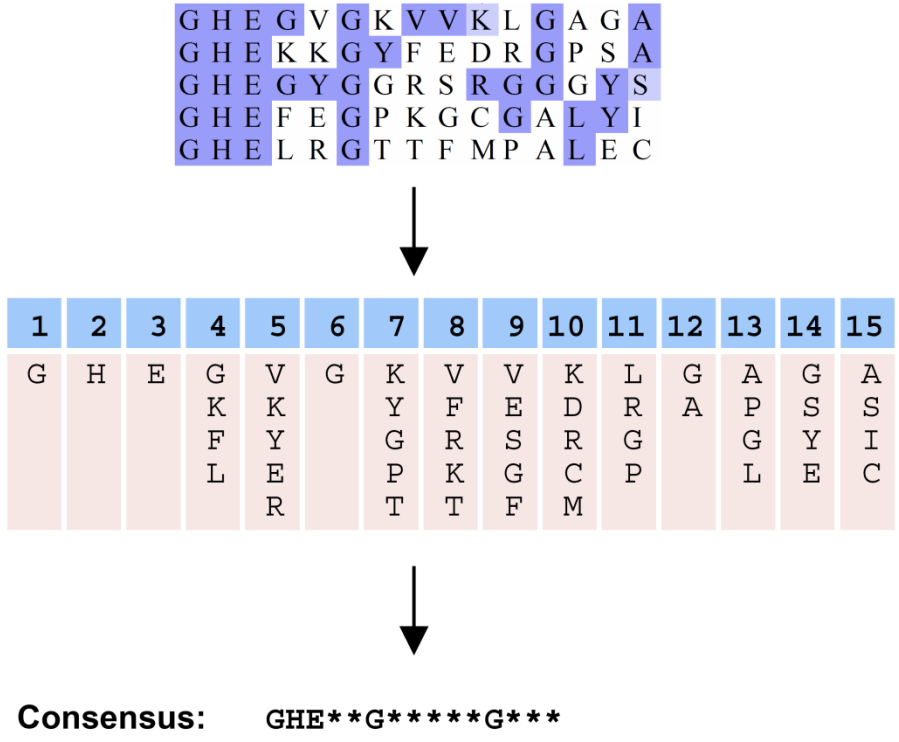
Motivos e perfis

Sequência de consenso

A partir do alinhamento múltiplo de uma família de sequências é possível determinar preferências **posicionais** para a ocorrência dos 20 a.a.

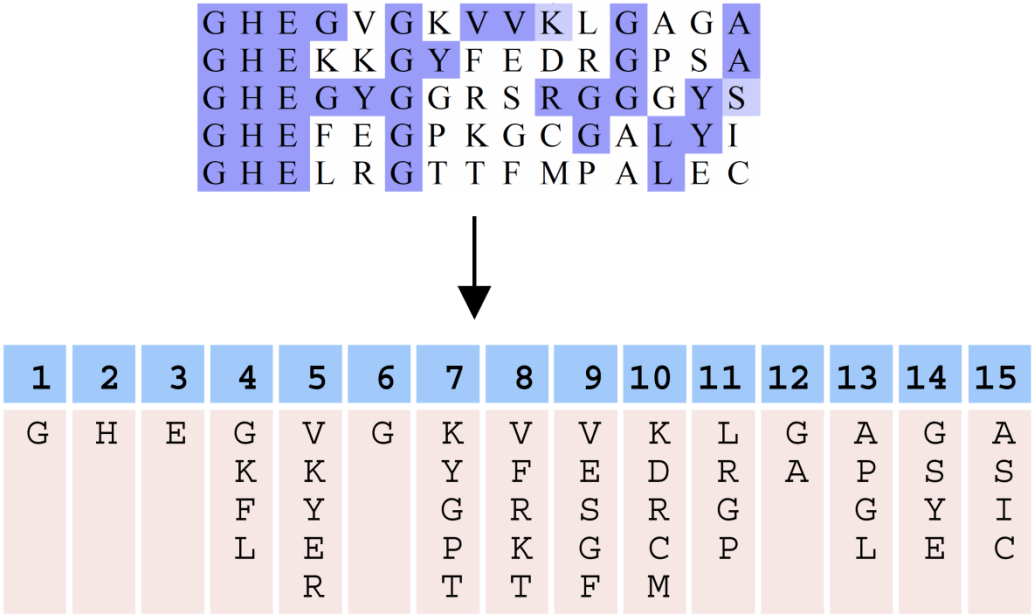
Exemplo:



Sequência de consenso: posições 100% conservadas no alihamento.

Protocolo de alinhamentos múltiplos

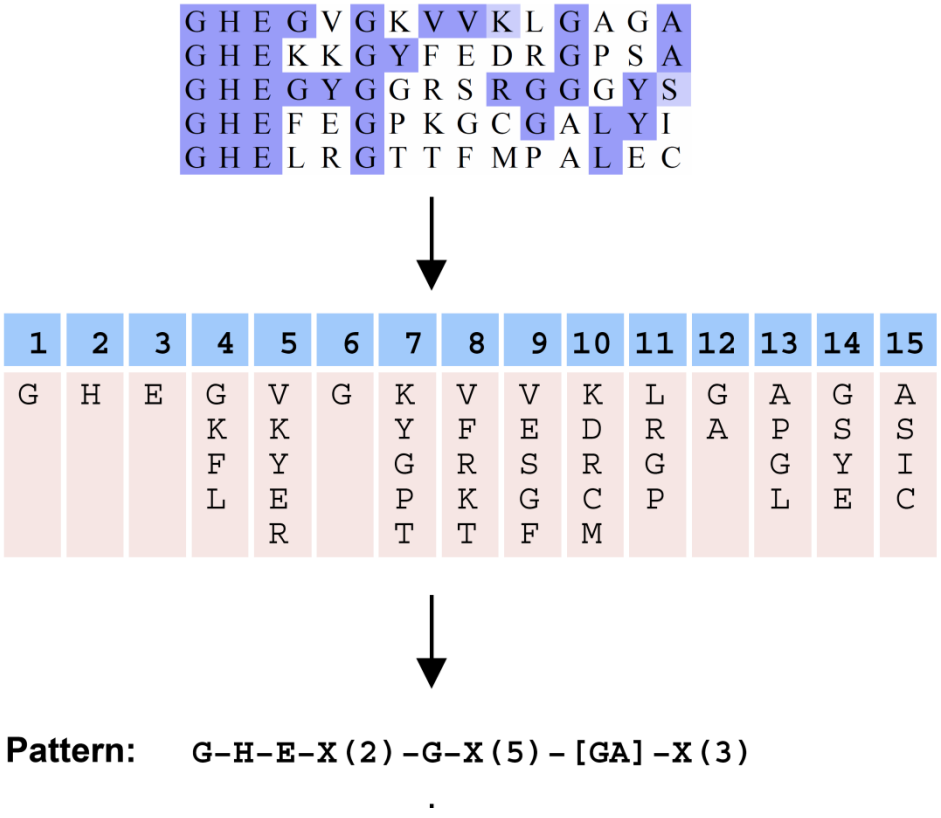
A sequência de consenso não permite descrever preferências não-integrais (<100%).



Na posição 12 podem ocorrer dois resíduos (G e A). Como representar este tipo de situação ?

Construção de Motivos

A variedade composicional do alinhamento múltiplo pode ser descrita por um **motivo**.



O **motivo** ou padrão representa de forma simbólica as características de conservação da zona alinhada. Muitos domínios funcionais apresentam motivos (ou assinaturas) característicos.

Formato dos padrões PROSITE

A sintaxe dos padrões PROSITE rege-se pelas seguintes regras:

1.Códigos IUPAC para aminóacidos (1-letter)

2.Os elementos são separados por “-”

3.“X” representa qualquer aminoácido

4.Ambiguidades representadas por “[]” (Ex. [AG] = A **ou** G)

5.Aminoácidos proibidos entre “{}” (Ex: {AG} todos os a.a. excepto A ou G)

6.Repetições são representadas por “()” (Ex: X(2) dois a.a. quaisquer, [AG](2,4) A ou G de 2 a 4 vezes

7.Para o C- e N-term (match no início ou fim) usam-se os símbolos “<“ ou “>”

Exemplo:

G-H-E-X(2)-G-X(5)-[GA]-X(3)

Exemplo de padrão PROSITE

O seguinte padrão:

<A-x-[ST](2)-x(0,1)-{V}

significa:

1. Uma Ala (A) no N-terminal,
2. Seguida de um aminoácido qualquer,
3. Seguida de uma Ser(S) ou Thr(T) duas vezes,
4. Seguida de zero ou um aminoácidos quaisquer,
5. Seguida de qualquer aminoácido menos Valina (V).

Motivos PROSITE

Padrões PROSITE que descrevem motivos de sequência primária característicos de locais de reconhecimento ou de famílias de proteínas. Associados a aspectos **funcionais** e **estruturais**.

Exemplos (usando o formato PROSITE):

• site de fosforilação das proteínas cinases:
[RK](2)-x-[ST]

• local de glicosilação:
S-G-x-G

• "Zipper" de leucina:
L-x(6)-L-x(6)-L-x(6)-L

• Família das proteases de serina, histidina do centro activo:
[LIVM]-[ST]-A-[STAG]-H-C

• Local de γ-carboxilação dependente da vitamina-K:
x(12)-E-x(3)-E-x-C-x(6)-[DEN]-x-[LIVMFY]-x(9)-[FYW]

Base de dados PROSITE

- Base de dados de famílias e domínios proteícos
- Apesar do elevado número de proteínas conhecidas, a maioria pode ser agrupada num número limitado de famílias, com base na simillaridade
- Proteínas ou domínios proteicos pertencendo a uma determinada família têm geralmente uma mesma função e um ancestral comum
- O estudo das famílias de proteínas indica que a conservação não é constante ao longo da sequência
- As zonas mais conservadas têm geralmente importância funcional
- Comparação das zonas conservadas permite derivar uma **assinatura**, ou motivo, que distingue os membros dessa família de outras proteínas
- PROSITE contem **motivos** e **perfis** para mais de 1000 famílias de proteínas
- Contem patterns (motivos) e profiles (perfis) que são formas diferentes de descrever assinaturas de uma sequência

Profiles, PWM, PSWM, P | ScanProsite | www.uniprot.org/uniprot | ExPASy - PROSITE

Paulo

prosite.expasy.org

Apps | Bookmarks | D pmartel | UALG Acad | Drug Design | Programming | Bioinformatics | Molecular Modelling | Other bookmarks

ExPASy Bioinformatics Resource Portal

PROSITE

Home | Contact

Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding

Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 20.131 of 27-Oct-2016 contains 1773 documentation entries, 1309 patterns, 1172 profiles and 1193 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Search

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] Examples

Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format

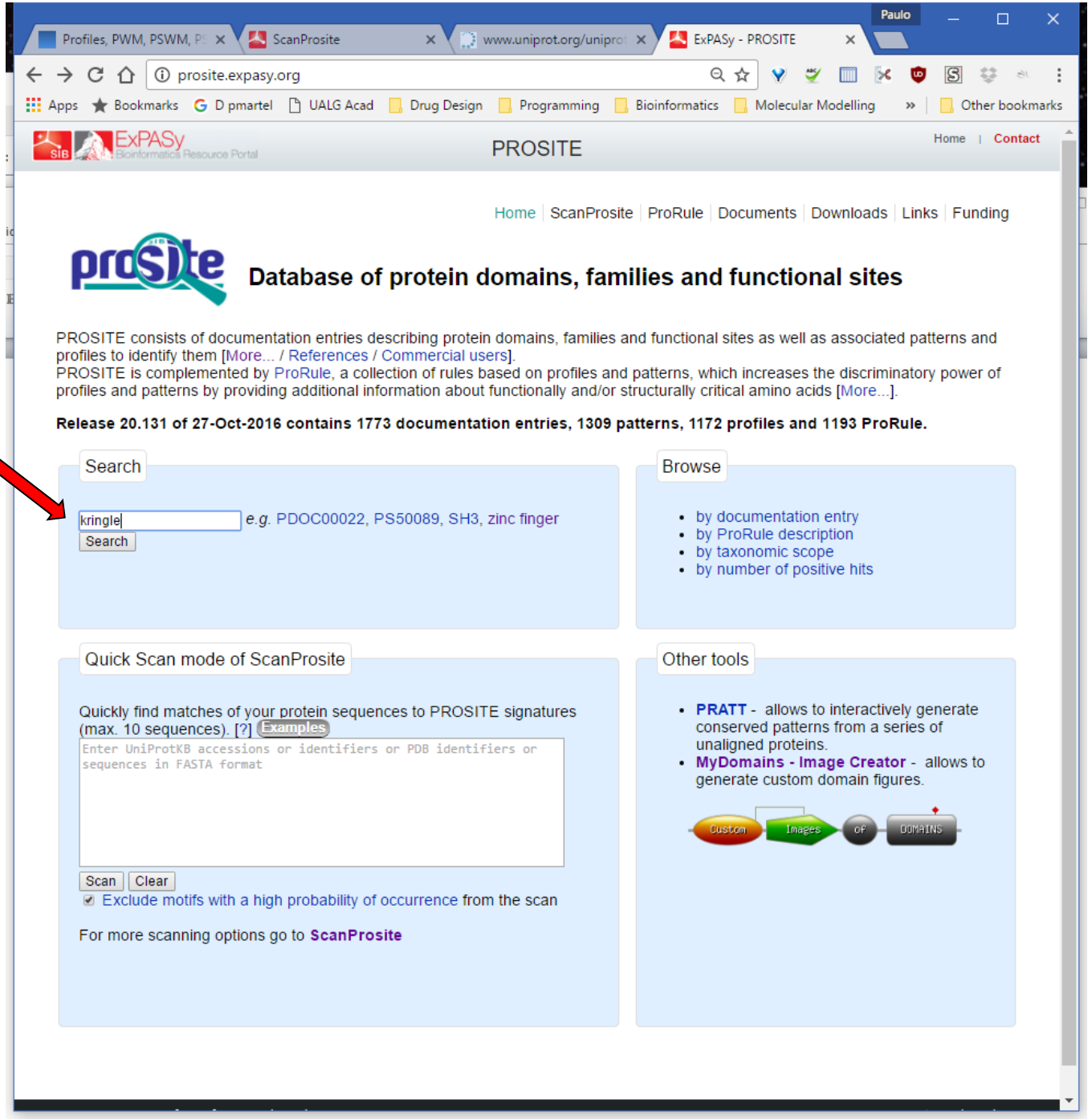
Scan | Clear

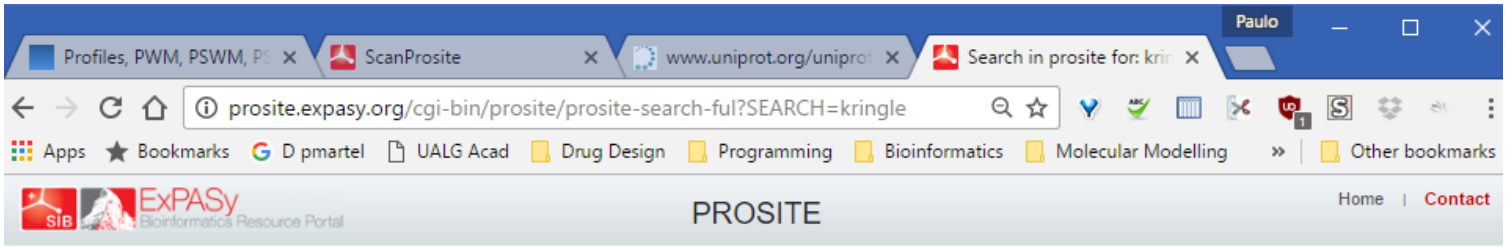
Exclude motifs with a high probability of occurrence from the scan

For more scanning options go to **ScanProsite**

Other tools

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.


Profiles, PWM, PSWM, P
ScanProsite
www.uniprot.org/uniprot
ExPASy - PROSITE
Paulo
prosite.expasy.org
Apps
Bookmarks
D pmartel
UALG Acad
Drug Design
Programming
Bioinformatics
Molecular Modelling
Other bookmarks
ExPASy
Bioinformatics Resource Portal
PROSITE
Home | Contact
Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding
Database of protein domains, families and functional sites
PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].
Release 20.131 of 27-Oct-2016 contains 1773 documentation entries, 1309 patterns, 1172 profiles and 1193 ProRule.
Search
kringle
e.g. PDOC00022, PS50089, SH3, zinc finger
Search
Browse
by documentation entry
by ProRule description
by taxonomic scope
by number of positive hits
Quick Scan mode of ScanProsite
Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] Examples
Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format
Scan
Clear
Exclude motifs with a high probability of occurrence from the scan
For more scanning options go to ScanProsite
Other tools
PRATT - allows to interactively generate conserved patterns from a series of unaligned proteins.
MyDomains - Image Creator - allows to generate custom domain figures.
Custom
Images
of
DOMAINS



Prosite search results

Search in PROSITE for: kringle

(Release 20.131, of 27-Oct-2016)

Enter search terms:

kringle

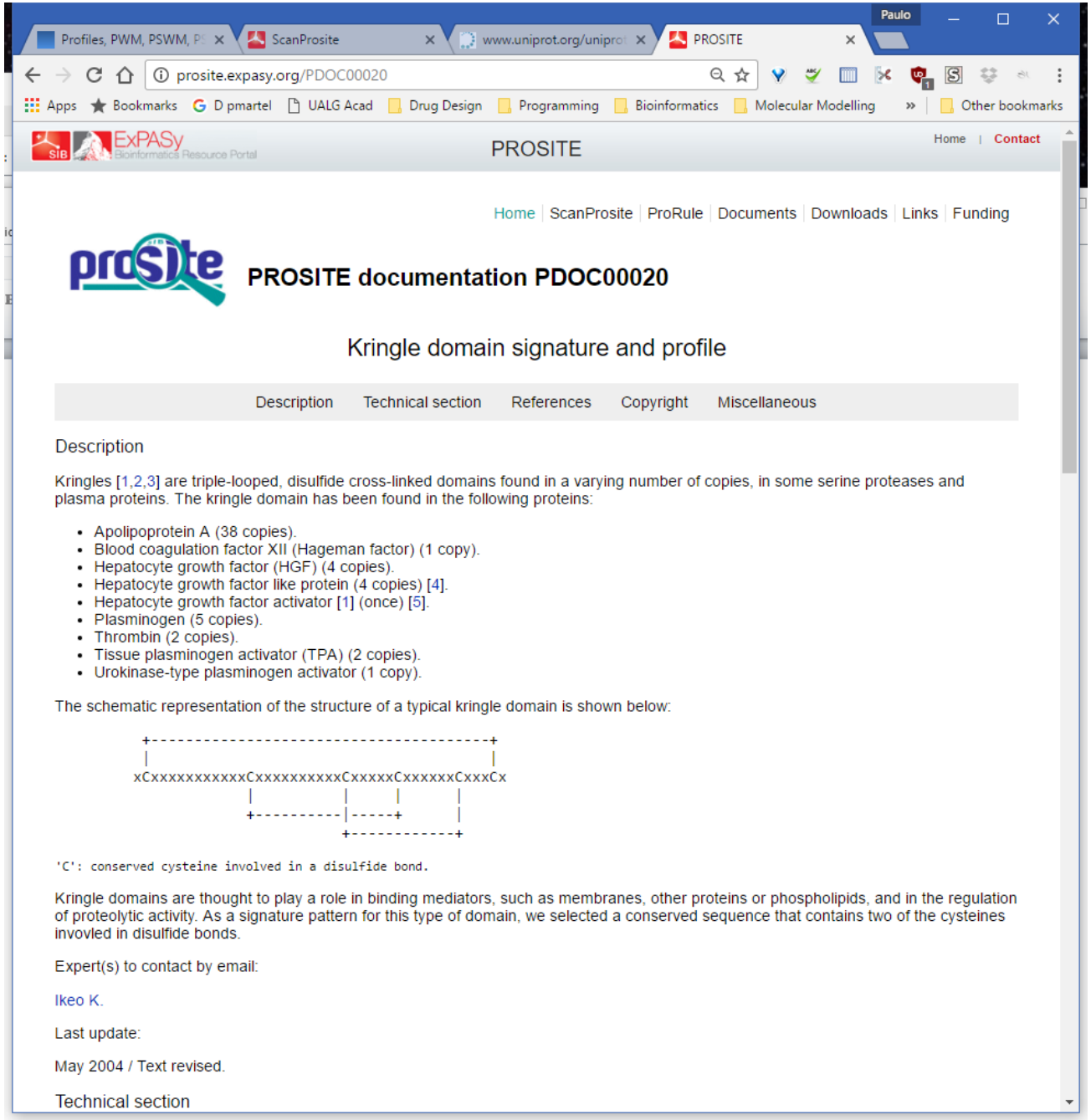
Prefix and append wildcard '*' to words.

new search clear

By default, this search engine searches for complete words only. If you did not find what you expected, and would try to do a substring match, you should perform a new search and select 'prefix and append wildcard to words'.

Number of documents in PROSITE containing the search term:4

- PDOC00537 C-type lectin domain signature and profile
- PDOC00965 Fibronectin type-I domain signature and profile
- PDOC00020 Kringle domain signature and profile
- PDOC51212 WSC domain profile


Profiles, PWM, PSWM, P
ScanProsite
www.uniprot.org/uniprot
PROSITE
Paulo
prosite.expasy.org/PDOC00020
Apps
Bookmarks
D pmartel
UALG Acad
Drug Design
Programming
Bioinformatics
Molecular Modelling
Other bookmarks
ExPASy
Bioinformatics Resource Portal
PROSITE
Home | Contact
Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding
PROSITE documentation PDOC00020
Kringle domain signature and profile
Description | Technical section | References | Copyright | Miscellaneous
Description
Kringles [1,2,3] are triple-looped, disulfide cross-linked domains found in a varying number of copies, in some serine proteases and plasma proteins. The kringle domain has been found in the following proteins:
Apolipoprotein A (38 copies).
Blood coagulation factor XII (Hageman factor) (1 copy).
Hepatocyte growth factor (HGF) (4 copies).
Hepatocyte growth factor like protein (4 copies) [4].
Hepatocyte growth factor activator [1] (once) [5].
Plasminogen (5 copies).
Thrombin (2 copies).
Tissue plasminogen activator (TPA) (2 copies).
Urokinase-type plasminogen activator (1 copy).
The schematic representation of the structure of a typical kringle domain is shown below:
'C': conserved cysteine involved in a disulfide bond.
Kringle domains are thought to play a role in binding mediators, such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity. As a signature pattern for this type of domain, we selected a conserved sequence that contains two of the cysteines invovled in disulfide bonds.
Expert(s) to contact by email:
Ikeo K.
Last update:
May 2004 / Text revised.
Technical section

May 2004 / Text revised.

Technical section

PROSITE methods (with tools and information) covered by this documentation:

KRINGLE_2, PS50070; Kringle domain profile (MATRIX)

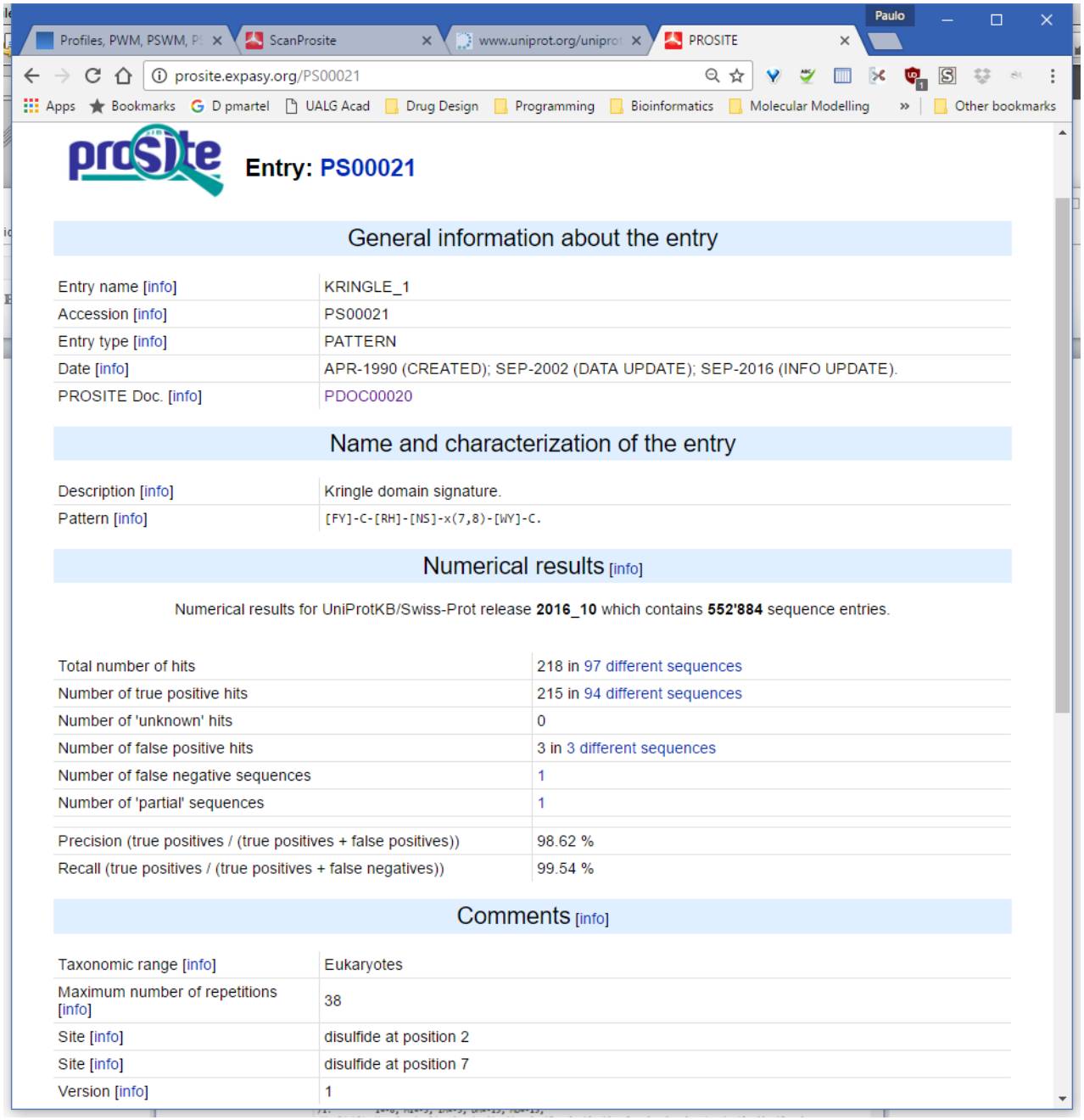
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 96
  - detected by PS50070: 95 (true positives)
  - undetected by PS50070: 1 (0 false negative and 1 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50070:
  NONE.
- Domain architecture view of Swiss-Prot proteins matching PS50070

- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50070
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50070
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50070
- View ligand binding statistics of PS50070
- Matching PDB structures: 1A0H 1B2I 1BHT 1CEA ... [ALL]

KRINGLE_1, PS00021; Kringle domain signature (PATTERN)

- Consensus pattern:
  [FY]-C-[RH]-[NS]-x(7,8)-[WY]-C
  The 2 C's are involved in a disulfide bonds
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 96
  - detected by PS00021: 94 (true positives)
  - undetected by PS00021: 2 (1 false negative and 1 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00021:
  3 false positives.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00021
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00021
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00021
- View ligand binding statistics of PS00021
- Matching PDB structures: 1A0H 1B2I 1BHT 1CEA ... [ALL]

References



Entry: PS00021

General information about the entry

Entry name [info]	KRINGLE_1
Accession [info]	PS00021
Entry type [info]	PATTERN
Date [info]	APR-1990 (CREATED); SEP-2002 (DATA UPDATE); SEP-2016 (INFO UPDATE).
PROSITE Doc. [info]	PDOC00020

Name and characterization of the entry

Description [info]	Kringle domain signature.
Pattern [info]	`[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C.`

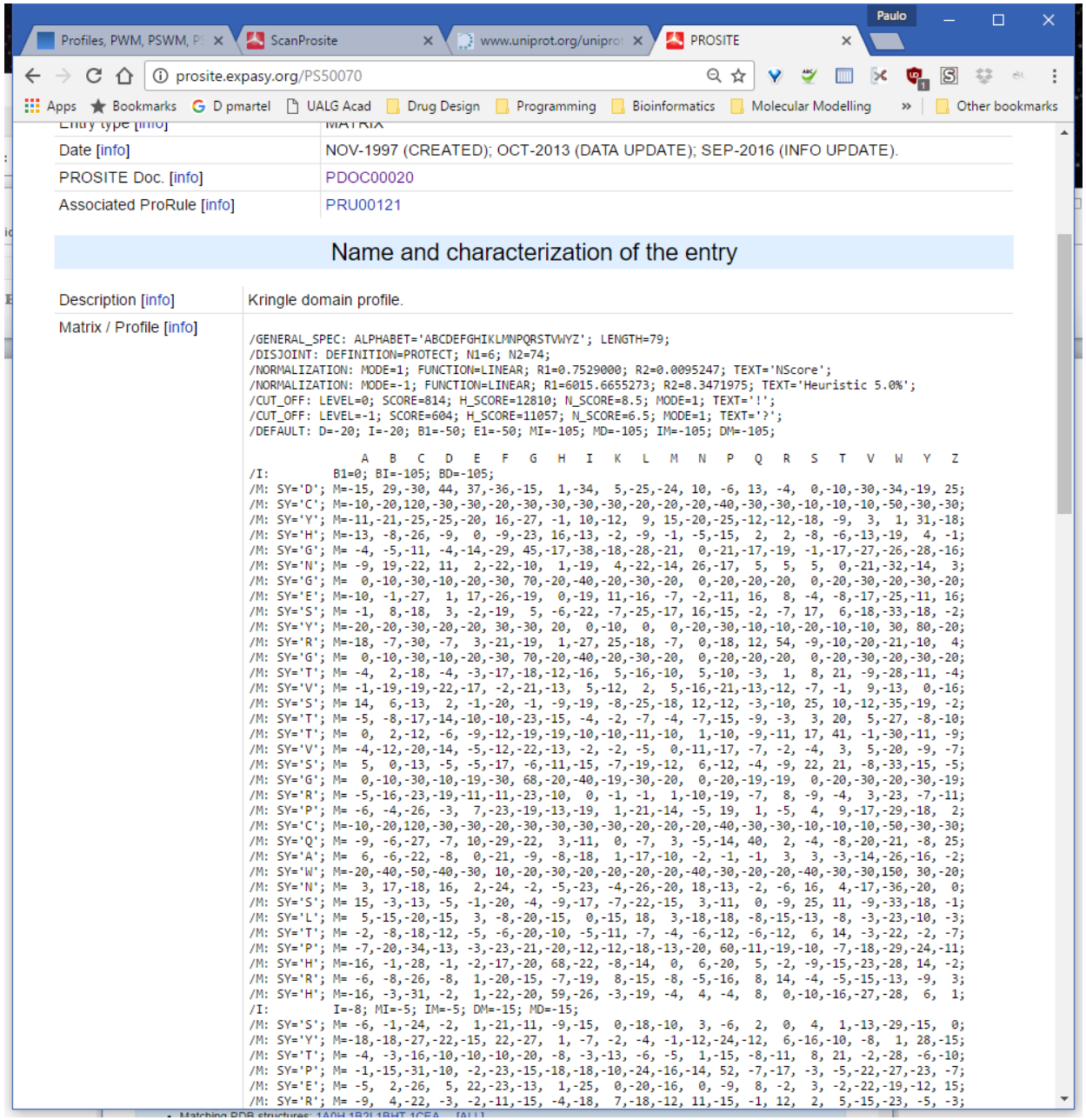
Numerical results [info]

Numerical results for UniProtKB/Swiss-Prot release **2016_10** which contains **552'884** sequence entries.

Total number of hits	218 in 97 different sequences
Number of true positive hits	215 in 94 different sequences
Number of 'unknown' hits	0
Number of false positive hits	3 in 3 different sequences
Number of false negative sequences	1
Number of 'partial' sequences	1
Precision (true positives / (true positives + false positives))	98.62 %
Recall (true positives / (true positives + false negatives))	99.54 %

Comments [info]

Taxonomic range [info]	Eukaryotes
Maximum number of repetitions [info]	38
Site [info]	disulfide at position 2
Site [info]	disulfide at position 7
Version [info]	1



Entry type [info]	MATRIX
Date [info]	NOV-1997 (CREATED); OCT-2013 (DATA UPDATE); SEP-2016 (INFO UPDATE).
PROSITE Doc. [info]	PDOC00020
Associated ProRule [info]	PRU00121

Name and characterization of the entry

Description [info]	Kringle domain profile.
Matrix / Profile [info]	

```
/GENERAL_SPEC: ALPHABET='ABCDEFGHIKLMNPQRSTVWYZ'; LENGTH=79;
/DISJOINT: DEFINITION=PROTECT; N1=6; N2=74;
/NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=0.7529000; R2=0.0095247; TEXT='NScore';
/NORMALIZATION: MODE=-1; FUNCTION=LINEAR; R1=6015.6655273; R2=8.3471975; TEXT='Heuristic 5.0%';
/CUT_OFF: LEVEL=0; SCORE=814; H_SCORE=12810; N_SCORE=8.5; MODE=1; TEXT='!';
/CUT_OFF: LEVEL=-1; SCORE=604; H_SCORE=11057; N_SCORE=6.5; MODE=1; TEXT='?';
/DEFAULT: D=-20; I=-20; B1=-50; E1=-50; MI=-105; MD=-105; IM=-105; DM=-105;

               A   B   C   D   E   F   G   H   I   K   L   M   N   P   Q   R   S   T   V   W   Y   Z
/I:         B1=0; BI=-105; BD=-105;
/M: SY='D'; M=-15, 29,-30, 44, 37,-36,-15,  1,-34,  5,-25,-24, 10, -6, 13, -4,  0,-10,-30,-34,-19, 25;
/M: SY='C'; M=-10,-20,120,-30,-30,-20,-30,-30,-30,-30,-20,-20,-20,-40,-30,-30,-10,-10,-10,-50,-30,-30;
/M: SY='Y'; M=-11,-21,-25,-25,-20, 16,-27, -1, 10,-12,  9, 15,-20,-25,-12,-12,-18, -9,  3,  1, 31,-18;
/M: SY='H'; M=-13, -8,-26, -9,  0, -9,-23, 16,-13, -2, -9, -1, -5,-15,  2,  2, -8, -6,-13,-19,  4, -1;
/M: SY='G'; M= -4, -5,-11, -4,-14,-29, 45,-17,-38,-18,-28,-21,  0,-21,-17,-19, -1,-17,-27,-26,-28,-16;
/M: SY='N'; M= -9, 19,-22, 11,  2,-22,-10,  1,-19,  4,-22,-14, 26,-17,  5,  5,  5,  0,-21,-32,-14,  3;
/M: SY='G'; M=  0,-10,-30,-10,-20,-30, 70,-20,-40,-20,-30,-20,  0,-20,-20,-20,  0,-20,-30,-20,-30,-20;
/M: SY='E'; M=-10, -1,-27,  1, 17,-26,-19,  0,-19, 11,-16, -7, -2,-11, 16,  8, -4, -8,-17,-25,-11, 16;
/M: SY='S'; M= -1,  8,-18,  3, -2,-19,  5, -6,-22, -7,-25,-17, 16,-15, -2, -7, 17,  6,-18,-33,-18, -2;
/M: SY='Y'; M=-20,-20,-30,-20,-20, 30,-30, 20,  0,-10,  0,  0,-20,-30,-10,-10,-20,-10,-10, 30, 80,-20;
/M: SY='R'; M=-18, -7,-30, -7,  3,-21,-19,  1,-27, 25,-18, -7,  0,-18, 12, 54, -9,-10,-20,-21,-10,  4;
/M: SY='G'; M=  0,-10,-30,-10,-20,-30, 70,-20,-40,-20,-30,-20,  0,-20,-20,-20,  0,-20,-30,-20,-30,-20;
/M: SY='T'; M= -4,  2,-18, -4, -3,-17,-18,-12,-16,  5,-16,-10,  5,-10, -3,  1,  8, 21, -9,-28,-11, -4;
/M: SY='V'; M= -1,-19,-19,-22,-17, -2,-21,-13,  5,-12,  2,  5,-16,-21,-13,-12, -7, -1,  9,-13,  0,-16;
/M: SY='S'; M= 14,  6,-13,  2, -1,-20, -1, -9,-19, -8,-25,-18, 12,-12, -3,-10, 25, 10,-12,-35,-19, -2;
/M: SY='T'; M= -5, -8,-17,-14,-10,-10,-23,-15, -4, -2, -7, -4, -7,-15, -9, -3,  3, 20,  5,-27, -8,-10;
/M: SY='T'; M=  0,  2,-12, -6, -9,-12,-19,-19,-10,-10,-11,-10,  1,-10, -9,-11, 17, 41, -1,-30,-11, -9;
/M: SY='V'; M= -4,-12,-20,-14, -5,-12,-22,-13, -2, -2, -5,  0,-11,-17, -7, -2, -4,  3,  5,-20, -9, -7;
/M: SY='S'; M=  5,  0,-13, -5, -5,-17, -6,-11,-15, -7,-19,-12,  6,-12, -4, -9, 22, 21, -8,-33,-15, -5;
/M: SY='G'; M=  0,-10,-30,-10,-19,-30, 68,-20,-40,-19,-30,-20,  0,-20,-19,-19,  0,-20,-30,-20,-30,-19;
/M: SY='R'; M= -5,-16,-23,-19,-11,-11,-23,-10,  0, -1, -1,  1,-10,-19, -7,  8, -9, -4,  3,-23, -7,-11;
/M: SY='P'; M= -6, -4,-26, -3,  7,-23,-19,-13,-19,  1,-21,-14, -5, 19,  1, -5,  4,  9,-17,-29,-18,  2;
/M: SY='C'; M=-10,-20,120,-30,-30,-20,-30,-30,-30,-30,-20,-20,-20,-40,-30,-30,-10,-10,-10,-50,-30,-30;
/M: SY='Q'; M= -9, -6,-27, -7, 10,-29,-22,  3,-11,  0, -7,  3, -5,-14, 40,  2, -4, -8,-20,-21, -8, 25;
/M: SY='A'; M=  6, -6,-22, -8,  0,-21, -9, -8,-18,  1,-17,-10, -2, -1, -1,  3,  3, -3,-14,-26,-16, -2;
/M: SY='W'; M=-20,-40,-50,-40,-30, 10,-20,-30,-20,-20,-20,-40,-30,-20,-20,-40,-30,-30,150, 30,-20;
/M: SY='N'; M=  3, 17,-18, 16,  2,-24, -2, -5,-23, -4,-26,-20, 18,-13, -2, -6, 16,  4,-17,-36,-20,  0;
/M: SY='S'; M= 15, -3,-13, -5, -1,-20, -4, -9,-17, -7,-22,-15,  3,-11,  0, -9, 25, 11, -9,-33,-18, -1;
/M: SY='L'; M=  5,-15,-20,-15,  3, -8,-20,-15,  0,-15, 18,  3,-18,-18, -8,-15,-13, -8, -3,-23,-10, -3;
/M: SY='T'; M= -2, -8,-18,-12, -5, -6,-20,-10, -5,-11, -7, -4, -6,-12, -6,-12,  6, 14, -3,-22, -2, -7;
/M: SY='P'; M= -7,-20,-34,-13, -3,-23,-21,-20,-12,-12,-18,-13,-20, 60,-11,-19,-10, -7,-18,-29,-24,-11;
/M: SY='H'; M=-16, -1,-28, -1, -2,-17,-20, 68,-22, -8,-14,  0,  6,-20,  5, -2, -9,-15,-23,-28, 14, -2;
/M: SY='R'; M= -6, -8,-26, -8,  1,-20,-15, -7,-19,  8,-15, -8, -5,-16,  8, 14, -4, -5,-15,-13, -9,  3;
/M: SY='H'; M=-16, -3,-31, -2,  1,-22,-20, 59,-26, -3,-19, -4,  4, -4,  8,  0,-10,-16,-27,-28,  6,  1;
/I:         I=-8; MI=-5; IM=-5; DM=-15; MD=-15;
/M: SY='S'; M= -6, -1,-24, -2,  1,-21,-11, -9,-15,  0,-18,-10,  3, -6,  2,  0,  4,  1,-13,-29,-15,  0;
/M: SY='Y'; M=-18,-18,-27,-22,-15, 22,-27,  1, -7, -2, -4, -1,-12,-24,-12,  6,-16,-10, -8,  1, 28,-15;
/M: SY='T'; M= -4, -3,-16,-10,-10,-10,-20, -8, -3,-13, -6, -5,  1,-15, -8,-11,  8, 21, -2,-28, -6,-10;
/M: SY='P'; M= -1,-15,-31,-10, -2,-23,-15,-18,-18,-10,-24,-16,-14, 52, -7,-17, -3, -5,-22,-27,-23, -7;
/M: SY='E'; M= -5,  2,-26,  5, 22,-23,-13,  1,-25,  0,-20,-16,  0, -9,  8, -2,  3, -2,-22,-19,-12, 15;
/M: SY='R'; M= -9,  4,-22, -3, -2,-11,-15, -4,-18,  7,-18,-12, 11,-15, -1, 12,  2,  5,-15,-23, -5, -3;
```

May 2004 / Text revised.

Technical section

PROSITE methods (with tools and information) covered by this documentation:

KRINGLE_2, PS50070; Kringle domain profile (MATRIX)

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 96
  - detected by PS50070: 95 (true positives)
  - undetected by PS50070: 1 (0 false negative and 1 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50070:
  NONE.
- Domain architecture view of Swiss-Prot proteins matching PS50070

- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50070
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50070
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50070
- View ligand binding statistics of PS50070
- Matching PDB structures: 1A0H 1B2I 1BHT 1CEA ... [ALL]

KRINGLE_1, PS00021; Kringle domain signature (PATTERN)

- Consensus pattern:
  [FY]-C-[RH]-[NS]-x(7,8)-[WY]-C
  The 2 C's are involved in a disulfide bonds
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 96
  - detected by PS00021: 94 (true positives)
  - undetected by PS00021: 2 (1 false negative and 1 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00021:
  3 false positives.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
  Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00021
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00021
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00021
- View ligand binding statistics of PS00021
- Matching PDB structures: 1A0H 1B2I 1BHT 1CEA ... [ALL]

References

PROSITE

prosite.expasy.org/cgi-bin/prosite/sequence_logo.cgi?ac=PS00021

Apps | Bookmarks | D pmartel | UALG Acad | Drug Design | Programming | Bioinformatics | Molecular Modelling | Science | Enzymology | To watch and read | Misc | GW2 | Other bookmarks

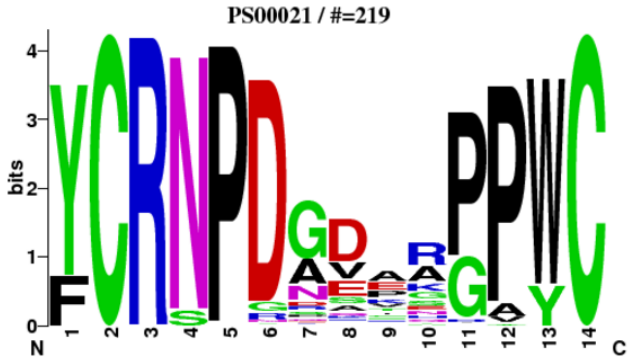
ExPASy Bioinformatics Resource Portal

PROSITE

Home | Contact

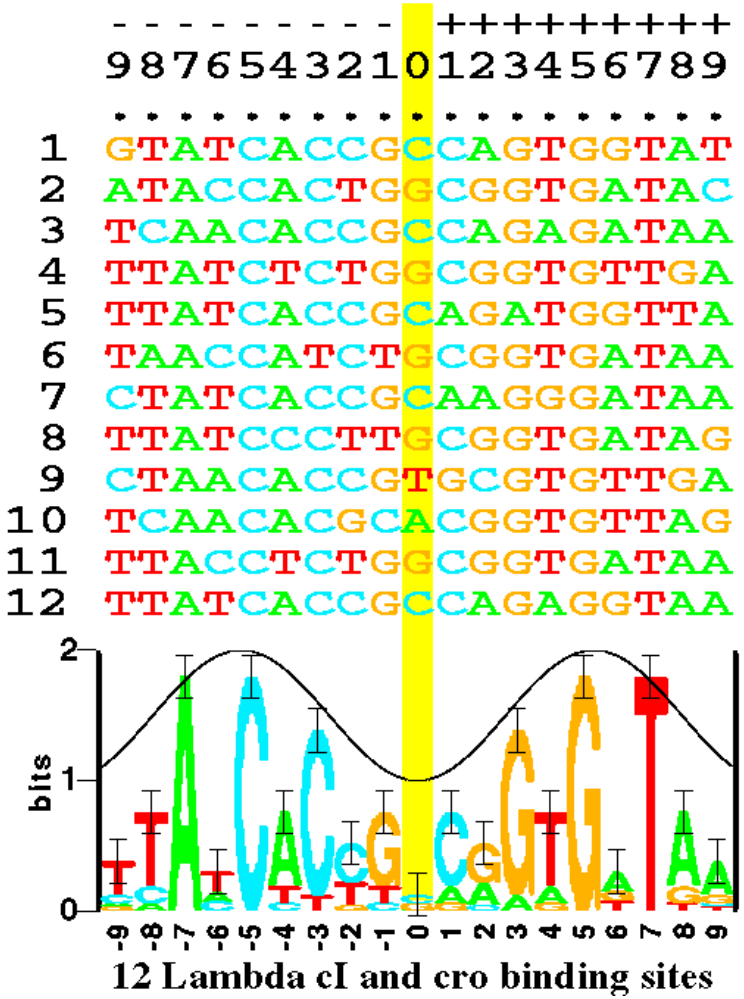
Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding

Sequence logo for PS00021



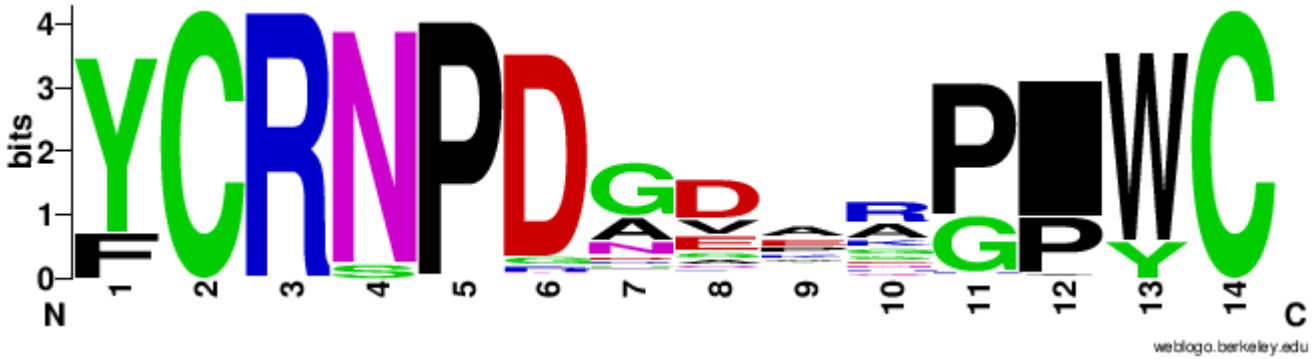
Number of UniProtKB/Swiss-Prot true positive hits used to build the logo:219.
Go to UniProtKB/Swiss-Prot true positive sequences.
Go to the list of all PROSITE motifs.
Go to the sequence logo help document.

[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C



http://weblogo.berkeley.edu/

Logo *versus* padrão PROSITE



[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C

ExPASy - PROSITE

prosite.expasy.org

Apps | Bookmarks | D pmartel | UALG Acad | Drug Design | Programming | Bioinformatics | Molecular Modelling | Science | Enzymology | To watch and read | Misc | GW2 | Other bookmarks

Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 20.131 of 27-Oct-2016 contains 1773 documentation entries, 1309 patterns, 1172 profiles and 1193 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Search

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] Examples

P00748

P00748 – Factor de coagulação F12

Scan | Clear

Exclude motifs with a high probability of occurrence from the scan

For more scanning options go to **ScanProsite**

Other tools

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

Profiles, PWM, PSWM, PS | ScanProsite | www.uniprot.org/uniprot | PROSITE | ScanProsite | Paulo

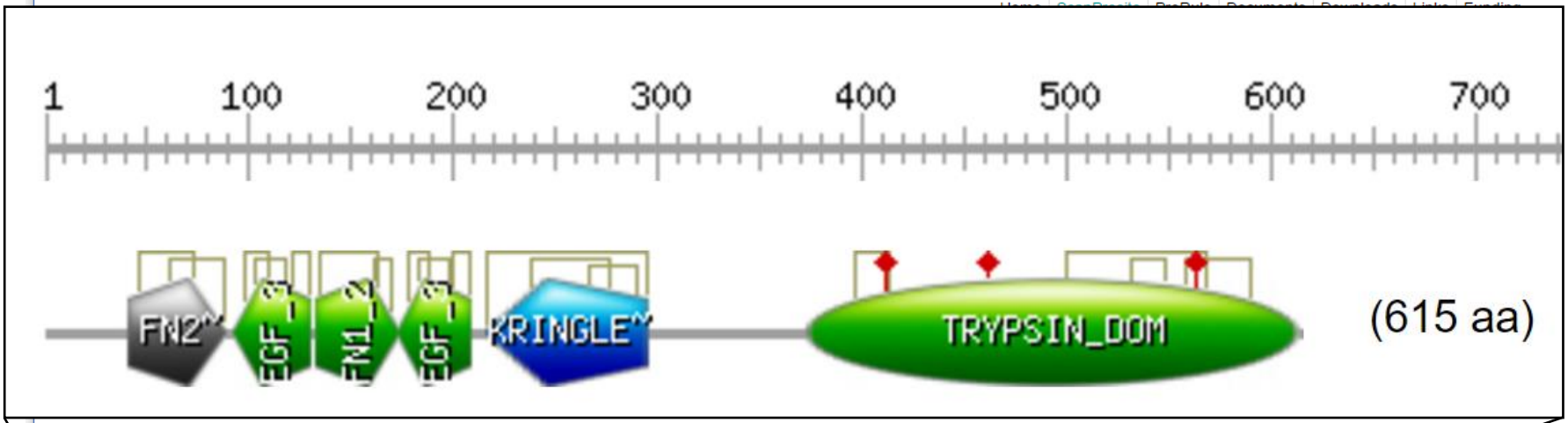
prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=156895430889.scan.gz

Apps | Bookmarks | D pmartel | UALG Acad | Drug Design | Programming | Bioinformatics | Molecular Modelling | Science | Enzymology | To watch and read | Misc | GW2 | Other bookmarks

ExPASy Bioinformatics Resource Portal

PROSITE

Home | Contact



Legend:

disulfide bridge | active site | other 'ranges' | other sites

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to http://prosite.expasy.org/mydomains/.

hits by profiles: [6 hits (by 5 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

ruler: 1 100 200 300 400 500 600 700 800 900 1000

sp-P00748-FA12_HUMAN (sp-P00748-FA12_HUMAN) (615 aa)

PS51092 **FN2_2** *Fibronectin type-II collagen-binding domain profile* :

42 - 90: **score** = 19.967

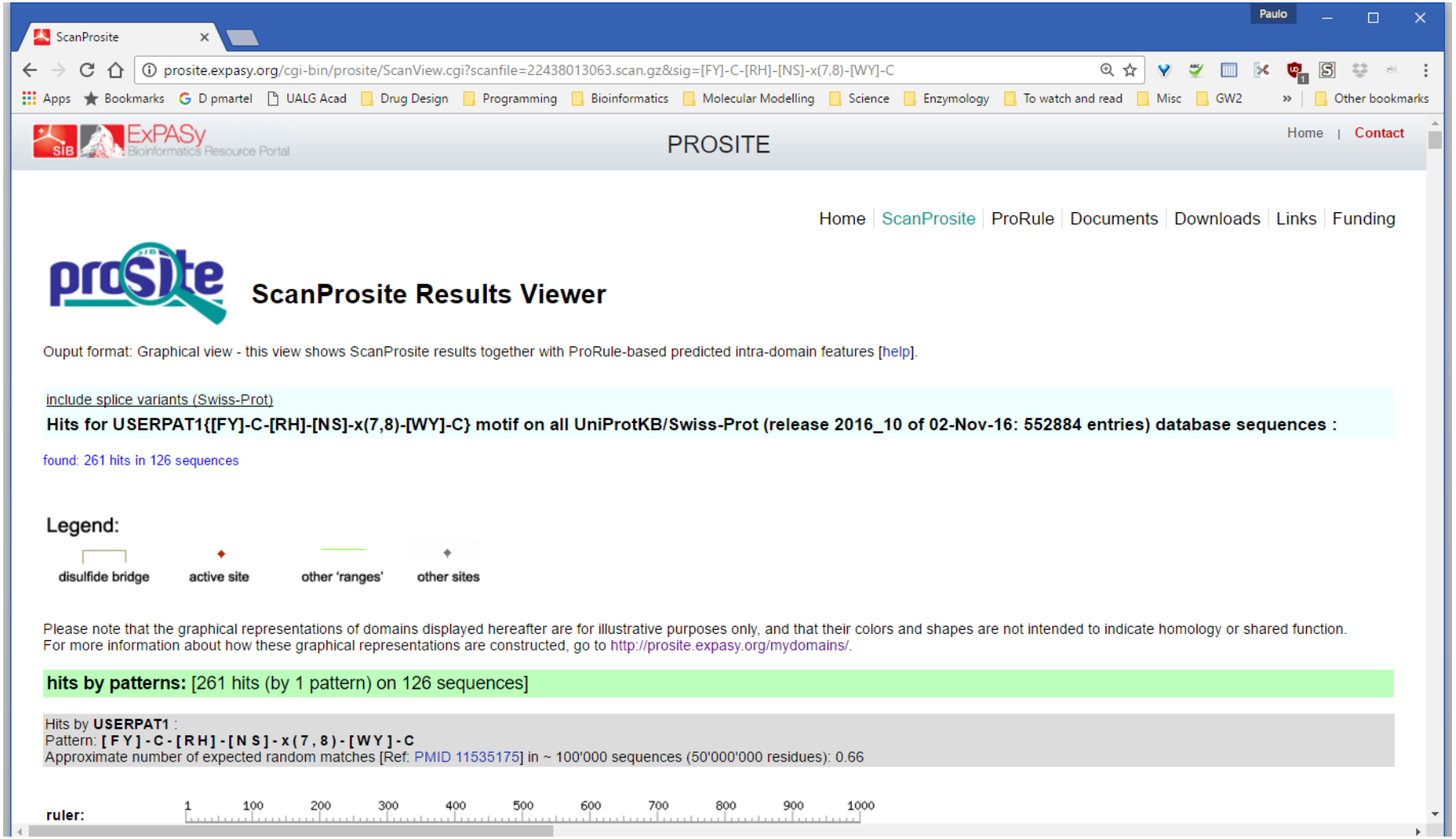
VTGEPCHFPFQYHRQLYHKCTHKGRPGPQPWCATTPNFDQDQRWGYCLE

Predicted features:

DOMAIN	42	90	Fibronectin type-II	[condition: none]
DISULFID	47	73		[condition: C-x*-C]

Pesquisa de motivos com Prosite

http://prosite.expasy.org/scanprosite/


ScanProsite
Paulo
prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=22438013063.scan.gz&sig=[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C
Apps
Bookmarks
D pmartel
UALG Acad
Drug Design
Programming
Bioinformatics
Molecular Modelling
Science
Enzymology
To watch and read
Misc
GW2
Other bookmarks
ExPASy
Bioinformatics Resource Portal
PROSITE
Home | Contact
Home ScanProsite ProRule Documents Downloads Links Funding
ScanProsite Results Viewer
Ouput format: Graphical view - this view shows ScanProsite results together with ProRule-based predicted intra-domain features [help].
include splice variants (Swiss-Prot)
Hits for USERPAT1{[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C} motif on all UniProtKB/Swiss-Prot (release 2016_10 of 02-Nov-16: 552884 entries) database sequences :
found: 261 hits in 126 sequences
Legend:
disulfide bridge
active site
other 'ranges'
other sites
Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to http://prosite.expasy.org/mydomains/.
hits by patterns: [261 hits (by 1 pattern) on 126 sequences]
Hits by USERPAT1 :
Pattern: [FY]-C-[RH]-[NS]-x(7,8)-[WY]-C
Approximate number of expected random matches [Ref: PMID 11535175] in ~ 100'000 sequences (50'000'000 residues): 0.66
ruler:
1 100 200 300 400 500 600 700 800 900 1000

ScanProsite

prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=22438013063.scan.gz&sig=[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to http://prosite.expasy.org/mydomains/.

hits by patterns: [261 hits (by 1 pattern) on 126 sequences]

Hits by **USERPAT1** :
Pattern: **[FY]-C-[RH]-[NS]-x(7,8)-[WY]-C**
Approximate number of expected random matches [Ref: PMID 11535175] in ~ 100'000 sequences (50'000'000 residues): 0.66

ruler: 1 100 200 300 400 500 600 700 800 900 1000

P08519
(APOA_HUMAN)

Apolipoprotein(a) (Apo(a)) (Lp(a)) (EC 3.4.21.-). *Homo sapiens (Human)*

76 - 88:	YCRNpdavaap.YC
190 - 202:	YCRNpdavaap.YC
304 - 316:	YCRNpdavaap.YC
418 - 430:	YCRNpdavaap.YC
532 - 544:	YCRNpdavaap.YC
646 - 658:	YCRNpdavaap.YC
760 - 772:	YCRNpdavaap.YC
874 - 886:	YCRNpdavaap.YC
988 - 1000:	YCRNpdavaap.YC
1102 - 1114:	YCRNpdavaap.YC
1216 - 1228:	YCRNpdavaap.YC

Exemplo de entrada na base PROSITE (motivo)

```
ID   CUTINASE_1; PATTERN.
AC   PS00155;
DT   APR-1990 (CREATED); NOV-1997 (DATA UPDATE); MAR-2005 (INFO UPDATE).
DE   Cutinase, serine active site.
PA   P-x-[STA]-x-[LIV]-[IVT]-x-[GS]-G-Y-S-[QL]-G.
NR   /RELEASE=46.4,178022;
NR   /TOTAL=20(20); /POSITIVE=20(20); /UNKNOWN=0(0); /FALSE_POS=0(0);
NR   /FALSE_NEG=0; /PARTIAL=0;
CC   /TAXO-RANGE=??EP?; /MAX-REPEAT=1;
CC   /SITE=11,active_site;
DR   P63880, CUT1_MYCBO , T; P63879, CUT1_MYCTU , T; P63882, CUT2_MYCBO , T;
DR   P63881, CUT2_MYCTU , T; P0A537, CUT3_MYCBO , T; P0A536, CUT3_MYCTU , T;
DR   P00590, CUTI1_FUSSO, T; Q96UT0, CUTI2_FUSSO, T; Q96US9, CUTI3_FUSSO, T;
DR   P41744, CUTI_ALTBR , T; P29292, CUTI_ASCRA , T; P52956, CUTI_ASPOR , T;
DR   Q00298, CUTI_BOTCI , T; P10951, CUTI_COLCA , T; P11373, CUTI_COLGL , T;
DR   Q8X1P1, CUTI_ERYGR , T; Q99174, CUTI_FUSSC , T; P30272, CUTI_MAGGR , T;
DR   Q8TGB8, CUTI_MONFR , T; Q9Y7G8, CUTI_PYRBR , T;
3D   1AGY; 1CEX; 1CUA; 1CUB; 1CUC; 1CUD; 1CUE; 1CUF; 1CUG; 1CUH; 1CUS; 1CUU;
3D   1CUV; 1CUW; 1CUY; 1CUZ; 1FFA; 1FFB; 1FFC; 1FFD; 1FFE; 1OXM; 1XZA; 1XZB;
3D   1XZC; 1XZD; 1XZE; 1XZF; 1XZG; 1XZH; 1XZJ; 1XZK; 1XZL; 1XZM; 2CUT;
DO   PDOC00140;
//
```

http://www.expasy.org/prosite

Exemplo de entrada na base PROSITE (perfil)

```
ID   HSP20; MATRIX.
AC   PS01031;
DT   JUN-1994 (CREATED); DEC-2001 (DATA UPDATE); MAR-2005 (INFO UPDATE).
DE   Heat shock hsp20 proteins family profile.
MA   /GENERAL_SPEC: ALPHABET='ABCDEFGHIKLMNPQRSTVWYZ'; LENGTH=88;
MA   /DISJOINT: DEFINITION=PROTECT; N1=6; N2=83;
MA   /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=-0.7971325; R2=0.0157729; TEXT='-LogE';
MA   /CUT_OFF: LEVEL=0; SCORE=590; N_SCORE=8.5; MODE=1; TEXT='!';
MA   /CUT_OFF: LEVEL=-1; SCORE=463; N_SCORE=6.5; MODE=1; TEXT='?';
MA   /DEFAULT: M0=-8; D=-20; I=-20; B1=-50; E1=-50; MI=-105; MD=-105; IM=-105; DM=-105;
MA   /I: B1=0; BI=-105; BD=-105;
MA   /M: SY='D'; M=-10,26,-29,38,34,-34,-14,-2,-33,7,-24,-23,8,-6,8,-4,0,-9,-27,-33,-19,21;
MA   /M: SY='I'; M=-8,-31,-23,-35,-28,7,-32,-27,27,-24,15,13,-27,-26,-24,-23,-20,-9,25,-4,2,-27;
MA   /M: SY='R'; M=-11,-12,-26,-12,-1,-13,-23,-1,-8,1,-7,-3,-8,-11,-2,8,-9,-6,-8,-22,-3,-4;
MA   /M: SY='E'; M=-11,17,-27,23,29,-24,-15,-3,-27,1,-22,-20,9,-1,6,-6,3,-4,-25,-32,-17,17;
MA   /M: SY='D'; M=-7,10,-23,11,2,-25,0,-6,-26,-4,-23,-18,7,-6,-5,-8,7,7,-20,-31,-17,-2;
MA   /I: I=-4; MD=-22;
MA   /M: SY='D'; M=-8,17,-27,25,19,-30,-13,-5,-28,6,-25,-20,7,3,4,-1,0,-7,-24,-30,-19,10; D=-4;
MA   /I: I=-4; MI=0; MD=-22; IM=0; DM=-22;
MA   /M: SY='D'; M=-11,20,-25,24,16,-29,-12,-1,-27,14,-25,-16,14,-9,10,5,1,-6,-23,-28,-14,13; D=-4;
MA   /I: I=-4; DM=-22;
..
... Some lines omitted..
..
MA   /M: SY='K'; M=-9,-5,-25,-6,0,-22,-21,-12,-17,30,-21,-6,-3,-16,1,23,-9,-7,-6,-23,-11,0;
MA   /I: E1=0; IE=-105; DE=-105;
NR   /RELEASE=46.4,178022;
NR   /TOTAL=195(194); /POSITIVE=190(189); /UNKNOWN=5(5); /FALSE_POS=0(0);
NR   /FALSE_NEG=1; /PARTIAL=8;
CC   /MATRIX_TYPE=protein_domain;
CC   /SCALING_DB=reversed;
CC   /AUTHOR=P_Bucher;
CC   /TAXO-RANGE=A?EP?; /MAX-REPEAT=2;
CC   /FT_KEY=DOMAIN; /FT_DESC=HSP20;
DR   P0A5B8, 14KD_MYCBO , T; P0A5B7, 14KD_MYCTU , T; P46729, 18K1_MYCAV , T;
DR   P46730, 18K1_MYCIT , T; P46731, 18K2_MYCAV , T; P46732, 18K2_MYCIT , T;
DR   P12809, 18KD_MYCLE , T; P80485, ASP1_STRTR , T; O30851, ASP2_STRTR , T;
..
... Some lines omitted..
```

Perfis (profiles)

Um **perfil** é uma descrição do padrão subjacente a um alinhamento múltiplo e reflecte a probabilidade de ocorrência de cada tipo de resíduo numa dada posição. Tem várias aplicações:

• Permite uma maior precisão no alinhamento de sequências distantes da mesma família

• Os padrões emergentes são úteis para a **classificação** de sub-famílias dentro de um conjunto de sequências homólogas.

• O alinhamento de uma sequência a um perfil é geralmente mais fiável e melhora o processo de **modelação estrutural** por homologia

• Os perfis permite **pesquisas de elevada sensibilidade** para a detecção de parentes distantes de uma dada família de proteínas

O alinhamento de uma sequência a um perfil é condicionado pela sua natureza e pelo seu grau de conservação. Assim, resíduos altamente conservados no perfil terão um score mais alto, e resíduos pouco conservados um score mais baixo. Este processo impõe uma tendência para alinhar em primeiro lugar as *zonas mais conservadas*.

Geração de perfis a partir de alinhamentos múltiplos

```
Q3IC08|DNAK_PSEHT    ( 358) GKEPRKDVNPDEAVAVGAAIQGGVLAGD
Q3KIA0|DNAK_PSEPF    ( 358) GKEARKDVNPDEAVAMGAAIQGAVLAGD
Q4ZNP7|DNAK_PSEU2    ( 358) GKEARKDVNPDEAVAMGAAIQGAVLAGD
Q4FPS9|DNAK_PSYAR    ( 357) GQEPRKDVNPDEAVAAGAAIQGAVLSGE
Q46XI7|DNAK_RALEJ    ( 359) GKEARKDVNPDEAVAVGAAIQGSVLSGD
Q3IYM7|DNAK_RHOS4    ( 354) GKEPHKGVNPDEVVALGAAIQAGVLQGD
Q4UJK7|DNAK_RICFE    ( 352) GREPHKGVNPDEVVALGAAIQGGVLNKE
Q57TP3|DNAK_SALCH    ( 358) GKEPRKDVNPDEAVAIGAAVQGGVLTGD
Q5PDJ5|DNAK_SALPA    ( 358) GKEPRKDVNPDEAVAIGAAVQGGVLTGD
Q326K7|DNAK_SHIBS    ( 358) GKEPRKDVNPDEAVAIGAAVQGGVLTGD
Q32KA5|DNAK_SHIDS    ( 358) GKEPRKDVNPDEAVAIGAAVQGGVLTGD
Q3Z601|DNAK_SHISS    ( 358) GKEPRKDVNPDEAVAIGAAVQGGVLTGD
Q5LWJ6|DNAK_SILPO    ( 353) GKEPHKGVNPDEVVAMGAAIQAGVLQGD
Q5HFI0|DNAK_STAAC    ( 328) GKEPNKGVNPDEVVAMGAAIQGGVITGD
Q5HNW6|DNAK_STAEQ    ( 328) GKEPHKGVNPDEVVAMGAAIQAGVITGD
Q4L6T0|DNAK_STAHJ    ( 328) GKDPHKGVNPDEVVAMGAAIQGGVITGD
Q49Y22|DNAK_STAS1    ( 328) GKDPHKGVNPDEVVAMGAAIQGGVITGD
Q3K3T2|DNAK_STRA1    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVITGD
P0A3J3|DNAK_STRA5    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVITGD
P0A3J4|DNAK_STRAG    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVITGD
P0C0C6|DNAK_STRP1    ( 327) GKEPNKSVNPDEVVAMGAAIQGGVITGD
P68837|DNAK_STRP8    ( 327) GKEPNKSVNPDEVVAMGAAIQGGVITGD
Q48RR3|DNAK_STRPM    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVITGD
Q5M1T8|DNAK_STRT1    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVISGD
Q5M6D1|DNAK_STRT2    ( 328) GKEPNKSVNPDEVVAMGAAIQGGVISGD
Q47TI0|DNAK_THEFY    ( 330) GKEPNKGVNPDEVVAVGAALQAGVLKGD
```

Cons	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z	Gap	Len
I	8	3	-2	5	4	5	5	-4	24	0	15	13	1	1	1	-7	2	22	21	-18	-6	4	100	100
T	13	19	-5	24	18	-18	19	7	1	7	-7	-4	14	11	10	-1	9	29	3	-28	-14	15	100	100
L	5	5	-5	3	4	13	4	2	8	-4	14	12	8	-5	0	-10	0	10	10	-1	5	2	22	22
S	17	14	17	13	10	-12	29	-5	-5	6	-14	-9	12	10	0	-2	34	19	1	-8	-15	4	100	100
T	15	3	22	0	-1	-5	12	-2	7	-3	-8	-6	5	7	-8	-7	16	29	9	-22	6	-4	100	100
T	8	-1	12	-2	0	5	6	-4	19	-4	8	5	-1	2	-8	-8	7	22	19	-15	4	-3	100	100
C	17	0	24	-1	-3	11	8	-1	7	-10	1	-2	1	-3	-8	-14	8	5	9	-5	14	-7	100	100
V	11	0	18	-1	-2	2	14	-10	26	-4	9	7	-3	7	-7	-7	21	10	31	-19	-5	-5	100	100
C	10	-8	15	-11	-11	6	8	-7	11	-10	4	3	-7	0	-11	-4	11	5	15	-22	14	-11	100	100
V	7	7	-3	8	8	-3	11	1	20	-1	14	10	4	2	8	-5	0	5	26	-24	-6	8	100	100

PIR
Protein Information Resource
A UniProt CONSORTIUM MEMBER

About PIR | Databases | Search/Analysis | Download | Support

HOME / Search / *Pattern Search*

http://pir.georgetown.edu/pirwww/search/pattern.shtml

Pattern Search Forms ?

Search a query pattern against a UniProt database

1. Select a database: (•) **UniProtKB** (or restricted by organism/taxon group)
( ) **UniRef100**

2. Insert a user-defined pattern below:

x(12)-E-x(3)-E-x-C-x(6)-[DEN]-x-[LIVMFY]-x(9)-[FYW]

Or, alternatively, enter a valid PROSITE code for a query pattern:

Submit Reset

Example: PS00888 (annotated output)

Search your query sequence against the PROSITE database

Insert a query sequence below using the single letter amino acid code:

Or, alternatively, enter a UniProtKB identifier:

Submit Reset

Example: O05689 (annotated output)

Matrizes PSSM

- PSSM (position specific scoring matrix) é um tipo de matriz cujos scores por aminoácido são dependentes da **posição do aminoácido** na sequência.

- As matrizes PSSM são semelhantes a **perfis** e permitem armazenar informação para uma determinada **assinatura** ou motivo de uma família de proteínas.

- Os valores de uma matriz PSSM são calculados a partir regiões conservadas de alinhamentos de sequências.

- As matrizes PSSM permitem detectar padrões e similaridades fracas

Criação de uma matriz PSSM

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0	0	0	0	0	0	0	0	0	0	0	2	1	0	2
C	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	5	0	1	0	0	0	1	0	0	0	0	1	0
F	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
G	5	0	0	2	0	5	1	0	1	0	2	3	1	1	0
H	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
K	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0
M	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0
S	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
T	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
V	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0

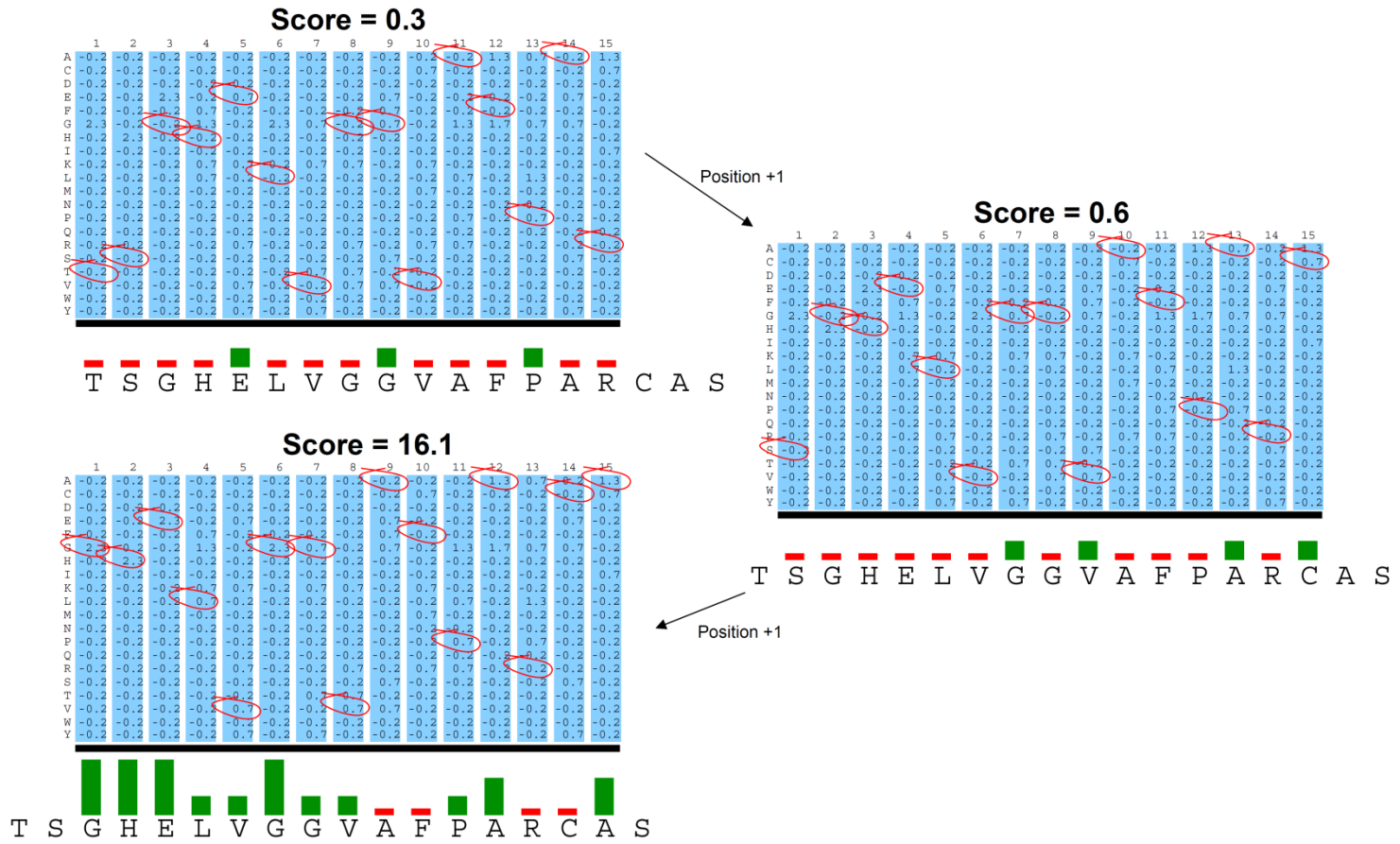
As frequências de ocorrência de cada aminoácido são contadas para cada **posição** do alinhamento. As contagens são normalizadas e convertidas numa matriz log odds semelhante a uma matriz de score.

Criação de uma matriz PSSM

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	1.3	0.7	-0.2	1.3
C	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7
D	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
E	-0.2	-0.2	2.3	-0.2	0.7	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2
F	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
G	2.3	-0.2	-0.2	1.3	-0.2	2.3	0.7	-0.2	0.7	-0.2	1.3	1.7	0.7	0.7	-0.2
H	-0.2	2.3	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
I	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7
K	-0.2	-0.2	-0.2	0.7	0.7	-0.2	0.7	0.7	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2
L	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	1.3	-0.2	-0.2
M	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2
N	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
P	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	0.7	-0.2	-0.2
Q	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
R	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	0.7	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2
S	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7	-0.2
T	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
V	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	0.7	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
W	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
Y	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2

Matriz PSSM calculada a partir do exemplo do slide anterior.

Uso de matriz PSSM



A matriz PSSM “desliza” sobre a sequência em análise, produzindo um score para cada posição. O score máximo indicará a posição do padrão codificado pela matriz PSSM.

PSI-BLAST

PSI-BLAST

- O variant PSI do programa BLAST combina um model PSSM com uma esquema de penalidades afins para “gaps
- Princípio do algoritmo:
  1. Fazer uma pesquisa BLAST standard contra uma base de dados usando uma matriz de alinhamento (por exemplo BLOSUM)
  2. Um modelo PSSM é construído automaticamente a partir do alinhamento múltiplo dos “hits” de score mais elevado
  3. O modelo PSSM é usado em vez da matriz de score original para realizar uma nova pesquisa BLAST
  4. Novas sequências obtidas em 3. são usadas para construir um novo alinhamento múltiplo, e a partir destes uma nova matriz PSSM
  5. Os passos 3. e 4. são repetidos, sendo em cada repetição adicionadas novas sequêncais à lista de hits
  6. Considera-se que o algoritmo convergiu quando aparecerem sequências novas

Bases de dados de motivos e domínios

- **PROSITE**
  http://www.expasy.ch/prosite
- **PRINTS**
  http://www.bioinf.man.ac.uk/dbbrowser/PRINTS
- **PFAM**
  http://www.sanger.ac.uk/Software/Pfam
- **PRODOM**
  http://prodom.prabi.fr
- **SMART**
  http://smart.embl-heidelberg.de

Pesquisa em múltiplas bases de dados:

INTERPRO
http://www.ebi.ac.uk/interpro