Journal of Computer-Aided Molecular Design (2005) 19: 421–443
DOI 10.1007/s10822-005-7969-7

Design of ligand binding to an engineered protein cavity using virtual screening and thermal up-shift evaluation

Claudia Machicado[a,c], Jon López-Llano[a,c], Santiago Cuesta-López[b,c], Marta Bueno[a,c] & Javier Sancho[a,c,*]
[a]*Departamento de Bioquímica y Biología Molecular y Celular;* [b]*Departamento de Física de la Materia Condensada Facultad de Ciencias. Universidad de Zaragoza, 50009, Zaragoza, Spain;* [c]*Biocomputation and Complex Systems Physics Institute-BIFI Universidad de Zaragoza, 50009, Zaragoza, Spain*

Received 25 February 2005; accepted 25 May 2005

Key words: flavodoxin, flexible docking, ligand binding, protein cavity, protein stability, reverse docking, virtual screening

Summary

Proteins could be used to carry and deliver small compounds. As a tool for designing ligand binding sites in protein cores, a three-step virtual screening method is presented that has been optimised using existing data on T4 lysozyme complexes and tested in a newly engineered cavity in flavodoxin. The method can pinpoint, in large databases, ligands of specific protein cavities. In the first step, physico-chemical filters are used to screen the library and discard a majority of compounds. In the second step, a flexible, fast docking procedure is used to score and select a smaller number of compounds as potential binders. In the third step, a finer method is used to dock promising molecules of the hit list into the protein cavity, and an optimised free energy function allows discarding the few false positives by calculating the affinity of the modelled complexes. To demonstrate the portability of the method, several cavities have been designed and engineered in the flavodoxin from *Anabaena* PCC 7119, and the W66F/L44A double mutant has been selected as a suitable host protein. The NCI database has then been screened for potential binders, and the binding to the engineered cavity of five promising compounds and three tentative non-binders has been experimentally tested by thermal up-shift assays and spectroscopic titrations. The five tentative binders (some apolar and some polar), unlike the three tentative non-binders, are shown to bind to the host mutant and, importantly, not to bind to the wild type protein. The three-step virtual screening method developed can thus be used to identify ligands of buried protein cavities. We anticipate that the method could also be used, in a reverse manner, to identify natural or engineerable protein cavities for the hosting of ligands of interest.

Introduction

The theoretical prediction of the association of ligands with protein receptors may seem deceitfully simple because it only requires an efficient sampling of the ligand and receptor conformational space and a sufficiently accurate free energy function [1–4]. In practical terms the task is difficult, one contributing reason being the simplicity of some of the functions used. On the other hand, the proliferation of new compounds from combinatorial chemistry and the availability of small-molecule libraries for virtual or biological screening have fostered the generation of lead compounds [5–8], and rational drug design strategies have become common [9–12]. For proteins of known structure, computational screening of three-dimensional molecular databases can, in principle, identify binder molecules that, sometimes, are readily

*To whom correspondence should be addressed. e-mail: jsancho@unizar.es

available [13–19]. It is clear that while massive virtual screening is very useful, fine docking strategies are needed to refine the output of massive screening and to guide subsequent experiments.

One major complication in protein/ligand docking arises from protein flexibility, which is typically associated to protein surfaces. In contrast, protein cores are much more rigid and, interestingly, they sometimes contain cavities (natural or engineered [20, 21]) that bind small molecules [22, 23]. Crystallographic analysis of the protein response to cavity creating mutations [20] has revealed that, in many cases, the residues surrounding the cavity only experience slight readjustments. In addition, it has been recently shown that the structure of cavity-bearing mutant proteins can be precisely modelled from the corresponding wild type structures using simple procedures [24]. It seems therefore that protein cavities are suitable scenarios for designing ligand binding sites with specific characteristics not shared by solvent exposed binding sites.

To provide tools for designing internal protein-binding sites, a three-step virtual screening method is presented here that can reliably identify ligands of specific protein cavities by scanning small-compound three-dimensional databases. The method applies several physicochemical filters and docking algorithms in a sequential way. To address the problem of false positives in the final screening step, a flexible docking protocol has been developed, based on existing Monte Carlo grid docking methods, that accurately predicts the X-ray structure of several known protein–ligand complexes and can discriminate most non-binders (decoys) by expelling them from the binding site. The remaining false positives are identified using an optimised free energy function that explicitly considers all the relevant enthalpic and entropic contributions to the binding-free energy. A full discrimination of non-binders is achieved in this way. The three-step virtual screening method successfully identifies, among the National Cancer Institute (NCI) database compounds, the known binders of a T4 lysozyme cavity mutant [25].

To test the portability of the method to other proteins, five cavity-containing mutants of the flavodoxin from *Anabaena* PCC 7119 have been designed, engineered and characterised. Based on its structural integrity and appropriate stability, the W66F/L44A mutant has been selected to act as protein receptor, and potential binders of the cavity present in this flavodoxin variant have been identified by screening the NCI database. The actual binding of the selected candidates has been experimentally tested by thermal up-shift experiments that show that the compounds predicted to bind to the cavity increase the temperature of mid-denaturation of the host protein, unlike the control compounds predicted not to bind. The binding of the selected candidates to W66F/L44A flavodoxin has been further confirmed by performing spectroscopic titrations of protein intrinsic fluorescence. The three-step screening method developed can thus effectively identify binders to specific protein cavities within large libraries of compounds. We discuss how the method could be used in a reverse manner to identify host cavities suitable to bind ligands of choice.

Methods

3D structures and T4 lysozyme 17 binders and 23 non-binders training datasets

High-resolution crystal structures of the L99A mutant T4 lysozyme and of its complexes with nine ligands [20, 23] have been used as modelling targets. The Protein Data Bank (PDB) codes of the T4 lysozyme proteins used are: 1L90 [L99A mutant], 181L [L99A complex with benzene], 182L [with benzofuran], 183L [with indene], 184L [with *iso*-butylbenzene], 185L [with indole], 186L [with *n*-butylbenzene], 187L [with *p*-xylene], 188L [with *o*-xylene], and 1NHB [with ethylbenzene]. The structures of these nine ligands were obtained from the Okanagan Database (http://people.ouc.bc.ca/woodcock/molecule/molecule.html) or built using *Biopolymer* (*Accelrys Inc.*). In either case, hydrogen atoms were added to the ligands and the structures were energy minimised, using the consistent valence force field (CVFF) [26], inside a solvent sphere of a 10 Å radius with an energy convergence criterion of 0.001 kcal mol^{-1}. In addition, the structures of eight molecules that have been experimentally shown to bind to the L99A cavity (*m*-xylene, propylbenzene, *o*-ethyltoluene, *m*-ethyltoluene, *p*-ethyltoluene, toluene, styrene, and naphthalene) and of 23 molecules known not to bind it (phenol, *p*-cresol, benzyl alcohol, methanol, ethanol, propanol, heptanol, octanol,

quinoline, aniline, benzaldehyde, benzoic acid, azulene, mesitylene, *tert*-butylbenzene, camphene, camphor, cyclohexane, pyridine, 1-phenylheptane, *N*,*N*′-dimethylaniline, furan, and *trans*-cinnamaldehyde) [25] were similarly obtained and subjected to energy minimisation.

To develop the screening method, these 40 training compounds have been added to the NCI database. This library hosts 250251 compounds. The two most stable conformers of each of them have been generated using *Catalyst 4.7* (*Accelrys Inc.*).

Development of a three-step, massive, virtual screening method and its bench-marking using L99A lysozyme-ligand complexes. First and second steps: initial filtering and fast massive docking

The screening method developed, links existing software packages (*Cerius*^{2} *4.7* and *InsightII* from *Accelrys Inc.*, and *Vega* [27]) and involves three steps (Figure 1). In the first one, a pre-selection of compounds, based on their physico-chemical properties, is performed through a Linux Shell Script that combines several AWK programs and *Vega*. *Vega* calculates the volume and polarity of each compound, which are then used to execute the specified discrimination. The volume and polarity filters were established according to the values computed for the L99A mutant lysozyme ligands used in the training (volumes < 160 Å^3, Crippen lipoles [28] < 1.7, and polar accessible surface areas (ASA_p) < 90 Å^2). Using this triple filter, a 2% of the compounds in the NCI database were selected as potential ligands of the L99A lysozyme cavity. Since the volume and hydrophobicity of this cavity, and those of the W66F/L44A mutant flavodoxin cavity that will be used to experimentally test the portability of the method to other proteins, are similar (volumes of 170 and 230 Å^3, and hydrophobicity indexes of 44.3 and 43.3, respectively) the NCI compounds selected in

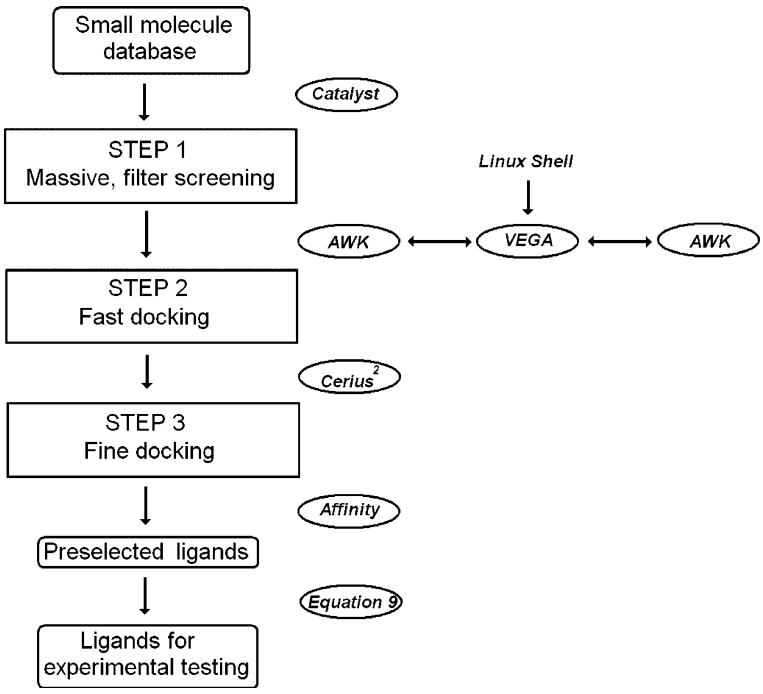

Figure 1. Three-step screening method devised for the discovery of potential ligands in large small-molecule libraries. The different programs used in the screening are indicated. Initially, two conformers of each compound stored in the database used (NCI) where generated using *Catalyst*. In the first step of the screening a pre-selection of compounds is performed based on physical-chemical properties using a Linux Shell linking *Vega* and various AWK programs. In the second step, fast docking of the selected compounds into a given protein cavity is carried out using *Cerius*, and the complexes are ranked according to their calculated energy using and a scoring function (Equation 1, see *Methods*). In the third step, a fine docking is performed, using *Affinity*, between highly ranked compounds in the hit list and a given cavity protein. Remaining false positives are identified and discarded using a comprehensive, optimise binding-free-energy function (Equation 9, see *Methods*).

this first step as potential binders of L99A lysozyme were also considered to be potential binders of W66F/L44A flavodoxin.

In the second step, the compounds preselected were docked into the L99A lysozyme cavity with *LigandFit* (implemented in *Cerius*^{2} *4.7*). This docking procedure uses a Monte Carlo conformational search of the ligand. For each conformation, fittings are carried out in ligand orientations selected using the non-mass-weighted principle moment of inertia [29], and the corresponding docking scores are computed. The docking was optimised using a training set of 6 L99A T4 lysozyme binders (ethylbenzene, *o*-ethyltoluene, benzofuran, indole, toluene, and *n*-butylbenzene) and six non-binders (benzylalcohol, *t*-butylbenzene, heptanol, camphene, mesytilene, azulene), and the L99A mutant structure modelled with a method previously described [24]. Atom charges were assigned with the Gasteiger–Marsili charging algorithm [30]. Three different force fields: *Dreiding* [31], *CFF91* [32] and CVFF95 [33] were tested and *Dreiding* was selected. Five functions were tested for scoring: *LigScore1* (including a van der Waals, *vdW* term with a soft Lennard–Jones 6–9 potential, a $C+Pol$ term related to attractive protein–ligand interactions, and a *Totpol*^{2} term that computes the attractive and repulsive protein–ligand interaction) [34], *LigScore2* (with the same vdW and $C+Pol$ terms as *LigScore1* plus a *BuryPol*^{2} term that approximates protein and ligand desolvation, see www.accelrys.com/products/datasheets/ds_ligandscore_data.pdf), *Piecewise Linear Potential* (*PLP,* [35]), *Jain* [36], and *Potential of Mean Force (PMF*, [37]).

An automatic evaluation of the complexes affinity was initially performed using some of the above terms with a *Consensus Scoring* function implemented in Cerius^2 [38] but, when this *Consensus Scoring* was applied to evaluate the NCI compounds selected in the first step of the screening, it failed to place the known binders among the best ranked compounds. We therefore tried various combinations of the different scoring terms available and defined the following *Scoring* function:

$$\begin{aligned}\text{Scoring} &= \text{vdW} + \text{LigScore2} + 0.3^{*} \\ &\text{BuryPol} + 1.5^{*}C + Pol + 2^{*} \\ &(\text{PLP1} + \text{PLP2} + \text{Jain} + \text{PMF})\end{aligned} \quad (1)$$

This *Scoring* function gave favourable scores to the 17 binders and unfavourable scores to most non-binders. More importantly, when applied to the NCI compounds that had been selected in the first screening step, it gave rise to a large enrichment in known binders within the best ranked compounds (see Results and discussion).

Third step: fine docking procedure

To further reduce the number of potential binders obtained in the second step of the screening, a third step was introduced and optimised that combines a finer docking procedure with the use of a comprehensive, purposely developed binding-free energy function to compute the affinity of the docked complexes.

The finer flexible docking of small molecules into the L99A T4 lysozyme cavity mutant was performed using *Affinity* (implemented in *Insight II*). Nine ligands whose complexes with the L99A mutant protein are available (see above) and six non-binders: *p*-cresol, *trans*-cinnamaldehyde, octanol, azulene, quinoline, and benzoic acid were used to implement the docking and to test its performance. Partial charges were defined in protein and ligands using the CVFF force field. As only small accommodation changes of the residues near the protein cavity are expected [20, 24], the ''bulk'' of the receptor was taken as rigid while the residues with atoms at the cavity surface and the ligand were considered flexible, which significantly saves computation time. Non-bonded interactions were calculated using a grid-based approach developed by Luty et al. [39]. A solvation grid was incorporated, according to the method described by Stouten et al. [40], to describe the contribution of the bulk region, while a Monte Carlo search strategy [39, 41] was used in conjunction with the CVFF force field to treat binding site and ligand. The initial position of a ligand in the cavity was determined by performing an interactive ligand–protein docking to alleviate unfavourable steric contacts. Docking grids with a 0.5 Å spacing were generated using the grid utility within *Affinity*. Different sets of parameters were evaluated by their ability to predict the structure of the complexes with low Root mean square deviation (RMSD, calculated using [42]).

The selected parameters are listed in Table 1. In the final protocol, at each Monte Carlo step, the conformation is energy minimised and the new conformation is accepted or rejected according to a Metropolis criterion with algorithm temperature of 600 K, which helps to explore the conformational space.

Third step: discriminating, comprehensive binding-free energy function

The affinity of the complexes modelled in the third step was computed using a purposely developed an optimised binding-free energy function that is specific for the binding of ligands in internal protein cavities and considers all the relevant enthalpic and entropic terms:

$$\Delta G_b = \Delta H_P + \Delta H_L + \Delta H_{PL} + \Delta H_{Lw} - T\Delta S_P - T\Delta S_L - T\Delta S_{PL} - T\Delta S_w \quad (2)$$

In this function, ΔH_P is the protein binding site enthalpy change, ΔH_L is the ligand enthalpy change, ΔH_{PL} is the protein–ligand interaction enthalpy change from the productive conformation, ΔH_{Lw} is the ligand–water interaction enthalpy change, ΔS_P is the protein binding site conformational entropy change, ΔS_L is the ligand conformational entropy change, ΔS_{PL} is the rotational/translational entropy change of protein and ligand upon complex formation, and ΔS_w is the water entropy change, which is associated to the hydrophobic effect that contributes to the binding-free energy. Since ΔH_{Lw} represents the desolvation enthalpy of the ligand, it will be termed ΔH_{desolv} hereinafter. If a term accounting for the change in protein/water solvation enthalpy were added to Equation 2, the function would describe the binding at a protein surface. Equation 2 can be reorganised by clustering the enthalpic terms and all the protein and ligand entropic terms as:

$$\Delta G_b = \Delta H - T\Delta S_{P\&L} - T\Delta S_w \quad (3)$$

Alternatively, it can be reorganised by clustering the enthalpic and entropic terms of each of the

Table 1. Parameters of the docking simulations.

Second-step, fast docking		Third-step, fine docking	
Condition	Value/method	Condition	Value/method
Docking	Flexible	Binding site	Surface cavity residues
Force field	Dreiding	Ligand initial structure	Minimised in sphere of solvent
Energy grid	Force field method	Flexible molecules	Ligand and binding site
Monte Carlo trials	2000	Force field	CVFF
Conformations saved	10 per ligand	Solvation grid	Included
Soft potential energy	Trilinear interpolation value	Grid point-spacing	0.5 Å
Ligand internal energy	vdW and electrostatics	Dielectric constant	1
Non-bond cutoff	11 Å	Dielectric method	Constant
Dielectric constant	1	Calculated structures	30
Conformation arrangement	Clustering	Minimisation method	Conjugate gradients
Clustering method	Complete linkage	Number of minimisation steps	300
Saved conformation	Two best conformers per cluster	Summation method for non-bonded interactions	Atom-based
Energy minimisation of ligand	100 steps	Energy test	Metropolis
Consensus Scoring function	vdW + CPol + LigScore2 + $BPol^2$ + Jain	Temperature	600 K
Scoring function (Equation 1)	vdW + LigScore2 + 0.3$BPol^2$ + 1.5Cpol + 2(PLP1 + PLP2 + Jain + PMF)	Maximum angle change of ligand	360°
		Maximum distance change of ligand	100 Å
		Energy tolerance	10^{-6} kcal mol^{-1}

species (protein, ligand, and complex) into free energy change terms (with $\Delta G_{desolv} = \Delta H_{desolv} - T\Delta S_w$):

$$\Delta G_b = \Delta G_P + \Delta G_L + \Delta G_{PL} + \Delta G_{desolv} \quad (4)$$

The individual enthalpic terms, H_L, H_{LP} required for calculating ΔH_L and ΔH_{PL} in Equation 2 have been approximated by energy calculations performed with the CVFF force field (using an 11 Å cut-off for non-bonded interactions). However, for the binding site enthalpies we compared the performance of two force fields, CVFF [26] and AMBER [43]. CVFF gave unrealistically large values, so we decided to use AMBER. Enthalpy differences (ΔH_P and ΔH_L) were calculated as the energies of the protein-binding site or of the ligand in the complex minus those of the corresponding entities in the solvent, while ΔH_{PL} was taken as the value of H_{PL} in the complex. The desolvation enthalpy was approximated by three different methods. First, a Molecular Dynamics calculation using the CVFF force field was used after energy minimisation of the solvated ligand, leading to non-realistic results. Subsequently, we tested a Quantum Chemistry method based on *Vega* [27], with partial atomic charges calculated by CM2 mapping of Mulliken charges obtained from the semi-empirical calculation with AM2 Hamiltonian (i.e. AM1–CM2). The *ChemSol* program [44] was used to simulate the solvation of the ligands when transferred from gas phase to an aqueous medium. Solvation enthalpies were computed by using a solvation model of Langevin Dipoles. A number of iterations (in average 25) were applied on the system to relax solute and solvent. In addition to a value for the solvation enthalpy, these calculations provided values for ligand–water interactions, and solvation free energy. The third approach used for calculating ΔH_{desolv} was based on a correlation proposed by Rashin and Namboodiri [45]:

$$\Delta H_{desolv} = -621.48 - 25.890\text{Area} \quad (5)$$

where ΔH_{desolv} is the enthalpy of non-electrostatic hydration in cal mol^{-1} of a molecule with its molecular surface area expressed in Å^2. The molecular surface area was computed by *Insight II* using the Connolly algorithm [46] with a probe radius of 1.4 Å. The missing electrostatic hydration term can be computed using *ChemSol*, as indicated above. Equation 5 was finally used to compute the ΔH_{desolv} for Equation 9, as it performed better (see results).

The entropic contributions to the binding-free energy were derived from empirical parameterisations. The protein conformational entropy change, ΔS_P, was calculated assuming that only the freedom of residues at the binding site is restricted upon ligand binding. The contributions to the binding-free energy due to freezing of internal degrees of freedom of the ligand and of the protein (ΔS_P and ΔS_L) were calculated as proposed [47–48]:

$$-T\Delta S = \text{TNR ln3} \quad (6)$$

where N is the number of side-chain rotatable bonds leading to altered positions of non-hydrogen atoms. This assumption considers a contribution of $-T\Delta S = 0.6$ kcal mol^{-1} per rotatable bond. The loss of translational and rotational entropy upon complex formation (ΔS_{PL}) has been estimated at 7–15 kcal mol^{-1} [47, 50–51]. We tried three values (7, 11 and 15 kcal mol^{-1}) for this term. They performed similarly, but that of 15 kcal mol^{-1} was slightly better. Finally, the entropic contribution of water (ΔS_w) was calculated from the empirical correlation with solvent-accessible surfaces (SAS) proposed by D'Aquino and co-workers [52]:

$$\Delta S_w = 0.45\Delta\text{ASA}_{np}\ln(T/384.15) - 0.26\Delta\text{ASA}_p \ln(T/335.15) \quad (7)$$

where ΔASA is the difference between the ASA [np: non-polar surface or p: polar, in Å^2] of the free and complexed molecules (calculated using the algorithm of Lee and Richards [53]), T is the temperature in K, and the entropy change is given in cal K^{-1} mol^{-1}. For the different ligands, the ΔASA was computed by subtracting the ASA of the solvated ligand from the ASA of the ligand inside the protein cavity. The latter value was assumed to be zero because the L99A cavity is fully buried so that the bound ligands are not accessible to solvent. For the same reason, the ΔASA for the protein is assumed to be zero because the surface cavity residues are always buried and no water molecules have been observed there. For protein cavities where evidence exists for the presence of buried water molecules, the

calculation of the protein ΔASA should take them into account.

As expressed by Equation 4, the ligand desolvation free energy is calculated by adding up the desolvation enthalpy and the hydrophobic effect. However, the desolvation free energy of a molecule can be directly calculated using other methods. One is the parameterisation of the change in solvation free energy proposed by Eisenberg and McLachlan [54] for protein folding and ligand binding processes, which considers octanol to water transfer:

$$\Delta G_{solv} = \Sigma \Delta \sigma_i \mathrm{ASAi} \quad (8)$$

where σ_i is the atomic solvation coefficient [54] and ASA_i the accessible-solvent area of each atom type: carbon, oxygen (neutral and charged), nitrogen (neutral and charged), and sulphur. The sign of the calculated ΔG_{solv} values was changed to represent the ΔG_{desolv} term in Equation 4. In addition, we tested Equation 8 using the solvation coefficients proposed by Wesson and Eisenberg [55] for the transfer from vapour phase to water. A second method used to approximate ΔG_{solv} was to simulate the process using Quantum Chemistry calculations with *ChemSol*. We used Equation 8 to calculate the ΔG_{solv} term in Equation 10, as it performed best.

The different terms in Equation 2, as calculated by the different methods indicated above, were used to calculate the grouped terms in Equations 3 and 4. Then, these free-energy functions were optimised by determining empirical weighting coefficients for their terms. To that end, the experimentally determined binding-free energies [25] of a training group of six known binders (*n*-butylbenzene: −6.7 kcal mol^{-1}, *iso*-butylbenzene: −6.5 kcal mol^{-1} ethylbenzene: −5.8 kcal mol^{-1}, *m*-ethyltoluene: −5.1 kcal mol^{-1}, indole: −4.9 kcal mol^{-1} and o-ethyltoluene: −4.5 kcal mol^{-1}) were fitted to Equations 9 and 10 so that the best *a–c* and *a–d* weighting coefficients were determined by performing multiparametrics fits with *Origin 6.0* (OriginLab Co).

$$\Delta G_b = a\Delta H_{total} - bT\Delta S_w - cT\Delta S_{P\&L} \quad (9)$$

$$\Delta G_b = a\Delta G_P + b\Delta G_L + c\Delta G_{desolv} + d\Delta G_{PL} \quad (10)$$

Design of mutant cavity-containing proteins, site-directed mutagenesis, protein expression and purification, preparation of apoproteins and spectroscopic characterisation

Single and double truncating mutations were introduced *in silico* on the wild type apo and holoflavodoxin structures (PDB-codes: 1FTG and 1FLV, respectively). Cavity containing mutants of *Anabaena* PCC 7119 flavodoxin were designed using a method previously developed in our laboratory [24]. Percentages of volume reduction of the modelled cavities were calculated as $100(V_t\text{-}V_x)/V_t$, where V_t is the theoretical cavity volume after *in silico* replacement of the mutated side-chain, and V_x is the volume calculated from the modelled coordinates of the mutant after minimisation. Hydrophobicity indexes were computed using the *WebLab ViewerPro* software (*Accelrys* Inc.), considering the cavity surface residues. Side-chain solvent accessibilities were computed with a 1.4 Å radius probe, using *Getarea* [56].

Several flavodoxin variants bearing truncating mutations aimed at generating buried cavities on the flavodoxin structure were prepared (W66F, W66F/L105A, W66F/L44A, I109A, W66F/I109A) using the *QuickChange* site directed mutagenesis kit from *Stratagene*, and the mutations were identified by sequencing the entire flavodoxin gene [57]. The expression and purification of wild type and mutant flavodoxins were performed as described [58]. To obtain the apoflavodoxins, the flavin mononucleotide group, FMN, was removed from the holoflavodoxins by trichloroacetic acid precipitation as described [59]. The concentrations of wild type and mutant holo and apoflavodoxins were determined from their absorbances at 464 and 278 nm, respectively, using extinction coefficients of 8.8 mM^{-1} cm^{-1} for the holoprotein and of 34.3 mM^{-1} cm^{-1} for the apoprotein [58]. For the variants containing the W66F mutation, extinction coefficients of 9.0 and of 27.6 mM^{-1} cm^{-1} were used for the holo and apoproteins [Dr S. Zorrilla, unpublished].

UV-Visible absorbance, fluorescence emission and far and near-UV circular dichroism (CD) measurements were carried out at 25 ± 0.1 ° C and pH 7.0 in either 10 mM sodium phosphate buffer (for far-UV CD, UV-Visible absorbance, and fluorescence) or 50 mM MOPS (for near-UV CD).

Thermal stability measurements

The stability of the apoflavodoxin variants was evaluated from thermal unfolding curves followed by fluorescence emission and far-UV CD, as described [60]. The buffer was 10 mM sodium phosphate, pH 7.0, and the protein concentration 2 μM for emission fluorescence and 20 μM for far-UV CD. To minimise the temperature dependence of fluorescence, thermal unfolding curves were registered as a ratio of the emissions at 320 and 360 nm. Thermal denaturation data were fitted to a two-state equation:

$$F = \frac{(F_N + m_N T) + (F_U + m_U T)e^{-(\Delta G/RT)}}{1 + e^{-(\Delta G/RT)}} \tag{11}$$

where F is the spectroscopic signal, F_N and F_U are the spectroscopic signals of the native and denatured states, m_N and m_U are the temperature dependences of the spectroscopic signals of the native and denatured state and ΔG is the unfolding free energy change:

$$\Delta G = \Delta H_{T_m}(1 - T/T_m) - \Delta C_p((T_m - T) + T \ln(T/T_m)) \tag{12}$$

where $\Delta H_{T\mathrm{m}}$ is the enthalpy difference at the melting temperature, T_m is the temperature in the transition mid point, and ΔC_p is the unfolding heat capacity change. Fitting thermal unfolding data to Equation 11 allows an accurate determination of T_m and a reasonable determination of $\Delta H_{T\mathrm{m}}$. However, the ΔC_p values so calculated are less reliable [61] and will not be reported. We notice that the T_m values calculated using the two-state equation (Equation 11) are just apparent temperatures because the thermal unfolding of apoflavodoxin is not two-state [62]. We have chosen to use apparent temperatures, instead of the microscopic T_ms that can be derived from global fittings of fluorescence and CD thermal unfolding curves [62], because apparent temperatures are most accurately calculated and therefore are ideal for thermal up-shift assays.

After evaluating the stability and conformational integrity of the different flavodoxin variants, the W66F/L44A mutant was chosen. All subsequent thermal up-shift experiments performed in this mutant, and the corresponding controls with the wild type protein, have focused in detecting changes in the apparent T_ms determined from the far-UV CD unfolding curves. This is so because, due to the presence of an intermediate in the thermal unfolding of apoflavodoxin [62], which could also appear in the destabilised W66F/L44A holoprotein, the apparent temperatures of mid-denaturation derived from fluorescence unfolding curves tend to reflect the differential binding energetic of the native state relative to a thermal intermediate. Since the intermediate displays a native-like structure in the vicinity of the cavity present in the W66F/L44A mutant [63] no thermal up-shift effects are expected in the fluorescence curves upon ligand binding. In contrast, the CD curves reflect the unfolding of the intermediate conformation. Thermal upshifts related to ligand binding are thus expected to be observed in the CD curves.

Binding detection by up-shift of the thermal denaturation temperature

Five representative compounds that were predicted to bind to the W66F/L44A flavodoxin cavity by using the three-step massive screening method developed were experimentally tested (benzylamine, pyridine, furan, benzene and toluene). They were selected to exhibit low extinction coefficients at 222 and 280 nm, which was required for the spectroscopically followed thermal up-shift assays. In addition, we selected three compounds that had been predicted not to bind to the W66F/L44A flavodoxin cavity (2-phenyl-propanol, butanol and choline).

The interaction of a ligand with a folded protein results in an increase in the denaturation temperature [64]. The binding of the eight compounds selected was first investigated using a thermal unfolding assay [65] in which the melting temperatures of the protein in the presence and absence of ligand were compared. To that end, the thermal unfolding of W66F/L44A flavodoxin in 10 mM sodium phosphate buffer, pH 7.0, was monitored by far-UV CD (222 nm) in a 0.1-cm cuvette, using an in-cell thermal probe. Aliquots of protein solutions were added with a Hamilton syringe to buffered solutions of different ligands and the mixtures were allowed to reach equilibrium at 25 ± 0.1 °C for 14 ± 2 h, in darkness. The final holoflavodoxin concentration was 20 μM and

that of ligand 10 mM. The thermal denaturation curves were fitted to a two-state model (Equation 11) and T_ms and van't Hoff enthalpy changes were calculated. Each thermal denaturation experiment was performed twice. To detect potential non-specific effects of the ligands tested on the stability of the protein, independent controls were performed using wild type flavodoxin (that lacks a suitable binding cavity). In this way, any unforeseen effect of a ligand on wild type stability can be subtracted from the observed thermal up-shift measured for the cavity mutant. Relative thermal up-shifts are thus also reported.

Determination of dissociation constants by ligand titration

The binding to W66F/L44A of the three ligands exhibiting larger relative thermal up-shifts (benzylamine, benzene and furan) was further investigated by ligand titration. The dissociation constants of the holoflavodoxin–ligand complexes were determined by fluorimetry (excitation at 280 and emission at 336 nm) in an *Aminco-Bowman Series 2* fluorimeter at 25 ± 0.1 °C (in darkness), using protein/ligand mixtures that had been equilibrated for 14 ± 2 h. These samples were prepared by mixing 900 μl of 1–300 mM ligand with 100 μl aliquots of 40 μM holoprotein, in 10 mM sodium phosphate, pH 7.0. Binding of the ligands to the cavity significantly quenched the fluorescence emission of the protein. The dissociation constants were calculated by fitting the emission fluorescence to Equation 13 [66], that assumes a 1:1 stoichiometry:

$$F = F_p + \frac{F_c - F_p}{2C_p}(C_p + C_l + K_d - \sqrt{(C_p + C_l + K_d)^2 - 4C_pC_l}) + bC_l \quad (13)$$

where F is the observed fluorescence emission intensity for each solution, F_p and F_c are the emissions of the protein and complex, respectively, C_p and C_l are the total protein and ligand concentrations in the solution, K_d the dissociation constant of the flavodoxin–ligand complex, and b a proportionality constant required to fit the experimental data. The robustness of the fits is indicated by the fact that, when the known concentration of protein was treated as an unknown, it was correctly predicted in all cases. Each ligand titration was performed twice.

Results and discussion

Validation of the fast-docking, second step of the screening method

The screening method developed is three-step. First, simple filters are used to substantially reduce the number of compounds that will be subjected to further screening. Then, a fast docking and ranking of the selected molecules is performed. In the third step, highly ranked potential ligands are subjected to a fine docking and binding-free energy evaluation that discards false positives.

The fast docking intermediate step has been implemented and validated (see Methods) using L99A lysozyme–ligand complexes for which binding data and X-ray structures are available [20, 23, 25]. The L99A lysozyme mutant is a very useful model to study binding to protein cavities because it has been thoroughly characterised both structurally [20], in its ability to bind ligands [25, 67], and as a tool for testing scoring functions and new docking algorithms [68, 69]

The best docking and scoring conditions found for a training set of 12 L99A lysozyme ligands (6 binders and 6 non-binders) are listed in Table 1. The *Consensus Scoring* function provided by the *Cerius*^{2} program (see Methods) performed well with the training set and then with the 40 lysozyme-related ligands (see Methods), but when applied to the compounds in the NCI database (supplemented with the 40 lysozyme-related ligands) that had been selected using the first step filters, it produced only a small enrichment. This prompted us to develop and alternative scoring function (Equation 1) that combines different energy terms provided by the *Cerius*^{2} program. The performance of Equation 1 in discriminating within the 40 lysozyme-related ligands was just similar to that of the *Consensus Scoring* but led to a much higher enrichment when applied to ranking the NCI compounds selected after the first step of the screening (see below).

Table 2. Root mean square deviation (RMSD) of modelled and X-ray complexes.

PDB	Ligand	RMSD (Å)[a]
181L	Benzene	0.22
182L	Benzofuran	0.34
183L	Indene	0.35
184L	*iso*-Butylbenzene	0.45
185L	Indole	0.35
186L	*n*-Butylbenzene	0.31
187L	*p*-Xylene	0.29
188L	*o*-Xylene	0.39
1NHB	Ethylbenzene	0.25

[a]RMSD calculated as indicated in the *Materials and Methods* section, considering the heavy atoms of ligand and cavity surface residues.

Validation of the fine docking, third step of the screening method. I: Accurate prediction of the structure of docked complexes at the L99A cavity

The third step of the screening aims at a full discrimination between binders and non-binders. It is performed using a fine docking protocol that produces accurate structures of the known complexes and expels most decoys from the docking cavity. The remaining false positives are subsequently identified by computing the binding affinity of the docked complexes.

The fine docking protocol (described in Table 1) combines algorithms already applied to other docking targets [70–74] and has been trained to model 17 ligand–protein complexes between

Table 3. Binding-free energies calculated with Equation 9 for ligands that bind to L99A T4 lysozyme (binders) and ligands that do not. Comparison with experimental binding energies[a].

	ΔH	$T\Delta S_w$	$T\Delta S_{L\&P}$	ΔG_b	ΔG_e
Binders					
Ethylbenzene	1.1	−7.4	0.3	−6.0	−5.8
o-Xylene	1.9	−6.7	0.3	−4.6	−4.6
m-Xylene	1.8	−6.9	0.3	−4.8	−4.7
p-Xylene	1.6	−7.3	0.3	−5.4	−4.7
Propylbenzene	1.1	−8.1	0.3	−6.7	−6.5
o-Ethyltoluene	2.5	−7.5	0.3	−4.7	−4.5
m-Ethyltoluene	2.4	−8.6	0.3	−5.9	−5.1
p-Ethyltoluene	1.2	−7.3	0.3	−5.9	−5.4
Benzofuran	0.9	−7.0	0.3	−5.8	−5.5
Indene	1.5	−7.5	0.3	−5.8	−5.1
Indole	1.0	−6.1	0.3	−4.8	−4.9
Benzene	1.1	−6.5	0.3	−5.1	−5.2
Toluene	0.9	−6.4	0.3	−5.3	−5.5
Iso-Buthylbenzene	2.6	−8.7	0.3	−5.8	−6.5
n-Buthylbenzene	2.0	−8.7	0.3	−6.4	−6.7
Styrene	0.8	−7.8	0.3	−6.7	–
Naphthalene	2.2	−7.7	0.3	−5.2	–
Average ± SD	1.6 ± 0.6	−7.4 ± 0.8	0.3 ± 0.0	−5.5 ± 0.7	−5.4 ± 0.7
Non-binders					
p-cresol	1.4	−5.4	0.3	−3.8	
Benzyl alcohol	1.0	−5.1	0.3	−3.9	
Octanol	10.9	−7.9	0.3	3.2	
Azulene	4.6	−8.7	0.3	−3.9	
tert-Buthylbenzene	6.6	−8.1	0.3	−1.3	
1-Phenylheptane	9.8	−11.1	0.3	−1.1	
Average ± SD	5.7 ± 4.2	−7.7 ± 2.2	0.3 ± 0.0	−1.8 ± 2.8	

[a]ΔG_b are binding-free energies calculated using Equation 9 and the modelled structures of the corresponding complexes. ΔG_e are experimentally determined binding-free energies [25]. The ligands used to optimise Equation 9 are shown in italics. The non-binders in the Table re the only false positives of 23 non-binders that were docked onto the protein.

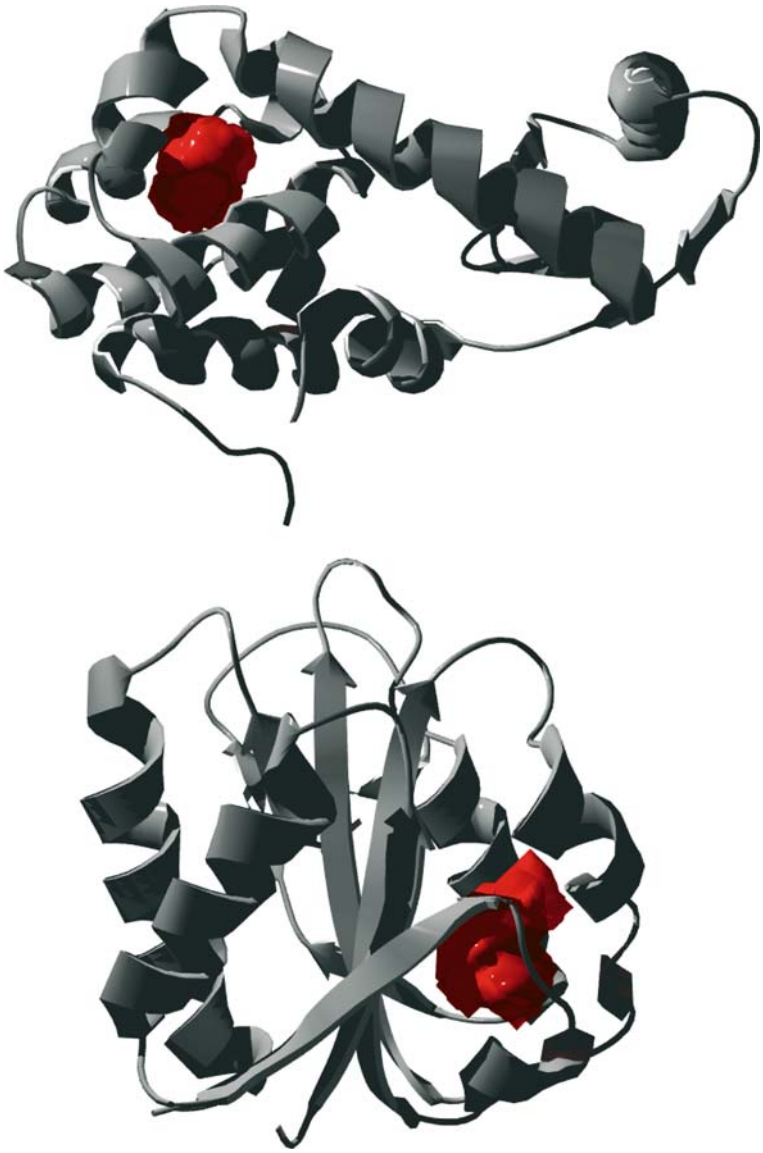

Figure 2. Ribbon diagrams of the L99A T4 lysozyme (above) and W66F/L44A flavodoxin (below) mutant modelled structures showing in red the engineered cavities that have been used, respectively, to develop and to experimentally testing the virtual screening method.

L99A mutant lysozyme (comprising 15 apolar and 2 slightly polar ligands, as listed in Table 3). Some additional ligand-lysozyme L99A complexes and some non-binders listed by Morton et al. [25] were not modelled since the ligands contained atom types for which the CVFF force field is not able to define partial charges. For each ligand, the best scored-complex, based on the lowest final potential energy, was selected from a total of 30 different docked conformations. Nine of the 17 complexes modelled had been previously crystallised [23] and their structures are available. As shown in Table 2, their structures could be accurately modelled, with RMSD between the experimental and modelled complexes ranging from 0.22 to 0.45 Å for the cavity-surface atoms and the ligand. Figure 2 shows, as an example, a superposition of the experimental and modelled structures of the ethylbenzene complex, and allows a visual evaluation of the quality of the modelling.

In addition to provide accurate structures of the complexes formed by the binding ligands, a

useful docking procedure should identify the ligands that do not bind to a given protein-binding site. To investigate whether the procedure developed was able to perform this task efficiently, we modelled into the L99A cavity a group of 23 compounds that have been described as non-binders [25]. Only 26 % false-positives (*p*-cresol, benzyl alcohol, octanol, azulene, *tert*-butylbenzene, and 1-phenylheptane) were obtained from the docking as the other non-binding compounds were expelled from the cavity. The docking protocol implemented thus allows a precise calculation of the three-dimensional structure of the complexes analysed and identifies two thirds of the decoys tested. Nevertheless, it still gives rise to one fourth of false positives. Since we aim at using this docking as the final step in the screening of large libraries, a more stringent performance is required.

Validation of the fine docking, third step of the screening method. II: Calculation of free energies of binding and discrimination of false positives.

The tool needed to identify the remaining false positives is provided here by a comprehensive binding-free energy function (Equation 2) that we have optimised to accurately predict the known binding affinities of the training complexes from their modelled structures. In this work we have chosen to calculate the affinity of the modelled L99A T4 lysozyme complexes by using a binding equation as complete as possible. Since eight different terms appear in the equation, which can be calculated or estimated in various manners, an optimisation of the equation seems required that we have addressed by calculating the weighting coefficients (Equations 9 and 10) that lead to a better affinity prediction. The different approaches that have been considered for calculating the values of some of the terms (see Methods) are in this way evaluated. Rather that fitting eight weighting coefficients, we have clustered the terms in Equation 2 in two different manners (Equations 9 and 10). We have interpreted that, whenever a negative weighting coefficient is obtained for any of the terms in Equations 9 and 10, the particular combination of approaches used to calculate the different terms is not appropriate, irrespectively of whether the global correlation obtained for the calculated and experimentally determined binding energies of the complexes is good or otherwise. By sticking to this criterion we hope to increase the portability of the final optimised equation to other proteins (see below).

In Equation 9 all the calculated terms are grouped in just three terms: total enthalpy change, water entropy change, and conformational, rotational and translational entropy change. The desolvation enthalpy included in the enthalpy change was approximated by three different methods and the Rashin's approach [45] gave the best fit leading to a correlation coefficient between the experimental and predicted ΔG_b values of the six selected training complexes of $r = 0.83$ (SD of $\pm$ 0.43 kcal mol^{-1}). Extrapolation of the coefficients obtained to the seventeen ligands that are known to bind (Figure 3; Table 3) gives $r = 0.83$, with a SD of ± 0.38 kcal mol^{-1}. In addition, the non-binders ΔG_b values are clearly less negative than those of the binders: in average 3.7 kcal mol^{-1}; Table 3. In Equation 10 we have clustered enthalpy and entropy terms into free energy terms. The best estimation of the desolvation free energy was provided by the parameterisation proposed by Eisenberg and McLachlan [54], which gave rise to $r = 0.89$ (SD = ± 0.38 kcal mol^{-1}) for the six training complexes. Extrapolation of the coefficients to the seventeen known binders (see Figure 3; Table 4) gives $r = 0.81$, with a SD of ± 0.38 kcal mol^{-1}. As with the optimised Equation 9, the non-binders ΔG_b values were clearly less negative than those of the binders: in average 3.4 kcal mol^{-1} (Table 4). To test the robustness of the fits to Equations 9 and 10 several training sets were considered (four different groups of six ligands) for which very similar *a–c* or *a–d* coefficients and correlation coefficients were obtained. We notice that the particular values of the translational and rotational entropies used were almost irrelevant for the quality of the fits to Equations 9 and 10. A value of 15 kcal mol^{-1}, however, leads to a slightly better fit and was thus selected. Although the binding-free energies calculated using Equations 9 and 10 are very similar (see Tables 4 and 5), the two equations provide complementary clues about the basis of the discrimination between binders and non-binders.

In the best fits to Equation 9, using the Rashin's approach, the binding enthalpy change is weighted (Table 5) by a factor of 0.18 and gives rise to an unfavourable contribution to the binding-free energy difference. Conversely, the water entropy change is weighted by a factor of 0.80 and

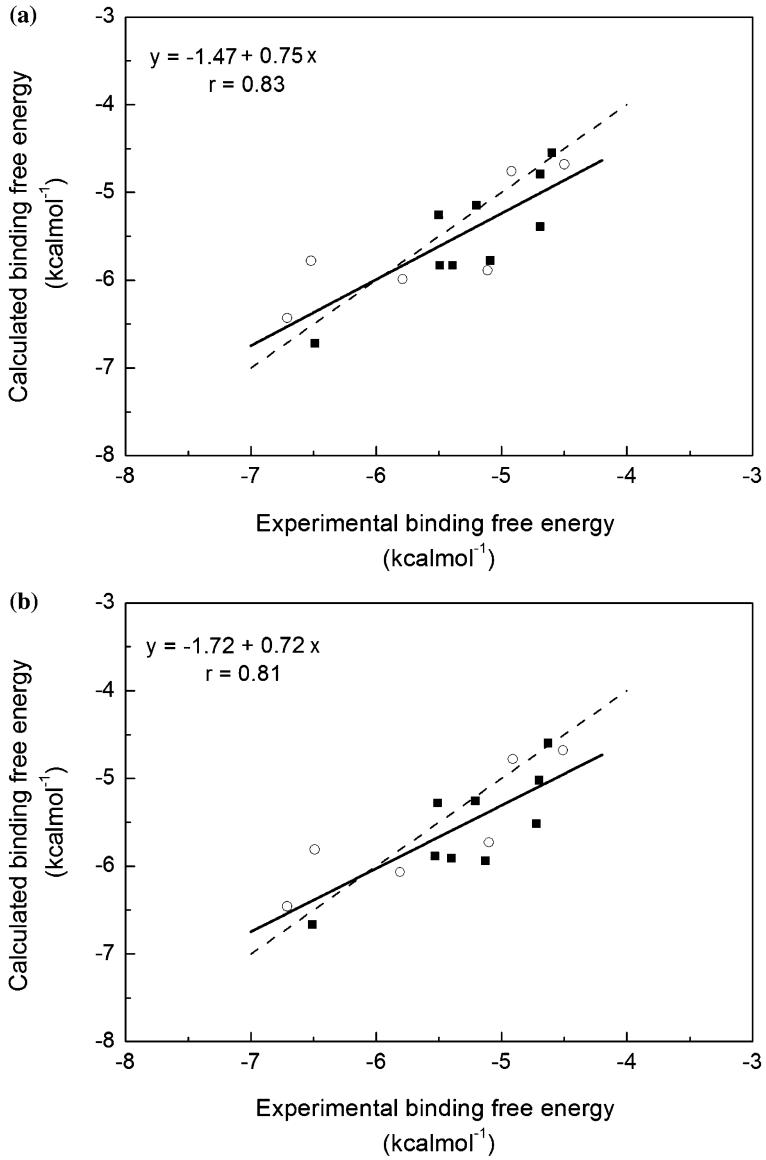

Figure 3. Calculated binding-free energies versus experimentally determined ones (from [25]) (a) Binding energies calculated with Equation 9. (b) Binding energies calculated with Equation 10. Open circles indicate the training set of positive controls used for fitting the *a, b, c,* (*d*) parameters in Equations 9 and 10. Black squares indicate the additional known binders. Solid lines are linear fits of all the data to the equations shown in the corresponding figure. Dashed lines are also shown that correspond to perfect unity fits. The slope-unity fit for the upper panel is $y = -0.18 + x$ ($r = 0.82$), and for the lower panel: $y = -0.14 + x$ ($r = 0.82$).

yields a large and favourable contribution to the binding-free energy change. Finally, the conformational, translational and rotational entropy, clustered as $-T\Delta S_{P\&L}$, are weighted by 0.02 and contribute unfavourably, but little, to the binding-free energy. Thus the only stabilising contribution to ligand binding is related to the entropy of water, which contributes, in average, −7.4 kcal mol^{-1} to the binding-free energy function. Interestingly, although the major contribution to binding is the hydrophobic effect, this doesn't contribute to discrimination because, in average, it is very similar in binders and non-binders (−7.4 and −7.7 kcal mol^{-1}, respectively). We remark that discrimination is exerted at the enthalpy level, where non-binders show, in average, an enthalpy change about 3.9 kcal mol^{-1} greater (more unfavourable) than binders. This is particularly true of

Table 4. Binding-free energies calculated with Equation 10 for ligands that bind to L99A T4 lysozyme (binders) and ligands that do not. Comparison with experimental binding energies[a].

	ΔG_P	ΔG_{desolv}	ΔG_L	ΔG_{PL}	ΔG_b	ΔG_e
Binders						
Ethylbenzene	1.7	−7.1	0.4	−1.0	−6.1	−5.8
o-Xylene	1.8	−6.5	0.9	−0.8	−4.6	−4.6
m-Xylene	1.8	−6.6	0.5	−0.7	−5.0	−4.7
p-Xylene	1.8	−7.0	0.6	−0.9	−5.5	−4.7
Propylbenzene	1.7	−7.8	0.6	−1.2	−6.7	−6.5
o-Ethyltoluene	1.8	−7.2	1.0	−0.3	−4.7	−4.5
m-Ethyltoluene	1.8	−8.2	1.2	−0.5	−5.7	−5.1
p-Ethyltoluene	1.7	−7.1	0.7	−1.4	−5.9	−5.4
Benzofuran	1.7	−6.7	0.6	−1.5	−5.9	−5.5
Indene	1.8	−7.3	0.6	−1.1	−5.9	−5.1
Indole	1.7	−5.8	0.6	−1.3	−4.8	−4.9
Benzene	1.7	−6.3	0.1	−0.8	−5.3	−5.2
Toluene	1.7	−6.2	0.4	−1.2	−5.3	−5.5
Iso-Buthylbenzene	1.8	−8.3	1.0	−0.3	−5.8	−6.5
n-Buthylbenzene	1.9	−8.3	1.1	−1.1	−6.5	−6.7
Styrene	1.7	−7.5	0.1	−1.1	−6.8	–
Naphthalene	1.7	−7.4	1.0	−0.4	−5.1	–
Average ± SD	1.8 ± 0.1	−7.2 ± 0.8	0.7 ± 0.3	−0.9 ± 0.4	−5.6 ± 0.7	−5.4 ± 0.7
Non-binders						
p-Cresol	1.7	−5.2	0.5	−0.8	−3.8	
Benzyl alcohol	1.7	−5.0	0.3	−1.0	−4.0	
Octanol	4.7	−7.3	2.2	0.7	0.3	
Azulene	1.8	−8.3	3.6	−0.8	−3.7	
tert-Buthylbenzene	2.0	−7.8	1.5	3.0	−1.3	
1-Phenylheptane	2.4	−10.7	4.6	3.5	−0.2	
Average ± SD	2.4 ± 1.2	−7.4 ± *2.1*	2.1 ± 1.7	0.8 ± 2.0	−2.1 ± *2.0*	

[a]ΔG_b are binding-free energies calculated using Equation 10. ΔG_e are experimentally determined binding-free energies [25]. ΔG_e are experimentally determined binding-free energies [25]. The ligands used to optimise Equation 10 are shown in italics. The non-binders in the Table are the only false positives of 23 non-binders that were docked onto the protein.

apolar non-binders, which shows that the discrimination achieved is not simply due to differences in polarity. While this picture is in qualitative agreement with the observed fact that enthalpy effects contribute to stabilise the complexes [25] we notice that our calculation does not capture the actual sign of the observed enthalpy change, possibly due to inaccuracies in at least one of the terms grouped in the overall enthalpy change term (possibly the protein enthalpy change for which fairly large positive values have been calculated for most of the complexes; not shown).

Table 5. Optimal coefficients for the binding-free energy functions given by Equations 9 and 10[a].

Equation 9		Equation 10	
Contributing term	Fitting coefficient (value)	Contributing term	Fitting coefficient (value)
Enthalpy (ΔH)	*a* (0.18)	Protein (ΔG_P)	*a* (0.07)
Water entropy ($T\Delta S_w$)	*b* (0.80)	Desolvation (ΔG_{desolv})	*b* (1.64)
Conf., rot. and transl. Entr. ($T\Delta S_{p\&L}$)	*c* (0.02)	Ligand (ΔG_L)	*c* (0.28)
		Protein–ligand (ΔG_{PL})	*d* (0.19)

[a]The errors associated to the fitted coefficients of Equation 9 were ≤ 0.01, and those of the coefficients in equation 10 ≤ 0.02.

On the other hand, in the best fit to Equation 10 (Table 5), using Eisenberg's solvation parameters, the change in binding site free energy and ligand free energy were weighted by factors of 0.07 and 0.28, respectively, and give rise to unfavourable contributions to the binding-free energy difference. Conversely, the ligand desolvation free energy is weighted by a factor of 1.64 yielding a large and favourable contribution to the binding-free energy change. Finally, the change in protein–ligand free energy is weighted by 0.19 and also contributes favourably to the binding-free energy. The binding site and ligand free energy changes thus oppose the formation of the complexes by contributing, in average, 1.8 and 0.7 kcal mol^{-1}, respectively (this is expected because, in order to form the complexes, both protein and ligand need to readjust their initial equilibrium conformations). The desolvation free energy of the ligand and the protein–ligand free energy are stabilising factors contributing −7.2 and −0.9 kcal mol^{-1}, respectively, to the binding-free energy. We notice that ligand desolvation is not discriminating because it is similar in binders and in non-binders (−7.2 and −7.4 kcal mol^{-1}, respectively). Discrimination takes place as a consequence of the poor complementarity of non-binders and binding site. Non-binders are more destabilised in their bound conformation than binders. The same applies to the binding cavity. In addition, the interaction enthalpy change between non-binders and protein, from their interacting conformations, is destabilising, unlike in the modelled complexes of the binder compounds.

It is thus clear that discrimination between binders and non-binders is not just related to differences in polarity. According to our analysis, the overall poorer affinity of the non-binders is related neither to a different solvent entropy change nor to a significantly different ligand desolvation free energy. Instead, it derives from a greater destabilisation of the protein-binding site and of the ligand when they bind, together with a poorer interaction between the two molecules in the complex. The combination of these effects gives average binding energies of -5.6 ± 0.7 kcal mol^{-1} for binders and of -2.0 ± 2.4 kcal mol^{-1} for non-binders, using either Equation 9 or 10. There is thus a clear cut between the two groups and the combined use of the docking procedure and of the optimised free energy functions allows a full discrimination between binders and non-binders. It seems that, although the T4 lysozyme ligands data base used to train the screening is small, it contains enough diversity (among the non-binders there are both polar and apolar molecules) so that the method developed can discriminate beyond simple polarity considerations.

Recent work by Wei et al. [68, 69] has demonstrated the relevance of massive screening methods to pinpoint potential ligands from a database composed by 172,118 molecules (the Available Chemical Directory). These authors have also shown the usefulness of the L99A T4 lysozyme model for improving scoring functions. Our contribution in the present work is a full discrimination between binders and non-binders, which arises from the combination of a flexible docking method and an accurate parameterisation of a comprehensive binding-free energy function (Equations 9 and 10). This full discrimination is a very useful tool in order to address the experimental testing of predictions, as will be illustrated in the testing of the portability of the method to a non-related protein (see below).

Overall performance of the three-step method. Selecting ligands of the T4 lysozyme L99A mutant cavity within the NCI database

The overall performance of the three-step screening method has been tested by screening the NCI Database (250251 compounds) supplemented with the 40 L99A lysozyme ligands used in this work The first-step, based on volume and polarity filters, selected 4913 compounds, which included all the known binders (17 compounds) and 16 out of the 23 non-binders analysed (seven non-binders were filtered out in this step). The 4913 selected compounds were then submitted to the second step and docked against the L99A lysozyme cavity, using the fast docking method. With the standard *Consensus Scoring* evaluation provided by the sofware, the 17 known binders of the L99A lysozyme were found within the 3824 best-scoring compounds, together with 12 out of the 16 decoys pre-selected in the filtering step. Since this discrimination was poor, we used a different *Scoring* function (Equation 1) to rank the remaining 3824 compounds, and the 17 known binders where ranked within the 469 best-scoring compounds,

together with only 3 decoys (cyclohexane, furan and azulene).

To evaluate the performance of the filtering and fast-docking steps, enrichment factors (concentration of binders in subset/concentration of binders in database) were calculated [69] as: b_sD/bD_s, where b_s is the number of known binders found within the D_s candidate compounds selected in a given step, and b is the total number of known binders in the D compounds of the database. The enrichment factor provided by the filtering step was 51 (calculated with $D_s = 4913$). In the second step, the *Consensus Scoring* function only raised the overall enrichment to 65 (with $D_s = 3824$), but using Equation 1, it reached a value of 534 (calculated with $D_s = 469$). As explained, all the known non-binders of the L99A lysozyme cavity mutant can be unequivocally identified, and thus discarded, using a subsequent docking procedure. Therefore, the strategy to screen for binders to a protein cavity will be simply to subject to the third step (fine docking plus energy evaluation of the modelled complexes using equations 9 or 10) compounds highly ranked in the second step.

The initial filtering step of the NCI database can be done in 3 h of CPU time. As for the second step, the fast docking of 5000 compounds will take 1 week of CPU time in a Silicon Altix 3000-type machine. The final fine docking step can be performed in the same machine at a rate of 1 compound every 25 min.

Testing the portability of the three-step screening method to a different protein. Design of cavity mutant flavodoxins and identification of a cavity-bearing flavodoxin suitable for hosting ligands

A most important issue that deserves consideration is that of the portability of the method to other proteins. Computational methods that are optimised using a particular dataset (i.e. a particular protein) are, sometimes, difficult to use in a different context. Although we have optimised the second and third steps of the screening method using a training set and then shown that it works similarly well for the rest of the known binders of the model cavity-bearing lysozyme mutant, it could be argued that the method might not work in a different protein. We have addressed this question by designing and engineering a new cavity in a non-related protein (flavodoxin) where the binding of several ligands that we predict either as binders or non-binders is tested. To that end, several mutant flavodoxins were initially designed and modelled from the known wild type X-ray structure [75]. The suitability of the W66F mutant to host additional mutations was explored because a small cavity was observed to be already present in its modelled structure and because the mutant protein, already available in the laboratory, was known to be well-folded and reasonably stable. Most truncated mutations lead to moderate cavity reductions, although in the I109A, W66FI109A, and L105A mutants small volume increases were

Table 6. Modeled flavodoxin cavity mutants.

Mutant	Theoretical cavity volume[a]	Modelled cavity volume[b]	% Cavity vol. reduction[c]
W66F/L44A[d]	251	232	8
W66F/L105A[d]	255	217	15
W66F/I109A[d]	205	210	−2
W66F[d]	173	140	19
L44A	151	137	9
L105A/L44A	151	137	9
I109A[d]	111	116	−5
I52A	97	91	6
L105A	52	61	−17

[a]Calculated after *in silico* mutagenesis, without minimisation; they represent the volumes in Å^3 of the newly created cavities before relaxing to their final conformation.
[b]Calculated after minimisation, as described in *Methods*. Volume given in Å^3.
[c]Calculated as $100(V_t-V_x)/V_t$, where V_t and V_x are, respectively, the theoretical and modelled cavity volumes.
[d]Mutations engineered into the flavodoxin gene, and proteins expressed and evaluated in order to chose an appropriate cavity-bearing model protein as host of small molecules.

observed (see Table 6). Cavity collapses were not observed possibly due to the low flexibility of the region. The modelled virtual mutants predicted to contain larger cavities (Table 6) were produced and expressed (W66F, W66F/L105A, W66F/L44A and W66F/I109). In addition, we prepared the I109A mutant because we felt curious about a water molecule observed in the crystal structure near I109, and wanted to determine whether the protein would tolerate a mutation in that region.

The overall structure of the mutants was analysed by fluorescence emission and CD. Although we planned to use the holoproteins to host the ligands, we chose to analyse also the apo forms because, due to their lower stability, they could more easily reveal unexpected departures from the modelled, close-to-wild type, structures. The fluorescence emission spectra of the W66F, W66F/L44A, and W66F/L105A mutant proteins were similar to that of wild type (with emission maxima at 336 nm), while those of I109A and W66F/I109A showed a red-shifted emission maximum at 340 nm and notoriously lower fluorescence intensities (not shown). For the holoproteins, the emission maximum was at 332 nm in the wild type protein and in the W66F, W66F/L44A and W66F/L105A mutants, while the I109A and W66F/I109A mutants showed a blue-shifted emission maximum at 330 and 328 nm, respectively, and their emission intensity was slightly lower (not shown). The fluorescence data thus suggests that the I109A and W66F/

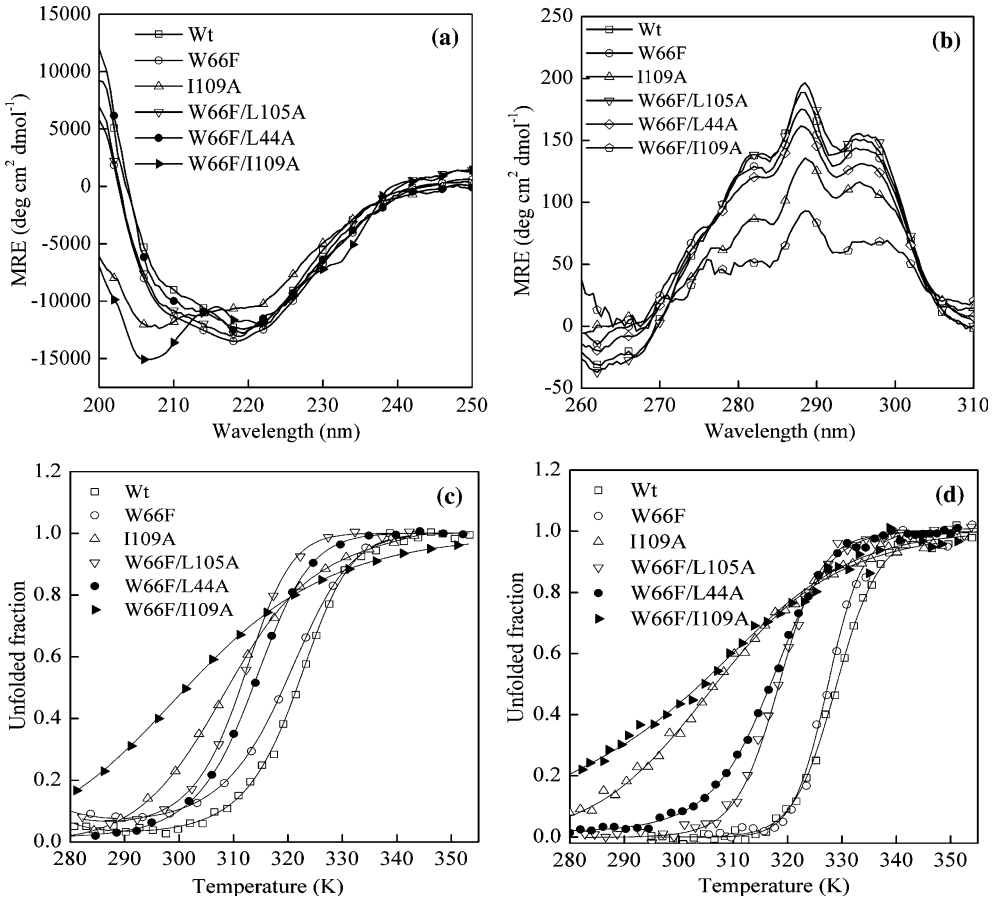

Figure 4. Spectroscopic and thermodynamic characterisation of mutant apoflavodoxins. (a) Far-UV CD spectra at pH 7.0 and 25 °C. Protein concentration: 2 μM, in 10 mM phosphate buffer. (b) Near-UV CD spectra at pH 7.0 and 25 °C. Protein concentration: 10 μM, in 50 mM MOPS buffer. (c) Thermal unfolding of wild type and mutant apoflavodoxins in 10 mM phosphate, pH 7.0. The unfolding has been followed by fluorescence emission (excitation at 280 nm) using 2 μM protein solutions The solid lines represent fractions of unfolded protein as functions of temperature, calculated after fitting of experimental curves to Equation 11. (d) Thermal unfolding followed by far-UV CD (at 222 nm) using 20 μM protein solutions.

I109A mutants might show some departure from the wild type structure. Similarly, the CD spectra in the far-UV of most mutants were very close to that of the wild type protein (Figure 4a). In contrast, the spectra of I109A and W66F/I109A showed clear indications of partial unfolding at 25.0 °C. As for the near-UV CD spectra (Figure 4b), those of W66F, W66F/L44A, and W66F/L105A apoflavodoxins are very similar to that of wild type while those of I109A and W66F/I109A show a less intense spectra (Figure 4b), indicative of a debilitation of their tertiary interactions at 25.0 °C. All the spectroscopic evidence available indicates that the global fold of the protein is not affected in some of the mutants (W66F, W66F/L44A and W66F/L105A), while there are clear indications of partial unfolding in both I109A and W66F/I109A.

To ensure that the mutant flavodoxin selected as a host protein is reasonably stable we have determined the thermostability of the candidates. Flavodoxin (in both the apo and holo forms) undergoes, at neutral pH, thermal transitions that can be followed using different spectroscopic techniques [62]. Thermal denaturation of apoflavodoxin leads to a decrease in fluorescence emission intensity and to a marked loss of secondary structure. Although the apoflavodoxin thermal unfolding is a three-state equilibrium [62], both the fluorescence and far-UV CD curves are often simple sigmoidals (this is the case of all the mutants analysed here) that can be fitted to a two-state equation to calculate apparent transition temperatures. This is more sensitive than using the globally fitted microscopic T_{m1} and T_{m2} of the N-to-I and I-to-D equilibria and has therefore been preferred. For the holoprotein, far UV CD apparent T_m values have been used to detect ligand binding by thermal up-shift assays (see below).

Table 7. Apparent denaturation temperatures of wild type and mutant apoflavodoxins, as measured by fluorescence emission and far-UV CD[a].

Protein	Fluorescence	Far-UV CD
Wt	322.9 ± 0.2	329.1 ± 0.1
W66F	319.2 ± 0.3	327.3 ± 0.1
W66F/L105A	312.3 ± 0.3	318.9 ± 0.1
W66F/L44A	313.6 ± 0.1	318.2 ± 0.2
I109A	307.6 ± 0.4	306.8 ± 0.6
W66F/I109A	301.6 ± 0.8	303.9 ± 1.4

[a]Parent T_ms (in K) ± standard errors calculated using Equation 11 (see *Materials and Methods*).

Fractions of folded protein versus temperature, deriving from fluorescence unfolding curves, are represented in Figure 4c for the wild type and the five mutant apoflavodoxins. The curves indicate that creating internal cavities in the protein markedly decreases the stability (Table 7), as has been reported for other proteins [25, 76]. The more notorious destabilising effects are observed in the I109A and W66F/I109A mutants, with decreases of 15° and 21°, respectively. Similarly, the apparent temperatures, calculated from the far-UV CD transitions (Figure 4d) confirm that the replacement of large side chains by smaller ones markedly decreases the thermal stability of flavodoxin (Table 7). The largest destabilisations are observed in the I109A and W66F/I109A mutants (of up to 25°). The thermal unfolding data thus confirm that the I109A mutation is very destabilising (both by itself and combined with W66F). Based in the spectroscopic characterisation and in the thermal unfolding analysis it seems that I109A and W66F/I109A are not suitable as ligand hosting proteins because they are quite unstable and partly unfolded at room temperature. Of the remaining stable mutants, W66F shows a significantly smaller cavity than those of W66F/L105A and W66F/L44A. Based on its reasonable stability and on its larger cavity volume, we have finally selected W66F/L44A flavodoxin as the hosting protein.

Identification of candidate ligands of the W66F/L44A flavodoxin cavity

The NCI database (supplemented with 40 small compounds, as explained) has been screened for binders of W66F/L44A holoflavodoxin, using the three-step method developed. From the hit list produced after the second step, nine compounds have been selected (furan, toluene, pyridine, benzylamine, benzene, ethylbenzene, cyclohexane, methanol, and propanol) that are ranked within the 731 best-scoring molecules. In addition to availability, one important criterion for the selection here is that the compounds must not interfere with the thermal up-shift assays used to detect binding; therefore, they had to display reasonably low extinction coefficients at the wavelengths of analysis. This limitation would not apply, for example, if isothermal titration calorimetry is used

to detect binding, although with this technique binders of low affinity may pass undetected. Following our screening scheme, the nine compounds selected have been docked into the flavodoxin cavity using the fine docking procedure (third step) and the calculated affinities of the modelled complexes (Equation 9) have been used to discard two of them (propanol and cyclohexane). Of the seven potential binders we have selected five for experimental testing, based in their spectroscopic properties and solubility (benzene, toluene, pyridine, furan, and benzylamine). In addition we have selected three decoys from the worst-scoring compounds in the hit-list (butanol, choline, and 2-phenyl-propanol) for comparison. The five potential binders selected and the three decoys have been experimentally tested for binding to the W66F/L44A flavodoxin cavity.

Binding of candidate ligands to W66F/L44A flavodoxin detected by up-shift of denaturation temperatures and by direct ligand titration

Thermal up-shift measurements is a simple way to detect ligand binding to proteins. The advantages and limitations of the method have been discussed [65, 77]. To experimentally testing the predictions of the screening, we have performed thermal up-shift experiments on the W66F/L44A flavodoxin mutant, whose temperature of mid denaturation has been measured in the presence and absence of the selected ligands and decoys (Figures 5a, b). The holoprotein, rather than the apo form, has been used to perform the up-shift assays because, in preliminary studies, we detected binding of some of the selected compounds at the empty FMN binding site of the apo form (not shown). In order to get finer temperature up-shift values, the wild type flavodoxin has been used as a control. Since it lacks suitable buried cavities, any T_m up-shift detected in wild type flavodoxin should be attributed to an unexpected binding on the protein surface. Since such a binding would most likely occur also in the W66F/L44A mutant, relative up-shifts values (up-shift in cavity mutant minus up-shift in wild type) are reported, that allow discounting surface binding effects. Relative up-shift values are also advantageous because they can additionally subtract any solvation effect on protein stability that could arise from the very presence of some of the ligands in the solution.

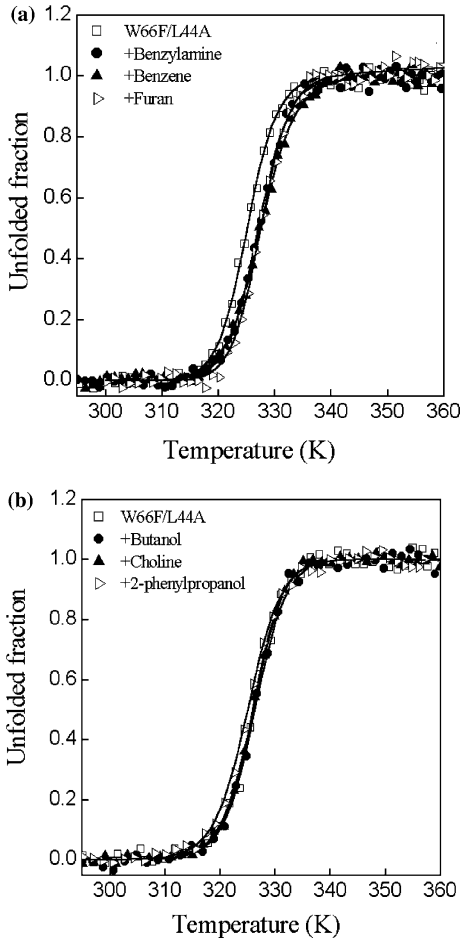

Figure 5. Detection of ligand binding to W66F/L44A holoflavodoxin by thermal up-shift (increase of T_m) followed by far-UV CD, in 10 mM phosphate buffer, pH 7.0. (a) W66F/L44A alone and in the presence of three candidate binders. (b) W66F/L44A alone and in the presence of three expected non-binders. Protein and ligand concentrations were 2 μM and 10 mM, respectively. Continuous lines represent the data fit to Equation 11 (*Methods*).

The thermal up-shifts measured for the predicted binders and non-binders are reported in Table 8. The up-shifts reported for the wild type control protein (devoid of a suitable cavity) are all very small, often smaller than or very close to the standard errors. This probably indicates that the ligands tested do not bind to the flavodoxin protein surface, nor they significantly modify the stability of the protein by their interactions with the solvent (only one binder, toluene, and one non-binder, choline, display thermal up-shifts for the wild type protein, that are slightly above the standard errors). A *t*-Student test performed on

Table 8. Thermal up-shift assays of ligand binding to the W66F/L44A holoflavodoxin cavity.

Ligand	Ranking in step 2[a]	ΔG_b in step 3[b]	Absolute thermal up-shift				Relative upshift
			Wild type		W66F/L44A		
			T_m[c]	ΔT_m[d]	T_m[c]	ΔT_m[d]	ΔT_m[e]
No ligand			340.9 ± 0.40	–	325.2 ± 0.10	–	–
Predicted binders[f]							
Benzene	355	−7.5	340.85 ± 0.35	−0.05 ± 0.53	327.35 ± 0.15	2.15 ± 0.18	2.20 ± 0.56
Benzylamine	66	−6.5	341.30 ± 0.10	0.40 ± 0.41	327.80 ± 0.70	2.60 ± 0.71	2.20 ± 0.82
Furan	435	−5.6	341.40 ± 0.50	0.50 ± 0.64	327.60 ± 0.10	2.40 ± 0.14	1.90 ± 0.66
Pyridine	324	−6.7	341.20 ± 0.20	0.30 ± 0.45	327.10 ± 0.10	1.90 ± 0.14	1.60 ± 0.47
Toluene	585	−7.4	341.70 ± 0.40	0.80 ± 0.57	327.05 ± 0.15	1.85 ± 0.18	1.05 ± 0.60
Predicted non-binders[g]							
2-phenyl-propanol	4895	−3.5	340.15 ± 0.45	−0.75 ± 0.60	325.05 ± 0.05	−0.15 ± 0.11	0.60 ± 0.61
Butanol	2312	−3.0	341.55 ± 0.25	0.65 ± 0.47	326.45 ± 0.05	1.25 ± 0.11	0.60 ± 0.48
Choline	2803	N.a.	342.25 ± 0.05	1.35 ± 0.40	325.80 ± 0.10	0.60 ± 0.14	−0.75 ± 0.42

[a]Ranking after step 2 of the NCI database screening.
[b]Binding-free energy calculated using Equation 9. N.a. identifies a ligand that did not remained in the cavity after the fine docking (step 3).
[c]Apparent temperatures of mid-denaturation (in K) calculated by fitting of thermal unfolding curves followed by far-UV CD to a two-state equation. Means of two experiments ± SE are reported.
[d]Absolute thermal up-shifts calculated as the difference between the T_m of the protein in the presence of a 10 mM ligand concentration and in its absence. SE propagated from those of the T_m*s*.
[e]Relative thermal up-shifts calculated as the thermal up-shift observed in the cavity containing mutant (W66F/L44A flavodoxin) minus that observed in the wild type protein. In relative thermal up-shifts, potential effects of the ligand on protein stability exerted through a modification of water properties or through unexpected binding to the protein surface are subtracted out. SE propagated from those of the individual thermal up-shifts of the two proteins.
[f]Ligands found in high positions in the hit list for binding to the W66F/L44A cavity at the end of the second step of the screening and that have successfully passed the third step with high calculated binding energy. They have been selected for spectroscopic convenience.
[g]Ligands found in low positions in the hit list for binding to the W66F/L44A cavity at the end of the second step of the screening and/or that have not passed the third step with high calculated binding-free energy. They have been additionally selected for spectroscopic convenience.

the wild type data clearly fails to distinguish between the binders and the non-binders ($p \leq 0.96$). In contrast, the thermal up-shifts of the W66F/L44A cavity flavodoxin mutant show significant differences between the binders and the non-binders (Figures 5a, b). While the non-binders up-shifts are low, the up-shifts determined for the five binders tested are all above or close to 2 °C (2.2 °C degrees in average), which is similar to the thermal up-shift values that have been reported for other proteins [25, 65, 68]. The *t*-Student test clearly distinguishes the binders from the non-binders of W66F/L44A flavodoxin ($p \leq 0.008$). Since the up-shifts observed in wild type flavodoxin are small, the relative up-shifts for W66F/L44A flavodoxin are not very different from the absolute ones. For the non-binders, all the relative up-shifts are small and close to the standard errors, while for the binders the up-shifts are larger and clearly above the standard errors. A *t*-Student test performed on the relative up-shifts indicates that the two groups of ligands are clearly different ($p \leq 0.015$). In fact, each up-shift was measured twice and when the Student test was applied to the individual relative up-shifts (rather than to the averages reported in Table 8) it leaded to $p \leq 0.0001$. The thermal up-shift analysis thus clearly validates the predictions of the massive screening by showing that the five predicted ligands bind to the protein cavity (only the binding of toluene could be questioned due to the small relative up-shift), while none of the three predicted non-binders do. This means that the three-step massive screening method developed and computationally tested on existing T4 lysozyme mutants has been successful in selecting binders of a different, non-previously characterised protein that was purposely designed and engineered to

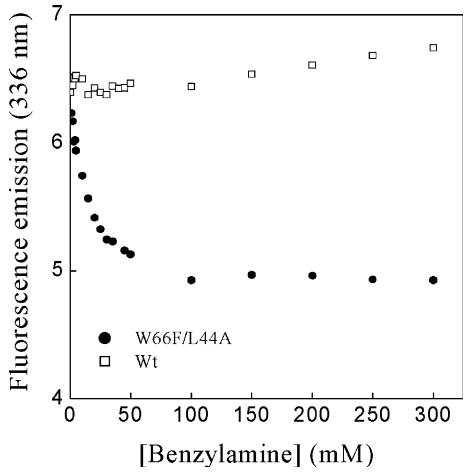

Figure 6. Titration of W66F/L44A emission fluorescence (at 336 nm) with benzylamine (filled circles). As a control, the fluorescence of wild type holoflavodoxin in the presence of benzylamine is also shown (open squares). The experiment was performed incubating the protein with ligand solutions of different concentrations for 14 h in darkness before the fluorescence was recorded at 25.0 ± 0.1 °C.

test the portability of the method. We notice that, among the binders of the flavodoxin cavity, there are also polar compounds (such as benzylamine), which illustrates that the screening method goes beyond performing a mere discrimination based in polarity.

As an additional confirmation of the binding of the flavodoxin ligands selected by the screening method, we have performed spectroscopic titrations of intrinsic protein fluorescence. Clear saturation curves are obtained for the W66F/L44A flavodoxin mutant when it is incubated with increasing concentrations of benzene, benzylamine or furan (the benzylamine titration in shown in Figure 6), which are in contrasts with the small and not saturating quenching effects produced on the wild type protein. From the titration curves we calculate, using Equation 13, that the dissociation constants of the W66F/L44A flavodoxin complexes with benzene, benzylamine and furan are of 43, 9, and 36 mM, respectively. We notice that these dissociation constants correspond to binding-free energies of −2.0, −2.8, and −1.8 kcal mol^{-1}, for benzene, benzylamine and furan, respectively, which are clearly lower than those calculated using Equation 9 (Table 8). This means that this theoretical equation, derived solely from an analysis of the binding of ligands to a cavity in the L99A lysozyme mutant, is able to discriminate between binders and non-binders in other protein cavities, but needs refining in order to quantitatively reproduce the affinity of the complexes.

Could it be done the other way round? Protein suits for chosen ligands by reverse docking

A three-step computational method for screening large small-molecule databases to identify ligands of natural or engineered protein cavities has been developed. The method is portable and has been successfully used to identify binders of a model cavity newly engineered in flavodoxin. We notice that applying the method in a reverse way could be particularly interesting. To that end, libraries of cavity proteins should be developed. Probably an efficient way to do it would be performing massive *in silico* point mutations (one by one, or combining two or three mutations) in the selected proteins (ideally a PDB subset of proteins that can be recombinantly expressed). Then, the structure of the cavities so generated should be modelled (see [24] for a method specifically designed for the fast modelling of protein cavities from wild type structures). Finally the three-step docking method would be performed to dock a pre-selected ligand of interest into the cavity library in order to find a suitable mutant host protein that would be then engineered, expressed and tested. In this way, ligands of moderate size could be hosted in proteins that, after appropriate tailoring, could be used to transport and deliver the ligand *in vivo*.

Acknowledgement

We acknowledge financial support from grants BFU2004–01411/BMC (DGI, Spain), BFM 2002–00113 DGES (Spain), and by the Aragon Government (DGA – Group of Non Linear and Statistical Physics). CM was supported by an AECI fellowship, JLLL by a fellowship from the Bask Government, and SCL and MB by MEC fellowships. We thank Dr. E. Freire for discussions.

References

1. Abagyan, R. and Totrov, M., Curr. Opin. Chem. Biol., 5 (2001) 375.

2. Bamborough, P. and Cohen, F.E., Curr. Opin. Struct. Biol., 6 (1996) 236.
3. Morris, G.M., Goodsel, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K and Olson, A.J., J. Comput. Chem., 19 (1998) 1639.
4. Totrov, M. and Abagyan, R., Proteins (Suppl.), 1 (1997) 215.
5. McKie, J.H., Douglas, K.T., Chan, C., Roser, S.A., Yates, R., Read, M., Hyde, J.E., Dascombe, M.J., Yuthavong, Y. and Sirawaraporn, W., J. Med. Chem., 41 (1998) 1367.
6. Iyer, S., Kontoyiannis, D., Chevrier, D., Woo, J., Mori, N., Cornejo, M., Kollias, G. and Buelow, R., J. Biol. Chem., 275 (2000) 17051.
7. Gruneberg, S., Stubbs, M.T. and Klebe, G., J. Med. Chem., 45 (2002) 3588.
8. Schapira, M., Abagyan, R. and Totrov, M., J. Med. Chem., 46 (2003) 3045.
9. vonItzstein, M., Dyason, J.C., Oliver, S.W., White, H.F., Wu, W.Y., Kok, G.B. and Pegg, M.S., J. Med. Chem., 39 (1996) 388.
10. PerrierV., , Wallace, A.C., Kaneko, K., Safar, J., Prusiner, S.B. and Cohen, F.E., Proc. Natl. Acad. Sci. U.S.A., 97 (2000) 6073.
11. Khan, M.O.F., Austin, S.E., Chan, C., Yin, H., Marks, D., Vaghjiani, S.N., Kendrick, H., Yardley, V., Croft, S.L. and Douglas, K.T., J. Med. Chem., 43 (2001) 3148.
12. Pavao, F., Castilho, M.S., Pupo, M.T., Dias, R.L., Correa, A.G., Fernandes, J.B., da Silva, M.F., Mafezoli, J., Vieira, P.C. and Oliva, G., FEBS Lett., 520 (2002) 13.
13. Milne, G.W.A., Nicklaus, M.C., Driscoll, J.S., Wang, S. and Zaharevitz, D.J., Chem. Inf. Comput. Sci., 34 (1994) 1219.
14. Schapira, M., Raaka, B.M., Samuels, H.H. and Abagyan, R., Proc. Natl. Acad. Sci. U.S.A., 97 (2000) 1008.
15. Gottlin, E.B., Benson, R.E., Conary, S., Antonio, B., Duke, K., Payne, E.S., Ashraf, S.S. and Christensen, D.J., J. Biomol. Screen., 8 (2003) 332.
16. Cox, B., Denyer, J.C., Binnie, A., Donnelly, M.C., Evans, B., Green, D.V., Lewis, J.A., Mander, T.H., Merritt, A.T., Valler, M.J. and Watson, S.P., Prog. Med. Chem., 37 (2000) 83.
17. Waszkowycz, B., Perkins, T.D.J., Sykes, R.A. and Li, J., IBM Syst. J., 40 (2001) 360.
18. Pickett, S.D., Sherborne, B.S., Wilkinson, T., Bennett, J., Borkakoti, N., Broadhurst, M., Hurst, D., Kilford, I., McKinnell, M. and Jones, P.S., Bioorg. Med. Chem. Lett., 13 (2003) 1691.
19. Jenkins, J.L., Kao, R.Y. and Shapiro, R., Proteins, 50 (2003) 81.
20. Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P. and Matthews, B.W, Science, 255 (1992) 178.
21. Xu, J., Baase, W.A., Baldwin, E. and Matthews, B.W., Protein Sci. , 7 (1998) 158.
22. Baldwin, E., Baase, W.A., Zhang, X.J., Feher, V. and Matthews, B.W., J. Mol. Biol., 277 (1998) 468.
23. Morton, A. and Matthews, B.W., Biochemistry, 34 (1995) 8576.
24. Machicado, C., Bueno, M. and Sancho, J., Prot. Eng., 15 (2002) 669.
25. Morton, A., Baase, W.A. and Matthews, B.W., Biochemistry, 34 (1995) 8564.
26. Dauber-Ogusthorpe, P., Roberts, V.A., Ogusthorpe, D.J., Wolff, J., Genest, M., and Hagler, A.T., Proteins Struct. Funct. Genet., 4 (1998) 31.
27. Pedretti, A., Villa, L. and Vistoli, G., J. Mol. Graph., 21 (2002) 47.
28. Ghose, A.K. and Crippen, G.M., J. Chem. Inf. Comput. Sci., 27 (1987) 21.
29. Hill, T.L. (Ed.) Introduction to Statistical Thermodynamics. Addison-Wesley, New-York, 1960.
30. Gasteiger, J. and Marsali, M., Tetrahedron, 36 (1980) 3219.
31. Mayo, S.L., Olafson, B.D. and Goddard, W.A. III, J. Phys. Chem., 94 (1990) 8897.
32. Maple, J.R., Hwang, M.J., Jalkanen, K.J., Stockæsch, T.P. and Hagler, A.T., J. Comput. Chem., 19 (1998) 430.
33. Hagler, A.T., Huler, E. and Lifson, S., J. Am. Chem. Soc., 96 (1974) 5319.
34. Venkatachalam, C.M., Jiang, X., Oldfield, T. and Waldman, M.J., Mol. Graphics Modell., 21 (2003) 289.
35. Gehlhaar, D.K., Bouzida, D. and Rejto, P.A. (Ed) Rational Drug Design: Novel Methodology and Practical Applications. American Chemical Society, Washington, D.C, 1999, pp. 292–311.
36. Jain, A.M., J. Comput. Aided Mol. Design, 10 (1996) 427.
37. Muegge, I. and Martin, Y.C., J. Med. Chem., 42 (1999) 791.
38. Charifson, P.S., Corkery, J.J., Murcko, M.A. and Walters, W.P., J. Med. Chem., 42 (1999) 5100.
39. Luty, B.A., Wasserman, Z.R., Stouten, P.F.W., Hodge, C.N., Zacharias, M. and McCammon, J.A., J. Comput. Chem., 16 (1995) 454.
40. Stouten, P.F.W., Frommel, C., Nakamura, H. and Sander, C., Mol. Simul., 10 (1993) 97.
41. Li, Z. and Scheraga, H.A., Proc. Natl. Acad. Sci. U.S.A., 84 (1987) 6611.
42. Guex, N. and Peitsch, M.C., Protein Data Bank Quat. Newslett., (1996) 77.
43. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M. Jr., Spellmeyer, D.C., Fo, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179.
44. Florián, J. and Warshel, A., J. Phys. Chem., B 101 (1997) 5583.
45. Rashin, A.A. and Namboodiri, K., J. Phys. Chem., 91 (1987) 6003.
46. Connolly, M.L., J. Appl. Cryst., 16 (1983) 548.
47. Finkelstein, A. and Janin, J., Protein Eng., 3 (1989) 1.
48. Jain, A.V., J. Med. Chem., 46 (2003) 499.
49. Novotny, J., Bruccoleri, R.E. and Saul, F.A., Biochemistry, 28 (1989) 4735.
50. Erickson, H.P., J. Mol. Biol., 206 (1989) 465.
51. Horton, N. and Lewis, M., Protein Sci., 1 (1992) 169.
52. D'Aquino, J.A., Gomez, J., Hilser, V.J., Lee, K.H., Amzel, L.M. and Freire, E., Proteins Struct. Funct. Genet., 25 (1996) 143.
53. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
54. Eisenberg, D. and McLachlan, A.D., Nature, 319 (1989) 199.
55. Wesson, L. and Eisenberg, D., Protein Sci., 2 (1992) 227.
56. Fraczkiewicz, R. and Braun, W., J. Comput. Chem., 19 (1998) 319.
57. Sanger, F., Nicklen, S. and Coulson, A.R., Proc. Natl. Acad. Sci. U.S.A., 74 (1977) 5463.

58. Genzor, C.G., Beldarrain, A., Gomez-Moreno, C., Lopez-Lacomba, J.L., Cortijo, M. and Sancho, J., Protein Sci., 5 (1996) 1376.
59. Edmonson, D.E. and Tollin, G., Biochemistry, 10 (1971) 124.
60. Irun, M.P., Maldonado, S. and Sancho, J., Protein. Eng., 14 (2001) 173.
61. Pace , C.N., Shirley, B.A., and Thompson J.A. (Eds), Protein Structure, a Practical Approach, IRL Press, Oxford, 1989.
62. Irun, M.P., Garcia-Mira, M.M., Sanchez-Ruiz, J.M. and Sancho, J., J. Mol. Biol., 306 (2001) 877.
63. Campos, L.A., Bueno, M., Lopez-Llano, J., Jimenez, M.A. and Sancho, J., J. Mol. Biol., 344 (2004) 239.
64. Schellman, J.A., Biopolymers, 15 (1976) 999.
65. Pantoliano, M., Petrella, E., Kwasnoski, J., Lobanov, V., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B., Lane, P. and Salemme, F., J. Biomol. Screen., 6 (2001) 429.
66. Lostao, A., Gomez-Moreno, C., Mayhew, S.G. and Sancho, J., Biochemistry, 36 (1997) 14334.
67. Su, A.I., Lorber, D.M., Weston, G.S., Baase, W.A., Matthews, B.W. and Shoichet, B.K., Proteins Struct. Funct. Genet., 42 (2001) 279.
68. Wei, B.Q., Baase, W.A., Weaver, L.H., Matthews, B.W. and Shoichet, B.K., J. Mol. Biol., 322 (2002) 339.
69. Wei, B.Q., Weaver, L.H., Ferrari, A.M., Matthews, B.W. and Shoichet, B.K., J. Mol. Biol., 337 (2004) 1161.
70. Akaho, E., Morris, G.M., Goodsell, D.S., Wong, D. and Olson, A.J., J. Chem. Sofware, 7 (2001) 103.
71. Apostolakis, J., Pluckthun, A. and Caflisch, A., J. Comput. Chem., 19 (1998) 21.
72. Holloway, K., Wai, J., Halgren, T., Fitzgerald, P., Vacca, J., Dorsey, B., Levin, R., Thompson, W., Chen, L., deSolms, S., Gaffin, N., Ghosh, A., Giuliani, E., Graham, S., Guare, J., Hungate, R., Lyle, T., Sanders, W., Tucker, T., Wiggins, M., Wiscount, C., Woltersdorf, O., Young, S., Darke, P. and Zugay, J., J. Med. Chem., 38 (1995) 305.
73. Rarey, M., Kramer, B. and Lengauer, T., Proteins (Suppl), 1 (1997) 221.
74. Sudbeck, E.A., Mao, C., Vig, R., Venkatachalam, T.K., Tuel-Ahlgren, L. and Uckun, F.M., Antimicrob. Agents Chemother., 42 (1998) 3225.
75. Rao, S.T., Shaffie, F., Yu, C., Satyshur, K.A., Stockman, B.J., Markley, J.L. and Sundarlingam, M., Protein Sci., 11 (1992) 1413.
76. Buckle, A.M., Cramer, P. and Fersht, A.R., Biochemistry, 35 (1996) 4298.
77. Lo, MC, Aulabaugh, A, Jin, G, Cowling, R, Bard, J, Malamas, M and Ellestad, G., Anal. Biochem., 332 (2004) 153.