Comparative Binding Energy (COMBINE) Analysis of HumanNeutrophil Elastase Inhibition by Pyridone-containing Trifluoromethylketones

Carmen Cuevas,[†] Manuel Pastor,[#] Carlos Pérez[†] and Federico Gago[*]

Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain

[†] *Present address: Pharma Mar S.A., Cantoblanco, 28760 Tres Cantos, Madrid, Spain*

[#] *Present address: Departament de Ciències Experimentals i de la Salut, Universitat Pompeu, Fabra, Dr. Aiguader 80, E-08003 Barcelona, Spain*

Abstract: The complexes of human neutrophil elastase with a series of 40 N3-substituted trifluoromethylketone-based pyridone inhibitors have been modelled. The series spans three orders of magnitude in inhibition constants despite the fact that it was originally developed in an attempt to improve the oral activity of a lead compound. Ligand-receptor interaction energies calculated using molecular mechanics did not correlate well with the experimental activities. A good correlation with activity was found, however, when a COMBINE analysis of the same data was carried out, which allowed a quantitative interpretation of the modelled complexes. The essence of this method is to partition the ligand-receptor interaction energies into individual residue-based van der Waals and electrostatic contributions, and to subject the resulting energy matrix to partial least squares analysis. Incorporation of two additional descriptors representing the electrostatic energy contributions to the partial desolvation of both the receptor and the ligands improved the QSAR model, as did the replacement of the distance-dependent electrostatic contributions with solvent-screened electrostatic interactions calculated by numerically solving the Poisson-Boltzmann equation. The model was validated both internally (cross-validation) and externally, using a set of twelve 6-phenyl-pyridopyrimidine analogs. The analysis reveals the subtle interplay of binding forces which occurs within the enzyme active site and provides objective information that can be interpreted in the light of the receptor structure. This information, gained from a series of real compounds, can be easily translated into 3D real or virtual database queries in the search for more active derivatives.

Keywords: Quantitative Structure-Activity Relationship, COMBINE Analysis, molecular mechanics, molecular modeling, structure-based drug design, virtual screening

INTRODUCTION

The exponentially increasing availability of three-dimensional structures for many macromolecular drug targets and rapid advances in combinatorial chemistry, both *in vitro* and *in silico*, provide fertile ground for the development and testing of modern computational methods of quantitative structure-activity relationships (QSAR) and structure-based drug design [1,2]. The automatic evaluation of large collections of compounds (the so-called chemical libraries) by computer programs has come to be known as virtual screening [3] and the field is developing very rapidly. The main goal of these computational screening methodologies is to limit the amount of synthetic effort that is wasted on inactive or weakly active molecules. Former limitations [4] such as insufficient speed, inadequate treatment of ligand and receptor flexibility, inappropriate docking methods, and inaccurate assessment of binding affinities are being gradually overcome [5-8].

*Address correspondence to this author at the Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain; Telephone: +34-918 854 514; Fax: +34-918 854 591; e-mail: federico.gago@uah.es

The opportunities for the successful contribution of computational methods to the ligand design process relies on an optimized management of available chemical information [9]. In principle, more success is likely in those cases where structural information for the target is also at hand. When this is not the case, a structure-affinity relationship can still be derived from the information inherent in a series of compounds of graded affinity towards a given receptor. To this end, a plethora of methods of similarity analysis, pharmacophore searching, and 3D-QSAR [10,11] exist that use both physico-chemical descriptors (e.g. the well-known triad of hydrophobic, electronic and steric properties of the substituents) and molecular interaction fields computed for different probes representative of the functional groups expected to be present in the binding site [12]. This variety of approaches reflects the enormous range of possibilities that can guide a given medicinal chemistry project.

When the structure of the macromolecular target is known, the design of the combinatorial library can be customized to suit the geometry of the binding site [13]. A focussed combinatorial library of small organic molecules typically consists of a common central template or scaffold

1386-2073/01 $28.00+.00

onto which different sets of substituents are attached at one or several positions [14]. One remaining theoretical hurdle, however, is what scoring function to employ for both docking and ranking the ligands. Scoring functions can be tailored to a specific problem, in which case a simple count of hydrogen bonds and the number of hydrophobic contacts might suffice [15], or can be much more detailed and serve a more general purpose [6-8,16-18].

When the structure of the receptor has been solved in a complex with one or several ligands, there are opportunities for molecular modeling techniques to provide structural information about a whole set of complexes for further analysis. Despite progress in alternative scoring methods [6,7], molecular mechanics force fields representing van der Waals and Coulombic interactions on the basis of empirical energy functions arguably remain the more widely used tools to calculate ligand-receptor interaction energies in these complexes. Although differences in binding enthalpies thus calculated occasionally reflect fairly well the differences in experimental free energies [19-21], this is not a general trend. Among the likely causes, we may well consider inappropriate modelling of the screening effect of the solvent in the computation of the electrostatic interactions and the fact that desolvation and entropic contributions are almost systematically ignored. If solvent molecules are explicitly included, free energy perturbation methods [22] can generally yield good agreement between experimental and calculated free energy differences for two ligands binding to a common protein [23]. Nevertheless, the accuracy of this approach is related to the very long times necessary for computation, which make it impractical for the comparative study of large series of receptor-ligand complexes. A slightly less costly alternative is the linear interaction energy method [24], which is based on linear response theory, but no formula appears to be universal for these calculations [25].

The molecular recognition problem is intrinsically complicated because the free energy difference that drives the association results from a subtle interplay of binding forces that take place within the protein binding site, usually in the face of competition with water molecules. Molecular modelers are also well aware that the modelled complexes for a full series of ligands might not be sufficiently accurate and that the final geometries are highly dependent on both an adequate parametrization and the energy minimization protocol employed in their refinement [21]. Besides, if just one single conformation is used to represent the bound and unbound states of the binding partners, the entropic contributions, which are rarely constant for a series of ligands, are invariably omitted unless implicitly included in some other term.

Some years ago, Ortiz et al. made a proposal to simplify this problem [26]. They showed that it was possible to derive a QSAR model, even if no initial correlation between calculated ligand-receptor energies and activities is apparent, through adequate weighting of selected pairwise interactions between the ligand (or parts of the ligand) and residues of the protein. The approach was termed Comparative Binding Energy (COMBINE) analysis, and since then, several successful applications on different systems have been reported in the literature [20,27,28]. What is needed to perform this analysis is a set of ligand-receptor complexes and a molecular mechanics force field to compute ligand-receptor interaction energies. Then, rather than using a global term to characterize the interaction, the intermolecular binding energy is partitioned on the basis of individual protein residues and further decomposed into van der Waals and electrostatic contributions. The essence of the COMBINE method is to project the resulting matrix of energy terms on to a small number of orthogonal "latent variables" (or principal components, PC) using partial least squares (PLS) analysis. After this projection, the original energy terms are given weights, according to their importance in the model, in the form of PLS pseudo-coefficients. Operating this way, it is expected that most of the "noise" present in the dataset (arising, for example, from errors in the modeling work or in the energetic description of the complexes) can be filtered out by means of the subsequent chemometric analysis. The extensive decomposition allows those components that are predictive of binding free energy (i.e., the real “signal”) to be detected, and these might implicitly represent other physically important interactions or even entropic terms. Therefore, the *COMBINE* acronym [29] refers to combinations in terms of both data and techniques:

- data calculated on the ligand-receptor structures are combined with the measured activities for the series of ligands;

- molecular mechanics and multivariate statistics are combined for the analysis.

As a new test of the COMBINE methodology, we present a study on human neutrophil elastase (HNE) inhibitors originally developed at Zeneca (now Astra-Zeneca). The example can be taken as representative of a typical medicinal chemistry program aimed at improving the oral activity of a lead compound by introduction of different substituents onto a common chemical scaffold [30]. HNE is a highly basic serine protease present in the azurophile granule of human polymorphonuclear neutrophils bound to a matrix of the glycosaminoglycan chondroitin-4-sulfate. HNE is essential for phagocytosis and defense against infection by invading microorganisms. It also clears elastin, a flexible protein highly cross-linked with pyridinium-containing amino acids (desmosine and isodesmosine) and abundant in the lungs, arteries, skin, and ligaments. HNE activity is tightly regulated by a number of endogenous inhibitors (e.g. α_1-antitrypsin and α_2-macroglobulin) but imbalance can result in abnormal degradation of connective tissue structural proteins. For this reason it is thought that HNE can play a causal or contributory role in several disease states such as pulmonary emphysema, cystic fibrosis, rheumatoid arthritis, chronic bronchitis or acute respiratory distress syndrome [31]. These pathological implications make HNE an important target for drug development [32], and the chronicity of these conditions makes oral activity a highly desirable goal for HNE inhibitors.

From the structural viewpoint, HNE contains four disulfide bridges and shows an extended substrate binding site, composed of a set of pockets (S1-S5) along the crevice formed by a domain interface on the acyl side of the catalytic

triad (Asp-102, His-57, and Ser-195) [33]. These additional target locations are occupied by the side chains of amino acids in peptide inhibitors and are also available for exploitation by non-peptidic inhibitors. This communication between remote and proximal subsites is important for drug design, as is the fact that some of the remote sites can be occupied in the crystal structures by residues from a symmetry-related neighbour molecule [21,33]. In common with other serine proteases, only minor conformational changes are observed crystallographically in elastases upon binding of substrates or inhibitors, which represents an obvious advantage for modelling purposes. This is particularly important in the present case because, in contrast with porcine pancreatic elastase (PPE), HNE either alone or in complex with small ligands does not yield crystals suitable for x-ray crystallography [33]. Fortunately, the structure of HNE has been solved as a complex with turkey ovomucoid inhibitor [34] and also with two specific chloromethyl ketone irreversible inhibitors [35,36].

A regular spacing of backbone hydrogen bonding donor and acceptor groups on the acyl side of the scissile bond allows peptidic inhibitors to form an antiparallel β-sheet structure with the peptide backbone of HNE [33]. Intense efforts to overcome the unsuitable bioavailability properties of most, but not all [37], peptidic compounds, have led to the development of a large number of non-peptidic inhibitors [38]. For example, pyridone trifluoromethylketone-based inhibitors interact with the enzyme in a substrate-like fashion occupying the S3-S1 subsites and retaining the ability to participate in many of the key binding interactions with the enzyme as observed for peptidic inhibitors. Thus, the 3-amino group and the carbonyl group of the pyridone ring are able to form antiparallel hydrogen bonding interactions with Val-216 of the enzyme (Fig. **1**). Incorporation of a phenyl substituent in the 6-position of the pyridine ring increases the affinity by occupying the S2 subsite [30]. Additional substitution at the 3-position can modulate both the affinity and the oral activity, and it was precisely when efforts were made to improve the bioavailability of these compounds that an inhibitor with a subnanomolar inhibition constant (K_i) was unexpectedly discovered (K13, K_i = 0.7 nM). The 40 molecules in the reported set (Table 1) encompass a wide range of physico-chemical parameters and span three orders of magnitude in inhibition constants.

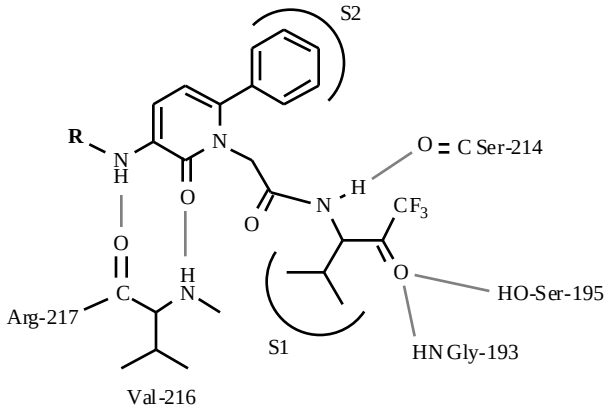


Fig. (1). Scheme of the binding mode of pyridone-containing trifluoromethyl ketones to HNE.

METHODOLOGY

a) Model Building of the Complexes and Parametrization of the Inhibitors

Coordinates for the whole set of inhibitors in their pre-covalent complexes with HNE were modelled on the crystallographic structures of HNE complexed with methoxysuccinyl-Ala-Ala-Pro-Ala chloromethyl ketone (PDB entry 1HNE) [36] and the complex between PPE and inhibitor **D13** solved at 1.8 Å resolution (PDB entry 1EAS) [30] taking full advantage of the great similarity in the active site regions of both enzymes (when the backbone atoms of the catalytic triad and the residues forming the central core of the binding site are superimposed the rms deviation is only 0.26 Å) [33]. The inhibitor molecules were model-built using standard bond lengths and angles based on the crystallographic structure of **D13**, and they were manually docked into the active site using the interactive graphics program Insight II [39]. Atom-centered charges for all the inhibitors were derived [40] by fitting the molecular electrostatic potential calculated with the AM1 Hamiltonian to a monopole-monopole expression [41]. The same sets of charges and radii were used in programs AMBER [42] and DelPhi, described below. Covalent and non-bonded parameters for the inhibitors were derived, by analogy or through interpolation, from those already present in the AMBER database [43]. Torsional parameters for $COO-CH_3$, $-OCH_3$, and $-SO_2-NHCH_3$, not present in the AMBER database, were fitted to reproduce the corresponding torsional barriers calculated *ab initio* [44] from model systems.

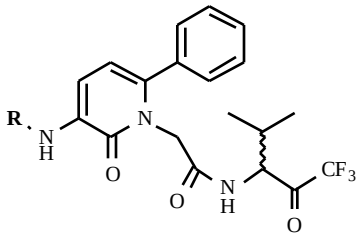
b) Molecular Mechanics Calculations

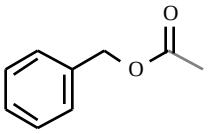
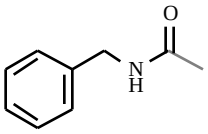
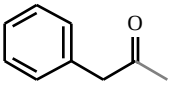
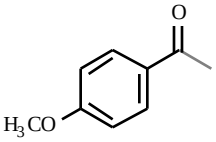
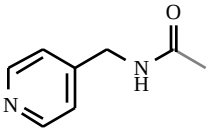
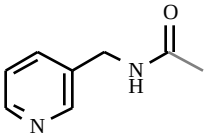
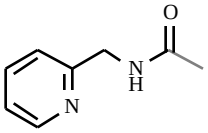
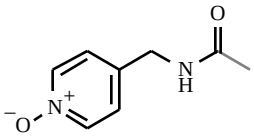
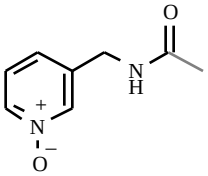
All-atom AMBER force field parameters were used for both the inhibitors and the protein. The geometry of each inhibitor was optimized while the protein atoms were held fixed. The energy minimizations proceeded using the steepest descent algorithm for 500 steps and then switching to the conjugate gradient minimizer until the root-mean-square value of the potential energy gradient was below 0.1 kcal mol^{-1} $Å^{-1}$. A cutoff of 15.0 Å and a distance-dependent dielectric constant ($\varepsilon = r_{ij}$) were used throughout.

c) Partitioning of the Intermolecular Interaction Energy and Pretreatment of the Resulting Energy Matrix

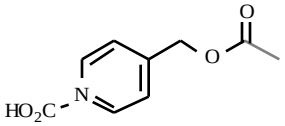
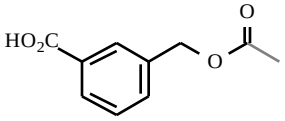
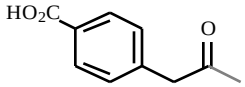
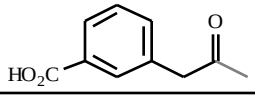
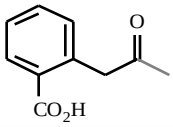
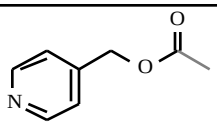
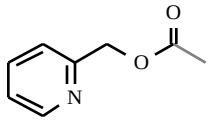
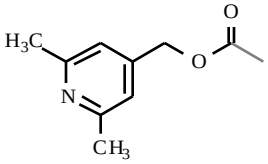
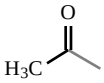
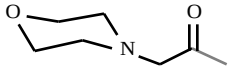
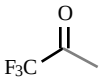
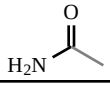
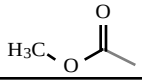
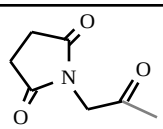
The individual residue contributions to the calculated ligand-receptor interaction energies in the refined complexes were obtained from the ANAL module in AMBER. Each inhibitor was regarded as a single fragment, all atom pairs were included in the calculation, and the same distance-dependent dielectric used in the energy refinement was employed for the analysis. Intramolecular energies, which are a potential source of noise, were not included as they appear to result in little improvement in the regression models [20,26,29]. Since there are 218 amino acids in the protein and 2 energy contributions (van der Waals and electrostatic) are considered for each residue, 436 variables were used to characterize each complex. These energy descriptors made up the matrix for the chemometrics program Q^2 [45].

Table 1. Human Neutrophil Elastase Inhibitors Included in the Training Set [30]

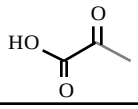
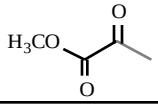
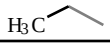
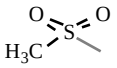
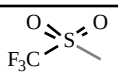
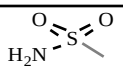
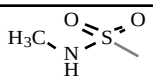
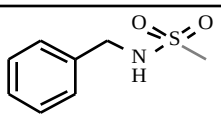
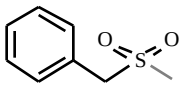
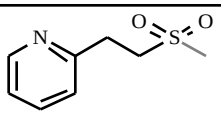
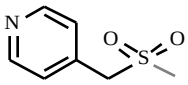
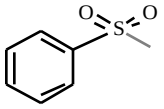
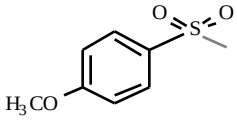
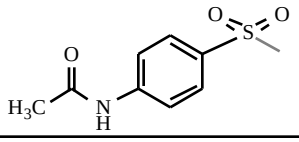
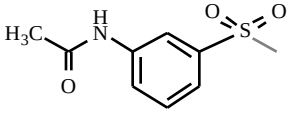


Inhibitor	**R** substituent	Activity (K_i)	ΔG_{desolv}^{R}	ΔG_{desolv}^{L}
I2A		4.5 ± 0.8	7.76	2.88
I2B		7.0 ± 1.6	8.30	4.24
I2C		15.0 ± 2.6	8.12	2.56
I2D		480.0 ± 170	8.42	3.22
I2E		16.0 ± 1	8.05	4.12
I2F		15.0 ± 3	8.08	4.03
I2G		15.0 ± 3	8.26	4.51
I2H		52.0 ± 5	8.73	4.27
I2I		52.0 ± 1	8.12	4.26

(Table 1). contd.....

Inhibitor	**R** substituent	Activity (K_i)	ΔG_{desolv}^{R}	ΔG_{desolv}^{L}
I2J		5.7 ± 1.5	7.64	3.06
I2K		4.8 ± 0.8	7.58	3.79
I2L		13.0 ± 2	7.67	2.66
I2M		23.0 ± 2	8.74	2.58
I2N		18.0 ± 4	7.88	2.64
I2O		8.0 ± 1	7.43	2.64
I2P		8.8 ± 1.9	7.48	2.81
I2Q		4.1 ± 2.1	7.34	3.21
I2R		57.0 ± 10	7.54	2.59
I2S		469.0 ± 190	8.29	2.88
I2T		39.0 ± 11	7.43	2.31
I2U		26.0 ± 3	7.06	2.76
I2V		43.0 ± 6	7.00	4.13
I2W		15.0 ± 3	7.24	2.66
I2X		30.0 ± 7	8.17	2.60

(Table 1). contd.....

Inhibitor	**R** substituent	Activity (K_i)	ΔG^{R}_{desolv}	ΔG^{L}_{desolv}
I2Y		34.0 ± 15	7.61	3.22
I2Z		80.0 ± 30	7.88	2.67
I07		62.0 ± 6	6.15	2.36
I10		8.7 ± 3.7	7.31	2.82
A13		9.3 ± 4.5	7.10	4.28
B13		190.0 ± 90	7.33	2.81
C13		71.0 ± 42	7.13	3.75
D13		18.0 ± 3	7.59	3.34
E13		1.8 ± 0.7	7.40	3.45
F13		1.4 ± 0.2	7.17	4.03
G13		6.9 ± 1.7	7.76	3.69
H13		13.0 ± 5	7.35	3.36
I13		6.0 ± 2	7.64	3.51
J13		8.4 ± 3.3	7.83	3.92
K13		0.7 ± 0.2	7.88	4.48
L13		9.4 ± 4.7	7.82	3.29

[a] experimental inhibition constant (nM); [b] electrostatic contribution to the free energy of desolvation of the receptor upon complex formation (kcal/mol); [c] electrostatic contribution to the free energy of desolvation of the ligand upon complex formation (kcal/mol).

d) Continuum Electrostatics Calculations

A finite difference method [46], as implemented in program DelPhi [39], was used to solve the linear form of the Poisson-Boltzmann equation and to describe the electrostatic effects of ligand binding in a format appropriate for use in COMBINE analysis. The procedure has been described in detail previously [20, 21]. In summary, the binding process is divided in two stages, the first one consisting in desolvating the apposing surfaces of both the ligand and the receptor, and a second one consisting in letting the charges of the two molecules interact. It is thus possible to separate the change in electrostatic free energy on molecular association (ΔG_{ele}) into three components: (i) the change in solvation energy of the ligand upon binding (ΔG^{L}_{desolv}), (ii) the change in solvation energy of the receptor upon binding (ΔG^{R}_{desolv}), and (iii) the ligand-receptor interaction energy in the presence of the surrounding solvent. The first two terms are calculated by considering the effects on the respective electrostatic free energies of replacing the high dielectric medium of the solvent with the low dielectric medium of the other molecule in those regions that are occupied by the binding partner in the complex. To calculate the residue-based ligand-receptor interaction energies, the solvent-corrected potential generated by the charges on the ligand is computed at the positions of each of the uncharged atoms of the receptor.

The atomic coordinates employed were those of the AMBER-optimized complexes. The surrounding solvent was treated as a high dielectric medium (ε = 80) with ionic strength of 0.145 M whereas the interior of the protein, the ligands and the complexes was considered a low dielectric medium (ε = 4). Cubic grids with a resolution of 0.5 Å were centered on the molecular systems considered, and the charges were distributed onto the grid points. Solvent-accessible surfaces, calculated with a spherical probe of 1.4 Å radius, defined the solute boundaries, and a minimum separation of 11 Å was left between any solute atom and the borders of the box. The potentials at the grid points delimiting the box were calculated analytically by treating each charged atom as a Debye-Hückel sphere.

e) Partial Least Squares (PLS) Analysis

No scaling or variable selection was carried out except for a mild pretreatment that consisted of removing any variables with a standard deviation below 0.01 kcal mol^{-1}. This procedure reduced the number of variables that entered the PLS analysis from 436 (or 438 when the desolvation energies of ligand and receptor were included) to around 350. The PLS analysis was carried out as implemented in the Q^2 program, and the models were validated using both internal and external validation tests. For internal validation, two methods of cross-validation were employed: the widely used leave-one-out procedure and a more stringent method that assigns the compounds randomly to one of five groups of approximately the same size, excludes from the analysis each group in turn, and repeats the whole procedure 20 times.

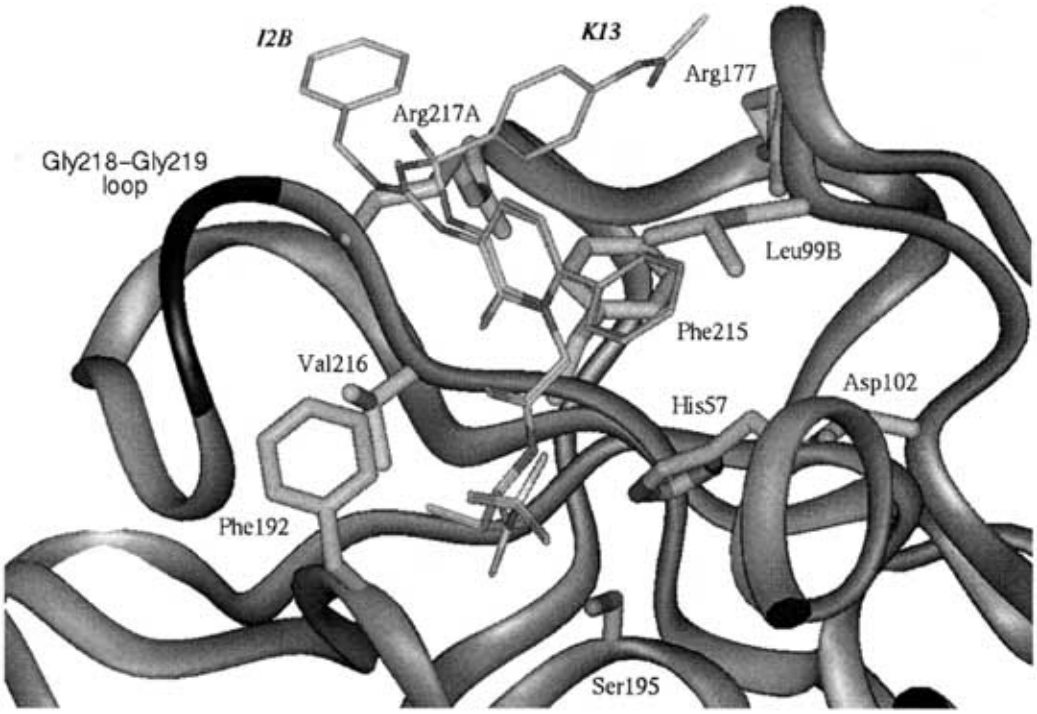


Fig. (2). Detail of the inhibitor binding site of human neutrophil elastase. The α-carbon trace of the protein is displayed as a ribbon, with the portion corresponding to Gly218 and Gly219 coloured darker. The side chains of important residues are shown as thick sticks and are labelled. For simplicity only two inhibitors, I2B and K13 (each representative of a different class), are shown as thin sticks in their respective pre-covalent complexes with the enzyme.

The latter method gives more conservative results: a smaller cross-validated correlation coefficient (q^2) and a higher cross-validated Standard Deviation Error of Predictions (SDEPcv). For external validation, the PLS models obtained were used to predict the biological activity of 12 related 6-phenyl-pyridopyrimidine analogs [47] (prediction set) not included in the training set. The quality of the predictions was assessed by the value of the external SDEP (SDEPex). The optimal dimensionality of each model was determined from the evolution of the cross-validation indexes as a function of the number of latent variables extracted.

RESULTS AND DISCUSSION

a) Structural Description of the Complexes

The common framework of all the inhibitors is the pyridone unit that acts as an effective dipeptide mimetic (Table 1 and Fig. **1**). In all the complexes, the carbonyl oxygen of the trifluoromethyl ketones hydrogen bonds to the backbone NH groups of Gly193 and Ser195 (oxyanion hole) while the isopropyl group occupies the primary specificity pocket (S1). The pyridone ring carbonyl O and the 3-position amine NH are found facing, respectively, the backbone NH and carbonyl O of Val216, and the 6-phenyl substituent nestles into the pocket defined by Leu99, Phe215 and His57 (S2 subsite). Despite this common theme, differences in the degree of fitting can be noted when all the complexes are superimposed on the graphics display. This is a consequence of the chemical diversity of the substituents on the 3-amino group, many of which are directed away from the S5-S4 subsite and cannot readily take full advantage of the binding opportunities afforded by this hydrophobic pocket of the enzyme (Fig. **2**). In this regard, it is worth mentioning that early modeling work at Zeneca suggested these substituents could nestle onto the surface of the enzyme near the loop made up by Gly218 and Gly219 [30]. Molecular dynamics simulations in water by the same group, on the other hand, suggested that the N3 substituents would be mostly exposed to the aqueous environment. What we find, however, is that "straight" inhibitors having an sp^2 carbon attached to the N3 have a propensity to clash sterically with Arg-217 whereas "bent" inhibitors having a sulfonyl group attached to N3 can direct substituents towards the S5-S4 subsites thereby establishing improved interactions with a number of different protein residues.

b) Correlation of Inhibitory Potencies with Calculated Binding Energies

From the statements above it can be inferred that the structural variations on the N3 substituents are likely to co-

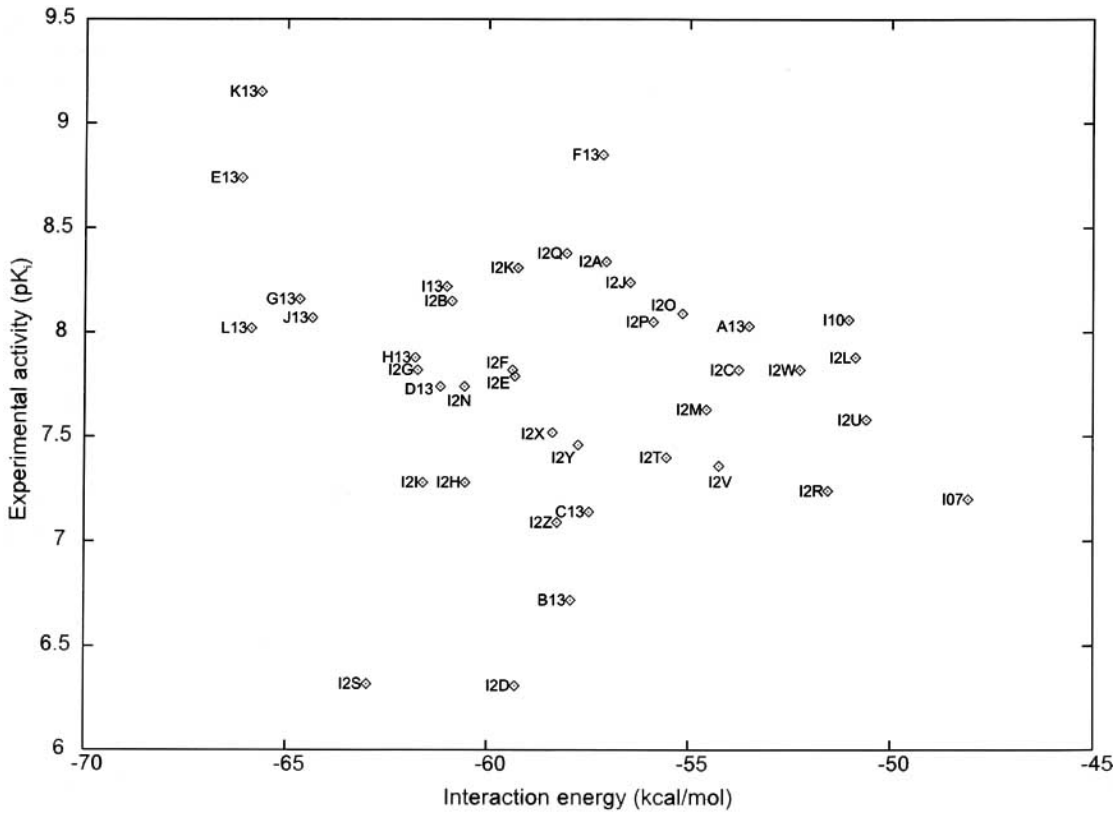


Fig. (3). Scatterplot relating the experimental activities to the global intermolecular interaction energies calculated in the molecular mechanics program (linear regression model).

operatively affect the binding of the rest of the molecule. Due to the many variables involved, recognizing the relation between the change in structure and the change in activity is not obvious. This is clearly manifest when a plot is produced of the biological activities of these inhibitors versus the intermolecular interaction energies calculated with AMBER (Fig. **3**). Apart from revealing a notable scattering of the data and the absence of a meaningful correlation, little additional information can be gained from this type of representation, and even identification of outliers is not straightforward.

On the other hand, breaking down the same calculated interaction energies between ligands and protein into residue-based van der Waals and electrostatic contributions, as described in the Methodology section, allows several things:

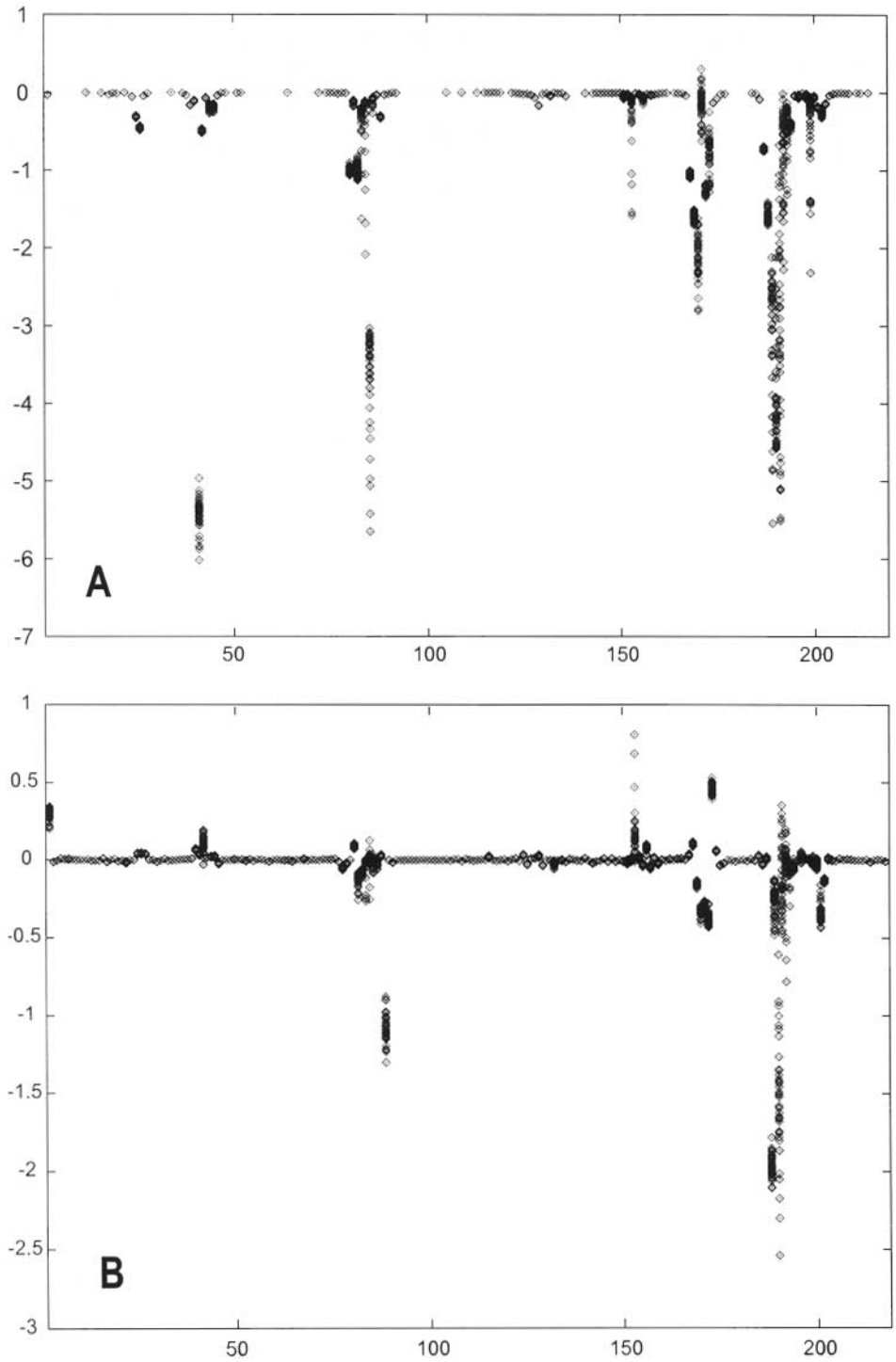


Fig. (4). Graphical overview of the energy matrix after removing those variables with a standard deviation below 0.01 kcal mol^{-1}: (A) AMBER van der Waals term, (B) DelPhi electrostatic term. Each diamond represents a value and interactions for the 40 compounds are plotted. Note that variable number 1 corresponds to the first amino acid residue, Ile-16.

Table 2. Evolution of the Chemometric Indexes as the Number of Principal Components in the COMBINE Model is Increased

Principal Components	XvarExp [a]	Xaccum [b]	r^2	$SDEP_{LOO}$ [c]	q^2_{LOO} [c]	$SDEP_{RG}$ [d]	q^2_{RG} [d]
1	27.9	27.9	0.54	0.45	0.42	0.45	0.43
2	31.3	59.3	0.62	0.44	0.46	0.42	0.49
3	17.4	76.7	0.69	0.41	0.52	0.41	0.52
4	1.4	78.1	0.80	0.44	0.44	0.45	0.42
5	2.2	80.2	0.84	0.40	0.54	0.44	0.44

[a] variance in X explained, [b] accumulated explained variance, [c] leave-one-out method, [d] five random groups.

first, a simple plot of the variation in interaction energy values for all the protein residues (Fig. **4**) is useful in highlighting those amino acids whose interaction with the ligand varies more from complex to complex. Second, the same plot also allows the detection of residues which give rise to unfavorable interaction energies. In this case, only Gly-193, in the S1 subsite, shows a slightly repulsive van der Waals interaction with a few of the inhibitors, but in other cases the plot can well unveil sources of docking errors that could pass otherwise unnoticed. Last, it allows a rapid assessment and comparison of the different variation in the van der Waals and electrostatic blocks.

Following energy decomposition, the energy matrix was analyzed using the PLS method implemented in the Q^2 program [45]. Inclusion of the two variables containing information about the electrostatic contribution to the desolvation of both ligand and receptor upon complex formation had a positive influence on the quality of the model, which was further improved when the electrostatic interactions between ligand and individual residues calculated with AMBER were replaced with those calculated using the continuum method, which now take into account the reaction field of the surrounding solvent around the complexes. For the discussion we will concentrate on this final model which contains the van der Waals interactions from AMBER and the electrostatic terms from DelPhi.

It can be seen how fitting of the model consistently improves as the number of principal components extracted is increased, but the optimal dimensionality is determined from the evolution of the cross-validation indexes (Table 2). In the present case, further increase of the number of model dimensions beyond 3 principal components is clearly not justified as it does not lead to any additional improvement. The first component accounts for almost 30 per cent of the variance in the data; the second one provides an additional 30 per cent and is used to account mostly for the least active compounds (I2D and I2S); and the third one improves the fit of compounds I2A and I2B, and highlights the hydrophobic interaction between the unsubstituted phenyl ring present in these two compounds (Table 1) and the aliphatic part of the side chain of Arg-217. The experimental activities fit to the resulting 3PC model (Fig. **5**) notably better than to the previous linear regression model.

In addition to providing a visual impression of the quality of the correlation between experimental and calculated activities, the plot shown in Fig. (**5A**) identifies B13 as an outlier. This triflamide compound is 20-fold less active relative to A13 and it has been suggested that this is due to deprotonation of the relatively acidic NH at the slighly basic pH of the inhibition assay with the resulting loss of the hydrogen bond with the carbonyl of Val216 [30]. Elimination of this compound results in a significant improvement of the model, both in fitting and in predictive ability (Table 3).

To obtain further insight into the correlation, the PLS model was transformed into a pseudo-multiple linear regression model where each pseudo-coefficient summarizes the contribution to the activity of each variable in the three principal components extracted. Fig. (**6**) graphically describes those protein residues whose van der Waals and/or

Table 3. Evolution of the Chemometric Indexes as the Number of Principal Components in the COMBINE Model is Increased after Exclusion of Compound B13

Principal Components	XvarExp [a]	Xaccum [b]	r^2	$SDEP_{LOO}$ [c]	q^2_{LOO} [c]	$SDEP_{RG}$ [d]	q^2_{RG} [d]
1	27.47	27.47	0.60	0.42	0.48	0.43	0.43
2	30.12	57.59	0.68	0.40	0.51	0.41	0.49
3	19.25	76.85	0.75	0.37	0.59	0.39	0.54
4	2.39	79.24	0.81	0.39	0.55	0.41	0.50
5	1.29	80.53	0.86	0.37	0.58	0.41	0.50

[a] variance in X explained, [b] accumulated explained variance, [c] leave-one-out method, [d] five random groups.

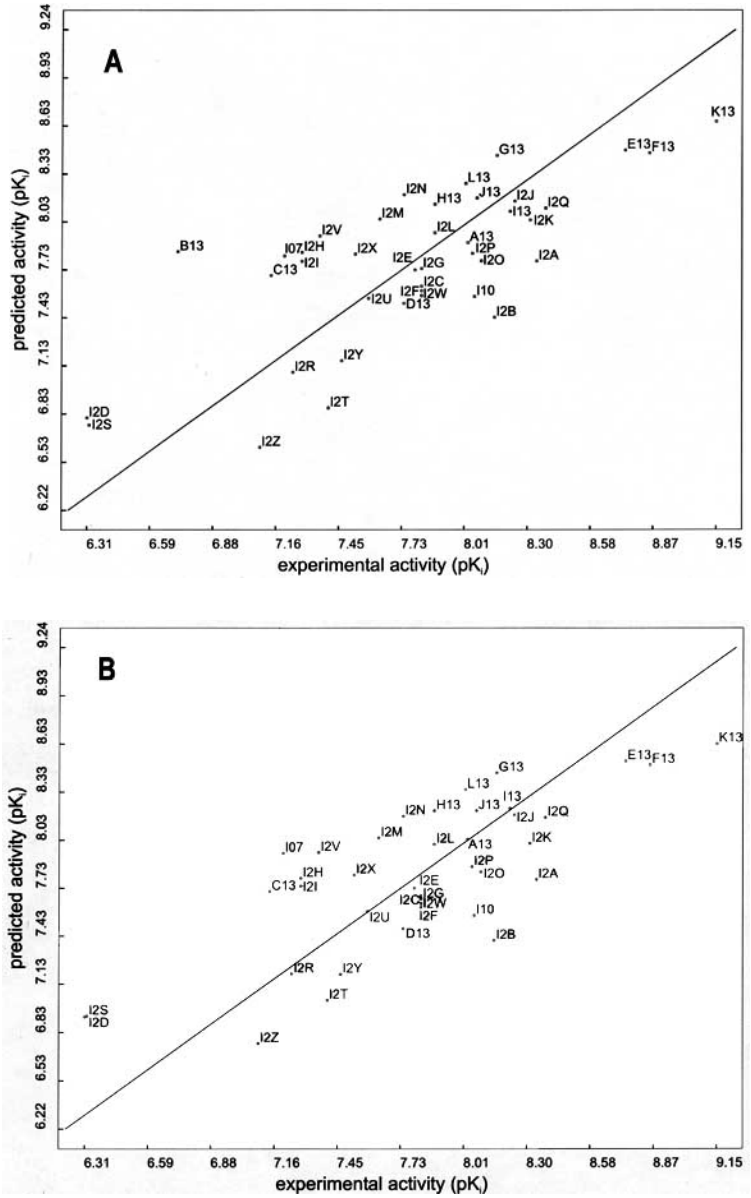


Fig. (5). Scatterplot comparing experimental vs. predicted activities in the final COMBINE models for (A) the 40 compounds of the training series, (B) all compounds except B13.

electrostatic interactions with the ligands are more important for explaining the variation in activity within the series. Around 25 energy terms account for almost 80 per cent of the differences in activity, and van der Waals interactions outnumber electrostatic interactions. The more significant fragments selected include a number of amino acids that are clustered around the regions defined from Val-99 to Asn-99B and from Phe-215 to Gly-219.

The negative coefficients in the van der Waals block highlight that favorable interactions with residues Val-99, Asn-99A, Leu-99B, and Arg-177, as occur in highly active compounds K13 and L13, and others common to other members of the series, such as with Phe-215 and Arg-217, are beneficial for activity. The positive coefficients, on the other hand, highlight that favorable interactions with Phe-192, Gly-218 and Gly-219, are correlated with a decrease in activity. Interaction with Phe-192, in particular, is weighted mostly to characterize the weakest compounds, I2D and I2S, and it may represent other more physically meaningful interactions. Interaction with Tyr-224 is mostly used to account for the decrease in activity of the N-oxide compounds I2H and I2I relative to I2E and I2F, respectively.

In the electrostatic block, small differences in favorable interactions with the backbone atoms of the loop containing Val-99 to Leu-99B appear to be important, and they are rather uniquely exploited by K13, the most active compound in the series. On the other hand, a favorable interaction with Arg-217 and Gly-219 is interpreted to result in a decrease in activity, as exemplified by compounds I2D and I2S. As regards the electrostatic contributions to the desolvation energies, it appears that the larger the desolvation of the receptor, the better for activity, whereas the converse is true

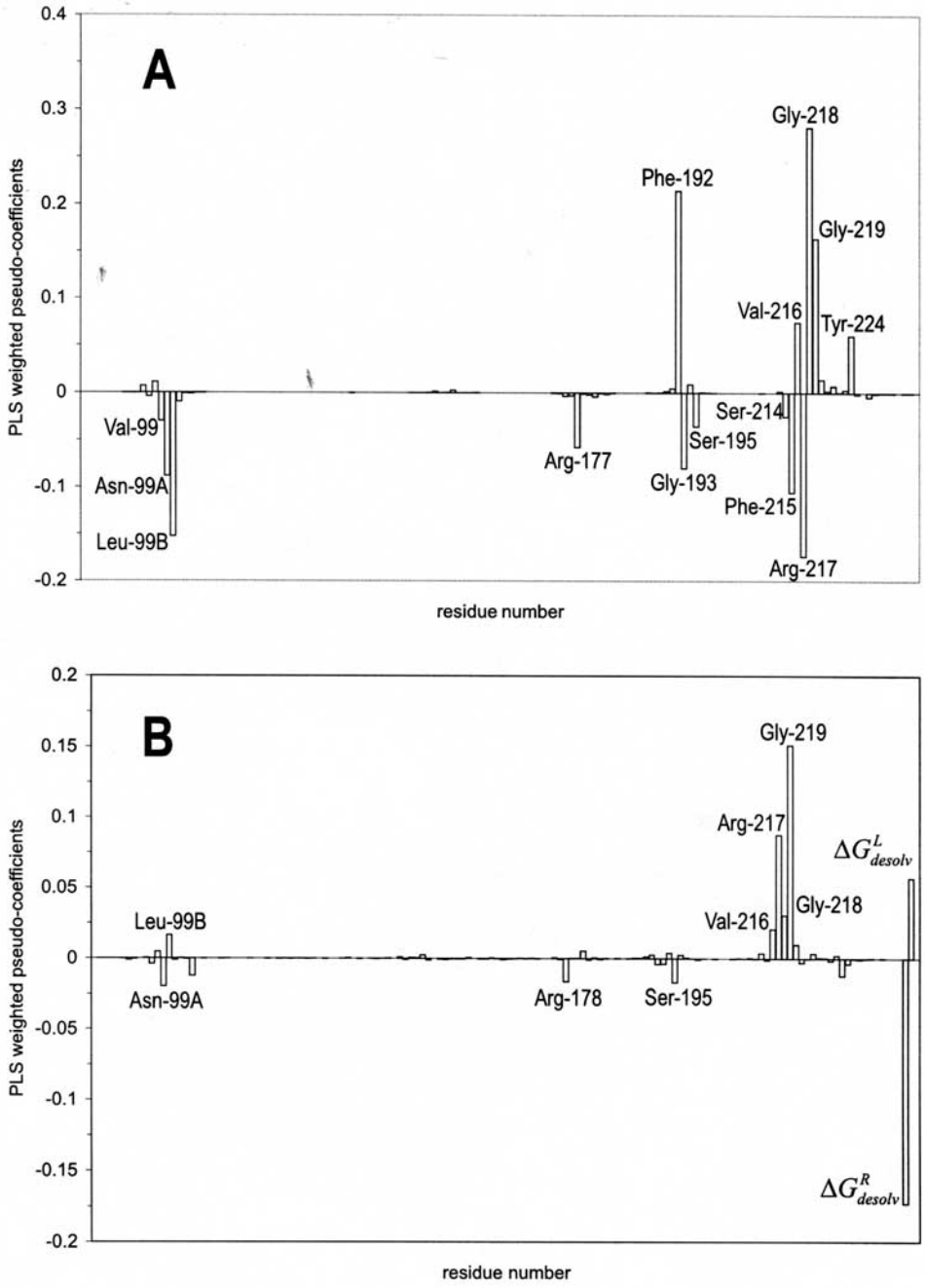


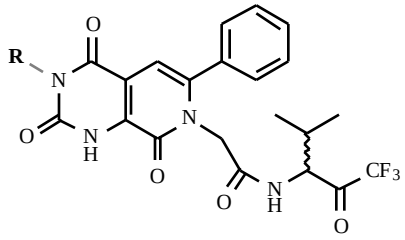
Fig. (6). Weighted PLS pseudo-coefficients (A: van der Waals; B: electrostatic) of the final COMBINE model for each of the variables studied. Compound B13 has been excluded, and the electrostatic contributions are those calculated by means of the continuum method.

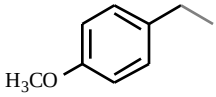
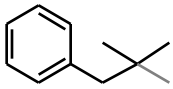
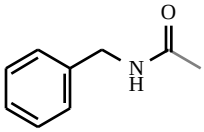
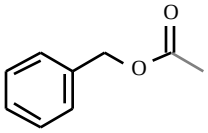
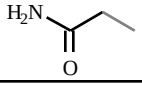
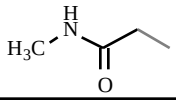
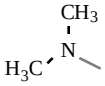
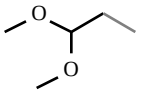
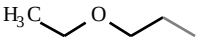
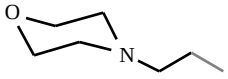
for the ligand. This might reflect differences in the extent of polar and apolar surface area in each species that is buried from the solvent upon complex formation.

c) External Validation

The definitive quality of a COMBINE model, as any other QSAR model, is predictiveness. The model endures that is precise when applied to the prediction of compounds not included in the training series. In order to assess the predictive ability of the final 3-PC COMBINE model (Table3, Fig. **5B**) beyond cross-validation, we attempted the prediction of the biological activity of the series of 6-phenyl-pyridopyrimidine trifluoromethyl ketones [47] reported in Table 4. The energy variables for these compounds were obtained following the procedure described above for the pyridone derivatives. Since the structures in both sets are

Table 4. Human Neutrophil Elastase Inhibitors included in the Prediction Set [47]



Inhibitor	R substituent	Activity (K_i)[a]	ΔG^{R}_{desolv} [b]	ΔG^{L}_{desolv} [c]
A24		15 ± 4.0	8.98	3.26
B24		10 ± 2.6	9.00	3.21
C24		5.7 ± 1.5	10.75	4.19
D24		0.95 ± 0.38	9.42	3.19
E24		9.9 ± 2.3	9.75	4.32
F24		2.8 ± 1.4	9.56	3.29
G24		15 ± 3.4	8.96	4.13
H24		22 ± 4.0	9.17	3.14
I24		19 ± 4.0	9.11	3.15
J24		12 ± 6.0	9.27	3.33
K24		7.5 ± 0.7	8.89	3.40
L24		8.4 ± 1.1	9.07	3.39

[a] experimental inhibition constant (nM); [b] electrostatic contribution to the free energy of desolvation of the receptor upon complex formation (kcal/mol); [c] electrostatic contribution to the free energy of desolvation of the ligand upon complex formation (kcal/mol).

significantly different, the maximum and minimum values found in the training set were used as a cutoff, in order to prevent the extrapolation of the model.

The results of the predictions are represented in Fig. (**7**) on the same plot that shows the recalculated versus experimental activities for the training set. The Standard Deviation Error of the Prediction (SDEP) for the external prediction set is 0.44, which compares quite well with the SDEP of 0.39 obtained in cross-validation (random groups) for the training set (Table 3). Moreover, the predictions are reasonably good for all the compounds except for compound D24, which is underpredicted by about one order of magnitude (experimental: 9.02, predicted: 8.14). Among the

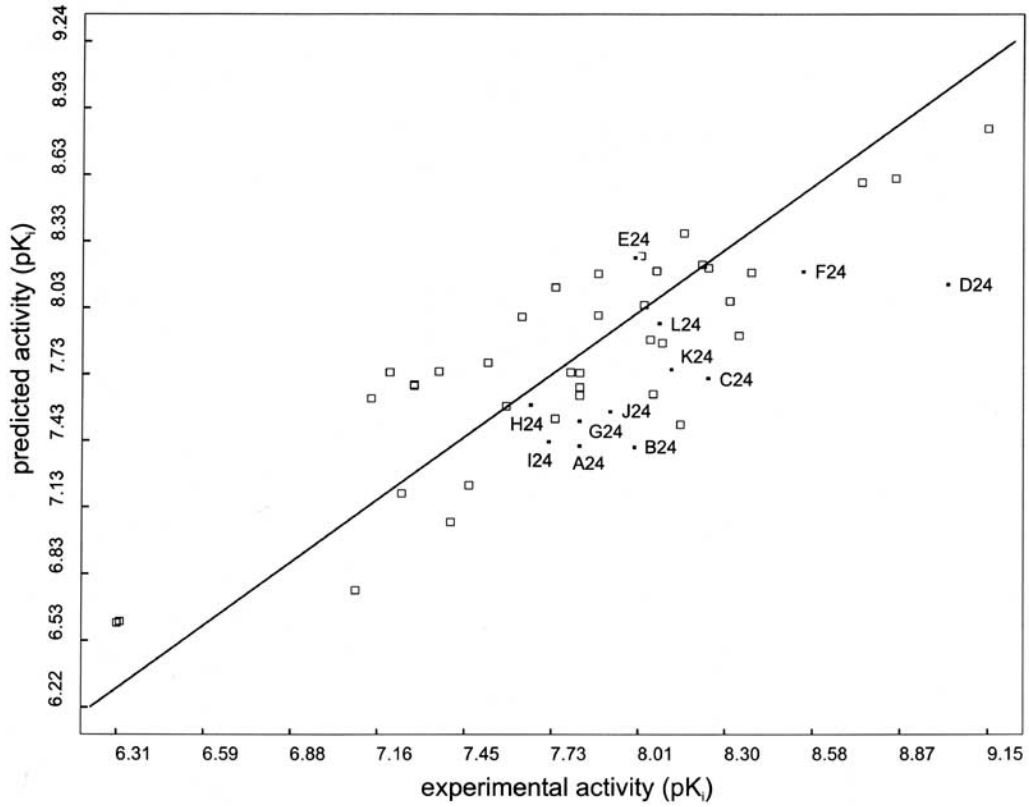


Fig. (7). Plot of recalculated versus experimental activities for the training set (open squares) showing the predictions for the compounds in the test set (filled squares, series 24, Table 4) using the final COMBINE model.

possible explanations, we would suggest that the bulky substituent on D24 makes use of hydrophobic and van der Waals interactions which are probably not sufficiently weighted in the previous model, and does not make use of polar interactions in that region (as defined, for example, by K13) that do have assigned weights. This finding underscores the fact that a COMBINE model can only make use of information already contained in the training series, and for this reason optimal results are to be expected with adequately designed series [20].

CONCLUSIONS

Structural variation in the series of HNE inhibitors under examination is exclusively concentrated on the N3-substituent attached to the pyridone unit. The series studied was designed with the aim of improving oral bioavailability and is less than ideal for chemometric analysis. On the one hand, the activities have large experimental errors and are not uniformly distributed among the members of the series: there is a large majority of medium-activity compounds, two are only weakly active and only a few can be classified as highly active. On the other hand, the suggested variety of binding modes, which is always a caveat for molecular modelling purposes, highlights the need for a tool that can assist in evaluating the goodness of the proposed models and the docking process itself. All in all, the series represents a serious challenge for any QSAR method, but COMBINE analysis appears to perform reasonably well and can be considered a good candidate for such a tool.

For the pyridone-based inhibitors studied, the inhibitory potency is proposed to depend to a large extent on a close fit of van der Waals interaction surfaces between the inhibitor structures and parts of the convoluted surface of the enzyme. Interaction with other residues, on the contrary, is interpreted as unfavorable for activity. This sort of information can be used to advantage by focusing the design of new structural changes with a view to improving the binding affinity of the next round of compounds.

From this and previous examples [20,26,27,28], it seems that COMBINE analysis is a relatively inexpensive tool suitable to derive robust QSAR models and to make proper assessment of existing data. Given the need to assign charges and parameters to all the ligands and the current implementation within a molecular mechanics force field for energy refinement and decomposition, the method in its present state is best suited for the final stages of ligand development and optimization. Predictivity can be extended to other classes of ligands provided they bind in the same region explored by the training series [20]. It will be interesting to see whether the COMBINE methodology can also be used to direct combinatorial libraries iteratively by focusing chemical diversity to the areas highlighted by the analysis.

ACKNOWLEDGMENTS

Fellowships to C. Cuevas (Basque Government) and C. Pérez (Spanish Ministry of Education and Culture) are gratefully acknowledged. This research has been financed by the Spanish CICYT (Project SAF98-0092). The critical and constructive comments of the reviewers are appreciated.

REFERENCES

[1] Greer, J.; Erickson, J.W.; Baldwin, J.J.; Varney, M.D. *J. Med. Chem.*, **1994**, *37*, 1035.

[2] Whittle, P.J.; Blundell, T.L. *Ann. Rev. Biophys. Biomol. Struct*., **1994**, *23*, 349.

[3] Walters, W.P; Stahl, M.T.; Murcko, M.A. *Drug Discov. Today*, **1998**, *3*, 160.

[4] Ajay; Murcko, M. A. *J. Med. Chem*., **1995**, *38*, 4953.

[5] Schnecke, V.; Kuhn, L. *Persp. Drug Discov. Des*., **2000**, *20*, 171.

[6] Marrone, T. J.; Luty, B. A.; Rose, P. W. *Persp. Drug Discov. Des.*, **2000**, 20, 209.

[7] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol*., **2000**, *295*, 337.

[8] Moret, E. E.; van Wijk, M. C.; Kostense, A. S.; Gillies, M. B. *Med. Chem. Res*., **1999**, *9*, 604.

[9] Oprea, T. I.; Gottfries, J.; Sherbukhin, V.; Svensson, P.; Kühler, T. C. *J. Mol. Graph. Model*., **2000**, *18*, 512.

[10] Cruciani, G.; Watson, K.A. *J. Med. Chem*. **1994**, *37*, 2589.

[11] Pastor, M.; Cruciani, G.; Clementi, S. *J. Med. Chem*., **1997**, *40*, 1455.

[12] Goodford, P.J. *J. Med. Chem*., **1985**, *28*, 849.

[13] Valler, M.J.; Green, D. *Drug Discov. Today*, **2000**, *5*, 286.

[14] Li, J.; Murray, C.W.; Waszkowycz, B.; Young, S.C. *Drug Discov. Today*, **1998**, *3*, 105.

[15] Bohacek, R.S.; McMartin, C. *J. Med. Chem.*, **1992**, *35*, 1671.

[16] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comp. Chem*., **1998**, *19*, 1639.

[17] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comp. Chem*., **1992**, *13*, 505.

[18] Böhm, H. J. *J. Comput. Aided Mol. Des*., **1994**, *8*, 243.

[19] Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. *J. Med. Chem.*, **1995**, *38*, 305.

[20] Pérez, C.; Ortiz, A. R.; Pastor, M.; Gago, F. *J. Med. Chem.*, **1998**, *41*, 836.

[21] Checa, A.; Ortiz, A.R.; de Pascual-Teresa, B.; Gago, F. *J. Med. Chem.*, **1997**, *40*, 4136.

[22] Kollman, P. *Chem. Rev.*, **1993**, *93*, 2395.

[23] Barril, X.; Orozco, M.; Luque, F. J. *J. Med. Chem*., **1999**, *42*, 5110.

[24] Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Prot. Eng*., **1994**, *7*, 385.

[25] Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. *J. Med. Chem.*, **1999**, *42*, 5142.

[26] Ortiz, A.R.; Pisabarro, M.T.; Gago, F.; Wade, R.C. *J. Med. Chem.*, **1995**, *38*, 2681.

[27] Tomic, S.; Nilsson, L.; Wade, R.C. *J. Med. Chem.*, **2000**, *43*, 1780.

[28] Lozano, J.J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N.B.; Gago, F.; Sanz, F. *J. Comp.-Aided Mol. Des*., **2000**, *14*, 341.

[29] Wade, R.C.; Ortiz, A.R.; Gago, F. In *3D QSAR in Drug Design, Ligand-Protein Interactions and Molecular Similarity*; Kubinyi, H., Folkers, G., Martin, Y.C., Eds.; Kluwer Academic Publishers: Dordrecht, **1998**; Vol. *2*, pp. 19-34.

[30] Bernstein, P.R.; Andisik, D.; Bradley, P.K.; Bryant, C.B.; Ceccarelli, C.; Damewood, J.R. Jr; Earley, R.; Edwards, P.D.; Feeney, S.; Gomes, B.C.; et al *J. Med. Chem*., **1994** , *37*, 3313.

[31] Barnes, P.J. *Trends in Pharmacol. Sci*., **1998**, *19*, 415.

[32] Skiles, J.W.; Jeng, A.Y. *Exp. Op. Ther. Pat*., **1999**, *9*, 869.

[33] Bode, W.; Meyer, E. Jr.; Powers, J.C. *Biochemistry*, **1989**, *28*, 1951.

[34] Bode, W.; Wei, A.Z.; Huber, R.; Meyer, E.; Travis, J.; Neumann, S. *EMBO J.*, **1986**, *5*, 2453.

[35] Wei, A.Z.; Mayr, I.; Bode, W. *FEBS Lett*., **1988**, *234*, 367.

[36] Navia, M.A.; McKeever, B.M.; Springer, J.P.; Lin, T.-Y.Lin, Williams, H.R.; Fluder, E.M.; Dorn, C.P.; Hoogsteen, K. *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 7.

[37] Veale, C.A.; Bernstein, P.R.; Bohnert, C.M.; Brown, F.J.; Bryant, C.; Damewood, J.R. Jr.; Earley, R.; Feeney, S.W.; Edwards, P.D.; Gomes, B.; Hulsizer, J.M.; Kosmider, B.J.; Krell, R.D.; Moore, G.; Salcedo, T.W.; Shaw, A.; Silberstein, D.S.; Steelman, G.B.; Stein, M.; Strimpler, A.; Thomas, R.M.; Vacek, E.P.; Williams, J.C.; Wolanin, D.J.; Woolson, S. *J. Med. Chem.*, **1997**, *40*, 3173.

[38] Edwards, P.D.; Bernstein, P.R. *Med. Res. Rev.*, **1994**, *14*, 127.

[39] Insight II (**1998**), Molecular Simulations, 9685 Scranton Road, San Diego, CA 92121-3752.

[40] Stewart, J.J.P. MOPAC 93 (**1993**), Fujitsu Ltd., Tokyo, Japan.

[41] Besler, B.H.; Merz, K.M.; Kollman, P.A. *J. Comput. Chem*., **1990**, *11*, 431.

[42] AMBER (UCSF): *Assisted Model Building with Energy Refinement*, version 4.1, **1995**. Department of Pharmaceutical Chemistry, University of California, San Francisco.

[43] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.*, **1995**, *117*, 5179.

[44] Gaussian 94, Revision E.1 (**1995**), Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A., Gaussian, Inc., Pittsburgh PA.

[45] Q^2 version 4.5 (**2000**), Multivariate Infometric Analysis, S.r.l. Perugia, Italy.

[46] Nicholls, A.; Honig, B. *J. Comput. Chem*., **1991**, *12*, 435.

[47] Edwards, P.D.; Andisik, D.W.; Strimpler, A.M.; Gomes, B.; Tuthill, P.A. *J. Med. Chem.*, **1996**, *39*, 1112.

Received: 17 July 2000 **Accepted: 11 December 2000**