Published in final edited form as:

J Comput Chem. 2008 August ; 29(11): 1753–1761. doi:10.1002/jcc.20936.

Empirical Entropic Contributions in Computational Docking: Evaluation in APS Reductase Complexes

Max Chang⁽¹⁾, Richard K. Belew^{(1),(2)}, Kate S. Carroll⁽³⁾, Arthur J. Olson⁽⁴⁾, and David S. Goodsell⁽⁴⁾

⁽¹⁾ Bioinformatics Program, University of California at San Diego, La Jolla, California 92093

⁽²⁾ Department of Cognitive Science, University of California at San Diego, La Jolla, California 92093

⁽³⁾ Department of Chemistry and the Life Sciences Institute, University of Michigan at Ann Arbor, Ann Arbor, Michigan 48109

⁽⁴⁾ Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Abstract

The results from reiterated docking experiments may be used to evaluate an empirical configurational entropy of binding in ligand-protein complexes. We have tested several methods for evaluating the configurational entropy of binding of 22 nucleotide analogues to the enzyme APS reductase. These include two cluster size methods that measure the probability of finding a particular conformation, a method that estimates the extent of the local energetic well by looking at the scatter of conformations within clustered results, and an RMSD-based method that uses the overall scatter and clustering of all conformations. We have also directly characterized the local energy landscape by randomly sampling around docked conformations. The simple cluster size method shows the best performance, greatly improving the ranking of conformations in multiple docking experiments.

Keywords

AutoDock; empirical free energy force fields; conformational entropy; computational docking

INTRODUCTION

The AutoDock3 and AutoDock4 empirical free energy force fields have been calibrated against a set of several hundred ligand-protein complexes of known structure and binding constants^{1,2}. In our experience, this force field has been effective for the prediction of binding constants with tight-binding complexes, but we have noticed two significant problems.

First, we often find an incorrect conformation with slightly more favorable energy than the experimentally-observed conformation. However, these incorrect conformations are found with very low frequency when multiple docking experiments are performed: incorrect, low energy conformations will be found in ~1% of docking experiments, and the correct conformation will be found in 25–100% of the experiments. Thus, in these cases, a simple

procedure that chooses the conformation of best energy from a set of multiple docking experiments will yield an incorrect conformation.

Second, the current force field poorly predicts the free energy of binding of weakly-interacting molecules. An example from APS reductase (adenosine 5'-phosphosulfate reductase), the subject of this report, highlights the problem. Experimentally, 5'-AMP binds tightly but 3'-AMP, which has a similar number of atoms and functional groups, binds weakly. However, in AutoDock both are predicted to bind tightly with similar binding constants. However, by looking at the frequency that a given conformation is found in reiterated docking experiments, a difference may be seen, as shown in Figure 1. When these compounds are docked multiple times, AutoDock finds a consistent conformation for 5'-AMP in many docking experiments, whereas 3'-AMP adopts many different conformations and the low energy conformations of 3'-AMP are only found in a small fraction of docking experiments.

We have observed this many times in other systems: if a given molecule shows a consistent conformation in many docking simulations, we have far more confidence in the result. Our current hypothesis is that the frequency of finding a given conformation is providing information on the energy landscape of binding, and that a high frequency is a measure of favorable entropy in the binding process. Recent work has shown that the energetic contribution of this configurational entropy will be high. A recent study by Chang, Chen, and Gilson³ has estimated that the configurational entropy of binding of amprenavir to HIV-1 protease is 26.4 kcal/mol, of which 1.8 kcal/mol is due to the loss of conformational entropy when the molecule moves from freely flexible in solution to its constrained position in the active site, and the bulk of the penalty is due to loss of vibrational entropy in the restrictive binding site.

Ideally, we would like to quantify the binding energy of the entire range of conformations available to the ligand and protein, and use this explicitly to evaluate the conformational entropy. However, these types of calculations, such as the Mining Minima calculation employed by Chang, Chen and Gilson, are too computationally expensive for typical docking studies. Instead, several laboratories are exploring methods for using information from the docking simulation or from inexpensive approximations of the range of conformations to evaluate this entropic component.

Many of these methods perform multiple docking experiments, cluster the resulting conformations by similarity, and then use a measure of the cluster size to estimate the conformational entropy. The assumption is that the docking protocol is providing information on the characteristics of the local energy landscape, and that large clusters of conformations are indicative of favorable entropic characteristics of this landscape. In previous experiments using AutoDock, we have used a cluster size measure to discriminate binders from non-binders in an artificial active site⁴. Bottegoni, Cavalli and Recanatini⁵ have explored several clustering methods combined with a number of popular docking methods to identify significantly populated clusters, showing that they tend to be associated with the observed binding mode. Ruvinsky and Kozintsev⁶ have used the size of clusters in AutoDock to approximate a vibrational entropy, and in later work⁷, to estimate a probability distribution function of conformations from multiple docking experiments. Xiang, Soto and Honig⁸ have used RMSD values between different conformations from a genetic algorithm search to approximate this entropic contribution in protein loop prediction.

In this report, we evaluate these methods for their ability to predict correctly the expected bound conformations of nucleotide analogues in APS reductase⁹⁻¹¹. All of these methods seek to characterize the local energy landscape and use this information to estimate an

entropic contribution to the binding free energy. The methods perform a sparse sampling of the landscape by reiterated docking or random sampling, making the implicit assumption that the sampled points will represent the features of the entire local landscape. We have found that APS reductase is an excellent test for these methods because experimental binding constants are available for a series of compounds of similar size and chemical composition, but with a wide range of binding constants. This provides a more critical test set than the typical databases used in most studies, which typically include a diverse collection of ligand-protein complexes, but all are specific, tight-binding complexes.

METHODS

Motivation

In the most general case, we seek to evaluate the entropic contribution of binding (ΔS) through use of a conformational integral⁶:

$$\Delta S = \ln \left(\frac{\sigma_l \sigma_p}{\sigma_{pl}} \frac{c_0 N_0}{8\pi^2 (2\pi)^{n_{tor}}} V_B \right)$$

where the conformational integral V is:

$$V_B = \int_{\Gamma} \exp \left[(U_{pl}(\mathbf{r}, \boldsymbol{\Omega}) - E_{pl}) / RT \right] d\mathbf{r} d\boldsymbol{\Omega}$$

In these equations, the σ terms account for any symmetry in the molecules, with values of 1 for asymmetric molecules, $c_0 = 1$ mol/L, N_a is Avogadro's constant, n_{tor} is the number of torsional degrees of freedom in the ligand, $U_{pl}(\mathbf{r}, \boldsymbol{\Omega})$ is the energy of each complex conformation, Γ is the region of integration (typically a small space that includes conformations with similar binding modes) and E_{pl} is the minimum energy of the complexes within the region Γ . The vectors $\mathbf{r}$ and $\boldsymbol{\Omega}$ define the 3 translational and the 3+ n_{tor} rotational motions of each complex. As noted in the Ruvinsky presentation, these calculations are appropriate for relative protein-ligand motions.

In this manuscript, we test several simple approximations to this integral, based on conformations obtained in reiterated AutoDock docking experiments and by directly sampling the local energy landscape. Our goal is to provide an efficient empirical method for estimating this entropic contribution. We seek to improve the estimation of binding constants by rescoring trial docked conformations, combining this estimated conformational entropy, which is derived from reiterated docking experiments, with predicted enthalpic and desolvation contributions used during the docking simulation of each conformation.

In all of these methods, we begin with a set of conformations obtained from docking simulation or from random sampling, and we assume that these sparse samples may be used to characterize the entire local energy landscape. It is important to keep in mind that the evolutionary search method used in AutoDock, which combines a genetic algorithm with a local search¹, is not designed to be a uniform (Monte Carlo) sampling process, but instead to be successful at finding extreme (minimum) values of the energy function. Thus, it is not directly giving the information needed to estimate the conformational integral, but may be used to infer properties of the energy landscape and conformational entropies. Note also that: (1) the method is heuristic and stochastic, and thus does not guarantee convergence, so the search must be repeated multiple, statistically independent times, and (2) it generates a history of the search process as a by-product. Both of these properties provide opportunities

and limitations for use in estimation of entropic contributions, and help to motivate our random sampling experiments, below.

Method of Ruvinsky and Kozintsev

Ruvinsky and Kozintsev have reported a method for using the results from multiple docking experiments to evaluate a vibrational entropy⁶. We will refer to this as the “RK” method. They begin by clustering the docked conformations, then they evaluate the entropy based on the conformational space spanned within the cluster. They choose the lowest energy conformation in each cluster as the representative position, and assume that the members of the cluster provide a representative snapshot of the motions available to representative position. For each cluster i , they estimate this conformational volume $V(r_i, \Omega_i)$ by evaluating the maximum and minimum values in each of the dimensions of r and Ω , and then calculating the product of these maximum-to-minimum intervals:

$$V_B(r_i, \Omega_i) \cong V_{RK}(r_i, \Omega_i) = \prod_{j=1}^3 (\max(r_i^j) - \min(r_i^j)) \prod_{k=1}^{3+N_{tor}} (\max(\Omega_i^k) - \min(\Omega_i^k))$$

Rather than explicitly incorporating a dependence on inter-sample distances, as in the Colony method or the weighted RMSD method described below, the RK method is evaluated within 1Å or 2Å clusters.

This method has one potential conceptual limitation: in cases where the number of conformations is small, the value of V_B may be more an indication of the sample size than the extent of the favorable conformational space. We performed a random value test (Figure 2) to determine the possible influence of the small number of conformations in a typical cluster on the value of this volume. It was performed by picking random values for the max and min values of each conformational variable within a constant volume of conformational space. For the experiment, random values were chosen within uniform bounds of -1.29 to 1.29 (Å or rad) for 12 conformational variables (3 translation, 3 rotation, and 6 torsion). During analysis of our docked conformations, a value of roughly $T\Delta S_{RK}=12$ kcal/mol was found for the largest clusters, so the value of 1.29 was chosen to provide a value of 12 kcal/mol for random distributions with many points, that maximally fill the entire available conformational volume.

Colony Method

The “Colony” method⁸ uses a RMSD-dependent term to evaluate a conformational free energy. The method was developed for use in protein loop prediction. The authors generated a set of 2,000 random backbone conformations, added sidechains using a rotamer approach, performed an energy minimization, and then retained the 1,000 lowest energy conformations for use in the free energy estimation. Here, we modify the approach for use with ligand-protein complexes, using 100 conformations from reiterated AutoDock simulations. The energy is based on the RMSD from each conformation to all other conformations obtained in the simulation:

$$\Delta G_i = -RT * \ln \left[\sum_j \exp \left(-\Delta E_j / RT - d_{ij}^2 / 6L \right) \right] + C$$

where ΔE values are the predicted energies of each conformation j , d_{ij} is the RMSD between the two conformations, L is number of amino acids in the peptide loops used in the original

study and $6L \text{ \AA}^3$ is an empirical factor defining the approximate volume of conformational space represented by each loop conformation, and C is a constant term that defines the magnitude of the entire conformational space. For use in rescoring multiple conformations of a single ligand, the constant term may be neglected. For this work, we used a value $L=2$ since these compounds have a similar number of torsional degrees of freedom as a dipeptide linker.

Cluster Size Method

We have tested two methods of using the cluster size as an estimate of the conformational integral. In these methods, the hypothesis is that the probability of finding a conformation in a given cluster is capturing information on the local energy landscape and thus is proportional to the conformational entropy. The first is a probability based on a simple conformation-centered RMSD, which we will refer to as the “RMSD” method. For each conformation i , RMSD values $d_{i,j}$ is calculated over all conformations j not equal to i , and the fraction less than a given threshold $d_{\max}$ is evaluated. In this work, we used a threshold of $d_{\max}=2 \text{ \AA}$ RMSD.

$$P_i^{\text{RMSD}} = \frac{N_{d_{j \neq i} \leq d_{\max}}}{N}$$

Where the numerator is the number of conformations with RMSD less than the threshold and N is the total number of conformations. The second is a probability based on a distance-weighted RMSD, which we will refer to as the “wRMSD” method:

$$P_i^{\text{wRMSD}} = \frac{\sum_{j \neq i} \exp(-d_{i,j}^2/2\sigma^2)}{N}$$

where the constant $\sigma = 2 \text{ \AA}$. If we assume that the favorable region of conformational space is proportional to these probabilities, then the conformational entropy may be estimated as:

$$\Delta S_i = -W^{\text{RMSD}} RT \ln(P_i^{\text{RMSD}})$$

where W^{RMSD} is an empirically-determined weight.

Random Sampling Method

We also estimated a value of the conformational integral based on a random sampling of the local energy landscape around each docked conformation. 100,000 conformations were generated with small random displacements from the docked conformation. Translational displacements were chosen from a random distribution with bounds -0.5 to $0.5 \text{ \AA}$, rotational displacements were generated by picking a random axis and rotating by a random angular displacement with bounds -0.5 to 0.5 rad , and torsional displacements were generated with a random angular displacement with bounds of -0.5 to 0.5 rad .

The conformational integral was calculated as:

$$\tilde{V}_B(r_i, \Omega_i) = \frac{\sum_j \exp(\Delta E(r_j, \Omega_j) - \Delta E(r_i, \Omega_i))}{N}$$

where the ΔE values are predicted energies from AutoDock and the summation is performed over the $N=100,000$ samples j around the conformation of minimum energy i .

Binding Constants for Ligands with APS Reductase

Binding constants are available for 22 nucleotide analogues bound to APS reductase (Table 1). Values of K_i were determined under single turnover conditions from the dependence of the observed rate constant (k_{obs}) at a given inhibitor concentration under conditions of subsaturating APS, such that K_i is equal to the K_d ^{9,12}. Kinetic data were nonlinear-least squares fit to a model of competitive inhibition. Each K_d reflects the average of at least two independent experiments, and the standard deviation was less than 10% of the value of the mean. The synthesis, characterization and biochemical analysis of the analogues used in this computational study will be reported elsewhere¹¹.

Docking with AutoDock4

Docked conformations and predicted free energies of association were obtained for 22 nucleotide analogues using AutoDock4 (<http://autodock.scripps.edu>). Coordinates for APS reductase were obtained from C. David Stout prior to release—they are identical with subunit B in entry 2goy at the Protein Data Bank¹⁰. Coordinates for the enzyme were processed in AutoDockTools by adding all hydrogens, assigning charges with the Gasteiger method^{2,13}, and merging non-polar hydrogen atoms. Coordinates for the nucleotides were constructed in InsightII starting with the conformation of the APS nucleotide bound at subunit B in the crystallographic structure. Charges were assigned in ADT and non-polar hydrogen atoms merged. Docking experiments were then performed in AutoDock4 using the default docking parameters, with 2,500,000 energy evaluations for each docking experiment and finding 100 separate docked conformations for each nucleotide.

A test of the role of sugar conformation in the nucleotide was performed using 5'-ADP conformations from entries 1e19, 1m7g, 1o0h and 1rdq from the Protein Data Bank (<http://www.pdb.org>), which were judged to have different sugar conformations based on the distance between C5' and N9, and the torsion angle through atoms C5'-C4'-C1'-N9. These ADP coordinates were prepared and docked similarly to the other nucleotides.

Since crystallographic results are only available for the ligand APS, RMSD values were calculated based on the distance between the nucleotide atoms and the modeled nucleotide, which was created to overlap the analogous atoms in the crystallographic conformation of APS. Thus, the RMSD values in this paper refer to the similarity of the binding modes to the observed mode of APS.

Calibration of Empirical Terms

Linear regressions and statistical analysis were performed using the free software R (<http://www.r-project.org>), forcing the regression to include the origin in all cases.

RESULTS

Role of Sugar Conformation

The APS nucleotide conformation found in the crystallographic structure has an unusual sugar conformation, which orients the adenine base and phosphate-sulfur group unusually close to one another. To test the importance of this conformation, we performed docking experiments with four different conformations of 5'-ADP. All four of these conformations did not yield docked conformations within 2 Å RMSD of the observed nucleotide conformation. So for all work, we used exclusively the sugar conformation found in the APS reductase structure.

Docking of Nucleotides to APS Reductase

For each of the 22 nucleotides, we performed 100 docking experiments, and clustered the resulting conformations using a 2 Å threshold. The results, shown in Table 1, are typical of results of AutoDock docking experiments. In 3/22 compounds, the conformation with best energy was in the proper position, but in the remaining 19, they were greater than 2 Å RMSD different than the crystallographic position. If, however, we look at the best conformation in the largest cluster, 18/22 conformations are within 2 Å of the expected location.

These types of results, which are commonly obtained for AutoDock experiments, are the motivation for the current work. Tight binding ligands, such as 5'-AMP (Figure 1a), show excellent clustering and weakly binding ligands, such as 3'-AMP (Figure 1b), show poor clustering, although both show similar predicted binding energies. The docking protocol, as revealed in the clustering, is capturing some aspect of the binding energetics that is missing from the current empirical free energy force field.

Conformational Entropies From Cluster Size

Table 2 includes results from regression analysis. Observed binding energies were fit with models that included the predicted AutoDock4 energy and one of the two clustering models: the 2 Å threshold model RMSD or the distance weighted model wRMSD. In both cases, modest improvement was seen. The standard error of the predicted binding energy was reduced slightly, and the multiple R-squared increased.

Table 3 shows the effectiveness of the cluster size models in rescoring. The first column shows the poor predictive ability of the basic AutoDock4 method: when looking at only the conformation of best energy, only 3/22 identify the proper conformation (these results are also shown in Table 1). The second and third columns show the results when the cluster size measure is included. Both methods show excellent predictive ability, ranking the expected conformation as the best in 15/22 or 16/22 cases.

The significance of this result may be estimated by comparison with a statistical method based on Bernoulli trials. We calculated the fraction of dockings with RMSD less than 2 Å for each compound, which ranges from 0.00 for 3'-phospho-5'-AMP to 0.99 for 3'-deoxyadenosine. Using these fractions, we can estimate the expected number of correct conformations we would obtain by randomly choosing a conformation for each compound. This analysis estimates that random choice would give a correct answer in 12.35 cases, with a standard deviation of 1.72, out of the 22 compounds.

Conformational Entropies from the RK Method

We also tested the RK method for evaluating the entropic component for ranking the binding energies. The entropy calculated by this method is typically on the order of tens of kcal/mol, so if it is simply added to the predicted energy from AutoDock4, the binding

energy is predicted to be positive in all cases. However, these energies are effective for reranking, as shown in the fourth column in Table 3 (marked “best RK”), correctly predicting the conformation of 14/22 compounds.

A regression analysis of the RK entropy showed a slightly worse prediction of the binding energy, as shown in Table 2, and a very slight improvement in reranking, correctly predicting 5/22 compounds.

In our work with the RK method, we have noticed one potential limitation: the evaluation of V_{RK} is strongly influenced by the number of observations in each cluster. The method attempts to evaluate the extent of favorable regions of the local energy landscape by evaluating the bounds on each cluster. However, since these clusters are composed of a small number of individual observations, clusters with few individuals will give smaller extents just through simple statistics. Figure 2 shows a scatter plot of cluster size vs. the RK entropy. The two lines show the upper and lower bounds of clusters composed of a set of points randomly distributed within a small constant volume of conformational space, and then used to calculate the RK entropy. The random distribution captures many of the features of the RK data, so we might expect that the RK method, when applied to clusters generated by AutoDock, more a reflection of the cluster size than of the extent of the local energy landscape. Notably, in a recent publication¹⁴, a direct measure of cluster size was used instead of the estimated conformational extent in a similar study to estimate conformational entropies in the context of several scoring methods.

Conformational Free Energies from the Colony Method

The “Colony” method⁸, originally developed to score candidate loop conformations in protein structure prediction, uses the RMSD values to the constellation of neighboring conformations to evaluate a conformational free energy. In cases where a conformation has many close neighbors, the Colony energy will be more favorable. As shown in Table 3, the Colony energy performs well for the rescoring of docked conformations in APS reductase, predicting the expected conformation in 13/22 cases. The use of the Colony method for prediction of binding free energies is problematic, since the constant term in the Colony equation is difficult to calculate. However, if we assume that this term is approximately the same across our test set of compounds, a regression analysis of Colony energies with the observed binding energies shows a moderate improvement over the predicted energies from AutoDock4, as shown in Table 2.

Conformational Entropies from the Local Conformational Integral

Ideally, we could like to be able to start with a single docked conformation and, by analyzing the local energy landscape, evaluate this entropic contribution to the binding strength. As a first step towards this goal, we have randomly sampled the conformational space around each docked conformation and calculated a conformational integral based on the energy landscape. This is partially effective for improving the prediction of free energies and in reranking. The regression showed a small improvement in the standard error, and the method was able to rank 10/22 compounds.

Comparing two of the compounds from this study, we can see how these conformational integrals capture the underlying landscape. 5'-AMP and 3'-AMP have the same number and type of atoms and the same number of torsional degrees of freedom, but widely different experimental binding constants. In docking, 5'-AMP gives a tight cluster of 61/100 docked conformations in the expected location, whereas weaker-binding 3'-AMP shows a scatter of different, small cluster conformations.

Looking at the energy landscape around the docked conformation, as shown in Figures 3 and 4, we find that 5'-AMP has a broader energy well than 3'-AMP. Thus, small motions of 3'-AMP will run up against large steric contacts, whereas small motions of 5'-AMP do not encounter bad contacts.

Unfortunately, these types of correlations were difficult to extract for other compounds, where the structural similarity was not as great. Looking at the entire set, the greatest trend was a strong correlation between the value of the conformational integral and the number of torsional degrees of freedom in the molecule. This is not a surprise, since this merely reflects the magnitude of the entropy involved in freezing these torsional degrees of freedom into a confined space of the active site. The more subtle effect of the local shape of that active site, as seen in the 5'-AMP vs. 3'-AMP landscapes, is overshadowed by this larger effect.

DISCUSSION

The ultimate goal of this work is to find a computationally tractable method to evaluate the conformational entropy of binding, and thus improve our predicted binding energies. This is essential for the future success of docking in computer-aided drug design, where the common presence of false positives and false negatives during virtual screening is a major problem in current studies.

The results presented here suggest that the cluster size is an effective and cheap method for evaluating these entropies, and may be used to improve both the ranking of different complexes, and for the identification of proper binding modes within a single complex. The Colony method shows a similar success, which should come as no surprise since it also uses a measure of cluster size to evaluate the entropic contribution.

These cluster size methods, however, are not satisfying from a conceptual level, since they are relying on some unknown combination of the overall energy landscape and the details of the docking protocol. Ideally, we would like to develop a method that analyses the energy landscape, both locally and globally, and uses that information to identify the major binding modes and affinities. The RK method was designed with this goal in mind. It uses a simple method to estimate the local conformational area that is available to a given ligand. Unfortunately, the results shown in Figure 2 suggest that the method is primarily a reflection of the cluster size, as opposed to a quantification of the local energy landscape. The RK term is highly dependent on the actual number of observations in each cluster and their orientation relative to one another in conformational space, not just the overall conformational volume available to the conformations.

Our attempt to characterize the local energy landscape through random sampling has provided some provocative, but not definitive, results. The results presented in Figure 3 show that there are significant differences in the local energy landscape for two forms of AMP, differences that correlate strongly with the large difference in binding constants between these two compounds. However, this principle did not generalize over the entire set. Our current hypothesis is that the docking analysis, and thus the clustering, is capturing information over a larger area of conformational space than we sampled in this work, and that sampling of this larger space will be necessary to develop an effective method for directly evaluating the conformational entropy contribution to binding. However, use of the cluster size in multiple docking experiments is a fast and easy way to estimate this contribution, and is a viable method for improving current docking results.

Acknowledgments

The authors wish to thank Garrett M. Morris for helpful discussion. This work was funded by grant R01 GM069832 from the National Institutes of Health. This is manuscript 18872 from the Scripps Research Institute.

References Cited

1. Morris GM, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comp Chem* 1998;19:1639–1662.
2. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 2006;28:1145–1152. [PubMed: 17274016]
3. Chang CA, Chen W, Gilson MK. Ligand configurational entropy and protein binding. *Proc Natl Acad Sci USA* 2007;104:1534–1539. [PubMed: 17242351]
4. Rosenfeld RJ, et al. Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *J Comp Aided Mol Design* 2003;17:525–536.
5. Bottegoni G, Cavalli A, Recanatini M. A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs. *J Chem Inf Model* 2006;46:852–862. [PubMed: 16563017]
6. Ruvinsky AM, Kozintsev AV. New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *J Comput Chem* 2005;26:1089–1095. [PubMed: 15929088]
7. Ruvinsky AM, Kozintsev AV. Novel statistical-thermodynamic methods to predict protein-ligand binding positions using probability distribution functions. *Proteins: Struct Funct Genet* 2006;62:202–208. [PubMed: 16287127]
8. Xiang Z, Soto CS, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc Natl Acad Sci USA* 2002;99:7432–7437. [PubMed: 12032300]
9. Carroll KS, et al. A conserved mechanism for sulfonucleotide reduction. *PLoS Biology* 2005;3:e250. [PubMed: 16008502]
10. Chartron J, et al. Substrate recognition, protein dynamics, and iron-sulfur cluster in *Pseudomonas aeruginosa* adenosine 5' phosphosulfate reductase. *J Mol Biol* 2006;364:152–169. [PubMed: 17010373]
11. Carroll, KS. Probing binding determinants of the *Mycobacterium tuberculosis* APS reductase active site. 2007. manuscript in preparation
12. Fersht, A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W. H. Freeman Company; New York: 1999.
13. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 1980;36:3219–3228.
14. Ruvinsky AM. Role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions. *J Comput Chem*. 2007 in press.
15. Sammon JW. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* 1969;C-18:401–409.

Figure 1.

Clusters analysis of docking for 5'-AMP and 3'-AMP. The graphs on the left use Sammon mapping¹⁵ to preserve the approximate separation in conformational space between clusters. Each circle represents a cluster of conformations within 2 Å RMSD of each other, and the size of the circle is proportional to the number of conformations in the cluster. The expected bound conformation is shown with a diamond. The images on the right show all of the docked conformations. 5'-AMP binds tightly, and many of the docked conformations cluster into one large group at the expected conformation. 3'-AMP, however, binds weakly and shows a wide scattering of small clusters.

Figure 2.

Analysis of RK conformational entropy results. Each point represents one cluster of docked conformations, with X for clusters within 2 Å RMSD of the expected conformation, and a dot for conformations docked incorrectly. Results for all 22 compounds docked to APS reductase are included. The lines show the upper and lower values obtained for a random sample of points distributed within a small constant volume. The random samples account for much of the range seen in the RK analysis of the docking results.

Figure 3.

Analysis of the local energy landscape. Each point represents a small random change in conformation away from the most favorable bound conformation. RMSD values are calculated between the perturbed conformation and the starting conformation. 5'-AMP shows a wide basin, with very few unfavorable conformations until they are a distance of about 0.5 Å RMSD from the bound conformation. 3'-AMP shows a narrower basin, with many unfavorable conformations as distances less than 0.25 Å RMSD.

Figure 4.

Analysis of the local translational energy landscape. Conformations were sampled in the range of $-1\text{\AA}$ to $+1\text{\AA}$ in the x and y directions around the most favorable bound conformation. The energy of the sampled conformations is shown here, with the outer contour at -1.5 kcal/mol and additional contours at -1.5 kcal/mol increments.

Table 1

Results of Docking

	ΔG_{obs}	Best Energy			Largest Cluster		
		N	AGAD4	RMSD	N	AGAD4	RMSD ntor
5' AMP	-8.07	2	-8.73	4.04	61	-7.96	0.81 6
7-deazaAMP	-7.51	1	-8.36	3.46	77	-8.14	0.81 6
5' ADP	-7.29	3	-10.07	3.03	41	-9.98	0.78 8
3' deoxyAMP	-7.21	3	-8.29	3.13	81	-8.29	0.81 5
5' PMP	-6.30	1	-8.37	3.33	60	-7.73	0.90 6
NimethylAMP	-5.97	45	-8.24	0.82			6
8-aminoAMP	-4.95	50	-8.29	1.63			6
2-aminoAMP	-4.76	2	-9.28	3.99	20	-8.19	0.96 6
3' phosphoAMP	-4.76	4	-9.07	3.21			7
2-methoxyAMP	-4.57	2	-8.51	3.56	17	-8.13	1.31 6
bmethAPS	-4.22	57	-9.34	0.81			8
2' deoxyAMP	-4.13	1	-8.75	4.10	31	-7.24	1.10 5
adenosine	-3.93	5	-5.58	3.64	27	-4.44	0.69 5
dimethylAMP	-3.90	13	-8.15	3.03			6
5' IMP	-3.44	2	-8.60	3.12	23	-7.26	1.48 7
3' deoxyadenosine	-3.17	1	-5.48	4.74	99	-5.27	0.61 4
5' phosphoribose	-2.73	2	-6.93	3.62	7	-6.06	1.93 5
3' AMP	-2.27	8	-9.25	3.87			6
2' deoxyadenosine	-2.00	4	-5.90	4.78	13	-4.90	3.17 4
ribose	-1.77	2	-3.65	9.99	56	-4.29	1.73 4
adenine	-1.76	21	-4.13	2.81	23	-3.81	1.49 0
5' IDP	-1.54	1	-9.90	3.90	13	-9.00	0.90 9

ΔG_{obs} is the experimental free energy of binding; N is the number of docked conformations in the cluster of best energy; AGAD4: Predicted free energy of binding from AutoDock; RMSD is the root mean square difference in coordinates between docked conformation and analogous atoms in the crystallographic structure; ntor is the number of torsional degrees of freedom in the molecule; and N, ΔG_{calc} , and RMSD are provided for the cluster of best energy and the largest cluster.

Full names of the compounds, in the order presented here are: 5'-adenosine monophosphate, 7-deaza-5'-adenosine monophosphate, 5'-adenosine diphosphate, 3'-deoxy-5'-adenosine monophosphate, 5'-purine monophosphate, N6-methyl-5'-adenosine monophosphate, 8-amino-5'-adenosine monophosphate, 2-amino-5'-adenosine monophosphate, 3'-phospho-5'-adenosine monophosphate, 2'-methoxy-5'-adenosine monophosphate, β -methylene adenosine 5'-phosphosulfate, 2'-deoxy-5'-adenosine monophosphate, adenosine, N6,N6-dimethyl-5'-adenosine monophosphate, 5'-inosine monophosphate, 3'-deoxyadenosine, 5'-phosphoribose, 3'-adenosine monophosphate, ribose, adenine, 5'-inosine diphosphate.

Table 2

Results of Regression

	SdtErr	MultRsqr	coeff(t value) ΔGAD4 term	coeff(t value) cluster term
Δ GAD4	1.806	0.864	0.577(11.6)	n/a
Δ GAD4+RMSD	1.737	0.881	0.658(9.6)	1.148(1.6)
Δ GAD4+wRMSD	1.709	0.884	0.658(10.2)	1.098(1.9)
Δ GAD4+RK	1.841	0.866	0.380(2.2)	0.084(0.6)
Colony	1.703	0.880	n/a	0.483(12.4)
Δ GAD4+Vb	1.691	0.887	1.381(3.4)	1.030(2.0)

Table 3

Results of Rescoring

	AD4	wRMSD	RMSD	best RK	fit RK	best Colony	fit Vb	lowest RMSD
5' AMP	4.04	0.83	0.81	0.81	4.04	0.93	0.95	0.78
7deazaAMP	3.46	0.80	0.81	0.81	0.87	0.80	0.85	0.77
5' ADP	3.03	0.78	0.78	0.78	0.80	0.77	2.72	0.69
3' deoxyAMP	3.13	0.90	0.90	0.81	3.47	3.13	0.81	0.77
5' PMP	3.33	0.89	0.89	0.90	3.34	0.89	0.93	0.81
NimethylAMP	0.82	0.84	0.82	0.82	2.37	0.87	2.42	0.78
8aminoAMP	1.63	1.63	1.63	1.63	1.78	1.58	0.80	1.52
2aminoAMP	3.99	0.96	0.97	0.96	3.99	1.79	3.87	0.81
3' phosphoAMP	3.21	3.21	3.20	3.48	3.50	3.80	3.08	2.76
2methoxyAMP	3.56	1.31	1.31	1.31	3.18	2.42	3.89	0.81
bmethAPS	0.81	0.81	0.81	0.81	0.78	0.73	2.61	0.63
2' deoxyAMP	4.10	1.10	1.39	3.45	3.89	3.89	2.72	0.94
adenosine	3.64	0.68	0.69	3.64	2.60	3.37	4.81	0.59
dimethylAMP	3.03	2.69	3.22	3.03	4.41	3.84	2.71	1.35
5' IMP	3.12	2.83	1.48	0.83	4.08	0.93	0.88	0.99
3' deoxyadenosine	4.74	0.59	0.61	0.61	0.61	0.58	5.16	0.56
5' phosphoribose	3.62	3.78	3.78	3.89	2.83	3.77	0.93	1.75
3' AMP	3.87	3.82	3.87	3.87	3.95	5.16	3.96	3.02
2' deoxyadenosine	4.78	4.78	4.78	4.69	2.39	2.79	1.79	1.12
ribose	10.62	1.72	1.73	1.62	10.62	1.71	1.55	1.55
adenine	2.81	2.81	2.81	6.44	2.97	1.49	1.58	1.48
5' IDP	3.90	0.98	0.98	0.80	2.26	0.74	2.56	0.81
	3	15	16	14	5	13	10	20/22

RMSD values are given for the docked conformation of best energy as determined by each method, with values >2.00Å in bold. The final line gives the number of conformations in each column with RMSD<2.00Å.