J-CAMD 247

# The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure

Hans-Joachim Böhm

*BASF AG, Central Research, D-67056 Ludwigshafen, Germany*

Received 20 September 1993
Accepted 15 December 1993

*Key words:* Proteins; Protein–ligand interaction; De novo design; Scoring function

## SUMMARY

A new simple empirical function has been developed that estimates the free energy of binding for a given protein–ligand complex of known 3D structure. The function takes into account hydrogen bonds, ionic interactions, the lipophilic protein–ligand contact surface and the number of rotatable bonds in the ligand. The dataset for the calibration of the function consists of 45 protein–ligand complexes. The new energy function reproduces the binding constants (ranging from $2.5 \cdot 10^{-2}$ to $4 \cdot 10^{-14}$ M, corresponding to binding energies between –9 and –76 kJ/mol) of the dataset with a standard deviation of 7.9 kJ/mol, corresponding to 1.4 orders of magnitude in binding affinity. The individual contributions to protein–ligand binding obtained from the scoring function are: ideal neutral hydrogen bond: –4.7 kJ/mol; ideal ionic interaction: –8.3 kJ/mol; lipophilic contact: –0.17 kJ/mol Å$^2$; one rotatable bond in the ligand: +1.4 kJ/mol. The function also contains a constant contribution (+5.4 kJ/mol) which may be rationalized as loss of translational and rotational entropy. The function can be evaluated very fast and is therefore also suitable for application in a 3D database search or de novo ligand design program such as LUDI.

## INTRODUCTION

The accurate and fast prediction of the binding constant for a protein–ligand complex of known three-dimensional structure is of paramount importance in structure-based drug design. The current approaches for 3D database searching or de novo design of protein ligands all attempt to prioritize the retrieved or generated structures [1–6]. Following the pioneering ideas of Goodford [7,8], a number of groups use a force field precalculated on grid points to score putative ligands [3–5]. In some approaches the force field is augmented by a term describing the solvation effects. In addition, Moon and Howe also include a conformational energy contribution for prioritization [6]. The first version of LUDI [1,2] uses a simple empirical function that accounts for hydrogen bonds and hydrophobic interactions. So far, none of these methods attempts to provide a quantitative estimate for the binding energy.

In this communication a new simple function is described that computes an estimate for the free energy of binding $\Delta G$ for a protein–ligand complex of known 3D structure. In the development of the energy function it was required that it should be fast enough to be applicable as scoring function in the de novo ligand design program LUDI [1,2]. Therefore, the computation time to evaluate the energy function for a given protein with more than 100 different putative ligands should be less than one minute on a workstation. A second requirement was that the function should be able to reject putative protein–ligand complexes with bad hydrogen-bond geometries.

The present approach is believed to have a number of advantages compared to the scoring functions that have been used so far. First, it is extremely fast. For a given protein it allows the scoring of 10 ligands per second on a current single-processor UNIX workstation. Second, the scoring function was devised in such a way that it can cope with small uncertainties in the atomic coordinates of the protein structure. Third, it accounts for the major entropic contributions to the binding affinity. Previous work by Williams and co-workers [9–11] and others [12,13] proposed methods to estimate binding constants. The present approach builds partly upon this earlier work. We now use a fairly large number of 45 protein–ligand complexes as basis to fit a free energy function to the experimental binding data. In comparison with the work by Horton and Lewis [12] it should be noted that the main focus of our investigation is on the binding of small and flexible ligands to proteins.

## METHODS

The following free energy function was used:

$$\begin{aligned}\Delta G_{\text{binding}} = \Delta G_0 &+ \Delta G_{\text{hb}} \Sigma_{\text{h-bonds}} f(\Delta R,\Delta\alpha) + \Delta G_{\text{ionic}} \Sigma_{\text{ionic int.}} f(\Delta R,\Delta\alpha)\\ &+ \Delta G_{\text{lipo}} |A_{\text{lipo}}| + \Delta G_{\text{rot}}\, \text{NROT}\end{aligned}$$

$$f(\Delta R,\Delta\alpha) = f1(\Delta R)\, f2(\Delta\alpha)$$

$$f1(\Delta R) = \begin{cases} 1 & \Delta R \le 0.2\ \text{Å} \\ 1-(\Delta R-0.2)/0.4 & \Delta R \le 0.6\ \text{Å} \\ 0 & \Delta R > 0.6\ \text{Å} \end{cases}$$

$$f2(\Delta\alpha) = \begin{cases} 1 & \Delta\alpha \le 30^\circ \\ 1-(\Delta\alpha-30)/50 & \Delta\alpha \le 80^\circ \\ 0 & \Delta\alpha > 80^\circ \end{cases}$$

$f(\Delta R,\Delta\alpha)$ is a penalty function which accounts for large deviations of the hydrogen-bond geometry from ideality. The functional form of $f(\Delta R,\Delta\alpha)$ is the same as in our previous work [2]. $\Delta R$ is the deviation of the H··O/N hydrogen-bond length from the ideal value of 1.9 Å; $\Delta\alpha$ is the deviation of the hydrogen-bond angle $\angle_{\text{N/O-H··O/N}}$ from its ideal value of 180°. The function tolerates small deviations of up to 0.2 Å and 30° from the ideal geometry; such deviations are often due to small uncertainties in the X-ray structure. In order to assess the influence of the penalty function $f(\Delta R,\Delta\alpha)$ on the values of the individual contributions to $\Delta G$ in the scoring function, we have also investigated an energy function with $f(\Delta R,\Delta\alpha)$ set to 1.

The present function does not include a term accounting for distortions in the hydrogen-bond angle $\angle_{C=O\cdot\cdot O/N}$. The statistical analysis of experimental $C=O\cdot\cdot H-N$ hydrogen-bond geometries reveals a weak preference for the angle $\angle_{C=O\cdot\cdot N}$ at about 120°. However, the whole range from 110–180° is significantly populated. Therefore, in view of this large observed spread and the much tighter distributions of bond lengths and $\angle_{N/O-H\cdot\cdot O/N}$ angles it was decided not to penalize deviations from 120° of the hydrogen-bond angle $\angle_{C=O\cdot\cdot O/N}$. It should also be kept in mind that other hydrogen-acceptor groups such as –O– or =N– show different preferences for the corresponding angle (see Ref. 2 for a short review of experimental data). The incorporation of these differences would require a significantly more complex scoring function.

$\Delta G_0$ is a contribution to the binding energy that does not directly depend on any specific interactions with the protein. It may be rationalized as a reduction of binding energy due to overall loss of translational and rotational entropy of the ligand.

$\Delta G_{hb}$ describes the contribution from an ideal hydrogen bond. $\Delta G_{ionic}$ represents the contribution from an unperturbed ionic interaction. The same geometric dependency is assumed for uncharged and charged interactions. For interactions with a zinc or iron atom, $f2(\Delta\alpha)$ is set to 1 and the ideal distance metal··O/N is set to 2 Å.

$\Delta G_{lipo}$ represents the contribution from lipophilic interactions. It is assumed that such interactions are proportional to $A_{lipo}$, the lipophilic contact surface between the protein and the ligand. We have developed a fast approximate method to calculate $A_{lipo}$. The algorithm is based on a cubic grid with 1 Å grid spacing. This rather coarse grid was chosen to speed up the calculation (which is the most time-consuming part of the present scoring function). In this grid, all cubes are marked that overlap with the ligand. This criterion is fulfilled if any of the ligand atoms is closer to the center of the cube than the van der Waals radius of the ligand atom. Then all cubes adjacent to the ligand are marked and this list is searched for those cubes that overlap with the protein. The number of overlapping cubes is proportional to the size of the protein–ligand contact surface. The actual value of the contact surface is then determined by multiplying the number of contact cubes with a calibration factor, derived from a comparison with surface areas calculated by Connolly's MS program [14]. We use a calibration factor of 0.71 to convert the calculated number of contact cubes into an approximate surface area. For example, for benzene, benzamidine and retinol the present approach yields 180, 211 and 457 surface cubes which can be translated into surface areas of 128, 151 and 326 Å$^2$, respectively. The corresponding surface areas obtained from the MS program are 121, 157 and 352 Å$^2$. The protein–ligand contact surfaces calculated with the present approach are 118 Å$^2$ for the complex trypsin–benzamidine (78% of the ligand surface is calculated to be buried) and 295 Å$^2$ for the complex retinol-binding protein–retinol (90% of the ligand surface buried). Each grid point is marked as being either lipophilic or polar, thus the lipophilic part of the contact surface can easily be determined.

The final parameter $\Delta G_{rot}$ describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand. The number of rotatable bonds NROT is taken as the number of acyclic $sp^3$-$sp^3$ and $sp^3$-$sp^2$ bonds. Rotations of terminal -$CH_3$ or -$NH_3$ groups are not taken into account. The flexibility of cyclic portions of the ligand is ignored at present.

The list of protein–ligand complexes used in the present study is given in Table 1. The coordinates of the protein–ligand complex were taken from the Brookhaven Protein Data Bank

TABLE 1
PROTEIN–LIGAND COMPLEXES USED FOR THE CALIBRATION OF THE FREE ENERGY FUNCTION

| Protein–ligand complex | PDB entry | –log $K_i$ pred. | –log $K_i$ expt. | Ref. |
|---|---|---|---|---|
| Trypsin–benzamidine | 3PTB | 5.49 | 4.74 | 16 |
| Trypsin–butylamine | 3PTB[a] | 3.05 | 2.82 | 16 |
| Trypsin–benzylamine | 3PTB[a] | 3.63 | 3.42 | 16 |
| Trypsin–phenylguanidine | 3PTB[a] | 4.83 | 4.14 | 16 |
| Thrombin–NAPAP | –[b] | 7.92 | 8.52[c] | 17 |
| Thrombin–MQPA | –[b] | 5.80 | 7.40 | 18 |
| Thrombin–TAPAP | –[b] | 6.11 | 6.19 | 19 |
| Thrombin–benzamidine | 1DWB | 3.92 | 2.92 | 19 |
| Thrombin–amidinopiperidine | 1DWB[a] | 4.09 | 3.82 | 20 |
| Chymotrypsin–benzene | 4CHA[a] | 2.04 | 1.60 | 21 |
| Chymotrypsin–phenole | 4CHA[a] | 2.54 | 2.19 | 21 |
| Chymotrypsin–indole | 4CHA[a] | 2.74 | 3.10 | 21 |
| Chymotrypsin–benzoquinoline | 4CHA[a] | 3.25 | 4.20 | 21 |
| Thermolysin–ZF$^{P}$LA | 4TMN | 9.66 | 10.19 | 22 |
| Thermolysin–ZG$^{P}$LL | 5TMN | 6.81 | 8.04 | 22 |
| Thermolysin–phosphoramidon | 1TLP | 6.42 | 7.55 | 22 |
| Thermolysin–CLT | 1TMN | 9.16 | 7.30 | 22 |
| Thermolysin–thiorphan | 5TMN[d] | 6.15 | 5.64 | 22 |
| Thermolysin–Leu-NHOH | 4TLN | 3.62 | 3.72 | 22 |
| Renin–CGP38560 | 1RNE | 8.20 | 8.70[e] | 23 |
| Endothiapepsin–H256 | 2ER6 | 5.75 | 7.22 | 24 |
| Endothiapepsin–Ac-pepstatin | 4ER2[f] | 5.68 | 8.04 | 25 |
| Endothiapepsin–H142 | 4ER4 | 8.05 | 6.79 | 26 |
| HIV protease–A74704 | 9HVP | 8.85 | 8.35 | 27 |
| HIV protease–L700,417 | 4PHV | 11.37 | 9.15[e] | 28 |
| HIV protease–MVT101 | 4HVP | 9.06 | 6.15 | 29 |
| DHFR–methotrexate | 4DFR | 9.05 | 9.70 | 30 |
| DHFR–2,4-diaminopteridine | 4DFR[g] | 5.50 | 6.00 | 31 |

(PDB) [15] if not specified otherwise. We have selected 45 structures with known $K_i$ values [16–46] as basis for the calibration of the energy function. The experimental binding data were taken from the literature, without any further modifications. We did not check the experimental data for differences in temperature or salt concentrations during measurement. In the selection of the structures, we have attempted to cover a broad spectrum of different types of protein–ligand complexes. The experimentally observed $K_i$ values range from 25 mM (chymotrypsin–benzene) to 40 fM (streptavidin–biotin) and cover 12 orders of magnitude. The molecular weights of the ligands vary between 66 and 1047. The number of rotatable bonds in the ligands varies between 0 and 29 (endothiapepsin–H142). Some of the ligands bind completely via lipophilic interactions (e.g. retinol-binding protein–retinol, chymotrypsin–benzene). Other ligands have very little lipophilic interactions with the protein and achieve their binding predominantly through hydrogen bonding (e.g. galactose-binding protein–galactose, concanavalin A–α-methyl-mannosid). The majority of the protein–ligand complexes exhibits both substantial lipophilic contacts and a number of hydrogen bonds and ionic interactions. However, the list is by no means complete and may reflect a certain bias. Some trypsin–inhibitor and thermolysin–inhibitor complexes were excluded from the calibration dataset in order to keep the bias towards these enzymes low. These structures were later used to test the new scoring function. In some cases

TABLE 1
(continued)

| Protein–ligand complex | PDB entry | $-\log K_i$ pred. | $-\log K_i$ expt. | Ref. |
|---|---|---|---|---|
| Thymidylate synthase–CB3717 | 2TSC | 6.04 | 8.52 | 32 |
| Thymidylate synthase–Cmpd 3 | 2TSC[g] | 6.16 | 5.40 | 32 |
| Streptavidin–biotin | 1STP | 10.75 | 13.40 | 33 |
| Retinol-binding protein–retinol | 1RBP | 7.49 | 6.72 | 34 |
| Fatty acid-binding protein–$C_{15}COOH$ | 2IFB | 5.28 | 5.43 | 35 |
| Galactose-binding protein–galactose | 2GBP | 6.63 | 7.60 | 36 |
| FKFB–FK506 | 1FKF | 7.15 | 9.70 | 37 |
| Virus coat protein–Cmpd 4 | 2R04 | 5.74 | 6.22[h] | 38 |
| PHBH–*p*-hydroxybenzoic acid | 2PHH | 7.02 | 4.68 | 39 |
| Concanavalin A–methylmannosid | 4CNA | 3.47 | 2.00 | 40 |
| Myoglobin–imidazole | 1MBI | 3.49 | 1.88 | 41 |
| Hemagglutinin–sialic acid | 4HMG | 3.34 | 2.55 | 42 |
| TIM–phosphoglycolic acid | 2YPI | 3.34 | 4.82 | 25 |
| Carboxypeptidase A–GT | 3CPA | 5.94 | 3.88 | 43 |
| Carboxypeptidase A–$ZAA^P(O)F$ | 6CPA | 10.89 | 11.52 | 44 |
| Xylose isomerase–xylitol | 2XIS | 6.40 | 5.82 | 25 |
| PNP–guanine | 1ULB | 5.12 | 5.30 | 25 |

The predicted $-\log K_i$ values refer to results obtained with function #2.

[a] The ligand was positioned into the protein binding site using the program LUDI [1,2].

[b] The protein structure determined by Brandstetter et al. [45] was used.

[c] The reported $K_i$ of 6 nM was obtained for the racemate. We used $K_i$ = 3 nM, as only one enantiomer binds to the protein.

[d] The protein coordinates were taken from 5TMN and the ligand coordinates were taken from Ref. 46. The S··Zn interaction present in this complex was counted as an ionic interaction.

[e] $IC_{50}$ value.

[f] The ligand was taken from the X-ray structure and the acetyl group was appended.

[g] The ligand was taken from the X-ray structure and part of the ligand was removed in order to obtain the specified structure.

[h] MIC: minimum inhibitory concentration.

the ligand was docked into the protein binding sites to obtain a reasonable 3D structure of the complex. The structures were taken as stored in the Brookhaven PDB. Water molecules were not taken into account. Hydrogens were added using the graphics program INSIGHT [47]. For the side-chain hydroxyl groups of serine, threonine and tyrosine, the hydrogen atoms were positioned according to the hydrogen-bond pattern observed in the structure. In all other cases (when the position of the hydrogen could not be deduced from the surrounding atoms), the hydrogen atoms were positioned in the trans orientation in an H–O–C–C fragment. No energy minimization was carried out on any of the structures used in the present study. The amino acids aspartic acid, glutamic acid, lysine and arginine were assumed to be charged if not stated otherwise by the authors of the structure. Histidine side chains were protonated as indicated in the original publications. Accordingly, in some cases histidine was assumed to be charged. In the ligands, the following groups were assumed to be charged: phosphate, carboxylate, guanidinium, amidinium and amine.

A special problem arose with the complex streptavidin–biotin. The ureido group of biotin is formally uncharged. However, it was suggested by Weber et al. [33] that, due to polarization, the ureido group carries a negative charge on the oxygen atom and a positive charge spread over the heterocycle. Therefore, in the derivation of the energy function the interaction between the ureido moiety of biotin and streptavidin was assumed to be partly formed by ionic interactions.

In eight cases the 3D structure of the protein–ligand complex was constructed by taking the protein structure from the PDB and docking the ligand with LUDI. This procedure was carried out for the trypsin ligands phenylguanidine, benzylamine and butylamine, the thrombin ligand amidinopiperidine and the chymotrypsin ligands benzene, phenole, indole and benzo[*f*]quinoline. Recently, we have shown for a number of cases with known 3D structure of the protein–ligand complex that the ligands positioned by LUDI are very close to the experimentally observed structures [1]. In three cases the 3D structure of a protein–ligand complex was constructed by taking the protein structure from the PDB and generating the position of the ligand by modifying a closely related structure. In two cases this involved simply removing a side chain from the ligand with known 3D structure (dihydrofolate reductase–2,4-diaminopteridine, thymidylate synthase – compound 3 from Ref. 32). In both cases, the unchanged protein–ligand complex from the PDB was included in the dataset as well. In the third case, an acetyl group was appended to pepstatin in a reasonable geometry to generate the complex endothiapepsin–acetylpepstatin. In this case we did not use the original protein–ligand complex, because we could not find binding data for it. In all three cases identical binding modes were assumed for the known protein–ligand complex and the modelled structure. The protein part of the complex was left unchanged.

The present work primarily aims at the description of the binding of small ligands to proteins. We have therefore not included proteins with macromolecular ligands (e.g. trypsin–pancreatic trypsin inhibitor). The physics of binding is clearly the same as for low-molecular-weight ligands. However, if the ligand is a protein it will usually fold into a stable conformation before binding. In this situation, the number of rotatable bonds in the ligand is not a realistic measure for the entropy loss, because these ligands will more or less behave as rigid bodies. It has been shown recently by Horton and Lewis [12] that the binding of macromolecular ligands can be described without accounting for the internal flexibility of the ligand.

TABLE 2

INDIVIDUAL CONTRIBUTIONS, STANDARD DEVIATIONS s AND CORRELATION COEFFICIENTS r OBTAINED FROM A FIT OF FREE ENERGY FUNCTIONS #1–#4 TO EXPERIMENTAL BINDING CONSTANTS OF 45 PROTEIN–LIGAND COMPLEXES

| Function | $\Delta G_0$ | $\Delta G_{hb}$ | $\Delta G_{ionic}$ | $\Delta G_{lipo}$ | $\Delta G_{rot}$ | s | r |
|---|---|---|---|---|---|---|---|
| 1 | +5.3 | -5.7 | -5.7[a] | –0.18 | 1.8 | 8.5 | 0.848 |
| 2[b] | +5.4 | -4.7 | -8.3 | –0.17 | 1.4 | 7.9 | 0.873 |
| 2a | +4.7 | –4.7 | -8.3 | –0.17 | 1.4 | 8.7 | 0.827 |
| 3 | +3.2 | -3.0[c] | -6.8[c] | –0.15 | 1.2 | 9.3 | 0.821 |
| 4 | +41.8[d] | -9.5 | -12.9 | –0.34 | 3.4 | 14.0 | 0.862 |
| 5 | 0.0[e] | –4.0 | -7.6 | –0.14 | 1.1 | 8.1 | 0.872 |

All values except r are given in kJ/mol. Function #2a was obtained from a subset of 37 protein–ligand complexes (see text).

[a] This value is set to be equal to $\Delta G_{hb}$.

[b] This function has been implemented in the de novo design program LUDI [1,2]. The 95% confidence intervals are: $\Delta G_0$: ±150%, $\Delta G_{hb}$: ±36%, $\Delta G_{ionic}$: ±27%, $\Delta G_{lipo}$: ±27%, $\Delta G_{rot}$: ±40%.

[c] $f(\Delta R, \Delta\alpha)$ set to 1.

[d] In function #4, $\Delta G_0$ was set to 41.8 kJ/mol.

[e] In function #5, $\Delta G_0$ was set to 0.0 kJ/mol.

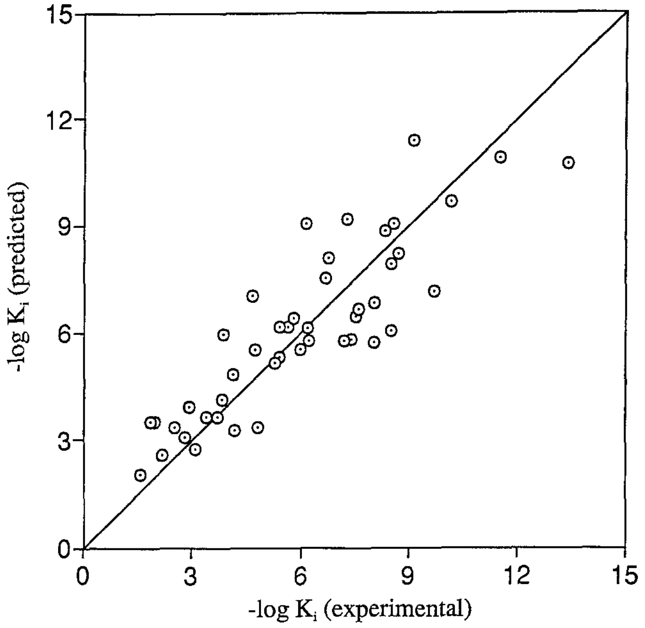

Fig. 1. Plot of estimated binding constants $K_i$, obtained from function #2, versus experimentally observed values for 45 protein–ligand complexes.

The experimental $K_i$ values were converted into free energies of binding $\Delta G$ using $\Delta G = -RT \ln K_i$ ($T = 298$ K). A least-squares fit was then performed to obtain the adjustable parameters in the free energy function.

## RESULTS

We have investigated a number of different functions as representations of $K_i$ or binding energies, respectively. The results are summarized in Table 2. The function #1 contains four adjustable parameters, i.e., $\Delta G_0$, $\Delta G_{hb,}$ $\Delta G_{lipo}$ and $\Delta G_{rot}$. Here, it is assumed that $\Delta G_{hb} = \Delta G_{ionic}$. The standard deviation is 8.5 kJ/mol, corresponding to an uncertainty of 1.5 orders of magnitude in $K_i$. The function #2, with five adjustable parameters $\Delta G_0$, $\Delta G_{hb,}$ $\Delta G_{ionic}$, $\Delta G_{lipo}$ and $\Delta G_{rot}$, yields a slightly smaller standard deviation of 7.9 kJ/mol, corresponding to 1.4 orders of magnitude in $K_i$ (number of structures $n = 45$, correlation coefficient $r = 0.873$, standard deviation $s = 7.9$ kJ/mol, Fisher significance ratio $F = 32.1$). The individual contributions to the binding are also given in Table 2. A plot of the predicted $K_i$ values versus the experimentally observed data is shown for function #2 in Fig. 1. We have also tested an alternative form of function #2, with $f(\Delta R, \Delta\alpha)$ set to 1 for all polar interactions. This function (#3) therefore ignores the effect of distortions on the strength of polar interactions. The standard deviation obtained in the fit of function #3 to the experimental binding data is 9.3 kJ/mol.

The present dataset contains eight structures where the ligand was docked with LUDI. In order to assess the influence of these eight structures on the parameters in the energy function, we have also investigated a function fitted to the dataset with the eight LUDI structures omitted. The individual contributions obtained from this fit to the binding constants of 37 protein–ligand complexes (see function #2a in Table 2) do not differ by more than 5% from those of function #2, with the exception of $\Delta G_0$ which is 15% smaller. We therefore conclude that the use of some structures with the ligand docked by LUDI does not introduce artefacts or biases into the energy function.

In addition, we have investigated a modified form of function #2, in which $\Delta G_0$ was assumed to be proportional to the logarithm of the molecular mass. The form was chosen to account roughly for the mass dependence of the translational and rotational entropy (cf. Fig. 7 in Ref. 9). However, this function did not yield an improved description of the binding energies (standard deviation 8.1 kJ/mol) compared to function #2 and was therefore not further considered.

In two other modifications of function #2, $\Delta G_0$ was first fixed at 41.8 kJ/mol (function #4) and then set to zero (function #5). Function #4 yields a large standard deviation of 14 kJ/mol. Function #5, with $\Delta G_0 = 0.0$ kJ/mol, is only marginally worse than function #2 with a standard deviation of 8.1 kJ/mol.

Function #2 was further examined using cross-validation. This is a technique where each object is eliminated once (leave-one-out) from the dataset and its affinity is predicted by the model derived from all other objects. The process is repeated n times (n = number of objects). The cross-validated values

$$r^2_{\text{press}} = q^2 = 1 - \Sigma(y_{\text{pred}} - y_{\text{obs}})^2 / \Sigma(y_{\text{obs}} - y_{\text{mean}})^2 = 0.696$$

($y_{\text{mean}}$ = the mean of the y value included to derive the corresponding model), and

$$s_{\text{press}} = \text{SQRT}(\Sigma(y_{\text{pred}} - y_{\text{obs}})^2 / (n - k - 1)) = 9.3 \text{ kJ/mol}$$

(k = number of variables) give evidence for the stability and the predictive power of the derived model.

Function #2 was further applied to nine protein–ligand complexes of known 3D structure which were not included in the calibration dataset. The results are summarized in Table 3. The experimentally determined binding constants [22,48–51] are predicted by function #2 with an rms deviation of 1.67 (log $K_i$), corresponding to an error of 9.6 kJ/mol.

We then used the enzyme dihydrofolate reductase from *Lactobacillus casei* as a further test case

TABLE 3
PROTEIN–LIGAND COMPLEXES USED TO TEST THE FREE ENERGY FUNCTION

| Protein–ligand complex | PDB entry | –log $K_i$ pred. | –log $K_i$ expt. | Ref. |
|---|---|---|---|---|
| Trypsin–NAPAP | 1PPC | 7.46 | 6.46 | 48 |
| Trypsin–3-TAPAP | 1PPH | 6.89 | 6.22 | 48 |
| Cytochrome P450–4-Phe-imidazole | 1PHF | 4.23 | 4.40 | 49 |
| Cytochrome P450–metyrapone | 1PHG | 6.20 | 8.66 | 49 |
| Cytochrome P450–camphore | 2CPP | 4.14 | 6.07 | 50 |
| Cytochrome P450–adamantone | 5CPP | 3.25 | 5.88 | 50 |
| Carboxypeptidase–ZFV$^P$(O)F | 7CPA | 11.4 | 14.0 | 51 |
| Thermolysin–PLN | 2TMN | 5.22 | 4.67 | 22 |
| Thermolysin–HONH-BAGN | 5TLN | 6.20 | 6.37 | 22 |

The predicted –log $K_i$ values refer to results obtained with function #2.

TABLE 4
COMPARISON OF PREDICTED BINDING CONSTANTS OF 5-(SUBSTITUTED BENZYL)-2,4-DIAMINOPYRIMIDINES AS INHIBITORS OF DHFR FROM *L. casei* WITH EXPERIMENTAL DATA FROM SELASSIE ET AL. [52]

| Substituent | $-\log K_i$ pred. | $-\log K_i$ expt. |
|---|---|---|
| H | 6.23 | 5.20 |
| 3,4,5-$(OCH_3)_3$ | 7.74 | 6.88 |
| 3,5-$(OCH_3)_2$, 4-$C(CH_3)$=$CH_2$ | 8.16 | 7.34 |
| 4-$O(CH_2)_6CH_3$ | 5.36 | 5.38 |
| 3,4,5-$(CH_2CH_3)_3$ | 7.10 | 6.88 |

for our new scoring function. Selassie et al. [52] recently reported the activities of 5-(substituted benzyl)-2,4-diaminopyrimidines as inhibitors of DHFR. X-ray structures of the protein–ligand complexes are not available. We have therefore docked a number of the structures reported by Selassie et al. [52] into the active site of DHFR (PDB reference code 3DFR), using computer graphics. We selected five compounds from the work of Selassie et al., including the compounds showing the strongest and the weakest binding, trimethoprim and a compound with a very flexible side chain. In the initial placement of the ligands in the binding site, it was assumed that the binding mode and ligand conformation are similar to those of trimethoprim [53]. The 2,4-diaminopyrimidine moiety was taken from the inhibitor present in the X-ray structure 3DFR, without modification. The position of the substituted benzyl side chain was then optimized using the CVFF force field [54]. The protein was kept rigid during the geometry optimization of the ligand. In Table 4, the results obtained from function #2 are compared with the experimental binding data. Very good agreement is obtained for all five compounds. The standard deviation is 0.71 ($\log K_i$), corresponding to an error in binding energies of 4.0 kJ/mol. The very good performance of the scoring function is not entirely unexpected, because the closely related complex DHFR–2,4-diaminopteridine was included in the calibration dataset with a very small difference between the calculated and experimental binding constants ($\log K_i = -5.5$ and $-6.0$, respectively).

## DISCUSSION AND CONCLUSIONS

We have presented a set of simple empirical functions that estimate the free energy of binding for a given protein–ligand complex with known 3D structure. The functions were calibrated using a set of 45 protein–ligand complexes with known binding constants. The best representation of the binding data is obtained with function #2, which reproduces the binding energies of the 45 protein–ligand complexes with a standard deviation of 7.9 kJ/mol. The function was further applied to nine protein–ligand complexes not included in the calibration dataset. Good agreement is obtained between the predicted and the experimental binding constants. It is further demonstrated that function #2 correctly predicts the binding affinities of a series of related inhibitors of DHFR. In view of the simple form of the scoring function, this good performance is quite surprising. The function can be evaluated very fast. In a test calculation on the specificity pocket of chymotrypsin, the function was calculated 168 times which took 20 s on a Silicon Graphics Indigo R4000 workstation (0.12 s per evaluation of the energy function). Therefore, the function is well suited as scoring function in a 3D database search or de novo ligand design program.

The values for the neutral hydrogen bond $\Delta G_{hb}$ and the ionic interaction $\Delta G_{ionic}$ are close to those reported by Fersht [55] and Shirley et al. [56]. The contribution due to lipophilic contacts $\Delta G_{lipo}$ is predicted to be –0.17 kJ/mol Å$^2$. This is larger than the value of –0.10 kJ/mol Å$^2$ estimated by Richards [57] and closer to a recent estimate of –0.20 kJ/mol Å$^2$ given by Sharp et al. [58]. The contribution from rotatable bonds $\Delta G_{rot}$ (1.4 kJ/mol) is slightly smaller than previous estimates of 1.6–3.6 kJ/mol given by Williams et al. [11]. This low value is probably due to an averaging over protein–ligand complexes with a preorganized ligand and complexes with a ligand that has to change its conformation upon binding to the protein. In principle, the number of rotatable bonds NROT should be the number of rotatable bonds that are freely rotatable in the free ligand and fixed in the bound one. However, it has been shown for a number of examples that molecules with a large number of rotatable bonds may be fairly rigid in aqueous solution [59,60].

Function #2 contains a constant contribution $\Delta G_0$ amounting to +5.4 kJ/mol. The rationale to include this term was to account for the loss of translational and rotational entropy upon ligand binding. However, some caution seems to be advisable in the physical interpretation of $\Delta G_0$. First, we note that the present value is much smaller than an early estimate of 58 kJ/mol by Andrews [61], based on work by Page [62]. On the other hand, it was shown recently by Williams et al. [10,11] that the loss of translational and rotational entropy upon binding is significantly smaller. Williams estimates this term to lie 'anywhere in the range 9 to 45 kJ/mol' [11]. In comparing the present result with earlier estimates, it should also be noted that the 95% confidence interval for $\Delta G_0$ is ±8 kJ/mol, which is much larger than for all other parameters. Therefore, $\Delta G_0$ is much less well defined than the other parameters. In fact, if $\Delta G_0$ is set to zero (function #5), the fit is only marginally degraded. However, if on the other hand we fix $\Delta G_0$ to 41.8 kJ/mol, which is closer to what is generally viewed as an acceptable value, the fit to the experimental data is poor, with a standard deviation of 14 kJ/mol. For example, function #4 predicts a positive $\Delta G$ for the binding of benzene to chymotrypsin.

The comparison of the energy functions #2 and #3 seems to indicate that it is not necessary to account for hydrogen-bond distortions. The good performance of function #3 in comparison with function #2 indicates that small distortions of hydrogen-bond geometries in protein–ligand complexes as found in X-ray structures do not correspond to reduced binding affinities and are more likely due to small experimental uncertainties. However, one should keep in mind that the dataset of structures used in the calibration contains ligands that all show a good complementarity with the protein structure. A scoring function for a 3D database search must be able to differentiate between structures forming good hydrogen bonds and those that can only form perturbed ones. Therefore, we have decided to use function #2 in the de novo ligand design program LUDI [1,2].

The following example serves to illustrate the advantage of using tolerances in the scoring function. We have compared two different trypsin structures, obtained as a complex with the inhibitors benzamidine (PDB reference code 3PTB) and bovine pancreatic trypsin inhibitor, BPTI (PDB reference code 2PTC). The peptidic inhibitor places a lysine side chain into the specificity pocket of the enzyme. This acyclic side chain is slimmer than benzamidine. As a consequence, trypsin slightly adjusts its structure to accommodate the different ligands. If one tries to dock benzamidine into the specificity pocket of trypsin taken from the complex with BPTI and keeps the protein fixed, one will inevitably get worse hydrogen-bond geometries. For example, using the CVFF force field [54], an interaction energy of –280 kJ/mol is obtained with trypsin, using the coordinates from 3PTB, and −185 kJ/mol using the coordinates from 2PTC. Using the present scoring function, we obtain very similar $K_i$ values of 3.3 μM (3PTB) and 3.0 μM (2PTC) (experimental value 18 μM [16]). This indicates that in this case the built-in tolerance in the present scoring function is able to account for the distorted hydrogen bonds, which are due only to the use of an inappropriate protein structure. However, it should be noted that the scoring function can only tolerate small conformational differences. If the conformational change of the protein due to an induced fit is significantly larger than in the case of trypsin, the scoring function will also run into problems.

The largest deviations between calculated and measured binding affinities are found for the complexes streptavidin–biotin (affinity underestimated by 15 kJ/mol) and HIV protease–MVT101 (affinity overestimated by 16.5 kJ/mol). As discussed by Weber et al. [33], the binding of biotin to streptavidin is enthalpically driven ($\Delta G = -77$ kJ/mol, $\Delta H = -134$ kJ/mol). The ureido oxygen is tetrahedrally coordinated and can form more and stronger hydrogen bonds than in water. Therefore, the streptavidin–biotin structure should be viewed as a special case. The overestimation of the binding affinity of MVT101 to HIV protease is caused by two problems. First, MVT101 contains 29 rotatable bonds. The contribution from one rotatable bond in function #2 is 1.4 kJ/mol. As discussed above, this value is probably slightly too small. In addition, a close inspection of the X-ray structure of the HIV protease–MVT101 complex reveals a number of unfavorable dihedrals in some of the side chains of MVT101. At present, the conformational energy of the ligand is not taken into account in the energy function.

The dataset contains small ligands, like benzene, that were positioned in the protein binding site using LUDI. One might suspect that benzene shows multiple binding modes. However, in the present scoring function the only contribution to the binding of benzene originates from lipophilic contacts. The measured $K_i$ for the benzene–chymotrypsin complex, 25 mM (corresponding to $\Delta G = -9.1$ kJ/mol) is remarkably close to the measured binding affinities of benzene with a macrocyclic host [63], cyclodextrins [64] or a lysozyme mutant with an artificial internal lipophilic pocket [65]. In all cases the benzene molecule is buried in a lipophilic pocket. Therefore, even if a second binding site for benzene would exist, it would yield roughly the same contribution to the binding affinity.

There are several limitations to the applicability of function #2. The function does not account for differences in binding strengths between various neutral hydrogen bonds or ionic interactions. For example, it is well known that solvent-accessible salt bridges contribute very little to protein stability, whereas buried ionic interactions yield a large contribution [66]. A similar behaviour can be expected for protein–ligand interactions. Also, it has been demonstrated that the removal of a hydrogen bond may leave the free energy of binding unaffected [67]. This behaviour is not reproduced with the current scoring function. Recently, it was demonstrated that interactions between a quaternary ammonium group and aromatic rings may contribute significantly to the binding affinity (cation–π interaction) [68]. This effect is ignored in our scoring function. As already discussed above, another important limitation is the neglect of the contribution from the internal conformational energy of both the ligand and the protein. The intramolecular strain in a particular ligand conformation may compensate in part for the favorable protein–ligand interaction. However, this conformational energy is difficult to calculate on the fly and was therefore not included in the present energy function. The neglect of these contributions may lead to some overestimations of the binding strength by the scoring function. Some flexible ligands may bind worse than predicted by our scoring function. However, this problem can easily be solved by performing a conformational analysis on the top scoring ligands after the design process. Another problem with the present scoring function occurs with very short hydrogen bonds, which are observed in some protein–ligand complexes. For example, in the complexes thermolysin–ZF$^{P}$LA and thermolysin–ZG$^{P}$LL, a very short hydrogen bond is observed between the protonated side chain of Glu$^{143}$ and the P=O group of the inhibitor, with hydrogen-bond lengths $R_{O\cdot\cdot O}$ of 2.3 Å. A similar hydrogen bond is found in the complex carboxypeptidase A–ZAA$^{P}$(O)F. Our scoring function does not recognize such hydrogen bonds as very strong and is therefore expected to slightly underestimate the binding affinity of the ligands.

It should also be noted that the present scoring function accounts for perturbed hydrogen bonds or ionic interactions, but it does not penalize repulsive interactions between the protein and the ligand. Therefore, the successful application of the scoring function requires a prescreen to detect those ligands that will form some sort of repulsive interaction with the protein, either due to steric problems or to electrostatic repulsion between polar groups. In LUDI [1,2], this check is carried out in advance for every putative ligand structure; only those structures that pass this test are then subjected to the scoring function.

In the derivation of the scoring function, water molecules were not taken into account. It is well known that water molecules can play an important role in mediating protein–ligand interactions. For example, dihydrofolate reductase [30] and HIV protease [27–29] both contain crucial water molecules that form hydrogen bonds both with the protein and the inhibitor. Therefore, one might anticipate that water-mediated hydrogen bonds contribute significantly to the binding. However, our attempts to incorporate crystallographically observed water molecules did not yield an improved model. One possible reason for this failure is that only crystallographically observed waters were taken into account. However, it is quite likely that not all important water molecules were determined in the X-ray diffraction experiment. Therefore, if we just use the waters present in the X-ray structure we end up with an inconsistent picture. It appears necessary to generate the positions and orientations of all water molecules in the vicinity of the ligand and then use them in the derivation of the scoring function. However, this procedure would then also be required for any putative ligand. At present, we consider this too time-consuming and therefore refrain from accounting for water-mediated hydrogen bonds.

Finally, it is clear that the accuracy of any computational approach to predict $\Delta G$ depends on the accuracy of the experimentally determined binding energies. For a number of cases contained in the present dataset, several measurements of binding constants were published that show a spread of the experimental values of up to a factor of five [25,27,69]. This uncertainty of the experimental binding data poses a limit for the accuracy of any theoretical description of the binding data.

The de novo design of protein ligands is a new area of research. A number of different methods have been recently proposed [1,2,4,6,70]. The scoring of putative ligands is a vital component of these approaches. As these programs retrieve or construct on the order of 1000 or more putative ligand structures, it is absolutely necessary to have a fast scoring algorithm. The present scoring function was developed with this application in mind. It contains only five adjustable parameters. In view of the limited size and accuracy of the dataset, we refrained from using a more sophisticated function. However, the inclusion of a larger number of protein–ligand complexes may justify the use of a larger number of adjustable parameters.

The standard deviation of function #2 corresponds to an error in $K_i$ of about a factor of 25. We do not yet consider this sufficient for a reliable quantitative prediction of binding constants. Nevertheless, it is hoped that it will be useful in prioritizing the hits from a 3D search or from a de novo ligand design program such as LUDI [71].

## ACKNOWLEDGEMENTS

I would like to thank Gerhard Klebe for many helpful discussions and Hugo Kubinyi for his help with the statistical analysis of the data.

## REFERENCES

1 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
2 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 593.
3 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
4 Rotstein, S.H. and Murcko, M.A., J. Med. Chem., 36 (1993) 1700.
5 Tomioka, N., Itai, A. and Iitaka, Y., J. Comput.-Aided Mol. Design, 1 (1987) 197.
6 Moon, J.B. and Howe, W.J., Proteins, 11 (1991) 314.
7 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
8 Boobyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
9 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.
10 Williams, D.H., Searle, M.S., Mackay, J.P., Gerhard, U. and Maplestone, R.A., Proc. Natl. Acad. Sci. USA, 90 (1993) 1172.
11 Searle, M.S. and Williams, D.H., J. Am. Chem. Soc., 114 (1992) 10690.
12 Horton, N. and Lewis, M., Protein Sci., 1 (1992) 169.
13 Bohacek, R.S. and McMartin, C., J. Med. Chem., 35 (1992) 1671.
14 Connolly, M.L., Science, 221 (1983) 458.
15 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, T., J. Mol. Biol., 112 (1977) 535.
16 Mares-Guia, M. and Shaw, E., J. Biol. Chem., 240 (1965) 1579.
17 Bode, W., Turk, D. and Stürzebecher, J., Eur. J. Biochem., 193 (1990) 175.
18 Kikumoto, R., Tamao, Y., Tezuka, T., Tonomura, S., Hara, H., Ninomiya, K., Hijikata, A. and Okamoto, S., Biochemistry, 23 (1984) 85.
19 Stürzebecher, J., Walsmann, P., Voigt, B. and Wagner, G., Thromb. Res., 36 (1984) 457.
20 Gubernator, K., private communication, 1993.
21 Wallace, R.A., Kurtz, A.N. and Niemann, C., Biochemistry, 2 (1963) 824.
22 Matthews, B.W., Acc. Chem. Res., 21 (1988) 333.
23 Rahuel, J., Priestle, J.P. and Grütter, M.G., J. Struct. Biol., 107 (1991) 227.
24 Cooper, J., Foundling, S., Hemmings, A. and Blundell, T., Eur. J. Biochem., 169 (1987) 215.
25 Zollner, H., Handbook of Enzyme Inhibitors, VCH Publishers, Weinheim, 1993.
26 Blundell, T.L., Cooper, J., Foundling, S.I., Jones, D.M., Atrash, B. and Szelke, M., Biochemistry, 26 (1987) 5585.
27 Erickson, J., Neidhart, D.J., VanDrie, 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.
28 Bone, R., Vacca, J.P., Anderson, P.S. and Holloway, M.K., J. Am. Chem. Soc., 113 (1991) 9382.
29 Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, S., Science, 246 (1989) 1149.
30 Bolin, J.T., Filman, D.A., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
31 Blaney, J.M., Hansch, C., Silipo, C. and Villon, A., Chem. Rev., 84 (1984) 333.
32 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.

33 Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. and Salemme, F.R., J. Am. Chem. Soc., 114 (1992) 3197.

34 Cowan, S.W., Newcomer, M.E. and Jones, T.A., Proteins, 8 (1990) 44.

35 Lowe, J.B., Sacchettini, J.C., Laposata, M., McQuillan, J.J. and Gordon, J.I., J. Biol. Chem., 262 (1987) 5931.

36 Miller, D.M., Olson, J.S., Pflugrath, J.W. and Quiocho, F.A., J. Biol. Chem., 258 (1983) 13665.

37 Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., Science, 252 (1991) 839.

38 Badger, J., Minor, I., Kremer, M.J., Oliveira, M.O., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishnaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.D., Dutko, F.J., Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. USA, 85 (1988) 3304.

39 Entsch, B., Ballou, D.P. and Massey, V., J. Biol. Chem., 251 (1976) 2550.

40 Dani, M., Manca, F. and Rialdi, G., Biochim. Biophys. Acta, 667 (1981) 108.

41 Bolognesi, M., Cannilo, E., Ascenzi, P., Giacometti, G.M., Merli, A. and Brunori, M., J. Mol. Biol., 158 (1982) 305.

42 Sauter, N.K., Bednarski, M.D., Wurzburg, B.A., Hanson, J.E., Whitesides, G.M., Skehel, J.J. and Wiley, D.C., Biochemistry, 28 (1989) 8388.

43 Bunting, J.W. and Myer, C.D., Can. J. Chem., 53 (1975) 1993.

44 Kim, H. and Lipscomb, W.N., Biochemistry, 29 (1990) 5546.

45 Brandstetter, H., Turk, D., Hoeffken, H.W., Grosse, D., Stürzebecher, J., Martin, P.D., Edwards, B.F.P. and Bode, W., J. Mol. Biol., 226 (1992) 1085.

46 Roderick, S.L., Fournie-Zuliski, M.C., Roques, B.P. and Matthews, B.W., Biochemistry, 28 (1989) 1493.

47 Program INSIGHT, Biosym Technologies, Inc., San Diego, CA, 1993.

48 Turk, D., Stürzebecher, J. and Bode, W., FEBS Lett., 287 (1991) 133.

49 Lipscomb, J.D., Biochemistry, 19 (1980) 3590.

50 Fisher, M.T. and Sligar, S.G., J. Am. Chem. Soc., 107 (1985) 5018.

51 Kim, H. and Lipscomb, W.N., Biochemistry, 30 (1991) 8171.

52 Selassie, C.D., Fang, Z.X., Li, R.L., Hansch, C., Debnath, G., Klein, T.E., Langridge, R. and Kaufman, B.T., J. Med. Chem., 32 (1989) 1895.

53 Roth, B. and Stammers, D.K., In Bedell, C.R. (Ed.) The Design of Drugs to Macromolecular Targets, Wiley, New York, NY, 1992, pp. 85–118.

54 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Proteins, 4 (1988) 31.

55 Fersht, A.R., Shi, J.P., 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.

56 Shirley, B.A., Stanssens, P., Hahn, U. and Pace, C.N., Biochemistry, 31 (1992) 725.

57 Richards, F.M., Annu. Rev. Biophys. Bioeng., 6 (1977) 151.

58 Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B., Biochemistry, 30 (1991) 9686.

59 Hoffmann, R.W., Angew. Chem., 104 (1992) 1147.

60 Lim, M.S.L., Johnston, E.R. and Kettner, C.A., J. Med. Chem., 36 (1993) 1831.

61 Andrews, P.R., Craik, D.J. and Martin, J.L., J. Med. Chem., 27 (1984) 1648.

62 Page, M.I., Angew. Chem., Int. Ed. Engl., 16 (1977) 49.

63 Jorgensen, W.L., Nguyen, T.B., Sanford, E.M., Chao, I., Houk, K.N. and Diederich, F., J. Am. Chem. Soc., 114 (1992) 4003.

64 Inoue, Y., Hakshi, T., Liu, Y., Tong, L.H., Shen, B.J. and Jin, D.S., J. Am. Chem. Soc., 115 (1993) 475.

65 Eriksson, A.E., Baase, W.A., Wozniak, J.A. and Matthews, B.W., Nature, 355 (1992) 371.

66 Dao-Pin, S., Nicholson, H., Baase, W.A., Zhang, X.-J., Wozniak, J.A. and Matthews, B.W., In Chadwich, D.J. (Ed.) Protein Conformation, Ciba Foundation Symposium, Vol. 161, Wiley, Chichester, 1991, pp. 52–62.

67 Morgan, B.P., Scholtz, J.M., Ballinger, M.D., Zipkin, I.D. and Bartlett, P.A., J. Am. Chem. Soc., 113 (1991) 297.

68 Dougherty, D.A. and Stauffer, D.A., Science, 250 (1990) 1558.

69 Baker, B.R. and Erickson, E.H., J. Med. Chem., 10 (1967) 1123.

70 Nishibata, Y. and Itai, A., Tetrahedron, 47 (1991) 8985.

71 LUDI is available from Biosym Technologies, Inc., San Diego, CA.