

ABC作业规范

基本规范

- 邮件标题：
 - 单人：s1a1_h1
 - 小组：s1a1a2b1b2_h1
- 作业名称(以第一组为例)：
 - 单人：pku09s1a1_h1a.doc
 - 小组：pku09s1a1a2b1b2_h1a.doc

基本规范

- 作业基本格式

- 页眉:

- 单人: 姓名, 作业号, 座位号, 组号 (组员),

- 宋新文, pku09s1a1_h1a, s1a1, G1 (s1a1,s1a2,s1b1,s1b2)

- 小组: 姓名, 作业号, 组号 (组员)

- 宋新文, 陈亮, 沈璧蓉, 赵文力, pku09s1a1a2b1b2_h1a, G1(s1a1,s1a2,s1b1,s1b2)

- 页脚: 页码

- 标题: **ABC**第一次作业第一部分

- 副标题: 该次作业的小课题名称

- 作者 (以贡献大小), 单位 (**实验室)

基本规范

- 正文结构
 - 摘要
 - 简介（必须）
 - 方法与结果（必须）
 - 总结（必须）
 - 讨论
 - 小组分工（必须）与致谢
- 正文要求：
 - 正文使用五号宋体，单倍行距；
 - 各级标题使用较大号的宋体或加粗区别出来；
 - 适当空行（两大部分之间等）

基本规范

- 工具使用描述规范
 - 需要写出的信息：工具名称，版本号，网址（如果有），输入数据和输出数据，各项参数
- 图表表示规范
 - 居中或其他合适的编排
 - 需要有编号
 - 图表注（注明表格或图片的内容）

(1) 使用 DAS-TMfilter server (<http://mendel.imp.ac.at/sat/DAS/DAS.html>) 预测
参数如下: Output format: short; Evaluation: trusted; TM-library size: 8;
Version: last update: May 27 2003, 结果如图 2 所示。

```
Calculating prediction for the following proteins
with reference library 08:

>gi7C1453366797Cref7CNP_175700.27C catalytic 5BArabidopsis thaliana5D
... Done.

*** List of predicted non-TM-protein codes ***

>gi7C1453366797Cref7CNP_175700.27C catalytic 5BArabidopsis thaliana5D

*** List of predicted TM-protein codes ***

none

=== Result of the prediction ===

>gi7C1453366797Cref7CNP_175700.27C catalytic 5BArabidopsis thaliana5D
# TMH: 0 Q: trusted !!! Warning! Non-TM protein!

<----- end of list ----->
```

图 2: DAS 预测结果

Good example

生物信息作业

张旭

10811082

pku09s1d1

EBI-SRS 使用实例和 swiss-pdbviewer 使用实例

张旭

北京大学生命科学学院干细胞与再生生物学实验室 北京 100871

摘要: SRS(Sequence Retrieval System)是由欧洲生物信息研究所(EBI)开发的以 WWW 界面运行的数据库检索及导航系统,是目前生物信息界应用最为广泛的数据库系统。Swiss-pdbviewer 是著名的查看 pdb 格式文件的软件,它功能强大,可通过友好的界面来查看蛋白结构的各种数据。本文以实例来讲述 EBI-SRS 和 swiss-pdbviewer 的使用方法。

关键词: SRS, sequence retrieval system, PDB, swiss-pdbviewer, 生物信息, 蛋白结构

SRS, 即 Sequence Retrieval System, 是检索分子生物学序列和其它数据最重要的数据库之一。它把众多的数据库,如 RefSeq genome, UniProtKB/Swiss-Prot 等整合成一个导航系统。用户可以通过它方便地检索这些众多数据库中的信息,并利用 SRS 自带的分析工具对检索得到的数据进行分析。

Swiss pdbviewer 是一个可同时查看和分析多个蛋白质结构的软件。该软件以友好的用户界面,提供多种可视化视图,使得用户可以方便地分析蛋白质的空间构象,氨基酸序列及突变, H 键, 原子间距离等多种信息。是同类软件中功能最强大,最为著名的一个。

1 材料与amp;方法

使用 EBI 的 SRS (<http://srs.ebi.ac.uk>) 来查找 POU 转录因子家族的所有蛋白,并将它们的 fasta 格式的数据保存到本地,然后利用 weblab/Macro/phylogenetic 进行分析。

以 POU 转录因子家族的一个蛋白: PO2F1_mouse 为例,来学习 swiss-pdbviewer 的使用方法。

Bad example

Nucleotide sequence databases

all EMBL Patent DNA EMBL (Contig)
 EMBL (Contigs expanded) EMBL (Annotated Cons) EMBL (Coding Sequences)

click the link

Gene
 Nucleotide

You can browse through

fill the blank

and

there are found 18 alpha subunits entries results, and search beta there are 41 entries. While most part of them are partial, mutant, subunit variant and truncated. I chose the full length :

<input type="checkbox"/>	EMBL (Coding Sequences):CAA23752	V00493.1	Homo sapiens (human)	Homo sapiens (human) hemoglobin alpha chain	429
--------------------------	--	--------------------------	----------------------	---	-----

There are too many partial hemoglobin beta, so I add

there are 12 entries, some of them are the same after needle program. I chose this one.

Bad example

Swiss-PDBviewer 的应用↵

本次上机主要用 [swiss-PDBviewer](#) 对几个蛋白的结构进行了展示。↵

↵

本次上机的蛋白功能的介绍：“该蛋白主要的功能是与 [cul1](#)——[roc1](#) 形成一个紧密的核心层，阻止 E3 的的接聚。该蛋白的晶体结构显示，该蛋白适配一个高度螺旋的结果，与 [cul1](#) 的伸长的 [scf](#) 碧形成夹合。最后，β 发卡结构，又可以部分的占据 [cul1](#) 的的结合位点，阻止与 [skp1](#) 的适配蛋白的结合。另一方面，另外两个 HEAT 的重复结构与 [cul1](#) 的保守序列堆积，并且，掩盖掉 [cul1](#) 的 [lys](#) 的位点被修饰。这些结构共同作用的结果就是：共同调节了 [scf](#) 和其他的 [cullin](#) 依赖的 E3 复合体的聚合，和接聚~”（译自 PDB 网站）；↵

↵

为了用 [swiss-pdbviewer](#) 的 alignment 功能，还对其中一个功能相似的蛋白质也进行了结构的展示。“该蛋白也是 E3 泛素连接酶家族的一份子。这里可以得到的是 [cul1-rbx-skp1](#) 的复合体结构。”↵

Bad example

对其序列进行多项比对进行分析 distance matrix 分析得出↵

Distance Matrix↵

-----↵

↵

Uncorrected for Multiple Substitutions↵

Gap weighting is 0.000000↵

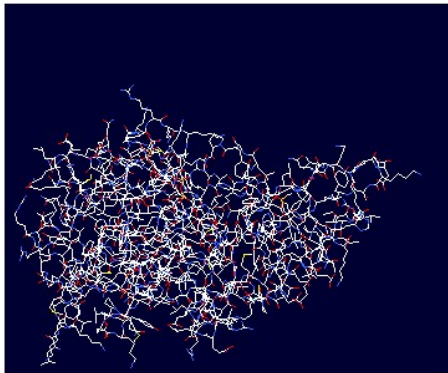
↵

	1	2	3↵	
	0.00	21.83	54.29	SWISSPROT HBA_HUMAN 1↵
		0.00	57.14	SWISSPROT HBA_RAT 2↵
			0.00	SWISSPROT HBB1_MOUSE 3↵

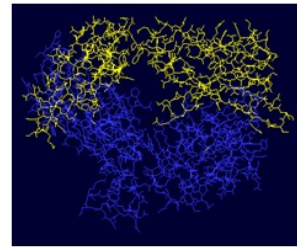
↵

所以仅根据此蛋白的序列比对结果得出人与大鼠关系更近，而小鼠远一些。
要得出更科学的结果还需对其其他蛋白进行分析。↵

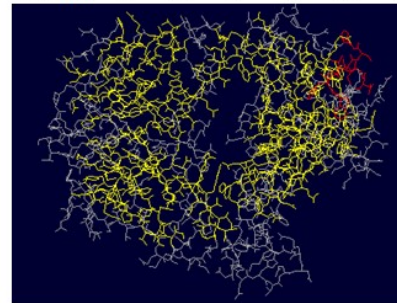
Bad example



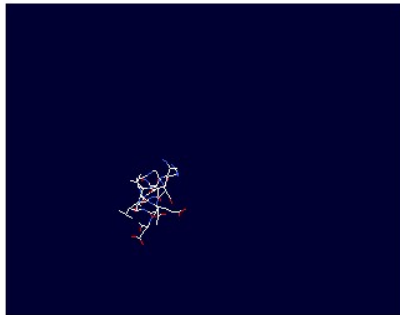
group	+	+	+	show	side	label
A	LEU-6	v	v			
A	TYR-5	v	v			
A	LYS-4	v	v			
A	h LYS-3	v	v			
A	h ALA-2	v	v			
A	h GLY-1	v	v			
A	h LEU0	v	v			
A	MET1	v	v			
A	VAL2	v	v			
A	ASP3	v	v			
A	LYS4	v	v			
A	s LYSS	v	v			
A	s LEU6	v	v			
A	s VAL7	v	v			
A	s VAL8	v	v			
A	s VAL9	v	v			
A	s PHE10	v	v			
A	GLY11	v	v			



此蛋白就有两条链构成。Select secondary structure 红色标记螺旋 (h), 黄色标记折叠(s), 无规卷曲则标记为灰色。



Color by type 有四种颜色标注, 蓝色代表带正电, 红色代表带负电, 黄色代表不带电, 灰色代表无极性。



A	GLY11	v	
A	GLY12	v	
A	THR13	v	
A	GLY14	v	
A	h ALA15	v	
A	h GLN16	v	
A	h GLY17	v	
A	h GLY18	v	
A	h SER19	v	
A	h VAL20	v	
A	h ALA21	v	
A	h ARG22	v	
A	h THR23	v	
A	h LEU24	v	
A	h LEU25	v	
A	h GLU26	v	
A	h ASP27	v	
A	GLY28	v	

