Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis

Angel R. Ortiz,*,†,‡ M. Teresa Pisabarro,†,‡ Federico Gago,‡ and Rebecca C. Wade*,†

European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany, and Departamento de Fisiología y Farmacología, Universidad de Alcalá de Henares, 28871 Madrid, Spain

Received November 29, 1994®

A new computational method for deducing quantitative structure–activity relationships (QSARs) using structural data from ligand–macromolecule complexes is presented. First, the ligand–macromolecule interaction energy is computed for a set of ligands using molecular mechanics calculations. Then, by selecting and scaling components of the ligand–macromolecule interaction energy that show good predictive ability, a regression equation is obtained in which activity is correlated with the interaction energies of parts of the ligands and key regions of the macromolecule. Application to the interaction of the human synovial fluid phospholipase A_2 with 26 inhibitors indicates that the derived QSAR has good predictive ability and provides insight into the mechanism of enzyme inhibition. The method, which we term comparative binding energy (COMBINE) analysis, is expected to be applicable to ligand–receptor interactions in a range of contexts including rational drug design, host–guest systems, and protein engineering.

Introduction

A primary goal of any drug design strategy is to predict the activity of new compounds,[1] and a promising route to achieve this is to exploit the information contained in the structures of critical biological macromolecules.[2] The ever accelerating rate at which the three-dimensional structures of biomacromolecules and their complexes with ligands are being solved means that it is becoming increasingly common to have this information available in drug-design projects.[3] One of the main challenges for structure-based drug design is thus to generate compounds that are more active than those previously synthesized and tested, on the basis of the receptor structure.[4]

Even when the three-dimensional structure of a pharmacologically interesting macromolecule is available, the chemical modifications necessary to improve binding are often not obvious.[5] The ability to readily calculate accurate ligand affinities is currently an elusive goal. A promising technique is the free energy perturbation method,[6–8] but at present this requires considerable computational resources, is restricted to closely related ligands, and often lacks the accuracy required[7] due to the complexity of the free energy surface of the ligand–receptor complex.[6–8] On the other hand, empirical correlations between binding affinities and a set of physicochemical descriptors of a series of ligands, quantitative structure–activity relationships (QSARs), have long been used in drug design,[9] and extension to the three-dimensional properties of the ligands (3D-QSAR) have proven useful.[10] Current QSAR approaches are, however, of limited value in structure-based drug design, as they make little use of structural information about ligand–receptor complexes in the derivation of correlations. Hansch and Klein suggested using qualitative molecular graphics analysis to confirm QSAR results.[11] The aim of the present work is to develop a more systematic and quantitative approach to using macromolecular structural data in the derivation of QSARs. This is achieved by using calculated ligand–receptor interaction energies *directly* as regressors in the QSAR.

Enzyme inhibition is often found to be a function of the variance of certain physicochemical properties at specific sites.[9] Thus, within a set of related compounds, it can be expected that only a subset of the terms of the Hamiltonian describing the binding energy will account for most of the variance in ligand affinity. In fact, several authors have shown good correlations between particular energy components and biological activity.[12] Therefore, in this work, variable selection procedures are used during the statistical analysis to systematically separate these "energetic signals" from the "background noise" in order to obtain a correlation between binding free energies and a subset of weighted energy terms.

Compared with classical molecular mechanics calculations of binding energies, the advantages of subjecting ligand–receptor interaction energies to statistical analysis are that the noise due to inaccuracies in the potential energy functions and molecular models can be reduced and mechanistically important interaction terms can be identified. Compared to more traditional QSAR analysis, this approach can be expected to be more predictive, as it incorporates more physically relevant information about the energetics of the ligand–receptor interaction.

This new approach, referred to as Comparative Binding Energy Analysis (COMBINE analysis), is here applied to and tested on a series of 26 inhibitors (see Figure 1 and Table 1) of the human synovial fluid phospholipase A_2 (HSF-PLA_2; Figure 2).[13,14] This enzyme catalyzes the hydrolysis of the *sn*-2 acyl side chain of phosphoglycerides, resulting in the subsequent liberation of arachidonic acid from the membrane.[15] Arachidonic acid is a precursor of inflammatory mediators such as prostaglandins, prostacyclins, and leukotrienes.[16] HSF-PLA_2 has been found in high concentrations in the synovial fluid of patients with rheumatoid arthritis,[17] and therefore, pharmacological control of the activity of this enzyme is of considerable interest.[18] The three-dimensional structure of the HSF-PLA_2 has been solved

* Authors to whom correspondence should be addressed.
† Germany.
‡ Spain
® Abstract published in *Advance ACS Abstracts*, June 15, 1995.

0022-2623/95/1838-2681$09.00/0

Table 1. Chemical Formulas and Activities of the HSF-PLA_2 Inhibitors[a]

Name	XLM	YLM	ZLM	SN1	SN2	SN3	SN4	SN5	RLM	GLI	% inhibn
LM1166	CH_2	CONH	OPO_2O	$(CH_2)_3CH_3$	C_7H_{14}	CH_2CH_2OH	$(CH_2)_3CH_3$				61.0
LM1192	CH_2	CONH	OPO_2O	$(CH_2)_3CH_3$	C_7H_{14}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$		6.0
LM1216	CH_2	CONH	OPO_2O	$(CH_2)_3CH_3$	C_7H_{14}	CH_2CHNH_3COO	$(CH_2)_3CH_3$				31.0
LM1220	O	PO_2O	OPO_2O	$(CH_2)_4$	C_6H_{12}	$CH_2CH_2NH_3$	CH_3	$(CH_2)_3CH_3$		*R*	33.0
LM1228	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$	*R*	74.0
LM1230	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2OH	$(CH_2)_3CH_3$				49.0
LM1240	O	PO_2O	OPO_2O	$(CH_2)_4$	C_6H_{12}	$CH_2CH_2N(CH_3)_3$	CH_3	$(CH_2)_3CH_3$		*R*	9.0
LM1245	CH_2	CONH	$OPOOCH_3CH_2$	$(CH_2)_3CH_3$	C_7H_{14}	$CH_2CH_2CH_3$	$(CH_2)_3CH_3$				24.0
LM1246	CH_2	CONH	OPO_2CH_2	$(CH_2)_3CH_3$	C_7H_{14}	$CH_2CH_2CH_3$	$(CH_2)_3CH_3$				45.0
LM1258	CH_2	SO_2NH	OCH_2	$(CH_2)_3CH_3$	C_6H_{12}	CF_3	$(CH_2)_3CH_3$				0.0
LM1261	CH_2	CONH	OPO_2O	$(CH_2)_3CH_3$	C_7H_{14}	$(CH_2)_3CH_3$	$(CH_2)_3CH_3$				80.0
LM1265	CH_2	SO_2NH	OPO_2CH_2	$(CH_2)_3CH_3$	C_6H_{12}	$CH_2CH_2CH_3$	$(CH_2)_3CH_3$				30.0
LM1277	CH_2	SO_2NH	OPO_2CH_2	$(CH_2)_3CH_3$	C_6H_{12}	CH_3	$(CH_2)_3CH_3$				33.0
LM1283	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$			*R*	45.0
LM1284	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$	*S*	12.0
LM1292	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2OH	$(CH_2)_3CH_3$			*R*	44.0
LM1293	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2OH	$(CH_2)_3CH_3$			*S*	40.0
LM1298	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	$CH_2CH_2NH_3$	$(CH_2)_3CH_3$				4.0
LM1299	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	$(CH_2)_3CH_3$	$(CH_2)_3CH_3$				0.0
LM1300	CH_2	SO_2NH	OPO_2O	C_6H_5	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$		24.0
LM1304	CH_2	SO_2NH	OSO_2CH_2	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$		28.0
LM1309	CH_2	CONH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	$CH_2CH_2NH_3$	$(CH_2)_3CH_3$		$CH_2C_6H_5$		28.0
LM1313	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$				36.0
LM1338	CH_2	SO_2NH	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$	*S*	46.0
LM1339	CH_2	SO_2CH_2	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$		$CH_2C_6H_5$		79.0
LM1340	CH_2	SO_2CH_2	OPO_2O	$(CH_2)_3CH_3$	C_6H_{12}	CH_2CH_2O	$(CH_2)_3CH_3$				29.0

[a] Reference 13; see Figure 1 for a schematic diagram of the general structure. The activity of each of the molecules is expressed as percentage of enzyme inhibition. XI(50) data available for a small number of these compounds (unpublished) are consistent within experimental error with the percentage inhibition data. The GLI fragment corresponds to the glycerol backbone. Its chirality is specified as follows: *R* indicates that both experiments and modeling were performed with the *R* structure; likewise for *S*. For the remaining compounds, a racemic mixture was used in the experiments, but the *R* form was modeled.

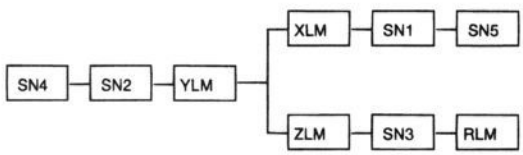

Figure 1. Schematic diagram of the HSF-PLA_2 inhibitors, showing the fragments into which they were divided for the analysis (see also Table 1). The glycerol backbone, corresponding to fragment GLI, is not labeled for clarity.

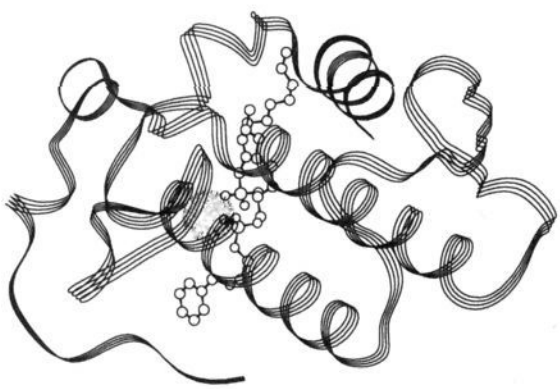

Figure 2. View of HSF-PLA_2 complexed with a representative inhibitor (LM1228). The active site calcium ion is represented by a Van der Waals dot sphere.

by X-ray crystallography both in its native form[19,20] and in a complex with a transition state analogue (LM1220, Table 1).[20]

Methods

A. Outline of the COMBINE Analysis Procedure. Statistical analysis is used to derive a relationship of the following form (see Appendix) between the binding free energy, ΔG (or biological activity), and a selected set of interactions in the ligand−receptor complexes of a family of compounds:

$$\Delta G = \sum_{i=1}^{n} w_i \Delta u_i^{\text{rep}} + C \qquad (1)$$

The n terms, Δu_i^{rep}, of the ligand−receptor binding energy, ΔU, are selected, and the coefficients w_i and constant C are determined by the statistical analysis. ΔU is calculated for representative conformations of the ligand−receptor complexes and the unbound ligands and receptor using a molecular mechanics force field. The ligands are divided into n_l fragments and the receptor into n_r regions, e.g., amino acid residues, and thus

$$\Delta U = \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{\text{VDW}} + \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{\text{ELE}} + \sum_{i=1}^{n_l} \Delta u_i^{\text{B,L}} + \sum_{i=1}^{n_l} \Delta u_i^{\text{A,L}} + \sum_{i=1}^{n_l} \Delta u_i^{\text{T,L}} + \sum_{i<i'} \Delta u_{ii'}^{\text{NB,L}} + \sum_{j=1}^{n_r} \Delta u_j^{\text{B,R}} + \sum_{j=1}^{n_r} \Delta u_j^{\text{A,R}} + \sum_{j=1}^{n_r} \Delta u_j^{\text{T,R}} + \sum_{j<j'} \Delta u_{jj'}^{\text{NB,R}} \qquad (2)$$

The first two terms on the right-hand side describe the intermolecular interaction energies between each fragment i of the ligand and each region j of the receptor. The next four terms describe changes in the bonded (bond, angle, and torsion) and the nonbonded (Lennard-Jones and electrostatic) energies of the ligand fragments upon binding to the receptor, and the last four terms account for changes in the bonded and nonbonded energies of the receptor regions upon binding of the ligand.

A matrix is built with columns representing each of these energy terms and rows representing each compound in the series. A final column containing inhibitory activities is then added to the matrix. A stable QSAR can be obtained from this highly underdetermined matrix using the partial least squares (PLS) method.[22] PLS "solves" the equation by performing rotations of the matrix of energy terms that maximize the linear correlation between the independent (energy terms) and the dependent (activities) variables. Some of the energy terms may not contribute to the differences in binding and

may therefore add "noise" to the matrix of energy terms. To exclude these from the QSAR, a variable selection procedure is carried out in which the effects of individual variables on the model predictivity are evaluated iteratively using a combination of D-optimal and fractional factorial designs,[23] as implemented in the GOLPE program.[38] GOLPE is an advanced variable selection procedure aimed at obtaining PLS regression models with the highest prediction ability.

B. Molecular Model Building. Computer models of all the complexes between the inhibitors in Table 1 and the HSF-PLA_2 were generated. The 3D structure of HSF-PLA_2 cocrystallized with a transition state analogue (LM1220, Table 1) and solved at 2.1 Å resolution[20] was employed. Both subunits in the asymmetric unit cell were superimposed. Significant differences between them were not found, and subunit A was chosen for study. The structures of each of the inhibitors in Table 1 containing either a phosphonate or a sulphonamide at the YLM position were generated by modification of LM1220 using the molecular modeling tools within the INSIGHT-II software package.[24] For the inhibitors containing an amide at the YLM position, the structure of the amide inhibitor solved in a complex with the porcine pancreatic PLA_2[25] (Brookhaven Data Bank pdb entry[26] 5p2p) was used as the basis for further modeling, after docking this inhibitor in the active site of HSF-PLA_2. Bond length and bond angle parameters not included in the cff91 DISCOVER force field[27] were taken from the CHARMm 22 force field.[28] Torsion parameters were either derived using the method of Hopfinger and Pearlstein,[29] based on AM1[30] potential energy surfaces obtained with the MOPAC program,[31] or taken from previous work.[14] Atom-centered charges were either taken from the literature[32,33] or calculated so as to reproduce the quantum mechanical molecular electrostatic potential (MEP). The MEP was generated with the MNDO Hamiltonian in the minimum energy conformation, using the keyword ESP in MOPAC 6.0.[34]

All hydrogen atoms were added to every complex. For those inhibitors with torsional degrees of freedom not present in LM1220, an adiabatic mapping calculation[35] using steepest descent energy minimization was performed with an angle size of 30°, and the conformer with the lowest potential energy was selected. Then, each complex was surrounded by a 5 Å shell of TIP3P water molecules[36] and optimized using the cff91 DISCOVER force field with a 10 Å cutoff and a distance-dependent dielectric constant.[37] Hydrogen atoms were first reoriented and optimized by steepest descent energy minimization (200 steps) while keeping the heavy atoms frozen. Then, steepest descent minimization (200 steps) was applied to the hydrogen atoms and water molecules. After this, the energy of the whole system was minimized to a gradient norm of less than 0.01 kcal mol^{-1} $Å^{-1}$. As an example, the structure of the minimized complex between HSF-PLA_2 and LM1228 is shown in Figure 2.

In order to calculate the energy differences, the conformations of the free enzyme and the free inhibitors were modeled. The crystal structure of the free enzyme[19] (Brookhaven Data Bank pdb entry 1bbc), was optimized using the same protocol as for the complexes after including the calcium ion in the calcium binding site. The calcium ion was not present in this enzyme due to the crystallization conditions. To obtain conformations for the free inhibitors, each compound was positioned in the local minimum nearest to the conformation found in the complex by 1500 steps of conjugate gradient energy minimization.

C. Statistical Analysis. The coordinates of the optimized complexes, free protein, and free inhibitors were analyzed in order to obtain a matrix (the **X** matrix) containing the terms on the right-hand side of eq 2 for each inhibitor. A program was written in FORTRAN-77 to generate this matrix from the output of the DISCOVER program in the appropriate input format for the chemometrics program GOLPE.[38] This program performs three main tasks:

(1) It divides the inhibitors into fragments. The same number of fragments is assigned to all compounds, and "dummy" fragments are added to those inhibitors that lack a particular fragment. The fragments are chosen according to their spatial location in the protein binding site rather than their chemical identity, but care is taken to ensure that dipoles are not split. The fragments defined for the HSF-PLA_2 inhibitors are shown in Figure 1 and Table 1.

(2) It calculates the terms in eq 2 describing the energies of each protein residue and each ligand fragment and the interaction energies between all pairs of residues/fragments for the complexes and for the free inhibitors and protein. In this study, a distance-dependent dielectric constant was used for the electrostatic interactions, and there was no cutoff for the nonbonded interactions.

(3) It eliminates the columns in the **X** matrix corresponding to intrareceptor energy terms that have a standard deviation below a certain threshold (0.02 kcal/mol in this study). This preselection is particularly important for simplifying the analysis of the intramolecular nonbonded interactions of the receptor because, even for small proteins, there can be a very large number of terms representing these interactions.

The **X** matrix prepared for analysis with the PLS module of the GOLPE package[38] contained 3310 columns of X variables corresponding to the energy terms. A column of Y variables containing the experimental activities, expressed as percentage inhibition,[13] was added. The matrix contained 26 rows, one for each compound studied. In GOLPE, the **X** matrix was first pruned by eliminating all columns with standard deviations less than 0.05 kcal/mol. The **X** matrix was then transformed so that each column of data had an average of zero and a standard deviation of one. Then the **X** matrix was reduced in size through D-optimal selection[39] in the partial weight space, using a leave-one-out cross-validation procedure for all structures considering up to five latent variables. The dimensionality of the data matrix was thus reduced while keeping the amount of information lost to a minimum. The number of latent variables chosen for the model was that yielding the best cross-validated performance. At this stage, the predictive ability of the regression models was $Q^2 \sim 0.3$ on average (see below for a detailed explanation of Q^2 and validation procedures). The D-optimized matrix was then reduced using the iterative "fixing and exclusion" of variables procedure through the application of design matrices processed using fractional factorial designs.[23,38] The predictivity of the generated matrices was used to determine which variables improved predictive correlation. Variables which were determined to be noise using this procedure were excluded, and variables which were found to be uncertain were retained. Cross-validation was done by leave-one-out cross-validation analysis allowing up to five latent variables, and the weights were recalculated after exclusion of the objects. The ratio of true variables to dummies was set to 2, with a design combination to variables ratio of 2. The final models were tested by cross-validation using 20 random groups of 5 molecules as the test sets. The model yielding the highest cross-validated performance was retained. The cross-validated performance was characterized by the Q^2 value, calculated as implemented in the GOLPE package as follows:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_{\mathrm{exp}(i)} - y_{\mathrm{pred}(i)})^2}{\sum_{i=1}^{n}(y_{\mathrm{exp}(i)} - \langle y_{\mathrm{exp}} \rangle)^2} \qquad (3)$$

where $y_{\mathrm{pred}(i)}$ corresponds to the activity predicted with the regression model for compound i, $y_{\mathrm{exp}(i)}$ is the experimental activity of compound i, and $\langle y_{\mathrm{exp}} \rangle$ is the average experimental activity of the complete set of n compounds.

D. Validation Procedures. Three different tests were performed in order to study the robustness of the methodology.

In the first test, the effects of chance correlations on the variable selection procedure were investigated. Random numbers, generated using the subroutine RAN1 of numerical recipes,[40] were assigned to the **Y** vector, and then the variable selection procedure was performed for the **X** matrix as described in section C, in order to find out if it was possible to obtain a PLS model with significant predictive ability. Four

runs with different random numbers were performed. Care was taken to ensure that the **Y** vectors created with the random number generator had a Spearman correlation coefficient with the experimental activities below 0.001.

In the second test, the biological activities of the molecules in the **Y** vector were randomly exchanged (scrambled) among the different molecules using the random number generator RAN1[40] before carrying out the regression analysis as described above. This was repeated for 20 different combinations of scrambled activity data in order to obtain statistically significant results.

In the third test, the significance of the cross-validated correlation coefficients obtained after the variable selection procedure was explored, following the recommendations of Wold.[41] The self-consistency of the regression model was tested by using a "blind cross-validation" test. Pairs of randomly grouped molecules were left out of the original data set. A regression model was derived for the rest of the molecules, as described in section C. The model was then used to predict the activity of the left-out molecules. This process was repeated so that the activity of all the molecules in the data set was predicted. A "predictive" cross-validated correlation coefficient (P^2) was then calculated. This is defined by the same formula as Q^2, but it is calculated for "test" compounds about which no information is included in the variable selection procedure. Thus, P^2 differs from the Q^2 parameter calculated within the GOLPE package in order to estimate the predictive ability. For Q^2, the $y_{pred(i)}$ values are obtained for molecules that, although not in the data set used to calculate the current model, are used in other stages of the cross-validation procedure in order to "guide" the variable selection procedure. In other words, the final model contains some information about the molecules used in order to test its predictive ability with the Q^2 parameter.

Results and Discussion

A. Predictive power of the COMBINE analysis. The HSF-PLA_2 system provides an excellent test of the methodology. On the basis of crystallographic studies of inhibitor–PLA_2 complexes, it has been suggested that a substrate or inhibitor embedded in the membrane is transferred to the active site of the enzyme by a facilitated diffusion process through a hydrophobic channel whose opening is on the interfacial binding surface.[42] This process may be facilitated by two factors: the small conformational changes required in the ligand in order to achieve the conformation imposed by the enzyme binding site and the limited solvation effects during the transfer of the ligand from the membrane to the binding site. Thus, an isostructural and desolvated transfer of the ligand from the aggregate to the protein binding site has been suggested.[42] Under these conditions, it can be expected that the main factor responsible for the differences in binding affinities among the inhibitors is the difference in the enthalpic interactions between the inhibitors and the enzyme.

The total ligand–enzyme binding energy can be calculated according to eq 6a in the Appendix. However, the binding energies calculated in this manner for the set of HSF-PLA_2 inhibitors are only weakly correlated with the experimental activities (correlation coefficient $R = 0.21$; see Figure 3). This is not wholly unexpected since inaccuracies in both the molecular mechanics parameters and the molecular models may contribute to this poor correlation. When, however, the PLS method coupled to the GOLPE variable selection procedure is applied to the components of the binding energies, good correlations with activity are obtained. Table 2 shows the predictivity values obtained after the variable selection procedure. The fitted regression coefficient (R^2) is typically about 0.92 (Figure 4a), with a linear regression coefficient $R = 0.95$. The cross-validated regression coefficient (Q^2), which provides a measure of the predictive power of the derived model, is typically 0.82 (Figure 4b), with a linear regression coefficient $R = 0.93$.

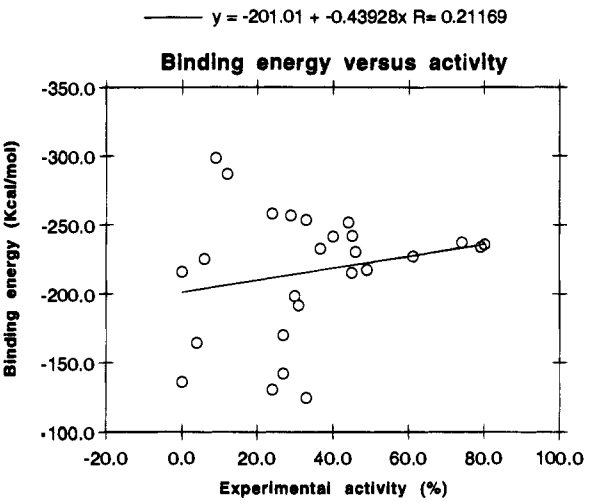

Figure 3. Calculated binding energy of the HSF-PLA_2 inhibitors to the enzyme versus inhibitory activity against HSF-PLA_2 expressed as percentage inhibition.

Table 2. COMBINE Analysis Performance Data

run no.	activity[a]	LV[b]	SDEP[c]	Q^2 [d]	SDEC[e]	R^2 [e]	X_{sel}[f]
1	E	2	9.35	0.82	6.95	0.90	41
2	E	2	9.23	0.83	6.07	0.92	49
3	E	2	9.33	0.82	6.23	0.92	45
4	E	2	9.73	0.81	6.99	0.90	55
5	E	2	8.93	0.83	5.70	0.93	56
6	E	2	9.18	0.83	6.26	0.92	43
7	R	2	19.85	0.31	16.27	0.50	3
8	R						0
9	R						0
10	R						0
11–30	S	2	19.20	0.23	11.96	0.72	204

[a] E, the experimental activities were used in the calculations; R, the **Y** vector was generated using a random number generator; S, the **Y** vector was generated by scrambling experimental activities amongst the inhibitors. Average values are given for the models derived in the 20 scrambling runs which all had two latent variables. [b] Optimum number of latent variables in the final model at the end of each run. [c] Standard deviation in cross-validated prediction, SDEP = $[\sum_{i=1}^{n}(y_{exp(i)} - y_{pred(i)})^2/n]^{1/2}$. [d] Q^2 for cross-validated predictive performance is given by eq 3. [e] SDEC and R^2 are the equivalent of SDEP and Q^2 calculated for fitting. [f] Number of variables selected in the final model.

The robustness of the correlation was ascertained using three different tests. In the first test, activities were replaced by random numbers. In three of the four runs, the variable selection procedure eliminated all the variables in the matrix (Table 2). A model was only obtained in one case: this had two latent variables formed by three energy components and a $Q^2 = 0.31$. This value of Q^2 is small enough to be judged a chance correlation.[43] In the second test, the activities were scrambled among the compounds. The R^2 and Q^2 values obtained in 20 models are shown in Figure 5. The average predictivity is given by $\langle Q^2 \rangle_s = 0.23$, significantly less than the Q^2 values for the models derived with the real activities. In the third test, "blind cross-validation" was performed. The correlation obtained ($P^2 = 0.52$, $R = 0.72$; Figure 6) was less good than that obtained with the standard cross-validation technique used in GOLPE ($Q^2 = 0.82$, $R = 0.95$; Figure 4b). This is because the model derived with the standard method

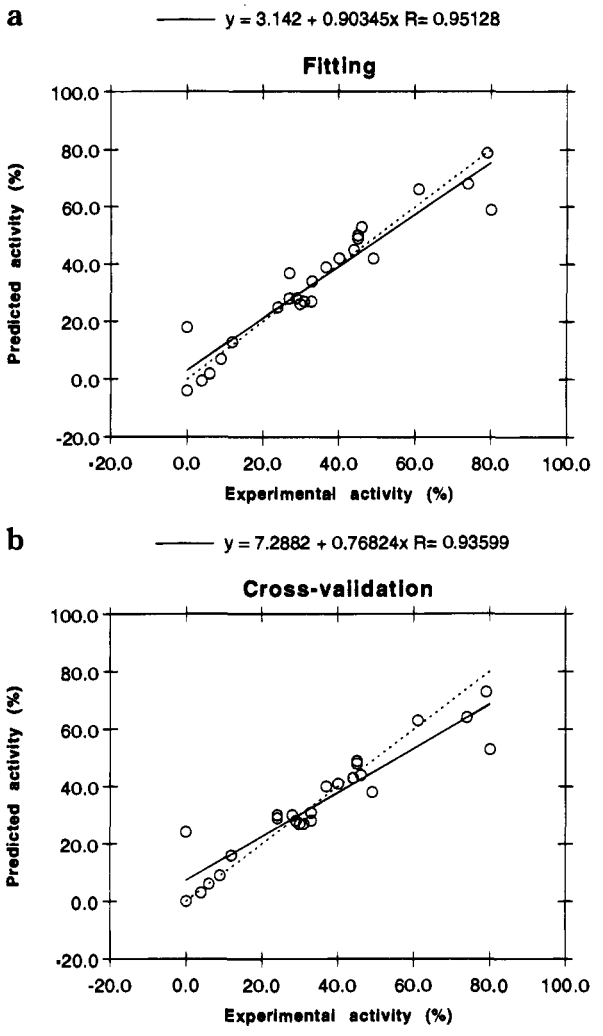

Figure 4. Predicted versus experimental activity for the HSF-PLA_2 inhibitory activity. The predictive model was derived using two latent variables yielding a fitted $R^2 = 0.92$ and a cross-validated $Q^2 = 0.82$ (data from Table 1). (a) Fitted versus experimental activity. (b) Cross-validated versus experimental activity. The dotted line corresponds to a perfect fit, and the solid line shows the regression fit given by the eq 1.

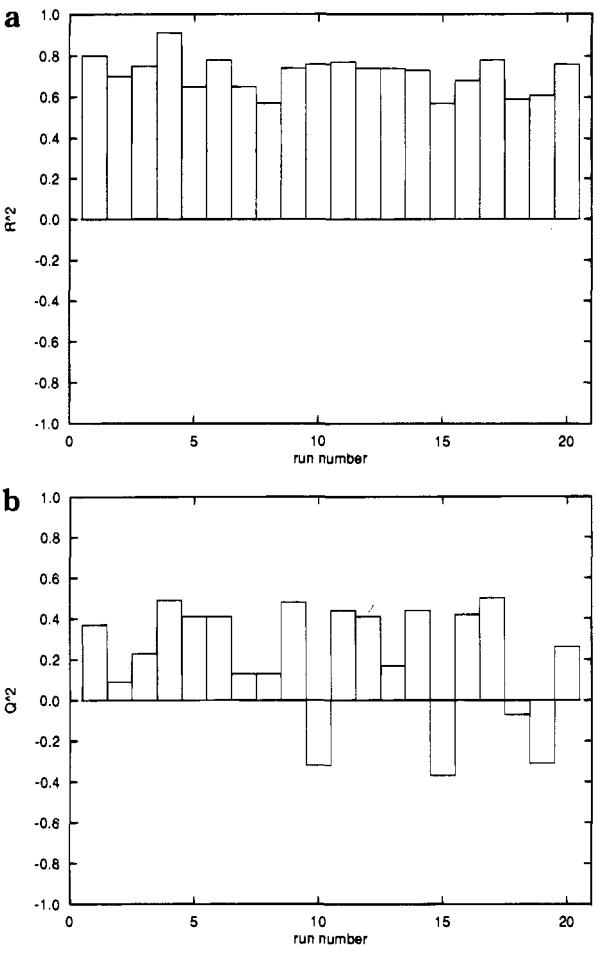

Figure 5. Performance of 20 regression models obtained after randomly scrambling the activity data: (a) R^2 value; (b) cross-validated Q^2 value.

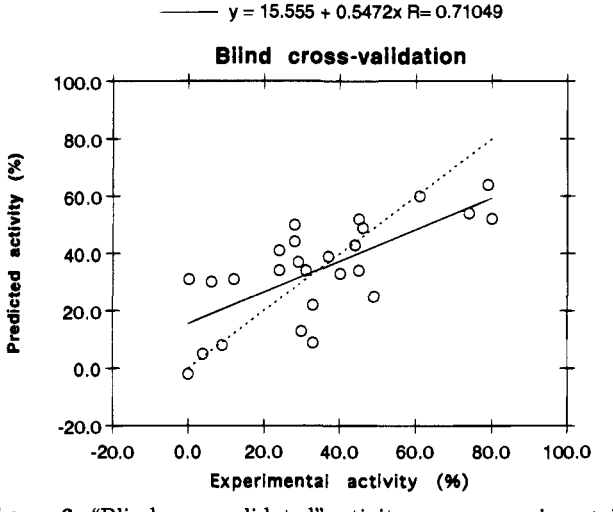

Figure 6. "Blind cross-validated" activity versus experimental activity (see text). The dotted line corresponds to a perfect fit, and the solid line shows the regression fit given by eq 1.

includes some information about the left-out molecules as each variable in the matrix is judged, and subsequently selected or eliminated, according to its ability to improve the correlation with the data set *in cross-validation*. The value of P^2 (=0.52) is similar to that of $Q^2 - \langle Q^2 \rangle_s$ (=0.59), suggesting that the results of the second and third tests are consistent with each other and that the regression models are not the result of spurious correlations.

The evolution of the predictivity of the X-matrix for the "real" activity data as a function of the number of energy variables selected and the number of latent variables is shown in Figure 7. This figure indicates that many of the energy terms contributing to ΔU in eq 2 are not important in determining the differences in experimental activity, since they are poorly correlated with activity. Moreover, a rather limited number of interactions between some parts of the inhibitors and certain residues of the enzyme seems to be responsible for most of the differences in inhibitory activity. This effect can be physically understood by considering the complexity of the conformational space of proteins, which consists of multiple minima with energy differences on the order of 0.04 kcal mol^{-1} (degree of freedom)$^{-1}$.[44] Given the number of degrees of freedom of a typical protein, these energy differences can dominate the difference in binding free energies of two ligands if the receptor occupies different parts of conformational space in the two ligand complexes. If upon modeling two different ligand–receptor complexes the receptor is trapped in different local minima that can actually be occupied in the presence of both ligands, the receptor energy differences may be considerable but irrelevant

Table 3. Selected Energy Variables for One of the Regression Models Derived by COMBINE Analysis[a]

variable number	energy description	residue/ fragment A	residue/ fragment B	weighted coefficient	regression coefficient	energy value
68	inter. coulomb	PHE 5	GLI	−0.0471	−11.270	0.100 93
212	inter. coulomb	LYS 16	ZLM	−0.0560	−3.768	−1.156 16
365	inter. vdw	PHE 24	SN2	−0.0273	−26.010	−0.245 81
530	inter. coulomb	ARG 34	ZLM	−0.0543	−2.004	−1.704 71
805	inter. vdw	TYR 52	XLM	−0.0500	−11.610	−0.509 39
934	inter. coulomb	GLY 67	ZLM	−0.0373	−21.760	−0.083 94
946	inter. coulomb	THR 68	SN3	−0.0517	−4.191	−0.403 08
967	inter. vdw	LYS 69	XLM	−0.0276	−6.611	−0.419 82
1262	inter. coulomb	ASP 99	ZLM	+0.0578	+1.131	3.663 42
1280	inter. coulomb	LYS 100	ZLM	−0.0518	−3.950	−0.918 59
1350	inter. coulomb	PHE 105	ZLM	−0.0626	−36.790	−0.129 92
1504	inter. coulomb	TYR 120	ZLM	+0.0651	+30.000	0.167 32
1568	inter. coulomb	CYS 126	SN3	+0.0536	+36.140	−0.007 11
1576	inter. coulomb	ARG 127	ZLM	−0.0559	−3.562	−1.066 56
1604	inter. coulomb	SER 129	SN3	−0.0677	−63.270	−0.002 72
1624	inter. coulomb	PRO 131	SN3	+0.0608	+9.873	−0.007 02
1666	inter. coulomb	CAL 801	ZLM	−0.0553	−0.036	−130.170 94
1713	inter. vdw	WAT 512	XLM	−0.0652	−24.900	−0.217 79
1782	intra. angle	TYR 22		−0.0811	−6.135	0.025 00
1808	intra. bond	VAL 31		+0.1034	+31.040	0.165 00
1822	intra. torsion	GLY 35		−0.0888	−21.060	0.019 00
1880	intra. bond	LEU 55		−0.0808	−29.150	−0.342 00
1902	intra. angle	THR 68		+0.0872	+3.551	−1.320 00
1904	intra. bond	LYS 69		+0.0518	+23.820	−0.477 00
2065	intra. torsion	LYS 124		−0.1084	−1.658	−4.552 00
2108	intra. nbond	ASN 1	SER 72	−0.0746	−3.759	0.171 98
2166	intra. nbond	ARG 7	GLU 17	+0.0774	+25.900	−0.534 57
2296	intra. nbond	GLU 17	ALA 18	−0.0806	−8.675	−1.573 67
2315	intra. nbond	ALA 18	GLY 23	−0.0676	−6.462	−0.389 45
2493	intra. nbond	ARG 34	LYS 116	+0.0504	+17.540	0.259 40
2693	intra. nbond	VAL 46	ARG 132	−0.0190	−6.754	−0.263 23
2745	intra. nbond	CYS 51	TYR 52	+0.0217	+1.905	0.440 65
2765	intra. nbond	LYS 53	ARG 54	−0.0602	−2.771	2.230 48
2766	intra. nbond	LYS 53	LEU 55	−0.0282	−3.070	0.080 02
2771	intra. nbond	LYS 53	LYS 69	+0.0400	+5.888	0.348 81
2778	intra. nbond	ARG 54	Leu 55	+0.0338	+2.802	1.315 72
2779	intra. nbond	ARG 54	GLU 56	−0.0551	−8.156	1.367 67
2781	intra. nbond	ARG 54	ARG 58	+0.0166	+0.983	0.715 87
2875	intra. nbond	GLY 67	LEU 95	+0.0816	+13.640	0.714 20
3054	intra. nbond	SER 93	CYS 96	+0.0356	+5.278	−0.000 82
3071	intra. nbond	GLU 97	ALA 101	+0.0321	+2.992	0.981 16
3193	intra. nbond	TYR 120	LYS 124	−0.0439	−9.525	−0.886 09
3256	intra. bond	ZLM		−0.0954	−26.130	−0.122 00
3259	intra. bond	SN3		+0.0834	+13.200	0.303 00
3305	intra. nbond	SN2	SN1	−0.0419	−4.760	0.291 80
3306	intra. nbond	SN2	RLM	+0.0990	+41.040	0.000 00

[a] The weighted coefficients indicate the relative importance of each variable. The energies are given in kcal/mol for the LM1166 compound. The activity of this compound (considered to be given by the binding free energy, ΔG) can be calculated by substituting in eq 1, the energies (Δu_i^{rep}) from column 7, the regression coefficients (w_i) from column 6, and $C = -19.94$.

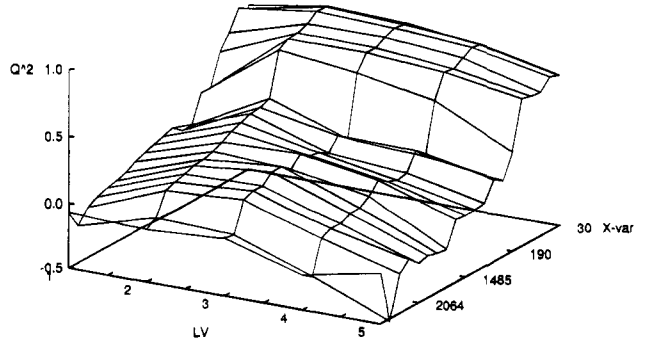

Figure 7. Evolution of the predictivity of the **X** matrix during the variable selection procedure. The predictivity is measured by Q^2 and given as a function of the number of energy variables selected (X-var) and the number of latent variables (LV).

to activity. In such cases, which can arise from incomplete sampling of phase space, comparative binding energy analysis offers a practical solution to the problem by detecting and eliminating the "noise" and properly weighting the remaining terms.

B. Mechanistic Information. Around 50 energy terms are selected from the 3310 initially in the **X** matrix. Thus, only about 2% of the energy terms are required in order to explain the differences in activity (Table 2). One reason for this small number is that some of the variables in the **X** matrix may be highly collinear, and therefore, the selected energy variables may represent combined effects and be susceptible to overinterpretation. For this reason, we refer to the final energy terms selected as "effective" energies. These effective energies may themselves lack physical meaning since they may act as statistical descriptors of other physically important interactions. Nevertheless, it is noteworthy that most of the intermolecular effective energies correspond to interactions with residues in the active site of the enzyme in close contact with the inhibitors. In particular, effective energies are selected for interactions between the XLM fragment and Tyr-52, Lys-69, and the crystallographic water molecule 512, which form a pocket where the *sn*-1 chain of the inhibitors binds (see Table 3). The effective energy

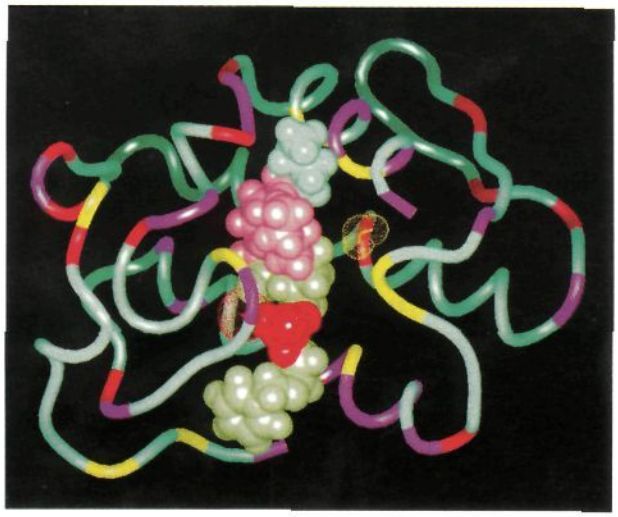

Figure 8. Distribution of the most important energy terms selected in all the regression models mapped onto the structure of the HSF-PLA_2 complexed with LM1228. Residues in the protein and fragments in the inhibitor whose energies are not found to be statistically significant for the differences in inhibitory activity are colored blue and all those that are statistically significant are shown in other colors. Important interactions between the protein and the ZLM fragment of the inhibitor are colored red. The rest of the important protein–inhibitor interactions are colored green in the inhibitor and yellow in the protein. Intraligand interactions are colored pale pink, and intraprotein interactions are colored magenta.

corresponding to the energetically dominating electrostatic interaction between the phosphate group in the ZLM fragment of the inhibitors and the calcium ion in the enzyme is also clearly important. Interestingly, several intramolecular effective energies are also selected, both within the ligands and within the protein. Within the ligands, there is an effective repulsive Lennard-Jones interaction between the *sn*-1 and *sn*-2 acyl chains of the inhibitors with increasing repulsion correlating with decreasing activity. Within the protein, the change in the effective energy corresponding to the nonbonded interaction between Asn-1 and Ser-72 indicates that when the ligand induces strain in this part of the protein, the activity of the inhibitor is reduced. The spatial distribution of the selected effective energies is shown in Figure 8.

The regression models indicate that the interaction of the enzyme with the calcium ion is important for inhibitory activity and that the binding site of the native enzyme is not large enough to accommodate the inhibitors. The enzyme must adapt to each inhibitor when it binds. If relaxation of the enzyme is hindered, then strain can be imposed on the inhibitor, resulting in decreased activity. This interpretation is in agreement with the crystallographic finding that the main difference between free and complexed enzymes is the widening of the hydrophobic channel that accommodates the substrate by means of motion of the N-terminal α-helix (residues 1–8) and residues 20–24, and it supports a previous suggestion by Wery et al.[19] that compounds with smaller cross-sections may be better inhibitors.

After this paper was submitted for publication, a paper appeared by Wheeler et al.[45] on the structure–activity relationships of phospholipid analogues which are substrates of HSF-PLA_2. On the basis of experiments conducted without a membrane surface, the

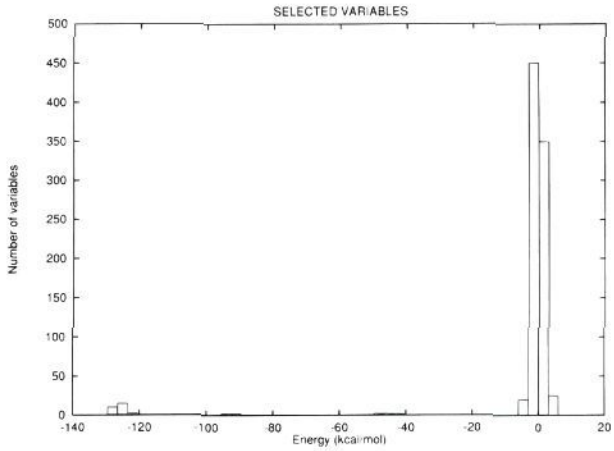

Figure 9. Histogram of the energy distribution of the selected variables for one representative regression model, calculated for all 26 compounds.

authors conclude, in agreement with our results, that the negative charge on the phosphate is necessary for activity. They also indicate that, in compounds in which the *sn*-1 chain has a single methylene connecting the glycerol with aromatic groups, the enzyme has decreased activity. These observations can be interpreted in the light of our regression models, which detect a pocket in the active site of the enzyme consisting of residues Tyr-52, Lys-69, Phe-5, and a crystallographic water molecule, which binds the first portion of the *sn*-1 chain of the inhibitors (the XLM fragment, see Figure 8 and Table 3). If the XLM fragment is too rigid or bulky, a poor fit into the pocket can be expected, leading to decreased activity.

The energy distribution of the variables selected in a typical regression model is shown in Figure 9. The effective energies responsible for the activity differences are generally rather small in magnitude, typically less than 3 kcal mol^{-1}. Such small energy differences imply that the detection of features important for differences in activity by analyzing the magnitude of binding energy components or by visual inspection of the modeled complexes is likely to have limited success. In fact, the structural changes among the different complexes are rather small, with an average rms deviation of the Cα atoms of the protein of about 0.5 Å.

C. Current Limitations. The results reported above show that comparative binding energy analysis can yield predictive models and that mechanistic insights can be obtained. However, it must be borne in mind that the regressions are derived from highly underdetermined equations and that other combinations of X variables may provide predictive models. The selection procedure using fractional factorial designs reduces this problem to some extent, as illustrated by the fact that the most important variables were selected in almost all of the runs performed. The problem is, however, most acute for the electrostatic energy terms, because the dielectric representation used here enhances the collinearity of the electrostatic interactions, as a result of neglecting the anisotropic nature of the dielectric constant in proteins and inadequately screening the charged groups at the protein surface. This is illustrated in Figure 10, where a projection of the **X** matrix in the partial weight space over the plane formed by the first two latent variables is shown. In the full **X**

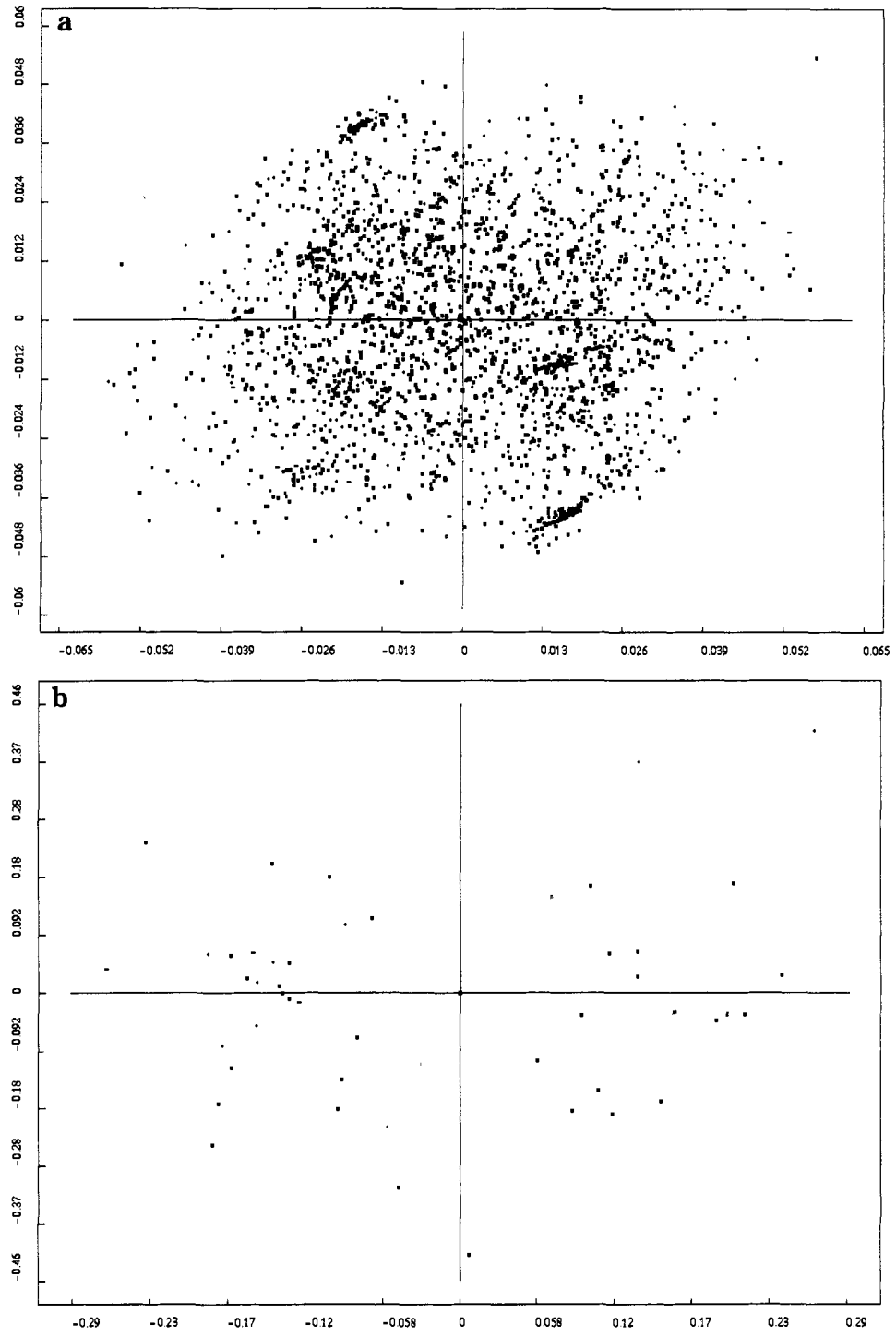

Figure 10. Distribution of the partial weights space of the X variables on the plane formed by the first two latent variables, represented by the axes: (a) before the variable selection procedure, (b) after the variable selection procedure.

matrix with all the variables present (Figure 10a), the clustering of variables in some parts of the space is clear, meaning that, within the first two latent variables, these energy components behave in a collinear fashion with one another. These clusters are formed mainly by electrostatic energy terms. After the variable selection procedure has been performed, these clusters disappear, as from the statistical view point, one or a few variables from the cluster are able to provide the same information as all the variables in the cluster. The selected variables are, therefore, fairly homogeneously distributed on the plane (Figure 10b). Thus, during D-optimal selection, almost *any* of these variables can be selected, although only a few variables from the cluster are *physically* meaningful, and therefore the physically important variables can be lost during the variable selection procedure.

The ZLM fragment provides a good example of the phenomenon described in the previous paragraph. In some inhibitors, this fragment contains the phosphate group in the *sn*-3 chain that interacts with the calcium ion, and this makes a large contribution to the activity, according to the COMBINE analysis (Tables 1 and 3). However, the calcium–ZLM interaction is not selected in all runs. In the cases where this interaction is not selected, different electrostatic interactions that are highly collinear with the calcium–ZLM one, but without clear physical meaning, are selected, such as the cou-

lombic interactions between the ZLM fragment and both Lys-16 and Arg-34. The problem of collinearity could be reduced by using a more realistic electrostatic model, e.g., from numerical solution of the Poisson–Boltzmann equation.[46] Nevertheless, the probability of mistaking some effective energies for physically meaningful terms can be considerably reduced by carrying out multiple runs and carefully examining the occurrence of different effective energies in the selected variables.

Conclusions

Interaction energies based on molecular mechanics calculations are one of the main tools in the study of ligand–receptor interactions.[47] When used with limited conformational sampling of the ligand–macromolecule complex, e.g., with energy minimization techniques, only qualitative conclusions can generally be derived. It is shown here, however, that when a series of related compounds is available, it is possible to relate the differences in activity to differences in components of the ligand binding energies corresponding to interactions of particular parts of the compounds and the receptor. The differences in these energy components are often small and are therefore difficult to detect when computing the total energy of each complex, particularly when compared to the noise introduced into the description of the potential energy as a result of inaccuracies in the molecular force fields and the molecular representation. Nevertheless, the important energy components can be detected by subjecting the molecular mechanics interaction energies to the statistical methods of multivariate analysis (PLS) and advanced variable selection procedures based on the iterative evaluation of the effects of individual energy terms on the model predictivity.

The predictive ability of COMBINE analysis can be expected to be significantly enhanced by improvements in the description of the electrostatic term, the inclusion of suitable descriptors for solvation and entropic effects, and the optimization of particular aspects of the methodology, such as the choice of ligand fragment definitions and the details of the variable selection protocol. Application of this methodology to different ligand–receptor systems is also required in order to validate the generality and usefulness of COMBINE analysis. With these caveats in mind though, it seems safe to say that this method can be expected to be particularly powerful when coupled to methods of positioning molecular fragments in binding sites[48–50] and data-base searching. It is also well-suited to guide the optimization of the pharmacological selectivity of one compound against two different macromolecular targets.

Acknowledgment. We thank Professor Paul B. Sigler for kindly providing us with the coordinates of the HSF-PLA_2–inhibitor complex and Dr. Gabriele Cruciani for his helpful comments on Chemometrics and provision of the GOLPE program. We also acknowledge Dr. Albert Palomer for bringing the problem of structure–activity relationships of PLA_2 inhibitors to our attention. Laboratorios Menarini, S. A. (Badalona, Spain) provided partial support for this research, as well as the experimental data used in this study. A.R.O. is the recipient of a predoctoral fellowship from the Comunidad Autónoma de Madrid.

Appendix

The binding of a ligand (L) to its receptor (R) can be described by the following equilibrium: $L + R \leftrightarrows LR$. Let us denote the unbound species as the system α and the bound species as the system β. In the canonical ensemble, the difference in Helmholtz free energy between these two systems, which corresponds to the ligand binding free energy, is given by[21]

$$\Delta F = F_\beta - F_\alpha = -k_B T \ln \frac{\int_v e^{-U_\beta/k_B T}\, dx_1 \ldots dx_N}{\int_v e^{-U_\alpha/k_B T}\, dx_1 \ldots dx_N} \quad (1a)$$

where $U(x_1, \ldots, x_N)$ is the potential energy function describing the interactions of a system of N atoms. $U(x_1, \ldots, x_N)$ can be considered as the sum of P interactions, u_k, between the N atoms of the system: $U(x_1, \ldots, x_N) = \sum_{k=1}^{P} u_k$.

The COMBINE analysis method is based on two assumptions. These can be considered valid if the statistical analysis results in a predictive model. First, it is assumed that there exists a set of n coefficients, w_i, with $n \ll P$, which allow a new effective potential energy function $U' = \sum_{i=1}^{n} w_i u_i$ to be defined in such a way that

$$\Delta F \sim -k_B T \ln \frac{\int_v e^{-U'_\beta/k_B T}\, dx_1 \ldots dx_N}{\int_v e^{-U'_\alpha/k_B T}\, dx_1 \ldots dx_N} \quad (2a)$$

Next, it is assumed that for both the bound and the unbound species, one conformation can be chosen as representative of the ensemble (note, however, that more than one representative structure could be chosen, if appropriate). In this case

$$\Delta F \sim -k_B T \ln (e^{-\Delta U'/k_B T}) = \Delta U' \quad (3a)$$

$$\Delta F \sim \sum_{i=1}^{n} w_i \Delta u_i^{rep} \quad (4a)$$

where $\Delta U' = U_\beta' - U_\alpha'$, and Δu_i^{rep} are terms of the binding energy calculated for the representative conformations of the bound and unbound species. In chemical and biochemical systems, the most interesting thermodynamic function is not the Helmholtz free energy but the Gibbs free energy. However, the difference is negligible for processes in solution.[21] The key issue is then to obtain the coefficients w_i in eq 4a and to select the n important interactions. This can be done by means of multivariate statistical analysis coupled to variable selection procedures. Thus, the equation for correlating the binding free energy with a set of interactions in the ligand–receptor complex is

$$\Delta G = \sum_{i=1}^{n} w_i \Delta u_i^{rep} + C \quad (5a)$$

where the constant C accounts for systematic contributions to the binding affinity and statistical errors in the fitting process.

A molecular mechanics force field is employed in order to calculate the energy terms Δu_i^{rep}. The total binding energy of a ligand to the receptor can be considered as

the sum of the following terms:

$$\Delta U = E_{lr} + \Delta E_l + \Delta E_r \quad (6a)$$

where E_{lr} is the ligand−receptor interaction energy, ΔE_l is the change in potential energy of the ligand upon formation of the complex, and ΔE_r is the change in the potential energy of the receptor upon formation of the complex. Each of the three terms on the right-hand side of eq 6a can be further subdivided into terms, Δu_i^{rep}, describing interactions for different parts of the ligands and receptors as in eq 2 of the Methods section.

References

(1) (a) Jolles, G., Wooldridge, K. R. H., Eds. *Drug Design: Fact or Fantasy?*; Academic Press: London, 1984. (b) Dean, P. M. *Molecular Foundations of Drug-Receptor Interactions*; Cambridge University Press: Cambridge, 1987.

(2) (a) Navia, M. A.; Murcko, M. A. Use of structural information in drug design. *Curr. Opin. Struct. Biol.* **1992**, *2*, 202. (b) Hol, W. J. Protein crystallography and computer graphics − toward rational drug design *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 767−778. (c) Goodford, P. J. Drug design by the method of receptor fit. J. *Med. Chem.* **1984**, *27* (5), 556−564.

(3) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078−1082.

(4) (a) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **1994**, *37* (8), 1035−1054. (b) Appelt, K.; Bacquet, R. J.; Barlett, C. A.; Booth, C. L. J.; Freer, S. T.; Fuhry, M. A. M.; Gehring, M. R.; Herrmann, S. M.; Howland, E. F.; Janson, C. A.; Jones, T. R.; Kan, C.-C.; Kathardekar, V.; Lewis, K. K.; Marzoni, G. P.; Matthews, D. A.; Mohr, C.; Moomaw, E. W.; Morse, C. A.; Oatley, S. J.; Ogden, R. C.; Reddy, M. R.; Reich, S. H.; Schoettlin, W. S.; Smith, W. W.; Varney, M. D.; Villafranca, J. E.; Ward, R. W.; Webber, S.; Webber, S. E.; Welsh, K. M.; White, J. Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* **1991**, *34* (7), 1925−1934.

(5) (a) Blaney, J. M.; Hansch, C. In *Comprehensive Medicinal Chemistry. Vol. IV: Quantitative Drug Design*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: New York, 1990. (b) Reich, S. H.; Webber, S. E. Structure-Based drug Design (SBDD): Every structure tells a story. *Perspect. Drug Discovery Des.* **1993**, *1*, 371−390.

(6) (a) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492. (b) Straatsma, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(7) van Gunsteren, W. F. In *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM Science Publishers B. V.: Leiden 1989; pp 27−59.

(8) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena *Chem. Rev.* **1993**, *93*, 2395−2417.

(9) Gupta, S. P. QSAR studies on enzyme inhibitors. *Chem. Rev.* **1987**, *87*, 1183−1253.

(10) Kubinyi, H., Ed. *3D-QSAR in Drug Design. Theory, Methods and Applications*; ESCOM Science Publishers B. V.: Leiden, 1993.

(11) Hansch, C.; Klein, T. E. Quantitative structure activity relationships and molecular graphics in evaluation of enzyme ligand interactions. *Methods Enzymol.* **1991**, *202*, 512−543.

(12) (a) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. Molecular mechanics simulation of protein-ligand interactions: Binding of Thyroid Hormone analogues to Prealbumin *J. Am. Chem. Soc.* **1982**, *104*, 6424−6434. (b) Menziani, M. C.; De Benedetti, P. G.; Gago, F.; Richards, W. G. The binding of benzene sulfonamides to carbonic anhydrase enzyme. A molecular mechanics study and quantitative structure-activity relationships. *J. Med. Chem.* **1989**, *32*, 951−956.

(13) (a) García, M. L. Manuscript in preparation. Cabré, F.; Carabaza, A.; García, A. M.; Gómez, M.; García, M. L.; Mauleón, D.; Carganico, G. VIIIth International Congress on Prostaglandins and Related Compounds, Montreal, 1992. (c) Carganico, G.; Mauleón, D.; García, M. L. Spanish Patent Application, ES 9301180.

(14) Pisabarro, M. T.; Ortiz, A. R.; Palomer, A.; Cabré, F.; García, M. L.; Wade, R. C.; Gago, F.; Mauleón, D.; Carganico, G. Rational modification of human synovial fluid phospholipase A_2 inhibitors. *J. Med. Chem.* **1994**, *37*, 337−341.

(15) (a) Dennis, E. A. Phospholipases. In *The Enzymes*; Boyer, P. D., Ed., Academic Press: New York, Vol. 16, pp 307−354. (b) Dennis, E. A. Regulation of eicosanoid production: role of phospholipase and inhibitors. Biotechnology **1987**, *5*, 1294−1300.

(16) Mobilio, D.; Marshall, L. A. Recent advances in the design and evaluation of inhibitors of PLA_2. *Annu. Rep. Med. Chem.* **1989**, *24*, 157−166.

(17) (a) Pruzanski, W.; Vadas, P.; Stefanski, E.; Urowitz, M. B. Phospholipase A_2 activity in sera and synovial fluids in rheumatoid arthritis and osteoarthritis. Its possible role as a proinflammatory enzyme. *J. Rheumatol.* **1988**, *12*, 211−216. (b) Hara, S.; Kudo, I.; Chang, H. W.; Matsuta, K.; Miyamoto, T.; Inoue, K. Purification and characterization of extracellular Phospholipase A_2 from human synovial fluid in rheumatoid arthritis. *J. Biochem.* **1989**, *105*, 395−399. (c) Vadas, P.; Pruzanski, W.; Kim, J.; Fornasier, V. The proinflammatory effect of inter-articular injection of soluble human and venom Phospholipase A_2. *Am. J. Pathol.* **1989**, *134*, 807−811.

(18) Glaser, K. B.; Mobilio, D.; Chang, J. Y.; Senko, N. Phospholipase A_2 enzymes: regulation and inhibition. *Trends Pharmacol. Sci.* **1993**, *14*, 92−98.

(19) Wery, J.-P.; Schevitz, R. W.; Clawson, D. K.; Bobbitt, J. L.; Dow, E. R.; Gamboa, G.; Goodson, T.; Hermann, R. B.; Kramer, R. M.; McClure, D. B.; Mihelich, E. D.; Putname, J. E.; Sharp, J. D.; Stark, D. H.; Teater, C.; Warrick, M. W.; Jones, N. D. Structure of Recombinant Human Rheumatoid Arthritic Synovial Fluid Phospholipase A_2 at 2.2 Å Resolution. *Nature* **1991**, *352*, 79−82.

(20) Scott, D. L.; White, S. P.; Browning, J. L.; Rosa, J. J.; Gelb, M. H.; Sigler, P. B. Structures of Free and Inhibited Human Secretory Phospholipase A_2 from Inflammatory Exudate. *Science* **1991**, *254*, 1007−1010.

(21) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.

(22) (a) Wold, S. Partial Least Squares Analysis. In *3D-QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: Leiden, Holland, 1993. (b) Wold, S.; Wold, H.; Ruhe, A.; Dunn, W. J., III. The collinearity problem in linear regression. The Partial Least Squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comp.* **1984**, *5*, 735−743. (c) Lindberg, W.; Persson, J. A.; Wold, S. Partial Least Squares method for spectrofluorometric analysis of mixtures of humic acid and ligninsulfonate. *Anal. Chem.* **1983**, *55*, 643−648.

(23) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. In *Statistics for Experimenters*; Wiley, J. and Sons: New York, 1978; Chapter 12.

(24) INSIGHT-II (version 2.1.2) 1993. Biosym Technologies, 9685 Scraton Road, San Diego, CA 92121-2777.

(25) Thunnissen, M. M. G. M.; Eiso, A. B.; Kalk, K. H.; Drenth, J.; Dijkstra, B. W.; Kuipers, O. P.; Dijkman, R.; de Haas, G. H.; Verheij, H. M. X-ray structure of Phospholipase A_2 complexed with a substrate-based inhibitor. *Nature* **1990**, *347*, 689−691.

(26) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.

(27) (a) Maple, J. R.; Dinur, U.; Hagler, A. T. Derivation of Force Fields for Molecular Mechanics and Dynamics fron *Ab initio* Energy Surfaces. *Proc. Nat. Acad. Sci. U.S.A.* **1988**, *85*, 5350. (b) Maple, J. R.; Thacher, T. S.; Dinur, U.; Hagler, A. T. Biosym Force Field Research Results in New Techniques for the Extraction of Inter- and Intramolecular Forces. *Chem. Des. Autom. News* **1990**, *5* (9), 5−10.

(28) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217. (b) *QUANTA 3.3 Parameter Handbook* 1992. Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA 02154.

(29) Hopfinger, A. J.; Pearlstein, R. A. Molecular mechanics force-field parameterization procedures. *J. Comput. Chem.* **1984**, *5*, 486−499.

(30) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(31) Stewart, J. J. P. MOPAC: A General Molecular Orbital Package, version 6.0; *QCPE 455*, Quantum Chemistry Program Exchange, Indiana, Bloomington, IN 47405.

(32) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. Accuracy of free energies of hydration for organic molecules from 6-31G*-derived partial charges. *J. Comput. Chem.* **1993**, *14* (10), 1240−1249.

(33) Nicholas, J. B.; Vance, R.; Martin, E.; Burke, B. J.; Hopfinger, A. J. Molecular mechanics valence force field for sulfonamides derived by *ab initio* methods. *J. Phys. Chem.* **1991**, *95*, 9803−9811.

(34) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431–439.
(35) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, UK, 1987.
(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
(37) (a) Guenot, J. M.; Kollman, P. A. Molecular dynamics studies of a DNA-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherichia coli trp* repressor. *Protein Sci.* **1992**, *1*, 1185–1205. (b) Guenot, J. M.; Kollman, P. A. Conformational and energetic effects of truncating nonbonded interactions in an aqueous protein dynamics simulation. *J. Comput. Chem.* **1993**, *14*, 295–311.
(38) Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.
(39) Mitchell, T. J. An algorithm for the construction of 'D-optimal' experimental designs. *Technometrics* **1974**, *16*, 203–210.
(40) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. In *Numerical Recipes*; Cambridge University Press: Cambridge, UK, 1989.
(41) Wold, S. Validation of QSAR's *Quant. Struct.-Act. Relat.* **1991**, *10*, 191–193.
(42) Scott, D. L.; White, S. P.; Otwinowski, Z.; Yuan, W.; Gelb, M. H.; Sigler, P. B. Interfacial catalysis: the mechanism of Phospholipase A_2. *Science* **1990**, *250*, 1541–1546.
(43) Clark, M.; Cramer, R. D., III The probability of chance correlation using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.
(44) (a) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Prespective on Structure, Dynamics and Thermodynamics*; Advances in Chemical Physics; Wiley: New York, 1988; Vol. 71. (b) Elber, R.; Karplus, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science* **1987**, *235*, 318–321.
(45) Wheeler, T. N.; Blanchard, S. G.; Andrews, R. C.; Fang, F.; Gray-Nunez, Y.; Harris, C. O.; Lambert, M. H.; Mehrotra, M. M.; Parks, D. J.; Ray, J. A.; Smalley, T. L. Substrate Specificity in Short-chain Phospholipid Analogs at the Active Site of Human Synovial Phospholipase A_2. *J. Med. Chem.* **1995**, *37*, 4118–4129.
(46) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(47) Seibel, G. L.; Kollman, P. A. Molecular Mechanics and the modeling of drug structures, in *Comprehensive Medicinal Chemistry. Vol. IV: Quantitative Drug Design*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: New York, 1990; pp 125–138.
(48) (a) GRIN, GRID and GRAB (version 11), 1994. Molecular Discovery Ltd., West Way House, Elms Parade, Oxford OX2 9LL, England. (b) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important molecules. *J. Med. Chem.* **1985**, *28*, 849–857. (c) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules pf known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094. (d) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140–147. (e) Wade, R. C.; Goodford, P. J. Further development of hydrogen-bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.
(49) (a) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288. (b) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementary as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31*, 722–729. (c) Shoichet, B. K.; Stroud, R. M.; Santi, D. N.; Kuntz, I. D.; Perry, K. M. Structure-based discovery of inhibitors of thymidilate synthase. *Science* **1993**, *259*, 1445–1450.
(50) Böhm, H. J. The computer program LUDI: A new simple method for the de-novo design of enzyme inhibitors. *J. Comput.-Aided. Mol. Des.* **1992**, *6*, 61–78.

JM9407318