

嗜热与嗜常温微生物的蛋白质 氨基酸组成比较

卢柏松 王国力 黄培堂

(军事医学科学院生物工程研究所 北京 100071)

摘 要 嗜热微生物的嗜热特性与其蛋白质的高度热稳定性紧密相关。为了探索嗜热蛋白质的热稳定机制,比较嗜热和嗜常温微生物的蛋白质在氨基酸组成上的差别,收集 110 对分别来自嗜热和嗜常温微生物的同源蛋白质序列,比较两组蛋白质各种氨基酸含量以及疏水性氨基酸组成、疏水性指数和荷电氨基酸组成的差别,结果两者在多种氨基酸含量上存在微小但统计学上显著的差别,嗜热蛋白质比嗜常温蛋白质具有较高的平均疏水性和荷电氨基酸组成。对两组蛋白质的“脂肪族氨基酸指数”进行分析,证明嗜热蛋白质之所以具有较高的脂肪族氨基酸指数是由于其亮氨酸含量较高,与影响该指数的其它几种氨基酸无关,从而认为该指数的意义值得怀疑。通过对大量同源嗜热蛋白质和嗜常温蛋白质氨基酸组成的比较,能够揭示一些有关蛋白质热稳定性的普遍规律。

关键词 嗜热蛋白质,疏水性,荷电氨基酸含量,脂肪族氨基酸指数

分类号 Q936

嗜热微生物(thermophilic microbes)的蛋白质具有较高的热稳定性,对少数已知三维结构的同源嗜热和嗜常温蛋白质的比较发现嗜热蛋白质的热稳定性是其一级结构上多处微小作用的综合结果^[1]。由于已知三维结构的蛋白质数目有限,许多作者直接比较同源嗜热蛋白质与嗜常温蛋白质的氨基酸组成,试图揭示嗜热蛋白质的热稳定性机制。但常常由于比较的序列资料有限,缺乏足够的说服力,而且不同作者有时得到相互矛盾的结论^[2]。分子生物学序列资料的迅速积累使收集并分析大量资料成为可能,为了系统地比较嗜热和嗜常温微生物蛋白质氨基酸组成的差别,我们收集 110 对分别来自嗜热和嗜常温微生物的同源蛋白质序列,比较两组蛋白质在单个氨基酸含量上存在的差别,结果表明嗜热微生物与嗜常温生物蛋白质氨基酸的组成除了文献报道的精氨酸含量不同^[3],其它多种氨基酸含量也存在差别:即嗜热微生物的蛋白质中有利于提高蛋白质热稳定性的氨基酸含量较高,而不利于其热稳定性的氨基酸含量较低,这反映了嗜热微生物长期适应高温环境的结果。

通过比较两组蛋白质荷电氨基酸组成、疏水性氨基酸组成及疏水性指数^[4],进一步分析嗜热蛋白质在这些指标上的一些特点,揭示其与嗜常温蛋白质存在差别:嗜热蛋白质的这些指标都显著高于嗜常温蛋白质,与嗜热蛋白质具有较高的热稳定性相吻合。Ikai 曾于

1981 年提出了“脂肪族氨基酸指数”^[5]的概念,认为该指数可以作为标定蛋白质热稳定性的一个指标,即脂肪族氨基酸指数高者蛋白质热稳定性高。在本资料中,嗜热蛋白质和嗜常温蛋白质的脂肪族氨基酸指数确实存在差别,但存在差别的原因在于嗜热蛋白质中 Leu 含量较高,与其它几种脂肪族氨基酸关系不大,因而认为“脂肪族氨基酸指数”的意义值得商榷。

1 材料和方法

1.1 嗜热和嗜常温蛋白质序列资料的收集

所有蛋白质序列来自分子生物学序列数据库“Entrez (Release 15.0)”,收集方法是用“thermophil”以截词方式检索该数据库的蛋白质序列库,从中选择最适生长温度在 55℃ 以上的嗜热微生物的蛋白质序列,排除重复、不完整和氨基酸数小于 100 的记录,同一功能的蛋白质只收集一个物种的序列,由此共得到 110 条来自嗜热微生物的蛋白质序列。同时利用“Entrez”数据库提供的“Neighboring”功能检索与每个嗜热蛋白质相关的序列,从中选择与之相似性最大的来自嗜常温生物(最适生长温度 < 40℃)的蛋白质序列作为其对应同源序列。物种的最适生长温度从序列相关文献或《伯杰氏鉴定细菌学手册》(第九版)获得。两组序列资料基本情况如表 1 (序列名、物种名和各物种最适生长温度略)。

表1 两组序列资料概况

Table 1 Summary of the two groups of sequences

	序列数 Sequences number	氨基酸数 Amino acid number	涉及物种数 Species number	最适生长温度 Optimum growth temperature(℃)
嗜热组 Thermophile	110	40948	22	55~95
嗜常温组 Mesophile	110	41469	44	<40

1.2 蛋白质氨基酸组成及其它指标的比较

在计算两组序列中每个蛋白质 20 种氨基酸组成的基础上,分别计算每个序列的荷电氨基酸组成 (Arg% + Lys% + Glu% + Asp%)、疏水性氨基酸组成 (Leu% + Ile% + Val% + Phe% + Met% + Ala%)、疏水性指数和脂肪族氨基酸指数。组氨酸在生理条件下通常不带电荷,故计算荷电氨基酸含量时不包括在内。疏水性指数^[4]的算法为 $H_i = (1.8 \times \text{Ala}\% + 4.2 \times \text{Val}\% + 3.8 \times \text{Leu}\% + 4.5 \times \text{Ile}\% + 2.8 \times \text{Phe}\% + 2.5 \times \text{Cys}\% + 1.9 \times \text{Met}\%) / n$,其中氨基酸前面的系数为相应氨基酸的疏水性系数^[6],n 为序列中氨基酸的个数。脂肪族氨基酸指数的算法为: $A_i = \{\text{Ala}\% + 2.9 \times \text{Val}\% + 3.9 \times (\text{Ile}\% + \text{Leu}\%)\} \times 100$, $i\%$ 即为氨基酸 i 在蛋白质序列中的百分含量。比较方法是分别计算两组蛋白中每个序列的上述各参数的值,然后用配对资料的 t 检验分析两组蛋白质上述指标在统计学上是否存在差别。所有运算利用自行编制的 Fortran 语言程序在 PC 机上进行。

2 结果

2.1 嗜热和嗜常温蛋白质 20 种氨基酸含量的比较

分别计算嗜热和嗜常温蛋白组 110 个序列中 20 种氨基酸的百分含量,用配对资料的 t

检验方法比较 20 种氨基酸的含量在嗜热蛋白组和嗜常温蛋白组中的差别,结果嗜热蛋白组 Leu、Pro、Glu 和 Arg 的含量显著高于嗜常温组($p < 0.0005$),而 Cys、Ser、Thr、Asn、Gln 和 Asp 的含量显著低于嗜常温组($p < 0.0005$)。这一统计结果与蛋白质化学理论研究和实验研究结果十分吻合,实验发现 Pro、Arg 等氨基酸对提高蛋白质的热稳定性有利^[7,8],而 Cys、Asp、Asn 和 Gln 等氨基酸在高温下不稳定,容易引起蛋白质的变性或断裂^[9,10]。嗜热蛋白质中有利于蛋白质热稳定性的氨基酸含量较高,而不利于蛋白质热稳定性的氨基酸含量较低,在一定程度上揭示了嗜热蛋白质在氨基酸组成水平的热稳定机制。

2.2 疏水性的比较

比较两组蛋白质的疏水性氨基酸组成(表 2)可见,嗜热蛋白质比嗜常温蛋白质具有显著较高的疏水性氨基酸组成($p < 0.0005$)。考虑到不同氨基酸具有不同的疏水性,同时比较了两组蛋白质的疏水性指数 Hi,结果嗜热蛋白组的平均疏水性指数显著高于嗜常温蛋白组($p < 0.0005$)。疏水性作用是稳定蛋白质构象的最重要因素,实验表明蛋白质内核的疏水性与其热稳定性直接相关,内部疏水区的疏水性增加则热稳定性提高,反之则降低^[11]。嗜热蛋白质比嗜常温蛋白质具有较高的疏水性氨基酸组成和疏水性指数,与其具有较高的热稳定性吻合,说明增加蛋白质的平均疏水性是嗜热微生物适应高温环境的策略之一。

表2 嗜热蛋白质和嗜常温蛋白质疏水性的比较

Table 2 Comparison of average hydropathy of thermophilic and mesophilic proteins

	m±s		m _d ±s _d	p值 p-value
	嗜热组 Thermophilic	嗜常温组 Mesophile		
疏水性氨基酸组成 Per. hydrophobic a.a	0.3895±0.0411	0.3791±0.0397	0.0104±0.0252	p<0.0005
疏水性指数(Hi) Hydropathy index(Hi)	1.1737±0.0750	1.1132±0.0745	0.0604±0.0646	p<0.0005

m表示平均值, s表示标准差, m_d表示同源嗜热和嗜常温蛋白质相应指标差值的平均值, s_d表示差值的标准差。
m, Mean value; s, Standard deviation; m_d, Mean of the differences of an index between homologous thermophilic and mesophilic proteins; s_d, Standard deviation of the differences.

2.3 荷电氨基酸组成的比较

比较两组蛋白质中四种生理条件下带电荷的氨基酸含量(表 3)可见,嗜热蛋白质荷电氨基酸的平均含量为 28.19%,而嗜常温蛋白质的含量为 26.7%,前者显著高于后者($p < 0.0005$)。本组序列平均每个蛋白质含 374 个氨基酸,这样平均每个嗜热蛋白质比其同源嗜常温蛋白质多 5~6 个荷电氨基酸,这意味着每个嗜热蛋白质可以多形成 2~3 个离子键。研究表明静电作用也是稳定蛋白质构象的因素之一^[12],嗜热铁氧还蛋白有较高的热稳定性就是由于该蛋白比其同源嗜常温蛋白质多形成了两个盐键^[13]。嗜热蛋白质总体上比嗜常温蛋白质具有较高的荷电氨基酸组成这一统计结果与上述结论是一致的,表明静电作用可能是嗜热微生物采取的提高其蛋白质热稳定性的策略之一。

表3 嗜热蛋白质和嗜常温蛋白质荷电氨基酸组成的比较

Table 1 Comparison of charged amino acids percentage of thermophilic and mesophilic proteins

	m±s		m _d ±s _d	p值 p-value
	嗜热组	嗜常温组		
	Thermophile	Mesophile		
荷电氨基酸百分比	0.2819±0.0465	0.2670±0.0464	0.0149±0.0280	p<0.0005
Per. charged a.a				

m、s、m_d和s_d的意义见表2注。
See note for table 2 for means of m, s, m_d and s_d.

2.4 脂肪族氨基酸指数的比较

对两组蛋白质的脂肪族氨基酸指数(Ai)进行了比较,结果如 Ikai^[5]报道的那样,嗜热蛋白质的 Ai 值显著高于嗜常温蛋白质(表 4)。Ikai 认为丙氨酸(Ala)、缬氨酸(Val)、异亮氨酸(Ile)和亮氨酸(Leu)四种脂肪族氨基酸对蛋白质的热稳定性具有较大的贡献,但 Ala、Val 和 Ile 三种氨基酸在本组资料中不存在差别,计算 Ai 值时如果不考虑 Leu 则两组蛋白质也不存在差别(表 4 中的 Ai*)。两组蛋白质的 Ai 值之所以有差别是因为嗜热蛋白质比嗜常温蛋白质具有较高的 Leu 含量。

表4 两组蛋白质脂肪族氨基酸含量及其脂肪族氨基酸指数(Ai)的比较

Table 4 Comparison of aliphatic amino acid composition and aliphatic index between thermophilic and mesophilic proteins

	m±s		m _d ±s _d	p值 p-value
	嗜热组	嗜常温组		
	Thermophile	Mesophile		
丙氨酸Ala	0.0907±0.0304	0.0890±0.0286	0.0028±0.0290	p>0.10
缬氨酸Val	0.0807±0.0248	0.0783±0.0198	0.0023±0.0181	p>0.05
异亮氨酸Ile	0.0633±0.0203	0.0612±0.0257	-0.0026±0.0244	p>0.10
亮氨酸Leu	0.0929±0.0265	0.0846±0.0209	0.00963±0.0226	p<0.0005
脂肪族氨基酸指数	93.79±11.96	90.12±11.58	3.67±7.60	p<0.0005
Aliphatic index(Ai)				
不含亮氨酸的指数(Ai*)	57.23±12.65	57.38±10.01	-0.15±8.7	p>0.25
Ai excluding Leu(Ai*)				

m、s、m_d和s_d的意义见表2注。
See note for table 2 for means of m, s, m_d and s_d.

3 讨论

比较嗜热蛋白质和嗜常温蛋白质 20 种氨基酸组成、疏水性氨基酸组成、疏水性指数以及荷电氨基酸组成上的差别,表明嗜热蛋白质与嗜常温蛋白质在氨基酸组成上有不同的特点,且嗜热蛋白质比嗜常温蛋白质具有较高的荷电氨基酸组成和平均疏水性。蛋白质突变研究和基于蛋白质三维结构的研究表明,Pro 能够增加蛋白质变性时的熵值,Leu 具有较强的疏水性和较大的侧链,Arg 和 Glu 分别比带同样电荷的氨基酸具有更大的侧

链,它们都有利于提高蛋白质的热稳定性^[3,7,8],相反,Cys、Ser、Thr、Asn、Gln和Asp则由于在高温下不稳定或者侧链太小而对蛋白质的热稳定性有不利影响^[8~10,14]。嗜热蛋白质氨基酸组成上的特点以及嗜热蛋白质具有较高的荷电氨基酸组成和平均疏水性都与嗜热蛋白质的热稳定性吻合,这一方面揭示了嗜热蛋白质氨基酸组成水平的热稳定机理,另一方面也表明通过对大量蛋白质序列资料的比较可以发现嗜热蛋白质热稳定性机制的一些普遍规律。

Ikai认为“脂肪族氨基酸指数”可以作为评价蛋白质热稳定性的一个指标,脂肪族氨基酸指数高者蛋白质的热稳定性高。比较110对同源嗜热和嗜常温蛋白质序列表明,嗜热蛋白质的脂肪族氨基酸指数确实高于嗜常温蛋白质,但这完全是Leu的贡献,如果脂肪族氨基酸指数中不包括Leu,则两组蛋白不存在差别。构成脂肪族氨基酸指数的Ala、Val和Ile等脂肪族氨基酸对蛋白质的热稳定性是否具有很大的贡献,值得进一步探讨,因而Ai指数的意义也存在疑问。

本工作通过对大量同源嗜热蛋白质和嗜常温蛋白质序列的比较揭示两者在氨基酸组成、荷电氨基酸含量和疏水性上存在差别,与已有实验研究结果一致。这里不同于以往类似研究^[2,4]的地方在于所采用的嗜热蛋白质序列数据较大,因而得到的结论更有普遍性。当然,比较结果也表明,尽管嗜热和嗜常温蛋白质的上述指标存在差别,但一一比较同源嗜热和嗜常温蛋白质的上述指标时,总有一部分同源蛋白质某种指标的大小关系与总体趋势不符(例如对于疏水性氨基酸含量、疏水性指数、荷电氨基酸含量该比例分别为33.6%、18.2%和29.1%)。不同嗜热蛋白质有其独特的热稳定性机制,一部分同源蛋白质某种指标的大小关系与总体趋势不符可能正是不同蛋白质有其不同的热稳定性机制的表现,因而要解开某个具体蛋白质的热稳定性机制,只有在三维结构已知的条件下才有可能。

参 考 文 献

- [1] Menendez-Arias L, Argos P. *J Mol Biol*, 1989, **206**(2): 397~406.
- [2] Britton K L, Baker P J, Borges K M M *et al.* *Eur J Biochem*, 1995, **229**(3): 688~695.
- [3] Argos P, Grau M G, Zuber U M *et al.* *Biochemistry*, 1979, **18**(25): 5698~5703.
- [4] Merkle D J, Farrington G K, Wedler F C. *Int J Pept Protein Res*, 1981, **18**(5): 430~442.
- [5] Ikai A. *J Biochem*, 1980, **88**(6): 1895~1898.
- [6] Kyte J, Doolittle R F, *J Mol Biol*, 1982, **157**(1): 105~132.
- [7] Mrabet N T, Van den Broeck A, Van den Broeck I *et al.* *Biochemistry*, 1992, **31**(8): 2239~2253.
- [8] Matthews B W, Nicholson H, Becktel W J. *Proc Natl Acad Sci, U S A*, 1987 **84**(19): 6663~6667.
- [9] Chen H M, Ford C, Reilly P J. *Protein Engineering*, 1995, **8**(7): 575~582.
- [10] Yusuke A, Hideo N, Tsuneo Y. *Applied Microbiol Biotechnology*, 1994, **40**(5): 664~668.
- [11] Matsumura M, Becktel W J, Matthews B W. *Nature*, 1988, **334**(6181): 406~410.
- [12] Ammendola S, Raucci G, Incani O *et al.* *Protein Engineering*, 1995, **8**(1)31~37.
- [13] Perutz M F, Raidt H. *Nature*, 1975, **255**(5505): 256~259.
- [14] McRee D E, Redford S M, Getzoff E D *et al.* *J Biol Chem*, 1990, **265**(24): 14234~14241.

A COMPARISON OF AMINO ACID COMPOSITION OF PROTEINS FROM THERMOPHILES AND MESOPHILES

Lu Baisong Wang Guoli Huang Peitang

(*Institute of Biotechnology, Academy of Military Medical Sciences, Beijing 100071*)

Abstract The thermophilic feature of thermophiles are due to the high thermostability of their proteins. In order to investigate the mechanism of the high thermostability of thermophilic proteins and compare the difference of amino acid composition between thermophilic and mesophilic proteins, 110 pairs of homologous protein sequences from thermophiles and mesophiles respectively are collected and the amino acid composition, percentage of hydrophobic amino acids, hydropathy index, percentage of charged amino acids of the two protein groups are compared. It reveals that the two groups have evidently different composition in several kinds of amino acids, and thermophilic proteins have higher average hydropathy and charged amino acids composition. A comparison of the aliphatic indices between the two protein groups indicates that the reason that thermophilic proteins have higher aliphatic index is that thermophilic proteins have higher leucine composition, and thus sheds doubt on the validity of aliphatic index. It shows that it is possible to disclose some general rules with regard to the mechanism of protein thermostability by comparing a large amount of protein sequence data.

Key words Thermophilic protein, Hydropathy, Charged amino acid composition, Aliphatic index