Simulation of electric charge distribution of all proteins in *Saccharomyces cerevisiae* proteome

Runcong Ke  
ke@bp.nuap.nagoya-u.ac.jp  

Shigeki Mitaku  
mitaku@nuap.nagoya-u.ac.jp

Department of Applied Physics, Graduate School of Engineering, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8603, Japan

**Keywords**: proteome, electric charge distribution, simulation

## 1 Introduction

It is widely believed that the protein structures are determined by the combination of various physical interactions, among which the electrostatic interaction between electric charges is one of the most well-defined physical factors. Therefore, the electrostatic effects have been studied for various individual proteins so far [1, 2]. However, the analysis of the electric charge distribution of total proteins in a proteome has little been studied yet.

The electric charge distribution of total sequences in a proteome is the result of evolutionary processes, and there must be two kinds of evolutionary pressures on the variation of charges in sequences: On one hand, the sequences of electric charges are randomized by mutations of amino acid sequences. On the other hand, some of charged residues are preserved because of their functional and structural importance. Thus, interesting questions arise how random is the sequences of charged residues in a proteome and whether the attractive interaction is dominant for the formation of globular proteins.

In this work, we calculated the autocorrelation function of electric charges in all amino acid sequences of the *Saccharomyces cerevisiae* proteome. Furthermore, we carried out simulations of amino acid sequences so that the autocorrelation function of electric charges can be reproduced by simple rule of the generation of charged residues. The results indicated that the repulsive interaction was surprisingly dominant in the total proteomes and this characteristic feature of a proteome is partially reproducible, assuming the local positive correlation of charged residues, but the long range repulsion could not be reproduced easily.

## 2 Method and Results

All amino acid sequences from *S. cerevisiae* proteome were obtained from NCBI [3, 4]. We calculated the electric charges of proteins, assuming that Lys, Arg and His residues are positively charged, Asp and Glu are negatively charged and other residues are neutral. The autocorrelation function of electric charges was calculated for all amino acid sequences, as shown in Fig. 1. The significant feature of the result is the repulsive interaction dominant in average throughout the interval of sequences longer than 100 residues. When completely random sequences with the same length distribution and propensity of residues as *S. cerevisiae* were generated, the autocorrelation function was zero throughout the all intervals. Therefore, the systematic correlation in Fig. 1 should have some biological and structural meaning of proteins in genome scale. The analysis of the autocorrelation indicated that there are two components of repulsive interaction whose characteristic lengths were about 80 and 6 residues, respectively.

Then, we tried to reproduce the positive correlation of charges in Fig. 1 by the simulation of amino acid sequences, keeping the length distribution and the propensities of positively as well as negatively charged residues the same as the actual sequences in the proteome of *S. cerevisiae*. The total propensity of positively charged residues was 0.1391 and those of negatively charged residues and neutral ones were 0.1230 and 0.7379, respectively. When the random amino acid sequences were generated using these propensities, the autocorrelation function became reasonably constant of zero. Then, we introduced the positive correlation between the neighboring charges. The sequences generated by the simple correlation...
between neighboring charges showed the autocorrelation function of Fig. 2. The decay curves by this simulation showed single exponential curves with the characteristic length of about 6 residues that is very similar to the shorter decay constant in Fig. 1. However, the region of the interval longer than about 20 residues did not show any repulsive interactions. Therefore, the simulation by this rule of the charge correlation is not enough to reproduce the autocorrelation function in Fig. 1.

3 Discussions

The previous work of the net electric charge distribution of all proteins in *S. cerevisiae* showed a Gaussian-like distribution, suggesting some random process generating the sequences of charges in proteins. This work studied the correlation of electric charges within amino acid sequences, revealing the positive correlation of charges that means the repulsive interaction within proteins. Therefore, the Gaussian distribution of net charges of total proteins in a proteome is not completely random. The charge in proteins loses the correlation only at the interval longer than 100 residues. The question why proteins have such large cluster of charges is obscure but the plausible reason is the stability of protein structures: Repulsive forces between domains will stabilize very condensed solution of proteins and prevent the amorphous aggregation of domains in proteins.

References


