Integrative Geometric-Hashing Approaches to Binding Site Modeling and Ligand-Protein Interaction Prediction

Joanna Lipinski-Kruszka[1] and Rahul Singh[2]

[1] Department of Biology
[2] Department of Computer Science, San Francisco State University,
San Francisco, CA 94132

Abstract. The function of a protein is dependent on whether and how it can interact with various ligands. Therefore, an accurate prediction of protein-ligand interactions is paramount to understanding proteins' biological mechanisms and hence to the development of therapeutic agents. A ligand is most likely to bind in the largest pocket on the surface of the protein. Moreover, it requires that the pocket meets certain structural and geometric criteria that allow the ligand to "anchor" in place by forming stabilizing interactions with the protein. Based on this logic, many geometry-based algorithms have been developed to predict protein-ligand interactions. Here we investigate a geometric-hashing based algorithm – to see how well it distinguishes proteins that do and do not bind a ligand, and propose enhancements that improve its robustness. We also introduce an alternative way of integrating geometric and biochemical properties of multiple binding mechanisms into a single representation.

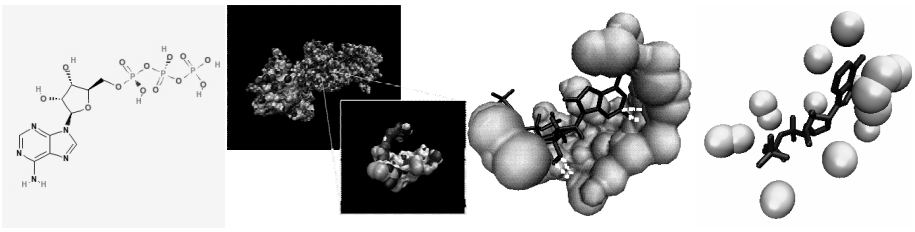
1 Introduction

Delivering a drug to market is a very lengthy process [3], much of which is spent in labs experimenting to find compounds that have good efficacy for the targeted protein. If this screening process could be expedited, drug development cost and time to market could be significantly reduced. This need has fueled the development of new computational approaches aimed at performing accurate virtual screening. A critical problem in this context is to correctly dock a ligand onto a receptor surface and involves both geometric and physicochemical reasoning.

Computational solutions to this problem are based on the postulate that one of the key requirements for binding is that the ligand has to be "anchored" in a small pocket on the surface of the protein, called the active site, through hydrogen or covalent bonds, or interactions such as hydrophobic or hydrophilic ones. Figure 1(c) illustrates an example of stabilization via hydrogen bonds. These stabilizing bonds and interactions can be formed only with the nearest atoms at specific locations. For this reason, the spatial distribution of atoms is thought to play a key role in ligand-protein binding, and hence suggests that the geometric definition of an active site is a critical and necessary [12] component in determining the protein's affinity for a ligand.

At the state-of-the-art, techniques that use geometry to screen for similarities between active sites can be broadly divided into two categories [12]: those that characterize general properties of an active site, and those that find sites that resemble an

G. Bebis et al. (Eds.): ISVC 2007, Part I, LNCS 4841, pp. 179–188, 2007.



a. b. c. d.

Fig. 1. (a) Structure of Adenosine Triphosphate (ATP). **(b)** A protein (PDB ID: 2HIX) and its active site. The close-up shows only atoms that are within 5 Å of the ligand (ligand not shown). **(c)** Example of ligand-protein interactions. ATP shown in black, active site of a protein in gray, some of the stabilizing hydrogen-bonds are shown as white, broken lines. **(d)** Model of ATP active site. Spheres represent atoms that make-up the model; ATP is shown as thick, dark lines.

already known active site. Here we focus on the latter, by considering approaches based on geometric hashing, an algorithm adopted from the field of computer vision, that has found significant applicability in this context.

Geometric hashing uses a model of a known active site to search for similar ones in other target proteins. One of the strengths of this approach is its ability to deal with noisy or incomplete data. One of its major drawbacks is its over-specificity due to a strong bias towards the geometry and physical properties of a single protein underlying the model. To overcome this limitation, we present an alternative that is based on robust, example-driven characterization of the binding site geometry. The proposed method starts by analyzing structurally diverse molecules which share their ability to bind a specific ligand but might differ in their mechanisms of interaction with it. Next, a voting-mechanism is used to extract from their binding sites characteristics that are common to multiple such proteins. Lastly, these features are integrated into a single arrangement of atoms representing all of the inspected binding mechanisms. This method not only limits the model's favoritism for any particular binding mechanism but, also, ensures that the pocket definition is deeply grounded and takes into account implicit geometric, physical, and chemical factors involved in the binding.

With this improved model, we then move on to investigate whether we can further improve on the robustness of geometric hashing for prediction of ligand-protein interactions. Because this approach can produce many possible solutions, careful scoring of "goodness" of matching is of key importance. Scoring is very sensitive to whether more emphasis is placed on the number of good matches or on their quality. How to choose the balance between these metrics depends on the data set. The former approach, which gives more weight to tightness of the fit, is more feasible if one wants to find partial regions that have an exact or very close match between them since it does a good job dealing with partial data; the latter approach is better if one wants to find the largest possible areas of similarity. Here we describe the pros and cons of a few approaches to finding the best possible balance between the two, and their biological significance.

2 Proposed Methods

Based on the geometric hashing algorithm, we created a model for an active site of Adenosine Triphosphate (ATP) (figure 1.a), a ubiquitous molecule playing a key role in intracellular transfer of energy. ATP was chosen because of the availability of structural data of proteins that are known to bind it at known locations [9].

We used unrelated, ATP-binding proteins to create our model and to evaluate our scoring methods. Using the Visual Molecular Dynamics (VMD) [11] tool, each of these proteins was stripped down to atoms within 3-5 Å from the bound ligand (figure 1.b). The rest of the atoms were ignored because they were assumed to be too far from the active site to significantly impact binding.

2.1 Base Creation and Transformation

Following the first step of geometric hashing, feature points were extracted; in this case, atoms of an active site were used for this purpose. Then, similarity between two binding sites – a model and a target – was investigated by comparing all possible transformations in 3-dimenssional space of one to that of the other. To do that, all possible permutations of three distinct atoms were found and used to create triplets to form triangular bases in 3D space. For each of the bases, transformation was done so that its vertices were placed: at the origin (0,0,0), on the X axis (x, 0, 0), and in the XY plane (x,y,0). The transform was then applied to all other atoms of the active site.

Every base of one protein was tested against all bases of another to find their best-fitting spatial confirmations. First, bases were tested for match. Two bases matched if after transformation onto the coordinate system they had: (1) the same atom types at the corresponding vertices, and (2) spatial location of their corresponding vertices within some specified threshold (set to 1 Å). If these conditions were met, transformations of the two proteins with respect to these bases were compared. Whenever atoms of the same type from the two proteins translated to approximately the same spatial location (also set to 1Å from each other) these atoms were considered matching. These matches were then used to evaluate the quality of the fit between the two bases.

2.2 Finding Best Fitting Bases

We evaluated four methods of scoring the fit between bases:

Overall Root Mean Square Deviation (RMSD): In this approach, after both proteins have been transformed, for every atom of the model protein a closest matching atom of the target protein was found. The distance between each atom pairs was used to compute the overall RMSD.

Top 80% RMSD: Our second approach, similarly to overall RMSD, found for every atom of one protein a closest atom in the other. This time, however, instead of taking all distances to compute the score, we only considered the top 80% best atom pairings. The worst scoring 20% were ignored.

RMSD of Matched Atoms: The third method computed RMSD for just atoms that matched. If a transformation found a close fit (noise threshold 1Å) between two atoms of the two proteins, the atom pair was included in the calculation of RMSD. Atoms that did not have a match within the allowable threshold were ignored. This method also required a minimum of 5 matching atoms.

Most Matched Atoms: Our last evaluation method gave best scores to transformations that resulted in the highest number of matched atoms and which had a global RMSD under 1.

2.3 Computation of the Binding Site Model

We used nine different ATP-binding proteins (2HIX, 2J9L, 2OOY, 2HF4, 2EWW, 2HVY, 2F43, 1Y64, and 1MO8) to build our model, which was created by finding the best fit of atoms of the active site of the template protein (2HIX) and of each of the other eight target proteins. Each best fit would cast a vote on atoms of the template that were matched to atoms of the target. In order to avoid bias towards any of the base selection algorithms, this process was repeated six times, using a different technique each time – one of the above described ones, plus two others that were only used for model creation but not for base fitting evaluation. These additional methods were (1) a modification of method *RMSD of Matched Atoms* which did not enforce an atom match minimum; and (2) a modification of *Most Matched Atoms* which removed the RMSD threshold constraint.

After performing comparisons using each of the techniques and with all of the proteins in the data set (this resulted in 48 alignments), votes were tallied. Atoms of the 2HIX active site that had at least 5 votes from at least two different evaluation methods were used as part of a model.

2.4 Evaluation of Scoring Methods

In order to evaluate our base matching techniques we tested them against two different data sets. For the first set, we selected active sites of thirteen known ATP-binding and -nonbinding proteins (table 2). The second data set was designed to test how well each technique aligned our model, which, as described above, consisted of a small subset of atoms from the active site of 2HIX, with a larger portion of the same active site. A total of 157 atoms were selected from the active site; each selection was done based on distance from ATP. The resulting pool of atoms was a superset of the model. Then, one by one, we removed from the set 10 atoms that were also present in the model (the model was unchanged, however). In this way, we created 11 new test data sets: 2HIX_all, which contained all 157 atoms, and 2HIX_less1 – 2HIX_less10, each of which had one to ten atoms removed (respective atom counts: 156 to 147). Together, these and the 13 shown in table 2 (total 24) were used as our test data.

In each evaluation of model-target pairs, two metrics of best fit were gathered: the number of atoms matched (N) and depending on the method of evaluation RMSD of either all, top 80%, or just the matched atoms (R). The final score was the product of N and the reciprocal of R (i.e., $S = N * 1/R$).

Table 1. Atoms making up the model. Only atoms that received at least 5 votes using at least two different scoring schemes were used. For spatial representation of the model see figure 1.d.

Summary of Votes						
# Votes	**# Methods**	**Residue**	**x**	**y**	**z**	**Symbol**
10	4	LYS	11.351	29.733	18.933	C
5	2	TYR	12.16	26.491	15.053	N
7	2	ARG	19.336	25.785	21.914	C
13	5	ARG	18.899	28.507	26.98	C
5	2	GLU	15.775	24.236	18.813	O
20	5	PHE	20.149	25.464	14.586	C
28	5	PHE	21.012	26.236	15.354	C
20	5	PHE	20.919	26.261	16.744	C
11	3	PHE	19.916	25.505	17.387	C
8	3	MET	15.763	31.245	15.317	C
6	3	LYS	19.976	29.852	13.22	N
23	6	ARG	17.982	34.78	25.432	C
10	4	ARG	17.428	33.79	24.751	N
19	6	LYS	14.799	34.08	21.473	C
9	3	LYS	14.381	32.963	22.409	N

Table 2. Proteins used to evaluate base-fitting methods. Relation refers to structural similarity to 2HIX, an ATP-dependent DNA ligase from S. Solfataricus [7].

Relation	Symbol	Description
neighbor, ATP	1A0I	ATP-dependent DNA ligase from Bacteriophage T7.
neighbor, no ATP	1VS0	Component of Mycobacterium DNA ligase D
unrelated, ATP	1XMJ	Human deltaf508 Nbd1 domain
unrelated, ATP	1V1B	2-keto-3-deoxygluconate kinase from thermus thermophilus
unrelated, ATP	2J9L	Cytoplasmic domain of the human chloride transporter
unrelated, ATP	1MO8	ATPase
unrelated, ATP	200Y	Adenylate sensor from AMP-activated protein kinase
unrelated, ATP	2HF4	Monomeric actin
unrelated, no ATP	1C97	Isocitrate complex of aconitase
unrelated, no ATP	2B3X	Orthorhombic crystal form of human cytosolic aconitase
unrelated, no ATP	2IPY	Iron regulatory protein
unrelated, no ATP	1Q5O	Hcn2j 443-645
unrelated, no ATP	1LB2	E. Coli Alpha C-Terminal Domain Of RNA Polymerase.

3 Results and Experimental Evaluations

We created a model of an active site for ATP docking based on several different, known ATP-binding proteins. We then used this model to evaluate four different alignment scoring techniques, each of which was based on geometric hashing.

3.1 Model

Because we choose random ATP-binding proteins to build our model, we expected that they would vary in their interactions and binding mechanism with their ligand. Visual inspection verified our prediction. The portions of proteins that were extracted around their ligands were of different shapes and sizes. Active sites of some proteins

interacted mostly with the base (the part of ATP consisting of two rings; see figure 1.a), others with phosphate group (the linear "tail" of ATP; see figure 1.a), but most with both parts (figure 2). The ligand was also positioned in each of these active sites in different poses (figure 3). These two factors contributed to a vast diversity of 3-dimentional conformations – virtually no two active sites were the same. Therefore, extraction of features common to all, and compilation of them into a single model was one of the key challenges. To address this, we utilized our geometric hashing-based voting approach and several unrelated proteins. In this way, we were able to combine into one model implicit geometric, physical, and chemical factors involved in a variety of mechanisms. The final model consisted of 15 atoms, which are listed in table 1 and rendered in figure 1.d.

3.2 Evaluation of Scoring Methods

Most biological applications that use geometric hashing for finding the best alignment between either atoms or feature points of two proteins evaluate the "goodness" of the match based on the overall RMSD [12]. In order to see if we could do better than RMSD, we evaluated each of the four here described methods. We tested how well each found the best base alignment and how well each filtered out active sites that did and did not bind ATP.

Overall RMS: First we evaluated how well the overall RMSD method picked the best base alignment of two segments of the same protein but of different sizes. To do this, we performed base pair alignment of our model and each of the following targets: 2HIX_all and 2HIX_less1 through 2HIX_less10. We expected that with all atoms present (2HIX_all) this method would be able to find the perfect alignment but that as atoms were removed and the number of overlapping points diminished, it would eventually fail. Our results verified our hypothesis. This method was able to find the perfect alignment with up to 7 atoms removed (47% removed).

Secondly, we evaluated this method based on how well it filtered out ATP-binding proteins from a pool of ATP-binding and -nonbinding targets. After aligning each of the 23 protein targets to the model, scores were computed and then used to rank the proteins (table 3.A). Not surprisingly, the best score was obtained by targets 2HIX_all through 2HIX_less7 (in that particular order), followed by 2HIX_less8 even though this method failed to find this protein's correct base alignment. The next best scoring protein was a 1A0I, which is structurally very closely related to 2HIX and which binds ATP. However, the scores of the remaining 12 proteins did not seem to follow any predictable pattern. Targets built based on 2HIX that had more than eight atoms removed scored lower than some of the unrelated proteins. Similarly, about half of the ATP-nonbinding proteins scored better than their ATP-binding counterparts (figure 4). These data indicate that this method does well only with highly conserved structures which most likely bind ligands in similar poses. Figure 3(a) shows an example of confirmations of ligands of the model and a closely related protein which was scored very well by this method.

Top 80% RMSD: We hypothesized that this approach would handle noisy data, such as missing data and imperfect fit, better and more consistently than the overall

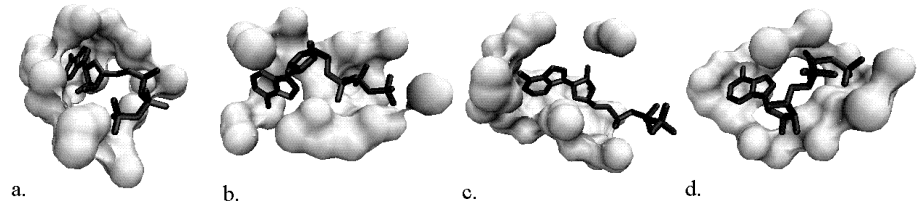


Fig. 2. Diversity of ligand interactions with active sites of proteins used to build and evaluate the model. Each shown active site includes atoms within 4Å of the ligand, ATP, which is shown as black, thick lines. **(a)** 2HIX – protein based on which the model was built; **(b)** 2EWW – used in voting to create model; **(c)** 1MO8 and **(d)** 2OOY were used for both model creation and method evaluation.

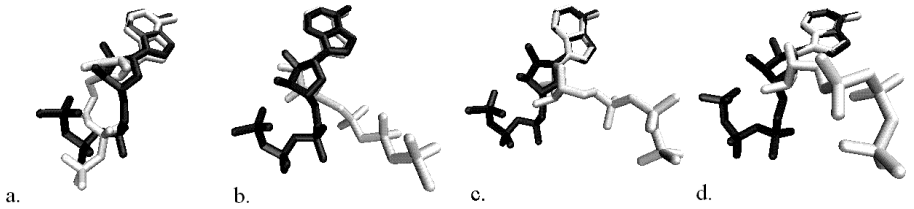


Fig. 3. Diversity of ligand binding conformations. The base portions of ligands were aligned; dark colored is ATP in its confirmation when docked in the active site of 2HIX, a protein based on which the model was created; light color shows ligands in their pose in active sites of proteins: **(a)** 1A0I – structurally closely related to 2HIX (note the similarity in ligand conformation), **(b)** 1V1B, **(c)** 2EWW, and **(d)** 2OOY. Proteins (a) and (b) were used only in evaluation; protein (c) just to build model; (d) used for both.

RMSD method. Surprisingly, it had the same threshold (7 atoms removed from the target) for finding the correct alignment. Similarly, it did equally well at filtering out structurally closely related proteins. It gave these proteins better scores than it gave to the 2HIX targets with more than 7 atoms removed. This suggests that it performs similarly well with missing data as it does with noisy data, but only if the model and target have high similarity (table 3.B). As was the case with overall RMSD, this method failed to distinguish the remaining ATP-binding from nonbinding targets. The rankings of the scores received by each target protein did not suggest that there is consistent preference for ATP-binding proteins (figure 4).

RMSD of Matched Atoms: This method was developed based on the assumption that not all atoms of the model were always required for binding. This was drawn from the fact that our model was built from a wide variety of ATP-binding proteins, and therefore, contained information relevant to several mechanisms. Because each ATP-binding protein had a somewhat different way of interacting with the ligand and hence anchoring via differently distributed atoms, the atoms contained in the model were unlikely to all be used simultaneously for binding. Based on this, we decided to remove the penalty for unmatched atoms that both the overall and top 80% RMSD methods had. The penalty originated from the inclusion in scoring, by both of these

methods, of atoms with no close fit. Each time a base alignment contained one or more such atoms, the total RMSD was negatively impacted. In order to avoid this we decided to compute RMSD only for atoms that did have a match. In order to ensure that the match is of significant length we added a constraint that at least a third of the atoms of the model have to have a close mate in the target.

We hypothesized that this method would perform well in finding partial regions with the tightest fit possible. And indeed, it did do a good job at performing perfect alignments between the model and the 2HIX targets with removed atoms. It was able to reliably find perfect fit for all of them, even the ones that contained as few as 5 of the atoms that were also contained in the models.

Moreover, this method did significantly better at filtering ATP-binding and -non-binding proteins. Out of the 24 target proteins that we tested, most that did use ATP as a ligand scored better than those that did not. There were two exceptions – one, 2OOY, an ATP-binding protein, obtained the lowest score of all of the test targets. The second exception was an ATP-nonbinding protein, 1Q5O. It obtained a score better than three of the ATP-binding proteins. It was also interesting to observe that it did not rank the structurally closely related proteins as highly as the other methods did. It did not give these proteins a score that was highest right after that of self-superset targets (2HIX-x) (table 3.C, figure 4). Investigation of ligand poses of proteins that scored well using this method revealed that it was able to handle a diversity of 3-dimentional confirmations. Figures 3(a) and 3(b) show poses of ligands of the two best scoring proteins.

Most Matched Atoms: Our last approach concentrated on evaluating bases based on the number of atoms that were matched. Our test verified our hypothesis that the tendencies of this approach were to pick alignments that found a "good enough" match for as many atoms as possible; it lacked sensitivity for shorter regions with very close match. This method turned out to perform quite similarly to overall RMDS. It was able to correctly align the model to the target with up to 8 of the atoms missing, at which point it ranked the target on par with other ATP-binding and -nonbinding proteins (table 3.D). Ranking of the scores did not result in any noticeably meaningful pattern and did not separate binders from non-binders (figure 4).

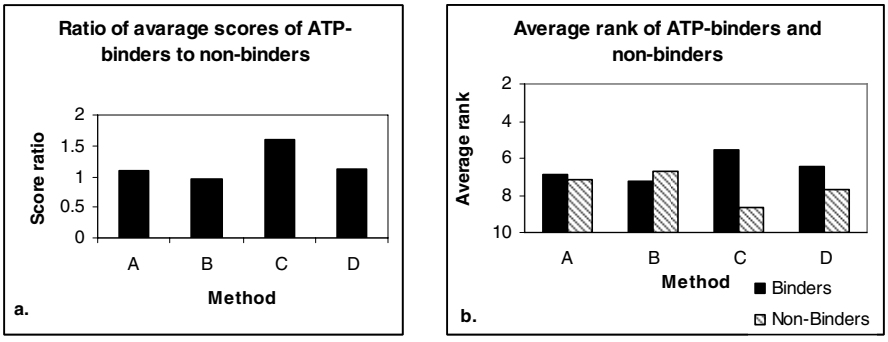


Fig. 4. Scores of ATP binders and nonbinders using method A: *Overall RMSD*, method B: *Top 80% RMSD*, C: *RMSD of Matched Atom,* and *D: Most Matched Atoms*. **(a)** Ratios of average scores obtained by ATP-binding proteins to those of ATP-nonbinding proteins. Ratio of 1 indicates that a binder cannot be distinguished from a nonbinder. **(b)** Average rank of ATP-binding and -nonbinding proteins.

Table 3. Results of evaluation of the four methods of base alignment scoring. Each sub-table contains ranked results of the matching between the model and each of the proteins listed in the left hand columns. Shaded rows contain proteins that do not bind ATP; clear rows contain proteins that do. Because scores in each of the sub-tables were computed using a different scoring technique, they should not be cross-compared. The cross-table scores by themselves are not informative enough to evaluate which method obtained a better alignment. This can only be done by further inspection. However, it is expected, in all categories, that proteins that do not bind ATP would score lower than those that do and, therefore, sorting by score would filter them down to the bottom of the table. Method (C), the *RMSD of Matched Atoms*, performs best.

A. Overall RMSD		B. Top 80% RMSD		C. RMSD of Matched		D. Longest Match	
protein	score	Protein	score	protein	score	protein	score
2HIX_less4	**27.99**	2HIX_less4	**93.22**	2HIX_less4	∞	2HIX_less4	∞
2HIX_less5	**20.24**	2HIX_less5	**42.02**	2HIX_less5	∞	2HIX_less5	∞
2HIX_less7	**11.763**	2HIX_less7	**16.994**	2HIX_less7	∞	2HIX_less7	∞
2HIX_less8	**9.86**	1A0I	**11.54**	2HIX_less8	∞	2J9L	**25.86**
1A0I	**7.99**	2HIX_less8	**10.62**	2HIX_less9	∞	1A0I	**18.32**
2HF4	**6.22**	2HIX_less9	**10.62**	1V1B	**70.42**	1C97	**17.86**
2HIX_less9	**6.08**	1V1B	**9.23**	1A0I	**58.77**	2HIX_less8	**16.81**
1VS0	**5.43**	2IPY	**8.97**	1XMJ	**41.67**	2HIX_less9	**16.81**
1V1B	**4.96**	2B3X	**7.61**	2J9L	**36.76**	1VS0	**16.16**
2IPY	**4.79**	1VS0	**7.60**	1Q5O	**34.48**	1Q5O	**15.90**
2B3X	**4.64**	1Q5O	**7.54**	1MO8	**30.49**	1V1B	**14.93**
1LB2	**4.07**	1XMJ	**7.07**	2HF4	**30.49**	2HF4	**14.44**
1XMJ	**3.48**	1MO8	**6.09**	1VS0	**27.37**	200Y	**13.55**
2J9L	**3.34**	2HF4	**5.28**	2IPY	**25.80**	1MO8	**13.49**
1Q5O	**3.27**	1LB2	**3.60**	2B3X	**22.73**	2B3X	**12.97**
1MO8	**2.41**	2J9L	**2.77**	1C97	**17.86**	1XMJ	**12.90**
1C97	**1.30**	200Y	**2.58**	1LB2	**17.21**	2IPY	**12.22**
200Y	**1.28**	1C97	**1.76**	200Y	**13.55**	1LB2	**10.99**

4 Conclusions

In this paper, we proposed and investigated multiple improvements to the robustness of geometric hashing based approaches to predicting ligand-protein interactions. We considered four methods of measuring the "goodness" of fit between a model and an active site. Our investigations demonstrate that the current most commonly used scoring technique – an *Overall RMSD* of the match – performs well with partial data and does well finding similarities between structurally closely related proteins. However, it performs poorly in terms of distinguishing between unrelated proteins that bind and those that do not bind ATP. Two of the other evaluation methods: *Top 80% RMSD* and *Most Matched Atoms*, also have similar problems. They were unable to predict which proteins did bind the ligand. The *RMSD of Matched Atoms* technique performed notably better. Ranking of scores received by various proteins revealed that this method performed better in distinguishing between binders and nonbinders.

The presented model of an active site was built utilizing a novel data-driven approach. Because it was built based on active sites of several, unrelated proteins it reflected properties of a diverse range of binding mechanisms. We postulate that the proposed approach and binding-site models derived from it present important

advantages in comparison to other methods in terms of their robustness to missing data, and their ability to handle variations in the 3D poses of ligands. Furthermore, being derived from a number of structurally diverse binding mechanisms, it provides a more general binding site definition than techniques that only analyze a pair of interacting molecules. Our future work is directed at comparing the quality of binding sites derived using the proposed method with those obtained using other techniques.

References

1. Brakoulias, A., Jackson, R.M.: Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. Proteins 56, 250–260
2. Comoglu, O., Kahveci, T., Singh, A.K.: Towards Index-based similarity search for protein structure databases. In: CSB 2003. Proceedings of the Computational Systems Bioinformatics (2003) 0-7695-2000-6/03
3. Coupez, B., Lewis, R.A.: Docking and Scoring – Theoretically Easy, Practically Impossible? Cur. Medicinal Chemistry 13, 2995–3003 (2006)
4. Edelsbrunner, H., Facello, M., Liang, J.: On the Definition and the Construction of Pockets in Macromolecules. Discrete Applied Mathematics 88, 83–102 (1998)
5. Laurie, A.T., Jackson, R.M.: Methods for the Prediction of Protein-Ligand Binding Sites for Structure-Based Drug Design and Virtual Screening. Current Protein and Peptide Science 7, 395–406 (2006)
6. Richard Jackson's group – ligand binding sites software, http://www.modelling.leeds.ac.uk/sb/
7. NCBI Protein Structure Database, ATP binding proteins: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?dopt=pccompound_structure&db=pccompound&cmd=Display&from_uid=5957
8. NCBI PubChem, ATP: http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=5957
9. PDB, http://www.rcsb.org/pdb/
10. Wang, J.T., Marr, T.G., Shasha, D., Shapiro, A.B., Chirn, G.-W.: Discovering active motifs in sets of related protein sequences and using them for classification. Nucleic Acids Res. 22(14), 2769–2775 (1994)
11. Visual Molecular Dynamics, Theoretical and Computational Biophysics Group, University or Illinois at Urbana-Champaign, http://www.ks.uiuc.edu/Research/vmd/
12. Rosen, M., Lin, S.L., Wolfson, H., Nussinov, R.: Molecular shape comparison in searches for active sites and functional similarity. Protein Engineering 11(4), 263–277 (1998)
13. Wolfson, H.: Geometric Hashing: an Overview. IEEE Computational Science and Engineering, 1070–9924 (1997)