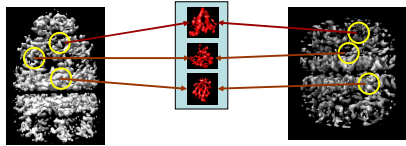


MOTIF-EM:

an Automated Computational Tool for
Identifying Conserved Domains in the CryoEM
Structures of Macromolecular Assemblies



Mitul Saha
Simbios Postdoctoral Scholar

NCMI National Center for
Macromolecular Imaging

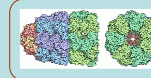
NIH Center for Biomedical
Computation at Stanford



Joint work with Michael Levitt and Wah Chiu

Large Macromolecular Assemblies (LMAs)

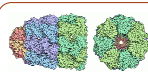
- Large macromolecular assemblies (such as ribosomes, chaperonins, viruses, etc.) critically influence essential biological processes, ranging from cell motility and signal transduction to information storage and processing.



Chaperonin: protein folding machine:
Molecular mass: ~1Mda

Large Macromolecular Assemblies (LMAs)

- Large macromolecular assemblies (such as ribosomes, chaperonins, viruses, etc.) critically influence essential biological processes, ranging from cell motility and signal transduction to information storage and processing.

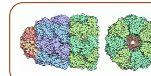


Chaperonin: protein folding machine:
Molecular mass: ~1Mda

- LMAs have well defined 3D structural forms which dictate their functional capabilities.

Large Macromolecular Assemblies (LMAs)

- Large macromolecular assemblies (such as ribosomes, chaperonins, viruses, etc.) critically influence essential biological processes, ranging from cell motility and signal transduction to information storage and processing.

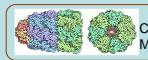


Chaperonin: protein folding machine:
Molecular mass: ~1Mda

- LMAs have well defined 3D structural forms which dictate their functional capabilities.
- Hence, detailed structural understanding of LMAs is essential for a complete understanding of cellular and systems biology.

Structure Determination of LMAs using CryoEM

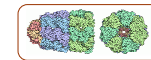
- LMAs have molecular masses in the range of 1-100 MDa, making it too difficult for conventional methods (X-ray crystallography and NMR) to determine their molecular structures.



Chaperonin: protein folding machine:
Molecular mass: ~1Mda

Structure Determination of LMAs using CryoEM

- LMAs have molecular masses in the range of 1-100 MDa, making it too difficult for conventional methods (X-ray crystallography and NMR) to determine their molecular structures.

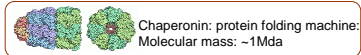


Chaperonin: protein folding machine:
Molecular mass: ~1Mda

- Fortunately, in recent years cryoEM has emerged as a single-most powerful means for determining the molecular structures of LMAs.

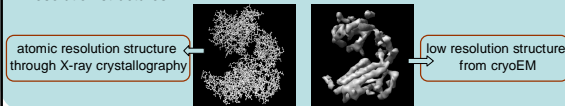
Structure Determination of LMAs using CryoEM

- LMAs have molecular masses in the range of 1-100 MDa, making it too difficult for conventional methods (X-ray crystallography and NMR) to determine their molecular structures.



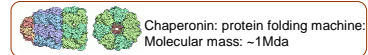
- Fortunately, in recent years cryoEM has emerged as a single-most powerful means for determining the molecular structures of LMAs.

➤ However, unlike conventional methods (X-ray and NMR), cryoEM yields low resolution structures.



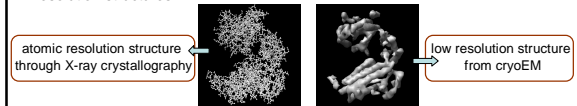
Structure Determination of LMAs using CryoEM

- LMAs have molecular masses in the range of 1-100 MDa, making it too difficult for conventional methods (X-ray crystallography and NMR) to determine their molecular structures.



- Fortunately, in recent years cryoEM has emerged as a single-most powerful means for determining the molecular structures of LMAs.

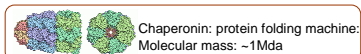
➤ However, unlike conventional methods (X-ray and NMR), cryoEM yields low resolution structures.



➤ Hence a significant portion of current cryoEM based research is focusing on building computational tools to counter this resolution gap and extract useful structural information from the low resolution cryoEM structures.

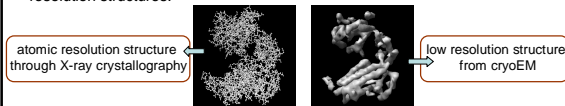
Structure Determination of LMAs using CryoEM

- LMAs have molecular masses in the range of 1-100 MDa, making it too difficult for conventional methods (X-ray crystallography and NMR) to determine their molecular structures.



- Fortunately, in recent years cryoEM has emerged as a single-most powerful means for determining the molecular structures of LMAs.

➤ However, unlike conventional methods (X-ray and NMR), cryoEM yields low resolution structures.

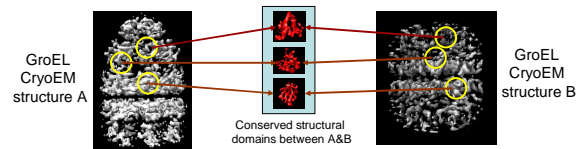


➤ Hence a significant portion of current cryoEM based research is focusing on building computational tools to counter this resolution gap and extract useful structural information from the low resolution cryoEM structures.

➤ The new tool MOTIF-EM that I will present today is one such computational tool.

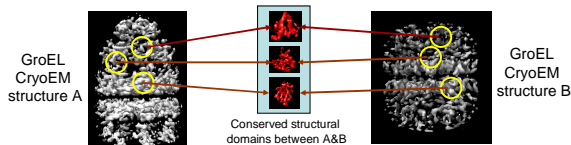
Problem Definition

➤ Find structurally conserved domains in a pair of cryoEM structures



Problem Definition

➤ Find structurally conserved domains in a pair of cryoEM structures

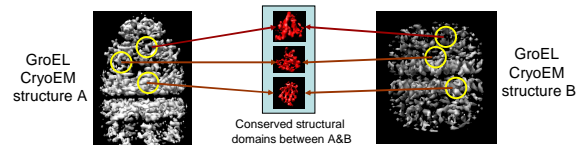


-This is a very hard computational problem: NP-complete.

Solving it exactly and efficiently is not an option.
Like for any other NP-complete problems,
here also we will try to approximately solve the problem.

Problem Definition

➤ Find structurally conserved domains in a pair of cryoEM structures



-This is a very hard computational problem: NP-complete.

Solving it exactly and efficiently is not an option.
Like for any other NP-complete problems,
here also we will try to approximately solve the problem.

- Solving this problem even approximately can help in:

- understanding molecular machineries/mechanisms, conformation changes, etc.
- propose atomic resolution models for the cryoEM structures
- build reduced articulated models for molecular dynamics simulations and meaningful morphing between known conformations
- show evolutionary relationship

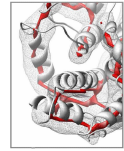
Related Work

- MOTIF-EM is the first automated computational tool to extract conserved structural domains from a pair of cryoEM maps.

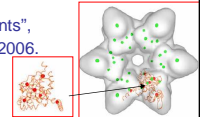
Related Work

- MOTIF-EM is the first automated computational tool to extract conserved structural domains from a pair of cryoEM maps. Other related work are represented by:

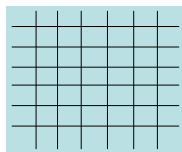
- "Identification of Secondary Structure Elements in Intermediate resolution Density Maps", M. L. Baker, T. Ju, and W. Chiu. Structure 2007.
- works for about 10% of the maps in the cryoEM structure database



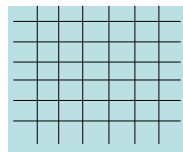
- "Multi-Resolution Anchor-Point Registration of Biomolecular Assemblies and their Components", S. Birmanns and W. Wriggers. J. Struc. Biol., 2006.
- need to know the atomic coordinates of the domain in advance



Conserved domain extraction method (MOTIF-EM) outline:



Input cryoEM structure grid #1

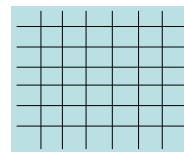


Input cryoEM structure grid #2

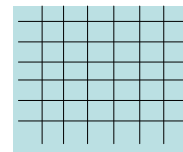
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point

Map pre-processing I: assigning local coordinate frames



Input cryoEM structure grid #1

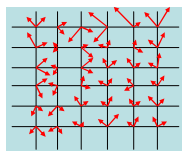


Input cryoEM structure grid #2

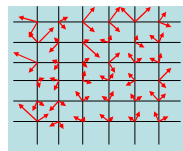
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point

Map pre-processing I: assigning local coordinate frames



Input cryoEM structure grid #1



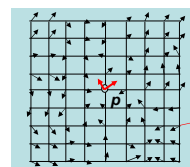
Input cryoEM structure grid #2

-principal direction along largest local density variation

Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor

Map pre-processing II: build local descriptors



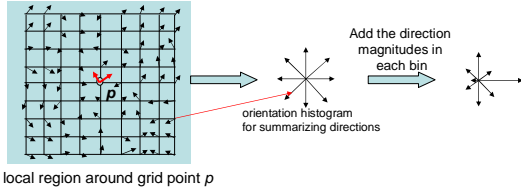
local region around grid point p



Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor

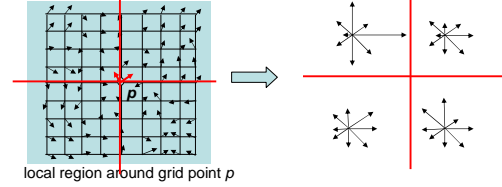
Map pre-processing II: build local descriptors



Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor

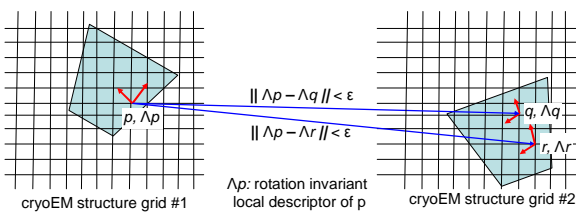
Map pre-processing II: build local descriptors



*** Adapted from SIFT technique for object recognition in computer vision [Lowe'04, J. Computer Vision]

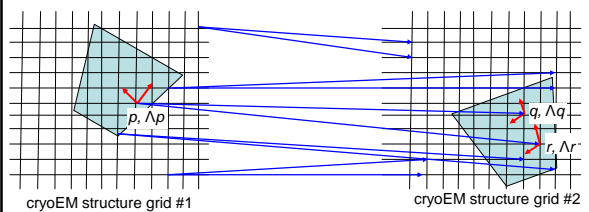
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor
- For each grid point find k matches in the other structure with similar local descriptor



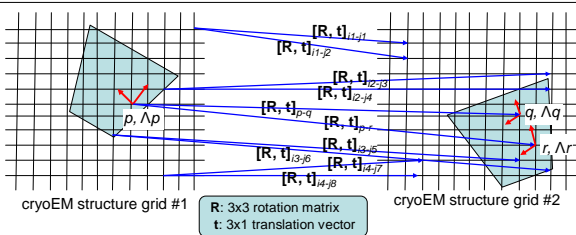
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor
- For each grid point find k matches in the other structure with similar local descriptor



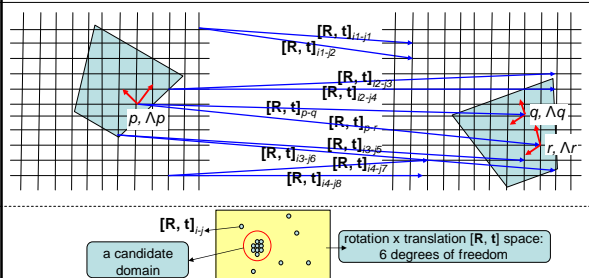
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor
- For each grid point find k matches in the other structure with similar local descriptor
- Cluster the matches using their respective transformations.



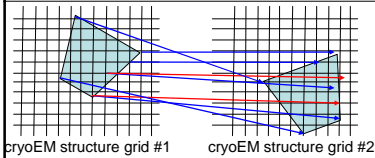
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
- For each grid point construct a rotationally invariant local descriptor
- For each grid point find k matches in the other structure with similar local descriptor
- Cluster the matches using their respective transformations.



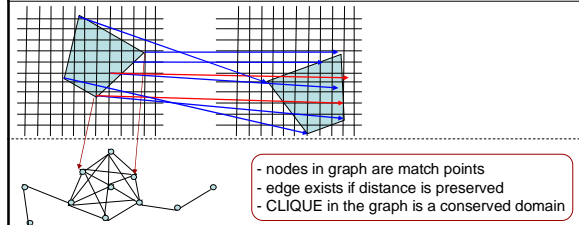
Conserved domain extraction method (MOTIF-EM) outline:

- Assign a coordinate frame to each grid point
 - For each grid point construct a rotationally invariant local descriptor
 - For each grid point find k matches in the other structure with similar local descriptor
 - Cluster the matches using their respective transformations.
- **The prominent cluster contains the conserved domain and false positives.**
The conserved domain are the matches with preserved inter-point distances .

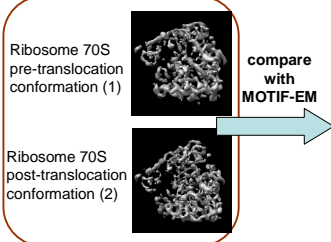


Conserved domain extraction method (MOTIF-EM) outline:

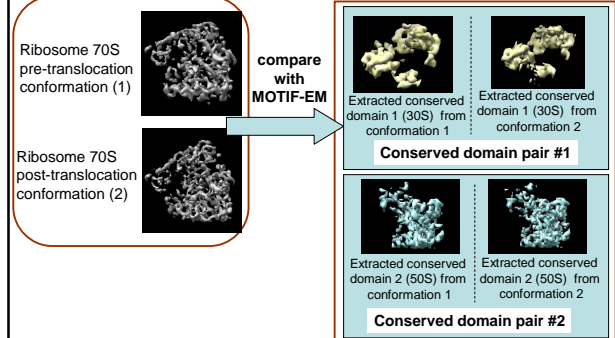
- Assign a coordinate frame to each grid point
 - For each grid point construct a rotationally invariant local descriptor
 - For each grid point find k matches in the other structure with similar local descriptor
 - Cluster the matches using their respective transformations.
- **The prominent cluster contains the conserved domain and false positives.**
The conserved domain are the matches with preserved inter-point distances .



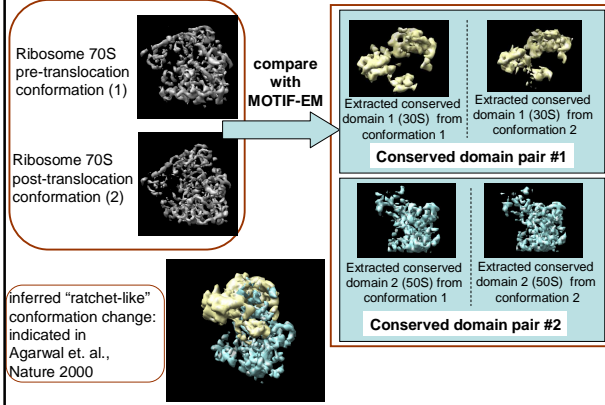
Inferring Conformation Change



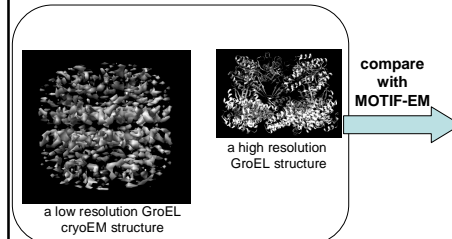
Inferring Conformation Change

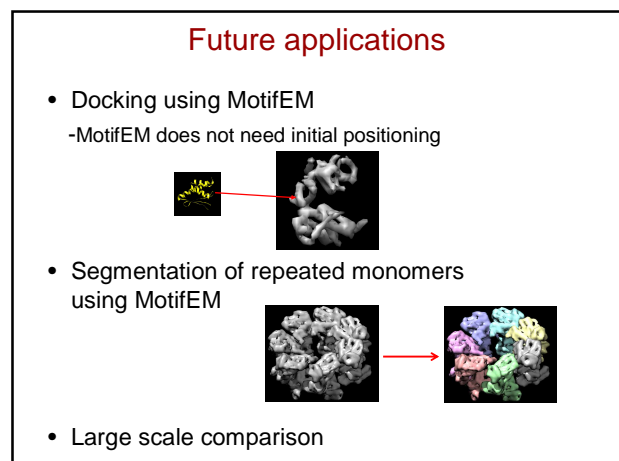
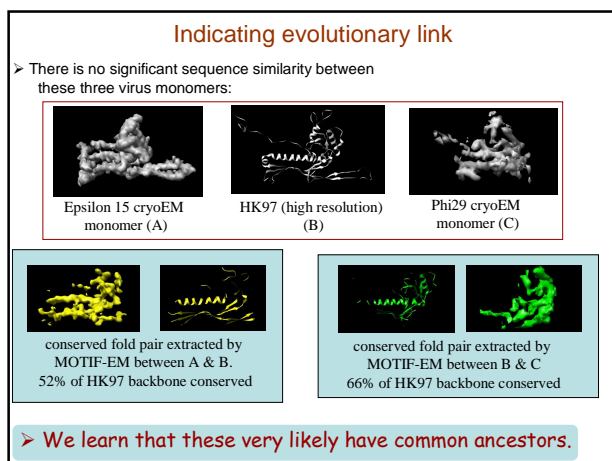
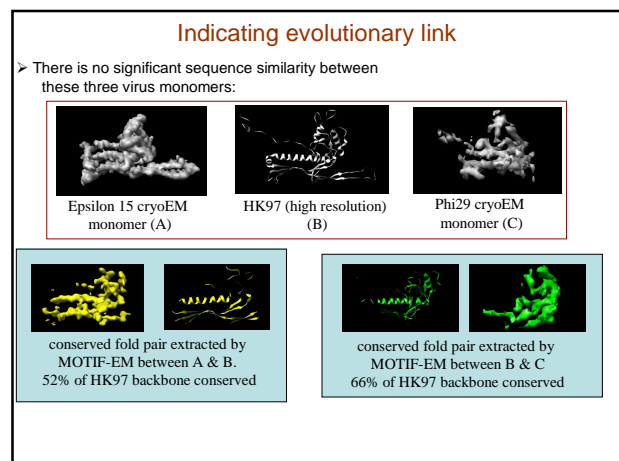
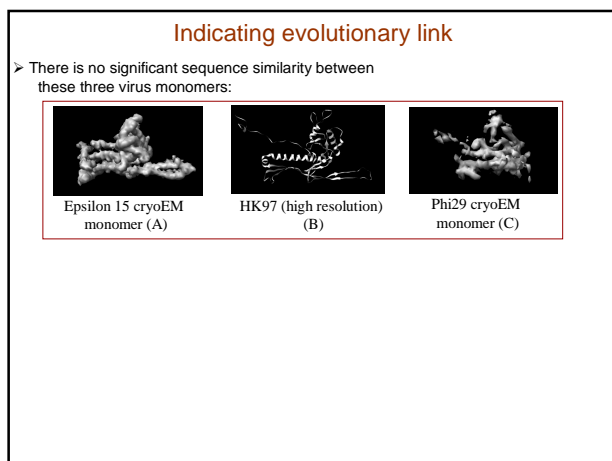
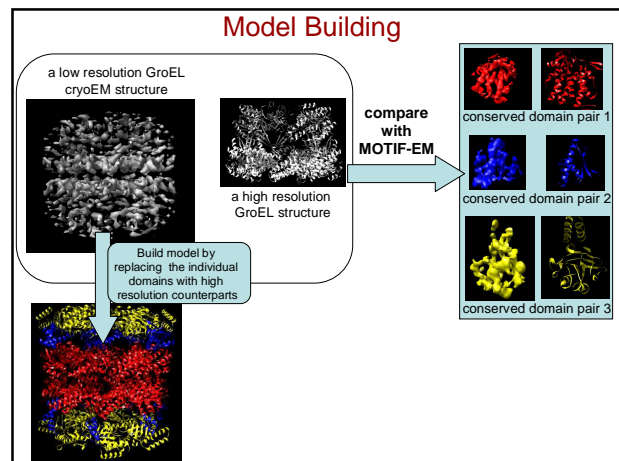
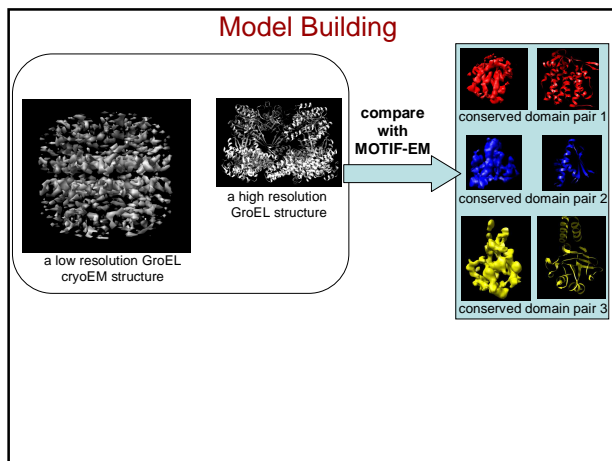


Inferring Conformation Change



Model Building





Conclusion

- CryoEM has emerged as a single-most powerful tool for determining structures of large molecular assemblies (ribosomes, chaperonins, viruses, etc.) which critically influence essential biological processes.

Conclusion

- CryoEM has emerged as a single-most powerful tool for determining structures of large molecular assemblies (ribosomes, chaperonins, viruses, etc.) which critically influence essential biological processes.
- However molecular structures from CryoEM have low resolution, compared to conventional methods (X-ray/NMR).

Conclusion

- CryoEM has emerged as a single-most powerful tool for determining structures of large molecular assemblies (ribosomes, chaperonins, viruses, etc.) which critically influence essential biological processes.
- However molecular structures from CryoEM have low resolution, compared to conventional methods (X-ray/NMR).
- A significant portion of current CryoEM research is about building new computational methods for extracting useful structural information from the low resolution cryoEM structures.
 - MOTIF-EM, presented here, is one such tool. Specifically it finds conserved structural domains between two cryoEM maps.
 - MOTIF-EM is the first automated computational tool to do so.

Conclusion

- CryoEM has emerged as a single-most powerful tool for determining structures of large molecular assemblies (ribosomes, chaperonins, viruses, etc.) which critically influence essential biological processes.
- However molecular structures from CryoEM have low resolution, compared to conventional methods (X-ray/NMR).
- A significant portion of current CryoEM research is about building new computational methods for extracting useful structural information from the low resolution cryoEM structures.
 - MOTIF-EM, presented here, is one such tool. Specifically it finds conserved structural domains between two cryoEM maps.
 - MOTIF-EM is the first automated computational tool to do so.
- In addition to finding conserved domains, MOTIF-EM can be used to:
 - infer conformation change towards understanding molecular machineries/mechanisms, etc.
 - propose atomic-resolution models for the low resolution cryoEM structures
 - segment repeated monomers
 - confirm evolutionary links

Conclusion

- CryoEM has emerged as a single-most powerful tool for determining structures of large molecular assemblies (ribosomes, chaperonins, viruses, etc.) which critically influence essential biological processes.
- However molecular structures from CryoEM have low resolution, compared to conventional methods (X-ray/NMR).
- A significant portion of current CryoEM research is about building new computational methods for extracting useful structural information from the low resolution cryoEM structures.
 - MOTIF-EM, presented here, is one such tool. Specifically it finds conserved structural domains between two cryoEM maps.
 - MOTIF-EM is the first automated computational tool to do so.
- In addition to finding conserved domains, MOTIF-EM can be used to:
 - infer conformation change towards understanding molecular machineries/mechanisms, etc.
 - propose atomic-resolution models for the low resolution cryoEM structures
 - segment repeated monomers
 - confirm evolutionary links
- MOTIF-EM would be soon available as a "CryoEM Map Processing Toolkit" at: <http://simtk.org/>

Acknowledgement

- Collaborators:
 - Simbios: Michael Levitt, Gunnar Schroeder, Emilio Rodriguez Antunez
 - Baylor College of Medicine: Wah Chiu, Steve Ludtke, Matthew Baker
 - UTMB: Marc Morais (Phi29 maps)
- Research Group:



- Funding: 
- Friends and family
- Thank you all for your kind attention!

