棕色棉查尔酮合成酶的生物信息学分析

小组成员：刘松江　　张　震

赵春德　　许二波

报告人：张　震

主要内容

一： 研究背景
二： 基因分析
三： 蛋白质分析

一：研究背景：
1：棉花

棉花是锦葵科，棉属植物的种籽纤维，是世界上主要的农作物之一，也是我国重要的经济作物，目前，我国棉花产量位于世界前列。

2：彩棉

- 彩棉用现代生物工程技术培育出来的一种在棉花吐絮时纤维就具有天然色彩的新型纺织原料。
- 彩色棉制品有利于人体健康，舒适：亲和皮肤，不起静电，不起球.透汗性好。
- 彩色棉制品在纺织过程中减少印染工序;迎合了减少了对环境污染。

3:查尔酮合成酶

查尔酮合成酶(chalcone synthase,CHS) 是生物类黄酮合成的关键酶，它催化丙二酰辅酶A的3个乙酸基和对羟苯丙烯酰辅酶A的1个乙酸基缩合，产生柚配基查尔酮和生松素查尔酮。

研究表明，通过荧光定量PCR检测显示：该基因在彩棉纤维细胞中优先表达，且在棕色棉纤维中的表达量远高于其近等基因系白色棉，由此推断，该基因可能在彩棉色素形成中发挥重要作用。

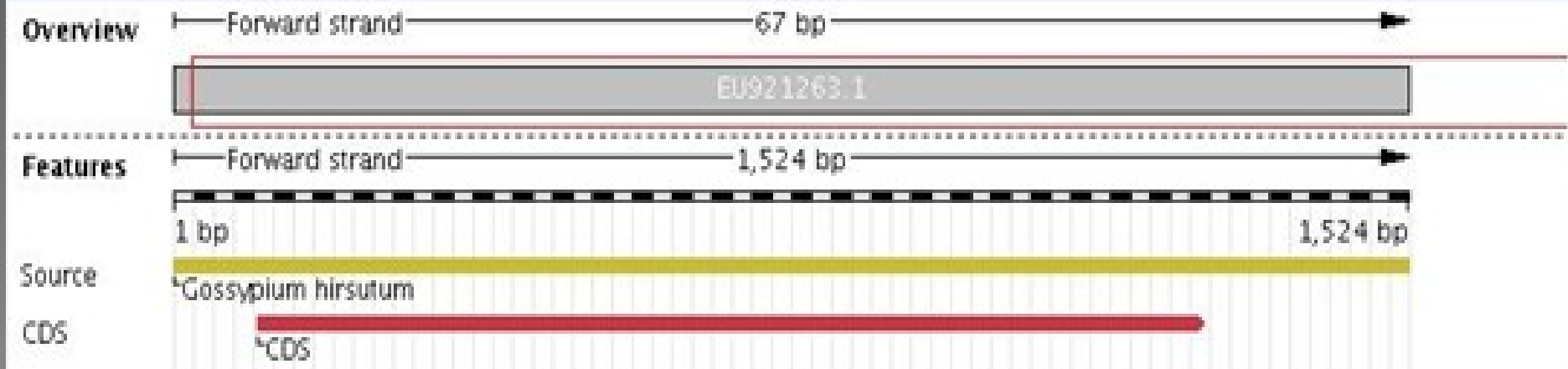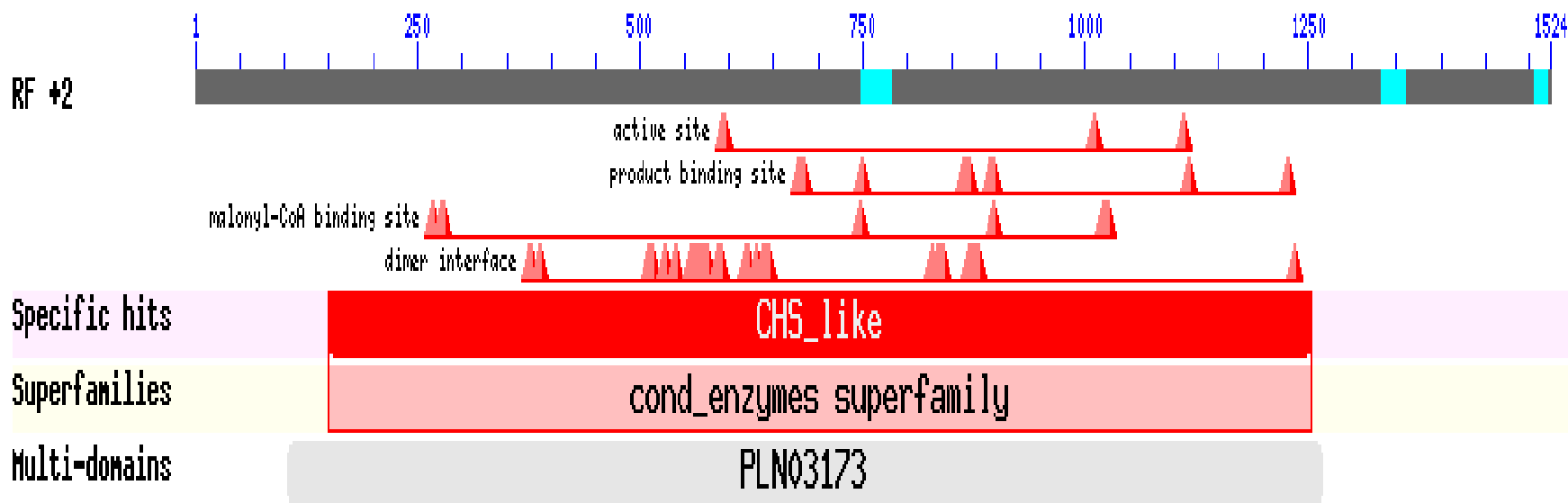对该基因进行进一步分析，有利于改善彩棉颜色单调，着色不均匀的的缺点，育成更好的彩色棉品种。

Graphical summary   show options »

| | RF +2 | | | | | | | |

Specific hits: CHS_like

Superfamilies: cond_enzymes superfamily

Multi-domains: PLN03173

[Search for similar domain architectures]   ?   [Refine search]   ?

List of domain hits

Description	Pssmld	Multi-dom	E-value
CHS_like[cd00831], Chalcone and stilbene synthases; plant-specific polyketide synthases (PKS) and related ...	238427	yes	2.46e-155
PLN03173[PLN03173], chalcone synthase; Provisional	178717	yes	0e+00

InterProScan

Chalcone/stilbene synthase, C-terminal

Domains and repeats
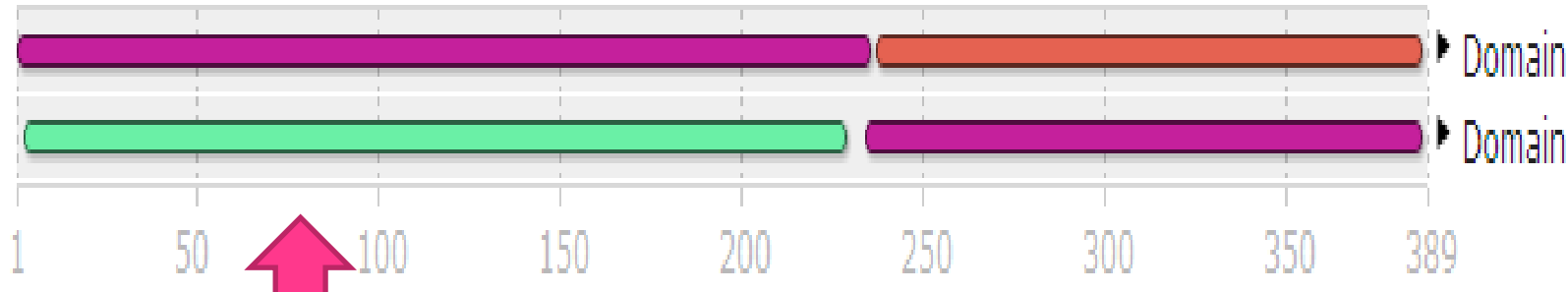
Chalcone/stilbene synthase, N-terminal

等电点和分子量分析

ExPASy
Bioinformatics Resource Portal
SIB

Compute pI/Mw tool

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
          10          20          30          40          50          60
MVTVEEVRKA QRAQGPATVL AIGTSTPPNC VDQSTYPDYY FRITNSEHKT ELKEKFKRMC
          70          80          90         100         110         120
EKSMIKKRYM YLTEEILKEN PNVCEYMAPS LDARQDMVVV EVPKLGKEAA TKAIKEWGQP
         130         140         150         160         170         180
KSKITHLVFC TTSGVDMPGA DYQLTKLLGL RPSVKRLMMY QQGCFAGGTV LRVAKDLAEN
         190         200         210         220         230         240
NKGARVLVVC SEITAVTFRG PSDTHLDSLV GQALFGDGAA AVIIGADPMP EIEKPMFEIV
         250         260         270         280         290         300
SVAQTILPDS DGAIDGHLRE VGLTFHLLKD VPGLISKNIE KSLVEAFQPL GISDWNSLFW
         310         320         330         340         350         360
IAHPGGPAIL DQVEAKLALK PEKLRATRHV LSEYGNMSSA CVLFILDEMR KKSREDGLQT
         370         380
TGEGLEWGVL FGFGPGLTVE TVVLHSVAA
```
Theoretical pI/Mw: 6.12 / 42668.39

4: 基因中可能的读码框
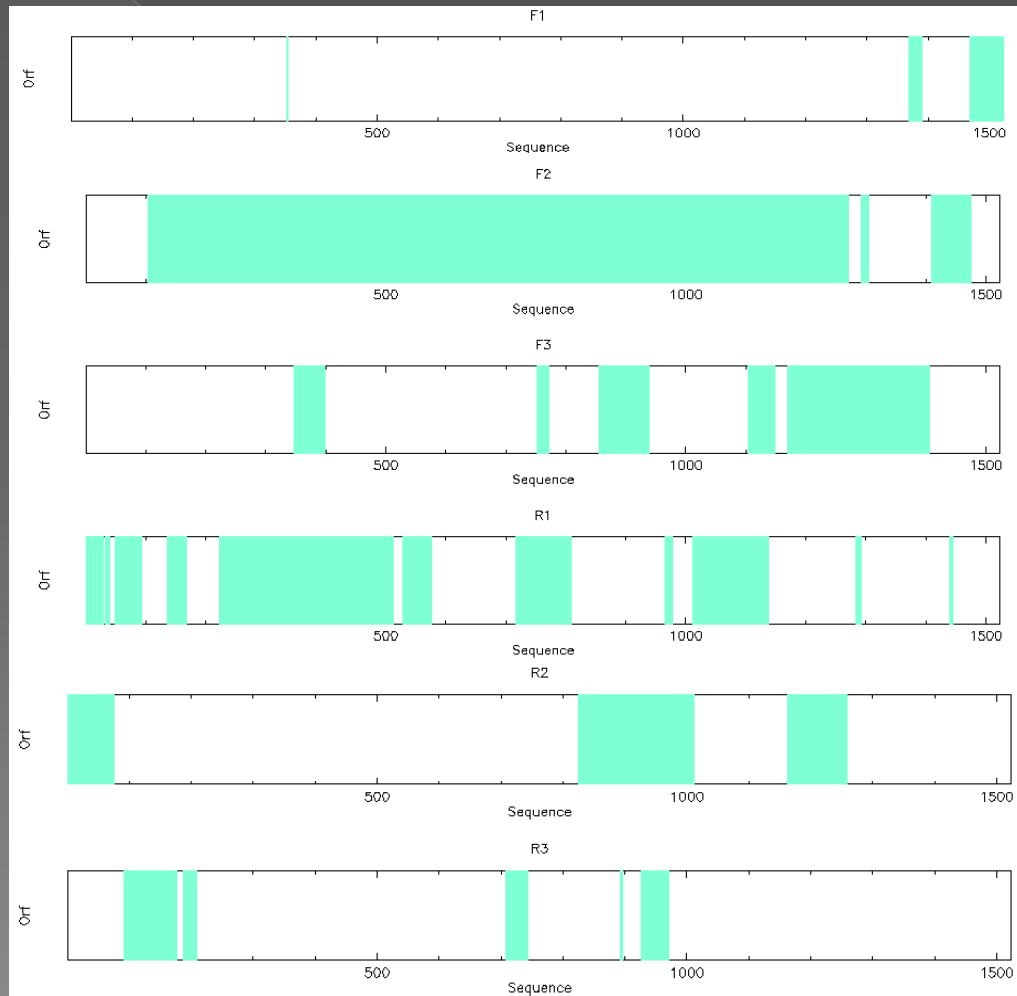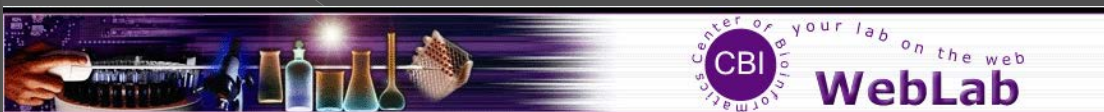
Service   sixpack(v6.0.1) - Display a DNA sequence with 6-frame translation and ORFs

>_2_ORF2 Translation of in frame 2, ORF 2, threshold 100, 407aa

STPPTPHHHWLFVIIRKRMVTVEEVRKAQRAQGPATVLAIGTSTPPNCVDQSTYPDYYFR

ITNSEHKTELKEKFKRMCEKSMIKKRYMYLTEEILKENPNVCEYMAPSLDARQDMVVVEV

PKLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQ

GCFAGGTVLRVAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAV

IIGADPMPEIEKPMFEIVSVAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKS

LVEAFQPLGISDWNSLFWIAHFGGPAILDQVEAKLALKFEKLRATRHVLSEYGNMSSACV

LFILDEMRKKSREDGLQTTGEGLEWGVLFGFGPGLTVETVVLHSVAA

用SixPack分析得到的氨基酸序列与GeneBank中提供的氨基酸序列进行比对

Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
407	2008.0	389/407 (95.6%)	389/407 (95.6%)	18/407 (4.4%)

Multiple sequence alignment

	Sequence
1. Q9M5B2 PETHY	----MVIVEEYRKAQRAEGPATVMAIGTATPTNCVDQSTYPDYYFRITNSEHKTDLKEKFKRMCEKSMIKKRYMHLTEEILKENPSMCEY
2. CHSY VITVI	----MVSVAEIRKAQRAEGPATVLAIGTATPANCVYQADYPDYYFRITNSEHMTELKEKFKRMCEKSMINKRYMHLTEEILKENPNVCAY
3. E9JWE4 PYRPY	----MVTVEEVRKAQRAEGPATVLAIGTATPPNCVDQATYPDYYFRITNSEHKTELKEKFQRMCDKSMIKTRYMYLTEEILKENPTVCEY
4. CHS5 PEA	----MVSVSEIRKAQRAEGPATILAIGTANPANCVEQSTYPDFYFRITNSEHKTELQKFQRMCDKSMINKRYMYLTEEILKENPSVCEY
5. CHS5 SOYBN	----MVSVEEIRQAQRAEGPATVMAIGTATPPNCVDQSTYPDYYFRITNSEHMTELKEKFKRMCDKSMIKKRYMYLNEEILKENPSVCAY
6. CHS8 MEDSA	----MVSVSEIRTAQRAEGPATILAIGTANPANCVEQSTYPDFYFKITNSEHKTELKEKFQRMCDKSMIKKRYMYLTEEILKENPSVCEY
7. CHS5 TRISU	----MVSVAEIRQAQRAEGPATILAIGTANPANKVEQATYPDFYFKITNSEHKVELKEKFQRMCDKSMIKSRYMYLTEEILKENPSVCEY
8. CHS5 SORBI	MAAATVTVAEVRKAQRAEGPATVLAIGTATPANCVHQADYPDYYFAITKSEHMTELKEKFKRMCDKSQIMKRYMHLTEEYLAENPNMCAY

	Sequence
1. Q9M5B2 PETHY	MAPSLDARQDIVVVEVPRLGKEAAQKAIKEWGQPKSKITHLVFCTTSGVDMPGCDYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAK
2. CHSY VITVI	MAPSLDARQDMVVVEVPRLGKEAAAKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAK
3. E9JWE4 PYRPY	MAPSLDARQDMVVVEVPRLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAK
4. CHS5 PEA	MAPSLDARQDMVVVEVPRLGKEAAVKAIKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTVLRLAK
5. CHS5 SOYBN	MAPSLDARQDMVVMEVPRLGKEAATKAIKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPSVKRYMMYQQGCFAGGTVLRLAK
6. CHS8 MEDSA	MAPSLDARQDMVVVEVPRLGKEAAVKAIKEWGQPKSKITHLIVCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTVLRLAK
7. CHS5 TRISU	MAPSLDARQDMVVVEVPRLGKEAAVKAIKEWGQPKSKITHLIFCTTSGVDMPGADYQLTKLLGLRPYVKRYMMYQQGCFAGGTVLRLAK
8. CHS5 SORBI	MAPSLDARQDIVVVEVPRLGKAAAHKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKMLGLRPSVNRLMMYQQGCFAGGTVLRVAK

	Sequence
1. Q9M5B2 PETHY	LAENNKGARVLVVCSEITAVTFRGPNDTHLDSLVGQALFGDGAGAIIIGSDPIEGVERPLFELVSAAQTLLPDSHGAIDGHLREVGLTFH
2. CHSY VITVI	LAENNAGSRVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAVIIGADPDTKIERPLFELVSAAQTILPDSEGAIDGHLREVGLTFH
3. E9JWE4 PYRPY	LAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAVIIGSDPVEIVERPLFELVSAAQTILPDSDGAIDGHLREVGLTFH
4. CHS5 PEA	LAENNKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAALIVGSDPIPEIEKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFH
5. CHS5 SOYBN	LAENNKGARVLVVCSEITAVTFRGPTDTHLDSLVGQALFGDGAAAVIVGSDPLE-VEKPLFQLVWTAQTILPDSEGAIDGHLREVGLTFH
6. CHS8 MEDSA	LAENNKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAALIVGSDPIPEIEKPIFEMVWTAQTIAPDSEGGIDGHLREAGLTFH
7. CHS5 TRISU	LAENNKGARVLVVCSEVTAVTFRGPSDTHLDSLVGQALFGDGAAALIVGSDPVPEIEKPIFEMVWTAQTIAPDSEGAIDGHLREAGLTFH
8. CHS5 SORBI	LAENNKGARVLVVCSEITAVTFRGPSESHLDSMVGQALFGDGAAAVIVGADPDEKVERPLFQLVSASQTILPDSEGAIDGHLREVGLTFH

序列比对

- 序列比对发现，棉花查尔酮CHS基因与葡萄、矮牵牛、梨、黄蜀葵等同源性较高。均达到90%以上。

蛋白质柔性分析

ProtScale Average flexibility

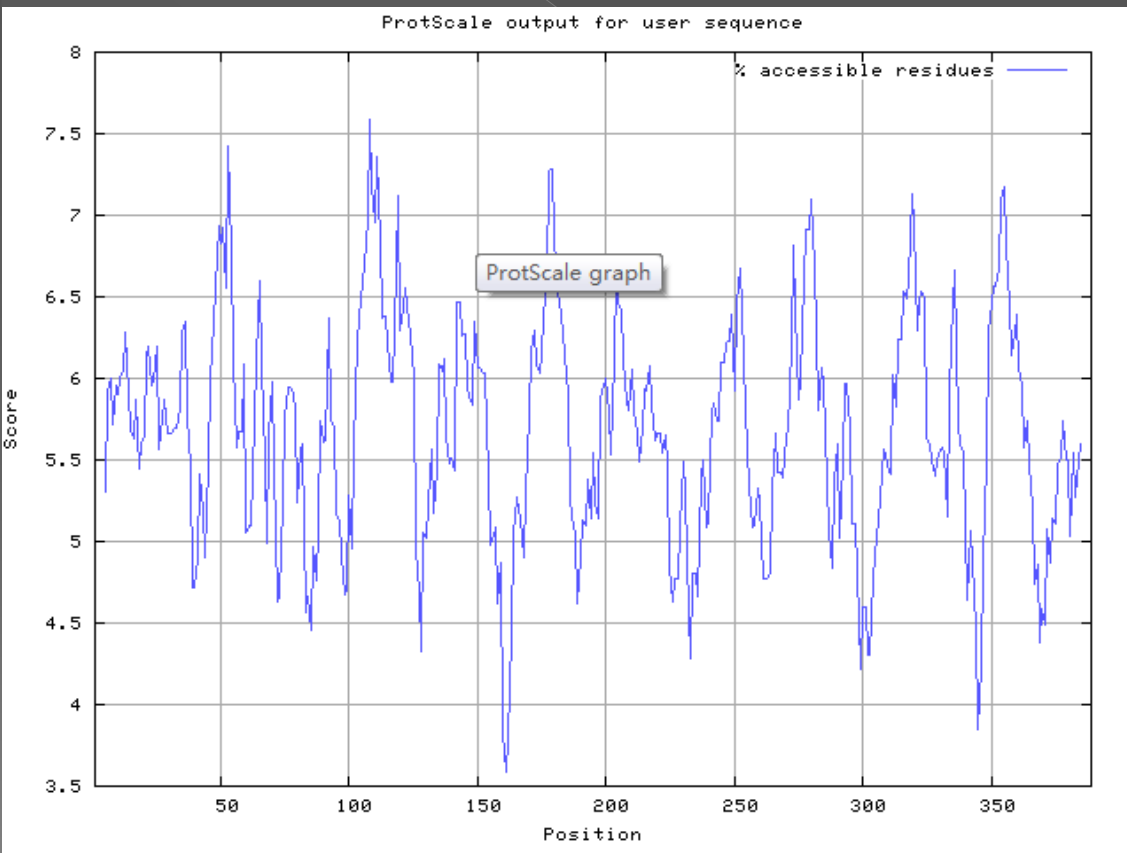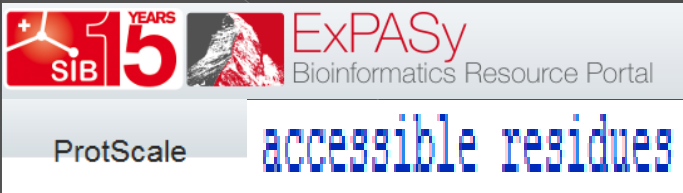ProtScale output for user sequence

scale values:
Ala: 0.360
Arg: 0.530
Asn: 0.460
Asp: 0.510
Cys: 0.350
Gln: 0.490
Glu: 0.500
Gly: 0.540
His: 0.320
Ile: 0.460
Leu: 0.370
Lys: 0.470
Met: 0.300
Phe: 0.310
Pro: 0.510
Ser: 0.510
Thr: 0.440
Trp: 0.310
 Tyr: 0.420
Val: 0.390

蛋白质柔性越高，蛋白质稳定性越好。

蛋白质溶剂可及性分析

scale values:
Ala:6.600
Arg:4.500
Asn:6.700
Asp:7.700
Cys:0.900
Gln:5.200
Glu:5.700
Gly:6.700
His:2.500
Ile: 2.800
Leu:4.800
Lys:10.300
Met:1.000
Phe:2.400
Pro:4.800
Ser:9.400
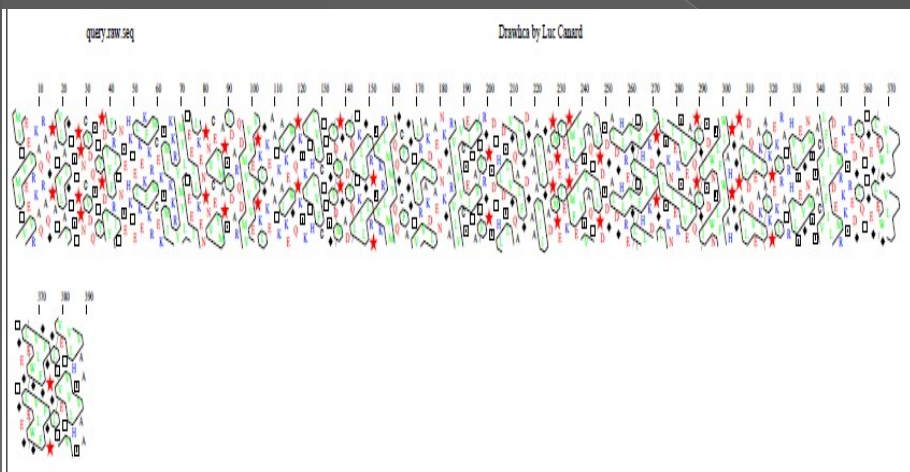Thr:7.000
Trp:1.400
Tyr:5.100
Val:4.500

溶剂可及性越高，越有利于与DNA结合

蛋白质中疏水氨基酸基团分析:
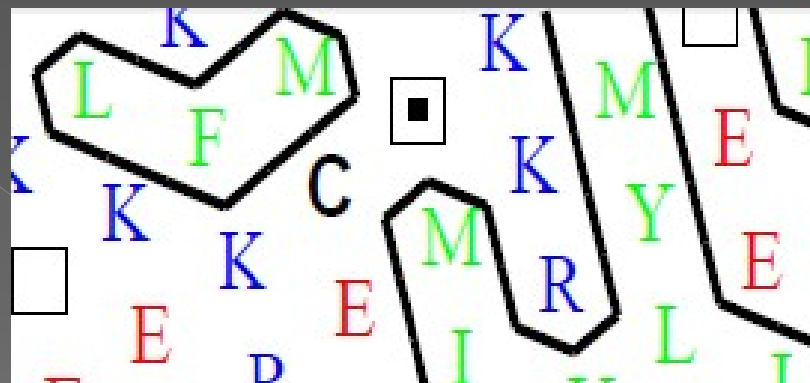
绿色：非极性氨基酸，也是疏水氨基酸。
蓝色：极性氨基酸，带正电的亲水氨基酸。
红色：极性氨基酸，带负电的亲水氨基酸。

蛋白质氨基酸组成分析

ExPASy
Bioinformatics Resource Portal

ProtScale   A.A. composition
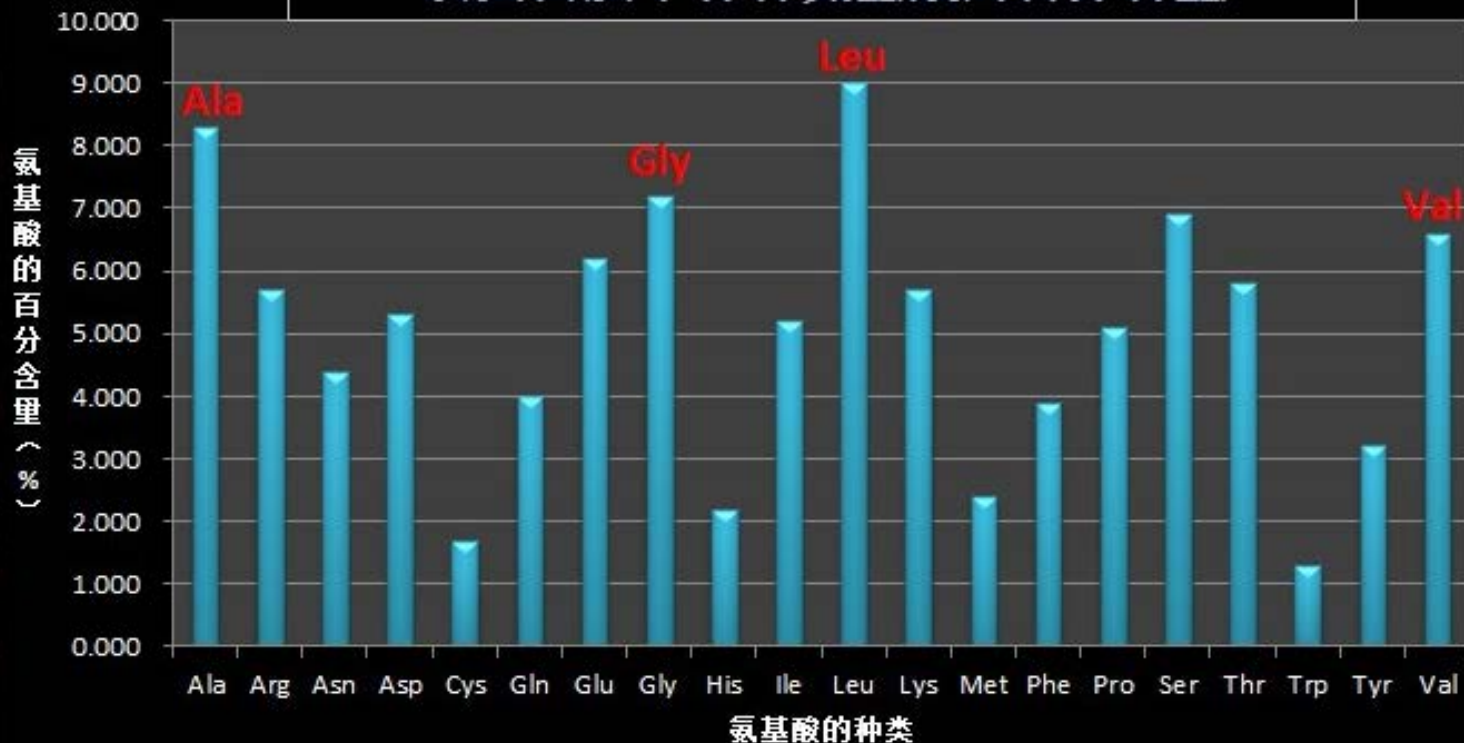
二级结构分析

Service

garnier(v6.0.1) - Predicts protein secondary structure using GOR method

```
            -      10      -      20      -      30      -      40      -      50
        MVTVEEVRKAQRAQGPATVLAIGTSTPPNCVDQSTYPDYYFRITNSEHKT
helix   HHHHHHHHHHHH                                                      HHH
sheet                  EEEEE              EEEEEE    EEEEE
turns               T           TT    TTT          TT
 coil                CCCC           CCC                        CCCC
            -      60      -      70      -      80      -      90      -     100
        ELKEKFKRMCEKSMIKKRYMYLTEEILKENPNVCEYMAPSLDARQDMVVV
helix   HHHHHHHHHHHHHHHHHHHH HHHHHHH          H     HHHHHHHHH
sheet                         E                 EEEE            EEEE
turns                                          TT   T
 coil                                            CC
            -     110      -     120      -     130      -     140      -     150
        EVPKLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGL
helix   HHHHHHHHHHHHHHHHHH                               HHHH
sheet                        EEEEE                           EEEEE
turns                     TTT           TTTTT        T                    T
 coil                     CCC                        CCC                  C
            -     160      -     170      -     180      -     190      -     200
        RPSVKRLMMYQQGCFAGGTVLRVAKDLAENNKGARVLVVCSEITAVTFRG
helix       HH                          HHHHHHHHH  H
sheet          EEEEEEE            EEE              EEEEEEEEEEEEE
turns     T            TTTTT                      T                       T
 coil   CC                         C                                    CC
            -     210      -     220      -     230      -     240      -     250
        PSDTHLDSLVGQQALFGDGAAAVIIGADPMPEIEKPMFEIVSVAQTILPDS
helix                      HHHH         HHHHHHHHHHHHHH
sheet          EEEEEEEEEE        EEEE                           EEEE
turns   T                                                                 T
 coil    CCCC              CC           CCCC                       C  C
            -     260      -     270      -     280      -     290      -     300
        DGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLVEAFQPLGISDWNSLFW
helix       HH      HHHHHHHHH        HHHHHHHHHHHHHHHHH
sheet             E                E                      E               E
turns   T                          TT                        TT  TT     T
 coil    CC    CC                C                          C  CC   CCC
            -     310      -     320      -     330      -     340      -     350
        IAHPGGPAILDQVEAKLALKPEKLRATRHVLSEYGNMSSACVLFILDEMR
helix             HHHHHHHHHHHHHHHHHHHHHH             HH   HHHHHHH
sheet   E         E                                EEE          EE
turns                                                   TTTTT
 coil   CCCCCCC                                           C    C
            -     360      -     370      -     380
        KKSREDGLQTTGEGLEWGVLFGFGPGLTVETVVLHSVAA
```
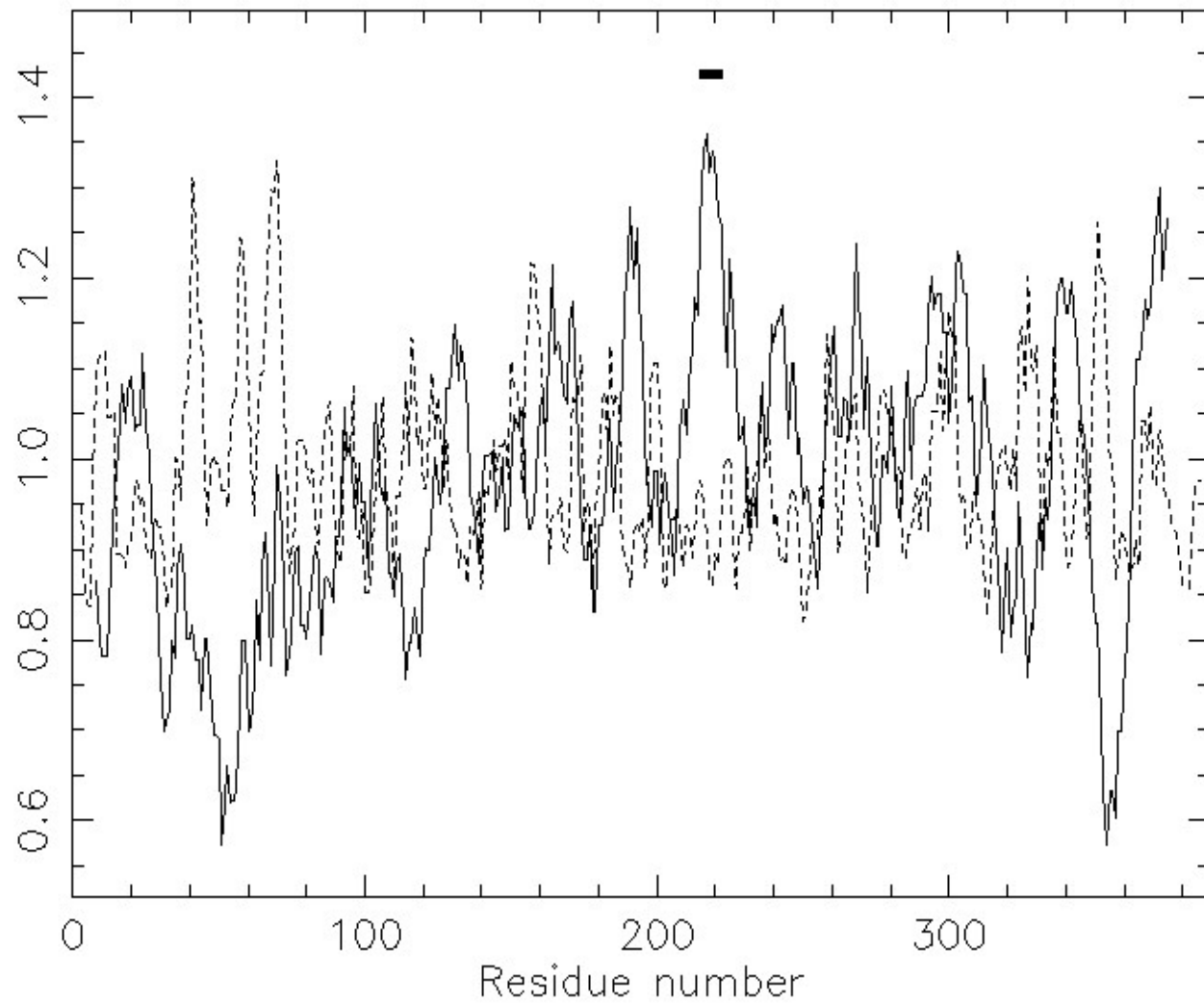
H： α－螺旋

S： β－折叠

T： β－转角

C： 卷曲

跨膜螺旋预测
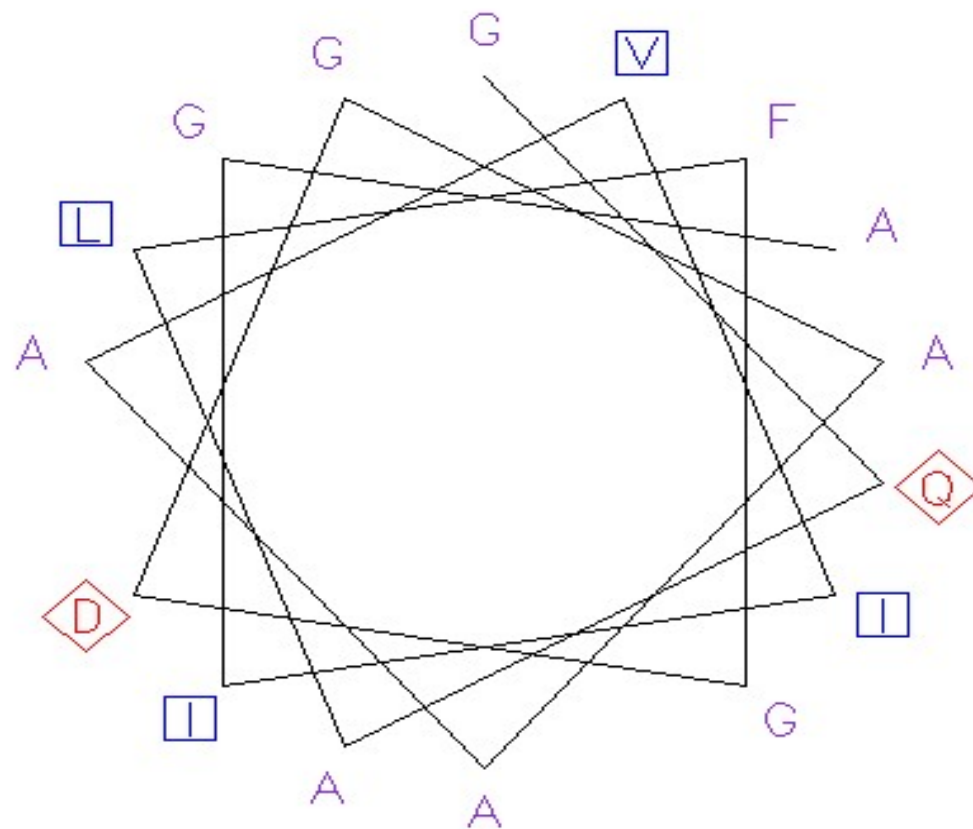
Start
211
End
226
TransMem
1
Sequence
GQALFGDGAAAVIIGA

Tmap

螺旋轮构建

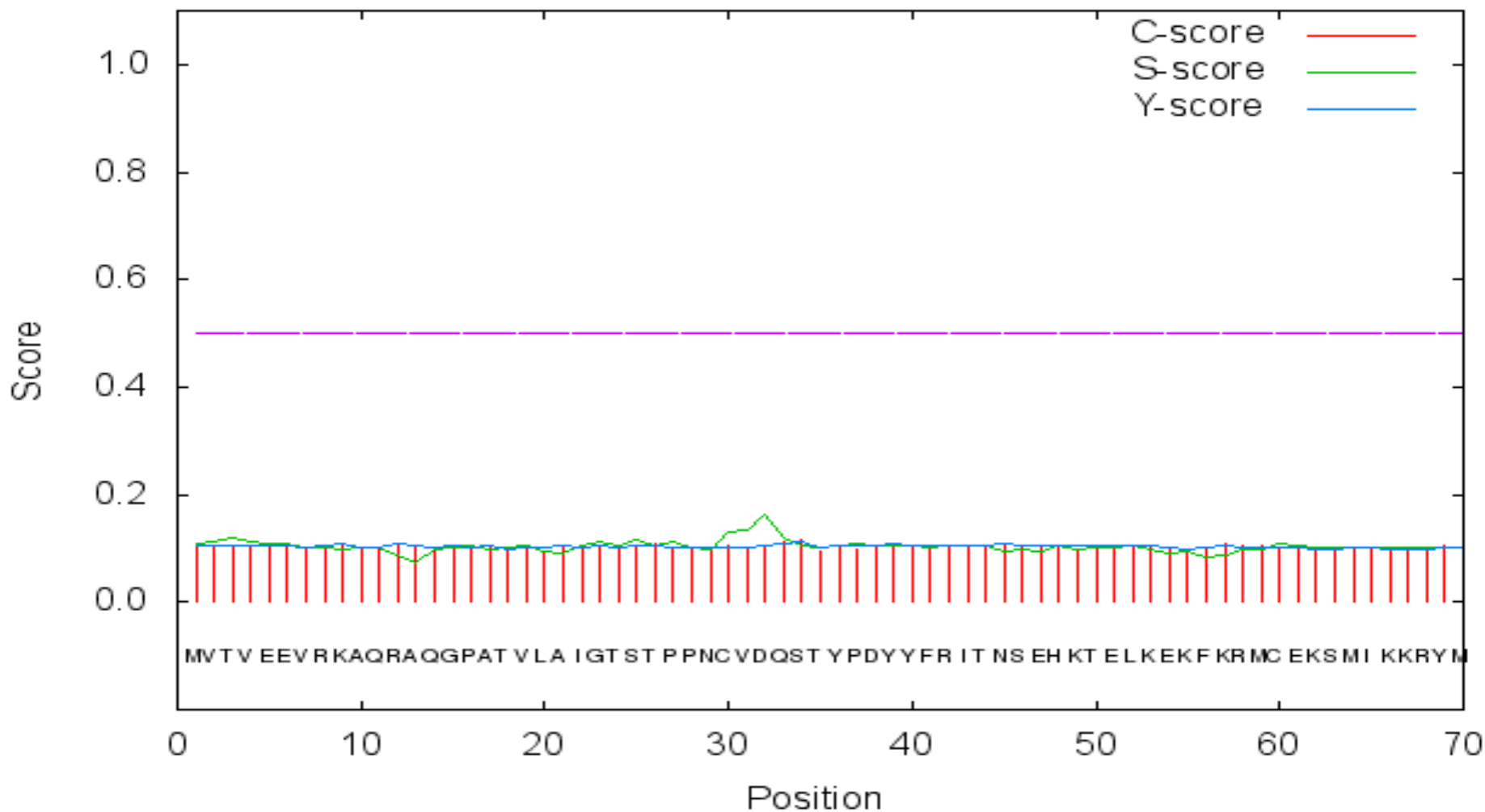Helical wheel of raw::635887
Thu 20 Jun 2013 14:01:57

SignalP-4.1 prediction (euk networks): Sequence

亚细胞定位

TargetP 1.1 Server - prediction results

Technical University of Denmark

```
### targetp v1.1 prediction results ###############################
Number of query sequences:  1
Cleavage site predictions included.
Using PLANT networks.

Name              Len   cTP    mTP     SP   other  Loc  RC  TPlen
----------------------------------------------------------------
Sequence          389  0.029  0.486  0.063  0.671   _    5     -
----------------------------------------------------------------
cutoff                 0.000  0.000  0.000  0.000
```

cTP：叶绿体

mTP：线粒体

SP：关联蛋白
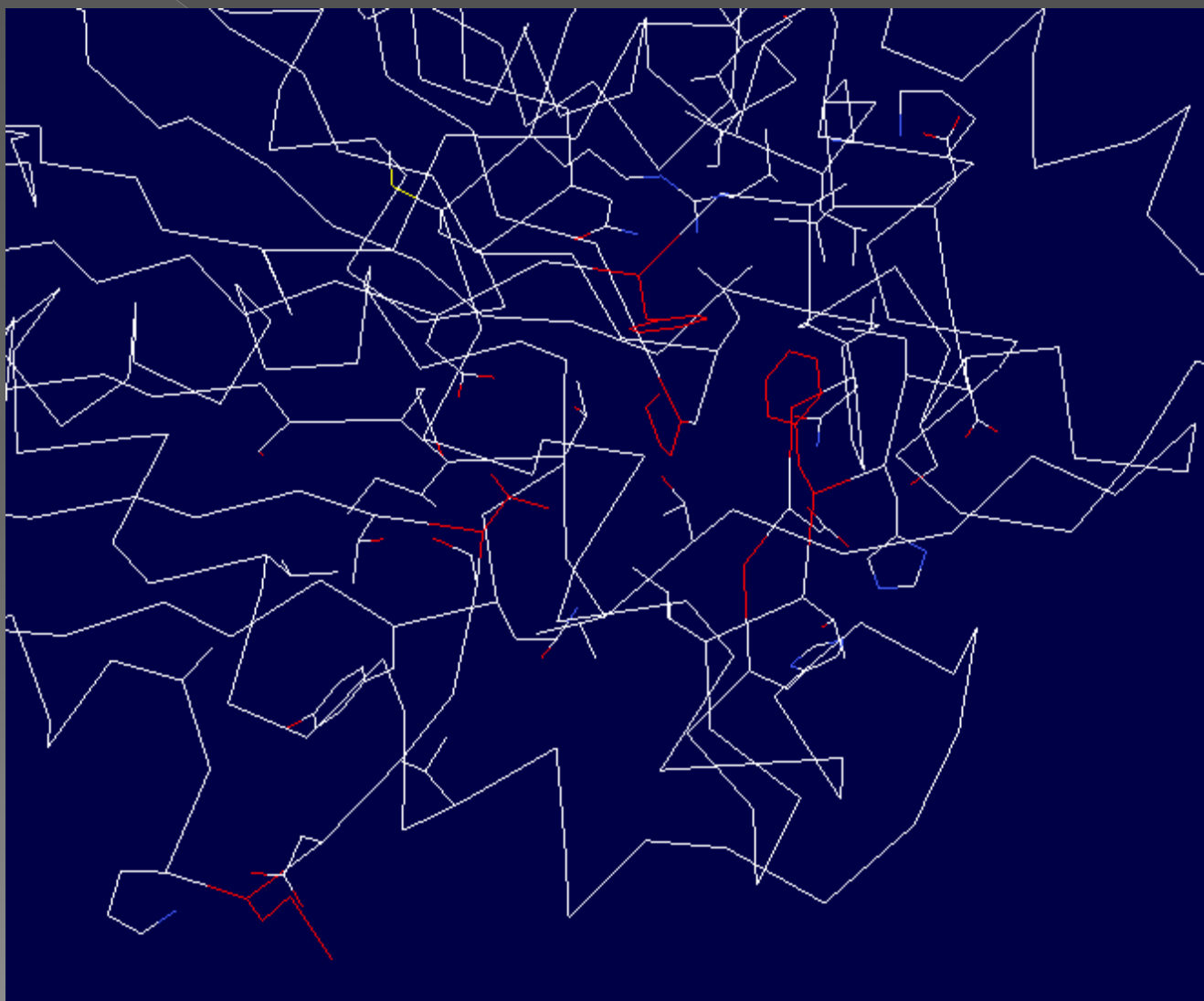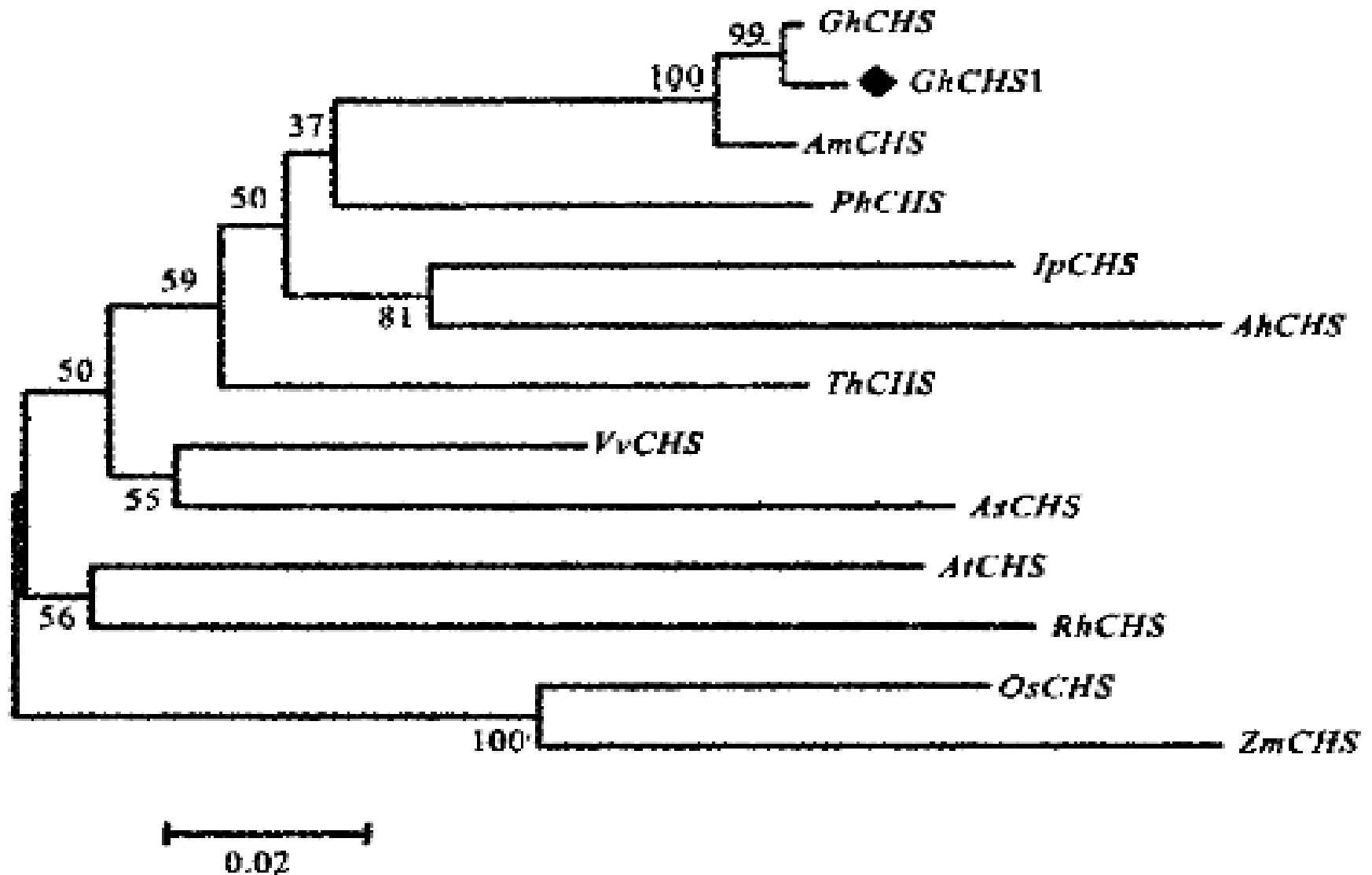
同源建模

四个酶活性位点

Ans336

His303

Phe125

Cys164
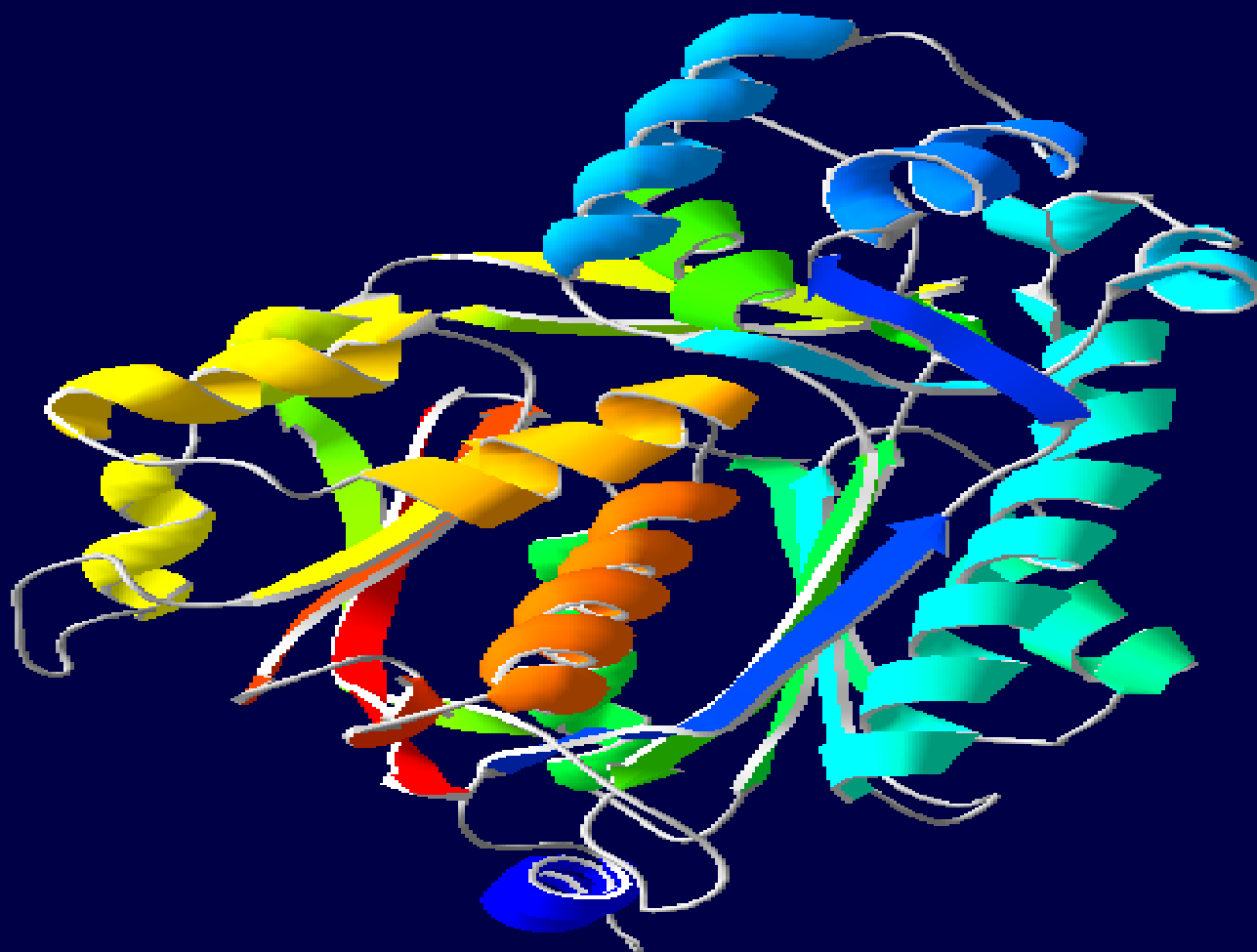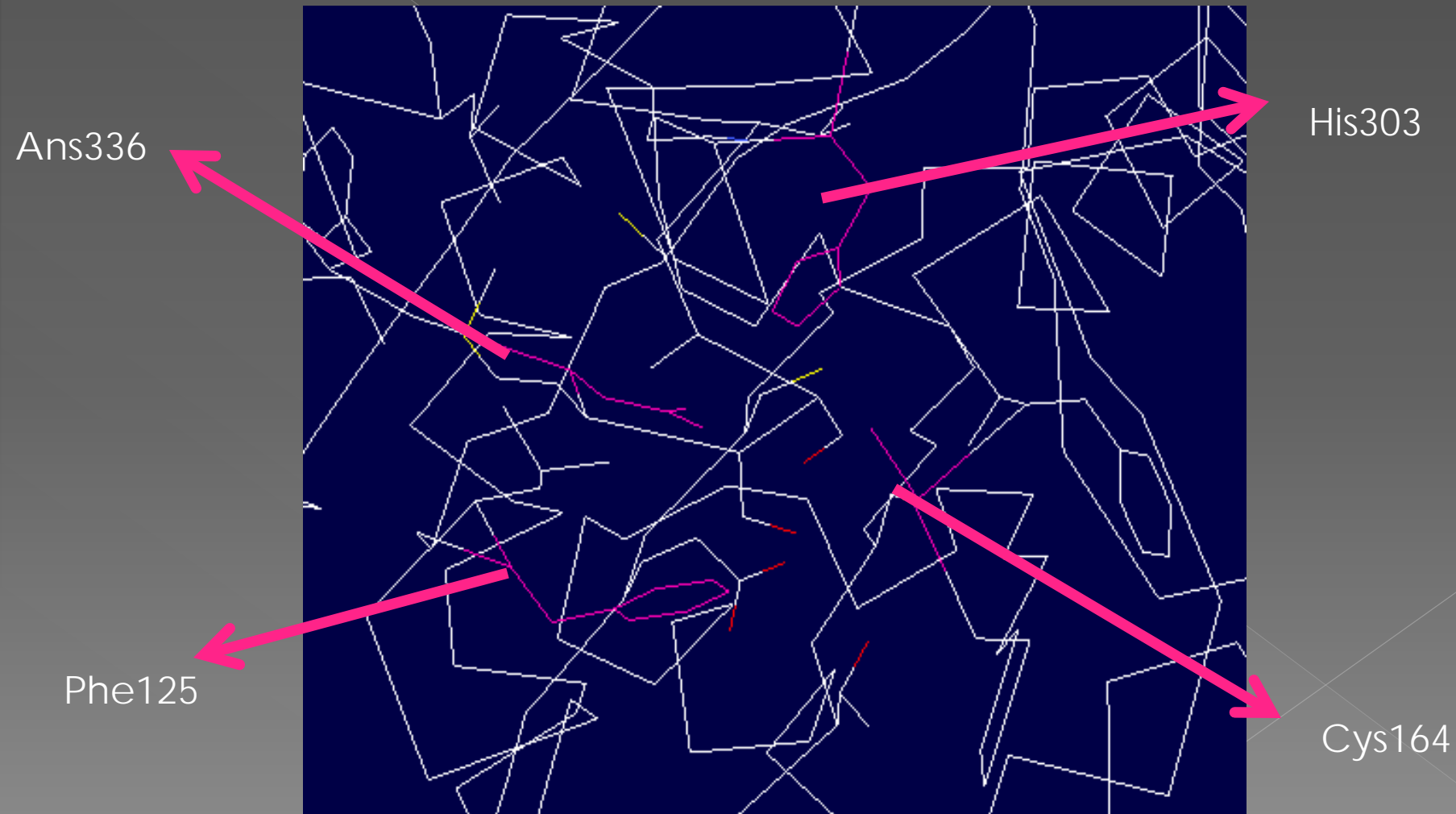
5个组成底物结合口袋位点
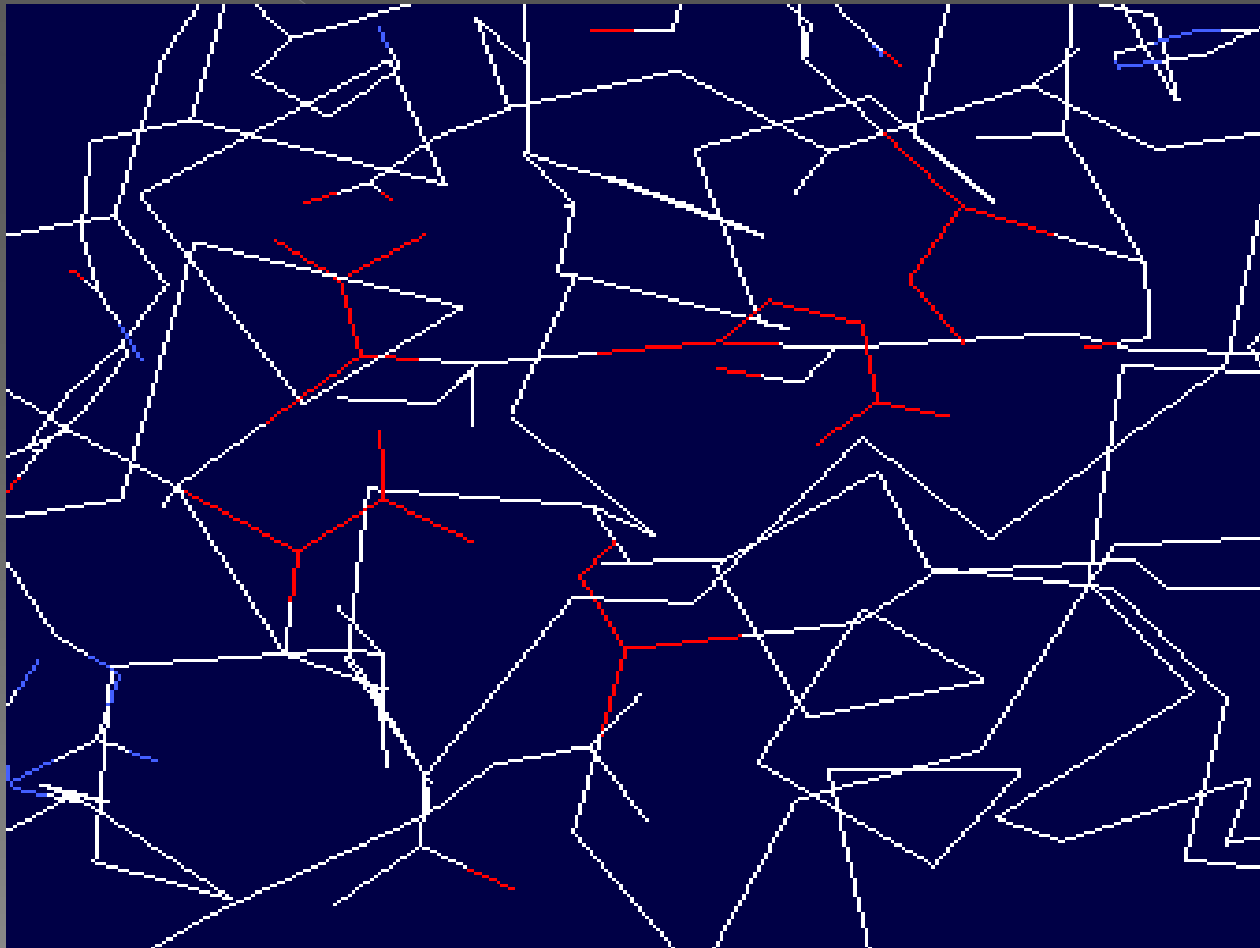
7个组成环化反应口袋的结合位点

构建进化树

CHS在进化上与陆地棉、黄蜀葵等最为接近，与水稻、玉米进化关系最远，可能与棉花和黄蜀葵属于棉葵科而且是双子叶植物 ， 而玉米、水稻是单子叶植物有关