibile danno informazioni circa le proprietà steriche delle molecole, $\log P$ è una tipica variabile QSAR, correlata con il profilo idrofobico/idrofilico dell'inibitore e nel nostro caso può essere correlato a questo tipo di interazione nel sito attivo, il lipolo totale è una misura della distribuzione lipofila nello spazio 3D ed è calcolato dalla somma atomica dei valori di $\log P$,⁴⁰ il numero di donatori/accettori di legami H fornisce altre informazioni circa la capacità dell'inibitore di stabilizzare la propria interazione nel "binding pocket" dell'enzima, il dipolo totale dà informazioni circa le caratteristiche elettroniche, E_{LUMO} e E_{HOMO} sono variabili

energetiche che classificano il set di inibitori in termini di abilità a comportarsi come nucleofili o elettrofili, come ci si aspetta per il processo di inibizione. Altre variabili energetiche sono il calore di formazione, che classifica il set di inibitori in termini di stabilità termodinamica relativa ed è ampiamente usato negli studi chemometrici e l'energia totale in acqua, che classifica il comportamento degli inibitori in ambiente fisiologico.

Tabella 2. Resistenza (R)^a/Susceptibilità (S)^b dei derivati inclusi nel database verso i mutanti di RT selezionati

Entry	Derivati	L100I	K103N	V106A	Y181C	Ref.
1	(+)-12-OXOCALANOLIDE	R	R		S	36
2	21-AAP-BHAP				S	37b
3	ADAM-II		R		R	37b
4	AN-29-PNU-32945					37b
5	BENZOTHIADIAZINE-2(R)					37c
6	BENZOTHIADIAZINE-2(S)					37c
7	BM-510836					37f
8	CALANOLIDE-A	R	R		S	37a
9	COMPD-1					37
10	COMPD-2					37a
11	COMPD-36				S	37b
12	COSTATOLIDE	R	R		S	36
13	DABO-12e					37a
14	DAMNI-6D					37g
15	DELAVIRDINE	S	R	R	R	38c, 38d
16	DIHYDROCOSTATOLIDE	R	R		S	36
17	DPC-082					37b
18	DPC-961		S			37b
19	E-EBU-dM			R		38a, 38b
20	EFAVIRENZ	R	R	S	S	36, 38d, 38c
21	GCA-186		R		R	37f
22	HBV-097		R			37f
23	INOPHYLLUM-B					37a
24	IPPH					37a
25	L-697661	S	S		R	38a, 38b
26	L-737126					37a
27	LOVIRIDE(R)		S		R	38a, 38b
28	LOVIRIDE(S)		S		R	38a, 38b
29	MKC-442		R		R	37a
30	MSA-300					37a
31	NEVIRAPINE	S	R	R	R	36
32	NPPS					37a
33	NSC-625487(R)					37a
34	NSC-625487(S)					37a
35	PBD					37a
36	PBO-13B(R)	S	S	S		37b
37	PBO-13B(S)	S	S	S		37b
38	PEQ					37a
39	PHS					37a
40	PNU-142721(R)				R	37b
41	PNU-142721(S)	S			S	37b
42	PNU-32945					37b
43	PPO-13I(R)	S	S	S		37b
44	PPO-13I(S)	S	S	S		37b
45	QM-99639	S	R		R	37b
46	S-2720			R		38a, 38b
47	SEFAVIRENZ					37b
48	TGG-II-23A(R)					37a
49	TGG-II-23A(S)					37a
50	TIVIRAPINE		R			38a, 38b
51	TNK-6123		S		S	37b
52	TTD-6G	S	R		R	37h
53	UC-781	R	R		S	37a
54	UC-84	R			R	38a, 38b
55	UK-129485					37a

| ^aR>15 resistant ratio (Attività sul ceppo mutante /Attività sul wt); ^bS<15 resistant ratio.

Tabella 3. Database dei descrittori molecolari utilizzati nel caso dei derivati NNRTI.

Entry	Descrittori										
	1	2	3	4	5	6	7	8	9	10	11
1	628.9	3.2	0	5	5.51	403.8	-143.82	-0.859	-8.922	8.45	-4748.8
2	3000	2.4	2	4	17.72	515.5	-8.08	-0.63	-8.831	5.59	-5692.5
3	1652.6	4.4	0	8	7.09	547.3	-282.18	-0.783	-9.42	2.42	
4	614.3	4.8	1	2	13.19	351.4	60.09	-0.16	-8.878	3.01	-3440.1
5	763.5	2.8	2	2	9.28	339.5	25.05	-0.672	-9.276	3.77	-3774.8
6	763.5	2.8	2	2	9.28	339.5	25.05	-0.672	-9.276	3.77	-3774.8
7	578.3	4.8	0	1	6.86	322.6	38.24	-0.599	-8.728	2.44	-3259
8	654.3	3.3	1	5	5.33	410.9	-167.61	-0.763	-8.826	6.25	-4776.9
9	351.4	1	0	3	4.4	265.5	-4.28	-0.706	-9.479	5.53	-3396.7
10	300.8	1.5	1	1	6.28	235	51.25	-0.311	-7.909	2.37	-2601.4
11	5434.5	2	3	4	19.97						
12	654.9	3.3	1	5	5.33	411.3	-167.61	-0.764	-8.827	6.26	-4776.9
13	691	3.5	1	3	8.15	358.2	-49.34	-0.23	-9.245	4.11	-3665.4
14	743.5	4.3	0	4	13.28	377.3	64.29	-0.87	-9.582	4.99	-4270.9
15	3450.6	0.7	3	4	10.59	492.3	10.13	-0.636	-8.559	7.65	-5603
16	623.4	3.4	1	5	4.93	415.5	-195.2	-0.705	-8.932	6.92	-4805.4
17	638.3	3	2	1	6.13	297.7	-198.04	-0.803	-9.318	3.69	-5102.7
18	735.3	3	2	1	7.08	300.7	-81.49	-0.529	-9.196	4.04	-4491.2
19	839.3	2.8	1	3	11.25	373.6	-102.35	-0.19	-9.374	4.12	-4011.1
20	674.7	3.6	1	2	5.48	297.6	-120.15	-0.749	-9.454	5.36	-4590.7
21	875	3.1	1	3	11.39	389.5	-104.31	-0.185	-9.37	3.9	-4166.5
22	1002.9	2.4	1	4	2.1	369.6	-63.41	-0.844	-8.454	5.45	-3959
23	712.4	3.8	1	5	8.93	426	-118.05	-0.86	-8.803	6.27	-5132.1
24	3281.9	4.8	0	4	7.2	455.3	131.06	-0.888	-8.641	3.99	-4650.8
25	1012.3	1.3	2	3	15.68	358.2	-0.77	-0.781	-8.197	3.36	-4265.1
26	807.7	1.8	2	3	10.77	311.9	-26.68	-0.99	-9.349	3.77	-4014.4
27	702.8	2.9	2	2	5.29	351.4	-38.49	-0.424	-8.671	2.65	-4199.5
28	702.8	2.9	2	2	5.29	351.4	-38.49	-0.424	-8.671	2.65	-4199.5
29	638.1	2.2	1	3	9.49	345.4	-89.08	-0.218	-9.406	3.56	-3854.8
30	2512.9	3.4	3	5	14.28	396.6	-3.11	-0.586	-8.369	5.42	-4705.8
31	421.9	1.9	1	3	1.98	287.4	87.9	-0.564	-8.902	3.19	-3258.8
32	601.2	2.8	0	4	7.1	244.5	6.97	-1.404	-10.395	9.37	-3338.6
33	375.9	4.1	0	1	4.69	279.5	29.32	-0.785	-8.972	3.21	-3633.9
34	375.9	4.1	0	1	4.69	279.5	29.32	-0.785	-8.972	3.21	-3633.9
35	1793.4	3	1	3	7.48	370.9	-11.47	-0.267	-8.458	1.73	-4310.4
36	804.1	4.1	0	3	5.2	356.7	10.96	-0.475	-8.964	2.31	-4115.2
37	806.7	4.1	0	3	5.2	356.7	10.96	-0.476	-8.964	2.32	-4115.3
38	876.9	3.3	2	2	9.88	339.8	115.59	-0.501	-8.922	2.15	-3808.9
39	584.6	2.8	0	4	9.14	366.5	-114.99	0.212	-8.984	3.8	-4082.7
40	1636.5	3.2	1	4	3.71	304.8	72.63	-0.472	-8.545	5.9	-3530.7
41	913.8	3.2	1	4	3.65	304.4	72.03	-0.427	-8.558	2.15	-3530.7
42	524.2	4.4	0	2	7.45	318.3	149.51	-0.279	-9.099	4.18	-3128.3
43	612.2	3.5	0	3	4.15	321.7	62.54	-0.746	-9.078	3.57	-3704
44	612.2	3.5	0	3	4.15	321.7	62.54	-0.746	-9.078	3.57	-3704
45	553.7	1.8	0	4	6.32	309.8	-32.82	-1.121	-9.528	5.65	-4385.8
46	539.3	2.8	1	3	5.95	322.4	2.74	-1.015	-8.747	2.84	-3620
47	737.5	4.7	1	2	4.64	311.9	-57.06	-1.07	-9.23	5.69	-4463.7
48	494.4	2.9	0	3	6.84	313.8	-14.2	-0.746	-9.285	3.55	-3284.9
49	494.4	2.9	0	3	6.84	313.8	-14.2	-0.746	-9.285	3.55	-3284.9
50	682.2	3.8	1	2	3.54	345.9	74.37	-0.616	-8.4	3.33	-3511.2
51	511.3	3.1	1	3	6.92	375.7	-131.61	-0.68	-9.054	3.7	-3978.7
52	451.3	1.7	0	4	5.96	304.6	12.13	-1.048	-9.584	5.35	-3913.7
53	1420.7	4.6	1	3	4.9	367.4	12.25	-0.948	-8.494	5.51	-3839.4
54	1610.2	1.1	1	4	13.32	370.7	-128.3	-0.419	-8.573	6.54	-4354.3
55	536.8	3.6	0	4	5.19	334	112.75	-0.599	-8.495	2.53	-3762.5

1. Volume ellissoidale (\AA^3), 2. log P, 3. numero di donatori di H, 4. numero di accettori di H, 5. lipolo totale, 6. area superficiale (\AA^2), 7. calore di formazione (kcal mol^{-1}), 8. E_{LUMO} (eV), 9. E_{HOMO} (eV), 10. momento di dipolo totale (Debye), 11. energia totale (eV).

Formattato: Giustificato, Interlinea singola

L'intero set di 11 descrittori e 55 composti è stato utilizzato per effettuare una PCA. La Tabella 4 riporta la matrice di 11 PC con la loro composizione in termini di variabili originali, insieme alla frazione di varianza spiegata, la frazione totale di varianza spiegata e l'autovalore della matrice di covarianza corrispondente a ogni componente, quest'ultimo valore è uguale alla frazione di varianza spiegata dal numero di variabili usate. Le prime 4 PC, che hanno autovalore superiore a 1 e che spiegano il 73.7% della varianza sono state selezionate per ulteriori calcoli. Nella prima PC le variabili che hanno maggiore importanza sono il volume ellissoidale (+0.39), l'area superficiale (+0.45) e l'energia totale in acqua (-0.46). Queste variabili chiaramente riflettono l'importanza dell'approccio sterico al binding pocket e il comportamento delle molecole all'interno del citoplasma. Nella seconda PC le caratteristiche elettroniche hanno un peso maggiore: E_{LUMO} (+0.45), E_{HOMO} (+0.45), dipolo totale (-0.47) riflettono l'importanza del π -stacking durante l'interazione farmaco-recettore. Log P e il calore di formazione hanno una grande importanza nella terza e nella quarta PC (+0.59, +0.44 e +0.38, -0.58 rispettivamente), analogamente il numero di donatori di H (-0.41) assume importanza nella terza PC e E_{LUMO} (+0.47) nella quarta. Ma dal momento che la frazione di varianza spiegata dalla terza e dalla quarta PC è piuttosto bassa queste variabili non sono certamente le più rilevanti.

Tabella 4. Risultati della PCA sul set di descrittori selezionato e sugli indici di similarità

<i>Descrittore</i>	PC1	PC2	PC3	PC4
1	0.386	0.238	0.259	-0.319
2	-0.140	0.034	0.589	0.438
3	0.278	0.302	-0.408	-0.150
4	0.331	-0.262	0.272	-0.027
5	0.288	0.223	-0.224	-0.083
6	0.450	0.124	0.267	0.213
7	-0.227	0.281	0.385	-0.581
8	-0.027	0.454	-0.150	0.473
9	0.094	0.453	0.225	-0.037
10	0.293	-0.467	0.025	-0.153
11	-0.457	0.075	-0.010	-0.203
Frazione di varianza spiegata	0.327	0.194	0.120	0.095
Varianza totale spiegata	0.327	0.521	0.642	0.737
Autovalore	3.602	2.137	1.325	1.046

<i>Shape</i>	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Frazione di varianza spiegata	0.529	0.118	0.091	0.045	0.034	0.029	0.020	0.019
Varianza totale spiegata	0.529	0.647	0.739	0.784	0.819	0.848	0.869	0.889
Autovalore	29.130	6.491	5.050	2.498	1.892	1.608	1.137	1.082

<i>Lipofilia</i>	PC1	PC2
Frazione di varianza spiegata	0.947	0.024
Varianza totale spiegata	0.947	0.971
Autovalore	52.1	1.33

Formattato: Tipo di carattere: 10 pt, Italiano (Italia)

Formattato: Tipo di carattere: 10 pt, Italiano (Italia)

Formattato: Tipo di carattere: 10 pt, Italiano (Italia)

Poiché la similarità molecolare è uno degli strumenti più utili nell'approccio computazionale per identificare molecole che si legano allo stesso sito recettoriale, sono stati calcolati anche gli indici di similarità in termini di caratteristiche steriche (shape) e potenziale lipofilo³⁹. Infatti la shape può fornire informazioni sull'accessibilità e sull'interazione con il sito attivo, mentre la lipofilia può essere importante non solo per l'interazione con il binding site, il quale è stabilizzato dalle caratteristiche idrofobiche dell'inibitore, ma anche per la biodisponibilità. Ancora una PCA è stata effettuata sulle matrici degli indici di similarità e il numero di PC significative è stato determinato come in precedenza (autovalore > 1): otto PC nel caso degli indici di similarità per shape e 2 nel caso degli indici di similarità secondo la lipofilia. Le componenti principali selezionate sono state incluse separatamente in una Discriminant Analysis con lo scopo di fornire l'assegnazione a una delle due classi: Sensibili (S) o Resistenti (R).

E' noto che una singola mutazione nel binding pocket degli NNRTI può condurre ad un elevato livello di resistenza. Le mutazioni che sono state selezionate sono K103N e Y181C, associate con un tipo di resistenza soprattutto nei confronti degli NNRTI di prima generazione^{37b}, inclusi i farmaci Nevirapina (31) e Delavirdina(15), e V106A e L100I che causano resistenza anche con i farmaci di nuova generazione, sebbene da bassa a intermedia nel caso di Delavirdina (15) e Efavirenz (20).^{41,42} Inizialmente sono stati inclusi nella DA soltanto i derivati per i quali sono disponibili dati di sensibilità/resistenza. I risultati sono schematizzati nella Tabella 5. Gli hit-rates ottenuti sono tutti elevati (≥ 80.9) specialmente per i composti sensibili: i derivati sono stati assegnati alla vera classe con poche eccezioni. Questi risultati sono particolarmente interessanti nel caso dei due mutanti più frequenti (L100I e K103N) quando vengono considerati le PC derivate dai descrittori e dagli indici di similarità per shape. Le PC derivate dagli indici di similarità per lipofilia hanno significatività statistica solo nel caso di L100I. Prendendo in considerazione le PC derivate dai descrittori, è rilevante osservare che nel caso di L100I e K103N, le prime due componenti principali evidenziano l'importanza delle caratteristiche steriche ed elettroniche nell'insorgere della resistenza. Nel caso di Y181C ancora una volta la prima PC è coinvolta insieme alla terza e alla quarta PC e quindi i descrittori log P, calore di formazione, ma anche E_{LUMO} e donatori di H. Nel caso della mutazione V106A la seconda, la terza, la quarta PC sono tutte di relativa importanza, così si può assumere che le caratteristiche elettroniche giocano un ruolo di primaria importanza.

Tabella 5. Risultati della DA basati sui PC per I descrittori e gli indici di similarità per le classi R e S

	Descrittori				Indice di similarità SHAPE				Indice di similarità LIPOFILIA					
	Predetti		Tot.	Hit Rate	Predetti		Tot.	Hit Rate	Predetti		Tot.	Hit Rate		
	R	S			R	S			R	S				
L	R	6	1	7	85.7%	7	0	7	100%	6	1	7	85.7%	
	S	0	10	10	100%	0	10	10	100%	1	9	10	90.0%	
	Totale	6	11	17	94.1%	7	10	17	100%	7	10	17	88.2%	
	0	<i>Variabili incluse</i> PC1, PC2, PC4				PC4, PC6, PC7				PC2				
	0	Chi2	13.24;	df=1,	p<0.001	17;	df=1,	p<0.001		9.74;	df=1,	p<0.01		
I	Cpro	51.5% min. hit-rate=64.3%			51.5% min. hit-rate=64.3%				51.5% min. hit-rate=64.3%					
	Cmax	58.8% min. hit-rate=73.5%			58.8% min. hit-rate=73.5%				58.8% min. hit-rate=73.5%					
	R	11	4	15	73.3%	14	2	16	87.5%	15	1	16	93.7%	
	S	0	8	8	100%	0	8	8	100%	7	1	8	12.5%	
1	Total	11	12	23^a	82.6%	14	10	24	91.6%	22	2	24	66.7%	
0	<i>Variabili incluse</i> PC1, PC2, PC4				PC3, PC4, PC5, PC8				PC1					
3	Chi2	11.24;	df=1,	p<0.001	13.24;	df=1,	p<0.001		0.27;	df=1,	p<1			
N	Cpro	52.7% min. hit-rate=65.8%			55.5% min. hit-rate=69.7%				55.5% min. hit-rate=69.7%					
	Cmax	65.2% min. hit-rate=81.5%			66.7% min. hit-rate=83.3%				66.7% min. hit-rate=83.3%					
V	R	4	0	4	100%	4	0	4	100%	2	2	4	50.0%	
	S	0	5	5	100%	0	5	5	100%	0	5	5	100%	
	1	Total	4	5	9	100%	4	5	9	100%	2	7	9	77.8%
	0	<i>Variabili incluse</i> PC2, PC3, PC4				PC3				PC1, C2				
	6	Chi2	9;	df=1,	p<0.01	9;	df=1, p	<0.01		3.21;	df=1,	p<0.1		
A	Cpro	50.5% min. hit-rate=63.1%			50.5% min. hit-rate=63.1%				50.5% min. hit-rate=63.1%					
	Cmax	55.5% min. hit-rate=69.4%			55.5% min. hit-rate=69.4%				55.5% min. hit-rate=69.4%					
Y	R	7	1	8	87.5%	5	3	8	62.5%	7	2	9	77.8%	
	S	1	11	12	91.6%	1	12	13	92.3%	7	6	13	46.1%	
	1	Total	8	12	20^{ab}	90.0%	6	15	21^b	80.9%	14	8	22	59.1%
	8	<i>Variabili incluse</i> PC1, PC3, PC4				PC2, PC3, PC5, PC6, PC7				PC1, PC2				
	1	Chi2	13.24;	df=1,	p<0.001	7.28;	df=1,	p<0.01		1.31;	df=1,	p<1		
C	Cpro	52.0% min. hit-rate=65.0%			52.8% min. hit-rate=66.0%				51.6% min. hit-rate=64.5%					
	Cmax	60.0% min. hit-rate=75.0%			61.9% min. hit-rate=77.3%				59.0% min. hit-rate=73.7%					

^aEntry esclusa: **3**. ^bEntry esclusa: **11**

Considerando la validità statistica dei risultati ottenuti, sono stati effettuati nuovi calcoli includendo questa volta anche i derivati per i quali i dati di sensibilità/resistenza non sono noti. La procedura di classificazione implica l'inserimento dei vari composti nelle classi S o R. I dati riportati in Tabella 6 mostrano che è possibile fornire previsioni di attività per un elevato numero di derivati assegnando fino a 10 composti alla classe R, mentre ci si può attendere che anche un elevato numero di composti inseriti nella classe S, può conservare la propria attività nei confronti dei ceppi mutanti. In generale i migliori risultati sono stati ottenuti nel caso di L100I e K103N per i quali si è ottenuto il maggior numero di assegnazioni soprattutto quando sono state utilizzate le PC provenienti dai descrittori. Mentre per V106A le previsioni sono meno significative probabilmente a causa del limitato numero di dati input.

I risultati ottenuti dalla DA possono essere così riassunti:

- Il paragone dei risultati per tutti i mutanti di RT mostra che una consistente classificazione come S è raggiunta nel caso dei derivati **33, 34, 40, 41, 48, 49,**

55 e come R nel caso dei composti **5, 6, 17, 51, 54**. I derivati **4** e **42** e **24** sono generalmente classificati come sensibili, ma le assegnazioni sono controverse contro K103N (**4, 42**) o contro V106A (**24**).

- I derivati **23** e **41** sono predetti sempre nella stessa classe.
- Per quanto riguarda i tre farmaci in commercio, la Delavirdina (**15**) viene generalmente identificata correttamente con la sola eccezione nel caso del mutante Y181C quando si considera l'indice di similarità per shape e; Efavirenz (**20**) è generalmente classificato correttamente dai PC derivati dai descrittori, mentre la Nevirapina (**31**) viene correttamente classificata nel caso del mutante V106A.
- Nel caso di PNU-142721, il solo composto sotto trials clinici per il quale sono disponibili dati di sensibilità/resistenza differenti per i due stereoisomeri [40 configurazione R; 41 configurazione S], le analisi che tengono conto dei descrittori e degli indici di similarità per shape riescono a discriminare tra i due isomeri. Per gli altri composti chirali inseriti nello studio, non vengono rilevate differenze sostanziali, confermando i dati biologici ottenuti sulla miscela di isomeri.

Tabella 6. Classificazione dell'intero set di NNRTI nelle classi R and S

Entry	DESCRITTORI				SHAPE			
	L100I	K103N	V106A	Y181C	L100I	K103N	V106A	Y181C
1	R	R	S	S	R	R		S
2	R	R	R	S	S		R	S
3						R		R
4		S	S			R		
5			R	R				
6			R	R				
7		S	S			R		R
8	R	R		S	R	R		S
9	S		R	R			R	
10		S	R					S
11								
12	R	R		S	R	R		S
13		S					R	
14	S	R	S	S				
15	S	R	R	R	S	R	R	R
16	R	R	S	S	R	R		S
17		R		R				
18		S		R		S		
19		S	R	R		R	R	R
20	R	R	S	S	R	R	S	S
21		R		R		R	R	R
22	S	R				R	R	R
23	R	R		S	R	R		S
24	S		S	S			R	
25	S	S	R	R	S	S		R
26	S		R	R				
27		S		R		S		R
28		S		R		S		R
29		R		R		R	R	R
30	S		R	R				
31	S	R	R	R	S	R	R	R
32	S	R		R				
33		S	S					
34		S	S					
35		S					R	S
36	S	S	S		S	S	S	
37	S	S	S		S	S	S	
38		S	R					
39	R	S	S	S				
40	S			R		S		R
41	S	S		S	S	S		S
42		S	S			R		
43	S	S	S		S	S	S	
44	S	S	S		S	S	S	
45	S	R		R	S	R		R
46	S		R				R	
47	R	R	S	S				
48	S							
49	S							
50		R	S			R	R	
51	R	R		R		R	R	R
52	S	R		R	S	R	S	R
53	R	R	S	S	R	R		S
54	R	R	R	R	R			R
55	S		S					

Corretto

Predetto

Misclassificato Classe Opposta

Misclassificato Classe No-Data

8.2 – Analisi multivariata applicata agli Inibitori della Proteasi

Lo studio sugli inibitori della Proteasi è stato effettuato invece su un set combinato di cinquantuno inibitori, sia di tipo peptidomimetico sia di tipo non peptidico.

Anche in questo caso la selezione include la maggior parte dei derivati attualmente presenti nel database del National Institute of Allergy and Infectious Diseases (NIAID)³⁶ e i derivati di letteratura per i quali è stata riportata l'attività inibitoria contro HIV-1.⁴³ (vedi Tabella 7) E' stato considerato solo il derivato più attivo appartenente ad ogni classe chimica.

Tabella 7. Resistenza (R)^a/Suscettibilità (S)^b verso i mutanti di PR selezionati

Entry	Derivati	V82A	V82F	I84V	Ref.
1	33_DER_CYCLOURETHAN				43a°
2	(R)6H-1,3-DIOS-4D				43b
3	AG_1350				43c
4	AHPBA_38				43d
5	AHPBA_48				43e
6	AMPRENAVIR	R	S	S	43f, 43g., 44a, 43q, 44b
7	A_77212				43h
8	BILA_2185_BS				43i, 44c
9	BMS_186318				43j, 44c
10	CAGEDIM_3C				43k
11	CYANGUAN_8I				43l
12	CYCLOSULFAMIDE_25				43m
13	CYCLOUREA_1				43m
14	CYCLOUREA_10A				43n
15	CYCPTR_11				43o
16	DHP_13Y				43p
17	DMP_450	R	R	R	43q, 43r 44d
18	DPC_681			S	43s
19	DUPLCTERM_4				43t
20	EGCg				43u
21	ESC_14				43v
22	HEA_19				43w
23	INDINAVIR	R	R	S	43x, 43y, 43q, 44b, 944e
24	ISOQUIN_URETH_ANALOG_2				43z
25	I_12		R		43aa
26	KNI_272				43ab, 43ac, 44c, 44f
27	LASINAVIR				43ad, 43ae
28	LC_3				43af
29	LOPINAVIR	R	R		43ag, 43s, 44g
30	LY_289612-ANAL1				43ah, 43ai
31	LY_314163				43aj
32	LY_326188				43ak
33	L_689502				43al, 43am
34	MANNAR_9F				43an
35	MANNOPYRANOSIDE_18				43ao
36	MW_583				43ap
37	NELFINAVIR	S	S	R	43aq, 43ar, 44h, 44c, 44f
38	PALINAVIR				43as, 44f
39	PARACYCLOPHANE_DER_1				43at
40	RITONAVIR	S	R	R	43au, 44f, 944c
41	R_87366				43av
42	SAQUINAVIR	S	S	S	43q, 43aw, 44e
43	SD_146	S	S	S	43q, 43ax,
44	SE_063				43ay
45	SYN_DIMERIC_DHP_17				43az
46	TELINAVIR				43ba,43bb 44c
47	THP_19				43bc
48	THP_20				43bc
49	TIPRANAVIR				43bd, 43be, 44c, 44f
50	TMC_126				43bf
51	UREATRICL_9				43bg

^aR>2.1 resistant ratio (Attività sul ceppo mutante /Attività sul wt); ^bS<2.1 resistant ratio.

Nel tentativo di classificare questo set di composti sulla base della loro attività contro i più comuni mutanti di PR^{43,44}, i derivati con resistant ratio > 2.1 sono stati considerati come resistenti (R), i composti con ratio < 2.1 sono stati considerati come suscettibili (S). Anche in questa occasione le molecole sono state disegnate inizialmente come rappresentazioni 2D utilizzando ChemDraw e poi ottimizzate come strutture tridimensionali attraverso metodi semiempirici (MNDO o PM3) utilizzando i moduli CORINA e COSMIC di TSAR 3.2.

Tabella 8. Database dei descrittori molecolari utilizzati nel caso degli inibitori della Proteasi.

Entry	Descrittori											
	1	2	3	4	5	6	7	8	9	10	11	12
1	7.76	543.72	2334.60	4.00	5.44	6.16	13.00	8	1.10	87.50	2	7
2	-19.12	538.06	4882.70	6.22	8.17	11.93	9.63	11	1.23	116.48	2	7
3	-4.81	522.66	2777.70	7.90	3.19	17.74	10.31	10	1.37	85.18	4	5
4	-28.03	588.82	3658.70	3.19	6.04	36.39	12.69	14	1.39	111.92	4	5
5	-14.37	465.65	1682.20	5.73	4.40	20.95	9.43	8	1.42	97.06	4	5
6	-18.36	449.62	2677.00	8.60	2.66	7.22	9.93	12	1.62	85.82	3	7
7	25.38	624.37	4785.00	5.47	7.02	8.69	15.36	21	1.45	114.67	4	10
8	-10.97	601.50	3440.20	6.97	4.42	14.50	12.26	13	1.33	95.38	3	6
9	-30.32	654.52	5672.60	2.84	3.16	16.49	15.43	18	1.50	114.33	5	9
10	114.97	503.97	1406.60	1.99	5.36	5.16	6.31	10	1.13	96.67	4	6
11	28.03	466.07	1683.90	3.61	7.40	2.21	9.24	9	1.29	78.50	2	4
12	20.70	554.47	2559.70	8.48	4.18	7.08	11.11	12	1.26	109.65	4	9
13	25.58	439.96	1276.40	6.40	5.69	11.37	9.46	10	1.21	89.67	2	5
14	35.88	527.54	2196.70	7.46	7.33	5.63	9.57	10	0.99	104.33	4	5
15	166.59	661.40	5609.20	1.56	4.71	4.69	13.08	17	1.06	126.67	6	6
16	-12.44	436.77	1231.40	6.12	4.24	2.12	7.56	7	1.60	74.47	3	4
17	-10.45	473.48	1190.30	6.75	4.94	7.52	8.81	8	1.24	90.00	4	3
18	-8.39	624.30	5044.50	6.71	4.51	12.81	13.91	17	1.65	117.57	5	7
19	34.39	757.32	5098.70	1.93	4.93	24.98	16.89	19	1.45	135.33	6	8
20	4.88	381.55	1360.60	3.40	3.02	3.36	6.09	4	1.34	99.83	8	11
21	9.55	531.55	3072.10	8.28	2.59	7.01	10.50	12	1.38	93.07	2	8
22	-30.20	634.33	4103.50	5.41	5.30	21.30	14.70	17	1.60	113.08	5	6
23	-17.27	574.53	3283.90	4.42	3.46	14.53	10.91	12	1.17	99.92	4	7
24	-20.97	553.13	2689.40	8.37	3.80	3.73	11.03	11	1.39	96.73	3	7
25	-58.66	667.91	4336.60	7.34	-1.42	20.15	15.74	18	1.25	149.93	7	14
26	-8.13	576.67	2848.10	6.42	1.71	19.20	13.01	14	1.35	108.68	4	7
27	-22.59	636.54	3387.50	4.34	3.34	21.48	16.96	20	2.28	112.75	4	9
28	-5.60	606.24	4290.80	5.70	6.43	35.81	11.65	13	1.11	120.75	3	6
29	-19.59	584.23	2282.60	4.02	4.99	19.42	13.33	15	1.39	104.50	4	5
30	-52.24	559.94	2513.60	9.56	4.02	20.76	11.37	13	1.39	113.57	4	6
31	-56.18	611.36	2293.60	9.86	2.80	13.97	12.71	14	1.46	123.22	4	9
32	-21.58	547.83	3624.10	9.05	2.28	8.61	11.63	12	1.38	95.00	3	7
33	-21.93	649.01	4697.40	5.31	4.77	25.50	13.24	16	1.17	110.08	4	8
34	38.26	648.52	2536.00	11.41	4.36	14.57	11.58	13	1.16	146.67	6	8
35	6.88	274.55	476.65	6.34	-0.22	8.65	5.40	4	1.86	58.00	4	6
36	3.89	517.52	2470.20	7.66	3.67	12.94	8.95	11	1.12	98.08	4	6
37	-4.60	554.21	3247.50	5.43	4.64	12.62	10.60	10	1.38	87.55	4	5
38	-4.95	650.88	5730.00	8.71	4.37	23.11	13.50	15	1.18	117.58	4	8
39	-29.83	641.01	2843.10	2.37	4.39	32.44	16.75	12	1.48	128.08	6	8
40	-21.89	654.06	4217.90	1.08	5.47	4.03	15.76	18	1.35	112.10	4	7
41	-24.57	526.63	2286.90	3.20	-0.02	22.42	11.09	12	1.38	115.58	5	8
42	-37.39	599.86	4227.50	2.94	2.99	16.21	12.19	13	1.21	114.58	5	7
43	-3.03	720.55	5298.90	10.55	7.60	5.18	11.92	12	0.79	141.33	6	7
44	20.33	474.24	1469.00	8.13	4.82	8.45	8.34	8	1.11	93.17	4	4
45	51.44	646.76	3011.00	8.40	6.15	4.93	12.44	17	1.17	127.00	2	8
46	-60.07	555.19	3836.70	4.86	2.62	19.34	12.53	14	1.67	108.42	5	6
47	4.44	492.98	2005.00	6.86	5.19	5.23	7.64	8	1.10	85.83	3	3
48	36.83	526.00	2441.60	5.29	5.07	8.24	7.29	8	1.00	89.00	3	4
49	-2.60	554.90	5695.80	6.29	6.03	9.64	9.67	11	1.23	114.48	2	6
50	2.79	502.48	3097.50	9.41	3.47	9.38	10.24	13	1.36	93.48	2	9
51	335.51	473.62	2805.20	8.25	3.49	12.31	6.62	7	0.97	86.92	3	5

1. Energia totale (eV); 2. Area superficiale accessibile (Å²); 3. Volume ellissoidale (Å³); 4. Momento di dipolo totale (Debye); 5. log P; 6. Lipolo totale; 7. Φ (flessibilità); 8. Numero di legami rotabili; 9. Indice di Balaban; 10. Indice Elettrotopologico di Kier; 11. Numero di donatori di H; 12. Numero di accettori di H.

È stato inizialmente calcolato un pool di più di 80 descrittori, tra i quali sono stati selezionati 12 descrittori da includere nel database che meglio identificano le caratteristiche molecolari che possono essere correlate con l'attività biologica. (Tabella 8)

La selezione di variabili strutturali è stata fatta considerando le numerose strutture cristallografiche di HIV-1 PR complessato con gli inibitori, le quali supportano il ben noto binding mode delle differenti classi di inibitori di PR. In particolare il volume ellissoidale e l'area superficiale accessibile per quanto riguarda le caratteristiche steriche, log P e lipolo totale per quanto riguarda le caratteristiche idrofobiche, il momento totale di dipolo e l'energia totale *in vacuo* per le caratteristiche energetiche, il numero di donatori e accettori di legami H. Sono stati scelti anche dei descrittori della classe degli indici topologici quali l'indice di flessibilità Φ il quale è correlato con il grado di linearità e la presenza di cicli o ramificazioni;⁴⁵ il numero di legami rotabili, l'indice di Balaban,⁴⁶ somma degli stati elettrotopologici di Kier,^{47, 48} e l'indice topologico di Kier-Hall. Quest'ultimo descrittore codifica informazioni circa la flessibilità, dimensioni, ramificazioni e forma delle molecole.

Su questo set di dodici descrittori è stata effettuata una PCA. I primi 4 PC, con autovalore > 1 e che spiegano il 77.2% della varianza, sono stati selezionati per ulteriori analisi. (vedi Tabella 9)

Tabella 9. Risultati della PCA sul set di descrittori selezionato e sull'indici di similarità (shape)

Descrittori	PC1	PC2	PC3	PC4
1	0.142	0.354	0.245	-0.376
2	-0.407	0.249	-0.008	0.050
3	-0.346	0.253	-0.044	0.069
4	0.092	0.054	0.426	0.722
5	0.022	0.562	-0.346	0.043
6	-0.246	-0.153	-0.243	-0.133
7	-0.417	-0.031	-0.194	0.072
8	-0.404	0.049	-0.140	0.137
9	-0.083	-0.538	-0.349	0.109
10	-0.388	0.159	0.269	-0.015
11	-0.224	-0.192	0.381	-0.514
12	-0.285	-0.228	0.426	0.099
Frazione di varianza spiegata	0.405	0.165	0.109	0.093
Varianza totale spiegata	0.405	0.570	0.679	0.772
Autovalore	4.859	1.979	1.307	1.117

Shape	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Frazione di varianza spiegata	0.209	0.160	0.132	0.058	0.047	0.040	0.033	0.027	0.026	0.022	0.022	0.020
Varianza totale spiegata	0.209	0.369	0.501	0.559	0.606	0.646	0.679	0.706	0.733	0.755	0.776	0.796
Autovalore	10.653	8.175	6.730	2.974	2.377	2.058	1.672	1.387	1.343	1.125	1.110	1.006

Nella prima PC le variabili che hanno maggiore importanza sono l'area superficiale accessibile (-0.41), Φ (-0.42) ed il numero di legami rotabili (-0.40). Questo riflette l'importanza dell'approccio sterico al binding pocket e in particolare la necessità di un alto grado di flessibilità per raggiungere interazione adatte. Nella seconda PC, log P (+0.56) sottolinea l'importanza della lipofilia nel binding mode dovuta alla presenza delle tasche idrofobiche S/S' che consentono la corretta sistemazione dell'inibitore. Nella stessa PC, l'indice di Balaban gioca un ruolo notevole (-0.54). PC3 è ben rappresentata in termini di momento di dipolo (+0.43) e numero di accettori di H (+0.43): questo mostra come le caratteristiche elettroniche, ma principalmente la capacità di formare legami H rappresenta un elemento fondamentale per l'inibizione del processo proteolitico. Queste considerazioni possono essere estese alla quarta PC dove i loadings più alti sono stati trovati per il momento di dipolo (+0.72) e il numero di donatori di H (-0.51).

Anche nel caso degli inibitori della proteasi sono stati calcolati gli indici di similarità per shape. Un'altra PCA è stata fatta sulla matrice degli indici di similarità e il numero di PC significative (12 PC) è stato determinato come sopra (autovalore > 1).(vedi Tabella 9)

Come per RT vengono considerati i dati di attività contro i più comuni ceppi mutanti.(vedi Tabella 7) I mutanti che sono stati scelti sono V82A, V82F e I84V poiché tutte queste mutazioni sono nel sito attivo dell'enzima e sono state riscontrate nella maggior parte dei pazienti che ricevono il trattamento con PI già approvati dal FDA. Inoltre per questi mutanti sono stati trovati in letteratura il maggior numero di dati in singola mutazione: questa scelta consente di costruire un modello più semplice. Prima la DA è stata effettuata considerando a turno solo i derivati per i quali i dati sensibilità/resistenza erano disponibili. I risultati della DA sono riportati nella Tabella 10, ed in generale mostrano una elevata significatività statistica dal momento che in quattro analisi su sei hit-rate = 100%: tutti gli inibitori sono assegnati alla classe vera di appartenenza. Solo nel caso delle mutazioni V82A e V82F, quando si considerano i descrittori, il test Chi2 non respinge la null-hypothesis, anche se i test di Huberty danno risultati statisticamente significativi. Per questi mutanti gli indici di similarità sembrano dare delle informazioni migliori per definire sensibilità o resistenza dei derivati.

Nel caso del mutante I84V, PC2 e PC3 supportano l'importanza della lipofilia e della capacità di formare legami H nello sviluppo della resistenza, mentre il peso del momento di dipolo (+0.43) può essere correlato alla rilevanza delle interazioni elettroniche e di non legame nel sito attivo. Tutti questi dati sono in accordo con le informazioni strutturali disponibili circa la natura della resistenza.⁴⁹

Considerando la validità statistica dei risultati ottenuti, sono stati effettuati nuovi calcoli includendo questa volta anche i derivati per i quali i dati di sensibilità/resistenza non sono noti. La procedura di classificazione implica l'inserimento dei vari composti nelle classi S o R. I dati riportati in Tabella 11 mostrano che è possibile fornire previsioni di attività per un elevato numero di derivati. I migliori risultati sono stati ottenuti nel caso del mutante I84V per il quale sono state trovate il maggior numero di assegnazioni (fino a 32 molecole classificate).

Tabella 10. Risultati della DA^a basati sui PC per i descrittori e gli indici di similarità per le classi R e S

		Descrittori				Shape			
		Predetti R	S	Tot.	Hit Rate	Predetti R	S	Tot.	Hit Rate
V 8	R	3	1	4	75.0%	4	0	4	100%
	S	1	3	4	75.0%	0	4	4	100%
	Totale	4	4	8	75.0%	4	4	8	100%
2 A	<i>Variabili incluse</i> PC1, PC2								
	Chi2	2;	df=1;	p<0.20		8;	df=1;	p<0.01	
	Cpro	50.0%	min. hit-rate=62.5%		50.0%	min. hit-rate=62.5%			
	Cmax	50.0%	min. hit-rate=62.5%		50.0%	min. hit-rate=62.5%			
V 8	R	5	0	5	100%	5	0	5	100%
	S	2	2	4	50.0%	0	4	4	100%
	Totale	7	2	9	77.8%	5	4	9	100%
2 F	<i>Variabili incluse</i> PC2, PC4								
	Chi2	3.2;	df=1;	p<0.10		9;	df=1;	p<0.01	
	Cpro	50.6%	min. hit-rate=63.2%		50.6%	min. hit-rate=63.2%			
	Cmax	55.5%	min. hit-rate=69.4%		55.5%	min. hit-rate=69.4%			
I 8	R	3	0	3	100%	3	0	3	100%
	S	0	5	5	100%	0	5	5	100%
	Totale	3	5	8	100%	3	5	8	100%
4 V	<i>Variabili incluse</i> PC2, PC3								
	Chi2	8;	df=1;	p<0.01		8;	df=1;	p<0.01	
	Cpro	53.1%	min. hit-rate=66.4%		53.1%	min. hit-rate=66.4%			
	Cmax	62.5%	min. hit-rate=78.1%		62.5%	min. hit-rate=78.1%			

I risultati possono essere riassunti come segue.

Il paragone dei risultati per tutti i mutanti della proteasi mostra che in numerosi casi è stata raggiunta una classificazione consistente come S (cfr. derivati **34, 38, 45, 50**) e come R nel caso dei composti **4, 10, 16, 44, 46**. Tali risultati possono essere di rilevante interesse dal momento che si può supporre che per i derivati predetti come S, la presenza di qualsivoglia mutazione non implica necessariamente l'insorgere di resistenza, mentre l'opposto diviene realmente rilevante per le molecole predette come R. Inoltre nel caso della mutazione I84V, i derivati **27** e **29** sono sempre classificati come R. Questi risultati possono essere giustificati considerando l'alto grado di similarità strutturale tra le molecole che sono classificate in maniera consistente. Ciò è confermato anche dall'analisi dei descrittori che hanno maggior importanza nelle PC incluse nella DA. Per quanto riguarda i derivati per i quali esistono dati in letteratura, nella maggior parte dei casi si ottiene una corretta classificazione confermando

così la rilevanza del modello. Nel caso dei descrittori, è stato possibile effettuare un maggior numero di assegnazioni, probabilmente poiché il set scelto rende conto in maniera ottimale dei fattori più importanti per determinare resistenza e suscettibilità.

Nel caso delle mutazioni V82A e V82F, i derivati con attività nota contro i mutanti selezionati sono classificati correttamente con poche eccezioni.

In particolare quando viene considerata la mutazione V82A, le molecole che sono classificate come **R** (escludendo solo il derivato **46**) sono non peptici e questo suggerisce che l'approccio non peptidico potrebbe essere poco proficuo per la progettazione di farmaci attivi contro ceppi che presentano questa mutazione per questa mutazione. Infatti tutte le molecole che sono classificate come sensibili sono di tipo peptidomimetico.

Per quanto concerne i farmaci disponibili in commercio, è interessante notare che LPV (**29**) e RTV (**40**) sono sempre classificati correttamente, APV (**6**) è classificato nella classe opposta solo nel caso della mutazione I84V.

Tabella 11. Classificazione dell'intero set di NNRTI nelle classi **R** and **S**. Legenda: vedi Tabella 6

Entry	Descrittori			Shape		
	V82A	V82F	I84V	V82A	V82F	I84V
1	R		R			
2	S	S	R			
3	R			S		
4	R	R	R			
5	R		R			
6	R	S	S	R	S	S
7	S	R	S			
8		S	R			
9	S	R	S			
10		R		R	R	
11	R		R		S	S
12		S	S			
13	R		R			S
14	S			R	S	S
15	S	R	S			
16	R		R	R	R	
17	R	R	R	R	R	R
18	S	R	S	S	S	S
19	S	R	S	S		
20		R	S			R
21		S		R		
22	S	R	R		S	S
23	R	R	S	R	R	S
24		S				
25		R	S		R	
26		S	S			
27	R	R	R	S		R
28	S	R	R			
29	R	R	R	R	R	R
30		S		S	R	R
31		S	S	S		R
32		S				
33	S	R	S			
34	S	S	S			
35	R		R			
36					R	
37	S	S	R	S	S	R
38	S	S	S	S	S	S
39		R	S			
40	S	R	R	S	R	R
41		R	S			
42	S	S	S	S	S	S
43	S	S	S	S	S	S
44	R			R		R
45	S	S	S			
46	R	R	R	R	R	
47	R		R			
48	R		R			
49	S	S	R			
50		S			S	S
51						

8.3 – Confronto dei dati di analisi multivariata e docking per NNRTI e inibitori di PR

Si è iniziato ricercando nel Brookhaven Protein Databank i files relativi alle strutture della PR e della RT di HIV-1 che presentassero singola mutazione. Come riportato nella Tabella 12 è stato impossibile selezionare una struttura idonea per i mutanti di V106A e L100I a causa della mancanza di strutture cristallografiche del singolo mutante nel PDB. Inoltre per mantenere l'omogeneità nelle conformazioni, in entrambi i casi, sono state scelte strutture che presentassero lo stesso ligando cristallizzato. Secondo la procedura riportata nella sezione materiali e metodi, sono state sfruttate le due modalità di ricerca dei siti al fine di valutare meglio l'interazione ligando-proteina. Esempi di cavità ottenute nel caso di RT e di Pr sono riportati in Figura 18. Come si può notare dalla Tabella 13 il rapporto fra la dimensione dei siti del ceppo mutante con il wild type indica una leggera variazione nell'ampiezza del binding site. Nella maggior parte dei casi la mutazione determina una riduzione del sito attivo, solo nel caso della mutazione V82F ed in modalità PROTEIN SHAPE si ha un netto incremento della cavità a disposizione del ligando.

Tabella 12. Strutture del Protein Data Bank selezionate per il docking.
In rosso le proteine mutate, in giallo i wild type

	LIGANDO	MUTAZIONE	PDB FILE	RISOLUZIONE
TRASCRIPTASI INVERSA	EFAVIRENZ	K103N	1IKV	3.00
	EFAVIRENZ	K103N	1FKO	2.90
	NEVIRAPINA	K103N	1FKP	2.90
	PNU142721	K103N	1KX	2.80
	EFAVIRENZ	Y181C	1JKH	2.50
	TNK-651	Y181C	1JLA	2.50
	NEVIRAPINA	Y181C	1JLB	3.00
	EFAVIRENZ	WILD TYPE	1FK9	2.50
PROTEASI	DUPONT XV638	V82F/I84V	1BWA	1.90
	DUPONT SD146	V82F/I84V	1BWB	1.80
	TIPRANAVIR	V82F/I84V	1D4S	2.50
	A77003	V82A	1HVS	2.25
	DMP450	I84V	1MER	1.90
	DMP323	I84V	1MES	1.90
	DMP323	V82F	1MET	1.90
	DMP323	V82F/I84V	1MEU	1.90
	A77003	WILD TYPE	1HVI	1.80
DMP323	WILD TYPE	1QBS	1.80	

Tabella 13. Dimensioni dei siti attivi calcolati con LIGANDFIT. In basso sono espressi i rapporti fra le dimensioni dei siti attivi delle proteine mutate e le dimensioni dei siti attivi dei WT

FREE POINTS				
	PDB ID	DOCKED LIGAND	PROTEIN SHAPE	CEPPO
RT	1FK9	1716	1979	WILD TYPE
	1FKO	1688	1856	K103N
	1JHK	1558	1774	Y181C
PR	1HVI	6934	1663	WILD TYPE
	1QBS	4153	2014	WILD TYPE
	1HVS	6954	1453	V82A
	1MES	4167	1997	I84V
	1MET	4029	3167	V82F

	MUTAZIONE	RATIO	DL	PS
RT	Y181C	1JKH/1FK9	0.910	0.900
	K103N	1FKO/1FK9	0.980	0.937
PR	I84V	1MES/1QBS	1.003	0.990
	V82F	1MET/1QBS	0.970	1.570
	V82A	1HVS/1HVI	1.003	0.870

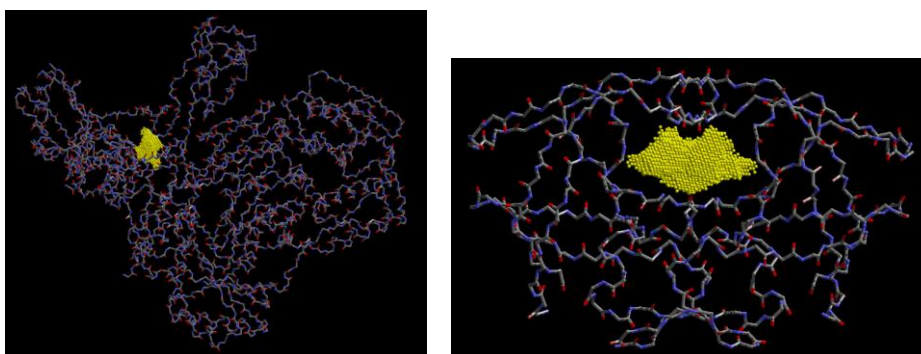


Figura 18. Esempi di siti attivi ottenuti con LIGANDFIT®

Si è passati poi alla fase di scoring. Sono state utilizzate tutte le funzioni di scoring disponibili in LIGANDFIT, per valutare tutti i possibili contributi energetici dell'interazione ligando proteina. (cfr. sezione 7.9) Normalmente, si procede prendendo in considerazione il conformero con il valore più alto, ma nel nostro caso anche i conformeri che presentano un valore di docking basso potrebbero indicare una bassa interazione ligando-proteina correlabile al possibile insorgere di resistenza. Per semplificare, allora, l'enorme quantità di dati prodotti dallo scoring delle molecole si è deciso di prendere in considerazione, per l'analisi comparata che si vuole effettuare in seguito, gli scores maggiori (MAX), quelli minori (MIN) e la media degli scores (MEDIA) relativi ad ogni inibitore analizzato.

Si è deciso poi di calcolare la somma algebrica degli scores:

$$Q = \text{score}(\text{proteina mutata}) - \text{score}(\text{proteina nativa})$$

In questo modo si possono classificare i derivati oggetto dello studio come suscettibili o resistenti. Infatti se $Q > 0$ si può ipotizzare che il composto dockato sia potenzialmente attivo (S) nei confronti della proteina mutata rispetto alla proteina nativa, se $Q < 0$ si può ipotizzare che il composto dockato sia potenzialmente non attivo (R) nei confronti della proteina mutata rispetto alla proteina nativa. Le funzioni di scoring vengono considerate separatamente, il composto è assegnato come R o S sulla base del fatto che almeno due dei tre valori Q siano concordanti. Questo tipo di interpretazione ha anche il vantaggio di poter confrontare i descrittori che definiscono le singole funzioni di scoring con i descrittori usati per l'analisi statistica multivariata. Nella Tabella 14 vengono riportate le percentuali di assegnazioni corrispondenti tra analisi multivariata e docking nel caso di RT, mentre in Tabella 15 vengono riportate quelle ottenute nel caso di PR.

Tabella 14. Percentuali di assegnazioni corrispondenti tra analisi multivariata e docking per RT

DOCKED LIGAND							
DESCRITTORI							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
K103N	42	47.60%	30.90%	50.00%	50.00%	47.60%	45.20%
Y181C	34	61.70%	61.70%	58.80%	52.90%	44.10%	44.10%
SIMILARITA'							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
K103N	31	48.40%	35.40%	41.90%	41.90%	29.00%	29.00%
Y181C	27	51.80%	59.20%	55.50%	51.80%	44.40%	37.00%

PROTEIN SHAPE							
DESCRITTORI							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
K103N	42	50.00%	54.70%	52.38%	40.40%	50.00%	57.14%
Y181C	34	61.70%	58.80%	50.00%	50.00%	35.20%	44.10%
SIMILARITA'							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
K103N	31	45.16%	58.00%	41.90%	29.00%	45.16%	48.40%
Y181C	27	59.20%	59.20%	55.50%	55.50%	37.00%	33.30%

In una prima analisi si nota che per gli inibitori di RT, essendo il sito di legame degli NNRTI molto circoscritto ed in genere completamente occupato dagli inibitori, i risultati sono paragonabili in entrambe le modalità di ricerca dei siti con l'eccezione dei dati ottenuti per la mutazione K103N in modalità DOCKED LIGAND a confronto con i dati ottenuti dall'analisi dei descrittori. Nel caso della RT, LIGSCORE 1 e 2 uniformemente forniscono gli scores più alti nel caso di Y181C e nel caso di K103N rispettivamente. Infatti tali funzioni tengono in considerazione contributi energetici di natura sterica ed elettronica in accordo con i descrittori aventi il maggior peso nelle PC. (cfr. Tabella 4)

Per PR le percentuali migliori sono ottenute in modalità DOCKED LIGAND. Questo può essere giustificato dal fatto che il ligando cristallizzato ed in genere tutti gli inibitori di PR non vanno ad occupare completamente il sito attivo dell'enzima, ma si limitano al massimo ad

occupare 4 delle 6 tasche disponibili, per cui la modalità protein shape calcolando un sito di binding di dimensioni più grandi fa sì che il docking dei ligandi fornisca valori di score variabili. LUDI è lo score che da i risultati maggiormente concordanti per V82A e V82F e questo è pienamente in accordo con i dati multivariati. Infatti La funzione LUDI comprende, come detto prima, contributi energetici derivanti da legami H ideali e interazioni lipofile, proprietà chimico fisiche espresse in log P, μ e donatori di H, cioè i descrittori con maggior peso nelle PC che definiscono ogni singola mutazione. Nel caso di I84V, ma in generale per tutte e tre le mutazioni considerate, le percentuali più elevate si ottengono mediamente quando i dati di docking vengono confrontati con quelli ottenuti dagli indici di similarità. (cfr. Tabella 10)

Tabella 15. Percentuali di assegnazioni corrispondenti tra analisi multivariata e docking per PR

DOCKED LIGAND							
DESCRITTORI							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
V82A	35	45.70%	45.70%	49.00%	45.70%	46.00%	60.00%
V82F	38	50.00%	47.30%	50.00%	50.00%	42.10%	57.80%
I84V	40	57.50%	60.00%	43.00%	53.00%	43.00%	50.00%
SIMILARITA'							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
V82A	21	33.30%	28.50%	43.00%	42.80%	48.00%	52.30%
V82F	20	60.00%	55.00%	60.00%	55.00%	50.00%	65.00%
I84V	20	45.00%	45.00%	45.00%	45.00%	70.00%	25.00%
PROTEIN SHAPE							
DESCRITTORI							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
V82A	35	51.40%	48.50%	43.00%	45.70%	43.00%	51.40%
V82F	38	52.60%	55.20%	36.80%	34.20%	42.10%	47.30%
I84V	40	45.00%	47.50%	55.00%	55.00%	47.50%	50.00%
SIMILARITA'							
	TOT	LIGSC1	LIGSC2	PLP1	PLP2	PMF	LUDI
V82A	21	57.10%	57.10%	38.00%	23.80%	52.00%	66.60%
V82F	20	40.00%	50.00%	40.00%	35.00%	50.00%	50.00%
I84V	20	35.00%	40.00%	60.00%	40.00%	35.00%	50.00%

I risultati riportati nelle Tabelle 14 e 15 possono essere meglio compresi se si visualizza il tipo di interazione che i derivati classificati rispettivamente come S e R instaurano con i siti attivi degli enzimi mutati, tenendo anche conto delle modificazioni avvenute nel sito attivo delle proteine. Nelle Figure 19 e 20 sono riportati l'insieme delle molecole classificate S e R nei confronti delle varie mutazioni oggetto di studio.

Come si può notare, la mutazione in posizione 103 della RT conduce ad un consistente restringimento del sito attivo, così le molecole che vengono coerentemente classificate come resistenti sono quelle più voluminose del set, mentre quelle che vengono classificate come sensibili presentano dimensioni ridotte e possono quindi assumere la tipica conformazione a

farfalla degli NNRTI necessaria per la comparsa dell'attività inibitoria. Nella mutazione in posizione 181 invece la perdita dell'anello aromatico della tirosina, oltre a condurre al restringimento del sito attivo, comporta la perdita da parte degli inibitori della capacità di effettuare interazioni π così come è stato dimostrato nel caso della Nevirapina e Delavirdina.⁵⁰ Infatti le molecole che vengono coerentemente classificate come R hanno una struttura analoga a quella dei due inibitori e tendono quindi ad assumere una conformazione simile ai due farmaci.

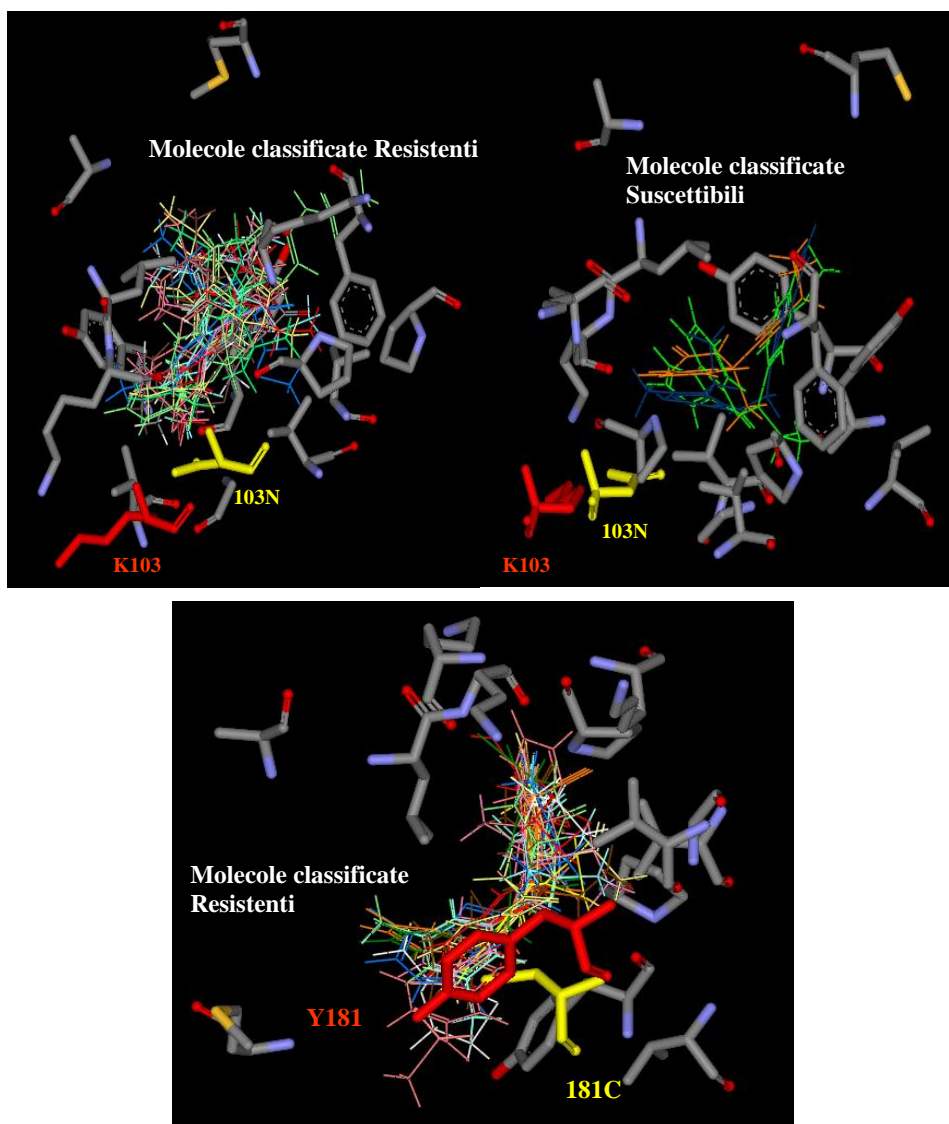


Figura 19. NNRTI: Interazioni delle molecole S e R nei siti attivi

Nel caso di PR le mutazioni V82A e I84V conducono ad una riduzione della lipofilia della tasca di legame così che le molecole che vengono classificate come resistenti, indipendentemente dal fatto che presentino struttura peptidica o non peptidica, incorporano delle grosse porzioni lipofile che riducono la loro capacità di interazione. Considerazioni di tipo opposto possono essere fatte nel caso della mutazione V82F. Infatti la presenza del residuo aromatico della fenilalanina aumenta la lipofilia del sito S/S' così che le molecole che presentano porzioni lipofile sono quelle che vengono coerentemente classificate come sensibili.

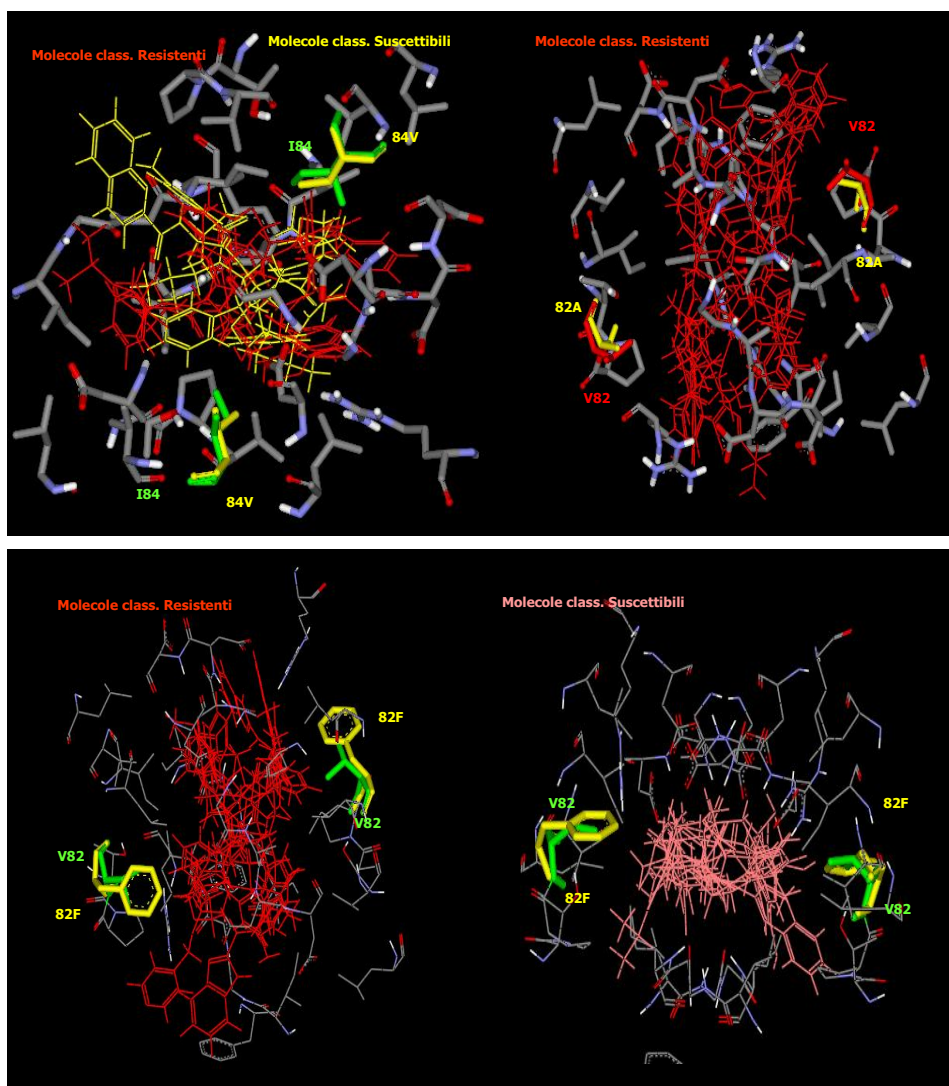


Figura 20. Inibitori di PR: Interazioni delle molecole S e R nei siti attivi

8.4 – Considerazioni finali all'analisi comparata tra metodi multivariati e docking

In questa prima parte del dottorato si è messo a punto un metodo per correlare i risultati ottenuti mediante QSAR e molecular docking per sets selezionati di inibitori della trascrittasi e della proteasi di HIV-1. E' stato messo in evidenza come generalmente sia possibile valutare l'eventuale insorgere di resistenza nei confronti dei mutanti selezionati che inducono modificazioni significative del sito di binding. Questi risultati contribuiranno sicuramente allo sviluppo della progettazione di nuovi inibitori potenzialmente attivi nei confronti di ceppi mutati.

8.5 – Analisi multivariata applicata agli Inibitori di IN

Ho iniziato il lavoro conducendo una ricerca bibliografica per selezionare le molecole, appartenenti al svariate classi strutturali, da cui è stato scelto un set di 450 composti, tra i più attivi nei confronti di HIV-1 IN (Cfr. Appendice III).

Dal momento che un tale set di strutture risulta inadeguato per gli studi QSAR, a causa della disomogeneità di dati da analizzare e della eterogeneità delle strutture a disposizione, si è deciso di applicare metodi di esplorazione multivariata (*Multivariate Exploratory Techniques*), allo scopo di individuare un numero opportuno e limitato di molecole su cui condurre le successive analisi chemiometriche. La tecnica utilizzata è stata CA, che permette di utilizzare il concetto di distanza per raggruppare elementi con proprietà simili all'interno di un cluster.

In una prima fase tutte le 450 molecole sono state inserite in uno storage del software CODESSA PRO. Le strutture vengono inizialmente ottimizzate per la visualizzazione in modalità 3D. Successivamente si è utilizzato il software MOPAC 7 per l'ottimizzazione della struttura attraverso un processo di minimizzazione dell'energia fino ad arrivare ad un valore di gradiente RMS preimpostato: in questo caso è stato applicato l'operatore Hamiltoniano PM3 con un gradiente RMS pari a 0.01. Infine per ciascuna molecola del set, sono stati calcolati tutti i descrittori disponibili, inserendoli in una matrice. I descrittori calcolati sono stati quindi utilizzati per condurre CA.

CODESSA PRO è in grado di calcolare per ciascuna struttura diverse centinaia di descrittori, tra i quali alcuni risultano ridondanti, in quanto derivano da manipolazioni matematiche di altri più rappresentativi. Altri, invece, sono correlati ad un particolare tipo di atomi (es. alogeni), e non risultano pertanto comuni a tutte le molecole. Si è perciò effettuata una selezione tra tutti quelli disponibili per mettere in evidenza i descrittori contenenti un'informazione utilizzabile nel caratterizzare le singole strutture. Si rende così possibile l'applicazione degli algoritmi di clustering alla matrice dei descrittori. Per effettuare questo processo di selezione ci si è basati inizialmente sul raggruppamento in classi proposto dal Theoretical Background di CODESSA PRO, che suddivide l'intera mole di descrittori nelle seguenti categorie:

- Costituzionali
- Topologici
- Geometrici
- Elettrostatici
- CPSA (Charged Partial Surface Area)
- Quanto-chimici
- Correlati agli Orbitali Molecolari
- Termodinamici

Successivamente si sono analizzati i descrittori inclusi in ciascuna classe, con particolare attenzione a quelli che forniscono informazioni circa le connettività degli atomi nella

molecola ed i parametri di distribuzione di carica o di energia, specie a livello delle aree superficiali. Con questo procedimento si è pervenuti infine alla selezione di 38 descrittori molecolari, comuni a tutte le molecole prese in esame, da utilizzare nella conduzione della CA come parametri esprimenti le peculiarità di ciascuna molecola. (Tabella 16)

Tabella 16. Descrittori selezionati

N_A	Numero totale di atomi
$N_{C,r}$	Numero relativo di atomi di C
N_B	Numero totale di legami
N_{Ar}	Numero di legami aromatici
W	Indice di Wiener
${}^m\chi$	Indice di connettività di Randic
${}^m\chi^v$	Indici topologici semplici di Kier e Hall
${}^m\kappa$	Parametro di forma di Kier ($m=1, 2, 3$)
Φ	Indice di flessibilità di Kier
J	Indice di connettività di Balaban
$I_{A,B,C}$	Momenti principali di inerzia
V_M	Volume Molecolare
P	Parametro di polarità
T_1^E	Indice topologico elettronico per gli atomi
T_2^E	Indice topologico elettronico per i legami
ΔH_f^0	Calore di formazione
ϵ_{HOMO}	Contenuto energetico dell' HOMO
ϵ_{LUMO}	Contenuto energetico del LUMO
μ	Momento dipolare della molecola
Min_{NAC}	Minima carica atomica netta
Max_{NAC}	Massima carica atomica netta
CSA	Area superficiale carica
PSA	Area superficiale parziale
$PSAM$	Area superficiale carica positivamente
$PPSAM$	Area superficiale parziale carica positivamente
$NSAM$	Area superficiale carica negativamente
$NPSAM$	Area superficiale parziale carica negativamente

In letteratura vengono forniti generalmente due tipi di dati di attività per gli inibitori di HIV-1 Integrase, espressi entrambi in μM , relativi sia alle reazioni di 3'-Processing che di Strand Transfer. Tuttavia sono presenti dei casi in cui è fornito un solo dato di attività relativo ad una delle due reazioni catalizzate dall'enzima. Non è stato perciò possibile generare clusters validi per lo studio di entrambi i tipi di attività enzimatica; ad un approccio incompleto si è pertanto preferito effettuare due CA distinte, una comprendente tutte le

molecole per le quali sono disponibili dati relativi al 3'-Processing, ed una per le molecole per cui si dispone dati relativi allo Strand Transfer.

Le due matrici dei descrittori (relativi a 375 molecole per 3'-Processing ed a 305 derivati per Strand Transfer) sono state in seguito importate nel pacchetto TSAR 3.2. Per entrambi i tipi di attività sono state eseguite quattro CA distinte, utilizzando tutti gli algoritmi di clustering messi a disposizione dal software, e mantenendo in tutte le analisi la misura della distanza euclidea. Si sono poi effettuate le analisi successive su campioni di molecole rappresentativi dell'intero set esaminato. Si sono scelte pertanto, attraverso l'esame del dendrogramma di ciascuna analisi, dei set di 30, 35, 40 e 50 molecole corrispondenti ai centrotipi di altrettanti clusters individuati, a vari valori di treshold di distanza euclidea. (Tabella 17)

Tabella 17. Valori di distanza euclidea

		WRNN	SL	CL	GA
Strutture 3'-Proc.	30	0.124	0.397	0.397	0.456
	35	0.118	0.417	0.417	0.442
	40	0.108	0.44	0.44	0.42
	50	0.08	0.37	0.37	0.382
Strutture Str.Tr.	30	0.126	0.445	0.244	0.338
	35	0.109	0.47	0.228	0.316
	40	0.1	0.486	0.21	0.296
	50	0.09	0.414	0.189	0.276

I set di molecole sono stati analizzati con CODESSA PRO, per costruire i modelli QSAR. I dati di attività relativi al saggio enzimatico *in vitro* sull'enzima purificato, sono stati espressi come pIC₅₀. Ad un approccio classico (del tipo Leo e Hansch), che prevede la correlazione dei dati di attività con una serie di descrittori scelti a priori, si è preferito utilizzare un approccio che sfruttasse al massimo le potenzialità del software CODESSA, che consente di cercare la migliore correlazione possibile, lineare o multilineare, attraverso l'utilizzo dei migliori descrittori selezionabili tra tutti quelli disponibili.

Attraverso questa procedura (BMLR, *Best Multilinear Regression*) sono state ottenute una serie di equazioni per tutti i set di molecole. Queste includono un numero variabile di descrittori molecolari e predicono le attività delle strutture corrispondenti con diverso grado di precisione e con qualche eccezione rappresentata dagli *outliers*. Non tutte le equazioni risultano tuttavia attendibili allo stesso modo, in quanto i valori di r e r^2_{cv} non conferivano validità statistica, cosicché si è preferito restringere il campo di ricerca ai set di 40 molecole per le attività di 3'-Processing e di 30 molecole per lo Strand Transfer, ottenuti mediante l'algoritmo del Complete Linkage. (Tabella 18) Questa procedura assicura che i cluster siano

scelti massimizzando la distanza euclidea tra i singoli casi, in maniera tale che i centrotipi sono rappresentativi del maggior numero possibile di casi. Inoltre questi set di molecole offrono il miglior compromesso tra il numero di descrittori utilizzati nella relazione, il numero di molecole esaminate ed il numero di casi non validi.

Tabella 18. Modelli QSAR selezionati

	Cluster	N° Strutture	N° Descr.	r^2	r^2_{cv}	N° Outliers
3'-Proc.	Com.Link.	40	8	0.81	0.7	0
Str. Tr.	Com.Link.	30	6	0.89	0.8	0

Il primo modello (Figura 21) è relativo alla reazione di 3'-Processing, e consta di 8 descrittori, per un totale di 40 molecole.

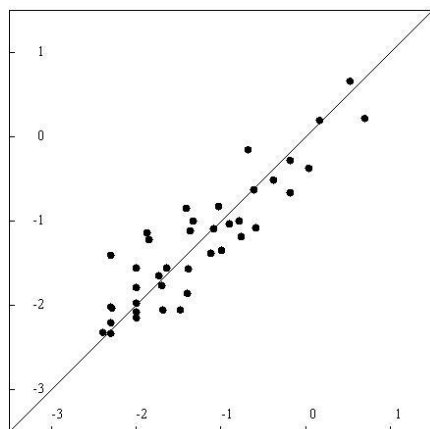


Figura 21. pIC₅₀ Osservati vs Predetti per 3'-Processing

Riesce a prevedere i valori di attività per tutte le molecole con buona validità statistica ($r^2=0.81$; $r^2_{cv}=0.7$), secondo la seguente equazione:

$$pIC_{50} = 74.429 + 0.177 \text{ HOLUGAP} - 15.321 \text{ MC_MAX} + 2.403 \text{ AIFCONT_1} - 0.312 \text{ NATM_N} \\ + 34.193 \text{ IRI_MAX} + 0.237 \text{ KSHP_2} - 0.348 \text{ HMIN} + 37.382 \text{ HDSA2TM_A}$$

dove:

HOLUGAP è la differenza in termini di energia tra HOMO e LUMO della molecola, grandezza indicativa della capacità di ionizzazione della molecola;

MC_MAX è la massima energia di scambio per il legame H-C;

AIFCONT_1 è l'indice di informazione sulla composizione atomica media di ordine 1, fondato sul concetto matematico di entropia e di contenuto di informazione, avente la proprietà di possedere un valore indipendente dalla numerazione dei vertici del grafo;

NATM_N è il numero di atomi di azoto;

IRI_MAX è l'indice di massima reattività elettronica per l'atomo di carbonio;

KSHP_2 è il parametro di forma di Kier di ordine 2, derivato dal grafo molecolare senza tenere conto dei legami ad idrogeno, normalizzato sul numero totale di atomi della molecola, che dà una rappresentazione dell'intera molecola;

HMIN è il numero minimo di siti accettori e donatori di legami ad H;

HDSA2TM_A è un indice, riferito al numero di siti accettori di legami ad idrogeno, del rapporto tra l'area superficiale dei siti H-donatori e l'area superficiale totale.

Dal tipo di descrittori utilizzati in questa equazione è possibile ipotizzare per la comparsa dell'attività inibitoria un notevole contributo di fattori di tipo elettronico, quali la distribuzione di carica all'interno della molecola, il grado di reattività dei singoli atomi o la composizione atomica media; emerge come determinante inoltre il ruolo del legame ad idrogeno, rappresentato sia dal parametro di distribuzione dei siti H-donatori all'interno della molecola, che dalla presenza di un numero minimo di siti accettori di legami ad idrogeno. Da non sottovalutare infine il ruolo degli atomi di azoto, il cui doppietto di non legame potrebbe fornire una base per legami aggiuntivi.

Il secondo modello (Figura 22), relativo alla reazione di Strand Transfer, consta di 6 descrittori.

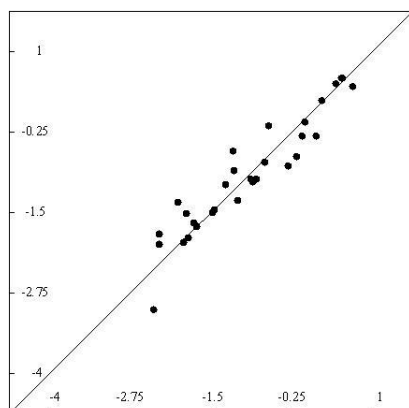


Figura 22. pIC₅₀ Osservati vs Predetti per per Strand Transfer

Esso può essere utilizzato per prevedere l'attività di un set di 30 molecole (Cfr. Appendice IV) con una buona approssimazione, valutata attraverso il valore di $r^2=0.89$ ($r^2_{cv}=0.8$).

L'equazione è la seguente:

$$\text{pIC}_{50} = -6.030 + 93.911 \text{ HDCA2TM_A} + 0.378 \text{ CO_CI_MAX} - 412.938 \text{ O_ERI_AVG} \\ - 0.942 \text{ HDCA2} + 0.007 \text{ WNSA1M} - 0.004 \text{ MNO_POL_A}$$

dove:

HDCA2TM_A è un indice, riferito al numero di siti accettori di legami ad idrogeno, del rapporto tra la superficie carica dei siti H-donatori e l'area superficiale totale;

CO_CI_MAX è il massimo valore di interazione coulombiana tra gli atomi di C ed O;

O_ERI_AVG è un parametro che esprime il valore medio di reattività elettrofilica per l'atomo di ossigeno;

HDCA2 è il rapporto tra la superficie carica dei siti donatori di legami ad idrogeno, calcolata secondo Zefirov, e la superficie molecolare totale;

WNSA1M è un parametro che esprime il rapporto pesato tra l'area superficiale carica negativamente e l'area superficiale totale;

MNO_POL_A è invece il fattore di polarizzabilità alfa, ossia il fattore di proporzionalità tra il momento dipolare indotto su una molecola ed il campo elettrico in cui la molecola è immersa.

Come nel caso precedente relativo alle attività di 3'-Processing, in questo modello sono presenti dei descrittori che forniscono informazioni sulla distribuzione dei siti donatori o accettori di legami ad idrogeno sulla superficie della molecola, come HDCA2TM_A o HDCA2. A differenza del precedente, però, nella reazione di Strand Transfer assumono maggiore importanza le forze di interazione tra cariche localizzate sugli atomi di C ed O, come lasciano intendere CO_CI_MAX ed O_ERI_AVG. Indicativa, a tal proposito, è inoltre la presenza di parametri che definiscono l'importanza delle superfici cariche negativamente o del momento dipolare, come WNSA1M o MNO_POL_A.

In questa seconda parte del lavoro di dottorato è stato possibile proporre un set di molecole, rappresentative delle classi più attive di inibitori di HIV-1 IN riportate in letteratura. Si sono sviluppati dei modelli QSAR che riescono a prevedere con buona approssimazione l'attività dei derivati selezionati in funzione di un buon numero di descrittori. I risultati di questo lavoro, quindi, oltre ad aver contribuito a chiarire l'interazione dell'enzima Integrasi con i suoi inibitori, possono fornire delle linee guida per la sintesi di nuovi composti che posseggano caratteristiche strutturali adeguate a massimizzarne la performance farmacologica.

9. CONCLUSIONI

Durante il mio dottorato di ricerca mi sono occupato dello studio degli inibitori dei tre principali enzimi di HIV-1 ossia RT, PR e IN nel tentativo di esplorare le caratteristiche che possono modulare l'attività biologica. Lo studio effettuato si è articolato in due parti. La prima ha riguardato RT e PR, la seconda IN.

Nella prima parte mi sono proposto di sviluppare un metodo per valutare l'eventuale insorgere di resistenza o suscettibilità di due diversi set di molecole contro i più comuni ceppi mutanti di RT (L100I, K103N, V106A e Y181C) e PR (V82A, V82F, I84V). Nell'analisi dei dati, ottenuti dalle tecniche chemiometriche utilizzate (PCA, DA) e dal docking molecolare (LIGANDFIT), si è potuta evidenziare una coincidenza tra i descrittori che assumono maggiore importanza nelle PCA e quelli che costituiscono i contributi energetici che determinano i valori di score. Inoltre è stato possibile osservare che nel caso degli inibitori di RT le molecole che vengono classificate come resistenti presentano dimensioni maggiori di quelle classificate come sensibili e che l'interazione π risulterebbe di fondamentale importanza nella modulazione dell'attività. Mentre nel caso degli inibitori di PR si è potuto osservare che in relazione al tipo di mutazione, la presenza di porzioni lipofile voluminose potrebbe definire il comportamento come suscettibili oppure resistenti: in questo caso è di fondamentale importanza la visualizzazione delle interazioni sul sito attivo così come possibile con le tecniche di docking.

Nella seconda parte mi sono occupato degli inibitori di IN. Non è stato possibile in questo caso effettuare uno studio sul possibile insorgere della resistenza in quanto non esistono ancora farmaci disponibili in commercio e non sono perciò noti dati riguardanti la possibile comparsa di mutazioni. L'obiettivo è stato allora quello di selezionare e classificare una serie di strutture rappresentative delle classi di inibitori più efficaci. Si sono sviluppati dei modelli QSAR che riescono a prevedere con buona approssimazione l'attività dei derivati selezionati in funzione di un buon numero di descrittori (in tutti i modelli $N_{Strut}/N_{Desc} \leq 5$), il cui significato può essere correlato alle caratteristiche strutturali della molecola. Si è evidenziata l'importanza di alcuni parametri, come l'abbondanza di siti donatori od accettori di legami ad idrogeno, che risultano necessari al fine di garantire una stabile interazione ligando-enzima.

Le tecniche chemiometriche e di docking utilizzate durante il mio dottorato di ricerca nello studio degli inibitori di HIV-1 si sono rivelate essere di particolare interesse e possono fornire delle linee guida per la sintesi di nuovi composti con caratteristiche strutturali adeguate a massimizzarne l'interazione farmaco-recettore e quindi di conseguenza l'attività biologica.

BIBLIOGRAFIA

- 1 Seydel, J.; Schaper, K. *Chemische Struktur und Biologische Aktivität von Wirkstoffen Weinheim*; Verlag Chemie 1979
- 2 Condra, J.; Emin, E.; Gotlib, L.; Graham, J.; Schlabach, A.; Wolfgang, J.; Colonna, R.; Sardana, V. *Antimicrob. Agents Chemother.*, 1992, 36, 441.
- 3 De Clercq, E. *Biochem. Pharmacol.*, 1994, 47, 155.
- 4 De Clercq, E. *Int. J. Immunother.*, 1994, 10, 145.
- 5 Nanni, R.; Ding, J.; Jacobo-Molina, A.; Hughes, S.; Arnold, E. *Perspect. Drug Discovery Design*, 1993, 1, 129.
- 6 Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R.; Boyer, P.; Hughes, S.; Pauwels, R.; Andries, K.; Jansen, P.; Arnold, E. *J. Mol. Biol.*, 1994, 243, 369.
- 7 Tong, L.; Cardozo, M.; Jones, P.-J.; Adams, J. *Bioorg. Med. Chem. Lett.*, 1993, 7, 721.
- 8 Asante-Appiah, E.; Skalka, A. M. *Antivir. Res.*, 1997, 36, 139
- 9 Pommier, Y.; Marchand, C.; Neamati, N. *Antivir. Res.*, 2000, 47, 139
- 10 Akimov, D. V.; Filimonov, D. A.; Poroikov, V. V. *Pharm. Chem. J.*, 2002, 36, 575
- 11 Sotriffer, C.; Ni, A.; McCammon, J.A. *J. Med. Chem.*, 2000, 43, 4109
- 12 Swanstrom, R.; Erona, J. *Pharm. Ther.*, 2000, 86, 145.
- 13 Menendez-Arias, L. *TRENDS in Pharm. Sci.*, 2002, 8, 381.
- 14 Leung, D.; Abbenante, G.; Fairlie, D. *J. Med. Chem.*, 2000, 43, 305.
- 15 Todeschini, R. *Introduzione alla Chemiometria*, EdiSesim, 1998
- 16 Manly, B. Factor Analysis. In *Multivariate statistical methods. A primer*. Chapman and Hall 1986, Chapter 7.
- 17 Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quant. Chem.* 1980, 17, 1185-1189.
- 18 *TSAR 3.2 Reference Guide*, Oxford Molecular, 1998
- 19 web address: <http://www.statsoft.com/textbook/stathome.html>
- 20 Darlington, R. B. *Regression and Linear Models*, McGraw-Hill, 1990
- 21 Neter, J.; Wassermann, W.; Kutner, M. H. *Applied Linear Regression Models (2nd ed.)*, Homewood, IL: Irwin. 1989
- 22 Eriksson, L.; Johansson, E.; Muller, M.; Wold, S. *J. Chemometr.*, 2000, 14, 599
- 23 Xu, Q. S.; Liang Y. Z. *Chemom. Intell. Lab. Syst.*, 2001, 56,1
- 24 Stone, M. *J. R. Stat. Soc.*, B. 1977, 38, 44
- 25 Martens, H. A.; Dardenne, P. "Validation and verification of regression in small data sets", 1998, 44, 99
- 26 Edington, E. S. *Randomization tests*, Marsel Dekker Inc., New York and Basel, 1981, 195
- 27 Huberty, C.; *Psychol. Bull.* 1984, 95, 156-169.
- 28 http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html
- 29 Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldaman, M. *J. Mol. Graph. Model.*, 2003, 21, 289
- 30 Oldfield, T. *Acta Cryst.* 2001, D57, 82
- 31 Oldfield, T. *Acta Cryst.* 2001, D57, 696
- 32 *LigandFit User Manual*, Accelerlys Inc, 2003
- 33 Verkhivker, G.; Bouzida, D.; Gehlhaar, D.; Rejto, P.; Arthurs, S.; Colson, A.; Freer, S.; Larson, V.; Luty, B.; Marrone, T.; Rose, P. *J. Comp. Aid. Mol. Des.*, 2000, 14, 731
- 34 Gehlhaar, D.; Verkhivker, G.; Rejto, P.; Sherman, C.; Fogel, D.; Fogel, L.; Freer, S. *Chem. Biol.*, 1995, 2, 317
- 35 Muegge, I.; Martin, Y. *J. Med. Chem.*, 1999, 42, 791
- 36 web address: <http://www.niaid.nih.gov/daids/dtpdb>
- 37 Dati ricavati dai seguenti articoli e note citate in essi: (a) De Clercq, E. *Med. Res. Rev.*, 1996, 16, 125. (b) Campiani, G.; Ramunno, A.; Maga, G.; Nacci, V.; Fattorusso, C.; Catalanotti, B.; Morelli, E.; Novellino, E. *Curr. Pharm. Design*, 2002, 8, 615. (c) Corbett, J.; Gearhart, L.; Ko, S.; Rodgers, J.; Cordova, B.; Klabe, R.; Erickson-Viitanen, S. *Bioorg. Med. Chem. Lett.*, 2000, 10, 193. (d) Tramontano, E.; Piras, G.; Mellors, J.; Putzolu, M.;

Bazmi, H.; La Colla, P. *Biochem. Pharm.*, 1998, 56, 1583. (e) Xu, Z.; Buckheit Jr., R.; Stup, T.; Flavin, M.; Khilevich, A.; Rizzo, J.; Lin, L.; Zembower, D. *Bioorg. Med. Chem. Lett.*, 1998, 8, 2179. (f) Balzarini, J. *Biochem. Pharm.*, 1999, 58, 1. (g) Silvestri, R.; Artico, M.; Massa, S.; Marceddu, T.; De Montis, F.; La Colla, P. *Bioorg. Med. Chem.*, 10, 2000, 253. (h) Arranz, M. *Bioorg. Med. Chem.*, 1999, 7, 2811. (i) [Buckheit, R.](#); [White, E.](#); [Fliakas-Boltz, V.](#); [Russell, J.](#); [Stup, T.](#); [Kinjerski, T.](#); [Osterling, M.](#); [Weigand, A.](#); [Bader, J.](#) *Antimicrob. Agents Chemother.*, 1999, 43, 1827. (l) [Zembower, D.](#); [Liao, S.](#); [Flavin, M.](#); [Xu, Z.](#); [Stup, T.](#); [Buckheit, R.](#); [Khilevich, A.](#); [Mar, A.](#); [Sheinkman, A.](#) *J. Med. Chem.*, 1997, 40, 1005.

38 Dati ricavati dai database disponibili nei seguenti siti web: (a) <http://www.mediscover.net/antiviralintro.cfm>; (b) http://resdb.lanl.gov/Resist_DB/default.htm; (c) <http://hivdb.stanford.edu/>; (d) <http://www.hivresistanceweb.com/>.

39 I descrittori molecolari sono stati calcolati secondo i metodi e le equazioni riportate su TSAR 3.2 Reference Guide, Oxford Molecular Limited (1998).

40 Viswanadhan, V.; Ghose, A.; Revankar, G.; Robins, R. *Atomic J. Chem. Inf. Comput. Sci.* 1989, 29, 163-172.

41 Bachelier, L.; Jeffrey, S.; Hanna, G., et al., *J Virol.* 75, 4999-5008 (2001).

42 Petropoulos, C. J., Parkin, N. T., Limoli, K.L., et al., *Antimicrob. Agents Chemother.* 44, 920-8 (2000).

43 Dati ricavati dai seguenti articoli e note citate in essi: (a) Ghosh, A.; Swanson, L.; Chunfeng, L.; Hussain, K.; Cho, H.; Walters, E.; Holland, L.; Buthod, J. *Bioorg. Med. Chem. Lett.* 2002., 12, 1993-1996. (b) Lee, Y.; Lee, Y.; Lee, J.; Kim, S.; Lee, C.; Park, H. *Bioorg. Med. Chem. Lett.* 2000, 10, 2625-2627. (c) Munroe, J.; Hornback, W.; Campbell, J.; Ouellette, M.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Baxter, A.; Su, K.; Campanale, K. *Bioorg. Med. Chem. Lett.* 1995, 5, 2885-2890. (d) Takashiro, E.; Nakamura, Y.; Miyamoto, S.; Ozawa, Y.; Sugiyama, A.; Fujimoto, K. *Bioorg. Med. Chem.* 1999, 7, 2105-2114. (e) Takashiro, E.; Hayakawa, I.; Nitta, T.; Kasuya, A.; Miyamoto, S.; Ozawa, Y.; Yagy, R.; Yamamoto, I.; Shibayama, T.; Nakagawa, A.; Yabe, Y. *Bioorg. Med. Chem.* 1999, 7, 2063-2072; (f) St Clair, M.; Millard, J.; Rooney, J.; Tisdale, M.; Parry, N.; Sadler, B.; Blum, M.; Painter, G. *Antiviral Res.* 1996, 29, 53-56. (g) Wilkerson, W.; Akamike, E.; Cheatham, W.; Hollis, A.; Collins, R.; De Lucca, I.; Lam, P.; Ru, Y. *J. Med. Chem.* 1996, 39, 4299-4312. (h) Kempf, D.; Marsh, K.; Paul, D.; Knigge, M.; Norbeck, D.; Kohlbrenner, W.; Codacovi, L.; Vasvanonda, S.; Bryant, P.; Wang, X.; Wideburg, N.; Clement, J.; Plattner, J.; Erickson, J. *Antimicrob. Agents Chemother.* 1991, 35, 2209-2214. (i) Doyon, L.; Croteau, G.; Thibeault, D.; Poulin, F.; Pilote, L.; Lamarre, D. *J. Virol.* 1996, 70, 3763-3769. (j) Bechtold, C.; Patick, A.; Alam, M.; Greytok, J.; Tino, J.; Chen, P.; Gordon, E.; Ahmad, S.; Barrish, J.; Zahler, R.; Lin, P.; Colonna, R. *Antimicrob. Agents Chemother.* 1995, 39, 374-379. (k) Hilgeroth, A.; Wiese, M.; Billich, J. *J. Med. Chem.* 1999, 42, 4729-4732. (l) Jhadv, P.; Woerner, F.; Lam, P.; Hodge, N.; Eyermann, C.; Man, H.; Daneker, W.; Bachelier, L.; Rayner, M.; Meek, J.; Erickson-Viitanen, S.; Jackson, D.; Calabrese, J.; Schadt, M.; Chang, C. *J. Med. Chem.* 1998, 41, 1446-1455. (m) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H.; Danielson, H.; Classon, B.; Unge, T.; Samuelsson, B.; Hulten, J.; Hallberg, A.; Karlen, A. *J. Med. Chem.* 2001, 44, 155-169. (n) Han, Q.; Chang, C.; Li, R.; Ru, Y.; Jadhav, P.; Lam, P. *J. Med. Chem.* 1998, 41, 2019-2028. (o) Martin, S.; Dorsey, G.; Gane, T.; Hillier, M. *J. Med. Chem.* 1998, 41, 1581-1597. (p) Hage, S.; Prasad, V.; Boyer, F.; Domagala, J.; Ellsworth, E.; Gajda, C.; Hamilton, H.; Markoski, L.; Steinbaugh, B.; Tait, B.; Lunney, E.; Tummino, P.; Ferguson, D.; Hupe, D.; Nouhan, C.; Grachek, S.; Saunders, J.; VanderRoest, S. *J. Med. Chem.* 1997, 40, 3707-3711. (q) Jadhav, P.; Ala, P.; Woerner, F.; Chang, C.; Garber, S.; Anton, E.; Bachelier, L. *J. Med. Chem.* 1997, 40, 181-191. (r) Hodge, C.; Aldrich, P.; Bachelier, L.; Chang, C.; Eyermann, C.; Garber, S.; Grubb, M.; Jackson, D.; Jadhav, P.; Korant, B.; Lam, P.; Maurin, M.; Meek, J.; Otto, M.; Rayner, M.; Reid, C.; Sharpe, T.; Shum, L.; Winslow, D.; Erickson-Viitanen, S. *Chem. Biol.* 1996,

Formattato: Italiano (Italia)

Codice campo modificato

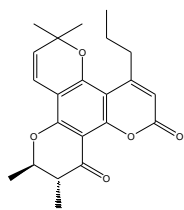
Codice campo modificato

Codice campo modificato

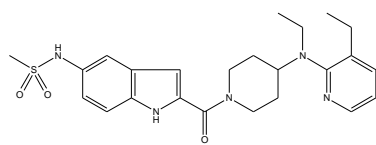
3, 301-314. (s) Kaltenbach, R. III; Trainor, G.; Getman, D.; Harris, G.; Garber, S.; Cordova, B.; Bacheler, L.; Jeffrey, S.; Logue, K.; Cawood, P.; Klabe, R.; Diamond, S.; Davies, M.; Saye, J.; Jona, J.; Erickson-Viitanen, S. *Antimicrob. Agents Chemother.* 2001, 45, 3021-3028. (t) Alterman, M.; Andersson, H.; Garg, N.; Ahlsen, G.; Lovgren, S.; Classon, B.; Danielson, H.; Kvarnstrom, I.; Vrang, L.; Unge, T.; Samuelsson, B.; Hallberg, A. *J. Med. Chem.* 1999, 42, 3835-3844. (u) Yamaguchi, K.; Honda, M.; Ikigai, H.; Yukihiko, H.; Shimamura, T.; Inhibitory Effects of (-)-*Antiviral Res.* 1992, 53, 19-34. (v) Ghosh, A.; Krishnan, K.; Walters, E.; Cho, W.; Cho, H.; Koo, Y.; Trevino, J.; Holland, L.; Buthod, J. *Bioorg. Med. Chem. Lett.* 1998, 8, 979-982. (w) Beaulieu, P.; Wernic, D.; Abraham, A.; Anderson, P.; Bogri, T.; Bousquet, Y.; Croteau, G.; Guse, I.; Lamarre, D.; Liard, F.; Paris, W.; Thibeault, D.; Pav, S.; Tong, L. *J. Med. Chem.* 1997, 40, 2164-2176. (x) Dorsey, B.; Levin, R.; Mcdaniel, S.; Vacca, J.; Guare, J.; Darke, P.; Zugay, J.; Emmini, E.; Schleif, W.; Quintero, J.; Lin, J.; Chen, I.; Ostovic, D.; Anderson, P.; Huff, J. *J. Med. Chem.* 1994, 37, 3443-3451. (y) Farese-Di Giorgio, A.; Rouquayrol, M.; Greiner, J.; Aubertin, A.; Vierling, P.; Guedj, R. *Antivir. Chem. Chemother.* 2000, 11, 97-110. (z) Ghosh, A.; Lee, H.; Thompson, W.; Culberson, C.; Holloway, M.; Mckee, S.; Munson, P.; Duong, T.; Smith, A.; Darke, P.; Zugay, J.; Emmini, E.; Schlieff, W.; Huff, J.; Anderson, P. *J. Med. Chem.* 1994, 37, 1177-1188. (aa) Nair, A.; Bonin, I.; Tossi, A.; Welsh, W.; Miertus, S. *J. Mol. Graph. Model.* 2002, 21, 171-179. (ab) Kageyama, S.; Mimoto, T.; Murakawa, Y.; Nomizu, M.; Ford, Jr H.; Shirasaka, T.; Gulnik, S.; Erickson, J.; Takada, K.; Hayashi, H.; Broder, S.; Kiso, Y.; Mitsuya, H. *Antimicrob. Agents Chemother.* 1993, 37, 810-817. (ac) Kato, R.; Mimoto, T.; Fukazawa, T.; Morohashi, N.; Kiso, Y. USP 5,932,550; *Chem Abst.* 2001, 134, 311436. (ad) Cozens, R.; Bold, G.; Capraio, H.; Fassler, A.; Mestan, J.; Lang, M.; Poncioni, B.; Stover, D.; Rosel, J. *Antivir. Chem. Chemother.* 1996, 7, 294-299; (ae) Lazdins, J.; Bold, G.; Capraro, H.; Cozens, R.; Fassler, A.; Flesch, G.; Klimkait, T.; Lang, M.; Mestan, J.; Poncioni, B.; Rosel, J.; Stover, D.; Tintelnot-Blomley, M.; Walker, M.; Woods-Cook, K. *Schweiz Med. Wochenschr* 1996, 43, 1849-1851. (af) Marastoni, M.; Bazzaro, M.; Salvatori, S.; Bortolotti, F.; Tomatis, R. *Bioorg. Med. Chem.* 2001, 9, 939-945. (ag) Molla, A.; Vasavanonda, S.; Kumar, G.; Sham, H.; Johnson, M.; Grabowski, B.; Denissen, J.; Kohlbrenner, W.; Plattner, J.; Leonard, J.; Norbeck, D.; Kempf, D. *Virology* 1998, 250, 255-262. (ah) Shepherd, T.; Jungheim, L.; Baxter, A. *Bioorg. Med. Chem. Lett.* 1994, 4, 1391-1396. (ai) Jungheim, L.; Shepherd, T.; Baxter, A.; Burgess, J.; Hatch, S.; Lubbenhusen, P.; Wiskerchen, M.; Muesing, M. *J. Med. Chem.* 1996, 39, 96-108. (aj) Munroe, J.; Shepherd, T.; Jungheim, L.; Hornback, W.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Su, K.; Campanale, K.; Baxter, A.; Colacino, J. *Bioorg. Med. Chem. Lett.* 1995, 5, 2897-2902. (ak) Hornback, W.; Munroe, J.; Shepherd, T.; Hatch, S.; Muesing, M.; Wiskerchen, M.; Colacino, J.; Baxter, A.; Su, K.; Campanale, K. *Bioorg. Med. Chem. Lett.* 1995, 5, 2891-2896. (al) Thompson, W.; Fitzgerald, P.; Holloway, M.; Emmini, E.; Darke, P.; Mckeever, B.; Schleif, W.; Quintero, J.; Zugay, J.; Tucker, T.; Schwering, J.; Homnick, C.; Nunberg, J.; Springer, J.; Huff, J. *J. Med. Chem.* 1992, 35, 1685-1701. (am) Ho, D.; Toyoshima, T.; Mo, H.; Kempf, D.; Norbeck, D.; Chen, C.; Wideburg, N.; Burt, S.; Erickson, J.; Singh, M. *J. Virol.* 1994, 68, 2016-2020. (an) Pyring, D.; Lindberg, J.; Rosenquist, A.; Zuccarello, G.; Kvarnstrom, I.; Zhang, H.; Vrang, L.; Unge, T.; Classon, B.; Hallberg, A.; Samuelsson, B. *J. Med. Chem.* 2001, 44, 3083-3091. (ao) Murphy, P.; O'Brien, J.; Gorey-Feret, L.; Smith III, A. *Bioorg. Med. Chem. Lett.* 2002, 12, 1763-1766. (ap) Smith III, A.; Hirschmann, R.; Pasternak, A.; Yao, W.; Sprengeler, P.; Holloway, K.; Kuo, L.; Chen, Z.; Darke, P.; Schleif, W. An Orally Bioavailable *J. Med. Chem.* 1997, 40, 2440-2444. (aq) Kaldor, S.; Kalish, V.; Davies, J.; Shetty, B.; Fritz, J.; Appelt, K.; Burgess, J.; Campanale, K.; Chirgadze, N.; Clawson, D.; Dressman, B.; Hatch, S.; Tatlock, J.; Et, A. *J. Med. Chem.* 1997, 40, 3979-3985. (ar) Fujiwara, T.; Sato, A.; El-Farrash, M.; Miki, S.; Abe, K.; Isaka, Y.; Kodama, M.; Wu, Y.; Chen, L.; Harada, H.; Sugimoto, H.; Hatanaka, M.; Hinuma, Y. *Antimicrob. Agents Chemother.* 1998, 42, 1340-1345. (as) Lamarre, D.; Croteau, G.;

- Wardrop, E.; Bourgon, L.; Thibeault, D.; Clouette, C.; Vaillancourt, M.; Cohen, E.; Pargellis, C.; Yoakim, C.; Anderson, P. *Antimicrob. Agents Chemother.* 1997, 41, 965-971. (at) Etmayer, P.; Billich, A.; Hecht, P.; Rosenwirth, B.; Gstach, H. *J. Med. Chem.* 1996, 39, 3291-3299. (au) Kempf, D.; Marsh, K.; Denissen, J.; McDonald, E.; Vasavanonda, S.; Flentge, C.; Green, B.; Fino, L.; Park, C.; Kong, X.; Wideburg, N.; Saldivar, A.; Ruiz, L.; Kati, W.; Sham, H.; Robins, T.; Stewart, K.; Hsu, A.; Plattner, J.; Leonard, J.; Norbeck, D. *Proc. Natl. Acad. Sci. USA* 1995, 92, 2484-2488. (av) Komai, T.; Yagi, R.; Suzuki-Sunagawa, H.; Sakurai, M.; Higashida, S.; Sugano, M.; Handa, H.; Mohri, H.; Yasuoka, A.; Oka, S.; Yabe, Y.; Nishigaki, T.; Kimura, S.; Shimada, K. *Biol. Pharm. Bull.* 1997, 20, 175-180. (aw) Billich, A.; Scholz, D.; Charpiot, B.; Gstach, H.; Lehr, P.; Peichl, P.; Rosenwirth, B. *Antivir. Chem. Chemother.* 1995, 6, 327-336. (ax) Ala, P.; Deloskey, R.; Huston, E.; Jadhav, P.; Lam, P.; Eyermann, C.; Hodge, C.; Schadt, M.; Lewandowski, F.; Weber, P.; McCabe, D.; Duke, J.; Chang, C. *J. Biol. Chem.* 1998, 273, 12325-12331. (ay) Patel, M.; Rodgers, J.; Mchugh, Jr. R.; Johnson, B.; Cordova, B.; Klabe, R.; Bachelier, L.; Erickson-Viitanen, S.; Ko, S. *Bioorg. Med. Chem. Lett.* 1999, 9, 3217-3220. (az) Hilgeroth, A.; Billich, A.; Lilie, H. *Eur. J. Med. Chem.* 2001, 36, 367-374. (ba) Getman, D.; Decrescenzo, G.; Heintz, R.; Reed, K.; Talley, J.; Bryant, M.; Clare, M.; Houseman, K.; Marr, J.; Mueller, R.; Vazquez, M.; Shieh, H.; Stallings, W.; Stegeman, R. *J. Med. Chem.* 1993, 36, 288-291. (bb) Bryant, M.; Getman, D.; Smidt, M.; Marr, J.; Clare, M.; Dillard, R.; Lansky, D.; Decrescenzo, G.; Heintz, R.; Houseman, K.; Reed, K.; Stolzenbach, J.; Talley, J.; Vazquez, M.; Mueller, R. *Antimicrob. Agents Chemother.* 1995, 39, 2229-2234. (bc) Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S. *Bioorg. Med. Chem.* 2001, 9, 3059-3063. (bd) Turner, S.; Strohbach, J.; Tommasi, R.; Aristoff, P.; Johnson, P.; Skulnick, H.; Dolak, L.; Seest, E.; Tomich, P.; Bohanon, M.; Horng, M.; Lynn, J.; Chong, K.; Hinshaw, R.; Watenpaugh, K.; Janakiraman, M.; Thaisrivongs, S. *J. Med. Chem.* 1998, 41, 3467-3476. (be) Molla, A.; Mo, H.; Vasavanonda, S.; Han, L.; Lin, C.; Hsu, A.; Kempf, D. *Antimicrob. Agents Chemother.* 2002, 46, 2249-2253. (bf) Ghosh, A.; Kincaid, J.; Cho, W.; Walters, D.; Krishnan, K.; Hussain, K.; Koo, Y.; Cho, H.; Rudall, C.; Holland, L.; Buthod, J. *Bioorg. Med. Chem. Lett.* 1998, 8, 687-690. (bg) Han, W.; Pelletier, J.; Hodge, N. Tricyclic Ureas: a New Class of HIV-1 Protease Inhibitors. *Bioorg. Med. Chem. Lett.* 1998, 8, 3615-3620
- 44 Dati ricavati dai seguenti articoli e dai database disponibili sui seguenti siti web: (a) Partaledis, J.; Yamaguchi, K.; Tisdale, M.; Blair, E.; Falcione, C.; Maschera, B.; Myers, R.; Pazhanisamy, S.; Futer, O.; Cullinan, A. *J. Virol.* 1995, 69, 5228-5235. (b) <http://www.hivresistanceweb.com/>; (c) http://srdata.nist.gov/hivdb/resistance2_losalamo_s_HIVPR_inhibitors.html; (d) Klabe, R.; Bachelier, L.; Ala, P.; Erickson-Viitanen, S.; Meck, J. *Biochemistry* 1998, 37, 8735-8742. (e) <http://hivdb.stanford.edu/cgi-bin/PIResiNote.cgi>; (f) <http://www.mediscover.net/antiviralintro.cfm>; (g) Sham, H.; Kempf, D.; Molla, A.; Marsh, K.; Kumar, G.; Chen, C.; Kati, W.; Stewart, K.; Lal, R.; Hsu, A.; Betebenner, D.; Korneyeva, M.; Vasavanonda, S.; McDonald, E.; Saldivar, A.; Wideburg, N.; Chen, X.; Niu, P.; Park, C.; Jayanti, V.; Grabowski, B.; Granneman, G.; Sun, E.; Japour, A.; Leonard, J.; Plattner, J.; Norbeck, D. *Antimicrob. Agents Chemother.* 1998, 42, 3218-3224. (h) Patick, A.; Mo, H.; Markowitz, M.; Appelt, K.; Wu, B.; Musick, L.; Kalish, V.; Kaldor, S.; Reich, S.; Ho, D.; Webber, S. *Antimicrob. Agents Chemother.* 1996, 40, 292-297.
- 45 Hall, L.; Kier, L. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modelling. In *Reviews in computational chemistry*. Lipkowitz and Boyd 1992, Chapter 9
- 46 Balaban, A. *Chem. Phys. Lett.* 1982, 89, 399-404.
- 47 Hall, L.; Mohny, B.; Kier, L. *J. Chem. Inf. Comput. Sci.* 1991, 31, 76-82.
- 48 Kier, L.; Hall, L. *Pharm. Res.* 1990, 7, 801-807.
- 49 Swanson, R.; Erona, J. *Pharmacol. Therap.* 2000, 86, 145-170.

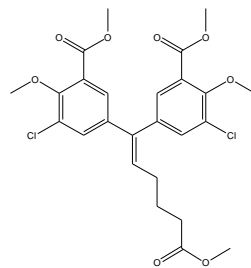
Appendice I – Inibitori della Trascrittasi



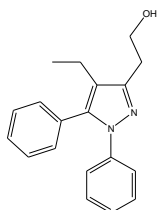
(1) 12-OXOCALANOLIDE



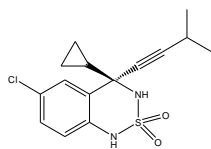
(2) 21-AAP-BHAP



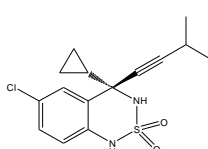
(3) ADAM-II



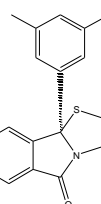
(4) AN-29-PNU-32945



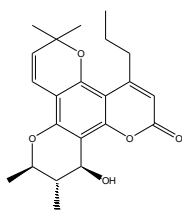
(5) BENZOTHIADIAZINE-2(R)



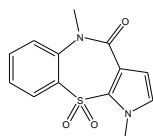
(6) BENZOTHIADIAZINE-2(S)



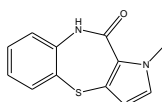
(7) BM-510836



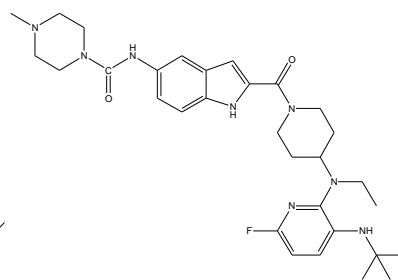
(8) CALANOLIDE-A



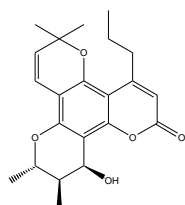
(9) COMPD-1



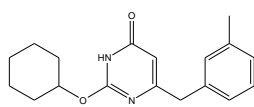
(10) COMPD-2



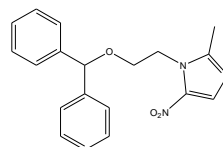
(11) COMPD-36



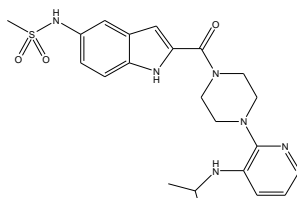
(12) COSTATOLIDE



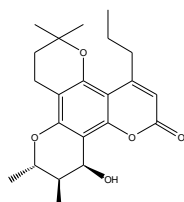
(13) DABO-12c



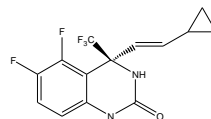
(14) DAMNI-6D



(15) DELAVIRDINE

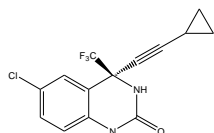


(16) DIHYDROSTATOLIDE

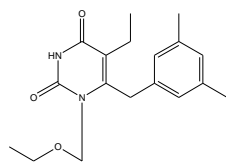


(17) DPC-082

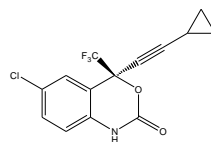
Appendice I – Continua



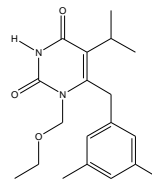
(18) DFC-961



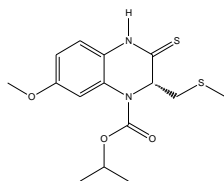
(19) E-EBU-dM



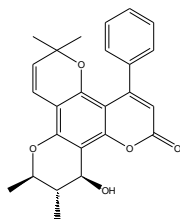
(20) EFAVIRENZ



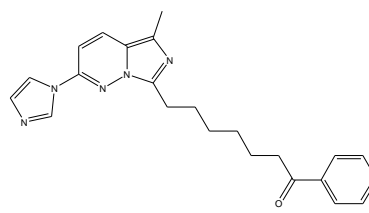
(21) GCA-186



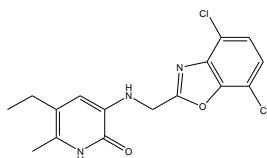
(22) HBY-097



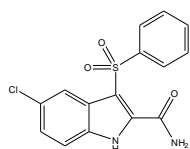
(23) INOPHYLLUM-B



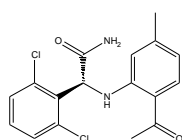
(24) IPPH



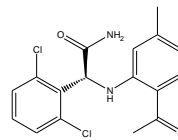
(25) L-697661



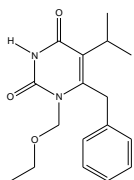
(26) L-737126



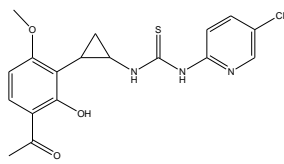
(27) LOVIRIDE(R)



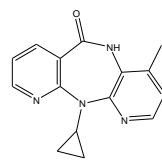
(28) LOVIRIDE(S)



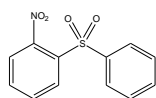
(29) MKC-442



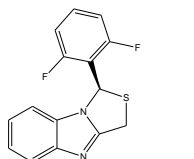
(30) MSA-300



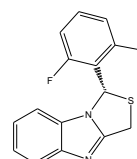
(31) NEVIRAPINE



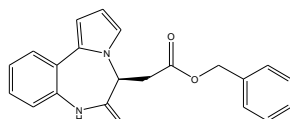
(32) NPPS



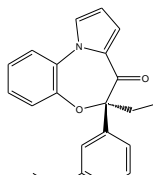
(33) NSC-625487(R)



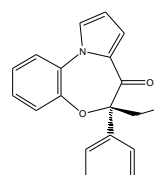
(34) NSC-625487(S)



(35) PBD

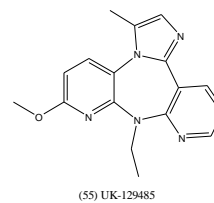
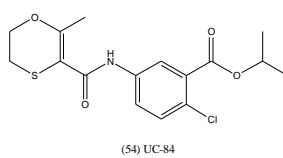
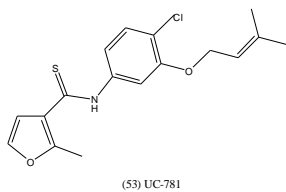
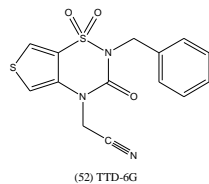
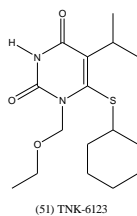
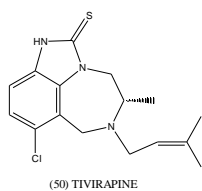
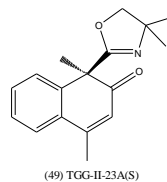
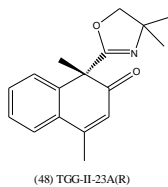
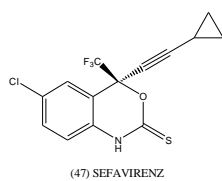
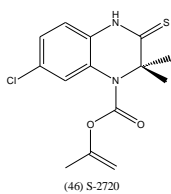
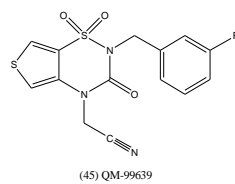
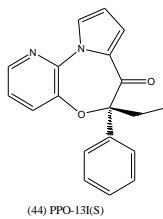
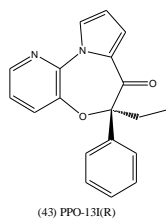
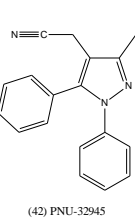
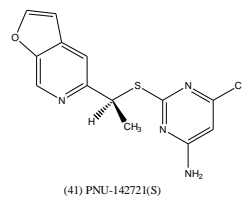
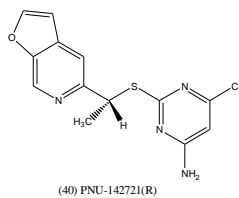
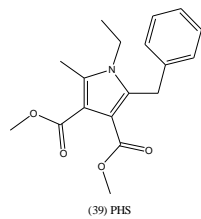
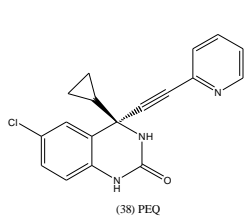


(36) PBO-13B(R)

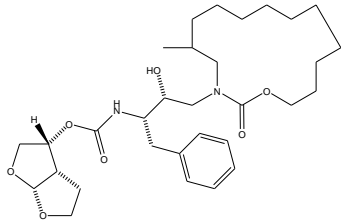


(37) PBO-13B(S)

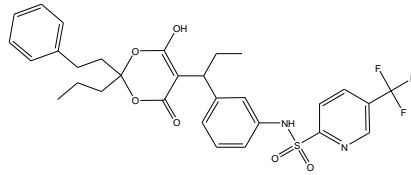
Appendix I – Continua



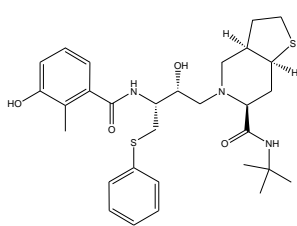
Appendice II – Inibitori della Proteasi



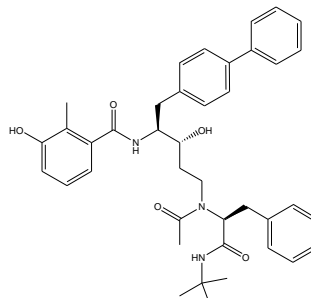
(1) 33_DER_CYCLOURETHAN



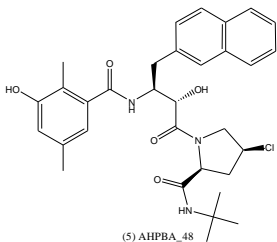
(2) 6H-1,3-DIOS-4D



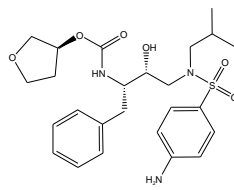
(3) AG_1350



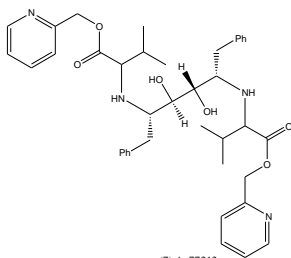
(4) AHPBA_38



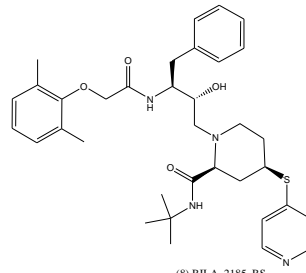
(5) AHPBA_48



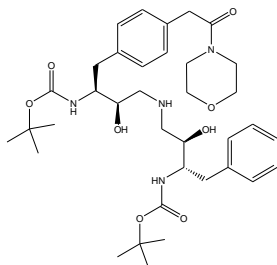
(6) AMPRENAVIR



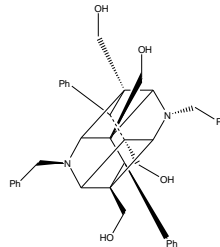
(7) A_77212



(8) BILA_2185_BS

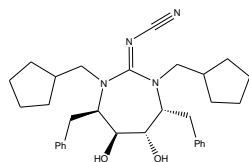


(9) BMS_186318

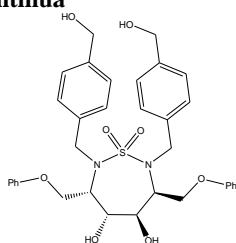


(10) CAGEDIM_3C

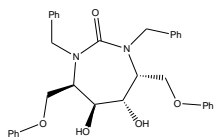
Appendix II – Continua



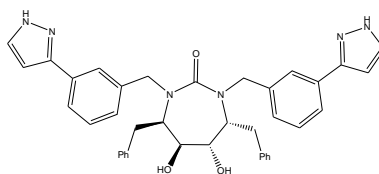
(11) CYANGUAN_81



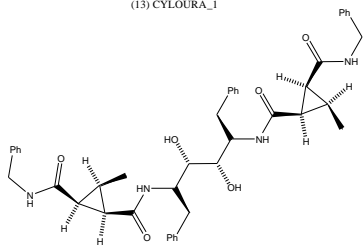
(12) CYCLOSULFAMIDE_25



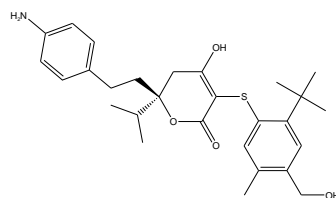
(13) CYLOURA_1



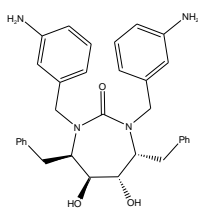
(14) CYCLOURA_10A



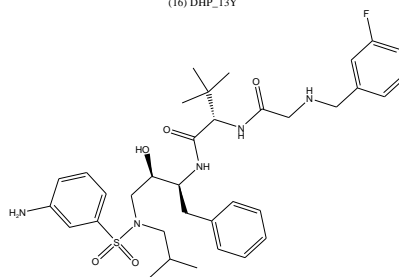
(15) CYCPTR_11



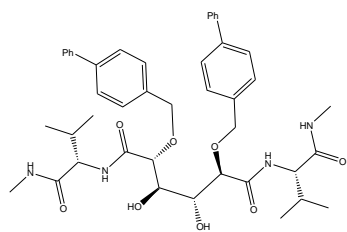
(16) DHP_13Y



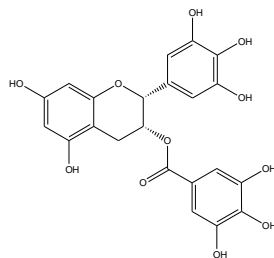
(17) DMP_450



(18) DPC_681

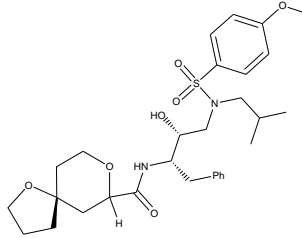


(19) DUPLCTERM_4

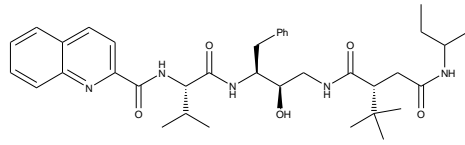


(20) EGCg

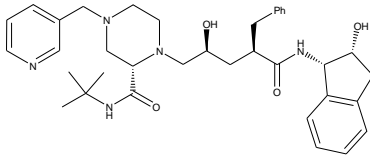
Appendix II – Continua



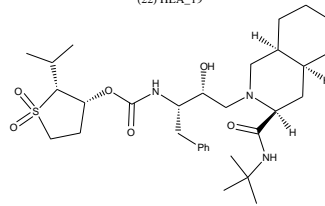
(21) ESC_14



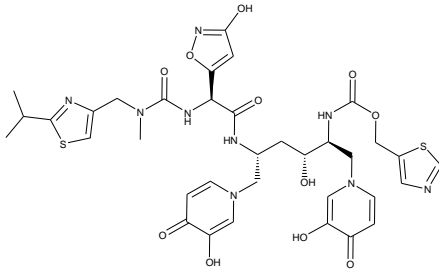
(22) HEA_19



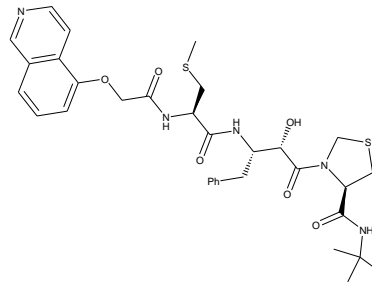
(23) INDINAVIR



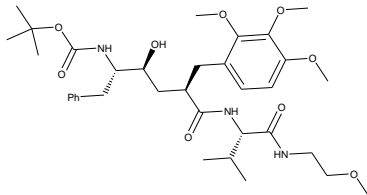
(24) ISOQUIN_URETHAN_ANALOG_2



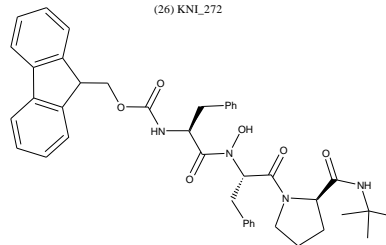
(25) I_12



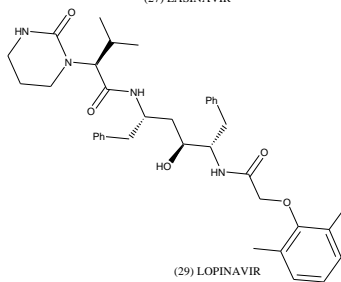
(26) KNL_272



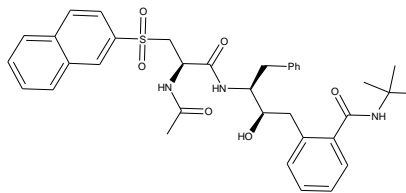
(27) LASINAVIR



(28) LC_3

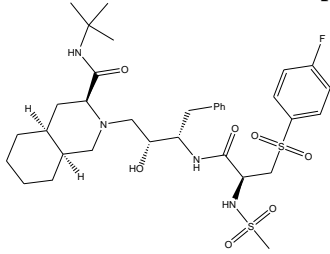


(29) LOPINAVIR

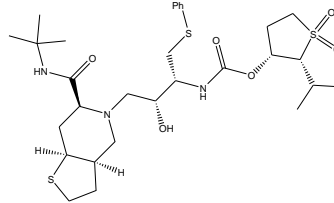


(30) LY_289612_ANAL1

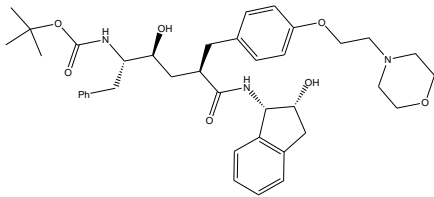
Appendix II – Continua



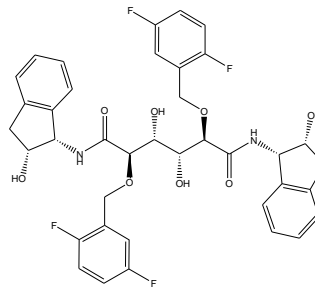
(31) LY_314163



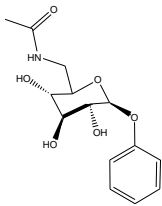
(32) LY_326188



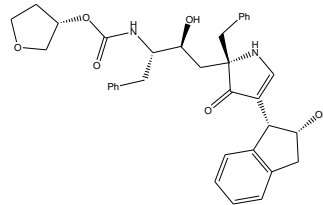
(33) L_689502



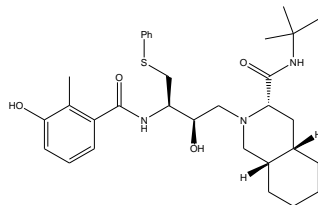
(34) MANNAR_9F



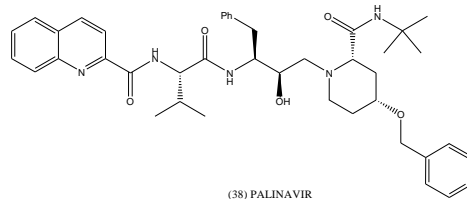
(35) MANNOPYRANOSIDE_18



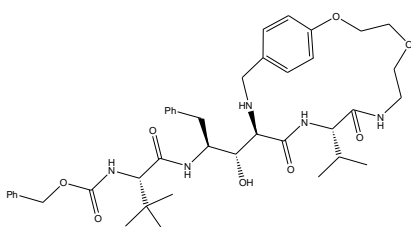
(36) MW_583



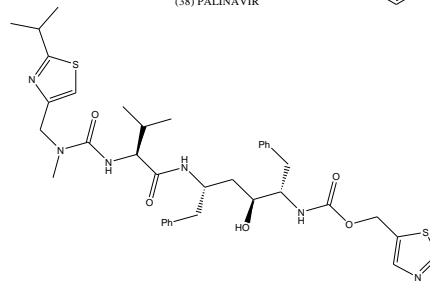
(37) NELFINAVIR



(38) PALINAVIR

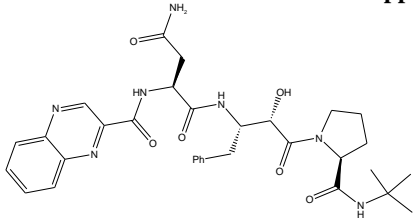


(39) PARACYCLOPHANE_DER_1

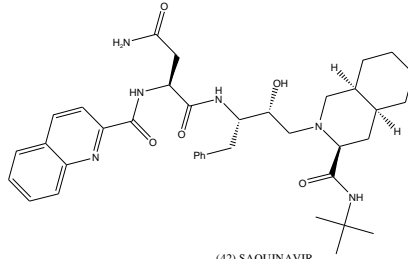


(40) RITONAVIR

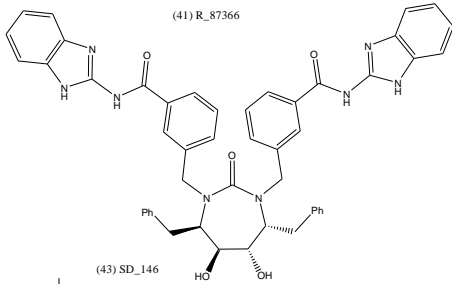
Appendix II – Continua



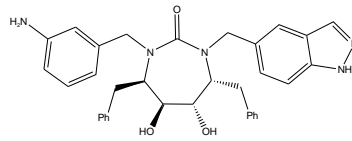
(41) R_87366



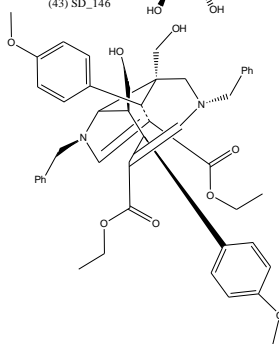
(42) SAQUINAVIR



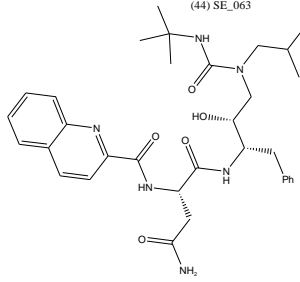
(43) SD_146



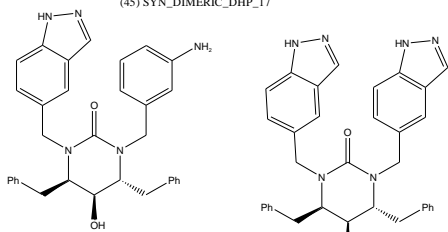
(44) SE_063



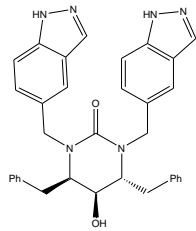
(45) SYN_DIMERIC_DHP_17



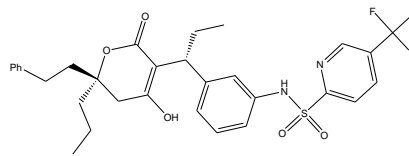
(46) TELINAVIR



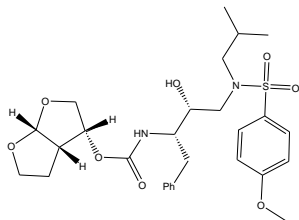
(47) THP_19



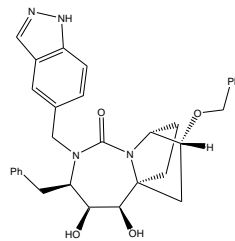
(48) THP_20



(49) TIPRANAVIR



(50) TMC_126



(51) UREATRICAL_9

Appendice III – Attività (µM) degli inibitori di Integrasi inclusi nello studio

ID	Derivati	3'-Proc.	Str. Tr.	Bibliografia
1	1,3-DCQA	0.68	1.08	1 2 3
2	1,4-DCQA	9.51	7.8	
3	1-MO-3,5-DCQA	0.47	0.47	
4	3,4-DCQA	0.25	0.46	
5	3,5-DCQA	0.64	0.66	
6	3-DCQA	87.8	45.8	
7	4,5-DCQA	0.79	0.54	
8	Bis-cinnamoyl-benzene 12a	0.8	Non Disponibile	4
9	Bis-cinnamoyl-benzene 12b	30	Non Disponibile	
10	Arylamide 97	33	33	1 5
11	Arylamide 98	350	350	
12	Arylamide 99	350	350	
13	ATC Derivate 125	4	Non Disponibile	1
14	ATC Derivate 126	7	Non Disponibile	
15	ATC Derivate 127	32.8	23	
16	ATC Derivate 128	49.7	43.9	
17	ATC Derivate 133	61.8	68.5	
18	ATC Derivate 134	25.6	27.2	
19	Aurintricarboxylic Acid (ATC)	10	Non Disponibile	1
20	Bis-cathecol 64	110	110	
21	Bis-cathecol 65	100	100	
22	Bis-cathecol 66	21.4	5.4	
23	Bis-cathecol 67	0.5	Non Disponibile	
24	Bis-cathecol 68	1.7	Non Disponibile	
25	Bis-cathecol 69	7	Non Disponibile	
26	Bis-cathecol 70	36	Non Disponibile	
27	Bis-cathecol 71	0.5	Non Disponibile	
28	Bis-cathecol 72	35	Non Disponibile	
29	Bis-cathecol 73	35	Non Disponibile	
30	Bisarylamide 105	0.98	0.81	1 5
31	Bisarylamide 106	200	200	
32	Bisarylamide 107	0.23	0.11	
33	Bisarylamide 108	200	200	1 2
34	Caffeic Acid	200	200	
35	Caffeic Acid Derivate 89	9	4	
36	Caffeic Acid Derivate 90	1.38	4.71	
37	CAPE	7	19	
38	CAPE-Amide 74	100	100	
39	CAPE-Amide 75	33	33	
40	CAPE-Amide 76	Non Disponibile	4	
41	Chalcone 10a	1.3	Non Disponibile	4
42	Chalcone 10b	20	Non Disponibile	
43	Chalcone 10c	100	Non Disponibile	
44	Chalcone 10d	2.3	Non Disponibile	
45	Chicoric Acid Amide 25	4.1	4.1	6
46	Chicoric Ester 6	9.8	6.2	
47	Chicoric Ester 7	33.3	4.6	
48	Chicoric Ester 13	38.4	7.8	
49	Chicoric Ester 14	24.8	7.6	
50	Chicoric Ester 15	0.4	0.4	
51	Chicoric Ester 16	27.5	7.5	
52	Chicoric Ester 17	27.5	6.5	
53	Chicoric Ester 18	10.1	10.1	

54	Chicoric Ester 19	2.5	3.2	
55	Chicoric Ester 20	2.8	3.8	
56	Chicoric Ester 24	2.1	2.8	
57	Chicoric Ester 27	4.2	1.7	
58	Chicoric Ester 28	3.3	1.7	
59	Chlorogenic Acid	142	142	2
60	Chloroparallelic Acid	2.2	1.6	
61	Chloroquine	13.1	5.7	
62	Cosalane	25	Non Disponibile	1
63	Cosalane Derivate 130	100	Non Disponibile	
64	Cosalane Derivate 131	2.2	Non Disponibile	
65	Cosalane Derivate 132	2.2	Non Disponibile	
66	Coumarin 8	43.4	38.8	
67	Coumarin 9	80.6	69.2	
68	Coumarin10	46	20.5	
69	Coumarin 11	34	11.1	
70	Coumarin 12	3	1.2	
71	Coumarin 13	141	75	
72	Coumarin 14	198	76	
73	Coumarin 15	100	50	
74	Coumarin 16	58	12	
75	Coumarin 17	66	54	
76	Coumarin 18	100	148	7
77	Coumarin 19	75	133	
78	Coumarin 20	29	23	
79	Coumarin 21	19	16	
80	Coumarin 22	14	13	
81	Coumarin 23	10	9.5	
82	Coumarin 24	5.5	3.7	
83	Coumarin 25	8.5	14	
84	Coumarin 26	8	Non Disponibile	
85	Coumarin 36	121	122	
86	Coumarin 37	35.7	22.5	
87	Coumarin 117	1.5	0.8	
88	Coumarin 118	0.4	0.3	
89	Coumarin 119	7.5	2.7	
90	Coumarin 120	10.5	12	1
91	Coumarin 121	4.2	3.5	7
92	Coumarin 122	7	1.8	
93	Coumarin 123	50	24	
94	Curcumin	95	40	
95	Curcumin Derivate 11a	100	Non Disponibile	
96	Curcumin Derivate 11b	0.7	Non Disponibile	4
97	Curcumin Derivate 11c	100	Non Disponibile	
98	Curcumin Derivate 11d	100	Non Disponibile	
99	Curcumin Derivate 79	6	3.1	8
100	Cyclovalone 7a	100	Non Disponibile	
101	Cyclovalone 7b	0.9	Non Disponibile	
102	Cyclovalone 7c	1	Non Disponibile	4
103	Cyclovalone 7d	0.2	Non Disponibile	
104	Cyclovalone 7e	0.2	Non Disponibile	
105	Cyclovalone 7f	0.2	Non Disponibile	
106	Derivate 139	6.1	6.1	
107	Derivate 140	8.3	4	
108	Derivate 141	11.6	4.9	1
109	Derivate 142	1.1	0.9	

110	Diarylsulfone 151	51.8	59	
111	Diarylsulfone 152	60	76	
112	Diarylsulfone 153	6.5	6.1	
113	Diarylsulfone 154	4.5	4.9	
114	Diarylsulfone 155	25.9	21.9	
115	Diarylsulfone 156	0.6	1.3	
116	Diarylsulfone 157	20.9	18.6	
117	Diarylsulfone 158	29.8	29.5	
118	Diarylsulfone 159	55	83	
119	Diarylsulfone 160	67.5	65.6	
120	Diarylsulfone 161	2.9	2.5	
121	Diarylsulfone 162	70	64.2	
122	Dihydranaphtic Acid	172	172	5
123	Dihydranaphtic Derivate 17	5.4	4.7	
124	Dihydranaphtic Derivate 19	53.3	62.4	
125	Dihydranaphtic Derivate 20	200	200	
126	Dihydranaphtic Derivate 21	200	200	
127	Dihydranaphtic Derivate 23	58.4	52.7	
128	Doxorubicin	0.9	2.4	1
129	Hydrazide 5	46	33	9
130	Hydrazide 18	100	100	
131	Hydrazide 24	70	57.5	
132	Hydrazide 25	100	100	
133	Hydrazide 26	100	100	
134	Hydrazide 109	2.07	0.73	
135	Hydrazide 110	9.1	5.8	
136	Hydrazide 111	2.3	1.1	
137	Hydrazide 112	200	200	
138	Hydrazide 113	100	100	
139	Hydrazide 114	100	100	
140	Hydrazide 115	6.7	5.2	
141	Hydrazide 116	100	100	
142	Hydroxylated Aromatic Derivate 77	0.8	0.1	1
143	Hydroxynaphthylquinone	5.7	2.5	4
144	Indane 8a	3	Non Disponible	
145	Indane 8b	86	Non Disponible	
146	L-Chicoric Acid	0.15	0.15	1
147	Mitoxantrone	3.8	8	10
148	NSC48240	0.026	0.021	
149	NSC49858	0.294	0.188	
150	NSC229488	0.229	0.202	
151	NSC246119	0.371	0.283	
152	NSC310217	0.002	0.002	
153	NSC371056	0.030	0.017	
154	NSC626436	0.177	0.103	
155	NSC632901	0.182	0.101	
156	NSC641547	0.224	0.135	
157	NSC642710	0.005	0.005	
158	Primaquine	15.3	3.6	1
159	Quercetin	19.4	11	
160	Quinic Acid	Non Disponible	Non Disponible	8
161	Rosmarinic Acid	9	4	
162	Salicylhydrazide 1	2.07	0.73	11
163	Salicylhydrazide 18	288	167	
164	Salicylhydrazide 19	192	155	
165	Salicylhydrazide 29	0.6	2.8	

166	Salicylhydrazide 30	0.9	0.6	
167	Salicylhydrazide 31	0.8	0.6	
168	Salicylhydrazide 32	1.4	2.6	
169	Salicylhydrazide 33	0.6	0.9	
170	Salicylhydrazide 34	0.9	7.4	
171	Salicylhydrazide 35	0.8	0.6	
172	Salicylhydrazide 36	0.5	0.5	
173	Salicylhydrazide 37	2	2	
174	Salicylhydrazide 38	2.7	2	
175	Salicylhydrazide 39	293	293	
176	Salicylhydrazide 40	293	293	
177	Stictic Acid	4.4	2.9	1
178	Styrylquinoline 12	100	100	12
179	Styrylquinoline 13	2.2	3.4	
180	Styrylquinoline 14	3.7	2.8	
181	Styrylquinoline 15	3.7	2.8	
182	Styrylquinoline 16	3	Non Disponibile	
183	Styrylquinoline 17	2.4	1	
184	Styrylquinoline 18	0.8	Non Disponibile	
185	Styrylquinoline 19	0.3	0.4	
186	Styrylquinoline 20	0.26	Non Disponibile	
187	Styrylquinoline 21	2.7	0.6	
188	Sulfonic Acid 144	0.47	0.29	1
189	Sulfonic Acid 145	3.4	2.7	
190	Sulfonic Acid 146	27.9	14.3	
191	Sulfonylamide 147	28.6	14	
192	Sulfonylamide 148	24	19	
193	Sulfonylamide 149	49	23.6	
194	Sulfonylamide 150	48.3	13.7	
195	Suramin	0.25	0.09	
196	Tetrahydroxycurcumin	300	300	8
197	Tyrphostin 80	1.9	0.8	1
198	Tyrphostin 81	0.4	0.2	
199	Tyrphostin 82	0.8	0.4	
200	Tyrphostin 83	3.3	1.9	
201	Tyrphostin 84	0.7	0.6	
202	Tyrphostin 85	0.5	0.2	
203	Tyrphostin 86	1.4	1.1	
204	Virensic Acid	4.6	6.5	
205	Virensic Acid Derivate 138	42.2	28.1	13
206	3-Aryl-1,3-Diketo Derivate 4a	100	2	
207	3-Aryl-1,3-Diketo Derivate 7a	7.8	1.83	
208	3-Aryl-1,3-Diketo Derivate 7b	7.2	1.28	
209	3-Aryl-1,3-Diketo Derivate 7c	82	6.5	
210	3-Aryl-1,3-Diketo Derivate 10b	65	0.52	
211	3-Aryl-1,3-Diketo Derivate 10e	Non Disponibile	0.48	
212	3-Aryl-1,3-Diketo Derivate 13b	35	0.65	
213	3-Aryl-1,3-Diketo Derivate 13h	40	40	
214	3-Aryl-1,3-Diketo Azo-derivate 2	100	0.48	14
215	3-Aryl-1,3-Diketo Azo-derivate 4	100	2	
216	3-Aryl-1,3-Diketo Azo-derivate 8	100	0.26	
217	3-Aryl-1,3-Diketo Azo-derivate 9	70	0.32	
218	3-Aryl-1,3-Diketo Azo-derivate 10	100	0.36	
219	3-Aryl-1,3-Diketo Azo-derivate 11	100	1.53	
220	3-Aryl-1,3-Diketo Azo-derivate 12	100	6.1	

221	3-Aryl-1,3-Diketo Azo-derivate 13	100	8.5	
222	3-Aryl-1,3-Diketo Azo-derivate 14	100	24	
223	3-Aryl-1,3-Diketo Azo-derivate 15	100	25	
224	3-Aryl-1,3-Diketo Azo-derivate 16	100	1.5	
225	3-Aryl-1,3-Diketo Azo-derivate 17	85	3.5	
226	4-Aryl-2,4-Dioxobutanoic Derivate 1	Non Disponibile	0.17	15
227	4-Aryl-2,4-Dioxobutanoic Derivate 2	Non Disponibile	5.67	
228	4-Aryl-2,4-Dioxobutanoic Derivate 3	Non Disponibile	0.22	
229	4-Aryl-2,4-Dioxobutanoic Derivate 4	Non Disponibile	0.18	
230	4-Aryl-2,4-Dioxobutanoic Derivate 5	Non Disponibile	0.1	
231	4-Aryl-2,4-Dioxobutanoic Derivate 6	Non Disponibile	0.16	
232	4-Aryl-2,4-Dioxobutanoic Derivate 7	Non Disponibile	0.5	
233	4-Aryl-2,4-Dioxobutanoic Derivate 8	Non Disponibile	0.5	
234	4-Aryl-2,4-Dioxobutanoic Derivate 9	Non Disponibile	0.1	
235	4-Aryl-2,4-Dioxobutanoic Derivate 10	Non Disponibile	0.1	
236	4-Aryl-2,4-Dioxobutanoic Derivate 11	Non Disponibile	1	
237	4-Aryl-2,4-Dioxobutanoic Derivate 12	Non Disponibile	0.1	
238	4-Aryl-2,4-Dioxobutanoic Derivate 13	Non Disponibile	0.25	
239	4-Aryl-2,4-Dioxobutanoic Derivate 14	Non Disponibile	0.1	
240	4-Aryl-2,4-Dioxobutanoic Derivate 15	Non Disponibile	0.15	
241	4-Aryl-2,4-Dioxobutanoic Derivate 16	Non Disponibile	0.14	
242	4-Aryl-2,4-Dioxobutanoic Derivate 17	Non Disponibile	0.1	
243	4-Aryl-2,4-Dioxobutanoic Derivate 18	Non Disponibile	0.1	
244	4-Aryl-2,4-Dioxobutanoic Derivate 19	Non Disponibile	0.1	
245	5-CITEP	168.1	690	16
246	6-Aryl-Dioxoesenoic Derivate 8a	7.9	7	17
247	6-Aryl-Dioxoesenoic Derivate 8b	50	65	
248	6-Aryl-Dioxoesenoic Derivate 8c	61	72	
249	6-Aryl-Dioxoesenoic Derivate 8d	56	67	
250	6-Aryl-Dioxoesenoic Derivate 8e	22	41	
251	6-Aryl-Dioxoesenoic Derivate 8f	76	92	
252	6-Aryl-Dioxoesenoic Derivate 11a	8.9	7.5	
253	6-Aryl-Dioxoesenoic Derivate 11b	85	90	
254	6-Aryl-Dioxoesenoic Derivate 11c	87	95	
255	6-Aryl-Dioxoesenoic Derivate 11d	73	88	
256	6-Aryl-Dioxoesenoic Derivate 11e	38	50	
257	6-Aryl-Dioxoesenoic Derivate 11f	57	45	
258	Bis-2,6-benzylidene 14i	0.5	0.9	18
259	Bis-2,6-benzylidene 14l	0.5	1.1	
260	Bis-2,6-benzylidene 14o	1.7	2.2	
261	Bis-2,6-benzylidene 15a	0.3	0.5	
262	Bis-2,6-benzylidene 15b	0.4	0.6	
263	Bis-2,6-benzylidene 15c	0.5	0.4	
264	Bis-2,6-benzylidene 15d	0.7	1.1	
265	Bis-2,6-benzylidene 15e	1.4	2.3	
266	Bis-2,6-benzylidene 15f	6	9	
267	Bis-2,6-benzylidene 15g	0.7	0.5	
268	Bis-2,6-benzylidene 16	1.6	0.9	
269	Bis-2,6-benzylidene 17a	0.3	0.7	
270	Bis-2,6-benzylidene 17b	0.2	0.5	
271	Bis-2,6-benzylidene 18a	3	4	
272	Bis-2,6-benzylidene 18b	1	1.6	
273	Bis-2,6-benzylidene 19a	0.2	0.4	
274	Bis-2,6-benzylidene 19b	1.3	1	
275	Bis-2,6-benzylidene 19c	0.2	0.3	
276	Bis-2,6-benzylidene 19d	0.2	0.3	
277	Bis-2,6-benzylidene 20a	2.8	4.3	

278	Bis-2,6-benzylidene 20b	0.7	1.2	19
279	Bis-2,6-benzylidene 20c	2.6	1.9	
280	Bis-2,6-benzylidene 20d	0.2	0.2	
281	Bis-cathecol 1	1.3	Non Disponibile	
282	Bis-cathecol 2	2.1	Non Disponibile	
283	Bis-cathecol 3	23	Non Disponibile	
284	Bis-cathecol 4	13	Non Disponibile	
285	Bis-cathecol 5	76	Non Disponibile	
286	Bis-cathecol 6	500	Non Disponibile	
287	Bis-cathecol 7	84	Non Disponibile	
288	Bis-cathecol 8	157	Non Disponibile	
289	Bis-cathecol 9	500	Non Disponibile	
290	Bis-cathecol 10	500	Non Disponibile	
291	Bis-cathecol 11	138	Non Disponibile	
292	Bis-cathecol 12	500	Non Disponibile	
293	Bis-cathecol 13	22	Non Disponibile	
294	Bis-cathecol 14	21	Non Disponibile	
295	Caffeoylpyrrolidine Derivate 1	39.2	Non Disponibile	20
296	Caffeoylpyrrolidine Derivate 2	25.2	Non Disponibile	
297	Caffeoylpyrrolidine Derivate 3	27.9	Non Disponibile	
298	Caffeoylpyrrolidine Derivate 4	33.8	Non Disponibile	
299	Caffeoyltetrahydrofuran Derivate 1	32.3	Non Disponibile	
300	Caffeoyltetrahydrofuran Derivate 2	47.4	Non Disponibile	
301	Caffeoyltetrahydrofuran Derivate 3	22.4	Non Disponibile	
302	Caffeoylnaphtalene Sulfonamide 3	11.9	Non Disponibile	21
303	Caffeoylnaphtalene Sulfonamide 4	24.6	Non Disponibile	
304	Caffeoylnaphtalene Sulfonamide 5	12.2	Non Disponibile	
305	Caffeoylnaphtalene Sulfonamide 6	8.6	Non Disponibile	
306	Caffeoylnaphtalene Sulfonamide 7	4.5	Non Disponibile	
307	Caffeoylnaphtalene Sulfonamide 8	7.9	Non Disponibile	
308	Caffeoylnaphtalene Sulfonamide 9	20.4	Non Disponibile	
309	Caffeoylnaphtalene Sulfonamide 10	19.5	Non Disponibile	
310	Caffeoylnaphtalene Sulfonamide 11	19.3	Non Disponibile	
311	Caffeoylnaphtalene Sulfonamide 12	20.4	Non Disponibile	
312	Caffeoylnaphtalene Sulfonamide 13	19.8	Non Disponibile	
313	Calcomine Derivate 2	5.8	7.3	22
314	Calcomine Derivate 3	2.7	3.2	
315	Calcomine Derivate 4	0.4	0.7	
316	Calcomine Derivate 5	0.5	0.7	
317	Calcomine Derivate 6	0.2	0.33	
318	Calcomine Derivate 7	1	1	
319	Calcomine Derivate 8	0.5	0.5	
320	Calcomine Derivate 9	0.5	1	
321	Calcomine Derivate 10	0.3	0.7	
322	Calcomine Derivate 11	2	4	
323	Calcomine Derivate 12	2.7	1.3	
324	Calcomine Derivate 13	4	0.6	
325	Calcomine Derivate 14	3	1	
326	Calcomine Derivate 15	5	2.3	
327	Calcomine Derivate 19	1	2.5	
328	Calcomine Derivate 20	0.8	1	
329	Calcomine Orange	0.35	0.35	
330	Carbonyl J	4	4	
331	Cathecol Maleic Anhydride 7	3.8	Non Disponibile	23
332	Cathecol Maleic Imide 6	15.5	Non Disponibile	
333	Cathecol Pyrrolidine 4	5.3	Non Disponibile	
334	Cathecol Succinic Derivate 3	23.6	Non Disponibile	

335	Cathecol Succinic Derivate 4	19.1	Non Disponibile	
336	Compound 4a	0.33	Non Disponibile	
337	Compound 4b	0.37	Non Disponibile	
338	Compound 4c	2.23	Non Disponibile	
339	Compound 4d	0.83	Non Disponibile	
340	Compound 9	1.97	Non Disponibile	
341	Compound 14	0.97	Non Disponibile	
342	Compound 19	10	Non Disponibile	24
343	Compound 22a	0.83	Non Disponibile	
344	Compound 23a	0.23	Non Disponibile	
345	Compound 27a	10	Non Disponibile	
346	Compound 27b	10	Non Disponibile	
347	Compound 27c	10	Non Disponibile	
348	Compound 29	10	Non Disponibile	
349	Compound V98	206	Non Disponibile	
350	Compound V99	74	Non Disponibile	
351	Compound V100	22.4	782	
352	Compound V101	250	250	
353	Compound V102	208	Non Disponibile	25
354	Compound V103	143.7	Non Disponibile	
355	Compound V104	0.9	201.3	
356	Compound V165	0.9	16.1	
357	Compound V545	20.5	3.5	
358	Galloylpyrrolidine 1	12.4	Non Disponibile	
359	Galloylpyrrolidine 2	15.6	Non Disponibile	20
360	Galloyltetrahydrofuran 1	7.4	Non Disponibile	
361	Galloyltetrahydrofuran 2	4.7	Non Disponibile	
362	Integracide 1	5	9	
363	Integracide 3	5	15	
364	Integracide 4	10	14	
365	Integracide 5	2	5.6	26
366	Integracide 24	5	12	
367	Integracide 25	Non Disponibile	12	
368	L-708,906	20.5	3.5	16
369	Lithospermic Acid	Non Disponibile	1.4	
370	Luteolin	Non Disponibile	11	27
371	NSC9169	200	200	
372	NSC44181	23	13	
373	NSC97318	13	14	
374	NSC103668	88	105	
375	NSC114690	60	69	
376	NSC127625	200	200	
377	NSC128439	32	39	
378	NSC134135	16	42	
379	NSC134142	39	50	
380	NSC134144	21	44	
381	NSC134173	60	59	28
382	NSC134191	55	65	
383	NSC134215	23	21	
384	NSC138557	30	30	
385	NSC156987	200	200	
386	NSC224283	200	200	
387	NSC348893	149	140	
388	NSC382139	200	200	
389	NSC621109	32	28	
390	NSC624906	200	200	

391	NSC642629	200	200	16
392	NSC686564	200	200	
393	P3 Compound	60	533	
394	P4 Compound	392	784	
395	P12 Compound	21.7	365	
396	P13 Compound	6.5	15.1	29
397	Quinoline 1	2.4	Non Disponible	
398	Quinoline 11	0.9	Non Disponible	
399	Quinoline 12	6.5	Non Disponible	
400	Quinoline 14	5	Non Disponible	
401	Quinoline 15	5	Non Disponible	
402	Quinoline 16	5	Non Disponible	
403	Quinoline 18	1.5	Non Disponible	
404	Quinoline 19	6.5	Non Disponible	
405	Quinoline 26	2.3	Non Disponible	
406	Quinoline 27	0.7	Non Disponible	30
407	Spiroindane 1	17	5	
408	Spiroindane 2	243	243	
409	Spiroindane 3	200	200	
410	Spiroindane 4	200	200	
411	Spiroindane 5	200	200	
412	Spiroindane 6	200	200	
413	Spiroindane 7	35	6	
414	Spiroindane 8	275	231	
415	Spiroindane 9	11	10	
416	Spiroindane 10	6.28	3.12	
417	Spiroindane 11	3.14	1.78	
418	Spiroindane 12	47	53	
419	Spiroindane 13	44	25	
420	Spiroindane 14	200	200	
421	Spiroindane 15	420	350	
422	Spiroindane 16	2.8	2	
423	Spiroindane 17	32	12	
424	Spiroindane 18	1.4	0.35	
425	Spiroindane 19	200	200	31
426	Styrylquinoline Derivate 9	100	100	
427	Styrylquinoline Derivate 10	5.3	2.1	
428	Styrylquinoline Derivate 11	1.9	5.1	
429	Styrylquinoline Derivate 12	3.4	3	
430	Styrylquinoline Derivate 13	4.1	11	
431	Styrylquinoline Derivate 14	1.2	1.7	
432	Styrylquinoline Derivate 15	3.5	2.2	
433	Styrylquinoline Derivate 16	1.4	1.2	
434	Styrylquinoline Derivate 17	1.6	1.6	
435	Styrylquinoline Derivate 18	3.2	3.2	
436	Styrylquinoline Derivate 19	3.7	2.8	
437	Styrylquinoline Derivate 20	2.4	1	
438	Styrylquinoline Derivate 21	2.8	3.7	
439	Styrylquinoline Derivate 22	0.9	3.3	
440	Styrylquinoline Derivate 23	0.3	0.4	
441	Styrylquinoline Derivate 24	0.7	1.7	
442	Styrylquinoline Derivate 25	4.9	4.5	
443	Styrylquinoline Derivate 26	1.3	1.2	
444	Styrylquinoline Derivate 27	4	4.9	
445	Styrylquinoline Derivate 28	100	100	
446	Styrylquinoline Derivate 29	100	100	
447	Styrylquinoline Derivate 30	2.3	1.5	32
448	Thalassiolin A	Non Disponible	0.4	

449	Thalassiolin B	Non Disponibile	43	
450	Thalassiolin C	Non Disponibile	28	

Bibliografia all'Appendice III

- 1 Neamati, N.; Sunder, S.; Pommier, Y. *Drug Discov. Today*, 1997, 2, 487.
- 2 McDougall, B.; King, P.J.; Wu, B. W.; Hostomsky, Z.; Reinecke, M. G.; Robinson, W. E. *Antimicrob. Agents Ch.*, 1998, 42, 140.
- 3 King, P. J.; Ma, G.; Miao, W.; Jia, Q.; McDougall, B. R.; Reinecke, M.; Cornell, C.; Kuan, J.; Kim, T.; Robinson, E. J. *Med. Chem.*, 1999, 42, 497
- 4 Artico, M.; Di Santo, R.; Costi, R.; Novellino, E.; Greco, G.; Massa, S.; Tramontano, E.; Marongiu, M.E.; De Montis, A.; La Colla, P. *J. Med. Chem.*, 1998, 41, 3948
- 5 Zhao, H.; Neamati, N.; Mazumder, A.; Sunder, S.; Pommier, Y.; Burke, T. R. *J. Med. Chem.*, 1997, 40, 1186.
- 6 Lin, Z.; Neamati, N.; Zhao, H.; Kiryu, Y.; Turpin, J. A.; Aberham, C.; Strelbel, K.; Kohn, K.; Witvrouw, M.; Pannecouque, C.; Debyser, Z.; De Clercq, E.; Rice, W. G.; Pommier, Y.; Burke, T. R. *J. Med. Chem.*, 1999, 42, 1401.
- 7 Zhao, H.; Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Milne, G. W. A.; Pommier, Y.; Burke, T. R. *J. Med. Chem.*, 1997, 40, 242.
- 8 Mazumder, A.; Neamati, N.; Sunder, S.; Schulz, J.; Pertz, H.; Eich, E.; Pommier, Y. *J. Med. Chem.*, 1997, 40, 3057.
- 9 Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S.; Milne, G. W. A.; Pommier, Y.; Burke, T.R. *J. Med. Chem.*, 1997, 40, 937.
- 10 Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M.; Mazumder, A.; Zhao, H.; Burke Jr., T.; Pommier, Y.; Milne, G. J. *Med. Chem.*, 1997, 40, 930.
- 11 Neamati, N.; Hong, H.; Owen, J. M.; Sunder, S.; Winslow, H. E.; Christensen, J. L.; Zhao, H.; Burke, T.; Milne, G. W. A.; Pommier, Y. *J. Med. Chem.*, 1998, 41, 3202.
- 12 Mekouar, K.; Mouscardet, J. -F. *J. Med. Chem.*, 1998, 41, 2846.
- 13 Pais, G. C. G.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E.S.; Pathak, V.K.; Tang, Y.; Micklaus, M.; Pommier, Y.; Burke, T.R. *J. Med. Chem.*, 2002, 45, 3184.
- 14 Zhang, X.; Pais, G. C. G.; Svarovskaia, E. S.; Marchand, C.; Johnson, A. A.; Karki, R. G.; Nicklaus, M. C.; Pathak, V. K.; Pommier, Y.; Burke, T. R. *Bioorg. Med. Chem. Lett.*, 2003, 13, 1215.
- 15 Wai, J.; Egbertson, M.; Payne, L.; Fisher, T.; Embrey, M.; Tran, L.; Melamed, J.; Langford, H.; Guare Jr., J.; Zhuang, L.; Grey, V.; Vacca, J.; Holloway, M.; Naylor-Olsen, A.; Hazuda, D.; Felock, P.; Wolfe, A.; Stillmock, K.; Schleif, W.; Gabryelski, L.; Young, S. *J. Med. Chem.*, 2000, 43, 4923.
- 16 Pluymers, W.; Pais, G.; Van Maele, B.; Pannecouque, C.; Fikkert, V.; Burke, T. R.; De Clercq, E.; Witvrouw, M.; Neamati, N.; Debyser, Z. *Antimicrob. Agents Ch.*, 2002, 46, 3292.
- 17 Costi, R.; Di Santo, R.; Artico, M.; Roux, A.; Ragno, R.; Massa, S.; Tramontano, E.; La Colla, M.; Loddo, R.; Marongiu, E.; Pani, A.; La Colla, P. *Bioorg. Med. Chem. Lett.*, 2004, 14, 1745.
- 18 Costi, R.; Di Santo, R.; Artico, M.; Massa, S.; Ragno, R.; Loddo, R.; La Colla, M.; Tramontano, E., La Colla, P.; Pani, A. *Bioorg. Med. Chem.*, 2004, 12, 199.
- 19 Dupont, R.; Jeansdon, L.; Mouscardet, J. -F.; Cotelle, P. *Bioorg. Med. Chem. Lett.*, 2001, 11, 3175.
- 20 Hwang, D. J.; Kim, S. N.; Choi, J. H.; Lee, Y. S. *Bioorg. Med. Chem.*, 2001, 9, 1429
- 21 Xu, Y. W.; Zhao, G. S.; Shin, C. G.; Zang, H. C.; Lee, C. K.; Lee, Y. S. *Bioorg. Med. Chem.*, 2003, 11, 3589.
- 22 Maurer, K.; Tang, A. H.; Kenyon, G. L.; Leavitt, A. D. *Bioorg. Chemistry*, 2000, 28, 140.
- 23 Lee, J. Y.; Yoon, K. J.; Lee, Y. S. *Bioorg. Med. Chem. Lett.*, 2003, 13, 4331.
- 24 Reinke, R. A.; King, P. J.; Victoria, J. G.; McDougall, B. R.; Ma, G.; Mao, Y.; Reinecke, M. G.; Robinson, E. *J. Med. Chem.*, 2002, 45, 3669.
- 25 Pannecouque, C.; Pluymers, W.; Van Maele, B.; Tetz, V.; Cherepanov, P.; De Clercq, E.; Witvrouw, M.; Debyser, Z. *Curr. Biol.*, 2002, 12, 1169.
- 26 Singh, S. B.; Ondeyka, J. G.; Schleif, W. A.; Felock, P.; Hazuda, D. *J. Nat. Prod.*, 66, 1338.
- 27 Tewtrakul, S.; Miyashiro, H.; Nakamura, N.; Hattori, M.; Kawahata, T.; Otake, T.; Yoshinaga, T.; Fujiwara, T.; Supavita, T.; Yuenyongsawad, S.; Rattanasuwon, P.; Dej-Adisai, S. *Phytother. Res.*, 2003, 27, 232.
- 28 Chen, I.; Neamati, N.; Nicklaus, M.; Orr, A.; Anderson, L.; Barchi Jr., J.; Kelley, J.; Pommier, Y.; MacKerell Jr A. *Bioorg. Med. Chem.*, 2000, 8, 2385.
- 29 Benard, C.; Zouhirim, F.; Normand, Bayle, M.; Danet, M.; Desmaele, D.; Leh, H.; Mouscardet, J. -F.; Mbemba, G.; Thomas, C. -M.; Bonenfant, S.; Le Bret, M.; D'Angelo, J. *Bioorg. Med. Chem. Lett.*, 2004, 14, 2473.
- 30 Molteni, V.; Rhodes, D.; Rubins, K.; Hansen, M.; Bushman, F. D.; Siegel, J.S. *J. Med. Chem.*, 2000, 43, 2031.
- 31 Zouhiri, F.; Mouscardet, J. -F. *J. Med. Chem.*, 2000, 43, 1533.
- 32 Rowley, D. C.; Hansen, M. S. T.; Rhodes, D.; Sotrier, C. A.; Ni, H.; McCammon, A.; Bushman, F.D.; Fenical, W. *Bioorg. Med. Chem.*, 2002, 10, 3619.

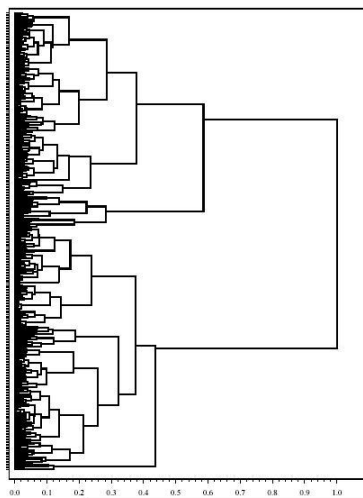
Appendice IV – Cluster Analysis

Algoritmo di clustering utilizzato: **Ward's Reciprocal Nearest Neighbors**

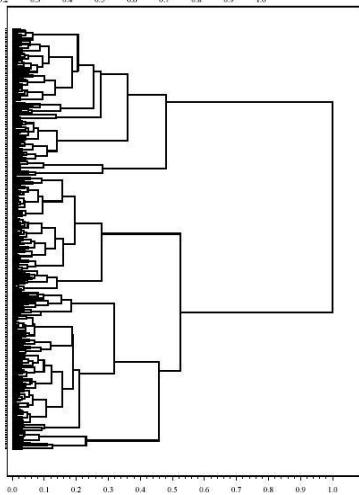
Distanza utilizzata: Distanza Euclidea

Standardizzazione dei dati: Media/Deviazione standard

3'-Processing



Strand Transfer



3'-Processing:

Distanza Massima di Amalgamazione: 60,946

Treshold per la selezione dei set:

30 strutture: 0.124
35 strutture: 0.118
40 strutture: 0.108
50 strutture: 0.08

Strand Transfer

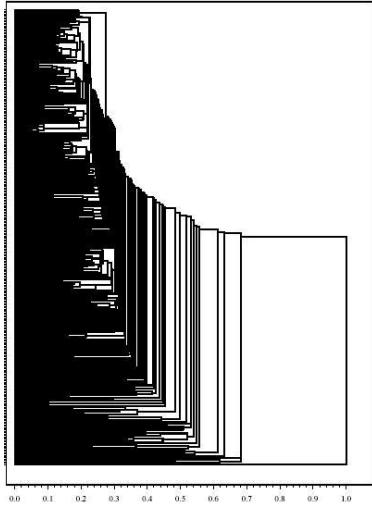
Distanza Massima di Amalgamazione: 55,659

Treshold per la selezione dei set:

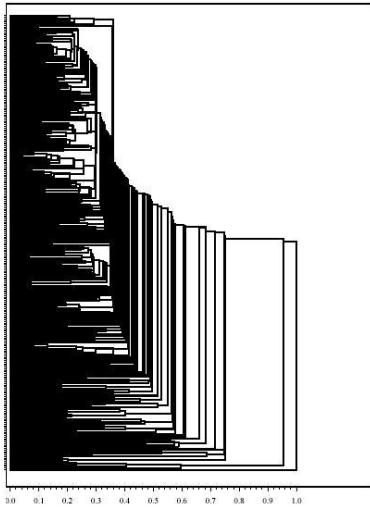
30 strutture: 0.126
35 strutture: 0.109
40 strutture: 0.10
50 strutture: 0.09

Algoritmo di clustering utilizzato: **Single Linkage**
Distanza utilizzata: Distanza Euclidea
Standardizzazione dei dati: Media/Deviazione standard

3'-Processing



Strand Transfer



3'-Processing:

Distanza Massima di Amalgamazione: 9,6692

Threshold per la selezione dei set:

30 strutture: 0,397

35 strutture: 0,417

40 strutture: 0,44
50 strutture: 0,37

Strand Transfer

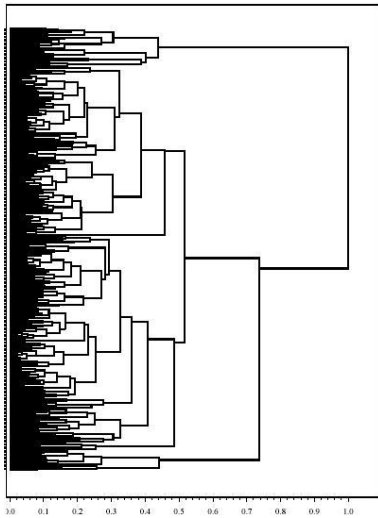
Distanza Massima di Amalgamazione: 8,3261

Treshold per la selezione dei set:

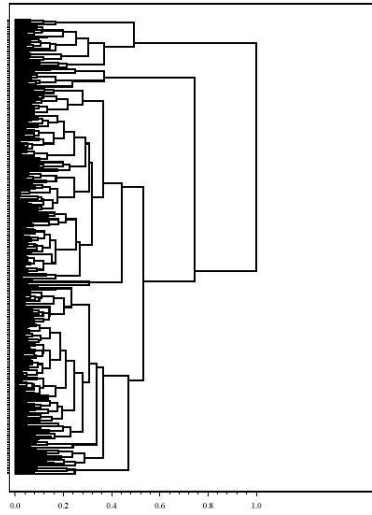
30 strutture: 0,445
35 strutture: 0,47
40 strutture: 0,486
50 strutture: 0,414

Algoritmo di clustering utilizzato: **Complete Linkage**
Distanza utilizzata: Distanza Euclidea
Standardizzazione dei dati: Media/Deviazione standard

3'-Processing



Strand Transfer



3'-Processing:

Distanza Massima di Amalgamazione: 9,6692

Threshold per la selezione dei set:

30 strutture: 0,397

35 strutture: 0,417

40 strutture: 0,44

50 strutture: 0,37

Strand Transfer

Distanza Massima di Amalgamazione: 30,246

Threshold per la selezione dei set:

30 strutture: 0,244

35 strutture: 0,228

40 strutture: 0,21

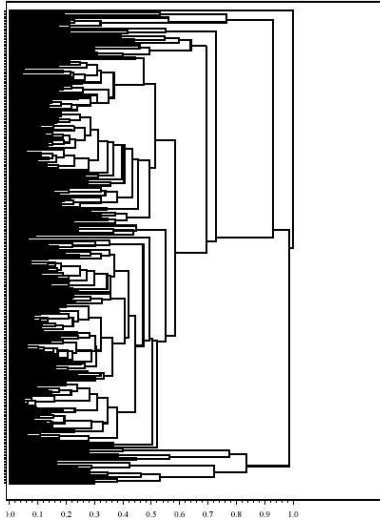
50 strutture: 0,189

Algoritmo di clustering utilizzato: **Group Average (Weighted Mean)**

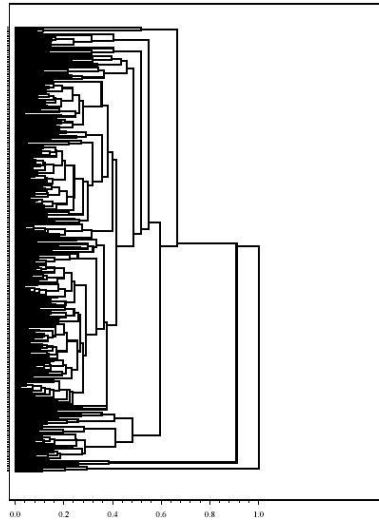
Distanza utilizzata: Distanza Euclidea

Standardizzazione dei dati: Media/Deviazione standard

3'-Processing



Strand Transfer



3'-Processing:

Distanza Massima di Amalgamazione: 12,803

Threshold per la selezione dei set:

30 strutture: 0,456

35 strutture: 0,442

40 strutture: 0,420

50 strutture: 0,382

Strand Transfer

Distanza Massima di Amalgamazione: 16,943

Threshold per la selezione dei set:

30 strutture: 0,338

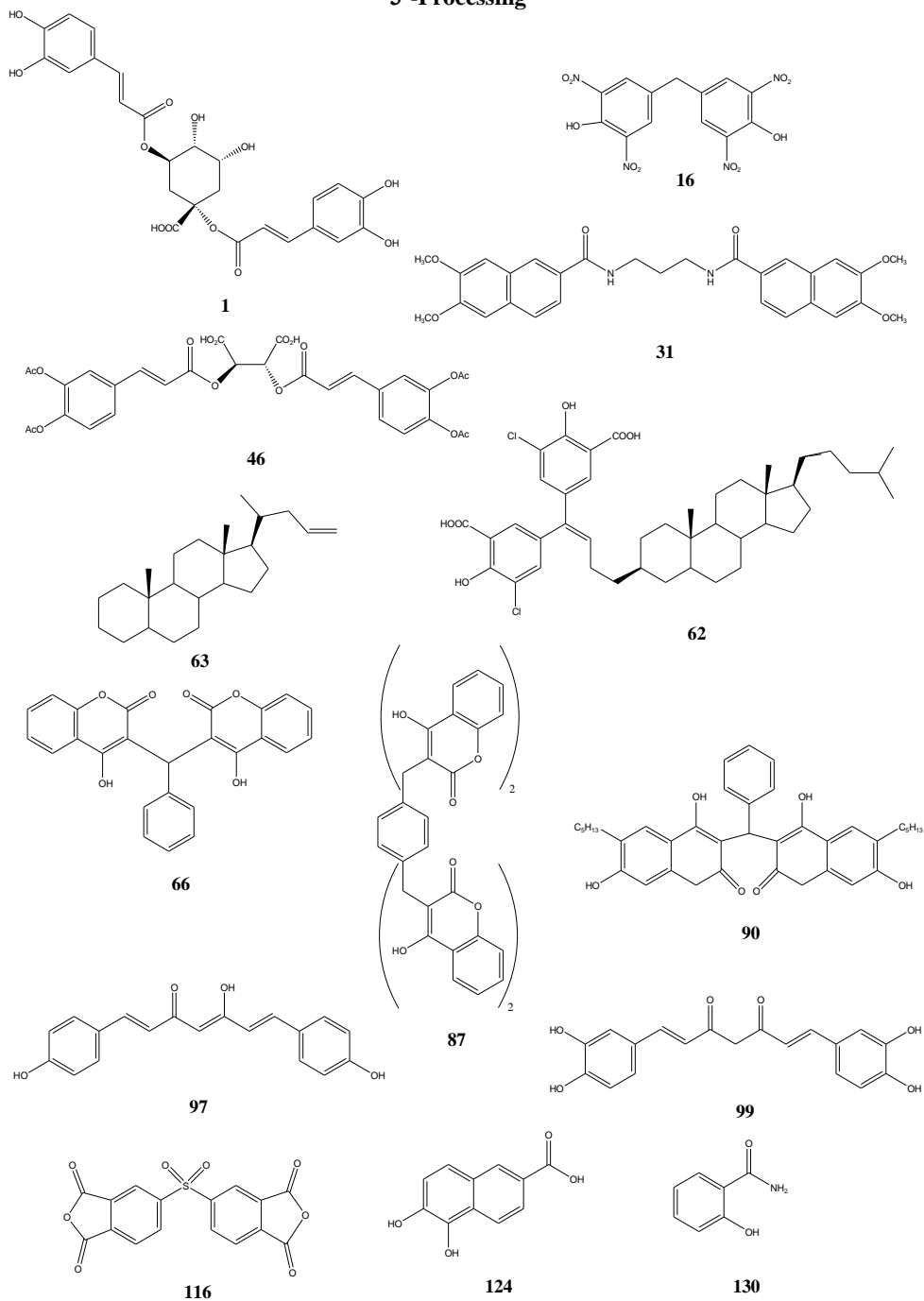
35 strutture: 0,316

40 strutture: 0,296

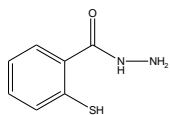
50 strutture: 0,276

Appendice V – Derivati inclusi nell'analisi QSAR

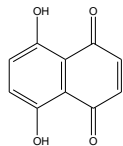
3'-Processing



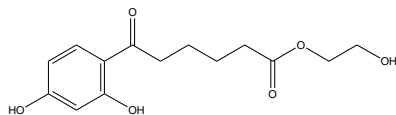
APPENDICE V - Continua



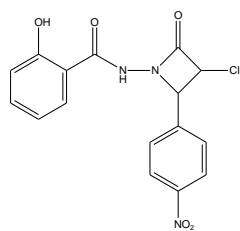
131



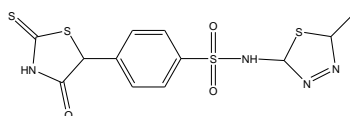
143



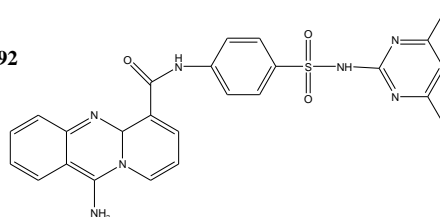
149



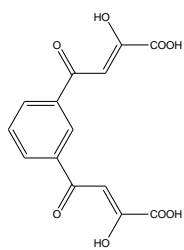
164



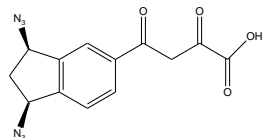
192



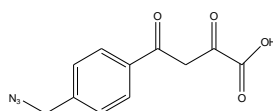
194



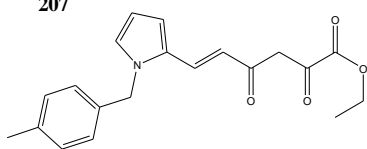
207



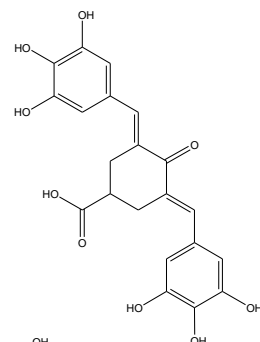
218



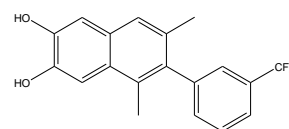
220



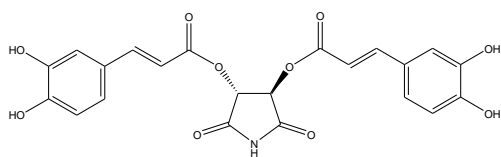
255



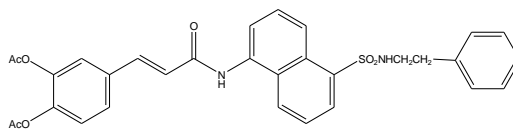
276



284

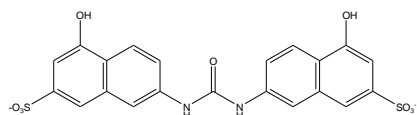


296

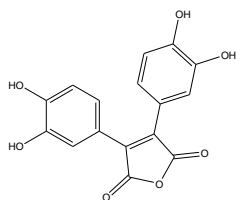


302

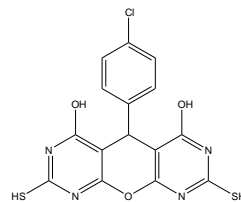
APPENDICE V - Continua



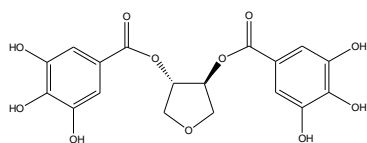
330



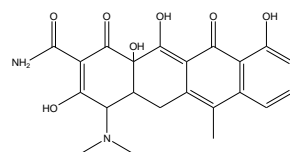
331



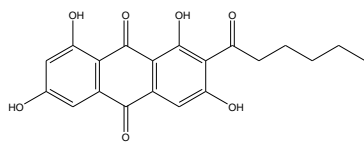
355



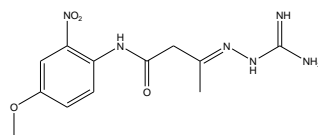
361



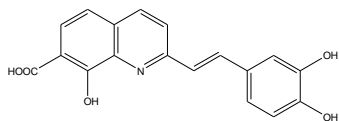
372



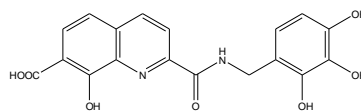
384



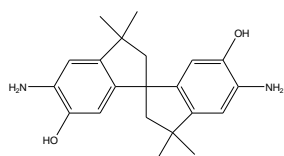
391



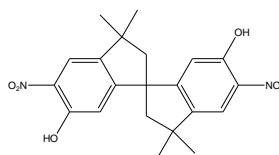
397



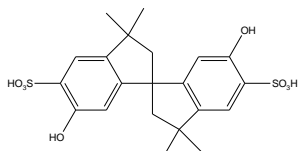
403



408



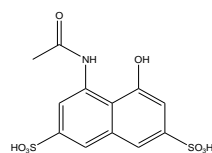
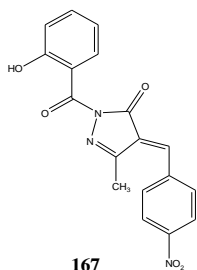
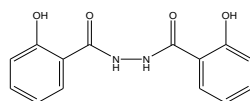
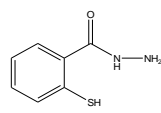
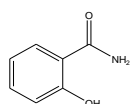
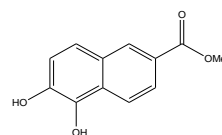
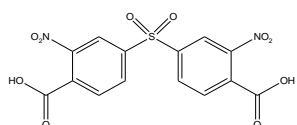
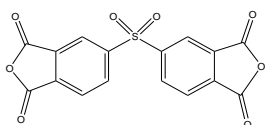
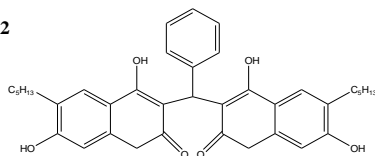
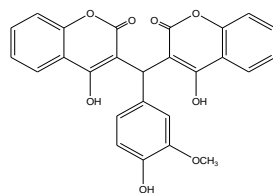
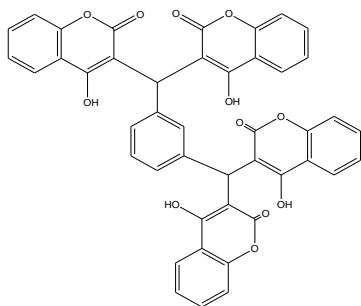
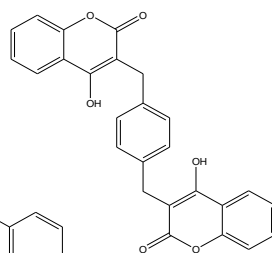
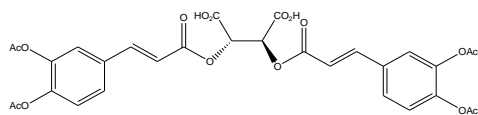
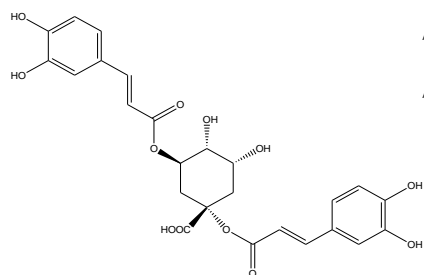
411



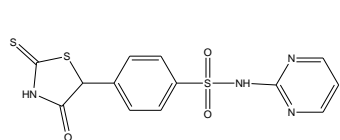
412

APPENDICE V - Continua

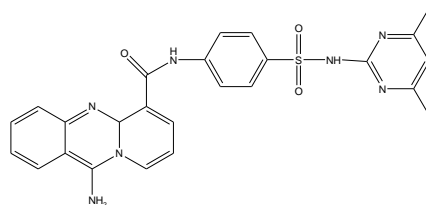
Strand Transfer



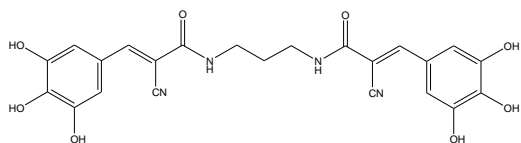
APPENDICE V - Continua



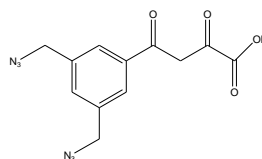
191



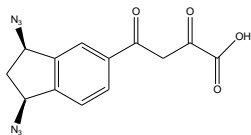
194



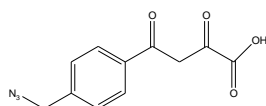
198



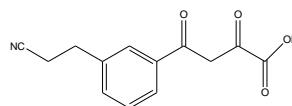
215



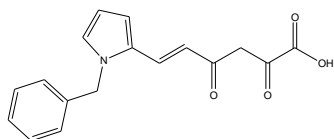
218



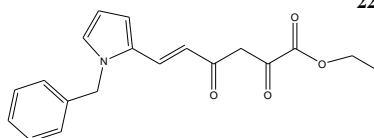
220



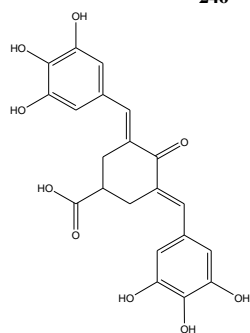
224



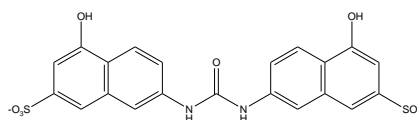
246



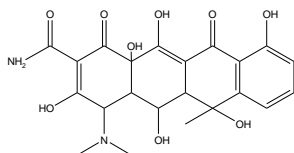
252



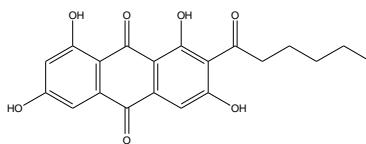
276



330

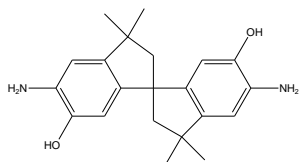


371

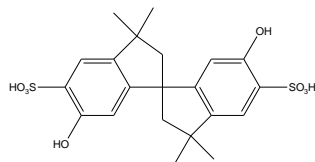


384

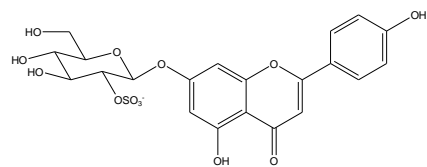
APPENDICE V - Continua



408



412



450