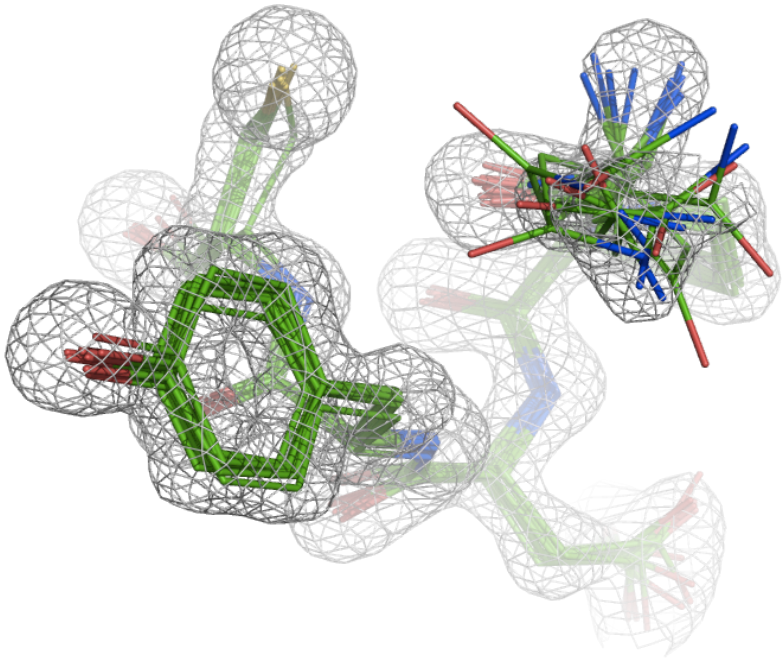
Ensemble refinement of protein crystal structures in PHENIX



Tom Burnley | Piet Gros

Universiteit Utrecht

NWO

Phenix

Incomplete modelling of disorder contributes to R factor gap

Only ~5% of residues in the PDB are modelled with more than one conformation (x-ray structures)

Multiple discrete models restricted due to increase in number of model parameters

Vitkup *et al.,* (2002) Lang *et al.,* (2010)

Incomplete modelling of disorder contributes to R factor gap

Only ~5% of residues in the PDB are modelled with more than one conformation (x-ray structures)

Multiple discrete models restricted due to increase in number of model parameters

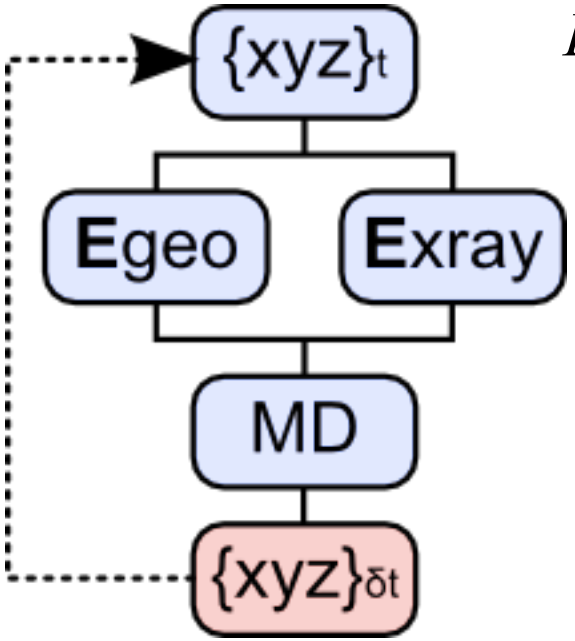
Molecular dynamics simulations produce a Boltzmann-weighted population of inter-related structures

MD simulations can be restrained with x-ray data

Simulated Annealing/MD refinement

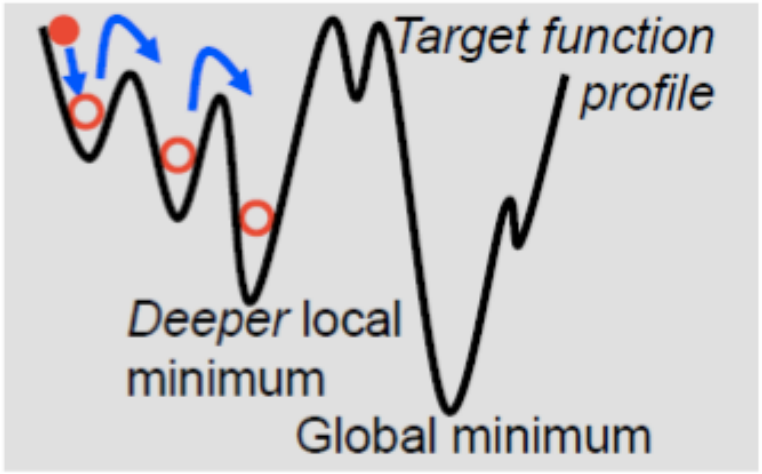
Geometric restraints:

E_{bond}
E_{angle}
E_{dih}
…



X-ray restraint:

$$E_{\text{X-ray}} = \sum_{hkl} w_{hkl} (|F_{\text{obs}}(hkl)| - k|F_{\text{calc}}(hkl)|)^2$$



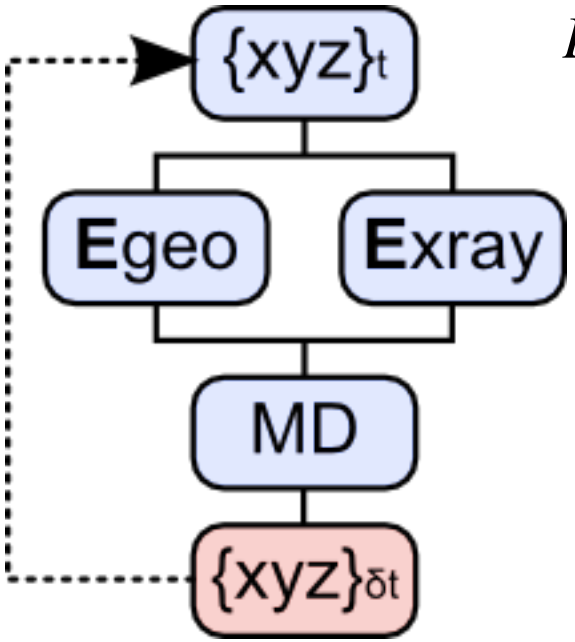
Simulation temperature >2000K
Trajectory resolves local minima
'Final model' = end structure

Brunger *et al.*, 1987

“Time-averaged” MD refinement

Geometric restraints:

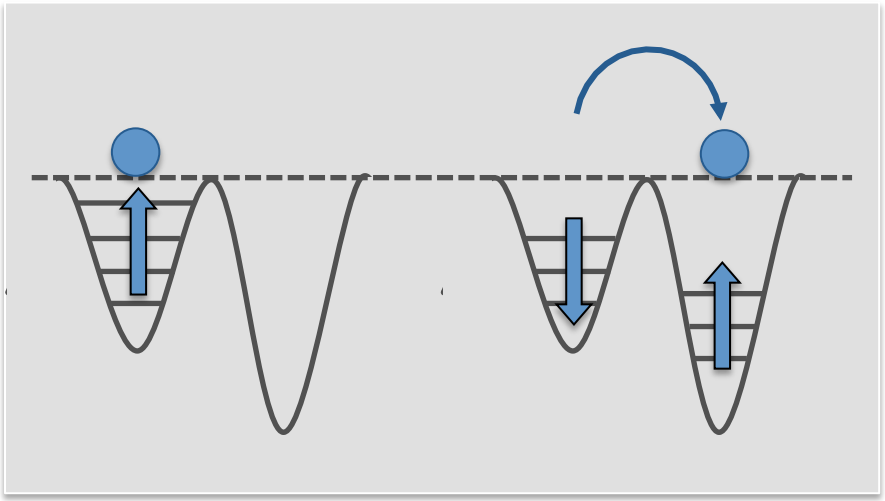
E_{bond}
E_{angle}
E_{dih}
…



X-ray restraint:

$$E_{\text{X-ray}} = \sum_{hkl} w_{hkl} (|F_{\text{obs}}(hkl)| - k|\langle F_{\text{calc}}(hkl)\rangle|)^2$$

Sampling



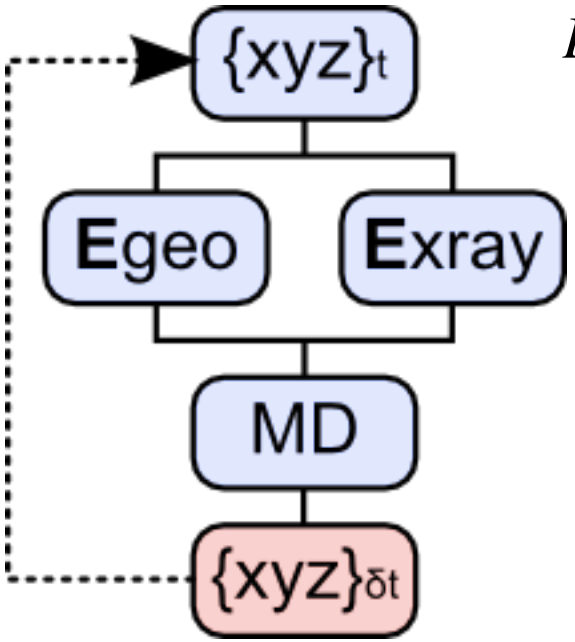
$$\langle F_{calc}\rangle_t = (1 - e^{-\Delta t/\tau_x}) F^t_{calc} + e^{-\Delta t/\tau_x} \langle F_{calc}\rangle_{t-\Delta t}$$

Gros *et al.*, 1990

"Time-averaged" MD refinement

Geometric restraints:

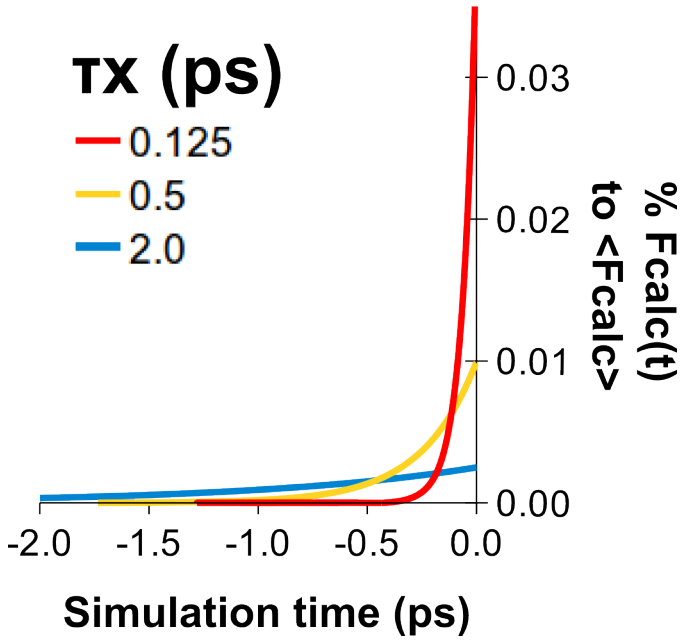
E_{bond}
E_{angle}
E_{dih}
$\ldots$



X-ray restraint:

$$E_{\text{X-ray}} = \sum_{hkl} w_{hkl} (|F_{\text{obs}}(hkl)| - k|\langle F_{\text{calc}}(hkl)\rangle|)^2$$

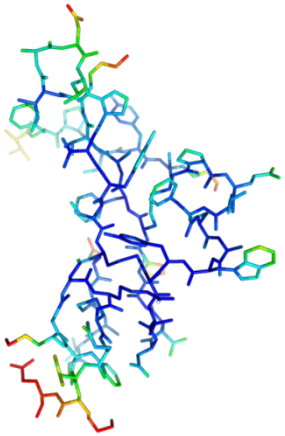
Running average

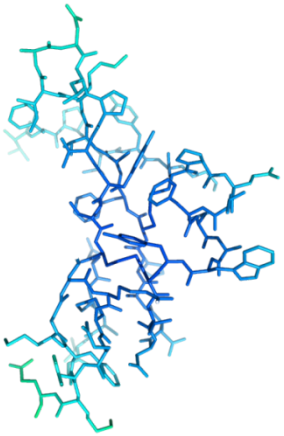


$$\langle F_{calc}\rangle_t = (1 - e^{-\Delta t/\tau_x}) F^t_{calc} + e^{-\Delta t/\tau_x} \langle F_{calc}\rangle_{t-\Delta t}$$

Simulation temperature = 300K
'Final model' = trajectory ensemble

Ensemble refinement with TA restraints

(1) 

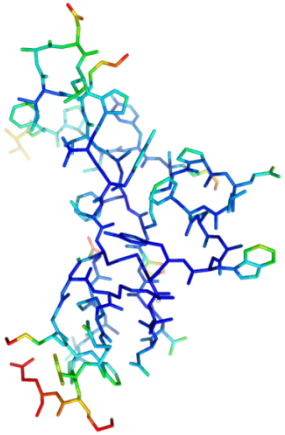
(2) 

(1) Start: ‘Traditional’ structure

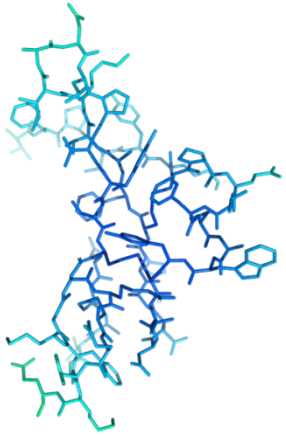
(2) Fit TLS / remove alt. conf.

Ensemble refinement with TA restraints

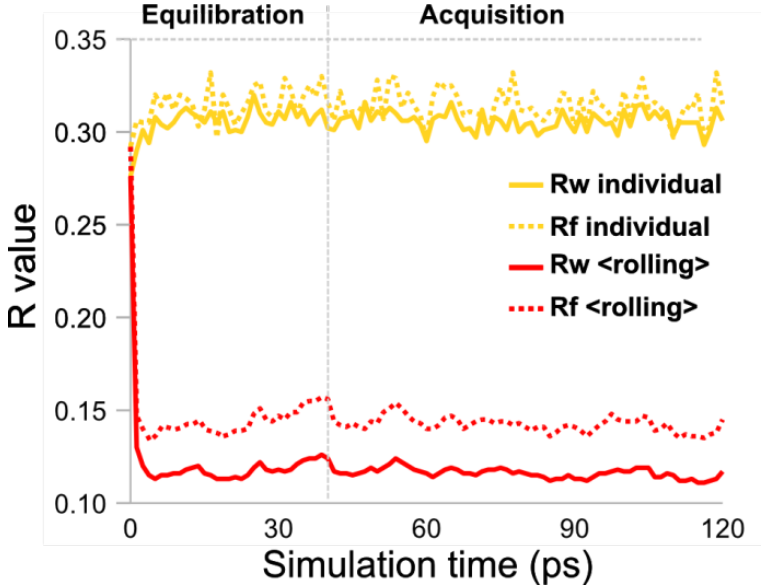
(1)



(2)



(3)



(1) Start: 'Traditional' structure

(2) Fit TLS / remove alt. conf.

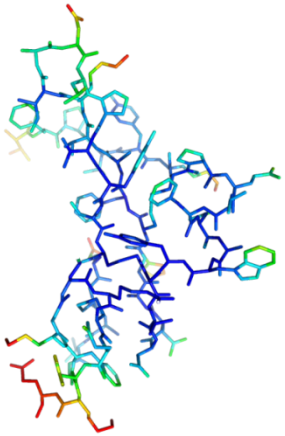
(3) TA restrained MD simulation

Collect structure / 0.04ps

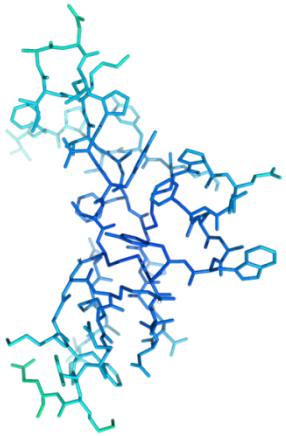
X-ray restraints accelerate sampling

Ensemble refinement with TA restraints

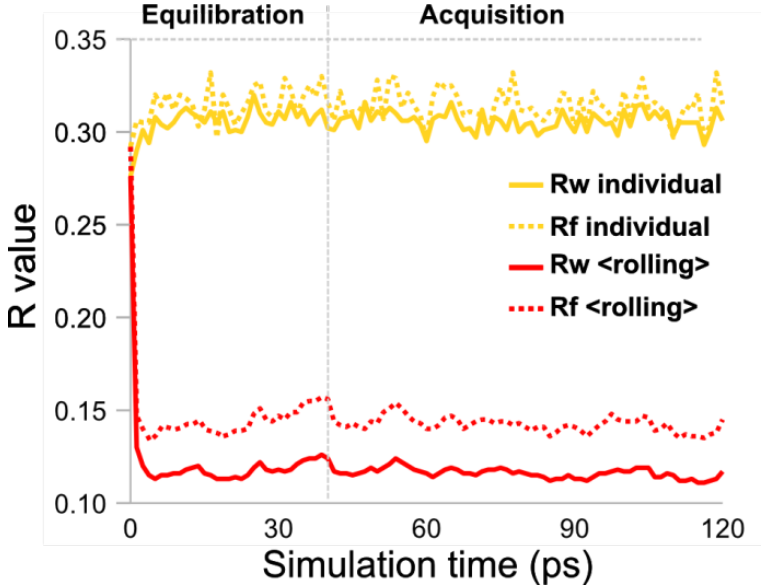
(1)



(2)



(3)



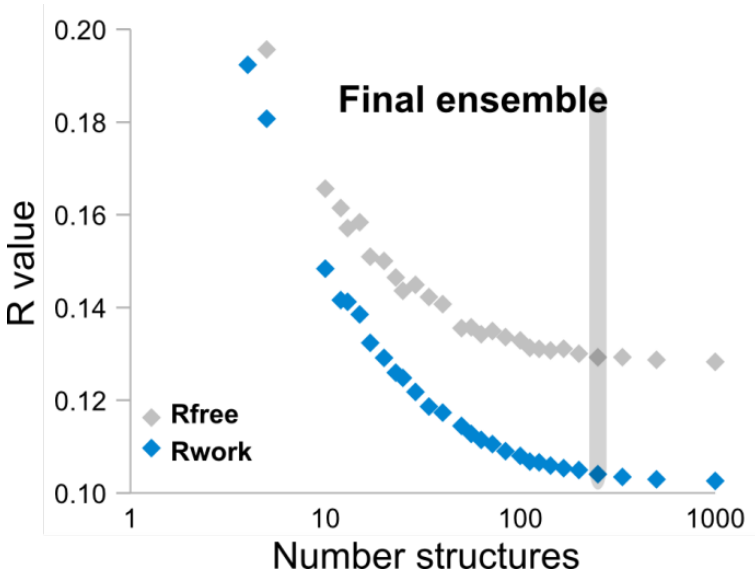
(1) Start: 'Traditional' structure

(2) Fit TLS / remove alt. conf.

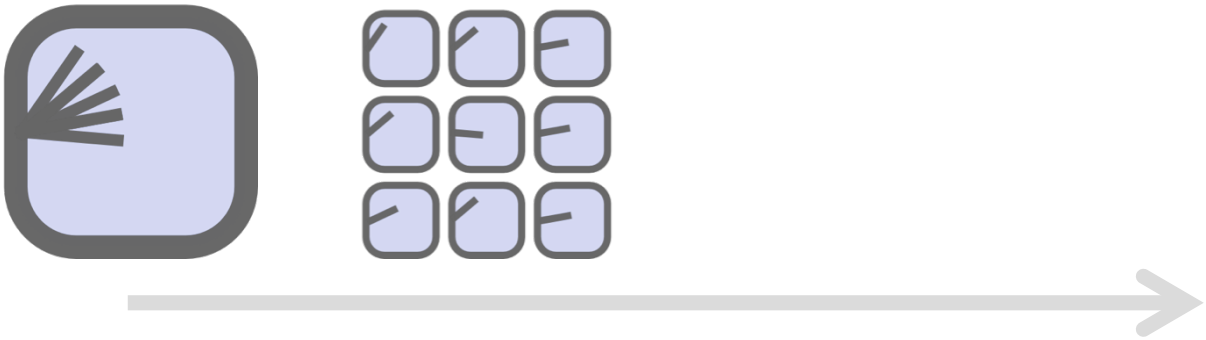
(3) TA restrained MD simulation

(4) Final ensemble

(4)



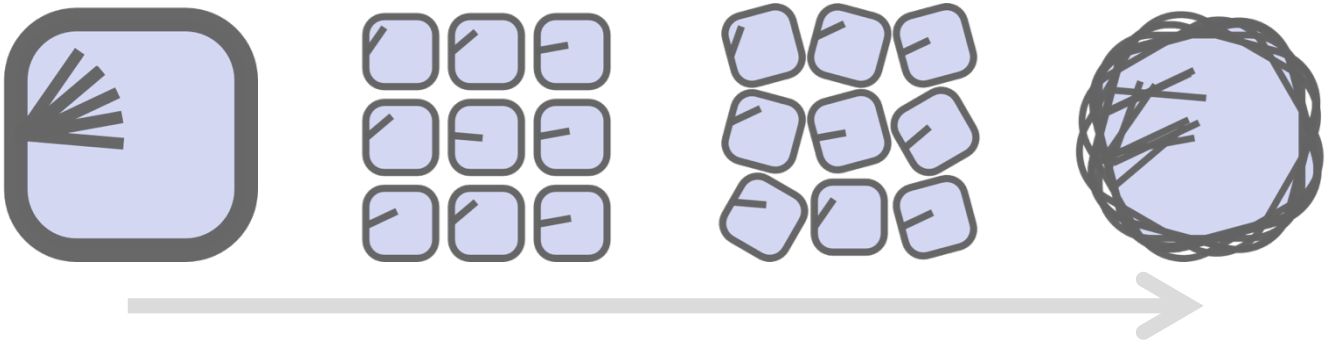
Molecular disorder / lattice disorder



Protein with local disorder

Perfect crystal lattice

Molecular disorder / lattice disorder



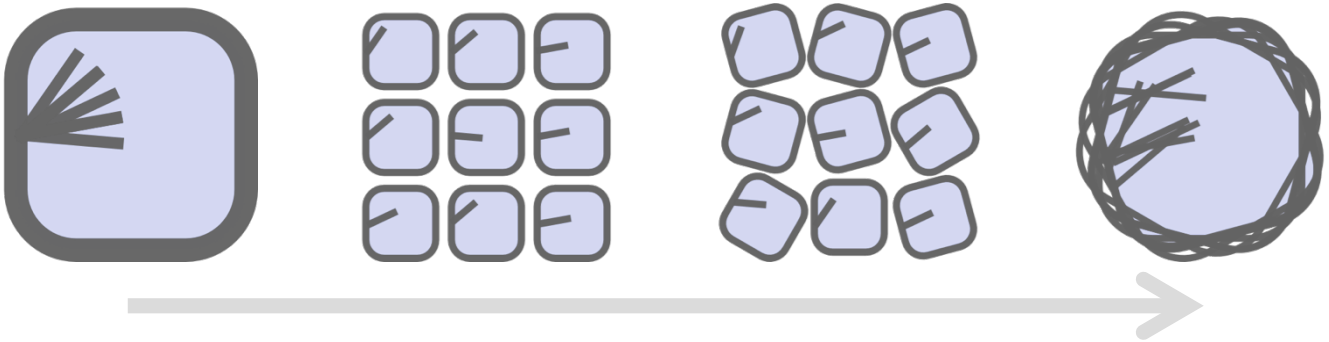
Protein with local disorder

Perfect crystal lattice

'Real' crystal lattice

Averaged data

Molecular disorder / lattice disorder



Protein with local disorder

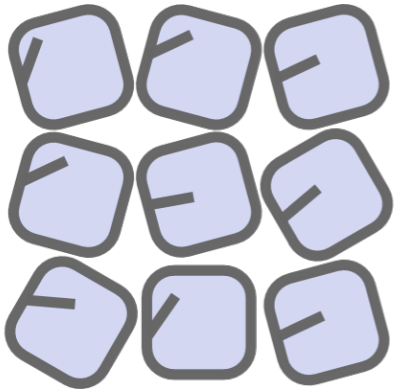
Perfect crystal lattice

'Real' crystal lattice

Averaged data

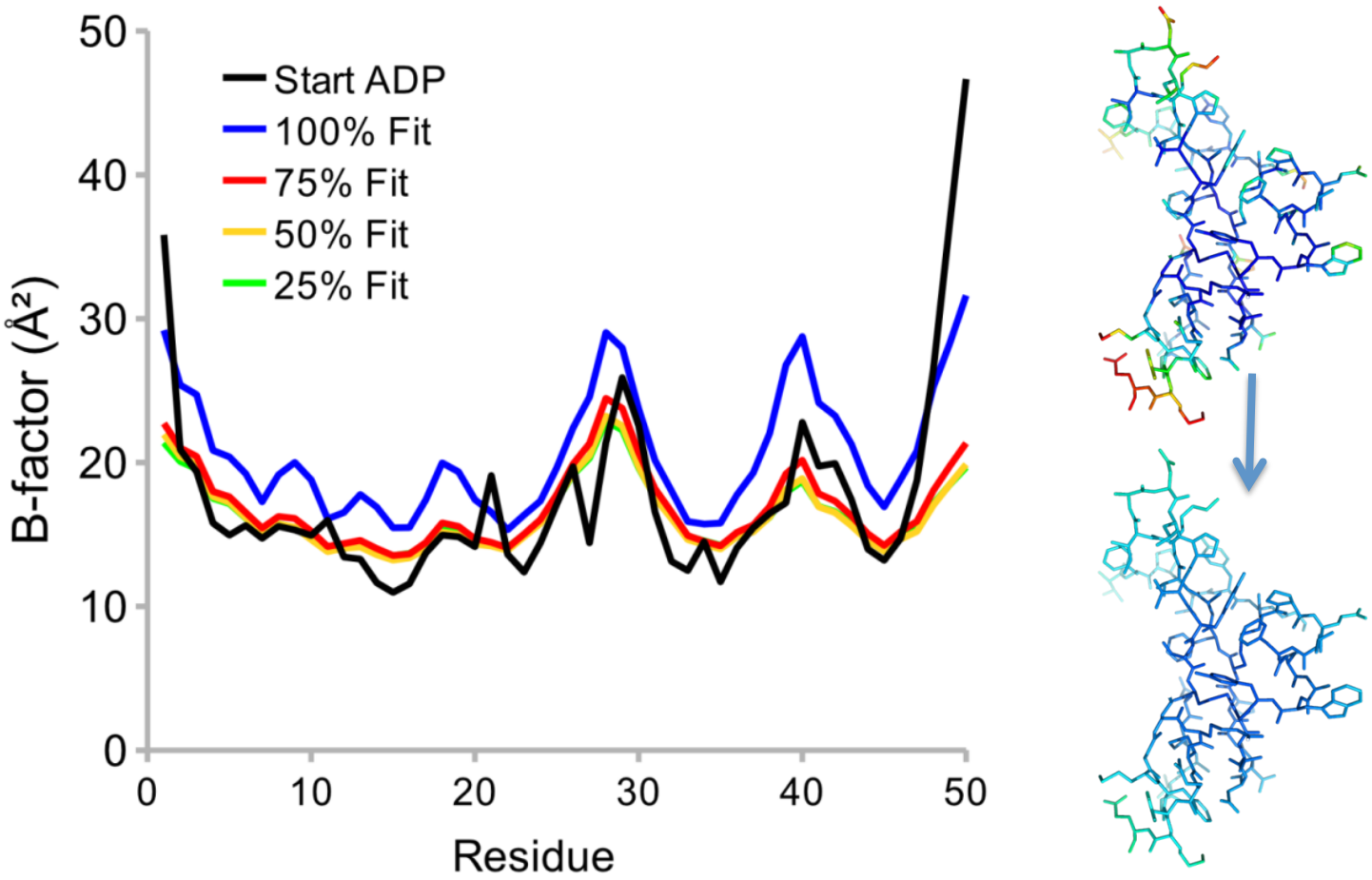
Deconvolute: molecular disorder from lattice disorder

Rigid body disorder modelled with TLS



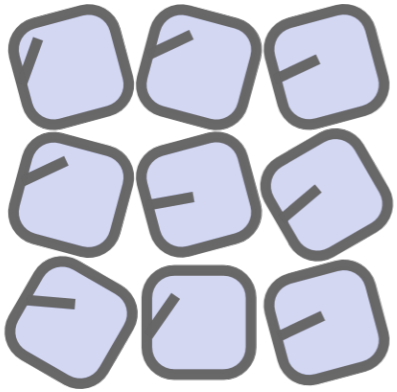
- **Lattice distortions, domain motions...**
- Model rigid body motions with ***TLS*** model

Rigid body disorder modelled with TLS

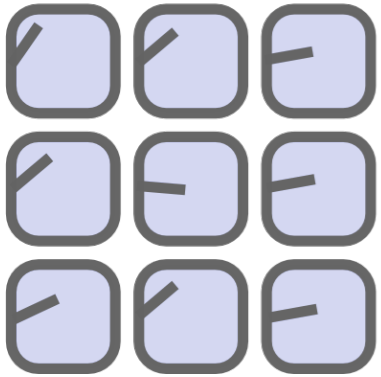


Extract core rigid body motion by excluding atoms with large local fluctuations (defined as deviations from B_{TLS}).

Local disorder sampled within restrained MD



- **Lattice distortions, domain motions...**
- Model rigid body motions with TLS model
  - 20 parameters per group
  - 1 group / domain
- Fit TLS to starting structure B-factors

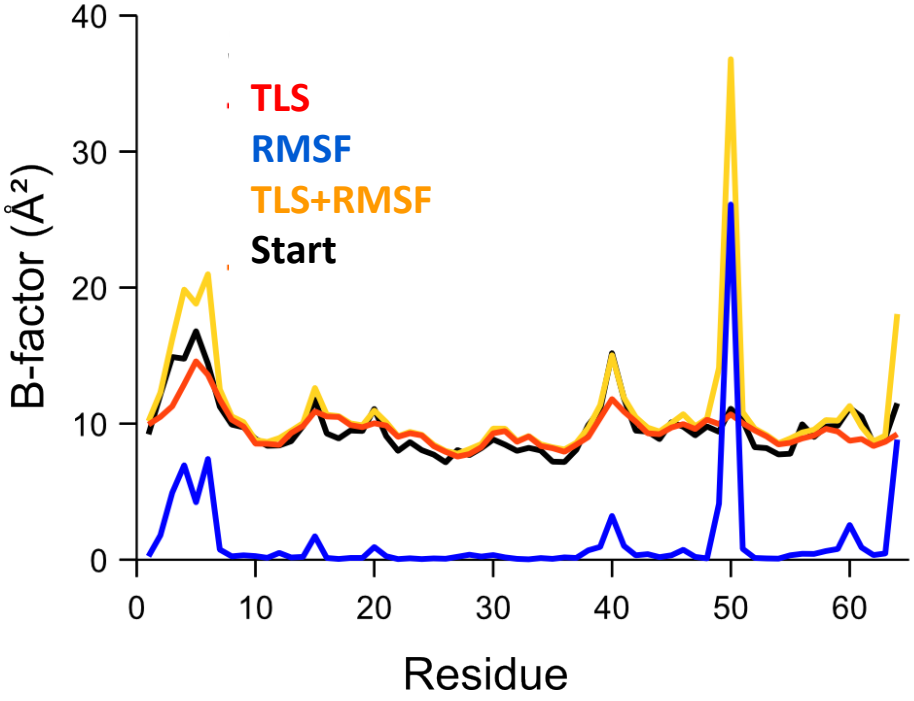


- **Side-chains, loops...**
- Local disorder sampled in MD
- MD simulation restrained with X-ray data

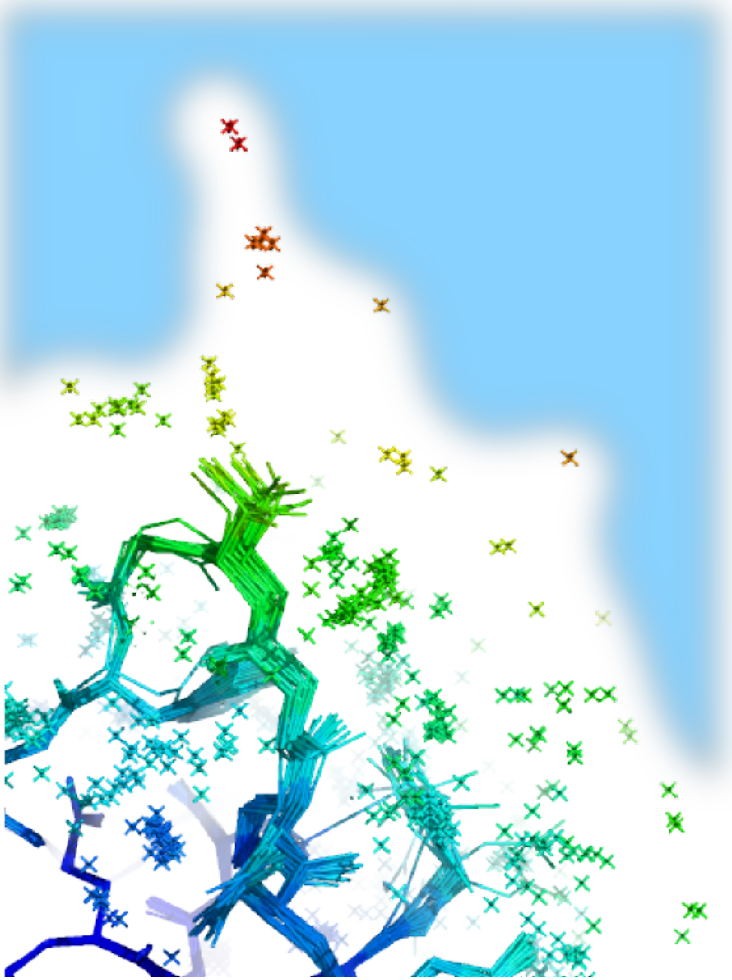
Composite B-factor sum of TLS and atomic fluctuations

- **TLS**
  - Core TLS model
- **RMSF**
  - Atomic fluctuations in ensemble
- **TLS+RMSF**
  - Total disorder in ensemble

- **Start**
  - Input single structure ADPs

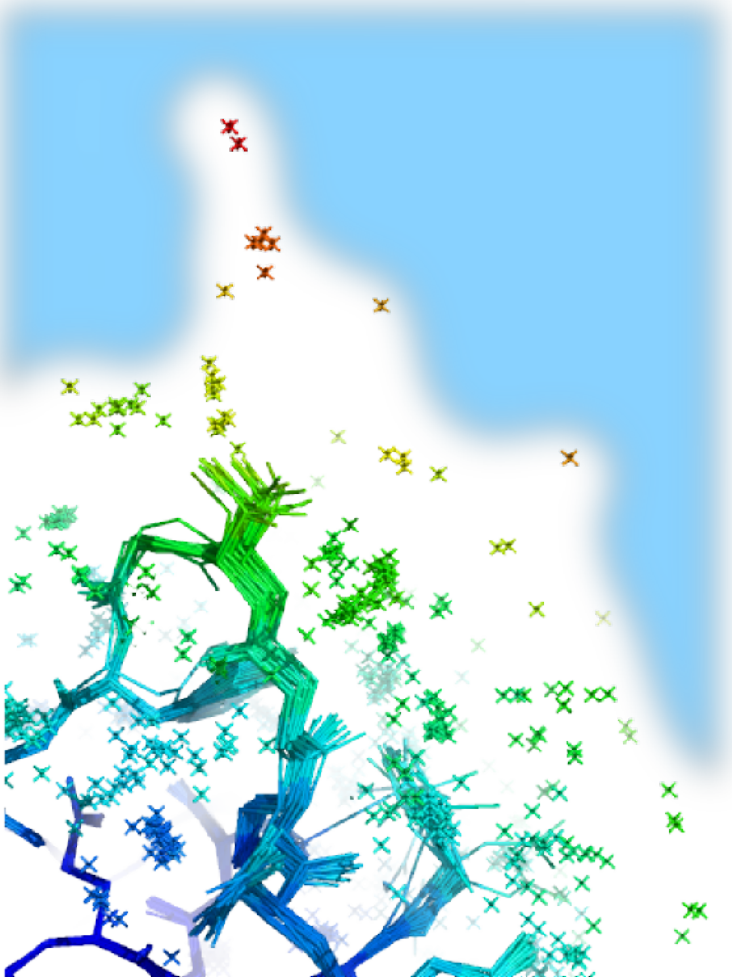


Dual explicit-bulk solvent model



- Explicit solvent
  - Model with explicit atoms
  - Water picked every 250 steps
    “standard” rules:
    > 3 σ in difference map
    < 3 Å distances
  - B-factor from nearest TLS group

Dual explicit-bulk solvent model



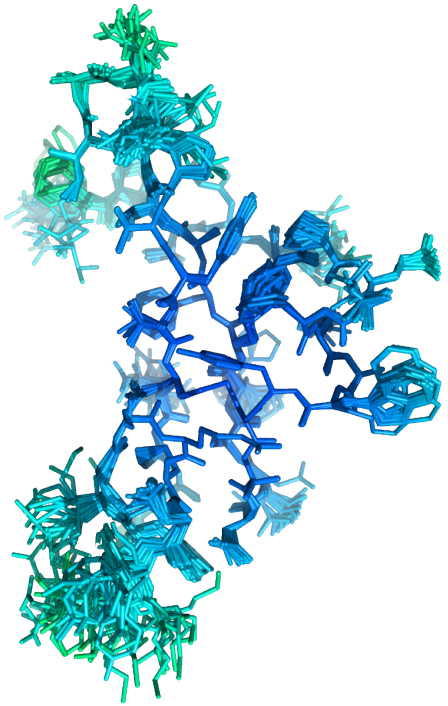
- Explicit solvent
  - Model with explicit atoms
  - Water picked every 250 steps
    - “standard” rules:
    - > 3 σ in difference map
    - < 3 Å distances
  - B-factor from nearest TLS group

- Bulk solvent
  - Model with ‘density mask’

$$\left\langle F_{mask} \right\rangle_t = (1 - e^{-\Delta t/\tau_x}) F^t_{mask} + e^{-\Delta t/\tau_x} \left\langle F_{mask} \right\rangle_{t-\Delta t}$$

Development of ensemble refinement

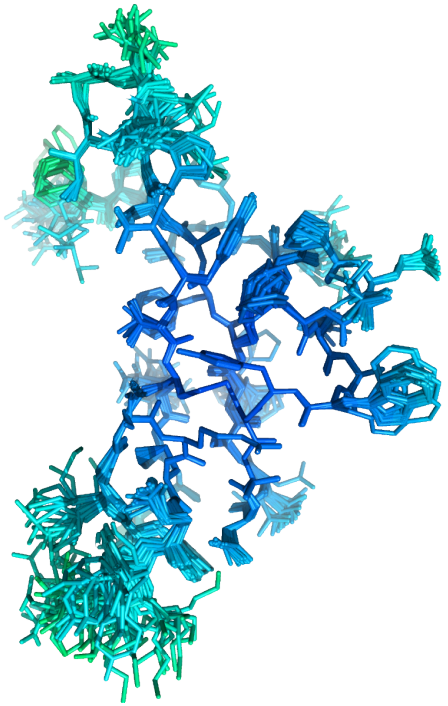
- **PHENIX**
- **Time-averaged x-ray restrained MD**
- **TLS fitting**
- **Explicit- and bulk- solvent model**
- **Maximum-likelihood target function**



Phenix

Development of ensemble refinement

- **Tested with 20 datasets**
- **Resolution: 1 - 3 Å**
- **ASU size: 50-1000 residues**
- **CPU time: 7 - 100 hours**
- **50 – 500 models / ensemble**



Phenix

Ensemble refinement reduces Rfree

- ***Rfree: ensemble vs phenix.refine***
  - **Rfree reduced in all cases**

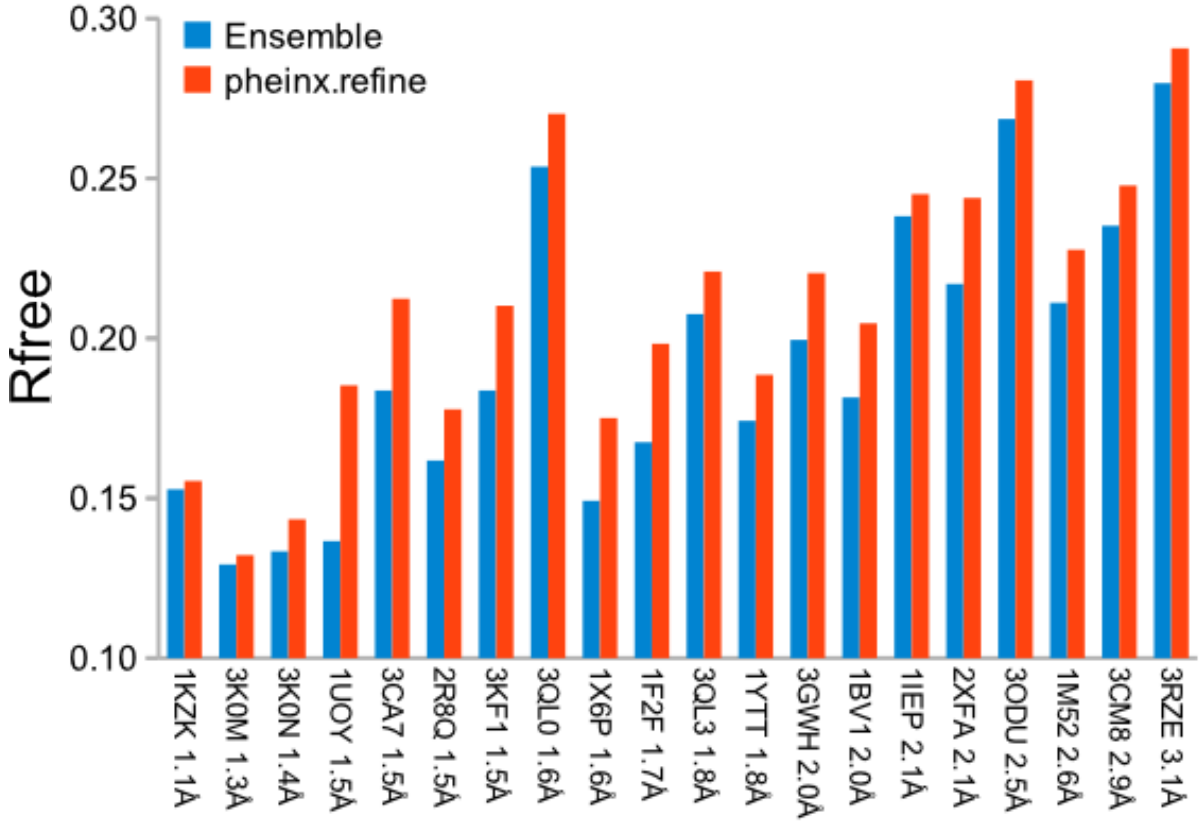
    - 4.9% (max)

    - 0.3% (min)

    - 1.8% (mean)
  - Rf/Rw ratio (mean):

    = 1.23 phenix.refine

    = 1.25 ensemble



Ensemble refinement reduces Rfree

- ***Rfree: ensemble vs phenix.refine***
  - **Rfree reduced in all cases**

    - 4.9% (max)

    - 0.3% (min)

    - 1.8% (mean)
  - Rf/Rw ratio (mean):

    = 1.23 phenix.refine

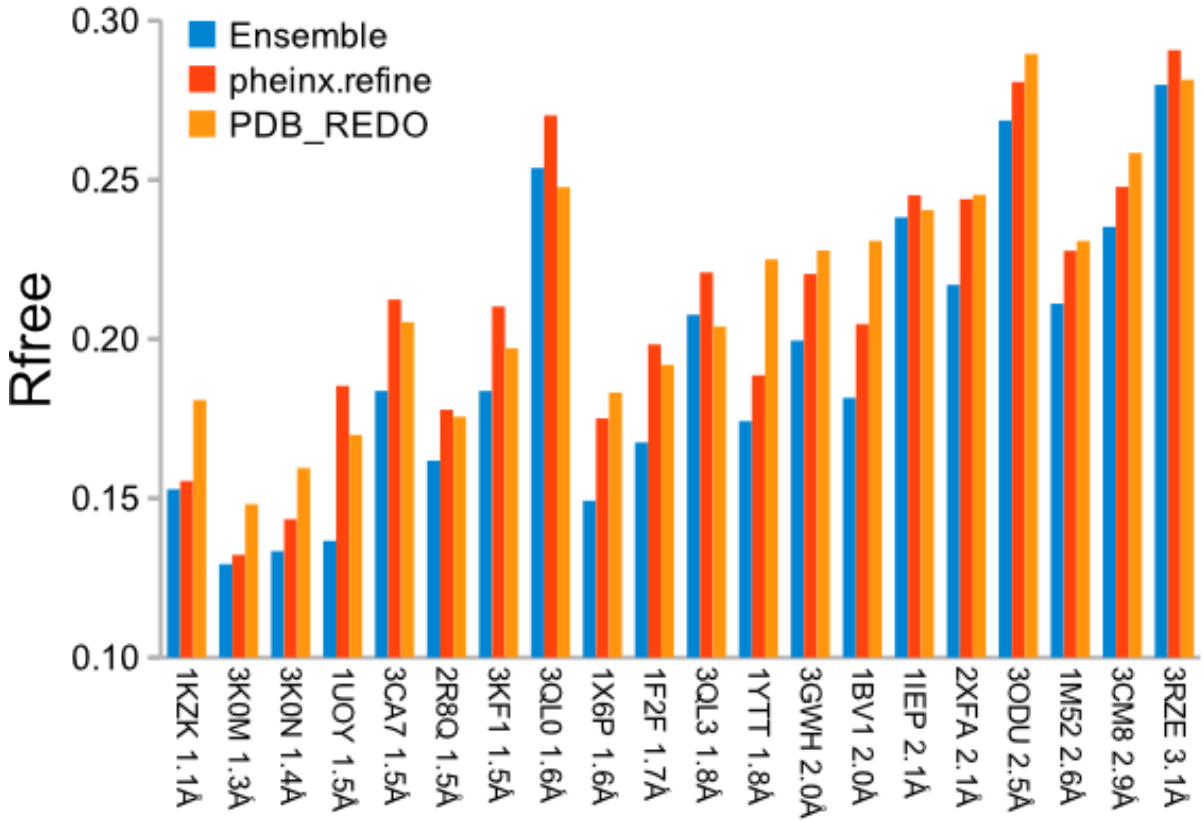
    = 1.25 ensemble

- ***Rfree: ensemble vs PDB_REDO***

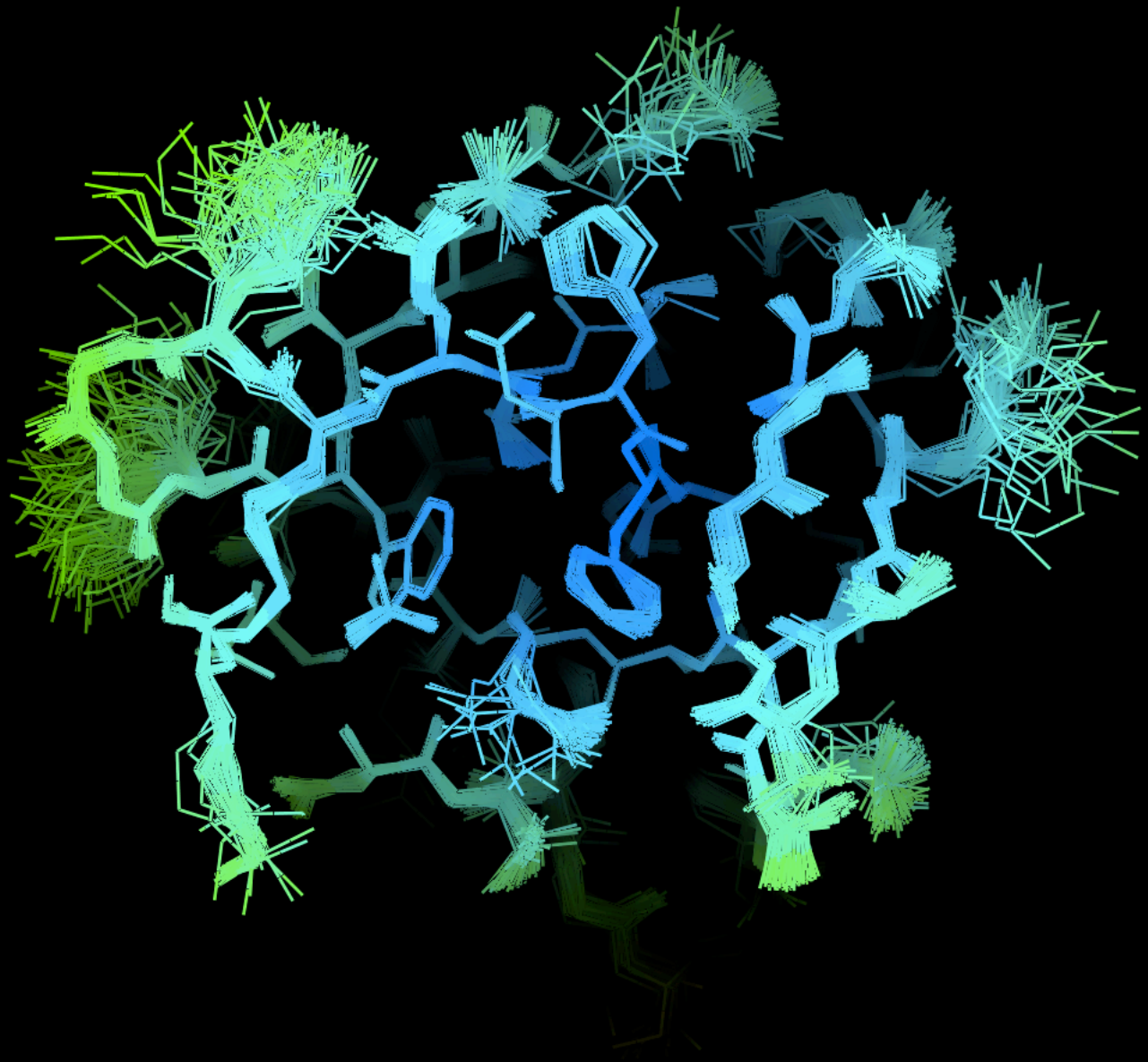
  - 5.1% (max)

  + 0.6% (min)

  - 2.1% (mean)



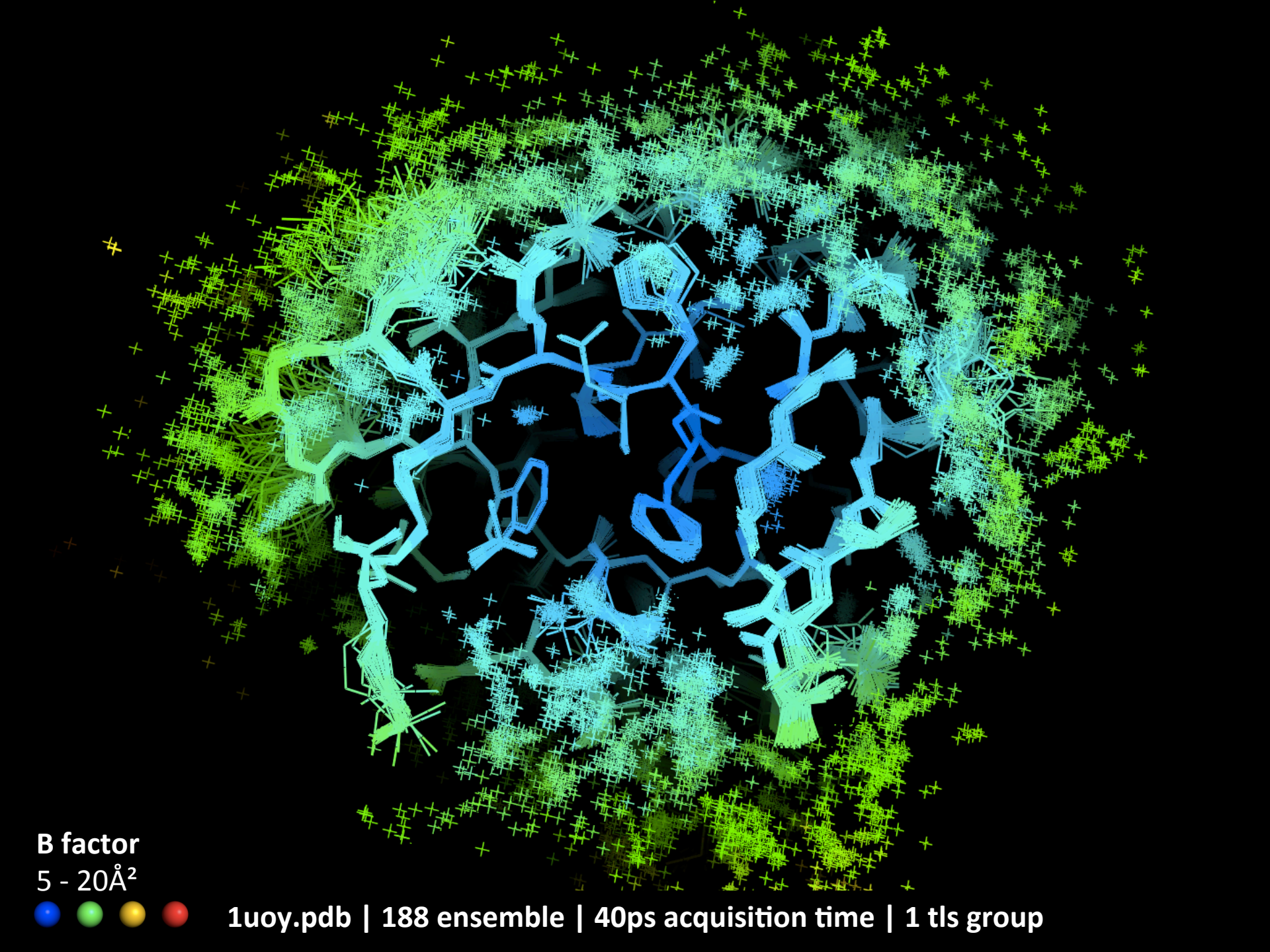
Joosten *et. al* (2011)

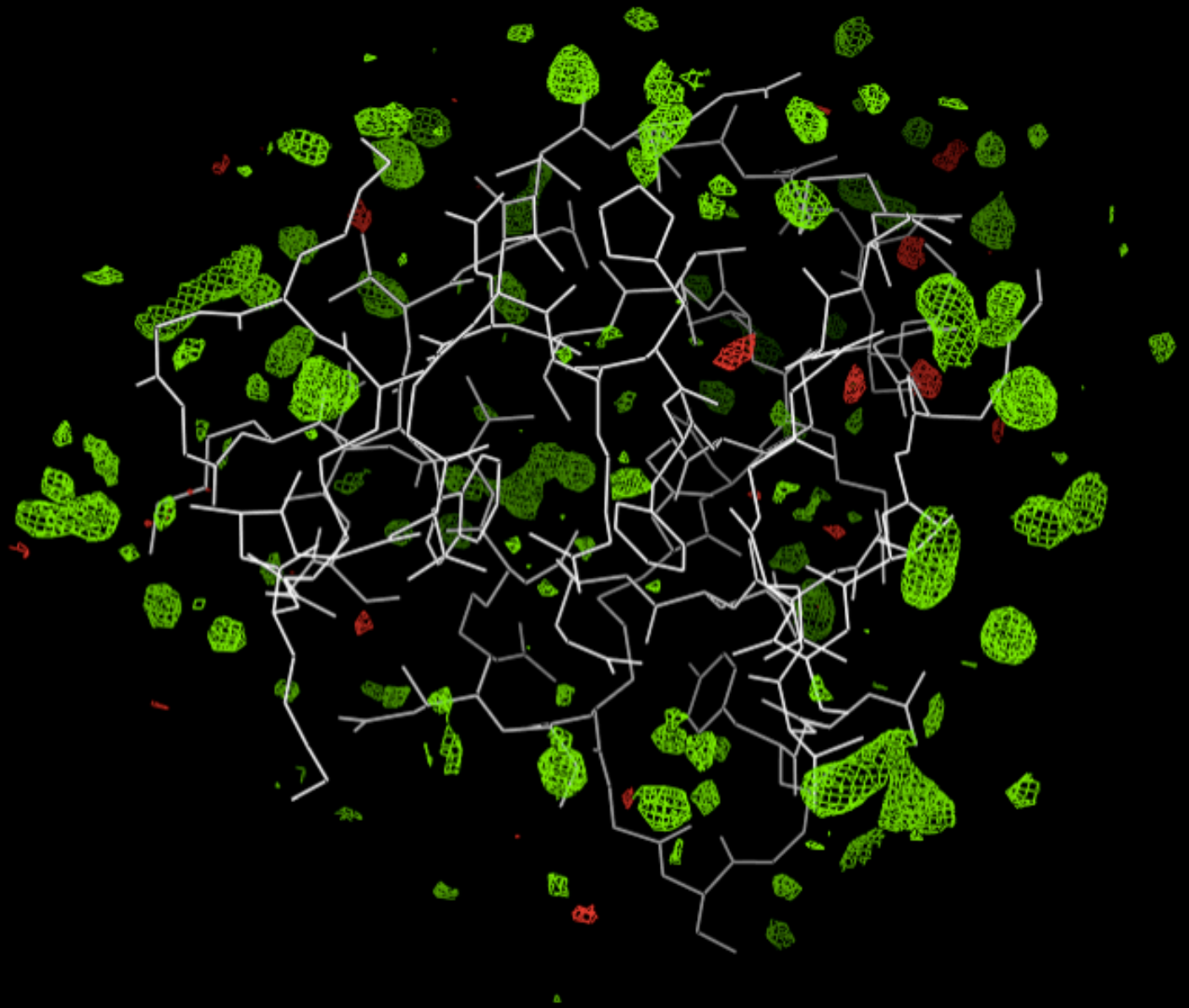


B factor

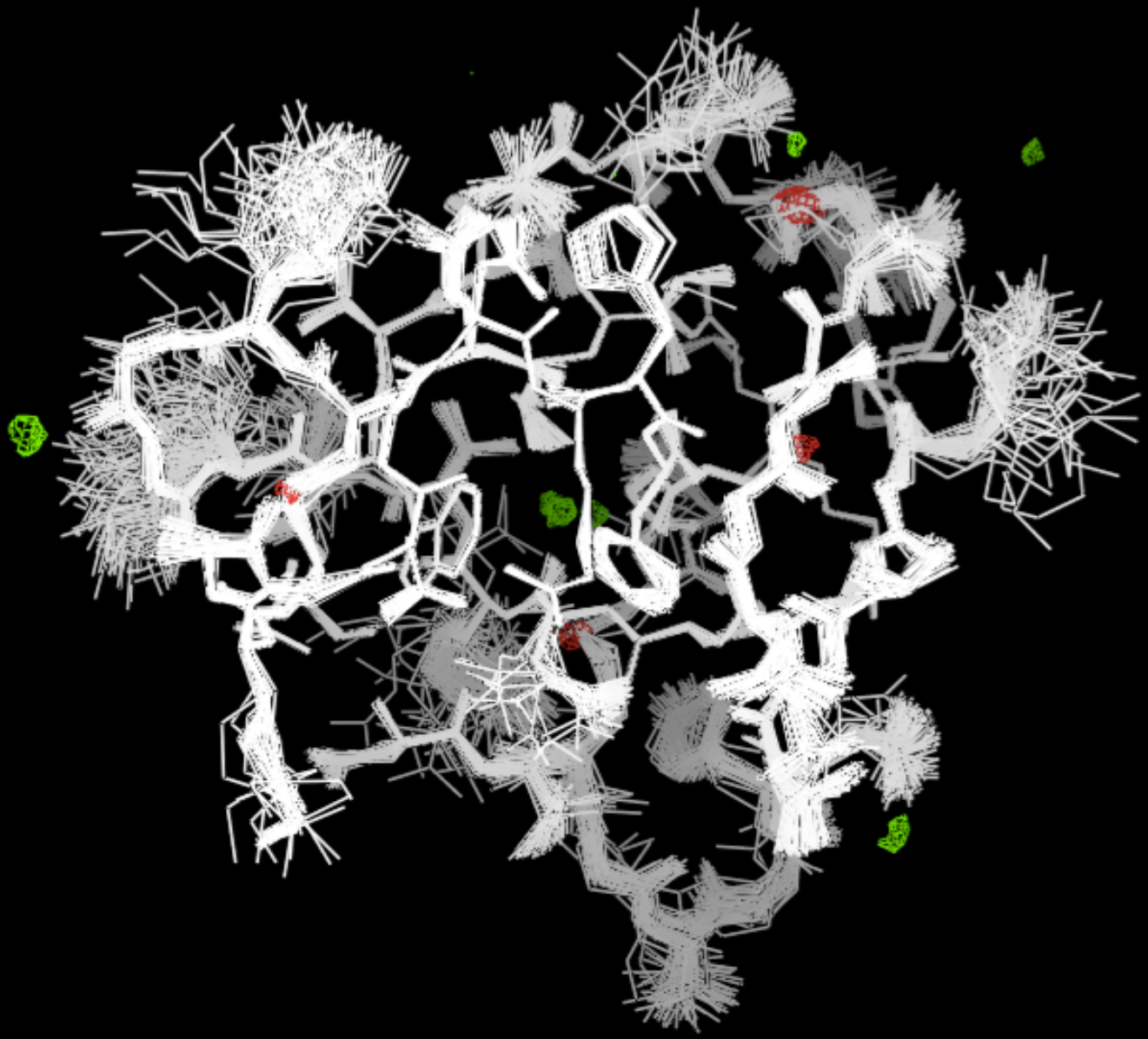
5 - 20Å²

1uoy.pdb | 188 ensemble | 40ps acquisition time | 1 tls group


B factor
5 - 20Å²
1uoy.pdb | 188 ensemble | 40ps acquisition time | 1 tls group



1uoy.pdb | phenix.refine | 1 tls group | mFo-DFc ±0.49 e/Å³ (3.00 σ)

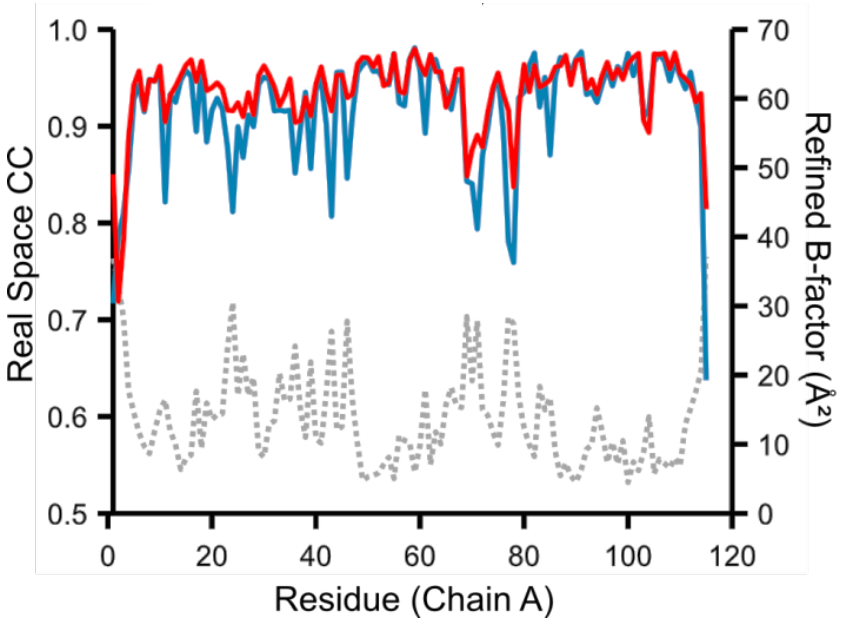


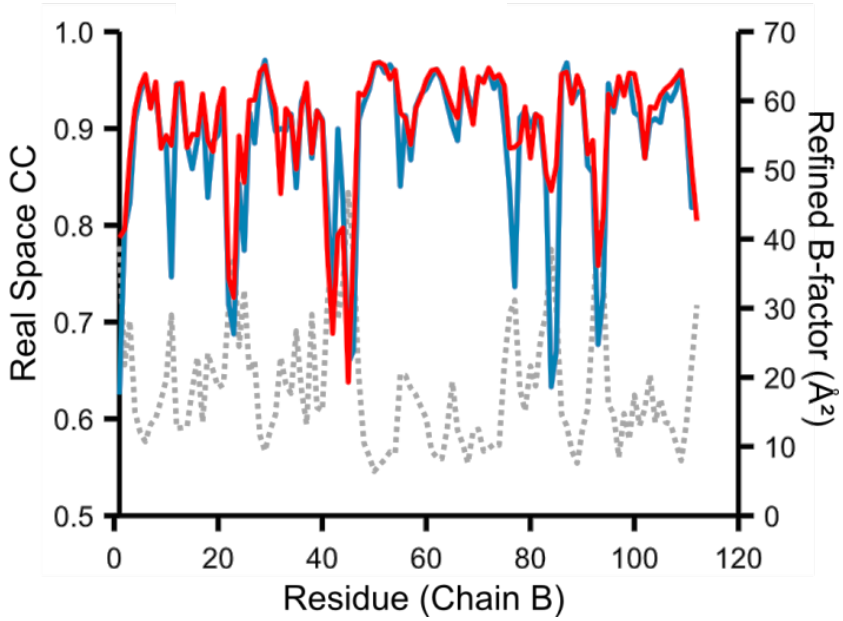
1uoy.pdb | 188 ensemble | 1 tls group | mFo-DFc ±0.49 e/Å³ (4.27 σ)

Improved real-space correlation

Burling *et al.* (1996):
Excellent experimentally phased data for MBP: 1YTT (1.8-Å res.)

Ensemble
Phenix.refine
B-factor

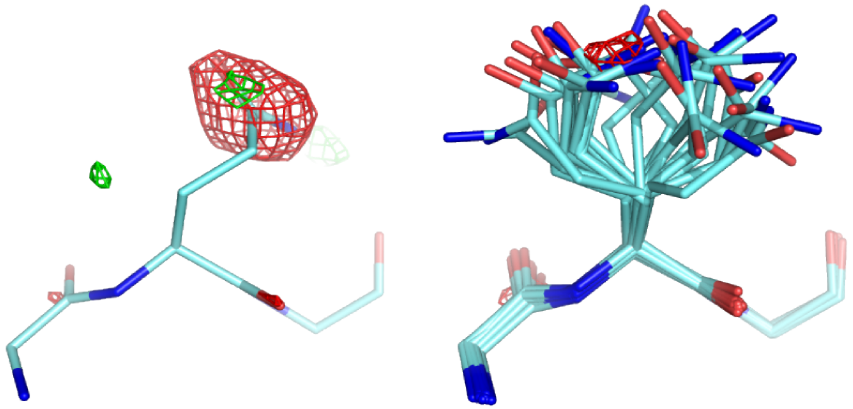




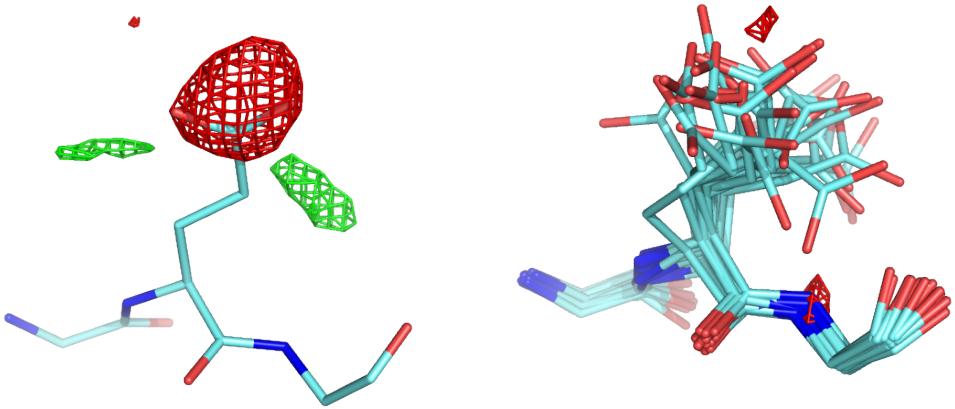
	Rwork	Rfree	Real Space CC
Ensemble	0.139	0.174	0.903
phenix.refine	0.166	0.189	0.895
PDB	0.185	0.206	0.873

Disordered side chain in MBP (1YTT)

Gln167 (A)



Glu117 (A)

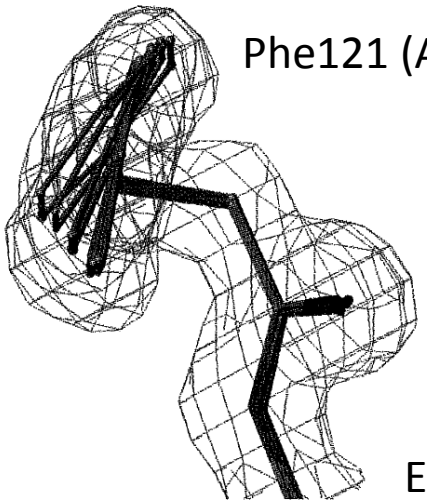


Phenix.refine
Diff. vector map (3σ)

Ensemble
Diff. vector map (3σ)

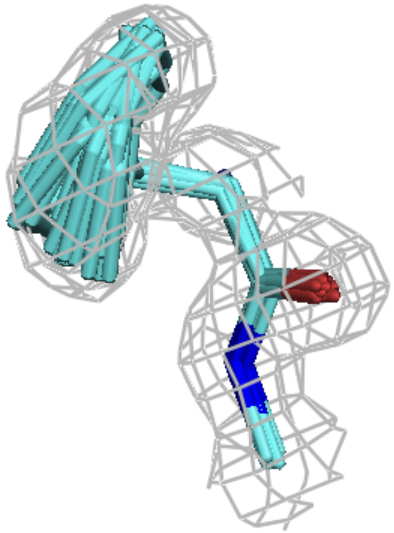
Anisotropic side chain in MBP (1YTT)

Burling *et al.* (1996)



Phe121 (A)

Experimental map (1.4σ)

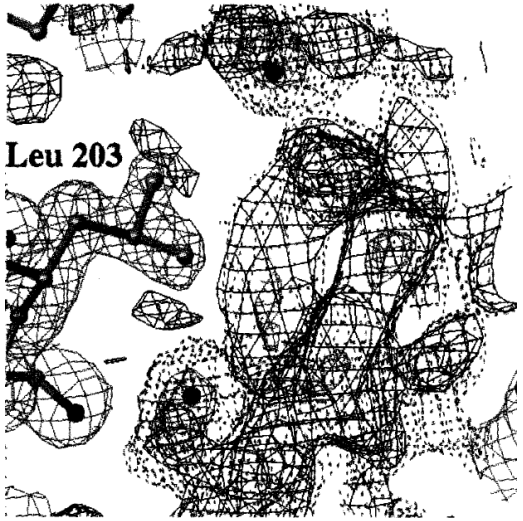


Multi-conformer
(Rfree 20.3%)

Ensemble
(Rfree 17.4%)

Diffuse solvent in MBP (1YTT)

Burling *et al.* (1996)



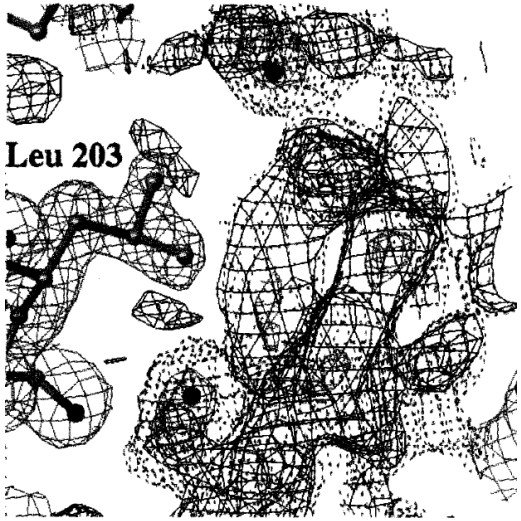
(Fig. 2B). The electron density around Leu^{203} in protomer A suggests a network of four to five partially disordered water molecules (23). The disorder is presum-

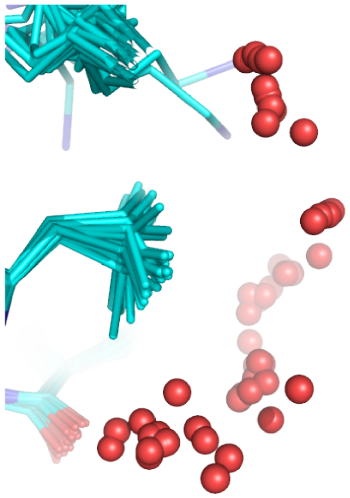
Published:

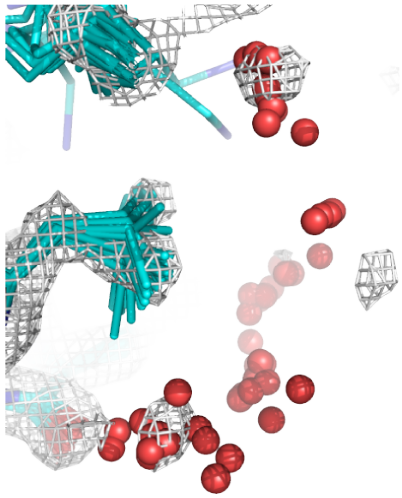
Experimental map
(1.4σ and 0.7σ)

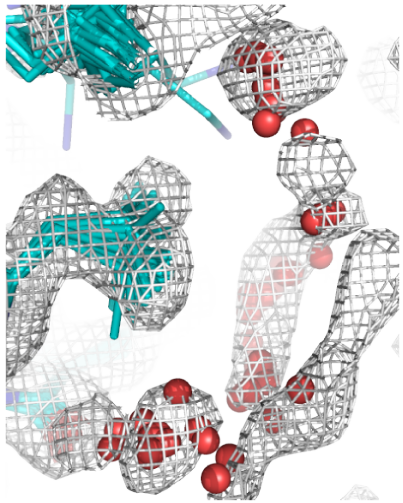
Diffuse solvent in MBP (1YTT)

Burling *et al.* (1996)









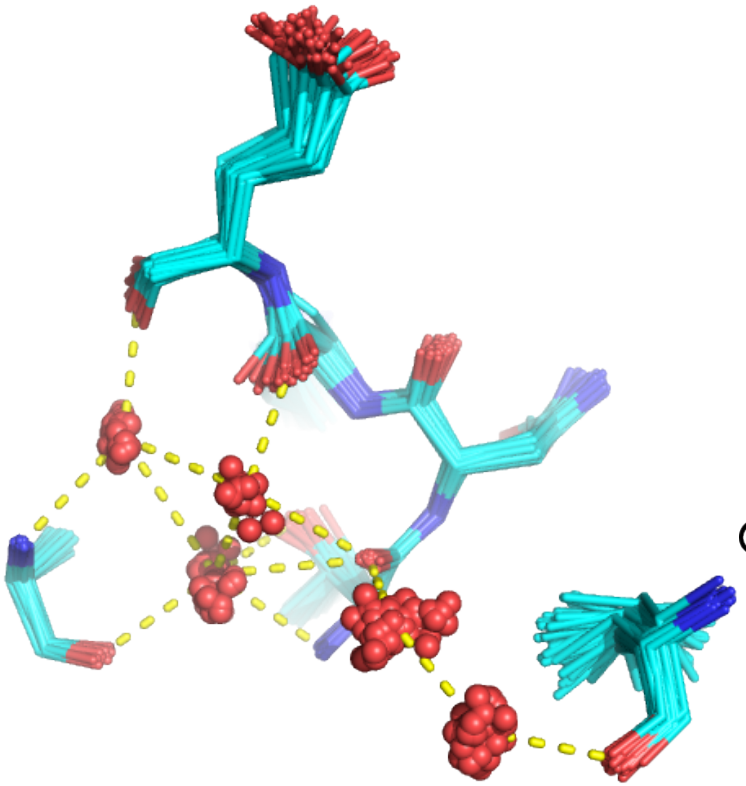
Published:

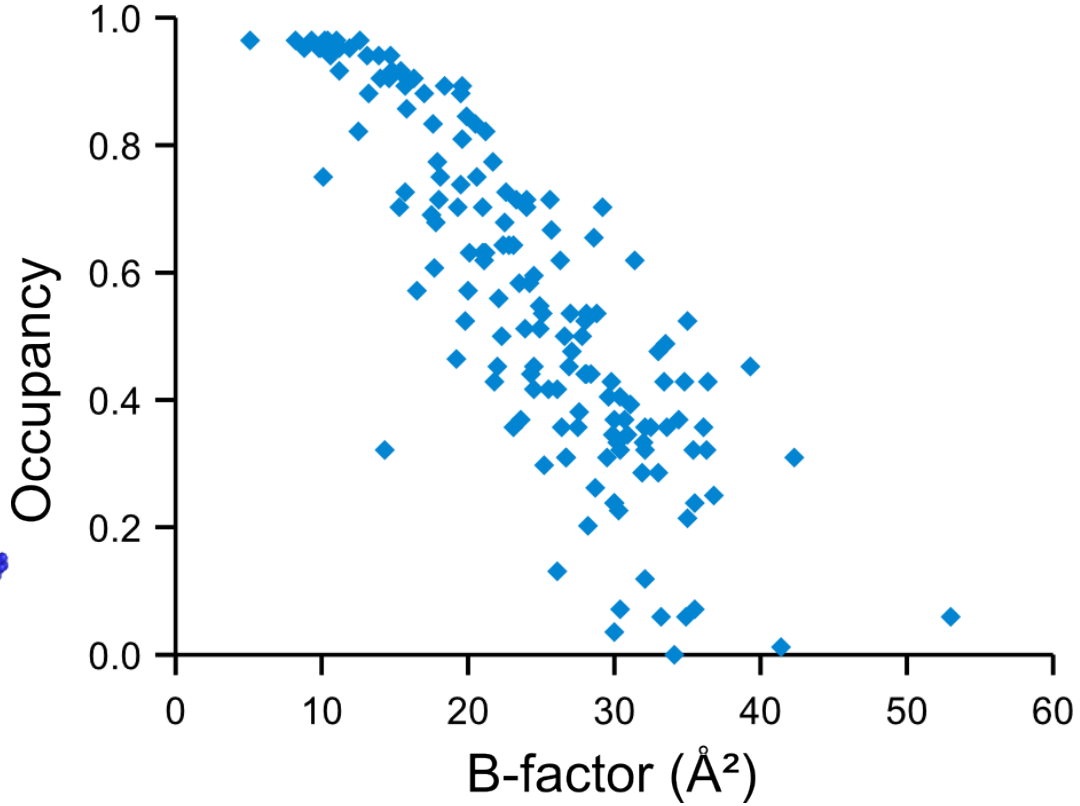
Experimental map
(1.4σ and 0.7σ)

Ensemble:

Experimental map (1.4 σ) (0.7 σ)

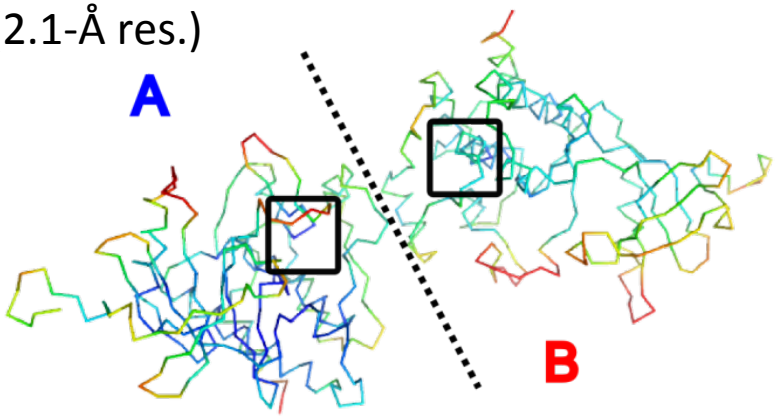
Explicit waters in MBP (1YTT)

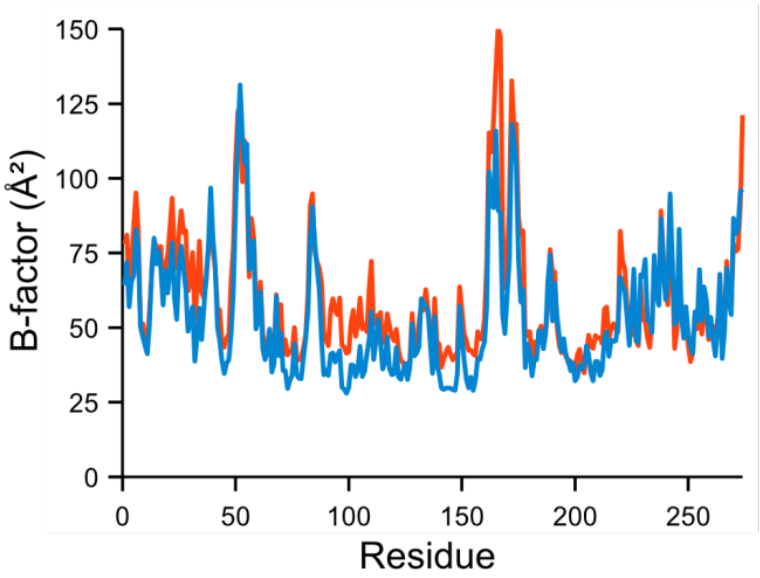


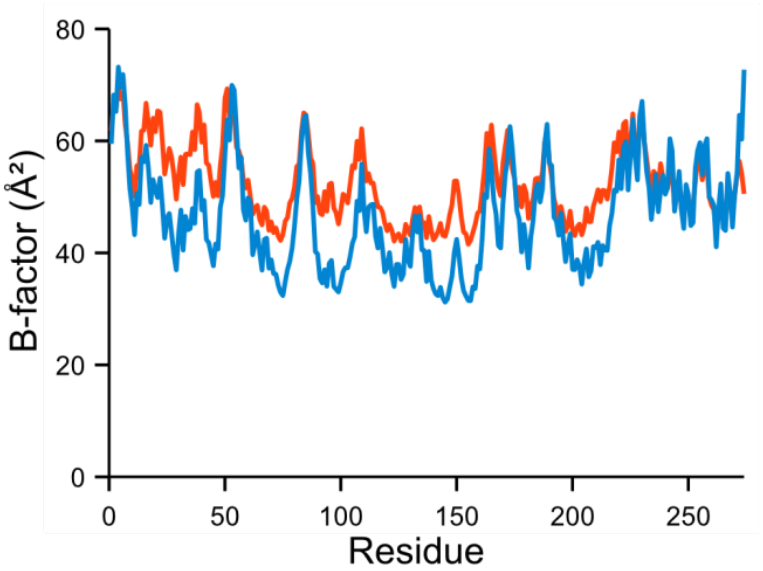


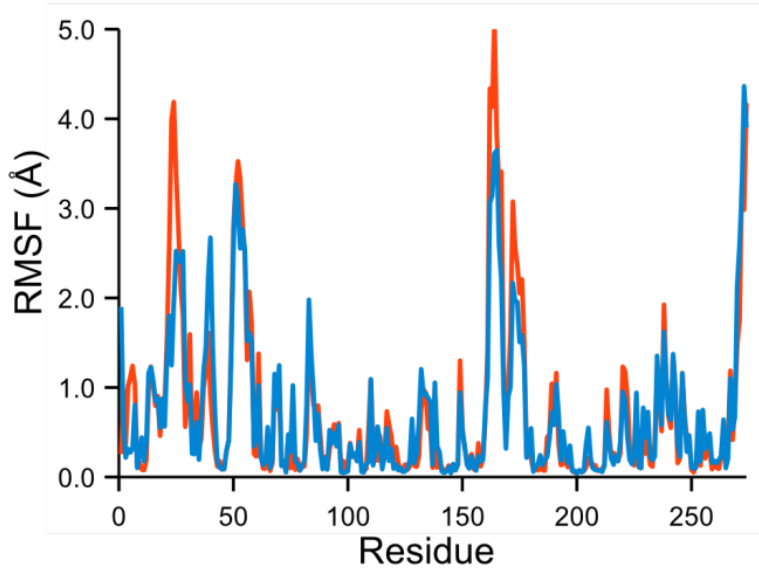
How similar are NCS copies?

1IEP (2.1-Å res.)

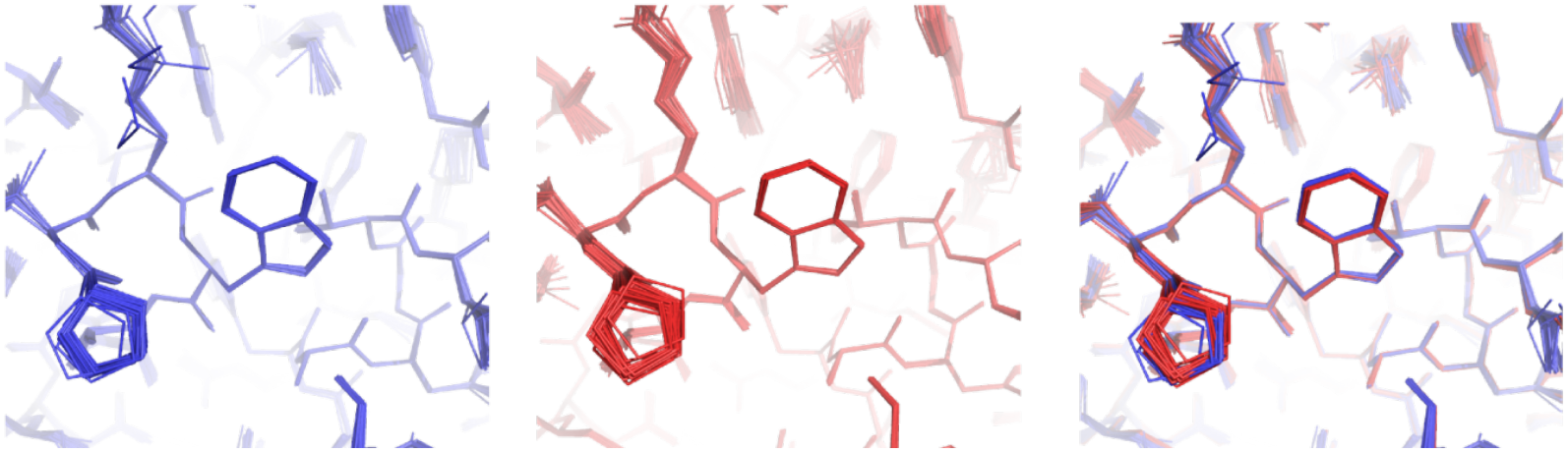








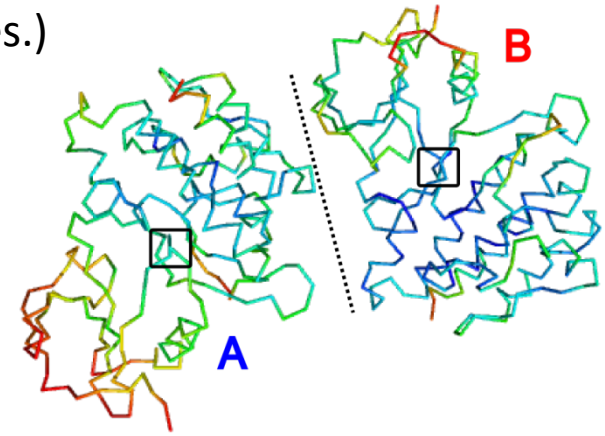
NCS copies show similar distributions

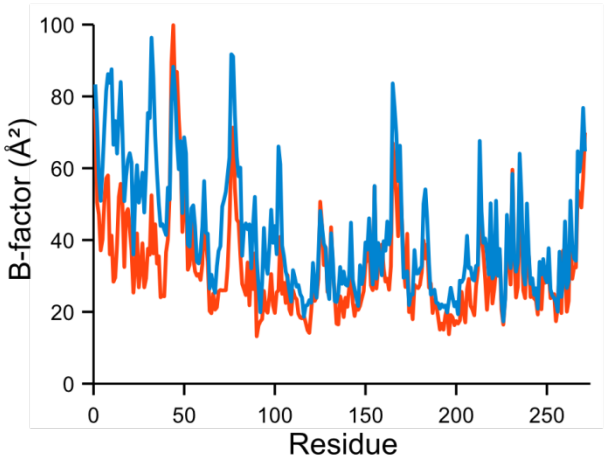


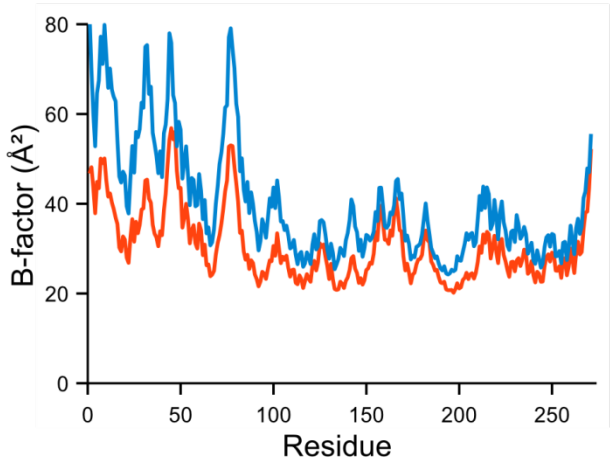
1IEP (2.1-Å res.)

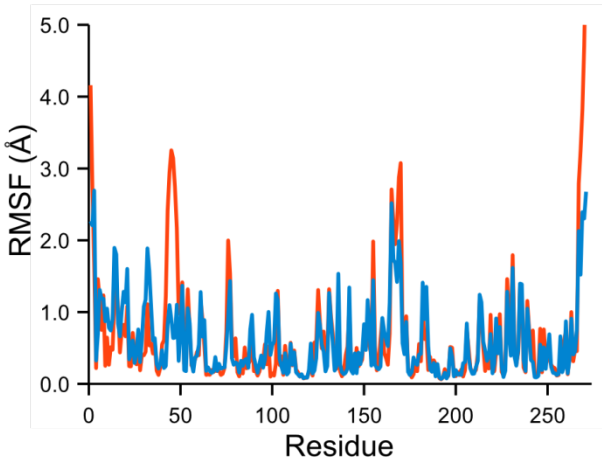
- Global TLS model accounts for differences in packing of copies
- Local fluctuations are similar between the two copies

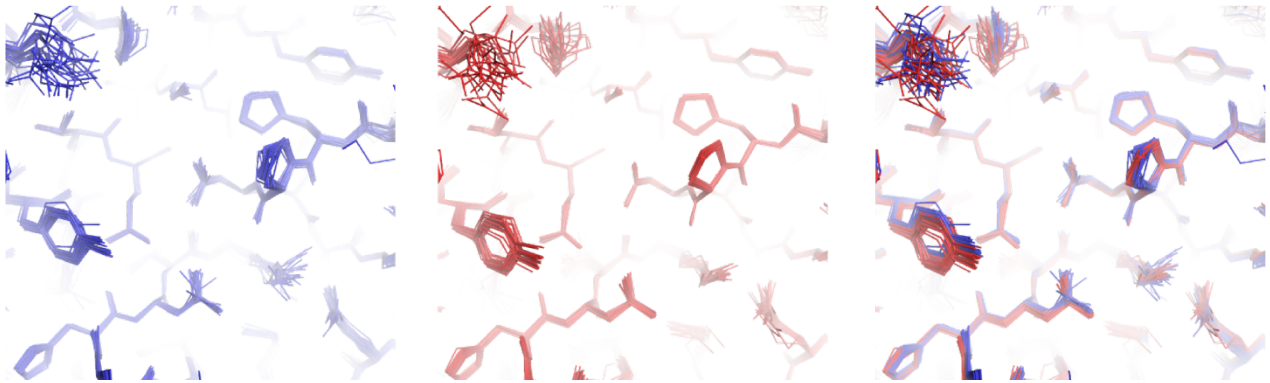
1M52 (2.6-Å res.)



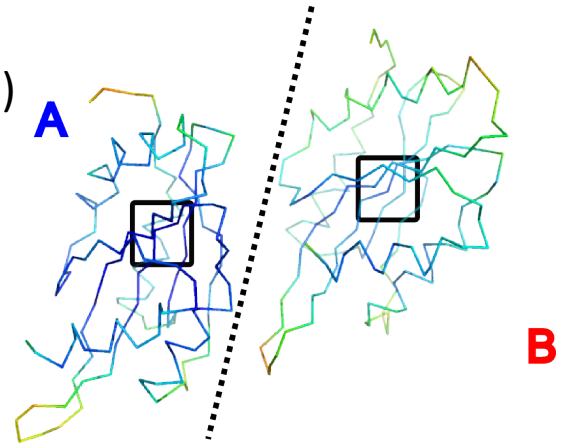


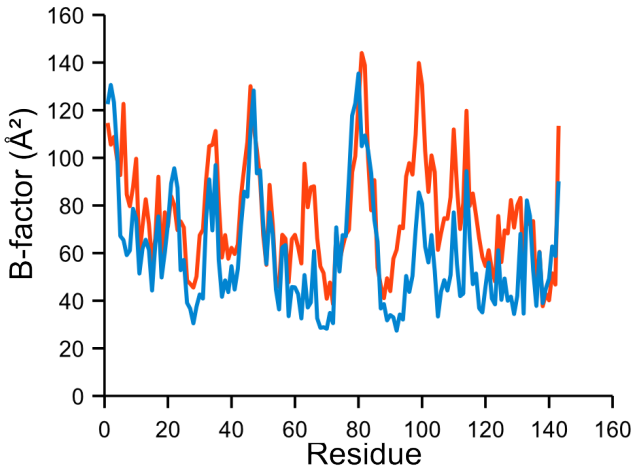


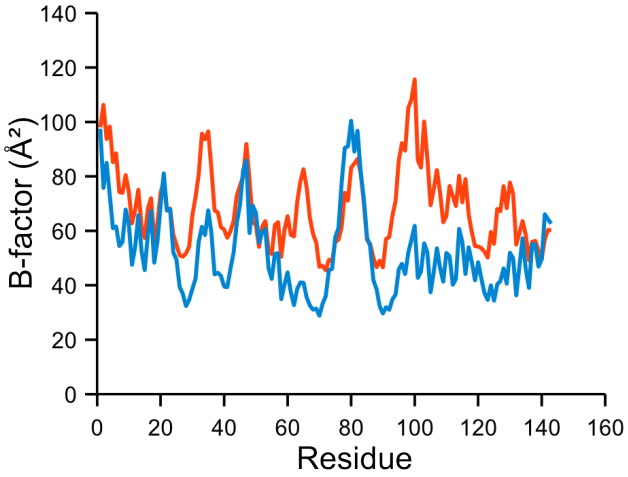


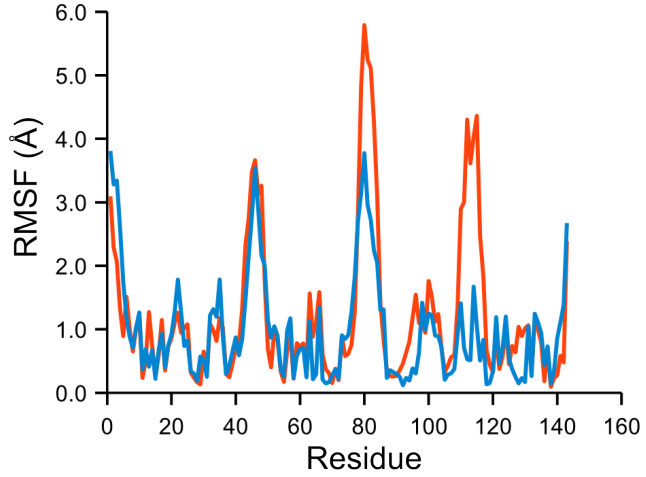


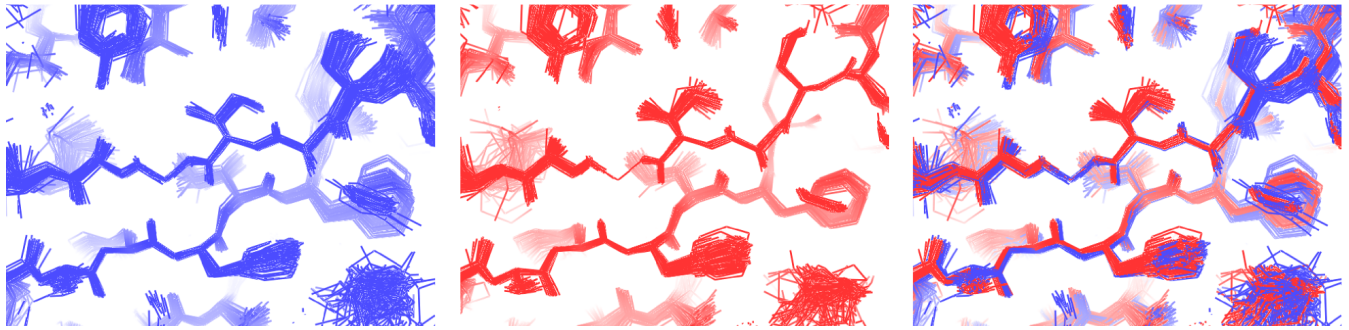
2XFA (2.1-Å res.)



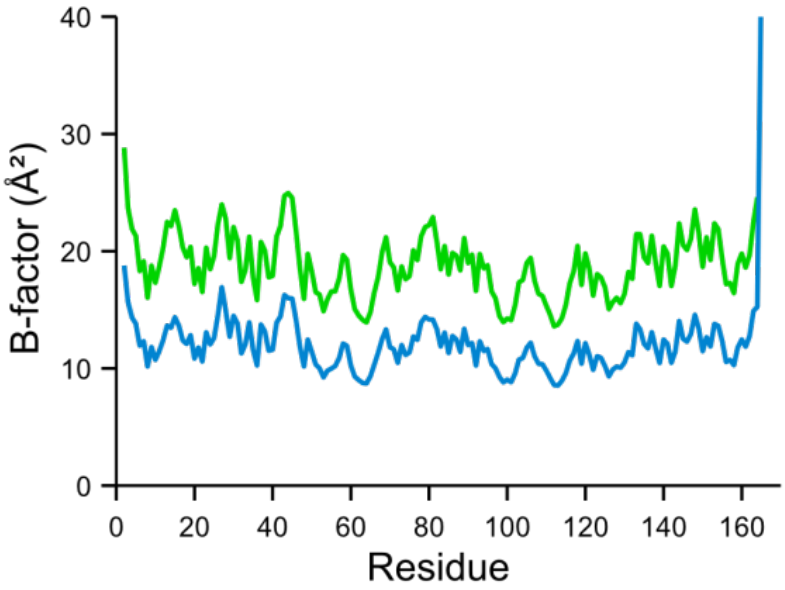


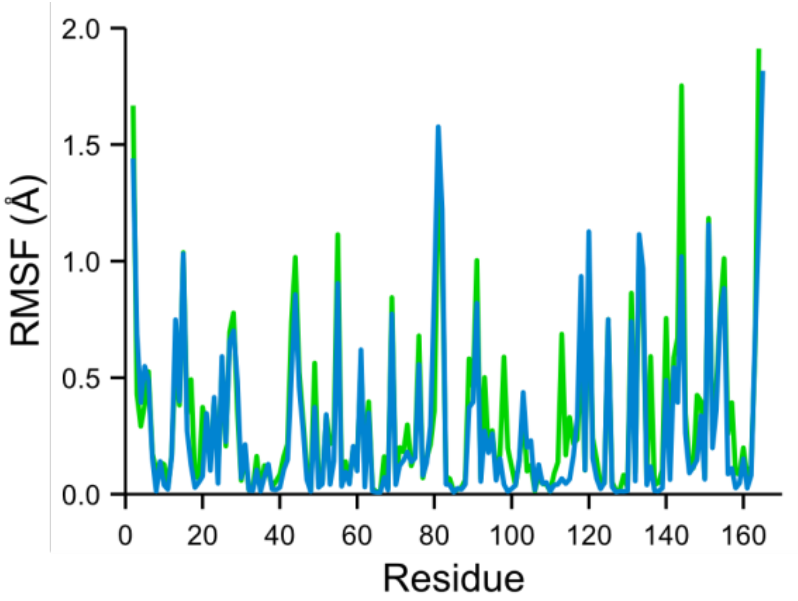






Isomorphous crystals at 100 and 288 K

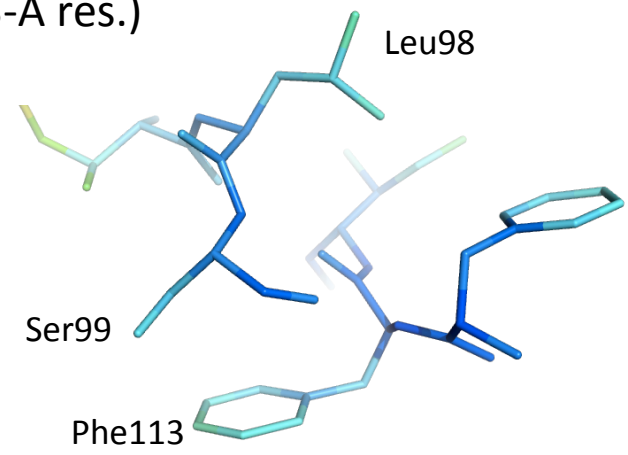


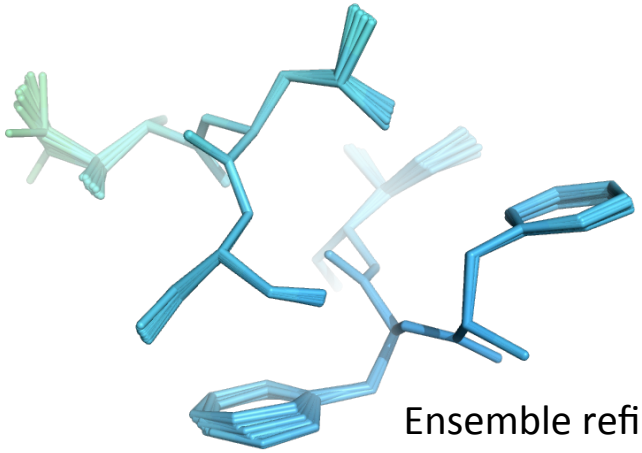


3K0N & 3K0M: Proline isomerase (Cyclophilin A)
Fraser *et al.* (2009)

Multi-conformers in active site

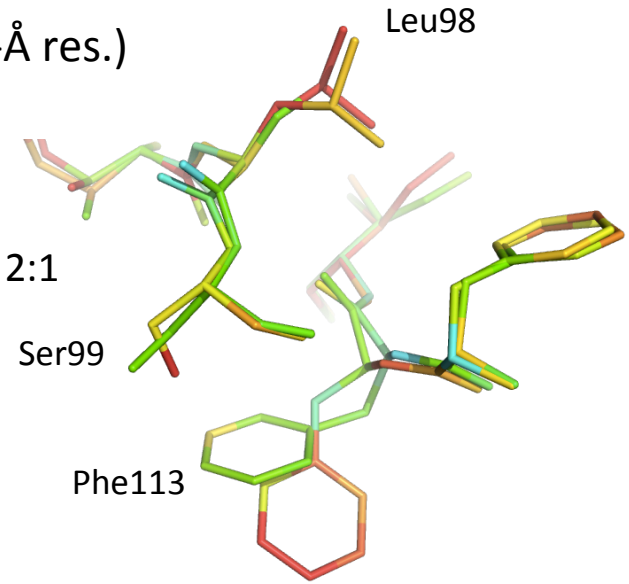
3K0M (1.3-Å res.)
100 K



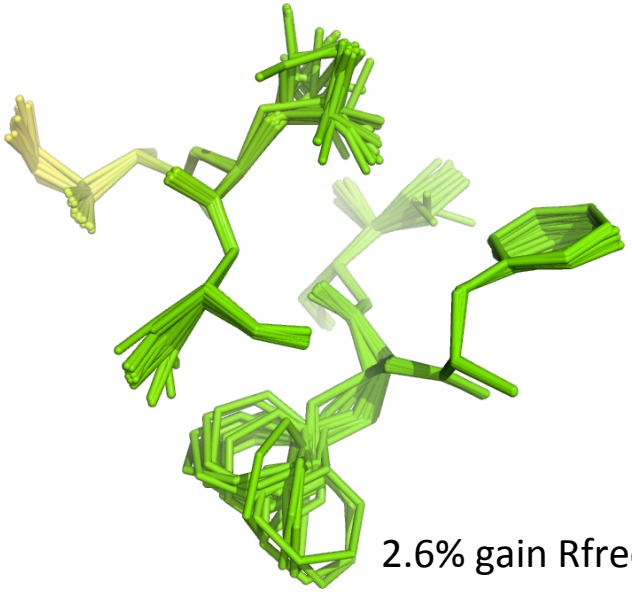


Ensemble refinement:
1.3% gain Rfree

3K0N (1.4-Å res.)
288 K



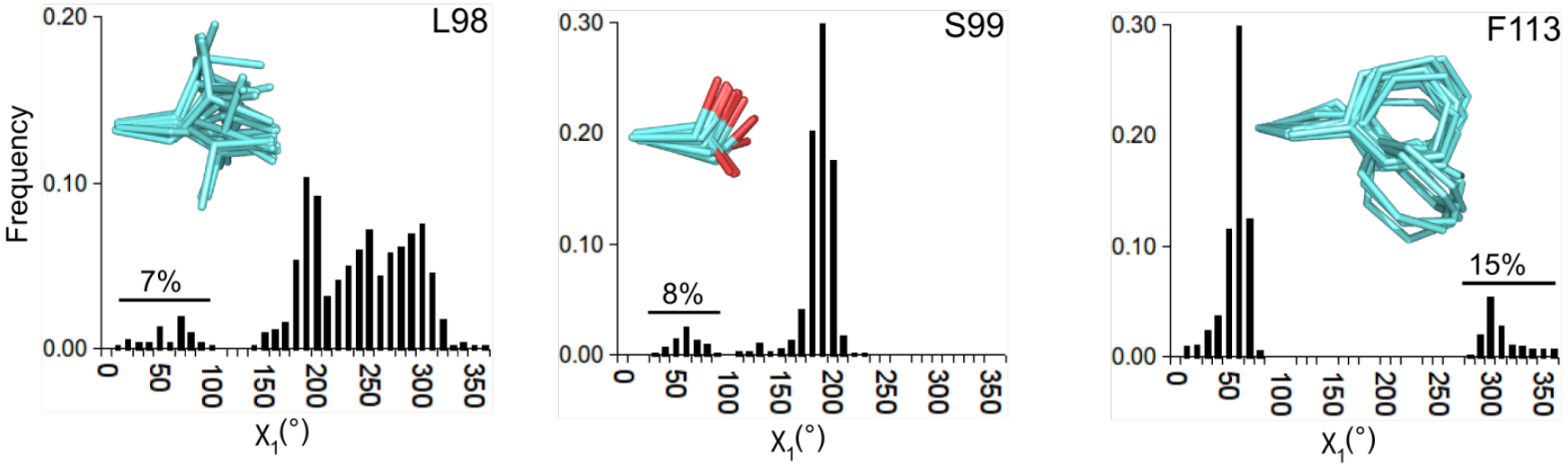
B-factor
5-25Å²



2.6% gain Rfree

Fraser *et al.* (2009), Eisenmesser *et al.* (2005)

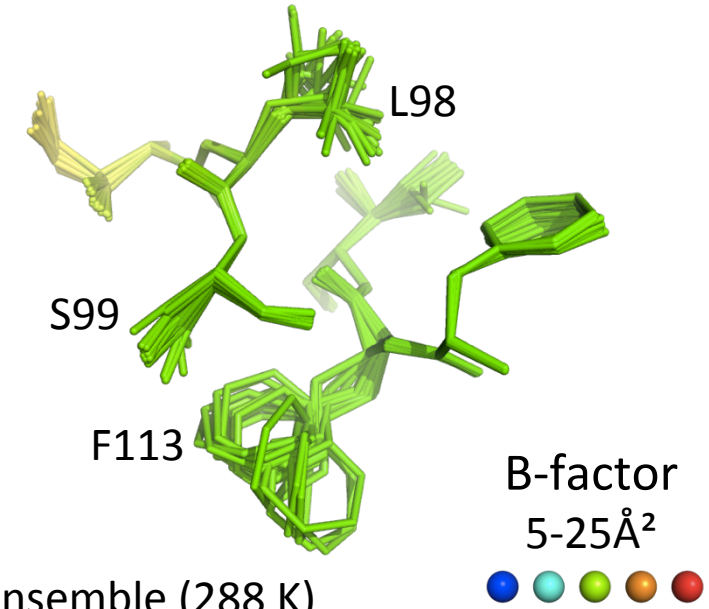
Occupancies agree with NMR data



- NMR relaxation dispersion (283 K)

  L98, S99, F113

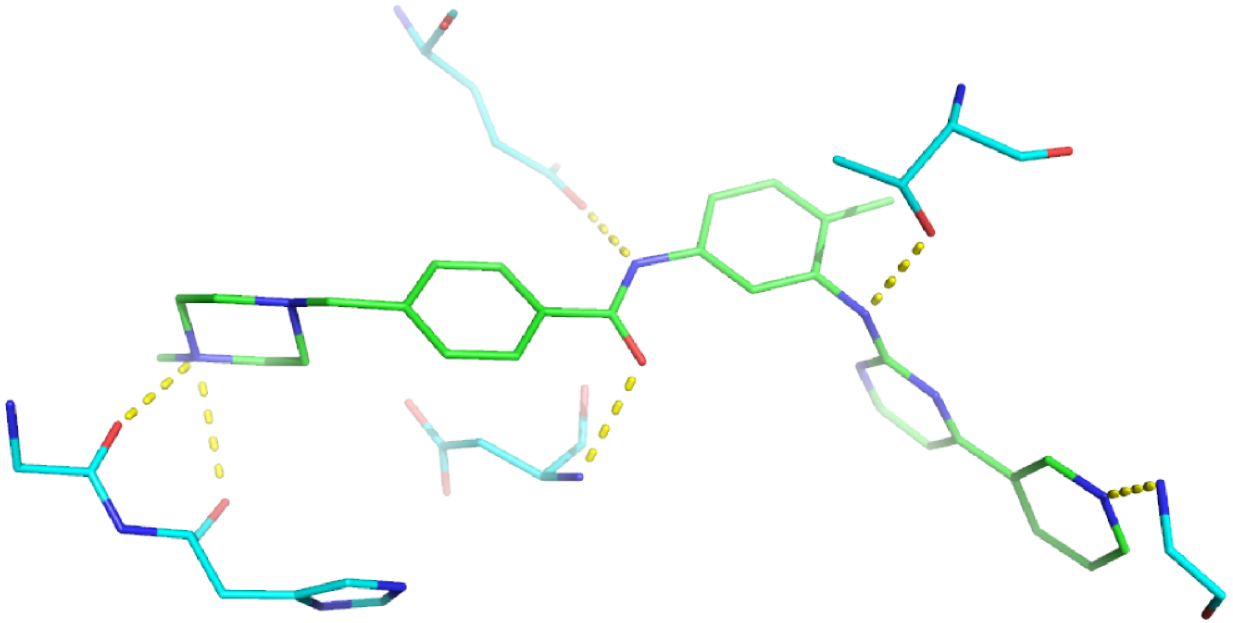
  Minor population ~10%



Ensemble (288 K)

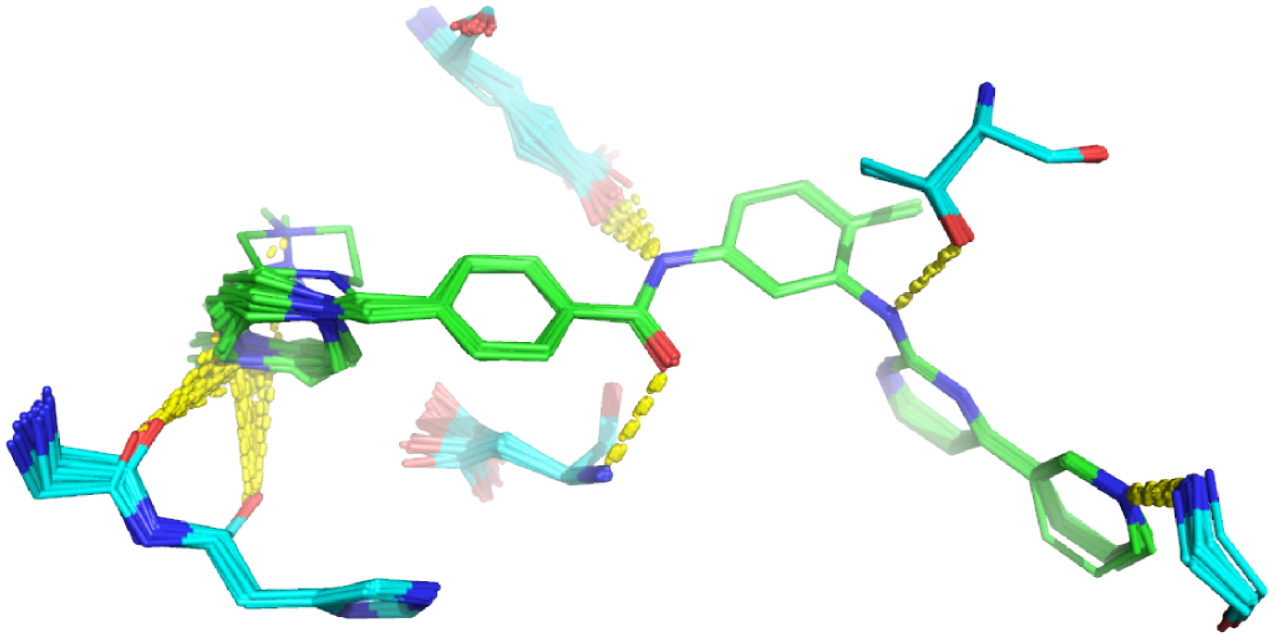
3K0N & 3K0M: Proline isomerase
Fraser *et al.* (2009), Eisenmesser *et al.* (2005)

Imatinib-ABL Tyrosine Kinase (1IEP)

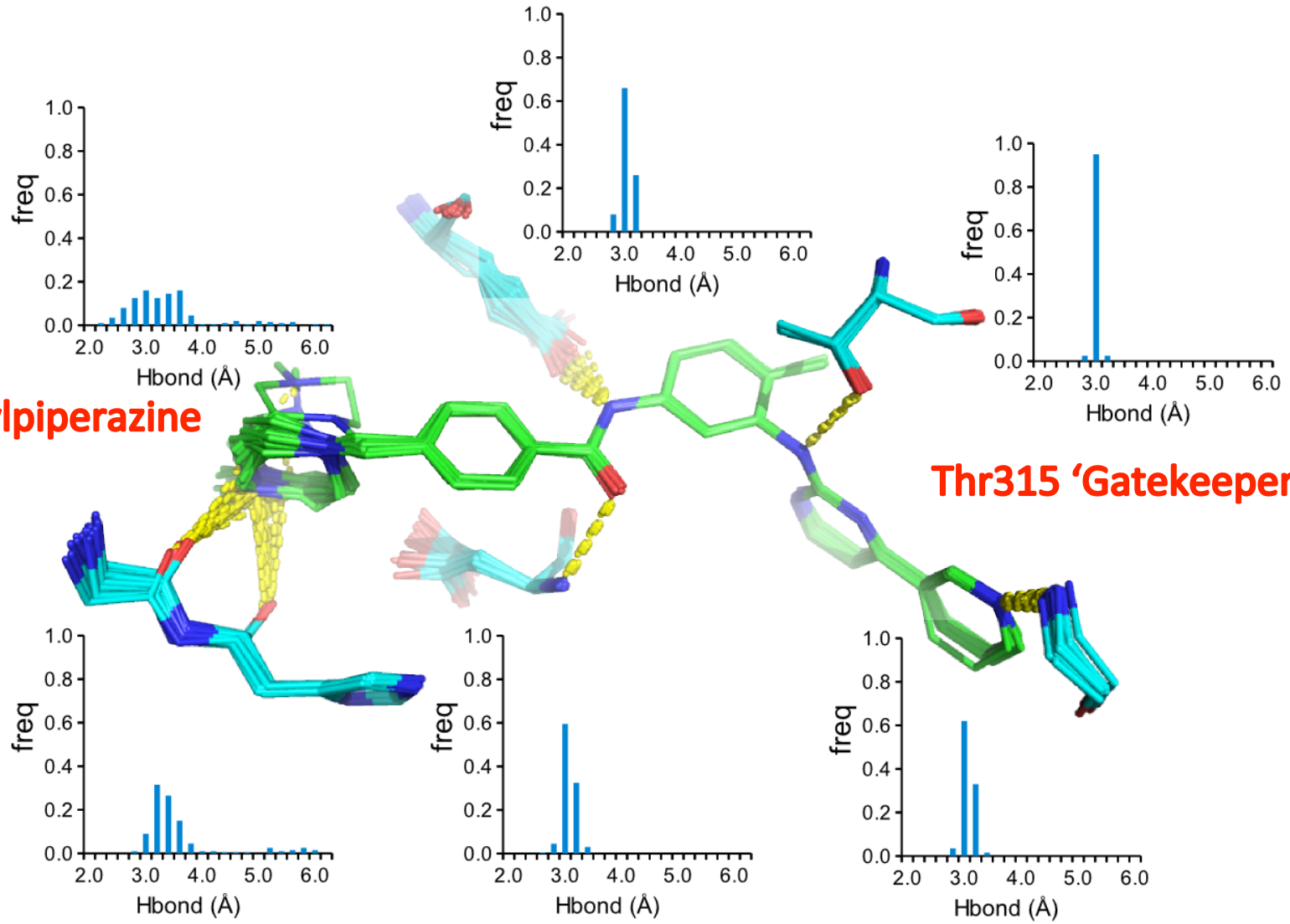


Nagar *et al.* (2002)

Imatinib-ABL Tyrosine Kinase (1IEP)

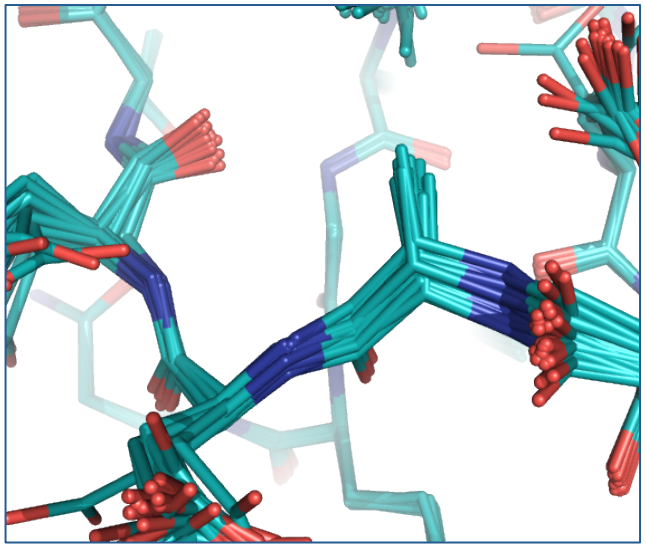


Imatinib-ABL Tyrosine Kinase (1IEP)

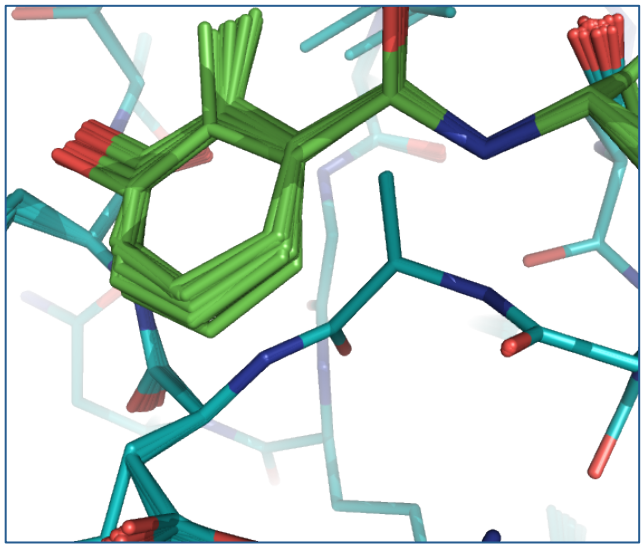


Nagar *et al* (2002), Johnson (2009)

Inhibitor binding to HIV protease



apo-protease (2PC0)



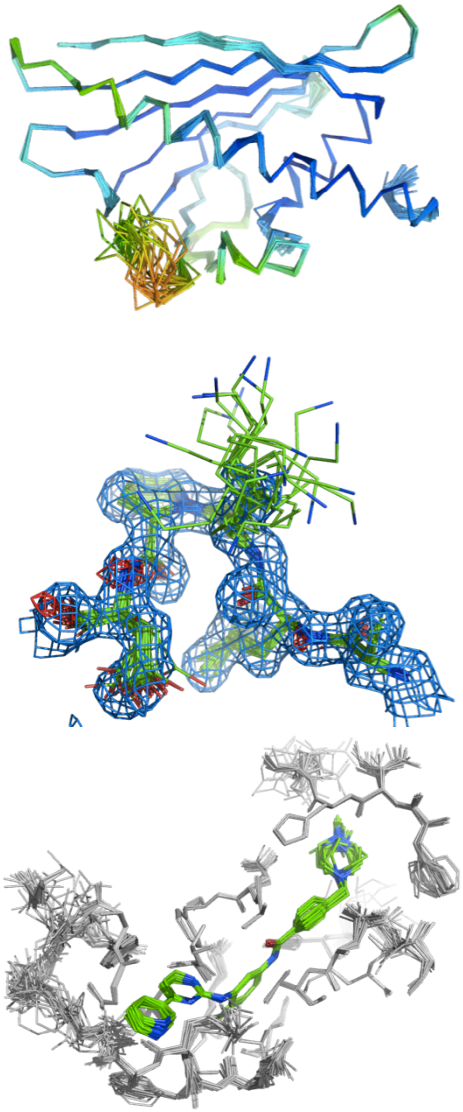
complex with JE-2147 (1KZK)

Comparative analysis of atomic flexibility:

- Isomorphous crystals eliminate differences due to Xtal contacts
- Non-isomorphous crystals allow evaluation of Xtal contact effects

Conclusions

- Global disorder modeled by TLS and local disorder by MD
- Ensemble refinement improves Rfree and electron density maps
- Suitable for a broad resolution range (1Å – 3Å)
- NCS copies show very similar fluctuations
- Clear representation of local disorder / uncertainty
- Distribution of atom positions allows further structural analysis
- Resolve the finer details of protein structure(s)



Acknowledgements

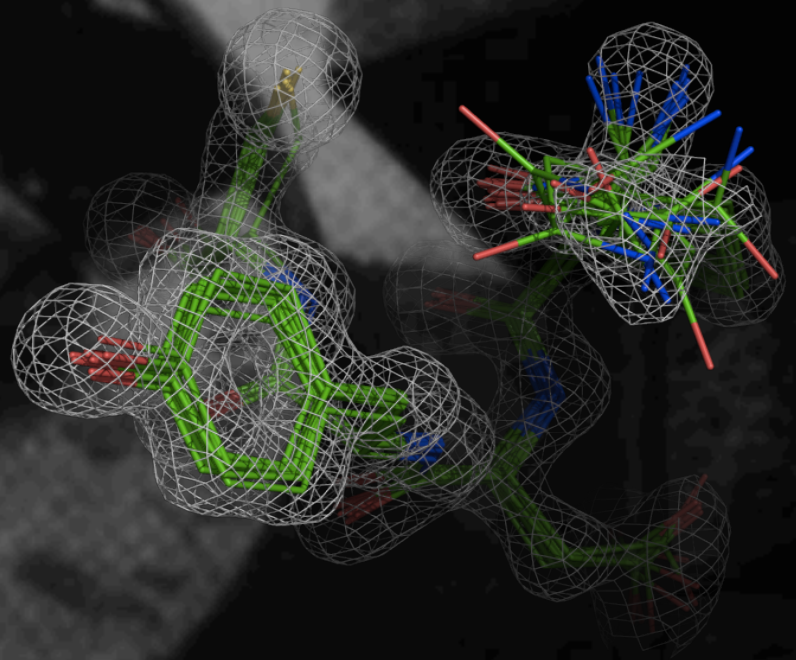
Piet Gros

Gros Laboratory

Pavel Afonine, Paul Adams

PHENIX Developers

Funding: Utrecht University / NWO



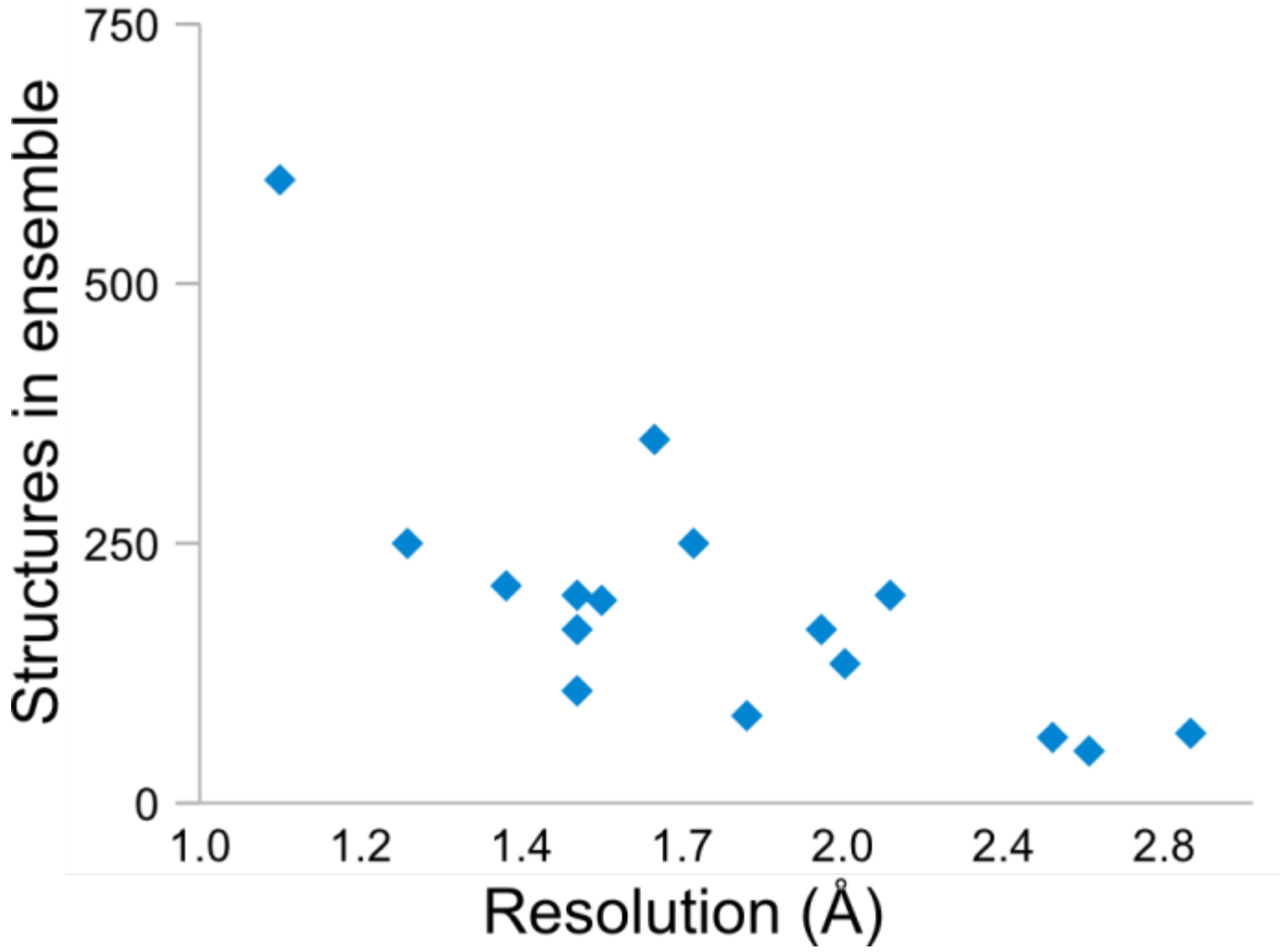
Universiteit Utrecht

BIJVOET CENTER

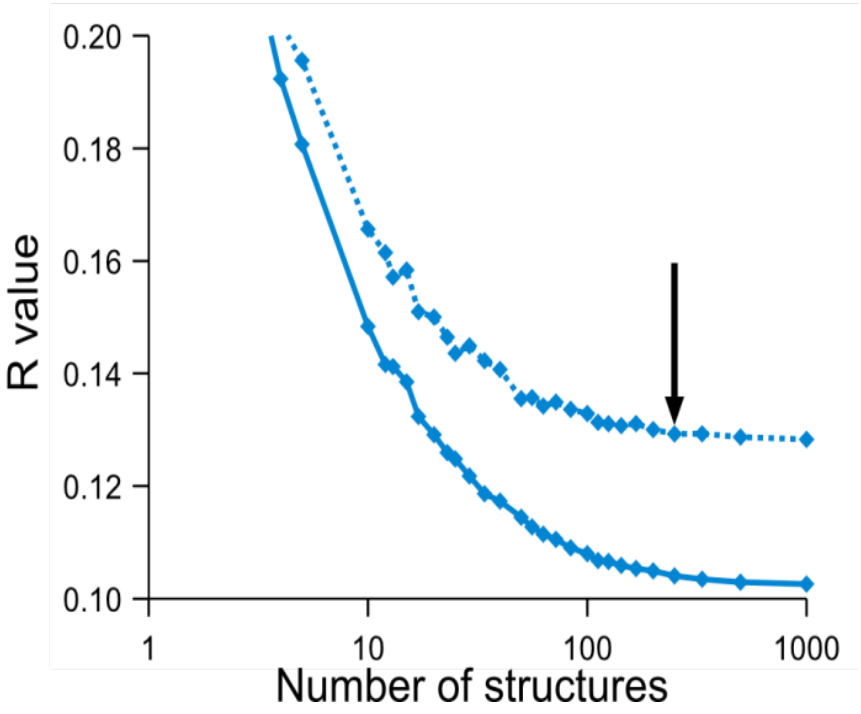
NWO

Phenix

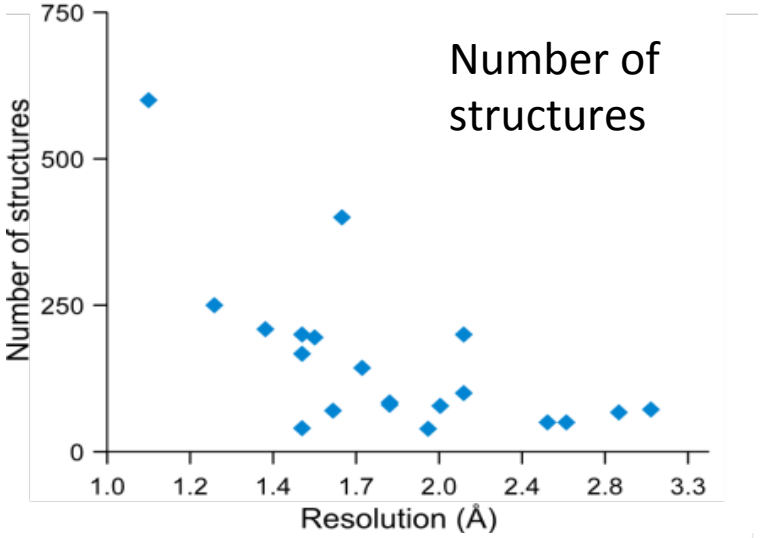
Number structures in ensemble vs resolution

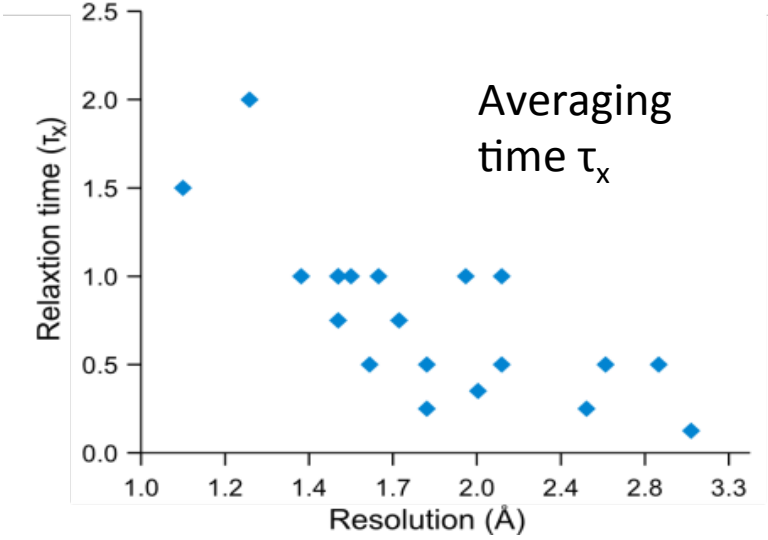


Number of structures in the ensemble

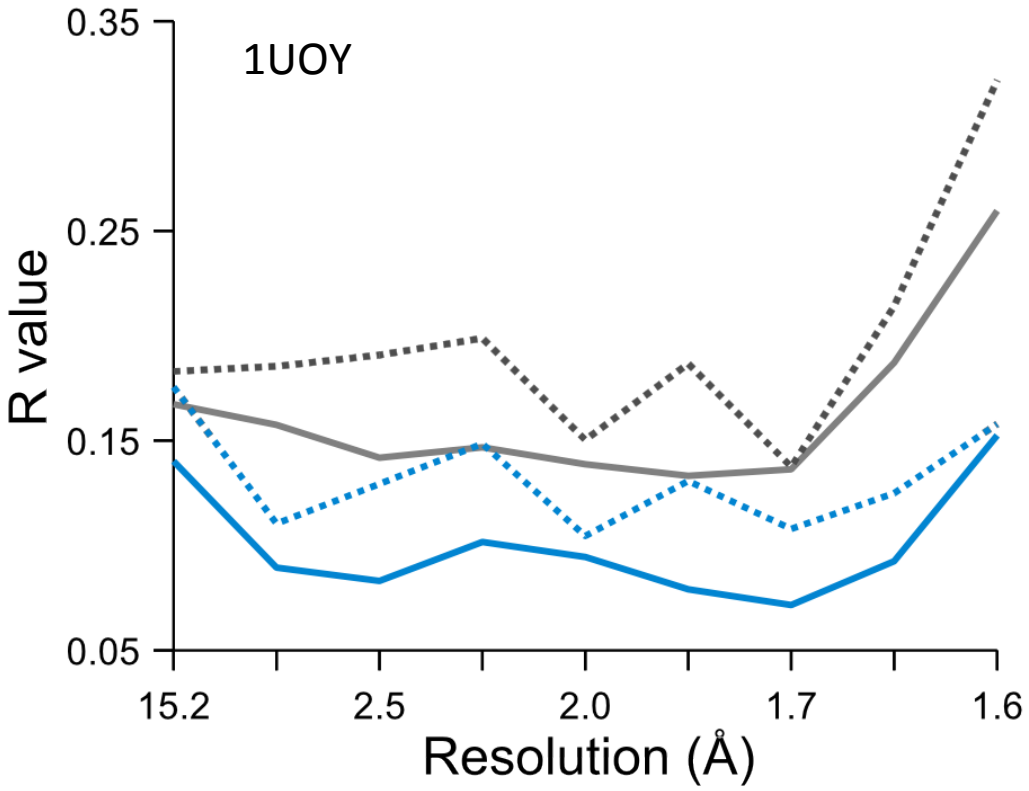


Structures taken equidistant in time to reproduce the Rwork within 0.1%

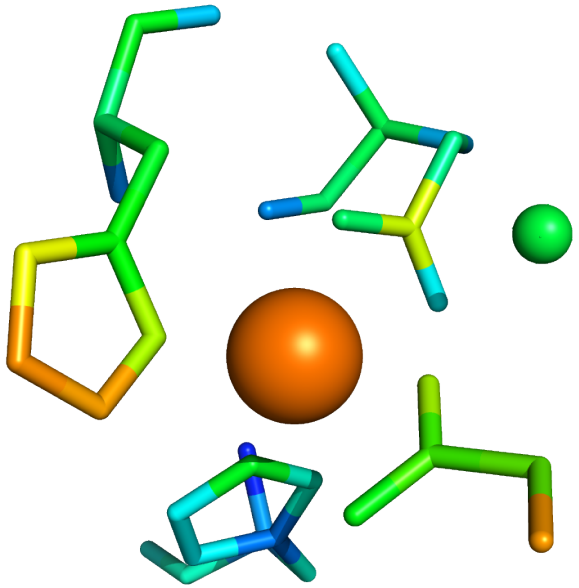




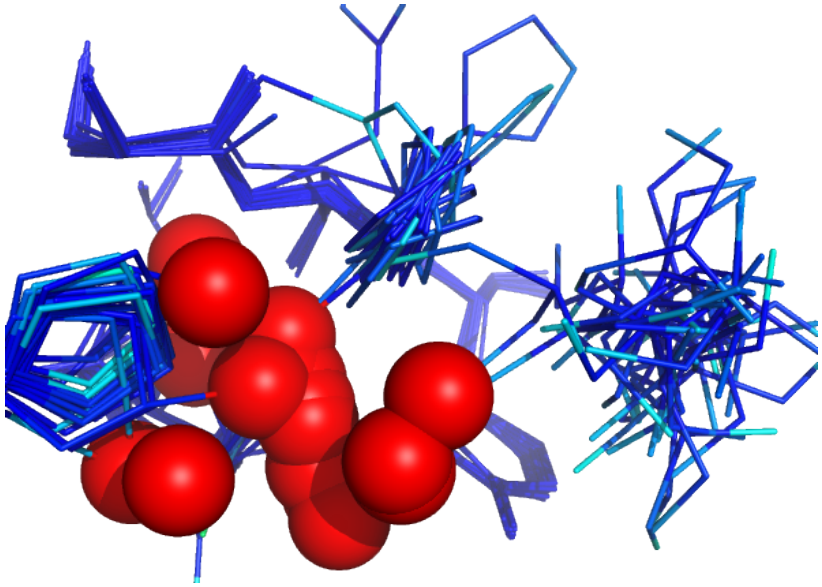
R-factor by resolution shell



Partial ligand/ion binding



B-factor of Cd^{2+}-ion more than twice its surrounding

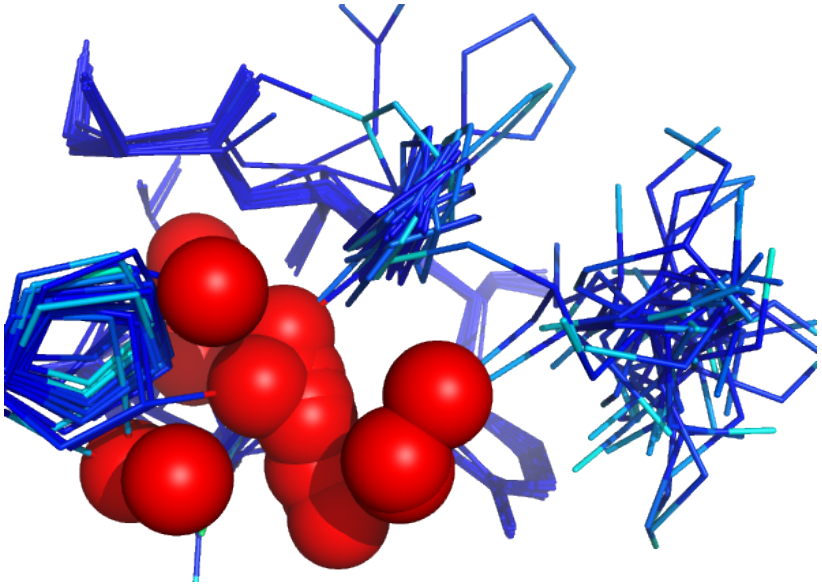


Simulation run at Q=1 for Cd^{2+}-ion (atoms coloured by kinetic energy)

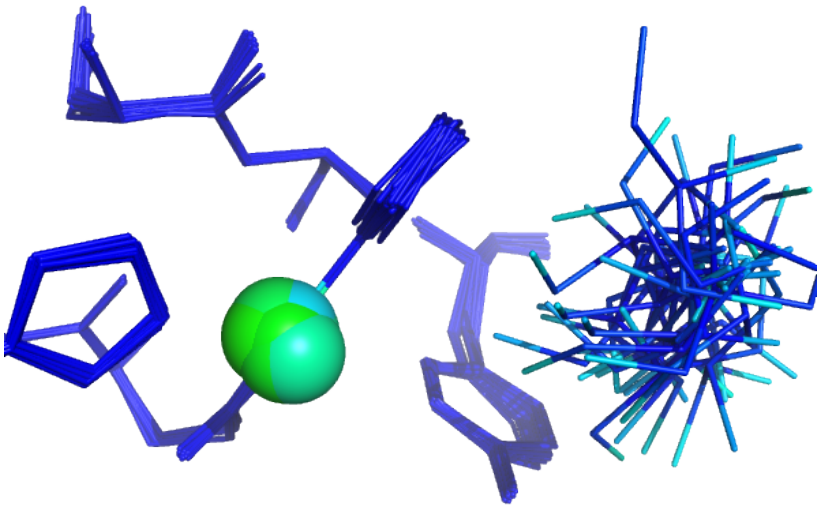
- High B-factor ligand/ion may indicate partial occupancy

Partial ligand/ion binding

Cd^{2+}: Q=1

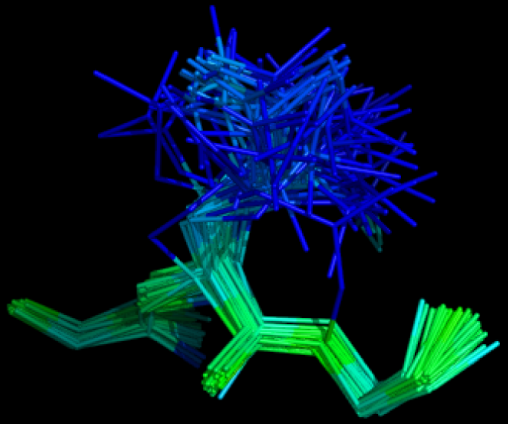


Cd^{2+}: Q=0.7

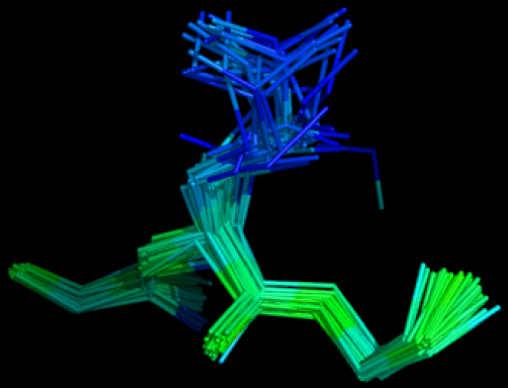


- Heat map (kinetic energy) is validation tool for Ensemble Refinement

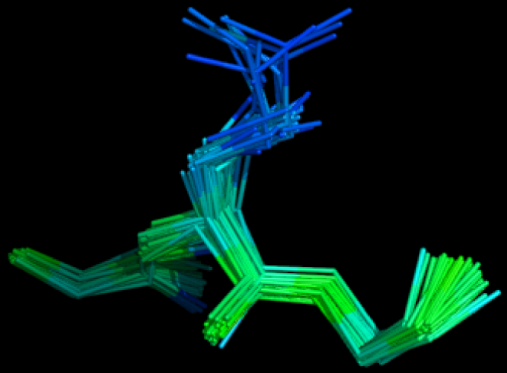
Ensemble



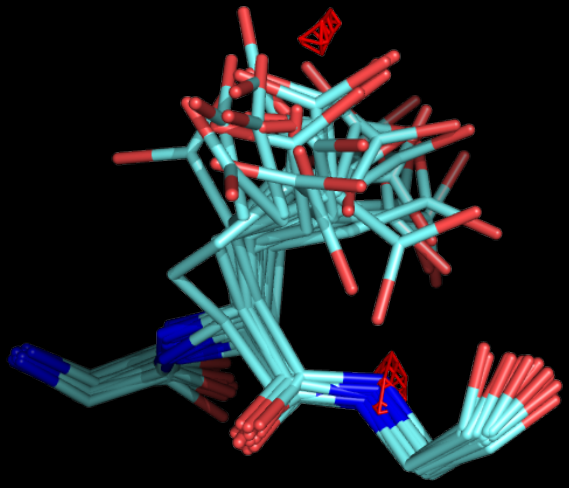
0–10 σ



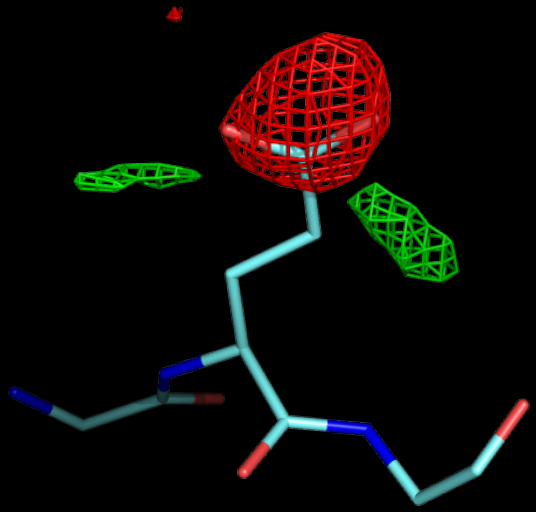
0.5–10 σ



1–10 σ



Ensemble

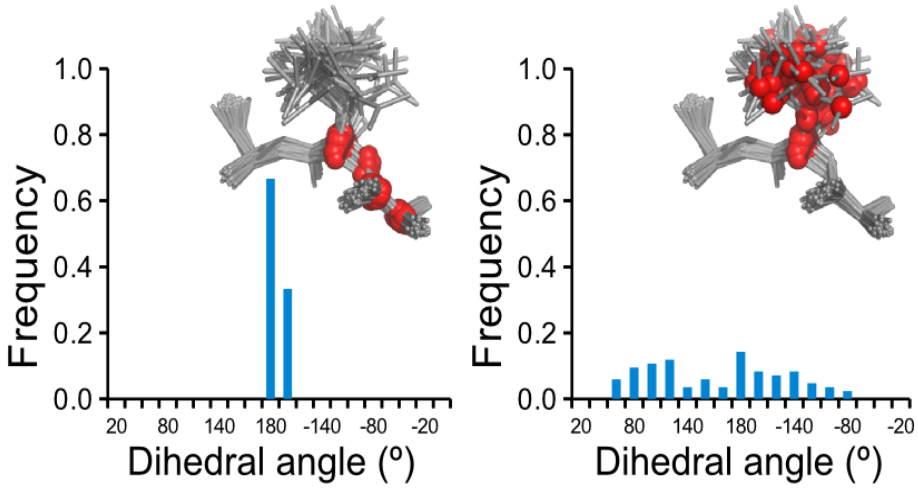


Single-structure

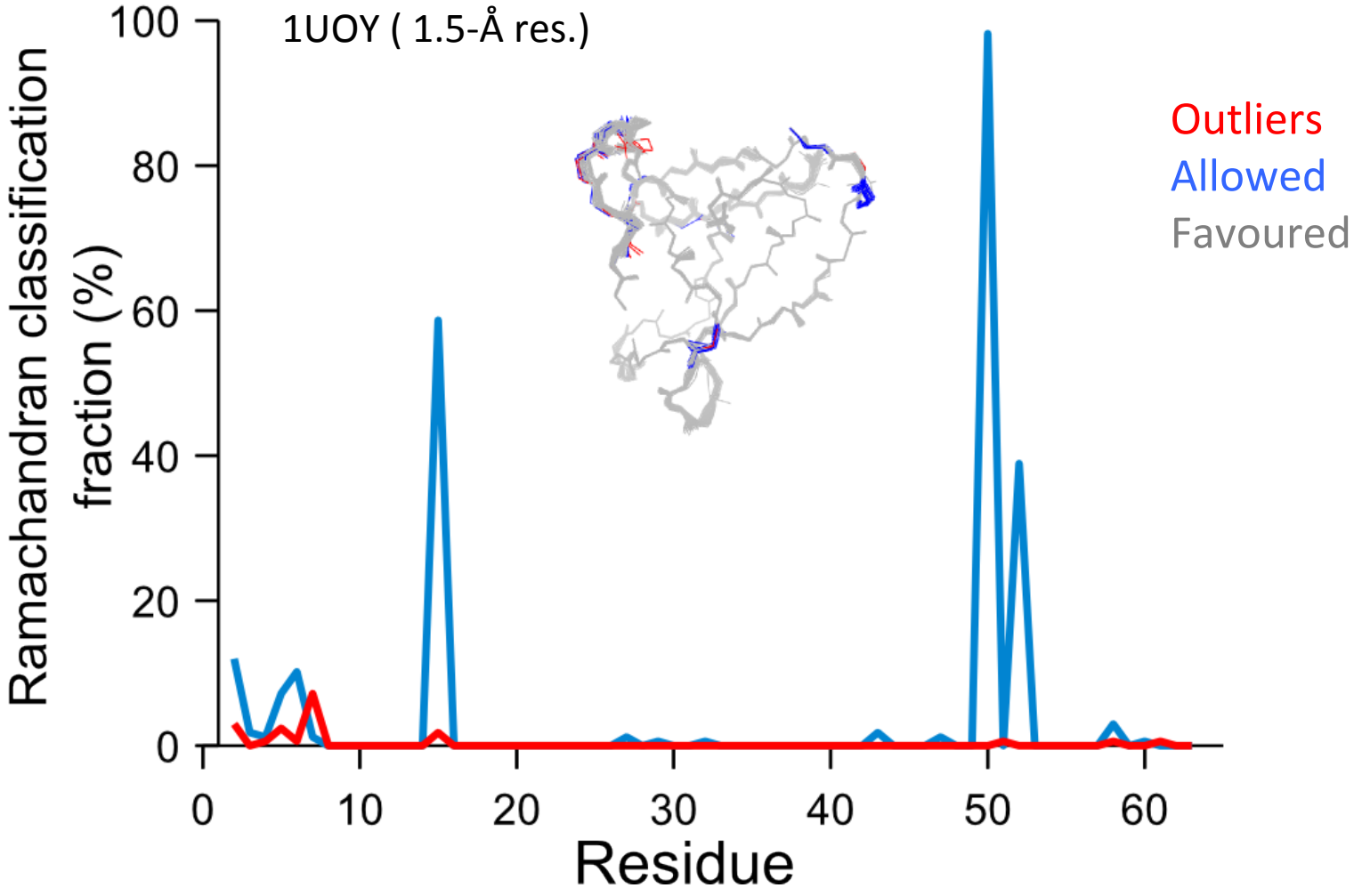
Geometric validation

	Bonds	Angles	Dih. angles
Single structure	0.010 Å	1.23°	15.1°
Ens: $\sqrt{\langle(x_{ideal}-\langle x_{model}\rangle)^2\rangle}$	-0.002	-0.28	-6.5
Ens: $\sqrt{\langle(x_{ideal}-x_{model})^2\rangle}$	+0.002	+0.30	+4.0

(statistics for all 20 cases)

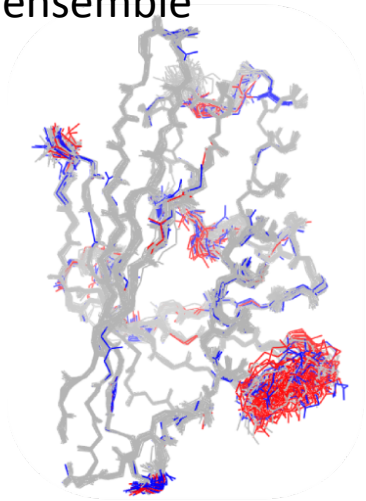


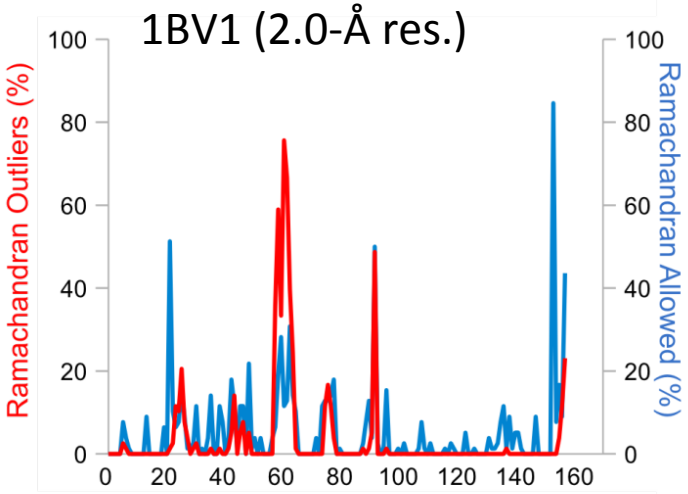
Ramachandran analysis

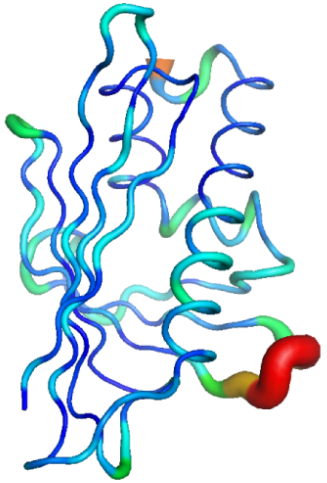


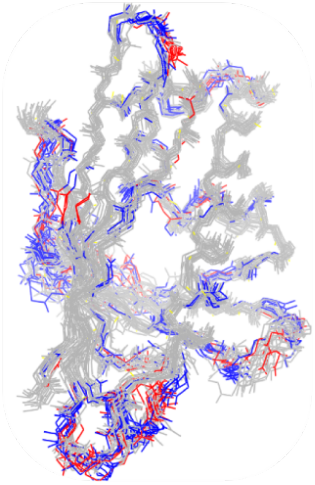
Outliers occur more in flexible regions

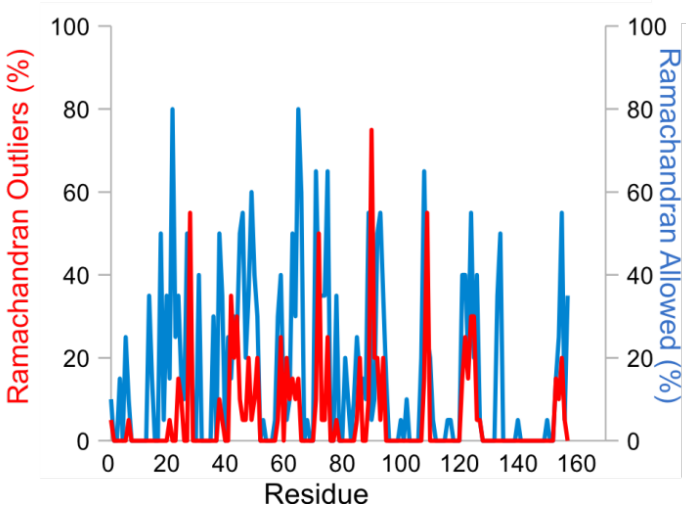
X-ray ensemble





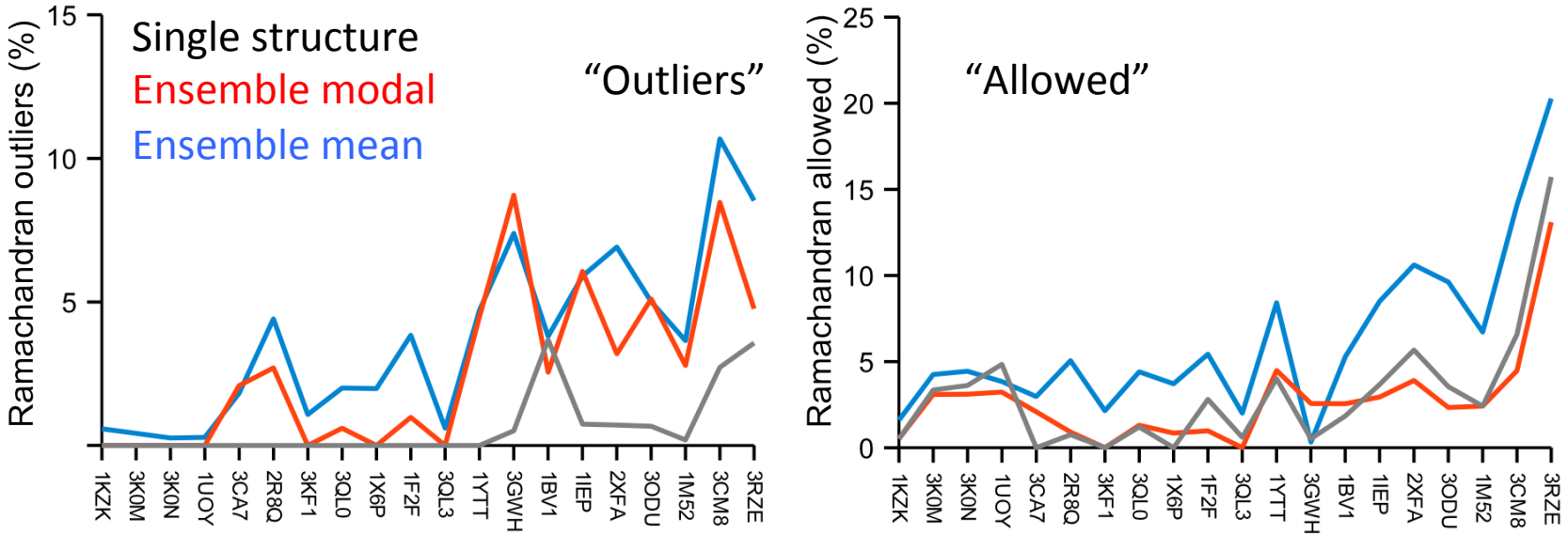






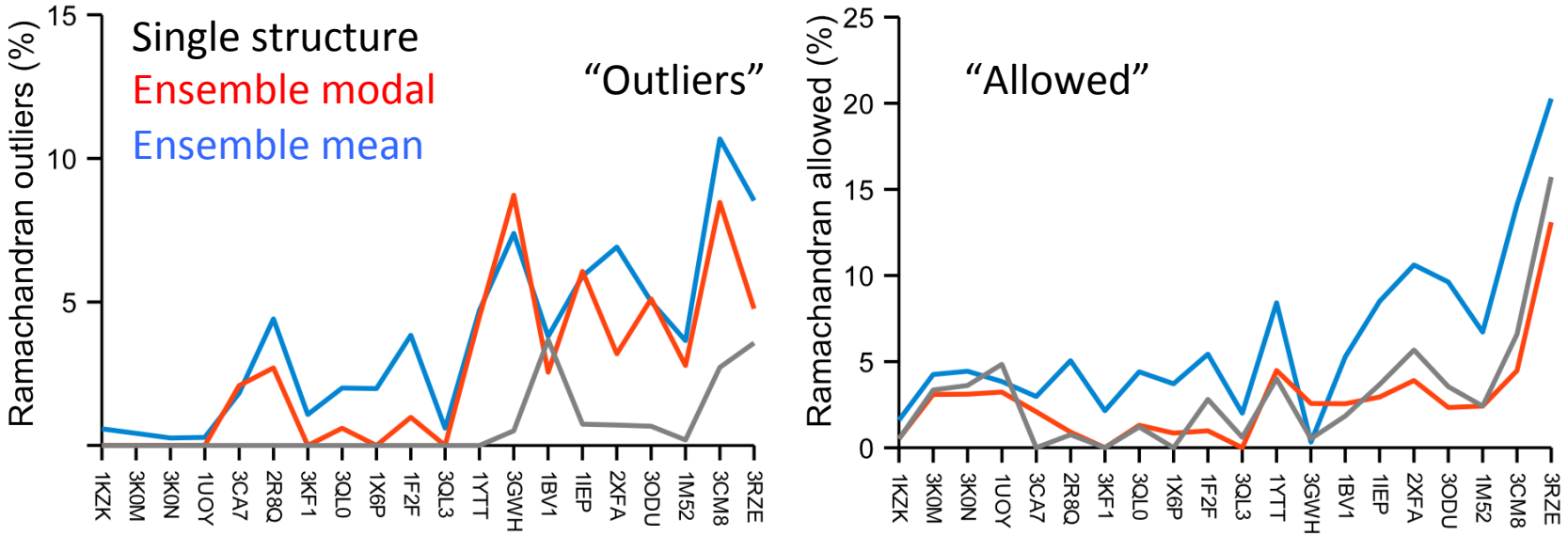
NMR ensemble

Ramachandran deviations by resolution



- More outliers are observed at lower resolution
- Geometric quality correlates with Rfree
- "Best" run selected by Rfree

Ramachandran deviations by resolution



- More outliers are observed at lower resolution
- Geometric quality correlates with Rfree
- "Best" run selected by Rfree