HMSN304 Structures des molécules

Projet

Projection d'alignements de séquences protéiques sur des structures 3D connues : développement de procédures en langage Jmol *scripting*.

protéine hypothétique **xy007** de *Mycobaterium …*

- fonction et structure inconnues
- séquence (acides aminés) obtenue par
  traduction conceptuelle d'une séquence génomique

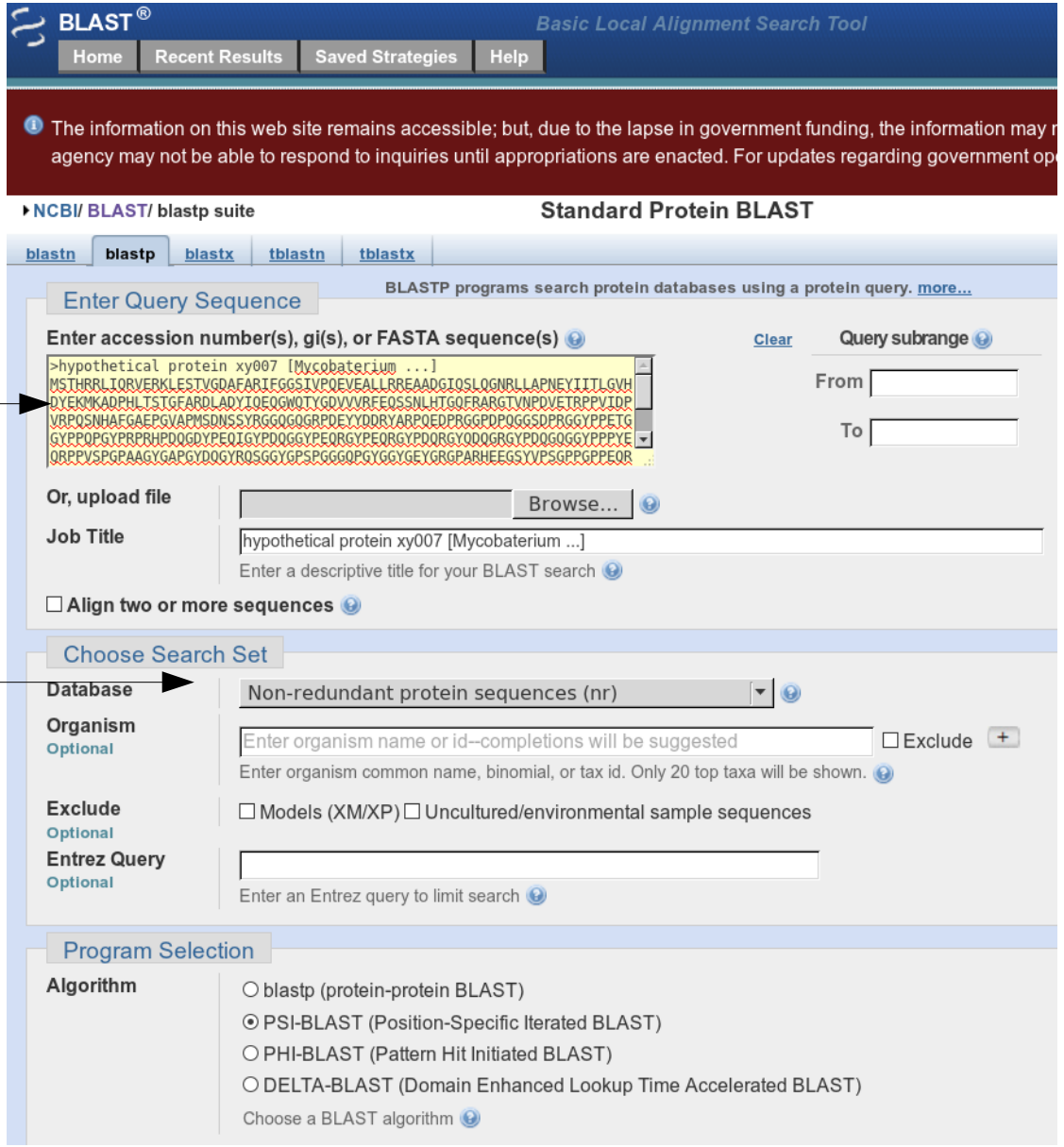
```
1
|
MSTHRRLIQRVERKLESTVGDAFARIFGGSIVPQEVEALLRREAADGIQSLQGNRLLAPN
EYIITLGVHDYEKMKADPHLTSTGFARDLADYIQEQGWQTYGDVVVRFEQSSNLHTGQFR
ARGTVNPDVETRPPVIDPVRPQSNHAFGAEPGVAPMSDNSSYRGGQGQGRPDEYYDDRYA
RPQEDPRGGPDPQGGSDPRGGYPPETGGYPPQPGYPRPRHPDQGDYPEQIGYPDQGGYPE
QRGYPEQRGYPDQRGYQDQGRGYPDQGQGGYPPPYEQRPPVSPGPAAGYGAPGYDQGYRQ
SGGYGPSPGGGQPGYGGYGEYGRGPARHEEGSYVPSGPPGPPEQRPAYPDQGGYDQGYQQ
GATTYGRQDYGGGADYTRYTESPRVPGYAPQGGGYAEPAGRDYDYGQSGAPDYGQPAPGG
YSGYGQGGYGSAGTSAYLRRDRAGSWWYLLASCENTAGRGFDSDIRLVDVTVSRNHAEIK
QNGGEYIIMDVNSTNGTYVQGARVNSVRIHDGDHIQLGKFEFEVRAG
                                                  |
                                                 527
```

Comparaison (alignement) de séquences de protéines

Recherche de séquences homologues par BLAST

séquence requête

base de données



Comparaison (alignement) de séquences de protéines

résultats BLAST (liste des intitulés des séquences et des leurs scores)

Sequences producing significant alignments with E-value BETTER than threshold

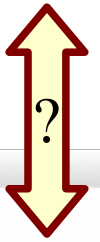
Accession	Description	Max score	Total score	Query coverage	E value	Ident	Links
NEW gi|490004265|WP_003907172.1	conserved hypothetical protein [Mycobacterium tuberculosis] >gi|289	427	427	48%	1e-142	94%	
NEW gi|15607162|NP_214534.1	Conserved protein with FHA domain, FhaA [Mycobacterium tuberculos	425	577	78%	1e-139	94%	G
NEW gi|494694404|YP_007959907.1	hypothetical protein J113_00140 [Mycobacterium tuberculosis CAS/NI	421	573	78%	3e-138	90%	G
NEW gi|15839394|NP_334431.1	hypothetical protein MT0023 [Mycobacterium tuberculosis CDC1551] >	417	569	78%	7e-137	94%	G
NEW gi|490002275|WP_003905210.1	signal peptide protein [Mycobacterium tuberculosis] >gi|289696666|gl	417	568	78%	1e-136	94%	
NEW gi|433640148|YP_007285907.1	Conserved hypothetical protein with Fha domain, TB39.8 [Mycobacteri	417	569	78%	1e-136	93%	G
NEW gi|340625053|YP_004743505.1	hypothetical protein MCAN_00191 [Mycobacterium canettii CIPT 1400	416	568	78%	3e-136	93%	G
NEW gi|121635930|YP_976153.1	hypothetical protein BCG_0050c [Mycobacterium bovis BCG str. Paste	416	565	78%	6e-136	93%	G
NEW gi|339630103|YP_004721745.1	hypothetical protein MAF_00200 [Mycobacterium africanum GM041182	415	567	78%	6e-136	93%	G

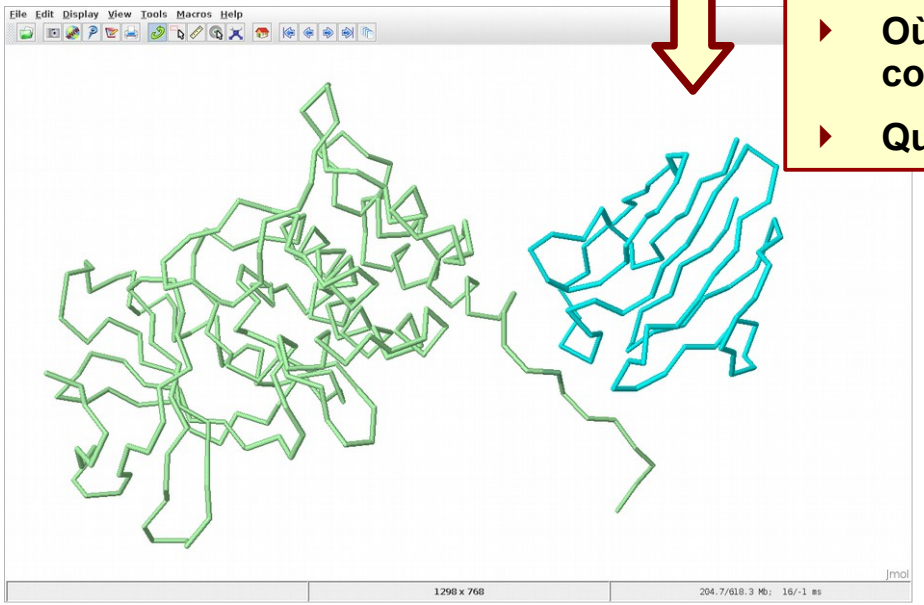
recouvrement de la séquence requête
(% de séquence qui a été aligné)
si très différent de 100% peut indiquer la présence de domaines

signification statistique
alignement significatif si *E*~0)

taux d'identité
entre les deux (segments de) séquences

```
>gi|374977508|pdb|3OUN|A Chain A, Crystal Structure Of The Fhaa Fha Domain Complexed Wit
The Intracellular Domain Of Rv3910
Length=157

 Score =   105 bits (262),  Expect = 3e-23, Method: Compositional matrix adjust.
 Identities = 84/138 (61%), Positives = 95/138 (69%), Gaps = 0/138 (0%)

Query  388  YAPQGGGYAEPAGRDYDYGQSGAPDygqpapggysgyggggygsagTSAYLRRDRAGSWW  447
            + PQGGGYAEPAGRDYDYGQSGAPDYGQPAPGGYSGYGQGGYGSAGTS  L+ D
Sbjct  20   HMPQGGGYAEPAGRDYDYGQSGAPDYGQPAPGGYSGYGQGGYGSAGTSVTLQLDDGSGRT  79

Query  448  YLLASCENTAGRGFDSDIRLVDVTVSRNHAEIKQNGGEYIIMDVNSTNGTYVQGARVNSV  507
            Y L    N  GRG D+  RL D  VSR H EI+ +G   ++ D+NSTNGT V  A V
Sbjct  80   YQLREGSNIIGRGQDAQFRLPDTGVSRRHLEIRWDGQVALLADLNSTNGTTVNNARVQEW  139

Query  508  RIHDGDHIQLGKFEFEVR  525
```





- **Où se trouve, dans la structure, le segment de séquence aligné ?**
- **Le résidu #*n* de la structure, à quel résidu correspond-il dans la séquence requête ?**
- **Où se situent, dans la structure, les résidus conservés ?**
- **Quel est le rôle des résidus conservés ?**

Projet

Projet (1/4)

Développer des procédures (fonctions) Jmol pour :

- définir une nouvelle propriété atomique

  `property_seqresno`

  correspondant aux numéros séquentiels (à partir de 1) des résidus tels qu'ils apparaissent dans les enregistrement SEQRES

  - ATTENTION :
    - à la numérotation des résidus qui ne commence pas forcement par 1
    - aux codes d'insertion dans la numérotation des résidus
    - aux résidus manquants
    - aux résidus non standard (modifiés)
  - AIDE : utiliser les informations obtenues par

    `getproperty('chainInfo')`

    ou **`getproperty('polymerInfo')`**

    Faire des tests sur **chacune** des chaînes de la structure **1a2c**

Projet (2/4)

- définir, pour chaque résidu de la structure, une nouvelle propriété atomique

  `property_qresno`

  correspondant au numéro du résidu de la séquence requête avec lequel le résidu a été aligné.

  - REMARQUE : définir
    - **`property_qresno = -1`** si le résidu ne fais pas partie du segment aligné
    - **`property_qresno = 0`** si le résidu est aligné avec un *indel* (*gap*)

- définir, pour chaque résidu de la structure, une nouvelle propriété atomique
  - **`property_qsimil = -1`** → résidu non aligné
  - **`property_qsimil = 0`** → résidu aligné dissimilaire
  - **`property_qsimil = 1`** → résidu aligné similaire
  - **`property_qsimil = 2`** → résidu aligné identique

Pour ce faire : des scripts contenants des alignements BLAST sont disponibles sur l'Espace Pédagogique

Comparaison (alignement) de séquences de protéines

résultats BLAST sauvegardés dans un fichier XML
(voir un exemple sur l'Espace Pédagogique)

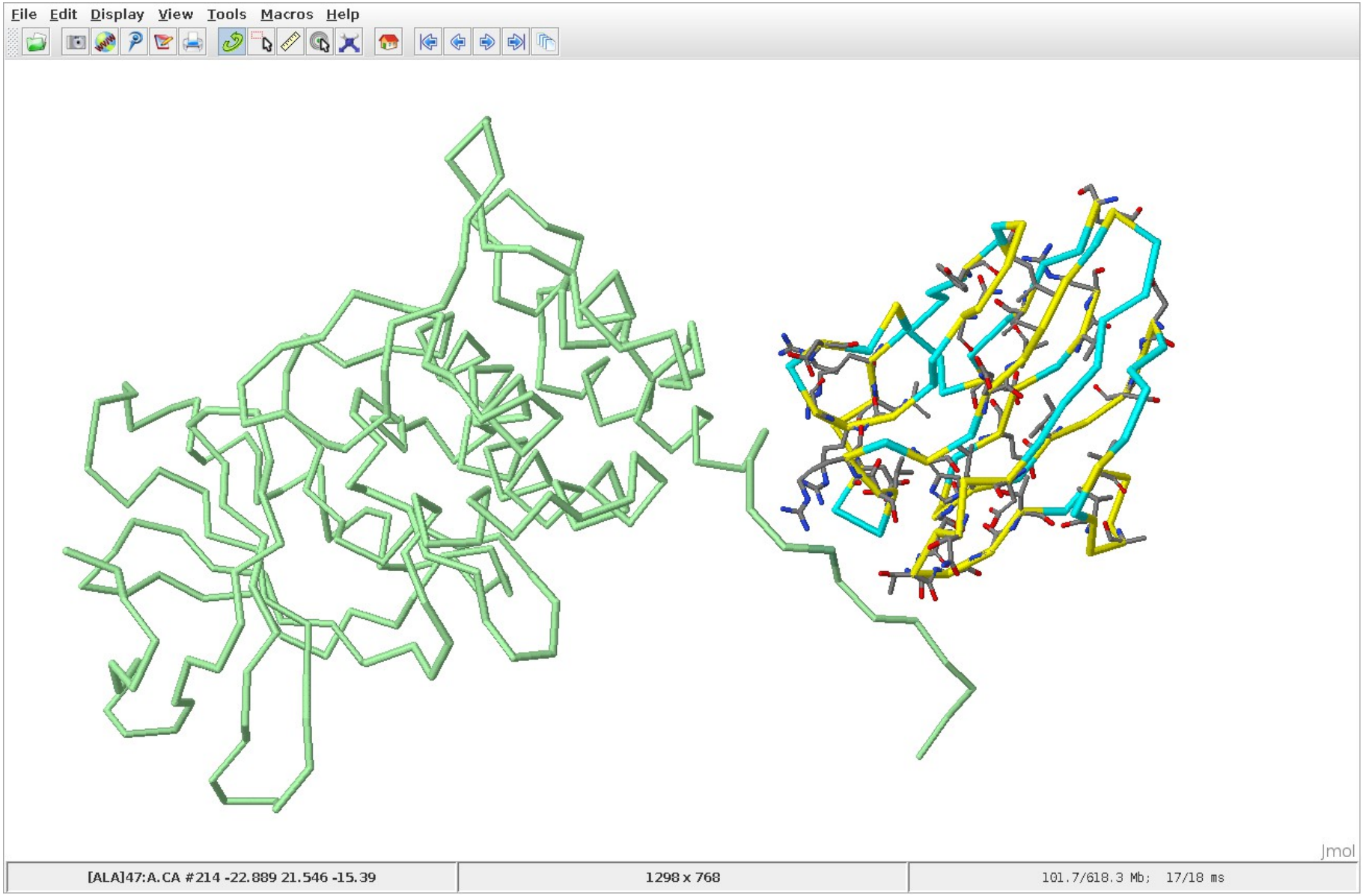

```
−<Hit>
  <Hit_num>10</Hit_num>
  <Hit_id>gi|374977508|pdb|3OUN|A</Hit_id>
 −<Hit_def>
    Chain A, Crystal Structure Of The Fhaa Fha Domain Complexed With The Intracellular Domain Of Rv3910
  </Hit_def>
  <Hit_accession>3OUN_A</Hit_accession>
  <Hit_len>157</Hit_len>
 −<Hit_hsps>
  −<Hsp>
     <Hsp_num>1</Hsp_num>
     <Hsp_bit-score>95.8033</Hsp_bit-score>
     <Hsp_score>237</Hsp_score>
     <Hsp_evalue>1.26809e-25</Hsp_evalue>
     <Hsp_query-from>434</Hsp_query-from>
     <Hsp_query-to>525</Hsp_query-to>
     <Hsp_hit-from>66</Hsp_hit-from>
     <Hsp_hit-to>157</Hsp_hit-to>
     <Hsp_query-frame>0</Hsp_query-frame>
     <Hsp_hit-frame>0</Hsp_hit-frame>
     <Hsp_identity>40</Hsp_identity>
     <Hsp_positive>55</Hsp_positive>
     <Hsp_gaps>0</Hsp_gaps>
     <Hsp_align-len>92</Hsp_align-len>
    −<Hsp_qseq>
       TSAYLRRDRAGSWWYLLASCENTAGRGFDSDIRLVDVTVSRNHAEIKQNGGEYIIMDVNSTNGTYVQGARVNSVRIHDGDHIQLGKFEFEVR
     </Hsp_qseq>
    −<Hsp_hseq>
       TSVTLQLDDGSGRTYQLREGSNIIGRGQDAQFRLPDTGVSRRHLEIRWDGQVALLADLNSTNGTTVNNAPVQEWQLADGDVIRLGHSEIIVR
     </Hsp_hseq>
    −<Hsp_midline>
       TS L D Y L +++N GRG D + RL+D VSR H EI++ G D+NSTNGT V++A+V + DGD+I LG+ E+ VR
     </Hsp_midline>
   </Hsp>
  </Hit_hsps>
 </Hit>
−<Hit>
  <Hit_num>11</Hit_num>
```

Comparaison (alignement) de séquences de protéines

résultats BLAST gardés dans des variables Jmol

`script` **`BLAST_alignement5_3OUN-A.jmol`** (voir l'Espace Pédagogique ; d'autres *scripts* sont aussi disponibles)

```
query =
"MSTHRRLIQRVERKLESTVGDAFARIFGGSIVPQEVEALLRREAADGIQSLQGNRLLAPNEYIITLGVHDYEKMKADPHLTSTGFARDLADYIQEQGWQTYGDVVVRFEQSSNLHT
GQFRARGTVNPDVETRPPVIDPVRPQSNHAFGAEPGVAPMSDNSSYRGGQGQGRPDEYYDDRYARPQEDPRGGPDPQGGSDPRGGYPPETGGYPPQPGYPRPRHPDQGDYPEQIGYP
DQGGYPEQRGYPEQRGYPDQRGYQDQGRGYPDQGQGGYPPPYEQRPPVSPGPAAGYGAPGYDQGYRQSGGYGPSPGGGQPGYGGYGEYGRGPARHEEGSYVPSGPPGPPEQRPAYPD
QGGYDQGYQQGATTYGRQDYGGGADYTRYTESPRVPGYAPQGGGYAEPAGRDYDYGQSGAPDYGQPAPGGYSGYGQGGYGSAGTSAYLRRDRAGSWWYLLASCENTAGRGFDSDIRL
VDVTVSRNHAEIKQNGGEYIIMDVNSTNGTYVQGARVNSVRIHDGDHIQLGKFEFEVRAG" ;

Hit_def    = "Chain A, Crystal Structure Of The Fhaa Fha Domain Complexed With The Intracellular Domain Of Rv3910" ;
Hit_PDB_ID     = "3OUN" ;
Hit_chain      = "A" ;
Hsp_score      = "237" ;
Hsp_Evalue     = "1.26809e-25" ;
Hsp_query_from = "434" ;
Hsp_query_to   = "525" ;
Hsp_hit_from   = "66" ;
Hsp_hit_to     = "157" ;
Hsp_identity   = "40" ;
Hsp_positive   = "55" ;
Hsp_gaps       = "0" ;
Hsp_align_len  = "92" ;
Hsp_qseq    = "TSAYLRRDRAGSWWYLLASCENTAGRGFDSDIRLVDVTVSRNHAEIKQNGGEYIIMDVNSTNGTYVQGARVNSVRIHDGDHIQLGKFEFEVR" ;
Hsp_midline = "TS  L  D      Y L +++N  GRG D + RL+D  VSR H EI++ G      D+NSTNGT V++A+V +    DGD+I LG+ E+ VR" ;
Hsp_hseq    = "TSVTLQLDDGSGRTYQLREGSNIIGRGQDAQFRLPDTGVSRRHLEIRWDGQVALLADLNSTNGTTVNNAPVQEWQLADGDVIRLGHSEIIVR" ;
```


File Edit Display View Tools Macros Help
Jmol
[ALA]47:A.CA #214 -22.889 21.546 -15.39
1298 x 768
101.7/618.3 Mb; 17/18 ms

Projet (3/4)

- définir une procédure pour superposer deux structures homologues à la séquence requête
  - REMARQUE :
    utiliser comme **points d'ancrage** pour la superposition les C^{α} des **résidus alignés en commun** entre les deux structures

Projet (4/4)

- Utiliser les procédures développées pour analyser les structures homologues de la protéines xy007