

用于系统地描绘多重比对的特征，these mark-up annotations are preceded by a 'magic' label，由四个不同的标签指定的内容组成注释。常被HMMER, Pfam和Belvu使用。

1. Header

首行包含了当前格式和版本信息

2. The sequence alignment

```
<seqname> <aligned sequence>
```

//

sequence, 表示序列名称，格式有 "name/start-edn"或"name"

// 表示比对终止

序列字符由任何字符表示，除了空格；gap由"."或"-"代表

3. The alignment mark-up

描述信息可包含除了空格意外任何字符，常用"_"来取代空格

```
#=GF<feature><Generic per-File annotation, free text>
```

```
#=GC<feature><Generic per-Column annotation, exactly 1 char per column>
```

```
#=GS<seqname><feature><Generic per-Sequence annotation, free text>
```

```
#=GR<seqname><feature><Generic per-Sequence AND per-Column markup, exactly 1 char per column>
```

Magic or recommended features:

```
#=GF
```

```
#=GF NH <tree in New Hampshire extended format>
```

```
#=GF TN<Unique identifier for the next tree>
```

注意：

一个树结构可能存在多重的#=GF NH 行，假如多个树结构存在与一个文件，每个树结构必须以#=GF TN line with a unique tree identifier 开头。如果只有一个树结构，#=GC TN 行可能被删除。

```
#=GC
```

相同的特征由#=GR with "_cons" 后缀表示，例如 ss_cons

```
#=GS
```

Pfam使用以下特征

Feature Description

```
AC <accession>: ACcession number
```

DE <freetext> : Description

DR <db>; <accession> : Database Reference

OS <organism> : Organism (species)

OC <clade> : Organism Classification (clade, etc.)

OL <look> : Look (Color, etc.)

#=GR

Feature Description Markup letters

SS: Secondary Structure: [HGIEBTSCX]

SA: Surface Accessibility: [0-9X]

(0=0%-10%; ...; 9=90%-100%)

TM: TransMembrane: [Mio]

PP: Posterior Probability: [0-9*]

(0=0.00-0.05; 1=0.05-1.05; *=0.95=1.00)

LI: Ligand binding: [*]

AS: Active Site: [*]

pAS: AS - Pfam predicted: [*]

sAS: AS - from SwissProt: [*]

IN: INtron (in or after): [0-2]

注意:

不用在多重行使用相同的#=GR 标签, 每个序列只有唯一的特征描述

"X"在SA和SS中表示" residue with unknown structure"

In SS the letters are taken from DSSP: H=alpha-helix, G=3/10-helix, I=p-helix, E=extended strand, B=residue in isolated b-bridge, T=turn, S=bend, C=coil/loop.)

推荐个magic line位置:

#=GF above the alignment

#=GC below the alignment

#=GS above the alignment or just below the corresponding sequence

#=GR just below the corresponding sequence

4. Size limits

没有大小限制

Line length: 10000.

<seqname>: 255

<feature>: 255

Example:

```
# STOCKHOLM 1.0
#=GF ID CBS
#=GF AC PF00571
#=GF DE CBS domain
#=GF AU Bateman A
#=GF CC CBS domains are small intracellular modules mostly found
#=GF CC in 2 or four copies within a protein.
#=GF SQ 67
#=GS 031698/18-71 AC 031698
#=GS 083071/192-246 AC 083071
#=GS 083071/259-312 AC 083071
#=GS 031698/88-139 AC 031698
#=GS 031698/88-139 OS Bacillus subtilis
083071/192-246      MTCRAQLIAVPRASSLAE..AIACAQKM...RVS RVPVYERS
#=GR 083071/192-246 SA 999887756453524252..55152525....36463774777
083071/259-312      MQHVSAPVFVFECTRLAY..VQHKLRAH....SRAVAIVLDEY
#=GR 083071/259-312 SS CCCCCHHHHHHHHHHHHHH..EEEEEEEE...EEEEEEEEEEE
031698/18-71        MIEADKVAHVQVGNNLEH..ALLVLT KT...GYTAIPVLDPS
#=GR 031698/18-71 SS CCCHHHHHHHHHHHHHHHH..EEEEEEEE...EEEEEEEEHHH
031698/88-139       EVMLTDIPRLHINDPIMK..GFGMVINN.....GFVCVENDE
#=GR 031698/88-139 SS CCCCCCHHHHHHHHHHHHH..EEEEEEEE...EEEEEEEEEEEH
#=GC SS_cons        CCCCCHHHHHHHHHHHHHH..EEEEEEEE...EEEEEEEEEEEH
031699/88-139       EVMLTDIPRLHINDPIMK..GFGMVINN.....GFVCVENDE
#=GR 031699/88-139 AS _____*_____
#=GR_031699/88-139_IN _____1_____2_____0_____
//
```