

文章编号: 1000-5013(2012)06-0655-05

# 嗜盐蛋白高级结构对其稳定性的影响

张光亚

(华侨大学 化工学院, 福建 厦门 361021)

**摘要:** 选择 30 对具有晶体结构的嗜盐和非嗜盐同源蛋白,通过计算氨基酸所处的结构状态,找出二者在统计学上具有显著差异的氨基酸. 研究表明:在总的氨基酸组成上,嗜盐蛋白中天冬氨酸含量更多,而苯丙氨酸和赖氨酸含量显著较少. 在不同二级结构区域, $\beta$ -折叠中两种蛋白氨基酸组成差异最大;而在不同溶剂可及表面,两种蛋白在分别表面氨基酸组成差异最大,嗜盐蛋白表面酸性氨基酸显著较多,而碱性氨基酸则较少;48 种高级结构结构参数中有 4 种存在显著差异,这 4 种结构参数都与可及性表面积有关.

**关键词:** 嗜盐蛋白;非嗜盐蛋白;二级结构;可及性表面积;溶剂可及性;嗜盐机制

**中图分类号:** O 517

**文献标志码:** A

存在于嗜盐菌细胞内的酶或蛋白在工农业及医药领域有着广泛用途,其在高盐环境下稳定性的机制也引起了众多研究者的关注. 有研究表明:嗜盐蛋白往往在其表面引入酸性氨基酸(D, E) 残基与碱性氨基酸(K, R) 残基并形成盐桥,消除了盐离子的屏蔽效应,使分子结构具有刚性,这对其结构的稳定起决定性的作用<sup>[1]</sup>. Panagiotis 等<sup>[2]</sup>通过同源建模的方法模拟了嗜盐及非嗜盐二氢叶酸还原酶的结构,表明较多的负电荷及弱的疏水核心是该酶在高盐环境下稳定的原因. Sandip 等<sup>[3]</sup>比较了几种嗜盐菌及非嗜盐菌基因组及蛋白质组的信息,发现二者在同义密码子使用上具有相似性,而在蛋白质组水平上,则主要表现为嗜盐蛋白中酸性氨基酸较多,同时更容易形成无规则卷曲(coil),不容易形成螺旋结构. 上述研究仅比较了一两个直系同源蛋白(orthologs),或者主要针对一维的氨基酸序列,而对其高级结构涉及较少. 大量研究证实,在蛋白质分子不同高级结构的区域存在不同的相互作用,这对蛋白质稳定性至关重要<sup>[4-5]</sup>. 由此可见,氨基酸在不同结构区域的分布会导致不同的物理化学相互作用,对蛋白质稳定性的贡献也可能存在差异. 因此,阐明蛋白质不同高级结构区域氨基酸组成的差异对了解其稳定性机制意义重大. 本文挑选了 30 对嗜盐和非嗜盐同源蛋白,系统分析了它们在不同结构区域氨基酸组成上的差异.

## 1 材料和方法

### 1.1 蛋白质晶体结构信息的获取

选取了 12 种嗜盐微生物(表 1),按以下 3 个步骤获取蛋白晶体结构. 1) 从蛋白质结构数据库(PDB, <http://www.pdb.org/>)<sup>[6]</sup>分别获取来源于这 12 种微生物的蛋白质晶体结构. 为了减少序列冗余,剔除了同一性大于 30% 的序列. 2) 利用获得具高级结构的蛋白质序列,分别用 Blast P 程序搜索蛋白质结构数据库(PDB),其  $E$  值为默认值  $1.0 \times 10^{-10}$ . 从返回的结果中寻找来源于嗜盐微生物的同源蛋白. 3) 若返回多条符合要求的结果,则选取和目标序列最为匹配且具有较高分辨率的蛋白;若未返回结果,则原始蛋白结构也舍弃. 最后,可以得到 30 对嗜盐和非嗜盐同源蛋白,其在 PDB 数据库的 ID 号及来源微生物,如表 1 所示.

**收稿日期:** 2012-02-05

**通信作者:** 张光亚(1975-),男,教授,主要从事酶学与生物信息学的研究. E-mail: zhgyghh@hqu.edu.cn.

**基金项目:** 国家自然科学基金资助项目(20806031);福建省自然科学基金资助项目(2009J01030);福建省高校新世纪优秀人才支持计划项目(07176C02)

表1 30对同源蛋白及其来源  
Tab.1 30 pairs of halophilic and non-halophilic proteins

嗜盐蛋白		非嗜盐蛋白		同一性 /%	E 值
ID号 <sup>①</sup>	菌名	ID号	菌名		
2X98A	<i>Halobacterium salinarum</i>	3E2DA	<i>Vibrio sp. g15-21</i>	33	$2.008\ 86 \times 10^{-13}$
1TJOA	<i>Halobacterium salinarium</i>	2VXXA	<i>Synechococcus elongatus</i>	35	$2.201\ 44 \times 10^{-21}$
1MOJA	<i>Halobacterium salinarium</i>	2C2FA	<i>Deinococcus radiodurans</i>	31	$2.565\ 61 \times 10^{-11}$
1O6ZA	<i>Haloarcula marismortui</i>	3H3JA	<i>Staphylococcus aureus</i>	38	$9.232\ 61 \times 10^{-47}$
1DOIA	<i>Haloarcula marismortui</i>	1EWYC	<i>Nostoc sp. pcc 7119</i>	51	$1.100\ 9 \times 10^{-14}$
2J5KA	<i>Haloarcula marismortui</i>	3D0OA	<i>Staphylococcus aureus</i>	38	$1.069\ 23 \times 10^{-46}$
2J5QA	<i>Haloarcula marismortui</i>	3D4PA	<i>Staphylococcus aureus</i>	38	$1.069\ 23 \times 10^{-46}$
1ITKA	<i>Haloarcula marismortui</i>	1MWVA	<i>Burkholderia pseudomallei</i>	57	0
2J5RA	<i>Haloarcula marismortui</i>	3D0OA	<i>Staphylococcus aureus</i>	38	$1.069\ 23 \times 10^{-46}$
2AZ3A	<i>Halobacterium salinarum</i>	3B54A	<i>Saccharomyces cerevisiae</i>	50	$7.083\ 49 \times 10^{-42}$
1TK6A	<i>Halobacterium salinarum</i>	2VXXA	<i>Synechococcus elongatus</i>	35	$2.201\ 44 \times 10^{-21}$
1TKPA	<i>Halobacterium salinarum</i>	2C2FA	<i>Deinococcus radiodurans</i>	31	$2.565\ 61 \times 10^{-11}$
2AZ1A	<i>Halobacterium salinarum</i>	1NDLA	<i>Drosophila melanogaster</i>	48	$7.953\ 93 \times 10^{-40}$
2ZUAA	<i>Haloarcula quadrata</i>	3R9LA	<i>Giardia lamblia</i>	59	$9.597\ 49 \times 10^{-49}$
1VDRA	<i>Haloferax volcanii</i>	3JW3A	<i>Bacillus anthracis</i>	34	$7.843\ 88 \times 10^{-22}$
2X0RA	<i>Haloarcula marismortui</i>	3D0OA	<i>Staphylococcus aureus</i>	38	$6.113\ 46 \times 10^{-47}$
1TKOA	<i>Halobacterium salinarum</i>	2C2FA	<i>Deinococcus radiodurans</i>	31	$2.565\ 61 \times 10^{-11}$
1D3AA	<i>Haloarcula marismortui</i>	1LLCA	<i>Lactobacillus casei</i>	34	$1.343\ 97 \times 10^{-46}$
3NEPX	<i>Salinibacter ruber</i>	1UXIA	<i>Chloroflexus aurantiacus</i>	53	$2.796\ 66 \times 10^{-90}$
3GYYA	<i>Halomonas elongata</i>	3FXBA	<i>Silicibacter pomeroyi</i>	59	$2.149\ 5 \times 10^{-103}$
3PAEA	<i>Acinetobacter baumannii</i>	1E3UA	<i>Pseudomonas aeruginosa</i>	39	$2.007\ 68 \times 10^{-39}$
3P7YA	<i>Enterobacter cloacae</i>	1OYBA	<i>Saccharomyces pastorianus</i>	36	$3.424\ 4 \times 10^{-54}$
3AFJA	<i>Cellvibrio gilvus</i>	1V7VA	<i>Vibrio proteolyticus</i>	34	$4.182\ 06 \times 10^{-131}$
3ACTA	<i>Cellvibrio gilvus</i>	1V7WA	<i>Vibrio proteolyticus</i>	34	$5.507\ 7 \times 10^{-131}$
3ACSA	<i>Cellvibrio gilvus</i>	1V7XA	<i>Vibrio proteolyticus</i>	34	$1.216\ 79 \times 10^{-130}$
2XHHA	<i>Cellvibrio japonicus</i>	2XFDA	<i>Escherichia coli</i>	86	$1.834\ 94 \times 10^{-48}$
2XHJA	<i>Cellvibrio japonicus</i>	2XFEA	<i>Escherichia coli</i>	86	$1.834\ 94 \times 10^{-48}$
3AB1A	<i>Chlorobaculum tepidum</i>	3LZWA	<i>Bacillus subtilis</i>	40	$9.184\ 19 \times 10^{-67}$
3OR5A	<i>Chlorobaculum tepidum</i>	3KCMA	<i>Geobacter metallireducens</i>	34	$1.625\ 26 \times 10^{-17}$
3P84A	<i>Enterobacter cloacae</i>	2R14A	<i>Pseudomonas putida</i>	52	$1.582\ 32 \times 10^{-98}$

① ID号里最后一位字母表示多肽链的序号

## 1.2 氨基酸所处区域的确定

PDB数据库得到的晶体结构数据已经包含了经实验所得的二级结构信息,该信息的读取通过DSSP<sup>[7]</sup>完成;溶剂可及性由PaleAle@UCD<sup>[8]</sup>完成;氨基酸所在区域的确定参照文献[9]进行.具体方法如下:当某个氨基酸溶剂可及性面积与其在PDB数据库相应氨基酸的最大溶剂可及性面积的比值趋近于0( $<4\%$ )时,该氨基酸被定义为内核区氨基酸;当比值较大( $>25\%$ )时,该氨基酸被定义为表面氨基酸,剩余的氨基酸(在 $4\% \sim 25\%$ 之间)在被定义为中间区域氨基酸.

## 1.3 高级结构参数的计算

提交晶体结构数据至VADAR服务器<sup>[10]</sup>(<http://vadar.wishartlab.com/index.html>),它通过晶体结构文件计算主链、侧链、氢键、范德华力、分子大小,以及3种二级结构含量等48个结构参数.

## 1.4 统计分析

为了比较嗜盐和非嗜盐蛋白在不同结构区域氨基酸的差异程度,进行 $t$ -检验.本研究中,自由度为58,自由度等于嗜盐蛋白和非嗜盐蛋白的数量减2. $t$ 检验中的 $p$ 值为结果可信程度的一个递减指标.其中: $0.05 \geq p > 0.01$ 差异具有统计学意义; $0.01 \geq p > 0.001$ 差异具有高度统计学意义;而 $p \leq 0.001$ 差异具有极高度统计学意义.

## 2 结果和分析

### 2.1 嗜盐和非嗜盐蛋白在不同二级结构氨基酸组成的差异

整体而言,两种类型蛋白氨基酸组成存在显著差异的氨基酸较少,仅有赖氨酸(K)、苯丙氨酸(F)、异亮氨酸(I)和天冬氨酸(D).众所周知,赖氨酸为碱性氨基酸,天冬氨酸为酸性氨基酸,苯丙氨酸和异亮氨酸为非极性疏水氨基酸.由于各种氨基酸自身的情况及周围环境的限制,造成在嗜盐菌与非嗜盐菌的蛋白质组成上的差异,如表 2 所示.由表 2 可知,嗜盐菌较非嗜盐菌天冬氨酸(D)含量多,赖氨酸(K)和苯丙氨酸(F)含量少,具有极高度统计学意义;而异亮氨酸(I)含量亦少较多,具有高度统计学意义.

表 2 嗜盐和非嗜盐蛋白氨基酸组成的差异

Tab. 2 Difference of amino acids in halophilic and non-halophilic proteins

氨基酸	二级结构			溶剂可及性表面			综合
	$\alpha$ -螺旋	$\beta$ -折叠	无规则卷曲	湮没	居中的	暴露	
Ala(A)	0.176 0	↑0.008 4 <sup>①</sup>	0.964 7	↑0.047 5	0.823 9	0.810 2	0.280 6
Cys(C)	↓0.039 9	↓0.044 3	0.051 9	↓0.044 1	↓0.007 4	0.374 1	0.500 2
Asp(D)	0.060 9	0.068 6	↑0.000 1	0.559 4	0.942 4	↑0.000 3	↑0.000 6
Glu(E)	0.115 4	0.179 5	0.956 8	0.220 07	0.967 1	↑0.002 2	0.979 4
Phe(F)	↓0.005 4	0.893 9	↓0.002 9	0.141 2	↓0.000 9	↓0.000 3	↓0.000 5
Gly(G)	0.072 6	0.118 0	0.378 3	0.605 0	0.170 2	0.077 0	0.052 3
His(H)	↑0.045 4	0.210 3	0.627 4	↑0.015 5	0.997 3	0.389 4	0.759 0
Ile(I)	0.736 0	↓0.023 1	0.657 1	↓0.001 6	0.760 2	0.144 1	↓0.046 4
Lys(K)	↓0.000 1	0.323 0	↓0.002 9	0.076 0	↑0.036 5	↓0.000 1	↓0.000 1
Leu(L)	0.708 0	0.096 4	0.713 9	0.871 4	0.200 6	0.417 0	0.453 7
Met(M)	0.376 3	↑0.006 1	0.874 9	0.936 8	0.727 6	0.327 1	0.699 2
Asn(N)	0.145 9	↓0.047 7	0.126 7	0.397 7	0.448 0	0.154 6	0.345 4
Pro(P)	0.254 0	0.970 7	0.591 9	0.854 9	0.878 5	0.349 0	0.372 0
Gln(Q)	0.346 4	0.818 6	0.482 1	0.633 2	0.275 1	↓0.007 0	0.209 1
Arg(R)	0.636 5	↑0.031 2	0.777 4	0.112 9	↑0.044 9	0.252 4	0.448 3
Ser(S)	0.076 0	↑0.001 1	0.420 8	0.550 0	↓0.010 2	0.205 9	0.945 5
Thr(T)	0.547 0	0.377 3	0.908 0	0.883 8	0.410 1	0.898 6	0.959 6
Val(V)	0.795 6	0.405 9	0.757 3	0.939 5	0.117 5	↓0.001 5	0.774 8
Trp(W)	0.233 3	0.966 9	0.874 6	0.730 6	0.468 0	0.367 1	0.518 0
Tyr(Y)	0.874 8	0.525 1	0.962 3	0.732 9	0.419 2	↓0.022 8	0.409 4

① ↑表示在嗜盐蛋白中含量较高;↓表示在嗜盐蛋白含量较少;数值为  $t$  检验的  $p$  值

在 3 种二级结构里,存在差异最多的是  $\beta$ -折叠结构,有 7 种氨基酸组成有统计学意义,其次为  $\alpha$ -螺旋和无规则卷曲,存在有统计学意义的氨基酸数量分别为 4 个和 3 个.在嗜盐菌蛋白中,亲水性氨基酸(L, M)和带正电荷氨基酸(K)是倾向于形成螺旋结构,而天冬氨酸等则偏向于形成无规则卷曲.与此同时,异亮氨酸显著的少,而缬氨酸与苏氨酸量较多,倾向于形成  $\beta$ -折叠结构.

嗜盐菌的  $\beta$ -折叠结构,相较于非嗜盐菌含有大量的 A(丙氨酸),M(甲硫氨酸),S(丝氨酸)和 R(精氨酸),较少的 C(半胱氨酸),I(异亮氨酸),D(天冬酰胺).A 为中性,支链较短小,排列紧密,是  $\beta$ -折叠结构起主要作用的成分;M, S 绝对含量亦相对较多,含有的羟基或硫会起到一定酸性的作用;R 绝对含量较少,会起到一定碱性作用.两种性质的结合,使得  $\beta$ -折叠结构具有两性性质而更易于适应环境的变化.

另一方面,嗜盐菌含有较于非嗜盐菌少的 C, I 和 N(天冬酰胺).C 中的一 SH 易与 Zn 离子结合.这对于蛋白质经常带这么一个金属离子在  $\beta$ -折叠结构边上是难以想象的,故其绝对含量也甚少是较为合理的;N 是酸性氨基酸,其含量少亦是其自身性质所限制的;I 相对含量虽少,但由于其支链长度大于 A,空间位阻较大,不易出现在  $\beta$ -折叠结构. $\beta$ -折叠片层的结构类似折叠的条状纸片侧向并排而成,它可赋予蛋白质分子更高的柔性<sup>[11]</sup>,从而加强其再高盐浓度下的稳定性.

### 2.2 嗜盐和非嗜盐蛋白在不同溶剂可及表面氨基酸组成的差异

相比二级结构而言,两种蛋白在不同溶剂可及表面中的组成差别更大.其中,差别最多的是分子表

面的氨基酸组成,存在统计学意义的氨基酸数量为8个,占总数的40%,而中间区域和内核区存在统计学意义的氨基酸数量分别为5个和4个.

由表2可知:D(天冬酰胺),E(谷氨酸)为酸性氨基酸,其在分子表面分布相当丰富;而K(赖氨酸),R(精氨酸),H(组氨酸)等碱性氨基酸,F,V(缬氨酸)等疏水氨基酸,Q(谷氨酰胺),Y(酪氨酸)等不带电的氨基酸也就相应的较少出现在嗜盐蛋白分子表面.这是由于D和E含有较多的羧基,在分子表面容易和水分子发生相互作用(氢键),即水合作用,它对维持蛋白质的稳定性非常重要<sup>[12]</sup>.这与相关报道吻合.此外,带负电荷的氨基酸也容易与溶液中的金属离子( $K^+$ , $Na^+$ )形成离子键,也可以很好解释为何嗜盐菌蛋白与周围水分子形成较非嗜盐菌蛋白多40%以上的氢键,以及其细胞内部积累了大量的无机离子(特别是 $K^+$ ).值得一提的是,这种蛋白表面的负电荷可以相互排斥,产生一种既能抵消凝聚,又能在低盐浓度下破坏蛋白结构的效应.

在蛋白质分子内核区,疏水残基A与I都容易出现在内核区,但二者含量在两种蛋白中相反.表明嗜盐蛋白在疏水性氨基酸的使用上存在某种偏好.由于A的侧链更短,分子量更小,可以让嗜盐蛋白内核区组装的更为紧密,这对嗜盐蛋白质稳定性有重要贡献<sup>[13]</sup>.中间区域中存在明显差异的氨基酸也比较容易解释,即由于分子表面有大量的D,E,所以分子内部就跟随着K,R,H等带正电的氨基酸,与分子表面带负电荷的氨基酸之间形成静电引力,即所谓的“盐桥”.这证实了之前的研究成果<sup>[1]</sup>.

### 2.3 嗜盐和非嗜盐蛋白在高级结构参数的差异

通过计算两种蛋白48种结构参数,然后进行*t*-检验,结果发现,有4种结构参数在两种蛋白中的差异达到显著水平,占总数的8.3%.这4个结构参数都与可及表面积有关.所谓可及表面积(accessible surface area, ASA)是指溶剂即水分子所能够接触到的蛋白质表面的面积.它是描述蛋白质疏水性的重要参数,而疏水性则是影响蛋白质折叠的重要物理因素.通过对ASA的计算,可分析蛋白质的疏水自由能及其结构稳定性<sup>[6]</sup>.

表3 4种差异显著的结构参数

Tab.3 Four structure parameters with significant differences

项目	$O^-$ 的ASA	非极性ASA比例	带电ASA比例	侧链ASA疏水性
嗜盐	$1\ 209.3 \pm 591.9$	$0.54 \pm 0.04$	$0.26 \pm 0.05$	$22.8 \pm 4.3$
非嗜盐	$918.3 \pm 478.7$	$0.58 \pm 0.04$	$0.21 \pm 0.04$	$27.0 \pm 6.4$
<i>p</i> 值	0.040 7	0.000 1	0.000 1	0.004 0

由表3可知,嗜盐蛋白在嗜盐菌蛋白氧离子( $O^-$ )的可及性表面积显著高于非嗜盐蛋白.这是由于嗜盐蛋白中存在很多的酸性氨基酸(天冬氨酸、谷氨酸),它们侧链上都多出一个羧基基团,这进一步证实了之前嗜盐蛋白分子表明存在更多的酸性氨基酸的结果.嗜盐蛋白非极性可及表面积的比例较少,与非嗜盐蛋白之间的差异达到极高度统计学意义,非极性区域与水分子作用较少,其对蛋白质稳定性的贡献也较少,故而在嗜盐蛋白中含量较少也就比较容易理解了.

带电可及表面的比较则相反,嗜盐蛋白明显高于非嗜盐蛋白,其差异也达极高度统计学意义.这也是由于在高盐条件下,嗜盐蛋白中带电荷的氨基酸一方面可以和水分子电离的稳定 $H^+$ 和 $OH^-$ 形成静电引力,另一方面也可以和细胞质中存在的 $Na^+$ 或者 $K^+$ 离子之间形成离子键,从而嗜盐蛋白的结构.这样同样解释了上面不同结构区域氨基酸组成的差异.侧链可及表面积疏水性嗜盐蛋白显著低于非嗜盐蛋白,而嗜盐菌中包含的疏水性氨基酸确实较非嗜盐菌少,也可以很好地解释这种差异.

## 3 结束语

选取30对具有晶体结构的嗜盐和非嗜盐蛋白,通过计算氨基酸所处的结构状态以其找出二者在统计学上具有显著差异的氨基酸.结果证实尽管在一级结构(序列)上它们具有较高的相似性,但在不同结构状态中氨基酸残基差异显著,其中在 $\beta$ -折叠结构及分子表面二者氨基酸组成差异最大.通过统计48中结构参数,结果表明,有4种结构参数在两类蛋白中存在显著差异.有趣的是这4种结构参数都与可及表面积有关.

## 参考文献:

- [1] 吕爱军,胡斌,温洪宇,等. 极端嗜盐菌的特性及其应用前景[J]. 微生物学杂志,2005,25(2):65-68.
- [2] PANAGIOTIS L,KASTRITIS N C,PAPANDREOU S J H. Haloadaptation: Insights from comparative modeling studies of halophilic archaeal DHFRs[J]. International Journal of Biological Macromolecules,2007,41(4):447-453.
- [3] PAUL S,BAG S K,DAS S,et al. Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes[J]. Genome Biology,2008,9(4):R70.
- [4] PACK S P,YOO Y J. Protein thermo-stability: Structure-based difference of amino acid between thermophilic and mesophilic proteins[J]. Journal of Biotechnology,2004,111(3):269-277.
- [5] 孙桂鸿,郭明雄,龚睿,等. 人朊病毒蛋白及其突变体的溶剂可及性和模拟分析[J]. 中国病毒学,2004,19(2):158-162.
- [6] BERMAN H M. The protein data bank: A historical perspective[J]. Acta Crystallographica,2008(A64):88-95.
- [7] KABSCH W,SANDER C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features[J]. Biopolymers,1983,22(12):2577-2637.
- [8] POIIASTRI G,JMMARTEN A,MOONEY C,et al. Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information[J]. BMC Bioinformatics,2007,8:201.
- [9] GLYAKINAL A V,GARBUZYNSKIY S O,LOBANOV M Y,et al. Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms[J]. Bioinformatics,2007,23(17):2231-2238.
- [10] WILLARD L,RANJAN A,ZHANG H,et al. VADAR: A web server for quantitative evaluation of protein structure quality[J]. Nucleic Acids Res,2003,31(13):3316-3319.
- [11] KABSCH W,SANDER C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features[J]. Biopolymer,1983,22(12):2577-2637.
- [12] KUMAR S,TAI C J,NUSSINOV R. Factors enhancing protein thermostability[J]. Protein Engineering,2000,13(3):179-191.
- [13] AKASAKO A,HARUKI M,OOBATAKE M,et al. Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid substitution at a cavity within the hydrophobic core[J]. Journal of Biological Chemistry,1997,272(30):18686-18693.

## Influence of Advanced Structures on Protein Halophilic Stability

ZHNAG Guang-ya

(College of Chemical Engineering, Huaqiao University, Xiamen 361021, China)

**Abstract:** To investigate the structural distribution responsible for protein halophilicity is of great significance for understanding the stability of halophilic protein and would help to develop a practical strategy for designing halophilic proteins. A systematic comparative analysis of 30 pairs of halophilic and non-halophilic proteins was reported. The residue structural states based on secondary structure and solvent accessibility were considered for analyzing the structural patterns of single amino acids. The statistical test revealed that higher frequency in overall of Asp at the expense of Phe and Lys, higher frequency in exposed state of acid amino acids at the expense of alkaline amino acids. The  $\beta$ -sheet and the exposed surface were the regions showed great differences. Among the 48 structure parameters, there were four parameters which showed significant difference between the two kinds of proteins. Interestingly, the four parameters were all related with the accessible surface area.

**Keywords:** halophilic protein; non-halophilic protein; secondary structure; accessible surface area; solvent accessibility; halophilic mechanism

(责任编辑: 陈志贤 英文审校: 刘源岗)