

Molecular Design

Saul Pierotti

March 25, 2019

Introduzione

- In questo corso noi faremo progettazione di molecole, non sintesi

Uso di modelli nelle scienze sperimentali

- I modelli sono impiegati in ogni settore scientifico, ma alcuni campi impiegano modelli hard ed altri soft
- Un esempio di modello hard è il modello ab initio, che permette di ottenere una proprietà del sistema in esame a partire da una sua configurazione
 - Uso modelli diversi per studiare proprietà diverse
- I risultati ottenuti tramite modelli hard sono approssimazioni della realtà
 - Le approssimazioni effettuate possono essere accettabili nella previsione di sistemi semplici (piccole molecole)
 - Questi modelli sono utili come previsione, o se non è possibile effettuare l'esperimento reale
 - Un modello hard usa dati solo calcolati
- Un modello soft è più accurato, ma necessita di poter ideare un esperimento adeguato in quanto necessita di dati sperimentali
- In questo corso utilizzeremo modelli che vanno dal soft al semi-hard
- La chemioinformatica, scienza che applica modelli informatici a sistemi molecolari, è stata fondata da Gasteiger, un chimico organico
 - Pur essendo un chimico organico ha sempre lavorato al pc e non ha mai fatto sintesi in laboratorio
 - Ha fondato la company Molecular Networks
 - Probabilmente verrà a fare una lezione a Perugia a fine aprile (!)

Cosa studieremo

- Come descrivere lo spazio biochimico
- I DBs utilizzati nella progettazione molecolare
- I metodi utilizzati per analizzare dati biochimici
- La relazione tra proprietà e struttura delle molecole
- Algoritmi sottostanti ai sistemi di previsione delle interazioni
- Intelligenza artificiale
 - E' nata con l'informatica
- Modelliamo le interazioni, non la struttura

Rappresentare un struttura molecolare

- Una cosa è compresa se esiste un suo modello
- Il nome dei composti deve essere in base alla struttura, non all'origine

Nomenclatura SMILES

- Trasformo la struttura in un grafo
 - Rimuovo gli idrogeni
 - Apro gli anelli ponendo un numero ad ogni rottura, che mi permette di identificare gli atomi separati
 - Il cicloesano può essere scritto C1CCCCC1
 - Se un atomo chiude 2 anelli gli si assegna 2 numeri consecutivi (es. C1CCCC2CCCC12)
 - Se voglio indicare più di 9 cicli premetto il simbolo % (l'atomo che chiude il ciclo 12 è C%12)
- Indico i legami in modo standard
 - Scrivo 2 atomi consecutivamente per un legame semplice (es. CC)
 - In alcuni casi è necessario esplicitare il legame con - (es. 2 cicli aromatici collegati tra loro)
 - = per doppi legami
 - # per triplo legame
 - \$ per quadruplo legame
 - . per un legame non esistente (es. [Na+].[Cl-])
 - : per un legame aromatico con parziale carattere di doppio legame
- I composti aromatici possono essere rappresentati in vari modi
 - Con i doppi legami alternati (Kekulé) C1=CC=CC=C1
 - Con il simbolo (:) C:1:C:C:C:C:1
 - Scrivendo i costituenti del ciclo in minuscolo c1ccccc1
- Le ramificazioni sono indicate con parentesi (es. acido acetico CC(=O)O)
- E' possibile indicare stereoisomeri
 - Per l'isomeria cis-trans indico con F/C=C/F (oppure F\C=C\F) l'isomero trans e F/C=C\F il cis (oppure F\C=C/F)
 - Gli stereoisomeri RS si indicano con @ se S e @@ se R (@ è una spirale antioraria!)
 - Il senso è quello rispetto al primo atomo elencato del centro chirale
 - L-Ala si indica N[C@@H](C)C(=O)O
- La codifica non è unica, una molecola può essere rappresentata in modi diversi
- Canonical SMILES è invece univoco
 - Dipende da un algoritmo di canonicalizzazione
 - * E' un problema complesso