

棉花线粒体ATP合酶beta亚基 的生物信息学分析

PKU09S2 Group4

(刘娟丽 李爽 王剑峽 侯英楠)

报告人：侯英楠

Outline

- 研究背景
- 序列分析
- 系统发生分析
- 三级结构预测与分析
- 结果与讨论

研究背景

- 棉花(cotton)

属于植物界、种子植物门、被子植物亚门、双子叶植物纲、五桠果亚纲、锦葵目 (Malvales)、锦葵科 (Malvaceae)、棉属 (Gossypium)。棉花是唯一由种子生产纤维的农作物。



棉花及棉纤维的地位

- 作为世界上最重要的经济作物之一，棉花是当今世界首要的纤维作物和重要的油料作物，是纺织品制造业的重要原料，也是我国出口创汇的重要产品。
- 我国是目前世界上最大的棉花生产国、消费国和纺织加工国，棉花种植业和棉纺业在我国国民经济中起着举足轻重的作用。

我国棉花生产存在的问题

- 首先，我国棉花生产品质单一，纤维强度偏低，纺高质纱的优质棉需大量进口。
- 其次，棉花病害是制约我国棉花生产的重要因素。

重要的生物学问题

- 棉花纤维的伸长机制
- 棉花抗病虫害的机理

棉花纤维发育过程



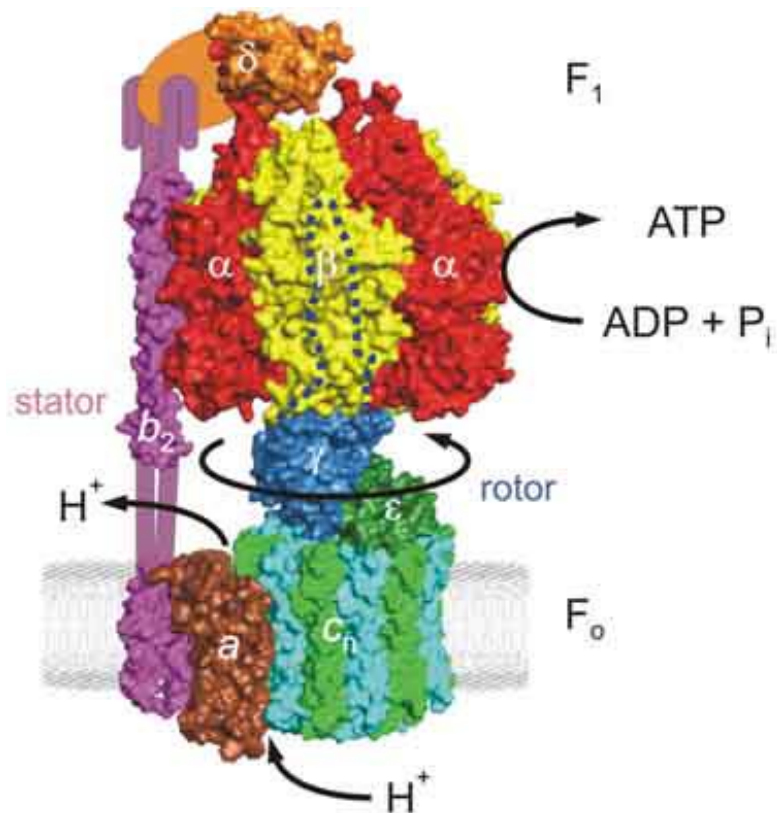
A. 开花期植株； B. 开花的棉花；
C.吐絮及未吐絮的棉桃。 D.成熟吐絮期植株。

图片引自中国作物种质资源信息网<http://icgr.caas.net.cn>。

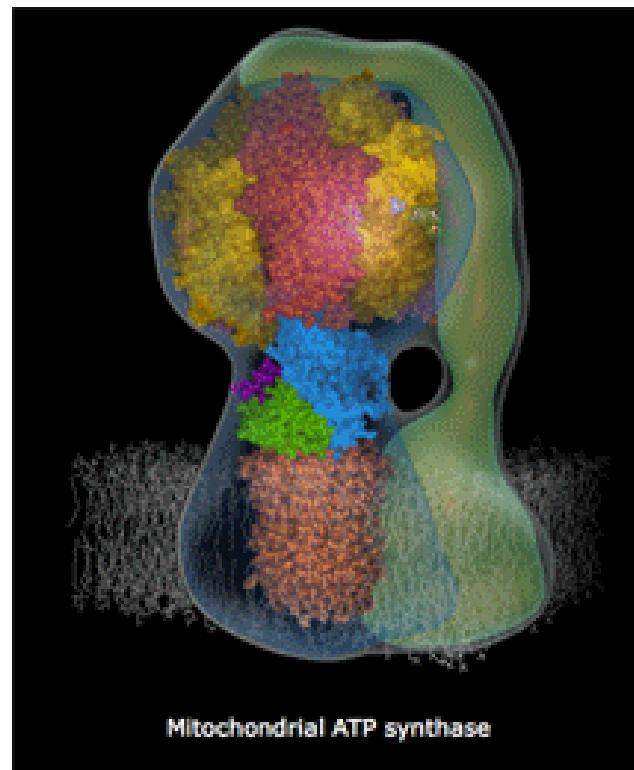
ATP合酶研究背景

- ATP合酶是利用跨膜质子泵催化ADP和磷酸反应生成ATP的一类蛋白复合体。
- 在线粒体中存在的主要是F₁-F₀型ATP合酶，它由膜外可溶性的球形结构域F₁和膜内结构域F₀两个复合体组成，而每个复合体又由多个小亚基组成。
- 其中，F₁结构域由 $\alpha_3\beta_3\gamma\delta\epsilon$ 五种肽链九个亚基组成，F₀部分是由a，b和c亚基组成。

真核生物ATP合酶结构图



Boyer et al., 1997

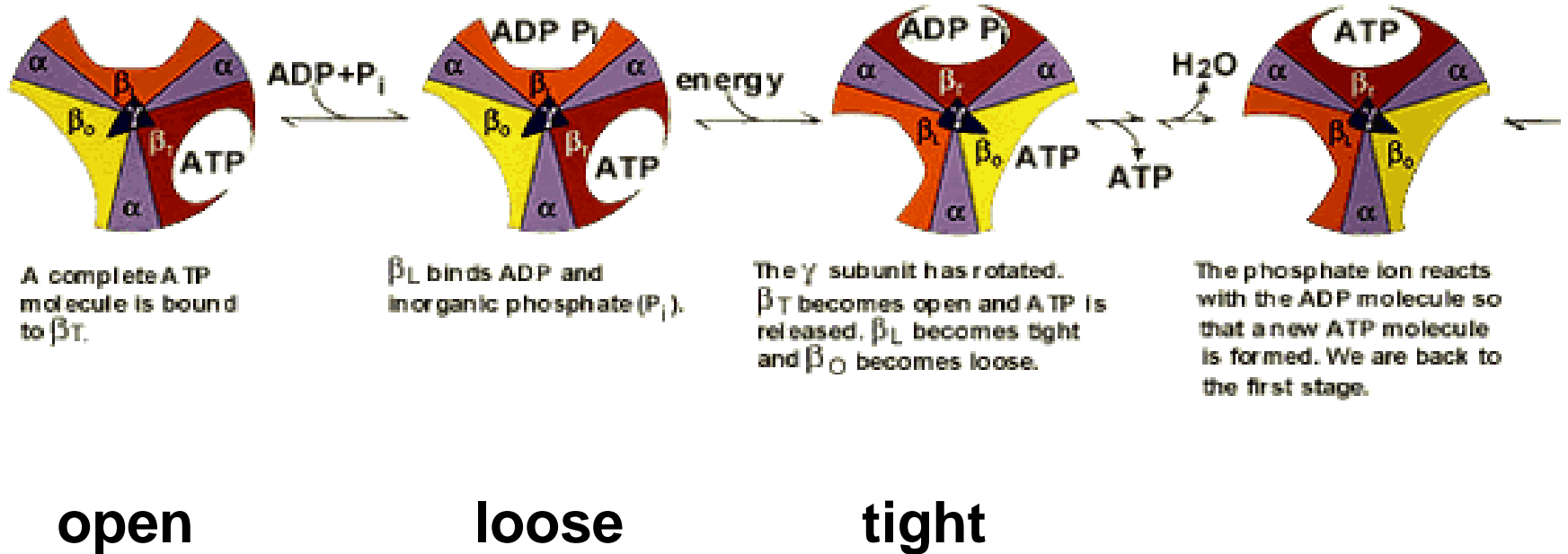


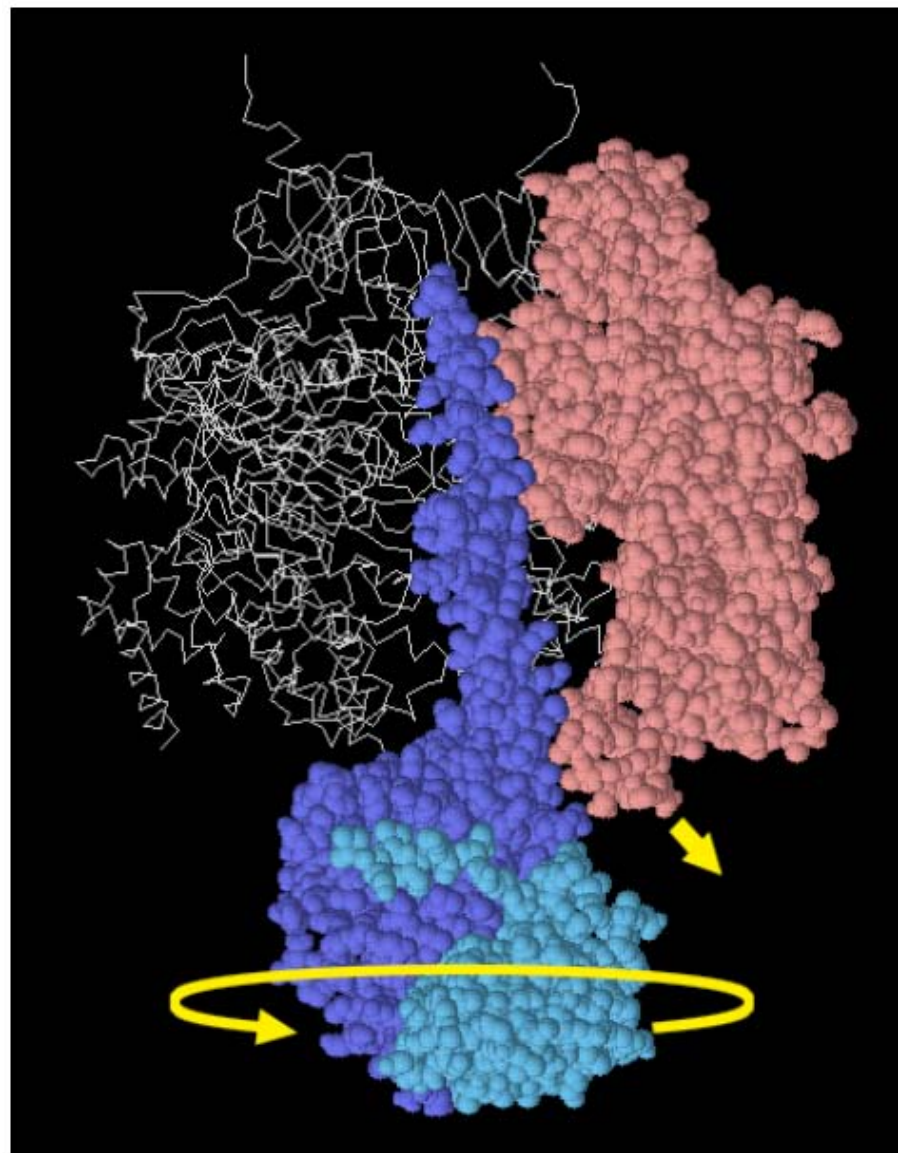
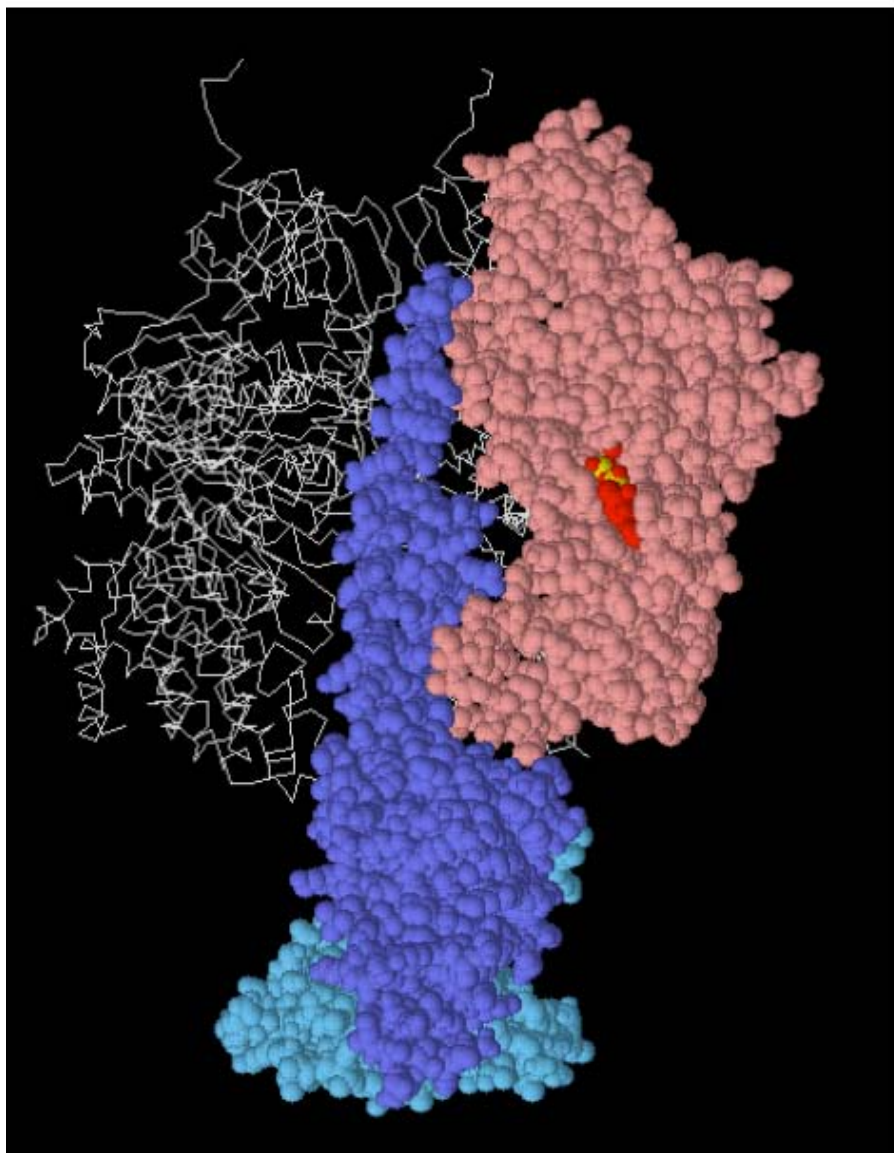
<http://www.mrc-mbu.cam.ac.uk/>

真核生物线粒体ATP合酶催化过程

“结合改变” (combination change) 或
“转换位点” (conversion position point) 假说
(Boyer et al., 1993) :

任一时刻，ATP 合酶上的三个催化部位（ β 亚基）的构象总是不同的。





获得序列

- 构建开花后5-10天的棉花纤维cDNA文库:大规模EST测序, 拼接得到了1万多条独立的陆地棉 (upland cotton) cDNA, 点制成cDNA芯片后通过与非快速伸长期的胚珠及无毛突变体相应时期芯片数据比较分析, 得到了一些棉花纤维发育相关基因 (Ji et al., 2003; Shi et al., 2006)。
- 在这些基因中, 我们发现1个在纤维发育时期表达上调的ATP合酶编码beta亚基的基因, 并进行生物信息学的分析。

序列分析

- RACE**测序得到该基因的 **cDNA**序列

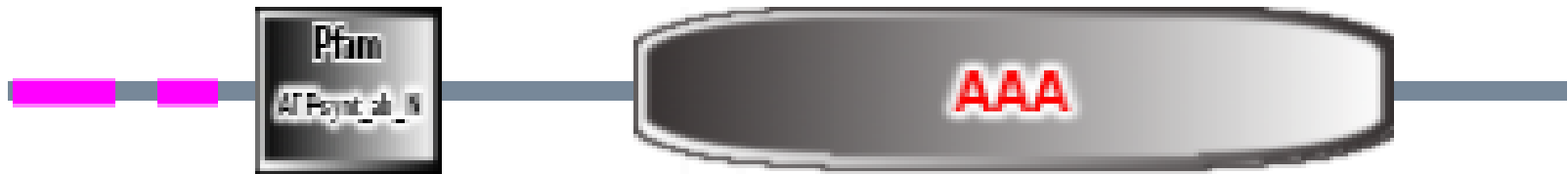
>cl3ct1411cn4173 COVERAGE:0.96; CRAWID: 2; TOTAL_ESTS:11; LENGTH:2241bp; ALT_CONSENSI:14

```
AAACCCTAGCCTCTATCTGGGAGCCCTTTCCAATGGCTTCGCGTTCGACTCCTCTCCTCTTTCCCTCCGCTCCAC
TTCTCGCCGATCTCTCTCCAAATCTCCCAGCCCTACCCCTCGCCTCCCTTACCTTCCCCCACGCGCCGCC
TTTTCCGTGCCCTACTTTCTCTCACGCGATGCTGAGTACTCTACATCTGCGGCTGCTGCCGCTTCTGCTCCA
TCGCAGCCTCCTCCCGCTAAGAAGGGCGGAAAAGGTAAGATCACTGATGAGTTCCTGGTGGTGGAGCCAT
CGGTCAGGTTTGTCAAGTAATCGGTGCCGTCGTCGATGTTAGATTCGATGAAGGCTTGCCTCCGATTTTGAC
TGCGCTTGAAGTGATGGACCACTCGATCCGGCTTGTGTTGGAAGTGGCTCAGCACTTGGGTGAAAACATGGT
TAGGACCATTGCTATGGATGGAACGGAGGGACTCGTCAGAGGGCAGCCGGTTCTCAACACCGGATCTCCTA
TCACTGTGCCTGTCGGTAGAGCTACTCTTGGTCGTATCATAAATGTCATTGGAGAGCCTATTGACGAGAAAG
GCGAACTCAAAACCGACCATTTCTTGCCCATACACAGAGAAGCACCAGCCTTTGTGATCAAGCTACTGAACA
ACAGATCCTTGTCACTGGTATTAAGGTTGTGGATCTCCTTGCACCATATCAAAGAGGAGGGAAAGATTGGACTT
TTTGGTGGTGCTGGTGTAGGGAAGACAGTGCTTATTATGGAAGTATCAACAATGTTGCTAAAGCCCACGGT
GGTTTCTCAGTGTTTGGCCGGTGTGGAGAACGTAAGTACTAGAGAGGGTAATGACTTGTATAGGGAAATGATTGAG
AGTGGTGTATTAAGCTAGGTGAAAAGCAGGCTGATAGCAAGTGTGCTCTTGTATTATGGTCAAATGAATGAGC
CCCCTGGTGCTCGTGTCTGCTGTTGGCCTTACTGGTCTGACTGTGGCTGAACATTTCCGTGACGCTGAAGGG
CAGGATGTGCTTCTCTTTCATTGACAACATTTTCCGTTTTACTCAAGCAAAGTCTGAAGTGTCTGCCTTGCTAGG
CCGTATTCCATCTGCTGTCTGTTACCAACCAACCTTGGCTACTGATCTTGGAGGTCTTCAGGAACGTATTACC
ACCACCAAGAAAGGTTCTATCACTTCTGTTCAAGCCATATATGTGCCTGCTGATGACTTGACAGATCCAGCTC
CTGCAACAACCTTTGCTCACTTGGATGCCACAACCTGTGTTGTCCCGACAGATCTCTGAGCTTGGTATTTATCC
TGCTGTGGATCCTTTGGATTCTACATCTCGTATGCTCTCTCCTCATATTTTGGGTGAGGAACACTACAACACT
GCTCGTGGTGTTTCAAGAGGTTCTCCAAAACACTACAAGAAGTTCGCAAGATATTATTGCTATTTTGGGGATGGACG
AGCTTAGTGAAGATGATAAATTAAGTGTGCACGTGCTCGTAAGATCCAAAGGTTCTTGAGCCAGCCTTTCCA
TGTGGCAGAAGTGTCACTGGTGCTCCTGGCAAGTACGTGGAGTTGAAGGAGAGCATCACCAGCTTCCAGG
GCGTGTGGATGGAAAGTATGATGACCTGCCCGAGCAGTCGTTTTACATGGTTGGAGGAATCGAGGAAGTG
ATTGCCAAGGCAGATAAGATTGCCAAGGAATCTGCAGCCTAATCAAATCCACATACCTTAATTTCTGCTGAGT
AATTTGAAAATAATCTAGGTGGTGTGGTAATGCCAGCAACAGAAGTTCATTTATTTTGGGGGTTTGTGTTGT
TTATGTTAATGATGTTATTCAAGAGTGTAAGAAGAGCACTTCCGCCCGCTTTTTTCTTTGCTTAAATAAGGGA
AATAGAAGGCGGATGTGTACATTTTTTTTGTTCGAAAACGCTGGCTGCTAAAGCCTTATCCTTTCTGACTGCC
GTTTATTTCGGAAGTACTGATGTTTTGAAATGCATTGAGACTAATGTTTCCTGAACCCATGGGGGTTTCAATTT
TCTTTTTCTTAATGAAGAAATTGGCGAATTTTATCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAACGGGGCCCCCTATTTCCCCCGGGGGGGGGGCCAGGGAAAAGGGACCCCAATTTCCCCTATTA
AGGGGGCTCTAATAAAAATTATACGGGGCGCGCCTTTTTACACAGGGCGGGAGGTGGGGGAAA
```

- 用**ExPasy**中的**Transeq**工具将核酸序列翻译成蛋白序列，命名为ATP beta1。
- > ATP beta1
MASRLLSSFLRSTSRRLSKSPSPTPRLPSPSPTRRPFPCPYFL
SRDAEYSTSAAAAASAPSQPPPAKKGGKGGKITDEFTGAGAIGQVC
QVIGAVVDVRFDEGLPPILTALEVMDHSIRLVLEVAQHLGENMVRT
IAMDGTEGLVRGQPVLNTGSPITVPVGRATLGRIINVIGEPIDEKGE
LKTDHFLPIHREAPAFVDQATEQQILVTGIKVVDLLAPYQRGGKIGL
FGGAGVGKTVLIMELINNVAKAHGGFSVFAGVGERTREGNDLYRE
MIESGVIKLGEKQADSKCALVYGQMNEPPGARARVGLTGLTVAEH
FRDAEGQDVLLFIDNIFRFTQANSEVSALLGRIPSAVGYQPTLATDL
GGLQERITTTKKGSITSVQAIYVPADDLTD PAPATTF AHL DATTVLS
RQISELGIYPAVDPLDSTSRMLSPHILGEEHYNTARGVQKVLQNYK
NLQDIIAILGMDELSEDDKLTVARARKIQRFLSQPFHVAEVFTGAPG
KYVELKESITSFQGVLDGKYDDLPEQSFYMGVGGIEEVIKADKIAK
ESAA

利用SMART网站对目的蛋白进行结构域分析

- 该蛋白具有一个AAA结构域(ATPases associated with a variety of cellular activities), 以及**ATP-synt_ab_N**结构域, 因此具有ATP合酶活性。



Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
low complexity	3	40	-
low complexity	54	75	-
Pfam:ATP-synt_ab_N	89	155	3.60e-25
AAA	224	496	2.15e-07

利用MEME对目的蛋白进行Motif分析

从Expasy获得序列ATP合酶beta亚基序列（共24条）

UniProt > UniProtKB Downloads • Contact • Documentation/Help

Search in Protein Knowledgebase (UniProtKB) **Query** atp AND synthase AND name:beta AND reviewed:yes AND fragment:no AND name:mitochondrial [Fields »](#)

1 - 25 of 38 results for **atp** AND **synthase** AND **name:beta** AND **reviewed:yes** AND **fragment:no** AND **name:mitochondrial** in UniProtKB sorted by **score** descending

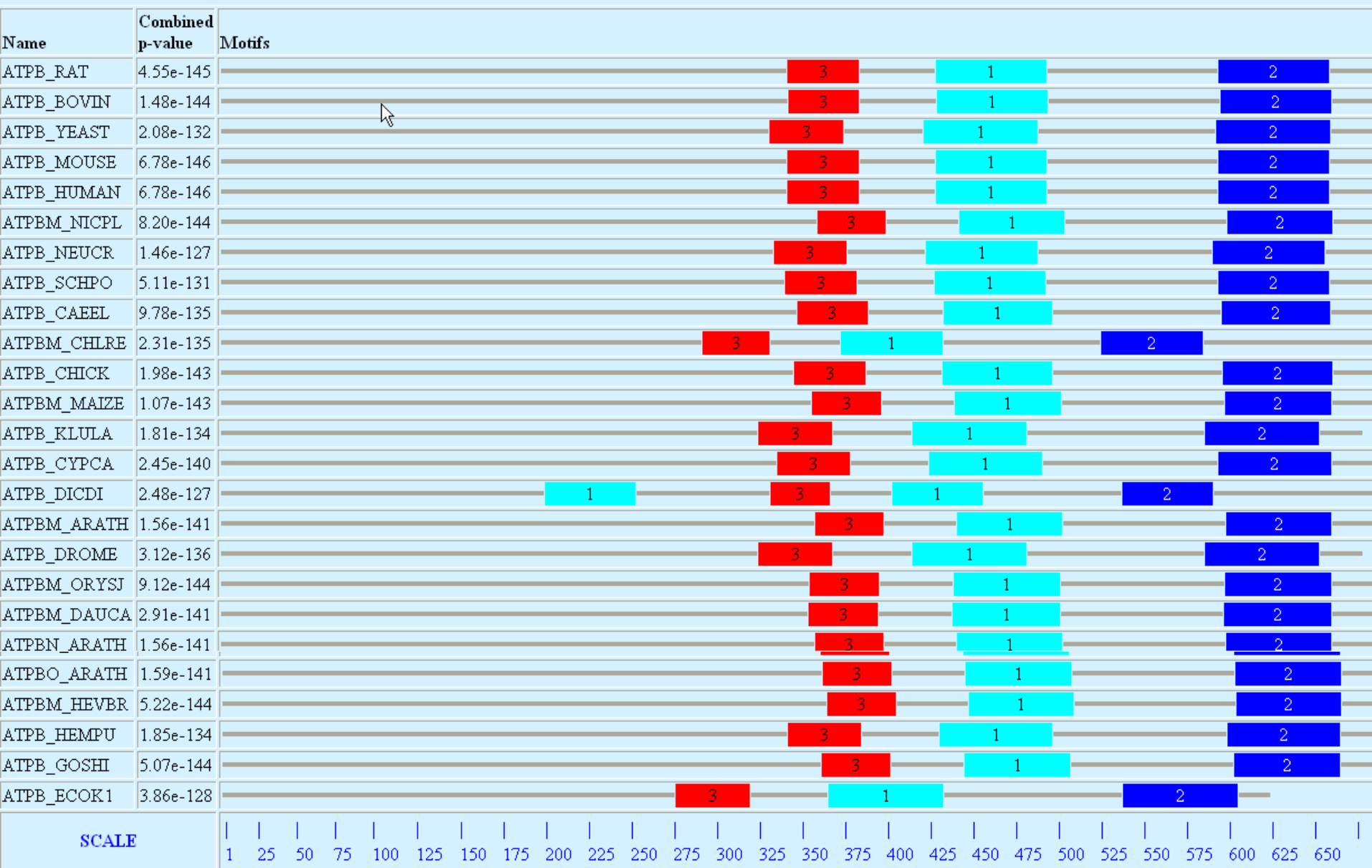
Browse by [taxonomy](#), [keyword](#), [gene ontology](#), [enzyme class](#) or [pathway](#) | Reduce sequence redundancy to [100%](#), [90%](#) or [50%](#) | [Customize display](#)

- > Quote terms: "atp synthase"
- > Restrict term "atp" to [domain](#), [protein family](#), [gene name](#), [gene ontology](#), [keyword](#), [protein name](#), [strain](#), [taxonomy](#)
- > Restrict term "synthase" to [domain](#), [protein family](#), [gene name](#), [gene ontology](#), [keyword](#), [protein name](#), [web resource](#)
- > Restrict term "atp" to [pathway](#)

Page 1 of 2 | [Next](#)

All	Accession	Entry name	Status	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	P10719	ATPB_RAT	★	ATP synthase subunit beta, mitochondrial (EC 3.6.3.14)	Atp5b	Rattus norvegicus (Rat)	529
<input type="checkbox"/>	P00829	ATPB_BOVIN	★	ATP synthase subunit beta, mitochondrial (EC 3.6.3.14)	ATP5B	Bos taurus (Bovine)	528
<input type="checkbox"/>	P56480	ATPB_MOUSE	★	ATP synthase subunit beta, mitochondrial (EC 3.6.3.14)	Atp5b	Mus musculus (Mouse)	529

三个保守Motif



Motif 1



MOTIF 1 width = 50 sites = 25 IIR = 3370 E-value = 3.3e-1023

SEQUENCE LOGO

Information Content

207.4 (bits)

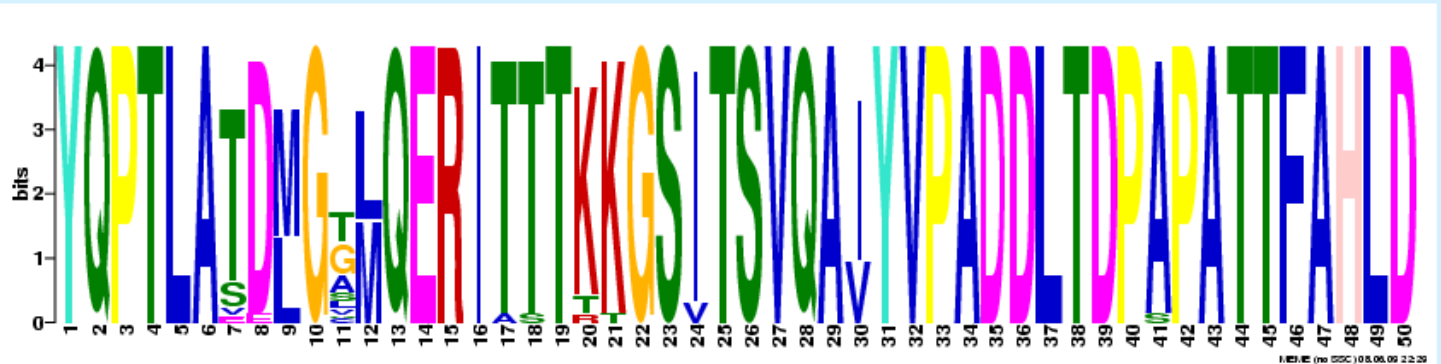
Relative Entropy

194.5 (bits)

Download LOGO

Without SSC: [\[EPS\]](#) [\[PNG\]](#)

With SSC: [\[EPS\]](#) [\[PNG\]](#)



NAME	START	P-VALUE	SITES
ATPB_HEMPU	325	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRGIA
ATPB_CYPCA	320	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_HUMAN	330	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_MOUSE	330	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_BOVIN	330	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_RAT	330	1.53e-62	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_DROME	307	1.02e-61	LLGRI PSAVG YQPTLATDMGSMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_CAEEL	339	1.02e-61	LLGRI PSAVG YQPTLATDMGSMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRGIA
ATPB_SCHPO	326	4.77e-61	LLGRI PSAVG YQPTLATDMGAMQERITTTKKGSI TSVQAVYV PADDLTD PAPATTFABLD ATTVLSRSIS
ATPB_CHICK	335	1.98e-60	LLGRI PSAVG YQPTLATDMGTMQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRAIA
ATPB_GOSHI	358	3.51e-60	LLGRI PSAVG YQPTLATDLGGLQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRQIS
ATPBM_HEVBR	363	3.51e-60	LLGRI PSAVG YQPTLATDLGGLQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRQIS
ATPBM_DAUCA	349	3.51e-60	LLGRI PSAVG YQPTLATDLGGLQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRQIS
ATPBM_ORYSJ	353	3.51e-60	LLGRI PSAVG YQPTLATDLGGLQERITTTKKGSI TSVQAIYV PADDLTD PAPATTFABLD ATTVLSRQIS

Motif 2



MOTIF 2 width = 50 sites = 25 lr = 3015 E-value = 9.3e-860

SEQUENCE LOGO

Information Content

180.5 (bits)

Relative Entropy

174.0 (bits)

Download LOGO

Without SSC: [\[EPS\]](#) [\[PNG\]](#)

With SSC: [\[EPS\]](#) [\[PNG\]](#)



NAME	START	P-VALUE	SITES
ATPB_HUMAN	459	1.25e-58	DKLTVSRARK IQRFLSQPFQVAEVFTGFMGKLVPLKETIKGFQQILAGEYDHLPEQAFYM VGPIEEAVAK
ATPB_MOUSE	459	1.25e-58	DKLTVSRARK IQRFLSQPFQVAEVFTGFMGKLVPLKETIKGFQQILAGEYDHLPEQAFYM VGPIEEAVAK
ATPB_CHICK	464	2.83e-58	DKLTVARARK IQRFLSQPFQVAEVFTGFMGKLVPLKETIKGFKQILAGEYDHLPEQAFYM VGPIEEAVAK
ATPB_GOSHI	487	4.55e-58	DKLTVARARK IQRFLSQPFHVAEVFTGAPGKYVELKESITSFQGVLDGKYDDLPEQSFYM VGGIEEVIK
ATPBM_HEVBR	492	4.55e-58	DKLTVARARK IQRFLSQPFHVAEVFTGAPGKYVELKESITSFQGVLDGKYDDLPEQSFYM VGGIDEVIK
ATPBM_NICPL	490	7.25e-58	DKMTVARARK IQRFLSQPFHVAEVFTGAPGKYVDLKEINSFQGVLDGKYDDLSEQSFYM VGGIDEVIK
ATPBM_ORYSJ	482	8.46e-58	DKLTVARARK IQRFLSQPFHVAEVFTGAPGKYVELKESVNSFQGVLDGKYDDLPEQSFYM VGGIEEVIK
ATPB_RAT	459	8.46e-58	DKLTVSRARK IQRFLSQPFQVAEVFTGFMGKLVPLKETIKGFQQILAGDYDHLPEQAFYM VGPIEEAVAK
ATPBM_MAIZE	483	9.85e-58	DKLTVARARK IQRFLSQPFHVAEVFTGAPGKYVELKESVKSFQGVLDGKYDDLPEQSFYM VGGIEEVIK
ATPB_BOVIN	459	2.79e-57	DKLTVSRARK IQRFLSQPFQVAEVFTGFMGKLVPLKETIKGFQQILAGEYDHLPEQAFYM VGPIEEAVAK
ATPBM_DAUCA	477	1.32e-56	DKLTVARARK IQRFLSQPFHVAEIFTGAPGKYVELKECVTSFQGVLDGKYDDLPEQSFYM LGGIEEVIK
ATPBO_ARATH	489	2.60e-56	DKLTVARARK IQRFLSQPFHVAEIFTGAPGKYVDLKENINSFQGLLDGKYDDLSEQSFYM VGGIDEVVK
ATPBN_ARATH	486	2.60e-56	DKLTVARARK IQRFLSQPFHVAEIFTGAPGKYVDLKENINSFQGLLDGKYDDLSEQSFYM VGGIDEVVK
ATPBM_ARATH	486	2.60e-56	DKLTVARARK IQRFLSQPFHVAEIFTGAPGKYVDLKENINSFQGLLDGKYDDLSEQSFYM VGGIDEVVK
ATPB_ECOK1	396	1.03e-54	DKLVVARARK IQRFLSQPFQVAEVFTGSPGKYVSLKDTIRGFKGIMEGEYDHLPEQAFYM VGSIEEAVEK

Motif 3



MOTIF 3 width = 32 sites = 25 hr = 2149 E-value = 1.3e-627

SEQUENCE LOGO

Information Content

131.9 (bits)

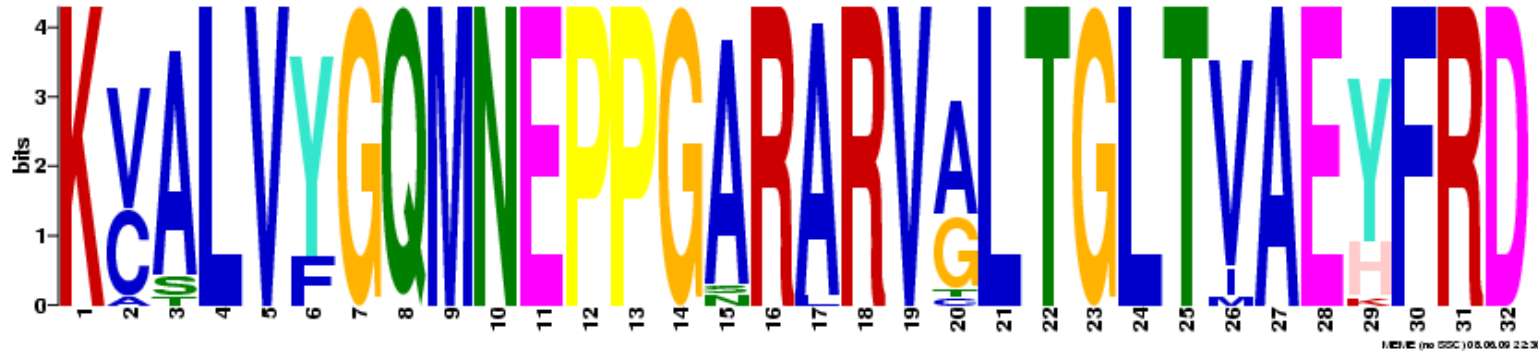
Relative Entropy

124.0 (bits)

Download LOGO

Without SSC: [\[EPS\]](#) [\[PNG\]](#)

With SSC: [\[EPS\]](#) [\[PNG\]](#)



NAME	START	P-VALUE	SITES
ATPBO_ARATH	293	2.33e-40	IKLGEKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBN_ARATH	290	2.33e-40	IKLGEKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_ARATH	290	2.33e-40	IKLGEKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPB_GOSHI	291	4.00e-40	IKLGEKQADS KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_HEVBR	296	4.00e-40	IKLGDKQADS KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_ORYSJ	286	4.00e-40	IKLGDKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_MAIZE	287	4.00e-40	IKLDDKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_NICPL	294	4.00e-40	IKLGEKQSES KCALVYQGMNEPPGARARVGLTGLTVAEYFRD AEGQDVLLFI
ATPBM_CHLRE	242	1.73e-39	IKLGDKRGS KCTLVYQGMNEPPGARARVGLTGLTVAEYFRD VEGQDVLLFV
ATPB_HEMPU	258	5.18e-39	GVISLKDDTS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_DROME	240	5.18e-39	GVISLKDKTS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_CYPCA	253	5.18e-39	GVINLKDTS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_CHICK	268	5.18e-39	GVINLKDATS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_HUMAN	263	5.18e-39	GVINLKDATS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_MOUSE	263	5.18e-39	GVINLKDATS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI
ATPB_BOVIN	263	5.18e-39	GVINLKDATS KVALVYQGMNEPPGARARVGLTGLTVAEYFRD QEGQDVLLFI

利用WEBLAB PEPSTATS进行氨基酸组成分析

- WEBLAB PEPSTATS of ATP beta1 from 1 to 557
- Molecular weight = 59853.16
- Residues = 557
- Average Residue Weight = 107.456
- Charge = -2.0
- Isoelectric Point = 6.2110
- A280 Molar Extinction Coefficient = 16640
- A280 Extinction Coefficient 1mg/ml = 0.28
- Improbability of expression in inclusion bodies = 0.689

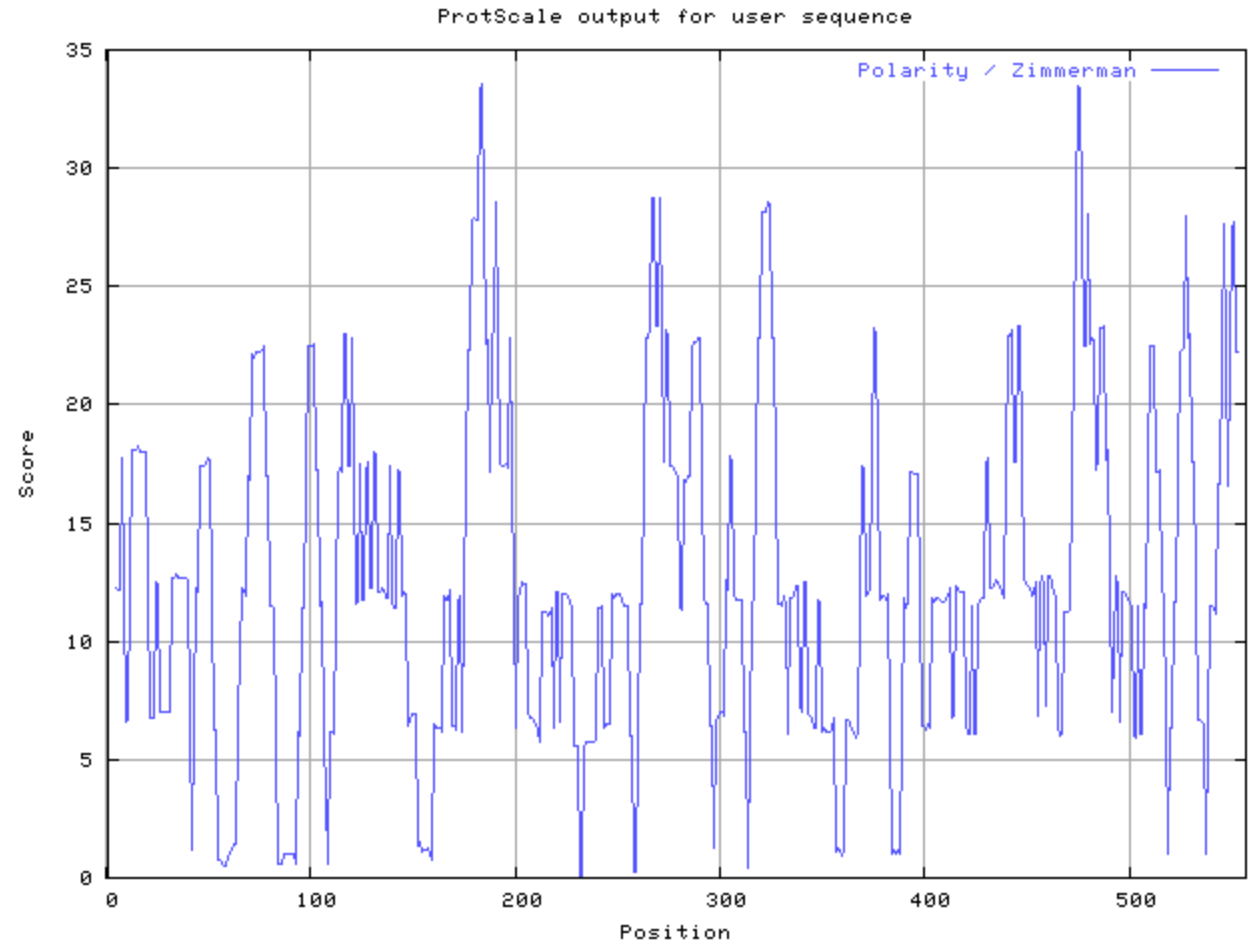
• Residue	Number	Mole%	DayhoffStat
• A = Ala	52	9.336	1.086
• B = Asx	0	0.000	0.000
• C = Cys	3	0.539	0.186
• D = Asp	30	5.386	0.979
• E = Glu	35	6.284	1.047
• F = Phe	20	3.591	0.997
• G = Gly	52	9.336	1.111
• H = His	10	1.795	0.898
• I = Ile	36	6.463	1.436
• J = ---	0	0.000	0.000
• K = Lys	26	4.668	0.707
• L = Leu	51	9.156	1.237
• M = Met	10	1.795	1.056
• N = Asn	12	2.154	0.501
• O = ---	0	0.000	0.000
• P = Pro	35	6.284	1.208
• Q = Gln	24	4.309	1.105
• R = Arg	32	5.745	1.172
• S = Ser	37	6.643	0.949
• T = Thr	36	6.463	1.060
• U = ---	0	0.000	0.000
• V = Val	43	7.720	1.170
• W = Trp	0	0.000	0.000
• X = Xaa	0	0.000	0.000
• Y = Tyr	13	2.334	0.686
• Z = Glx	0	0.000	0.000

• Property	Residues	Number	Mole%
• Tiny	(A+C+G+S+T)	180	32.316
• Small	(A+B+C+D+G+N+P+S+T+V)	300	53.860
• Aliphatic	(A+I+L+V)	182	32.675
• Aromatic	(F+H+W+Y)	43	7.720
• Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	315	56.553
• Polar	(D+E+H+K+N+Q+R+S+T+Z)	242	43.447
• Charged	(B+D+E+H+K+R+Z)	133	23.878
• Basic	(H+K+R)	68	12.208
• Acidic	(B+D+E+Z)	65	11.670

利用ProtScale对目的蛋白的理化性质分析

- 极性

该蛋白有六个
高级极性区

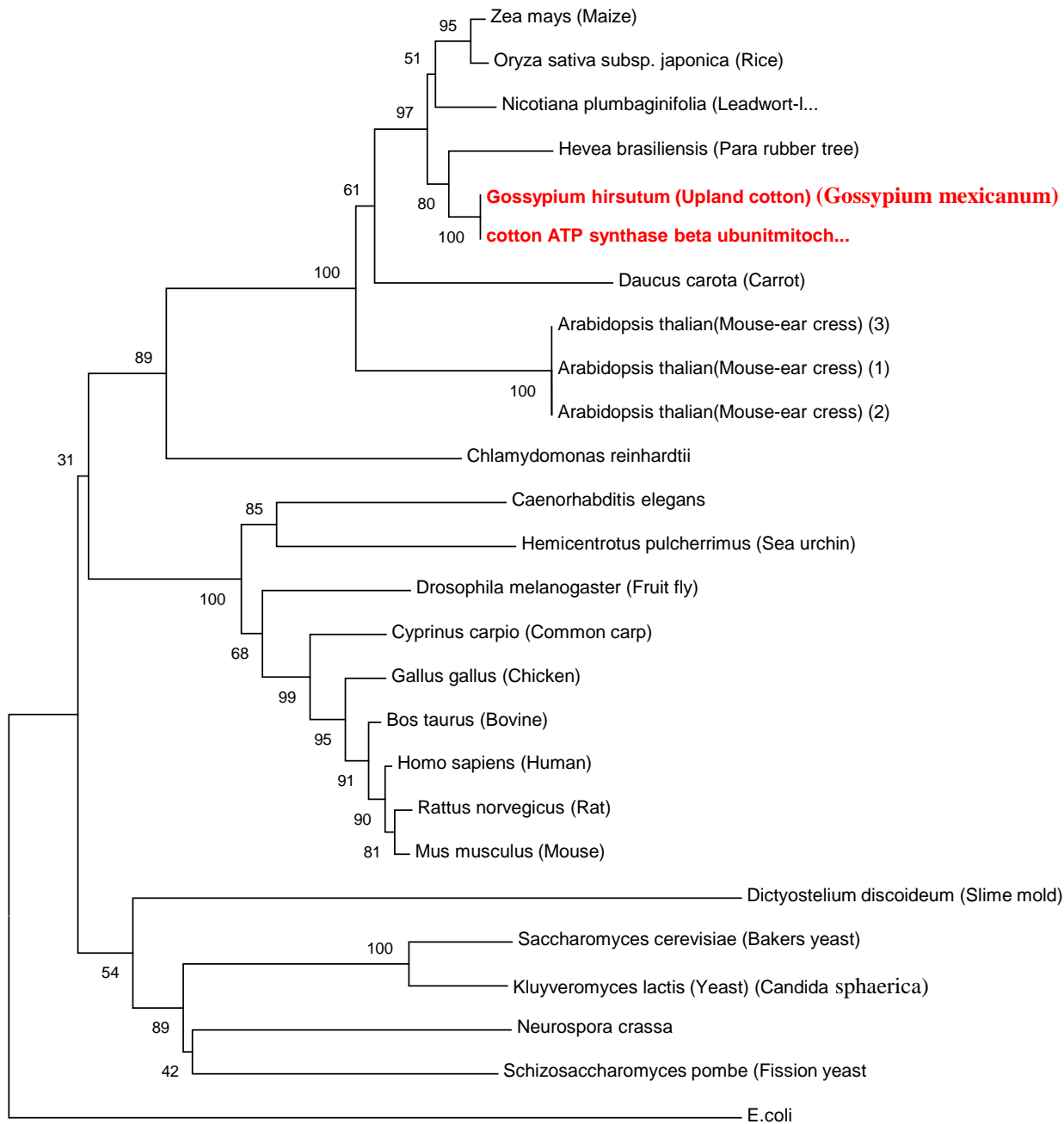


系统发生分析

特征序列的选择

- 在**Expasy**中得到的**38**条相关序列，**23**条为**ATP合酶beta**亚基序列
- 另外找到一条**ATPB_ECOK1 ATP synthase subunit beta**序列，作为外类群
- 用这**24**条序列和我们的目的序列进行比对，在**MEGA**中用**bootstrap NJ**构建系统发生树

- 从以上结果可以看出，**ATP合酶beta**亚基从细菌到高等动物中的序列相似性很高。



0.05

- 从上面看出，我们的序列和Gossypium hirsutum (Upland cotton) (Gossypium mexicanum) 在同一分支上，且它们之间的bootstrap值为100。将这两个序列进行needle全局比对，得到：

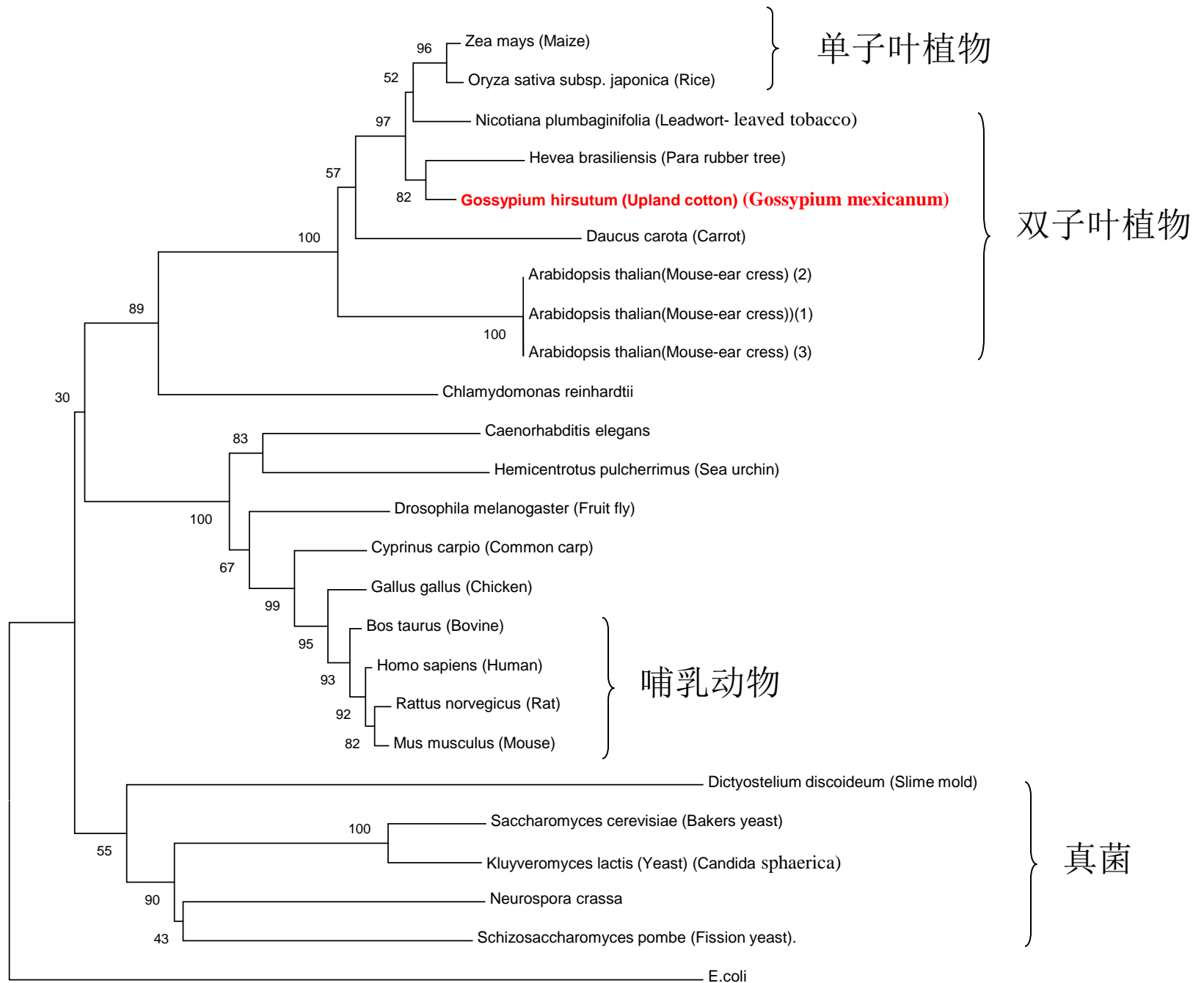
Result: View:

557	2818.0	557/557 (100.0%)	557/557 (100.0%)	0/557 (0.0%)
-----	--------	------------------	------------------	--------------

```

Gossypium      1 MASRRLSSFLRSTSRRLSKSPSPTRRLSPSPTRRPFPCPYFLSRDAE    50
  |||
cotton          1 MASRRLSSFLRSTSRRLSKSPSPTRRLSPSPTRRPFPCPYFLSRDAE    50
  |||
Gossypium     51 YSTSAAAAASAPSQPPPAKKGKIGKITDEFTGAGAIGQVCQVIGAVVDVR  100
  |||
cotton        51 YSTSAAAAASAPSQPPPAKKGKIGKITDEFTGAGAIGQVCQVIGAVVDVR  100
  |||
Gossypium    101 FDEGLPPILTALVMDHSIRLVLEVAQHLGENMVRTIAMDGTEGLVRGQP  150
  |||
cotton      101 FDEGLPPILTALVMDHSIRLVLEVAQHLGENMVRTIAMDGTEGLVRGQP  150
  |||
Gossypium    151 VLNTGSPITVPVGRATLGRINIVIGEPIDKGEKLTDFHFLPIHREAPAFV  200
  |||
  
```

- 可以看出，这两个序列是完全相同的，所以将其合并，重新作树如下：



目的蛋白三级结构分析


- 寻找结构已知的同源蛋白
- 预测目的蛋白的三维结构
- 保守序列和突变位点分析

寻找结构已知的同源蛋白

- 用翻译出的蛋白序列进行blastp的结果，与其氨基酸一致性超过60的序列多于500个。

Sequences producing significant alignments:	Score (Bits)	E Value	
emb CAN69512.1 hypothetical protein [Vitis vinifera]	<u>990</u>	0.0	
ref NP_001043900.1 Os01g0685800 [Oryza sativa (japonica cult...]	<u>983</u>	0.0	UG
sp P17614.1 ATPBM_NICPL RecName: Full=ATP synthase subunit be...	<u>983</u>	0.0	
qfb EEF30158.1 ATP synthase beta subunit, putative [Ricinus c...]	<u>981</u>	0.0	
qfb AAD03392.1 mitochondrial ATPase beta subunit [Nicotiana s...]	<u>979</u>	0.0	
ref NP_001056261.1 Os05g0553000 [Oryza sativa (japonica cult...]	<u>979</u>	0.0	UG
ref NP_001151807.1 ATP synthase beta chain [Zea mays] >qfb AC...	<u>977</u>	0.0	G
dbj BAA01372.1 mitochondrial F1-ATPase beta subunit [Oryza s...]	<u>974</u>	0.0	
ref XP_002280824.1 PREDICTED: hypothetical protein [Vitis vi...]	<u>973</u>	0.0	UG
qfb AAD03394.1 ATPase beta subunit [Nicotiana sylvestris]	<u>973</u>	0.0	
qfb AAD03393.1 ATPase beta subunit [Nicotiana sylvestris]	<u>971</u>	0.0	
ref XP_002283951.1 PREDICTED: hypothetical protein [Vitis vi...]	<u>969</u>	0.0	UG
emb CAO42228.1 unnamed protein product [Vitis vinifera]	<u>964</u>	0.0	
ref NP_001105340.1 mitochondrial F-1-ATPase subunit 2 [Zea m...]	<u>964</u>	0.0	UG
qfb AAD03391.1 mitochondrial ATPase beta subunit [Nicotiana s...]	<u>961</u>	0.0	
ref YP_002315002.1 mitochondrial beta subunit of F1 ATP synt	<u>954</u>	0.0	UG

选择分值最高的ATP合酶beta亚基

```
>  sp|P17614.1|ATPBM\_NICPL RecName: Full=ATP synthase subunit beta, mitochondrial; Flags:  
Precursor  
emb|CAA26620.1| ATP synthase beta subunit [Nicotiana plumbaginifolia]  
Length=560  
  
Score = 983 bits (2541), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 492/560 (87%), Positives = 520/560 (92%), Gaps = 3/560 (0%)
```

Nicotiana plumbaginifolia皱叶烟草

Search for

Format: **GenPept** [FASTA](#) [Graphics](#) [More Formats](#) ▼

[Download](#) ▼ [Save](#) ▼ [Links](#) ▼

Swiss-Prot P17614.1

RecName: Full=ATP synthase subunit beta, mitochondrial; Flags: Precursor

[Comment](#) [Features](#) [Sequence](#)

LOCUS P17614 560 aa linear PLN 03-MAR-2009
 DEFINITION RecName: Full=ATP synthase subunit beta, mitochondrial; Flags: Precursor.
 ACCESSION P17614
 VERSION P17614.1 GI:114421
 DBSOURCE UniProtKB: locus ATPBM_NICPL, accession [P17614](#);
 class: standard.
 created: Aug 1, 1990.
 sequence updated: Aug 1, 1990.
 annotation updated: Mar 3, 2009.
 xrefs: [X02868.1](#), [CAA26620.1](#), [A24355](#), [1PYV_A](#)
 xrefs (non-sequence databases): PDBsum:1PYV, SMR:P17614,
 BRENDA:3.6.3.14, GO:[0005743](#), GO:[0045261](#), GO:[0005524](#), GO:[0046933](#),
 GO:[0046961](#), GO:[0008553](#), GO:[0046872](#), GO:[0015986](#), InterPro:[IPR003593](#),
 InterPro:[IPR005722](#), InterPro:[IPR000793](#), InterPro:[IPR004100](#),
 InterPro:[IPR000194](#), PANTHER:PTHR15184:SF8, Pfam:PF00006,
 Pfam:PF00306, Pfam:PF02874, SMART:SM00382, TIGRFAMs:TIGR01039,
 PROSITE:PS00152
 KEYWORDS 3D-structure; ATP synthesis; ATP-binding; CF(1); Hydrogen ion transport; Hydrolase; Ion transport; Membrane; Mitochondrion; Mitochondrion inner membrane; Nucleotide-binding; Transit peptide; Transport.
 SOURCE *Nicotiana plumbaginifolia* (curled-leaved tobacco)
 ORGANISM [Nicotiana plumbaginifolia](#)

Change Region Shown

Customize View






BLAST Sequence

Find regions of similarity between sequences using BLAST.

Conserved Domains

View conserved domains detected using CD-search.

Recent Activity

-  RecName: Full=ATP synthase subunit beta, mitochondrial; Flags: Precursor
-  [ATPB_GOSHI ATP synthase s...](#) BLAS
-  [H.brasiliensis mR...\[gi:18830\]](#)
-  [225424141.18830.224099436...](#) (37) Nucleoti
-  [1](#) BLAS

- BLink
- Conserved Domains
- Full text in PMC
- Identical Proteins
- Related Structure**
- Related Sequences
- Domain Relatives
- PubMed
- Taxonomy
- LinkOut

2QE7为皱叶烟草 (Nicotiana plumbaginifolia) 的同源蛋白

Structures related to [\[gi114421|sp|P17614.1|ATPBM_NICPL\]](#)

RecName: Full=ATP synthase subunit beta, mitochondrial; Flags: Precursor

17 Low redundancy structures identified

View **Low redundancy** sequences, sort by **BLAST E-value** and display as **graphic** at **20** sequences per page [Go](#)

[« Previous page](#)

Jump to page **1** of 1

[Next page »](#)

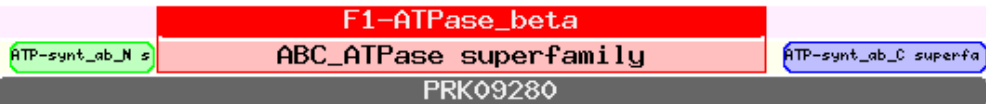


Protein Families

Specific hits

Superfamilies

Multi-domains



E-value

Structures



2QE7_D



1e-125



2QE7_E



1e-125



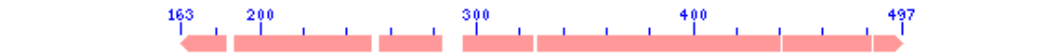
2QE7_F



1e-125



2OBL_A



1e-38



ATPbeta1与2QE7序列比对

- 将ATPbeta1的蛋白序列与**2QE7的D链(beta 亚基)**进行比对:

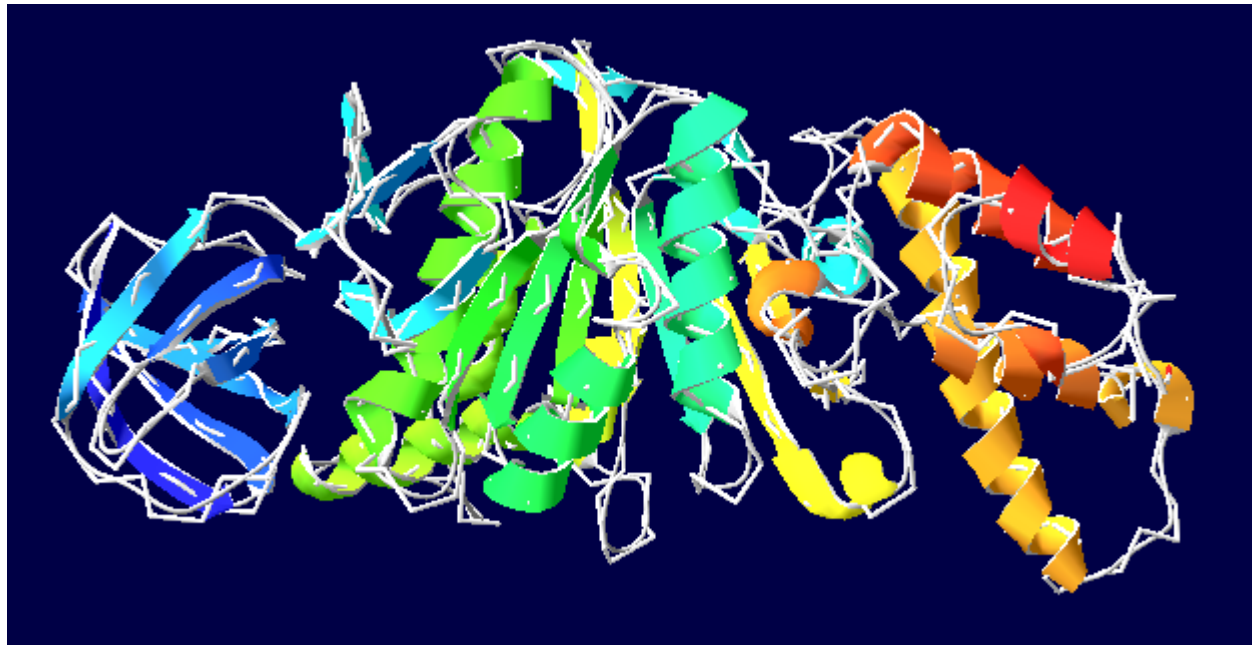
Pairwise Alignment Result

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
559	1667.5	326/559 (58.3%)	390/559 (69.8%)	99/559 (17.7%)

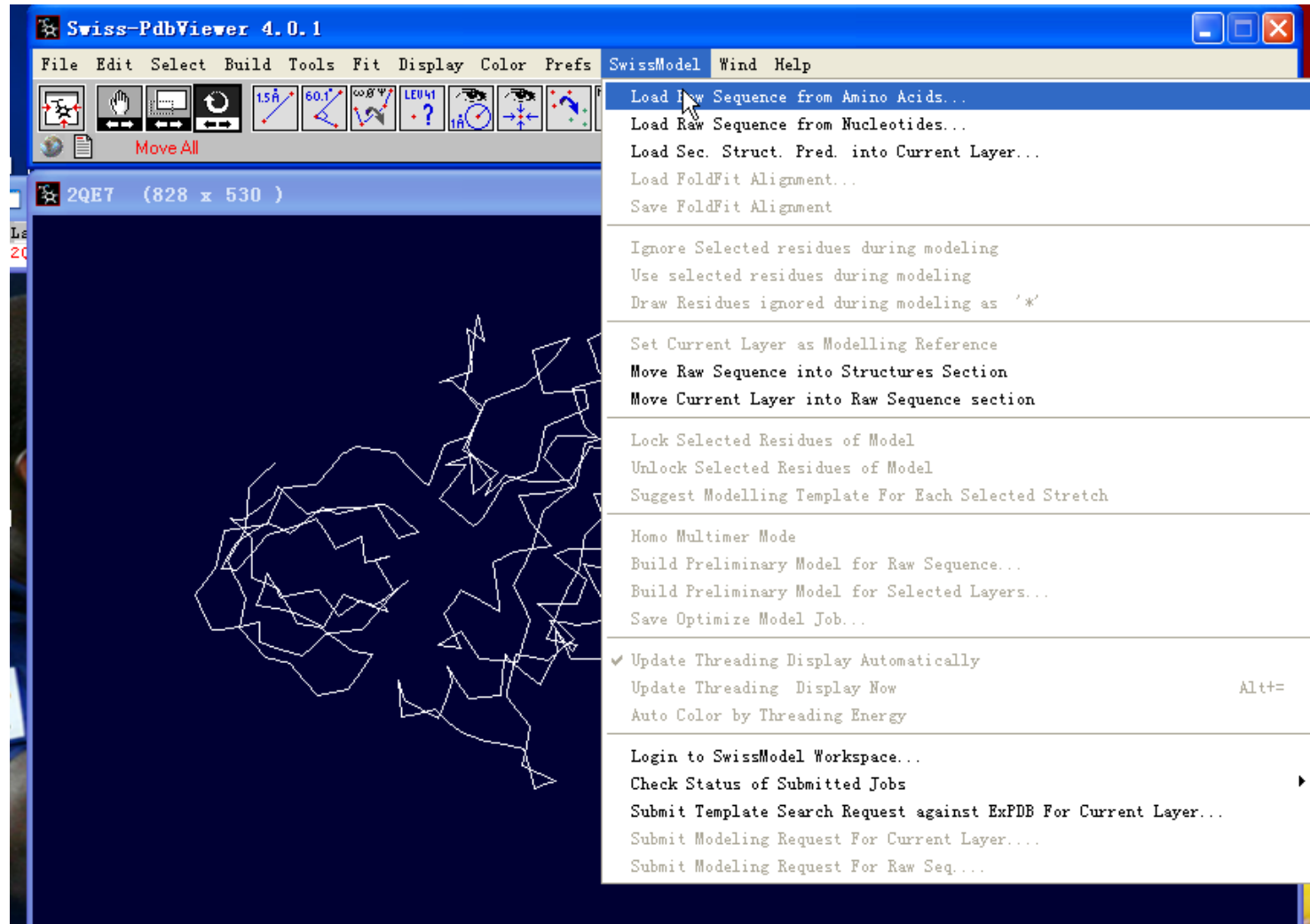
2QE7的D链与**ATPbeta1**的相似性很高，而且同为ATP合酶的beta 亚基，功能相同，所以两者为同源基因。

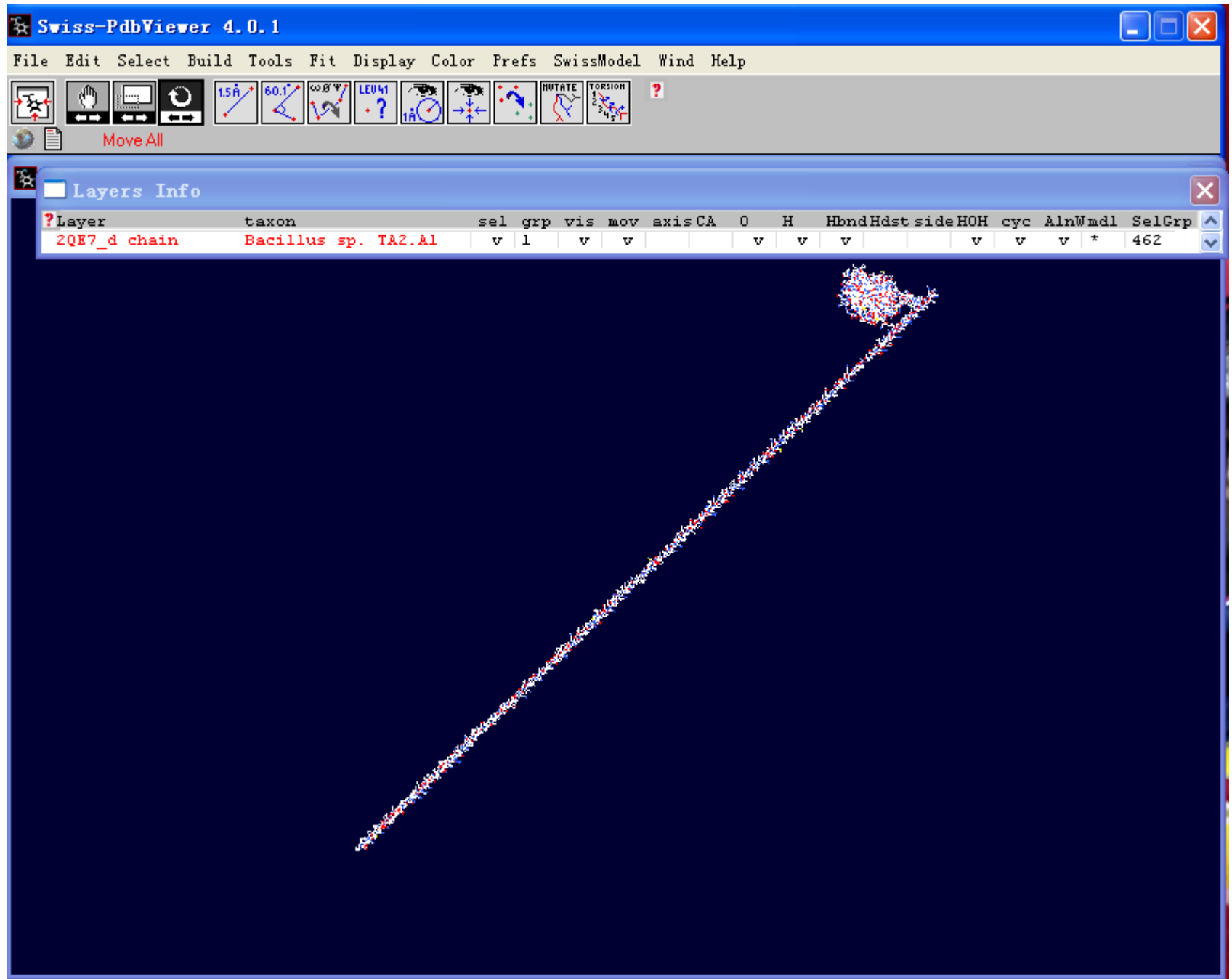
预测目的蛋白的三维结构

- 用SPDBV软件来构建目的蛋白的模型
 1. 从PDB网站下载2QE7的蛋白三维结构模型，保留D链



2. 利用SPDBV中的SwissModel工具进行结构模拟







Move All

Layers Info

? Layer	taxon
2QE7_d_chain	Bacillus
ATPE_GOSHI	N/A

Fit Raw Sequence
 Align Selected Columns (using MUSCLE)
 Align All
 Align Selected Sequences onto non Selected Sequences...
 Realign Raw Sequence Selection against Ref. Seq. Selection

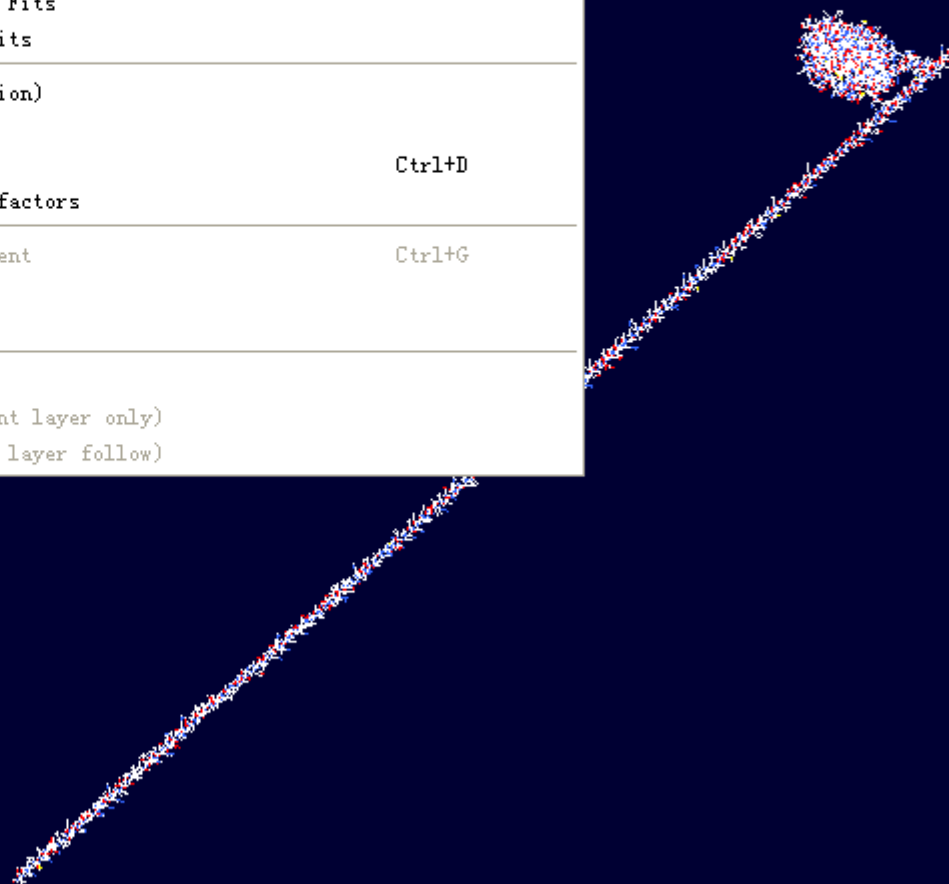
Magic Fit Ctrl+M
 Iterative Magic Fit Ctrl+Shift+M
 Explore Fragment Alternate Fits
 Explore Domain Alternate Fits

Fit Molecules (from selection)
 Improve Fit
 Calculate RMS... Ctrl+D
 Set Layer Std Dev. into B-factors

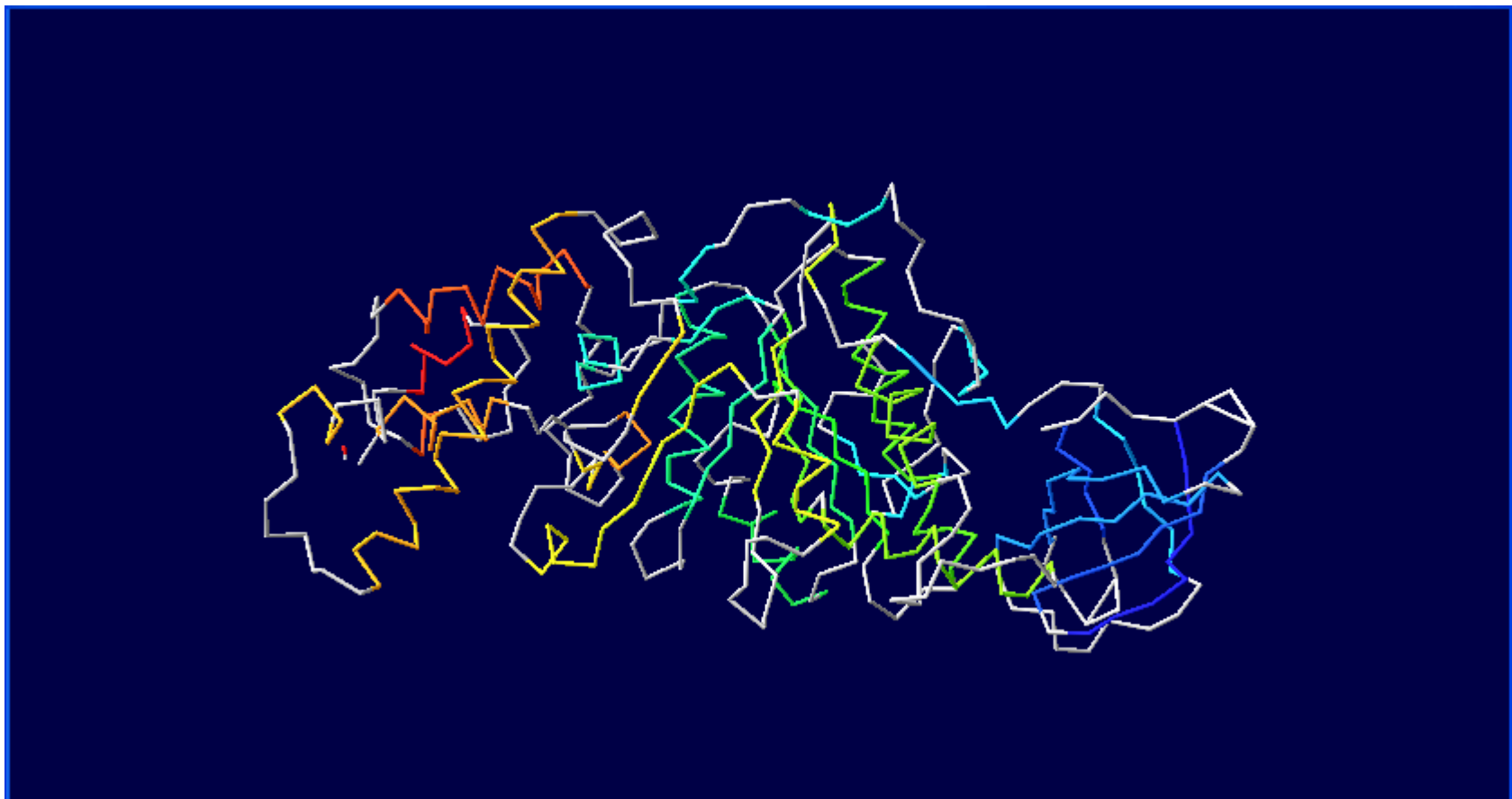
Generate Structural Alignment Ctrl+G
 Compress Gaps
 Reset Alignment

Validate Current Fit
 UnFit Current Layer (current layer only)
 UnFit Current Layer (other layer follow)

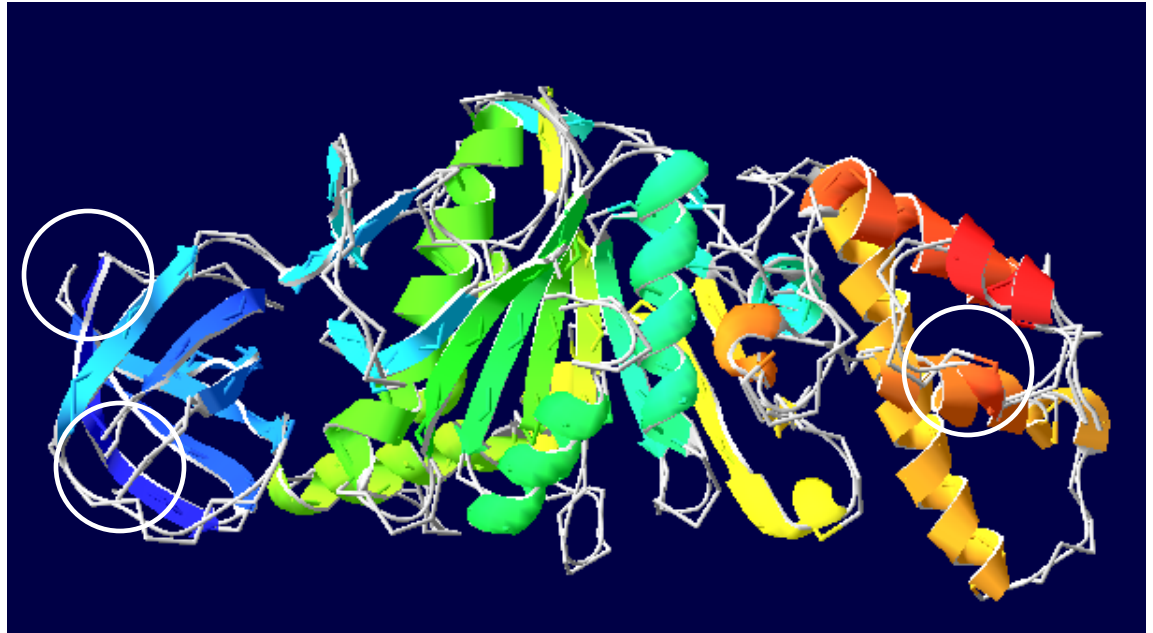
yc	AlnWmdl	SelGrp
v	v	* 462
v	v	0



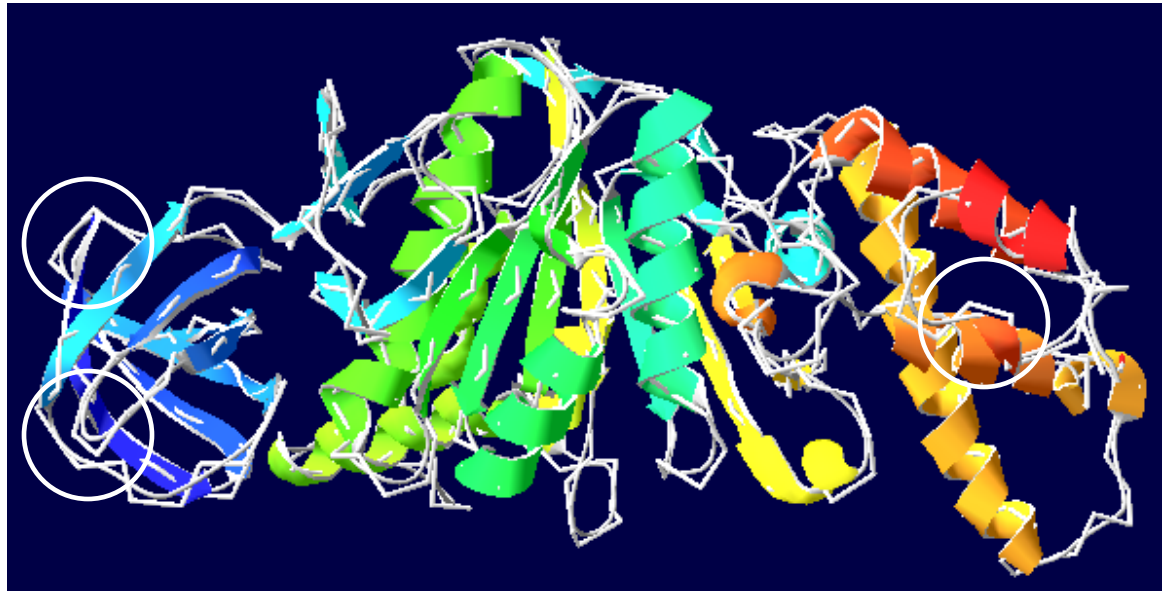
Magic fit结果:



ATP beta 1

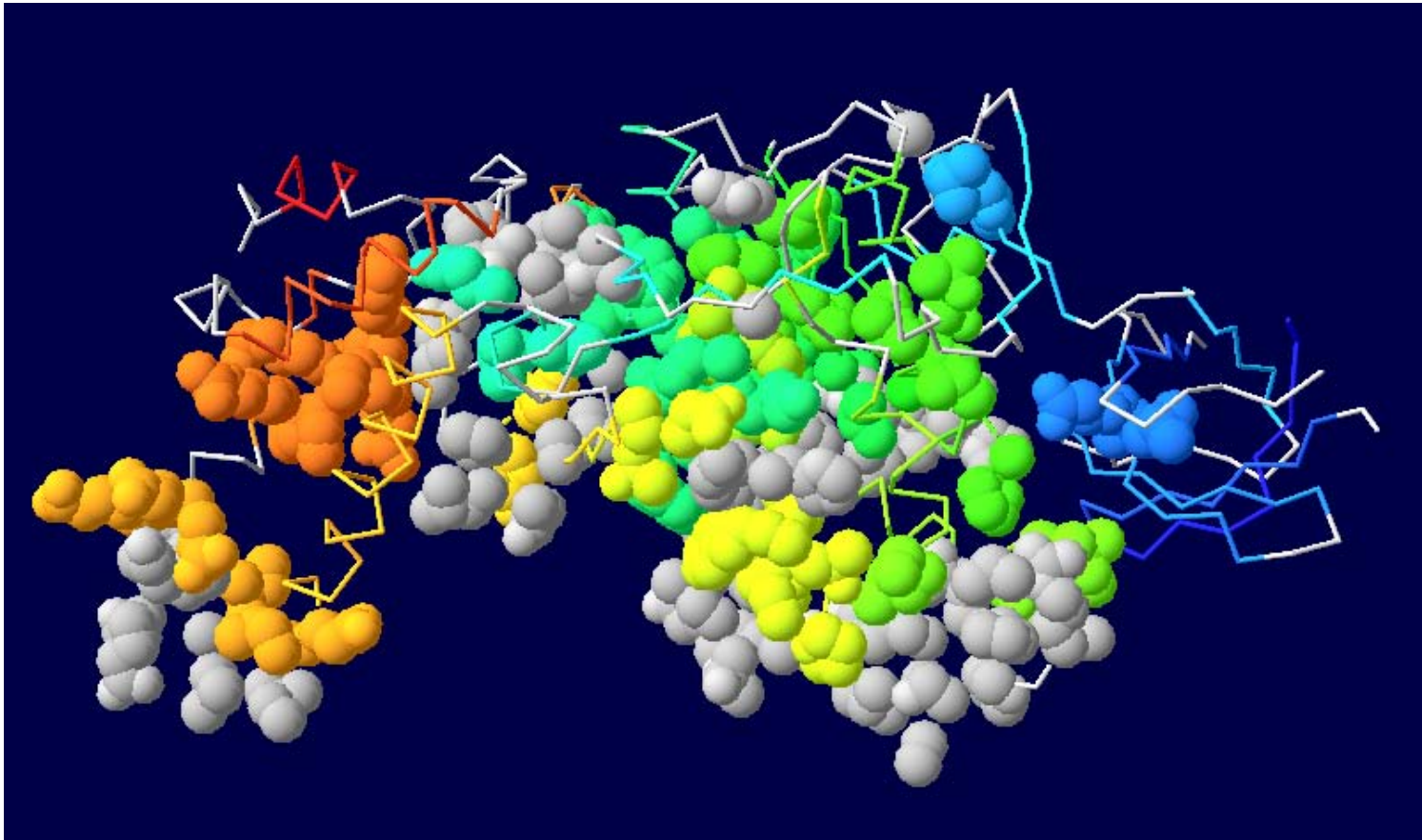


2QE7_D chain



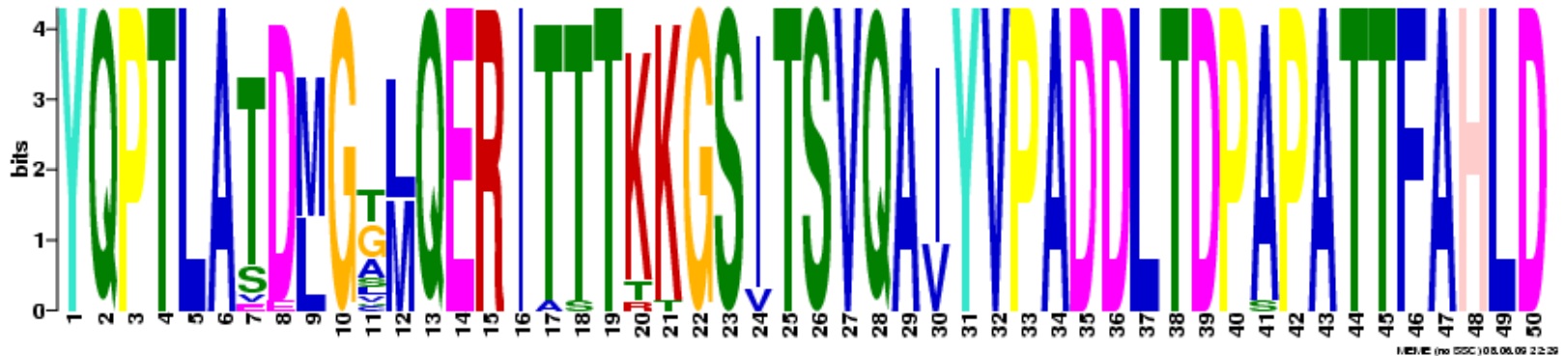
保守序列和突变位点分析

保守结构域（球状）



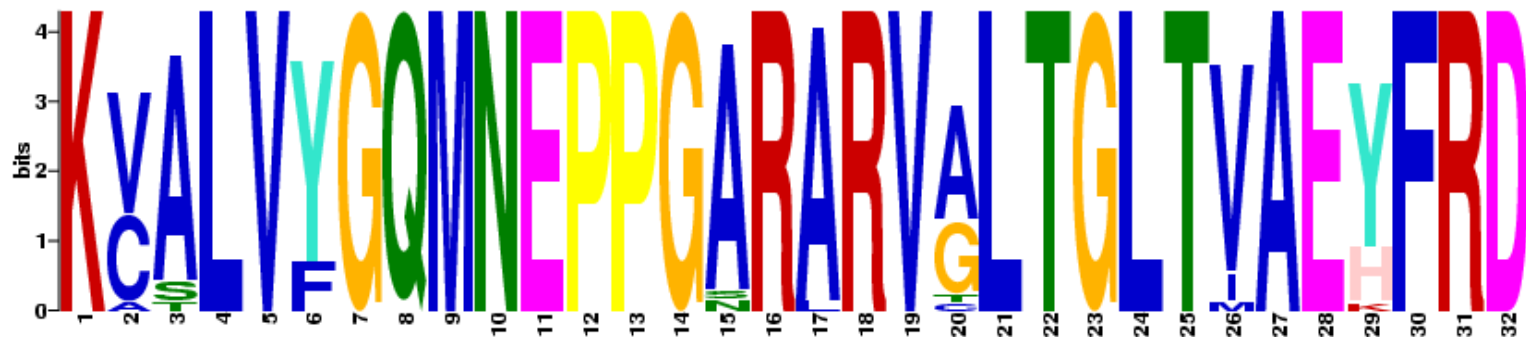
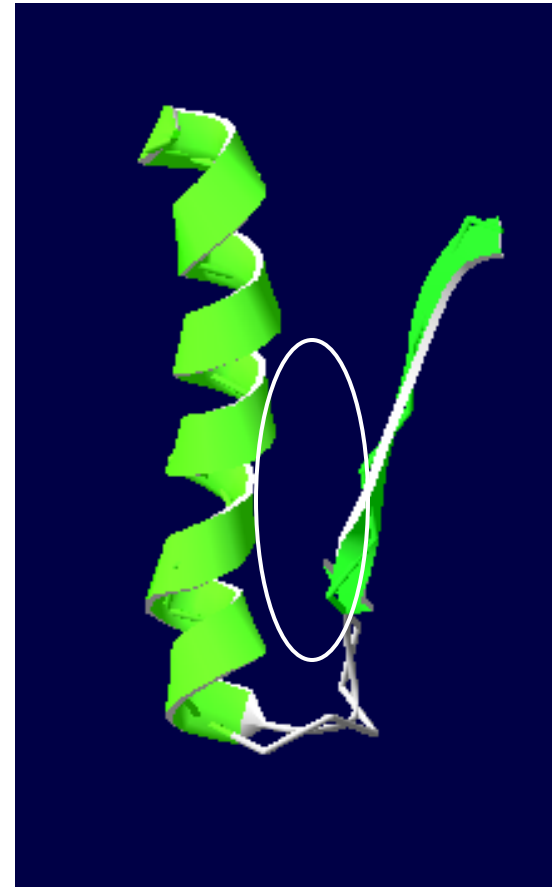
Motif1

Y359-D408,
富含Thr, Ser
和极性氨基酸,
且形态上为口
袋状, 可能为
催化中心。

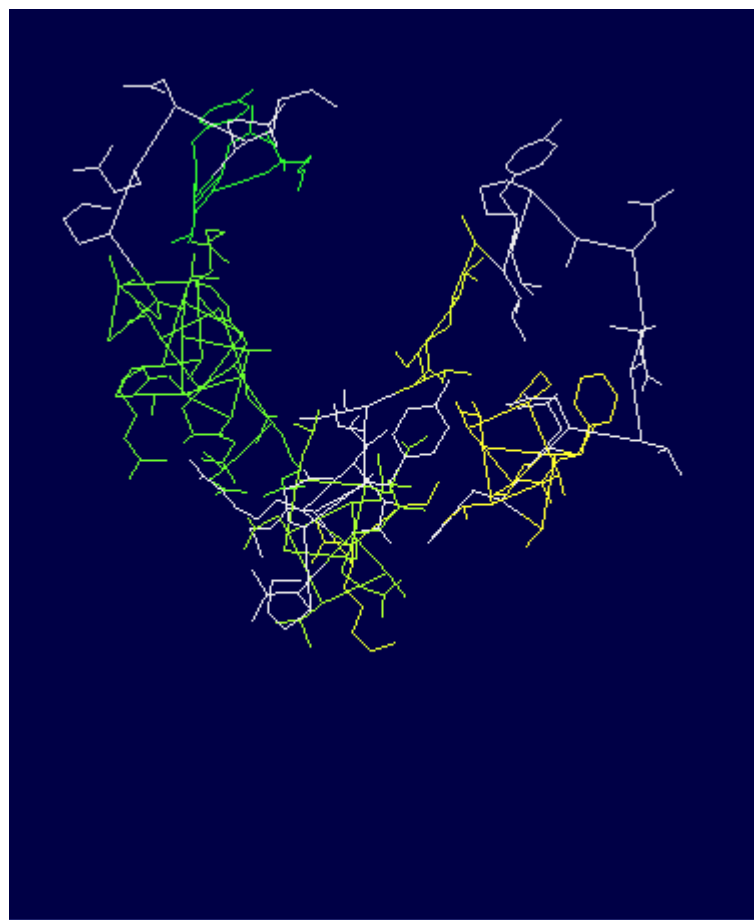
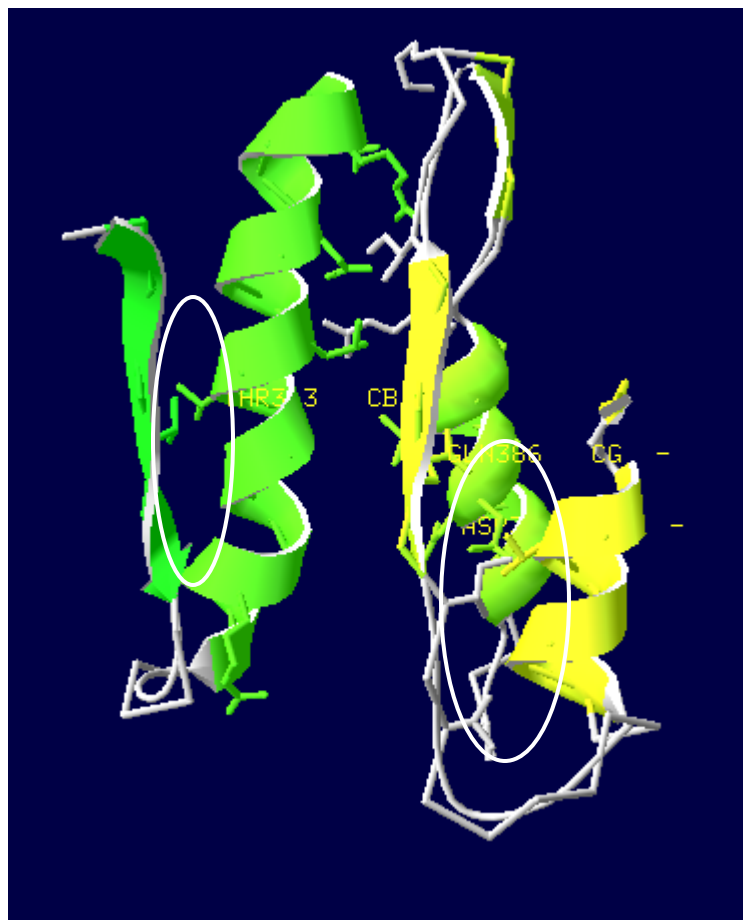


Motif3

K292-D323, 富含Thr, Ser和极性氨基酸, 且形态上为钳状, 可能为催化中心。



结构叠合



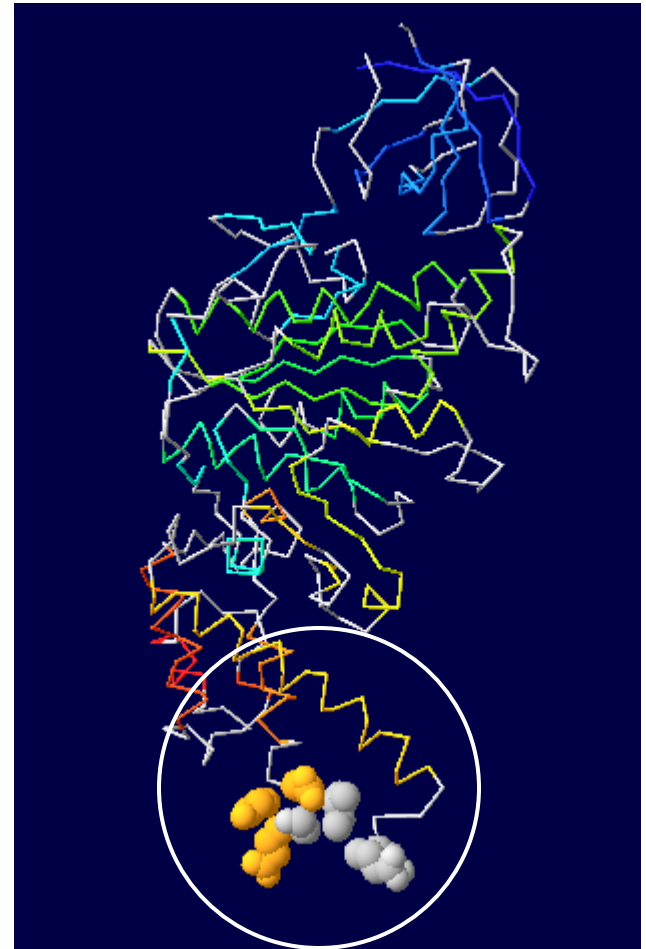
猜 测

- 最为保守的**Motif1**和**Motif3**组合成双臂结构分别结合磷酸分子和**ADP**分子，并催化其结合为**ATP**。

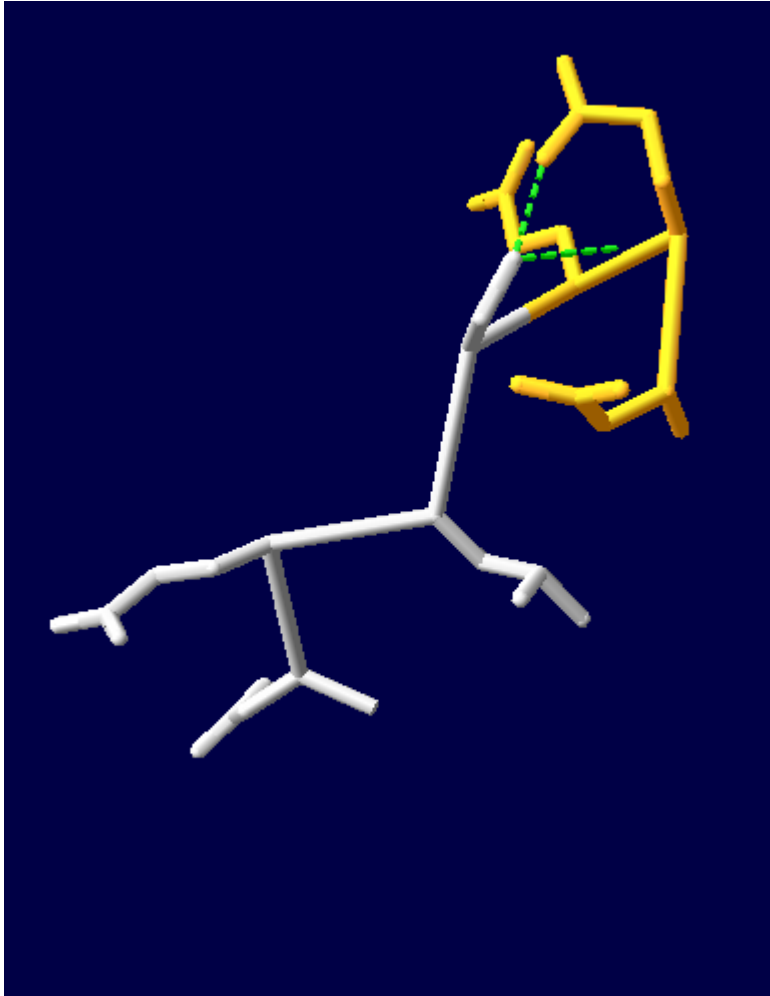
“DELSEED” motif

位于蛋白C端的“DELSEED”模体是与 γ 亚基结合的结构域。提供了头部F1转动的动力。

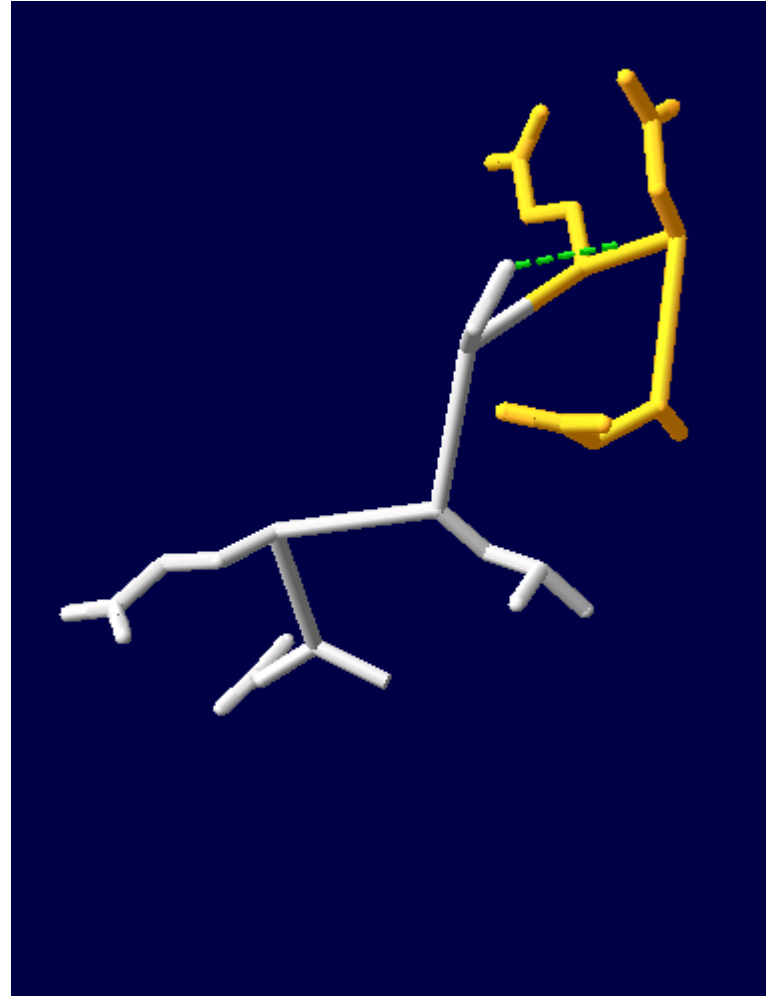
在高等植物中这一模体为“DELSEDD”。



DELSEED



DELSEDD



猜 测

- 在微生物及动物中， **DELSEED**模体第六位的**Glu**使其分子间多形成一个氢键，可能有助于与 γ 亚基的结合，产生更强的动力，催化**ATP**的合成。

结 论

- ATP合酶beta亚基是极度保守的蛋白，从细菌到高等动物中的序列相似性很高。
- ATP合酶beta亚基通过“DELSE(ED)D”模体与gamma亚基相结合，通过质子运动产生动力催化合成或分解ATP。
- 在微生物及动物中，因为在保守位点的氨基酸改变，ATP合酶可能具有更强活性。
- ATP合酶最为保守的**Motif1**和**Motif3**组合成双臂结构，可能是催化ATP合成的部位
- ATP合酶在棉花纤维发育时期表达上调，可能是为了产生更多能量，促进纤维细胞的伸长，及细胞内物质的合成。

实验设计

- 针对“**DELSEDD**”模体设计点突变（如**D→E**），分析纤维发育过程中**ATP/ADP**的比率变化及纤维细胞的长短变化。
- 针对**Motif1**双臂结构中的关键氨基酸（如**Asp366, Gln386, Thr313**）设计点突变，分析纤维发育过程中**ATP/ADP**的比率变化及纤维细胞的长短变化。
- 在酵母中表达点突变的蛋白，分析酵母细胞形态变化。
- 根据以上保守结构底物类似物，处理材料竞争性抑制底物结合，分析同上。

参考文献

- **K. Y. Hara, Y. Kato-Yamada, Y. Kikuchi, T. Hisabori, and M. Yoshida**
The Role of the beta DELSEED Motif of F1-ATPase.
J. Biol. Chem., June 22, 2001; 276(26): 23969 - 23973.
- **H Omote, MY Park, M Maeda and M Futai**
The alpha/beta subunit interaction in H(+)-ATPase (ATP synthase). An Escherichia coli alpha subunit mutation (Arg-alpha 296-->Cys) restores coupling efficiency to the deleterious beta subunit mutant (Ser-beta 174-->Phe)
J. Biol. Chem., Vol. 269, Issue 14, 10265-10269, 1994
- **M. K. Al-Shawi, C. J. Ketchum, and R. K. Nakamoto**
Energy Coupling, Turnover, and Stability of the F0F1 ATP Synthase Are Dependent on the Energy of Interaction between gamma and beta Subunits
J. Biol. Chem., January 24, 1997; 272(4): 2300 - 2306.
- **M. Nakanishi-Matsui, S. Kashiwagi, T. Ubukata, A. Iwamoto-Kihara, Y. Wada, and M. Futai**
Rotational Catalysis of Escherichia coli ATP Synthase F1 Sector: STOCHASTIC FLUCTUATION AND A KEY DOMAIN OF THE beta SUBUNIT
J. Biol. Chem., July 13, 2007; 282(28): 20698 - 20704.
- **逢宇(2008) 博士研究生学位论文**

致谢

- 罗静初 教授
- 王珺 逢宇
- **Pku09s2**班的全体同学

Thank you!