

## Fiche technique : étudier les protéines avec UniProt



<http://www.uniprot.org/>

**Objectifs :** ce site en ligne « UniProt », pour Universal Protein Resource, permet d'étudier les protéines dans un cadre évolutionniste. Il vous permet en particulier de :

1. Rechercher une protéine donnée, dans la banque, pour connaître sa séquence
2. Etudier en détail l'organisation moléculaire d'une protéine (niveau top ☺)
3. Comparer une même protéine mais provenant d'espèces animales ou végétales différentes dans le but d'établir le degré de parenté entre espèce
4. Rechercher dans la banque, le nom de protéine à partir d'une séquence d'acides aminés connue.

### Objectif 1 : rechercher une protéine connue



Dans cet exemple, on recherche la protéine **Sex-determining Region Y protein** (= SRY), impliquée dans la différenciation de l'embryon en mâle... sans elle que des femmes !!!

Dans la fenêtre de recherche, en haut, taper le nom de la protéine en anglais ; puis « Search ».

Dans la fenêtre de résultats, de très nombreuses réponses apparaissent mais qui ne correspondent pas forcément à la requête ! Il faut alors préciser votre recherche via « Advanced »



UniProtKB SRY

Advanced Search

BLAST Align Download Add to basket Columns

1 to 25 of 6,502 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
P07664	SRYD_DROME	Serendipity locus protein delta	Sry-delta, Sry-d, CG17958	Drosophila melanogaster (Fruit fly)	433
P07665	SRYB_DROME	Serendipity locus protein beta	Sry-beta, Sry-b, CG7938	Drosophila melanogaster (Fruit fly)	356
P07666	SRYA_DROME	Serendipity locus protein H-1 alpha	Sry-alpha, Sry-a, CG17957	Drosophila melanogaster (Fruit fly)	530
P15619	SRYC_DROME	Serendipity locus protein H-1	wdn, pqp, sry h-1, Sry-c, CG1454	Drosophila melanogaster (Fruit fly)	869
Q15599	NHRF2_HUMAN	Na(+)/H(+) exchange regulatory cofactor	SLC9A3R2, NHERF2	Homo sapiens (Human)	337
O42601	S17B1_XENLA	Transcription factor Sox-17-beta.1	sox17b.1, sox17b	Xenopus laevis (African clawed frog)	373
Q9UN79	SOX13_HUMAN	Transcription factor SOX-13	SOX13	Homo sapiens (Human)	622

En renseignant les options de recherche, vous pouvez cibler la recherche. Ici, on ne veut que la protéine SRY d'origine humaine, donc pour l'espèce Homo sapiens : Robert en nom scientifique !

Maintenant la liste est plus courte et on voit en 3<sup>ème</sup> ligne notre protéine SRY. Cliquer pour avoir les infos de la protéine. Noter tout de suite le nombre d'acides aminés.

Entry	Entry name	Protein names	Organism	Length
<input type="checkbox"/> Q15599	NHRF2_HUMAN	Na(+)/H(+) exchange regulator	Homo sapiens (Human)	337
<input type="checkbox"/> Q9UN79	SOX13_HUMAN	Transcription factor SOX-13	Homo sapiens (Human)	622
<input type="checkbox"/> Q05066	SRY_HUMAN	Sex-determining region Y protein	Homo sapiens (Human)	204
<input type="checkbox"/> Q7Z5C8	Q7Z5C8_HUMAN	SRY	Homo sapiens (Human)	84
<input type="checkbox"/> Q7Z5C4	Q7Z5C4_HUMAN	SRY	Homo sapiens (Human)	83

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Cross-references
- Publications
- Entry information
- Miscellaneous
- Similar proteins

Les différents menus, vous permettent d'obtenir toutes les informations concernant cette protéine. En particulier :

- La fonction de la protéine ;
- La séquence de la protéine ;

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing

## Function<sup>i</sup>

Transcriptional regulator that controls a genetic switch in male development. It is necessary and sufficient for init development of supporting cell precursors (pre-Sertoli cells) as Sertoli rather than granulosa cells (By similarity). motor functions of dopaminergic neurons (By similarity). Involved in different aspects of gene regulation including Promotes DNA bending. SRY HMG box recognizes DNA by partial intercalation in the minor groove. Also involved i sequence 5'-[AT]AACAA[AT]-3'. [By similarity](#) [4 Publications](#)

## Display

None

## Sequence<sup>i</sup>

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence

Sequence status<sup>i</sup>: Complete.

Q05066-1 [UniParc] [FASTA](#) [Add to basket](#)

« Hide

```

      10      20      30      40      50
MQSYASAMLS VFNSDDYSPA VQENIPALRR SSSFLCTESC NSKYQCETGE
      60      70      80      90     100
NSKGNVQDRV KRPMNAFIVW SRDQRRKMAL ENPRMRNSEI
      110     120     130     14
LTEAEKWPFQ QEAQKLQAMH REKYPNYKYR PRRKAKMLF
      160     170     180     19
SVLCSEVQLD NRLYRDDCTK ATHSRMEHQL GHLPPINAA

```

Chaque lettre représente un acide aminé. On connaît ainsi la séquence en acides aminés de la protéine.

L'option « FASTA » vous permet de l'éditer sans la numérotation.

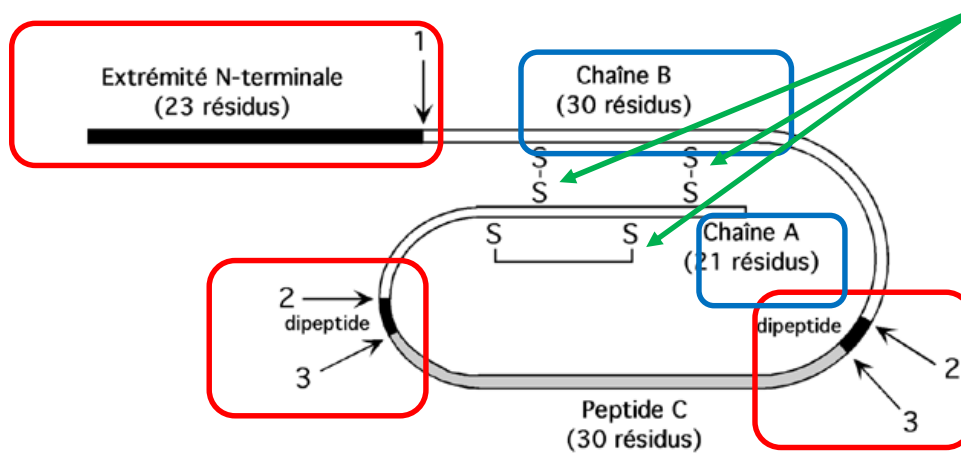
```

>sp|Q05066|SRY_HUMAN Sex-determining region Y protein OS=Homo sapiens GN=SRY PE=1 SV=1
MQSYASAMLSVFNSDDYSPAVQENIPALRRSSSFLCTESCNSKYQCETGENSKGNVQDRV
KRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFQEAQKLQAMH
REKYPNYKYRPRRKAKMLPKNCSSLPADPASVLCSEVQLDNRLYRDDCTKATHSRMEHQL
GHLPPINAASSPQQRDRYSHWTKL

```

## Objectif 2 : étudier en détail l'organisation moléculaire d'une protéine

Pour illustrer cet exemple, nous nous intéresserons à la protéine **insuline**. Il s'agit d'une hormone, produite par le pancréas et qui régule la quantité de sucre dans le sang. En cas de problème d'insuline, la personne souffre alors de **diabète**.



**Schéma de l'organisation de l'insuline – source « biochimiedesprotéine »**

La protéine insuline est fabriquée sous la forme d'un précurseur, la proinsuline, qui subit après des modifications post-transcriptionnelle. Ainsi :

- **Les 3 parties en rouge seront tout simplement éliminées !**
- **Ne seront conservées que les 2 parties bleues (chaîne A & B)**
- **3 ponts disulfures seront créés pour assurer la cohésion de la molécule**

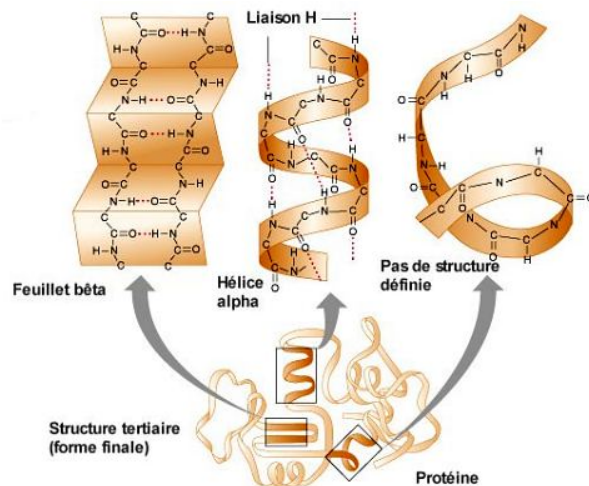
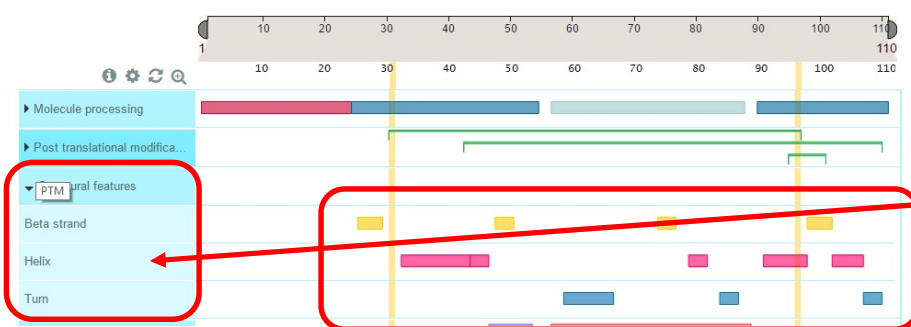
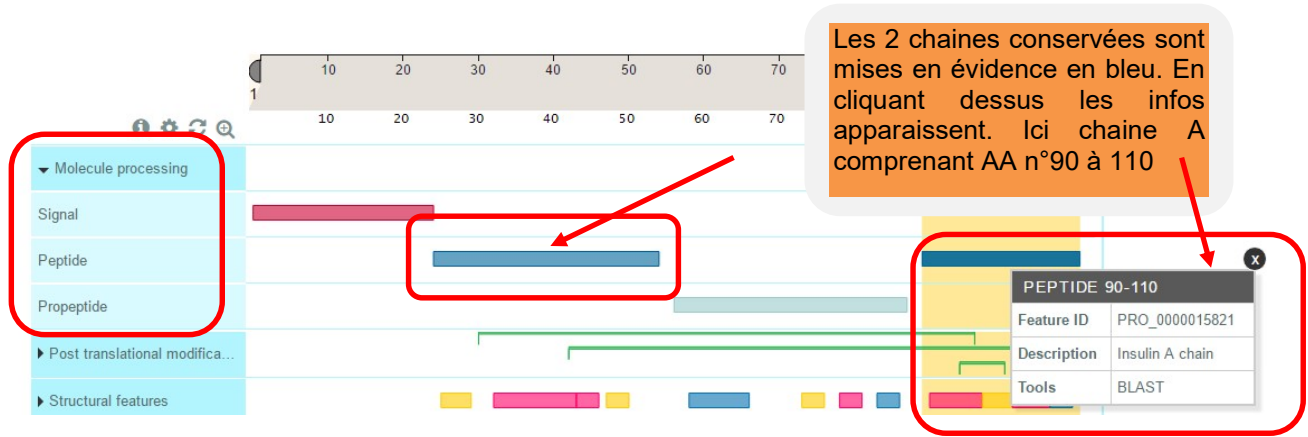
La capture d'écran montre la page UniProtKB pour l'insuline (P01308). Les annotations incluent :

- Le bouton « Feature viewer » est entouré d'un rectangle rouge.
- Le menu « Feature viewer » est également entouré d'un rectangle rouge.
- Le graphique de l'échelle des acides aminés numérotés (1 à 110) est visible.
- Le graphique de l'échelle des acides aminés numérotés est entouré d'un rectangle rouge.

Rechercher la protéine « insuline » comme précédemment. Dans le menu à gauche, cliquer alors sur l'option « Feature viewer ».

Echelle des acides aminés numérotés





**Pour information – les motifs élémentaires des protéines**



Cette option fait apparaître les mutations, en particulier celles à l'origine des maladies génétiques ! En cliquant dessus les informations apparaissent. Ici la mutation substitue en position n°47 à la place d'une Valine, une Glycine.

Human diseases ▾ Diabetes mellitus, permanent neonatal (PNDM)

mapping

### ases results

Download 1 to 1 of 1 Show 25 ▾

Repeat search in UniProtKB (3) ✕

**Disease**

- Diabetes mellitus, permanent neonatal
 

A rare form of diabetes distinct from childhood-onset autoimmune diabetes mellitus type 1. It is characterized by insulin-requiring hyperglycemia that is diagnosed within the first months of life. Permanent neonatal diabetes requires lifelong therapy.

UniProtKB (3)

En cliquant sur le lien, une page décrit la maladie







Enfin, vous pouvez traduire les ressemblances en arbre phylogénétique :

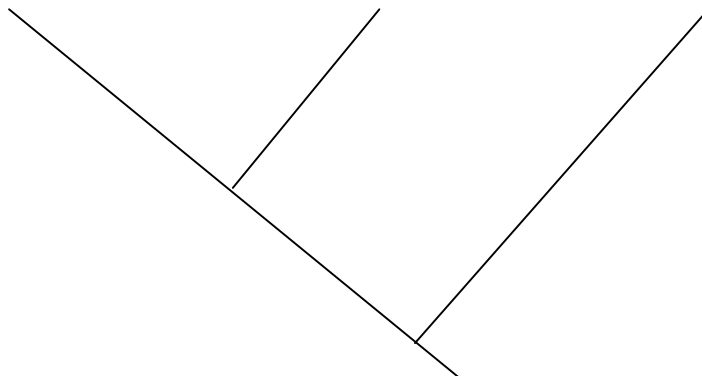
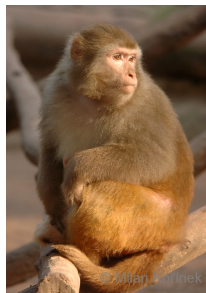
Plus le % d'identité est fort et plus la parenté entre 2 espèces est forte.

Tree



Highlight Taxonomy

Ce qui donne en image...



**Arbre traduisant la parenté de 3 vertébrés - obtenu par comparaison de l'hormone de croissance (Gh)**

Plus simple mais beaucoup moins de liberté...

<http://education.expasy.org/cgi-bin/philophylo/philophylo.cgi>

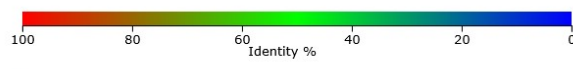
Autre possibilité, charger une des sélections de protéines déjà faite sur « Expasy » :

[http://education.expasy.org/cours/FLO/Liste\\_prot\\_evol.html](http://education.expasy.org/cours/FLO/Liste_prot_evol.html)

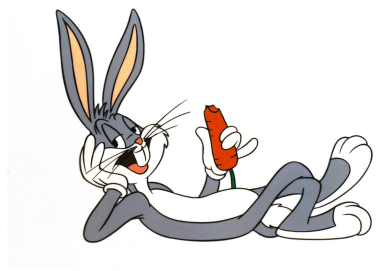
**Objectif 4 : Rechercher dans la banque, le nom de protéines à partir d'une séquence d'acides aminés connue.**

<http://www.uniprot.org/blast/>

Vous avez trouvé une protéine dont la séquence d'acides aminés est la suivante « CYIQNCPLG » : quelle est cette protéine ? A qui appartient-elle ?



Entry	Protein names	Match hit	Identity
P69057	Oxytocin (Tachyglossus aculeatus acule..)	██████████	100.0%
P69056	Oxytocin (Balaenoptera physalus)	██████████	100.0%
	Oxytocin (Hippopotamus amphibius)	██████████	100.0%
	Oxytocin (Oryctolagus cuniculus)	██████████	100.0%



Il s'agit donc de la protéine « ocytocine » composée de 9 aa. Elle s'exprime notamment chez le lapin. Cette protéine, intervient entre autre pendant la tété en permettant l'excrétion du lait.