

生物信息学初探

农科院生物信息一班 caas09s1

报告人：荆振宇

课题研究思路

来源于文献



第一篇 已知的理论基础

- 血红蛋白是红细胞的主要成分,它由珠蛋白四聚体结合四个含铁的血素,在成体细胞中,珠蛋白四聚体有两条相同的 α 链和 β 链构成。

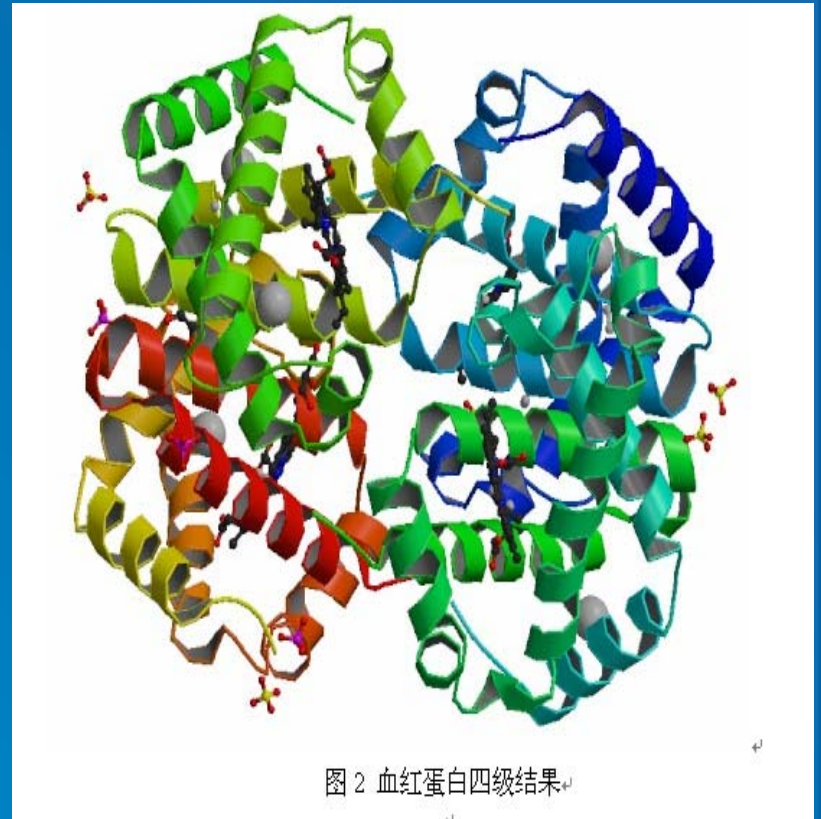
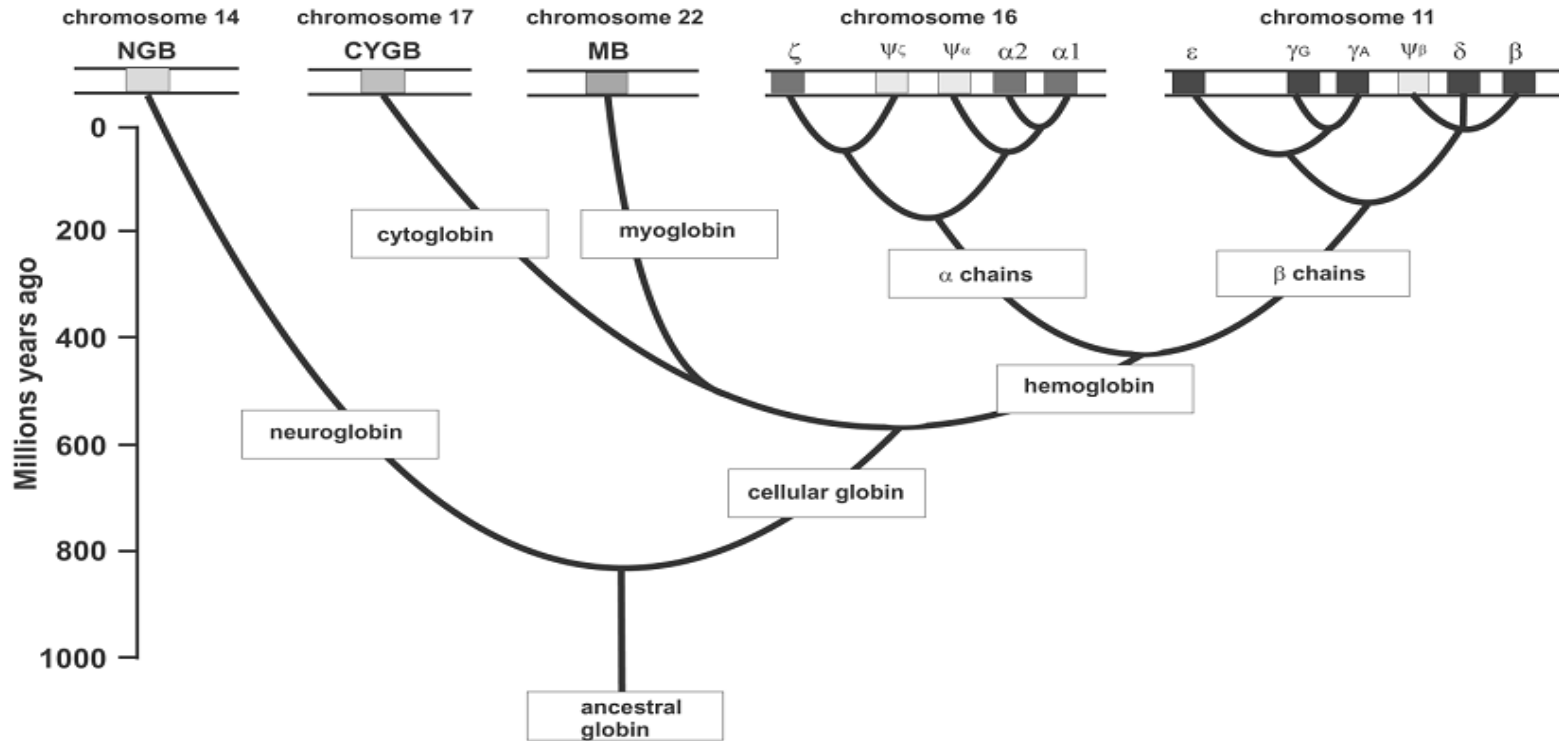


图 2 血红蛋白四级结构

脊椎动物珠蛋白演化模式如下：



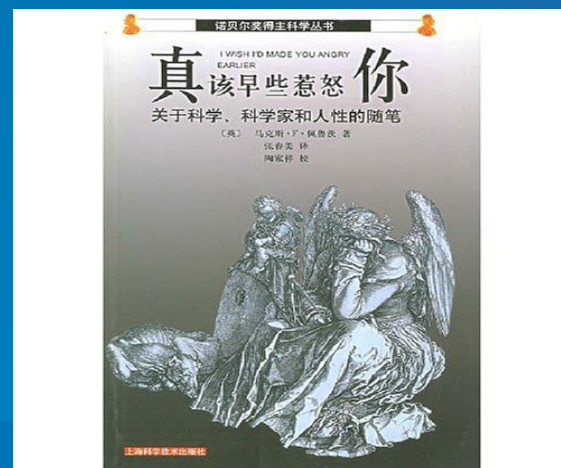
从上图可以看出，脊椎动物所有珠蛋白基因是由共同的祖先基因进化而来的， α 链珠蛋白基因和 β 链珠蛋白基因的分开发生的很晚，在300~400百万年前。 α 链珠蛋白基因包括 ζ 、 $\alpha 2$ 、 $\alpha 1$ 和和两条假基因，而 β 链珠蛋白包括 β 、 δ 、 ϵ 和两个 γ 何一条假基因。(Burmester T, 2002)

血红蛋白结构

- 血红蛋白的空间结构测定是由佩鲁茨及其同事于1963年完成的，在他的《真该早些惹怒你》一书中，创造性地提出了“同晶置换法”，即在所要分析的蛋白质晶体中引入适当的重金属原子，得到同晶置换晶体，通过对同晶置换晶体的x射线衍射结构分析，成功揭示了血红蛋白的三维结构。



佩鲁茨 (Perutz, Max Ferdinand)英国化学家，生物学家。



镰刀型血红蛋白

- 1956年,英格拉姆(Ingram)等人用胰蛋白酶把正常的血红蛋白(HbA)和镰形细胞的血红蛋白(HbS)在相同条件下切成肽段,通过对比二者的滤纸电泳双向层析谱,发现有一个肽段的位置不同。
- 这是 β 链N末端的一段肽链。也就是说,HbS和HbA的 α 链是完全相同的,所不同的只是"链上从N末端开始的第6位的氨基酸残基,在正常的HbA分子中是谷氨酸,在病态的HbS分子中却被缬氨酸所代替。

第二篇 生物信息学的预测与论证

第一部分 血红蛋白的起源

1

系统发育树的构建

2

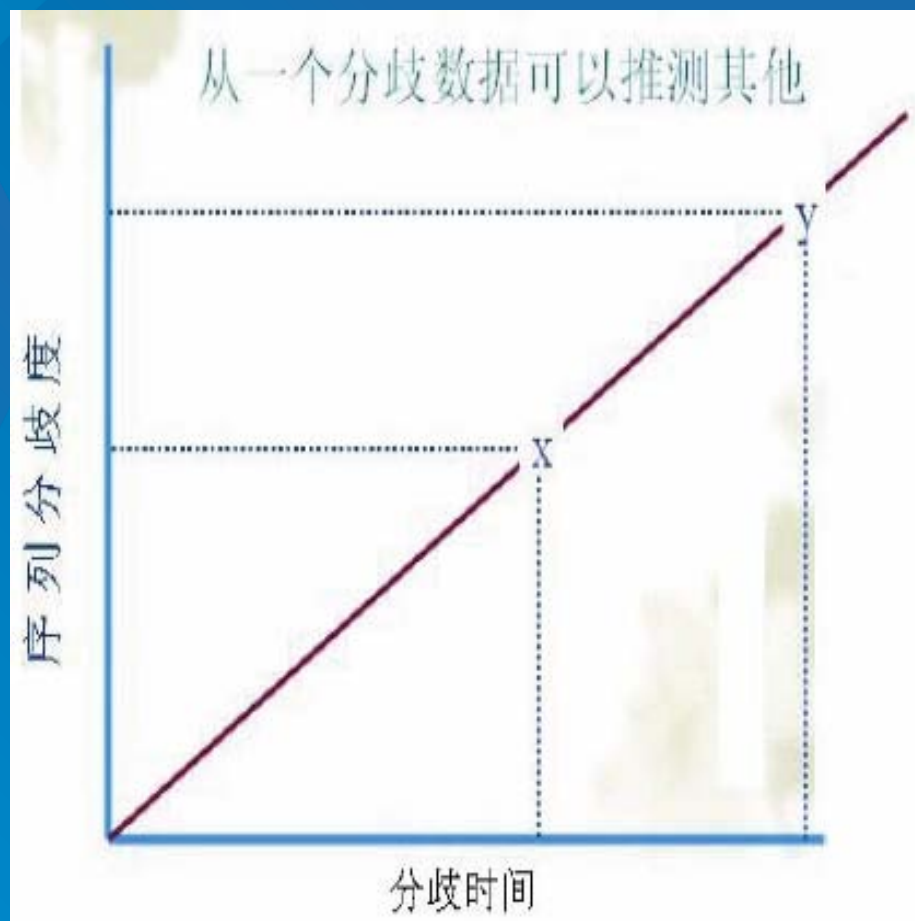
人血红蛋白在进化过程中的位置

3

物种起源与血红蛋白的分化

系统发育树构建的理论基础

- 系统发生分析一般是建立在分子钟基础上的。
- 分子钟:分子序列进化是按照一恒定速率进行的，所以积累突变的数量和进化时间成一定比例，基于这个假说，发生树上的树枝长度可以用来估算基因分离的时间。



进化树的构建

- 以人新发现的Neuroglobin为例利用mega软件对Neuroglobin是否属人类血红蛋白家族经分析。
- 判定利用生物信息学预测生物进化关系是否与通过传统的进化理论和实验室所得数据一致。

步骤

1

1、ngb_human
蛋白序列在
swiss-prot
中搜索



2

2、在NCBI
中Blast搜索
ngb_human
同源序列



3

3、用mega
软件进行序列
比对并构建进
化树



在ExPASy搜索NGB_HUMAN

UniProtKB > UniProtKB Downloads · Contact · Documentation/Help

Search in Protein Knowledgebase (UniProtKB) **Query** [Fields »](#)

★ **Reviewed, UniProtKB/Swiss-Prot Q9NPG2 (NGB_HUMAN)**

Last modified May 5, 2009. Version 66. [History...](#)

[Clusters with 100%, 90%, 50% identity](#) | [Documents \(4\)](#) | [Third-party data](#) | [Customize display](#)

[TEXT](#) [XML](#) [RDF/XML](#) [GFF](#) [FASTA](#)

[Names and origin](#) · [Protein attributes](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Sequence annotation \(Features\)](#) · [Sequences](#) · [References](#) · [Web resources](#) · [Cross-references](#) · [Entry information](#) · [Relevant documents](#)

保存蛋白质序列为.fasta格式

NCBI中Blast结果

① Your search is limited to records matching entrez query: txid9606 [ORGN].

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

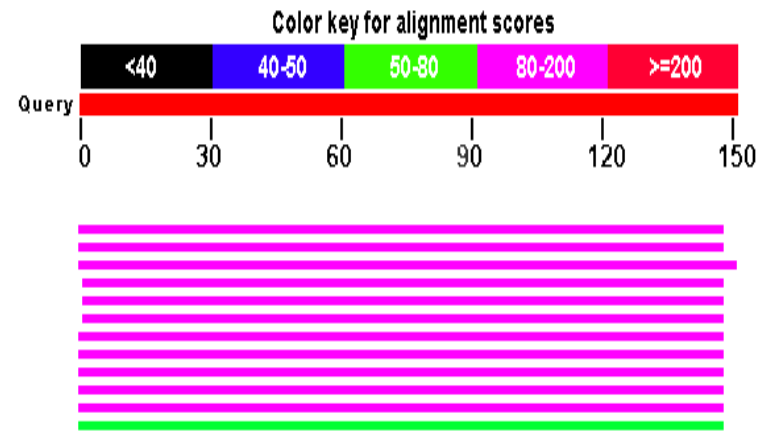
PSI blast Iteration 3

Neuroglobin

Query ID	lcl 71789	Database Name	swissprot
Description	Neuroglobin	Description	Non-redundant SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.2.20+ Citation
Query Length	151		


Distribution of 12 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments



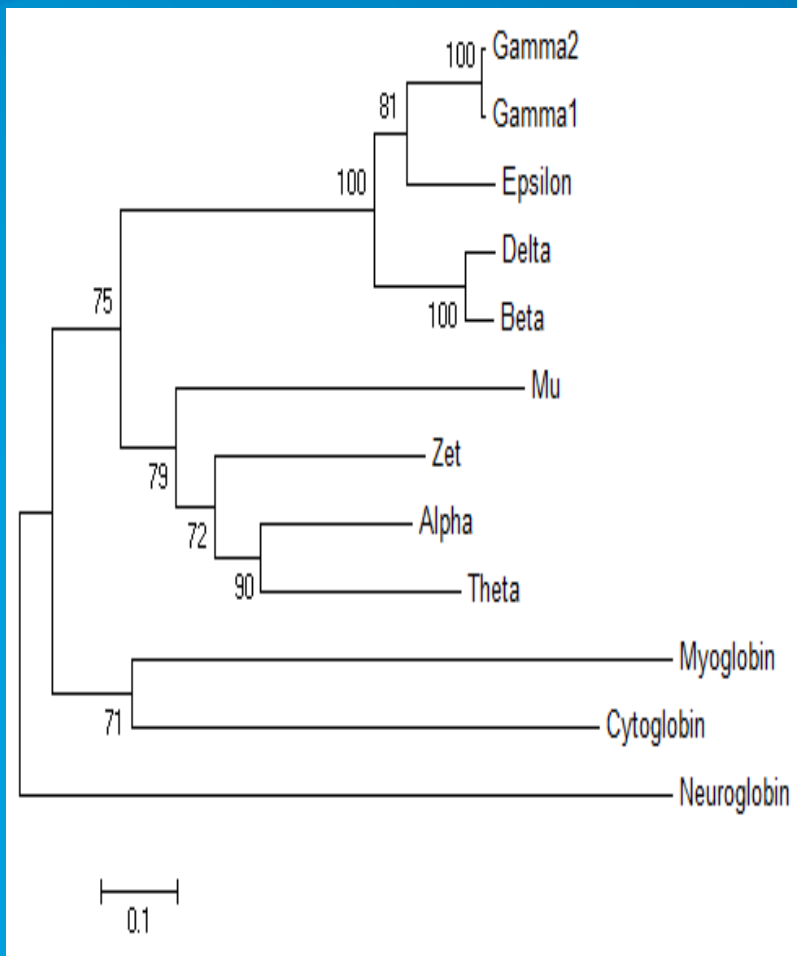
- 利用Blastp在swissprot proteinsequence的人类蛋白数据中搜索到12条保存为.fasta格式

view sequence data

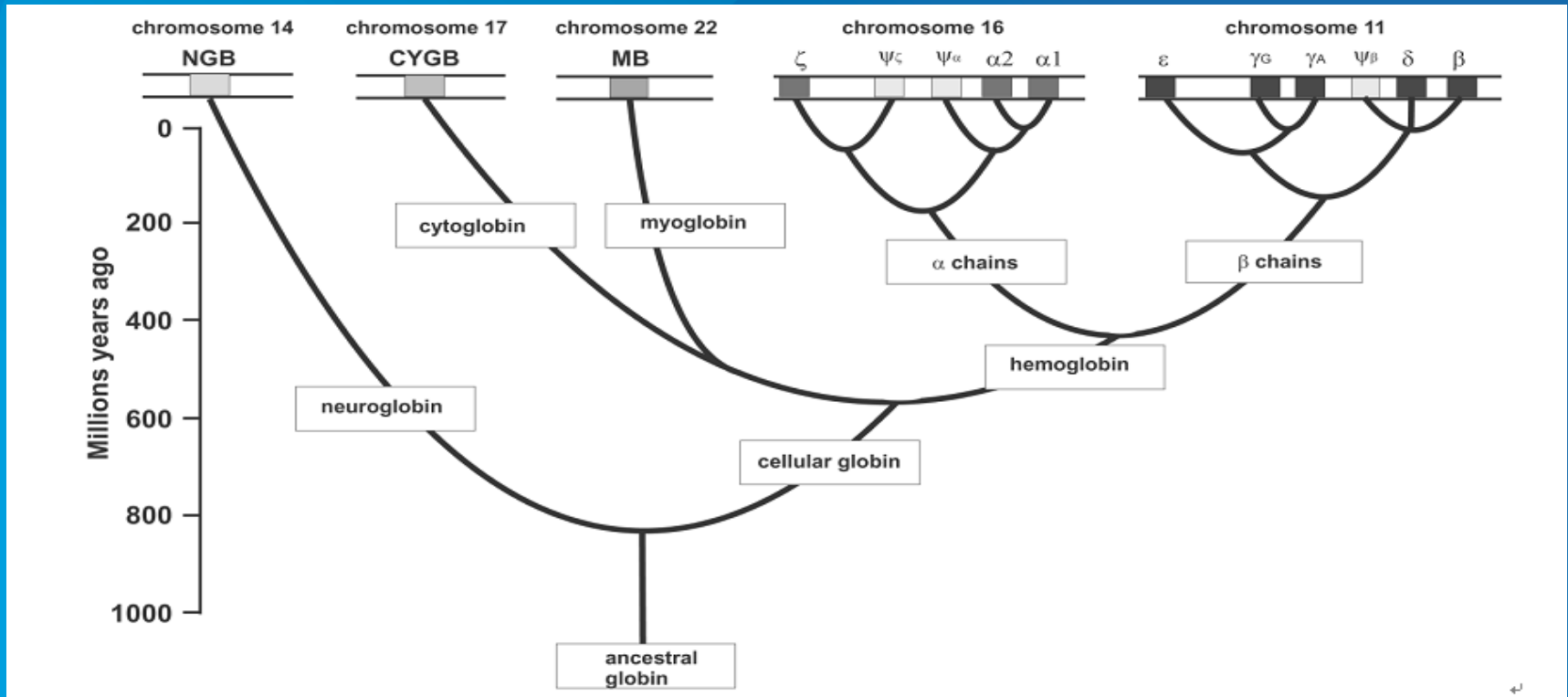
Data Display Highlight Statistics Help																																																													
<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> UUC <input type="checkbox"/> C V PI S 0 2 4 																																																													
	M S L T K T E R T I I V S M W A K I S T Q A D T I G T E T L E R L F L S H P Q T K T Y F																																																												
✓ 1. Zeta - gj12233	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-																																			
✓ 2. Alpha - gj15701	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	V	.	S	P	A	D	K	.	N	V	K	A	A	.	G	.	V	G	A	H	.	G	E	Y	.	A	.	A	.	.	M	.	.	.	F	.	T			
✓ 3. Delta - gj1227	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	V	H	.	.	P	E	.	K	.	A	V	N	A	L	.	G	.	V	N	-	-	V	.	A	V	.	G	.	A	.	G	.	.	L	V	V	Y	.	W	.	Q	R	F	.	
✓ 4. Gamma2 - gj15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	G	H	F	.	E	E	D	K	A	T	.	T	.	L	.	G	.	V	N	-	-	V	E	D	A	.	G	.	.	G	.	.	L	V	V	Y	.	W	.	Q	R	F	.		
✓ 5. Beta - gj158749	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	V	H	.	.	P	E	.	K	S	A	V	T	A	L	.	G	.	V	N	-	-	V	.	E	V	.	G	.	A	.	G	.	.	L	V	V	Y	.	W	.	Q	R	F	.	
✓ 6. Gamma1 - gj15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	G	H	F	.	E	E	D	K	A	T	.	T	.	L	.	G	.	V	N	-	-	V	E	D	A	.	G	.	.	G	.	.	L	V	V	Y	.	W	.	Q	R	F	.		
✓ 7. Theta - gj1223	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	A	.	S	A	E	D	.	A	L	V	R	A	L	.	K	.	L	G	S	N	V	G	V	Y	T	.	A	.	.	T	.	.	A	F	.	A	
✓ 8. Epsilon - gj122	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	V	H	F	.	A	E	.	K	A	A	V	T	.	L	.	S	.	M	N	-	-	V	E	E	A	.	G	.	A	.	G	.	.	L	V	V	Y	.	W	.	Q	R	F	.	
✓ 9. Mu - gj17474838	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	.	S	A	Q	.	.	A	Q	.	A	Q	V	.	D	L	.	A	G	H	E	A	Q	F	.	A	.	L	.	L	.	.	T	V	Y	.	S	.	.	V
✓ 10. Myoglobin - gj	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	G	.	S	D	G	.	W	Q	L	V	L	N	V	.	G	.	V	E	A	D	I	P	G	H	.	Q	.	V	.	I	.	.	K	G	.	.	E	.	L	E	K	.			
✓ 11. Cytoglobin - g	E	K	V	P	G	E	M	E	I	E	R	R	E	R	S	E	E	.	S	E	A	.	.	K	A	V	Q	A	.	.	R	L	Y	A	N	C	E	D	V	.	V	A	I	.	V	.	F	.	V	N	F	.	S	A	.	Q	.				
✓ 12. Neuroglobin -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	M	E	R	P	.	P	E	L	.	R	Q	S	.	R	A	V	.	R	S	P	L	E	H	.	.	V	L	F	A	.	.	.	A	L	E	.	D	L	L	P	L	.		

- C--保守位点 V--可变位点区
- PI--信息位点 S--单残基差异

选用Neighbor-Joining法建树

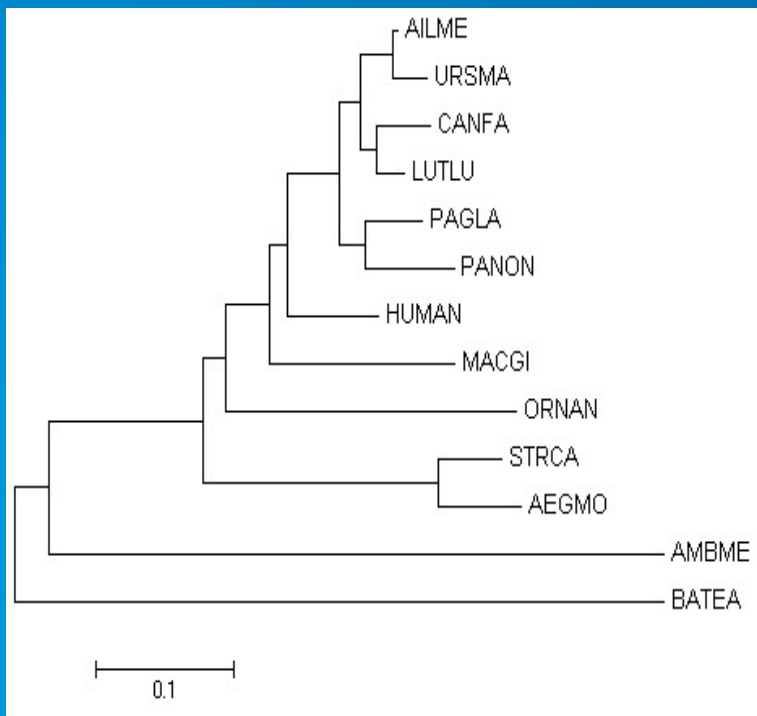


所得结果与文献中数据的对比



Thorsten Burmester Cytoglobin: A Novel Globin Type Ubiquitously Expressed in Vertebrate Tissues. *Mol. Biol. Evol.* 19(4):416–421. 2002

利用血红蛋白构建发育树的实际应用



氨基酸分析结果可以为生物的亲缘关系的研究提供更多的资料，目前研究最多的是血红蛋白和细胞色素C。

- 哺乳纲-食肉目-犬型亚目-熊科-熊猫亚科-大熊猫
- 两栖纲-有尾目-蝾螈科-蝾螈属-蝾螈
- 哺乳纲-食肉目-犬科-犬属-狗
- 哺乳纲-灵长目-人科-人属-智人种-人
- 哺乳纲-食肉目-鼬科-水獭属-水獭
- 哺乳纲-单孔目-鸭嘴兽科-鸭嘴兽属-鸭嘴兽
- 哺乳纲-食肉目-灵猫科-果子狸属-果子狸
- 鸟纲-新鸟亚纲-鸵形目-鸵形科-鸵鸟属-鸵鸟种
- 哺乳纲-食肉目-熊科-熊亚科-棕熊属-北极熊
- 鸟纲-今鸟亚纲-突胸总目-隼形目-鹰科-兀鹫亚科-秃鹫属
- 软骨鱼纲-板鳃亚纲-下孔总目-鳐目
- 哺乳纲-有袋目-袋鼠科-大袋鼠属-灰大袋鼠
- 哺乳纲-食肉目-猫科-豹亚科-豹属-美洲虎

第二部分 人血红蛋白分析

- 人的12种血红蛋白同源序列基因结构差异分析
- 人的12种血红蛋白同源序列相似性分析
- 38种哺乳动物血红蛋白alpha亚基相似性分析

人的12种血红蛋白同源序列基因结构差异分析

Sequences producing significant alignments:

		Score (Bits)	E Value	
<input checked="" type="checkbox"/>	sp P02008.2 HBAZ_HUMAN	RecName: Full=Hemoglobin subunit zeta;...	<u>182</u>	5e-47 G
<input checked="" type="checkbox"/>	sp P69892.2 HBG2_HUMAN	RecName: Full=Hemoglobin subunit gamma...	<u>180</u>	2e-46 G
<input checked="" type="checkbox"/>	sp P02042.2 HBD_HUMAN	RecName: Full=Hemoglobin subunit delta;...	<u>179</u>	5e-46 G
<input checked="" type="checkbox"/>	sp P69891.2 HBG1_HUMAN	RecName: Full=Hemoglobin subunit gamma...	<u>179</u>	6e-46 G
<input checked="" type="checkbox"/>	sp P68871.2 HBB_HUMAN	RecName: Full=Hemoglobin subunit beta; ...	<u>178</u>	1e-45 G
<input checked="" type="checkbox"/>	sp P69905.2 HBA_HUMAN	RecName: Full=Hemoglobin subunit alpha;...	<u>176</u>	3e-45 G
<input checked="" type="checkbox"/>	sp P02100.2 HBE_HUMAN	RecName: Full=Hemoglobin subunit epsilo...	<u>174</u>	1e-44 G
<input checked="" type="checkbox"/>	sp P09105.2 HBAT_HUMAN	RecName: Full=Hemoglobin subunit theta...	<u>171</u>	1e-43 G
<input checked="" type="checkbox"/>	sp Q6B0K9.1 HEM_HUMAN	RecName: Full=Hemoglobin subunit mu; Sh...	<u>168</u>	1e-42 G
<input checked="" type="checkbox"/>	sp P02144.2 MYG_HUMAN	RecName: Full=Myoglobin	<u>159</u>	4e-40 G
<input checked="" type="checkbox"/>	sp Q8WWM9.1 CYGB_HUMAN	RecName: Full=Cytoglobin; AltName: Ful...	<u>152</u>	7e-38 G
<input checked="" type="checkbox"/>	sp Q9NPG2.1 NGB_HUMAN	RecName: Full=Neuroglobin	<u>132</u>	7e-32 G

Query	Full Name	ID	Sequence length
P02008	Hemoglobin subunit zeta	HBAZ_HUMAN	142
P69892	Hemoglobin subunit gamma-2	HBG2_HUMAN	147
P02042	Hemoglobin subunit delta	HBD_HUMAN	147
P69891	Hemoglobin subunit gamma-1	HBG1_HUMAN	147
P68871	Hemoglobin subunit beta	HBB_HUMAN	147
P69905	Hemoglobin subunit alpha	HBA_HUMAN	142
P02100	Hemoglobin subunit epsilon	HBE_HUMAN	147
P09105	Hemoglobin subunit theta-1	HBAT_HUMAN	142
Q6B0K9	Hemoglobin subunit mu	HBM_HUMAN	141
P02144	Myoglobin	MYG_HUMAN	154
Q8WWM9	Cytoglobin	CYGB_HUMAN	190
Q9NPG2	Neuroglobin	NGB_HUMAN	151

用“HBA_HUMAN”在 ExPASy 中提取序列, 用 blastP 在人的基因组中检索到 12 条同源序列

所得的 12 种蛋白的基本信息

利用GSDS软件分析基因结构差异

- 用搜索到的人12条序列的ID在ExPASy中检索
- 用“Cross Reference”查找对应蛋白序列的全长CDS和完整基因核酸序列
- 将12条CDS核酸序列链接成一个FASTA文件，将12条基因序列链接成另一个FASTA文件
- 运行GSDS

The CDS sequence and the genomic sequence of the same gene **must** use the **same sequence ID** in fasta file

● **CDS sequences** (Fasta ?):

Load file:

or paste data:

```
>HBZ J00182 HBZ_CDS
atgtctctgaccaagactgagaggaccatcattgtgtcc
atgtggccaagatctccacgcaggccgacaccatcggc
accgagactctggagaggctcttcctcagccaccgcag
accaagacctacttccgcacttcgaacctgcaccgggg
```

● **Genomic sequences** (Fasta ?)

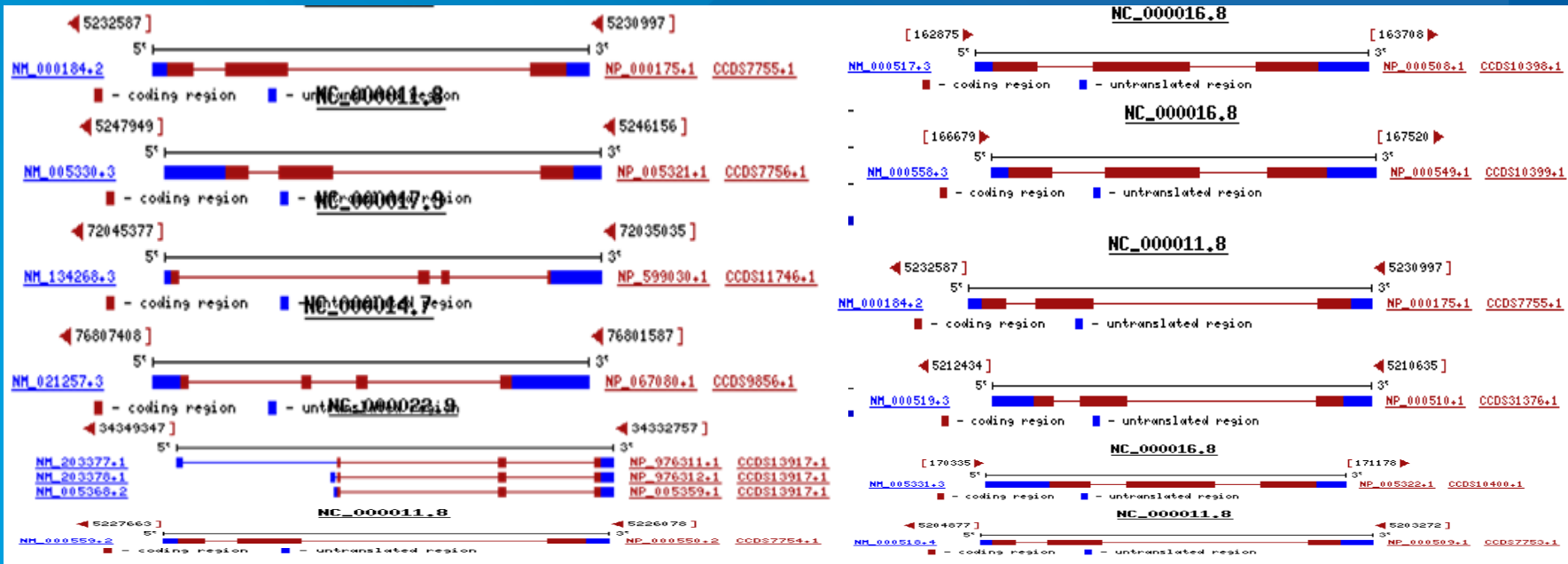
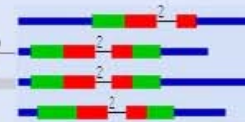
Load file:

or paste data:

```
>HBZ J00182 HBZ_gene
accgaaggccagtcctgagcaggcccaactccagtcgagc
tgcccacctgcccctatgtctctgaccaagactgagag
gaccatcattgtgtccatgtggccaagatctccacgca
ggccgacaccatcggcaccgagactctggagaggtgagt
```

GSDS

Gene Structure Display Server



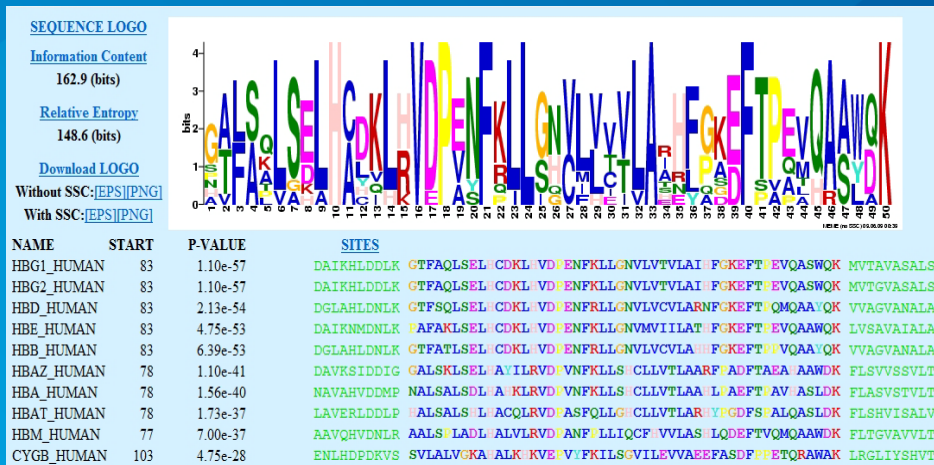
- 大部分含3个外显子，少数的含有一个外显子
- 内含子的序列长度差异很大
- 在进化的过程中，由同一祖先进化过来的基因内含子外显子产生了一定的变化，这有可能是物种进化的一种方式

人的12种血红蛋白同源序列相似性分析

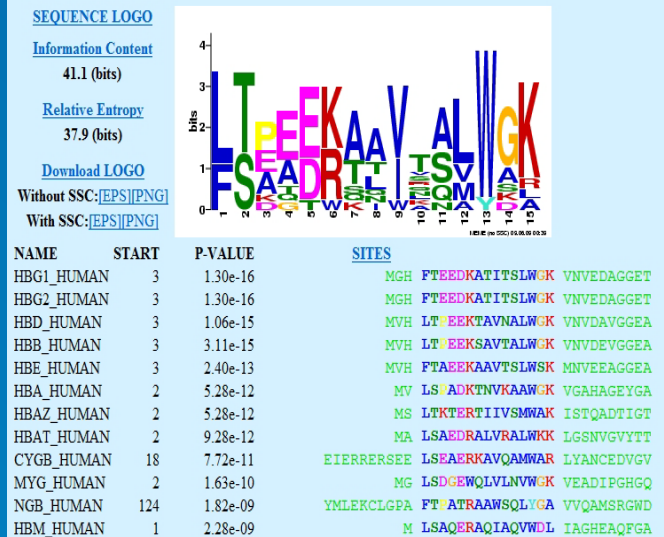
用HBA_human蛋白质序列进行PSI-BLAST，
找出人类基因组中全部同源序列，结果显示共有
12条

Sequences producing significant alignments:			Score	E	
			(Bits)	Value	
<input checked="" type="checkbox"/>	sp P02008.2 HBAZ_HUMAN	RecName: Full=Hemoglobin subunit zeta;...	<u>194</u>	1e-50	G
<input checked="" type="checkbox"/>	sp P69905.2 HBA_HUMAN	RecName: Full=Hemoglobin subunit alpha;...	<u>187</u>	1e-48	G
<input checked="" type="checkbox"/>	sp P02042.2 HBD_HUMAN	RecName: Full=Hemoglobin subunit delta;...	<u>186</u>	3e-48	G
<input checked="" type="checkbox"/>	sp P69892.2 HBG2_HUMAN	RecName: Full=Hemoglobin subunit gamma...	<u>186</u>	5e-48	G
<input checked="" type="checkbox"/>	sp P68871.2 HBB_HUMAN	RecName: Full=Hemoglobin subunit beta; ...	<u>184</u>	1e-47	G
<input checked="" type="checkbox"/>	sp P69891.2 HBG1_HUMAN	RecName: Full=Hemoglobin subunit gamma...	<u>184</u>	2e-47	G
<input checked="" type="checkbox"/>	sp P09105.2 HBAT_HUMAN	RecName: Full=Hemoglobin subunit theta...	<u>182</u>	7e-47	G
<input checked="" type="checkbox"/>	sp P02100.2 HBE_HUMAN	RecName: Full=Hemoglobin subunit epsilo...	<u>179</u>	6e-46	G
<input checked="" type="checkbox"/>	sp Q6B0K9.1 HBM_HUMAN	RecName: Full=Hemoglobin subunit mu; Sh...	<u>176</u>	3e-45	G
<input checked="" type="checkbox"/>	sp P02144.2 MYG_HUMAN	RecName: Full=Myoglobin	<u>161</u>	1e-40	G
<input checked="" type="checkbox"/>	sp Q8WWM9.1 CYGB_HUMAN	RecName: Full=Cytoglobin; AltName: Ful...	<u>161</u>	1e-40	G
NEW <input checked="" type="checkbox"/>	sp Q9NPG2.1 NGB_HUMAN	RecName: Full=Neuroglobin	<u>58.2</u>	2e-09	G

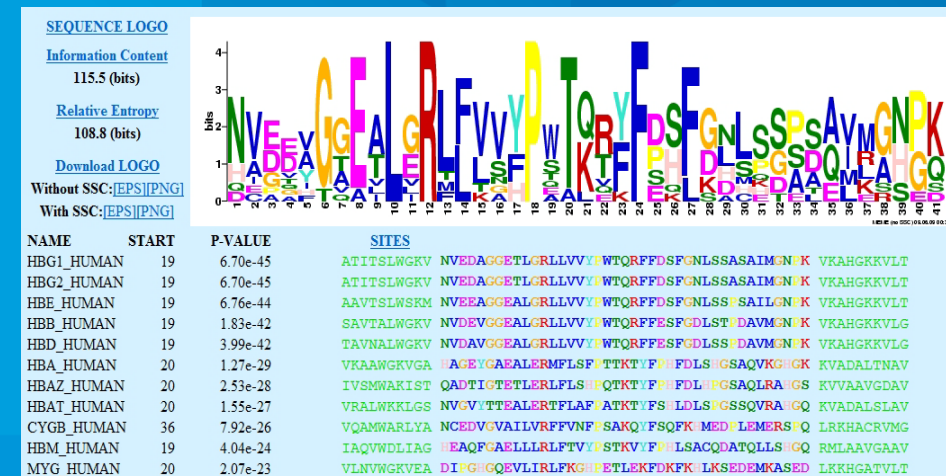
用MEME预测12条序列中的保守motif



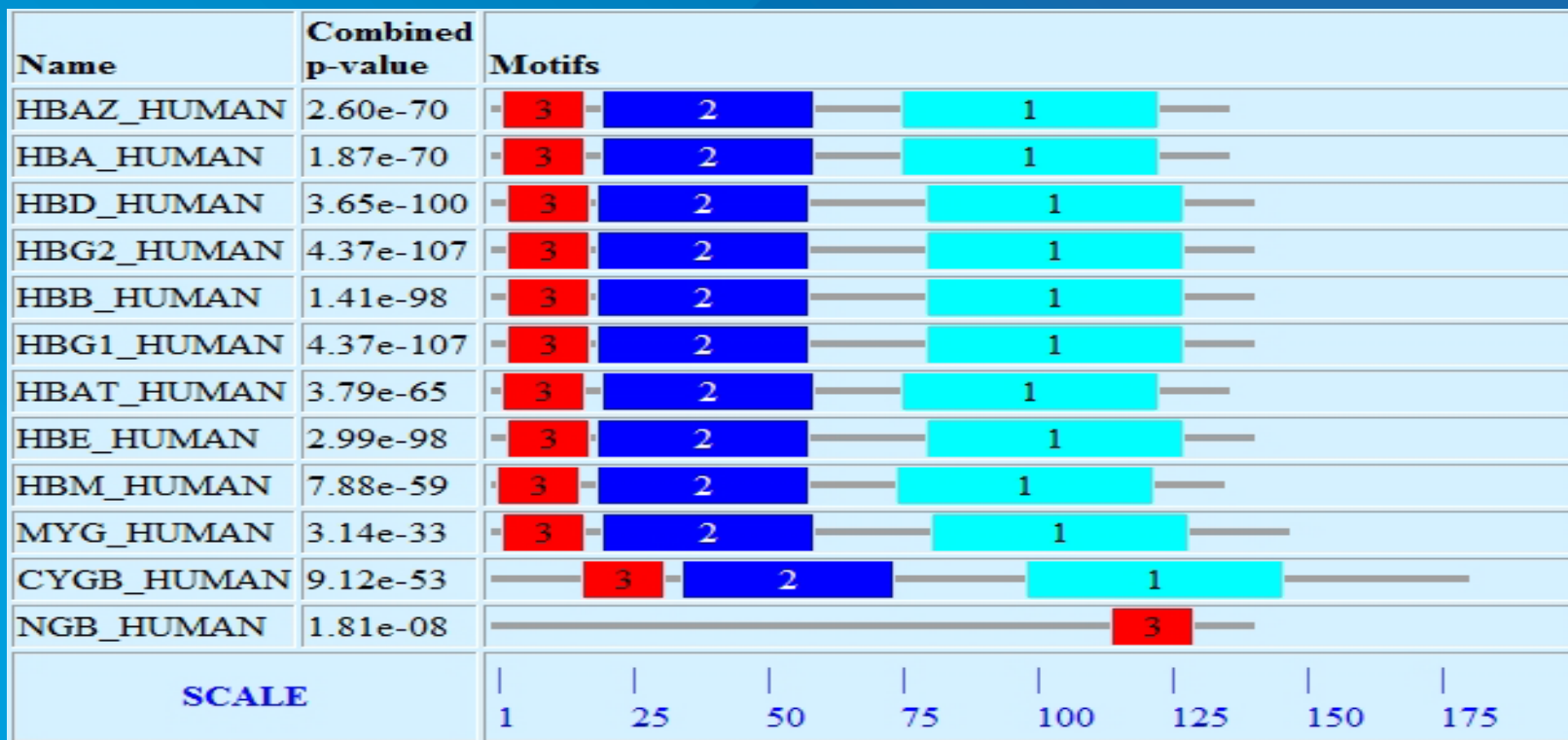
motif 1



motif 3



12条序列的motif组成结构：



为了形成各自特定的功能，而产生了特异性的分化，从而导致了基因结构的差异

38种哺乳动物血红蛋白alpha亚基相似性分析

在SRS中提取38种哺乳动物的血红蛋白alpha亚基蛋白序列

The screenshot shows the EMBL-EBI SRS search interface. At the top, there is a search bar with "All Databases" selected and "Enter Text Here" as a placeholder. Below the search bar are navigation tabs: Databases, Tools, EBI Groups, Training, Industry, About Us, Help, Site Index, and a logo. A secondary row of tabs includes Quick Search, Library Page, Query Form (selected), Tools, Results, Projects, Views, Databanks, and HEL. The main search area has a "Reset" button and the text "search UniProtKB/Swiss-Prot".

Search Options

Combine search terms with: & (AND)
Use wildcards
Get results of type: Entry

Fields you can search	Your search terms
<input type="button" value="i"/> AllText <input type="button" value="v"/>	"Hemoglobin subunit alpha"&"mammalia"
<input type="button" value="i"/> Sequence Length <input type="button" value="v"/>	142
<input type="button" value="i"/> AllText <input type="button" value="v"/>	
<input type="button" value="i"/> AllText <input type="button" value="v"/>	

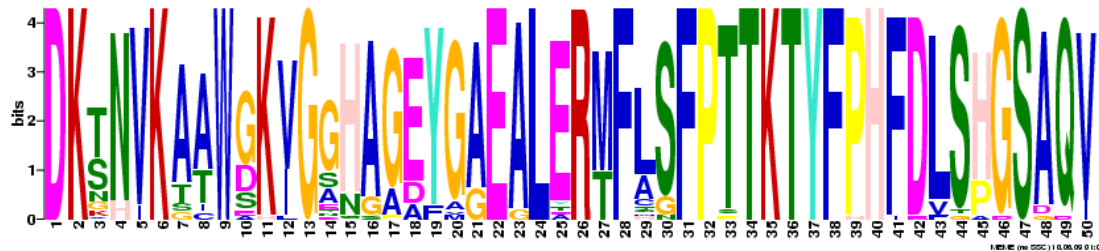
In a single field, you can separate multiple values by: &, | or !

用MEME预测38条序列中的保守motif

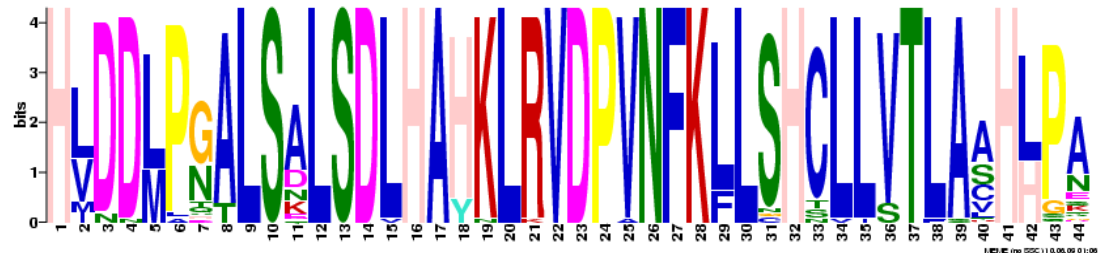
Name	Combined p-value	Motifs
spP18969HBA_AILFU	1.60e-115	1 2 3
spP18970HBA_AILME	1.47e-117	1 2 3
spP01971HBA_ALCAA	5.70e-112	1 2 3
spP14387HBA_ANTPA	1.64e-117	1 2 3
spP67817HBA_ATEGE	7.20e-121	1 2 3
spP18971HBA_BALAC	2.41e-104	1 2 3
spP01969HBA_BOSGF	1.18e-113	1 2 3
spP01966HBA_BOVIN	2.00e-114	1 2 3
spP60330HBA_CANLA	1.50e-112	1 2 3
spP01947HBA_CAVPO	3.53e-102	1 2 3
spP01923HBA_CEBAP	1.14e-120	1 2 3
spP01926HBA_CERAE	3.22e-120	1 2 3
spP01963HBA_CERSI	1.37e-114	1 2 3
spP01930HBA_COLBA	3.52e-119	1 2 3
spP19014HBA_CRIGA	3.19e-111	1 2 3
spP07419HBA_DASVI	1.15e-106	1 2 3
spP01919HBA_EQUAS	6.26e-115	1 2 3
spQ9X3E9HBA_EQUFR	5.77e-116	1 2 3
spP01960HBA_EQUZE	6.26e-115	1 2 3
spP01935HBA_HORSE	1.09e-117	1 2 3
spP69905HBA_HUMAN	2.39e-121	1 2 3
spP81043HBA_MACEU	1.09e-108	1 2 3
spP63107HBA_MACFU	1.56e-122	1 2 3
spP63108HBA_MACMU	1.56e-122	1 2 3
spQSHY34HBA_MONDO	1.53e-94	1 2 3
spP01942HBA_MOUSE	8.09e-118	1 2 3
spQOZA50HBA_PANHO	1.84e-114	1 2 3
spP18975HBA_PANLE	1.77e-118	1 2 3
spP63109HBA_PANON	1.77e-118	1 2 3
spP69906HBA_PANPA	2.39e-121	1 2 3
spP69907HBA_PANTR	2.39e-121	1 2 3
spP63111HBA_PAPAN	4.49e-116	1 2 3
spP63112HBA_PAPCY	4.49e-116	1 2 3
spP06635HBA_PONPY	2.98e-119	1 2 3
spP01957HBA_PROHA	1.78e-113	1 2 3
spP68235HBA_URSMA	9.72e-116	1 2 3
spP68237HBA_URSMML	9.72e-116	1 2 3
spP68236HBA_URSTH	9.72e-116	1 2 3

SCALE	1	25	50	75	100	125
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1
11	1	1	1	1	1	1
12	1	1	1	1	1	1
13	1	1	1	1	1	1
14	1	1	1	1	1	1
15	1	1	1	1	1	1
16	1	1	1	1	1	1
17	1	1	1	1	1	1
18	1	1	1	1	1	1
19	1	1	1	1	1	1
20	1	1	1	1	1	1
21	1	1	1	1	1	1

MOTIF 1 width = 50 sites = 38 llr = 4874 E-value = 1.6e-1630



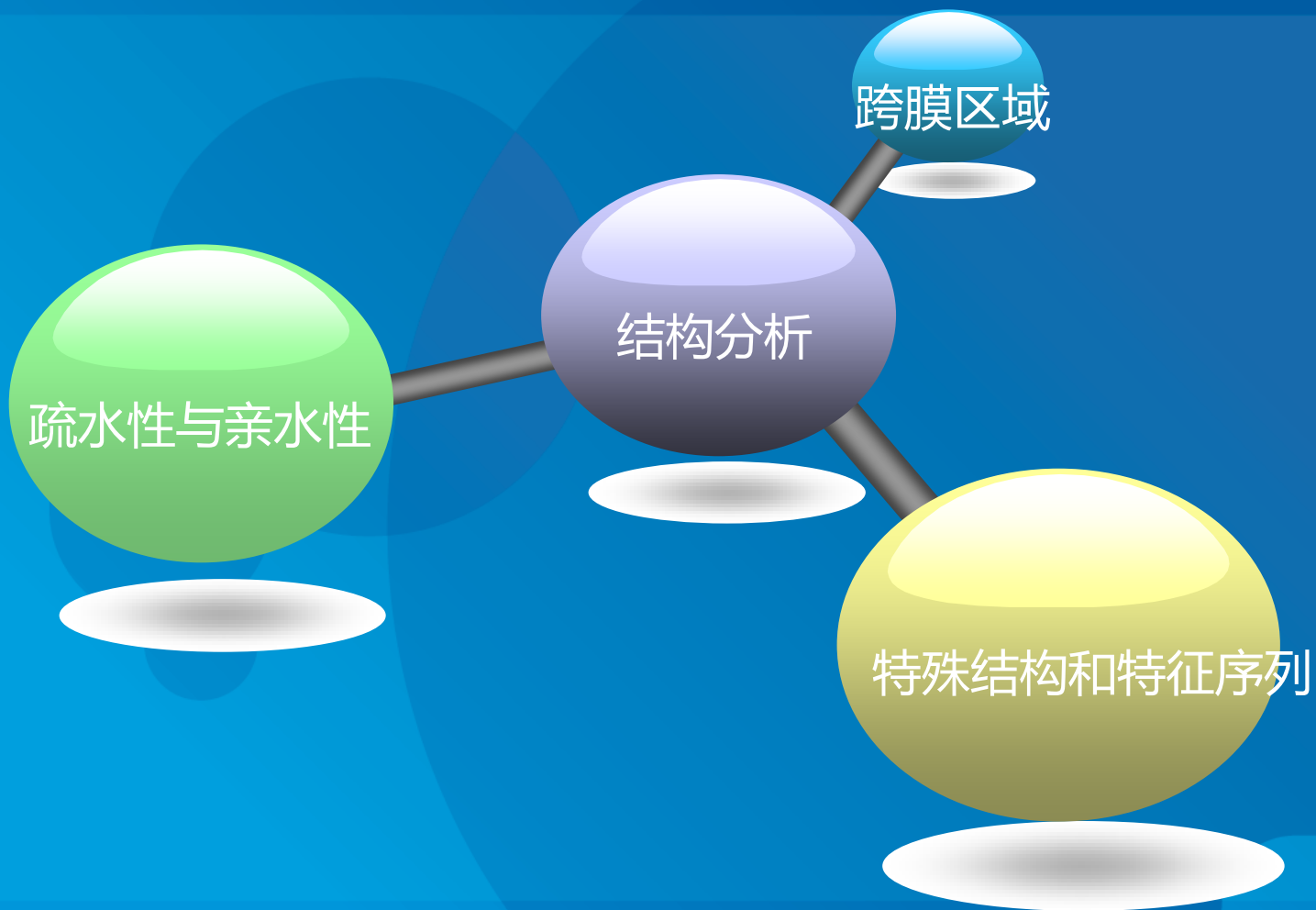
MOTIF 2 width = 44 sites = 38 llr = 3897 E-value = 1.4e-1239



MOTIF 3 width = 21 sites = 38 llr = 1952 E-value = 3.6e-593



第三部分 人血红蛋白alpha亚基结构分析



氨基酸残基疏水性与亲水性分析

首先用pepstats, 统计各种氨基酸残基的含量, 结果如下:

Residue	Number	Mole%	DayhoffStat
A = Ala	21	14.789	1.720
B = Asx	0	0.000	0.000
C = Cys	1	0.704	0.243
D = Asp	8	5.634	1.024
E = Glu	4	2.817	0.469
F = Phe	7	4.930	1.369
G = Gly	7	4.930	0.587
H = His	10	7.042	3.521
I = Ile	0	0.000	0.000
J = ---	0	0.000	0.000
K = Lys	11	7.746	1.174
L = Leu	18	12.676	1.713
M = Met	3	2.113	1.243
N = Asn	4	2.817	0.655
O = ---	0	0.000	0.000
P = Pro	7	4.930	0.948
Q = Gln	1	0.704	0.181
R = Arg	3	2.113	0.431
S = Ser	11	7.746	1.107
T = Thr	9	6.338	1.039
U = ---	0	0.000	0.000
V = Val	13	9.155	1.387
W = Trp	1	0.704	0.542
X = Xaa	0	0.000	0.000
Y = Tyr	3	2.113	0.621
Z = Glx	0	0.000	0.000

Molecular weight = 15257.51

Residues = 142

Charge = 7.0

Average Residue Weight = 107.447

Isoelectric Point = 9.0879

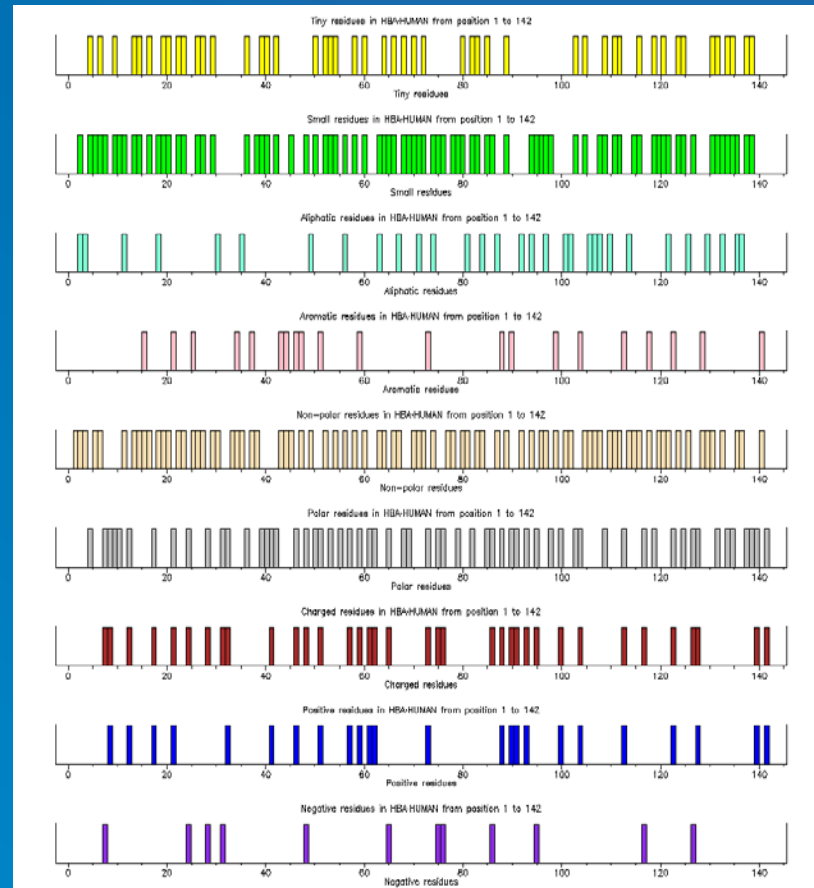
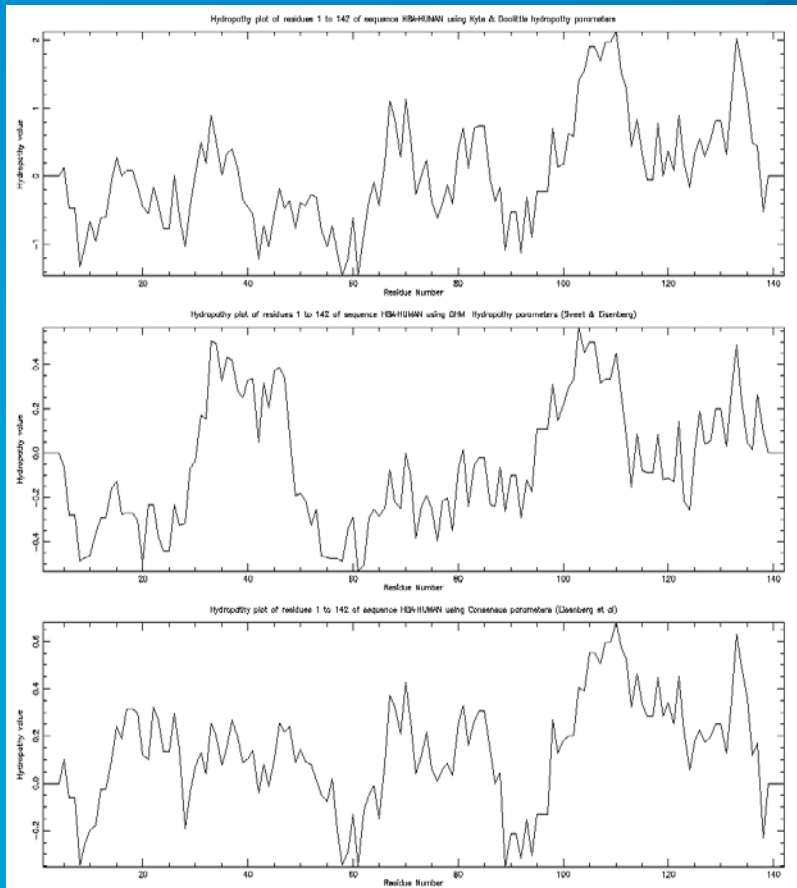
A280 Molar Extinction Coefficient = 9530 A280

Extinction Coefficient 1mg/ml = 0.62

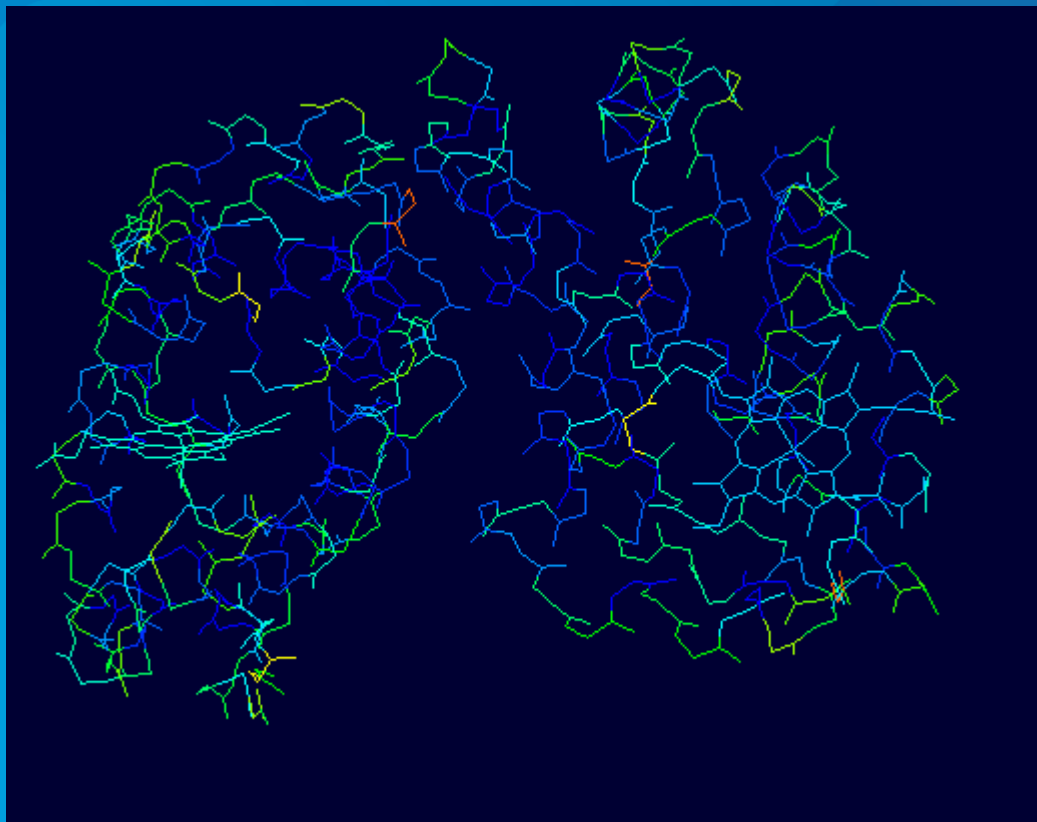
Improbability of expression in inclusion bodies = 0.724

Property	Residue	Number	Mole%
Tiny	(A+C+G+S+T)	49	34.507
Small	(A+B+C+D+G+N+P+S+T+V)	81	57.042
Aliphatic	(A+I+L+V)	52	36.620
Aromatic	(F+H+W+Y)	21	14.789
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	81	57.042
Polar	(D+E+H+K+N+Q+R+S+T+Z)	61	42.958
Charged	(B+D+E+H+K+R+Z)	36	25.352
Basic	(H+K+R)	24	16.901
Acidic	(B+D+E+Z)	12	8.451

用pepinfo，以图形方式显示不同的氨基酸残基含量

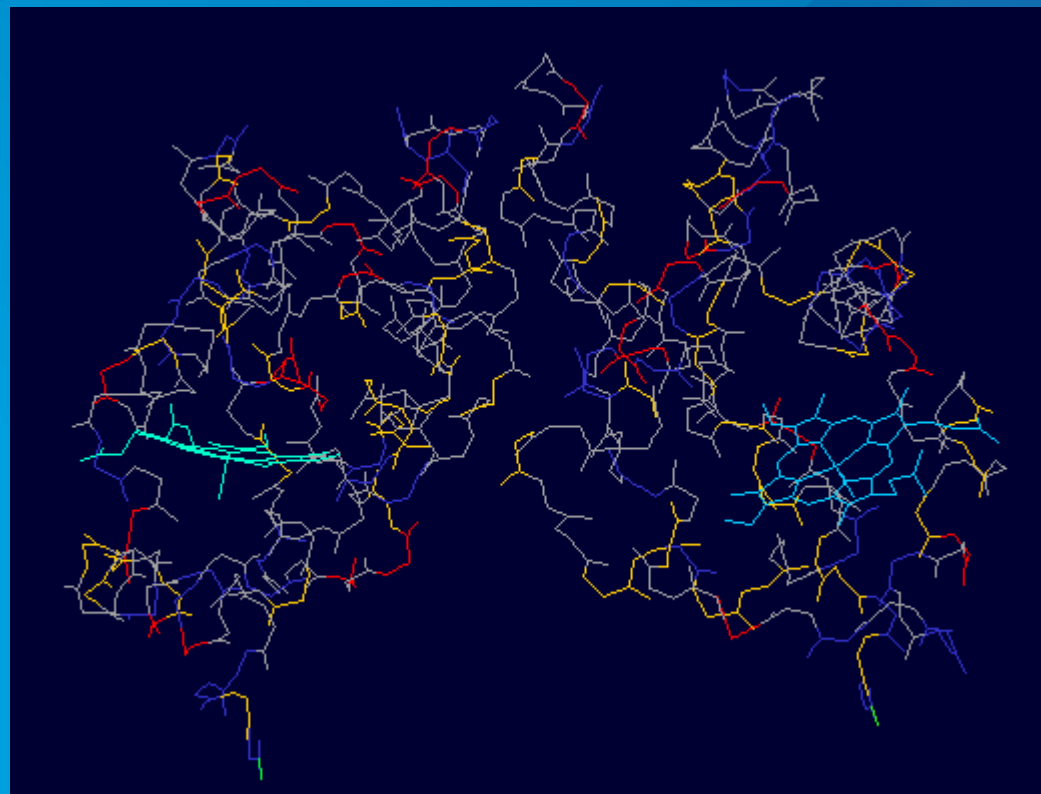


用SPDBV分析，在color中选择accessibility



Color---by Accessibility
在结构中每个氨基酸残基与周围溶剂接触程度的多少决定了残基的颜色。与溶剂接触最少的是蓝色，完全露在分子表面的是红色，接触介于2者之间的，用蓝色和红色中间的颜色表示，如蓝绿色和洋红色。

Color中选择type

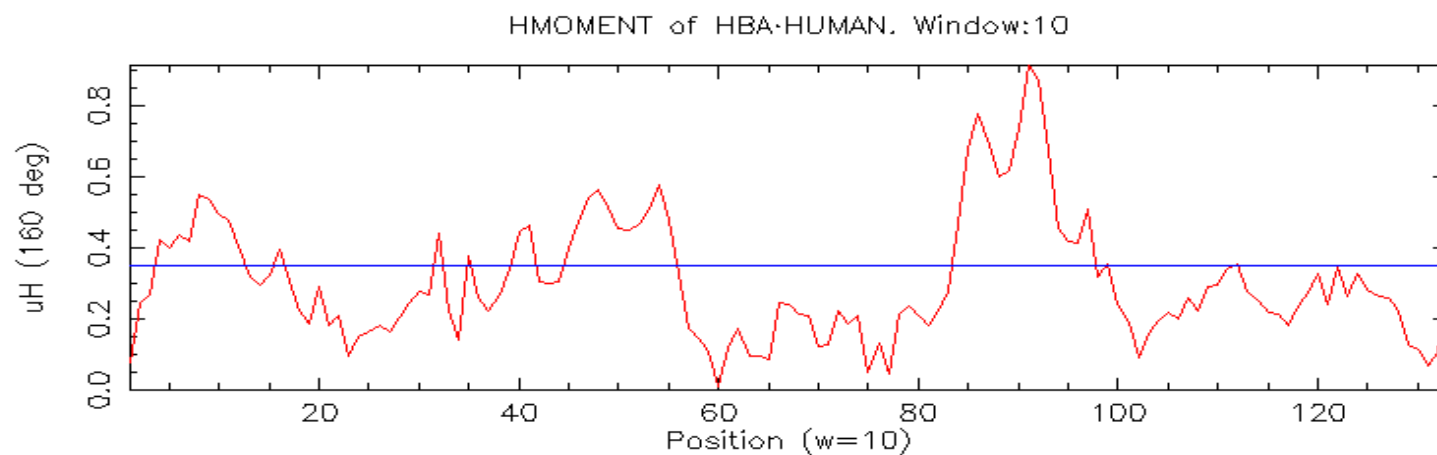
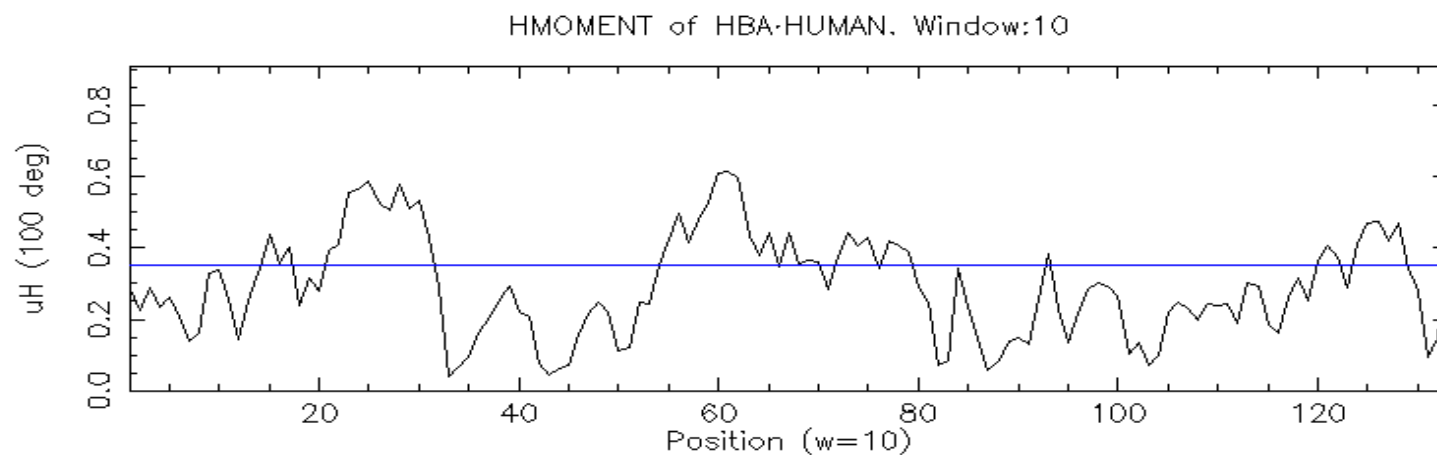


对结构模型染色的依据是残基的化学类型：带正电的用蓝色表示；带负电用红色表示；不带电的用黄色表示；无极性的用灰色表示。

接着用garnier，预测蛋白质序列中的二级结构

```
#-----  
      .   10   .   20   .   30   .   40   .   50  
      MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLS  
helix   HHHHHHHHHHHHHH      HHHHHHHHHHHHHH  
sheet  EEEE                                     EEE E  
turns                                     TTTT   T T  
coil   .         CCCCCC         CC         CC  
      .   60   .   70   .   80   .   90   .  100  
      HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFK  
helix   HHHH      HHHHHHHHHH  HHHHHHHHHHHHHHHHHHHH  HHHHH  
sheet   .         EE         EE         E EE  
turns   .         TT         .         T  
coil   CCC      CC  
      .  110   .  120   .  130   .  140  
      LLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTIVLISKYR  
helix  HHHHHHHHHHHHHH  HHHHHHHHHHHHHHHHHHHHHHHHHHHH  
sheet   .         EEEE  
turns   .         TTTT  
coil   .         C  
#-----  
#  
# Residue totals: H: 93   E: 19   T: 14   C: 16  
# percent: H: 73.8 E: 15.1 T: 11.1 C: 12.7
```


也可以用hmoment，以图形的方式预测蛋白质序列中可能形成 α 螺旋和 β 折叠的片段



用helixturnhelix和pepcoil，预测该序列当中是否有这些特殊的domain

```
#####
# Program: helixturnhelix
# Rundate: Wed 10 Jun 2009 16:47:15
# Commandline: helixturnhelix
#   -auto
#   -minsd 2.5
#   -mean 238.71
#   -sd 293.61
#   -sequence 256638
#   -outfile 257090
# Report_format: motif
# Report_file: 257090
#####
#=====
#
# Sequence: HBA_HUMAN      from: 1    to: 142
# HitCount: 0
#
# Hits above +2.50 SD (972.73)
#
#=====
Maximum_score_at at "*"
#-----
#-----
#-----
# Total_sequences: 1
# Total_hitcount: 0
#-----
```

```
#####
# Program: pepcoil
# Rundate: Wed 10 Jun 2009 16:50:32
# Commandline: pepcoil
#   -auto
#   -other N
#   -coil
#   -window 28
#   -sequence 256638
#   -outfile 257091
# Report_format: motif
# Report_file: 257091
#####
#=====
#
# Sequence: HBA_HUMAN      from: 1    to: 142
# HitCount: 0
#
# Window size: 28 residues
#
#=====
Max_coil_pos at "*"
#-----
#-----
```

- 结果显示：没有这些domain

最后用patmatmotif，预测该蛋白具有哪些特征序列

LENGTH	START	END	MOTIF
4	59	62	AMIDATION

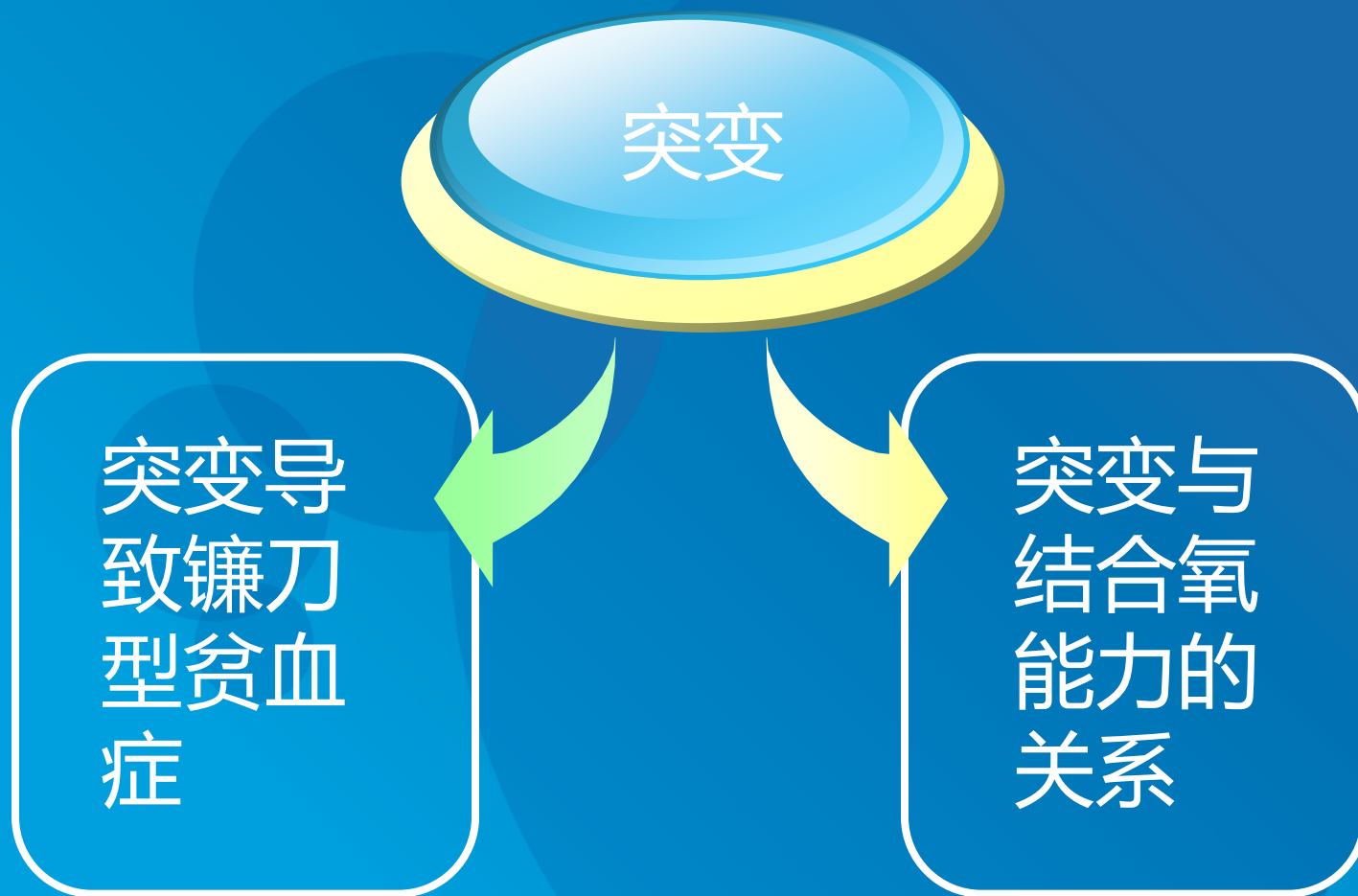
AQVKGHGKKVADAL

 | |

 59 62

结果显示：有1个Amidation酰胺化位点

第四部分 血红蛋白功能预测的验证



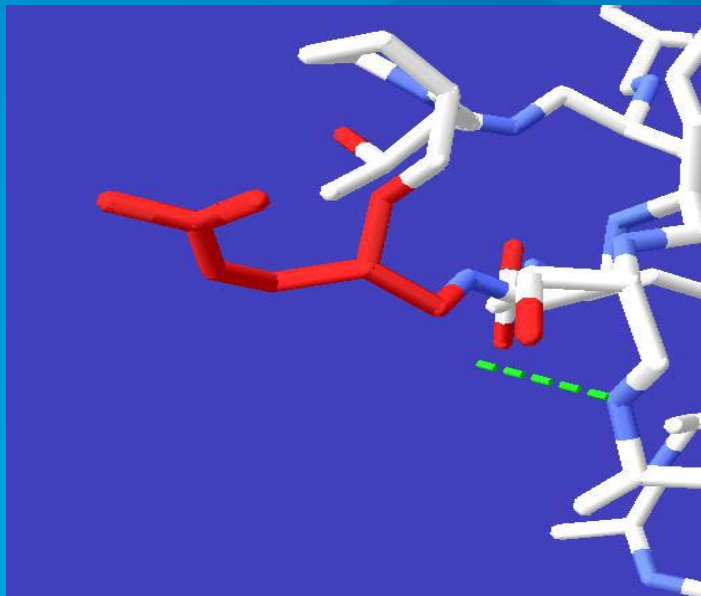
镰刀型红细胞贫血症

- 镰刀型红细胞贫血症是一种“分子病”，即分子结构、特别是蛋白质分子结构发生遗传性变化而造成的病变。
- 为什么会出现这样的遗传病呢？

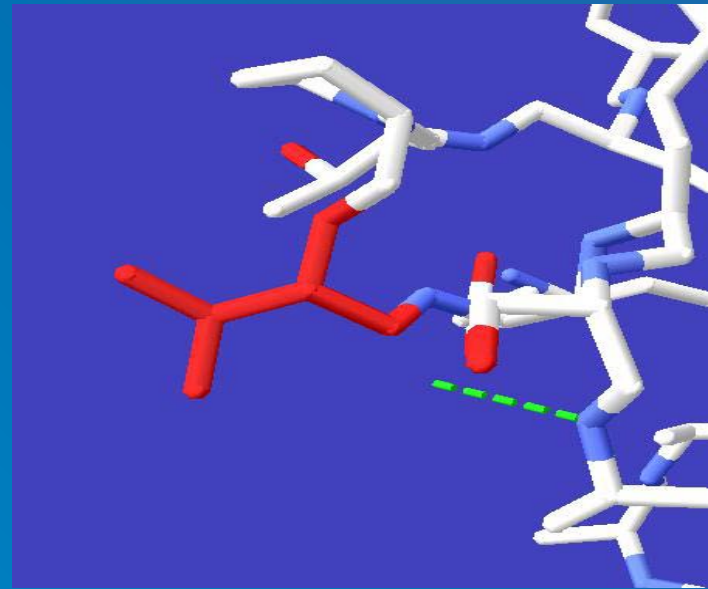
研究表明：出现镰刀型红细胞，原因在于：

- 异常血红蛋白 β 链的第6位谷氨酸被缬氨酸所代替。
- 由于带负电的极性亲水谷氨酸被不带电的非极性疏水缬氨酸所代替，致使血红蛋白的溶解度下降。在氧张力低的毛细血管区，血红蛋白形成管状凝胶结构(如棒状结构)，导致红细胞扭曲成镰刀状(即镰变)。

β 链的第6位谷氨酸被缬氨酸所代替

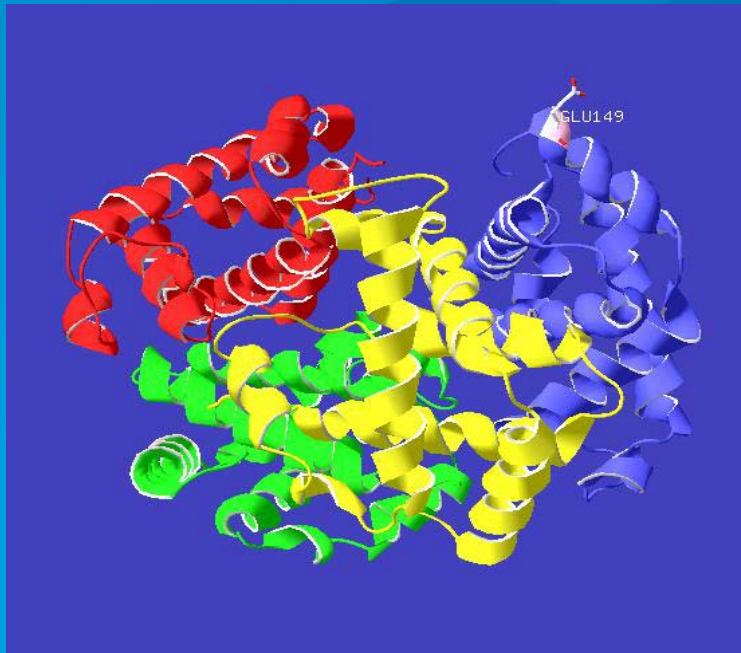


未突变的 β 链的第6位 Glu

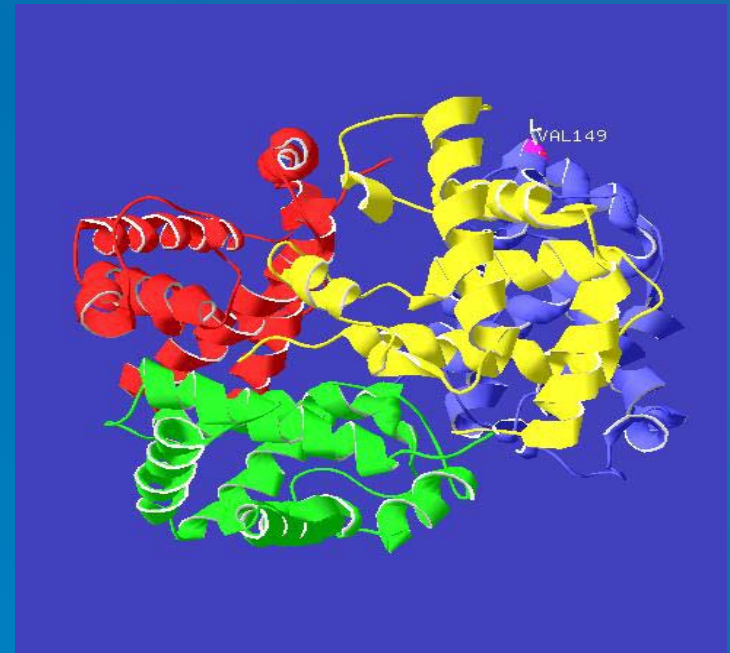


突变后 β 链的第6位 Val

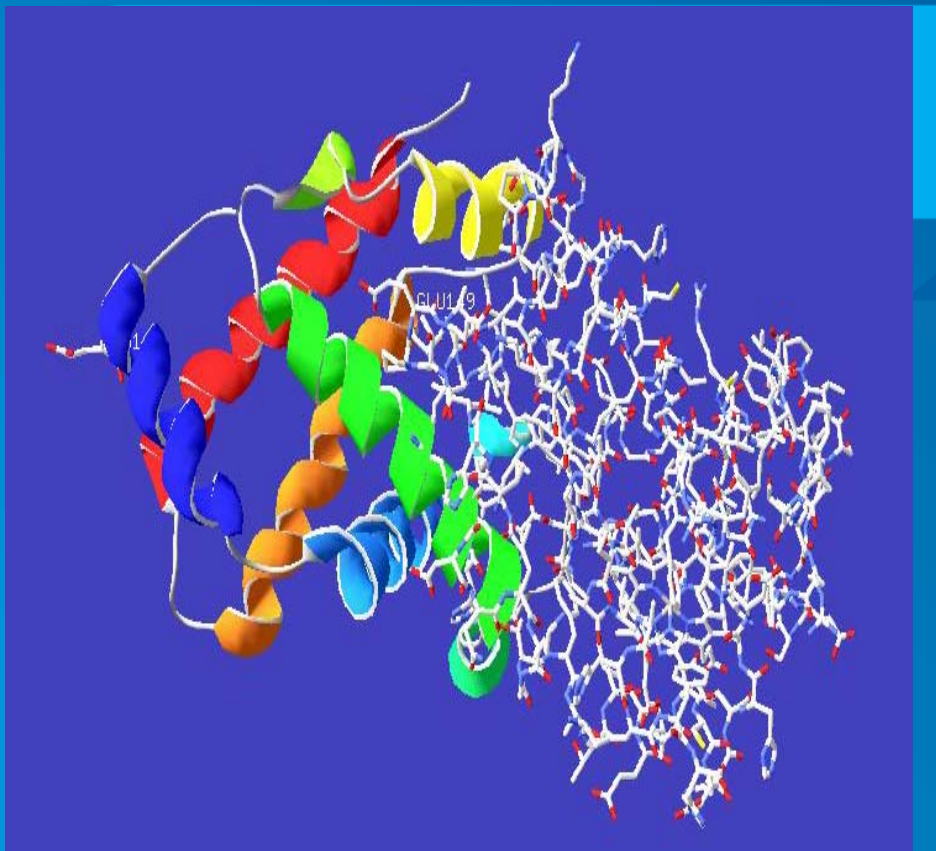
β 链的第6位谷氨酸被缬氨酸所代替



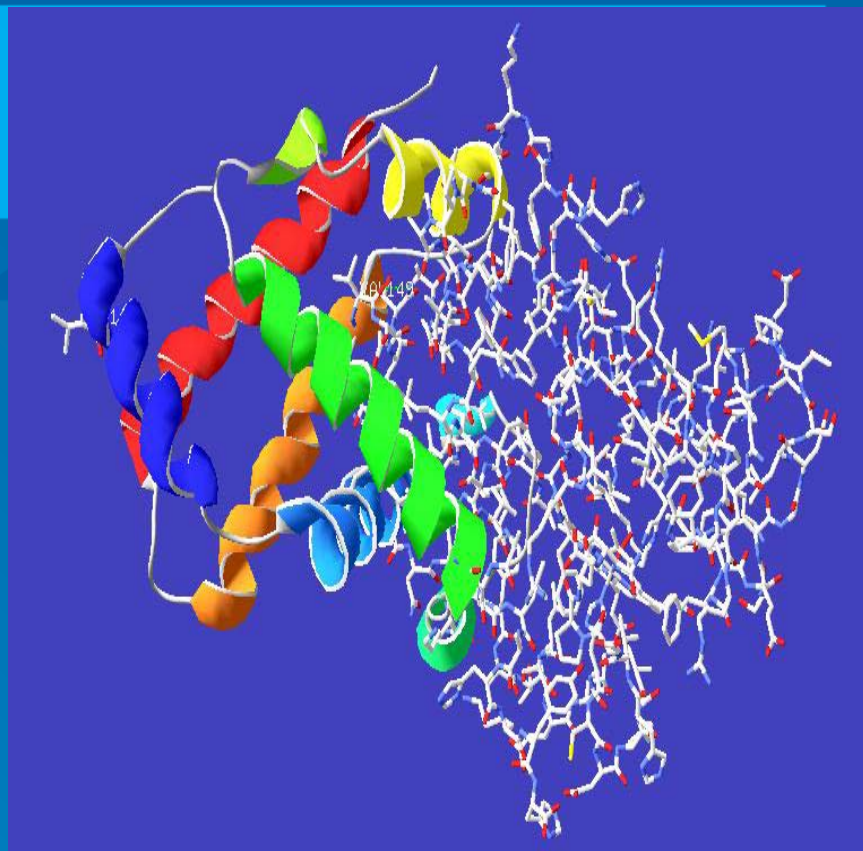
未突变的
hemoglobin



突变的 hemoglobin

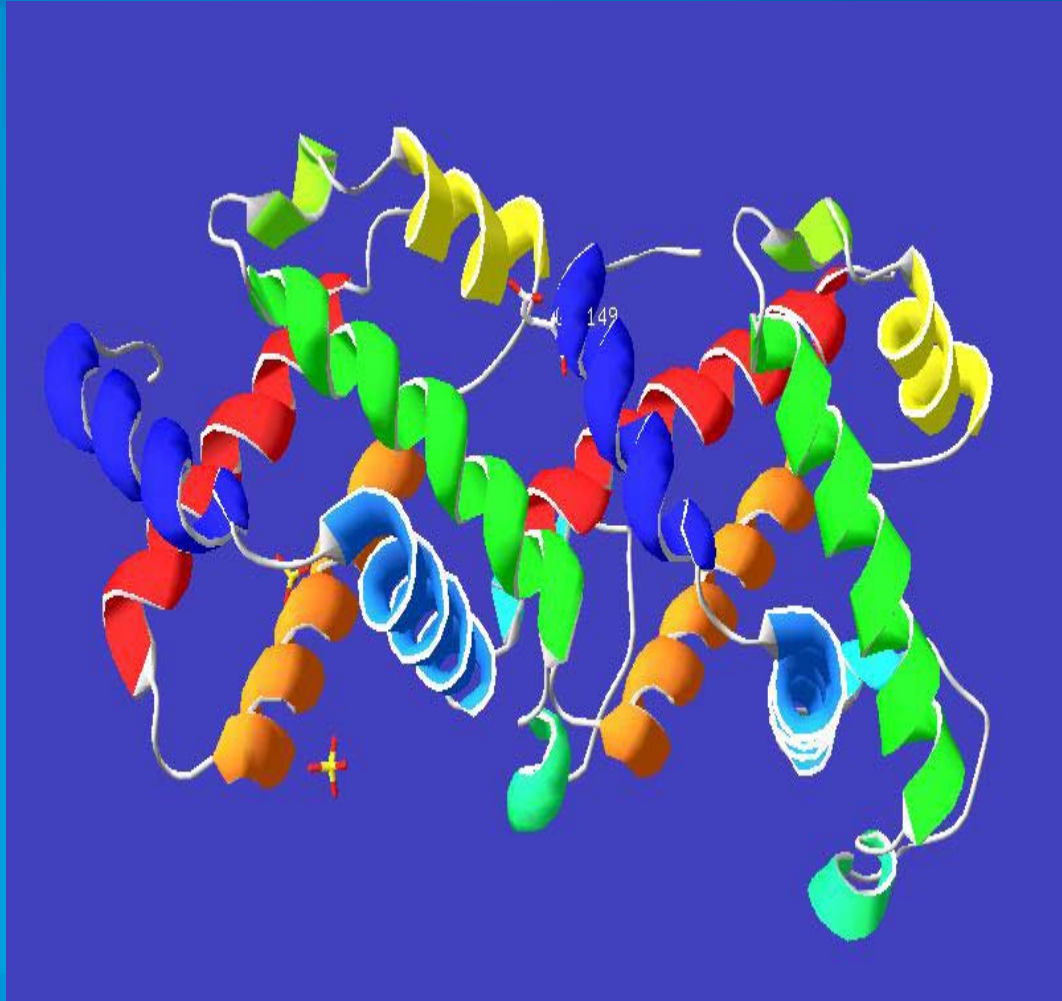


β 链的第6位Glu未突变时
与相邻的血红蛋白 β 亚基
不能相连



β 链的第6位Val突变时与
相邻突变的 β 亚基相连

为什么脱氧合血红蛋白镰刀型化而氧合血红蛋白（携氧）不镰刀型化？



通过图中的分析，可以简单解释为：在氧合形式中，血红蛋白亚基的重新排列使 β 链的口袋不能接受相邻的血红蛋白分子。

第五部分 斑头雁血红蛋白序列，结构和功能的分析



引言

斑头雁是候鸟中一个特殊的物种。它们夏天居住在青海湖，而在秋天会飞过青藏高原到达印度再在春天飞回。有趣的是，与斑头雁亲缘关系接近的灰雁长年居住在印度低地，并不迁徙。通过两者血红蛋白的序列对比发现仅有4处发生了替换。灰雁alpha亚基中第119位脯氨酸在斑头雁中变为丙氨酸。该残基位于alpha/beta的分界面上。目前研究认为，这种残基的改变导致脱氧状态时张力状态的变化使得alpha和beta亚基之间的接触减少，从而增强了对氧的亲合力。下面我们将用SPDBV软件来证明该结论。

1、序列比对

- 从NCBI上分别搜索1a4f和1faw可以得到斑头雁和灰雁的血红蛋白alpha和beta亚基的蛋白质序列各两条。
- 以fasta格式保存到本地计算机，并用wablab中的needle工具进行序列比对，注意的是alpha和beta亚基分别进行比对。

LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
141	711.0	138/141 (97.9%)	138/141 (97.9%)	0/141 (0.0%)
1FAWA	1	VLSAADKINVKGVFSKIGSHAE EYGAETLERMFTAYPQTKITYFPHF DLQH	50	
1A4FA	1	VLSAADKINVKGVFSKIGSHAE EYGAETLERMFTAYPQTKITYFPHF DLQH	50	
1FAWA	51	GSAQIKAHGKKVAAALVEAVNHIDDIA GALS KLS DL HAQKLRVDPVNF KF	100	
1A4FA	51	GSAQIKAHGKKVAAALVEAVNHIDDIA GALS KLS DL HAQKLRVDPVNF KF	100	
1FAWA	101	LGHCFLVVVAIHHP SALTAEVHASLDKFLCAVGT VLTAKYR	141	
1A4FA	101	LGHCFLVVVAIHHP SALTAEVHASLDKFLCAVGT VLTAKYR	141	
LENGTH	SCORE	IDENTITY	SIMILARITY	GAPS
146	763.0	145/146 (99.3%)	146/146 (100.0%)	0/146 (0.0%)
1FAWB	1	VHVSAAEEKQLITGLW GKVNVADCGAEALARLLIVYPWTQRFFSSFCNLSS	50	
1A1FB	1	VHVSAAEEKQLITGLW GKVNVADCGAEALARLLIVYPWTQRFFSSFCNLSS	50	
1FAWB	51	PTAILGNPMVRAH GKKVLT SF GDAVKNLDNINKNTFAQLSELHCDKLVDP	100	
1A4FB	51	PTAILGNPMVRAH GKKVLT SF GDAVKNLDNINKNTFAQLSELHCDKLVDP	100	
1FAWB	101	ENFRLLGDILITVLA AHFAKKEFTPE CQAAMQKLV RVVAHA LARKYH	146	
1A4FB	101	ENFRLLGDILITVLA AHFAKKEFTPE CQAAMQKLV RVVAHA LARKYH	146	

序列比对的结果显示，斑头雁和灰雁的血红蛋白仅有4个氨基酸的差异。

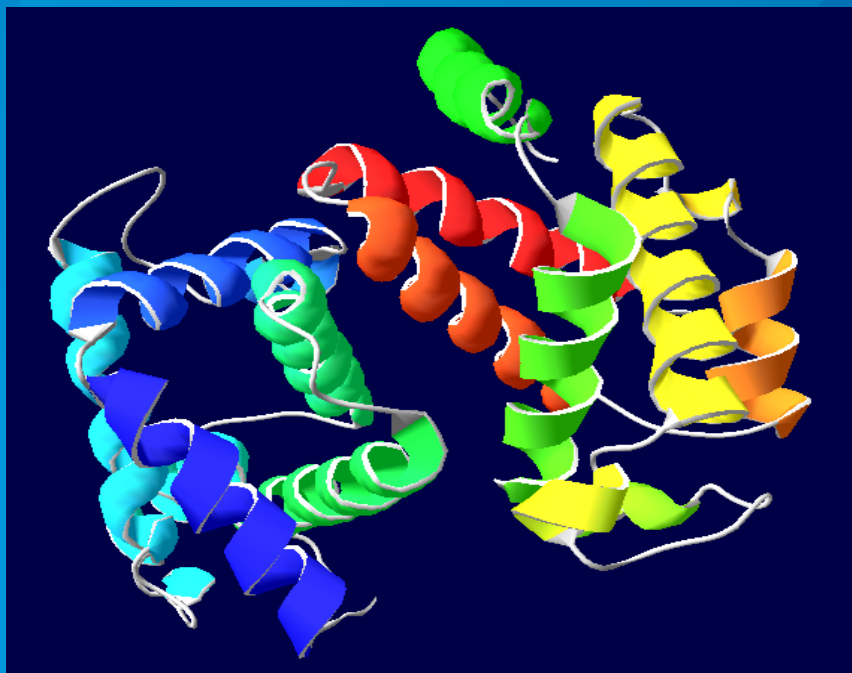
2.PDB文件的获取

- 首先下载并安装SPDB-Viewer软件.
- 进入PDB主页后，搜索1a4f和1faw分别得到斑头雁血和灰雁的血红蛋白3D结构
- 点击download图标，下载1a4f.pdb和1faw.pdb文件至本地计算机。

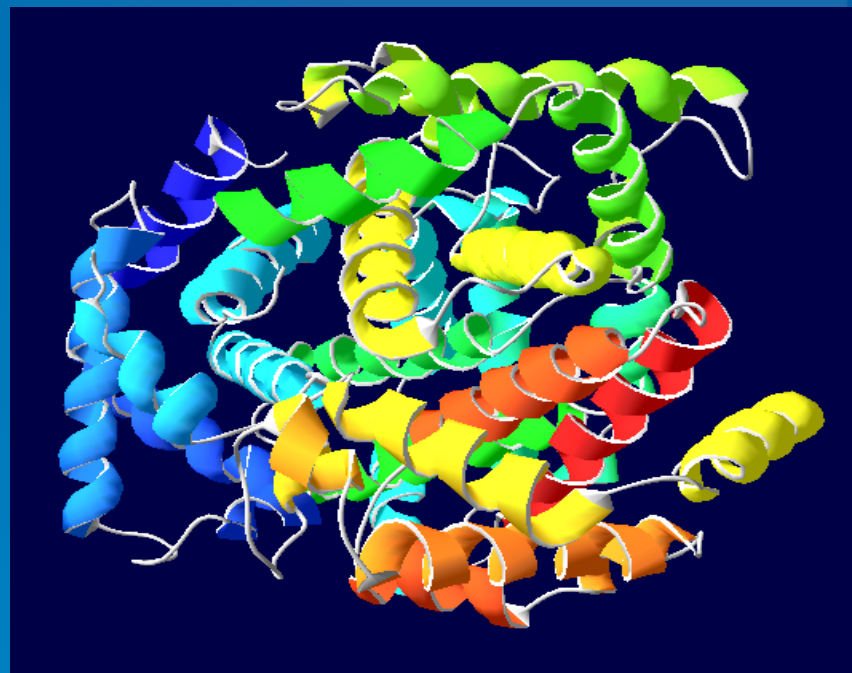
3、载入文件

- 载入蛋白质文件：在菜单file→open PDB file 选择目标pdb文件，即可载入。
- 载入蛋白质后，我们首先打开windows下的 control panel和layers infos。
- 去除主链和侧链，用ribon形式分别显示,并选择 color-Secondary Structure Succession对其染色。

4.3D结构的显示



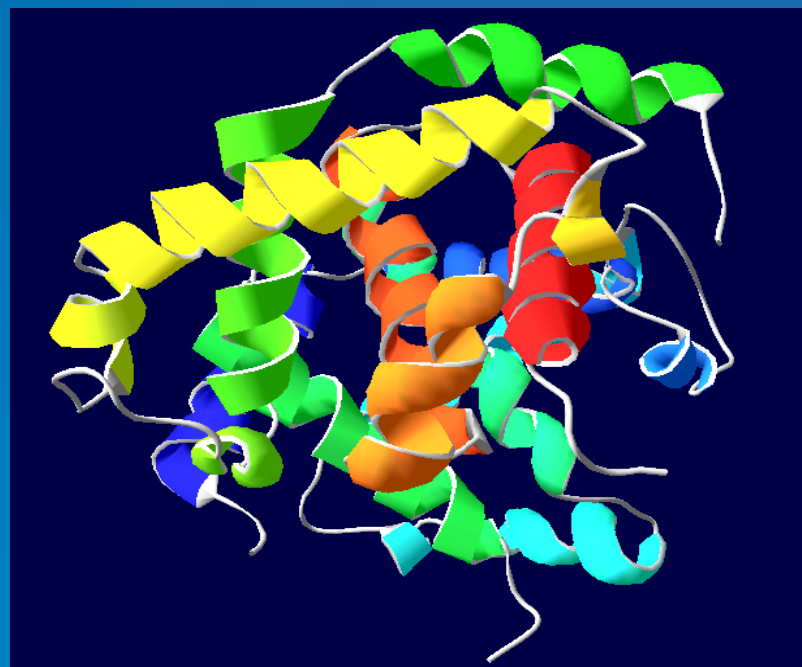
斑头雁血红蛋白的氧合形式1a4f



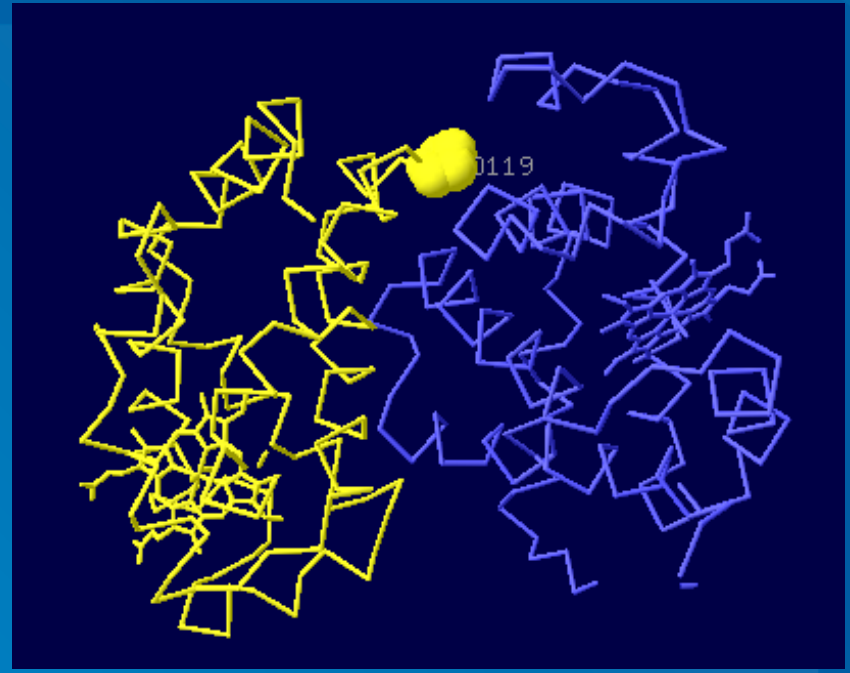
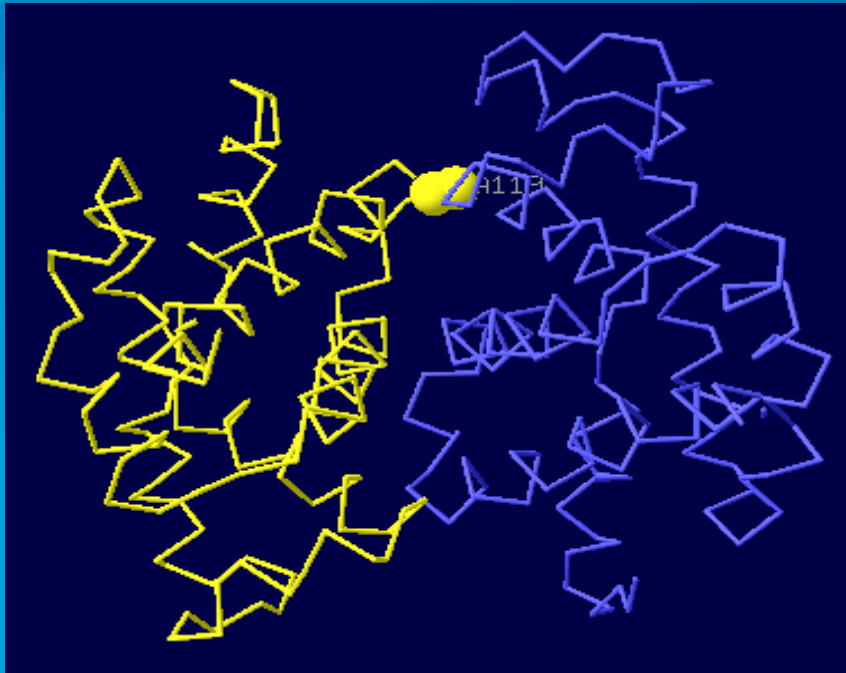
灰雁血红蛋白的氧合形式
1faw

5、选定所需序列

- 通过color-chain功能我们很容易看出1a4f由 α 和 β 两个亚基组成，而1faw则含有4个亚基。
- 选取1faw的 α 和 β 亚基，保存为1faw-ab，便于后面的操作。



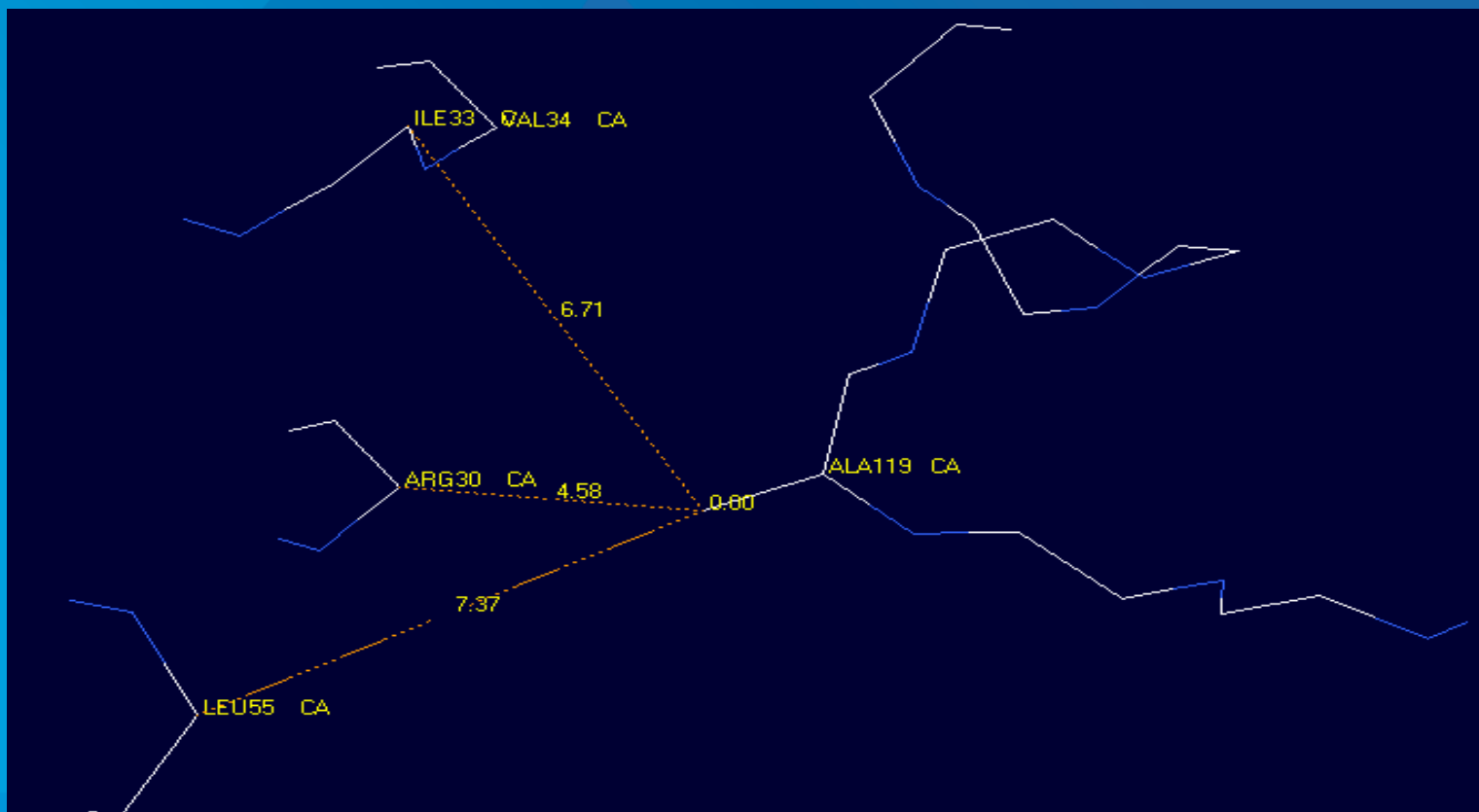
6.用SPDBV进行结构分析



由上图可以看出，斑头雁 α 链119位的Ala和灰雁 α 链119位的Pro都位于血红蛋白 α 链和 β 链的结合处。

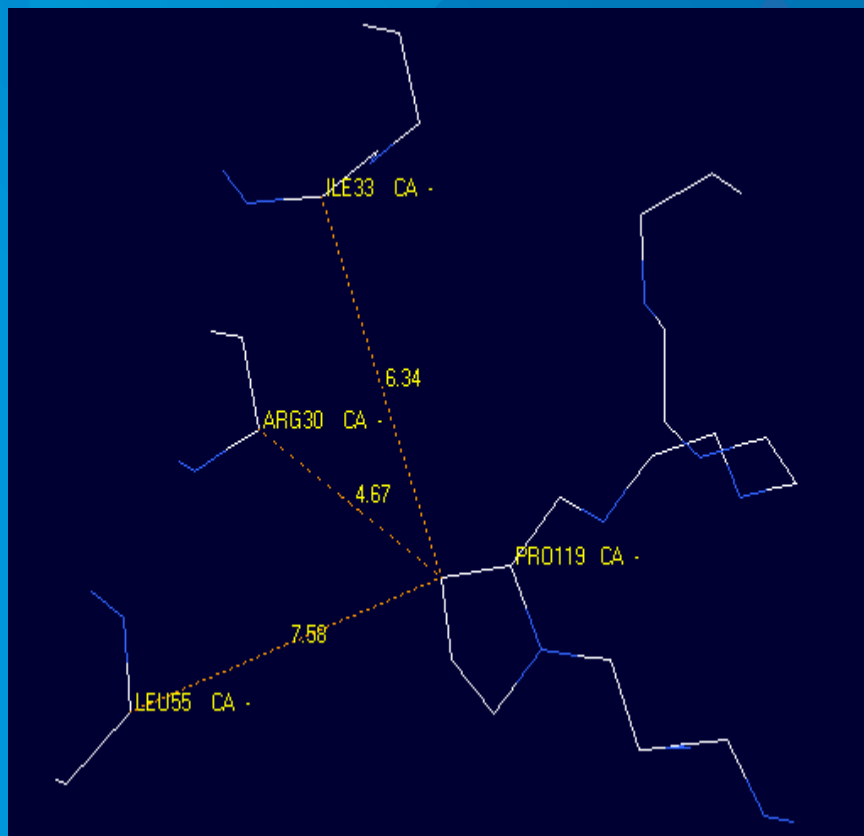
7、 α 链和 β 链之间距离的测定-斑头雁

- 显示斑头雁 α 链119位Ala周围6埃的氨基酸

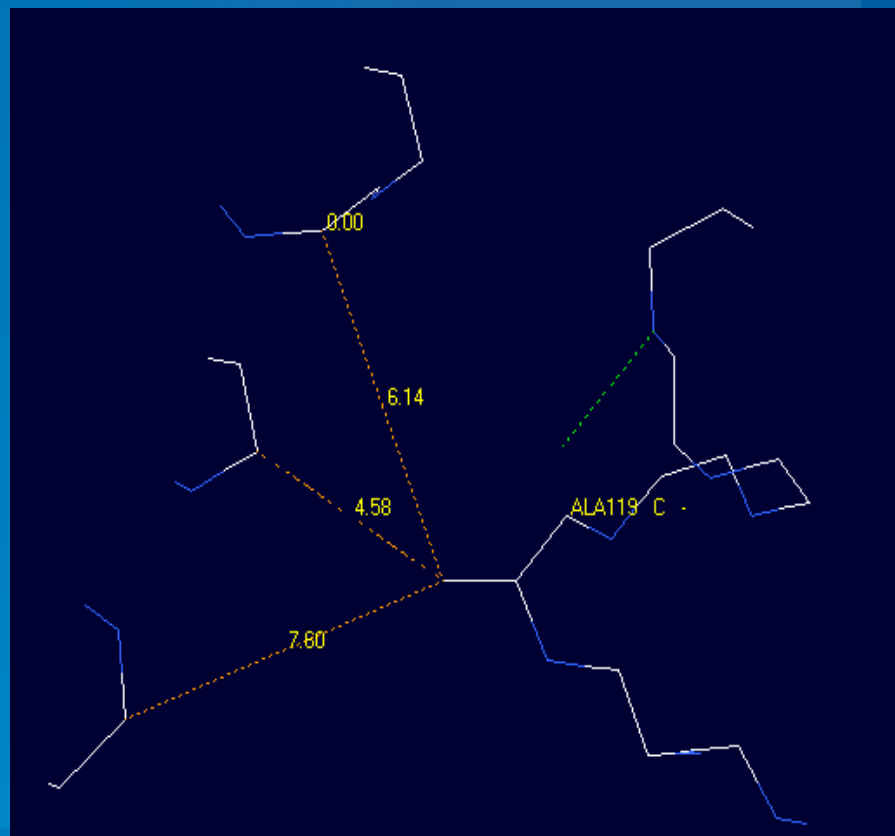


8. α 链和 β 链之间距离的测定-灰雁

未突变119位脯氨酸



突变119位丙氨酸



9、结论

- 和灰雁比较，斑头雁血红蛋白 α 链119位Ala使亚基间距离减小，导致分子疏松、亚基间牵制作用弱，从而更易结合氧气分子。推测其氧合曲线趋于平缓，这符合斑头雁在高原缺氧情况下结合氧气的需要。
- 以上现象说明 α -119 在血红蛋白氧亲和力上起到了非常重要的作用。因此可以通过改变这一残基来提高血红蛋白的氧亲和力。

Last, but not the least...

Thank you!