

# Bioinformatics 2: Lecture 9

- MSAs in MOE
- modeling outgaps
- Fixing errors.

# 9.1 MSAs in MOE

- finding SCRs

# Jargon word: SCR

- Structurally Conserved Regions (SCR) are assumed to be structurally invariant.
- SCRs should generally be 'fixed' during energy minimization. ( Initially fix all atoms, and finally just fix the backbone atoms. )



From a multiple structure alignment like this you can identify SCRs versus designated/variable Loops. SCRs can include multiple SSEs and the loops between them.

# Jargon word : Loop

Three types of loops:

- **Designated Loop:** coordinates derived from a *template*. Not variable from model to model. But *not fixed* during energy minimization.
- **Variable Loop:** coordinates derived from a *database search* or simply constructed. May be variable from model to model. Not fixed during energy minimization.
- **Outgap:** a variable loop at the end of a chain. May be derived from a secondary structure prediction or experimental data.

## exercise9.1

# Identify SCRs in a superfamily

- Open “exercise9.1.moe” from the course web page.
- This is a multi-structure alignment, showing the flexibility of this protein family over the course of evolution. SCR (structurally conserved regions) are places in the structure that evolution does not touch.
- Identify SCRs in a structure family.
- Why have these parts of the molecule been highly conserved for structure?

# 9.2 Outgaps

- disordered
- ordered

±	Tag	Chain	1	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	79	
	2GI9	1: 2GI9.A																		
	2PTL	2: 2PTL.A																		

MQYKLILNGKT-LKGETTTEAVDAATAEKVFKQYANDNGV---DGEWYDDATKTFTVTE  
ENKEETPETPETDSEEBVTIKANLIFANGSTQAEFKGTFE-KATSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFA

# Is it ordered or disordered?

- DISOPRED -- a program for predicting disorder/order.
- Intrinsically unstructured proteins are characterized by a low content of bulky hydrophobic amino acids and a *high proportion of polar and charged amino acids*.
- Example: N-term of 2ptl

> Code: 2PTL Chain: \_ 78 residues.

**ENKEETPETPETDSEEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAY**

**ADTLKKDNGEYTVDVADKGYTLNIKFAG**

[http://cms.cs.ucl.ac.uk/typo3/fileadmin/bioinf/Disopred/disopred\\_help.html#seq\\_help](http://cms.cs.ucl.ac.uk/typo3/fileadmin/bioinf/Disopred/disopred_help.html#seq_help)

or google DISOPRED.

# Ordered outgaps

1. Find a way to extend the domain.

Search SCOP or CATH for protein of similar topology with an additional SSE.

(examples in exercise9.1.moe\*)

2. Outgap is a separate domain.

Search PDB using BLAST, using the outgap sequence. Search CDD. \*\*Search PROSITE for coiled coil, kringle, zinc finger, etc.

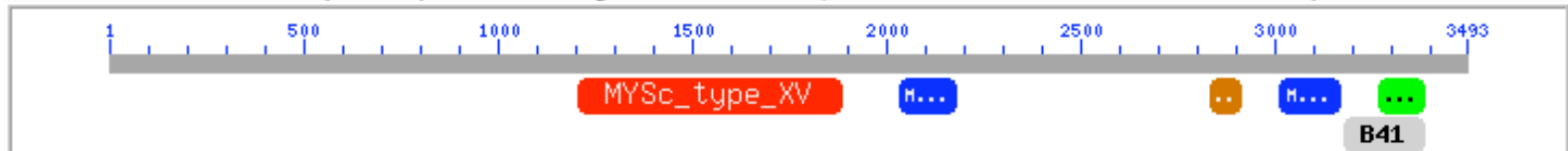
\*1GIT, 1GIS versus all others. C-terminal outgap is cys-rich beta.

\*\*1PWB, 1HUP, 1R13 vs all others. N-terminal outgap is probably coiled-coil.

Query sequence: [[gi|157041248|ref|NP\\_001096641.1|](#)]  
 myosin XV isoform 3 [Mus musculus]

Concise Result  Full Result  Show Search Information [?](#)

Click on the **colored bar** for a conserved domain to **view your query sequence** within the multiple sequence alignment for that domain. To see only the sequences used to generate the domain, click on its **PSSMID** in the tabular summary.




### Descriptions

	Title	Pssmid	Multi-Dom	E-value
[+]	<a href="#">cd01387</a> , MYSc_type_XV, Myosin motor domain, type XV myosins. In vertebrates, myosin XV...	<a href="#">30109</a>	Yes	< 1.0e-180
[+]	<a href="#">smart00139</a> , MyTH4, Domain in Myosin and Kinesin Tails; Domain present twice in myosin-...	<a href="#">47478</a>	No	8e-40
[+]	<a href="#">smart00139</a> , MyTH4, Domain in Myosin and Kinesin Tails; Domain present twice in myosin-...	<a href="#">47478</a>	No	2e-32
[+]	<a href="#">pfam00373</a> , FERM_M, FERM central domain. This domain is the central structural domain o...	<a href="#">84727</a>	No	4e-06
[+]	<a href="#">pfam07653</a> , SH3_2, Variant SH3 domain. SH3 (Src homology 3) domains are often indicativ...	<a href="#">87319</a>	No	0.001
[+]	<a href="#">smart00295</a> , B41, Band 4.1 homologues; Also known as ezrin/radixin/moesin (ERM) protein...	<a href="#">47624</a>	Yes	6e-05

[Search for similar domain architectures](#)

#### CD Search Reference:

 Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", *Nucleic Acids Res.*32(W)327-331.

# Exercise 9.2: a better loop search

- Download files linked to "Feb22"
- Find the modeling errors.
- Align target (gi 168041592) to template (1T15.A) using **SEQ: Alignment | Align**
- Display template as backbone atoms only.
- Select the template residues bordering each deletion (there are no insertions). **Ask *is it possible to put a deletion at that point in the structure?*** \_\_\_\_\_ . If not, why not?  
\_\_\_\_\_
- Move the deletion to a loop position. ↓

```
1 gi|168041592|ref|XP_001773275.1| predicted protein [Physcomitrella patens subsp. patens]
  SLFAGFYI·YPTPHVQPPLDTIVKLS EAAGGKVLKSLDEALQQNYVSHSIVLGGE···EDKSEVENAAKAGLRTFTGEWFMQAI·VKQKIDLDTNILQ
2 1T15.A ANTITUMOR PROTEIN
  IFRGLEICCYGPF1TNHPTDQLEW2MVQLCGASV3VVKEISSFTLGTGV4HP5IVVV6QPD7AWT8EDNGF9HAI10GQM11CEAP12VVTRE13WVLD14SV15ALY16QC17QEL18DTY19LIP20QIP
```



# The Easy Way. Too easy.

<http://swissmodel.expasy.org//SWISS-MODEL.html>

## SWISS MODEL WORKSPACE

[ [Workspace](#) ] [ [Modelling](#) ] [ [Tools](#) ]


[ [Repository](#) ] [ [General Info](#) ] [ [Links](#) ] [ [Help](#) ]

[ [login](#) ]

### SwissModel Automatic Modelling Mode

Email:

Project Title:

Provide a protein sequence or a UniProt AC Code: 

LFAGFYIYPTPHVQPPLDTIVKLSEAAGGKVLKSLDEALQONYVSHSIVLGGEEKSEVENAAKAGLRT  
PTGEWFMOAIVKQKIDLDTNILO

### Options:

BLAST E-value limit  to select

Use a specific template:

Read the SWISS-MODEL paper

## 9.3 Finding modeling errors

# Modeling errors\*

- Backbone errors

- indel in helix

realign

- indel in strand

realign

- loose loops

realign, MD

- missing H-bonds

restrain, minimize

- Sidechain errors

- buried charges, polar sidechains

realign

- too many exposed hydrophobics

realign

- $\phi > 0$  and not Gly, Asn

realign or  
minimize

- $(\phi < -90 \text{ or } \phi > 0)$  and Pro

realign or

- Voids

MD, rotamer search,  
minimize

\*Here we are omitting obvious errors: collisions, stretched bonds, distorted planar groups, etc.

# Homework 5

- Find and fix errors in a modeled structure.
- Due Feb 28. Turn in MOE files with comments/answers attached.