### Fiche technique : étudier les protéines avec UniProt

## UniProt

## http://www.uniprot.org/

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

- 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



۲	BLAST alig	n 🛃 Download 🕋 Ad	ld to b	asket 🖉 Columns >		<b>1</b> to <b>25</b> of <b>6,502</b>	► 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	Å	Serendi Serendipity locus protei	n H-1 <b>alpha</b> , 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	Å	Na(+)/H(+) exchange regulatory cofa	SLC9A3R2, NHERF2	Homo sapiens (Human)	337
	042601	S17B1_XENLA		Transcription factor Sox- 17-beta.1	<b>sox17b.1</b> , sox17b	Xenopus laevis (African clawed frog)	373
	Q9UN79	SOX13_HUMAN	<b>A</b> <sup>►</sup>	Transcription factor SOX-	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 !								

Searching in UniProtKB		Help     R	lobe	<u>nome capiene</u> .	×			
A		(	All UniProtein AC Entry name [ID] Protein name [DE Gene name [GN] Organism [OS] Taxonomy [OC] Virus host Protein Existence		Term Siy Ferm Homo sapiens [9606]			٩
-	Linu y V	Linery	Function					cing un v
	P07664	SRYD_	Pathology & Biote	ch	rendipity locus protein Ita	Sry-delta, Sry-d, CG17958	Drosophila melanogaster (Fruit fly)	433
	P07665	SRYB_	Expression Interaction		erendipity locus protein eta	Sry-beta, Sry-b, CG7938	Drosophila melanogaster (Fruit fly)	356
	P07666	SRYA_	Sequence Family and Doma	ins	erendipity locus protein pha	Sry-alpha, Sry-a, CG17957	Drosophila melanogaster (Fruit fly)	530
	P15619	SRYC_	Cross-references Web resource Date Of		rendipity locus protein	<b>wdn</b> , pqp, sry h-1, Sry-c, CG1454	Drosophila melanogaster (Fruit fly)	869
	Q15599	NHRF2	Gene Ontology [G Keyword [KW] Literature Citation	60]	a(+)/H(+) exchange gulatory cofa	SLC9A3R2, NHERF2	Homo sapiens (Human)	337
	042601	S17B1	Proteomes Cited For		anscription factor Sox- /-beta.1	<b>sox17b.1</b> , sox17b	Xenopus laevis (African clawed frog)	373
	Q9UN79	S0X13	Active UniRef ID		anscription factor SOX-	SOX13	Homo sapiens (Human)	622
	Q05066	SRY_H	UniParc ID		ex-determining region Y protein	SRY, TDF	Homo sapiens (Human)	204
	005738	SRV N			Sov-datarmining ragion V	Sev Tdf Tdy	Mue mueculue (Mouco)	205

S BLAST ≡ Align 🛨 Download 📾 Add to basket 🖉 Columns >						
🔲 Entry 🗘	Entry name 🖨	Protein names 🕈	, es			
Q15599	NHRF2_HUMAN	Na(+)/H(+) exchange regulator	iom			
Q9UN79	SOX13_HUMAN	Transcription factor SOX-13				
Q05066	SRY_HUMAN	Sex-determining region Y protein				
Q7Z5C8	Q7Z5C8_HUMAN	SRY				
Q7Z5C4	Q7Z5C4_HUMAN	SRY SRY				

ntenant la liste est plus courte et on voit en ligne notre protéine SRY. Cliquer pour avoir infos de la protéine. Noter tout de suite le nbre d'acides aminés.

Homo sapiens (Hun

Homo sapiens (Human)

Homo sapiens (Human)

Homo sapiens (Human)

an)

622

83

204

×

Les différents menus, vous permettent
d'obtenir toutes les informations concernant
antha martífica. En martínullar a

cette protéine. En particulier :

SOX13

SRY, TDF

SRY

SRY

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

Ton

✓ Function

**Expression** 

✓ Interaction

Structure

Sequence

Publications Entry information Miscellaneous Similar proteins

Family & Domains

Cross-references

✓ Names & Taxonomy Subcellular location ✓ Pathology & Biotech ✓ PTM / Processing

# Function

✓ Names & Taxonomy

✓ Function

- Subcellular location
- ✓ Pathology & Biotech
- PTM / Processing

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 N	one Sequence						_	
✓ Function	Sequence status	: Complete.						
✓ Names & Taxonomy	Q05066-1 [UniPa		🛱 Add to ba	sket				
Subcellular location	« Hide							
Pathology & Biotech								
PTM / Processing	10	20	30	40	50			
Expression	MQSYASAMLS 60	VFNSDDYSPA 70	VQENIPALRR	SSSFLCTESC 90	NSKYQCETGE 100			
✓ Interaction	NSKGNVQDRV	KRPMNAFIVW	SRDQRRKMAL	ENFRMRNSEI				
Structure	110 LTEAEKWPFF	120 QEAQKLQAMH	130 REKYPNYKYR	14 PRRKAKMLE	Chaque aminé.	lettre On	représente un connaît ain	acide si la
Family & Domains	160	170	180	19	séquenc	e en	acides aminés	de la
Sequence	SVLCSEVQLD	NRLYRDDCTK	ATHSRMEHQL	GHLPPINAA	protéine.			
					L'option l'éditer sa	« FA: ans la	STA » vous per numérotation.	met de

>sp|Q05066|SRY\_HUMAN Sex-determining region Y protein OS=Homo sapiens GN=SRY PE=1 SV=1 MQSYASAMLSVFNSDDYSPAVQENIPALRRSSSFLCTESCNSKYQCETGENSKGNVQDRV KRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMH REKYPNYKYRPRRKAKMLPKNCSLLPADPASVLCSEVQLDNRLYRDDCTKATHSRMEHQL 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 (chaine A & B)
- **3 ponts disulfures seront créés pour assurer la cohésion de la molécule**





					Cette option fait
▶ Proteomics					apparaitre les mutations,
✓ Variants		A day and			en particulier celles à l'origine des maladies
Eilter concernance	G	•		G	génétiques ! En cliquant
	A V			A	dessus les informations
Predicted deleterious	i	VARIANT 47-47			annaraissent Ici Ia
Predicted benign	s •	Source UniProt		s	mutation substitue on
Non-disease	T	Feature ID VAR_0637	30		mutation substitute en
Init, stop loss or gain	C  M	Variant G > V		M	position n°47 a la place
Filter data source	D • • •	Cross-references rs8035666	7 dbSNP   Ensembl	D	d'une Valine, une
UniProt reviewed	E	Disease Association		E	Glycine.
Large scale studies	G B	Disease Diabetes n	nellitus, permanent neonatal (PN	IDM)	
	к	OMIM 606	176	— н	
	H O		-		
	Υ		/	Y	
Human diseases - Dia mapping	ibetes mellitus, permanent neonatal (PNDM)		En cliquant une page maladie	sur le lie décrit	n, la
ases results					es
± Download				1 to 1 of 1 S	how 25 •
Repeat search i	n UniProtKB (3)				×
Disease		•			
Diabetes mellitus, perr A rare form of diabetes first months of life. Per	nanent neonatal s distinct from childhood-onset autoimmune diab manent neonatal diabetes requires lifelong thera	etes mellitus type 1. It is characterized py.	by insulin-requiring hypergly	cemia that is diagnosed	within the
UniProtKB (3)					

## Objectif 3 : comparer une même protéine ayant pour origine différentes espèces

On veut comparer différentes hormones de croissance (nommée « growth hormone ») qui proviennent d'espèces différentes :

- □ Homme (Homo sapiens)
- Saumon atlantique (Oncorhynchus keta)
- Singe macaque (Macaca mulatta)

Rechercher, suivant la méthodologie de l'objectif 1, ces protéines afin de les sélectionner.



	✓ Alignment	Alignment							
	✓ Tree	How to print an alignment in color							
	Result info	DOI 241 SOMA HUMAN	1	MATCEDTET I A SET OT DUT OF CEARDT DI EDI FONJAMI DAUDT NOT A FOTVOFFERA	60				
(	Highlight	OSII60 OSII60 ONCKE P33093 SOMA_MACMU	1 1	MATGSTSJLLEVSCOLEDFSCOLEGARENORLENTRAINLINGTHLIGTELE MCQVFLLMFVLLVSCFLSQGAAMENORLENTVNRVCHLHLLQQKHNDFEGT MAAGSRTSLLLAFALLCLPWLOEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEA * :** ** :: :*:::**:	53 60				
	Annotation Helix Alternative sequence	P01241 SOMA HUMAN Q91160 Q91160 ONCKE P33093 SOMA_MÄCMU	61 54 61	YIPKEQKYSFLQNPGTSLGFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLR LLSDERR-QLNKIFLDFCNSDSTVSPIDKQETQKSSVLKLHISFRLIESWEYPSQTL YIPKEQKYSFLQNPGTSLFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLR 	120 112 120				
	<ul> <li>Signal peptide</li> <li>Beta strand</li> <li>Disulfide bond</li> </ul>	P01241 SOMA HUMAN Q91160 Q91160 ONCKE P33093 SOMA_MACMU	121 113 121	SVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSH ISNSLMVR-NSNQISEKLSDLKVGINLLEGSQEGVLSLDDNDSOHLPPYGNYYQNL SVFANSLVYGISYSDVYDLLKDLEEGIQTLMGRLEDGSSRTGQIFKQTYSKFDTNSH ***:: **.	177 169 177				
	Sequence conflict  Natural variant  Metal binding  Chain	P01241 SOMA HUMAN Q91160 Q91160_ONCKE P33093 SOMA_MACMU	Da co	ns la fenêtre à gauche, vous pouvez cher différentes options, afin de mettre	217 210 217				
	☐ Modified residue ☐ Turn	You may add additional	en aci	évidence les propriétés de certains ides aminés, comme par exemple :					
l	Amino acid properties Similarity Hydrophobic			Acides aminés similaires, c'est à dire qui ne 'changent' pas la protéine					
				Acides aminés impliqués dans les ponts disulfures, jouant un rôle majeur dans la structure et le repliement de la protéine.					





Highlight Taxonomy

### Ce qui donne en image...



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

# 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 ?

