

Molecular modeling studies of 2-substituted cephalosporin esters binding to human leukocyte elastase

G Bandoli^{1*}, A Dolmella¹, S Gatto¹, F de Angelis²

¹Dipartimento di Scienze Farmaceutiche, Via Marzolo 5, 35131 Padova;

²Dipartimento di Chimica, Ingegneria Chimica e Materiali, Via Vetolo, 67010 Coppito Due, L'Aquila, Italy

(Received 10 June 1996; accepted 18 September 1996)

Summary — A series of 17 putative human leukocyte elastase (HLE) inhibitors belonging to the family of cephalosporin esters have been studied by means of molecular modeling techniques. The optimized conformation (molecular mechanics) of HLE was used as input for a series of simulated annealing calculations meant to locate a lower energy minimum than that identified by the previous minimization. Manual and automated docking experiments with all the 17 compounds have then been made with elastase in the lowest energy conformation found by the annealing phase. Further molecular dynamics studies of the complexes made by the enzyme with two different inhibitors, as well as a relative free-energy calculation of the two above-mentioned complexes, have been performed in order to get information about the recognition/binding process. The relationships between the steric, electrostatic and lipophilic descriptors (some of which were obtained with semiempirical MO calculations) of the inhibitors, their intermolecular non-bond interaction energies (INIs) and their IC₅₀s have been described with a series of statistical equations. PLS and MLR-like models explaining such relationships have been generated.

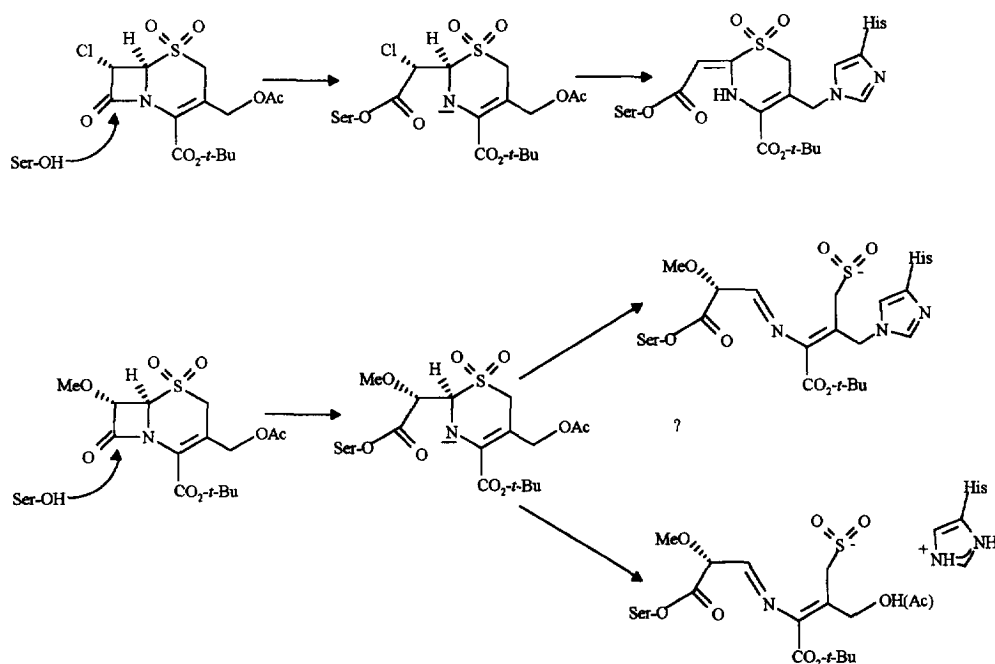
molecular modeling / elastase inhibitors / molecular dynamics / PLS model

Introduction

The human leukocyte elastase (HLE, EC.3.4.21.37) is one of the most important proteases released from polymorphonuclear leukocytes [1] and it is involved in a number of connective tissue diseases, including emphysema [2], rheumatoid arthritis [3], cystic fibrosis [4] and chronic bronchitis [5]. Many research groups are presently studying the biological role of HLE with the aim of developing molecules able to stop HLE's negative effects [6–10]. A promising family of compounds gathers some cephalosporins characterized by the high reactivity of the β -lactam ring. A possible reaction path for such compounds has already been suggested (scheme 1 [11]); it assumes that the C(8) carbonyl atom places near to the Ser 195 hydroxy group, in a position suitable for a nucleophilic attack on the β -lactam ring. From this point the reaction might be different owing to the presence/absence of a good leaving group at C(7). If the substituent

at C(7) is a chlorine atom, the reaction contemplates loss of the 3' acetate, loss of HCl and a Michael addition involving the N(1) atom of the His 57 imidazole ring. When the substituent at C(7) is a poor leaving group, such as methoxy, the inhibitor might react as above or without the expulsion of the 3' acetate, forming a salt bridge with the same His 57. In this paper we examine 17 esters of the cephalosporins family with a different substituent group bound to the carboxy residue at position 2 (table I). Our aim is to clarify the enzyme–inhibitor recognition process, and to elucidate the structure–activity relationships. Accordingly, we perform a detailed analysis of the protein–substrate interactions for all the compounds, in order to understand the steric and electrostatic requirements for binding and activity. Our approach contemplates a series of molecular modeling calculations (molecular mechanics and dynamics, semiempirical MO) meant to provide a number of steric/electrostatic descriptors, followed by PLS and MLR-like analyses in which these descriptors are related to the available experimental data. The results of these analyses are summarized in a series of statistical models describing the nature of the enzyme–inhibitor interactions leading to activity.

*Correspondence and reprints



Scheme 1.

Methods

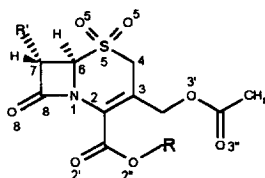
All molecular modeling calculations have been performed on a Silicon Graphics IRIS-4D 320 VGX computer system (IRIX version 4.0.5).

Molecular modeling

The structure of HLE bound with its inhibitor 2-MeO-succinyl-Ala-Ala-Pro-Val-chloromethyl acetone [13], deposited in the Brookhaven Data Bank, was used as initial geometry. Hydrogen positions were automatically fixed with an appropriate routine of the Insight (version 2.3.5) software [14]. The enzyme structure was optimized with the Tether option of the Discover (version 2.9.0) program package [15], using the Discover default force field. This calculation was repeated six times; in the first calculation a motion restraint (100 kcal/Å²) was applied to all non-hydrogen atoms; in the second the restraint was applied only to the protein backbone; in the third, fourth and fifth ones the restraint was limited only to the α -carbons, but its strength was lowered from 100 to 70 and finally to 30 kcal/Å². All calculations were stopped when the maximum absolute derivative (mad) value dropped below 0.1 kcal/molÅ. The sixth step was a fully unconstrained geometry optimization (mad < 0.01 kcal/molÅ). The optimized structure of

the enzyme was then submitted to simulated annealing molecular dynamics calculations (1 ps equilibration, 10 ps sampling at 900 K) in order to locate other low-energy structures; *cis-trans* inversion of peptide bonds in the high temperature dynamics was avoided by imposing a 5 kcal/rad² restraint to the O-C-N-H dihedral angles. Four frames were saved (one every 2.5 ps) and then cooled to 300 K in 5 ps with the TIMTMP criterion [15]. The resulting structures were then minimized (mad < 5 × 10⁻² kcal/molÅ).

Following this, compounds 1–17 were docked to the receptor binding site in HLE. The enzyme input geometry was the lowest energy conformation found by the previous phase of annealing. The geometry-optimized molecules have been positioned in the receptor active site by visually aligning the enzyme reacting residues with the designated anchor points on the substrates. Molecule 1 [11] was taken as reference compound because of the availability of the crystallographic coordinates of its covalently bonded complex with HLE [16]. The bad contacts between fitted molecules and HLE found by studying van der Waals and Connolly surfaces have been corrected by arbitrary modification of the side-chain dihedrals. The fitting was optimized by means of an automatic MonteCarlo-like conformational search. While a fully automated optimization was performed for molecule 1, compounds 2–17 were only superimposed on the cephalo-

Table I. Inhibitory activity against HLE and intermolecular non-bond interaction (INI) energies of cephalosporins.

Compound	R	IC_{50}^a (μM)	K_{obs}/I $M^{-1}s^{-1}$ ^b	INI (Kcal)		
				Van der Waals	Coulomb	Total
1 ^c	<i>t</i> -C ₄ H ₉	0.04	161 000	-45.93	-29.10	-75.03
2	<i>t</i> -C ₄ H ₉	1.0	16 000	-50.61	-23.56	-74.16
3	<i>i</i> -C ₃ H ₇	0.8		-51.81	-24.14	-75.96
4	C ₂ H ₅	0.2		-47.61	-23.15	-70.76
5	CH ₃	0.2	17 000	-48.71	-30.49	-79.20
6	CH ₂ CO ₂ C ₂ H ₅	0.1		-52.67	-20.71	-73.38
7	(CH ₂) ₂ CO ₂ CH ₃	0.5	7700	-52.09	-18.17	-70.26
8	(CH ₂) ₃ CO ₂ CH ₃	0.1	30 000	-52.97	-19.51	-72.48
9	CH ₂ C ₆ H ₅	0.07	d	-58.65	-21.03	-79.68
10	CH ₂ C ₆ H ₄ -4-OCH ₃	0.1	e	-62.94	-12.41	-75.35
11	CH ₂ C ₆ H ₄ -4- <i>t</i> -C ₄ H ₉	0.2	e	-72.53	-17.08	-89.62
12	CH ₂ C ₆ H ₄ -4-CO ₂ CH ₃	0.1	e	-70.89	-33.55	-104.44
13	CH ₂ C ₆ H ₄ -4-CO ₂ - <i>t</i> -C ₄ H ₉	0.4	d	-78.48	-38.89	-117.37
14	CH ₂ C ₆ H ₄ -3-CO ₂ CH ₃	0.04	e	-66.31	-25.13	-91.44
15	H	>50		-42.37	-24.07	-66.44
16	CH ₂ C ₆ H ₄ -4-COOH	0.4	57 000	-58.22	-23.31	-81.54
17	CH ₂ C ₆ H ₄ -3-COOH	0.4	62 000	-36.13	-18.23	-54.36

^aSee reference 12; ^bsee reference 11 for methodology; ^cin compound 1 R' = Cl; in compounds 2–17 R' = OCH₃; ^dno time-dependent inhibition was observed; ^ean inverse progress curve was observed.

sporanic ring of molecule 1 and then manually adjusted. All the enzyme–inhibitor complexes have been energy-minimized (mad < 10⁻² kcal/molÅ; 13 Å cutoff). The enzyme active site was also studied with molecular dynamics simulations (1 ps equilibration, 60 ps sampling). The enzyme–inhibitor complexes HLE-1 and HLE-2 were examined with dynamics as well for analysing how intermolecular interactions (notably, hydrogen bonds) can change. Twenty conformations were saved and minimized (mad < 1 kcal/molÅ, 13 Å cutoff). Furthermore, the lowest energy conformations of HLE-1 and HLE-2 were used for a relative free-energy calculation, using the finite difference thermodynamic (FDTI) algorithm [17]. This calculation was attempted only once because of our

limited computational resources. A number of other descriptors, especially electrostatic (partial charges, dipole moment, HOMO/LUMO) have been obtained by means of MOPAC [18] calculations performed on compounds 1–17. The PM3 hamiltonian and MOPAC default parameters were chosen. The data obtained from these calculations have been used only in the chemometric analysis.

Chemometrics

The correlations between steric, electrostatic and lipophilic properties of compounds 2–17, their INIs and their IC₅₀s were evaluated by means of statistical methods. Compound 1 was excluded because of the

diversity of its binding mode to the enzyme; compound **15** was also excluded because its activity level was ill-defined. Several cross-validated PLS and MLR-like analyses were performed with the Spectre [19] and Scan [20] software. Classical QSAR regression studies were discarded in favor of this alternative approach because we tried to achieve a model of greater predictive rather than descriptive ability. The steric, electrostatic and lipophilic descriptors of the inhibitors, many of which obtained during previous calculations, were correlated with the biological experimental data by means of a stepwise procedure. Only significant variables were retained in the final model. A non-obvious link between INI and IC_{50} was detected, which allowed the validation of the statistical model developed for the above-mentioned compounds. In the same way, we calculated three more models describing the correlations existing between the IC_{50} s and the steric/electrostatic descriptors of two other groups of inhibitors not included in the present modeling study.

Results and discussion

Molecular modeling

The conformation shown by the optimized enzyme was fairly close to that found in the solid state. The degree of similarity was measured by the root mean square (RMS) deviations of the calculated structure vs the X-ray one. The agreement between theoretical and experimental data is more than acceptable, even considering all non-hydrogen atoms (RMS = 2.36 Å). The agreement becomes better by considering only backbone atoms or just the α -carbons (RMS = 1.17 and 1.12 Å, respectively). This calculation shows the similarity of the gas-phase conformation vs the solid state one and that a simple minimization does not move the enzyme too far from the original coordinates. Accordingly, we felt that it was legitimate to use the optimized enzyme conformation for the subsequent docking calculations, in which our focus was of analysing the interaction network (in particular hydrogen bonds) that lead to a proper fit of the compounds into the enzyme active site. With respect to docking, experimental data relative to the activity of the 17 HLE inhibitors [11, 21] are compared in table I with the calculated binding energies. These data show that binding (INIs) and activity (IC_{50} s) are not correlated in an obvious manner. This makes sense, because the biological activity of the compounds depends on the formation of a covalent bond between the inhibitor and the enzyme, while docking and the INIs describe phenomena preceding such interactions. However, this point will be more

thoroughly examined later; here we will restrain our discussion on the features leading to optimal fit. For example, compound **15** (only a hydrogen atom attached to the carboxy group at position 2) does not fit appropriately into the enzyme active site and indeed shows poor binding and activity with respect to compounds bearing a bulkier residue, such as molecules **12** and **13** (table I). The examination of the binding fitness of all compounds allowed for a mapping of the enzyme active site; the areas relevant to the enzyme–inhibitor interactions are reported in table II. We should recall here that we defined a pocket within the active site by considering all the residues found within a 5 Å distance from the geometric centroid of each of the substituent groups. Thus, the S1 pocket hosts the group bound at position 7, while the S2 and S3 pockets those bound at position 2 and 3'. Table II also shows that in the active site the variable domain is (obviously) only the S2 pocket. From the examination of the residues involved in the S1 and S3 pockets it emerges that the S1 site is mainly hydrophobic, whereas the S3 site is less so. The variable S2 pocket also shows a balance between hydrophilic and hydrophobic residues, but it becomes increasingly hydrophobic when considering compounds with a large substituent group. Besides, it is interesting to note that in the series **2–17** the compounds with the best binding and activity are those able to hold the largest possible number of hydrophobic contacts (table II). So far we have only discussed the steric interactions, as regards the electrostatic ones, we have performed an analysis of all hydrogen bonds found for compounds **1–17**. The detailed results have been deposited as supplementary material (table A) and are available from the authors upon request; a shortened version of table A is presented in table III, where it is possible to appreciate the importance of Ser 195, Asp 102, Asp 194 and Ser 214. This was partially expected, as the so-called 'catalytic triad' of a large number of proteases does actually involve an Asp, a Ser and a His residue. Furthermore, in the homologous enzyme PPE (porcine pancreatic elastase) [11] the triad is formed just by Ser 195, Asp 102 and His 57. In our case, as we will see below, it is difficult to say which one of Asp 102 and Asp 194 is actually involved in the reaction. Instead, we would like to remark that His 57 does not look as if it is involved at this stage. The above-cited Ser 214 is directed towards the sulfone group, and so it is probably involved in the enzyme–inhibitor reciprocal recognition rather than in the biological response. From the available data, it may be thought that the inhibitor first enters the enzyme active site, then it is favorably positioned by some weak non-bond intermolecular interactions (mainly hydrophobic) and finally oriented by stronger hydrogen bonds. The observation about hydrogen bonds

Table II. Active site mapping^a.

<i>Enzyme pocket^b</i>	<i>Residues involved</i>	<i>Compounds</i>
S1	Val 190, Cys 191, Phe 192, Asp 194, Ser 195, Ala 213, Ser 214, Phe 215, Val 216	1–17
S2	Phe 41, His 57, Cys 58, Ser 195	1–5
S2	Cys 42, Ala 55, His 57, Cys 58, Phe 192, Ser 195	6–8
S2	Leu 33, Cys 42, Gly 43, Ser 54, Ala 55, Ala 56, His 57, Cys 58, Val 59, Val 104, Ser 195, Gly 193	9–14
S2	Cys 42, His 57, Cys 58, Phe 192, Gly 193, Asp 194, Ser 195	16, 17
S3	Ala 55, Ala 56, His 57, Cys 58, Tyr 94, Leu 99, Asp 102	1–17

^aThe residues defining the S2 pocket depend on the nature of the substituents hosted; ^bS1 hosts the substituent group bound at position 7; S2 and S3 host the substituent groups at positions 2 and 3, respectively.

Table III. Intermolecular hydrogen bonds (average plus esd) found for the 17 enzyme–inhibitor complexes (upper half) and during the dynamics of HLE-1 and HLE-2 (bottom half).

<i>Donor atom</i>	<i>Hydrogen atom</i>	<i>Acceptor atom</i>	<i>Incidence</i>	<i>H-A (Å)</i>	<i>D-A (Å)</i>	<i>Angle D-H-A (°)</i>
HLE						
Ser 214 OG	Ser 214 HG	O(5)	8	1.73 ± 0.13	2.68 ± 0.11	166.71 ± 6.95
Ser 195 OG	Ser 195 HG	O(7)	7	1.77 ± 0.08	2.70 ± 0.07	166.40 ± 14.05
Asp 102 OD2	Asp 102 HD2	O(3'')	6	1.78 ± 0.07	2.73 ± 0.07	169.19 ± 7.32
Ser 195 OG	Ser 195 HG	O(2'')	3	1.90 ± 0.07	2.79 ± 0.06	154.14 ± 6.56
Asp 102 OD2	Asp 102 HD2	O(3')	1	1.68	2.63	164.00
Cys 58 N	Cys 58 HN	O(2')	1	1.84	2.84	163.09
Gly 193 N	Gly 193 HN	O(8)	1	1.97	2.98	165.93
Ser 195 OG	Ser 195 HG	O(8)	1	1.65	2.60	164.24
HLE-1						
Gly 193 N	Gly 193 HN	O(2')	13	1.96 ± 0.05	2.94 ± 0.05	161.46 ± 11.05
Asp 194 OD2	Asp 194 HD2	O(8)	7	1.88 ± 0.06	2.83 ± 0.06	166.88 ± 2.91
Asp 194 N	Asp 194 HN	O(8)	3	1.91 ± 0.03	2.88 ± 0.02	154.16 ± 1.86
Asp 194 N	Asp 194 HN	O(2')	1	2.07	3.07	163.05
Ser 195 OG	Ser 195 HG	O(2'')	1	1.99	2.78	137.55
HLE-2						
Ser 214 OG	Ser 214 HG	O(5)	16	1.80 ± 0.08	2.72 ± 0.04	158.31 ± 11.14
Ser 195 OG	Ser 195 HG	O(8)	13	1.78 ± 0.08	2.70 ± 0.05	160.62 ± 9.32
Tyr 94 OH	Tyr 94 HH	O(3'')	5	1.83 ± 0.05	2.76 ± 0.05	165.50 ± 1.80
Asp 194 N	Asp 194 HN	O(8)	1	1.84	2.82	155.63
Gly 193 N	Gly 193 HN	O(8)	1	1.89	2.88	157.92
Ser 214 OG	Ser 214 HG	O(8)	1	1.71	2.66	153.77

For compounds **10** and **17** no hydrogen bonds were found.

involving the sulfone group can explain the higher reported activity for sulfone-bearing cephalosporins. The phase of His 57 attack (the one determining activity), instead, must occur afterwards, probably after a substrate-induced conformational change in the protein that must also bring the Asp 102/Asp 194 residues closer to the bound inhibitor. To elucidate this point, we carried out an examination of the hydrogen bonds found during the HLE-1 and HLE-2 complex dynamics (table III). The calculations on the HLE-1 show the importance of the interactions between the Gly 193 and Asp 194 residues by one side and the O(2') and O(8) atoms by the other one. The first anchors the O(2') of a carbonyl group and seems to help the attack of Ser 195, while the other one holds the β -lactam carbonyl oxygen and also seems to compete with Asp 102 in the catalytic triad; our calculations could not resolve this ambiguity. With respect to the HLE-2 complex, the most relevant interactions are those between Ser 214 and the sulfone group oxygens and those linking Ser 195 with the O(8). In addition, an H-bond between Tyr 94 and the O(3'') oxygen was also found. The first evidence reaffirms the importance of the sulfone group for activity. The other two hydrogen bonds indicate that the hydrogen bond network established by compounds with a methoxy

group in position 7 is different from the one shown by compound **1**, and that the ability to make new interactions could compensate for the activity loss determined by replacing the chlorine atom with a methoxy group. Similar considerations could also explain the low level of binding and activity shown by compound **15**, ie, a limited ability of establishing the correct interaction with the enzyme. In summary, these results indicate the substantial diversity of the binding mode of **1** and **2** and in particular that the higher degree of rotational freedom achieved by substituting the chlorine atom interacting at the S1 pocket (one bond to the cephalosporanic ring) with a methoxy group (two bonds) seems to be responsible for this change. With respect to the conformational analysis, the frames saved during the dynamics run did not differ appreciably from one another. The similarity was assessed by means of a RMS graph (fig 1) automatically generated by the Discover program. The dynamics of the active site (ie, all protein residues found within 5 Å from the geometric centroid of the inhibitor in the HLE-1 and HLE-2 complexes) showed instead a certain degree of freedom, as reported in table B in the supplementary material. Relevant features are described in table IV, which shows that the free enzyme is considerably more mobile than each of the complexes. Our calcula-

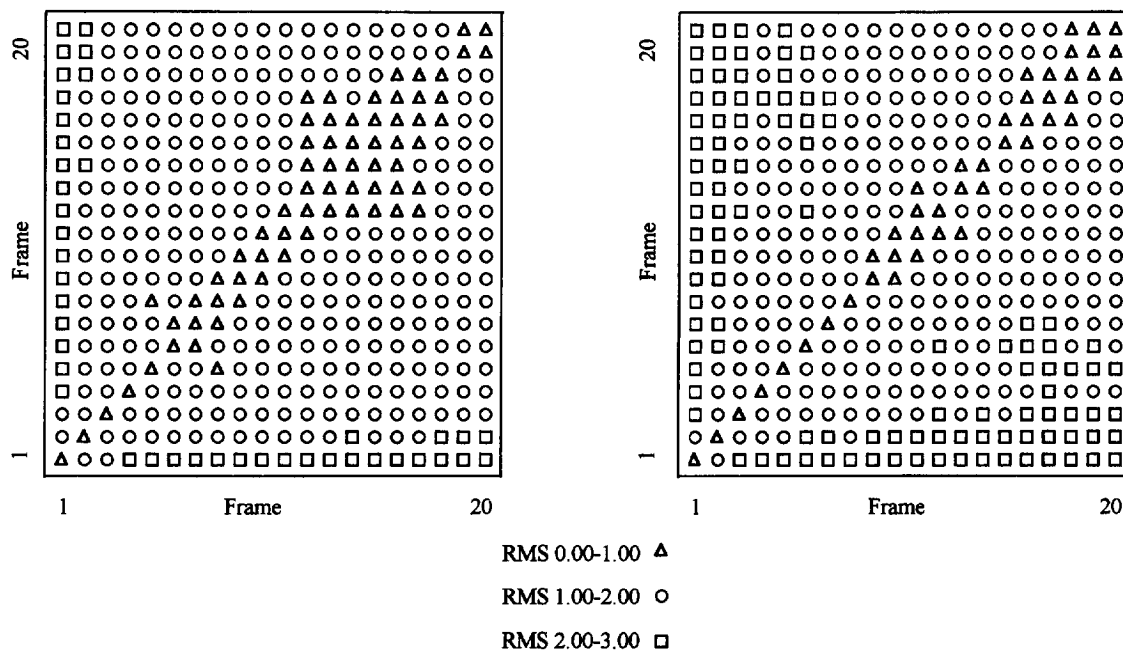


Fig 1. RMS graph for conformations stored during dynamics showing the high degree of similarity of saved frames. Left: HLE-1 complex; right: HLE-2 complex.

Table IV. Largest residue shifts for HLE-1 complex, HLE-2 complex and HLE (variation (Å) in parentheses).

<i>Residue</i>	<i>HLE-1</i>	<i>HLE-2</i>	<i>HLE</i>
Phe 41			5.40–10.10 (4.70)
Ala 55		5.25–7.99 (2.74)	
His 57		2.67–7.37 (4.70)	3.90–9.55 (5.65)
Cys 58		5.75–7.89 (2.14)	
Tyr 94	6.73–8.26 (1.53)	8.70–10.50 (1.80)	
Pro 96	8.70–10.8 (2.10)		9.00–14.50 (5.50)
Leu 99			6.80–12.10 (5.30)
Asp 102			4.63–9.31 (4.68)
Cys 191	7.08–8.41 (1.33)		
Phe 192	5.68–7.19 (1.51)	7.00–10.2 (3.20)	
Gly 193			
Asp 194	6.29–7.92 (1.63)		5.30–10.20 (4.90)
Ser 195	4.52–5.67 (1.15)		
Phe 215		6.37–8.89 (2.52)	

Conformational freedom measured as the variation of the distances between the geometric centroid of all the residues belonging to the active site and the geometric centroid of each of these residues during dynamics.

tions indicate that the active site without the inhibitor looks like an open pocket, while the presence of the inhibitor determines a series of conformational changes, including stiffening of the enzyme and active site closure. These changes allow Asp 102/Asp194 and His 57 to come closer to the inhibitor and also allow the enzymatic reaction to occur. Further insight in binding was also sought by means of thermodynamic calculations, in which we characterized the different behaviors of the HLE-1 and HLE-2 complexes. A molecular dynamics calculation was set in order to evaluate the ΔG for the transition HLE-1 \rightarrow HLE-2; ΔG was found positive by an amount of 29.63 ± 1.40 kcal/mol. This figure is in agreement with the different degree of activity shown by the two molecules, with the value of the second-order rate constant, (a measure of the rate of time-dependent inactivation; table I), and with the INIs. It is then ascertained also by theoretical methods that the modifications induced in the enzyme–inhibitor fit by the loss of the chlorine atom in position 7 are detrimental for binding and activity, and that the same activity level can be reached only if other interactions come into play. Furthermore, the agreement between calculated and experimental data suggests that the primary complex formation is correlated with biological activity (see below).

Chemometrics

At this point of the work, our idea was to rationalize all the observations made so far about recognition and binding, and we tried to clarify the correlations among the steric, electrostatic, lipophilic features of the inhibitors, their INIs and their IC_{50} s by means of statistical methods. Compounds **15** and **1** were not included in the chemometric calculations; the former because its IC_{50} value was ill-defined, and the latter because its reaction path [11] is probably different from the one followed by the other compounds. The analysis of the compounds **2–14** and **16, 17** has been independently performed by means of two chemometric software: Spectre [19] and Scan [20]. The calculations were performed twice because we wanted to compare results from both software and also because Spectre does not evaluate results by means of the classical statistical criteria (some embedded criteria are used [19]). Classical criteria (for example, F test) are also available within Scan. The Spectre models have been obtained with a cross-validated PLS analysis, and an MLR-like transformation of such models was then performed with Scan, according to literature data [22]. Cross-validation was achieved by the Spectre software by automatically selecting ten groups of three objects (leave-out three objects for ten

times), each of which was used in turn as the test set. The remaining 12 objects were used as the training set for each of the programmed (ten) iterations. This procedure leaves out each object at least twice, and it is performed by the program in an entirely unbiased, black-box manner. A stepwise automated procedure was used to select the most significant variables, a full listing of which is reported in table V together with predicted and observed values of IC_{50} . The a - g descriptors have been generated by means of MOPAC/PM3 calculations [18]. As recalled above, the two procedures gave strictly similar results; only the Scan models are reported here. The statistical criteria for the calculated models are listed in table VI, while the correlation matrices are reported in table C of the supplementary material. The first model developed was meant to identify the relationship between the IC_{50} and the descriptors of table V, and it is described by equation [1]:

$$\log 1/IC_{50} = -5.3 \times 10^{-6} - 95.70qC(3) + 1.82qN + 1.37qC(6) + 0.79qC(7) + 0.87Dist + 1.27Vol - 1.13HOMO + 0.57\Delta H + 0.52\mu - 1.48ClogP \quad [1]$$

It appears that electronic descriptors (six out of ten) are important, and it is interesting to note that all the charge parameters describe the β -lactam ring reactivity. This information is reinforced by the dipolar moment term (μ), that underlines the opportunity of a charge separation on the molecule, and also by the HOMO descriptor. The most important contributions to the HOMO, in fact, come from N(1) and C(3), two atoms involved in the resonating double-bond system which is fundamental for the reactivity of the β -lactam ring as well as of the substituent group at position 3. Besides, there are also clear indications about the weight of steric (Dist, Vol) and lipophilic (ClogP) parameters. The coefficients shown by the two steric descriptors predict that activity will increase with the dimensions of the substituent bound at position 2, both in terms of linear extent and volume. In fact, Dist is the distance between the geometric centroid of the CONH β -lactam group and the geometric centroid of the substituent in position 2, and Vol is the volume of the same substituent. However, we note that the Vol descriptor is not truly discriminant by itself (see also equation [4] below). What is apparently confusing (according to previous observations) is that the lipophilic term points to the opportunity of having a hydrophilic rather than a hydrophobic molecule. We must remember, however, that the ClogP is a descriptor carrying information about the molecule as a whole, and not only about the substituent at position 2. The ΔH term, finally, suggests that unstable molecules enhance activity; this information, in conjunction with the HOMO and charge terms, indicates that

enhanced reactivity (especially, but not exclusively, at the β -lactam ring) prompts the progress of the catalytic reaction. Before using this model for predicting the IC_{50} s of the compounds in our set, we decided to test its reliability, and it is at this point that we return to the correlation between INI and IC_{50} . As discussed above, biological data do not seem to be related to INIs; however, using the same procedure as for model [1], we have been able to see that there is a non-obvious link between INI and IC_{50} . This relationship is shown in equation [2]:

$$\log 1/IC_{50} = -3.8 \times 10^{-6} + 0.72qO(8) - 0.58qRes + 0.71Dist - 0.41HOMO + 0.79\mu - 0.85INI \quad [2]$$

Examination of models [1] and [2] and table I suggests a cumulative interpretation of the first two equations. The analysis shows that a favorable (more negative) INI value leads to lowering of the IC_{50} ; a good recognition is then essential for better activity. The μ term indicates the importance of the sulphone group (as in equation [1]), while the Dist term, in accordance with the INI (see table I), indicates the opportunity of an appropriate hydrophobic interaction with the S2 pocket (see also table II). The remaining electronic parameters ($qO(8)$, $qRes$, HOMO) in equation [2] seem to be important in a second phase, i.e., that of the enzyme nucleophilic attack on the inhibitor. In synthesis, the compounds undergo a shape recognition (INI, Dist), where the hydrophobic nature of the substituent at position 2 is relevant and electronic features are not determinant. The fit is then driven by overall properties (μ , ClogP), and finally efficacy of action is determined by presenting the appropriate reactivity (atomic charges, HOMO) towards the incoming reacting group (Asp 102/Asp 194 and His 57) of the enzyme. Another important conclusion is that the parameters relevant to calculate IC_{50} s in model [1] are also relevant to relate IC_{50} s to INIs. According to this, we have found it appropriate to test the reliability of model [1] by looking at our ability of correctly predicting the INIs. Equation [3] shows the relation:

$$INI(kcal) = -4.8 \times 10^{-6} - 4.69qN(1) - 2.03qC(2) + 0.22qS(5) - 2.86qC(6) - 1.55qC(7) + 0.43qO(8) - 1.73qRes + 0.50Vol + 0.29HOMO + 1.23\Delta H + 0.68\mu \quad [3]$$

and the plot of INI_{obsd} vs INI_{calc} is shown in figure 2. Thus, INI is mainly (nine terms out of eleven) a function of electronic parameters. By looking at equations [2] and [3] together, it is possible to say that the INI term in equation [2] collects all the contribution to activity of a number of electronic terms inherent to the cephalosporin ring. Look, for example, at the $qS(5)$ term that informs about the better recorded activity of

Table V. Parameters used to derive equations [1] and [2].

Compound	$qN(1)^a$	$qC(2)$	$qC(3)$	$qC(4)$	$qS(5)$	$qC(6)$	$qC(7)$	$qC(8)$	$qO(8)$	$qRes^b$	$Dist(\text{\AA})^c$	$Vol(\text{\AA}^3)^d$	$HOMO(eV)^e$	$LUMO(eV)^e$	$\Delta H(Kcal)^f$	$\mu(D)^g$	$ClogP^h$	IC_{50}^{obsd}	$IC_{50}^{calc,i}$
2	0.012	-0.080	-0.129	-0.491	2.189	-0.542	0.068	0.285	-0.251	0.039	4.13	142.6	-9.5621	-0.6853	-271.13	3.849	-0.98	1.00	1.16
3	0.014	-0.076	-0.131	-0.493	2.194	-0.548	0.071	0.284	-0.249	0.032	4.20	117.9	-9.5634	-0.6741	-267.57	3.703	-1.38	0.80	0.75
4	0.012	-0.075	-0.133	-0.495	2.195	-0.543	0.070	0.282	-0.247	0.029	4.66	93.2	-9.6000	-0.7048	-260.75	3.893	-1.68	0.20	0.19
5	0.010	-0.070	-0.135	-0.499	2.195	-0.546	0.074	0.281	-0.248	0.027	4.26	68.5	-9.6039	-0.7389	-255.82	3.759	-2.21	0.20	0.18
6	0.008	-0.074	-0.127	-0.497	2.196	-0.541	0.074	0.283	-0.249	0.016	5.08	154.8	-9.6646	-0.8170	-337.71	3.492	-1.98	0.10	0.12
7	0.013	-0.083	-0.118	-0.499	2.198	-0.545	0.072	0.280	-0.256	0.024	4.48	154.8	-9.6362	-0.7556	-339.45	4.060	-2.12	0.50	0.46
8	0.010	-0.072	-0.133	-0.496	2.193	-0.545	0.070	0.281	-0.248	0.023	4.98	179.5	-9.6351	-0.7547	-346.05	3.70	-2.20	0.10	0.10
9	0.015	-0.074	-0.134	-0.492	2.196	-0.547	0.072	0.280	-0.247	0.032	5.15	171.3	-9.5740	-0.7050	-226.75	3.811	-0.51	0.07	0.07
10	0.010	-0.069	-0.140	-0.488	2.192	-0.543	0.074	0.286	-0.250	0.028	7.63	205.2	-9.3073	-0.6855	-264.74	4.567	-0.59	0.10	0.10
11	0.005	-0.070	-0.133	-0.492	2.188	-0.538	0.072	0.287	-0.254	0.045	7.99	270.1	-9.5111	-0.6848	-250.84	4.142	1.31	0.20	0.20
12	0.013	-0.078	-0.125	-0.494	2.195	-0.545	0.074	0.282	-0.250	0.026	5.85	232.9	-9.6676	-0.8392	-307.58	2.442	-0.54	0.10	0.10
13	-0.005	-0.060	-0.142	-0.486	2.183	-0.525	0.068	0.295	-0.262	0.031	6.61	307.0	-9.5702	-0.7666	-319.45	2.154	0.70	0.40	0.38
14	0.014	-0.077	-0.126	-0.493	2.192	-0.542	0.072	0.282	-0.253	0.021	6.15	232.9	-9.5926	-0.7217	-308.86	4.457	-0.54	0.04	0.04
16	0.011	-0.077	-0.128	-0.492	2.193	-0.544	0.071	0.283	-0.249	0.024	5.51	208.2	-9.6758	-0.9130	-315.76	1.873	-0.77	0.40	0.37
17	-0.004	-0.066	-0.126	-0.499	2.208	-0.527	0.083	0.280	-0.250	0.006	3.35	208.2	-9.5466	-0.6515	-312.04	4.444	-0.77	0.40	0.39

^aPartial charges of the atoms of the cephalosporan ring; ^btotal charge of the substituent chain in position 2; ^cdistance between the geometric centroid of the CONH β -lactam group and the geometric centroid of the substituent in position 2; ^dsubstituent volume [23]; ^eHOMO and LUMO energies; ^fheat of formation; ^gdipole moment; ^hcalculated octanol/water partition coefficient [24]; ⁱcalculated using equation [1].

Table VI. Statistical criteria of the PLS models [1]–[6] and of their MLR-like transformations.

Criterion	(1)	(2)	(3)	(4)	(5)	(6)
PLS models [1]–[6]						
Dim ^a	10	5	11	9	9	6
PRESS ^b	1.14	3.67	0.08	3.83	0.10	0.13
PP(%) ^c	95.93	86.88	99.70	91.74	99.39	99.40
P–r ^d	0.979	0.932	0.998	0.958	0.997	0.997
<i>r</i>	0.995	0.943	0.999	0.963	0.999	0.999
<i>r</i> ² %	99.00	88.92	99.80	92.73	99.80	99.80
MLR-like transformations of PLS models [1]–[6]						
<i>n</i>	10	6	11	9	9	6
<i>r</i>	0.995	0.944	0.999	0.963	0.999	0.999
<i>s</i>	0.129	1.515	0.001	1.509	0.013	0.011
<i>F</i>	42.90	10.99	2708	17.23	457.1	419.8
<i>p</i>	0.0012	0.0029	0.0001	0.0001	0.0001	0.0002

^aDim = dimensionality of PLS model; ^bPRESS = predicted residual error sums of squares; ^cPP(%) = prediction power [22]. This entity is defined by the following expression:

$$PP\% = 100 \cdot \left(\frac{\sum_p \sum_j (\hat{y}_{pj} - y_{pj})^2}{\sum_p \sum_j (y_{pj} - \bar{y}_{pj})^2} \right)$$

where \hat{y}_{pj} are the calculated values from test set, y_{pj} are the correspondent observed values, \bar{y}_{pj} is the average value for all observed y , p is the number of objects in test-set and j is total number of responses for all objects; ^dP–r = predictive correlation coefficient.

sulfone derivatives. The weight of the electronic terms resumed by INI has then to be more finely balanced by some specific interaction taking place at the substituent groups (*q*Res). The problem of reliability was also raised for model [3], so we calculated an INI value for a cephalosporin derivative not belonging to our series, but modeled from scratch only for this purpose. This number (–78.99 kcal) was compared with the calculated docking value of –80.40 kcal. Since the two figures were in good agreement, we were confident about the predictive ability of model [3], and proceeding from the above discussion, we were also confident about the appropriateness of model [1]. Another interesting remark is that the parameters used to calculate the INIs have been generated with the MOPAC/PM3 program on the isolated (gas-phase) molecule. They turned out to be very descriptive, so it was decided to use them for studying other groups of cephalosporin inhibitors (table VII) bearing

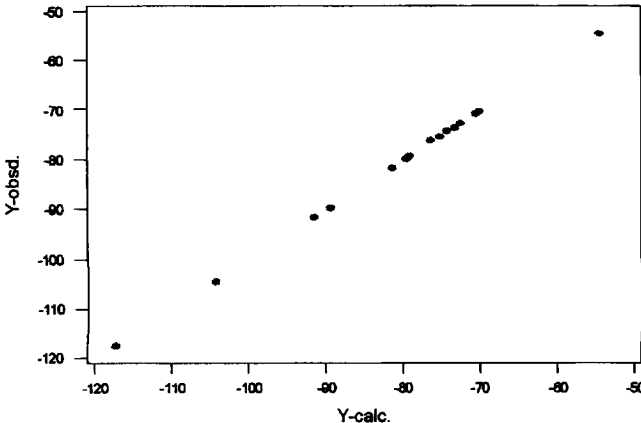
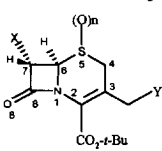

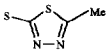


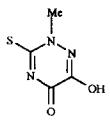


Fig 2. Plot of INI_{obsd} vs INI_{calc} (equation [3]).

Table VII. Compounds described by equations [4] and [5].


Compound	X	n	Y
<i>Equation [4]^a</i>			
4a	MeO	0	OC(O)CH ₃
5a	MeO	2	OC(O)CH ₃
5b	EtO	2	OC(O)CH ₃
5c	<i>n</i> -PrO	2	OC(O)CH ₃
5e	<i>i</i> -PrO	2	OC(O)CH ₃
5f	<i>s</i> -BuO	2	OC(O)CH ₃
5h	PhO	2	OC(O)CH ₃
6a	MeO	1(β)	OC(O)CH ₃
6b	MeO	1(β)	OC(O)CH ₃
11a	HCOO	0	OC(O)CH ₃
11b	MeCOO	0	OC(O)CH ₃
12a	HCOO	2	OC(O)CH ₃
12b	MeCOO	2	OC(O)CH ₃
12g	MeNHCOO	2	OC(O)CH ₃
14a	Et	2	OC(O)CH ₃
14b	<i>n</i> -Pr	2	OC(O)CH ₃
14c	<i>n</i> -Bu	2	OC(O)CH ₃
15	Cl	0	OC(O)CH ₃
16	Cl	2	OC(O)CH ₃
17	F	0	OC(O)CH ₃
18	F	2	OC(O)CH ₃
26b	HCONH	2	OC(O)CH ₃
<i>Equation [5]^b</i>			
1	MeO	2	OAc
6	MeO	2	Cl
7a	MeO	2	
7b	MeO	2	
7c	MeO	2	
7d	MeO	2	
7e	MeO	2	S(C-S)OEt
7f	MeO	2	
7h	MeO	2	SPh
7i	MeO	2	SOPh
7j	MeO	2	SO ₂ Ph
9c	MeO	2	PhCH ₂ CH(COOH)NHCOO
10	MeO	2	MeO
12	MeO	2	H

^aThe compound numbering used here is the same as that adopted in reference [11]; ^bthe compound numbering used here is the same as that adopted in reference [25].

different substituent groups in position 7 [11] or in position 3' [25] (models [4]–[6]). The resulting equations are reported below:

$$\log 1/IC_{50} = 1.85 \times 10^{-6} + 0.34qC(2) + 1.82qS(5) - 0.75qC(7) + 0.83\Delta H - 0.15HOMO + 0.32B1 - 0.68Vol + 0.14Dist - 1.53qRes \quad [4]$$

$$\log 1/IC_{50} = 3.96 \times 10^{-6} - 0.74qC(6) - 0.65qC(7) - 1.30qO(8) + 1.17Dist + 0.38Vol - 0.51HOMO + 1.04LUMO + 0.32\Delta H - 0.32\mu \quad [5]$$

$$\log 1/IC_{50} = 2.57 \times 10^{-6} + 2.01qN(1) - 0.55qC(7) + 1.77qC(8) - 0.53qRes + 0.87LUMO + 0.33\sigma_1 \quad [6]$$

The main features of model [1] have been in this case slightly modified; for example, in model [4] Sterimol parameters (L, B1, B2, B3, B4) [26] and a parameter *n* (number of oxygens atom bound to S(5)) have been added, while in model [6] also σ_1 [27, 28] was present. Only B1 in equation [4] and σ_1 in equation [6] were retained in the final models. Full details of data used in models [4]–[6] are reported in tables D and E of the supplementary material. A comparison of the last three models with equation [1] shows that the ΔH , HOMO, Vol and $qC(7)$ descriptors are always preserved and μ almost so. This occurrence gives us a good feeling about the transferability of the main features of our initial model to other classes of related compounds. Coming into the last three equations, model [4] highlights again the importance of the $qS(5)$ variable, while the combined variation of the steric parameters B1 and Vol gives further shape insight. In particular, it explains why thin phenol rings can accommodate quite nicely into the rather puckered S1 pocket and branched aliphatic chains find it difficult. For the set of compounds described in [25], model [5] (then modified in model [6]) shows that electronic parameters are dominant, and in particular that IC_{50} must grow with σ_1 . This is in agreement with the evidence that electron-withdrawing substituents (at position 3') increase β -lactam ring reactivity [29]. Since the last three equations are not supported by docking calculations, we do not make further comments on them at this point.

Conclusions

In this paper we have examined 17 putative HLE inhibitors in order to clarify the requirements for binding and activity. The available experimental data, together with the results of molecular modeling studies, indicate that the chlorine atom, despite the hydrophobic nature of the S1 pocket, is a more appropriate substituent at position 7 than the methoxy group. Compounds **9–14** show, in fact, a more favorable INI value due to the appropriate interaction of the

substituent bound at position 2 of these molecules with the hydrophobic residues defining the S2 pocket. This fact leads to a more efficient recognition and orientation of the compounds towards the enzyme reacting residues, and allows us to recover the activity level that was lost by changing the chlorine with the methoxy group. In the same way, it also explains the poor activity level shown by compound **15**, due to the loss of the favorable interaction made by the chlorine atom at position 7 that is not counterbalanced by new interactions made by the substituent group at position 2. This observation was reconfirmed by the docking studies and later rationalized by the statistical models. The docking and the ΔG calculations allowed us to clarify the non-negligible variation in the number of residues involved in binding (the variable extent of the S2 pocket) and the importance of subsidiary hydrogen bonds (especially those made by Ser 214, Ser 195, Asp 102, Asp 194, Gly 193 and Tyr 94) that co-operate in the recognition/binding process as well as in the catalytic reaction. ΔG studies, in particular, give an indication that the primary complex formation is correlated with biological activity; the dynamics of such a relationship was clarified by the statistical analysis. Furthermore, docking and molecular dynamics experiments suggest that the recognition/binding process precedes some kind of conformational change (ie, enzyme stiffening, closure of active site onto the substrate, Asp 102/Asp 194 and His 57 coming effectively into play). It is the ease with which the inhibitor suits this change that leads to different levels of biological activity. The existence of a two-phase process in which shape recognition and orientation driven by overall molecular properties (see docking and models [1]–[3]) is followed by a finely tuned electronic interaction (appropriate location of atomic charges and extent of reacting orbitals matching the conformational changes described above, see models [1]–[6]) is reconfirmed also by chemometric studies. The first significant result of these studies was the definition of the non-obvious link existing between the INI and the IC_{50} , that we used for a validation of equations [1] and [2] with equation [3]. Then, the models [1] and [4]–[6] clearly indicate the key role played by electronic parameters in order to achieve the biological effect. This is especially true in model [6], where only electronic parameters remain significant. However, models [4]–[6] are not supported by docking studies; then, the lack of shape/steric parameters in the equations was not surprising. The details of the models indicate that hydrophobic factors are important for the filling of the enzyme pocket S2 (models [1] and [2]), an issue that had been hypothesized by looking at the residues involved. Model [4] defines the importance of the

steric requirements of the substituent groups at position 7 (as well as the physical extension of the S1 pocket). From model [6] we understand the importance of the electron-withdrawing effect of the substituent group bound at position 3' of the cephalosporanic ring. Some small, but not trivial insights have thus been achieved, despite our inability to reduce the numbers of parameters involved in the models to a more manageable size. This could be an indication that activity is not strongly linked with a specific feature in the molecule, but rather to a tiny balance between steric/shape requirements and electron density on the molecule. Thus, the picture is still incomplete, and further work must be carried out, for example, in the field of the carbonylic β -lactam vibration frequencies in relation to the different features of the substituent groups close to it. This will probably help to elucidate the way in which substituents at the β -lactam ring influence reactivity.

References

- Dewald B, Rinder-Ludwig R, Bretz U, Baggiolini M (1975) *J Exp Med* 141, 709–723
- Morrison HM, Kramps JA, Afford SC, Burnett D, Dijkman JH, Stockley RA (1987) *Clin Sci* 73, 19–28
- Velvar M (1981) *Rheumatol Int* 1, 121–130
- Jackson AH, Hill SL, Afford SC, Stockley RA (1984) *Eur J Respir Dis* 65, 114–121
- Snider GL (1987) *Drug Dev Res* 10, 235–253
- Veale CA, Damewood JR, Steelman GB, Bryant C, Gomes B, Williams J, (1995) *J Med Chem* 38, 86–97
- Veale CA, Bernstein PR, Bryant C et al (1995) *J Med Chem* 38, 98–108
- Bernstein PR, Gomes BC, Kosmider BJ, Vacek EP, Williams JC (1995) *J Med Chem* 38, 212–215
- Vergely I, Boggetto N, Okochi V et al (1995) *Eur J Med Chem* 30, 199–208
- Subramanyan C, Bell MR, Ghose AK et al (1995) *Bioorg Med Chem Lett* 5, 325–330
- Doherty JB, Ashe BM, Barker PL et al (1990) *J Med Chem* 33, 2513–2521
- Doherty JB, Ashe BM, Argenbright LW et al (1986) *Nature* 322, 192–194
- Wei AZ, Mayr I, Bode W (1984) *FEBS Lett* 234, 367–373
- Insight II User Guide, version 2.3.5, Biosym Technologies, San Diego CA, 1994
- Discover User Guide, version 2.9.5, Biosym Technologies, San Diego CA, 1994
- Navia MA, Springer JP, Lin T et al (1987) *Nature* 327, 79–82
- Mezei M (1987) *J Chem Phys* 86, 7084–7088
- Stewart JJP version 6.0, 1990, Mopac Manual, A General Molecular Orbital Package, 6th edn
- Marsili M, Marengo E, Saller H (1988) *Anal Chim Acta* 210, 33–50
- Scan Software for Chemometric Analysis, Release 1.1, Minitab Inc, 3081 Enterprise Drive State College, PA, 16801-3008, USA, 1995
- Finke PE, Ashe BM, Knight WB et al (1990) *J Med Chem* 33, 2522–2528
- Katsumi H, Yoshida M, Kikuzono Y, Takayama C, Marsili M (1993) In: *Computer-Aided Innovation of New Materials II* (Doyama M, Kihara J, Tanaka M, Yamamoto R, eds) Elsevier Science, Amsterdam, 887–890
- Immirzi A, Perini B (1977) *Acta Cryst* A33, 216–218
- ClogP for Windows, version 1.0.0, BioByte Corp, Claremont CA, 1995
- Shah SK, Brause KA, Chandler GO et al (1990) *J Med Chem* 33, 2529–2535
- Hansch C, Leo A (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, New York
- Charton M (1987) *Prog Phys Org Chem* 16, 287–298
- Boyd DB (1984) *J Med Chem* 27, 63–66
- Nishikawa J, Tori K (1984) *J Med Chem* 27, 1657–1663