The bioinfomatic analysis of human toposoimerase 3β

G06 group
汇报人：李雯
小组成员：刘玉祥、邢孟坦、高胜贤、李雯
2014.01.10

Outline

- Background
- Research purpose
- Research contents
- Results and discussion
- Acknowledgement

Background

- DNA supercoil

  It refers to the over- or under-winding of a DNA strand, is an expression of the strain on that strand. It is important in a number of biological processes.

- DNA topology

  Three main types of topology: supercoiling, knotting and catenation.

- DNA toposoimerase

  Topoisomerases are “magicians of the DNA world,” working their ‘wizardry’ to solve topological problems in DNA metabolism.

Type I Type II topoisomerase

Background

- DNA topoisomerase IIIa

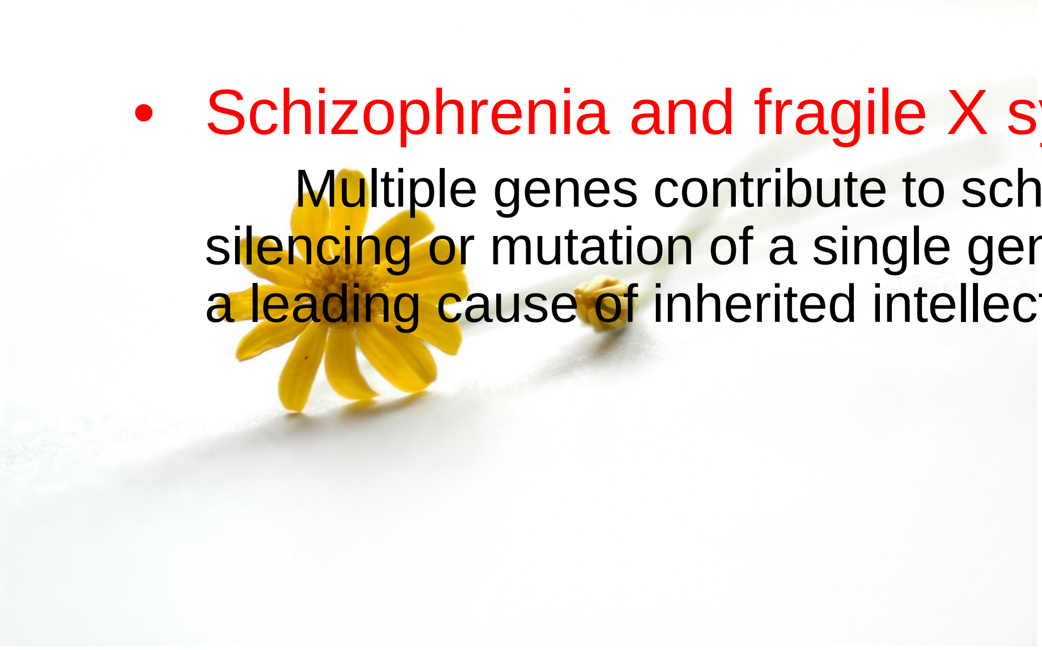
  It is part of the DNA ‘dissolvasome’ that resolves intermediates generated during repair and recombination

- DNA topoisomerase IIIb

  It's a type I topoisomerase which binds multiple mRNAs encoded by genes with neuronal functions linked to schizophrenia and autism.

- Schizophrenia and fragile X syndrome (FXS):

  Multiple genes contribute to schizophrenia, inappropriate silencing or mutation of a single gene, FMR1, causes FXS, which is a leading cause of inherited intellectual disability and autism.



	DNA	RNA
Polymerase	√	√
Nuclease	√	√
Ligase	√	√
Helicase	√	√
Topoisomerase	√	**?**

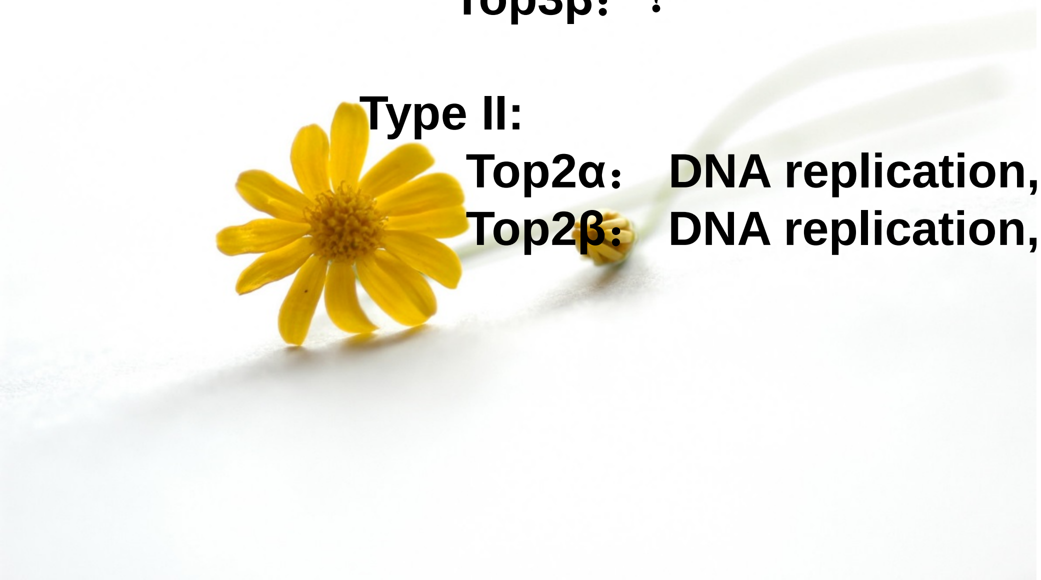
In eukaryotes:

Type I:

- **Topo I：trancription, DNA replication**
- **Top3α：DNA damage repair，chromosome segregation**
- **Top3β：？**

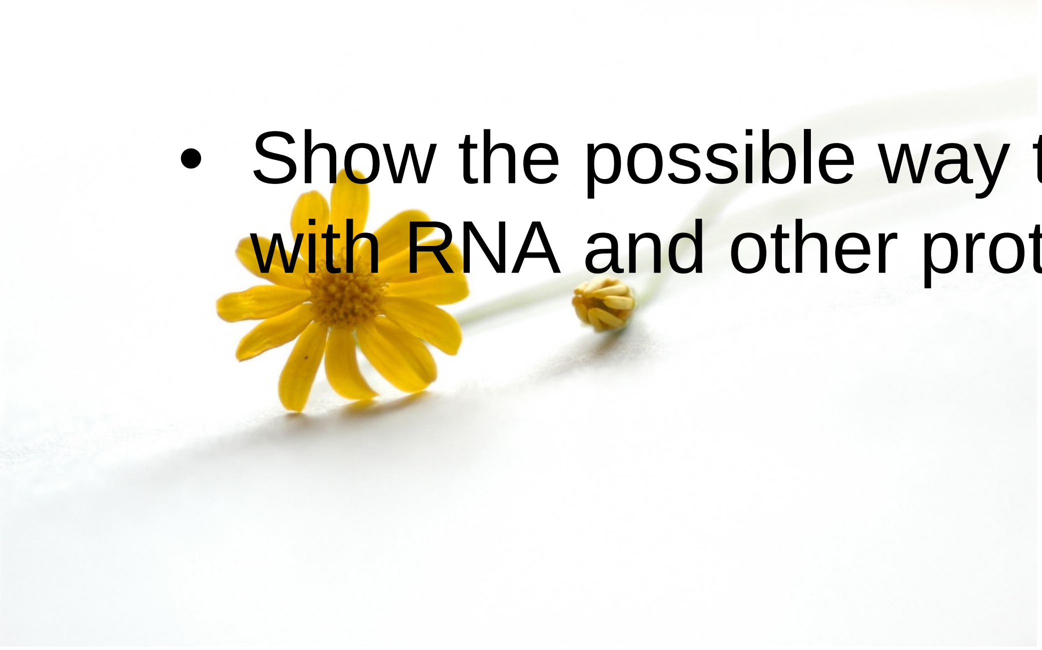
Type II:

- **Top2α：DNA replication, chromosome segregation**
- **Top2β：DNA replication, chromosome segregation**



Research purpose

- Preliminary analysis the evolution, gene structure and protein struction of top3b
- Know the possible function and improtant domains of top3b
- Show the possible way that how top3b inteact with RNA and other proteins



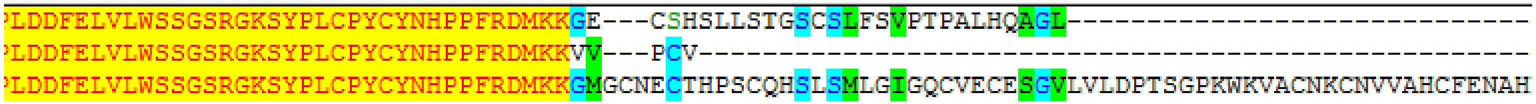
Research contents

1. 数据收集与特征分析
2. 序列比对与基因结构分析
3. 蛋白基本信息及氨基酸基本性质分析
4. 蛋白质二级结构及结构域预测
5. 蛋白质序列三级结构预测/同源建模
6. 结构特性分析及功能关系

1.数据收集与特征分析

UniProt › UniProtKB

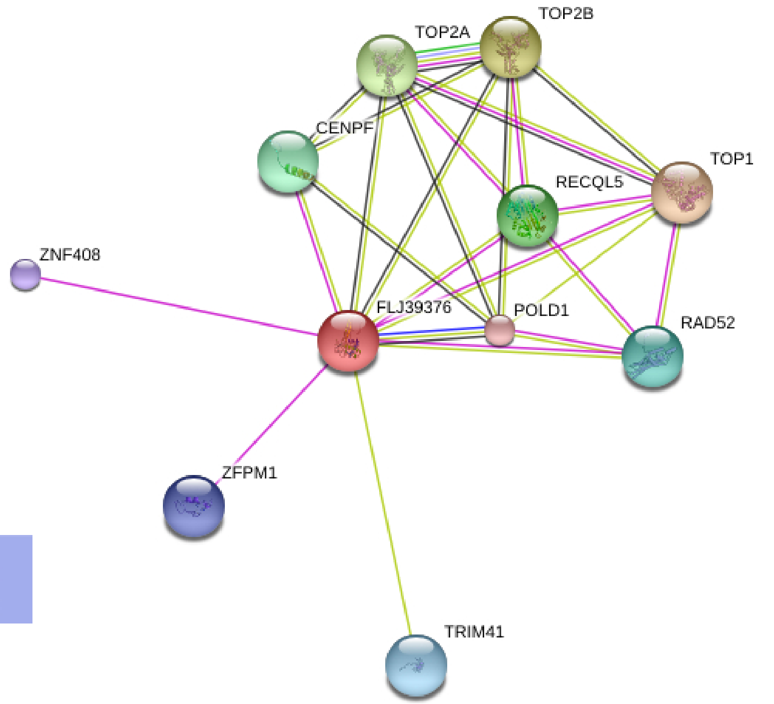
- 862AA,   CDS:2589bp
- type IA topoisomerase family
- 3 isoforms, C terminal different



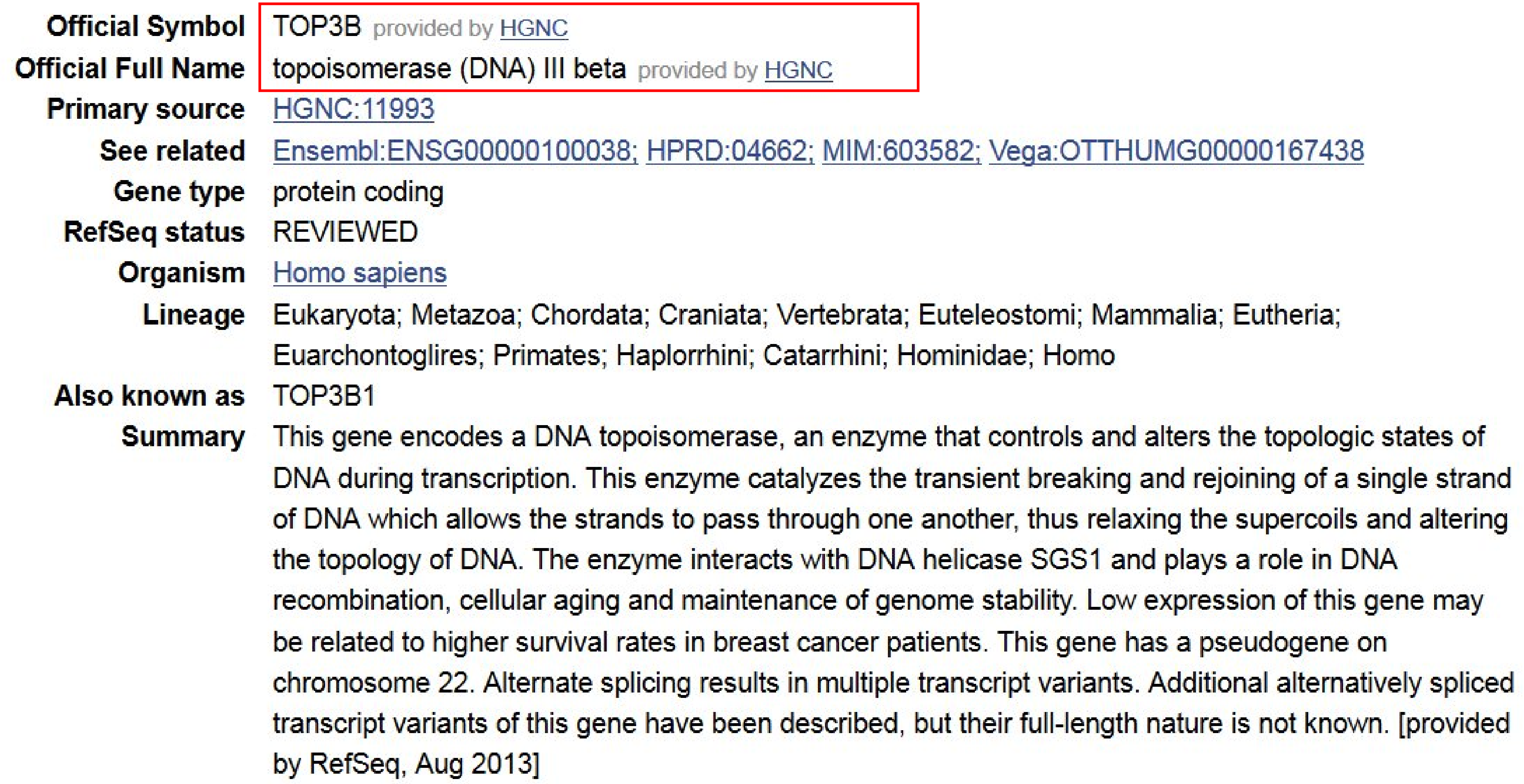
```
LDDFELVLWSSGSRGKSYPLCPYCYNHPPFRDMKKGE---CSHSLLSTGSCSLFSVPTPALHQAGL-----------------------------
LDDFELVLWSSGSRGKSYPLCPYCYNHPPFRDMKKVV---PCV----------------------------------------------------
LDDFELVLWSSGSRGKSYPLCPYCYNHPPFRDMKKGMGCNECTHPSCQHSLSMLGIGQCVECESGVLVLDPTSGPKWKVACNKCNVVAHCFENAH
```

- Interaction

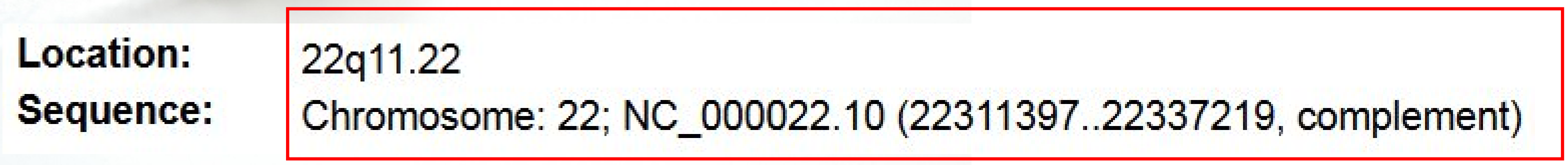
STRING 9.1



2.序列比对与基因结构分析

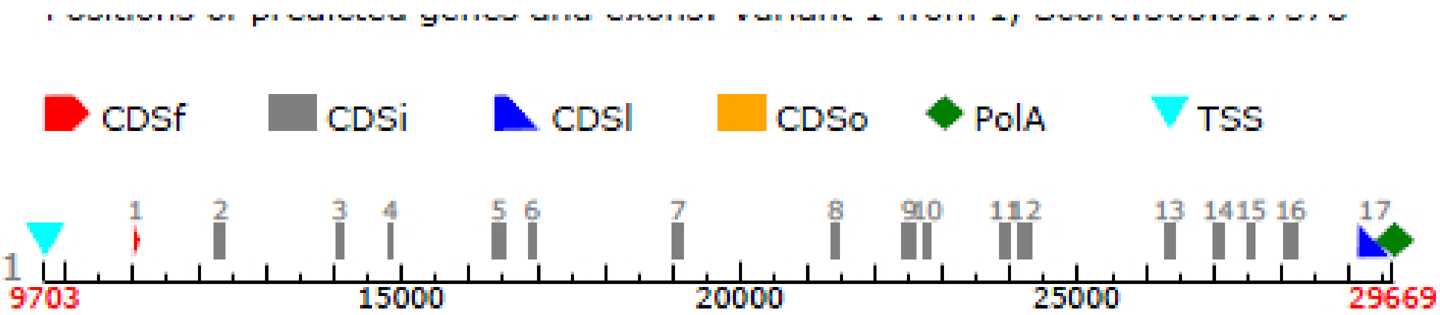


Official Symbol	TOP3B provided by HGNC
Official Full Name	topoisomerase (DNA) III beta provided by HGNC
Primary source	HGNC:11993
See related	Ensembl:ENSG00000100038; HPRD:04662; MIM:603582; Vega:OTTHUMG00000167438
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	TOP3B1
Summary	This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This enzyme catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus relaxing the supercoils and altering the topology of DNA. The enzyme interacts with DNA helicase SGS1 and plays a role in DNA recombination, cellular aging and maintenance of genome stability. Low expression of this gene may be related to higher survival rates in breast cancer patients. This gene has a pseudogene on chromosome 22. Alternate splicing results in multiple transcript variants. Additional alternatively spliced transcript variants of this gene have been described, but their full-length nature is not known. [provided by RefSeq, Aug 2013]

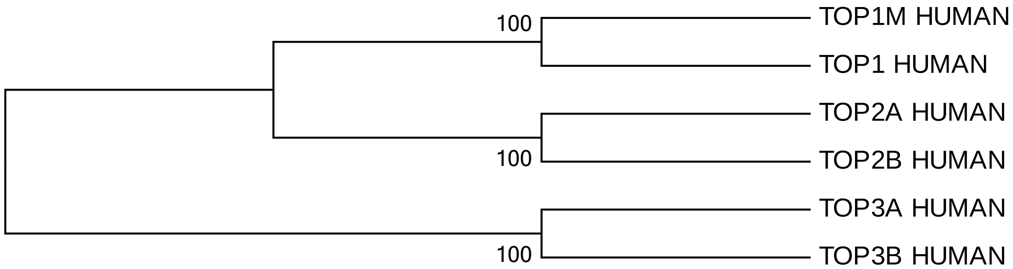


Location:	22q11.22
Sequence:	Chromosome: 22; NC_000022.10 (22311397..22337219, complement)

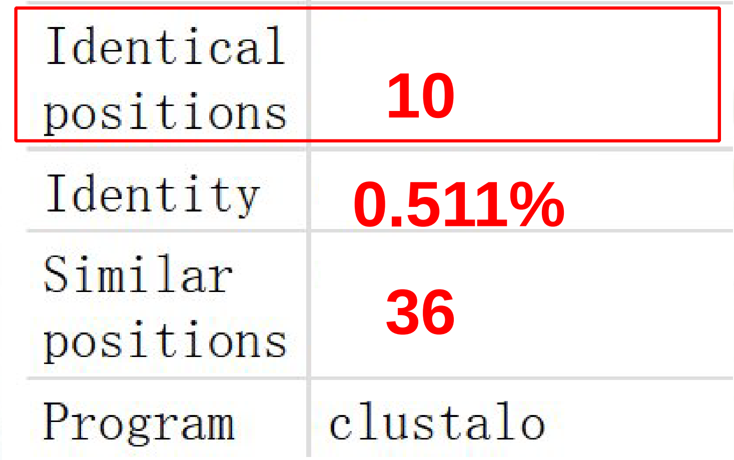
在NCBI中找到“Top3beta”的全长DNA序列，用softberry 进行基因结构预测，预测出17个外显子：



人类拓扑异构酶蛋白质家族序列比对



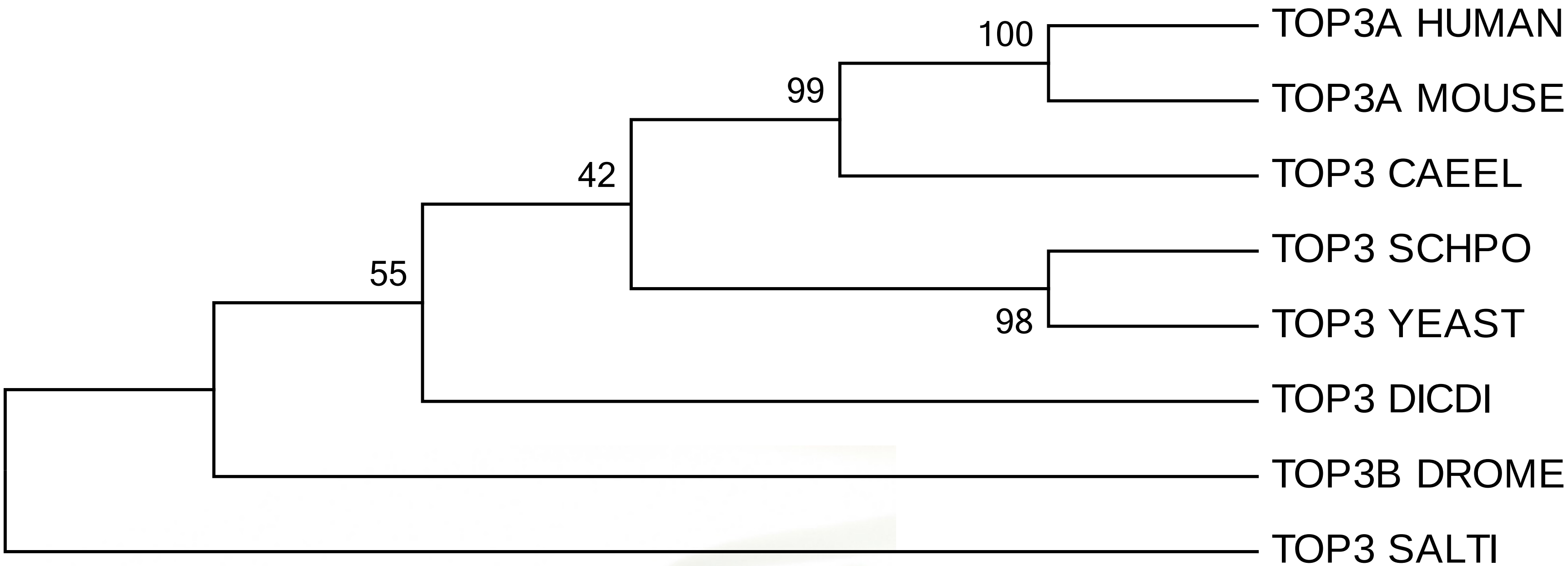
Entry	Entry name	Length
Q969P6	TOP1M_HUMAN	601
P11387	TOP1_HUMAN	765
P11388	TOP2A_HUMAN	1531
Q02880	TOP2B_HUMAN	1626
Q13472	TOP3A_HUMAN	1001
O95985	TOP3B_HUMAN	862



Identical positions	**10**
Identity	**0.511%**
Similar positions	**36**
Program	clustalo

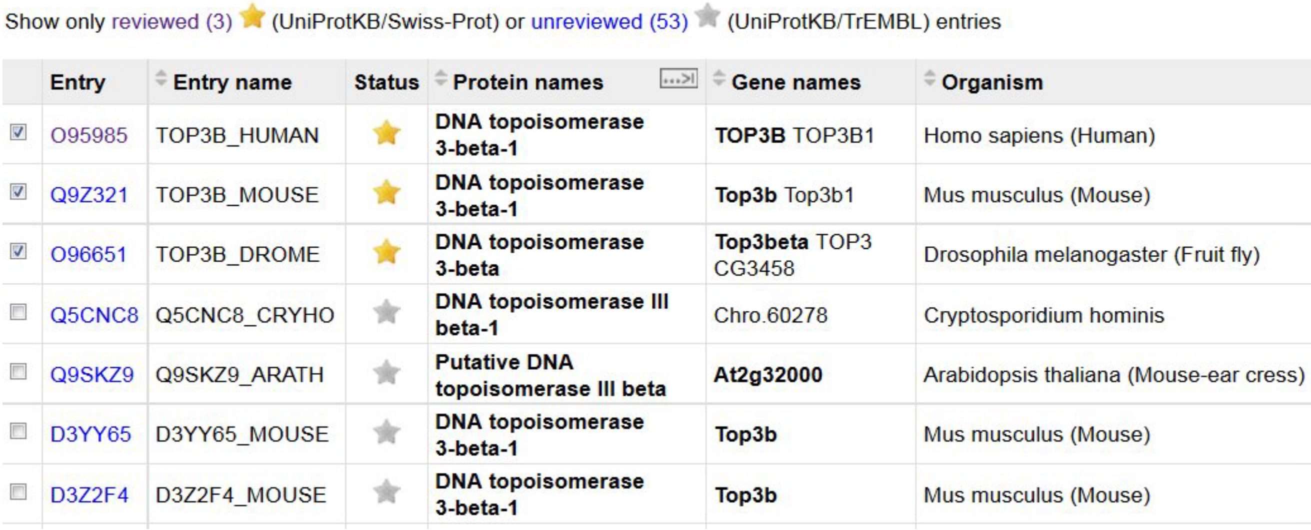
旁系同源基因：多序列比对相似性不高。

拓扑异构酶3蛋白质（TOP3）系统发育树构建



直系同源基因：分化较早

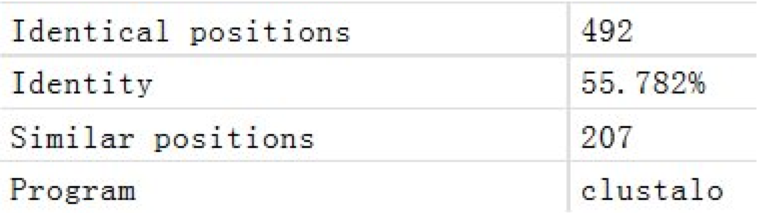
Top3-beta蛋白质多序列比对



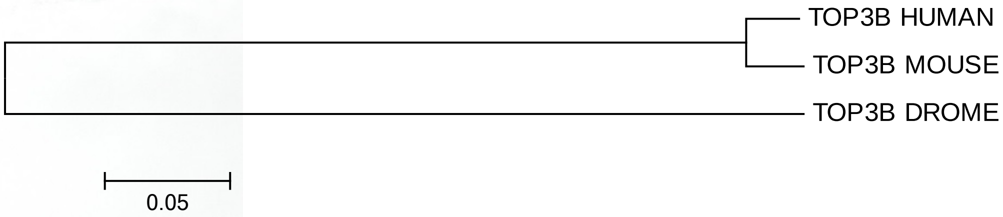
Show only reviewed (3) ★ (UniProtKB/Swiss-Prot) or unreviewed (53) ☆ (UniProtKB/TrEMBL) entries

	Entry	Entry name	Status	Protein names	Gene names	Organism
☑	O95985	TOP3B_HUMAN	★	**DNA topoisomerase 3-beta-1**	**TOP3B** TOP3B1	Homo sapiens (Human)
☑	Q9Z321	TOP3B_MOUSE	★	**DNA topoisomerase 3-beta-1**	**Top3b** Top3b1	Mus musculus (Mouse)
☑	O96651	TOP3B_DROME	★	**DNA topoisomerase 3-beta**	**Top3beta** TOP3 CG3458	Drosophila melanogaster (Fruit fly)
☐	Q5CNC8	Q5CNC8_CRYHO	☆	**DNA topoisomerase III beta-1**	Chro.60278	Cryptosporidium hominis
☐	Q9SKZ9	Q9SKZ9_ARATH	☆	**Putative DNA topoisomerase III beta**	**At2g32000**	Arabidopsis thaliana (Mouse-ear cress)
☐	D3YY65	D3YY65_MOUSE	☆	**DNA topoisomerase 3-beta-1**	**Top3b**	Mus musculus (Mouse)
☐	D3Z2F4	D3Z2F4_MOUSE	☆	**DNA topoisomerase 3-beta-1**	**Top3b**	Mus musculus (Mouse)

最大似然法建树（N-J法建树至少需要四条序列）

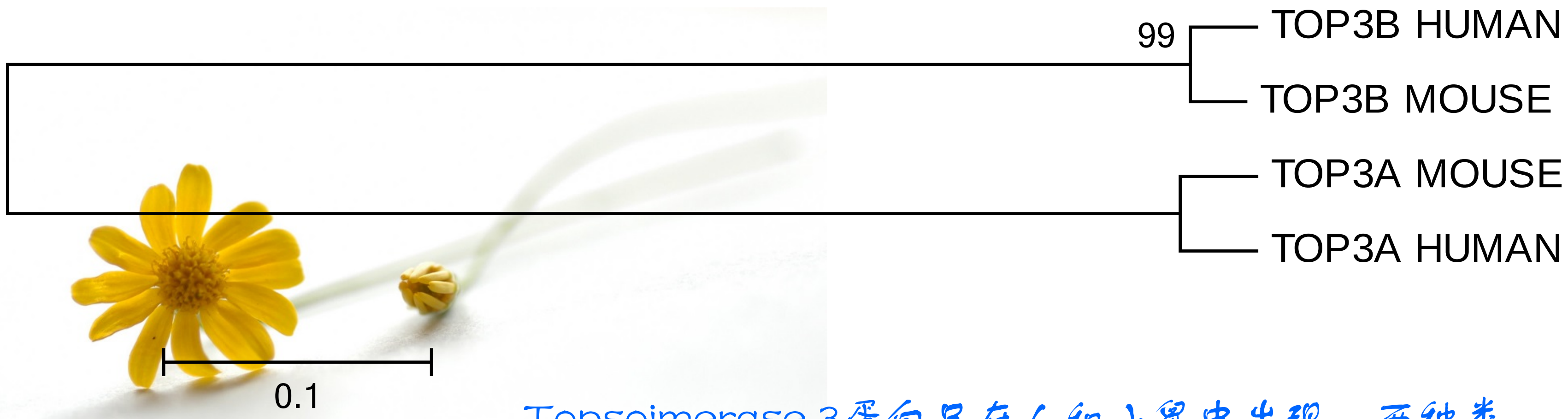


Identical positions	492
Identity	55.782%
Similar positions	207
Program	clustalo



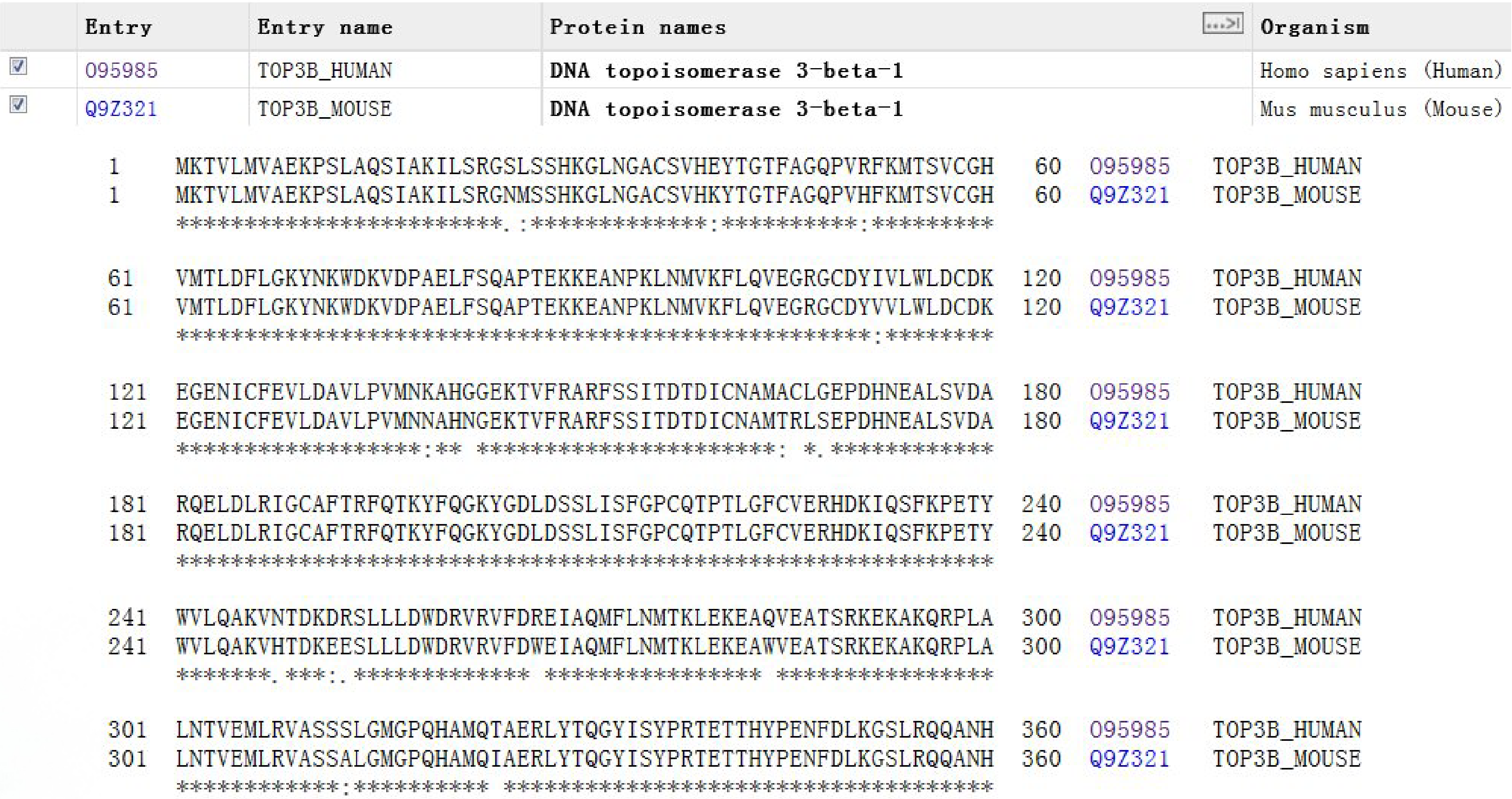
Top3-beta蛋白质NCBI blast结果

搜索条件organism: mammals (taxid:40674)； algorithm:PSI-BLAST; max target sequence=500, expect threshold=0.01, word size=2, matrix: BLOSUM 62, gap cost: existence=11 extension=1, PSI-BLAST Threshold 0.005, Pseudocount 0。导出cDNA序列及FASTA格式的文件，在Mega中建树，结果如下：



Topsoimerase 3蛋白只在人和小鼠中出现，两种类型a和b早很早之前分化，两蛋白序列相似但功能差别较大

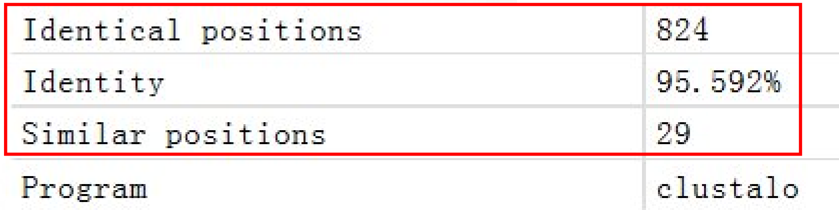
人和小鼠的Top3-beta蛋白质序列比对结果



	Entry	Entry name	Protein names	Organism
☑	O95985	TOP3B_HUMAN	**DNA topoisomerase 3-beta-1**	Homo sapiens (Human)
☑	Q9Z321	TOP3B_MOUSE	**DNA topoisomerase 3-beta-1**	Mus musculus (Mouse)

```
1     MKTVLMVAEKPSLAQSIAKILSRGSLSSHKGLNGACSVHEYTGTFAGQPVRFKMTSVCGH   60   O95985   TOP3B_HUMAN
1     MKTVLMVAEKPSLAQSIAKILSRGNMSSHKGLNGACSVHKYTGTFAGQPVHFKMTSVCGH   60   Q9Z321   TOP3B_MOUSE
      ************************.:*************:**********:*********

61    VMTLDFLGKYNKWDKVDPAELFSQAPTEKKEANPKLNMVKFLQVEGRGCDYIVLWLDCDK  120   O95985   TOP3B_HUMAN
61    VMTLDFLGKYNKWDKVDPAELFSQAPTEKKEANPKLNMVKFLQVEGRGCDYVVLWLDCDK  120   Q9Z321   TOP3B_MOUSE
      ***************************************************:********

121   EGENICFEVLDAVLPVMNKAHGGEKTVFRARFSSITDTDICNAMACLGEPDHNEALSVDA  180   O95985   TOP3B_HUMAN
121   EGENICFEVLDAVLPVMNNAHNGEKTVFRARFSSITDTDICNAMTRLSEPDHNEALSVDA  180   Q9Z321   TOP3B_MOUSE
      ******************:** **********************: *.************

181   RQELDLRIGCAFTRFQTKYFQGKYGDLDSSLISFGPCQTPTLGFCVERHDKIQSFKPETY  240   O95985   TOP3B_HUMAN
181   RQELDLRIGCAFTRFQTKYFQGKYGDLDSSLISFGPCQTPTLGFCVERHDKIQSFKPETY  240   Q9Z321   TOP3B_MOUSE
      ************************************************************

241   WVLQAKVNTDKDRSLLLDWDRVRVFDREIAQMFLNMTKLEKEAQVEATSRKEKAKQRPLA  300   O95985   TOP3B_HUMAN
241   WVLQAKVHTDKEESLLLDWDRVRVFDWEIAQMFLNMTKLEKEAWVEATSRKEKAKQRPLA  300   Q9Z321   TOP3B_MOUSE
      *******.***:.************* **************** ****************

301   LNTVEMLRVASSSLGMGPQHAMQTAERLYTQGYISYPRTETTHYPENFDLKGSLRQQANH  360   O95985   TOP3B_HUMAN
301   LNTVEMLRVASSALGMGPQHAMQIAERLYTQGYISYPRTETTHYPENFDLKGSLRQQANH  360   Q9Z321   TOP3B_MOUSE
      ************:********** ************************************
```

Length:862aa

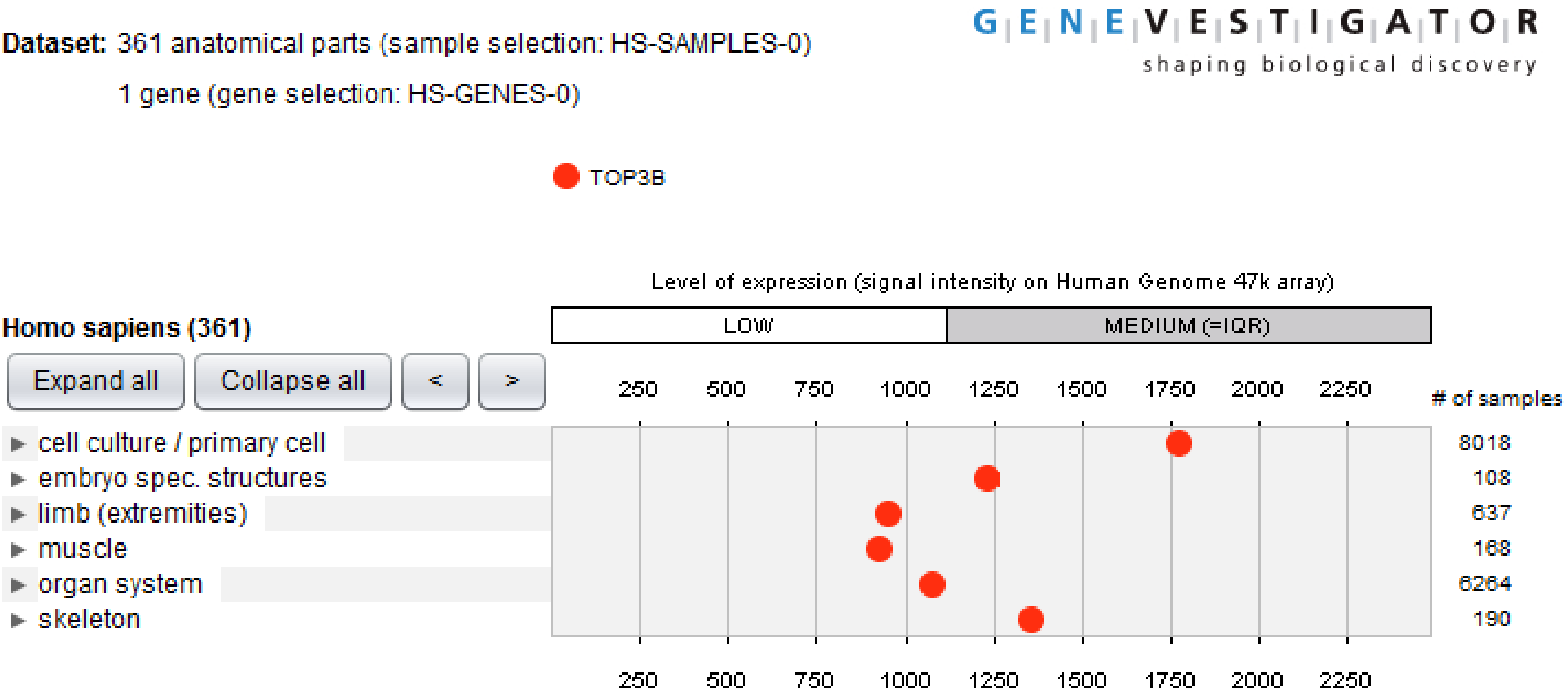


Identical positions	824
Identity	95.592%
Similar positions	29
Program	clustalo

结论：Top3-beta蛋白序列在人和小鼠中高度保守。

3.蛋白基本信息及氨基酸基本性质分析

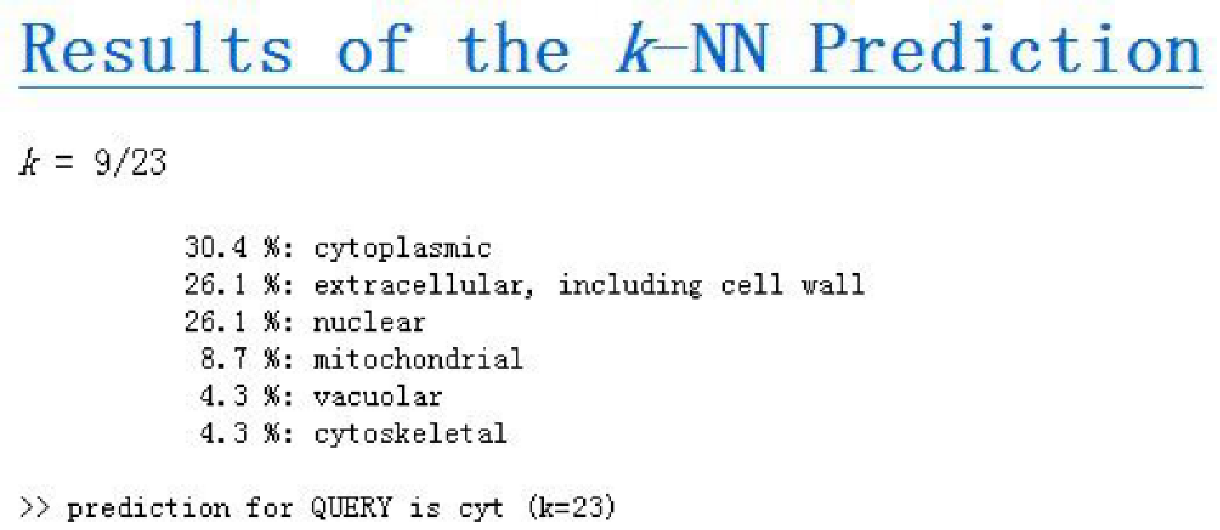
- Top3β在不同组织，细胞中的表达——Genevestigator



Dataset: 361 anatomical parts (sample selection: HS-SAMPLES-0)

1 gene (gene selection: HS-GENES-0)

- 亚细胞定位——psort



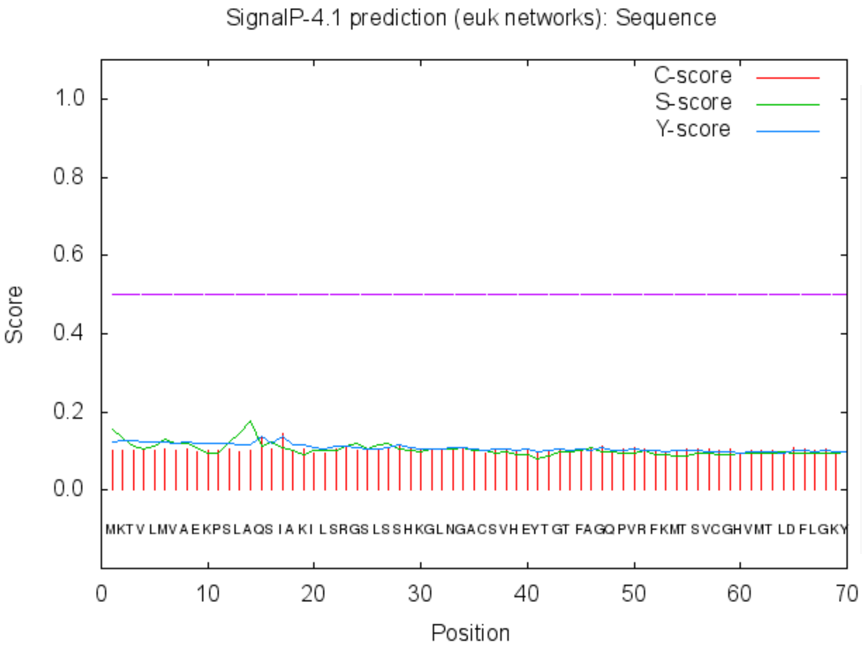
Results of the *k*-NN Prediction

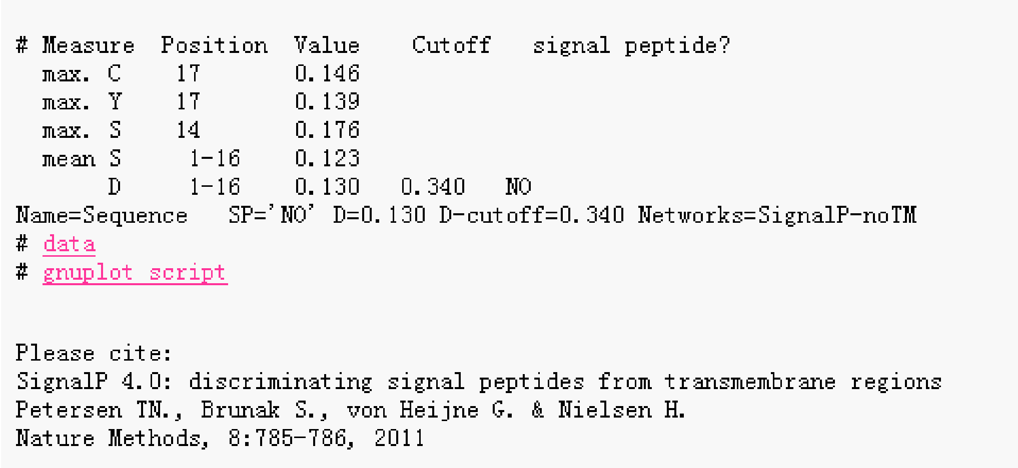
```
k = 9/23

        30.4 %: cytoplasmic
        26.1 %: extracellular, including cell wall
        26.1 %: nuclear
         8.7 %: mitochondrial
         4.3 %: vacuolar
         4.3 %: cytoskeletal

>> prediction for QUERY is cyt (k=23)
```

大部分在胞质中表达，其次在核内表达较多。

- 信号肽预测——SignalIP

SignalP 4.1 Server





```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    17       0.146
  max. Y    17       0.139
  max. S    14       0.176
  mean S     1-16    0.123
       D     1-16    0.130   0.340   NO
Name=Sequence   SP='NO' D=0.130 D-cutoff=0.340 Networks=SignalP-noTM
# data
# gnuplot script


Please cite:
SignalP 4.0: discriminating signal peptides from transmembrane regions
Petersen TN., Brunak S., von Heijne G. & Nielsen H.
Nature Methods, 8:785-786, 2011
```

没有明显的信号肽序列——入核可能是通过其他蛋白的帮助或有别的机制

人类不同类型的拓扑异构酶氨基酸残基分布

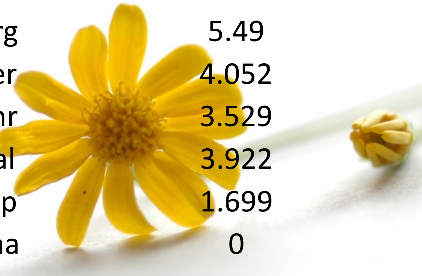
pepstats(v6.0.1)

	TOP1	TOP1B	TOP2A	TOP2B	TOP3A	TOP3B
Tiny	17.255	27.924	23.841	25.892	28.671	29.466
Small	38.17	49.54	46.832	48.831	49.95	50.464
Aliphatic	20	27.004	24.494	24.785	25.375	25.522
Aromatic	10.98	9.658	9.34	10.025	10.589	11.601
Non-polar	39.608	47.7	45.983	47.294	53.147	53.48
Polar	60.392	52.3	54.017	52.706	46.853	46.52
Charged	**44.706**	**27.858**	**32.397**	**31.98**	**26.973**	**27.494**
Basic	**26.144**	**15.046**	**17.636**	**16.974**	**15.584**	13. 90
Acidic	**18.562**	**12.812**	**14.762**	**15.006**	**11.389**	**11.369**

人类不同类型的拓扑异构酶氨基酸残基分布

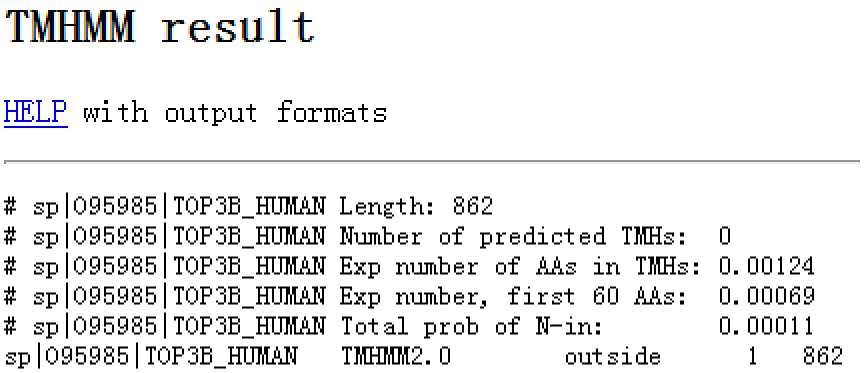
pepstats(v6.0.1)

Residue	TOP1	TOPB1	TOP2	TOP2B	TOP3A	TOP3B
Ala	5.229	6.307	4.703	5.904	6.613	6.613
Asx	0	0	0	0	0	0
Cys	1.046	2.431	0.849	1.046	3.944	3.944
Asp	7.451	4.534	6.728	7.257	5.22	5.22
Glu	11.111	8.279	8.034	7.749	6.148	6.148
Phe	3.399	3.285	3.788	4.428	3.944	3.944
Gly	3.399	4.599	5.487	6.458	6.961	6.961
His	3.137	2.957	1.568	1.599	3.48	3.48
Ile	4.706	5.059	5.487	4.982	3.48	3.48
Lys	17.516	7.556	11.692	11.501	6.845	6.845
Leu	6.144	9.001	8.034	7.565	8.933	8.933
Met	2.484	1.445	2.547	2.091	2.9	2.9
Asn	4.706	4.928	5.16	4.859	3.248	3.248
Pro	4.837	5.519	4.833	4.49	6.032	6.032
Gln	3.399	4.928	3.658	3.383	3.828	3.828
Arg	5.49	4.534	4.376	3.875	5.8	5.8
Ser	4.052	9.33	6.858	7.073	6.265	6.265
Thr	3.529	5.256	5.944	5.412	5.684	5.684
Val	3.922	6.636	6.27	6.335	6.497	6.497
Trp	1.699	1.117	1.176	1.169	1.044	1.044
Xaa	0	0	0	0	0	0
Tyr	2.745	2.3	2.809	2.829	3.132	3.132
Glx	0	0	0	0	0	0



4.蛋白质二级结构及结构域预测

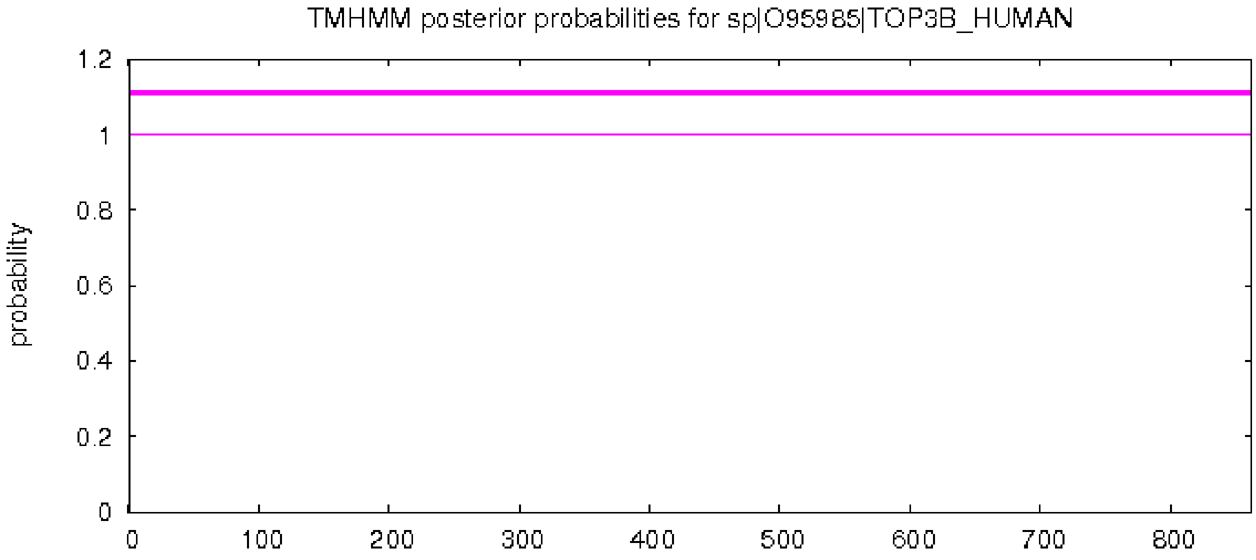
- 跨膜区预测——TMHMM



TMHMM result

HELP with output formats

```
# sp|O95985|TOP3B_HUMAN Length: 862
# sp|O95985|TOP3B_HUMAN Number of predicted TMHs:  0
# sp|O95985|TOP3B_HUMAN Exp number of AAs in TMHs: 0.00124
# sp|O95985|TOP3B_HUMAN Exp number, first 60 AAs:  0.00069
# sp|O95985|TOP3B_HUMAN Total prob of N-in:        0.00011
sp|O95985|TOP3B_HUMAN   TMHMM2.0        outside      1   862
```



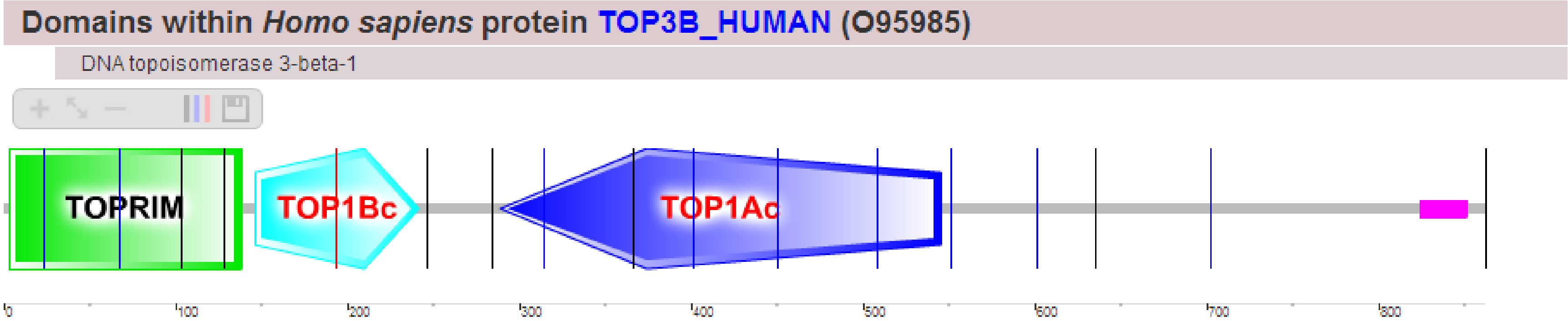
没有明显的跨膜区

SMART MODE:

NORMAL

GENOMIC

- 结构域预测

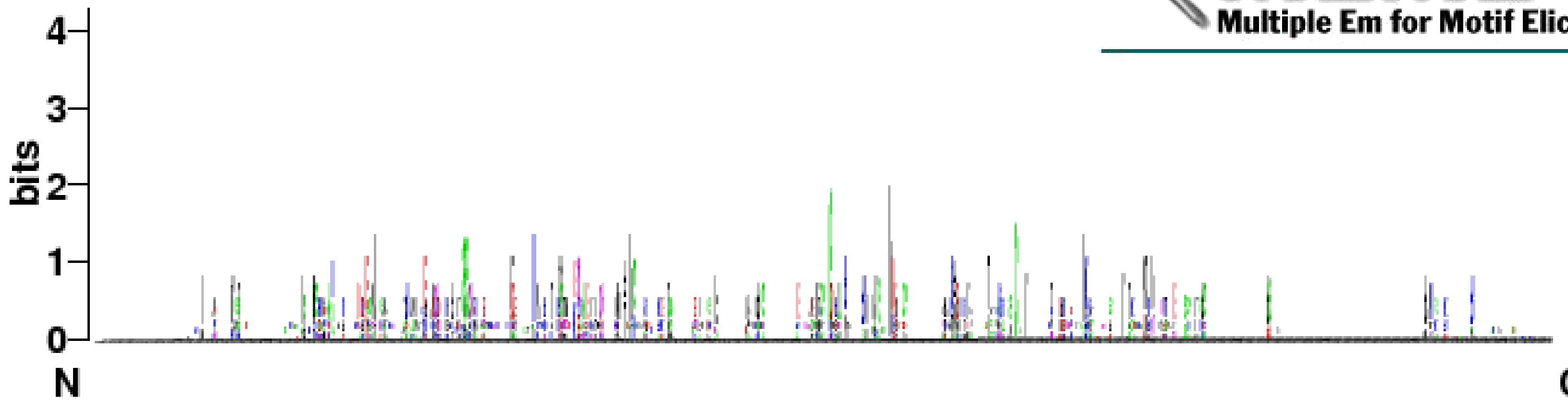


Domains within *Homo sapiens* protein TOP3B_HUMAN (O95985)

DNA topoisomerase 3-beta-1

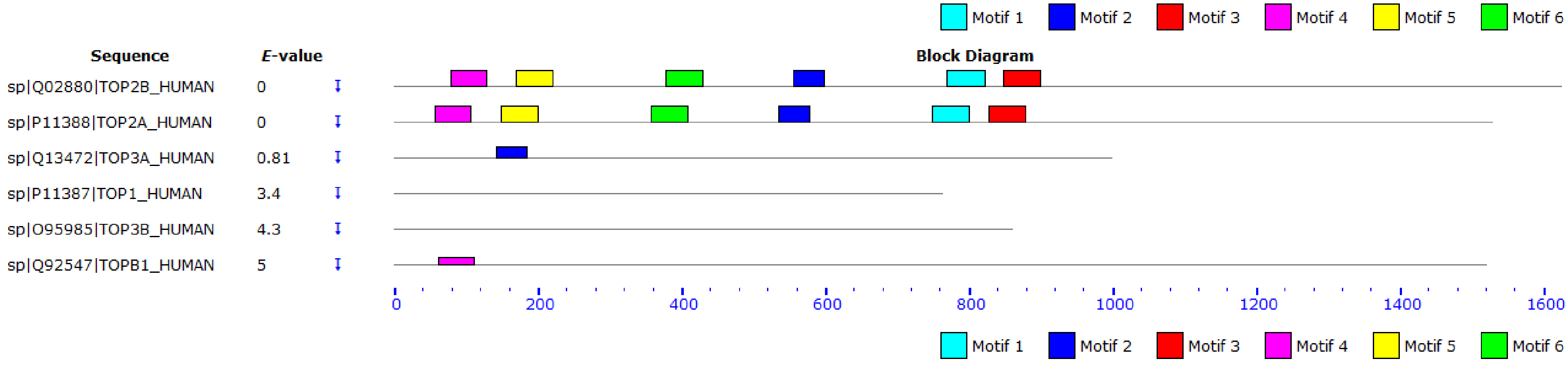
- Motif预测及保守性比较

MEME
Multiple Em for Motif Elicitation



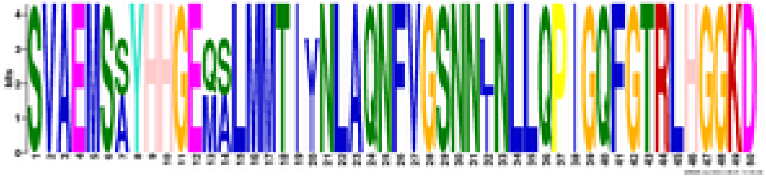
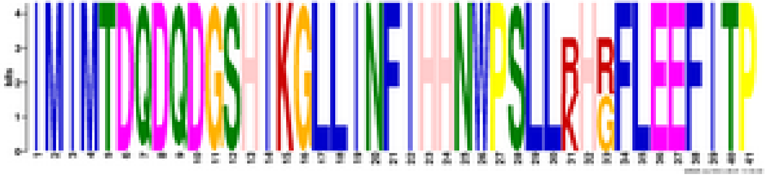
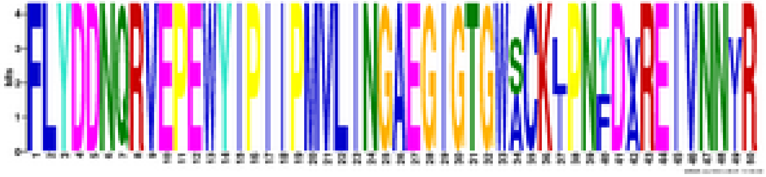
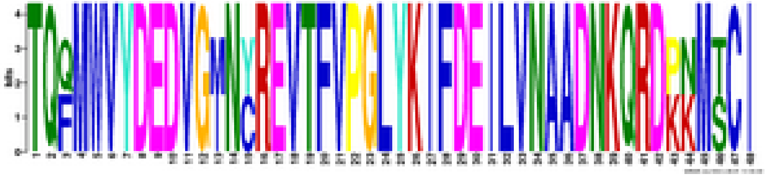
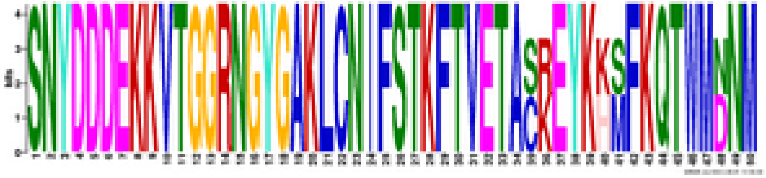
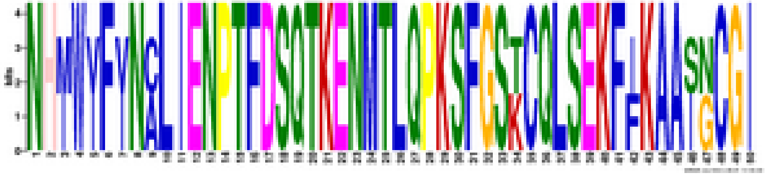
weblogo.berkeley.edu

全基因组保守性位点



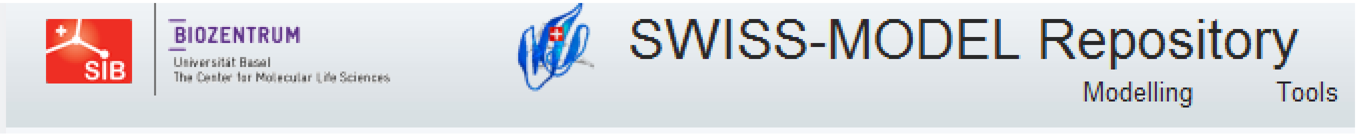
Sequence	*E*-value
sp|Q02880|TOP2B_HUMAN	0
sp|P11388|TOP2A_HUMAN	0
sp|Q13472|TOP3A_HUMAN	0.81
sp|P11387|TOP1_HUMAN	3.4
sp|O95985|TOP3B_HUMAN	4.3
sp|Q92547|TOPB1_HUMAN	5

人类拓扑异构酶Motif预测

Motif 1	- 2.7e-020 - 2 sites	
Motif 2	- 2.0e-019 - 2 sites	
Motif 3	- 1.4e-021 - 2 sites	
Motif 4	- 2.6e-019 - 2 sites	
Motif 5	- 9.6e-019 - 2 sites	
Motif 6	- 1.9e-017 - 2 sites	

5.蛋白质序列三级结构预测/同源建模

- 同源模建



SIB

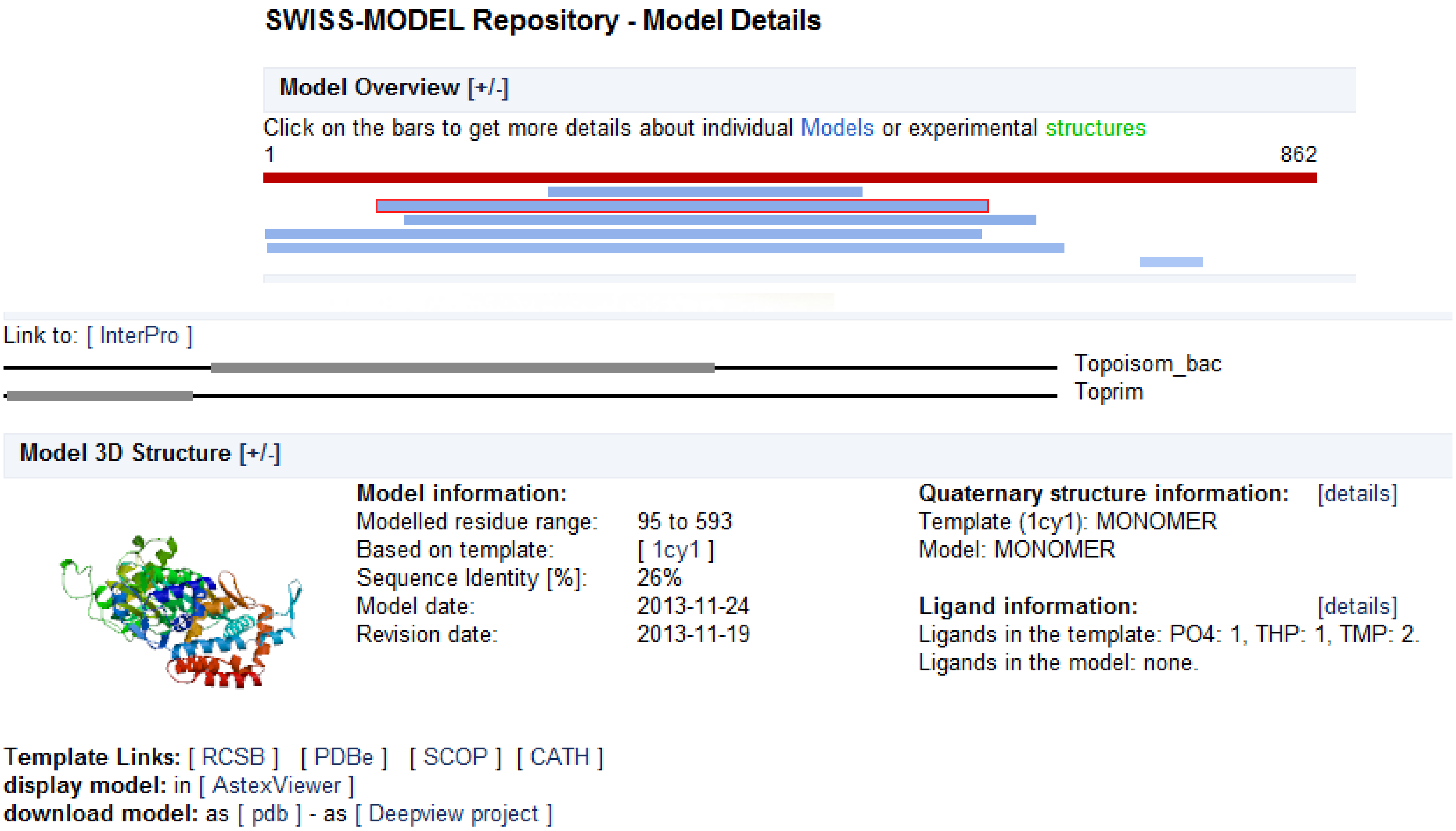
BIOZENTRUM

Universität Basel

The Center for Molecular Life Sciences

SWISS-MODEL Repository

Modelling Tools



SWISS-MODEL Repository - Model Details

Model Overview [+/-]

Click on the bars to get more details about individual Models or experimental structures

Link to: [InterPro]

Model 3D Structure [+/-]

Model information:

Modelled residue range: 95 to 593

Based on template: [1cy1]

Sequence Identity [%]: 26%

Model date: 2013-11-24

Revision date: 2013-11-19

Quaternary structure information: [details]

Template (1cy1): MONOMER

Model: MONOMER

Ligand information: [details]

Ligands in the template: PO4: 1, THP: 1, TMP: 2.

Ligands in the model: none.

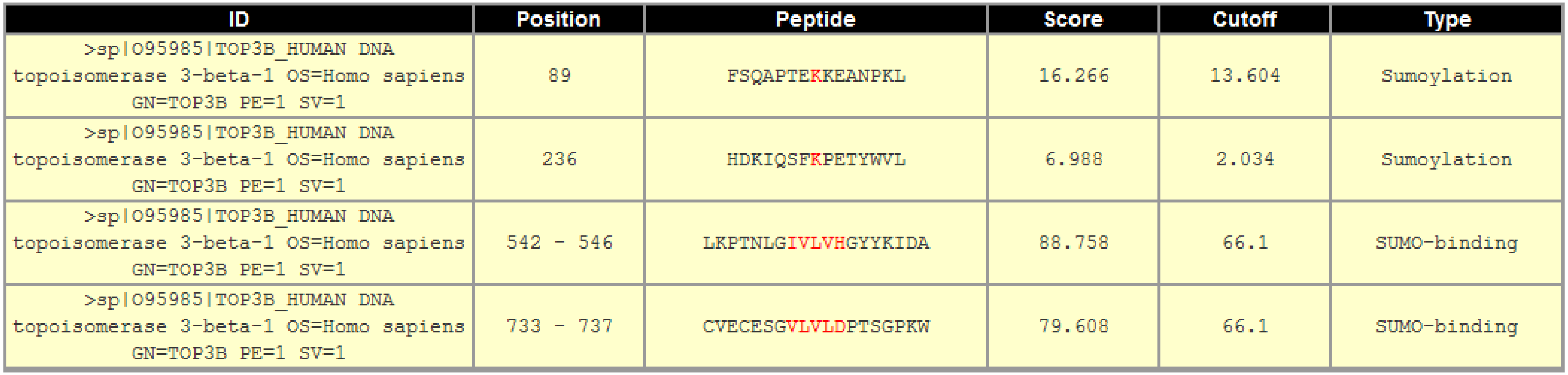
Template Links: [RCSB] [PDBe] [SCOP] [CATH]

display model: in [AstexViewer]

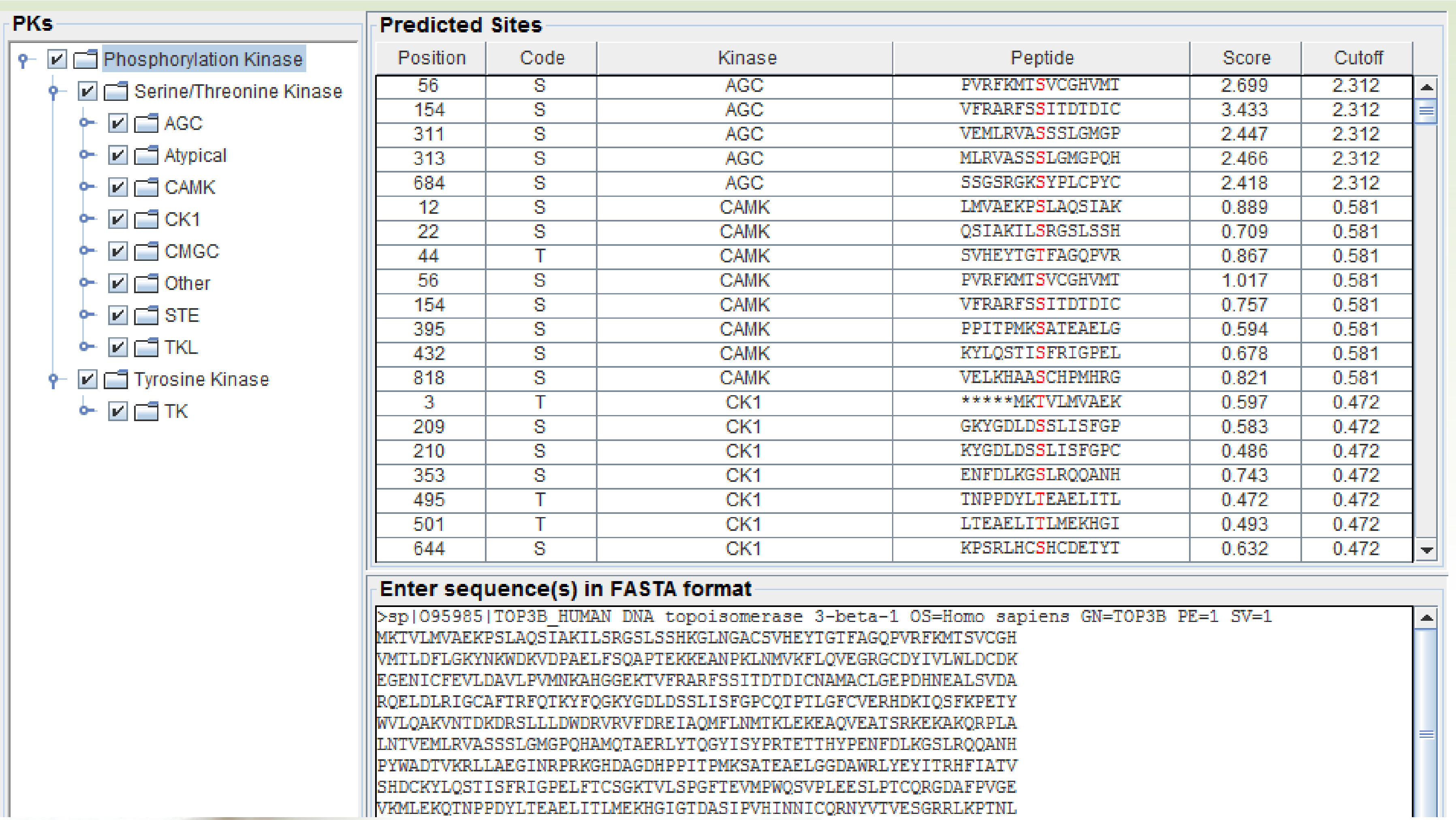
download model: as [pdb] - as [Deepview project]

6.结构特性分析及功能关系

- 磷酸化位点预测——GPS
- sumo预测SUMOsp2.0



ID	Position	Peptide	Score	Cutoff	Type
>sp|O95985|TOP3B_HUMAN DNA topoisomerase 3-beta-1 OS=Homo sapiens GN=TOP3B PE=1 SV=1	89	FSQAPTEKKEANPKL	16.266	13.604	Sumoylation
>sp|O95985|TOP3B_HUMAN DNA topoisomerase 3-beta-1 OS=Homo sapiens GN=TOP3B PE=1 SV=1	236	HDKIQSFKPETYWVL	6.988	2.034	Sumoylation
>sp|O95985|TOP3B_HUMAN DNA topoisomerase 3-beta-1 OS=Homo sapiens GN=TOP3B PE=1 SV=1	542 - 546	LKPTNLGIVLVHGYYKIDA	88.758	66.1	SUMO-binding
>sp|O95985|TOP3B_HUMAN DNA topoisomerase 3-beta-1 OS=Homo sapiens GN=TOP3B PE=1 SV=1	733 - 737	CVECESGVLVLDPTSGPKW	79.608	66.1	SUMO-binding



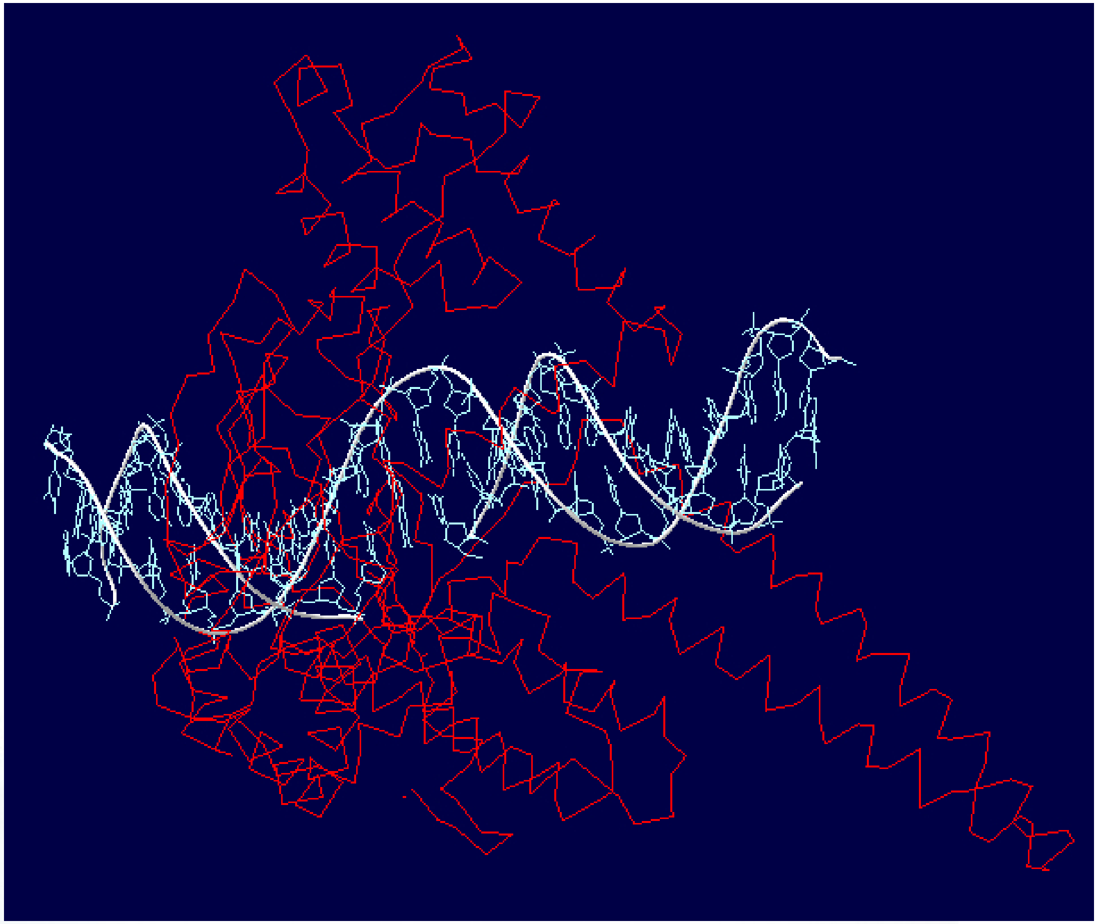
Predicted Sites

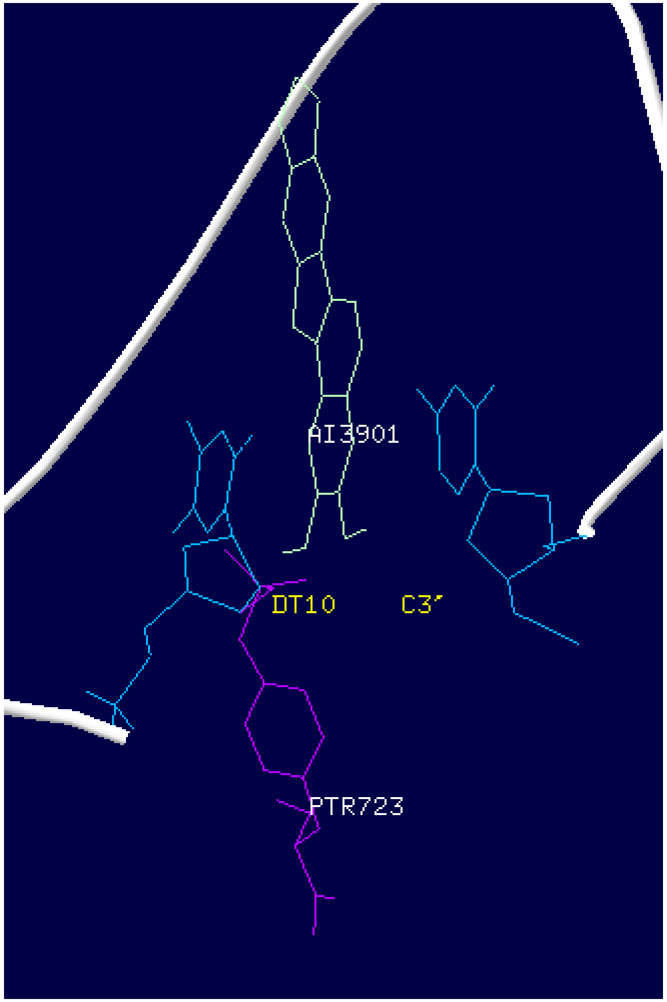
Position	Code	Kinase	Peptide	Score	Cutoff
56	S	AGC	PVRFKMTSVCGHVMT	2.699	2.312
154	S	AGC	VFRARFSSITDTDIC	3.433	2.312
311	S	AGC	VEMLRVASSSLGMGP	2.447	2.312
313	S	AGC	MLRVASSSLGMGPQH	2.466	2.312
684	S	AGC	SSGSRGKSYPLCPYC	2.418	2.312
12	S	CAMK	LMVAEKPSLAQSIAK	0.889	0.581
22	S	CAMK	QSIAKILSRGSLSSH	0.709	0.581
44	T	CAMK	SVHEYTGTFAGQPVR	0.867	0.581
56	S	CAMK	PVRFKMTSVCGHVMT	1.017	0.581
154	S	CAMK	VFRARFSSITDTDIC	0.757	0.581
395	S	CAMK	PPITPMKSATEAELG	0.594	0.581
432	S	CAMK	KYLQSTISFRIGPEL	0.678	0.581
818	S	CAMK	VELKHAASCHPMHRG	0.821	0.581
3	T	CK1	*****MKTVLMVAEK	0.597	0.472
209	S	CK1	GKYGDLDSSLISFGP	0.583	0.472
210	S	CK1	KYGDLDSSLISFGPC	0.486	0.472
353	S	CK1	ENFDLKGSLRQQANH	0.743	0.472
495	T	CK1	TNPPDYLTEAELITL	0.472	0.472
501	T	CK1	LTEAELITLMEKHGI	0.493	0.472
644	S	CK1	KPSRLHCSHCDETYT	0.632	0.472

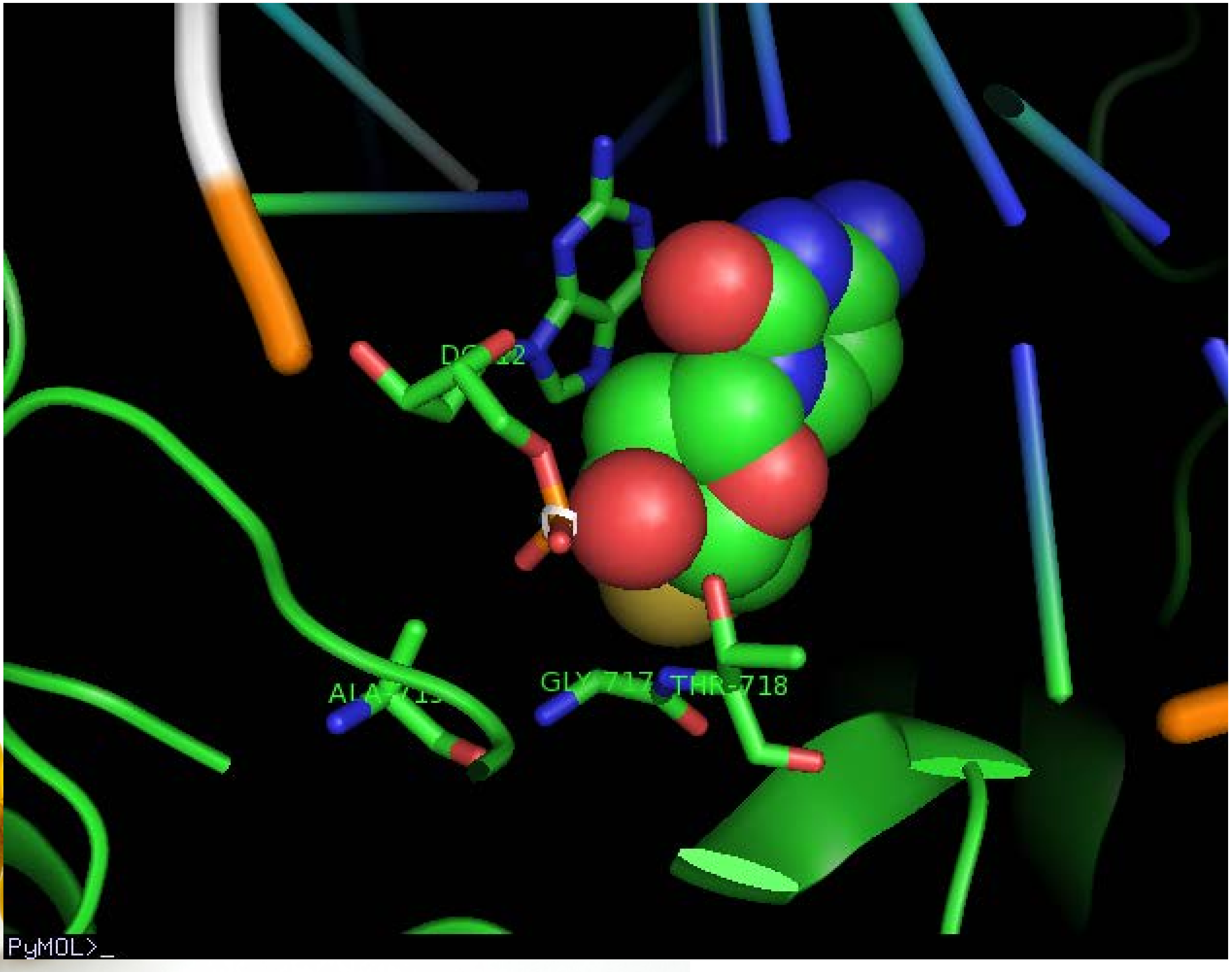
Enter sequence(s) in FASTA format

```
>sp|O95985|TOP3B_HUMAN DNA topoisomerase 3-beta-1 OS=Homo sapiens GN=TOP3B PE=1 SV=1
MKTVLMVAEKPSLAQSIAKILSRGSLSSHKGLNGACSVHEYTGTFAGQPVRFKMTSVCGH
VMTLDFLGKYNKWDKVDPAELFSQAPTEKKEANPKLNMVKFLQVEGRGCDYIVLWLDCDK
EGENICFEVLDAVLPVMNKAHGGEKTVFRARFSSITDTDICNAMACLGEPDHNEALSVDA
RQELDLRIGCAFTRFQTKYFQGKYGDLDSSLISFGPCQTPTLGFCVERHDKIQSFKPETY
WVLQAKVNTDKDRSLLLDWDRVRVFDREIAQMFLNMTKLEKEAQVEATSRKEKAKQRPLA
LNTVEMLRVASSSLGMGPQHAMQTAERLYTQGYISYPRTETTHYPENFDLKGSLRQQANH
PYWADTVKRLLAEGINRPRKGHDAGDHPPITPMKSATEAELGGDAWRLYEYITRHFIATV
SHDCKYLQSTISFRIGPELFTCSGKTVLSPGFTEVMPWQSVPLEESLPTCQRGDAFPVGE
VKMLEKQTNPPDYLTEAELITLMEKHGIGTDASIPVHINNICQRNYVTVESGRRLKPTNL
```

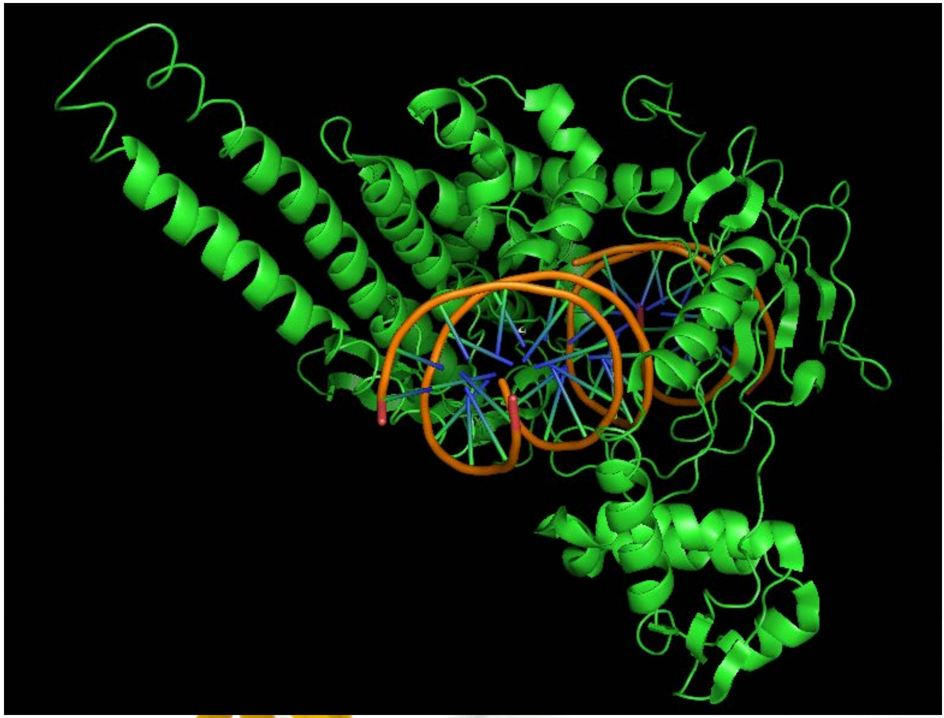
- top 1 in complex with DNA —single strand break (1TL8)
- DT10, TPC11,PTR723



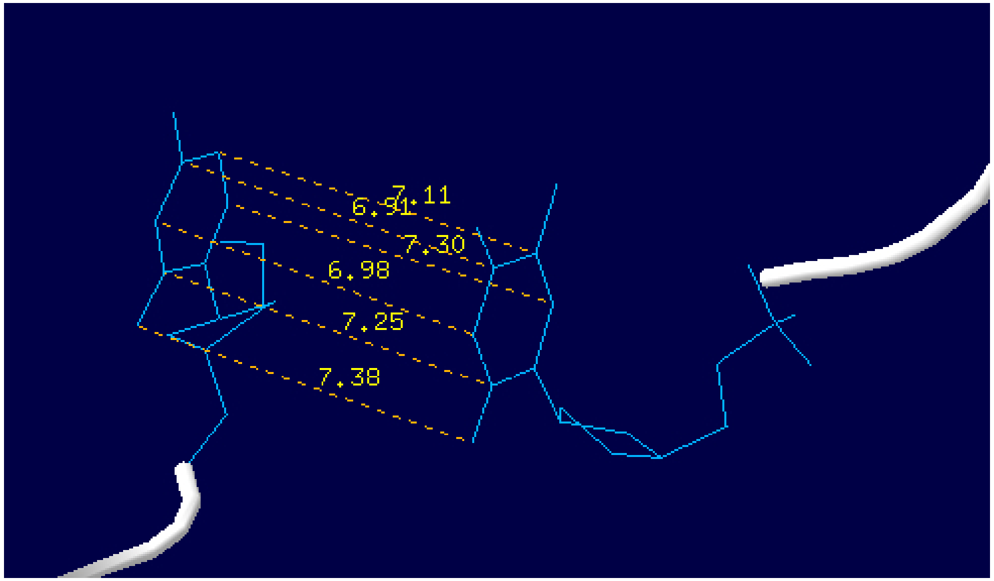


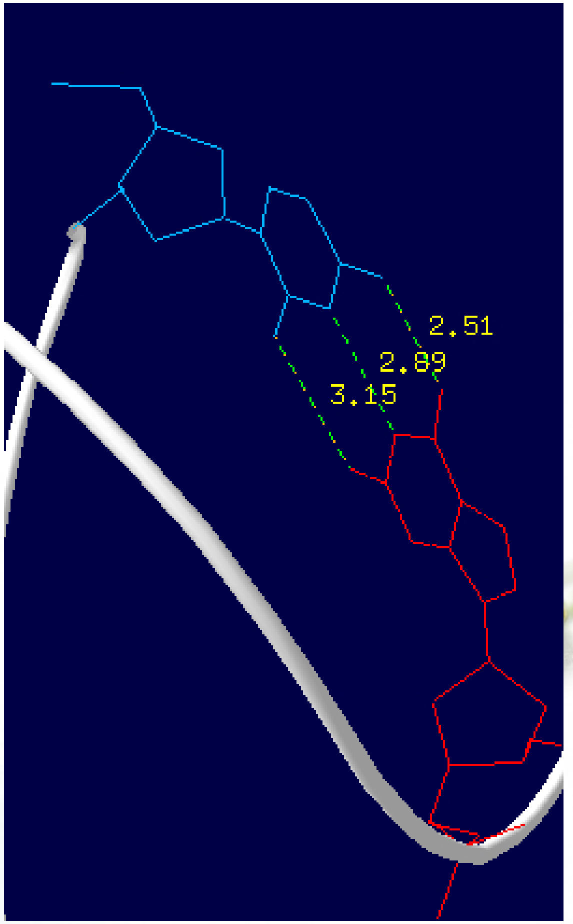


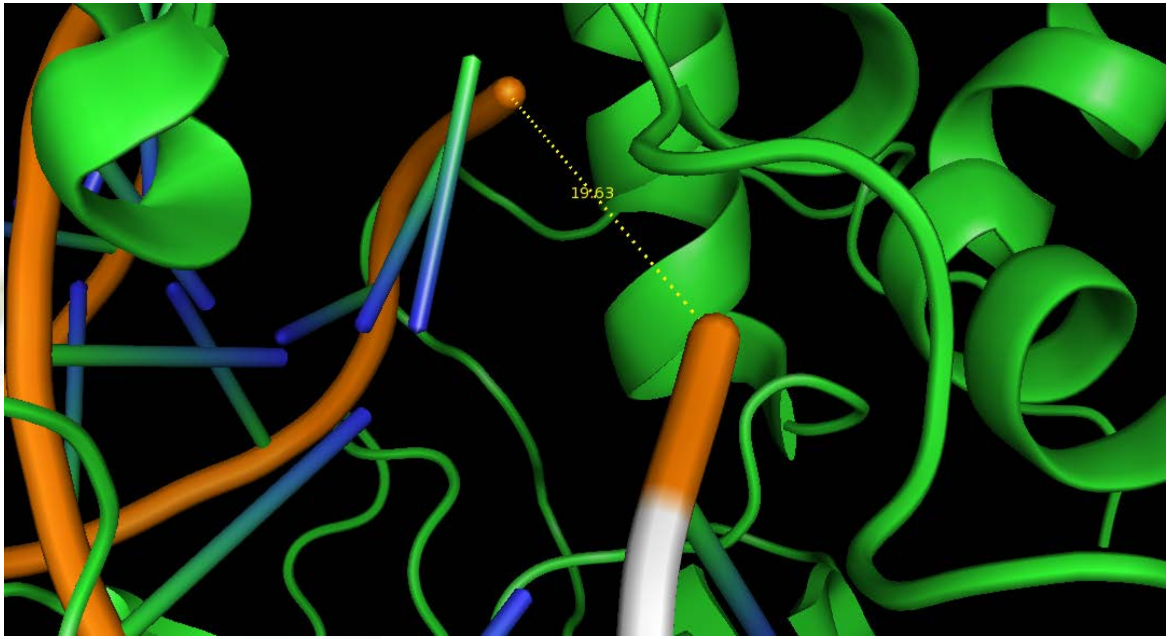
```
PyMOL>_
```



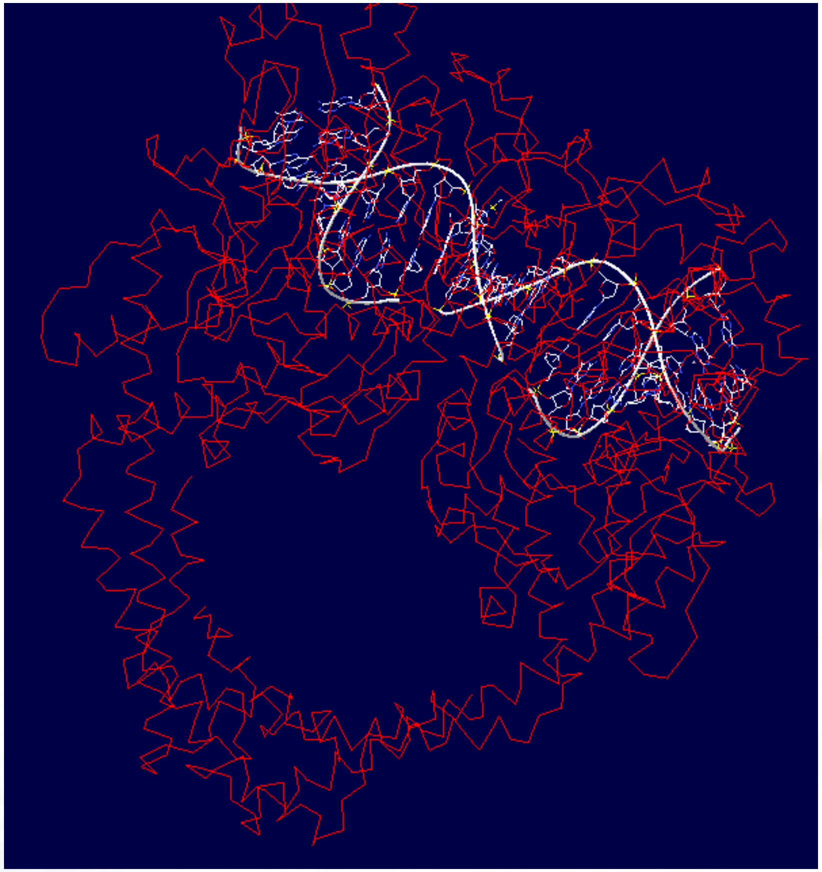
- The distance between two breaks: 6.91Å
- TPC11:DG112
- DT10:DA113

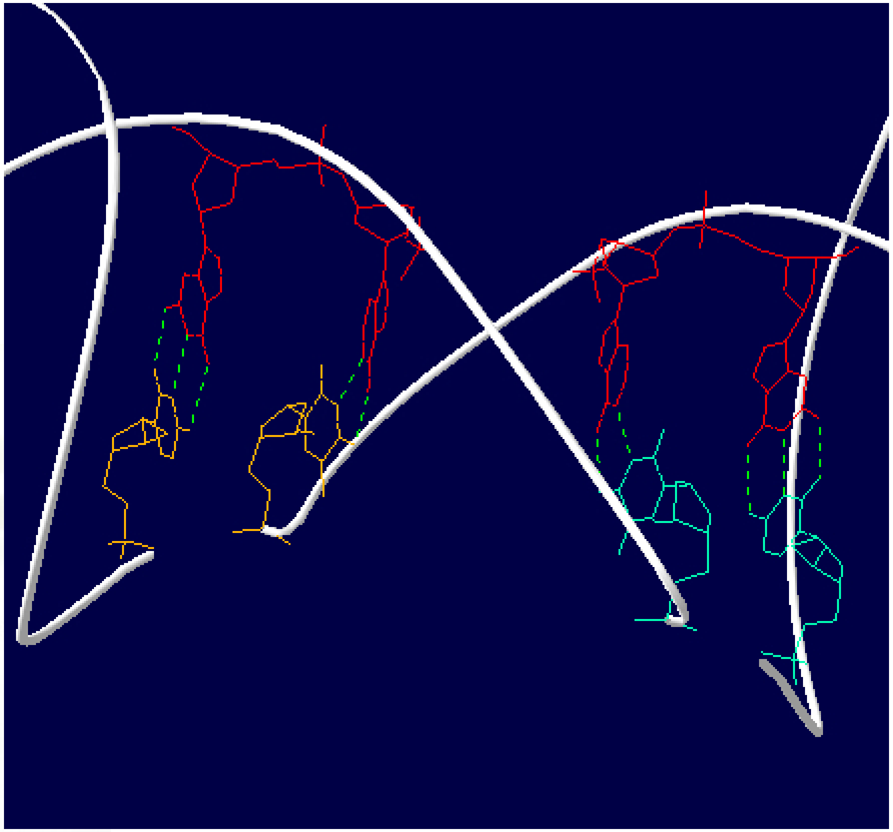






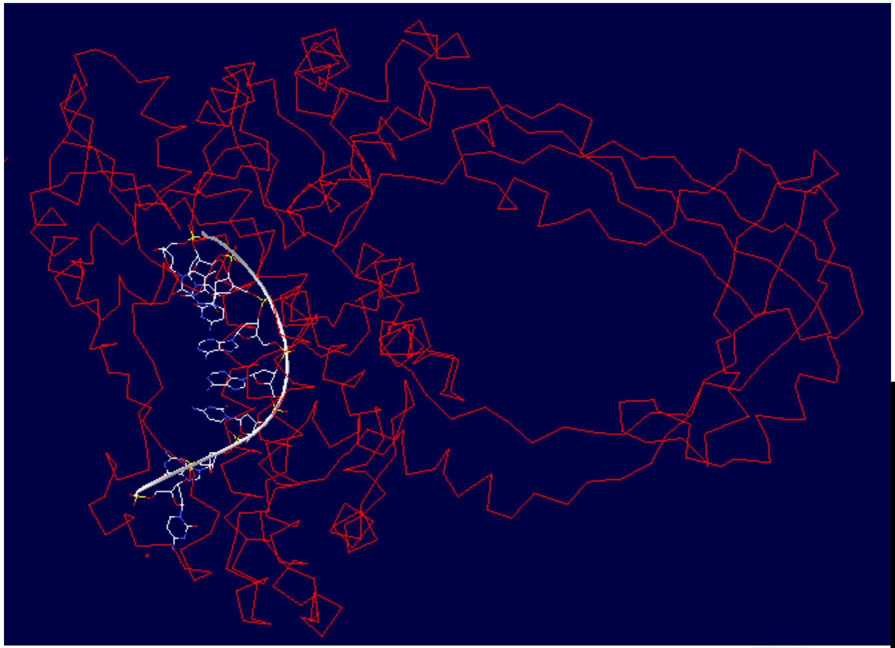
- top 2b in complex with DNA—two breaks (4G0U)
- DC8:DG13
- DT9:DA12





E.coli toposoimerase 3 (1I7D)

—with 8 base single strand DNA



Results and discussion

- 正确并适当的使用生物信息学工具有助于科研工作
- 生物信息学工具预测的结果需要“湿实验”加以验证；
- 科研工作—下结论需慎重；

Thanks!