Master 1 MABS

UE Bioinformatique des séquences

novembre - décembre 2013

APPLICATION DES ALIGNEMENTS MULTIPLES : L'IDENTIFICATION DE MOTIFS OU DE PATTERNS

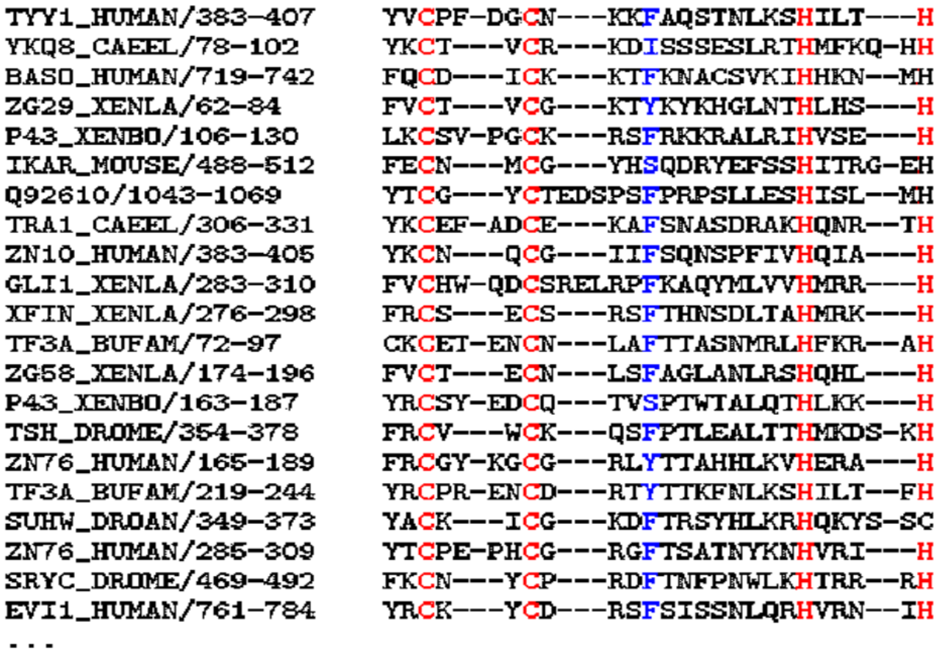
UN PEU DE VOCABULAIRE

Familles, domaines, motifs, pattern, etc...

- domaine protéique: unité structurale (et fonctionnelle) indépendante, évolutivement conservée (doigt de zinc, boucle,...)
- motifs protéiques: plus courts
  - site de modification post-traductionnelle
  - site de liaison (ADN, métal,...)
  - site actif d'enzyme
- famille protéique: ensemble de protéines évolutivement reliées par un ou plusieurs domaines protéiques communs

Domaine conservé

- Exemple doigt de zinc

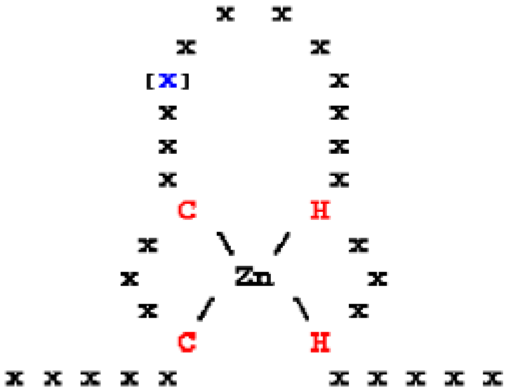


```
TYY1_HUMAN/383-407      YVCPF-DGCN---KKFAQSTNLKSHILT---H
YKQ8_CAEEL/78-102       YKCT---VCR---KDISSSESLRTHMFKQ-HH
BASO_HUMAN/719-742      FQCD---ICK---KTFKNACSVKIHHKN--MH
ZG29_XENLA/62-84        FVCT---VCG---KTYKYKHGLNTHLHS---H
P43_XENBO/106-130       LKCSV-PGCK---RSFRKKRALRIHVSE---H
IKAR_MOUSE/488-512      FECN---MCG---YHSQDRYEFSSHITRG-EH
Q92610/1043-1069        YTCG---YCTEDSPSFPRPSLLESHISL--MH
TRA1_CAEEL/306-331      YKCEF-ADCE---KAFSNASDRAKHQNR--TH
ZN10_HUMAN/383-405      YKCN---QCG---IIFSQNSPFIVHQIA---H
GLI1_XENLA/283-310      FVCHW-QDCSRELRPFKAQYMLVVHMRR---H
XFIN_XENLA/276-298      FRCS---ECS---RSFTHNSDLTAHMRK---H
TF3A_BUFAM/72-97        CKCET-ENCN---LAFTTASNMRLHFKR--AH
ZG58_XENLA/174-196      FVCT---ECN---LSFAGLANLRSHQHL---H
P43_XENBO/163-187       YRCSY-EDCQ---TVSPTWTALQTHLKK---H
TSH_DROME/354-378       FRCV---WCK---QSFPTLEALTTHMKDS-KH
ZN76_HUMAN/165-189      FRCGY-KGCG---RLYTTAHHLKVHERA---H
TF3A_BUFAM/219-244      YRCPR-ENCD---RTYTTKFNLKSHILT--FH
SUHW_DROAN/349-373      YACK---ICG---KDFTRSYHLKRHQKYS-SC
ZN76_HUMAN/285-309      YTCPE-PHCG---RGFTSATNYKNHVRI---H
SRYC_DROME/469-492      FKCN---YCP---RDFTNFPNWLKHTRR--RH
EVI1_HUMAN/761-784      YRCK---YCD---RSFSISSNLQRHVRN--IH
...
```

Extrait de Pfam, entrée zf-C2H2

Domaine conservé

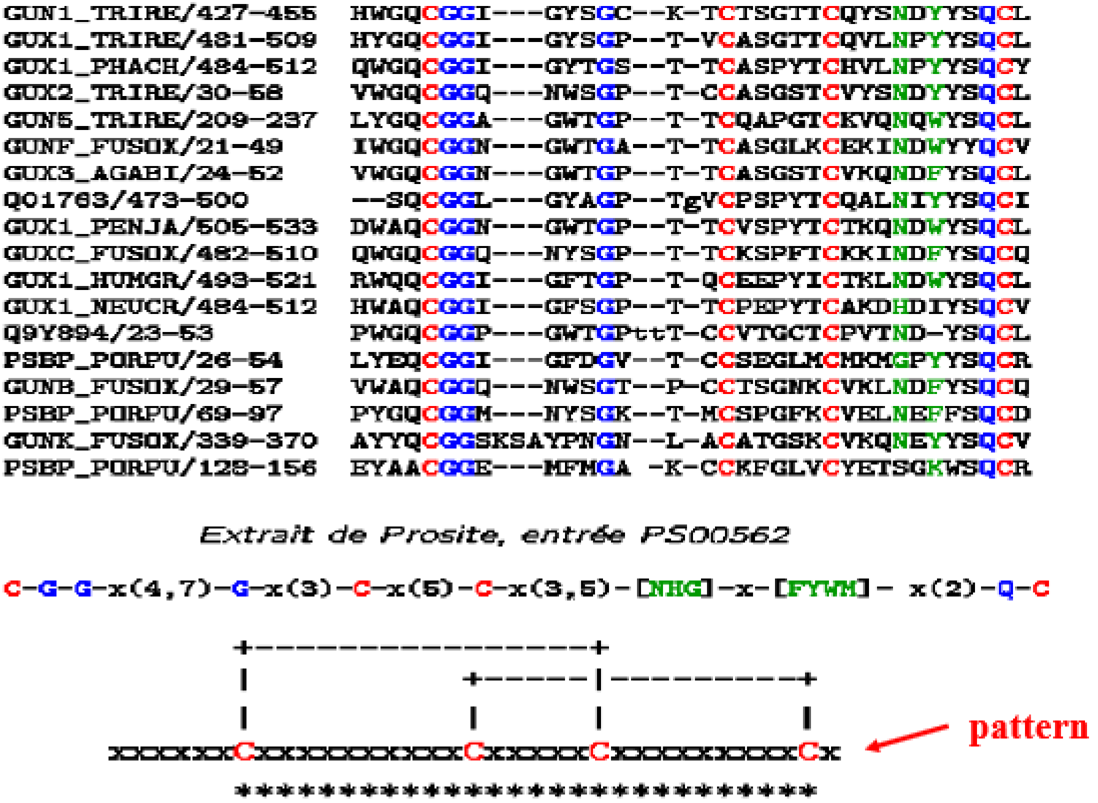
- Exemple doigt de zinc



Motif Prosite:

C-x (2,4)-C-x (3)-[LIVMFYWC]-x (8)-H-x (3,5)-H

MOTIF (PATTERN)



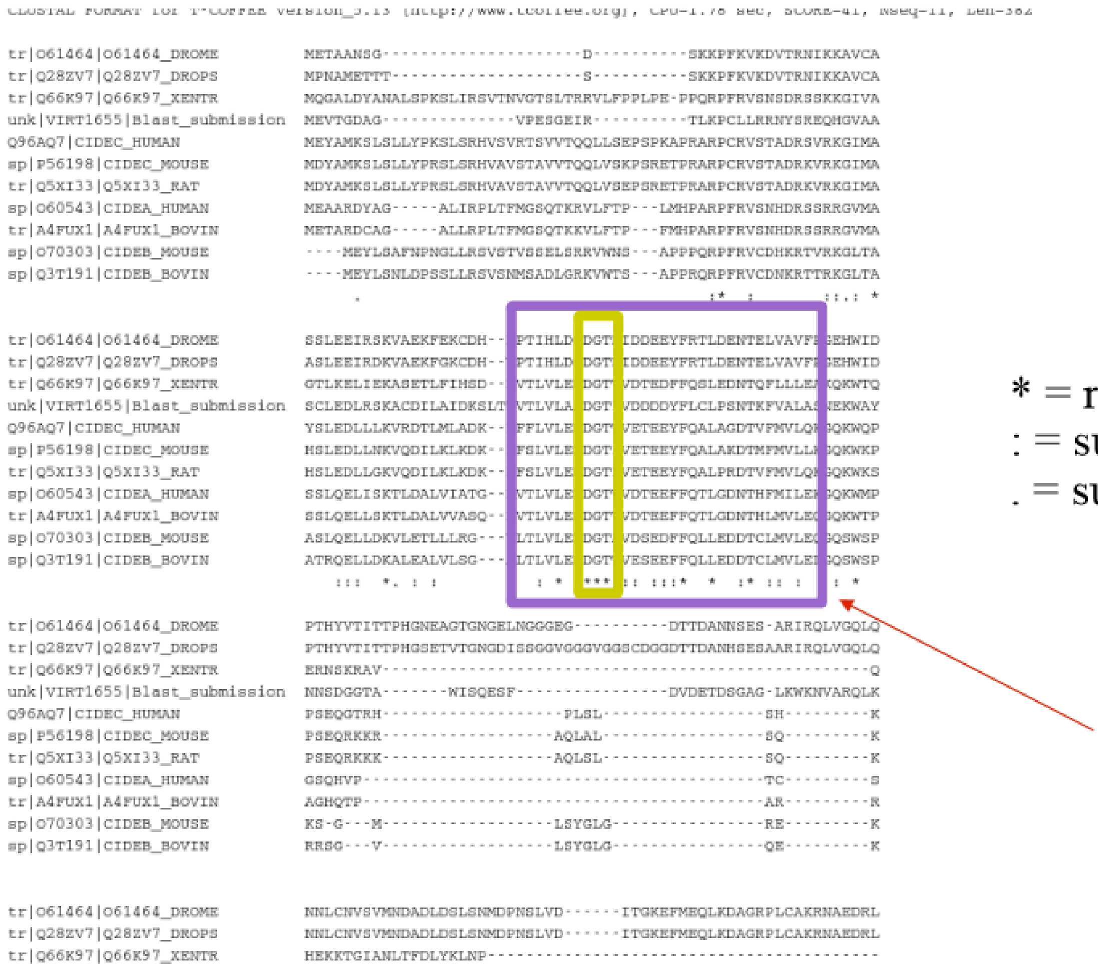
```
GUN1_TRIRE/427-455   HWGQCGGI---GYSGC--K-TCTSGTTCQYSNDYYSQCL
GUX1_TRIRE/481-509   HYGQCGGI---GYSGP--T-VCASGTTCQVLNPYYSQCL
GUX1_PHACH/484-512   QWGQCGGI---GYTGS--T-TCASPYTCHVLNPYYSQCY
GUX2_TRIRE/30-58     VWGQCGGQ---NWSGP--T-CCASGSTCVYSNDYYSQCL
GUN5_TRIRE/209-237   LYGQCGGA---GWTGP--T-TCQAPGTCKVQNQWYSQCL
GUNF_FUSOX/21-49     IWGQCGGN---GWTGA--T-TCASGLKCEKINDWYYQCV
GUX3_AGABI/24-52     VWGQCGGN---GWTGP--T-TCASGSTCVKQNDFYSQCL
Q01763/473-500       --SQCGGL---GYAGP--TgVCPSPYTCQALNIYYSQCI
GUX1_PENJA/505-533   DWAQCGGN---GWTGP--T-TCVSPYTCTKQNDWYSQCL
GUXC_FUSOX/482-510   QWGQCGGQ---NYSGP--T-TCKSPFTCKKINDFYSQCQ
GUX1_HUMGR/493-521   RWQQCGGI---GFTGP--T-QCEEPYICTKLNDWYSQCL
GUX1_NEUCR/484-512   HWAQCGGI---GFSGP--T-TCPEPYTCAKDHDIYSQCV
Q9Y894/23-53         PWGQCGGP---GWTGPttT-CCVTGCTCPVTND-YSQCL
PSBP_PORPU/26-54     LYEQCGGI---GFDGV--T-CCSEGLMCMKMGPYYSQCR
GUNB_FUSOX/29-57     VWAQCGGQ---NWSGT--P-CCTSGNKCVKLNDFYSQCQ
PSBP_PORPU/69-97     PYGQCGGM---NYSGK--T-MCSPGFKCVELNEFFSQCD
GUNK_FUSOX/339-370   AYYQCGGSKSAYPNGN--L-ACATGSKCVKQNEYYSQCV
PSBP_PORPU/128-156   EYAACGGE---MFMGA -K-CCKFGLVCYETSGKWSQCR
```

Extrait de Prosite, entrée PS00562

C-G-G-x(4,7)-G-x(3)-C-x(5)-C-x(3,5)-[NHG]-x-[FYWM]- x(2)-Q-C

```
      +-----------------+
      |           +-----|----------+
      |           |     |          |
xxxxxxCxxxxxxxxxxxCxxxxxCxxxxxxxxxxCx
      *****************************
```

Le profil de conservation issu de l’alignement multiple

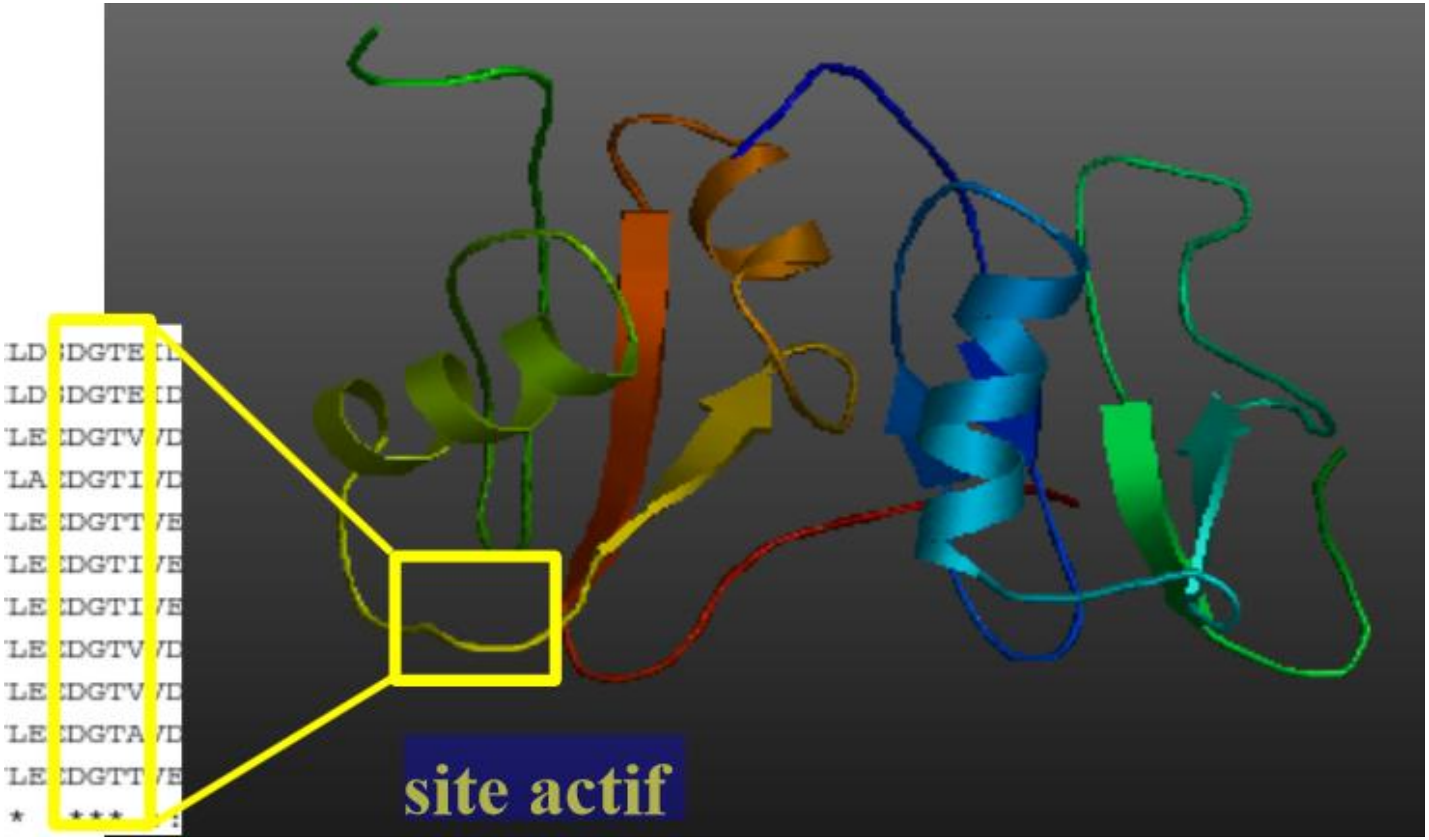


```
CLUSTAL FORMAT for T-COFFEE Version_5.13 [http://www.tcoffee.org], CPU=1.78 sec, SCORE=41, Nseq=11, Len=382

tr|O61464|O61464_DROME         METAANSG--------------------D----------SKKPFKVKDVTRNIKKAVCA
tr|Q28ZV7|Q28ZV7_DROPS         MPNAMETTT-------------------S----------SKKPFKVKDVTRNIKKAVCA
tr|Q66K97|Q66K97_XENTR         MQGALDYANALSPKSLIRSVTNVGTSLTRRVLFPPLPE-PPQRPFRVSNSDRSSKKGIVA
unk|VIRT1655|Blast_submission  MEVTGDAG-------------VPESGEIR---------TLKPCLLRRNYSREQHGVAA
Q96AQ7|CIDEC_HUMAN             MEYAMKSLSLLYPKSLSRHVSVRTSVVTQQLLSEPSPKAPRARPCRVSTADRSVRKGIMA
sp|P56198|CIDEC_MOUSE          MDYAMKSLSLLYPRSLSRHVAVSTAVVTQQLVSKPSRETPRARPCRVSTADRKVRKGIMA
tr|Q5XI33|Q5XI33_RAT           MDYAMKSLSLLYPRSLSRHVAVSTAVVTQQLVSEPSRETPRARPCRVSTADRKVRKGIMA
sp|O60543|CIDEA_HUMAN          MEAARDYAG-----ALIRPLTFMGSQTKRVLFTP---LMHPARPFRVSNHDRSSRRGVMA
tr|A4FUX1|A4FUX1_BOVIN         METARDCAG-----ALLRPLTFMGSQTKKVLFTP---FMHPARPFRVSNHDRSSRRGVMA
sp|O70303|CIDEB_MOUSE          ----MEYLSAFNPNGLLRSVSTVSSELSRRVWNS---APPPQRPFRVCDHKRTVRKGLTA
sp|Q3T191|CIDEB_BOVIN          ----MEYLSNLDPSSLLRSVSNMSADLGRKVWTS---APPRQRPFRVCDNKRTTRKGLTA
                                     .                                   :*  :       ::.: *

tr|O61464|O61464_DROME         SSLEEIRSKVAEKFEKCDH-- PTIHLD DGT IDDEEYFRTLDENTELVAVF EHWID
tr|Q28ZV7|Q28ZV7_DROPS         ASLEEIRDKVAEKFGKCDH-- PTIHLD DGT IDDEEYFRTLDENTELVAVF EHWID
tr|Q66K97|Q66K97_XENTR         GTLKELIEKASETLFIHSD-- VTLVLE DGT VDTEDFFQSLEDNTQFLLLE QKWTQ
unk|VIRT1655|Blast_submission  SCLEDLRSKACDILAIDKSLT VTLVLA DGT VDDDDYFLCLPSNTKFVALAS EKWAY
Q96AQ7|CIDEC_HUMAN             YSLEDLLLKVRDTLMLADK-- FFLVLE DGT VETEEYFQALAGDTVFMVLQ GQKWQP
sp|P56198|CIDEC_MOUSE          HSLEDLLNKVQDILKLKDK-- FSLVLE DGT VETEEYFQALAKDTMFMVLL GQKWKP
tr|Q5XI33|Q5XI33_RAT           HSLEDLLGKVQDILKLKDK-- FSLVLE DGT VETEEYFQALPRDTVFMVLQ GQKWKS
sp|O60543|CIDEA_HUMAN          SSLQELISKTLDALVIATG-- VTLVLE DGT VDTEEFFQTLGDNTHFMILE GQKWMP
tr|A4FUX1|A4FUX1_BOVIN         SSLQELLSKTLDALVVASQ-- VTLVLE DGT VDTEEFFQTLGDNTHLMVLE GQKWTP
sp|O70303|CIDEB_MOUSE          ASLQELLDKVLETLLLRG--- LTLVLE DGT VDSEDFFQLLEDDTCLMVLE GQSWSP
sp|Q3T191|CIDEB_BOVIN          ATRQELLDKALEALVLSG--- LTLVLE DGT VESEEFFQLLEDDTCLMVLE GQSWSP
                                  :::  *. : :          : *  *** :: :::*  *  :* :: :   : *

tr|O61464|O61464_DROME         PTHYVTITTPHGNEAGTGNGELNGGGEG----------DTTDANNSES-ARIRQLVGQLQ
tr|Q28ZV7|Q28ZV7_DROPS         PTHYVTITTPHGSETVTGNGDISSGGVGGGVGGSCDGGDTTDANHSESAARIRQLVGQLQ
tr|Q66K97|Q66K97_XENTR         ERNSKRAV---------------------------------------------------Q
unk|VIRT1655|Blast_submission  NNSDGGTA-------WISQESF----------------DVDETDSGAG-LKWKNVARQLK
Q96AQ7|CIDEC_HUMAN             PSEQGTRH-------------------PLSL-----------------SH---------K
sp|P56198|CIDEC_MOUSE          PSEQRKKR------------------AQLAL-----------------SQ---------K
tr|Q5XI33|Q5XI33_RAT           PSEQRKKK------------------AQLSL-----------------SQ---------K
sp|O60543|CIDEA_HUMAN          GSQHVP------------------------------------------TC---------S
tr|A4FUX1|A4FUX1_BOVIN         AGHQTP------------------------------------------AR---------R
sp|O70303|CIDEB_MOUSE          KS-G---M------------------LSYGLG----------------RE---------K
sp|Q3T191|CIDEB_BOVIN          RRSG---V------------------LSYGLG----------------QE---------K


tr|O61464|O61464_DROME         NNLCNVSVMNDADLDSLSNMDPNSLVD------ITGKEFMEQLKDAGRPLCAKRNAEDRL
tr|Q28ZV7|Q28ZV7_DROPS         NNLCNVSVMNDADLDSLSNMDPNSLVD------ITGKEFMEQLKDAGRPLCAKRNAEDRL
tr|Q66K97|Q66K97_XENTR         HEKKTGIANLTFDLYKLNP-----------------------------------------
```

* = résidu parfaitement conservé
: = substitution conservative
. = substitution semi-conservative

Qu'est ce qu'il y a de si spécial ici ???

SITE ACTIF : CELL DEATH ACTIVATOR PROTEIN FAMILY



```
GHEGVGKVVKLGAGA
GHEKKGYFEDRGPSA
GHEGYGGRSRGGGYS
GHEFEGPKGCGALYI
GHELRGTTFMPALEC
```

```
GHE--G---------
GHE--G-----G---
```

Consensus 100%

Consensus 60 %

```
GHE-x(2)-G-x(5)-[GA]
```

pattern

signature

profil

précision

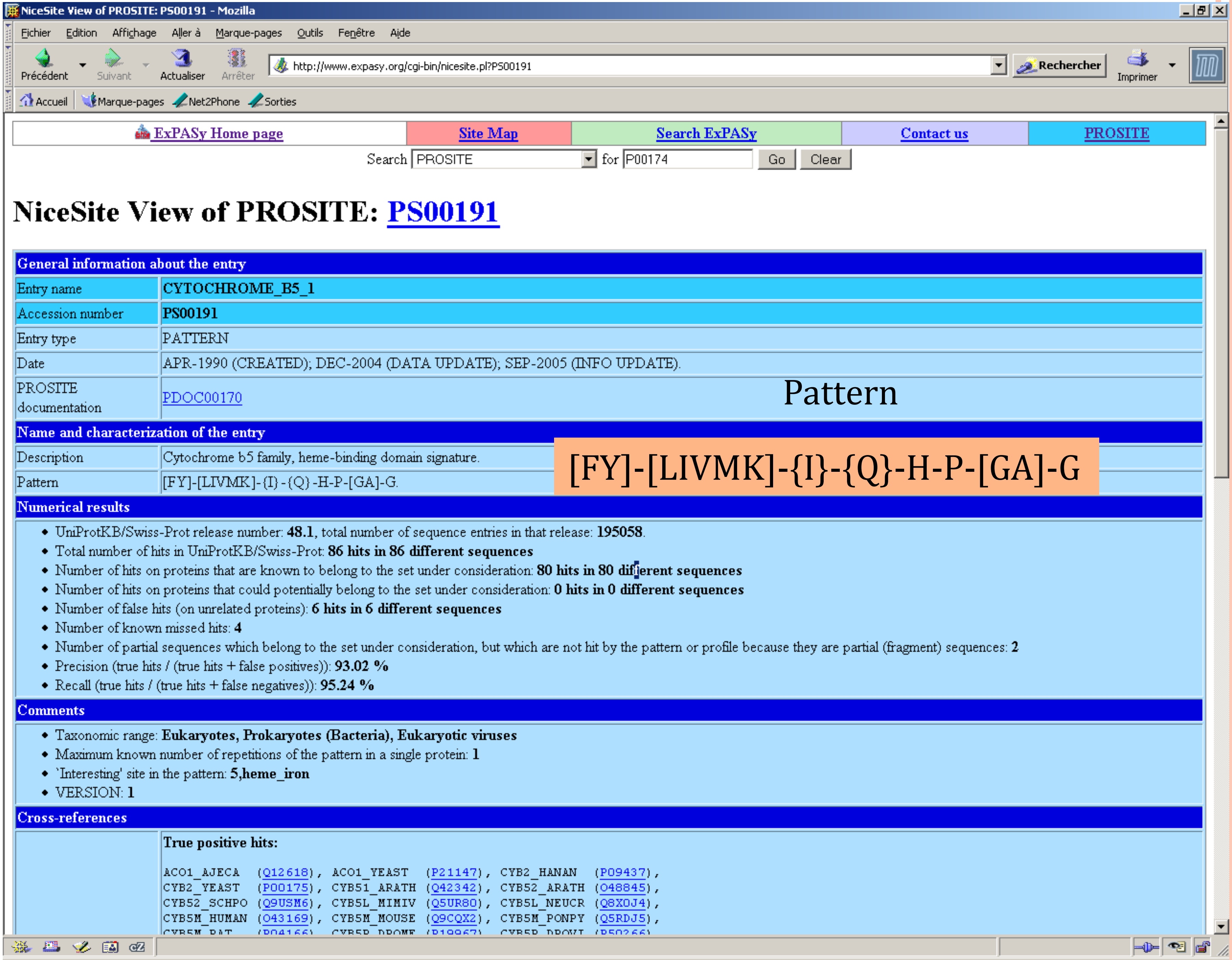
sensibilité

```
<A-x-[ST](2)-x(0,1)-[APTL]-x(4,10)-C-{V}
```

`<A`	en N terminal
`x`	= n’importe quel AA
`ST](2)`	= Ser ou Thr 2 fois
`x(0,1)`	1 aa ou aucun
`x(4,10)`	entre 4 et 10 aa quelconques
`{V}`	tout sauf une Val

Code IUPAC pour les nucléotides

Code	Description
A	Adénine
C	Cytosine
G	Guanine
T	Thymine
U	Uracile
R	Purine (A ou G)
Y	Pyrimidine (C, T, ou U)
M	C ou A
K	T, U, ou G
W	T, U, ou A
S	C ou G
B	C, T, U, ou G (pas A)
D	A, T, U, ou G (pas C)
H	A, T, U, or C (pas G)
V	A, C, or G (pas T, pas U)
N	Toutes les bases (A, C, G, T, ou U)


NiceSite View of PROSITE: PS00191 - Mozilla
Fichier Edition Affichage Aller à Marque-pages Outils Fenêtre Aide
Précédent
Suivant
Actualiser
Arrêter
http://www.expasy.org/cgi-bin/nicesite.pl?PS00191
Rechercher
Imprimer
Accueil
Marque-pages
Net2Phone
Sorties
ExPASy Home page
Site Map
Search ExPASy
Contact us
PROSITE
Search PROSITE for P00174 Go Clear

NiceSite View of PROSITE: PS00191

General information about the entry	
Entry name	**CYTOCHROME_B5_1**
Accession number	**PS00191**
Entry type	PATTERN
Date	APR-1990 (CREATED); DEC-2004 (DATA UPDATE); SEP-2005 (INFO UPDATE).
PROSITE documentation	PDOC00170

Name and characterization of the entry	
Description	Cytochrome b5 family, heme-binding domain signature.
Pattern	[FY]-[LIVMK]-{I}-{Q}-H-P-[GA]-G.

Pattern

[FY]-[LIVMK]-{I}-{Q}-H-P-[GA]-G

Numerical results

- UniProtKB/Swiss-Prot release number: **48.1**, total number of sequence entries in that release: **195058**.
- Total number of hits in UniProtKB/Swiss-Prot: **86 hits in 86 different sequences**
- Number of hits on proteins that are known to belong to the set under consideration: **80 hits in 80 different sequences**
- Number of hits on proteins that could potentially belong to the set under consideration: **0 hits in 0 different sequences**
- Number of false hits (on unrelated proteins): **6 hits in 6 different sequences**
- Number of known missed hits: **4**
- Number of partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences: **2**
- Precision (true hits / (true hits + false positives)): **93.02 %**
- Recall (true hits / (true hits + false negatives)): **95.24 %**

Comments

- Taxonomic range: **Eukaryotes, Prokaryotes (Bacteria), Eukaryotic viruses**
- Maximum known number of repetitions of the pattern in a single protein: **1**
- `Interesting' site in the pattern: **5,heme_iron**
- VERSION: **1**

Cross-references

True positive hits:

```
ACO1_AJECA  (Q12618), ACO1_YEAST  (P21147), CYB2_HANAN  (P09437),
CYB2_YEAST  (P00175), CYB51_ARATH (Q42342), CYB52_ARATH (O48845),
CYB52_SCHPO (Q9USM6), CYB5L_MIMIV (Q5UR80), CYB5L_NEUCR (Q8X0J4),
CYB5M_HUMAN (O43169), CYB5M_MOUSE (Q9CQX2), CYB5M_PONPY (Q5RDJ5),
CYB5M_RAT   (P04166), CYB5P_DROME (P19967), CYB5P_DROVI (P50266),
```

Alignement

AATTGA
AGGTCC
AGGATG
AGGCGT

Matrice de position

	1	2	3	4	5	6
A	4	1	0	1	0	1
C	0	0	0	1	1	1
G	0	3	3	0	2	1
T	0	0	1	2	1	1

Matrice de fréquences

	1	2	3	4	5	6
A	1	0.25	0	0.25	0	0.25
C	0	0	0	0.25	0.25	0.25
G	0	0.75	0.75	0	0.50	0.25
T	0	0	0.25	0.50	0.25	0.25

Matrice de fréquences $\xrightarrow{\log\left[\frac{f_b}{p_b}\right]}$ Matrice de poids de position (Position Weight Matrix ou Position Specific Scoring Matrix PSSM)

	1	**2**	**3**	**4**	**5**	**6**
A	1.2	0	-1.6	0	-1.6	0
C	-1.6	-1.6	-1.6	0	0	0
G	-1.6	0.96	0.96	-1.6	0.59	0
T	-1.6	-1.6	0	0.59	0	0

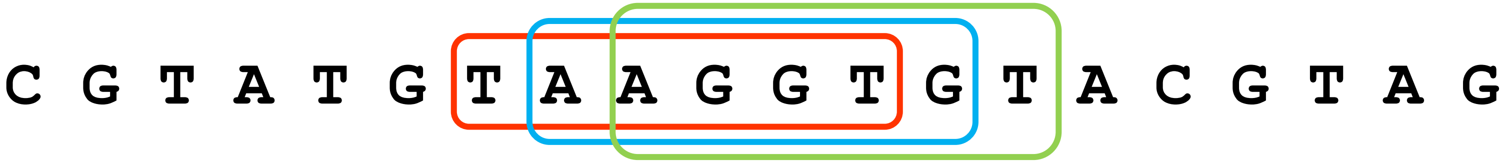
Soumission d'une séquence d'intérêt dans les bases de données du type PFAM

C G T A T G T A A G G T G T A C G T A G

Pour trouver si la séquence contient un motif, les bases de données appliquent sur les séquences soumises les matrices qu'elles ont correspondant toutes à des motifs décrits, et notamment la matrice ci-dessous.

	1	**2**	**3**	**4**	**5**	**6**
A	1.2	0	-1.6	0	-1.6	0
C	-1.6	-1.6	-1.6	0	0	0
G	-1.6	0.96	0.96	-1.6	0.59	0
T	-1.6	-1.6	0	0.59	0	0

Calcul de score par fenêtre glissante



C G T A T G T A A G G T G T A C G T A G

C G T A T G $\boxed{\text{T A A G G T}}$ G T A C G T A G

	1	2	3	4	5	6
A	1.2	$\boxed{0}$	$\boxed{-1.6}$	0	-1.6	0
C	-1.6	-1.6	-1.6	0	0	0
G	-1.6	0.96	0.96	$\boxed{-1.6}$	$\boxed{0.59}$	0
T	$\boxed{-1.6}$	-1.6	0	0.59	0	$\boxed{0}$

Score = -4.21

C G T A T G T $\boxed{\text{A A G G T G}}$ T A C G T A G

	1	2	3	4	5	6
A	$\boxed{1.2}$	$\boxed{0}$	-1.6	0	-1.6	0
C	-1.6	-1.6	-1.6	0	0	0
G	-1.6	0.96	$\boxed{0.96}$	$\boxed{-1.6}$	0.59	$\boxed{0}$
T	-1.6	-1.6	0	0.59	$\boxed{0}$	0

Score = 0.56

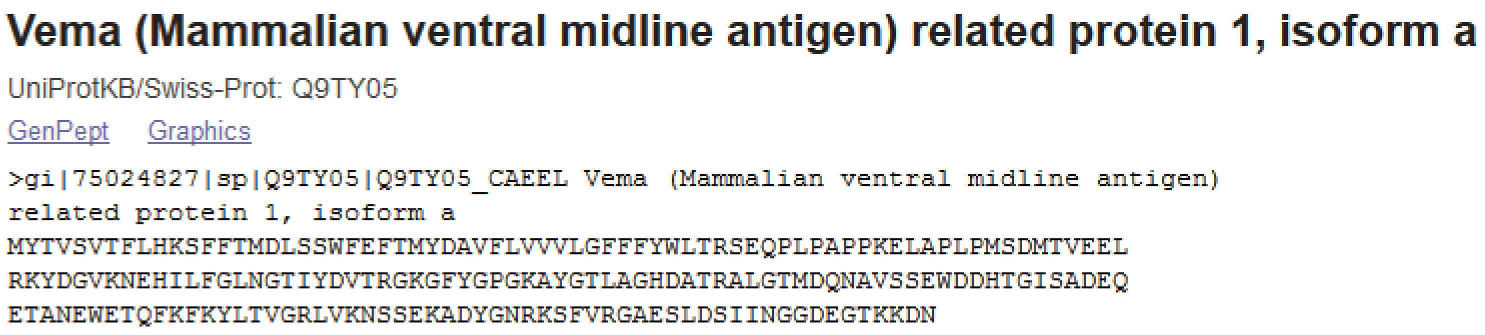
CGTATGTAAGGTGTACGTAG

	1	2	3	4	5	6
A	1.2	0	-1.6	0	-1.6	0
C	-1.6	-1.6	-1.6	0	0	0
G	-1.6	0.96	0.96	-1.6	0.59	0
T	-1.6	-1.6	0	0.59	0	0

Score = 4.3

Si dans la base de données interrogée, le seuil minimal de détection de motif est par exemple S = 4 alors seule la séquence AGGTGT sera considérée comme pattern

Exemple : recherche de motif sur une séquence VEM-1 de *Caenorhabditis elegans*

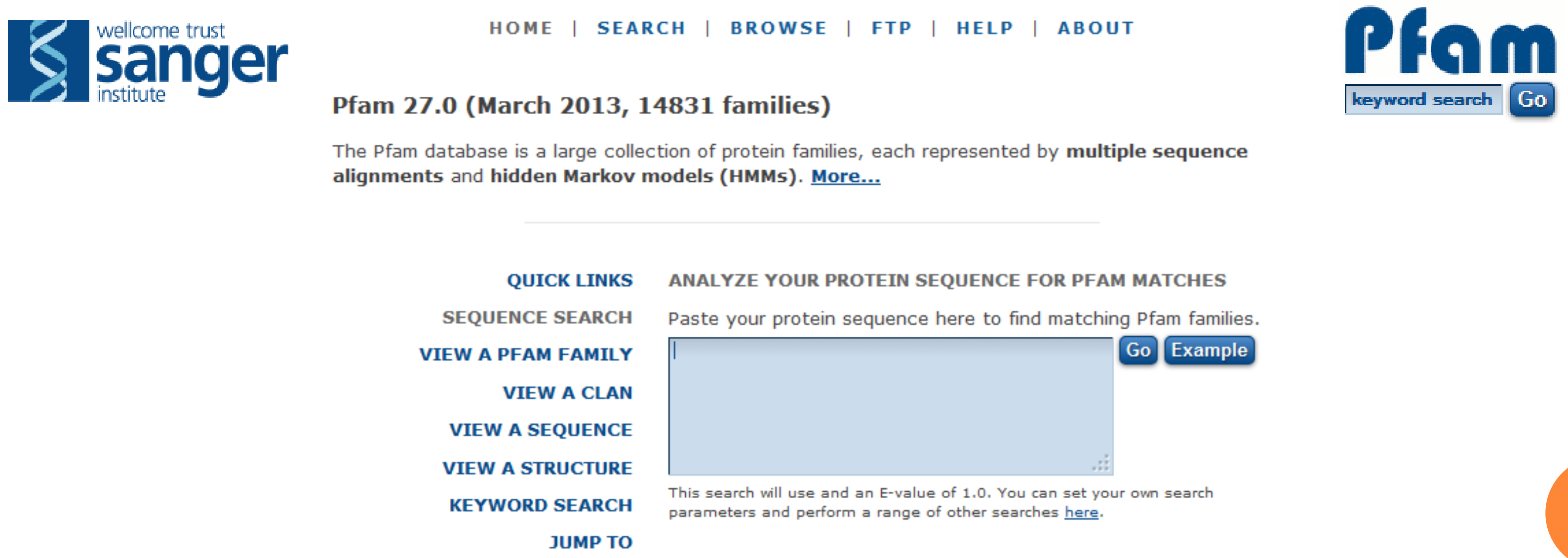


Vema (Mammalian ventral midline antigen) related protein 1, isoform a

UniProtKB/Swiss-Prot: Q9TY05

GenPept Graphics

```
>gi|75024827|sp|Q9TY05|Q9TY05_CAEEL Vema (Mammalian ventral midline antigen)
related protein 1, isoform a
MYTVSVTFLHKSFFTMDLSSWFEFTMYDAVFLVVVLGFFFYWLTRSEQPLPAPPKELAPLPMSDMTVEEL
RKYDGVKNEHILFGLNGTIYDVTRGKGFYGPGKAYGTLAGHDATRALGTMDQNAVSSEWDDHTGISADEQ
ETANEWETQFKFKYLTVGRLVKNSSEKADYGNRKSFVRGAESLDSIINGGDEGTKKDN
```



Pfam 27.0 (March 2013, 14831 families)

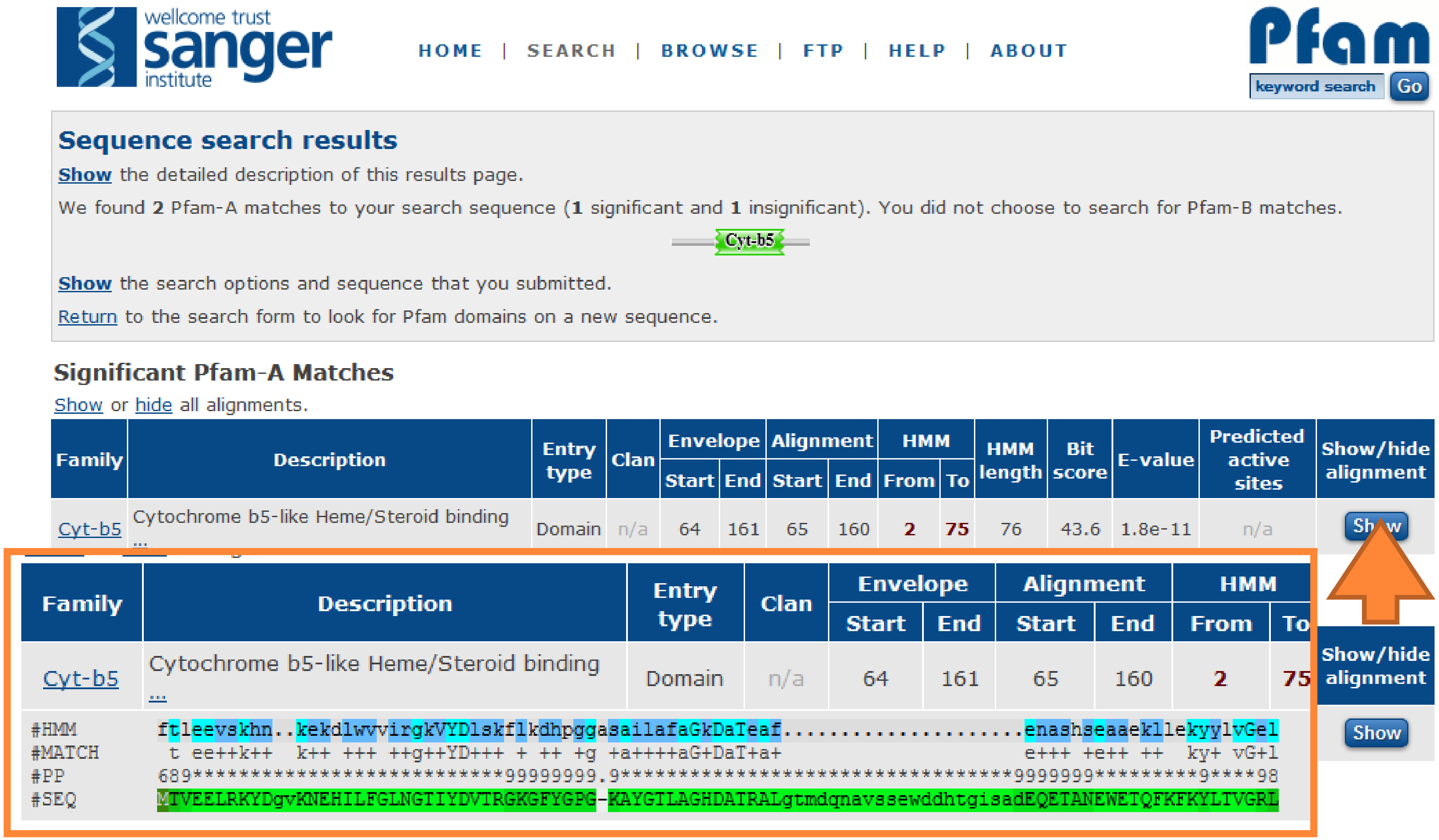
The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. More...

ANALYZE YOUR PROTEIN SEQUENCE FOR PFAM MATCHES

Paste your protein sequence here to find matching Pfam families.

This search will use and an E-value of 1.0. You can set your own search parameters and perform a range of other searches here.

Exemple : recherche de motif sur une séquence VEM-1 de *Caenorhabditis elegans*



Sequence search results

Show the detailed description of this results page.

We found **2** Pfam-A matches to your search sequence (**1** significant and **1** insignificant). You did not choose to search for Pfam-B matches.

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

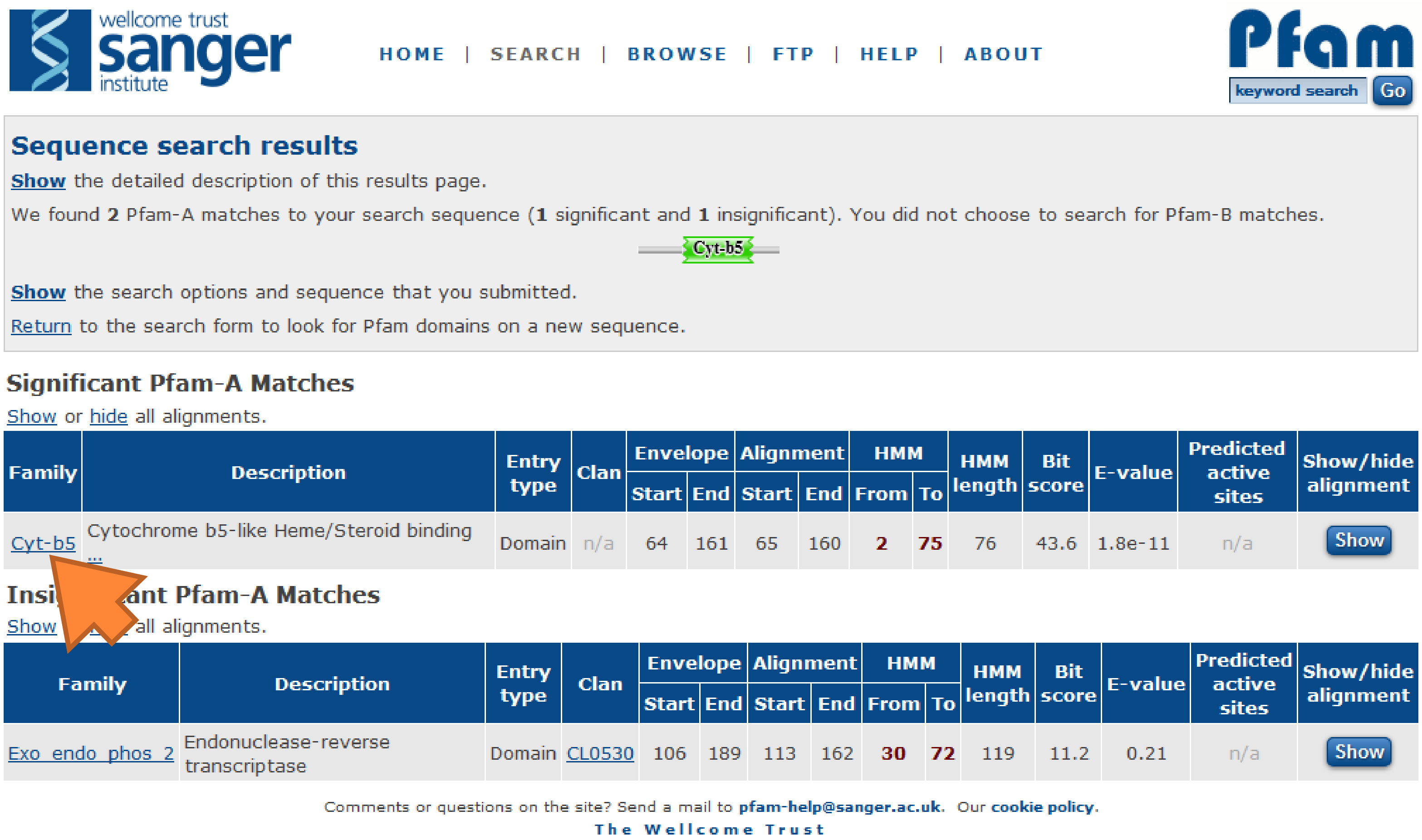
Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Cyt-b5	Cytochrome b5-like Heme/Steroid binding ...	Domain	n/a	64	161	65	160	2	75	76	43.6	1.8e-11	n/a	Show

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		Show/hide alignment
				Start	End	Start	End	From	To	
Cyt-b5	Cytochrome b5-like Heme/Steroid binding ...	Domain	n/a	64	161	65	160	2	75	Show

```
#HMM      ftleevskhn..kekdlwvvirgkVYDlskflkdhpggasailafaGkDaTeaf.....................enashseaaekllekyylvGel
#MATCH     t ee++k++  k++ +++ ++g++YD+++ + ++ +g +a+++++aG+DaT+a+                     e+++ +e++ ++  ky+ vG+l
#PP       689***************************99999999.9**********************************9999999*********9*****98
#SEQ      MTVEELRKYDgvKNEHILFGLNGTIYDVTRGKGFYGPG-KAYGTLAGHDATRALgtmdqnavssewddhtgisadEQETANEWETQFKFKYLTVGRL
```

Exemple: recherche de motif sur une séquence VEM-1 de *Caenorhabditis elegans*


wellcome trust sanger institute
HOME | SEARCH | BROWSE | FTP | HELP | ABOUT
Pfam
keyword search
Go
Cyt-b5

Sequence search results

Show the detailed description of this results page.

We found **2** Pfam-A matches to your search sequence (**1** significant and **1** insignificant). You did not choose to search for Pfam-B matches.

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Cyt-b5	Cytochrome b5-like Heme/Steroid binding ...	Domain	n/a	64	161	65	160	**2**	**75**	76	43.6	1.8e-11	n/a	Show

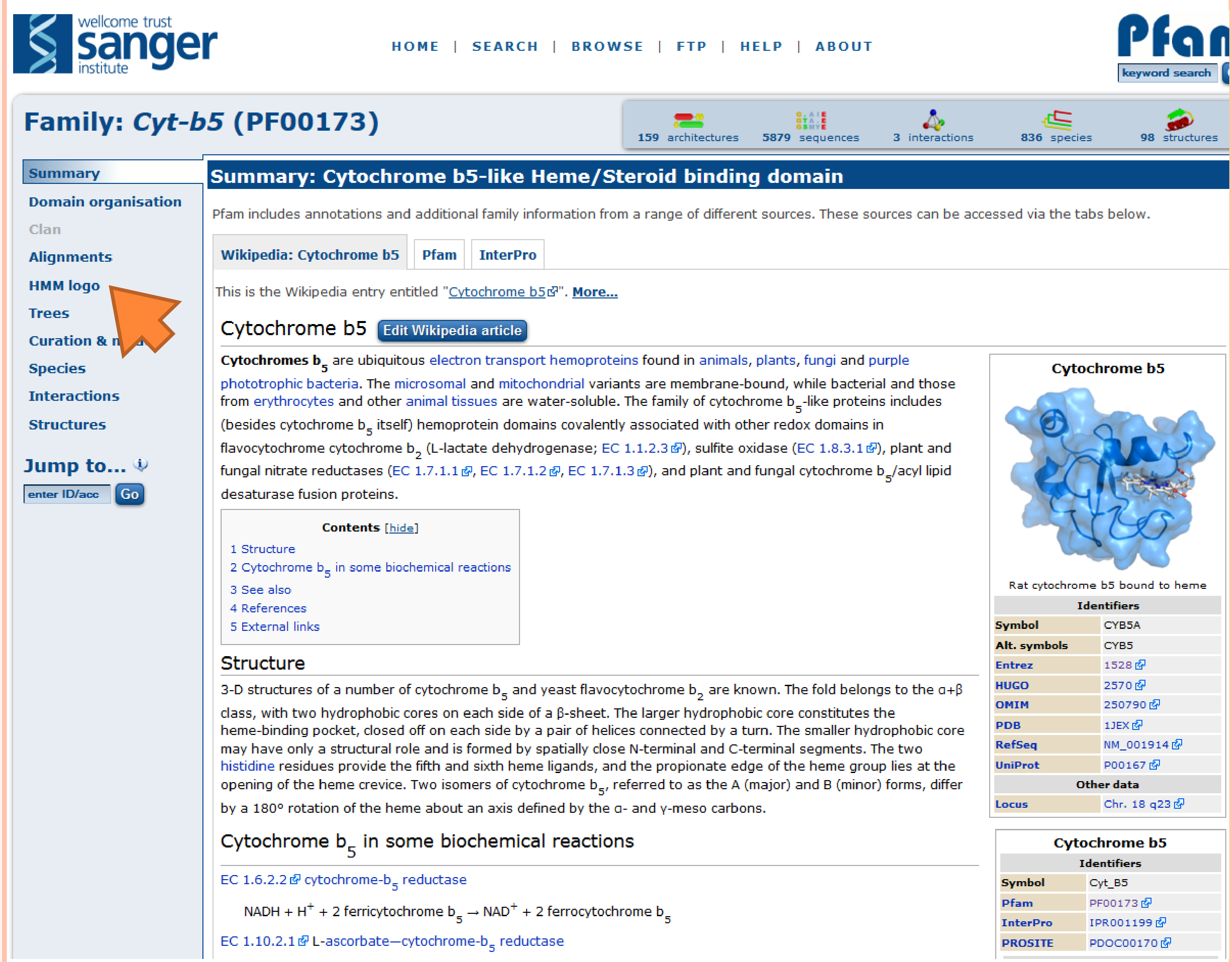
Insignificant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Exo_endo_phos_2	Endonuclease-reverse transcriptase	Domain	CL0530	106	189	113	162	**30**	**72**	119	11.2	0.21	n/a	Show

Comments or questions on the site? Send a mail to **pfam-help@sanger.ac.uk**. Our **cookie policy**.

The Wellcome Trust



Family: *Cyt-b5* (PF00173)

Summary: Cytochrome b5-like Heme/Steroid binding domain

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

This is the Wikipedia entry entitled "Cytochrome b5". More...

Cytochrome b5

Cytochromes b$_5$ are ubiquitous electron transport hemoproteins found in animals, plants, fungi and purple phototrophic bacteria. The microsomal and mitochondrial variants are membrane-bound, while bacterial and those from erythrocytes and other animal tissues are water-soluble. The family of cytochrome b$_5$-like proteins includes (besides cytochrome b$_5$ itself) hemoprotein domains covalently associated with other redox domains in flavocytochrome cytochrome b$_2$ (L-lactate dehydrogenase; EC 1.1.2.3), sulfite oxidase (EC 1.8.3.1), plant and fungal nitrate reductases (EC 1.7.1.1, EC 1.7.1.2, EC 1.7.1.3), and plant and fungal cytochrome b$_5$/acyl lipid desaturase fusion proteins.

Contents [hide]

Structure

3-D structures of a number of cytochrome b$_5$ and yeast flavocytochrome b$_2$ are known. The fold belongs to the α+β class, with two hydrophobic cores on each side of a β-sheet. The larger hydrophobic core constitutes the heme-binding pocket, closed off on each side by a pair of helices connected by a turn. The smaller hydrophobic core may have only a structural role and is formed by spatially close N-terminal and C-terminal segments. The two histidine residues provide the fifth and sixth heme ligands, and the propionate edge of the heme group lies at the opening of the heme crevice. Two isomers of cytochrome b$_5$, referred to as the A (major) and B (minor) forms, differ by a 180° rotation of the heme about an axis defined by the α- and γ-meso carbons.

Cytochrome b$_5$ in some biochemical reactions

EC 1.6.2.2 cytochrome-b$_5$ reductase

NADH + H$^+$ + 2 ferricytochrome b$_5$ → NAD$^+$ + 2 ferrocytochrome b$_5$

EC 1.10.2.1 L-ascorbate—cytochrome-b$_5$ reductase

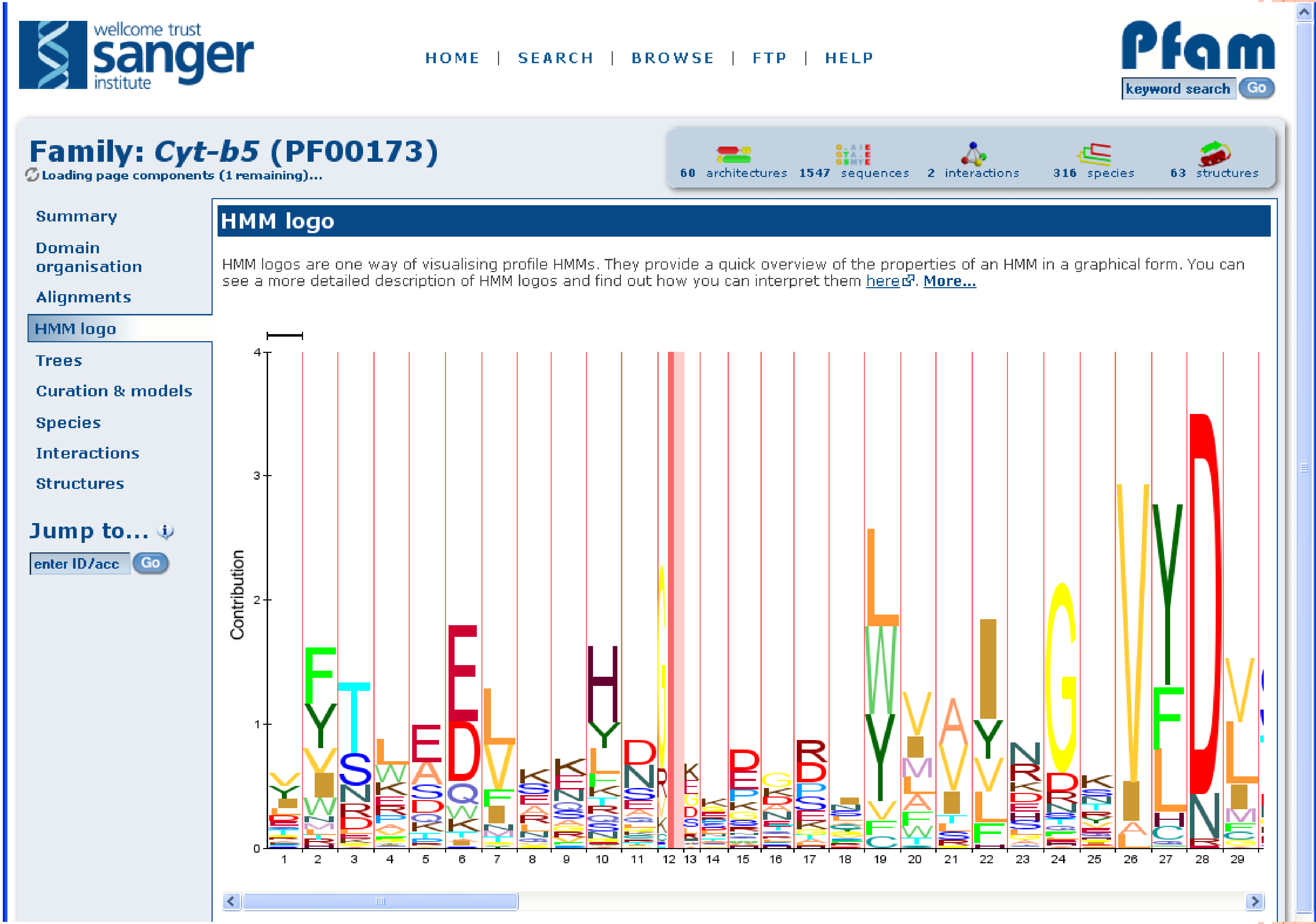
Cytochrome b5

Rat cytochrome b5 bound to heme

Identifiers	
Symbol	CYB5A
Alt. symbols	CYB5
Entrez	1528
HUGO	2570
OMIM	250790
PDB	1JEX
RefSeq	NM_001914
UniProt	P00167
Other data	
Locus	Chr. 18 q23

Cytochrome b5

Identifiers	
Symbol	Cyt_B5
Pfam	PF00173
InterPro	IPR001199
PROSITE	PDOC00170



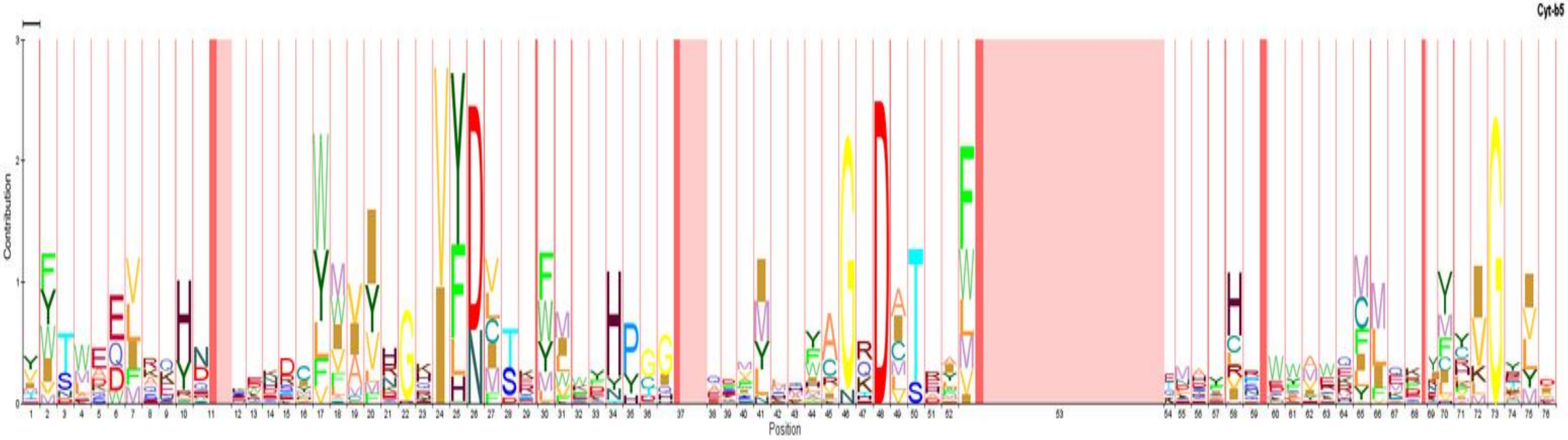
Family: *Cyt-b5* (PF00173)

Loading page components (1 remaining)...

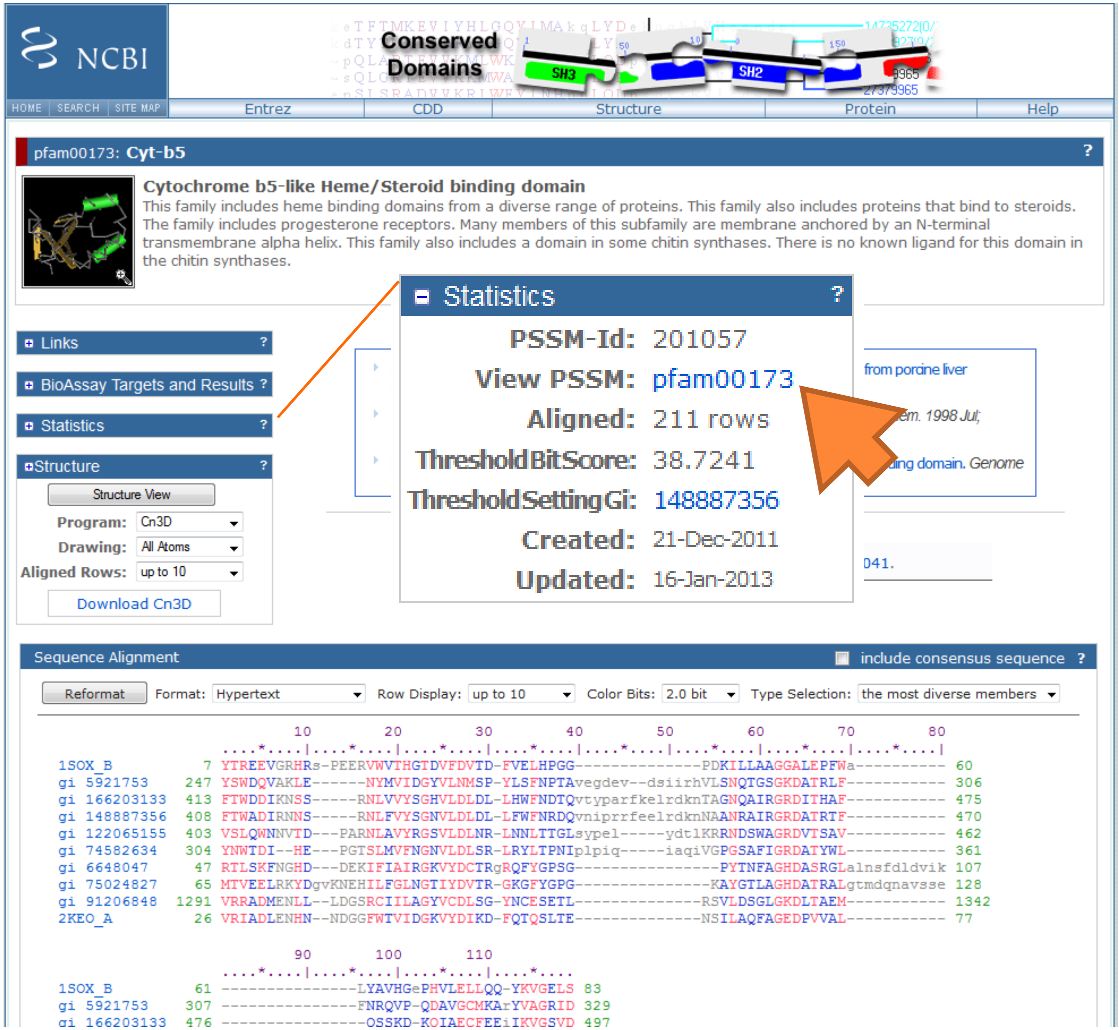
HMM logo

HMM logos are one way of visualising profile HMMs. They provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them here. More...

Web logo entier de cyt-b5



Description de cyt-b5 dans la Conserved Domains Database du NCBI



pfam00173: **Cyt-b5** ?

Cytochrome b5-like Heme/Steroid binding domain

This family includes heme binding domains from a diverse range of proteins. This family also includes proteins that bind to steroids. The family includes progesterone receptors. Many members of this subfamily are membrane anchored by an N-terminal transmembrane alpha helix. This family also includes a domain in some chitin synthases. There is no known ligand for this domain in the chitin synthases.

Program: Cn3D
Drawing: All Atoms
Aligned Rows: up to 10

from porcine liver
em. 1998 Jul;
ing domain. Genome
041.

Statistics ?

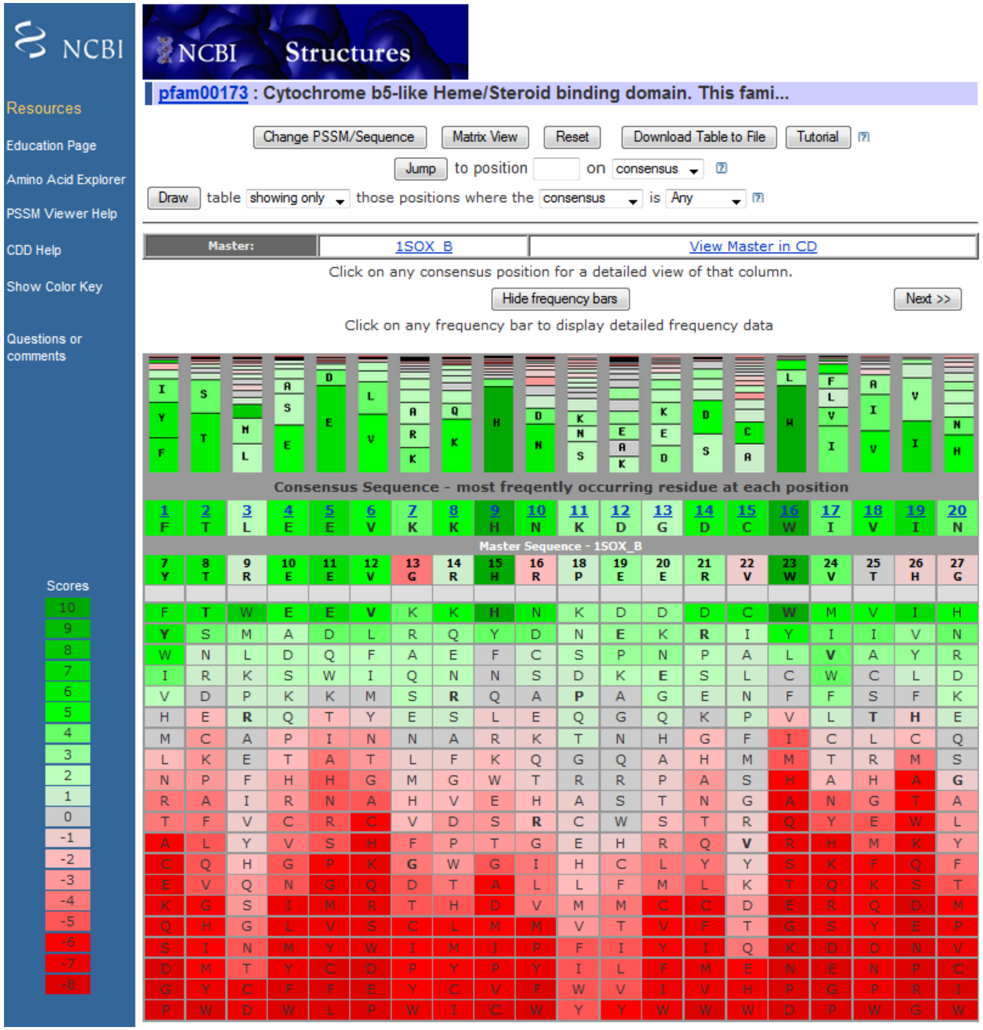
PSSM-Id:	201057
View PSSM:	pfam00173
Aligned:	211 rows
Threshold Bit Score:	38.7241
Threshold Setting Gi:	148887356
Created:	21-Dec-2011
Updated:	16-Jan-2013

Sequence Alignment

Format: Hypertext · Row Display: up to 10 · Color Bits: 2.0 bit · Type Selection: the most diverse members

```
                         10        20        30        40        50        60        70        80
                 ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
1SOX_B         7 YTREEVGRHRs-PEERVWVTHGTDVFDVTD-FVELHPGG--------------PDKILLAAGGALEPFWa----------- 60
gi 5921753   247 YSWDQVAKLE------NYMVIDGYVLNMSP-YLSFNPTAvegdev--dsiirhVLSNQTGSGKDATRLF----------- 306
gi 166203133 413 FTWDDIKNSS-----RNLVVYSGHVLDLDL-LHWFNDTQvtyparfkelrdknTAGNQAIRGRDITHAF----------- 475
gi 148887356 408 FTWADIRNNS-----RNLFVYSGNVLDLDL-LFWFNRDQvniprrfeelrdknNAANRAIRGRDATRTF----------- 470
gi 122065155 403 VSLQWNNVTD---PARNLAVYRGSVLDLNR-LNNLTTGLsypel-----ydtlKRRNDSWAGRDVTSAV----------- 462
gi 74582634  304 YNWTDI--HE---PGTSLMVFNGNVLDLSR-LRYLTPNIplpiq-----iaqiVGPGSAFIGRDATYWL----------- 361
gi 6648047    47 RTLSKFNGHD---DEKIFIAIRGKVYDCTRgRQFYGPSG----------------PYTNFAGHDASRGLalnsfdldvik 107
gi 75024827   65 MTVEELRKYDgvKNEHILFGLNGTIYDVTR-GKGFYGPG---------------KAYGTLAGHDATRALgtmdqnavsse 128
gi 91206848 1291 VRRADMENLL--LDGSRCIILAGYVCDLSG-YNCESETL--------------RSVLDSGLGKDLTAEM----------- 1342
2KEO_A        26 VRIADLENHN--NDGGFWTVIDGKVYDIKD-FQTQSLTE--------------NSILAQFAGEDPVVAL----------- 77

                         90       100       110
                 ....*....|....*....|....*....|....*....
1SOX_B        61 ---------------LYAVHGePHVLELLQQ-YKVGELS 83
gi 5921753   307 ---------------FNRQVP-QDAVGCMKArYVAGRID 329
gi 166203133 476 ----------------QSSKD-KQIAECFEEiIKVGSVD 497
```

PSSM



NCBI

Resources

Education Page

Amino Acid Explorer

PSSM Viewer Help

CDD Help

Show Color Key

Questions or comments

Scores

10

9

8

7

6

5

4

3

2

1

0

-1

-2

-3

-4

-5

-6

-7

-8

NCBI Structures

pfam00173 : Cytochrome b5-like Heme/Steroid binding domain. This fami...

Change PSSM/Sequence | Matrix View | Reset | Download Table to File | Tutorial

Jump to position on consensus

Draw table showing only those positions where the consensus is Any

Master: 1SOX_B View Master in CD

Click on any consensus position for a detailed view of that column.

Hide frequency bars | Next >>

Click on any frequency bar to display detailed frequency data

Consensus Sequence - most freqently occurring residue at each position

Master Sequence - 1SOX_B

1 F	2 T	3 L	4 E	5 E	6 V	7 K	8 K	9 H	10 N	11 K	12 D	13 G	14 D	15 C	16 W	17 I	18 V	19 I	20 N
7 Y	8 T	9 R	10 E	11 E	12 V	13 G	14 R	15 H	16 R	18 P	19 E	20 E	21 R	22 V	23 W	24 V	25 T	26 H	27 G
F	T	W	E	E	V	K	K	H	N	K	D	D	D	C	W	M	V	I	H
Y	S	M	A	D	L	R	Q	Y	D	N	E	K	R	I	Y	I	I	V	N
W	N	L	D	Q	F	A	E	F	C	S	P	N	P	A	L	V	A	Y	R
I	R	K	S	W	I	Q	N	N	S	D	K	E	S	L	C	W	C	L	D
V	D	P	K	K	M	S	R	Q	A	P	A	G	E	N	F	F	S	F	K
H	E	R	Q	T	Y	E	S	L	E	Q	G	Q	K	P	V	L	T	H	E
M	C	A	P	I	N	N	A	R	K	T	N	H	G	F	I	C	L	C	Q
L	K	E	T	A	T	L	F	K	Q	G	Q	A	H	M	M	T	R	M	S
N	P	F	H	H	G	M	G	W	T	R	R	P	A	S	H	A	H	A	G
R	A	I	R	N	A	H	V	E	H	A	S	T	N	G	A	N	G	T	A
T	F	V	C	R	C	V	D	S	R	C	W	S	T	R	Q	Y	E	W	L
A	L	Y	V	S	H	F	P	T	G	E	H	R	Q	V	R	H	M	K	Y
C	Q	H	G	P	K	G	W	G	I	H	C	L	Y	Y	S	K	F	Q	F
E	V	Q	N	G	Q	D	T	A	L	L	F	M	L	K	T	Q	K	S	T
K	G	S	I	M	R	T	H	D	V	M	M	C	C	D	E	R	Q	D	M
Q	H	G	L	V	S	C	L	M	M	V	T	V	C	T	G	S	Y	E	P
S	I	N	M	Y	W	I	M	I	P	F	I	Y	I	Q	K	D	D	N	V
D	M	T	Y	C	D	P	Y	P	Y	I	L	F	M	E	N	E	N	P	C
G	Y	C	F	F	E	Y	C	V	F	W	V	I	V	H	P	G	P	R	I
P	W	D	W	L	P	W	I	C	W	Y	Y	W	W	W	D	P	W	G	W

PSI-BLAST (Position-Specific Iterative)

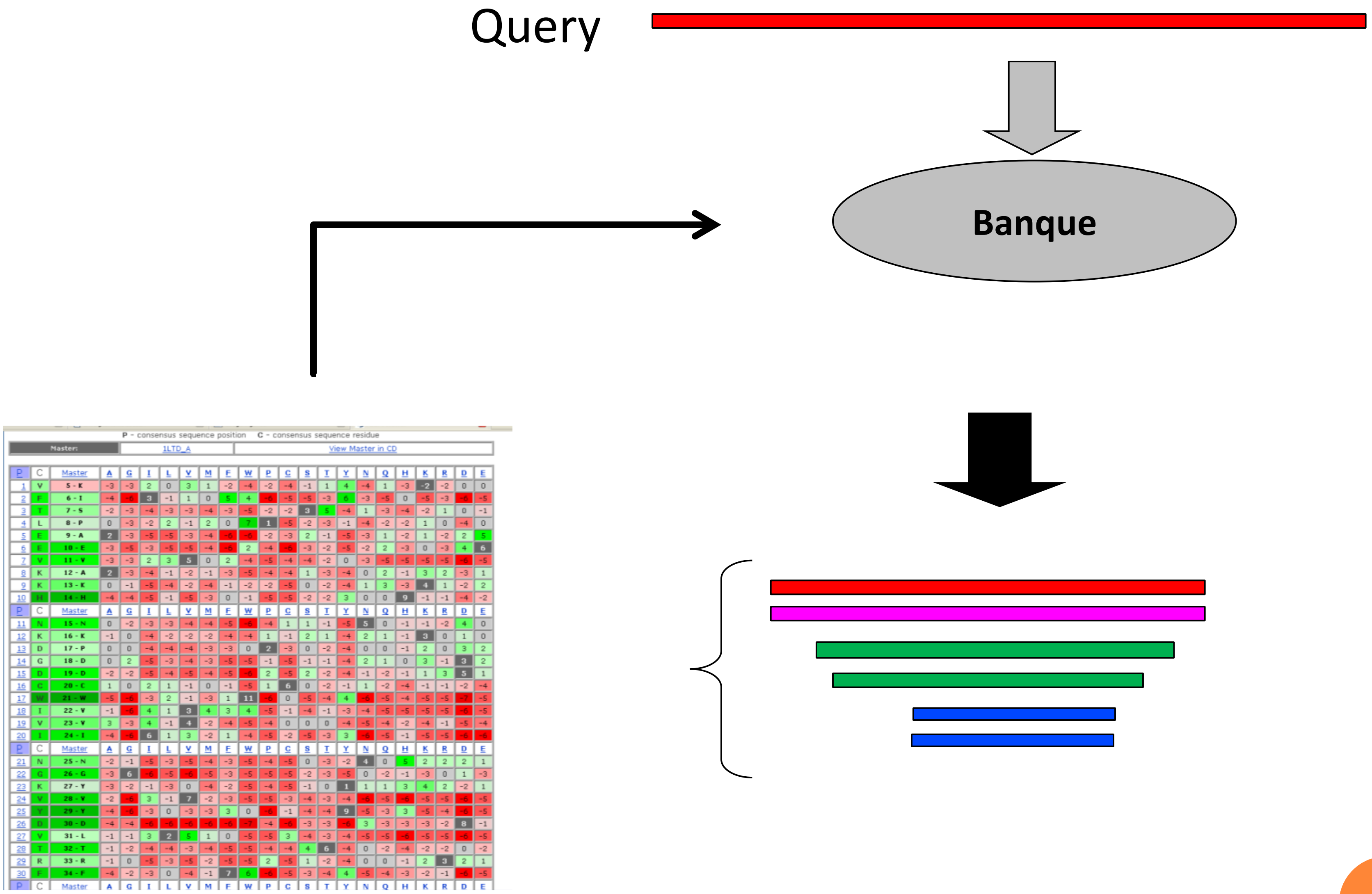
- alignements multiples de hits ayant les meilleurs scores dans un blast classique
- génération d'un profil en calculant un score pour chacune des positions de l'alignement (PSSM)
- utilisation façon itérative de ce profil pour faire de nouvelles recherches et affinage à chaque itération

PHI-BLAST (Pattern Hit Initiated BLAST)

Pattern donné par l'utilisateur puis PSI-BLAST

Intérêt : recherche de familles de protéines
détecter des membres que BLAST ne trouve pas

PSI-BLAST (Position-Specific Iterative)



Query

Banque

P - consensus sequence position   C - consensus sequence residue

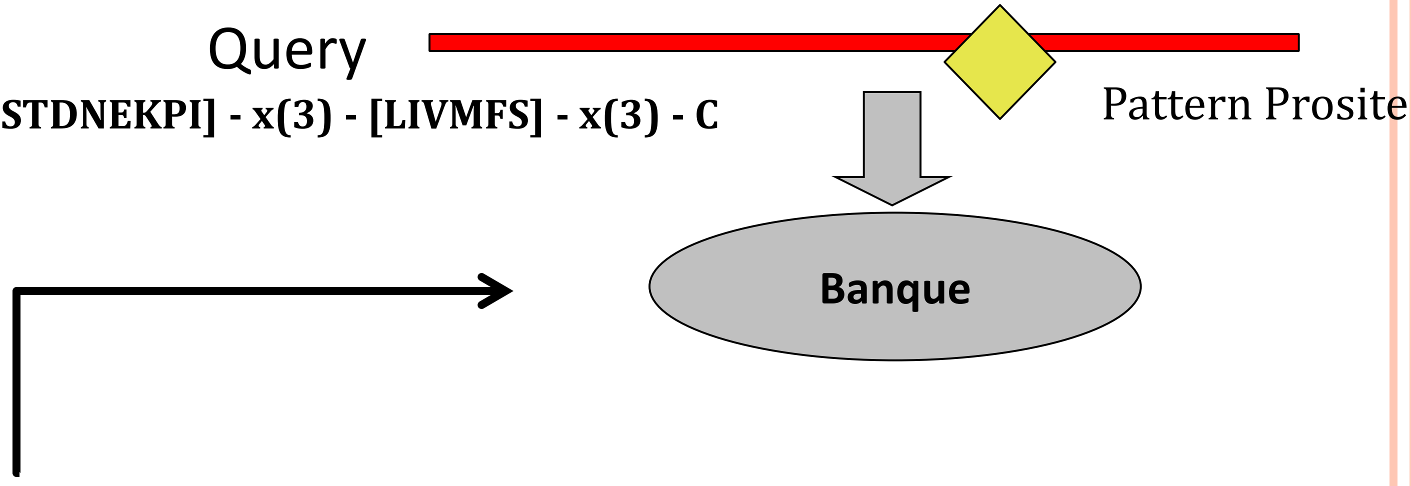
Master: 1LTD_A   View Master in CD

P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
1	V	5 - K	-3	-3	2	0	3	1	-2	-4	-2	-4	-1	1	4	-4	1	-3	-2	-2	0	0
2	F	6 - I	-4	-6	3	-1	1	0	5	4	-6	-5	-5	-3	6	-3	-5	0	-5	-3	-6	-5
3	T	7 - S	-2	-3	-4	-3	-3	-4	-3	-5	-2	-2	3	5	-4	1	-3	-4	-2	1	0	-1
4	L	8 - P	0	-3	-2	2	-1	2	0	7	1	-5	-2	-3	-1	-4	-2	-2	1	0	-4	0
5	E	9 - A	2	-3	-5	-5	-3	-4	-6	-6	-2	-3	2	-1	-5	-3	1	-2	1	-2	2	5
6	E	10 - E	-3	-5	-3	-5	-5	-4	-6	2	-4	-6	-3	-2	-5	-2	2	-3	0	-3	4	6
7	V	11 - V	-3	-3	2	3	5	0	2	-4	-5	-4	-4	-2	0	-3	-5	-5	-5	-5	-6	-5
8	K	12 - A	2	-3	-4	-1	-2	-1	-3	-5	-4	-4	1	-3	-4	0	2	-1	3	2	-3	1
9	K	13 - K	0	-1	-5	-4	-2	-4	-1	-2	-2	-5	0	-2	-4	1	3	-3	4	1	-2	2
10	H	14 - H	-4	-4	-5	-1	-5	-3	0	-1	-5	-5	-2	-2	3	0	0	9	-1	-1	-4	-2
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
11	N	15 - N	0	-2	-3	-3	-4	-4	-5	-6	-4	1	1	-1	-5	5	0	-1	-1	-2	4	0
12	K	16 - K	-1	0	-4	-2	-2	-2	-4	-4	1	-1	2	1	-4	2	1	-1	3	0	1	0
13	D	17 - P	0	0	-4	-4	-4	-3	-3	0	2	-3	0	-2	-4	0	0	-1	2	0	3	2
14	G	18 - D	0	2	-5	-3	-4	-3	-5	-5	-1	-5	-1	-1	-4	2	1	0	3	-1	3	2
15	D	19 - D	-2	-2	-5	-4	-5	-4	-5	-6	2	-5	2	-2	-4	-1	-2	-1	1	3	5	1
16	C	20 - C	1	0	2	1	-1	0	-1	-5	1	6	0	-2	-1	1	-2	-4	-1	-1	-2	-4
17	W	21 - W	-5	-6	-3	2	-1	-3	1	11	-6	0	-5	-4	4	-6	-5	-4	-5	-5	-7	-5
18	I	22 - V	-1	-6	4	1	3	4	3	4	-5	-1	-4	-1	-3	-4	-5	-5	-5	-5	-6	-5
19	V	23 - V	3	-3	4	-1	4	-2	-4	-5	-4	0	0	0	-4	-5	-4	-2	-4	-1	-5	-4
20	I	24 - I	-4	-6	6	1	3	-2	1	-4	-5	-2	-5	-3	3	-6	-5	-1	-5	-5	-6	-6
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
21	N	25 - N	-2	-1	-5	-3	-5	-4	-3	-5	-4	-5	0	-3	-2	4	0	5	2	2	2	1
22	G	26 - G	-3	6	-6	-5	-6	-5	-3	-5	-5	-5	-2	-3	-5	0	-2	-1	-3	0	1	-3
23	K	27 - Y	-3	-2	-1	-3	0	-4	-2	-5	-4	-5	-1	0	1	1	1	3	4	2	-2	1
24	V	28 - V	-2	-6	3	-1	7	-2	-3	-5	-5	-3	-4	-3	-4	-6	-5	-6	-5	-5	-6	-5
25	Y	29 - Y	-4	-6	-3	0	-3	-3	3	0	-6	-1	-4	-4	9	-5	-3	3	-5	-4	-6	-5
26	D	30 - D	-4	-4	-6	-6	-6	-6	-6	-7	-4	-6	-3	-3	-6	3	-3	-3	-3	-2	8	-1
27	V	31 - L	-1	-1	3	2	5	1	0	-5	-5	3	-4	-3	-4	-5	-5	-6	-5	-5	-6	-5
28	T	32 - T	-1	-2	-4	-4	-3	-4	-5	-5	-4	-4	4	6	-4	0	-2	-4	-2	-2	0	-2
29	R	33 - R	-1	0	-5	-3	-5	-2	-5	-5	2	-5	1	-2	-4	0	0	-1	2	3	2	1
30	F	34 - F	-4	-2	-3	0	-4	-1	7	6	-6	-5	-3	-4	4	-5	-4	-3	-2	-1	-6	-5
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E

PSSM *(position specific score matrix)*

(matrice de poids de position)

PHI-BLAST (Pattern Hit Initiated)

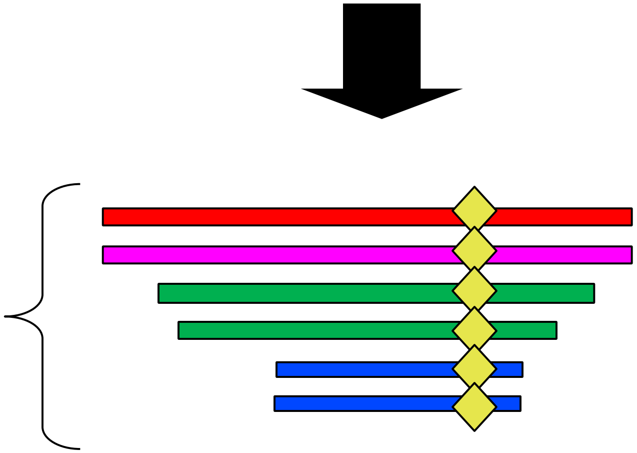


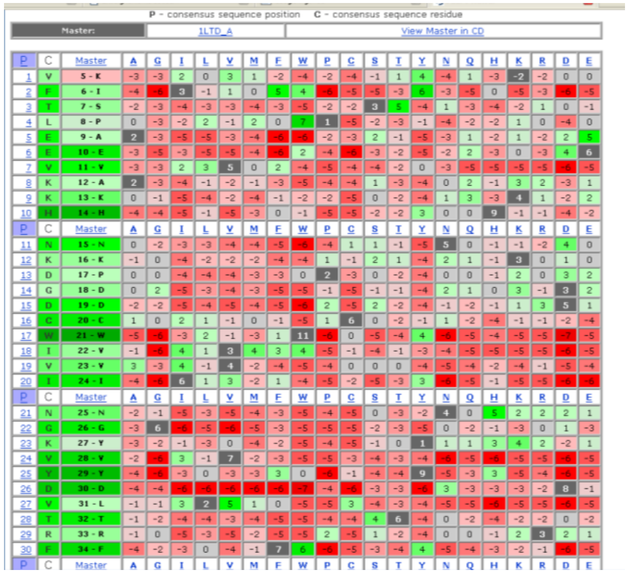
Query

C - C - {P} - {P} - x - C - [STDNEKPI] - x(3) - [LIVMFS] - x(3) - C

Pattern Prosite

Banque





P - consensus sequence position   C - consensus sequence residue

Master:	1LTD_A	View Master in CD

P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
1	V	5 - K	-3	-3	2	0	3	1	-2	-4	-2	-4	-1	1	4	-4	1	-3	-2	-2	0	0
2	F	6 - I	-4	-6	3	-1	1	0	5	4	-6	-5	-5	-3	6	-3	-5	0	-5	-3	-6	-5
3	T	7 - S	-2	-3	-4	-3	-3	-4	-3	-5	-2	-2	3	5	-4	1	-3	-4	-2	1	0	-1
4	L	8 - P	0	-3	-2	2	-1	2	0	7	1	-5	-2	-3	-1	-4	-2	-2	1	0	-4	0
5	E	9 - A	2	-3	-5	-5	-3	-4	-6	-6	-2	-3	2	-1	-5	-3	1	-2	1	-2	2	5
6	E	10 - E	-3	-5	-3	-5	-5	-4	-6	2	-4	-6	-3	-2	-5	-2	2	-3	0	-3	4	6
7	V	11 - V	-3	-3	2	3	5	0	2	-4	-5	-4	-4	-2	0	-3	-5	-5	-5	-5	-6	-5
8	K	12 - A	2	-3	-4	-1	-2	-1	-3	-5	-4	-4	1	-3	-4	0	2	-1	3	2	-3	1
9	K	13 - K	0	-1	-5	-4	-2	-4	-1	-2	-2	-5	0	-2	-4	1	3	-3	4	1	-2	2
10	H	14 - H	-4	-4	-5	-1	-5	-3	0	-1	-5	-5	-2	-2	3	0	0	9	-1	-1	-4	-2
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
11	N	15 - N	0	-2	-3	-3	-4	-4	-5	-6	-4	1	1	-1	-5	5	0	-1	-1	-2	4	0
12	K	16 - K	-1	0	-4	-2	-2	-2	-4	-4	1	-1	2	1	-4	2	1	-1	3	0	1	0
13	D	17 - P	0	0	-4	-4	-4	-3	-3	0	2	-3	0	-2	-4	0	0	-1	2	0	3	2
14	G	18 - D	0	2	-5	-3	-4	-3	-5	-5	-1	-5	-1	-1	-4	2	1	0	3	-1	3	2
15	D	19 - D	-2	-2	-5	-4	-5	-4	-5	-6	2	-5	2	-2	-4	-1	-2	-1	1	3	5	1
16	C	20 - C	1	0	2	1	-1	0	-1	-5	1	6	0	-2	-1	1	-2	-4	-1	-1	-2	-4
17	W	21 - W	-5	-6	-3	2	-1	-3	1	11	-6	0	-5	-4	4	-6	-5	-4	-5	-5	-7	-5
18	I	22 - V	-1	-6	4	1	3	4	3	4	-5	-1	-4	-1	-3	-4	-5	-5	-5	-5	-6	-5
19	V	23 - V	3	-3	4	-1	4	-2	-4	-5	-4	0	0	0	-4	-5	-4	-2	-4	-1	-5	-4
20	I	24 - I	-4	-6	6	1	3	-2	1	-4	-5	-2	-5	-3	3	-6	-5	-1	-5	-5	-6	-6
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
21	N	25 - N	-2	-1	-5	-3	-5	-4	-3	-5	-4	-5	0	-3	-2	4	0	5	2	2	2	1
22	G	26 - G	-3	6	-6	-5	-6	-5	-3	-5	-5	-5	-2	-3	-5	0	-2	-1	-3	0	1	-3
23	K	27 - Y	-3	-2	-1	-3	0	-4	-2	-5	-4	-5	-1	0	1	1	1	3	4	2	-2	1
24	V	28 - V	-2	-6	3	-1	7	-2	-3	-5	-5	-3	-4	-3	-4	-6	-5	-6	-5	-5	-6	-5
25	Y	29 - Y	-4	-6	-3	0	-3	-3	3	0	-6	-1	-4	-4	9	-5	-3	3	-5	-4	-6	-5
26	D	30 - D	-4	-4	-6	-6	-6	-6	-6	-7	-4	-6	-3	-3	-6	3	-3	-3	-3	-2	8	-1
27	V	31 - L	-1	-1	3	2	5	1	0	-5	-5	3	-4	-3	-4	-5	-5	-6	-5	-5	-6	-5
28	T	32 - T	-1	-2	-4	-4	-3	-4	-5	-5	-4	-4	4	6	-4	0	-2	-4	-2	-2	0	-2
29	R	33 - R	-1	0	-5	-3	-5	-2	-5	-5	2	-5	1	-2	-4	0	0	-1	2	3	2	1
30	F	34 - F	-4	-2	-3	0	-4	-1	7	6	-6	-5	-3	-4	4	-5	-4	-3	-2	-1	-6	-5
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E

PSSM *(position specific score matrix)*

(matrice de poids de position)