

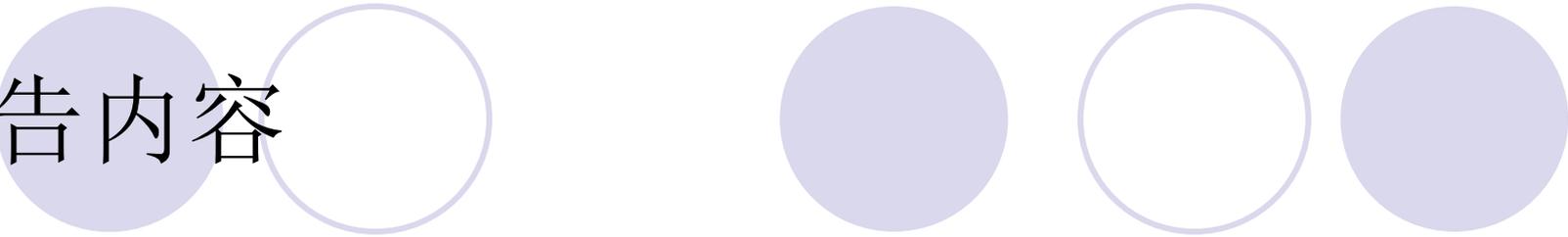
利什曼原虫去乙酰化酶SIR2的序列分析和 结构预测

Sequence analysis and structure prediction of SIR2 in *leishmania*

第一组：左云云
单兴娜
李建秋
陈新新

2013.6.21

报告内容



- 1、背景知识介绍及研究的目的意义
- 2、SIR2的进化分析，构建发育树
- 3、SIR2的核酸序列分析
- 4、SIR2的蛋白序列分析及结构预测

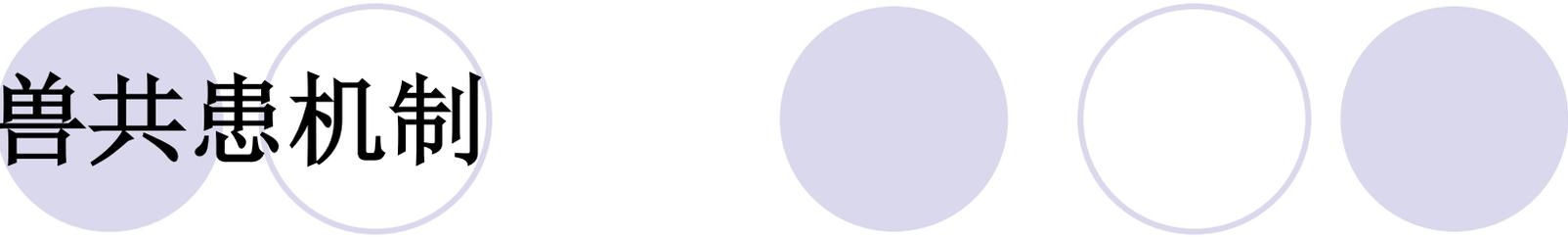
利什曼原虫病

- 利什曼病（**Leishmaniasis**）又名利什曼原虫病。是由利什曼原虫(*Leishmania*)寄生在人和犬、野生动物、爬行类动物中的一种常见的寄生虫病。该病能引起皮肤或内脏器官的严重损害甚至坏死。该病在大多数地区是人兽共患病。在不同国家与地区流行的利什曼病，其主要侵害部位不同.引起我国利什曼原虫病的病原是杜氏利什曼原虫。主要引起皮肤黑热病和内脏利什曼原虫病。

疾病病原

- 利什曼原虫与锥虫同属一个科，即锥虫科 (*Trypanosomatidae*)。利什曼原虫随着生活史的不同阶段而表现出不同的形态。按其生活史的共同特点有前鞭毛体 (*Promastigote*) 和无鞭毛体 (*amastigote*) 两个时期，前鞭毛体寄生在无脊椎动物的消化道内，其宿主为白蛉 (*Phlebotomus Spp.*)。无鞭毛体寄生在脊椎动物的网状内皮细胞内，其宿主为哺乳类和爬行类动物。

人兽共患机制



- 内脏利什曼病或黑热病病原贮藏在家犬体内。再由犬类传给人类，成为人兽共传的传染病。

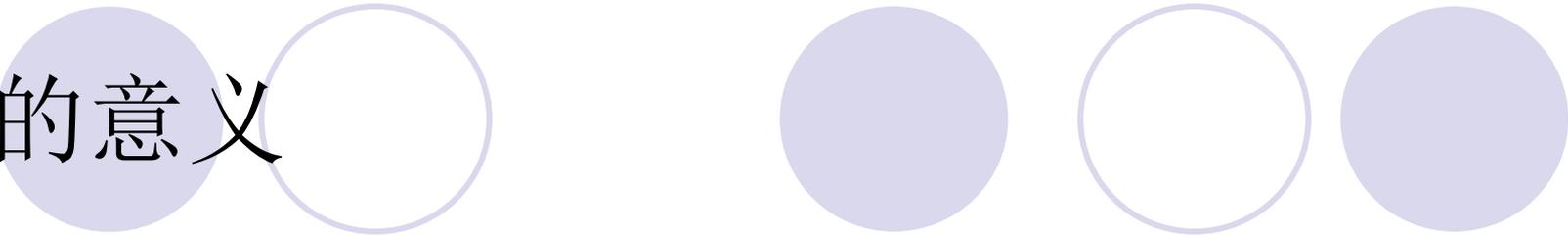
SIR2的背景知识介绍及研究的目的意义

- SIR2 (silence information regulator)基因家族是一种保守的从古细菌到哺乳动物都存在的NAD⁺依赖的组蛋白/非组蛋白去乙酰化酶。在酵母中，SIR2连同与它相互作用的几个蛋白质在基因沉默、基因组稳定性、细胞寿命以及代谢调节上起着不可缺少的作用。在哺乳动物中，有7个SIR2同源基因，分别命名为SIRT1 到SIRT7。
- SIR2催化的去乙酰反应方程式： NAD^+ 乙酰化赖氨酸 \rightarrow 烟酰胺 + 乙酰-ADP-核糖 + 去乙酰化的赖氨酸

参考生物-酿酒酵母

- 酵母**SIR2** 由两个结构域组成，大的结构域由**6** 个**a** 螺旋和**6** 个平行**b** 折叠组成，小的结构域由**3** 个反平行**b** 折叠，**2** 个**a** 螺旋和**1** 个锌原子组成；另外还有一系列环状结构与这两个结构域相连，并在其核心区域的中间形成一个裂缝。
- 乙酰化的赖氨酸残基和辅助因子**NAD+** 结合在裂缝的对面。
- **SIR2** 核心区域高度的序列保守性，暗示其保守的催化机制。

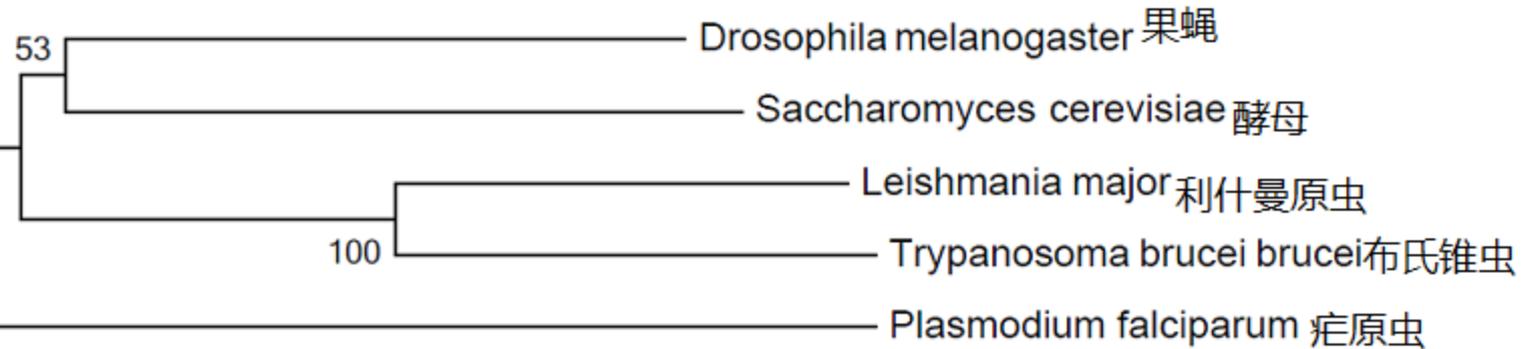
目的意义



寄生虫中这类蛋白研究较少。

利用生物信息软件分析利什曼原虫**SIR2**蛋白的相关序列和预测其蛋白结构，并分析**SIR2**蛋白与底物结合位点与特点，根据作用靶点找出其抑制剂，应用于抗虫药物研究。

2、SIR2的进化分析，构建发育树

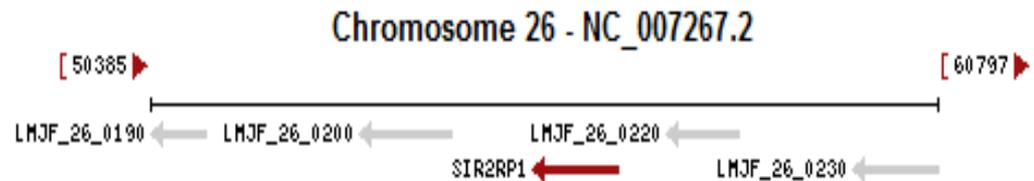


3、SIR2的核酸序列分析

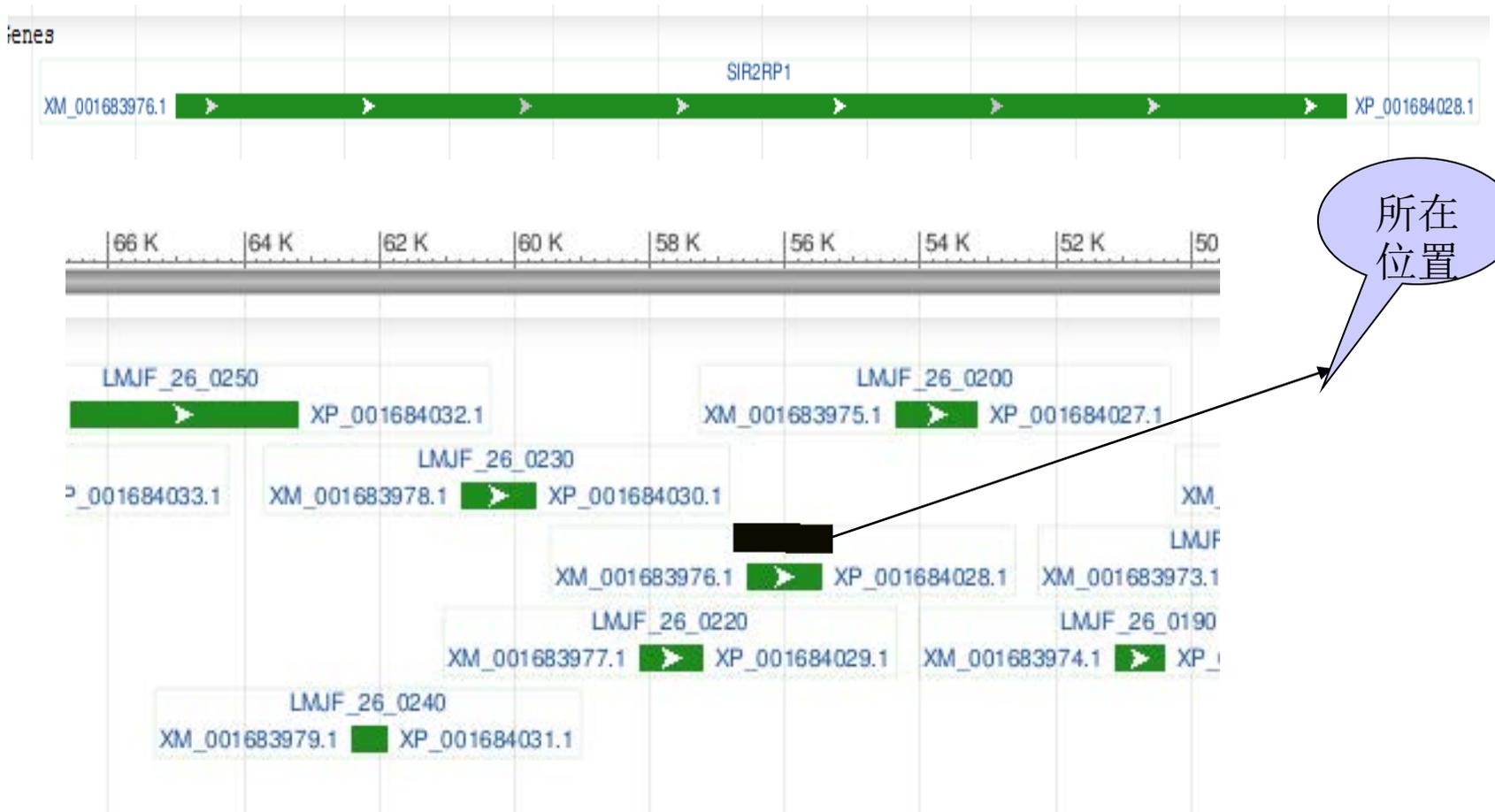
- 位于第26号染色体，从55441-56562，共1121个碱基

Location: chromosome: 26

Sequence: Chromosome: 26; NC_007267.2 (55441..56562, complement)



● 没有内含子



密码子使用分析cusp:

#CdsCount: 1

#Coding GC 62.83%

#1st letter GC 63.37%

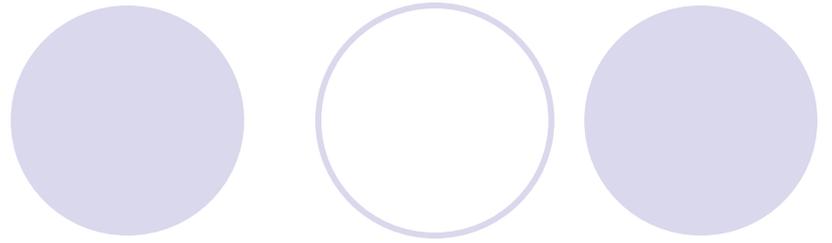
#2nd letter GC 47.33%

#3rd letter GC 77.81%

#Codon	AA	Fraction	Frequency	Number
GCA	A	0.135	13.369	5
GCC	A	0.351	34.759	13
GCG	A	0.432	42.781	16
GCT	A	0.081	8.021	3
TGC	C	1.000	26.738	10
TGT	C	0.000	0.000	0
GAC	D	0.696	42.781	16
GAT	D	0.304	18.717	7
GAA	E	0.179	13.369	5
GAG	E	0.821	61.497	23
TTC	F	0.706	32.086	12
TTT	F	0.294	13.369	5
GGA	G	0.100	8.021	3
GGC	G	0.467	37.433	14
GGG	G	0.267	21.390	8
GGT	G	0.167	13.369	5
CAC	H	0.667	21.390	8

- GC含量很高
- 双链稳定，不容易分开成为单链

4、蛋白质分析



- 成分
- 疏水性
- 信号肽预测
- 亚细胞定位
- 二级结构预测
- **3D**结构预测

成分

```
##
# The Expected frequencies are calculated on the (false) assumption that every
# word has equal frequency.
#
# The input sequences are:
#   SIR2_LEIMA
```

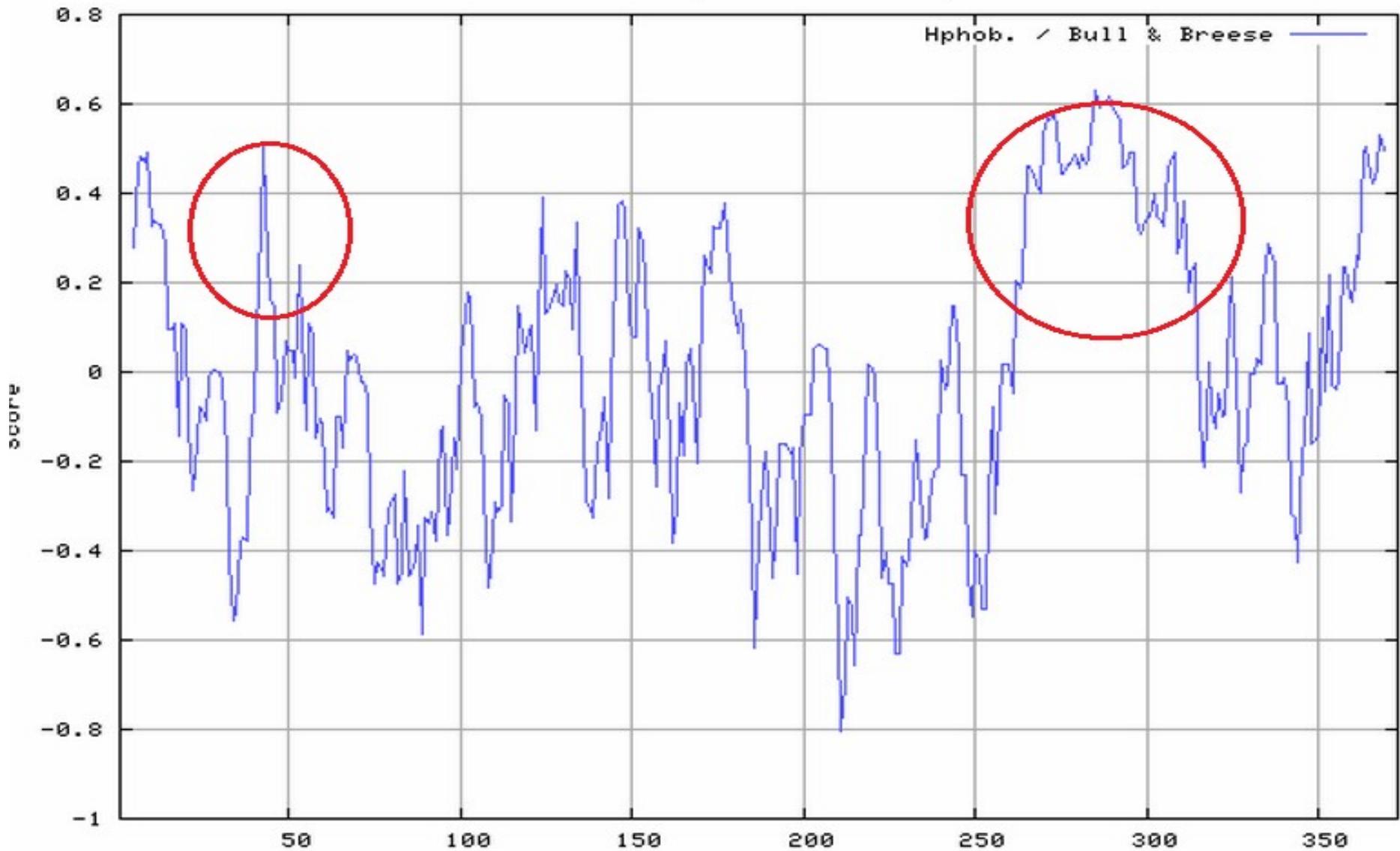
```
Word size      1
Total count    373
```

#	# Word	Obs Count	Obs Frequency	Exp Frequency	Obs/Exp Frequency
#	A	37	0.0991957	0.0476190	2.0831099
	C	10	0.0268097	0.0476190	0.5630027
	D	23	0.0616622	0.0476190	1.2949062
	E	28	0.0750670	0.0476190	1.5764075
	F	17	0.0455764	0.0476190	0.9571046
	G	30	0.0804290	0.0476190	1.6890080
	H	12	0.0321716	0.0476190	0.6756032
	I	17	0.0455764	0.0476190	0.9571046
	K	11	0.0294906	0.0476190	0.6193029
	L	33	0.0884718	0.0476190	1.8579088
	M	5	0.0134048	0.0476190	0.2815013
	N	10	0.0268097	0.0476190	0.5630027
	P	23	0.0616622	0.0476190	1.2949062
	Q	9	0.0241287	0.0476190	0.5067024
	R	26	0.0697051	0.0476190	1.4638070
	S	32	0.0857909	0.0476190	1.8016086
	T	17	0.0455764	0.0476190	0.9571046
	U	0	0.0000000	0.0476190	0.0000000
	V	23	0.0616622	0.0476190	1.2949062
	W	1	0.0026810	0.0476190	0.0563003
	Y	9	0.0241287	0.0476190	0.5067024
	Other	0	0.0000000	0.0000000	10000000000.0000000

- 从分析结果可以看出，丙氨酸、甘氨酸、亮氨酸这三个非极性疏水性氨基酸含量很高，说明该蛋白可能存在疏水区。

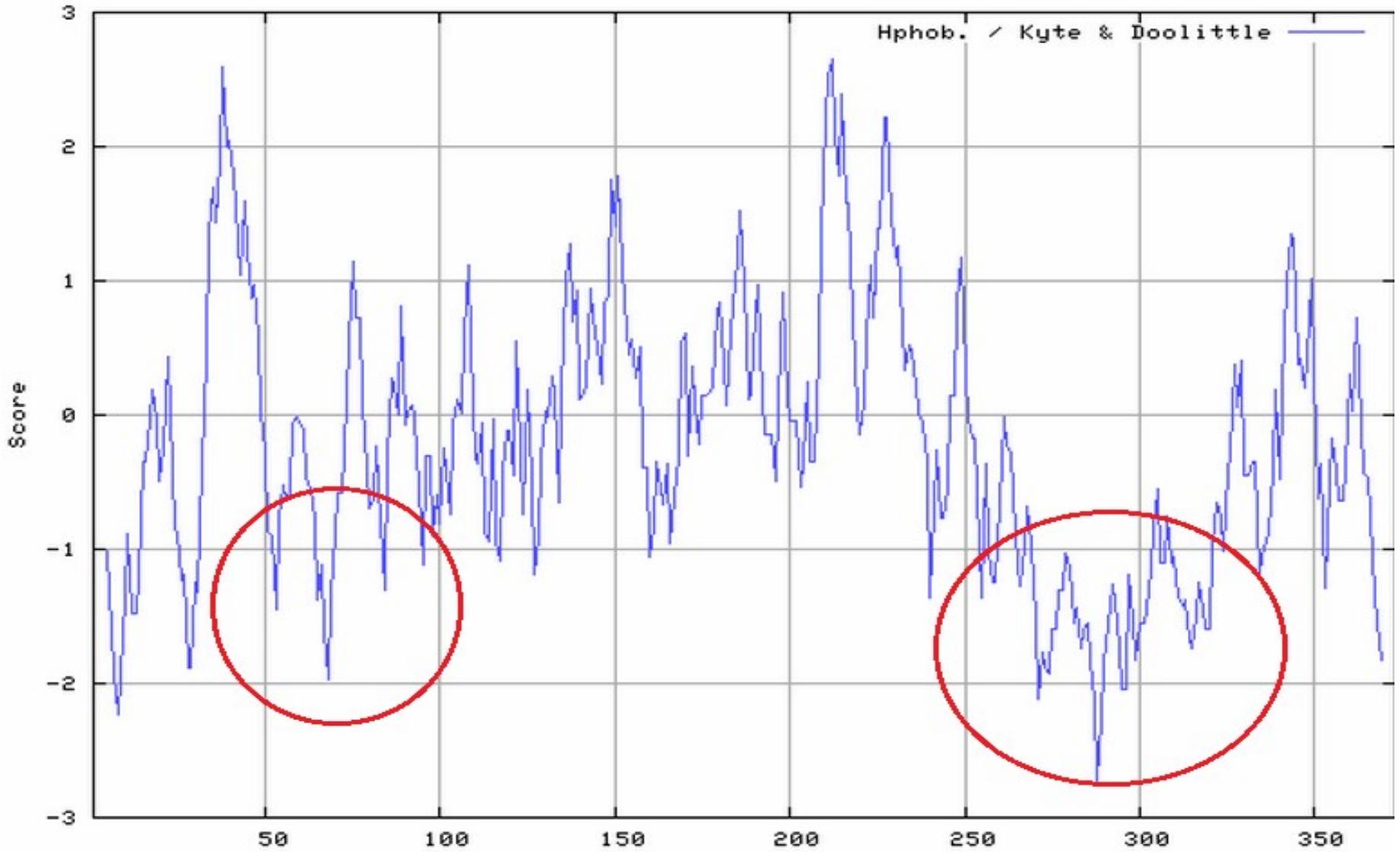
疏水性

Protoscale output for user sequence



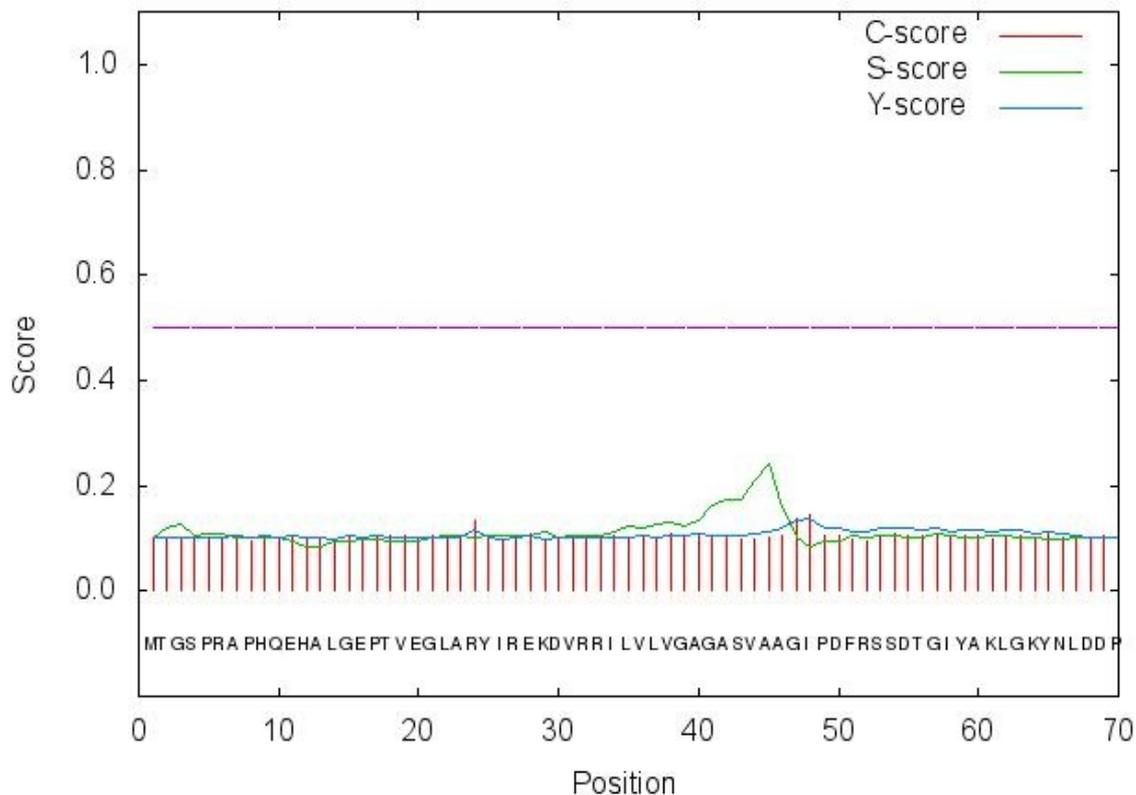
亲水性

ProtScale output for user sequence



信号肽预测

SignalP-4.1 prediction (euk networks): sp_Q25337_SIR2_LEIMA



- 从预测结果可以看出该蛋白很大程度上不含有信号肽。

# Measure	Position	Value	Cutoff	signal peptide?
max. C	48	0.143		
max. Y	48	0.138		
max. S	45	0.240		
mean S	1-47	0.116		
D	1-47	0.126	0.450	NO

Name=sp_Q25337_SIR2_LEIMA SP='NO' D=0.126 D-cutoff=0.450 Networks=SignalP-noTM

亚细胞定位

TargetP 1.1 Server - prediction results

Technical University of Denmark

CENTER FOR
BIOLOGICAL
SEQUENCE
ANALYSIS
CBS

```
### 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
sp_Q25337_SIR2_LEIMA	373	0.250	0.046	0.775	_	3
cutoff		0.000	0.000	0.000		

从结果分析得出：由于该蛋白定位于线粒体的可能性较小，是分泌蛋白的可能性也很小。所以最大可能是定位于细胞其他部位。

Results of the k -NN Prediction

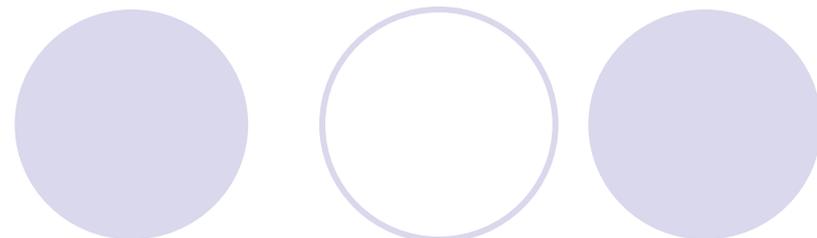
$k = 9/23$

65.2 %: cytoplasmic
13.0 %: nuclear
13.0 %: mitochondrial
4.3 %: Golgi
4.3 %: cytoskeletal

No.	seg. Length	subcellular Localization site	ID
0001	373a. a.	C (cytoplasmic)	None

- 从PSORT II、SOSUIGramN结果可以看出：该蛋白很大程度上定位于细胞质。

二级结构预测

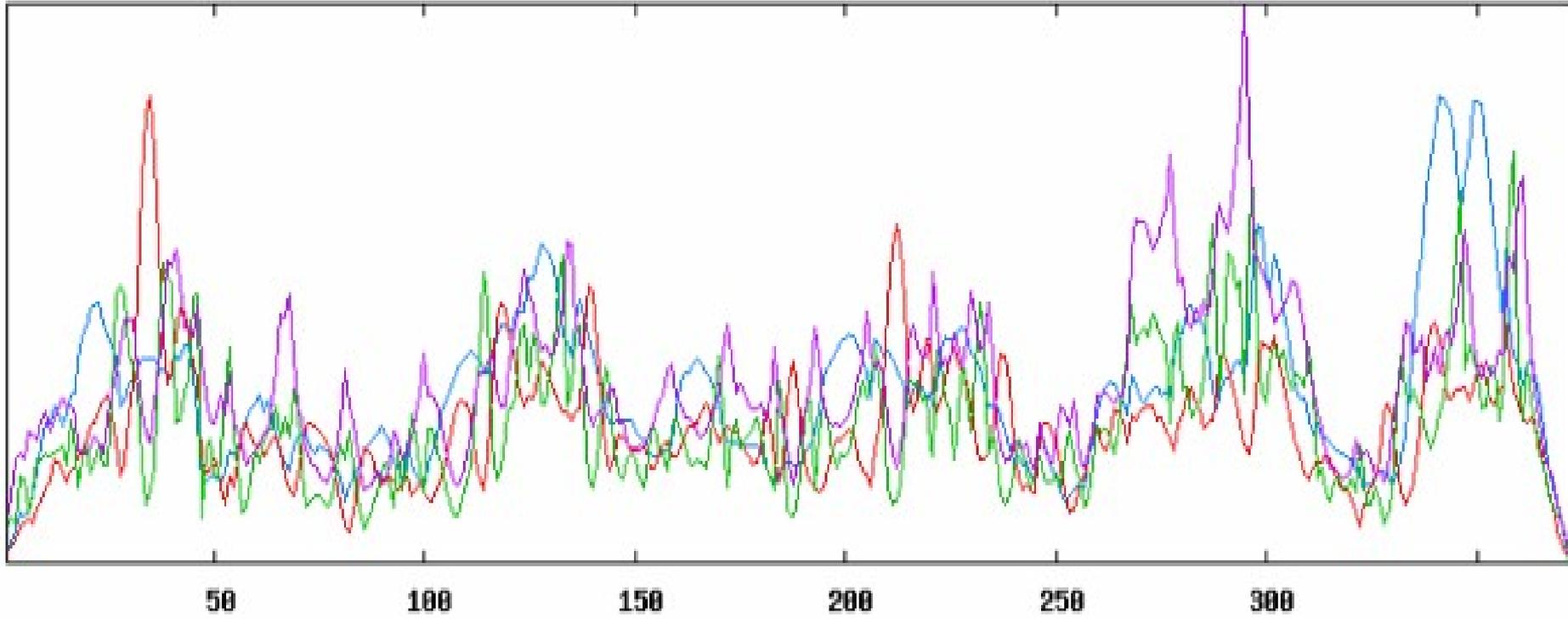
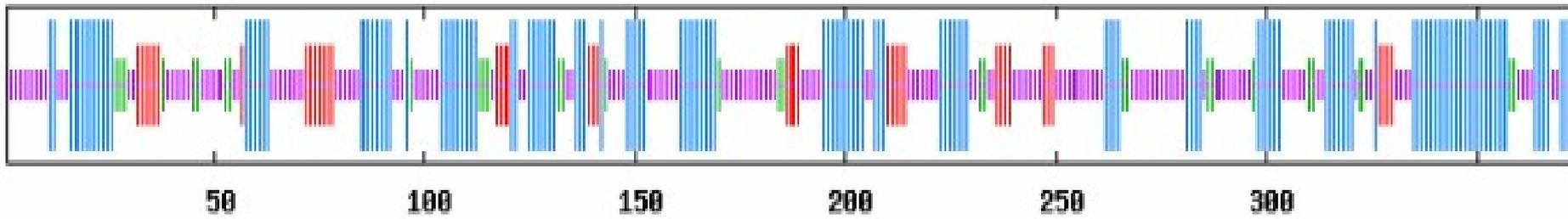


```
      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
MTG SPRAPHQEHALGEP TVEGLARYIREKDVRRLVLVVGAGASVAAGIPDFRSDTG IYAKLGKYNLDDP
ccccccccccccchccccchhhhhhhhhhhhtttceeeeeeetcccccoettcccccoettceehhhhhhhcccccc
TDAFSLTLLREKPEIFYSIARELNLWPGHFQPTAVHHFIRLLQDEGRLLRCCTQNIDGLEKAAGVSPELL
ceeeeeeeccccchhhhhhhhhccchtccccchhhhhhhhhhtttceeeehhccccchhhhhhhhttcchhhhe
VEAHG SFAAAACIECHTPF SIEQNYLEAMSG TVSRCS TCGG IVKPNVVFVFFGENLPDAFFDALHHDAP IAE
eehtccccchhhhhccccccccchhhhhhhhhhtccccccccccccccettceeeccccchhhhhhhhhhhhhcchhh
LVI IIGT SMQVHPFALLPCVVPK SIPRVLMMNRERVGGLLFRFPDDPLDT IHDDAVAKEGRSSSSSQSRSPS
eeeesccccccccchhhhhhhhccttceeeccccccccccccccccccccccccccccccccchhhhtcccccccccccc
ASARREEGG TEDGSSSPNEEVEDASTSSSSDGYGQYGDYYAHPDVCRDVFFFRGDCQENVLKLAECLGLRE
chhhhtttcccccccccccthhhhhhcccccccttcchhhhhhhcttccheeeccccchhhhhhhhhhhhhhhhhhh
ALAKRMRFSGAAPATARKTSNET
hhhhhhhhhttccechhhhechhh
```

Sequence length : 373

SOPMA :

Alpha helix	(Hh)	:	136	is	36.46%
3 ₁₀ helix	(Gg)	:	0	is	0.00%
Pi helix	(Ii)	:	0	is	0.00%
Beta bridge	(Bb)	:	0	is	0.00%
Extended strand	(Ee)	:	41	is	10.99%
Beta turn	(Tt)	:	31	is	8.31%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	165	is	44.24%
Ambiguous states (?)		:	0	is	0.00%
Other states		:	0	is	0.00%

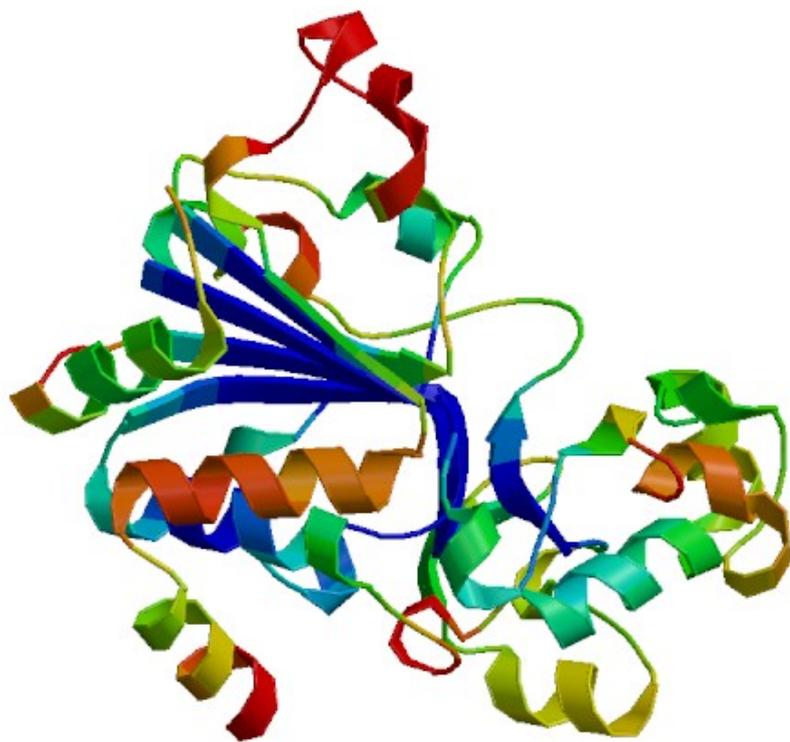


Parameters :

Window width : 17
Similarity threshold : 8
Number of states : 4

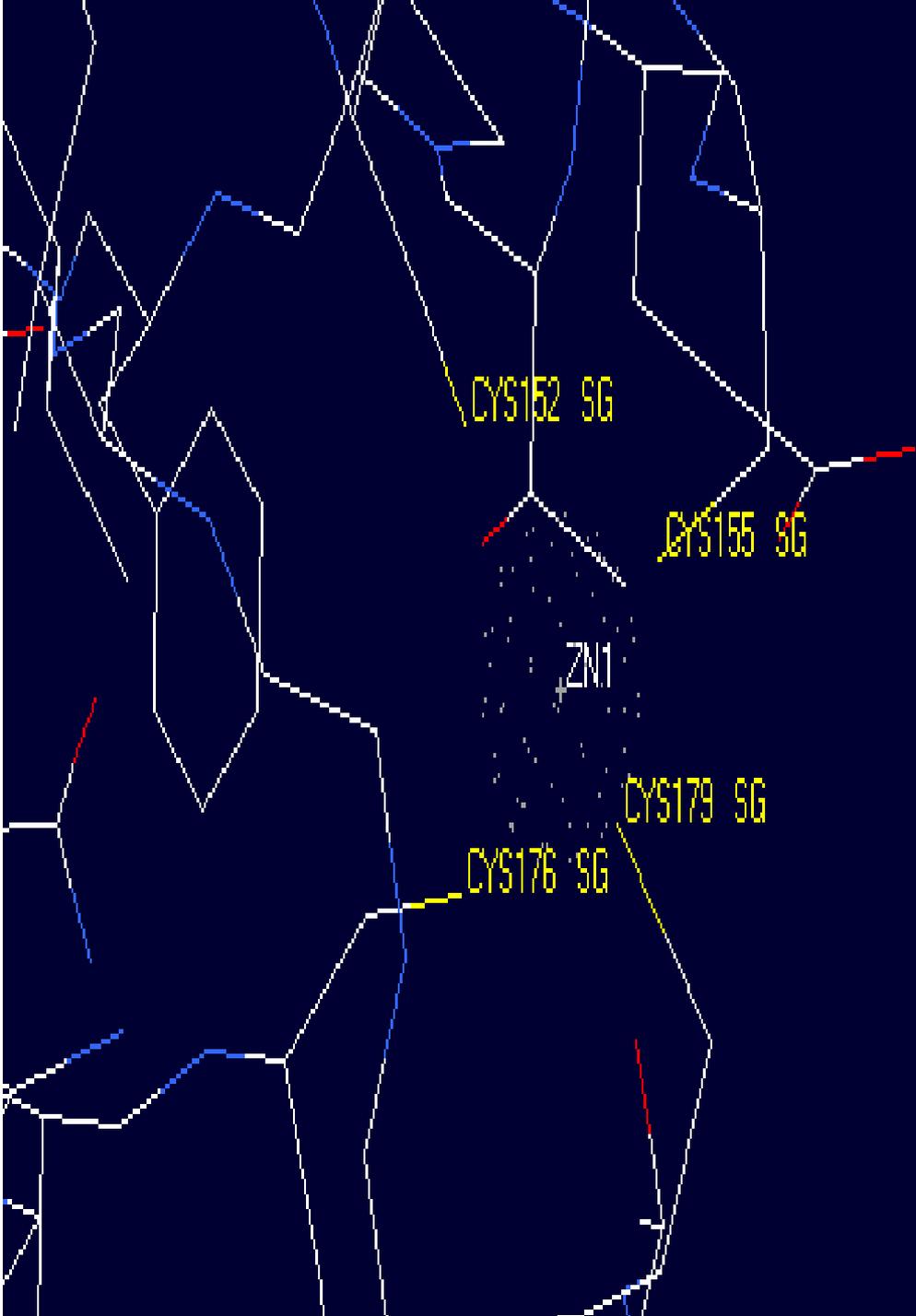
3D结构预测

- 利用**swiss-model**预测**SIR2**（15-295位）



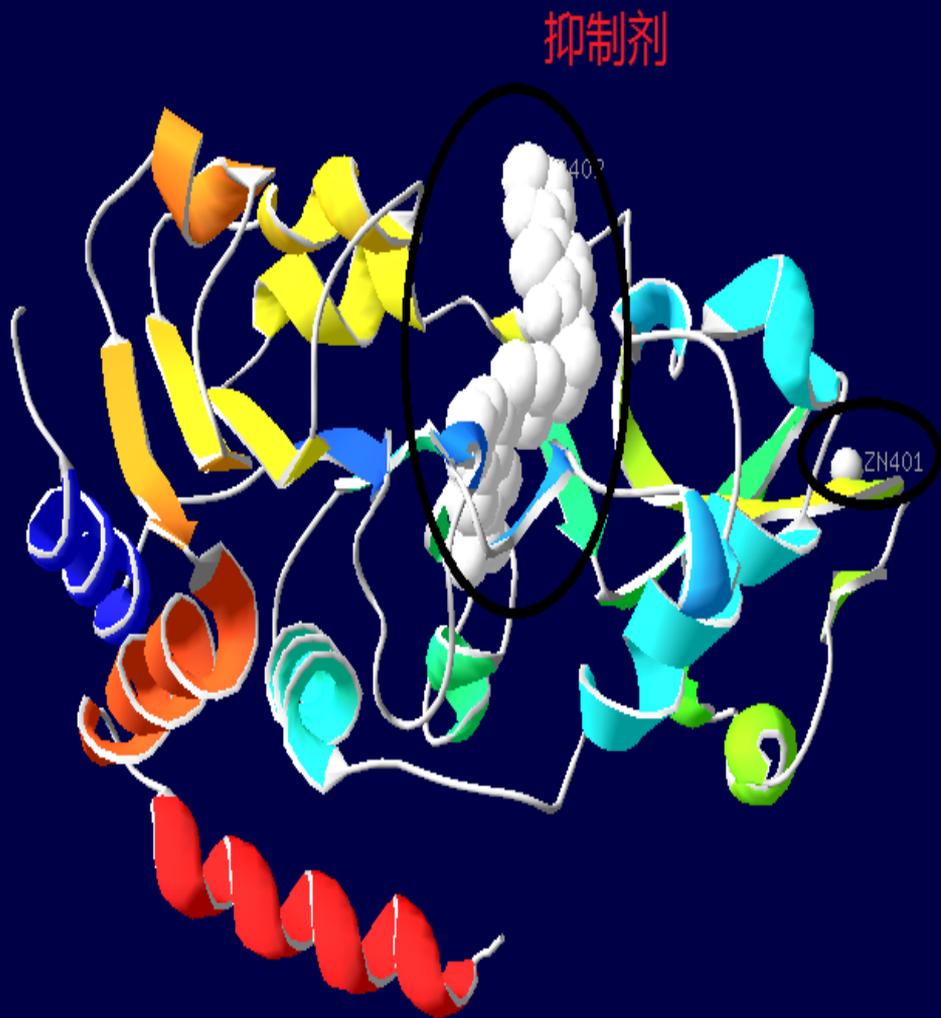
- **4jsrA**（121-402位）
作为模板





- 辅基Zn周围的氨基酸残基
- 152、155、176、179均为半胱氨酸。

与抑制剂作用

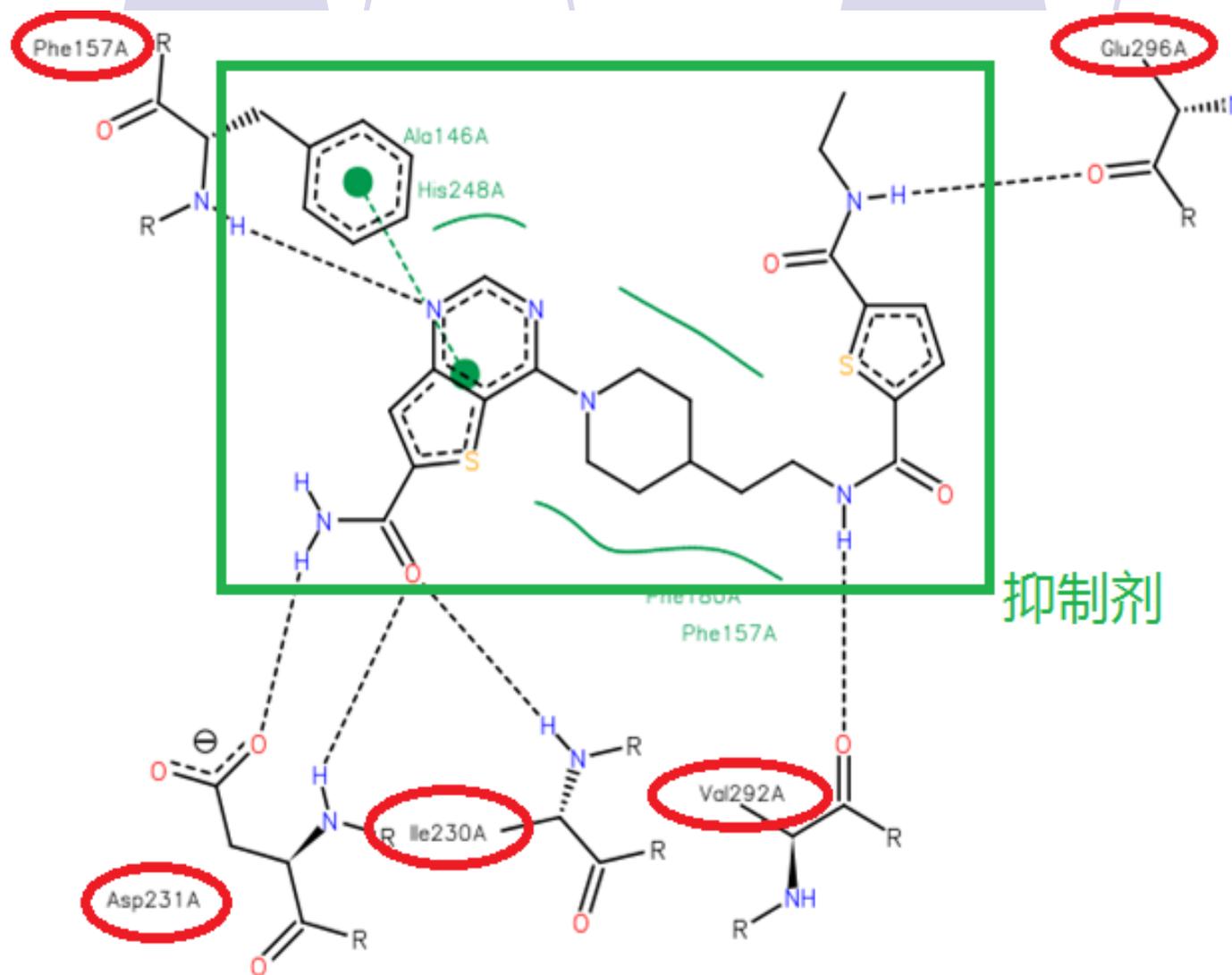


4jsr (模板) 与抑制剂



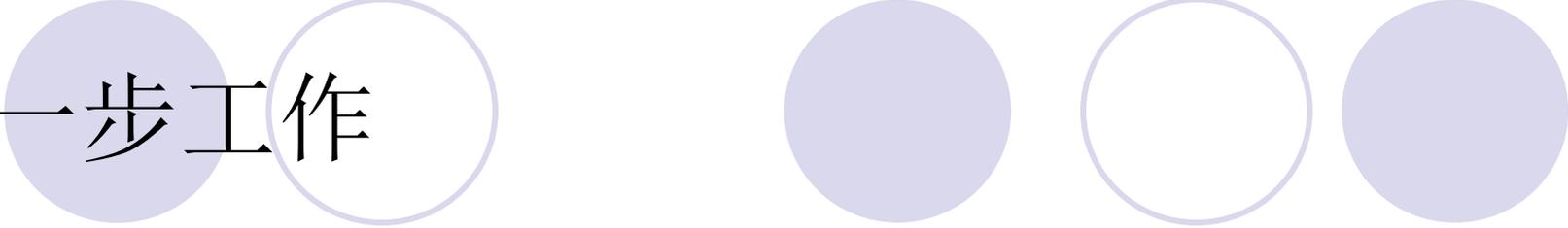
SIR2与NAD或抑制剂可能结合位点

抑制剂与周围的氨基酸残基（4jsr模板）

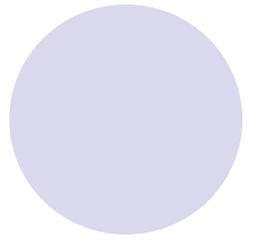
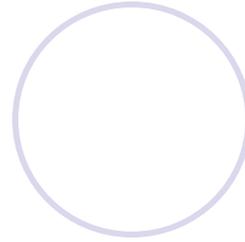
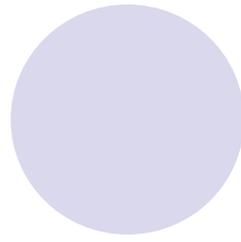
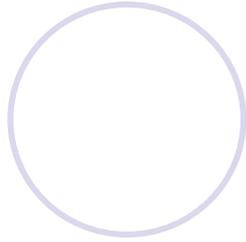
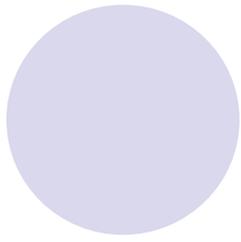


1NQ: N-{2-[1-(6-carbamoylthieno[3,2-d]pyrimidin-4-yl)piperidin-4-yl]ethyl}-N'-ethylthiophene-2,5-dicarboxamide

下一步工作



- **SIR2**是 NAD^+ 依赖性的蛋白酶，有关 NAD^+ 如何调节其活性有待研究
- **SIR2**如何与底物结合发挥作用，也需要对其蛋白质三维结构进一步分析。



Thanks for your attention!