

## 研究论文

### Research Article

# *Cupriavidus campinensis* BJ71 菌株 2,4-D 降解基因 *tfdA* 生物信息学分析

韩丽珍<sup>1,2</sup> 李翠翠<sup>2,3</sup> 赵德刚<sup>1,2,3\*</sup>

1 贵州大学绿色农药与农业生物工程国家重点实验室培育基地, 贵阳, 550025; 2 贵州大学生命科学学院山地植物保护与种质创新省部共建教育部重点实验室, 贵阳, 550025; 3 贵州大学贵州省农业生物工程重点实验室, 贵阳, 550025

\* 通讯作者, dgzhao@gzu.edu.cn

**摘要** 本文以 2,4-二氯苯氧乙酸高效降解菌 *Cupriavidus campinensis* BJ71 菌株中克隆的 *tfdA* 基因全序列为研究对象, 利用生物信息学方法分析该基因及其推测出的编码蛋白理化性质、蛋白的亲水性或疏水性及亚细胞定位、预测信号肽、跨膜结构域、蛋白二级结构及模体、三级结构等特性。结果表明, BJ71 菌株的 *tfdA* 基因开放阅读框全长 864 bp, 系统发育树表明其属于 族 *tfdA* 基因。TfdA 蛋白包含 287 个氨基酸, 分子量为 32 kD、理论等电点为 6.01, 是一种不含信号肽、定位于细胞质中的可溶性亲水蛋白。 $\alpha$  螺旋和无规卷曲是该蛋白主要的二级结构元件。用 RaptorX 进行三级结构建模、Ramachandran 评估显示获得了合理可靠的 TfdA 蛋白结构模型。

**关键词** 2,4-D 降解基因, *tfdA*, 系统发育树, 生物信息学分析

## Bioinformatics Analysis of 2, 4-D-Degrading Gene *tfdA* of *Cupriavidus campinensis* BJ71

Han Lizhen<sup>1,2</sup> Li Cuicui<sup>2,3</sup> Zhao Degang<sup>1,2,3\*</sup>

1 Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Guiyang, 550025; 2 Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Guizhou University), Ministry of Education, College of Life Sciences, Guizhou University, Guiyang, 550025; 3 Guizhou Key Lab of Agro-Bioengineering, Guizhou University, Guiyang, 550025

\* Corresponding author, dgzhao@gzu.edu.cn

DOI: 10.13417/j.gab.033.000478

**Abstract** Nucleotide and deduced amino acid sequence of *tfdA* gene, which cloned from an effective 2,4-D-degrading bacteria strain BJ71 identified as *Cupriavidus campinensis*, were assessed by using bioinformatic methods, which including analysis of physical and chemical characteristics, hydrophobicity or hydrophilicity, and subcellular localization, prediction of signal peptide, transmembrane spanning domain, secondary structure and tertiary structure according to protein sequence. The results showed that *tfdA* gene ORF was 864 bp, ascribed as Class *tfdA* gene according to analysis of phylogenetic tree. It encoded 287 amino acid polypeptide with a molecular weight of 32 kD and the isoelectric point of 6.01. It had no signal peptide and transmembrane region, and was a hydrophilic soluble protein located in the cytoplasm. Alpha helix and random coil were main secondary structure components. Tertiary structure was also obtained using RaptorX and Ramachandran analysis in Molprobity software.

**Keywords** 2,4-D-degrading gene, *tfdA*, Phylogenetic tree, Bioinformatical analysis

高浓度 2,4-二氯苯氧乙酸(2,4-D)作为一种防 业化生产以来已被广泛应用了六十多年(韩丽珍和赵  
除阔叶杂草的选择性除草剂,自 20 世纪 40 年代工 德刚, 2012)。尽管 2,4-D 的半衰期较短,但大量使用

基金项目:本研究由国家转基因生物新品种培育重大专项子课题《安全转基因技术研究-转基因农作物的“基因删除”技术和“基因拆分”技术》(2013ZX08010-003)、国家转基因生物新品种培育重大专项《安全转基因技术研究》子项目《保障转基因生物安全的外源基因清除技术研究》(2014ZX0801008B-002)和贵州省科技厅基金项目《除草剂 2,4-D 降解基因 *tfdA* 的克隆及降解机制研究》(黔科合 J 字[2013]2123)共同资助

有可能在土壤中蓄积、通过径流进入地表水,甚至渗透进入地下水中造成污染(Cox et al., 2000; Starrett et al., 2000; Johannesen and Aamand, 2003; Farenhorst et al., 2008)。环境中的很多微生物都可以降解或耐受 2,4-D,已经分离到的菌株基于进化和系统发育树被分为 3 群(Kamagata et al., 1997)。第 1 群包含在生长较快的  $\beta$ - 和  $\gamma$ - 变形菌纲中,基于 *tfdA* 基因的差异进一步分为 3 个族, 族以澳大利亚分离到的 *Cupriavidus necator* (原称为 *Alcagenes eutrophus*, *Ralstonia eutropha*, *Wautersia eutropha*) JMP134 菌株 *tfdA* 基因为代表, 族以 *Burkholderia* sp. RASC 菌株 *tfdA* 基因为代表, 族以 *Burkholderia cepacia* 2a 菌株 *tfdA* 基因为代表(McGowan et al., 1998); 第 2 群和第 3 群含有 *tfdA*-like 基因, 通常降解能力较弱(Kitagawa et al., 2002; Itoh et al., 2004; Zaprasis et al., 2010)。其中研究较为清楚的是 *C. necator* JMP134 菌株的 *tfdA* 基因, 编码产物为 Fe( )/ $\alpha$ - 酮戊二酸依赖性双加氧酶 TfdA, 与来源于 *Escherichia coli* 牛磺酸双加氧酶 TauD 同属于第二群  $\alpha$ - 酮戊二酸依赖性双加氧酶家族成员(Hogan et al., 2000; Müller et al., 2006)。研究表明, TfdA 蛋白在代谢 2, 4-D 为 2, 4-二氯苯酚(2, 4-DCP)和乙醛酸的同时,  $\alpha$ - 酮戊二酸被转变为琥珀酸(Fukumori and Hausinger, 1993)。Elkins 等(2002)利用 *E. coli* 牛磺酸双加氧酶 TauD 和肉桂酸合成酶 CAS1 的晶体结构(PDB ID: 1gqw, 1DS1)为模型预测了 *C. necator* JMP134 菌株 TfdA 蛋白的三

级结构,但后期的研究表明来源于 *E. coli* 和 *Pseudomonas putida* 的 TauD 并非二聚体结构,而是四聚体的晶体结构(Knauer et al., 2012)。

在前期研究中,我们从 2, 4-D 污染的麦田土壤中分离筛选到一株 2, 4-D 降解能力较强的 *Cupriavidus campinensis* BJ71 菌株,并克隆了存在于质粒上的 2, 4-D 降解基因 *tfdA* 全长序列,初步分析发现与 *C. necator* JMP134 菌株 *tfdA* 基因存在一定差异。本研究通过构建系统发育树对获得的 *tfdA* 基因进行分析,通过生物信息学方法对 *tfdA* 编码产物进行预测,可为进一步研究该基因的表达及突变奠定基础。

## 1 结果与分析

### 1.1 *tfdA* 基因保守结构域分析及系统发育树构建

前期克隆的 2,4-D 高效降解菌 *C. campinensis* BJ71 菌株 *tfdA* 基因全长序列 864 bp,通过 Bioedit 软件翻译成长度为 287 个氨基酸序列的蛋白质(图 1)。CD Search 分析表明 *tfdA* 基因编码产物为 2, 4-D/ $\alpha$ - 酮戊二酸依赖性双加氧酶 TfdA, 具有 TauD 多重结构域,属于 CAS 类似的超家族(图 2)。与目前报道的代表性的 族、族、族 *tfdA* 基因以及 *tfdA*-like 基因、*tauD* 基因共同构建系统发育树,结果可见我们从 BJ71 菌株中克隆得到的 2,4-D 降解基因 *tfdA* 与 *A. eutrophus* JMP134 菌株 *tfdA* 聚于一枝,属于 族 *tfdA* 基因(图 3)。该基因编码产物 TfdA 蛋白与 JMP134 菌株 TfdA 相比,存在 27 个氨基酸差异(图 4)。

1	ATG	AGC	GCA	GTT	GCC	AAC	CCT	CTT	CAC	CCT	CTT	TTC	GTC	GGA	GAG	GTC	GAG	GAC	ATC	GAC	CTT	CGA	TGC	69
1	M	S	A	V	A	N	P	L	H	P	L	F	V	G	E	V	E	D	I	D	L	R	C	23
70	CCC	CTG	GGC	CCG	GCC	GAG	GTC	CGG	GAG	ATC	GAA	CGC	CAG	ATG	GAC	GAG	ATG	TCG	GTC	CTG	GTG	TTC	CGC	138
24	P	L	G	P	A	E	V	R	E	I	E	R	Q	M	D	E	M	S	V	L	V	F	R	46
139	GGG	CAA	GCC	CTG	GAC	CAG	GAT	CAG	CAG	ATC	GCC	TTC	GCG	CGC	AAC	TTC	GGG	CCG	CTC	GAA	GGC	GGC	TTC	207
47	G	E	A	L	D	Q	D	Q	Q	I	A	F	A	R	N	F	G	P	L	E	G	G	F	69
208	ATC	AAG	GTC	AAC	CAG	AGG	CCT	TCG	AGA	TTC	AAG	TAC	GCC	GAG	TTG	GCG	GAC	ATC	TCG	AAC	GTG	AGT	CTC	276
70	I	K	V	N	Q	R	P	S	R	F	K	Y	A	E	L	A	D	I	S	N	V	S	L	92
277	GAC	GGC	AAG	GTC	GCG	CAA	GCG	GAC	GCG	GAA	GTC	GTC	GGT	AAT	TTC	GCC	AAC	CAG	CTG	CTG	TGG	CAC	AGC	345
93	D	G	K	V	A	Q	R	D	A	R	E	V	V	G	N	F	A	N	E	L	W	H	S	115
346	GAC	AGC	TCA	TTC	CAG	CAG	CCC	GCC	GCC	GCG	TAT	TCG	ATG	CTT	TCC	GCG	GTG	GTC	GTT	CCC	TCG	TCG	GGC	414
116	D	S	S	F	E	Q	P	A	A	R	Y	S	M	L	S	A	V	V	V	P	S	S	G	138
415	GGC	GAC	ACC	GAG	TTC	TGC	GAC	ATG	CGT	GCG	GCC	TAC	GAC	GCG	CTT	CCC	GCA	GAC	CTG	AAG	GGC	GAA	CTC	483
139	G	D	T	E	F	C	D	M	R	A	A	Y	D	A	L	P	A	D	L	K	A	E	L	161
484	GAA	GGG	CTG	CAG	GCC	GAG	CAC	TAG	GGG	CTG	AAC	TCC	CGC	TTC	CTT	CTG	GGC	GAC	ACC	GAC	TAC	TCC	GAG	552
162	E	G	L	Q	A	E	H	Y	A	L	N	S	R	F	L	L	G	D	T	D	Y	S	E	184
553	GCG	CAA	GCC	AAT	GCC	ATG	CCG	CCG	GTC	AGG	TGG	CGS	CTG	GTT	CGA	ACC	CAT	GCC	CTG	TGG	GGA	CGC	CGC	621
185	A	Q	R	N	A	M	P	P	V	R	W	P	L	V	R	T	H	A	G	S	G	R	R	207
622	TTT	CTC	TTC	ATC	GGT	GCC	CAC	GCG	GGC	CAC	ATC	GAA	GGC	CGC	CCG	GTC	GCC	GAA	GGC	CGG	ATG	CTG	CTC	690
208	F	L	F	I	G	A	H	A	G	H	I	E	G	R	P	V	A	E	G	R	M	L	L	230
691	GCG	GAG	CTG	CTC	GAA	CAC	GCG	ACG	CAA	CGC	CAG	TTC	GTG	TAC	CGC	CAT	CGC	TGG	CAG	GTC	GGG	GAC	CTG	759
231	A	E	L	L	E	H	A	I	Q	R	Q	F	V	Y	R	H	R	W	Q	V	G	D	L	253
760	GTG	ATG	TGG	GAC	AAC	CGT	TGC	GTT	CTC	CAC	CGC	GGG	CGA	CGC	TAC	GAC	ATC	TCG	GCC	AGG	CGC	GAG	CTG	828
254	V	M	W	D	N	R	C	V	L	H	R	G	R	R	Y	D	I	S	A	R	R	E	L	276
829	CGC	CGG	GCG	ACG	ACC	CTG	GAC	GCC	GCC	GTC	GCC	TAG												867
277	R	R	A	T	T	L	D	D	A	V	A	-												287

图 1 *Cupriavidus campinensis* BJ71 菌株 *tfdA* 基因及推导出的编码氨基酸序列

Figure 1 Sequences of *tfdA* gene and deduced amino acid of *Cupriavidus campinensis* BJ71



表 1 *C. campinensis* BJ71 菌株 TfdA 蛋白氨基酸组成  
Table 1 Amino acid composition of TfdA protein of *C. campinensis* BJ71

氨基酸 Amino acid	数量 No.	比例(%) Proportion (%)	氨基酸 Amino acid	数量 No.	比例(%) Proportion (%)
丙氨酸 Ala	33	11.5	精氨酸 Arg	30	10.5
天冬酰胺 Asn	9	3.1	天冬氨酸 Asp	20	7.0
半胱氨酸 Cys	3	1.0	谷氨酰胺 Gln	15	5.2
谷氨酸 Glu	19	6.6	甘氨酸 Gly	20	7.0
组氨酸 His	9	3.1	异亮氨酸 Ile	8	2.8
亮氨酸 Leu	28	9.8	赖氨酸 Lys	4	1.4
甲硫氨酸 Met	8	2.8	苯丙氨酸 Phe	13	4.5
脯氨酸 Pro	13	4.5	丝氨酸 Ser	16	5.6
苏氨酸 Thr	6	2.1	色氨酸 Trp	4	1.4
酪氨酸 Tyr	7	2.4	缬氨酸 Val	22	7.7

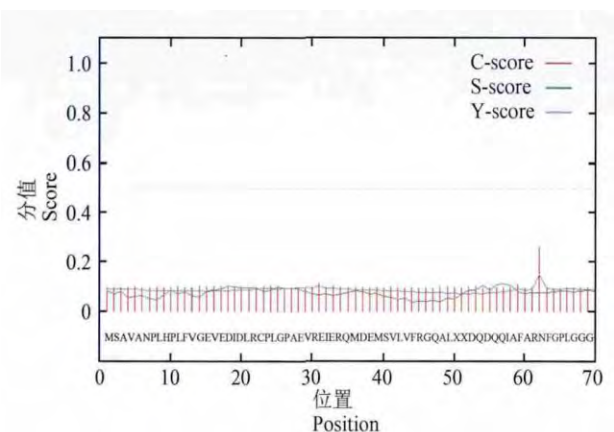


图 5 SignalP 4.1 预测 *C. campinensis* BJ71 菌株 TfdA 蛋白的信号肽  
Figure 5 Signal Peptide analysis of TfdA protein of *C. campinensis* BJ71 by SignalP 4.1

膜蛋白分为胞外区、跨膜部分及胞内区，跨膜区就是蛋白在细胞膜内的部分。利用 TMpred 和 SOSUI 分析跨膜性，结果表明该蛋白不存在跨膜蛋白，是一种可溶性蛋白(图 6)。进一步利用 PSORT 软件对 TfdA 蛋白的细胞内定位进行预测，结果见表 2，显示其定位在细胞质的可能性极大。

疏水作用是蛋白质折叠的主要驱动力，可为蛋白质二级结构预测提供佐证。利用 ProtScale 分析蛋

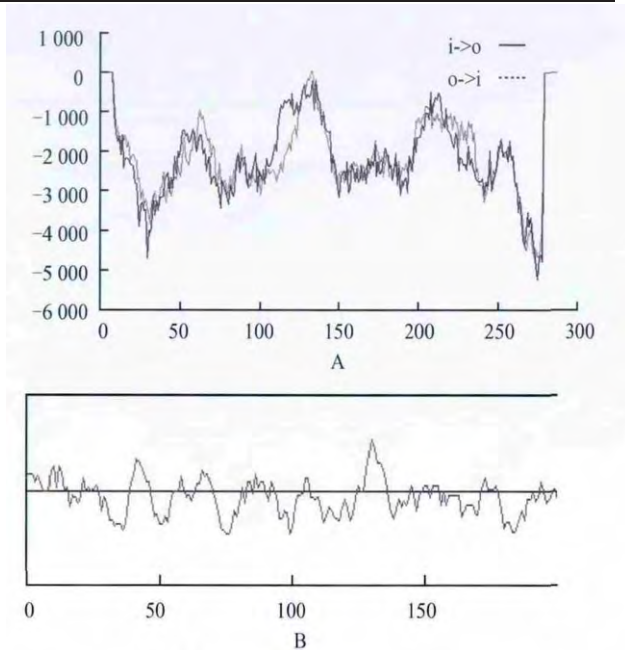


图 6 *C. campinensis* BJ71 菌株 TfdA 蛋白跨膜性分析  
注: A: TMpred 分析结果; B: SOSUI 分析结果  
Figure 6 Transmembrane analysis of TfdA protein of *C. campinensis* BJ71  
Note: A: Analysis by TMpred; B: Analysis by SOSUI

表 2 *C. campinensis* BJ71 菌株 TfdA 蛋白的亚细胞定位分析  
Table 2 Analysis of subcellular location of TfdA protein of *C. campinensis* BJ71

定位 Location	概率 Probability
细胞质 Cytoplasmic	9.97
细胞膜 Cytoplasmic membrane	0.01
细胞周质 Periplasmic	0.01
外膜 Outermembrane	0.00
胞外 Extracellular	0.00

注: 亚细胞定位分析是利用 PSORT 进行  
Note: Analysis of subcellular location by PSORT

白质的亲疏水性。287 个氨基酸组成的 TfdA 蛋白中，亲水氨基酸占 63.4%，而疏水氨基酸占 33.8%，亲水性最强的氨基酸为第 77 位的丝氨酸，其分值最低 (-2.311)，疏水性最强的氨基酸为第 130 位的丝氨酸，其分值最高(1.911)；其亲水性区域较多，预测 TfdA 蛋白为亲水性蛋白(图 7)。这与缺乏疏水性的跨膜结构域、不具跨膜功能的预测相一致。因而 *C. campinensis* BJ71 菌株的 TfdA 蛋白是一个不含信号肽、定位于细胞质中的可溶性亲水蛋白。

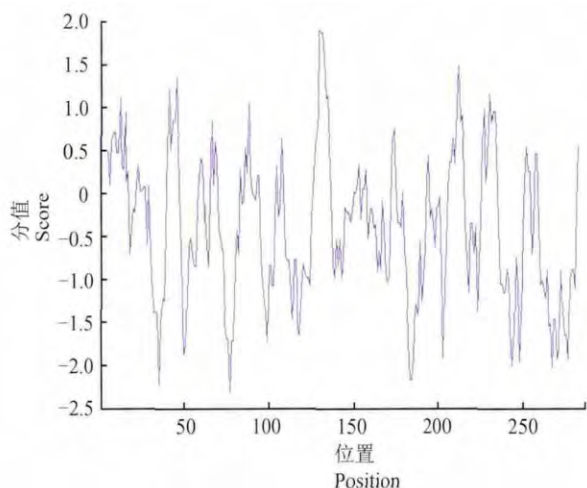


图 7 ProtScale 预测 *C. campinensis* BJ71 菌株 TfdA 蛋白的亲疏水性

Figure 7 Prediction of hydrophobicity or hydrophilicity of TfdA protein of *C. campinensis* BJ71 by ProtScale

#### 1.4 TfdA 蛋白二级结构及模体分析

以 *C. necator* JMP134 菌株的 TfdA 蛋白作为参照 利用 SOPMA 分析 2 个不同菌株 TfdA 蛋白的二级结构。结果表明(表 3) 2 个菌株构成的二级结构元件中  $\alpha$ -螺旋(H)和  $\beta$ -转角(T)比例存在微小差异。我

表 3 2 个菌株 TfdA 蛋白二级结构预测结果

Table 3 Predicted results of secondary structure of TfdA protein from two strains

二级结构元件	$\alpha$ -螺旋(%)	$\beta$ -转角(%)	无规卷曲(%)	延伸链(%)
Elements of secondary structure	$\alpha$ -helix (%)	$\beta$ -turn (%)	Random coil (%)	Extended strand (%)
<i>C. campinensis</i> BJ71 TfdA	31.71	2.79	47.39	18.12
<i>C. necator</i> JMP134 TfdA	30.66	3.83	47.39	18.12

注: 蛋白二级结构预测是利用 SOPMA 进行

Note: Predicted results of secondary structure by SOPMA

表 4 *C. campinensis* BJ71 菌株 TfdA 蛋白模体分析

Table 4 Motifs in TfdA protein of *C. campinensis* BJ71

模体(注释)	TfdA 氨基酸位置	TfdA 氨基酸残基
Motifs (annotation)	Sites of TfdA amino acid	Amino acid
N-糖基化位点 ASN_GLYCOSYLATION	89~92	NVSL
cAMP & cGMP 依赖性蛋白激酶磷酸化位点 CAMP_PHOSPHO_SITE	277~280	RRAT
蛋白激酶 C 磷酸化位点 PKC_PHOSPHO_SITE	204~206	SGR
	238~240	TQR
	271~273	SAR
酪蛋白激酶 磷酸化位点 CK2_PHOSPHO_SITE	137~140	SGGD
	280~283	TTLD
N-豆蔻酰基化位点 MYRISTYL	106~111	GNFANQ
酰胺化位点 AMIDATION	212~217	GAHAGH
	204~207	SGRR
	264~267	RGRR

注: TfdA 蛋白模体分析利用 Prosite 进行

Note: Motifs in TfdA protein by Prosite

们克隆获得的 BJ71 菌株 TfdA 蛋白  $\alpha$ -螺旋为 91 个,  $\beta$ -转角为 8 个, 无规卷曲(C)为 136 个, 延伸链(E)为 52 个  $\alpha$ -螺旋和无规卷曲是 TfdA 蛋白整体结构中的主要组成结构元件(图 8), 有利于稳定蛋白质的结构。

运用 PredictProtein 预测该蛋白是否有二硫键 (Ceroni et al., 2006), 结果显示, 该蛋白没有二硫键形成; Prosite 进行功能位点的分析 (Bairoch et al., 1997), 发现有 1 个糖基化位点、6 个磷酸化位点、2 个酰基化位点和 2 个酰胺化位点(表 4)。蛋白质翻译后修饰在酶和其它的重要功能分子活性的发挥、

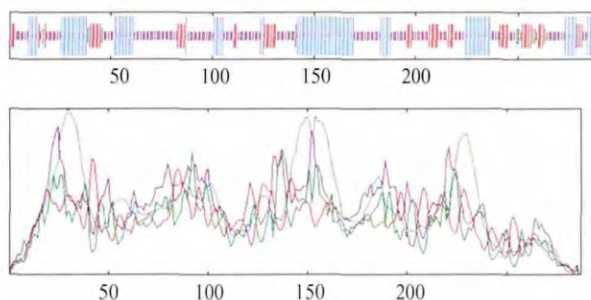


图 8 SOMPA 分析 *C. campinensis* BJ71 菌株 TfdA 蛋白的二级结构

Figure 8 Analysis of secondary structure of TfdA protein of *C. campinensis* BJ71 by SOMPA

第二信使传递和酶的级联作用中起到重要的作用,其中,蛋白质磷酸化是蛋白质翻译后修饰的重要内容,显然,该蛋白磷酸化位点数明显高于糖基化位点。

### 1.5 TfdA 蛋白卷曲螺旋域分析

卷曲螺旋是由 2~5 条  $\alpha$  螺旋相互缠绕而形成的平行或反平行超螺旋结构,含有卷曲螺旋结构的蛋白主要是一些转录因子、结构蛋白、膜蛋白和酶等(魏香等, 2004)。亮氨酸拉链即为典型的卷曲结构。虽然  $\alpha$  螺旋在该蛋白中的比例较大(31.71%),但 COILS 预测结果显示并不含卷曲螺旋域,也不形成亮氨酸拉链(图 9)。

### 1.6 TfdA 蛋白三级结构预测及分析

利用 RaptorX 软件进行三级结构预测,结果显示可以 PDB 数据库中 3pvjA、1gqwA 和 1oihA 为模板,最优模板为 3pvjA,将具有 27 个氨基酸差异的 BJ71 菌株 TfdA 蛋白以及 *C. necator* JMP134 TfdA 蛋白(NCBI 登录号: YP\_025400)同时建模,三级结构如图 10。可以看出 2 个蛋白的三级结构极其相似,但 JMP134 菌株的 TfdA 蛋白似乎具有更多的  $\beta$ -转角(图中箭头所示);前面利用 SOPMA 预测 JMP134 菌株 TfdA 蛋白二级结构的结果也表明  $\alpha$ -螺旋、 $\beta$ -转角比例分别略低于和高于 BJ71 菌株的 TfdA 蛋白,这也与预测的三级结构吻合。使用 MolProbity 软件中 Ramachandran 分析 BJ71 菌株 TfdA 的三级结构(图 11),显示构建的 TfdA 三级结构模型有 92.3% (263/285)的氨基酸位点都位于最佳区域内,有 97.9% (279/285)的氨基酸位点位于允许区域范围内,说明本次模型结构可靠准确。

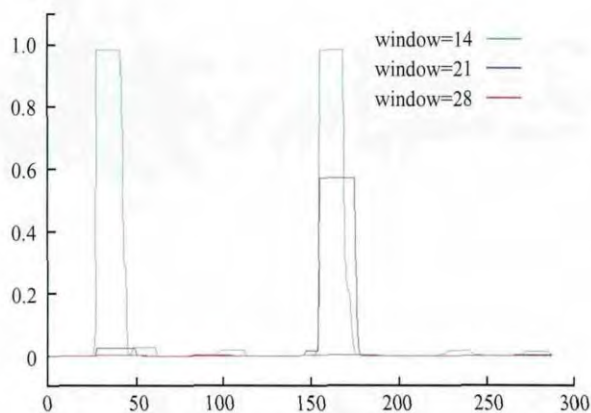


图 9 COILS 分析 *C. campinensis* BJ71 菌株 TfdA 蛋白的卷曲螺旋域

Figure 9 COILS analysis of TfdA protein of *C. campinensis* BJ71

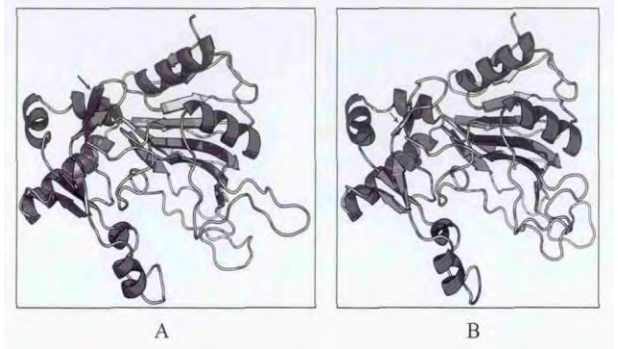


图 10 RaptorX 预测 TfdA 蛋白三级结构

注: A: *C. campinensis* BJ71; B: *C. necator* JMP134; 箭头所示结构差异

Figure 10 Tertiary structure of TfdA predicted by RaptorX

Note: A: *C. campinensis* BJ71; B: *C. necator* JMP134; arrow indicated structure difference

## 2 讨论

借助于生物信息学手段,我们对从 2,4-D 高效降解菌 *C. campinensis* BJ71 菌株中克隆获得的 2,4-D 降解基因 *tfdA* 及其编码产物进行了一系列分析。结果表明该基因全长 864 bp、属于  $\alpha$ -酮戊二酸依赖性双加氧酶 TfdA 蛋白,分子量为 32 KD、理论等电点 pI 为 6.01。该蛋白不含信号肽和跨膜结构域,是一种定位于细胞质中的可溶性亲水蛋白。 $\alpha$  螺旋和无规卷曲是其二级结构的主要组成元件,内部存在一系列磷酸化位点和较少的糖基化、酰基化和酰胺化位点。

蛋白质三级结构预测的方法通常有同源建模法、折叠识别法和从头预测法。同源建模法(homology modeling)是基于待建模蛋白在模板库中可找到序列较相似的蛋白而进行预测的,又称为比较建模(comparative modeling),其代表软件为 SWISS-MODEL,有研究显示如果序列相似性高于 75%则预测准确度较高,低于 40%则需要手动调整(Schwede et al., 2003; Arnold et al., 2006)。从头计算方法(ab initio)的理论依据是依据热力学理论,即求蛋白质能量最小的状态,但该方法并不实用,目前仅能计算低于 120 个氨基酸形成的结构。针对较低序列相似性的折叠识别法(fold recognition)又称为穿线法(threading method),以加拿大滑铁卢大学的 RAPTOR 为代表,该方法在 CASP (critical assessment of protein structure prediction)评估中被评价为最好的结构预测软件,可以有效优化蛋白穿线打分功能(Xu and Li, 2003)。与老版本的 RAPTOR 程序相比,RaptorX 有 3 个提高,第一



图 11 MolProbity Ramachandran 分析 *C. campinensis* BJ71 菌株 TfdA 蛋白三级结构

Figure 11 MolProbity Ramachandran analysis of tertiary structure of TfdA protein of *C. campinensis* BJ71 strain

是考虑了蛋白特征之间的相关性；第二是通过借力结构信息处理相似性不高的蛋白；第三是利用了新的多重模板打分运算以扩展模板结构相似性和提高比对准确性。这种基于多个模板进行折叠识别的特征，大大提高了预测结构的准确性(Peng and Xu, 2009; 2010; 2011; Källberg et al., 2012)。

本研究中，我们试图利用 SWISS-MODEL 进行同源建模，但发现与 PDB 数据库中的模板同源性仅有 30.822%，因而并不适合采用该方法建模，而 RaptorX 适合于同源性较低的蛋白序列三级结构的预测，应用该方法进行蛋白三维结构预测、经 Ramachandran 分析获得的结构合理可靠。Knauer 等 (2012) 的研究表明，*P. putida* KT2440 与 *E. coli* 的 TauD 晶体结构(PDB ID: 3pvjA, 1gqw)均为四聚体。以 PDB 数据库中 3pvjA 为模板，利用 RaptorX 进行建模，发现 *C. campinensis* BJ71 菌株 TfdA 蛋白与 *C. necator* JMP134 菌株的三级结构非常相似，但  $\alpha$ -螺旋和  $\beta$ -转角的比例略有不同，可能会造成 2 个 TfdA 蛋白在 2,4-D 降解能力上存在一定的差异。本研究可为该蛋白的原核表达、设计突变位点进行突变研究奠定基础。

### 3 材料与方法

#### 3.1 材料

分析用的 2,4-D 降解基因 *tfdA* (GenBank 登录号: KJ028765)核苷酸序列是从本实验室前期分离筛选到的 2,4-D 高效降解菌 *Cupriavidus campinensis* BJ71 菌株(保藏号: CCTCC M 2014006)中克隆得到的。

#### 3.2 *tfdA* 基因编码产物保守域及系统发生分析

利用 Bioedit 将克隆获得的 *tfdA* 基因核苷酸序列翻译为蛋白质序列，利用 NCBI 网站上 CD Search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)分析编码产物保守结构域，选择 *Alcaligenes eutrophus* *tfdA* (M16730)、*Burkholderia* sp. RASC *tfdA* (U25717)、*Burkholderia cepacia* 2a *tfdA* (AF029344)分别为 Class 、 、 *tfdA* 基因的代表，*Alphaproteobacterium* sp. HW13 *tfdA*  $\alpha$  (AB074492)为 *tfdA*-like 基因的代表，以 *Escherichia coli* str. K-12 substr. MG1665 *tauD* (NC\_000913)为外群，利用 ClustalW 2.0 (Larkin et al., 2007)进行核苷酸多重序列比对，用 MEGA5.0 软件采用邻近法(Neighbor-Joining)进行系统发育分析、构建系统发育树(Tamura et al., 2011)，通过重复取样 1 000 次

进行自展值(Bootstrap)分析来评估系统发育树的拓扑结构稳定性。

### 3.3 *tfdA* 基因编码产物的基本理化性质分析

在 ExPASy 网站上利用在线软件 ProtParam (<http://www.expasy.org/tools/protparam.html>)分析 *tfdA* 基因编码产物的基本理化性质(Gasteiger et al., 2005)。

### 3.4 *tfdA* 基因编码产物信号肽、跨膜区、定位及亲疏水性分析

利用 SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP>) (Petersen et al., 2011)预测 *tfdA* 基因编码产物有无信号肽序列; 利用 TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)) (Hofmann and Stoffel, 1993)分析蛋白跨膜区、SOSUI (<http://bp.nuap.nagoya-u.ac.jp/sosui/>) (Hirokawa et al., 1998)分析是否为可溶蛋白; 利用 PSORT (<http://psort.hgc.jp>)进行蛋白定位分析; 利用 ExPASy 网站上 ProtScale (<http://ca.expasy.org/tools/protscale.html>) (Gasteiger et al., 2005)分析亲疏水性。

### 3.5 *tfdA* 基因编码产物二级结构及模体分析

蛋白质二级结构是指多肽链有规则重复的构象,其结构组件为 $\alpha$ 螺旋、 $\beta$ 转角、无规卷曲和延伸链。利用 PBIL-IBCP Lyon-Gerland 网站上 SOMPA ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_sopma.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)) (Geourjon and Deléage, 1995)进行二级结构预测; 利用 PredictProtein(<http://www.predictprotein.org/>) (Rost et al., 2004)进行功能预测、二硫键及结构域; 利用 COILS ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) (Lupas et al., 1991)分析蛋白有无卷曲螺旋域结构。

### 3.6 *tfdA* 基因编码产物三级结构模拟

利用在线软件 RaptorX (<http://raptorx.uchicago.edu>) (Källberg et al., 2012)进行三级结构预测; 利用 MolProbity 软件 (<http://molprobity.biochem.duke.edu>; <http://kinemage.biochem.duke.edu>) (Davis et al., 2007; Chen et al., 2010)评估建模模型。

### 作者贡献

韩丽珍负责基因克隆、分析与论文撰写,李翠翠负责资料收集并参与分析,责任作者赵德刚负责课题整体设计、论文修改和指导。

### 致谢

本研究由国家转基因生物新品种培育重大专项子课题《安全转基因技术研究 - 转基因农作物的“基因删除”技术和“基因拆分”技术》(2013ZX08010-003)、国家转基因生物新品种培育重大专项《安全转基因技术研究》子项目《保障转基因生物安全的外源基因清除技术研究》(2014ZX0801008B-002)和贵州省科技厅基金项目《除草剂 2,4-D 降解基因 *tfdA* 的克隆及降解机制研究》(黔科合 J 字[2013]2123)共同资助。

### 参考文献

- Arnold K., Bordoli L., Kopp J., and Schwede T., 2006, The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling, *Bioinformatics*, 22 (2): 195-201
- Chen V.B., Arendall W.B., Headd J.J., Keedy D.A., Immormino R.M., Kapral G.J., Murray L.W., Richardson J.S., and Richardson D.C., 2010, Molprobity: All-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol. Crystallogr*, D66: 12-21
- Cox L., Celis R., Hermosin M.C., and Cornejo J., 2000, Natural soil colloids to retard simazine and 2,4-D leaching in soil, *J. Agric. Food Chem.*, 48(1): 93-99
- Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J., Wang X., Murray L.W., Arendall W.B., Snoeyink J., Richardson J.S., and Richardson D.C., 2007, MolProbity: All-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res.*, 35: W375-W383
- Elkins J.M., Ryle M.J., Clifton I.J., Hotopp J.C.D., Lloyd J.S., Burzlaff N.I., Baldwin J.E., Hausinger R.P., and Roach P.L., 2002, X-ray crystal structure of *Escherichia coli* Taurine/ $\alpha$ -ketoglutarate dioxygenase complexed to ferrous iron and substrates, *Biochemistry*, 41(16): 5185-5192
- Farenhorst A., Londry K.L., Nahar N., and Gaultier J., 2008, In-field variation in 2,4-D mineralization in relation to sorption and soil microbial communities, *J. Environ. Sci. Health B*, 43: 113-119
- Fukumori F., and Hausinger R.P., 1993, *Alcaligenes eutrophus* JMP134 “2,4-dichlorophenoxyacetate monooxygenase” is an  $\alpha$ -ketoglutarate-dependent dioxygenase, *J. Bacteriol.*, 175: 2083-2086
- Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and Bairoch A., 2005, Protein identification and analysis tools on the ExPASy Server, In: Walker J.M. (ed.), *The proteomics protocols handbook*, Humana Press, pp. 571-607



- Geourjon C., and Deléage G., 1995, SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Comput. Appl. Biosci.*, 11(6): 681-684
- Han L.Z., and Zhao D.G., 2012, Research progress of microbial degradation of the herbicide 2,4-dichlorophenoxyacetic acid, *Nongyao (Agrochemicals)*, 51(10): 710-714, 719 (韩丽珍, 赵德刚, 2012, 除草剂 2,4-滴微生物降解研究进展, *农药*, 51(10): 710-714, 719)
- Hirokawa T., Boon-Chieng S., and Mitaku S., 1998, SOSUI: Classification and secondary structure prediction system for membrane proteins, *Bioinformatics*, 14(4): 378-379
- Hofmann K., and Stoffel W., 1993, TMbase-A database of membrane spanning proteins segments, *Biol. Chem. Hoppe-Seyler*, 374: 166
- Hogan D.A., Smith S.R., Saari E.A., McCracken J., and Hausinger R.P., 2000, Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/ $\alpha$ -ketoglutarate dioxygenase, *J. Biol. Chem.*, 275(17): 12400-12409
- Itoh K., Tashiro Y., Uobe K., Kamagata Y., Suyama K., and Yamamoto H., 2004, Root nodule *Bradyrhizobium* spp. harbor *tfdA*  $\alpha$  and *cadA*, homologous with genes encoding 2,4-dichlorophenoxyacetic acid-degrading proteins, *Appl. Environ. Microbiol.*, 70(4): 2110-2118
- Johannesen H., and Aamand J., 2003, Mineralization of aged atrazine, terbutylazine, 2,4-D, and mecoprop in soil and aquifer, *Environ. Toxicol. Chem.*, 22: 722-729
- Källberg M., Wang H., Wang S., Peng J., Wang Z., Lu H., and Xu J., 2012, Template-based protein structure modeling using the RaptorX web server, *Nat. Protoc.*, 7(8): 1511-1522
- Kamagata Y., Fulthorpe R.R., Tamura K., Takami H., Forney L. J., and Tiedje J.M., 1997, Pristine environments harbor a new group of oligotrophic 2,4-dichlorophenoxyacetic acid-degrading bacteria, *Appl. Environ. Microbiol.*, 63(6): 2266-2272
- Kitagawa W., Takami S., Miyauchi K., Masai E., Kamagata Y., Tiedje J.M., and Fukuda M., 2002, Novel 2,4-dichlorophenoxyacetic acid degradation genes from oligotrophic *Bradyrhizobium* sp. strain HW13 isolated from a pristine environment, *J. Bacteriol.*, 184(2): 509-518
- Knauer S.H., Hartl-Spiegelhauer O., Schwarzingler S., Hänzelmann P., and Dobbek H., 2012, The Fe ( $\alpha$ )-ketoglutarate-dependent taurine dioxygenases from *Pseudomonas putida* and *Escherichia coli* are tetramers, *FEBS J.*, 279: 816-831
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., and Higgins D. G., 2007, Clustal W and Clustal X version 2.0, *Bioinformatics*, 23(21): 2947-2948
- Lupas a., van Dyke M., and Stock J., 1991, Predicting coiled coils from protein sequences, *Science*, 252(5009): 1162-1164
- McGowan C., Fulthorpe R., Wright A., and Tiedje J.M., 1998, Evidence for interspecies gene transfer in the evolution of 2,4-dichlorophenoxyacetic acid degraders, *Appl. Environ. Microbiol.*, 64(10): 4089-4092
- Müller T.A., Fleischmann T., van der Meer J.R., and Kohler H.P. E., 2006, Purification and characterization of two enantioselective  $\alpha$ -ketoglutarate-dependent dioxygenases, RdpA and SdpA, from *Sphingomonas herbicidovorans* MH, *Appl. Environ. Microbiol.*, 72(7): 4853-4861
- Peng J., and Xu J., 2009, Boosting protein threading accuracy, *Res. Comput. Mol. Biol.*, 5541: 31-45
- Peng J., and Xu J., 2010, Low-homology protein threading, *Bioinformatics*, 26(12): i294-i300
- Peng J., and Xu J., 2011, Raptor X: Exploiting structure information for protein alignment by statistical inference, *Proteins*, 79(Suppl 10): 161-171
- Petersen T.N., Brunak S., von Heijne G., and Nielsen H., 2011, SignalP 4.0: Discriminating signal peptides from transmembrane regions, *Nat. Methods*, 8: 785-786
- Rost B., Yachdav G., and Liu J., 2004, The PredictProtein server, *Nucleic Acids Res.*, 32: W321-W326
- Schwede T., Kopp J., Guex N., and Peitsch M.C., 2003, SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res.*, 31(13): 3381-3385
- Starrett S.K., Christians N.E., and Austin T.A.L., 2000, Movement of herbicides under two irrigation regimes applied to turfgrass, *Adv. Environ. Res.*, 4(2): 169-176
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S., 2011, MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, 28(10): 2731-2739
- Wei X., Zeng X.G., and Zhou H.M., 2004, Progress on the study of coiled coils in protein structures, *Zhongguo Shengwu Huaxue Yu Fenzi Shengwu Xuebao (Chinese Journal of Biochemistry and Molecular Biology)*, 20(5): 565-571 (魏香, 曾宪纲, 周海梦, 2004, 蛋白质结构中卷曲螺旋的研究进展, *中国生物化学与分子生物学报*, 20(5): 565-571)
- Xu J., and Li M., 2003, Assessment of RAPTOR's linear programming approach in CAFASP3, *Proteins*, 53: 579-584
- Zaprasias A., Liu Y.J., Liu S.J., Drake H.L., and Horn M.A., 2010, Abundance of novel and diverse *tfdA*-like genes, encoding putative phenoxyalkanoic acid herbicide-degrading dioxygenases, in soil, *Appl. Environ. Microbiol.*, 76(1): 119-128