

## ESERCITAZIONE (versione comprendente i risultati)

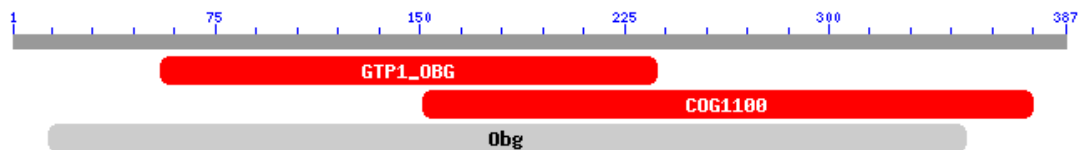
Costruzione di un modello di una proteina mediante homology modelling

Codice Encode della proteina da modellare (target): **AC006153**

- 1) Utilizzando il codice Encode fornito, effettuare innanzi tutto una ricerca nel database Ensemble ed analizzare l'output ottenuto.
- 2) Individuare le varianti di splicing e tra di esse quelle che vengono tradotte. (**R: vi sono 3 varianti di splicing**)
- 3) Costruire quindi il modello per ognuna delle varianti di splicing che vengono tradotte.

### A) Costruzione del modello della variante 1 (è la seq. proteica più lunga)

- 1) Utilizzando il database Pfam caratterizzare la sequenza da modellare per quanto riguarda la composizione in domini (**vengono predetti 2 domini in questa variante a differenza dell'altra in cui ne viene predetto solo 1**).
- 2) Effettuare una ricerca mediante PSI-Blast per individuare il/i templat(i) da utilizzare per la modellizzazione. (**R: al primo round si ha 1 templat(i): 1LNZ\_A**)



```
>gi|24158882|pdb|1LNZ|B S Chain B, Structure Of The Obg Gtp-Binding Protein  
gi|24158881|pdb|1LNZ|A S Chain A, Structure Of The Obg Gtp-Binding Protein
```

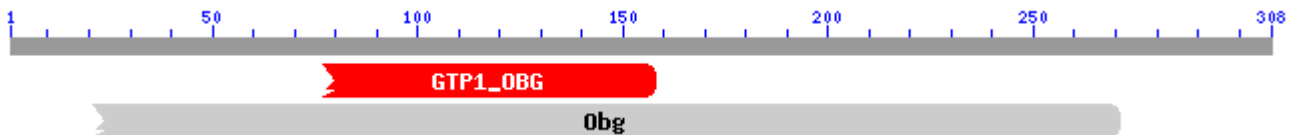
Score = 205 bits (521), Expect = 3e-51

Identities = 136/360 (37%), Positives = 189/360 (52%), Gaps = 59/360 (16%)

```
Query 15 FIDKLRLLFTRGGSGMGYPRL-----GGEGGKGGDVWVVAQNRMTLKQLKDRY 62  
F+D++++ +GG GG G GG+GGKGGDV V + L+ L D  
Sbjct 2 FVDQVKVYVKGDDGNGXVAFRREKYVPKGGPAGDGGKGGDV--VFEVDEGLRTLXDFR 59  
  
Query 63 PRKRFVAGVGANSKISALKSGKGDCEIPVPGISVTDENGK-IIGELNKENDRILVAQG 121  
+K F A G + G D I VP G VTD++ K +I +L + R ++A+G  
Sbjct 60 YKKHFKAIRGEHGXSKNQHGRNADDXVIKVPVPGTVVTDGDKQVIADLTEHGQRAVIARG 119  
  
Query 122 GLGGKLLTNFLPLK-----GQKRIIHLDLKLIADVGLVGFNPAGKSSLLSCVS 169  
G GG+ + F G++R I L+LK++ADVGLVGFNP+ GKS+LLS VS  
Sbjct 120 GRGGRGNSRFATPANPAPQLSENGEPGKERYIVLELKVADVGLVGFPSVGKSTLLSVVS 179  
  
Query 170 HAKPAIADYAFTTLKPELGKIMYSDFKQISVADLPGLIEGAHMNKGMGHFKKHIERTRQ 229  
AKP IADY FTTL P LG + D + ADLPGLIEGAH G+GH+FL+HIERTR  
Sbjct 180 SAKPKIADYHFTTLVPLNGXVETDDGRSFVXADLPGLIEGAHQVGLGHQFLRHIERTRV 239  
  
Query 230 LLFVVDISGFQLSSHTQYRTAFETIILLTKELELYKEELQTKPALLAVNKMDLPDAQDF 289  
++ V+D SG + R ++ + + +EL Y L +P ++ NK D P+A +  
Sbjct 240 IVHVIDXSGL-----EGRDPYDDYLTINQELSEYNLRLTERPQIIVANKXDXPEAAEN- 292  
  
Query 290 HELMSQLQNPKDFLHLFEKNMIPERTVEFQHIIPISAVTGEIEELKNCIRKSLDEQANQ 349  
L F+ E+ + + PISAVT EG+ EL L E ANQ  
Sbjct 293 -----LEAFK-----EKLTDYVPVPI SAVTREGLREL-----LFEVANQ 327
```

- 3) Effettuare un allineamento multiplo di tutte le sequenze proteiche della famiglia cui appartiene il template strutturale individuato in precedenza (incluso la sequenza target).
- 4) Scaricare le coordinate atomiche della proteina che useremo come template.
- 5) Analizzare in modo dettagliato la proteina template (catene A e B) e scegliere uno dei due monomeri da usare. **(utilizzare la catena A perché nella B mancano alcuni residui).**
- 6) Utilizzo del Programma Swisspdb Viewer (installazione, visualizzazione del file, comandi principali, editing, ecc).
- 7) Caricare la sequenza del target.
- 8) Editare l'allineamento manualmente.
- 9) Identificazione e modelling dei loop (tools di SPDBV).
- 10) Eventuale sottomissione a SwissModel (e/o server semi/automatici), o minimizzazione dell'energia.
- 11) Eventuale confronto.

## Costruzione del modello della variante 2



**Ripetere gli step effettuati per la costruzione del modello della variante 1, confrontando di volta in volta i risultati che si ottengono.**

```

Score = 137 bits (345), Expect = 6e-31
Identities = 98/260 (37%), Positives = 130/260 (50%), Gaps = 57/260 (21%)

Query 21  LFTRGGSSGGMGY-----PRLGGEGGKGGDVVVAQNRMTLKQLKDRYPRKRFVAG 70
+ RGG GG G P+L G G + ++V + LK L D V
Sbjct 115 VIARGRRGGGRNSRFATPANPAPQLSENGEPGKERYIVLE----LKVLD-----VGL 163

Query 71  VGANSKFPNAGKSSLLSCVSHAKPAIADYAFTTLKPELGKIMYSDFKQISVADLPGLIEG 130
VG FP+ GKS+LLS VS AKP IADY FTTL P LG + D + ADLPGLIEG
Sbjct 164 VG----FPSVGKSTLLSVSSAKPKIADYHFTTLVPLNGXVETDDGRSFVXADLPGLIEG 219

Query 131 AHMNKMGHGHKFLKHIERTRQLLFVVDISGFQLSSHTQYRTAFETIILLTKELELYKEELQ 190
AH G+GH+FL+HIERTR ++ V+D SG + R ++ + + +EL Y L
Sbjct 220 AHQGVGLGHQFLRHIERTRVIVHVIDXSGL-----EGRDPYDDYLTINQELSEYNLRLT 273

Query 191 TKPALLAVNKMDLPDAQDKFHELMSQLQNPKDFLHLFEKNMIPERTVEFQHIIPIISAVTG 250
+P ++ NK D P+A + P+L G G + ++V + LK L D V
Sbjct 274 ERPQIIIVANKXDXPEAAEN-----LEAFK----EKLTDYVPVFPISAVTR 314

Query 251 EGIEELKNCIRKSLDEQANQ 270
EG+ EL L E ANQ
Sbjct 315 EGLREL-----LFEVANQ 327

```

Alla fine della modellizzazione delle varianti effettuare un'analisi comparativa dei modelli generati.

## - Fare i modelli utilizzando HHpred

Sequenze:

>AC006153.3-001

MVHCSCVLFRKYGNFIDKLRLFTRGGSGMGYPRLGGEGGKGGDVWVVAQNRMTLKQL  
KDRYPRKRFVAGVGANSKISALKGSKGKDCEIPVPVGISVTDENGKIIGELNKENDRILVAQ  
GGLGGKLLTNFLPLKGQKRIIHLDLKLIADVGLVGFPNAGKSSLLSCVSHAKPAIADYAF  
TTLKPELGKIMYSDFKQISVADLPGLIEGAHMNKGMGHKFLKHIERTRQLLFVVDISGFQ  
LSSHTQYRTAFETIILLTKELELYKEELQTKPALLAVNKMDLPDAQDKFHELMSQLQNP  
DFLHLFEKNMIPERTVEFQHIIPISAVTGEGIEELKNCIRKSLDEQANQENDALHKKQLL  
NLWISDTMSSTEPPSKHAVTTSKMDII

>AC006153.3-002

MVHCSCVLFRKYGNFIDKLRLFTRGGSGMGYPRLGGEGGKGGDVWVVAQNRMTLKQL  
KD  
RYPRKRFVAGVGANSKFPNAGKSSLLSCVSHAKPAIADYAFTTLKPELGKIMYSDFKQIS  
VADLPGLIEGAHMNKGMGHKFLKHIERTRQLLFVVDISGFQLSSHTQYRTAFETIILLTK  
ELELYKEELQTKPALLAVNKMDLPDAQDKFHELMSQLQNPKDFLHLFEKNMIPERTVEFQ  
HIIPISAVTGEGIEELKNCIRKSLDEQANQENDALHKKQLLNLWISDTMSSTEPPSKHAV  
TTSKMDII