

# 结核分枝杆菌Rv1636蛋白的生物学信息学分析

汇报人： 崔保亮

17组成员： 崔保亮  
郭立平  
王香玲  
赵利明

# 汇报提纲

- 1.背景知识介绍
- 2.mRNA的分析
- 3.氨基端序列分析
- 4.蛋白质结构预测
- 5.参考文献
- 6.致谢

# 1.背景知识介绍

- 结核分枝杆菌(*Mycobacterium tuberculosis*) 是重要的全球性致病菌，由其引起的结核病是世界上致死率最高的传染病之一。
- 结核分枝杆菌H<sub>37</sub>Rv菌株是研究结核分枝杆菌的标准菌株。

- 1998 年，H<sub>37</sub>Rv 菌株全基因组测序完成，并发现了 **PE** 和 **PPE** 两大富含甘氨酸的蛋白家族，推测其功能可能与结核分枝杆菌抗原变异和干扰、抑制抗原递呈过程相关，具有极为重要的免疫学意义。

- Rv1636基因所编码的Rv1636蛋白属于普通应激蛋白A家族。
- 当结核杆菌处于应激条件下，该蛋白的表达会加强，它会增加结核杆菌在恶劣条件的存活几率，而且可能使结核杆菌具有“应激耐受”活性。

- 本报告通过对**Rv1636**进行生物信息学和结构的初步分析，以期能够对进一步深入研究该蛋白的免疫学特性，开发更有效的基于复合抗原的结核病免疫学诊断试剂，以及新型疫苗和新型药物研究提供理论依据。

# 2.mRNA分析

在GeneBank中找到Rv1636的mRNA序列

## Mycobacterium tuberculosis H37Rv complete genome; segment 6/13

GenBank: BX842577.1

[GenBank](#) [Graphics](#)

>gi|38684030:111194-111634 Mycobacterium tuberculosis H37Rv complete genome; segment 6/13

```
ATGAGCGCCTATAAGACCGTGGTGGTAGGAACCGACGGTTCGGACTCGTCGATGCGAGCGGTAGATCGCG  
CTGCCCAGATCGCCGGCGCAGACGCCAAGTTGATCATCGCCTCGGCATACCTACCTCAGCAGGAGGACGC  
TCGCGCCGCGGACATTCTGAAGGACGAAAGCTACAAGGTGACGGGCACCGCCCCGATCTACGAGATCTTG  
CACGACGCCAAGGAACGAGCGCACAAACGCCGGTGGCAAAAAACGTCGAGGAACGGCCGATCGTCGGCGCCC  
CGGTCGACCGGTTGGTGAACCTGGCCGATGAGGAGAAGGCGGACCTGCTGGTCGTGGCAATGTCGGTCT  
GAGCACGATCGCGGGTCGGCTGCTCGGATCGGTACCGGCCAATGTGTACGCGCGGGCCAAGGTCGACGTG  
CTGATCGTGCACACCACCTAG
```

# 用WebLab中的cusp软件显示密码子使用偏好性

- #CdsCount: 1

#Coding GC 63.72%

#1st letter GC 67.35%

#2nd letter GC 42.86%

#3rd letter GC 80.95%

- #Codon AA Fraction Frequency Number
- GCC A 0.583 95.238 14
- GAC D 0.833 68.027 10
- ATC I 0.889 54.422 8
- AAG K 0.875 47.619 7
- CTG L 0.583 47.619 7
- GTC V 0.412 47.619 7
- GTG V 0.412 47.619 7



### 3.氨基酸序列分析

- 在Uniprot中找到Rv1636的氨基酸序列

Sequence	Length	Mass (Da)	Tools
006153 [UniParc].	FASTA	146	15,312
Last modified July 1, 1997. Version 1. Checksum: 23FDE515DE3EF4ED			
<pre>      10      20      30      40      50      60 MSAYKTVVVG TDGSDSSMRA VDRAAQIAGA DAKLIIASAY LPQHEDARAA DILKDESYKV        70      80      90     100     110     120 TGTAPIYEIL HDAKERAHNA GAKNVEERPI VGAPVDALVN LADEEKADLL VVGNVGLSTI       130     140 AGRLLGSVPA NVSRRRAKVDV LIVHTT</pre>			

# 用Protparam对氨基酸序列进行理化性质分析

Number of amino acids: 146

Molecular weight: 15312.3

Theoretical pI: 5.51

Amino acid composition: [CSV](#)

Ala (A)	24	16.4%
Arg (R)	8	5.5%
Asn (N)	5	3.4%
Asp (D)	12	8.2%
Cys (C)	0	0.0%
Gln (Q)	2	1.4%
Glu (E)	8	5.5%
Gly (G)	10	6.8%
His (H)	4	2.7%
Ile (I)	9	6.2%
Leu (L)	12	8.2%
Lys (K)	8	5.5%
Met (M)	2	1.4%
Phe (F)	0	0.0%
Pro (P)	5	3.4%
Ser (S)	9	6.2%
Thr (T)	7	4.8%
Trp (W)	0	0.0%
Tyr (Y)	4	2.7%
Val (V)	17	11.6%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 20

Total number of positively charged residues (Arg + Lys): 16

**Atomic composition:**

Carbon	C	667
Hydrogen	H	1101
Nitrogen	N	193
Oxygen	O	214
Sulfur	S	2

**Formula:** C<sub>667</sub>H<sub>1101</sub>N<sub>193</sub>O<sub>214</sub>S<sub>2</sub>

**Total number of atoms:** 2177

**Extinction coefficients:**

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

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

Ext. coefficient            5960

Abs 0.1% (=1 g/l)        0.389

### 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 25.53  
This classifies the protein as stable.

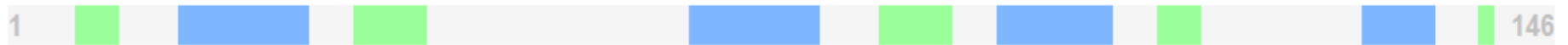
Aliphatic index: 106.30

Grand average of hydropathicity (GRAVY): 0.005

**ProtParam**没有考虑蛋白质翻译后修饰、蛋白质多聚体等情况，故用户在预测和分析此类特定蛋白质的基本理化性质时需要仔细审视反馈结果。

# 氨基酸序列的二级结构

## Secondary structure

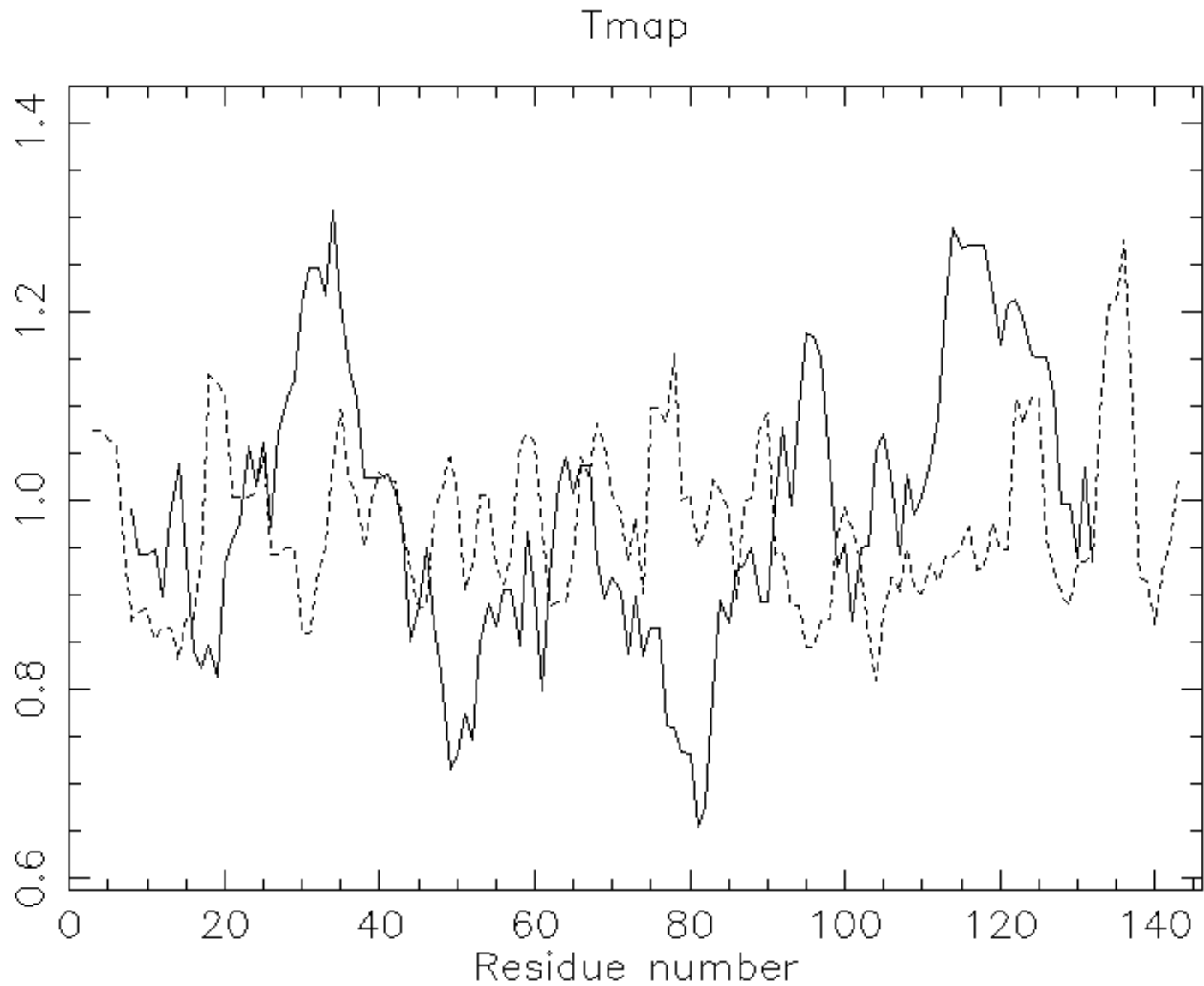


■ Helix ■ Strand ■ Turn

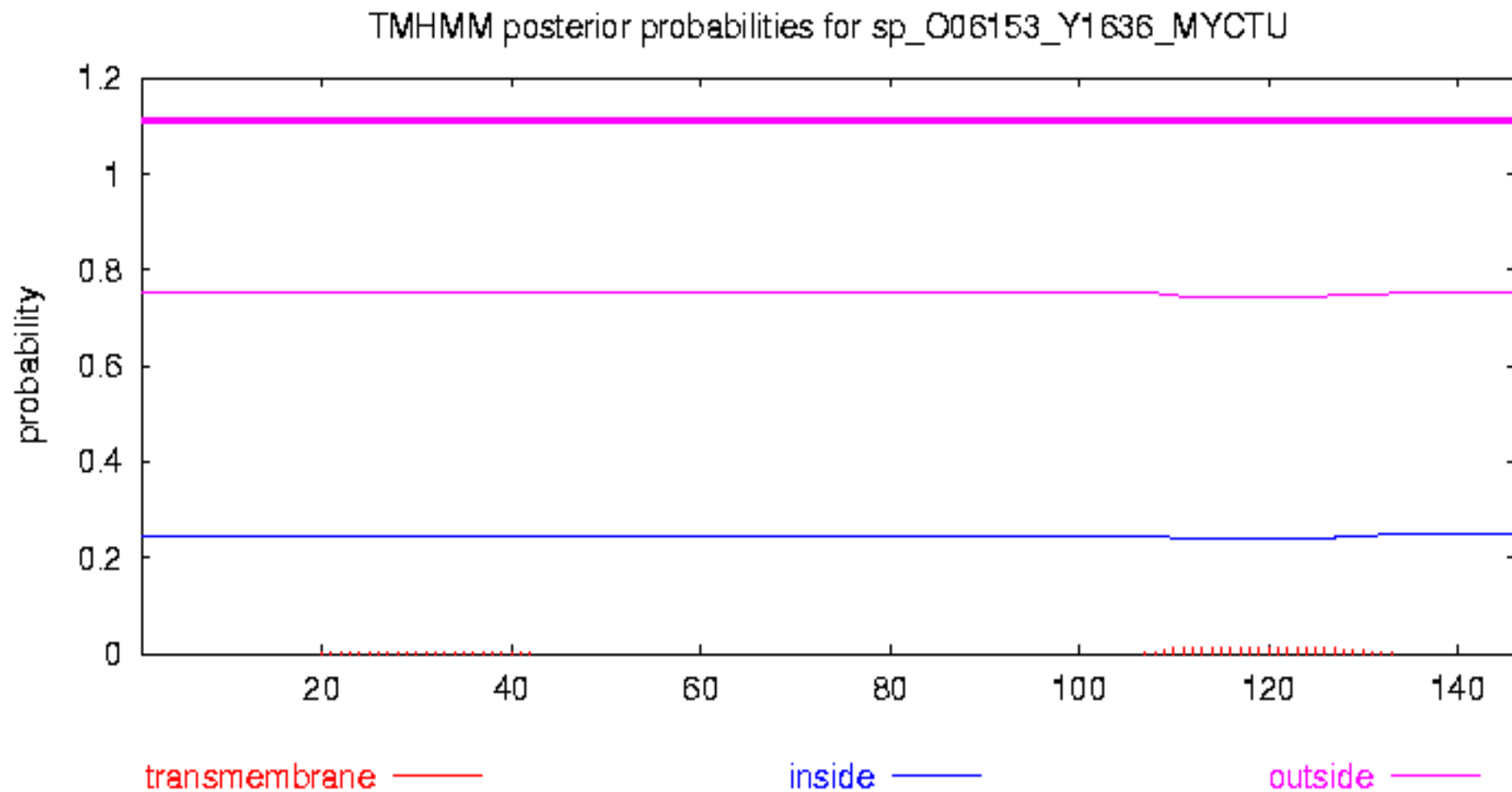
[Details...](#)

<input type="checkbox"/>	Beta strand	6 – 9	4	
<input type="checkbox"/>	Helix	15 – 27	13	
<input type="checkbox"/>	Beta strand	32 – 39	8	
<input type="checkbox"/>	Helix	66 – 78	13	
<input type="checkbox"/>	Beta strand	84 – 90	7	
<input type="checkbox"/>	Helix	94 – 104	11	
<input type="checkbox"/>	Beta strand	108 – 112	5	
<input type="checkbox"/>	Helix	128 – 135	8	
<input type="checkbox"/>	Beta strand	139 – 142	4	

# Tmap进行跨膜螺旋预测



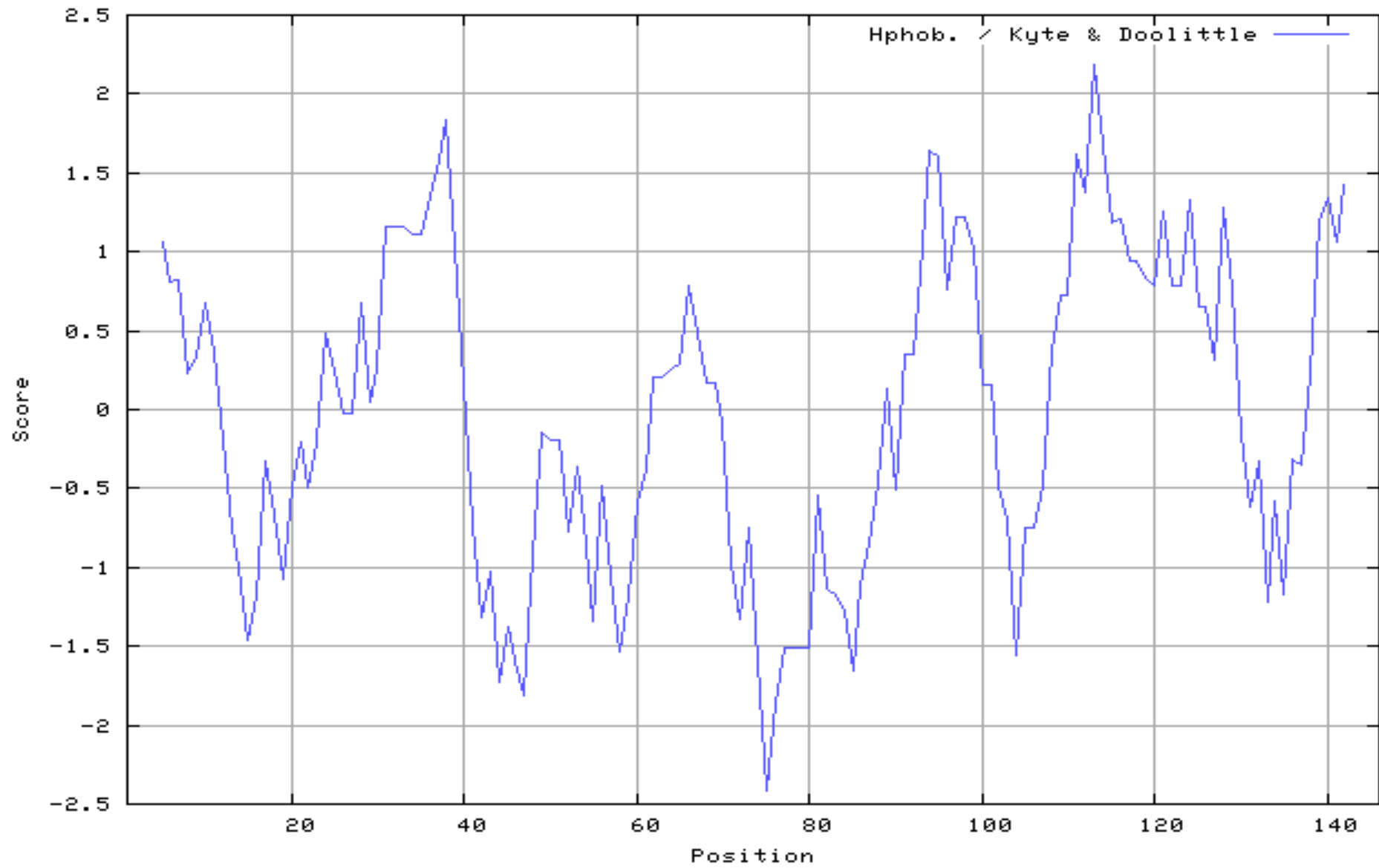
# TMHMM 预测跨膜螺旋



- 显示没有跨膜螺旋！

# ExPASy进行亲水性分析

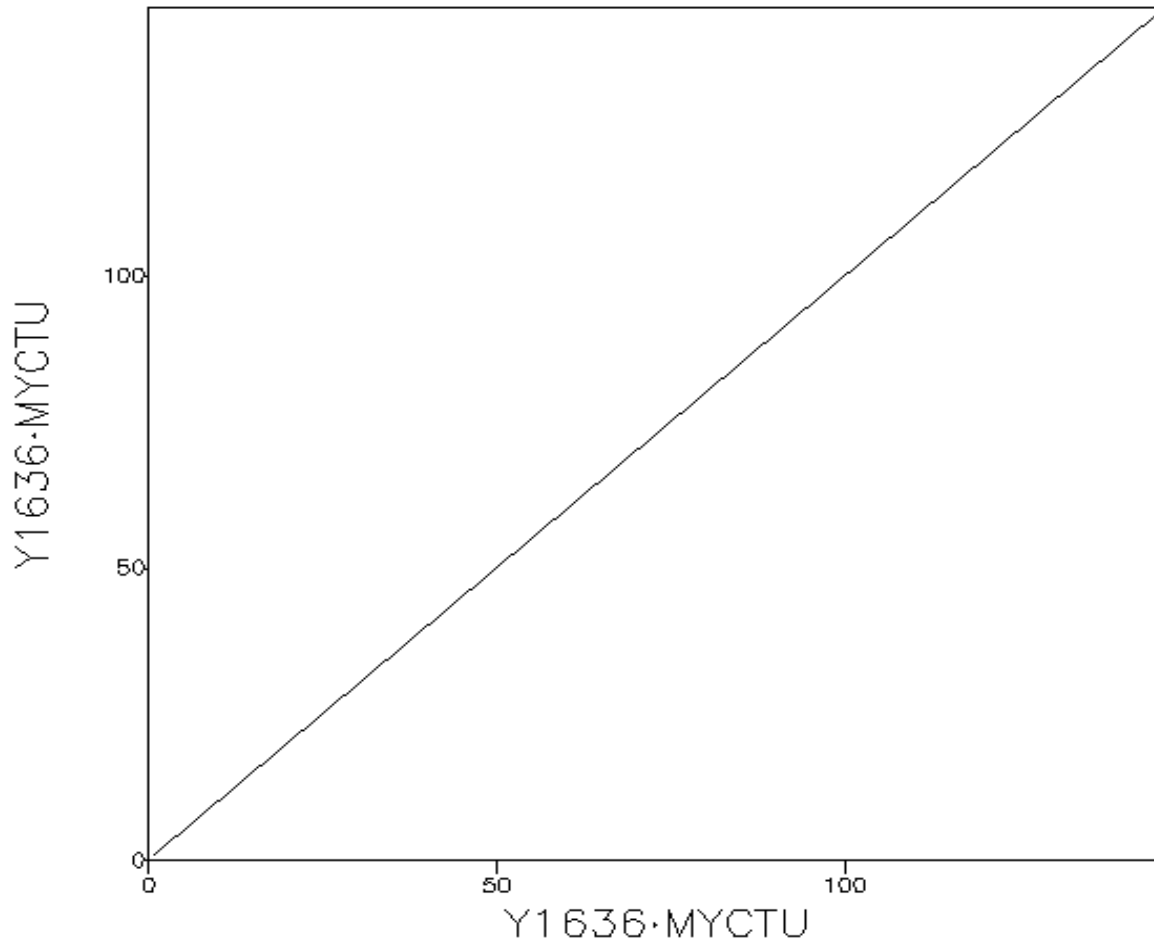
ProtScale output for user sequence





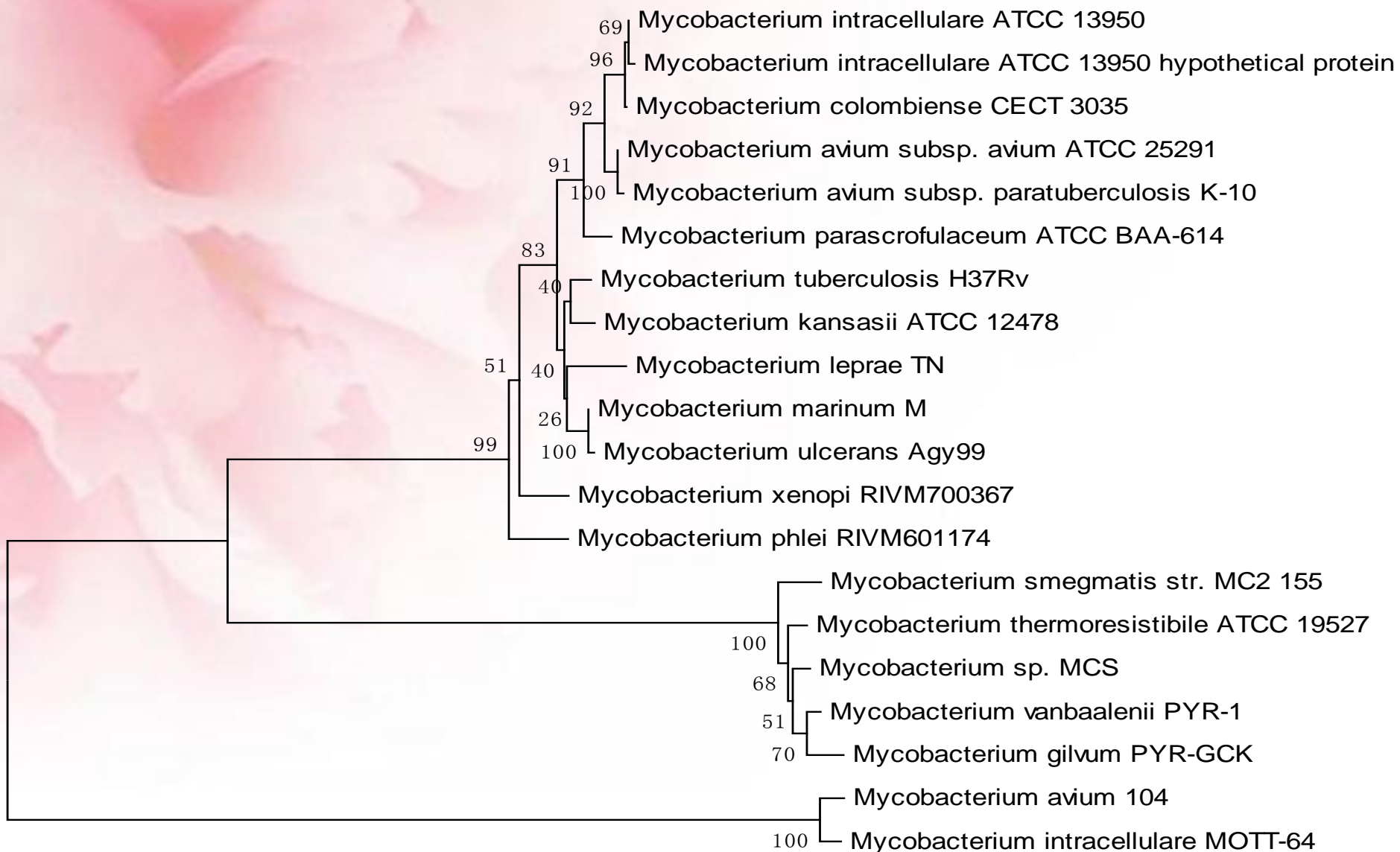
# 用dotmatcher对氨基酸重复序列进行分析

Dotmatcher: fasta::543170:Y1636·MYCTU vs fasta::543171:Y...  
(window size = 10, threshold = 23.00 06/06/12)





# 构建进化树

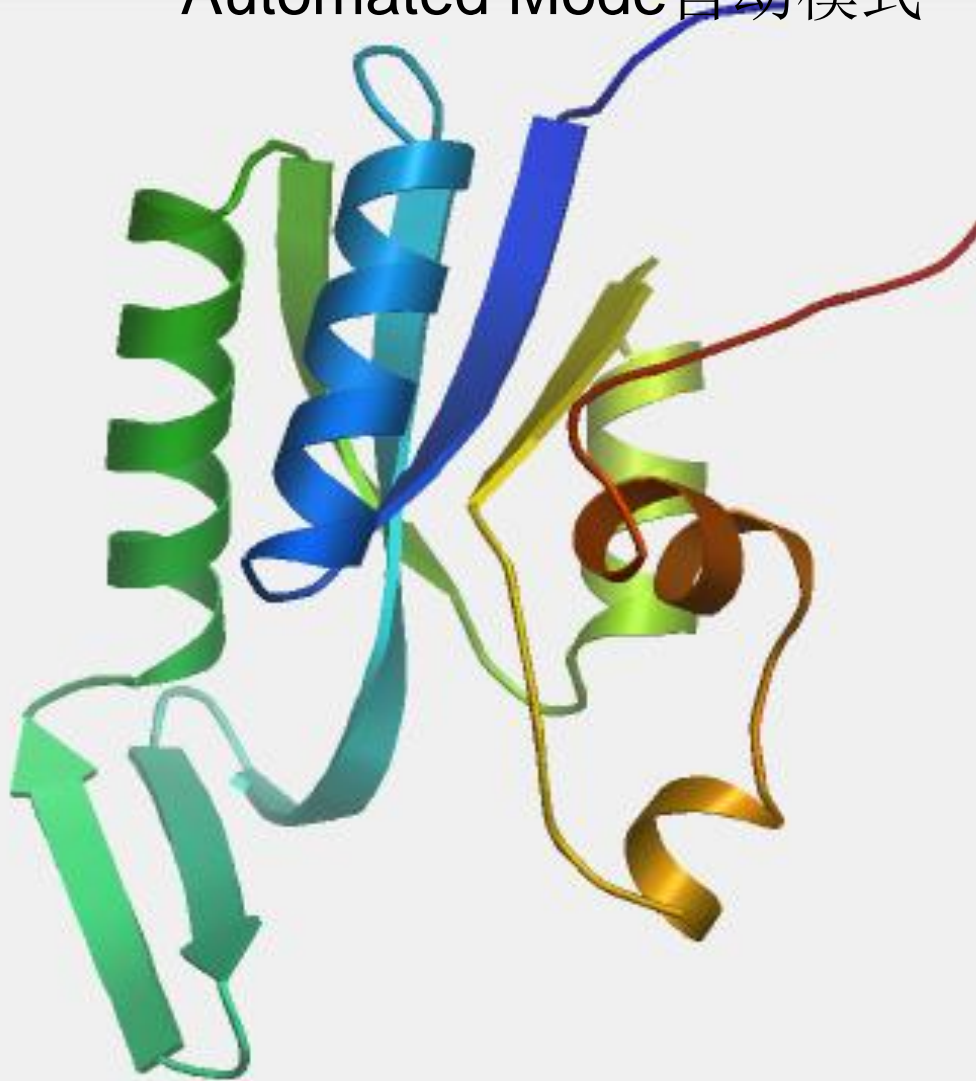


0.2

# 4.蛋白质结构预测

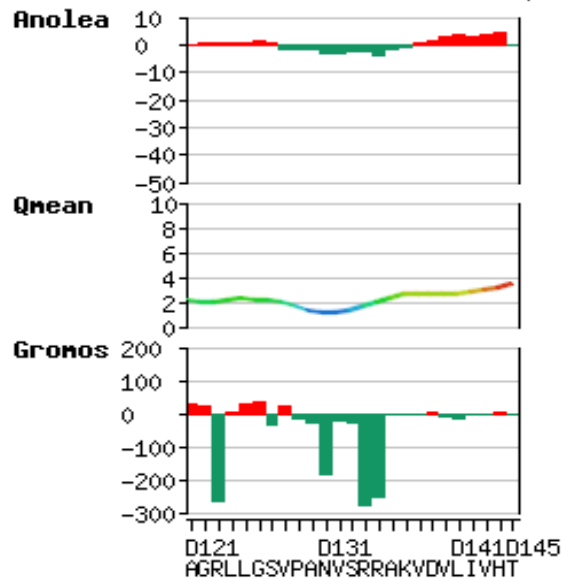
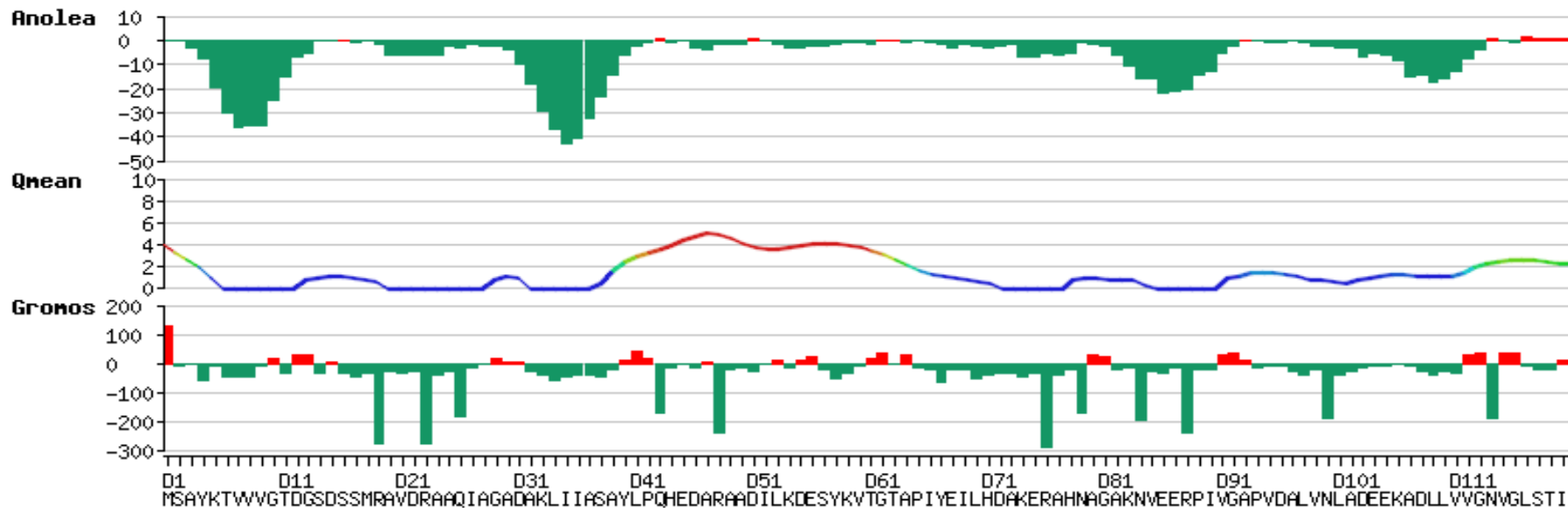
用Swiss-Model预测的结构

Automated Mode自动模式



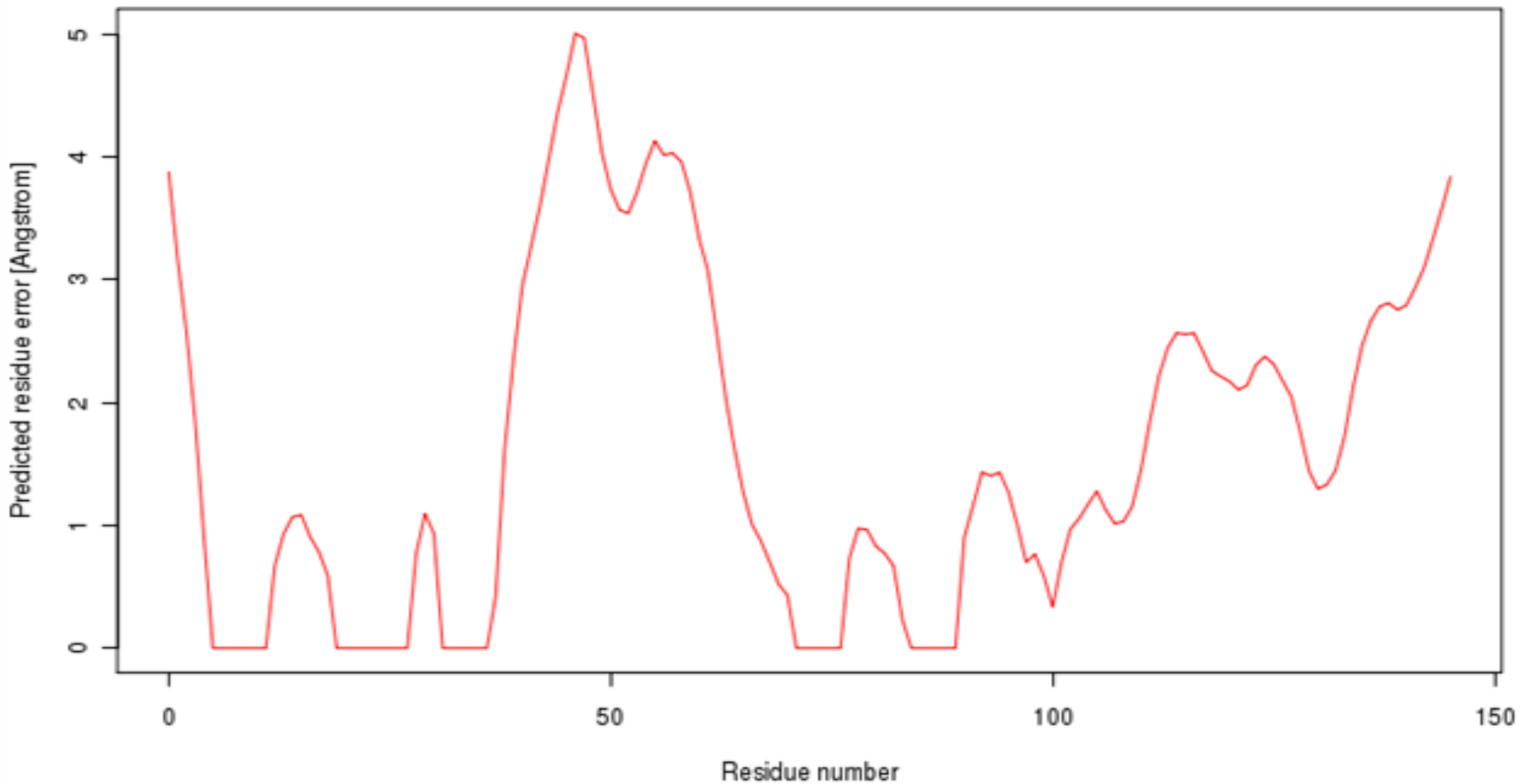
# 模型的评估: Anolea / QMEAN / Gromos:

anolea:  on  off    QMEAN:  on  off    gromos:  on  off

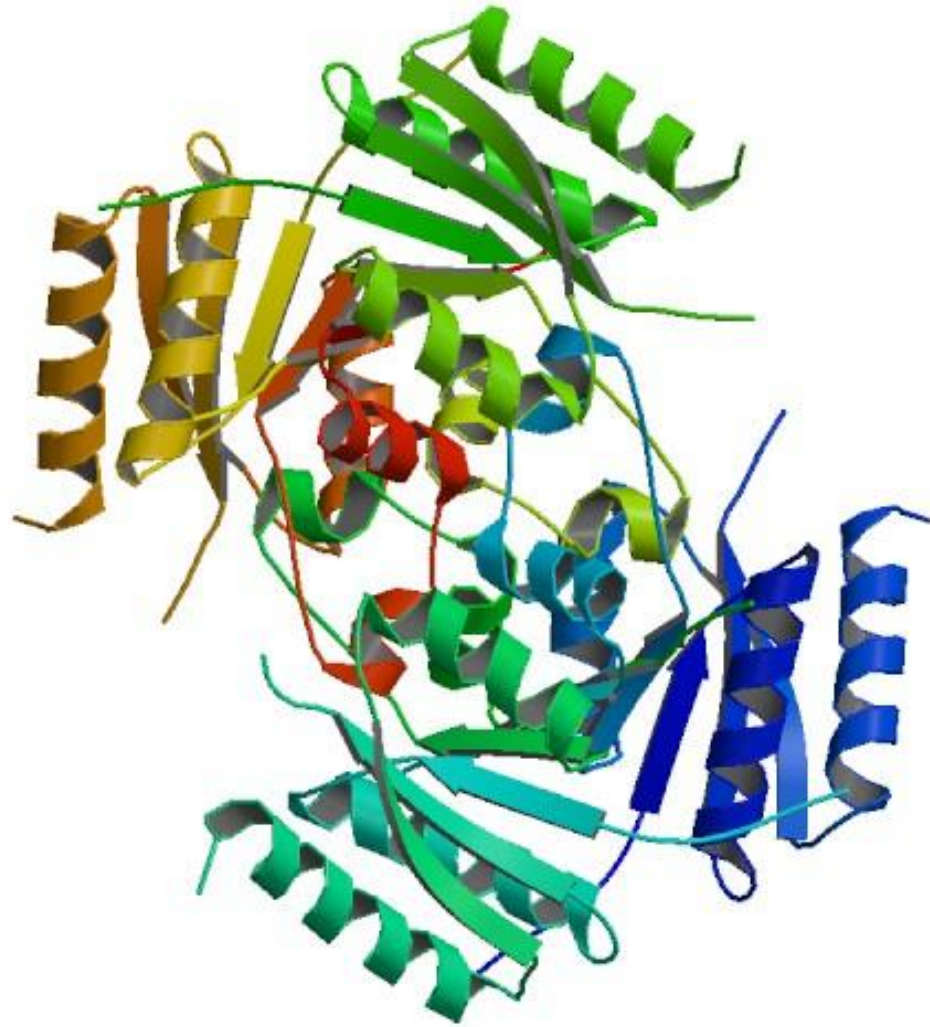


# Residue error plot

Predicted local error for structure Batch.1.short.pdb



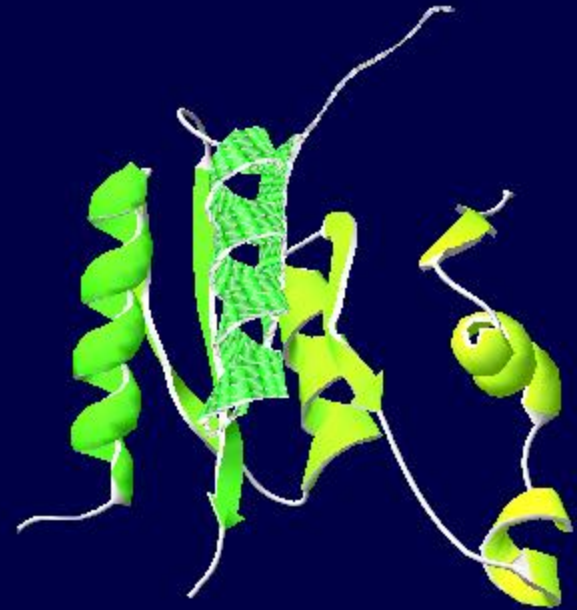
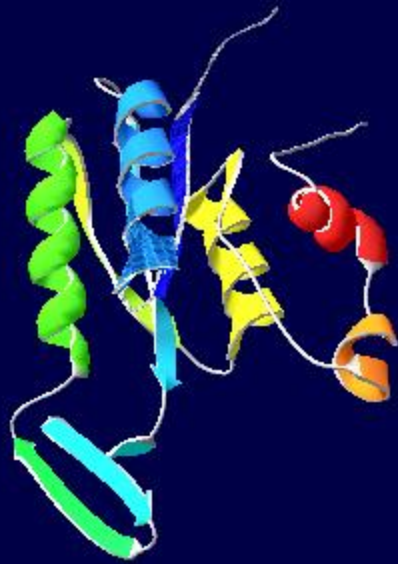
# 从PDB中下载的Rv1636的晶体结构



Chains: A, B,  
C, D,  
E, F

选取第4位的Tyr，第64位的Ala和第135位的Arg，与A亚基做Fit Molecular

- 比较重合之前的3D结构:

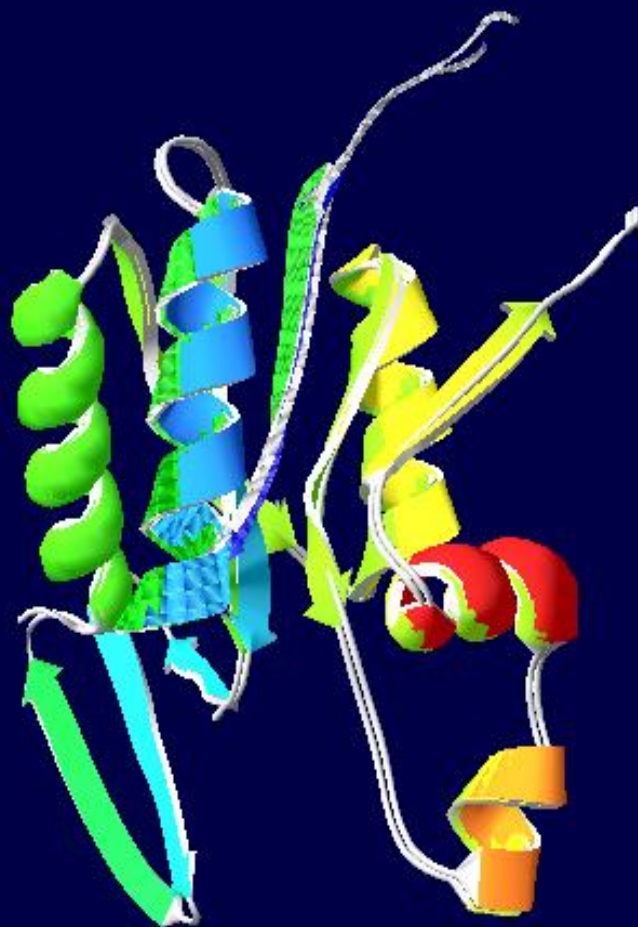


- 预测的

PDB中的



- 比较重合之后的3D结构:



- $\text{RMSD}=0.08$

## 4.参考文献

### 1.结核分枝杆菌Rv1168c 蛋白的基因表达、纯化及结构分析

余晓丽，孙战强，周晨俊，温子禄，陈军，孙庆文，王洪海，张舒林

### 2.结核分枝杆菌蛋白质组学的研究进展

吴雪琼，刘又宁

### 3.Crystal structure of hypothetical protein Rv1636 from Mycobacterium tuberculosis H37Rv

## 5.致谢

- 感谢罗老师！
- 感谢17组成员的合作！
- 感谢慈彦鹏和张利杰同学的帮助！

# 谢谢！