

第13章 同源序列分析

第14章 蛋白质理化性质分析



生命健康信息学院

解增言

第13章 同源序列分析

- BLAST+
- HMMER

BLAST+

- BLAST (Basic Local Alignment Search Tool) 是一组在蛋白质或核酸数据中搜索相似序列的分析工具，它能迅速将查询序列 (Query) 与序列数据库进行比较找出相似序列。
- BLAST有多个版本，其中最常用的是NCBI BLAST+。

BLAST+

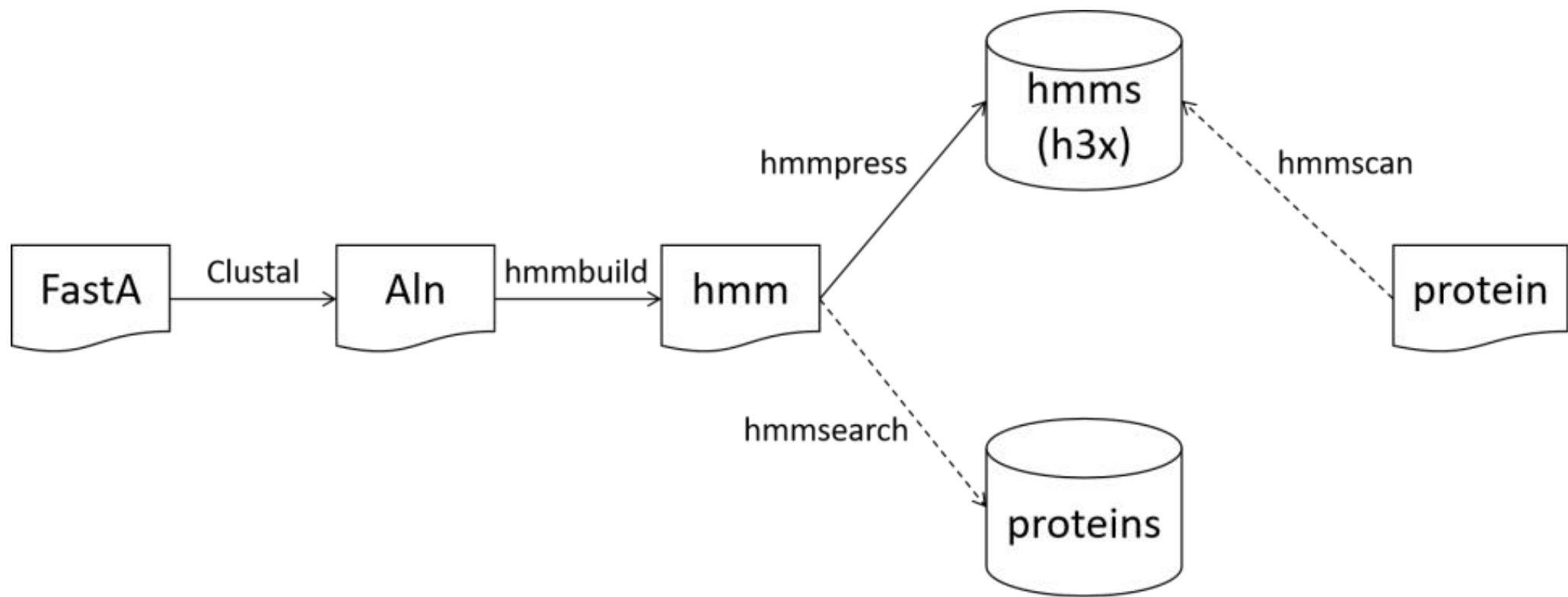
BLAST+程序	作用
makeblastdb	构建本地序列库
blastp	在蛋白质序列库中搜索蛋白质序列
blastn	在核酸序列库中搜索核酸序列
blastx	将给定的核酸序列按照六种阅读框翻译成蛋白质然后与蛋白质序列库中的序列进行比对
tblastn	将给定的蛋白质序列与核酸序列库中序列的六种阅读框进行比对
tblastx	将核酸序列和核酸序列库中的序列按不同的阅读框全部翻译成蛋白质序列，然后进行蛋白质序列比对
blastdbcmd	从序列库中取出指定序列

HMMER

- 蛋白质保守域（Protein Domain）是指蛋白质序列中的一段保守区，长短不一，有的蛋白质包含多个保守域。
- 蛋白质保守域分析最常用的工具是HMMER。HMMER是利用隐马尔科夫模型谱分析生物序列同源性的工具包。



HMMER分析流程



HMMER常用程序

HMMER程序	功能
hmmbuild	用多重比对序列构建HMM模型
hmmcompress	格式化HMM数据库，供hmmscan搜索使用
hmmsearch	使用HMM模型搜索序列库
hmmscan	使用序列搜索HMM模型库

第14章 蛋白质理化性质分析

- 蛋白质及其功能
- 蛋白质结构
- 蛋白质理化性质
- 蛋白质理化性质分析工具

蛋白质及其功能

➤ 蛋白质是一种由氨基酸串联而成的大分子。

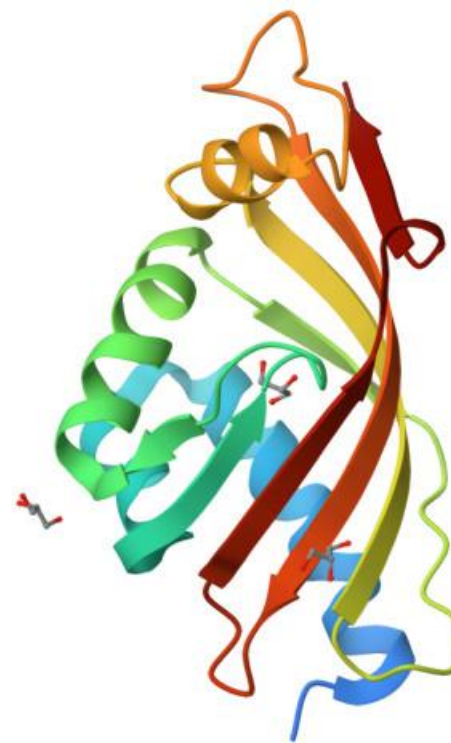
➤ 蛋白质功能：

- **催化：** 酶
- **结构与支持：** 胶原蛋白、角蛋白
- **运输与储存：** 血红蛋白、载脂蛋白、铁蛋白
- **运动：** 肌球蛋白
- **免疫防御：** 抗体
- **信号传递：** 激素、受体
- **调节：** 转录因子

蛋白质结构

- **一级结构**：氨基酸序列
- **二级结构**： α -螺旋、 β -折叠等
- **三级结构**：空间结构
- **四级结构**：亚基数量及排列

>9KW7_1|Chains A, B|MonBI|Streptomyces virginiae (1961)
MNHKVVHHHHHHIEGRHMNEFARAKRALEHSRRINAGDLDAIIDLY
APDAVLEDPVGLPPVTGHDALRAHYEPLLAHLREEAAEPVAGQD
ATHALIQISSVMDYLPVGPLYAERGWLKAPDAPGTARIHRTAMLVIR
MDASGLIRHLKSYWGTSDLTVLG



9KW7蛋白序列及结构卡通图

一种核转运因子2 (NTF2) 家族蛋白。该家族蛋白的标志是拥有一个保守的NTF2-like 结构域，其内部具有一个疏水性的空腔，用于结合RanGDP（一种小G蛋白），并将其从细胞质运入细胞核。

蛋白质理化性质

- 分子量
- 等电点
- 疏水性/亲水性
- 溶解性
- 稳定性
- 表面可及性
-

蛋白质理化性质分析工具

- infoseq: 序列信息统计
- pepstats: 蛋白质序列理化性质综合统计
- pepwindow: 疏水性图谱可视化

infoseq

```
$ infoseq -only -name -length /home/pub/seq/lde.fa
```

Display basic information about sequences

Name	Length
Musmu	550
Homsa	544
Anoga	558
Caeel	500
Drome	534
Danre	568
Xentr	534

pepstats

```
$ pepstats lde_hs.fa
```

```
Calculate statistics of protein properties
```

```
Pepstats program output file [homsa.pepstats]:
```

```
$ cat homsa.pepstats
```

```
PEPSTATS of Homsa from 1 to 544
```

```
Molecular weight = 61554.93      Residues = 544
```

```
Average Residue Weight  = 113.152      Charge   = -16.0
```

```
Isoelectric Point = 5.1441
```

```
A280 Molar Extinction Coefficients  = 67270 (reduced)    67895 (cystine bridges)
```

```
A280 Extinction Coefficients 1mg/ml = 1.093 (reduced)    1.103 (cystine bridges)
```

```
Probability of expression in inclusion bodies = 0.639
```

Residue	Number	Mole%	DayhoffStat
A = Ala	36	6.618	0.769
B = Asx	0	0.000	0.000
C = Cys	11	2.022	0.697
D = Asp	42	7.721	1.404
.....			
Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	143	26.287
Small	(A+B+C+D+G+N+P+S+T+V)	265	48.713
Aliphatic	(A+I+L+V)	142	26.103
Aromatic	(F+H+W+Y)	67	12.316
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	269	49.449
Polar	(D+E+H+K+N+Q+R+S+T+Z)	275	50.551
Charged	(B+D+E+H+K+R+Z)	167	30.699
Basic	(H+K+R)	81	14.890
Acidic	(B+D+E+Z)	86	15.809

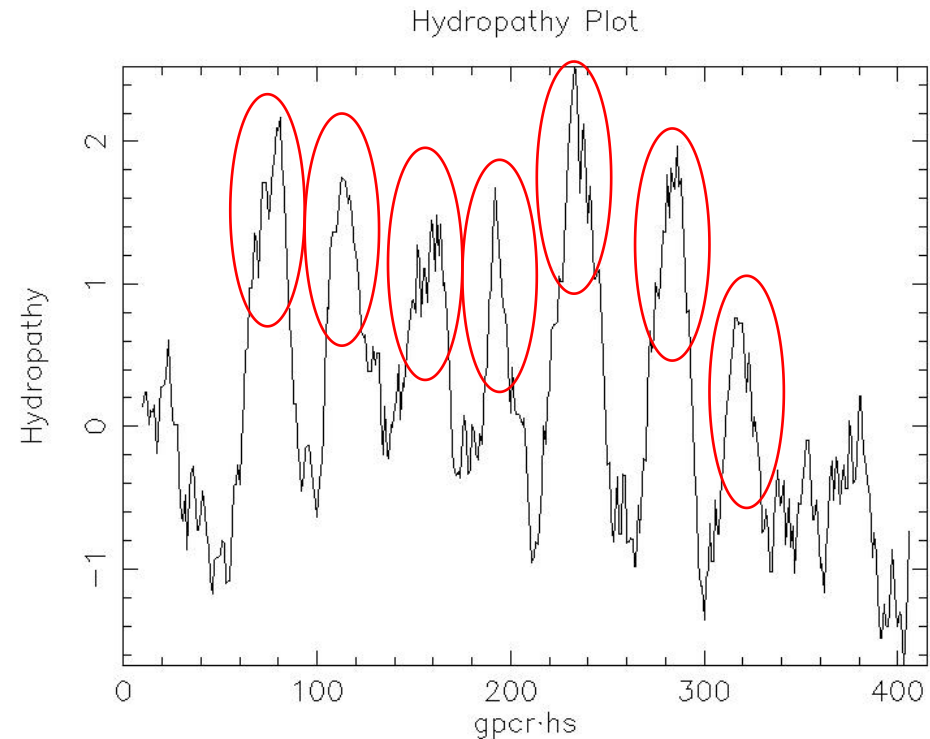
pepwindow

```
$ pepwindow GPCR_hs.fa
```

Draw a hydropathy plot for a protein sequence

Graph type [x11]: png

Created pepwindow.1.png



单元测试4