

甘蓝型油菜紫色酸性磷酸酶 基因类似物的生物信息学分析

邹春阳, 臧新, 杨冬之, 曹刚强, 焦思凯

(郑州大学 生命科学院, 河南 郑州 450001)

摘要:利用生物信息网络数据库,对甘蓝型油菜紫色酸性磷酸酶基因(PAP)类似物编码的蛋白进行生物信息学分析,以预测该蛋白的理化性质、分子结构与生物学功能。结果表明:该基因编码的蛋白含 481 个氨基酸,其蛋白的分子式为 $C_{2520}H_{3714}N_{666}O_{719}S_{19}$,分子量为 55.45 kDa,等电点为 6.35,为亲水性分泌型蛋白,并且不稳定;存在 3 个跨膜螺旋区域及 3 个卷曲螺旋结构,含有 1 个信号肽,39 个潜在磷酸化位点和 3 个 N-糖基化修饰,存在 6 种类型特定的功能位点,主要定位于高尔基体;氨基酸序列同源性分析表明,该基因编码的蛋白与甘蓝型油菜 PAP12 家族具有较近的亲缘关系,直系同源源于甘蓝型油菜 PAP12-5;二级结构主要为无规则卷曲、延伸链和 α -螺旋,具有较为明显的 10 个螺旋和 26 个折叠;具有 PAP 的保守结构域,属于金属磷酸二酯酶超家族成员;根据保守结构域及功能位点分析发现该基因属于紫色酸性磷酸酶类,可能属于甘蓝型油菜 PAP 基因家族成员,预测其编码产物在植物磷高效利用与吸收中发挥重要的生理功能。生物信息学分析及结合蛋白的结构与功能预测分析可为深入研究甘蓝型油菜 PAP 提供理论指导。

关键词:甘蓝型油菜;紫色酸性磷酸酶(PAP)基因;生物信息学分析;功能预测

中图分类号:S 634.303.6 **文献标识码:**A **文章编号:**1001-0009(2016)13-0093-05

甘蓝型油菜(*Brassica napus*)是世界上第二大油料作物,然而,在英国约有 8% 的油料作物种植区域缺乏磷元素^[1]。同样的,在中国主要的油料作物生产地区,尤其是长江中游区域,也严重地受到了有效磷匮乏的影响。土壤中有有效磷元素含量匮乏,已成为影响农作物产量的限制因素^[2]。由于油菜对缺磷敏感,所以提高磷吸收和代谢效率对于油菜具有重要意义。在弱酸条件下,酸性磷酸酶(apases)通过活化水解磷酸酐、磷酸酯能释放出无机磷(Pi),供植物生长发育所需。而紫色酸性磷酸酶(purple acid phosphatase, PAP)构成了 apases 中最大的成员,属于双核金属脱氢酶(binuclear metallohydro-lase)家族^[3]。植物紫色酸性磷酸酶(PAP)基因的成功分离对研究植物磷营养机制具有重要意义^[4]。目前已从数十种植物中将 PAP 基因分离出来^[5]。LU 等^[6-7]从甘蓝型油菜中成功克隆出 7 个 PAP12 基因家族,它们

都编码高分子量(HMW)PAP 蛋白;分子进化和系统发育分析结果表明,*Bn*PAP12 直系同源源于 *At*PAP12(拟南芥),与 *At*PAP12 一样,*Bn*PAP12 基因在微管和早期生殖器官的表达强烈,预测它们在低磷条件下与磷元素的转移有关,并且 *Bn*PAP12 基因与 *At*PAP12 基因具有相似的生物学功能。甘蓝型油菜基因组测序工作的完成^[8]为 PAP 基因的深入研究提供了可能。该研究利用生物信息学数据库对甘蓝型油菜 PAP 基因类似物 BnaA04g15420D 编码蛋白的分子结构、理化性质和生物学功能进行了预测分析,为深入研究甘蓝型油菜 PAP 对磷元素吸收和利用机制提供理论参考依据。同时,对于植物紫色酸性磷酸酶基因家族的研究将有助于揭示甘蓝型油菜磷代谢的分子机制,并为油菜进行转基因操作或分子标记辅助选择高磷吸收效率的品种提供机会。

1 材料与方法

1.1 试验材料

甘蓝型油菜 PAP 基因的同源序列 BnaA04g15420D (登录号:CDY48178.1)的数据资料通过 GenBank 数据库获得。

1.2 试验方法

将氨基酸序列利用 NCBI 数据库中的 BLASTP

第一作者简介:邹春阳(1992-),女,硕士研究生,研究方向为甘蓝型油菜酸性磷酸酶磷代谢机制。E-mail:2849605970@qq.com

责任作者:臧新(1968-),男,博士,副教授,硕士生导师,现主要从事红豆杉细胞大规模培养生产紫杉醇等研究工作。E-mail:zan-gxin@zzu.edu.cn

收稿日期:2015-11-02

(<http://www.ncbi.nlm.nih.gov/blastp>)程序进行同源比对,然后利用软件 MEGA 6.06 进行多序列比对并构建系统发育进化树。通过软件 ProtParam(<http://us.expasy.org/tools/protparam.html>)预测该基因编码蛋白的理化性质;利用 TMPred Server(<http://www.ch.embnet.org/software/TMPRED-form.html>)对该蛋白质的跨膜结构进行预测;利用 ProtScale(<http://www.expasy.org/cgi-bin/protscale.pl>)生物学软件,选择 Hphob./Kyte&Doolittle 模式预测该蛋白的亲水性/疏水性;利用 MultiLoc/TargetLoc (<http://abi.inf.uni-tuebingen.de/Services/MultiLoc>)选择 MultiLoc(plant)预测方法进行亚细胞定位预测;通过 SignalP 4.1 Server 预测该蛋白的信号肽结构;分别利用 NCBI 数据库中的 CDD(conserved domain database)和 SMART(<http://smart.embl-heidelberg.de/>)对蛋白序列的超家族及结构域进行分析预测;利用 NetPhos 2.0 Server(<http://www.cbs.dtu.dk/services/NetPhos/>)和 COILS Server (<http://www.ch.embnet.org/software/COILS-form.html>)生物学软件分别预测该蛋白质的潜在磷酸化位点和卷曲螺旋结构;利用 NetNGlyc 1.0 Server 对该蛋白的 N-糖基化位点进行预测;利用 PROSITE(<http://www.expasy.org/prosite/>)预测分析该蛋白的功能位点及模体数;利用在线软件 SOPMA(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)对该蛋白质的二级结构进行预测;利用 Phyre 2.0([http://www.sbg.bio.ic.](http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)

[ac.uk/phyre2/html/page.cgi?id=index](http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index))及 SWISS-MODEL(<http://swissmodel.expasy.org>)网站对蛋白质的三级结构进行预测分析,运用 AntheProt 3D viewer 查看该蛋白三级结构的预测结果。

2 结果与分析

2.1 氨基酸同源性分析及系统发育进化树的构建

对甘蓝型油菜 PAP 基因类似物编码的氨基酸序列进行同源分析后,选取同源性较高的 11 条序列。结果显示,该蛋白与来自甘蓝型油菜(*Brassica napus*) PAP12 基因家族编码的蛋白 BnPAP12-1(ABV89739)、BnPAP12-2(ABV89741)、BnPAP12-3(ABV89743)、BnPAP12-4(ABV89745)、BnPAP12-5(ABV89749)、BnPAP12-6(ABV89755)、BnPAP12-7(ABV89757)、白菜(*Brassica rapa* XP_009140744)PAP12 蛋白、深山南芥(*Arabidopsis lyrata* subsp. *lyrata* XP_002874764)PAP 前体蛋白和拟南芥(*Arabidopsis thaliana* NP_180287)PAP12 蛋白的同源性分别为 92%、93%、92%、87%、97%、89%、92%、93%、85%、85%。然后利用生物学软件 MEGA 6.06 对氨基酸序列进行多重比对并构建系统发育进化树(图 1)。通过上述结果可见甘蓝型油菜 PAP 基因类似物的氨基酸序列与其它植物 PAP 基因具有很高的同源性,其中与甘蓝型油菜 PAP12-5 的同源性最高,其次是甘蓝型油菜 PAP12-2 和白菜 PAP12。

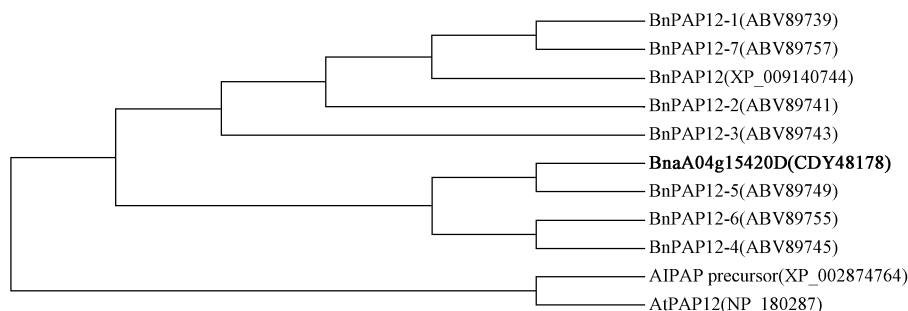


图 1 甘蓝型油菜 BnaA04g15420D 蛋白与其它植物 PAP 蛋白的系统进化树

Fig. 1 Phylogenetic tree of BnaA04g15420D and other plants PAP

2.2 蛋白质理化性质、跨膜结构及亲水性/疏水性分析

利用 ProtParam 在线工具对 PAP 基因类似物蛋白的理化性质进行预测,该蛋白的氨基酸个数为 481,分子量 55.45 kDa,等电点为 6.35,负电荷残基与正电荷残基个数分别为 52 和 47,多肽的原子组成为 $C_{2520}H_{3714}N_{666}O_{719}S_{19}$,预计在哺乳动物网状细胞中的半衰期为 30 h,在酵母中的半衰期大于 20 h,在大肠杆菌中的半衰期大于 10 h,该蛋白的不稳定系数为 41.85,由此说明其是一个不稳定蛋白。

该蛋白的跨膜结构预测结果显示,其具有 3 个跨膜

螺旋区域,对应的氨基酸位置分别在 7~26、78~96 与 282~300;每个跨膜结构域分别为由 20、19 与 19 个氨基酸残基组成的螺旋,并且是比较明显的 N 末端跨膜螺旋。

利用 ProtScale 对该蛋白的亲水性/疏水性进行预测和分析,其第 19 位亮氨酸(Leu)疏水性最强,为 2.889,第 448 位天冬氨酸(Asp)与 449 位甘氨酸(Gly)亲水性最强,为 -2.711,亲水性/疏水性分析显示亲水性负峰值个数多于疏水性正峰值个数,表明该蛋白具有较好的亲水性,为可溶性蛋白质。

2.3 亚细胞定位及信号肽的预测

应用 MultiLoc/TargetLoc 在线工具对该蛋白的亚细胞定位进行预测,结果表明,该蛋白最可能定位于高尔基体,可能性为 0.61。通过以上分析,推断该 PAP 基因类似物属于一种胞外蛋白,具有分泌性蛋白的特性。

信号肽预测结果显示该蛋白序列含 1 条信号肽。信号肽区域可能在第 1~28 位氨基酸之间,其剪切位点位于第 28 与 29 位氨基酸之间。说明该蛋白存在跨膜转移,为分泌性蛋白。

2.4 保守结构域预测

利用 CDD 分析预测该蛋白的保守结构域(图 2),发

现该蛋白在 163~461 氨基酸之间具有一个 PAP 保守家族结构域,属于典型的金属磷酸二酯酶超家族(MPP),包含 PAP 功能位点。利用 SMART 对蛋白进行在线结构域分析显示,在蛋白第 167~366 位,有 1 个 Metallophos(Pfam00149,金属磷酸酯酶家族)的结构域,该结构域具有水解酶活性,在不同种类的磷酸酯酶中均有发现^[9]。在第 391~457 位,有 1 个 Metallophos_C(Pfam14008,紫色酸性磷酸酶)的结构域,该结构域发现于拟南芥(*Arabidopsis thaliana*)PAP 蛋白的 C 末端^[10]。由此可知,预测该蛋白属于 PAP 家族,并且与 NCBI 中 BLAST 的结果吻合。

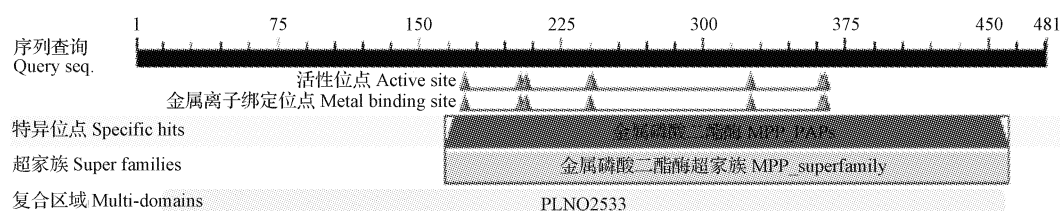


图 2 BnaA04g15420D 蛋白保守结构域分析

Fig. 2 Conserved domain analysis of BnaA04g15420D protein

2.5 潜在磷酸化位点及螺旋卷曲与 N-糖基化位点预测

对该蛋白进行潜在磷酸化位点分析,结果显示,该蛋白具有 19 个丝氨酸磷酸化位点(Ser),14 个酪氨酸磷酸化位点(Tyr),6 个苏氨酸磷酸化位点(Thr)。可以看出,丝氨酸磷酸化在该蛋白中占主导地位,推测丝氨酸磷酸化可能在甘蓝型油菜该蛋白行使功能的过程中发挥重要作用。螺旋卷曲预测该蛋白可能存在 3 个卷曲螺旋结构,而第 301~314 位的可能性指数最大,约为 0.067。预测该蛋白的 N-糖基化位点发现该蛋白可能存在 3 个 N-糖基化位点,分别在第 120 位的 NYTS、第 182 位的 NRTL 和第 313 位的 NRTE。N-糖基化是蛋白质重要的翻译后修饰作用,对蛋白质的功能调节有着非常重要的作用。

2.6 功能位点及模体预测

利用 PROSITE 对该蛋白的模体进行分析,结果发

现该蛋白具有 6 种类型特定的功能位点,包括 12 个蛋白激酶 C 磷酸化位点(protein kinase C phosphorylation site)、7 个 N-肉豆蔻酰基化位点(N-myristoylation site)、6 个酪蛋白激酶 II 磷酸化位点(casein kinase II phosphorylation site)、5 个 N-糖基化位点(N-glycosylation site)、1 个酪氨酸激酶磷酸化位点(tyrosine kinase phosphorylation site)和 1 个依赖于 cAMP 和 cGMP 的蛋白激酶磷酸化位点(cAMP-and cGMP-dependent protein kinase phosphorylation site)。

2.7 二级结构和三级结构预测

由图 3 可知,不规则卷曲(紫色区域)含量最高,约 40.96%,延伸链(红色区域)为 26.40%, α -螺旋(蓝色区域)约 20.37%, β -折叠(绿色区域)最低,约为 12.27%。延伸链和无规则卷曲构成该蛋白二级结构的主要结构元件,分布于整个蛋白结构中。

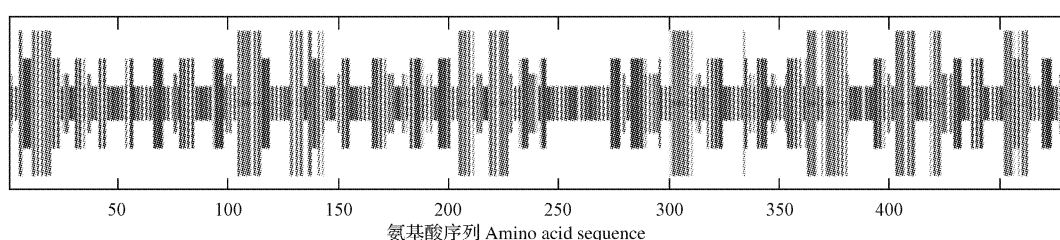


图 3 BnaA04g15420D 蛋白的二级结构预测

Fig. 3 The secondary structure prediction of sugarcane BnPA12-8

分别用 Phyre2.0 和 SWISS-MODEL 网站对甘蓝型油菜 BnaA04g15420D 编码蛋白的三级结构进行预测分

析(图 4、5),其结果具有良好的一致性和相似性。由图 4 推测该蛋白具有 10 个 α -螺旋和 26 个 β -折叠,通过图 5

可以看到该蛋白的整体三维立体结构,红色部分代表 α 螺旋组件,整体表明蛋白亲水性较好。

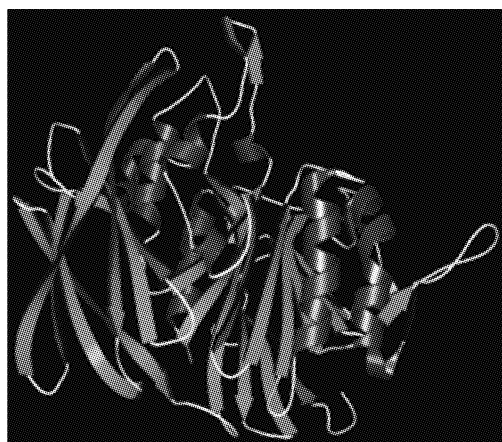


图4 BnaA04g15420D蛋白的彩带立体结构

Fig. 4 Ribbons stereogram of BnaA04g15420D

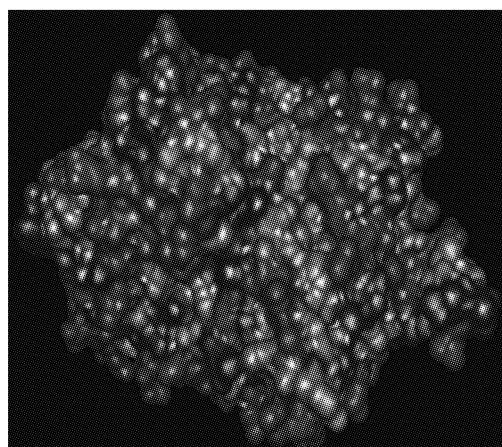


图5 BnaA04g15420D蛋白的疏水表面立体结构

Fig. 5 Hydrophobic model of BnaA04g15420D

3 讨论

近年来,已有许多PAP基因从植物中分离鉴定出来,推动了PAP研究的进展。WONGKAEW等^[11]从绿豆的幼苗子叶中克隆PAP基因(*VrPAP1*),其编码大分子量的PAP蛋白。李东屏等^[12]从拟南芥中分离出3个PAP基因,其编码产物在细胞质中发挥生物学功能。LU等^[7]从甘蓝型油菜中克隆出PAP17基因家族,包括3个成员,并且成员之间具有高度的相似性,都编码低分子量的PAP蛋白;包含金属磷酸酯保守结构域、5个保守模体和7个金属配体残基;它们主要在生殖器官表达,尤其在芽部位;在幼苗时期,这些基因在低磷胁迫时被强烈诱导,当供磷后恢复到基础水平,说明它们在甘蓝型油菜生殖发育和适应缺磷环境中具有重要作用。LU等^[6]发现BnPAP12成员在以器官特异性及低磷胁迫为基础表现出不同的表达模式。这些结果将有助

于揭示甘蓝型油菜磷吸收和利用的分子机制。

该研究利用生物信息学数据库及相关生物在线/离线软件,对PAP基因类似物CDY48178.1编码产物的理化性质和结构功能进行分析预测。通过该蛋白的结构和功能位点预测,推测该蛋白属于植物紫色酸性磷酸酶(PAP)家族。其理化性质、结构和功能位点与已分离出来的*BnPAP12*基因家族编码的蛋白非常相似。同源建模发现,该蛋白的三维结构与芸豆极为相似,一致性为69.71%,为同源四聚体蛋白,其活性位点的金属配体为Fe(III)-Zn(II)。研究发现芸豆PAP蛋白的活性位点位于羧基端的 $\beta\alpha\beta\alpha\beta$ 三明治结构的亚基之间,并且该酶的产物中富含磷酸盐,该类产物能够被植物吸收利用^[13]。综合亚细胞定位,稳定性预测,跨膜结构等信息可推测PAP基因类似物BnaA04g15420D的编码产物可能是位于高尔基体的一种PAP蛋白,预测在植物磷吸收和利用过程中发挥重要作用,由于该蛋白不稳定,这可能也是该基因难以从甘蓝型油菜分离出来的原因之一。但是为研究该基因具体的生物学功能,还需在甘蓝型油菜中进行表达模式分析以及克隆来进行验证。

参考文献

- [1] SKINNER R J, TODD A D. Twenty-five years of monitoring pH and nutrient status of soils in England and Wales[J]. Soil Use Man, 1998, 14: 162-169.
- [2] VANCE C P. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources[J]. Plant Physiology, 2001, 127(2): 390-397.
- [3] HUE T T, QIAN W Q, BRENDEN A. Biochemical and molecular characterization of AtPAP12 and AtPAP26, the predominant purple acid phosphatase isozymes secreted by phosphate-starved *Arabidopsis thaliana* [J]. Plant, Cell and Environment, 2010, 33: 1789-1803.
- [4] DUFF S M G, SARATH G, PLAXTON W C. The role of acid phosphatases in plant P metabolism[J]. Physiologia Plantarum, 1994, 90: 791-800.
- [5] 卢坤, 张凯, 柴友荣, 等. 甘蓝和白菜紫色酸性磷酸酶 17 基因家族的克隆和比较分析[J]. 作物学报, 2010, 36(3): 517-525.
- [6] LU K, CHAI Y R, ZHANG K, et al. Cloning and characterization of phosphorus starvation inducible *Brassica napus* PURPLE ACID PHOSPHATASE 12 gene family, and imprinting of a recently evolved MITE-minisatellite twin structure[J]. Theor Appl Genet, 2008, 117: 963-975.
- [7] LU K, LI J N, ZHONG W R, et al. Isolation, characterization and phosphate-starvation inducible expression of potential *Brassica napus* purple acid phosphatase 17 (*BnPAP17*) gene family[J]. Bot Stud, 2008, 49: 199-213.
- [8] CHALHOUB B, DENOEUDE F, LIU S, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome[J]. Science, 2014, 345(6199): 950-953.
- [9] ARAVIND L, KOONIN E V. Phosphoesterase domains associated with DNA polymerases of diverse origins[J]. Nucleic Acids Research, 1998, 26(16): 3746-3752.
- [10] LI D P, ZHU H F, LIU K F, et al. Purple acid phosphatases of *Arabidopsis thaliana* comparative analysis and differential regulation by phosphate deprivation[J]. The Journal of Biological Chemistry, 2002, 277(31): 27772-27781.

- [11] WONGKAEW A, SRINIVES P, NAKASATHIEN S. Isolation and characterization of purple acid phosphatase gene during seedling development in mungbean[J]. *Biologia Plantarum*, 2013, 57(2): 267-273.
- [12] 李东屏, 王道文. 3 个拟南芥紫色酸性磷酸酶基因的 cDNA 克隆及体外表达[J]. 湖南师范大学自然科学学报, 2003, 26(3): 78-82.

- [13] THOMAS K, NORBERT S, ROLAND F, et al. Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures[J]. *J Mol Biol*, 1996, 259: 737-748.

Bioinformatics Analysis of the *PAP* Gene Analogs in *Brassica napus*

ZOU Chunyang, ZANG Xin, YANG Dongzhi, CAO Gangqiang, JIAO Sikai

(School of Life Science, Zhengzhou University, Zhengzhou, Henan 450001)

Abstract: Bioinformatics network database and softwares were used to analyse and predict the biological properties of the *PAP* (purple acid phosphatase) gene analogs encoding protein in *Brassica napus*. The physical and chemical properties, structure and function of the protein were predicted, in the cause of providing a theoretical foundation for the further study on the adjustment mechanism of phosphorus absorption and utilization of *Brassica napus*. The results indicated that the gene encoded a length of 481 amino acid protein. The molecular formula was $C_{2520}H_{3714}N_{666}O_{719}S_{19}$. Its molecular weight was 55.45 kDa and the PI was 6.35. It was predicted that the protein was hydrophilic, secretory protein and instability, and there was a signal peptide. The protein had 3 transmembrane helical regions, 3 coiled coil structures, 39 potential phosphorylation sites and 3 N-glycosylated sites. It was mainly located in the Golgi apparatus. Homologous analysis showed that the gene had the close genetic relationship with *Brassica napus* *PAP12*, and orthologous to *PAP12-5*. Random coil, extend chains and α -helix were the main secondary structures of the protein, and it had obvious 10 helices and 26 folds. The protein had *PAP*'s conserved domains, which belonged to metal phosphodiesterase super family (MPP). The conserved domain analysis showed that it had two conserved motifs, seven metal-coordinating residues and six specific type functional sites, which predicted that the gene belonged to *Brassica napus* *PAP* gene family. Bioinformatics analysis and the protein structure and function prediction could provide theoretical guidance for further research on *Brassica napus* *PAP*.

Keywords: *Brassica napus*; *PAP* gene; bioinformatics; function prediction

生物信息学

知识窗

生物信息学是一门研究生物和生物相关系统中信息内容和信息流向的综合系统科学,是与计算机科学及应用数学等学科相互交叉而形成的一门新兴学科。广义地说,生物信息学从事对基因研究相关信息的获取、加工、储存、分配、分析和解释。这一定义包括两层含义:一是对大量数据的收集、整理与服务,也就是管理好这些数据;另一个是从中发现新的规律,也就是用好这些数据。具体地说,生物信息是把基因组 DNA 序列信息分析为源头,找到基因组序列中代表蛋白质和 RNA 基因的编码区。同时阐明基因组中大量存在的非编码区的信息实质,破译隐藏在 DNA 序列中的遗传语言规律,在此基础上,归纳整理与基因遗传信息、释放及其调控相关的转录谱和蛋白质谱的数据,从而认识代谢、发育、分化、进化的规律。

(来源:现代药名词手册)