

doi:10.3969/j.issn.1672-5565.2014.04.10

枯草芽孢杆菌同义密码子使用偏性对蛋白质折叠速率的影响

于志芬,李瑞芳*,黄 俏
(内蒙古师范大学,呼和浩特 010022)

摘要:目前,有关同义密码子使用偏性对蛋白质折叠的影响研究中,样本蛋白均来源于不同的物种。考虑到同义密码子使用偏性的物种差异性,选取枯草杆菌的核蛋白为研究对象。首先,将每条核蛋白按二级结构截取为 α 螺旋片段、 β 折叠片段和无规卷曲(α - β 混合)片段,并计算其蛋白质折叠速率。然后,整理每个片段相应的核酸序列信息,计算其同义密码子使用度。在此基础上,分析枯草芽孢杆菌核蛋白的同义密码子使用偏性与蛋白质折叠速率的相关性。发现对于不同二级结构的肽链片段,都有部分密码子的使用偏性与其对应的肽链折叠速率显著相关。进一步分析发现,与肽链片段折叠速率显著相关的密码子绝大部分为枯草杆菌全序列或核蛋白序列的每一组同义密码子中使用度最高的密码子。结果表明,在蛋白质的折叠过程中,枯草芽孢杆菌的同义密码子使用偏性起着重要作用。

关键词:枯草芽孢杆菌;同义密码子;蛋白质折叠速率;使用偏性

中图分类号:Q-3 **文献标志码:**A **文章编号:**1672-5565(2014)-04-292-08

Influence of synonymous codon usage bias on protein folding rate of bacillus subtilis

YU Zhifen, LI Ruifang*, HUANG Qiao

(College of Physics and Electronic Information, Inner Mongolia Normal University, Hohhot 010022, China)

Abstract: In the current study about the influence of synonymous codon usage bias on protein folding rate, the proteins are always derived from different species. In this article, considering the species differences of synonymous codon bias, the nucleoproteins of bacillus subtilis was selected as the samples. First, each of the nucleoprotein was cut into α -helices, β -strands and mixed-class segments according to their secondary structure. Then, the information of its corresponding mRNA sequence of each peptide segment was found, the synonymous codon usage bias and the protein folding rate of corresponding peptide segments were calculated. On this basis, the correlation between synonymous codon usage bias and the corresponding peptide folding rate of bacillus subtilis were analyzed. It was found that some of codon usage bias were significantly correlated with the protein folding rate of the peptide segments for different secondary structures. Further analysis found, most of codons significant correlated with peptide folding rate are the high RSCU value of bacillus subtilis complete sequences and ribosomal sequences. The results showed that the synonymous codon usage bias of bacillus subtilis plays an important role in protein folding.

Keywords: Bacillus subtilis; Synonymous codon; Protein folding rate; Using bias

蛋白质是由氨基酸按照一定的顺序组成的生物大分子,在生物体内担任着重要的角色^[1]。自然状态下,蛋白质通常可以从未折叠的状态快速可靠地折叠成具有三维结构的天然构象。正确的结构是功能的基础,蛋白质的错误折叠会形成无活性蛋白或

引起淀粉样纤维的聚集,引起阿尔茨海默综合症(Alzheimer's)^[2]、帕金森氏症(Parkinson's)、传染性海绵状脑病(CJD)等蛋白质折叠疾病^[3-4]。蛋白质折叠对于所有生物体系,都是最基本和最重要的过程,这个过程是一个复杂而深奥的基础理论问题,并

收稿日期:2014-06-13;修回日期:2014-10-10.

基金项目:内蒙古自治区自然科学基金(2012MS0115);内蒙古自治区研究生教育创新计划(S20141013523);内蒙古师范大学研究生科研创新基金(CXJJS14061)资助。

作者简介:于志芬,女,硕士研究生,研究方向:理论生物物理;E-mail: yuzhifensh@163.com.

* 通信作者:李瑞芳,女,博士,副教授,研究方向:理论生物物理;E-mail: liruiyang@imnu.edu.cn.

且是生命科学中最基本的问题之一,但是对于人类而言,仍是个未解之谜。蛋白质的折叠问题是分子生物学的中心法则仍未解决的一个重大的生物学问题,它被列为二十一世纪生物物理学的重要课题。

核糖核蛋白体(Ribosome,也称为核蛋白体)是一种重要的细胞器。细胞内核蛋白是蛋白质生物合成的基地。无论原核细胞(细菌)或真核细胞(动植物),凡有蛋白质合成,都一定有核糖核蛋白体的参加^[5]。它在阐明遗传密码子的编排工作中,发挥了非常重要的作用,对分子生物学的发展有着重大的贡献。

核糖核蛋白体 RNA 占核糖核蛋白体重量的三分之二,占细胞全部 RNA 的 80%。核糖核蛋白体 RNA 在核糖核蛋白体的自动装配和表现生物功能方面起着重要的作用。因而研究核糖核蛋白体 RNA 与研究核糖核蛋白体蛋白质一样,是深入认识核糖核蛋白体结构与功能的必要前提。

蛋白质折叠是指一个蛋白质分子从它的变性状态转变到其特定的具有生物活性的天然构象的过程。这一过程遵循自由能减小规律,即蛋白质在一定时间内沿着某些特定路径(过渡态系综)达到其自由能最小(极小)的天然构象^[6]。在蛋白质中,它们的折叠速率又有着很大的差异,有些蛋白质在几微秒内就能完成折叠过程,而有些则需要几个小时^[7]。目前,普遍认为决定蛋白质折叠速率的因素主要来源于构成蛋白质的氨基酸序列和各级结构以及环境和温度^[8]。此外,决定蛋白质折叠速率的信息还来自于蛋白质编码序列,它们在调节蛋白质折叠过程中起到重要作用^[9]。

同义密码子编码同一氨基酸,假如没有任何选择压力或突变偏好,每个氨基酸位点上的核苷酸突变都是随机的,同义密码子出现的概率应该是一样的^[10]。但实际上,同义密码子的使用并非随机的,

而是不同的物种或同一物种的不同基因在编码氨基酸时倾向使用某些特定的同义密码子,这种现象称为同义密码子的使用偏好性(Synonymous codon usage bias)^[11]。同义密码子的使用偏好性广泛地存在于细菌、真菌、植物、动物及人类中,而且对于不同的生物,影响密码子使用偏性的因素也不尽相同,因此,产生了多种解释密码子偏性使用的理论^[12]。在同义密码子的使用偏性对蛋白质折叠速率影响的研究中,如果选取不同物种的蛋白,就避免不了同义密码子物种的差异对研究结果的影响。由此,如果去除物种本身对同义密码子使用偏性的影响,选择单一物种作为研究对象,这样可能会得到更令人满意的结果。

基于这样的想法,以枯草芽孢杆菌的核蛋白作为研究样本,首先按二级结构将每条蛋白进行截取并分类。然后找到每个片段的核酸序列信息,计算其同义密码子使用度和相应肽链折叠速率。以此为研究基础,分析同义密码子的使用偏性与相应肽链折叠速率间的相关性,探索同义密码子的使用偏性对蛋白质折叠速率的影响。

1 材料选取

以枯草芽孢杆菌的核蛋白为研究对象,蛋白质信息及其相对应的核酸序列均取自核蛋白库(<http://ribosome.miyazaki-med.ac.jp/>)。首先,提取出枯草芽孢杆菌的核蛋白,选取八个不同的蛋白质二级结构预测软件,预测每条核蛋白并截取出 α 螺旋片段、 β 折叠片段和无规卷曲(α - β 混合)片段。然后,运用蛋白质折叠速率预测软件(<http://psfs.cbrc.jp/fold-rate>)计算各肽链片段的折叠速率,相关信息见表1。

表1 151条肽链片段的基本信息
Table 1 The set of 151 peptide segments

NO.	Name	$\ln(k_f)$	Length	Structureclass	NO.	Name	$\ln(k_f)$	Length	Structureclass
1	bsu0052	12.60	32	α	1	bsu0052	56.10	45	β
2	bsu0099	9.72	10	α	2	bsu0099	38.00	7	β
3	bsu0102	31.70	49	α	3	bsu0102	63.40	27	β
4	bsu0103	17.20	79	α	4	bsu0103	45.90	46	β
5	bsu0104	16.30	121	α	5	bsu0104	83.70	9	β
6	bsu0105	13.90	91	α	6	bsu0110	53.70	24	β
7	bsu0110	0.62	13	α	7	bsu0111	68.60	6	β
8	bsu0111	16.20	95	α	8	bsu0115	82.00	8	β
9	bsu0115	11.50	54	α	9	bsu0116	70.30	56	β
10	bsu0116	20.50	7	α	10	bsu0117	65.90	29	β
11	bsu0117	8.55	55	α	11	bsu0118	33.30	26	β
12	bsu0118	20.90	21	α	12	bsu0119	59.40	66	β
13	bsu0120	19.30	10	α	13	bsu0120	45.50	18	β
14	bsu0121	-2.79	49	α	14	bsu0121	91.30	14	β
15	bsu0122	16.50	75	α	15	bsu0122	67.60	36	β
16	bsu0123	8.67	47	α	16	bsu0123	68.40	31	β

续(表1)

NO.	Name	$\ln(k_f)$	Length	Structureclass	NO.	Name	$\ln(k_f)$	Length	Structureclass
17	bsu0124	20.60	53	α	17	bsu0125	62.50	20	β
18	bsu0125	44.80	11	α	18	bsu0126	68.00	39	β
19	bsu0126	25.40	13	α	19	bsu0127	37.10	46	β
20	bsu0128	10.40	65	α	20	bsu0128	53.80	19	β
21	bsu0130	-1.19	39	α	21	bsu0130	77.30	29	β
22	bsu0131	22.50	26	α	22	bsu0131	28.90	55	β
23	bsu0132	9.88	47	α	23	bsu0132	59.70	13	β
24	bsu0133	15.60	71	α	24	bsu0133	57.70	17	β
25	bsu0134	11.60	13	α	25	bsu0134	51.10	11	β
26	bsu0135	0.433	25	α	26	bsu0135	15.30	15	β
27	bsu0141	16.70	44	α	27	bsu0140	87.10	10	β
28	bsu0142	8.20	39	α	28	bsu0141	93.70	15	β
29	bsu0144	10.80	67	α	29	bsu0142	43.10	29	β
30	bsu0149	13.20	45	α	30	bsu0144	65.30	9	β
31	bsu0150	9.34	40	α	31	bsu0149	78.40	19	β
32	bsu0887	7.25	50	α	32	bsu0150	48.50	25	β
33	bsu1510	7.24	10	α	33	bsu1583	68.20	16	β
34	bsu1600	21.70	31	α	34	bsu1600	69.70	15	β
35	bsu1605	7.86	33	α	35	bsu1605	44.60	34	β
36	bsu1650	18.60	126	α	36	bsu1650	111.00	34	β
37	bsu1669	32.10	55	α	37	bsu1669	126.00	6	β
38	bsu2287	30.90	84	α	38	bsu2287	55.00	86	β
39	bsu2537	12.30	35	α	39	bsu2790	59.60	22	β
40	bsu2551	-8.73	63	α	40	bsu2792	53.30	37	β
41	bsu2790	10.50	12	α	41	bsu2882	30.40	4	β
42	bsu2792	17.30	11	α	42	bsu2962	62.00	9	β
43	bsu2881	14.30	88	α	43	bsu3705	26.90	10	β
44	bsu2882	4.78	28	α	44	bsu4047	108.00	10	β
45	bsu2962	17.30	81	α	45	bsu4086	77.20	6	β
46	bsu3705	37.00	7	α	46	bsu4088	97.10	12	β
47	bsu4047	8.79	63	α	47	bsu4103	82.70	4	β
48	bsu4086	1.70	28	α	27	bsu0134	27.00	19	$\alpha-\beta$
49	bsu4088	20.90	41	α	28	bsu0135	23.00	92	$\alpha-\beta$
50	bsu4103	43.90	4	α	29	bsu0140	47.40	19	$\alpha-\beta$
1	bsu0052	24.80	94	$\alpha-\beta$	30	bsu0141	37.50	56	$\alpha-\beta$
2	bsu0099	-2.86	27	$\alpha-\beta$	31	bsu0142	32.80	57	$\alpha-\beta$
3	bsu0102	20.40	52	$\alpha-\beta$	32	bsu0144	20.80	33	$\alpha-\beta$
4	bsu0103	31.90	89	$\alpha-\beta$	33	bsu0149	30.00	61	$\alpha-\beta$
5	bsu0104	16.80	24	$\alpha-\beta$	34	bsu0150	20.60	56	$\alpha-\beta$
6	bsu0105	16.00	22	$\alpha-\beta$	35	bsu0887	45.20	32	$\alpha-\beta$
7	bsu0110	60.90	91	$\alpha-\beta$	36	bsu1510	-17.50	30	$\alpha-\beta$
8	bsu0111	45.60	47	$\alpha-\beta$	37	bsu1583	43.40	36	$\alpha-\beta$
9	bsu0115	28.50	33	$\alpha-\beta$	38	bsu1600	46.70	36	$\alpha-\beta$
10	bsu0116	34.90	119	$\alpha-\beta$	39	bsu1605	14.90	41	$\alpha-\beta$
11	bsu0117	21.70	89	$\alpha-\beta$	40	bsu1650	11.10	67	$\alpha-\beta$
12	bsu0118	17.10	41	$\alpha-\beta$	41	bsu1669	52.70	21	$\alpha-\beta$
13	bsu0119	18.30	176	$\alpha-\beta$	42	bsu2287	29.10	121	$\alpha-\beta$
14	bsu0120	23.30	62	$\alpha-\beta$	43	bsu2537	27.10	8	$\alpha-\beta$
15	bsu0121	46.00	40	$\alpha-\beta$	44	bsu2551	35.60	19	$\alpha-\beta$
16	bsu0122	14.00	91	$\alpha-\beta$	45	bsu2790	32.60	38	$\alpha-\beta$
17	bsu0123	23.80	56	$\alpha-\beta$	46	bsu2792	46.80	40	$\alpha-\beta$
18	bsu0124	7.26	7	$\alpha-\beta$	47	bsu2881	21.60	19	$\alpha-\beta$
19	bsu0125	22.70	39	$\alpha-\beta$	48	bsu2882	28.90	23	$\alpha-\beta$
20	bsu0126	39.00	55	$\alpha-\beta$	49	bsu2962	37.20	81	$\alpha-\beta$
21	bsu0127	25.60	53	$\alpha-\beta$	50	bsu3705	-18.90	46	$\alpha-\beta$
22	bsu0128	12.50	76	$\alpha-\beta$	51	bsu4047	16.20	52	$\alpha-\beta$
23	bsu0130	25.10	54	$\alpha-\beta$	52	bsu4086	18.20	39	$\alpha-\beta$
24	bsu0131	31.80	74	$\alpha-\beta$	53	bsu4088	8.64	31	$\alpha-\beta$
25	bsu0132	28.20	49	$\alpha-\beta$	54	bsu4103	63.40	31	$\alpha-\beta$
26	bsu0133	23.30	54	$\alpha-\beta$					

2 方法

2.1 肽链片段二级结构的获得

以往研究表明,不同二级结构类蛋白质折叠速率与不同密码子的使用偏性有关。而目前准确率高的蛋白质二级结构预测软件很多。本文为了准确分类,采用八种不同的二级结构预测软件(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html)。运用八种不同的蛋白质二级结构预测软件同时对每一条蛋白进行预测。对于预测结果,本文截取了至少有三种软件预测结果相同的二级结构片段,然后将每一条肽链截取出的片段按不同二级结构分别连接在一起。例如,核蛋白 *bsu0052*,用蛋白质二级结构预测软件分析,截取后得到四个 α 类肽链片段,将这四个片段连接得到一个长度为 32 的 α 类蛋白(命名为 *bsu0052 α* 类蛋白)。同样地,我们分别得到了长度为 45 的 β 类蛋白以及长度为 94 的 α - β 混合类蛋白,分类保存作为研究片段。我们将全部核蛋白按照这样的方法整理分为 α 螺旋、 β 折叠和无规卷曲(α - β 混合)三类不同肽链折叠片段作为研究样本。

2.2 蛋白质折叠速率

为了计算出每一个肽链的蛋白质折叠速率,我们尝试了许多方法,发现 Gromiha 等提出的基于一级序列直接预测蛋白质折叠速率的方法。Gromiha 等认为蛋白质的折叠速率是由残基间相互作用决定的,而相互作用又受氨基酸的物理、化学等性质影响,提出根据氨基酸属性来预测蛋白质折叠速率^[13]。氨基酸的平均属性用 Pare 来表示,计算公式如下:

$$Pare = \frac{1}{N} \sum_{j=1}^N P(j) \quad (1)$$

式中, N 是氨基酸序列的残基数, $p(j)$ 是氨基酸序列中第 j 个残基的属性。

对于 α 螺旋类片段,Gromiha 采用以下线性回归公式计算:

$$\ln(k_f) = -33.191\alpha_c + 20.195 \quad (2)$$

式中, $\ln(k_f)$ 是蛋白质折叠速率的自然对数, α_c ^[14-15] 是 α 螺旋 C 端强度(The power to be at the C-terminal)。

对于 β 折叠类片段,线性回归公式为:

$$\ln(k_f) = -81.48K^0 + 163.08P_\beta + 79.92R_\alpha - 134.99\Delta ASA - 13.18 \quad (3)$$

式中, K^0 为残基的可压缩性^[16-17], P_β 为 β 折叠趋势^[14], R_α 和 ΔASA 是和溶剂溶解有关的参量。

对于无规卷曲(α - β 混合)类片段,线性回归公式是:

$$\ln(k_f) = -102.60K^0 - 90.33R_\alpha + 131.80\Delta ASA - 90.82\Delta GhD - 38.53 \quad (4)$$

式中, ΔGhD 是对于变性蛋白的水合作用的吉布斯自由能变化量^[18],其它参量与公式(3)相同。

用以上公式对枯草杆菌全部核蛋白中统计出的 53 个 α 螺旋片段、47 个 β 折叠片段和 52 个无规卷曲(α - β 混合)片段计算它们的蛋白质折叠速率,结果详见表 1。

2.3 密码子使用度

本文用密码子使用度(Relative synonymous codon usage, *RSCU*)来描述密码子的偏性,它的计算方法如公式(1)。由于该指标计算简单,且除去了氨基酸组成对密码子使用的影响,所以能比较直观地反映出同义密码子的使用偏好。

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n} \sum_{i=1}^n X_{ij}} \quad (5)$$

式中, $RSCU_{ij}$ 是编码第 i 个氨基酸的第 j 个密码子的密码子使用度, x_{ij} 是编码第 i 个氨基酸的第 j 个密码子的出现次数, n_i 是编码第 i 个氨基酸的同义密码子的数量(值为 1~6)。本文运用了 Codon W 软件对密码子使用进行了分析,该软件可以对密码子使用进行多变量分析,还可以对密码子使用与氨基酸的相关性进行分析。通过使用该软件,我们得到了 53 个 α 螺旋片段、47 个 β 折叠片段和 52 个无规卷曲(α - β 混合)片段的密码子使用度。

3 结果分析

3.1 α 螺旋片段相关性分析

对枯草芽孢杆菌的 54 条核蛋白按蛋白质二级结构分类后得到了 53 个 α 螺旋片段,计算了每一条肽链的折叠速率,并整理了每条肽链相应的核酸序列,分析了它们的密码子使用度。对 α 螺旋片段的密码子使用度与肽链折叠速率进行相关性分析,结果如表 2。

表2 53个 α 螺旋片段相应编码序列RSCU值与其肽链折叠速率之间的线性回归分析结果

Table 2 Linear regression analysis between the peptide segment folding rates and the RSCU values for 53 α -helices

	Condon	R	P
Ala	GCA	-0.33	0.02
	GCG	-0.20	0.17
	GCT	-0.07	0.62
	GCC	-0.16	0.28
	AGA	-0.34	0.02
Arg	CGC	-0.39	0.01
	AGG	-0.09	0.51
	CGG	-0.09	0.55
	CGA	-0.14	0.35
	CGT	-0.15	0.31
Ser	TCT	-0.29	0.04
	AGT	0.12	0.42
	AGC	0.01	0.97
	TCG	-0.05	0.75
	TCA	-0.07	0.62
Gln	TCC	-0.17	0.23
	CAA	-0.34	0.02
	CAG	-0.17	0.24

对 α 螺旋片段的密码子使用度与其肽链折叠速率进行相关性分析,由表2的结果中我们看到了一些相关性,对于50个 α 螺旋片段蛋白,五个密码子的使用与蛋白质折叠速率显著相关。共有四种氨基酸,分别为丙氨酸、精氨酸、丝氨酸和谷氨酰胺。表中R值有正负之分,若R值为正,则这个密码子的使用促进蛋白质的折叠;若R值为负,那么这个密码子的使用阻碍蛋白质的折叠。例如,密码子GCA(编码丙氨酸Ala)的使用与肽链的折叠速率呈负相关,即这个密码子的使用阻碍蛋白质的折叠。

3.2 β 折叠片段相关性分析

按照 α 螺旋片段整理方法,对枯草芽孢杆菌的54条核蛋白按蛋白质二级结构分类后得到了47个 β 折叠片段,对肽链片段相应核酸序列,分析其密码子使用度与肽链折叠速率的相关性,结果见表3。

对 β 折叠片段密码子使用度与肽链折叠速率进行相关性分析,共有47个 β 折叠片段蛋白,四个密码子的使用与肽链折叠速率显著相关。共三种氨基酸分别为谷氨酸、苏氨酸和缬氨酸。表中,密码子GTT(编码缬氨酸Val)的使用与肽链的折叠速率呈正相关,即这个密码子的使用促进蛋白质的折叠。

表3 47个 β 折叠片段相应编码序列RSCU值与其肽链折叠速率之间的线性回归分析结果

Table 3 Linear regression analysis between the peptide segment folding rates and the RSCU values for 47 β -strands

	Condon	R	P
Glu	GAA	-0.30	0.04
	GAG	-0.19	0.20
	GTA	-0.33	0.02
Val	GTT	0.33	0.03
	GTG	-0.02	0.89
	GTC	-0.17	0.27
Thr	ACT	-0.36	0.01
	ACC	-0.05	0.72
	ACG	0.06	0.69
	ACA	-0.10	0.53

3.3 α - β 混合片段相关性分析

同样地,对枯草芽孢杆菌的54条核蛋白按蛋白质二级结构分类后得到了52个无规卷曲(α - β 混合)片段,分析这些片段相应的核酸序列的密码子使用度与肽链折叠速率的相关性,见表4。

表4 52个混合类片段相应编码序列RSCU值与其肽链折叠速率之间的线性回归分析结果

Table 4 Linear regression analysis between the peptide segment folding rates and the RSCU values for 52 mixed-class segments

	Condon	R	P
Val	GTT	0.30	0.03
	GTG	0.15	0.29
	GTA	-0.25	0.06
	GTC	-0.05	0.74
Arg	AGG	0.32	0.02
	AGA	0.02	0.90
	CGC	0.07	0.60
	CGG	1.64×10^{-8}	1.00
Asn	CGA	0.23	0.10
	CGT	-0.02	0.89
	AAT	-0.29	0.03
	AAC	0.21	0.13
Cys	TGC	-0.35	0.01
	TGT	-0.22	0.10
Leu	CTT	-0.30	0.03
	TTG	0.03	0.81
	TTA	0.15	0.29
	CTG	0.10	0.49
	CTA	-0.07	0.63
	CTC	0.001	1.00

对 α - β 混合片段核酸序列的密码子使用度与肽链片段折叠速率进行相关性分析,共有54个混合类片段蛋白,有五个密码子的使用与蛋白质折叠速率显著相关。例如,密码子GTT(编码缬氨酸Val)的使用与蛋白质的折叠速率呈正相关,促进蛋白质的折叠;而密码子CTT(编码亮氨酸Leu)的使用与肽链的折叠速率呈负相关,阻碍蛋白质的折叠。

3.4 全序列和核蛋白序列的密码子使用度与蛋白质不同二级结构比较分析

为了更深入分析,本文统计了枯草杆菌全序列与核蛋白序列的密码子使用度(见表5),表5中列出了枯草芽孢杆菌的全序列和核蛋白序列的64组RSCU值。然后,将三类研究片段的密码子使用度与其肽链折叠速率之间的相关性较好的密码子(即P值小于0.05的密码子)列在表6中。将相关性较好的这些密码子与它们在枯草芽孢杆菌全序列及核蛋白序列的RSCU值进行比较。

表5 枯草杆菌全序列与核蛋白序列的RSCU值

Table 5 The RSCU values of *Bacillus subtilis* complete sequence and ribosomal protein sequence

	Codon	Complete	Rib		Codon	Complete	Rib	
	TTG	1.32	0.72		CCG	1.20	0.56	
	TTA	1.08	1.50		CCA	1.04	1.60	
Leu	CTG	1.08	0.42	Pro	CCT	1.04	1.76	
	CTA	0.42	0.48		CCC	0.72	0.08	
	CTT	1.38	2.76		GGG	0.68	0.08	
	CTC	0.72	0.18		Gly	GGA	1.28	1.28
	AGG	0.90	0.06			GGT	0.84	1.80
Arg	AGA	1.26	0.54		GGC	1.20	0.80	
	CGG	1.08	0.00	Glu	GAG	0.64	0.46	
	CGA	0.96	0.06		GAA	1.36	1.54	
	CCT	0.84	3.48	GAT	1.30	1.06		
	CGC	0.96	1.80	Asp	GAC	0.70	0.94	
AGT	0.60	0.48	AAG		0.70	0.36		
Ser	AGC	1.20	0.90	Lys	AAA	1.30	1.64	
	TCG	0.78	0.12		AAT	1.22	0.54	
	TCA	1.44	1.20	Asn	AAC	0.78	1.46	
	TCT	1.08	3.00		Cys	TGT	1.00	0.62
	TCC	0.96	0.30	Cys	TGC	1.00	1.38	
Ala	GCG	0.80	0.56	His	CAT	1.32	0.88	
	GCA	1.08	1.08		CAC	0.68	1.12	
	GCT	1.20	2.20	Tyr	TAT	1.32	0.50	
	GCC	0.88	0.16		TAC	0.68	1.50	
	GTG	0.84	0.44	ATA	0.90	0.03		
Val	GTA	0.84	1.32	Ile	ATT	1.17	1.23	
	GTT	1.36	1.92		ATC	0.96	1.74	
	GTC	0.96	0.32	TGA	1.44	0.00		
Thr	ACG	0.92	0.60	End	TAG	0.45	0.21	
	ACA	1.44	1.40		TAA	1.11	2.79	
	ACT	0.84	1.92	Gln	CAG	0.90	0.44	
	ACC	0.80	0.08		CAA	1.10	1.56	
	Phe	TTT	1.24	0.62	Trp	TGG	1.00	1.00
TTC		0.76	1.38	Met	ATG	1.00	1.00	

表6 三类二级结构片段相应编码序列RSCU值与其肽链折叠速率之间的线性回归分析结果

Table 6 Linear regression analysis between the protein folding rates and the RSCU values for all

	Condon	R	Complete	Ribosomal	Structure class
Ala	GCA	-0.33	1.08	1.08	α
Arg	AGA	-0.34	1.26	0.54	α
Ser	TCT	-0.30	1.08	3.00	α
Arg	CGC	-0.39	0.96	1.80	α
Gln	CAA	-0.34	1.10	0.56	α
Glu	GAA	-0.30	1.36	1.54	β
Val	GTA	-0.33	0.84	1.32	β
Val	GTT	0.33	1.36	1.92	β
Thr	ACT	-0.36	0.84	1.92	β
Val	GTT	0.30	1.36	1.92	α - β
Arg	AGG	0.32	0.90	0.06	α - β
Asn	AAT	-0.29	1.22	0.54	α - β
Cys	TGC	-0.35	1.00	1.38	α - β
Leu	CTT	-0.30	1.38	2.76	α - β

表6为三类二级结构片段肽链折叠速率与相应编码序列密码子使用度相关性分析结果,其中列出的密码子全部为显著相关的密码子,以及这些密码子对应的全序列和核蛋白序列的RSCU值。比较分析这些密码子的数据发现,共十四个相关性较好的密码子,有八个密码子在枯草杆菌全序列的同义密码子中使用度最高,同样也有八个密码子在核蛋白序列的同义密码子中使用度最高。

4 讨论

在翻译过程中,在一维序列的水平上,mRNA编码区以一个三联体密码子编码一个氨基酸的方式来指导合成初生肽链的一维序列。蛋白质的折叠速率受翻译起始速率和肽链延伸速率的影响,而翻译起始速率是由核糖体和mRNA相互结合的速率决定的^[19]。同义密码子的使用会影响到mRNA的结构,而结构必定会影响到mRNA相互结合的速率,从而影响到蛋白质的折叠速率^[20-24]。这样看来,密码子的使用偏性与蛋白质折叠速率应该存在某种普遍规律。

在结果中,发现了一些关于同义密码子使用偏性与肽链折叠速率之间的相关性,而且有些同义密码子对肽链折叠速率的影响是相反的,例如,在 β 折叠片段中,GTA(编码Val)的使用阻碍蛋白质的折

叠,相关系数为-0.33,而 GTT(编码 Val)的使用则是促进蛋白质的折叠,相关系数为 0.33。而分析我们得到的与肽链折叠速率显著相关的密码子,发现它们在枯草杆菌全序列和核蛋白序列的同义密码子使用度中的值也是相对较高的。

之前我们研究了一些蛋白质的同义密码子使用偏性与蛋白质折叠速率的关系^[25]。但是样本为来自不同物种的数据,本文滤去物种本身对同义密码子使用偏性的影响,针对枯草芽孢杆菌的核蛋白进行研究。我们发现,枯草芽孢杆菌同义密码子使用度与蛋白质折叠速率显著相关的密码子有十四个,并且这些相关性较好的密码子绝大部分为枯草杆菌全序列或核蛋白序列每一组同义密码子中使用度最高的密码子。本文存在的不足之处是所使用的蛋白质折叠速率数据都是计算得到的,这是由于目前我们可以搜集整理到的有折叠速率实验数据的蛋白质仅有一百多个,而且这些数据来自不同物种^[26],未来我们希望得到更多的折叠速率实验数据,进而详细分析同义密码子的使用偏性对蛋白质折叠速率的影响。

参考文献(References)

- [1] 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 Biophys Sin, 2013, 29(3): 192-202.
- [2] TAUBES G. Misfolding the way to disease[J]. Science, 1996, 271(5255): 1493-1495.
- [3] 王明,李学周,符兆英. 蛋白质错误折叠与蛋白质构象病[J]. 延安大学学报(医学科学版), 2009, 7(2): 12-13, 16.
WANG Ming, LI Xuezhou, FU Zhaoying. Protein misfolding and conformational disease[J]. J Yanan Univ (Med Sci), 2009, 7(2): 12-13, 16.
- [4] LIN M M, ZEWAİL A H. Protein folding-simplicity in complexity[J]. Annalen der Physik, 2012, 524(8): 379-391.
- [5] 刘望夷,秦士良. 核糖核蛋白体的结构与功能[J]. 生物学通报, 1982, (3): 31-35.
LIU Wangyi, QIN Shiliang. The structure and function of ribosome[J]. Bulletin of Biology, 1982, (3): 31-35.
- [6] HUANG J T, XING D J, HUANG W. Relationship between protein folding kinetics and amino acid properties[J]. Amino Acids, 2012, 43(2): 567-572.
- [7] 李瑞芳,李宏. 蛋白质折叠速率与其氨基酸序列极性的关系[J]. 生物信息学, 2009, 7(4): 299-291.
LI Ruifang, LI Hong. The relationship between the protein folding rate and its average polarity[J]. Chinese Journal of Bioinformatics, 2009, 7(4): 299-291.
- [8] LI R F, LI H. The influence of protein coding sequences on protein folding rates of all- β proteins[J]. General Physiology and Biophysics, 2011, 30(2): 154-161.
- [9] LI R F, LI H. Study on the influences of palindromes in mRNA sequences on the folding rates of nascent peptide chains[J]. Protein Pept Lett, 2010, 17(7): 881-888.
- [10] 刘汉梅,何瑞,张怀渝,等. 拟南芥和水稻转录因子 WRKY 同义密码子的偏好性分析[J]. 四川农业大学学报, 2010, 18(3): 456-461.
LIU Hanmei, HE Rui, ZHANG Huaiyu, et al. Analysis of wrky transcriptional factors on synonymous codon bias in arabidopsis and rice[J]. Journal of Sichuan Agricultural University, 2010, 28(1): 20-27.
- [11] 刘汉梅,何瑞. 玉米同义密码子偏爱性分析[J]. 农业生物技术学报, 2010, 18(3): 456-461.
LIU Hanmei, HE Rui. Analysis of synonymous codon bias in maize[J]. Journal of Agricultural Biotechnology, 2010, 18(3): 456-461.
- [12] IKEMURA T. Codon usage and tRNA content in unicellular and multicellular organisms[J]. Mol Bio Evol, 1985, 2(1): 13-34.
- [13] GROMIHA M M. A statistical model for predicting protein folding rates from amino acid sequence with structural class information[J]. J Chem Inf Model, 2005, 45(2): 494-501.
- [14] GROMIHA M M, OOBATAKE M, SARIAI A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins[J]. Biophys Chem, 1999, 82(1): 51-67.
- [15] CHOU P Y, FASMAN G D. Prediction of the secondary structure of proteins from their amino acid sequence[J]. Adv Enzym, 1978, 47: 145-148.
- [16] IQBAL M, VERRALL R E. Implications of protein folding. Additivity schemes for volumes and compressibilities[J]. Biol Chem, 1988, 263(9): 4159-4165.
- [17] GEKKO K, NOGUCHI H. Compressibility of globular proteins in water at 25 degrees C[J]. J Phys Chem, 1979, 83(21): 2706-2714.
- [18] PONNUSWAMY P K, PRABHAKARAN M, MANAVANLAN P. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins[J]. Biochim Biophys Acta, 1980, 623(2): 301-316.
- [19] VARENNE S, BUC J, LLOUBES R, et al. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains[J]. J Mol Biol, 1984, 180(3): 549-576.
- [20] BALDWIN R L. Intermediates in protein folding reactions

- and the mechanism of protein folding[J]. *Annu Rev Biochem*, 1975, 44(3): 453-475.
- [21] KUDLICK W, ODOM O W, KRAMER G, et al. Chaperone-dependent folding and activation of ribosome-bound nascent rhodanese[J]. *Analysis by fluorescence. J Mol Biol*, 1994, 244(3): 319-331.
- [22] HARDESTRY B, TSALKOVA T, KRAMER G. Co-translational folding[J]. *Curr Opin Struct Biol*, 1999, 9(1): 111-114.
- [23] THANARAJ T A, ARGOS P. Ribosome-mediated translational pause and protein domain organization[J]. *Protein Sci*, 1996, 5(8): 1594-1612.
- [24] THANARAJ T A, ARGOS P. Protein secondary structural types are differentially coded on messenger RNA[J]. *Protein Sci*, 1996, 5(10): 1973-1983.
- [25] 于志芬, 李瑞芳. 同义密码子的使用偏性对蛋白质折叠速率的影响[J]. *生物物理学报*, 2013, 8(29): 603-613.
- YU Zhifen, LI Ruifang. The influence of synonymous codon bias on protein folding rates[J]. *Acta Biophys Sin*, 2013, 8(29): 603-613.
- [26] 董蕊, 胡晓红, 吕军. 蛋白质折叠速率数据集的构建及分析[J]. *生物物理学报*, 2012, 28(6): 509-519.
- DONG Rui, HU Xiaohong, LÜ Jun. The constitution and analysis of the protein folding rate dataset[J]. *Acta Biophys Sin*, 2012, 28(6): 509-519.