

绿盲蝽气味结合蛋白OBP-7的功能鉴定

Identification of structure of odorant binding protein7 in *Apolygus lucorum*

第八组报告人：朱晓强

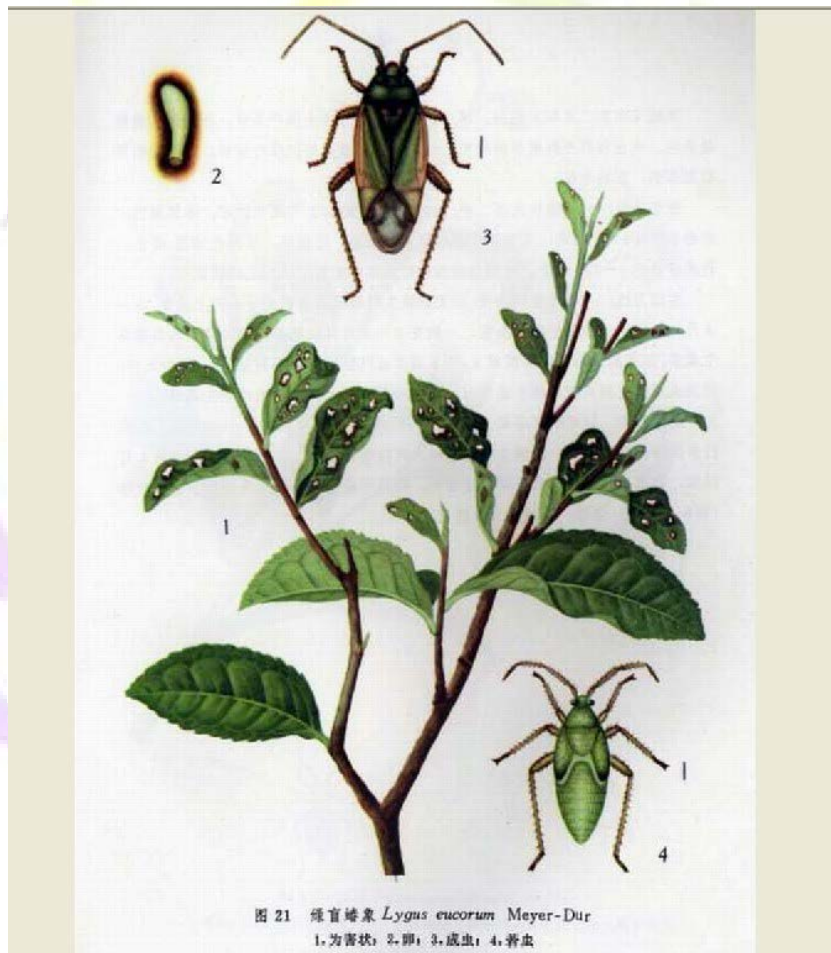
组员：朱晓强 陈金奕 王金鑫 张重阳

植物保护研究所

内容概括

- 1. 研究目的
- 2. 序列获得
- 3. *AlucOBP7*的序列分析
- 4. 序列比对和系统进化树的构建
- 5. *AlucOBP7*的三维结构
- 6. 总结

1. 研究目的



- **OBP(odorant binding protein):** 气味结合蛋白
- **PBP(pheromone binding protein):** 性信息素结合蛋白

2. 序列获得

- **EST**: 基因表达序列标签 (Expressed Sequences Tags, ESTs), 一般是从cDNA文库中随机挑选单克隆进行测序得到的序列片段, 序列长度一般为60 bp-500 bp。与全基因组测序相比, EST序列具有快速、简便、成本低的优势。
- 1993年NCBI专门建立了EST数据**dbEST**(database of EST), 系统的收集和保存所有的EST数据。在dbEST中记录了每个EST的序列号、测序引物、碱基序列、cDNA文库构建方法、组织来源等详细资料。由于EST来源于生物体不同组织或同一组织的不同发育阶段的cDNA文库, 因此通过对这些EST数据进行比较可以发现组织间或同一组织不同发育阶段特异转录表达的基因。

AlucOBP-7的ORF

ATGAACCCACTCATTCTCATACTGTTGGTAGTATTTGC
TGCTGCGACGAGGGGAGAAGAACAAGCCAACGCCT
TGGTAGCCAAAGCTTTCAACAAATGCTTTGGAGAATT
TCCTCTCGGGGACGATGAAATGAAAGAAGTGAAGGA
TAAATCGACTGTCCCAAGCTCACATAATGCCAAATGC
CTCATGGCGTGTCATGTTGAAAGAAGGGAGAATTTTG
AGGGGTGGGAAGTACGAGTTAGAGAATGCAATTTTG
ATGGCTGACGTACTCAACAAAATGATCATGCAGCAA
CCGACAAGGCGAAGCAGTTGATCGAAACGTGTGCT
GCCCAAGTGGGTACTGACGCCAGCGCAGACGAGTG
CGAATTCGCCTACAAAATGGCTCTTTGTGCCTCAGAT
GAAGCTAAGAAGCTTGGAGTTCGTCCGCCAGATTTC
TGA

3. Aluc0BP7的序列分析

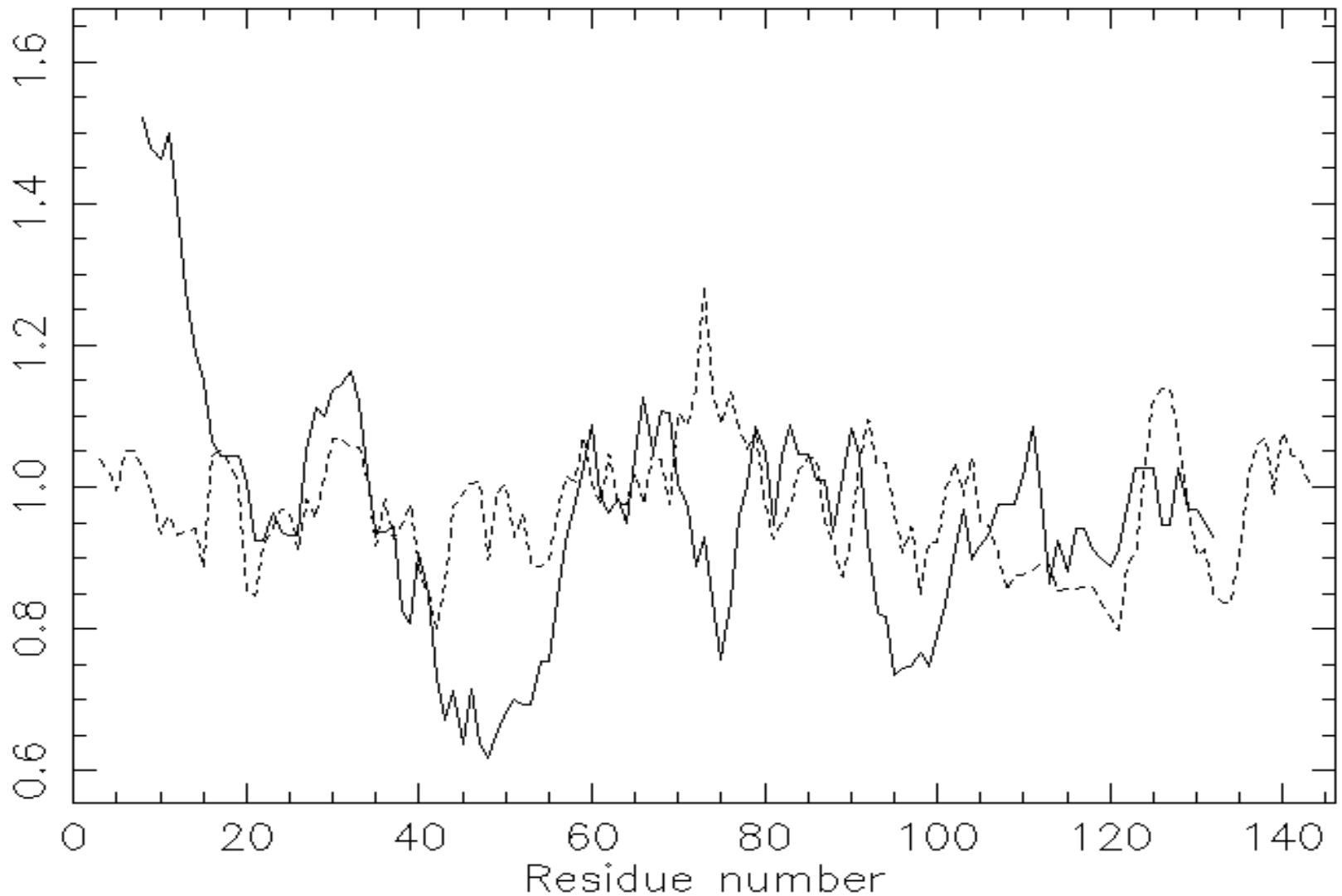
(1).用WebLab的transseq将CDS翻译成氨基酸序列

>_1

```
MNFLILILLVFAAATRGEEQANALVAKAFNKCFGEFPLGDDEMKEVKDKSTVPSSHNAK  
CLMACMLKEGRILRGGKYELENAILMADVLNKNDHAATDKAKQLIETCAAQVGTDASADE  
CEFAYKMALCASDEAKKLGVRPPDF*
```

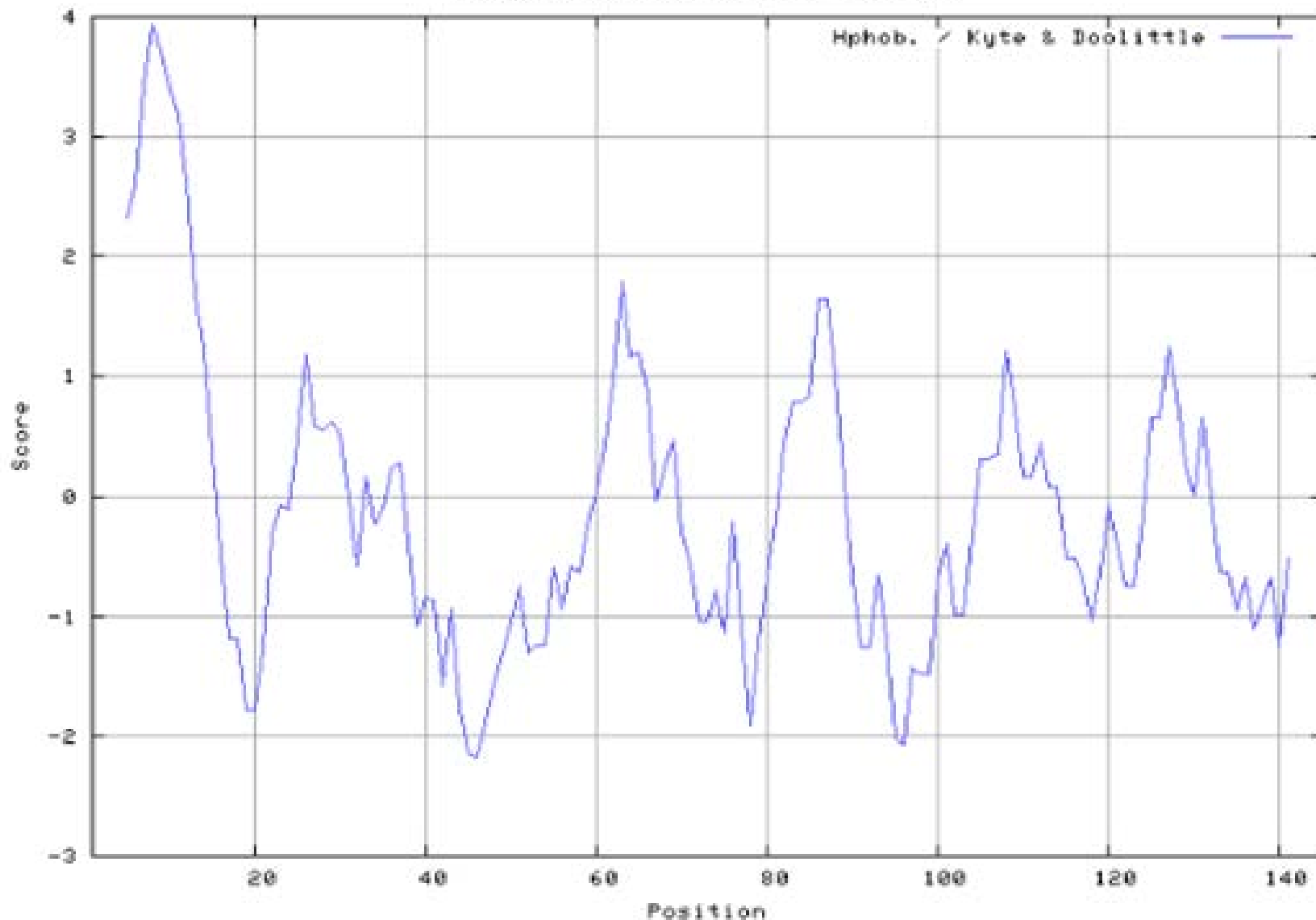
(2). 用WebLab的tmap预测跨膜螺旋

Tmap



(3). 用ExPASy的ProtScale预测疏水性

ProtScale output for user sequence



(4). 用CBS的TargetP进行亚细胞定位



TargetP 1.1 Server - prediction results

Technical University of Denmark

```
### targetp v1.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using NON-PLANT networks.
```

Name	Len	mTP	SP	other	Loc	RC
AlucOBP_7	145	0.038	0.935	0.105	S	1
cutoff		0.000	0.000	0.000		

[Explain](#) the output. Go [back](#).

Results of the k -NN Prediction

$k = 9/23$

```
55.6 %: extracellular, including cell wall  
22.2 %: vacuolar  
11.1 %: endoplasmic reticulum  
11.1 %: mitochondrial
```

>> prediction for QUERY is exc (k=9)

(5). 用WebLab的sigcleave预测信号肽

(1) Score 11.510 length 13 at residues 6->18

Sequence: LILLVVFAAATRG

 | |
 6 18

mature_peptide: EEQANALVAKAFNKCFGEFPLGDDEMKEVKDKSTVPSSHNAKCLMACMLK

(2) Score 7.527 length 13 at residues 4->16

Sequence: LILILLVVFAAAT

 | |
 4 16

mature_peptide: RGEEQANALVAKAFNKCFGEFPLGDDEMKEVKDKSTVPSSHNAKCLMACM

(3) Score 5.113 length 13 at residues 3->15

Sequence: PLILILLVVFAAA

 | |
 3 15

mature_peptide: TRGEEQANALVAKAFNKCFGEFPLGDDEMKEVKDKSTVPSSHNAKCLMAC

(6). 用WebLab的garnier预测二级结构

```
      .   10      .   20      .   30      .   40      .   50
MNPLILILLVVFAAATRGEEQANALVAKAFNKCFGEFPLGDDEMKEVKDK
helix  HHHH      HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH      HHHHHHHHHHHH
sheet  EEEE
turns
coil  CC                                  TTTT          CCCC
      .   60      .   70      .   80      .   90      .  100
STVPSSHNAKCLMACMLKEGRILRGGKYELENAILMADVLNKNDHAATDK
helix  HHHHHHHHHHHHHHHHHHHHH      HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
sheet
turns  T TT                                  T
coil  CCCC C                                  CCCC
      .  110      .  120      .  130      .  140
AKQLIETCAAQVGTDADECEFAKMLCASDEAKKLGVRPPDF
helix  HHHHHHHHHHHHHHH      HHHHHHHHHHHHHHHHHHHHHHHHHHHHH
sheet
turns  T                                  TTT      TTT      EE
coil  CCCC                                  C
```

(7).用SMART预测结构域

SMART

Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864
Letunic et al. (2012) *Nucleic Acids Res*, doi:10.1093/nar/gkr931

SMART MODE:
NORMAL
GENOMIC

Simple
Modular
Architecture
Research
Tool

HOME SETUP FAQ ABOUT GLOSSARY WHAT'S NEW FEEDBACK

The domain within your query sequence starts at position 29 and ends at position 135; the E-value for the PhBP domain shown below is 9.77716539587537e-8.

AFNKCFFGEFFLGDDDEMKEVKDKSTVPSSHNARCLMACMLKEGRILRGKKYELENAILMAD
VLNKNDDHAATDKAKQLIETCAAQVGTASADECEFAKMLCASDEA

BLAST with Domain Align your sequence against the SMART alignment

PhBP
Insect pheromone/odorant binding protein domains.

SMART accession number: SM00708

Description:

Interpro abstract (IPR006625): PhBP is the insect pheromone/odorant binding protein [(PUBMED:7612612)].

GO function: odorant binding (GO:0005549)

Family alignment: View Alignment consensus sequence or Family alignment in CHROMA format

There are 780 PhBP domains in 746 proteins in SMART's nrdb database.

Click on the following links for more information.

- ▶ Evolution (species in which this domain is found)
- ▶ Literature (relevant references for this domain)
- ▶ Structure (3D structures containing this domain)
- ▶ Links (links to other resources describing this domain)

PhBP domain

This is a SMART **PhBP** domain ([full annotation](#)).



Position: 29 to 135

E-value: 9.77716539587537e-8 ([HMMER2](#))

SMART ACC: [SM000708](#)

Definition: Insect pheromone/odorant binding protein domains.

Description:

Interpro abstract ([IPR006625](#)): PhBP is the insect pheromone/odorant binding protein [([PUBMED:7612612](#))].

GO function: odorant binding ([GO:0005549](#))

PhBP domain sequence (107 aa):

[Submit to BLAST](#)

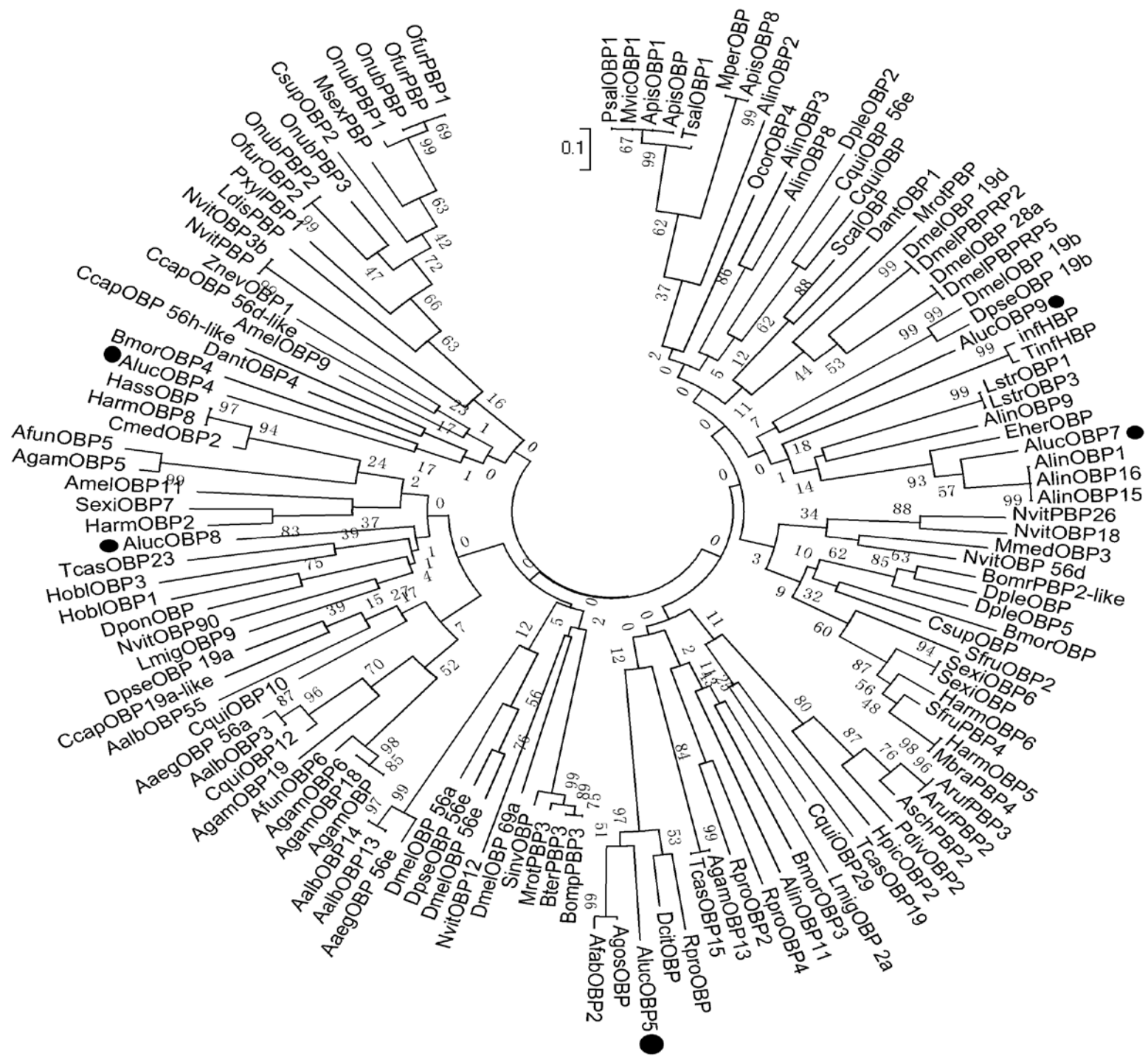
[Align with the SMART alignment](#)

[Copy to clipboard](#)

```
AFNKCFGEFPLGDDEMKEVKDKSTVPSSHNAKCLMACMLKEGRILRGGKYELENAILMAD  
VLNKNDDHAATDKAKQLLETCAAQVGTDADECEFAFKMALCASDEA
```

4. 序列比对和系统进化树的构建

- >AlucOBP-4
- MEVAACLVLALLAALTAAVEEGRPLCKAPTTAPRKLEKVINQCQEEIKYALLQEAPSV
LGETVGLKTALTRNRSKRETFTGEERRIAGCLLQCVYRKMALDETGFPTATGLVKIY
SEGVEDRNYYLATIQGVQRCLSRELQSRNTNPSIVKAEGYSCDVAYDMFNCVSEQIE
QLCGTSP
- >AlucOBP-5
- MNSIIVLCLVASAVTLSQGNPTTPNPSTSHVSSSAGITVSGVSKSPEEIKLKIKEQVATL
TGACKTQTKLTGEQAKIVASQAIPKTEAEKCFLECIYQGLQLTKDGKFNPAARAWAQ
KRFGNAPEDLQKANTMIDICVKEVVVDENEKCALGRLIRECFVKNGAKINFFPKP
- >AlucOBP-7
- MNPLILILLVFAAATRGEEQANALVAKAFNKCFGEFPLGDDEMKEVKDKSTVPSSH
AKCLMACMLKEGRILRGGKYELENAILMADVLNKNKNDHAATDKAKQLIETCAAQVGT
ASADECEFAKMLCASDEAKKLGVRPPDF
- >AlucOBP-8
- MVLKMKQILVVFVALQVLISTTEAVMTQAQMKQAMKTVRNMCIKSGVDKEALAKMV
EGEFDESQKLCYLGCVLGMMQAVKNNKINLTMVKNQISKMLAPEQQQRILAAFE
GCATVTGDDNCDLAFKFAKCIYDTDKELLFQAFIVP
- >AlucOBP-9
- MKSFVGLIFAVALVEFASAITKEYHDRAVAAKDAKCLKKHPSIKESDVQEFLLKHKLPET
DDGKCMACIACMEEMNLNADGKINVEEAKKTNSDKYDGEPPDNKELADKLIDHCSSQV
SPDGMSKCEYAYQISKCGLEYGMKNGLTPPKMYEEQRR



进化树

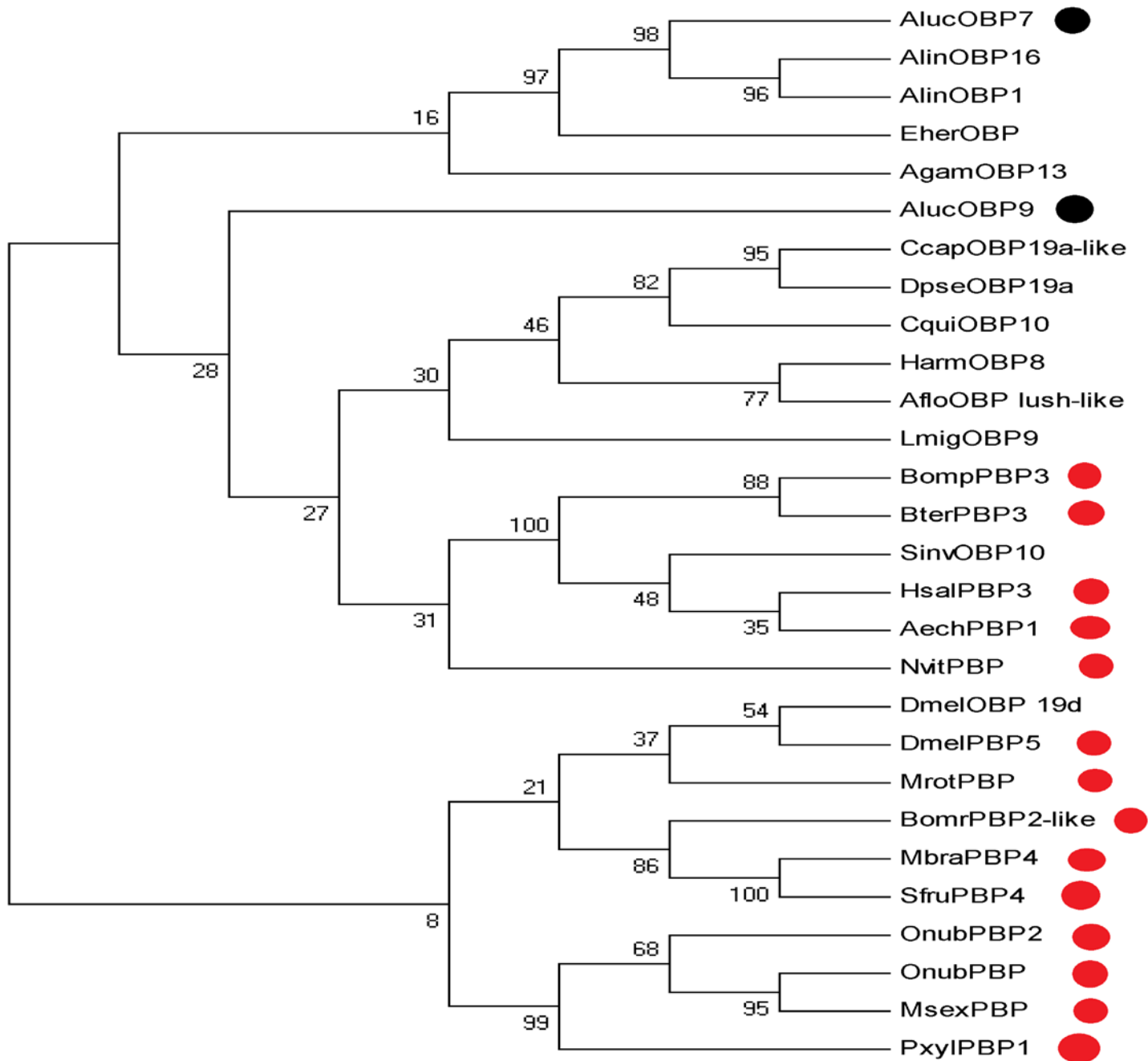
Protein Sequences

Species/Abbrv	*	*	*
1. AlucOBP7	NALVAKAFNKCFGEFPL	GDEEMKSVKDKSIV--FSHNAKCLMACMLKEGRILRGGKY	
2. AgamOBP13	ENAKKLMTECKLKV	GADADVSEALKMHQMFEFEFGFCMLFCVFDLAKIML-DGK	
3. DmelOBP_19d	RDFAAELANECAKAEI	GADDEVSEFLMSHDLFERHEAKCLLFAACVMKKLIMDESGL	
4. AlucOBP9	HDFAVAAKDAKCLKKHPFSIK	-ESLVEEFLKKHKLLEFETDDGKCMIAICYMEEMNLMA-DGKI	
5. DmelPBP5	--LMESAESCMPSEV	GADADLDEMVKKQEAFTYAGKCLLFAACVMKNIGIL-DANGK	
6. CcapOBP19a-like	IAAGKLMRDVCLPKFF	KVPLDVADEGKDGIVFEIKVEVKCYINCIEMMQLMK	
7. CquiOBP10	LKSGMMRSVCMYKA	KPSLEQVEGLPSGKFFFEKKEIMCYANCVLELMQGMK	
8. HarmOBP8	RNLGKMMRKSCKFKV	NVADEEIDPIAEGVFNEDKEVKCYMACIMKMANIK	
9. DpseOBP19a	VSAGKLMRDVCLPKYF	KVTVFVADGIRNGIFNNKDSNCYINCVLEMMAIK	
10. EherOBP	KAKIMKVFNTCKEKHPV	TDDELAAFR-KADIGFGYSHDAKCLMACMLEEGKMLK-DGKY	
11. BompPBP3	IAIAAASVVNACQIQI	GVAIVDI EAVRNGQWFEIRQLKCYMYCLWEQFGLVDDKRE	
12. SinvOBP10	IAIAAASVVNACQIQI	GVAIVDI EAVRNGQWFEIRQLKCYMYCLWEQFGLVDDKRE	
13. LmigOBP9	RGLKIVRNMCLKKT	GVDLALIVEGIEGQFEDVDLKCVMCGAMVLR	
14. BomrPBP2-like	KIRFKLVMKCKKHPVDMSS	--SLMLLQQLIAFKKTESKCLLACAYKLVGVMFSGGI	
15. BterPBP3	IAIAAASVVNACQIQI	GVAIVDI EAVRNGQWFEIRQLKCYMYCLWEQFGLVDDKRE	
16. HsalPBP3	IAIAAASVVNACQIQI	GVAIVDI EAVRNGQWFEIRQLKCYMYCLWEQFGLVDDKRE	
17. AechPBP1	IAIAAASVVNACQIQI	GVAIVDI EAVRNGQWFEIRQLKCYMYCLWEQFGLVDDKRE	
18. AfloOBP_lush-like	EKLAKMRRKSCVQKI	AILEELVGMRRGFEFDDHDLQCYITCIMKLLRFK	
19. MrotPBP	--LMKSAIQACAKSEVGYIDGNAFSIFDKEMVLG	--FENATCLFAACVMKMGAFMHHKIKI	
20. NvitPBP	GAMVVGVBDKCHREIG	VDI EHVDRIVEGYFHFSELLGQCYFSCIFNHFNLLDNDG	
21. OnuhPRP2	KLRFKAVFVCAKSEVDTDFAAAGATMNFVKGVVV	--DFAGATLFTSSTKLTITLRFSG	

Species/Abbrv	*	*	*
1. AlucOBP7	NDHAAIDRAKQLIETCAAQVGI--DA--ADECEFAFKMALCASDEAKKRLGVREFDF		
2. AgamOBP13	--FKFLIGDDKAKLESLKLSAICEELGDG	EDKCEIAKRLVECVIKKNGKIHGFVFFERE	
3. DmelOBP_19d	DAEKED--DAPA--EVVAKCEAIEIF	DDHCDAAFAYEECIYEQMKEHGLELEH	
4. AlucOBP9	EPDNRKELADKLIIDHCSSQVSP	DGMKCEYAYQISKGLLEYGMKNGLIHKMYEEQR	
5. DmelPBP5	EKAKQYVGNPAKLIKIALDIGETCAAILV	DDHCEAAEAYGTICFRGEAKKHGLL	
6. CcapOBP19a-like	DLLMFPDDYKDQY--KSGFAACKDSPNGIKN	NCDASYALLICMR	
7. CquiOBP10	DLLLIPPEYAEPT--KRAFDIRHAGDGVKN	NCEVAYALLKCLH	
8. HarmOBP8	DLLLFPDDIKEP--AKEAITACRKYADAYK	DICDASFHIKCIY	
9. DpseOBP19a	DVLLFPDSYKPEY--RQGINNCKDATAGLKNVA	FNCDFAHALLSCLKNLK--VFVHF	
10. EherOBP	--DDFDEAAKARKVIENCSEVVS--EVGDEQCEFAFKMAVCGANESKLLGMKEHDFFE		
11. BompPBP3	--MLLFFFRIPAYFAEVKAISECKGIGKYLAKG	DNCEYAYRFNKCAYAESSPRTYYLF	
12. SinvOBP10	--MLLFFFRIPAYFAEVEKAISECKGLGNYLAKG	DNCEYAYAFNKCAYALSPRTYYLF	
13. LmigOBP9	DKMLFPDLKGRFID--MLDACSFRGDGVDD	DCEMAYGLIKCSYETDKEIFLFF	
14. BomrPBP2-like	--DEKFLKNGKKVADICVAVNDVEVSDGEGKGEFAALIFKCTLENAPKFGFKV		
15. BterPBP3	--MLLFFFRIPAYFAEVKAISECKGIGKYLAKG	DNCEYAYRFNKCAYAESSPRTYYLF	
16. HsalPBP3	--MLLFFFRIPAYFAEVEKAISECKGI--AKG	DNCEYAYRFNKCAYALSPRTYYLF	
17. AechPBP1	--MLLFFFRIPAYFAEVEKAISECKGI--AKG	DNCEYAYAFNKCAYALSPREDLFFI	
18. AfloOBP_lush-like	EILMP--PEEVAIGKEIVEACR--NEEYLG	DDCQKTYQYVQCHYKQFPEKFFFF	
19. MrotPBP	IFLGVVHGENPELLKLEKGAECIDKVVDD--IADECKKAYVYIQFLDKH		
20. NvitPBP	DVVKVVVIPPSFKDHADEMIAACKLILGK	DPCDASVIVQCFKILPAVRIMYKRVPIIFLLI	
21. OnuhPRP2	--DAMAAHHTVTVHACR--VPR--RNTMATGTMTMFRFTLTNAARFTMTFMMAHMA		

去除信号肽的28条不同昆虫OBP序列比对

DDHCEAAEAYGTICFRGEAKKHGLL 详细>
 没有汉英互译结果
 请尝试网页搜索



去除信号肽后构建系统发育树

5. Swiss-model 进行三维结构预测，
并用SPDBV软件进行三维结构分析

[myWorkspace]

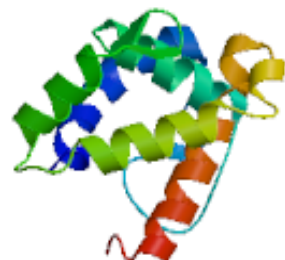
[login]

Workunit: P000001 Aluc-obp-7 - Overview



Print/Save this page as

Model Summary



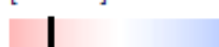
Model information:

Modelled residue range: 21 to 143
 Based on template: [2wcjA] (1.40 Å)
 Sequence Identity [%]: 20.47
 Evaluate: 5.50e-19

Quality information:

QMEAN Z-Score: -2.53

[details] ▾



Quaternary structure information: [details] ▾

Template (2wcj): MONOMER
 Model built: SINGLE CHAIN

Ligand information:

[details] ▾

Ligands in the template: M21: 1, MG: 1.
 Ligands in the model: none.

logs: [Templates] ▾ [Alignment] ▾ [Modelling] ▾

display model: as [pdb] ▾ - as [DeepView project] ▾ - in [AstexViewer] ▾

download model: as [pdb] ⚡ - as [Deepview project] ⚡ - as [text] ⚡

Global Model Quality Estimation [+/-]

QMEAN4 global scores:

QMEANscore4

Estimated absolute model quality

Score components

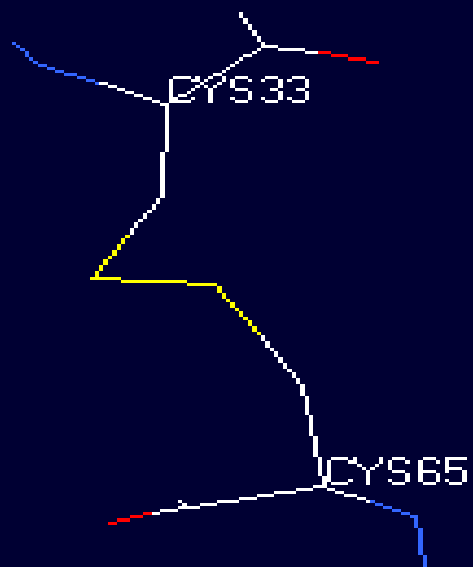
Local scores

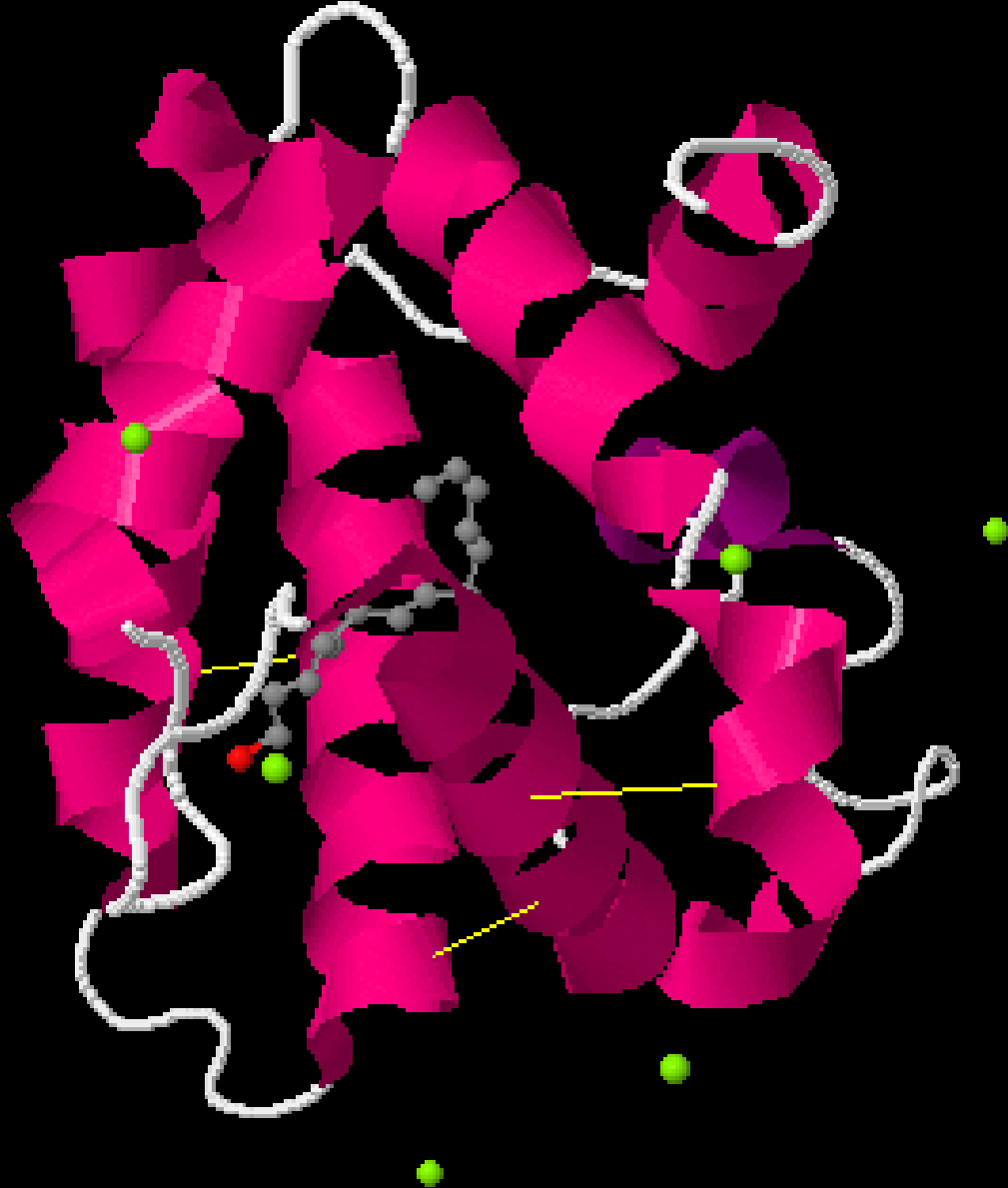
Coloring by residue error

Residue error plot









家蚕的 PBP1 的三维 结构图



6 总结

- 首先，必须感谢罗老师，将我们领进了生物信息学的殿堂，为我们开启了这个浩大的图书馆的大门。
- 生物信息学的应用会让我们的实验事半功倍，所以我们今后应该多加学习！