

大麻 *CBDA1* 基因的生物信息学分析

常丽, 唐慧娟, 李建军, 黄思齐, 陈安国, 赵立宁*, 李德芳* (中国农业科学院麻类研究所, 湖南长沙 410205)

摘要 [目的]研究大麻二酚酸合成酶基因(*CBDA1*)编码的蛋白质序列所包含的生物学信息。[方法]利用生物信息学在线工具及软件分析大麻二酚酸合成酶(CBDAS)的理化性质、亲/疏水性、跨膜结构、信号肽、motif 结构域及空间结构等。[结果] CBDAS 由 544 个氨基酸组成, 分子量为 62 168.42, 理论等电点为 8.81, 是一种稳定的亲水性分泌蛋白, N 末端包含 1 个由 28 个氨基酸残基组成的信号肽, 该蛋白的亚细胞定位在胞外。CBDAS 属于氧化还原酶家族, FDA 是该酶活性的必需辅因子, CBDAS 蛋白中只含有 1 个低复杂度区域, 含有 23 个磷酸化位点和 6 个 N-糖基化位点, 其二级结构主要由 α -螺旋、 β -转角和无规卷曲组成, 三级结构与四氢大麻酚酸合成酶(THCAS)同源性最高。[结论]该研究结果可为今后深入研究 CBDAS 蛋白的结构特征和功能提供理论参考。

关键词 大麻; 大麻二酚酸合成酶; 生物信息学

中图分类号 S563.3 **文献标识码** A **文章编号** 0517-6611(2017)29-0144-05

Bioinformatic Analysis of *CBDA1* Gene in *Cannabis sativa*

CHANG Li, TANG Hui-juan, LI Jian-jun, ZHAO Li-ning*, LI De-fang* et al (Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, Hunan 410205)

Abstract [Objective] To study the biological information contained in the protein sequence encoded by cannabidiolic acid synthase gene (*CBDA1*). [Method] The physicochemical properties, hydrophilicity and hydrophobicity, transmembrane structure domain, the signal peptide, the motif and the spatial structure of cannabidiolic acid synthase (CBDAS) were analyzed by bioinformatic sever and online tools. [Result] CBDAS consisted of 544 amino acids with molecular weight of 62 168.42 and a theoretical isoelectric point of 8.81. It was a stable hydrophilic secretory protein and the N-terminal contained a signal peptide of 28 amino acid residues, moreover, the subcellular localization of CBDAS was extracellular. It was presumed that CBDAS belonged to the family of oxidoreductases and FDA was a necessary cofactor for the activity of CBDAS. In addition, CBDAS contained one low complexity region, twenty three phosphorylation sites and six N-glycosylation sites. The secondary structure mainly included α -helix, β -turn and random coil. Furthermore, the tertiary structure was the most homologous to tetrahydrocannabinol synthase (THCAS). [Conclusion] The results could provide a theoretical reference for further study of the structure and functions of CBDAS.

Key words *Cannabis sativa*; CBDAS; Bioinformatics

大麻(*Cannabis sativa*)是一年生草本植物,起源于我国,在南北朝时期人们就开始种植和使用大麻,大麻是我国传统经济作物,主要用于纺织、建材、造纸、药用、食用、饲料、工业原料等方面^[1]。大麻植株中含有多种活性物质,主要分为两大类,即大麻酚类化合物和非大麻酚类化合物。目前,研究最多的是大麻酚类化合物,主要包括四氢大麻酚(tetrahydrocannabinol, THC)、大麻酚(cannabinol, CBN)、大麻二酚(cannabidiol, CBD)、大麻萜酚(cannabigerol, CBG)、大麻环萜酚(cannabichromene, CBC)等。THC 是由以色列 Weizmann 科学研究所的 Yechiel Gaoni 和 Raphael Mechoulam 于 1963 年首次分离得到^[2],并于次年确定其化学结构。THC 是大麻中最重要的活性物质,具有神经保护作用,可用于治疗癌症引起的呕吐^[3],但 THC 具有致幻作用,因此大麻在多国被禁止种植。四氢大麻酚酸合成酶(Tetrahydrocannabinolic acid synthase, THCAS)是 THC 合成途径中的关键酶,最早在 1995 年由 Taura 等^[4]从大麻幼叶中分离出,并于 2004 年成功克隆了该酶的基因^[5],随后在 2012 年研究了该酶的结构和功能^[6]。与 THC 不同的是, CBD 是大麻中的非成瘾性成分,能阻碍 THC 对人体神经系统影响,并具有治疗癫痫、抗痉挛、抗炎、抗焦虑等药理活性^[7-10]。因此,高 CBD 含量的药用大麻成

为当今研究的一个热点。而大麻二酚酸合成酶(Cannabidiolic acid synthase, CBDAS)是 CBD 合成途径中的关键酶,最早在 1996 年由 Taura 等^[11]从墨西哥纤维大麻中分离得到,并于 2007 年通过逆转录获得其 cDNA^[12]。但是 CBDAS 的结构和功能至今还未报道,该研究以 CBDA 合成酶基因为研究对象,采用生物信息学方法对 CBDA 合成酶基因编码蛋白质序列的理化性质、结构特征、修饰位点等进行预测和分析,以期今后深入研究和利用 CBDA 合成酶提供重要的理论依据和研究基础,同时为大麻作物遗传改良提供参考。

1 材料与方法

1.1 材料 以大麻品种 Carmen 的大麻二酚酸合成酶基因(*CBDA1*)(LOCUS KJ469374)为研究对象,对其完整的 CDS 序列编码的氨基酸序列、蛋白质理化性及功能结构域进行预测与分析。

1.2 方法 利用 ExPASy 软件中的 ProtParam 程序对 CBDAS 蛋白的氨基酸序列长度、分子量大小及等电点等进行分析;利用 ProtScale 工具分析 CBDAS 蛋白的亲疏水性;利用 TM-HMM Server v. 2.0 和 Signal IP 4.1 工具分析 CBDAS 蛋白的跨膜结构域及信号肽;利用 ProtComp v. 9.0 工具对 CBDAS 蛋白的亚细胞定位进行分析;利用 PROSITE 模体数据库对 CBDAS 蛋白进行 motif 预测;利用 SMART 工具分析 CBDAS 蛋白的保守功能域;利用 NetPhos 2.0 Server 和 NetNGlyc 1.0 Server 分析其蛋白质翻译后修饰位点;利用 GOR(Garnier-Osguthorpe-Robson Method)对蛋白的二级结构进行性分析;利用 SWISS-MODEL 服务器同源模拟构建 CBDAS 的三级结构。

基金项目 国家麻类产业体系分子育种项目(CARS-19-E04)。
作者简介 常丽(1984—),女,河南焦作人,助理研究员,博士,从事麻类生物技术研究。*通讯作者:赵立宁,研究员,硕士生导师,从事麻类生物技术研究;李德芳,研究员,博士生导师,从事一年生麻类遗传改良研究。
收稿日期 2017-07-31

2 结果与分析

2.1 CBDA1 基因编码蛋白的氨基酸组成 氨基酸的种类、排列顺序及数量直接影响蛋白质的功能。CBDA1 基因的 CDS 序列编码蛋白质的氨基酸序列为：

001 MKCSTFSFWF VCKIFFFFS FNIQTSIANP REN-
FLKCF SQ YIPNNATNLK LVYTQNPNLY
061 MSVLNSTIHN LRFSSDTPK PLVIVTPSHV SHIQG-
TILCS KKVGLQIRTR SGGHDSEGM S
121 YISQVPFVIV DLRNMRSIKI DVHSQTAWVE AGATL-
GEVYY WVNEKNESLS LAAGYCPTVC
181 AGGHFGGGGY GPLMRSYGLA ADNIIDAHLV NVH-
GKVLDRK SMGEDLFWAL RGGGAESFGI
241 IVAWKIRLVA VPKSTMFSVK KIMEIHELK LVNK-
WQNIAY KYDKDLLMT HFITRNITDN
301 QGKNKTAIHT YFSSVFLGGV DSLVDLMNKS FPEL-
GIKKT DCRQLSWIDTI IFYSGVVNYD
361 TDNFNKEILL DRSAGQNGAF KIKLDYVKKP IPES-
VFVQIL EKLYEEDIGA GMYALPYGG
421 IMDEISESAI PFPHRAGILY ELWYICSWEK QED-
NEKHLNW IRNIYNFMTP YVSQNPRLAY
481 LNYRDLDIGI NDPKNPNNTY QARIWGEKYF GKN-
FDRLVKV KTLVDPNFF RNEQSIPPLP
541 RHRH

CBDAS 由 544 个氨基酸组成,分子式为 C₂₈₃₄H₄₃₄₃N₇₄₃O₇₉₂S₂₁, 分子量为 62 168.42,理论等电点为 8.81。CBDAS 包含 20 种常见氨基酸(表 1),其中疏水性氨基酸占 48.8%,亲水性氨基酸占 51.2%,碱性氨基酸占 13.6%,酸性氨基酸占 9.4%,且含有 21 个含硫氨基酸,说明该蛋白中存在二硫键。由于 CBDAS 序列的 N 末端是 Met,该蛋白估计半衰期为 30 h(哺乳动物网织红细胞,体外)、> 20 h(酵母,体内)、> 10 h(大肠杆菌,体内)。CBDAS 的不稳定指数 II 为 30.57,属于稳定蛋白^[13]。脂肪族氨基酸指数为 88.31。

表 1 CBDA1 编码 CBDAS 的氨基酸组成
Table 1 The amino acid composition of CBDAS encoded by CBDA1

氨基酸 Amino acid	百分比 Ratio//%	氨基酸 Amino acid	百分比 Ratio//%
Ala (A)	4.6	Leu (L)	8.3
Arg (R)	4.0	Lys (K)	6.8
Asn (N)	7.2	Met (M)	2.4
Asp (D)	5.0	Phe (F)	5.5
Cys (C)	1.5	Pro (P)	4.4
Gln (Q)	2.9	Ser (S)	7.0
Glu (E)	4.4	Thr (T)	4.6
Gly (G)	6.8	Trp (W)	2.0
His (H)	2.8	Tyr (Y)	5.0
Ile (I)	8.3	Val (V)	6.6

2.2 CBDA1 基因编码蛋白的亲/疏水性分析 疏水作用能驱动蛋白质的肽链压缩成球状结构,对于维持蛋白质的空间构象十分重要。氨基酸发生变化可导致蛋白质亲/疏水性的改变,而亲/疏水性的变化直接影响蛋白质的结构以及功能。此外,通过了解肽链中不同肽段的疏水性,可以对跨膜蛋白

的跨膜结构域进行预测,为蛋白二级结构的预测及功能结构域的分选提供重要的参考依据。因此,分析蛋白质的亲/疏水性具有十分重要的意义。通过 ProtScale 在线工具对 CBDAS 进行亲/疏水性分析,结果见图 1,在第 15 位氨基酸出现最高值 2.566,即疏水性最强,在第 453 位氨基酸出现最低值 -3.556,即亲水性最强。整体看 CBDAS 的疏水性和亲水性氨基酸分布均衡,但预测结果显示 CBDAS 的亲水性指数平均值(GRAVY,表示蛋白质的溶解度)为 -0.202,所以 CBDAS 更偏向是一个亲水蛋白^[14]。由图 1 可知,在前 29 个氨基酸位置出现一个较强的疏水区域(score > 1.5),且疏水区域较宽,在这个位置有可能出现一个跨膜结构。

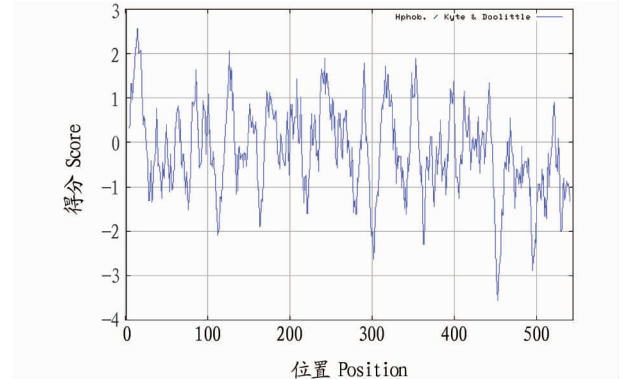


图 1 CBDAS 的亲/疏水性分析

Fig.1 Hydrophilicity and hydrophobicity analysis of CBDAS
2.3 CBDA1 编码蛋白的跨膜结构分析 跨膜结构是蛋白质通过与膜内在蛋白的静电相互作用和氢键键合作用与膜结合的一段氨基酸片段,一般由 20 个左右的疏水氨基酸残基组成,主要形成 α -螺旋。跨膜结构域是膜中蛋白与膜脂相结合的主要部位,固着于细胞膜上起“锚定”作用^[15]。跨膜结构域的预测和分析对于了解蛋白质的结构、功能以及在细胞中的作用部位具有重要意义。在目前的基因组数据中,有 20%~30% 的基因产物被预测为膜蛋白,它们在生物体中担负着多种功能。因此,有效、准确地预测跨膜区和跨膜的方向对指导跨膜蛋白的结构和功能的研究具有重要意义。利用跨膜预测服务器 TMHMM Server v. 2.0 对 CBDAS 进行分析,结果见图 2,该蛋白存在一个潜在的跨膜区(第 1~28 位氨基酸),其中第 1~4 位氨基酸位于膜内,第 5~27 位氨基酸为跨膜的螺旋结构,第 28 位以后的肽链主要在细胞膜外发挥其生物学功能。由于该跨膜结构位于蛋白质的 N 端,推测其极可能为一个信号肽结构。蛋白质序列的其他位置不存在跨膜结构,因此,该蛋白属于跨膜蛋白。

2.4 CBDA1 基因编码蛋白的信号肽分析 信号肽是蛋白质的一个片段,一般由 5~30 个氨基酸残基组成^[16],并大致分为 3 个区段:N 端为带正电荷的氨基酸;中间为由 20 个或更多的以中性氨基酸为主组成的疏水核心区,能够形成一段 α -螺旋;C 端含有小分子氨基酸,是被信号肽酶裂解的部位,亦称加工区。信号肽在蛋白分泌的过程中起重要作用^[17],主要负责引导新合成蛋白质的跨膜、转移和定位,把蛋白质引导到细胞不同的亚细胞器内发挥其生物学功能。

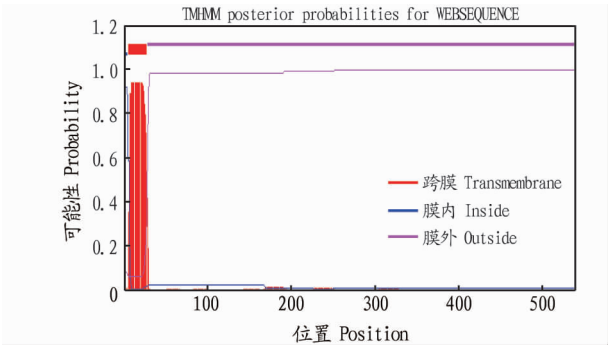


图2 CBDAS 的跨膜结构分析

Fig. 2 Analysis of transmembrane structure of CBDAS
通过 Signal IP 4.1 工具进行分析^[18],结果表明(图3),CBDAS 的 N 末端包含 1 个由 28 个氨基酸残基组成的信号肽,切割位点在第 28 和 29 个氨基酸残基之间,其平均值 S 为 0.801,当平均值 $S > 0.500$ 时,可判断该蛋白为分泌蛋白,说明 CBDAS 是一种分泌蛋白。

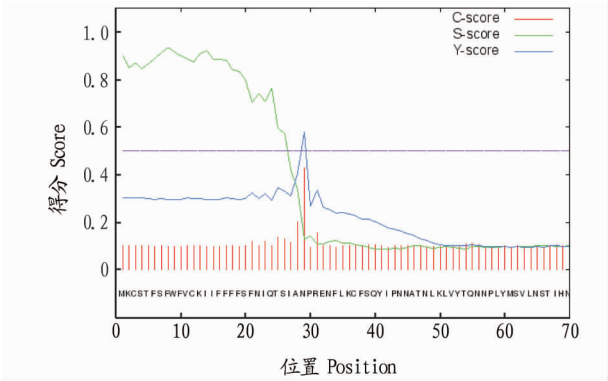


图3 CBDAS 的信号肽位点分析

Fig. 3 Analysis of signal peptide sites in CBDAS
2.5 CBDAS 基因编码蛋白的亚细胞定位 细胞中蛋白质合成后经蛋白质分选信号引导被转运到特定的细胞器中,部分蛋白质则被分泌到细胞外或留在细胞质中,只有转运到正确的部位才能参与细胞的各种生命活动^[19],如果定位发生

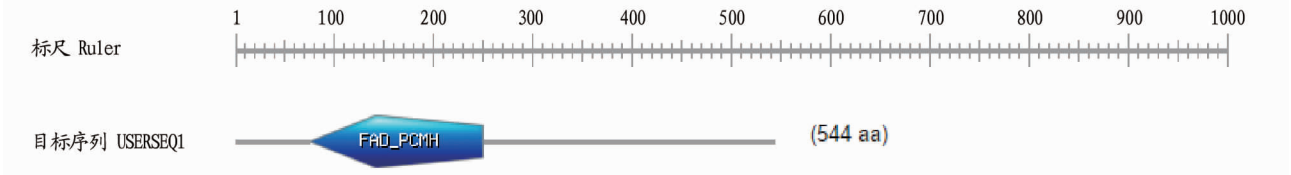


图4 CBDAS 的 motif 分析

Fig. 4 The motif analysis of CBDAS
2.7 CBDAS 基因编码蛋白的保守功能域分析 保守结构域指生物进化或 1 个蛋白家族中不变或相同的结构域,具有重要功能。采用 SMART 工具推测,CBDAS 蛋白中只含有 1 个低复杂度区域 (low complexity region, LCR): GGHF-GGGGYG,位于第 182 ~ 191 位氨基酸。
2.8 CBDAS 蛋白翻译后修饰位点分析 真核生物中的多肽及蛋白质分子经核糖体合成后大多需翻译后修饰,才能确保蛋白质发挥其正常的生物学功能^[23]。常见的蛋白质翻译后修饰有磷酸化和糖基化 2 种。磷酸化是由蛋白质激酶催

偏差,将会对细胞功能甚至生命产生重大影响。了解蛋白质的亚细胞定位信息,可以为推断蛋白质的生物学功能提供必要的帮助,同时对蛋白质的其他研究如相互作用、进化等也能提供必要的信息。利用 ProtComp v. 9.0 对 CBDAS 进行亚细胞定位分析,结果显示,该蛋白质位置的积分预测为细胞外(分泌),得分 9.4,说明该蛋白主要在细胞外发挥其生物学功能。

2.6 CBDAS 基因编码蛋白 motif 分析 PROSITE 数据库收集了生物学有显著意义的蛋白质位点和序列模式,并能根据这些位点和模式快速、可靠地鉴别一个未知功能的蛋白质序列应该属于哪一个蛋白质家族。利用 PROSITE 对 CBDAS 编码蛋白进行 motif 预测,结果如图 4 所示,CBDAS 含有 1 个 FAD - PCMH 结合域,位于第 77 ~ 251 位氨基酸 (TTPK-PLVIVTPSHVSHIQGTILCSKKVGLQIRTRSGGHDSEGmsYISQV--PFVIVDLRNMRSI-KIDVHSQTAWVEAGATLGEVYYWvnEK-NESLSLAAGYCPTVCAGGHFGGGGYGPLM-RSYGLAADNIIDAHLVNVHGVKVL---DRKSMGEDLFWALRGGGAESFGIIVAWKIRLVAV)。CMH 型 FAD 结合结构域是由 2 个 α - β 亚结构域组成:1 个由 α 螺旋包围的 3 个平行的 β 链 (B1 ~ B3) 组成,并被包含在含有 5 个反平行 β 链的第 2 子结构域 (B4 ~ B8)^[20]。2 个子域可以适应它们之间的 FAD 辅因子^[21]。在 PCMH 蛋白中,辅酶 FAD 也共价连接到位于 C 末端催化结构域 FAD 结合结构域之外的酪氨酸^[22]。除 CBDAS 外,目前发现大麻的四氢大麻酚酸合成酶 (THCAS)、细菌 UDP - N - 乙炔烯醇丙酮酰葡萄糖还原酶 (UDP - N - acetylenolpyruvoyl-glucosamine reductase, EC 1.1.1.158)、脊椎动物烷基二羟基丙二酸合酶 (alkyldihydroxyacetonephosphate synthase, EC 2.5.1.26)、真核乳酸脱氢酶 D (D lactate dehydrogenase, EC 1.1.2.4) 和细菌一氧化碳脱氢酶 (Carbon monoxide dehydrogenase, EC 1.2.99.2) 的结构中也含有 PCMH 型 FAD 结合结构域。推测 CBDAS 同 THCAS 一样属于氧化还原酶家族, FDA 是 CBDAS 酶活性的必需辅因子。

化将 ATP 或 GTP γ 位的磷酸基转移到底物蛋白质氨基酸残基 (Ser、Thr、Tyr) 上,是生物体内一种普通的调节方式^[24],蛋白质磷酸化修饰的作用主要体现在以下 3 个方面:一是通过磷酸化修饰改变了受体蛋白质的活性,蛋白质磷酸化或去磷酸化修饰起到开启或关闭蛋白质活性的作用;二是磷酸化蛋白质参与植物体内信号的传导;三是影响蛋白质间的互作,由于在氨基酸残基上结合或失去了磷酸基团,从而改变了受体蛋白质的结构,影响了该受体蛋白质与其他蛋白质间的互作。细胞中蛋白质磷酸化水平是一个动态的变化过程,其细

微差异都可能导致细胞代谢水平上的变化。因此,蛋白质磷酸化对植物生长发育的影响是全方位的。糖基化通常修饰天冬酰胺的 N 端,其氨基酸特征序列为 Asn - X - Ser - Thr (X 是除 Pro 外的任一种类氨基酸)^[25]。N - 糖基化与植物蛋白质正确折叠、细胞凋亡、器官发育及信号转导等生物学功能密切相关^[26]。通常胞外分泌蛋白、膜整合蛋白及构成内膜系统的可溶性驻留蛋白大多需要经过 N - 糖基化修饰。利用 NetPhos 2.0 和 NetNGlyc 1.0 对 CBDAS 进行预测,结果表明该蛋白存在 23 个磷酸化位点、6 个 N - 糖基化位点(表 2、3)。

2.9 CBDA1 基因编码蛋白的二级、三级结构分析 目前最好的单序列预测程序能够达 70% 左右,比如基于 information theory 的 GOR 准确度达 69.7%^[27],利用 GOR IV 对 CBDAS 的二级结构进行预测,结果如图 5 显示,CBDAS 蛋白由 α - 螺旋、β - 折叠和无规卷曲组成,分别占整个肽链的 21.88%、26.29% 和 51.84%。

利用 SWISS-MODEL 蛋白质三维结构建模工具构建的 CBDAS 的三维结构模型,如图 5 所示。建模过程中共有 168 条模板和目标序列相匹配,通过启发式分析过滤得到 29 个模板,主要有 Tetrahydrocannabinolic acid synthase (四氢大麻酚酸合成酶)、Pollen allergen Phl p (花粉过敏原 Phl p)、berberine bridge-forming enzyme (小檗碱桥形成酶)、Reticuline oxidase (纤维素氧化酶)、alkyl dihydroxyacetone phosphate synthase, peroxisomal (烷基二羟基乙酸磷酸酯合成酶,过氧化物酶)。CBDAS 的三级结构也是参考这 29 个模板模拟构建的,其中与 THCAS^[28] 的同源性最高,为 83.95%。

表 2 CBDAS 的预测磷酸化位点

Table 2 Predict phosphorylation sites of CBDAS				
序号 No.	位点 Site	序列 Sequence	得分 Score	预测氨基酸 Predicted amino acid
1	26	NIQTSIANP	0.683	* S *
2	74	NLRFSSDTT	0.940	* S *
3	91	PSHVSHIQG	0.815	* S *
4	111	IRTRSGGHD	0.828	* S *
5	120	SEGMSYISQ	0.980	* S *
6	144	IDVHSQTAW	0.850	* S *
7	170	NESLSLAAG	0.709	* S *
8	221	MDRKSMGED	0.996	* S *
9	254	AVPKSTMFS	0.530	* S *
10	258	STMFSVKKI	0.949	* S *
11	345	CRQLSWIDT	0.998	* S *
12	426	MDISESAI	0.769	* S *
13	78	SSDTPPKPL	0.989	* T *
14	154	EAGATLGEV	0.781	* T *
15	469	YNFMTYPVS	0.888	* T *
16	522	VKVKTLDVP	0.651	* T *
17	41	CFSQYIPNN	0.606	* Y *
18	60	NNPLYMSVL	0.967	* Y *
19	160	GEVYYWVNE	0.623	* Y *
20	368	IKLDYVKKP	0.936	* Y *
21	444	YELWYICSW	0.662	* Y *
22	471	FMTYPVVSQN	0.968	* Y *
23	509	WGEKYFGKN	0.632	* Y *

表 3 CBDAS 的预测糖基化位点

Table 3 Predict glycosylation sites of CBDAS				
序号 No.	位点 Site	序列 Sequence	得分 Score	预测形式 Predicted form
1	45	NATN	0.671	N - Glyc
2	65	NSTI	0.663	N - Glyc
3	296	NITD	0.704	N - Glyc
4	304	NKTA	0.539	N - Glyc
5	328	NKSF	0.506	N - Glyc
6	498	NYTQ	0.583	N - Glyc



图 5 CBDAS 的二级结构

Fig. 5 The secondary structure of CBDAS

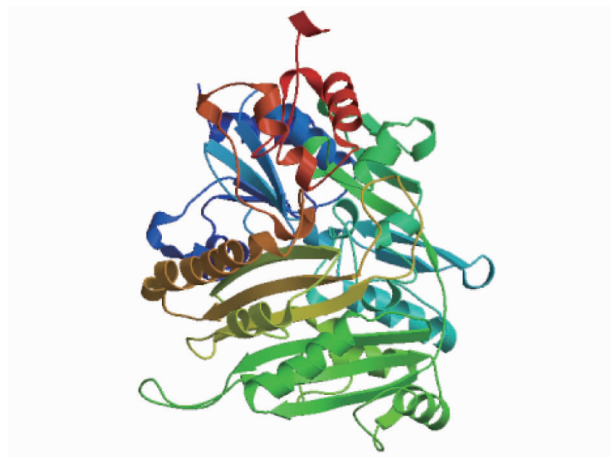


图6 CBDAS 的三级结构模型

Fig.6 The tertiary structure model of CBDAS

3 讨论与结论

利用生物信息学对目的基因进行功能预测是当前国际上研究的热点之一,也是发现和研究新基因的一个重要手段。生物信息学与传统的通过 RT-PCR 方法进行克隆分析基因的方法相比,具有快捷、针对性强、成本低等优点。生物信息学能针对未知功能基因,采集数据,归纳分析,预测基因功能,挖掘基因潜在的研究线索,可为科学研究提供启示和方向指导。对于蛋白质而言,其生物学功能才是最终的研究目的。通过多种生物信息学工具分析 *CBD1* 基因编码的蛋白序列,发现该基因编码 544 个氨基酸,等电点为 8.81, N 端包含 1 个信号肽,而含有信号肽的蛋白质一般都是分泌到细胞外。CBDAS 的亚细胞定位结果也证实了该蛋白是一种稳定的分泌蛋白,主要在胞外发挥其生物学功能。THCAS 的二级结构丰富,包含了 α -螺旋- β -折叠和无规卷曲,含有许多蛋白质修饰及活化位点,如磷酸化位点、糖基化位点、FDA 结合位点等,暗示该蛋白可能在体内受多种因子的调控,具有接受细胞信号并做出反应,实现其生物学功能的潜能。这些结果对正确认识和理解蛋白质结构、定位、功能等均有重要的指导意义。

参考文献

- [1] 梁晓红. 大麻的生物学特性及用途[J]. 现代农业科技, 2014(13): 48-50.
- [2] GAONI Y, MECHOULAM R. Isolation, structure, and partial synthesis of an active constituent of hashish[J]. Journal of the American chemical society, 1964, 86(8): 1646-1647.
- [3] ABRAMS D I, GUZMAN M. Cannabis in cancer care[J]. Clinical pharmacology and therapeutics, 2015, 97(6): 575-586.
- [4] TAURA F, MORIMOTO S, SHOYAMA Y, et al. First direct evidence for the mechanism of $\Delta 1$ -tetrahydrocannabinolic acid biosynthesis[J]. Journal of the American chemical society, 1995, 117(38): 9766-9767.
- [5] SIRIKANTARAMAS S, MORIMOTO S, SHOYAMA Y, et al. The gene controlling marijuana psychoactivity[J]. Journal of biological chemistry, 2004, 279(38): 39767-39774.
- [6] SHOYAMA Y, TAMADA T, KURIHARA K, et al. Structure and function of $\Delta 1$ -tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling

- the psychoactivity of *Cannabis sativa* [J]. Journal of molecular biology, 2012, 423(1): 96-105.
- [7] KLEIN B D, JACOBSON C A, METCALF C S, et al. Evaluation of cannabidiol in animal seizure models by the Epilepsy Therapy Screening Program (ETSP) [J]. Neurochemical research, 2017, 42(7): 1939-1948.
- [8] BORRELLI F, AVIELLO G, ROMANO B, et al. Cannabidiol, a safe and non-psychotropic ingredient of the marijuana plant *Cannabis sativa*, is protective in a murine model of colitis [J]. Journal of molecular medicine, 2009, 87: 1111-1121.
- [9] SYED Y Y, MCKEAGE K, SCOTT L J. Delta-9-tetrahydrocannabinol/cannabidiol (Sativex®): A review of its use in patients with moderate to severe spasticity due to multiple sclerosis [J]. Botany and biotechnology, 2014, 74(5): 563-578.
- [10] BLESSING E M, STEENKAMP M M, MANZANARES J, et al. Cannabidiol as a potential treatment for anxiety disorders [J]. Neurotherapeutics, 2015, 12(4): 825-836.
- [11] TAURA F, MORIMOTO S, SHOYAMA Y, et al. Purification and characterization of cannabidiolic-acid from *Cannabis sativa* L. [J]. Journal of biological chemistry, 1996, 271(29): 17411-17416.
- [12] TAURA F, SIRIKANTARAMAS S, SHOYAMA Y, et al. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa* [J]. FEBS letters, 2007, 581(16): 2929-2934.
- [13] 于欣, 杨震, 楚元奎, 等. IL-6 基因结构和功能生物信息学预测 [J]. 国际检验医学杂志, 2016, 37(21): 2959-2960, 2963.
- [14] 丁帅, 熊勇, 李正涛, 等. 菊花 *rbcl* 基因电子克隆及生物信息学、适应性进化分析 [J]. 种子, 2015, 34(10): 24-30.
- [15] KROGH A, LARSSON B, VON HEIJNE G, et al. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes [J]. Journal of molecular biology, 2001, 305(3): 567-580.
- [16] YAN S, WU G. Signal peptide of cellulase [J]. Applied microbiology and biotechnology, 2014, 98(12): 5329-5362.
- [17] IZARD J W, DOUGHTY M B, KENDALL D A. Physical and conformational properties of synthetic idealized signal sequences parallel their biological functional [J]. Biochemistry, 1995, 34(31): 9904-9912.
- [18] PETERSEN T N, BRUNAK S, VON HEIJNE G, et al. SignalP 4.0: Discriminating signal peptides from transmembrane regions [J]. Nature methods, 2011, 8(10): 785-786.
- [19] 张松, 黄波, 复学峰, 等. 蛋白质亚细胞定位的生物信息学研究 [J]. 生物化学与生物物理进展, 2007, 34(6): 573-579.
- [20] CUNANE L M, CHEN Z W, SHAMALA N, et al. Structures of the flavocytochrome p-cresol methylhydroxylase and its enzyme-substrate complex: Gated substrate entry and proton relays support the proposed catalytic mechanism [J]. Journal of molecular biology, 2000, 295(2): 357-374.
- [21] FRAAIJE M W, MATTEVI A. Flavoenzymes: Diverse catalysts with recurrent features [J]. Trends in biochemical sciences, 2000, 25(3): 126-132.
- [22] MCINTIRE W, EDMONDSON D E, HOPPER D J, et al. 8 alpha-(O-Tyrosyl)flavin adenine dinucleotide, the prosthetic group of bacterial p-cresol methylhydroxylase [J]. Biochemistry, 1981, 20(11): 3068-3075.
- [23] EISENHABER B, EISENHABER F. Prediction of posttranslational modification of proteins from their amino acid sequence [J]. Methods in molecular biology, 2010, 609: 365-384.
- [24] 刘秋林, 钟月仙, 万伟峰, 等. 植物磷酸化蛋白质组学研究进展 [J]. 福建农林大学学报(自然科学版), 2015, 44(3): 225-231.
- [25] 叶强, 金晓琴, 刘伟娜, 等. 植物蛋白质 N-糖基化修饰研究进展 [J]. 浙江师范大学学报(自然科学版), 2016, 39(1): 80-86.
- [26] LEROUGE P, CABANES-MACHETEAU M, RAYON C, et al. N-Glycoprotein biosynthesis in plants: Recent developments and future trends [J]. Plant molecular biology, 1998, 38(1/2): 31-48.
- [27] SEN T Z, JERNIGAN R L, GARNIER J, et al. GOR V server for protein secondary structure prediction [J]. Bioinformatics, 2005, 21(11): 2787-2788.
- [28] SHOYAMA Y, TAMADA T, KURIHARA K, et al. Structure and function of $\Delta 1$ -tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of *Cannabis sativa* [J]. Journal of molecular biology, 2012, 423(1): 96-105.