Thermal Stabilization Improvement of Adenylate Kinase by Residue Correlation Analysis and Sequence Entropy Comparison

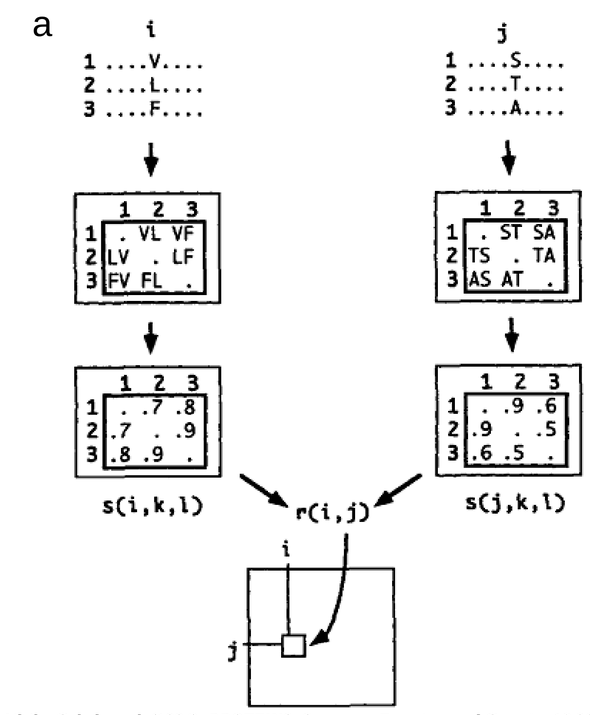
Jian Chang*, Chengxin Zhang†, Jialin Yang‡ and Yan-Wen Tan*
** State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai 200433, China*
† Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
‡ School of Life Science, Fudan University, Shanghai 200433, China

Introduction

Proteins with high thermal stability are of substantial importance in both industrial and laboratory settings. However, redesigning proteins for thermal stabilization improvement still remains challenging. Among all used techniques, bioinformatic methods are aimed to provide theoretical guidance for protein redesign without need of numerous experiment screenings, but almost all of them rely on tertiary structure information. In our study, a primary-sequence-based bioinformatic method was proposed for elevation of proteins' melting temperature. In this technique, residue correlation analysis (RCA) is applied to detect the coevolving residues related to enzymatic activity. Then sequence entropy comparison is used to determine the mutated residues with quantified divergences between amino acid composition of thermophilic protein sequence profile and mesophilic protein sequence profile. The scheme was applied to adenylate kinase (AK) from *Escherichia coli*. Four single mutants of AK were observed an increase of at least 4°C in melting temperature without drastic loss of catalysis activity, which indicated the usefulness of this technique.

Experimental Results

(1)



Ulrike Gobel, *et al*. (1994). *PROTEINS: Structure, Function, and Genetics* **18**,309-317

$$r_{ij} = \frac{2}{N(N-1)} \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} \frac{(s_{ikl} - \langle s_i \rangle)(s_{jkl} - \langle s_j \rangle)}{\sigma_i \sigma_j}$$

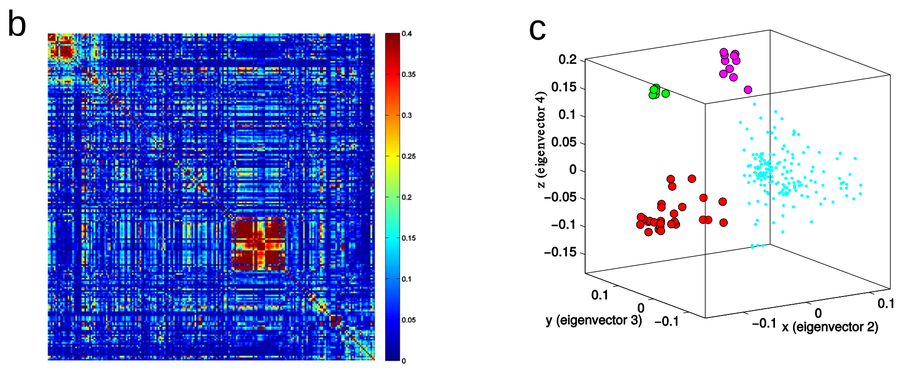
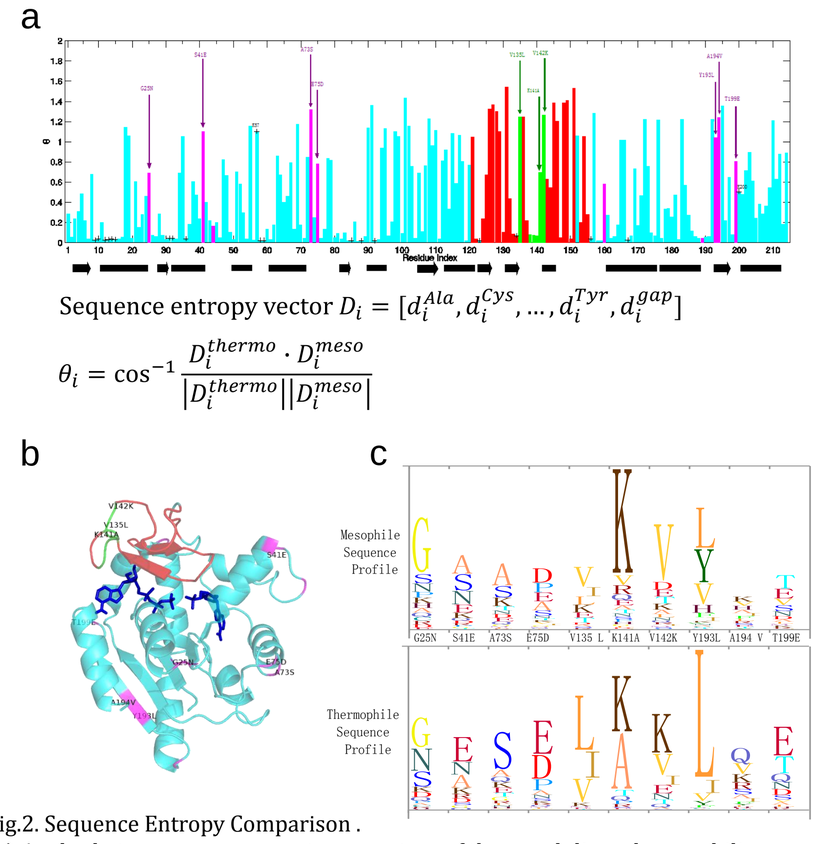


Fig.1. Residue Correlation Analysis.
(a) A protein family is presented as a multiple sequence alignment (series of horizontal lines). Mutational behavior at each single position is summarized in a mutation matrix. The mutation matrix contains the amino acid similarity of any pair of residues at that position (indices k, l run over proteins in the family; i, j run over common positions in the sequences). Correlation between position i and j could be quantified by Pearson Correlation Coefficient.
(b) Heat map representation of RCA matrix r_{ij} for adenylate kinase from *E. coli*. The largest region with high correlation indicates the lid domain of the protein AK.
(c) Principal component analysis of r_{ij}. A spectral decomposition of r_{ij} offers 214 eigenvalues and eigenvectors, which have different contributions to the correlations. Only the top 4 modes of r_{ij} are informative. The second, third and fourth eigenvectors are applied to a three dimensional scatter plot. All the 214 positions could be clustered into four sectors, colored green, red, magenta, and cyan.

Experimental Results

(2)



Sequence entropy vector $D_i = [d_i^{Ala}, d_i^{Cys}, \ldots, d_i^{Tyr}, d_i^{gap}]$

$$\theta_i = \cos^{-1} \frac{D_i^{thermo} \cdot D_i^{meso}}{|D_i^{thermo}||D_i^{meso}|}$$

Fig.2. Sequence Entropy Comparison .
(a) Angles between sequence entropy vectors of thermophilic and mesophilic sequence profile at each position. Bars were colored by sectors. Residues in direct contact with substrates were marked by black crossing. Selected mutation sites marked by arrow heads. d_i^{aa}, the relative entropy of amino acid aa at position i. The angle θ_i represented how divergent the distribution of amino acid types at position i was for thermophilic sequences and mesophilic sequences.
(b) Position of mutation sites at tertiary structure of AK.
(c) Amino acid distribution of mesophilic sequences (upper panel) and thermophilic sequences (lower panel) at mutation positions.

(3)

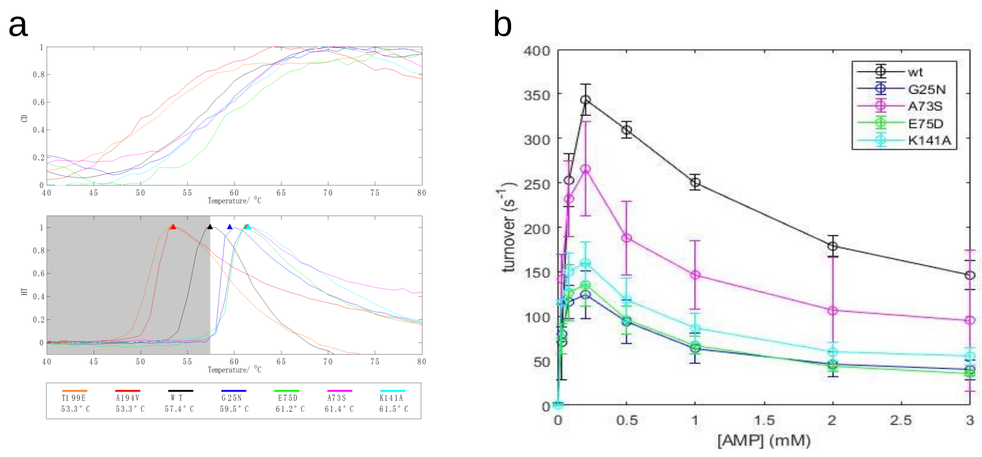


Fig.3. Thermostability and enzymatic activity of wild type and selected mutants of AK.
(a) Circular dichroism (CD) at 222nm of AK wild type and mutants T199E, A194V, G25N, E75D, A73S, and K141A from 40°C to 80°C. All data were normalized so that the maximum and minimum were one and zero.
(b) Forward enzymatic activity of AK wild type and six mutants.

Conclusions

Combining residue correlation analysis and sequence entropy comparison, we succeed in improving the thermal stability of adenylate kinase based on the primary sequence and with no need of any tertiary structure information. The thermally stable mutants exhibit enzymatic activities without dramatic loss compared to the wild type protein.