American Physical Society, Northwest Section, May 17 2007

# Single-Point Mutations and Protein Structure Deviations

Roy Campbell
Physics Department, Walla Walla University

## Introduction

The V4 proteins, LmcA and LmcB, in the social amoeba Dictyostelium are believed to play an important role in the growth to development transition of the organism. Whether or not a normal transition occurs depends on whether one or the other or both of the proteins is expressed [1]. The proteins only differ by two residues. In order to understand the possible differences in protein structure and function we performed an *ab initio* structure prediction computation for each protein.

## Structure Prediction

The predicted structures of the V4 proteins of Dictyostelium were calculated using the Monte Carlo fragment insertion method implemented by Rosetta [2], a suite of programs, scripts and files developed by David Baker and others at the University of Washington. The computation ran on the Robetta Server [3] and the results are shown in Figure 1. We were surprised to see such a large change in structure since the proteins only differed by one non-conserved residue

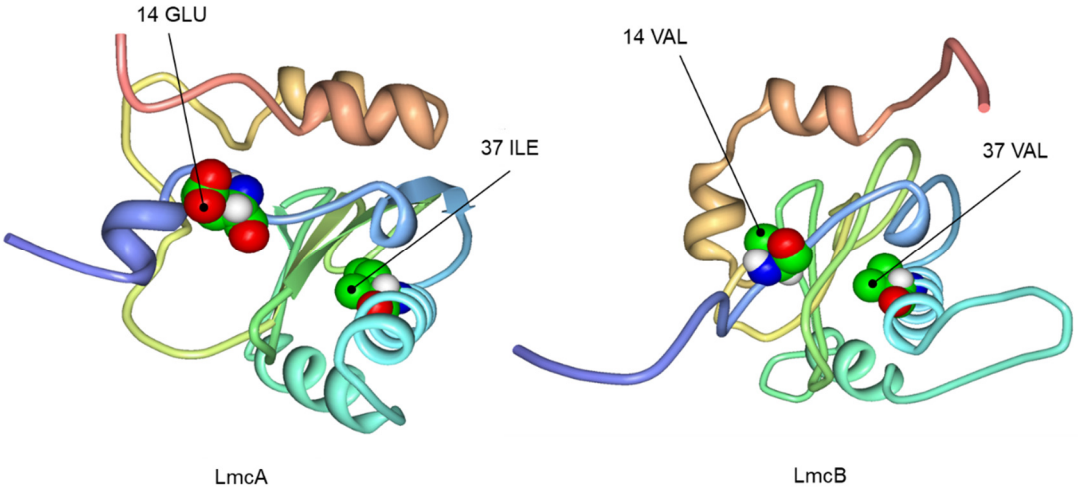


Figure 1. The predicted structures for the V4 proteins of Dictyostelium. LmcA and LmcB differ by two residues. The Glutamic acid at position 14 and the Isoleucine at position 37 in LmcA are each replaced by Valine in LmcB. Since Isoleucine and Valine are both hydrophobic this substitution is said to be conserved and is not expected to have a significant effect on structure. On the other hand, Glutamic Acid is polar, not hydrophobic, and this substitution is expected to be significant. The rms difference in atomic positions for these proteins is 3.02 angstroms.

Single-point Mutations

We did not expect the fragment substitution algorithm to be sensitive to a single mutation. However, it's possible that if a radical substitution, such as Valine for Glutamic Acid, occurred in a high probability folding initiation site, the structure modification early in the folding process could lead to significant structure differences [6]. To find examples of single-point mutations with significant structure differences for known structures, A comprehensive survey of the RCSB Protein Data Base (PDB) [7]. We searched the non-redundant PDB for single-point mutations. Approximately 5800 pairs of proteins differing by one residue were found. The structure files were compared and the root-mean-square (rms) deviations in atomic coordinates were calculated using a combinatorial extension method [8].

The distribution of mutations with respect to the resulting rms deviations, Figure 3, was found to have a simple exponential behavior. In the tangentially related field of theoretical population genetics, it has been shown that the distribution of beneficial fitness effects at a gene is exponential [9]. However, to my knowledge, the exponential distribution of mutations with respect to deviations in structure has not been previously reported.

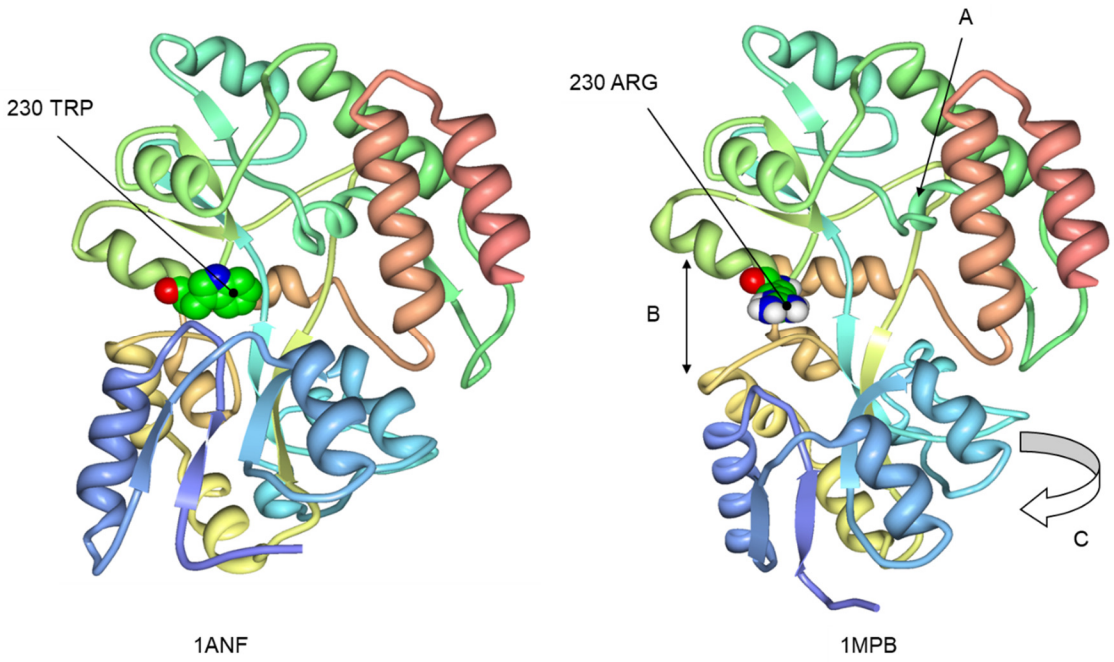


Figure 2. 1ANF [4] is a maltodextrin binding protein and 1MPB [5] is its mutant with arginine replacing tryptophan at position 230. This is an example of how a single-point mutation can affect structure. The two diagrams were oriented so that the upper parts of the structures were aligned as closely as possible. The side chains for the residues at position 230 are shown. There are three apparent changes in structure: The alpha helix labeled A is significantly distorted in 1MPB. The gap labeled B is wider in 1MPB and the lower part of the 1MPB structure is twisted with respect to the lower part of 1ANF; this is indicated by label C. The rms difference in atomic positions for these proteins is 3.13 angstroms.

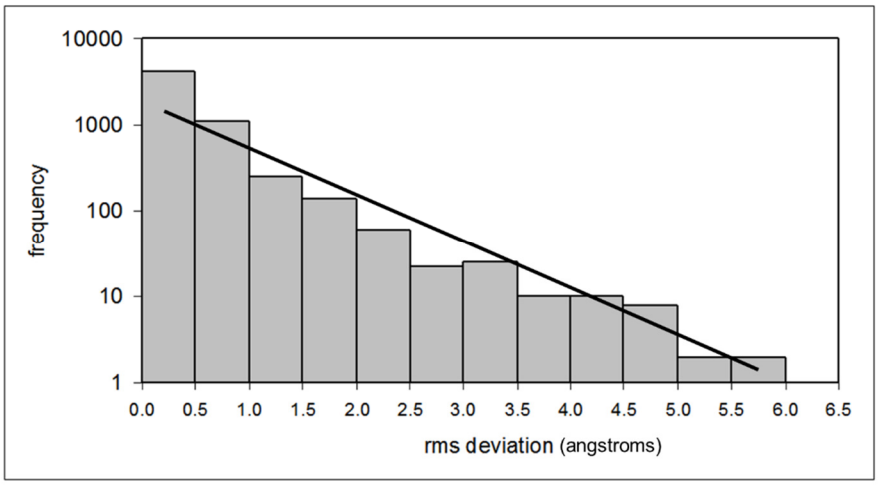


Figure 3. The frequency of mutations verses the resulting rms deviation in structure for the 5800 single point mutations found in the RCSB Protein Data Bank. The line shows an exponential fit to the distribution.

We can interpret this exponential result as follows. The probability that a mutation will result in a protein that will **not** be found in the PDB is proportional to the resulting rms deviation in atomic coordinates.

Folding Pathways

One common method for calculating the effect of a mutation on protein structure starts with the folded structure of the non-mutated protein, residues are substituted and changes in structure or stability are calculated [10]. This computational method ignores the possibility that a mutation may result in a small change when applied to an already folded state but could cause a dramatic difference in structure by effecting the folding pathway early in the folding process. This is, in the language of the Rosetta group, equivalent to the mutation occurring in a high-probability folding initiation site.

Figure 4 illustrates this possibility for a hypothetical protein. Folding pathways are shown following the valleys of the energy landscape, searching for the lowest energy state. This landscape is shown as a function of just two coordinates. For a real protein the energy is a function of as few as three and as many as seven angles for each amino acid residue. In order to understand and be able to predict protein structure it is absolutely necessary to understand folding pathways [11].

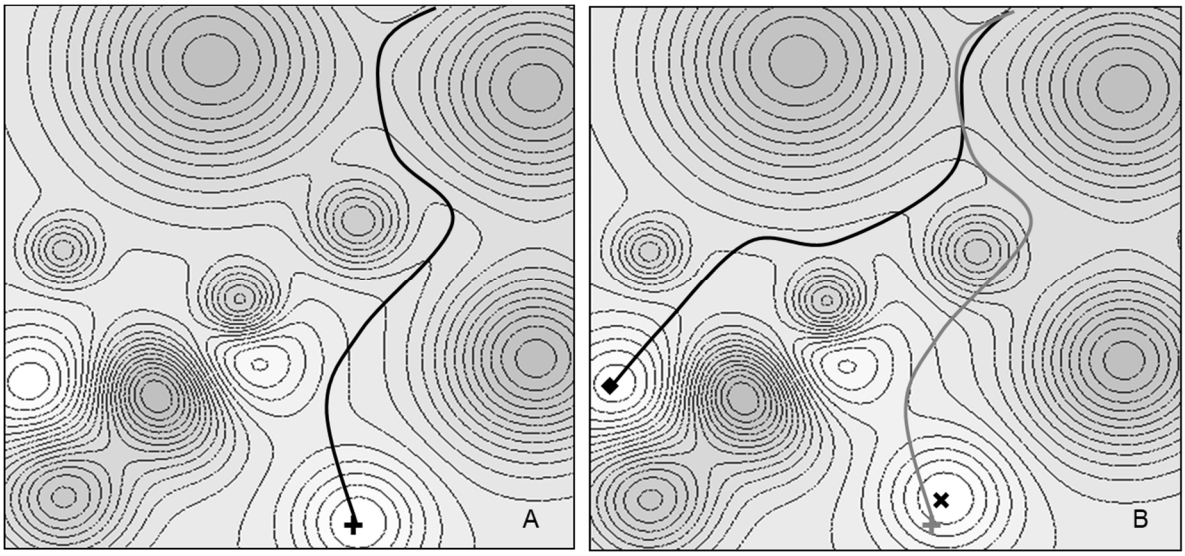


Figure 4. Map A represents the energy landscape of a hypothetical protein showing the folding pathway (▬) leading to the native folded state, indicated with a cross (✚). Map B represents the energy landscape after a mutation. The differences are subtle. The effect of the mutation on the folded state of the non-mutated protein (✖) and the folding pathway of the mutated protein (▬) are shown leading to the actual folded state of the mutated protein (◆).

## Conclusions

1) The following hypothesis is reasonable and should be tested further. *In vivo, the probability that a single-point mutation results in an **unsuitable** protein is proportional to the resulting rms deviation in atomic coordinates.*

2) It is worthwhile to determine the structures of the V4 proteins experimentally. To that end we are currently in the process of isolating and purifying the proteins.

## References

[1] David Lindsey, (Private Communication)

[2] R. Bonneau, C.E. Strauss, C.A. Rohl, D. Chivian, P. Bradley, L. Malmstrom, T. Robertson, and D. Baker, J. Mol. Biol. **322**, 65 (2002).

[3] David E. Kim, Dylan Chivian, and David Baker, Nucleic Acids Res. **32**, W526 (2004).

[4] F.A. Quiocho, J.C. Spurlino, and L.E. Rodseth, Structure, **5**, 997 (1997).

[5] B.H. Shilton, H.A. Shuman, and S.L. Mowbray, J. Mol. Biol. **264**, 364 (1996).

[6] Chris Bystroff, (Private Communication).

[7] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, Nucleic Acids Res. **28**, 235 (2000).

[8] I.N. Shindyalov, P.E. Bourne, Protein Engineering, **11**, 739 (1998).

[9] H. Allen Orr, Genetics, **163**, 1519 (2003).

[10] Vijaya Parthiban, M. Michael Gromiha, and Dietmar Schomburg, Nucleic Acids Res. **34**, W239 (2006).

[11] Jeffrey Skolnick, Proc Natl. Acad. Sci. USA, **102**, 2265 (2005).