

Molecular Design

Saul Pierotti

May 18, 2019

Cosa studieremo

- Progettazione di molecole, non sintesi
- Come descrivere lo spazio biochimico in termini di strutture e reazioni
- I DBs utilizzati nella progettazione molecolare
- I metodi utilizzati per analizzare dati biochimici, ossia la chemiometria
- La relazione tra proprietà e struttura delle molecole (QSPR)
- Algoritmi sottostanti ai sistemi di previsione delle interazioni

Note sparse

- In chemioinformatica sfruttiamo modelli per predire le interazioni tra molecole, non la loro struttura
- Una cosa si può definire compresa, conosciuta, se ne esiste un modello sufficientemente accurato
- Lo scopo della sintesi non è la produzione di composti, ma di proprietà
- I farmaci subiscono un iter di approvazione di più di 15 anni
- Nel settore farmaceutico guadagni e perdite sono enormi
- Poter tagliare fuori candidati problematici ad uno stadio precoce ha un potenziale enorme
- Goodford, collega del Prof, ha fondato la Wellcome Trust
- L'arginina nelle proteine è spesso usata come amminoacido idrofobico (!)
- Le simulazioni di docking hanno una precisione sull'ordine degli Angstrom
- Le drugs tendono ad avere simile LogD a pH 7.4, anche se hanno profili molto diversi
- Il pH dei fluidi biologici è attorno a 7.4

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 (!)
- In questo campo sono molto usati modelli basati sull'intelligenza artificiale
 - L'AI è una tecnica nata con l'informatica stessa

Descrivere sinteticamente una molecola

- Il nome di un composto è più utile che venga definito in base alla sua struttura, non alla sua origine
- Una nomenclatura efficiente migliora la produttività dei chimici
- La nomenclatura IUPAC permette una nomenclatura univoca e descrittiva dei composti chimici

Formato dei dati chimici, biologici e farmaceutici

- I dati sono sia input che output dei programmi di modelling
- Un dato chimico deve contenere la composizione chimica, i legami e la geometria molecolare
- Per piccole molecole il formato preferito è mol2
- Per le proteine ed altre macromolecole si utilizza il pdb
 - Il record è l'atomo, indicato in coordinate xyz
 - E' indicato anche l'aminoacido di appartenenza
 - Sono anche riportate le molecole d'acqua legata
 - La precisione è riportata tramite il B-factor
 - * La vibrazione termica causa incertezza nella misura
 - ATOM indica tutti gli atomi che partecipano alla struttura di aminoacidi
 - HETATM indica atomi di solvente, piccole molecole, ecc.
 - Gli H non sono mai presenti perché non visibili ai raggi X, ma possono essere inseriti su modelli virtuali
- I dati sono memorizzati in databases come il PDB e il CCDC (piccole molecole)
- Il Cambridge Crystallographic Data Centre (CCDC) è un DB a pagamento di strutture di piccole molecole
- Presenta molte più strutture di PDB
- Le banche dati pubbliche contengono sui ⁵ composti, quelle private farmaceutiche più di 10⁶
- Tutti i formati riportano delle informazioni essenziali, e altre non essenziali
 - Il tipo di atomo
 - Le coordinate atomiche xyz
 - Non è necessario inserire i legami, poiché sono dedotti dalle distanze atomiche
 - * In alcuni formati è comunque riportata una matrice di connettività
 - Spesso sono riportate informazioni sulla confidenza della posizione
 - * La confidenza è assoluta per strutture calcolate
 - Può essere riportata la densità di carica elettronica dei vari atomi
 - * Viene salvata per evitare di ricalcolarla
 - Può essere riportata la geometria molecolare, ossia lo stato di ibridazione dei vari atomi

Notazione SMILES

- E' una rappresentazione che consente di convertire una molecola in una stringa
- E' il formato più usato in banche dati chimiche e farmaceutiche
- 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 assegnano 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
 - Questo crea problemi nel mining dei databases e per la ricerca di sottostrutture
- Canonical SMILES è invece univoco
 - Dipende da un algoritmo di canonicalizzazione
 - * E' un problema complesso

Ottenere la struttura 3D

- La stringa SMILES viene convertita in rappresentazione 2D, che è poi usata per ottenere una 3D approssimata
- La struttura è migliorata con metodi di minimizzazione energetica o semiempirici
- I metodi semiempirici e di meccanica molecolare sono troppo lenti
 - Si parla di secondi, ma se i composti sono milioni è un problema
- Il software CONCORD riesce a convertire 1D in 3D in tempi ragionevoli
 - Spezza la molecola in frammenti a struttura nota, di cui ha un database interno
 - Ottiene una struttura 3D approssimata, che poi migliora con metodi di minimizzazione energetica
 - Si valuta come l'energia varia al variare della posizione, fino ad arrivare ad un minimo
 - Il minimo di energia di solito corrisponde alla struttura cristallografica

Metodi teorici e sperimentali

- I metodi teorici possono essere di varie tipologie
 - La predizione ab initio usa modelli hard basati esclusivamente su modelli quantistici, ossia risolve l'equazione di Schroedinger
 - * Non sempre è computazionalmente possibile
 - I metodi semiempirici richiedono alcuni parametri sperimentali, e risolvono la funzione d'onda solo per gli elettroni di valenza
 - * Sono più approssimati, ma più veloci
 - * Usano modelli di meccanica classica per gli altri elettroni
 - Metodi di meccanica molecolare ignorano gli effetti quantistici
- I metodi sperimentali permettono di ottenere strutture e conformazioni
 - Uno dei metodi sperimentali più usato è la cristallografia ai raggi X, specie per macromolecole
 - Per piccole molecole si usa più NMR
 - * E' usato per ottenere la struttura, più che la conformazione
 - * Oggi sono usati spettrometri NMR anche per le proteine, anche se non è sempre applicabile
 - * Nella spettrometria NMR non serve il cristallo (!)
 - Per misurare precisamente gli H si usa la cristallografia a diffrazione neutronica

Meccanica molecolare

- I sistemi che studiamo sono complessi, con molti elettroni e nuclei, e non possono essere predetti *ab initio*
- La meccanica molecolare (MM) è il metodo di predizione più veloce delle proprietà molecolari
- E' applicabile allo stato fondamentale ma non a quello eccitato
- Non permette di prevedere la distribuzione elettronica di una molecola
- Permette di prevedere le proprietà cinetiche e termodinamiche e l'energia di una conformazione
- Sfrutta l'approssimazione di Born-Oppenheimer, considera i nuclei fermi nelle transizioni elettroniche
 - Assume quindi la distanza tra nuclei costante
- Considera gli atomi come sfere legati da forze elastiche, con carica netta o parziale
- Descrive le interazioni come potenziali, e determina l'energia di una conformazione in base a questi
- L'insieme dei parametri e delle funzioni potenziali è definito Force-Field (FF)
- Le forze intra- ed inter-molecolari sono definite da 4 contributi nei FF
 - $E = E_{stretching} + E_{bending} + E_{torsion} + E_{non-bonding}$
 - Possono essere considerati anche contributi aggiuntivi
 - L'energia totale non ha significato assoluto, ma le differenze energetiche sono significative
 - Lo scopo di un FF di MM è la predizione della struttura di una molecola
- L'energia di stretching è modellata come elastica (legge di Hooke) attorno ad una lunghezza di equilibrio, tipica del legame
 - $E_{stretching} = \sum k_b(r - r_0)^2$
 - k_b modella la rigidità del legame
 - E' un'approssimazione, l'energia non ha un andamento parabolico ma di ordine superiore
- L'energia di bending considera la deformazione in modo elastico attorno all'angolo di equilibrio
 - In questo caso la costante è assegnata non ad un legame ma ad una tripletta di atomi
- L'energia torsionale è modellata da una funzione periodica
 - $E_{torsion} = \sum A[1 + \cos(n\tau - \phi)]$
 - Il parametro A controlla l'ampiezza della curvatura, n la periodicità
 - La parametrizzazione coinvolge quartetti atomici
- L'energia di non legame considera le interazioni di Lennard-Jones ed elettrostatiche
 - $E_{non-bonding} = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \frac{q_i q_j}{r_{ij}}$
 - E' il calcolo computazionalmente più impegnativo
 - La curva presenta una distanza optimum a cui l'energia è minima
 - A distanze inferiori si ha interazione repulsiva, a distanze superiori attrattiva
 - L'energia dell'interazione è 0 per distanze infinite, diviene negativa all'avvicinarsi dei nuclei, incontra un minimo e poi aumenta verso infinito per distanza tendente a 0
 - Il parametro A indica la polarizzabilità dell'atomo, B la durezza del guscio atomico
 - B viene determinato per cristallografia
 - Le cariche della componente elettrostatica sono pre-assegnate o calcolate
 - La costante dielettrica è un termine preponderante nell'interazione coulombica, in H_2O l'interazione è 80 volte inferiore che nel vuoto
- Altri possibili contributi all'energia totale sono i legami idrogeno e l'effetto del solvente
- Nei FF comunemente usati sono necessari da 2000 a 50000 parametri misurati, per modellare 60 tipi di atomi unici

Cristallografia

- Per la cristallografia è importante avere un campione proteico puro
 - Un cristallo è un array periodico di molecole
 - I cristalli proteici contengono canali e buchi pieni di solvente
 - E' molto difficile ottenere il cristallo
 - Una volta era un processo artigianale, ora è automatizzato
 - * Si modulano sali, acqua, metalli pesanti, tensioattivi
- Il cristallo viene colpito da raggi X, e l'analisi del pattern di diffrazione prodotto permette di generare

- una mappa di densità elettronica
 - Dal reticolo di diffrazione non ottengo la posizione ma la densità elettronica, non discerno bene atomi da gruppi di atomi
 - Fare il cristallo serve ad amplificare il segnale (!)
 - * Più del 99% del raggio incidente non viene deviato
 - Nella zona centrale dell'immagine ho il fascio diretto, mentre attorno il pattern di diffrazione
- La sequenza della proteina viene adattata alla mappa di densità elettronica ruotando attorno ai vari legami, in un processo detto fitting
 - Più la mappa ha risoluzione elevata, meno ambiguità conformazionali vi sono
 - Si considera alta risoluzione 1.5Å, bassa risoluzione 5Å
 - * Per poterci lavorare bene deve essere almeno 2.5Å
- Oggi la cristallografia si fa in pochi centri specializzati
 - In Europa si fa a Grenoble dove c'è un sincrotrone (European Synchrotron Radiation Facility, ESRF)
 - Contiene camere di analisi poste tangenzialmente all'anello del sincrotrone
- Oggi si sta iniziando ad usare la diffrazione a raggi X per medical imaging e microchirurgia
- Si conoscono più di 10^5 proteine, ma solo 10^4 strutture cristallografiche
- Vi sono grandi investimenti sullo studio delle proteine
- L'utilizzo di enzimi permette di effettuare reazioni chimiche estremamente selettive
- I detersivi per lavatrice hanno una grossa componente enzimatica
 - Le proteine sono stabilizzate con ponti SS e altri legami per farle resistere nelle condizioni di utilizzo

Molecular Interaction Fields (MIF)

- Il target dei MIF può essere qualsiasi molecola a struttura 3D nota
- Si costruisce un reticolato di punti che circonda completamente il target
- Si posiziona un probe in ogni punto del grid, valutandone le interazioni col target
- I contributi dell'interazione del probe col target sono simili a quelli descritti per la meccanica molecolare, ma non includono i legami di bonding e hanno componenti aggiuntivi
 - $E = E_{Lendar-Jones} + E_{coulomb} + E_{h-bond} + E_{entropy}$
 - Per l'interazione coulombica, in biologia si usa un'equazione modificata e più complessa
 - * Modella l'azione dell'acqua nelle proteine, in cui la costante dielettrica non è più costante (!)
 - Il legame H è elettrostatico ma non è descritto adeguatamente dall'interazione coulombica
 - * La direzionalità modifica l'interazione al punto da doverla descrivere separatamente
 - * Dipende dalla geometria delle molecole interagenti e dei loro orbitali
 - * Ha una componente simile alla Lendar Jones, un contributo coulombiano ed una componente geometrica
 - * Screenando DBs per la posizione di molecole d'acqua per un particolare gruppo chimico in diversi contesti trovo gli angoli di interazione più favorevoli
 - * Il carbonile forma legami H soprattutto in direzione dei lone pairs, ma un po' anche tra di essi
 - Tra i lone pairs è possibile per H₂O formare 2 legami con entrambi, ma la sovrapposizione orbitalica è inferiore
 - Viene considerato il contributo entropico
 - * Le molecole di solvente scalzate da una superficie vanno a fare interazione tra loro stesse, favorendo la propria rimozione
 - * Una molecola di acqua libera forma in media 3 legami H, con un guadagno di -0.9 kcal/mol
 - * Questo è particolarmente importante con probes idrofobici
- Le interazioni deboli vanno da 0 a 10 kcal/mol
 - La Leidar-Jones è di circa 1 kcal/mol
 - Il legame H è di circa 4 kcal/mol
 - Le interazioni ioniche possono raggiungere i 15 kcal/mol
 - La componente entropica è sulle 1.5 kcal/mol

- Il target dei MIFs è l'insieme dell'interazione, non le singole componenti
- La zona di interesse può essere l'intera molecola o una particolare porzione dello spazio
 - La zona di interesse è definita con una griglia
- In tale regione metto un **probe** chimico, con cui la scanso muovendolo per righe, colonne e piani
- La sonda chimica da utilizzare può essere scelta in base alle necessità
 - Posso usare ioni, piccole molecole, ecc...
 - Posso anche usare parti fittizie di molecole, ad esempio un gruppo OH non legato a nulla
 - La scelta della sonda dipende dal tipo di interazioni con la proteina che voglio studiare
- Per simulare l'interazione calcolo la sommatoria dell'interazione del probe
 - Una risultante negativa indica attrazione, una positiva repulsione del probe
 - L'interazione è valutata con tutta la proteina, anche nelle zone al di fuori della regione di interesse (!)
- Nel Flexible MIF la proteina ed il probe si muovono per minimizzare l'energia libera
 - Oltre a legami diretti col probe, si valutano anche tutti i legami indotti all'interno della proteina stessa
 - Questo può alterare la conformazione in zone distanti della proteina (!)
 - Spesso una molecola d'acqua può fare da ponte per trasmettere un legame H
 - Il risultato ottenuto è molto più accurato di un MIF statico se sto studiando una proteina con conformazione flessibile
 - E' importante notare che le possibili interazioni predette sono in molti casi mutualmente esclusive (!)
 - Questi campi sono usati per studiare come una certa regione possa alterare le proprie proprietà in virtù della conformazione
- Il software **GRID** è gratis per l'accademia e a pagamento per i *for profit*, impiega i MIF
 - L'interazione è valutata come il lavoro necessario a portare il probe da infinito al punto in esame
 - I probe sono spesso anisometrici e sono liberi di ruotare
 - Assume l'acqua come solvente di target e probe
 - Considera la presenza di tautomeri

Importanza delle interazioni tra proteine e piccole molecole

- Spesso quella idrofobica è l'interazione più importante per il numero di interazioni che si formano
 - Un farmaco idrofobico è spesso più potente di uno idrofilico perchè la sua interazione col target è favorita
 - E' molto più tollerante ad imprecisioni nella previsione del legame, poichè è poco direzionale
- In biologia la selettività è data dalle interazioni polari, la potenza dell'interazione da quelle idrofobiche
- Posso utilizzare un probe idrofobico, che spiazzale delle molecole d'acqua di solvatazione da una superficie apolare
 - Queste molecole d'acqua subiscono un guadagno energetico di circa 0.9 kcal/mol dovuto alle maggiori interazioni che possono formare quando non solvatano la superficie idrofobica
 - Si ha un ulteriore guadagno per interazioni di Lennard-Jones di circa 1 kcal/mol tra la superficie ed il probe
- Consideriamo ora lo stesso probe idrofobico su una superficie polare
 - Si hanno le stesse componenti di prima che determinano -1.9 kcal/mol, ossia guadagno entropico del solvente e interazioni Lennard-Jones
 - Si ha una perdita di energia dovuta alla rottura dell'interazione tra il gruppo polare e le molecole di acqua, di circa +2.5 kcal/mol
 - Il processo non è spontaneo in quanto ha un'energia di circa +0.6 kcal/mol

Progettare una molecola con GRID

- Il primo step è trovare il sito o i siti di modulazione/legame
- Tramite vari probes vedo dove certi gruppi sono favoriti
 - Posso usare gruppi fosfato, metile, carbonile

- Un singolo composto potrebbe interagire con più di una tasca nella stessa proteina
- E' possibile descrivere un pocketoma che raccolga le tasche note, e predica le interazioni di una molecola
 - E' rappresentato come network che misura la distanza di fitting di 2 tasche
- Si fanno simulazioni di fitting provando le varie tasche disponibili
- GRID può essere impostato sia con MIF rigidi che flessibili

Progettare lo scaffold

- GRID mi dice dove devo mettere le mie decorazioni
- Il chimico si preoccupa di creare uno scaffold per posizionare tali gruppi in modo appropriato
- Oggi esistono anche software in grado di suggerire lo scaffold appropriato
- Nota che tutti questi stilemi progettano ligandi, non farmaci, non è detto che le molecole che trovo abbiano attività biologica (!)

Uso dei MIF per produrre un farmaco anti-influenzale

- In 7 anni, senza il cristallo della neuraminidasi, si è riuscito a sviluppare un farmaco capace di inibirla
 - Questo composto aveva poca affinità, con K di circa $1 \mu\text{mol}$
- A seguito della pubblicazione del cristallo del farmaco nella proteina si è cercato di aumentarne la potenza con poco successo
- Il prof Cruciani a Oxford ha usato i MIF per migliorarlo
 - Ha visto che usando un particolare probe vi era una zona ad alta affinità nella tasca di legame
 - Si è quindi inserito tale gruppo in posizione consona, producendo lo Zanamivir
- La GSK ha comprato la ditta che lo produceva, rinominando il farmaco Relenza
- La Gilead ha copiato il farmaco aggiungendo un estere di un gruppo carbossilico, creando il Tamiflu
 - L'estere viene idrolizzato a livello gastrico riproducendo il gruppo carbossilico
 - Lo ha fatto per evadere il brevetto e per aumentarne la permeabilità di membrana
 - Ha anche sostituito una porzione idrofilica con una idrofobica, usando una previsione fatta con GRID
- La Roche ha comprato il Tamiflu a 100 milioni di dollari, ottenendoci guadagni enormi
- Il Tamiflu fino a qualche anno fa aveva il 90% di share mentre il Relenza il 10%
- Adesso si sta sviluppando resistenza al Tamiflu, mentre non vi è ancora resistenza a Relenza
 - Questo perché il Tamiflu è stato abusato mentre Relenza no

Scelta dei probe e dei livelli energetici

- Il probe H_2O (wat) con energia attrattiva (-3 kcal/mol) mi definisce aree idrofiliche dove è possibile formare legami H
- Il probe wat con energia repulsiva ($+0.2 \text{ kcal/mol}$) mi definisce l'ingombro sterico della molecola
 - E' interessante usare wat poiché le interazioni avvengono in acqua in biologia, mi interessa l'ingombro in acqua
 - Questo mi permette di calcolare il volume e l'ingombro sterico della molecola
- Il probe idrofobico (dry) mi definisce le porzioni apolari a -1 kcal/mol
 - Le strutture delocalizzate sono molto apolari poiché polarizzabili
- Posso usare probe anfilici per evidenziare dove avviene la transizione idrofobico-idrofilico
- Possiamo attuare le tecniche precedenti anche con piccole molecole, ad esempio lipidi
- Studiando con probe idrofobico, wat e anfilico il colesterolo posso predire come questo si posiziona sulle membrane
- In una fosfatidilcolina sorprendentemente l' N^+ quaternario non è idrofilico (!)
 - Osserviamo una porzione idrofobica sulla coda alifatica, una idrofilica a livello del fosfato e una intermedia a livello dell'azoto
- I diacilgliceroli in acqua si dispongono con le code adiacenti per minimizzare la superficie idrofobica esposta

- L'idrofobicità della molecola con le code appaiate è inferiore a quella della stessa con code separate (!)
- In un trigliceride similmente le code sono disposte aggrovigliate tra loro
- Minimizzare una struttura significa trovare la sua conformazione di minimo energetico

Previsione della pK_a di composto (MoKa)

- La pK_a è essenziale per determinare la permeabilità di molecole a livello delle membrane biologiche
 - I farmaci tendono ad essere carichi a pH fisiologico
- Per convenzione si usa come parametro sempre la pK_a , anche per le basi
- La pK_a dipende da alcuni effetti
 - L'elemento a cui l' H^+ acido è legato
 - * Più questo è elettronegativo più il protone è acido
 - La base coniugata è più stabile con un elemento fortemente elettronattrattore, che può accettare l'elettrone in eccesso
 - * Più l'atomo è grande più è acido poiché distribuisce la carica su un volume maggiore
 - Stabilizza la base coniugata
 - Effetto induttivo mediato dai legami σ
 - * Altri atomi elettronegativi presenti nella molecola influenzano l'acidità sottraendo elettroni e quindi diffondendo la carica
 - Effetto di risonanza mediato da legami π
 - Ibridazione dell'atomo a cui il protone è legato
 - * All'aumentare del carattere s aumenta l'acidità
 - * Un maggior carattere s significa elettroni più vicini al nucleo, e quindi base coniugata più stabile
 - Effetto prossimità, dovuto alla vicinanza di un gruppo che forma legami H intramolecolari
- Una pK_a sperimentale è sempre preferibile ad una calcolata
 - Non sempre è possibile la determinazione sperimentale per motivi tecnici o economici, e quando ho molte molecole da analizzare
- La pK_a può essere determinata sperimentalmente in vari modi
 - Elettroforesi capillare, in cui il tasso di migrazione dipende dalla carica del composto ad un dato pH
 - Titolazione spettrofotometrica se il composto ha una variazione di assorbimento a seguito della dissociazione
 - Titolazione potenziometrica usando un pH-metro
- I metodi ab initio per la previsione della pK_a sono troppo poco precisi e richiedono molto tempo
- I metodi sperimentali richiedono di avere a disposizione un campione, e non sempre è possibile
- La QSAR (quantitative structure-activity relationship) è la tecnica più usata per questa previsione
 - Il software usato è MoKa
 - Molto veloce ed accurata, considera le molecole poliprotiche
 - E' indipendente dalla rappresentazione esplicita degli H
 - Calcola il LogP (partizione ottanolo- H_2O) e LogD (LogP in funzione del pH) del composto in esame
 - Può essere allenato usando un modello
 - Si basa sui campi GRID, dove un probe si muove e valuta le interazioni con la molecola
 - * Si stanno ora sviluppando anche probes poliatomici per migliorare la previsione
 - Si creano dei livelli allontanandosi dall'atomo di riferimento (non ho ben capito questa parte)
 - * Il livello 0 è l'atomo stesso, e si pongono tutti i bit a 0 eccetto quello per N
 - * La rappresentazione è non più binaria
 - * Si crea un fingerprint concatenando questi bit
 - Si crea una tabella che correla il fingerprint con la pK_a , usata come modello di training
 - * Usa il metodo dei partial least squares (PLS)
 - Bisogna considerare i tautomeri del composto, poiché ognuno ha una sua pK_a (!)

QSPR predictions (Volsurf)

- Volsurf usa descrittori molecolari per effettuare previsioni QSPR
 - In generale sono volumi di interazione con un probe sopra o sotto ad un certo cutoff energetico, e vettori momento
 - E' importante disegnare descrittori non correlati tra loro, altrimenti descrivo 2 volte la stessa cosa (!)
- Descrittori a energia positiva (volume e superficie)
 - Il volume della molecola viene determinato come volume con interazione $> +0.2$ kcal/mol
 - La superficie viene determinata sempre a $+0.2$ kcal/mol
 - Il rapporto volume/superficie è un altro descrittore
 - La globularità della molecola indica quanto questa devii da una sfera di pari volume
 - * $Glob = S_{tot}/S_e$
 - * S_e è il volume di una sfera di volume pari a V_{tot}
- Descrittori di interazione con H_2O (WAT)
 - Volumi di 8 livelli energetici di interazione col probe WAT (-0.2, -0.5, -1, -2, -3, -4, -5, -6 kcal/mol)
 - Integy moment
 - * E' un vettore che va dal centro di massa della molecola al centro di volume dell'interazione
 - * Se l'idrofilità è diffusa anziché localizzata è più facile che un composto passi la BBB
 - * Vi sono 8 integy moments, uno per livello di interazione
 - * Non vi è correlazione tra momento idrofilico e forza delle interazioni idrofiliiche
 - Capacità, data dal rapporto tra volume di interazione e superficie della molecola
 - * 8 descrittori, uno per livello
 - Capacità di formare legami H su 8 livelli
- Descrittori di interazioni idrofobiche
 - Volumi di 8 livelli di interazione col probe DRY (-0.2, -0.4, -0.6, -0.8, -1, -1.2, -1.4, -1.6 kcal/mol)
 - Integy moment agli 8 livelli
- Descrittori misti
 - Bilanciamento idrofobico-idrofilico
 - * E' il rapporto tra volumi idrofobici ed idrofiliici
 - Momento anfifilico
 - * E' il vettore tra i centri di massa idrofobico ed idrofilico
 - Critical packing
 - * E' una parametrizzazione della concentrazione micellare critica
 - E' un parametro usato da chi studia tensioattivi e micelle, inventato da un russo
 - * E' un fattore geometrico che indica il fattore critico di impacchettamento
 - * Alla concentrazione micellare critica una molecola anfifilica forma micelle anziché stare in soluzione
 - * E' dato da un'equazione che considera lunghezza lipofila della molecola, volume lipofilo e superficie idrofilica
 - * Il risultato è adimensionale, è un volume diviso una superficie per una lunghezza (ossia un volume)
 - Diffusività molecolare in acqua
 - * Usa l'equazione di Stokes-Einstein modificata, che predice la diffusività in acqua usando solo parametri fisici
 - * Un modello del prof di machine learning usa i vari descrittori per predire la diffusività
 - La diffusività è poi trattata come un altro descrittore
 - E' in accordo con i dati sperimentali
 - Elongation
 - * E' la lunghezza più probabile che una molecola flessibile assume in soluzione
 - * In una molecola flessibile è diversa dalla lunghezza della molecola disegnata (!)
 - * Può essere calcolata con un'equazione
 - LogP
 - * E' il Log del rapporto delle concentrazioni della molecola in n-ottanolo e acqua

- Approssima il comportamento della molecola sulle membrane biologiche
- $P = [n - ott]/[H_2O]$
- Una volta si usava olio di oliva ma poi si è standardizzato con n-ottanolo
- L'n-ottanolo è poco miscibile in acqua
- Se LogP è 0 significa che $P=1$ e quindi la molecola ha la stessa solubilità in acqua e n-ottanolo
- Se è LogP è 1 il composto è 10 volte più solubile in n-ottanolo che in acqua, se il LogP è -1 viceversa
- * Dipende dalla conformazione assunta dalla molecola
 - Il valore sperimentale è una media pesata del LogP delle varie conformazioni
- * La variabilità della stima è ora nel range di errore sperimentale (+/- 0.7)
- * Come calcolare il LogP
 - Riesco a correlare bene il LogP con la conformazione, ma vi sono outliers
 - Gli outliers sono molecole molto flessibili che non hanno una conformazione ben definita
 - Quale conformazione scelgo per la previsione?
- * Posso dire al software di modificare la molecola per adattarla al pH di lavoro
 - Modifica lo stato di protonazione in base al pH
 - Usa l'algoritmo di MoKa
 - Il LogP varia molto con il pH (!)
- * Altri descrittori non trattati
 - Energia minima di interazione locale
 - Distanza dell'interazione locale
 - Polarizzabilità
- Volsurf usa un modello multivariato per correlare i descrittori a proprietà sperimentali, ad esempio la biodisponibilità

Metabolismo degli xenobiotici

- Gli xenobiotici sono composti che penetrano nell'organismo ma non svolgono funzione fisiologica
- Sono assorbiti, metabolizzati ed escreti
- Il metabolismo ha una fase I di funzionalizzazione ed una fase II di coniugazione
- Il metabolismo avviene in fegato, BBB, reni, intestino, microflora, polmoni, plasma, pelle, e molti altri siti
- Un modello sperimentale molto usato per lo studio del metabolismo sono i microsomi
 - Si centrifuga un omogenato di tessuto (fegato di solito) a 10000 g per 10 minuti
 - * Questo separa nuclei, mitocondri e debris
 - Si centrifuga il supernatante a 100000 g per 1 ora
 - Nel pellet ho i microsomi, ossia parte del ER contenente i CYP
- La maggior parte delle drugs fallisce per problemi di metabolismo
- Gli organoidi sono usati per studiare il metabolismo di xenobiotici in un sistema più rappresentativo
- I profarmaci sono metabolizzati a farmaco attivo dai CYP
- Il CYP3A4 e il CYP2D6 compiono da soli quasi 3/4 delle funzionalizzazioni
- La finestra terapeutica è il tempo in cui la concentrazione di farmaco rimane sopra alla concentrazione minima attiva
- Le reazioni avverse ai farmaci sono la 4° causa di morte negli US
- La predizione dei siti di metabolismo prima della produzione di un farmaco ha un'importanza chiave
- Posso modificare una molecola per alterarne il metabolismo e la biodisponibilità
 - Il fluoro è isosterico all'idrogeno, e posso introdurlo in molecole per alterarne la reattività
 - * Il fluoro è spesso utilizzato per gestire il metabolismo dei farmaci
 - * Il legame F-C è molto forte e difficile da rompere
 - * Essenzialmente impedisce l'astrazione del H che rimpiazza
- Il metabolismo di un composto è influenzato da vari fattori
 - Fattori sterici

- Orientamento nel sito attivo, ossia quale parte è esposta al centro catalitico
 - * E' dovuto sia a fattori sterici che termodinamici
 - La cinetica e termodinamica della reazione stessa
 - * Di solito, significa quanto è facile astrarre un certo protone
 - I CYP presentano un centro catalitico con eme legato ad un ossigeno
 - La forma della tasca varia tra i CYP, e ne determina la specificità
 - Metasite predice i siti di metabolismo effettuando una simulazione di docking
 - Valuta l'ingombro sterico per capire se la molecola entra nella tasca enzimatica
 - Predice quale parte della molecola sarà vicina al sito attivo
 - Contiene pre-calcolati i MIF GRID dei CYP, la struttura dei CYP, il loro fingerprint
 - Prende in input una SMILES o una struttura mol2, genera i conformeri, ne calcola il fingerprint e la reattività
 - Non richiede training, non è un QSAR e non fa docking
 - Cerca di sovrapporre i MIF di tasca e substrato
 - La probabilità di un sito di metabolismo è data dal prodotto tra l'esposizione al sito attivo e la reattività della posizione
 - Le reazioni più comunemente svolte dai CYP sono
 - Idrossilazione di un C alifatico
 - * L'eme ossidato astrae un protone dal C formando un radicale
 - * L'eme cede l'OH al radicale rigenerandosi
 - Idrossilazione di un C aromatico
 - Epossidazione di doppi legami
 - Demetilazione e deidrogenazione
 - Rottura di esteri
 - La piridina si ossida su N a dare un N-ossido, molto polare
 - Le posizioni benziliche (un C di distanza da un fenolo) sono vulnerabili, ma comunque la loro reattività dipende dall'esposizione al centro catalitico
 - Il radicale benzilico è un toluene senza un elettrone
 - * E' uno dei radicali più stabili, e quindi uno di quelli che si formano più facilmente
 - * Per questo una volta si pensava fosse sempre uno dei siti più reattivi per le ossidazioni CYP, che sono radicaliche
 - * In realtà comunque dipende molto dall'esposizione del sito
 - Se voglio una prova del metabolismo posso fare una LC-MS/MS prima e dopo il metabolismo con microsomi
 - La differenza di massa mi dà un'idea di che modifiche sono avvenute
 - Non sempre è possibile identificare univocamente il metabolismo avvenuto
-