Hunting PP

CNB/CSIC

José R. Valverde
Britt Mellström
José R. Naranjo

FreeBIT/CYTED

Outline

Hunting Protein-Protein interactions in 13 steps (novel DREAM dimerization)

- Step 0: Read bibliography
- Step 1: Search for homologs
- Step 2: Multiple Alignment
- Step 3: Coevolution Analysis
- Step 4: Predict interating proteins
- Step 5: Predict interacting a. a.
- Step 6: Homology modeling
- Step 7: *ab initio* modeling
- Step 8: Build ionic models
- Step 9: Molecular refinement
- Step 10: Protein-Protein docking
- Step 11: Dimer analysis
- Step 12: Dimer refinement
- Step 13: Region analysis

0: Read Bibliography

- Should be the initial step in all cases
- Should have been already done
- Likely to be neglected
  - It is funnier to play from the start
- Guides all subsequent analysis and experiment
- Allows taking a decision
  - Is it worth the trouble?

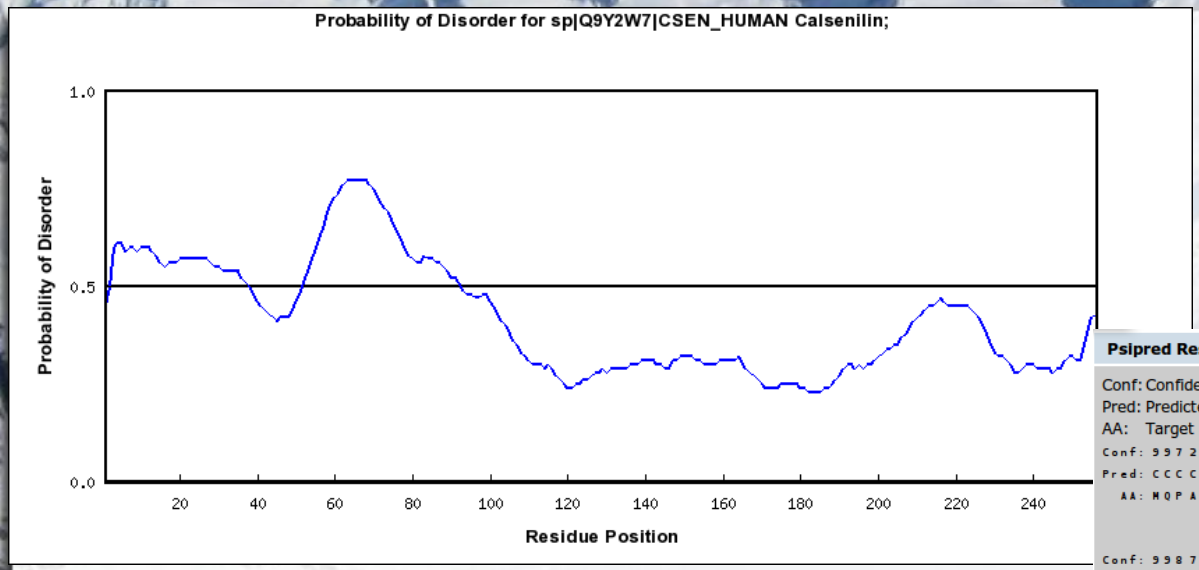
Step 0: What was known

- From bibliography, Calsenilin:
  - Has four EF hands, two of them bind Ca^{++}
  - Full-length protein is 100% insoluble (expose hydrophobic domain)
  - Ca^{++} dimerizes the protein ($L_{155}+L_{159} \leftrightarrow L_{251}$) and increases 7% alpha helix (likely covering hydrophobic domain)
  - Co-expressed with PSEN2 moves to membrane, alone to cytosol
  - N-term unstable random coil (NMR), cleaved by Caspase3 (DXXD)
  - L_{155}XXLL involved in vitamin D interaction (surface)
  - Mg^{++} binds EF-II and is involved in DNA interaction
  - Solvent exposed groove with F_{100}, F_{114}, I_{117}, Y_{118}, F_{121}, F_{122}, Y_{151} and L_{155} implicated in target DNA recognition

Step 0: what wasn't written

- From experiment we learn that
  - $C_{46}C_{47}$ -> $S_{46}S_{47}$ increases DNA binding strength
  - In presence of Mg^{++} W_{50} interacts with Y_{203} in region between EF-III and EF-IV
  - With Mg^{++} it dimerizes in a different way from the known Ca^{++} dependent one (insoluble)
- **Which are the key interacting amino acids in presence of Mg^{++}?**

Step 0: preliminary analysis



Non disordered region around W_{50} likely in alpha-helix conformation.

Psipred Residue Predictions

Conf: Confidence 0 =low 9 =high
Pred: Predicted secondary structure H =helix E =strand C =coil
AA: Target sequence

```
Conf: 997221000368864589999998642111224688750567665555667652037999999
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCCCCCC
  AA: MQPAKEVTKASDGSLLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKWILSSTAPQGS
                10        20        30        40        50        60

Conf: 998711122546899888999999826999999999999999811999998798989999973
Pred: CCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHH
  AA: DSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGFKNECPTGLVDEDTFKLIYAQ
                70        80        90       100       110       120

Conf: 089998226788888642469998311999999999874079998899877431125998 7
Pred: HCCCCCCHHHHHHHHHHHHCCCCCCCEEHHHHHHHHHHHCCCCHHHHHHHHHHCCCCCCCC
  AA: FFPQGDATTYAHFLFNAFDADGNGAIHFEDFVVGLSILLRGTVHEKLKWAFNLYDINKDG
                130       140       150       160       170       180

Conf: 4159999999999999860578899999566799999999998369999986069999999997
Pred: CCCHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHH
  AA: YITKEEMLAIMKSIYDMMGRHTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQ
                190       200       210       220       230       240

Conf: 19889860324547899
Pred: CCHHHHHCCCCCCCCCC
  AA: KDENIMSSMQLFENVI
                250
```

Step 1: Search for homologs

- Protein
  - Human
  - Bovin
  - Rat
  - Mouse
- PDB
  - Human (EF 3-4)
  - Mouse (EF1-4)

- Nucleic Acids

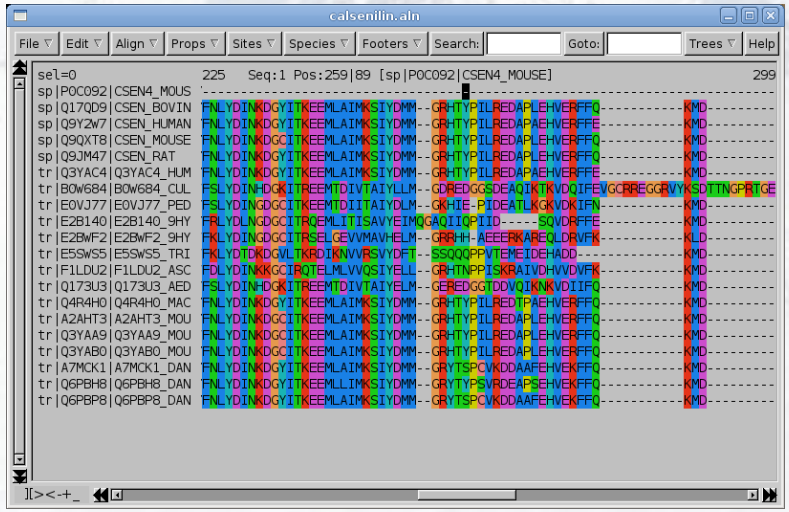
+ Crab-eating macaque
+ Chicken
+ Beetle (Cowpea bruchid)
+ Fish (Gold, Zebra,Sable...)
+ Coral (Acropora millepora)
+ Ascaris suum
+ Trichinella, Schistosoma

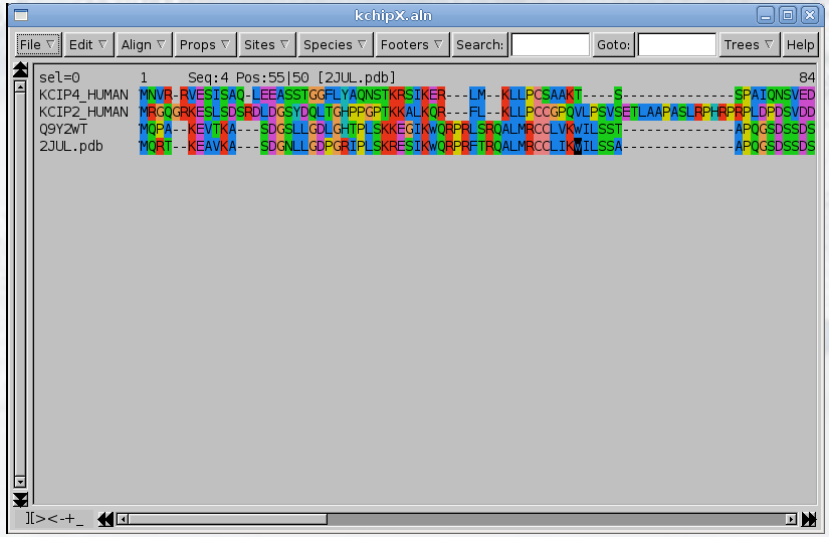
Step 1: same family

- KChIP-1/KCNIP-1: K_v channel-interacting protein 1 (recoverin family)
- KChIP-2/KCNIP-2: K_v channel-interacting protein 2 (recoverin family)
- **KChIP 3/KCNIP-3: Calsenilin, DREAM (neuronal calcium sensor family)**
- KChIP 4/KCNIP-4: K_v channel-interacting protein 4 (recoverin family)

Step 2: Multiple Alignment

- Homologue proteins
  - Too high conservation
- Same family
  - Too little conservation





Step 3: coevolution analysis

- CAPS results (protein family)
  - Unable to find coevolution traces before aa 68 or around Y_{202}
    - N-term function is privative of DREAM
    - Any role is likely due to conformational freedom

Step 4: Predict interacting proteins

- PIPs
  - Lists 7 interactions for DREAM. It is possible that some of them share common amino acids, so they may be investigated by docking or else
- PRISM
  - Lists several other putative interactions, which are worth saving for further analysis

Step 5: Predict interacting a. a.

- All report a patch in the region of interest covering 8-9 consecutive a. a. but cannot resolve most relevant ones
  - ConsPPISP
  - MetaPPISP
  - Polyview
  - PPI-Pred
  - ProMate
  - ConSurf
  - 3d_Partner

K	A	192	0.265	N
S	A	193	0.855	P
I	A	194	0.989	P
Y	A	195	0.884	P
D	A	196	0.980	P
M	A	197	0.994	P
M	A	198	0.991	P
G	A	199	0.977	P
R	A	200	0.996	P
H	A	201	0.984	P
T	A	202	0.993	P
Y	A	203	0.992	P
P	A	204	0.208	N
I	A	205	0.989	P
L	A	206	0.926	P

Step 6: Homology modeling

- In order to proceed we need a 3D structure
  - Homology modeling
    - Likely to fail with the N-term as we know it holds no great similarity with known sequences
  - Threading
    - We may be lucky and find some small patch in the N-term that can be assigned
  - **There are many servers available**
    - **<u>Which one should we use?</u>**

Step 6: Try everything!

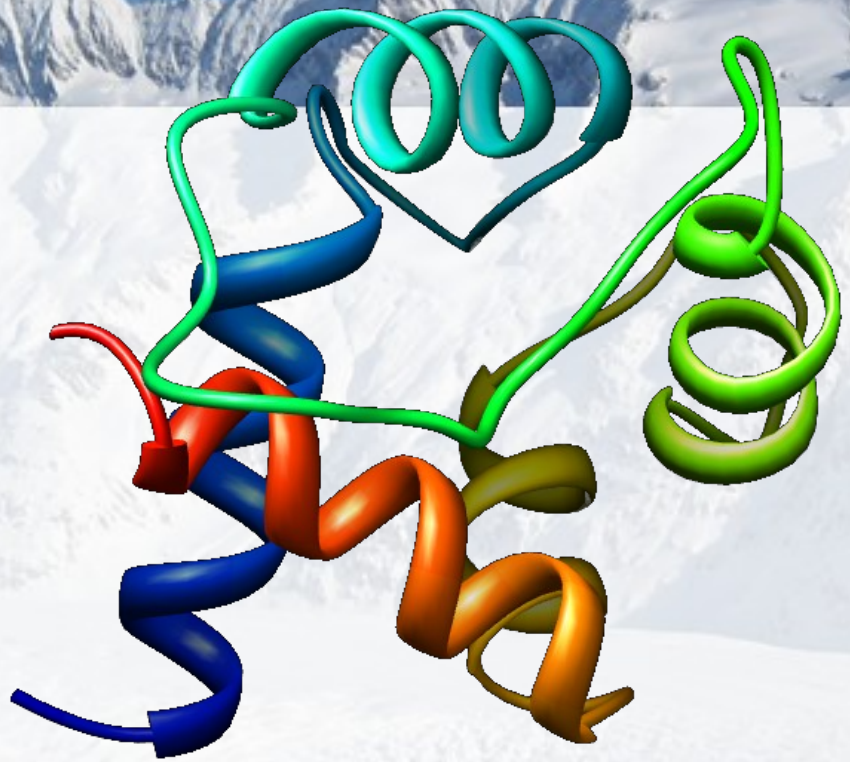
- **Servers**
- CPHmodels
- HHpred
- LOOPP (adds heuristics)
- MUSTER
- Phyre and Phyre2
- $(ps)^2$
- PsiPred
- I-Tasser

- **Metaservers**
- GeneSilico: blastp, compass, ffas, fugue, HHsearch, jmbrank, pcons5, pdbblast, phyre, PRC, sparks
- LoMets: PSIpred, MUSTER, hhsearch, SAM-T02, Sparks-2, SP3, PROSPECT2, PPA-I, FUGUE
- PMP: SwissModel, M4T, ModWeb

27 servers 5-10 models per server

Homology model samples

SwissModel



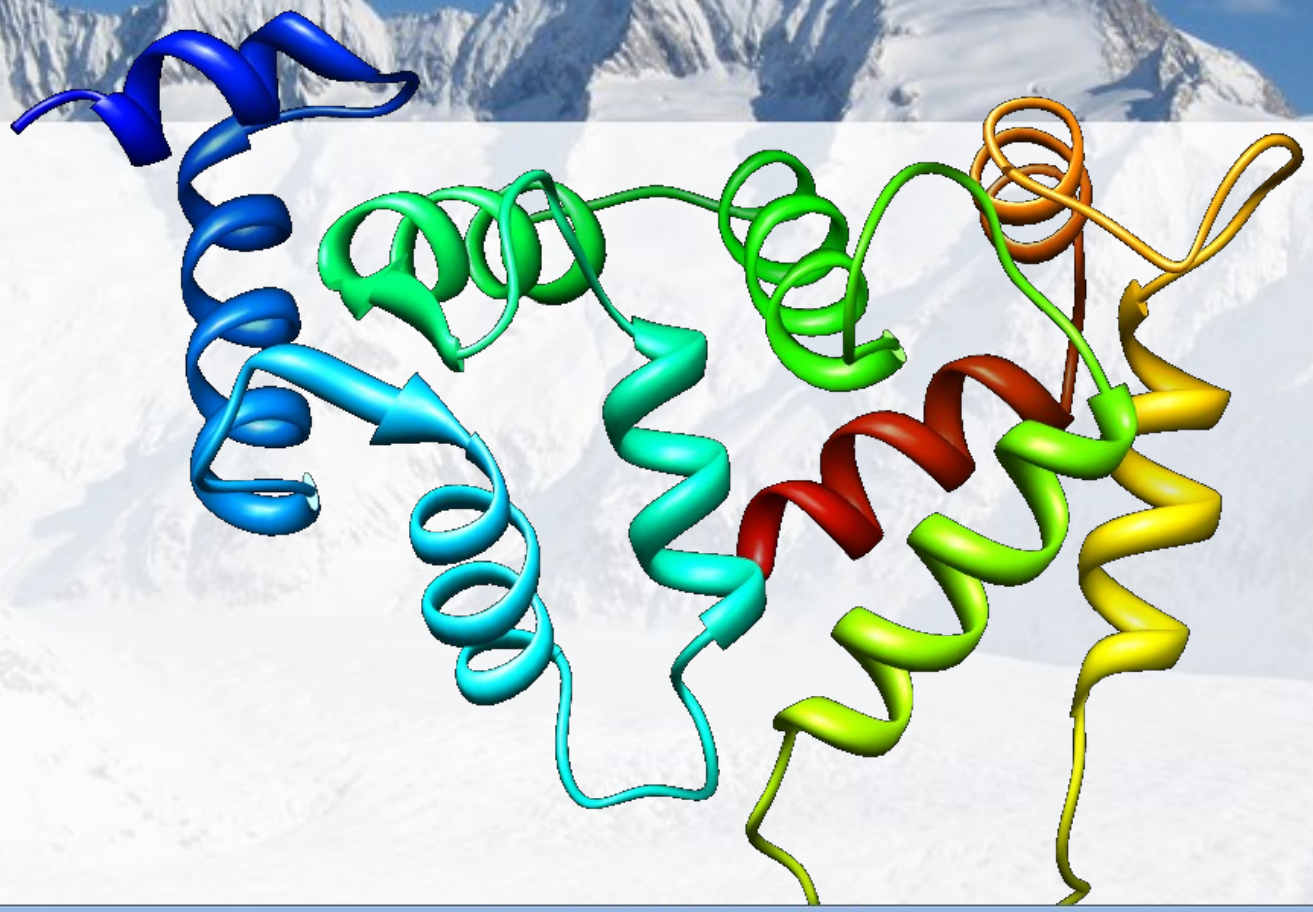
swissmodel.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

swissmodel.pdb...principal chain 161 GTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGRHTYPILREDA

swissmodel.pdb...principal chain 211 PAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQLFENVI

CPHmodels



CPHmodels.pdb (#0) principal chain

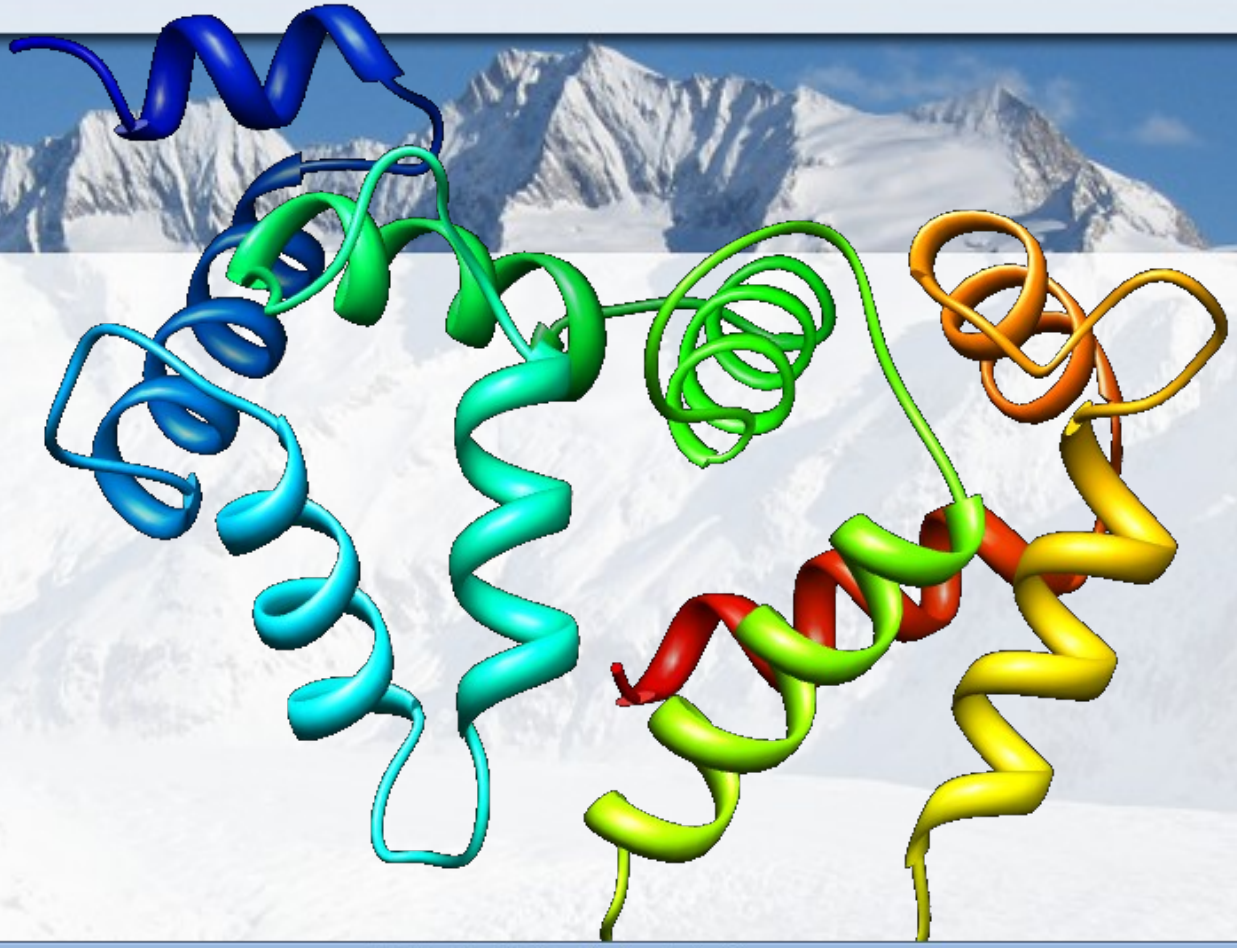
File Edit Structure Headers Numberings Tree Tools Preferences

CPHmodels.pdb ...principal chain 76 PEGLDQLQAQTKFTKKELQSLYRGFKNECPTGLVDEDTFKLIYAQFFPQG

CPHmodels.pdb ...principal chain 126 DATTYAHFLFNAFDADGNGAIHFEDFVVGLSILLRGTVHEKLKWAFNLYD

CPHmodels.pdb ...principal chain 176 INKDGYITKEEMLAIMKSIYDMMGRHTYPILREDAPAEHVERFFEKMDRN

M4T

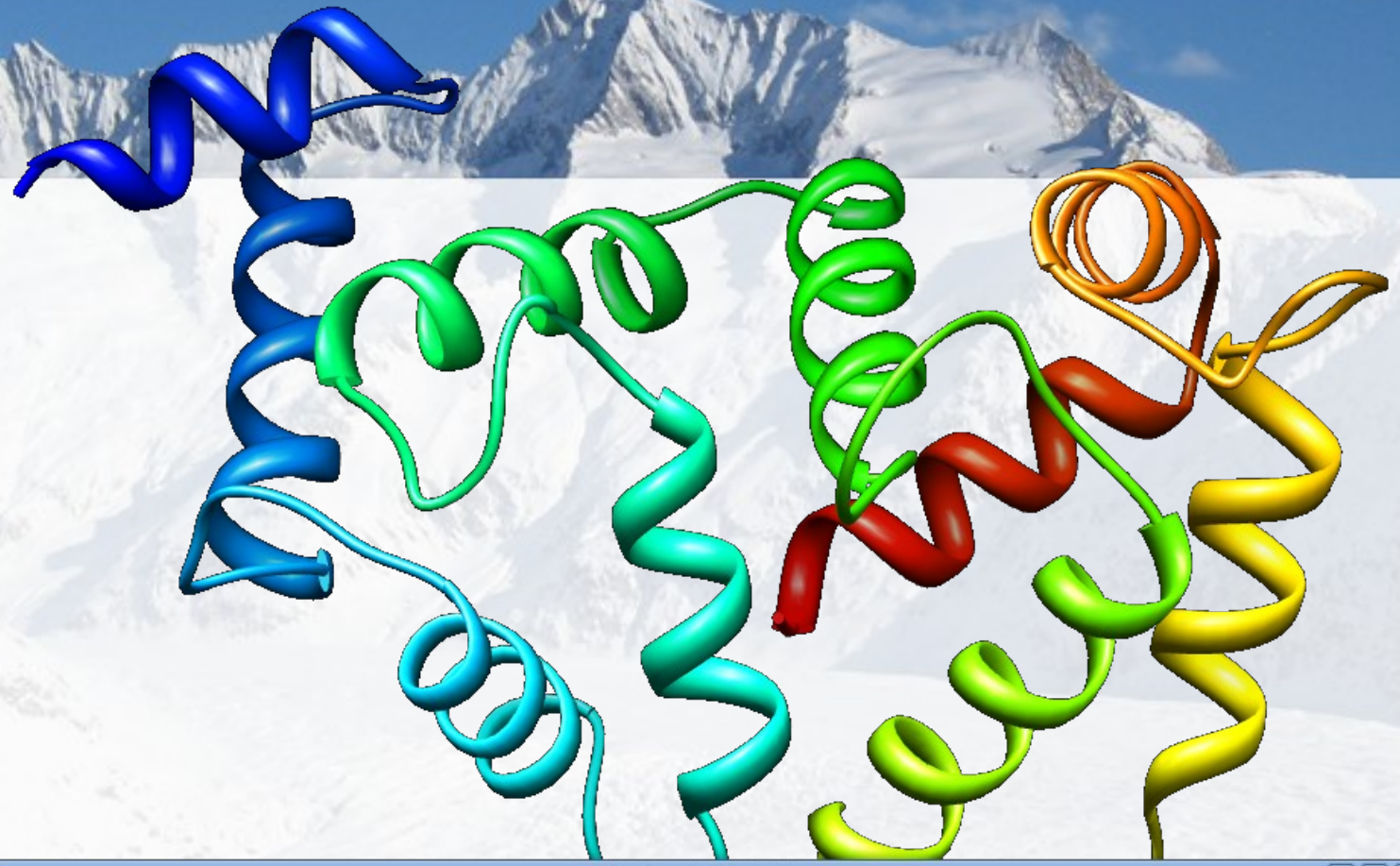


M4T.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

```
M4T.pdb (#0) principal chain   76 PEGLDQLQAQTKFTKKELQSLYRGFKNECPTGLVDEDTFKLIYAQFFPQG
M4T.pdb (#0) principal chain  126 DATTYAHFLFNAFDADGNGAIHFEDFVVGLSILLRGTVHEKLKWAFNLYD
M4T.pdb (#0) principal chain  176 INKDGYITKEEMLAIMKSIYDMMGRHTYPILREDAPAEHVERFFEKMDRN
M4T.pdb (#0) principal chain  226 QDGVVTIEEFLEACQKDENIMSSMQLFENVI
```

$(ps)^2$

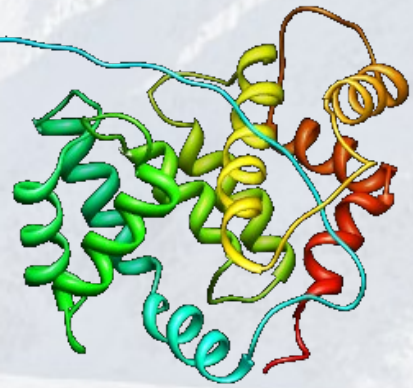


ps2.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

ps2.pdb (#0) principal chain	1	PEGLDQLQAQTKFTKKELQSLYRGFKNECPTGLVDEDTFKLIYAQFFPQG
ps2.pdb (#0) principal chain	51	DATTYAHFLFNAFDADGNGAIHFEDFVVGLSILLRGTVHEKLKWAFNLYD
ps2.pdb (#0) principal chain	101	INKDGYITKEEMLAIMKSIYDMVGRHTYPILREDAPAEHVERFFEKMDRN
ps2.pdb (#0) principal chain	151	QDGVVTIEEFLEACQKDENIMSSMQLFENVI

MUSTER

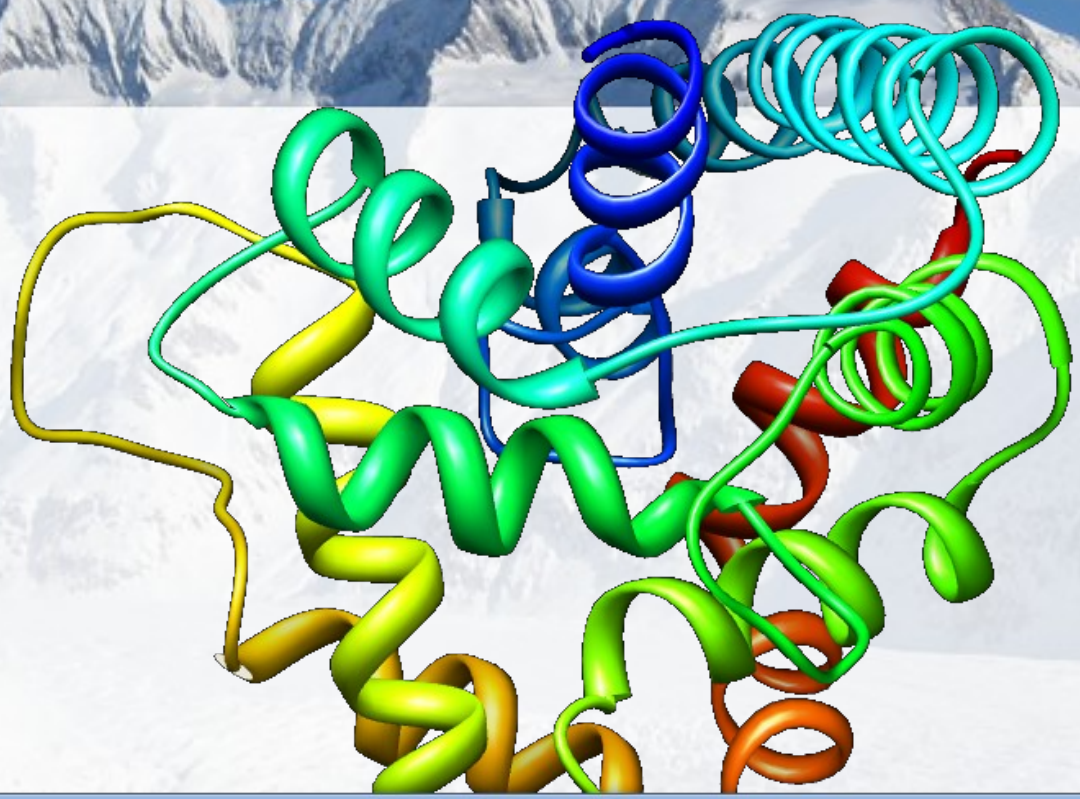


MUSTER.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

```
MUSTER.pdb (#0...principal chain   1 MQPAKEVTKASDGSLLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKW
MUSTER.pdb (#0...principal chain  51 ILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGF
MUSTER.pdb (#0...principal chain 101 KNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAFDADGNGAIHFED
MUSTER.pdb (#0...principal chain 151 FVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGR
MUSTER.pdb (#0...principal chain 201 HTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQ
MUSTER.pdb (#0...principal chain 251 LFENVI
```

LOOPP LP9

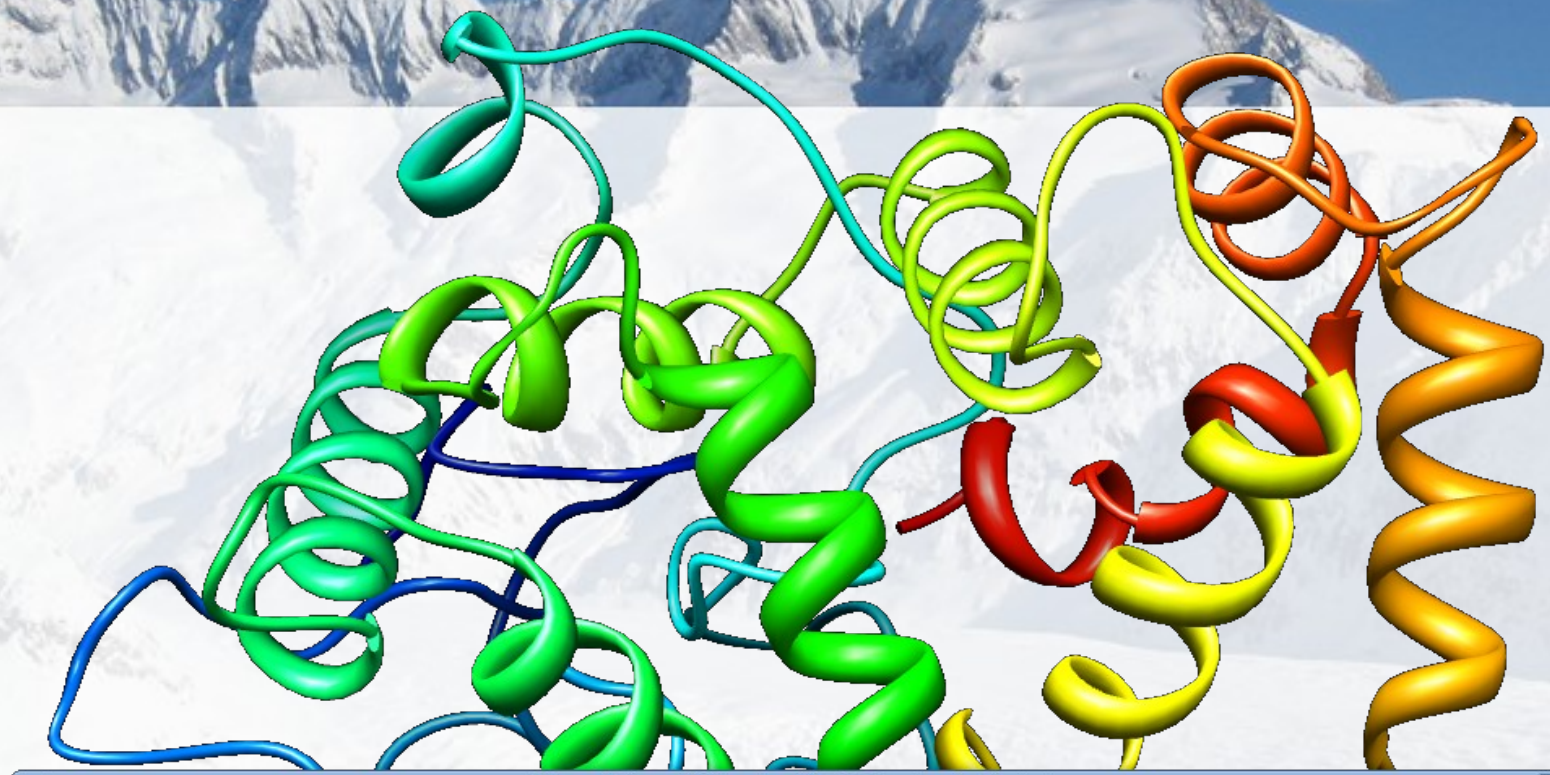


LOOPP_LP9.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

```
LOOPP_LP9.pdb ...principal chain   38 RQALMRCCLVKWILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKF
LOOPP_LP9.pdb ...principal chain   89 TKKELQSLYRGFKNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAF
LOOPP_LP9.pdb ...principal chain  139 DADGNGAIHFEDFVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEML
LOOPP_LP9.pdb ...principal chain  189 AIMKSIYDMMGRHTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEA
LOOPP_LP9.pdb ...principal chain  239 CQKDENIMSSMQLFENVI
```

HHpred (handpicked templates)



HHpred_handpicked.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

HHpred_handpic...principal chain	1	MQPAKEVTKASDGSLLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKW
HHpred_handpic...principal chain	51	ILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGF
HHpred_handpic...principal chain	101	KNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAFDADGNGAIHFED
HHpred_handpic...principal chain	151	FVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGR
HHpred_handpic...principal chain	201	HTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQ
HHpred_handpic...principal chain	251	LFENVI

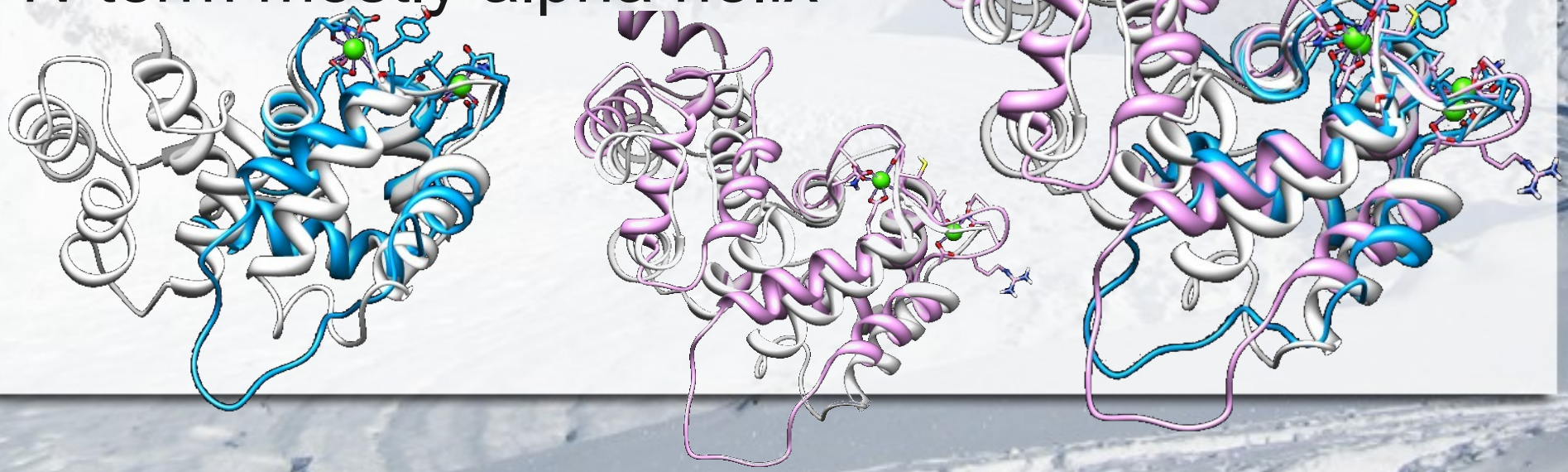
Step 7: *ab initio* modeling

- Phyre2 (fast/slow)
- QUARK (I-Tasser)
- Rosetta (Robetta)

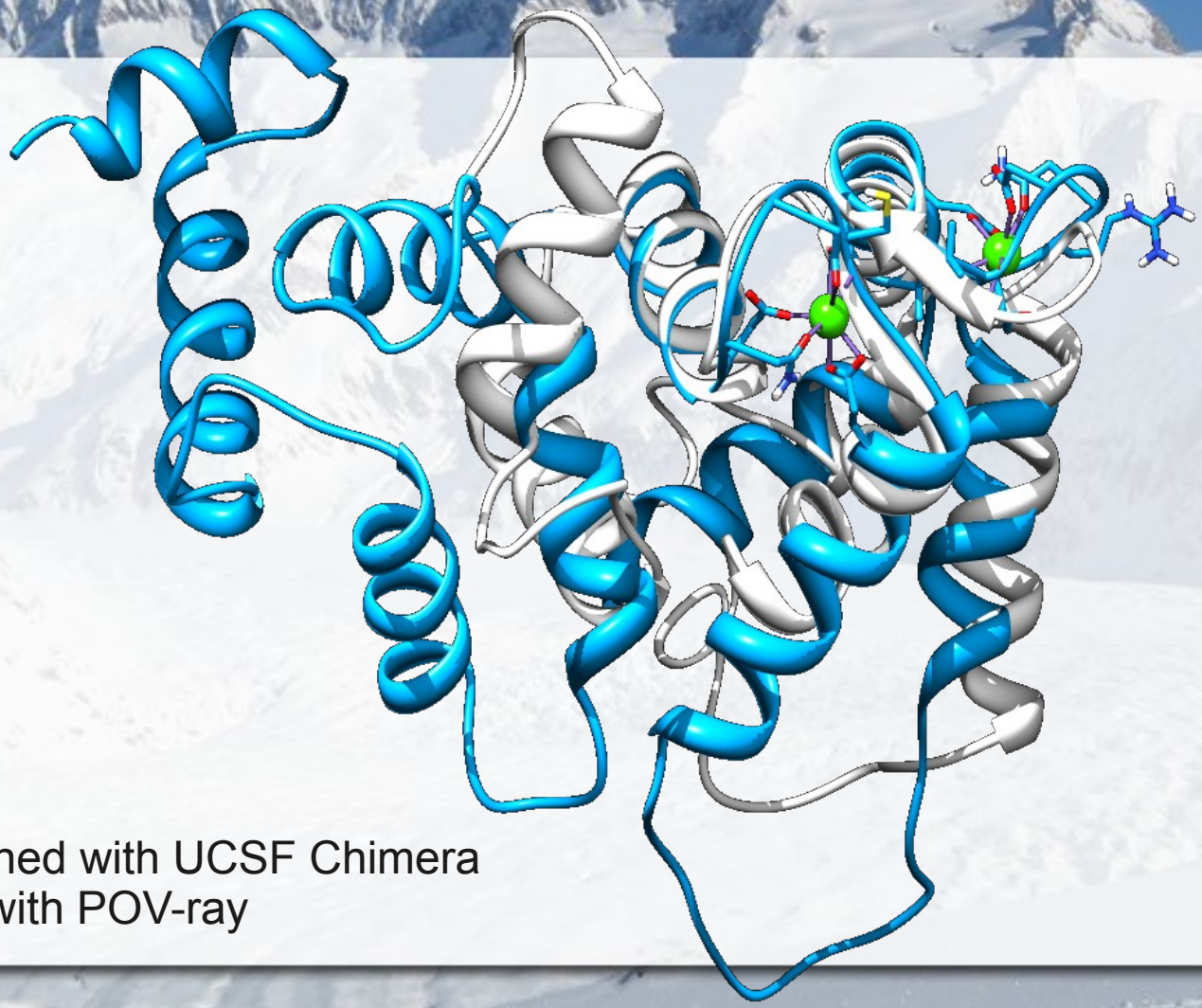
- Works well for “short” sequences (up to 150-200 aa)
- Combine with homology modeling

Step 7: *ab initio* validation

- Build models for N-term and C-term
- Compare C-term against known structure
  - Good prediction quality
- N-term mostly alpha helix

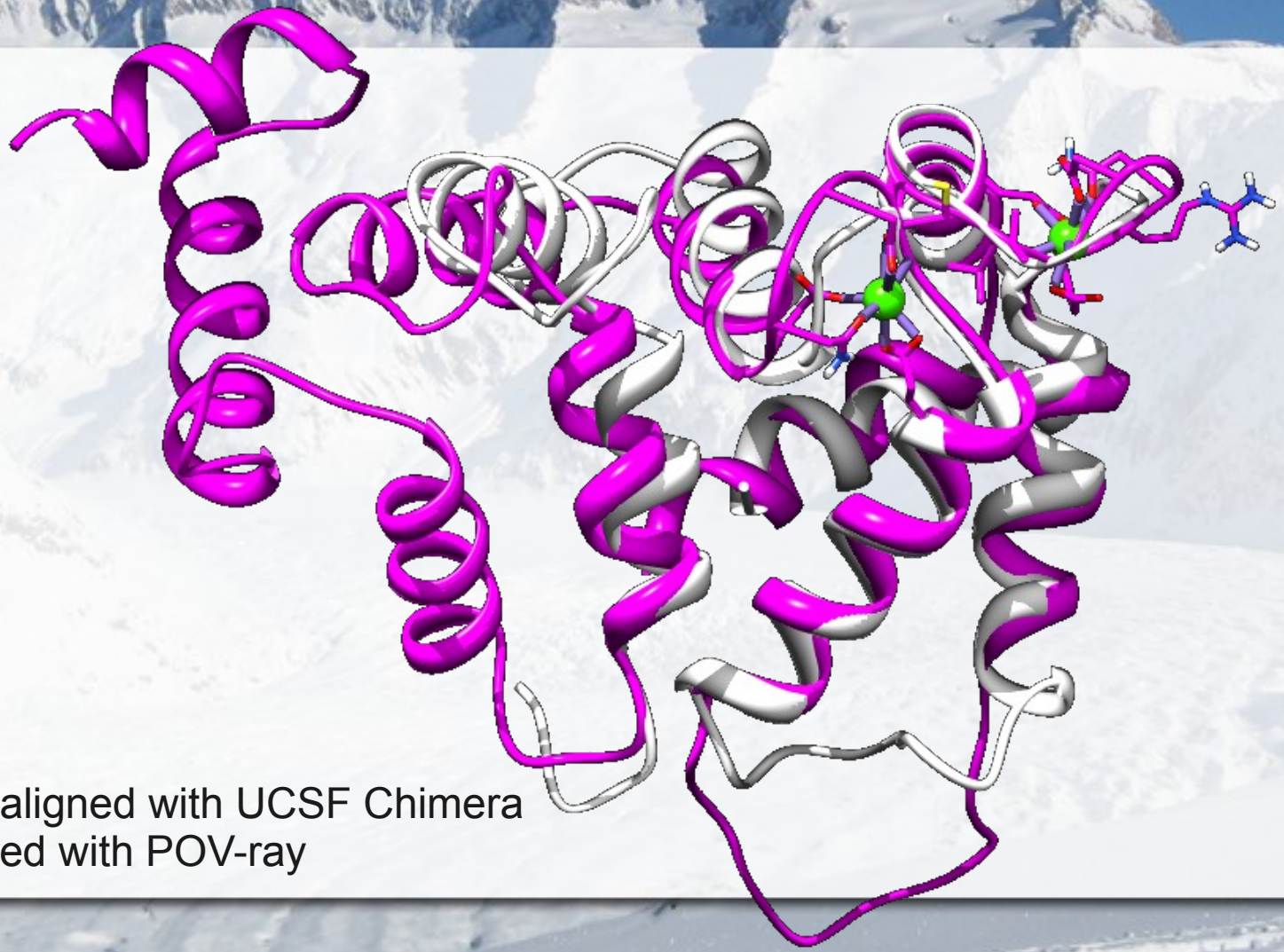


Quark *ab initio* 1 vs 2jul-A



Structures aligned with UCSF Chimera
and rendered with POV-ray

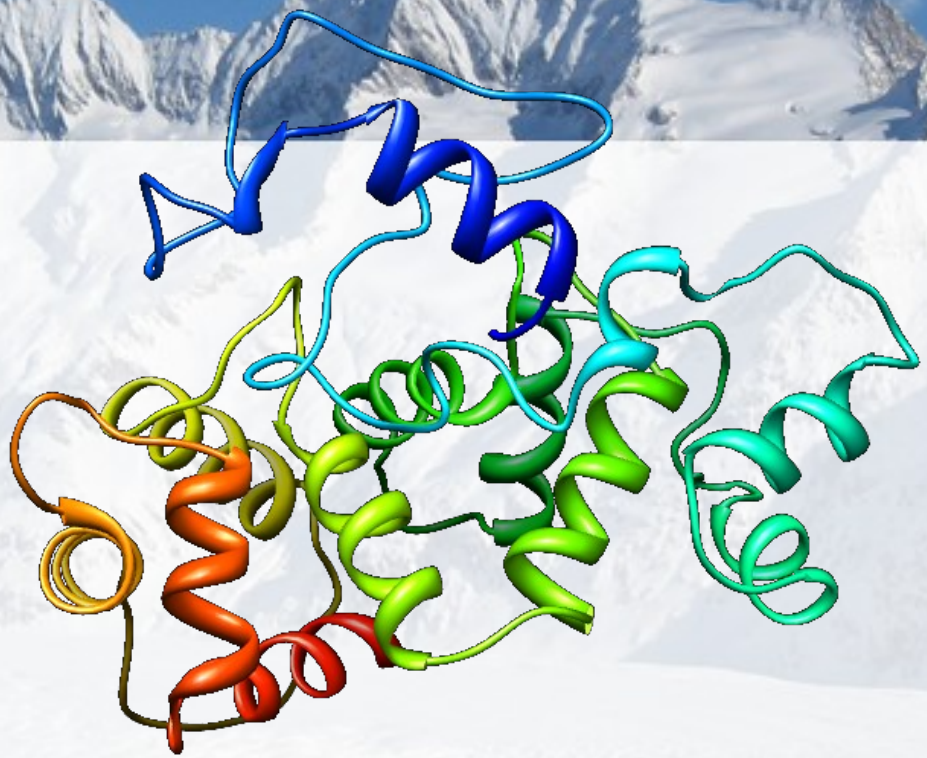
Rosetta *ab initio* 1 vs 2jul-A



Structures aligned with UCSF Chimera
and rendered with POV-ray

Ab initio + homology model samples

Phyre2 (slow)

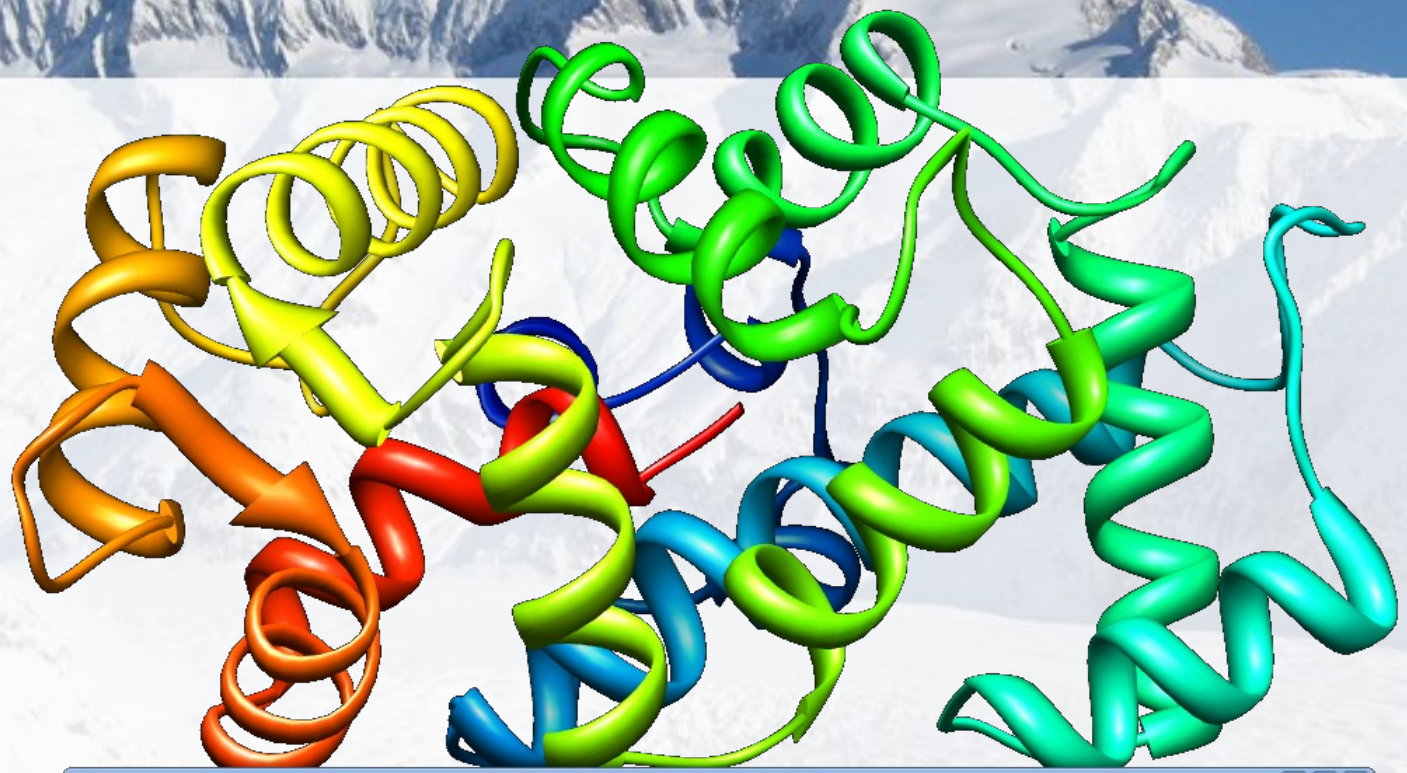


Phyre2slow.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

Phyre2slow.pdb...principal chain	1	MQPAKEVTKASDGSLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKW
Phyre2slow.pdb...principal chain	51	ILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGF
Phyre2slow.pdb...principal chain	101	KNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAFDADGNGAIHFED
Phyre2slow.pdb...principal chain	151	FVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGR
Phyre2slow.pdb...principal chain	201	HTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQ
Phyre2slow.pdb...principal chain	251	LFENVI

i-tasser (quark)

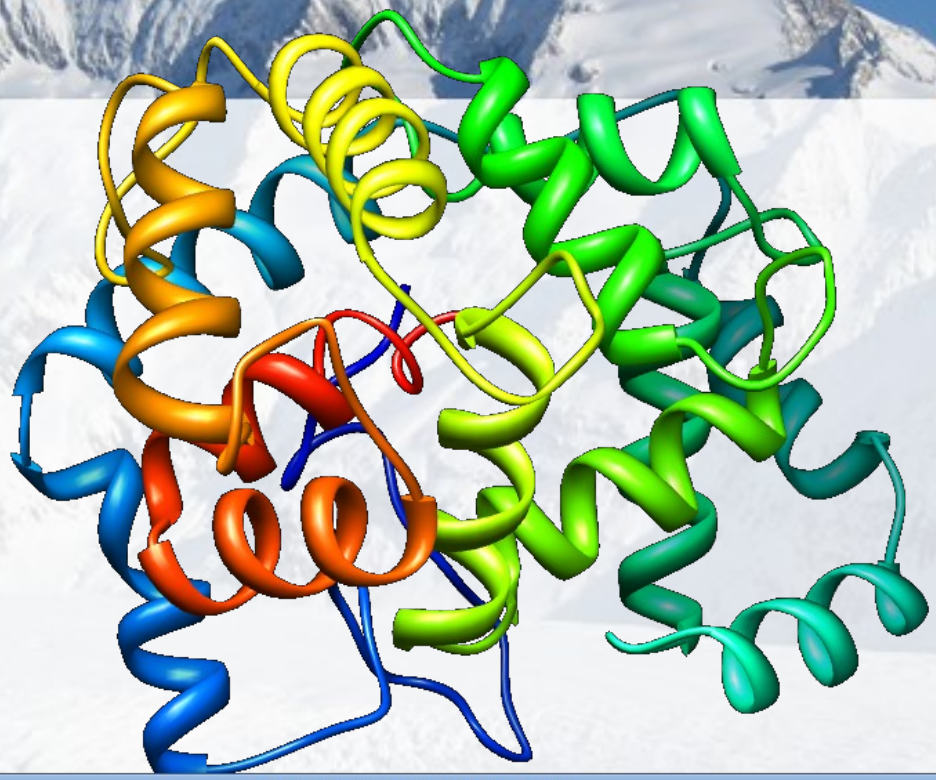


i-tasser.pdb (#0) chain A

File Edit Structure Headers Numberings Tree Tools Preferences

```
i-tasser.pdb (#0) chain A   1 MQPAKEVTKASDGSLLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKW
i-tasser.pdb (#0) chain A  51 ILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGF
i-tasser.pdb (#0) chain A 101 KNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAFDADGNGAIHFED
i-tasser.pdb (#0) chain A 151 FVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGR
i-tasser.pdb (#0) chain A 201 HTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQ
i-tasser.pdb (#0) chain A 251 LFENVI
```

Robetta (rosetta)



robetta.pdb (#0) principal chain

File Edit Structure Headers Numberings Tree Tools Preferences

robetta.pdb (#...principal chain	1	MQPAKEVTKASDGSLLGDLGHTPLSKKEGIKWQRPRLSRQALMRCCLVKW
robetta.pdb (#...principal chain	51	ILSSTAPQGSDSSDSELELSTVRHQPEGLDQLQAQTKFTKKELQSLYRGF
robetta.pdb (#...principal chain	101	KNECPTGLVDEDTFKLIYAQFFPQGDATTYAHFLFNAFDADGNGAIHFED
robetta.pdb (#...principal chain	151	FVVGLSILLRGTVHEKLKWAFNLYDINKDGYITKEEMLAIMKSIYDMMGR
robetta.pdb (#...principal chain	201	HTYPILREDAPAEHVERFFEKMDRNQDGVVTIEEFLEACQKDENIMSSMQ
robetta.pdb (#...principal chain	251	LFENVI

Step 7: Model selection

- ~200-400 models generated
- As expected neither homology nor threading models include a structured N-term with W_{50}
- Homology + *ab initio* produce complete models but no N-term common structure (barring W_{50} alpha helix)
  - HHpred (auto and hand-selected templates): **1**
  - I-Tasser (also predicts Ca^{++} binding sites): **5**
  - Loopp (heuristics L9 +/- redconf): **1**
  - Phyre-2 (slow and fast): **2**
  - Robetta: **5**

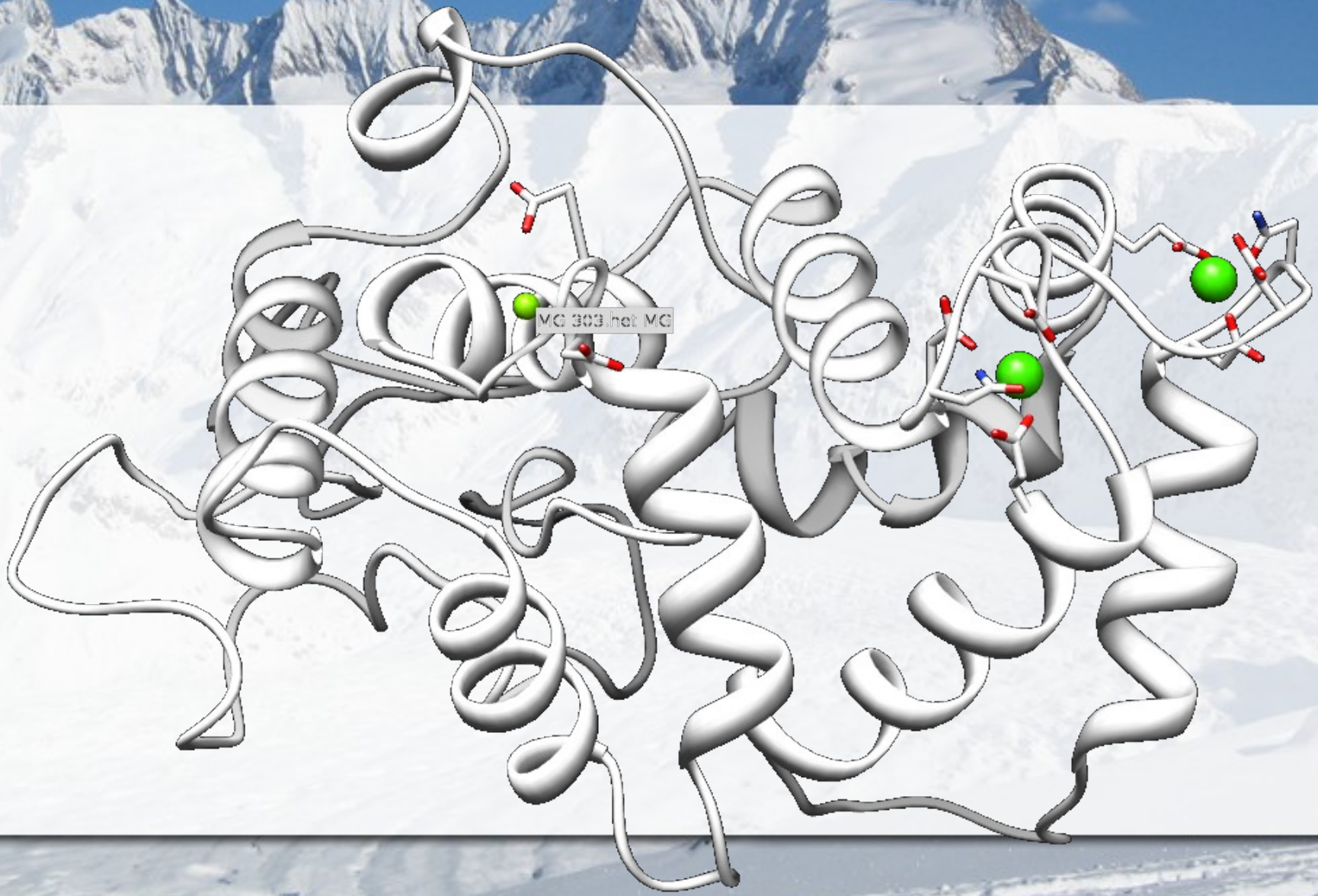
Step 8: build ionic models

- Use 2JUL (mouse) to add Ca^{++} ions to models in EF-hand 3 and 4
- Add Mg^{++} to EF-Hand 2

- No ions
- 2 Ca^{++}
- 1 Mg^{++} 2 Ca^{++}
- 1 Mg^{++}

14 models x 4 conformations

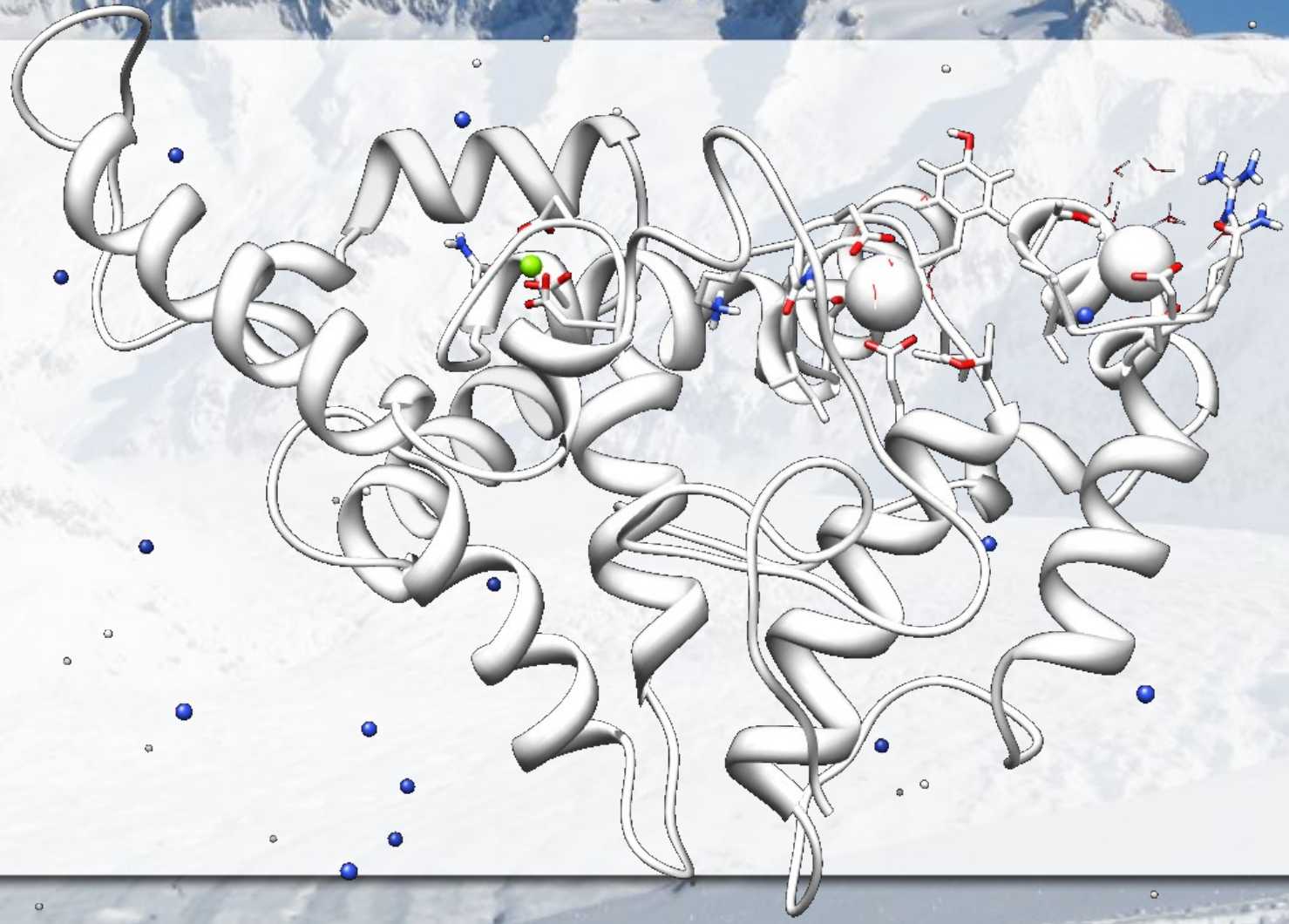
HHpred (handpicked) 1 Mg^{++} 2 Ca^{++}



Step 9: Molecular simulation

- Minimize/Optimize structure
  - GROMACS + gromos53a6 force field
- Full MD runs *in vacuo* and in intracellular physiologic saline solution
  - SPC/E water
  - *intracellular ionic strength* 215 mM (Mouat & Manchester (1998) *Comp Hematol Int*, **8**:56)
  - Equilibration NVT, NPT (300ºK, 100ps in 2fs steps)
  - Production NPT (300ºK, 10ns in 2fs steps)

Robetta3 1Mg^{++} 2Ca^{++} MD in solution



Step 9: validate model dynamics

- Verify final models agree with known data
  - Check for known differences between conformations
  - Verify final structure agrees with experiment
- Analyze dynamic behaviour
  - Understand flexibility of different regions

Step 10: Docking

- GRAMM
- HEX
- ZDOCK
- PatchDock/FireDock
  - No assumptions / Hint L / Hint W
- SymmDock
  - No assumptions / Hint L / Hint W

14 models x 4 conformations x 5 dockings

Step 11: Dimer refinement

- Some dimers may enhance if relaxed
- Simulate each dimer *in vacuo* and in physiologic intracellular solution (GROMACS with gromos53a6 and spc/e water)
- Analyze interactions (H-bonding, H_2O H-bond networks) using UCSF Chimera for visualization.
- Generate mutants for identified amino acids and review dimerization changes (TRITON + Modeller + Gromacs + docking)

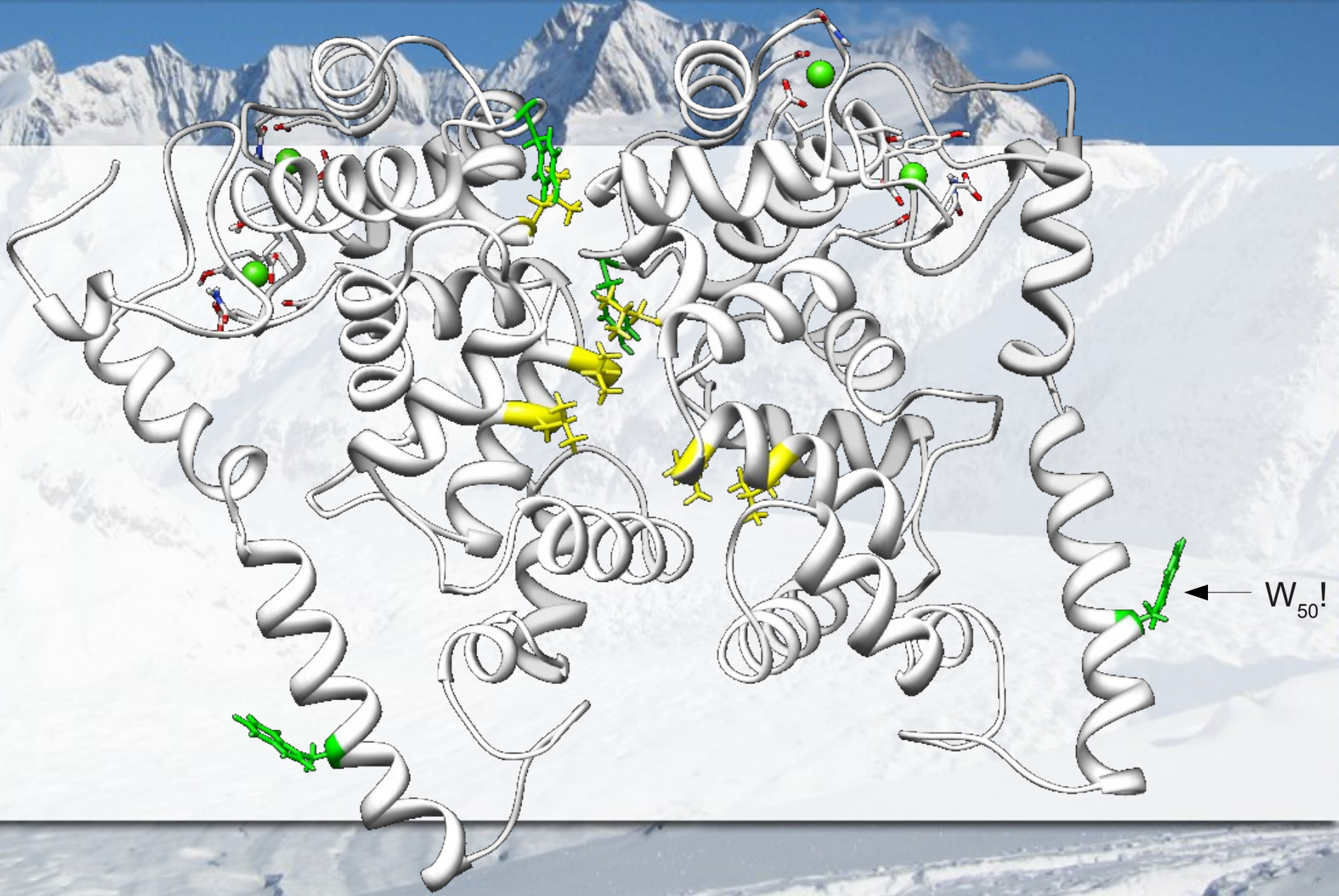
Step 12: Dimer analysis

- 2 Ca^{++}
  - Dimerize around L_{155}, L_{159}, L_{251}
  - itasser3 - phyre2fast,slow - robetta3,4
- 1 Mg^{++} 2 Ca^{++}
  - Dimer involving $W_{50} \leftrightarrow Y_{203}$
  - Hhpred-hand - itasser1- phyre2fast,slow, robetta1,3,4
- Functional site accessibility
  - $L_{155}XXLL$, $D_{61}XXD_{64}$, differential solubility...

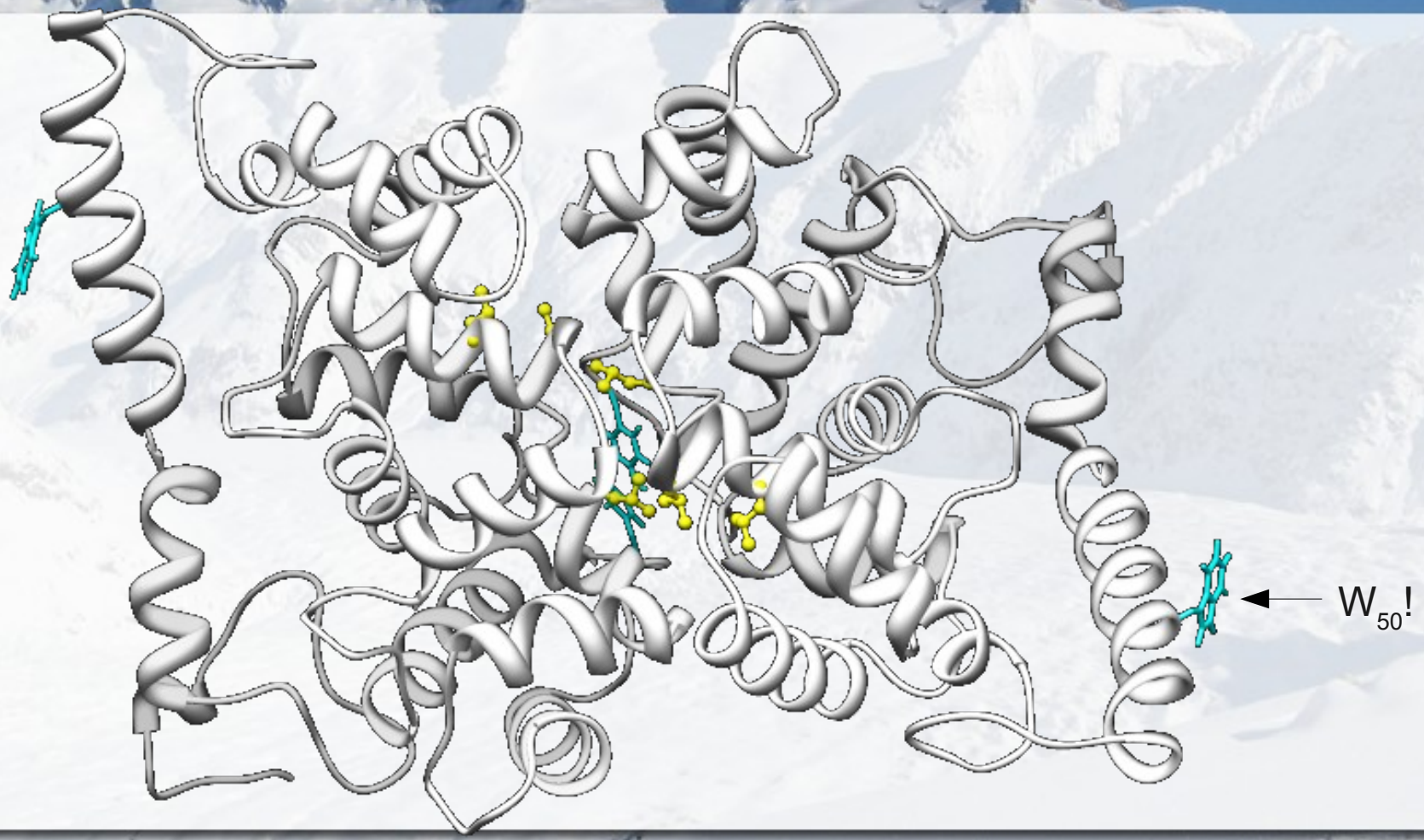
14 models x 4 conformations x 5 dockings x 10-20 dimers

A dimer sampler

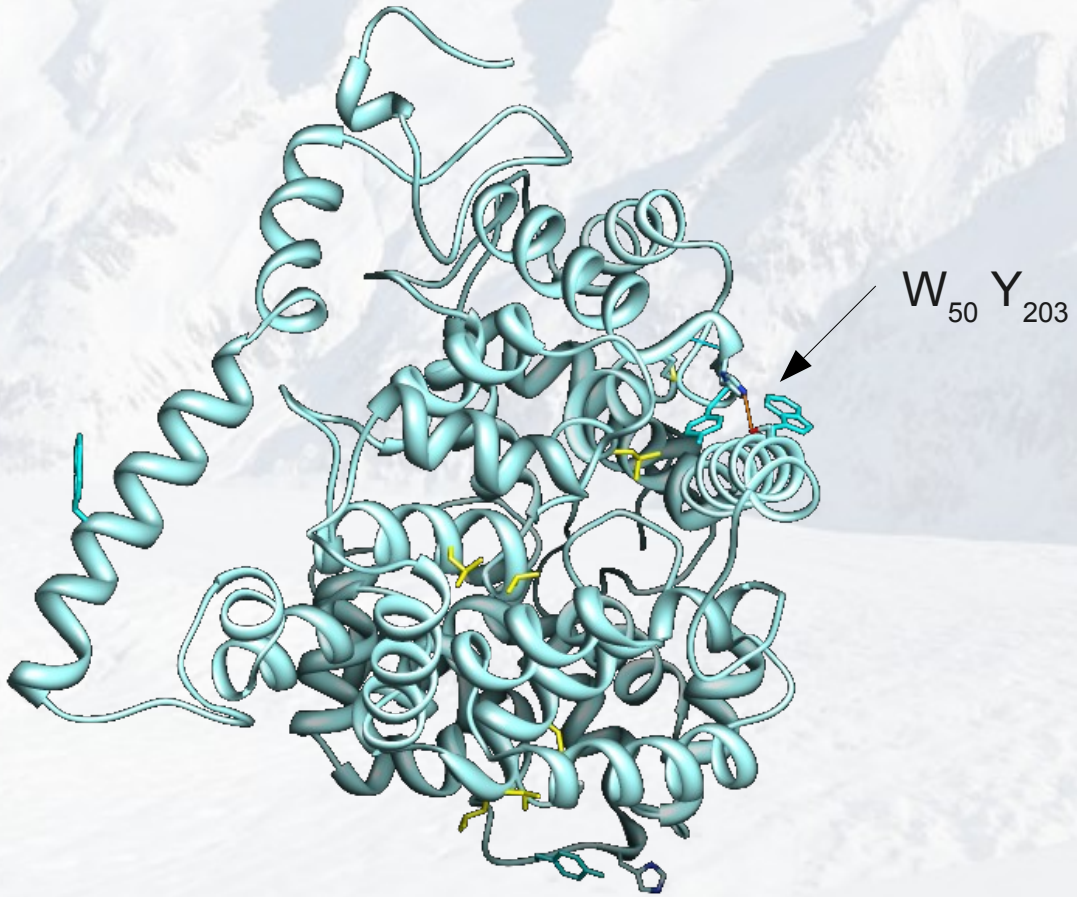
$2Ca^{++}$ with W_{50} favored



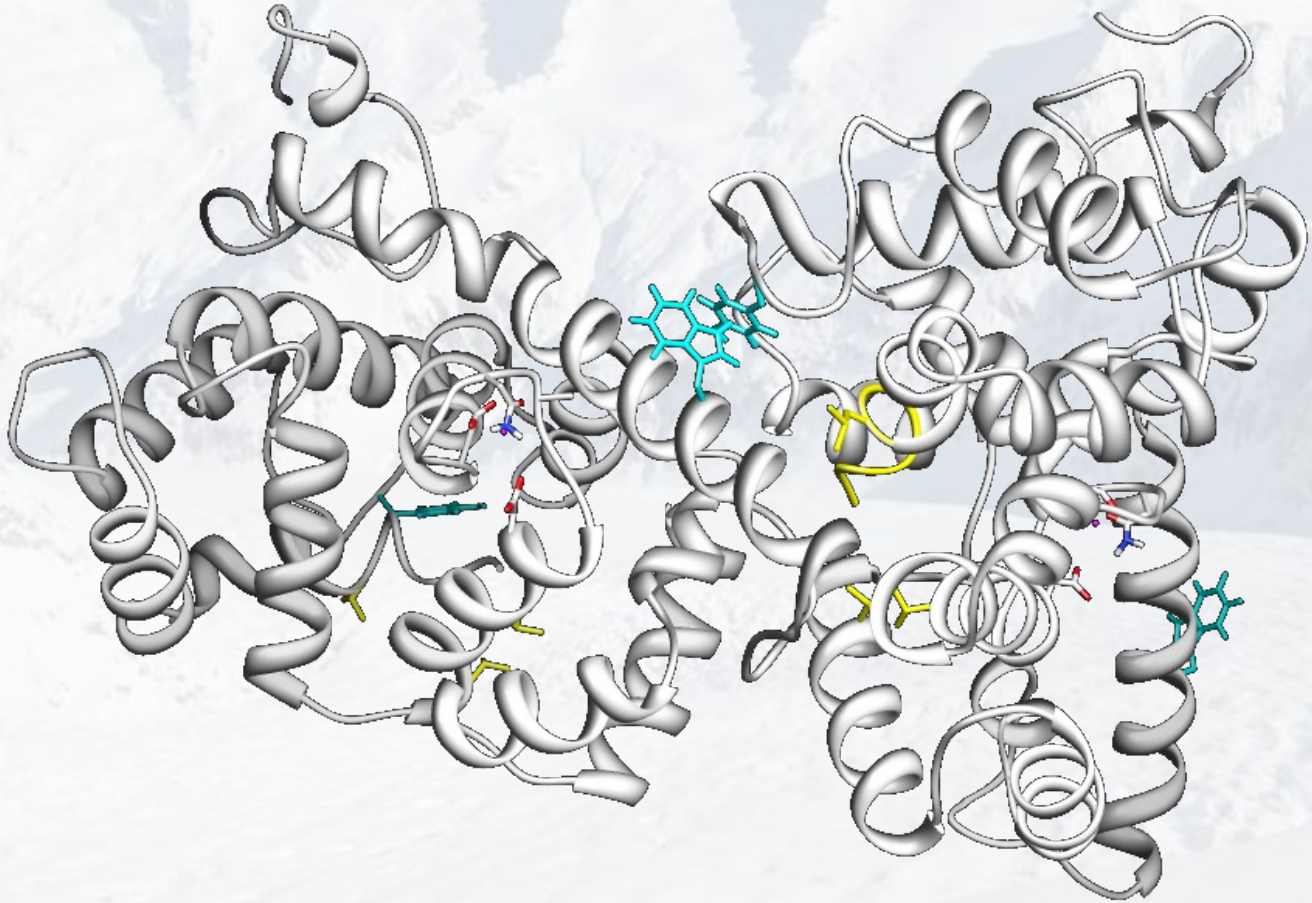
$2Ca^{++}$ with W_{50} favoured (min)



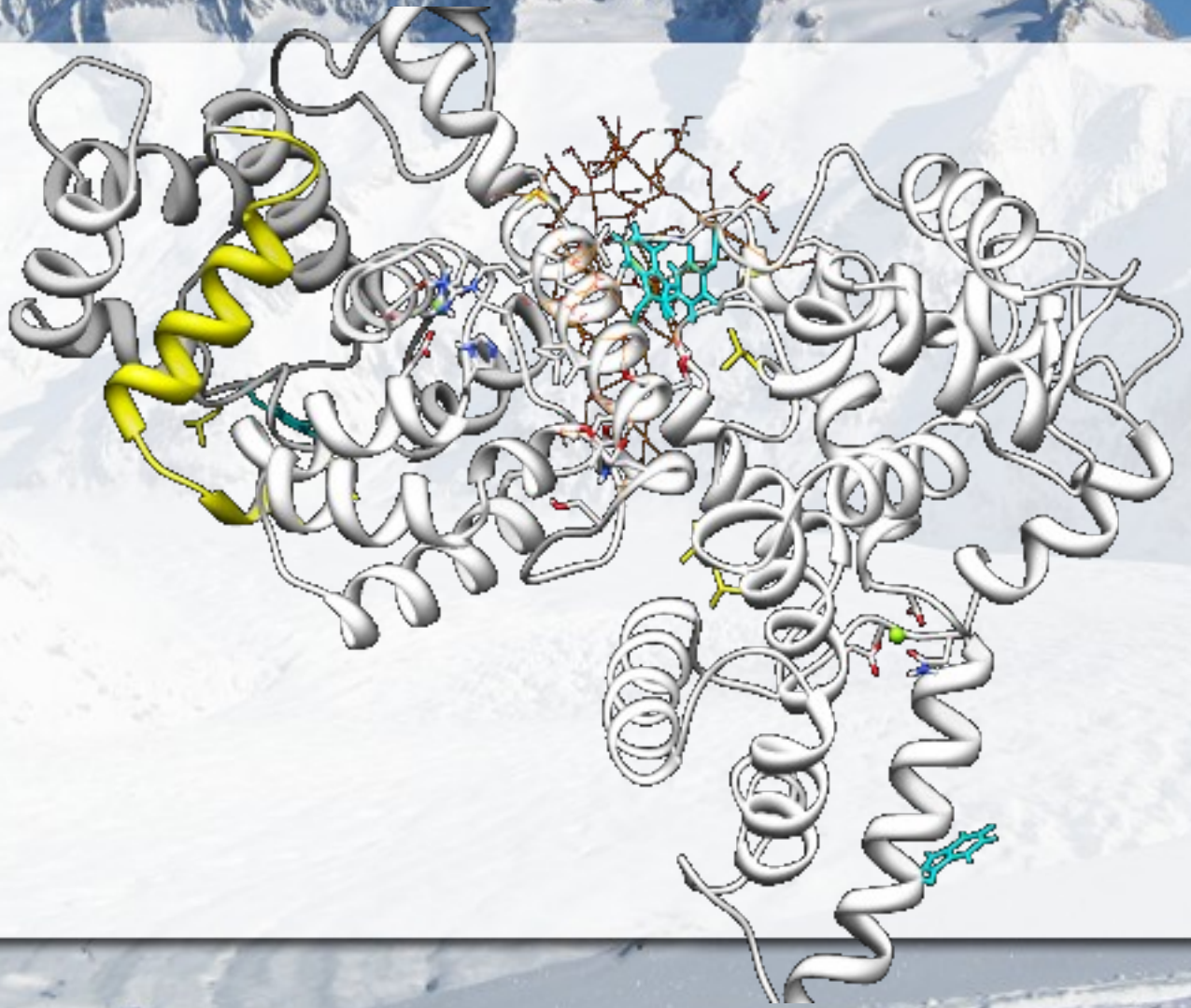
1Mg++ 2Ca++



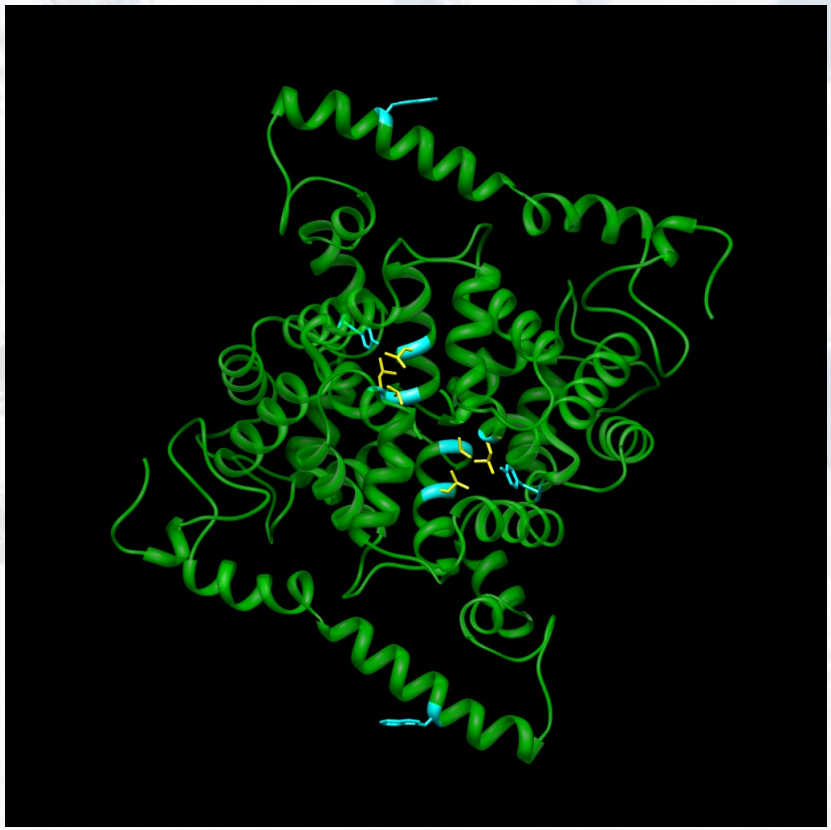
1Mg++ 2Ca++ (min)



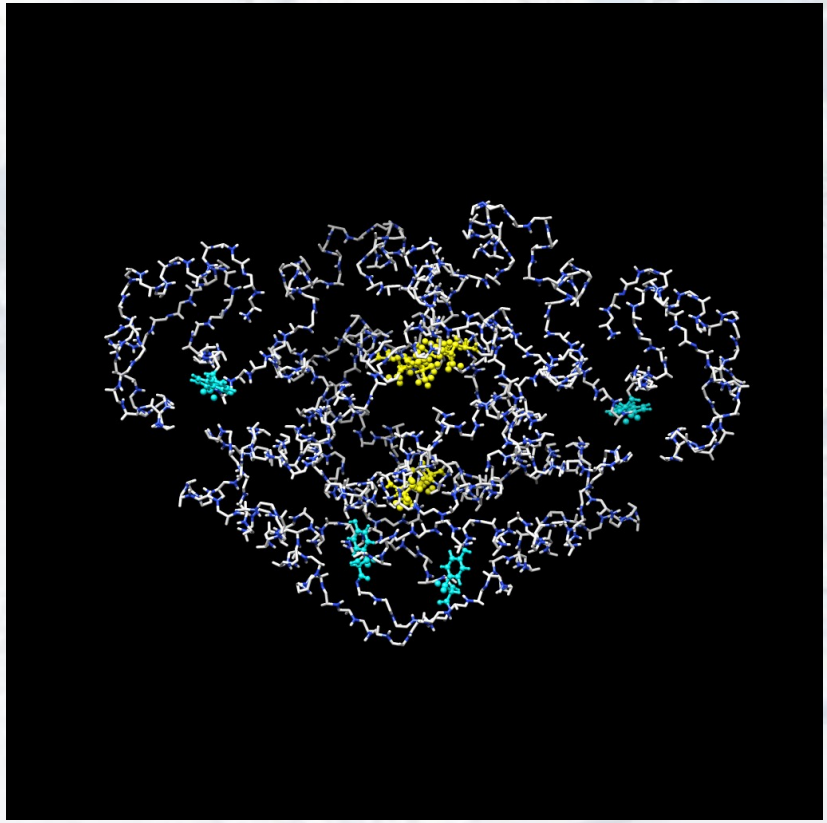
1Mg++ 2Ca++ in solution



Site-directed mutants



$R_{200}A$

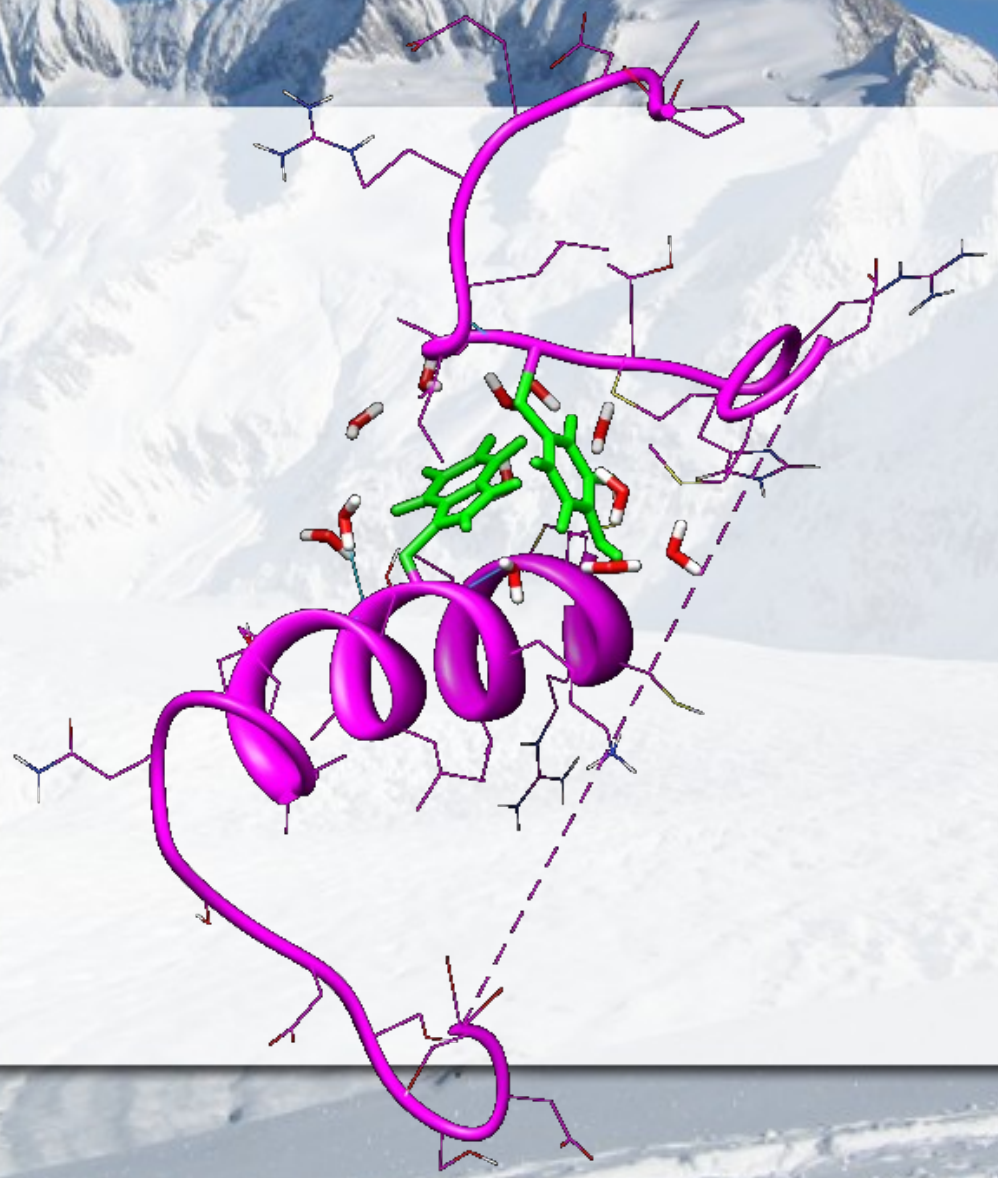


$T_{202}A$

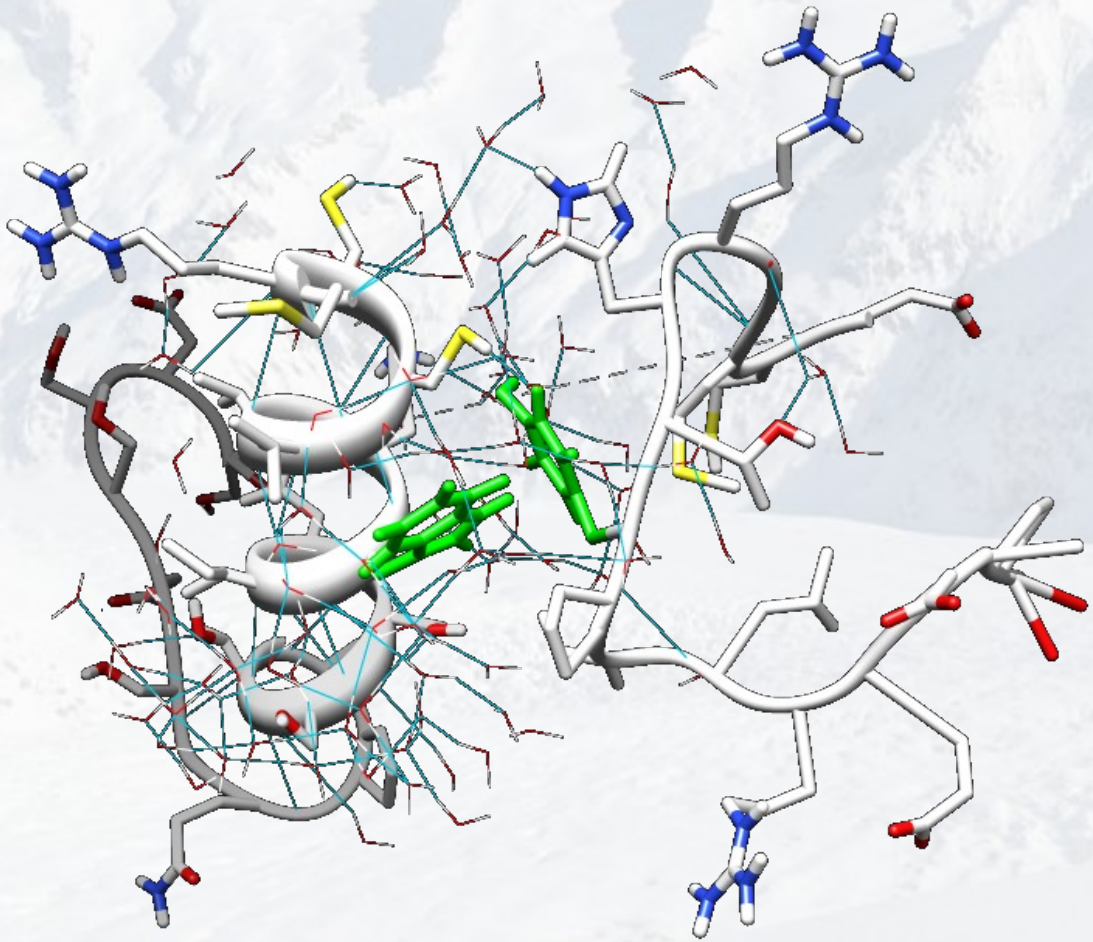
Step 13: Region analysis

- Concentrate on interface region and analyze in detail
  - W_{50} is likely in an alpha helix
  - Y_{203} is in extended/loop conformation
  - Take best generated interfaces and relax using MD
  - Analyze interactions
  - 1SCF/DFT Quantum analysis of electron density distribution
    - Analyze cost/benefit
    - Visualization challenges

Robetta3 PatchDock08 fragment



Interaction network in solution



Conclusion

- **We have an extensive toolset to predict macromolecular interactions**
  - Interaction prediction may be as easy as reading the bibliography or doing an MSA
  - Difficulty increases with lack of knowledge
  - Yet we can proceed a long way forward
- **Predictions need experimental validation**
  - Site-directed mutagenesis is being carried out at J. R. Naranjo's lab by B. Mellström

Short description

"Given the low sequence conservation of the N-terminal region, we generated *in silico* models of native and mutant DREAM 3D structures, predicted likely dimer conformations and analyzed in detail the candidate interactions of W_{50} in presence and absence of Mg^{++}; the results of these computer simulations were used to guide subsequent site-directed mutagenesis in the vicinity of Y_{203}."

Thanks

- Medline, EMBL, SwissProt, PDB, EBI, NCBI
- RONN, PsiPred, Jpred, CDM, YASPIN
- Mafft, ClustalW, SeaView
- CAPS, PIPs, PRISM
- ConsPPISP, MetaPPISP, Polyview, PPI-Pred, ProMate, ConSurf, 3d_Partner
- CPHmodels, HHPred, LOOPP, MUSTER, Phyre, Phyre^2, (ps)^2, PsiPred, I-tasser, GeneSilico, blastp, compass, ffas, FUGUE, HHsearch, jmbrank, pcons5
- PDBblast, PRC, sparksLoMets, SAM-T02, Sparks-2, SP3, PROSPECT2, PPA-I, PMP, SwissModel, M4T, ModWeb
- QUARK, Rosetta, Robetta
- UCSF Chimera, PyMol
- Gramm, Hex, Zdock, PatchDock/FireDock, SymmDock
- Gromacs, Triton, Modeller, MOPAC2009

More thanks

- To Britt Mellström and José R. Naranjo
- To CYTED and FreeBIT
- To CSIC for “trueno”
- To CESGA for “finisterrae”
- **To all of you** for coming