



UNIVERSITÀ DEGLI STUDI DI ROMA  
“La Sapienza”

Facoltà di Farmacia  
Corso di Laurea in Chimica e Tecnologie Farmaceutiche

Tesi Sperimentale di Laurea in Chimica Farmaceutica e Tossicologica

**Inibitori dell'enzima Istone Deacilasi (HDAC1):  
Relazioni Quantitative Struttura-Attività di tipo Tridimensionale  
(3-D QSAR)**

Relatore:

Chiar.mo Dott. Rino Ragno

Laureanda:

Manuela Kossuth  
Matr: 06032224

Anno Accademico 2004-2005

Nel dominio delle scienze chimiche e discipline affini, l'utilizzazione delle potenzialità derivanti dall'informatica ha consentito lo sviluppo di nuove branche quali la Chemiometria e la Modellistica Molecolare (MM). Entrambe affrontano, seppure in modo sostanzialmente diverso, le problematiche di interscambio informativo tra le varie discipline del settore chimico.

In particolare la Chemiometria, considerata come una disciplina derivata dall'applicazione della statistica alla chimica, ha permesso di razionalizzare in modo innovativo problemi complessi della medicina e della biologia, la Modellistica Molecolare, invece, raggruppa una serie di tecniche, procedure e applicazioni, che spaziano dalla fisica teorica, all'informatica fino alla biologia molecolare, rivolte all'analisi strutturale delle molecole e delle proprietà chimiche, fisiche e biologiche.

Attualmente, i metodi chemiometrici, integrati da elementi di MM consentono di effettuare la QSAR.

Con questo termine vengono definiti i modelli quantitativi che mettono in relazione la variazione dell'attività biologica con la diversa struttura chimica di una serie di composti a comune meccanismo d'azione. Duplice obiettivo di uno studio QSAR è quello di poter predire l'attività biologica di una serie di nuove molecole non ancora valutate sperimentalmente e di aiutare alla comprensione dei meccanismi che sottendono tale risultato.

Nell'ambito della progettazione di nuovi farmaci, particolarmente utili sono le informazioni riguardanti il *binding-mode* farmaco-recettore. Questo tipo di informazioni si estrinsecano principalmente dalla valutazione di interazioni di tipo sterico ed elettrostatico.

Nella 3-D QSAR, evoluzione della QSAR classica di Hansch, queste interazioni vengono simulate immergendo le molecole in una griglia tridimensionale ai nodi della quale sono posizionati un atomo sonda o raggruppamenti atomici di varia natura chimica (*probes*). Le interazioni di non legame vengono quindi calcolate per ogni molecola della serie in oggetto e per tutti i nodi della griglia.

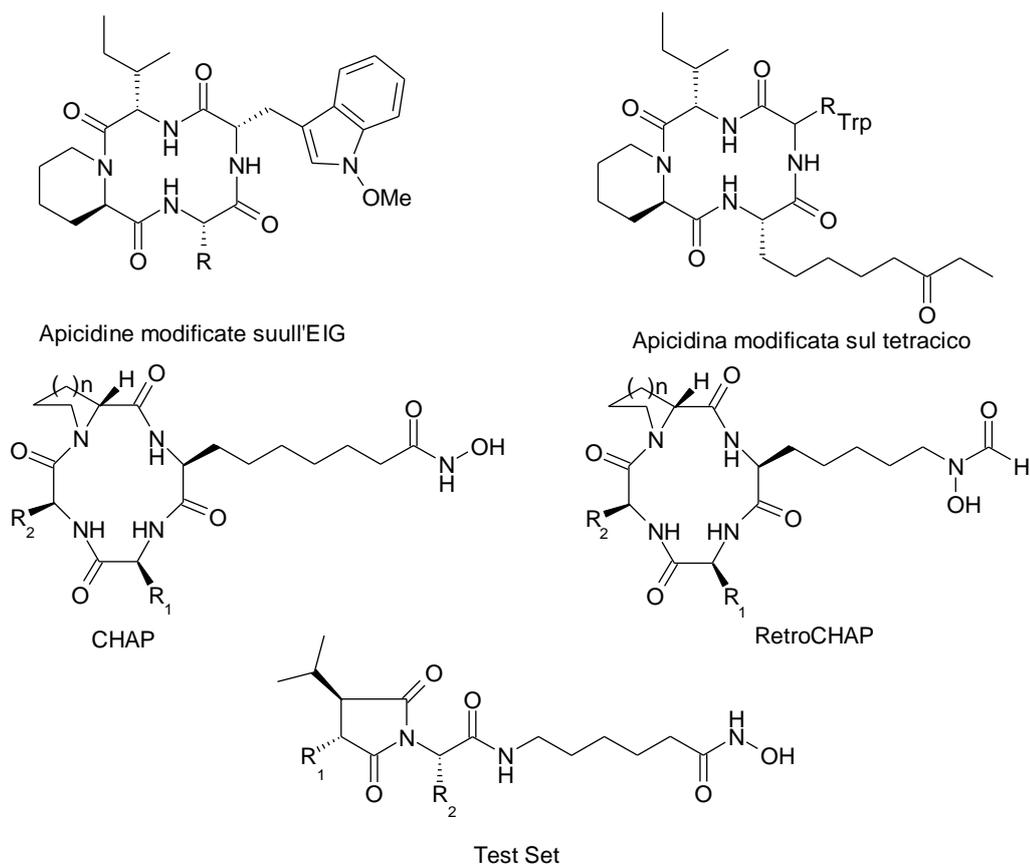
In questa tesi sperimentale, la 3-D QSAR è stata applicata a diverse classi di inibitori dell'istone deacetilasi (HDAC), importante enzima in grado di catalizzare la reazione di deacetilazione degli  $\epsilon$ -ammino gruppi terminali di lisina presenti nelle code istoniche del DNA. Tale trasformazione influenza l'accessibilità alla lettura del codice genetico, per questo sembra che l'HDAC sia implicato nello sviluppo di diverse forme tumorali. Negli ultimi anni, sulle più note riviste scientifiche, è stato riportato un numero sempre maggiore di derivati soprattutto sintetici in grado di inibire gli HDACs.

Utilizzando i dati strutturali degli inibitori noti è possibile elaborare un modello strutturale (farmacoforo comune). Questo farmacoforo è costituito da un dominio che lega il metallo detto EIG (*Enzyme Inhibiting Group*), un dominio "linker", che occupa il canale e un dominio di ricognizione superficiale, che interagisce con i residui all'estremità del sito attivo chiamato "Cap Group" (CAP). Fino ad oggi, comunque, pur essendo di notevoli dimensioni l'attenzione rivolta verso gli inibitori HDAC, nessuno studio di 3-D SAR o 3-D QSAR ad ampia applicazione è stato ancora pubblicato. Nell'ambito di un progetto in grado di fornire una relazione quantitativa tra struttura chimica-attività biologica, abbiamo deciso di realizzare una 3-D QSAR considerando solo gli inibitori con "Cap Group" estremamente grande. Il *Training Set* scelto per questo studio è composto da 85 molecole, divise in quattro gruppi di inibitori ottenuti da diverse pubblicazioni: Apicidine modificate sull'EIG, Apicidine modificate sulla porzione tetraciclica, i tetrapeptidi ciclici contenenti un gruppo idrossammico (CHAP) e i tetrapeptidi ciclici contenenti un gruppo retroidrossammico (RetroCHAP), in ultimo dall'unione dei quattro gruppi precedenti è stato generato un gruppo finale, chiamato CAP-GROUP.

Le molecole sono state modellate usando il programma Sybyl e ottimizzate con un semplice protocollo di dinamica molecolare presente nel programma Macromodel. Successivamente, avendo a disposizione le coordinate dell'istone deacetilasi, è stato realizzato tramite *docking* molecolare, mediante il programma Autodock 3.0, un allineamento *Best Cluster* e un allineamento *Best Docked*. In base al diverso allineamento sono stati sviluppati studi in parallelo per i quattro gruppi e per il gruppo unito, ottenendo in totale dieci modelli.

Il programma GRID, sviluppato da Goodford, è stato utilizzato per calcolare i campi d'interazione molecolare (MIF) tra le molecole e gli atomi sonda presenti a livello dei nodi della griglia, le cui dimensioni sono state impostate per includere tutte le molecole ( $X=24 \text{ \AA}$ ,  $Y=13 \text{ \AA}$ ,  $Z=16 \text{ \AA}$ ). Per la determinazione dell'atomo sonda è stata effettuata un'iniziale analisi *multi-probes* e conseguentemente ai risultati è stata scelta la *probe* OH<sub>2</sub> (*water probe*) in quanto è risultato essere quello che meglio descriveva la variabilità strutturale dei vari *training set*. Per le impostazioni scelte, GRID ha permesso di calcolare 28350 variabili iniziali per ogni gruppo.

La determinazione di un modello matematico di correlazione struttura-attività che tenga conto di un numero così grande d'informazioni richiede l'utilizzazione di un *software* che sia in grado di individuare i dati veramente significativi evitando di considerare quelli meno importanti. Per la trattazione statistica e per lo sviluppo dei dieci modelli 3-D QSAR è stato utilizzato un metodo PLS implementato nel programma GOLPE. Mediante questo programma è possibile effettuare sia un pretrattamento dei dati, per eliminare sin da subito le variabili non importanti, che la selezione delle variabili più significative. Per la generazione dei modelli PLS finali nelle selezioni FFD è stato utilizzato l'algoritmo SRD (*Smart Region Definition*). Per scegliere il numero di componenti principali da estrarre, è stata eseguita una *cross-validation* (CV) usando il metodo *Leave Some Out* (LSO) a 5 gruppi. Tale approccio è stato applicato ad ogni analisi PLS, dalla preliminare (senza pretrattamento e selezione delle variabili) all'ultima (dopo il pretrattamento dei dati e quattro FFD). Con la CV si ottiene un valore chiamato  $q^2$  che riflette il grado di predizione interna del modello stesso. I valori del coefficiente di correlazione  $r^2$  e di *cross-validation*  $q^2$  devono essere quanto più possibili vicino ad uno che rappresenta il valore di massima correlazione per il primo e di massima predizione interna per il secondo. I dieci modelli 3-D QSAR sono stati generati ottenendo ottimi risultati come si può osservare dalla seguente tabella riassuntiva.



Per ognuno dei modelli 3-D QSAR sono state generate le mappe di correlazione struttura-attività e analizzate le loro corrispondenze con i derivati della serie molecolari. Di particolare riguardo è la validazione dei modelli che è stata effettuata con gruppi di molecole esterne ai *training set* e quindi non utilizzate nella formulazione dei modelli stessi. Attraverso le validazioni esterne i dieci modelli si sono dimostrati discretamente predittivi, se considerata anche la notevole differenza strutturale tra le molecole del *TestSet* e quelle del *Training Set*.