



小鼠蛋白激酶C λ 亚型的生物信息学分析

汇报人：段瑶芸
G04组员：刘若飞 秦青青 卞展





汇报内容

1. 研究背景
2. 核酸序列分析
3. 氨基酸序列分析
4. 蛋白结构预测及分析



研究背景

蛋白激酶C λ (PKC lambda) 是非典型蛋白激酶C的一种，其主要通过参与和介入多种信息传导通路的组成和功能，对细胞的凋亡、细胞周期、增殖等多方面产生重要的调节作用，对体内众多生理病理过程发挥广泛而独特的影响，从而日益受到关注；对其进行深入研究可使我们加深对这个PKC亚家族年轻成员的认识。

研究背景

哺乳动物组织中的PKC亚型

亚 类	氨基酸残基	分子量 (kDa)	激 活 剂	表达组织
A组：经典PKC				
α	672	76	PS、Ca ²⁺ 、DAG、FFA、LysoPC	广泛
β I	671	76	同 上	某些组织
β II	673	76	同 上	多种组织
γ	697	78	同 上	脑
B组：新型PKC				
δ	673	77	PS、DAG	广 泛
ϵ	737	83	PS、DAG、FFA	脑 等
η (L)	683	77	?	肺、皮肤、心脏
θ	707	81	?	骨骼肌
C组：非典型PKC				
ζ	592	67	PS、FFA	广 泛
λ	586	67	?	卵巢、睾丸等

注：PS：磷脂酰丝氨酸 FFA：游离脂肪酸

DAG：二酰基甘油 LysoPC：溶血磷脂酰胆碱

研究背景

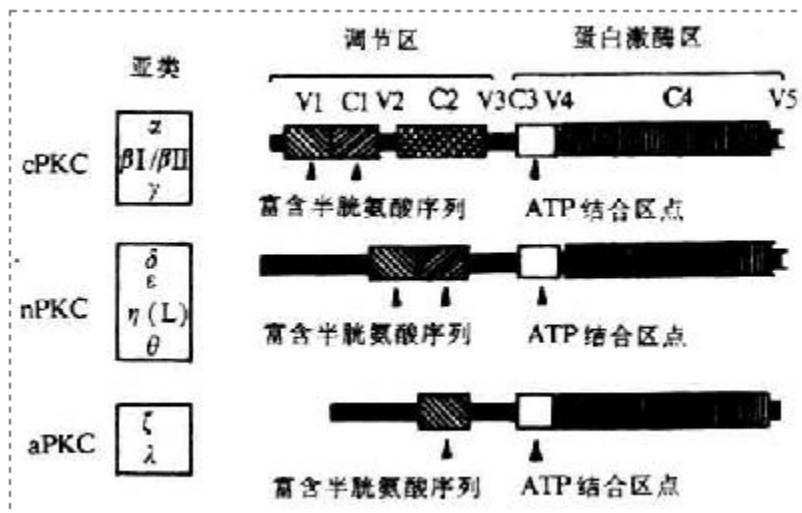
PKC的结构

PKC的所有亚类都由一条单肽链组成（图8-11），分子量大约为67-83kDa，其结构可分为四个保守区C1-C4（mPKC和αPKC缺少C2区）和五个可变区V1-V5。

其中C1区可能是膜结合区，并且含有富含半胱氨酸的随机重复序列Cys-X₂-Cys-X₁₃₍₁₄₎-Cys-X₂-Cys-X₇-Cys-X₇-Cys（X代表任何一种氨基酸），这段顺序与在许多金属-蛋白质及转录调节有关的DNA结合蛋白中的半胱氨酸-锌-DNA结合指形区（cysteine-Zinc-DNAbinding finger）保守顺序Cys-X₂-Cys-X₁₃-Cys-X₂-Cys相似。最近对PKC的多肽片段进行分析发现，该序列与佛波酯和二酰基甘油（DAG）的结合有关。C2区与PKC对Ca²⁺的敏感性有关。C1和C2在结构上不同于其它蛋白激酶，能结合Ca²⁺、磷脂、DAG和TPA，因此C1和C2区又称为调节区。

C3区包括一个ATP结合序列Gly-X-Gly-X-X-Gly-Lys，该区域与其它蛋白激酶的ATP结合位点具有很高的同源性，又称催化区。C4区包含一个底物结合区，是识别磷酸化底物所必需的。

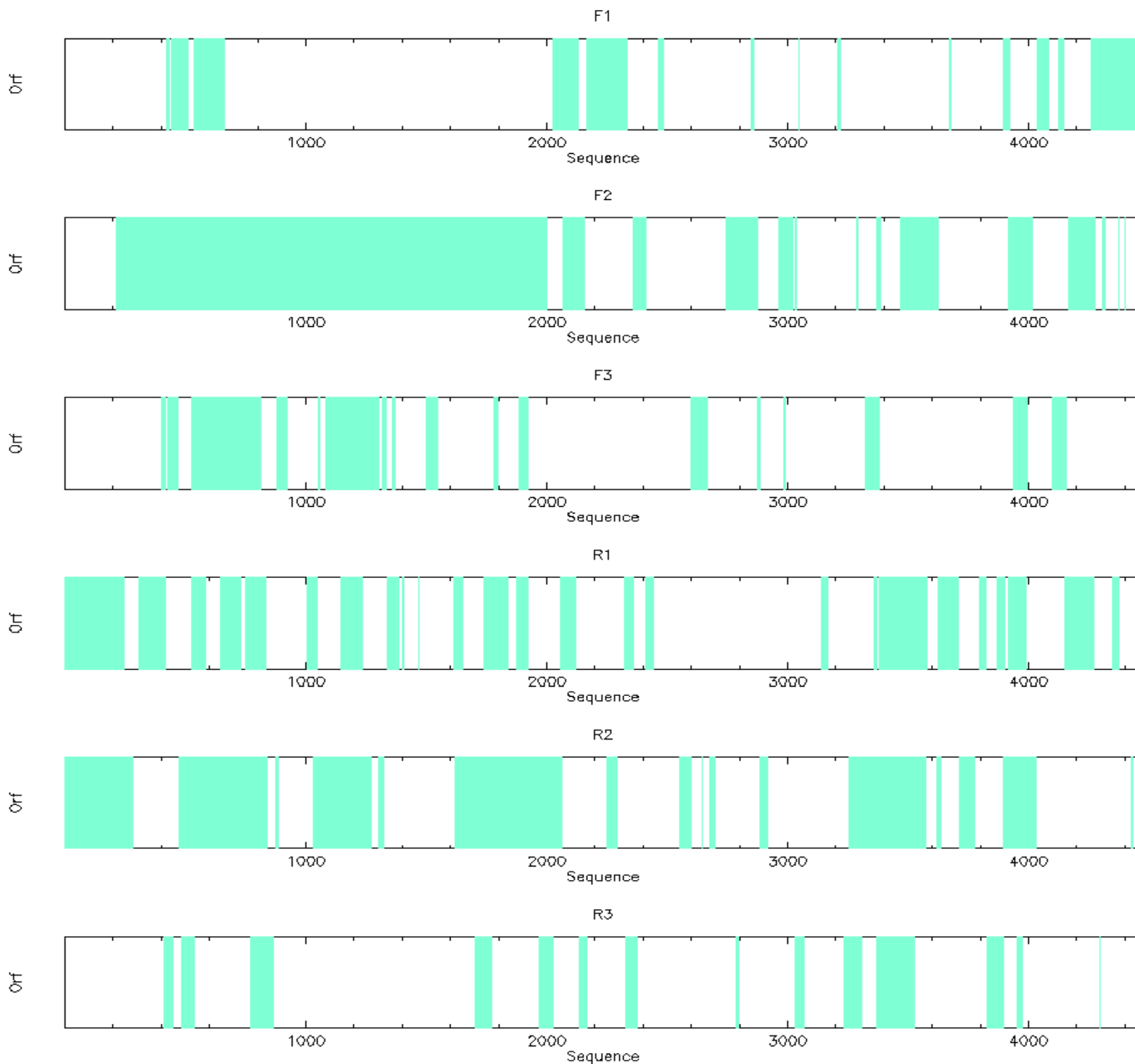
研究背景



ORF分析-PlotORF

PlotORF 从 6 个 frame着手寻找mRNA的开放阅读框。F2是最有可能的开放阅读框。

可看出正向从第一个碱基开始读起才能在215-2002处读到最完整的可连续编码氨基酸序列的读码框。



ORF分析-ShowORF

```
-----|-----|-----|-----|-----|
201 GGGGGAGTGAGGAGATGCCGACCCAGAGGGACAGCAGCACCATGTCCCAC 250
F2   68  G  S  E  E  M  P  T  Q  R  D  S  S  T  M  S  H  83

-----|-----|-----|-----|-----|
1951 AGGTTTTGAGTATATCAACCCCTCTTGATGTCTGCAGAAGAGTGTGTCT 2000
F2   651 G  F  E  Y  I  N  P  L  L  M  S  A  E  E  C  V  * 666

-----|-----|-----|-----|-----|
2001 GATTCTGCTTGCCGCCATTCAGTGCATGGATATCCACCCGTCAGCCTGCT 2050
F2   1   F  C  L  P  P  F  S  A  W  I  S  T  R  Q  P  A  16
```

ShowORF用特殊的格式陈列全长mRNA核酸序列和翻译的蛋白质序列，从215bp到2002bp之间的开放阅读框最长、最可信。1-215虽然也能连续编码氨基酸，但不是由起始密码子开始的，2002之前的三个密码子是终止密码子，因此此后的序列不属于编码区。

ORF分析-SixPack

```
>NM_008857.3_2_ORF1 Translation of NM_008857.3 in frame 2, ORF 1, threshold 1, 595aa
MPTQRDSSSTMSHTVACGGGGDHS HQVRVKAYYRGDIMITHFEPSISFEGLCSEVRDMCSF
DNEQPFTMKWIDEEGDPCTVSSQLELEEAFLYELNKDSELLIHVFPCVPERPGMPCPGE
DKSIYRRGARRWRKLYCANGHTFQAKRFNRAHCAICTDRIWGLGRQGYKCINCKLLVHK
KCHKLV TIECGRHSLPPEPMPMDQTMHPDHTQTVI PYNPSSHESLDQVGEEKEAMNTRE
SGKASSSLGLQDFDLLRVIGRGSYAKVLLVRLKKTDR IYAMKVVKELVNDDEDIDWVQT
EKHVFEQASNH PFLVGLHSCFQTESRLFFVIEYVNGGDL MFHMQRQRKLPEEHARFYSAE
ISLALNYLHERGI IYRDLKLDNVLLDSEGH IKLTDYGMCKEGLRPGDTTSTFCGTPNYIA
PEILRGEDYGF SVDWWALGVLMFEMMAGRSPFDIVGSSDNP DQNTEDYLFQVILEKQIRI
PRLSVKAASVLK SFLNKDPKERLGCHPQTGFADIQGH PFFRNVDWDMMEQKQVVPFFKP
NISGEFGLDNFDS QFTNEPVQLTPDDDDIVRKIDQSEFEGFEY INPLLMSAEECV
```

SixPack程序用于寻找开放阅读框，以多肽序列呈现。需要将最后一条参数（ORF start with an M）设置为Yes，运行后有多个结果，其中一条结果最可信（如图），有595个氨基酸残基。

ORF分析-GetORF

```
>NM_008857.3_10 [215 - 1999] Mus musculus protein kinase C, iota (Prkci), mRNA
ATGCCGACCCAGAGGGACAGCAGCACCATGTCCCACACGGTCGCGTGC GGCGGGCGGGCGGG
GACCATTCCCACCAGGTCCGGGTGAAAGCCTACTACCGCGGGGATATTATGATAACACAC
TTTGAGCCTTCCATCTCCTTTGAGGGACTTTGCAGTGAGGTTTCGAGATATGTGTTCTTTT
GACAATGAGCAGCCGTTCAACATGAAATGGATAGATGAGGAAGGAGACCCATGTACAGTG
TCTTCTCAGTTTGAGTTAGAAGAGGCTTTTCAGGCTGTACGAGCTGAACAAGGATTCTGAA
CTCTTGATTTCATGTATTTCCATGTGTACCAGAGCGTCCTGGAATGCCTTGCCCAGGGGAA
GACAAGTCCATTTACCGGAGAGGGGCACGCCGGTGGAGAAAGCTGTATTTGTGCAAATGGC
CACACTTTTCAAGCCAAACGTTTCAATAGGCGCGCCACTGTGCCATCTGCACAGACAGA
ATCTGGGGCCTCGGACGACAAGGATACAAGTGCATCAACTGCAAACCTGCTGGTTCATAAG
AAGTGCCACAAGCTGGTCACAATTGAGTGTGGGCGGCACTCTTTGCCACCGGAACCCATG
ATGCCAATGGACCAGACCATGCATCCAGACCACACACAGACAGTAATTCCATATAATCCT
TCAAGTCATGAGAGTTTGGACCAAGTTGGTGAAGAAAAGGAGGCAATGAACACCAGGGAG
AGTGGTAAAGCGTCGTCCAGTCTAGGTCTGCAGGATTTTCGATTTGCTTCGAGTTATAGGG
AGGGGAAGTTACGCCAAAGTACTGTTGGTTCGGTTAAAGAAAACAGATCGCATTTATGCA
ATGAAAGTTGTGAAGAAAGAGCTCGTCAATGACGATGAGGATATCGATTGGGTGCAGACA
GAGAAGCATGTGTTTGAGCAGGCGTCCAATCACCCCTTTTCTTGTCGGGCTGCATTCTTGC
TTCCAGACAGAAAGCAGGTTGTTTTTTTGTTCATAGAATATGTAAATGGAGGGGACCTCATG
TTTTCATATGCAGCGACAGAGAAAACCTTCTTGAAGAGCATGCCAGGTTTTACTCTGCAGAA
ATCAGTCTAGCACTGAATTATCTTCATGAGCGAGGGATAAATTATAGAGATTTGAAGTTG
GACAATGTACTGCTAGACTCTGAAGGACACATTAACCTCACTGACTACGGCATGTGTAAG
GAAGGATTGCGGCCTGGAGACACAACCAGCACTTTCTGCGGCACTCCCAATTACATTGCT
CCAGAGATCTTAAGAGGAGAAGATTATGGCTTCAGCGTTGACTGGTGGGCTCTTGGAGTG
CTTATGTTTGAGATGATGGCGGGAAGGTCTCCGTTTGATATCGTTGGGAGCTCTGACAAAT
CCTGACCAAACACAGAGGATTATCTATTCCAAAGTCATTTTGGAAAAGCAGATCCGCATC
CCGCGTTCTCTGTCTGTAAAAGCAGCAAGTGTACTGAAGAGTTTTCTCAACAAGGACCCA
AAGGAACGATTGGGTTGTCACCCTCAAACCTGGATTTGCTGACATTC AAGGACATCCATTC
TTCAGAAATGTGGACTGGGACATGATGGAGCAAAGCAGGTGGTTCCACCCTTTAAACCA
AACATTTCTGGAGAATTTGGTTTTGGATAAATTCGATTCAGTTTACTAATGAACCAGTC
CAGTTCACCTCAGATGATGACATTTGTGAGGAAGATTGATCAGTCTGAATTTGAAGGT
TTTTGAGTATATCAACCCCTCTTGATGTCTGCAGAAGAGTGTGTC
```

getorf(v6.0.1):该软件用于寻找和提取开放阅读框。将输出类型（Type of output）设置为“在起始密码子和终止密码子之间的核酸序列”（Nucleic sequences between START and STOP codon），其它参数不变,输出的是可能的编码区结果。编码区长度不同，选取最长的那条（如图），该结果有1785个碱基。

ORF分析-GetORF

```
>NM_008857.3_10 [215 - 1999] Mus musculus protein kinase C, iota (Prkci), mRNA
MPTQRDSSTMSHTVACGGGGDHS HQVRVKAYYRGDIMITHFEP S ISFEGLCSEVRDMCSF
DNEQPFTMKWIDEEGDPCTVSSQLELEEA FRLYELNKDSELLIHVFPCVPERPGMPCPGE
DKSIYRRGARRWRKLYCANGHTFQAKRFN RRAHCAICTDRIWGLGRQGYKCINCKLLVHK
KCHKLV TIECGRHSLPPEPMPMDQTMHPDHTQTVI PYNPSSHESLDQVGEEKEAMNTRE
SGKASSSLGLQDFDLLRVIGRGSYAKVLLVRLKKTDR IYAMKVVKELVNDDDEDIDWVQT
EKHVFEQASNHPFLVGLHSCFQTESRLFFVIEYVNGG DLMFHMQRQRKLPEEHARFYSAE
ISLALNYLHERGIIYRDLKLDNVLLDSEGH IKLTDYGMCKEGLRPGD TTSTFCGTPNYIA
PEILRGEDYGF SVDWWALGVL MFEMMAGRSPFDIVGSSDNP DQNTEDYLFQVILEKQIRI
PRSLSVKAASVLKSFLNKDPKERLGCHPQTGFADIQGH PFFRNVDWDMMEQKQVVPFFKP
NISGEFGLDNFDSQFTNEPVQLTPDDDDIVRKIDQSEFEGFEYINPLLMSAE ECV
```

将输出类型（Type of output）设置为“在起始密码子和终止密码子之间翻译”（Translation of regions between START and STOP codon），其它参数不变。有多个结果，翻译的蛋白质长度不同，选取最长的那条（如图），共595氨基酸残基。

密码子使用特征分析

```
#CdsCount: 1

#Coding GC 47.15%
#1st letter GC 53.52%
#2nd letter GC 37.42%
#3rd letter GC 50.50%

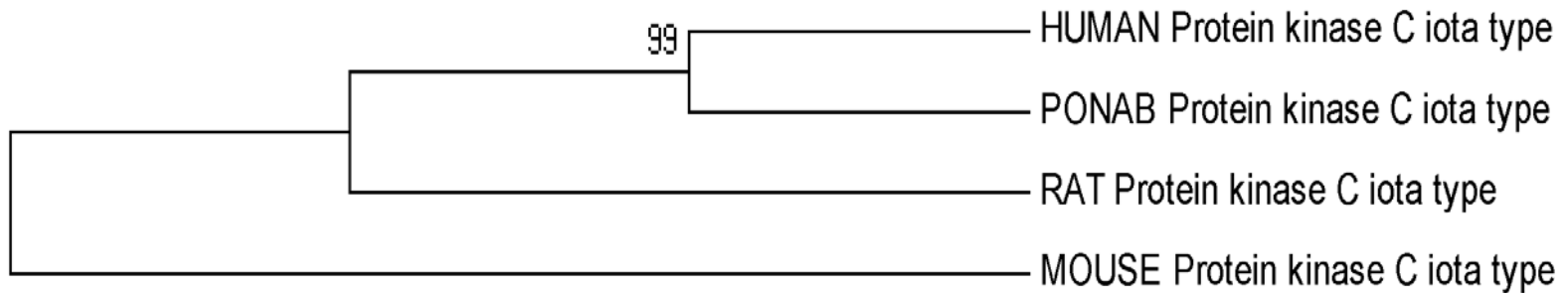
# CHIPS codon usage statistics
Nc = 56.391

#Codon AA Fraction Frequency Number
GCA A 0.391 15.101 9
GCC A 0.261 10.067 6
GCG A 0.174 6.711 4
GCT A 0.174 6.711 4
TGC C 0.500 15.101 9
TGT C 0.500 15.101 9
```

N_c数在20到61之间，数值越大代表密码子偏好程度越低，越小代表密码子偏好程度越强。N_c=56.391表明密码子使用的偏好度低。

GC含量约为47.15%。

系统进化树及氨基酸序列比对



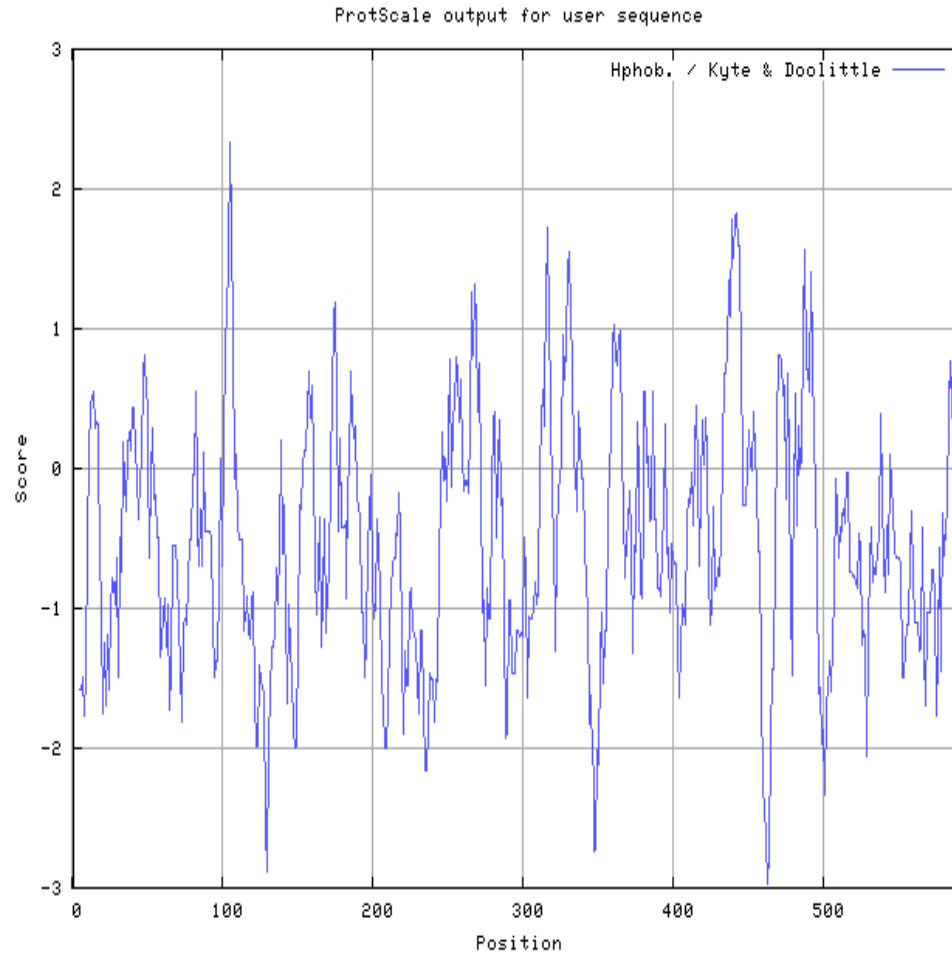
Identical positions	586
Identity	98.322%
Similar positions	5

氨基酸组成分析

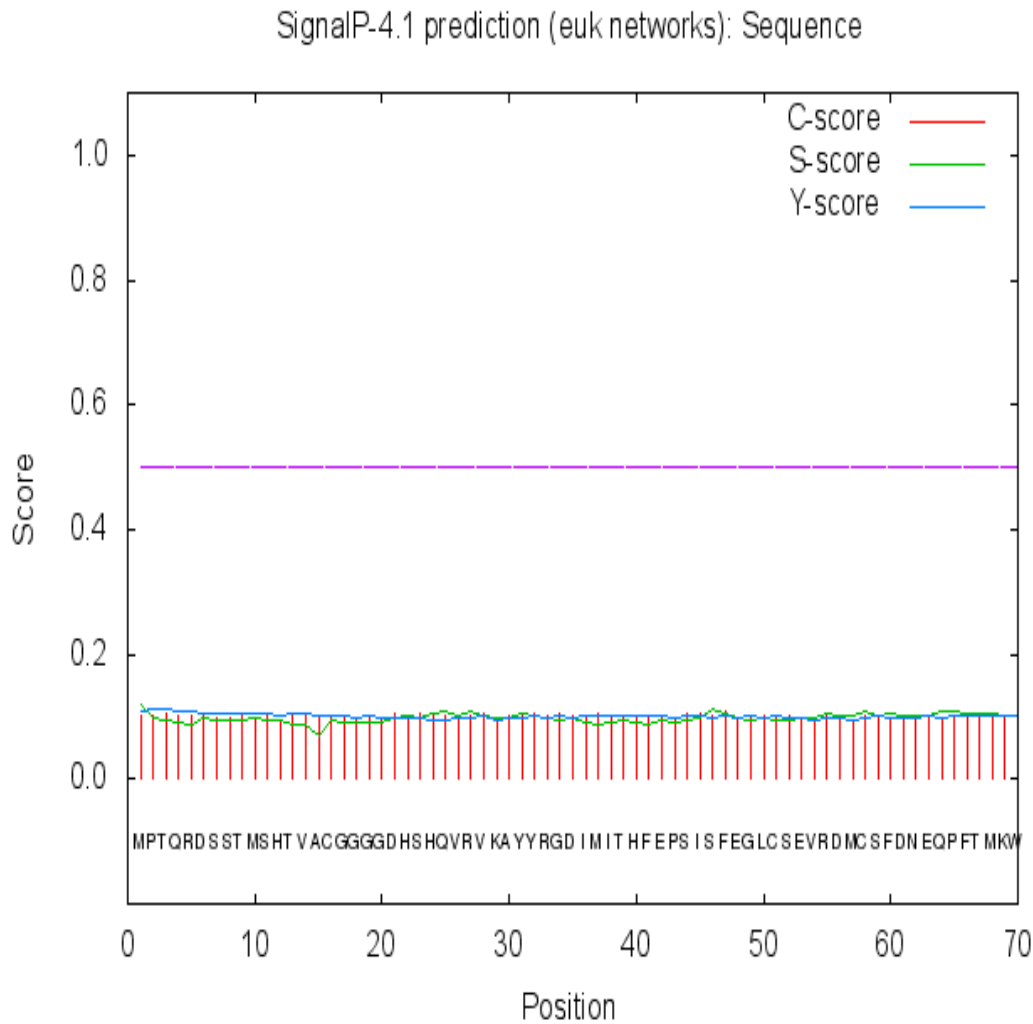
利用WebLab中的
pepstats，显示PKC- λ
的20种氨基酸组成如右
图：

Residue	Number	Mole%	DayhoffStat
A = Ala	23	3.866	0.449
B = Asx	0	0.000	0.000
C = Cys	18	3.025	1.043
D = Asp	43	7.227	1.314
E = Glu	46	7.731	1.289
F = Phe	31	5.210	1.447
G = Gly	40	6.723	0.800
H = His	22	3.697	1.849
I = Ile	30	5.042	1.120
J = ---	0	0.000	0.000
K = Lys	32	5.378	0.815
L = Leu	49	8.235	1.113
M = Met	21	3.529	2.076
N = Asn	21	3.529	0.821
O = ---	0	0.000	0.000
P = Pro	33	5.546	1.067
Q = Gln	25	4.202	1.077
R = Arg	36	6.050	1.235
S = Ser	40	6.723	0.960
T = Thr	25	4.202	0.689
U = ---	0	0.000	0.000
V = Val	35	5.882	0.891
W = Trp	7	1.176	0.905
X = Xaa	0	0.000	0.000
Y = Tyr	18	3.025	0.890
Z = Glx	0	0.000	0.000

ProtScale 在线分析亲疏水性

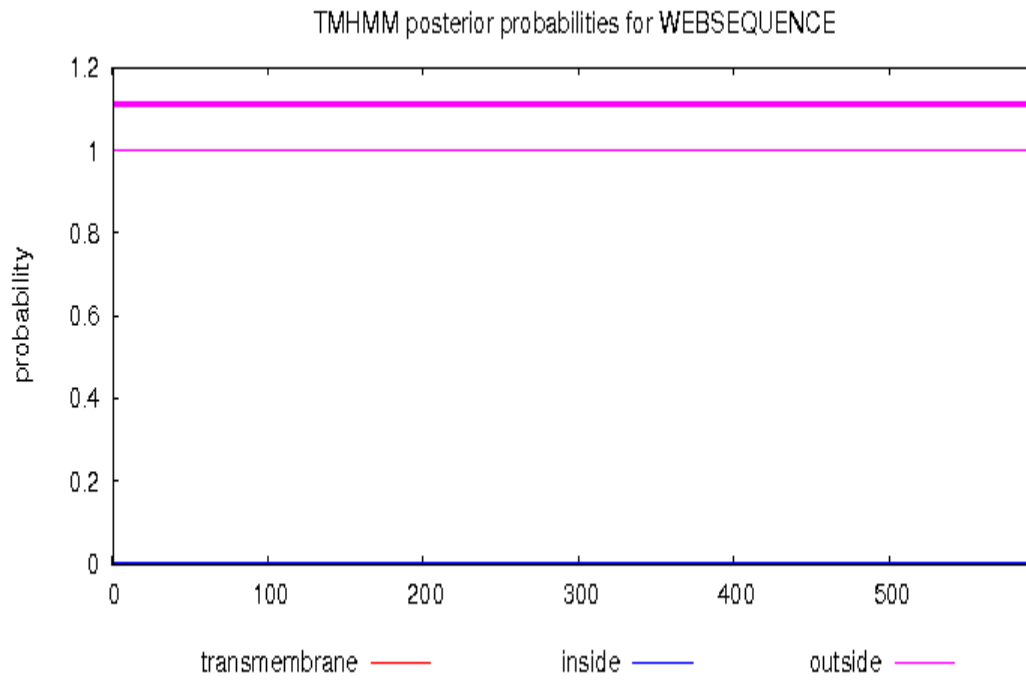


SignalP 4.1 预测信号肽



结果显示
无信号肽

TMHMM 预测跨膜区



结果显示
无跨膜区

TargetP 1.1 软件预测亚细胞定位

```
### targetp v1.1 prediction results #####  
Number of query sequences: 1  
Cleavage site predictions not included.  
Using NON-PLANT networks.
```

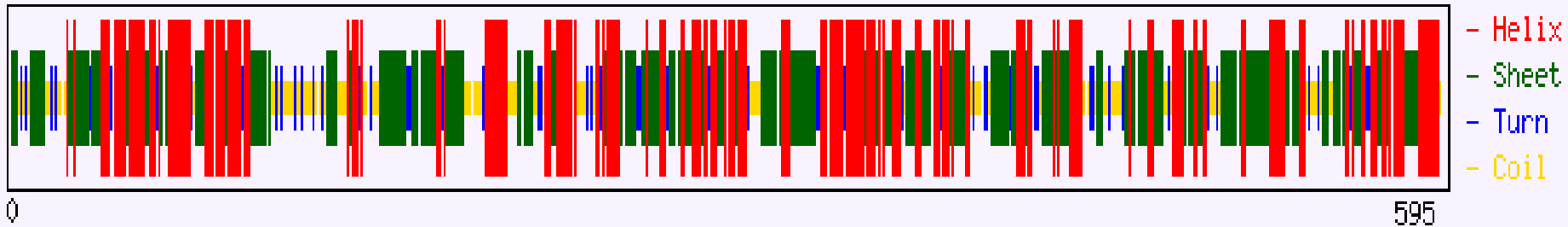
Name	Len	mTP	SP	other	Loc	RC
Sequence	595	0.226	0.034	0.786	_	3
cutoff		0.000	0.000	0.000		

结果显示其定位于细胞质

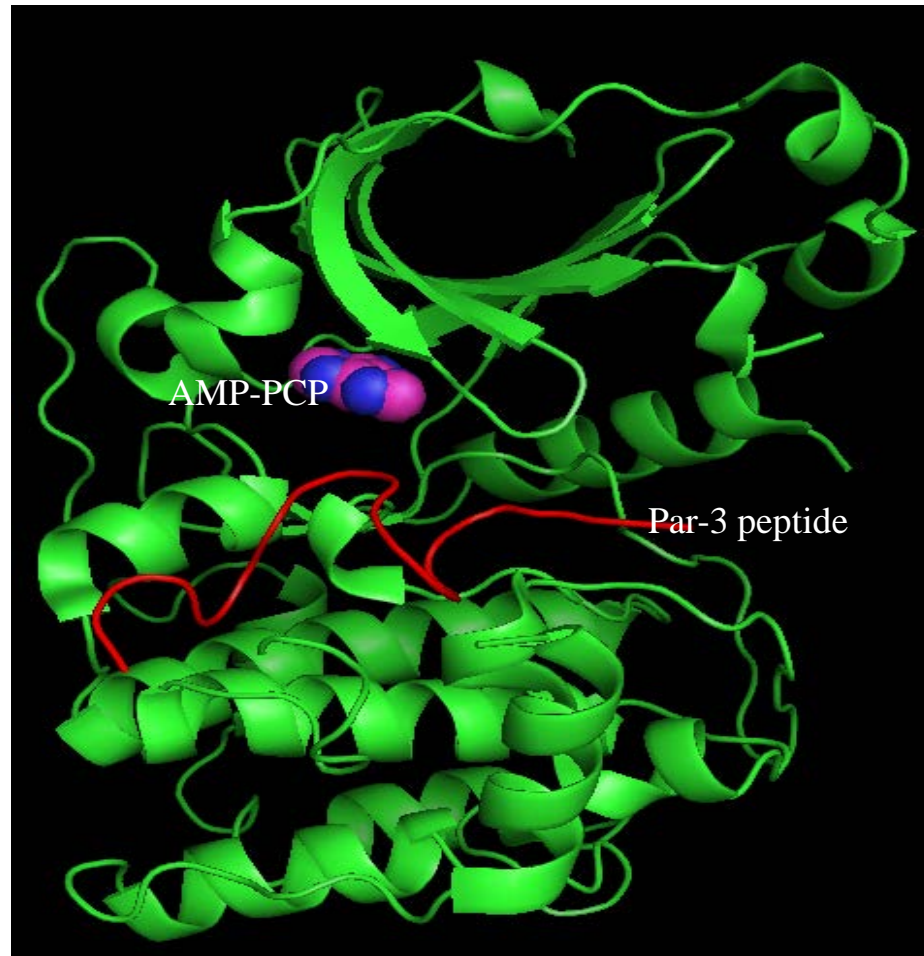
蛋白质二级结构分析

PKC蛋白的分子量为68203.24Da，等电点为5.6767。

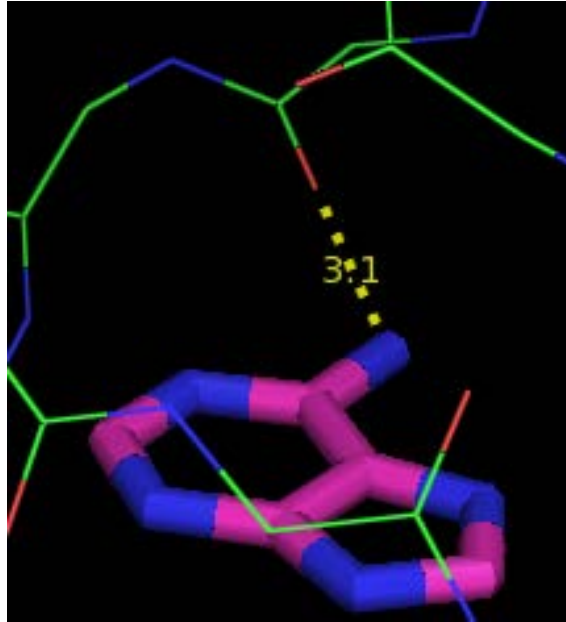
CFSSP (<http://www.biogem.org/tool/chou-fasman/>) 对二级结构进行预测：



蛋白质高级结构分析

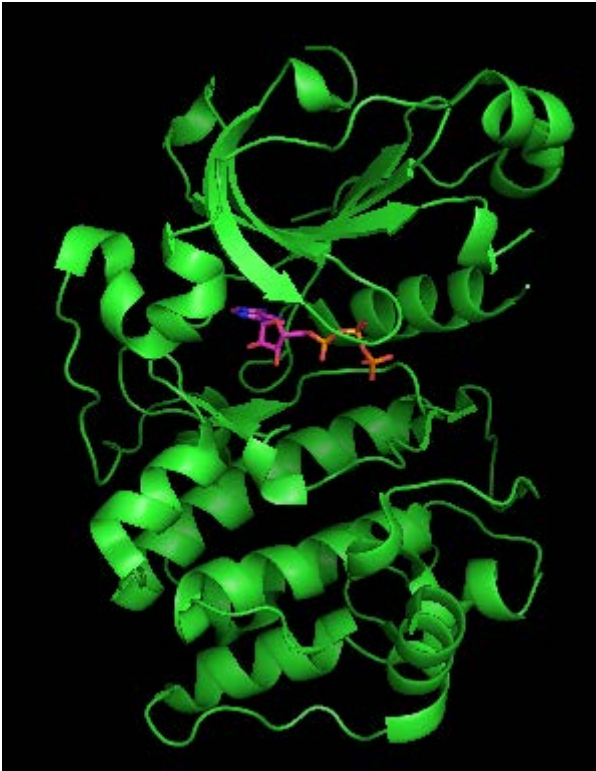


PDB ID: 4DC2 Resolution: 2.44
Residue: 239-577



AMP-PCP与蛋白质之间的最短距离为3.1埃

两种空间结构： close ---open



Close



Open

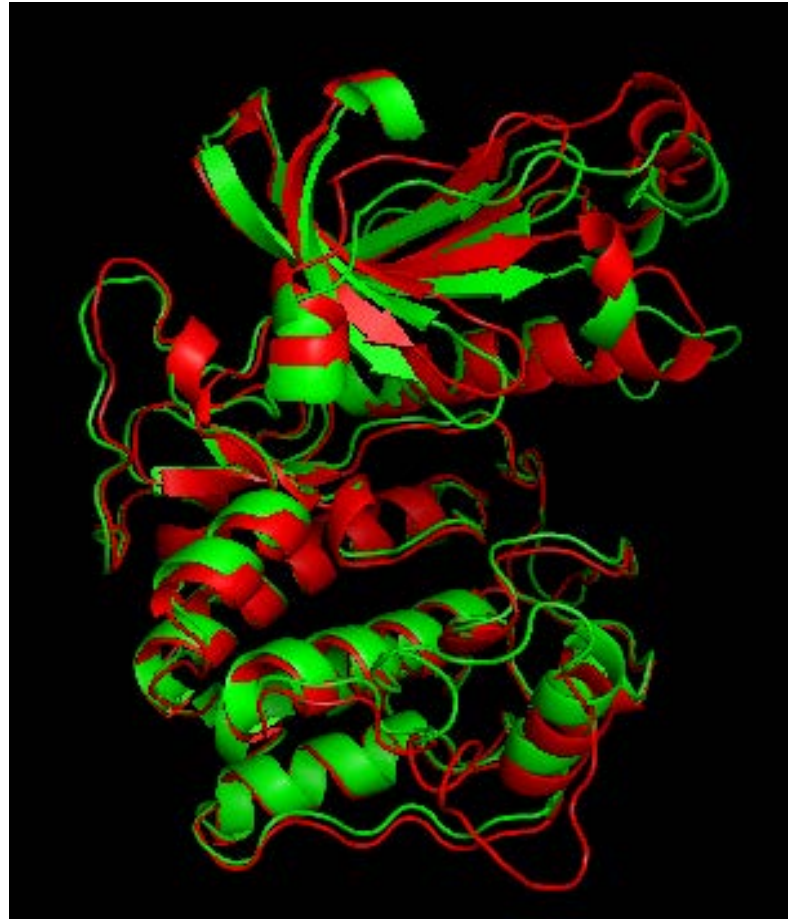
Compare PKC-lambda/Par-3 with open state of PKC-lambda



Red: open state Magenta: PKC-lambda/Par-3

RMS=1.360

Swiss Model predict



Red: automatic model 3A8X Residue: 248-587

Green: definite model 4DC2 Residue: 248-586

RMS=1.338

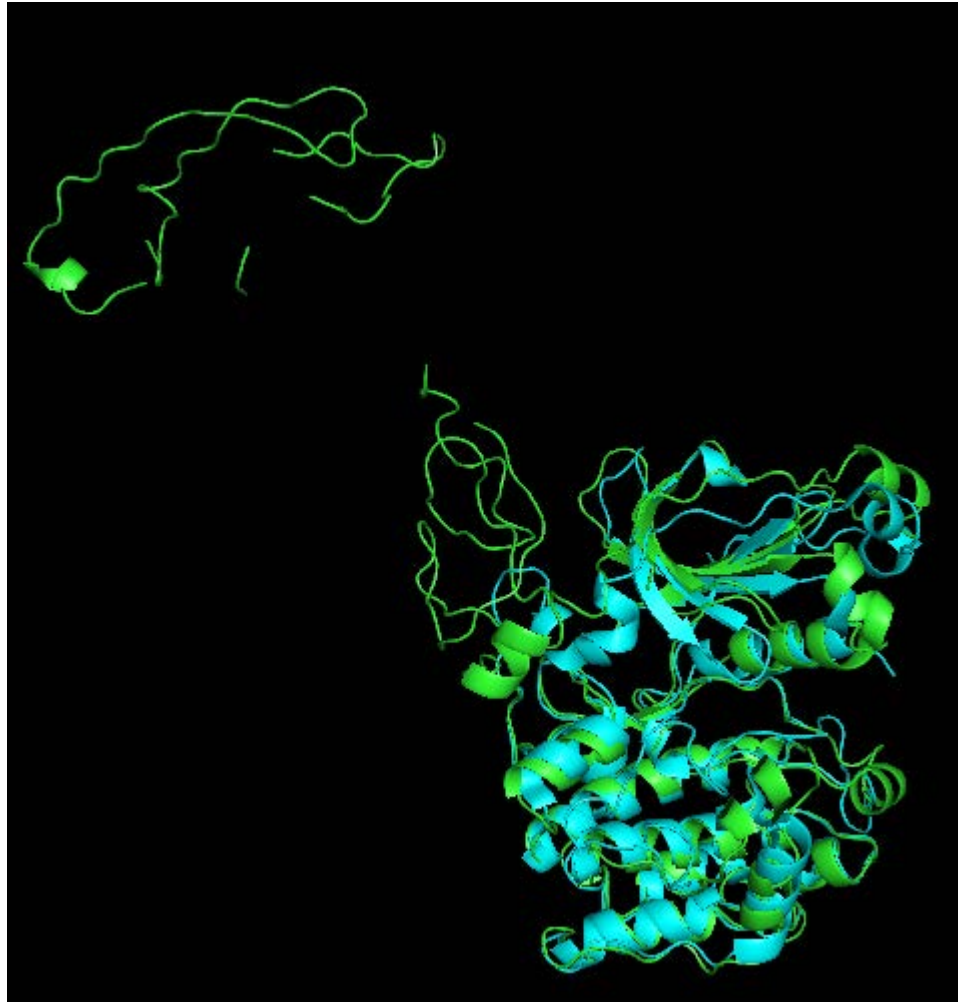
Phyre 2 Predict - Nomal



Confidence: **100.0%** Coverage: **73%**

432 residues (141-585, 73% of sequence) have been modelled with 100.0% confidence by the single highest scoring template. Template: c3pfq.

Compare with 4DC2

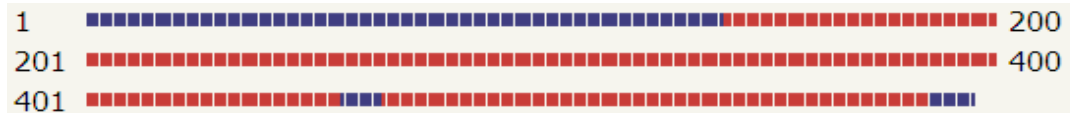
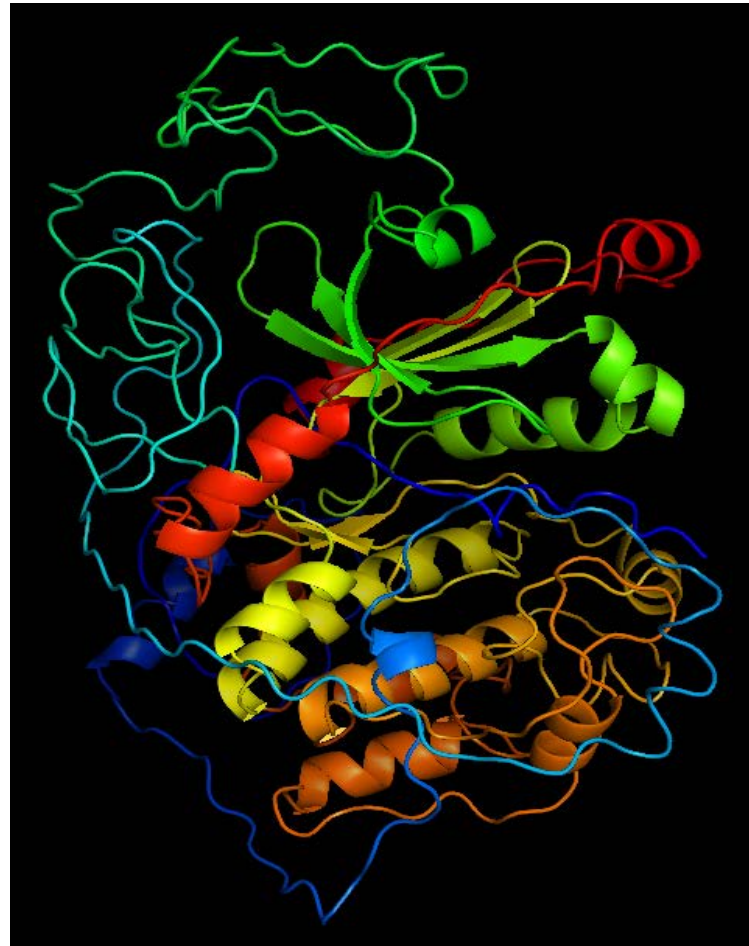


Green: predicted structure Blue: 4DC2

RMS=1.382

Phyre 2 Predict-Intensive

All residues



Confidence Key
High(9)  Low (0)

73% of residues modelled at >90% confidence ([Details](#))

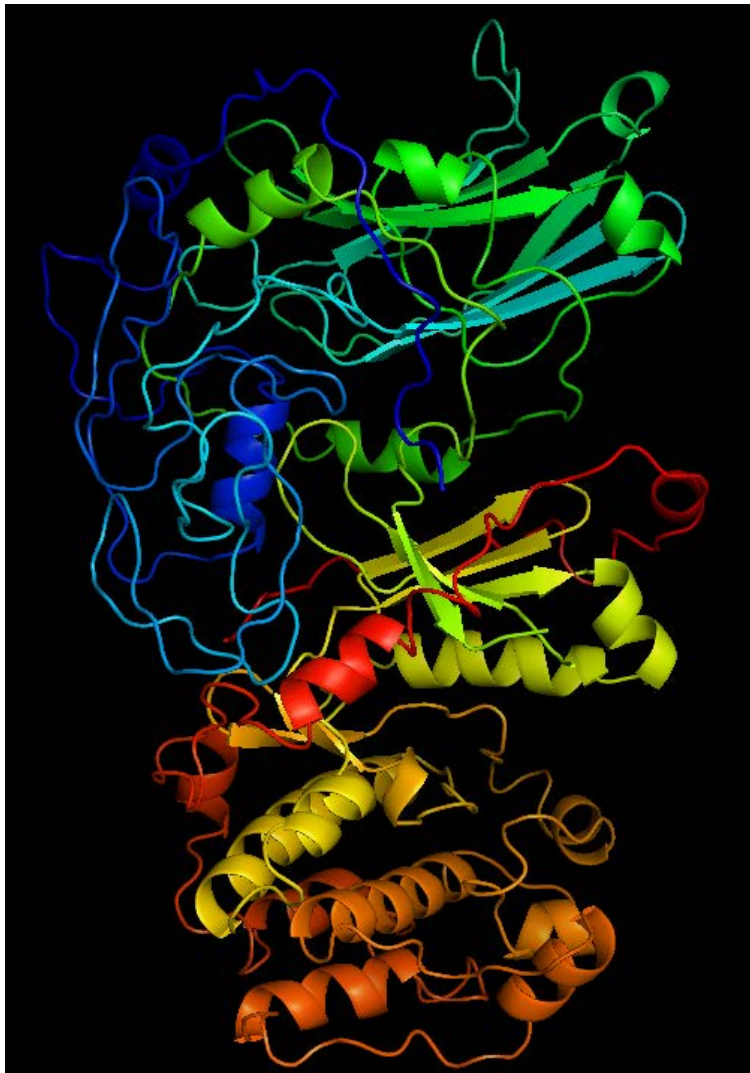
Compare with 4DC2



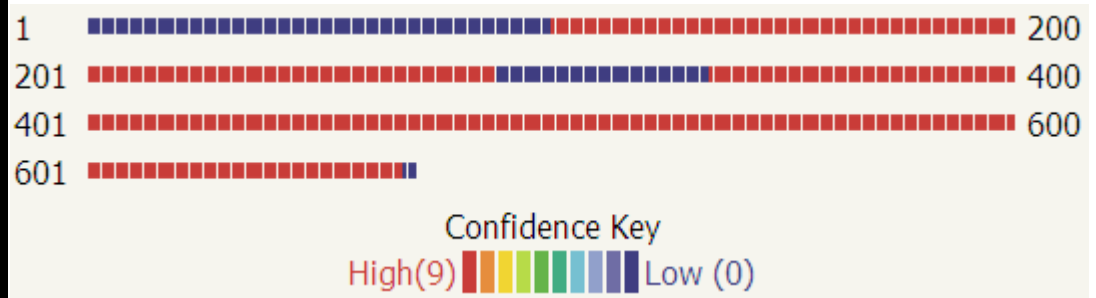
Red: predicted structure Blue: 4DC2

RMS=1.401

Predict PKC-beta (classical PKC)

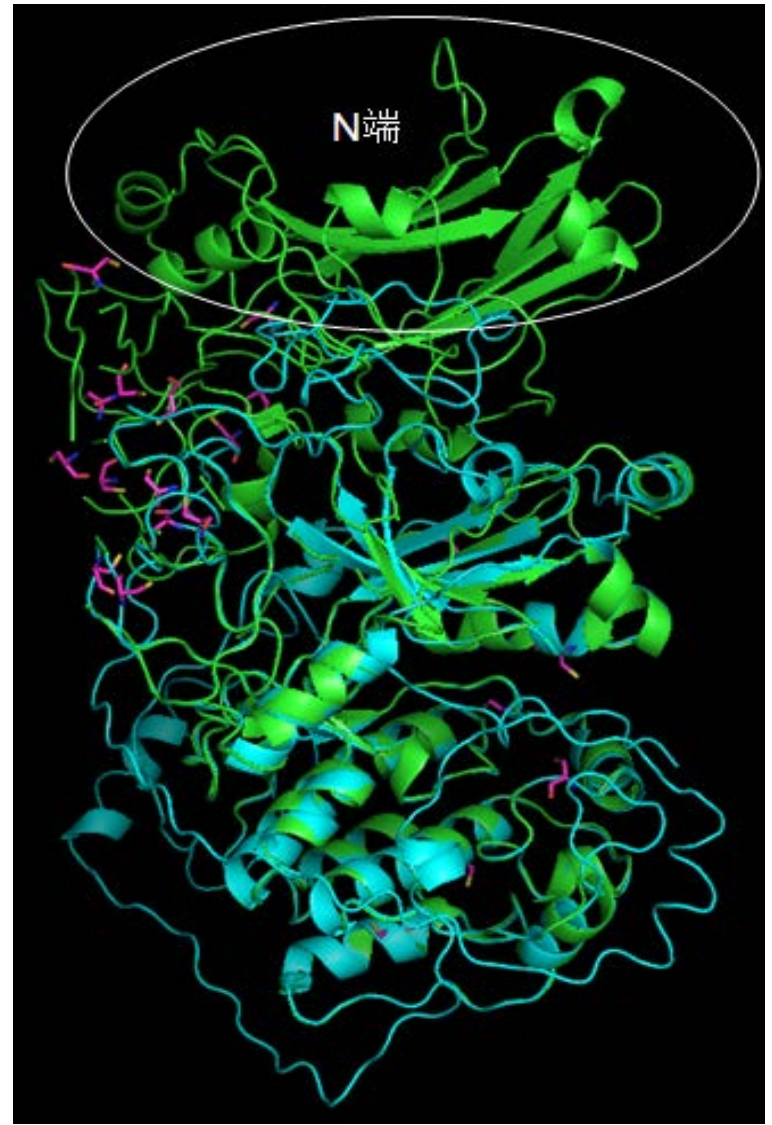


Phyre2-Intensive

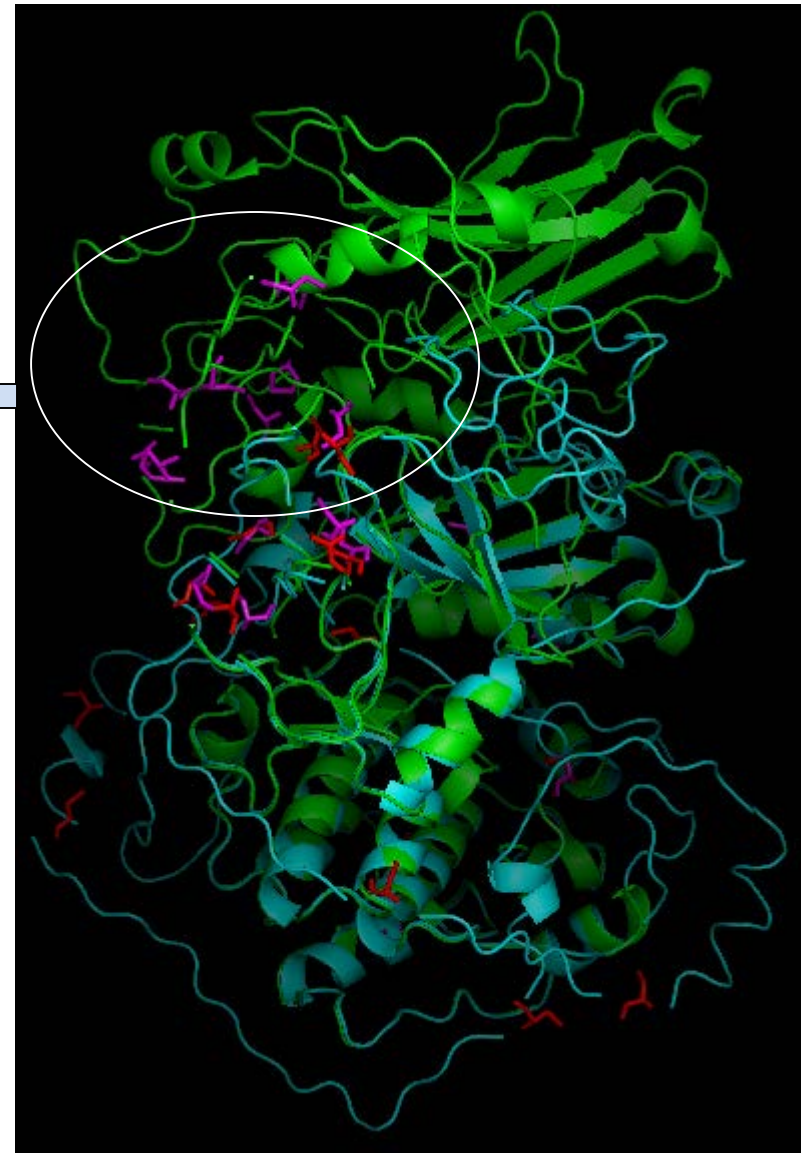
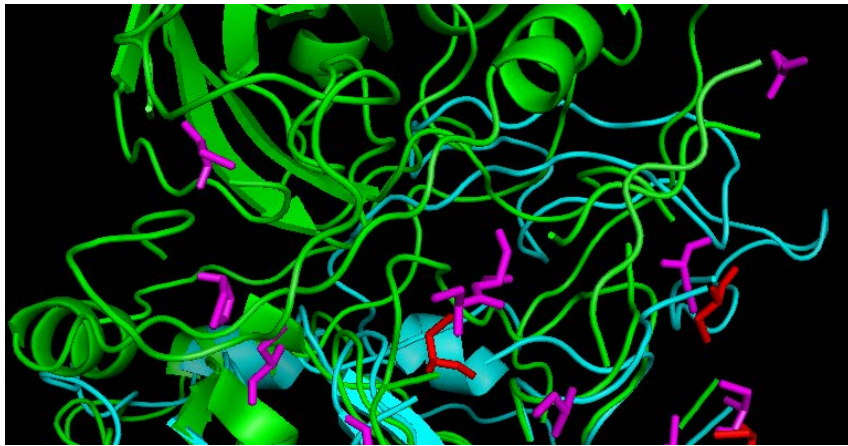




PKC-beta Cys distribution



Compare PKC-beta with PKC-lambda



Compare PKC-beta with
PKC-lambda

Green: PKC-beta

Blue: PKC-lambda



致谢

衷心感谢罗老师对我们的悉心指导和谆谆教诲，让我们发现和体会到了生物信息学的价值！

非常感谢组员及同学们的支持和帮助！





Thank you for your attention!

