

---

## “实用生物信息技术”课程小组讨论总结报告

组：G6 次：3 组长：冯睿桥 执笔：冯睿桥、凌子涵

### 1. 时间

2026年3月30日、4月1日、4月12日

### 2. 方式

线上线下交流与讨论

### 3. 主题

序列比对及 blast 课后近期学习内容交流与互助

### 4. 内容

#### A.问题讨论

#### 1. 序列分析基本概念

(1) 举例说明序列相似性和同源性之间的差别和联系。

序列相似性 (similarity) 是可量化的指标，表示两个序列在比对后相似程度，用百分比、分值等表示；同源性 (homology) 是进化概念，表示两个序列是否来源于共同祖先，不可量化。例如人类血红蛋白  $\alpha$  链和小鼠血红蛋白  $\alpha$  链有约 85% 的相似性，因此可以推断它们是同源的。

(2) 举例说明全局比对和局部比对的适用范围。

全局比对是对整个序列进行比对（端到端），适用于长度相近、整体相似的序列；局部比对是找到最相似的局部片段，适用于长度差异大或只共享某个结构域。

(3) 举例说明 BLOSUM 计分矩阵和 PAM 计分矩阵的构建方法和特点。

打分是怎么来的。PAM 矩阵的出发点是“模拟进化过程”。先选取一批非常相似的蛋白序列（通常相似性  $>85\%$ ），因为这类序列之间的差异基本可以看作是刚刚发生的点突变。统计这些序列之间氨基酸是如何发生替换的，然后得到一个“单位进化距离”的替换概率矩阵 (PAM1)。然后通过数学外推 (矩阵幂运算) 推导出更远进化距离的情况，比如 PAM250。这就意味着 PAM 矩阵本质上是一个从近到远推演出来的进化模型。它的特点是数值越大表示进化距离越远，因此 PAM250 用于远缘序列，而 PAM30 更适合近缘序列。

相比之下，BLOSUM 的思路完全不同，它不去假设进化路径，而是直接从数据中“看见”替换模式。从数据库中提取出一批保守结构域 (blocks)，这些区域在不同蛋白中功能重要且较稳定，然后对这些序列进行聚类，比如把相似性高于 62% 的序列合并 (这就是 BLOSUM62 中的“62”)。在去除冗余之后，直接统计这些保守区域中氨基酸替换出现的频率，再转化为对数得分。也就是说 BLOSUM 是一个经验统计矩阵，它反映的是“自然界中真实发生过的替换”。它的数值含义和 PAM 正好相反：BLOSUM 数字越大 (如 80)，表示序列越接近；数字越小 (如 45)，

表示适用于更远的序列。

(4) 举例说明空位罚分的意义和用法。

在序列比对中，引入空位 (gap) 表示插入/缺失 (indel)。但如果不加限制，会出现“乱插空位”的问题。空位罚分用于惩罚不合理的插入/缺失，提高比对生物学合理性。

(5) 了解动态规划双序列全局比对和局部比对算法的基本思路。

核心是把复杂问题拆分为子问题，用矩阵逐步求解最优路径。无论是全局比对还是局部比对，本质上都在解决同一个问题，即在所有可能的对齐方式中找到一个总得分最高的方案。但因为可能的对齐方式是指数级的，不能暴力枚举，所以采用动态规划。全局比对和局部比对的差别，其实只是在比对允许不允许中途放弃。

比对还是局部比对，都是采用动态规划算法。动态规划算法是指在给定计分矩阵和空位罚分的条件下，通过插入适当空位，使比对结果的总分值最高，即找到最优解。无论是 Needleman-Wunsch 算法或者是 Smith-Waterman 算法，都采用计算机领域中常用的动态规划 (Dynamic Programming) 算法。动态规划算法的核心思想，是把一个复杂问题分解为若干子问题，并通过寻找子问题的解，最终找到初始复杂问题的解。

## 2. Blast 数据库搜索

基于序列相似性的数据库搜索，顾名思义，就是以某个核酸或蛋白质序列作为查询序列 (Query sequence)，与核酸或蛋白质序列数据库中的序列进行比较，找出数据库中与查询序列相似性较高的序列，即目标序列 (Subject sequence)。

### (1) Blast 搜索的主要程序

BLASTP: 以蛋白质序列为查询序列，搜索数据库中的相似蛋白质。

BLASTN: 即用核酸序列作为查询序列，搜索核酸序列数据库；搜索步骤和 BLASTP 基本相同，只是种子串长度、计分矩阵和空位罚分等参数不同。

BLASTX: 用于以核酸序列为查询序列，搜索蛋白质序列数据库。查询序列为蛋白质编码序列

TBLASTN: 与 BLASTX 相反，TBLASTN 以蛋白质为查询序列，搜索核酸序列数据库，如 NCBI 核酸参考序列数据库 (Reference RNA, RefSeq\_RNA)

### (2) Blast 主要用途

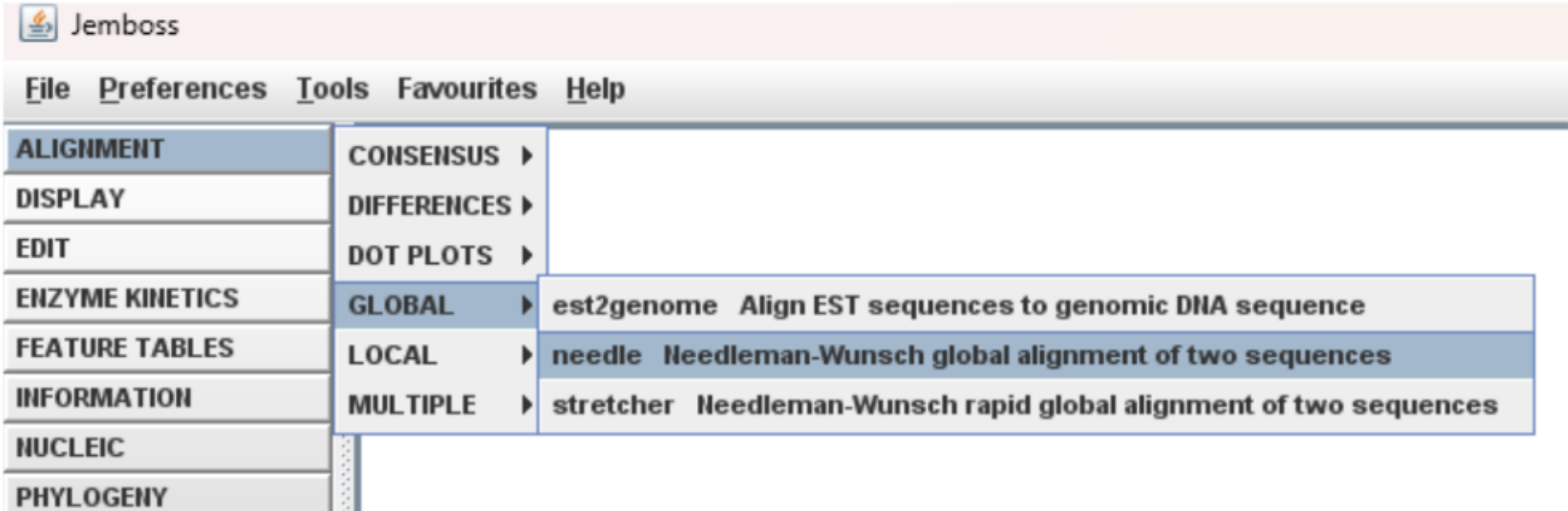
BLAST 数据库搜索的实际应用包括许多方面。例如，对于新测定的或功能未知的核酸序列，通过数据库搜索，可获得数据库中已经收录的目标序列。根据相似性程度，推断该未知序列与目标序列是否为同源序列，并进一步根据目标序列的注释信息，推测该未知序列可能属于哪个基因家族，具有哪些生物学功能。

## B. 课堂内容复习与互助

### 1. Uniprot 蛋白质数据库检索条目

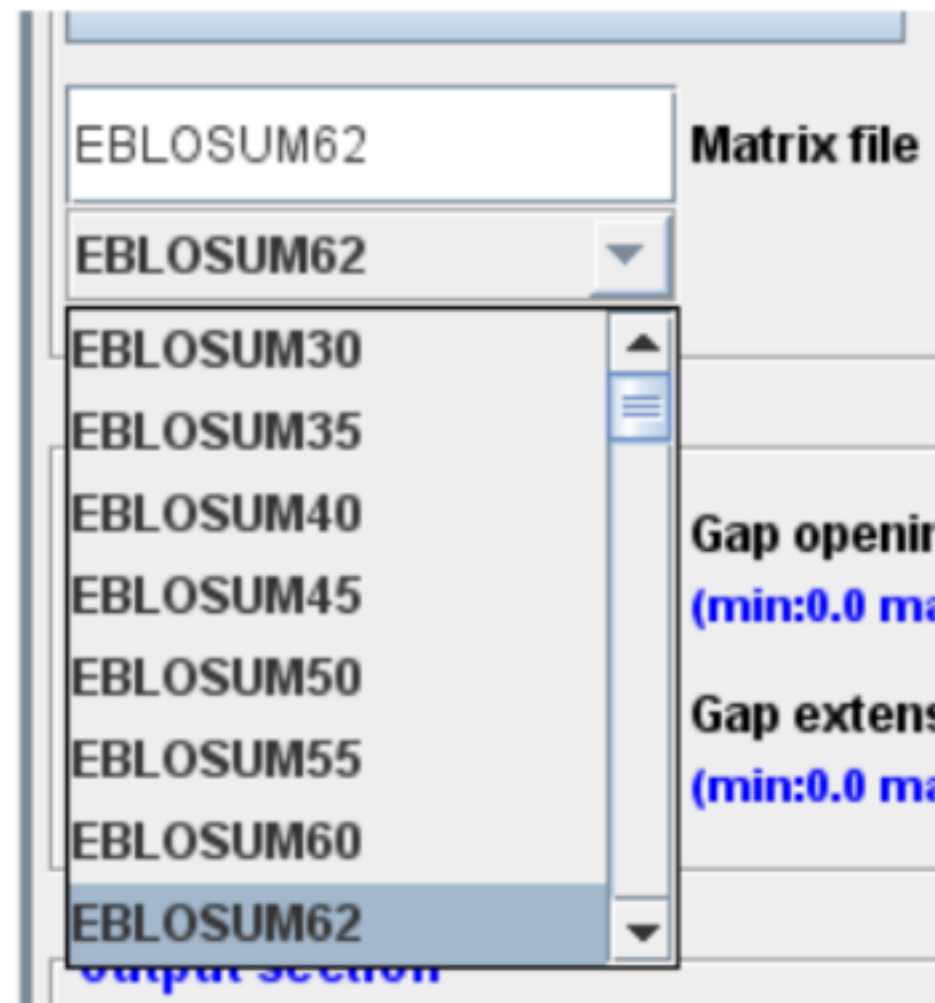
根据已知蛋白条目信息进行搜索，得到目标蛋白质序列，以便于后续进行序列比对

### 2. emboss 全局序列比对软件的使用（以人源 HBA 蛋白与鼠源 HBA 蛋白为例）

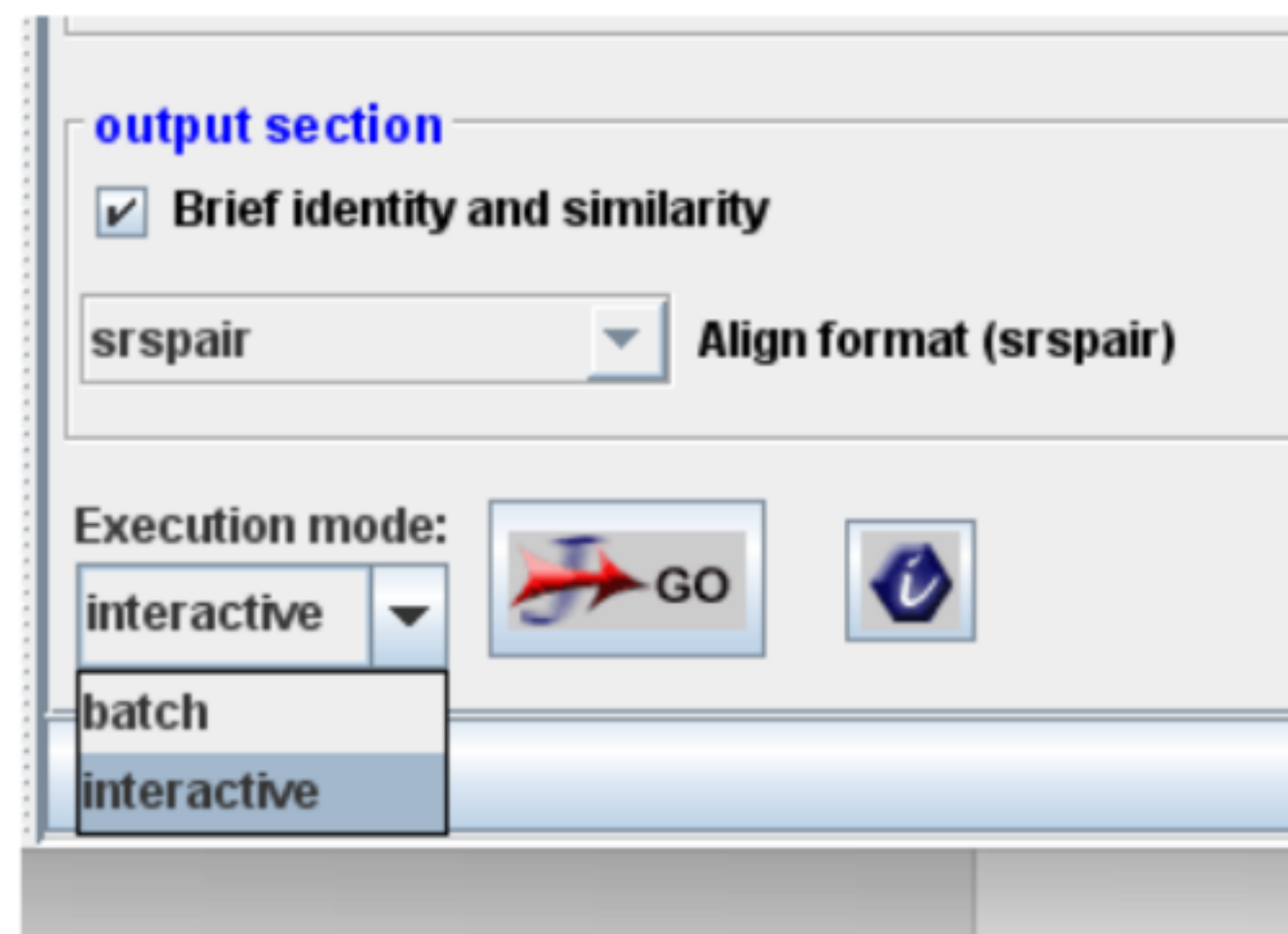


打开 JEMBOSS, ALIGNMENT→GLOBAL→needle (全局比对)

点击，都选择 paste，然后将序列粘贴到框里，这里粘贴的是人源 HBA 蛋白与鼠源 HBA 蛋白序列



选择默认计分矩阵 EBLOSUM62



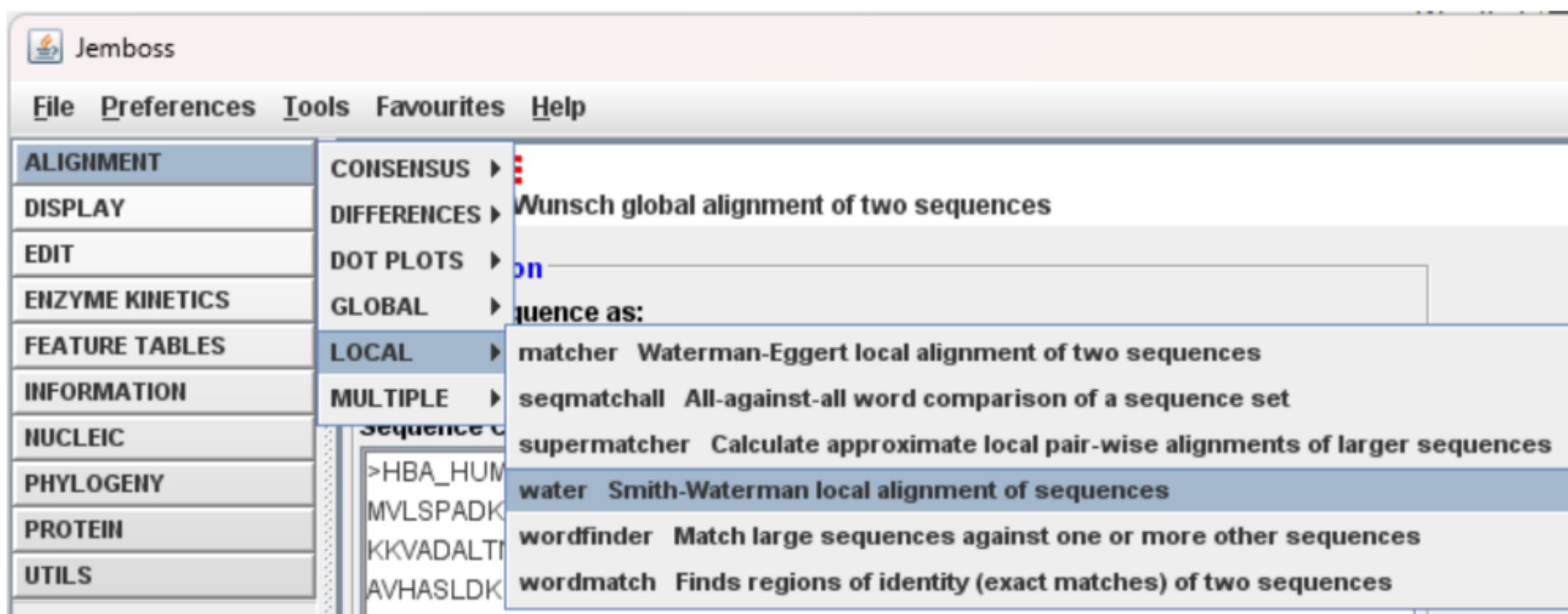
Execution mode 选择 interactive, 然后 GO!

```
..
# Aligned_sequences: 2
# 1: HBA_HUMAN
# 2: HBA_MOUSE
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 142
# Identity:    122/142 (85.9%)
# Similarity:  131/142 (92.3%)
# Gaps:       0/142 ( 0.0%)
# Score: 648.0
#
```

显示比对结果并分析，一致性 85.9%，通常蛋白序列一致性大于 50%便可认定为同源蛋白，这两个跨物种的蛋白结构和功能高度一致，几乎没有发生性质改变；相似性 92.3%，进一步证明高度同源性；空位 0%，这说明人源和鼠源的 HBA 蛋白长度完全一致（均为 142 个氨基酸），且在全长范围内完美对位，没有缺失或插入突变；比对总得分 648，对于 142 长度的序列

属高分，说明比对算法非常匹配，同源性极强。

讨论与拓展: water 是局部比对工具, 与 needle 全局比对对应, 局部比对只比对两条序列中同源性最高的区域, 适合序列长度差异大、全局比对空位多的序列, 比如 CEAM5/6 这类家族蛋白, 具体选项如图, ALIGNMENT→LOCAL→waste



### 3、抗菌肽和蜘蛛毒素序列比对

The image shows the wwPDB website for entry 1DKC. The entry is titled '1DKC | pdb\_0001dkc' and is described as 'SOLUTION STRUCTURE OF PAFP-S, AN ANTIFUNGAL PEPTIDE FROM PHYTOLACCA AMERICANA'. The classification is 'ANTIFUNGAL PROTEIN' and the organism is 'Phytolacca americana'. The deposition date is 1999-12-07 and the release date is 2000-12-13. The deposition author(s) are Wang, D.C., Gao, G.H., Shao, F., Dai, J.X., Wang, J.F. The experimental data snapshot shows the method as SOLUTION NMR, with 30 conformers calculated and 10 submitted. The selection criteria are structures with the lowest energy. The wwPDB Validation metrics are shown in a table:

Metric	Percentile Ranks	Value
Clashscore		28
Ramachandran outliers		4.2%
Sidechain outliers		25.0%

The validation metrics are visualized with horizontal bars and a legend indicating the percentile relative to all structures and all NMR structures.

1QK6 | pdb\_00001qk6

Solution structure of huwentoxin-I by NMR  
 PDB DOI: <https://doi.org/10.2210/pdb1QK6/pdb>

Classification: TOXIN  
 Organism(s): *Cyriopagopus schmidtii*  
 Mutation(s): No  
 Membrane Protein: Yes **OPM**

Deposited: 1999-07-10 Released: 1999-08-20  
 Deposition Author(s): Qu, Y., Liang, S., Ding, J., Liu, X., Zhang, R., Gu, X.

Experimental Data Snapshot  
 Method: SOLUTION NMR  
 Conformers Calculated: 50  
 Conformers Submitted: 10  
 Selection Criteria: STEREOCHEMICAL ENERGY AND RESULT OF PROCHECK

wwPDB Validation **3D Report** Full Report

Metric	Percentile Ranks	Value
Clashscore	8	8
Ramachandran outliers	0	0
Sidechain outliers	21.7%	21.7%

Worse | Percentile relative to all structures | Better

在蛋白质结构数据库 PDB 中找到这两个蛋白质序列，打开 FASTA 文件复制

← → ↻ 🏠 **不安全** <https://bioinformatics.nl/emboss-explorer/>

[ [sort alphabetically](#) ]

**ALIGNMENT**  
[extractalign](#)

**ALIGNMENT**  
**CONSENSUS**  
[cons](#)  
[consambig](#)  
[megamerger](#)  
[merger](#)

**ALIGNMENT**  
**DIFFERENCES**  
[diffseq](#)

**ALIGNMENT**  
**DOT PLOTS**  
[dotmatcher](#)  
[dotpath](#)  
[dottup](#)  
[polydot](#)

**ALIGNMENT**  
**GLOBAL**  
[est2genome](#)  
[needle](#)  
[needleall](#)  
[stretcher](#)

**ALIGNMENT**  
**LOCAL**  
[domainalign](#)

EMBOSS explorer

## needle

Needleman-Wunsch global alignment of two sequences ([read the manual](#))

Unshaded fields are optional and can safely be ignored. ([hide optional fields](#))

Input section

Select an input sequence. Use one of the following three fields:

- To access a sequence from a database, enter the USA here:
- To upload a sequence from your local computer, select it here:

```
>IDKC_1|Chain A|ANTIFUNGAL PEPTIDE|Phytolacca americana (3527)
AGCIKNGGRCNASAGPPYCCSSYCFQIAGQSYGVCKNR
```

- To enter the sequence data manually, type here:

Select an input sequence. Use one of the following three fields:

- To access a sequence from a database, enter the USA here:
- To upload a sequence from your local computer, select it here:

```
>1QK6_1|Chain A|HUWENTOXIN-I|SELENOCOSMIA HUWENA (29017)
ACKGVFDACTPGKNECCPNRVCSDKHKWCKWKL
```

- To enter the sequence data manually, type here:

打开 EMBOSS explorer 分析平台，左侧导航菜单中找到 ALIGNMENT GLOBAL，选择 needle 全局比对，将两个蛋白质序列粘贴到框里

**OUTPUT FILE** [outfile](#)

```
#####
# Program: needle
# Rundate: Tue 7 Apr 2026 10:36:14
# Commandline: needle
#   -auto
#   -asequence /var/lib/emboss-explorer/output/041364/.asequence
#   -bsequence /var/lib/emboss-explorer/output/041364/.bsequence
#   -noendweight
#   -brief
#   -outfile outfile
#   -aformat3 srspair
# Align_format: srspair
# Report_file: outfile
#####

#-----
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_002
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 44
# Identity:      6/44 (13.6%)
# Similarity:    12/44 (27.3%)
# Gaps:          17/44 (38.6%)
# Score: 17.0
#
#
#-----

EMBOSS_001      1 AGCIKNGGRCN-----ASAGPPYCCSSYCFQIAGQSYGVCKNR-   38
                  .|.      .:. |... ||.:  ::...:.. ||.:
EMBOSS_002      1 -----ACKGVFDCTPGKNECCPN---RVCSDKHKWCKWKL   33

#-----
#-----
```

## Run needle 获得比对结果

Matrix file. Use one of the following two fields:

1. To access a standard EMBOSS data file, enter the name here:

(default is EBLOSUM62 for protein, EDNAFULL for nucleic)

2. To upload a data file from your local computer, select it here:

PAM 系列计分矩阵中，PAM250 适用于相似性较低的序列比对，因此尝试改变计分矩阵，选择框中输入计分矩阵 PAM250，并适当降低空位罚分值，优化输出结果

## OUTPUT FILE [outfile](#)

```
#####
# Program: needle
# Rundate: Tue 7 Apr 2026 11:06:56
# Commandline: needle
#   -auto
#   -asequence /var/lib/emboss-explorer/output/576234/.asequence
#   -bsequence /var/lib/emboss-explorer/output/576234/.bsequence
#   -datafile EPAM250
#   -gapopen 5
#   -gapextend 0.5
#   -noendweight
#   -endopen 10
#   -endextend 0.5
#   -brief
#   -outfile outfile
#   -aformat3 srspair
# Align_format: srspair
# Report_file: outfile
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_002
# Matrix: EPAM250
# Gap_penalty: 5.0
# Extend_penalty: 0.5
#
# Length: 42
# Identity:      9/42 (21.4%)
# Similarity:   18/42 (42.9%)
# Gaps:         13/42 (31.0%)
# Score: 51.5
#
#=====

EMBOSS_001      1 AGCIKNG--GRCNASAGPPYCCSS-YCFQIAGQSYGVCKNR-      38
                  :|  :|  ::|  ::|...||::|  ...:..  ||.:
EMBOSS_002      1 -AC--KGVFDAC--TPGKNECCPNRVC--SDKHKW--CKWKL      33
```

可以看出调整过后的结果更加理想

## 4、Blast 个别练习互助与讨论

### (1) BLASTP 练习

1. 以人血红蛋白 alpha 亚基 HBA\_HUMAN 为搜索序列。可以在 Uniprot 搜索得到序列

## 2. 利用中国国家生物信息中心 CNCB 蛋白质数据库搜索程序 BlastP

计算分析 查看所有计算分析资源 >

**序列比对(BLAST)**

基于BLAST算法比对核酸或蛋白序列，鉴定序列间的相似性和同源性，帮助分析序列功能和进化关系

**云平台(BiT)**

基于云的生物信息学工具包，集成常用生物信息算法、软件与图形绘制方法，提供免费的数据分析云服务

**基因组组装与注释**

将测序数据组装成连续序列及完整基因组，并对基因组中基因的结构与功能等特征进行注释和分析

**单细胞与空间组学**

单细胞分类、细胞的组织空间分布及相互作用分析，实现高分辨率的细胞异质性和空间关系表征

**系统发生与分子进化**

通过比较核酸、蛋白质等分子信息，分析物种演化与遗传变异，揭示进化历程、遗传多样性与适应性

**健康与疾病检测**

多种定量定性检测分析，以评估和检测个体健康状况、发现疾病严重程度、监测治疗进展等情况

<http://www.cncb.ac.cn/tools>

**Nucleotide BLASTN**  
nucleotide → nucleotide  
e.g., SARS-CoV-2 (C\_AA019067.1)

**Protein BLASTP**  
protein → protein  
e.g., TP53 (NP\_001394193.1)

**BLASTX**  
translated nucleotide → protein  
e.g., HBB (NM\_000518.5)

**TBLASTN**  
protein → translated nucleotide  
e.g., SCN2A (NP\_001035232.1)

输入搜索序列

请输入序列: [示例](#) [清空](#)

```
>HBA_HUMAN - P69905, Hemoglobin subunit alpha, HBA1; J Luo, 2016-08-21
MVLSPADKTNVKAAWGKVGAGHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNVAHAVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLTKFLASVSTVLTSKYR
```

或上传文件: [选择文件](#)

任务名字:

选择目标:  与数据库比对  与输入序列比对

查询子区间:

选择搜索数据库

标准数据库(nr等)

数据库:

物种:  (包含)

3. 搜索 SwissProt 数据库，找出灵长目动物 Primates 中 alpha 血红蛋白
4. 配置算法参数

配置算法参数 恢复默认搜索参数

<p><b>常规参数</b></p> <p>* 输出格式: <input type="text" value="Tabular"/></p> <p>* 最大目标序列: <input type="text" value="100"/></p> <p>* 期望阈值: <input type="text" value="0.050"/></p> <p>* 字长: <input type="text" value="3"/></p>	<p><b>计分参数</b></p> <p>* 计分矩阵: <input type="text" value="BLOSUM62"/></p> <p>* 空位罚分: <input type="text" value="Existence: 11 Extension: 2"/></p> <p>* 矩阵校正: <input type="text" value="Conditional compositional scor..."/></p>	<p><b>过滤和屏蔽参数</b></p> <p><input checked="" type="checkbox"/> 软屏蔽</p> <p><input checked="" type="checkbox"/> 屏蔽小写字母</p>
--	--	--

比对数据库: 
 比对程序: 
 任务完成时邮件通知您
   
 在新窗口中显示结果

## 搜索结果

以下表格里展示的是比对显著的结果序列 下载 选择列 显示条目

100 条序列已选中 距离树 多序列比对

<input checked="" type="checkbox"/>	序列描述	物种命名	最高得分	总得分	覆盖率	期望值	相似度	序列长	序列号
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Homo sapiens</a>	317	317	100.00%	1.96637e-110	100.00%	142	<a href="#">P69905.2</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Gorilla gorilla gorilla</a>	312	312	99.00%	1.59706e-108	99.29%	141	<a href="#">P01923.1</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha-1; AltName:...</a>	<a href="#">Hylobates lar</a>	311	311	100.00%	4.00779e-108	98.59%	142	<a href="#">Q9TS35.2</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Pongo pygmaeus</a>	311	311	100.00%	6.29289e-108	97.89%	142	<a href="#">P06635.2</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Semnopithecus entellus</a>	308	308	99.00%	6.50533e-107	97.87%	141	<a href="#">P01924.1</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Macaca fuscata fuscata</a>	308	308	100.00%	7.47879e-107	97.18%	142	<a href="#">P63107.2</a>
<input checked="" type="checkbox"/>	<a href="#">RecName: Full=Hemoglobin subunit alpha; AltName: F...</a>	<a href="#">Ateles geoffroyi</a>	306	306	100.00%	2.966e-106	96.48%	142	<a href="#">P67817.2</a>

可以通过调整物种名，相似度等对结果进行筛选

[< 编辑搜索](#) | [保存搜索策略](#) | [搜索策略概览](#)

任务名称	HBA_HUMAN - P69905, Hemoglobin
请求 ID	DUC8R5GCMPB <a href="#">结果下载</a> ▼
程序	BLASTP <a href="#">引用</a> ▼
数据库	NCBI_SwissProt
查询 ID	Query_1
描述	HBA_HUMAN__P69905__Hemoglobin_subunit_a...
分子类型	Protein
查询长度	142
其他报告	<a href="#">距离树</a> <a href="#">多序列比对</a> <a href="#">?</a>

**筛选结果**

物种

排除 [?](#)

---

相似度 [?](#)      期望值 [?](#)      覆盖率 [?](#)

至        至        至

[过滤](#)   [重置](#)

## 2. TBLASTN 练习

以人的血红蛋白 alpha 亚基 HBA\_HUMAN 为搜索序列，用 NCBI tBlastN 搜索参考序列数据库 RefSeq 中人的珠蛋白家族 mRNA 序列。



### Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

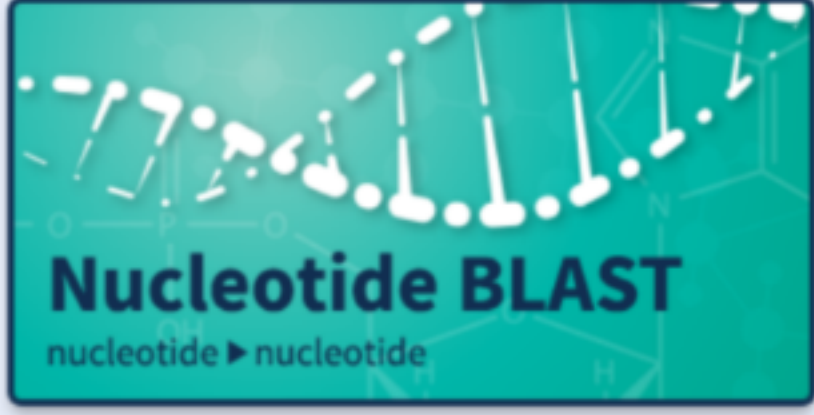
[SNP](#)

[Gene](#)

[Protein](#)

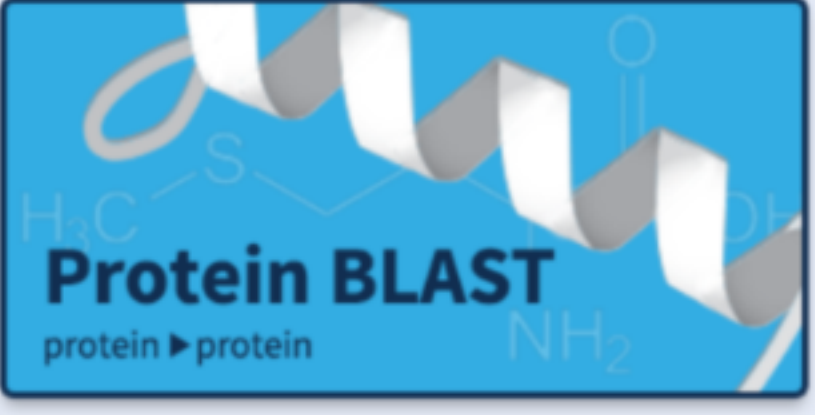
[PubChem](#)

**Web BLAST**



**Nucleotide BLAST**  
nucleotide ▶ nucleotide

**blastx**  
translated nucleotide ▶ protein



**Protein BLAST**  
protein ▶ protein

**tblastn**  
protein ▶ translated nucleotide

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

GSAQVKGHG  
KKVADALTNAVAHVDDMPNALSALSSDLHAHKLKRVDPVNFKLLSHCLLVTLAAH  
LPAEFTP  
AVHASLDKFLASVSTVLTSKYR

From   
To

Or, upload file   [?](#)

Job Title   
Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Database  [?](#)

Organism Optional   exclude   
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to Optional  Sequences from type material

Entrez<sup>®</sup> Query Optional     
Enter an Entrez query to limit search [?](#)

参数设置：期望阈值 E 为 1，字长 Word size 为 2，计分矩阵 Matrix 为 PAM250，空位罚分 Gap cost 为起始空位罚分 13，延伸空位罚分 2，其余参数默认

[Restore default search parameters](#)

**General Parameters**

Max target sequences  [?](#)  
Select the maximum number of aligned sequences to display [?](#)

Expect threshold  [?](#)

Word size  [?](#)

Max matches in a query range  [?](#)

**Scoring Parameters**

Matrix  [?](#)

Gap Costs  [?](#)

Compositional adjustments  [?](#)

## 结果

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit alpha 2 (HBA2), mRNA</a>	<a href="#">Homo sapiens</a>	216	216	100%	8e-78	100.00%	576	<a href="#">NM_000517.6</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit alpha 1 (HBA1), mRNA</a>	<a href="#">Homo sapiens</a>	216	216	100%	8e-78	100.00%	577	<a href="#">NM_000558.5</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit theta 1 (HBQ1), mRNA</a>	<a href="#">Homo sapiens</a>	155	155	100%	6e-53	61.97%	528	<a href="#">NM_005331.5</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit zeta (HBZ), mRNA</a>	<a href="#">Homo sapiens</a>	151	151	100%	1e-50	59.86%	798	<a href="#">NM_005332.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit mu (HBM), mRNA</a>	<a href="#">Homo sapiens</a>	127	127	99%	2e-41	45.39%	502	<a href="#">NM_001003938.4</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit gamma 2 (HBG2), mRNA</a>	<a href="#">Homo sapiens</a>	103	103	99%	1e-31	40.41%	586	<a href="#">NM_000184.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit beta (HBB), mRNA</a>	<a href="#">Homo sapiens</a>	103	103	99%	3e-31	43.15%	628	<a href="#">NM_000518.5</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit gamma 1 (HBG1), mRNA</a>	<a href="#">Homo sapiens</a>	102	102	99%	4e-31	40.41%	587	<a href="#">NM_000559.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit delta (HBD), mRNA</a>	<a href="#">Homo sapiens</a>	101	101	99%	2e-30	43.15%	620	<a href="#">NM_000519.4</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens hemoglobin subunit epsilon 1 (HBE1), mRNA</a>	<a href="#">Homo sapiens</a>	98.4	98.4	99%	3e-29	36.99%	623	<a href="#">NM_005330.4</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens cytoglobin (CYGB), mRNA</a>	<a href="#">Homo sapiens</a>	69.3	69.3	96%	1e-15	27.89%	1962	<a href="#">NM_134268.5</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens myoglobin (MB), transcript variant 1, mRNA</a>	<a href="#">Homo sapiens</a>	57.2	57.2	100%	8e-12	27.52%	1070	<a href="#">NM_005368.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens neuroglobin (NGB), mRNA</a>	<a href="#">Homo sapiens</a>	32.0	32.0	96%	0.043	21.38%	1778	<a href="#">NM_021257.4</a>

NCBI blast 利用已知序列片段进行比对查找相似序列，在实际应用中可以通过多序列比对进行物种分析或建树，可结合 MEGA 等软件进行后续工作。