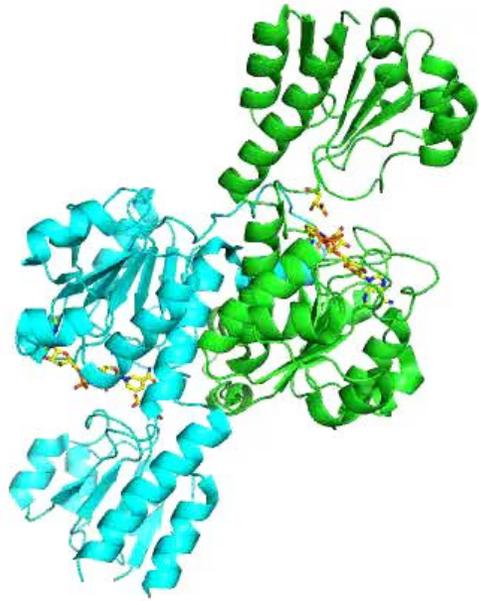


# The Application of Bioinformatics to Research on D-3-Phosphoglycerate Dehydrogenase

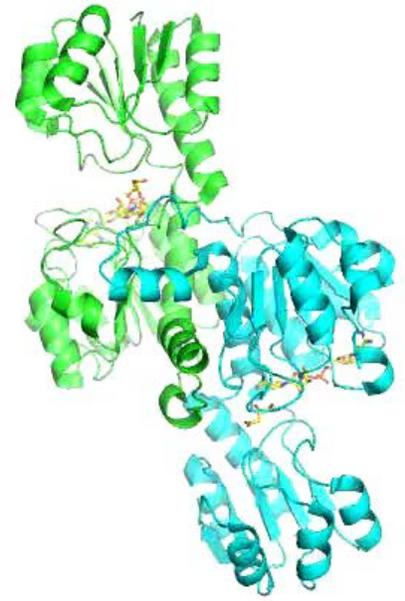
## 生物信息学在磷酸甘油酸脱氢酶研究中的应用



小组：G11&G12

报告人：刘培

组员：景誉庆、吴洋、梁振、米丽拜尔、  
王玉生、刘培



2014-6-19

# 概要



背景介绍—磷酸甘油酸脱氢酶



课题意义



生物信息学的应用

- ✓ 第一部分 一级机构分析
- ✓ 第二部分 二级结构预测
- ✓ 第三部分 三级结构预测



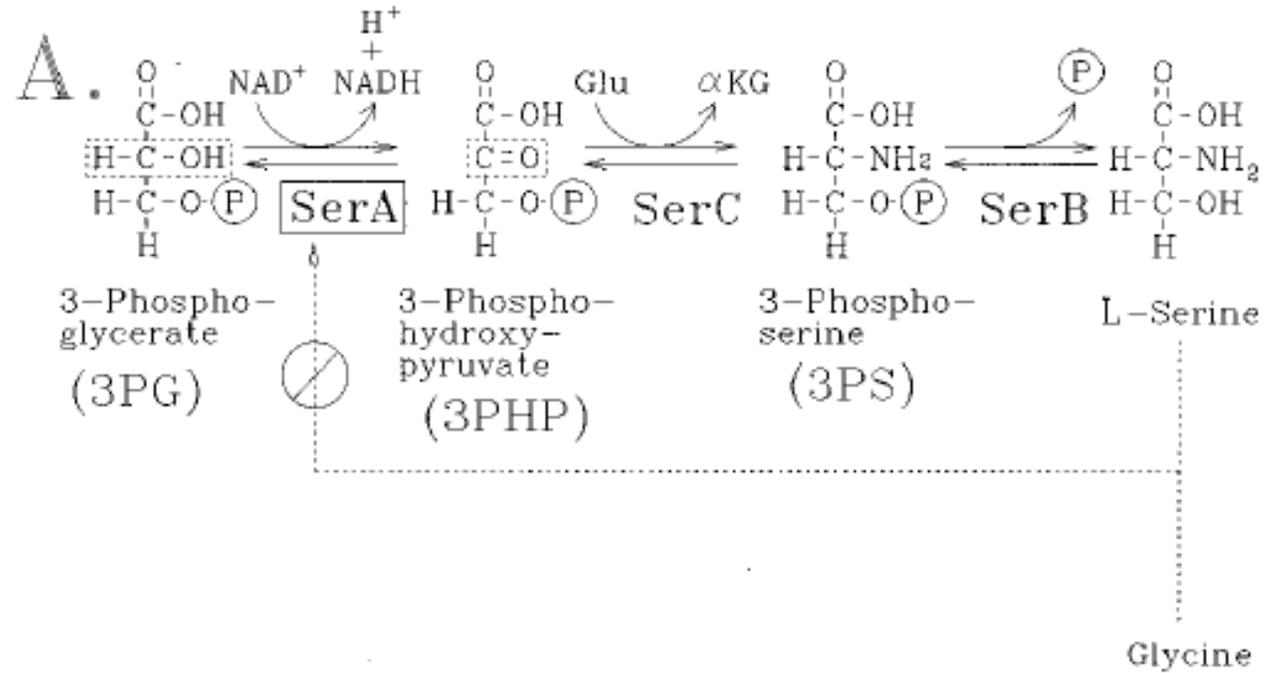
结果分析及讨论



致谢

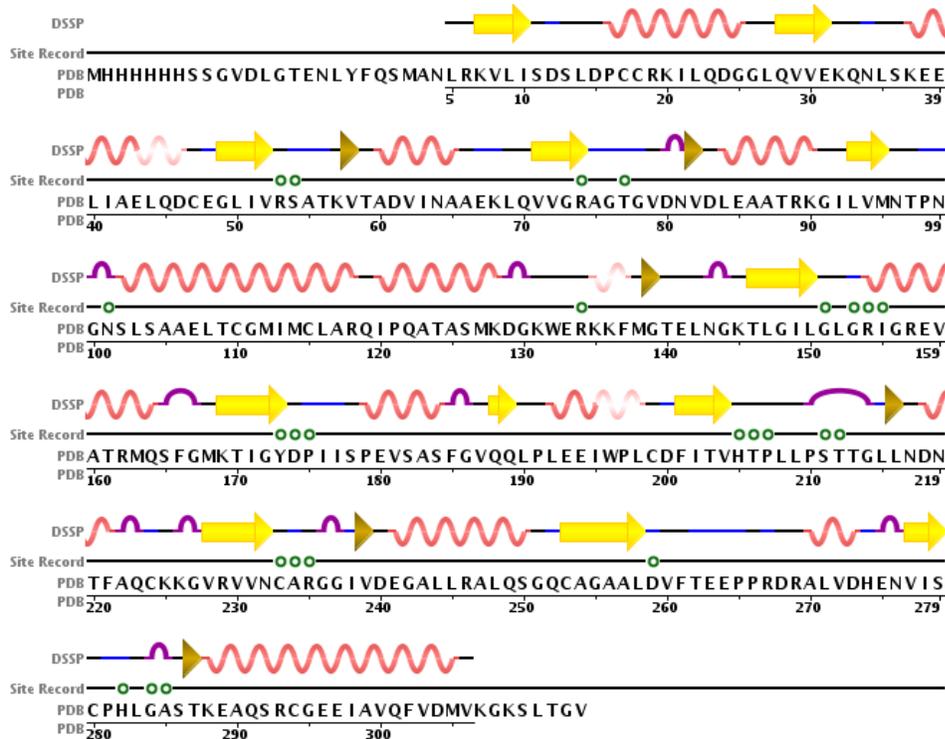
# 一、背景

## 磷酸甘油酸脱氢酶



# 一、背景

# 磷酸甘油酸脱氢酶—PDB结构

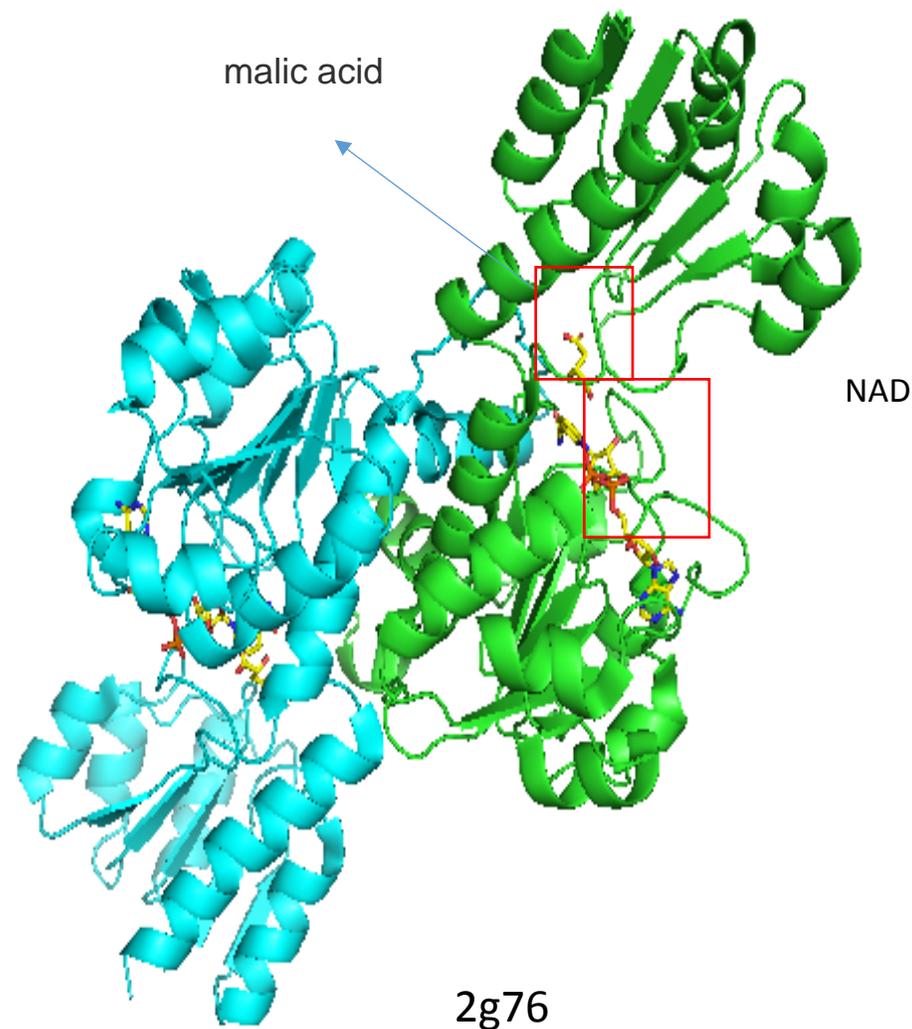


### Site Record Legend

- BINDING SITE FOR RESIDUE MLT A 503 (SOFTWARE)
- BINDING SITE FOR RESIDUE NAD A 501 (SOFTWARE)

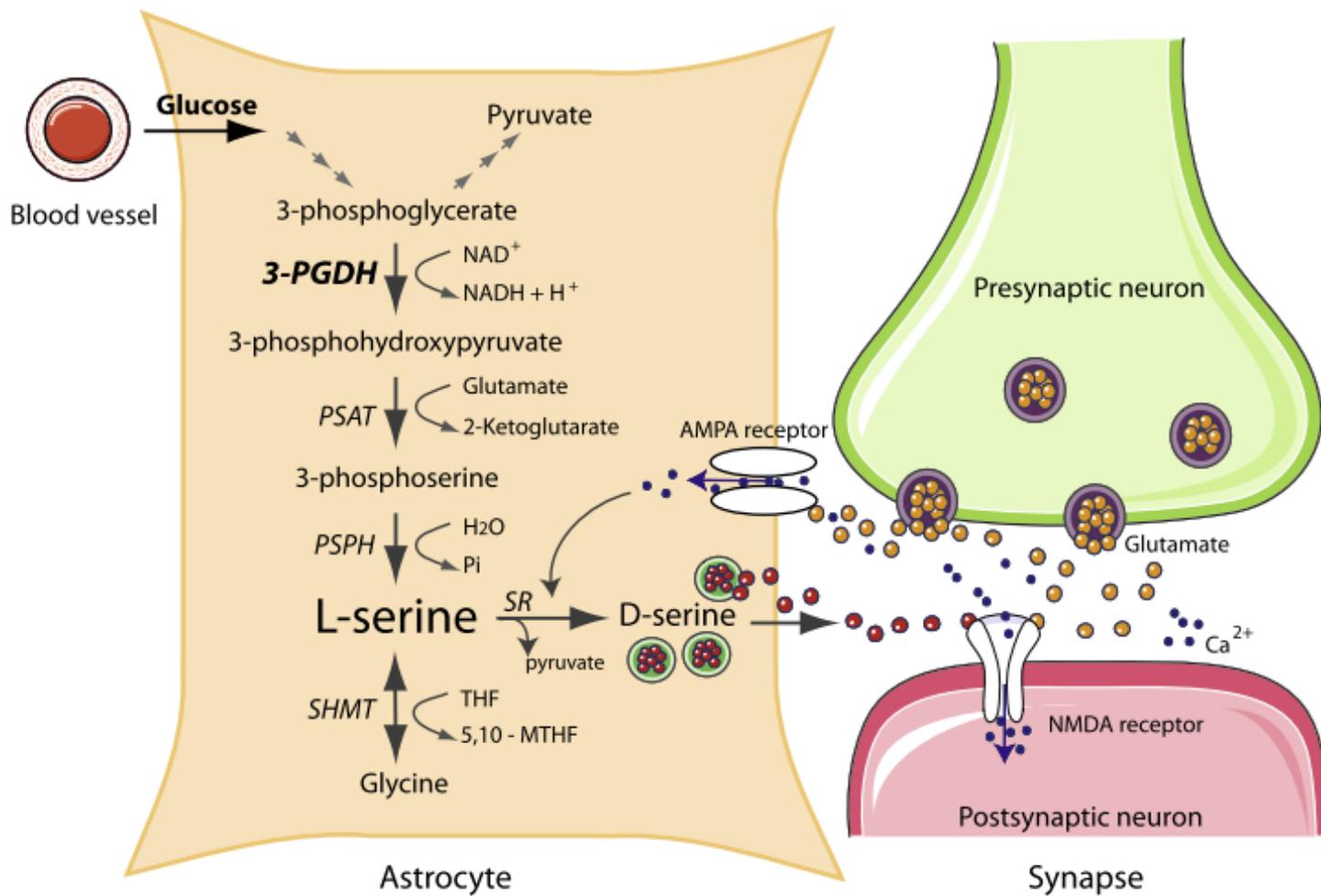
### DSSP Legend

- T: turn
- E: beta strand
- empty: no secondary structure assigned
- G: 3/10-helix
- B: beta bridge
- S: bend
- H: alpha helix



# 一、背景

## 磷酸甘油酸脱氢酶缺乏症



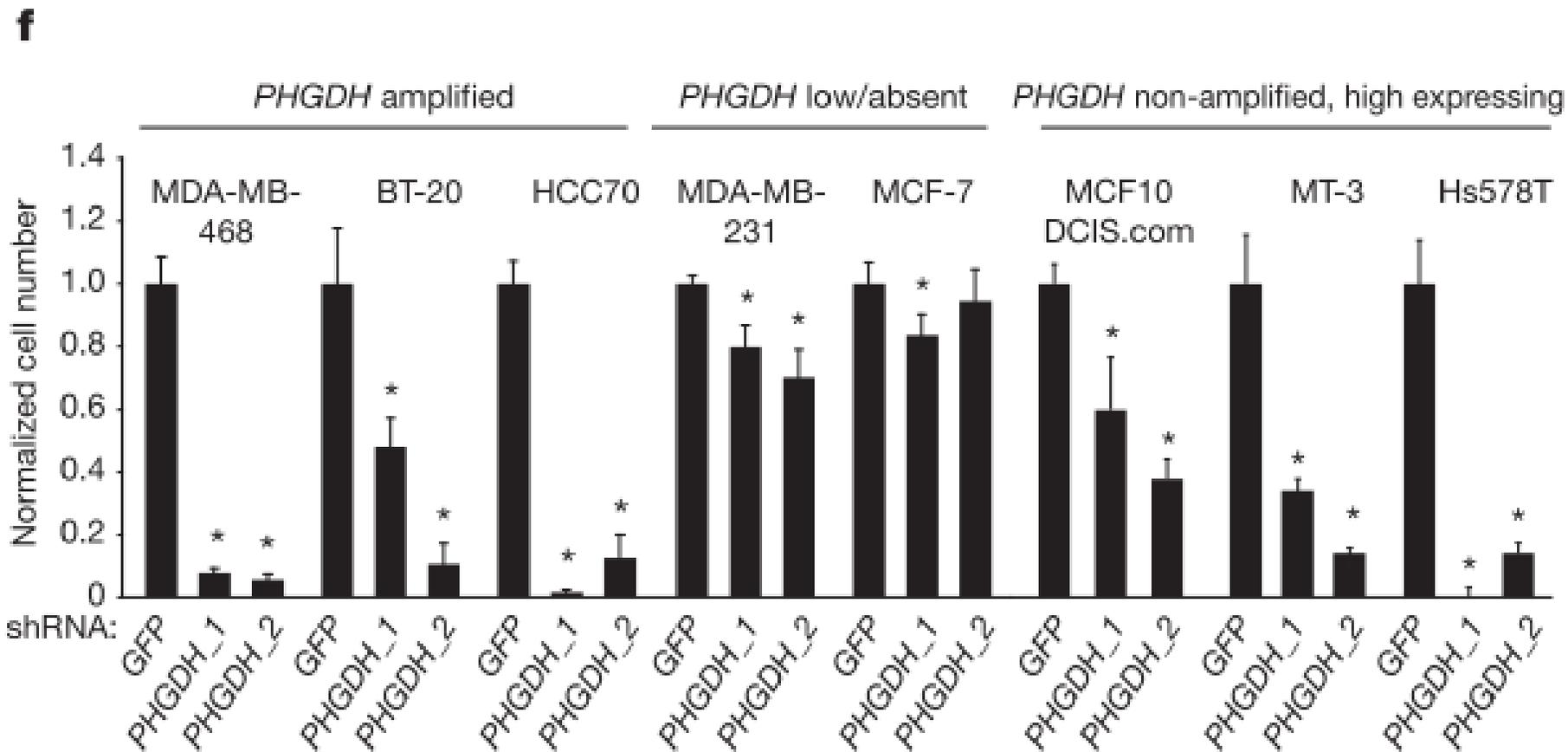
3-磷酸甘油酸脱氢酶(3-PGDH)缺乏症是一种L-丝氨酸生物合成紊乱所致的疾病,以先天性的小头畸形、严重的精神运动性迟滞及顽固的癫痫发作为特征。

尽管丝氨酸是一种非必须氨基酸,但是由于人大脑中的血脑屏障,外源补充的丝氨酸无法通过,人自身合成丝氨酸的途径必不可少。

丝氨酸在神经中枢系统中的代谢

# 一、背景

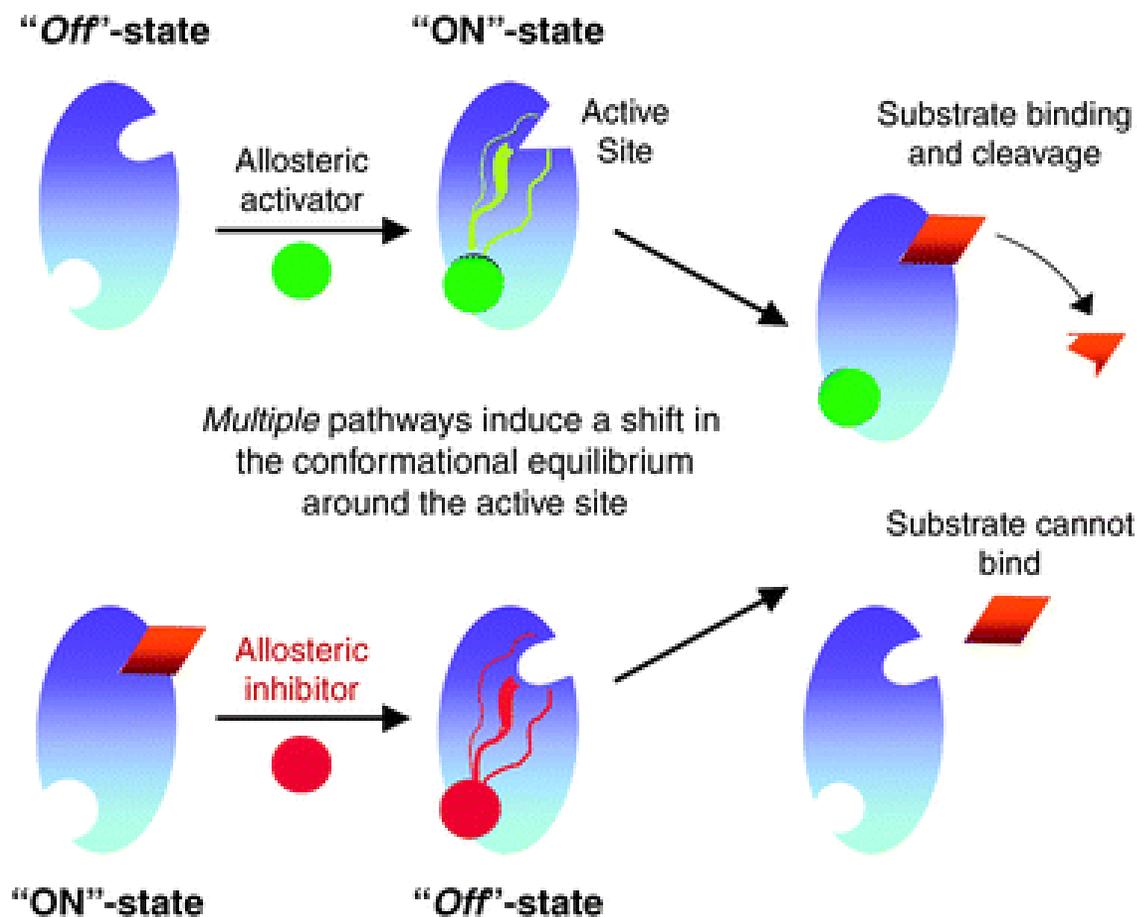
## 磷酸甘油酸脱氢酶和癌症



在黑色素瘤细胞和乳腺癌细胞中，磷酸甘油酸脱氢酶过表达

## 二、课题意义

- 1、基于结构的药物设计
- 2、别构抑制剂的设计



别构效应

### 三、生物信息学的应用

## 第一部分 一级结构序列分析

- 一级序列
- 氨基酸成分分析
- 蛋白质理化性质分析
- 疏水性分析
- 重复序列分析

# 第一部分 一级结构序列分析 一级序列

From Uniprot

226 个氨基酸的结构是缺失的

From PDB

6

```

10      20      30      40      50      60
MAFANLRKVL ISDSLDPCCR KILQDGGGLQV VEKQNLKSEE LIAELQDCEG LIVRSATKVT

70      80      90      100     110     120
ADVINA AEKL QVVGRAGTGV DNV DLEAATR KGILVMNTPN GNSLSAAELT CGMIMCLARQ

130     140     150     160     170     180
IPQATASMKD GKWERKKFMG TELNGKTLGI LGLGRIGREV ATRMQSFGMK TIGYDPIISP

190     200     210     220     230     240
EVSASFGVQQ LPLEEIWPLC DFITVHTPLL PSTTGLLNDN TFAQCKKGVR VVNCARGGIV

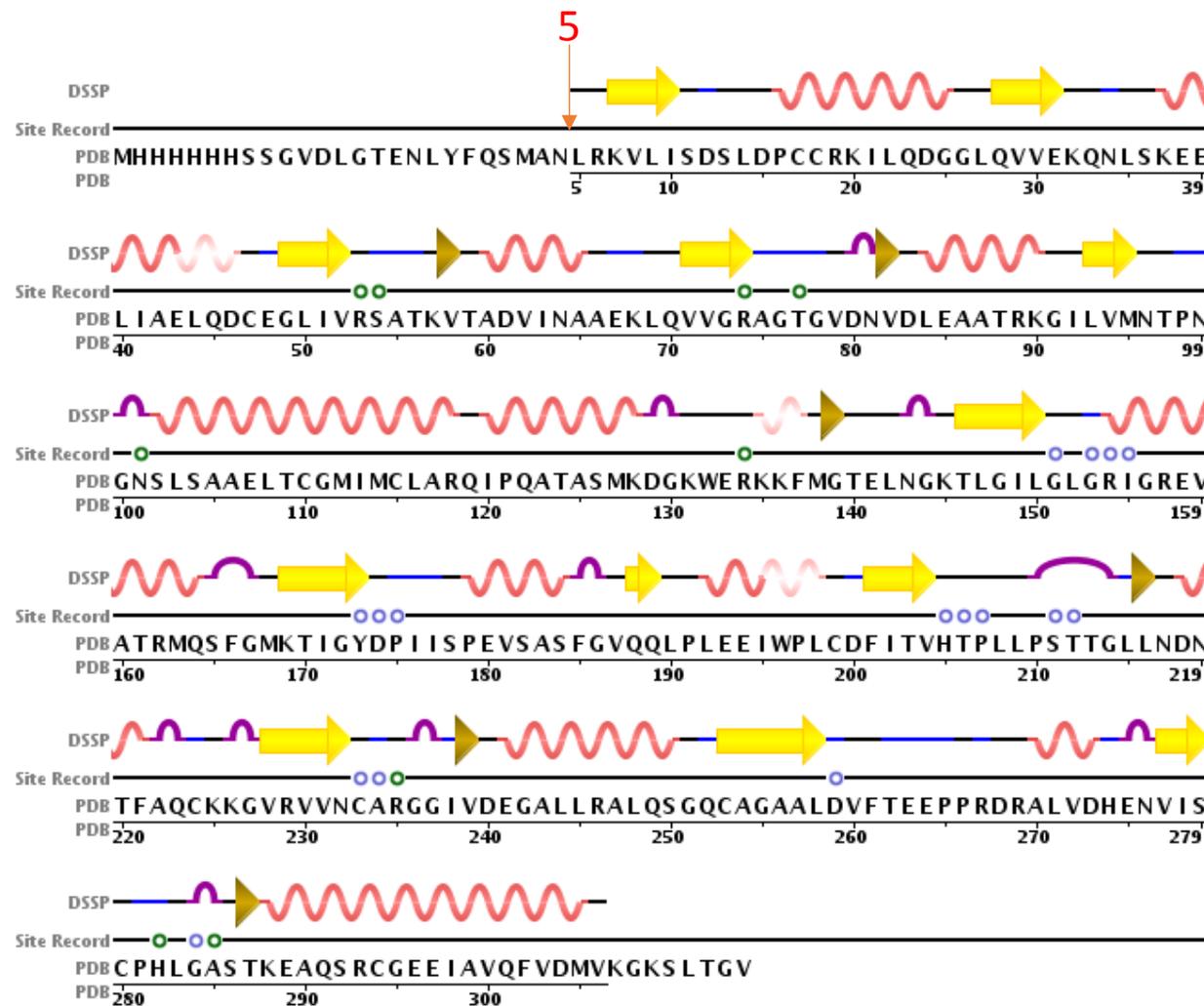
250     260     270     280     290     300
DEGALLRALQ SGQCAGAALD VFTEPPRDR ALVDHENVIS CPHLGASTKE AQSRCGEEIA

310     320     330     340     350     360
VQFVDMVKGK SLTGVVNAQA LTSAFSPHTK PWIGLAEALG TLMRAWAGSP KGTIQVITQG

370     380     390     400     410     420
TSLKNAGNCL SPAVIVGLLK EASKQADVNL VNAKLLVKEA GLNVTTS HSP AAPGEQGFGE

430     440     450     460     470     480
CLLAVALAGA PYQAVGLVQG TTPVLQGLNG AVFRPEVPLR RDLPLLLFRT QTSDPAMLPT

490     500     510     520     530
MIGLLAEAGV RLLSYQTSLV SDGETWHVMG ISSLLPSLEA WKQHVTEAFQ FHF
    
```



Site Record Legend

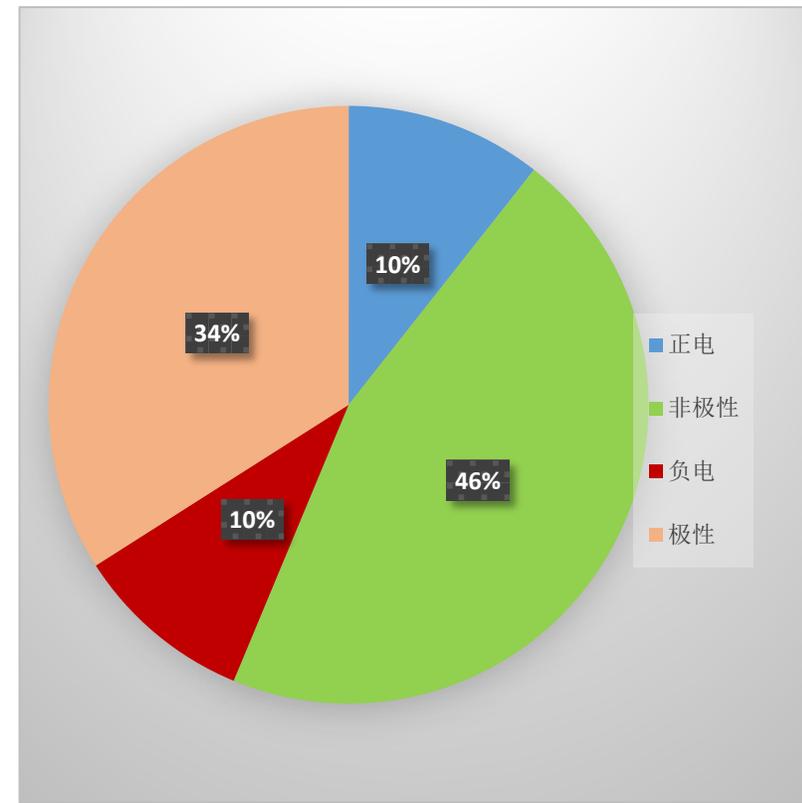
- BINDING SITE FOR RESIDUE NAD A 501 (SOFTWARE)
- BINDING SITE FOR RESIDUE NAD A 502 (SOFTWARE)

# 第一部分 一级结构序列分析

## 氨基酸组成分析

使用ProtParam分析氨基酸成分：

氨基酸	数目	比例	氨基酸	数目	比例
His (H)	8	1.50%	Asp (D)	20	3.80%
Arg (R)	22	4.10%	Glu (E)	31	5.80%
Lys (K)	26	4.90%	Cys (C)	13	2.40%
Ala (A)	54	10.10%	Gln (Q)	27	5.10%
Ile (I)	24	4.50%	Gly (G)	50	9.40%
Leu (L)	65	12.20%	Ser (S)	32	6.00%
Met (M)	13	2.40%	Thr (T)	35	6.60%
Phe (F)	15	2.80%	Asn (N)	19	3.60%
Val (V)	44	8.30%	Tyr (Y)	3	0.60%
Pro (P)	26	4.90%	Trp (W)	6	1.10%



非极性和极性氨基酸占据了绝大部分，可能和一些domain链接的地方，比如ASB domain和底物domain的多个gly。还有，倾向helix以及折叠等二级结构。

# 第一部分 一级结构序列分析

## 理化性质分析

利用PEPSTATS 分析,

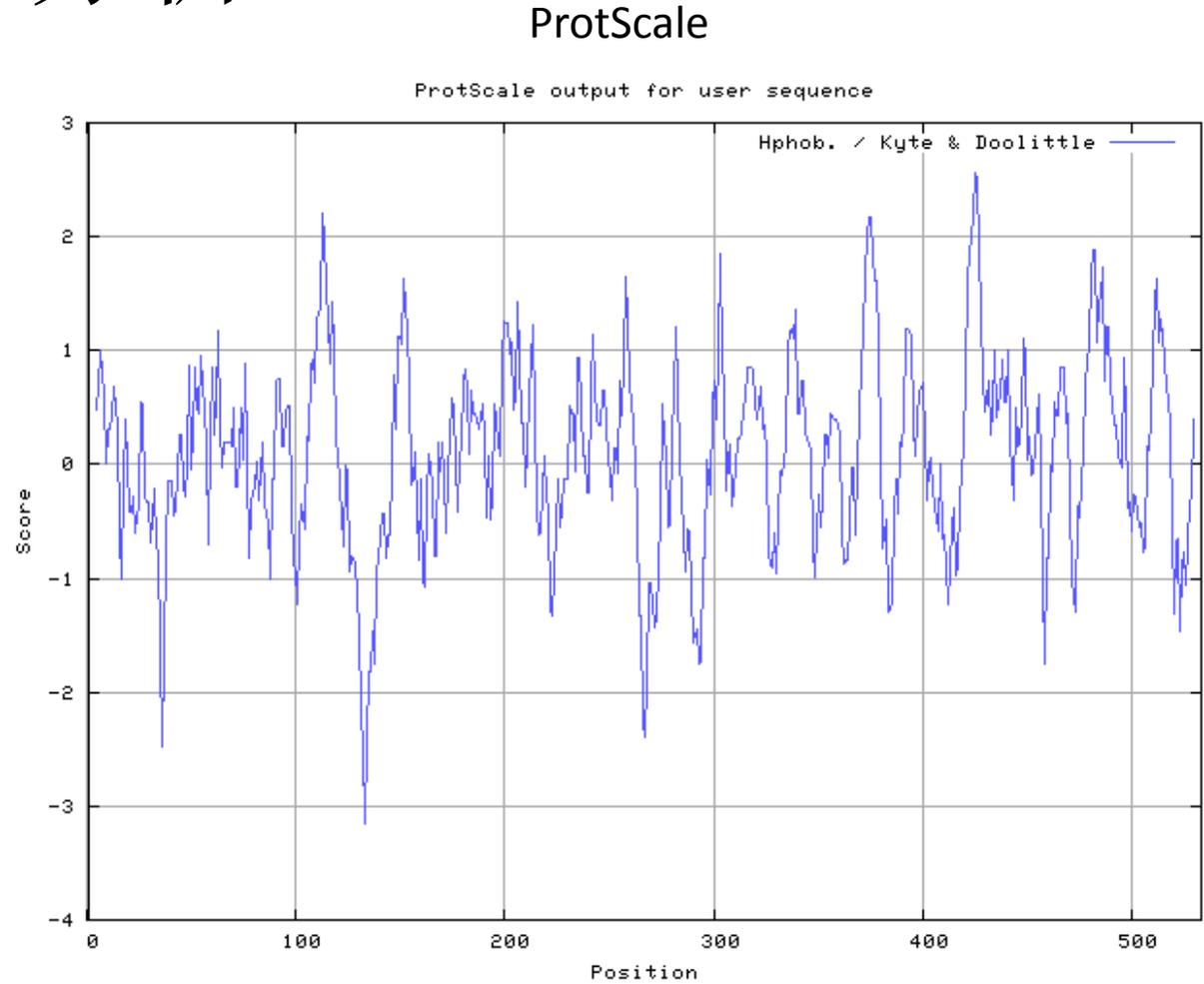
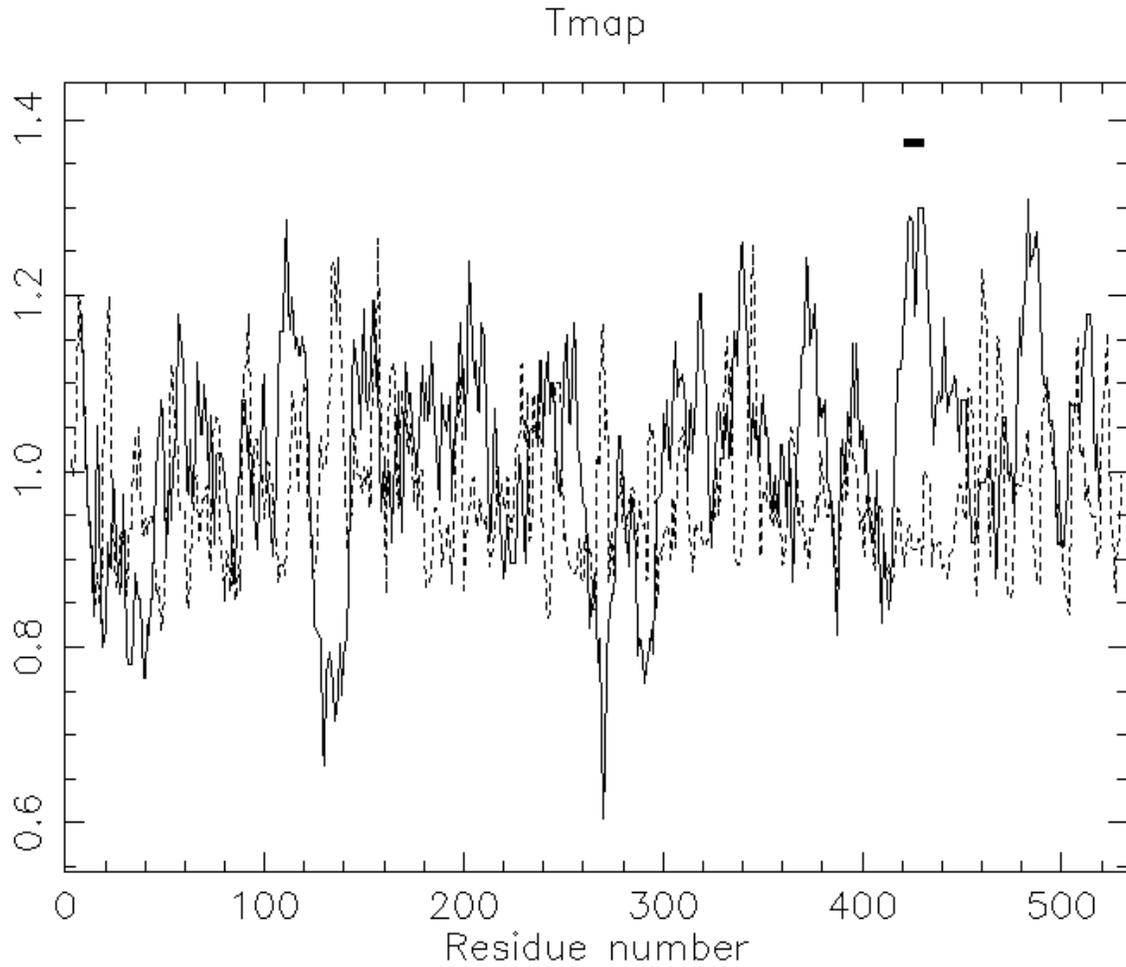
Molecular weight = 56650.34

Property	Residues	Number	Mole%
Tiny	( A+C+G+S+T )	184	34.522
Small	( A+B+C+D+G+N+P+S+T+V )	293	54.972
Aliphatic	( I+L+V )	133	24.953
Aromatic	( F+H+W+Y )	32	6.004
Non-polar	( A+C+F+G+I+L+M+P+V+W+Y )	313	58.724
Polar	( D+E+H+K+N+Q+R+S+T+Z )	220	41.276
Charged	( B+D+E+H+K+R+Z )	107	20.075
Basic	( H+K+R )	56	10.507
Acidic	( B+D+E+Z )	51	9.568

进一步分析, 体积较小的氨基酸含量较多

# 第一部分 一级结构序列分析

## 疏水性分析

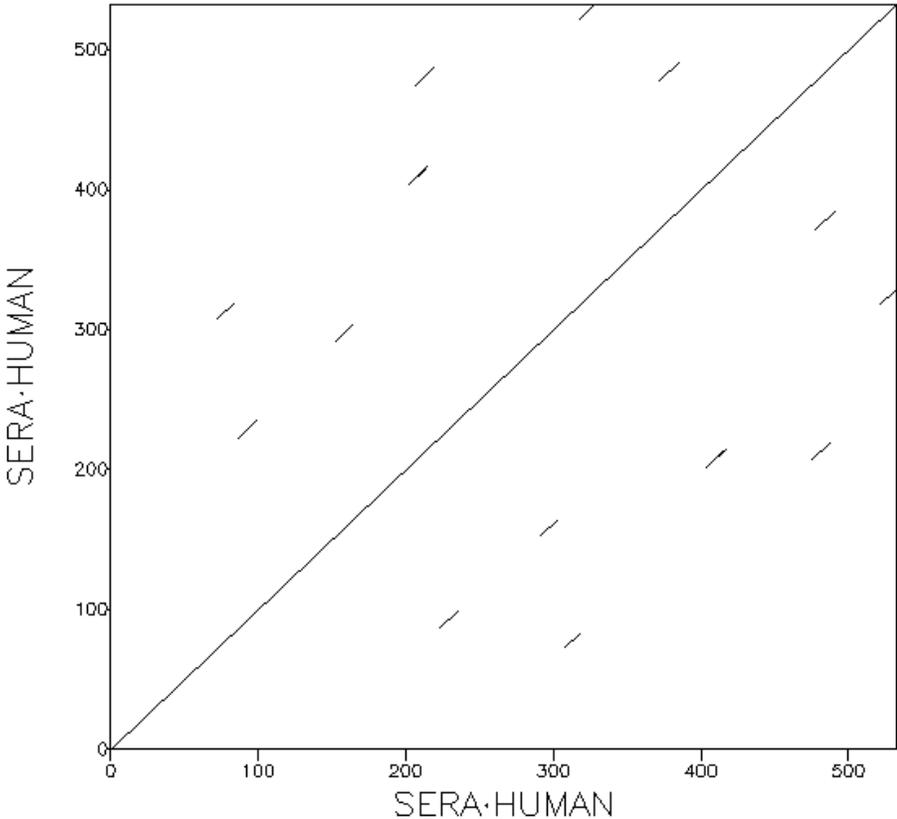


没有明显的跨膜螺旋，蛋白处于细胞质中。亲疏水性间隔，有较多的二级结构

# 第一部分 一级结构序列分析

## 重复序列分析

Dotmatcher: fasta::748474:SERA·HUMAN vs fasta::748474:SE...  
(windowsize = 10, threshold = 23.00 18/06/14)



## 第一部分 一级结构序列分析

# 第一部分 小结

- ✓ PDB数据库的小错误，在做突变等实验时需要注意
- ✓ 成分上疏水的、以及体积小的氨基酸较多
- ✓ 序列上有7段小的重复序列

## 第二部分 二级结构预测

# 第二部分 二级结构预测

- 蛋白domain分析
- 二级结构预测
- Weblogo

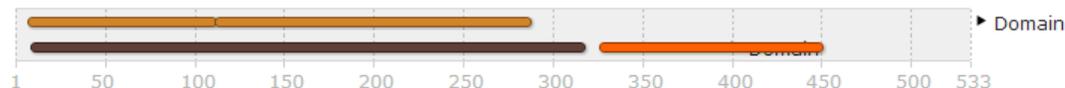
## 第二部分 二级结构预测

# 蛋白domain分析

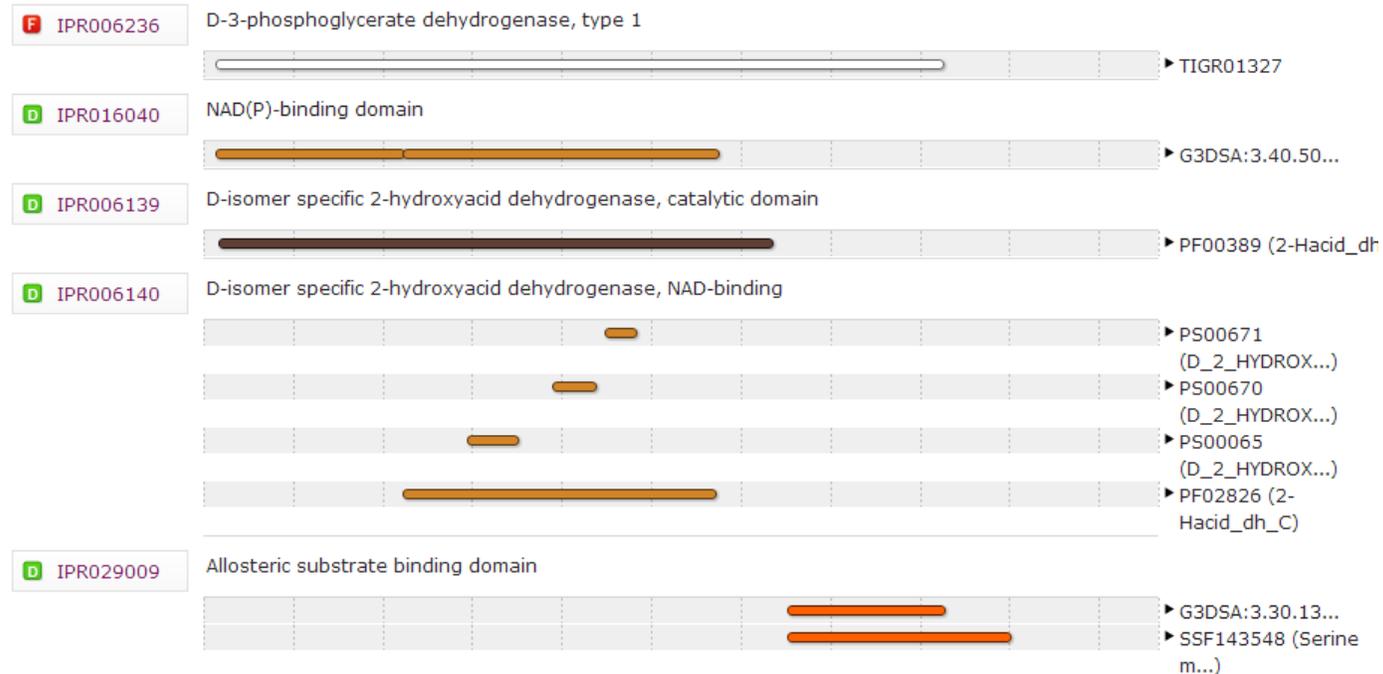
### Protein family membership

■ D-3-phosphoglycerate dehydrogenase, type 1 (IPR006236)

### Domains and repeats



### Detailed signature matches

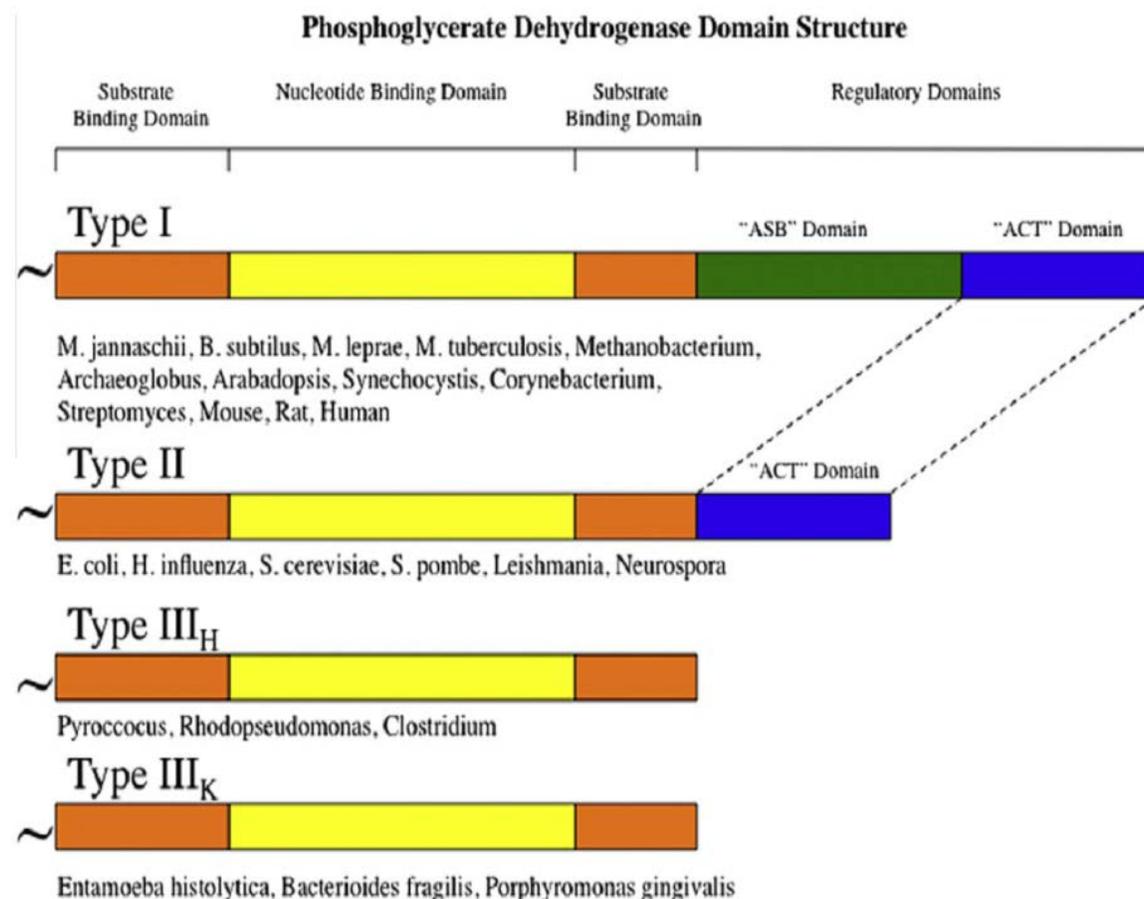


可以检测到分为3个domain, NAD binding domain, Substrate domain, ASB domain

缺少ACTdomain

## 第二部分 二级结构预测

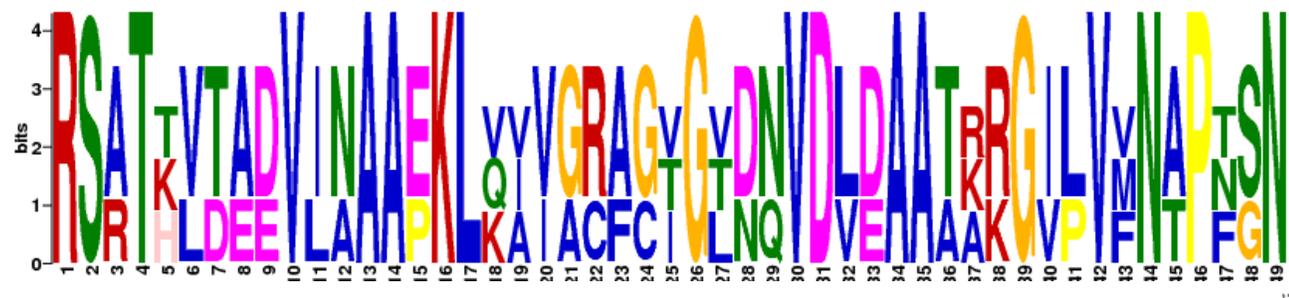
# 蛋白domain分析



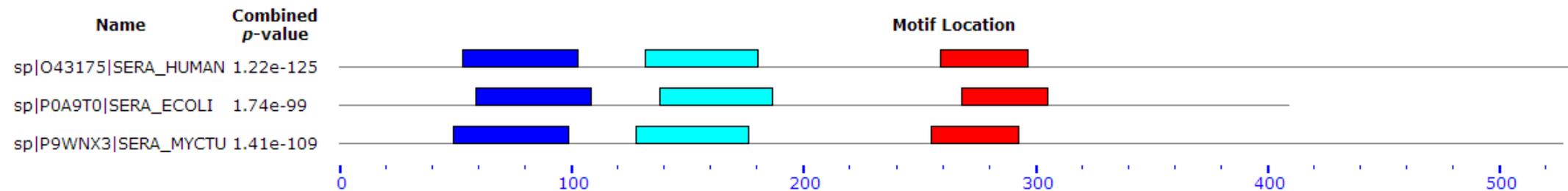


# 第二部分 二级结构预测

# Weblogo



Motif 1 Motif 2 Motif 3



# 第二部分 小结

- ✓从DOMAIN预测中可以看到，PHGDH三个特征的domain，因此同源预测未解出的结构部分是可能的
- ✓从二级结构可以看出，未解出结构的308-533位氨基酸结构含有大量的二级结构，其预测的可靠性比较高。这样的结果，对于引入新的药物设计位点十分重要，也比较可行
- ✓从weblogo的结果来看，3个motif被识别

## 第三部分 三级结构预测

# 第三部分 三级结构预测

Swiss-model

Phyre2

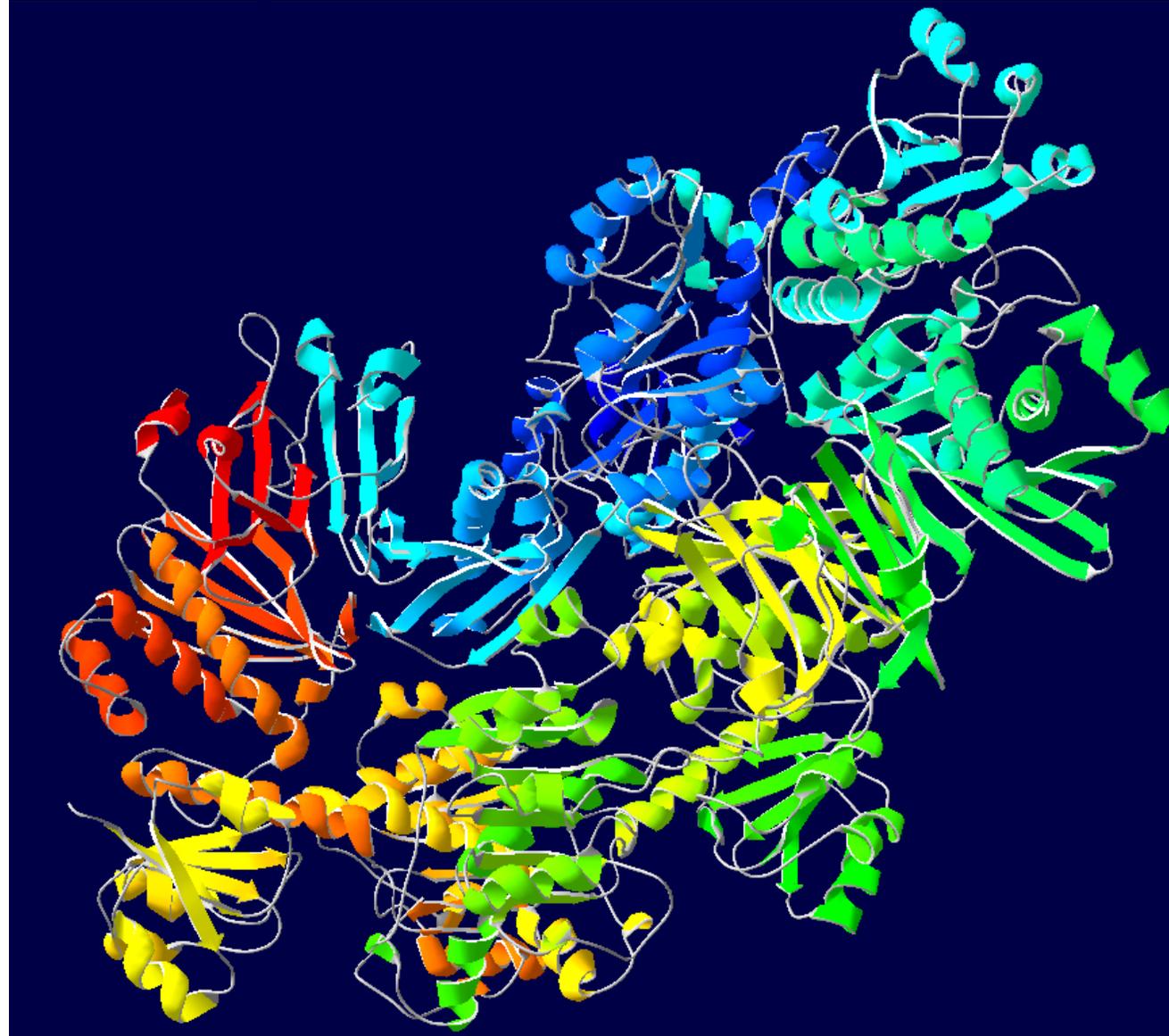
### 第三部分 三级结构预测

## SWISS-MODEL **VS** Phyre2

项目	Phyre2	Swiss-model
模板	<u>1ygy. A</u>	<u>3ddn. 1. A/1ygy. A/ 1ygy. B</u>
Identity	34.06%	34.06%
花费时间 (h)	7	4
模式	Intensive mode	Automated mode

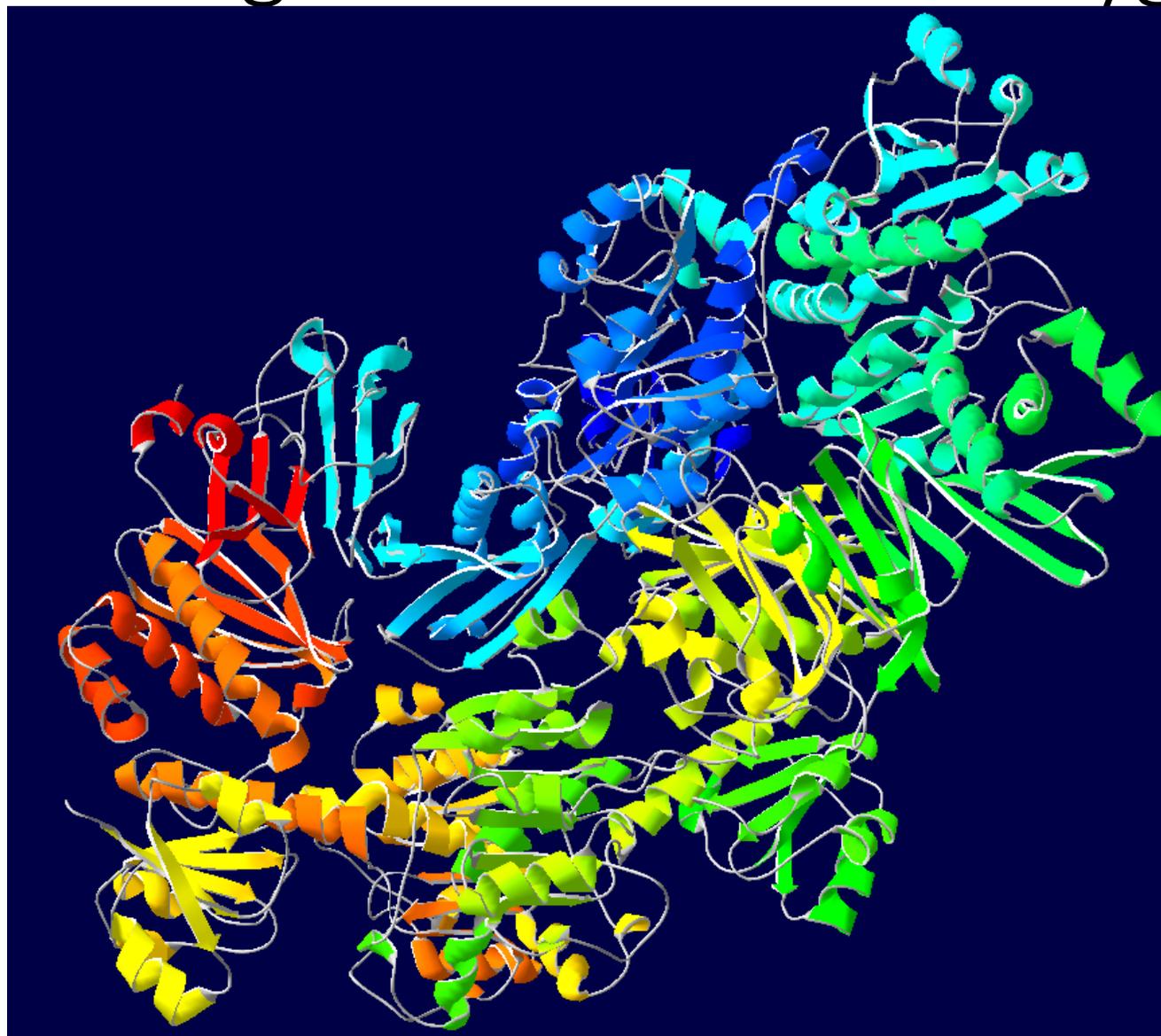
### 第三部分 三级结构预测

PHGDH-full length-SWISS-MODEL-3ddn.A



### 第三部分 三级结构预测

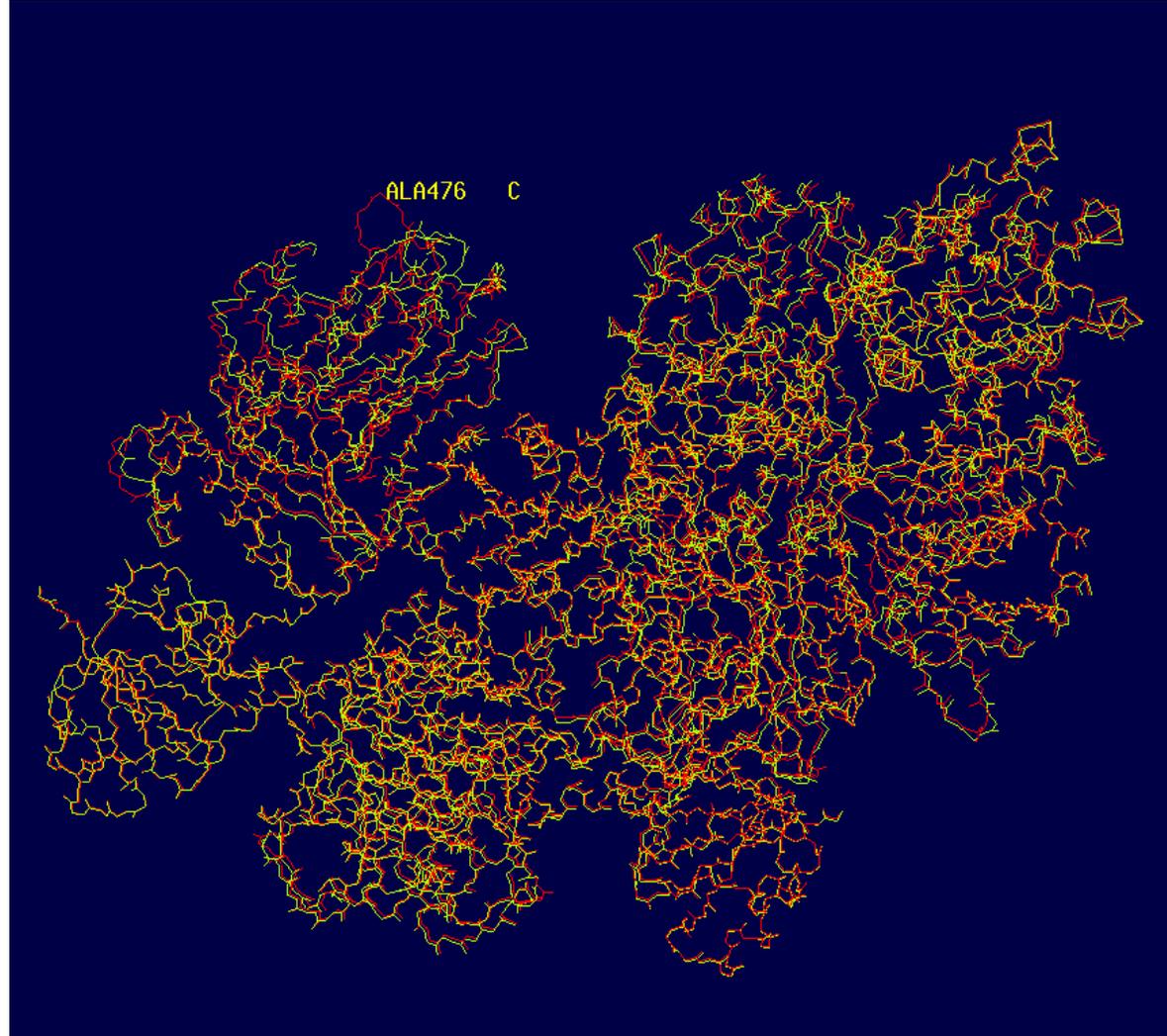
PHGDH-full length-SWISS-MODEL-1ygy.A



### 第三部分 三级结构预测

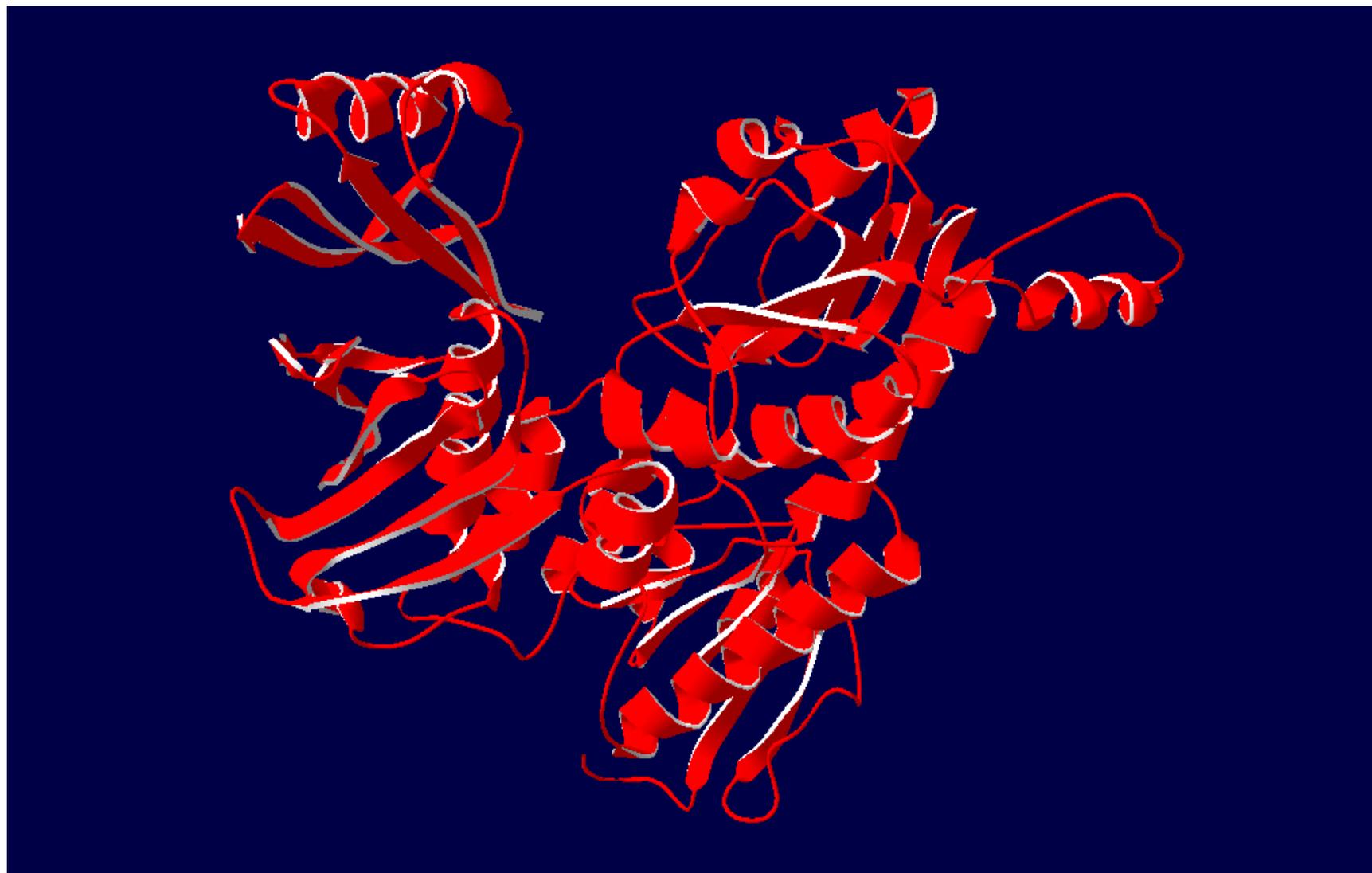
# SWISS-MODEL-3ddn.A VS SWISS-MODEL-1ygy.A

Red VS Green



RMSD: 1.02 A

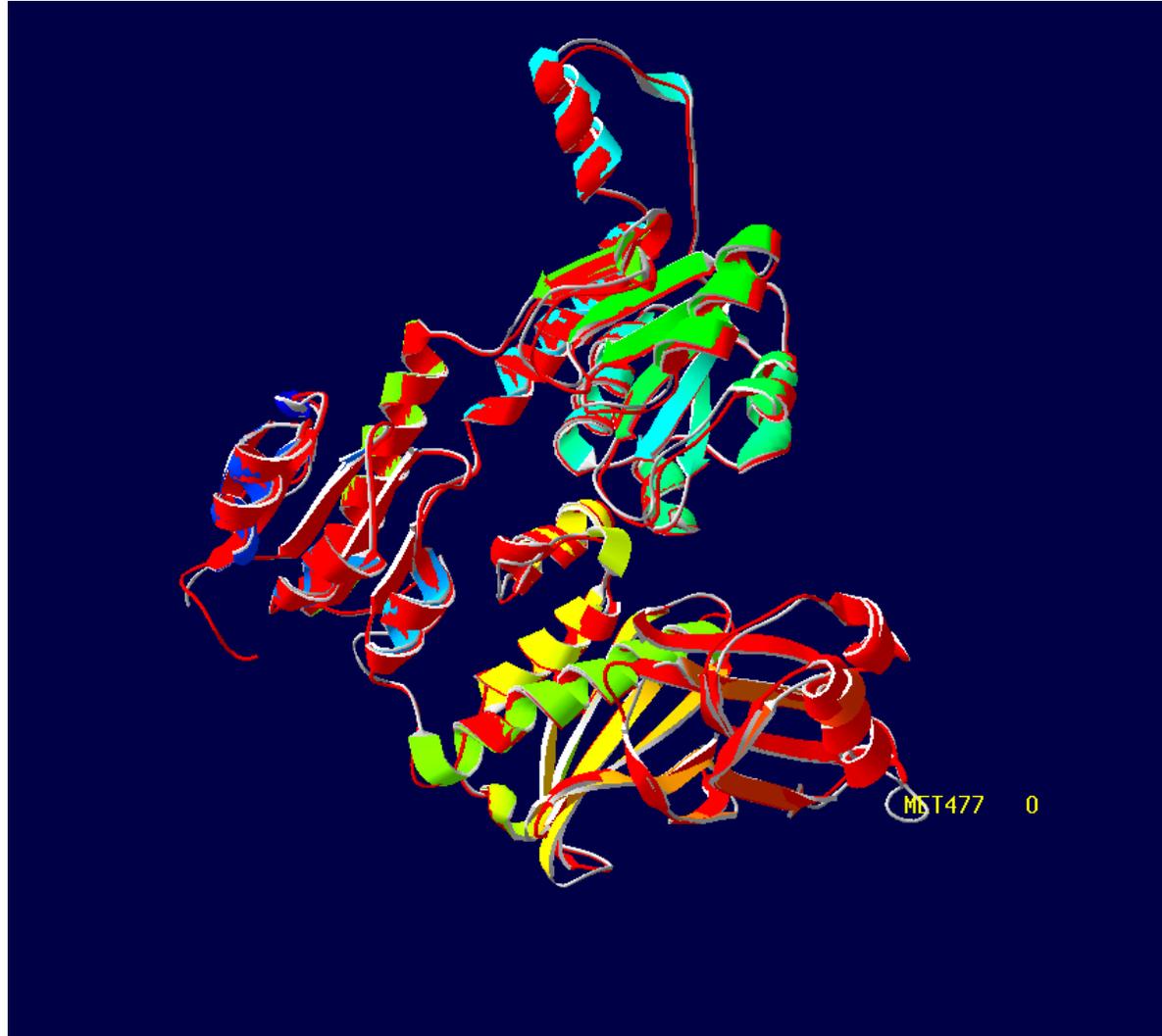
# PHGDH-full length-Phyre2



### 第三部分 三级结构预测

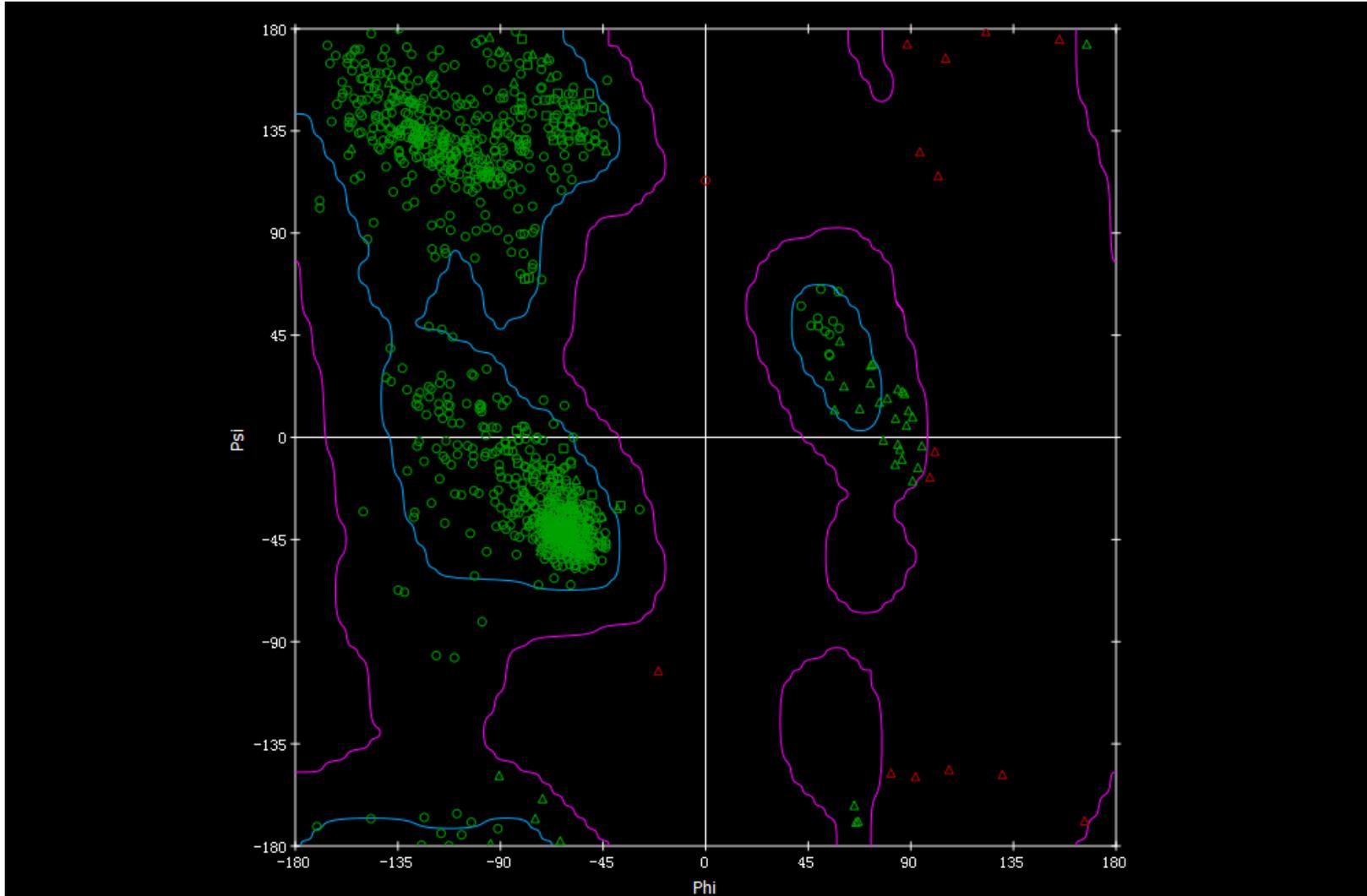
# SWISS-MODEL-3ddn.A VS Phyre2-1ygy.A

Color VS Red



RMDS: 1.75 A

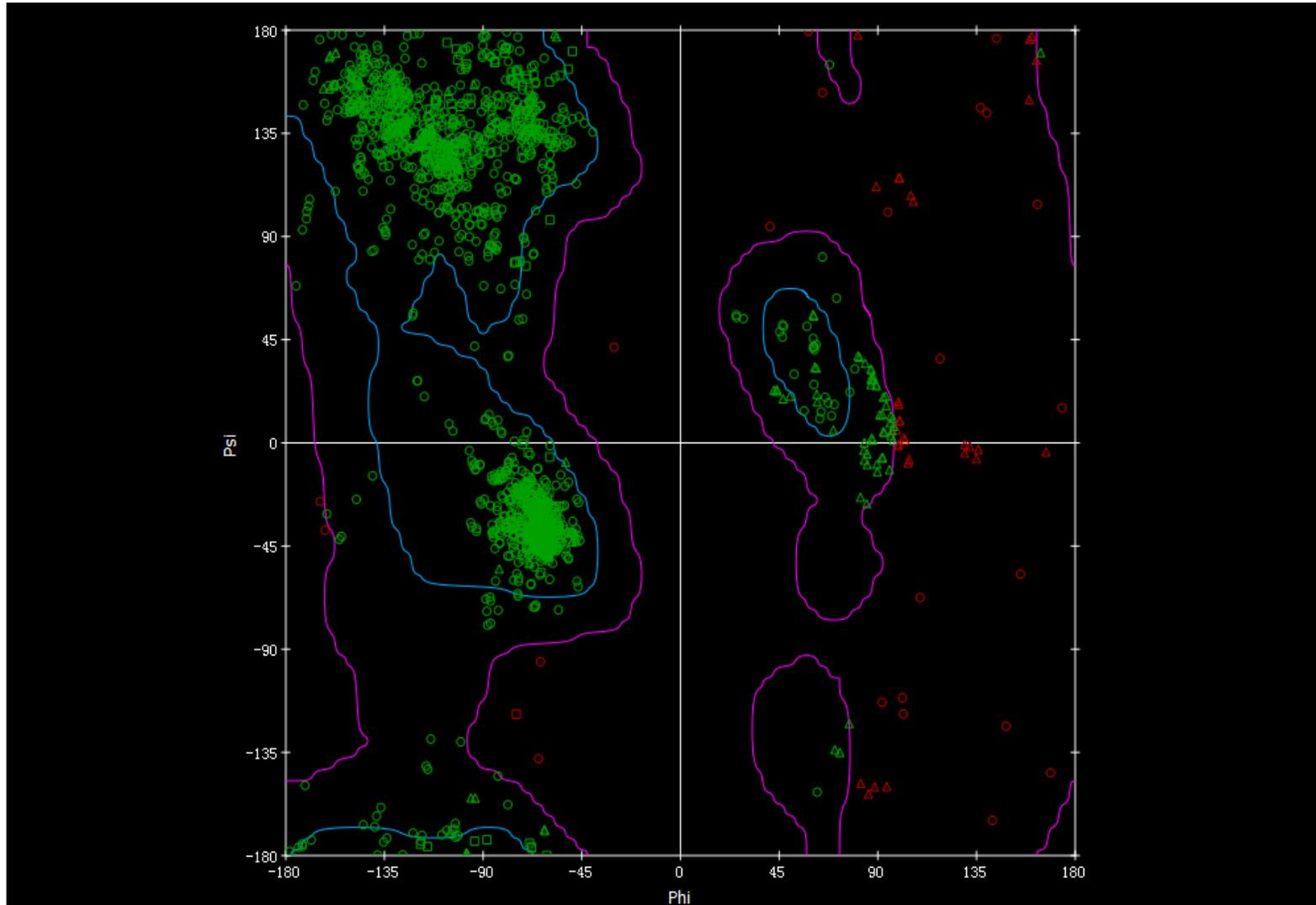
# 模型评估



PHGDH-full length-phyre2

拉氏图

## 模型评估



PHGDH-full length-SWISS-MODEL-3ddn.A

拉氏图

# 第三部分 小结

- ✓ 在Identity较低时，不适合使用自动建模
- ✓ 自动建模的不能按照自己的需求进行，如果要建立满意可靠的结果，最好自己调节参数进行建模
- ✓ Phyre2的intensive模式的模建结果可靠性高于Swiss-model
- ✓ Swiss-model和Phyre2所做的优化比较少

# 结论与展望

## 总结

在分析一级结构和二级结构的基础上，使用Swiss-model以及Phyre2进行了自动建模，并使用Modeler初步建模

## 展望

在使用Modeler建模的基础上，根据文献报道的重要位点精修（比如ASB Domain的关键残基），使用分子动力学进一步能量优化，得到合理的结构。在此基础上，进行进一步的药物设计，得到可能的抗癌的药物

# 致谢

- 罗老师精彩的讲课，耐心的指导
- 来鲁华教授，刘莹副教授，各位组员的指导老师
- 所有的组员（景誉庆、梁振、米丽拜尔、吴洋、王玉生）在这一学期的讨论，和最后的报告中细节的讨论和付出
- 全班所有同学—pku2014

**Thanks!**