Acta Cryst. (1995). D**51**, 560–566

Correlation of Binding Affinities with Non-bonded Interaction Energies of Thrombin–Inhibitor Complexes

BY PETER D. J. GROOTENHUIS AND PHILIP J. M. VAN GALEN

Department of Computational Medicinal Chemistry, NV Organon, The Netherlands

(*Received* 6 *July* 1994; *accepted* 14 *October* 1994)

Abstract

Several empirical modeling protocols are evaluated allowing a quantification of the interaction between an enzyme and a series of inhibitors. The evaluation and optimization of the modeling protocols used a database of 35 non-covalently bound inhibitors of human thrombin. Intermolecular interaction energies were calculated with *CHARMm* after energy minimization of the modeled complexes using various dielectric functions and constraining strategies. These calculated binding energies were correlated with the experimentally obtained binding constants of the inhibitors. The best protocols resulted in linear correlations with correlation coefficients > 0.80. In the best protocols the enzyme was fully constrained, the ligand was allowed to move freely and electrostatics were scaled down drastically or fully neglected during the energy minimization. For the interaction energy evaluation step, a distance-dependent dielectric function $\varepsilon = R$ proved to be optimal. This simple empirical protocol, that neglects solvation or entropy effects, can be implemented readily in other force field packages and may be applied on various enzyme–inhibitor complexes, providing a tool for the evaluation and rank-ordering of newly designed inhibitors.

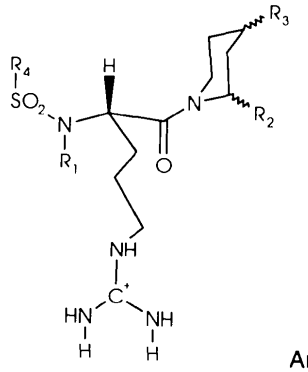
Introduction

Three-dimensional structures of proteins (*e.g.* generated by X-ray crystallography or NMR measurements) are becoming increasingly important in the design of novel drugs (Greer, Erickson, Baldwin & Varney, 1994). They enable the medicinal chemist to directly inspect those structural properties of the target protein that are essential for interaction with an enzyme inhibitor (or receptor ligand), in principle allowing for a rationalization of the design process sometimes referred to as structure-based design (Kuntz, 1992; Bugg, Carson & Montgomery, 1993; Verlinde & Hol, 1994). New leads may originate from modification of known inhibitors, three-dimensional database searches or *de novo* methods (Kuntz, Meng & Shoichet, 1994; Rotstein & Murcko, 1993). In particular the two last approaches typically result in many (10–1000) candidate molecules that need to be evaluated. The choice of which compounds will actually be synthesized often depends on the chemist's intuition, as well as on synthetic feasibility. Thus, there is a clear need for an objective method to rank suitable candidates for synthesis, preferably a method that will predict binding affinities in a quantitative way. Ideally, other properties like selectivity and various pharmacokinetic parameters would also be taken into account. This is currently a considerable scientific challenge. With respect to the calculation of binding affinities, substantial progress has been made with free-energy perturbation methods (Kollman, 1993). However, these methods are less suited for dissimilar compounds and are computationally very intensive, and therefore less practical in an industrial setting, where the synthetic chemist requires fast feedback from the modeling department. Furthermore, the general predictive (as opposed to retrospective) power of this method has, to the best of our knowledge, not been documented. Partial least-squares methods, *e.g.* as implemented in *CoMFA* (Cramer, Patterson & Bunce, 1988), are also promising, at least with regard to optimization within a series of related compounds (Grootenhuis & van Helden, 1994; Folkers, Merz & Rognan, 1993).

Apart from more rigorous approaches towards estimating the binding between receptors and ligands (Williams, Searle, Mackay, Gerhard & Maplestone, 1993), empirical approaches have been followed by various groups (*e.g.* Kuntz, Meng & Shoichet, 1994; Blaney & Dixon, 1993). Such approaches try to estimate the binding affinity of a ligand directly from the interaction energy with its receptor, as calculated by simplified, often grid-based molecular mechanics potentials. Here, we evaluate various *CHARMm*-based modeling protocols for the calculation of interaction energies, and the correlation of these interaction energy values with experimentally determined binding affinities.

As a test case for this approach, we have used the crystal (or modeled) structures of thrombin–inhibitor complexes. Thrombin is a serine protease that plays a central role in the coagulation cascade, and inhibitors of this enzyme are potential antithrombotic drugs (Berliner, 1992). The mechanism of action of thrombin and related serine proteases is well understood, and a considerable number of crystal structures of the enzyme in complex

ISSN 0907-4449



Argatroban series

Compound	R_1	R_2	R_3	R_4	pK_i
(1)	H	COOH	Me	1,2,3,4-tetrahydro-8-quinolinyl	7.72
(2)	H	COOH	Me (S)	1,2,3,4-tetrahydro-8-quinolinyl	6.62
(3)	H	COOH (S)	Me	1,2,3,4-tetrahydro-8-quinolinyl	5.72
(4)	H	COOH (S)	Me (S)	1,2,3,4-tetrahydro-8-quinolinyl	3.55
(5)	H	COOMe	Me	1,2,3,4-tetrahydro-8-quinolinyl	7.85
(6)	H	COOH	H	1,2,3,4-tetrahydro-8-quinolinyl	6.51
(7)	H	COOH (S)	H	1,2,3,4-tetrahydro-8-quinolinyl	4.24
(8)	H	H	Me	1,2,3,4-tetrahydro-8-quinolinyl	7.52
(9)	H	H	H	1,2,3,4-tetrahydro-8-quinolinyl	6.88
(10)	H	COOH	Me	2-methoxynaphthyl	6.70
(11)	Me	COOH	Me	2-methoxynaphthyl	4.22
(12)	H	H	Me	2-methoxynaphthyl	7.14
(13)	H	H	Ph	2-methoxynaphthyl	4.81
(14)	H	H	Me	naphthyl	6.70
(15)	H	H	Ph	naphthyl	4.70
(16)	H	H	Et	5-dimethylamino-1-naphthyl	7.70
(17)	H	H	Ph	5-dimethylamino-1-naphthyl	5.30

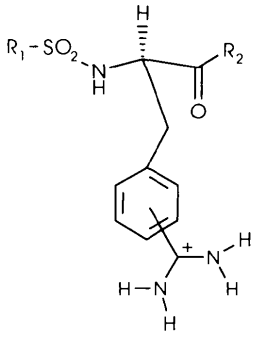
(*a*)

with various inhibitors are available (Stubbs & Bode, 1993). The binding site of thrombin is a well defined groove with both polar and apolar sites, making it an ideal target for structure-based drug design.

Computational methods

We have generated a database of 35 non-covalently bound thrombin inhibitors taken from the literature, comprising three structurally different classes of inhibitors [Figs. 1(*a*), 1(*b*) and 1(*c*)]. The 35 inhibitors exhibit a wide range of binding affinities towards thrombin with pK_i's falling in the range between 3 and 8. Experimentally determined structures for the parent compounds in these series (*viz.* Argatroban, NAPAP and 4-TAPAP) have been published (Bode, Turk & Stürzebecher, 1990; Banner & Hadváry, 1991; Brandstetter *et al.*, 1992). Fig. 2 gives a schematic overview of the interactions that contribute to the high affinity of these inhibitors.

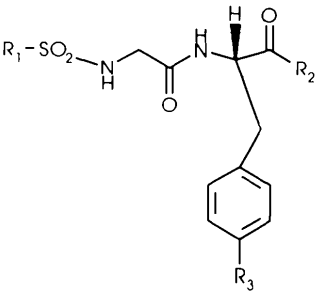
The structures of the 32 remaining derivatives were modeled. Energy minimization and interaction energy evaluations were carried out using *CHARMm*. Variables that were studied include the flexibility of the system (by putting positional and/or harmonic constraints on the enzyme) and the relative contribution of van der Waals and electrostatic components to the non-bonded interaction energies. A preliminary account of some of the results presented here with a smaller database of Argatroban derivatives only, has been published elsewhere (Grootenhuis & van Helden, 1994).



TAPAP series

Compound	R_1	R_2	Amidino	pK_i
(18)	*p*-toluyl	piperidinyl	*meta*	6.77
(19)	α-naphthyl	piperidinyl	*meta*	6.96
(20)	β-naphthyl	piperidinyl	*meta*	7.04
(21)	*p*-toluyl	piperidinyl	*para*	6.19
(22)	α-naphthyl	piperidinyl	*para*	6.26
(23)	β-naphthyl	piperidinyl	*para*	6.68
(24)	*p*-toluyl	morpholinyl	*para*	5.57
(25)	*p*-toluyl	pyrrolidinyl	*para*	5.54
(26)	*p*-toluyl	phenyl	*para*	4.28

(*b*)



NAPAP series

Compound	R_1	R_2	R_3	pK_i
(27)	β-naphthyl	piperidinyl	amidino	8.15
(28)	β-naphthyl	piperidinyl	amino	4.03
(29)	β-naphthyl	piperidinyl	aminomethyl	5.02
(30)	β-naphthyl	piperidinyl	cyano	4.65
(31)	β-naphthyl	morpholinyl	amidino	6.94
(32)	β-naphthyl	piperidinyl	amidino	8.52
(33)	β-naphthyl	piperidinyl	*N*-hydroxyamidino	6.74
(34)	β-naphthyl	pyrrolidinyl	amidino	8.22
(35)	*p*-toluyl	piperidinyl	amidino	7.62

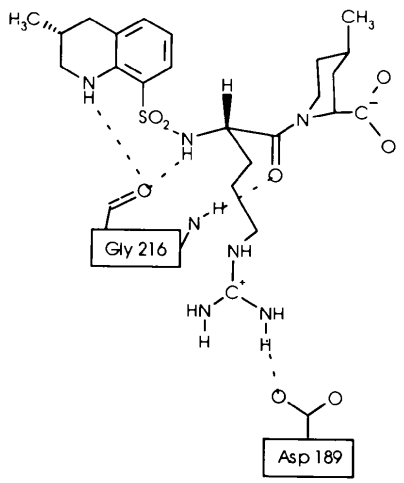
(*c*)

Fig. 1. (*a*) Structures and pK_i values for Argatroban-based inhibitors used in this study. Data taken from Kikumoto *et al.* (1984), Hijikata-Okunomiya *et al.* (1987) and Stürzebecher, Vieweg, Wikström, Turk & Bode (1992). (*b*) Stuctures and pK_i values for TAPAP-based inhibitors used in this study. Data taken from Brandstetter *et al.* (1992) and Stürzebecher, Walsmann, Voigt & Wagner (1984); Stürzebecher *et al.* (1987); Stürzebecher *et al.* (1989); Stürzebecher *et al.* (1992); Stürzebecher, Prasa & Taby (1993). (*c*) Structures and pK_i values for NAPAP-based inhibitors used in this study. Data taken from Stürzebecher, Markwardt, Walsmann, Voigt & Wagner (1983); Stürzebecher *et al.* (1984); Stürzebecher *et al.* (1988).

Crystal structures of the complexes between thrombin and Argatroban, 4-TAPAP and NAPAP were obtained from Dr W. Bode (Max Planck Institut, Martinsried, Germany) and are also available from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). Water molecules were removed from the complex structures and His57 was left unprotonated. H atoms were included on the protease at calculated positions using the *HBUILD* module present in *CHARMm* 22.2 (Molecular Simulations Inc., 1994). The active site was defined by the amino-acid residues 16–17, 41–43, 55–60F, 94–102, 142–148, 158–160, 172–176, 180–184, 187–197 and 212–229 (chymotrypsin numbering).

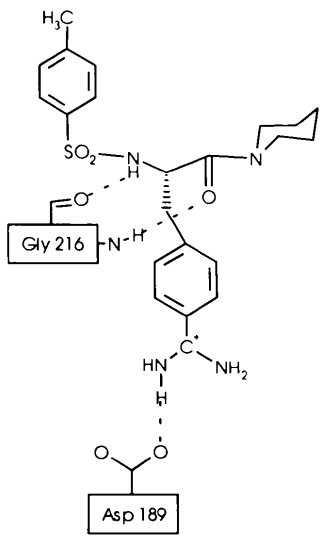
The three-dimensional molecular editor of *QUANTA* Versions 3.3.1 and 4.0 (Molecular Simulations Inc., 1994) was used to construct the various thrombin inhibitors, starting from the most closely related crystal structure. Criteria for the selection of the thrombin inhibitors were: (i) they should be non-covalent inhibitors; (ii) the experimentally determined K_i's should be available; and (iii) it should be possible to readily model the inhibitor from one of the parent compounds (*i.e.* the binding mode is expected to be similar and conformational changes or uncertainties in the modeled compound should be minimal). Partial atomic charges for the inhibitors were calculated by the charge-template method implemented in *QUANTA*'s molecular editor; residual charges were put on the C and non-polar H atoms of the inhibitor.

The (unminimized) inhibitors were transferred into the binding site in the same position and orientation as the parent inhibitors. The 35 complexes thus obtained were energy minimized with *CHARMm* 22.2 using the adopted-basis Newton–Raphson algorithm. The all-atom force field and parameters as implemented in *QUANTA* 4.0 were used. The non-bonded cut-off amounted to 15 Å; switching (van der Waals) and shifting (electrostatics) functions were turned on between 11 and 14 Å. The default heuristic non-bonded list-update method was used. Energy minimizations continued until the root-mean-square energy gradient was less than 0.10 kcal $Å^{-1}$. A number of different settings for the dielectric function and/or positional constraints were applied (Tables 1 and 2). Subsequent calculation of the interaction energies between enzyme and inhibitor was performed separately.



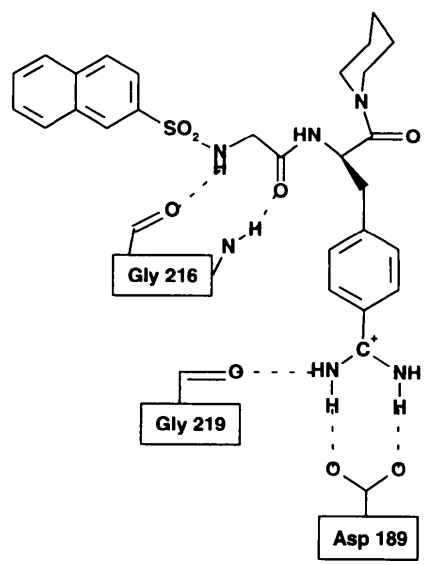
Argatroban (1)

(*a*)



4-TAPAP (23)

(*b*)



NAPAP (27)

(*c*)

Fig. 2. Interactions of Argatroban, NAPAP and 4-TAPAP with thrombin, as observed in the crystal structures of the protease-inhibitor complexes (Bode, Turk & Stürzebecher, 1990; Banner & Hadváry, 1991; Brandstetter *et al.*, 1992).

Table 1. *Correlation between the calculated and experimentally determined pK_i values: influence of the flexibility of the protein*

Positional constraints were put on the bulk protease and/or the active site. In all cases, the inhibitor was left unconstrained, and a dielectric function of $\varepsilon = 4R$ was used during energy minimization. The interaction energy evaluation was performed using $\varepsilon = R$. For the rationale behind the dielectric functions used, *vide infra* $N = 32$.

	Constraints on bulk protease	Constraints on active site (kcal Å^{-1})	r	r^2_{cross}	$F_{1,30}$
1	No constraints	No constraints	0.54	0.18	12.1
2	Fixed	No constraints	0.60	0.28	16.8
3	Fixed	Harmonic constraint 2.5	0.72	0.46	32.7
4	Fixed	Harmonic constraint 5.0	0.76	0.52	40.3
5	Fixed	Harmonic constraint 10.0	0.74	0.49	36.3
6	Fixed	Fixed	0.80	0.59	55.2

Table 2. *Correlation between calculated and experimentally determined pK_i values: balance of van der Waals and electrostatic contributions*

In all cases, bulk protease and active site are fixed, whereas the inhibitor is left unconstrained. For the rationale behind this, *vide supra* $N = 32$.

	Minimization	Evaluation	r	r^2_{cross}	$F_{1,30}$
7	Only vdW	Only vdW	0.26	−0.03	2.2
8	Only vdW	$\varepsilon = R$	0.81	0.60	57.4
9	$\varepsilon = R$	$\varepsilon = R$	0.73	0.47	34.8
10	$\varepsilon = 4R$	$\varepsilon = 4R$	0.64	0.34	21.2
11	$\varepsilon = 4R$	$\varepsilon = 1$	0.68	0.39	26.1

The non-bonded potential-energy equation used in *CHARMm* (Brooks *et al.*, 1983) has the general form,

$$E_{nonb} = \sum(A/R^{12} - B/R^6 + q_1q_2/DR),$$

in which R denotes the interparticle distance, A and B the Lennard–Jones parameters, q_1 and q_2 the partial atomic charges, and D the effective dielectric function.

Statistical analyses were performed by carrying out linear-regression analyses of the calculated interaction energies *versus* the experimental pK_i values; by using these linear correlations calculated pK_i values were derived for each compound. The experimentally determined K_i values of the Argatroban derivatives were taken directly from the literature. Since the binding data for the NAPAP and TAPAP series of compounds were determined on racemic mixtures, the K_i's were divided by 2, assuming that all activity is due to one enantiomer only. Evaluations were based on the correlation coefficients r, F values and cross-validated r^2_{cross} values. The latter values were calculated by a leave-one-out cross-validation procedure as suggested and explained by Cramer, DePriest, Patterson & Hecht (1993). They reflect the predictive power of the model. It is accepted that when $r^2_{cross} > 0.3$ chance correlations are virtually ruled out and useful models result; perfect predictions would give $r^2_{cross} = 1$ (Clark, Cramer, Jones, Patterson & Simeroth, 1990). From its definition, r^2_{cross} is negative when the model performs less than when the mean affinity would be used as prediction for each member of the test set.

Results

They influence of the flexibility of the protein on the correlations between calculated and experimental binding data was studied by putting positional constraints on the bulk protease and/or the active site during the minimization (Table 1).

In all cases, the inhibitor was left unconstrained. Since it soon became clear that three NAPAP derivatives [(28), (29), (30)] lacking the amidinium moiety that interacts with Asp189 of the specificity pocket are outlyers in all protocols studied, we initially focussed on the 32 remaining inhibitors. From Table 1, it is obvious that no constraints at all on bulk protease and active site (protocol 1) yields the poorest correlation between the calculated interaction energies and experimentally determined pK_i values. Increasing the number of fixed or harmonically constrained atoms generally improves the correlation, the best being protocol 6. In this protocol only the inhibitor is optimized in the context of a fixed protease: the correlation coefficient is 0.80 while the r^2_{cross} value of 0.59 indicates a useful model. Since some of the compounds [*e.g.* (13) and (15)] have large substituents that may induce changes in active-site residues, we anticipated that a harmonically constrained active site could be advantageous. Although this turned out to be the case for the compounds mentioned, it can be deduced from Table 1 that harmonic constraints on the active site in combination with a fixed bulk protease (protocols 3–5) in general do not give improvement relative to a fixed active site (protocol 6).

Interaction energies as calculated by *CHARMm* consist of van der Waals and electrostatic contributions. The balance between the two was studied by employing a number of different values for the dielectric function including $\varepsilon = 1$ (vacuum) and the distance-dependent functions $\varepsilon = R$ and $\varepsilon = 4R$. *CHARMm* also allows for different values of ε to be used during the energy minimization and the phase in which the final interaction energies are calculated. In Table 2 it is shown that not taking electrostatics into account during both minimization and evaluation (protocol 7) gives a poor correlation.

Much better correlations are found when electrostatics are taken into account during the evaluation step, albeit that downplaying their influence by using a distance-dependent dielectric constant ($\varepsilon = R$, protocol 10) gives better results than when calculations are performed *in vacuo* ($\varepsilon = 1$, protocol 11). Surprisingly, during the minimization, the contribution of electrostatics should be greatly reduced ($\varepsilon = 4R$, protocol 10) or not taken into account at all (8) for optimal correlation. Protocol 8 ($E_{inter} = -4.23pK_i - 54.74$, $N = 32$) gave the best overall correlation and the highest predictive value as indicated by the correlation coefficient and cross-

Table 3. *Correlation between calculated and experimentally determined pK_i values using protocol 8 on various subsets*

Subset	N	r	r^2_{cross}	$F_{1,N-2}$	Standard deviation
All compounds	35	0.64	0.30	22.4	1.45
All amidines/guanidines	32	0.81	0.60	57.0	0.97
Argatroban series	17	0.77	0.46	22.4	0.96
NAPAP series	9	0.49	−0.63	2.2	2.44
TAPAP series	9	0.72	0.07	7.5	0.86

Table 4. *Energy decomposition and differences between experimentally determined and calculated pK_i values using the correlation with protocol 8 and $N = 32$*

E_{tot}, E_{vdw} and E_{elec} denote the calculated total interaction energy, van der Waals and electrostatic components.

Compound	E_{tot} (kcal)	E_{elec} (kcal)	E_{vdw} (kcal)	pK_i (exp)	pK_i (calc)	Difference
(1)	−89.1	−34.0	−55.1	7.72	8.13	−0.41
(2)	−82.2	−29.2	−52.9	6.62	6.49	0.13
(3)	−78.9	−26.3	−52.6	5.72	5.72	0.00
(4)	−76.4	−24.1	−52.3	3.55	5.14	−1.58
(5)	−86.0	−29.4	−56.6	7.85	7.40	0.45
(6)	−83.9	−31.9	−52.0	6.51	6.90	−0.40
(7)	−75.8	−26.3	−49.5	4.24	4.99	−0.75
(8)	−84.1	−32.5	−51.5	7.52	6.94	0.58
(9)	−77.0	−28.8	−48.2	6.88	5.26	1.62
(10)	−82.5	−25.6	−56.9	6.70	6.58	0.12
(11)	−75.8	−24.6	−51.2	4.22	4.97	−0.75
(12)	−78.7	−25.7	−53.0	7.14	5.66	1.48
(13)	−73.1	−18.8	−54.3	4.81	4.34	0.48
(14)	−83.9	−32.5	−51.4	6.70	6.91	−0.21
(15)	−65.3	−16.0	−49.3	4.70	2.50	2.20
(16)	−87.6	−30.4	−57.2	7.70	7.77	−0.08
(17)	−72.1	−17.9	−54.3	5.30	4.12	1.18
(18)	−89.6	−40.9	−48.7	6.77	8.25	−1.48
(19)	−84.1	−38.4	−45.7	6.96	6.94	0.02
(20)	−86.0	−41.1	−44.9	7.04	7.40	−0.36
(21)	−79.1	−29.6	−49.5	6.19	5.76	0.42
(22)	−77.5	−29.2	−48.3	6.26	5.38	0.88
(23)	−78.3	−28.2	−50.1	6.68	5.57	1.10
(24)	−79.0	−29.3	−49.7	5.57	5.74	−0.17
(25)	−76.2	−28.5	−47.7	5.54	5.07	0.47
(26)	−74.9	−25.7	−49.2	4.00	4.78	−0.78
(27)	−92.7	−39.7	−53.0	8.15	8.97	−0.82
(28)	−92.9	−41.7	−51.3	4.03	9.04	−5.00
(29)	−86.1	−31.5	−54.6	5.02	7.42	−2.40
(30)	−59.4	−9.2	−50.2	4.65	1.10	3.55
(31)	−92.9	−39.0	−53.9	6.94	9.03	−2.09
(32)	−93.8	−39.1	−54.6	8.52	9.24	−0.72
(33)	−80.9	−26.2	−54.7	6.74	6.19	0.55
(34)	−90.2	−38.3	−52.0	8.22	8.39	−0.17
(35)	−90.9	−40.2	−50.7	7.62	8.55	−0.93

validation statistics of 0.81 and 0.60, respectively. The standard deviation is just below 1 (0.97) corresponding with an estimated error of a factor of approximately 10 in the K_i values. Protocol 8 takes 2–5 c.p.u. min per compound on a Silicon Graphics Indigo equipped with a 75/150 MHz R4400 processor. An analysis of the results on the various subsets using protocol 8 is given in Table 3.

For the larger subsets with $N \geq 17$ useful models are obtained while for the NAPAP ($N = 9$) and TAPAP ($N = 9$) subsets the correlations in terms of r^2_{cross} are not significant. The poor performance of the NAPAP series is mainly due to the three compounds, (28), (29) and (30), that lack the amidine moiety. The differences between experimentally determined and calculated pK_i values for all 35 compounds, as calculated with the presently preferred protocol 8 is listed in Table 4 and plotted in Fig. 3.

Five compounds deviate more than two standard deviations from the experimental K_i values including the NAPAP derivatives (28), (29) and (30), which all lack an amidinium or guanidinium moiety. It is tempting to speculate that these compounds have desolvation energies significantly different from the amidinium/guanidinium compounds, accounting for the gross errors in the estimated pK_i values of these compounds, since no solvation term is included in the present protocol. Argatroban derivative (15) contains a bulky substituent that is forced into a small pocket, which may explain the poor prediction. The morpholinyl analogue [(31)] of NAPAP is also seriously overestimated by protocol 8. However, the rather poor activity reported for compound (31) is hard to rationalize. Although the more hydrophilic morpholinyl moiety is less suited to be positioned in the hydrophobic S_2 pocket than the corresponding piperidinyl moiety of NAPAP, this effect is much smaller in the TAPAP series [compare (21) and (24)].

Apart from the outliers the agreement between calculation and experiment is reasonable and it is tempting to start interpreting some of the observed structure–activity relationships in terms of the calculated interaction energies. Thus, we note that the observed differences in binding of the four Argatroban derivatives (1), (2), (3) and (4) may be explained by electrostatic rather than van der Waals interactions. In contrast to the modeling experiments of Banner (Banner & Hadváry, 1991), we find that no drastic conformational changes in

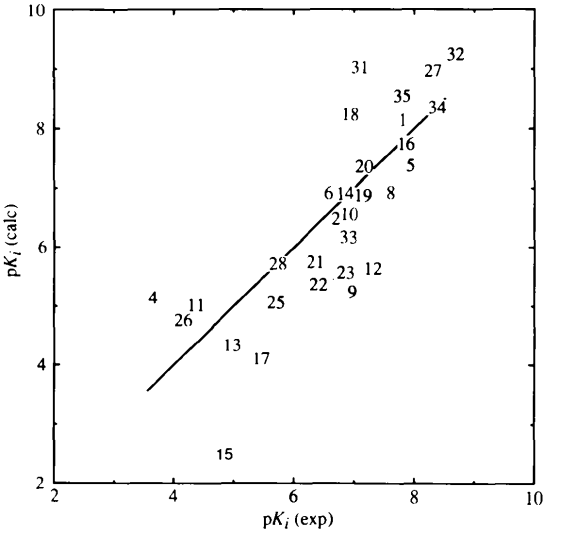


Fig. 3. Plot of the experimental *versus* the calculated binding constants (protocol 8, $N = 32$).

the pipacolic acid moiety are needed in order to qualitatively reproduce the observed binding differences in derivatives (1)–(4).

Discussion

We are aware that the current approach is subject to many assumptions and shortcomings. For instance, it is assumed that the mode of binding in the protease–inhibitor complexes has been modeled correctly, and conformational stress in the ligands has been neglected. Furthermore, entropy, polarization and (de)solvation effects were not included. Despite all this, the results obtained here are likely to be relevant for other systems as well.

From the results with the use of positional constraints it is evident that – at least in the case of thrombin – the best results are obtained when fairly large constraints are put on the active site and the rest of the protease is kept fixed in the conformation observed in the crystal structure. In other words, it appears that the ligand should adapt its conformation to the shape of the protein, rather than that the protein undergoing a conformational change in order to accommodate the ligand. Dixon has recently studied a large number of receptor–ligand complexes for which crystallographic data is available in the Protein Data Bank (Dixon, 1994). He found that in the majority of the cases, the conformation of the host molecule changed very little upon binding of the ligand, fully in agreement with our finding of a lack of induced fit in thrombin. Although we have obtained similar results with another serine protease (factor Xa) and a series of inhibitors (Grootenhuis, 1994), it is conceivable that the protocol should be adapted in individual cases where host molecules are known to undergo induced fit, *e.g.* HIV proteases.

The results obtained with the use of various dielectric constants suggest that the driving force in the formation of the ligand–receptor complex (*i.e.* during the energy minimization), is dominated by van der Waals interactions. However, the contribution of electrostatic interactions to the overall interaction energy calculated in the subsequent evaluation step is quite important, a distance-dependent dielectric function $\varepsilon = R$ giving the best results.

It has been pointed out by Kuntz that rank ordering binding energies using force fields and empirical energy functions is difficult, and in his experience any degree of accuracy can only be reached within a family of closely related compounds with little conformational flexibility (Kuntz, Meng & Shoichet, 1994). Although our protocol is obviously still an oversimplification, for most compounds in three structurally diverse groups of thrombin inhibitors good agreement between calculated and experimentally determined affinity is found.

There are, however, a few cases in which the error in the estimated pK_i is rather large, most notably when the amidino group of NAPAP is replaced. This suggests that some further modifications or additions will have to be made. Two major contributors to $\Delta G_{binding}$ are currently not taken into account: the loss of entropy of the ligand upon binding and the influence of solvation and desolvation. For both contributions approximate methods are available in principle. The loss of entropy can be approximated by imposing a certain penalty for every rotatable bond in the ligand. Böhm (1994) has recently used this approach for the empirical scoring function that is used in the *de novo* design program *LUDI* (*vide infra*). His estimate is that the loss of one rotatable bond will cost $1.4\,kJ\,mol^{-1}$. The errors in the calculated pK_i values of the amidino-modified compounds however, are more likely the result of neglecting the influence of solvation/desolvation energy. Unfortunately, the effect of (de)solvation can't be implemented easily. In our hands, preliminary investigations using Eisenberg's solvation function (Wesson & Eisenberg, 1992) in the *CHARMm* calculations yielded no significant improvement (data not shown). However, it is noted that Horton & Lewis reported promising results for a series of 24 protein complexes when the Eisenberg solvation of the polar and apolar atoms was fitted separately to the free energies of binding (Horton & Lewis, 1991).

Böhm (1994) has very recently developed a five-parameter scoring function that accounts for hydrogen bonds, ionic interactions, lipophilic protein–ligand contact surface and the number of rotatable bonds in the ligand. Böhm's scoring function does not include an explicit solvation term. Applied on a data set of 45 (mostly experimentally studied) protein–ligand complexes the correlation between calculated and experimental binding energies results in a correlation coefficient of 0.87 ($N = 45$, $F = 32.1$, $r^2_{cross} = 0.67$; error in K_i approximately a factor of 25). Since the calibration databases and philosophies behind Böhm's and our method are different, a critical evaluation is hard to make at this point, although it appears that the predictive value and computational efficiency of Böhm's scoring function is higher than ours. In addition it has been shown to be applicable to a wide variety of protein–ligand complexes. However, we note that our approach has no need for special software, but can be readily implemented in force-field packages that have non-bonded potential-energy functions similar to the one used in *CHARMm*.

Concluding remarks

Here, we document the usefulness of a straightforward energy-minimization protocol for the (semi-quantitative) prediction of the binding affinity of a ligand for a protein. In the case of the blood coagulation enzyme thrombin, the optimal protocol only allows for an optimization of the ligand in the context of a fixed protease, while neglecting electrostatics. During the final interaction

energy evaluation a distance-dependent dielectric function was shown to be optimal. Although this method needs further improvement, we feel that it may be useful in prioritizing candidates for synthesis.

The authors thank Dr S. P. van Helden for his assistance with performing the statistical analyses.

Note added in proof: Recently Kurinov & Harrison have described a related approach towards predicting the binding between trypsin and 18 different compounds (Kurinov & Harrison, 1994).

References

BANNER, D. W. & HADVÁRY, P. (1991). *J. Biol. Chem.* **30**, 20085–20093.

BERLINER, L. J. (1992). Editor. *Thrombin, Structure and Function.* New York: Plenum Press.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BLANEY, J. M. & DIXON, J. S. (1993). *Persp. Drug Disc. Des.* **1**, 301–319.

BODE, W. T., TURK, D. & STÜRZEBECHER, J. (1990). *Eur. J. Biochem.* **193**, 175–182.

BÖHM, H. J. (1994). *J. Comput. Aid. Mol. Des.* **8**, 243–256.

BRANDSTETTER, H., TURK, D., HOEFFKEN, W., GROSSE, D., STÜRZEBECHER, J., MARTIN, P. D., EDWARDS, B. F. P. & BODE, W. (1992). *J. Mol. Biol.* **226**, 1085–1099.

BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S. & KARPLUS, M. (1983). *J. Comput. Chem.* **4**, 187–217.

BUGG, C. E., CARSON, W. M. & MONTGOMERY, J. A. (1993). *Sci. Am.* pp. 60–66.

CLARK, M., CRAMER, R. D. III, JONES, D. M., PATTERSON, D. E. & SIMEROTH, P. E. (1990). *Tetrahedron Comput. Methodol.* **3**, 47–59.

CRAMER, R. D. III, DEPRIEST, S. A., PATTERSON, D. E. & HECHT, P. (1993). *3D QSAR in Drug Design. Theory, Methods and Applications*, edited by J. KUBINYI, pp. 443–485. Leiden: ESCOM.

CRAMER, R. D. III, PATTERSON, D. E. & BUNCE, J. D. (1988). *J. Am. Chem. Soc.* **110**, 5959–5967.

DIXON, S. (1994). Sixth Cyprus Conference on New Methods in Drug research, Limasol, Cyprus, 8–14 May, 1994.

FOLKERS, G., MERZ, A. & ROGNAN, D. (1993). *3D QSAR in Drug Design. Theory, Methods and Applications*, edited by J. KUBINYI, pp. 583–618, Leiden: ESCOM.

GREER, J., ERICKSON, J. W., BALDWIN, J. J. & VARNEY, M. D. (1994). *J. Med. Chem.* **37**, 1035–1054.

GROOTENHUIS, P. D. J. (1994). Unpublished results.

GROOTENHUIS, P. D. J. & VAN HELDEN, S. P. (1994). *Computational Approaches in Supramolecular Chemistry*, edited by G. WIPFF, pp. 137–149. Dordrecht: Kluwer Academic Publishers.

HORTON, N. & LEWIS, M. (1991). *Protein Sci.* **3**, 169–181.

HIJIKATA-OKUNOMIYA, A., OKAMOTO, S., KIKUMOTO, R., TAMAO, Y., OHKUBO, K., TEZUKA, T., TONOMURA, S. & MATSUMOTO, O. (1987). *Thromb. Res.* **45**, 451–462.

KIKUMOTO, R., TAMAO, Y., TEZUKA, T., TONOMURA, S., HARA, H., NINOMIYA, K., HIJIKATA, A. & OKAMOTO, S. (1984). *Biochemistry*, **23**, 85–90.

KOLLMAN, P. (1993). *Chem. Rev.* **93**, 2395–2417.

KUNTZ, I. D. (1992). *Science*, **257**, 1078–1082.

KUNTZ, I. D., MENG, E. C. & SHOICHET, B. K. (1994). *Acc. Chem. Res.* **27**, 117–123.

KURINOV, I. V. & HARRISON, R. W. (1994). *Nature Struct. Biol.* **1**, 735–743.

Molecular Simulations Inc. (1994). *CHARMm* and *QUANTA*. Molecular Simulations Inc., Burlington, MA, USA.

ROTSTEIN, S. H. & MURCKO, M. A. (1993). *J. Med. Chem.* **36**, 1700–1710.

STUBBS, M. T. & BODE, W. (1993). *Thromb. Res.* **69**, 1–58.

STÜRZEBECHER, J., HORN, H., WALSMANN, P., VOIGHT, B., MARKWARDT, F. & WAGNER, G. (1988). *Pharmazie*, **43**, 782–783.

STÜRZEBECHER, J., MARKWARDT, F., WALSMANN, P., VOIGT, B. & WAGNER, G. (1983). *Thromb. Res.* **29**, 635–642.

STÜRZEBECHER, J., PRASA, D. & TABY, O. (1993). *Thromb. Res.* **69**, 533–539.

STÜRZEBECHER, J., STÜRZEBECHER, U., VIEWEG, G., WAGNER, J., HAUPTMANN, J. & MARKWARDT, F. (1989). *Thromb. Res.* **54**, 245–252.

STÜRZEBECHER, J., VIEWEG, H., WIKSTRÖM, P., TURK, D. & BODE, W. (1992). *Biol. Chem. Hoppe-Seyler*, **373**, 491–496.

STÜRZEBECHER, J., VOIGT, B. & WAGNER, G., WALSMANN, P., VOIGT, B. & WAGNER, G. (1987). *Pharmazie*, **42**, 114–116.

STÜRZEBECHER, J., WALSMANN, P., VOIGT, B. & WAGNER, G. (1984). *Thromb. Res.* **36**, 114–116.

VERLINDE, C. L. M. J. & HOL, W. G. J. (1994). *Structure*, **2**, 577–587.

WESSON, L. & EISENBERG, D. (1992). *Protein Sci.* **1**, 227–235.

WILLIAMS, D. H., SEARLE, M. S., MACKAY, J. P., GERHARD, U. & MAPLESTONE, R. A. (1993). *Proc. Natl Acad. Sci.* **90**, 1172–1178.