Applied Bioinformatics Course Final Seminar

Class 1, Peking University

Speaker: Changzu Cai

2010-06-13

Overview of Groups in Class 1

Group 1

Speaker: Jieru Li

bZIP Transcription Factor

Plant transcription factor database

Phylogenetic tree

Transmembrane domain prediction

Group 2

Speaker: Yibin Peng

RuBicCO

Swiss-PDB viewer-find active site

Phylogenetic tree

Group 3

Speaker: Shuaipeng Ma

FAIM1-CTP

Homology Modeling

Group 4

Speaker: Huaqian Yang

Calmodulin

Phylogenetic tree-MEGA

Structure Comparison with and without Ca^+

Group 5

Speaker: Zhaorui Shen

Subcellular localization

Comparison of a series of databases

Group 6

Speaker: Changzu Cai

HCV NS3 helicase-Pymol

Group 7

Speaker: Yue Huang

Cotton UER gene

Phylogenetic tree

Group 8

Speaker: Yu Hua

T7 SSB

Structure and function analysis

Human CD20 and Rituximab Interaction Analyzing and A CD20 Homologous Protein Modeling

Rituximab

- Rituxan or Rituximab is a monoclonal antibody directed against the CD20 antigen found on B cells.
- It was reported to be used in the treatment of B cell lymphomas, autoimmune diseases, graft rejection, RA (rheumatoid arthritis) and so on .
- Rituxan was first MoAb approved by FDA for the treatment of non-Hodgkin's lymphoma in 1997.
- Kill CD20+ cells by Complement dependent cytotoxicity, antibody-dependent cell-mediated cytotoxicity and apoptosis
- Binding of Rituximab to CD20 is not sufficient to kill many lymphoma cells, indicating that there are mechanisms of resistance.

CD20

- A non-glycosylated phosphoprotein

- Encoded by *MS4A1* gene

- Expressed on the surface of all mature B-cells, but not on either pro-B cells or plasma cells

- Have no known natural ligand and its function is unclear

- Target of Rituximab, which is an active agent in the treatment of all B cell lymphomas and leukemias
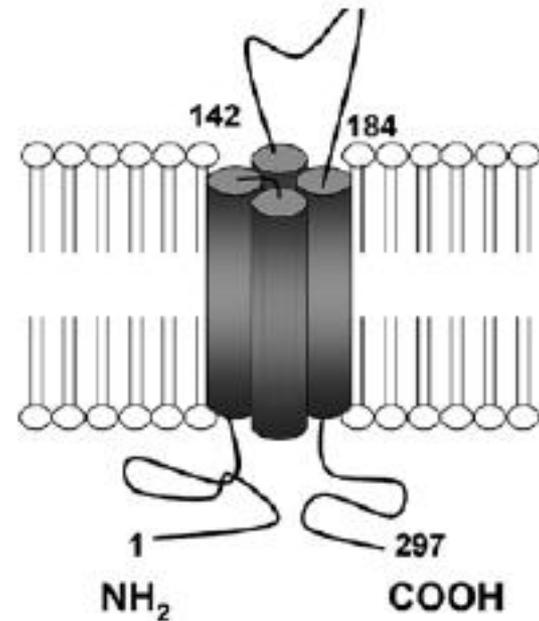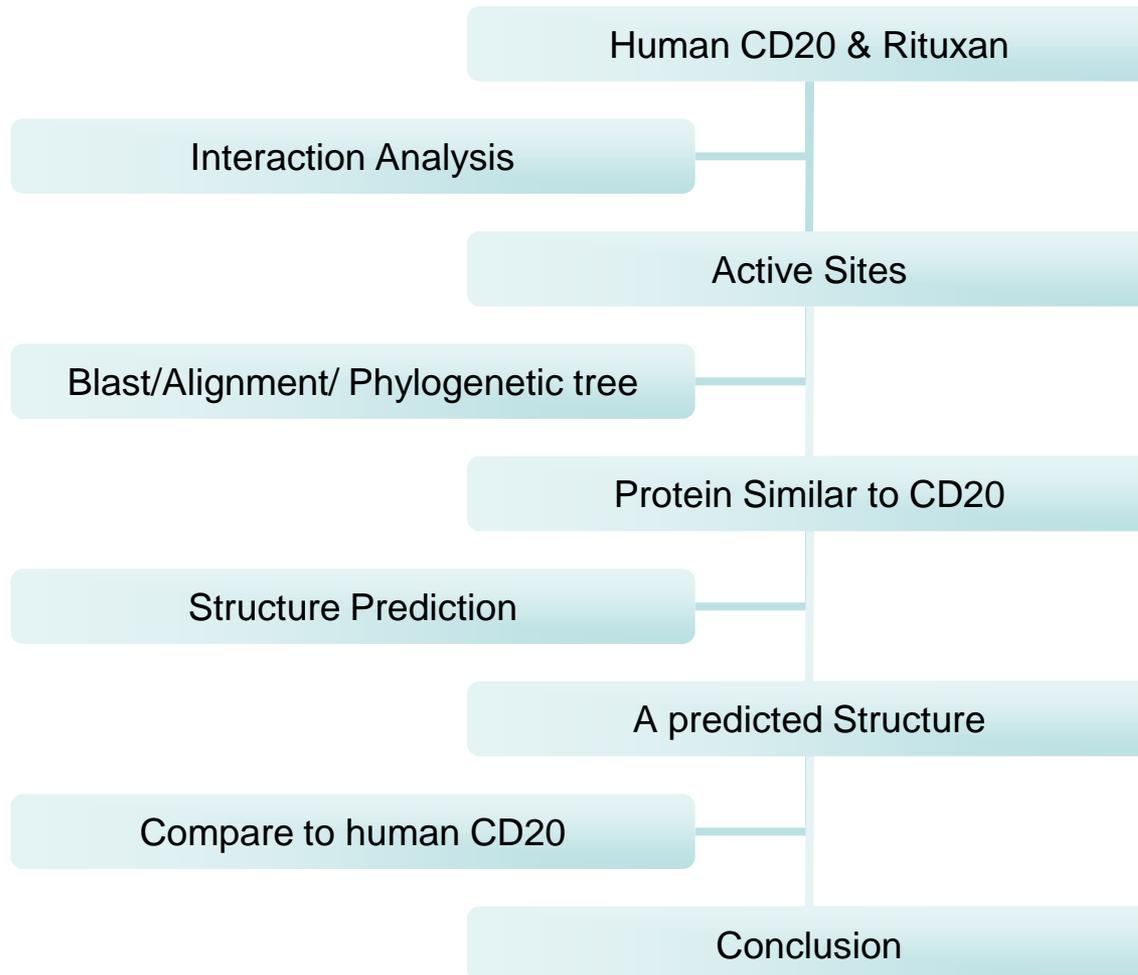
CD 20 in UniProt

Topological domain	1 – 63	63	Cytoplasmic Potential
Transmembrane	64 – 84	21	Potential
Transmembrane	85 – 105	21	Potential
Transmembrane	121 – 141	21	Potential
Transmembrane	189 – 209	21	Potential
Topological domain	210 – 297	88	Cytoplasmic Potential
Region	74 – 80	7	Epitope 1
Region	146 – 160	15	Epitope 2
Region	168 – 175	8	Epitope 3 (recognized by antibodies, including Rituximab)

Residue 168-175 can be used to analyze the interaction between CD20 and Rituximab.

Model organism is needed in the research of Rituximab resistance mechanism.

Methodology

Human CD20 & Rituxan

Interaction Analysis

Active Sites

Blast/Alignment/ Phylogenetic tree

Protein Similar to CD20

Structure Prediction

A predicted Structure

Compare to human CD20

Conclusion

PyMOL

uA Swiss-PDBViewer-like software used mainly for **protein structure visulization**
uWritten by Warren Lyford Delano

u**Powerful Graphic Functionalities**: **25%** of all the 3D protein structures in scientific papers

uPyMOL = Python(A programming language easy to learn and use) + Molecular
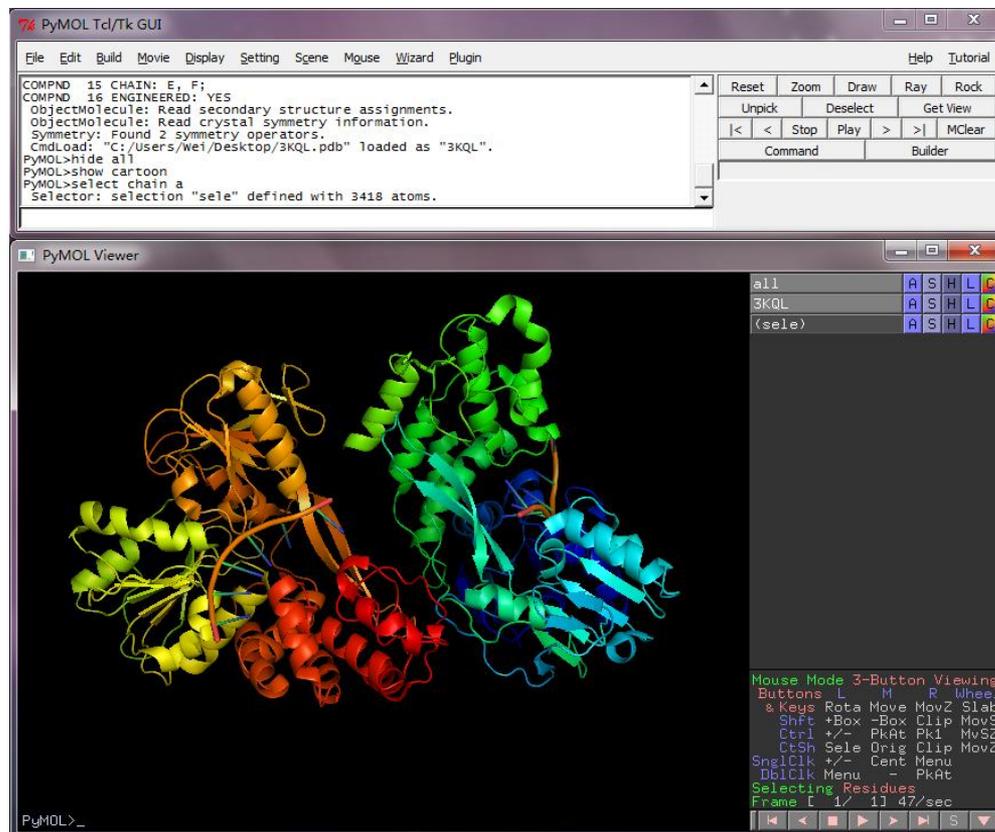
Targeting DNA, BIOCHEMIE ↗, July 1, 2008 Cover, Vol. 90.

Harnessing Helices, Chemical & Engineering News ↗, June 2, 2008 Cover, Vol. 86, Issue 22.

Fusarium head blight (FHB), JBC ↗, January 18, 2008 Cover, Vol. 283, Issue 3.
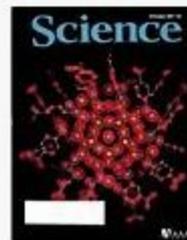
Pumping Ions, Nature ↗, Dec. 13th, 2007.
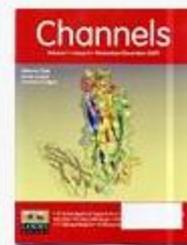
[3] ↗, Dec 2007.

Science ↗, Nov. 23rd, 2007.

Sensing Calciums[4] ↗, Nov 20th, 2007.

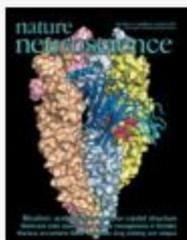Aminoacyl-tRNA synthetases from human mitochondria[5] ↗, Nov. 13th, 2007.
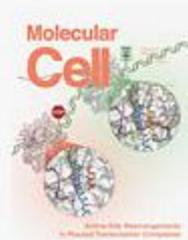
Science ↗, Oct. 19th, 2007.

Sensing Acid, Nature ↗, Sept. 20th, 2007.
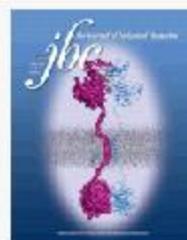
Form Finds Function?, Nature , Aug. 16thm 2007.

Nature Neuroscience ↗ Vol. 10 No. 8 Aug. 2007

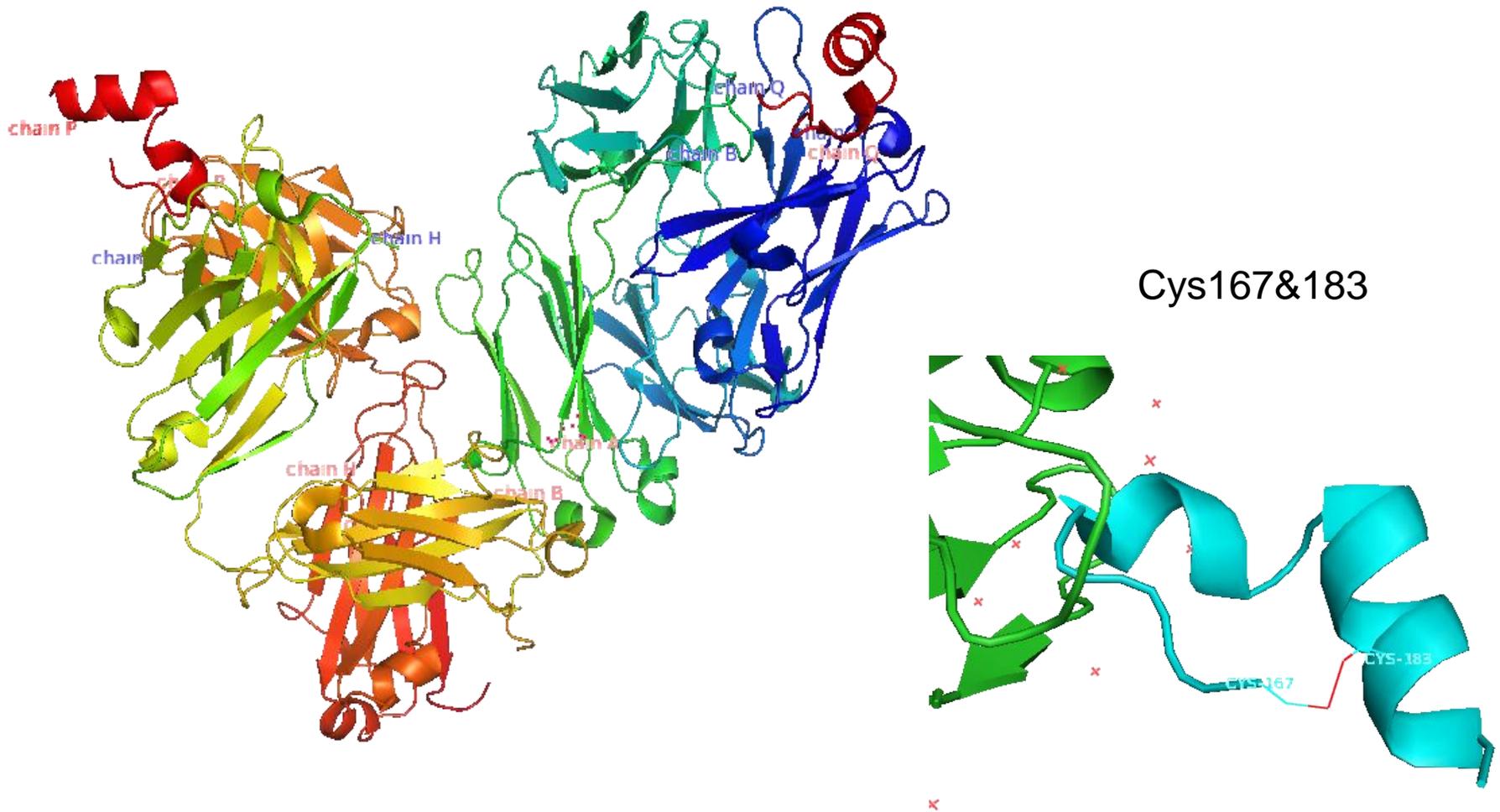Paused transcription complexes, Molecular Cell , August 3 2007
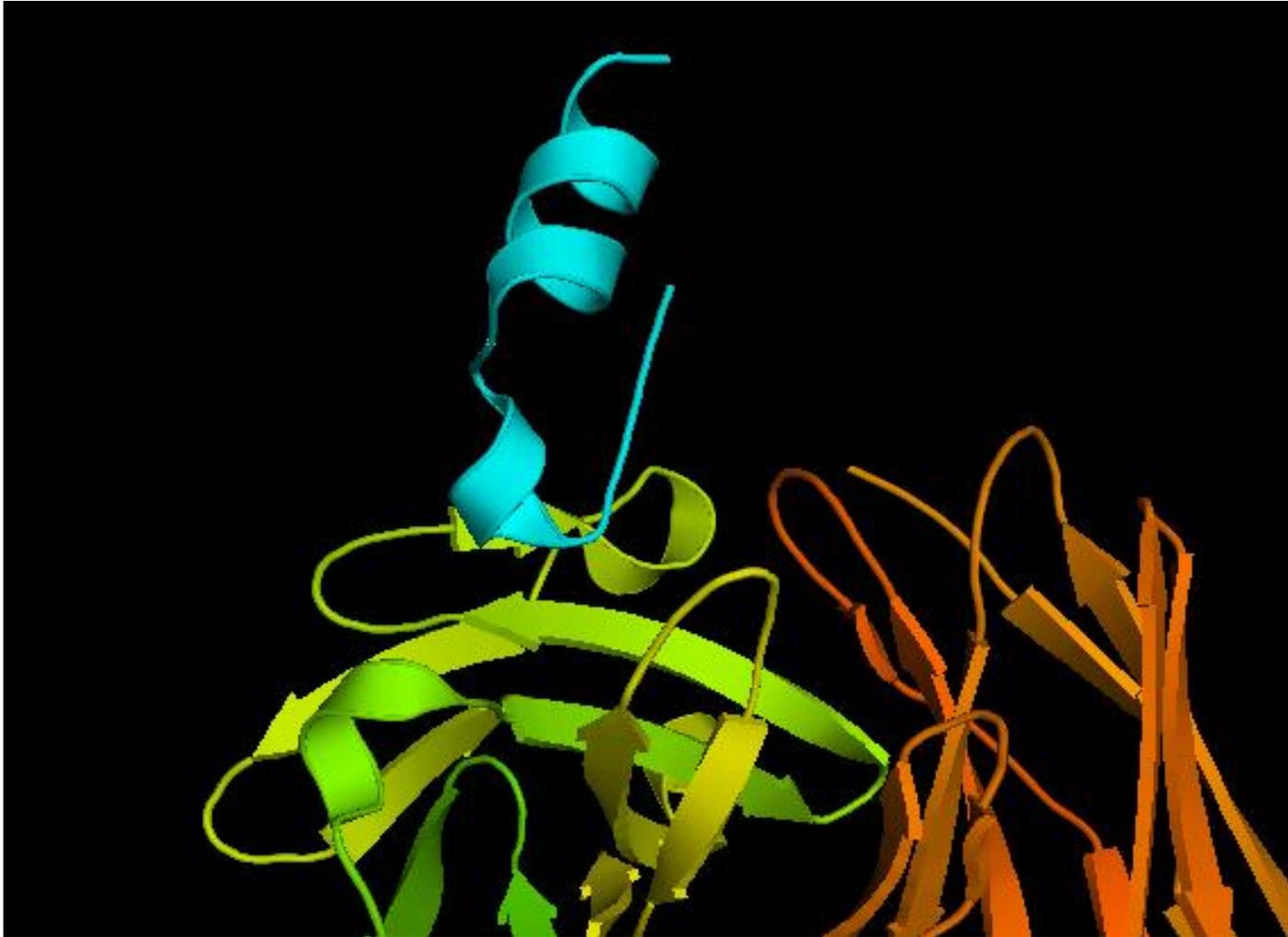
Vitamin D, ABB ↗, April 14 2007 Cover, Vol. 460, Issue 2.

Mycobacterium tuberculosis PknD dimerization JBC ↗, April

PyMOL's Windows

Human CD20 and Rituximab Interaction Analysis

Cys167&183

Human CD20 and Rituximab Interaction Analysis-Overview

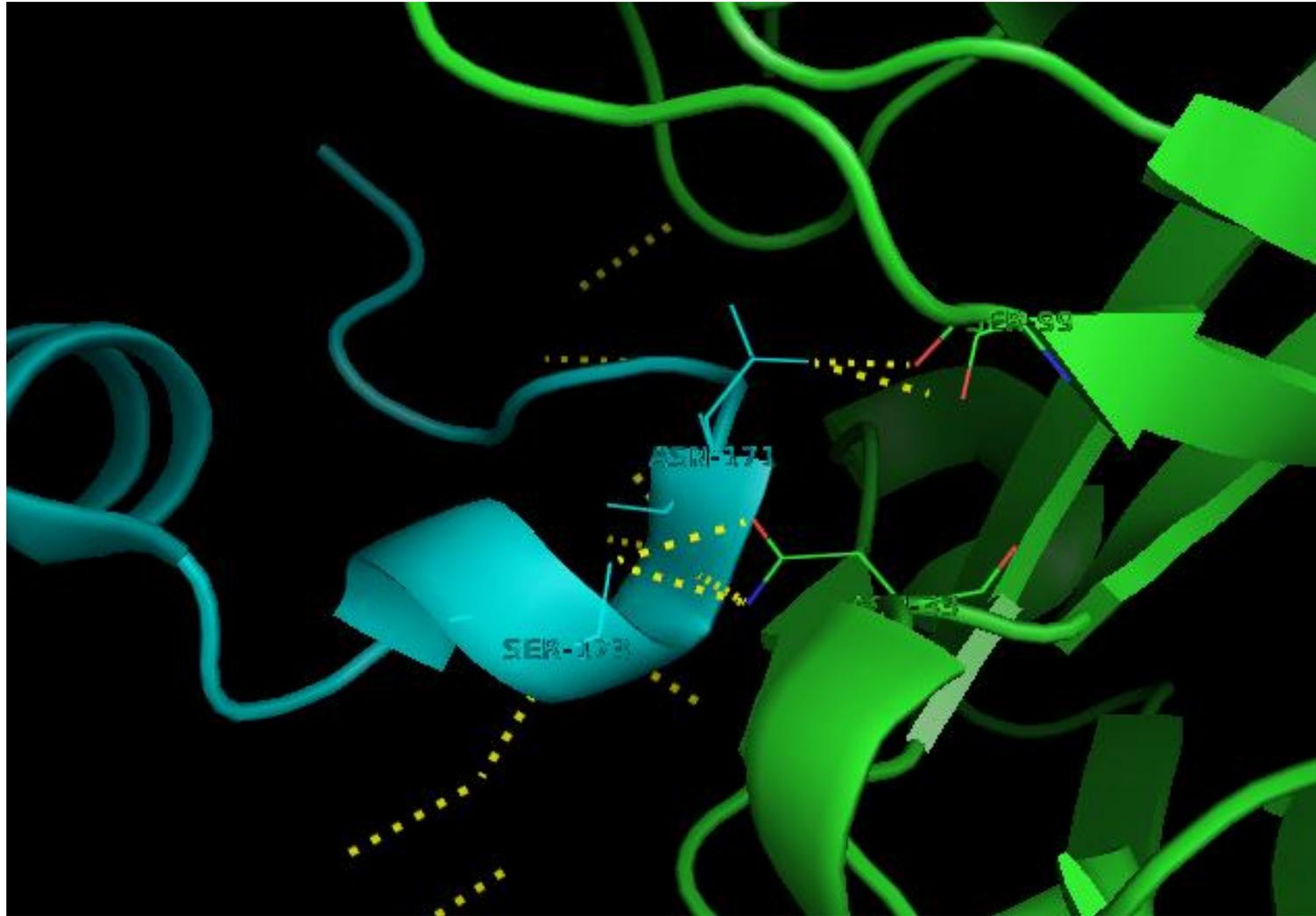# Human CD20 and Rituximab Interaction Analysis-Polar Contacts

Asn 171, Ser 173

Asn166, Glu168, Pro169

Homologous protein Searching and Prediction
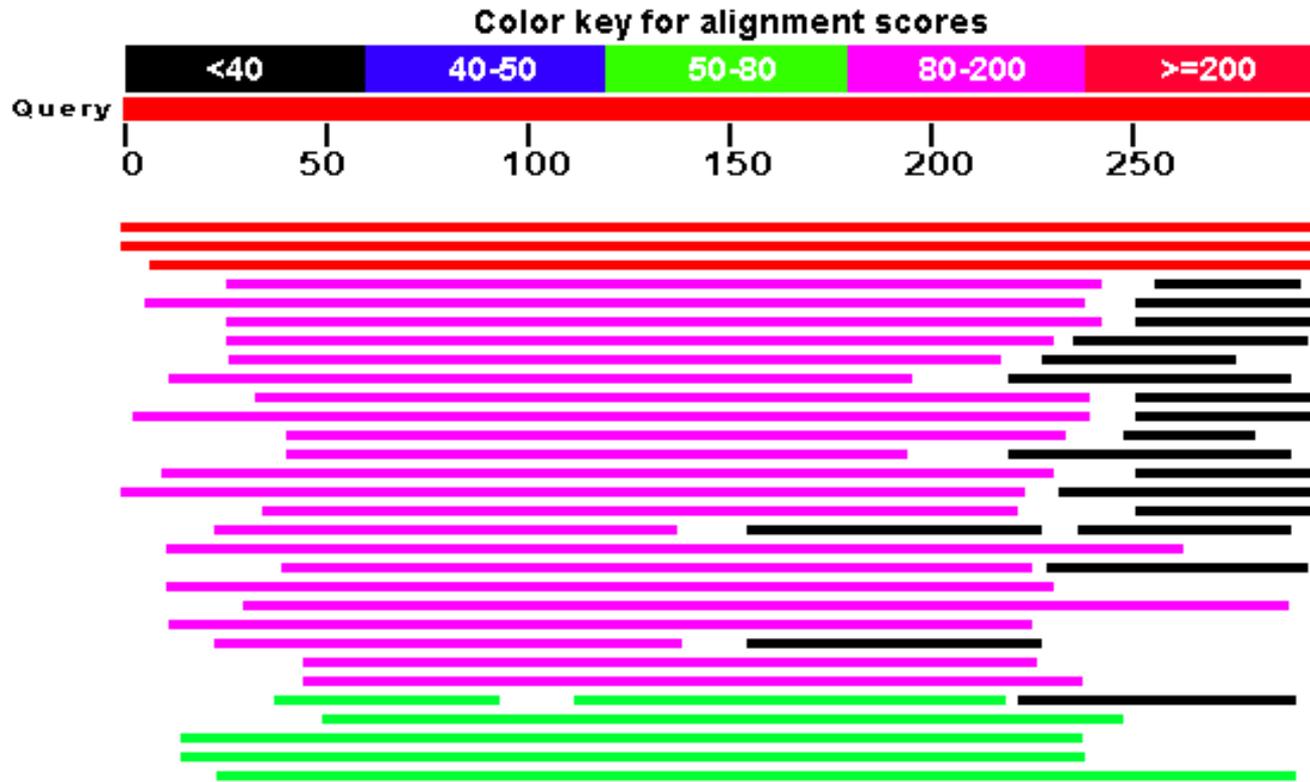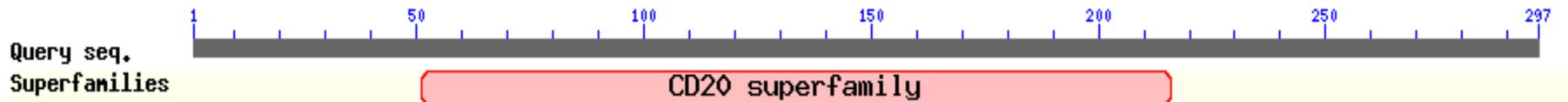
Comparison of Different Phylogenetic Tree Construction Methods

Method	Advantages	Disadvantages	应用情形
UPGMA	Very Fast	Real data rarely meet the assumption	分类数目巨大,粗略分析(DNA芯片)
NJ	Fast	Not guaranteed to find the best tree	分类数目较多,序列差异较大,对结果的准确性要求不太高
ME	More accurate than NJ	Slow	分类数目不太多
MP	More accurate when sequence divergence is low	Less effective when sequence divergence is high Long-branch attraction	分类数目不太多且序列差异较小,对结果的准确性要求较高
ML	Vigorous and Exhaustive	Very Slow Depends on the model	分类数目较少,对结果的准确性要求高

Phylogenetic Tree

```
                                                +------FCERB RAT
                                        +-49.0-|
                                +--------49.0-|    +------FCERB MOUS
                                |            |
                                |            +-----------FCERB HUMA
                        +-27.0-|
                        |       |            +------M4A8A MOUS
                        |       |    +--------49.0-|
                        |       |    |            +------M4A8B HUMA
                        |       +-48.0-|
```

```
        |              |                                               +------F1891 MOUS
+-39.5-|              |                                       +-49.0-|
   |    |              |  +-----------------------------------------------49.0-|   +------F1891 HUMA
   |    |              |  |                                                        |
   |    |              |  |                                                        +------------M4A14 HUMA
+------|              |  |
   |    |              |  +--------------------------------------------------------------CD20 MOUSE
   |    |              |
   |    |              +-----------------------------------------------------------------CD20 HUMAN
   |
+-----------------------------------------------------------------------------------CD20 CANFA
```
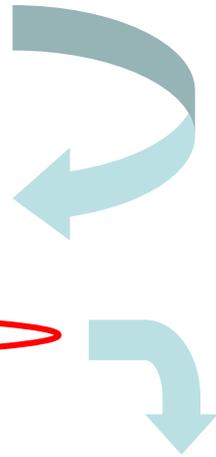
```
                            +-46.0-|    +------M4A6Z MOUS
                            |    |    |
                    +-22.0-|    +------------M4A7A HUMA
            +-26.5-|    |  
            |       |  +------------36.5-|    +------------MS4A7 HUMA
            |       |                    |
            |       |                    +------M4A6C MOUS
            |       +------------24.0-|
   +-49.0-|       |                    +------M4A6E HUMA
   |       |       +------------T176B HUMA
   |       |                    +------M4A10 MOUS
   |       +------------49.0-|
   |                            |  +------M4A10 HUMA
+-49.0-|                    +-49.0-|
   |                            +------M4A10 CONA
   |                    +------F1891 MOUS
+-39.5-|            +-49.0-|
   |       |            +------------F1891 HUMA
+------|            +------------M4A14 HUMA
   |                    -------------CD20 MOUSE
   |    -------------CD20 HUMAN
+-----------CD20 CANFA
remember: this is an unrooted tree!
```

CANFA

HUMAN

MOUSE

Alignment of Binding Domain
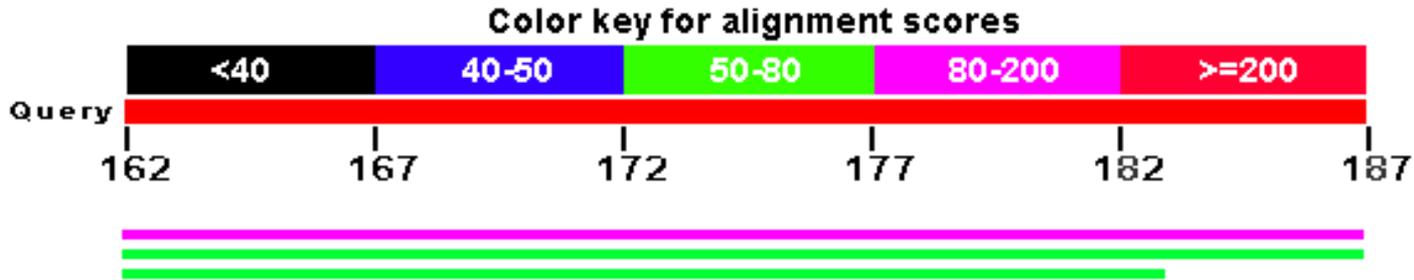
Mouse CD20

Mouse CD20 Structure Prediction

Available Databases

Tertiary structure prediction

Homology modeling

- SWISS-MODEL 🦠 - An automated knowledge-based protein modelling server
- 3Djigsaw - Three-dimensional models for proteins based on homologues of known structure
- CPHmodels - Automated neural-network based protein modelling server
- ESyPred3D - Automated homology modeling program using neural networks
- Geno3d - Automatic modelling of protein three-dimensional structure
- SDSC1 - Protein Structure Homology Modeling Server

Threading

- Phyre (Successor of 3D-PSSM) - Automated 3D model building using profile-profile matching and secondary structure
- Fugue - Sequence-structure homology recognition
- HHpred - Protein homology detection and structure prediction by HMM-HMM comparison
- Libellula - Neural network approach to evaluate fold recognition results
- LOOPP - Sequence to sequence, sequence to structure, and structure to structure alignment
- SAM-T02 - HMM-based Protein Structure Prediction
- Threader - Protein fold recognition

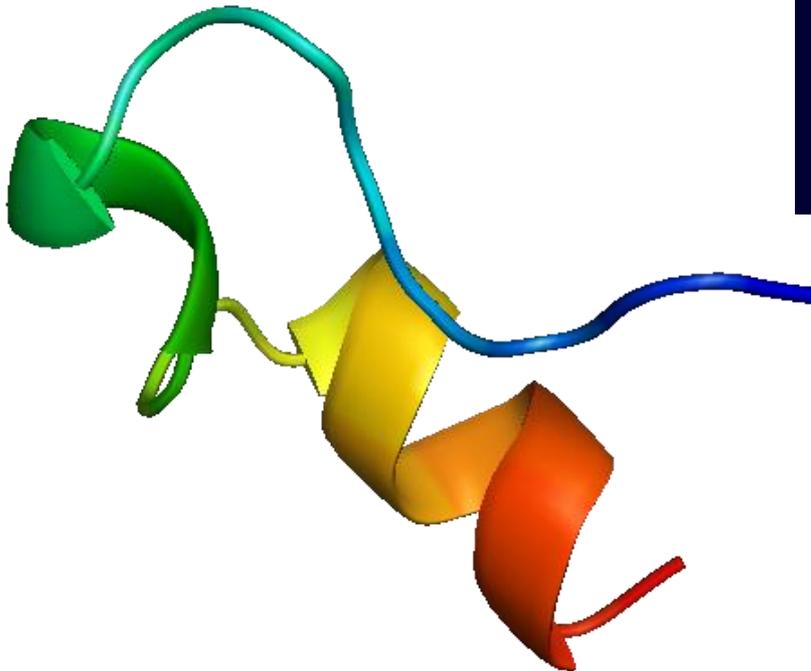- SWEET - Constructing 3D models of saccharides from their sequences

Ab initio

- HMMSTR/Rosetta - Prediction of protein structure from sequence

Three Modes in SWISS-MODEL

Automated Mode

Alignment Mode

Project Mode √
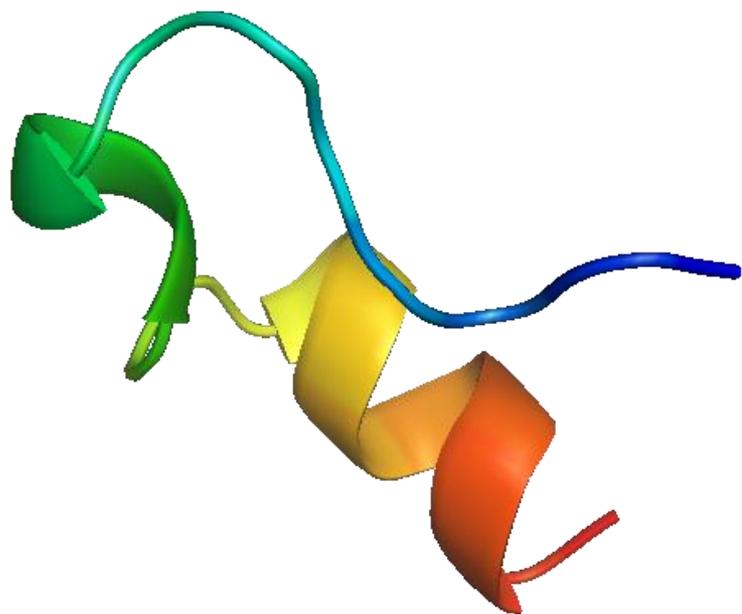
Predicted by Phyre (Version 0.2)

| E-mail Address | |
| Optional Job description | |

Amino Acid Sequence

Quick Phyre Search

View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText
	c2oslP_ (length:25) 80% i.d.	Jmol MDL	21	10 %	n/a
	c2oslQ_ (length:25) 80% i.d.	Jmol MDL	21	10 %	n/a

Alignment of Predicted Structures by Different Methods
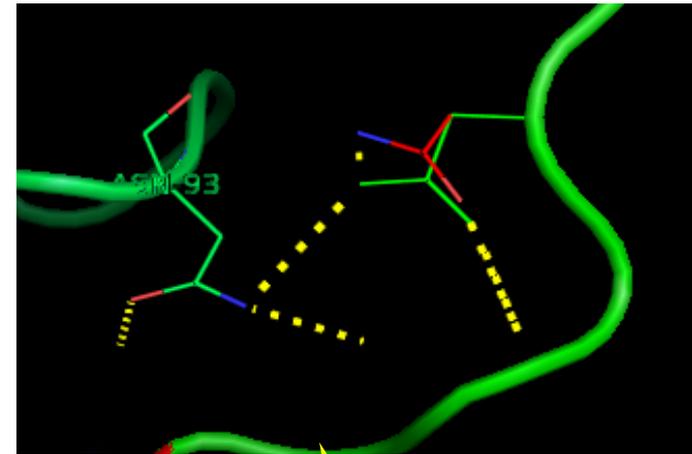

Swiss Model
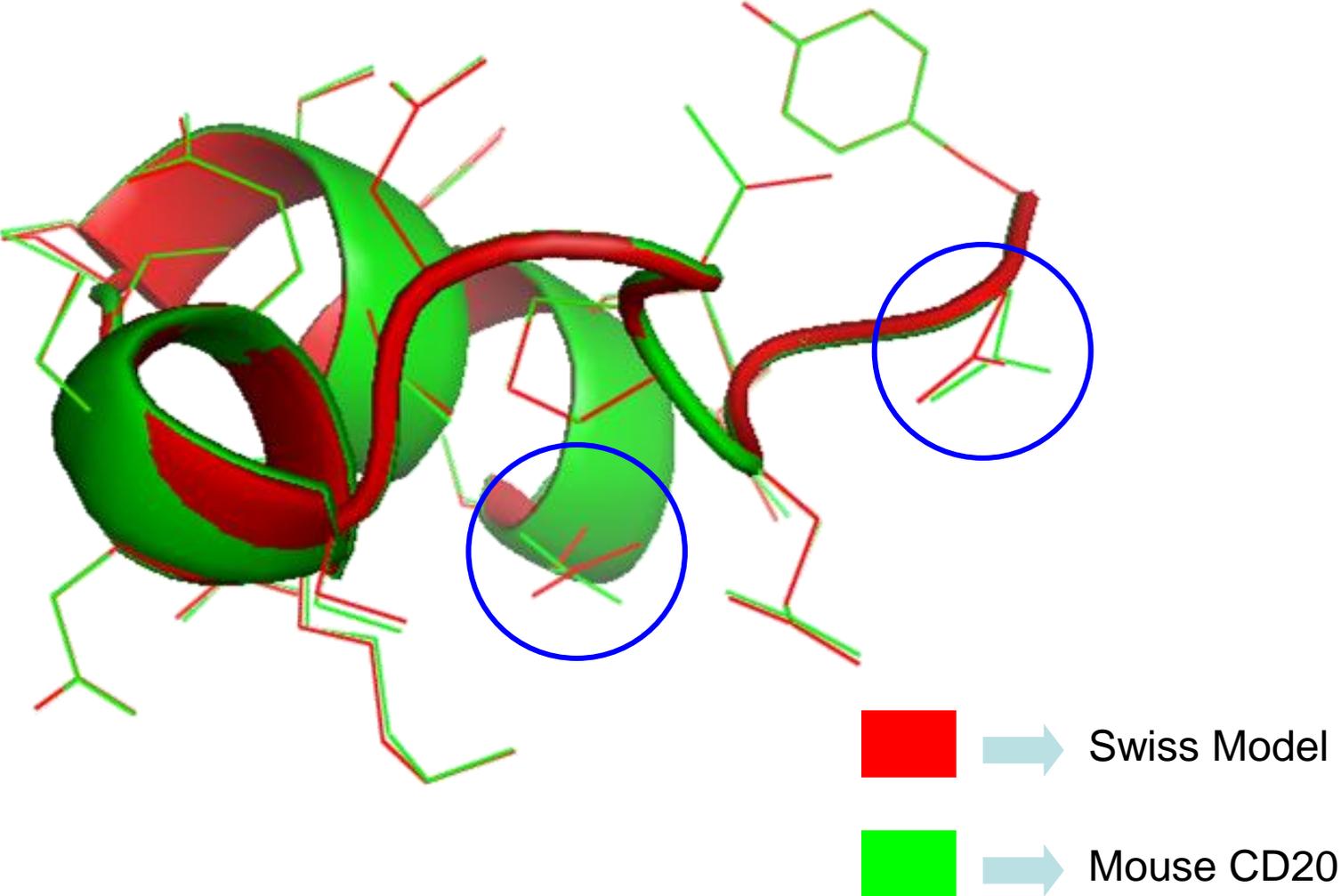
Mouse CD20

Alignment of Predicted Mouse CD20 to Human CD20
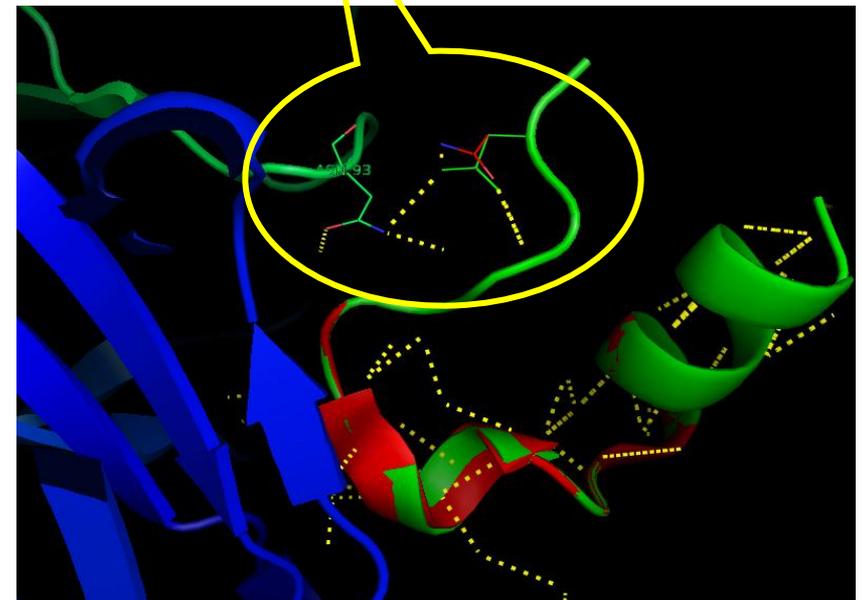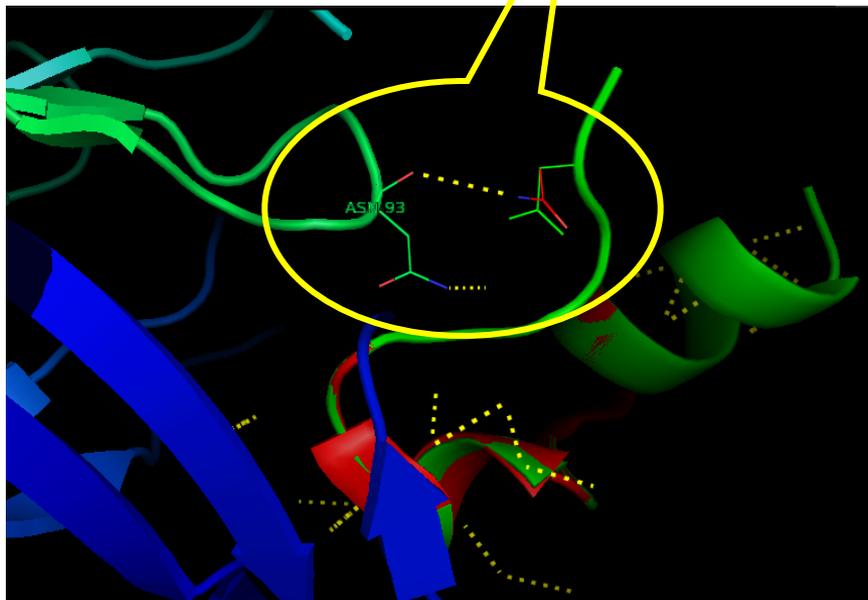

Human CD20

Phyre
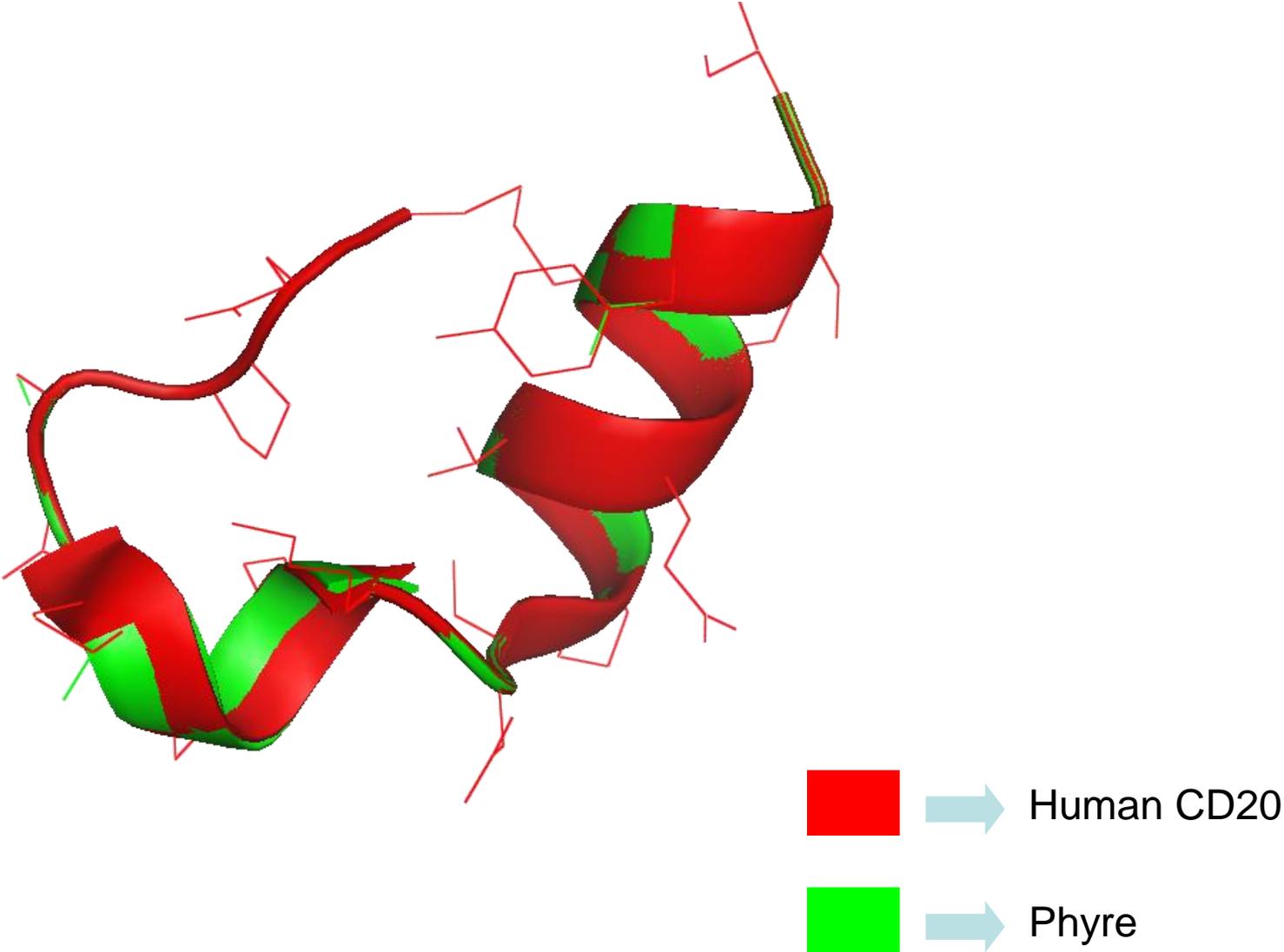
Binding form of Human and Mouse CD20 residue 166 to ASN193
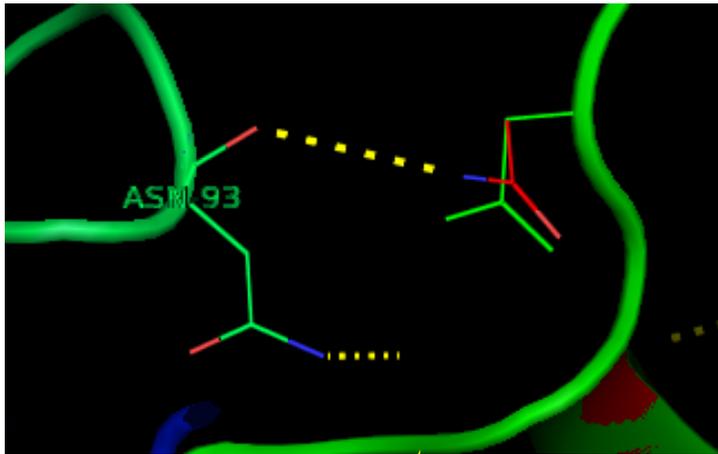
Human-Asn

Mouse-Asp

Discussion and Conclusion

Mouse can be used as a model organism to detect the mechanism of Rituximab

Differences between PyMOL and Swiss-PDB Viewer

PyMOL	Swiss-PDB Viewer
Beautiful Image	Protein Modeling
Movie Maker	Integrated with Online Databases
Command Line	

博学之，审问之，慎思之，
明辨之，笃行之

——《中庸》

道不行，乘桴浮于海

——《论语》

靡不有初，鲜克有终

——《诗经》

Acknowledgement

Yong Liu, Shuaipeng Ma in protein structure prediction

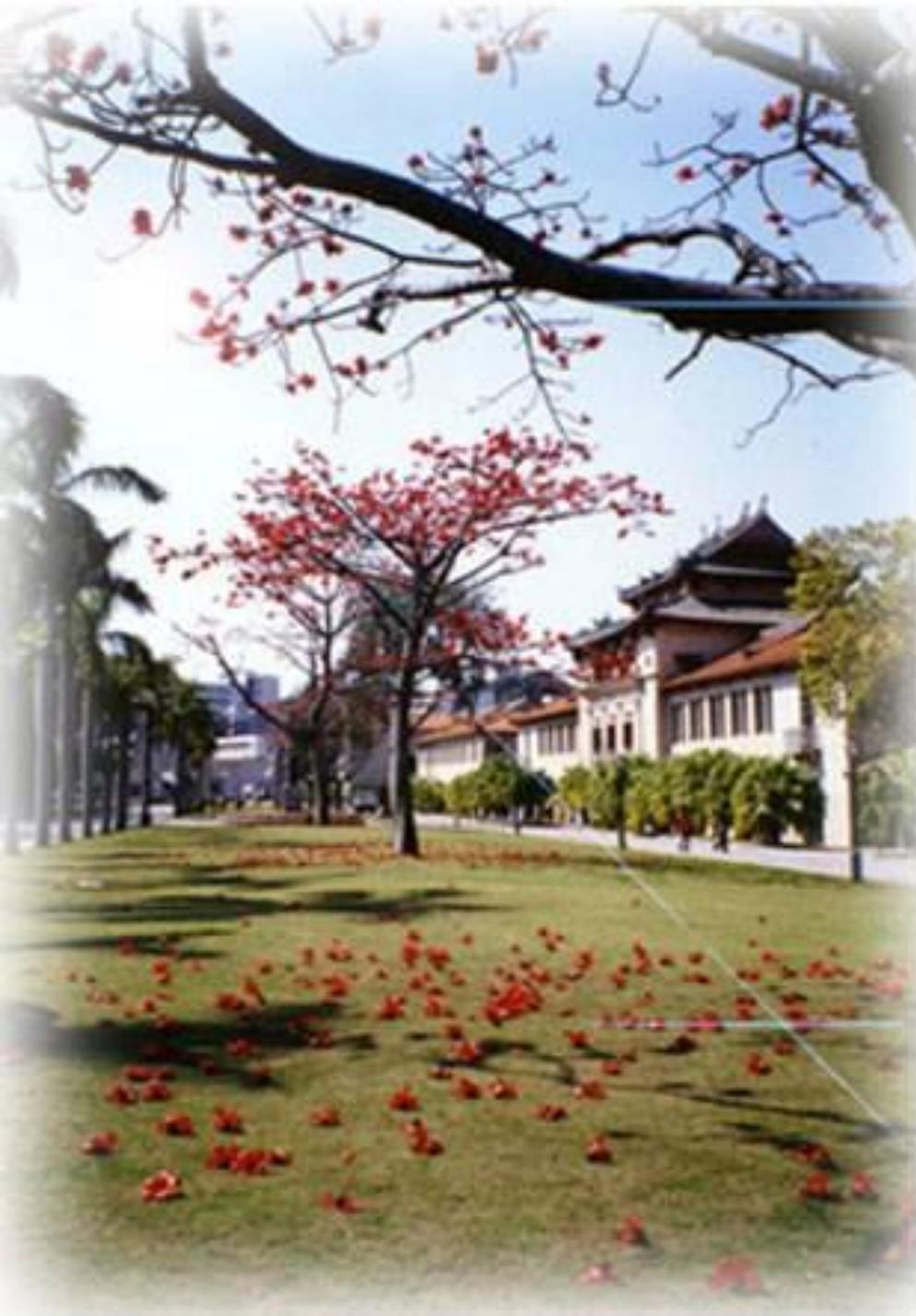Yue Huang, Zhen Liu and Jieru Li in Phylogenetic Tree

My Team Member: Hua Gao, Yang Ding and Weizhen Zhou

Teaching Assistantship: Yi Zhao

All my classmates in Class 1

Especially For——

Prof. Luo

Thank you !