拟南芥中 PAP 特异磷酸酶 (AtAHL) 的折叠识别与结构预测*

彭 艺 曾宗浩**

(中国科学院生物物理研究所，北京 100101)

摘要 通过现有的序列同源性比较、二级结构预测、三维结构预测和模拟方法，得到了拟南芥中 PAP 特异磷酸酶的三维结构. 这是一种与酵母中的 Hal2p 蛋白质类似，并且 N 端为 α+β，C 端为 α/β 结构域的多结构域蛋白质. 分析预测所得结构，发现了 Mg^{2+} 等金属离子的结合位点，推测了对 Na^+ 敏感的结构基础. 这些结合位点与其生化功能相关. 而且，通过结构与功能分析，讨论了蛋白质数据库 (PDB) 中同一个酶已有理论结构的不合理性.

关键词 PAP 特异磷酸酶 (AtAHL)，蛋白质三维结构，结构预测，拟南芥

学科分类号 Q6

土壤的盐碱化是农业生产中的一个重要问题. 在盐碱化的地区需要发展有更强耐盐性的作物. 认识对盐敏感的新陈代谢反应是理解与调控耐盐性的关键. 在硫酸盐的吸收中，无机硫酸盐通过两步激活，同时消耗 2 个 ATP 分子，最后以 3′-磷酸腺苷-5′-磷酰硫酸 (PAPS) 的形式积累起来. 当这个被激活的硫酸盐还原为亚硫酸盐或者转化为其他分子时，会形成一个核苷酸，即 3′-磷酸腺苷-5′-磷酸 (PAP)[1]. 积累的 PAP 会对细胞产生毒性. 为了阻止 PAP 的积累，PAP 磷酸酶催化 PAP 转化为 AMP 和无机磷酸[2]. 已经有报道表明，酵母中的 *HAL2* 基因编码的蛋白质 (Hal2p) 是一种把 PAP 水解为 AMP 的核苷酸酶，并且是 Li^+ 和 Na^+ 毒性在生物体内的主要作用目标. Hal2p 需要 Mg^{2+} 帮助催化，也被低浓度的 Li^+ 和 Na^+ 抑制. 在盐胁迫的情况下，Hal2p 的抑制导致细胞中的 PAP 积累，从而产生了各种毒效应. 例如，酵母中硫吸收的途径中断，RNA 加工酶被抑制. 而这种抑制致使细胞死亡[1,3].

另外已经在许多植物中发现了与酵母 *HAL2* 类似的基因家族，其表达产物是 *HAL2* 基因编码蛋白质 (Hal2p) 的同源类似物. 其中包括拟南芥中的 AtSAL1[4]，AtSAL2 和 AtAHL[5]，还有水稻中的 OsRHL[6]. 这些酶在体外均不同程度地表现出对阳离子敏感. 这些蛋白质的催化作用都需要二价金属离子的辅助，却被单价 Li^+ 抑制[7]，关于它们改进植物耐盐能力的研究还在进行. 另一方面，研究也发现拟南芥与酵母 *HAL2* 类似的基因家族中有一个 Na^+ 敏感的 PAP 磷酸酶，即 AtAHL. AtAHL 的 PAP 磷酸酶活性对生理浓度的 Na^+ 高度敏感，对 Li^+ 适度敏感. AtAHL 是对 PAP 特异的，并且能够与植物的磺基转移酶协同作用，也是植物中盐毒性的目标蛋白质[5]. 对耐盐性相关蛋白质结构的了解可以更好地理解其生化功能，改进作物的耐盐性. 但是直到现在，拟南芥中的 PAP 特异磷酸酶 (AtAHL)，还没有实验上的结构信息可用. 当某个蛋白质序列存在三维结构已知的同源蛋白质时，预测出这个序列正确的三维结构是可能的[8]. 因此，对于一个蛋白质家族而言，二级结构预测、折叠识别的算法、已知的蛋白质结构及与结构有关的实验数据，都可以用来推测未知结构蛋白质的可能折叠类型. 使用一些已有的计算方法，我们在这里给出 AtAHL 的结构预测结果，以帮助研究者更好地研究植物的耐盐性.

1 方 法

采用常规的蛋白质建模方案[8]来预测 AtAHL 的结构. 用 FASTA (version 3.3t09，May 18，2001)[9] 和 BLASTP (version 2.2.6，April 9，2003)[10]在 PDB (Protein Data Bank)[11]和 Swiss-Prot 中搜索 AtAHL 的同源类似物. 并且用 CLUSTAL W (release 1.6)[12]进行多序列比对. 使用 CE (cl.sdsc.edu/ce.html)[13]检测 AtAHL 同

* 国家自然科学基金资助项目 (30080004).

** 通讯联系人.

Tel：010-64855592，E-mail：zzh@sun5.ibp.ac.cn

收稿日期：2003-02-27，接受日期：2003-04-24

源类似物的结构邻居. 采用几种现存的方法（例如：GOR4、APSSP2、nnPredict、PredictProtein、Pro Predict、SSpro 等）[14~19]进行二级结构预测. 以上二级结构预测工具来自于互联网（www. expasy. org/tools）. 结构分类信息和序列对各种蛋白质折叠的适合程度分别采用了互联网上的服务器 SMART[20,21]（smart. embl-heidelberg. de/）和 3D-PSSM[22]. 根据这些结果选择出合适的结构模板蛋白质，由 CLUSTAL W（release 1. 6）[12]得到待定结构蛋白质与模板蛋白质序列的初步比对. 根据模板蛋白质的二级结构，参照对待定结构蛋白质的二级结构预测，结合 AtAHL 的生化信息以及保守残基区域[5]，最终得到用于三维结构预测和模拟的序列比对. 将最终的序列比对输入程序 MODELLER（release 6）[23]中就可以得到预测的三维结构. 用 PROCHECK（version 3. 5）[24]来检查所得预测结构的立体化学特性. 用 O（xray. bmc. uu. se/alwyn/）检查侧链构象. 最后使用 CNS（version 1. 1）[25]优化局部结构.

2 结 果

2. 1 AtAHL 的同源类似物

拟南芥中的 PAP 特异磷酸酶（AtAHL）全序列有 373 个残基. 在 PDB 中已有一个理论结构（1FLF）. 这个结构整体上二级结构含量极少，尤其在 C 端甚至还有一个可疑的结（图 1a）. 1FLF 的这些结构特征与常见多结构域蛋白质不符合. 使用 FASTA 在 PDB 中寻找 AtAHL 同源序列，发现还有 15 个与 AtAHL 有较强同源性的序列. 其中 5 个序列相同，E 值均为 7.1×10^{-16}，都来自于酵母（*Saccharomyces cerevisiae*）中的 Hal2p 蛋白质，其代表结构（1QGX _A）见图 1b. 还有另外 10 个序列也相同，E 值均为 0.49，都来自于真菌（*Archaeoglobus fulgidus*）FBPASE/IMPASE，其代表结构（1LBZ _A）见图 1c. 另外，用 BLASTP 搜索 PDB，得到 3 组与 AtAHL 有很强同源性的序列（每组内的序列均相同）. 从 3 组序列中选择一个代表，其 PDB 代码分别是 1QGX、1IMA、2HHM，相应的 E 值分别是 1e-42、2e-4、2e-4. 其中 1IMA 和 2HHM 都是属于肌醇一磷酸酶. 同时得到另外 2 组序列，其与 AtAHL 的同源性比以上 3 组弱，其中代表序列 PDB 代码分别是 1LBZ 和 1JP4. 这些蛋白质和 AtAHL 均属于磷酸酶家族，而且都是既含有 α 螺旋又含有 β 折叠的多结构域蛋白质. E 值越低，两个序列同源性越强. 因此，分析以上两种方法搜索的结果，1QGX 是 AtAHL 在 PDB 中较好的同源类似物. 另一方面，对蛋白质序列库 Swiss-Prot 进行搜索，得到 AtAHL 同源序列，都是磷酸酶家族蛋白质序列. 然后，用 CLUSTAL W 对 AtAHL 同源序列进行多序列比对，帮助确定用于结构预测的序列比对. 根据以上 AtAHL 同源类似物的生物功能和晶体结构的初步分析，可以判断现存的理论结构 1FLF 的可靠性不高.

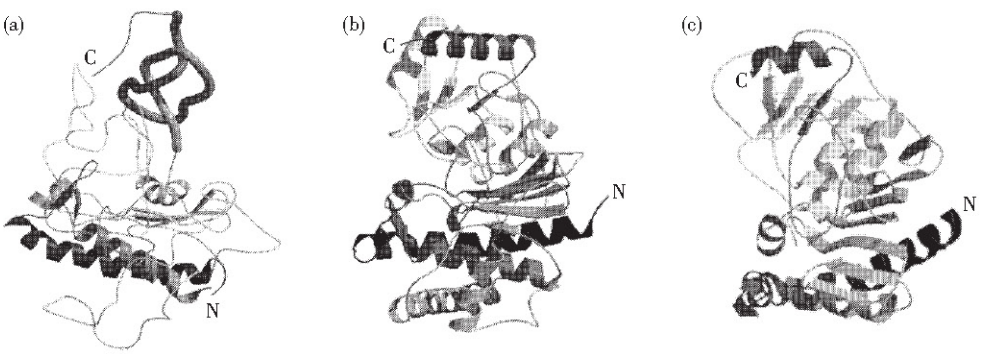

Fig. 1 The Structures of AtAHL and its homologues

(a) theoretical structure of AtAHL in PDB (1FLF); (b) crystal structure of Hal2p (1QGX); (c) crystal structure of FBPASE/IMPASE (1LBZ).

2. 2 结构邻居与折叠识别

用 CE 检测 AtAHL 同源类似物 1QGX 的结构邻居，发现用 FASTA 和 BLASTP 初次鉴定的 AtAHL 同源蛋白质 1IMA、2HHM、1JP4 与 1QGX 是高得分的结构邻居. 将待定结构序列发送到 3D-PSSM 服务器上进行折叠识别. 在具有 95%准确性的 E 值范

围内，得到的结果中包括 1QGX（表 1）. 所有排名在前的蛋白质都是既含有 α 螺旋又含有 β 折叠的多结构域蛋白质，其中 1QGX 排名第一.

2.3 二级结构预测

使用各种方法进行二级结构预测，发现 α 螺旋与 β 折叠都存在. 而且，这个结构可能含有 9～14 个 α 螺旋，11～17 个 β 折叠. 用 SMART 所作分析表明，我们的序列在 SCOP 中也被分类为既有 α 螺旋又有 β 折叠的多结构域蛋白，而且 N 端是 α＋β 结构域，C 端是 α/β 结构域.

Table 1 Validity of the sequence of *Arabidopsis thaliana* PAP-specific phosphatase (AtAHL) to various structures

PDB code	*E*	Class	Fold	Sequence identity
1QGX _A	0.00274	Multi-domain proteins (alpha and beta)	Sugar phosphatases	34%
2HMM _A	0.00412	Multi-domain proteins (alpha and beta)	Sugar phosphatases	21%
1JP4 _A	0.00663	Multi-domain proteins (alpha and beta)	Sugar phosphatases	22%
1INP	0.0238	Multi-domain proteins (alpha and beta)	Sugar phosphatases	17%
1DK4 _A	0.0551	not in SCOP 1.53	hydrolase	19%
1LBZ _A	0.0753	not in SCOP 1.53	hydrolase	23%

* Results from 3D-PSSM Theading Server. The lower the value of *E* is, the more valid the structure is to the sequence.

2.4 模板选择和用于结构预测的序列比对

用 MODELLER 预测三维结构需要有恰当的三维结构模板及其与待定结构的序列比对. 从上面的分析，发现 1QGX 是一个合适的三维结构模板. 用 CLUSTAL W 可以得到 1QGX 与 AtAHL 的序列比对. 在将这个比对直接输入 MODELLER 之前，需要检查比对中插入的空隙. 各种序列比对方法，为了得到最高的序列匹配指标，都会在序列中加入空隙. 在包含空隙的地方，预示着两个序列的三维结构有较大的区别. 从结构上来看，如果这些空隙发生在模板上连续规则的二级结构中间，这些空隙的加入就可能是不合理的. 应该小心地检查待定结构是否有可能在这些地方发生二级结构的中断. 可以作为判断依据的信息有：模板的二级结构和这些二级结构的堆积方式，待定结构的序列在该区域是否有不形成规则二级结构的倾向，等等. 例如，在 β 链中插入空隙或插入更多的残基有可能使整个 β 片层解体；而螺旋的中断可能导致螺旋的堆积方式改变. 这些都是三维结构的较大变化. 如果对模板结构和待定结构之间的相似性有较大的信心，就应该避免这种变化的发生. 因此，在将序列比对输入 MODELLER 之前，检查和调整用各种自动方法得到的序列比对是得到可靠结构的必需步骤.

以 AtAHL 的第 80 位到第 120 位的氨基酸序列为例（图 2）. 在模板蛋白 Hal2p 的结构

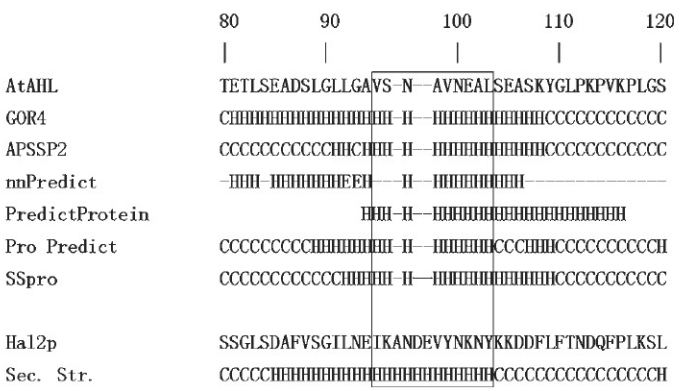

Fig. 2 Predicted secondary structure of AtAHL by different methods and alignment of the two sequences of AtAHL and Hal2p

Below the sequence of Hal2p is its actual secondary structure. H: helix; E: strand; C: coil; blank: loop; -: no-prediction.

(1QGX) 中这段序列对应一个长 α 螺旋. CLUSTAL W 给出的匹配在这段序列的中部加入了几个空隙. 这是否意味着在 PAP 磷酸酶中这一条螺旋应折为两段? 采用各种方法所作二级结构预测都表明方框中的这段序列 (图 2) 是一个连续的 α 螺旋. 没有理由认为在 AtAHL 中由于序列的改变, 这条螺旋会变成不连续的两段. 连续的螺旋仍然是最优的选择. 所以在用于结构预测的序列比对中, 应把相应的空隙删除. 根据同样类似的步骤, 还删除了其他几处空隙. 同时, 根据上文提到的多序列比对的结果, 在不影响 AtAHL 序列保守区的序列比对前提下, 相应地调整了空隙删除处附近区域的序列比对. 图 3 示出了经过调整后用于结构预测的序列比对.

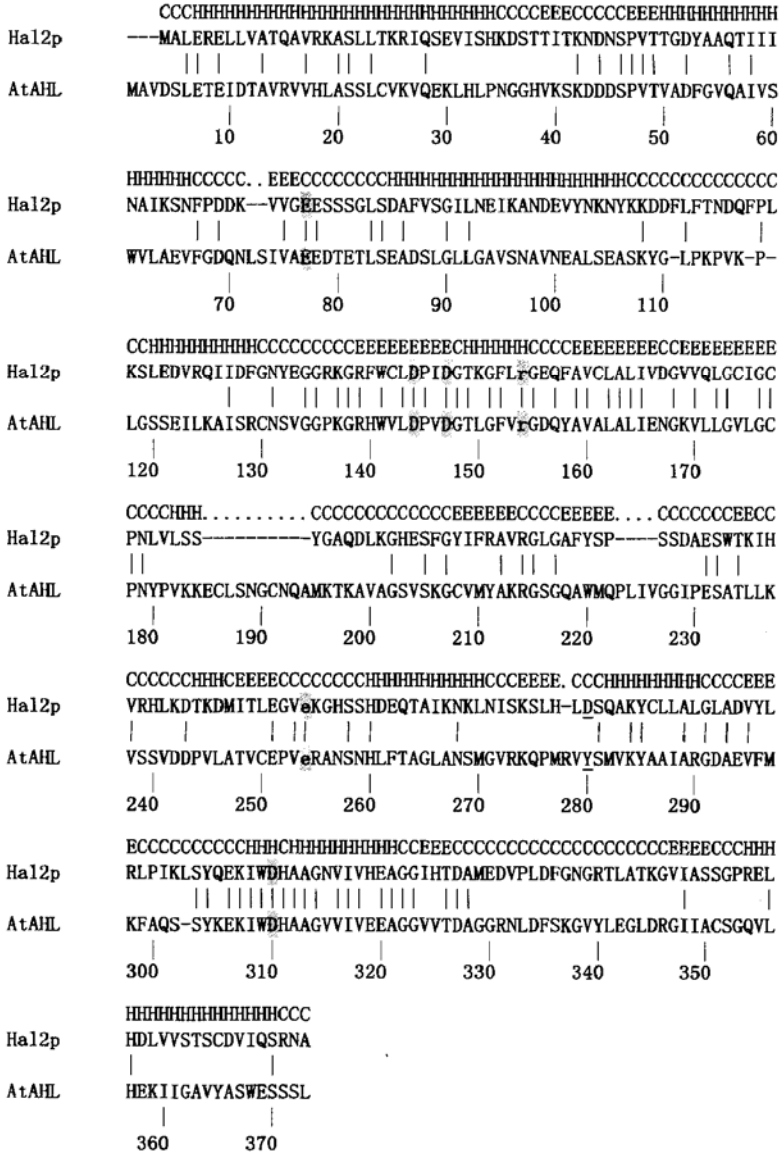

Fig. 3 Sequence alignment of Hal2p and AtAHL used for 3D structure prediction

Secondary structure is calculated by PROCHECK from the structure 1QGX of Hal2p. H: helix; E: strand; C: loop. Residues are numbered by the sequence of AtAHL.

2.5 同源模型的建立与检查

采用 Hal2p 的结构 (1QGX) 为模板, 配以图 3 所示的待定结构序列与模板序列的比对, 通过 MODELLER 得到了 AtAHL 的三维结构. 用 PROCHECK 检查, 这个结构的主链构象有 95.3% 的残基在最适合区域, 4.4% 的残基在附加允许区域 (图 4), 可以说这个结构有很好的立体化学指标. 使用 CNS 进行优化, 有小部分的改进.

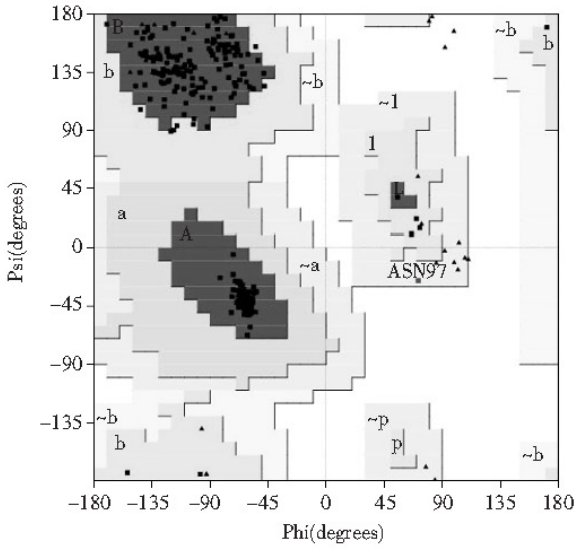

Fig. 4 Ramachandran plot of the predicted structure of AtAIIL

95.3% of the residues are in the most favored, 4.4% in additional allowed and 0.3% in generously allowed regions.

预测的结构 N 端是 α+β 结构域，C 端是 α/β 结构域（图 5）. 预测结构与 IIal2p 主链的比较见图 6.

与之相比，在已有的理论结构 1FLF 中 80.0%的残基在最适合区域，16.0%在附加允许区域，并且只有很低的二级结构含量（约 23.3%），外加一个不大可能形成的结（图 1a）. 这表明，1FLF 的立体化学指标没有我们现在通过各种方法详尽分析之后再用 MODELLER 得到的结果好.

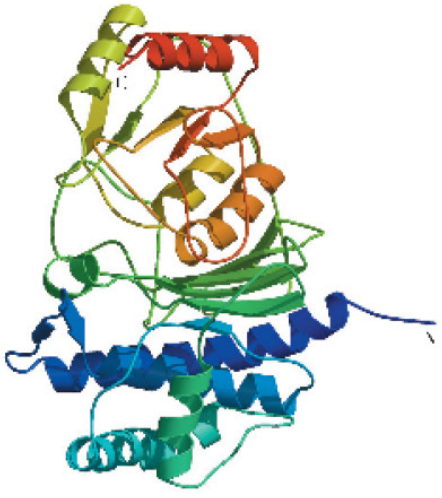

Fig. 5 Cartoon diagram of the predicted structure for AtAIIL

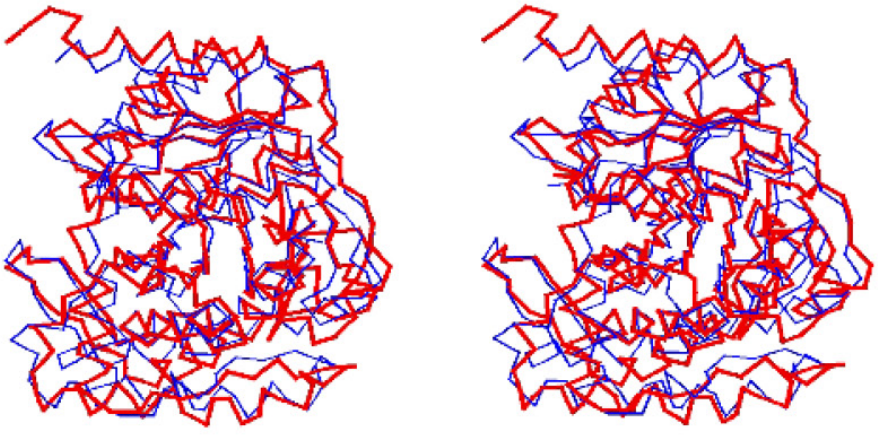

Fig. 6 Stereo view of C_α trace superposition of 1QGX (in blue) with the predicted structure of AtAIIL (in red).

3 讨 论

根据已有的 AtAIIL 功能信息，进一步讨论我们所得预测结构与 PDB 中已经存在的 AtAIIL 理论结构（1FLF）的区别. 1FLF 也是一个理论预测的结构，在它的 C 端有一个结，这是引起我们质疑这个结构的主要原因. 而且通过各种方式检查 1FLF，其立体化学指标不好. 为了解决这个问题，可以进行 AtAIIL 生化功能的分析. 下面我们讨论它的功能性结合位点.

AtAIIL 是一种类似于酵母中的 IIal2p 蛋白. 它们都有 PAP 磷酸酶的活性. IIal2p 对低浓度的 Li^+ 和 Na^+ 敏感；AtAIIL 对生理浓度的 Na^+ 高度敏感，对 Li^+ 适度敏感. 它们在 Mg^{2+} 帮助下催化

PAP 转化为 AMP 和无机磷酸[5]. 那么，AtAHL 应该有与 Hal2p 相似的 Mg^{2+} 和 Li^{+} 等金属离子结合位点，及对 Na^{+} 敏感的结构基础[26]. Hal2p 晶体结构[26]显示 Mg^{2+} 和其他金属离子的结合位点主要由 Glu72、Asp142、Asp145 和 Asp294 以及 PAP 水解产物 AMP 和 Pi 形成. 在 AtAHL 上，与这些残基匹配的是 Glu77、Asp144、Asp147 和 Asp310（图 3 中带灰度底纹大写字母表示）. 根据这些保守残基可以预测 AtAHL 的结合位点. 比较预测结构与 1QGX 的结构，发现在 AMP、Pi 与 Mg^{2+} 位置不变的条件下，Glu77、Asp144、Asp147 和 Asp310 形成类似于 Hal2p 中由 Glu72、Asp142、Asp145 和 Asp294 组成的金属离子结合位点（S1，S2，S3）（图 7）. S1 与 S3 是 Mg^{2+} 的结合位点，S2 是单价金属离子结合位点. Hal2p 中

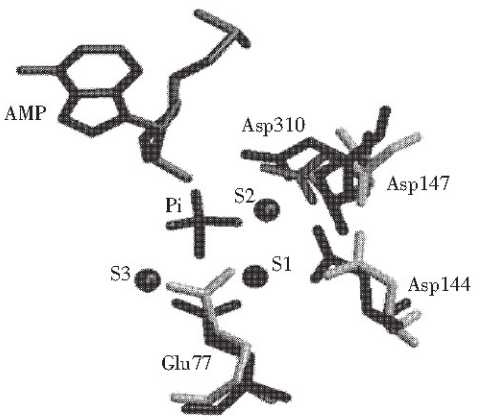

Fig. 7 Comparision of the metal binding site in 1QGX (in black) with that in the predicted structure (in gray)

Amino acid numbering is for the predicted structure. Shown are the residues conserved in both proteins and AMP , Pi and metal ion binding site are those in 1QGX.

Arg152 与 Asp263 之间的盐桥与 Hal2p 对 Li^{+} 和 Na^{+} 敏感的特性有关，这个盐桥可以改变 S2 的尺寸. Arg152 与 Asp263 之间的离子对由一个涉及相邻 Glu154 和 Glu238 的离子网进一步稳定. 从图 3 可以发现在 AtAHL 中 Hal2p 中的 Asp263 被取代为 Tyr（图 3 中带下划线字母），因此在 AtAHL 中不一定形成类似于 Hal2p 的盐桥. 在 Hal2p 中 Arg152 和 Glu238 可能形成盐桥. 从图 3 可以发现 AtAHL 上的 Arg154 和 Glu254 与 Hal2p 上的 Arg152 和 Glu238 匹配（图 3 带灰度底纹小写字母表示）. 预测结构与 1QGX 的结构比较，可以发现 AtAHL 上的 Arg154 和 Glu254 与 Hal2p 上的 Arg152 和 Glu238 几乎重合（图 8）. 因此 AtAHL 上的 Arg154 与 Glu254 可能形成一个盐桥. AtAHL 对 Na^{+} 高度敏感，对 Li^{+} 适度敏感，可能是由于这个盐桥的形成使金属结合位点 S2 变大，造成比较小的 Li^{+} 不能稳定结合（Albert A，Martinez-Ripoll M，Structural Biology of the Salt Tolerance；www. xtal. iqfr. csic. es/CMBE/articulos/salt_tolerance. pdb）. 然而，在 1FLF 的结构中不存在这两个关键结构区域.

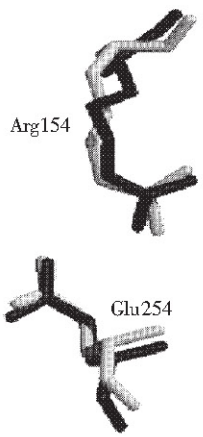

Fig. 8 Comparision of the two salt bridges each in one of the two structures, 1QGX (in black) and the predicted structure (in gray)

Amino acid numbering is for the predicted structure.

引起我们对 AtAHL 结构产生兴趣的原因之一，是其理论结构 1FLF 中结（三叶草结，如图 1a）的存在. 根据早期的观点，蛋白质结构中不会有结存在[27]. 在现存结构中，只发现一些似是而非的结. 从 20 世纪 90 年代中期起，陆续发现了一些较深的结[28]. 根据我们未发表的系统分析，现存 PDB 中（2003 年 3 月）存在三种结：活结（例如 1ED8）、三叶草结（例如 1K3R）和 8 字形结（例如 1YVE）. 这些结表现为规则二级结构以成结方式堆积的结果. 而不是像理论结构 1FLF，由没有二级结构的多肽链形成. 事实上，仅仅根据 AtAHL 与 1QGX 的序列比对，得到了 34% 序列同源性这个结果，就能根据经验断定 AtAHL 与 1QGX 整体上有相似的三维结构. 根据上述基于序列同源性的结构预测，也证明 AtAHL 应该是一个无结的结构. 由于与 1FLF 相关的文章未发表，不能断定其理论结构产生方法中的问题. 但是，根据

经验推测，1FLF 理论结构产生方法中的问题可能有以下几点：a. 没有找到较强同源性的结构模型；b. 过分依赖各种自动化程序，尤其是高温模拟退火程序；c. 对最终结果的合理性未仔细分析. 这些都是结构预测中应该注意的问题. 本文所提供的结构预测信息，有助于拟南芥中的 PAP 磷酸酶生化特性和生物功能进一步的研究. 我们期待完整的 X 射线晶体结构的发表，以验证我们所得到预测结构的正确程度.

参 考 文 献

1 Murguia J R, Belles J M, Serrano R. The yeast HAL2 nucleotidase is an *in vivo* target of salt toxicity. J Biol Chem, 1996, **271** (46): 29029～29033

2 Murguia J R, Belles J M, Serrano R. A salt-sensitive 3′ (2′), 5′-bisphosphate nucleotidase involved in sulfate activation. Science, 1995, **267** (5195): 232～234

3 Patel S, Martinez-Ripoll M, Blundell T L, *et al*. Structural enzymology of Li^+-sensitive/Mg^{2+} dependent phosphatases. J Mol Biol, 2002, **320** (5): 1087～1094

4 Quintero F J, Garciadeblas B, Rodriguez-Navarro A. The SAL1 gene of *Arabidopsis*, encoding an enzyme with3′ (2′), 5′-biophosphate nucleotidase and inositol polyphosphate 1-phosphatase activities, increases salttolerance in yeast. Plant Cell, 1996, **8** (3): 529～537

5 Gil-Mascarell R, Lopez-Coronado J M, Belles J M, *et al*. The *Arabidopsis* HAL2-like gene family includes a novel sodium-sensitive phosphatase. Plant J, 1999, **17** (4): 373～383

6 Peng Z, Verma D P S. A rice HAL2-like gene encodes a Ca^{2+}-sensitive 3′ (2′), 5′-biophosphate nucleoside 3′ (2′)-phosphohydrolase and complements yeast met22 and *Escherichia coli* cysQ mutations. J Biol Chem, 1995, **270** (49): 29105～29110

7 York J D, Ponder J W, Majerus P W. Definition of a metal-dependent Li^+-inhibited phosphomonoesterase protein family based upon a conserved 3-dimensional core structure. Proc Natl Acad Aci USA, 1995, **92** (11): 5149～5153

8 吉伯斯，詹姆贝克 著. 孙超，等译. 生物信息学中的计算机技术. 北京：中国电力出版社，2002. 287～307
Gibas C, Jambeck P. Translated by Sun C, *et al*. Developing Bioinformatics Computer Skills. Beijing: Chinese Electric Power Press, 2002. 287～307

9 Pearson W R, Lipman D J. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA, 1988, **8** (8): 2444～2448

10 Altschul S F, Madden T L, Schaffer A A, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 1997, **25** (17): 3389～3402

11 Berman H M, Westbrook J, Feng Z, *et al*. The protein data bank. Nucleic Acids Res, 2000, **28** (1): 235～242

12 Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994, **22** (22): 4673～4680

13 Shindyalov I N, Bourne P E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering, 1998, **11** (9): 739～747

14 Rost B, Sander C. Prediction of protein structure at better than 70% accuracy. J Mol Biol, 1993, **232** (2): 584～599

15 Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins, 1994, **19** (1): 55～72

16 Kneller D G, Cohen F E, Langridge R. Improvements in protein secondary structure prediction by an enhanced neural network. J Mol Biol, 1990, **214** (1): 171～182

17 Combet C, Blanchet C, Geourjon C, *et al*. NPS@: Network protein sequence analysis. Trends Biochem Sci, 2000, **25** (3): 147～150

18 Baldi P, Brunak S, Frasconi P, *et al*. Exploiting the past and the future in protein secondary structure prediction. Bioinformatics, 1999, **15** (11): 937～946

19 Gilbrat J F, Garnier J, Robson B. Further developments of the protein secondary structure prediction using information theory, new parameters and consideration of residue pairs. J Mol Biol, 1987, **198** (3): 425～443

20 Schultz J, Milpetz F, Bork P, *et al*. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA, 1998, **95** (11): 5857～5864

21 Letunic I, Goodstadt L, Dickens N J, *et al*. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res, 2002, **30** (1): 242～244

22 Kelley L A, MacCallum R M, Sternberg M J E. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol, 2000, **299** (2): 499～520

23 Sali A, Blundell T L. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol, 1993, **234** (3): 779～815

24 Laskowski R A, MacArthur M W, Moss D S, *et al*. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst, 1993, **26** (Pt 2): 283～291

25 Brunger A T, Adams P D, Clore G M, *et al*. Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr, 1998, **54** (Pt 5): 905～921

26 Albert A, Yenush L, Gil-Mascarell M R, *et al*. X-ray structure of yeast Hal2p, a major target of lithium and sodium toxicity, and identification of framework interactions determining cation sensitivity. J Mol Biol, 2000, **295** (4): 927～938

27 Schulz G E, Schirmer R H. Principles of Protein Structure. New York: Springer-Verlag, 1979. 84～87

28 Taylor W R. A deeply knotted protein structure and how it might fold. Nature, 2000, **406** (6798): 916～919

Fold Recognition and Structure Prediction for *Arabidopsis thaliana* PAP-specific Phosphatase (AtAHL)*

Peng Yi, Zeng Zong-Hao**

(*Institute of Biophysics, The Chinese Academy of Sciences, Beijing* 100101, China)

Abstract The three dimensional structure of *Arabidopsis thaliana* PAP-specific phosphatase was predicted by use of various existing methods on sequence comparison, secondary structure prediction, three dimensional structure prediction and simulation. It was a structure similar with that of Hal2p in *Saccharomyces cerevisiae*, consisting of an α+β N-terminal domain and an α/β C-terminal domain. In the predicted structure, possible binding sites for Mg^{2+}, as well as for other metal ions, and the structural base sensitive to Na^{+} were found. These sites were related with the biochemical function of *Arabidopsis thaliana* PAP-specific phosphatase. The structural and functional analysis suggested that the theoretical structure of *Arabidopsis thaliana* PAP-specific phosphatase, having been deposited in PDB, is not reasonable.

Key words PAP-specific phosphatase (AtAHL), three dimensional structure, structure prediction, *Arabidopsis thaliana*

---

* This work was supported by a grant from The National Natural Science Foundation of China (30080004).

** Corresponding author. Tel: 86-10-64855592, E-mail: zzh@sun5.ibp.ac.cn

Received: February 27, 2003 Accepted: April 24, 2003