

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

Werner G. Krebs and Phil Bourne<sup>1,\*</sup>

Integrative Biosciences  
San Diego Supercomputer Center MC 0505  
National Partnership for Advanced Computational Infrastructure  
<sup>1</sup>Department of Pharmacology  
University of California, San Diego,  
9500 Gilman Drive  
La Jolla, CA 92093-0505, USA

Authors' Email: [wkrebs@sdsc.edu](mailto:wkrebs@sdsc.edu), [bourne@sdsc.edu](mailto:bourne@sdsc.edu)\*

\*To whom correspondence should be addressed.

Keywords: Proteomics, protein mutations, conformational change, data visualization, morph movies, PDB mutation selection bias, Protein Mutant Resource, rational drug design, protein engineering, statistical data mining

Last modified: 3/30/2003 8:23:06 PM

## Abstract

Structural studies of the effects of non-silent mutations on protein conformational change are an important key in deciphering the language that relates protein amino acid primary structure to tertiary structure. Elsewhere, we presented the Protein Mutant Resource (PMR) database, a set of online tools that systematically identified groups of related mutant structures in the Protein DataBank (PDB), accurately inferred mutant classifications in the Gene Ontology using an innovative, statistically rigorous data-mining algorithm with more general applicability, and illustrated the relationship of these mutant structures via an intuitive user interface. Here, we perform a comprehensive statistical analysis of the effect of PMR mutations on protein tertiary structure. We find that, although the PMR does contain spectacular examples of conformational change, in general there is a counter-intuitive inverse relationship between conformational change (measured as C- $\alpha$  displacement or RMS of the core structure) and the number of mutations in a structure. That is, post-translational modifications by structural biologists present in the PDB contrast naturally evolved mutations. We compare the frequency of mutations in the PMR/PDB datasets against the accepted PAM250 natural amino acid mutation frequency to confirm these observations. We generated morph movies from PMR structure pairs using technology previously developed for the Macromolecular Motions Database (<http://molmovdb.org>), allowing bioinformaticians, geneticists, protein engineers, and rational drug designers to analyze visually the mechanisms of protein conformational change and distinguish between conformational change due to motions (e.g., ligand binding) and mutations. The PMR morph movies and statistics can be freely viewed from the PMR website, <http://pmr.sdsc.edu>.

## 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<sup>1-4</sup>.

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<sup>5-7</sup>, 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<sup>7,8</sup>.

Although databases of mutant gene products<sup>9-12</sup> as well as specialized databases of mutant protein structures have previously been developed<sup>13-15</sup>, the PMR<sup>16</sup> was the first PDB<sup>17</sup>-wide 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 mutation 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 post-translationally modify 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<sup>18,19</sup>. 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<sup>20,21</sup>. PMR database entries interacted with a number of external databases (MolMovDB<sup>22-25</sup>, GO<sup>26</sup>, PubMed/Entrez<sup>27</sup>, PDBsum<sup>28</sup>) 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 generate morph movies assist scientists in visually discriminating between conformational changes caused by protein motions<sup>25</sup> from those caused primarily by changes in protein sequence<sup>16</sup>. 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>).

## Materials and Methods

Elsewhere<sup>16</sup>, the PMR was generated by automatically clustering the PDB<sup>17</sup> at 95% sequence identity using the highly efficient, state-of-the-art **cd-hit** sequence clustering algorithm<sup>29</sup>. 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 the scientific literature. Software was developed to automatically find and add new PDB entries to existing PMR families where appropriate. The resulting data was loaded into Oracle tables and made freely accessible via a web interface, developed in Perl<sup>30</sup>, Oracle SQL, Chime, javascript, and HTML.

Here, we applied morph movie technology<sup>22,23</sup> originally developed for the Database of Macromolecular Motions<sup>22,24,25,31,32</sup> (<http://molmovdb.org>) to PMR data to make four-dimensional visual illustrations<sup>22,23,33</sup> of the PMR conformation changes available on the PMR website (<http://pmr.sdsc.edu>; Figure 1) and to compute a number of useful statistics, which we have tabulated (Table 1). The number of mutations per structure in PMR data was first tabulated (Figure 2). Then, morph analysis and an estimate of the structural change (in terms of C $\alpha$  displacement) within each possible, non-redundant wild-type and mutant structure pair in each PMR family was computed using our morph server algorithm and software<sup>23</sup>. Mean structural change (in terms of C $\alpha$  displacement) by mutant residue type in PMR data is given (Figure 3). The frequency of amino acid mutations in PMR data was tabulated and compared with the accepted PAM250 natural amino acid mutation frequencies<sup>34</sup> (Table 2 and Figure 4). A scatter-plot depicting the

relationship of number of mutations with structural change (as measured by previously computed C $\alpha$  displacements) is shown (Figure 5).

## Results

A morph analysis of all possible structure pair combinations within PMR families would have had  $O(N^2)$  complexity with size of family in terms of both disk space and CPU time and was not computationally tractable for us. Instead, we limited our analysis to non-redundant combinations of the members of each family and its wild-type, which had  $O(N)$  complexity and reduced the task by several orders of magnitude. Even so, approximately 4.3 gigabytes of disk storage were required to store the resulting morph movies and statistical data. Running in parallel, our morph server software required three days of CPU time on our sixteen-CPU cluster of four-CPU Sun Ultra-80 servers running SunOS 5.7. Rapid re-clustering of the PDB by **cd-hit**<sup>29</sup> together with many other algorithm design decisions within the morph server software pipeline<sup>23</sup> and elsewhere make tractable the automatic analysis of wild-type/new-structure pairs resulting from PDB updates.

Our computed morph movie results can be used to visually analyze the outliers shown in the PMR data (Figure 5). The largest structural change in the PMR occurs in TAQ DNA Polymerase<sup>35</sup>. 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 140 Å C $\alpha$  displacement, which makes a spectacular morph movie. Crystal contact artifacts may largely be ruled out in this case because 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 C $\alpha$  displacement. The 140 Å C $\alpha$  displacement structures are all bound with DNA, whereas the 1TAQ wild-type<sup>35</sup> is an unbound structure, suggesting 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. Our Internet-accessible morph movies confirm this and depict a small domain swinging a large distance on a hinge, presumably as part of the DNA binding process.

Analysis based on our morph movie results can also be applied to the second largest structural change between 7ADH<sup>36</sup> and the designated wild-type alcohol dehydrogenase structure 1ADG<sup>37</sup>. This has a displacement of 66 Å caused by structural changes resulting from 22 mutations. Our Internet-accessible morph movie shows chain breakages and backbone elements passing through each other, confirming that, unlike TAQ DNA polymerase, the structural change here is entirely due to mutation.

We find greater than 99% of PMR (and hence PDB) mutant structures have fewer than 10 mutations (Figure 2). PMR data suggests that alanine is the residue most commonly involved in mutations by crystallographers (Table 1) while single mutations involving glycine result in the largest structural changes (Figure 3). These results may provide a statistical characterization of mutations carried out in the course of structural biological experiments.

Most importantly, there is a significant difference in the mutational frequencies for PMR versus the natural PAM250 frequencies (Figure 4 and Table 2). Contrary to what one would expect, structural change in the PDB is inversely related to the number of mutations (Figure 5), further confirming a selection bias in PMR mutant structure data.

## Discussion and Conclusion

Our addition of morph analysis technology to the PMR to allow data visualization of conformational change illustrates the usefulness of this approach. Morph movies often make it easier to distinguish between conformational change induced by mutation (indicated in the movie as refolding, 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 plausible morph movies). We have also computed a number of statistics potentially useful to crystallographers, including mean structural change by mutant residue type (Figure 3).

Clearly, crystallographic mutation processes are different from natural mutation processes (Table 2 and Figures 4 and 5). Natural mutations are thought to be selected for preservation or enhancement of function. Crystallographers often mutate proteins specifically to induce crystallization<sup>38</sup> and select mutants that will produce interesting structures. Mutants that cannot be crystallized are absent from PMR data.

The selection of structures, particularly mutant structures, by x-ray crystallographers appears biased and qualitatively and quantitatively different from the evolutionary processes that select mutant proteins for inclusion in living species. The advent of high-throughput structural genomics and the long-standing trend of exponential growth in the number of solved structures each year<sup>39,40</sup>, may make it possible in the near future to design large scale mutagenic crystallization studies that select structural mutants in a manner more consistent with the practices of nature using the data of this and other statistical studies. Exhaustive structural mutagenesis studies would be interesting not only for the insight they provide to theoretical structural biologists but also because they provide valuable raw data for 'brute force' database-based approaches capable of combining data from many variant protein structures to predict more accurately the structure of novel proteins from primary sequence data. Future such studies can potentially use these and other results to describe and possibly compensate for this mutant structure selection bias.

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<sup>29</sup> and design decisions<sup>23</sup> within our software pipeline. These new results and methods complement existing, innovative statistical inference algorithms<sup>20,21</sup> used to generate data for other areas of the PMR website.

## 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.

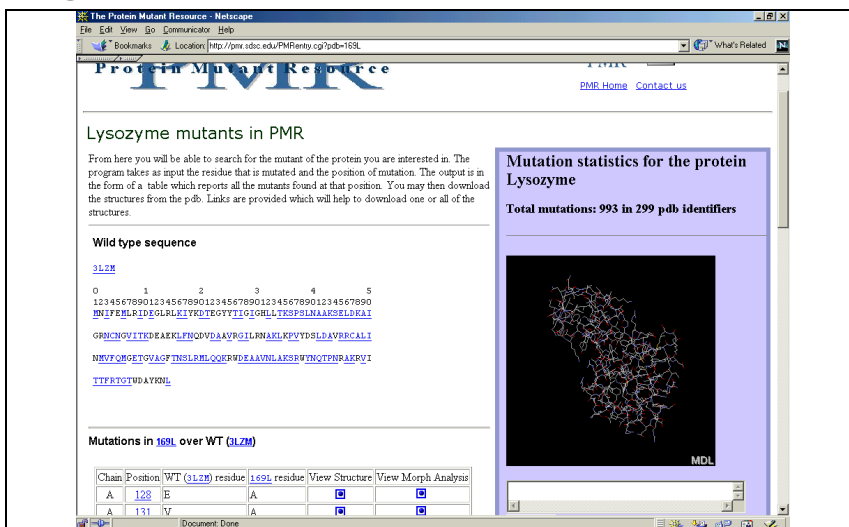
## References

1. Sabb, A. L. et al. Discovery of a highly potent, functionally-selective muscarinic M1 agonist, WAY-132983 using rational drug design and receptor modelling. *Bioorg Med Chem Lett* **9**, 1895-900. (1999).
2. Hakala, J. M. & Vihinen, M. Modelling the structure of the calcitonin gene-related peptide. *Protein Eng* **7**, 1069-75. (1994).

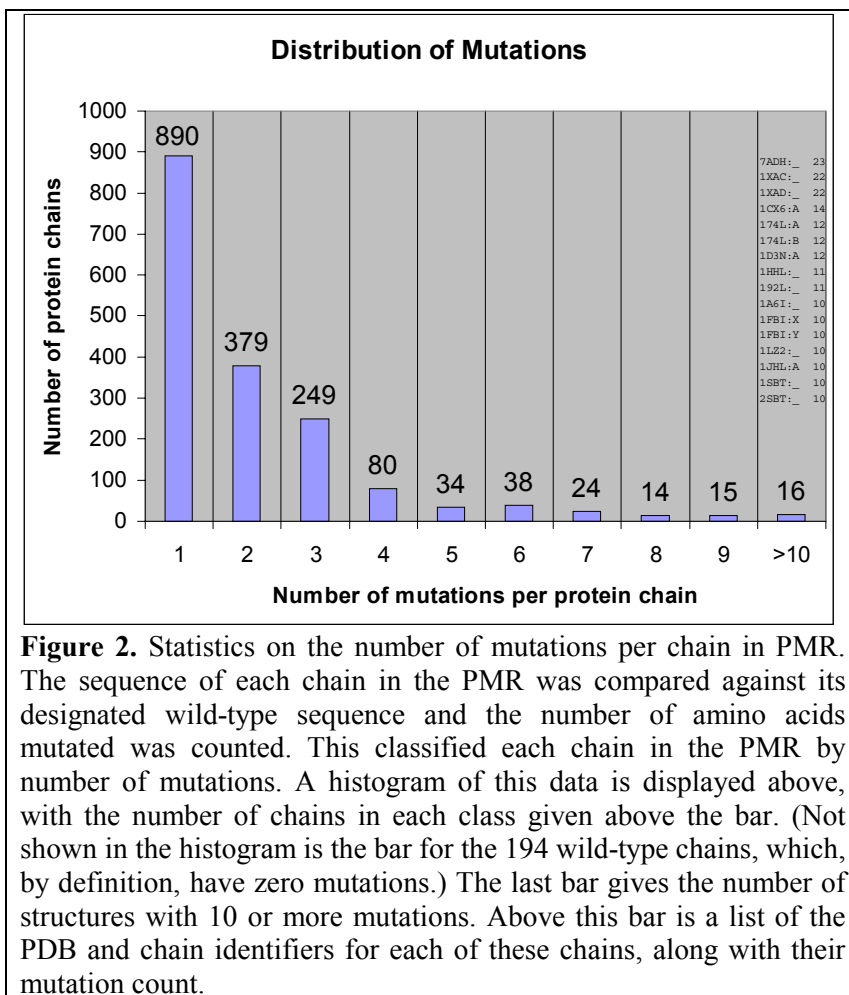
3. McKie, J. H. Homology modelling of the dihydrofolate reductase-thymidylate synthase bifunctional enzyme of *Leishmania major*, a potential target for rational drug design in leishmaniasis. *Drug Des Discov* **11**, 269-88. (1994).
4. Vriend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56 (1990).
5. Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* **92**, 10869-73. (1995).
6. Govindarajan, S. & Goldstein, R. A. Optimal local propensities for model proteins. *Proteins* **22**, 413-8. (1995).
7. Unger, R. & Moulton, J. Local interactions dominate folding in a simple protein model. *J Mol Biol* **259**, 988-94. (1996).
8. Li, H., Carrion-Vazquez, M., Oberhauser, A. F., Marszalek, P. E. & Fernandez, J. M. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat Struct Biol* **7**, 1117-20. (2000).
9. Beukers, M. W., Kristiansen, I., AP, I. J. & Edvardsen, I. TinyGRAP database: a bioinformatics tool to mine G-protein-coupled receptor mutant data. *Trends Pharmacol Sci* **20**, 475-7. (1999).
10. Kristiansen, K., Dahl, S. G. & Edvardsen, O. A database of mutants and effects of site-directed mutagenesis experiments on G protein-coupled receptors. *Proteins* **26**, 81-94. (1996).
11. Edvardsen, I. & Kristiansen, K. Computerization of mutant data: the tinyGRAP mutant database. *7TM journal* **6**, 1-6 (1997).
12. Maurer, S. M., Firestone, R. B. & Sriver, C. R. Science's neglected legacy. *Nature* **405**, 117-20. (2000).
13. Nishikawa, K. et al. Constructing a protein mutant database. *Protein Eng* **7**, 773 (1994).
14. Kawabata, T., Ota, M. & Nishikawa, K. The Protein Mutant Database. *Nucleic Acids Res* **27**, 355-7. (1999).
15. Gromiha, M. M. et al. ProTherm, Thermodynamic Database for Proteins and Mutants: developments in version 3.0. *Nucleic Acids Res* **30**, 301-2. (2002).
16. Krebs, W. G., Nair, T. M. & Bourne, P. E. The Protein Mutant Resource: A Tool for Protein Engineering. *Manuscript in preparation*. (2003).
17. Berman, H., 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.* **28**, 235-242 (2000).
18. Shindyalov, I. N. & Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* **11**, 739-47. (1998).
19. Levitt, M. *STRUCTAL. A structural alignment program* (Stanford University, 1994).
20. Krebs, W. G. & Bourne, P. E. Statistically Rigorous Automated Gene Annotation and Classification and its Application to Protein Data Bank Sequences using Gene Ontology Terms. *Manuscript in peer-review*. (2003).
21. Krebs, W. G. & Bourne, P. E. (U.S. Patent Pending).
22. Krebs, W. et al. Progress in Protein Motions: Experimental and Mathematical Methods, Tools, Databases, and Hierarchical Classification Schemes. *Methods Enzymol* **In press** (2003).
23. Krebs, W. G. & Gerstein, M. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res* **28**, 1665-1675 (2000).
24. Echols, N., Milburn, D. & Gerstein, M. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res* **31**, 478-82. (2003).
25. Gerstein, M. & Krebs, W. A Database of Macromolecular Movements. *Nucl. Acids Res* **26**, 4280 (1998).
26. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
27. Schuler, G. D., Epstein, J. A., Ohkawa, H. & Kans, J. A. Entrez: molecular biology database and retrieval system. *Methods Enzymol* **266**, 141-62 (1996).
28. Laskowski, R. A., 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.* **22**, 488-490 (1997).
29. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282-3. (2001).

30. Wall, L., Christiansen, D. & Schwartz, R. *Programming Perl* (O'Reilly and Associates, Sebastapol, CA, 1996).
31. Krebs, W. G. et al. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins* **48**, 682-95. (2002).
32. Gerstein, M. B., Jansen, R., Johnson, T., Park, B. & Krebs, W. in *Rigidity theory and applications* (eds. Thorpe, M. F. & Duxbury, P. M.) 401-442 (Kluwer Academic/Plenum press, 1999).
33. Martz, E. (URL: <http://www.umass.edu/microbio/chime/explorer/index.htm>, 1999).
34. Pearson, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* **183**, 63-98 (1990).
35. Kim, Y. et al. Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376**, 612-6. (1995).
36. Plapp, B. V., Eklund, H., Jones, T. A. & Branden, C. I. Three-dimensional structure of isonicotinimidylated liver alcohol dehydrogenase. *J Biol Chem* **258**, 5537-47. (1983).
37. Li, H. et al. Crystallographic studies of two alcohol dehydrogenase-bound analogues of thiazole-4-carboxamide adenine dinucleotide (TAD), the active anabolite of the antitumor agent tiazofurin. *Biochemistry* **33**, 23-32. (1994).
38. Heinz, D. W. & Matthews, B. W. Rapid crystallization of T4 lysozyme by intermolecular disulfide cross-linking. *Protein Eng* **7**, 301-7. (1994).
39. Orengo, C. A., Jones, D. T. & Thornton, J. M. Protein superfamilies and domain superfolds. *Nature* **372**, 631-634 (1994).
40. Holm, L. & Sander, C. Mapping the Protein Universe. *Science* **273**, 595-602 (1996).

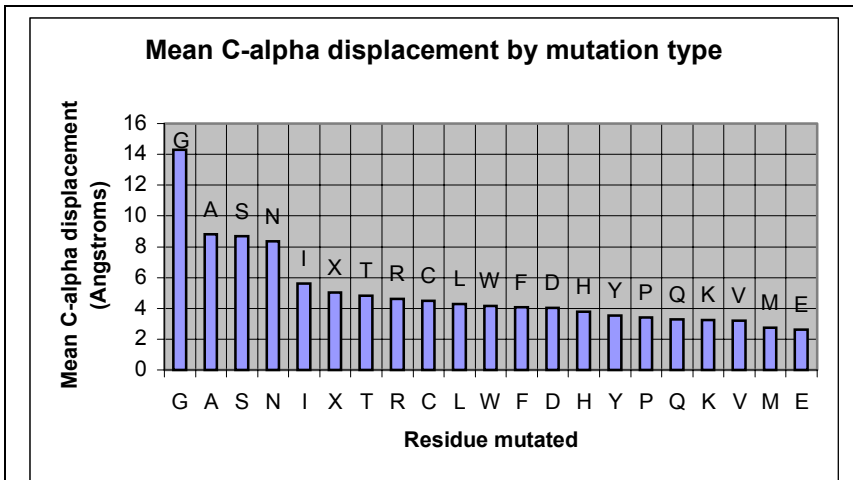
## Figures



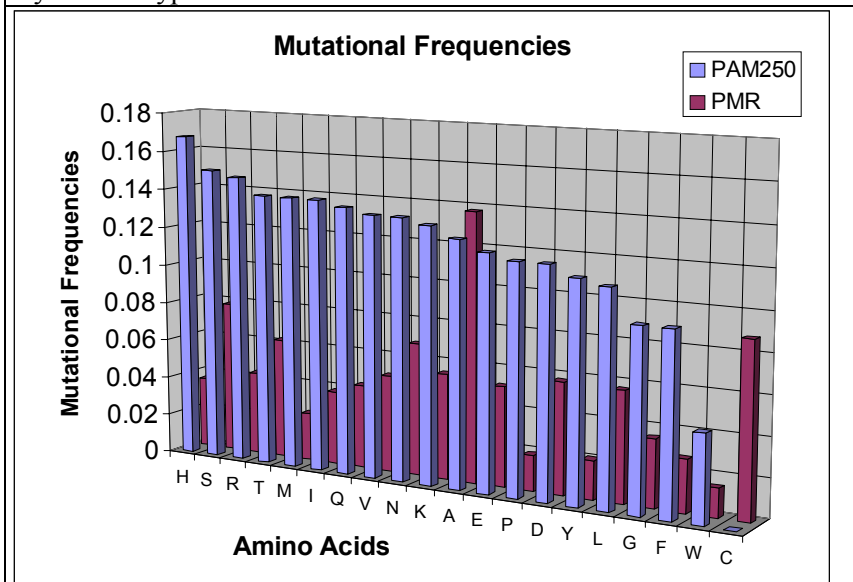
**Figure 1.** Web screen shot of PMR mutant browser. This figure gives a screen shot of the interface allowing the user to select a specific mutational site to list the mutants available at that site. Also shown is a listing of the mutations for a particular mutant PDB ID. The user may request a number of analyses on the selected PDB ID, including 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.



**Figure 2.** Statistics on the number of mutations per chain in PMR. The sequence of each chain in the PMR was compared against its designated wild-type sequence and the number of amino acids mutated was counted. This classified each chain in the PMR by number of mutations. A histogram of this data is displayed above, with the number of chains in each class given above the bar. (Not shown in the histogram is the bar for the 194 wild-type chains, which, by definition, have zero mutations.) The last bar gives the number of structures with 10 or more mutations. Above this bar is a list of the PDB and chain identifiers for each of these chains, along with their mutation count.



**Figure 3.** Mean C $\alpha$  displacement by mutant residue. The C $\alpha$  displacement of chains with only one mutation separating their sequences from their wild-type structure was determined and averaged by residue type.



**Figure 4.** Bar chart of mutational frequencies of amino acids in the PMR and the accepted natural PAM250 amino acid mutation rate<sup>34</sup>. 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, which are given by the blue bars in the foreground. The red bars in the background give the PMR values.

## Mutations versus Structural Change

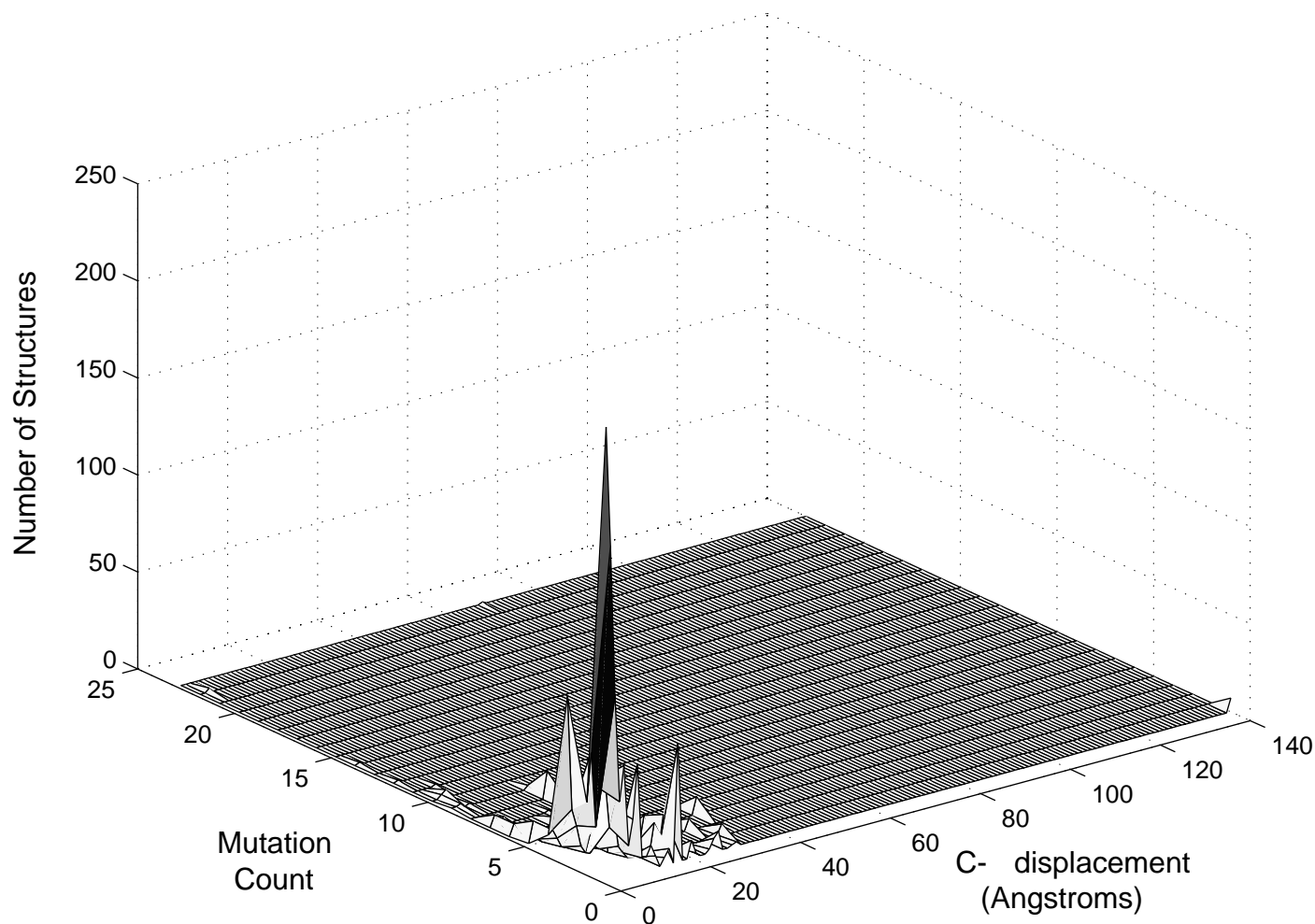


Figure 5. Number of mutations versus structural change (displacement). Chains in the PMR were sorted by the number of mutated amino acids between their sequence and their designated wild-type sequence. A histogram was then computed for each class using a consistent set of bins. The resulting data was then plotted in 3D. The X-axis gives the number of mutations between each chain and its wild-type structure, while the Y-axis gives the C displacement, a measure of structural change. The Z-axis gives the density of structures at any point on the graph. Examination of the X- and Y-axes reveals that the PMR shows an inverse relationship between the number of mutations and the amount of structural change. As discussed in the paper, this is consistent with the crystallographic experimental practices that generated the raw PMR data but inconsistent with the natural selection processes that evolved protein structure. Examination of the Z-axis shows that there are several hotspots in the dataset where many structures are concentrated.

## Tables

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

**Table 1.** Entry statistics for the PMR database. The PMR identifies structures as protein chains consisting of a PDB ID and a PDB chain identifier, of which there are 3343 in the PMR database. Of these, 194 have been designated wild-type. The PMR identifies 1157 mutation sites, giving an average of 2.9 mutations for each mutation site. Alanine is the most common amino acid found at mutation sites.

	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 *

**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<sup>34</sup>. Lower (more negative) numbers indicate an observed reduced tendency to mutate. PMR values along the diagonal are undefined, so only PAM250 values are given along the diagonal. Values in which the difference between the PAM250 accepted mutation rate and the PMR rate exceed more than half of the range are shown in yellow; about 5% of the values are designated this way. The table shows that the interconversion of phenylalanine to tyrosine seems disfavored by crystallographers despite a relatively high frequency of mutation in natural processes. In general, however, the yellow cells indicate mutations preferred by crystallographers that occur very rarely in natural mutation processes, such as the conversion of alanine to cysteine, which is used to stabilize crystal structures.