PREDICTION OF PROTEIN MUTANT STABILITY BY COMBINING OF SUPPORT VECTOR MACHINE WITH A COARSE GRAINED MODEL CALCUALTION

Ge Tan[1,2], Bingchuan Liu[2], Cheng Zhang[1], Bairong Shen[1,2,*]

[1]*Center for Systems Biology, Soochow University, 215006, China*

[2]*Department of Biomedical Engineering, Tongji University, 200092, China*

Email:bairong.shen@suda.edu.cn

A model combines of support vector machine with a coarse grained model is proposed to predict the stability changes of protein mutant. With optimized kernel function and attributes, a final accuracy of 84.61% was obtained with only 8 parameters in the task of predicting the ΔΔG sign which reflects the stability change of the protein.

1. BACKGROUND:

The prediction of the effect of amino acid substitutions to the protein structure stability provides invaluable information for the novel protein design, the protein biological function assignment and the understanding of the molecular mechanisms of diseases. Data about amino acid alterations, including rare amino acid mutations and common single nucleotide polymorphisms is accumulated rapidly due to the completion of the human genome project and the development of high throughput biological technology. To understand the effects of these alterations, computational models are preferred over the time-consuming and expensive experimental methods. Several models based on machine learning methods have been proposed to predict the stability change of mutated protein when a single amino acid mutation. However, these models, which mainly use the primary sequence information and the 20 amino acid codes, may become over-fitted when only limited data available. A model with reasonable number of parameters and structure information are necessary.

2. RESULTS:

Table 1 Comparison of different models

Models	Number of mutations	Number of parameters	Accuracy	MCC	References
Capriotti (SVM)	2048	42	0.77	0.42	[1]
Capriotti (NN)	1615	22	0.74	0.24	[2]
Cheng (SVM)	388	140	0.86	0.26	[3]
FoldX	388	8	0.75	0.25	[3]
Cheng (SVM)	1023	140	0.74	0.13	[3]
Cheng (SVM)	1135	140	0.78	0.31	[3]
Cheng (SVM)	1496	142	0.84	0.59	[3]
Shen (SVM)	1448	508	0.80	0.39	[4]
FoldX	1514	9	0.79	0.46	[5]
Ours (SVM)	1514	8	0.84	0.50	

To decrease the number of model parameters, we take the three dimensional structure information into account in our model by the calculation of the contact energy change with a coarse-grained model, then combine with the physicochemical features of the amino acids to develop a support vector machine model.

* Corresponding author.

The comparison of our model with others are listed in the Table I.

The eight parameters used in our model are hydrophobicity, isotropic surface area, electronic charge concentration, volume, contact energycalculated with RankviaContact [6] solvent accessibility (ASA), pH value and experimental temperature (T). As indicated in the Table I, with optimized kernel function and attributes, a final accuracy of 84.61% and MCC of 0.50 are obtained with only 8 parameters in the predicting the ΔΔG sign which performs much better than the previous machine learning methods and energy-functions-based methods such as FoldX.

Conclusions:

The problem of protein stability change is still appealing because of its significance in the understanding of diseases and the novel protein design in protein engineering. Although several models have been proposed to predict the stability change, the models are still not satisfactory. Our model improve the previous model by combining a quantative calculation based on structure information with more reasonable physicochemical attributes when train the support vector machine model.

Acknowledgment:

THIS WORK IS SUPPORTED BY THE NATIONAL NATURE SCIENCE FOUNDATION OF CHINA (20872107); THE AUTHOR GRATEFULLY ACKNOLOWLEDGES THE SUPPORT OF K.C WONG EDUCTATIONFOUNDATION, HONG KONG

1. CAPRIOTTI, E., ET AL., *PREDICTING PROTEIN STABILITY CHANGES FROM SEQUENCES USING SUPPORT VECTOR MACHINES*. BIOINFORMATICS, 2005. **21 SUPPL 2**: P. II54-8.

2. CAPRIOTTI, E., P. FARISELLI, AND R. CASADIO, *A NEURAL-NETWORK-BASED METHOD FOR PREDICTING PROTEIN STABILITY CHANGES UPON SINGLE POINT MUTATIONS*. BIOINFORMATICS, 2004. **20 SUPPL 1**: P. I63-8.

3. CHENG, J., A. RANDALL, AND P. BALDI, *PREDICTION OF PROTEIN STABILITY CHANGES FOR SINGLE-SITE MUTATIONS USING SUPPORT VECTOR MACHINES*. PROTEINS, 2006. **62**(4): P. 1125-32.

4. SHEN, B., J. BAI, AND M. VIHINEN, *PHYSICOCHEMICAL FEATURE-BASED CLASSIFICATION OF AMINO ACID MUTATIONS*. PROTEIN ENG DES SEL, 2008. **21**(1): P. 37-44.

5. GUEROIS, R., J.E. NIELSEN, AND L. SERRANO, *PREDICTING CHANGES IN THE STABILITY OF PROTEINS AND PROTEIN COMPLEXES: A STUDY OF MORE THAN 1000 MUTATIONS*. J MOL BIOL, 2002. **320**(2): P. 369-87.

6. SHEN, B. AND M. VIHINEN, *RANKVIACONTACT: RANKING AND VISUALIZATION OF AMINO ACID CONTACTS*. BIOINFORMATICS, 2003. **19**(16): P. 2161-2.