

DOI: 10.11931/guihaia.gxzw201506023

齐小琼,葛亚飞,李大卫. 水蕨血红蛋白基因的分子克隆和序列分析[J]. 广西植物, 2016, 36(2):216-223

QI XQ, GE YF, LI DW. Molecular clone and sequence analysis of hemoglobin gene in *Ceratopteris thalictroides*[J]. *Guihaia*, 2016, 36(2):216-223

水蕨血红蛋白基因的分子克隆和序列分析

齐小琼¹, 葛亚飞¹, 李大卫^{2*}

(1. 湖北第二师范学院 化学与生命科学学院/植物抗癌活性物质提纯与应用湖北省重点实验室, 武汉 430205;
2. 中国科学院武汉植物园 植物种质创新与特色农业重点实验室, 武汉 430074)

摘要: 植物血红蛋白(Hemoglobin)是一类由珠蛋白(Globin)和血红素(Ferroheme)组成的结合蛋白,在植物中广泛分布,迄今已在苔藓植物、裸子植物和被子植物中克隆到血红蛋白基因序列,但在蕨类植物中相关研究还未见报道。该研究采用热不对称交错PCR(TAIL-PCR)方法克隆了水蕨血红蛋白基因的全长序列。该基因的序列总长为949 bp,包含4个外显子和3个内含子,编码189个氨基酸。预测的蛋白质(命名为CtHb)的分子量为21.14 kDa,等电点(pI)为7.81。三维结构模拟表明CtHb具有植物血红蛋白典型的三级结构:即含有A、B、C、E、F、G和H螺旋,形成了3-on-3的“三明治”结构。和水稻血红蛋白的三级结构相比,CtHb的大部分结构(包括具有远端和近端组氨酸定位的E螺旋和F螺旋的位置等)同水稻的结构极为相似。两者的不同之处主要表现在:(1)CtHb含有较长的N-端区域;(2)两者CD-loop的折叠方式不同;(3)两者螺旋B和螺旋C的连接方式不同,CtHb是通过卷曲连接的,而水稻中借助的是螺旋。结构进化分析揭示了植物血红蛋白从非共生到共生进化过程中的一些关键改变,这些改变可能有助于非共生血红蛋白向共生血红蛋白结构的转变,特别是有助于豆血红蛋白共生功能的实现。

关键词: 植物血红蛋白, 序列分析, 三级结构模拟, 结构进化, 水蕨

中图分类号: Q943.2, Q75 **文献标识码:** A **文章编号:** 1000-3142(2016)02-0216-08

Molecular clone and sequence analysis of hemoglobin gene in *Ceratopteris thalictroides*

QI Xiao-Qiong¹, GE Ya-Fei¹, LI Da-Wei^{2*}

(1. School of Chemistry and Life Sciences, Hubei University of Education/Hubei Key Laboratory of Purification and Application of Plant Anticancer Active Ingredients, Wuhan 430205, China; 2. Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China)

Abstract: Plants, like humans, contain hemoglobins (Hbs), which consist mostly of protein subunits (the “globin” molecules) and heme groups. Phylogenomic analysis shows that Hbs are widely distributed in higher plants, and by comparing sequences, expression patterns, and ligand-binding properties, it is evident that three distinct types of Hb exist in plants: symbiotic, non-symbiotic, and truncated Hbs. In the recent years, the genes of Hb have been cloned and identified in various species of bryophytes, gymnosperms and angiosperms, but no study has been reported in the ferns until

收稿日期: 2015-06-24 修回日期: 2015-12-04

基金项目: 国家自然科学基金(31201242, C130405);湖北省科学研究计划项目(Q20143004);湖北省自然科学基金(2015CFB508);植物抗癌活性物质提纯与应用湖北省重点实验室开放课题(HLPAI 2014004)[Supported by the National Natural Science Foundation of China(31201242, C130405); Scientific Research Project of Educational Commission of Hubei Province(Q20143004); Natural Science Foundation of Hubei Province(2015CFB508); Open Research Fund of Hubei Key Laboratory of Purification and Application of Plant Anticancer Active Ingredients(HLPAI 2014004)].

作者简介: 齐小琼(1981-),女,山东昌邑人,博士,讲师,主要从事进化遗传学研究,(E-mail)108816564@qq.com.

***通讯作者:** 李大卫,博士,副研究员,主要从事果树遗传育种方向,(E-mail) david.lee1983@163.com.

now. In the present study, a full-length Hb gene from *Ceratopteris thalictroides* was cloned by Thermal Asymmetric Inter-laced (TAIL) PCR, and sequence analysis was conducted by bioinformatics. The results revealed that the gene was 949 bp long, containing 4 exons interrupted by 3 introns. The deduced protein (named CtHb) with 189 amino acid residues had a predicted isoelectric point (PI) of 7.81 and a calculated molecular mass 21.14 kDa. Modeling of the tertiary structure indicated that CtHb possesses the typical tertiary structure as plant Hbs (including helices A, B, C, E, F, G and H, these together forms a “sandwich” structure of 3-on-3). Comparative structure analysis of CtHb with rice NsHb1 revealed that most of the CtHb structure was quite similar to that of rice NsHb1, including the positions of helices E and F, where distal and proximal His were located, respectively. However, the structure differences were as follows: (1)The N-terminus region was longer in CtHb than in rice NsHb1; (2)The CD-loop folded differently; (3)Helices to coils transition at the connect regions of helices B and C were different. Structural comparison revealed major evolutionary changes during plant Hbs evolution. Some of these structural changes would be helpful for the transition from nonsymbiotic Hb to symbiotic Hb and for the specialization of a symbiotic function for symbiotic Hb.

Key words: plant hemoglobin, sequence analysis, tertiary structure modeling, structural evolution, *Ceratopteris thalictroides*

植物血红蛋白 (Hemoglobin) 是一类由珠蛋白 (Globin) 和血红素 (Ferroheme) 组成的结合蛋白, 广泛存在于植物界, 在植物固氮及抵御胁迫中具有重要作用。随着研究的不断深入, 根据血红蛋白的序列、表达方式以及配体结合性质把植物中的血红蛋白分为三类: 共生的血红蛋白 (Symbiotic hemoglobin, sHb)、非共生的血红蛋白 (Nonsymbiotic hemoglobin, NsHb) 和截短的血红蛋白 (Truncated hemoglobin, tHb)。共生的血红蛋白主要存在于豆类和一些非豆类植物 (主要是放线菌结瘤植物和榆科植物某些属) 固氮根瘤菌侵染的根瘤细胞中。非共生的血红蛋白比共生的血红蛋白分布更广泛, 不仅存在于含有共生血红蛋白的植物中, 而且也存在于其他植物中; 并可在多种器官中表达, 如正常生长条件和胁迫条件下植物的根、茎、叶、花和种子中 (Anderson et al, 1996; Lira-Ruan et al, 2001; Ross et al, 2004; Taylor et al, 1994; Trevaskis et al, 1997)。截短的血红蛋白首先在原核生物、原生动物和真核藻类中发现, 后来在一些植物中也发现了该蛋白 (Watts et al, 2001)。其序列比其他血红蛋白短 (Pesce et al, 2000), 三维结构是由 α -螺旋通过 2-on-2 的方式排列而成, 不同于 3-on-3 标准的血红蛋白折叠。通过系统发育分析表明, 植物中的 tHb 与 sHb、NsHb 形成了不同的谱系 (Lineages), tHb 是通过基因水平转移从细菌中获得 (Vinogradov et al, 2006), 具体的功能还不是很清楚。

血红蛋白基因在植物中普遍存在, 就目前积累的数据来看, 大部分序列来自于被子植物, 在裸子植物和苔藓植物中有几条代表序列, 但蕨类和石松类

植物中尚未见有相关报道, 不完全的序列数据影响了该基因家族的分子进化研究。本研究通过分子克隆手段获得蕨类植物水蕨的血红蛋白基因的完整编码区序列, 并进行生物信息学分析、三级结构模拟及进化分析, 试图阐明该基因在序列组成、基因结构、三级结构等方面的特征, 并推断其可能的功能; 通过与其他分类群中的血红蛋白基因进行比较分析, 探讨其在不同分类群中较独特的进化样式。

1 材料与方 法

1.1 植物材料

蕨类植物水蕨的健康幼嫩叶片采自中国科学院武汉植物园。

1.2 主要分子生物学试剂

普通 Taq DNA 聚合酶、高保真 LA Taq 酶、dNTP 等试剂购自 TaKaRa; DNA 凝胶纯化回收试剂盒购于 Axygen 生物技术有限公司。

1.3 试验方法

1.3.1 水蕨血红蛋白基因部分片段的扩增 分别用拟南芥 (*Arabidopsis thaliana*) *AHB1* 和 *AHB2* 基因 (登录号: AAD26949, AAB82770) 的蛋白序列作为查询序列, 通过对数据库中已报道的蕨类序列进行相似性搜索, 在 *Ceratopteris richardii* 中得到一条同目标序列相似性较高的 EST 序列 (登录号: BQ087262)。根据这条序列设计一对特异引物, 进行 PCR 扩增。引物序列为 F-cera: 5'-TCACCGAGC-ATCTGCGTCAAT-3'; R-cera: 5'-TCAATAGGCAGTA-CCTGACGA-3'。

PCR 反应体系为 20 μL , 引物各 40 pmol, dNTP 各 1 pmol, 10 \times PCR 反应缓冲液 2.0 μL , Taq 酶 0.2 μL (5 U/ μL)。反应条件为 95 $^{\circ}\text{C}$ 预变性 5 min; 接着 94 $^{\circ}\text{C}$ 变性 50 s, 53 $^{\circ}\text{C}$ 退火 50 s, 72 $^{\circ}\text{C}$ 延伸 60 s, 共 36 个循环; 最后 72 $^{\circ}\text{C}$ 延伸 8 min。

1.3.2 水蕨血红蛋白基因 3' 侧翼序列的扩增 以提取的总 DNA 为模板, 用 3 个巢式单引物进行 3 轮 TAIL-PCR 扩增, 引物序列为 3'TAIL-1: 5'-TCTTAATGTGAGTTCGCCTCTGTGA-3'; 3' TAIL-2: 5'-CTTTGAGCAGCACAGGAATGGTCTC-3'; 3' TAIL-3: 5'-CGGTGTAGTGGATGAGCATTATGAG-3'。

TAIL-PCR 的反应体系为 50 μL , 简并引物 AD (5'-TC (G/C) TICGNACIT (A/T) GGA-3') (Liu and Whittier, 1995)。各成分及用量如下: LA buffer (Mg²⁺ plus) 5.0 μL ; dNTP Mix (2.5 mmol/L) 8.0 μL ; AD 引物 (20 $\mu\text{mol/L}$) 5.0 μL ; 基因特异性引物 (5 $\mu\text{mol/L}$) 2.0 μL ; LA Taq 酶 (5 U/ μL) 0.5 μL ; DNA (50 ng/ μL) 5.0 μL ; H₂O 24.5 μL 。

第一轮反应引物采用 3'TAIL-1 和 AD, 条件为 94 $^{\circ}\text{C}$ 预变性 1 min, 98 $^{\circ}\text{C}$ 预变性 1 min; 接着 94 $^{\circ}\text{C}$ 变性 30 s, 65 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 共 5 个循环; 紧接着 94 $^{\circ}\text{C}$ 变性 30 s, 25 $^{\circ}\text{C}$ 退火 3 min, 在 3 min 内攀升到 72 $^{\circ}\text{C}$, 72 $^{\circ}\text{C}$ 延伸 2 min, 进行一个循环; 接着 94 $^{\circ}\text{C}$ 变性 30 s, 65 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 94 $^{\circ}\text{C}$ 变性 30 s, 65 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 94 $^{\circ}\text{C}$ 变性 30 s, 44 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 共 15 个循环; 最后 72 $^{\circ}\text{C}$ 延伸 10 min。TAIL-PCR 第二轮反应引物采用 3'TAIL-2 和 AD, 将第一轮反应产物稀释 50 倍, 取 1 μL 作为第二轮反应的模板, 其它成分与第一轮反应相同。反应条件为 94 $^{\circ}\text{C}$ 预变性 1 min; 接着 94 $^{\circ}\text{C}$ 变性 30 s, 65 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 94 $^{\circ}\text{C}$ 变性 30 s, 65 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 94 $^{\circ}\text{C}$ 变性 30 s, 44 $^{\circ}\text{C}$ 退火 1 min, 72 $^{\circ}\text{C}$ 延伸 2 min, 共 15 个循环; 最后 72 $^{\circ}\text{C}$ 延伸 10 min。第三轮反应引物采用 3'TAIL-3 和 AD, 将第二轮反应产物稀释 50 倍, 取 1 μL 作为第三轮反应的模板, 其它成分与第二轮反应相同。反应条件也同第二轮相同。

1.3.3 产物回收、连接、转化、克隆、测序 扩增片段凝胶回收采用凝胶回收试剂盒, 参照说明书进行; 产物连接采用 TaKaRa 公司 pMD19-T simple Vector, 参考说明书, 反应体系稍作优化。大肠杆菌感受态细胞制备采用 CaCl₂ 法, 将目的基因通过热激法转化

到大肠杆菌感受态细胞中, 阳性克隆采用蓝白斑筛选方法并进一步采用 PCR 方法, 扩增产物长度与目的片段相当者为阳性克隆。每个目的片段, 随机选取至少 3 个阳性克隆, 送华大基因测序。序列测定采用 ABI PRISM 3730 DNA 测序仪用 BigDye Terminator Kit (Applied Biosystems) 进行双向克隆测序。

1.4 数据分析

采用 BioEdit 软件 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) (North Carolina State University) 对克隆测序得到的片段进行拼接。采用 FGENESH 和 NCBI (<http://linux1.softberry.com/berry.phtml>) 的 ORF finder 进行开放阅读框确定。运用 DNASTAR 软件推测氨基酸序列, 分析蛋白质的分子量、等电点和氨基酸组成。使用 SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) 分析蛋白 N-端氨基酸序列, 根据 SignalP 分析得到的 Cmax 值 (the cleavage site score)、Smean (the signal peptide score) 和 HMM 值 (Hidden Markov Model) 判别信号肽序列, 确定是否存在信号肽。使用 TargetP Server v1.01 (<http://www.cbs.dtu.dk/services/TargetP/>) 确定分泌蛋白的分泌途径和亚细胞定位。蛋白质保守结构功能域用 PROSITE、SMART 软件和 PFAM 数据库 (<http://pfam.wustl.edu/>) 进行预测。多条序列的一致性分析用 Bioedit 软件 sequence identity matrix 进行。使用在线软件 SOPMA 预测蛋白二级结构并计算各种二级结构所占的百分比。将推导的氨基酸序列提交瑞士生物信息研究所的 Swiss-model (<http://www.isb-sib.ch/>) (Arnold et al, 2006; Schwede et al, 2003), 基于同源建模的原理, 用水稻的 NsHb1 作为模板 (PDB code: 1D8UA) 进行水蕨血红蛋白三级结构的预测。预测结果用 PyMOL 软件展示和编辑。

运用 Blast 工具 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), 以水蕨血红蛋白作为查询序列, 进行序列相似性搜索, 并选取相似性较高的序列用 MEGA6 软件 (Tamura et al, 2013) 中的邻接算法进行进化树的构建, 其中多重序列比对采取 Clustalx 软件进行。

2 结果与分析

2.1 水蕨血红蛋白基因的克隆

以水蕨的总 DNA 为模板, 用一对特异引物 (F-cera 和 R-cera) 扩增得到一段约 820 bp 的片段 (图

1:a)。经过分析,该序列的 5'端是完整的,因此只需通过染色体步移技术获得 3'侧翼序列。基于获得的序列片段,进一步设计了一组巢式引物通过 TAIL-PCR 来获得 3'侧翼序列。在 TAIL-PCR 的第二轮和第三轮分别获得了几条清晰的条带,根据巢式引物设计的相对位置,选取这两轮反应的产物大小相差大约 100 bp 的条带(图 1:b)作为目的条带。将第三轮的目的条带回收、克隆测序,经分析,得到的序列含有同已知序列完全一致的重叠区。将上述实验获得的序列拼接最终得到水蕨血红蛋白基因的完整编码区序列。

将得到的序列同 *C. richardii* 的 EST(登录号: BQ087262)序列进行比对,发现本研究获得的序列比数据库中的序列多了三段区域,其它部分都完全匹配,推测这三部分是基因的内含子区域。由此获知该基因含有 4 个外显子,3 个内含子,这与其他植物血红蛋白的基因结构相同。说明血红蛋白基因的外显子-内含子结构在植物进化中是保守的。

2.2 全长基因序列及编码蛋白的基本性质分析

对已获得的克隆基因进行分析研究后,得到水

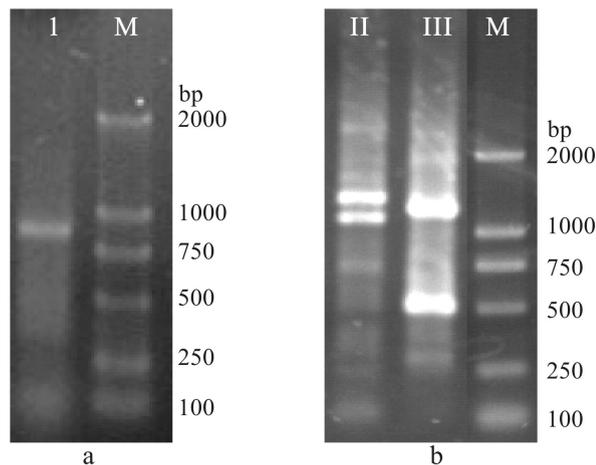


图 1 特异 PCR(a)、TAIL-PCR(b)琼脂糖电泳检测 M.2000 bp 分子量标记; a I. 特异引物 F-cera 和 R-cera 扩增得到的大约 820 bp 的产物; b II. 引物 3'TAIL-2 通过 TAIL-PCR 第二轮反应得到的条带; b III. 引物 3'TAIL-3 通过 TAIL-PCR 第三轮反应得到的条带。

Fig. 1 Agarose gel electrophoresis of the specific PCR(a), TAIL-PCR(b) products M. Molecular weight marker (DL2000); a I. About 820 bp fragment generated by specific PCR with primer F-cera and R-cera; b II. Several bands produced by the secondary reaction of TAIL-PCR with primer 3'TAIL-2; b III. Two clear bands produced by the tertiary reaction of TAIL-PCR with primer 3'TAIL-3.

表 1 序列一致性分析

Table 1 Analysis of percent sequence identify

序列名称 Sequence name	棉花 Ns1 Cotton Ns1	拟南芥 Ns1 AHB1	大麦 Ns1 Barley Ns1	玉米 Ns1 Maize Ns1	水稻 Ns1 Rice Ns1	苜蓿 Lb Alfalfa Lb	百脉根 Lb Lotus Lb	大豆 Lb Soybean Lb	羽扇豆 Lb1 Lupinus Lb1	拟南芥 Ns2 AHB2	棉花 Ns2 Cotton Ns2	小立碗藓 Physcomitrella	角齿藓 Ceratodon	石松 Selaginella
拟南芥 Ns1 AHB1	0.762													
大麦 Ns1 Barley Ns1	0.709	0.674												
玉米 Ns1 Maize Ns1	0.702	0.668	0.824											
水稻 Ns1 Rice Ns1	0.650	0.644	0.777	0.789										
苜蓿 Lb Alfalfa Lb	0.417	0.431	0.370	0.381	0.403									
百脉根 Lb Lotus Lb	0.392	0.393	0.364	0.369	0.385	0.719								
大豆 Lb Soybean Lb	0.390	0.378	0.345	0.339	0.367	0.628	0.655							
羽扇豆 Lb1 Lupinus Lb	0.414	0.403	0.388	0.400	0.409	0.541	0.561	0.538						
拟南芥 Ns2 AHB2	0.487	0.500	0.497	0.476	0.508	0.518	0.474	0.481	0.531					
棉花 Ns2 Cotton Ns2	0.527	0.527	0.518	0.502	0.511	0.522	0.522	0.462	0.550	0.722				
小立碗藓 Physcomitrella	0.450	0.411	0.401	0.412	0.417	0.316	0.322	0.292	0.348	0.366	0.408			
角齿藓 Ceratodon	0.455	0.432	0.416	0.427	0.438	0.325	0.325	0.301	0.346	0.381	0.413	0.828		
石松 Selaginella	0.400	0.373	0.397	0.397	0.397	0.273	0.268	0.252	0.315	0.357	0.361	0.410	0.389	
水蕨 Ceratopteris	0.373	0.365	0.343	0.375	0.380	0.269	0.259	0.243	0.285	0.342	0.329	0.307	0.303	0.341

注: AHB1 和 AHB2 分别代表拟南芥非共生血红蛋白 1 和 2; Ns1 代表非共生血红蛋白 1; Lb 代表豆血红蛋白(共生血红蛋白)。

Note: AHB1 and AHB2 represent nonsymbiotic hemoglobin 1 and 2 from *Arabidopsis thaliana*, respectively; Ns1 indicates nonsymbiotic hemoglobin-1 (NsHb-1); while Lb indicates leghemoglobin (symbiotic hemoglobin).

蕨血红蛋白(命名为 CtHb)基因序列全长(图 2)为 949 bp,该基因包括 4 个外显子和 3 个内含子,编码 189 个氨基酸。氨基酸组成中含量最高的为丙氨酸(Ala, 11.6%),其次缬氨酸(Val, 9.0%)、丝氨酸

(Ser, 8.5%)、亮氨酸(Leu, 7.9%)和赖氨酸(Lys, 7.9%)。碱性氨基酸(Arg, Lys) 24 个,强酸性氨基酸(Asp, Glu) 23 个,疏水氨基酸(Ala, Ile, Leu, Phe, Trp, Val) 71 个,不带电荷的极性氨基酸(Asn, Cys,

```

ATGCGTCCCATAGAAAAGACGCGATGTATTTCCGGTTTCAATGCCGGTGTGCGATACACAAG 60
M R P I E R R D V F P V S M P V S I H K
AATTCAAGAGGAAGTAAAGTGATTGCTGCTGCATTGTCGTCTTTCACGAAGGAGCAGGCT 120
N S R G S K V I A A A L S S F T K E Q A
GCTCTCGTGAAGGACACATGGGCCGTTCTTAAGAAAAGATGCAAGCCGCCACGCAATGGAT 180
A L V K D T W A V L K K D A S R H A M D
TTCTTCTTAATgtgagttcgcctctgtgatttaagcggcggttgttgtgcattcgttaacg 240
F F L M
caatggcacatggtgagacaaggatacgatgacatgaatttgtgcagGGTTTTTGTATGTT 300
                                     V F D V
GCACAGGCCGCAAAGAAGCTCTTCTCCTTCTTAAGGATAGCGACGAGCCCCTTGAGAAG 360
A Q A A K K L F S F L K D S D E P L E K
AATGCGAAGTTGAAGGCACATGCACTGCAAGTTTTTGTGATTgtgtgactctctctctct 420
N A K L K A H A L Q V F V I
catatatttccatcatgccattattgaagaaccatcaatgtccttcaaagtctctgtaatt 480
ctgtctcagtatTTGATATCTCGCATCGAAACATATATATGTATCGTACGTCGCTAACA 540
ttgcttattcccagATCTGCGAGTCTGCTGCATCTTTGAGCAGCACAGGAACGGTCTCAA 600
                                     I C E S A A S L S S T G T V S T
CTCCAGGTTCAACCATGAAAGATATGGCCAGAGCTCATTACCTAGCCGGTGTAGTGGACG 660
P G S T M K D M A R A H Y L A G V V D E
AGCATTATGAGGTgagcatcaaagtgtgcatgttcacgttgcagttttagttcaagttgc 720
H Y E V
ggaaatgtcagcgcagatgattttttctcagccctcgcactttcgccatggcctgataagg 780
atgcggttgttctgctgttgcaggtCGTCAGGTACTGCCTATTGAAGACCATTTCAGGCAGG 840
                                     V R Y C L L K T I Q A G
CCTTCCAGAGGGAATGTGGAATGAGGATGTGAAGCAGGCTTGGCATCATGCATACGATGA 900
L P E G M W N E D V K Q A W H H A Y D E
ACTTGTACGGGAAATGAGGTGCAAGAGCCCCAAACTCCTGGACAGTGA 949
L V R E M R S Q E P Q T P G Q *

```

图2 水蕨血红蛋白基因的核酸序列和推导的氨基酸序列 推导的氨基酸序列标注在开放阅读框下面,小写字母示内含子部分。

Fig. 2 Nucleotide and deduced amino acid sequences of hemoglobin gene from *Ceratopteris thalictroides* The predicted amino acid sequence is shown below its open reading frame, and the introns are present in lowercase letters.

Gln, Ser, Thr, Tyr) 41 个。预测的蛋白质分子量为 21.14 kDa, 理论等电点 (pI) 为 7.81。

2.3 蛋白质的信号肽预测及亚细胞定位分析

采用 SignalP 对该蛋白序列及其他分类群的代表序列 (包括苔藓植物 2 条 (*Physcomitrella patens* 和 *Ceratodon purpureus*)、石松类 1 条 (*Selaginellae moellendorffii*) 和被子植物 11 条 (具体的物种和序列名称见表 1) 进行在线搜索, 发现这些蛋白存在信号肽的概率都极低, 不存在信号肽酶切位点, 说明这些血红蛋白很可能为非分泌蛋白。进一步分析这些蛋白的亚细胞定位, 发现大部分蛋白是细胞质蛋白, 只有小立碗藓和江南卷柏的血红蛋白具有较高分值的叶绿体定位信号, 可能会定位于叶绿体中。

2.4 序列一致性分析

表 1 给出的是 15 条植物血红蛋白序列两两之间的一致性指数, 其中包括 2 条苔藓植物序列 (*Phy-*

scomitrella patens 和 *Ceratodon purpureus*), 1 条石松类植物序列 (*Selaginellae moellendorffii*), 1 条蕨类植物序列 (*Ceratopteris thalictroides*) 及 11 条被子植物序列。本研究发现 CtHb 与被子植物 NsHb1 的序列一致性偏高, 与豆科植物 sHb 的序列一致性偏低, 说明 CtHb 属于 NsHb1; 同时比较 CtHb 和其他分类群 NsHb 的一致性发现, 与古老的苔藓、石松类植物的 NsHb 序列相比, CtHb 在进化上更类似于被子植物 NsHb。

2.5 水蕨血红蛋白高级结构预测

在二级结构预测中, CtHb 主要由 61.38% 的 α -螺旋, 6.35% 的延伸片层, 5.82% 的 β -转角和 26.46% 无规卷曲组成。其中 α -螺旋和卷曲是 CtHb 二级结构的主要构成成分。

了解水蕨血红蛋白的结构性质对于阐明植物血红蛋白的进化多样性及其在植物中发挥的作用都有

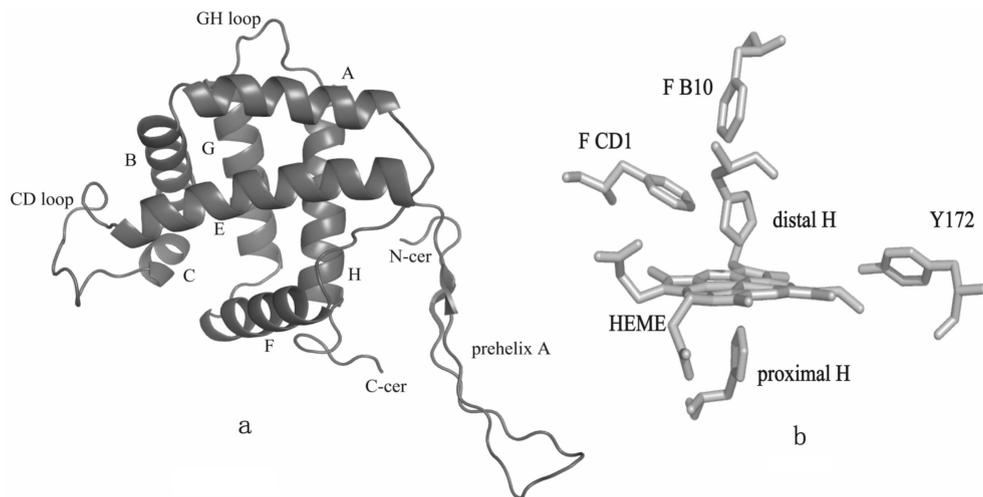


图 3 a. 预测的 CtHb 三级结构; b. 选取的 CtHb 血红素口袋中的氨基酸三级结构 每个螺旋分别用字母 A, B, C, E, F 和 H 表示。

Fig. 3 a. Predicted tertiary structure of CtHb; b. Tertiary structure of selected amino acids in heme pocket of CtHb Helices (including to prehelix A) are indicated with letters A, B, C, E, F and H.

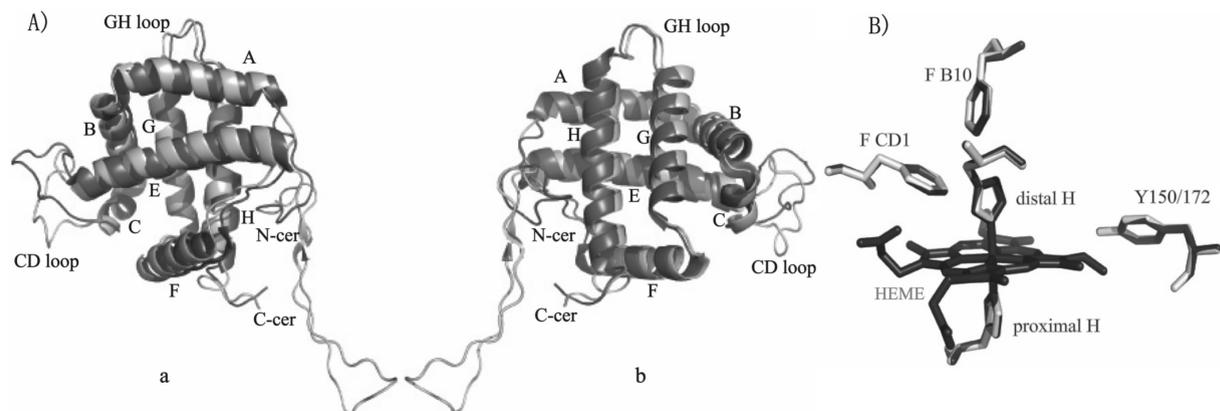


图 4 A. CtHb(灰色表示)和水稻 NsHb1(黑色表示)三级结构对比图; B. 选取在 CtHb(灰色表示)和水稻 NsHb1(黑色表示)血红素口袋中的氨基酸位点的三级结构对比图 每个螺旋分别用 A, B, C, E, F 和 H 表示,其中(a)为正面,(b)为背面。

Fig. 4 A. Overlay of predicted tertiary structure of CtHb(gray) and native rice NsHb1(black) tertiary structures; B. Tertiary structure overlay of selected amino acids in heme pocket of CtHb(gray) and native rice NsHb1(black) Helices(including to prehelix A) are indicated with letters A, B, C, E, F and H. (a) is a front view, while (b) is a back view.

着重要的意义。通过计算机模拟可以从模板预测出高质量的蛋白质三级结构,通过预测出的结构可以考察血红蛋白的进化模式,在结构和功能方面理解进化问题。由于 CtHb 和水稻 NsHb1 具有接近 50% 的相似性,所以可以用水稻 NsHb1(1D8UA)作为预测 CtHb 三级结构的模板。

图 3 给出了 CtHb 三维结构模拟结果。图中显示,CtHb 具有植物血红蛋白典型的三级结构模式,即含有 A, B, C, E, F, G 和 H 螺旋,形成了 3-on-3 的“三明治”结构。将 CtHb 和水稻 NsHb1 的三级结构进行 overlay 比较,结果如图 4: A 所示:预测得到的

CtHb 三级结构具有典型球蛋白的折叠方式,它的大部分结构(包括具有远端和近端组氨酸定位的 E 和 F 螺旋的位置等)同水稻的结构非常相似;不同之处主要表现在:(1) CtHb 含有较长的 N-端区域;(2) 两者 CD-loop 的折叠方式不同;(3) 两者螺旋 B 和螺旋 C 的连接方式不同, CtHb 是通过卷曲连接的,而水稻 NsHb1 借助的是螺旋。

在对比图 5 中,发现苔藓植物、石松类和蕨类植物 NsHb 中都含有一段延长的 N-端序列, Ross et al (2002) 在苔藓植物 NsHb 中检测到了类似于前导肽的蛋白位点,推测植物 NsHb 的祖先可能是定位于

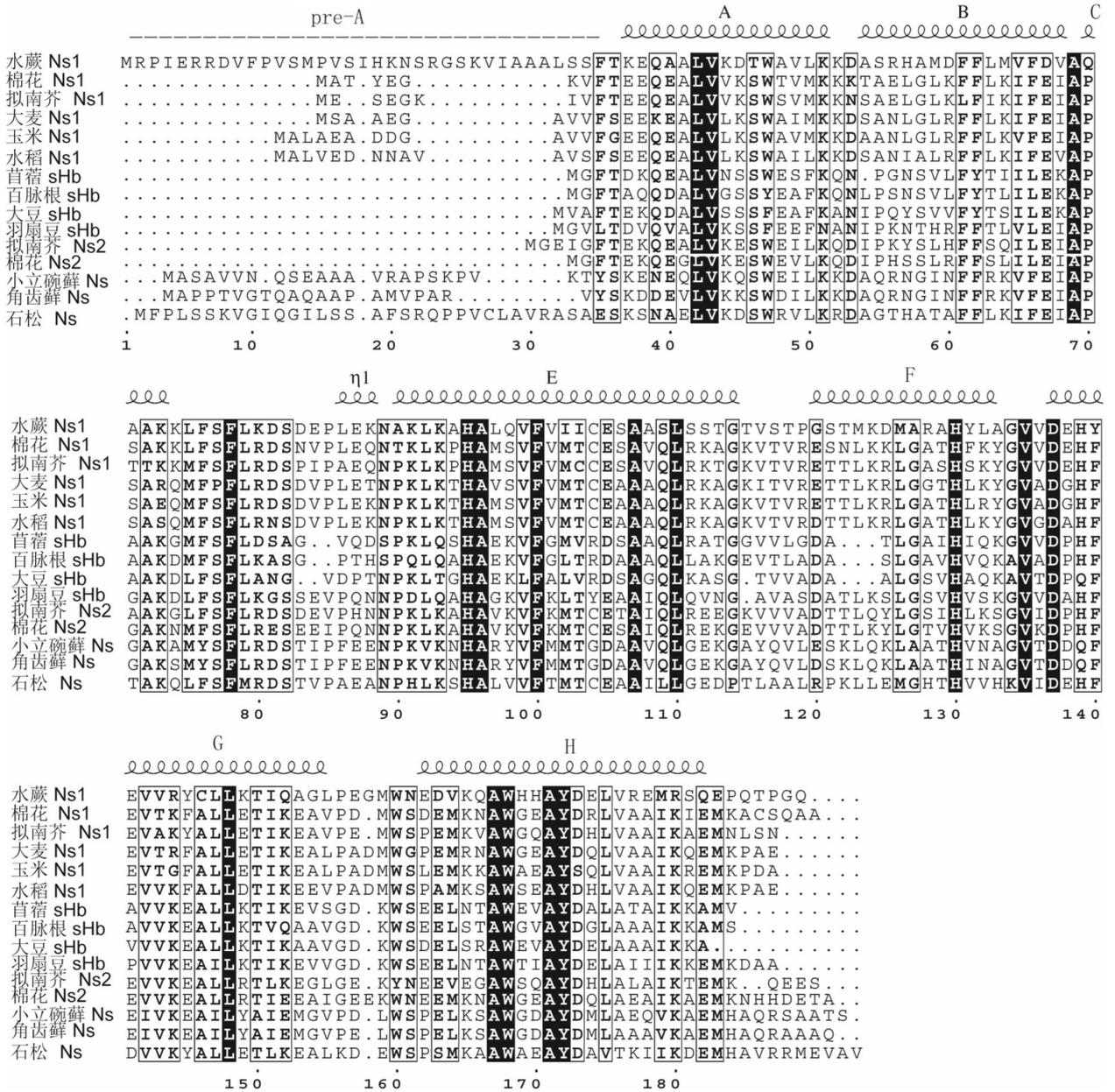


图5 植物血红蛋白的多序列比对 序列的保守性通过 ESPrpt 化学等价策略进行可视化,设置相似性阈值为0.8,以强调严格保守的位置;不变残基添加黑色背景,理化性质相同的残基用灰色矩形框标注;预测的 CtHb 二级结构元件标注在序列比对的的最上面。

Fig. 5 Multiple alignment of plant hemoglobins Sequence conservation is visualized according to the ESPrpt chemical equivalence measure, with a similarity threshold set to 0.8 to emphasize strictly conserved positions; Invariant residues are boxed in black, physico-chemical equivalent residues are boxed in gray; The deduced secondary structure of CtHb is shown above the alignment.

细胞器中的,这同软件预测的小立碗藓 NsHb 的亚细胞定位结果一致。推测这段延长的 N-端氨基酸可能有助于蛋白的转运 (Garrocho-villegas & Arredondo-peter, 2008)。在 CtHb 中,A 螺旋之前存在较长的一段序列,可能与苔藓植物延长的 N-端氨基酸有着相似的作用,即与蛋白的转运有关,但信号肽和亚细胞定位预测都没有支持这一推断。这可能与采用的生物信息学的预测方法有关,需要进一步的实

验证据来证明。如果这些都被验证是正确的,那么从进化的角度讲,NsHb 可能从细胞器的蛋白进化成了细胞质的蛋白 (Ross et al, 2002)。

本研究也检测了参与配基结合的一些重要位点的氨基酸组成和相对位置 (图 4: B),结果表明近端组氨酸 (His), 远端组氨酸,螺旋 B 第 10 位的苯丙氨酸 (Phe), CD-loop 第 1 位的苯丙氨酸及 172/150 的酪氨酸 (Tyr) 在水蕨和水稻 NsHb1 中都非常一

致,位置完全吻合。这表明 CtHb 具有水稻 NsHb1 类似的配基结合动力学性质,即具有很低的 O_2 解离常数,很高的 O_2 亲和力。

4.6 血红蛋白的序列和结构进化

通过将 CtHb 与其他分类群的代表序列比较发现(图 5): CtHb 包含非常保守的近端组氨酸(H95),远端组氨酸(H130)和位于 B 螺旋第 10 位的苯丙氨酸(F62);同时也发现,在 CD-loop 的第 1 位上只有苔藓植物是酪氨酸(Tyr)而其他物种都是苯丙氨酸(Phe)。这说明 CD1 在非共生血红蛋白的祖先序列中是被酪氨酸占据的,而在进化过程中被苯丙氨酸所取代。

图 6 不同物种 GGPPS 蛋白序列的同源比较采用 Neighbor-joining 算法,在 MEGA6 软件平台上构

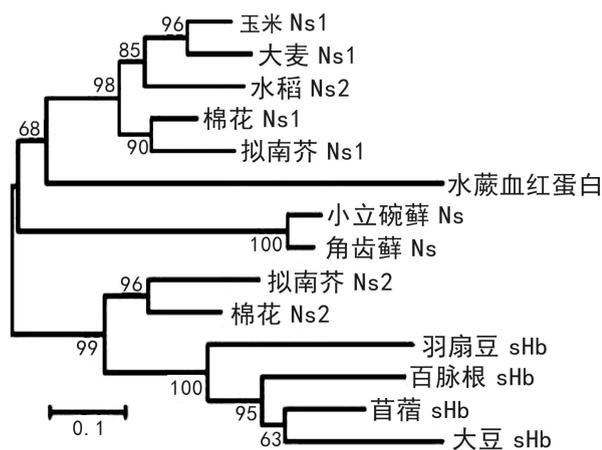


图 6 不同物种的系统进化树

Fig. 6 Phylogenetic tree of different species

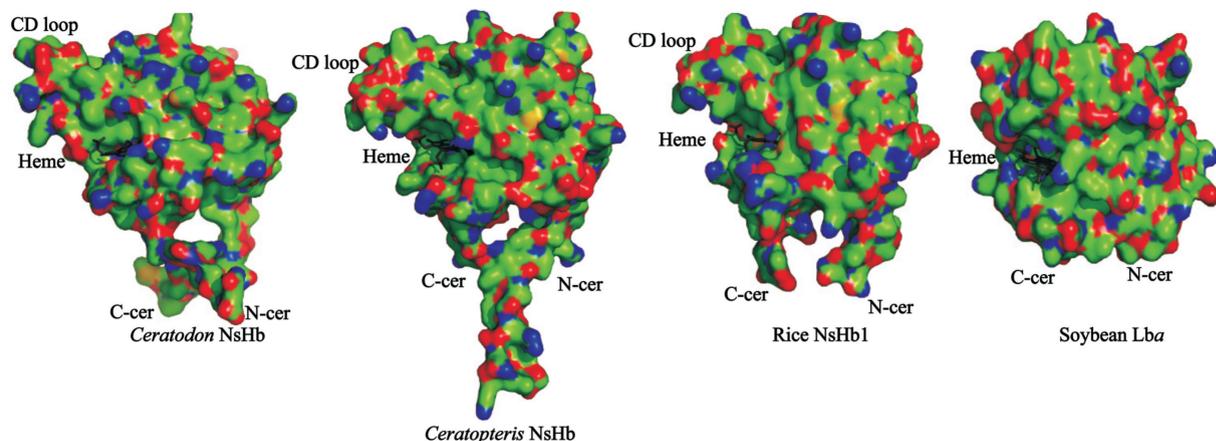


图 7 角齿藓、水蕨、水稻非共生血红蛋白和大豆血红蛋白三级结构和表面电荷分布比较 带有正电荷和负电荷的氨基酸分别用蓝色和红色表示;小立碗藓的三级结构坐标数据(ID number:PM0075013)从蛋白质模型数据库(Protein Model Database)(<http://mi.caspar.it/PMDB/>)下载,水稻和大豆的三级结构坐标数据(ID numbers:1D8U 和 1BIN)从 RCSB 蛋白质数据库(RCSB Protein Data Bank)下载。

Fig. 7 Comparison of the predicted nonsymbiotic hemoglobin from *Ceratodon purpureus*, *Ceratopteris thalictroides*, native rice NsHb1 and soybean Lba tertiary structures and surface charge distribution Blue and red colors represent positively and negatively charged amino acid residues, respectively; Coordinates for nonsymbiotic hemoglobin from *Ceratodon purpureus* was obtained from Protein Model Database (<http://mi.caspar.it/PMDB/>) under the ID number PM0075013. Coordinates for the rice NsHb1 and soybean Lba structures were obtained from RCSB Protein Data Bank with the ID numbers 1D8U and 1BIN, respectively.

建了植物血红蛋白的分子系统进化树(图 6)。根据序列的差异分为两大分支—非共生血红蛋白 1(Ns1)、古老的血红蛋白(Ns)组成的分支和非共生血红蛋白 2(Ns2)、共生血红蛋白(sHb)组成的分支;CtHb 位于非共生血红蛋白 1 的基部,说明其可能是非共生血红蛋白 1 的共同祖先,并且其进化地位比古老的苔藓植物血红蛋白稍微高一点,这跟物种的进化关系是吻合的。

本研究也比较了角齿藓、水蕨、水稻非共生血红蛋白及大豆的豆血红蛋白三维结构的差异,揭示了

植物血红蛋白从非共生到共生进化过程中的一些关键改变(图 7)。这些改变包括:(1)血红素辅基从六配位体(Hexacoordinate)到五配位体(Pentacoordinate)的转变(Gopalasubramaniam et al, 2008);(2)CD-loop, N-端和 C-端区域长度的缩短;(3)蛋白结构更加紧凑,形成球状结构。这些改变可能有助于非共生血红蛋白向共生血红蛋白结构的转变,特别是有助于豆血红蛋白共生功能的实现(Gopalasubramaniam et al, 2008)。

(下转第 161 页 Continue on page 161)