Theor Chem Acc (1999) 101:138–142
DOI 10.1007/s00214980m111

Theoretical
Chemistry Accounts


*Regular article*

# Examining ligand-protein interactions with binding-energy landscapes*

**Paul A. Rejto, Djamal Bouzida, Gennady M. Verkhivker**

Agouron Pharmaceuticals, Inc., 3301 North Torrey Pines Court, La Jolla, CA 92037, USA

Received: 5 May 1998 / Accepted: 3 September 1998 / Published online: 17 December 1998

**Abstract.** Binding-energy landscapes are used to investigate the thermodynamics of molecular recognition for the pteridine ring, a recognition anchor in binding with dihydrofolate reductase, and two molecules with the same shape but different heteroatom substitutions. The relative importance of hydrogen bonding and hydrophobic interactions in this system is analyzed by comparing these three different decorations of the pteridine scaffold.

**Key words:** Ligand-protein binding – Energy landscapes – Molecular recognition – Molecular anchor

## 1 Introduction

Structure-based drug design of enzyme inhibitors, where the atomic structure of the binding site is used to guide the development of novel compounds, has resulted in several effective therapies including a number of HIV-1 protease inhibitors [1]. While computational methods play an important role in this process, structure-based drug design continues to rely on intuition and serendipity. In part, difficulties arise because the mechanisms of molecular recognition are not well understood and principles to guide the drug discovery process remain elusive. The goal of this work is to begin to unravel some of the critical features of tight-binding ligand-protein complexes and to provide a rationale for a structure-based drug design strategy. It is not an attempt to accurately reproduce the binding energetics of a specific ligand-protein complex, as in methods based on detailed force field models [2], but is rather an attempt to focus on general principles of molecular recognition. In this sense, our approach is similar to that in lattice models of protein folding [3] where it is hoped that the simplified model retains enough physics that the results are meaningful. An advantage of this approach is that extraneous detail is eliminated and a complete statistical mechanical analysis can be performed, making universal aspects of the phenomenon more apparent.

The assertion that there is a fundamental relationship between protein folding and molecular recognition mechanisms has been put forward recently in the statistical mechanical analysis of binding-energy landscapes [4]. Computational studies of FKBP12-ligand complexes [5] and streptavidin-peptide complexes [6] suggest that specific recognition may be fulfilled by a relatively rigid, evolutionarily conserved portion of the receptor active site interacting with a core fragment of the ligand. This core acts as a receptor-specific recognition nucleus in the binding process, a molecular anchor. Two criteria which distinguish known recognition motifs from random fragments were established; structural consensus, where a single binding mode is highly favored, and structural harmony, where the bound structure of the recognition core is consistent with its conformation when embedded in larger ligands that incorporate this motif.

In this work, we examine three closely related fragments of methotrexate (MTX), an antifolate that is a high-affinity inhibitor of dihydrofolate reductase (DHFR), and one of the earliest ligand-protein complexes whose crystal structure was determined at high-resolution [7]. MTX has three major components: a pteridine ring, a *p*-aminobenzoyl group, and a glutamate portion. Of these three substructures, the glutamate is dominated by strong electrostatic interactions with the guanidinium group of an arginine residue. By contrast, the pteridine ring is the most buried and forms more than half of the hydrogen bonds with DHFR.

In order to assess the relative contribution of steric and hydrogen-bonding interactions, we analyze three variants of the pteridine ring with a simplified energy model that has been developed for predicting the structure of ligand-protein complexes [8]. The three fragments are 1. The pteridine ring. 2. An all-carbon variant, which has approximately the same shape as the pteridine ring

* Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

*Correspondence to*: P.A. Rejto
e-mail: Rejto@agouron.com

but no ability to form hydrogen bonds. 3. An *anti*-pteridine ring, where hydrogen-bond donors in the pteridine ring are replaced with hydrogen-bond acceptors, and acceptors in the pteridine ring with hydrogen-bond donors. We compute the temperature-dependent free energy for the three molecules with the weighted histogram analysis method (WHAM) [9] as a function of the root mean square (RMS) deviation from the crystallographic conformation of the pteridine ring. Analysis of the resulting binding-energy landscapes characterizes the relative importance of nonspecific hydrophobic interactions and specific hydrogen bonds, and by comparing the landscapes, we can identify features that are important for molecular recognition.

## 2 Methods

### *2.1 Energy function*

The energy function contains an intermolecular term between the ligand and the protein binding site and an intramolecular term for the ligand, which consists of the van der Waals and torsional strain terms of the Dreiding force field [10]. The protein is held fixed during all simulations. The intermolecular term has three interaction types, a hydrogen-bonding interaction between donors and acceptors, a repulsive interaction for donor-donor and acceptor-acceptor contacts, and a generic dispersion term for nonpolar interactions [11]. Each pair of interacting atoms is assigned one type of interaction, whose parameters were derived from the distances found in high-quality crystal structures and then optimized.

### *2.2 Binding-energy landscapes*

We perform dynamically optimized Monte Carlo simulations [12] at a series of six temperatures, $T = 300, 600, 1000, 1500, 3000$, and 5000 K, and use a generalization of WHAM [9] to compute the binding-energy landscape. Based on the multiple-histogram method [13], WHAM optimally combines data from equilibrium simulations at different temperatures to estimate the density of states, which can then be used to compute properties at arbitrary temperature. After tabulating two-dimensional histograms $H_i(E,R)$ as a function of energy $E$ and RMS deviation from the crystallographic conformation $R$ during each constant-temperature equilibrium simulation $i$, the self-consistent multiple-histogram equations [13] are solved for the density of states $W(E,R)$,

$$W(E,R) = \frac{\sum_{i=1}^{M} g_i^{-1} H_i(E,R)}{\sum_{j=1}^{M} g_j^{-1} n_j \exp[-(E-F_j)/k_B T_j]}, \quad (1)$$

where

$$\exp[-F_j/k_B T_j] = \sum_E W(E) \exp[-E/k_B T_j]$$

and the density of states

$$W(E) = \sum_R W(E,R), \quad (2)$$

with $g_j = 1 + 2\tau_j$, $\tau_j$ the correlation time, $M$ the number of simulations, and $n_j$ the number of samples at $T_j$.

Using Eq. (2), the one-dimensional equations can be solved to determine the free energies $F_j$, and then used in Eq. (1) to calculate the multi-dimensional density of states $W(E,R)$. Computing the multi-dimensional density of states as a function of $E$ and $R$ requires no additional effort beyond tabulating the histogram as a function of both reaction coordinate and energy; the only complication is that more sampling is required to ensure adequate statistics. From $W(E,R)$, the potential of mean force $F(R,T)$ as a function of RMS from a reference coordinate $R_c$ and temperature is given by

$$F(R,T) = -k_B T \ln \frac{P(R,T)}{P(R_c,T)},$$

where

$$P(R,T) = \sum_E W(E,R) \exp[-E/k_B T]$$

and the native state is chosen to be the reference state, so $R_c = 0.0$. We term the potential of mean force $F(R,T)$ the binding-energy landscape.

## 3 Results

### *3.1 Methotrexate/dihydrofolate reductase*

We first examined the binding modes predicted by a docking simulation [8] for the glutamate portion of MTX, the *p*-aminobenzoyl group, as well as the pteridine ring. The glutamate portion and the pteridine ring are predicted consistently to bind in their crystallographic locations, while the *p*-aminobenzoyl group is predicted to bind to the same site that the pteridine ring binds to in the MTX/DHFR complex. Because the pteridine ring complements the active site of DHFR both in overall shape as well as the formation of six specific hydrogen bonds (Fig. 1), it combines steric and hydrogen-bonding interactions, and we focus on this substructure in this study. Docking simulations correctly

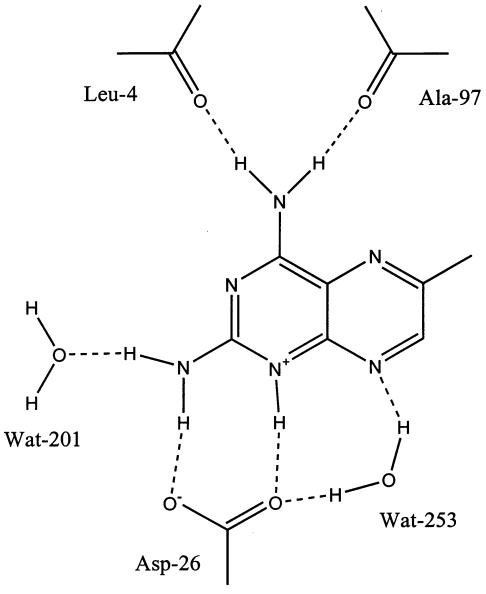


**Fig. 1.** The pteridine ring substructure of methotrexate with all polar hydrogens and the protonated N1. Hydrogen bonds formed with dihydrofolate reductase and two water molecules in the crystal structure are indicated by *dotted lines*

predict the crystallographic binding mode of this substructure, but do not provide insight into the relative importance of these terms. To address the issue, we consider the pteridine ring and two variations, an all-carbon variant, and a variant where the pteridine ring hydrogen-bond acceptors are replaced with donors, and donors with acceptors (Fig. 2). All three molecules have approximately the same shape, so a comparison of the binding-energy landscapes of the two pteridine ring variants with the standard pteridine ring can illuminate the relative importance of steric interactions and hydrogen bonding.

### 3.2 Docking simulations

There are significant differences in the predicted binding modes of the three pteridine rings. For the standard pteridine ring, three binding modes are found in the course of 100 docking simulations (Fig. 3), with the lowest energy mode corresponding to the crystallographic structure. Because the pteridine ring has a dominant predicted binding mode that is obtained with high probability, it satisfies the criterion of structural consensus. Furthermore, the predicted conformation of the bound pteridine ring is in structural harmony with its conformation when embedded in the bound structure of MTX. For the all-carbon ring, the crystallographic structure continues to be the predicted binding mode; surprisingly, the probability of predicting this mode is actually higher for the all-carbon ring than for the standard ring. Finally, the *anti*-pteridine ring has three predicted binding modes, none of which correspond to the crystallographic conformation of the ring.

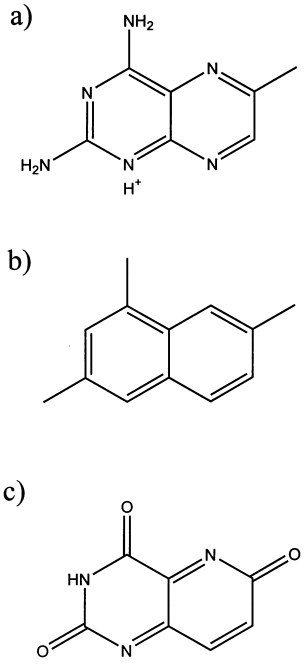


**Fig. 2a–c.** Polar hydrogen models of **a** the standard pteridine ring, **b** the all-carbon variant, and **c** the *anti*-pteridine variant. N1 on the standard pteridine ring is protonated, as suggested by crystallographic and other experimental evidence [7]

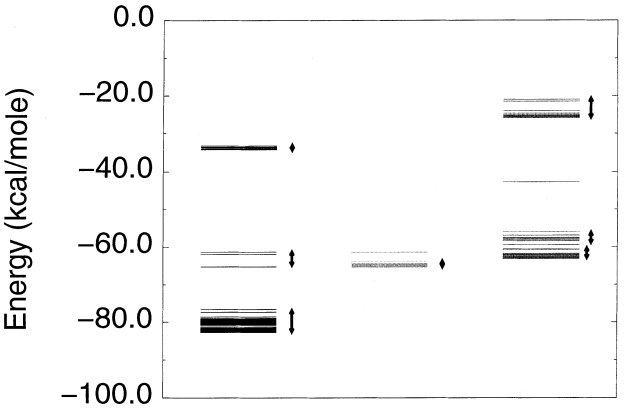


**Fig. 3.** Energy spectra obtained from docking simulations for the pteridine ring, the all-carbon variant, and the *anti*-pteridine ring. *Arrows* indicate the energies of complexes in a common binding mode. The standard pteridine ring has two high-energy binding modes in addition to the low-energy crystallographic binding mode. Only the crystallographic binding mode is found for the all-carbon variant, while the *anti*-pteridine variant has three binding modes, none of which correspond to the crystal structure

### 3.3 Binding energy landscapes

To examine the thermodynamics of the binding process, we compute binding-energy landscapes $F(R,T)$ for the three molecules. These landscapes do not measure binding affinity, but rather the relative free energy of various binding modes. As such, they determine how strongly a molecule is anchored [14] in a particular region of the active site, but do not directly determine the free energy of bound conformations relative to unbound conformations.

The binding-energy landscape for the pteridine ring is smooth and characterized by a single stable state located near $R = 0.0$, the crystallographic binding mode (Fig. 4a). Even at relatively high temperatures such as $T = 1000$ K, alternate binding modes are not stable; only at much higher temperatures do alternate binding modes begin to compete with the crystallographic state.

The landscape for the all-carbon ring has a low-energy binding mode corresponding to the crystallographic orientation, which is energetically favored at low temperatures, below $T = 400$ K (Fig. 4b). Two additional binding modes are stable at intermediate temperatures, one near 2.8 Å and a second near 4.8 Å. The second alternate binding mode is dominant at high temperature, suggesting that it is favored entropically. This variant illustrates one of the difficulties with using kinetic docking simulations to infer the shape of binding-energy landscapes. While the docking simulations reveal a single favorable binding mode, which is the correct result at low temperature, there are two additional low-energy binding modes accessible at higher temperatures. As a consequence, the crystallographic binding mode of the all-carbon variant is much less stable relative to alternate bound conformations than that of the standard pteridine ring.

Like the all-carbon ring, the landscape for the *anti*-pteridine ring is characterized by three low-energy

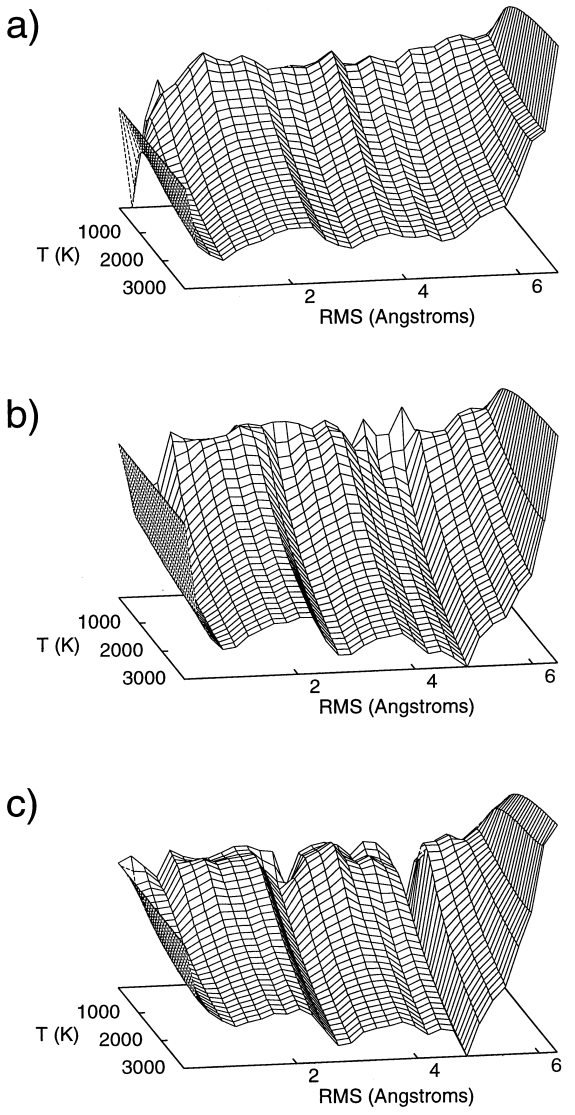


**Fig. 4a–c.** The binding-free-energy landscapes for **a** the pteridine ring, **b** the all-carbon variant, and **c** the *anti*-pteridine variant

binding modes, at approximately 0.5, 2.8, and 4.8 Å (Fig. 4c). In this case, alternate binding modes are more stable than the crystallographic binding mode at all temperatures and the crystallographic binding mode, while a local minimum, is never thermodynamically stable even at very low temperature. Again, analysis of the landscape shows that docking simulations provide an incomplete picture: they do not identify the crystallographic conformation as a metastable state.

The landscapes of the three pteridine rings illustrate the role of steric interactions and hydrogen bonding in molecular recognition of MTX by DHFR. The underlying shape of all three landscapes is similar and is apparently determined by the steric interactions, with local minima in all cases at 0.5 Å, and at least a weak local minimum at 2.8 and 4.8 Å. While the shape of the pteridine ring is sufficient to ensure that the crystallographic binding mode is a local energy minimum, it does not guarantee that it is thermodynamically stable: the hydrogen bonds ensure that the crystallographic binding mode has the lowest free energy.

The landscape for the standard pteridine fragment is characteristic of a recognition nucleus with a single dominant binding domain, which results from the energy gap between the native binding mode and alternative low-energy conformations. As a result, the native binding mode for the pteridine fragment is thermodynamically favorable at high temperature and does not compete with alternate binding modes. The native binding mode for the all-carbon variant is stable at low temperature, but much less so than the standard pteridine ring. The hydrogen-bonding pattern for the *anti*-pteridine ring is sufficiently different from the standard pteridine ring that the crystallographic binding mode is no longer a stable minimum even at low temperature. While the shape of the ring is sufficient to make the crystallographic mode a weak local minimum, the fact that the hydrogen bonds are anti-complementary ensures that the crystallographic mode is never stable.

## 4 Conclusions

A molecular anchor can be identified from the shape of its binding-energy landscape $F(R,T)$, and may be defined as a compound with a single dominant binding mode. Analysis of the binding-energy landscape of the pteridine ring in MTX reveals that this substructure acts as an anchor in binding to DHFR. However, the binding-energy landscape only measures the free-energy difference between alternate bound conformations and does not refer to the bound-unbound equilibrium process: it is independent of the nature of the solvent. For example, it is energetically favorable to remove the all-carbon ring from water because it is strongly hydrophobic. By contrast, the pteridine ring has many hydrogen-bonding groups and less hydrophobic driving force. Under conditions that the hydrophobic repulsion from the solvent dominates the equilibrium, the all-carbon ring may actually have higher binding affinity than the standard pteridine ring. However, the binding mode is not specific and therefore the all-carbon ring is not a good anchor.

The crystallographic binding mode is unstable for the *anti*-pteridine ring and only marginally stable for the all-carbon ring; neither variant provides an anchor point for the attachment of additional groups. Hence, while steric complementarity appears to be a necessary condition for binding and is sufficient to ensure that a particular binding mode is favorable at low temperature, a complementary pattern of specific hydrogen bonds stabilizes the favored binding mode and ensures its stability over a broader range of temperature. The inverse pattern of hydrogen bonds is sufficient to disrupt the binding mode at all temperatures.

The high stability of the pteridine ring ensures that it is a good anchor and is amenable to subsequent elaboration. Accordingly, it is possible to design larger inhibitors that incorporate the pteridine ring, where the

pteridine ring remains in the same crystallographic binding mode. MTX is an example of such an inhibitor, which also contains a second anchoring substructure, the glutamate moiety. One may speculate that the primary function of the *p*-aminobenzoyl group is to link the pteridine ring with the glutamate, and that part of the reason that MTX is such a good inhibitor is that it is composed of two distinct anchor moieties. Other anchor moieties could be identified by an analogous analysis of substructures of other inhibitors of DHFR, and might provide further insight into the important features of the critical fragment.

An important advantage of molecules that contain an anchor is that the resulting binding mode can be predicted with high accuracy. As a result, peripheral groups that are designed to interact favorably with specific regions within the active site can be attached to the anchor [15]. By contrast, lack of an anchor fragment implies that the binding mode is not highly stable and may change depending on the type of groups added, complicating the design process considerably. After the anchor has been identified, one can increase binding affinity by adding hydrophobic groups [16]. Alternatively, focused combinatorial libraries based on an anchor fragment will have increased probability of success because they ensure specific molecular recognition with the active site.

*Acknowledgements*. We thank the members of the computational chemistry department at Agouron Pharmaceuticals, Inc., in particular D. Gehlhaar, for help developing the software used in this study.

## References

1. Wlodawer A, Erickson JW (1993) Annu Rev Biochem 62:543–585
2. Hünenberger PH, van Gunsteren WF (1997) In: van Gunsteren WF, Weiner PK, Wilkinson, AJ (eds) Computer simulation of biomolecular systems. Theoretical and experimental applications. Kluwer Dordrecht, pp 3–82
3. Yue K, Fiebig KM, Thomas PD, Chan HS, Shakhnovich EI, Dill KA (1995) Proc Natl Acad Sci USA 92:325–329
4. Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rejto PA, Rose PW, Verkhivker GM In: Altman RB, Dunker AK, Hunter L, Klein TE (eds) Pacific Symposium on Biocomputing '99. World Scientific, Singapore (in press)
5. Rejto PA, Verkhivker GM (1996) Proc Natl Acad Sci USA 92:325–329
6. Shah NK, Rejto PA, Verkhivker GM (1997) Proteins Struct Funct Genet 28:421–433
7. Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J (1982) J Biol Chem 22:13650–13662
8. Gehlhaar DG, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST (1995) Chem Biol 2:317–324
9. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) J Comput Chem 13:1011–1021
10. Mayo SL, Olafson BD, Goddard WA III (1990) J Phys Chem 94:8897–8909
11. Gehlhaar DG, Bouzida D, Rejto PA In: ACS Symposium Series on Rational Drug Design (in press)
12. Bouzida D, Kumar S, Swendsen RH (1992) Phys Rev A 45:8894–8901
13. Ferrenberg AM, Swendsen RH (1989) Phys Rev Lett 63: 1195–1198
14. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) J Mol Biol 229:105–124
15. Miranker A, Karplus M (1991) Proteins Struct Funct Genet 11:29–34
16. Rejto PA, Verkhivker GM (1997) In: Altman RB, Dunker AK, Hunter L, Klein TE (eds) Pacific Symposium on Biocomputing '98. World Scientific, Singapore, pp 362–373