

利用基本生物信息学工具对蓝藻 Synechococcus.PCC.7002中DNA binding protein HU的结构、功能分析和结构预测

报告人：王志敏

组 员：刘晓琴 邹 盼
陶建立 王志敏

Outline

▶ 课题背景：

蓝藻

Synechococcus sp. PCC 7002 的基因组

蓝藻中DNA同源重组

histone-like (HU) DNA 结合蛋白

▶ 研究课题介绍

研究内容

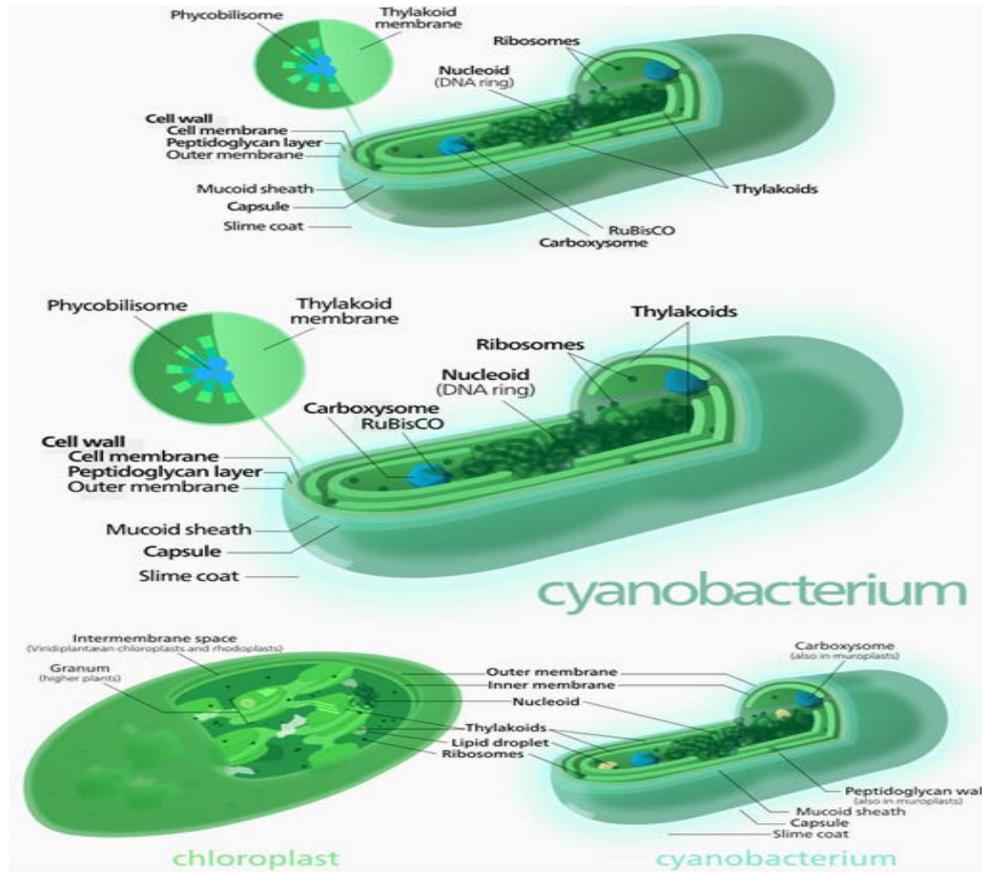
研究意义

利用生物信息学工具的初步研究和结果

研究思路和方法

▶ 总结和讨论

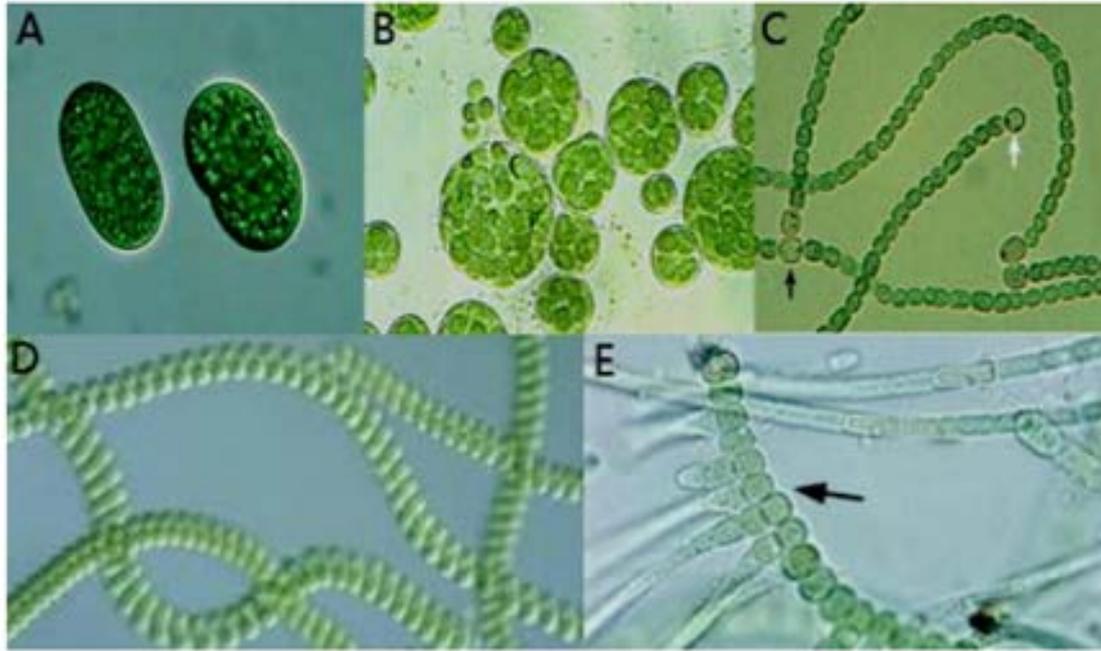
蓝藻(cyanobacteria)



- ▶ 来自35亿年前、前寒武纪的化石记录显示那时蓝藻就已经在地球上存在

——<http://en.wikipedia.org/wiki/File:Cyanobacterium-inline.svg>

蓝藻(cyanobacteria)

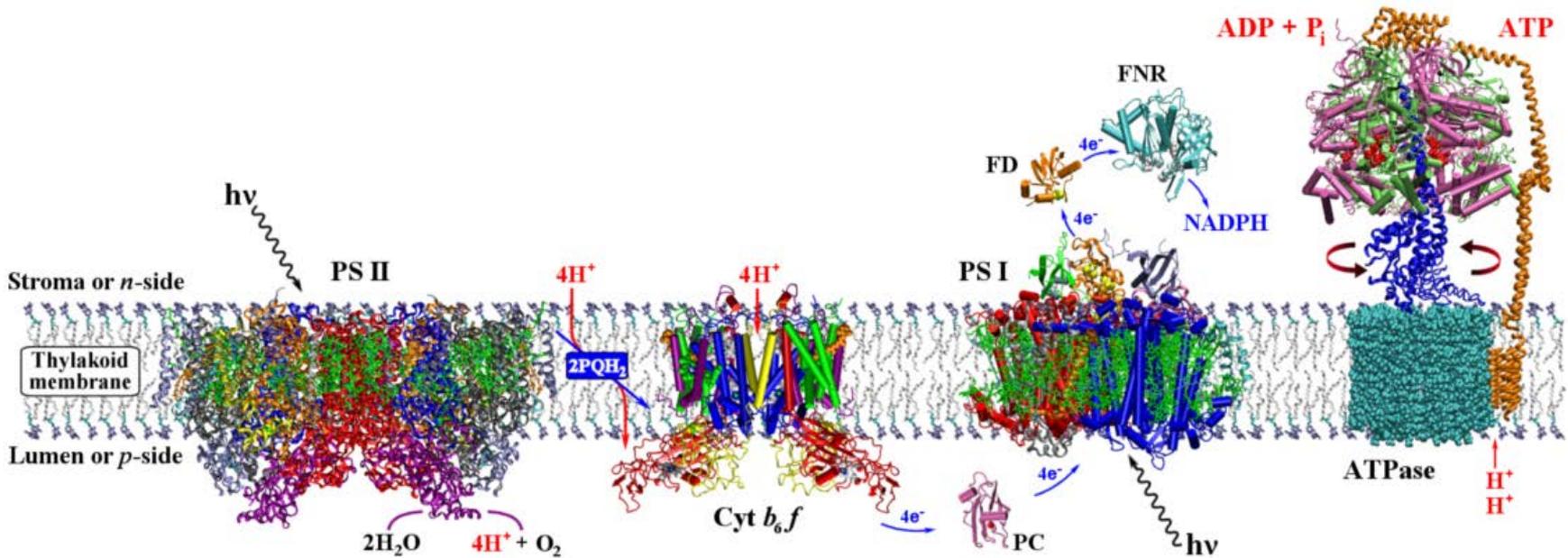


蓝细菌的五个目及其代表种类

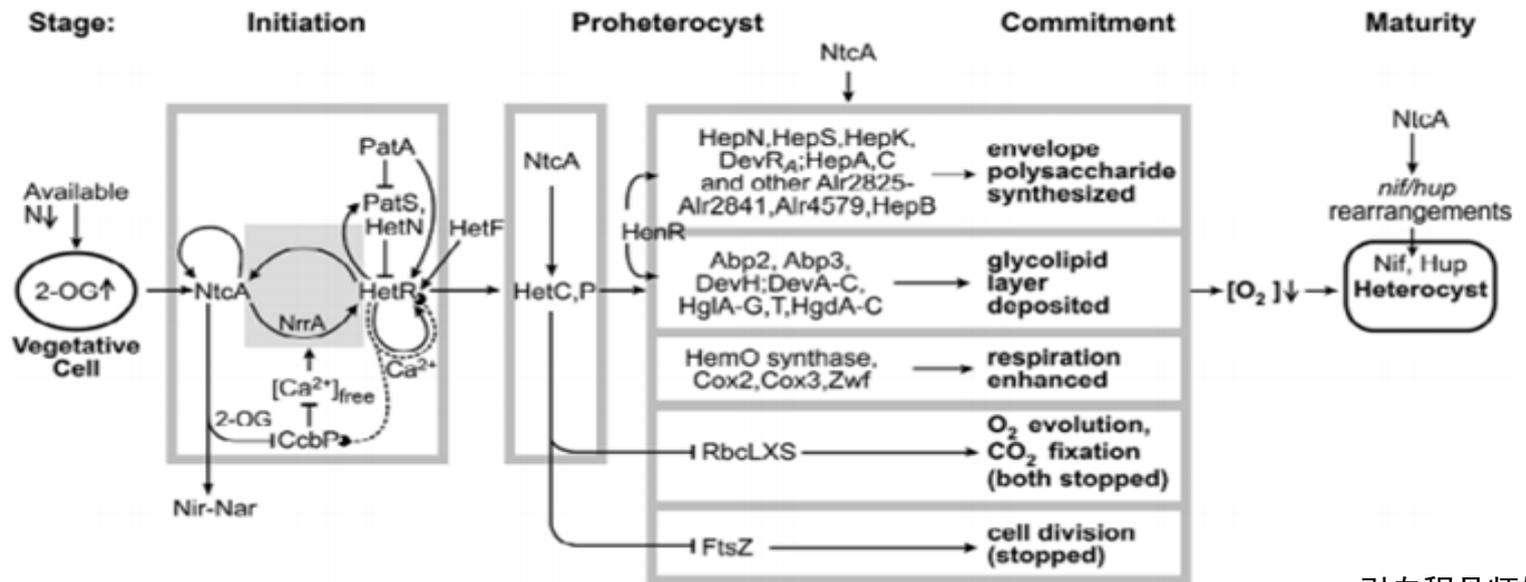
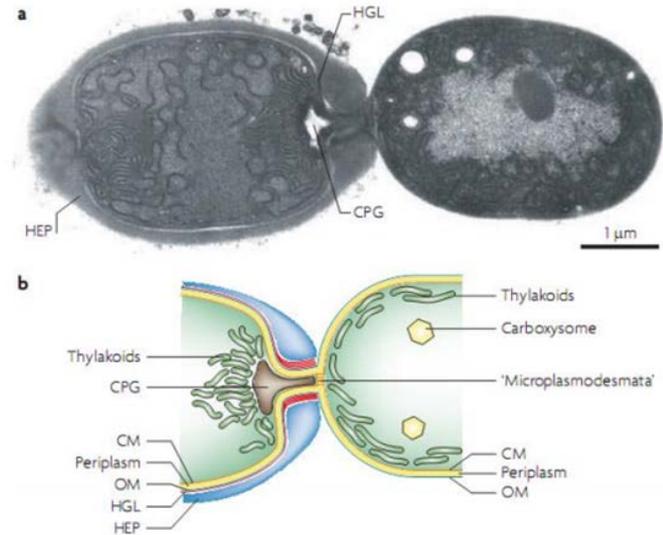
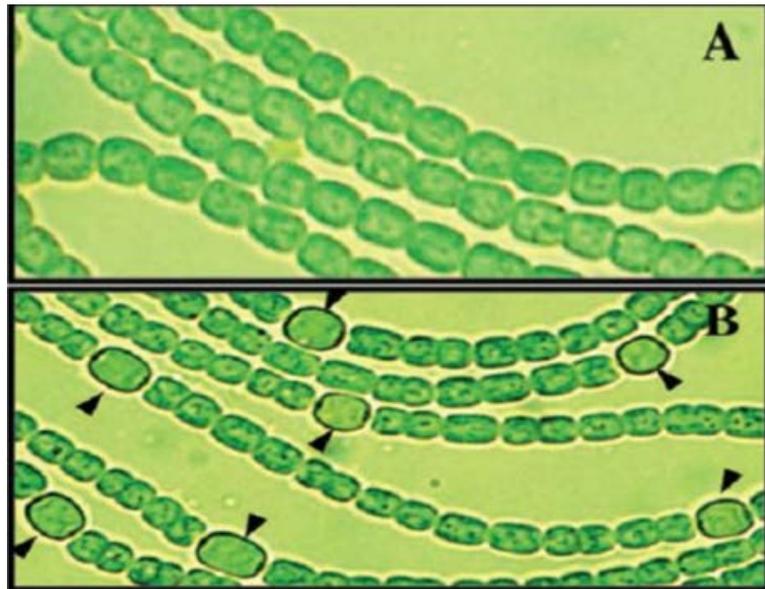
- ▶ 相对于真核生物，蓝藻遗传结构简单，生长周期短，分化速度快，形态和结构上相对简单。一直是研究光合作用和固氮作用的模型之一

蓝藻光合系统

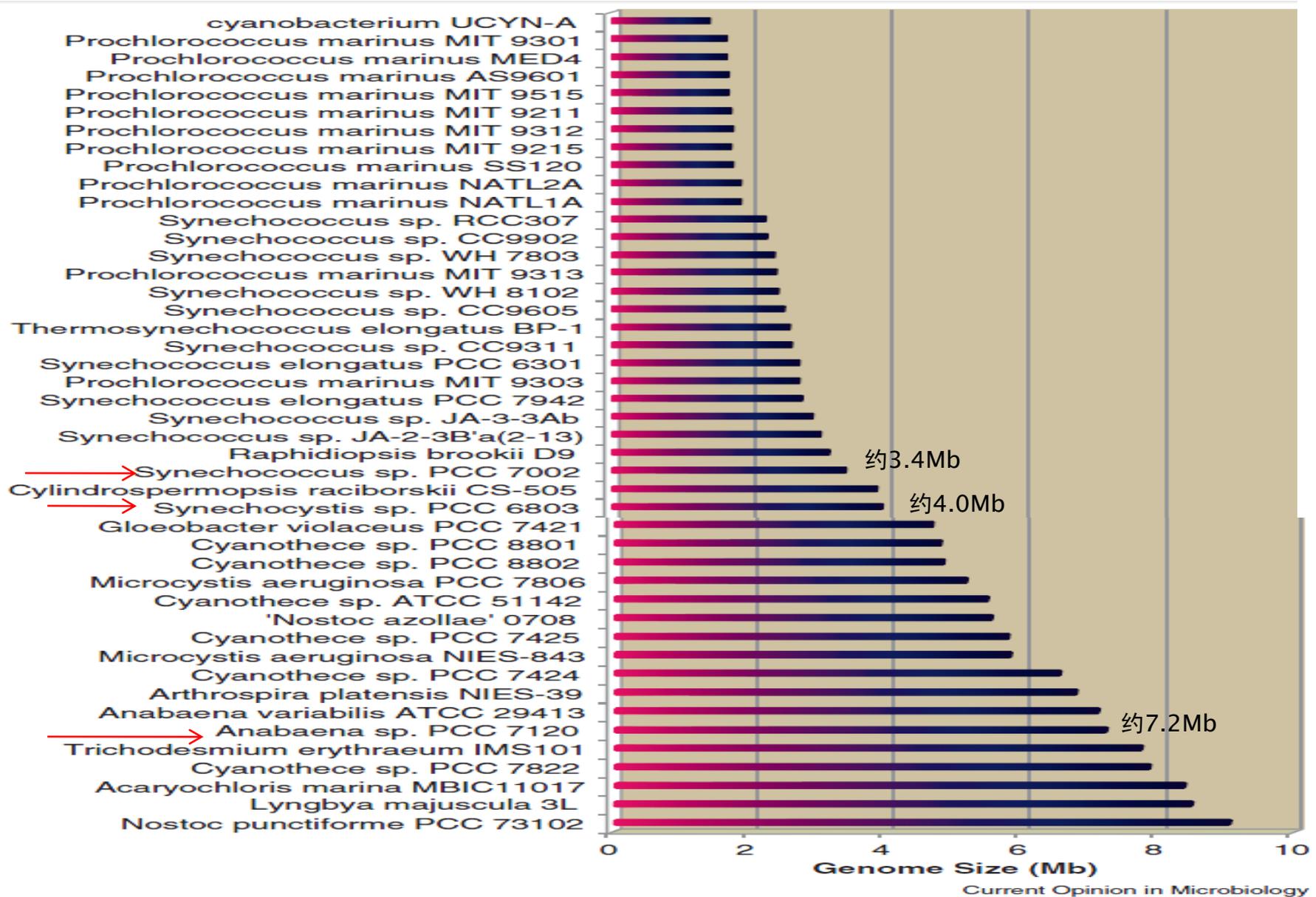
The photosynthetic membrane



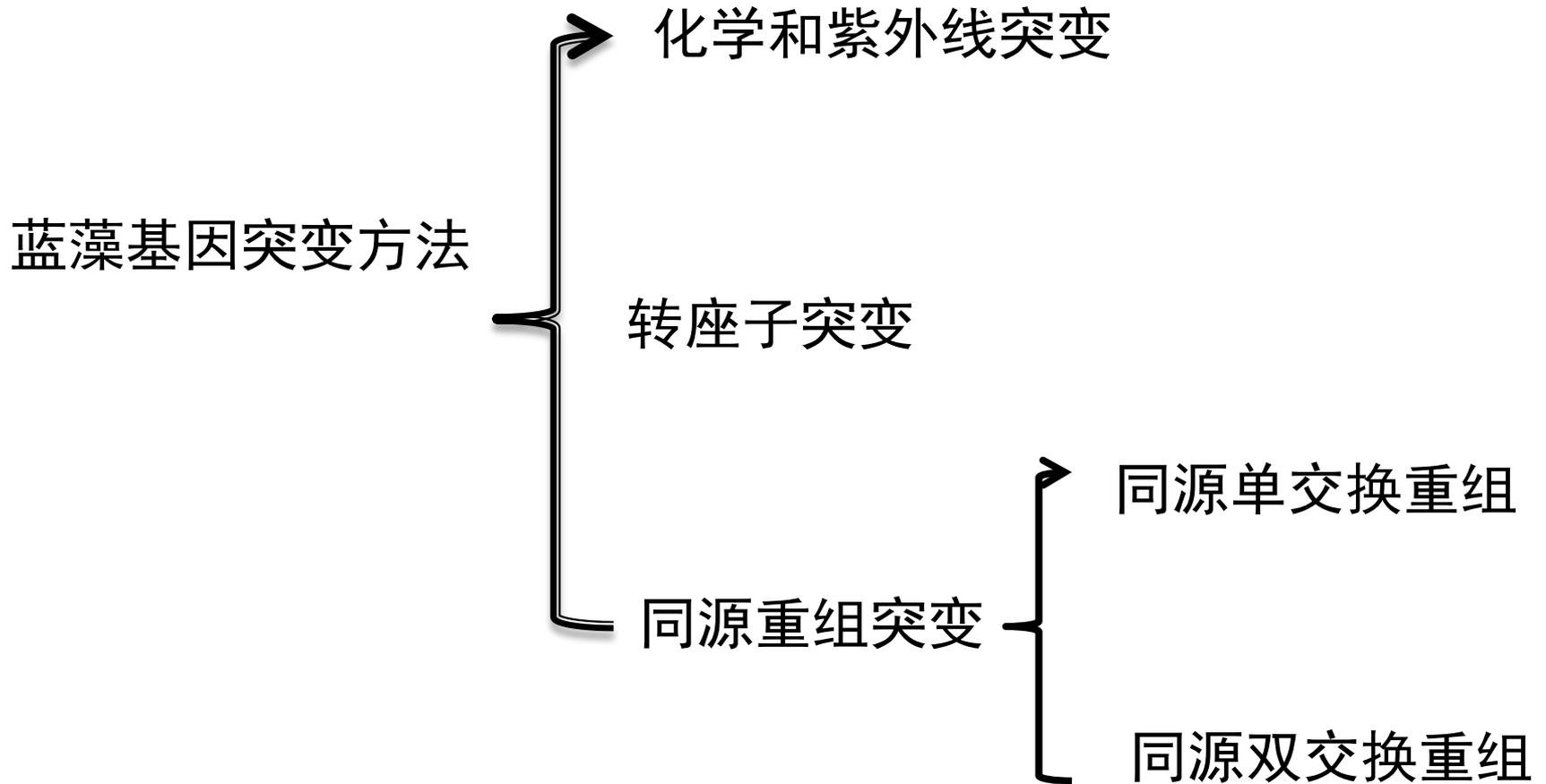
异形胞分化和格式形成



蓝藻(cyanobacteria)



蓝藻中DNA同源重组



同源重组突变

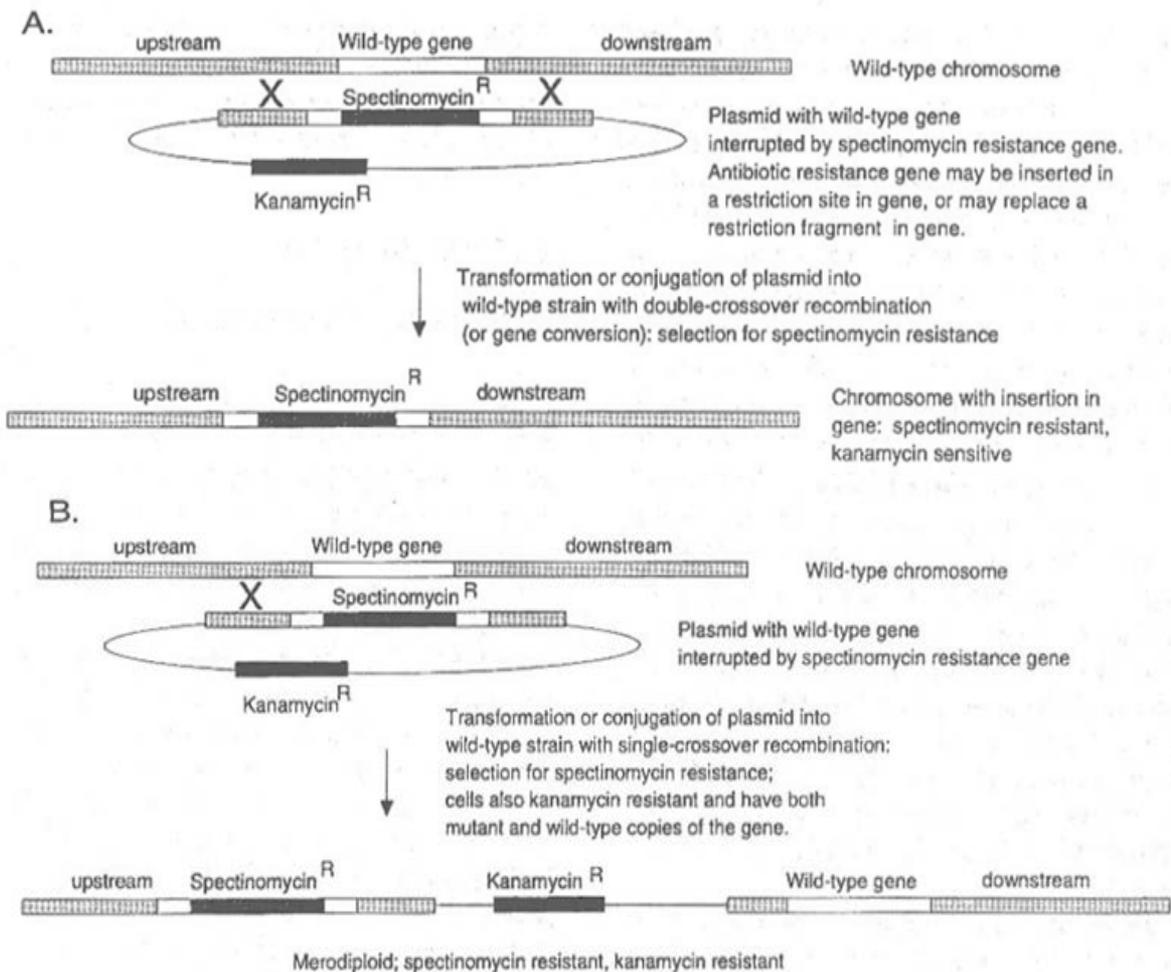
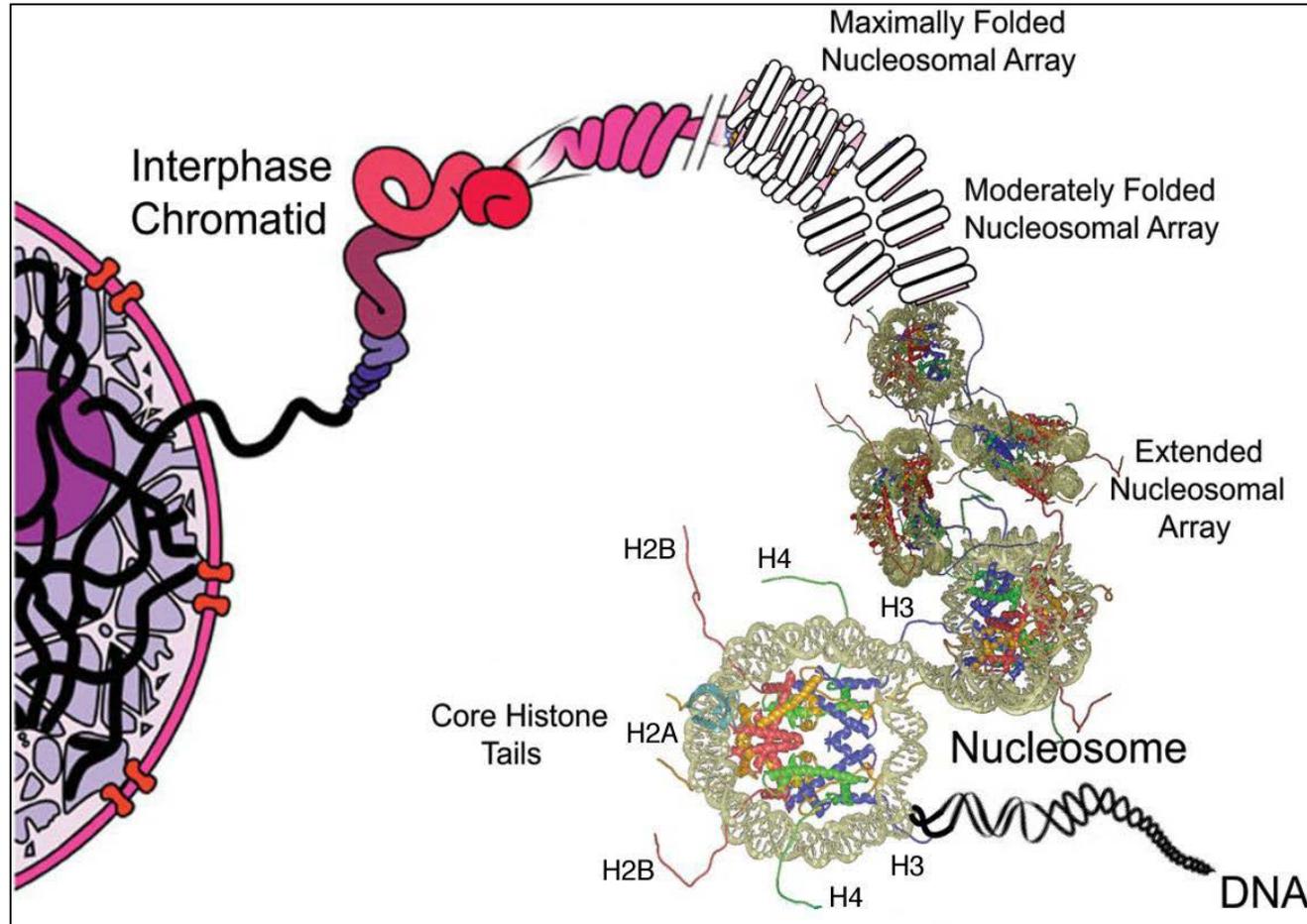


图 1-4 质粒 DNA 与蓝细菌染色体 DNA 重组示意图

Epigenetics的研究简介



Synechococcus sp. PCC 7002 Chromosome 1: SYNCC7002_A1965



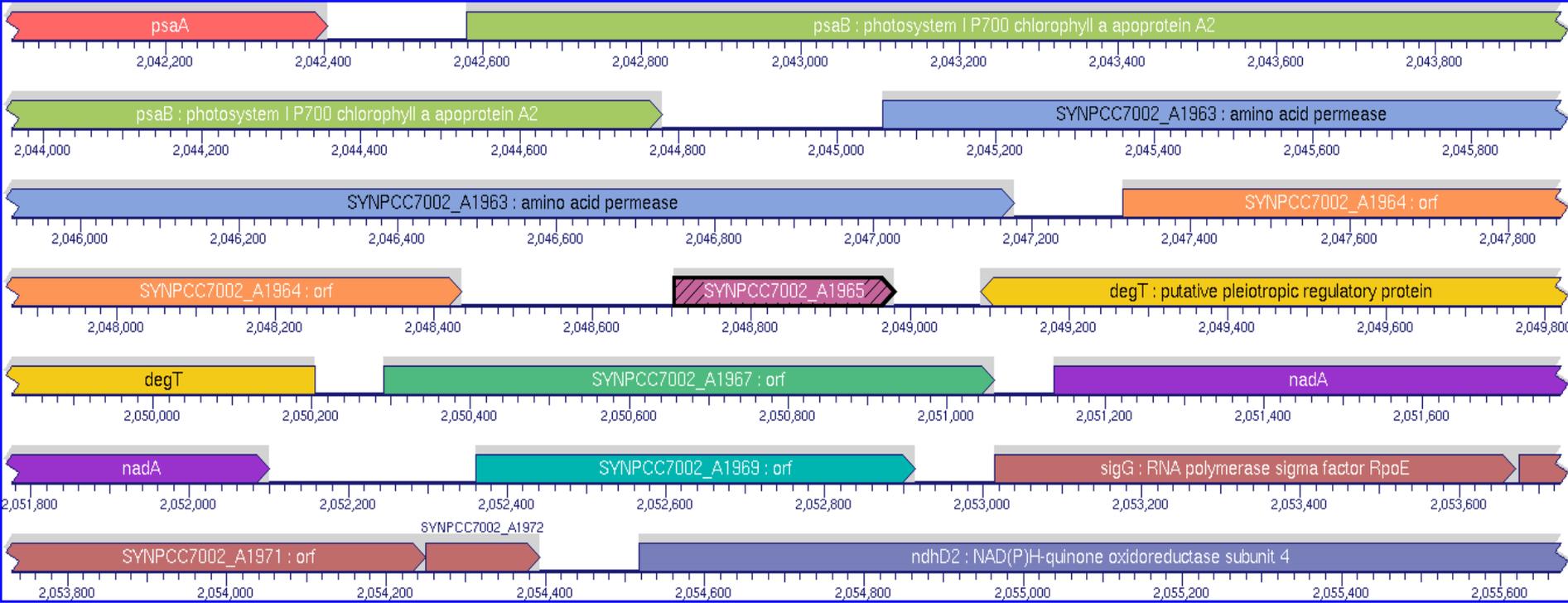
genome 4
operons 3
genes 2
sites 1

Zoom Out
Left Right
Zoom In

Start (bp): 2042006 End (bp): 2055676 Go Gene name: Show Tracks

Legend: Protein gene RNA gene Transcription Start Terminator

Gene color indicates operon membership.
Mouse over genes and operons for more information.
To center gene in display, click on tick mark under it.



DNA 结合蛋白histone-like (HU)研究简介

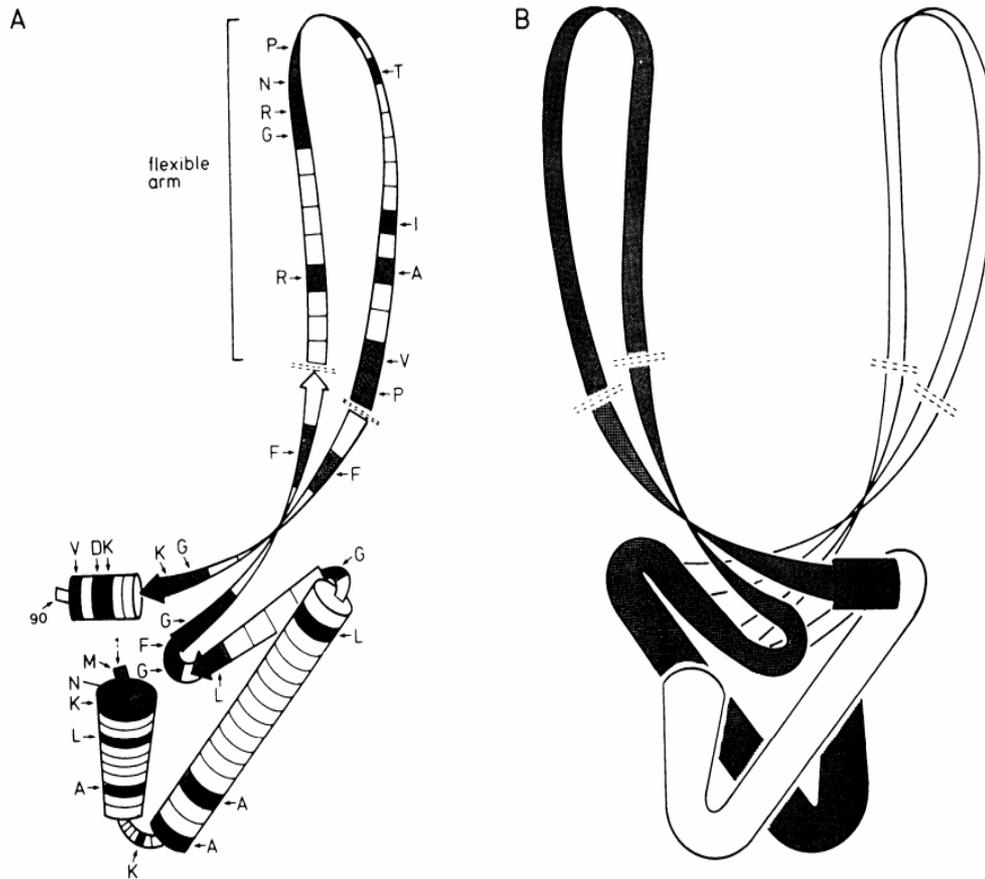
利用InterPro对HU进行分析：

所属蛋白家族： Histone-like bacterial DNA-binding protein family



- 该家族细菌合成的一类基本的、含量丰富的蛋白，约90AA，能结合DNA，在不同种类的真细菌、古细菌、蓝藻甚至一些真核藻类的叶绿体中有发现。
- 目前有HU protein, Integration Host Factor, Protein H, Protein H1, FirA五种，其中HU、IHF的研究较多。

DNA 结合蛋白histone-like (HU)研究简介

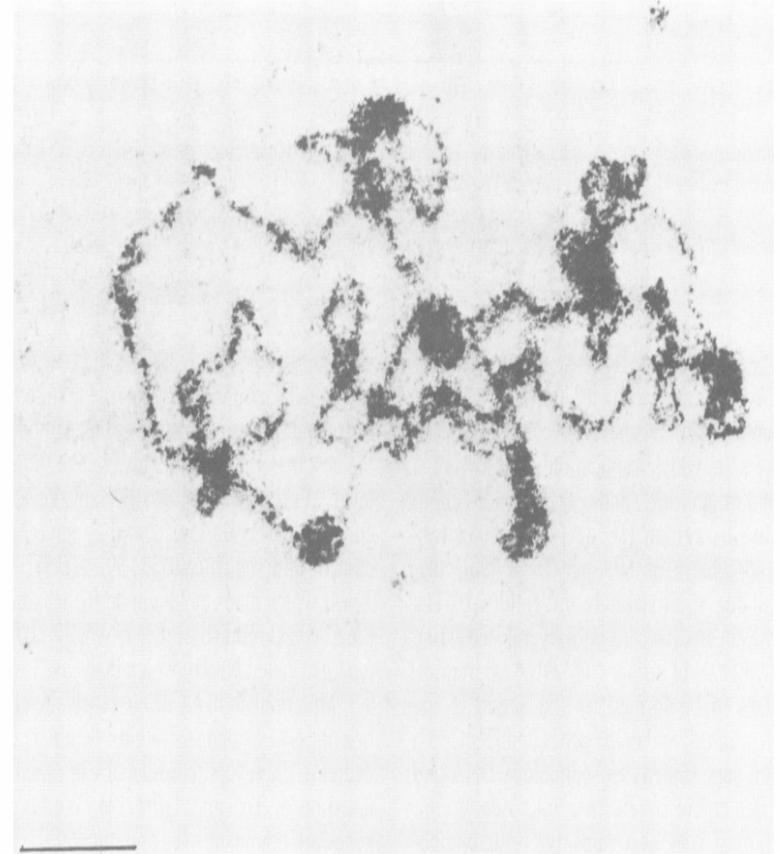
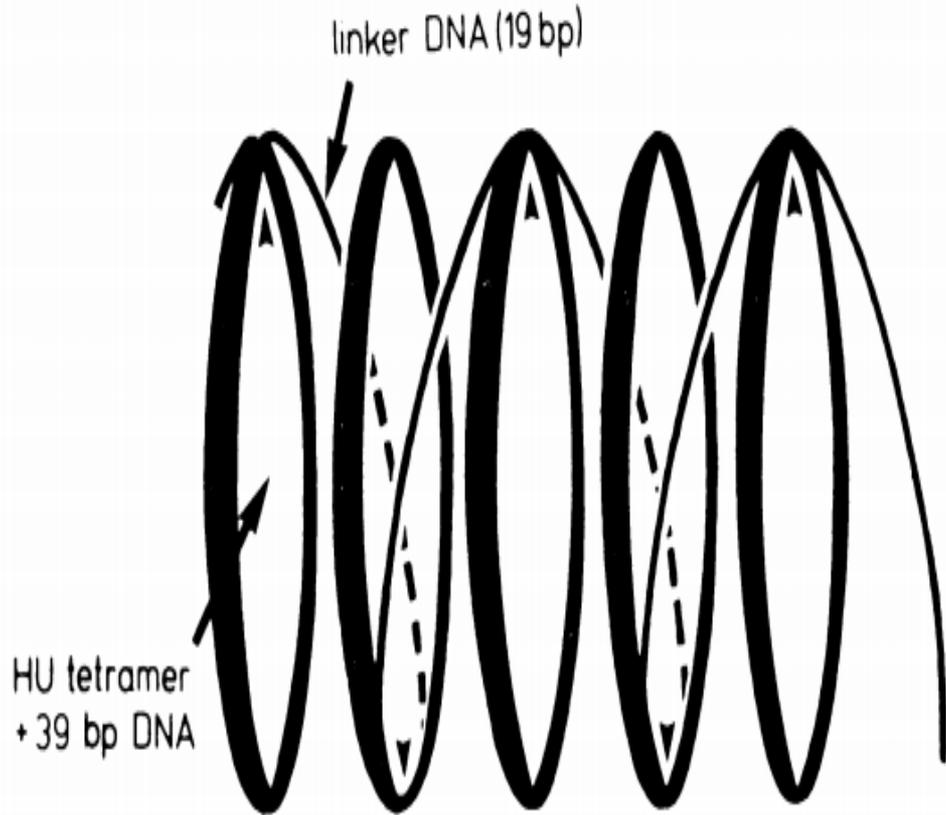


研究显示HU可能有如下的功能：
结合、压缩、弯曲DNA成类似核小体的结构；
参与转录调控：既有激活的证据也有抑制的证据；
参与一些类型的位点特异性重组；
体外实验显示，HU参与DNA复制起始，既有激活证据也有抑制证据

.....

Figure Structure of HU in Bacillus:(A)monomer,(B)dimer.

DNA 结合蛋白 histone-like (HU) 研究简介



——K Drlica and J Rouviere-Yaniv Microbiol. Rev. 1987, 51(3):301.

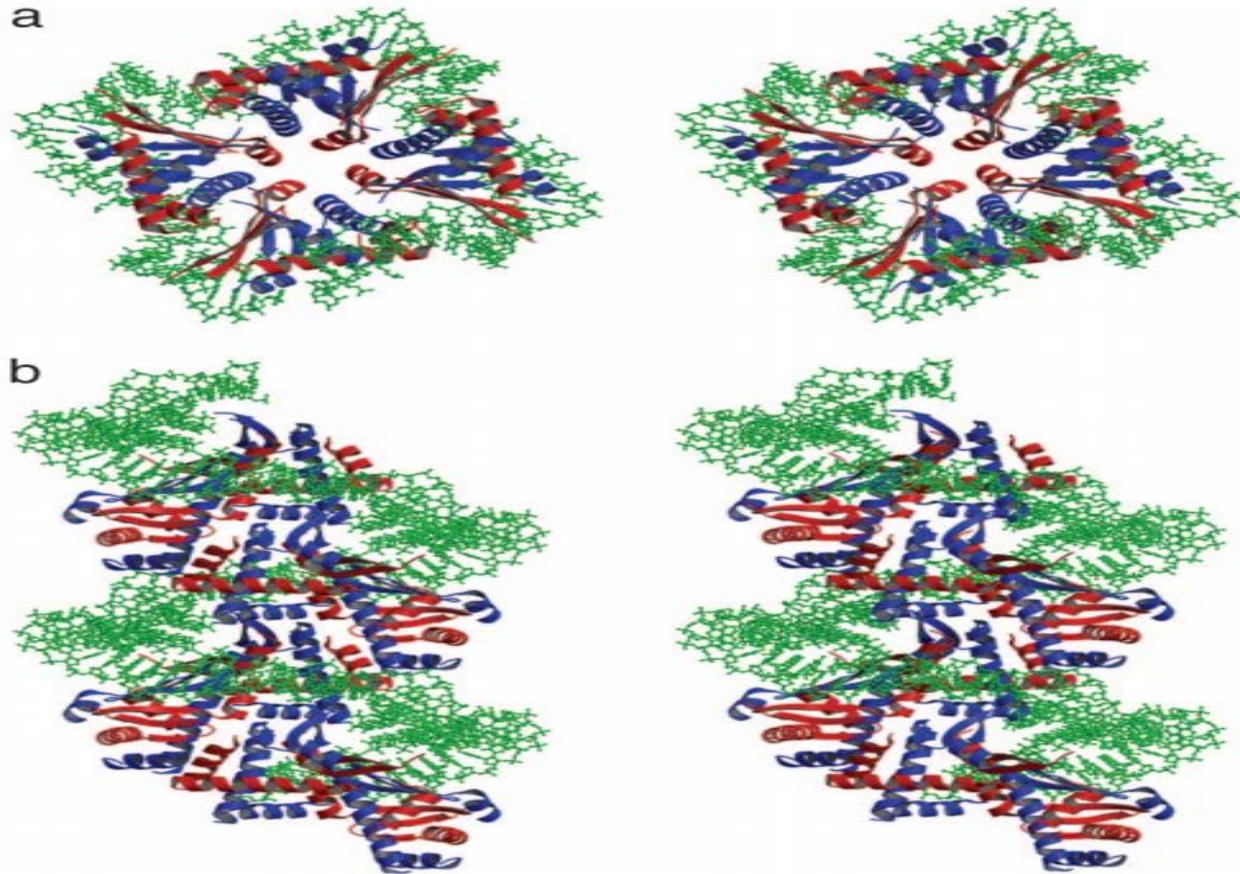


Fig. 5. Stereoviews of modeled $HU\alpha\beta$ octameric unit of the left-handed multimer with DNA fragment as a repeating unit of a spiral structure. The α -subunit is in red, and the β -subunit is in blue. (a) Top view. (b) Side view. The $HU\alpha\beta$ multimeric spiral is wrapped by DNA in a left-handed solenoidal structure.

研究内容

- ▶ histone-like DNA binding protein (HU) 在蓝藻*Synechococcus sp.PCC 7002*中的功能;
- ▶ 初步探究与HU互作的、可能参与多拷贝基因组稳定存在的蛋白，对稳定的分子机制有最基本的了解。
- ▶ 本次汇报着眼于利用基本生物信息学工具对*Synechococcus sp.PCC 7002* HU进行初步探究，以期对HU的演化情况、序列保守性，结构特征有一定认识，进而对*Synechococcus sp.PCC 7002* HU进行结构模拟和分析。

研究意义

- ▶ 对理解蓝藻中广泛存在的基因组多拷贝稳定存在现象的分子机制做初步探索；
- ▶ 对DNA甲基化、组蛋白修饰等在真核生物尤其是高等动植物中广泛存在的Epigenetic 修饰，其相类似的修饰在蓝藻的存在与否及可能发挥的功能进行初步探索；
- ▶ 对蓝藻人工进行基因组层面的操作给予分子机制上的支持。
- ▶ 本次汇报通过基本生物信息学工具对HU进行初步探究，对HU蛋白序列保守性、结构特征有初步认识，对*Synechococcus sp. PCC 7002* HU进行结构模拟，以期为研究其可能发生的修饰和功能奠定一定基础。

生物信息学工具的初步研究和结果

▶ **Uniprot** 数据库中对 *Synechococcus sp. PCC 7002* HU 相关信息的检索：



◆ **蛋白名称：** DNA-binding protein HU

◆ **蛋白大小：** 92AA

◆ **生物学功能：** 1. 参与DNA组装与压缩



2. DNA的结合

◆ **参考文献：**

"Complete sequence of *Synechococcus sp. PCC 7002*."

Li T., Zhao J., Zhao C., Liu Z., Zhao F., Marquardt J., Nomura C.T., Persson S., Detter J.C., Richardson P.M., Lanz C., Schuster S.C., Wang J., Li S., Huang X., Cai T., Yu Z., Luo J., Zhao J., Bryant D.A.

Submitted (FEB-2008) to the EMBL/GenBank/DDBJ databases

Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

Strain: ATCC 27264 / PCC 7002 / PR-6.

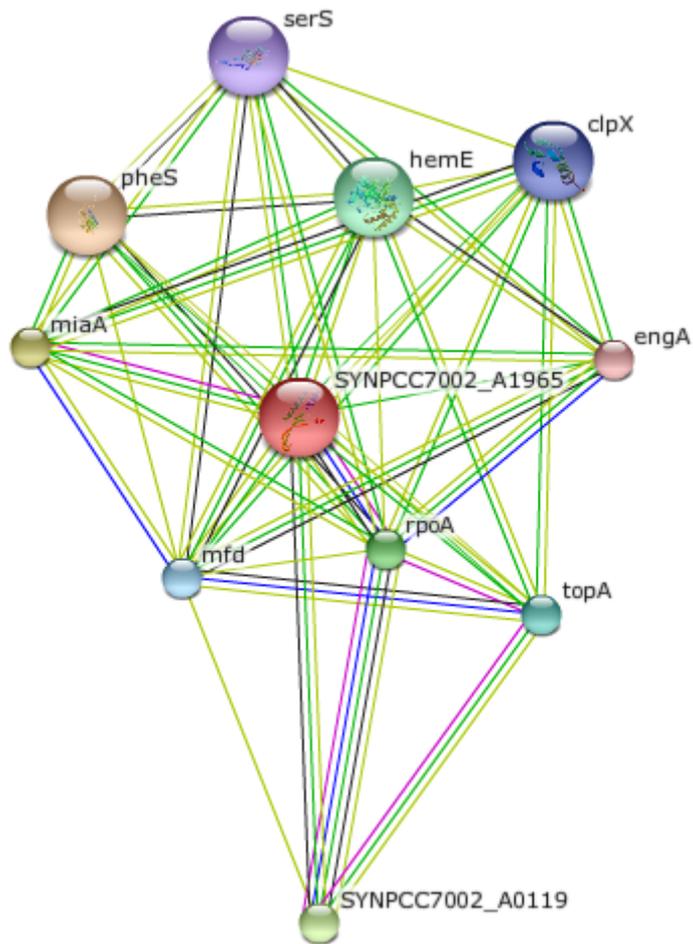
KEGG数据库中相关结果：

KEGG **Synechococcus sp. PCC7002: SYNPC7002_A1965** Help

Entry	SYNPC7002_A1965 CDS T00664
Definition	DNA-binding protein HU
Orthology	K03530 DNA-binding protein HU-beta
Organism	syp Synechococcus sp. PCC7002
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: Bac_DNA_binding DUF4496 HU-DNA_bdg Motif
Other DBs	NCBI-GI: 170078570 NCBI-GeneID: 6055703 CyanoBase: SYNPC7002_A1965 UniProt: B1XQQ7
Position	2048702..2048980 Genome map
AA seq	92 aa AA seq DB search MNRGELVDLVAEKAGISKQADSVISATVEAIMETVANGDKVTLVGFSGFEPHRKAREG RNPKTNEKMQIPATTVPAFSAGKQFKEMVAPK
NT seq	279 nt NT seq +upstream <input type="text" value="0"/> nt +downstream <input type="text" value="0"/> nt atgaataaaggcgaattagtcgatttagtcgctgagaaagctggtatcagcaaaaagcag gcagattctgtgattagtgccaccgtagaggcaatcatggaaaccggtgccaatggtgac aaagtaaccctcgtcggttttggtcctcttgaaccccgccaccgcaaagcgagagaaggt cgcaaccctaagaccaacgaaaaaatgcaaatccagcaaccacagtgcccgcatcttct gctggtaaacagtttaaggaaatggttgcccctaaatag

DBGET integrated database retrieval system

STRING数据库中相关结果：



可能的相互作用蛋白：

- 1.pheS: 苯丙酰胺tRNA合成酶亚基 α , 分值: 0.885;
- 2.miaA: tRNA Δ (2) 异苯丙焦磷酸转移酶, 分值: 0.875;
- 3.SYNPC7002_A0119: 单链DNA结合蛋白, 分值: 0.864;
- 4.rpoA: 以DNA为模板的RNA多聚酶亚基 α , 分值: 0.846;
- 5.hemE: 尿卟啉原脱羧酶, 分值: 0.845

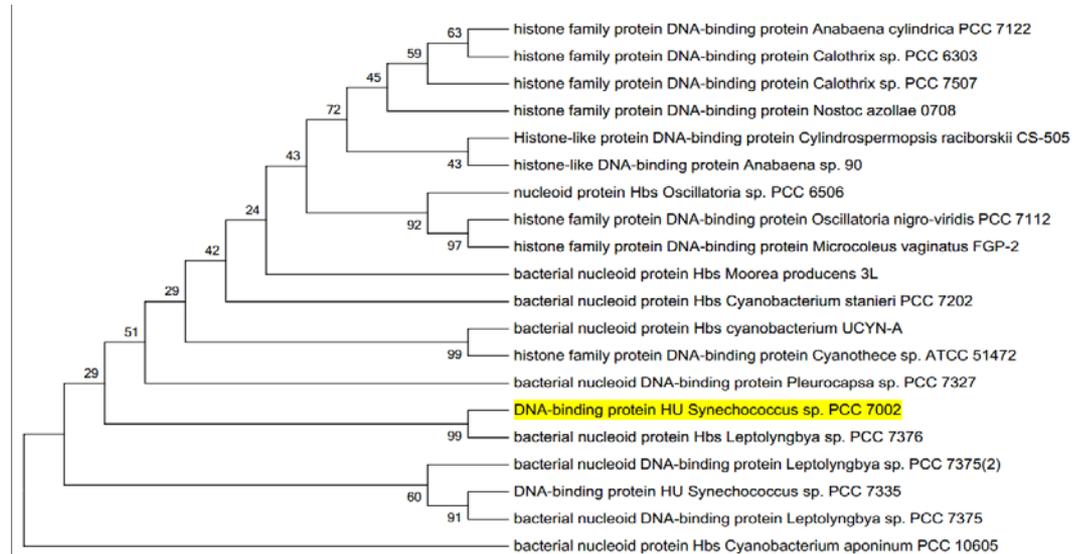
这些蛋白大部分都是通过其序列信息预测得到的，缺乏有力的实验数据！

BLAST结果：

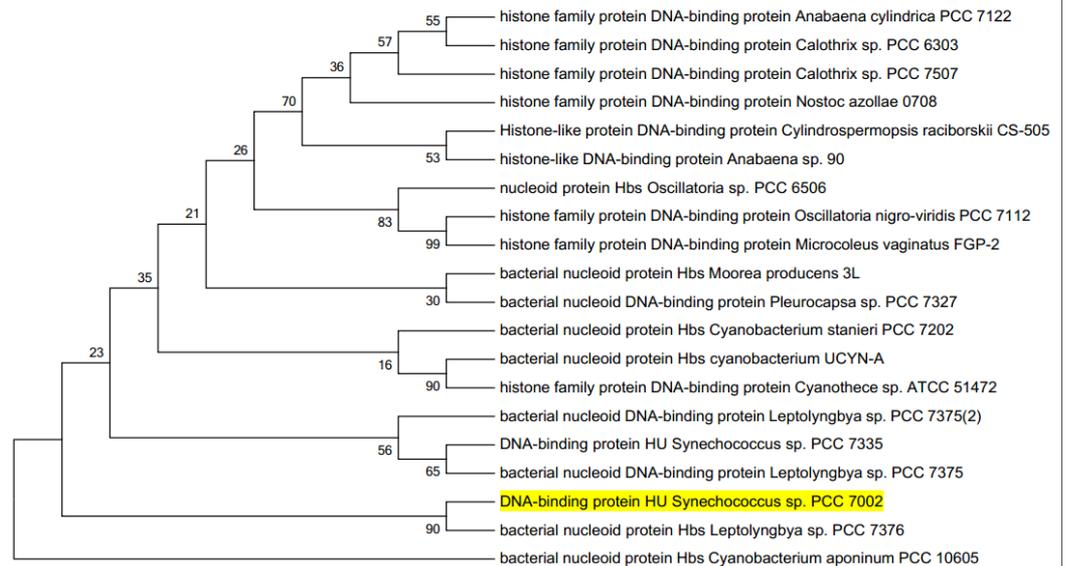
由于HU蛋白在不同种类原核生物中广泛存在，设置最多显示100条，所以在NCBI中用SYNPCC7002_A1965HU氨基酸序列进行BLAST，我们得到了100条序列，并对前20条序列进行序列比对和系统树的构建。

系统发育树构建：

Neighbor-Joining tree:



Maximum Likelihood tree:



进化距离计算：

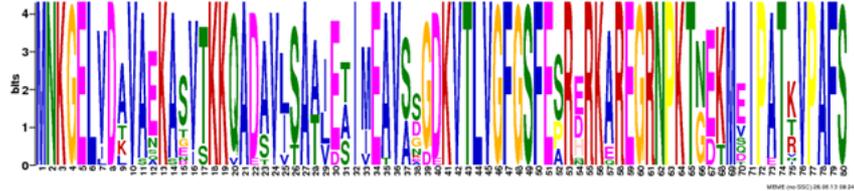
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. DNA-binding protein HU <i>Synechococcus</i> sp. PCC 7002		0.027	0.036	0.025	0.042	0.044	0.042	0.039	0.041	0.041	0.041	0.042	0.041	0.044	0.044	0.040	0.043	0.044	0.041	0.040
2. bacterial nucleoid protein Hbs <i>Leptolyngbya</i> sp. PCC 7376	0.077		0.039	0.026	0.040	0.042	0.040	0.038	0.039	0.039	0.039	0.040	0.041	0.042	0.042	0.040	0.041	0.043	0.039	0.041
3. bacterial nucleoid protein Hbs <i>Cyanobacterium stanieri</i> PCC 7202	0.154	0.165		0.026	0.039	0.041	0.043	0.033	0.032	0.031	0.032	0.033	0.037	0.036	0.036	0.033	0.037	0.040	0.031	0.043
4. bacterial nucleoid protein Hbs <i>Cyanobacterium aponinum</i> PCC 10605	0.923	0.912	0.912		0.025	0.025	0.025	0.028	0.028	0.028	0.028	0.026	0.028	0.026	0.026	0.026	0.025	0.026	0.028	0.025
5. bacterial nucleoid protein Hbs <i>cyanobacterium UCYN-A</i>	0.198	0.209	0.165	0.923		0.049	0.043	0.038	0.043	0.043	0.044	0.043	0.041	0.045	0.045	0.040	0.027	0.042	0.041	0.043
6. DNA-binding protein HU <i>Synechococcus</i> sp. PCC 7335	0.198	0.198	0.165	0.923	0.253		0.031	0.040	0.039	0.039	0.038	0.040	0.044	0.039	0.039	0.043	0.048	0.046	0.039	0.044
7. bacterial nucleoid DNA-binding protein <i>Leptolyngbya</i> sp. PCC 7375	0.209	0.220	0.198	0.923	0.242	0.110		0.035	0.040	0.039	0.041	0.040	0.039	0.037	0.037	0.042	0.043	0.041	0.036	0.042
8. histone family protein DNA-binding protein <i>Nostoc azollae</i> 0708	0.209	0.187	0.121	0.901	0.176	0.220	0.198		0.019	0.016	0.023	0.027	0.025	0.032	0.032	0.031	0.037	0.039	0.019	0.044
9. histone family protein DNA-binding protein <i>Anabaena cylindrica</i> PCC 7122	0.220	0.198	0.121	0.901	0.209	0.209	0.231	0.033		0.011	0.014	0.026	0.031	0.031	0.031	0.031	0.041	0.042	0.023	0.043
10. histone family protein DNA-binding protein <i>Calothrix</i> sp. PCC 7507	0.220	0.198	0.110	0.901	0.198	0.209	0.220	0.022	0.011		0.018	0.024	0.029	0.029	0.029	0.031	0.041	0.040	0.020	0.043
11. histone family protein DNA-binding protein <i>Calothrix</i> sp. PCC 6303	0.231	0.209	0.132	0.901	0.231	0.209	0.231	0.055	0.022	0.033		0.023	0.032	0.031	0.031	0.032	0.042	0.042	0.025	0.042
12. nucleoid protein Hbs <i>Oscillatoria</i> sp. PCC 6506	0.220	0.198	0.099	0.912	0.209	0.176	0.176	0.088	0.077	0.066	0.077		0.035	0.019	0.019	0.033	0.042	0.042	0.031	0.044
13. Histone-like protein DNA-binding protein <i>Cylindrospermopsis raciborskii</i> CS-50E	0.220	0.209	0.154	0.901	0.187	0.231	0.220	0.066	0.099	0.088	0.121	0.143		0.036	0.036	0.035	0.042	0.040	0.026	0.041
14. histone family protein DNA-binding protein <i>Oscillatoria nigro-viridis</i> PCC 7112	0.220	0.198	0.110	0.912	0.231	0.176	0.176	0.110	0.099	0.088	0.110	0.033	0.143		0.000	0.037	0.045	0.043	0.031	0.047
15. histone family protein DNA-binding protein <i>Microcoleus vaginatus</i> FGP-2	0.220	0.198	0.110	0.912	0.231	0.176	0.176	0.110	0.099	0.088	0.110	0.033	0.143	0.000		0.037	0.045	0.043	0.031	0.047
16. bacterial nucleoid protein Hbs <i>Moorea producens</i> 3L	0.209	0.198	0.121	0.912	0.187	0.209	0.209	0.110	0.110	0.110	0.110	0.099	0.154	0.121	0.121		0.040	0.035	0.031	0.042
17. histone family protein DNA-binding protein <i>Cyanothece</i> sp. ATCC 51472	0.209	0.209	0.143	0.923	0.077	0.253	0.242	0.154	0.187	0.176	0.209	0.187	0.176	0.209	0.209	0.176		0.045	0.041	0.045
18. bacterial nucleoid DNA-binding protein <i>Pleurocapsa</i> sp. PCC 7327	0.220	0.209	0.187	0.912	0.198	0.220	0.198	0.176	0.187	0.176	0.176	0.176	0.176	0.176	0.176	0.143	0.209		0.039	0.045
19. histone-like DNA-binding protein <i>Anabaena</i> sp. 90	0.220	0.198	0.110	0.901	0.198	0.209	0.209	0.044	0.055	0.044	0.077	0.099	0.066	0.099	0.099	0.121	0.176	0.165		0.042
20. bacterial nucleoid DNA-binding protein <i>Leptolyngbya</i> sp. PCC 7375(2)	0.220	0.220	0.209	0.923	0.231	0.176	0.209	0.253	0.253	0.253	0.253	0.220	0.220	0.242	0.242	0.209	0.242	0.242	0.242	

- ▶ 结论：*Synechococcus* sp.PCC 中的DNA结合蛋白HU与蓝藻中的颤藻目瘦鞘丝藻属蓝藻PCC 7376 Hbs氨基酸序列保守关系最近。

MEME软件结果：

Motif 1

- 4.5e-1059
- 20 sites



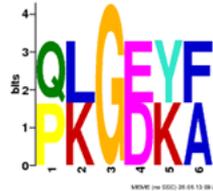
Motif 2

- 1.1e-090
- 20 sites



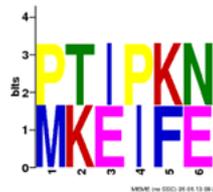
Motif 3

- 1.6e+004
- 2 sites



Motif 4

- 4.3e+005
- 2 sites



使用SMART对*sp.PCC 7002* HU的氨基酸序列进行结构域预测



Length	92 aa
Source database	UniProt
Identifiers	B1XQQ7_SYNP2, B1XQQ7



Confidently predicted domains, repeats, motifs and features:

Name	Start ▲	End	E-value
BHL	1	90	1.39e-43

Outlier homologues and homologues of known structure

Name	Sequence	Start ▲	End
Blast:BHL	DBH_ANASP 1-90	1	90
PDB:1P78 B	1p78	1	91
SCOP:d1exea_	d1exea_	1	89

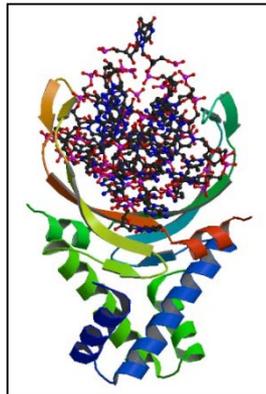
Features NOT shown in the diagram: ?

Name	Start ▲	End	E-value	Reason
Pfam:Bac_DNA_...	1	90	3.5e-30	overlap

InterPro数据库中在 *Synechococcus sp. PCC 7002* HU 注释信息中有关蛋白结构数据显示结构模拟的信息:



1P51



Molecule:	DNA-binding protein HU
Polymer:	2 Type:
Length:	94
Chains:	A, B, C, D
Organism	Anabaena sp.
Gene Names	hup hanA asr3935

在PDB数据库中输入DNA binding protein HU,搜索已知结构

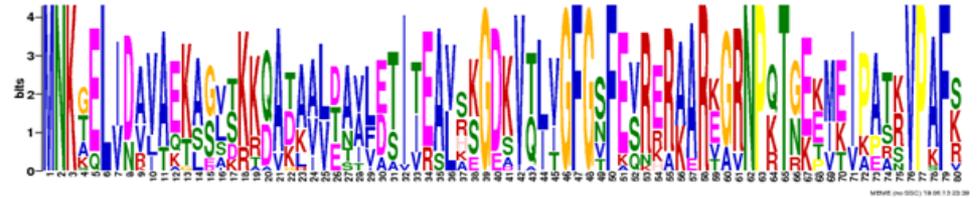
总共搜索到15个蛋白结构,我们将其中部分蛋白的信息总结于下:

1HUE	1HUU	1B8Z	1P51	1P71	1P78
Histone-like protein	Dna-binding protein hu from bacillus stearothermophilus	Hu from thermotoga maritima	Anabaena hu-dna cocrystal structure (ahu6)	Anabaena HU-DNA corcrystal structure (TR3)	Anabaena HU-DNA cocrystal structure (AHU2)
1RIY	2NP2	1MUL	2O97	3C4I	4DKY
HU mutant V42I from Thermotoga maritima	Hbb-DNA complex	Crystal structure of the E. coli HU alpha2 protein	Crystal Structure of E. coli HU heterodimer	Histone like protein-DNA binding) from Mycobacterium tuberculosis	Histone like protein-DNA binding) from Mycobacterium tuberculosis

使用MEME对结构已知的15条HU氨基酸序列进行motif的检索

Motif 1

- 2.5e-991
- 23 sites



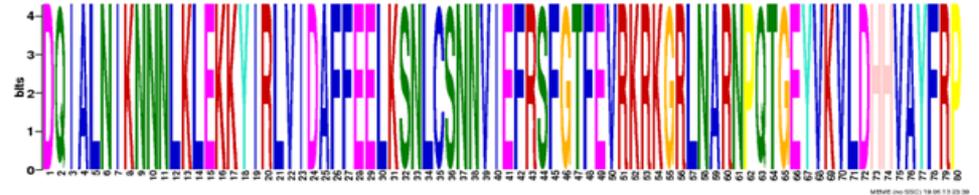
Motif 2

- 6.4e-055
- 35 sites



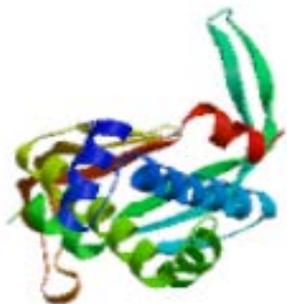
Motif 3

- 3.1e-037
- 2 sites



使用Swiss—Model对*Synechococcus sp.PCC 7002* HU进行结构模拟

Model Summary:



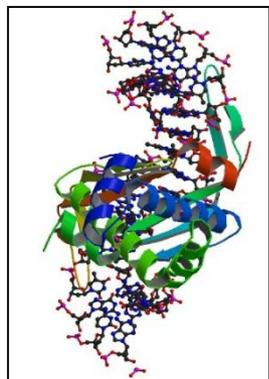
Model information:

Modelled residue range: 1 to 91
Based on template: 1p78B (2.25 Å)
Sequence Identity [%]: 76.923
Evaluate: 1.43718e-27

Quality information:

QMEAN Z-Score: 0.266

1P78



Molecule: DNA-binding protein HU
Polymer: 2 Type:
Length: 94
Chains: A, B
Organism: Anabaena sp.
Gene Names: hup hanA asr3935

使用Swiss-PDB Viewer对1P51 & 1P78蛋白结构模型进行分析

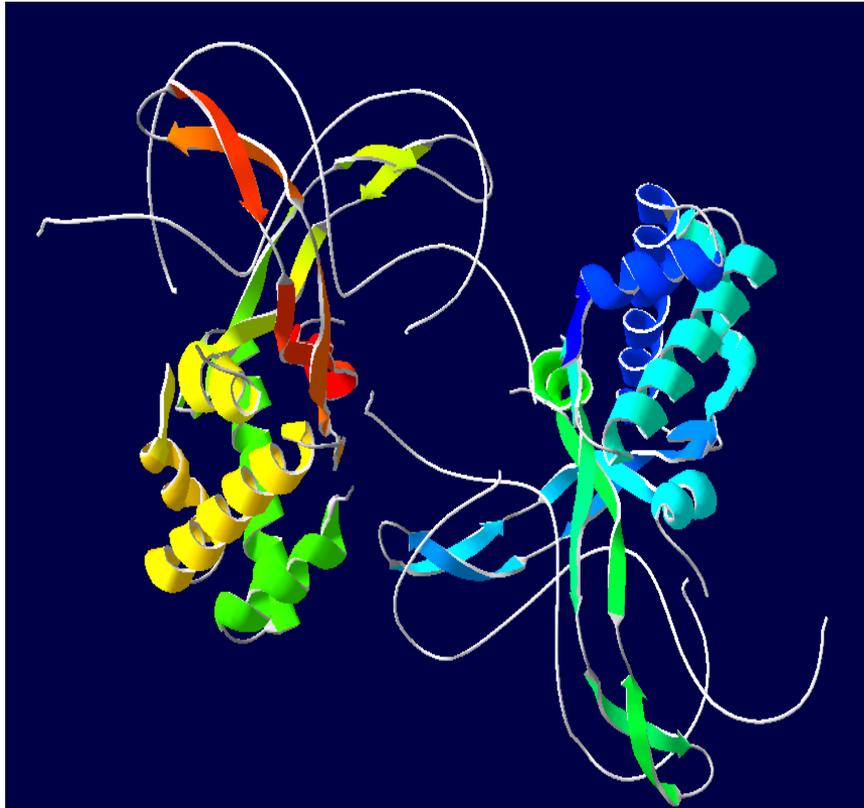


Fig1. **1P51**: Anabaena HU-DNA cocrystal structure(AHU6).

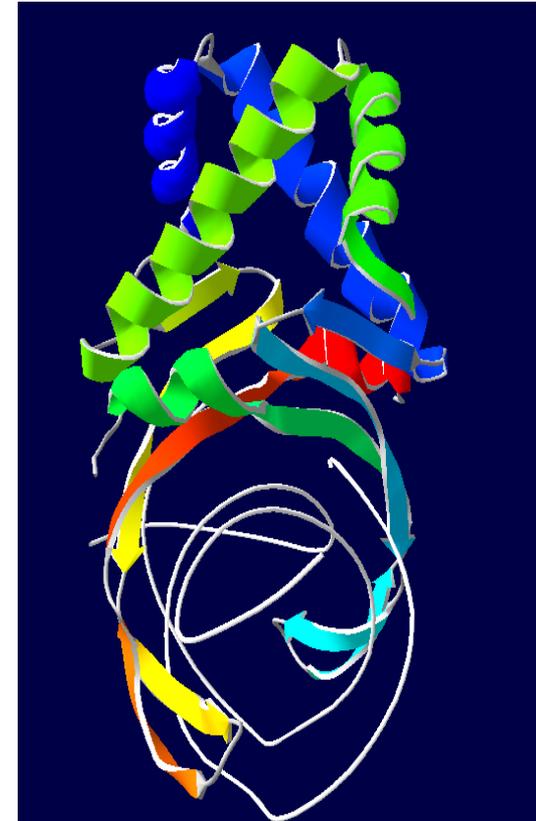


Fig2. **1P78**:Anabaena HU-DNA cocrystal structure (AHU2).

使用Swiss-PDB Viewer软件用不同模型对7002_HU蛋白结构进行分析

Fit:RMSD: 0.06

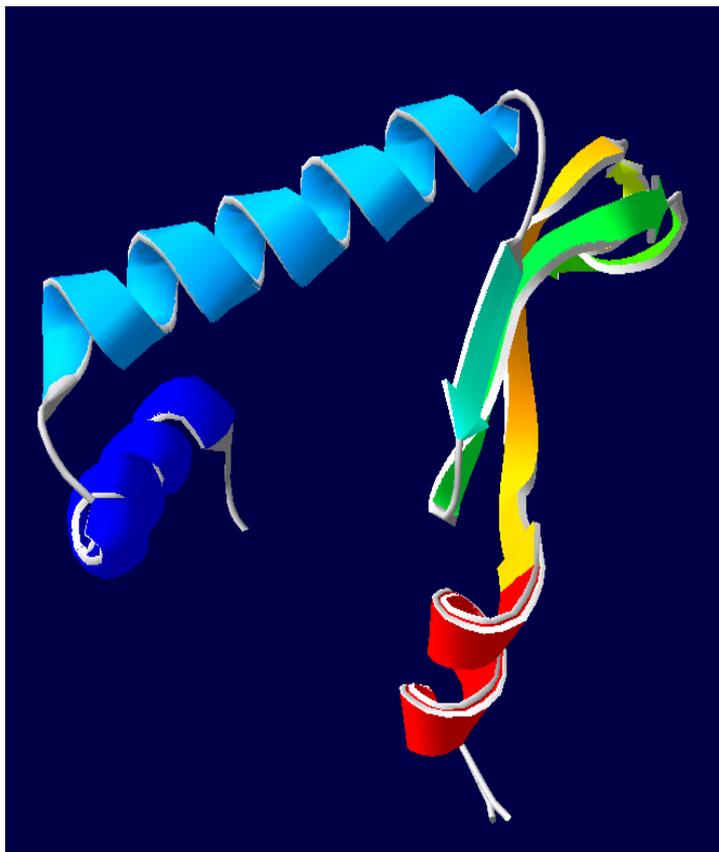


Fig1. 以1P51D与为模板构建7002_HU结构模型

0.07



Fig2.以1P78为模板构建的7002_HU结构 模型

总结和讨论

- ▶ DNA binding protein HU是原核生物中普遍存在的一类蛋白，在不同种类的蓝藻中也普遍存在，且氨基酸序列相对保守；
- ▶ 从结构上来看，DNA binding protein HU蛋白结构也相对保守的，大致结构为monemer由三个 α 螺旋和三个 β 片层组成，dimer呈现龙虾状结构 (a lobster-shaped structure)。这种结构有利于结合DNA。
- ▶ SMART和MEME进行motif检索显示，DNA binding protein HU 类蛋白仅有一个大的BHL(bacteria (prokary) histone like domain)结构域，用于结合DNA.目前参阅的文献显示HU没有特异性的DNA结合位点。
- ▶ 进一步需要进行HU上有关的修饰作用的文献检索和阅读，以及生物信息学分析。

研究思路和方法

- ▶ 基本的生物信息学的研究：序列比对，结构预测，HU可能处于的 pathway 分析，与HU互作蛋白的预测等；
- ▶ 体外实验：HU的表达、纯化；RNA表达量分析；RT-PCR；pull-down；western blot；DNA binding 实验等等。
- ▶ 体内功能实验：构建HU缺失突变体、点突变体、HU过表达突变体，与荧光蛋白融合，观察突变体蓝藻细胞的表型；对缺失突变体和点突变进行rescue等实验；荧光免疫杂交等实验对HU进行定位等等。

可行性分析

- ▶ 实验室长期从事蓝藻分子生物学研究，形成一套研究蛋白功能的成熟体系和方法；
- ▶ 实验室完成了对蓝藻*Synechococcus sp. PCC 7002*全基因组的注释，为研究HU功能提供了可靠的支撑；
- ▶ 实验室有用于蓝藻分子生物研究的先进仪器设备，提供了完备的硬件支持；

参考文献：

- ▶ K Drlica and J Rouviere–Yaniv.(1987) Histone-like proteins of bacteria.Microbiol.Rev 51(3):301.
- ▶ David E. Pettjohn.(1988) Histone-like Proteins and Bacterial Chromosome Structure.The Journal Of Biological Chemistry 263(26).
- ▶ Wolfgang R Hess.(2011) Cyanobacterial genomics for ecology and biotechnology. Current Opinion in Microbiology,14:608 – 614.
- ▶ Kerren K.Swinger et al.(2003) Flexible DNA bending in HU±DNA cocystal structures. The EMBO JournalVol(22): 3749–3760.
- ▶ Fusheng Guo, Sankar Adhya.(2007) Spiral structure of Escherichia coli HU provides foundation for DNA supercoiling. PNAS (104):4309 – 4314

致谢：

- ▶ 感谢罗老师一个学期以来认真、热情的讲解和指导；希望以后还能得到您的指导和帮助；
- ▶ 感谢小组成员积极参与讨论、相互启发学习和帮助；
- ▶ 感谢助教郭羽同学的帮助和指导；
- ▶ 感谢班级其他同学的相互学习、帮助；
- ▶ 希望大家以后还能相互帮助，祝大家科研上越走越顺，身体健康，生活快乐！

Thank you!

