

Identification and Free Energy Simulations of Correlated Mutations in Proteins

Adam Gorecki,¹ Jacek Leluk,² Bogdan Lesyng,³

Keywords: correlated mutations, proteins, free energy simulations, molecular dynamics, molecular evolution

1 Abstract

Correlated mutations are frequently observed in homologous proteins when carrying out multiple alignment analyses. In practical terms this means that residue types in two or more aligned positions appear in a correlated way. If the protein structure is known, positions of the correlated mutations can be mapped into its 3D-model, this in turn allows to classify such mutations by distance between them. These are nearest-neighbor or long-range (remote) mutations. The first class can be explained by required, conserved molecular contacts. However, about 30-50% of all observable mutations belong to the second class and they are of particular interest, since they cannot be interpreted in terms of direct contacts. The purpose of our study is to identify correlated mutations in a test set of proteins, and to analyze the mutations using methods of statistical physics, in particular to compute the free energy changes when computationally changing one residue into another (computational alchemy). The simulation was focused on the members of lysozyme family (EC 3.2.1.17). We found a few sets of the correlated sites, and one set was selected which involved two positions. We chose `3lzt.pdb` as the reference protein molecule with the well known 3D structure, and selected homologous sequences using NCBI BLAST service, with the SwissProt database. Sequences were aligned using MultAlign service ([1]). Then, we used a Corm program, developed in our group (freely accessible as a Java applet from <http://bioinfo.icm.edu.pl/corm/>), to identify the sites suspected for being correlated. The Corm program searches positions being potentially correlated with other ones, by analysis of sets of residues observed in the other positions. Let $S(i)$ denotes a set of residues observed at position i in multiple alignment, $S(j|i, A)$ - set of residues at position j observed under condition that there is a residue A at position i (gap is also treated as a residue). The program assumes that the positions i, j are correlated when:

$$S(j|i, A) \cap S(j|i, B) = \emptyset \text{ for all } A, B \in S(i), A \neq B.$$

One position i can be correlated with many other positions j . Residues A, B should also satisfy a frequency condition - namely, the residue A should be observed in some minimum of observations, but the alignment column shouldn't be too conservative.

The selected positions from the multiple alignment procedure were mapped to 3D- positions in the protein chain of the PDB structure. We tried to identify correlation involving a low number of remote positions in 3D model. Among many hits we chose a pair of sequence

¹Department of Biophysics, Faculty of Physics, Warsaw University, Zwirki i Wigury 93, 02-089 Warsaw, Poland e-mail: agorecki@icm.edu.pl

²Interdisciplinary Centre for Mathematical and Computational Modelling, Zwirki i Wigury 93, 02-089 Warsaw, Poland e-mail: lulu@icm.edu.pl

³Interdisciplinary Centre for Mathematical and Computational Modelling and Faculty of Physics, Warsaw University, Zwirki i Wigury 93, 02-089 Warsaw, Poland e-mail: lesyng@icm.edu.pl

alignment positions, 103 and 122, corresponding to 99 and 118 respectively in PDB. Table 1 presents the correlated sites.

Statistics given by Corm for selected correlations are shown in Table 1. We have been simulating the free energy changes of the solvated protein 3lzt.pdb, using a virtual mutation procedures: VAL99 to LEU99, THR118 to ALA118, and both mutations simultaneously. The structure 3lzt.pdb was virtually titrated at $pH = 5.5$ using the UHBD software package. Next it was solvated in a water box, relaxed and thermalized at $T = 300K$ using the NAMD package ([2]). Residue mutations were carried out using a Free Energy Perturbation module, applying a dual topology structure for the initial and mutated residues. The free energy changes were computed using the following formula:

$$\Delta G_{a \rightarrow b} = -k_B T \sum_{k=1}^N \log \left\langle \exp \left[-\frac{\mathcal{H}(\mathbf{r}, \mathbf{p}; \lambda_{k+1}) - \mathcal{H}(\mathbf{r}, \mathbf{p}; \lambda_k)}{k_B T} \right] \right\rangle_k$$

where $\mathcal{H}(\lambda) = \mathcal{H}_i + \lambda \mathcal{H}_a + (1 - \lambda) \mathcal{H}_b$ - is a hybrid Hamiltonian, a, b - denotes two different protein structures - before and after mutation. Dual topology residues has two noninteracting side chains and common C_α atoms. Because of some numerical instabilities the coupling parameter λ varies between 0.001 and 0.999 and the final results are being extrapolated to 0 and 1, respectively. The free energy differences are of the order of 10 kcal/mol. The most important question is whether the free energy changes at the two selected sites are additive or not. Non-additive effects would indicate long-range cooperative effects in the proteins, which could have an evolutionary interpretation. First simulations show that some non-additivity exists. Currently we verify these results and continue our study by simulating other mutations.

Residues in 103(99)	Residues in 122 (118)
L (24 hits)	-ADE
V (71 hits)	HKQRST

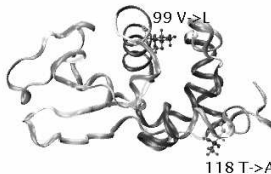


Table 1: Correlated mutation selected for simulation and protein 3lzt structure with virtual dual-topology residues in corresponding positions

These studies are supported by a PhD grant of A.Gorecki provided by Faculty of Physics, and by ICM funds of Warsaw University.

References

- [1] Corpet, F., 1988, Multiple sequence alignment with hierarchical clustering, *Nucl. Acids Res.*, 16 (22): 10881-10890.
- [2] Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., Schulten, K., 1999, NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.*, 151:283-312. (<http://prodes.toulouse.inra.fr/multalin/multalin.html>).
- [3] Humphrey, W., Dalke, A. and Schulten, K., 1996, VMD - Visual Molecular Dynamics', *J. Molec. Graphics*, 14.1 : 33-38.