Pacific Symposium on Biocomputing 7:673-684 (2002)

DETECTING NATIVE PROTEIN FOLDS AMONG LARGE DECOY SETS WITH HYDROPHOBIC MOMENT PROFILING

RUHONG ZHOU and B. DAVID SILVERMAN
IBM Thomas J. Watson Research Center
Route 134 & PO Box 218, Yorktown Heights, NY 10598

Abstract

A new hydrophobic score will be presented in this paper for detecting native-like folds from a large set of decoy structures. A recent paper (B. D. Silverman, PNAS 98, 4996, 2001) had revealed that for globular proteins there exists a relatively universal constant of 0.74 for a hydrophobic ratio, which is defined as the ratio of radii from the protein centroid at which the second order hydrophobic moment and the zero order moment vanishes. This paper further defines a new hydrophobic score which will be used to examine protein decoys, in particular, the Holm & Sander, Park & Levitt and Baker decoy sets. It will be shown that the hydrophobic score and profile shapes can provide useful information that should be complementary to the information provided by other procedures, such as free energy calculations.

1 Introduction

The ability to recognize native protein conformations from misfolded ones is a problem of fundamental importance in the development of methods for protein structure prediction. Decoy structures of proteins have provided test sets for the evaluation of scoring functions in threading and homology modeling, as well as energy functions used in ab-initio predictions. While an ideal objective would be the determination of a free energy function that selects structures that are native structures or minimally displaced spatially from the native structures, success has not been forthcoming. One suspects that the difficulty in the determination of an appropriate free energy function is related to the approximate manner in which the calculations treat the entropic character of solvation. One global structural feature arising from solvation is the ubiquitous hydrophobic core and hydrophilic exterior of soluble globular proteins. This feature has been used to identify protein structures that might be candidates that approximate the native structure or used to eliminate candidate structures that might not.[1–3] Considerations of hydrophobicity together with free energy approaches[4,5,8] can provide a more selective procedure than the use of either alone.

The universal spatial transition from the hydrophobic core to the hydrophilic exterior of globular proteins motivated detailed spatial profiling calculations of

Table 1: Eisenburg hydrophobicity consensus values for each amino acid.

res	hydro.	res	hydro.	res	hydro.	res	hydro.
ARG	-1.76	GLU	-0.62	TYR	0.02	TRP	0.37
LYS	-1.10	HIS	-0.40	CYS	0.04	LEU	0.53
ASP	-0.72	SER	-0.26	GLY	0.16	VAL	0.54
GLN	-0.69	THR	-0.18	ALA	0.25	PHE	0.61
ASN	-0.64	PRO	-0.07	MET	0.26	ILE	0.73

this transition. With an ellipsoidal characterization of protein shape, an appropriate scaling of residue hydrophobicity and a second-order ellipsoidal moment, it was shown that thirty or more diverse globular proteins shared detailed spatial features of this transition, with a quasi-invariant hydrophobic ratio (defined in next section) of 0.74 $\pm$ 0.05 for the protein structures examined. Since the small protein decoys of Park and Levitt, and those of the Baker group, have been central to ab-initio procedures in discriminating decoys from native structures, it is of interest to see if moment profiling could yield useful supplemental information, even in the regime of profile irregularities due to the small size of proteins. The Holm & Sander decoy sets, which include larger-sized proteins, are also examined. The results show that useful discrimination between native and decoy structures can be obtained.

2 Methodology

For proteins, each residue exhibits a different degree of hydrophobicity or hydrophilicity, based upon its solubility in water. A value of hydrophobicity, h_i, can then be assigned to each residue of type, i. Table 1 lists the Eisenberg hydrophobicity consensus values for each amino acid.[6]

Since the distribution of hydrophobicity is profiled from the protein interior to the exterior of globular proteins, an ellipsoidal profiling shape had been chosen with axes determined by the moments-of-geometry of the residue distribution,

$$I_{jk} = \int_V \rho(\vec{r})(r^2\delta_{jk} - x_j x_k)dV, \tag{1}$$

where I_{jk} is the moment-of-geometry terms and $\rho(\vec{r})$ is the density of the residue centroids of unit mass. Diagonizing the moment-of-geometry matrix, one obtains the three principal axes as well as the moments of geometry. The x, y, and z axes are then aligned with the principal axes. The moments of geometry are designated as g_1, g_2 and g_3, with $g_1 < g_2 < g_3$. The ellipsoidal representation generated by these moments will be,

$$x^2 + g_2' y^2 + g_3' z^2 = d^2, \tag{2}$$

where $g_2' = g_2/g_1$, $g_3' = g_3/g_1$. The value d is the major axis of the ellipsoid and can be considered as a generalized ellipsoidal radius.

The Eisenburg hydrophobicity distribution is shifted such that the net hydrophobicity of each protein vanishes. The distribution is then normalized to yield a standard deviation of one. Shifting the residue hydrophobicity distribution for each protein selects a common structural reference and thus enables the quantitative comparison of protein profile shapes and profile features such as the hydrophobic ratio. The zero-order hydrophobic moment of the residue distribution within the ellipsoidal surface specified by d is then written,

$$H_0(d) = \sum_{r<d} h_i^{'} = \sum_{r<d} (h_i - \bar{h}) / < (h_j - \bar{h})^2 >^{1/2}, \tag{3}$$

The prime designates the value of hydrophobicity of each residue after shifting and normalizing the distribution, and $\bar{h}$ is the mean of the h_i for all residues in the protein. Therefore, when the value of d is just sufficiently large enough to collect all of the residues, the net hydrophobicity of the protein vanishes. This value of d_0, for which $H_0(d)$ vanishes assigns a surface as common structural reference for each protein.

Second-order moments amplify the differences between hydrophobic and hydrophilic residues that contribute to the spatial profile of the hydrophobicity distribution. The second-order hydrophobic moment is defined as,

$$H_2(d) = \sum_{r<d} h_i'(x_i^2 + g_2' y_i^2 + g_3' z_i^2), \tag{4}$$

where the (x_i, y_i, z_i) denote the position of a residue centroids. For native globular protein structures, the zero and second-order moments are positive when d is small. Both increase with distance, d, within the region of the hydrophobic core. The increase of both the zero- and second-order moments with distance then slows and turns around decreasing with increasing d. Since the second order moment amplifies differences in the distribution, this moment will cross zero, becoming negative at a distance below the value of, d_0, at which the zero-order moment vanishes. The location at which the second-order moment vanishes is defined as d_2. The hydrophobic-ratio is then defined as,

$$R_H = d_2/d_0. \tag{5}$$

The paper by Silverman[7] had shown the hydrophobic-ratio to be a quasi-invariant, 0.74, for all of the native protein structures that had been examined. This ratio will also be shown to characterize native and near native structures in the following section. Such ratio, however, cannot always be defined for arbitrary

protein structures. This is particularly true if the second-order moment profile does not exhibit the smooth generic behavior expected. The hydrophobic ratio would then be unable to provide a continuous score with respect to how deviant a decoy profile would be with respect to its native profile. To provide such continuous ranking of each decoy profile with respect to its native profile, a new scoring function will have to be defined.

3 Results and Discussion

The Holm & Sander, Park & Levitt[2] and Baker decoy sets[8] examined in this study have been downloaded from the web (http://dd.standford.edu for the Holm & Sander and Park & Levitt set, and http://depts.washington.edu/bakerpg for the David Baker set). Since the hydrophobic moments and ratios involve the spatial profiling of the residue distribution, and this distribution is discretely distributed in space, a typical window of 1Å in generalized ellipsoidal radius, d, had been used to generate the nested ellipsoidal surfaces. This provided reasonable resolution in obtaining the generally smooth moment profiles over the range of protein sizes previously investigated. Protein size imposes a constraint upon the ability to generate relatively smooth profiles. It is found that a relatively smooth profile can be obtained for proteins with residue number greater than 100. This number of residues, namely 100, was the lower limit chosen in the investigation of the Holm Sander decoys, which resulted in a total of 14 decoy sets out of total 26, with protein size ranging from 107 residues to 317 residues. The Park & Levitt and Baker decoys range in size well below this limit so proteins chosen for the present study are limited to have a residue number of no less than 60.

For the Baker decoy sets, we have also applied two other criteria to eliminate decoy sets from the total of 92: (1) those decoy sets where 10% or less of the decoys have RMSD's from the native structure that are less than 8 Å will be eliminated; and (2) those decoy sets having the smallest RMSD larger than 4 Å will be eliminated. The objective is to examine decoys with a broad range of RMSD's and hence a broad range of "similarity" to their native structures. Decoys significantly displaced in RMSD from their native structures are not included. This selects the decoys that should be more difficult to distinguish from their native structure. This decoy set elimination together with the residue number limitation reduces the number of Baker sets studied to 11 from the total of 92. The residue number restriction imposed on the Park & Levitt decoy sets reduces the number of sets examined to 4 from a total of 7 (one decoy set has outdated native PDB structures, which has also been eliminated). The numbers of residues of these proteins range from 60 to 75. These protein sizes are insufficient, in most cases, to yield smooth hydrophobic moment profiles. It will,

however, be shown that even subject to this limitation, the moment profiling can provide useful complementary information to that obtained from energy minimization procedures. The RMSD values for the Park & Levitt decoy sets are supplied by the authors on their web site. These are RMSDs for C_α atoms. The RMSD values for the Baker decoy sets are not available from the web site and are, therefore, recomputed with the IMPACT program[9] for all backbone atoms. The RMSD values based on the C_α atoms, backbone atoms, or all of the atoms will be slightly different, but for the case at hand, they should be equally instructive.

All native second-order moment profiles of the proteins in Holm & Sander, Park & Levitt, and Baker decoy sets selected here (total 29) show a hydrophobic core and a sharp plunge to negative values in the transition from hydrophobic core to hydrophilic exterior. Similar to previous results,[7] the native decoy structures have R_H values that range from 0.640 to 0.791, with a mean of 0.73.

Holm Sander decoys had been generated to test their solvation preference method[1] designated to distinguish native(correct) from decoy(incorrect) structures. Figure 1 shows the second-order hydrophobic moment profiles for 6 such decoys (all 14 decoys show the same behavior basically). All native structures exhibit a second-order profile shape that had been previously found for thirty PDB structures of diverse size and fold. All of the decoy structures, on the other hand, do not show the significant separation between the hydrophobic residues forming the native core and hydrophilic exterior. The second-order moments fluctuate around zero on the radial axis. The hydrophobic ratio cannot be defined for these decoy structures.

The second order moment profiles of the thousands of Park & Levitt and Baker decoy structures do not, however, always exhibit easy patterns to be discriminated against as in the Holm & Sander single decoy sets. It is also not feasible to plot thousands of profiles in one or a few figures. Therefore, a new scoring function is needed to quantitatively rank each decoy profile with respect to an expected native profile. Examination of a few of the decoy profiles will reveal several of the issues involved in defining the scoring function. In general, the structures of smaller RMSD ($< 2.0\mathring{A}$, native-like) approximate the native profile more closely. The decoy structures with larger RMSD have hydrophobic peaks that are either not well-defined as shown in Holm & Sander's decoy profiles (Figure 1) or less pronounced than their native structures. The hydrophilic ranges are also generally extended out to greater distances and are not as negative as seen in the native structure profiles. This suggested that the area under the hydrophobic peak and that under hydrophilic exterior could play a role in discriminating the native from the decoy structures. On the other hand, a significant increase in protein extent of the decoy could yield a spurious contribution from the area under the negative moment profile. This contribu-

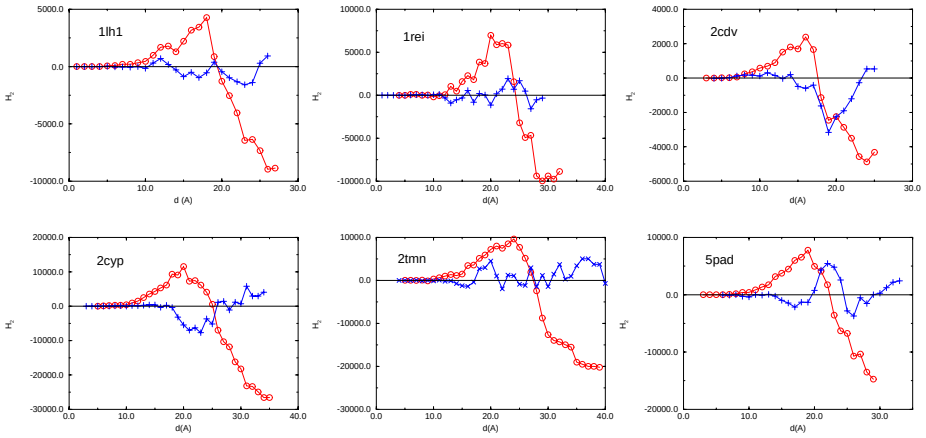

Figure 1: Second-order moments for the native and decoy structures of the Holm & Sander single decoy sets (circles: native; plus: decoy).

tion could, however, be reduced by scaling the native and decoy structures by the value of protein extent, namely, by d_0. The abscissa on the moment plot was, therefore, divided by d_0 and the second-order moment divided by d_0^2. Such comparison does not take differences in residue number into account. For the present case, however, the decoys and their corresponding native structures have the same number of residues.

The proposed hydrophobic score, S_H, which ranks the quality of the decoys with respect to an expected native profile, is then chosen as the integral of the area under the normalized second-order hydrophobic moment profiles,

$$\begin{aligned} \tilde{H}_2 &= H_2/d_0^2 \\ \tilde{r} &= r/d_0. \end{aligned} \tag{6}$$

The absolute value of $\tilde{H}_2$ is integrated over the normalized distance,

$$S_H = \int_0^1 |\tilde{H}_2| d\tilde{r}. \tag{7}$$

This score not only measures the prominence of the hydrophobic core, but also the prominence of the hydrophilic exterior by taking into account the rapidity of decrease of the profile in the hydrophilic exterior.

Figure 2 shows the hydrophobic scores versus the RMSDs for the four Park & Levitt decoy sets. Almost all decoys have lower hydrophobic scores than their corresponding native structures. Table 2 lists the number and percentage

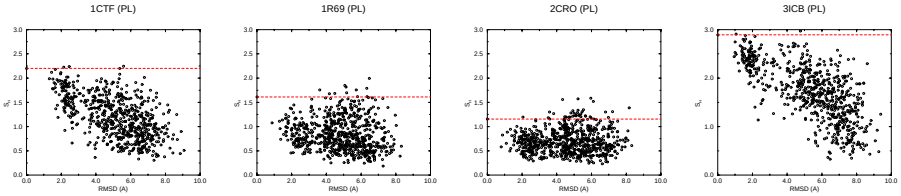

Figure 2: Hydrophobic score versus RMSD for Park & and Levitt decoys.

Table 2: Performance of the hydrophobic score: the percentage of decoy structures which have lower hydrophobic score than their native ones (denoted as "low scores" in the table).

Decoy Set	PDB entry	low scores	total decoys	%
Park/Levitt	3icb	651	654	99.5
	1ctf	627	631	99.4
	1r69	664	676	98.2
	2cro	637	675	94.4
Baker	2ezh	957	1000	95.7
	1mzm	864	1000	86.4
	1nkl	848	1000	84.8
	1ctf	816	1000	81.6
	1r69	656	1000	65.6
	2fow	627	1000	62.7
	2ptl	619	1000	61.9
	1sro	559	1000	55.9
	1c5a	493	991	49.8
	1hsn	245	970	25.4
	1leb	253	1000	25.3

of decoys out of the total which have lower hydrophobic scores than their native proteins. 99.5%, 99.4%, 98.2%, and 94.4% of the decoys have hydrophobic scores below their native benchmarks for 3icb, 1ctf, 1r69, and 2cro, respectively. Proteins 3icb and 1ctf which show native profiles accentuating the hydrophobic and hydrophilic regions have fewer than 0.5–0.6% of decoys with a score that is greater than that of the native structures. One also notes significant correlation in their decoy distributions, namely, decoys with greater RMSD generally have smaller hydrophobic area or score. Proteins 2cro and 1r69 with native profiles that do not accentuate the hydrophobic and hydrophilic regions as observed for proteins 1ctf and 3icb show slightly greater numbers of decoys with greater scores than their native structures, and their distribution of decoy scores does not exhibit the correlation found for 1ctf and 3icb. The decoy scores of 1r69 and 2cro appear to be essentially uniformly distributed about the RMSD values.

Little or no correlation of hydrophobic score with RMSD might arise from

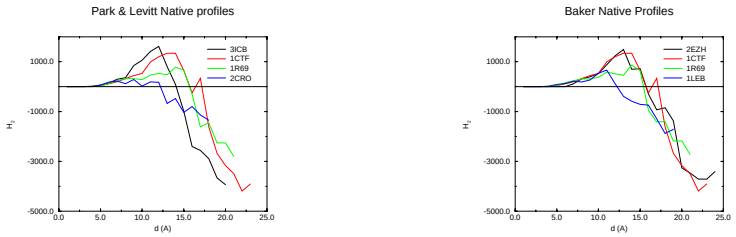

Figure 3: (a) Left: the four native structure profiles in the Park & Levitt decoy set, 3icb, 1ctf, 1r69, and 2cro. (b) Right: the four native structure profiles in the David Baker decoy set, 2ezh, 1ctf, 1r69, and 1leb.

native structures with profiles that do not accentuate the core and hydrophilic regions. It is then less restrictive for a decoy to score well with respect to the native structure. Figure 3(a) shows the native profiles of the four decoy sets of Park & Levitt, namely, 3icb, 1ctf, 1r69 and 2cro. It is clear that 1r69 and 2cro have native profiles with hydrophobic and hydrophilic regions of lesser prominence than found for 1ctf and 3icb.

Figure 4 shows the hydrophobic scores for the four representative Baker decoy sets: two decoys 1ctf and 1r69 which are also in the Park & Levitt set, and the other two 2ezh and 1leb which have the highest and lowest percentage of decoys with scores below their native structure scores respectively (see Table 2). In contrast to the Park & Levitt decoy sets, the Baker decoy sets show a much broader distribution of hydrophobic scores. The percentage of decoys which have scores below their native benchmark scores ranges from 25.3% (1leb) to 95.7% (2ezh), with the majority in the range of $60 - 80\%$. Also, most of these decoy sets do not exhibit the correlation with RMSD that the 1ctf and 3icb Park & Levitt decoys show. The four shown decoy sets 2ezh, 1ctf, 1r69, and 1leb have a percentage of decoys with scores below the native to be 95.7%, 81.6%, 65.6% and 25.3%, respectively. Interestingly, 2ezh and 1ctf (higher percentages 95.7% and 81.6%), show a more prominent native structure profile than 1r69 and 1leb (lower percentages 65.6% and 25.3%), as can be seen from the Figure 3(b). Other decoys in the Baker set show similar behavior. The numbers of decoys with a higher percentage below the native score (2ezh, 1mzm, 1nkl, 1ctf, etc) show more pronounced native structure profiles than decoys with a lower percentage (1hsn, 1leb, etc). As discussed earlier, for the Park & Levitt decoys 1r69 and 2cro, this correspondence between a higher percentage scoring poorly with the less prominent native profiles appears reasonable. It is easier for decoys to score well against native structures that exhibit reduced separation of hydrophobic and hydrophilic residues with consequent low score.

The relatively large number of Baker decoys with high hydrophobic score

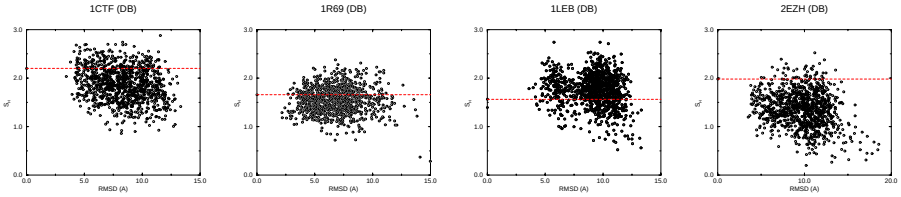

Figure 4: Hydrophobic score versus RMSD for Baker decoys.

compared with the Park Levitt decoys might be related to the manner in which the decoys were generated and selected. In particular, a significant fraction of the decoys of 1leb Baker set clearly show greater spatial segregation of the hydrophobic and hydrophilic residues than observed for the native structure. The scores of the 1r69 and 1ctf Park & Levitt and Baker decoy sets can be compared from Figure 2 and Figure 4. The Baker decoys clearly show a greater number of structures with scores that are higher than their native scores when compared with the Park & Levitt decoy scores. Calculation of the radii of gyration (Rg) found the Baker sets to have slightly larger Rg's compared with the Park and Levitt sets. Decoys of 1r69 have Rg of 12.00 ± 0.81 Å in Baker set and 10.99 ± 0.53 Å in the Park & Levitt set. Similarly, decoys of 1ctf have Rg of 11.65 ± 0.66 Å in the Baker set and 11.19 ± 0.59 Å in Park & Levitt set. Perhaps larger radius of gyration provides greater spatial freedom to segregate the hydrophobic from hydrophilic residues. A point of greater relevance is related to the way Baker and coworkers have selected these ab-initio decoys. One of the fundamental assumptions underlying their program Rosetta[3,8] is that the distribution of conformations sampled for a given nine residue segment of the chain is reasonably well approximated by the distributions in known protein structures in the PDB Databank. Fragment libraries for each 3- and 9-residue segment of the chain are extracted from the protein structure database using a sequence profile-profile comparison method. The conformational space defined by these fragments is then searched using a Monte Carlo procedure with an energy function that favors compact structures with paired β strands and buried hydrophobic residue. The favoring of buried hydrophobic residues in the energy function should provide the Baker sets with greater segregation of hydrophobic and hydrophilic residues from the protein core to exterior and consequently provide higher hydrophobic scores than achieved by the Park & Levitt decoy sets.

It is also interesting to note that there are low RMSD structures which have low hydrophobic scores even among the decoys of the well correlated sets, such as 3icb. Figure 5(a) shows several hydrophobic moment profiles for 3icb decoy

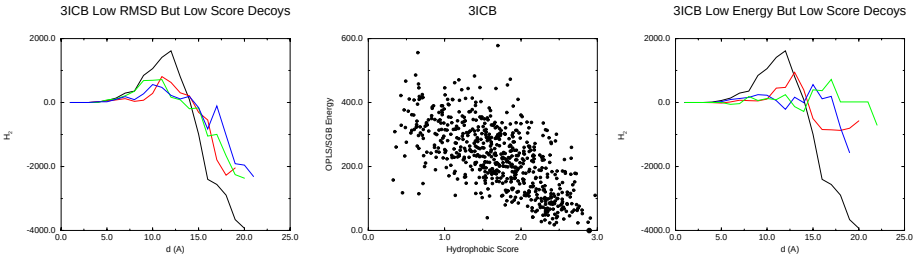

Figure 5: For Park & Levitt decoy set 3icb: (a) left: Hydrophobic moment profiles for some low RMSD structures but with low hydrophobic scores (the thick dark line is from the native structure for comparison). (b) Middle: hydrophobic score versus OPLSAA/SGB energy[5] (the native one is marked with a bigger circle). (c) Right: Hydrophobic moment profiles for some of the low OPLSAA/SGB energy structures but with low hydrophobic scores.

structures with less than a 3.0Å RMSD and less than a 1.5 hydrophobic score (decoy index a587, a591, and a8110, to name a few). The native score is 2.89 for this case. These decoy structures have fewer hydrophobic residues in the protein interior and consequently fewer hydrophilic residues in the protein exterior than expected for native structures. The hydrophobic residues and hydrophilic residues are more spatially mixed. Might these structures be less favorable candidates as near native structures? From the reported OPLSAA/SGB free energies,[5] they are indeed energetically unfavorable structures. The three decoys plotted, a587, a591, and a8110, are 206.98, 116.94, 110.14 kcal/mol higher than the native structure. The OPLSAA/SGB energies are obtained from Levy and coworkers (see below for more details). This indicates that the simple hydrophobic score should provide useful information in discriminating decoy structures from native structures.

Levy and coworkers[5] have calculated the energies of the Park & Levitt decoys using the OPLSAA force field[10] and a Surface Generalized Born (SGB) model[11] for a continuum solvent. They found that without the continuum solvation free energy, the OPLSAA gas phase energies are not sufficient to distinguish native-like from non native-like structures. Figure 5(b) plots the OPLSAA/SGB energy (the energy of the native structure is set zero) versus the hydrophobic score for the protein 3icb of the Park & Levitt set. The OPLSAA/SGB energies have been kindly supplied by the Levy group. It should be noted that in the Levy energy calculations, the decoy structures are minimized first to remove bad contacts (otherwise the energies could be huge and meaningless). Thus, the structures used in the Levy energy calculations are slightly different from ours; however, we don't believe that this should affect the hydrophobic scores. This is an advantage of the method of hydrophobic scoring. Differences in structure that would affect the free energy values significantly will not affect the hydrophobic

scores much. One need not even add hydrogen atoms to the PDB structures for most of the calculations. Free energy calculations, on the other hand, are not only sensitive to the presence or absence of hydrogen atoms, but extremely sensitive to small differences in structure. Figure 5(b) shows the correlation between the OPLSAA/SGB energy and the hydrophobic score, i.e., decoys with smaller or poorer scores have higher energies compared with the native energy, and those with higher or better scores are closer in energy to the native structures. Similar to 3icb, protein 1ctf also shows a significant correlation between the OPLSAA/SGB energy and the hydrophobic score, whereas 1r69 and 2cro show a weaker correlation. This weak correlation for the 1r69 and 2cro decoys reflects their weak correlation between hydrophobic score and RMSD as described previously.

Interestingly, there are decoy structures with low OPLSAA/SGB free energies that do not have high hydrophobic scores. This is found even for the decoys of 3icb, which show a strong correlation between hydrophobic score and RMSD. The decoy sets showing poorer correlation have a greater number of decoys exhibiting this behavior. Figure 5(c) also shows several representative profiles of 3icb decoy structures with low free energies and also low hydrophobic scores. These decoys are not the same as those with low RMSD and low score as discussed previously and shown in Figure 5(a). The bad or low hydrophobic scores indicate that the structures have a poorly formed hydrophobic core and hydrophilic exterior even though the free energy is low. By comparison with the native profile (the thick dark line in the figure), it is evident that the hydrophobic core of these decoys has been "damaged". The region of positive moment that might be identified as a core region is shifted out to greater distances than found for the native structure. Furthermore, none of the decoys exhibit the sharp plunge to negative values in the protein exterior expected for a native structure. This yields a low score or unfavorable protein structure. This example demonstrates the value of the hydrophobic score in providing complementary information to that obtained from the free energy calculations. Previously we had shown that a low RMSD does not necessarily guarantee a good hydrophobic score, and here we have shown that a low free energy does not guarantee a good hydrophobic score either.

It should be pointed out that the resulting second-order profiling pattern, hydrophobic ratio and hydrophobic score are for globular proteins only. Other type of proteins, such as DNA-binding proteins and membrane proteins might have hydrophobic residues in the exterior region, thus the profiling and ratios will be different.

This approach should be very useful for pre-screens in various structure prediction or refinement algorithms, since it is extremely fast. It takes less than a minute on a typical PC to calculate the second-order moment profiling and

the hydrophobic score for each structure.

4 Conclusions

The present paper has examined the utility of molecular moment hydrophobicity profiling in discriminating between near native protein structures and incorrect decoy structures for the widely used Holm & Sander, Park & Levitt and Baker decoy sets. Subject to the conditions that limit the type of small structures examined, the moment profiling and a subsequent hydrophobic score, which is the integral of the area under the normalized second-order hydrophobic moment profile, enable one to distinguish the decoys from near native structures of globular proteins. It is also found that the hydrophobic score can suggest that certain structures with small RMSD from the native structure should be eliminated as candidates due to profiles displaced significantly from their native hydrophobicity profiles. Interestingly, some decoys with low free energies, such as OPLSAA/SGB energy, can also be eliminated by the hydrophobic moment profiling and consequent hydrophobic score, since they show little or no hydrophobic core and hydrophilic exterior compared with their native profiles. This shows that the simple hydrophobic score can provide information that complements that obtained by the more rigorous free energy approach.

References

1. L. Holm and C. Sander. *J. Mol. Biol.*, 225:93–105, 1992.
2. E. S. Huang, S. Subbiah, and M. Levitt. *J. Mol. Biol.*, 252:709–720, 1995.
3. R. Bonneau, C. E. M. Strauss, and D. Baker. *Proteins: Structure, Function and Genetics*, 43:1–11, 2001.
4. B. Park and M. Levitt. *J. Mol. Biol.*, 258:367–392, 1996.
5. A. K. Felts, A. Wallqvist, E. Gallicchio, R. Levy, D. Bassolino, and S. R. Krystek. submitted, 2001.
6. D. Eisenberg, R. M. Weiss, and T. C. Terwilliger. *Nature (London)*, 299:371–374, 1982.
7. B. David Silverman. *Proc. Natl, Acad. Sci. USA*, 98:4996–5001, 2001.
8. K. T. Simons, R. Bonneau, I. Ruczinski, and D. Baker. *Proteins: Structure, Function and Genetics*, 37 S3:171–176, 1999.
9. F. Figueirido, R. Zhou, R. Levy, and B. J. Berne. *J. Chem. Phys.*, 106:9835, 1997.
10. D. Maxwell W. L. Jorgensen and J. Tirado-Rives. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
11. A. Ghosh, C. S. Rapp, and R. A. Friesner. *J. Phys. Chem.*, 102:10983–10990, 1998.