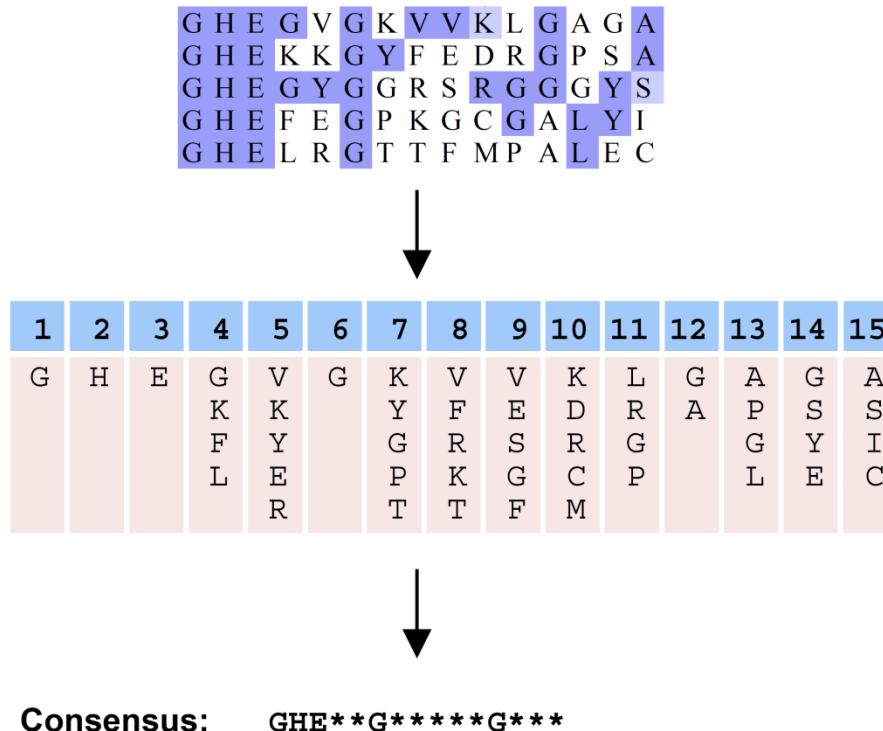


Motivos e perfis

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.

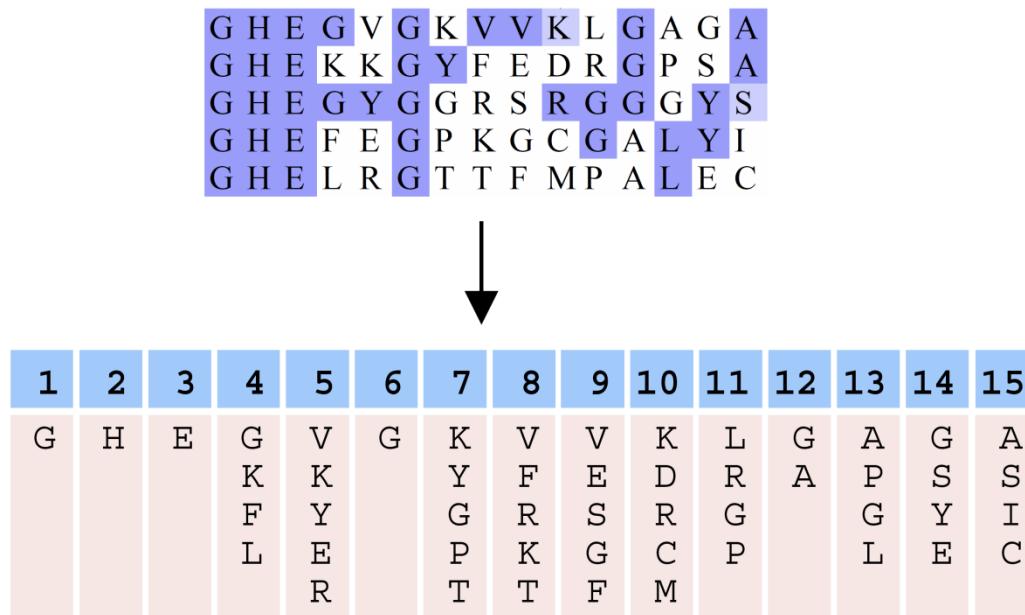
Exemplo:



Sequência de consenso: posições 100% conservadas no alinhamento.

Protocolo de alinhamentos múltiplos

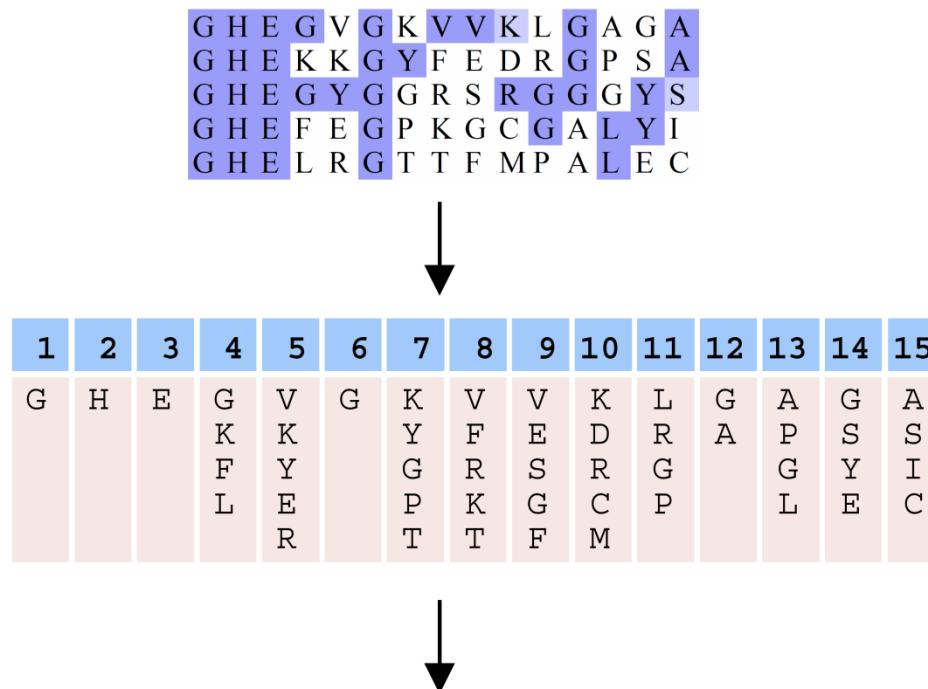
A sequência de consenso não permite descrever preferências não-integrais (<100%).



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**.



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

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 aminoácidos (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:

$\langle A-x-[ST](2)-x(0,1)-\{V\} \rangle$

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 proteicos
- Apesar do elevado número de proteínas conhecidas, a maioria pode ser agrupada num número limitado de famílias, com base na similaridade
- 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, PE 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

prosite 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

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

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, Po × ScanPosite × 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

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 ScanPosite

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 [ScanPosite](#)

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

Profiles, PWM, PSWM, PS... ScanProsite www.uniprot.org/unipro... Search in prosite for: kringle

prosite.expasy.org/cgi-bin/prosite/prosite-search-fu?SEARCH=kringle

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

ExPASY Bioinformatics Resource Portal PROSITE Home Contact



Prosite search results

Search in PROSITE for: kringle

(Release 20.131, of 27-Oct-2016)

Enter search terms:

kringle

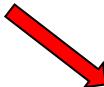
Prefix and append wildcard '*' to words.

[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, PS X ScanProsite www.uniprot.org/uniprot PROSITE

prosite.expasy.org/PDOC00020

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

Home Contact

PROSITE

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

prosite 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 involved in disulfide bonds.

Expert(s) to contact by email:
Ikeo K.

Last update:
May 2004 / Text revised.

Technical section

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

prosite.expasy.org/PDOC00020

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

Description Technical section References Copyright Miscellaneous

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

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

prosite.expasy.org/PS00021

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

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

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

prosite.expasy.org/PS50070

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

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]	<pre> /GENERAL_SPEC: ALPHABET='ABCDEFGHIJKLMNPQRSTVWYZ'; 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; 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; B2=-105; B3=-105; /M: SY=D'; M=-15, 29, -30, 44, 37, -36, -15, 1, -24, 10, -6, 13, -4, 0, -10, -30, -34, -19, 25; /M: SY=C'; M=-10, -20, 120, -30, -20, -30, -30, -30, -20, -20, -20, -40, -30, -30, -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= -9, -10, -30, -10, -20, -30, 70, -20, -40, -20, -30, -20, 0, -20, -20, -20, 0, -20, -30, -20, -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, -20, -30, -10, -10, -20, -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= -10, -30, -10, -20, -30, 70, -20, -40, -20, -30, -20, 0, -20, -20, -20, 0, -20, -30, -20, -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=V'; 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= -10, -30, -10, -19, -30, -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, -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, -40, -30, -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; /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, -18, -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, -18, -18, -10, -24, -16, 14, 12, -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; </pre>

Matching PDB structures: 1A04 1B21 1B71 1CEA 1AII

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

prosite.expsy.org/PDOC00020

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

Description Technical section References Copyright Miscellaneous

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.expsy.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

SIB ExPASy Bioinformatics Resource Portal PROSITE Home Contact

Home ScanProsite ProRule Documents Downloads Links Funding

Sequence logo for PS00021

PS00021 / #=219

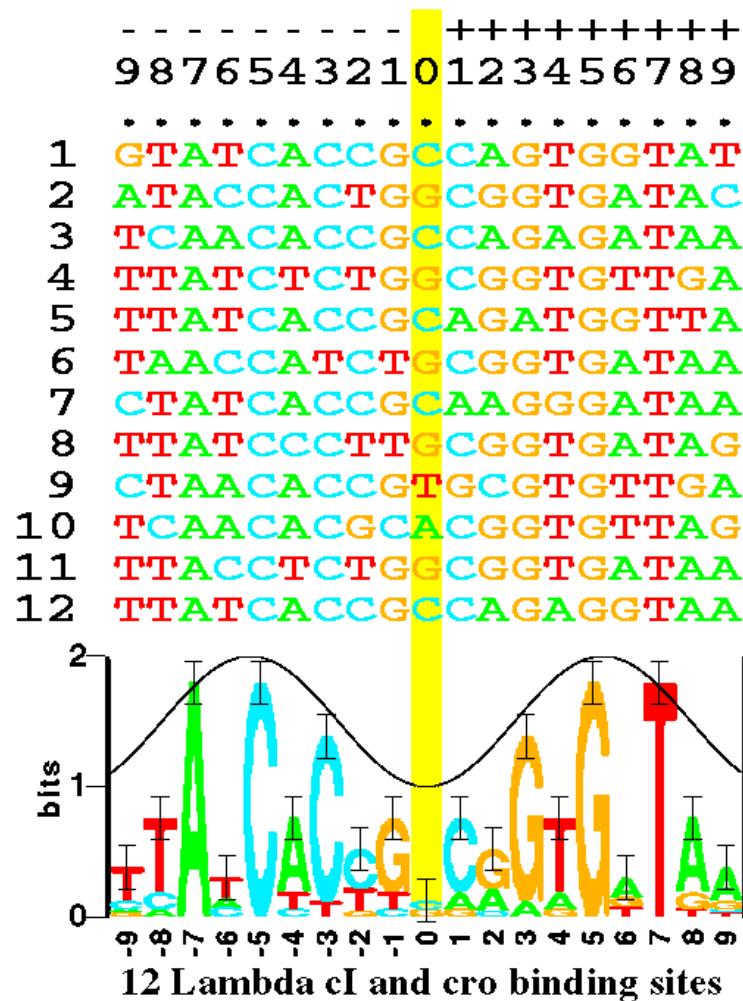
bits

0 1 2 3 4

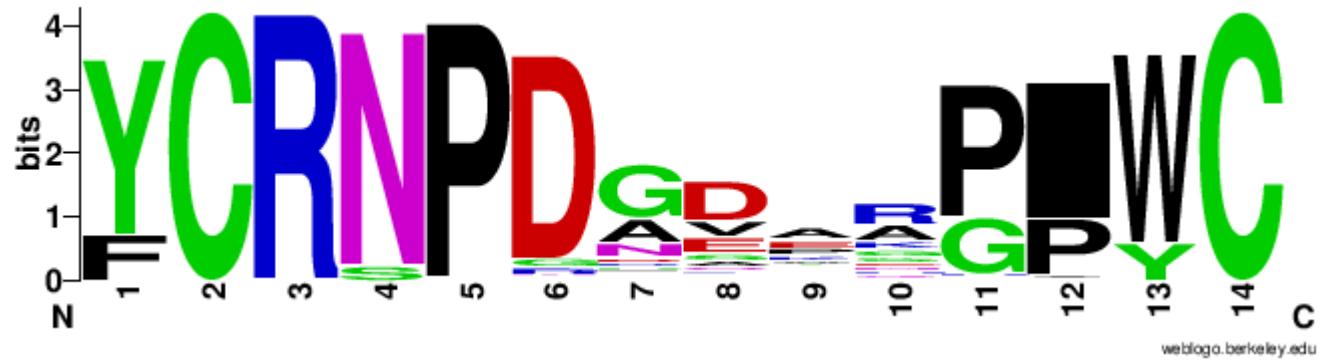
N 1 2 3 4 5 6 7 8 9 10 11 12 13 14 C

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



Logo versus padrão PROSITE



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

ExPASy - PROSITE X

prosite.expasy.org

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

PROSITE 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

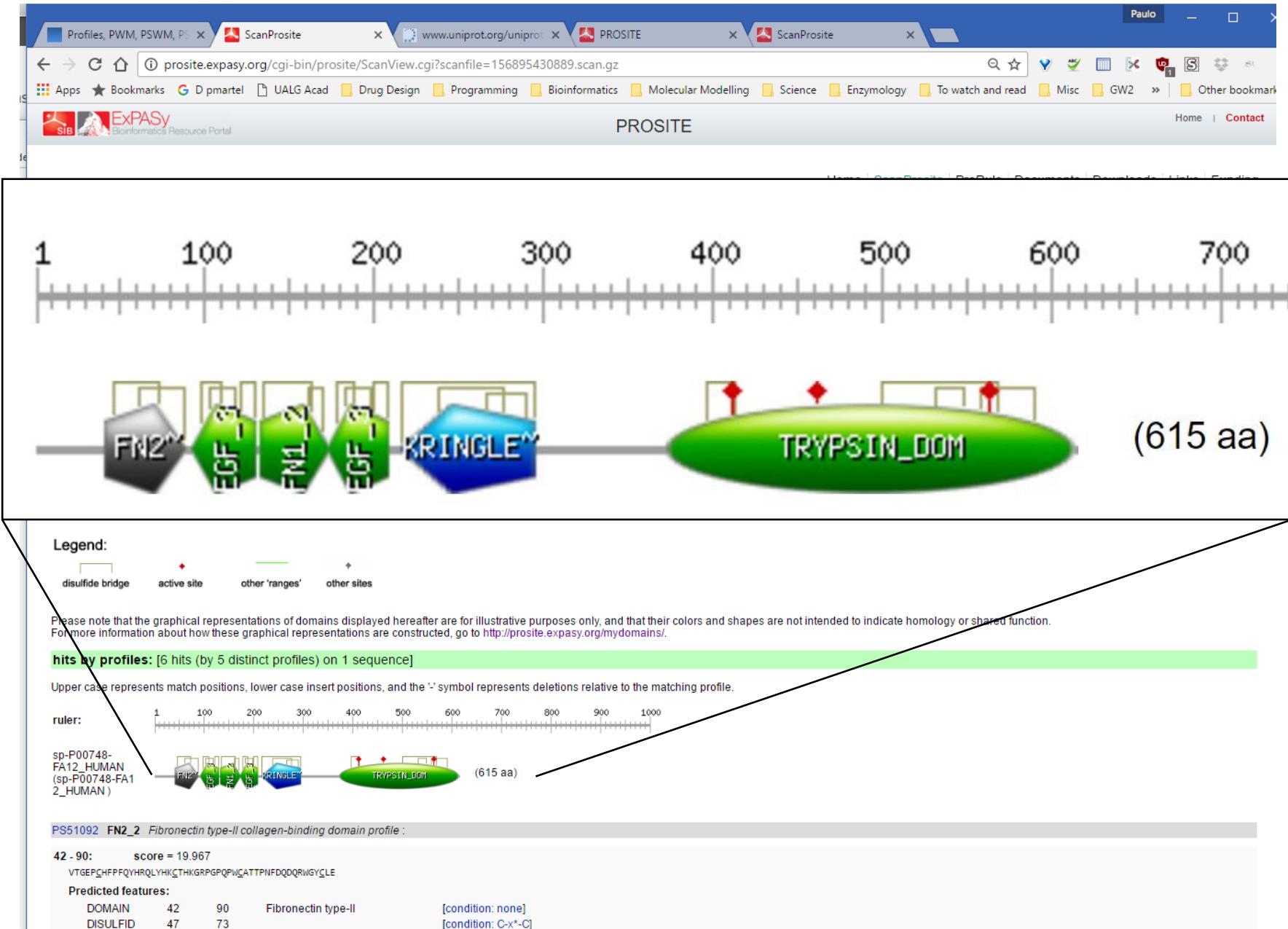
Exclude motifs with a high probability of occurrence from the scan

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.

A diagram illustrating a process flow: 'Custom' leads to 'Images', which then leads to 'Of', which finally leads to 'DOMAINS'. Arrows connect the boxes, and a small red arrow points upwards from the 'Of' box towards the 'DOMAINS' box.



Pesquisa de motivos com Prosite

ScanProsite

proSite ScanProsite tool

This form allows you to scan proteins for matches against the PROSITE collection of motifs as well as against your own patterns.

Option 1 - Submit PROTEIN sequences to scan them against the PROSITE collection of motifs.
 Option 2 - Submit MOTIFS to scan them against a PROTEIN sequence database.
 Option 3 - Submit PROTEIN sequences and MOTIFS to scan them against each other.

STEP 1 - Enter a MOTIF or a combination of MOTIFS [Examples](#) [\[help\]](#)

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

Supported input:

- A PROSITE accession e.g. PS50240 or identifier e.g. TRYPSIN_DOM
- Your own pattern e.g. P-x(2)-G-E-S-G(2)-[AS]

[» More](#)

[» Options](#) [\[help\]](#)

STEP2 - Select a PROTEIN sequence database [\[help\]](#)

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

ScanProsite

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

PROSITE

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

prosite ScanProsite Results Viewer

Output 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:

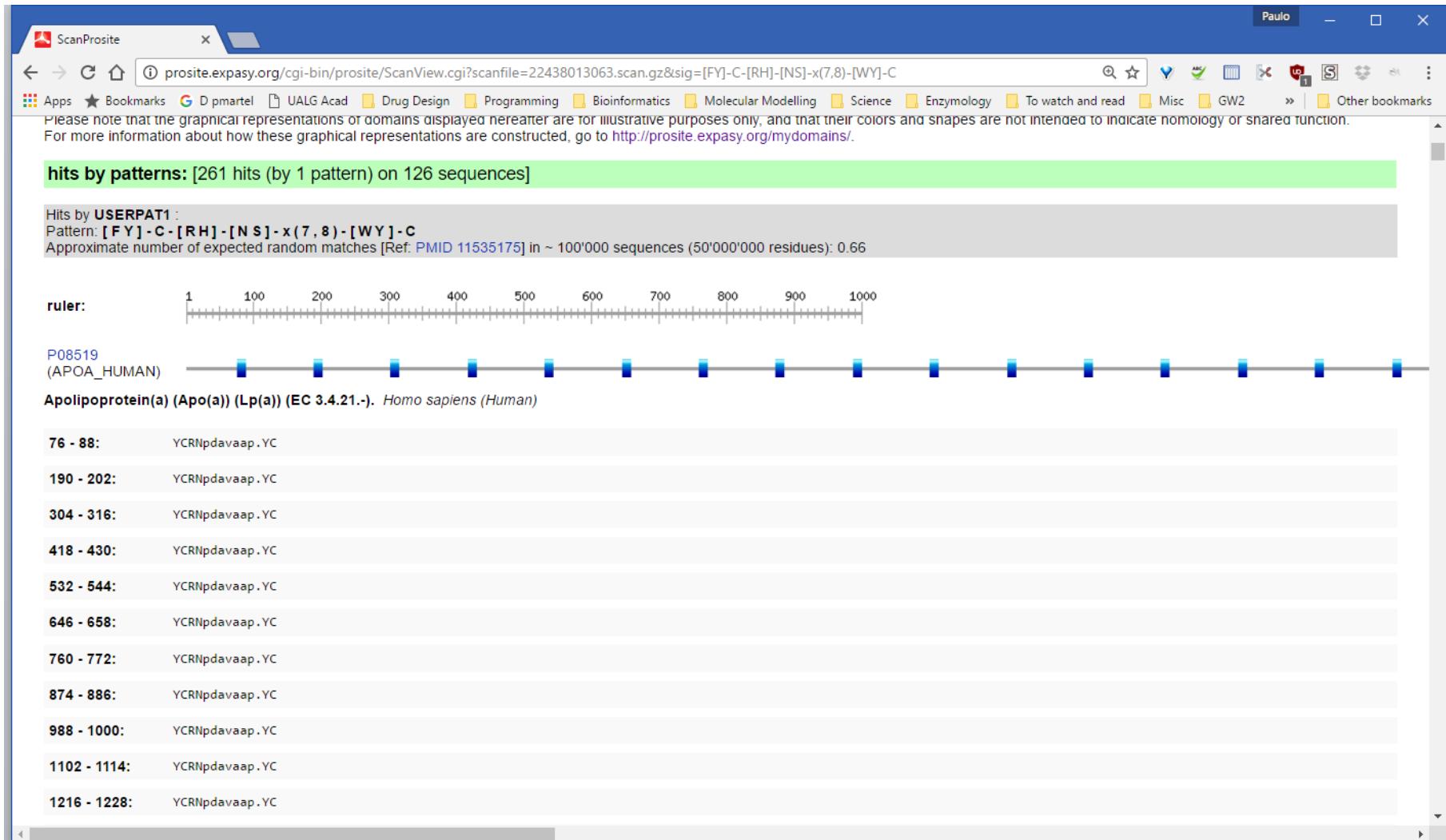
- disulfide bridge
- active site
- other 'ranges'
- other sites

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: **[F Y]-C-[R H]-[N S]-x(7,8)-[W Y]-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



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;
//
```

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='ABCDEFGHIJKLMNPQRSTVWXYZ'; 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 permitem **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) GKEPRKDVNPD PDEAVAVGAAIQGGVLAGD
Q3KIA0 DNAK PSEPF	(358) GKEARKDVNP PDEAVAMGAAIQGAVLAGD
Q4ZNP7 DNAK PSEU2	(358) GKEARKDVNP PDEAVAMGAAIQGAVLAGD
Q4FPS9 DNAK PSYAR	(357) GQEPRKDVNPD EAVAAAGAAIQGAVLSGE
Q46X17 DNAK RALEJ	(359) GKEARKDVNP DEAVAVGAAIQGSVLSGD
Q3IYM7 DNAK RHOS4	(354) GKEPHKGVN PDEVVALGAAIQAGVLQGD
Q4UJK7 DNAK RICFE	(352) GREGPHKGVN PDEVVALGAAIQGGVLNKE
Q57TP3 DNAK SALCH	(358) GKEPRKDVNPD EAVAIAGAAVQGGVLITGD
Q5PDJ5 DNAK SALPA	(358) GKEPRKDVNPD EAVAIAGAAVQGGVLITGD
Q326K7 DNAK SHIBS	(358) GKEPRKDVNPD EAVAIAGAAVQGGVLITGD
Q32KA5 DNAK SHIDS	(358) GKEPRKDVNPD EAVAIAGAAVQGGVLITGD
Q3Z601 DNAK SHISS	(358) GKEPRKDVNPD EAVAIAGAAVQGGVLITGD
Q5LWJ6 DNAK SILPO	(353) GKEPHKGVN PDEVVAMGAAIQAGVLQGD
Q5HFI0 DNAK STAAC	(328) GKEPNKGVN PDEVVAMGAAIQGGVITGD
Q5HNW6 DNAK STAEQ	(328) GKEPHKGVN PDEVVAMGAAIQAGVITGD
Q4L6T0 DNAK STAHAJ	(328) GKDPHKGVN PDEVVAMGAAIQGGVITGD
Q49Y22 DNAK STAS1	(328) GKDPHKGVN PDEVVAMGAAIQGGVITGD
Q3K3T2 DNAK STRA1	(328) GKEPNKSVN PDEVVAMGAAIQGGVITGD
P0A3J3 DNAK STRAS	(328) GKEPNKSVN PDEVVAMGAAIQGGVITGD
P0A3J4 DNAK STRAG	(328) GKEPNKSVN PDEVVAMGAAIQGGVITGD
P0C0C6 DNAK STRP1	(327) GKEPNKSVN PDEVVAMGAAIQGGVITGD
P68837 DNAK STRP8	(327) GKEPNKSVN PDEVVAMGAAIQGGVITGD
Q48RR3 DNAK STRPM	(328) GKEPNKSVN PDEVVAMGAAIQGGVITGD
Q5M1T8 DNAK STRT1	(328) GKEPNKSVN PDEVVAMGAAIQGGVISGD
Q5M6D1 DNAK STRT2	(328) GKEPNKSVN PDEVVAMGAAIQGGVISGD
Q47TI0 DNAK THEFY	(330) GKEPNKGVN PDEVVAVGAAIQAGVLKGD

	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	<u>24</u>	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	<u>29</u>	3	-28	-14	15	100	100
L		5	5	-5	3	4	13	4	2	8	-4	<u>14</u>	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	<u>34</u>	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	<u>29</u>	9	-22	6	-4	100	100
T		8	-1	12	-2	0	5	6	-4	19	-4	8	5	-1	2	-8	-8	7	<u>22</u>	19	-15	4	-3	100	100
C		17	0	<u>24</u>	-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	<u>31</u>	-19	-5	-5	100	100
C		10	-8	<u>15</u>	-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	<u>26</u>	-24	-6	8	100	100

Q3IC08 | DNAK PSEHT
Q3KIA0 | DNAK PSEPF
Q4ZNP7 | DNAK PSEU2
Q4FPS9 | DNAK PSYAR
Q46X17 | DNAK RALEJ
Q3IYM7 | DNAK RHOS4
Q4UJK7 | DNAK RICFE
Q57TP3 | DNAK SALCH
Q5PDJ5 | DNAK SALPA
Q326K7 | DNAK SHIBS
Q32KA5 | DNAK SHIDS
Q3Z601 | DNAK SHISS
Q5LWJ6 | DNAK SILPO
Q5HFI0 | DNAK STAAC
Q5HNW6 | DNAK STAEQ
Q4L6T0 | DNAK STAHQ
Q49Y22 | DNAK STAS1
Q3K3T2 | DNAK STRA1
P0A3J3 | DNAK STRAS
P0A3J4 | DNAK STRAG
P0C0C6 | DNAK STRP1
P68837 | DNAK STRP8
Q48RR3 | DNAK STRPM
Q5M1T8 | DNAK STRT1
Q5M6D1 | DNAK STRT2
Q47TI0 | DNAK THEFY

(358) GKEPRKDVPN PDEAVAVGAAI QGGVLAGD
 (358) GKEARKDVNPDEAVAMGAAI QGAVLAGD
 (358) GKEARKDVNPDEAVAMGAAI QGAVLAGD
 (357) GQEPRKDVPN PDEAVAAGAAI QGAVLSDG
 (359) GKEARKDVNPDEAVAVGAAI QGSVLSGD
 (354) GKEPHKGVN PDEVVALGAAI QAGVLQGD
 (352) GREGPHKGVN PDEVVALGAAI QGGVLNKE
 (358) GKEPRKDVPN PDEAVAIGAAV QGGVLITGD
 (358) GKEPRKDVPN PDEVVAMGAAI QAGVLQGD
 (328) GKEPNKGVN PDEVVAMGAAI QGGVITGD
 (328) GKEPHKGVN PDEVVAMGAAI QAGVITGD
 (328) GKDPHKGVNPDEVVAMGAAI QGGVITGD
 (328) GKDPHKGVNPDEVVAMGAAI QGGVITGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (327) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (327) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVITGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVISGD
 (328) GKEPNKSVNPDEVVAMGAAI QGGVISGD
 (330) GKEPNKGVN PDEVVAVGAAI QAGVLKGD

	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	<u>24</u>	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	<u>29</u>	3	-28	-14	15	100	100
L		5	5	-5	3	4	13	4	2	8	-4	<u>14</u>	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	<u>34</u>	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	<u>29</u>	9	-22	6	-4	100	100
T		8	-1	12	-2	0	5	6	-4	19	-4	8	5	-1	2	-8	-8	7	<u>22</u>	19	-15	4	-3	100	100
C		17	0	<u>24</u>	-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	<u>31</u>	-19	-5	-5	100	100
C		10	-8	<u>15</u>	-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	<u>26</u>	-24	-6	8	100	100



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:

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](#):

Example: O05689 ([annotated output](#))



Matrizes PSSM

Matrizes PSSM

- PSSM (**P**osition **S**pecific **S**coring **M**atrix) 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
- Os scores produzidos por uma matriz PSSM refletem as particularidades da família de proteínas a partir da qual foram construídas (valores totalmente diferente de matrizes BLOSUM ou PAM)

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	2	1	0	2	
C	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	
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	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

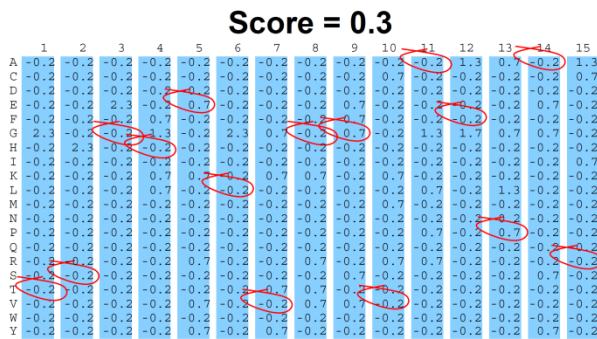
As frequências de ocorrência de cada aminoácido são contadas para cada **posição** do alinhamento. As contagens são normalizadas e convertidas numa matriz *log odds* semelhante a uma matriz de score.

Criação de uma matriz PSSM

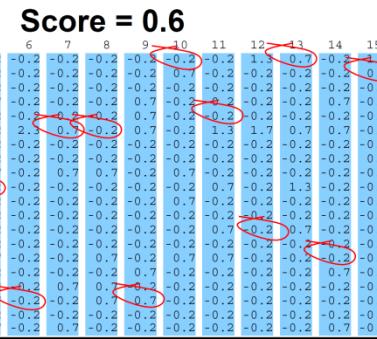
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	1.3	0.7	-0.2	1.3
C	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	0.7
D	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
E	-0.2	-0.2	2.3	-0.2	0.7	-0.2	-0.2	-0.2	0.7	-0.2	-0.2	-0.2	-0.2	-0.2	0.7
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.

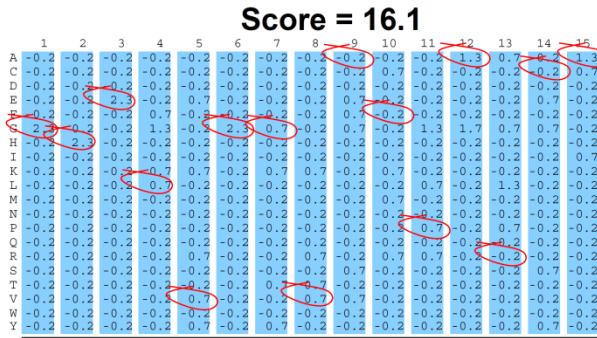
Uso de matriz PSSM



Position +1



T S G H E L V G G V A F P A R C A S

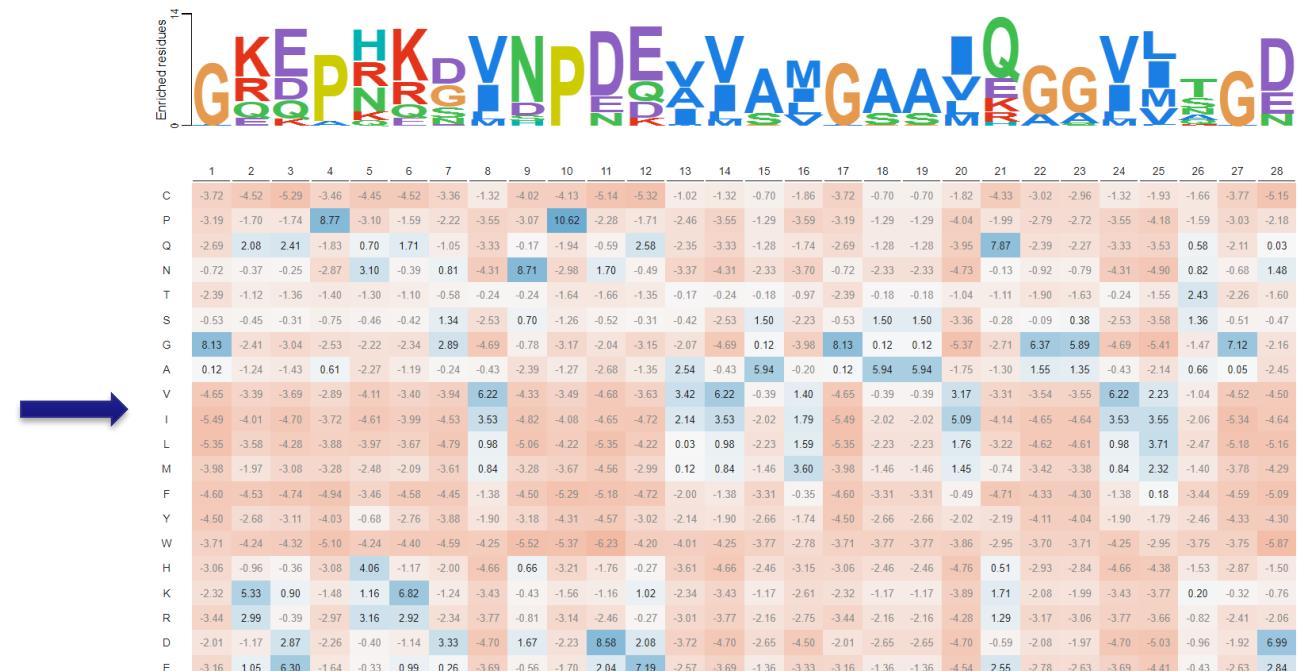


\leftarrow Position +1

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.

Geração de PSSMs com pssmsearch

Q3IC08	DNAK_PSEHT
Q3KIA0	DNAK_PSEPF
Q4ZNP7	DNAK_PSEU2
Q4FPS9	DNAK_PSYAR
Q46X17	DNAK_RALEJ
Q3IYM7	DNAK_RHOS4
Q4UJK7	DNAK_RICFE
Q57FB3	DNAK_SALCH
Q5PDJ5	DNAK_SALPA
Q326R7	DNAK_SHIBS
Q32KA5	DNAK_SHIDS
Q3Z6O1	DNAK_SHISS
Q5LW62	DNAK_SILO
Q5HF10	DNAK_STAAC
Q5HNW6	DNAK_STAEQ
Q4L6T0	DNAK_STAHJ
Q49Y22	DNAK_STAS1
Q3K3T2	DNAK_STRA1
P0A3J3	DNAK_STRA5
P0A3J4	DNAK_STRA6
P0C0C6	DNAK_STRP1
P68837	DNAK_STRP8
Q48RR3	DNAK_STRPM
Q5M1T8	DNAK_STRT1
Q5M6D1	DNAK_STRT2
Q47T10	DNAK_THEFY



<http://slim.icr.ac.uk/pssmsearch/>



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
C	-3.72	-4.52	-5.29	-3.46	-4.45	-4.52	-3.36	-1.32	-4.02	-4.13	-5.14	-5.32	-1.02	-1.32	-0.70	-1.86	-3.72	-0.70	-0.70	-1.82	-4.33	-3.02	-2.96	-1.32	-1.93	-1.66	-3.77	-5.15
P	-3.19	-1.70	-1.74	8.77	-3.10	-1.59	-2.22	-3.55	-3.07	10.62	-2.28	-1.71	-2.46	-3.55	-1.29	-3.59	-3.19	-1.29	-1.29	-4.04	-1.99	-2.79	-2.72	-3.55	-4.18	-1.59	-3.03	-2.18
Q	-2.69	2.08	2.41	-1.83	0.70	1.71	-1.05	-3.33	-0.17	-1.94	-0.59	2.58	-2.35	-3.33	-1.28	-1.74	-2.69	-1.28	-1.28	-3.95	7.87	-2.39	-2.27	-3.33	-3.53	0.58	-2.11	0.03
N	-0.72	-0.37	-0.25	-2.87	3.10	-0.39	0.81	-4.31	8.71	-2.98	1.70	-0.49	-3.37	-4.31	-2.33	-3.70	-0.72	-2.33	-2.33	-4.73	-0.13	-0.92	-0.79	-4.31	-4.90	0.82	-0.68	1.48
T	-2.39	-1.12	-1.36	-1.40	-1.30	-1.10	-0.58	-0.24	-0.24	-1.64	-1.66	-1.35	-0.17	-0.24	-0.18	-0.97	-2.39	-0.18	-0.18	-1.04	-1.11	-1.90	-1.63	-0.24	-1.55	2.43	-2.26	-1.60
S	-0.53	-0.45	-0.31	-0.75	-0.46	-0.42	1.34	-2.53	0.70	-1.26	-0.52	-0.31	-0.42	-2.53	1.50	-2.23	-0.53	1.50	1.50	-3.36	-0.28	-0.09	0.38	-2.53	-3.58	1.36	-0.51	-0.47
G	8.13	-2.41	-3.04	-2.53	-2.22	-2.34	2.89	-4.69	-0.78	-3.17	-2.04	-3.15	-2.07	-4.69	0.12	-3.98	8.13	0.12	0.12	-5.37	-2.71	6.37	5.89	-4.69	-5.41	-1.47	7.12	-2.16
A	0.12	-1.24	-1.43	0.61	-2.27	-1.19	-0.24	-0.43	-2.39	-1.27	-2.68	-1.35	2.54	-0.43	5.94	-0.20	0.12	5.94	5.94	-1.75	-1.30	1.55	1.35	-0.43	-2.14	0.66	0.05	-2.45
V	-4.65	-3.39	-3.69	-2.89	-4.11	-3.40	-3.94	6.22	-4.33	-3.49	-4.68	-3.63	3.42	6.22	-0.39	1.40	-4.65	-0.39	-0.39	3.17	-3.31	-3.54	-3.55	6.22	2.23	-1.04	-4.52	-4.50
I	-5.49	-4.01	-4.70	-3.72	-4.61	-3.99	-4.53	3.53	-4.82	-4.08	-4.65	-4.72	2.14	3.53	-2.02	1.79	-5.49	-2.02	-2.02	5.09	-4.14	-4.65	-4.64	3.53	3.55	-2.06	-5.34	-4.64
L	-5.35	-3.58	-4.28	-3.88	-3.97	-3.67	-4.79	0.98	-5.06	-4.22	-5.35	-4.22	0.03	0.98	-2.23	1.59	-5.35	-2.23	-2.23	1.76	-3.22	-4.62	-4.61	0.98	3.71	-2.47	-5.18	-5.16
M	-3.98	-1.97	-3.08	-3.28	-2.48	-2.09	-3.61	0.84	-3.28	-3.67	-4.56	-2.99	0.12	0.84	-1.46	3.60	-3.98	-1.46	-1.46	1.45	-0.74	-3.42	-3.38	0.84	2.32	-1.40	-3.78	-4.29
F	-4.60	-4.53	-4.74	-4.94	-3.46	-4.58	-4.45	-1.38	-4.50	-5.29	-5.18	-4.72	-2.00	-1.38	-3.31	-0.35	-4.60	-3.31	-3.31	-0.49	-4.71	-4.33	-4.30	-1.38	0.18	-3.44	-4.59	-5.09
Y	-4.50	-2.68	-3.11	-4.03	-0.68	-2.76	-3.88	-1.90	-3.18	-4.31	-4.57	-3.02	-2.14	-1.90	-2.66	-1.74	-4.50	-2.66	-2.66	-2.02	-2.19	-4.11	-4.04	-1.90	-1.79	-2.46	-4.33	-4.30
W	-3.71	-4.24	-4.32	-5.10	-4.24	-4.40	-4.59	-4.25	-5.52	-5.37	-6.23	-4.20	-4.01	-4.25	-3.77	-2.78	-3.71	-3.77	-3.77	-3.86	-2.95	-3.70	-3.71	-4.25	-2.95	-3.75	-3.75	-5.87
H	-3.06	-0.96	-0.36	-3.08	4.06	-1.17	-2.00	-4.66	0.66	-3.21	-1.76	-0.27	-3.61	-4.66	-2.46	-3.15	-3.06	-2.46	-2.46	-4.76	0.51	-2.93	-2.84	-4.66	-4.38	-1.53	-2.87	-1.50
K	-2.32	5.33	0.90	-1.48	1.16	6.82	-1.24	-3.43	-0.43	-1.56	-1.16	1.02	-2.34	-3.43	-1.17	-2.61	-2.32	-1.17	-1.17	-3.89	1.71	-2.08	-1.99	-3.43	-3.77	0.20	-0.32	-0.76
R	-3.44	2.99	-0.39	-2.97	3.16	2.92	-2.34	-3.77	-0.81	-3.14	-2.46	-0.27	-3.01	-3.77	-2.16	-2.75	-3.44	-2.16	-2.16	-4.28	1.29	-3.17	-3.06	-3.77	-3.66	-0.82	-2.41	-2.06
D	-2.01	-1.17	2.87	-2.26	-0.40	-1.14	3.33	-4.70	1.67	-2.23	8.58	2.08	-3.72	-4.70	-2.65	-4.50	-2.01	-2.65	-2.65	-4.70	-0.59	-2.08	-1.97	-4.70	-5.03	-0.96	-1.92	6.99
E	-3.16	1.05	6.30	-1.64	-0.33	0.99	0.26	-3.69	-0.56	-1.70	2.04	7.19	-2.57	-3.69	-1.36	-3.33	-3.16	-1.36	-4.54	2.55	-2.78	-2.63	-3.69	-4.41	-0.43	-2.63	2.84	

PSI-BLAST

PSI-BLAST

A variante PSI (**P**osition **S**pecific **I**terated) do programa BLAST combina um modelo 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 standard (por exemplo BLOSUM62).
 2. Um modelo PSSM é construído automaticamente a partir do alinhamento múltiplo dos “hits” de score mais elevado (E-value inferior ao valor do parâmetro *PSI-BLAST Thresold*).
 3. O modelo PSSM é usado em vez da matriz de score original para realizar uma nova pesquisa BLAST. Como resultado, poderão ser adicionadas novas sequências de E-value inferior ao valor de *PSI-BLAST Thresold*.
 4. O novo conjunto de sequências obtido em 3. é usado para construir um novo alinhamento múltiplo e, a partir deste, uma nova matriz PSSM
 5. Os passos 3. e 4. são repetidos enquanto sejam adicionadas, em cada iteração, novas sequências à lista de hits.
 6. Considera-se que o algoritmo convergiu quando não são adicionadas mais sequências à nova iteração.

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>

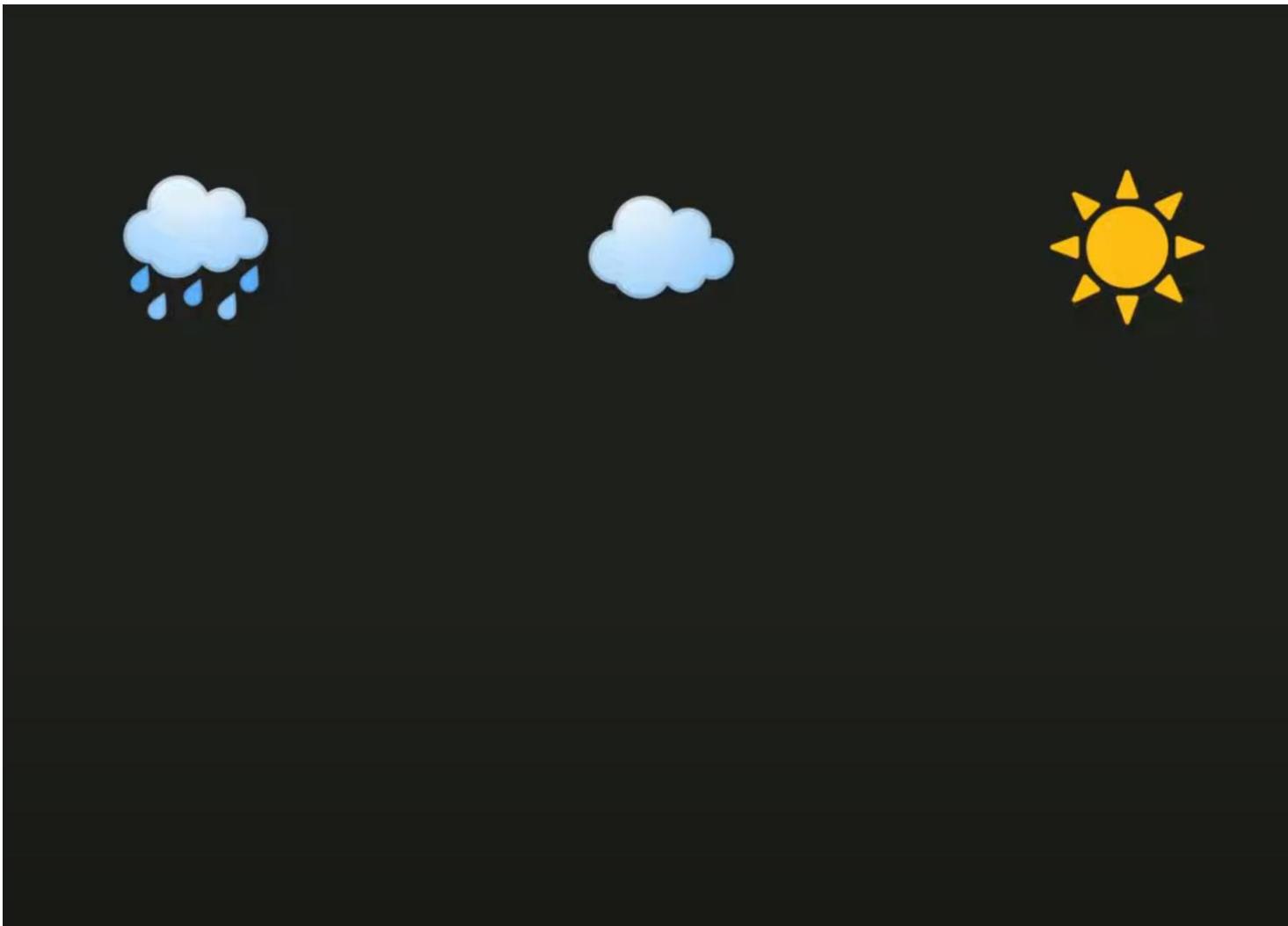
Pesquisa em múltiplas bases de dados:

INTERPRO

<http://www.ebi.ac.uk/interpro>

Hidden Markov Models (HMMs)

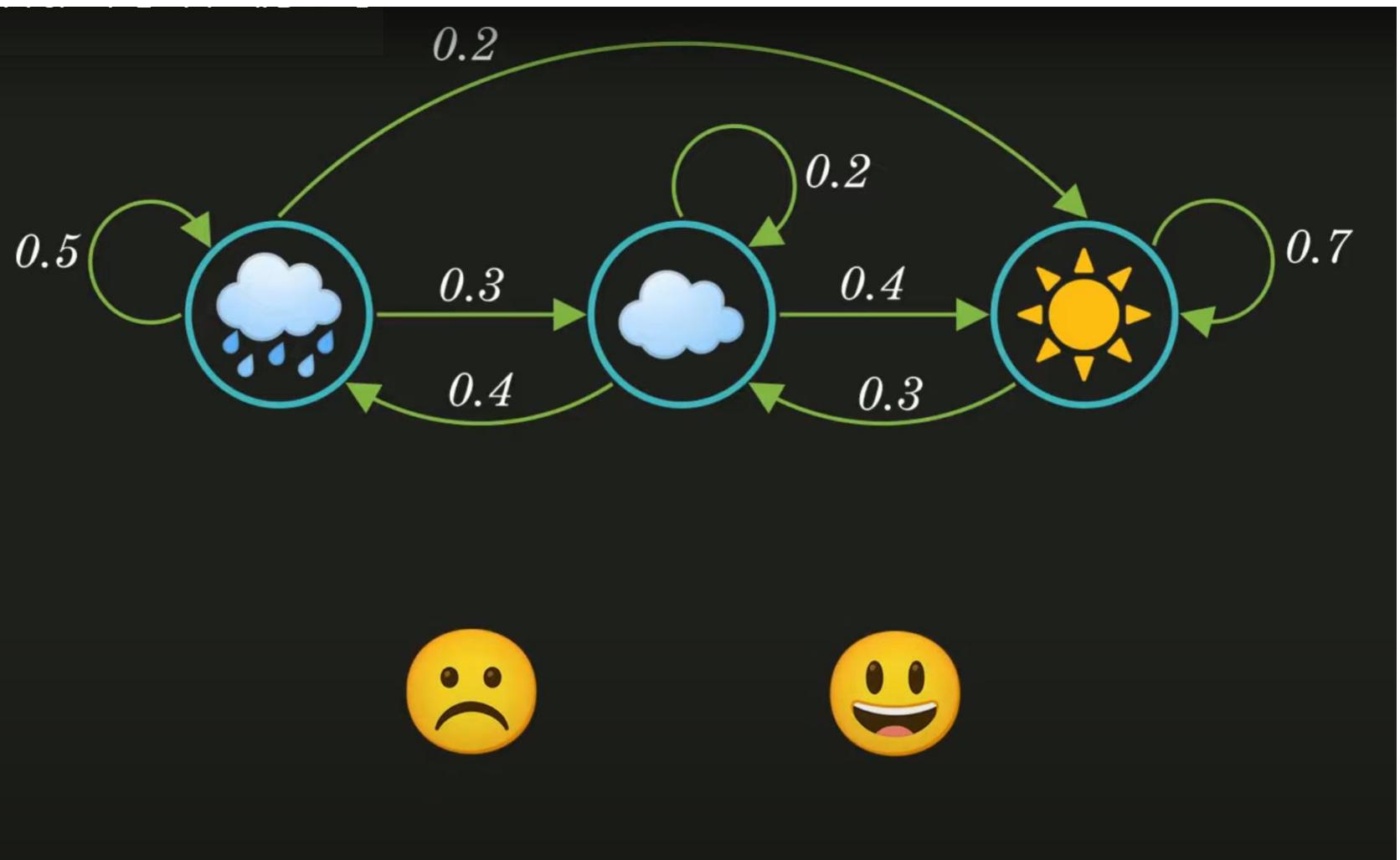
- Os HMMs são modelos estocásticos que representam a estrutura de uma sequência em termos de distribuições de probabilidade de transição entre diferentes estados.
- As variáveis observáveis são determinadas pelos estados do sistema, estes tipicamente ocultos.
- Através da análise de uma ou mais sequências, é possível estimar os parâmetros internos de um dado modelo.
- Conhecidos os parâmetros, é possível estimar a probabilidade de um modelo produzir uma certa sequência.
- Os HMMs permitem obter representações probabilísticas de conjuntos de sequências que capturam as preferências posicionais da composição aminoacídica.
- Os HMMs são um dos métodos mais sensíveis para deteção de similaridades fracas entre sequências.



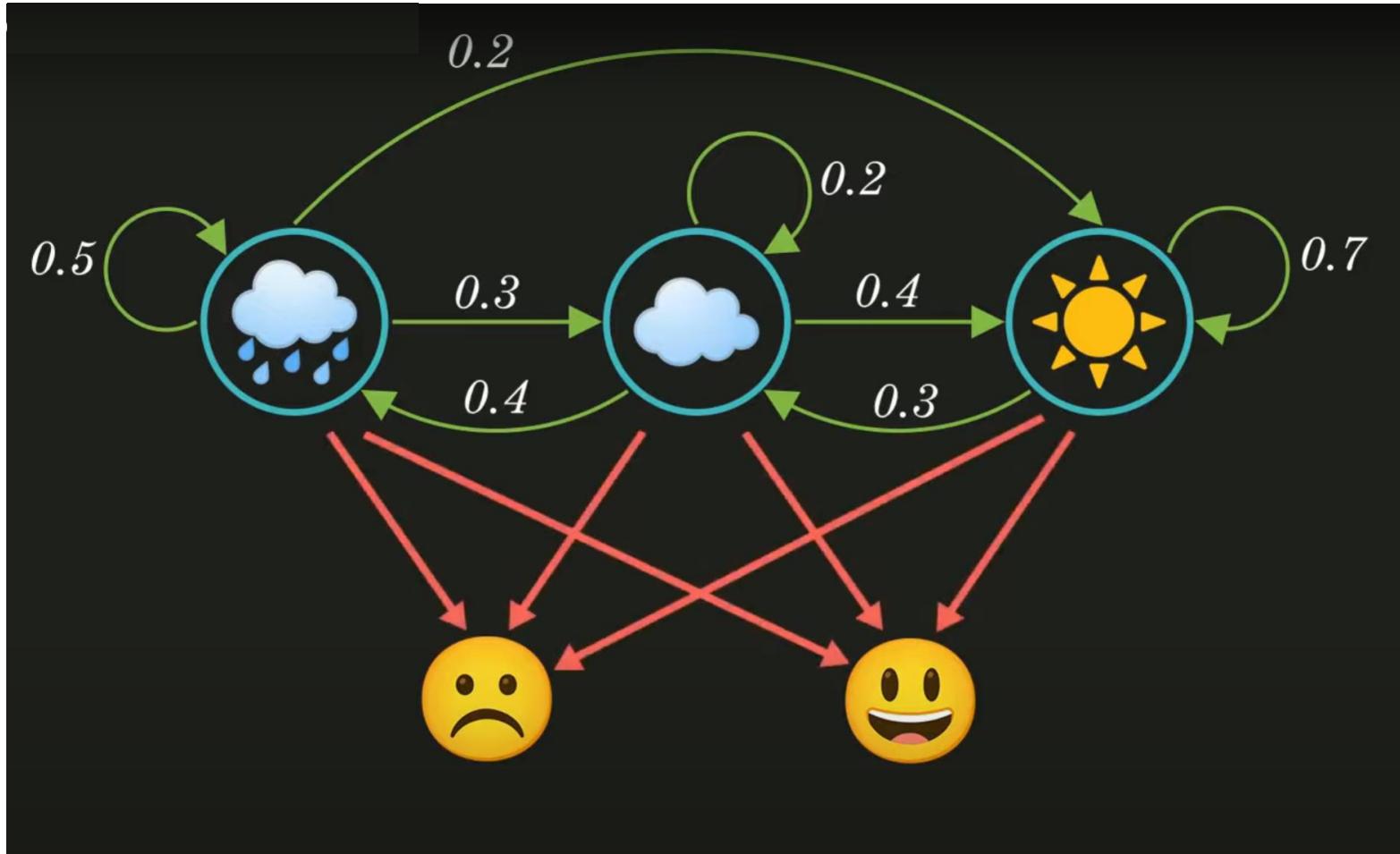
<https://youtu.be/RWkHJnFj5rY>

Model Clearly Explained! Part - 5

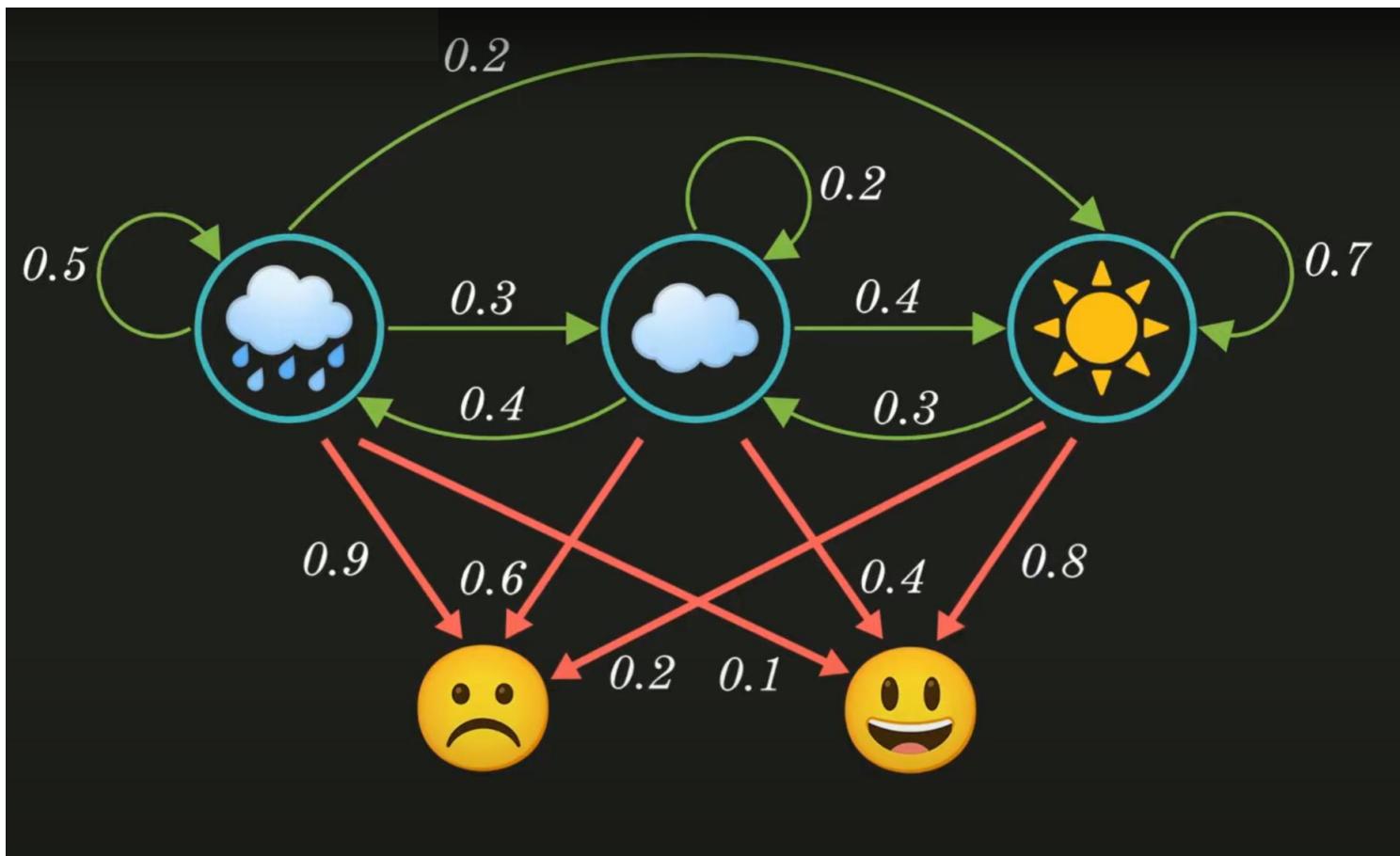




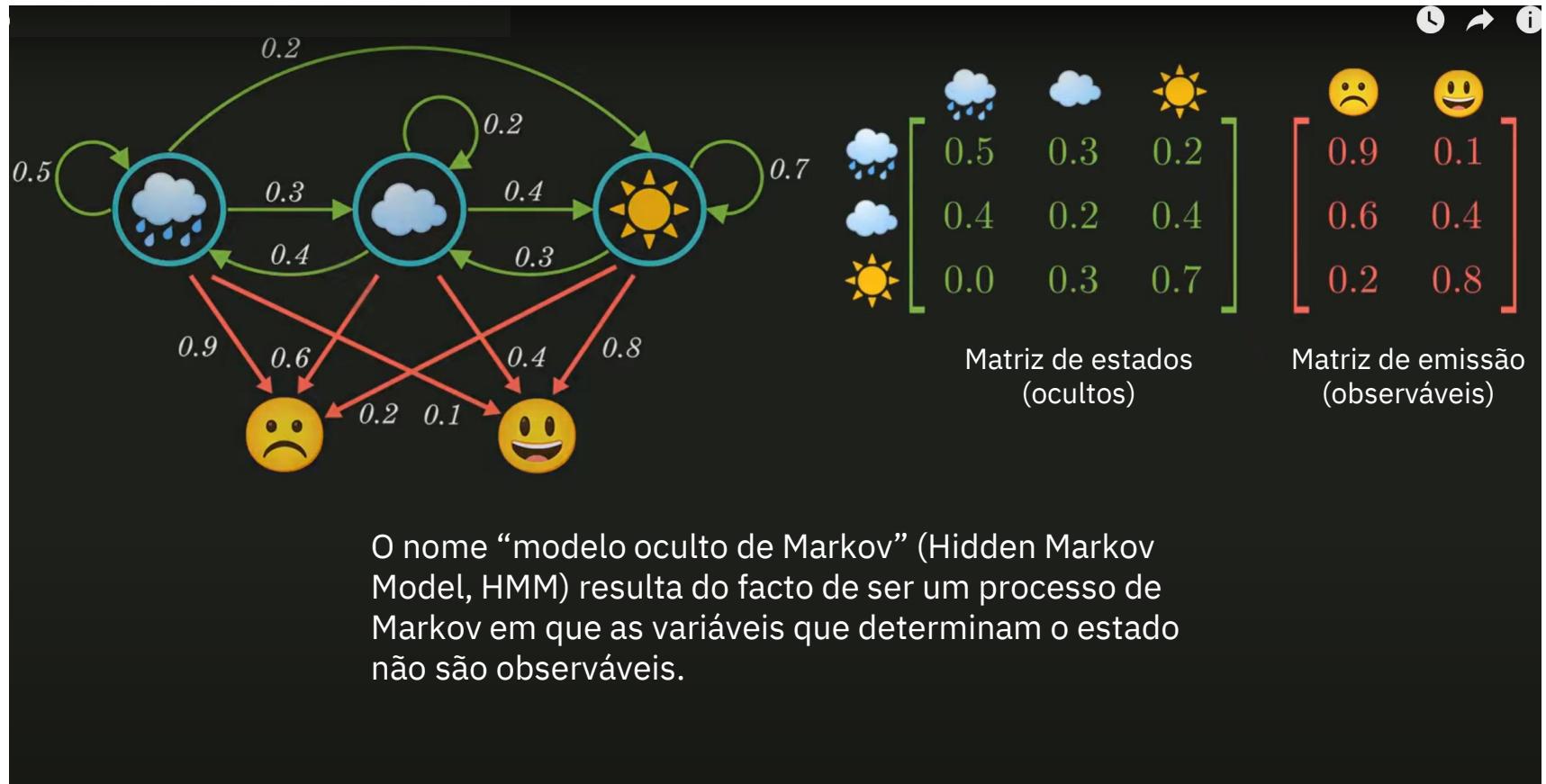
<https://youtu.be/RWkHJnFj5rY>



<https://youtu.be/RWkHJnFj5rY>



<https://youtu.be/RWkHJnFj5rY>



<https://youtu.be/RWkHJnFj5rY>



	0.5	0.3	0.2
	0.4	0.2	0.4
	0.0	0.3	0.7

	0.9	0.1
	0.6	0.4
	0.2	0.8

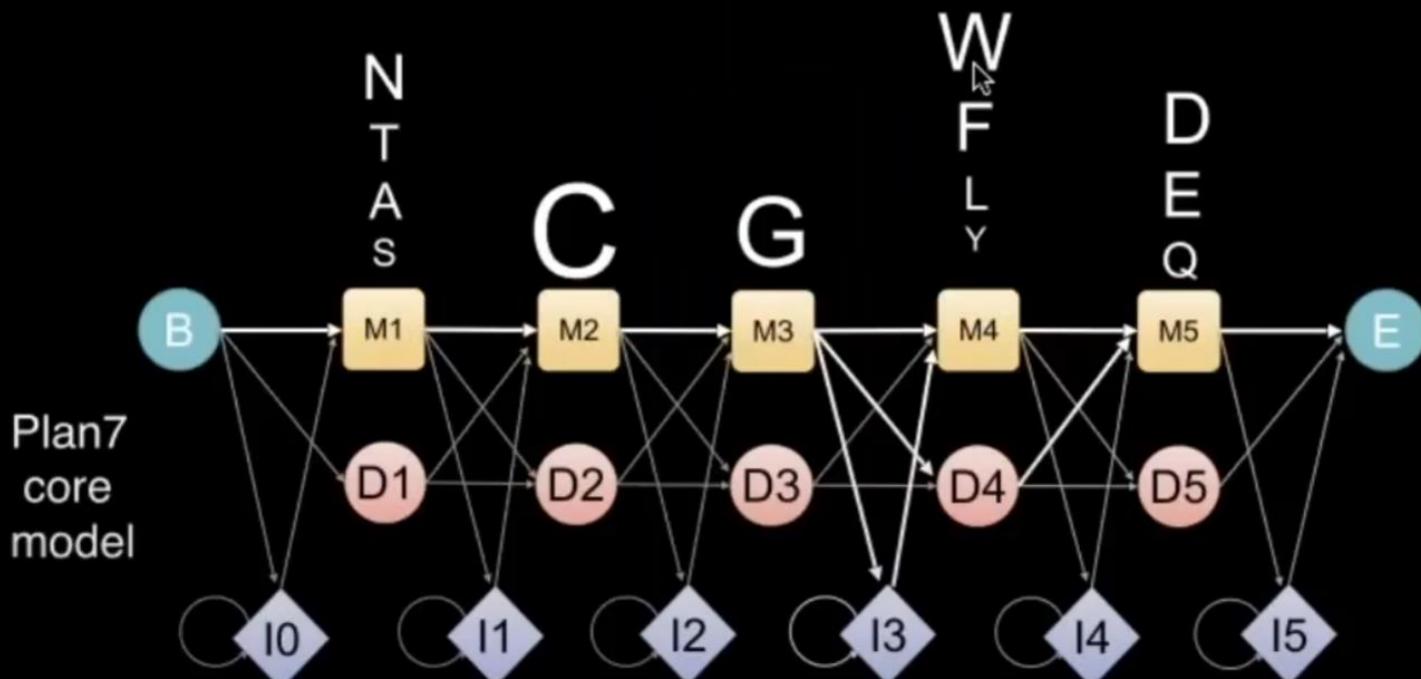
Qual a probabilidade da sequência de emojis,
assumindo a sequência de condições atmosféricas?

$$P(Y = \text{😊 😊 ☹} , X = \text{☀️ ☁ ☀️})$$

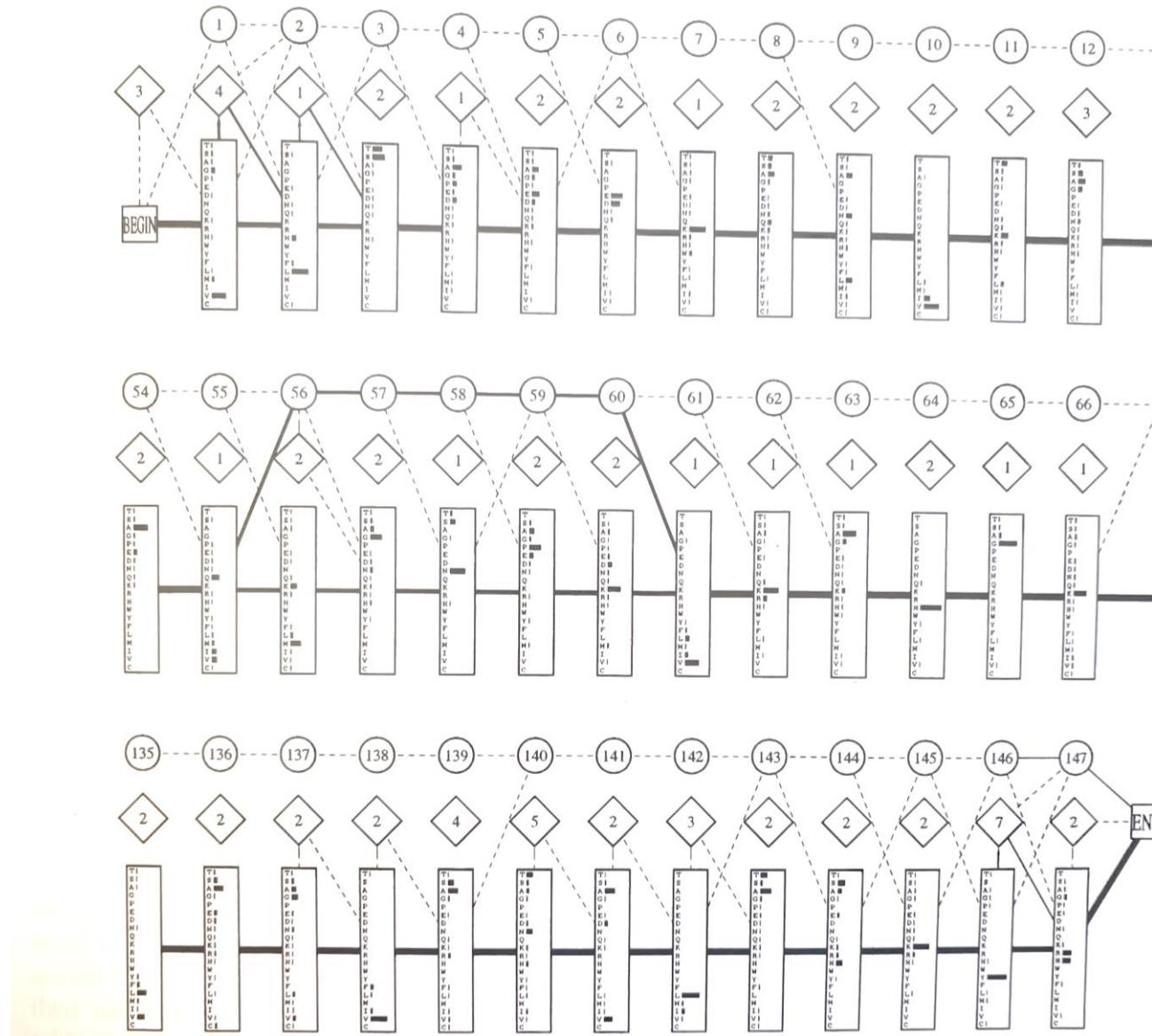
Representação HMM de um alinhamento

Input multiple alignment:

seq1	ACG-LD	Consensus columns assigned
seq2	SCG--E	Defining inserts and deletes:
Seq3	NCGgFD	
Seq4	TCG-WQ	
	123-45	



Perfil HMM das globulinas



HMMER@EBI

The screenshot shows the HMMER@EBI homepage. At the top, there's a dark header bar with links for EMBL-EBI, Services, Research, Training, About us, and a search icon. To the right is the EMBL-EBI Hinxton logo. Below the header is a large teal banner featuring the HMMER logo (a stylized 'H' made of colored squares) and the text "HMMER Biosequence analysis using profile hidden Markov Models". A navigation menu below the banner includes Home, Search, Results, Software, Help, About, and Contact. A yellow callout box in the center of the page contains a "Service Update: Improvements Underway for Enhanced Reliability" message. The message states: "We understand that our service has experienced issues in recent months, and we sincerely apologize for any inconvenience caused. We want to assure you that we are actively working to address these challenges and enhance the overall reliability of our service. Please bear with us as we make significant improvements to minimize disruptions. Your patience and understanding are greatly appreciated during this transition."

Quick search

Paste in your sequence or use the [example](#)

Enter your sequence

Reference Proteomes UniProtKB SwissProt Pfam

Submit **Reset** **Clean**

Alternative search options

The HMMER web server: fast and sensitive homology searches. This site has been designed to provide near **interactive searches** for most queries, coupled with **intuitive and interactive results** visualisations.



Quickstart tutorial



Online documentation

News

February, 2022

HmmerWeb Pfam 35.0 Release

HmmerWeb 2.43 includes support for the most recent Pfam 35.0 and its 19632 protein families, used to annotate functional domains in all our hosted sequences. Find more in the official Pfam website.

Download HMMER

v3.3.2

[Download Source](#)

User's Guide: [PDF, pages]

Alternative Download Options

Papers

[HMMER web server: 2018 update](#)

S.C. POTTER, A. LUCIANI, S.R. EDDY Y. PARK, R. LOPEZ and R.D. FINN,
Nucleic Acids Research (2018) Web Server Issue 46:W200-W204. PDF

[HMMER web server: 2015 update](#)

<https://www.ebi.ac.uk/Tools/hmmer/>

Ferramentas de HMMER

- **phmmmer** - sequência de proteína contra uma base de dados de sequências de proteína (pouco mais sensível que uma pesquisa normal com BLAST)
- **hmmscan** - sequência de proteína contra uma livraria de perfis de HMM (pfam, TIGRFAM, Gene3D, Superfamily, PIRSF, TreeFam)
- **hmmpress** - alinhamento múltiplo ou perfil HMM contra uma base de dados de sequências de proteínas
- **jackhammer** - pesquisa iterativa iniciada com uma única sequência, perfil HMM ou alinhamento múltiplo. Cada nova interação usa as sequências encontradas para refinar o perfil HMM.

HMMER@EBI

The screenshot shows the HMMER@EBI homepage. At the top, there's a navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and a search icon. On the right, it says "EMBL-EBI Hinxtion". Below the navigation is the HMMER logo and the text "Biosequence analysis using profile hidden Markov Models". A horizontal menu bar includes Home, Search, Results, Software, Help, About, and Contact. A yellow callout box in the center states: "Service Update: Improvements Underway for Enhanced Reliability. We understand that our service has experienced issues in recent months, and we sincerely apologize for any inconvenience caused. We want to assure you that we are actively working to address these challenges and enhance the overall reliability of our service. Please bear with us as we make significant improvements to minimize disruptions. Your patience and understanding are greatly appreciated during this transition." Below this, there's a "Quick search" section where users can paste their sequence or use an example. It includes a sequence editor with a scroll bar, and buttons for Reference Proteomes, UniProtKB, SwissProt, and Pfam. There are "Submit" and "Reset" buttons, and a link for "Alternative search options". To the right, there's a brief description of the service, a "Quickstart tutorial" button, and an "Online documentation" button.

News

HmmrWeb Pfam 35.0 Release

HmmrWeb 2.43 includes support for the most recent Pfam 35.0 and its 19632 protein families, used to annotate functional domains in all our hosted sequences. Find more in the official Pfam website.

February, 2022

Download HMMER

v3.3.2



User's Guide: [PDF, pages]
Alternative Download Options

Papers

HMMER web server: 2018 update

S.C. POTTER, A. LUCIANI, S.R. EDDY Y. PARK, R. LOPEZ and R.D. FINN,
Nucleic Acids Research (2018) Web Server Issue 46:W200-W204. PDF

HMMER web server: 2015 update

phhmer



protein sequence vs protein sequence database

Paste a Sequence | Upload a File | Accession Search

Paste in your sequence or use the example ⓘ

```
>sp|P02239|LGB1_LUPLU Leghemoglobin-1 OS=Lupinus luteus OX=3873 PE=1 SV=3
MGVLTDQVALVKSSFEEFNANIPKNTHRFFFLVLEIAPGAKDLFSFLKGSSSEVPQNNPD
LQAHAGKVFKLTYEARIQLOVNGAVASDATLRLSLSLGSVHVSKGVVDAAHFVVKEAILKTIK
EVVGDKWSEELNTAWTIIAYDELAIITIKKENMDAA
```

Submit **Reset**

▼ Sequence Database ⓘ

Frequently used databases: [Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [AlphaFold](#) [Ensembl](#)

Current database selection:

Reference Proteomes

▼ Restrict by Taxonomy ⓘ

Taxon search Pre-defined representatives

Organism:

|

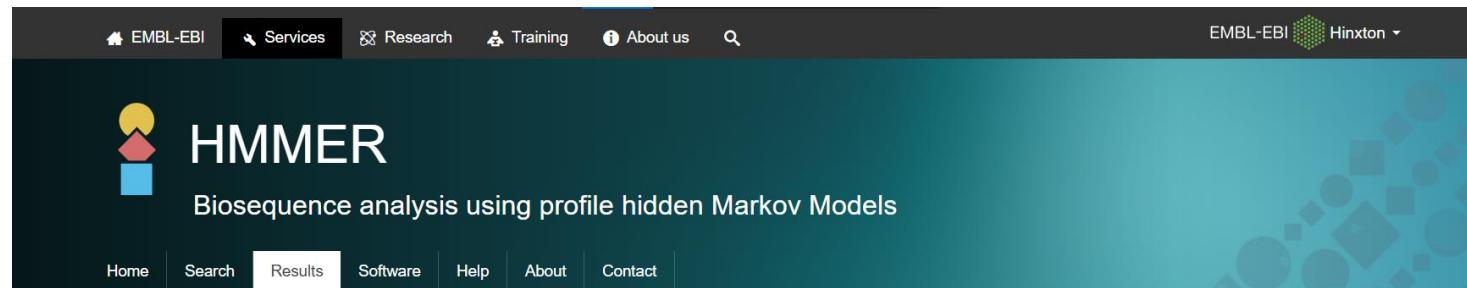
[Include all taxa](#)

Homo sapiens (taxid: 9606)

x

To only remove taxa, make sure to clear your current selection then click the "Include all taxa" button, and finally search for the taxa you wish to remove

phhmer



The banner features the HMMER logo (a stylized orange and blue graphic) and the word "HMMER" in large white letters. Below it, the tagline "Biosequence analysis using profile hidden Markov Models" is displayed. A navigation bar at the bottom includes links for Home, Search, Results (which is highlighted), Software, Help, About, and Contact.

Service Update: Improvements Underway for Enhanced Reliability

We understand that our service has experienced issues in recent months, and we sincerely apologize for any inconvenience caused. We want to assure you that we are actively working to address these challenges and enhance the overall reliability of our service.

Please bear with us as we make significant improvements to minimize disruptions. Your patience and understanding are greatly appreciated during this transition.

PHMMER Results

[Search Again](#)



Distribution of Significant Hits



Did you know? Clicking the button customise, in the table header below, gives you the opportunity toggle up to twelve columns of data in this table.

[hide this X](#)

Also, have a look at the new **Cross-references** column, showing references to other resources at the EBI.

phhmer

Home | Search | Results | Software | Help | About | Contact

Show hit details

Distribution of Significant Hits ⓘ



■ Bacteria ■ Eukaryota ■ Archaea ■ Viruses ■ Unclassified Sequences ■ Other Sequences

« First « Previous Page 1 of 102 Next » Last »

Significant Query Matches (4036) in uniprotkb (v.2021_04)

Customise

	Target	Species	E-value
>	A0A833QD73_9POAL ⓘ	Carex littledalei ⓘ	8.3e-112
>	A0A443PME4_9MAGN ⓘ	Cinnamomum micranthum f. kanehirae ⓘ	6.2e-104
>	A0A443PMG7_9MAGN ⓘ	Cinnamomum micranthum f. kanehirae ⓘ	2.1e-103
>	A0A7M7G143_STRPU ⓘ	Strongylocentrotus purpuratus ⓘ	1.5e-96
>	A0A0D9Z4F0_9ORYZ ⓘ	Oryza glumipatula ⓘ	2.9e-95
>	LGB1_LUPLU ⓘ	Lupinus luteus ⓘ	2.7e-94
>	A0A7M7HNK1_STRPU ⓘ	Strongylocentrotus purpuratus ⓘ	2.7e-89
>	A0A7M7LT58_STRPU ⓘ	Strongylocentrotus purpuratus ⓘ	6.1e-89
>	A0A833RM36_9POAL ⓘ	Carex littledalei ⓘ	5.9e-85
>	A0A833QA41_9POAL ⓘ	Carex littledalei ⓘ	3.6e-84
>	A0A394DEX3_LUPAN ⓘ	Lupinus angustifolius ⓘ	2.1e-83
>	A0A6A4P2K0_LUPAL ⓘ	Lupinus albus ⓘ	1.4e-82
>	A0A0D3FFX0_9ORYZ ⓘ	Oryza barthii ⓘ	2.3e-82
>	A0A7M7HLE9_STRPU ⓘ	Strongylocentrotus purpuratus ⓘ	8.8e-82
>	A0A7M7HID4_STRPU ⓘ	Strongylocentrotus purpuratus ⓘ	9.6e-82
>	LGB2_LUPLU ⓘ	Lupinus luteus ⓘ	6.2e-81
>	Q6LBG6_LUPLU ⓘ	Lupinus luteus ⓘ	1.9e-79

(show all) alignments

> A0A1I7GN75_LUPAN ⓘ

Your search took: 6.44 secs

Lupinus angustifolius ⓘ

showing rows 1 - 100 of 10125

2.2e-78

jackhmmer

 HMMER
Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

Service Update: Improvements Underway for Enhanced Reliability
We understand that our service has experienced issues in recent months, and we sincerely apologize for any inconvenience caused. We want to assure you that we are actively working to address these challenges and enhance the overall reliability of our service.
Please bear with us as we make significant improvements to minimize disruptions. Your patience and understanding are greatly appreciated during this transition.

phmmmer hmmscan hmmsearch jackhmmer

iterative search vs protein sequence database

Paste a Sequence or an Alignment | Upload a File | Accession Search

Paste in your sequence (example), HMM (example) or multiple sequence alignment (example) ⓘ

```
>sp|P02239|LGB1_LUPLU Leghemoglobin-1 OS=Lupinus luteus OX=3873 PE=1 SV=3  
MGVLTDVQVALVKSSFEENANIPKNTNHRFFTLVLEIAPGAKDLFSFLKGSSVPQNNPD  
LQAHAGKVFKLTYEAAIQLQVNGAVASDATALKSLGSVHVSKGVVDAHPVVKEAILKTIK  
EVVGDKWSEELNTAWTIAYDELAIIKKEMKDAA
```

Submit Reset

▼ Sequence Database ⓘ

Frequently used databases: Reference Proteomes UniProtKB SwissProt PDB AlphaFold Ensembl

Current database selection: Reference Proteomes

▼ Restrict by Taxonomy ⓘ

Taxon search Pre-defined representatives

Organism:

Include all taxa

Homo sapiens (taxid: 9606) x

jackhmmer

The screenshot shows the JACKHMMER Results page. At the top, there's a navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and a search icon. On the right, it says "EMBL-EBI Hinxton". Below the navigation is the HMMER logo and the text "Biosequence analysis using profile hidden Markov Models". A horizontal menu bar includes Home, Search, Results (which is highlighted), Software, Help, About, and Contact. A yellow banner at the top of the main content area reads "Service Update: Improvements Underway for Enhanced Reliability". It continues: "We understand that our service has experienced issues in recent months, and we sincerely apologize for any inconvenience caused. We want to assure you that we are actively working to address these challenges and enhance the overall reliability of our service. Please bear with us as we make significant improvements to minimize disruptions. Your patience and understanding are greatly appreciated during this transition."

JACKHMMER Results

[Search Again](#)

Jackhmmer Summary

Iteration	Results	Hits			
		New	Lost <small>?</small>	Dropped <small>?</small>	Total
1	D58B99EC-8D1D-11EE-9C44-E7C8F9E0C6C4.1	+1	-	-	1
2	D58B99EC-8D1D-11EE-9C44-E7C8F9E0C6C4.2	+11	-	-	12
3	D58B99EC-8D1D-11EE-9C44-E7C8F9E0C6C4.3	+16	-	-	28
4	D58B99EC-8D1D-11EE-9C44-E7C8F9E0C6C4.4	+2	-	-	30
5	D58B99EC-8D1D-11EE-9C44-E7C8F9E0C6C4.5	-	-	-	30

Your search has converged. No more iterations will be run.

HmmWeb version 2.41.2

Next release in more than a month

Comments or questions about the site? Click here to use our contact form

jackhmmer

Home | Search | Results | Software | Help | About | Contact |

Distribution of Significant Hits ⓘ

■ Bacteria ■ Eukaryota ■ Archaea ■ Viruses ■ Unclassified Sequences ■ Other Sequences

Significant Query Matches (30) in uniprotrefprot (v.2021_04)

Customise

	Target	Species	E-value
>	A0A2R8Y7X9_HUMAN ⓘ	Homo sapiens ⓘ	9.0e-61
>	HBG1_HUMAN ⓘ	Homo sapiens ⓘ	2.4e-56
>	HBG2_HUMAN ⓘ	Homo sapiens ⓘ	3.4e-56
>	HBD_HUMAN ⓘ	Homo sapiens ⓘ	1.4e-54
>	HBE_HUMAN ⓘ	Homo sapiens ⓘ	3.3e-54
>	HBB_HUMAN ⓘ	Homo sapiens ⓘ	5.7e-54
>	CYGB_HUMAN ⓘ	Homo sapiens ⓘ	1.8e-50
>	HBA_HUMAN ⓘ	Homo sapiens ⓘ	3.4e-47
>	HBAZ_HUMAN ⓘ	Homo sapiens ⓘ	3.4e-47
>	MYG_HUMAN ⓘ	Homo sapiens ⓘ	6.3e-46
>	B0QYF8_HUMAN ⓘ	Homo sapiens ⓘ	3.2e-45
>	HBAT_HUMAN ⓘ	Homo sapiens ⓘ	9.7e-44
>	HBM_HUMAN ⓘ	Homo sapiens ⓘ	3.2e-42
>	NGB_HUMAN ⓘ	Homo sapiens ⓘ	5.9e-39
>	E9PFT6_HUMAN ⓘ	Homo sapiens ⓘ	6.5e-39
>	A0A2R8Y7R2_HUMAN ⓘ	Homo sapiens ⓘ	1.2e-37
>	E9PEW8_HUMAN ⓘ	Homo sapiens ⓘ	2.5e-37
>	K7EMC7_HUMAN ⓘ	Homo sapiens ⓘ	1.5e-35
>	G3V1N2_HUMAN ⓘ	Homo sapiens ⓘ	1.1e-33
>	E9PBW4_HUMAN ⓘ	Homo sapiens ⓘ	4.5e-33
>	F2Z2F1_HUMAN ⓘ	Homo sapiens ⓘ	8.3e-32
>	K7EIM9_HUMAN ⓘ	Homo sapiens ⓘ	2.2e-30
>	F8W6P5_HUMAN ⓘ	Homo sapiens ⓘ	6.7e-29
>	A0A2R8Y7C0_HUMAN ⓘ	Homo sapiens ⓘ	1.9e-27