

doi:10.3969/j.issn.1672-5565.2014.03.12

# 蛋白质序列的混合特征值对折叠速率的影响

李欣颖,白凤兰\*

(大连交通大学理学院,辽宁 大连 116028)

**摘要:** 鉴于蛋白质折叠速率预测对研究其蛋白质功能的重要性,许多的科研工作者都开始对影响蛋白质折叠速率的因素进行研究。各种预测参数和方法被提出。利用蛋白质编码序列的不同特征参数,不同的二级结构及不同的折叠类的蛋白质对折叠速率的不同影响,我们选取蛋白质编码序列的新的特征值,即选取蛋白质序列的 *LZ* 复杂度,等电点等特征值。然后把把这些特征值与 20 种氨基酸的属性  $\alpha_c$ 、 $C_\alpha$ 、 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $PI$ 、 $\Delta G_{hd}$ 、 $N_m$ 、 $LZ$ 、 $Mu$ 、 $El$  融合,建立多元线性回归模型,并利用回归模型计算了 13 个全  $\alpha$  类蛋白质、18 个全  $\beta$  类蛋白质、13 个混合类蛋白质和 39 个未分类蛋白质的  $\ln(k_f)$  与预测值之间的相关系数分别达到 0.89、0.93、0.98、0.86。在 Jack-knife 方法的验证下发现在不同的结构中混合特征值与相应折叠速率有很好的相关性。结果表明,在蛋白质折叠过程中,蛋白质序列的 *LZ* 复杂度、等电点等特征值可能影响蛋白质的折叠速率及其结构。

**关键词:** 蛋白质序列;特征值;折叠速率;相关系数

**中图分类号:** Q523      **文献标志码:** A      **文章编号:** 1672-5565(2014)-03-227-05

## The impact on the folding rate by mixed eigenvalues of protein sequences

LI Xinying, BAI Fenglan\*

(Dalian Jiaotong University school of science, Dalian 116028, China)

**Abstract:** Given the importance of protein folding rate prediction on protein function analysis, many researchers have begun to study the influential factors of protein folding rate. Several prediction parameters and methods have been proposed due to the different eigenvalues of protein-coding sequences, using the different effects on the folding rate by different secondary structures and folding classes of proteins, we selected some characteristic values of protein coding sequence the characteristics of *LZ* complexity of protein-coding sequence and isoelectric point. Then we integrated these eigenvalues and properties of 20 kinds of amino acids  $\alpha_c$ 、 $C_\alpha$ 、 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $PI$ 、 $\Delta G_{hd}$ 、 $N_m$ 、 $LZ$ 、 $Mu$ 、 $El$ , and established a linear regression model. Using the regression model to calculate the correlation coefficient between  $\ln(k_f)$  and predicted values for 13 all- $\alpha$  proteins, 18 all- $\beta$  proteins, 13 mixed class proteins, 39 unclassified proteins as 0.89, 0.93, 0.98, 0.86, respectively. We used Jack-knife to test our model and it was found that the classification of protein into different structural classes reflected a good correlation between mixed eigenvalues and protein folding rates. The result shows that the characteristics of *LZ* complexity of protein-coding sequence and isoelectric point may affect the structure and the protein in the protein folding rate.

**Keywords:** Protein sequence; Eigenvalue; Folding rate; Correlation coefficient

由于蛋白质折叠速率预测对研究其蛋白质功能的重要性,许多的科研工作者都开始对影响蛋白质折叠速率的因素进行研究。各种预测参数和方法被提出,已有的预测方法大致可以分为三大类<sup>[1-2]</sup>:第一类是基于三级结构的预测方法,如接触序

(Contact order, CO)<sup>[3]</sup>、总接触距离(The total contact distance, TCD)等<sup>[4]</sup>。然而,要得到这类方法所必须的三级结构信息,由于成本高、周期长,很难达到快速预测的目的;第二类是基于二级结构的预测方法<sup>[5-6]</sup>,而这类方法所需的二级结构信息,也是由分

收稿日期:2014-05-08;修回日期:2014-06-02.

作者简介:李欣颖,女,硕士生,研究方向:计算数学;E-mail:lxystation@163.com.

\*通信作者:白凤兰,女,教授,研究方向:计算数学;E-mail:bfl0219@163.com.

子实验得到,或从一级序列预测得到,而这种预测方法受到二级结构的限制;第三类是基于一级序列的预测方法<sup>[7-8]</sup>,这类方法是直接从氨基酸序列出发预测蛋白质的折叠速率。这类方法主要概括以下几个方面<sup>[9]</sup>:(1)基于蛋白质肽链长度的预测方法<sup>[6,8]</sup>;(2)基于氨基酸的组成信息的预测方法<sup>[10-11]</sup>;(3)基于氨基酸的物化性的预测方法<sup>[12-13]</sup>;(4)基于氨基酸相互作用的预测方法<sup>[14]</sup>。这类方法的共同点就是从蛋白质的编码序列提取对折叠速率有影响的特征信息,然后用这些有效特征信息从不同角度描述蛋白质的折叠进程。然而蛋白质结构复杂并含有丰富的生物学信息,仅从这些信息来研究蛋白质的折叠速率远远不够,只有不断地从多方面挖掘对折叠速率有效的特征参数,并融合多种特征参数,就有可能取得好的结果。本文在已有研究蛋白质序列的特征参数对蛋白质折叠速率的影响基础上,为了从多个角度印证蛋白质序列的特征值对蛋白质折叠速率的影响,提取蛋白质序列的多种特征值,分析了这些特征值与不同的结构及不同折叠类蛋白质的折叠速率之间的相关性,印证了这些特征值可能对蛋白质折叠速率有影响。

## 1 材料和方法

### 1.1 数据

本文从文献[15-18]中选取了83个蛋白质,氨基酸序列从PDB库取得<sup>[19]</sup>。在选取44个蛋白质中包括13个全 $\alpha$ 类蛋白质,18个全 $\beta$ 类蛋白质和13个混合类蛋白质,以及39个未分类蛋白质。

### 1.2 蛋白质编码序列的特征值提取

氨基酸的标准化属性 $P_{\text{norm}}(i)$ 的计算公式为:

$$P_{\text{norm}}(i) = \frac{P(i) - P_{\min}}{P_{\max} - P_{\min}} \quad (1)$$

其中, $P_{\text{norm}}(i)$ 是氨基酸的标准化属性, $P(i)$ 是氨基酸序列中第 $i$ 个残基的属性, $P_{\max}$ 和 $P_{\min}$ 分别表示氨基酸属性中的最大值和最小值。

蛋白质序列中氨基酸的平均属性 $P_{\text{ave}}$ 的计算公式:

$$P_{\text{ave}} = \sum_{j=1}^N P(j)/N \quad (2)$$

其中, $P_{\text{ave}}$ 是蛋白质的氨基酸平均属性, $P(j)$ 是氨基酸序列中第 $j$ 个残基的属性, $N$ 是氨基酸序列的残基数。

蛋白质序列的复杂度 $LZ_c$ 计算公式:

$$LZ_c = \frac{1}{100}c(S) \quad (3)$$

其中, $S$ 表示的是序列, $c(S)$ 是序列 $S$ 的复杂度<sup>[11]</sup>。

20个氨基酸 $\alpha_c$ 、 $C_\alpha$ 、 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $PI$ 、 $\Delta G_{hd}$ 、 $N_m$ 、 $Mu$ 、 $EL$ 属性利用公式(1)计算出标准化后的值。

其中, $\alpha_c$ 是 $\alpha$ 螺旋的C端动力<sup>[20-21]</sup>, $C_\alpha$ 是 $\alpha$ 螺旋接触面积<sup>[15]</sup>, $K^0$ 是可压缩性<sup>[22-23]</sup>, $P_\beta$ 是 $\beta$ 折叠趋势<sup>[21]</sup>, $R_a$ 是在溶剂中的收缩率<sup>[24]</sup>, $\Delta ASA$ 是溶剂可及表面积<sup>[25]</sup>, $PI$ (at 25°C)表示氨基酸的等电点<sup>[26]</sup>, $\Delta G_{hd}$ 是吉布斯自由能变性蛋白水化的变化 $N_m$ 是平均中程接触, $Mu$ 是折射率, $EL$ 是长距离的非键能<sup>[15]</sup>。

利用20个氨基酸标准化后的值和公式(2)、(3)计算了13个全 $\alpha$ 类蛋白质,18个全 $\beta$ 类蛋白质和13个混合类蛋白质 $\alpha_c$ 、 $C_\alpha$ 、 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $PI$ 、 $\Delta G_{hd}$ 、 $N_m$ 、 $LZ_c$ 、 $Mu$ 、 $EL$ 的特征值,以及39个蛋白质的 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $Mu$ 、 $EL$ 的特征值,由于数据多没列在文章里。

## 2 结果与讨论

首先,利用多元线性回归函数分别计算了13个全 $\alpha$ 类蛋白质、18个全 $\beta$ 类蛋白质、13个混合类蛋白质和未分类的39个蛋白质的12种特征值与折叠速率之间的相关性,实验值与预测值之间的相关系数分别达到了0.99、0.96、0.99、0.865,但是用Jack-knife方法检验,都得出 $p > 0.05$ 。由此可知,12种特征值当中某些特征值对蛋白质的折叠速率没有影响,这样经过多次试验,对于全 $\alpha$ 类蛋白质选取有效特征值 $C_\alpha$ 、 $R_a$ 、 $LZ_c$ ,对于全 $\beta$ 类蛋白质选取有效特征值 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $N_m$ ,对于混合类蛋白质选取有效特征值 $K^0$ 、 $\Delta ASA$ 、 $PI$ ,对于未分类的蛋白质选取有效特征值 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $Mu$ 、 $EL$ ,计算这些特征值与折叠速率 $\ln(k_f)$ 之间的相关性。

其次,分别对全 $\alpha$ 类蛋白质、全 $\beta$ 类蛋白质、混合类蛋白质和未分类的蛋白质的有效特征值与折叠速率做相关性分析,并与其它方法进行比较。

选取13个全 $\alpha$ 类蛋白质 $C_\alpha$ 、 $R_a$ 、 $LZ_c$ 3个特征值与折叠速率 $\ln(k_f)$ 做回归方程:

$$\ln(k_f) = 43.9 - 32.18C_\alpha - 44.46R_a - 19.37LZ_c \quad (4)$$

用 $p$ 值检验了方程(4)每一项特征值,每一特征值对应的 $p$ 值都小于0.05。实验值与预测值之间的相关系数 $R = 0.89$ 。用Jack-knife方法检验,得出 $R = 0.77$ 、 $t = 4.04$ 、 $p < 0.05$ 。

Gromiha文章中选取了6个全 $\alpha$ 类蛋白质的1个特征值 $\alpha_c$ 作线性回归,本文在6个数据的基础上

增加到13个全 $\alpha$ 类蛋白质,用特征值 $\alpha_c$ 作线性回归,得到回归方程:

$$\ln(k_f) = 8.39 - 1.64\alpha_c. \quad (5)$$

本文用 $p$ 值检验了得到的方程(5), $p>0.05$ 。实验值与预测值之间的相关系数 $R=0.03$ , $t=-0.097$ 。在选取13个全 $\alpha$ 类蛋白质的 $C_\alpha$ 、 $R_a$ 、 $LZ_c$ 3个特征值中,得到回归方程的相关系数为0.89,说明选取的特征值 $C_\alpha$ 、 $R_a$ 、 $LZ_c$ 对全 $\alpha$ 类蛋白质有影响。对比结果见表1。

表1 全 $\alpha$ 类蛋白质回归分析结果

Table 1 The results of all- $\alpha$  proteins regression analysis

Model	数据	特征值	$R$
Gromiha	13	$\alpha_c$	0.03
Our	13	$C_\alpha, R_a, LZ_c$	0.89

选取18个全 $\beta$ 类蛋白质的 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $N_m$ 5个特征值与折叠速率 $\ln(k_f)$ 做回归方程:

$$\ln(k_f) = 18.31 - 85.75K^0 + 89.12P_\beta + 94.95R_a - 78.09\Delta ASA - 45.27N_m. \quad (6)$$

用 $p$ 值检验了以上方程的每一项特征值,每一特征值对应的 $p$ 值都小于0.05。实验值与预测值之间的相关系数 $R=0.93$ 。用Jack-knife方法检验,得出 $R=0.78$ , $t=4.93$ , $p<0.001$ 。

同样18个全 $\beta$ 类蛋白质,选取Gromiha文章中的 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 4个特征值作线性回归,得到方程:

$$\ln(k_f) = -8.81 - 60.58K^0 + 138.43P_\beta + 68.85R_a - 133.48\Delta ASA. \quad (7)$$

用 $p$ 值检验了以上方程的每一项特征值,其对应的 $p$ 值都小于0.05。实验值与预测值之间的相关系数 $R=0.89$ 。用Jack-knife方法检验,得出 $R=0.73$ , $t=4.28$ , $p<0.001$ 。这说明本文选取的特征值 $K^0$ 、 $P_\beta$ 、 $R_a$ 、 $\Delta ASA$ 、 $N_m$ 对全 $\beta$ 类蛋白质有影响。对比结果见表2。

表2 全 $\beta$ 类蛋白质回归分析结果

Table 2 The results of all- $\beta$  proteins regression analysis

Model	数据	特征值	Jack-knife			
			$R$	$R$	$t$	$P$
Gromiha	18	$K^0, P_\beta, R_a, \Delta ASA$	0.89	0.73	4.28	<0.001
Our	18	$K^0, P_\beta, R_a, \Delta ASA, N_m$	0.93	0.78	4.93	<0.001

选取13个混合类蛋白质 $K^0$ 、 $\Delta ASA$ 、 $PI$ 三个特征值与折叠速率 $\ln(k_f)$ 做回归方程:

$$\ln(k_f) = 17.67 - 84.85K^0 + 27.36\Delta ASA + 24.87PI. \quad (8)$$

用 $p$ 值检验了以上方程的每一项特征值,每一特征值对应的 $p$ 值都小于0.05。实验值与预测值之间的

相关系数 $R=0.98$ 。用Jack-knife方法检验,得出 $R=0.97$ , $t=13.46$ , $p<0.001$ 。

同样的13个混合类蛋白质,选取Gromiha文章中的 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $\Delta G_{hd}$ 4个特征值作线性回归,得到方程:

$$\ln(k_f) = 18.73 - 66.63K^0 - 6.39R_a + 22.96\Delta ASA - 0.24\Delta G_{hd}. \quad (9)$$

用 $p$ 值检验了以上方程的每一项特征值,每一特征值对应的 $p$ 值都小于0.05。实验值与预测值之间的相关系数 $R=0.96$ 。用Jack-knife方法检验,得出 $R=0.91$ , $t=7.07$ , $p<0.001$ 。这说明本文选取的特征值 $K^0$ 、 $\Delta ASA$ 、 $PI$ 对混合类蛋白质有影响。对比结果见表3。

表3 混合类蛋白质回归分析结果

Table 3 The results of mixed class proteins regression analysis

Model	数据	特征值	Jack-knife			
			$R$	$R$	$t$	$P$
Gromiha	13	$K^0, R_a, \Delta ASA, \Delta G_{hd}$	0.96	0.91	7.07	<0.001
Our	13	$K^0, \Delta ASA, PI$	0.98	0.97	13.46	<0.001

对于未分类的39个蛋白质,选取5个特征值 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $Mu$ 、 $El$ 与折叠速率 $\ln(k_f)$ 作回归方程:

$$\ln(k_f) = 41.46 - 26.88K^0 - 47.89R_a + 136.13\Delta ASA - 60.15Mu - 95.33El. \quad (10)$$

用 $p$ 值检验了以上方程的每一项特征值,每一特征值对应的 $p$ 值都小于0.05。实验值与预测值之间的相关系数 $R=0.86$ ,用Jack-knife方法检验,得出 $R=0.81$ , $t=8.32$ , $p<0.001$ 。

同样的39个未分类蛋白质,选取Gromiha文章中的 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $\Delta G_{hd}$ 4个特征值作线性回归,得到方程:

$$\ln(k_f) = 21.69 - 13.7K^0 - 93.44R_a + 69.41\Delta ASA + 1.88\Delta G_{hd}. \quad (11)$$

用 $p$ 值检验了以上方程的每一项特征值,其对应的 $p$ 值都大于0.05。实验值与预测值之间的相关系数 $R=0.697$ 。用Jack-knife方法检验,得出 $R=0.48$ , $t=3.37$ , $p>0.001$ 。这说明选取的特征值 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $Mu$ 、 $El$ 对未分类蛋白质有影响。对比结果见表4。

表4 未分类蛋白质回归分析结果

Table 4 The results of unclassified proteins regression analysis

Model	数据	特征值	Jack-knife			
			$R$	$R$	$t$	$P$
Gromiha	39	$K^0, R_a, \Delta ASA, \Delta G_{hd}$	0.70	0.48	3.37	>0.001
Our	39	$K^0, R_a, \Delta ASA, Mu, El$	0.86	0.81	8.32	<0.001

通过实验计算得出,对于未分类蛋白质选取5个特征值 $K^0$ 、 $R_a$ 、 $\Delta ASA$ 、 $Mu$ 、 $El$ 计算蛋白质折叠速率

预测值,与  $\ln(k_f)$  有良好的相关性。

对于不同类别的蛋白质,其折叠速率有很大的区别。本文研究不同的特征值对不同类别的蛋白质折叠速率的影响,以及特征值对未分类蛋白质折叠速率的影响。从本文的实验结果发现,  $C_\alpha$ 、 $R_a$ 、 $LZ_c$  3个特征值对全  $\alpha$  类蛋白质折叠有一定的影响,对于全  $\beta$  类蛋白质,在 Gromiha 文章选取的 4 个特征值基础上增加了  $N_m$  特征值,使得全  $\beta$  类蛋白质的折叠速率有所提高,相关系数达到 0.93。为了说明  $N_m$  特征值对全  $\beta$  类蛋白质的折叠速率有影响,本文又选取了 13 个数据进行验证。由实验结果可知,在增加  $N_m$  特征值后全  $\beta$  类蛋白质的折叠速率确实有所提高(见表 2)。由此可见,增加的特征值是有效特征值。对于混合类蛋白质,本文选取了 3 个特征值  $K^0$ 、 $\Delta ASA$ 、 $PI$  其相关系数比 Gromiha 文章选取特征值得到的相关系数要高(见表 3)。由研究结果发现,特征值  $K^0$ 、 $\Delta ASA$  对全  $\beta$  类蛋白质和混合类蛋白质的折叠速率都有影响。选取  $K^0$ 、 $R_\alpha$ 、 $\Delta ASA$ 、 $Mu$ 、 $EIS$  个特征值对未分类的蛋白质的折叠有一定的影响。

### 3 结 论

蛋白质在生物体内占有重要的地位,是一个生物大分子,由 20 个氨基酸以肽键的形式形成肽链。肽链在空间结构中通过卷曲形成特定的空间结构,如二级结构和三级结构。氨基酸残基及周围介质之间的相互作用决定了蛋白质的结构和折叠速率。由于蛋白质折叠速率对蛋白质功能有一定的影响,近些年来,已有很多方法来预测蛋白质折叠速率。有很多研究工作者从蛋白质的二级结构和三级结构来进行预测蛋白质的折叠速率,但是由于蛋白质的二级结构和三级结构影响因子单一,结构复杂,因此越来越多的研究工作者们开始从蛋白质的一级结构来预测蛋白质的折叠速率。本文就是研究蛋白质的一级结构信息对蛋白质折叠速率的影响,运用生物统计学和生物信息学的方法,选取了蛋白质编码序列的一些特征值,通过实验验证了这些特征值对不同类别的蛋白质折叠速率的影响不同。

本文对于全  $\alpha$  类蛋白质,全  $\beta$  类蛋白质,混合类蛋白质和未分类蛋白质分别得到 4 个线性回归方程。利用这些回归方程研究了所选取的特征值与蛋白质折叠速率之间的相关性,得到了较好的结果,比 Gromiha 文章选取的特征值相关系数都有所提高。不同的数据集对结果有一定的影响,如何减少数据集对结果的影响会在后续工作中进行更深入研究。

### 参考文献(References)

- [1] GUO Jianxiu, MA Binguang, ZHANG Hongyu. Progress in protein folding rate prediction [J], *Acta Biophysica Sinica*, 2006, 22(2): 89-95.  
郭建秀,马彬广,张红雨.蛋白质折叠速率预测研究进展[J], *生物物理学报*, 2006, 22(2): 89-95.
- [2] GROMIHA M M, SELVARAJ S. Bioinformatics approaches for understanding and predicting protein folding rates [J]. *Current Bioinformatics*, 2008, 3(1): 1-9.
- [3] PLAXCO K W, SIMONS K T, BAKER D. Contact order, transition state placement and the refolding rates of single domain proteins [J]. *Journal of Molecular Biology*, 1998, 277(4): 985-944.
- [4] ZHOU H, ZHOU Y. Folding rate prediction using total contact distance [J]. *Biophysical Journal*, 2002, 829(1), 458-463.
- [5] GONG H, ISOM D G, SRINIVASAN R, et al. Local secondary structure content predicts folding rates for simple two-state proteins [J]. *J Mol Biol*, 2003, 327(5): 1149-1154.
- [6] IVANKOV D N, FINKELSTEIN A V. Prediction of protein folding rates from the amino acid sequence-predicted secondary structure [J]. *Proc Nat Acad Sci USA*, 2004, 101(24): 8942-8944.
- [7] SHAO H, PENG Y, ZENG Z H. A simple parameter relating sequences with folding rates of small helical proteins [J]. *Protein Pept Lett*, 2003, 10(3): 277-280.
- [8] GALZITSKAYA O V, GARBUZYNSKIY S O, IVANKOV D N, et al. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics [J]. *Proteins*, 2003, 51(2): 162-166.
- [9] 徐宏睿,马彬广.蛋白质折叠速率决定因素与预测方法的研究进展 [J], *生物物理学报*, 2013, 29(3): 192-202.  
XU Hongrui, MA Binguang. Progress in the study on determinants of protein folding rate and method of folding rate prediction [J]. *Acta Biophysica Sinica*, 2013, 29(3): 192-202.
- [10] MA B G, GUO J X, ZHANG H Y. Direct correlation between proteins' folding rates and their amino acid compositions: an ab initio folding rate prediction [J]. *Proteins*, 2006, 65(2): 362-372.
- [11] HUANG J T, XING D J, HUANG W. Relationship between protein folding kinetics and amino acid properties [J]. *Amino Acids*, 2012, 43: 567-572.
- [12] GROMIHAM M, THANGAKANI A M, SELVARAJ S. FOLD-RATE: prediction of protein folding rates from amino acid sequence [J]. *Nucleic Acids Res*, 2006, 34 (suppl\_2): 70-74.

- [13] HUANG L T, GROMIHA M M. Analysis and prediction of protein folding rates using quadratic response surface models[J]. *J Comput Chem*, 2008, 29(10): 1675-1683.
- [14] GOU J X, RAO N N, LIU G X, et al. Predicting protein folding rate from amino acid sequence[J]. *Prog Biochem Biophys*, 2011, 37(12): 1331-1338.
- [15] GROMIHA M M. A statistical model for predicting protein folding rates from amino acid sequence with structural class information[J]. *Chem Inf Model*, 2005, 45(2): 494-501.
- [16] 于志芬, 李瑞芳. 同义密码子的使用偏好性对蛋白质折叠速率的影响[J], *生物物理学报*, 2013, 29(8): 603-613.  
YU Zhifen, LI Ruifang. The influence of synonymous codon bias on protein folding rates[J]. *Acta Biophysica Sinica*, 2013, 29(8): 603-613.
- [17] 胡睿, 史小红, 李晋惠. 基于序列疏水值震荡的折叠速率预测[J]. *生物信息学*, 2013, 11(2): 86-89.  
HU Rui, SHI Xiaohong, LI Jinhui. Prediction of protein folding-rate based on the hydrophobic value vibration[J]. *Chinese Journal of Bioinformatics*, 2013, 11(2): 86-89.
- [18] 胡睿, 史小红, 基于残基接触数的蛋白质折叠速率预测[J]. *西安工业大学学报*, 2013, 33(2): 146-150.  
HU Rui, SHI Xiaohong. Prediction of protein folding-rate based on the residues contact number[J]. *Journal of Xi'an Technological University*, 2013, 33(2): 146-150.
- [19] BEMAN H M, WESTBROOK J, FENG Z, et al. The protein Databank[J]. *Nucleic Acids Res*, 2000, 28(1): 235-242.
- [20] GROMIHA M M, SARAI O A. Important amino acid properties for enhanced thermostability from mesospheric to hemophilic protein[J]. *Biophys Chem*, 1999, 82: 51-67.
- [21] CHOU P Y, FASMAN G D. Prediction of the secondary structure of proteins from their amino acid sequence[J]. *Adv Enzym*, 1978, 47: 45-148.
- [22] IQBAL M, VERRALL R E. Implications of protein folding. Additivity schemes for volumes and compressibilities[J]. *Biol Chem*, 1988, 263(9): 4159-4165.
- [23] GEKKO K, NOGUCHI H. Compressibility of globular proteins in water at 25 degree C[J]. *Phys Chem*, 1979, 83(21): 2706-2714.
- [24] PONNUSWAMY P K, PRABHAKARAN M, MANAVALAN P. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins[J]. *Biochim Biophys Acta*, 1980, 623(2): 301-316.
- [25] OOBATAKE M, OOI T. Hydration and heat stability effects on protein unfolding[J]. *Prog Biophys Mol Biol*, 1993, 59(3): 237-284.
- [26] 李丹, 基于蛋白质图形表示的膜蛋白跨膜区预测[D]. 杭州: 浙江理工大学, 2012.  
LI Dan. The prediction of transmembrane domains based on the graphical representation of protein sequences[D]. HANG zhou: Zhejiang Sci-Tech University, 2012.