



细胞肿瘤抗原P53蛋白的序列, 结构和功能分析

Sequence, structure and function analysis of the cellular tumor antigen P53

G09小组: 代博涛_G09A

林珮璐_G09B

张 正_G09C

沈 可_G09D

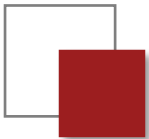
张芷妍_G09E

汇报人: 代博涛

2023-02-09

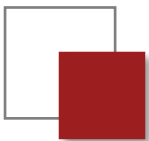
Content

- 1 背景介绍
- 2 序列和功能分析
- 3 系统发生学分析
- 4 蛋白结构特征和预测
- 5 课程总结与感悟



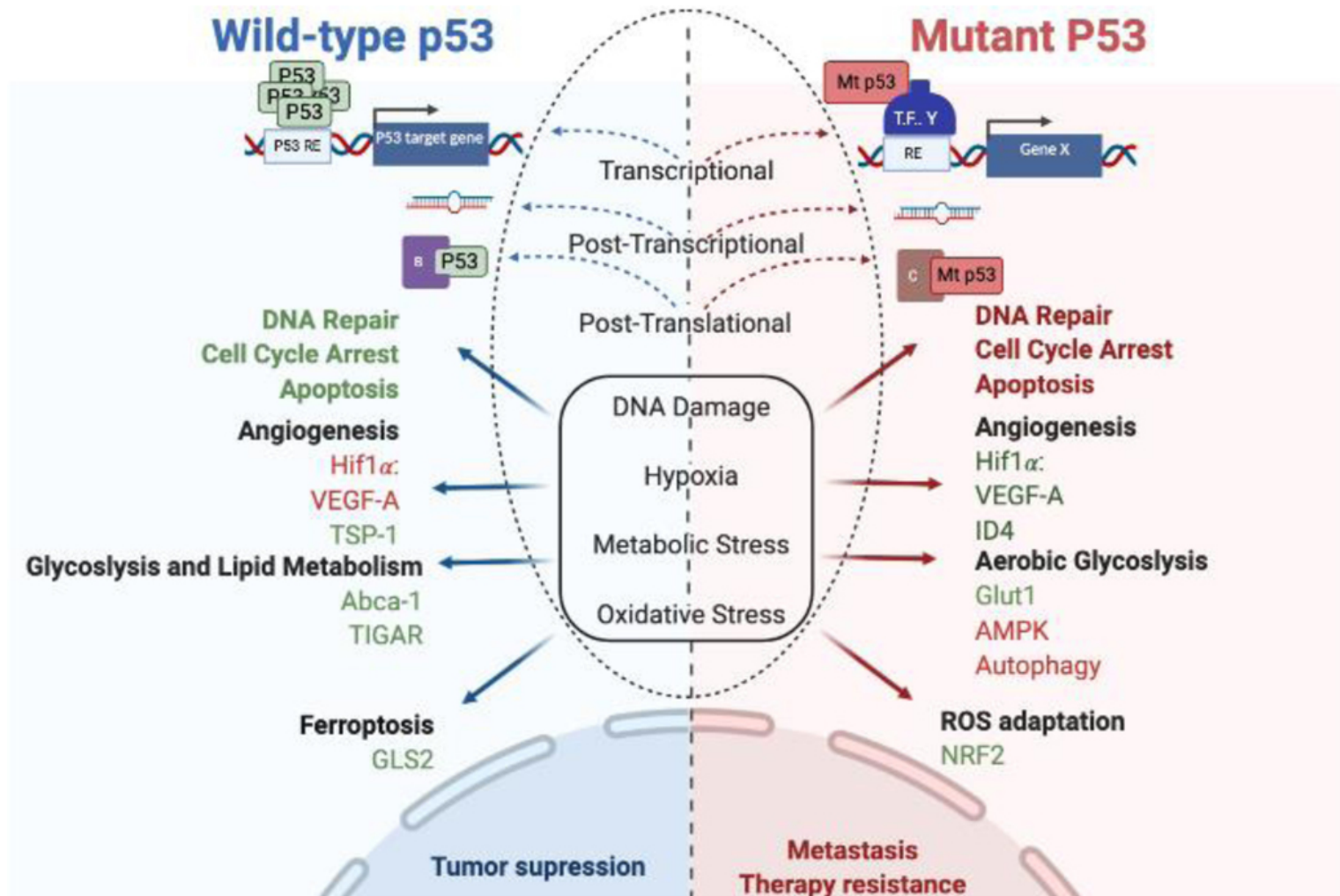
细胞守护者——P53

- P53 基因位于人类 **17p13.1** 染色体短臂上，全长约 **16-20 kb**，由11个外显子和10个内含子构成，是一种核结合蛋白，大多数肿瘤都与 P53 的突变有关。
- P53肿瘤抑制因子**通常水平较低**，但当检测到DNA损伤时，**p53水平上升**并启动保护措施。P53与基因组中的许多调控位点结合，开始产生阻止细胞分裂的蛋白质，直到损伤得到修复。如果损伤太严重，p53会启动**细胞程序性死亡或凋亡**的过程，它会引导细胞自杀，永久性地消除损伤。
- 癌细胞通常包含两种类型的突变：一种是导致细胞**不受控制的生长和增殖的突变**，另一种是**阻止防止非自然生长的正常防御的突变**。P53属于第二类，P53基因的突变导致了大约一半的人类癌症病例。其中大多数是**错义突变**，改变了DNA中某个位置的信息，导致细胞错误地构建p53，在蛋白质链的某个位置替换了不正确的氨基酸。在这些突变体中，p53的正常功能被阻断，会导致其无法阻止受损细胞中的增殖。如果细胞有其他突变导致不受控制的生长，细胞就会**发展成肿瘤**。



1. 背景介绍

野生型和突变型p53在调控癌症生物学中的相反作用



2.1 人P53蛋白的序列和功能特征

🌟 P04637 · P53_HUMAN

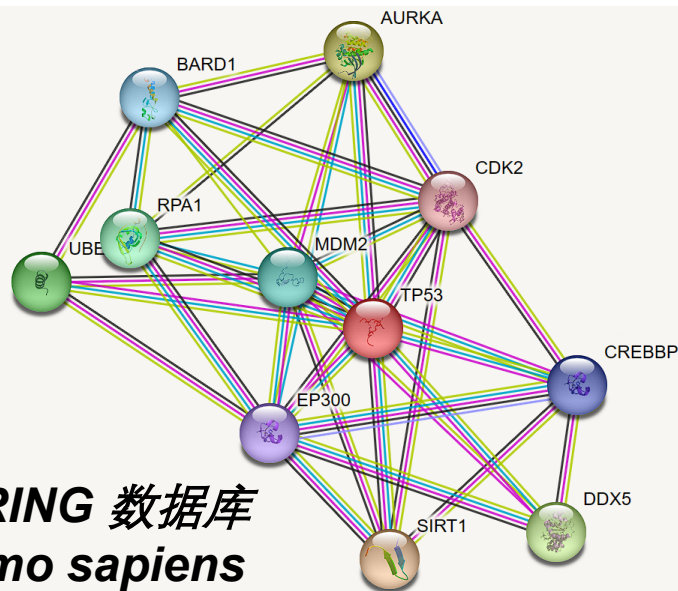
Protein ⁱ	Cellular tumor antigen p53	Amino acids	393
Gene ⁱ	TP53	Protein existence ⁱ	Evidence at protein level
Status ⁱ	UniProtKB reviewed (Swiss-Prot)	Annotation score ⁱ	5/5
Organism ⁱ	Homo sapiens (Human)		

- Site 120: 与DNA的作用位点
- DNA binding 102-292: DNA结合区
- Binding site: (锌指结构) C176, H179, C238, C242
- PTM: 主要为磷酸化修饰

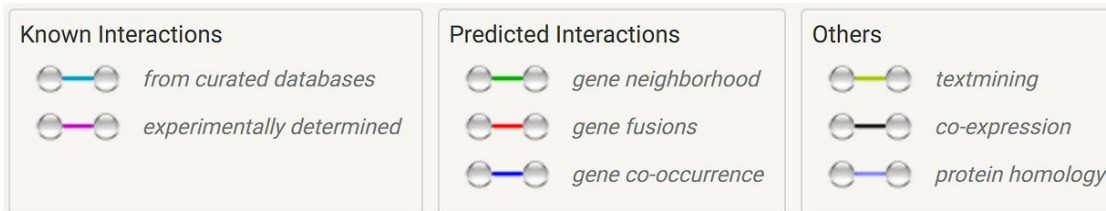
Uniprot 数据库 Feature viewer



2.2 人P53蛋白-蛋白互作网络特征

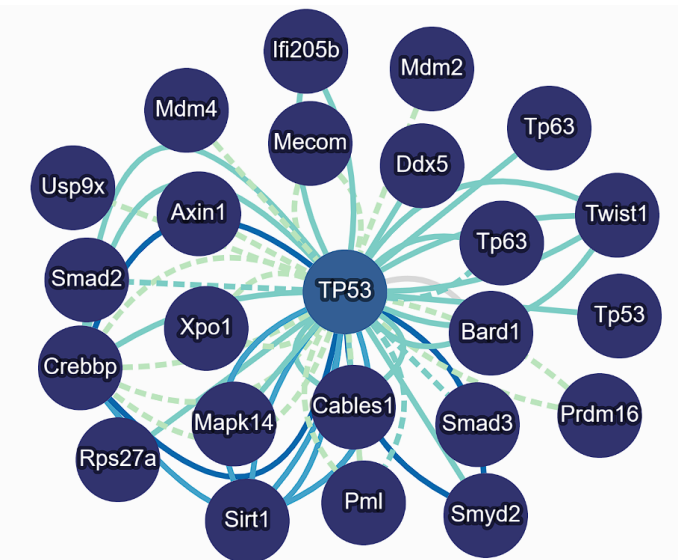


STRING 数据库
Homo sapiens



Predicted Functional Partners:

SIRT1	<i>NAD-dependent protein deacetylase sirtuin-1; NAD-dependent protein deacetylase that links transcriptional regulation dir...</i>
AURKA	<i>Aurora kinase A; Mitotic serine/threonine kinase that contributes to the regulation of cell cycle progression. Associates ...</i>
DDX5	<i>Probable ATP-dependent RNA helicase DDX5; Involved in the alternative regulation of pre-mRNA splicing; its RNA helicase...</i>
UBE3A	<i>Ubiquitin-protein ligase E3A; E3 ubiquitin-protein ligase which accepts ubiquitin from an E2 ubiquitin-conjugating enzym...</i>
RPA1	<i>Replication protein A 70 kDa DNA-binding subunit; As part of the heterotrimeric replication protein A complex (RPA/RP-A...</i>
MDM2	<i>E3 ubiquitin-protein ligase Mdm2; E3 ubiquitin-protein ligase that mediates ubiquitination of p53/TP53, leading to its deg...</i>
BARD1	<i>BRCA1-associated RING domain protein 1; E3 ubiquitin-protein ligase. The BRCA1-BARD1 heterodimer specifically media...</i>
CREBBP	<i>CREB-binding protein; Acetylates histones, giving a specific tag for transcriptional activation. Also acetylates non-histon...</i>
EP300	<i>Histone acetyltransferase p300; Functions as histone acetyltransferase and regulates transcription via chromatin remod...</i>
CDK2	<i>Cyclin-dependent kinase 2; Serine/threonine-protein kinase involved in the control of the cell cycle; essential for meiosis, ...</i>



Nodes

Color ~ Species

- Homo sapiens
- Mus musculus

Shape ~ Type

- protein

Edges

Color ~ Type

- colocalization
- association
- physical association
- direct interaction
- enzymatic reaction

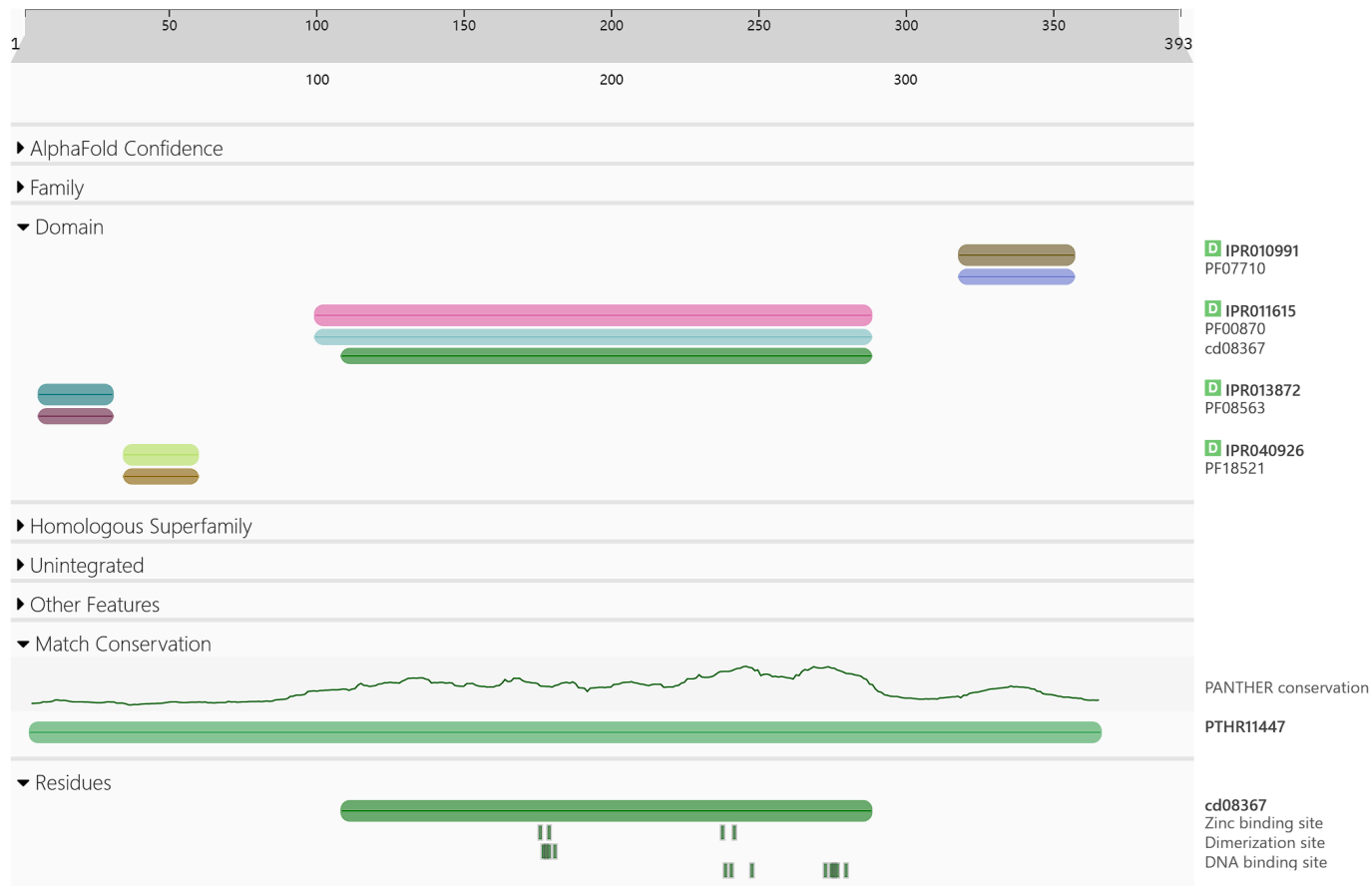
Shape ~ Expansion method

- spoke expansion

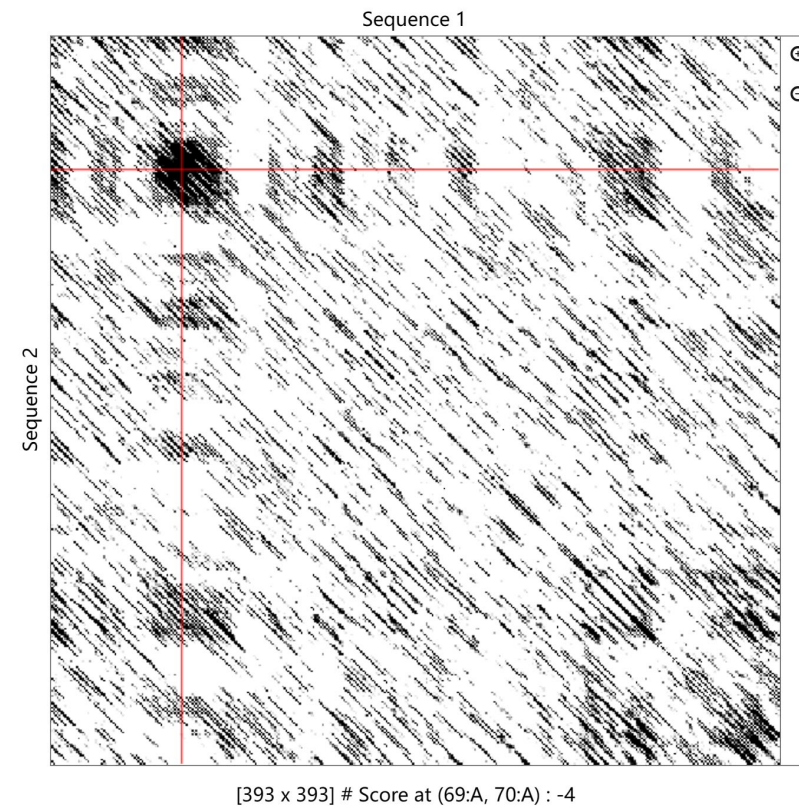
IntAct 数据库
Homo sapiens
&
Mus musculus

2.3 人P53蛋白保守域分析

InterPro数据库 p53蛋白家族序列保守域分析



Dotlet p53_human序列自身重复性分析

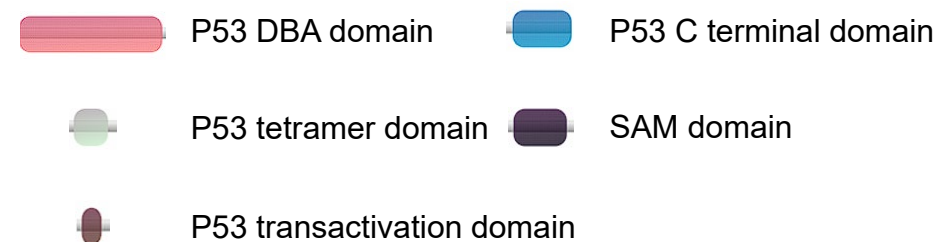
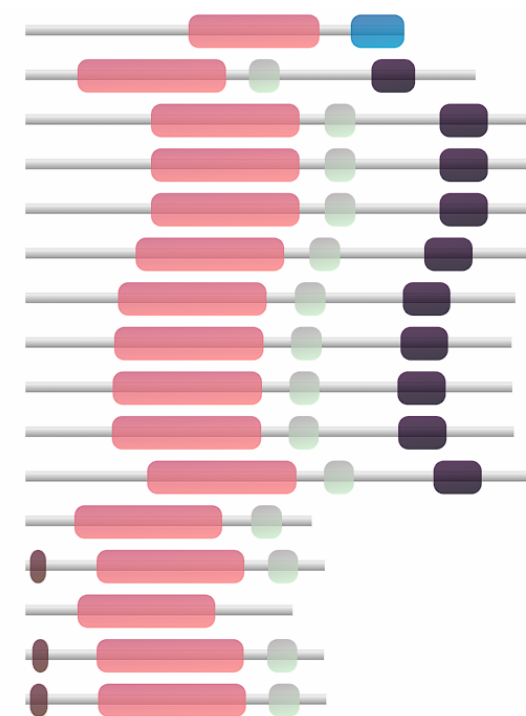
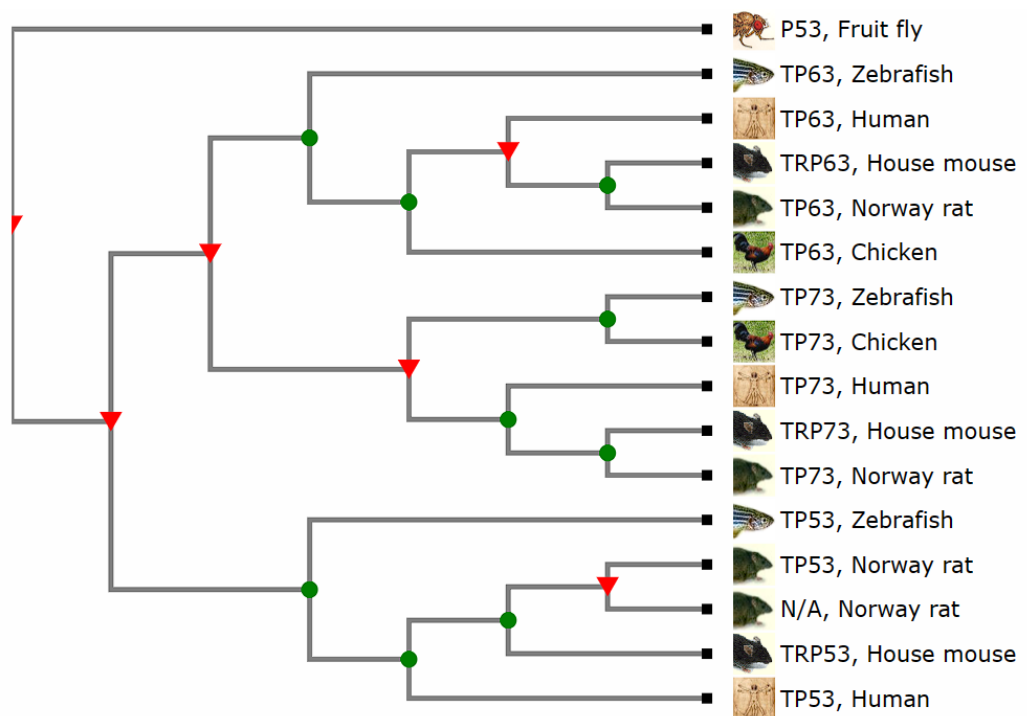


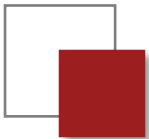
- P53_human序列自身存在重复片段，但片段长度不长

- P53家族主要的保守域为反式激活区、DNA结合区、四聚化区

2.3 P53蛋白家族进化树和保守域分析

TreeFam 数据库





2.4 蛋白质序列比对——Emboss



#####

```

# Program: needle
# Rundate: Mon 6 Feb 2023 03:02:50
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20230206-030412-0929-64386536-p1m.asequence
# -bsequence emboss_needle-I20230206-030412-0929-64386536-p1m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout

```

#####

=====

```

#
# Aligned_sequences: 2
# 1: P53_HUMAN
# 2: P53_MOUSE
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 396
# Identity: 304/396 (76.8%)
# Similarity: 326/396 (82.3%)
# Gaps: 9/396 ( 2.3%)
# Score: 1558.0
#
#
#
=====

```

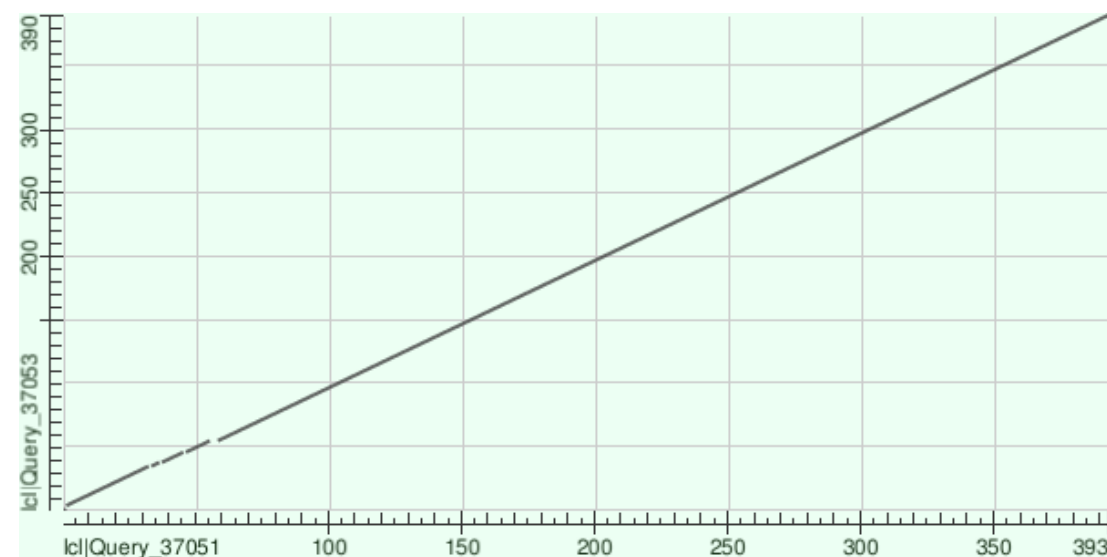
```

P53_HUMAN      1  —MEEPQSDPSVEPPLSQETFSDLWKLLENVLSPLPSQAMDDLMLSF  47
      ||| ||| |:| ||||| ||||| :|| | | | | | | |
P53_MOUSE      1  MTAMEESQSDISLELPLSQETFSGLWKLPPEDIL-PSP-HCDDLLLL-P  47
      ||| ||| |:| ||||| ||||| :|| | | | | | | |
P53_HUMAN     48  DDIEQWFTEDPGFDEAPRMPEAAPPVAPAPAAPTAAAPAPAPSWPLSSSV  97
      .:|:|:| | | | | | | | | | | | | | | | | | | |
P53_MOUSE     48  QDVVEFF—EGPSEALRVSGAPAAQDPVTETPGPVAPAPATPWPLSSFV  94
      ||| ||| |:| ||||| ||||| :|| | | | | | | |
P53_HUMAN     98  PSQKTYQGSYGFRGLHSGTAKSVTCTYSPALNEMFCQLAKTCPVQLWV  147
      ||||| ||| ||| ||||| ||||| ||| | | | | | | | |
P53_MOUSE     95  PSQKTYQGNVGFHLGLQSGTAKSVMCTYSPPLNKLFCQLAKTCPVQLWV  144
      ||||| ||| ||| ||||| ||||| ||| | | | | | | | |
P53_HUMAN    148  DSTPPPGTRVRAMAIYKQSQHMTVEVRRCPHHERCSDSDGLAPPQHILRV  197
      .: ||| |: | | | | | | | | | | | | | | | | | | |
P53_MOUSE    145  SATPPAGSRVRAMAIYKQSQHMTVEVRRCPHHERCSDGDGLAPPQHILRV  194
      .: ||| |: | | | | | | | | | | | | | | | | | | |
P53_HUMAN    198  EGNLRVEYLDNRNTRFRHSVVPYEPPEVGSDCCTTIHYNYMCNSSCMGGMN  247
      |||| . |||: || | | | | | | | | | | | | | | | | |
P53_MOUSE    195  EGNLYPEYLEDRQTFRHSVVPYEPPEAGSEYTTIHYKVMCNSSCMGGMN  244
      |||| . |||: || | | | | | | | | | | | | | | | | |
P53_HUMAN    248  RRPILTIITLEDSSGNLLGRNSFEVRCACPGDRDRTEENLRKKGEPHH  297
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
P53_MOUSE    245  RRPILTIITLEDSSGNLLGRDSFEVRCACPGDRDRTEENFRKKEVLCF  294
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
P53_HUMAN    298  ELPPGSTKRALPNNTSSSPQPKKPLDGEYFTLQIRGRERFEMFRELNEA  347
      ||||| ||||| . ||: || . ||||| ||||| ||||| ||||| |||||
P53_MOUSE    295  ELPPGSAKRALPTCTSASPPQKKPLDGEYFTLQIRGRERFEMFRELNEA  344
      ||||| ||||| . ||: || . ||||| ||||| ||||| ||||| |||||
P53_HUMAN    348  LELKDAQAGKEPGGSAHSSHLKSKKQSTSRHKKLMFKTEGPDSD  393
      ||||| . |:| | | | | | | | | | | | | | | | | | |
P53_MOUSE    345  LELKDAHATEESGDSRAHSSYLKTKKQSTSRHKKTMVKKVGPDSD  390
      ||||| . |:| | | | | | | | | | | | | | | | | | |

```

2.4 蛋白质序列比对——NCBI Blast(global alignment)

NW Score	Identities	Positives	Gaps
1535	304/396(77%)	326/396(82%)	9/396(2%)
Query 1	M---EEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTED	57	
Sbjct 1	M EE QSD S+E PLSQETF S LWKLLP ++L P P MDDL+L P D+E++F	55	
Query 58	PGPDEAPRMPEAAAPPVAPAPAAPTPAAPAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSG	117	
Sbjct 56	-GPSEALRVSCAPAAQDPVTETPGPVAPAPATPWPLSSFVPSQKTYQGNVYGFHLGFLQSG	114	
Query 118	TAKSVICTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVRRCP	177	
Sbjct 115	TAKSV CTYSP LNK+FCQLAKTCPVQLWV +TPP G+RVRAMAIYK+SQHMTEVRRCP	174	
Query 178	HHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVWVWPYEPPEVGS DCTTIHYNM	237	
Sbjct 175	HHERCSD DGLAPPQHLIRVEGNL EYL+DR TFRHSVWVWPYEPPE GS+ TTIHY YM	234	
Query 238	CNSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRRRTEEENLRKKGEPHH	297	
Sbjct 235	CNSSCMGGMNRRPILTIITLEDSSGNLLGR+SFEVRVCACPGRRRTEEEN RKK	294	
Query 298	ELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGK	357	
Sbjct 295	ELPPGS KRALP TS+SP KKKPLDGEYFTL+IRGR+RFEMFRELNEALELKDA A +	354	
Query 358	EPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	393	
Sbjct 355	E G SRAHSS+LK+KKGQSTSRHKK M K GPDSD	390	



➤ 人和小鼠p53蛋白双序列BLAST比对点阵图

2.5 人P53蛋白理化性质分析

ProtParam 数据库

Number of amino acids:	393	Amino acid composition:	
Molecular weight:	43653.18	Ala (A)	24 6.1%
Theoretical pI:	6.33	Arg (R)	26 6.6%
Carbon	C 1898	Asn (N)	14 3.6%
Hydrogen	H 2980	Asp (D)	20 5.1%
Nitrogen	N 548	Cys (C)	10 2.5%
Oxygen	O 592	Gln (Q)	15 3.8%
Sulfur	S 22	Glu (E)	30 7.6%
Formula:	$C_{1898}H_{2980}N_{548}O_{592}S_{22}$	Gly (G)	23 5.9%
Total number of atoms:	6040	His (H)	12 3.1%
		Ile (I)	8 2.0%
		Leu (L)	32 8.1%
		Lys (K)	20 5.1%
		Met (M)	12 3.1%
		Phe (F)	11 2.8%
		Pro (P)	45 11.5%
		Ser (S)	38 9.7%
		Thr (T)	22 5.6%
		Trp (W)	4 1.0%
		Tyr (Y)	9 2.3%
		Val (V)	18 4.6%
		Pyl (O)	0 0.0%
		Sec (U)	0 0.0%
Total number of negatively charged residues (Asp + Glu):		50	
Total number of positively charged residues (Arg + Lys):		46	

2.5 人P53蛋白理化性质分析

➤ 极性分析

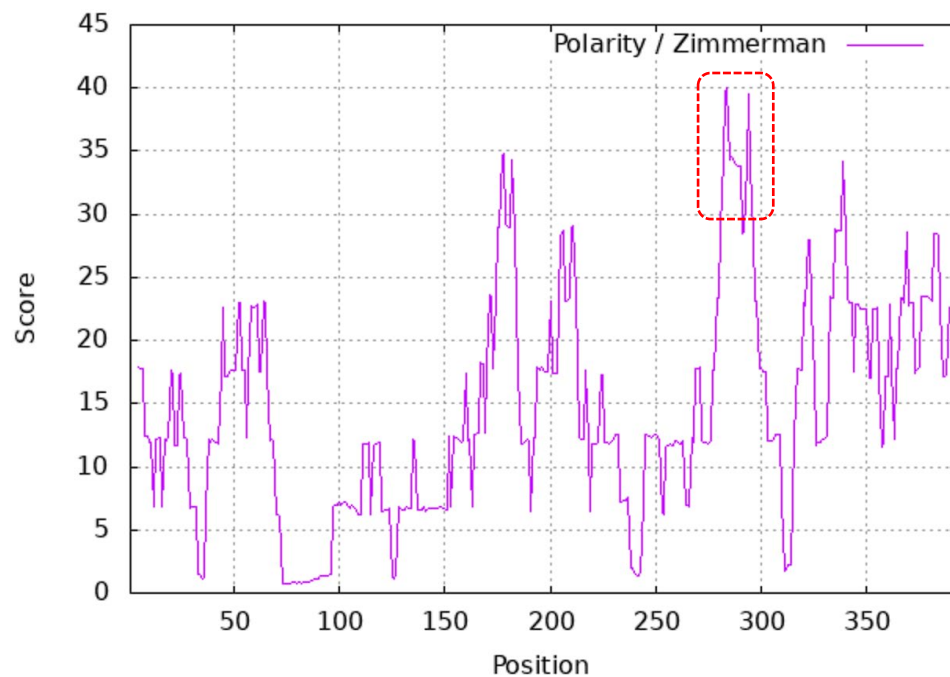
Using the scale **Polarity / Zimmerman**, the individual values for the 20 amino acids are:

Ala: 0.000 Arg: 52.000 Asn: 3.380 Asp: 49.700 Cys: 1.480 Gln: 3.530
 Glu: 49.900 Gly: 0.000 His: 51.600 Ile: 0.130 Leu: 0.130 Lys: 49.500
 Met: 1.430 Phe: 0.350 Pro: 1.580 Ser: 1.670 Thr: 1.660 Trp: 2.100
 Tyr: 1.610 Val: 0.130 : 26.540 : 26.715 : 13.594

Weights for window positions 1,...,9, using **linear weight variation model**:

1	2	3	4	5	6	7	8	9
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
edge				center				edge

ProtScale output for P53_HUMAN



➤ 亲疏水性分析

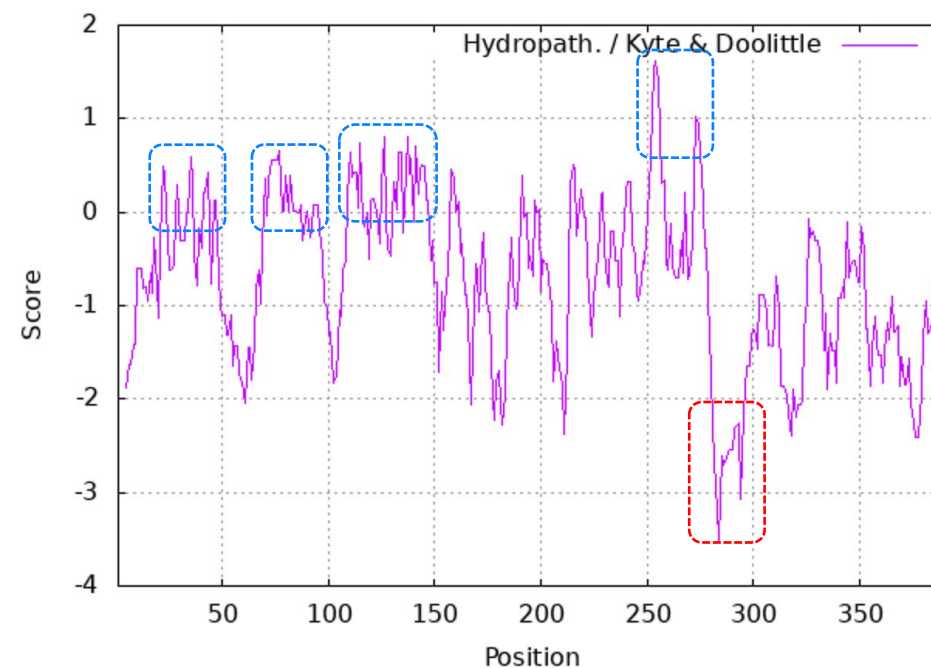
Using the scale **Hydropath. / Kyte & Doolittle**, the individual values for the 20 amino acids are:

Ala: 1.800 Arg: -4.500 Asn: -3.500 Asp: -3.500 Cys: 2.500 Gln: -3.500
 Glu: -3.500 Gly: -0.400 His: -3.200 Ile: 4.500 Leu: 3.800 Lys: -3.900
 Met: 1.900 Phe: 2.800 Pro: -1.600 Ser: -0.800 Thr: -0.700 Trp: -0.900
 Tyr: -1.300 Val: 4.200 : -3.500 : -3.500 : -0.490

Weights for window positions 1,...,9, using **linear weight variation model**:

1	2	3	4	5	6	7	8	9
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
edge				center				edge

ProtScale output for P53_HUMAN

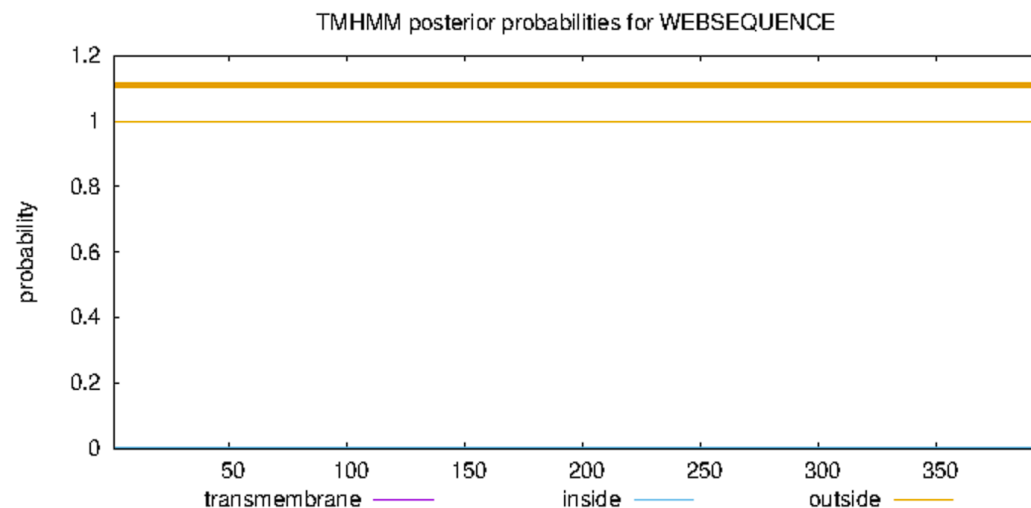


2.5 人P53蛋白理化性质分析

➤ 跨膜螺旋预测

TMHMM result

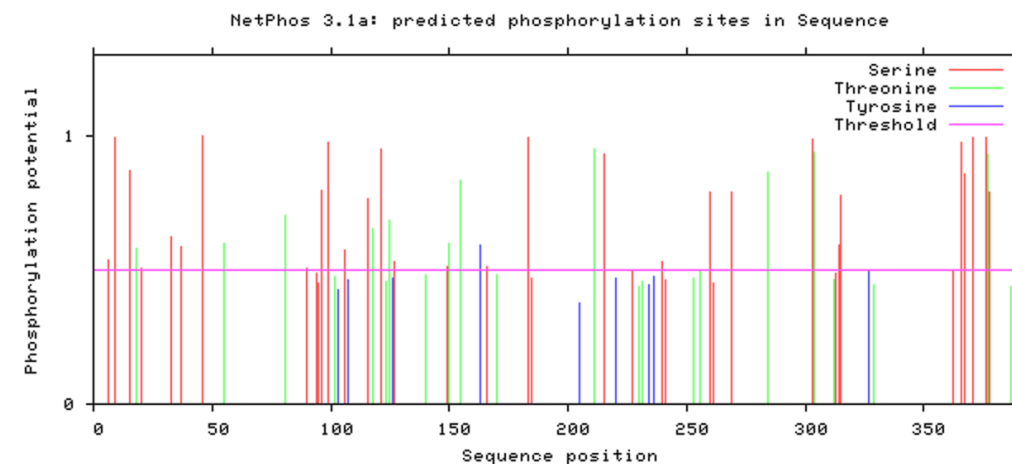
```
# WEBSEQUENCE Length: 393
# WEBSEQUENCE Number of predicted TMHs: 0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00506
# WEBSEQUENCE Exp number, first 60 AAs: 0
# WEBSEQUENCE Total prob of N-in: 0.00308
WEBSEQUENCE TMHMM2.0 outside 1 393
```



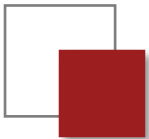
- 使用 **TMHMM** 预测蛋白质跨膜区和跨膜方向，结果如图所示，由图可知该蛋白无跨膜螺旋区域，所有氨基酸都分布在细胞膜外侧。

➤ 磷酸化位点预测

```
%1 .....S..S.....S..T.S.....S...S.....S... # 50
%1 ....T.....T.....S.....S..S. # 100
%1 .....S.....S.T..S...T.S.....ST # 150
%1 ....T.....Y..S.....S..... # 200
%1 .....T..S.....S..... # 250
%1 .....S.....S.....T..... # 300
%1 ..ST.....SS..... # 350
%1 .....SS...S...STS.....S.
```



- 使用 **NetPhos 3.1 Server** 预测蛋白中潜在的磷酸化位点，结果如图所示，其中洋红色的线表示阈值，超过此条线的位点氨基酸可能会发生磷酸化。



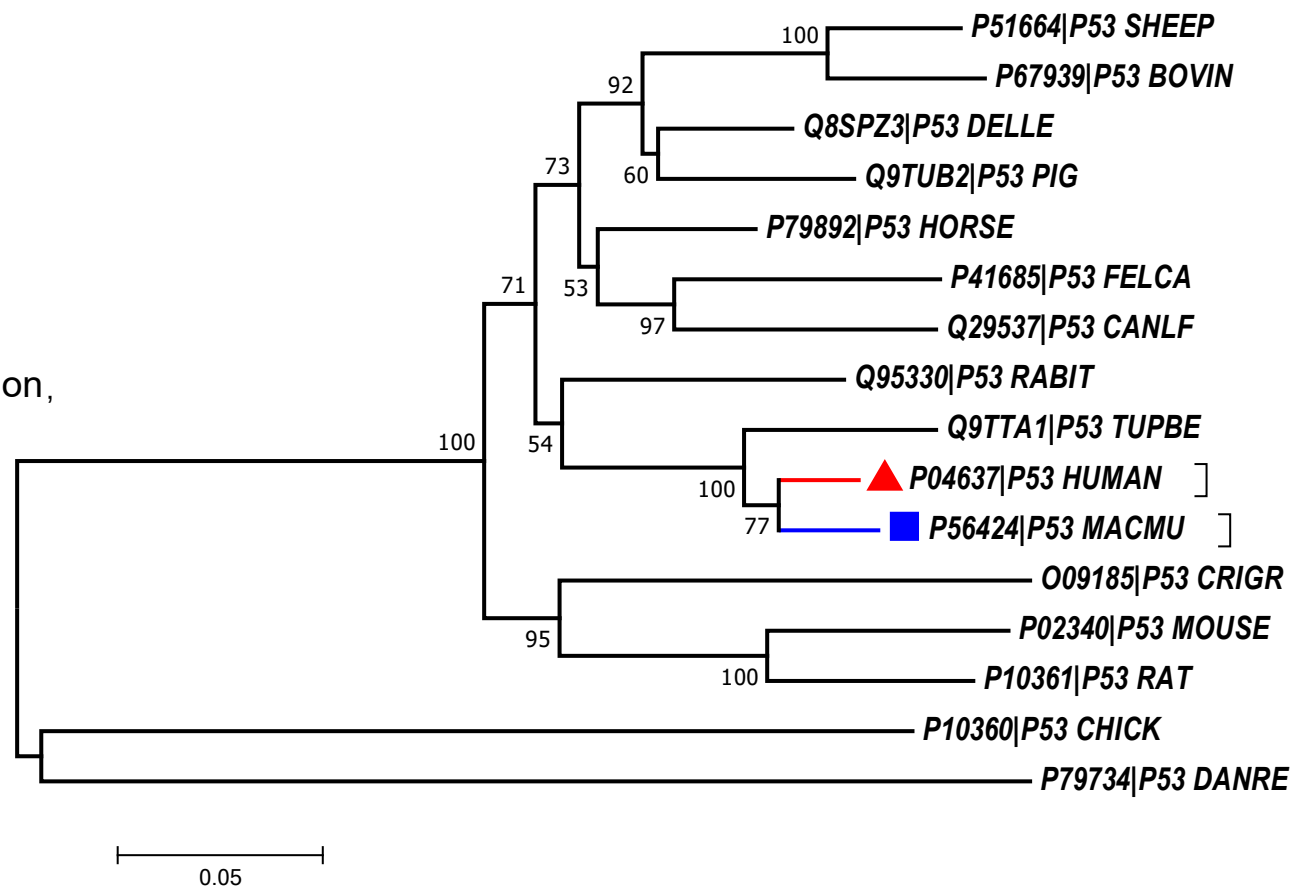
3. P53基因的系统进化树构建

登录号	物种	序列号	长度
P53_HUMAN	Homo sapiens (Human)	P04637	393
P53_MOUSE	Mus musculus (Mouse)	P02340	390
P53_RAT	Rattus norvegicus (Rat)	P10361	391
P53_MACMU	Macaca mulatta (Rhesus macaque)	P56424	393
P53_HORSE	Equus caballus (Horse)	P79892	280
P53_CANLF	Canis lupus familiaris (Dog)	Q29537	381
P53_PIG	Sus scrofa (Pig)	Q9TUB2	386
P53_BOVIN	Bos taurus (Bovine)	P67939	386
P53_CRIGR	Cricetulus griseus (Chinese hamster)	O09185	393
P53_DELLE	Delphinapterus leucas (Beluga whale)	Q8SPZ3	387
P53_TUPBE	Tupaia belangeri (Common tree shrew)	Q9TTA1	393
P53_DANRE	Danio rerio (Zebrafish)	P79734	373
P53_SHEEP	Ovis aries (Sheep)	P51664	382
P53_CHICK	Gallus gallus (Chicken)	P10360	367
P53_RABIT	Oryctolagus cuniculus (Rabbit)	Q95330	391
P53_FELCA	Felis catus (Cat)	P41685	386

3. P53基因的系统进化树构建

MEGA 7

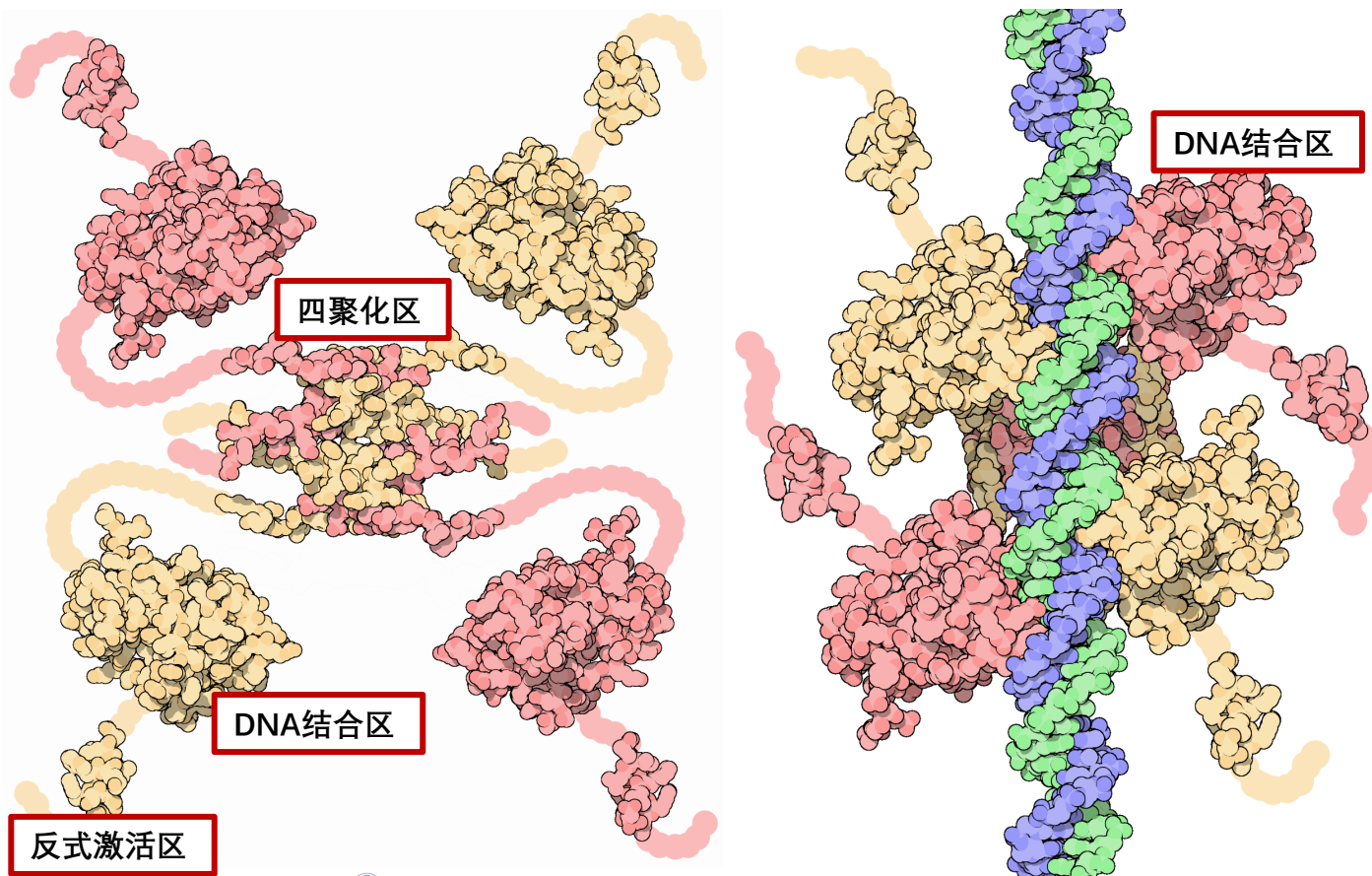
- 序列比对算法: Muscle
- 构建系统发生树: 邻接法
 - bootstrap设为1000,
 - Model设为p-distance,
 - Gaps采用Partial deletion,
 - cutoff设为50,
 - 其余采用默认参数



➤ 由构建的进化树不难看出，13个物种来源的p53蛋白起源于同一个祖先分子，且同属于一个有相似功能的蛋白家族。另外可以看出，人和猕猴在进化关系上非常接近，两者的p53蛋白基因与共同祖先的距离最近；而鸡和斑马鱼的p53蛋白基因与其他物种都有较大的差异，这可能是由于在进化过程中它们的功能发生了重大变化。这也提示我们鸡，斑马鱼这类生物可能不具备与人类相同的、由进化发展出的p53调控网络，在研究人p53蛋白时不适合使用鸡和斑马鱼等物种作为模式生物，使用猕猴可能较为适合。

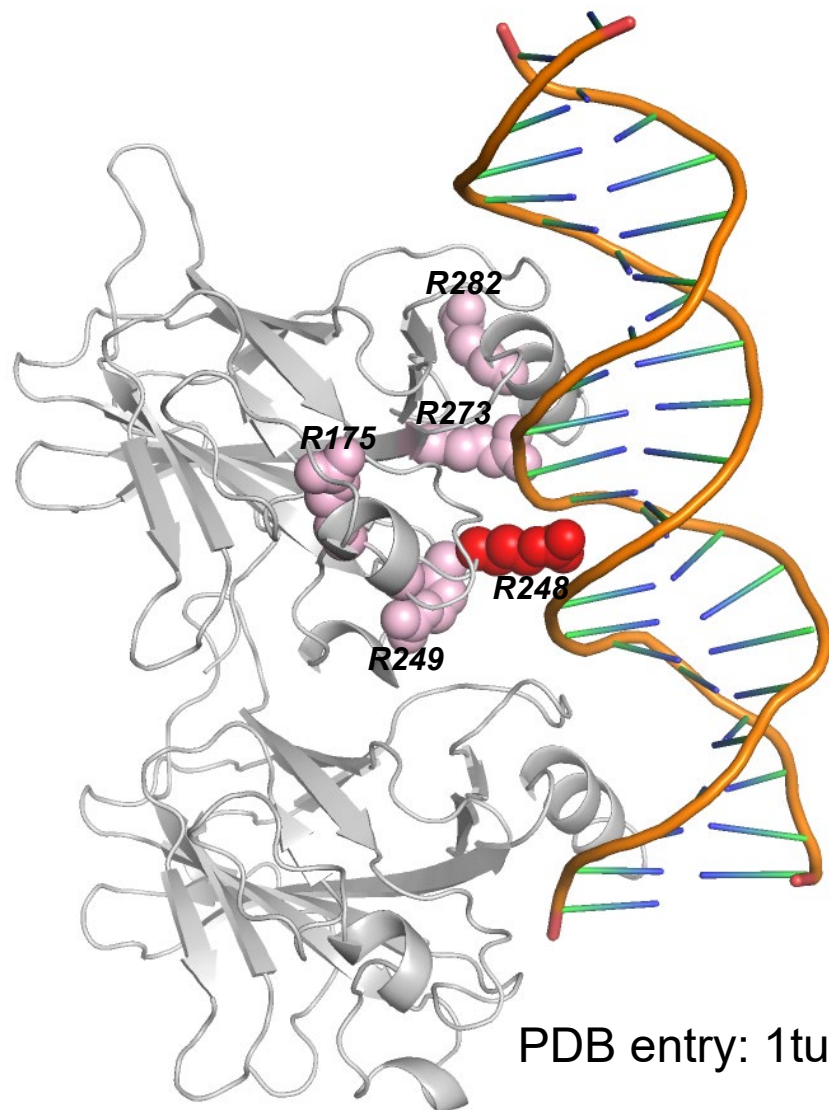
4.1 人P53蛋白的结构特征

- P53蛋白是由四个相同的蛋白质链组成的柔性分子。
- 整个柔性分子可分成三个紧凑的球状“域”和无规卷曲链区域。
- P53的中心是一个四聚化结构域 (PDB: 1olg), 它将四个链连接在一起。
- 每条链中的一个长的柔性区域连接到第二个稳定域: 一个大的 DNA 结合域 (PDB: 1tup), 它富含与 DNA 相互作用的精氨酸。该域识别 DNA 上的特定调节位点。
- 第三个稳定结构域是反式激活结构域 (PDB: 1ycq), 位于每条手臂末端附近, 可激活 DNA 读取机制。



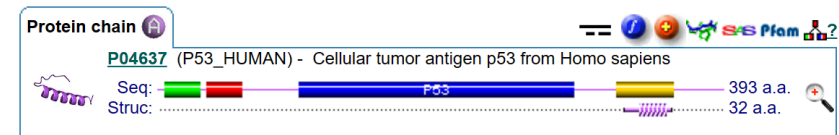
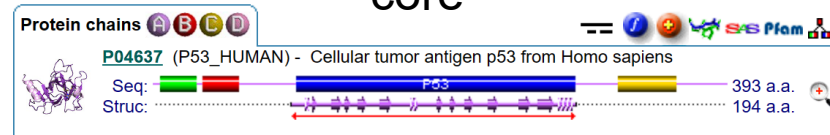
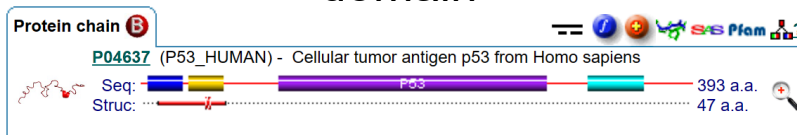
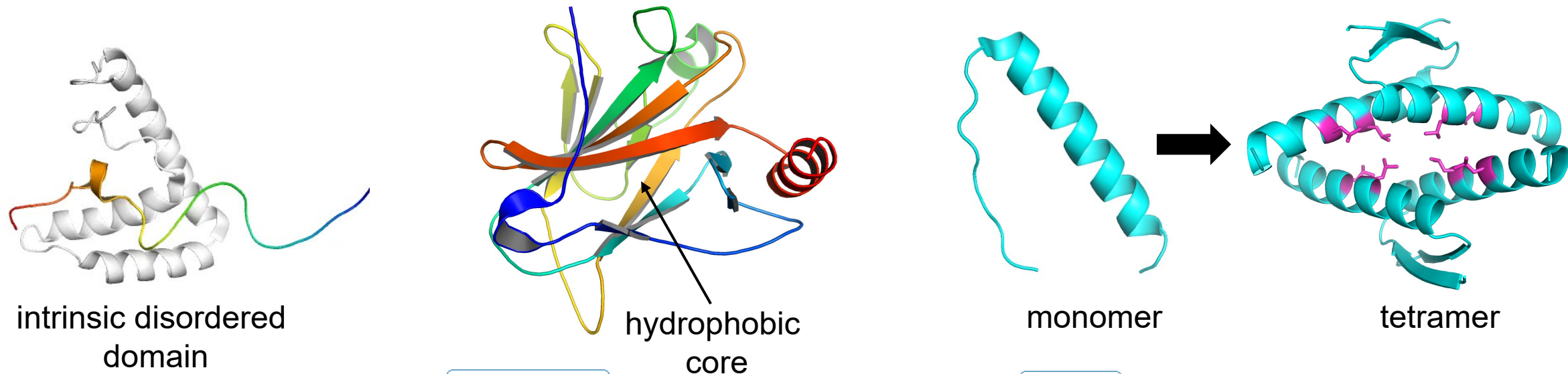
4.1 人P53蛋白的结构特征

- 大多数导致癌症的 p53 突变都存在于 DNA 结合域中。此处使用 PDB 条目 1tup 显示了最常见的突变。
- 这些突变存在于蛋白质的 DNA 结合表面及其周围。最常见的突变是 **R248**（红色球型）。它插入到DNA的小沟，与周围残基形成稳定的相互作用。当突变为另一种氨基酸时，这种相互作用就会消失。其他关键突变位点以粉红色显示，包括 **R175**、**R249**、**R273** 和 **R282**。
- 还有一些精氨酸直接与 DNA 接触，另一些则参与定位其他 DNA 结合氨基酸。



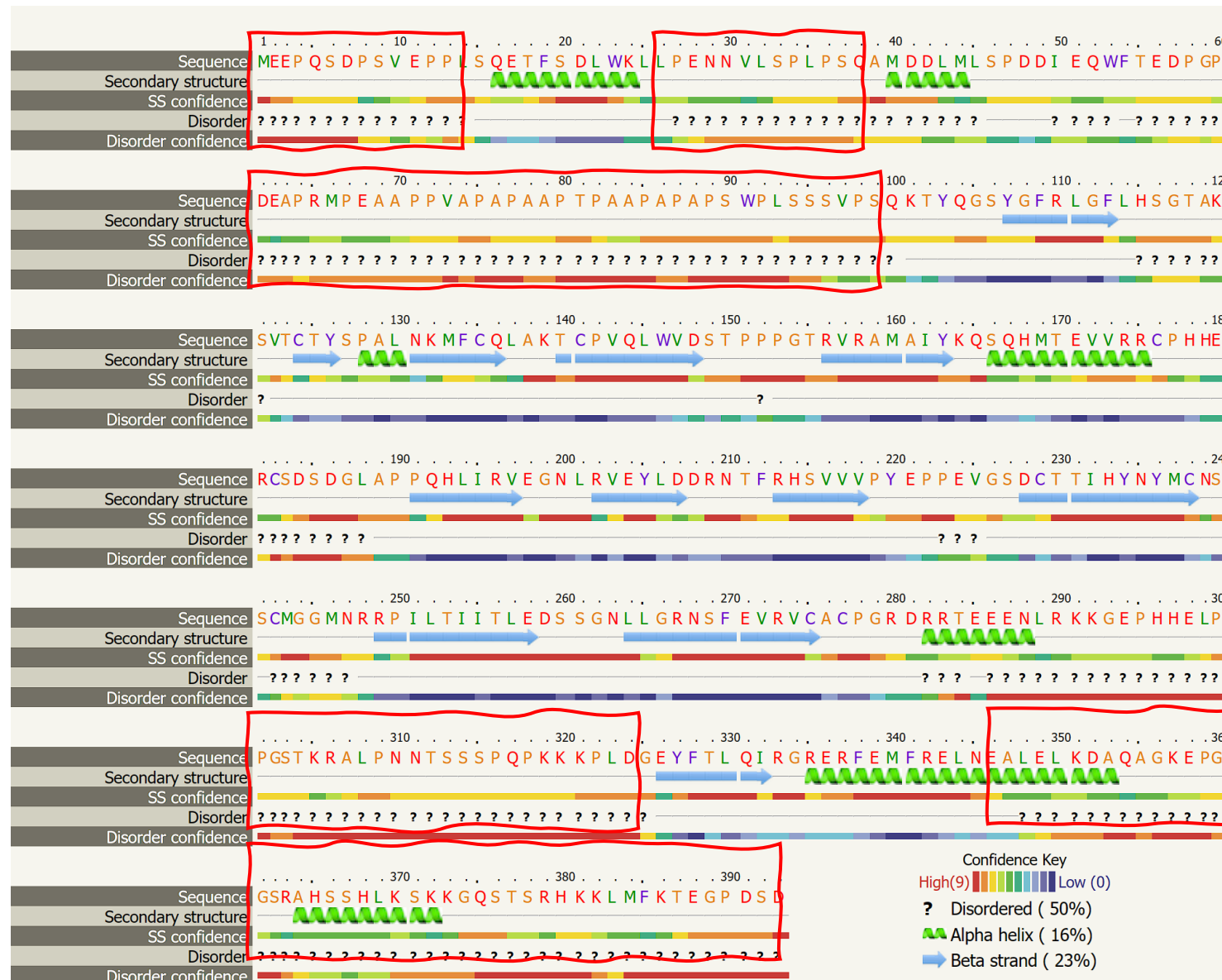
4.1 人P53蛋白的结构特征

- P53蛋白N端部分包含1-93位的氨基酸，生理状态下呈未折叠态。DNA结合区域包含94-302位的氨基酸，与特定DNA结合，是整个蛋白质的核心区域。
- 四聚化结构域包含325-356位的氨基酸，由一个短的 β 折叠(326-333)和一个 α 螺旋(335-354)组成，中间由一个保守的甘氨酸连接。两个单体通过反平行的分子间 β 折叠和 α 螺旋堆积，形成原始二聚体。
- 两个二聚体通过平行螺旋相互作用形成四聚体，四聚体作用界面主要通过Leu344和Leu348的疏水作用稳定。



4.2 人P53蛋白的二级结构预测

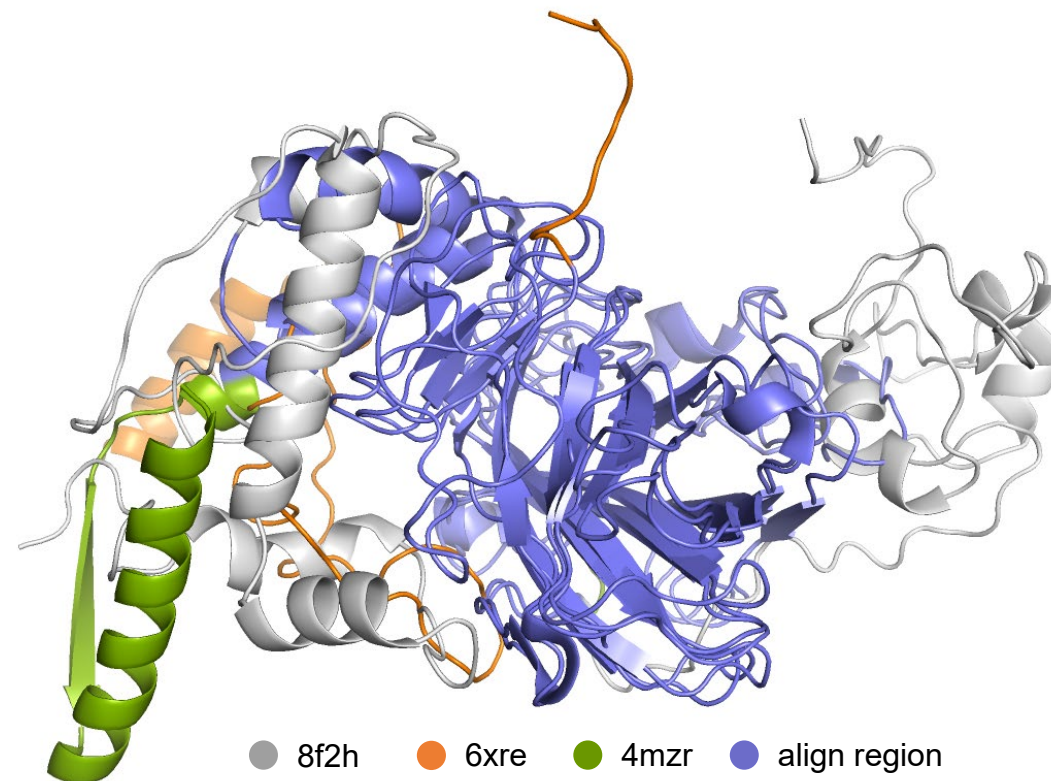
- Phyre²预测出的P53蛋白N端和C端属于内在无序区。
- 预测出的 α 螺旋， β 折叠和无序区比例分别为16%，23%，50%，整个分子非常柔性。



Swiss-model

Template	Sequence Identity	Description	Length
8f2h	100.00	Wild type P53 dimer structure from human cancer cells	1-393
6xre	100.00	Structure of the p53/RNA polymerase II assembly	1-393
4mzr	90.72	Crystal structure of a polypeptide p53 mutant bound to DNA	92-357
3q05	92.70	An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity	92-357

- P53蛋白的柔性N端和C端与不同蛋白的结合构象存在较大差异，这有利于其结合不同核酸或蛋白，从而发挥功能的多样性。



Aligned residues: 97-292, 325-355

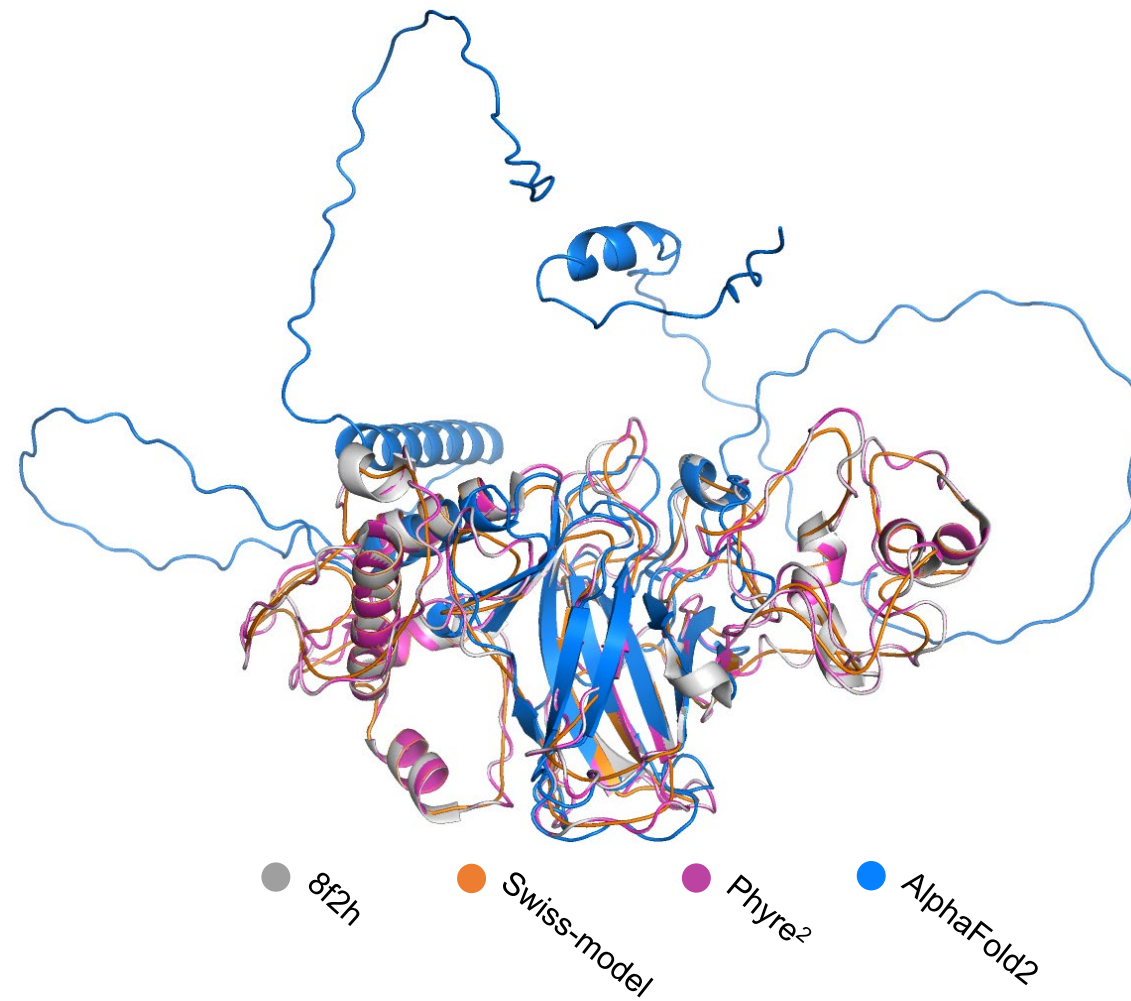
RMSD: 8f2h vs 4mzr = 1.6 Å

RMSD: 6xre vs 4mzr = 3.2 Å

4.3 人P53蛋白的三级结构预测

Predicted structures

- Aligned residues: 97-292, 325-355
- Swiss-model和Phyre² 都基于8f2h模版建模，预测出的结构与8f2h非常接近，Ca RMSD分别为0.5 Å、0.1 Å，但低估了p53蛋白的柔性。
- AlphaFold2 对 p53 蛋白的稳定结合域能预测出高置信度结构，与 8f2h 的Ca RMSD 为 1.78 Å，同时能区分出内在无序区。



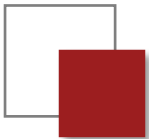
5. 总结与感悟

总结

- P53 在细胞生长周期中扮演负调节因子的角色，它是很重要的抑癌基因，和细胞周期的调控、DNA 修复、细胞分化、细胞凋亡等重要的生物功能有关。
- 与人p53基因蛋白质序列最相似的物种是黑猩猩，其次是人与猕猴，推测是因为人是由灵长类动物进化而来所有保留高度的相似性。
- P53蛋白有一个较强的亲水区域，位于 280~290 aa，同时有两个较强的疏水区域，位于 250~255 aa 和 265~275 aa，蛋白无信号肽，没有发现跨膜结构域。
- P53蛋白约16%为 α 螺旋，23%为 β 折叠，50%为无规则卷曲，含有大量天然未折叠区域。

课程感悟

- 生物信息学学科非常交叉，需要对数学，物理，化学，生物都有一定基础。
- 生信分析工具众多，对于同一个问题有许多工具可以使用，但学习难度不高。
- 熟能生巧，针对自己的课题找到合适的工具，勤加练习，熟练掌握。



5. 致谢 Acknowledgements



北京大学
PEKING UNIVERSITY

感谢罗老师的耐心教学和指导！

感谢G09小组成员的讨论和帮助！

感谢大家的聆听，请大家批评指正！