Los Alamos NATIONAL LABORATORY

Molecular Dynamics Simulation Methods for Macromolecular Crystallography

David Wych
CCS-3, CNLS

OpenEye Cup XXI | 3/17/22

LA-UR-22-22059

Molecular Dynamics Simulation Methods for Macromolecular Crystallography

Background

- *Me*
- *Crystallography*
- *Crystalline Molecular Dynamics*

Introduction and Methods

- *Previous Work*
- *MD-MX Analysis Procedure*
- *Methods*

Results

- *Density Comparison*
- *Water Building*
- *Protein Remodeling*

Discussion

- *Implications*
- *Insights*
- *Future directions*

Appreciation

Background: me

Orbiting OpenEye since ~2012

- ***B.A.* — Physics ’15 CMC**
  - Paul Nerenberg (now @ CSLA)
- ***Ph.D.* — Pharmaceutical Sciences ’21 UCI**
  - David Mobley
  - Mike Wall
- ***Post-doc* — LANL (CCS-3/CNLS)**
  - Crystallography + Diffuse Scattering
  - Exascale Computing
  - Full-protein + solvent QM/MM

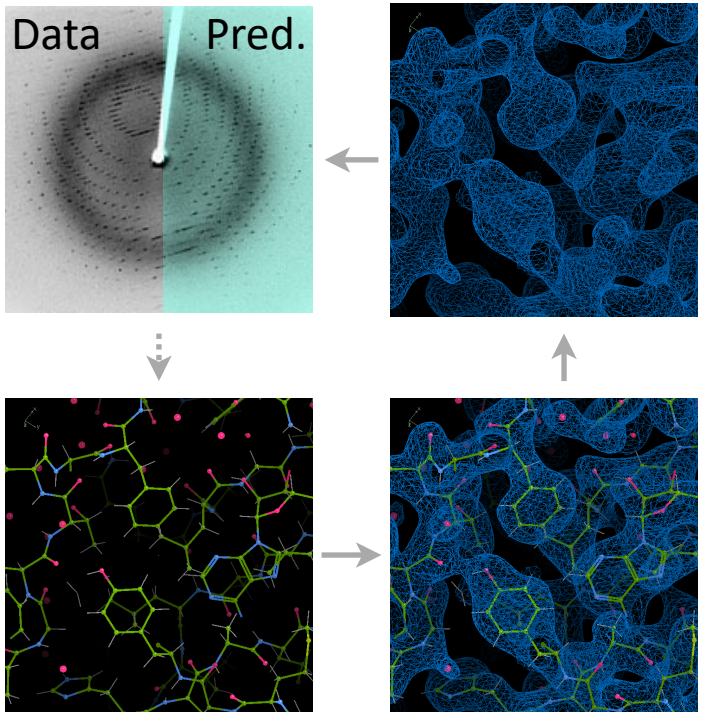
Claremont McKenna College

UCI University of California, Irvine

Los Alamos National Laboratory

Macromolecular Crystallography (MX): “all models are wrong…”

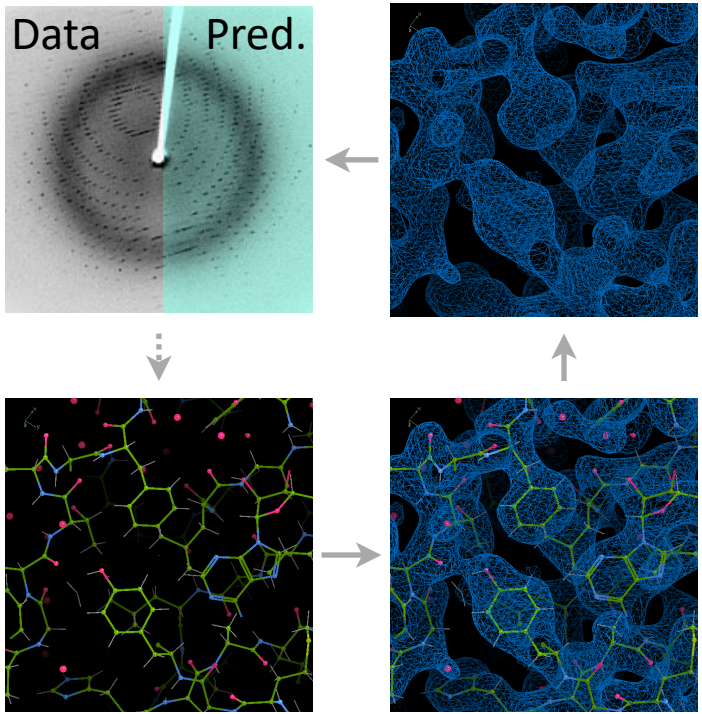
- Responsible for most known structures of proteins



https://news.cancerresearchuk.org/2014/12/23/the-shape-of-all-things-the-year-of-crystallography/

Macromolecular Crystallography (MX): “all models are wrong…”

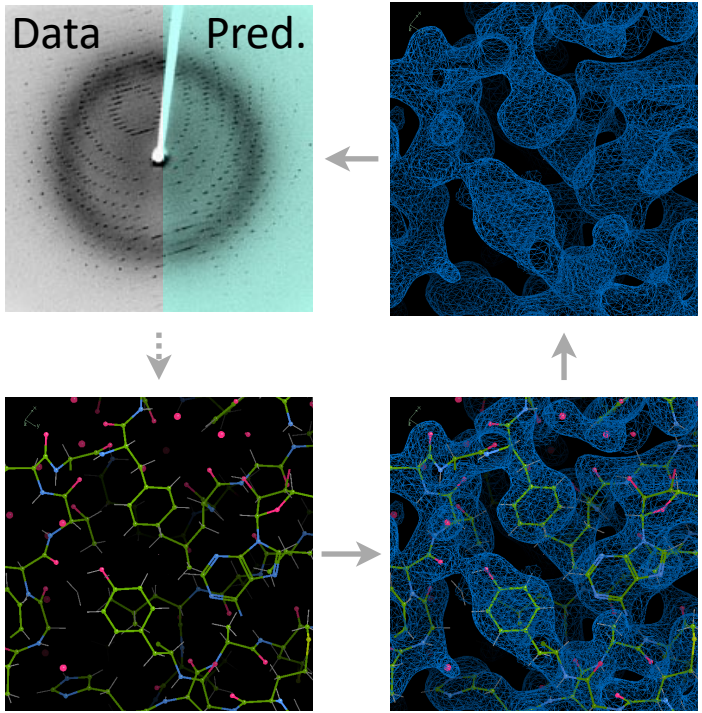
- Responsible for most known structures of proteins
- Models have **historically been single structure** but there are now various ways of incorporating **structural heterogeneity**
  - Multi-conformer modeling
  - Ensemble refinement



https://news.cancerresearchuk.org/2014/12/23/the-shape-of-all-things-the-year-of-crystallography/

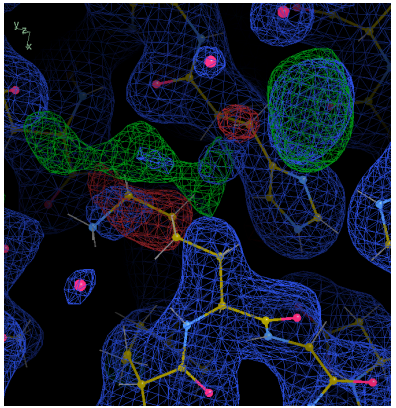
Macromolecular Crystallography (MX): “all models are wrong…”

- Responsible for most known structures of proteins
- Models have **historically been single structure** but there are now various ways of incorporating **structural heterogeneity**
  - Multi-conformer modeling
  - Ensemble refinement
- **Important to remember:** *It’s just a **model***
  - *Data doesn’t come with labels*
  - Can fail in regions where density is **difficult to interpret** (protein-solvent interface)



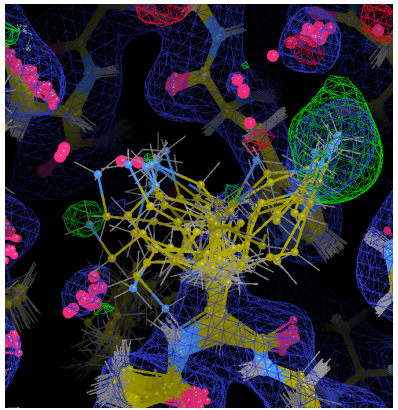
Macromolecular Crystallography (MX): “all models are wrong…”

- Responsible for most known structures of proteins
- Models have **historically been single structure** but there are now various ways of incorporating **structural heterogeneity**
  - Multi-conformer modeling
  - Ensemble refinement
- **Important to remember:** *It’s just a **model***
  - *Data doesn’t come with labels*
  - Can fail in regions where density is **difficult to interpret** (protein-solvent interface)



Single Structure Refinement

- 2Fo-Fc (1σ)
- Fo-Fc (3σ)
  - pos.
  - neg.



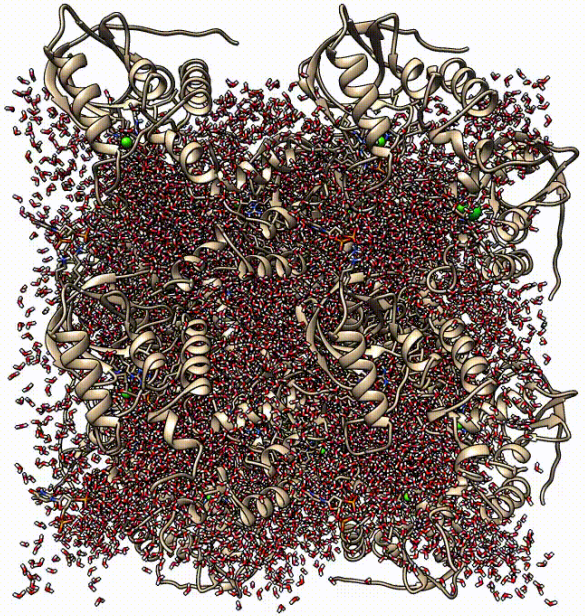
Ensemble Refinement

- 2Fo-Fc (1σ)
- Fo-Fc (3σ)
  - pos.
  - neg.

How can MD help in crystallographic modeling?

- Look enormous… **Why crystalline?** (why not just do solution single protein?)
  - Allows for direct comparison with **crystallographic data**
  - **Efficiency** and **sampling**
- Crystalline MD methods have improved substantially, in large part thanks to studies of **diffuse scattering**
- MD simulations can deviate from the crystal structure if not **restrained**

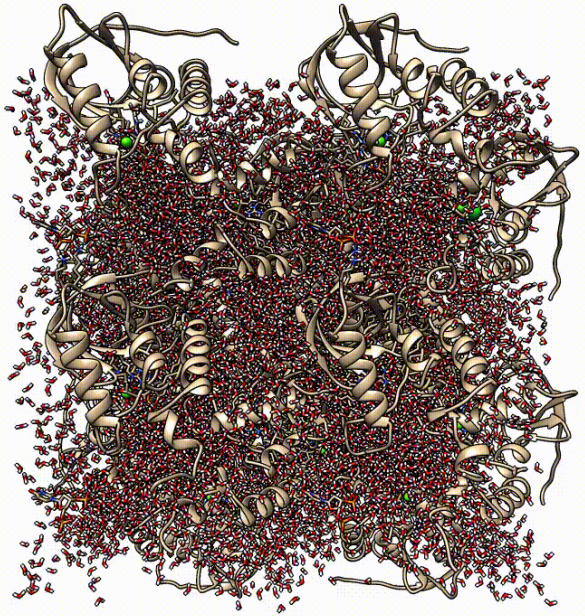
Crystalline Molecular Dynamics (MD)



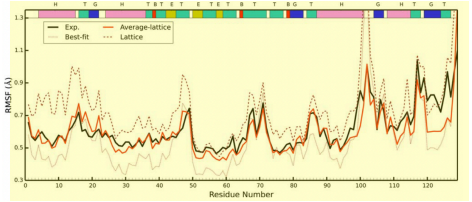
How can MD help in crystallographic modeling?

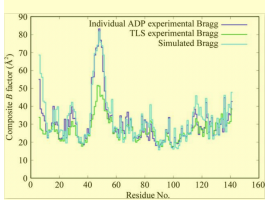
- Look enormous… **Why crystalline?** (why not just do solution single protein?)
  - Allows for direct comparison with **crystallographic data**
  - **Efficiency** and **sampling**
- Crystalline MD methods have improved substantially, in large part thanks to studies of **diffuse scattering**
- MD simulations can deviate from the crystal structure if not **restrained**

Crystalline Molecular Dynamics (MD)

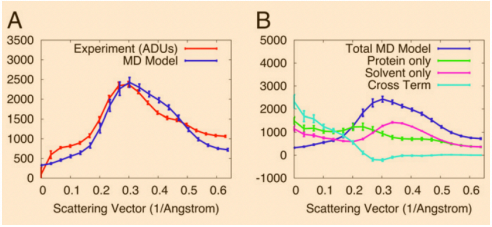


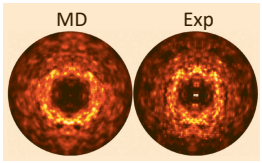
Previous work: *to restrain or not to restrain?*





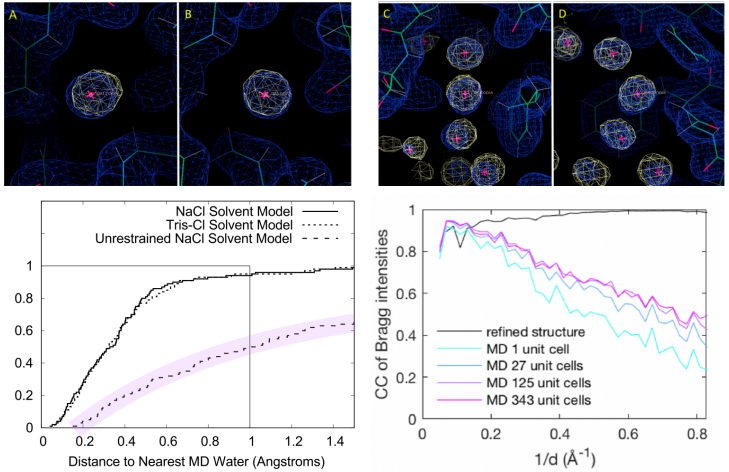
B-factors





Diffuse Scattering

Janowski, P.A., et al.*Prot. Sci.*(2016); Wall, M.E., *IUCrJ* 5.2 (2018); Wall, M.E. *et al.*, PNAS (2014)



Wall *et al.*, JACS (2019); Meisburger *et al.*, Nat. Comm. (2020)

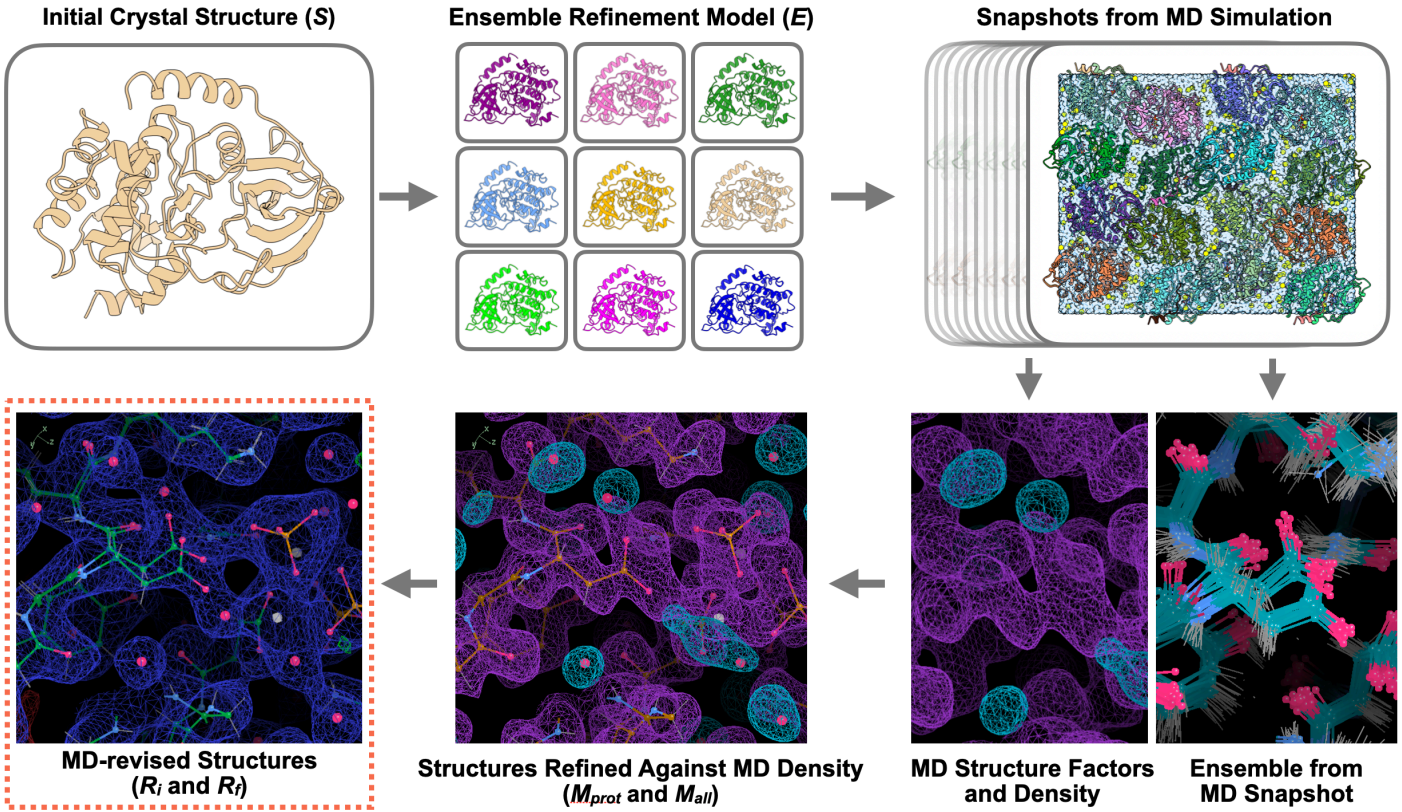
Dynamics

- **Unrestrained simulations**
  - Excellent reproduction of **B-factors**
  - Reproduction and contextualization of the **"solvent ring"**
  - Reproducing **anisotropic diffuse scattering**

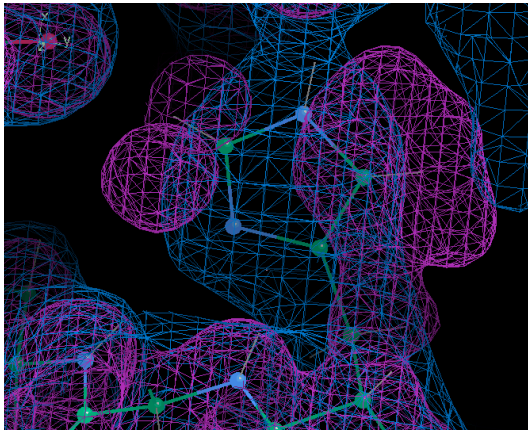
Average Structure

- **Restrained Simulations:**
  - Excellent reproduction of **ordered water structure**
- **Consistent degradation** ("**drift**") of the crystal lattice when restraints are relaxed

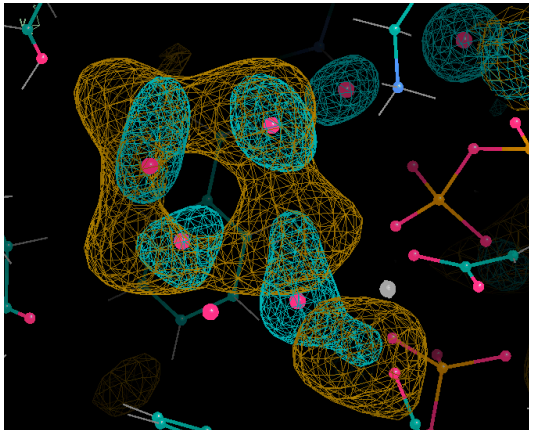
The MD-MX Analysis Procedure



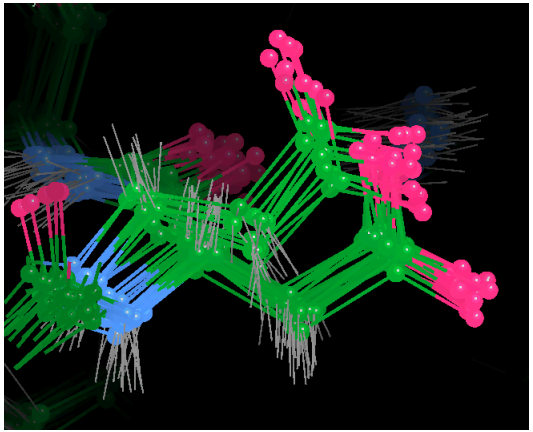
Methods to improve interpretation and modeling of crystallographic density



Density Comparison



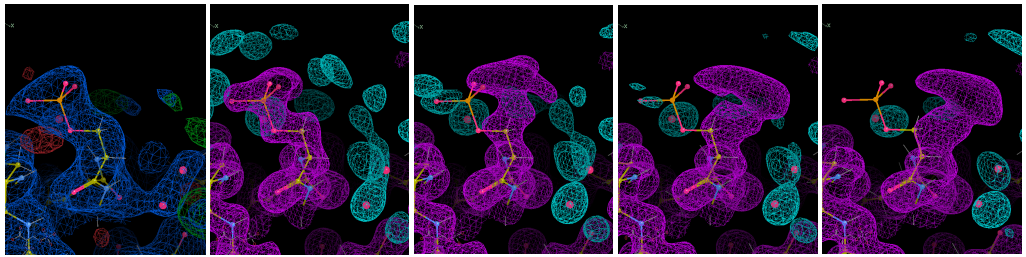
Water Building / Solvent Modeling

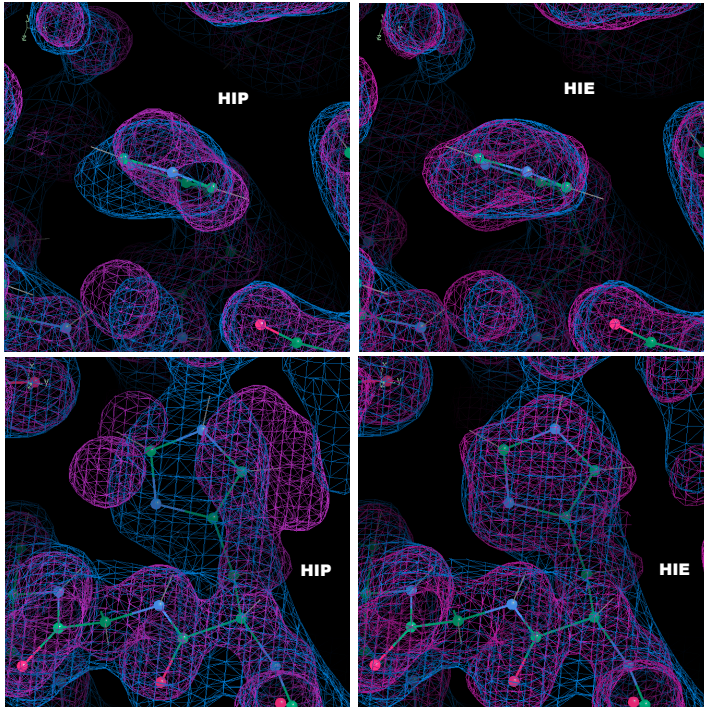


Protein Re-modelling

Density comparison for making direct contact with experimental data

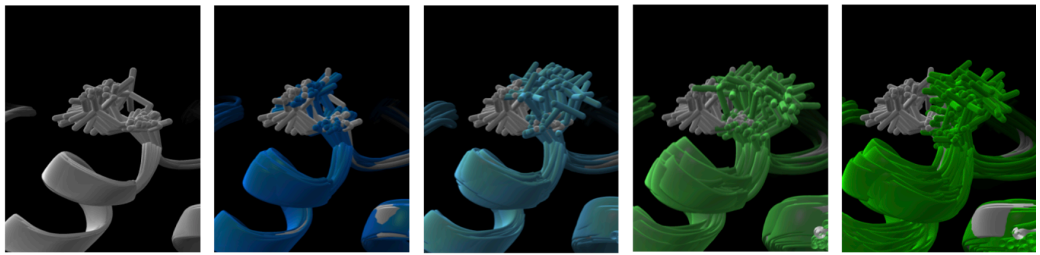
- Calculate the structure factors/density from **any component** of the system
  - **Protein, water, ions, etc.**

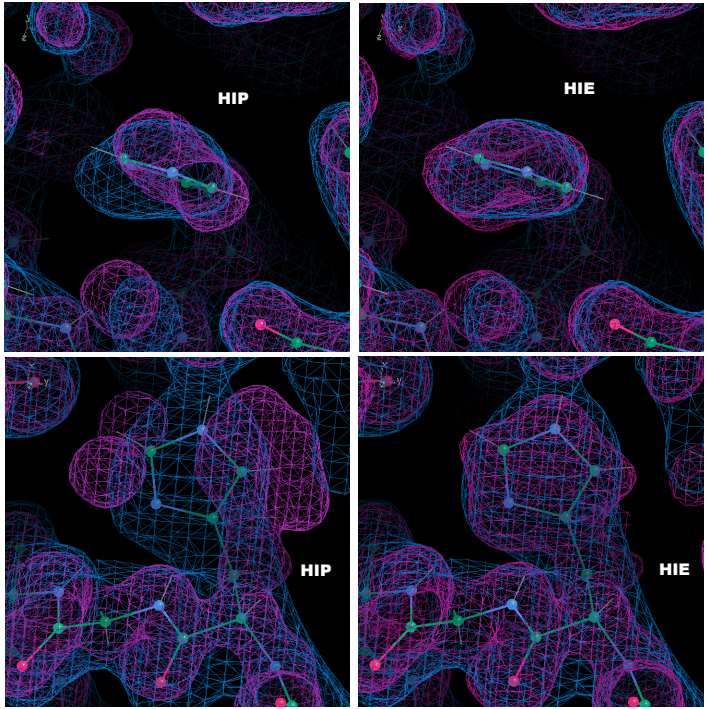




Density comparison for making direct contact with experimental data

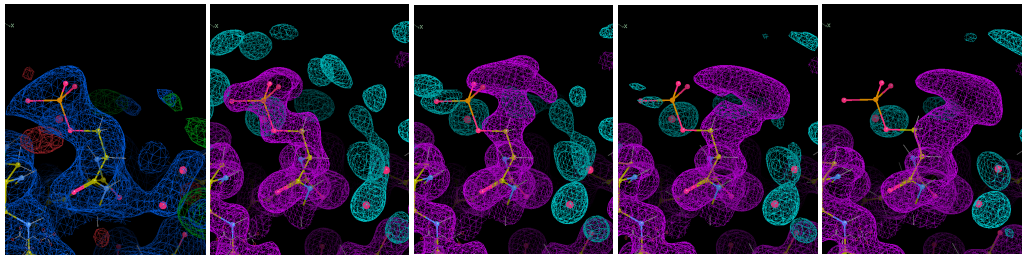
- Calculate the structure factors/density from **any component** of the system
  - **Protein, water, ions, etc.**
- **Average picture** over an otherwise inscrutable ensemble of structures

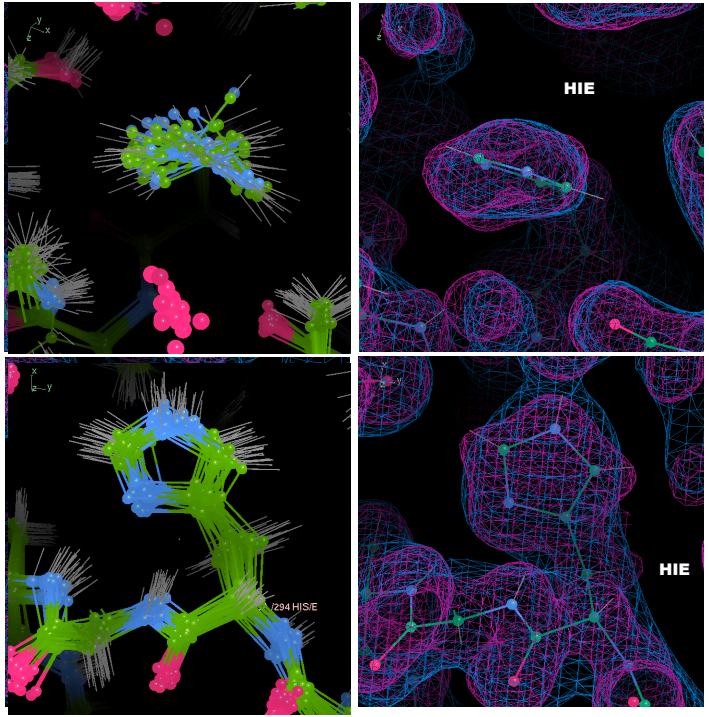




Density comparison for making direct contact with experimental data

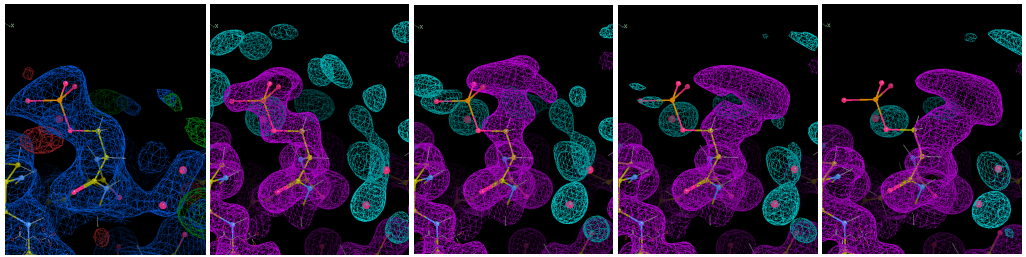
- Calculate the structure factors/density from **any component** of the system
  - **Protein, water, ions, etc.**
- **Average picture** over an otherwise inscrutable ensemble of structures

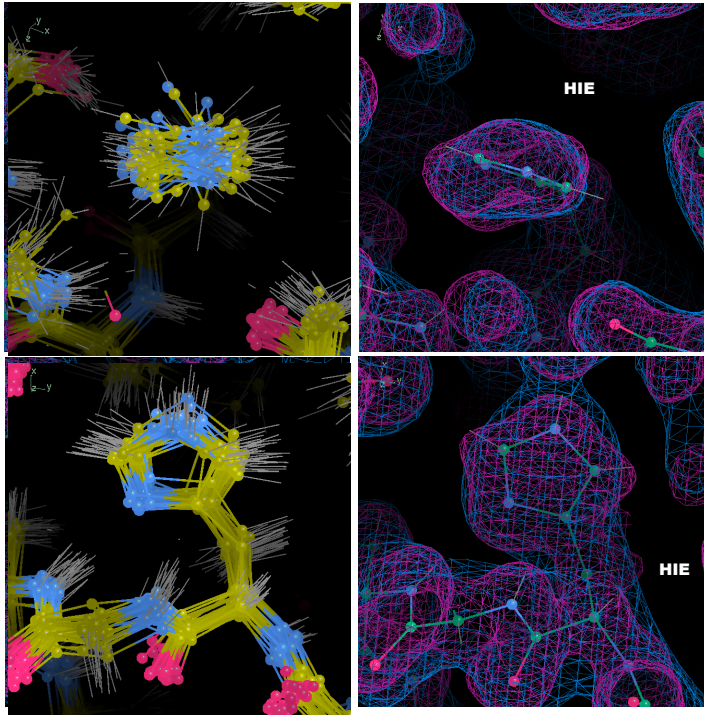




Density comparison for making direct contact with experimental data

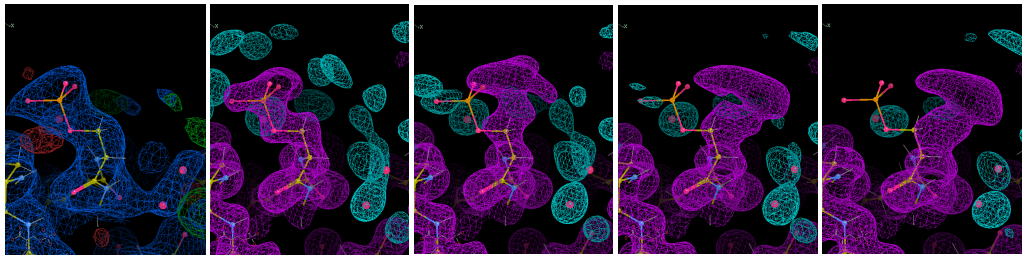
- Calculate the structure factors/density from **any component** of the system
  - **Protein, water, ions, etc.**
- **Average picture** over an otherwise inscrutable ensemble of structures

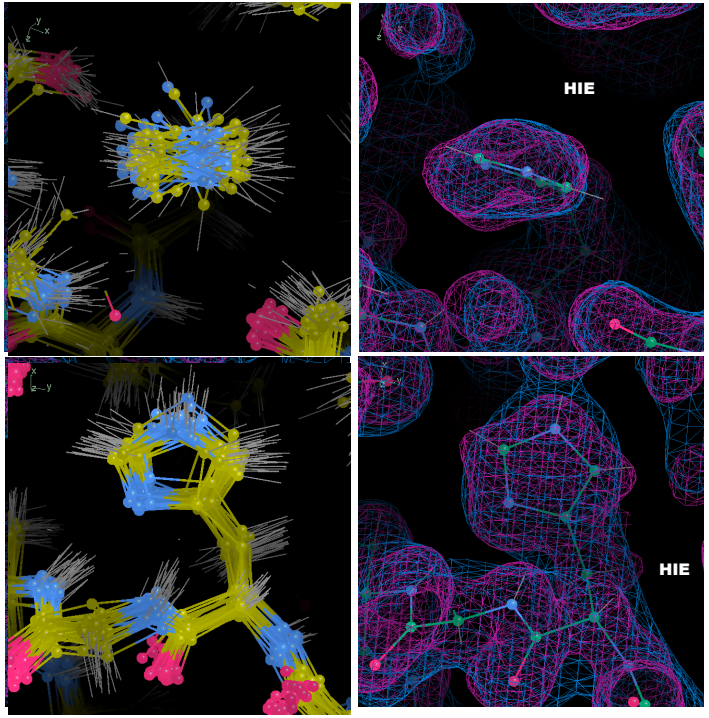




Density comparison for making direct contact with experimental data

- Calculate the structure factors/density from **any component** of the system
  - **Protein, water, ions, etc.**
- **Average picture** over an otherwise inscrutable ensemble of structures
- Spot **sensitivity to protonation**, **"sketchy" parameters**, and more

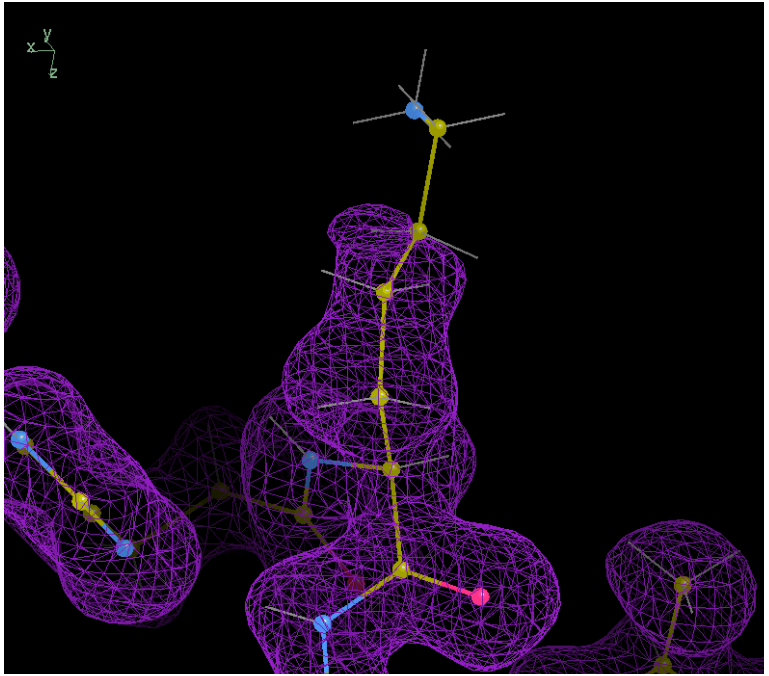




Water building for generating an alternative solvent model

Leveraging the separability of the MD density

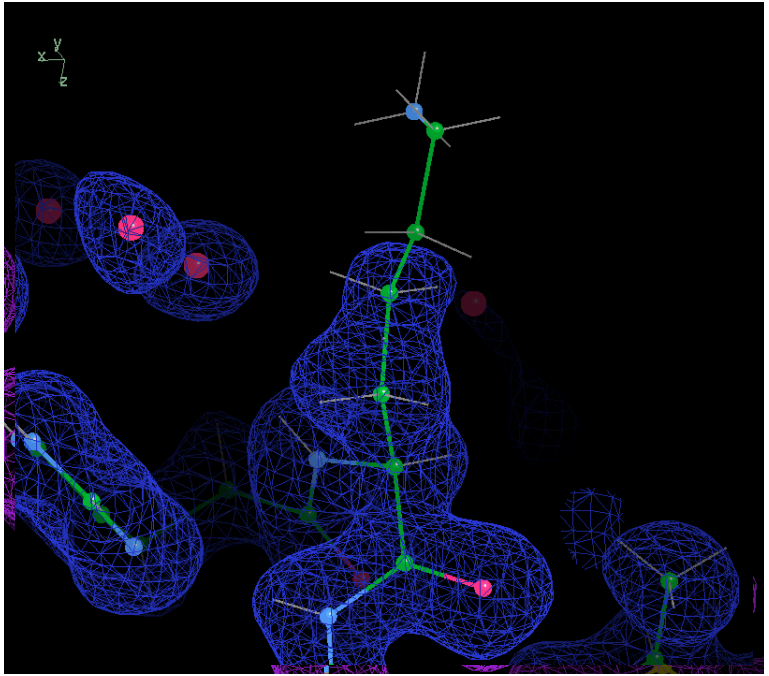
- **Protein-first refinement**
  - Refine against **non-solvent MD density**



Water building for generating an alternative solvent model

Leveraging the separability of the MD density

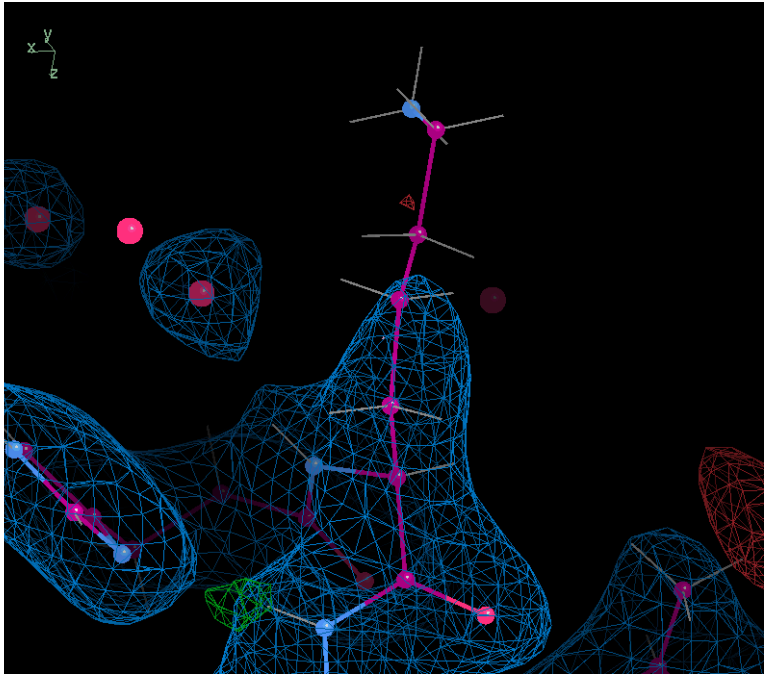
- **Protein-first refinement**
  - Refine against **non-solvent MD density**
  - Refine against **total MD density**
    - With water picking



Water building for generating an alternative solvent model

Leveraging the separability of the MD density

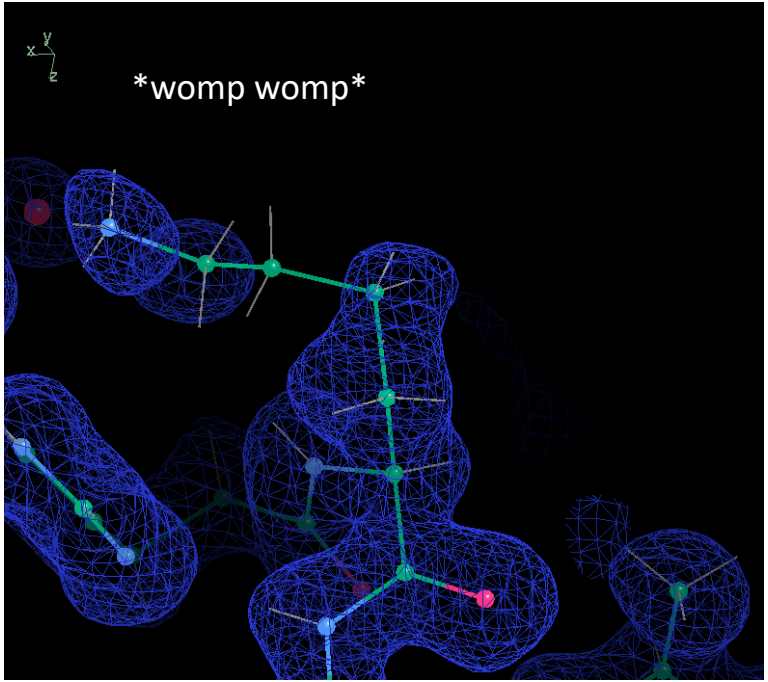
- **Protein-first refinement**
  - Refine against **non-solvent MD density**
  - Refine against **total MD density**
    - With water picking
  - Refine against **experimental data**



Water building for generating an alternative solvent model

Leveraging the separability of the MD density

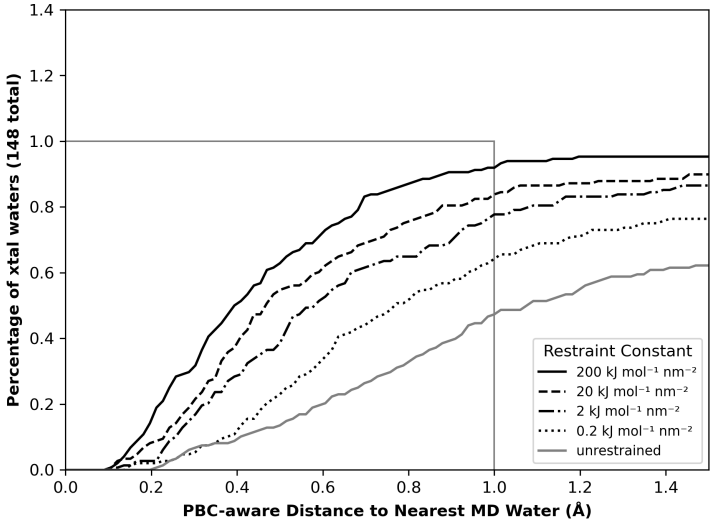
- **Protein-first refinement**
  - Refine against **non-solvent MD density**
  - Refine against **total MD density**
    - With water picking
  - Refine against **experimental data**
- Some obvious problems with refining against ***total* structure factors**



Water building for generating an alternative solvent model

Good agreement with experimentally-refined model and many more waters

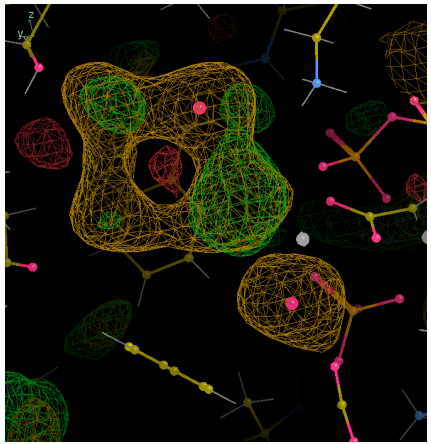
- Reproduces **92%** of waters from single structure refinement to **within 1Å**
  - Recovery degrades progressively as restraints are relaxed



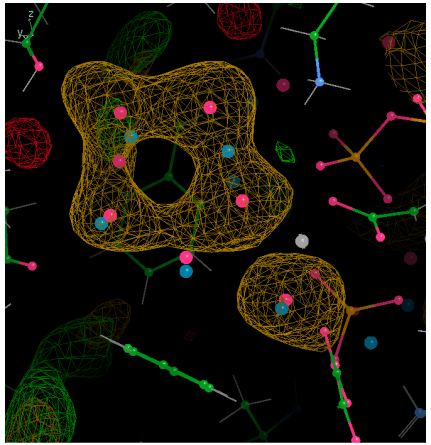
Water building for generating an alternative solvent model

Good agreement with experimentally-refined model and many more waters

- Reproduces **92%** of waters from single structure refinement to **within 1Å**
  - Recovery degrades progressively as restraints are relaxed
- **Many additional waters**
  - Around the active site, additional waters are supported by Polder density and XR-ND data
  - **Many beyond first hydration shell**



Single Structure Refinement

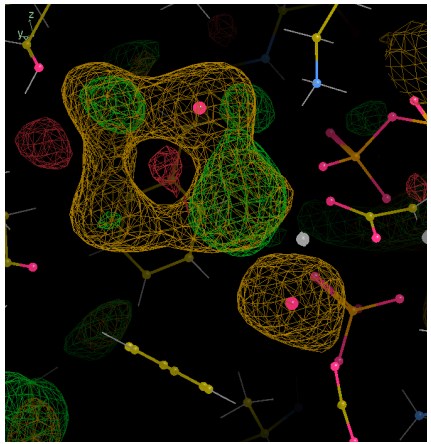


Protein-first Refinement

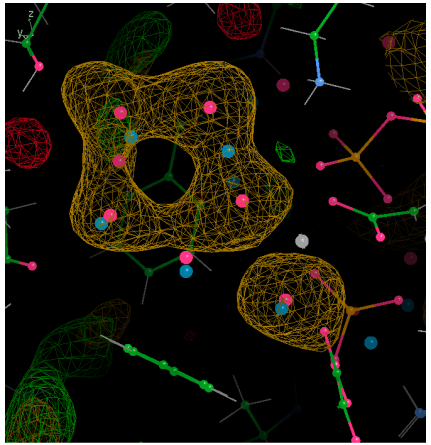
Water building for generating an alternative solvent model

Good agreement with experimentally-refined model and many more waters

- Reproduces **92%** of waters from single structure refinement to **within 1Å**
  - Recovery degrades progressively as restraints are relaxed
- **Many additional waters**
  - Around the active site, additional waters are supported by Polder density and XR-ND data
  - **Many beyond first hydration shell**
- **… and lower R-factors**
  - Just from a pass through refinement against the MD data



Single Structure Refinement



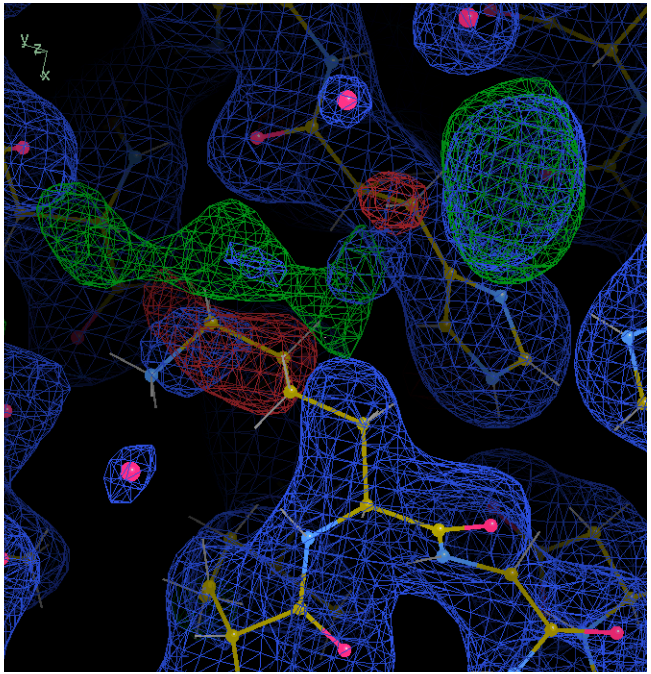
Protein-first Refinement

	R-work	R-free
Single Structure Refinement	0.1557	0.1838
Protein-first Refinement	0.1328	0.1777

Protein remodeling for making sense of uncertain density

Modeling ordered solvent misinterpreted by ensemble refinement

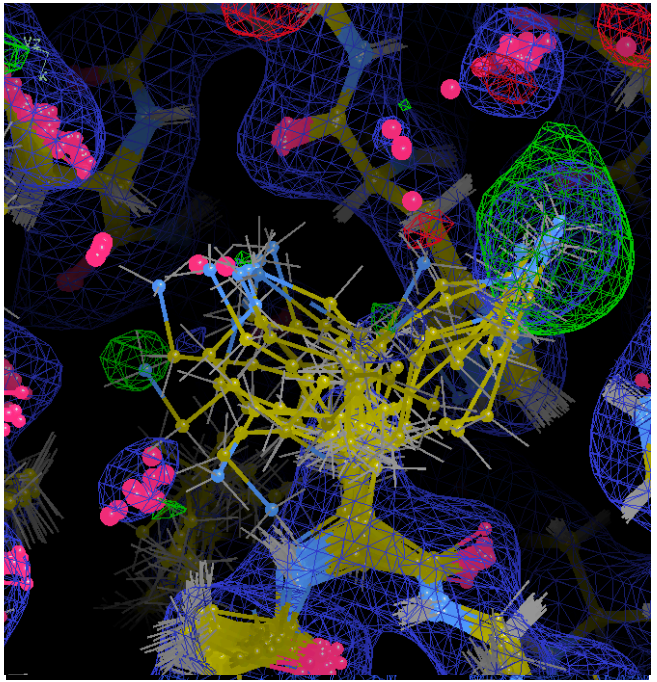
- Single structure refinement is a mess



Protein remodeling for making sense of uncertain density

Modeling ordered solvent misinterpreted by ensemble refinement

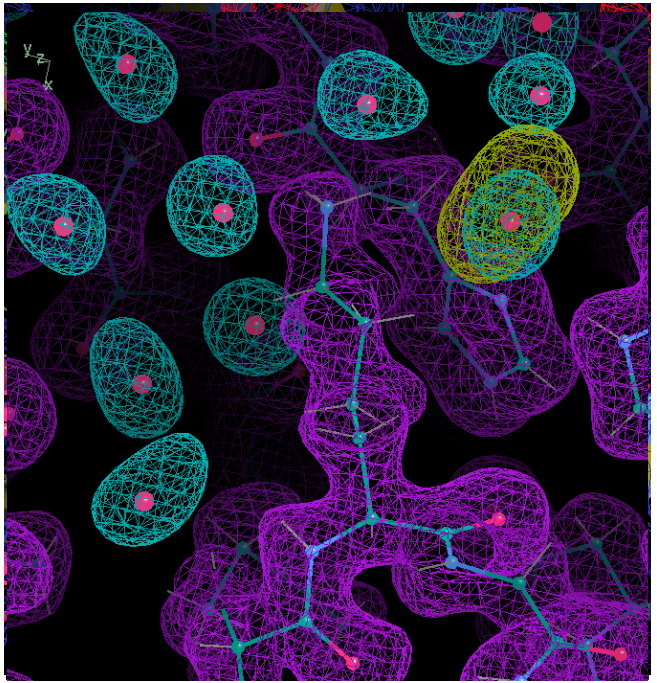
- Single structure refinement is a mess
- Ensemble refinement is… better…?



Protein remodeling for making sense of uncertain density

Modeling ordered solvent misinterpreted by ensemble refinement

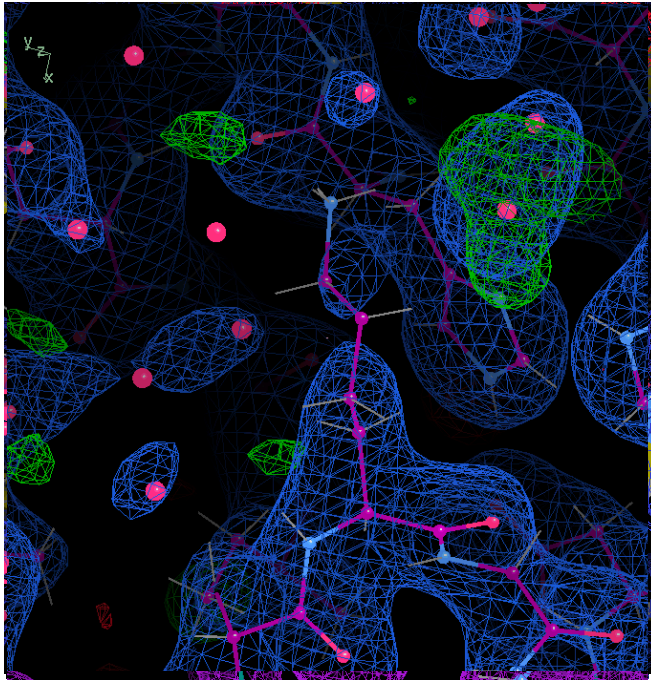
- Single structure refinement is a mess
- Ensemble refinement is… better…?
- MD densities reveal a fairly concerted side chain position, and water or chloride density in the off-backbone spot into which ensemble refinement had placed side chain amino groups



Protein remodeling for making sense of uncertain density

Modeling ordered solvent misinterpreted by ensemble refinement

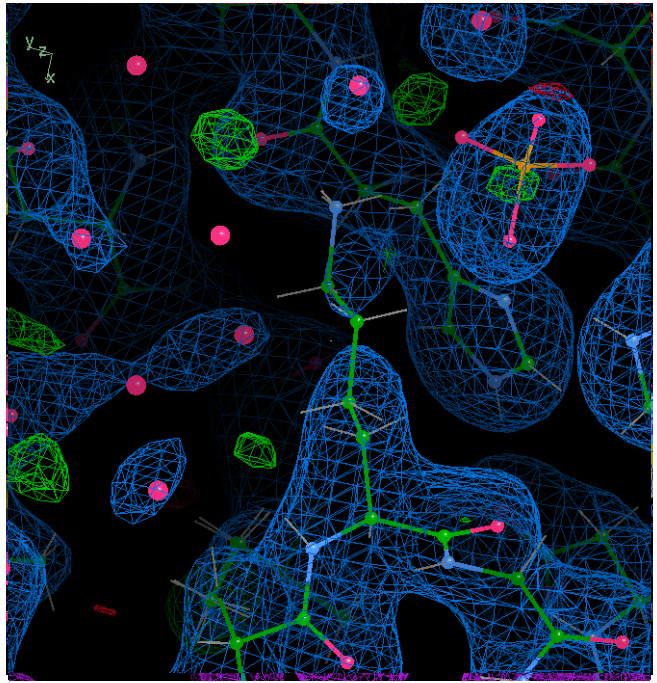
- Single structure refinement is a mess
- Ensemble refinement is… better…?
- MD densities reveal a fairly concerted side chain position, and water or chloride density in the off-backbone spot into which ensemble refinement had placed side chain amino groups
- Modeling with a chloride or water left difference density
- Although not modeled explicitly in the MD, phosphate was present in the crystallization buffer



Protein remodeling for making sense of uncertain density

Modeling ordered solvent misinterpreted by ensemble refinement

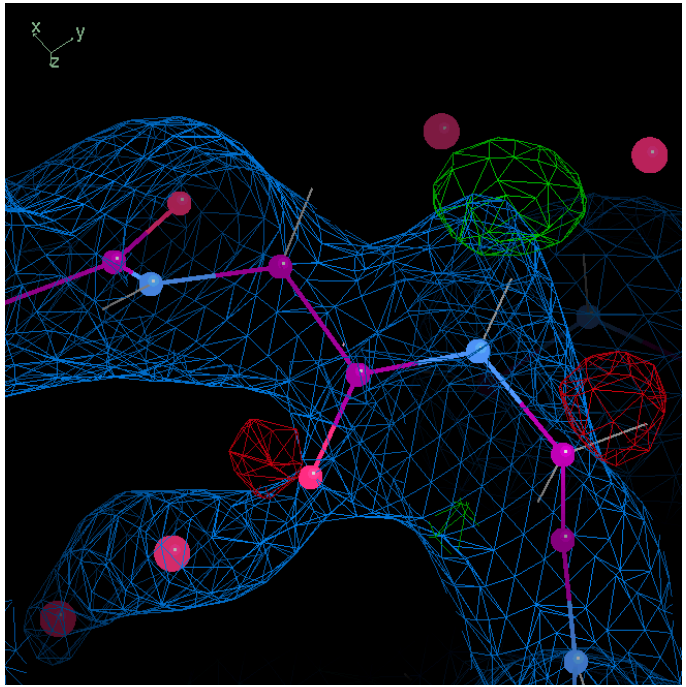
- Single structure refinement is a mess
- Ensemble refinement is… better…?
- MD densities reveal a fairly concerted side chain position, and water or chloride density in the off-backbone spot into which ensemble refinement had placed side chain amino groups
- Modeling with a chloride or water left difference density
- Although not modeled explicitly in the MD, phosphate was present in the crystallization buffer
- When modeled in, difference density all but disappears



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

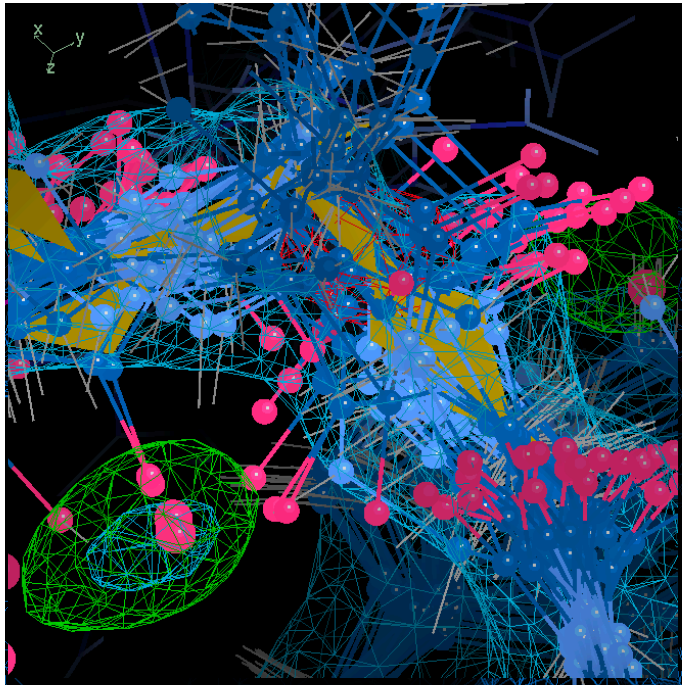
- Single structure refinement makes a peptide-flip multi conformer model look likely



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

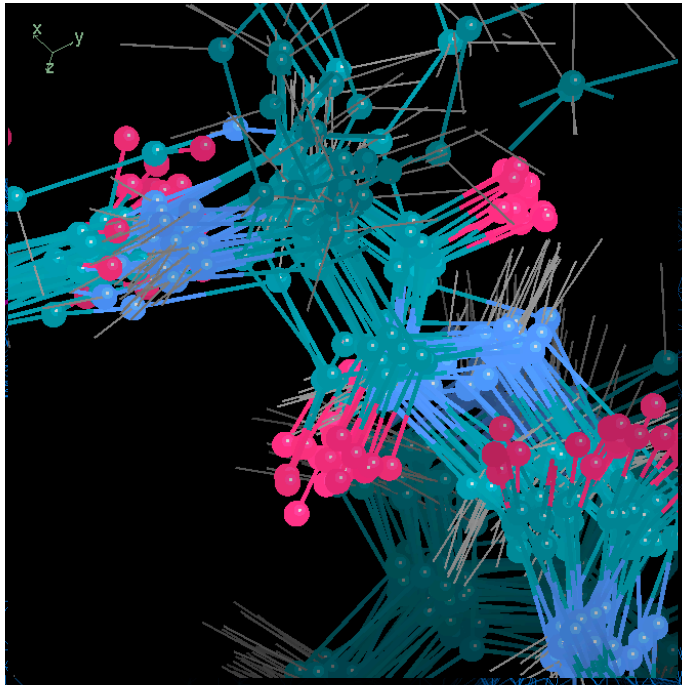
- Single structure refinement makes a peptide-flip multi conformer model look likely
- Ensemble model is a mess



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

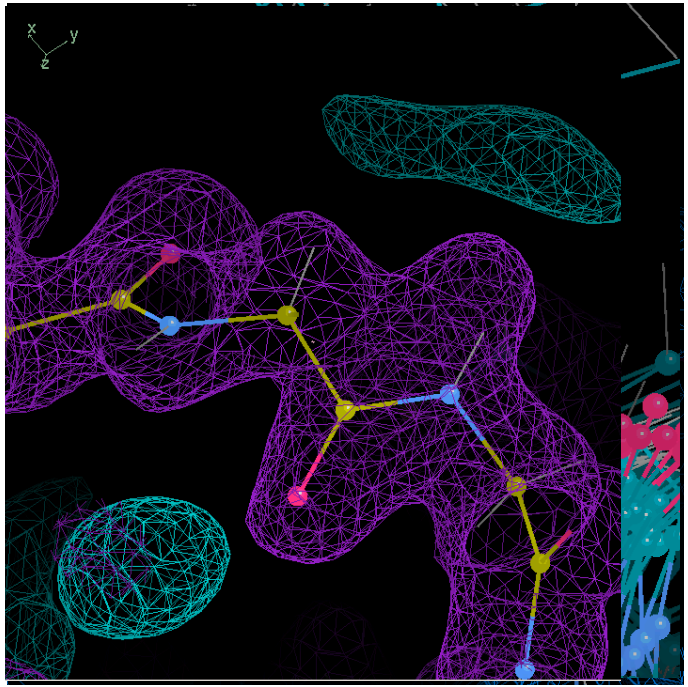
- Single structure refinement makes a peptide-flip multi conformer model look likely
- Ensemble model is a mess
- MD ensemble and density provide additional justification for the peptide flip



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

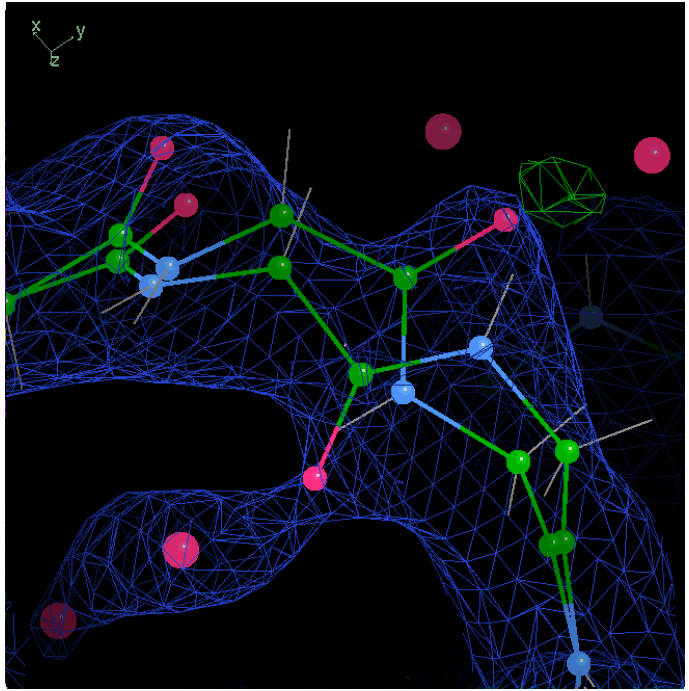
- Single structure refinement makes a peptide-flip multi conformer model look likely
- Ensemble model is a mess
- MD ensemble and density provide additional justification for the peptide flip



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

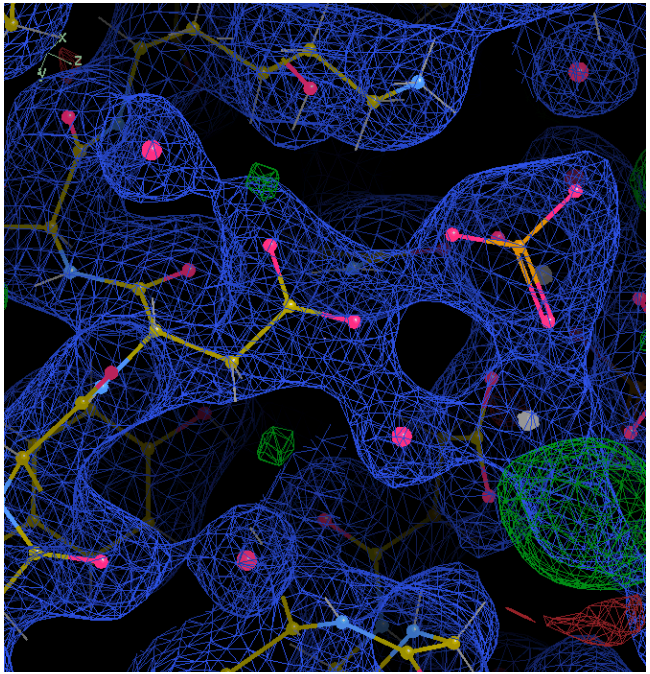
- Single structure refinement makes a peptide-flip multi conformer model look likely
- Ensemble model is a mess
- MD ensemble and density provide additional justification for the peptide flip
- B conformation pulled from MD ensemble
- Occupancies are similar to populations form MD ensemble



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

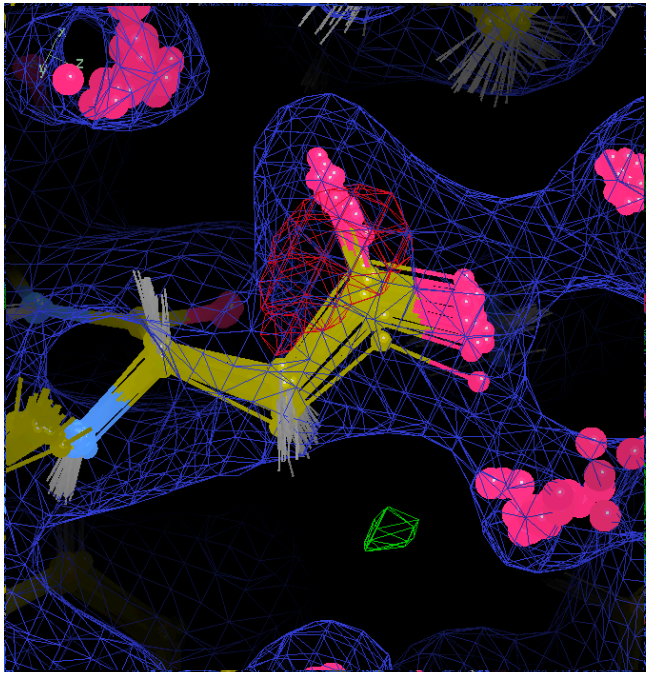
- Single structure refinement looks fairly reasonable… a bit of suspicious difference density



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

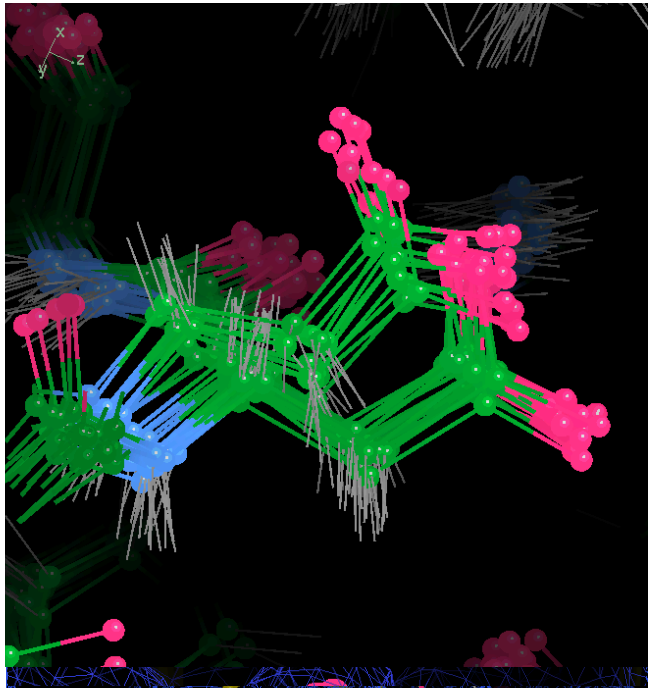
- Single structure refinement looks fairly reasonable… a bit of suspicious difference density
- Ensemble refinement model looks even more suspicious….



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

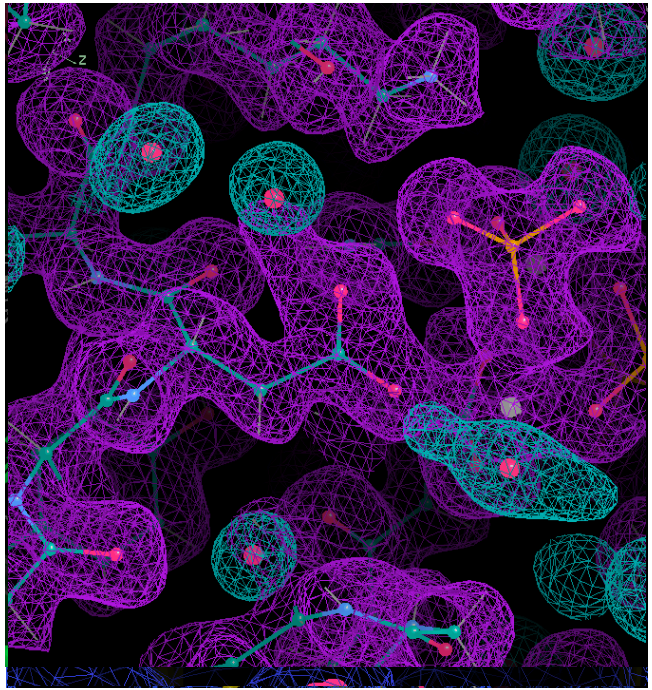
- Single structure refinement looks fairly reasonable… a bit of suspicious difference density
- Ensemble refinement model looks even more suspicious….
- The MD ensemble shows a clear two-state ensemble



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

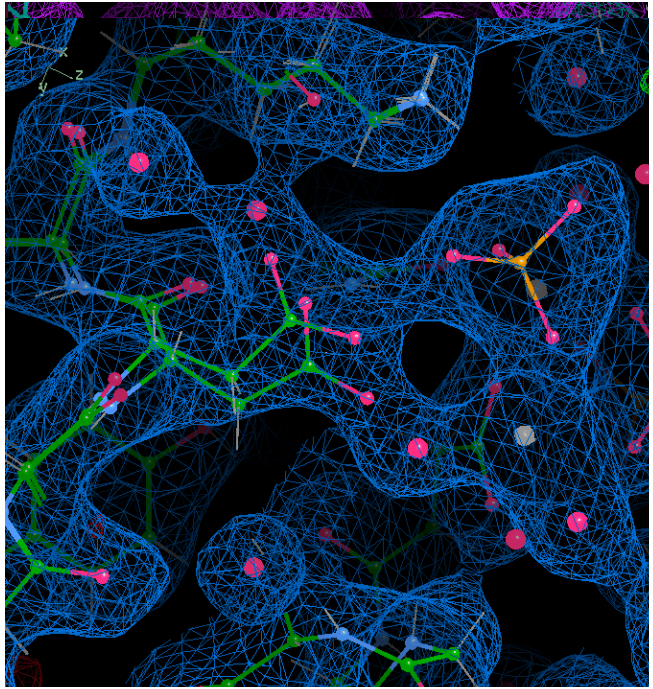
- Single structure refinement looks fairly reasonable… a bit of suspicious difference density
- Ensemble refinement model looks even more suspicious….
- The MD ensemble shows a clear two-state ensemble
- But there's also coordinated waters identified by the MD density



Protein remodeling for identifying and modeling structural heterogeneity and more

MD ensemble and density pick out two-state multi-conformer model

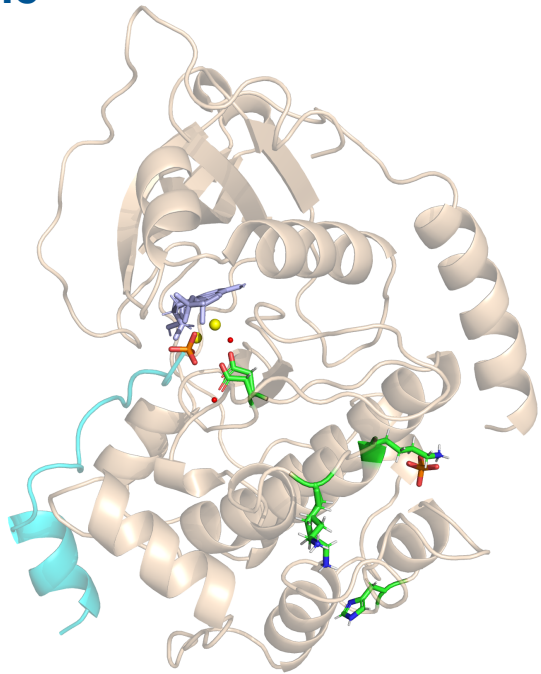
- Single structure refinement looks fairly reasonable… a bit of suspicious difference density
- Ensemble refinement model looks even more suspicious….
- The MD ensemble shows a clear two-state ensemble
- But there's also coordinated waters identified by the MD density
- We folded the waters in to the multi conformer model
- No difference density remaining in the region (> 3σ)



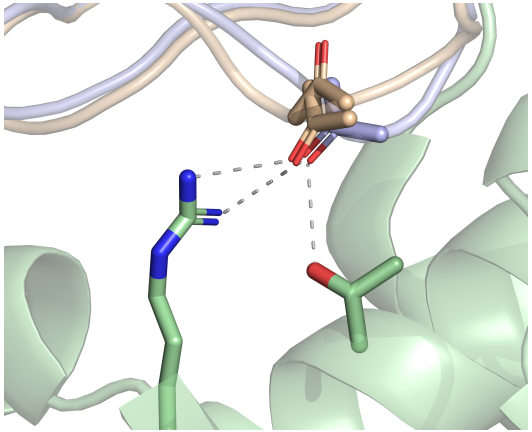
The System: Protein Kinase A Catalytic Subunit (PKA-C)

Well-studied kinase: blueprint for the human “kinome”

- In the past, crystalline MD people worked with mostly **model systems** (e.g. staph nuclease, lysozyme)
- PKA is involved in **metabolism, development, memory,** and **immune response**.
- Worked with Susan Taylor’s Lab (UCSD)
  - **Room temperature** crystal growth and data collection
  - **Single structure** and **ensemble** refinement

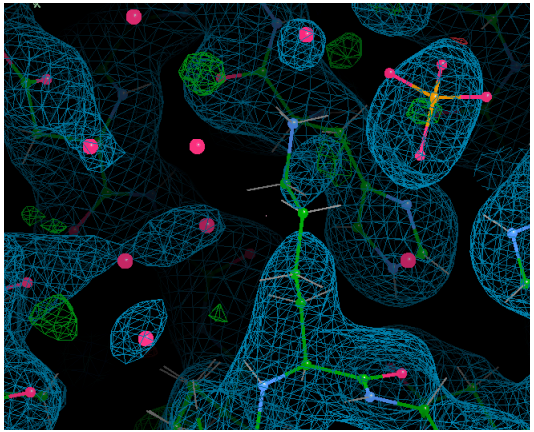


Mechanistic insights in to PKA biology from our analysis



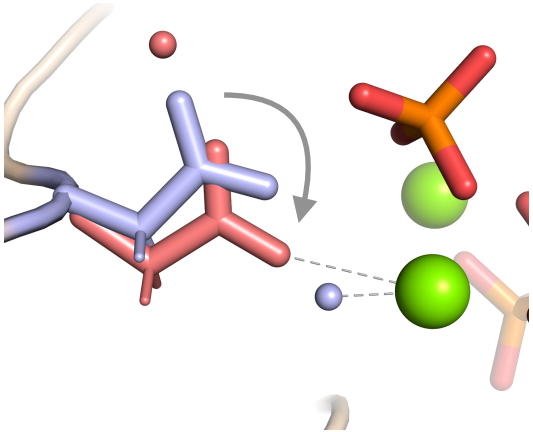
Conformational Selection

- Although the **regulatory subunit wasn't bound** in our crystals, the backbone of the residue **still samples the backbone conformation** for binding



Allosteric Modulation

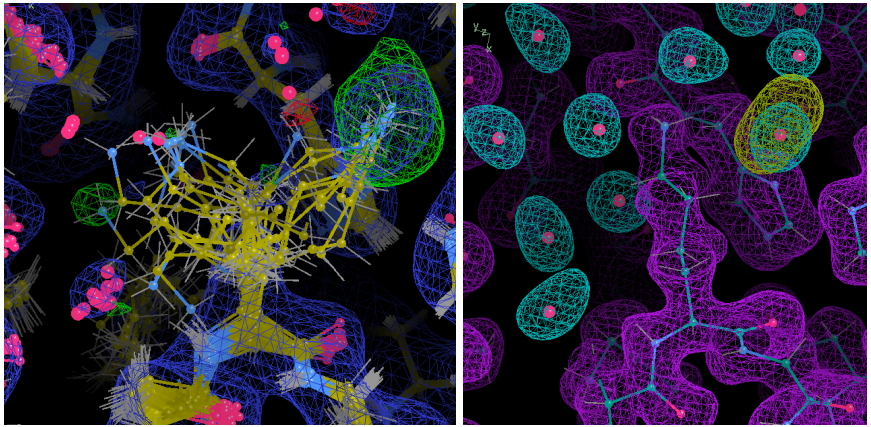
- Bound phosphate has not been observed before
- May play some roll in allosteric modulation

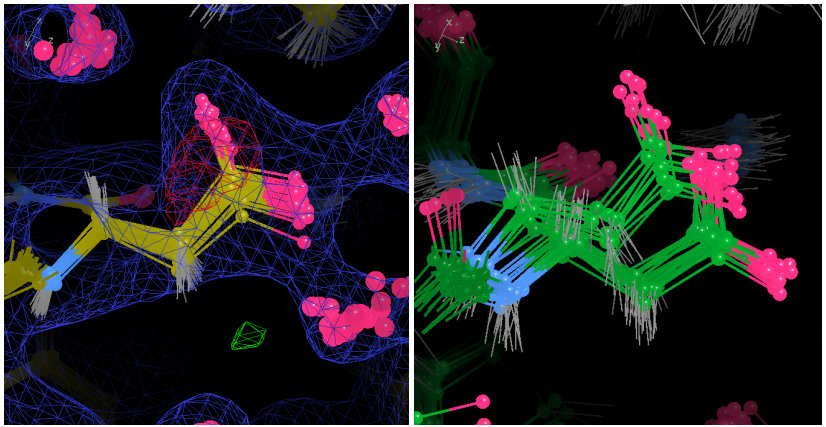


Progression of Catalysis

- Room temperature crystal growth and data collection
- Unique intermediate catalytic state
- Multiconformer: pre + post catalysis?
- Priming Mg^{2+} for exit from active site

Implications





Simultaneous refinement of protein and solvent

- There are fundamental issues to refining structures against density that is not separable into components
- MD can provide a jumping off point

Fully leveraging MD for crystallography

- Ensemble refinement currently does not seem to be using MD to its fullest potential
- Explicit solvent appears to be important

Future directions

Still a good bit of work to be done in ***automating*** these methods

- **"Easy"**
  - Iterative solvation and equilibration
    - Turning up the pressure
    - NVT
  - Density calculations
    - *xtraj.py* (parallel, blazing fast)
  - MD ensembles from trajectories
  - Multi-conformer *identification*
  - Identifying areas of interest

- **Hard**
  - Unrestrained/unbiased simulation
    - Force-field people… help
  - Multi-conformer *modeling*
    - protein + solvent
  - State of the art experiments (XFEL)
    - Out of equilibrium experiments (e.g. mix-and-inject, temperature jump, etc.)

Appreciation

- **OpenEye!**

- Mike Wall
- David Mobley
- Phillip Aoto
- Lily Vu
- James Fraser
- Alex Wolff
- Susan Taylor
- The Mobley Lab