Journal of Computer-Aided Molecular Design, 11 (1997) 425–445.
KLUWER/ESCOM

J-CAMD 406

Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes

Matthew D. Eldridge*, Christopher W. Murray**, Timothy R. Auton, Gaia V. Paolini*** and Roger P. Mee

Proteus Molecular Design Ltd., Proteus House, Lyme Green Business Park, Macclesfield, Cheshire SK11 0JL, U.K.

Received 31 January 1997
Accepted 20 April 1997

Keywords: Binding affinity prediction; Empirical scoring function; De novo molecular design; Protein–ligand complexes

Summary

This paper describes the development of a simple empirical scoring function designed to estimate the free energy of binding for a protein–ligand complex when the 3D structure of the complex is known or can be approximated. The function uses simple contact terms to estimate lipophilic and metal–ligand binding contributions, a simple explicit form for hydrogen bonds and a term which penalises flexibility. The coefficients of each term are obtained using a regression based on 82 ligand–receptor complexes for which the binding affinity is known. The function reproduces the binding affinity of the complexes with a cross-validated error of 8.68 kJ/mol. Tests on internal consistency indicate that the coefficients obtained are stable to changes in the composition of the training set. The function is also tested on two test sets containing a further 20 and 10 complexes, respectively. The deficiencies of this type of function are discussed and it is compared to approaches by other workers.

Introduction

Structure-based drug design attempts to use the structure of proteins as a basis for designing new ligands by applying accepted principles of molecular recognition. The assumption is that by designing ligands which form good contacts with the receptor, one will obtain molecules with a high binding affinity. Many methods exist for designing or obtaining potential new ligands and most of these methods propose a binding mode for the ligand in the receptor. It would therefore be extremely useful if one could use these proposed binding modes to predict approximate binding affinities for a diverse set of ligands. This estimate could then be used in screening database hits, for scoring molecules produced by de novo design packages, or, after appropriate modification, as an objective function in molecular docking.

Recently, Böhm implemented a new approach which is suitable for application to a large number of ligands produced by a de novo design program [1]. It built on the work of others [2–6] in trying to estimate contributions to binding affinity, but was the first attempt at producing an extremely fast, all-purpose estimate of ligand–receptor binding for drug-like molecules. In his method, four terms based on properties known to be important in drug binding were defined, and an equation for binding affinity was constructed from them. Multiple linear regression was then used to optimise the coefficients multiplying the terms using a training set of 45 protein–ligand complexes for which the binding affinity was known and the binding mode was experimentally determined (or could be reliably inferred). An equation was obtained which had a cross-validated estimate of error of 9.3 kJ/mol across the training set. (A value of 5.7 kJ/mol is equivalent to about 1 order of magnitude in the affinity.) This work represented a considerable step forward for molecular design applications where previously there had been a reliance on empirical functions with a less firm basis or on force fields

*Present address: EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K.
**To whom correspondence should be addressed.
***Present address: Department of Chemistry, University of Liverpool, Liverpool L69 3BX, U.K.

TABLE 1
PROTEIN–LIGAND COMPLEXES IN CLASS 1 (ASPARTIC PROTEASES) OF THE TRAINING SET

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
HIV-1 protease–hydroxyethylene	1AAQ	–47.91	–49.77	16
Penicillopepsin–pepstatin analogue	1APT	–53.62	–42.33	7
Penicillopepsin–pepstatin analogue	1APU	–43.92	–36.67	7
HIV-1 protease–SB203238	1HBV	–36.32	–33.62	17
HIV-1 protease–VX-478	1HPV	–52.61	–38.34	18
HIV-1 protease–GR126045	1HTF	–46.18	–40.60	19
HIV-1 protease–GR137615	1HTG	–55.21	–52.74	19
HIV-1 protease–A-77003	1HVI	–57.48	–61.64	20
HIV-1 protease–A-78791	1HVJ	–59.65	–58.80	20
HIV-1 protease–A-76928	1HVK	–57.70	–59.11	20
HIV-1 Protease–A-76889	1HVL	–51.35	–60.72	20
HIV-1 Protease–XK263	1HVR	–54.25	–59.15	21
Cathepsin D–pepstatin	1LYB	–65.15	–48.40	22
Penicillopepsin–phospho analogue	1PPK	–43.69	–39.09	23
HIV-1 protease–MVT-101	4HVP	–34.85	–53.93	24
HIV-1 protease–acetyl-pepstatin	5HVP	–43.92	–49.79	25
HIV-1 protease–JG-365	7HVP	–54.88	–53.61	26

which were slow and often unreliable when comparing ligands with different chemistries.

Since then, Head et al. [7] have proposed another binding affinity equation (termed VALIDATE) involving 12 terms in which the coefficients were obtained using partial least squares (PLS). Their training set consisted of 51 complexes and they obtained an excellent cross-validated estimate of error of 6.5 kJ/mol and also demonstrated good predictions for test sets. The VALIDATE equation involves several terms that are reasonably difficult to calculate and is over 1000 times slower to evaluate than the Böhm function, making it unsuitable for some situations where speed is important.

The focus of this paper is to develop a scoring function suitable for several applications in molecular design. Primarily, we would like to use the scoring function for application in our synthetically constrained de novo design program, PRO_SELECT, which constructs molecular designs consistent with both synthetic constraints and constraints imposed by a protein structure [8,9]. The program typically screens thousands of diverse structures overnight and scores many different conformations for each structure. Speed is therefore of great importance in our application. In the future, we would also like to apply the scoring function during conformational analysis of protein–ligand complexes, which again places massive speed constraints on the function. Initial experience of applying an implementation of Böhm's scoring function

TABLE 2
PROTEIN–LIGAND COMPLEXES IN CLASS 2 (SERINE PROTEASES) OF THE TRAINING SET

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
Trypsin mutant–benzamidine	1BRA	–10.41	–23.17	27
Thrombin–MQPA	1ETR	–42.21	–38.00	28
Thrombin–NAPAP	1ETS	–48.62	–46.82	28
Thrombin–4-TAPAP	1ETT	–35.3	–35.12	28
Trypsin–NAPAP	1PPC	–36.83	–36.93	29
Trypsin–3-TAPAP	1PPH	–35.5	–37.81	30
Thrombin–D-Phe-Pro-Arg	1TMT	–35.58	–46.75	31
Trypsin–aminomethylcyclohexane	1TNG	–16.73	–21.51	32
Trypsin–4-fluorobenzylamine	1TNH	–19.21	–20.48	32
Trypsin–4-phenylbutylamine	1TNI	–9.69	–18.95	32
Trypsin–2-phenylethylamine	1TNJ	–11.17	–18.83	32
Trypsin–3-phenylpropylamine	1TNK	–8.49	–18.88	32
Trypsin–*t*-2-phenylcyclopropylamine	1TNL	–10.7	–20.91	32
Trypsin–benzamidine	3PTB	–27.04	–23.77	1
Thrombin–D-Phe-Pro-agmatine	TMT1	–39.7	–40.87	9

to drug design [9] indicated that it had merit as a coarse screening function, but predicted differences of an order of magnitude between one design and another should not be relied upon as a basis for preferring one design over another. Further, classically based QSAR must therefore be anticipated during the course of a drug design project [10] so as to improve the estimates of binding affinity provided by empirical scoring functions. We therefore place great emphasis on physical interpretation of the terms and coefficients in the scoring function and also on graphical visualisation of those terms [8]. These provide useful information in choosing synthesis candidates and also form a basis for subsequent QSAR analysis.

We identify the following facets of a good scoring function which this work addresses. We cannot stress strongly enough that many of these points may not be important to other workers and are certainly not important in all situations. For instance, there is still a need for a simple empirical function which yields as accurate an estimate of the binding affinity as possible and this would currently be provided by VALIDATE [7].

Accuracy A scoring function should give acceptable and useful accuracy especially in the context of its proposed usage. In this case, the function should be accurate when applied to a selection of promising *designs* with reasonable binding modes. The first application of an empirical scoring function to this type of training set is given in the next paper of this series [10]. However, in practice, accuracy is most easily tested on complexes where the structure and binding affinity are known.

Range of applicability It is important that the scoring function be applicable to a wide diversity of chemistries without interference or interaction from the user. It should not be so tightly parameterised that it fails to give meaningful results because new functional groups or chemistries are encountered.

Robustness of regression In QSAR applications, a major problem exists in applying QSAR equations to molecules which are not sufficiently like molecules in the training set. This problem is acute here because 3D structures are only available for good binders, but we would like to apply the equation to a large number of different designs, the majority of which will not bind at all. Additionally, the scoring function will be applied to many geometries and chemistries for the ligand–receptor complexes that are not well represented in the database. In our view, this problem can be ameliorated by having as large a database of structures as possible and by performing cross-validation studies to ensure that the regression coefficients are stable with respect to changes in the form of the training set.

Speed issues and appropriateness for docking It is desirable that the scoring function be sufficiently quick to allow application to high-throughput *de novo* design and to use in docking with minor modifications.

Physical interpretability We believe it is preferable to use a small number of terms in the regression and have been strict in the requirements for adding new terms. It is also important to ensure that terms which are significantly correlated (with each other) are not included in the same regression and that all terms and *the coefficients which multiply them* have a physical and interpretable basis. Obviously, these views reflect our desire for a simple equation with a physical interpretation for the terms which can be easily amended in follow-up QSAR. For this reason, we have eschewed PLS with a large number of terms, but accept that such approaches have merit.

This paper describes an approach to the construction of a scoring function which goes some way to meeting the above aims. It extends the approach of Böhm to a much larger training set of 82 compounds and also includes a further 30 complexes as test sets. The compounds in the training set are sorted into classes and the robustness of the coefficients is demonstrated using cross-validation with these classes. This is the first study to go beyond leave-one-out cross-validation which can give misleading results when applied to training sets containing classes of similar compounds occupying similar regions of property space [11].

Only four separate terms (excluding the intercept) are included in the equation and the physical motivation for the terms is similar to Böhm's. However, the hydrogen bond term includes water-mediated contacts, and it draws no distinction between charged and uncharged contacts. The lipophilic term is contact based, rather than being an area term, and does not consider hydrogen atoms explicitly. There is a term for contact of the ligand with metals. The final term considers contributions from rotatable bonds which are frozen because of contact with the receptor. Each rotatable bond makes a different contribution depending on the percentage of non-lipophilic atoms on either side of the rotatable bond. The result is a function that is robust and fairly accurate and which differs considerably from previous work. The avoidance of area terms makes the function suitable for adaption to application of molecular docking.

The second paper [10] in this series will look at the application of the scoring function to a set of designs produced by PRO_SELECT which were later synthesised and tested. The designs were chosen using considerations of diversity, graphical visualisation, empirical score and a simplistic estimate of strain energy [9]. These designs represent a very appropriate test set for the function and allow an assessment of the problems encountered in applications of empirical functions to real molecular design situations. The results indicate the need for follow-up QSAR during the progression of a drug design project employing simple empirical scoring functions.

The next section describes the binding energy terms used in this paper. It also details the test sets and training

TABLE 3
PROTEIN–LIGAND COMPLEXES IN CLASS 3 (METALLOPROTEASES) OF THE TRAINING SET

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
CPA–L-benzylsuccinate	1CBX	–36.21	–46.56	33
Neutrophil collagenase–hydroxamate	1MNC	–51.35	–41.57	34
Thermolysin–phosphoramidon	1TLP	–43.09	–36.72	35
Thermolysin–peptidomimetic	1TMN	–41.65	–47.89	36
Thermolysin–P-Leu-NH_2	2TMN	–33.58	–33.98	35
CPA–GY	3CPA	–22.17	–33.25	37
Thermolysin–VW	3TMN	–33.66	–31.88	7
Thermolysin–Leu-NHOH	4TLN	–21.23	–20.35	1
Thermolysin–ZFP(O)LA	4TMN	–58.13	–58.23	1
Thermolysin–nitroanilide	5TLN	–36.32	–34.04	38
Thermolysin–thorphan	5TMN	–45.87	–43.79	1
CPA–ZAAP(O)F	6CPA	–65.74	–62.15	39
Thermolysin–ZGP(O)LL	6TMN	–28.79	–40.78	40
CPA–BZ-FVP(O)F	7CPA	–79.64	–67.17	41
CPA–BZ-AGP(O)F	8CPA	–52.19	–65.29	41

sets employed, which are all available in the Brookhaven Protein Databank (PDB) [12]. The training sets and associated binding affinity data represent a useful contribution to the scientific community. The section also describes the strategy adopted in cleaning and relaxing (by energy minimisation) the PDB files. The Results section looks at the fitted values for the scoring function, and goes on to describe the cross-validation results for the five different classes defined from the training set. Results for the two test sets of PDB complexes are also given. A Discussion section follows, which interprets and discusses the method and results, and compares them with those of other workers. There is a separate discussion of complexes omitted from the analysis and also of difficulties with the overall approach. Finally, there is a section outlining the conclusions.

Methods

Construction of training set and test set

The regression approach to the construction of an empirical scoring function requires a database of protein–ligand complexes for which the binding geometry is reliable and the binding affinity is known. The database that has been constructed contains only small non-covalently bound ligands of the type which de novo design programs would be expected to produce. For this reason, we have excluded ligands with large cycles, large peptide ligands and ligands containing several sugar monomers. It was also decided to use only PDB structures so the database contains no docked structures or structures obtained from proprietary or privileged sources. Three additional structures (TMT1, TSC2 and DFR4) were obtained by remov-

TABLE 4
PROTEIN–LIGAND COMPLEXES IN CLASS 4 (SUGAR-BINDING PROTEINS) OF THE TRAINING SET

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
ABP–L-arabinose	1ABE	–40.04	–34.79	42
ABP–D-fucose	1ABF	–30.92	–36.68	42
ABP(P254G)–D-fucose	1APB	–33.21	–35.75	42
ABP(P254G)–L-arabinose	1BAP	–39.10	–34.78	42
Glucoamylase–1-deoxynojirimycin	1DOG	–22.92	–21.57	43
Immunoglobulin–D-gal-D-abe-D-man	1MFE	–30.30	–16.23	44
Neuraminidase–sialic acid	1NSC	–17.12	–31.06	45
Neuraminidase–DANA	1NSD	–30.24	–32.06	45
GBP–D-glucose	2GBP	–43.36	–36.55	1
Xylose isomerase–xylitol	2XIS	–33.23	–29.70	1
ABP–D-galactose	5ABP	–37.87	–39.55	42
Concanavilin A–O1-methyl-mannose	5CNA	–11.41	–23.59	1
ABP(M108L)–L-arabinose	6ABP	–36.26	–30.56	46
ABP(M108L)–D-fucose	7ABP	–36.83	–36.30	46
ABP(M108L)–D-galactose	8ABP	–45.64	–36.14	46
ABP(P254G)–D-galactose	9ABP	–45.64	–39.55	46

TABLE 5
PROTEIN–LIGAND COMPLEXES IN CLASS 5 OF THE TRAINING SET

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
Alcohol dehydrogenase–CNAD	1ADB	–47.91	–51.65	47
Dihydrodipicolinate R–NADPH	1DIH	–32.77	–40.25	48
Enolase–phosphonoacetohydroxamate	1EBG	–61.75	–51.96	49
Histidine-binding protein–H	1HSL	–41.65	–30.44	50
Myoglobin–imidazole	1MBI	–10.76	–21.66	1
6-PGDH–6-phosphogluconic acid	1PGP	–32.51	–23.12	51
Cytochrome P450cam–4-phenylimidazole	1PHF	–25.09	–29.12	52
Cytochrome P450cam–metyrapone	1PHG	–49.39	–43.17	52
Retinol-binding protein–retinol	1RBP	–38.35	–39.27	53
Immunoglobulin–GAS	2CGR	–41.51	–42.62	54
Cytochrome P450cam–camphor	2CPP	–34.63	–30.02	55
FABP–palmitic acid	2IFB	–30.99	–39.26	1
Thymidylate synthase–CB3717	2TSC	–48.62	–46.17	56
TP isomerase–2-phosphoglycolate	2YPI	–27.52	–30.18	1
DHFR–methotrexate	4DFR	–55.33	–36.77	1
Cytochrome P450cam–adamantanone	5CPP	–33.58	–30.01	55
DHFR–folate	7DFR	–42.21	–37.15	7
DHFR–2,4-diaminopteridine	DFR4	–34.23	–23.64	1
Thymidylate synthase–compound 3	TSC2	–30.80	–37.05	56

ing part of the ligand in a PDB structure (1TMT, 2TSC and 4DFR, respectively) in the same manner as in Böhm's work.

Crystal structures are of varying reliability according to the resolution obtained and, on consultation, it was decided to use a cutoff on resolution of 2.5 Å (L. Brady, personal communication). Applying this restriction leaves 82 structures for the training set. However, so as not to lose the valuable information contained in structures with a resolution greater than 2.5 Å, 20 structures with resolutions between 2.5 and 3.2 Å were used as test set 1. Another test set, test set 2, was constructed which contains 10 endothiapepsin complexes; the reasons for separating out the endothiapepsins will be addressed in detail later. It should be noted that three outliers were studied, but have been left out of the analysis. One of these (2ADA) would fall into the training set; the other two (1STP and 1ADC) are outside the resolution cutoff and would fall into test

TABLE 6
PROTEIN–LIGAND COMPLEXES IN TEST SET 1 (COMPLEXES WITH >2.5 Å RESOLUTION)

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
Alcohol dehydrogenase β–TAD	1ADF	–26.11	–25.07	57
DB3–progesterone	1DBB	–51.35	–36.80	58
DB3–aetiocholanolone	1DBJ	–43.8	–40.23	58
DB3–5-β-androstanedione	1DBK	–46.19	–37.96	58
DB3–progesterone analogue	1DBM	–53.88	–43.75	58
Thrombin–benzamidine	1DWB	–16.66	–22.61	1
Thrombin–NAPAP	1DWD	–48.39	–45.61	28
Glutamine synthetase–AMP	1LGR	–17.52	–6.52	59
Immunoglobulin–peptide	1MCB	–27.6	–22.76	60
Immunoglobulin–peptide	1MCF	–29.36	–25.35	60
Immunoglobulin–peptide	1MCH	–29.36	–21.54	60
Immunoglobulin–peptide	1MCJ	–21.59	–27.68	60
Immunoglobulin–peptide	1MCS	–27.6	–34.51	60
Neuraminidase–DANA	1NNB	–30.24	–25.70	45
PNP–guanine	1ULB	–30.24	–22.04	1
DB3–pregnane analogue	2DBL	–49.63	–38.58	58
PHBH–*p*-hydroxybenzoate	2PHH	–26.69	–31.16	1
Virus coat protein–compound IV	2R04	–35.49	–44.39	1
Hemagglutinin–sialic acid	4HMG	–14.56	–15.43	1
HIV-1 protease–A-74704	9HVP	–47.62	–61.84	61

TABLE 7
PROTEIN–LIGAND COMPLEXES IN TEST SET 2 (ENDOTHIAPEPSINS)

Protein–ligand complex	PDB code	Experimental affinity (kJ/mol)	Predicted affinity (kJ/mol)	Reference
Endothiapepsin–PD125754	1EED	–27.33	–38.13	62
Endothiapepsin–CP-81,282	1EPO	–45.40	–43.60	63
Endothiapepsin–PD-130,693	1EPP	–40.86	–45.67	64
Endothiapepsin–H-256	2ER6	–41.20	–45.75	65
Endothiapepsin–H-261	2ER7	–51.35	–56.20	66
Endothiapepsin–L363,564	2ER9	–44.48	–57.94	65
Endothiapepsin–CP-71,362	3ER3	–40.52	–52.75	7
Endothiapepsin–PD125967	4ER1	–37.75	–58.52	62
Endothiapepsin–H-142	4ER4	–38.77	–57.27	65
Endothiapepsin–CP-69,799	5ER2	–37.47	–54.55	67

set 1. The outliers will be discussed later. The complexes in the training set (number = 82) have been split into five classes:

(1) aspartic proteases (17 complexes)
(2) serine proteases (15 complexes)
(3) metalloproteases (15 complexes)
(4) sugar-binding proteins (16 complexes)
(5) others (19 complexes).

In the case of distinguishing between sugar-binding proteins and 'others', some subjective judgement was required – where a sugar-like moiety was a fragment of a much larger molecule (such as a nucleoside), then the complex was assigned to 'others'. Tables 1–5 give important information on the classes of the training set, and Tables 6 and 7 give information on test set 1 and test set 2. The information given is the protein–ligand complex, the PDB entry, the experimental K_i (in kJ/mol), the predicted K_i and a reference to the paper where the K_i, that we used, was taken from. We hope these tables will be useful to other workers as a database for comparison of future scoring functions. At the outset of this work, it was expected that inclusion of well-positioned hydrogen atoms would be required to obtain good results. It was therefore necessary to establish a protocol to position the hydrogens. It was also felt that since the scoring function was to be applied to modelled structures of one sort or another, the PDB structures should be locally relaxed in the region of the ligand–receptor interface. This might produce a scoring function more appropriate for modelled complexes and could perhaps reduce the subjectivity associated with hydrogen positioning. Subjectivity still remains because, for some complexes, the results are influenced by the initial positioning of the hydrogens. In such cases, the initial positioning must be done carefully if full recovery of hydrogen bonds is to be obtained during relaxation. The basic protocol used is as follows:

(1) Read PDB structures into INSIGHT [13].
(2) Prepare for DISCOVER [14] job.
    Assign hydrogen positions automatically – amending by hand if necessary.
    Assign ionisation states at the interface between ligand and receptor.
    Assign and check ligand atom types [15] and charges.
(3) Set up and run DISCOVER job. This uses a two-step minimisation procedure. The first minimises the hydrogen positions only and stops after a hundred steps. It is designed to remove any particularly bad clashes induced by the hydrogen atom positioning. The second minimisation continues to completion and uses three zones:
    The inner zone includes all residues 10 Å or closer to the ligand. In this region, heavy atoms are tethered and hydrogens move freely.
    The mid-zone includes all residues 15 Å or closer, but which are not in the inner zone. Here all atoms are tethered.
    The outer zone contains all other residues. Here the atoms are fixed and so they are not used in the calculation of the non-bonded energies.
(4) A separate routine compares the distances moved during the minimisation procedure and the scale of these distances can be displayed graphically in INSIGHT.
(5) Another routine identifies hydrogen bonds and potential hydrogen bonds between the ligand and the receptor. It also provides a macro which displays, in INSIGHT, the changes in hydrogen bond patterns before and after minimisation. The crystallographic paper is used as a guide in interpreting whether changes are acceptable or not. Generally, the user is hoping to see a regularisation in the hydrogen bond pattern without *any* significant changes in conformation.
(6) Modify and rerun the process (or reject structure) if analysis identifies problems with the minimised complex.

Many of the steps in this procedure are automated. However, so far it has not proved possible to automate the whole approach. The preparation of the complexes, and

the control on their quality, is fairly time consuming – on average, each complex takes over half a day to model, although sets of related complexes take less time.

The regression terms

The scoring function assigns general atom types to all ligand atoms and receptor atoms in contact with the ligand. The atom types are:

Lipophilic Chlorine, bromine and iodine atoms which are not ions; sulphurs which are not acceptor or polar types; carbons which are not polar type;

H-bond donor Nitrogens with hydrogen attached; hydrogens attached to oxygen or nitrogen;

H-bond donor/acceptor Oxygens attached to hydrogen atoms; special case of imine nitrogen (i.e. C=NH nitrogen);

H-bond acceptor Oxygens not attached to hydrogen; nitrogens with no hydrogens attached and one or two connections; halogens which are ions; sulphurs with only one connection (e.g. thioureas);

Polar (non H-bonding) Nitrogens with no hydrogens attached and more than two connections; phosphorus; sulphurs attached to one or more polar atoms (including H-bonding atoms and not including polar carbon atoms or fluorine atoms); carbons attached to two or more polar atoms (including H-bonding atoms and not including polar carbon atoms or fluorine atoms); carbons in nitriles or carbonyls; N atoms with no hydrogens and four connections; fluorine atoms; and

Metal Metal atoms.

It is hoped that these atom types are reasonable in most applications of the function, and they are thought to be reasonable compromises across the molecules in the ligand–receptor database. The polar atom type is not used in any of the terms described here but was useful for prototyping alternative terms, and is useful in the atom assignment routines in establishing the lipophilic atom type. The assignment of heteroatoms as H-bond donors and H-bond donor/acceptors reflects our desire to avoid the explicit use of hydrogens in future docking applications. However, in this work hydrogen bond energy expressions are only evaluated when the hydrogen bond donor atom involved in the H-bond is H.

The empirical scoring function can be written in the form

$$\begin{aligned}\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} \sum_{iI} g_1(\Delta r) g_2(\Delta\alpha) \\ + \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{lL} f(r_{lL}) \\ + \Delta G_{rot} H_{rot}\end{aligned} \quad (1)$$

The ΔG coefficients are unknown and will be obtained by multiple linear regression. The hydrogen bond term, $\sum_{iI} g_1 g_2$, is calculated for all complementary possibilities of hydrogen bonds between ligand atoms, i, and receptor atoms, I. The functions g_1 and g_2 are of the same form as used by Böhm:

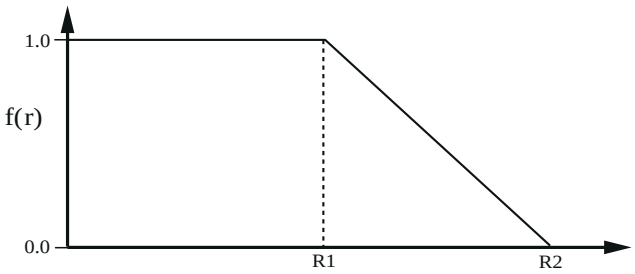


Fig. 1. Form of the function f(r) given in Eq. 1. The parameters R1 and R2 have different values for the different terms in Eq. 1.

$$g_1(\Delta r) = \begin{cases} 1 & \text{if } \Delta r \le 0.25\,\text{Å} \\ 1 - (\Delta r - 0.25)/0.4 & \text{if } 0.25\,\text{Å} < \Delta r \le 0.65\,\text{Å} \\ 0 & \text{if } \Delta r > 0.65\,\text{Å} \end{cases}$$

and

$$g_2(\Delta\alpha) = \begin{cases} 1 & \text{if } \Delta\alpha \le 30^\circ \\ 1 - (\Delta\alpha - 30)/50 & \text{if } 30^\circ < \Delta\alpha \le 80^\circ \\ 0 & \text{if } \Delta\alpha > 80^\circ \end{cases}$$

Δr is the deviation of the H···O/N hydrogen bond length from 1.85 Å and $\Delta\alpha$ is the deviation of the hydrogen bond angle N/O-H···O/N from its ideal value of 180°. Water molecules are scored for contact with the receptor and those possessing more than one hydrogen bond to the receptor are treated as if they are part of the receptor when contacts between the ligand and receptor are evaluated. It should also be noted that there is no distinction between ionic and non-ionic hydrogen bonds; hydrogen bonds between atoms with formal charges are given the same weight as those without charges.

The metal term, $\sum_{aM} f(r_{aM})$, is calculated for all acceptor and acceptor/donor atoms, a, in the ligand and any metal atoms, M, in the receptor. The function f(r) is a simple contact term and its functional form is illustrated in Fig. 1. The parameters R1 and R2 defined by Fig. 1 are 2.2 and 2.6 Å, respectively. r_{aM} is the distance between ligand and receptor atoms of the appropriate type.

The lipophilic term, $\sum_{lL} f(r_{lL})$, is calculated for all lipophilic ligand atoms, l, and all lipophilic receptor atoms, L. The parameters R1 and R2 defined by Fig. 1 are $r_l^{vw} + r_L^{vw} + 0.5$ and R1 + 3.0, respectively, where r_l^{vw} is the van der Waals radius of atom l and distances are measured in Å. It should be noted that this term is fairly long range. This appears to be important for obtaining good results and will be discussed later.

The final term first identifies frozen rotatable bonds. The definition of rotatable bonds is any sp^3-sp^3 bond and any sp^2-sp^3 bond, excluding bonds to terminal CH_3, CF_3, NH_2 or NH_3 groups. Bonds are considered frozen if atoms on both sides of the rotatable bond are in contact with the receptor (i.e. the distance between any two heavy

TABLE 8
STATISTICAL QUANTITIES FOR REGRESSIONS USING DIFFERENT TRAINING SETS

Description of training set	Number in training set	r^2	s	q^2	s_{PRESS}
Full training set	82	0.710	7.99	0.658	8.68
Class 1	65	0.719	7.72	0.653	8.58
Class 2	67	0.645	8.04	0.570	8.85
Class 3	67	0.691	8.02	0.615	8.94
Class 4	66	0.733	8.12	0.677	8.93
Class 5	63	0.750	7.90	0.675	9.01
Classes 1, 4 and 5	30	0.858	7.01	0.785	8.64
Classes 1 and 5	46	0.789	7.20	0.716	8.36
Classes 1 and 4	49	0.765	7.75	0.700	8.77

Note that when the training set descriptions give a specified class, this means that the class is *absent* from the training set during the regression. The coefficients for the terms in the corresponding regressions are given in Table 9.

atoms is less than the sum of the relevant van der Waals radii plus 0.5 Å). The rationale is that if one side of the rotatable bond is not in contact, then rotation about that bond is only partially impaired and therefore the entropic penalty is not as large. Initial experience with this term indicated that the rotatable bond penalty for large lipophilic ligands could often be unrealistically large. For instance, it is not uncommon for peptide ligands of aspartic proteases to have more than 20 frozen rotatable bonds. The main reason why this is unrealistic is that lipophilic peptides are likely to fold into a low-energy conformation in solution and, therefore, the effective number of freely rotatable bonds is greatly reduced. Additionally, small lipophilic side chains in contact with the receptor are likely to retain some residual entropy associated with the non-local nature of lipophilic interactions. Both these effects will tend to reduce the contribution from rotatable bonds which have mainly lipophilic fragments on either side of the rotatable bond. It is to be expected that the effect will be less important for small numbers of rotatable bonds. The following function is used to estimate the flexibility penalty for molecules possessing frozen rotatable bonds:

$$H_{rot} = 1 + (1 - 1/N_{rot}) \sum_r (P_{nl}(r) + P'_{nl}(r))/2 \qquad (2)$$

where N_{rot} is the number of frozen rotatable bonds, the summation is over frozen rotatable bonds and $P_{nl}(r)$ and $P'_{nl}(r)$ are the percentages of non-lipophilic heavy atoms on either side of the rotatable bond. This term is considered in more detail in the Discussion section.

Regression

The multiple linear regression method used in this work is standard. The methods used for assessing the various complexes deserve some explanation though. In the assessment of the quality of fitting to the training set, we shall use s (the standard deviation or standard error) and r^2 (the amount of variance in the training set explained by the scoring function). The leave-one-out cross-validation quantities q^2 and s_{PRESS} are also used, where

TABLE 9
COEFFICIENTS OBTAINED FOR EACH TERM IN A VARIETY OF DIFFERENT REGRESSIONS USING DIFFERENT TRAINING SETS

Description of training set	Intercept[a]	H-bond coefficient (kJ/mol)	Metal coefficient (kJ/mol)	Lipophilic coefficient (kJ/mol)	Rotatable bond coefficient (kJ/mol)
Full training set	−5.48	−3.34	−6.03	−0.117	2.56
Class 1	−2.92	−3.63	−6.08	−0.119	2.51
Class 2	−10.8	−3.05	−5.04	−0.107	2.65
Class 3	−5.40	−3.39	−7.41	−0.117	2.61
Class 4	−4.47	−3.67	−6.01	−0.122	3.16
Class 5	−4.28	−2.96	−6.12	−0.122	1.96
Classes 1, 4 and 5	+2.32	−2.82	−8.42	−0.168	2.76[b]
Classes 1 and 5	+1.54	−3.32	−5.61	−0.128	1.08[b]
Classes 1 and 4	−1.96	−4.17	−6.02	−0.123	3.49

Note that when the training set descriptions give a specified class, this means that the class is *absent* from the training set during the regression. The statistical quantities showing the performance of the corresponding regressions are given in Table 8.

[a] The value assigned to the intercept has very little statistical significance in the regressions.

[b] The coefficient assigned to the rotatable bond term has little statistical significance in this regression.

$$q^2 = 1 - \sum(y_{pred} - y_{obs})^2 \Big/ \sum(y_{obs} - y_{mean})^2 \qquad (3)$$

and

$$s_{PRESS} = \sqrt{\sum\left(y_{pred} - y_{obs}\right)^2 \Big/ (N - k - 1)} \qquad (4)$$

with summations occurring over the number of complexes in the training set, N, and k being the number of terms in the regression. In judging the quality of the regression, especially in comparison with other workers, reliance should be placed mainly on s_{PRESS} or s, since the r^2 values will depend heavily on the composition of the training set. Cross-validation studies have also been applied using the different classes defined in Tables 1–5. Regressions are repeated with a class (or classes) absent from the training set. The change in the coefficients of the regression and the performance of the equation is analysed with respect to the equation obtained from the full training set. In particular, the predictive r^2 and the rms residual for predictions, s_{pred}, are calculated for the missing class (or classes) using the equation established with the training set minus themselves. In this paper, the predictive r^2 is defined by

$$r^2_{pred} = \left(\sum(y_{obs} - y_{mean})^2 - \sum(y_{pred} - y_{obs})^2\right) \Big/ \sum(y_{obs} - y_{mean})^2 \qquad (5)$$

where summations occur over the set to be predicted but y_{mean} is the mean of the training set used in establishing the relevant regression equation. Pearson's correlation coefficient, r_p, is also used to give an indication of how well-predicted complexes are ranked amongst themselves. Poor values of r_p will be observed when the variance of activities in the predicted class is small; r^2_{pred} will not suffer from this problem, although it is affected by the range of activities in the training set. Similar quantities are used to judge the quality of predictions with the scoring function from the full training set on test set 1 and test set 2.

Results

Information on the statistical performance of the regression on the training set is given as the first line in Table 8. The corresponding coefficients for the training set are given in the first line of Table 9. This is our preferred regression. The values for the fitted binding affinities in the training set are given next to the experimental values in Tables 1–5. A scatter plot showing the performance on the training set is given in Fig. 2. Overall, the performance of the scoring function is quite good; giving cross-validated estimates of error (s_{PRESS}) of about 1.4 orders of magnitude (8.68 kJ/mol).

Each of the classes was removed in turn from the training set and the regression repeated. The results are shown in Table 8, together with the results for three other experiments. The results indicate that the coefficients are, for the most part, fairly stable to changes in the training set. In particular, the changes in predictions will be small

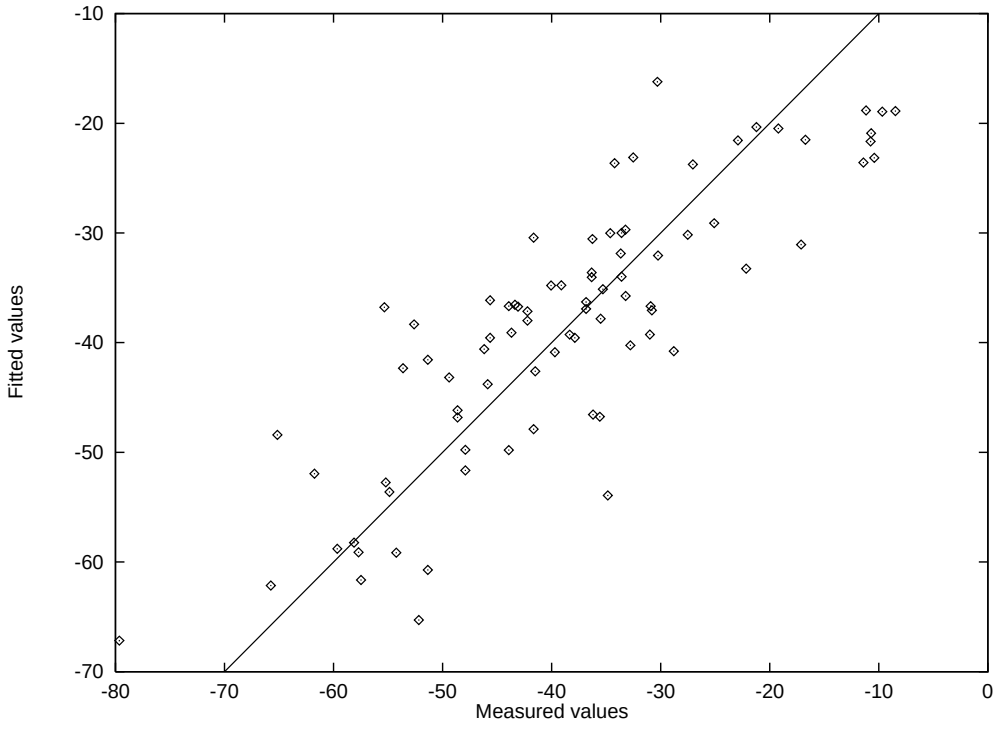


Fig. 2. Scatter plot showing the fitted binding affinities for the training set versus the experimental binding affinities in kJ/mol. The fitted values were obtained using the preferred regression equation.

TABLE 10
QUALITY OF THE PREDICTION FOR DIFFERENT CLASSES IN THE TRAINING SET

Description of predicted set	Number in predicted set	Rms residual	Mean residual	r_p	r^2_{pred}	Range of ΔG in predicted set
Class 1	17	9.22	–0.62	0.462	0.595	–65.2 to –34.9
		(8.63)	(–1.20)	(0.497)	(0.645)	
Class 2	15	9.10	6.82	0.933	0.758	–48.6 to –8.5
		(6.87)	(4.11)	(0.930)	(0.862)	
Class 3	15	8.38	3.19	0.867	0.736	–79.6 to –21.3
		(7.76)	(0.94)	(0.869)	(0.774)	
Class 4	16	7.74	–1.64	0.615	0.469	–45.6 to –11.4
		(7.30)	(–1.20)	(0.640)	(0.527)	
Class 5	19	8.71	–2.52	0.691	0.418	–61.8 to –10.8
		(7.92)	(–1.90)	(0.739)	(0.519)	
Full training set	82	(7.75)	(0.00)	(0.843)	(0.710)	–79.6 to –8.5
Test set 1	20	7.93	–2.22	0.805	0.630	–53.9 to –14.6
Test set 2	10	12.58	10.52	0.420	–3.00	–51.4 to –27.3

The predictions were obtained from regression equations in which the specified class (or classes) was removed from the training set. The fitting for the classes using the preferred regression equation from the full training set is given in parentheses – in this case, r^2_{pred} is defined as in Eq. 5, although strictly speaking this is not a predictive r^2. The results of using the preferred equation (derived from the entire training set) to predict the binding affinities of test set 1 and test set 2 are also given in the table. The results for the full training set are reproduced for convenience.

with each of the different regressions shown in the tables. This means that the scoring function produced is expected to be robust as more complexes are added to the training set. It is interesting to note that some regressions appear to perform better than others. For example, the regression in which classes 1 and 5 are excluded gives an r^2 of 0.789, which is considerably better than the full training set. This shows that some care must be taken when comparing the performance of different scoring functions when they are being tested using different training sets.

Table 10 gives an overview of how each class is predicted using the regression equations given in Table 8 in which the relevant class was excluded from the regression equation. Also shown in parentheses are the corresponding *fitted* values for each of the classes using the preferred regression equation which was derived from the full

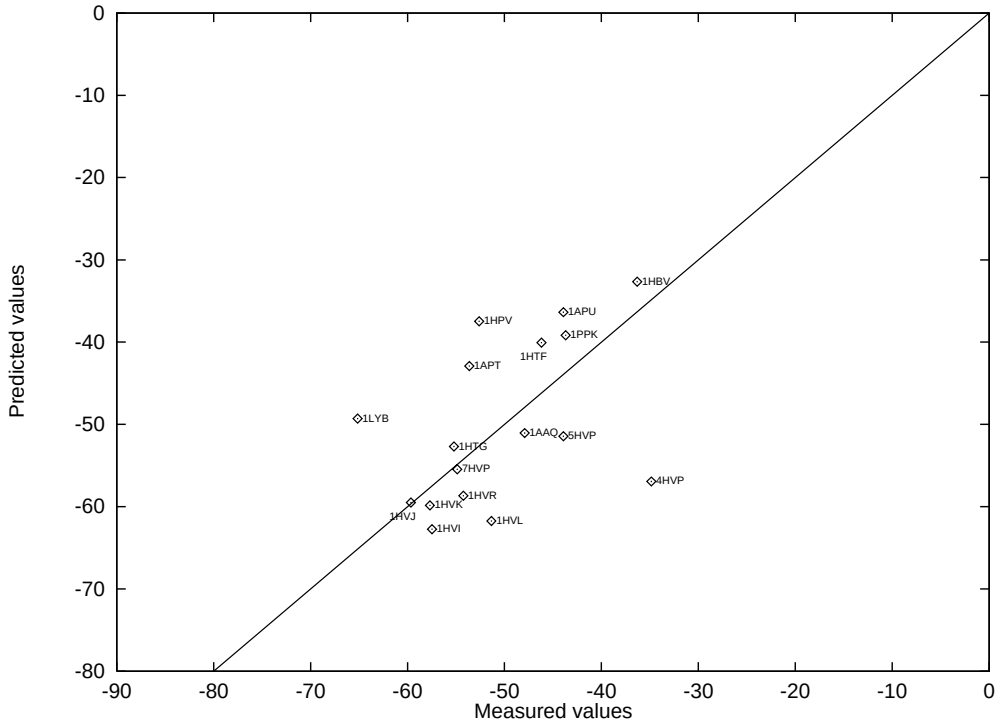


Fig. 3. Scatter plot showing the predicted binding affinities for class 1 (the aspartic proteases) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the regression equation constructed without class 1.

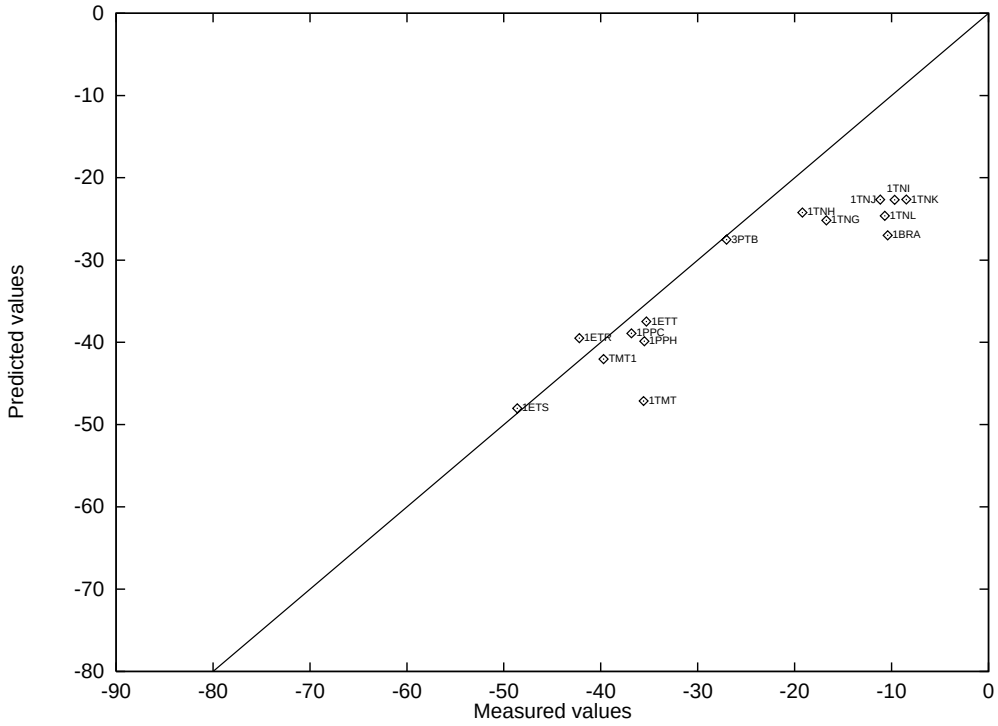


Fig. 4. Scatter plot showing the predicted binding affinities for class 2 (the serine proteases) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the regression equation constructed without class 2.

training set. It can be seen that all the classes are predicted reasonably well, although classes 2 and 3 (the serine proteases and metalloproteases) seem to be slightly better fitted/predicted than the rest. The r^2_{pred} values are less evenly distributed, reflecting that some classes have a narrow range of activities and that this range may or may not be close to the mean activity of the training set. It can be seen that the rms residuals for the predictions

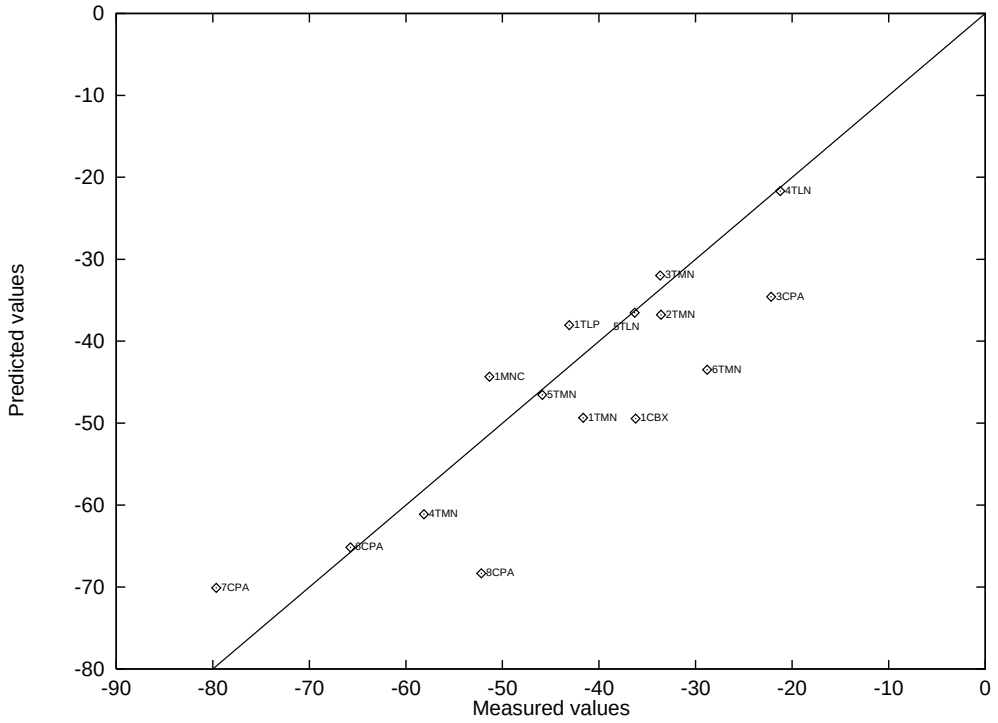


Fig. 5. Scatter plot showing the predicted binding affinities for class 3 (the metalloproteases) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the regression equation constructed without class 3.

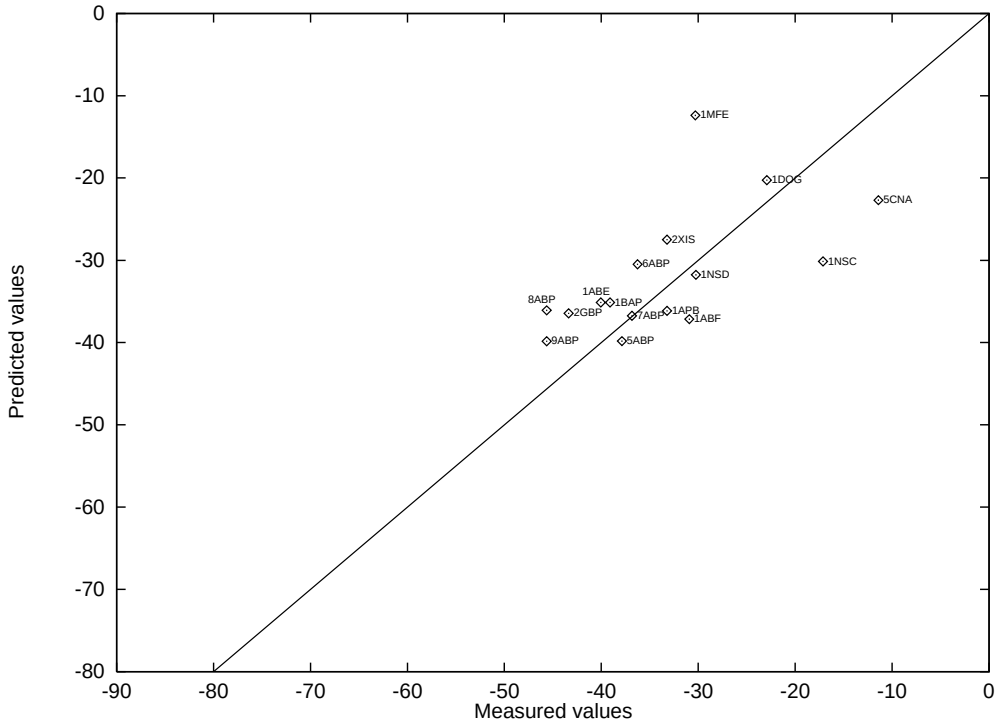


Fig. 6. Scatter plot showing the predicted binding affinities for class 4 (the sugar-binding proteins) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the regression equation constructed without class 4.

are less than 2.5 kJ/mol worse than those for the fitting. This level of degradation is acceptable and shows the robustness of the training set and the terms chosen for the regression. It is interesting to note that some classes have non-zero mean residuals, indicating a systematic tendency to under- or overpredict. This can also be seen in the scatter plots for the predictions on the classes shown in Figs. 3–7. It is noticeable that the serine proteases are sys-

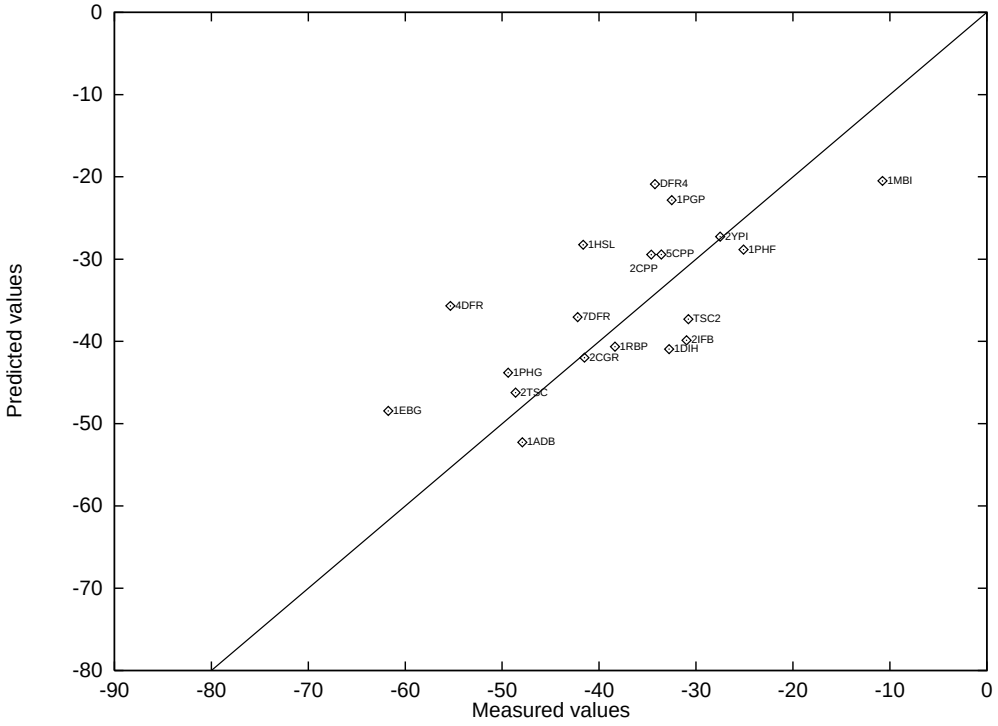


Fig. 7. Scatter plot showing the predicted binding affinities for class 5 ('other' complexes) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the regression equation constructed without class 5.

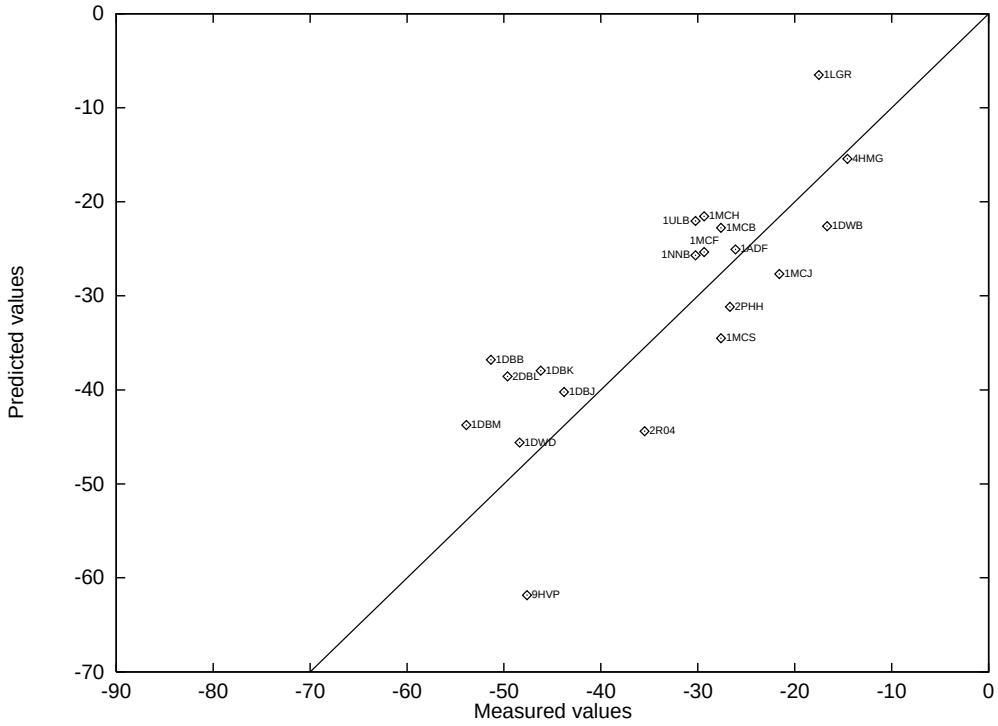


Fig. 8. Scatter plot showing the predicted binding affinities for test set 1 ('other' complexes) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the preferred regression equation.

tematically overpredicted (Fig. 4). This could be due to the presence of a number of small molecules binding to trypsin in the low-activity region of the scatter plot – these molecules have fairly odd geometries, indicating that some account of strain energy might bring them closer to the line.

Table 10 also gives the results of using the preferred regression on the two test sets, and the corresponding scatter plots are shown in Figs. 8 and 9. Test set 1 consists of the 20 complexes of resolution poorer than 2.5 Å. The results are of similar quality to the fitting results for the training set using the same regression. Again, this demonstrates the robustness of the equation when applied to complexes derived in a similar way to the training set. The results also indicate that the resolution cutoff, although useful and physically justified, makes very little difference to the quality of the results. Test set 2 consists of 10 endothiapepsin complexes, all derived from the same research group [62–67]. It is clear from Table 10 and Fig. 9 that they are systematically overpredicted by about 2 orders of magnitude. Although the actual prediction for the endothiapepsin complexes is bad, the ranking of the molecules (as evidenced by the correlation coefficient, r_p) is fairly respectable. It is noted that Head et al. [7] comment on an early form of the VALIDATE scoring function where they observed good correlation among ligands associated with the same receptor, as compared to ligands from different receptor types. They addressed this by scaling computed surface areas by the size of the active site of the receptor compared to HIV-1 protease. We are uncertain of the physical significance of this approach and, given our prejudice against using a PLS-based method with a large number of terms, decided not to correct for the endothiapepsins with a new or an altered term in the regression but to retain them as a second test set. It is not obvious why there should be a difference in the behaviour of the endothiapepsins compared with other aspartic protease complexes, and this could indicate a weakness of our entire approach. Another explanation lies in the experimental binding affinities for the endothiapepsins which are all measured at a pH of 3.1 (or slightly higher). This could lead to a lowering of the measured K_i values relative to other values in the ligand–receptor database, and thus produce a tendency to overpredict. Support for this explanation lies in the behaviour of the K_i's of HIV-1 protease inhibitors on changing the pH. Hosur et al. [20] give the K_i of 1HVI, 1HVJ, 1HVK and 1HVL at pH values of 6.2 and 4.7, where the lower pH leads to a less good binding affinity by just under an order of magnitude. Although this is an attractive explanation, it is to be expected that the pH dependency of ligands and receptors will be highly dependent on the nature of the ligand and the receptor. The important point is that on a drug design project targeted at endothiapepsins, the ranking of ligands to endothiapepsins is expected to be fairly accurate on the basis of these data, regardless of the absolute value of binding affinity. It would be a simple matter to perform a follow-up QSAR after the initial assay results were obtained and correct for the overprediction.

Discussion

A new scoring function has been developed which relates binding affinities to a number of simple terms based on the principles of molecular recognition. The approach and philosophy are most like those adopted by Böhm [1], who was the first to apply this approach to the high-throughput scoring of general small molecules in receptor sites. This work differs primarily in the definition of a much larger database of protein–ligand complexes, all of which are directly derived from public domain PDB structures. The scoring function produced should be more reliable and robust. (It should be noted that near the completion of this work, Böhm presented a new scoring function at the 1996 ACS meeting in Orlando. A database of 82 complexes was used to fit a function with seven terms with an s value of 7 kJ/mol. The main improvement over his earlier work [1] was presented as increased robustness, similar to the conclusions we have reached with a similarly sized database. Further comparison of our work with Böhm's latest unpublished work is difficult and will not be attempted.) The s_{PRESS} values from this work also compare favourably with Böhm's work (8.68 kJ/mol versus 9.3 kJ/mol). Comparing the terms in the two regressions, we have adopted a contact-based lipophilic term rather than an area-based term. During the development of the scoring function, we progressed from an area-based term to a contact term partly because improved area terms were taking longer and longer range interactions into account. We ascribed this to the need to use the general environment of an area in assigning it as lipophilic or not, rather than just the closest atom. The closest atom can be misleading because some very hydrophobic pockets contain, for example, amide groups which sit tangential to the surface that defines the pocket. It is important to consider more distant atoms to properly describe the area in the region of such atoms as predominantly lipophilic. We subsequently discovered that a longer range contact term was just as effective as the more complicated and computationally less attractive area terms.

The hydrogen bond term has a similar explicit form to Böhm's. We have investigated replacing this with a simpler contact-based term, $\sum_{il} f(r_{il})$, where f(r) is defined as in Fig. 1, and R1 and R2 are given by 3.0 and 3.2 Å, respectively. Our implementation of this term has the advantage that hydrogens are not considered explicitly and this may provide a smoother potential energy surface for docking applications; it also means that it is possible to use PDB files directly without significant and time-consuming preprocessing. Across the training set, the correlation between the explicit hydrogen bond term and the simpler contact term is strong ($r^2_{CV} = 0.901$; $s_{PRESS} = 1.1$). Full regression using the simpler contact term generally gives slightly superior results for all the tests given in the results section. (For example, regression across the training set gives $r^2_{CV} = 0.678$ and $s_{PRESS} = 8.43$ which is to be compared with the preferred regression results in Table

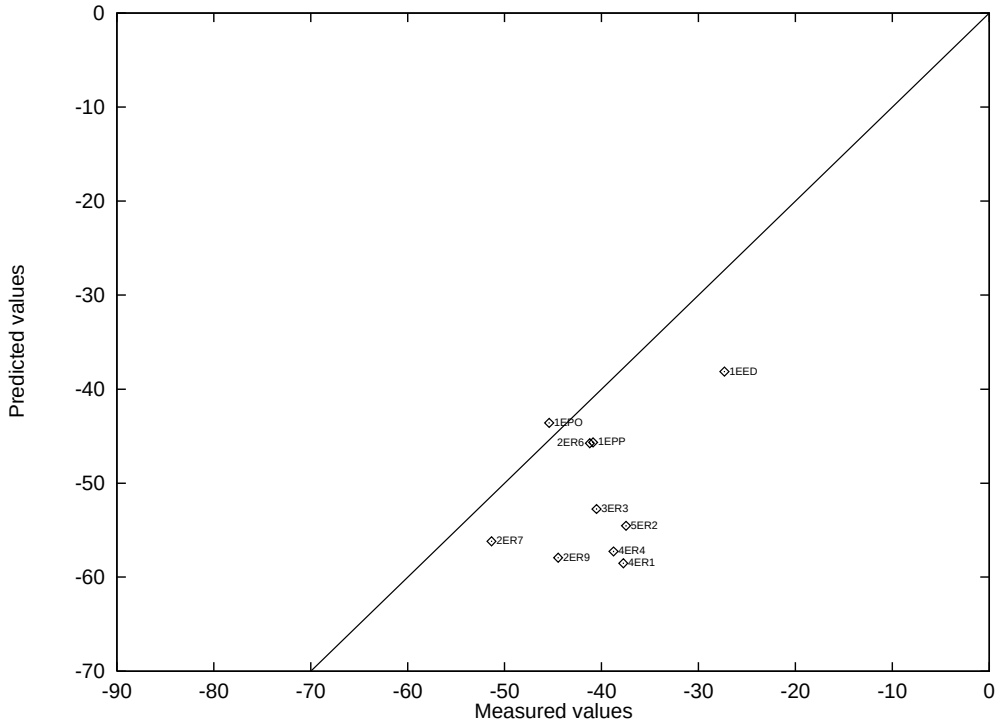


Fig. 9. Scatter plot showing the predicted binding affinities for test set 2 (the endothiapepsins) versus the experimental binding affinities in kJ/mol. The predicted values were obtained using the preferred regression equation.

8 of $r^2_{CV} = 0.658$ and $s_{PRESS} = 8.68$.) The results using the simple contact term on the unminimised PDB files are also respectable ($r^2_{CV} = 0.653$; $s_{PRESS} = 8.74$), indicating that the improvement on minimising the structures in the training set is only modest. The good results obtained using the simpler contact term are curious, since it might be expected that not taking account of hydrogen positions would be a severe approximation. Interpretation of this is difficult; clearly, for some complexes involving many contacting polar atoms (e.g. 2XIS), the contact term gives more contacts/hydrogen bonds than could be satisfied by any specific orientation of the relevant hydrogen atoms. In some cases, this may be bad, but in, say, the aspartic proteases the simpler contact term will score more contacts with the catalytic aspartate groups than the explicit term, and this may reflect a physical effect which yields improved binding affinity. We prefer the explicit term because experience with the simpler form in real design situations has been unsatisfactory. In particular, we have used the function to assess the quality of docked database hits where the putative binding geometry for the ligand–receptor complex is often suspect. The simple contact term often scores geometries with unrealistic polar contacts favourably and accentuates the problem of false positives, as compared with the explicit form. This experience underlines the problem of relying too heavily on the performance of the scoring function on a set of known ligand–receptor complexes, and not paying enough attention to the performance (or likely performance) of the function in real applications. The form of function required to avoid the proliferation of false positives has been discussed extensively and lucidly by Jain [68].

A difference in the hydrogen bond term of this work compared to that used by Böhm is that charged or 'ionic' hydrogen bonding contacts are scored the same as uncharged ones. The basis for this is purely empirical and it is certainly true that such contacts are vitally important in some enzymes [69] (e.g. the S1 pocket of trypsin will not tolerate uncharged contacts with Asp^{189}). However, it is our observation that separating out the ionic and neutral hydrogen bonds does not significantly help the performance of *our* scoring function and certainly does not justify the introduction of a new term in the regression. For example, if hydrogen bonds in which donor and acceptor atoms are charged are treated separately from other hydrogen bonds, no improvement is seen for the training set ($r^2_{CV} = 0.667$; $s_{PRESS} = 8.63$). More worryingly, the coefficient of the ionic term is slightly smaller than that for the normal hydrogen bonds, indicating that the slight improvement in the regression is probably artifactual. Similarly, poor results were observed with other definitions of what constitutes an ionic hydrogen bond. Another difference in the hydrogen bond term is the inclusion of water-mediated contacts in the scoring function. These were introduced when it was apparent that several underpredicted complexes contained trapped water molecules while, in general, overpredicted complexes did not. The regression is improved by the inclusion of water-mediated terms (water-mediated terms absent: $r^2_{CV} = 0.618$; $s_{PRESS} = 9.17$).

The final difference between our terms and those used by Böhm is the different form of the flexibility term. In our hands, a simple count of rotatable bonds proved to give unreliable results. The results were only marginally improved by the definition of frozen rotatable bonds ($r^2_{CV} = 0.623$, $s_{PRESS} = 9.11$) since this only changes the rotatable bond count for a small number of complexes in the training set. The unreliability of the rotatable bond term is manifested by the sensitivity of the results to the composition of the training set. For example, if a simple count of frozen rotatable bonds is used and the training set is changed so as to exclude no complexes (N = 82), exclude complexes with more than 15 frozen rotatable bonds (N = 69), exclude complexes with more than 12 frozen rotatable bonds (N = 58) and finally to exclude complexes with more than 10 frozen rotatable bonds (N = 54), then the coefficient of the rotatable bond term changes as 0.635, 0.924, 1.08 and 1.71, respectively, with fairly large changes in the other terms to compensate. In particular, the intercept becomes progressively more positive (–3.32 goes to +2.86) and the favourable contribution from the other terms increases. The coefficients of the last regression (+2.86, –4.12, –8.16, –0.150, +1.71; $r^2_{CV} = 0.663$, $s_{PRESS} = 7.72$, N = 54) are quite similar to Böhm's original ones [1]. The variation in coefficients as the composition of the training set changes is somewhat alarming and again reveals the dangers of placing too much reliance on empirical scoring functions. One option is to simply exclude the larger complexes from the regression – this makes sense because the intended application is small molecules with modest numbers of rotatable bonds. However, we chose to look for a more stable definition of flexibility which corrects the rotatable bond count for large peptides. The hope was that a regression that is applicable to the more flexible molecules will contain coefficients that are more transferable in general. The rotatable bond count becomes unreliable because ligands with large numbers of rotatable bonds, especially if they are hydrophobic, will tend to fold in solution and thus the loss in entropy on binding will not be as large as anticipated by the bond count. Additionally, due to the non-local nature of hydrophobic contacts, rotatable bonds which are frozen as a result of hydrophobic contacts will often retain residual entropy (e.g. the side chains of valine or leucine might be expected to retain entropy when bound in a lipophilic pocket). The term used in this work (Eq. 3) attempts approximately to correct for these two problems by scaling each rotatable bond with the percentage of non-lipophilic character on each side of the rotatable bond. The result is a more complex term (which is a disadvantage),

but a term that is more transferable across the database. For instance, in the above experiment on the amended training sets (N = 82, N = 69, N = 58 and N = 54), the coefficients of the amended rotatable bond term are 2.56, 2.35, 2.18 and 2.97, respectively. The other terms are also more stable to these experiments and the performance of the regression is superior with the full training set. It is for these reasons that we use the flexibility term given in Eq. 3. However, it has to be admitted that neither the rotatable bond count nor the flexibility term are particularly satisfactory estimates of the entropic penalty. The work of Nicklaus et al. [70] is particularly interesting in this regard. They have performed a study of the conformations of 33 compounds in the Brookhaven database and have estimated the strain energy *without inclusion of solvation effects*. They have found a fair correlation (r^2 = 0.633) between the strain energy and the number of rotatable bonds. This work indicates that some of the strain energy may be implicitly included in the flexibility term, again clouding the physical interpretation of the term.

Comparison of the current work with the VALIDATE scoring function [7] is less relevant than the comparison to Böhm's work since the philosophies used in obtaining the functions are very different. We have focused on a small number of terms with physical interpretations for their functional form and weights, and this focusing will compromise accuracy. For example, it seems likely that a combination of all the different terms we have tried for lipophilic area together with all the other area terms we have investigated would lead to a more accurate scoring function. We have not followed this route for reasons outlined in the Introduction section; primarily the reasons are simplicity, speed, interpretability and appropriateness for application in follow-up QSAR. To illustrate this, an early problem of physical interpretation in the training set was the hydrophobic–hydrophilic contact area. Initial physical ideas would mark this as a term unfavourable, or at best neutral, to binding affinity. However, in prototype regressions, this term was helpful to the QSAR and came in with a negative coefficient (i.e. more mixed contact was favourable to binding). We ascribed this to the correlation of the mixed area with molecular volume, which was also useful in prototype regressions. Generally there is a trend in the training set for larger ligands to have larger binding affinities. Clearly, large ligands chosen at random will not fit this trend, and so this correlation is to some extent artifactual and is related to the type of molecules which are chosen for crystallisation. This indicates the dangers inherent in the training set and underlines the need to test the scoring function on a training set more representative of the required applications. The second paper in this series will illustrate the application of the scoring function to a set of promising designs [10] and hopefully goes some way to addressing the problems inherent in the training set and test sets. The VALIDATE scoring function also includes modelling data in the regression, in particular the van der Waals interaction energy, the electrostatic interaction energy and an estimate of strain energy. We have avoided the use of modelling data for reasons of computational speed, especially since we wish to use an amended form of the scoring function in docking applications. It is to be expected that including modelling data will give better results, and, in particular, a good calculation of the strain energy should be very helpful. VALIDATE uses the best calculation possible, without severely compromising speed, and this could explain its superior performance. Another important drug binding term included in the VALIDATE function is Log P [71]. We have made some attempts to test the effect of including estimates of Log P in our regression, but have not found an improvement. In our case, this can be attributed to a poor estimate of Log P [72] for largish peptides where the calculation of Log P is notoriously difficult. In conclusion, the VALIDATE scoring function is significantly more accurate (s_{PRESS} = 6.50 kJ/mol) than our scoring function due to the larger number of terms and the incorporation of modelling data, however, our function is more suited to the applications it was designed to address, and has been built upon a larger training set.

Whilst this manuscript was in preparation, Jain [68] published an article describing a new empirical scoring function. Thirty-four complexes are included in his training set and he fitted a large number of parameters using a non-linear technique. Due to the nature of his fitting technique, it is difficult to compare his results with this or other work, but his rms errors appear to be more accurate than VALIDATE's and hence significantly more accurate than those presented in this work. However, the large number of parameters used in the fitting procedure would appear to indicate that there has been some overfitting of the data, and it would be important to see the performance of the scoring function on a larger test/training set in order to get a better feeling for the true accuracy. An interesting aspect of Jain's work is the approach adopted in preparing the training set and fitting the parameters. Once the functional form was established, the orientation of the ligand in the receptor and the conformation of the ligand were locally optimised. The results were fed back into the fitting procedure and used to obtain the free parameters in the scoring function. This is a good idea for docking applications since the minimum of the scoring function potential energy surface will be sampled during molecular docking and it is this configuration which should be used in fitting the function in the first place (provided it is close to the crystallographic minimum). The training set we have used is locally minimised using DISCOVER [14] and so should be reasonably well balanced for applications to similarly modelled complexes. Another interesting aspect of Jain's work is the careful consideration given to the problem of false positives which

provides motivation for the functional forms used in the fitting procedure.

One of the longer term goals of the current work is to establish a scoring function which could be used in docking applications. Clearly, the terms given here would need to have repulsive terms included before application to docking. The size and nature of the repulsive terms would be determined by the performance in docking tests. Also of interest is the more general suitability of our energy form to docking studies. It is important that the potential energy surface defined by a docking function does not contain a large number of false minima. Such frustration in the surface will lead to poor convergence in docking runs and the location of false binding modes. We have assessed the tendency of our function to identify some classes of false positives by studying geometries stemming from molecular modelling of database hits. It was these studies that led us to retain the angular dependence of the hydrogen bond function despite the poorer performance of this function on the training set. Another potential worry is that the polar contacts in our function are comparatively short range, and so docking algorithms might find it difficult to locate good contacts. However, it is questionable whether longer range functions would perform better. For example, a Coulomb-based term leads to substantial frustration in molecular mechanics functions because long-range attractions occur from polar atoms throughout the protein. Clearly, the problem of determining an optimal docking function involves making a difficult compromise, the results of which can only be properly tested in a large-scale docking study. However, we believe that our function is a good starting point and note that Rarey et al. [73] have successfully applied a function with similar characteristics to a large number of examples.

Other workers have attempted to fit binding affinities to geometric considerations of ligand–receptor binding [2–6,74–76] or to modelling data on the binding geometries [32,77–83]. Many of these have restricted their analysis to particular classes of receptors [2,3,5,6,32,74,75, 77–79,81,83]. It would be beyond the scope of this work to go into detail on these methods and the reader is referred to the recent article by Ajay and Murcko [84] which gives a good overview of computational approaches to the estimate of binding affinity. However, early forerunners to the approach adopted in this paper are worthy of specific mention. Williams and co-workers [2,3] tried to establish the general principles of drug–receptor binding from experimental information on the binding of small molecules to vancomycin and related antibiotics. Horton and Lewis [4] analysed a fairly small number of protein–ligand complexes (the majority of which contained large peptidic ligands) and demonstrated excellent correlations between the binding affinities and buried solvent-accessible areas coupled with solvation parameters. Krystek et al. [6] established a scoring function for serine proteases which used hydrophobic surface areas, a flexibility term and an electrostatic term, although its appropriateness for application to small-molecule ligands in other receptors is questionable.

One way to improve regressions is to remove complexes which do not fit the regression equation. It is unfortunate that it is often easy to rationalise why a particular complex is not appropriate for a study or, indeed, to introduce corrections to make it fit the study better. Such an approach is perfectly proper, but it is important to discuss where it has been used since this points the way to future improvements and provides useful caveats to other workers. In developing our preferred QSAR equation, no amendments were made to any of the terms described in order to get larger or smaller contributions from particular complexes. However, the following compounds were omitted from the study: 1ADC, 2ADA, 1STP. The binding affinities and predicted affinities for these compounds are (–26.69 [47]; –52.14), (–71.44 [85]; –31.14) and (–76.45 [1]; –40.64), respectively, in kJ/mol. These are bad outliers and their inclusion in the regression is deleterious.

1ADC is one of a related set of three complexes: 1ADB, 1ADC and 1ADF. The ligand structures for these complexes are shown in Fig. 10. The point of the original work [47,57] on these complexes was to show how small changes in the ligand lead to large changes in binding affinity, and to see how these changes are reflected in the crystal structure. In 1ADB, the pyridine nitrogen binds to a zinc ion and, presumably, this is favourable for binding. The 1ADB complex is predicted by the preferred regression to within an order of magnitude. 1ADC contains a ligand in which the pyridine nitrogen is moved to a position where it cannot bind to the zinc without affecting the binding of other moieties in the ligand. In other words, a favourable contact has been replaced by an unfavourable contact, since the zinc atom in the 1ADC structure is now not coordinated properly. It is the unfavourable aspect of the binding which cannot be taken into account in the scoring function, and this leads to a massive overprediction of binding affinity. It would be very difficult to change the scoring function in a physically justifiable way to improve the prediction, since the size of such unfavourable contacts could range from small to very large and there will not be enough information in the training set to make such a term transferable. This illustrates a serious weakness in the function which will be encountered when the function is applied to analogues of good binders, where the analogues contain unfavourable features. 1ADF is interesting because the ligand has a very similar structure to 1ADC, but in this case the crystallographic binding mode has changed considerably. Clearly, the old binding mode (i.e. 1ADC's binding mode) must be of similar energy to the new one, but the new binding mode contains no large unfavourable terms. Therefore, 1ADF is predicted well.

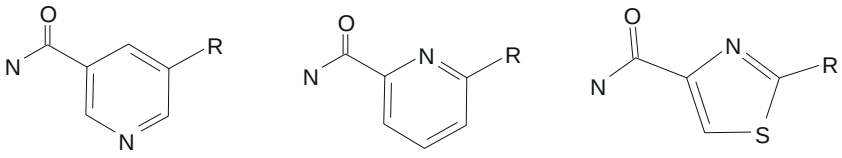


Fig. 10. Ligands in 1ADB, 1ADC and 1ADF.

2ADA contains a small ligand (see Fig. 11) with a very high binding affinity for adenosine deaminase. The binding affinity is inferred from an equilibrium constant and data for a related ligand, because the ketone form of the purine is normally favoured [85]. It is possible that this process of obtaining the binding constant is suspect, but this seems unlikely. The hydroxy compound is a perfect transition-state mimic for the enzyme, and it should be compared with the analogue where the hydroxy group is replaced by a hydrogen atom. The latter compound has a binding affinity of –28.53 kJ/mol, i.e. 8 orders of magnitude less [85]! It is difficult to imagine how any empirical scoring function could be expected to predict the scores of 2ADA *and* its dehydroxy analogue correctly. Again, the conclusion we come to is that one of the structures contains significant unfavourable terms which cannot be well represented in the empirical scoring function. Here the argument must be different than with 1ADC. The assumption must be that the original enzyme contains some unfavourable terms – perhaps the zinc atom which binds to the hydroxy group is not effectively coordinated. Binding of the dehydroxy compound makes no difference to the unfavourable terms and is fairly well predicted by the scoring function (if it has a similar binding mode to 2ADA, then the predicted score can be inferred). However, binding of the transition-state analogue causes the unfavourable situation to be removed and replaces it with a favourable one, the former effect being important in this case. It should be noted that if such large and *localised* unfavourable terms did exist in the unbound enzyme, then it might be expected that the bound conformation of the enzyme would probably *not* be the normal form. There would be an induced fit effect (as there is in 2ADA [86]) which converted some initial conformation into the bound conformation. The energy change for the shift in conformation would have to be reasonably low but would allow the creation of a *localised* unfavourable strain in the enzyme balanced by favourable interactions. The unfavourable strain energy could be relieved on binding of the transition-state analogue, leading to a large increase in affinity. Needless to say, if this type of mechanism were in operation in any of the complexes, the complex would be poorly predicted. It is known that mutations in the binding site regions of receptors can lead to a stabilisation of the receptor, supporting the idea of a 'generalised' strain energy in the apo-receptor which when relieved provides an extra component to the binding affinity [87]. The idea of generalised strain energy being important in enzymes has been discussed by Fersht [88]. Whilst we find the above argument attractive, it is clearly speculative and there may be other explanations for the large underprediction of 2ADA. Alternative explanations would need to provide a reason for the reduced binding affinity of the dehydroxy analogue.

The ligand in 1STP is biotin (see Fig. 12) and the unusually high binding affinity of biotin in streptavidin and avidin has been the subject of discussion in several previous papers [89–91]. There is also a need to explain the analogue data for biotin because small structural changes to the ureido ring have profound effects on the binding affinity [91]. Weber et al. [90] point out that the ureido oxygen of biotin is in an oxyanion hole in which it is hydrogen-bonded to three receptor atoms in a tetrahedral sp^3 arrangement. This indicates that it may be charged and that there is a large enthalpic contribution to the binding affinity as a result, as is supported by the calorimetric data [90]. An important part of Weber et al.'s argument is that the ureido oxygen is normally uncharged, meaning that the desolvation penalty for burying a charged group does not have to be paid and biotin binds more tightly than other molecules. The argument is persuasive and must account for some of the missing binding affinity. Böhm uses this argument to justify treating interactions with the ureido group as if they are ionic in nature and this causes the group to make a much larger contribution in his scoring function. As explained

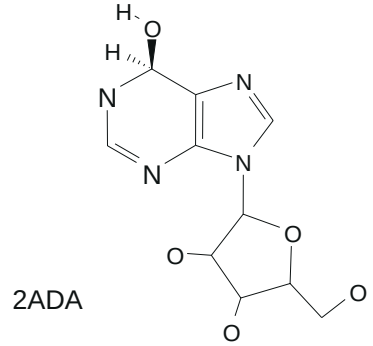


Fig. 11. Compound 6-hydroxy-1,6-dihydropurine ribonucleoside, which is the ligand in 2ADA.

earlier, our approach in such cases is to regard the complex as an outlier and simply not include it in the regression (in any case, our scoring function does not separate ionic and non-ionic contributions). It is also tempting to argue that there might be some generalised receptor strain energy which is relieved on the binding of biotin to streptavidin. This could stem from the imperfect solvation of the oxyanion hole prior to binding of biotin. In conclusion, the high binding affinity of biotin compared to other analogues is not well predicted by the scoring function for a variety of possible reasons, and perhaps new generations of scoring functions will find convincing ways to improve the situation.

The discussion of outliers leads to a more general discussion on the deficiencies of empirical scoring functions of the type given in this paper. Some of these will be discussed below.

Negative data The training set and test sets used in this paper only contain molecules which bind to the receptor. Usually in a design situation, one has a selection of molecules which all bind plausibly and one would like to rank them according to their predicted scores. This is a very different situation to a selection of crystallographically determined complexes for good ligands. It is likely that the empirical function will perform much more poorly in the real design situation than when testing internal consistency on other crystallographically determined complexes. The next paper in the series tests the function on a set of design candidates and illustrates the performance obtained [10]. It indicates the need for further QSAR in trying to assess the binding affinity of the next round of molecular designs.

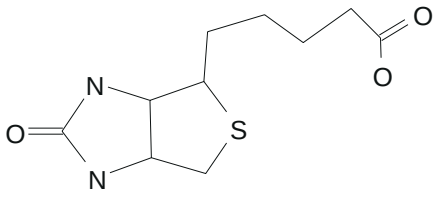
Binding mode The training set contains molecules where the binding geometry is known. However, in a real design situation the binding geometry must be inferred, and clearly a large range of empirical scores will be obtained from a set of alternative ligand–receptor geometries. The error bars inherent in the scoring function make it difficult to use the scoring function to decide on the best binding geometry. This can be particularly worrying when one wishes to calculate the binding affinity for a set of analogues where one often feels secure in assuming a similar binding mode. However, as with 1ADC above, if the wrong change is made to the starting structure, then an unfavourable energy term will be introduced that will not be picked up by the scoring function, or, as with 1ADF above, the binding geometry will change and this will not be picked up by the methods used to predict the binding geometry.

Binding affinities The regression analysis requires an accurate set of binding affinities. It also requires that the binding affinities be consistent with each other. For example, if the binding affinities are different at different values of pH, then all the binding affinities in the database must be at an equivalent pH. This problem was discussed earlier, where fairly small changes in pH led to quite different K_i measurements in the case of HIV-1 protease [20]. It is commonplace for different laboratories to report different measurements for the same binding affinity [1] due to a multiplicity of potential differences in experimental set-up. We have noticed a number of differences between the affinities reported in Tables 1–7 and the affinities collected by other groups, and this is why we have given references to the papers from which the binding affinities were extracted. Uncertainties in the experimental free energy of binding will place an intrinsic limit on the accuracy of empirical scoring functions of this type. Since we expect errors in the experimental K_i of about an order of magnitude, we cannot expect the scoring function to be more accurate than this.

Exclusion of important terms Clearly, the scoring function misses out some important criteria in binding affinity, and some of these have been discussed in other parts of the paper. Among the most important of these is the strain energy. However, there are a whole class of terms which have a deleterious effect on binding affinity which are poorly represented in the scoring function. If they are not present in the training set, it is impossible to arrive at a coefficient for even the most plausible of functional forms. Obviously, this point is connected to the problems of not having enough/any negative data in the training set. Additionally, there is no consideration of the receptor conformation prior to binding, and this is a problem if the receptor has some generalised strain energy built in, as has been argued in the case of 2ADA.

Problems with terms Scoring functions tend to assume an average contribution from each type of interaction. For example, in our case this means that a lipophilic contact always scores the same amount independent of the detailed nature of the environment. Clearly, contacts, and especially electrostatic contacts, are going to change in importance from one enzyme to the next, e.g. the S1 subsite of trypsin appears to be less tolerant of uncharged species than other trypsin-like serine proteases. These effects can in part be traced to inadequate treatments of long-range electrostatic and bulk partitioning effects.

It is our belief that all empirical scoring functions of the type discussed in this paper share the overall deficiencies outlined above. However, empirical scoring functions are



1STP

Fig. 12. Biotin, the ligand in 1STP.

still of enormous utility in de novo design, receptor-based database screening and docking [73,92]. Previous methods were simply not as good, and our experience of using an implementation of Böhm's original scoring function in applications of synthetically constrained de novo design demonstrated the practical value of such methodologies [9].

Conclusions

This paper has described the development of a simple empirical scoring function designed to estimate the free energy of binding for a protein–ligand complex when the 3D structure of the complex is known or can be approximated. The cross-validated prediction on a large training set of 82 complexes is 8.68 kJ/mol. It has also been applied to two test sets containing 20 and 10 complexes, respectively, and there have been extensive tests on the robustness of the function. Taken with the large number of complexes used in the analysis, then it is to be expected that the function is more robust than other empirical functions with similar philosophies of construction. A further publication will apply the function to a set of molecular designs which were later synthesised and assayed and will illustrate how follow-up QSAR can be used to improve the quality of binding affinity predictions during the course of a drug discovery research project.

Acknowledgements

C.W.M. and M.D.E. would like to thank David Frenkel for many useful discussions about this paper and the application of QSAR methodology.

References

1 Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
2 Williams, D.H., Cox, J.P.L., Doig, A.J., Gardner, M., Gerhard, U., Kaye, P.T., Lal, A.R., Nicholls, I.A., Salter, C.J. and Mitchell, R.C., J. Am. Chem. Soc., 113 (1991) 7020.
3 Searle, M.S., Williams, D.H. and Gerhard, U., J. Am. Chem. Soc., 114 (1992) 10697.
4 Horton, N. and Lewis, M., Protein Sci., 1 (1992) 169.
5 Bohacek, R.S. and McMartin, C., J. Med. Chem., 35 (1992) 1671.
6 Krystek, S., Stouch, T. and Novotny, J., J. Mol. Biol., 234 (1993) 661.
7 Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Green, S.M. and Marshall, G.R., J. Am. Chem. Soc., 118 (1996) 3959.
8 Murray, C.W., Clark, D.E., Auton, T.R., Firth, M.A., Li, J., Sykes, R.A., Waszkowycz, B., Westhead, D.R. and Young, S.C., J. Comput.-Aided Mol. Design, 11 (1997) 193.
9 Young, S.C., Waszkowycz, B.W., Clark, D.E., Li, J., Liebeschuetz, J.W., Lowe, R., Mahler, J., Martin, H., Murray, C.W., Rimmer, A.D. and Westhead, D.R., J. Med. Chem., to be submitted.
10 Murray, C.W., Auton, T.R. and Eldridge, M.D., J. Comput.-Aided Mol. Design, to be submitted.
11 Cruciani, G., Clementi, S. and Baroni, M., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 551–566.
12 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
13 INSIGHT, Molecular Simulations Inc., San Diego, CA, U.S.A.
14 DISCOVER, v. 2.9.5, Molecular Simulations Inc., San Diego, CA, U.S.A.
15 CFF95 force field, implemented in DISCOVER 2.9.5, Molecular Simulations Inc., San Diego, CA, U.S.A.
16 Dreyer, G.B., Lambert, D.M., Meek, T.D., Carr, T.J., Tomaszek Jr., T.A., Fernandez, A.V., Bartus, H., Cacciavillani, E., Hassell, A.M., Minnich, M., Petteway Jr., S.R. and Metcalf, B.W., Biochemistry, 31 (1992) 6646.
17 Hoog, S.S., Zhao, B., Winborne, E., Fisher, S., Green, D.W., Desjarlais, R.L., Newlander, K.A., Callahan, J.F., Moore, M.L., Huffman, W.F. and Abdel-Meguid, S., J. Med. Chem., 38 (1995) 3246.
18 Kim, E.E., Baker, C.T., Dwyer, M.D., Murcko, M.A., Rao, B.G., Tung, R.D. and Navia, M.A., J. Am. Chem. Soc., 117 (1995) 1181.
19 Jhoti, H., Singh, O.M.P., Weir, M.P., Cooke, R., Murray-Rust, P. and Wonacott, A., Biochemistry, 33 (1994) 8417.
20 Hosur, M.V., Bhat, T.N., Kempf, D., Baldwin, E.T., Liu, B., Gulnik, S., Wideburg, N.E., Norbeck, D.W., Appelt, K. and Erickson, J.W., J. Am. Chem. Soc., 116 (1994) 847.
21 Lam, P., Jadhav, P., Eyermann, C., Hodge, N., Ru, Y., Bacheler, L., Meek, J., Otto, M., Rayner, M., Wong, Y., Chang, C.-H., Weber, P., Jackson, D., Sharpe, T. and Erickson-Viitanen, S., Science, 263 (1994) 380.
22 Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder II, R.C., Cachau, R.E., Collins, J., Silva, A.M. and Erickson, J.W., Proc. Natl. Acad. Sci. USA, 90 (1993) 6796.
23 Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E. and James, M.N.G., Biochemistry, 31 (1992) 5201.
24 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A., Science, 246 (1989) 1149.
25 Fitzgerald, P.M.D., McKeever, B.M., Van Middlesworth, J.F., Springer, J.P., Heimbach, J.C., Leu, C.-T., Herber, W.K., Dixon, R.A.F. and Darke, P.L., J. Biol. Chem., 265 (1990) 14209.
26 Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H. and Wlodawer, A., Proc. Natl. Acad. Sci. USA, 87 (1990) 8805.
27 Perona, J.J., Tsu, C.A., Craik, C.S. and Fletterick, R.J., J. Mol. Biol., 230 (1993) 934.
28 Brandstetter, H., Turk, D., Hoeffken, H.W., Grosse, D., Stuerzebecher, J., Martin, P.D., Edwards, B.F.P. and Bode, W., J. Mol. Biol., 226 (1992) 1085.
29 Bode, W., Turk, D. and Stuerzebecher, J., Eur. J. Biochem., 193 (1990) 175.
30 Turk, D., Stuerzebecher, J. and Bode, W., FEBS Lett., 287 (1991) 133.
31 Nienaber, V.L., Mersinger, L.J. and Kettner, C.A., Biochemistry, 35 (1996) 9690.
32 Kurinov, I. and Harrison, R.W., Nat. Struct. Biol., 1 (1994) 735.
33 Mangani, S., Orioli, P. and Carloni, P., J. Mol. Biol., 223 (1992) 573.
34 Stams, T., Spurlino, J.C., Smith, D.L., Wahl, R.C., Ho, T.F., Qoronfleh, M.W., Banks, T.M. and Rubin, B., Nat. Struct. Biol., 1 (1994) 119.
35 Tronrud, D.E., Monzingo, A.F. and Matthews, B.W., Eur. J. Biochem., 157 (1986) 261.
36 Mozingo, A.F. and Matthews, B.W., Biochemistry, 23 (1984) 5724.

37 Christianson, D.W. and Lipscomb, W.N., Proc. Natl. Acad. Sci. USA, 83 (1986) 7568.
38 Holmes, M.A. and Matthews, B.W., Biochemistry, 20 (1981) 6912.
39 Kim, H. and Lipscomb, W.N., Biochemistry, 29 (1990) 5546.
40 Matthews, B.W., Acc. Chem. Res., 21 (1988) 333.
41 Kim, H. and Lipscomb, W.N., Biochemistry, 30 (1991) 8171.
42 Vermersch, P.S., Tesmer, J.J.G., Lemon, D.D. and Quiocho, F.A., J. Biol. Chem., 265 (1990) 16592.
43 Harris, E., Aleshin, A., Firsov, L. and Honzatko, R.B., Biochemistry, 32 (1993) 1618.
44 Cygler, M., Rose, D.R. and Bundle, D.R., Science, 253 (1991) 442.
45 Burmeister, W.P., Henrissat, B., Bosso, C., Cusack, S. and Ruigrok, R.W.H., Structure, 1 (1993) 19.
46 Vermersch, P.S., Tesmer, J.J.G., Lemon, D.D. and Quiocho, F.A., Biochemistry, 30 (1991) 6861.
47 Goldstein, B.M., Li, H., Jones, J.P., Ellis Bell, J., Zeidler, J., Pankiewicz, K.W. and Watanabe, K.A., J. Med. Chem., 37 (1994) 392.
48 Scapin, G., Blanchard, J.S. and Sacchettini, J.C., Biochemistry, 34 (1995) 3502.
49 Wedekind, J.E., Reed, G.H. and Rayment, I., Biochemistry, 33 (1994) 9333.
50 Yao, N., Trakhanov, S. and Quiocho, F.A., Biochemistry, 33 (1994) 4769.
51 Adams, M.J., Ellis, G.H., Gover, S., Naylor, C.E. and Phillips, C., Structure, 2 (1994) 651.
52 Lipscomb, J.D., Biochemistry, 19 (1980) 3590.
53 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Proteins Struct. Funct. Genet., 8 (1990) 44.
54 Guddat, L.W., J. Mol. Biol., 236 (1993) 247.
55 Fisher, M.T. and Sligar, S.G., J. Am. Chem. Soc., 107 (1985) 5081.
56 Appelt, K., Bacquet, R.J., Bartlett, 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. and White, J., J. Med. Chem., 34 (1991) 1834.
57 Goldstein, B.M., Ellis Bell, J. and Marquez, V.E., J. Med. Chem., 33 (1990) 1123.
58 Arevalo, J.H., Taussig, M.J. and Wilson, I.A., Nature, 365 (1993) 859.
59 Liaw, S.-H., Jun, G. and Eisenberg, D., Biochemistry, 33 (1994) 11184.
60 Edmundson, A.B., Harris, D.L., Fan, Z.-C., Guddat, L.W., Schley, B.T., Hanson, B.L., Tribbick, G. and Geysen, H.M., Proteins Struct. Funct. Genet., 16 (1993) 246.
61 Erikson, J., Neidhart, D.J., Van Drie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., Science, 249 (1990) 527.
62 Cooper, J.B., Quail, W., Frazao, C., Foundling, S.I., Blundell, T.L., Humblet, C., Lunney, E.A., Lowther, W.T. and Dunn, B.M., Biochemistry, 31 (1992) 8142.
63 Veerapandian, B., Cooper, J.B., Sali, A., Blundell, T.L., Rosati, R.L., Dominy, B.W., Damon, D.B. and Hoover, D.J., Protein Sci., 1 (1992) 322.
64 Lunney, E.A., Hamilton, H.W., Hodges, J.C., Kaltenbrohn, J.S., Repine, J.T., Badasso, M., Cooper, J.B., Dealwis, C., Wallace, B.A., Lowther, W.T., Dunn, B.M. and Humblet, C., J. Med. Chem., 36 (1993) 3809.
65 Foundling, S.I., Cooper, J.B., Watson, F.E., Cleasby, A., Pearl, L.H., Sibanda, B.L., Hemmings, A., Wood, S.P., Blundell, T.L., Valler, M.J., Norey, C.G., Kay, J., Boger, J., Dunn, B.M., Leckie, B.J., Jones, D.M., Atrash, B., Hallet, A. and Szelke, M., Nature, 327 (1987) 349.
66 Veerapandian, B., Cooper, J.B., Sali, M.A. and Blundell, T.L., J. Mol. Biol., 216 (1990) 1017.
67 Sali, A., Veerapandian, B., Cooper, J.B., Foundling, S.I., Hoover, D.J. and Blundell, T.L., EMBO J., 8 (1989) 2179.
68 Jain, A.J., J. Comput.-Aided Mol. Design, 10 (1996) 427.
69 Fersht, A.R., Shi, J., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. and Winter, G., Nature, 314 (1985) 235.
70 Nicklaus, M.C., Wang, S., Milne, G.W.A. and Driscoll, J.S., Bioorg. Med. Chem., 3 (1995) 411.
71 Wireko, F.C., Kellogg, G.E. and Abraham, D.J., J. Med. Chem., 34 (1991) 758.
72 Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
73 Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
74 Nauchitel, V., Villaverde, M.C. and Sussman, F., Protein Sci., 4 (1995) 1356.
75 Verkhivker, G., Appelt, K., Freer, S.T. and Villafranca, J.E., Protein Eng., 8 (1995) 677.
76 Wallqvist, A., Jernigan, R.L. and Covell, D.G., Protein Sci., 4 (1995) 1881.
77 Carson, M., Yang, Z., Babu, Y.S. and Montgomery, J.A., Acta Crystallogr., D51 (1995) 536.
78 Grootenhuis, P.D.J. and Van Galen, P.J.M., Acta Crystallogr., D51 (1995) 560.
79 Vajda, S., Weng, Z.P., Rosenfeld, R. and Delisi, C., Biochemistry, 33 (1994) 13977.
80 Aqvist, J., Medina, C. and Samuelsson, J.E., Protein Eng., 7 (1994) 385.
81 Jedrzejas, M.J., Singh, S., Brouillette, W.J., Air, G.M. and Luo, M., Proteins Struct. Funct. Genet., 23 (1995) 263.
82 Ortiz, A.R., Pisabarro, M.T., Gago, F. and Wade, R.C., J. Med. Chem., 38 (1995) 2681.
83 Holloway, M.K., Wai, J.M., Halgren, T.A., Fitzgerald, P.M.D., Vacca, J.P., Dorsey, B.D., Levin, R.B., Thompson, W.J., Chen, L.J., DeSolms, S.J., Gaffin, N., Ghosh, A.K., Giuliana, E.A., Graham, S.L., Guare, J.P., Hungate, R.W., Lyle, T.A., Sanders, W.M., Tucker, T.J., Wiggins, M., Wiscount, C.M., Woltersdorf, O.W., Young, S.D., Darke, P.L. and Zugay, J.A., J. Med. Chem., 38 (1995) 305.
84 Ajay and Murcko, M.A., J. Med. Chem., 38 (1995) 4953.
85 Kati, W.M. and Wolfenden, R., Science, 243 (1989) 1591.
86 Wilson, D.K., Rudolph, F.B. and Quiocho, F.A., Science, 252 (1991) 1278.
87 Shoichet, B.K., Baase, W.A., Kuroki, R. and Matthews, B.W., Proc. Natl. Acad. Sci. USA, 92 (1995) 452.
88 Fersht, A.R., Enzyme Structure and Mechanism, 2nd ed., Freeman, New York, NY, U.S.A., 1984.
89 Weber, P.C., Pantoliano, M.W. and Thompson, L.D., Biochemistry, 31 (1992) 9350.
90 Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. and Salemme, F.R., J. Am. Chem. Soc., 114 (1992) 3197.
91 Pugliese, L., Coda, A., Malcovati, M. and Bolognesi, M., J. Mol. Biol., 231 (1993) 698.
92 Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.