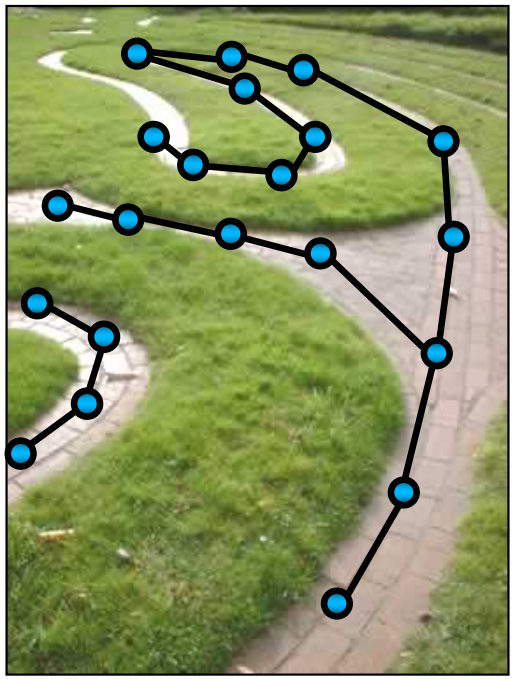
PathRover - Sampling Protein Motion in an Energy Maze



Barak Raveh - Furman Lab

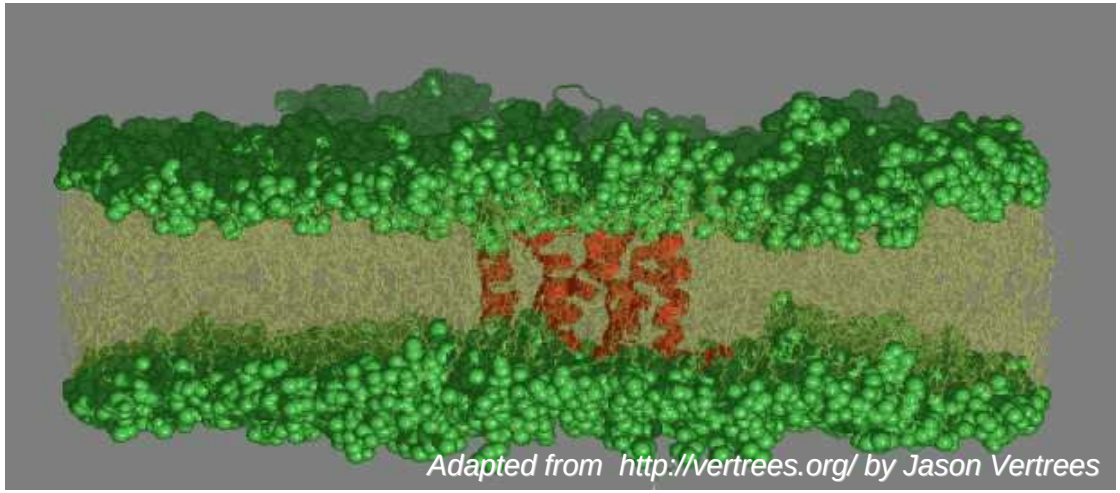
Hadassah Medical School, Hebrew University, Jerusalem

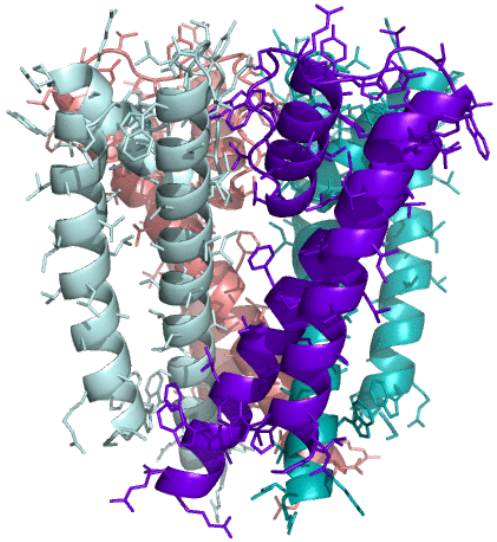
School of Computer Science, Tel-Aviv University

Joint work with Angela Enosh, Nir Ben-Tal and Dan Halperin

Rosetta Conference, 2008

PathRover - From Structure Prediction to Motion Prediction in Rosetta



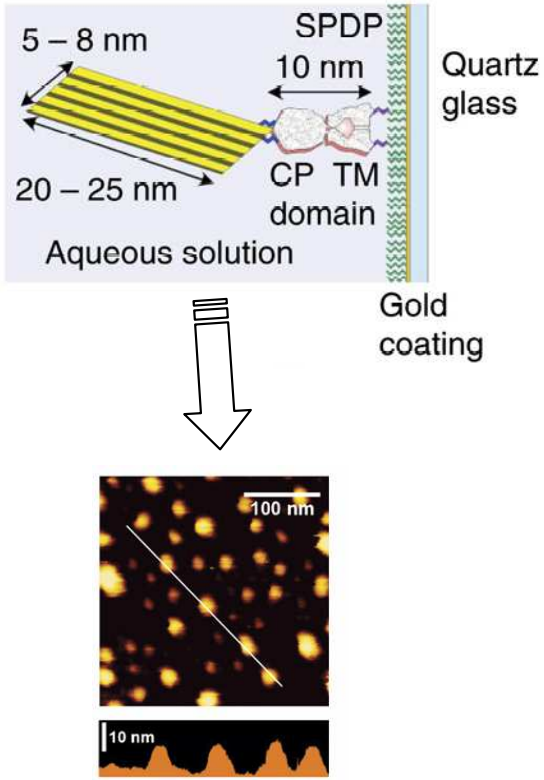


From Protein Structure to Protein Motion

Bad news: experimental methods for observing protein motion in high-resolution are still a dream

Good news: there's some hope

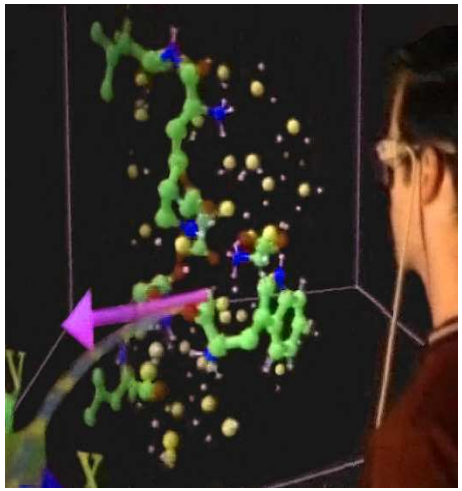
- **Nano-labeling with gold particles**
- **Spectroscopic Methods**
  - RDC, PRE, FRET, SAXS, etc.
- **Homologues & Alternative Conformations**



Shimizu et al., Cell, 2008

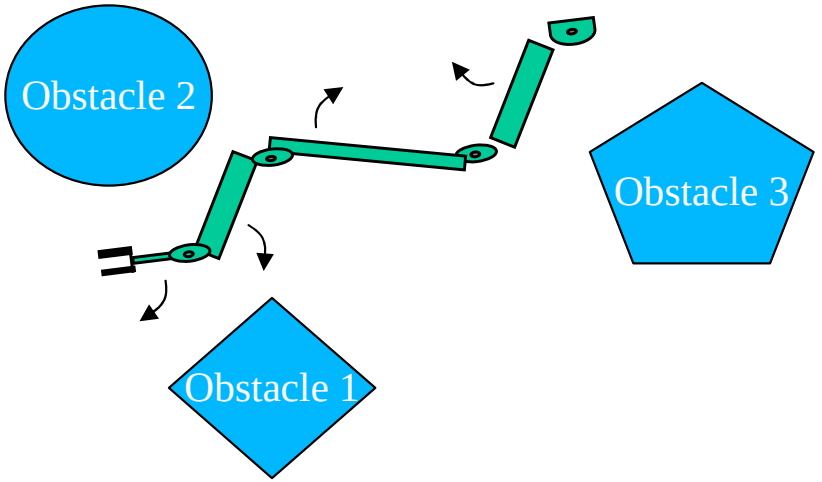
Outline

- Introduction to Sampling-Based Motion Planning
- Some Results
- Conclusions and Future Work

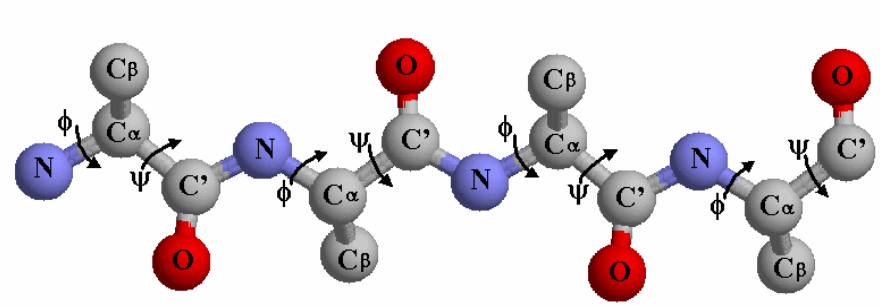


Motion Planning Techniques – From Robots to Molecules

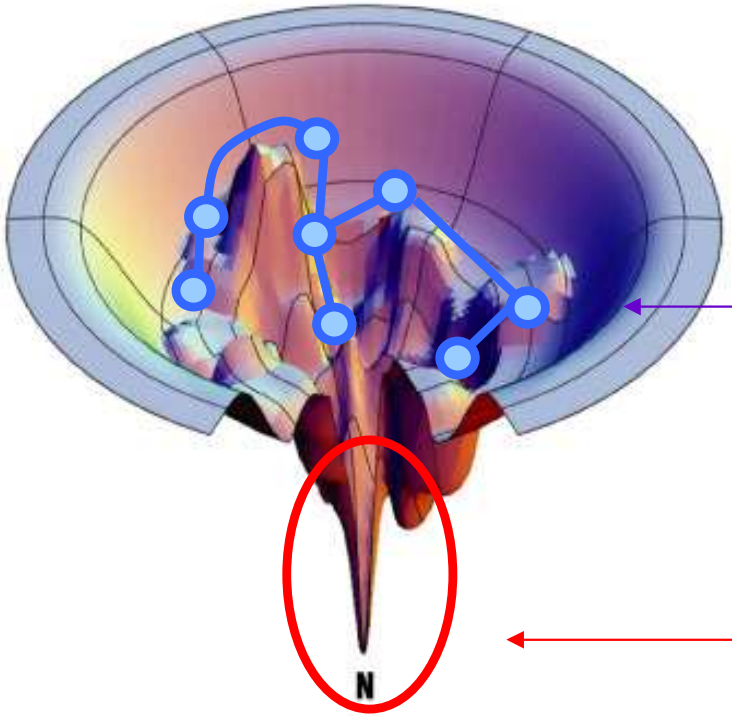
Robot Motion Planning: Given a *robot* with *k* degrees of freedom, in an environment with *obstacles*, find a *collision free path* from an initial state to a goal state



- *Robot* ➔ **peptide chain**
- *Obstacles* ➔ **steric clashes between atoms**
- *Collision-free path* ➔ a low-energy **motion pathway**, free of steric clashes



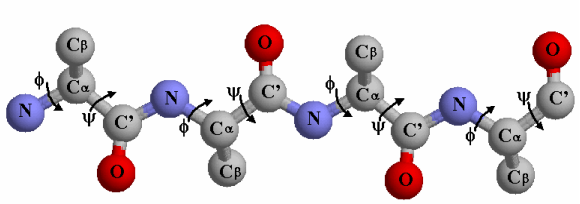
Sampling Motion Paths in the Energy Landscape

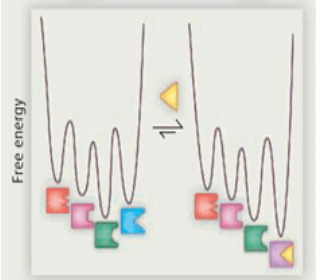


Folding

Conformational Changes

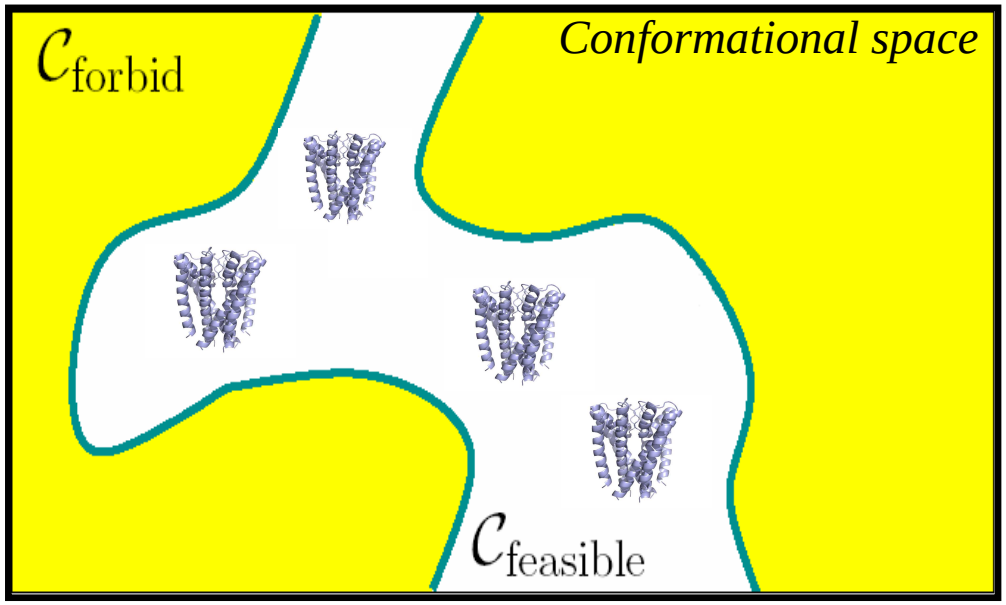
http://www.dillgroup.ucsf.edu





Boehr & Wright, Science 2008

Feasible and Forbidden Space

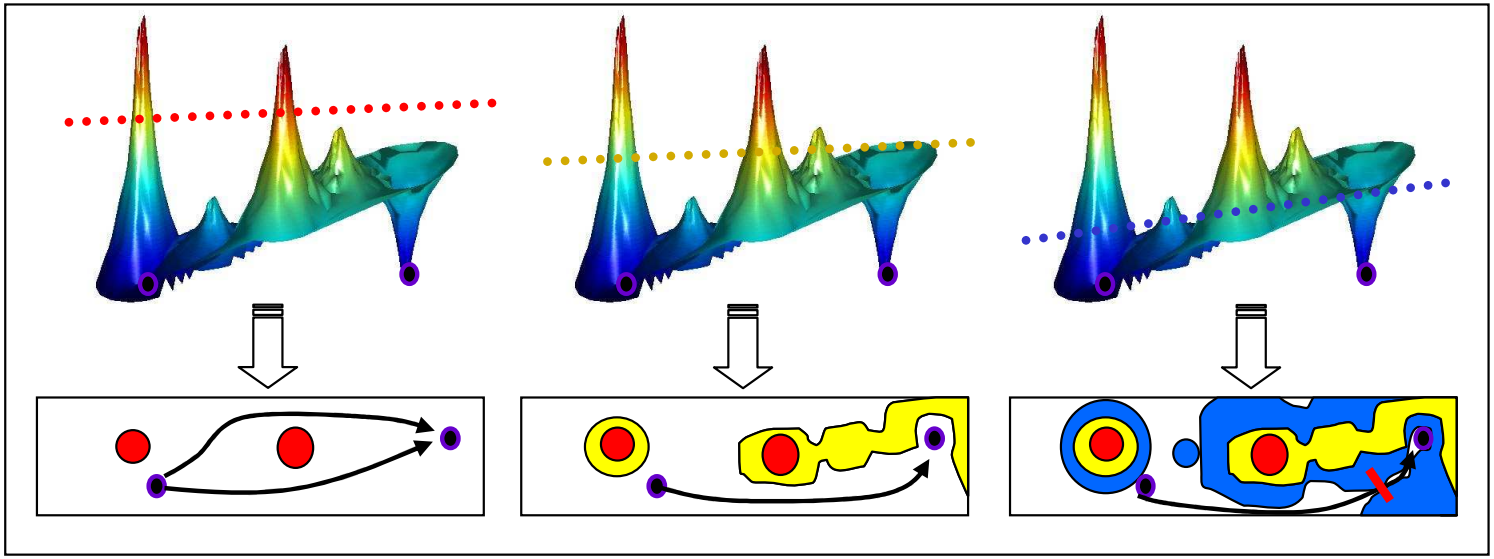


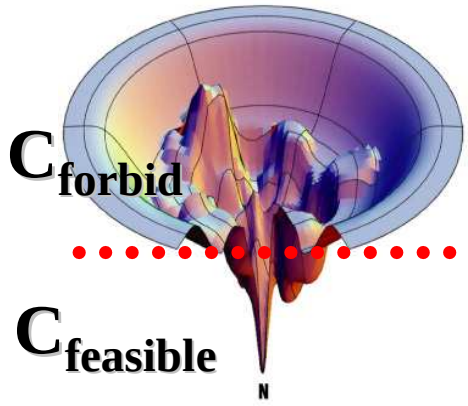
C_{forbid} = Energy > Threshold

e.g., conformations with steric clashes or conformations with poor solvation

$C_{feasible}$ = Energy < Threshold

e.g., clash-free conformations

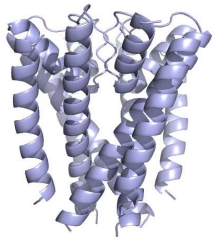




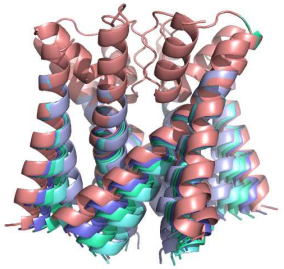
Rapidly-exploring Random Trees (RRT)

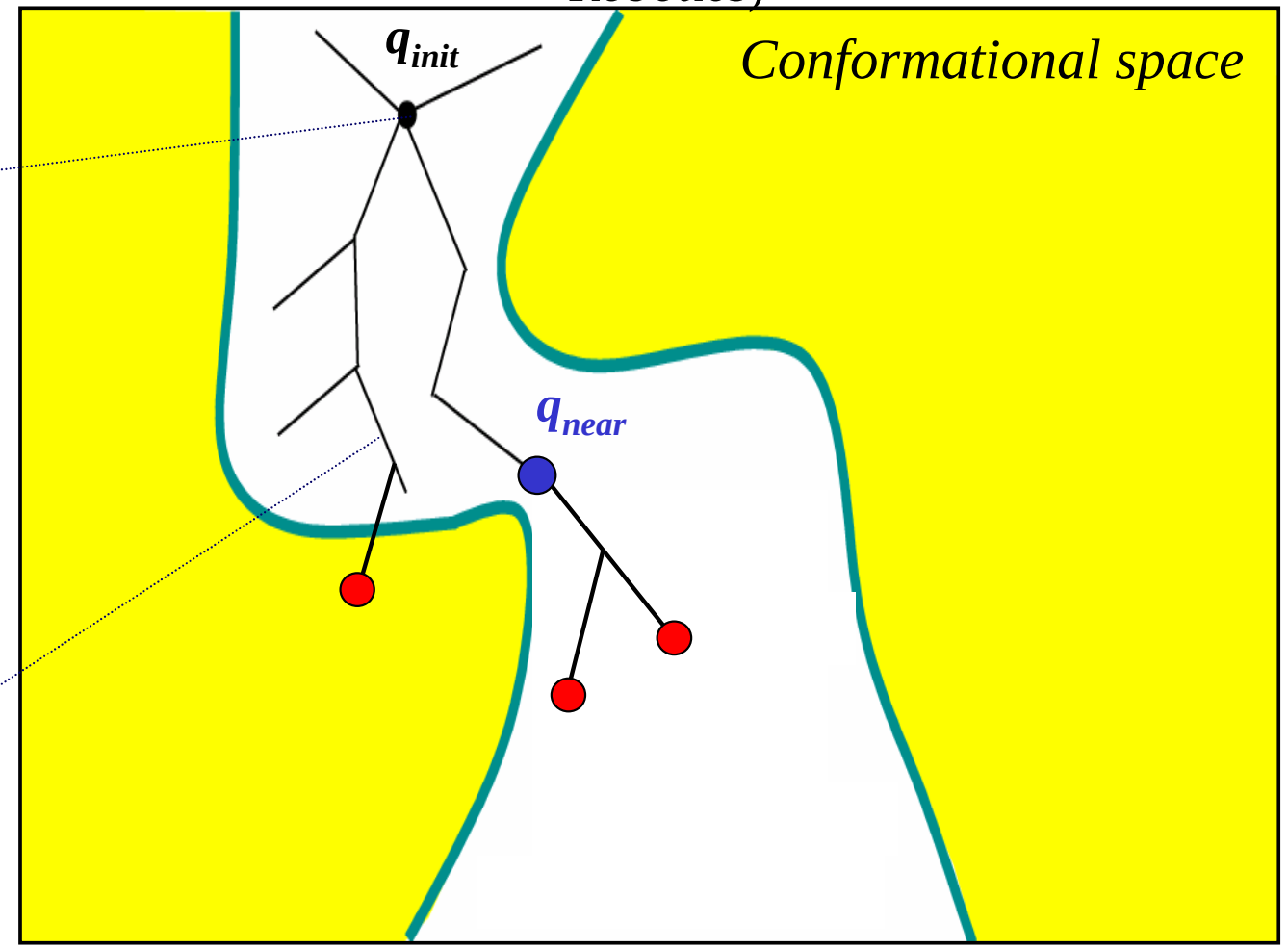
Mapping the feasible conformation space (LaValle & Kuffner, 2001, *Robotics*)

Nodes = conformations



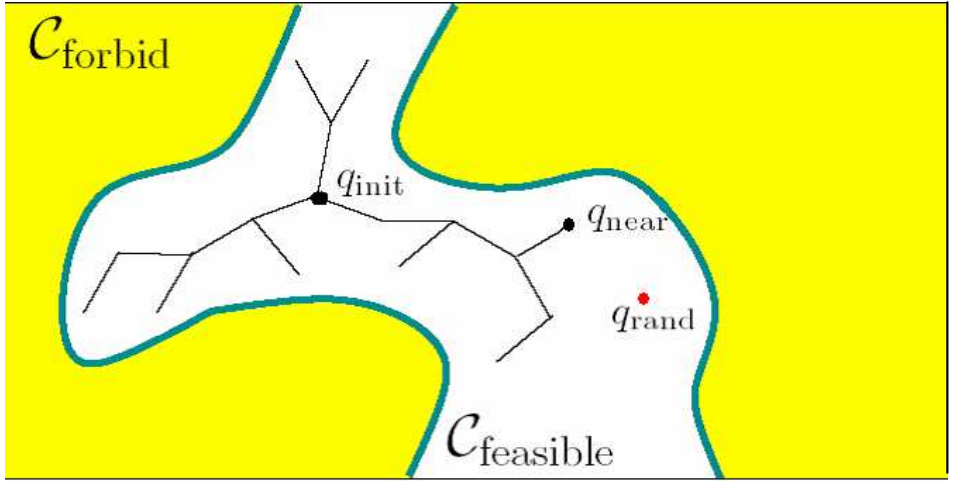
Edges = motion in feasible space





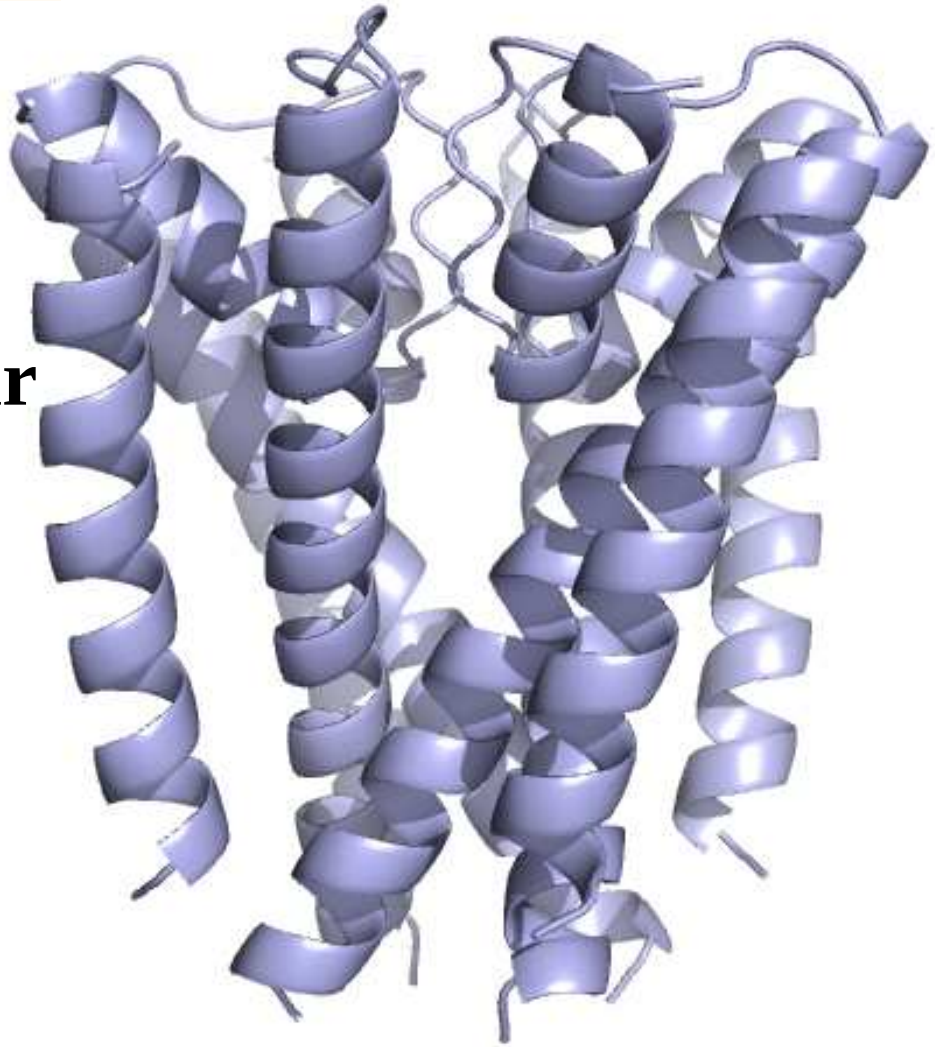
$C_{feasible}$ – plausible conformations

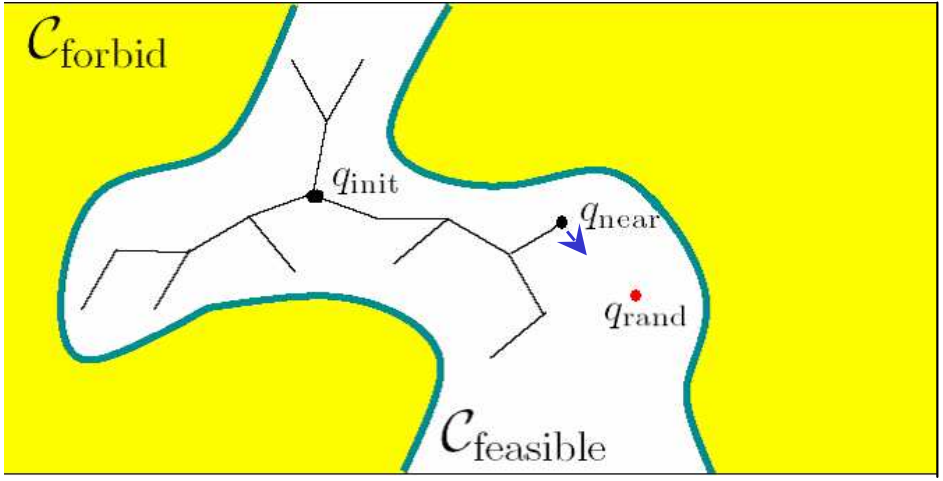
$C_{forbidden}$ – high energy conformations



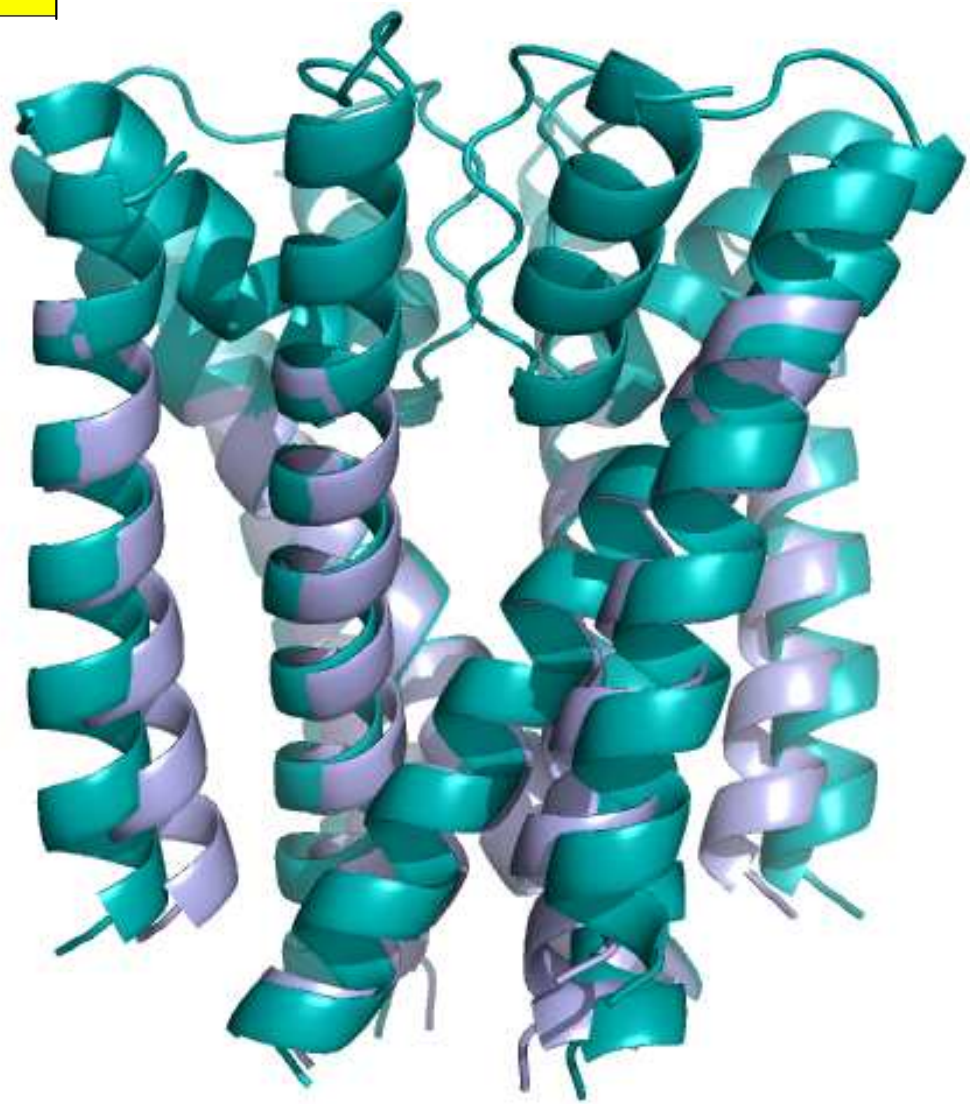
Dense sampling → continuous clash-free motion

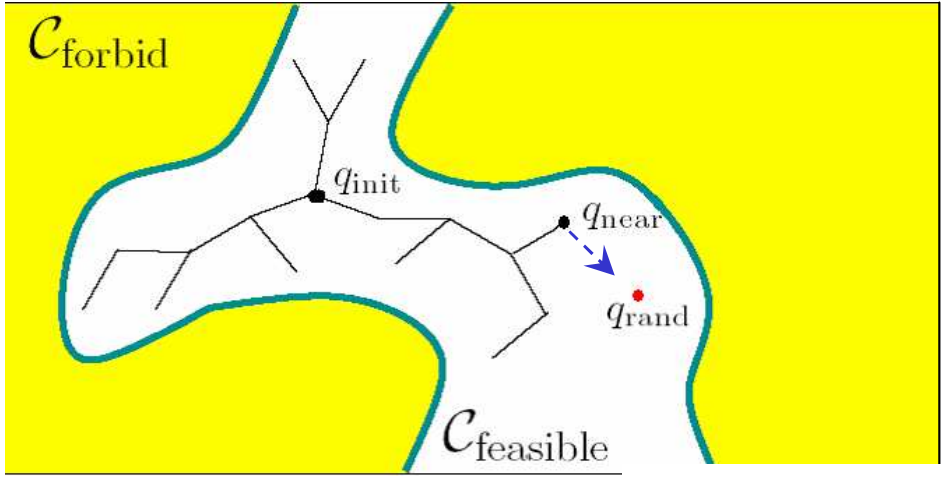
q_{near}



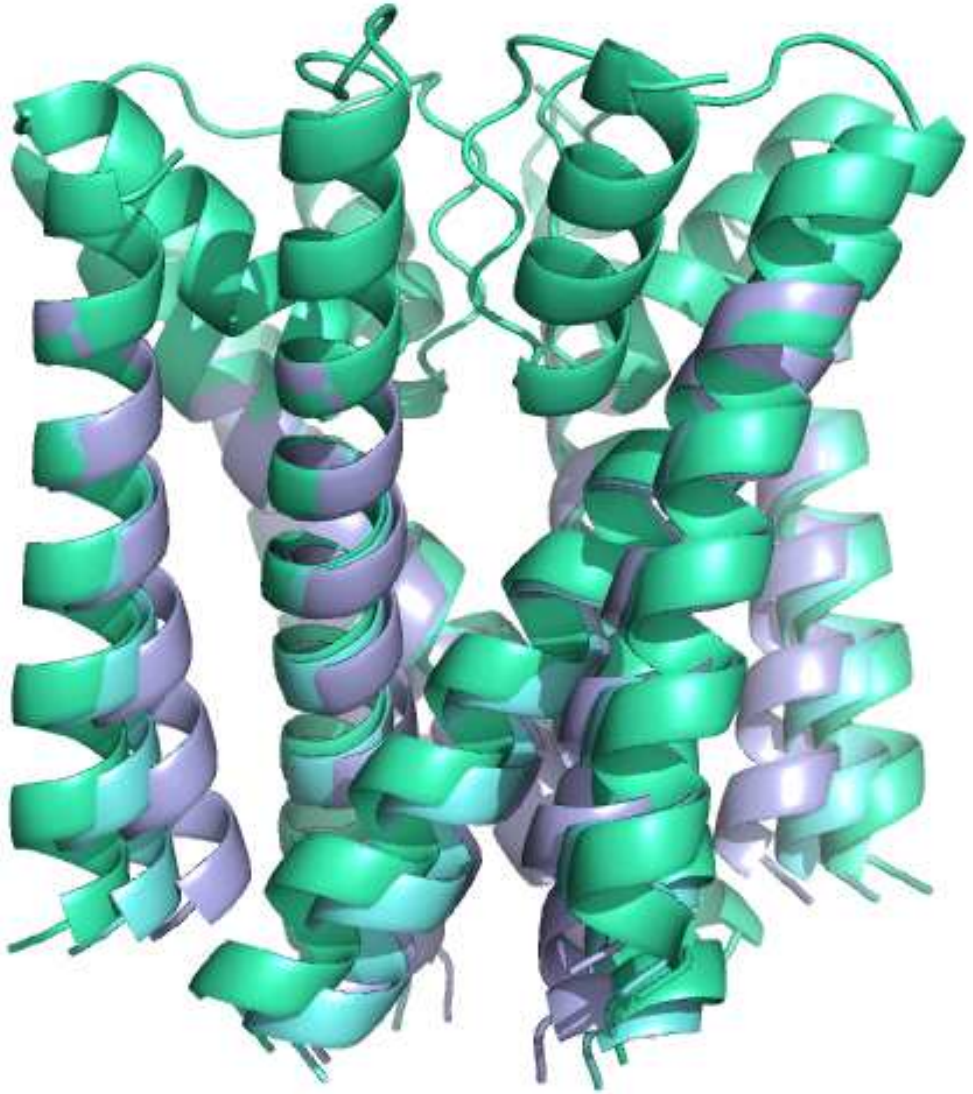


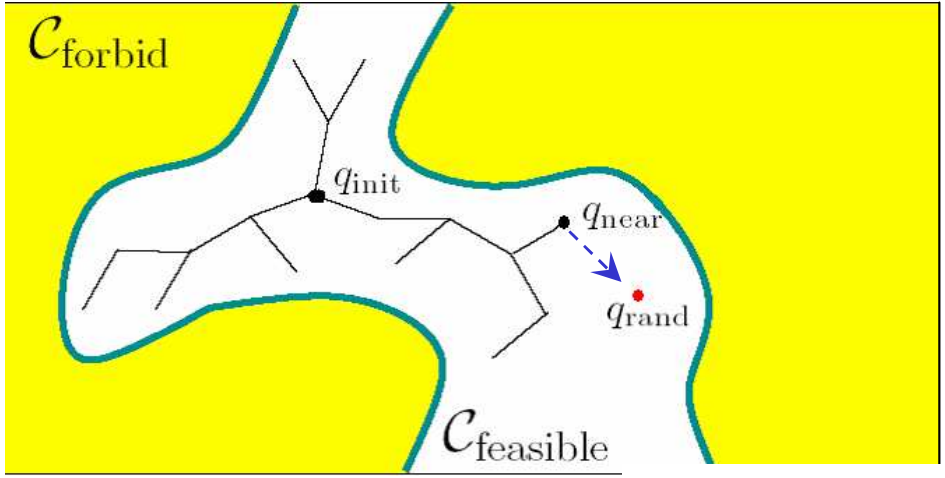
Dense sampling → continuous clash-free motion



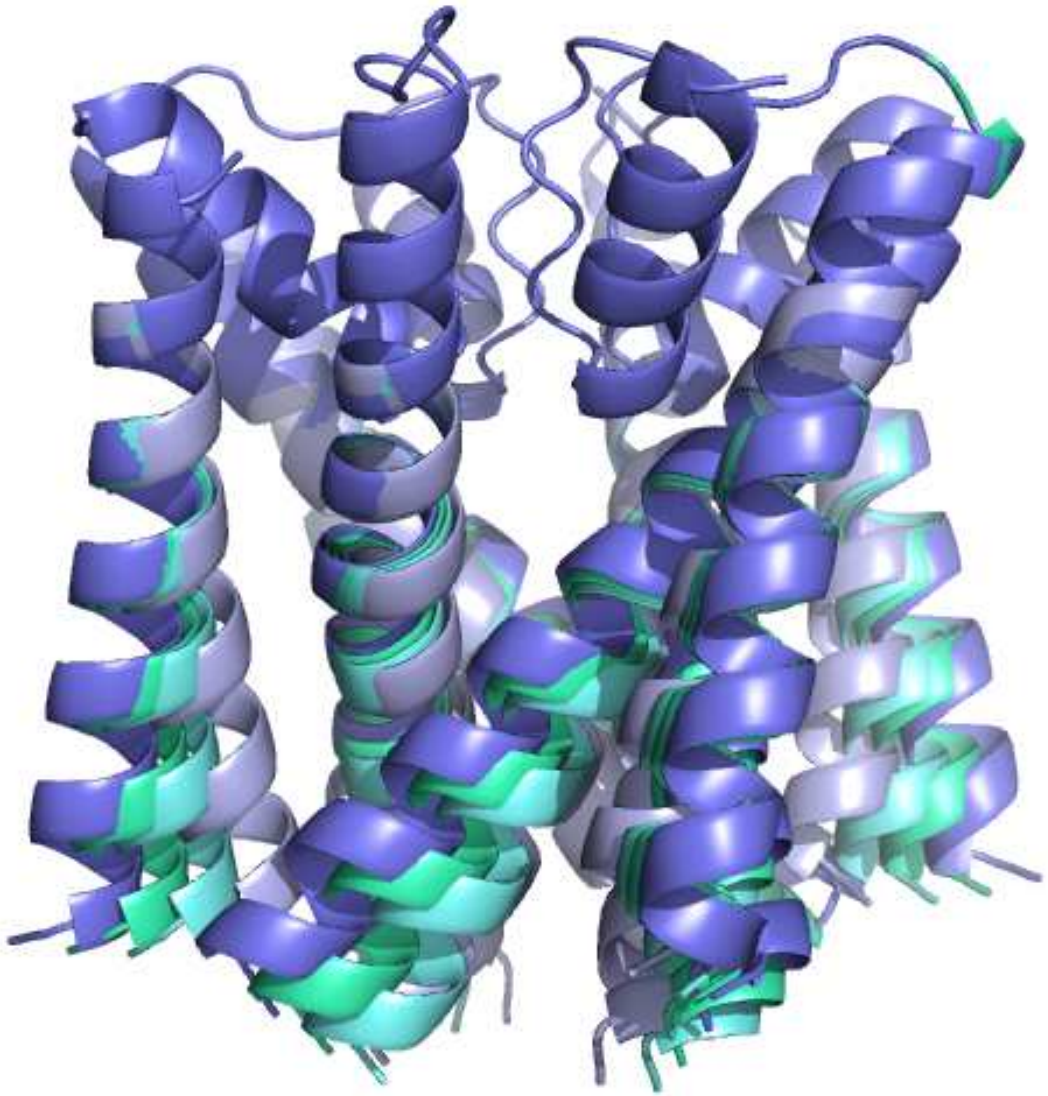


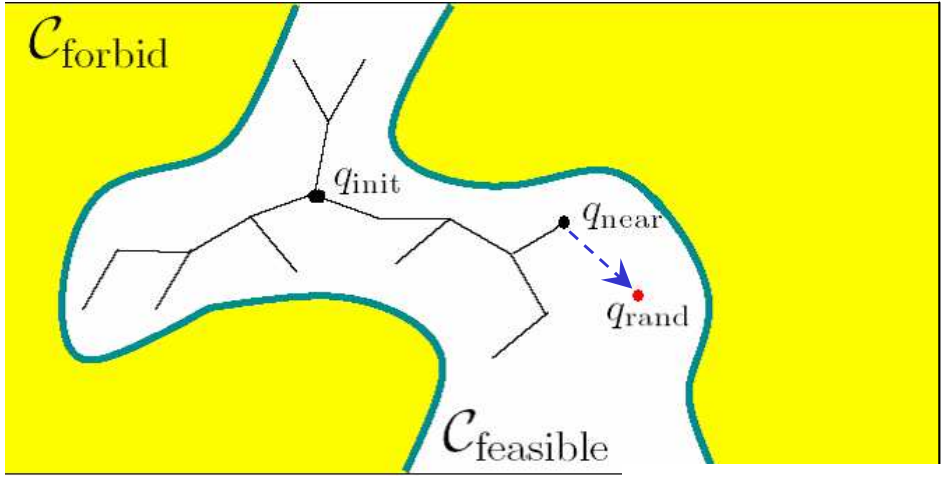
Dense sampling → continuous clash-free motion



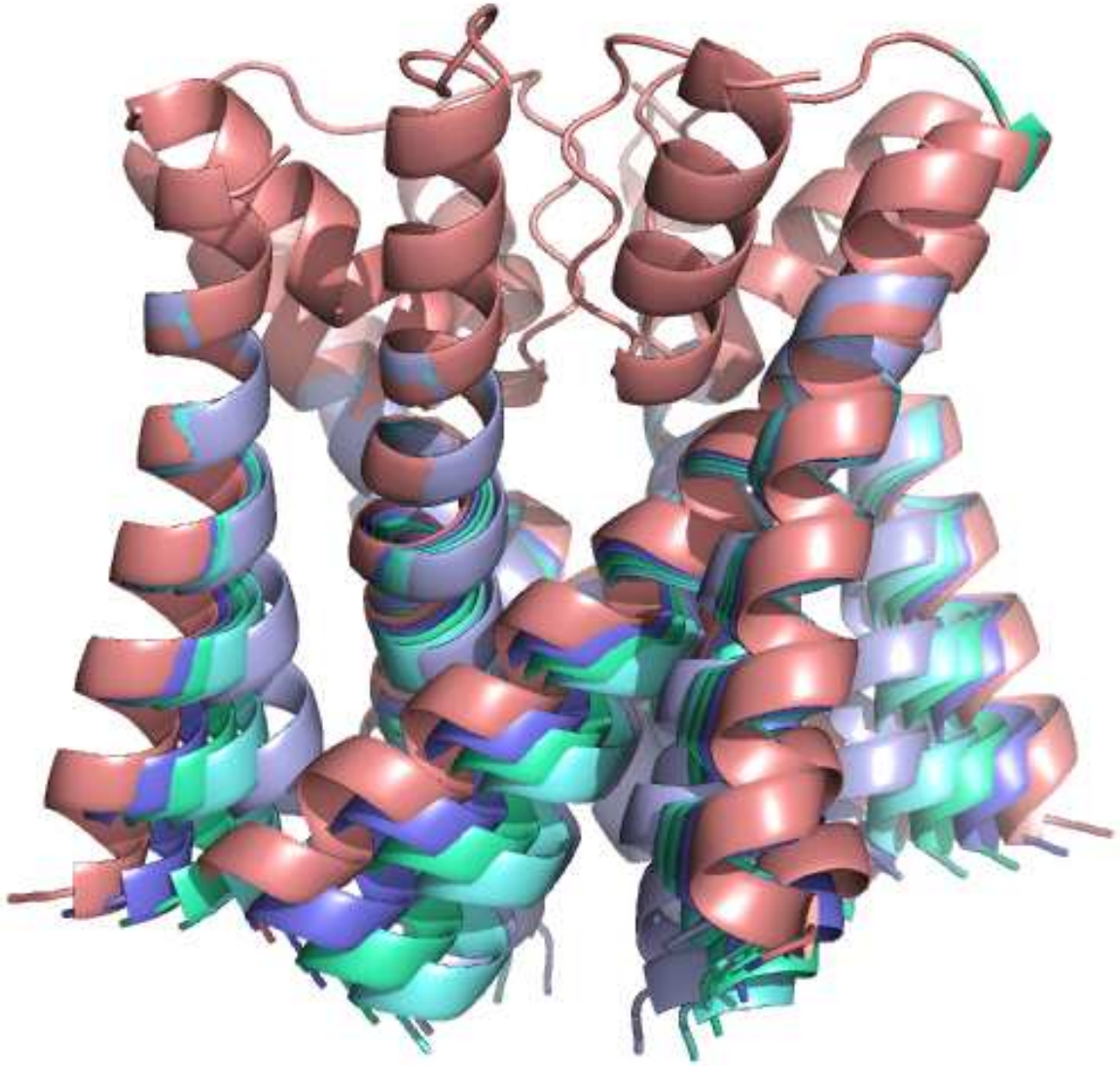


Dense sampling → continuous clash-free motion





Dense sampling → continuous clash-free motion

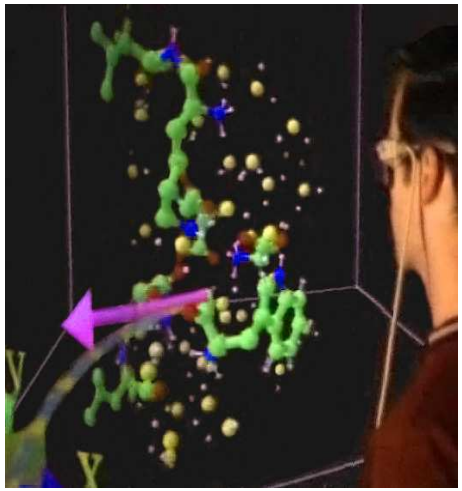


Sampling-Based Motion-Planning – Some Previous Work

- Protein folding pathways Amato, Dill & Song, 2003
- Protein loop motion Cortés, Siméon, Remaud-Siméon and Tran, 2004
- Large-amplitude conformational changes
  Cortés, Siméon et al., 2005
  - Integration with Normal-Mode Analysis
    Kirillova, Cortés, Stefaniu and Siméon, 2008
- Ligand binding Singh, Latombe and Brutlag, 1999; Cortés, Jaillet and Siméon, 2007
- Motion of transmembrane helices
  Enosh, Fleishman, Ben-Tal & Halperin, 2007

Outline

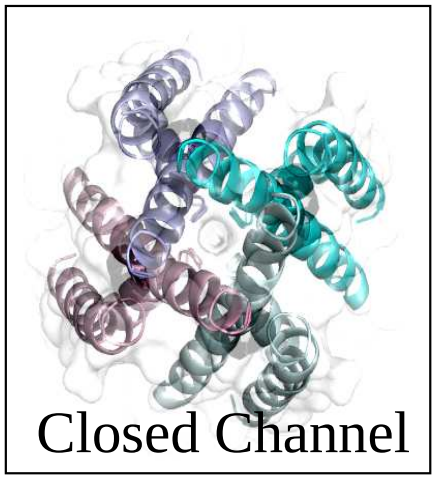
- Introduction to Sampling-Based Motion Planning
- Some Results
- Conclusions and Future Work



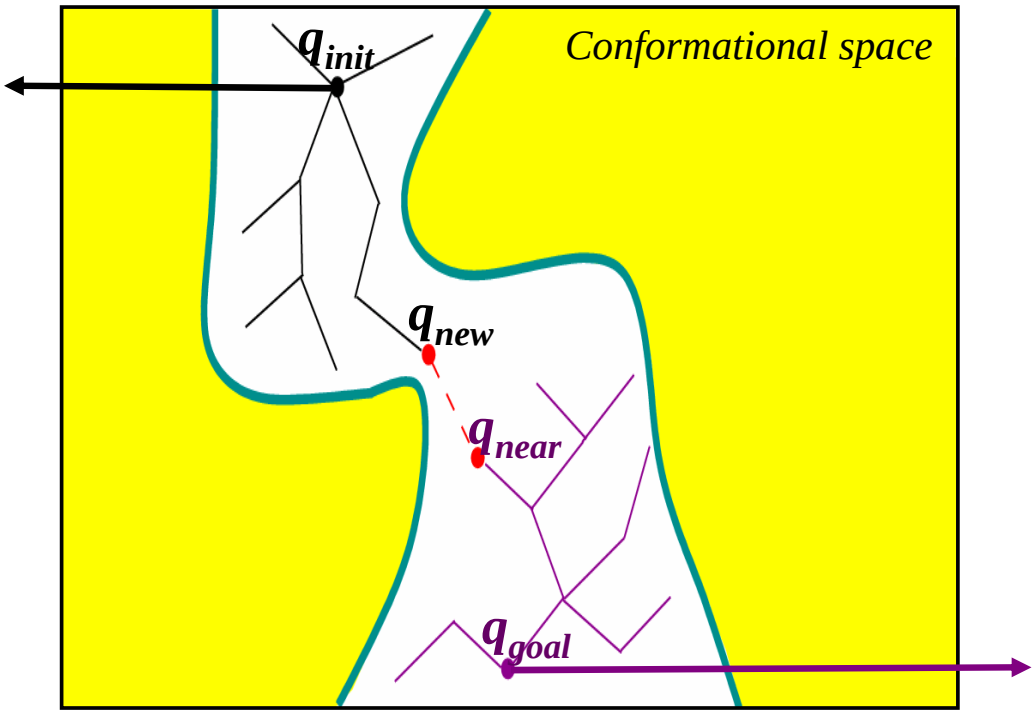
KcsA Channel

Case Study I: Modeling the Motion of the Transmembrane Potassium Channel KcsA

(Enosh, Raveh, *et al.*, J Biophysics – to appear)

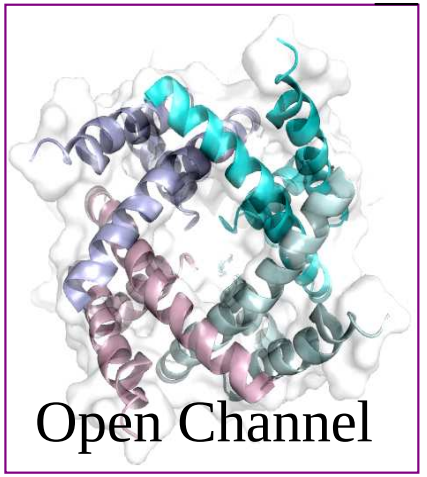


Closed Channel



plausible conformations

high energy conformations



Open Channel

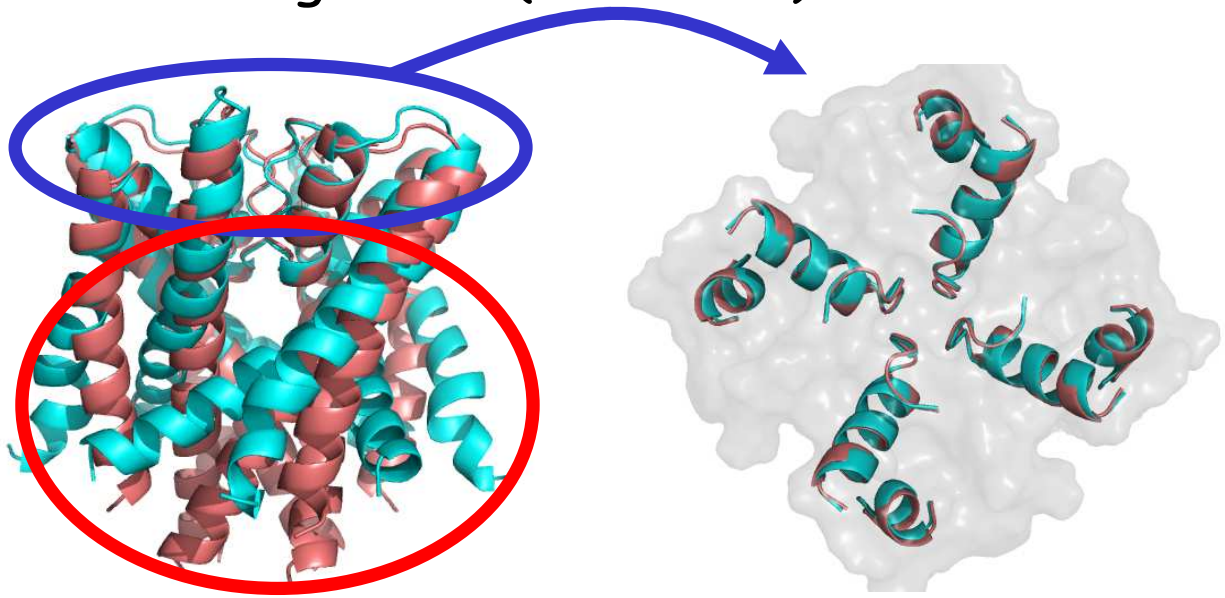
Open state = an homology model of the *KvAP* channel

Definition of Flexible Regions for KcsA

- Flexible structural alignment (FlexProt*)

Constant part: 0.3Å deviation only

Flexible part: 104 backbone degrees of freedom + side-chains



Additional Tools:

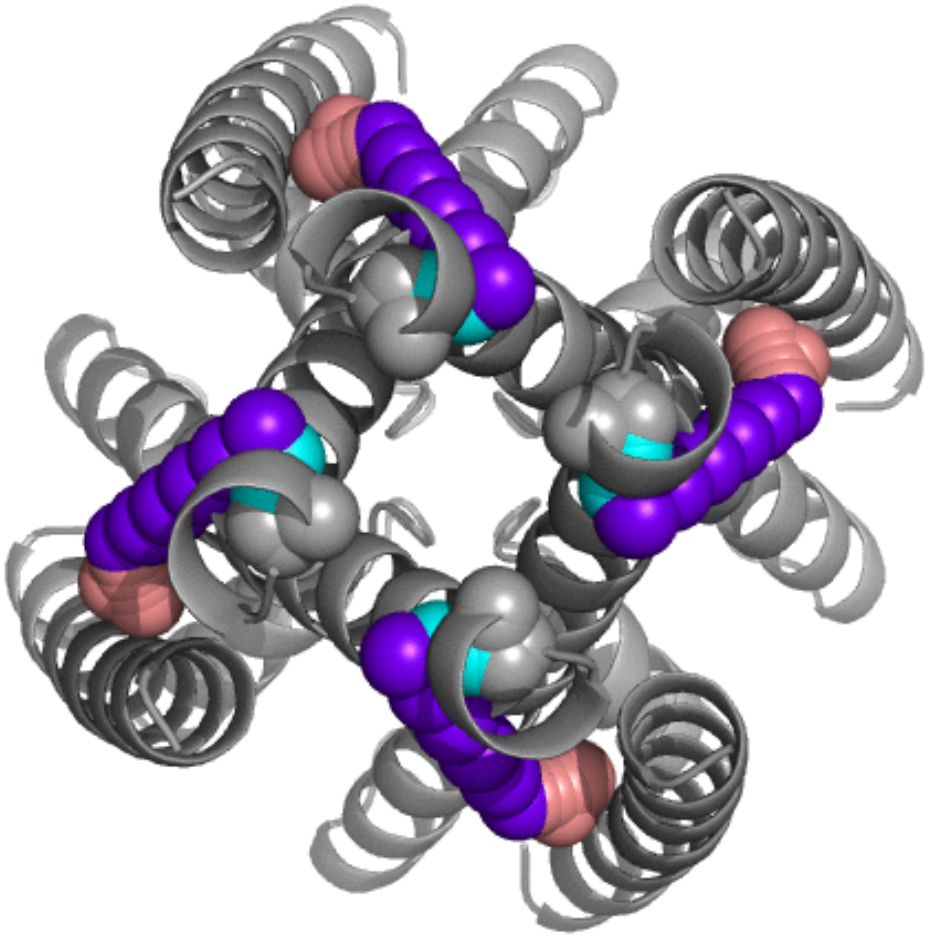
- Comparison of phi / psi values
- Hinge-Prediction Tools (elastic-network models, etc.)

* Shatski et al., *Proteins* 2002

Technical Run Details for KcsA

- 100 paths generated from 12 independent runs
  - Running time: ~1 hour for each run on a single CPU
  - Tested ~50,000 conformations in each run
  - Added ~30,000 valid conformations to each tree
- Final output path of 23 conformations
- Energy Function
  - Full-atom *score12*, excluding solvation & pair terms
  - Mainly prevents clashes
  - Should be compared to a membrane-tailored scoring function (e.g., Barth, Schonbrun and Baker, *PNAS* 2007)

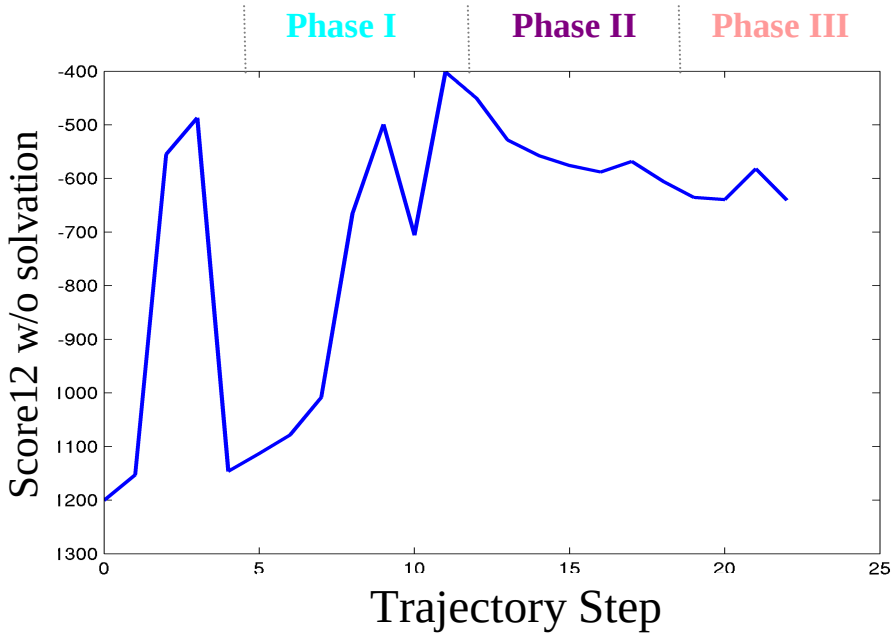
Result: KcsA Motion Prediction

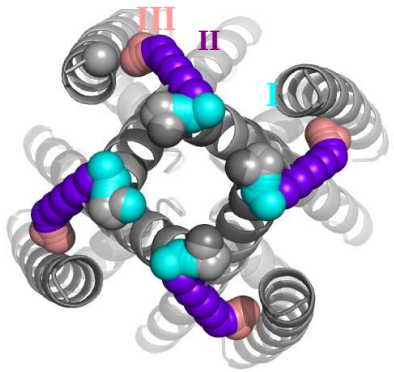


Three-Phase Safe-Lock Mechanism?

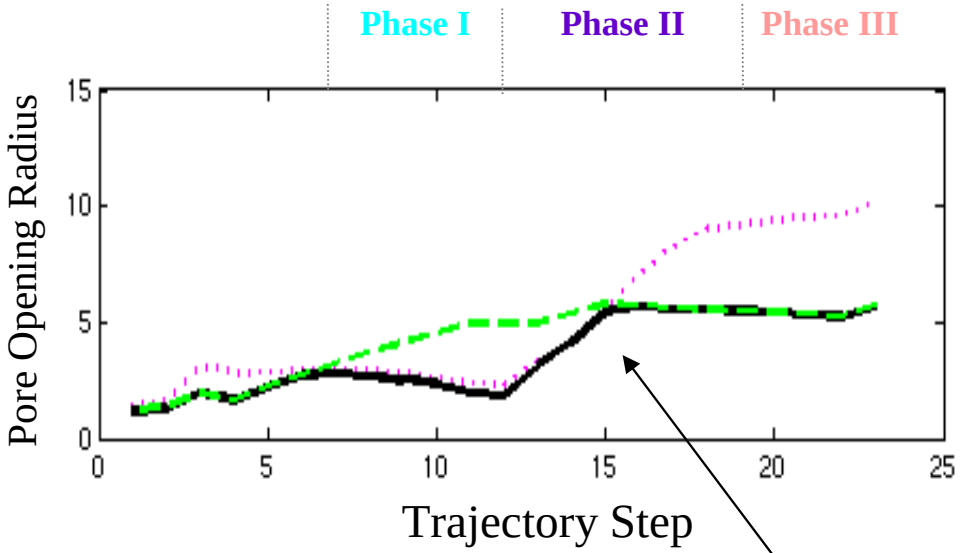
unlock → open → relock

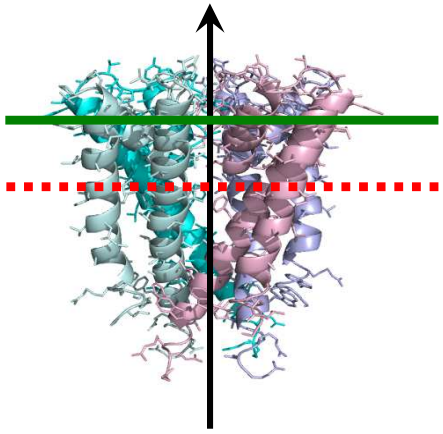
Energy Along Pathway
(score12 without solvation & pair terms)





Radius of Pore Opening during Simulation (HOLE program)





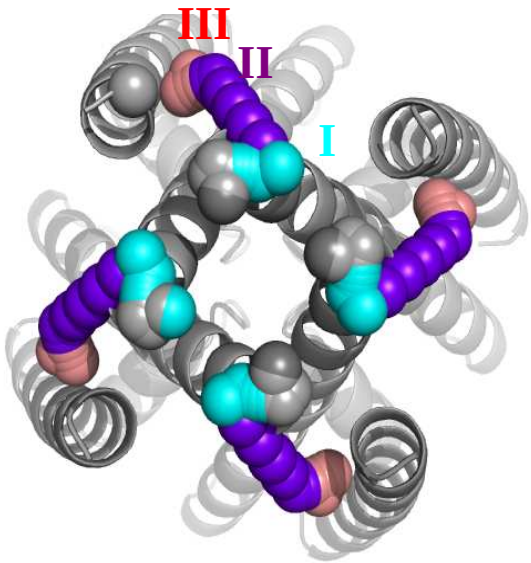
black line →
minimal opening radius of the channel pore

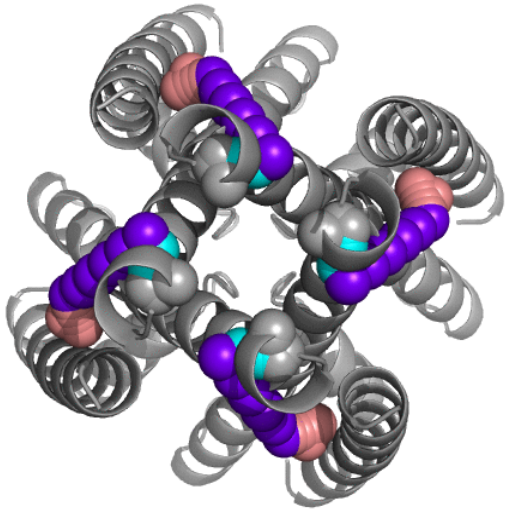
Three-Phase Safe-Lock Mechanism?

unlock → open → relock

Following a short phase of Brownian motion (*grey*):

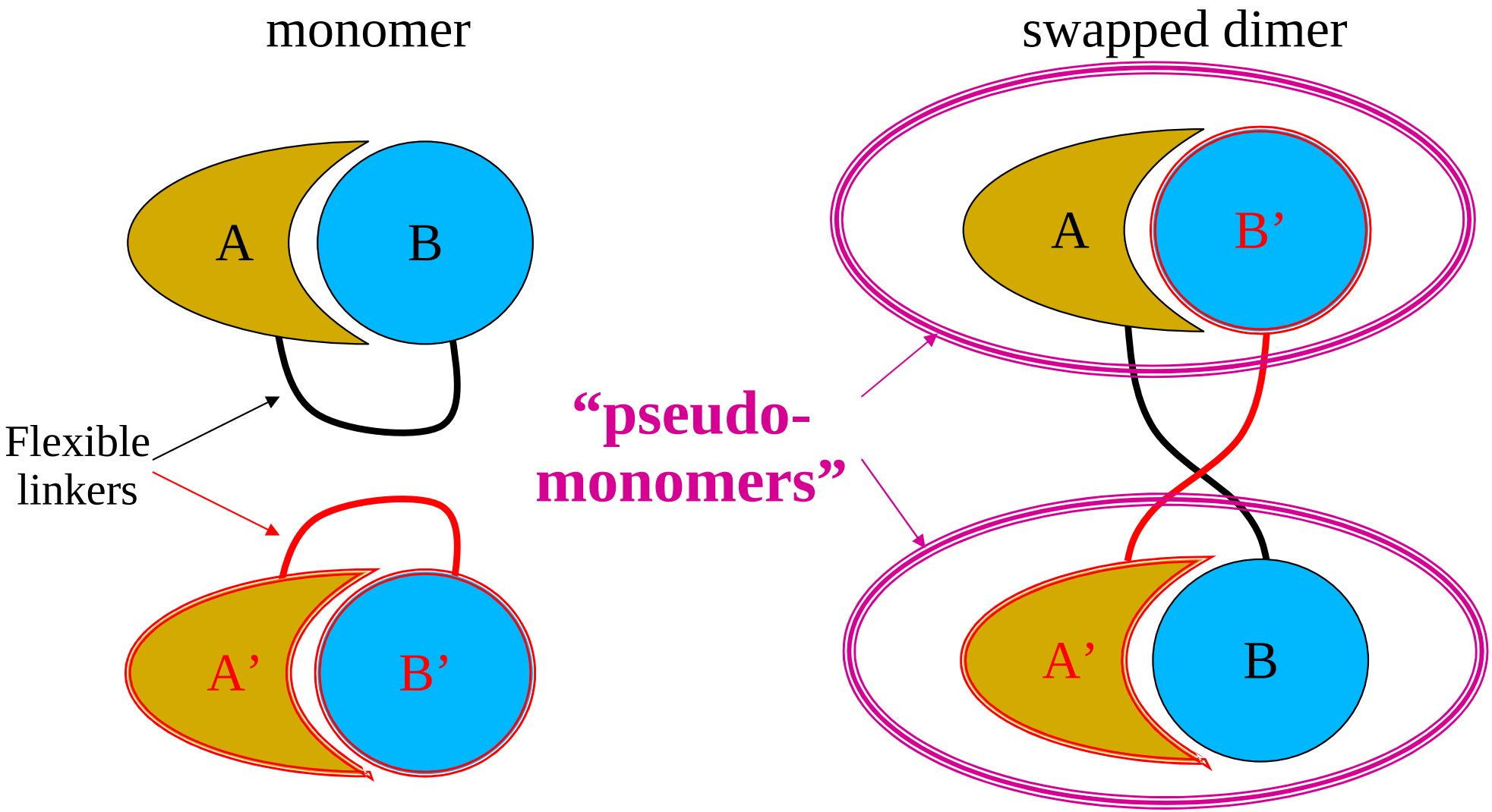
- Unlocking from closed state by a slight clockwise movement.
- Opening: the inner helices slide over each other, moving counter-clockwise and laterally away from the pore axis.
- Locking in the open conformation by a counter-clockwise motion.





Case Study II: Domain Swapping – Oligomerization Mechanism

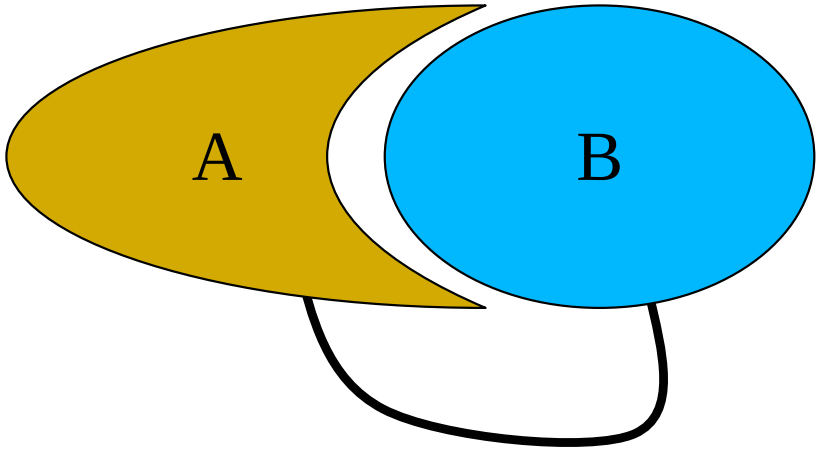
(Raveh, Enosh et al., Submitted)



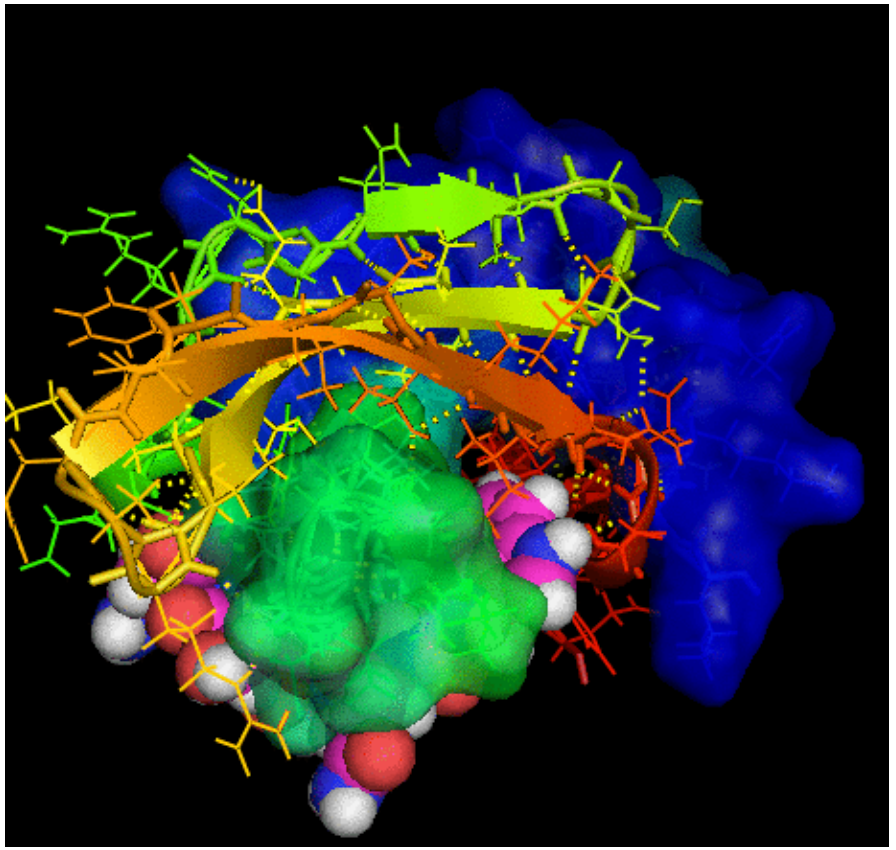
Cyanovirin-N Domain Swapping

(Full-atom mode, Score12)

Flexible Alignment*, Normal-Mode Analysis** → one central hinge loop



The central hinge loop does not allow domain swapping
(only in full-atom mode!)

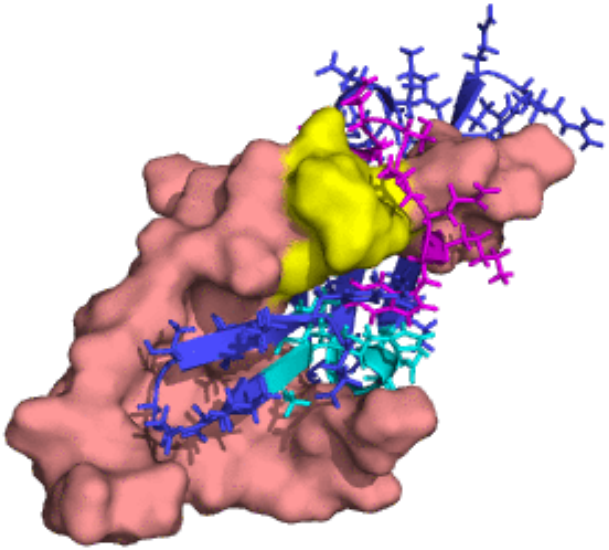


* Shatski et al., *Proteins* 2002

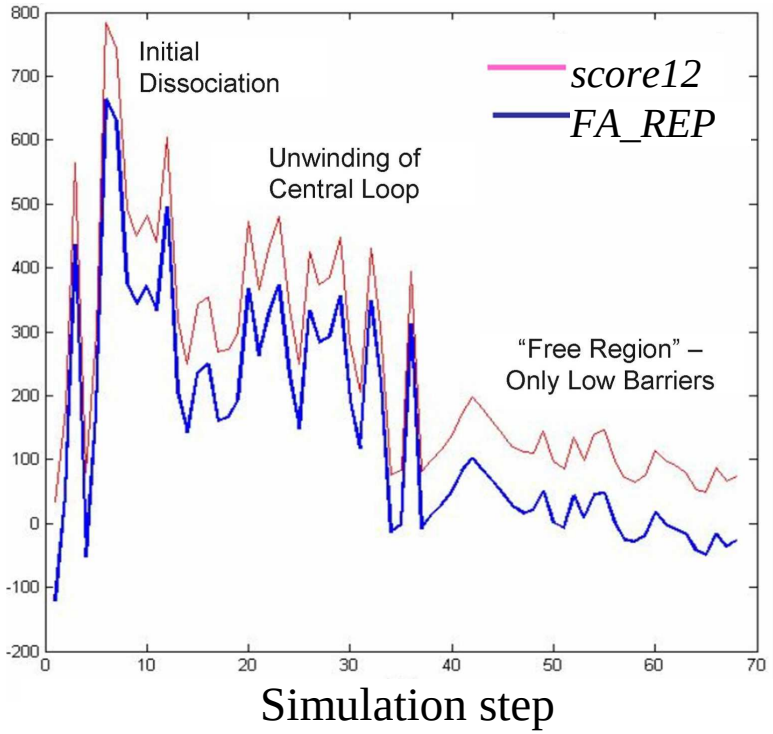
** Emeki et al., *Proteins* 2008

Cyanovirin-N Domain Swapping - Two Additional Hinge Loops Allow the Motion

Phi / Psi comparison → secondary hinge loops



Score 12 and FA_REP



Cyanovirin-N Domain Swapping - Two Additional Hinges Allow the Motion

Conclusions from Cyanovirin-N domain swapping example:

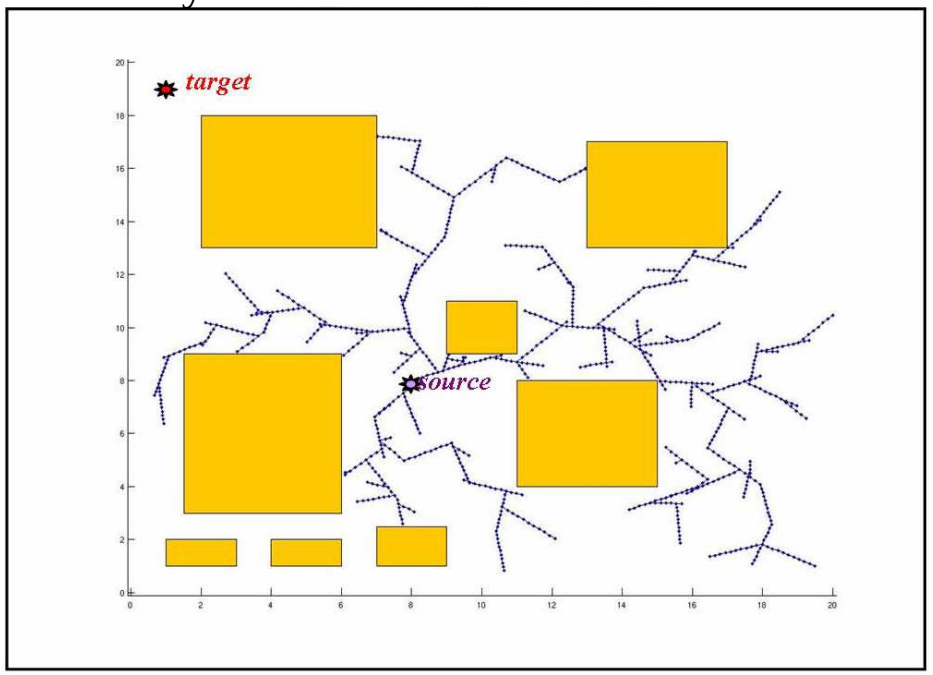
- Certain degrees of freedom may be the key to protein motion
- Not surprisingly, side-chains may play a crucial role in “locking” the protein

Incorporation of Prior Information Constraints in PathRover

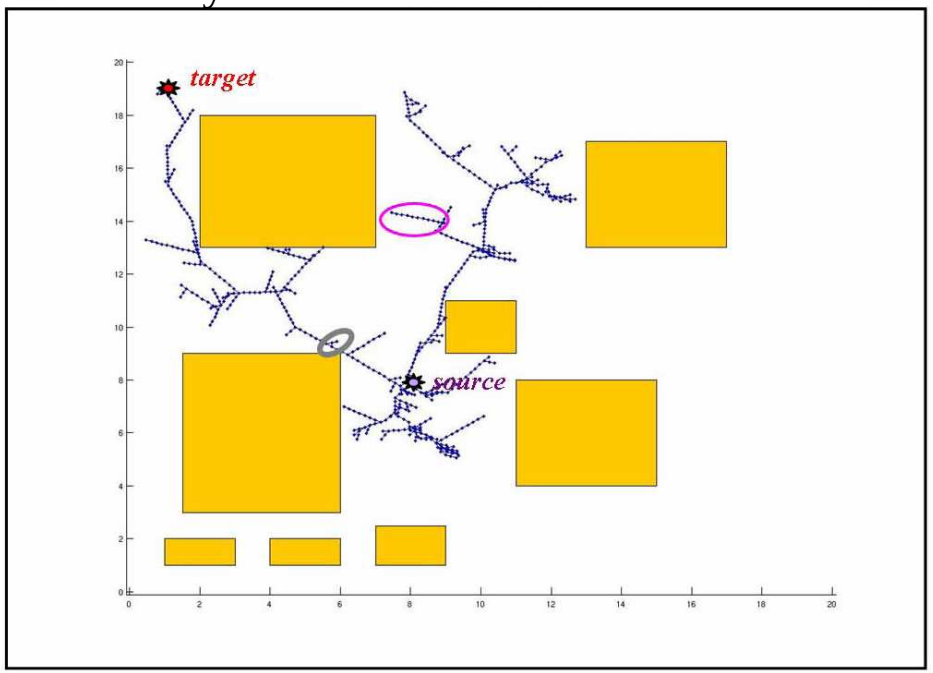
(Raveh, Enosh et al., Submitted)

Experimental Knowledge / Expert Intuition / etc.

2D toy model **without** Partial Information

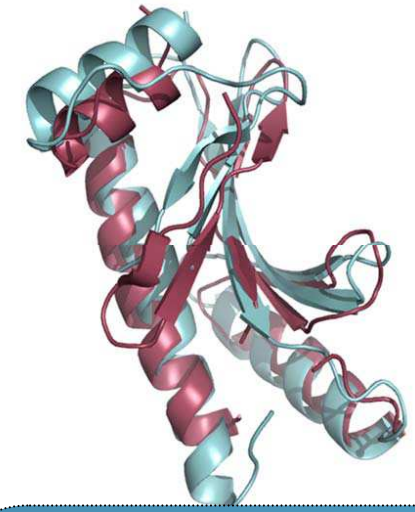


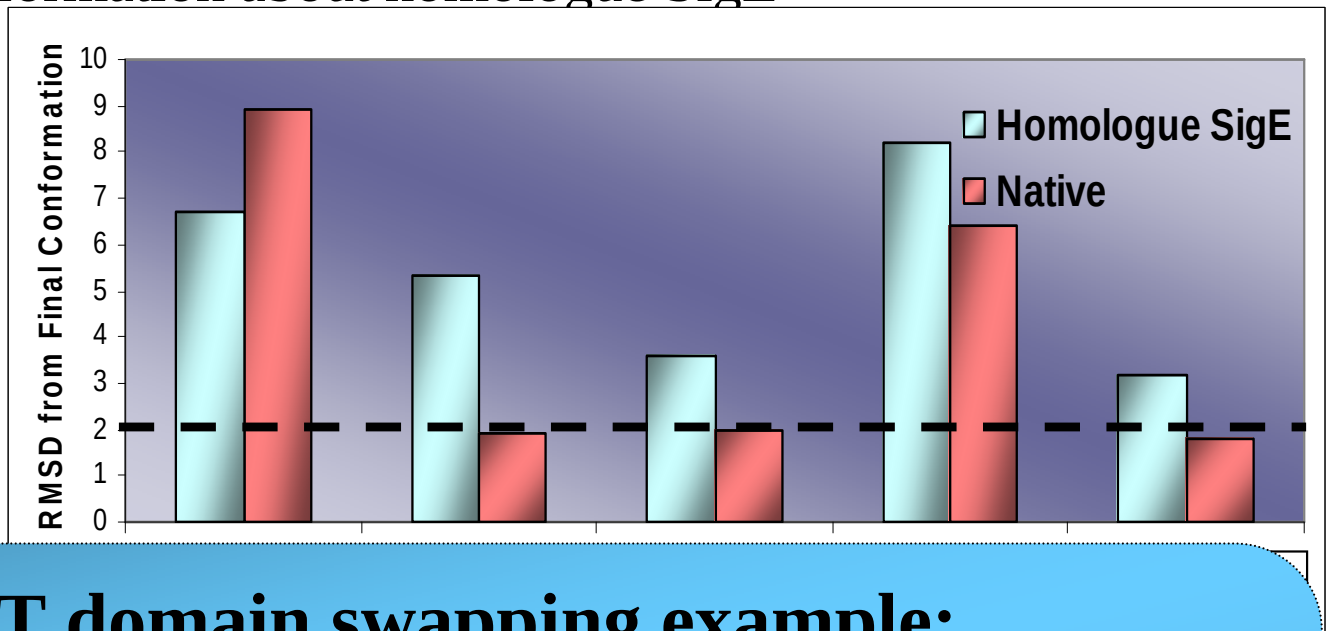
2D toy model **with** Partial Information



Incorporation of Partial Information Contributes to Predictions

CesT domain swapping in centroid mode using partial information about homologue SigE



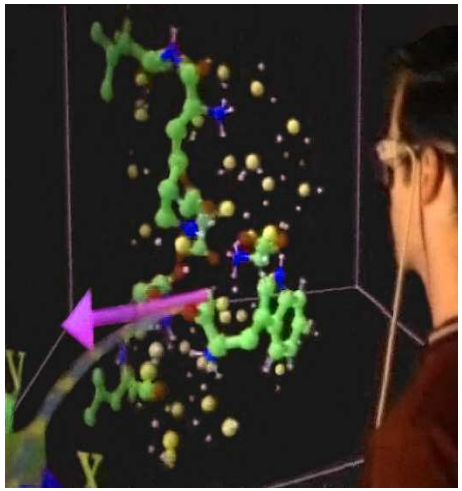


Conclusion from CesT domain swapping example:

- A certain amount of partial information may contribute to predictions
- The energy function prevents over-biasing towards partial information

Outline

- Introduction to Sampling-Based Motion Planning
- Some Results
- Conclusions and Future Work



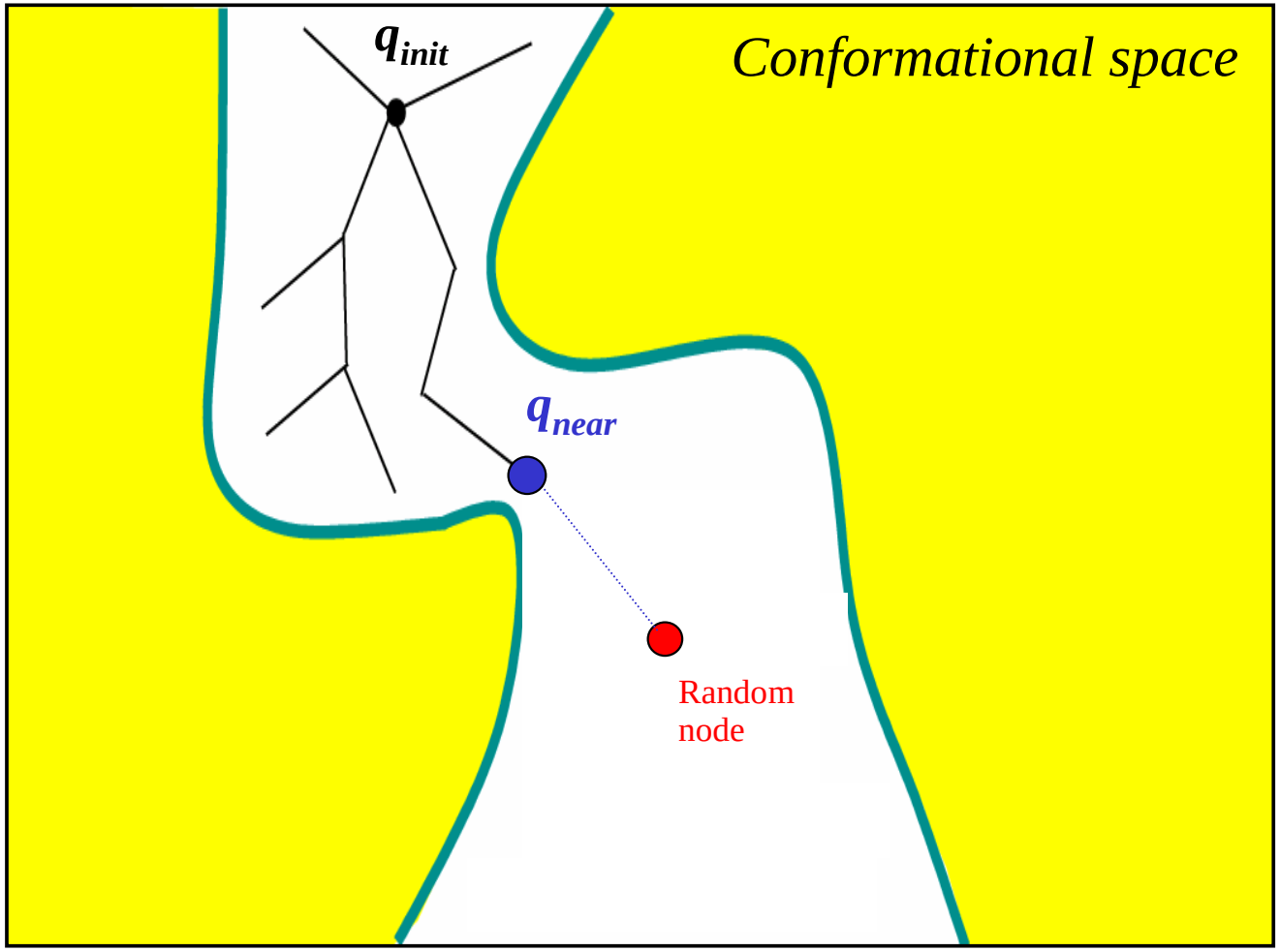
Sampling-based Motion Planning - Conclusions

- Generation of collision-free low-energy motion pathways
  - Fast: minutes (centroid mode) to hours (full-atom mode) per run
  - Partial information constraints may guide towards correct solutions
- Additional validation is still needed

Rosetta PathRover – Back to the future

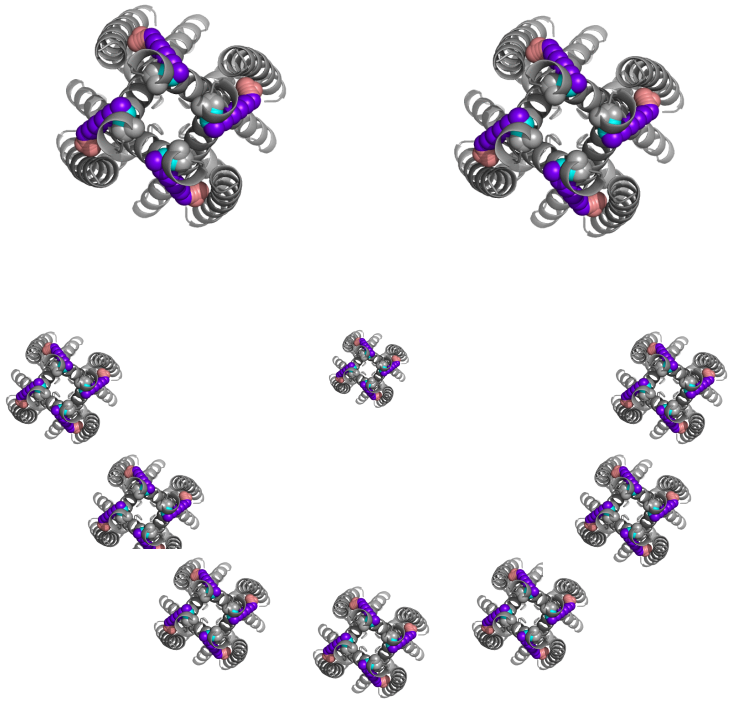
- Incorporating "classical" Rosetta moves
  - Fragment insertion
  - Backrub
  - Loop-closure
  - Etc.
- Experimental validations of predictions
  - Spectroscopic methods
  - Mutation analysis of transition conformations
- Application to Types of Protein Motion:
  - Protein-Peptide Interactions
  - Docking
  - Allostery
  - etc.

Mapping the Conformation Space → Sampling with Memory?



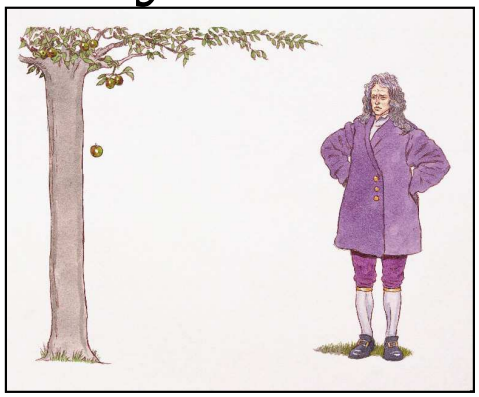
Acknowledgements

- Angela Enosh
- Nir Ben-Tal (KcsA)
- Ora Schueler-Furman & co.
- Dan Halperin & co.
- All Rosetta Developers

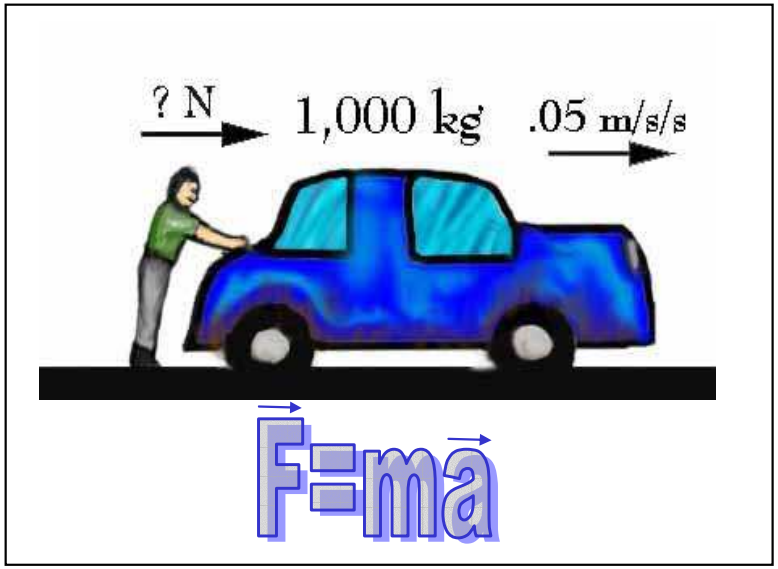


Thank You

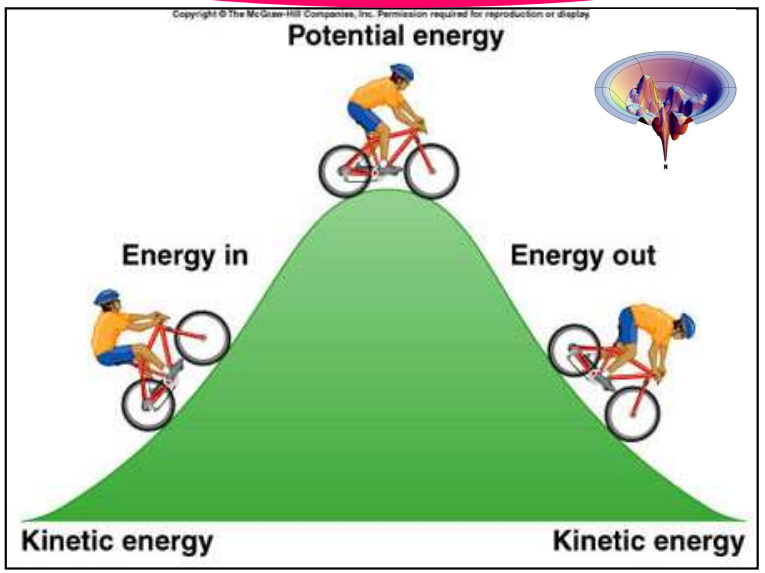
Conceptual Difference from Molecular Dynamics



Force Calculations



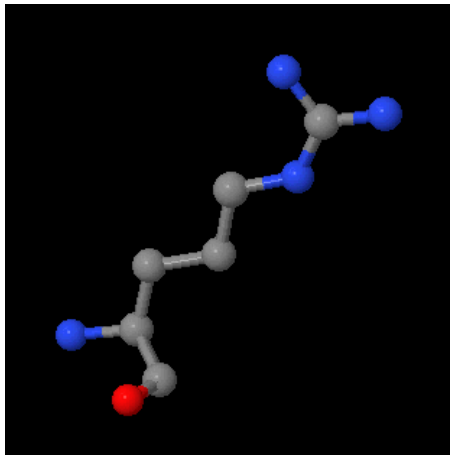
Energy Considerations



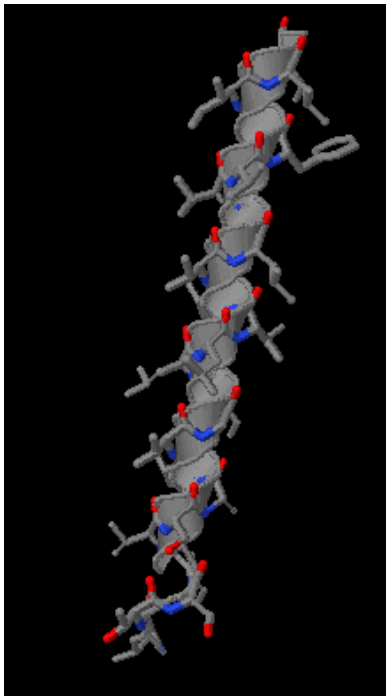
Types of Motions

Side-chain movements

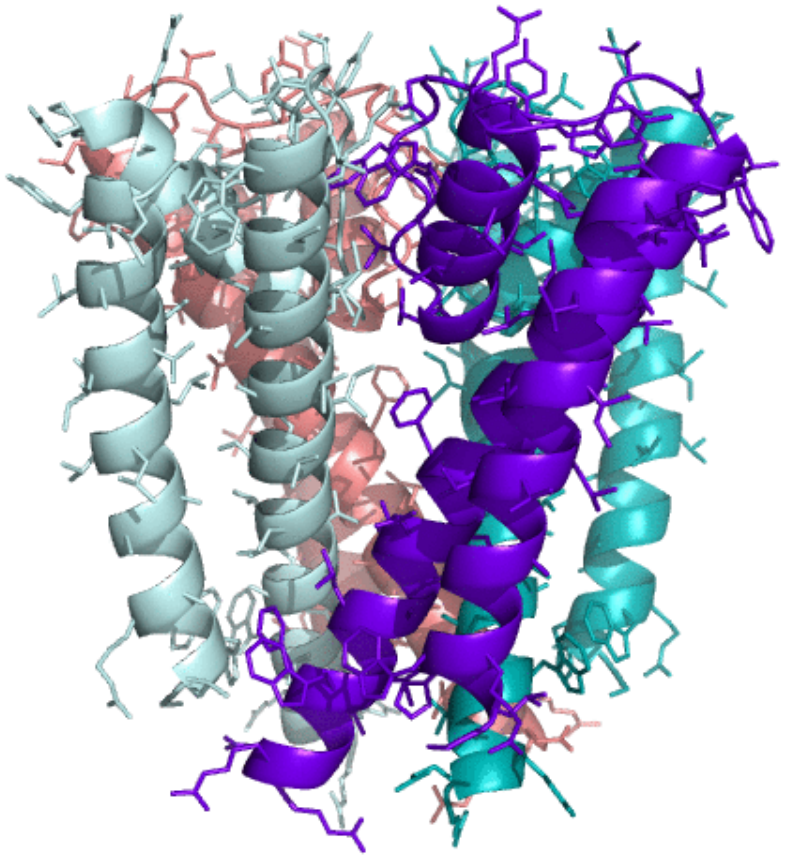
Torsion angle movements



+



=



Previous Work K^+ channels

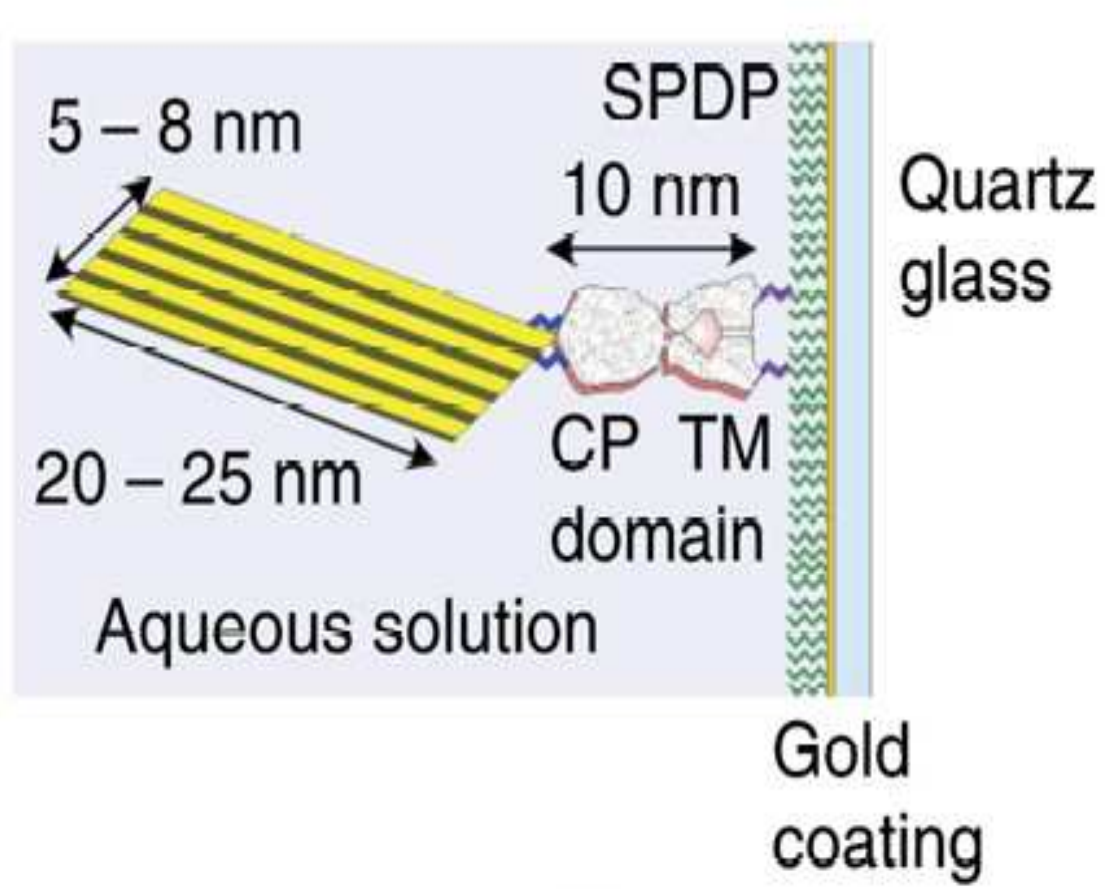
Biggin and Sansom (2002) - steered molecular dynamics.

Tikhonov and Zhorov (2004) - Monte-Carlo simulation.

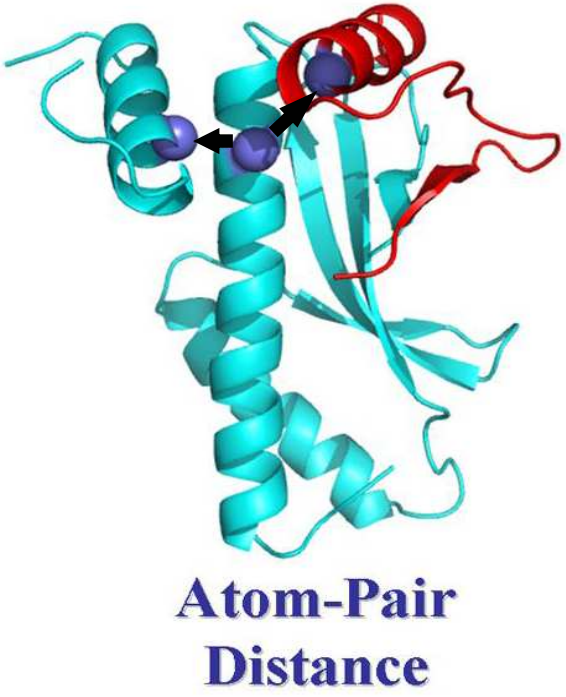
Shrivastava and Bahar (2006) - Gaussian network method.
channel opening follows a corkscrew motion of the intracellular regions of the channel.

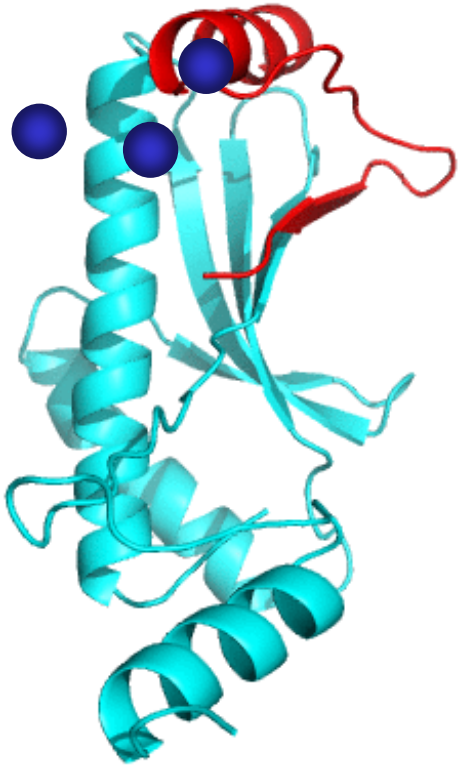
Shimizu et at. (2008) - single molecule studies.
Rotational mechanism of the intracellular ends of TM2.

Coarse Experimental Validation for Circular Motion (Shimizu et al., Cell 2007)

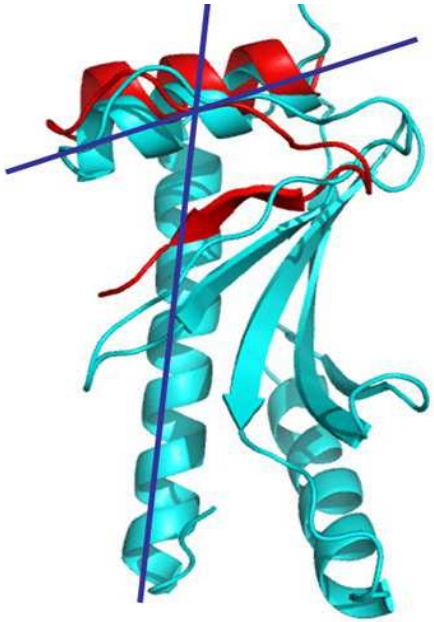


Movie – CesT simulation with Atom-Distance Constraint

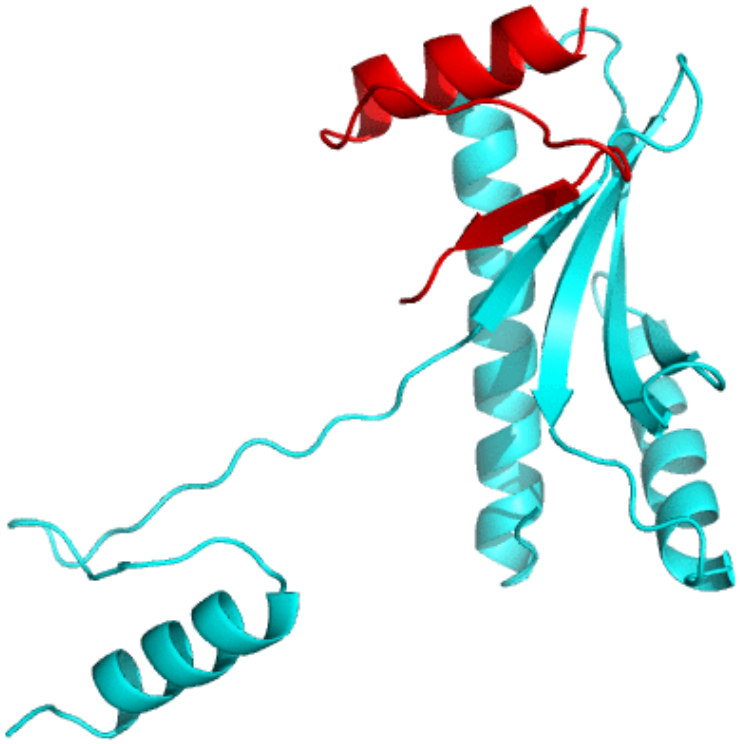




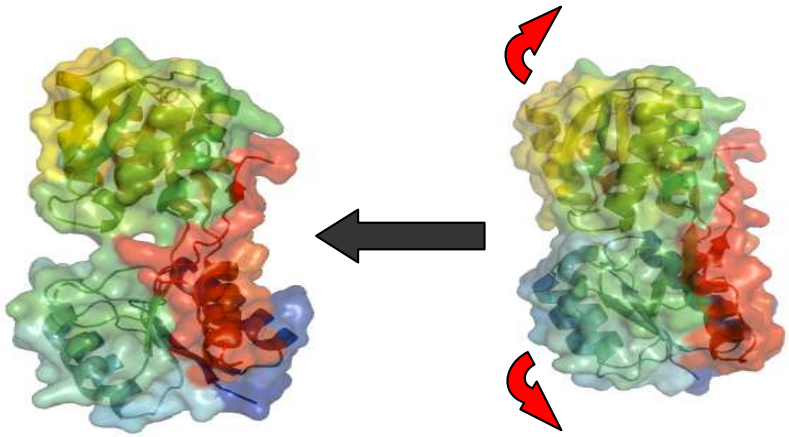
Movie – CesT simulation with Helix-Orientation Constraint

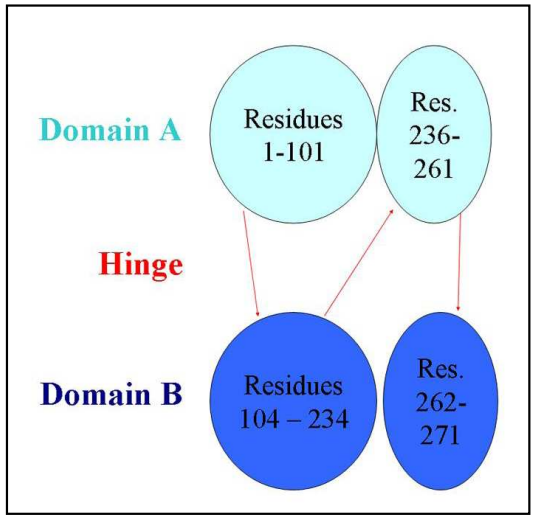


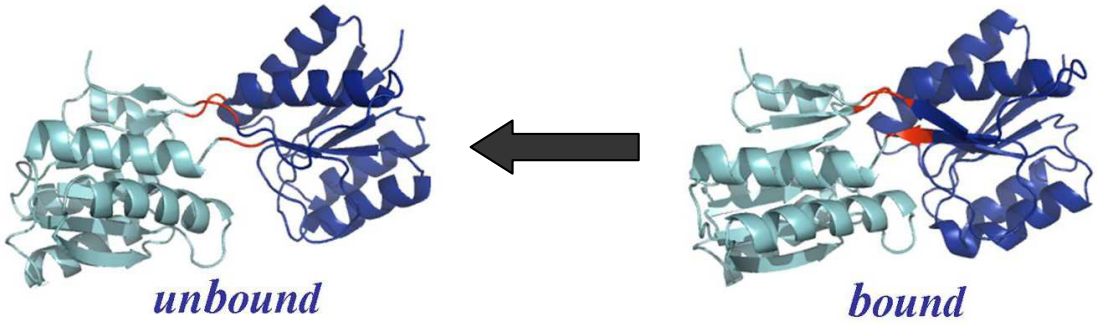
Helix-Angle and Helix-Distance



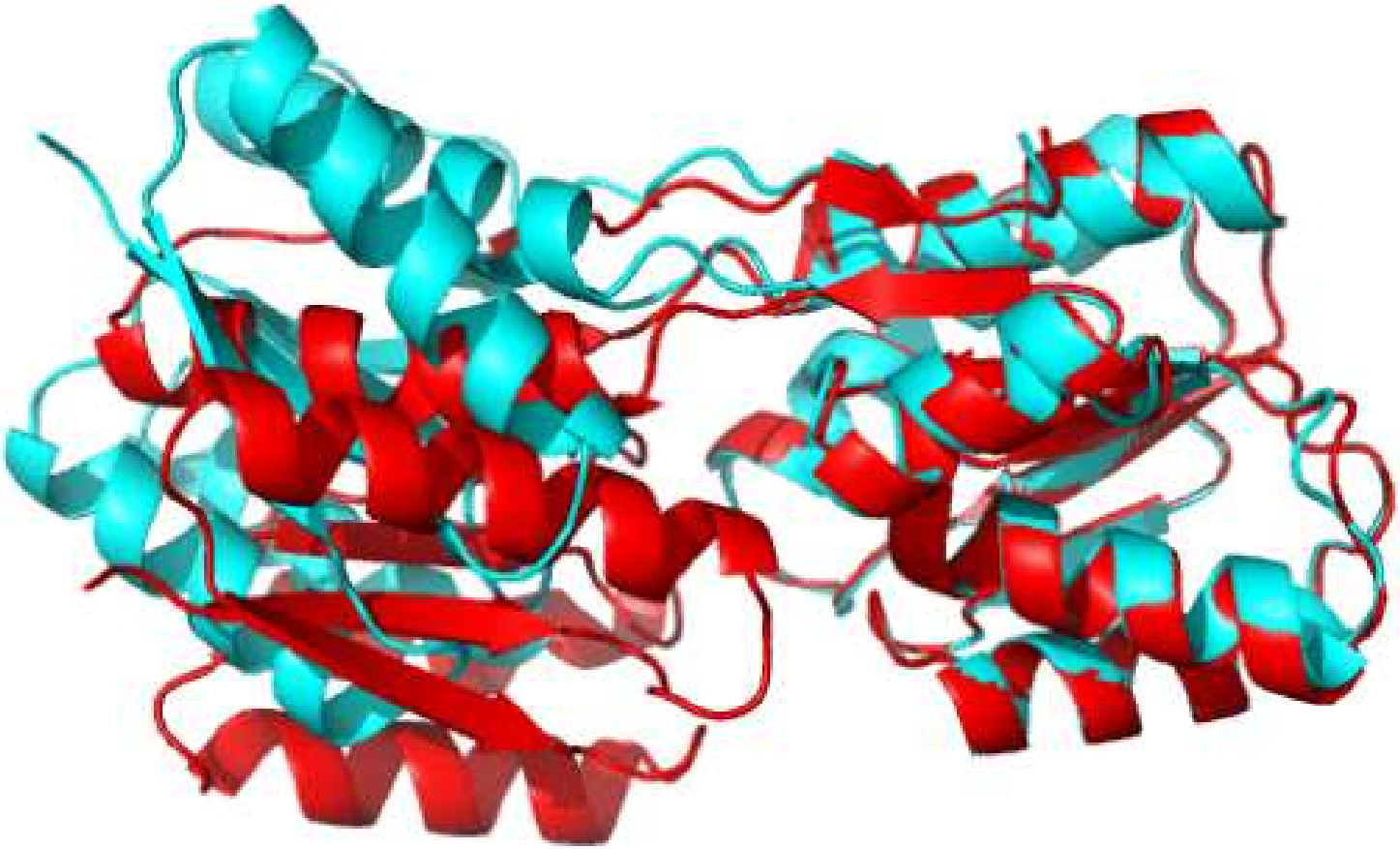
Ribose-binding Protein: Triple-Hinge = Problems?





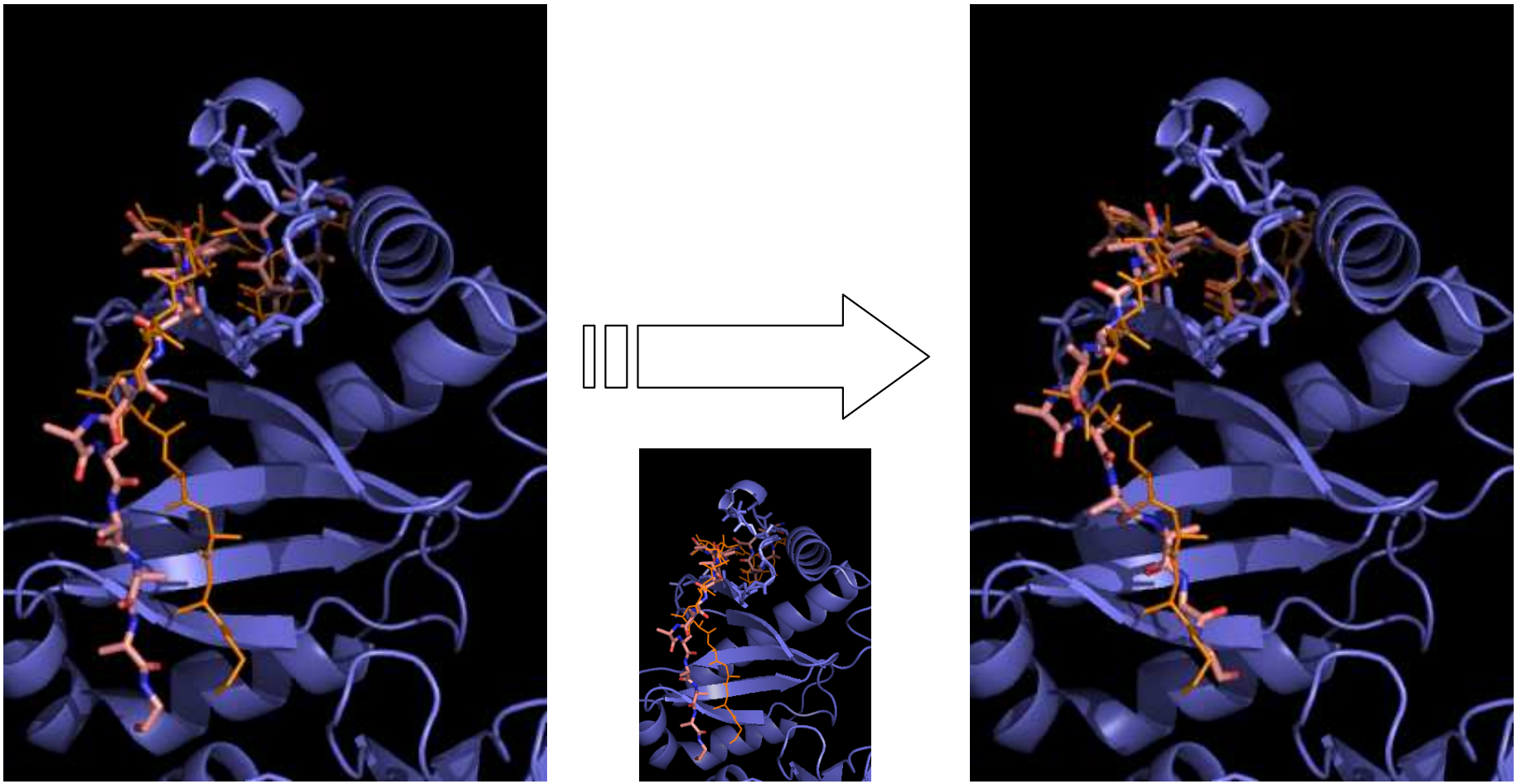


Triple Hinges – Partial Information can Force Coordinated Motion



Sampling with RRT for Peptide Docking

(See poster on peptide docking)



a

Partial Data / Evaluated on	(i)	(ii)†			(iii)	(iv)	(v)
		LLD	LLA††	CMD			
Start Conformation (*CesT*)	39.1	25.1	91°	47.2	13.8	7.8	13.8
Target Conformation (*SigE*)	6.1	8.8	279°	15.0	0	0	0
Final Conformation	6.1	8.6	275°	11.1	1.2	1.2	1.8

b

RMSD to Final Conf.	*SigE*	**6.7**	**5.3**			**3.6**	**8.2**	**3.2**
	pseudo-monomer	**8.9**	**1.9**			**2.0**	**6.4**	**1.8**

† LLD / LLA = Least Mean Square Line Distance / Angle ; CMD = Center of Mass Distance

†† LLA measured in angles