Statistical and Visual Morph Movie Analysis of Crystallographic Mutant Selection Bias in Protein Mutation Resource Data

Werner G. Krebs and Philip E. Bourne[1,]*
Integrative Biosciences
San Diego Supercomputer Center MC 0505
National Partnership for Advanced Computational Infrastructure
[1]Department of Pharmacology
University of California, San Diego,
9500 Gilman Drive
La Jolla, CA 92093-0505, USA
**To whom correspondence should be addressed.*
wkrebs@sdsc.edu, bourne@sdsc.edu*

Abstract

The relationship between protein mutations and conformational change can potentially decipher the language relating sequence to structure. Elsewhere, we presented the Protein Mutant Resource (PMR), an online tool that systematically identified related mutants in the Protein DataBank (PDB), inferred mutant Gene Ontology classifications using data-mining, and allowed intuitive exploration of relationships between mutant structures. Here, we perform a comprehensive statistical analysis of PMR mutants. Although the PMR contains spectacular conformational changes, generally there is a counter-intuitive inverse relationship between conformational change and the number of mutations. That is, PDB mutations contrast naturally evolved mutations. We compare the frequencies of mutations in the PMR/PDB datasets against the PAM250 natural mutation frequencies to confirm this. We make available morph movies from PMR structure pairs, allowing visual analysis of conformational change and the ability to distinguish visually between conformational change due to motions (e.g., ligand binding) and mutations. The PMR is at http://pmr.sdsc.edu.

1. Introduction

Rational drug design seeks to use knowledge of a protein's three-dimensional chemical structure as a target against which to design new drugs. This particular application of x-ray crystallography has likely been one of the principal economic factors driving growth in the experimental field. However, mutant proteins may occur naturally in the target host population or, in the case of drugs such as antibiotics, may evolve in the parasite as a form of drug resistance. Scientists would like to have an understanding of how a putative drug interacts not only with the wild-type protein, but also with its likely mutants. The cost of experimental determination of mutant structures is often still prohibitive. An extensive structural database of proteins and neighboring mutants can be expected to assist scientists in visualizing likely structural changes brought about by mutation and would likely find immediate application in rational drug design through improved homology modeling, which is of importance in the pharmaceutical industry [1-4].

The deduction of a detailed, three-dimensional chemical structure of a protein from its genetic sequence is a fundamental and long-studied problem in structural biology [5-7], of importance in *de novo* structure prediction, protein folding, crystallographic refinement phasing, molecular dynamics, and computational chemistry. At present, it is solved principally through the labor-intensive but effective process of X-ray crystallography and NMR. A direct study of existing data on the effects of protein mutation on protein structure can have immediate payoffs [7, 8].

Although databases of mutant gene products [9-12] as well as specialized databases of mutant protein structures have previously been developed [13-15], the PMR [16] was the first PDB-wide [17] database of mutant protein structures. Entering the PDB ID of a structure in the PMR into the entry form on the PMR home page brought up the sequence of the wild-type structure for that mutant family along with a listing of the differences in amino acid sequence between the wild-type and the selected PDB ID (Figure 1). Users could click on any of the muta-

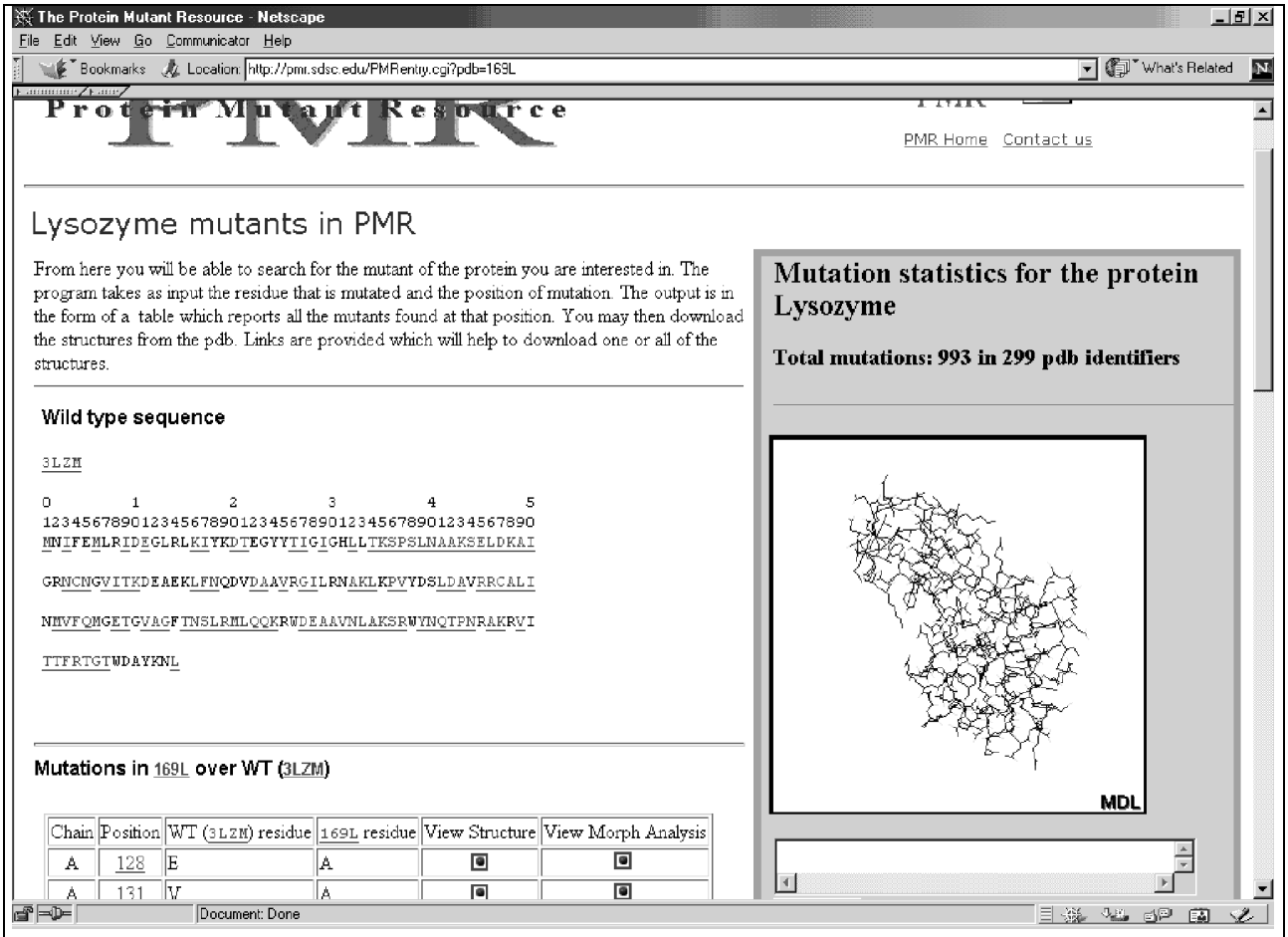

Figure 1. Web screen shot of the PMR mutant browser. Upon selecting a specific mutational site from the sequence all mutations at that site are shown and the site is shown on the 3-D structure (requires Chime). A torsion angle analysis and a morph movie analysis displaying the conformational change between the selected PDB ID and the designated wild-type for the particular mutant family can also be displayed.

tion sites listed for the wild-type structure to obtain a listing of the available mutant structures with modifications at the amino acid position (Figure 1). An anticipated use of the PMR was in protein engineering. Scientists seeking to modify the sequence of an existing protein to express a slightly different structure could search the PMR for sequences matching the protein's current sequence and then examine the stored mutations for structure variants [18, 19]. The PMR website had a number of other innovations; in particular, the PMR GO classification feature utilized an improved means of and database-wide statistically rigorous gene annotation and data-mining with widespread applicability [20, 21]. PMR database entries interacted with a number of external databases (MolMovDB [22-25], GO [26], PubMed/Entrez [27], PDBsum [28]) as well as the PDB. Consequently, the PMR could be used as a portal by those studying families of proteins of closely related sequence within the PDB.

Here, we characterized the effect of PMR mutations on protein tertiary structure statistically and detected a potential selective bias in available PMR/PDB structures of mutant proteins. To confirm this, we compared the frequency of mutations in PMR/PDB datasets against accepted PAM 250 natural amino acid mutation frequencies. We further improved on the PMR web interface by generating and making morph movies of the conformation changes available on the web. These automatically generated morph movies assist scientists in visually discriminating between conformational changes caused by protein motions [25] from those caused primarily by

changes in protein sequence [16]. We believe both our statistical analysis of PMR/PDB data and our morph movies of PMR data will be of general interest to the structural bioinformatics community. Our morph movies are freely available off the PMR website (http://pmr.sdsc.edu).

Table 1. Mutation statistics for the PMR database taken from 1157 PDB structures.

Total number of mutations (chains)	3343
Wild-type PDB chains	194
Non wild-type PDB chains	3149
Number of PDB IDs associated with mutations	
Mutation Sites	1157
Average number of residues mutated per chain	2.9
Most commonly mutated amino acid in PMR	Alanine

2. Materials and Methods

As described elsewhere [16], the PMR was generated by automatically clustering the PDB [17] at 95% sequence identity using the CD-HIT sequence clustering approach [29]. CD-HIT uses a greedy algorithm [30] to sort and process sequences in order of decreasing length; the longest sequence in each cluster becomes its represen-

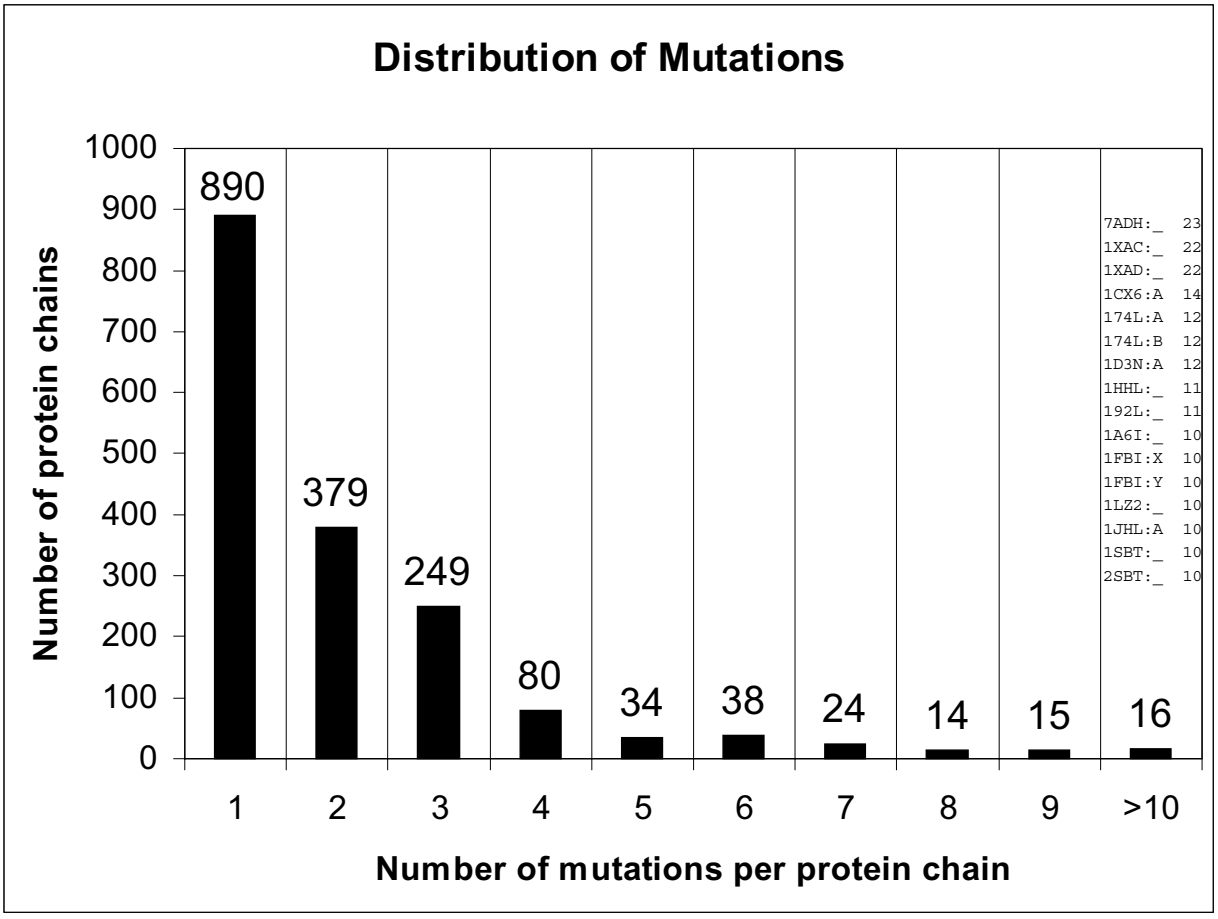

Figure 2. Distribution of the number of mutations per polypeptide chain in the PMR. The specific PDB and chain identifiers for each polypeptide chain with > 10 mutations are shown.

tative. Efficiency is achieved because sequences are processed by comparing them only against the representative sequences for each established cluster to decide whether they should be added to an existing cluster or become the representative for a new cluster.

The resulting clusters were then manually filtered into species-based families and a 'wild-type' PDB chain was manually selected from each family by inspection of

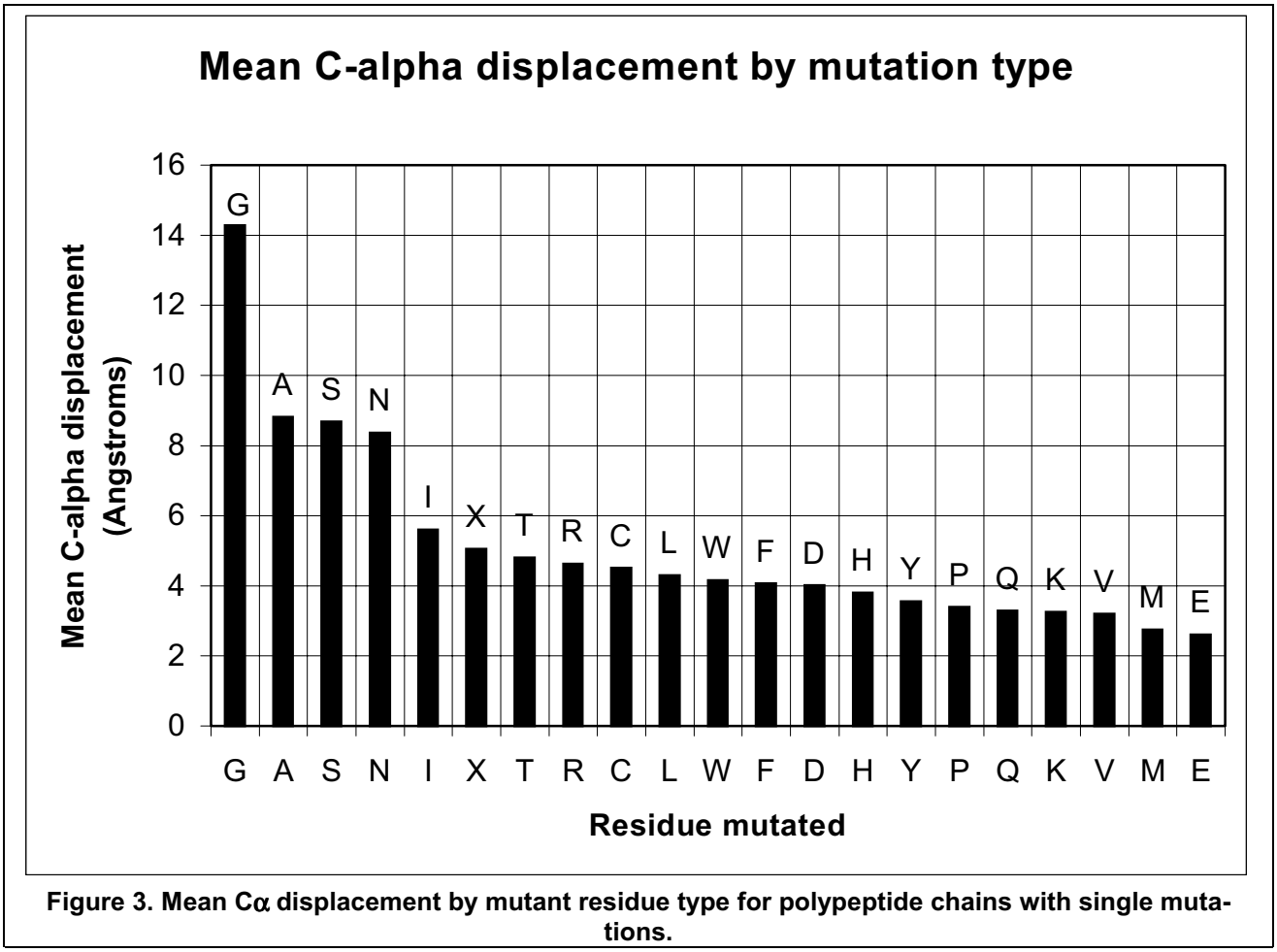

Figure 3. Mean Cα displacement by mutant residue type for polypeptide chains with single mutations.

the scientific literature [16]. Software was developed to automatically find and add new PDB entries to existing PMR families on a regular basis. The resulting data were loaded into Oracle tables and made freely accessible via a web interface, developed in Perl [31], Oracle SQL, Chime, JavaScript, and HTML.

Morph movie technology [22, 23] originally developed for the Database of Macromolecular Motions [22, 24, 25, 32, 33] (http://molmovdb.org) was applied to PMR data to make four-dimensional visual illustrations [22, 23, 34] of the conformation changes induced by mutation. Structural change is given in terms of Cα displacement for each possible, non-redundant wild-type and mutant structure pair in each PMR family using a sieve-fitting superposition technique [35] which interpolates between solved structures to provide a visual rendering. Protein motions are available as animated GIF images on the PMR website (http://pmr.sdsc.edu; Figure 1). Summary statistics on the types of mutation and the motions induced are given in Table 1. The number of mutations per structure is given in Figure 2. Cα displacements by mutant residue type in PMR data are given in Figure 3. The frequency of amino acid mutations in PMR data were

Table 2. PAM250 mutation frequencies compared with PMR mutation frequencies. The rightmost value in each cell gives the mutational frequency in the PMR between any two amino acids. This was computed by tabulating the occurrences of that particular mutation data between wild-type and mutant chains throughout the PMR, and then linearly normalizing that fractional value to the same scale as PAM250. The leftmost value in each cell gives the raw PAM250 accepted natural amino acid mutation rate [36]. Lower (more negative) numbers indicate an observed reduced tendency to mutate. PMR values along the diagonal are undefined, so only PAM250 values are given. Values in which the difference between the PAM250 accepted mutation rate and the PMR rate exceed more than half of the range are shaded; about 5% of the values are designated this way.

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	2 *																			
C	-2 7	4 *																		
D	0 -5	-5 -7	4 *																	
E	0 -5	-5 -7	3 -4	4 *																
F	-4 -6	-4 -7	-6 -7	-5 -7	9 *															
G	1 -5	-3 -7	1 -5	0 -6	-5 -7	5 *														
H	-1 -6	-3 -6	1 -7	1 -7	-2 -7	-2 -6	6 *													
I	-1 -5	-2 -7	-2 -7	-2 -7	1 -7	-3 -7	-2 -7	5 *												
K	-1 -4	-5 -7	0 -7	0 -5	-5 -7	-2 -7	0 -6	-2 -7	5 *											
L	-2 -3	-6 -7	-4 -7	-3 -7	2 -5	-4 -6	-2 -6	2 -4	-3 -7	6 *										
M	-1 -6	-5 -7	-3 -7	-2 -7	0 -6	-3 -7	-2 -7	2 -6	0 -7	4 -5	6 *									
N	0 -5	-4 -7	2 1	1 -7	-4 -7	0 -7	2 -5	-2 -7	1 -6	-3 -6	-2 -7	2 *								
P	1 -6	-3 -7	-1 -7	-1 -7	-5 -7	-1 -6	0 -7	-2 -7	-1 -7	-3 -5	-2 -7	-1 -7	6 *							
Q	0 -6	-5 -7	2 -7	2 -1	-5 -7	-1 -7	3 -5	-2 -7	1 -4	-2 -6	-1 -7	1 -6	0 -7	4 *						
R	-2 -5	-4 -7	-1 -7	-1 -6	-4 -7	-3 -7	2 -6	-2 -7	3 -1	-3 -6	0 -7	0 -7	0 -7	1 -6	6 *					
S	1 -3	0 -2	0 -6	0 -6	-3 -7	1 -6	-1 -6	-1 -7	0 -6	-3 -6	-2 -7	1 -1	1 -6	-1 -7	0 -5	3 *				
T	1 -3	-2 1	0 -7	0 -6	-2 -7	0 -6	-1 -6	0 -6	0 -7	-2 -6	-1 -7	0 -7	0 -7	-1 -7	-1 -7	1 -4	3 *			
V	0 -3	-2 -7	-2 -7	-2 -7	-1 -5	-1 -7	-2 -6	4 -2	-2 -7	2 -5	2 -5	-2 -7	-1 -7	-2 -7	-2 -7	-1 -7	0 -6	4 *		
W	-6 -6	-8 -7	-7 -7	-7 -7	0 -6	-7 -6	-3 -7	-5 -7	-3 -7	-2 -7	-4 -7	-4 -7	-6 -7	-5 -7	2 -7	-2 -7	-5 -7	-6 -7	17 *	
Y	-3 -7	0 -7	-4 -7	-4 -7	7 -4	-5 -7	0 -7	-1 -7	-4 -6	-1 -7	-2 -7	-2 -7	-5 -7	-4 -7	-4 -7	-3 -7	-3 -7	-2 -7	0 -6	10 *

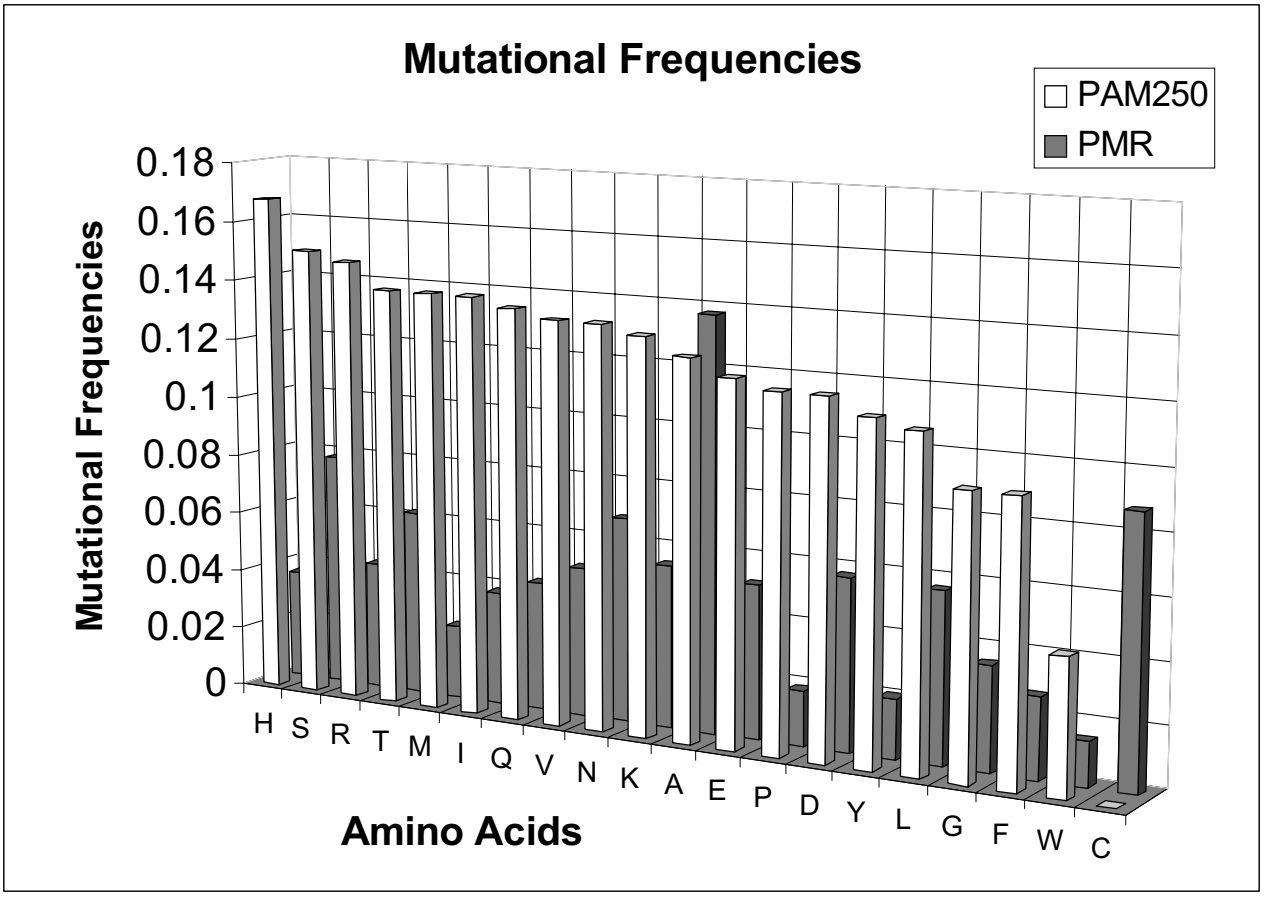

Figure 4. Bar chart of mutational frequencies of amino acids in the PMR and the accepted natural PAM250 amino acid mutation rate [36]. Mutational frequencies for PMR and PAM250 datasets were summed vertically and horizontally (Table 2) for each amino acid to generate a mutational frequency for each amino acid. Mutational frequencies were then sorted by the PAM250 values (foreground).

tabulated and compared with the accepted PAM250 natural amino acid mutation frequencies [36] and are shown in Table 2 and Figure 4. A scatter-plot depicting the relationship of number of mutations with structural change (as measured by previously computed Cα displacements) is shown in Figure 5.

3. Results

A morph analysis of all possible structure pair combinations within PMR families would have $O(N^2)$ complexity with the size of family in terms of both disk space and CPU time and was not computationally tractable. Instead, we limited our analysis to non-redundant combinations of the members of each family and its wild-type. This has $O(N)$ complexity and reduced the task by several orders of magnitude. Running in parallel, our morph server software required three days of CPU time on a sixteen-CPU cluster of four-CPU Sun Ultra-80 servers running SunOS 5.7. The resulting morph movies and statistical data require 4.3 gigabytes of disk storage.

The distribution of numbers of mutations per polypeptide chain is given in Figure 2. Greater than 99% of the PMR mutant structures have 9 or fewer mutations when compared to the wild-type. The rapid fall-off in number of mutations is expected as these structures are usually studied to understand the impact of single or a small number of correlated mutations. Alanine is the residue most commonly used to mutate structures (Table 1) presumably to change the functional role of a given residue and through the presence of a C-beta carbon still confer side chain directionality and some sense of side chain

Mutations versus Structural Change

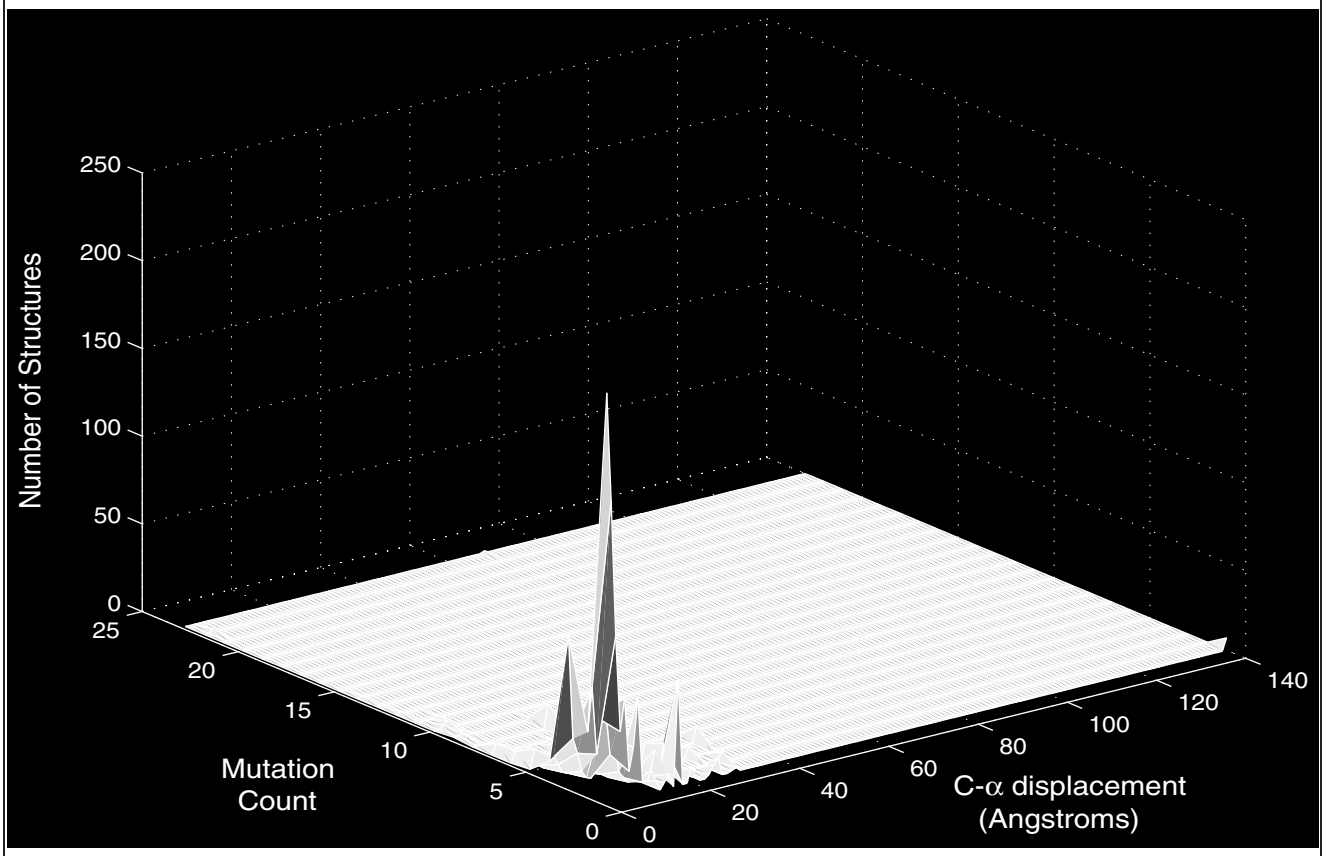

Figure 5. Number of mutations per polypeptide chain versus structural change (C-alpha displacement).

volume in a neutral substitution. Single mutations involving glycine result in the largest structural changes (Figure 3) presumably as a result of significant stereochemical change in side chain volume and possibly physicochemical property.

There is a significant difference in the mutational frequencies for PMR versus the natural PAM250 frequencies (Figure 4 and Table 2). For example, the table shows that the mutation of phenylalanine to tyrosine seems disfavored by structural biologists despite a relatively high frequency during the course of evolution. Conversely, the conversion of alanine to cysteine, uncommon during the course of evolution, is favored by structural biologists, presumably to add stability to a structure under study. Divergence from natural frequencies of mutation is not surprising since structural biology is often concerned with the engineering of proteins to test function. Not surprisingly, loss of function is less common in nature than it is in the laboratory. It is not difficult to engineer a protein that retains its structure, but which is biologically inactivate or has a significantly reduced activity.

Given the complete body of structural data showing mutation from the wildtype, contrary to what one would expect, structural change shows an inverse relationship to the number of mutations (Figure 5) – large structural changes are induced by small numbers of mutations, whereas a large number of mutations can lead to relatively small changes. Consider several specific cases shown in the PMR data (Figure 5). The largest structural change in the PMR occurs in TAQ DNA Polymerase [37]. A number of structures of TAQ DNA polymerase have been solved with an alanine instead of the wild-type glycine at position 152. These all show an unusually large 140Å Cα displacement. Crystal contacts as a cause may be ruled out since the mutant structure has been solved in a number of different space groups, including the same space group as the wild-type, and all structures give an identical Cα displacement. The 140 Å Cα displacement structures are all bound to DNA, whereas the 1TAQ wildtype [37] is an unbound structure indicating that the unusually large conformation change here is likely due to the combined effect of a DNA clamping protein motion as well as the single-point mutation. Thus the PMR does not distinguish between conformational changes induced by ligand binding or complex formation and that induced simply by point mutation. However, large changes induced by point mutations alone do occur. Analysis of the structural change between 7ADH [38] and the designated wild-type alcohol dehydrogenase structure 1ADG [39] shows a displacement of 66Å caused by structural changes resulting from 22 mutations. Visualization of the morphing shows chain breakages and backbone elements passing through each other. That is, unlike TAQ DNA polymerase where the structural change represents an observable physiological change, here there are discreet and distinct states representing different structures.

4. Discussion and Conclusion

The addition of a morph analysis to the PMR permits the visualization of conformational change induced by changes to the wildtype protein. However care and a review of the original PDB files is needed to distinguish between conformational change induced by mutation (often indicated by alternative folding, chain breakage, or the passage of backbone and sidechains atoms through one another) and protein motions caused by ligand binding rather than mutation, which appear as smooth transitions. It is necessary to often refer back to the original PDB files to understand the cause of conformational change. With improvements to ligand descriptions within the PDB we anticipate better annotating and classifying specific motions in the future.

Induced mutation while retaining structure shows a different patter of substitutions that that observed in nature (Table 2 and Figures 4 and 5). Evolutionary sequence drift usually preserves structure and function whereas structural biologists often mutate proteins specifically to change structure and function, or indeed to induce better structure formation to better nature [40]. A useful addition to the PMR would be mutations that prevented a structure from being observed. Such negative data has not traditionally found its way to the literature or to public databases, but that will likely change with the advent of structural genomics.

Our computation of morph movie representations enables the broader structural bioinformatics community to analyze and represent protein mutation conformational change visually. Computation of these biologically interesting results was made tractable by a number of new algorithms [29] and design decisions [23] within our software pipeline. These results and methods complement existing algorithms [20, 21] used to generate data for other areas of the PMR website.

5. Acknowledgements

The PMR was supported by the U.S. National Science Foundation (NSF) National Partnership for Advanced Computational Infrastructure (NPACI) and the U.S NSF Division of Biological Infrastructure grant DBI 0111710. The authors would also like to thank Dr. T. Murlidharan Nair for work on the original PMR dataset and Drs. Ilya Shindalov and Wolfgang Bluhm for useful discussions.

6. References

[1] A. L. Sabb, G. M. Husbands, J. Tokolics, R. P. Stein, R. P. Tasse, C. A. Boast, J. A. Moyer, and M. Abou-Gharbia, "Discovery of a highly potent, functionally-selective muscarinic M1 agonist, WAY-132983 using rational drug design and receptor modelling," *Bioorg Med Chem Lett*, vol. 9, pp. 1895-900., 1999.

[2] J. M. Hakala and M. Vihinen, "Modelling the structure of the calcitonin gene-related peptide," *Protein Eng*, vol. 7, pp. 1069-75., 1994.

[3] J. H. McKie, "Homology modelling of the dihydrofolate reductase-thymidylate synthase bifunctional enzyme of Leishmania major, a potential target for rational drug design in leishmaniasis," *Drug Des Discov*, vol. 11, pp. 269-88., 1994.

[4] G. Vriend, "WHAT IF: A molecular modeling and drug design program," *J. Mol. Graph.*, vol. 8, pp. 52-56, 1990.

[5] A. R. Fersht, "Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications," *Proc Natl Acad Sci U S A*, vol. 92, pp. 10869-73., 1995.

[6] S. Govindarajan and R. A. Goldstein, "Optimal local propensities for model proteins," *Proteins*, vol. 22, pp. 413-8., 1995.

[7] R. Unger and J. Moult, "Local interactions dominate folding in a simple protein model," *J Mol Biol*, vol. 259, pp. 988-94., 1996.

[8] H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez, "Point mutations alter the mechanical stability of immunoglobulin modules," *Nat Struct Biol*, vol. 7, pp. 1117-20., 2000.

[9] M. W. Beukers, I. Kristiansen, I. J. AP, and I. Edvardsen, "TinyGRAP database: a bioinformatics tool to mine G-protein-coupled receptor mutant data," *Trends Pharmacol Sci*, vol. 20, pp. 475-7., 1999.

[10] K. Kristiansen, S. G. Dahl, and O. Edvardsen, "A database of mutants and effects of site-directed mutagenesis experiments on G protein-coupled receptors," *Proteins*, vol. 26, pp. 81-94., 1996.

[11] I. Edvardsen and K. Kristiansen, "Computerization of mutant data: the tinyGRAP mutant database," *7TM journal*, vol. 6, pp. 1-6, 1997.

[12] S. M. Maurer, R. B. Firestone, and C. R. Scriver, "Science's neglected legacy," *Nature*, vol. 405, pp. 117-20., 2000.

[13] K. Nishikawa, S. Ishino, H. Takenaka, N. Norioka, T. Hirai, T. Yao, and Y. Seto, "Constructing a protein mutant database," *Protein Eng*, vol. 7, pp. 773, 1994.

[14] T. Kawabata, M. Ota, and K. Nishikawa, "The Protein Mutant Database," *Nucleic Acids Res*, vol. 27, pp. 355-7., 1999.

[15] M. M. Gromiha, H. Uedaira, J. An, S. Selvaraj, P. Prabakaran, and A. Sarai, "ProTherm, Thermodynamic Database for Proteins and Mutants: developments in version 3.0," *Nucleic Acids Res*, vol. 30, pp. 301-2., 2002.

[16] W. G. Krebs, T. M. Nair, and P. E. Bourne, "The Protein Mutant Resource: A Tool for Protein Engineering," *Manuscript in preparation.*, 2003.

[17] H. Berman, M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, pp. 235-242, 2000.

[18] I. N. Shindyalov and P. E. Bourne, "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path," *Protein Eng*, vol. 11, pp. 739-47., 1998.

[19] M. Levitt, *STRUCTAL. A structural alignment program*: Stanford University, 1994.

[20] W. G. Krebs and P. E. Bourne, "Statistically Rigorous Automated Gene Annotation and Classification and its Application to Protein Data Bank Sequences using Gene Ontology Terms," *submitted to Bioinformatics*, 2003.

[21] W. G. Krebs and P. E. Bourne, "Statistically Rigorous Automated Gene Annotation and Classification and its Application to Protein Data Bank Sequences using Gene Ontology Terms,". U.S. Patent Pending.

[22] W. Krebs, J. Tsai, V. Alexandrov, J. Junker, R. Jansen, and M. Gerstein, "Studying Protein Flexibility in a Statistical Framework: Tools and Databases for Analyzing Structures and Approaches for Mapping this onto Sequences," *Methods Enzymol*, vol. 374, 2003.

[23] W. G. Krebs and M. Gerstein, "The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework," *Nucleic Acids Res*, vol. 28, pp. 1665-1675, 2000.

[24] N. Echols, D. Milburn, and M. Gerstein, "MolMovDB: analysis and visualization of conformational change and structural flexibility," *Nucleic Acids Res*, vol. 31, pp. 478-82., 2003.

[25] M. Gerstein and W. Krebs, "A Database of Macromolecular Movements," *Nucl. Acids Res*, vol. 26, pp. 4280, 1998.

[26] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet*, vol. 25, pp. 25-9, 2000.

[27] G. D. Schuler, J. A. Epstein, H. Ohkawa, and J. A. Kans, "Entrez: molecular biology database and retrieval system," *Methods Enzymol*, vol. 266, pp. 141-62, 1996.

[28] R. A. Laskowski, Hutchinson, E.G., Michie. A.D., Wallace, A.C., Jones, M.L., Thornton, J.M., "PDBsum: a Web-based database of summaries and analyses of all PDB structures," *Trends Biochem. Sci.*, vol. 22, pp. 488-490, 1997.

[29] W. Li, L. Jaroszewski, and A. Godzik, "Clustering of highly homologous sequences to reduce the size of large protein databases," *Bioinformatics*, vol. 17, pp. 282-3., 2001.

[30] L. Holm and C. Sander, "Removing near-neighbour redundancy from large protein sequence collections," *Bioinformatics*, vol. 14, pp. 423-9., 1998.

[31] L. Wall, D. Christiansen, and R. Schwartz, *Programming Perl*. Sebastapol, CA: O'Reilly and Associates, 1996.

[32] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Yu, and M. Gerstein, "Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic," *Proteins*, vol. 48, pp. 682-95., 2002.

[33] M. B. Gerstein, R. Jansen, T. Johnson, B. Park, and W. Krebs, "Studying Macromolecular Motions in a Database Framework: From Structure to Sequence," in *Rigidity theory and applications*, *Fundamental Materials Science*, M. F. Thorpe and P. M. Duxbury, Eds. New York: Kluwer Academic/Plenum press, 1999, pp. 401-442.

[34] E. Martz, "Protein Explorer (software package),".: URL: http://www.umass.edu/microbio/chime/explorer/index.htm, 1999.

[35] A. M. Lesk and C. Chothia, "Mechanisms of Domain Closure in Proteins," *J. Mol. Biol.*, vol. 174, pp. 175-91, 1984.

[36] W. R. Pearson, "Rapid and sensitive sequence comparison with FASTP and FASTA," *Methods Enzymol*, vol. 183, pp. 63-98, 1990.

[37] Y. Kim, S. H. Eom, J. Wang, D. S. Lee, S. W. Suh, and T. A. Steitz, "Crystal structure of Thermus aquaticus DNA polymerase," *Nature*, vol. 376, pp. 612-6., 1995.

[38] B. V. Plapp, H. Eklund, T. A. Jones, and C. I. Branden, "Three-dimensional structure of isonicotinimidylated liver alcohol dehydrogenase," *J Biol Chem*, vol. 258, pp. 5537-47., 1983.

[39] H. Li, W. H. Hallows, J. S. Punzi, V. E. Marquez, H. L. Carrell, K. W. Pankiewicz, K. A. Watanabe, and B. M. Goldstein, "Crystallographic studies of two alcohol dehydrogenase-bound analogues of thiazole-4-carboxamide adenine dinucleotide (TAD), the active anabolite of the antitumor agent tiazofurin," *Biochemistry*, vol. 33, pp. 23-32., 1994.

[40] D. W. Heinz and B. W. Matthews, "Rapid crystallization of T4 lysozyme by intermolecular disulfide cross- linking," *Protein Eng*, vol. 7, pp. 301-7., 1994.