

# 茶树咖啡碱合成酶序列和结构分析

---

## Sequence and structure analysis of Tea Caffeine Synthase

报告人：王熠  
中国农业科学院研究生院  
2022级硕士18班

2022年5月15日

# 小组成员

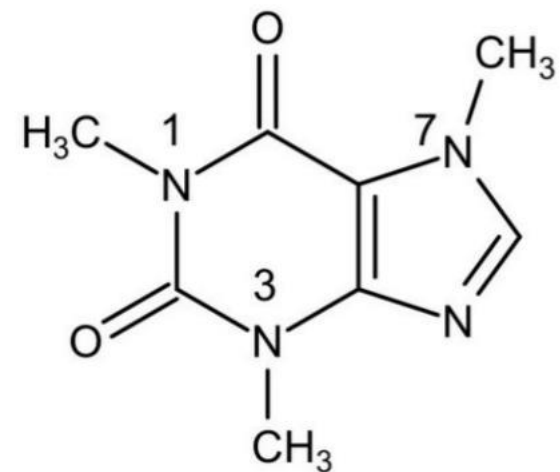
编号	姓名	研究所	导师	研究方向
3G05A	陈慧	茶叶所	付建玉	茶树种质资源与育种
3G05B	夏晴	茶叶所	孙亮	茶树害虫化学通讯和功能
3G05C	张舒然	茶叶所	陈亮	茶树种质资源与育种
3G05D	王熠	茶叶所	金基强	茶树种质资源与育种

# 主要内容

- 1、研究背景
- 2、蛋白质序列分析
- 3、蛋白质结构预测

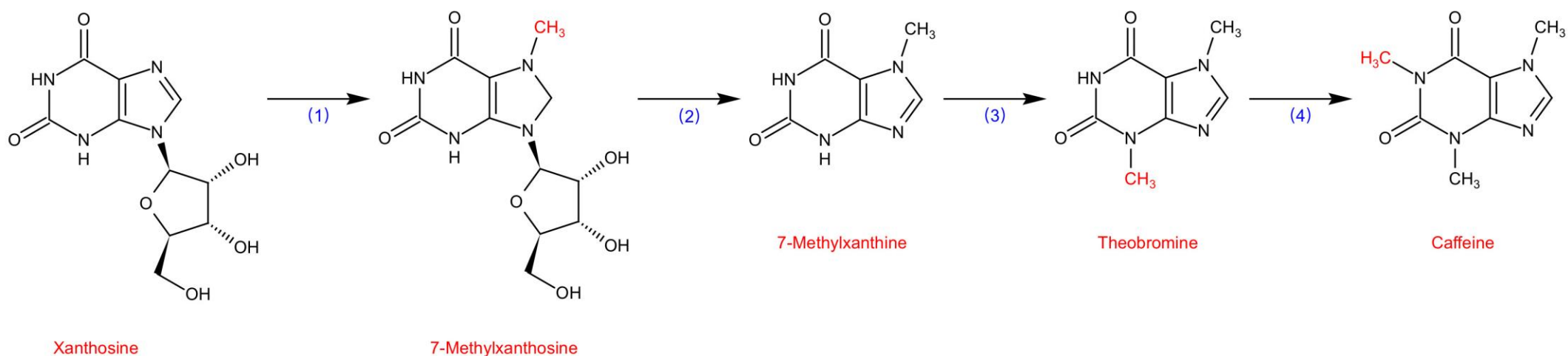
# 1、研究背景

茶、咖啡和可可是世界上最受欢迎的三大无酒精饮料，它们都含有嘌呤生物碱，如咖啡碱、可可碱等。嘌呤生物碱是茶树重要的次级代谢产物，参与一些生物和非生物胁迫。大部分茶树资源咖啡碱含量较高，一般占干物质含量的2-5%。咖啡碱不仅有利于茶叶风味形成，还有很好的保健功效，能够刺激中枢神经系统兴奋，有助于提高记忆和识别能力，还可以用于治疗 and 防御一些疾病。具有如此多生理活性的咖啡碱一直是茶树研究的一个重点和热点。



$C_8H_{10}N_4O_2$   
1,3,7-trimethylxanthine  
Caffeine

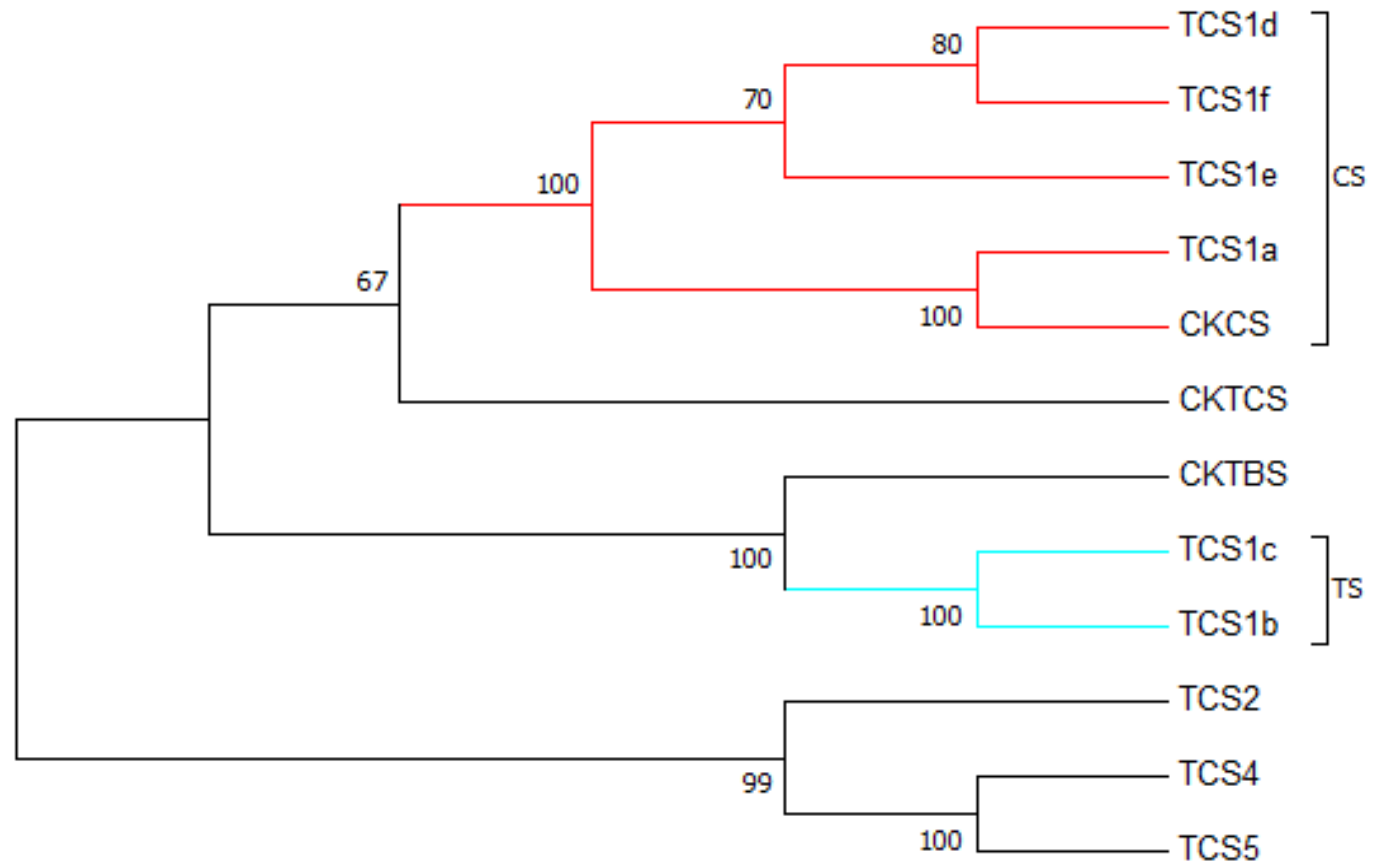
以往研究发现，可以合成咖啡碱等嘌呤生物碱的植物（茶、咖啡和可可）都具有如下图所示的咖啡碱核心合成途径，即由黄嘌呤核苷通过三步甲基化和一步脱核糖反应生成咖啡碱。其中，三步甲基化所需要的氮甲基转移酶（NMT）是咖啡碱等嘌呤生物碱合成的关键酶类。



## 2、蛋白质序列分析

## 2.1、山茶属植物NMTs

现已从山茶属中分离克隆到20条左右的NMT基因。茶树中有多个NMT基因，研究表明茶树中咖啡碱合成的关键基因是*TCS1*，其他的基因无生物学功能或表达水平极低。现已克隆出6种NMT基因（*TCS1-TCS6*），并发现*TCS1*在茶组植物中存在变异，克隆出6种*TCS1*等位基因。





## 2.2、TCS1稀有等位 变异氨基酸序列分析

```
TCS1b -----MGKQNEVLFMNRGEGEISYAQNSAFTQKVASMAMPALENAVETLFSKDFHLLQAL 55
TCS1c -----MGKQNEVLFMNRGEGEISYAQNSAFTQKVASMAMPALENAVETLFSKDFHLLQAL 55
TCS1e ME-LAMGKQNEVLFMNRGEGESSYAQNSSTTQQVASMARPALENAVKTLFSKDFHL-QAL 58
TCS1a MELATAGKQNEVLFMNRGEGESSYAQNSSTTQQVASMAQPALENAVETLFSRDFHL-QAL 59
TCS1d MELATTGKQNEVLFMNRGEGESSYAQNSSTTQQVASMATPALENAVETLFSKDFHL-QAL 59
TCS1f MELATTGKQNEVLFMNRGEGESSYAQNSSTTQQVASMATLALENAVETLFSKDFHL-QAL 59
*****:***:***:**** *****:**** **

TCS1b TAADLGAAGPNTFAVISTIKRMMEKKCRELYCQTELEQVYLNDLFGNDFNTLTKGLSSE 115
TCS1c TAADLGAAGPNTFAVISTIKRMMEKKCRELYCQTELEQVYLNDLFGNDFNTLTKGLSSQ 115
TCS1e NAADLGAAGPNTFAVISTIKRMMEKKCRELNCQTELEQVYLNDLFGNDFNTLTKGLSSE 118
TCS1a NAADLGAAGPNTFAVISTIKRMMEKKCRELNCQTELEQVYLNDLFGNDFNTLTKGLSSE 119
TCS1d NAVDLGAAGPNTFAVISTIKRMMEKKCRELNCQTELEQVYLNDLFGNDFNTLTKGLSSE 119
TCS1f NATDLGAAGPNTFAVISTIKRMMEKKCRELNCQTELEQVYLNDLFGNDFNTLTKGLSSE 119
*.*****:***:***:**** *****:**** **

TCS1b VVGNKCEEVSCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTSREGLALNKG 175
TCS1c VVGNKCEEVSCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTSREGLALNKG 175
TCS1e VVGNKCEEVPCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTSREGLALNKG 178
TCS1a VIGNKCEEVPCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTSREGLALNKG 179
TCS1d VIGNKCEEVPCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTNREGLALNKG 179
TCS1f VIGNKCEEVPCYVMGVPGSFHGRLEPFRNSLHLVHSSYSVHWLTQAPKGLTNREGLALNKG 179
*:*****:***:***:**** *****:**** **

TCS1b KIYISKTSPPVVKHAYLSQFHEDFTMFLNARSQEVVPNGCMVLILHGRQSSDPSEMESCFC 235
TCS1c KIYISKTSPPVVKHAYLSQFHEDFTMFLNARSQEVVPNGCMVLILHGRQSSDPSEMESCFC 235
TCS1e KIYISKTSPPVVRHAYLSQFHEDFTMFLNARSQEVVPNGCMVLILRGKASDPDMESCFC 238
TCS1a KIYISKTSPPVVRHAYLSQFHEDFTMFLNARSQEVVPNGCMVLILRGRQCSDPDMQSCFC 239
TCS1d KIYISKTSPPIVRHAAYLTQFHEDFTMFLNARSQEVVPNGCMVLILRSRQSSDPDMQSCFC 239
TCS1f KIYISKTSPPIVRHAAYLSQFHEDFTMFLNARSQEVVPNGCMVLILRGRQSSDPDMQSCFC 239
*****:***:***:**** *****:**** **

TCS1b TWELLAIAIAELVSQGLIDEDKLDTFNVPSYWPSLEEVKDIVERDGSFTIDHLEGFELDS 295
TCS1c TWELLAIAIAELVSQGLIDKLDKLDTFNVPSYWPSLEEVKDIVERDGSFTIDHLEGFELDS 295
TCS1e TWELLAIAIAELVSQGLIDEDKLDTFNIPCYFPSLEEVKDIVERDGSFTIDHMEGFELDS 298
TCS1a TWELLAMAIAELVSQGLIDEDKLDTFNIPSYFASLEEVKDIVERDGSFTIDHIEGFELDS 299
TCS1d TWELLAKAIAELVSQGLIDEDKLDAFNIPCYFPSLEEVKDIVERDGSFTIDHMEGFGLDS 299
TCS1f TWELLAIAIAELVSQGLIDEDKLDTFNIPCYFPSLEEVKDIVERDGSFTIDHMEGFELDS 299
*****:***:***:**** *****:**** **

TCS1b LEMQENDKWVRGDKFAKMVRAFTEPIISNQFGHEIMDKLYDKFTHILVSDLEAKLPKTT 355
TCS1c LEMQEDDKWVRGDKFAKMVRAFTEPIISNQFGQHEIMDKLYDKFTHILVSDLEAKLPKTT 355
TCS1e LQMQENDKWVRGENFTKVVRAFTEPIISNQFGHEIMGKLYDKFTHIVSDLEAKLPKTT 358
TCS1a VEMQENDKWVRGEKFTKVVRAFTEPIISNQFGPEIMDKLYDKFTHIVSDLEAKLPKTT 359
TCS1d LQMEENDKWVRGEKFTKVVRAFTEPIISNQFGHEIMDKLYDKFTHIVSDFEAKLPKTT 359
TCS1f LQMQENDKWVRGENFTKVVRAFTEPIISNQFGHEIMDKLYDKFTHIVSDLEAKLPKTT 359
::*:*****:***:***:**** *****:**** **

TCS1b IILVLSKIVG 365
TCS1c IILVLSKIVG 365
TCS1e IILVLSKIDG 368
TCS1a IILVLSKIDG 369
TCS1d IILVLSKIDG 369
TCS1f IILVLSKIDG 369
***** *
```

# 3、蛋白质结构预测

# 3.1 咖啡碱合成酶TCS1一级结构分析

## Protparam分析序列理化性质

Number of amino acids: 369

Molecular weight: 41272.07

Theoretical pI: 5.04

Amino acid composition:

CSV format

Ala (A)	23	6.2%
Arg (R)	14	3.8%
Asn (N)	18	4.9%
Asp (D)	20	5.4%
Cys (C)	8	2.2%
Gln (Q)	15	4.1%
Glu (E)	29	7.9%
Gly (G)	22	6.0%
His (H)	8	2.2%
Ile (I)	19	5.1%
Leu (L)	37	10.0%
Lys (K)	20	5.4%
Met (M)	12	3.3%
Phe (F)	22	6.0%
Pro (P)	14	3.8%
Ser (S)	31	8.4%
Thr (T)	19	5.1%
Trp (W)	3	0.8%
Tyr (Y)	8	2.2%
Val (V)	27	7.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 49

Total number of positively charged residues (Arg + Lys): 34

Atomic composition:

Carbon	C	1834
Hydrogen	H	2871
Nitrogen	N	483
Oxygen	O	559
Sulfur	S	20

Formula:  $C_{1834}H_{2871}N_{483}O_{559}S_{20}$

Total number of atoms: 5767

Extinction coefficients:

Extinction coefficients are in units of  $M^{-1} cm^{-1}$ , at 280 nm measured in water.

Ext. coefficient 28920

Abs 0.1% (=1 g/l) 0.701, assuming all pairs of Cys residues form cystines

Ext. coefficient 28420

Abs 0.1% (=1 g/l) 0.689, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 35.83

This classifies the protein as stable.

Aliphatic index: 86.64

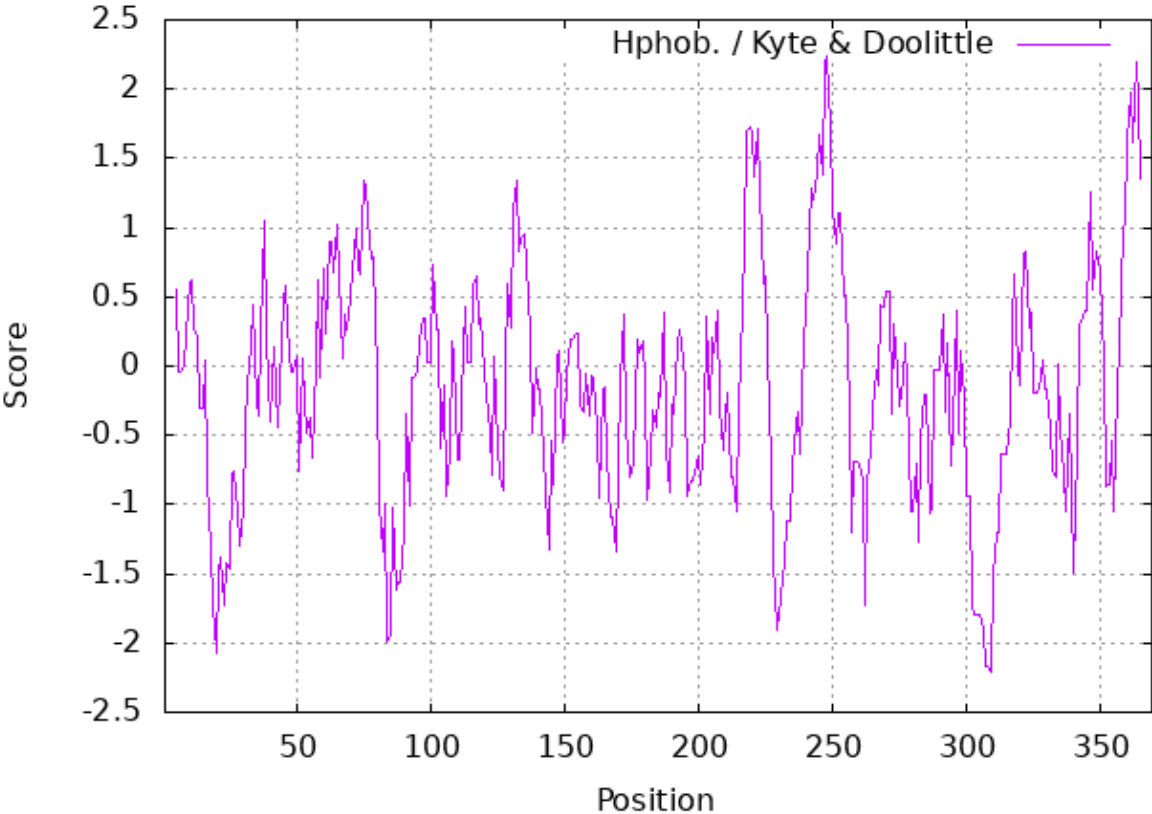
Grand average of hydropathicity (GRAVY): -0.137

# Protparam分析序列理化性质

<b>Number of amino acids</b>	<b>369</b>
<b>Molecular weight</b>	<b>41272.07</b>
<b>Theoretical pI</b>	<b>5.04</b>
<b>Total number of negatively charged residues (Asp+Glu)</b>	<b>49</b>
<b>Total number of positively charged residues (Arg+Lys)</b>	<b>34</b>
<b>Aliphatic index</b>	<b>86.64</b>
<b>Grand average of hydropathicity (GRAVY)</b>	<b>-0.137</b>

# ProtScale分析亲疏水性、TMHMM分析跨膜螺旋

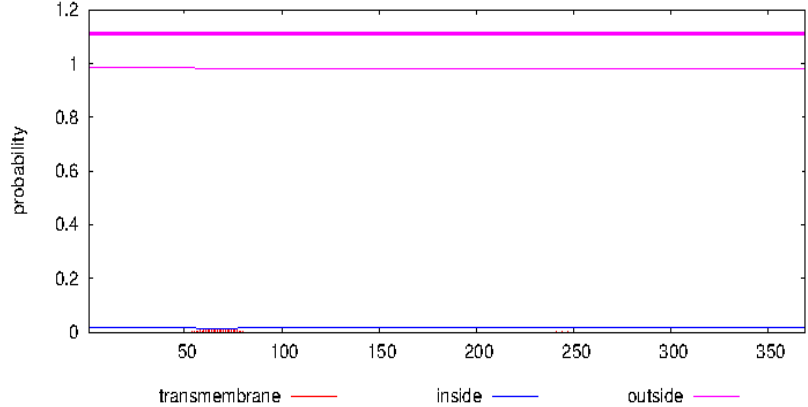
ProtScale output for TCS1\_CAMSI



## TMHMM result

```
# WEBSEQUENCE Length: 369
# WEBSEQUENCE Number of predicted TMHs: 0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.16553
# WEBSEQUENCE Exp number, first 60 AAs: 0.02844
# WEBSEQUENCE Total prob of N-in: 0.01544
WEBSEQUENCE TMHMM2.0 outside 1 369
```

TMHMM posterior probabilities for WEBSEQUENCE

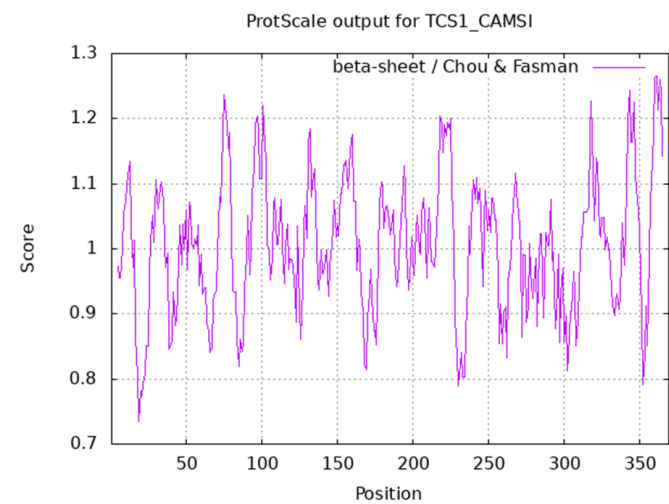
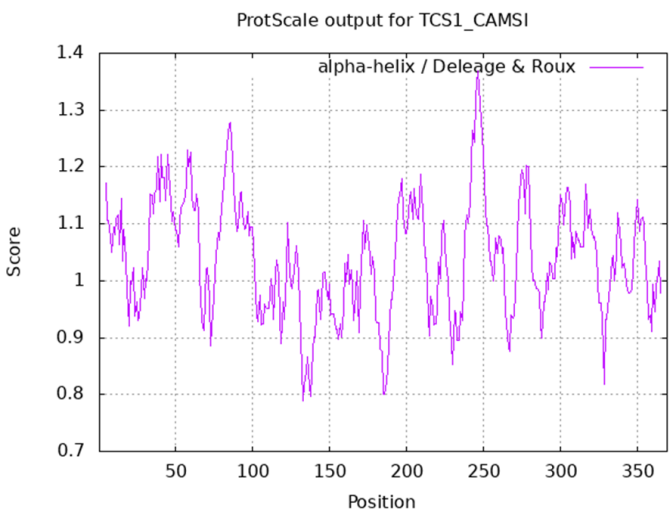


# EMBOSS Explorer分析密码子偏好

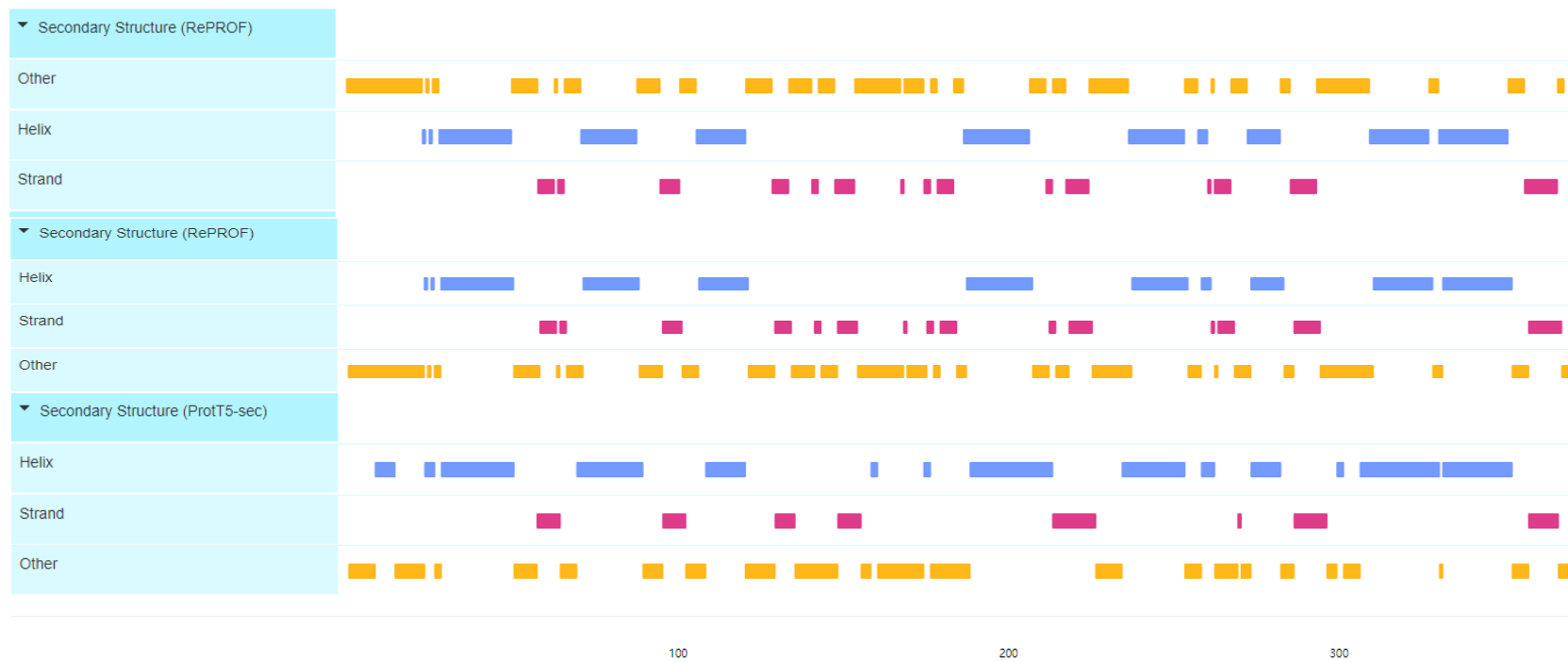
<b>Codon adaptation index, CAI</b>	<b>0.645</b>
<b>Effective number of codon, ENC</b>	<b>48.288</b>
<b>Observed number of occurrences of codon 'i', Obsi</b>	<b>Frequency</b>
<b>TTT</b>	<b>72.653</b>
<b>AAA</b>	<b>67.690</b>
<b>AAT</b>	<b>48.736</b>

# 3.2 咖啡碱合成酶TCS1二级结构分析

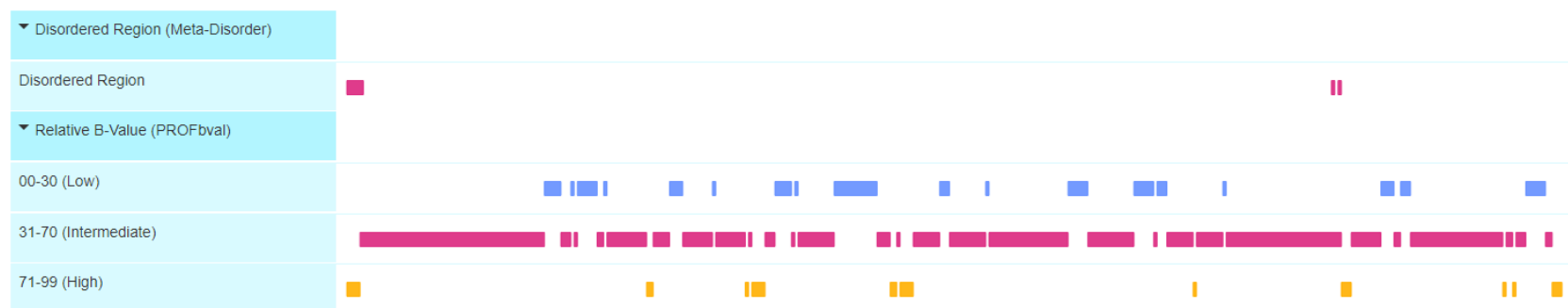
## PredictProtein预测蛋白质二级结构



**A**



**B**



# PredictProtein预测亚细胞定位



细胞质

Predicted localization for the Eukarya domain: Cytoplasm (GO term ID: [GO:0005737](#)) Prediction confidence 26

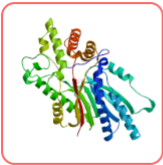


# 3.2 咖啡碱合成酶TCS1三级结构预测

## Swiss-Model查找蛋白质结构模型

### Template Results

Templates	Quaternary Structure	Sequence Similarity	Alignment	More				
Sort	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	6lyh.1.A N-methyltransferase CkTcS <i>Crystal structure of tea N9-methyltransferase CkTcS in complex with SAH and 1,3,7-trimethyluric acid</i>	0.90	0.61	91.97	X-ray, 3.2Å	homo-dimer Δ	2 x SAH, 2 x EXU
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	6lyh.1.A N-methyltransferase CkTcS <i>Crystal structure of tea N9-methyltransferase CkTcS in complex with SAH and 1,3,7-trimethyluric acid</i>	0.90	0.61	92.80	X-ray, 3.2Å	homo-dimer Δ	2 x SAH, 2 x EXU
<input type="checkbox"/>	<input type="checkbox"/>	6lyi.1.B N-methyltransferase CkTbS <i>Crystal structure of a N-methyltransferase CkTbS from Camellia assamica var. kucha</i>	0.90	0.62	91.06	X-ray, 2.5Å	homo-dimer Δ	None
<input type="checkbox"/>	<input type="checkbox"/>	6lyi.1.B N-methyltransferase CkTbS <i>Crystal structure of a N-methyltransferase CkTbS from Camellia assamica var. kucha</i>	0.89	0.62	91.60	X-ray, 2.5Å	homo-dimer Δ	None
<input type="checkbox"/>	<input type="checkbox"/>	6lyi.1.A N-methyltransferase CkTbS <i>Crystal structure of a N-methyltransferase CkTbS from Camellia assamica var. kucha</i>	0.88	0.60	91.06	X-ray, 2.5Å	homo-dimer Δ	None
<input type="checkbox"/>	<input type="checkbox"/>	6lyi.1.A N-methyltransferase CkTbS <i>Crystal structure of a N-methyltransferase CkTbS from Camellia assamica var. kucha</i>	0.88	0.60	91.60	X-ray, 2.5Å	homo-dimer Δ	None



Model 02

Structure Assessment

**Oligo-State** Monomer

**Ligands** 1 x SAH (matching prediction)

1 x S-ADENOSYL-L-HOMOCYSTEINE

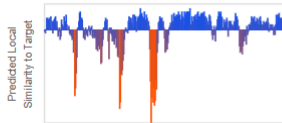
SAH.1: 21 residues within 4Å

7 PLIP interactions

**GMQE** 0.89

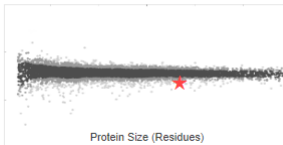
**QMEANDisCo Global** 0.84 ± 0.05

**QMEANDisCo Local** Local Quality Estimate



**QMEAN Z-Scores**

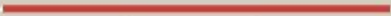



QMEAN	-2.49
Cβ	-2.03
All Atom	-0.60
solvation	0.69
torsion	-2.45



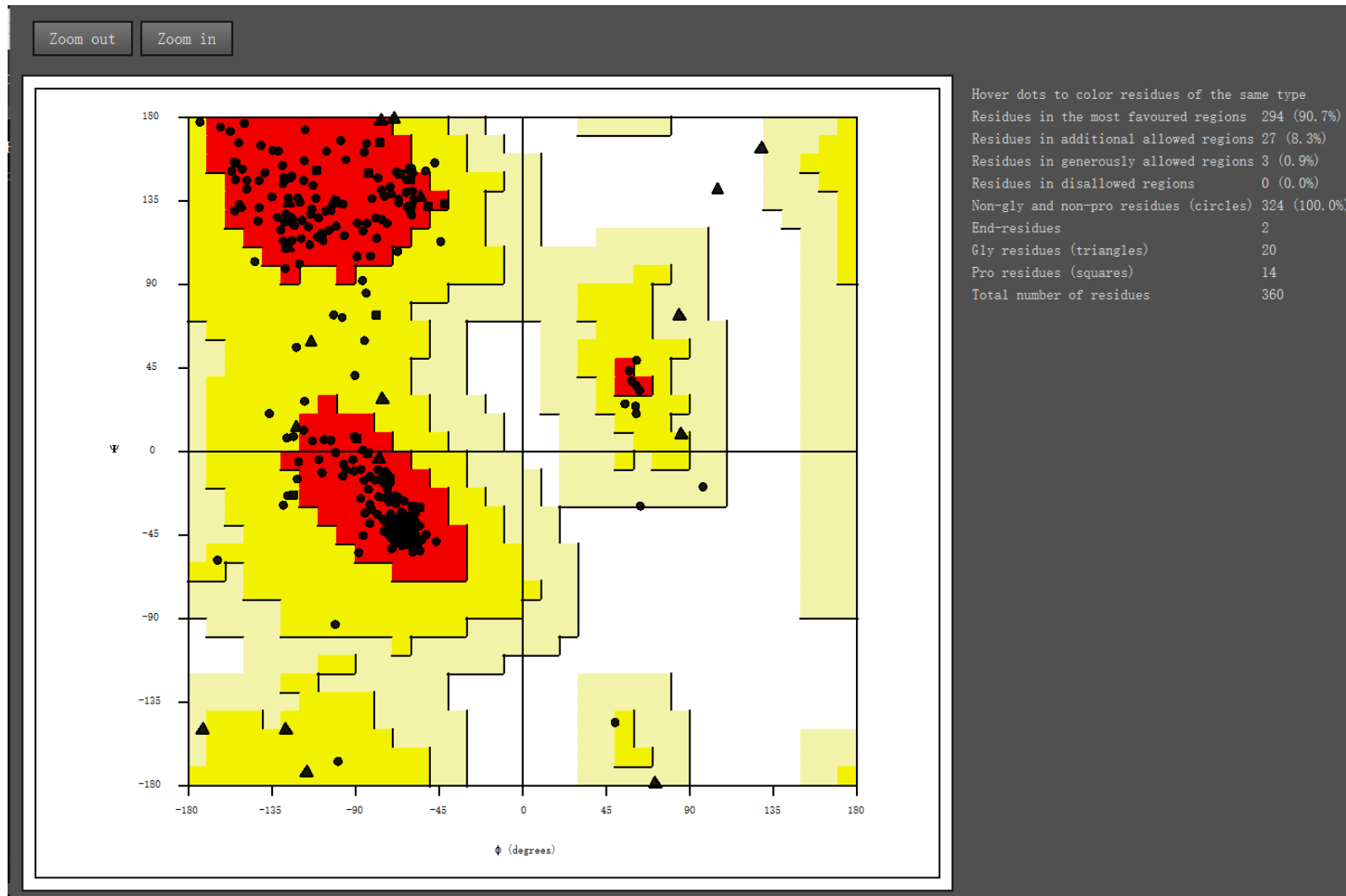
Template	Seq Identity	Coverage	Description
6lyh.1.A	92.80%	<div style="width: 100%; height: 10px; background-color: blue;"></div>	N-methyltransferase CkTcS <i>Crystal structure of tea N9-methyltransferase CkTcS in complex with SAH and 1,3,7-trimethyluric acid</i>

Model-Template Alignment

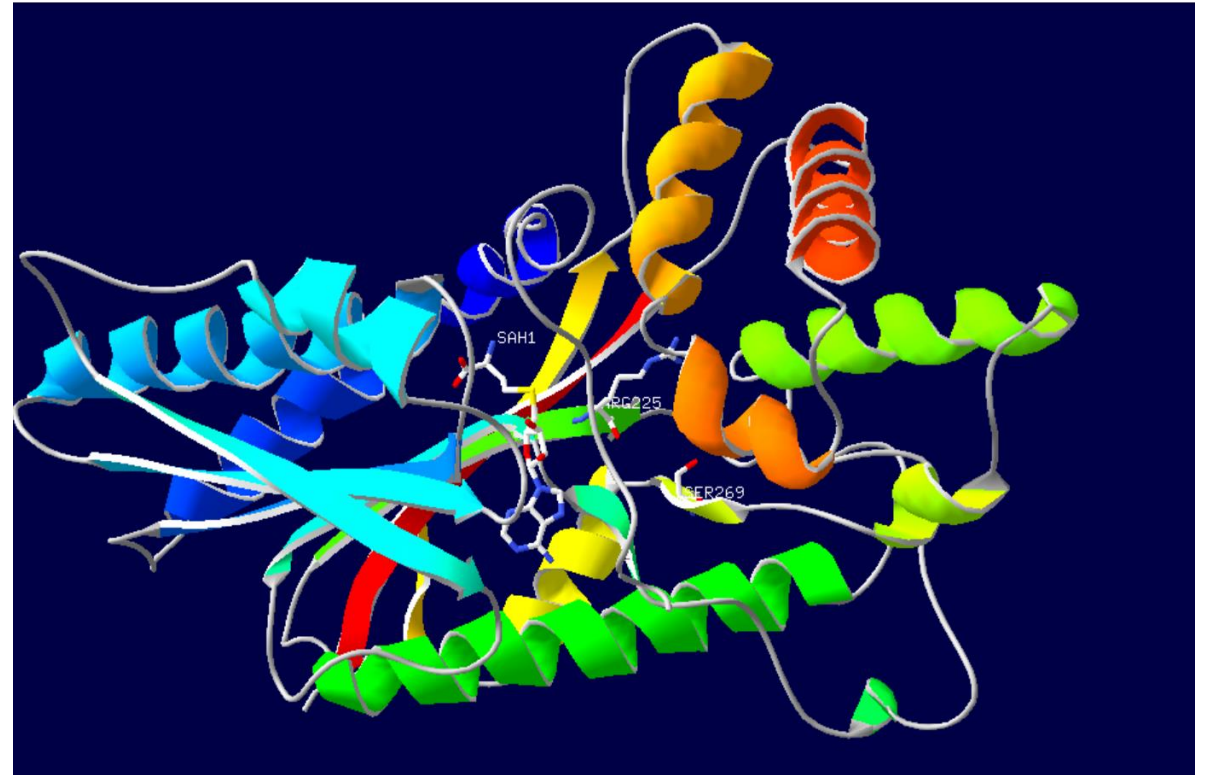
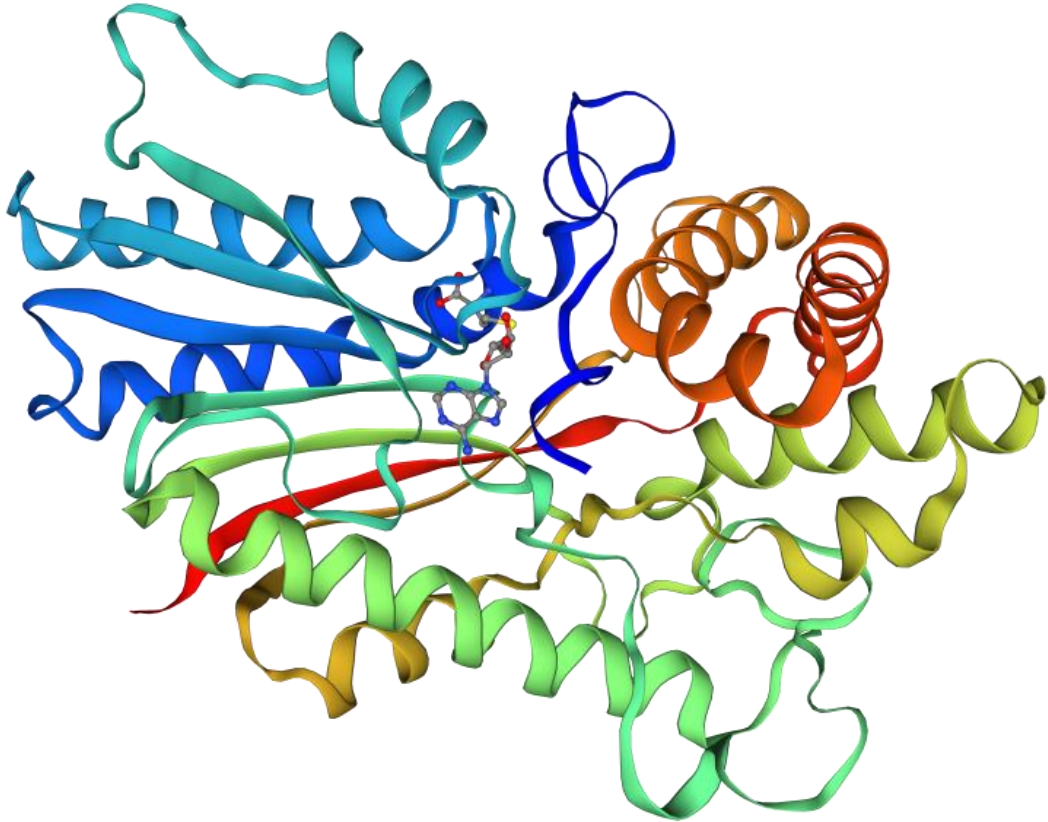
# Phyre<sup>2</sup>网站建模

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	<a href="#">c6lyhE</a> <input type="radio"/> <input type="checkbox"/>	 Alignment		100.0	91	<b>PDB header:</b> transferase <b>Chain:</b> E: <b>PDB Molecule:</b> n-methyltransferase cktcs; <b>PDBTitle:</b> crystal structure of tea n9-methyltransferase cktcs in com with2 sah and 1,3,7-trimethyluric acid <b>PDB Entry:</b> <a href="#">PDBe</a> <a href="#">RCSB</a> <a href="#">PDBj</a>
2	<a href="#">d1m6ex</a> <input type="radio"/> <input type="checkbox"/>	 Alignment		100.0	41	<b>Fold:</b> S-adenosyl-L-methionine-dependent methyltransferases <b>Superfamily:</b> S-adenosyl-L-methionine-dependent methyltransferases <b>Family:</b> Salicylic acid carboxyl methyltransferase (SAMT) <b>PDB entry:</b> <a href="#">PDBe</a> <a href="#">RCSB</a> <a href="#">PDBj</a>

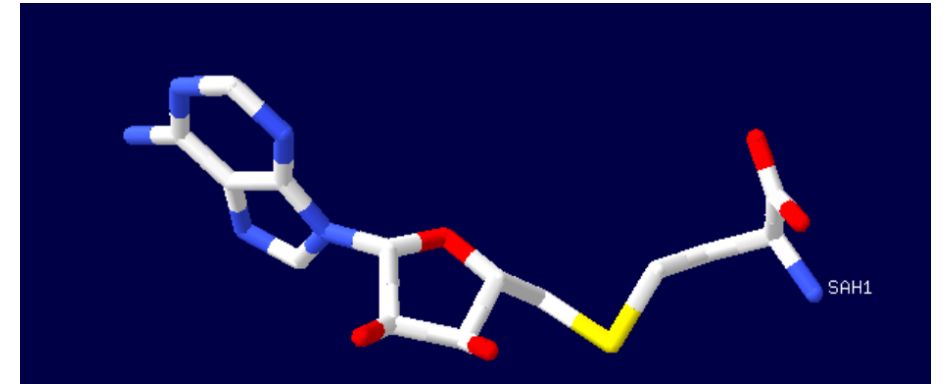
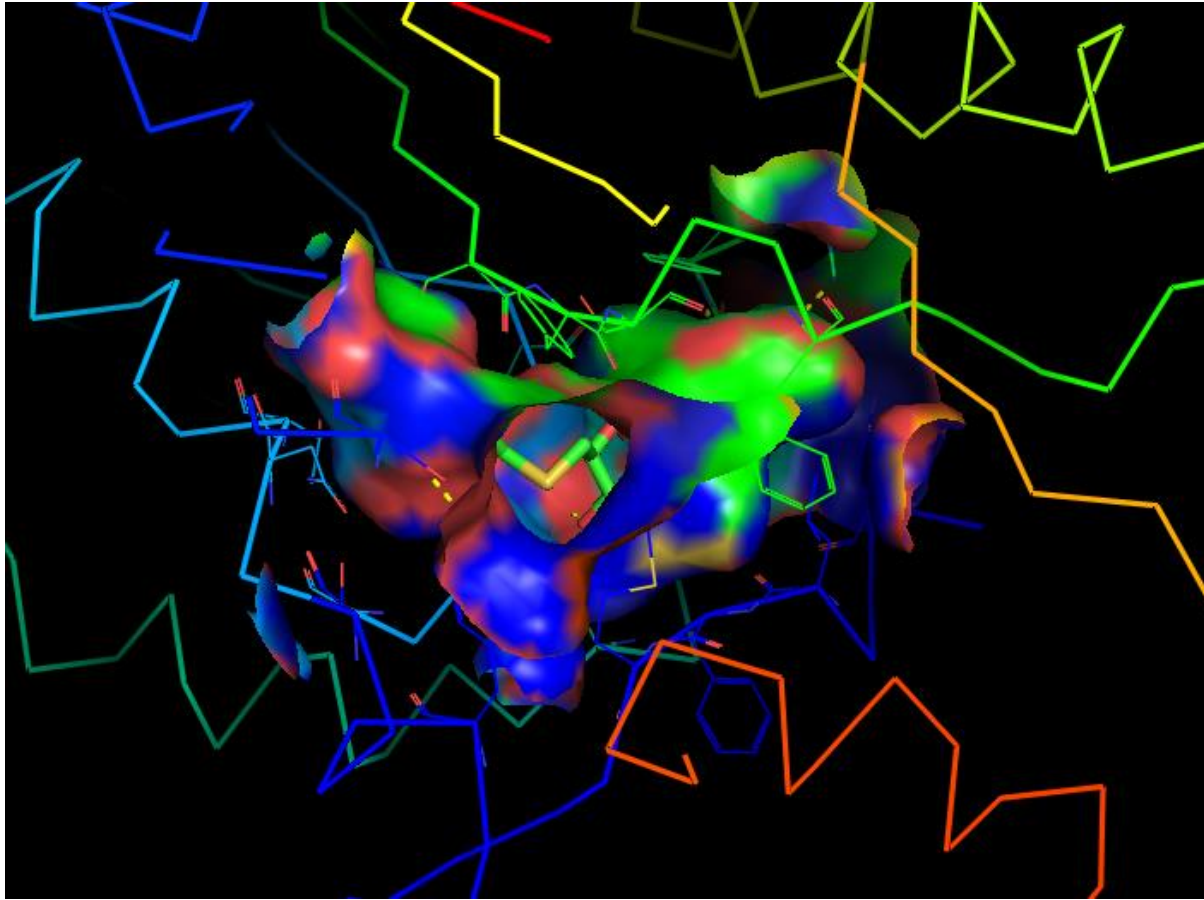
# Ramachandran Plot (拉氏图)



# TCS1三级结构预测结果



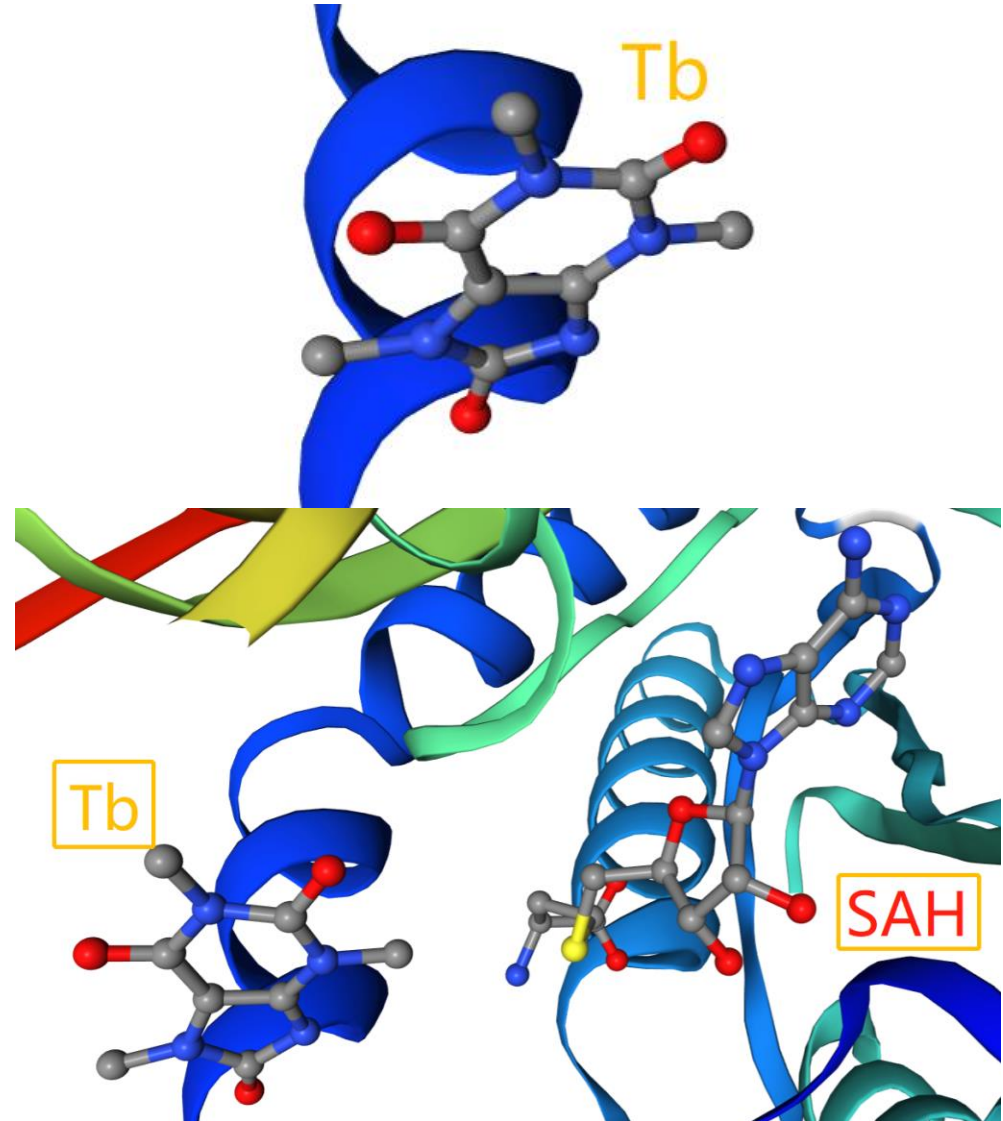
# TCS1特殊三级结构



**SAH**

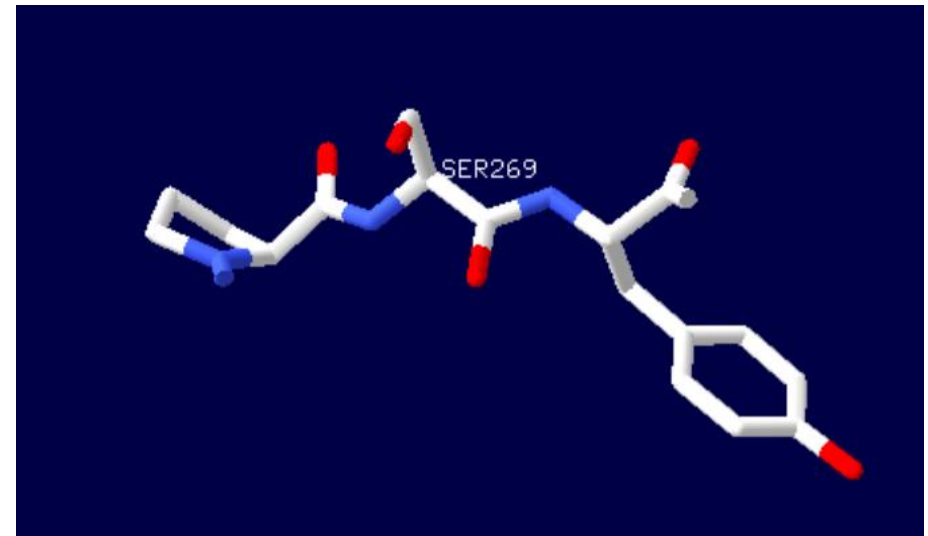
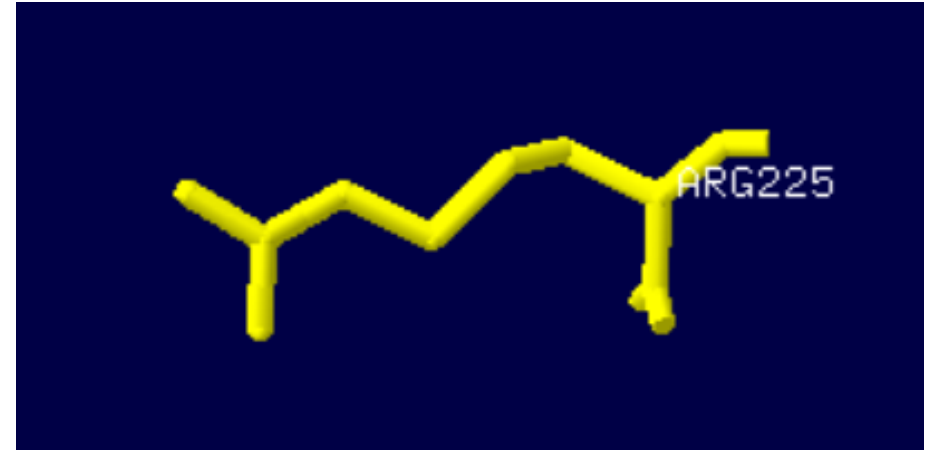
# 底物结构与底物结合位点

茶树咖啡碱合成酶1 (TCS1) 具有催化N-3和N-1甲基化的活性，能够催化7-甲基黄嘌呤 (7mX) 与S-腺苷-L-蛋氨酸 (SAM) 生成可可碱 (theobromine)，进而生成咖啡碱。



## TCS1关键氨基酸位点

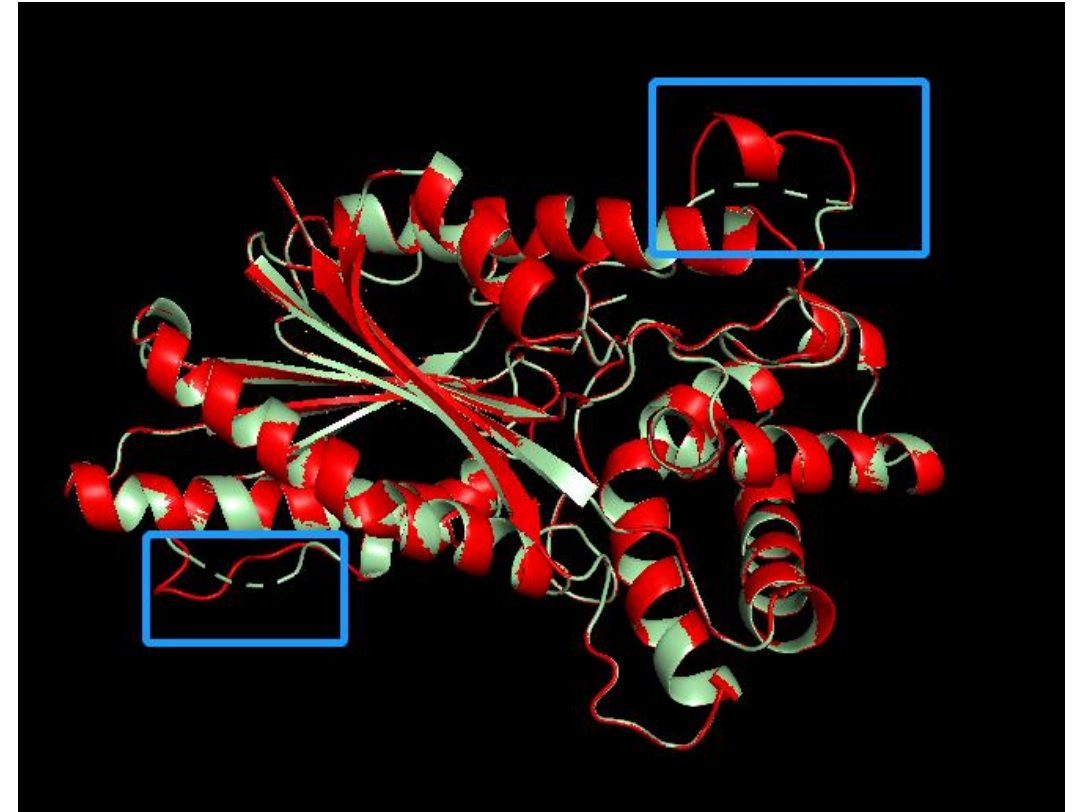
与 *TCS1a-c* 相比, *TCS1d-f* 具有 Ser269Cys 变化, *TCS1a* 和 *TCS1d* 中 269 位氨基酸残基分别为丝氨酸和半胱氨酸, 进行定点诱变实验证明 TCS1 中的第 225、269 位氨基酸残基在 TCS (咖啡碱合成酶) 活性和底物识别中起重要作用。



# TCS1与CkTcS三级结构比较

N-methyltransferase CkTcS

Seqres	SVKEVLFMNTGEGESSYVQNSSFTEKVASMAMPALENAVETLFSKDFHLFQAINAADLGCATGPN	65
61yh.1.A	VKEVLFMNTGEGESSYVQNS(SFTEKVASMAMPALENAVETLFSKDFHLFQAINAADLGCATGPN	65
61yh.1.B	VKEVLFMNTGEGESSYVQNSSFTEKVASMAMPALENAVETLFSKDFHLFQAINAADLGCATGPN	65
Seqres	TFAVISTIKRMMKKRELNCQTLELQVYMNDFLFGNDFNTLFGKLSXVIGNKCEEVSCYVMGVF	130
61yh.1.A	TFAVISTIKRMMKKRELNCQTLELQVYMNDFLFGNDFNTLFGKLSXVI-----EEVSCYVMGVF	130
61yh.1.B	TFAVISTIKRMMKKRELN--TLELQVYMNDFLFGNDFNTLFGKLS--SCYVMGVF	130
Seqres	GSPFHGRFPFRNSLHLVHSSYSVHWLTQAPKGLTSREGLALNKGRIYISKTSPPVVREAYLSQFHE	195
61yh.1.A	GSPFHGRFPFRNSLHLVHSSYSVHWLTQAPK-----LALNKGRIYISKTSPPVVREAYLSQFHE	195
61yh.1.B	GSPFHGRFPFRNSLHLVHSSYSVHWLTQAPKGLTSRE-LALNKGRIYISKTSPPVVREAYLSQFHE	195
Seqres	DFTMFLNARSQEVVPNGCMVLILRGRQSSDPSDMQSCFIWELLAIATAELVSQGLIDEDKLDTFN	260
61yh.1.A	DFTMFLNARSQEVVPNGCMVLILRGRQSSDPSDMQSCFIWELLAIATAELVSQGLIDEDKLDTFN	260
61yh.1.B	DFTMFLNARSQEVVPNGCMVLILRGRQSSDPSDMQSCFIWELLAIATAELVSQGLIDEDKLDTFN	260
Seqres	IPCYFPSLEEVDKIVERDGSFTIDHMEGFELDSLQMQENDKWVRGEKFAKIVRAFTEPIISNQFG	325
61yh.1.A	IPCYFPSLEEVDKIVERDGSFTIDHMEGFEDDSLQMQENDKWVRGEKFAKIVRAFTEPIISNQFG	325
61yh.1.B	IPCYFPSLEEVDKIVERDGSFTIDHMEGFEDDSLQMQENDKWVRGEKFAKIVRAFTEPIISNQFG	325
Seqres	HEIMDKLYDKFTHIVVSDLEAKLPKTTSIILVLSKIVG	363
61yh.1.A	HEIMDKLYDKFTHIVVSDLEAKLPKTTSIILVLSKIV-	362
61yh.1.B	HEIMDKLYDKFTHIVVSDLEAKLPKTTSIILVLSKIV-	362



注：红色为TCS1a  
绿色为CkTcS



# 总结

a) 将新鉴定的N-甲基转移酶与已知的N-甲基转移酶进行系统发生分析建树，确定它们的进化与分枝情况，根据结果分析N-甲基转移酶是否通过趋同进化从两个独立的起源进化而来，从而初步推断酶功能。

b) 对酶的蛋白质一级结构进行理化分析，为我们提供酶的基础信息用于鉴定、设计底物活性实验及酶的基础分离实验。

c) 蛋白质三级结构预测可以为我们提供酶与底物及辅因子的结合情况，判断其存在形式、核心区域、氢键作用、保守结构域、活性结构等基础信息，对于酶的功能活性研究，及后续的可变位点或关键位点的定点突变实验的设计具有重要意义。

# 总结

d) 在蛋白质的结构分析中找到主要作用残基位点，底物结合特异性，重要的相互作用位点，设计实验判断定向底物的位点,综上对一个未知蛋白进行系统的挖掘验证。

e) 茶树相关数据库如TAIP，拟南芥相关数据库TAIR和转录因子数据库Plant TFDB，为进一步研究调控咖啡碱合成酶基因的启动子及相关转录因子提供基础信息。根据已知元件进行转录水平研究，从转录水平分析影响茶树咖啡碱代谢合成的调控机制。

# 参考文献

- [1] Jin, Ji-Qiang, Ma, et al. Natural allelic variations of TCS1 play a crucial role in caffeine biosynthesis of tea plant and its related species[J]. *Plant Physiology & Biochemistry*, 2016.
- [2] Zhang YH, Li YF, Wang Y , et al. Identification and characterization of N9-methyltransferase involved in converting caffeine into non-stimulatory theacrine in tea. *Nat Commun*. 2020 Mar 19;11(1):1473. doi: 10.1038/s41467-020-15324-7.
- [3] Jin JQ, Chai YF, Liu YF , et al. Hongyacha, a Naturally Caffeine-Free Tea Plant from Fujian, China. *J Agric Food Chem*. 2018 Oct 31;66(43):11311-11319. doi: 10.1021/acs.jafc.8b03433.

**敬请批评指正!**