



Solubilità e lipofilia in funzione del pH

di Elena Fioravanzo, Filippo Magnaguagno, Massimo Mabilia

Due software che utilizzano un recente algoritmo per la predizione di proprietà chimico-fisiche quali la solubilità e la lipofilia anche in funzione del pH. Diversi set di molecole organiche vengono utilizzati per verificare la bontà delle predizioni del Log P e confrontata l'accuratezza raggiunta nel calcolo del Log P con quella ottenibile con la predizione del Log D, valore di Log P che tiene conto della ionizzazione delle molecole e dipende quindi dal pH.

La lipofilia gioca spesso un ruolo significativo nel determinare l'attività biologica di piccole molecole organiche. Influenza, infatti, sia le interazioni idrofobiche tra il recettore e il ligando sia le sue proprietà farmacocinetiche quali ad esempio l'assorbimento o il trasporto attraverso la membrana cellulare. La lipofilia è quindi, al di là dell'esatto meccanismo di intervento, un termine fondamentale nelle relazioni quantitative attività biologica-struttura molecolare (Quantitative Structure - Activity Relationships, QSAR).

Nella maggior parte dei casi la lipofilia è misurata con il coefficiente di partizione ottanolo-acqua cioè il rapporto tra la concentrazione di un soluto in ottanolo rispetto a quella in acqua in un sistema a due fasi all'equilibrio. Dato che i valori misurati del coefficiente di partizione variano da meno di 10^{-4} a più di 10^8 (almeno dodici ordini di grandezza), si usa comunemente il logaritmo della costante di equilibrio della partizione di un soluto, Log P, per caratterizzarne i valori.

Esiste un rinnovato interesse per lo sviluppo di metodi per il calcolo del valore del Log P [1] visto che la lipofilia, esercitando una notevole influenza sia sul trasporto del ligando sia sull'interazione di questo con il recettore, ha un grande valore pratico per la derivazione di modelli quantitativi per l'attività biologica. Tra i metodi disponibili per la stima del Log P ecco i più utilizzati:

- il metodo della costante f di Rekker [2];
- l'approccio frammentale di Leo e Hansch [3];
- l'approccio di Broto e Moreau [4].

Rekker definisce dapprima un arbitrario set di frammenti molecolari utilizzando un database di circa mille composti di cui è noto il Log P. Effettua poi un'analisi di regressione lineare utilizzando i valori di Log P come variabile dipendente. I coefficienti ottenuti sono utilizzati come contributi dei vari frammenti. Per

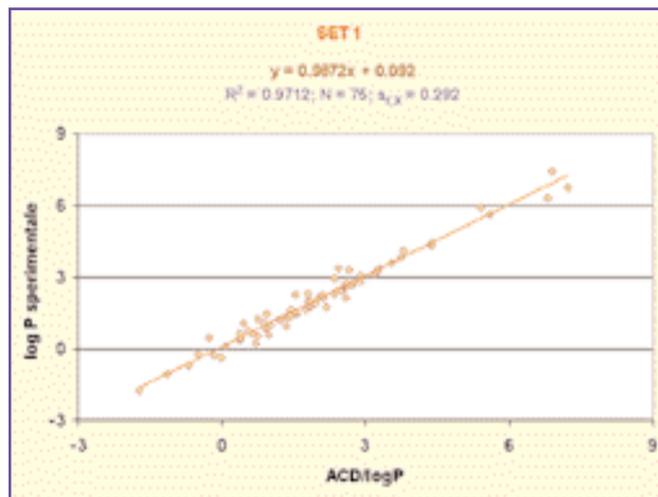


Figura 1 - Rappresentazione grafica e relativi dati statistici della relazione tra valori di Log P sperimentali e calcolati da ACD/LogP DB per le 75 molecole del set 1

stimare il coefficiente di partizione di un composto, quindi, si suddivide la molecola in frammenti e si sommano i diversi contributi tenendo conto anche di appropriati fattori di correzione sviluppati e introdotti in seguito.

La filosofia di Leo e Hansch è invece di determinare in modo molto accurato i valori di Log P di un set di piccole molecole e di calcolare il contributo dei vari frammenti da questi dati. Anche questo sistema include molti fattori di correzione.

Dato che la frammentazione di una molecola, utilizzata dai due approcci precedenti, non è univoca, Broto e Moreau hanno sviluppato un metodo alternativo che si basa su contributi atomici al valore totale di Log P del composto. In questo modo si evita il passaggio critico della suddivisione della molecola in frammenti utili, ma ci si trova nella necessità di utilizzare un elevato numero di "atom type" non sempre facilmente definibili. Il contributo di un singolo atomo alla lipofilia totale di una molecola dipende, infatti, non solo dal tipo di atomo considerato, ma anche dall'intorno in cui esso si trova nella molecola.

Il Log D

Il Log P è un parametro teorico che caratterizza solamente le strutture neutre. Spesso però le molecole, ad esempio i farmaci in condizioni fisiologiche, esistono come elettroliti ionizzati. Per questa ragione il Log D, che tiene conto della ionizzazione delle molecole, è meglio correlato all'attività biologica o a proprietà macroscopiche e può essere messo direttamente in relazione con il trasporto di molecole ionizzabili o con la loro interazione con il recettore. Queste considerazioni sono peraltro conferma-





te dal recente paragone, riportato in letteratura, delle correlazioni di Log P e Log D con costanti di inibizione di enzimi e costanti di "binding" [5]. In tutti i casi analizzati utilizzando il Log P si ottengono coefficienti di correlazione $R < 0,90$, mentre con il Log D si hanno valori di $R > 0,99$. L'unico vantaggio dell'utilizzo del parametro Log P invece di Log D è che quest'ultimo è più complesso da calcolare con analoga accuratezza. In questo articolo sono presentati due software che utilizzano un recente algoritmo per la predizione del Log P e del Log D. Si utilizzano diversi set di molecole organiche per verificare la bontà delle predizioni del Log P e si confronta l'accuratezza raggiunta nel calcolo del Log P con quella ottenibile con la predizione del Log D.

Materiali e metodi

Data-set

• Set 1

Il primo set di molecole in esame è formato da 75 molecole organiche [6] i cui pesi molecolari variano da un minimo di 42 a un massimo di 361, con un picco di oltre 30 molecole fra 150 e 200, di cui si dispongono i valori sperimentali di Log P. Le strutture bidimensionali sono state generate, a partire dai nomi di uso comune, con CS ChemDraw Ultra (modulo di ChemOffice di CambridgeSoft Co. [7]) e inserite in un foglio di calcolo di Accord for Excel (di Synopsys Ltd. [8]) da cui risulta agevole l'esportazione di file in formato SD da passare ai programmi per il calcolo automatico del valore teorico di Log P.

• Set 2

Sono state selezionate venti molecole di interesse biologico appartenenti a diverse categorie terapeutiche per le quali sono disponibili i dati di Log P* [9] (Tabella). Le strutture bidimensionali e i dati relativi alle categorie terapeutiche sono stati tratti da Dictionary (ACD/ChemSketch di AcD Inc. [10]) e inseriti, insieme ai dati di Log P*, in un database creato con ChemFinder (ChemOffice di CambridgeSoft Co. [7]). È stato quindi esportato un file SD utilizzato per il calcolo del Log P.

• Set 3

Il set è composto da 21 molecole di interesse biologico, 19 delle quali fanno anche parte del set 2. Per queste molecole sono disponibili valori di lipofilia determinati sperimentalmente a diversi valori di pH [11] (Tabella). Per alcune di queste

Valori sperimentali di Log D e Log P* confrontati rispettivamente con i valori calcolati da ACD/LogD e ACD/LogP

Molecola	pH	Log D ^(a)		Log P	
		sperim.	ACD/LogD	sperim.	ACD/LogP
Atropine	7,4	-0,44	-0,94	1,83	1,53
Chloramphenicol	7,4	1,00	1,02	1,14	1,02
Chlorothiazide	da 1 a 3	-0,10	-0,03	-0,1	-0,03
	5	-0,19	-0,04		
	6,5	-0,52	-0,23		
Chlorpromazine	5	1,13	0,92 ^b	5,35	5,36
	7,4	3,15	3,36		
	8	4,04	3,92		
Cimetidine	7,4	0,33	0,21	0,4	0,36
Diazepam	2	1,20	1,56	2,8	2,96
	2,5	1,66	2,02		
	4	2,80	2,87		
	7,4	2,80	2,96		
Diltiazem	7,4	1,29	2,12	2,7	3,63
Diphenhydramine	7,4	1,61	2,29	3,27	3,66
	9,2	3,20	3,53		
Disopyramide	7,4	0,38	0,31	2,58	2,86
Flufenamic acid	2	5,62	5,61	5,25	5,62
	2	5,25	5,61		
	7,4	2,02	2,03		
	8	1,74	1,71		
Haloperidol	6	1,57	1,87	3,36	4,06
	7,4	2,69	3,16		
Imipramine	6	1,21	1,51	4,8	4,47
	7	2,16	2,07		
	7,4	2,43	2,41		
	11	4,68	4,45		
Lidocaine	6,4	0,72	0,27	2,26	2,36
	7,4	1,13	1,20		
Phenobarbital	7	1,32	1,62	1,47	1,71
	7,4	1,53	1,51		
Phenytoin	3,5	2,46	2,52	2,47	2,52
	7	2,46	2,49		
	7,4	2,29	2,44		
	11,2	0,14	-0,57		
Procainamide	7,4	-1,16	-1,14	0,88	1,23
Propranolol	4	0,38	0,00	3,56	3,10
	7	0,73	0,99		
	7,4	1,20	1,36		
	9	2,62	2,72		
	13	3,28	3,04		
Tetracaine	7,4	2,73	2,86	3,73	3,75
	8	3,00	3,32		
Trimethoprim	7,4	0,64	0,52	0,91	0,79
	13	0,87	0,79		
Furosemide	7,4	-0,92	-0,99		
Verapamil	6,1	1,34	2,32		
	7	2,28	3,02		
	9	3,79	4,71		

(a) valore di Log D calcolato e misurato al pH indicato

(b) valore di Log D calcolato senza ripartizione ionica (ion-pair partitioning)





molecole sono disponibili anche i valori sperimentali di pKa [12]. Le strutture sono state generate con l'aiuto di Dictionary (ACD/ChemSketch di AcD Inc. [10]) e inserite in un foglio di calcolo di Accord for Excel (di Synopsys Ltd [8]).

Infine sono state considerate due molecole che rappresentano dei casi particolari: la furosemide per la quale lo storico valore sperimentale di Log P proposto sembra non essere corretto e la fenilalanina, a rappresentare gli amminoacidi in generale.

Software

• ACD/LogP

L'algoritmo utilizzato da ACD/LogP è stato sviluppato utilizzando un "training set" che contiene oltre 5.000 valori di Log P determinati sperimentalmente. Questo algoritmo include contributi di atomi separati, di oltre 500 frammenti strutturali e di oltre 2.000 tipi di interazioni intramolecolari tra diversi frammenti. ACD/LogP offre anche la possibilità di includere "nuovi" frammenti nel database in modo che l'utente possa migliorarne le capacità predittive utilizzando i propri dati sperimentali ("learning system"). È sufficiente fornire al programma una nuova struttura con il relativo valore sperimentale di Log P: la struttura sarà automaticamente suddivisa in frammenti per i quali sarà calcolato l'incremento additivo a partire dal dato sperimentale. Questi nuovi dati saranno aggiunti al database già disponibile ed eventualmente utilizzati per successivi calcoli, in base alla loro rilevanza statistica. È generalmente sufficiente aggiungere al sistema il dato misurato per una sola struttura che appartiene a una nuova classe di composti per migliorare ulteriormente le predizioni di Log P di altre molecole appartenenti alla stessa classe. Il programma è quindi personalizzabile dall'utente in base alle proprie esigenze.

• ACD/pKa

Il pKa è una misura della tendenza di una molecola o ione a trattenere un protone, H⁺ sui suoi centri di ionizzazione. Più facilmente una molecola si ionizza, più questo composto tenderà a rimanere in una soluzione acquosa dato che l'acqua è un potente solvente polare. Se una molecola si ionizza difficilmente questa tenderà a distribuirsi maggiormente in un solvente non polare quali cicloesano o ottanolo. Il valore di Log P di un composto è quindi strettamente dipendente dal suo valore di pKa. In termini biologici il pKa è un fattore importante nel determinare se una molecola tenderà a essere assorbita dai componenti acquosi del tessuto cellulare oppure dalla membrana lipidica.

ACD/pKa (di AcD Inc. [10]) calcola accurate costanti di ionizzazione acido-base (valori di pKa) a 25 °C e forza ionica zero in soluzione acquosa di molecole organiche. Il programma utilizza un approccio frammentale e un database interno contenente 8.000 strutture con oltre 23.000 valori sperimentali di pKa. Anche ACD/pKa permette l'implementazione del database interno con dati sperimentali forniti dall'utente. Il programma calcola sia le costanti di dissociazione osservate o macroscopiche che le costanti di dissociazioni microscopiche di specifiche forme ioniche che potrebbero essere presenti in un sistema all'equilibrio.

• ACD/LogD che contiene ACD/LogP e ACD/pKa

Questo programma (di AcD Inc. [10]), unendo le caratteristiche di ACD/LogP a quelle di ACD/pKa DB, è in grado di calcolare sia il valore di Log P apparente che quello della solubilità in acqua in funzione del pH.

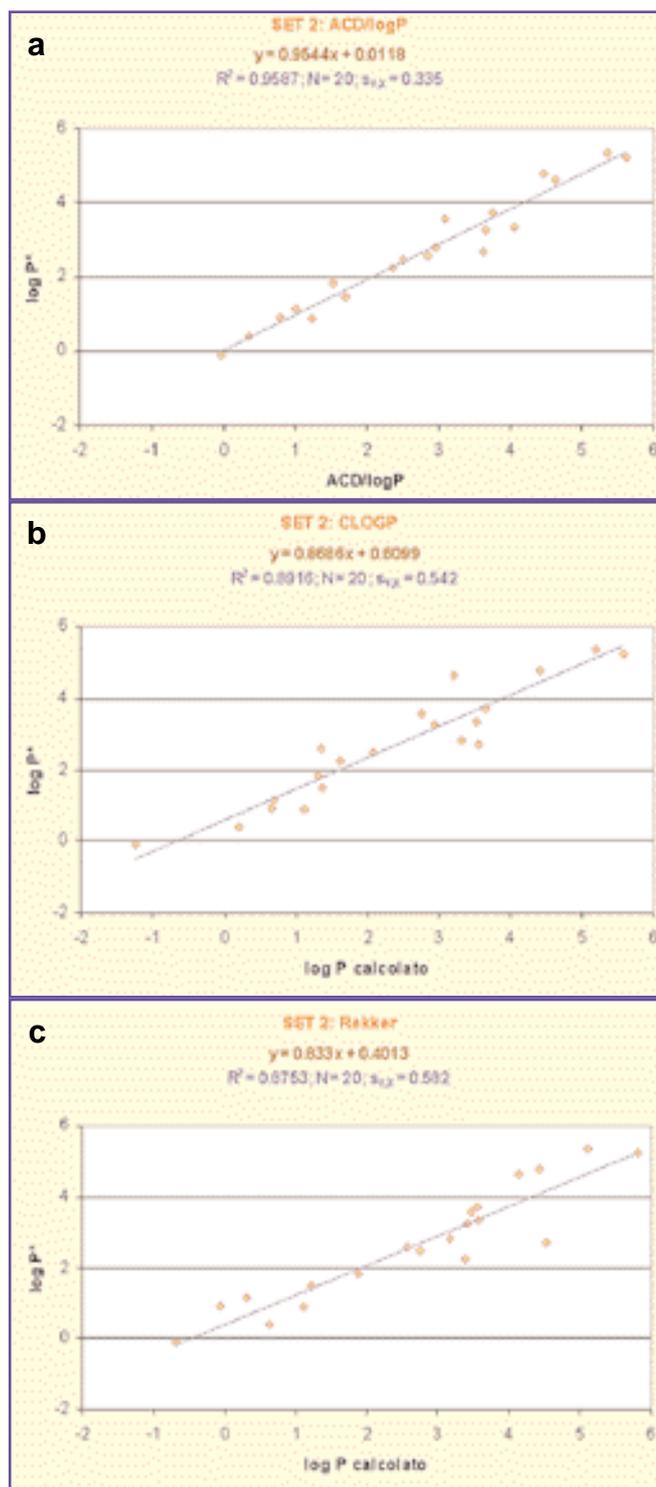


Figura 2 - Rappresentazione grafica e relativi dati statistici della relazione tra valori di Log P sperimentali e calcolati con il metodo implementato in ACD/LogP DB (a), il metodo di Leo e Hansch (b) e il metodo di Rekker (c) per le 20 molecole del set 2

Risultati e discussione

• Set 1

L'errore standard della stima ($s_{y,x}$) ottenuto per le 75 molecole in esame è 0,29. Gli errori commessi da ACD/logP variano da un massimo di 0,95 a un errore minimo <0,01 per ben 17 com-



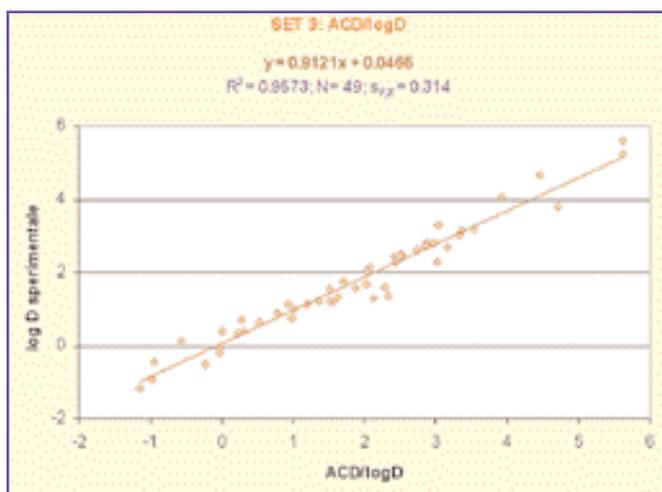


Figura 3 - Rappresentazione grafica e relativi dati statistici della relazione tra valori di Log P sperimentali determinati a diversi valori di pH per le molecole del set 3 confrontati con i valori calcolati da ACD/LogD

posti. Il coefficiente di correlazione R^2 tra i valori sperimentali e calcolati di Log P è 0,97.

In Figura 1 viene rappresentato il grafico dei valori sperimentali di lipofilia contro i valori predetti.

• Set 2

In Tabella sono riportati i valori di Log P sperimentali ($\log P^*$ [13]) e calcolati con il metodo frammentale applicato da ACD/LogP. Per le stesse molecole sono disponibili anche i valori di lipofilia calcolati con altri due metodi: il metodo di Leo Hansch [13] utilizzato per calcolare il valore di CLOGP, il metodo di Rekker [9].

Con il metodo implementato nel software ACD applicato al set di molecole in esame si sono ottenuti risultati migliori rispetto alle predizioni effettuate con i due metodi tradizionali presi in considerazione. I coefficienti di correlazione R^2 tra dato sperimentale e valore calcolato sono infatti di 0,96 per il programma ACD, 0,89 per il metodo di Hansch e Leo e 0,88 per il metodo di Rekker che corrispondono, rispettivamente a un errore standard della stima ($s_{y,x}$) di 0,33, 0,54 e 0,58 (Figura 2).

• Set 3

In Tabella è riportato il Log D sperimentale misurato a diversi valori di pH e il rispettivo valore calcolato da ACD/LogD [10]. Il coefficiente di correlazione R^2 ottenuto tra valore sperimentale e valore calcolato è di 0,96 mentre l'errore standard della stima ($s_{y,x}$) è di 0,31 (Figura 3).

Paragonando questo risultato con l'errore standard commesso da ACD/LogP, 0,33, su di un set di molecole analogo a questo (set 2) nel predire i valori di Log P si potrebbe concludere che il Log D è predetto con analoga o addirittura maggiore accuratezza del valore di Log P. Questo risultato è alquanto inaspettato dato che il calcolo del parametro Log D è molto più complesso del calcolo del semplice valore di Log P. La predizione del Log D prevede infatti il calcolo sia dei valori individuali di Log P, per tutte le forme ioniche della molecola, sia del valore della costante di dissociazione acido-base, pKa, della molecola stessa. È lecito quindi aspettarsi che l'errore commesso da una predizione di Log D sia maggiore di quello commesso nel predire il Log P per la molecola neutra.

Una possibile spiegazione [11] di quanto qui osservato è che i valori di Log P sono determinati sperimentalmente in modo meno accurato dei corrispondenti valori di Log D. Il Log P deve essere infatti ottenuto in condizioni tali da annullare la ionizzazione delle molecole essendo un parametro relativo alla molecola neutra. Questo significa in molti casi dover lavorare a valori di pH estremi ($\text{pH} > 10$ o $\text{pH} < 2$) condizioni in cui la lipofilia potrebbe essere influenzata da altri effetti non osservabili nelle condizioni cui sono ottenuti la maggior parte dei valori di Log D (pH tra 5 e 8).

In Figura 4 sono invece riportati i valori sperimentali di pKa disponibili per le 21 strutture del set 3 e i relativi valori predetti da ACD/pKa. Il coefficiente di correlazione ottenuto è 0,97 mentre l'errore standard della stima commesso è 0,45.

Due casi particolari

Furosemide

Sia Rekker [14] sia Leo [15] sono d'accordo nell'affermare che il valore di Log P di -0,83 inizialmente indicato per la furosemide non sia corretto. Questo valore è stato originariamente misurato [16] e riportato senza indicazioni di un controllo del pH. Si può quindi sospettare che il valore sia stato ottenuto a pH 7 e che non sia stata applicata una correzione adeguata alla forma ionizzata della molecola. Possiamo verificare questa ipotesi utilizzando ACD/LogD che effettivamente calcola un valore di -0,83 proprio a pH 7,05, valore cui l'acido della furosemide è ionizzato (Figura 5). Questi risultati sono convalidati anche da successive misurazioni che indicano valori sperimentali di Log P uguale a -1,06 [17a] o a -0,92 [17b] a pH 7,4, e un valore di -0,99 a pH 7,4 calcolato da ACD/LogD.

Per quanto riguarda il Log P (relativo cioè alla molecola neutra), si può quindi sostituire all'iniziale valore di -0,83, il valore sperimentale di 2,03 determinato a pH 2,0 [17a] che è possibile paragonare al valore 2,88 calcolato da ACD/LogD allo stesso pH.

Fenilalanina

Quest'ultimo esempio è particolarmente significativo per illustrare l'utilità e spesso la necessità di calcolare valori di lipofilia

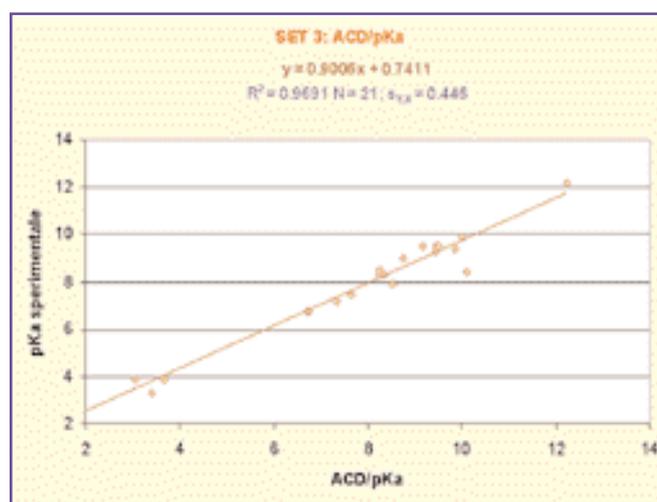


Figura 4 - Rappresentazione grafica e relativi dati statistici della relazione tra valori di pKa sperimentali e i valori calcolati da ACD/LogD



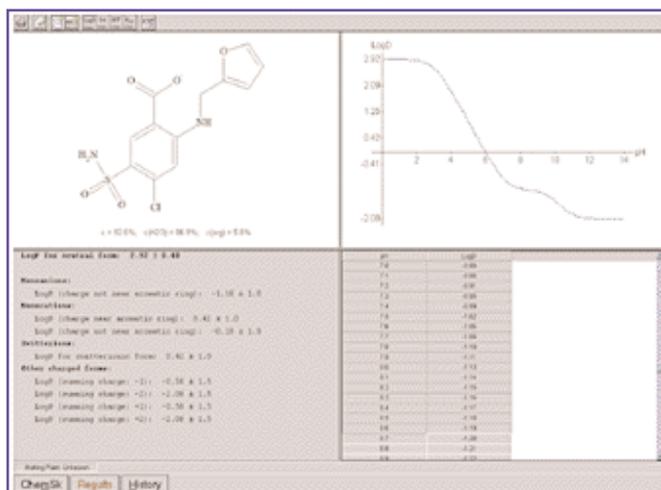


Figura 5 - Calcolo della lipofilia della furosemide con ACD/LogD

in funzione del pH. La fenilalanina, come tutti gli aminoacidi, esiste principalmente come zwitterione (circa da pH 2 a pH 9), quindi con il gruppo amminico protonato e l'acido carbossilico deprotonato, mentre non esiste mai, da pH 0 a pH 14, con entrambi i gruppi ionizzabili neutri. ACD/LogP, come altri programmi analoghi, presuppone l'utilizzo di molecole con i gruppi funzionali ionizzabili non carichi (neutri): calcola perciò un valore di lipofilia (1,11) relativo a una specie non esistente commettendo di fatto un errore di 2,6 unità logaritmiche nella stima del Log P sperimentale (-1,52). Per contro, ACD/LogD è in grado di calcolare la lipofilia dello zwitterione con un'ottima approssimazione (-1,39), Figura 6.

Conclusioni

Il metodo frammentale implementato nei software ACD per la predizione della lipofilia utilizza nuovi algoritmi rispetto ai metodi frammentali tradizionali e si basa su di un vasto database di dati sperimentali che è in continuo sviluppo. Questo metodo si è dimostrato in grado di calcolare il Log P (ACD/LogP DB) di molecole organiche contenenti una vasta gamma di gruppi funzionali (set 1) con un notevole grado di precisione ottenendo per 75 molecole i parametri: $R^2=0,97$ $s_{y,x}=0,29$.

ACD/LogD, inoltre, avvalendosi di ACD/pKa DB, è in grado di predire valori di lipofilia in funzione del pH con lo stesso grado di accuratezza ($s_{y,x}=0,31$) con il quale vengono predetti i valori di Log P per molecole neutre (set 2 e 3) da ACD/LogP ($s_{y,x}=0,33$), e addirittura superiore alle predizioni di Log P ottenute con i metodi di Rekker ($s_{y,x}=0,58$) e di Hansch e Leo ($s_{y,x}=0,54$). Va infine ricordato che, grazie al calcolo della pKa e della lipofilia, ACD/LogD è in grado anche di stimare la solubilità in acqua di molecole organiche in funzione del pH.

Bibliografia

- [1] Ghose A.K., Crippen G.M., *J. Comp. Chem.*, **7**, 565, (1986). (b) Viswanadhan V.N., Ghose A.K., Ravankar G.R., Robins R.K., *J. Chem. Inf. Comput. Sci.*, **29**, 163, (1989). (c) Klopman G., Wang, S., *J. Comp. Chem.*, **12**, 1025, (1991).
- [2] Rekker R.E., *The Hydrophobic Fragment Constant*, Elsevier, Amsterdam, 1976.
- [3] Leo A.J., Jow P.Y.C., Silipo C., Hansch, C. *J. Med. Chem.*,

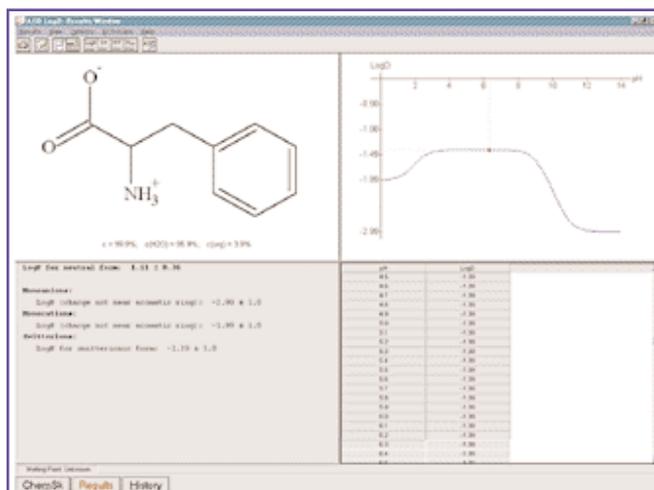


Figura 6 - Calcolo della lipofilia della fenilalanina con ACD/LogD

18, 865, (1975).

- [4] Broto P., Moreau G., VanDycke C., *Eur. J. Med. Chem.*, **19**, 71, (1984).
- [5] E.Kolovanov A., Petrauskas M., Vinogradova, Comparison of LogP and LogD Correlations With Inhibition and Binding Constants of Enzymes, to be published.
- [6] Breindl A., Beck B., Clark T., *J. Mol. Model.*, **3**, 142, (1997).
- [7] CambridgeSoft Corporation, 100 Cambridge Park Drive, Cambridge, MA 02140 Usa, Software distribuito da S.IN-Soluzioni Informatiche, www.s-in.it.
- [8] Synopsys Scientific Systems Ltd., 5 North hill Road, Leeds, LS6 2EN, United Kingdom, Software distribuito da S.IN-Soluzioni Informatiche, www.s-in.it.
- [9] Rekker R.F., Mannhold R., Calculation of Drug Lipophilicity; Vch Publishers, Inc. New York, NY (Usa).
- [10] Advanced Chemistry Development Inc., 90 Adelaide Street West, Suite 702, Toronto, ON, M5H 3V9 Canada, S.IN-Soluzioni Informatiche, www.s-in.it.
- [11] Eduard A. Kolovanov, Alanas A. Petrauskas, *Comparison of the Accuracy of Log P and Log D. Calculations for 22 Drugs*, www.acdlabs.com/publish/publ_pres.html.
- [12] (a) *Clarke's Isolation and Identification of Drugs*, Moffat A.C. et al., Eds, The Pharmaceutical Press, London, 1986. (b) *Sirius Technical Application Notes*, **1** (7), 104 (1994). (c) A. E. Martell, R.M. Smith, "Critical Stability Constants", 1974-1989, v. 1-6. (d) B. Hansen, *Acta. Chem. Scand.*, **12**, 324 (1958). (e) *Sirius Technical Application Notes*, **2** (6), 93 (1995).
- [13] Craig P.N., *Drug Compendium*, in: Vol. VI, Cumulative Subject Index & Drug Compendium (Vol. Ed. Drayton, C.J.) of: *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds* (Eds. Hansch, C.; Sammes, P.G.; Taylor, J.B.), Pergamon Press-Oxford, 1990.
- [14] Rekker R.F., ter Laak A.M., Mannhold R., *Quant. Struct. - Act. Relat.*, **12**, 152, (1993).
- [15] Leo A.J., *Chem. Pharm. Bull.*, **43**, 512, (1995).
- [16] Horstmann H., Moller E., Wehinger E., Meng K, *ACS Symposium Series*, **83**, 125, (1978).
- [17] (a) Mork N., Bundgaard H., Shalmi M., Christensen S., *Int. J. Pharm.*, **60** (2), 163, (1990). (b) Masereel B., Lohrmann E., Schynts M., Piroette B., Greger R., Delarge J., *J. Pharm. Pharmacol.*, **44** (7), 589, (1992).

