de novo structure prediction methods for transmembrane proteins and their application to the divisome

J. M. Crowet, L. Lins

Laboratoire de Biophysique Moléculaire aux Interfaces, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium
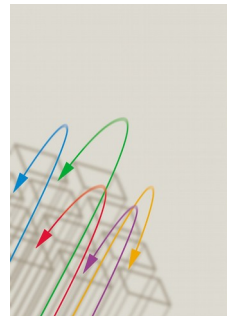
Membrane proteins

- Approximately 30 % of the genes of a typical genome
- Crucial roles in a diverse range of essential biological processes

  Transport of ions and small molecules, intercellular communication and signal transduction

- About 60 % of current drug targets are membrane proteins (MP)
- Represents about 1 % of the PDB structures
- Technical difficulties with purification and structure determination
- Computational structure prediction is an interesting approach

Structure prediction methods for membrane proteins

■ Homology modelling

Needs a suitable template

Membrane protein-specific tools can be used

Fold preservation in TM regions requires less sequence conservation than for GP

Relatively accurate 3D models even at low sequence identity (<20 %)

Small number of solved membrane protein structures

■ *de novo* prediction

It only uses protein sequence and knowledge-based potentials

Methods :

   - **RosettaMembrane** (Yarov-Yarovoy 2006, Barth 2007, 2009)

   - **BCL::MP-Fold** method (BioChemical Library ; Weiner 2013)

   - **Film3** (Folding In Lipid Membrane ; Nugent 2012)

de novo structure prediction methods for membrane proteins

RosettaMembrane

Implicit representation of the membrane with terms in the energy function that describe intraprotein and protein–solvent interactions

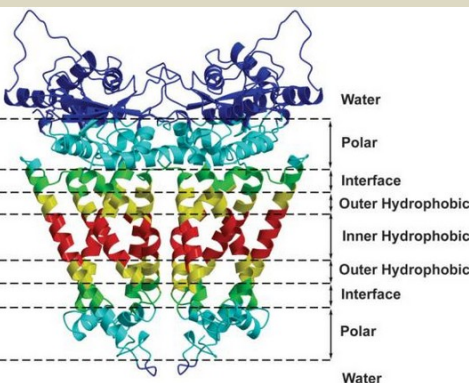Assembles fragments of known structures

A topology is needed and TMs are restrained in the membrane boundaries

Scores are computed from a FF that uses a CG representation of the side chains
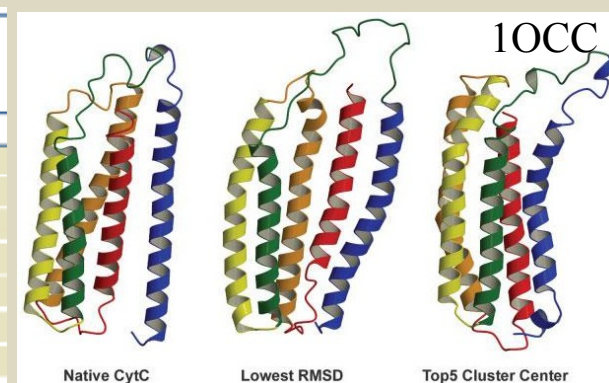
100.000 structures are generated and clusters are computed

A structure close to the native one is usually found among the 5 best clusters

Predicts the structures of 12 membrane proteins (< 150 AA ; rmsd < 4Å)

Water
Polar
Interface
Outer Hydrophobic
Inner Hydrophobic
Outer Hydrophobic
Interface
Polar
Water

PDB ID Code	TMH	Domain	Residues	Rosetta-Membrane			
				RMSD100		Score	
				Best	Top 5%	Best	Top 5%
Traditional Folding							
2BG9	3	A: 211–301	91	3.9	5.5	10.7	10.2
1PY6	4	77–199	123	2.2	3.2	2.8	5.7
2PNO	4	A: 2–131	130	4.0	5.1	9.4	8.2
1OCC	5	C: 71–261	191	7.3	9.2	10.8	11.3
2NR9	6	4–195	192	5.6	7.6	10.5	10.5
1PY6	7	5–231	227	3.3	5.8	8.4	8.7
3HD6	12	6–448	403	8.0	9.5	11.9	11.7

1OCC

Native CytC   Lowest RMSD   Top5 Cluster Center

de novo structure prediction methods for membrane proteins

■ BCL::MP-Fold method

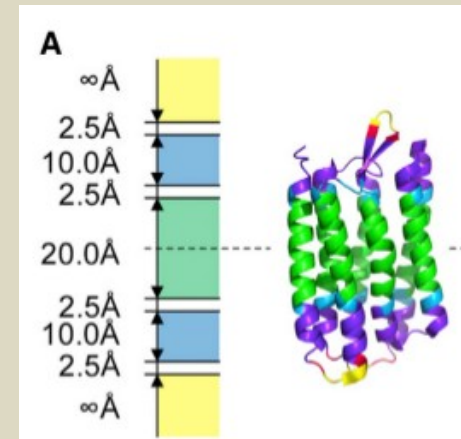Rapidly assembles secondary structure elements (SSEs)

1000 structures are generated through a 6 stages process

Models are built into a static membrane object

Scoring function accounts for the membrane environment

A symmetry folding mode can predict homomultimers

The correct topolgy is sampled when the RMSD100 is lower than 8Å

PDB ID Code	TMH	Domain	Residues	BCL::MP-Fold				Rosetta-Membrane			
				RMSD100		Score		RMSD100		Score	
				Best	Top 5%	Best	Top 5%	Best	Top 5%	Best	Top 5%
Traditional Folding											
2BG9	3	A: 211–301	91	2.8	3.4	9.9	6.7	3.9	5.5	10.7	10.2
1PY6	4	77–199	123	3.9	4.7	5.4	6.4	2.2	3.2	2.8	5.7
2PNO	4	A: 2–131	130	5.0	6.7	5.4	8.6	4.0	5.1	9.4	8.2
1OCC	5	C: 71–261	191	4.6	5.9	8.5	8.0	7.3	9.2	10.8	11.3
2NR9	6	4–195	192	5.7	7.2	8.7	9.5	5.6	7.6	10.5	10.5
1PY6	7	5–231	227	4.8	5.9	6.1	8.4	3.3	5.8	8.4	8.7
3HD6	12	6–448	403	7.2	8.2	11.0	10.3	8.0	9.5	11.9	11.7

1OCC
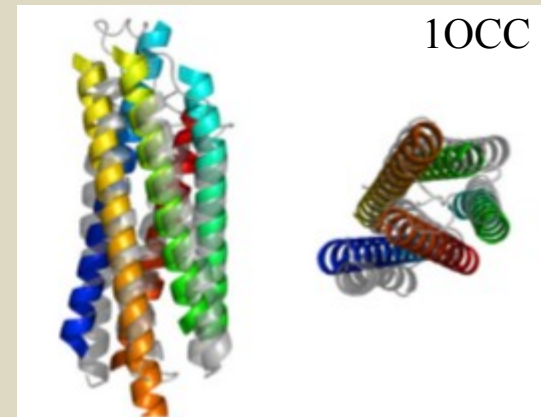
de novo structure prediction methods for membrane proteins

■ Film3

Identifying residue-residue contacts from multiple sequence alignments (MSAs)

- Detection of correlated mutations between sites in MSAs

- Residues proximal in 3D space are likely to impose constraints on each other

- If mutation to complementary pairing, the contact is more likely to be retained

- Due to rapid growth in the size of sequence databases (x10 in 5 years)

Uses predicted secondary structure and transmembrane topology

Conformational search using SSEs to assemble the tertiary fold

FILM3 uses predicted contacts for scoring and adds a membrane potential

It uses modeller to recombine candidate structures to generate a final model

Model of the Cytochrome c oxidase with a rmsd of 5.7Å (514AA and 12TMs)

Qualtity assessment

- Number of contacts, predicted TM score, temperature factor, TMs inside
  membrane

Aim of the molecular modelling study

- Except for MraY the structures of the membrane proteins from the divisome are not known and there is no homolog

- Even if it seems easier than for globular proteins, their *ab initio* prediction remains a difficult task

- Use the modelling methods available to predict the structures of FtsX, FtsK and FtsW

FtsX : predictions by RosettaMembrane

- FtsX from *E coli* is a protein of 352 AA

- FtsX is predicted to have 4 TMs

- Rosetta predictions :

  62462 structures generated and the score of the best model is -133

Score vs RMSD (Å)
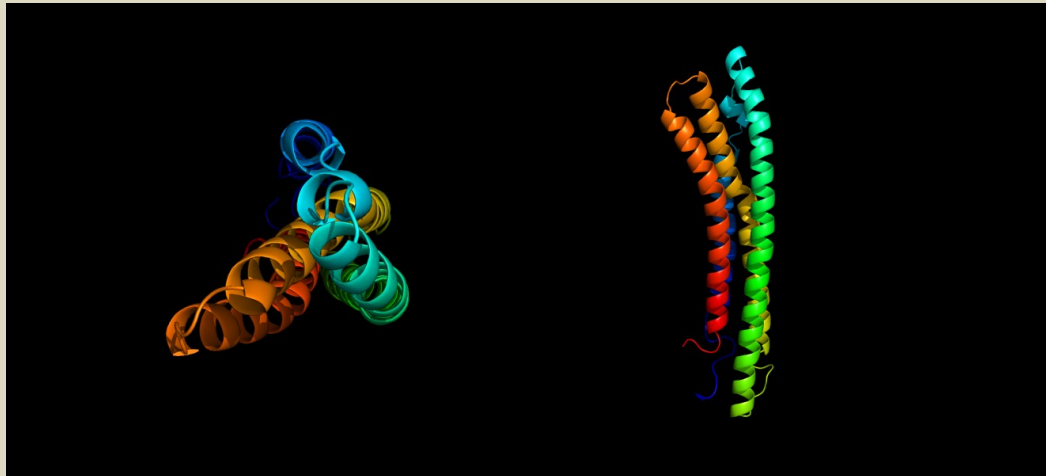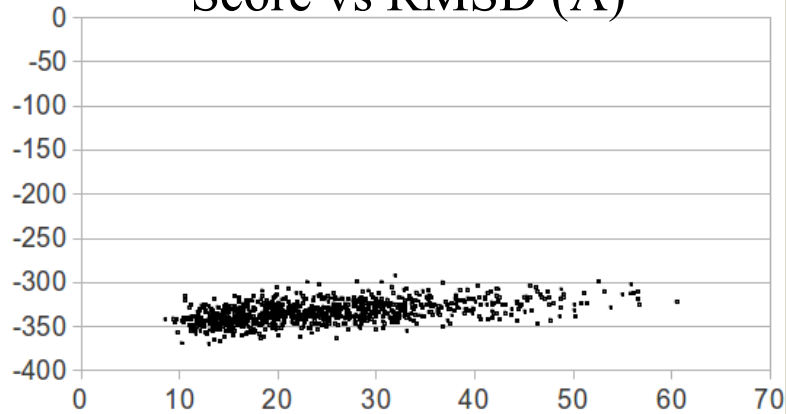
FtsX : predictions by BCL::MP-Fold method

- The model 825 has the lowest score (-14026)
- The clustering shows 2 clusters with more than 100 structures
- We can assume that the native structure has been sampled
- It is not possible from the clustering to define a good model
- Tuning the TMs definition could give better results

Instead of using SPOCTOPUS

clusters	Model N°	Nb (/1000)	Middle RMSD
Clus 1	226	106	4.06 Å
Clus 2	39	101	4.07 Å
Clus 3	670	61	4.15 Å
Clus 4	82	56	3.92 Å
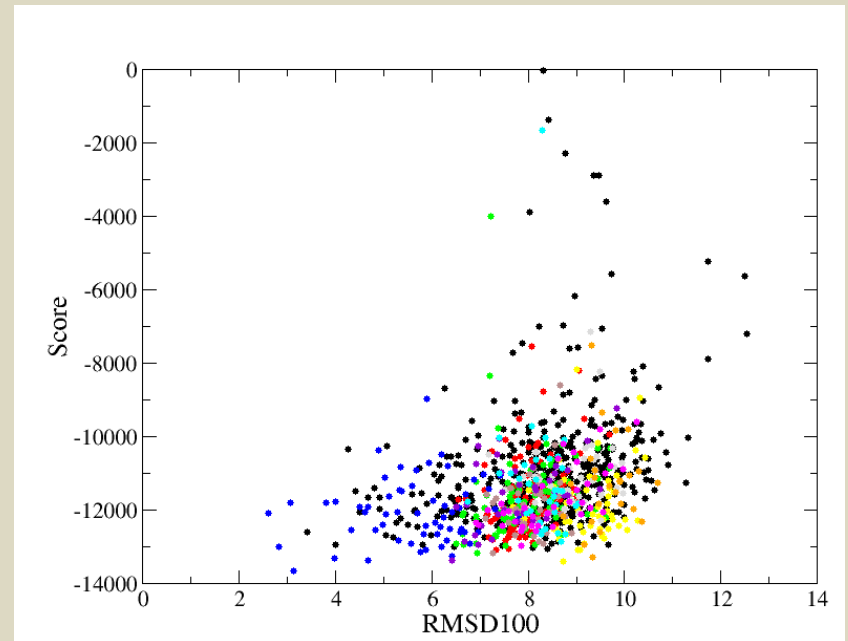
FtsX : predictions by Film3

- Computations have been done with and then without the loops

- 108 models have been built with Z restraints and 108 without

- MSA with 7815 aligned sequences

  59969 Contacts, P > 0.5 for **110**

- The largest cluster = 4 % (100 best)

  Mean TM-Score = 0.22

  Mean RMSD = 13.8

- 24 Contacts with P > 0.5 within 5Å

- 100 models are used to refine the model

- Expected final TM-score is 0.43

- TMs are not spanning the membrane
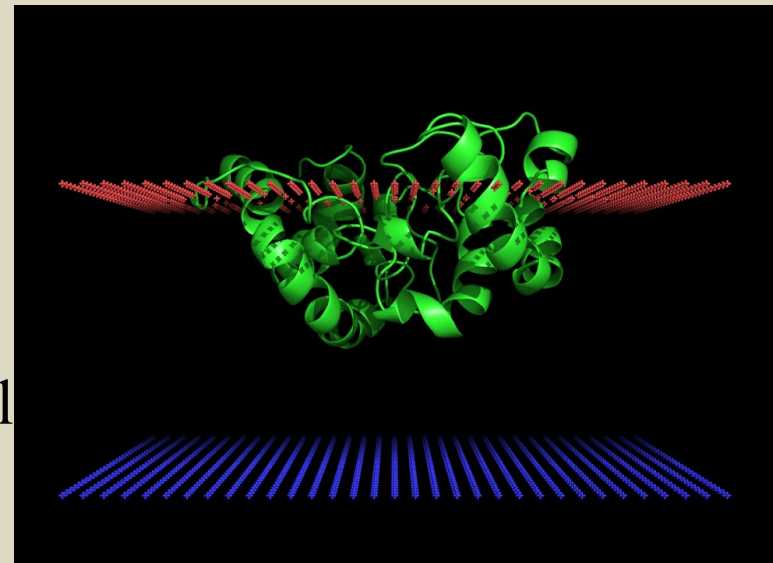
FtsK

- FtsK from *E coli* is a protein of 1329 AA
- It has 4 TMs which have been validated experimentally by Dorazi in 2000
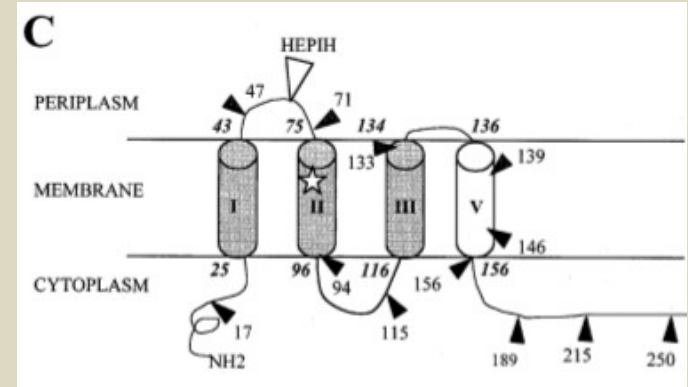
FtsK : predictions by RosettaMembrane

- FtsK from *E coli* is a protein of 1329 AA

- It has 4 TMs which have been validated

  experimentally by Dorazi in 2000

- Rosetta predictions :

  17208 structures generated and the score of the best model is -145
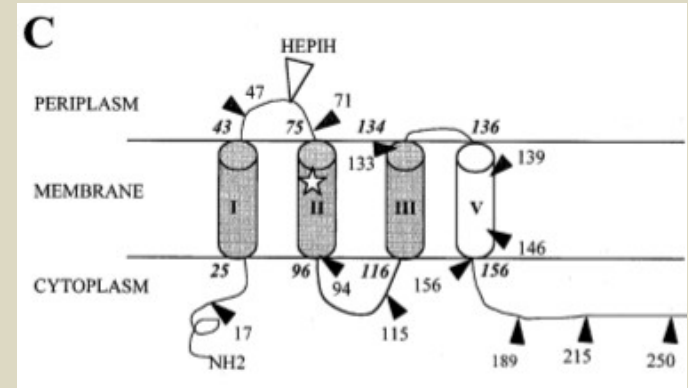
Score vs RMSD (Å)
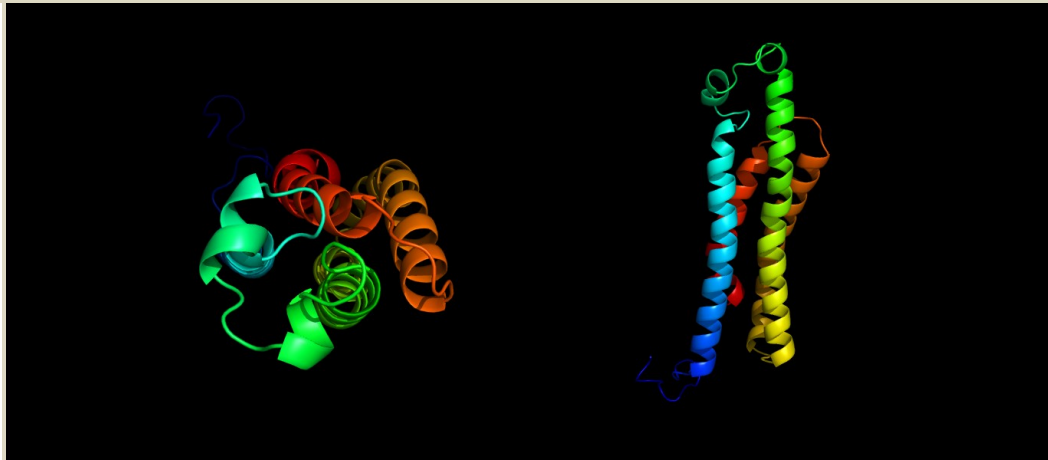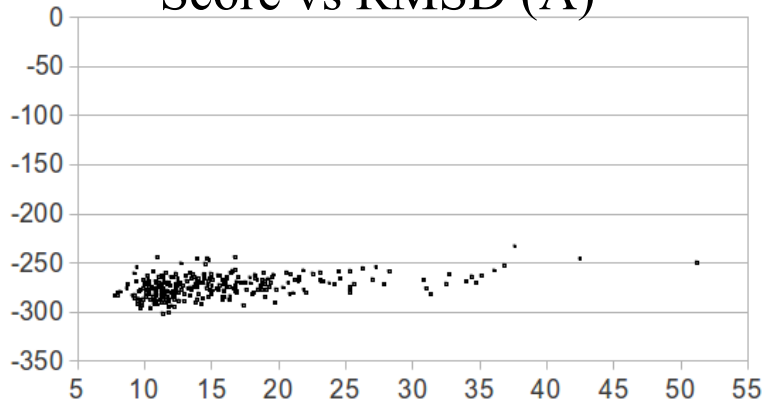
FtsK : predictions by BCL::MP-Fold method

- The model 828 has the lowest score (-11532)
- There are 4 clusters that represent 75% of the sampled structures
- These are then good models for the FtsK structure.

Clusters	Model N°	Nb (/1000)	Middle RMSD
Clus 1	29	202	3.63 Å
Clus 2	863	185	3.54 Å
Clus 3	198	184	3.47 Å
Clus 4	945	176	3.26 Å
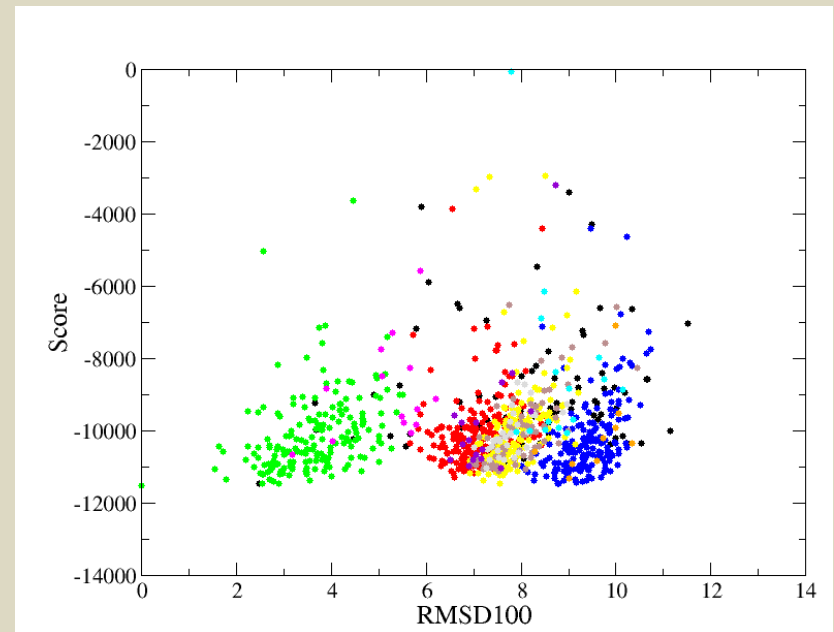
FtsK : predictions by BCL::MP-Fold method

- The 4 clusters present different topologies
- Several interactions between helices are conserved
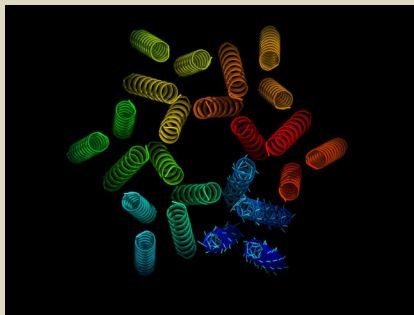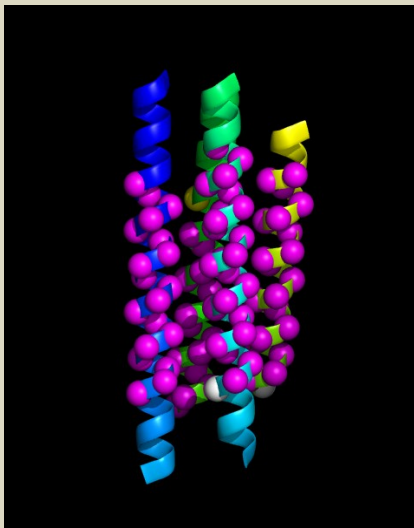
Clus 1
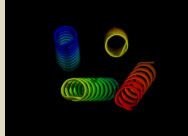
Clus 2
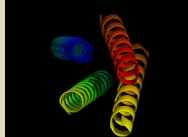
Clus 3
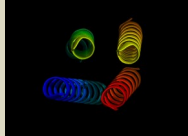
Clus 4
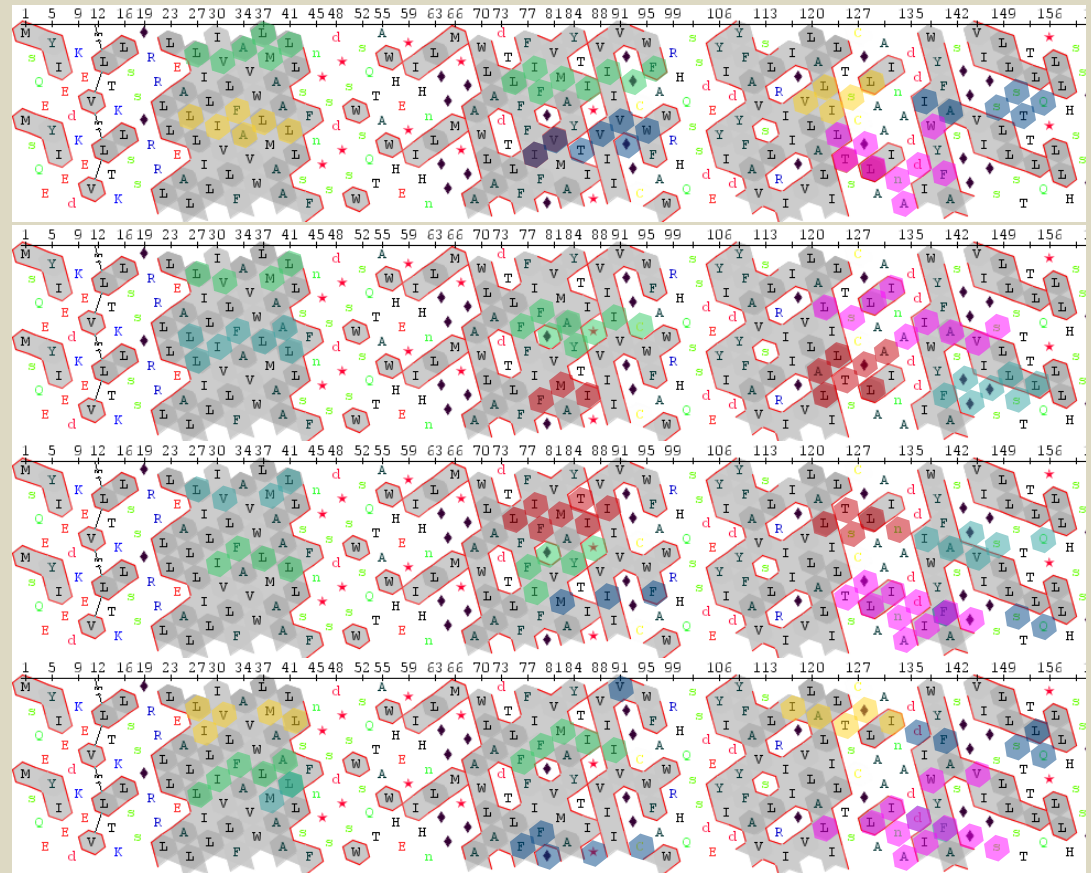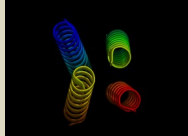
H1-H2   H1-H3   H1-H4   H2-H3   H2-H4   H3-H4

FtsK : predictions by Film3

- Not enough aligned sequence to make the covariance analysis (109)
- Moreover, FtsK is an hexamer

MraY

- MraY from *E coli* is a protein of 359 AA

- It has 10 TMs

- The structure of MraY from *Aquifex aeolicus* has been resolved in 2013 by Chung *et al*. (4J72)

- This protein has 48% identity with the MraY from *E. coli*
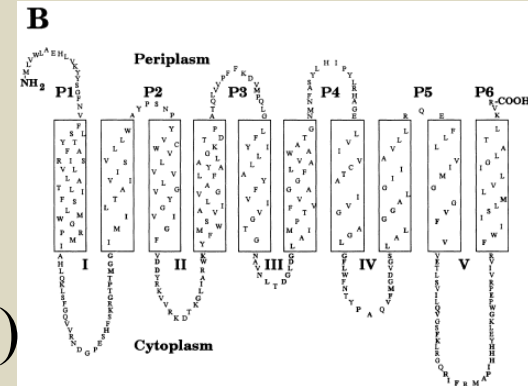
MraY : predictions by RosettaMembrane

- MraY from *E coli* is a protein of 359 AA

- It has 10 TMs

- The structure of MraY from *Aquifex aeolicus*

  has been resolved in 2013 by Chung et al. (4J72)

- This protein has 48% identity with the MraY from *E. coli*

- Rosetta predictions :

5673 structures generated and the score of the best model is -158
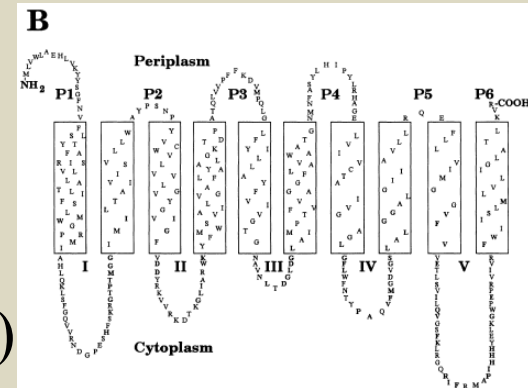
Score vs RMSD (Å)
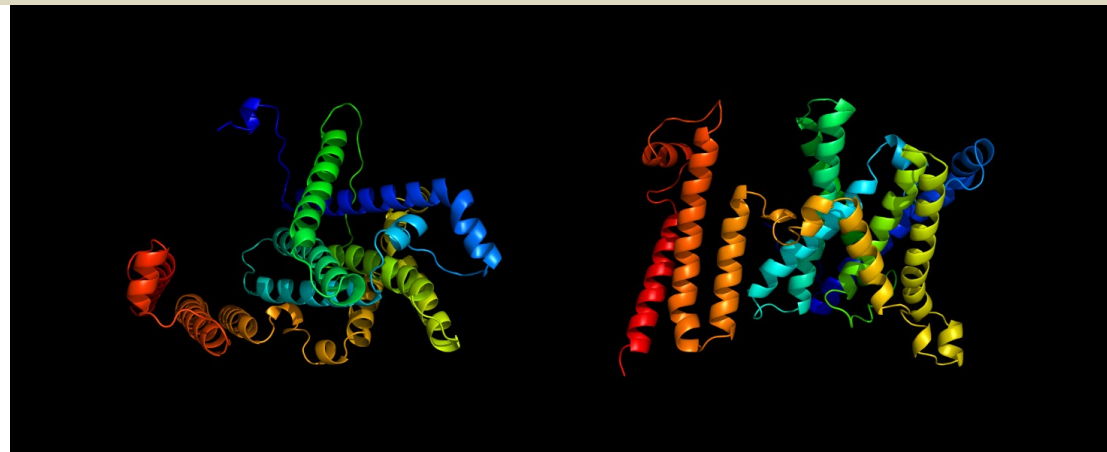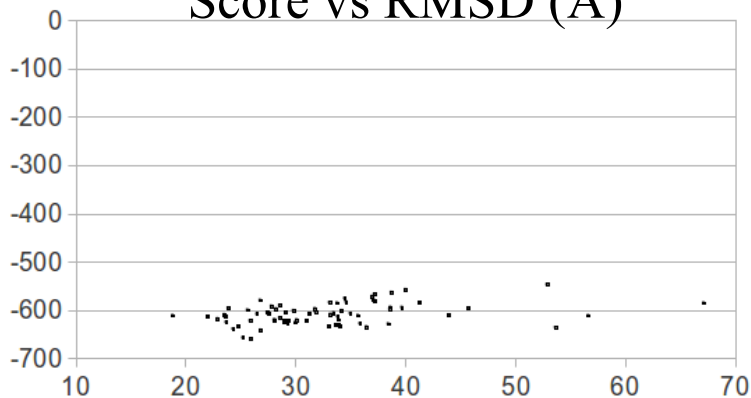
MraY : predictions by BCL::MP-Fold method

- 1000 models have been built

- The model 984 has the lowest score (-28679)

- The sampled structures are too distant from each other

- No clusters can be defined
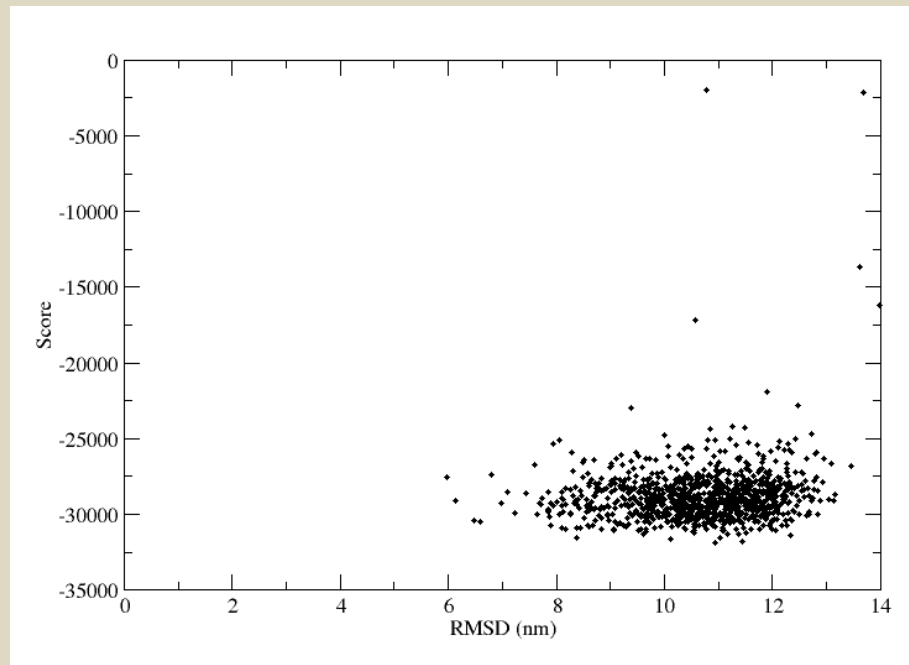
MraY : predictions with Film3

- 108 models have been built with Z restraints and 108 without

- MSA with 2714 aligned sequences

  60427 Contacts, P > 0.5 for **178**

- The largest cluster = 39 % (100 best)

  Mean TM-Score = 0.32

  Mean RMSD = 15.5

- The best model has a score of -391

- 62 Contacts with P > 0.5 within 5Å

- 100 models are used to refine the model

- Expected final TM-score is 0.5

PDB

Best model
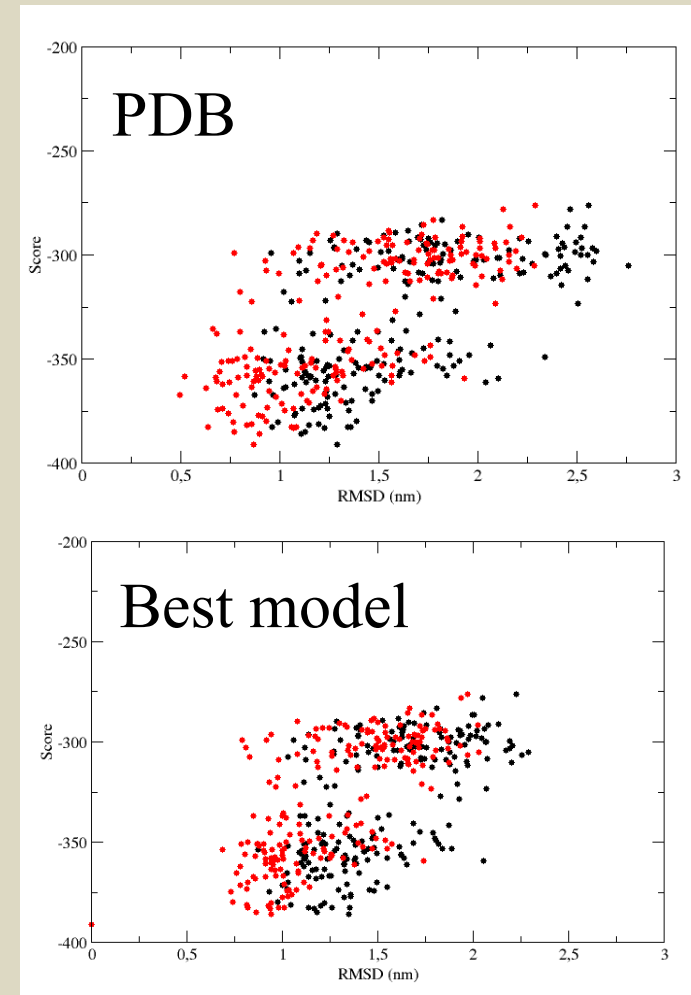
# *MraY : predictions with Film3*

- The TMs are inside the membrane

- After refinement, the model has 77 Contacts with P > 0.5 within 5Å

- PDB has 88 Contacts and a RMSD of 9.9Å and 7.6Å for TMs
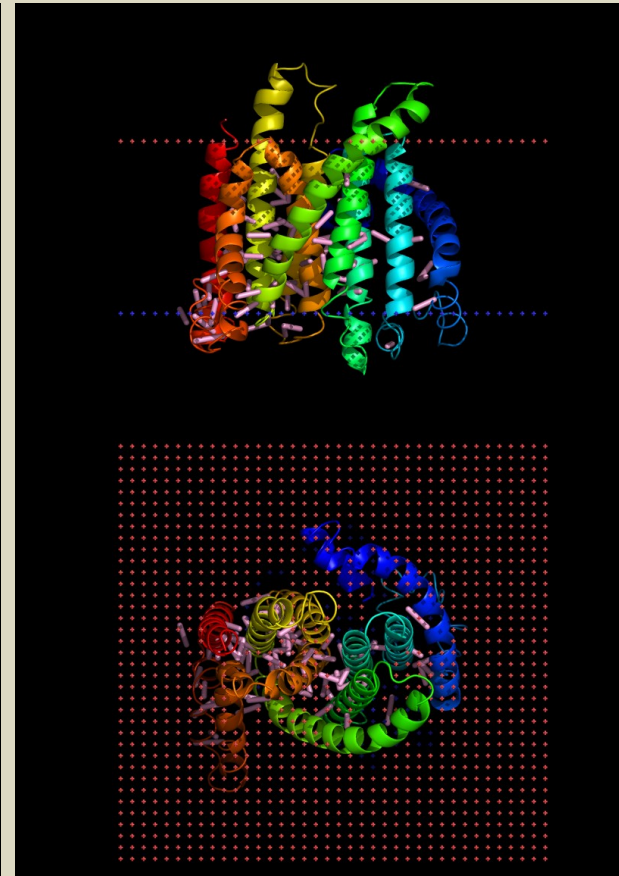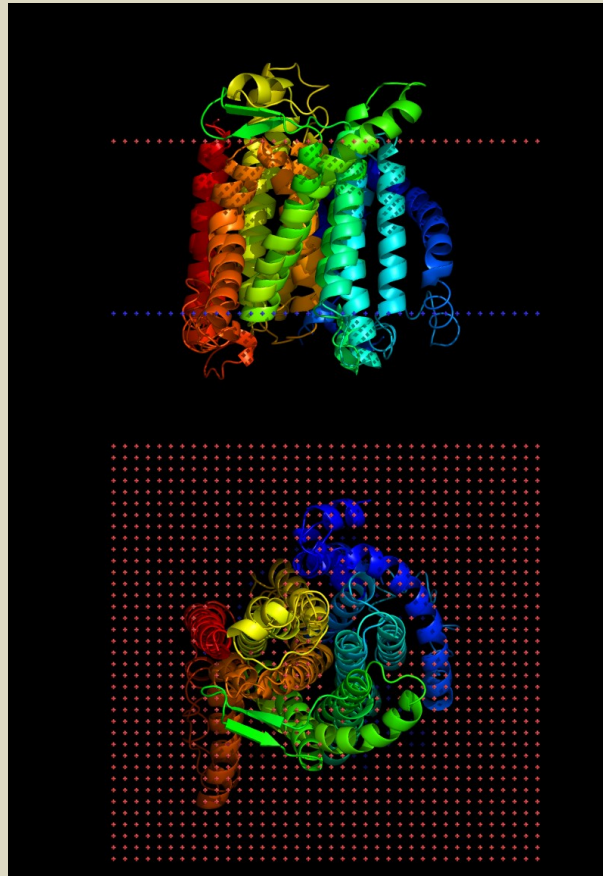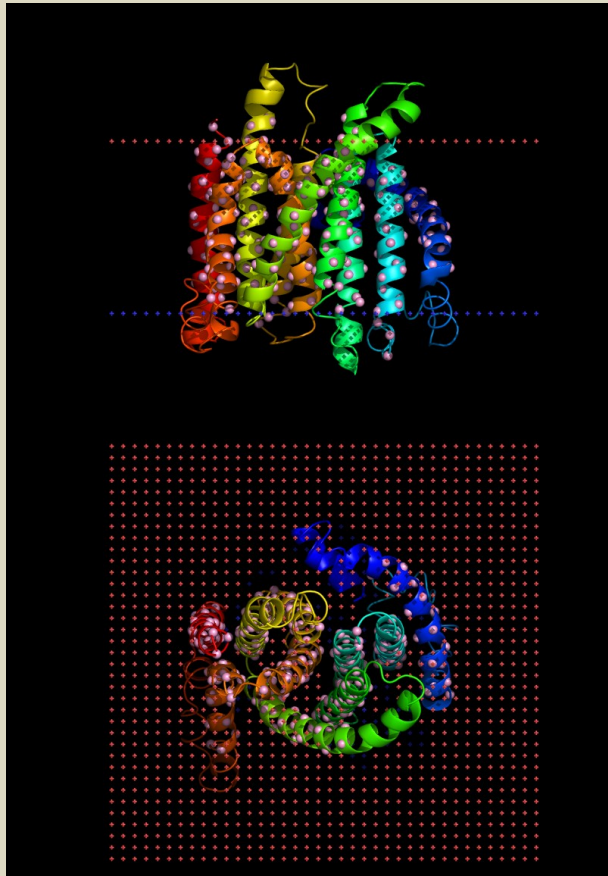
MraY : predictions with Film3

- Additional 108 models have been built with Z restraints
- The largest cluster has 59 % of the 100 best structures

  Mean TM-Score = 0.35

  Mean RMSD = 14.3

- The best model has a score of -391.0
- It has 62 Contacts with P > 0.5 within 5Å
- 100 models are used to refine the model
- Expected final TM-score is 0.52
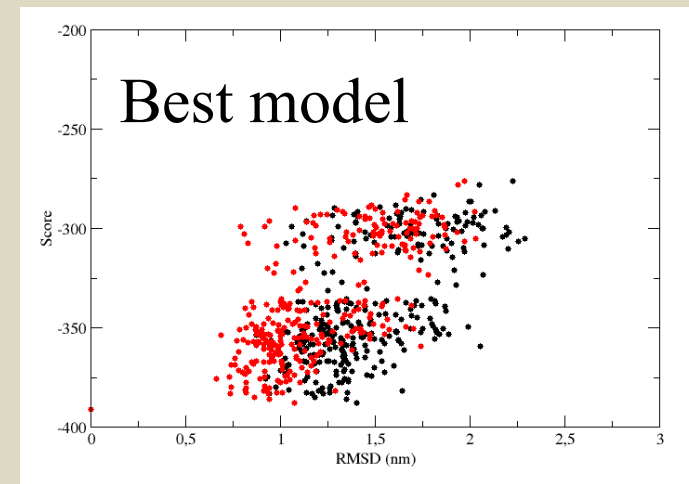
Best model

MraY : predictions with Film3

- The TMs are inside the membrane

- After refinement, the model has 76 Contacts with P > 0.5 within 5Å

- PDB has 88 Contacts and a RMSD of 8.1Å and 4.2Å for TMs
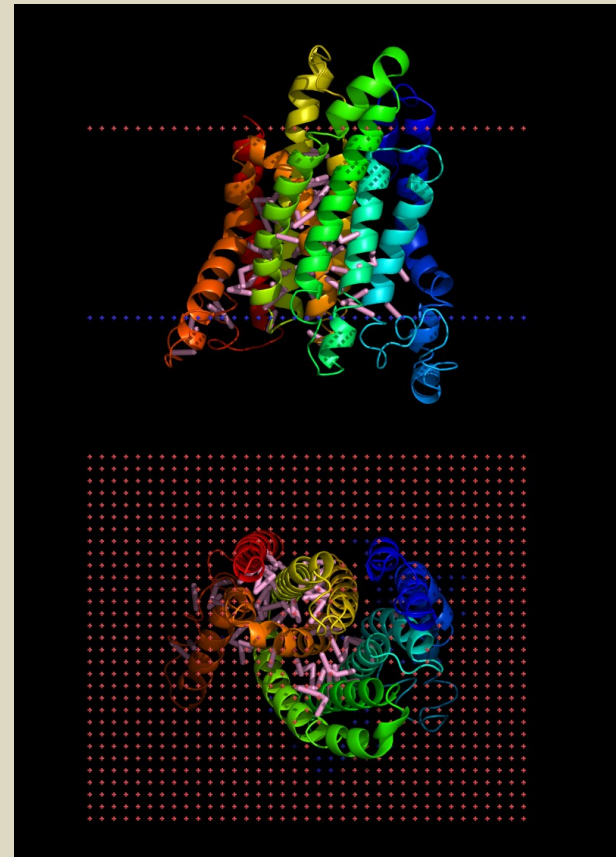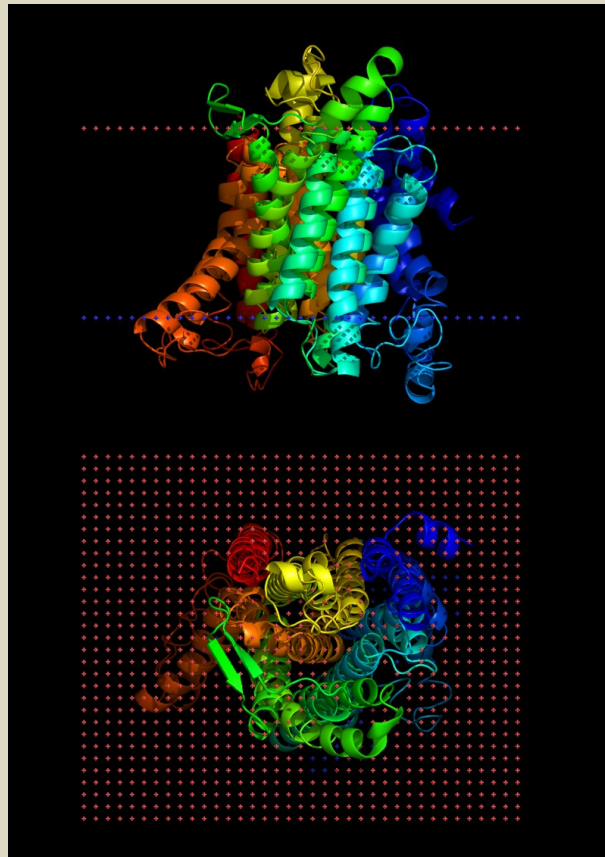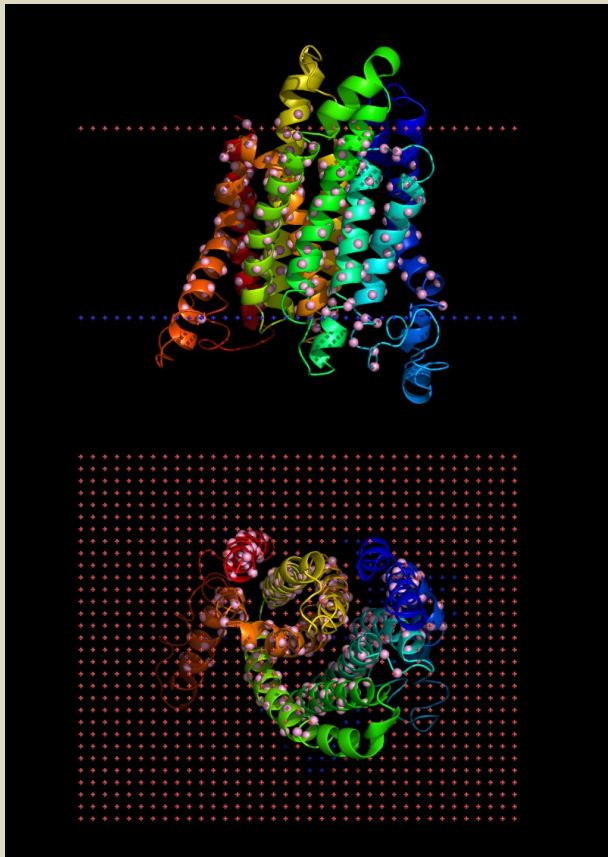
FtsW

- FtsW from *E coli* is a protein of 414 AA
- It has 10 TMs which have been validated experimentally by Lara in 2002
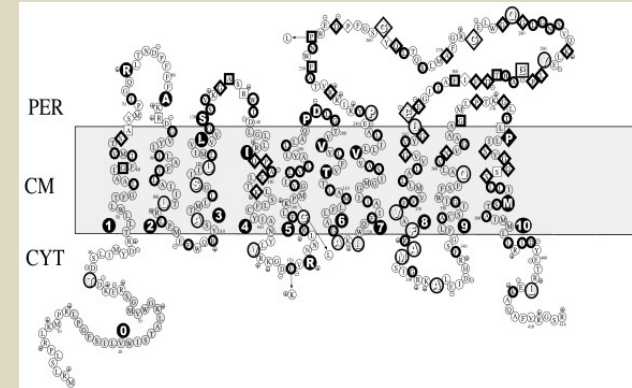
FtsW : predictions by RosettaMembrane

- FtsW from *E coli* is a protein of 414 AA
- It has 10 TMs which have been validated experimentally by Lara in 2002
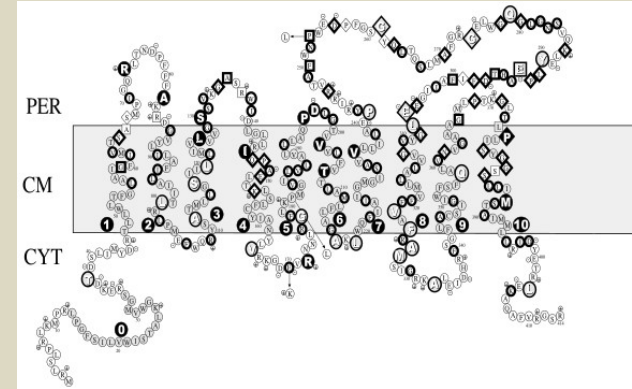
- Rosetta predictions :

14092 structures generated and the score of the best model is -202

Score vs RMSD (Å)
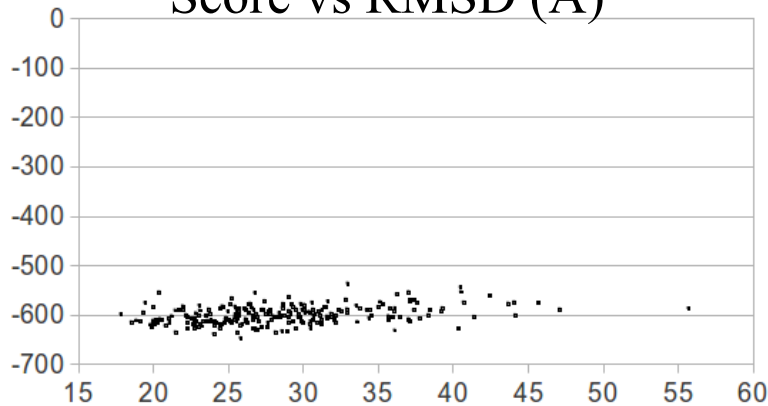
FtsW : predictions by BCL::MP-Fold method

- 1000 models have been built
- The model 118 has the lowest score (-27885)
- The sampled structures are too distant from each other
- No clusters can be defined
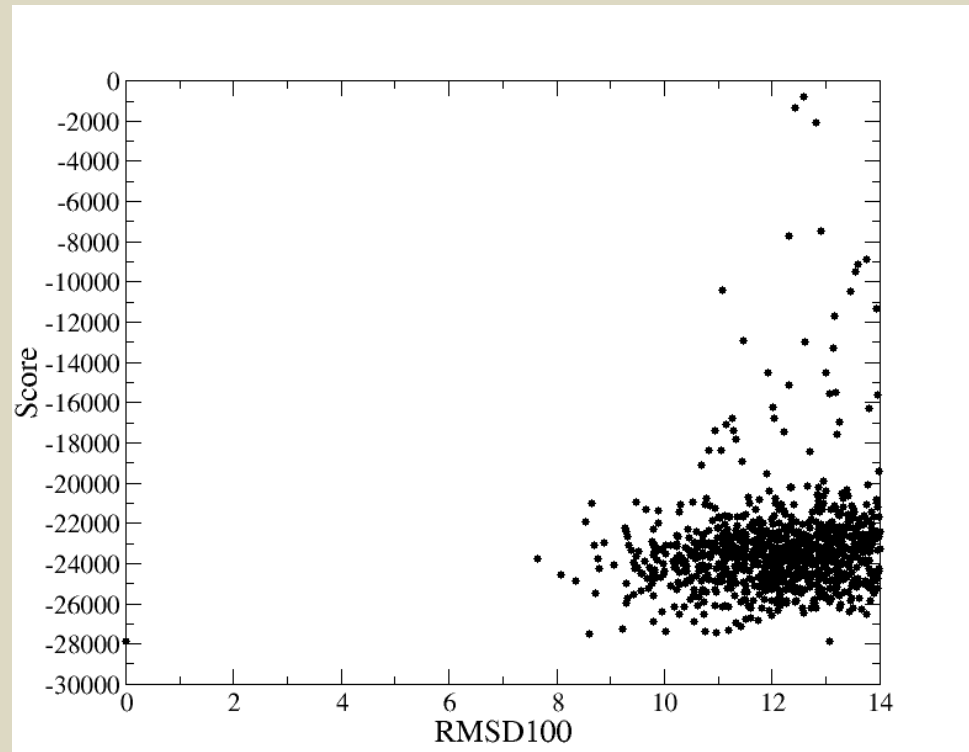
FtsW : predictions with Film3

- 216 models have been built with Z restraints and 108 without

- MSA with 3293 aligned sequences

  81395 Contacts, **248** with a proability > 0.5

- The largest cluster has 93 % of the 100 best structures

  Mean TM-Score = 0.44

  Mean RMSD = 12.7
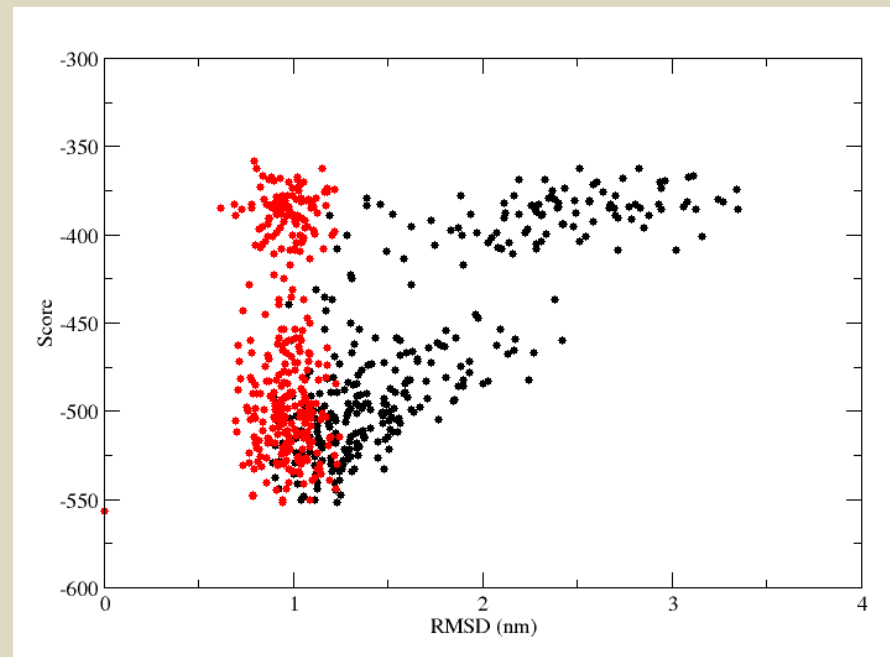
FtsW : predictions with Film3

- The best model has a score of -556.4

- 105 Contacts with P > 0.5 within 5Å

- 100 models are used to refine the model

- Expected final TM-score is 0.58

- The TMs are inside the membrane
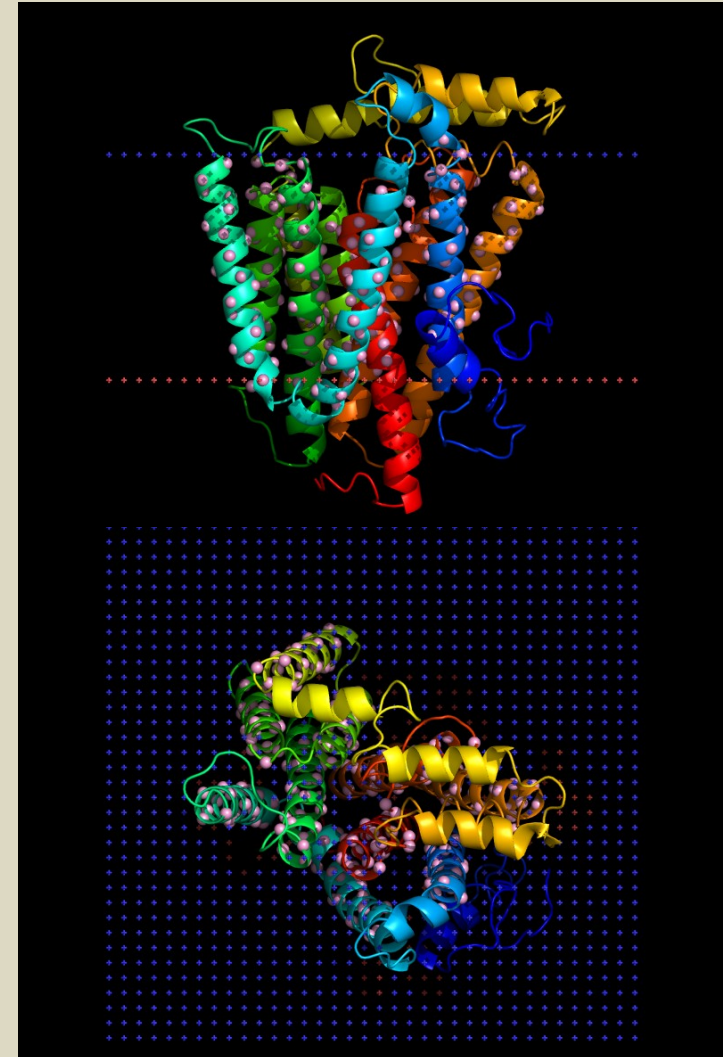
FtsW : predictions with Film3

- The best model has a score of -556.4

- 105 Contacts with P > 0.5 within 5Å

- 100 models are used to refine the model

- Expected final TM-score is 0.58

- The TMs are inside the membrane

- After refinement, the model has

  127 Contacts with P > 0.5 within 5Å
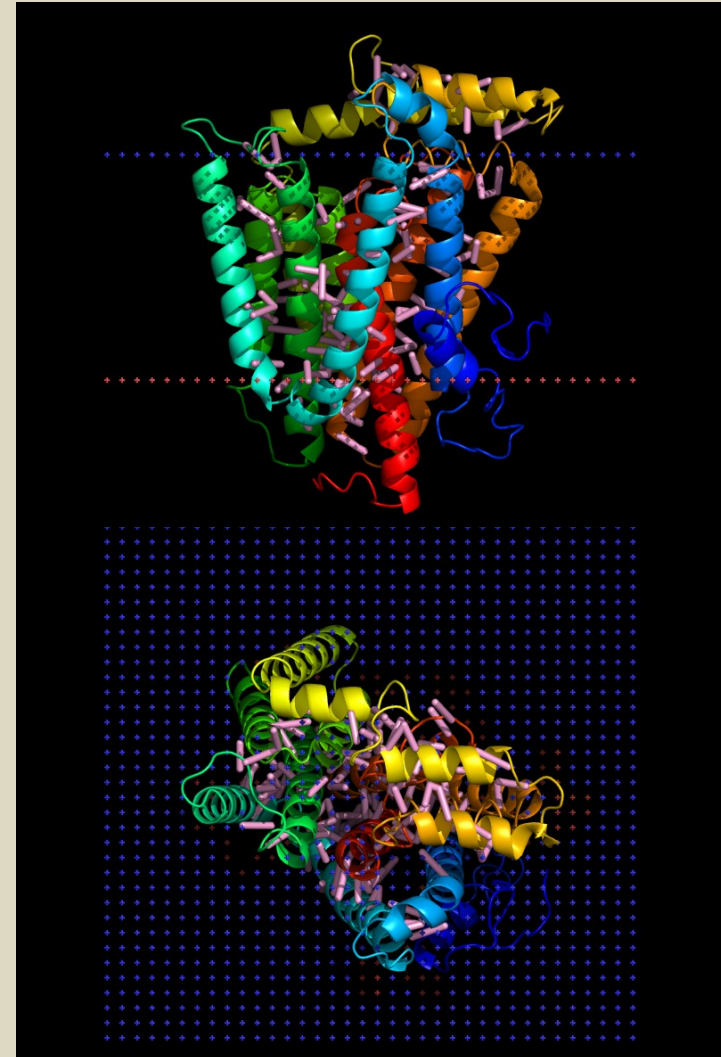
FtsW : MD simulations

- Topologies for PG and PE are available in AT and CG
- Topologies for the lipid II are available in AT
- Topology for the lipid II has been made for CG
- Builder and CG2AT tool have been adapted for the Lipid II
- Aim of the MD study :

Test the model stability

Test the transport ability

of the model
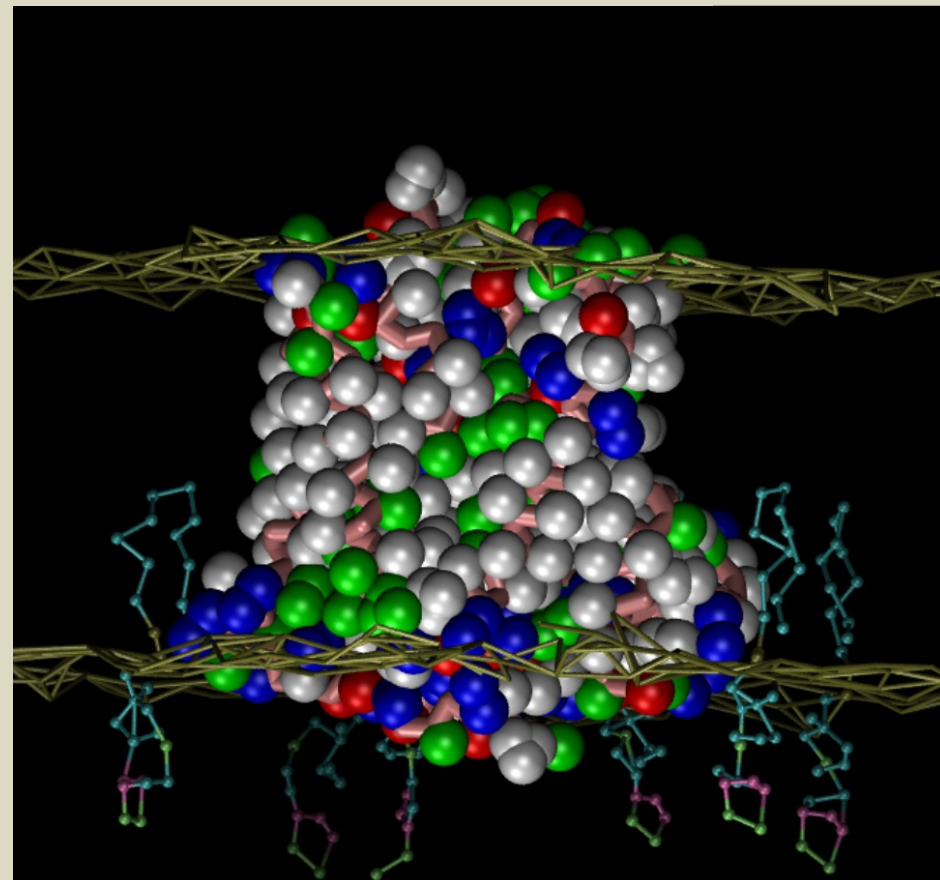
Conclusions

- *de novo* prediction of membrane protein structures remains a difficult task

- Several method are available and have been tested on the membrane proteins from the divisome

- The BCL and Film3 methods give the best results

- Good models have been proposed for FtsK (, MraY) and FtsW

Perspectives

■ Build full models for FtsX and FtsK

■ Continue the MD study of FtsW

Stability of the FtsW model structures

Refine models with modeller or by MD

Test the interaction with the Lipid II

Test the interaction of FtsW with other TMs from the divisome
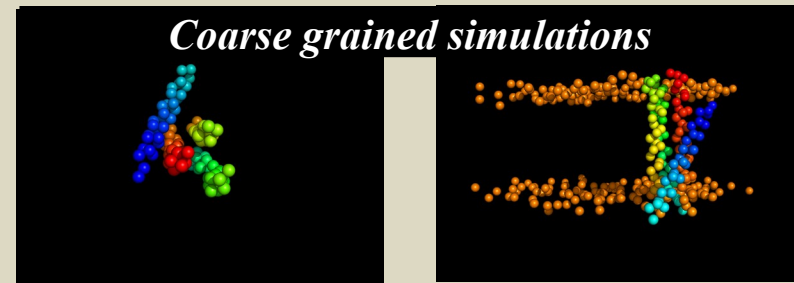
BlaR

- BlaR from Bacilus Licheniformis is a protein of X AA

- It has 4 TMs which have been validate experimentally by Hardt 1997
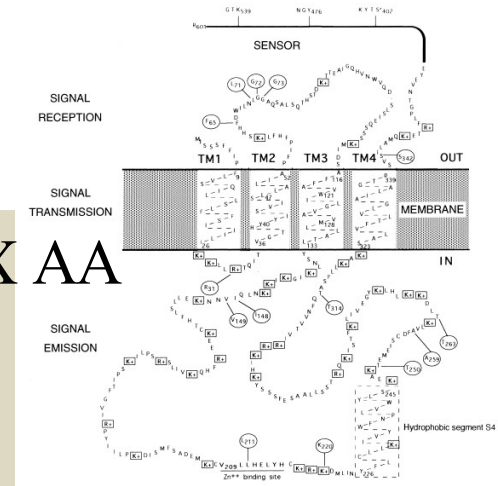
- Rosetta predictions :

  8604 structures generated and

  the score of the best model is -125
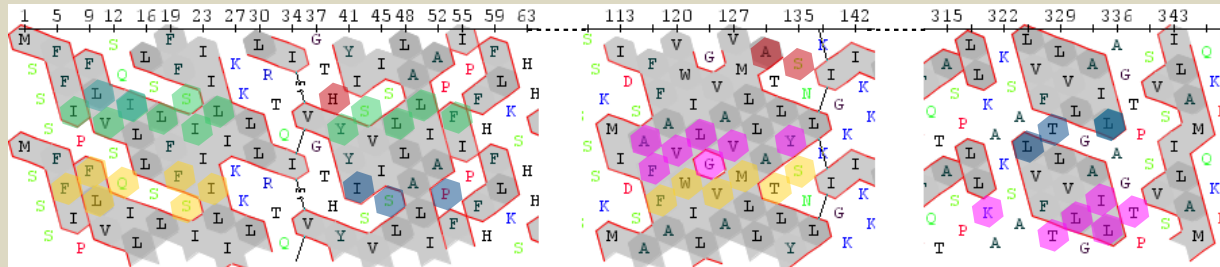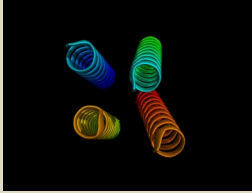

Coarse grained simulations

R1 R2 R3

Score vs RMSD

Rosetta prediction
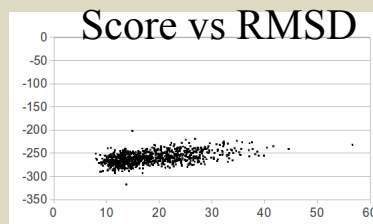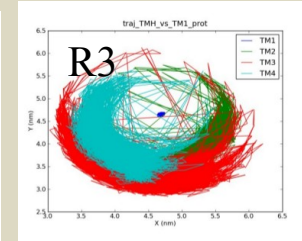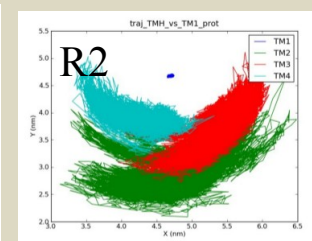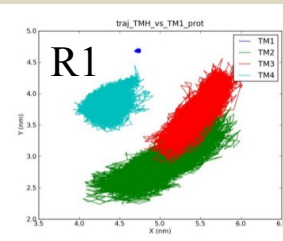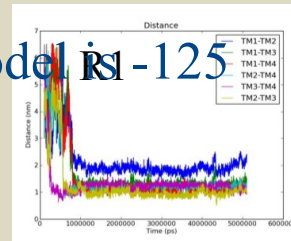
BlaR

- The model 781 has the lowest score (-13080)
- There is 4 clusters that represents 71% of the sampled structures
- These are then good models for the BlaR structure.

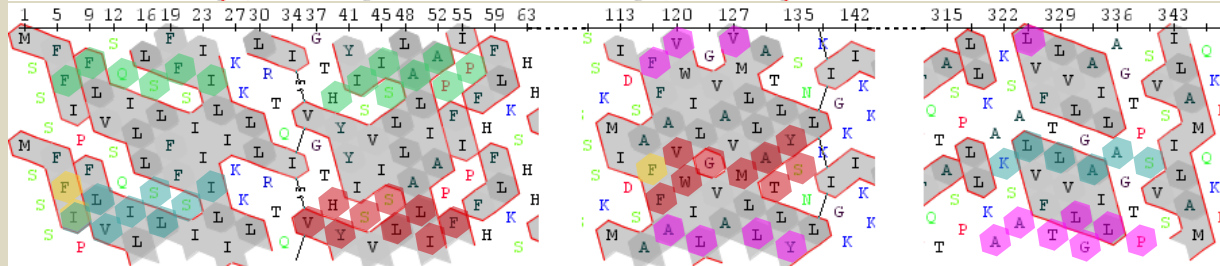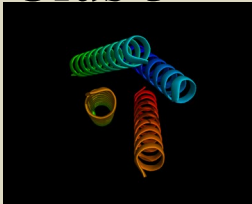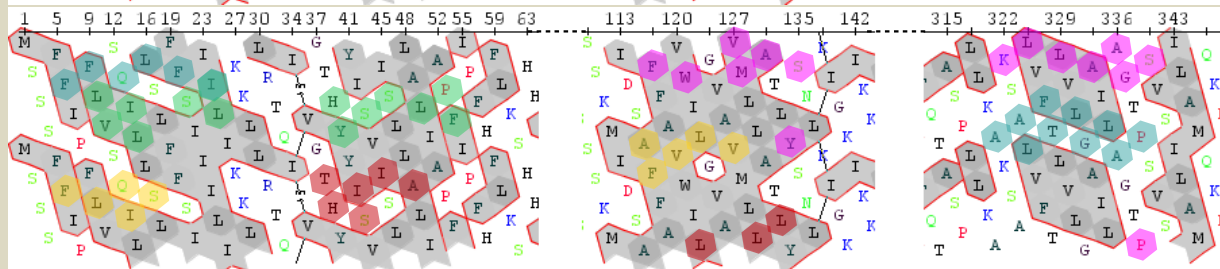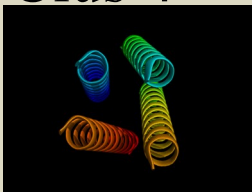| clusters	Model N°	Nb (/1000)	Middle RMSD
Clus 1	142	248	3.31 Å
Clus 2	891	227	3.46 Å
Clus 3	593	125	3.13 Å
Clus 4	248	108	3.31 Å
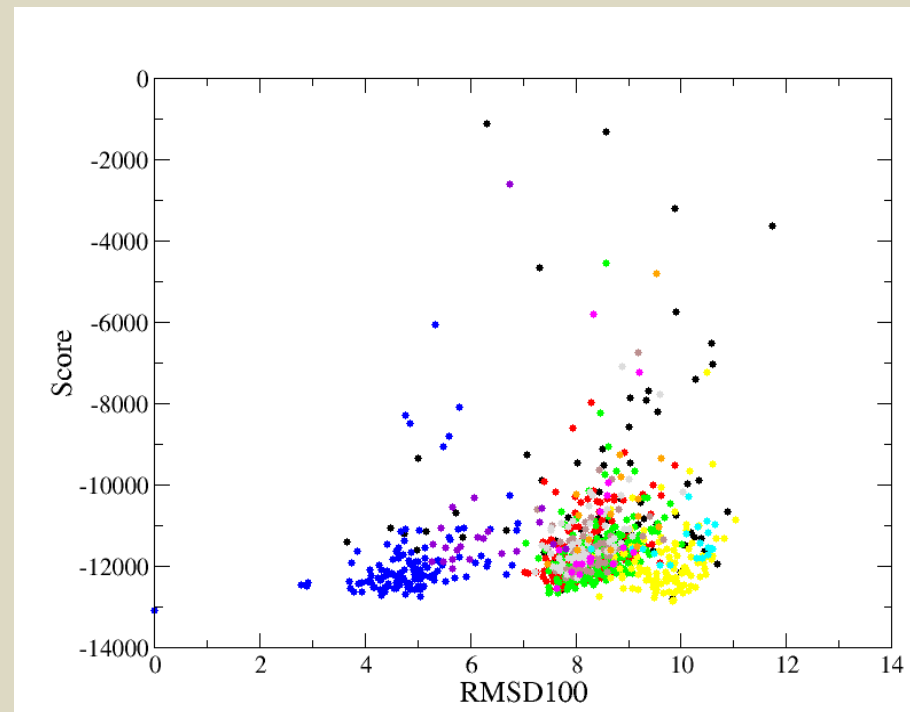
Clus 1

Clus 2
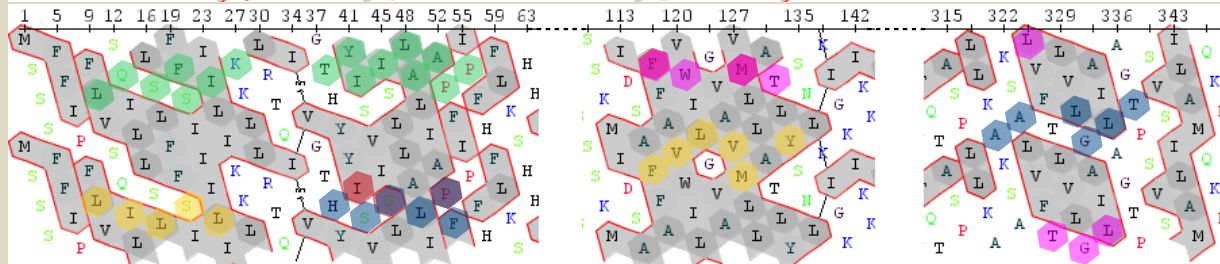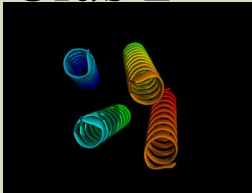
Clus 3

Clus 4

BlaR

TMs are not spanning the membrane