Molecular anchors with large stability gaps ensure linear binding free energy relationships for hydrophobic substituents

Paul A. Rejto and Gennady M. Verkhivker
Agouron Pharmaceuticals, Inc.
3565 General Atomics Court
San Diego, CA 92121

Ligand–protein docking simulations are employed to analyze the binding energy landscape of the pipecolinyl fragment that serves as a recognition core of the FK506 ligand in binding with the FKBP12 protein. This fragment acts as a molecular anchor that specifically binds within the protein active site in a unique binding mode, in harmony with the structure of the FK506–FKBP12 complex. Molecular anchors are characterized by a large stability gap, defined to be the free energy of a ligand bound in the native binding mode relative to the free energy of alternative binding modes. For ligands that share a common anchor fragment, a linear binding free energy relationship may be expected for hydrophobic substituents provided they do not abrogate the anchor binding mode. Changes in solvent–accessible surface area for these peripheral groups are used to rationalize the relative binding affinities of a series of FKBP12–ligand complexes which share the pipecolinyl anchor fragment. A series of benzene derivatives that bind to a mutant form of T4 lysozyme is also analyzed, and implications for structure–based drug design are discussed.

1 Introduction

Bimolecular association is a critical step for many chemical and biological processes including chemical reactions, catalysis and cell signaling. Specificity and affinity are essential components of recognition and understanding their interrelation is an important problem in molecular biology and protein engineering. Binding affinity is the free energy of bound ligand–protein complexes relative to their unbound forms, while binding specificity is governed by the relative affinity of a given ligand to a particular receptor with respect to another receptor, such as a mutant form of the protein, or differences in the binding affinity of a given receptor with different ligands.

The binding free energy landscape of ligand–protein complexation is typically characterized by a multitude of energetically similar but structurally different binding modes reflecting the fact that the complex can adopt a variety of alternative structures. These alternative minima arise because there are competing interactions and, for example, it is not possible for the complex to satisfy all possible favorable ligand–protein interactions while simultaneously maintaining a low energy internal conformation for the ligand. In complex systems, energy landscapes with competing interactions are ubiquitous and are

termed 'frustrated' [1]. Consequently, a simplified model of ligand–protein interactions that can adequately reproduce differences in the energetics of native and non–native binding modes is desirable for the analysis of molecular recognition [2], even when it is unable to estimate precisely binding free energies. This approach has been used to rationalize experimental differences in receptor–specific binding to HIV–1 protease relative to HIV–2 protease [3]. It has also been used to address the origins of specificity in ternary FKBP12–FK506–CN and FKBP12–rapamycin–FRAP complexes, and the lack of receptor specificity in binary complexes of FK506 and rapamycin with wild–type FKBP12 protein relative to FKBP12 protein mutants [4].

Structural studies of FKBP12–ligand complexes [5,6,7], as well as mutant T4 lysozyme–ligand complexes [8] and streptavidin–peptide complexes [9,10,11], suggest that molecular recognition may be fulfilled by a conserved, relatively rigid portion of the receptor active site interacting with an anchoring fragment of the ligand acting as its recognition core in the binding process. Molecular anchors are small molecules or molecular fragments that bind within the enzyme active site in a specific binding mode that is structurally stable and characterized by a pronounced gap between the free energy of the favorable binding mode and that of alternative binding modes. The stabilty gap is a signature of a minimally frustrated binding energy landscape that has a dominant free energy minimum. This gap ensures thermodynamic stability of the favorable binding mode and has been shown to be an important prerequisite for specificity of ligand–protein recognition [3].

Criteria which distinguish known molecular anchors from random molecular fragments were established in computational studies of molecular recognition with the FKBP12 protein [12] and streptavidin [13] based on a statistical analysis of the binding energy landscapes. Structural harmony arises when the bound structure of the molecular anchor shares the same binding mode even when embedded in larger ligands that incorporate this motif. Structural harmony and the structural consensus of molecular anchors in multiple docking simulations can be used to characterize the overall shape of the binding energy landscape and distinguish these recognition motifs from non–specific molecular fragments [12]. These features of the energy landscape reflect a more general minimum frustration principle [1] and are reminiscent of structural harmony in proteins, where local interactions that stabilize secondary structures must be consistent with tertiary interactions that stabilize complete protein folds [14].

Molecular anchors, while responsible for specific binding, are usually not sufficient to provide tight binding unless they are incorporated in a larger ligand. When the relative energy of the native binding mode compared to alternative modes, the stability gap, is localized in an anchor fragment, ap-

propriate hydrophobic modifications of peripheral ligand groups may confer additional binding affinity. If the peripheral group alters the binding mode of the molecular anchor, estimation of the free energy of binding is difficult. However, when the binding mode is preserved, the ligand can be subdivided into two quasi–independent subsystems representing the anchor fragment and the peripheral groups. Estimation of binding free energy differences in this case is greatly facilitated.

In this study, ligand–protein docking simulations are performed to explore the binding energy landscape of the pipecolinyl moiety, a putative molecular anchor in FKBP12. In addition, derivatives of this compound are analyzed by docking simulations and are shown to have a high degree of structural harmony with a common binding mode for the molecular anchor. The binding free energy of these compounds is predicted to correlate linearly with changes in the solvent–accessible surface area of peripheral hydrophobic substituents. The affinity of a series of compounds that bind to a mutant T4 lysozyme is also analyzed. The structures of these complexes have been determined by X–ray crystallography [15]; one set of derivatives has structural harmony and obey a linear binding free energy relationship, while a second set does not exhibit structural harmony and does not follow a linear binding free energy relationship.

2 Molecular docking

The structure of the FKBP12 protein used for these simulations was obtained from a crystal structure of FKBP12 bound with the FK506 ligand [5]. Ligand conformations and orientations are searched by a simulated evolution algorithm in a parallelepiped that encompasses the binding site obtained from the crystallographic structure of the FKBP12–FK506 complex with a 2.0 Å cushion added to every side of this box [16]. A constant energy penalty of 200.0 kcal/mole is added to every ligand atom outside the box, but no assumptions regarding either favorable ligand orientations or any specific ligand–protein interactions are made. The protein conformation is held fixed and all crystallographic water molecules are included as a part of the protein structure. The structure with the lowest energy obtained from multiple docking simulations defines the predicted structure for the FKBP12–ligand complex. Bonds linking sp^3 hybridized atoms to either sp^3 or sp^2 hybridized atoms and non–conjugated single bonds linking two sp^2 hybridized atoms are allowed to rotate, while bond distances, bond angles and all other torsional angles are fixed during the docking simulations.

The molecular recognition model used in this study includes both in-

Figure 1: a) The function form of the ligand–protein interaction energy. b) The hydrogen bond interaction energy is multiplied by the hydrogen bond strength term, which is a function of the angle θ determined by the relative orientation of the protein and ligand atoms. The range of θ is between 0 and 180°. c) A protein donor atom D bound to one hydrogen atom H makes an angle θ with the ligand atom L. d) A protein donor atom D bound to two hydrogen atoms H makes an angle θ with the ligand atom L. e) A protein acceptor atom A makes an angle θ with the ligand atom L.

Table 1: Interaction types for various protein and ligand atom types. Primary and secondary amines are defined to be donors while oxygen and nitrogen atoms with no bound hydrogens are defined to be acceptors. Crystallographic water molecules and hydroxyl groups are defined to be both donor and acceptor, and carbon atoms are defined to be nonpolar.

	Protein			
Ligand	*Donor*	*Acceptor*	*Both*	*Nonpolar*
Donor	Repulsive	H–bond	H–bond	Steric
Acceptor	H–bond	Repulsive	H–bond	Steric
Both	H–bond	H–bond	H–bond	Steric
Nonpolar	Steric	Steric	Steric	Steric

Table 2: Parameters for the three different ligand–protein interactions. The hydrogen bond function is multipied by the three–body hydrogen bond strength term (Fig. 1). Parameters C and D are not relevant for the repulsive interaction term, and are indicated by dashes. The units of A, B, C, and D are Å. The units of E, F, and G are kcal/mole.

Interaction	A	B	C	D	E	F	G
Steric	3.3	3.6	4.5	5.4	–0.1	0.0	20.0
H–bond	2.3	2.6	3.1	3.4	–4.0	0.0	20.0
Repulsive	3.2	6.0	—	—	0.0	2.0	20.0

tramolecular energy terms for the ligand, given by torsional and non–bonded functions [17], and simplified intermolecular ligand–protein interaction terms. The intermolecular terms represent steric interactions, favorable hydrogen bond interactions between donors and acceptors, and the repulsive interaction between pairs of donors or acceptors. The parameters of the energy function were obtained by requiring the crystallographic conformation of a set of ligand–protein complexes to be the global potential energy minimum [16]. These functions are designed for structure prediction of ligand–protein complexes, and are not intended to be complete force fields. The contributions are calculated from a piecewise linear potential summed over all protein and ligand heavy atoms, together with an angular dependence for the hydrogen bond interaction (Fig. 1). Hydrogen atoms are not included in the calculation. The parameters of the pairwise potential depend on the four different atom types: hydrogen–bond donor, hydrogen–bond acceptor, both donor and acceptor, and nonpolar (Table 1). All three potentials have the same functional form with different parameters (Table 2).

3 Binding affinity and molecular anchors

The binding free energy of an enzymatic protein P with a ligand L is given by

$$\Delta G = -k_B T \ln \left(\frac{[LP]}{[L][P]} \right),$$

where $[L]$, $[P]$, and $[LP]$ are the concentrations of unbound ligand, unbound protein, and bound ligand–protein complex, respectively. Estimating binding free energy difference due to chemical modification of a ligand, $\delta\Delta G$ requires significant computational effort and adequate sampling of the entire configurational space available to the system. When a chemical modification affects the recognition motif of the ligand, substantial changes in the resulting energy spectrum are expected and the effect of such a chemical modification on the binding energetics is complicated. The binding free energy assessment is simplified considerably if the binding mode of the recognition motif is structurally stable and a series of ligands which contain this molecular anchor along with additional peripheral group retain the binding mode favorable for the molecular anchor. A necessary condition for this structural stability within the active site is that the stability gap $\Delta_{anchor} \gg k_B T$, where

$$\Delta_{anchor} = -k_B T \ln \left(\frac{[AP_{non-native}]}{[AP_{native}]} \right)$$

governs the free energy of the native binding mode of the anchor, AP_{native}, relative to other, non–native binding modes, $AP_{non-native}$. If the stability gap Δ_{anchor} is not large, then the bound form of the molecular anchor is not dominated by the native binding mode, and addition of peripheral groups can easily alter the binding mode of the complete ligand. Provided that this stability gap is sufficiently large and that peripheral groups do not abrogate the native binding mode, the free energy ΔG may be written as the sum of one term due to the anchor and a second due to the peripheral groups. If the ligand peripheral groups are hydrophobic and do not form specific interactions within the binding site, one expects their contribution to the total free energy of inhibition to scale with the solvent–accessible surface area σ of the peripheral group, and

$$\Delta G \approx \Delta G_{anchor} + \Delta G_{peripheral} \approx \Delta G_{anchor} + \alpha^{protein}_{aqueous} \sigma \qquad (1)$$

The coefficient $\alpha^{protein}_{aqueous}$ depends only on the protein environment within the native binding site. For a series of compounds where the native binding mode is unaffected by the peripheral group and which differ only in the number of hydrophobic atoms, then using Eq. 1,

$$\delta\Delta G \approx \alpha^{protein}_{aqueous}\delta\sigma.$$

The binding process may be decomposed into two parts[8]: the first corresponding to transfer from aqueous solution to cyclohexane, followed by transfer from cyclohexane to the protein binding site, so

$$\alpha^{protein}_{aqueous} = \alpha^{cyclohexane}_{aqueous} + \alpha^{protein}_{cyclohexane}. \quad (2)$$

Peripheral ligand groups typically interact with flexible regions of the protein[12], and one may anticipate that the binding site resembles a time–averaged cyclohexane environment[8]. In this case, $\alpha^{protein}_{cyclohexane} \approx 0$, and so $\alpha^{protein}_{aqueous} \approx \alpha^{cyclohexane}_{aqueous}$.

3.1 FKBP12–ligand complexes

The binding affinity for a series of FKBP12–ligand complexes that share a common core, the pipecolinyl fragment, has been measured experimentally[18]. Importantly, there is no crystallographic information regarding the bound conformation of the ligands: the suitability of the proposed model was tested by predicting the binding affinity of the FKBP12–ligand complexes. First, the stability gap of the pipecolinyl fragment, which contains 14 heavy atoms (Fig. 2), was studied by performing 100 docking simulations for the bare anchor. The docking simulations reveal a single favorable binding mode (Fig. 3), within 1.5 Å of the location of the pipecolinyl fragment in the FKBP12–FK506 complex[5]. There are few alternate binding modes, all of which have ligand–protein interaction energies considerably higher than that of the native binding mode. Hence, this fragment appears to have a large stability gap with $\Delta_{anchor} \gg k_B T$.

To investigate the ability of the molecular anchor to retain its binding mode when embedded in ligands containing hydrophobic peripheral groups, flexible docking simulations for a series of seven derivatives of the pipecolinyl fragment were performed. The structure with the lowest energy obtained from 100 docking simulations defines the predicted structure for the FKBP12–ligand complex. The predicted structures of the bound ligands reveal a pattern consistent with the pipecolinyl fragment (Fig. 4), indicative of structural harmony in this series of related ligands. Molecular recognition of these ligands by FKBP12 is apparently controlled by specific interactions formed by the pipecolinyl anchor fragment. The stability gap, Δ_{anchor}, which governs the free energy of the native binding mode within the active site relative to alternative binding modes within the same active site, is large enough to preserve

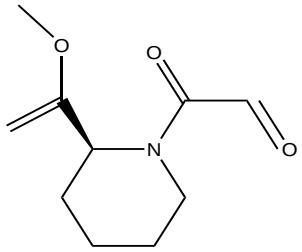

Figure 2: Structure of the 14 heavy atom pipecolinyl fragment, with hydrogens suppressed.

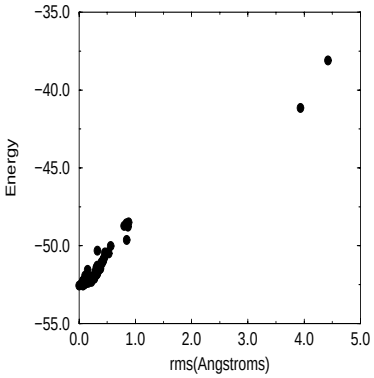

Figure 3: Energy vs. root mean square (rms) deviation from the predicted structure obtained from 100 docking simulations of the pipecolinyl fragment in FKBP12.

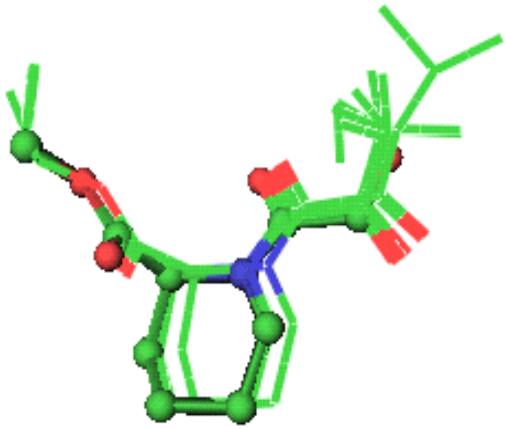

Figure 4: Predicted structures for each of six compounds (thin lines) derived from the pipecolinyl fragment (ball and stick) superimposed within the FKBP12 binding site. For clarity, hydrogens and the protein are not shown: a common binding mode for the pipecolinyl fragment is revealed.

a favorable binding mode. According to the proposed model, then, changes in solvent–accessible surface area are expected to correlate linearly with changes in binding affinity for this series of hydrophobic derivatives (Fig. 5). The experimental binding affinity for this series of inhibitors is well approximated by the effect of non–specific binding for peripheral hydrophobic groups, with $\alpha_{aqueous}^{FKBP12} = -21.7$ cal/mole/Å^2.

3.2 T4 lysozyme–ligand complexes

A similar linear free energy relationship arises in a series of ligands bound to a designed cavity of T4 lysozyme [8]. These ligands share a common benzene ring but have peripheral hydrophobic chains of different lengths. Crystallographic structure determination revealed structural harmony for these molecules [15]: the addition of up to four peripheral carbon atoms has no significant effect on the binding mode of the benzene fragment. For this series of ligands, $\alpha_{aqueous}^{lysozyme} = -11.5$ cal/mole/Å^2 (Fig. 6). In a study of solvation free energies, Simonson and Brünger [19] found that the atomic surface tension for neutral amino acids was $\alpha_{aqueous}^{cyclohexane} = -19.0 \pm 6.0$ cal/mole/Å^2, which is in agreement with experimental estimates for n–alkanes. From Eq. 2, the magnitudes of both $\alpha_{cyclohexane}^{FKBP}$ and $\alpha_{cyclohexane}^{lysozyme}$ are near zero, as expected if the protein environment for the peripheral atoms is similar to that of cyclohexane.

A second series of compounds bound to the mutant T4 lysozyme shows that binding free energies can be determined by subtle differences in the electronic structure of the ligand. The binding modes for the isosteric ligands indene, indole and benzofuran are altered relative to that of the anchor benzene molecule and do not obey a simple binding free energy relationship. Differences in the binding affinity for these compounds are not dominated by differences in their hydrophobicity, as was intended in their original design, but rather appear to be determined by differences in their interaction with the protein. Even in this series of closely related ligands, small differences in ligand composition alter the energy distribution of the binding modes and no single favorable binding mode is shared by all of these compounds. The favorable binding mode differs for each chemical modification in this series and no structural harmony is observed; as a result no linear free energy relationship is found or expected. A linear free energy relationship is only anticipated for ligands with hydrophobic substituents and a common anchor fragment that is structurally stable.

Figure 6: Changes in measured binding affinities [8] vs. changes in solvent–accessible surface area [19] for a series of six ligands that bind a mutant T4 lysozyme, relative to compound 1, with structures indicated. The coefficient $\alpha_{aqueous}^{lysozyme} = -11.5$ cal/mole/Å^2 with a linear correlation coefficient $r^2 = 0.96$.

4 Implications for structure–based drug design

This study demonstrates that computational prediction of binding affinity is considerably simplified for ligands within a series that contains a common anchor fragment. Excellent correlation was obtained using a simple binding affinity model based on changes in solvent–accessible surface area, even in the absence of explicit crystallographic information about the ligands. Given an anchoring fragment with a significant stability gap Δ_{anchor}, peripheral hydrophobic groups that do not abrogate the initial binding mode lead to ligands with improved binding affinity and which may be described by a simple linear free energy relationship. For such ligands, precise knowledge of interaction energies may therefore not be necessary since transfer free energies between cyclohexane and water could serve as a good estimate of the binding free energies. For polar substituents, the presence of an anchor fragment will still simplify the computational prediction of binding affinity, although the linear correlation between changes in solvent–accessible surface area and changes in binding free energy shown in this work would no longer be expected.

A lead molecule with a large stability gap for a native binding mode is expected to be specific for a given protein active site and consequently is amenable to further optimization. Chemical modifications of the peripheral groups can confer additional affinity in a series of related compounds where the native binding mode is preserved. When searching for a new drug lead, molecules with anchor fragments that are characterized by large stability gaps for the favored binding mode but possibly modest binding affinity may be preferable to molecules with a number of isoenergetic and structurally different binding modes but relatively high binding affinity. This concept has been exploited in the technique called SAR by NMR [20], where small molecules that bind to a specific site in a receptor with low affinity are identified. Then, in the presence of saturating concentrations of this small molecule, a second small molecule is identified that binds to a nearby site. These two small molecules are linked to form a single ligand, which can bind with significantly improved binding affinity [21]. The identification of small anchor molecules may also be useful for a receptor–biased combinatorial chemistry that begins with computational discovery of anchor fragments. Subsequent production of chemical libraries that employ a diversity of peripheral groups to interact with flexible portions of the active site may be an efficient method for the generation of novel drug leads.

References

1. J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes, *Proteins* **21**, 167 (1995).
2. J. Cherfils and J. Janin, *Curr. Opin. Struct. Biol.* **3**, 265 (1993).
3. G.M. Verkhivker and P.A. Rejto, *Proc. Natl. Acad. Sci. USA* **93**, 60 (1996).
4. P.A. Rejto and G.M. Verkhivker, *Proteins* **28**, 313 (1997).
5. G.D. Van Duyne, R.F. Standaert, P.A. Karplus, S.L. Schreiber and J. Clardy, *Science* **252**, 839 (1991).
6. K.P. Wilson *et al*, *Acta Crystallogr.* D **51**, 511 (1995).
7. R.E. Babine *et al*, *Bioorg. Med. Chem. Lett.* **5**, 1719 (1995).
8. A. Morton and B.W. Matthews, *Biochem.* **34**, 8576 (1995).
9. P.C. Weber, M.W. Pantoliano and L.D. Thompson, *Biochem.* **31**, 9350 (1992).
10. B.A. Katz, *Biochem.* **34**, 15421 (1995).
11. T.G.M. Schmidt, J. Koepke, R. Frank and A. Skerra, *J. Mol. Biol.* **255**, 753 (1996).
12. P.A. Rejto and G.M. Verkhivker, *Proc. Natl. Acad. Sci. USA* **93**, 8945 (1996).
13. N.K. Shah, P.A. Rejto and G.M. Verkhivker, *Proteins* **28**, 421 (1997).
14. N. Gō, *Annu. Rev. Biophys. Bioeng.* **12**, 183 (1983).
15. A. Morton, W.A. Baase and B.W. Matthews, *Biochem.* **34**, 8564 (1995).
16. D.K. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel and S.T. Freer, *Chem. Biol.* **2**, 317 (1995).
17. S.L. Mayo, B.D. Olafson and W.A. Goddard, III, *J. Phys. Chem.* **94**, 8897 (1990).
18. D.A. Holt *et al*, *Biorg. Med. Chem. Lett.* **4**, 315 (1994).
19. T. Simonson and A.T. Brünger, *J. Phys. Chem.* **98**, 4683 (1994).
20. S.B. Shuker, P.J. Hajduk, R.P. Meadows and S.W. Fesik, *Science* **274**, 1531 (1996).
21. P.J. Hajduk *et al, J. Am. Chem. Soc.* **119**, 5818 (1997).