

ExPASy使用初探

—以斑头雁血红蛋白为例

组长：任娇

组员：申立平 孔寅飞 李东旭

2009.03.18

工作分工

- **a2任姣：** 斑头雁的物种介绍，ExPASy介绍，整理全组的工作，查漏补缺，编辑全文。
- **b1申立平：** 对斑头雁血红蛋白研究背景包括现阶段的进展做一个详尽的介绍。
- **a1孔寅飞：** 尽可能多的运用ExPASy的各种工具对该蛋白的DNA序列进行分析。
- **b2李东旭：** 尽可能多的运用ExPASy的各种工具对该蛋白的蛋白质序列进行分析。

ExPASy简介

- ExPASy是Expert Protein Analysis System的缩写，即为专业蛋白质分析系统。ExPASy由瑞士生物信息学研究所维护（[Swiss Institute of Bioinformatics](http://www.expasy.org)），提供从序列（Swiss-Prot）到结构(Swiss-Model)，以及2-D PAGE等蛋白质操作相关的全套服务。ExPASy汇总的蛋白质组学在线实用分析工具包，涉及蛋白分类、翻译、结构预测、相似检索、序列比对等等。

ExPASy Proteomics tools

- [Protein identification and characterization] (MS相关)
 - [DNA -> Protein]
 - [Similarity searches]
 - [Sequence alignment]
- } (序列相关)
- [Pattern and profile searches] (domain/motif相关)
 - [Post-translational modification prediction]
 - [Topology prediction] (定位, 跨膜相关)
 - [Primary structure analysis]
 - [Secondary structure prediction]
 - [Tertiary structure]
- } (结构相关)
- [Phylogenetic analysis] (进化相关)
 - [Biological text analysis]

第一部分

斑头雁背景知识介绍

基本知识:

- ***Anser indicus* Latham**, 世界上飞的最高的鸟类
- 别名: 白头雁
- 英文名: Bar-headed Goose
- 斑头雁, 雁形目, 鸭科, 雁属。大雁的一种, 因头上有两条黑色条斑而得名。它体形如鹅, 重约2500-3500克, 长着桔红色嘴巴、黑色嘴唇、桔红色腿脚, 灰褐色的羽翅, 虹膜一褐色; 嘴一鹅黄, 嘴尖黑; 脚一橙黄。飞行高度可达8000米以上, 每日飞行300到500公里, 是鸟类飞行的一级运动员。



灰雁是斑头雁的近亲, 生活在欧洲和亚洲大陆。栖息于湖泊、河滩水域地带。它们却不能承受缺氧环境。所以怀疑可能是斑头雁的血红蛋白和灰雁的血红蛋白结构有所不同, 从而导致它结合氧的能力不同。

斑头雁血红蛋白研究背景及进展：

- 1979年 [Braunitzer G](#), [Oberthür W](#). 比对了灰雁和鸡的血红蛋白的 α 链和 β 链，发现灰雁的氨基酸序列相对于鸡的有差异，证明血红蛋白确实存在突变，为上面的对斑头雁的推测奠定了依据。
- 1980年 [Black CP](#), [Tenney SM](#) 做了一个研究工作，把斑头雁和北京鸭（体型和斑头雁相当，但生长在地面，不能飞的鸟类）的耐低氧能力做了一个比较，发现当把二者都置在高空中时，斑头雁没有明显的动作变化，但北京鸭的结合氧的能力却没有增强。
- 1982年，[Oberthür W](#), [Braunitzer G](#), [Würdinger I](#). 等人把斑头雁和灰雁的血红蛋白氨基酸序列进行了分析，其中斑头雁 α 链119位Ala在灰雁中变成了Pro，而斑头雁 β 链125位Asp变成了Glu，这两个氨基酸都在血红蛋白 α 链和 β 链的结合处。并试图解释了此变化造成斑头雁结合氧能力强的原因。

斑头雁血红蛋白研究背景及进展：

- 1993年，[Weber RE](#), 等人将人的血红蛋白进行了类似斑头雁血红蛋白的突变，发现，突变后的血红蛋白确实能够增强和氧的结合能力。
- 1999年北京大学Zhang J, Hua Z.等人解析了斑头雁的血红蛋白晶体结构，证明了斑头雁的高结合氧能力确实是和Pro变为Ala有关。

通过近二三十年的对斑头雁血红蛋白的研究，现在已经清楚的了解斑头雁血红蛋白的结构，以及由于结构的变化而导致斑头雁结合氧能力增强能在海拔9000多米飞翔的原因。斑头雁血红蛋白的研究也成为一个比较有代表性的例子，为其它近亲物种由于结构变异引起的功能差异奠定了很好的基础与参考。

第二部分

斑头雁血红蛋白的蛋白质分析

(一) 蛋白序列基本特征分析

- ProtParam tool
- Compute pI/Mw
- PeptideCutter
- PeptideMass
- Isotopident

1. **ProtParam tool:** 蛋白质的氨基酸组成、分子量、PI等基本特征的分析。

将斑头雁的血红蛋白序列输入该工具中，可以得到如下结果：

Number of amino acids: 142

Molecular weight: 15467.8

Theoretical pI: 8.54

Amino acid composition:

Ala (A)	19	13.4%
Arg (R)	3	2.1%
Asn (N)	3	2.1%
Asp (D)	7	4.9%

Total number of negatively charged residues (Asp + Glu): 13

Total number of positively charged residues (Arg + Lys): 15

Atomic composition:

Carbon	C	702
Hydrogen	H	1102
Nitrogen	N	190
Oxygen	O	196
Sulfur	S	4

Formula: C₇₀₂H₁₁₀₂N₁₉₀O₁₉₆S₄

Total number of atoms: 2194

Extinction coefficients/Estimated half-life /Instability index /Aliphatic index

该软件的分析结果包括了蛋白的氨基酸组成、原子组成、分子量、半衰期、等电点等全面的基本信息。

2.Compute pI/Mw: 通过与UniProt Knowledgebase中蛋白的比较计算目的蛋白的理论分子量与等电点。

以斑头雁的血红蛋白为例，结果如下：

Theoretical pI/Mw (average) for the user-entered sequence:

```
      10      20      30      40      50      60
MVLSAADKTN VKGVFSKISG HAE EYGAETL ERMFTAYPQT KTYFPHF DLQ HGSAQIKAHG
      70      80      90     100     110     120
KKVVAALVEA VNHIDDIAGA LSKLSDLHAQ KLRVDPVNFK FLGHCFLVVV AIHHPSALTA
      130     140
EVHASLDKFL CAVGTVLTAK YR
Theoretical pI/Mw: 8.54 / 15467.87
```

该软件分析结果只能得到蛋白质分子量和理论等电点的信息，但是与上面的分析结果是一致的。

3.PeptideCutter: 蛋白质酶切分析

以斑头雁的血红蛋白为例，

参数设置：

有多种酶可供选择，此例中选择Trypsin；

Trypsin 和Chymotrypsin酶还可以选择酶切效率，此例中选择酶切最低效率为100%；
结果可以表格形式、图谱形式等显示，此例中选择以表格形式显示；

Available enzymes

The enzyme(s) that you have chosen:

- Trypsin

You have chosen to display all possible cleaving enzymes.

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	6	8 (100%) 12 (100%) 32 (100%) 91 (100%) 93 (100%) 100 (100%)

选择酶切最低效率为90%时：

Name of enzyme	No. of cleavages	Positions of cleavage sites
Trypsin	12	8 (100%) 12 (94.4%) 17 (100%) 32 (100%) 41 (93.1%) 57 (95.3%) 61 (92.9%) 83 (100%) 91 (100%) 93 (100%) 100 (100%) 140 (100%)

4.PeptideMass: 蛋白酶切得到的肽段的分子量计算（也可以包含翻译后修饰）

以斑头雁的血红蛋白为例，

参数设置:

选择Trypsin作酶解分析；

Cysteines在蛋白预处理过程中的处理方法也可以进行选择，此例中选择“nothing（in reduced form）”；

也可以对“missed cleavages”进行设置，此例中设置为“0”；

还可以对肽段的显示方式进行选择，包括显示内容和顺序，此例中选择显示>500 Da的肽段，并按其大小顺序排列；

结果如下:

The peptide masses from your sequence are:

[Theoretical pI: 8.72 / Mw (average mass): 15257.55 / Mw (monoisotopic mass): 15247.93]

mass	position	#MC	modifications	peptide sequence
2996.4894	63-91	0		VADALTNAVAHVDDMPNALS ALSDLHAHK
2967.6124	101-128	0		LLSHCLLVTLAAHLPAEFTP AVHASLDK
1833.8918	42-57	0		TYFPHFDLSHGSAQVK
1529.7342	18-32	0		VGAHAGEYGAEALER
1252.7147	129-140	0		FLASVSTVLTSK
1071.5543	33-41	0		MFLSFPTTK
860.4546	1-8	0		MVLSPADK
818.4406	94-100	0		VDPVNFK
532.2878	13-17	0		AAWGK

90.8% of sequence covered (you may modify the input parameters to display also peptides < 500 Da):

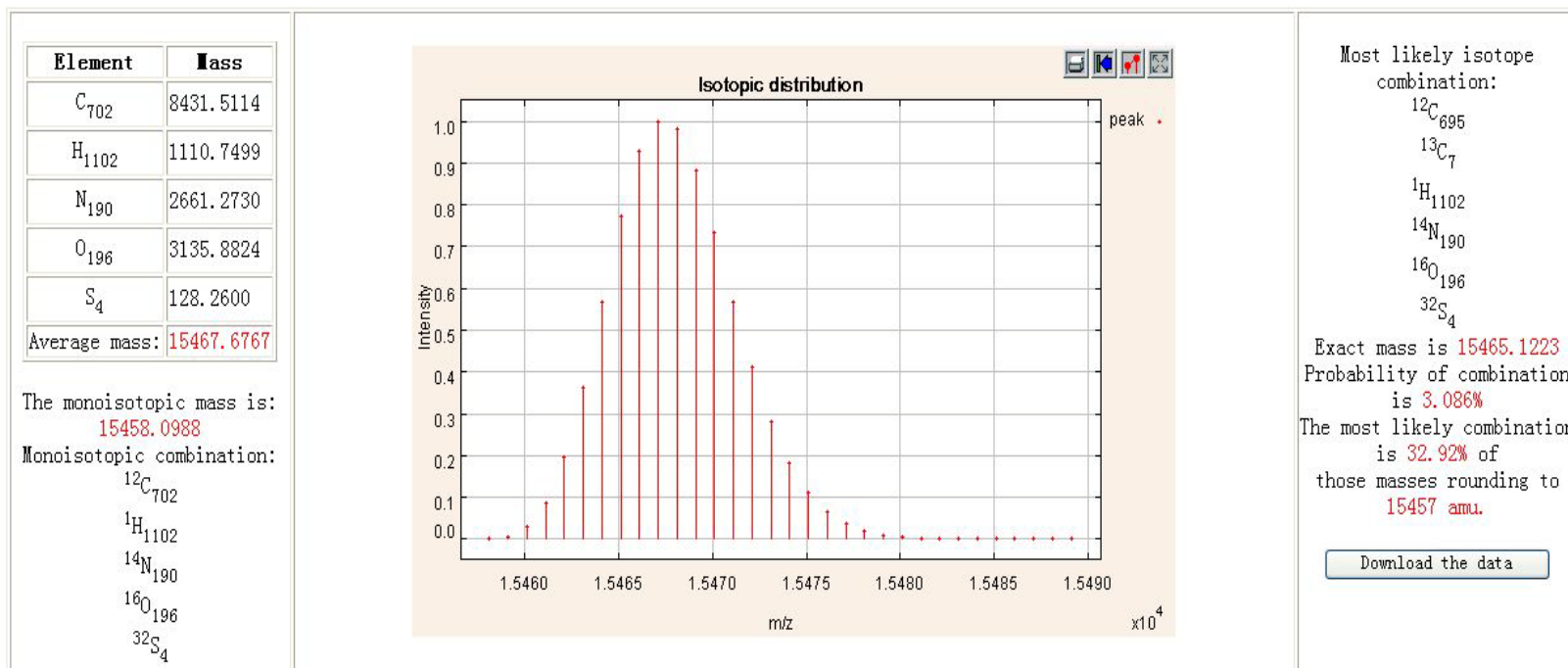
```

      10      20      30      40      50      60
MVLSPADKtn vkAAWGKVGGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS HGSAQVKghg
      70      80      90     100     110     120
kkVADALTNA VAHVDDMPNA LSALSDLHAH KlrVDPVNFK LLSHCLLVTL AAHLPAEFTP
      130     140
AVHASLDKFL ASVSTVLTSK yr
  
```

与**PeptideCutter**比较分析：**PeptideMass**所得的结果较直观，并且以表格形式集中显示出所得酶切肽段的分子量、氨基酸序列及肽段在整个蛋白中的位置，而且还可以对“missed cleavages”进行设置。另外，酶切肽段的显示方式也比**PeptideCutter**更加完善。

5.Isotopident:用来估测蛋白或肽段或多聚核苷酸链中的同位素分布情况，还可以计算出该蛋白或肽段的单一同位素峰的分子量以及自然情况下最有可能存在的同位素分布组合并计算出其分子量。这在质谱数据分析中很有用。

以斑头雁的血红蛋白为例，结果如下：



(二) Protein → DNA

- **Reverse Translate**
- **Genewise**

1.Reverse Translate:用来从蛋白质序列推测DNA序列，可以用来对还不知道基因序列的已知蛋白进行引物设计。

以斑头雁的血红蛋白分析为例，

参数设置: 采用default codon usage table,

结果如下:

>reverse translation of Untitled to a 426 base sequence of **most likely codons.**

```
atggtgctgagcgcggcggataaaaccaacgtgaaaggcgtgttagcaaaattagcggc
catgcggaagaatatggcgcggaaaccctggaacgcatgtttaccgcgtatccgcagacc
aaaacctattttccgcattttgatctgcagcatggcagcgcgcagattaaagcgcattggc
aaaaaagtggcggcggcgtggtggaagcggtaaccatattgatgatattgcgggcgcg
ctgagcaaactgagcgtatctgcatgcgcagaaactgcgcgtggatccggtgaactttaa
ttctgggccattgctttctggtggtggcgcgattcatcatccgagcgcgctgaccgcg
gaagtgcattgcgagcctggataaatttctgtgcgcggtgggcaccgtgctgaccgcgaa
tatcgc
```

>reverse translation of Untitled to a 426 base sequence of **consensus codons.**

```
atggtnytnwsngcngcngayaaracnaaygtnaargngntnttywsnaarathwsnggn
caycngargartayggngcngaracnytngarmgnatgttyacngcntayccncaracn
aaracntayttyccncayttygayytnarcayggcnwsngcncarathaargcncayggc
aaraargtngtngcngcnytngtngargcngtnaaycayathgaygayathgcnngngcn
ytnwsnaarytnwsngayytncaaygncaraarytnmgngtngayccngtnaayttyaar
ttyytnngncaytgytyytngtngtngtngcnathcaycayccnwsngcnytnacngcn
gargtncaygcnwsnytnngayaarttyytnngygcngtnggnacngtnytnacngcnaar
taymgn
```

同时它还给出了，每个氨基酸所对应的密码子的每个位置的核苷酸的出现概率，如下：

Graph of base probabilities:

```
.....  
3_L_first  
g 0.00  
a 0.00  
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 0.26  
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 0.74  
3_L_second  
g 0.00  
a 0.00  
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1.00  
C 0.00  
3_L_third  
gggggggggggggggggggggggggggggg 0.62  
aaaaaaaaaaaaaaaaaaa 0.17  
TTTTTTTTTTTTT 0.11  
CCCCCCCCCCC 0.10  
.....
```

该软件除了可以选择所使用的密码子的优点外，还具有在结果显示每个氨基酸所对应的密码子的每个位置的核苷酸的出现概率的优点。

2.Genewise: 可以将一个蛋白序列与某一个基因组DNA序列进行比较，适合寻找还不知道基因的蛋白的同源基因。

将斑头雁的血红蛋白序列与与Rat的血红蛋白基因的CDs序列进行比对，得到如下结果：

```
EMBOSS_001      1  MVL$AADKTNVKGVFSKISGHAEYGAETLERMFTAYPQTKTYFPHFDL ..
                  MVL$A DKTIN+K  + KI GH  EYG E  L+RMF A+P TKTYF H+D+ ..
                  MVL$ADDKTNIKNCWKGKIGGHGGEYGEEALQRMFAAFPTTKTYF$HIDV ..
EMBOSS_001      1  agctgggaaaaaattgaaggcggggtggggccaatgggtcaaaatttcagg ..
                  tttccaaacataaggggatggaggaagaactagttcctcccacatcatat ..
                  ggctatcaccgccgggttctttatcggcagggctcccccgccctetta ..
..
..
EMBOSS_001      50  QHGS$AQIKAHGKKVVAALVEAVNHIDDIAGALSKL$DLHAQKLRVDPVN ..
                  G$AQ+KAHGKKV  AL +A +H++D+ GALS L$DLHA KLRVDPVN ..
                  SPG$AQVKAHGKKVADALAKAADHVEDLPGAL$TL$DLHAHKLRVDPVN ..
EMBOSS_001      148 acgtgcgagcgaaggggtgagggcgggcccggctacagccgcaccggcga ..
                  gcgccatacagaatcactcaccataaatcgctcctgatacaatgtacta ..
                  ccctcgcggtccggtttctgcatacccacgttcgcgtgcgcgtccagtgttcc ..
..
..
EMBOSS_001      99  FKFLGHCFLVVVAIHHP$ALTAEVHASL$DKFLCAVGI$VLTAKYR ..
                  FKFL HC+LV +A HHP  T  +HASL$DKFL +V TVLT+KYR ..
                  FKFL$HCLLVTLACHHPGDFTPAMHASL$DKFLASV$TVLTSKYR ..
EMBOSS_001      295 tatcactccgatgtcccggtacgacgtcgatcgatgaagcatatc ..
                  tattgagtttctcgaacgatccctacctaattcctgcttccaag ..
                  cgcgccccggcggtccctatcaccgcctgcactctgctggccgct ..
..
..
//.
Gene 1..
Gene 1 426 |.
      Exon 1 426 phase 0..
```

(三) 蛋白序列比对

- BLAST
- Fasta3
- MPsrch

1.BLAST: 用来寻找蛋白的同源序列，再根据同源蛋白的已知信息来推测未知蛋白的相关信息。**BLAST**根据所使用数据库和不同软件有多种方法，在此仅以在ExPASy中对斑头雁的血红蛋白序列进行**BLAST**为例：

Db	AC	Description	Score	E-value
<input type="checkbox"/>	sp P01990	HBA_ANSIN Hemoglobin subunit alpha-A (Hemoglobin alpha...	285	8e-76
<input type="checkbox"/>	sp P01989	HBA_ANSAN Hemoglobin subunit alpha-A (Hemoglobin alpha...	280	2e-74
<input type="checkbox"/>	sp P01991	HBA_BRACA Hemoglobin subunit alpha-A (Hemoglobin alpha...	275	8e-73
<input type="checkbox"/>	sp P01987	HBA_CAIMO Hemoglobin subunit alpha-A (Hemoglobin alpha...	274	2e-72
<input type="checkbox"/>	sp P22740	HBA_ANAPE Hemoglobin subunit alpha-A (Hemoglobin alpha...	273	4e-72
<input type="checkbox"/>	sp P01988	HBA_ANAPP Hemoglobin subunit alpha-A (Hemoglobin alpha...	272	7e-72
<input type="checkbox"/>	tr Q5XME0	_OXYVI Alpha A hemoglobin [Oxyura vittata (Lake duck)]	272	7e-72
<input type="checkbox"/>	tr Q5XMD9	_OXYJA Alpha A hemoglobin [Oxyura jamaicensis jamaicensis]	272	7e-72
<input type="checkbox"/>	tr Q5XMD8	_9AVES Alpha A hemoglobin [Oxyura australis (blue-bille...	272	7e-72
<input type="checkbox"/>	tr Q5XMD7	_OXYLE Alpha A hemoglobin [Oxyura leucocephala (White-h...	272	7e-72
<input type="checkbox"/>	tr Q5XMD6	_9AVES Alpha A hemoglobin [Oxyura maccoa]	272	7e-72
<input type="checkbox"/>	sp P01992	HBA_CYGOL Hemoglobin subunit alpha-A (Hemoglobin alpha...	271	9e-72

从**BLAST**结果中可以看到，目的蛋白跟**Hemoglobin subunit**有很高的同源性，若我们不知道目的蛋白是斑头雁的血红蛋白的话，便由此推测目的蛋白可能与血红蛋白在结构或功能上相似。

2.Fasta3: 用FASTA和SSEARCH programs进行蛋白相似序列的搜索。
以斑头雁的血红蛋白为例：

Alignment	DB:ID	Source	Length	Identity%	Similar%	Overlap	E()
1 <input type="checkbox"/>	UNIPROT:HBA_ANSIN	Hemoglobin subunit alpha-A OS	142	100.0	100.0	142	6.7e-59
2 <input type="checkbox"/>	UNIPROT:HBA_ANSAN	Hemoglobin subunit alpha-A OS	142	97.9	99.3	142	9.9e-58
3 <input type="checkbox"/>	UNIPROT:HBA_BRACA	Hemoglobin subunit alpha-A OS	141	96.5	99.3	141	1.4e-56
4 <input type="checkbox"/>	UNIPROT:HBA_CAIMO	Hemoglobin subunit alpha-A OS	142	95.1	98.6	142	1.7e-56
5 <input type="checkbox"/>	UNIPROT:HBA_ANAPE	Hemoglobin subunit alpha-A OS	142	95.1	98.6	142	4e-56
6 <input type="checkbox"/>	UNIPROT:Q5XME0_OXYVI	Alpha A hemoglobin OS=Oxyu	142	93.7	99.3	142	5.6e-56
7 <input type="checkbox"/>	UNIPROT:Q5XMD7_OXYLE	Alpha A hemoglobin OS=Oxyu	142	93.7	99.3	142	5.6e-56
8 <input type="checkbox"/>	UNIPROT:Q5XMD9_OXYJA	Alpha A hemoglobin OS=Oxyu	142	93.7	99.3	142	5.6e-56
9 <input type="checkbox"/>	UNIPROT:Q5XMD6_9AVES	Alpha A hemoglobin OS=Oxyu	142	93.7	99.3	142	5.6e-56
10 <input type="checkbox"/>	UNIPROT:Q5XMD8_9AVES	Alpha A hemoglobin OS=Oxyu	142	93.7	99.3	142	5.6e-56

3.MPsrch: a biological sequence comparison tool that implements the true Smith and Waterman algorithm.

以斑头雁的血红蛋白为例:

Alignment	DB:ID	Description	Length	Match%	Query Match%	Score	Pred.No.
1 <input type="checkbox"/>	UNIPROT:HBA_ANSIN	RecName: Full=Hemoglobin subunit alpha-A;AltName: Full=Hemoglobin alpha-A chain;AltName: Full=Alpha-A-globin;	142	100.0	100.0	1212	4.37e-296
2 <input type="checkbox"/>	UNIPROT:HBA_ANSAN	RecName: Full=Hemoglobin subunit alpha-A;AltName: Full=Hemoglobin alpha-A chain;AltName: Full=Alpha-A-globin;	142	97.9	98.5	1194	4.94e-291
3 <input type="checkbox"/>	UNIPROT:HBA_CAIMO	RecName: Full=Hemoglobin subunit alpha-A;AltName: Full=Hemoglobin alpha-A chain;AltName: Full=Alpha-A-globin;	142	95.1	96.7	1172	7.36e-285
4 <input type="checkbox"/>	UNIPROT:HBA_BRACA	RecName: Full=Hemoglobin subunit alpha-A;AltName: Full=Hemoglobin alpha-A chain;AltName: Full=Alpha-A-globin;	141	96.5	96.5	1170	2.68e-284
5 <input type="checkbox"/>	UNIPROT:Q5XMD6_9AVES	SubName: Full=Alpha A hemoglobin;	142	93.7	96.2	1166	3.54e-283
6 <input type="checkbox"/>	UNIPROT:Q5XMD7_OXYLE	SubName: Full=Alpha A hemoglobin;	142	93.7	96.2	1166	3.54e-283
7 <input type="checkbox"/>	UNIPROT:Q5XMD8_9AVES	SubName: Full=Alpha A hemoglobin;	142	93.7	96.2	1166	3.54e-283
8 <input type="checkbox"/>	UNIPROT:Q5XMD9_OXYJA	SubName: Full=Alpha A hemoglobin;	142	93.7	96.2	1166	3.54e-283
9 <input type="checkbox"/>	UNIPROT:Q5XME0_OXYVI	SubName: Full=Alpha A hemoglobin;	142	93.7	96.2	1166	3.54e-283
10 <input type="checkbox"/>	UNIPROT:HBA_ANAPE	RecName: Full=Hemoglobin subunit alpha-A;AltName: Full=Hemoglobin alpha-A chain;AltName: Full=Alpha-A-globin;	142	95.1	96.1	1165	6.76e-283

上述三种工具的比较:

Fasta3所使用的SSEARCH严格按照Smith-Waterman原理进行相似序列搜索，FASTA则对于低相似性的长序列分析比对具有专一性；MPsrch目前被认为是最灵敏的序列比对工具，也是使用Smith-Waterman原理，但它利用了exhaustive algorithm，而BLAST and FASTA仅仅是启发式的。

(四) 一级结构分析

- Three-/one-letter amino acid converter
- ScanSite pI/Mw

1.Three-/one-letter amino acid converter: 可以将蛋白氨基酸序列由三个字母显示转换成一个字母显示，反之亦然。

以斑头雁的血红蛋白为例：

One letter sequence:

```
> 142 aminoacids; Mw=15469.79Da
MVLSAADKTNVKGVFSKISG
HAEEYGAETLERMFTAYPQT
KTYFPHFDLQHGSAQIKAHG
KKVVAALVEAVNHIDDIAGA
LSKLSDLHAQKLRVDPVNFK
FLGHCFLVVVAIHHPALTA
EVHASLDKFLCAVGTVLTA
YR*
```

Three letter sequence:

```
> 142 aminoacids; Mw=15469.79Da
MetValLeuSerAlaAlaAspLysThrAsnValLysGlyValPheSerLysIleSerGly
HisAlaGluGluTyrGlyAlaGluThrLeuGluArgMetPheThrAlaTyrProGlnThr
LysThrTyrPheProHisPheAspLeuGlnHisGlySerAlaGlnIleLysAlaHisGly
LysLysValValAlaAlaLeuValGluAlaValAsnHisIleAspAspIleAlaGlyAla
LeuSerLysLeuSerAspLeuHisAlaGlnLysLeuArgValAspProValAsnPheLys
PheLeuGlyHisCysPheLeuValValAlaIleHisHisProSerAlaLeuThrAla
GluValHisAlaSerLeuAspLysPheLeuCysAlaValGlyThrValLeuThrAlaLys
TyrArg***
```

蛋白质结构数据库（PDB）用三字符缩写表示，而蛋白质序列、家族、功能等数据库用单字符代码表示，以减少存储空间，提高数据库搜寻速度。常用序列分析软件也用单字符表示。因此，该软件使用频率较高，是蛋白分析的基本软件。

2.ScanSite pI/Mw: 可以计算蛋白在各种磷酸化状态下的分子量及等电点。

以斑头雁的血红蛋白为例：

# Phosphates	Molecular Weight	Isoelectric Point
0	15479.7836	8.54
1	15557.7476	7.38
2	15635.7116	6.89
3	15713.6756	6.62
4	15791.6396	6.42
5	15869.6036	6.25
6	15947.5676	6.09
7	16025.5316	5.95
8	16103.4956	5.81

该软件可用于等点聚焦分离具有多种磷酸化修饰状态的磷酸化蛋白的分析。

(五) 二级结构分析

- GOR
- SOPMA

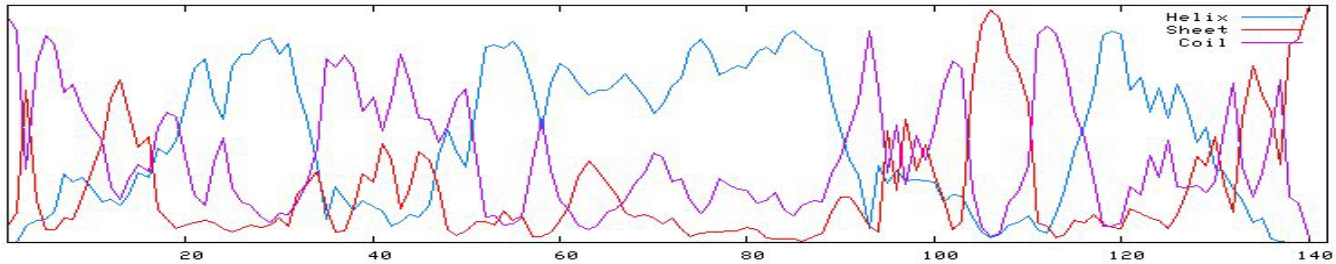
1.GOR: 蛋白质二级结构预测。斑头雁的血红蛋白为例:

```
      10      20      30      40      50      60      70
MVLSAADKTNVKGVFSKISGHAEEYGAETLERMFTAYPQIKTYFPHPFDLQHGSAQIKAHGKKVVAALVEA
ccccccccccccceeeecchhhhhhhhhhhhhhhcccccccccccccccccccccccccccccccccccccccc
VNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPALTAEVHASLDKFLCAVGTVLTAK
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
YR
ec
```

Sequence length : 142

GOR4 :

Alpha helix (Hh) :	68 is	47.89%
3 ₁₀ helix (Gg) :	0 is	0.00%
Pi helix (Ii) :	0 is	0.00%
Beta bridge (Bb) :	0 is	0.00%
Extended strand (Ee) :	18 is	12.68%
Beta turn (Tt) :	0 is	0.00%
Bend region (Ss) :	0 is	0.00%
Random coil (Cc) :	56 is	39.44%
Ambiguous states (?) :	0 is	0.00%
Other states :	0 is	0.00%



该软件可以预测分析蛋白的helix, sheet, coil等二级结构。

2.SOPMA: 蛋白质二级结构预测。以斑头雁的血红蛋白为例:

```

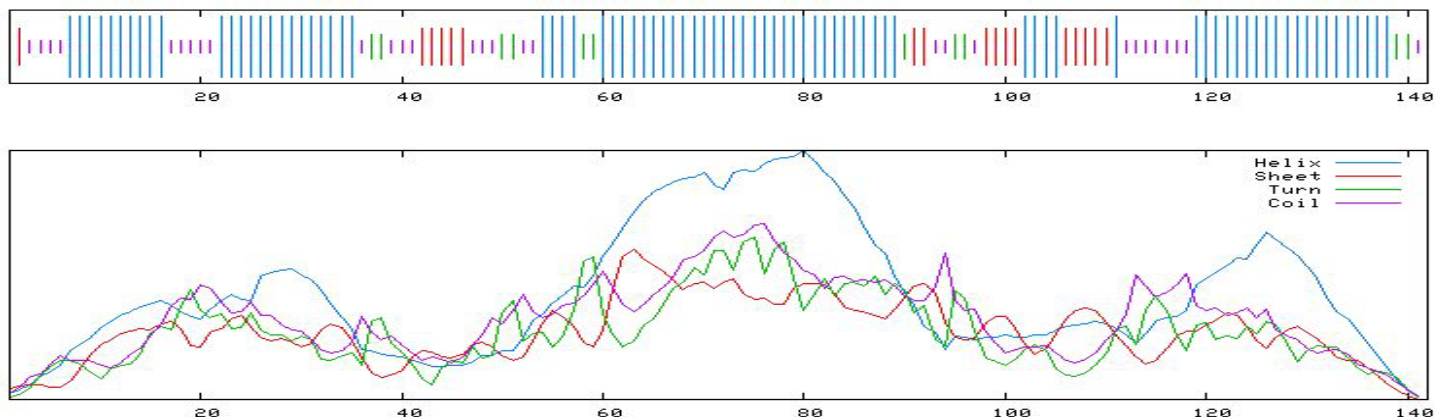
      10      20      30      40      50      60      70
      |      |      |      |      |      |
MVLSAADKTNVKGVFSKISGHAEYGAETLERMFTAYPQTKTYFPHF DLQHGSAQIKAHGKKVVAALVEA
eeecccc hhhhhhhhhhhhh ccccc hhhhhhhhhhhhhhhhh tcccc eeeee ecccc tccc hhhh t hhhhhhhhhhh
VNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHP SALTAEVHASLDKFLCAVGTVLTAK
hhhhhhhhhhhhhhhhhhhh t eec ctt c e e e e h h h e e e e h c c c c c h h h h h h h h h h h h t
YR
tc

```

Sequence length : 142

SOPMA :

Alpha helix	(Hh) :	83 is	58.45%
₃ 10 helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	19 is	13.38%
Beta turn	(Tt) :	11 is	7.75%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	29 is	20.42%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%



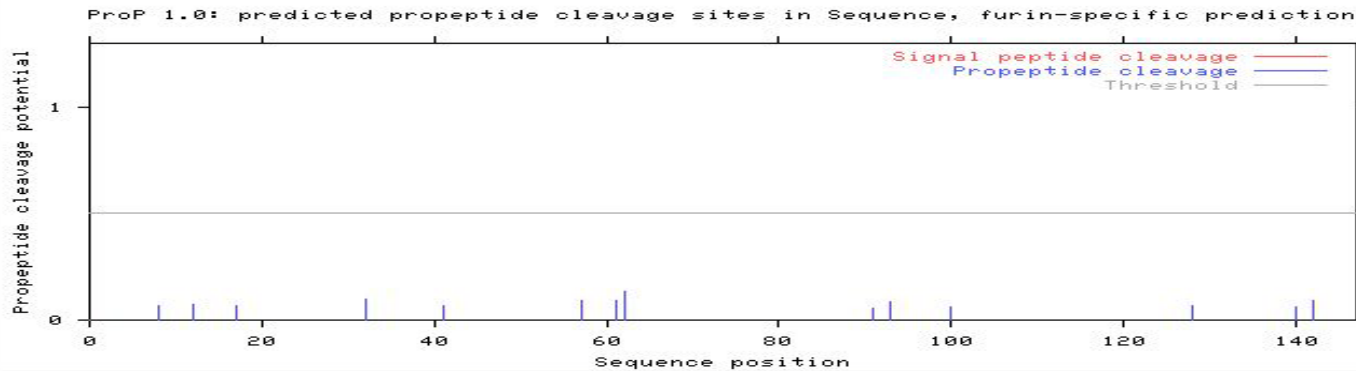
该软件功能与GOR功能差不多，所得结果稍微有些差别，可能是由于原理的差别而造成的。

(六) 翻译后修饰分析

- ProP
- NetPhos

1.ProP: 分析蛋白质中赖氨酸和精氨酸的断开位点，以斑头雁的血红蛋白为例：

Name	Pos	Context	Score	Pred
Sequence	8	VLSPADK TN	0.066	.
Sequence	12	ADKINVK AA	0.074	.
Sequence	17	VKAAWGK VG	0.068	.
Sequence	32	GAEALER MF	0.099	.
Sequence	41	LSFPTTK TY	0.068	.
Sequence	57	HGSAQVK GH	0.094	.
Sequence	61	QVKGHGK KV	0.091	.
Sequence	62	VKGHGKK VA	0.136	.
Sequence	91	SDLHAHK LR	0.055	.
Sequence	93	LHAHKLR VD	0.085	.
Sequence	100	VDPVNFK LL	0.064	.
Sequence	128	VHASLDK FL	0.065	.
Sequence	140	STVLTSK YR	0.063	.
Sequence	142	VLTSKYR --	0.089	.



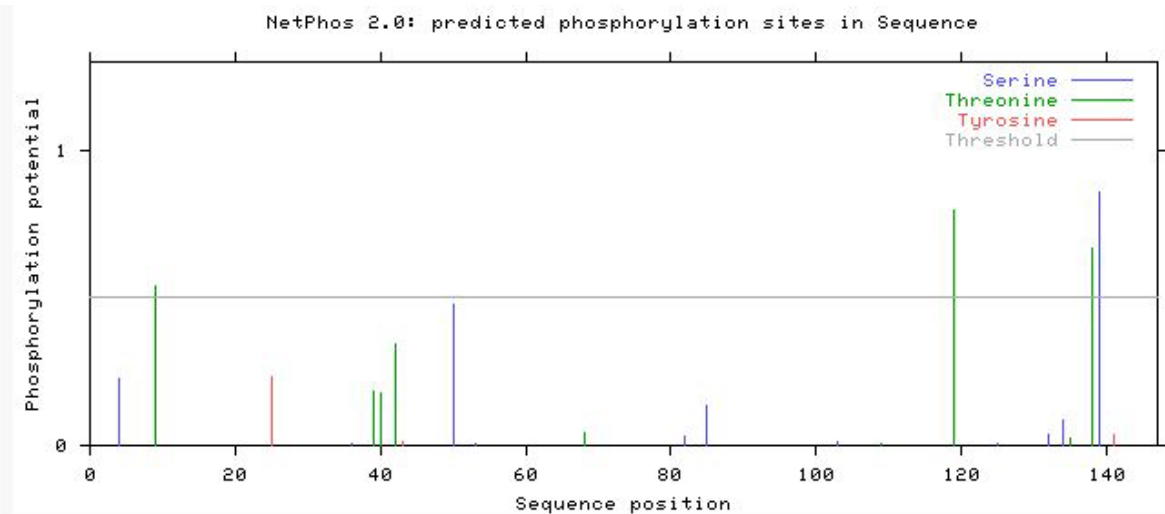
该软件功能与Trypsin酶切功能有些相似。

2.NetPhos: 分析蛋白质中serine, threonine及tyrosine的磷酸化位点。 以斑头雁的血红蛋白为例:

Serine predictions				
Name	Pos	Context	Score	Pred
v				
Sequence	4	-MVLSPADK	0.226	.
Sequence	36	RMFLSFPTT	0.009	.
Sequence	50	HFDLSHGSA	0.477	.
Sequence	53	LSHGSAQVK	0.005	.
Sequence	82	PNALSALSD	0.028	.
Sequence	85	LSALSDLHA	0.133	.
Sequence	103	FKLLSHCLL	0.011	.
Sequence	125	AVHASLDKF	0.009	.
Sequence	132	KFLASVSTV	0.037	.
Sequence	134	LASVSTVLT	0.087	.
Sequence	139	TVLTSKYR-	0.859	*S*
^				

Threonine predictions				
Name	Pos	Context	Score	Pred
v				
Sequence	9	PADKINVKA	0.540	*T*
Sequence	39	LSFPITKTY	0.184	.
Sequence	40	SFPITKTYF	0.177	.
Sequence	42	PTTKTYFPH	0.341	.
Sequence	68	ADALTNVA	0.041	.
Sequence	109	CLLVTLAAH	0.005	.
Sequence	119	PAEFTPAVH	0.800	*T*
Sequence	135	ASVSTVLT	0.022	.
Sequence	138	STVLTISKYR	0.668	*T*
^				

Tyrosine predictions				
Name	Pos	Context	Score	Pred
v				
Sequence	25	HAGEYGAEA	0.234	.
Sequence	43	TIKTYFPHF	0.010	.
Sequence	141	LTSKYR---	0.035	.
^				



该软件对于蛋白质的磷酸化位点分析及预测很有价值。

第三部分

斑头雁血红蛋白DNA序列分析

考虑到Expasy是主要分析蛋白质序列的，我们对该网站上可以用以分析DNA的软件也进行了尝试，在这里只是略讲。

主要有一下工具：

Codon Plot

Codon Usage

CpG Islands

DNA Molecular Weight

DNA Pattern Find

ORF FINDER

Pairwise Align DNA

Codon Plot可以对DNA序列的密码子组成进行分析，也可以分析出相应的氨基酸的出现频率以及该密码子出现的概率。

Codon Usage可以得出密码子出现的概率，也可以得出各个氨基酸的简并密码子的频率。

CpG Islands可以分析GC含量，可以通过该软件的分析设计一定的片段克隆的引物，并判断引物的退火温度。

DNA Molecular Weight分析DNA的分子质量。

DNA Pattern Find可以用来分析连续氨基酸出现的频率，可进而分析氨基酸Group在该蛋白中的作用。

ORF FINDER可以分析不同的翻译可能性，直接翻译出相应的蛋白。

Pairwise Align DNA可以用来进行保守性分析，进而分析功能结构上的进化同源性。

致谢

- 罗老师
- 任娇
- 申立平
- 孔寅飞