PROTEINS: Structure, Function, and Genetics 27:336–344 (1997)

Mutation Matrices and Physical-Chemical Properties: Correlations and Implications

Jeffrey M. Koshi[1] **and Richard A. Goldstein**[1,2]*
[1]*Biophysics Research Division and* [2]*Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055*

ABSTRACT **To investigate how the properties of individual amino acids result in proteins with particular structures and functions, we have examined the correlations between previously derived structure-dependent mutation rates and changes in various physical-chemical properties of the amino acids such as volume, charge, α-helical and β-sheet propensity, and hydrophobicity. In most cases we found the Δ*G* of transfer from octanol to water to be the best model for evolutionary constraints, in contrast to the much weaker correlation with the Δ*G* of transfer from cyclohexane to water, a property found to be highly correlated to changes in stability in site-directed mutagenesis studies. This suggests that natural evolution may follow different rules than those suggested by results obtained in the laboratory. A high degree of conservation of a surface residue's relative hydrophobicity was also observed, a fact that cannot be explained by constraints on protein stability but that may reflect the consequences of the reverse-hydrophobic effect. Local propensity, especially α-helical propensity, is rather poorly conserved during evolution, indicating that non-local interactions dominate protein structure formation. We found that changes in volume were important in specific cases, most significantly in transitions among the hydrophobic residues in buried locations. To demonstrate how these techniques could be used to understand particular protein families, we derived and analyzed mutation matrices for the hypervariable and framework regions of antibody light chain V regions. We found a surprisingly high conservation of hydrophobicity in the hypervariable region, possibly indicating an important role for hydrophobicity in antigen recognition. Proteins 27: 336–344, 1997.** © **1997 Wiley-Liss, Inc.**

Key words: hydrophobicity; molecular evolution; local propensities; reverse hydrophobic effect; protein stability

INTRODUCTION

The characteristics of proteins are determined by their constituent amino acids. Each of the 20 naturally occurring amino acids has distinct attributes; natural selection takes advantage of these differences to construct proteins that fulfill numerous criteria such as stability, foldability, and functionality. In spite of the sizable database of solved protein structures, it is still not known which attributes of the amino acids—volume, charge, hydrophobicity, etc.—are the most important factors in various parts of the protein, or even what criteria constrain the choice of amino acids at different locations in the sequence.[1,2]

The dominant approach toward answering such questions has been through site-directed mutagenesis—mutating specific amino acids within a protein and testing the effect of those mutations on protein characteristics.[3–9] Changes in the characteristics of proteins can then be correlated with the changes in amino acid attributes. For instance, researchers such as Pace,[10] Rose and Wolfenden,[11] and Pielak et al.[12] have interpreted changes in stability resulting from site mutations based on the Δ*G* of transfer of the amino acids from octanol and cyclohexane to water. There are, however, several major difficulties faced in such studies. The first is the need to verify that the mutant protein does not have a significantly different tertiary structure, which can only be done by time-consuming methods such as nuclear magnetic resonance (NMR) or X-ray crystallography. A larger problem is the limited range of mutational combinations that can be studied. While researchers often have the ability to make any mutations they choose, the number of possible mutants makes it difficult to look at all the single mutations at a given site, much less all the double and triple mutations possible if neighboring amino acids are considered. This means that researchers can either use the technique of random mutagenesis and sample an extremely small random subset of possible mutations, or choose a limited number of presumably important mutations to examine, with their choices

*Correspondence to: Richard A. Goldstein, Dept. of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055.
Received 1 August 1996; accepted 20 September 1996.

necessarily based upon a priori assumptions about what is important in protein structures.

In contrast to the few years biochemists have been studying directed mutations, nature has had billions of years to do similar studies. The result is a vast database of evolutionary information representing proteins that are continually evolving, yet retaining their functions and structures over geological time scales. This implies that evolution must be selecting changes that preserve important characteristics of the protein, allowing structure and function to remain relatively constant. Thus, by identifying what characteristics are conserved in mutations allowed by evolution, we can determine what factors are important in different local environments of proteins (i.e., positions with various secondary structures and surface accessibilities). This approach makes no prior assumptions about what factors are important in determining protein structure, and also has the advantages that *all* possible mutations are considered, with the resulting proteins known to be viable in an in vivo environment.

The primary method by which researchers have tapped into the vast database provided by evolution has been to create mutation matrices. The first matrices were published by Dayhoff and Eck[13] in 1968, based on pairs of closely aligned sequences. Subsequent developments in the field have focused primarily on refining the Dayhoff approach, including comparing homologous fragments of proteins or choosing alignments based on matching three dimensional structures,[14–19] and applying the Dayhoff method to data sets restricted to certain types of proteins.[20] Some researchers have created matrices based on various other properties of the amino acids, but these were not intended to model the evolutionary process as much as provide a tool for sequence comparisons.[21–27] Also, these approaches do not result in matrices optimal for quantitative applications.[28]

The basic limitation of approaches based on the Dayhoff method is the absence of knowledge about ancestral sequences, or any rigorous method to infer that information. With this problem, such methods can only derive symmetric mutation matrices representing short periods of evolutionary time. It is, however, over longer periods of evolutionary time that the effects of evolutionary constraints are most strongly felt.[29] To avoid the constraints imposed by the Dayhoff approach, we developed a method to derive mutation matrices using estimation-maximization techniques, which allow the use of more distantly related sequences by creating a probabilistic reconstruction of the ancestral sequences.[30] By using data sets consisting of proteins of known structure or data sets limited to specific types of proteins, we were able to derive optimal mutation matrices for various secondary structure and surface accessibility classes, as well as optimal mutation matrices for the evolution of specific types of proteins. In this paper, we make use of our previously published optimal structure-dependent mutation matrices to determine how mutation rates correlate with changes in physical-chemical parameters. By identifying which correlations are significant, we can see which characteristics are most conserved during evolution, and thus are presumably most important. By analyzing our structure-dependent mutation matrices, we can also study how the requirements placed on amino acids vary with local environment. We then demonstrate how we can apply these methods to the mutational process in a particular class of proteins by constructing and analyzing mutation matrices for the framework and hypervariable region of the light chain V region of antibody (Ab) molecules.

In our analysis, we find that a residue's relative hydrophobicity is the most conserved quantity, even in exposed regions of the protein (i.e., hydrophobic residues tend to be replaced by hydrophobic residues, and hydrophilic residues mutate to other hydrophilic residues). We also find that secondary structure propensity and charge are of less importance, and volume plays a key role in specific situations. Perhaps the most important conclusion from our study stems from the contrast between our findings and those from site-directed mutagenesis studies. This contrast seems to imply that mutations allowed by evolution may follow different rules than mutations made in the laboratory. Finally, we note interesting differences between the mutation rates in the framework and hypervariable regions. A preliminary version of some of these results has been presented in a conference proceedings.[31]

METHODS AND RESULTS

As discussed above, our analysis was based on optimal structure-dependent and Ab-specific mutation matrices created using methods described previously.[30] For the structure-dependent matrices, 84 sets of homologous proteins were aligned and phylogenetic trees constructed with the program ClustalV.[32] The probability that any mutation matrix would result in this particular set of homologous sequences was computed, and the matrix most likely to result in the current sequences was derived. As the data set consisted of proteins of known structure, we were able to generate optimal mutation matrices for different combinations of secondary structure (α-helix, β-sheet, turn, and coil) and surface accessibility (buried or exposed), as well as a general matrix for transitions independent of local structure.[30] In the case of the Ab matrices, the data set was made up of 16 groups, taken from the KABAT database. Each group consisted of 10–56 aligned sequences of antibody light chain V regions from various subgroups

and species. Mutation matrices were optimized separately for the framework and hypervariable regions of the light chain V region. For the hypervariable region matrix, mutations rates to and from Arg (R), Asp (D), Cys (C), and Gly (G) were fixed at initial values and not included in the optimizations or the analysis, as not enough data existed to optimize these particular rates. Both the structure-dependent matrices and Ab matrices are available over the world wide web.

In this analysis we report the correlations of our various mutation matrices with changes in several physical-chemical parameters. Hundreds of physical-chemical parameters of the amino acids have been characterized, but we chose to focus first on several whose importance has been widely debated in the scientific literature.[33,34] One of these quantities is hydrophobicity, as measured by the ΔG of transfer from cyclohexane and octanol to water (ΔG_{oct} and ΔG_{chx}, respectively).[35,36] The importance of a general hydrophobic force as opposed to specific interactions such as hydrogen bonding has been under debate since Kauzmann first argued for the importance of hydrophobicity in protein folding.[37,38] In addition, the predictive values of these two particular indices of hydrophobicity have been previously studied.[11,12] Two other parameters chosen were amino acid volume and charge.[39] Both are known to be important characteristics in determining protein structure,[40,41] but their relative importance in the process of evolution is not well understood.[39,42] The last parameter we chose to examine was local structure propensity.[43–45] In this case, some researchers hold that local propensity is a dominant force in protein folding;[46] others believe it to provide only a minor contribution.[47–49] We recognize that all of these parameter scales represent averaged values for the amino acids, as different environments are undoubtedly characterized by slightly different scales. However, even with the generalized nature of these scales, they can still have good predictive value, as we demonstrate.

We specifically looked at the correlations between $|\Delta q|$ and $\ln (M_{a_1 a_2} M_{a_2 a_1})$, where $|\Delta q|$ is the absolute value of the difference in parameter value between amino acid a_1 and amino acid a_2, and $M_{a_1 a_2}$ represents the probability of amino acid a_1 mutating to amino acid a_2 in some fixed period of evolutionary time. The functional form of these correlations was motivated by the empirical observation that the best correlations were observed against the logarithm of our mutation matrices, implying an exponential relation involving fitness and mutation rate. This functional form was also supported by previous work involving theoretical models for evolution.[31] Correlations were examined for transitions within each structure-dependent mutation matrix, and for transitions within and between subsets of amino acids: hydrocarbon (LIVAG), hydrophobic and non-hydrogen bonding (LIVAMCG), neutral and polar (YWTSHQNF), and charged plus proline (RDEKP). The placement of Gly(G) and Pro(P) was motivated by work done by Thompson and Goldstein,[50] to reflect the optimal substitution classes derived in their work. The placement of the aromatic amino acids Phe(F) and Trp(W) is also somewhat nebulous. This is due to the delocalized π electrons of the aromatic ring structures, which give these residues a partially polar nature. Both Phe(F) and Trp(W) were placed in the netural and polar subset, as we found the highest correlation coefficients for all subsets were obtained with this placement.

In addition to the correlation coefficients, we also calculated the probability that a random, uncorrelated sample with the same number of data points would give that correlation coefficient or higher. As the number of data points differs for each case, it is this probability that is actually the more important value for determining which correlations are significant. The correlation coefficients (r), and probabilities of a random distribution matching or exceeding that correlation coefficient (P_r), are shown for various cases in Tables I and II and Figure 1.

DISCUSSION AND CONCLUSIONS

For the cases in which significant correlations existed between the structure-dependent mutation matrices and changes in physical-chemical parameters, correlations with our matrices were typically much higher than with the Dayhoff matrix. This is consistent with the results of Benner et al.,[29] who showed that short-time molecular evolution (as represented by the Dayhoff matrix) is more indicative of the underlying DNA mutation rates, while longer time behavior is more influenced by considerations at the amino acid level. This also indicates the accuracy of our mutation matrices, in that it is unlikely that less accurate matrices would be better correlated with changes in physical-chemical parameters.

One of the most obvious results we found was the high correlation of our structure-dependent matrices with changes in ΔG_{oct}, as shown in Table I and Figure 1a. The strong correlation with our matrix for buried residues is similar to the findings of Rose and Wolfenden[11] and Pielak et al.,[12] who found ΔG_{oct} to be a good indication of changes in stability for most amino acid substitutions in the protein core. The most likely explanation for this high correlation is that ΔG_{oct} serves as a good model for moving residues from the aqueous environment to the hydrophobic core during folding. This interpretation is supported by Pielak et al.'s[12] observation that mutation matrices are highly correlated with changes in stability for these substitutions.

We also observed a high correlation between changes in ΔG_{oct} and the mutation matrix for *exposed* residues (Table I). As the environment of these

TABLE I. Correlations of Mutation Matrices With $\Delta\Delta G_{oct}$ and $\Delta\Delta G_{chx}$*

	$\Delta\Delta G_{oct}$		$\Delta\Delta G_{oct}$ (within polar)		$\Delta\Delta G_{oct}$ (between polar and charged)		$\Delta\Delta G_{chx}$		$\Delta\Delta G_{chx}$ (within hydrocarbon)	
Matrix	r	P_r	r	P_r	r	P_r	r	P_r	r	P_r
All residues	−0.601	3.33×10^{-15}	−0.829	7.77×10^{-09}	−0.830	7.78×10^{-08}	−0.364	4.22×10^{-07}	−0.916	1.41×10^{-05}
Exposed	−0.625	1.32×10^{-16}	−0.846	2.02×10^{-09}	−0.845	2.83×10^{-08}	−0.313	1.42×10^{-05}	−0.904	2.69×10^{-05}
Buried	−0.536	5.93×10^{-12}	−0.791	9.57×10^{-08}	−0.811	2.52×10^{-07}	−0.312	1.50×10^{-05}	−0.911	1.93×10^{-05}
Alpha helix	−0.551	1.21×10^{-12}	−0.771	3.08×10^{-07}	−0.803	3.94×10^{-07}	−0.360	5.93×10^{-07}	−0.918	1.28×10^{-05}
Beta sheet	−0.552	1.10×10^{-12}	−0.842	2.86×10^{-09}	−0.797	5.60×10^{-07}	−0.345	1.66×10^{-06}	−0.928	6.96×10^{-06}
Turn	−0.563	3.28×10^{-13}	−0.760	5.54×10^{-07}	−0.863	7.22×10^{-09}	−0.313	1.40×10^{-05}	−0.879	8.25×10^{-05}
Coil	−0.583	3.34×10^{-14}	−0.808	3.42×10^{-08}	−0.779	1.37×10^{-06}	−0.347	1.50×10^{-06}	−0.928	6.90×10^{-06}
Exposed alpha helix	−0.479	1.39×10^{-09}	−0.710	5.55×10^{-06}	−0.815	2.00×10^{-07}	−0.264	2.20×10^{-04}	−0.878	8.72×10^{-05}
Beta sheet	−0.460	7.04×10^{-09}	−0.732	2.11×10^{-06}	−0.659	1.26×10^{-04}	−0.297	3.61×10^{-05}	−0.839	3.26×10^{-04}
Turn	−0.588	1.72×10^{-14}	−0.805	4.14×10^{-08}	−0.853	1.57×10^{-08}	−0.277	1.11×10^{-04}	−0.899	3.48×10^{-05}
Coil	−0.536	6.57×10^{-12}	−0.803	4.63×10^{-08}	−0.786	9.71×10^{-07}	−0.270	1.69×10^{-04}	−0.888	5.76×10^{-05}
Buried alpha helix	−0.493	4.08×10^{-10}	−0.660	3.62×10^{-05}	−0.594	1.82×10^{-03}	−0.327	5.59×10^{-06}	−0.916	1.44×10^{-05}
Beta sheet	−0.479	1.43×10^{-09}	−0.750	9.08×10^{-07}	−0.735	9.66×10^{-06}	−0.341	2.16×10^{-06}	−0.943	2.20×10^{-06}
Turn	−0.480	1.23×10^{-09}	−0.524	1.48×10^{-03}	−0.833	6.64×10^{-08}	−0.287	6.42×10^{-05}	−0.830	4.21×10^{-04}
Coil	−0.522	2.70×10^{-11}	−0.750	9.19×10^{-07}	−0.436	0.013	−0.319	9.76×10^{-06}	−0.895	4.18×10^{-05}
Dayhoff PAM	−0.451	1.44×10^{-08}	−0.617	1.41×10^{-04}	−0.622	3.43×10^{-04}	−0.201	3.98×10^{-03}	−0.768	1.78×10^{-03}
Ab framework	−0.357	8.55×10^{-06}	−0.422	0.010	−0.561	1.44×10^{-03}	−0.063	0.204	0.095	0.384
Ab hypervariable	−0.365	5.66×10^{-05}	−0.573	4.63×10^{-04}	−0.617	3.20×10^{-03}	0.113	0.123	−0.179	0.335

*Correlation coefficients (r) and probability that a correlation coefficient of equal or higher value could arise from uncorrelated data (P_r) are given for the various matrices vs. the $\Delta\Delta G$ of transfer from octanol to water for all transitions, for transitions within the polar amino acids (YWTSHQNF), and for transitions between the polar and charged (RDEKP) amino acids. Similar results are also shown for the $\Delta\Delta G$ of transfer from cyclohexane to water for all transitions, and for transitions within the hydrocarbon (LIVAG) amino acids.

TABLE II. Correlations of Mutation Matrices With Δ Local Structure Propensity, Δ Charge, and Δ Volume*

	Δα-helical propensity		Δβ-sheet propensity		Δ Charge		Δ Volume		Δ Volume (within hydrophobic)	
Matrix	r	P_r	r	P_r	r	P_r	r	P_r	r	P_r
All residues	0.066	0.181	−0.323	2.47×10^{-06}	−0.020	0.393	−0.165	0.011	−0.884	1.16×10^{-08}
Exposed	0.103	0.077	−0.283	3.46×10^{-05}	0.020	0.393	−0.125	0.042	−0.844	2.11×10^{-07}
Buried	0.037	0.308	−0.222	9.91×10^{-04}	−0.105	0.075	−0.167	0.010	−0.891	6.18×10^{-09}
Alpha helix	0.110	0.065	−0.276	5.47×10^{-05}	−0.006	0.465	−0.125	0.043	−0.840	2.69×10^{-07}
Beta sheet	0.033	0.327	−0.303	9.62×10^{-06}	−0.084	0.123	−0.208	1.90×10^{-03}	−0.855	1.05×10^{-07}
Turn	0.094	0.097	−0.309	6.56×10^{-06}	−0.010	0.448	−0.132	0.034	−0.803	1.97×10^{-06}
Coil	0.075	0.150	−0.209	1.86×10^{-03}	−0.065	0.185	−0.168	9.98×10^{-03}	−0.823	7.15×10^{-07}
Exposed alpha helix	0.063	0.191	−0.327	1.78×10^{-06}	0.018	0.404	−0.145	0.022	−0.621	7.92×10^{-04}
Beta sheet	0.024	0.369	−0.184	5.36×10^{-03}	−0.043	0.279	−0.161	0.013	−0.750	1.89×10^{-05}
Turn	0.103	0.077	−0.297	1.47×10^{-05}	0.027	0.356	−0.036	0.311	−0.771	8.45×10^{-06}
Coil	0.098	0.088	−0.176	7.26×10^{-03}	0.031	0.337	−0.133	0.033	−0.658	3.25×10^{-04}
Buried alpha helix	0.006	0.466	−0.201	2.64×10^{-03}	−0.103	0.077	−0.096	0.093	−0.857	9.13×10^{-08}
Beta sheet	0.039	0.296	−0.251	2.28×10^{-04}	−0.195	3.41×10^{-03}	−0.233	5.71×10^{-04}	−0.812	1.28×10^{-06}
Turn	0.060	0.203	−0.216	1.31×10^{-03}	−0.113	0.059	−0.179	6.38×10^{-03}	−0.675	2.06×10^{-04}
Coil	0.037	0.304	−0.208	1.90×10^{-03}	−0.155	0.016	−0.126	0.040	−0.833	4.04×10^{-07}
Dayhoff PAM	−0.077	0.145	−0.165	0.011	0.047	0.258	−0.135	0.031	−0.503	7.25×10^{-03}
Ab framework	0.014	0.424	−0.074	0.154	−0.086	0.188	−0.053	0.231	−0.504	7.14×10^{-03}
Ab hypervariable	−0.034	0.356	−0.132	0.074	−0.043	0.319	−0.180	0.024	−0.279	0.190

*Correlation coefficients (r) and probability that a correlation coefficient of equal or higher value could arise from uncorrelated data (P_r) are given for the various mutation matrices vs. the Δ α-helical propensity, Δ β-sheet propensity, Δ charge, Δ volume for all transitions, and Δ volume for transitions within the hydrophobic (LIVAMCG) amino acids.

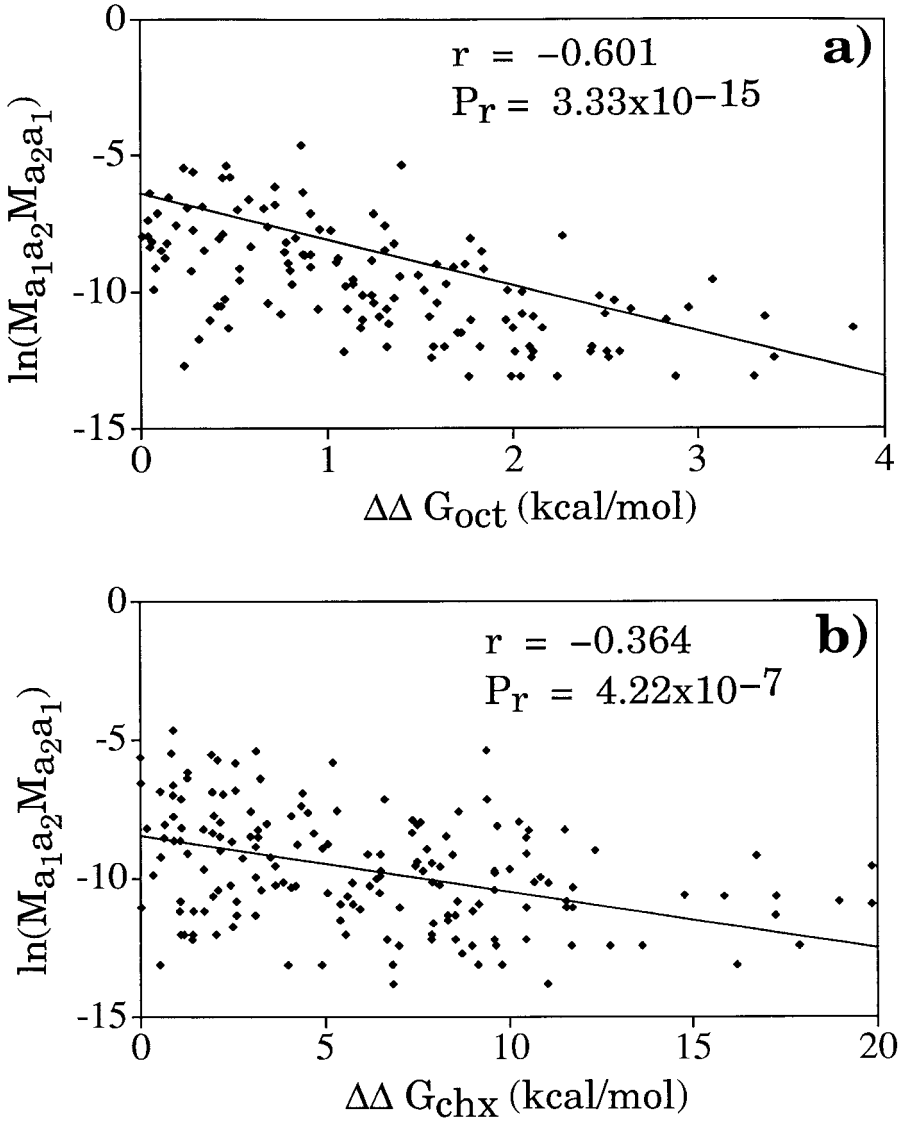

Fig. 1. **a:** Scatter plot of the $\Delta\Delta G$ of transfer of the amino acids from octanol to water vs. ln ($M_{a_1a_2}$ $M_{a_2a_1}$), the logarithm of the product of the transition probabilities. The correlation coefficient (r), probability that a correlation coefficient of equal or higher value could arise from uncorrelated data (P_r), and best fit line are also shown. **b:** Similar plot for the $\Delta\Delta G$ of transfer of the amino acids from cyclohexane to water.

residues remains roughly constant during folding, this correlation cannot be easily explained by stabilization of the folded conformation. Similarly, the highest and most significant correlation coefficients for ΔG_{oct} are against transitions within the polar residues and between the polar and charged residues, amino acids generally found on the surface of proteins. We can find a likely explanation for these correlations in the "reverse hydrophobic effect."[51,52] One of the major factors in efficient folding of the protein is the destabilization of incorrect conformations; the polar nature of surface residues prevents stabilization of alternatively folded states in which these residues are buried.

Surprisingly, correlations between ΔG_{chx} and any of our matrices were much less significant than those between ΔG_{chx} and our matrices (Table I and Fig. 1b), even the correlation between ΔG_{chx} and our matrix for buried residues. Rose and Wolfenden[11] and Pielak et al.[12] found ΔG_{chx} to be an excellent model, superior to ΔG_{oct}, for predicting the effect of mutations occurring in the protein core. This suggests that the correlation of ΔG_{chx} and our matrix for buried residues should have been at least equal to the correlation of that matrix with ΔG_{oct}. A more significant correlation was not found even for transitions among only the hydrocarbon amino acids, as would be expected based on the findings of site-directed muta-

genesis studies.[11,12] A reasonable explanation for these contrasting results can be found by examining the nature of the two solvents. Cyclohexane cannot form hydrogen bonds and as a result might be a good model for artifical site mutations, in which nature is not able to maximize the positive contributions of factors like hydrogen bonding. Evolutionarily constrained mutations, however, are likely to occur when the substituted residue can take advantage of hydrogen bond donors or acceptors or make use of subtle structural changes or compensatory mutations elsewhere in the sequence to optimize positive contributions to folding. The effects of such mutations are better modeled using octanol, a solvent with a limited ability to form hydrogen bonds.

In addition to ΔG_{oct} and ΔG_{chx}, we also examined α-helical and β-sheet propensity for correlations with our matrices. These results appear in Table II. Pielak et al.[12] found negligible correlation between α-helical propensity and changes in stability; in a similar fashion, we found no significant correlations of α-helical propensity with the various mutation matrices. This suggests that helical propensity is not generally conserved during mutations and thus is not an especially important factor in determining structure or function. This conclusion is supported by the results of researchers such as Chakrabartty et al.[53] and Govindarajan and Goldstein,[49] who found local propensity not to be a dominating factor in protein folding.[48,54,55] It has also been found that patterns of hydrophobicity are prevalent in α-helical structures[56] and are sufficient to induce helix formation.[57] These results suggest that it is patterns of hydrophobicity, rather than α-helical propensity, that dominate the formation of α-helices.

β-sheet propensity showed a higher correlation with our structure-dependent matrices. This higher correlation was not simply a dependence on physical-chemical properties such as volume or hydrophobicity, as we found no correlation between β-sheet propensity and these characteristics. These results agree with those of West and Hecht,[56] who found that characteristic patterns of hydrophobicity were less prevalent in β-sheets than in α-helices. This result suggests that other factors such as secondary structure propensity play a larger role in maintaining β-sheets. We also noted that buried β-sheets had a higher correlation than exposed β-sheets, again consistent with their observations that exposed β-sheets tended to contain more patterns of hydrophobicity than buried β-sheets.

Correlations of our matrices and changes in charge and volume were also explored (Table II). Change in charge showed no significant correlations with any of our matrices, but we determined that volume was an important parameter for specific subsets of transitions in specific environments. While correlations between transitions and changes in volume averaged over all locations were only modest, we did observe stronger correlations with mutations occurring in buried β-sheets and buried turns. This is not surprising, as volume is an important factor in turns, where steric clashes are a major constraint, and in buried positions, where internal packing plays an important role.[40] All correlations of changes in volume with transitions among the hydrophobic amino acids were significant. The correlation coefficients observed were much larger than those seen with transitions among the other groupings of amino acids. The strongest of these correlations, not surprisingly, was with the buried matrix. Thus, we can determine that volume is of key importance in specific situations: mutations from one hydrophobic residue to another, especially in buried positions.

The correlations of the Ab matrices for the framework and hypervariable regions of the light chain V region with ΔG_{oct} and ΔG_{chx} are similar to those of the structure-dependent matrices, but do have a few surprises of their own. As with the structure-dependent matrices, correlations of the Ab matrices with ΔG_{chx} were much lower than with ΔG_{oct}. Interestingly, among polar residues, it was mutations in the hypervariable region and not the framework region that showed a significant correlation with ΔG_{oct}. At first glance, this is surprising given that the hypervariable region Ab matrix is a matrix derived from predominantly solvent-exposed coil positions, a structure that normally imposes few restraints on residue characteristics. However, when the important functional nature of the hypervariable region in antigen recognition is considered, the correlation of the hypervariable region matrix with ΔG_{oct} is not quite as unexpected; hydrophobicity may play a key role in molecular recognition. The fact that the framework region Ab matrix showed such a low correlation is also of interest. In fact, for transitions among the polar residues, the framework region Ab matrix showed no strong correlations with any of the amino acid indices we examined. This could be a result of the stabilizing effect of the disulfide bond found in the structure of the light chain, or it could argue for the existence of other key amino acid characteristics that are not as well recognized as those like hydrophobicity or volume. Correlations of β-sheet propensity and size with our Ab matrices were also not significant, suggesting that such factors are not important in antibody molecules.

Lastly, we examined the correlations of the Ab matrices with the minimum number of base changes necessary to mutate from one amino acid to another. Interestingly, the framework and hypervariable region matrices showed a distinct difference in their degree of correlation. Neglecting the transitions that could not be fixed in the hypervariable region matrix, the framework region matrix had a correlation coefficient of -0.596 against the minimum base change matrix, while the hypervariable region matrix was more highly correlated, with an r value of -0.647.

This difference in r values corresponds to a difference in 6 orders of magnitude in P_r (1.85×10^{-26} vs. 2.45×10^{-32}), indicating a significant dissimilarity in the mutational processes in these two different regions. This is not a surprising observation, given that the hypervariable regions mutate some 1,000 times faster than normal proteins.

ACKNOWLEDGMENTS

We wouls like to thank Kurt Hillig and Jim Raines for computational assistance. Also, Michael Thompson and Gary Pielak deserve thanks for their thoughtful insights. Financial support was provided by the College of Literature, Science, and the Arts, the Program in Protein Structure and Design, the Horace H. Rackham School of Graduate Studies, NIH grants GM08270 and LM0577, and NSF equipment grant BIR9512955.

REFERENCES

1. Matthews, B.W. Genetic and structural analysis of the protein stability problem. Biochemistry 26:6885–6888, 1987.
2. Shoichet, B.K., Baase, W.A., Kuroki, R., Matthews, B.W. A relationship between protein stability and protein function. Proc. Natl. Acad. Sci. U.S.A. 92:452–456, 1995.
3. Nicholson, H., Becktel, W.J., Matthews, B.W. Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. Nature 336:651–656, 1988.
4. Lim, W.A., Farruggio, D.C., Sauer, R.T. Structural and energetic consequences of disruptive mutations in a protein core. Biochemistry 31:4324–4333, 1992.
5. Hurley, J.H., Baase, W.A., Matthews, B.W. Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. J. Mol. Biol. 224:1143–1159, 1992.
6. Hellinga, H.W., Wynn, R., Richards, F.M. The hydrophobic core of *Escherichia coli* thioredoxin shows a high tolerance to nonconservative single amino acid substitutions. Biochemistry 31:11203–11209, 1992.
7. Zhang, X.-J., Baase, W.A., Matthews, B.W. Multiple alanine replacements within alpha-helix 126–134 of T4 lysozyme have independent, additive effects on both structure and stability. Protein Sci. 1:761–776, 1992.
8. Eriksson, A.E., Baase, W.A., Matthews, B.W. Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. J. Mol. Biol. 229:747–769, 1993.
9. Lim, W.A., Hodel, A., Sauer, R.T., Richards, F.M. The crystal structure of a mutant protein with altered but improved hydrophobic core packing. Proc. Natl. Acad. Sci. U.S.A. 91:423–427, 1994.
10. Pace, C.N. Contribution of the hydrophobic effect to globular protein stability. J. Mol. Biol. 226:29–35, 1992.
11. Rose, G.D., Wolfenden, R. Hydrogen bonding, hydrophobicity, packing and protein folding. Ann. Rev. Biophys. Biomol. Struct. 22:381–409, 1993.
12. Pielak, G.J., Auld, D.S., Beasley, J.R., Betz, S.F., Cohen, D.S., Doyle, D.F., Finger, S.A., Fredericks, Z.L., Hilgen-Willis, S., Saunders, A.J., Trojak, S.K. Protein thermal denaturation, side-chain models, and evolution: Amino acid substitutions at a conserved helix-helix interface. Biochemistry 34:3268–3276, 1995.
13. Dayhoff, M.O., Eck, R.V. A model of evolutionary change in proteins. In: "Atlas of Protein Sequence and Structure." Vol. 3. Dayhoff, M.O., Eck, R.V. (eds.). Silver Spring, MD: National Biomedical Research Foundation, 1968:33–41.
14. McLachlan, A.D. Tests for comparing related amino-acid sequences. J. Mol. Biol. 61:409–424, 1971.
15. Henikoff, S., Henikoff, J.G. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. 89:10915–10919, 1992.
16. Jones, D.T., Taylor, W.R., Thornton, J.M. A new approach to protein fold recognition. Nature 358:86–89, 1992.
17. Risler, J.L., Delorme, M.O., Delacroix, H., Henaut, A. Amino acid substitutions in structurally related proteins. J. Mol. Biol. 204:1019–1029, 1988.
18. Altschul, S.F. Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219:555–565, 1991.
19. Overington, J., Donnelly, D., Johnson, M.S., Šali, A., Blundell, T.L. Environment-specific amino-acid substitution tables: Tertiary templates and prediction of protein folds. Protein Sci. 1:216–226, 1992.
20. Jones, D.T., Taylor, W.R., Thornton, J.M. A mutation data matrix for transmembrane proteins. FEBS Lett. 339:269–275, 1994.
21. Levin, J.M., Robson, B., Garnier, J. An algorithm for secondary structure determination in proteins based on sequence similarity. FEBS Lett. 205:303–308, 1986.
22. Fitch, W.M. An improved method of testing for evolutionary homology. J. Mol. Biol. 16:9–16, 1966.
23. Grantham, R. Amino acid difference formula to help explain protein evolution. Science 185:862–864, 1974.
24. Miyata, T., Miyazawa, S., Yasunaga, T. Two types of amino acid substitutions in protein evolution. J. Mol. Evol. 12:219–236, 1979.
25. Feng, D.F., Johnson, M.S., Doolittle, R.F. Aligning amino-acid sequences: A comparison of commonly used methods. J. Mol. Evol. 21:112–125, 1985.
26. Rao, J.K.M. New scoring matrix for amino acid residue exchange based on residue characteristic physical parameters. Int. J. Pept. Protein Res. 29:276–281, 1987.
27. Miyazawa, S., Jernigan, R.L. A new substitution matrix for protein sequence searches based on contact frequencies in protein structures. Protein Eng. 6:267–278, 1993.
28. Johnson, M.S., Overington, J.P. A structural basis for sequence comparisons. J. Mol. Biol. 233:716–738, 1993.
29. Benner, S.A., Cohen, M.A., Gerloff, D.L. Amino acid substitution during functionally constrained divergent evolution of protein sequences. Protein Eng. 7:1323–1332, 1994.
30. Koshi, J.M., Goldstein, R.A. Context-dependent optimal substitution matrices derived using bayesian statistics and phylogenetic trees. Protein Eng. 8:641–645, 1995.
31. Koshi, J.M., Goldstein, R.A. Correlating mutation matrices with thermodynamic and physical-chemical properties. In: Pacific Symposium on Biocomputing '96. Hunter, L., Klein, T. (eds.). Singapore: World Scientific, 1995:488–499.
32. Higgins, D.G., Bleasby, A.J., Fuchs, R. Clustal v: Improved software for multiple sequence alignment. CABIOS 8:189–191, 1992.
33. Kidera, A., Konishi, Y., Oka, M., Oi, T., Scheraga, H.A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. J. Protein Chem. 4:23–55, 1985.
34. Tomii, K., Kanehisa, M. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. Protein Eng. 9:27–36, 1996.
35. Fauchere, J., Pliska, V. Hydrophobic parameters of amino acid-side chains from the partitioning of n-acetyl-amino-acid amides. Eur. J. Med. Chem. 18:369–375, 1983.
36. Radzicka, A., Wolfenden, R. Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochemistry 27:1664–1670, 1988.
37. Kauzmann, W. Denaturation of proteins and enzymes. In: "The Mechanism of Enzyme Action." McElroy, W.D., Glass, B. (eds.). Baltimore: Johns Hopkins Press, 1954:70–120.
38. Kauzmann, W. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14:1–63, 1959.
39. Gerstein, M., Sonnhammer, E.L.L., Chothia, C. Volume changes in protein evolution. J. Mol. Biol. 236:1067–1078, 1994.
40. Richards, F.M. Areas, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. 6:151–176, 1977.
41. Abler, T. Stabilization energies of protein conformation. In:

"Prediction of Protein Structure and Principles of Protein Conformation." Fasman, G.D. (ed.). New York: Plenum Press, 1989:161–192.
42. Schueler, O., Margalit, H. Conservation of salt bridges in protein families. J. Mol. Biol. 248:125–135, 1995.
43. Chou, P.Y. Prediction of protein structural classes from amino acid compositions. In: "Prediction of Protein Structure and the Principles of Protein Conformation." Fasman, G.D. (ed.). New York: Plenum Press, 1989:549–586.
44. Creamer, T.P., Rose, G.D. Alpha-helix-forming propensities in peptides and proteins. Proteins 19:85–97, 1994.
45. Minor, D.L., Kim, P.S. Measurement of the beta-sheet-forming propensities of amino acids. Nature 367:660–663, 1994.
46. Zwanzig, R., Sxabo, A., Bagchi, B. Levinthal's paradox. Proc. Natl. Acad. Sci. U.S.A. 89:20–22, 1992.
47. Dill, K.A. Dominant forces in protein folding. Biochemistry 29:7133–7155, 1990.
48. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. Principles of protein folding—a perspective from exact simple models. Protein Science 4:561–602, 1995.
49. Govindarajan, S., Goldstein, R.A. Optimal local propensities for model proteins. Proteins 22:413–418, 1995.
50. Thompson, M.J., Goldstein, R.A. Constructing amino-acid residue substitution classes maximally indicative of local protein structure. Proteins 25:28–37, 1996.
51. Pakula, A.A., Sauer, R.T. Reverse hydrophobic effects relieved by amino acid substitutions at a protein surface. Nature 344:363–364, 1990.
52. Bowler, B.E., May, K., Zaragoza, T., York, P., Dong, A., Caughey, W.S. Destabilizing effects of replacing a surface lysine of cytochrome c with aromatic amino acids: Implications for the denatured state. Biochemistry 32:183–190, 1993.
53. Chakrabartty, A., Kortemme, T., Baldwin, R.L. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci. 3:843–852, 1994.
54. Pinker, R.J., Lin, L., Rose, G.D., Kallenback, N.R. Effects of alanine substitutions in alpha-helices of sperm whale myoglobin on protein stability. Protein Sci. 2:1099–1105, 1993.
55. Blaber, M., Zhang, X.J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A., Matthews, B.W. Determination of alpha-helix propensity within the context of a folded protein. sites 44 and 131 in bacteriophage t4 lysozyme. J. Mol. Biol. 235:600–624, 1994.
56. West, M.W., Hecht, M.H. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. Protein Sci. 4:2032–2039, 1995.
57. Ziong, H., Buckwalter, B.L., Shieh, H., Hecht, M.H. Periodicity of polar and non-polar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. Proc. Natl. Acad. Sci. U.S.A. 92:6349–6353, 1995.