BIOINFORMATICS

Vol. 00 no. 00 2007
Pages 1–7

Four-Body Scoring Function for Mutagenesis

Chris Deutsch [a] and Bala Krishnamoorthy [a]

[a]Department of Mathematics, Washington State University

ABSTRACT

Motivation: There is a need for an efficient and accurate computational method to identify the effects of single and multiple residue mutations on the stability and reactivity of proteins. Such a method should ideally be consistent and yet applicable in a widespread manner, i.e., it should be applied to various proteins under the same parameter settings, and have good predictive power for all of them.

Results: We study the applicability of the Delaunay tessellation-based four-body scoring function for predicting the effects of single and multiple residue mutations on the stability of proteins. Unlike previous studies, we test the scoring function on a diverse set of single and multiple residue mutations, collected from over twenty four different studies. Out of the 236 single/multiple residue mutants analyzed, the four-body scoring function correctly predicted the changes to the stability or reactivity of 186 (79%) of the proteins. For the mutants which had the changes in stability/reactivity quantified (using reaction rates, temperatures etc.), an average Spearman rank correlation coefficient of 0.67 was achieved with the four-body scores. We also develop an efficient method for screening huge numbers of mutants of a single protein, called *combinatorial mutagenesis*. In one study, 64 million mutants of a cold-shock nucleus binding domain protein 1CSQ, with six of its residues being changed to all possible (20) amino acids, were screened within a few hours on a PC, and all five stabilizing mutants reported were correctly identified as stabilizing by combinatorial mutagenesis.

Availability: The full list of mutants, and executables of programs developed as part of this study are available from the author's web page: http://www.wsu.edu/∼kbala/Mutate.html.

Contact: kbala@wsu.edu

1 INTRODUCTION AND PREVIOUS WORK

Mutagenesis is the process of replacing one or more amino acids in a wild-type (WT) protein by alternate amino acids to generate a mutant protein. The goal of the process is to create a protein with certain desirable biochemical properties that are lacking in the WT protein. For instance, a protein that is more reactive, or more stable, in a particular reaction than the WT can often be generated by altering the identity of a single key residue (to generate a single-point mutant). Mutagenesis finds applications naturally in protein design, drug discovery, and other similar areas (see Section 2.2 for a listing of many applications of mutagenesis).

The experimental process of creating mutants could often be expensive and time-consuming. To start with, it is often not straightforward to identify the key residue(s) that need to be mutated in order to achieve the desired biochemical properties. Once the critical residue positions are identified, it could still be non-trivial to decide what the new amino acids should be. Hence biochemists often end up having to create and analyze a large number of mutants in order to identify a handful of desirable ones. In a typical example (1CSQ, one of the proteins included in our research), six amino acid positions were identified as desirable mutation sites. The experimentalists wanted to try all possible alternate amino acid combinations for these six residues, changing at least one amino acid in each case. The total number of single- and multiple-point mutants that they could have considered is 64 million ($20^6 - 1$ to be exact)! Of course, they only tried a few hundreds of the mutants, and reported five of them that had the desired properties. As illustrated by this example, it is quite desirable to use computational methods to reduce the number of mutants that need to be generated experimentally.

One of the key steps in most protein structure prediction methods is the screening of multiple candidate conformations to select the best one(s), and a scoring function is used for this purpose. Scoring functions have been used for protein fold recognition for several years – many of them are studied in the following references (Miyazawa and Jernigan, 1985; Sippl, 1990; Park and Levitt, 1996; Krishnamoorthy and Tropsha, 2003). One could possibly study the use of any such scoring function for the purpose of virtual mutagenesis – score the original (wild-type) protein, and then score the original protein after making the proposed changes to the sequence while keeping the structure unchanged. We could then correlate the change in score (i.e., difference between the two scores as calculated above) and the effects of the mutations on various properties of the original protein, thus developing a *predictor* for the effects of mutations. In spite of the apparent simplicity, only a few such studies have been undertaken so far. Carter et al. (2001) obtained high correlations between changes in the four-body scores and the free energy changes (measured as $\Delta\Delta G$ values) resulting from mutations to residues in the hydrophobic cores of five different proteins. More recently, Masso, Lu, and Vaisman (2006) used the same four-body scoring function to study the structure-function correlations of the mutants of for HIV-1 protease and T4 lysozyme. Korkegian et al. (2005) used computational techniques to determine the best mutations that increase the thermostabilization of an enzyme.

Our main goal in this study is to develop and test a scoring function to predict the changes to the stability and reactivity of proteins due to mutations that is robust (i.e., applicable to a wide variety of proteins) and accurate. One of the main reasons for the lack of such scoring functions is the difficulty in obtaining experimental mutagenesis data. We have performed an exhaustive search of the literature to identify a comprehensive list of proteins and their mutants, along with experimental data that quantifies the change in stability or reactivity of the WT protein. This test set of 236 mutants is collected from twenty four different experimental studies. The scoring function we use is developed from the four-body scoring function that is based on the Delaunay tessellation of proteins. The latest (and most accurate) version of the four-body scoring function as used for protein decoy discrimination was proposed by the author previously

(Krishnamoorthy and Tropsha, 2003), and has been tested exhaustively on many test sets of decoys. Though two previous studies (Carter et al., 2001; Masso et al., 2006) used the four-body scoring function to analyze mutagenesis, they both tested the scoring function only on a small set of proteins. The first of these two studies considered mutations that are made only in the hydrophobic core of five proteins. The second study considers most possible mutations for two different proteins. Further, they both use an older version of the scoring function, and the second study uses different settings for scoring the mutants of the two proteins considered. We address these and other shortcomings when defining our four-body scoring function. We correctly predict the effects of the mutations (single- or multiple-point) for 186 out of the 236 mutants considered (79 %). For several sets of the mutants, the original authors had quantified the effects on the WT proteins by providing melting temperatures, reaction rates, free energy changes, or other appropriate data for the mutants. For these sets of mutants, we obtained an average Spearman rank correlation coefficient of 0.67 between the quantifying data and the four-body scores. We also propose an efficient method to evaluate huge number of mutants by working with only the Delaunay tetrahedra that the mutated residues participate (as opposed to scoring the WT protein repeatedly).

2 METHODS AND MATERIALS

We describe the details of the four-body scoring function used for mutagenesis, and the test set of single- and multiple-residue mutants that we have assembled and applied our scoring function on.

2.1 Four-body scoring function

The idea of a scoring function for protein fold recognition that is built on the Delaunay tessellation of proteins was first proposed by Tropsha et al. (1996). Singh, Tropsha, and co-workers subsequently developed the scoring function (Munson and Singh, 1997; Tropsha et al., 1998), and also explored the possibility of using the same for *ab initio* protein folding (Gan et al., 2001). The formulation of the scoring function was improved by Krishnamoorthy and Tropsha (2003), and the applicability for decoy discrimination was tested on various decoy sets. Further extensive testing of the scoring function has been conducted recently by Krishnamoorthy et al. (Fowler et al., 2007; Krishnamoorthy and Stratton, 2007). Since we use the version of the scoring function given in this paper for scoring mutations, we reproduce this formulation here, which defines the scoring function as a log-likelihood ratio:

$$Q^{\alpha}_{ijkl} = \log\left[\frac{f^{\alpha}_{ijkl}}{p^{\alpha}_{ijkl}}\right]. \qquad (1)$$

$i, j, k,$ and l represent the residue identities of the four amino acids (20 possibilities) in a Delaunay tetrahedron from the tessellation of the protein. Each amino acid is represented by a single point located at the centroid of the atoms in its side-chain (including the back-bone C_α atom). α represents the *type* of the tetrahedron based on the back-bone chain connectivity of the four participating amino acids. There are five tetrahedron types possible, and α takes one of the values $0, 1, 2, 3$ or 4 corresponding to these types. The total score (or simply, the *score*) of a protein is then defined as the sum of the log-likelihood ratios of all tetrahedra in its Delaunay tessellation. A cut-off value of 10 Å (Angstroms) was used for the length of any edge of the tetrahedra that are scored, thus eliminating biochemically irrelevant tetrahedra with huge edge lengths from consideration. Simply put, the score of a protein gives a measure of how *well-packed* its residues are (hence it was also called the Simplicial Neighborhood Analysis of Protein Packing (SNAPP) score). Further, the correct way to interpret the score is in a *relative* sense, i.e., we can compare the scores of two otherwise similar conformations to quantify how better the packing of one structure is compared to that of the other.

The back-bone chain connectivity of the tetrahedra was not considered in the previous version of this scoring function when Carter et al. (2001) correlated the changes in total score of proteins to the free energy changes resulting from mutating one or more amino acids in the WT protein. The more recent study by Masso et al. (2006) also ignores this factor in the form of the four-body scoring function used. Further, in both these studies, there is some ambiguity regarding the choice of side-chain centers of residues versus back-bone C_α atom coordinates that should be used to represent each amino acid. If we are proposing a scoring function that can be used for various proteins, it is not desirable to change the settings and other parameters when scoring each individual protein. In fact, Masso et al. use two different settings for the two proteins that they study. The justification they give is the robustness of the four-body score of a protein under small perturbations of the points representing each amino acid. The claim made is that total score of a protein does not change *by much* when the representation of the amino acids is changed from C_α to side-chain centers. The question of robustness of the Delaunay tessellation of proteins (and point sets in general) was addressed by Bandyopadhyay and Snoeyink (2004) – they defined the concept of *almost Delaunay* simplices, where the positions of the points defining the simplex are allowed to vary in a controlled range (as opposed to being fixed). The four-body scoring function was tested under the almost-Delaunay setting to obtain decoy discrimination results that are roughly comparable to those obtained by the original scoring function (as described in Equation (1)). Still, the results obtained using the side-chain center representation were markedly better than those obtained using C_α representation by Krishnamoorthy and Tropsha (2003) – to such an extent that the C_α results were not reported at all. Hence it was suggested that side-chain centers be used always in order to obtain the most accurate results. This suggestion has been further validated by recent results obtained by Krishnamoorthy et al. (Fowler et al., 2007; Krishnamoorthy and Stratton, 2007).

Another factor critical to the robustness of the four-body scoring function is that this result holds for the *total score* of the protein, and might not hold for the case of *change* in the total score – especially when the change is small. Several of the mutants that have been analyzed previously, and in this paper, are single-point mutations. The identity of a single residue is changed, and hence only a small subset of the full set of Delaunay tetrahedra are affected by such changes. The change in the total score in this case might be sensitive to the way the residues are represented, and also to perturbations in the positions of these residues under a single representation. The residues that are in the inner portions of the protein (buried) participate in many more tetrahedra than those that are on the outside (surface), and hence the robustness result might apply more for the case of the buried residues. All mutations studied by Carter et al. (2001) are performed on hydrophobic core residues. On the other hand, many mutations that we considered involve changes to surface residues,

each of which typically does not participate in a large number of Delaunay tetrahedra. Hence we suggest the consistent use of side-chain centers when scoring mutations using the four-body scoring function. Similarly, we do not use weights for the scores of different classes of tetrahedra (Krishnamoorthy and Tropsha, 2003), as these weights are more appropriate for scoring the entire protein and not for individual, or a subset of, the tetrahedra.

We also observed that some key long-range interactions between amino acids are missed out by the use of a 10 Å cut-off on the Delaunay edges, especially for the case of surface residues. Hence we use an increased cut-off of 12 Å when scoring mutations using the four-body scoring function. Notice that the contacts made by most buried residues remain unchanged from before, as most such contacts are well within the 10 Å range. At the same time, several key interactions of surface residues that are left out by the 10 Å cut-off are now included in the calculations, thus making the scoring function more accurate for the purpose of virtual mutagenesis.

Under the settings described above, we calculate the change in total score between the WT and the mutant protein (mutant score - WT score). A positive change (i.e., the mutant score is more than the WT score) indicates that the mutant is more stable (equivalently, less reactive) than the WT, while a negative change indicates lower stability or increased activity of the mutant.

2.2 A test set of mutants

In order to test the four-body scoring function for mutagenesis, we needed to assemble a set of mutations that were analyzed experimentally for changes in stability or reactivity. Even though the technique of changing one or more residues has been used for various purposes, the results from these studies have not been collected together, or presented in a unified manner suitable for our purpose. We have searched the literature to identify a comprehensive list of such mutations. Overall, there are 236 mutants taken from 24 different papers. 170 of the mutants had decreased stability (or increased activity) than the WTs, while the remaining 66 mutants were reported to have increased stability. Many of the papers provide only qualitative information regarding the properties of the mutants – they report whether the mutant has higher (or lower) stability or reactivity than the WT protein. On the other hand, several of the studies also quantify the changes effected by the mutations by providing reaction rates, free energy changes, or melting temperatures as appropriate for the WT as well as the mutant proteins. The whole data set, along with the performance of the four-body scoring function on the mutants, is presented in Table 1. We describe the various types of mutations assembled briefly.

One group of mutations either destabilized or increased the reactivity of the WT protein. For instance, Bonander et al. (2000) reported destabilization of azurin protein due to the destruction of disulfide bonds (with cystines mutated to other residues). de Antonio et al. (2000) analyzed the effect of mutating tryptophan residues (to phenylalanines) on the functional properties of the ribotoxin α-sarcin, and reported decreased thermostability. Takano et al. (1999) observed destabilization of human lysozyme when foreign terminal residues were introduced, due to the disruption of hydrogen bonding. Ribonucleotide reductase mutations considered by Ormö et al. (1995) had similar destabilizing effects. Chen and Gouaux (1997) achieved increased reactivity in an aqueous solution of bacteriorhodopsin by introducing polar residues in the enzyme surface in Helix D. Carter et al. (2001) presented a fairly large data set of mutants (76 mutants of five proteins) collected from various previous studies. All of them were destabilizing substitutions made to one or more residues in the hydrophobic core of the proteins in question. We did not include three mutants of T4 lysozyme (PDB code 1L63) from our test set. One of them considered "replacing" residue 71, which is a Valine, with a Valine itself (it is not clear how a SNAPP score was recorded by the authors for this "non-mutant"). Then there were two different sets of data recorded for mutating residue 71 to an Alanine. Even though we identified this change (V71 to A) correctly as a destabilizing one (as indicated by the mutant having a smaller four-body score), it is ambiguous which $\Delta\Delta G$ value to use. Hence we decided to leave out both the V71A mutants. Erwin et al. (1990) engineered salt bridges in subtilisin BPN, resulting in the protein being destabilized. We must note here that a few mutations from this study were not included in our list, as they resulted in mutually disagreeing effects – for instance, these mutants had increased stability and increased enzyme reactivity as well. Lastly, mutations from a study on evolutionally conserved aspartic acid residues in human glutathione S-transferase P1-1 (Kong et al., 1993) were chosen due to the authors reports of decreased thermostability.

Another category of mutants included in our list had stabilizing effects on the WT protein, or decreased its chemical activity. Martin et al. (2001) analyzed mutants that increased the thermostability of a cold shock protein. Pathange et al. (2006) introduced histidine residues on the surface of a protein in order to decrease its enzymatic activity. Hahn et al. (1995) analyzed mutants of glucanase that showed decrease enzymatic activity. In one interesting case, Korkegian et al. (2005) achieved thermostability of cytosine deaminase with the aid of computational techniques. Similarly, Almog et al. (2002) studied thermostabilized mutations of subtilisin bpn. Siadat et al. (2006) engineered disulfide bonds in Drosophila melanogaster acetylcholinesterase to stabilize the it. However, some mutants from this study were not included due to multiple disagreeing effects.

A third group of mutants were selected from papers that reported some increased stability (or reactivity) mutants and some other mutants that showed the opposite effect(s) for the same WT protein. For instance, Ge et al. (2003) studied the activity of a rice lipid transfer protein. We included those mutants that showed increased enzymatic activity. At the same time, mutants whose secondary structure was reported to be destroyed were categorized as destabilizing as well. Certain other mutants were not included in our set due to unspecified "similar activity" to wild type. Suresh et al. (2006) analyzed the effects of C-reactive protein on protecting mice from pneumococcal infection. Two mutations were included from this study: 175A reduced enzymatic activity, while 114A increased the same. Funahashi et al. (2002) reported several mutants of human lysozyme protein. They included mutants with both increased and decreased stability as compared to the WT.

The last group of mutants were selected from studies that used with mutagenesis scanning. This is a process in which specific residues (often referred to as *sites*) of the WT protein are mutated with a goal to differentiate functions of particular parts of the protein. Sun et al. (2001) performed active site mutagenesis of methylamine dehydrogenase and reported two groups of mutants – stabilizing and destabilizing. In another case, Huang et al. (1996) performed site-directed mutagenesis of phosphatase 1. Some mutations were categorized as stabilizing due to decreased enzymatic activity and/or increased stability to tested drugs. Others were categorized as destabilizing due to increased enzymatic activity or decreased stability

to the tested drugs. Furthermore, some mutations were excluded due to disagreeing effects of the aforementioned cases. Oppermann et al. (1997) conducted site-directed mutagenesis of hydroxysteroid dehyrdogenase, and reported the mutations to be either stabilizing or destabilizing based on their enzymatic activity. Mutations that were listed as resulting in "little or no effect" were not considered for our list. Similarly, the mutants considered by Dvir et al. (2003) for human acid-beta-glucosidase using site-directed mutagenesis belonged to two groups: one for decreased enzymatic activity and one for increased enzymatic activity or secondary structure destruction. However, some mutations were required to be excluded based on multiple disagreeing affects or non-reporting of stabilities. Mutants of porphobilinogen deaminase reported by Brownlie et al. (1994), those of lactose permease created by Braun et al. (1997) using site-directed alanine insertion, and those of ribonuclease A reported by Köditz et al. (2004) were all selected into these two groups in a similar manner.

2.3 Combinatorial Mutagenesis

There are two choices that are made for performing mutagenesis on a protein – what residues (sites) are to be mutated, and what (new) residues to which these sites are to be mutated. The first choice is relatively easy to make based on various biochemical and structural properties of the WT protein. Even after the potential mutation sites are identified, it is often not straightforward to decide the new residues. Experimentalists might want to try several amino acids for each mutation site, and hence are faced with the task of generating a large number of mutants. For example, if three potential mutation sites have been identified, and we want to try the residues Ala, Val, Ile for the first site, Val and Leu for the second site, and Cys, His, Lys, and Arg for the third site. The total number of mutants we have to analyze is $3 \times 2 \times 4 = 24$. In one of the studies that we included in the list of mutants, Martin et al. (2001) identified six amino acid sites of the protein 1CSQ to be mutated to *all other* amino acids. The result is a staggering 64,000,000 proteins to be analyzed (including the WT). A scoring function such as ours is most valuable for such mutagenesis experiments – we could identify (computationally) a relatively small set of mutants that are potentially the most suitable for our purpose, and experimentally generate these mutants before considering others. If the scoring function is accurate, chances are high that we will identify one (or a few) "good" mutants from this set of mutants, thus helping to reduce the total time and effort spent in the lab to analyze potential mutants. Since we consider all possible combinations of mutations at the individual sites, we term the process of scoring all possible mutants using the four-body scoring function as *combinatorial mutagenesis*.

As seen by the example of 1CSQ, the number of combinations could be quite huge, and for such cases, the usual way of scoring the mutants turns out to be highly inefficient. By default, we would score the WT protein once, and then score the same again for each mutant, with the appropriate changes made in the amino acid sequence. Each call to the four-body scoring function (to find the total score) involves the computation of the Delaunay tessellation of the protein, which proves to be the bottleneck as far as the overall running time of the algorithm is concerned. The most efficient algorithms for computing Delaunay tessellations have a worst-case running time of $O(n \log n)$, where n is the number of points (see Edelsbrunner, 2001, Chapters 1,5). The average running times in practice also follow the same bounds. At the same time, we notice that the structure of the WT is not altered in any of the mutants, and hence the Delaunay tessellation is the *same* as well – the residue numbers of the four amino acids forming each tetrahedron remains unaltered, even though the identities of some of the amino acids are changed. Hence we calculate the Delaunay tessellation only *once* as part of combinatorial mutagenesis, when scoring the WT protein. We just need to change the amino acid identities corresponding to each mutant.

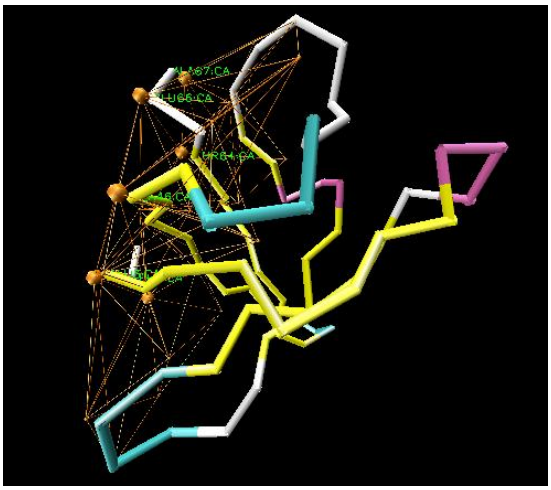

Fig. 1. 1CSQ (backbone trace shown) along with the Delaunay tetrahedra that the six mutation sites participate in. The six residues considered for mutation are 2-Leu, 3-Glu, 46-Ala, 64-Thr, 66-Glu, and 67-Ala. The orange spheres are at the side-chain centers of these residues, and the lines are the edges of the 58 tetrahedra that contain at least one of these six residues. If we consider all 67 amino acids, we get a total of 256 Delaunay tetrahedra (at 12 Å distance cut-off). This image was created using VMD (Humphrey et al., 1996).

We also notice that only those tetrahedra that involve one or more of the mutation sites see changes to the sequence identities of the participating amino acids. The remaining tetrahedra are identical in all respects in the WT and the mutant. For instance, the six mutation sites of 1CSQ participate in (i.e., at least one of the six is in) 58 Delaunay tetrahedra, which is only a fraction of the total of 256 tetrahedra formed by the entire protein (see Figure 1). From the theoretical point of view, there are some bounds for the number of Delaunay triangles that each point (out of the total n points) participates in when we consider the two-dimensional case (Edelsbrunner, 2001), but no analogous results are known for three dimensions. From among the 4,000-odd proteins that we analyzed, the maximum number of tetrahedra that a single amino acid participated in is 48 (43-Arg in 2BOQ) in we must note that the typical number of tetrahedra is much smaller (average is 17.65). This number is even smaller if the amino acid in question is on the surface of the protein. Hence the number of tetrahedra that see changes to the sequence identities could be much smaller than the total number. So we identify the set of tetrahedra that the mutation sites participate in (by searching the Delaunay tessellation of the WT). For each mutant, we calculate the difference in the sum of the log-likelihood ratios for these tetrahedra alone in the WT and in the mutant, and we use this difference to score and rank all the mutants. This implementation of combinatorial mutagenesis proves to be far more efficient

Table 1. Test set of mutations studied. Each mutation (if given) is indicated by the residue number and the new amino acid to which it is mutated. If there are multiple changes defining a single mutant, these are separated by "/", and all the mutations enclosed inside braces. Pred gives the number of correctly predicted mutants, out of the total number given by TOT. The complete list of mutants are available from the author's web page: http://www.wsu.edu/∼kbala/Mutate.html.

#	Article	Study	Mutants scored	Pred	TOT
1	Bonander et al. (2000)	Disulfide bond-deficient azurin mutant	(3A/26I), (3A/26A), (3A/25R/26A/27R)	3	3
2	Martin et al. (2001)	In-vitro selection of highly stabilized mutants with optimized Surface	G-1 (66L/67P), G-2 (2I/3S/46Q/64L/66L/67P), G-3 (46L/66L/67H), G-4 (2Y/3R/46L), G-5 (2Y/3I/46Q/64L/66L/67P)	5	5
3	de Antonio et al. (2000)	Contribution of tryptophan to the properties of ribotoxin α-sarcin	4F, 51F, (4F/51F)	0	3
4	Chen and Gouaux (1997)	Reduction of hydrophobicity in bacteriorhodopsin	Q1 (113Q), Q2 (113Q/116Q), Q3 (113Q/116Q/120Q), Q4 (113Q/116Q/120Q/124Q), Q4D (113Q/116Q/117D/120Q/124Q)	5	5
5	Huang et al. (1996)	Mutagenesis of protein phosphatase 1 for catalysis and inhibitor binding	96A, 124D, 248N, 221S, 395A, 208A	2	6
6	Ge et al. (2003)	Antifungal activity of a rice lipid transfer protein	45A, 46A, 72L	2	3
7	Suresh et al. (2006)	C-reactive protein mutants and pneumococcal infection in mice	175A, 114A	2	2
8	Oppermann et al. (1997)	Mutagenesis of hydroxysteroid dehydrogenase	12A, 87A, 138A	2	3
9	Pathange et al. (2006)	Correlation between protein binding strength	29 single-point mutants from 82-93H, 136-155H	26	29
10	Funahashi et al. (2002)	Surface hydration and stability of lysozyme	2G, 2A, 2L, 2M, 2F, 2S, 2Y, 2D, 2N, 2R, 2I 74G, 74A, 74F, 74S, 74Y, 74D, 74N, 74R, 74M, 74L, 74I, 110F, 110Y, 110R, 110N, 110D, 110M, 110L, 110I, 110A, 110G (original residue in all cases is V)	24	32
11	Takano et al. (1999)	N-terminal residues and conformational stability of human lysozyme	1M, 1A	0	2
12	Hahn et al. (1995)	Mutagenesis of Glucanase	101Y, 103Q, 105N, 105K, 107D, 101F, 103D	5	7
13	Sun et al. (2001)	Mutations and stability of methylamine dehydrogenase	76N, 122A, 122C, 119F, 119E, 119K	4	6
14	Dvir et al. (2003)	Human acid-β-glucosidase and Gaucher disease	370S, 394L, 463C, 496H	2	4
15	Korkegian et al. (2005)	Thermostabilization of an enzyme	(23L/140L/108I), (23L/140L), 23L, 140L, 10T, 67E, 69L	5	7
16	Brownlie et al. (1994)	Structures of the mutants of porphobilinogen deaminase	26H, 149L, 173W, 31T, 34K, 116T, 116W, 116Q, 177R, 223K, 250K, 252T, 93F, 201W, 247F, 256N, 167W	12	17
17	Braun et al. (1997)	Alanine insertion in lactose permease transmembrane helices	83A, 87A, 90A, 91A, 93A, 79A, 96A	5	7
18	Siadat et al. (2006)	Disulfide bonds and stability of an acetylcholinesterase	M2 (327C/375C), M3 (354C/456C), M4 (369C/476C), M6 (452C/533C), M7 (464C/543C)	4	5
19	Almog et al. (2002)	Stabilizing mutations in subtilisin BPN	S63 (41A/50F/73L/206W/217K/218S/221C/271E), S88 (2K/3C/5S/43N/50F/73L/206C/217K/218S/271E)	2	2
20	Erwin et al. (1990)	Salt bridges and stability of subtilisin BPN	271E, 51K, 164R	3	3
21	Köditz et al. (2004)	Mutagenesis of the unfolding region of ribonuclease A	35S, 35A, 46Y, (31A/33S), (35S/46Y), (35A/46Y), (31A/33S/46Y)	5	7
22	Ormö et al. (1995)	Radical stability in ribonucleotide reductase from E-coli	212W, 234N	2	2
23	Kong et al. (1993)	Evolutionally conserved aspartic acid residues in human glutathione S-transferase P1-1	57A, 98A, 152A	0	3
24	Carter et al. (2001)	Hydrophobic core mutations and stability	73 mutants (five different proteins)	66	73

than repeated calls to the default four-body scoring function for each mutant (see Section 3.1).

3 RESULTS AND DISCUSSION

We say that the four-body scoring function predicts the effect of a mutation *correctly* if an increase (correspondingly, a decrease) in the four-body total score is observed for mutations that are experimentally observed to be stabilizing or decreasing the reactivity (de-stabilizing or increasing the reactivity) of the WT protein. The performance of the scoring function is summarized in the last two columns of Table 1. Overall, 79% of the mutants were identified correctly. We are currently undertaking a detailed examination of how the scoring function performed for each of the twenty four mutant sets, and especially the cases of de Antonio et al. (2000), Takano et al. (1999), and Kong et al. (1993) (articles #3, #11, and #23 in

Table 1), for which the scoring function failed on all the mutants considered from each set. For the human lysozyme mutants studied by Takano et al., only the N-terminal residue is mutated, which does not form enough Delaunay tetrahedra (being on one end of the chain, and on the surface of the protein).

The effects of the mutants are quantified for ten of the mutant sets, five of them being different proteins studied by Carter et al. (2001). Even though the authors calculated linear correlation coefficients between four-body scores and free energy changes of these mutants, there is no clear evidence to suggest that the four-body scores follow a linear relationship with *all*, or any, for that matter, of the experimental quantities reported. (Masso et al. (2006) also report linear correlation coefficients, but see Section 3.2 for discussion on their work). The Spearman rank correlation coefficient between the four-body scores and the experimental values seems more appropriate, as there is no assumption of linear relationships involved. We present the rank correlation coefficients for the ten mutant sets in Table 2. The overall (average weighted with the number of mutants) Spearman rank correlation coefficient is 0.67. The rank correlations for the set of mutants studied by Carter et al. are markedly high – the average for these five mutant sets is 0.77. This result is not surprising, as all these mutations are done on sites in the hydrophobic cores of the proteins in question. The log-likelihood ratios (Equation (1)) are the most accurate for the tetrahedra formed by hydrophobic residues, as many more of such tetrahedra are observed in the training set of proteins used to develop the scoring function than, say, surface tetrahedra (Tropsha et al., 1996, 1998; Krishnamoorthy and Tropsha, 2003).

Table 2. Spearman rank correlation coefficients for mutant sets whose change in stability/reactivity has been quantified. # Mut gives the number of mutants, and RC gives the rank correlation coefficient. *: indicates that the WT is also included in the rank calculations. †: We could use only five of the mutants reported by Hahn et al. out of a total of seven (see Entry 12 in Table 1), as the mutants 103Q and 105K were listed as unmeasurable. The last five sets of mutants were all reported by Carter et al. (2001).

Study	Experimental quantity	# Mut	RC
Martin et al. (2001)	melting temperature	5	0.90
Chen and Gouaux (1997)	enthalpy of activation	6^*	1.00
Oppermann et al. (1997)	reaction rates (K)	6^*	-0.37
Funahashi et al. (2002)	stability z-values	32	0.67
Hahn et al. (1995)	WT/mutant reactivity	5 †	-0.10
barnase	$\Delta\Delta G_{\text{unfold}}$	9	0.90
chymotripsin inhibitor	$\Delta\Delta G_{\text{unfold}}$	9	0.82
staphylococcal nuclease	$\Delta\Delta G_{\text{unfold}}$	19	0.87
calbindin	$\Delta\Delta G_{\text{unfold}}$	9	0.78
T4 lysozyme	$\Delta\Delta G_{\text{unfold}}$	27	0.63

3.1 Combinatorial mutagenesis: an example

Our implementation of combinatorial mutagenesis (Section 2.3) scored all 64,000,000 mutants of 1CSQ (including the WT) within 6 hours (on a typical PC). In comparison, calling the four-body scoring function separately for each mutant did not finish in 24 hours. The original authors reported only five stabilizing mutants. It is not clear from their paper how many mutants were actually created by them, but it is hard to imagine this number being more than a few hundreds. Combinatorial mutagenesis predicted all five of them correctly. Furthermore, they were in the top 17.7% (of 64 million mutants). Analysis of the top-scoring mutants shows that high scores are assigned for mutants with Cystines in the selected sites. As illustrated in Figure 1, the six mutation sites participate in several tetrahedra together, i.e., they are linked to each other. The occurrences of stabilizing disulfide bonds between cystines is scored among the highest by the four-body scoring function, and hence the mutants with two or more Cystines are naturally scored high. This example shows that the experimentalist could analyze the list of the highest scoring mutants as identified by the four-body scoring function to determine a small set of highly likely candidates, which could be created first before exploring other mutants.

3.2 Comments on the work of Masso et al. (2006)

Masso, Lu, and Vaisman analyzed mutants reported in three different papers (and some more mutants that were not reported in these papers) using an earlier version of the four-body scoring function. In the first of these papers, Loeb et al. (1989) studied mutants of HIV-1 Protease. They reported a western blot assay analysis as well as an enzyme activity analysis. We could not consider many of the reported mutations, as they had disagreeing effects on the two analyses as far as stability and reactivity are concerned (both increase or both decrease). Only those mutations that had a reduced enzymatic activity and a WT western blot assay were considered. Most mutations were reported to have ambiguous "WT-like" behavior. Furthermore, only mutations of the wild type western blot assay classification were considered, as they were the only group for which the production of the protein was explicitly proven. This is an important factor, as the authors were mutating an operon, which leads to the production of multiple HIV proteins. Our results of this study were poorer than the other reported results. Out Of 56 mutants, only 21 were correctly scored. This lower accuracy might be accounted for by the unreported global destabilization of the enzyme. In another paper, Wrobel et al. (1998), analyzed mutants of HIV-1 reverse transcriptase. The analysis as well as the results presented were similar to those presented by Loeb et al., and a subset of appropriate mutations was selected in a similar manner. From among the 105 selected mutants, the four-body scoring function correctly predicted 51 mutants. Once again, the incorrect scoring of the remaining mutants in this study could very well be due to the global destabilization of the protein, which the authors did not report explicitly.

4 CONCLUSIONS

The strengths of the four-body scoring function for predicting stability and reactivity effects of mutations are widespread applicability, consistency (one setting works for all cases), computational efficiency (combinatorial mutagenesis), and accuracy. The idea of combinatorial mutagenesis can in principle be used even for a single mutation, or a few of them, but the gain in computational efficiency might not be noticeable. The test set of mutants assembled here can be used by researchers to study the applicability of other scoring functions to predict the effects of mutations.

Our test set is comprehensive, but is not *complete* – we plan to further explore previous as well as forthcoming literature to add

new sets of mutants to the current ones. Even though the same settings (in terms of the use of side-chain centers of the residues, 12 Å cut-off etc.), are recommended for applying our scoring function to all proteins, we could customize it with different settings specifically for certain classes of proteins, thus increasing its accuracy (of course, the scoring function will not perform as well under such customized settings for other classes of proteins). Another idea for increasing the accuracy of the scoring function is to use different weights for various quadruplets (rather than simply adding them all up). We could divide the set of mutants into training and test sets, determine the weights by learning from the training set and then validate them on the test set. We are currently investigating these and other ideas.

REFERENCES

Almog, O., Gallagher, D. T., Ladner, J. E., Strausberg, S., Alexander, P., Bryan, P., and Gilliland, G. L. (2002) Structural basis of thermostability analysis of stabilizing mutations in subtilisin bpn. *The Journal of biological chemistry*, **277**(30), 27553–27558.

Bandyopadhyay, D. and Snoeyink, J. (2004) Almost-Delaunay simplices: nearest neighbor relations for imprecise points. In *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, 410–419, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Bonander, N., Leckner, J., Guo, H., Karlsson, B. G., and Sjölin, L. (2000) Crystal structure of the disulfide bond-deficient azurin mutant c3a/c26a. *European Journal of Biochemistry*, **267**, 4511–4519.

Braun, P., Persson, B., Kaback, H. R., and von Heijne, G. (1997) Alanine insertion scanning mutagenesis of lactose permease transmembrane helices. *Journal of Biological Chemistry*, **272**(47), 29566–29571.

Brownlie, P. D., Lambert, R., Louie, G. V., Jordan, P. M., Blundell, T. L., Warren, M. J., Cooper, J. B., and Wood, S. P. (1994) The three-dimensional structures of mutants of porphobilinogen deaminase: Toward an understanding of the structural basis of acute intermittent porphyria. *Protein Science*, **3**(10), 1644–1650.

Carter, Jr, C. W., LeFebvre, B., Cammer, S. A., Tropsha, A., and Edgell, M. H. (2001) Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. *Journal of Molecular Biology*, **311**, 625–638.

Chen, G.-Q. and Gouaux, E. (1997) Reduction of membrane protein hydrophobicity by site-directed mutagenesis: introduction of multiple polar residues in helix d of bacteriorhodopsin. *Protein Engineering*, **10**, 1061–1066.

de Antonio, C., del Pozo, A. M., Mancheño, J. M., Oñaderra, M., Lacadena, J., Martínez-Ruiz, A., Pérez-Cañadillas, J. M., Bruix, M., and Gavilanes, J. G. (2000) Assignment of the contribution of the tryptophan residues to the spectroscopic and functional properties of the ribotoxin alpha-sarcin. *Proteins: Structure, Function, and Genetics*, **41**(3), 350–361.

Dvir, H., Harel, M., McCarthy, A. A., Toker, L., Silman, I., Futerman, A. H., and Sussman, J. L. (2003) X-ray structure of human acid-β-glucosidase, the defective enzyme in gaucher disease. *EMBO Reports*, **4**(7), 704–709.

Edelsbrunner, H. (2001) *Geometry and Topology for Mesh Generation*. Cambridge University Press, England.

Erwin, C., Barnett, B., Oliver, J., and Sullivan, J. (1990) Effects of engineered salt bridges on the stability of subtilisin bpn. *Protein Engineering*, **4**(1), 87–97.

Fowler, A., Krishnamoorthy, B., and Stratton, K. (2007) A hierarchy of scoring functions for protein decoy discrimination based on Delaunay tessellation of proteins. *Bioinformatics*, under review.

Funahashi, J., Takano, K., Yamagata, Y., and Yutani, K. (2002) Positive contribution of hydration structure on the surface of human lysozyme to the conformational stability. *The Journal of biological chemistry*, **277**(24), 21792–21800.

Gan, H. H., Tropsha, A., and Schlick, T. (2001) Lattice protein folding with two and four-body statistical potentials. *PROTEINS:Structure, Function, and Genetics*, **43**, 161–174.

Ge, X., Chen, J., Sun, C., and Cao, K. (2003) Preliminary study on the structural basis of the antifungal activity of a rice lipid transfer protein. *Protein engineering*, **16**(6), 387–390.

Hahn, M., Olsen, O., Politz, O., Boriss, R., and Heinemann, U. (1995) Crystal structure and site-directed mutagenesis of bacillus macerans endo-1, 3-1, 4 -beta glucanase. *Journal Biology and Chemistry*, **270**, 3081–3088.

Huang, H.-B., Horiuchi, A., Goldberg, J., Greengard, P., and Nairn, A. C. (1996) Site-directed mutagenesis of amino acid residues of protein phosphatase 1 involved in catalysis and inhibitor binding. *The American Society for Biochemistry and Molecular Biology, Inc*, **271**(5), 2574–2577.

Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, **14**, 33–38.

Köditz, J., Ulbrich-Hofmann, R., and Arnold, U. (2004) Probing the unfolding region of ribonuclease a by site-directed mutagenesis. *European Journal of Biochemistry*, **271**, 4147–4156.

Kong, K.-H., Inoue, H., and Takahashi, K. (1993) Site-directed mutagenesis study on the roles of evolutionally conserved aspartic acid residues in human glutathione s-transferase p1-1. *Protein Engineering*, **6**(1), 93–99.

Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme. *Science*, **308**(5723), 857–860.

Krishnamoorthy, B. and Stratton, K. (2007) Ranking CASP predictions using a four-body scoring function. *Proteins: Structure, Function, and Bioinformatics*, under review.

Krishnamoorthy, B. and Tropsha, A. (2003) Development of a four-body statistical pseudo-potential to discriminate native from non-native protein conformations. *Bioinformatics*, **19**(12), 1540–1548.

Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A. (1989) Complete mutagenesis of hiv-1 protease. *Nature*, **340**(6232), 397–400.

Martin, A., Sieber, V., and Schmid, F. (2001) In-vitro selection of highly stabilized protein variants with optimized surface. *Journal of Molecular Biology*, **309**(3), 717–726.

Masso, M., Lu, Z., and Vaisman, I. I. (2006) Computational mutagenesis studies of protein structure-function correlations. *Proteins: Structure, Function, and Bioinformatics*, **64**(1), 234–245.

Miyazawa, S. and Jernigan, R. L. (1985) Estimation of effective inter-residue contact energies from protein crystal structures: A quasi-chemical approximation,. *Macromolecules*, **18**, 534–552.

Munson, P. J. and Singh, R. K. (1997) Statistical significance of hierarchical multi-body potentials based on Delaunay tessellation and their application in sequence-structure alignment. *Protein Science*, **6**, 1467–1481.

Oppermann, U. C. T., Filling, C., Berndt, K. D., Persson, B., Benach, J., Ladenstein, R., , and JÖrnvall, H. (1997) Active site directed mutagenesis of $3^{\beta}/17^{\beta}$-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions. *Biochemistry*, **36**(1), 34–40.

Ormö, M., Regnström, K., Wang, Z., Jr., L. Q., Sahlin, M., and Sjöberg, B.-M. (1995) Residues important for radical stability in ribonucleotide reductase from escherichia coli. *Journal of Biological Chemistry*, **270**(12), 6570–6576.

Park, B. and Levitt, M. (1996) Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. *Journal of Molecular Biology*, **258**, 367–392.

Pathange, L. P., Bevan, D. R., Larson, T. J., and Zhang, C. (2006) Correlation between protein binding strength on immobilized metal affinity chromatography and the histidine-related protein surface structure. *Analytical Chemistry*, **78**(13), 4443–4449.

Siadat, O. R., Lougarre, A., Lamouroux, L., Ladurantie, C., and Fournier, D. (2006) The effect of engineered disulfide bonds on the stability of drosophila melanogaster acetylcholinesterase. *Journal of Biological Chemistry*, **7**(12).

Sippl, M. (1990) Calculation of conformational ensembles from potentials of mean force. *Journal of Molecular Biology*, **213**, 859–883.

Sun, D., Jones, L. H., Mathews, F. S., and Davidson, V. L. (2001) Active-site residues are critical for the folding and stability of methylamine dehydrogenase. *Protein Engineering*, **14**(9), 675–681.

Suresh, M. V., Singh, S. K., Jr., D. A. F., and Agarwal, A. (2006) Role of the property of c-reactive protein to activate the classical pathway of complement in protecting mice from pneumococcal infection. *The Journal of Immunology*, **176**, 4369–4374.

Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (1999) Effect of foreign n-terminal residues on the conformational stability of human lysozyme. *European Journal of Biochemistry*, **266**, 675–682.

Tropsha, A., Singh, R. K., and Vaisman, I. I. (1996) Delaunay tessellation of proteins: Four body nearest neighbor propensities of amino acid residues. *Journal of Computational Biology*, **3**(2), 213–222.

Tropsha, A., Vaisman, I. I., and Zheng, W. (1998) Compositional preferences in quadruplets of nearest neighbor residues in protein structures: statistical geometry analysis. In *IEEE Symposia on Intelligence and Systems*, 163–168.

Wrobel, J. A., Chao, S.-F., Conrad, M. J., Merker, J. D., Swanstrom, R., Pielak, G. J., and Hutchison, C. A. (1998) A genetic approach for identifying critical residues in the fingers and palm subdomains of hiv-1 reverse transcriptase. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 638–645.