Bioinformatic Analysis of CD38

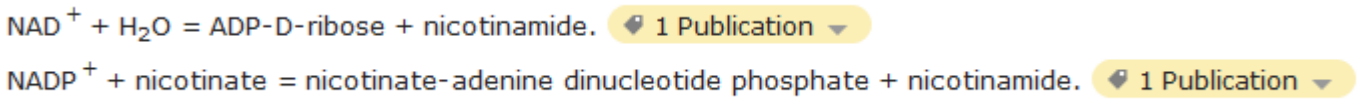
CD38的生物信息学分析

成员： G03：何卫明、刘思宇、陆正昌、潘治翔
G17：陈娅琦、马小燕、李惟芬
报告人：何卫明

目录

一、CD38简介

- CD38：定位于膜上的糖蛋白，催化环腺苷二磷酸核糖（cADPR, cyclic ADP-ribose）的合成和降解。



- cADPR是核苷酸的代谢产物，通过作用于ryanodine（RyRs）等受体参与细胞内钙库的钙动员。

- 多发性骨髓瘤细胞：表面高度表达CD38

- daratumumab：anti-CD38的抗癌单抗，用于治疗多发性骨髓瘤

解析全球第一个anti-CD38的抗癌单抗daratumumab
http://www.chinairn.com/news/20151120/140912808.shtml

二、表达分析

- RNA水平： GTEx-RNAseq

- 从570个捐赠者不同的53个组织中获得的8555个样本蛋白，
- 用高通量技术对它们相同量的转录组(mRNA，smallRNA，and NONcodingRNA, pseudogene）进行测序，用中位数反映它们在不同组织中的表达水平。

结果：超链接资料\UCSC Genome Browser GTEx Track Settings.html

- RNA水平和蛋白质水平：The Human Protein Atlas project

简称HPA数据库，它致力于提供全部24,000种人类蛋白质的组织和细胞分布信息。

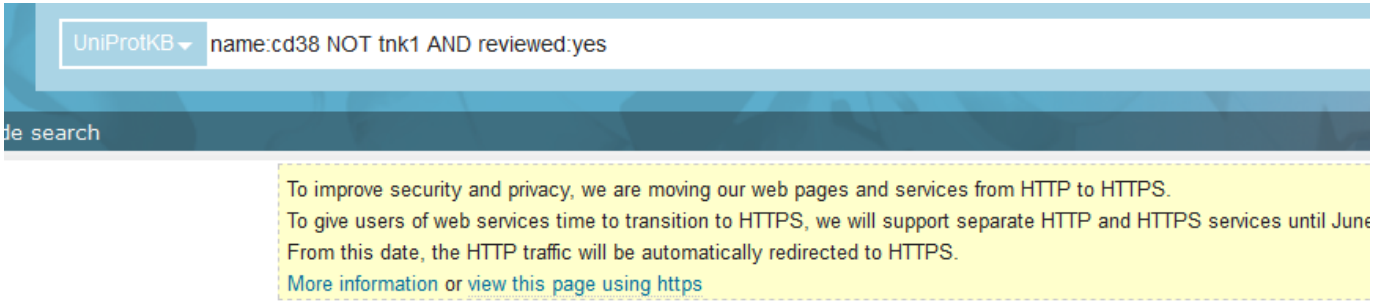
使用特制的抗体，运用免疫组化的技术。

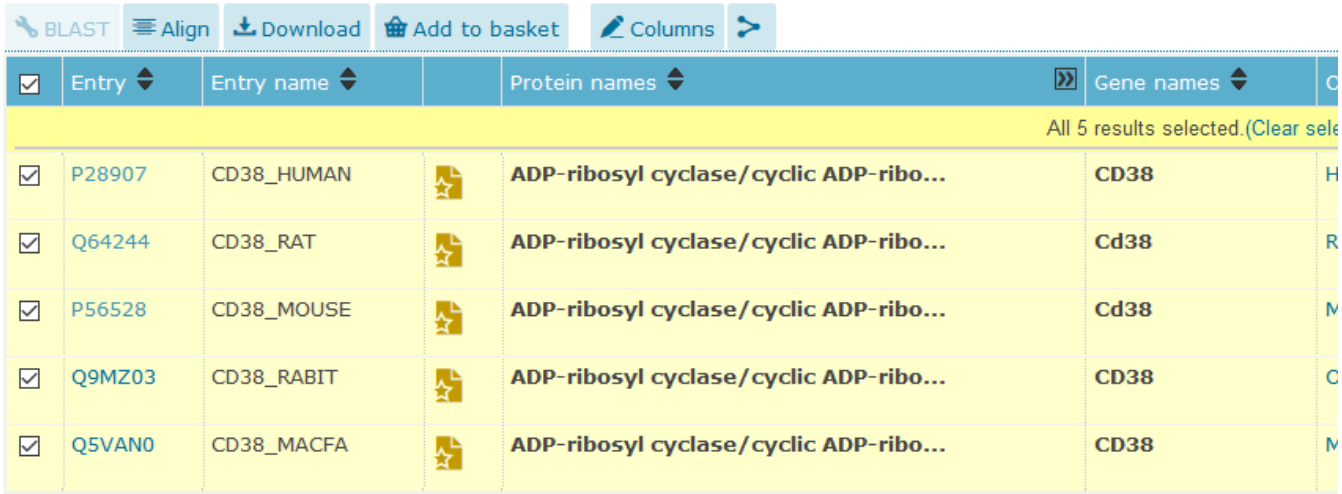
结果：超链接资料\Search cd38 - The Human Protein Atlas.html
Cell : 超链接资料\Cell atlas - CD38 - The Human Protein Atlas.html
Tissue: 超链接资料\Tissue expression of CD38 - Summary - The Human Protein Atlas.html
Pathology : 超链接资料\Expression of CD38 in cancer - Summary - The Human Protein Atlas.html

超链接资料

三、利用UniProt对蛋白质做综合分析

（1）过程：





（2）结果：超链接资料\Alignment [completed].html

超链接资料


P28907 CD38_HUMAN 1 MANCEFSPVSGDKPCCRLSRRAQLCLGVSILVL-ILVVVLAV--VVPRWRQQWSGPGTTK 57
Q64244 CD38_RAT 1 MANYEFSQVSEDRPGCRLTRKAQIGLGVGLLLLVALVVV-VVIVLWPRSPLVWKGKPTTK 59
P56528 CD38_MOUSE 1 MANYEFSQVSGDRPGCRLSRKAQIGLGVGLLVLIALVVGIVVILLRPRSLLVWTGEPTTK 60
Q9MZ03 CD38_RABIT 1 MPDYEFSPASGDRPRSWISKQVLIVLGVCLPVILALAIWVGV-LTWRQS-----SMGATD 54
Q5VAN0 CD38_MACFA 1 MANCEFSPVSGDKPCCRLSRRAQVCLGVCLLVLLILVVVVAV--VLPRWRQQWSGSGTTS 58
P28907 CD38_HUMAN 58 RFPETVLARCVKYTE-IHPEMRHVDCQSVWDAFKGAFISKHPCNITEEDYQPLMKLGTQT 116
Q64244 CD38_RAT 60 HFADIILGRCLIYTQILRPEMRDQDCKKILSTFKRGFISKNPCNITNEDYAPLVKLVTQT 119
P56528 CD38_MOUSE 61 HFSDIFLGRCLIYTQILRPEMRDQNCQEILSTFKGAFVSKNPCNITREDYAPLVKLVTQT 120
Q9MZ03 CD38_RABIT 55 HVSAIVLGRCLTYTRNMHPELRNQDCKKILNTFTSAFVSKDPCNITKEDYQPLIDLVTQT 114
Q5VAN0 CD38_MACFA 59 RFPETVLARCVKYTE-VHPEMRHVDCQSVWDAFKGAFISKYPCNITEEDYQPLVKLGTQT 117
P28907 CD38_HUMAN 117 VPCNKILLWSRIKDLAHQFTQVQRDMFTLEDTLLGYLADDLTWCGEFNTSKINYQSCPDW 176
Q64244 CD38_RAT 120 IPCNKTLFWSKSKHLAHQYTWIQGKMFTLEDTLLGYIADDLRWCGDPSTSDMNYDSCPHW 179
P56528 CD38_MOUSE 121 IPCNKTLFWSKSKHLAHQYTWIQGKMFTLEDTLLGYIADDLRWCGDPSTSDMNYVSCPHW 180
Q9MZ03 CD38_RABIT 115 VPCNKTLFWSRSKELAHQYSGIQKEMFTLEDTLLGYIADNLVWCGDPRTSEVKEEFCPYR 174
Q5VAN0 CD38_MACFA 118 VPCNKTLLWSRIKDLAHQFTQVQRDMFTLEDMLLGYLADDLTWCGEFNTFEINYQSCPDW 177
P28907 CD38_HUMAN 177 RKDCSNNPVSVFWKTVSRRFAEAACDVVHVMLNGSRSKIFDKNSTFGSVEVHNLQPEKVQ 236
Q64244 CD38_RAT 180 SENCPNNPVAVFWNVISQKFAEDACGVVQVMLNGSLSEPFYRNSTFGSVEVFNLDPNKVH 239
P56528 CD38_MOUSE 181 SENCPNNPITVFWKVISQKFAEDACGVVQVMLNGSLREPFYKNSTFGSVEVFSLDPNKVH 240
Q9MZ03 CD38_RABIT 175 NENCSSTATSVFWTVVSQKFAESACGTVYVMLNGSRTTAFSKASTFGSVEVFNLHPDRVH 234
Q5VAN0 CD38_MACFA 178 RKDCSNNPVSVFWKTVSRRFAETACGVVHVMLNGSRSKIFDKNSTFGSVEVHNLQPEKVQ 237
P28907 CD38_HUMAN 237 TLEAWVIHGGREDSRDLCQDPTIKELESIISKRNIQFSCKNIYRPDKFLQCVKNPEDSSC 296
Q64244 CD38_RAT 240 KLQAWVMHDIKGTSSNACSSPSINELKSIVNKRNMIFACQDNYRPVRFLQCVKNPEHPSC 299
P56528 CD38_MOUSE 241 KLQAWVMHDIEGASSNACSSSSLNELKMIVQKRNMIFACVDNYRPARFLQCVKNPEHPSC 300
Q9MZ03 CD38_RABIT 235 TLHAWVMHDIGGVERDSCLGSSIKELKSIVNQRNISFFCQDDYRPARFVQCVRHPEHPSC 294
Q5VAN0 CD38_MACFA 238 ALEAWVIHGGREDSRDLCQDPTIKELESIISKRNIRFFCKNIYRPDKFLQCVKNPEDSSC 297
P28907 CD38_HUMAN 297 TSEI 300
Q64244 CD38_RAT 300 RLNV 303
P56528 CD38_MOUSE 301 RLNT 304
Q9MZ03 CD38_RABIT 295 SVLM 298
Q5VAN0 CD38_MACFA 298 LSGI 301

四、蛋白质二级结构预测

预测蛋白质的二级结构主要基于如下两个方面：

- 一是如何有效地获取蛋白质的特征信息。

许多二级结构预测都是基于蛋白质的氨基酸组成特征信息构建的。

- 二是采用何种预测算法以及如何利用这些算法构建模式识别分类器。

目前,有许多算法都被应用于蛋白质的二级结构预测中。如神经网络、马尔科夫模型、支持向量机、K最近邻、粗糙集、模糊聚类分析、信息偏差、贝叶斯分类算法等。

基于蛋白质结构数据的二级结构计算

DSSP(hydrogen bond estimation algorithm)是用于对蛋白质结构中的氨基酸残基进行二级结构构像分类的标准化算法，使用PDB格式的原子级分辨率的蛋白质三维结构坐标集数据，从而得到每个氨基酸残基的二级结构构象参数。

DSSP ：基于氢键
STRIDE：基于氢键及二面角统计分布

八种二级结构

- H = α-helix，α-螺旋
- B = residue in isolated β-bridge，β-桥
- E = extended strand, participates in β ladder，折叠
- G = 3-helix (310 helix)，螺旋
- I = 5 helix (π-helix)，螺旋-5
- T = hydrogen bonded turn，转角
- S = bend ，卷曲
- L= coil，环

- 主要的蛋白质二级结构预测

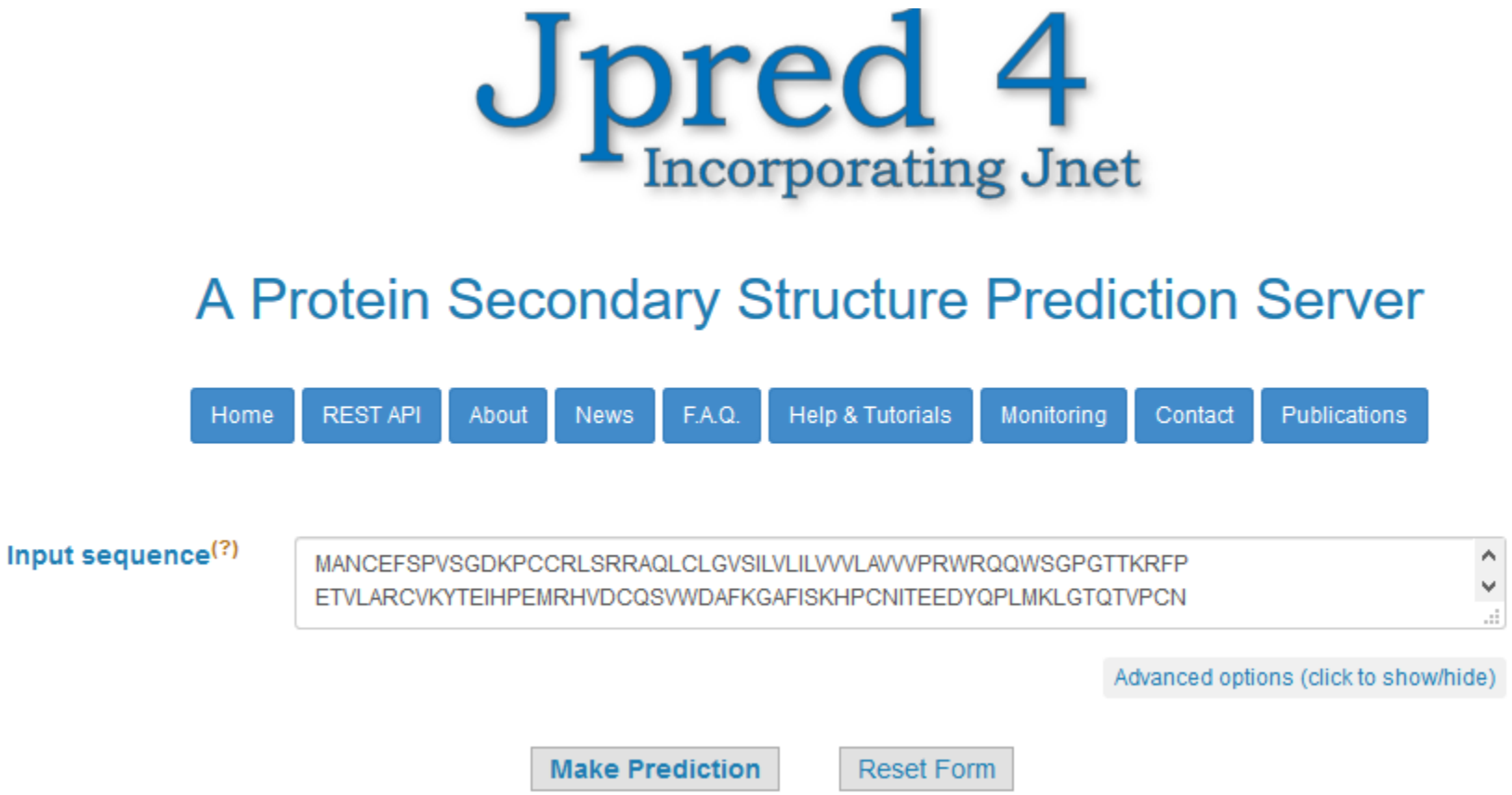
三类二级结构预测：螺旋、折叠和卷曲

- 代表的算法有PHD、PSIPRED和JPred等，而且有相关很成熟的免费服务器供科研工作者使用。

JPred

现在已经升级到JPred(version4):
其包含Jnet 算法——使用神经网络算法预测蛋白质二级结构，平均预测准确率>82%。

预测：CD38_HUMAN（P28907）



http://www.compbio.dundee.ac.uk/jpred/

（2）结果首先显示可以在PDB数据库中匹配的已经有实验结果

Match found in PDB

The sequence you submitted is similar to those with known structure. These may provide a more accurate secondary structure assignment than a JPred prediction.

If you still want to carry out a Jpred prediction click continue

Hits found

Show 25 entries

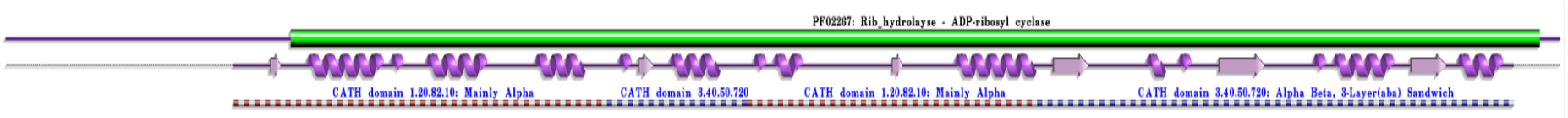
PDB	Chain	Description	Blast E-value
2ef1	B	ADP-ribosyl cyclase 1	e-154
2ef1	A	ADP-ribosyl cyclase 1	e-154
4cmh	A	ADP-RIBOSYL CYCLASE 1	e-152
1zvm	D	ADP-ribosyl cyclase 1	e-152
1zvm	C	ADP-ribosyl cyclase 1	e-152
1zvm	B	ADP-ribosyl cyclase 1	e-152
1zvm	A	ADP-ribosyl cyclase 1	e-152
5f1k	B	ADP-ribosyl cyclase/cyclic ADP-ri...	e-152
5f1k	A	ADP-ribosyl cyclase/cyclic ADP-ri...	e-152

（3）如果还继续做预测，得到如下结果。

```
MANCEFSPVSGDKPCCRLSRRAQLCLGVSILVLILVVVLAVVVPRWRQQWSGPGTTKRFPETVLARCVKYTEIHPEMRHVDCQSVWDAFKGAFISKHPCNITEEDYQPLMKLGTQTVPC
-----------------HHHHHHHHHHHHHHHHHHHHHHHHH--------------HHHHHHHHHHH-HHEEEE-------HHHHHHHHHHHHH--------HHHHHHHHHH-------

IWCGEFNTSKINYQSCPDWRKDCSNNPVSVFWKTVSRRFAEAACDVVHVMLNGSRSKIFDKNSTFGSVEVHNLQPEKVQTLEAWVIHGGREDSRDLCQDPTIKELESIISKRNIQ
-------------------------HHHHHHHHHHHHHHHHH--EEEEEE----------------EEE----------EEEEEEEE------------HHHHHHHHHHHH---E
```

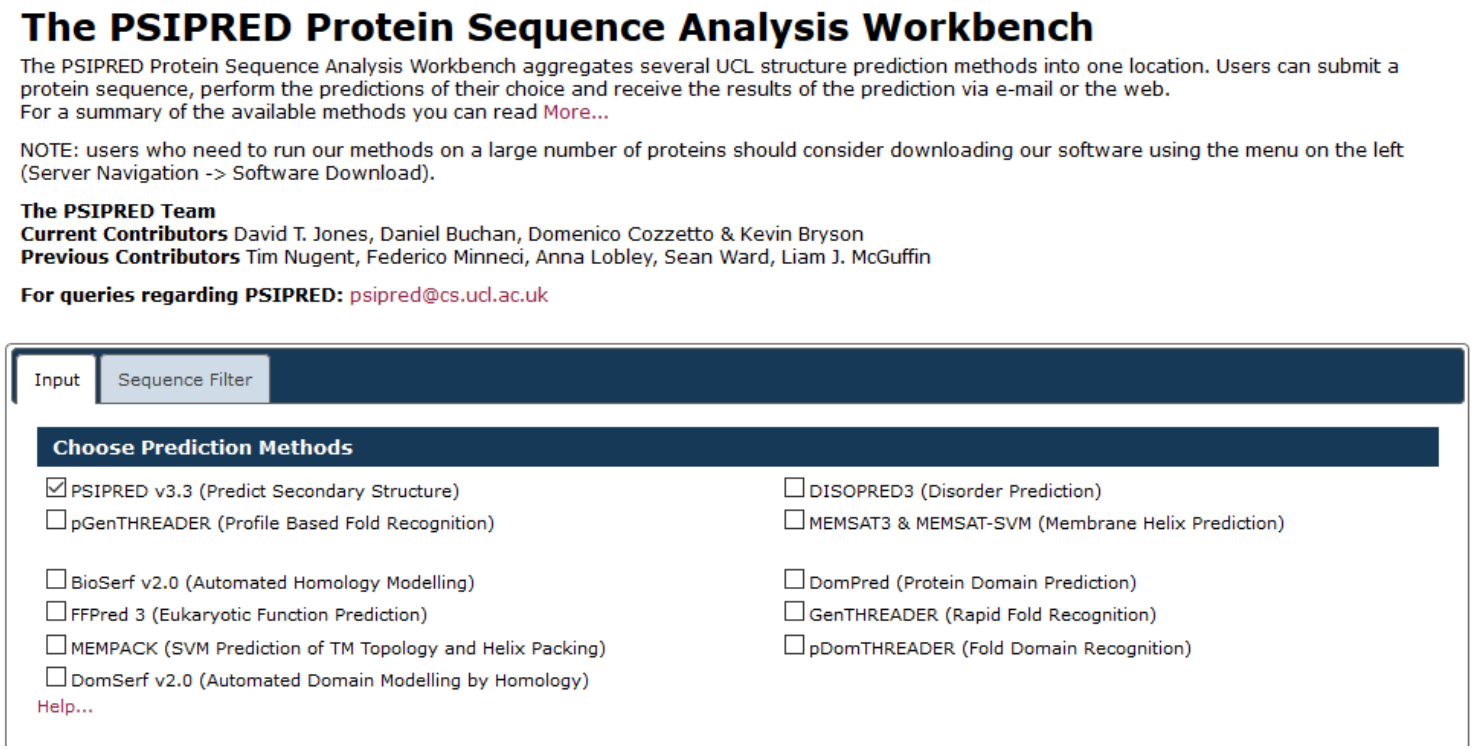
PDB id: 2ef1



PSIPRED

两阶段神经网络，使用PSI-BLAST输出来预测二级结构。

PSIPRED V3.0目前对二级结构的预测最高的Q3为81.4%（±0.6%）。



http://bioinf.cs.ucl.ac.uk/psipred/

Secondary Structure Map

Feature predictions are colour coded onto the sequence according to the sequence feature key shown below.

KEY	Helix	Sheet	Disordered	Disordered protein binding	Dompred Boundary	DomSSEA Boundary
Annotations	H	E	D	D	A	S

Sequence Resubmission

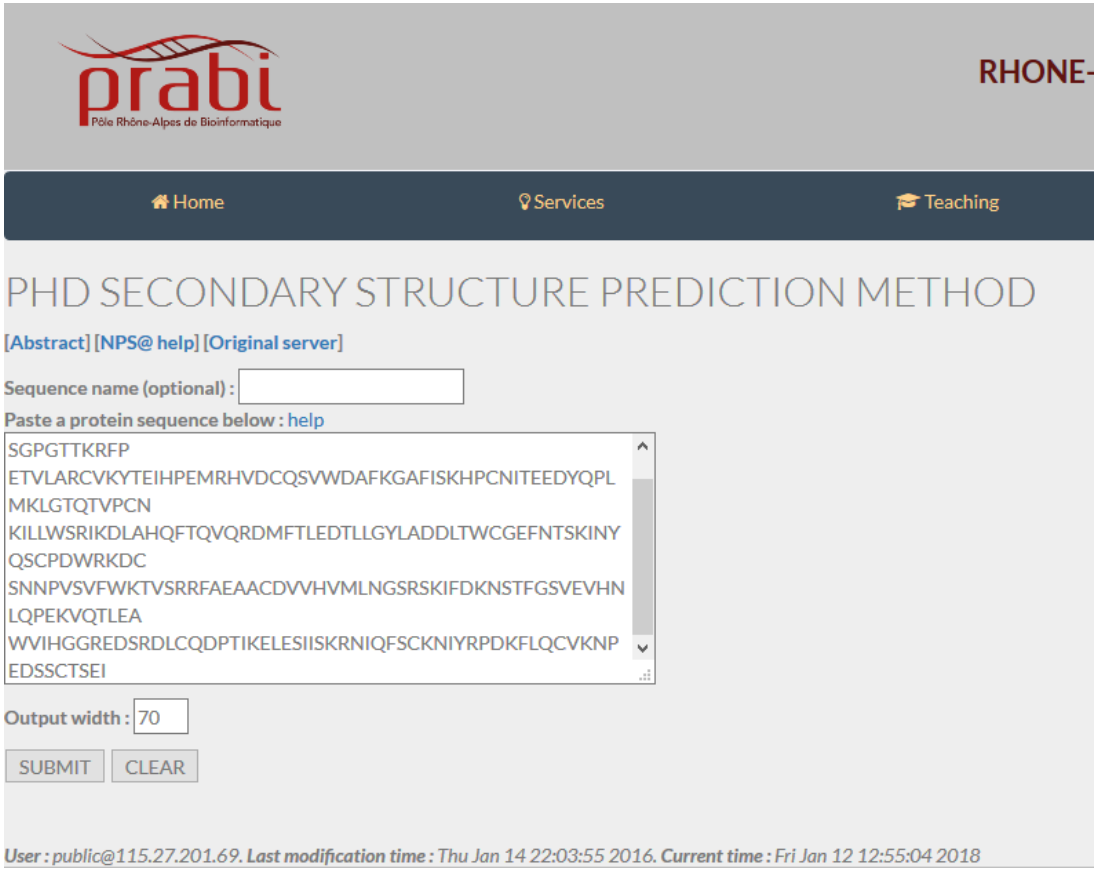
Start: 1

Stop: 300

Select Methods

- PHD

· 神经网络系统（“序列到结构水平”和“结构到结构水平”）
· 通过氢键预测。

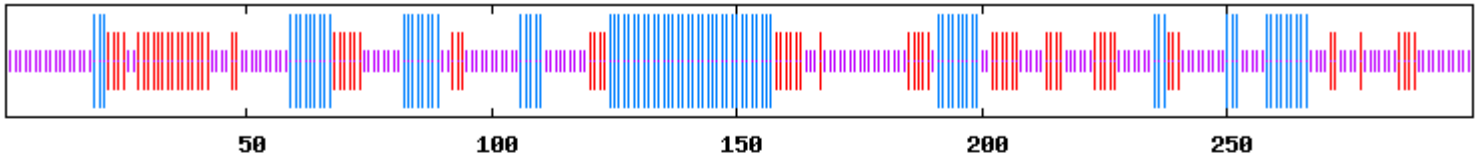


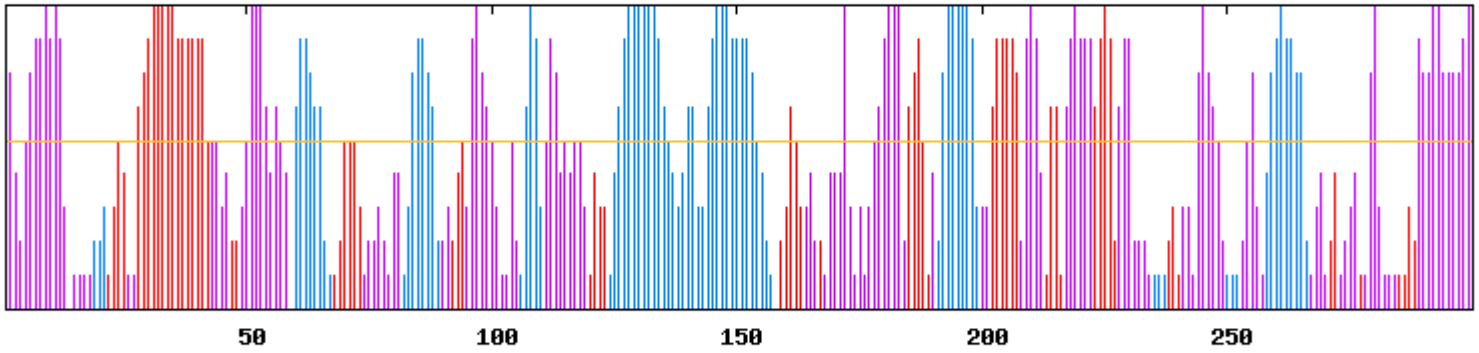
https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MANCEFSPVSGDKPCCRLSRRAQLCLGVSILVLILVVVLAVVVPRWRQQWSGPGTTKRFPETVLARCVKY
CCCccCCCCCCCCcccccchhheeEeccEEEEEEEEEEEEEECCcceecCCCCCcCCchHHHHHHhhee
TEIHPEMRHVDCQSVWDAFKGAFISKHPCNITEEDYQPLMKLGTQTVPCNKILLWSRIKDLAHQFTQVQR
EEEeccccccchhHHHHHhcceeEcCCCCCcccCchHHHhCCCcCcCCceeeehhHHHHHHHHHHHHhh
DMFTLEDTLLGYLADDLTWCGEFNTSKINYQSCPDWRKDCSNNPVSVFWKTVSRRFAEAACDVVHVMLNG
HHhhHHHHHHHHHHHhhheeeEEecccecccCccccCCCCCcEEEEechHHHHHHHhccEEEEEEcC
SRSKIFDKNSTFGSVEVHNLQPEKVQTLEAWVIHGGREDSRDLCQDPTIKELESIISKRNIQFSCKNIYR
CCceEEeCCCCCCEEEEeCCCcccchhheeecccCCCCCchhhcCCcchHHHHHHHhccccceeccccecC
PDKFLQCVKNPEDSSCTSEI
CcccceeeeCCCCCCCCCC
```

```
Sequence length :   300
```

```
PHD :
   Alpha helix     (Hh) :    83 is  27.67%
   3₁₀ helix       (Gg) :     0 is   0.00%
   Pi helix        (Ii) :     0 is   0.00%
   Beta bridge     (Bb) :     0 is   0.00%
   Extended strand (Ee) :    71 is  23.67%
   Beta turn       (Tt) :     0 is   0.00%
   Bend region     (Ss) :     0 is   0.00%
   Random coil     (Cc) :   146 is  48.67%
   Ambiguous states (?)  :     0 is   0.00%
   Other states         :     0 is   0.00%
```





Residues with a scale reliability index of prediction of 5 and over (uppercase letters) are predicted at better than 82%.

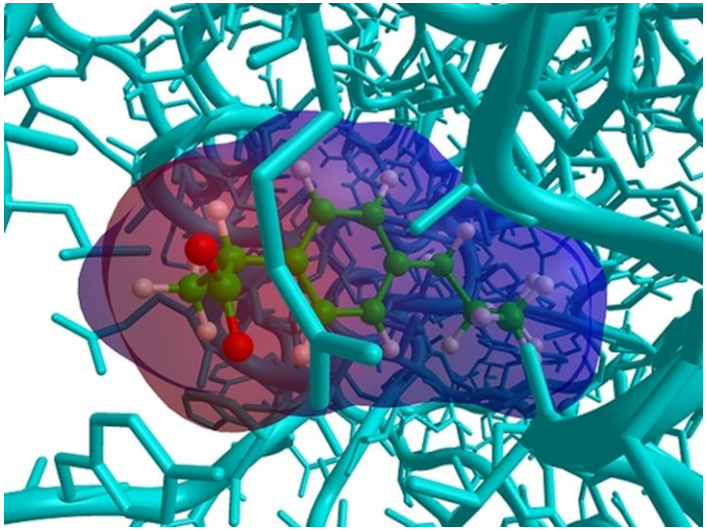
Prediction result file (text): [PHD]
Intermediate result file (text): [BLASTP on NRPROT]

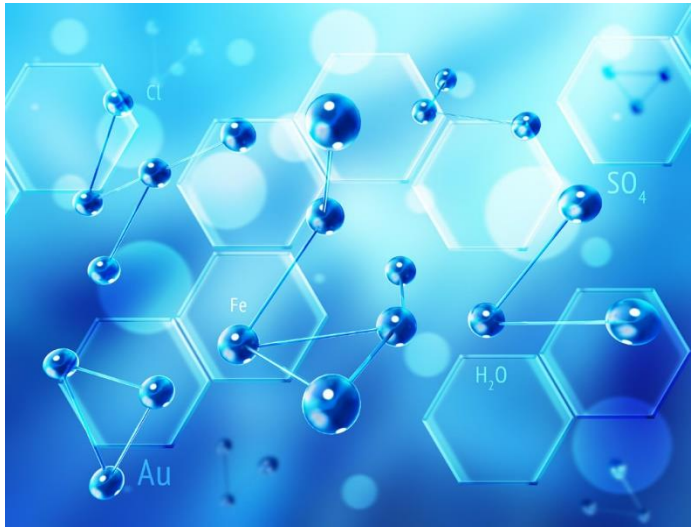
五、分子对接

定义

分子对接是通过研究配体小分子和受体生物大分子的相互作用，预测其亲和力，实现基于结构的药物设计的一种重要方法。

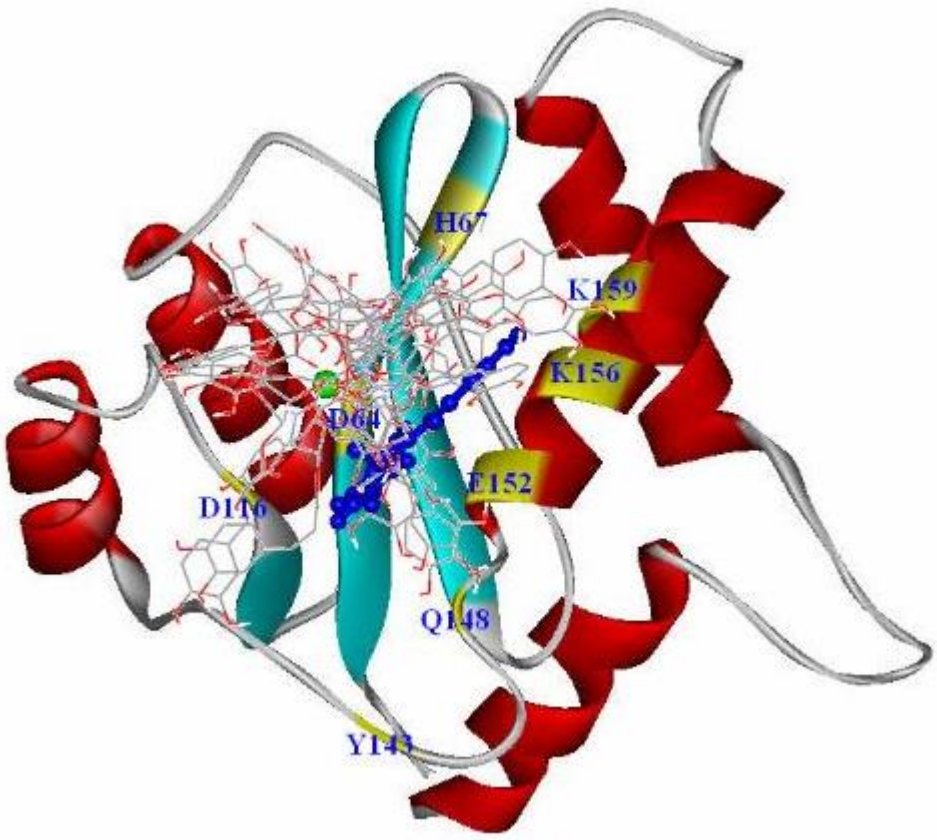
整体上考虑配体与受体的结合效果，较好的避免局部作用、整体结合欠佳的情况。





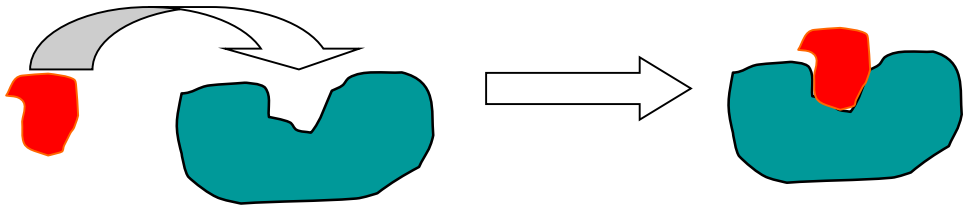
基本原理

- 配体与受体结合时，彼此存在静电相互作用、氢键相互作用、范德华力相互作用和疏水作用力。
- 配体与受体结合必须满足互相匹配原则，即配体与受体几何形状互补匹配、静电相互作用互补匹配、氢键相互作用互补匹配、疏水相互作用互补匹配。



经典模型

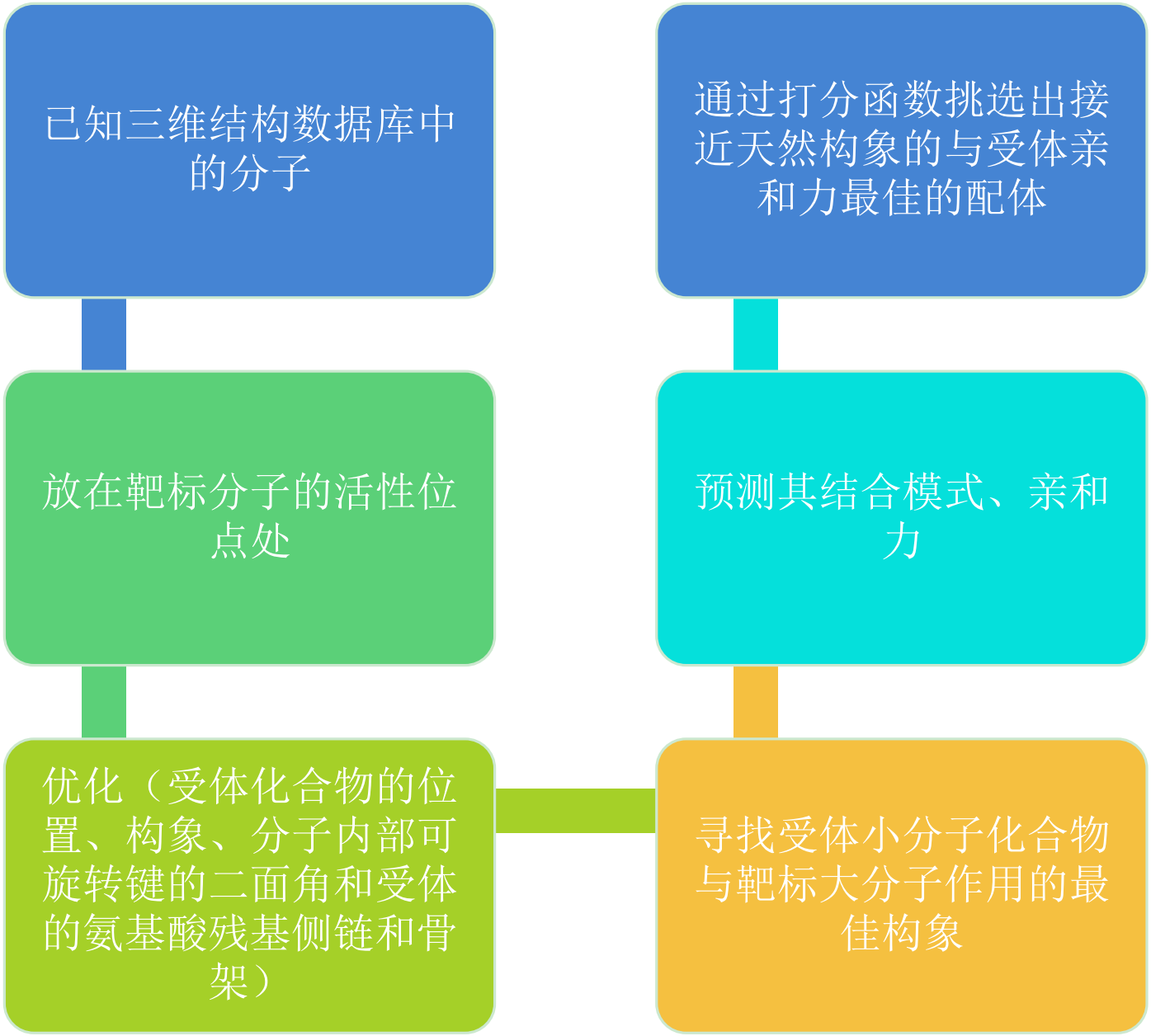
分子对接的最初思想起源于Fisher E提出的“锁和钥匙模型”，即受体与配体的相互识别首要条件是空间结构的匹配。



配体　　受体　　复合物

受体－配体的锁和钥匙模型

主要步骤

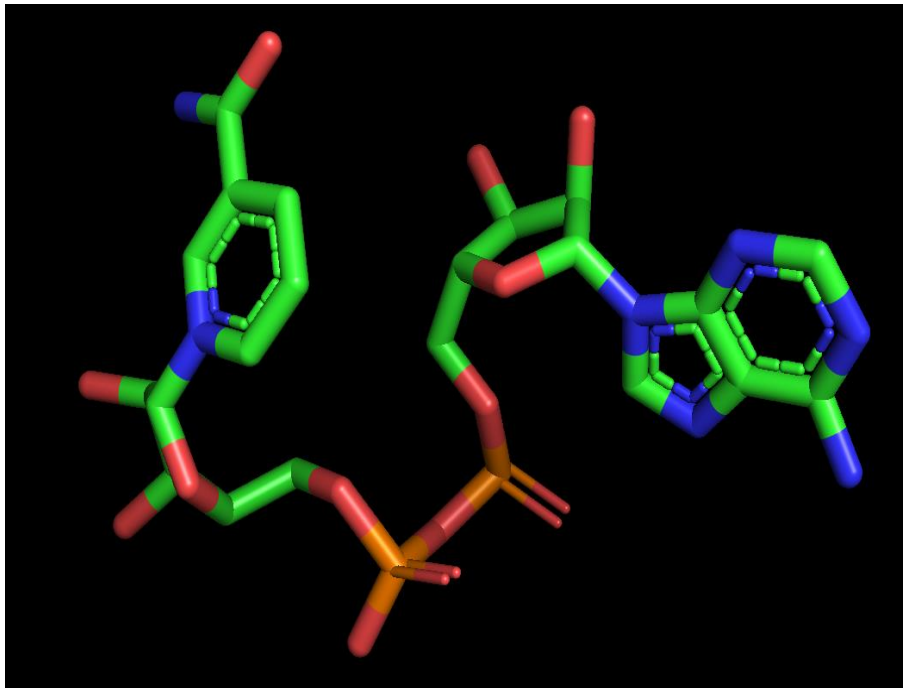

已知三维结构数据库中的分子
放在靶标分子的活性位点处
优化（受体化合物的位置、构象、分子内部可旋转键的二面角和受体的氨基酸残基侧链和骨架）
寻找受体小分子化合物与靶标大分子作用的最佳构象
预测其结合模式、亲和力
通过打分函数挑选出接近天然构象的与受体亲和力最佳的配体

代表性软件

名称	构象搜索方法	结合评价方法	速度
Flex X (Sybyl)	片段生长法	半经验自由能	快
LigandFit(Cerius2)	蒙地卡罗模拟	半经验自由能	快
Glide (薛定谔软件)	系统搜索	半经验自由能	一般
Gold	遗传算法	半经验自由能	快
Affinity（InsightII)	蒙地卡罗/MM/MD	分子力场	慢
AutoDock	遗传算法	半经验自由能	一般
Dock	片段生长法	分子力场	快
ICM-Dock	随机全局优化	半经验自由能	快
Fred (openeye)	系统搜索	半经验自由能	快

以CD38为例子

CD 38



NAD

软件：Autodock Vina程序

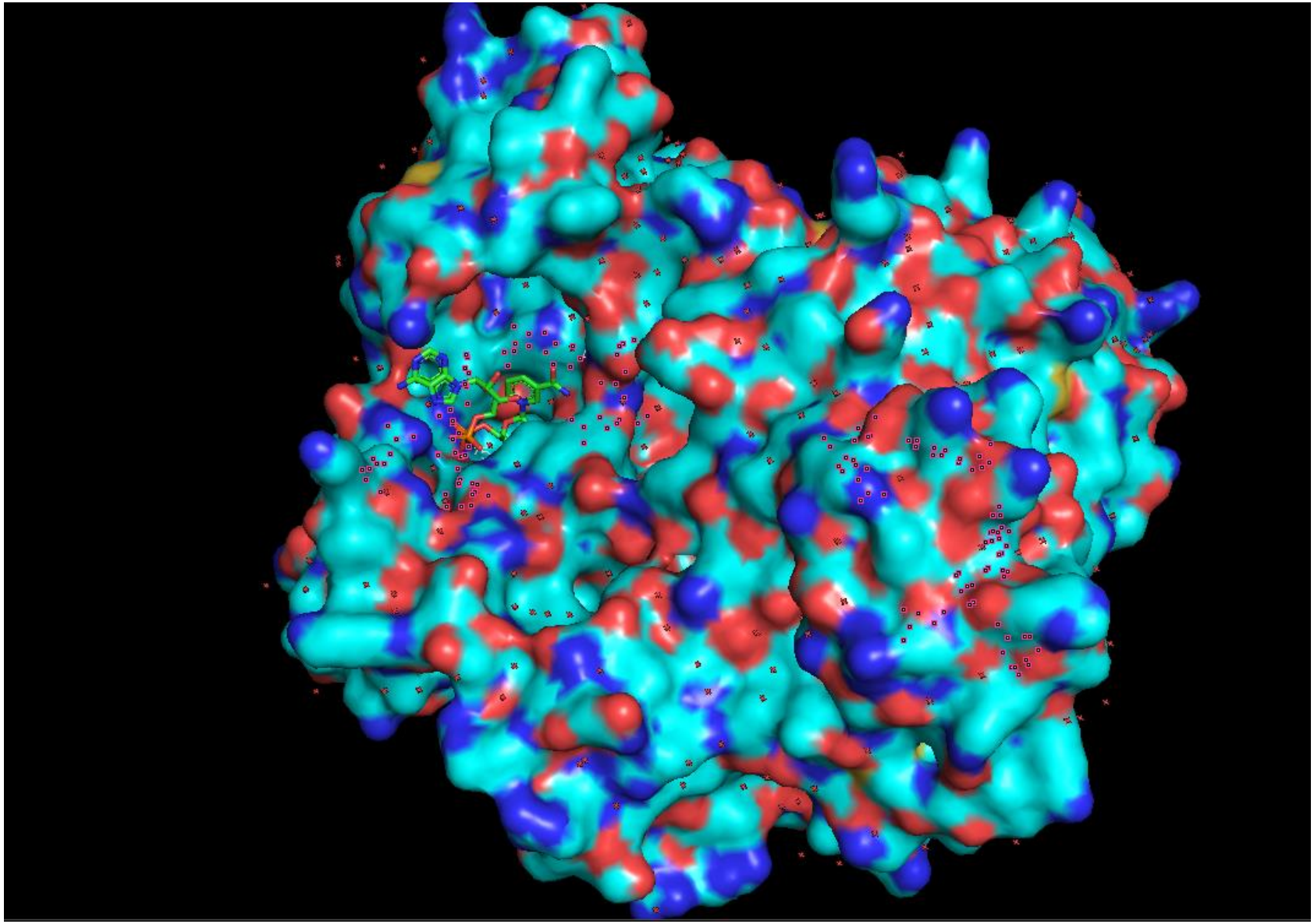
a) 使用 MGLTools 准备蛋白质和小分子文件 （蛋白质文件要重新计算原子部分电荷），以小分子位置为对接中心， 盒子大小定为 15*15*15埃。

b) 使用 Vina进行对接， 输出可能的蛋白质-小分子结合构象。

c) 利用 PYMOL 显示对接结果， 将对接后的小分子构象与原始的小分子构象进行比较， 与蛋白质在一起作图显示。

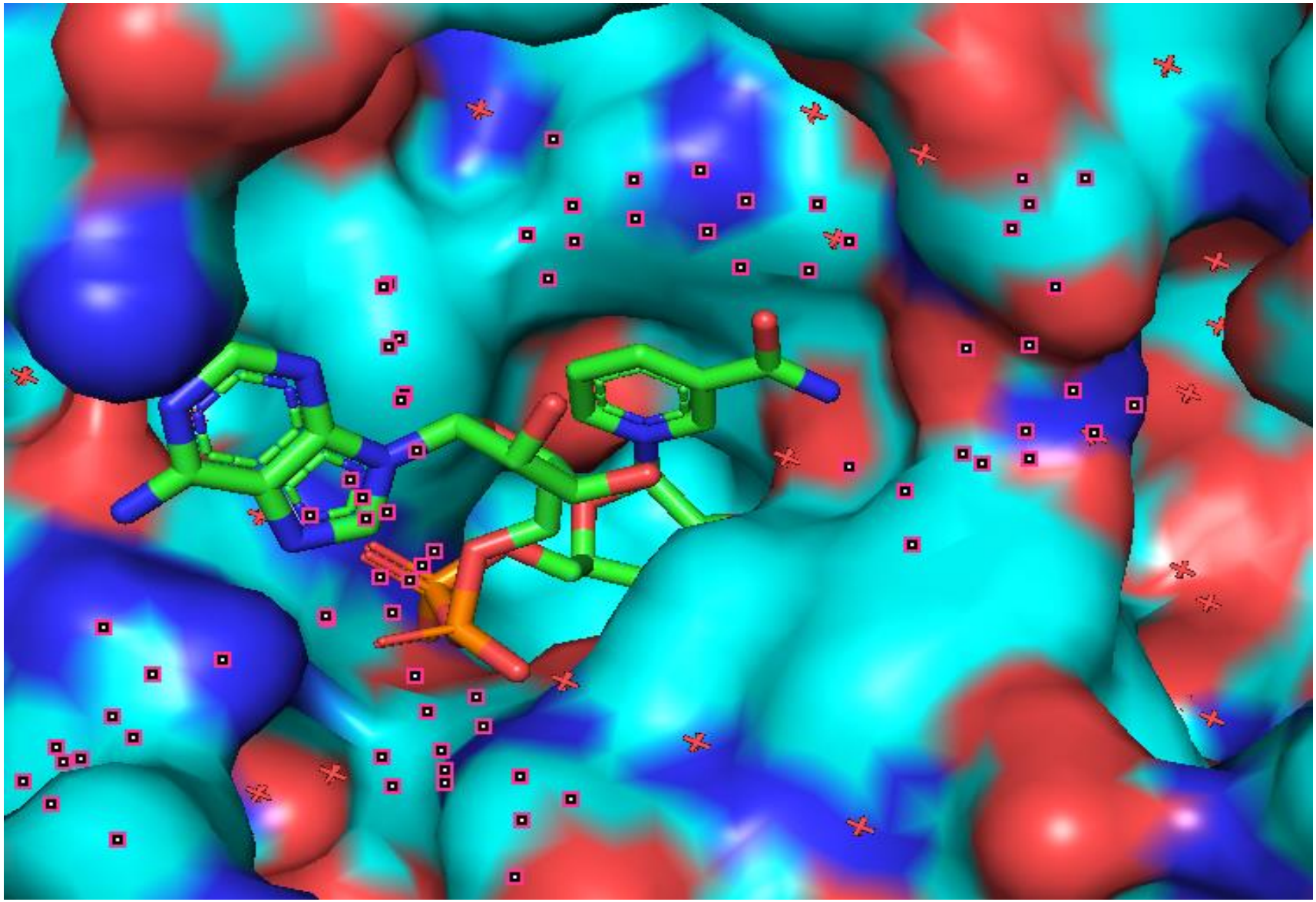
Qun Liu et al. Structural Basis for Enzymatic Evolution from a Dedicated ADP-ribosyl Cyclase to a Multifunctional NAD Hydrolase*. October 2, 2009. The Journal of Biological Chemistry. 284, 27637-27645.

结果分析



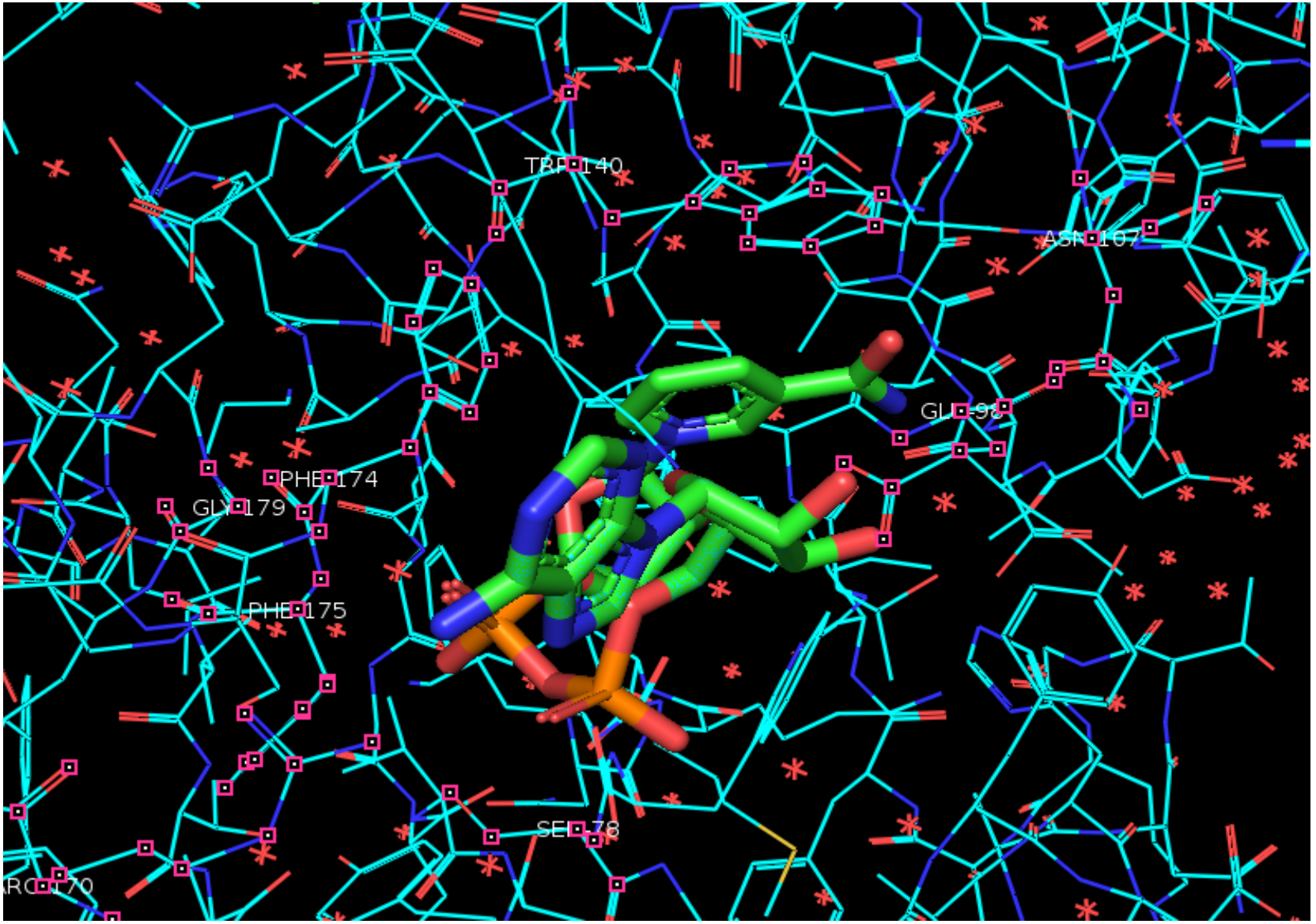
Qun Liu et al. Structural Basis for Enzymatic Evolution from a Dedicated ADP-ribosyl Cyclase to a Multifunctional NAD Hydrolase*. October 2, 2009. The Journal of Biological Chemistry. 284, 27637-27645.

结果分析



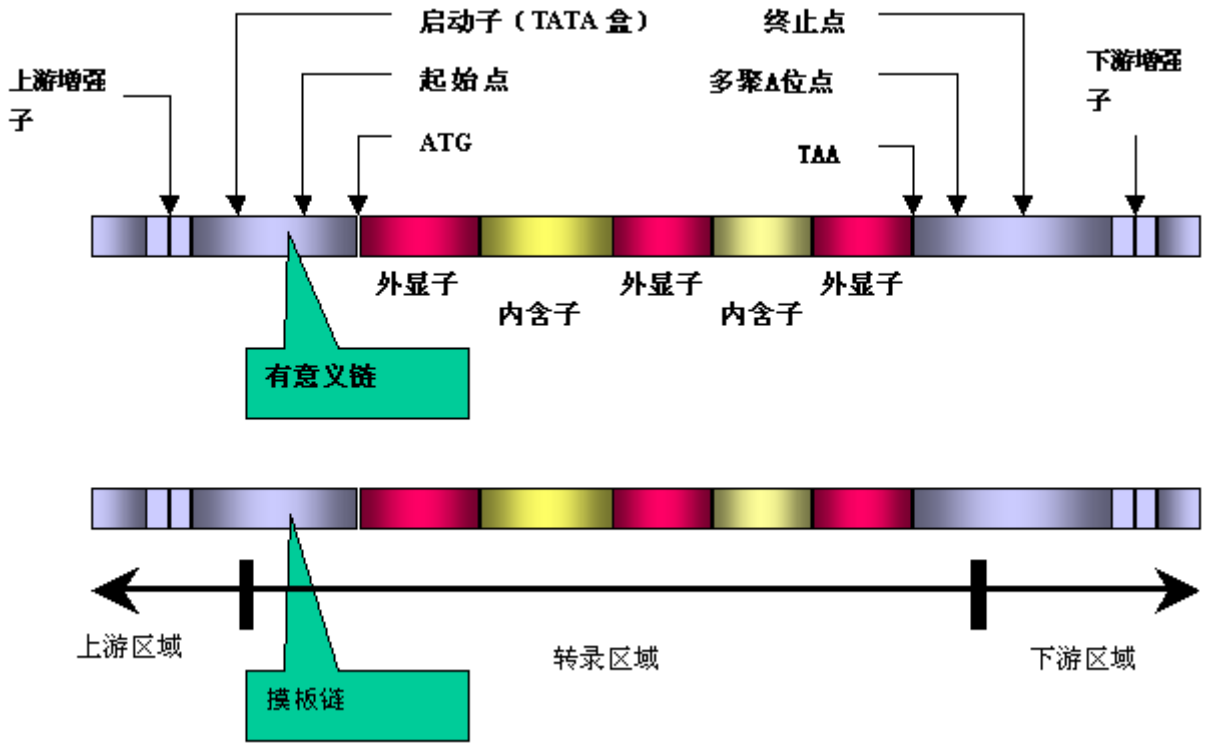
Qun Liu et al. Structural Basis for Enzymatic Evolution from a Dedicated ADP-ribosyl Cyclase to a Multifunctional NAD Hydrolase*. October 2, 2009. The Journal of Biological Chemistry. 284, 27637-27645.

结果分析



Qun Liu et al. Structural Basis for Enzymatic Evolution from a Dedicated ADP-ribosyl Cyclase to a Multifunctional NAD Hydrolase*. October 2, 2009. The Journal of Biological Chemistry. 284, 27637-27645.

六、启动子分析


启动子（TATA 盒）
终止点
上游增强子
起始点
多聚A位点
下游增强子
ATG
TAA
外显子
内含子
外显子
内含子
外显子
有意义链
上游区域
转录区域
下游区域
模板链

- 启动子一般可分为两类:
- (1)一类是RNA聚合酶可以直接识别的启动子。

- (2)另一类启动子在和聚合酶结和时需要有蛋白质辅助因子的存在。

- 启动子预测软件大体分为三类
- 第一类是启发式的方法

利用模型描述几种转录因子结合部位定向及其侧翼结构特点，它具有挺高的特异性，但未提供通用的启动子预测方法。

- 第二类是根据启动子与转录因子结合的特性

从转录因子结合部位的密度推测出启动子区域，这方法存在较高的假阳性。

- 第三类是根据启动子区自身的特征来进行测定

这种方法的准确性比较高。同时，还可以结合是否存在CpG岛，而对启动子预测的准确性做出辅助性的推测。

常用启动子预测软件

- PromoterScan：

根据转录因子结合部位在基因组中分布的不平衡性,将转录因子结合部位分布密度与TATA 盒的权重矩阵(weight matrix) 结合起来,从基因组DNA中识别出启动子区。但上述程序预测的假阳性率较高, 平均每19kb出现一个假阳性。

- Promoter 2.0：

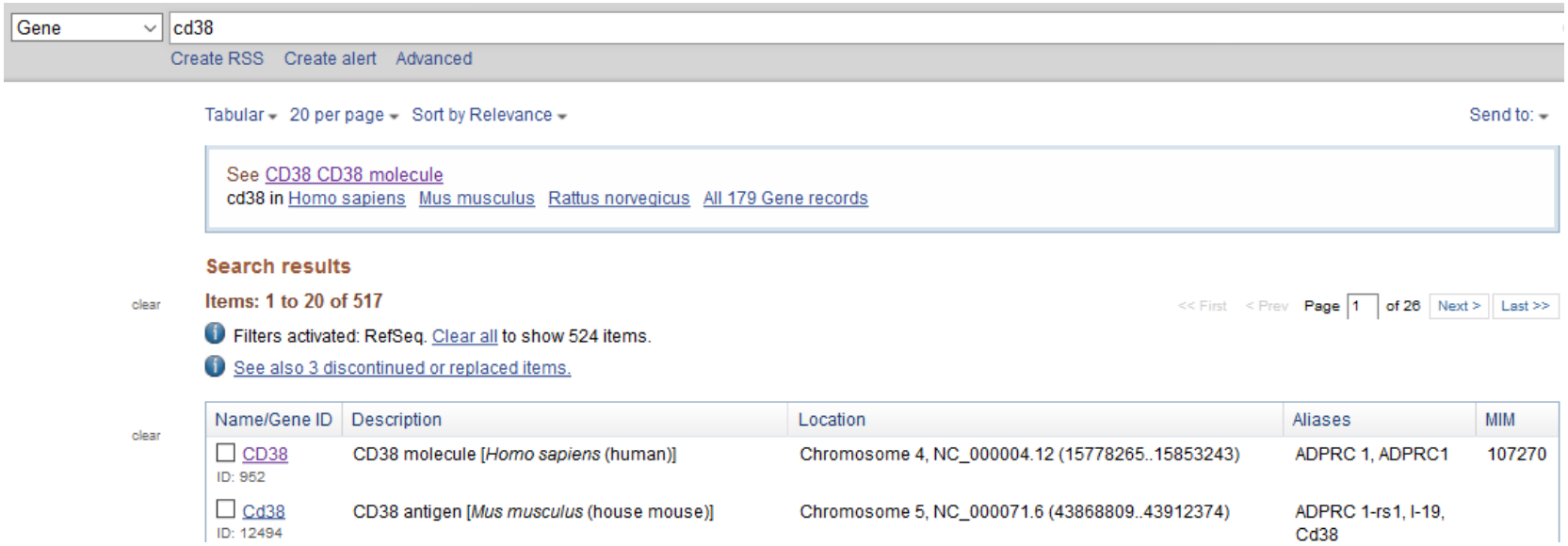
用神经网络方法确定TATA 盒、CCAAT盒、加帽位点(cap site) 和GC 盒(GCbox) 的位置和距离, 识别含TATA 盒的启动子。

- FirstEF：

结合启动子上的CpG 岛(CpG islands) 信息

- 应用举例

（1）找到基因在基因组上的位置



CD38 CD38 molecule [*Homo sapiens* (human)]

Gene ID: 952, updated on 21-Dec-2017

Summary

Official Symbol CD38 provided by HGNC
Official Full Name CD38 molecule provided by HGNC
Primary source HGNC:HGNC:1667
See related Ensembl:ENSG00000004468 MIM:107270; Vega:OTTHUMG00000048206
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 ADPRC1; ADPRC 1
Summary The protein encoded by this gene is a non-lineage-restricted, type II transmembrane glycoprotein that synthesizes and hydrolyzes cyclic adenosine 5'-diphosphate-ribose, an intracellular calcium ion mobilizing messenger. The release of soluble protein and the ability of membrane-bound protein to become internalized indicate both extracellular and intracellular functions for the protein. This protein has an N-terminal cytoplasmic tail, a single membrane-spanning domain, and a C-terminal extracellular region with four N-glycosylation sites. Crystal structure analysis demonstrates that the functional molecule is a dimer, with the central portion containing the catalytic site. It is used as a prognostic marker for patients with chronic lymphocytic leukemia. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Sep 2015]
Expression Broad expression in lymph node (RPKM 11.2), brain (RPKM 10.3) and 17 other tissues See more
Orthologs mouse all

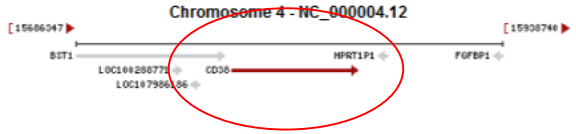
Genomic context

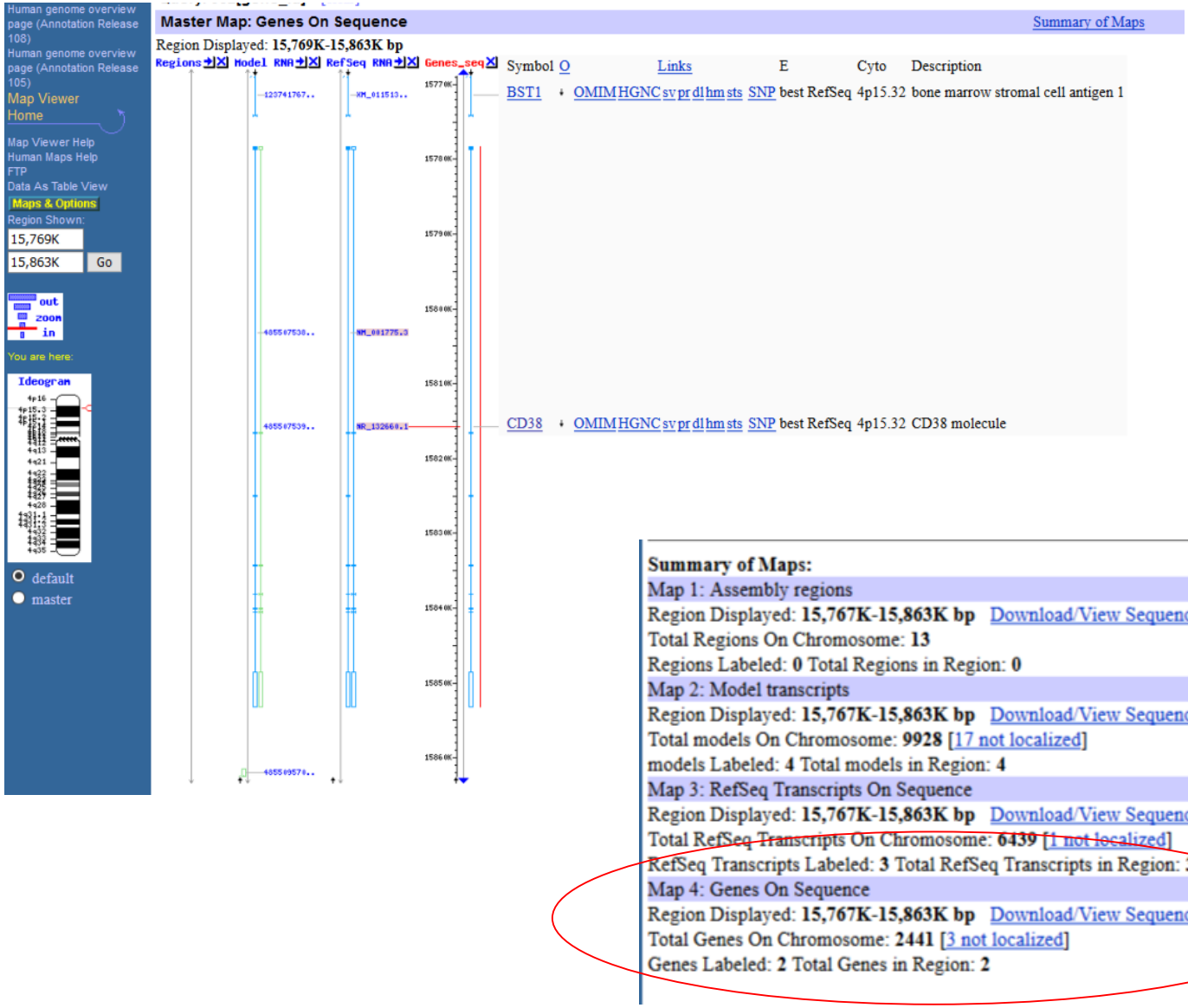
Location: 4p15.32

Exon count: 8

See CD38 in Genome Data Viewer Map Viewer

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7 (GCF_000001405.33)	4	NC_000004.12 (15778265..15853243)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	4	NC_000004.11 (15779921..15850706)




Human genome overview page (Annotation Release 108)
Human genome overview page (Annotation Release 105)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Region Shown:
15,769K
15,863K
Go
out
zoom
in
You are here:
Ideogram
default
master
Master Map: Genes On Sequence
Summary of Maps
Region Displayed: 15,769K-15,863K bp
Regions
Model RNA
RefSeq RNA
Genes_seq
Symbol O
Links
E
Cyto
Description
BST1 OMIM HGNC sv pr dl hm sts SNP best RefSeq 4p15.32 bone marrow stromal cell antigen 1
CD38 OMIM HGNC sv pr dl hm sts SNP best RefSeq 4p15.32 CD38 molecule
Summary of Maps:
Map 1: Assembly regions
Table View
Region Displayed: 15,767K-15,863K bp
Download/View Sequence/Evidence
Total Regions On Chromosome: 13
Regions Labeled: 0 Total Regions in Region: 0
Map 2: Model transcripts
Table View
Region Displayed: 15,767K-15,863K bp
Download/View Sequence/Evidence
Total models On Chromosome: 9928 [17 not localized]
models Labeled: 4 Total models in Region: 4
Map 3: RefSeq Transcripts On Sequence
Table View
Region Displayed: 15,767K-15,863K bp
Download/View Sequence/Evidence
Total RefSeq Transcripts On Chromosome: 6439 [1 not localized]
RefSeq Transcripts Labeled: 3 Total RefSeq Transcripts in Region: 3
Map 4: Genes On Sequence
Table View
Region Displayed: 15,767K-15,863K bp
Download/View Sequence/Evidence
Total Genes On Chromosome: 2441 [3 not localized]
Genes Labeled: 2 Total Genes in Region: 2

（2）下载启动子序列

Homo sapiens (human) (Annotation Release 108)
Region to retrieve (in chromosome coordinates):
Chromosome: 4 Strand: plus
from: 15768892 adjust by: -0K
to: 15862615 adjust by: +0K Change Region/Strand

Sequence Format: FASTA

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_006316.17	6952415	7046138	+	Display Save to Disk

Homo sapiens (human) (Annotation Release 108)
Region to retrieve (in chromosome coordinates):
Chromosome: 4 Strand: plus
from: 15766692 adjust by: -0K
to: 15768892 adjust by: +0K Change Region/Strand

Sequence Format: FASTA

This chromosome region corresponds to the contig region(s):

Contig	start	stop	strand	
NT_006316.17	6952415	7046138	+	Display Save to Disk

（3）选择软件进行分析

Proscan: Version 1.7
Processed Sequence: 2201 Base Pairs

Promoter region predicted on reverse strand in 1207 to 957
Promoter Score: 72.62 (Promoter Cutoff = 53.000000)
TATA found at 974, Est.TSS = 942
Significant Signals:

Name	Strand	Location	Weight
UCE.2	-	1185	1.278000
(SRF)	-	1084	8.606000
hsp70_US	-	1083	8.606000
EivF	-	1031	1.613000
c-fos_US5	-	1031	1.912000
EivF/CREB	-	1031	1.304000
ATF	-	1030	1.157000
CREB	+	1026	1.147000
E4F1	+	1025	1.912000
ATF/CREB	+	1025	1.138000
E4F1	+	1025	1.613000
CREB	+	1024	2.549000

Promoter Scan： https://www-bimas.cit.nih.gov/molbio/proscan/

参考文献

[1]Gao Y1, Mehta K. N-linked glycosylation of CD38 is required for its structure stabilization but not for membrane localization. Mol Cell Biochem. 2007 Jan;295(1-2):1-7. Epub 2006 Jul 14.

[2] Liu J1, Zhao YJ2, Li WH1, Hou YN1, Li T1, Zhao ZY1, Fang C1, Li SL1, Lee HC2. Cytosolic interaction of type III human CD38 with CIB1 modulates cellular cyclic ADP-ribose levels. Proc Natl Acad Sci U S A. 2017 Jul 18. pii: 201703718. doi: 10.1073/pnas.1703718114. [Epub ahead of print]

[3]Babaei S, Geranmayeh A, Seyyedsalehi SA.Protein secondarystructure prediction using modular reciprocal bidirectional recurrentneural networks[J].Comput Methods Programs Biomed, 2010, 100 (3) :237-247

[4]ROST B, SANDER C.Prediction of protein secondary structure at better than 70%accuracy[J].Journal of Molecular Biology, 1993, 232 (2) :584-599.

[5]BUCHAN D W, WARD S, LOBLEY A E, et al.Protein annotation and modelling servers at university college london[J].Nucleic Acids Research, 2010, 38 (Web Server issue) :W563-W568.

[6]DROZDETSKIY A, COLE C, PROCTER J, et al.JPred4:a protein secondary structure prediction server[J].Nucleic Acids Research, 2015, 43 (1) :389-394.

THANKS !