

Q|R – Quantum Refinement: progress update Jan 15-Mar 24, 2016

**Pavel Afonine
Mark Waller
Min Zheng**

**March 25th, 2016
ICQMS, Shanghai University, PR China**

Refinement in a nutshell

- Refinement - optimization of T w.r.t. model parameters:

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$

From
diffraction
experiment

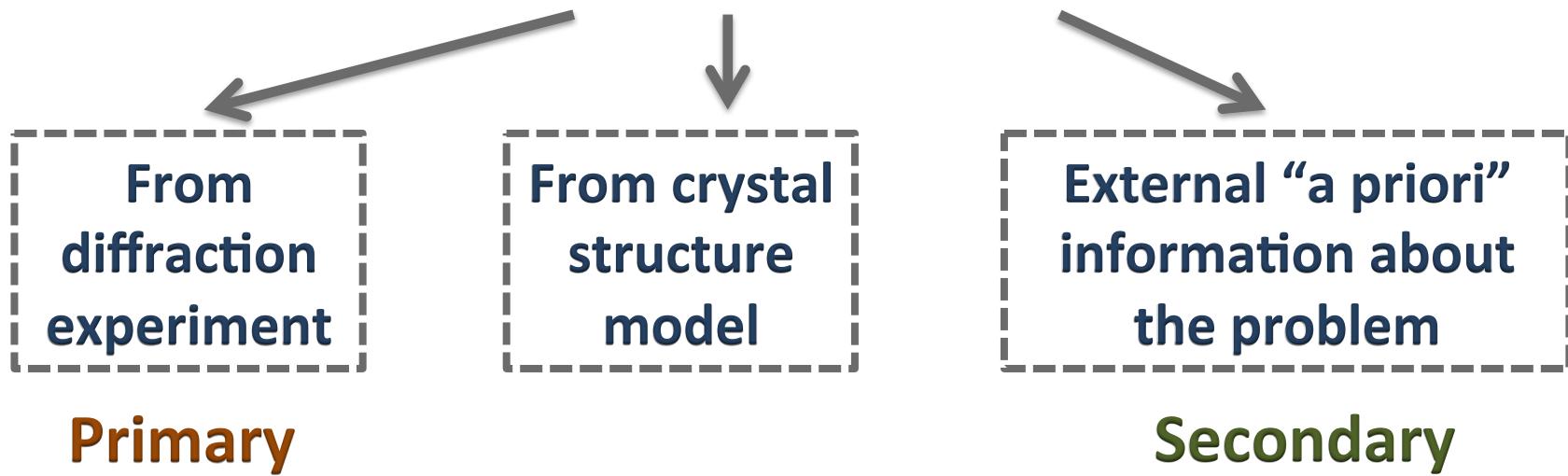
From crystal
structure
model

External “a priori”
information about
the problem

Refinement in a nutshell

- Refinement - optimization of T w.r.t. model parameters:

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$



We want to fit model to data as good as possible without violating available “a priori” knowledge

Refinement in a nutshell

- Refinement - optimization of T w.r.t. model parameters:

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$



$$T_{\text{RESTRAINTS}} = T_{\text{BOND}} + T_{\text{ANGLE}} + T_{\text{DIHEDRAL}} + T_{\text{PLANARITY}} + T_{\text{NONBONDED}} + T_{\text{CHIRALITY}}$$

$$T_{\text{BOND}} = \sum_{\text{all bonded pairs}} w(d_{\text{ideal}} - d_{\text{model}})^2, \quad T_{\text{ANGLE}} = \dots, \dots$$

We call it *geometry restraints*...

... also known as *potential function*, *energy term*, *force field* etc.

Refinement in a nutshell

- Refinement - optimization of T w.r.t. model parameters:

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$



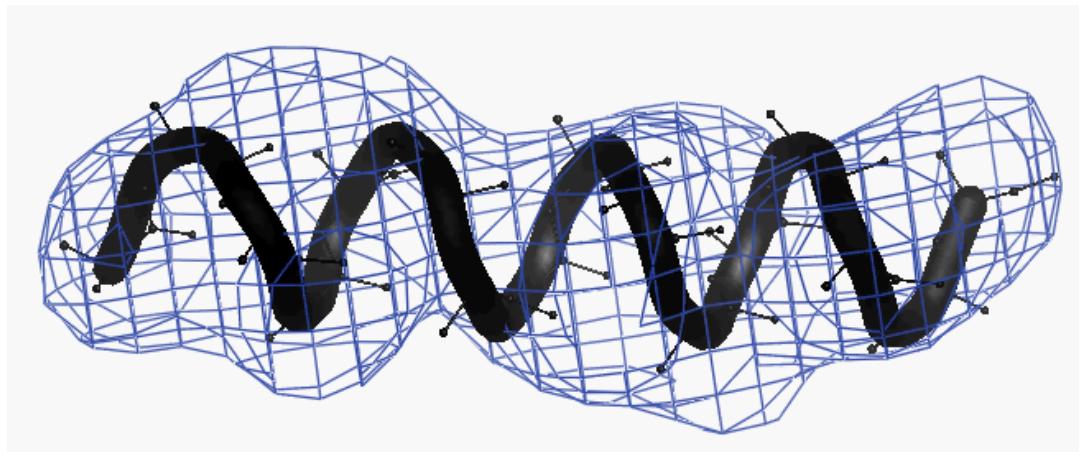
$$T_{\text{RESTRAINTS}} = T_{\text{BOND}} + T_{\text{ANGLE}} + T_{\text{DIHEDRAL}} + T_{\text{PLANARITY}} + T_{\text{NONBONDED}} + T_{\text{CHIRALITY}}$$

$$T_{\text{BOND}} = \sum_{\text{all bonded pairs}} w(d_{\text{ideal}} - d_{\text{model}})^2, T_{\text{ANGLE}} = \dots, \dots$$

Geometry restraints are too simple to provide comprehensive description of chemistry should it is needed to compensate for low resolution of experimental data.

Refinement in a nutshell

- Example: Refinement of perfect α -helix against low-resolution data
 - Standard restraints are insufficient
 - Helix geometry deteriorates during refinement

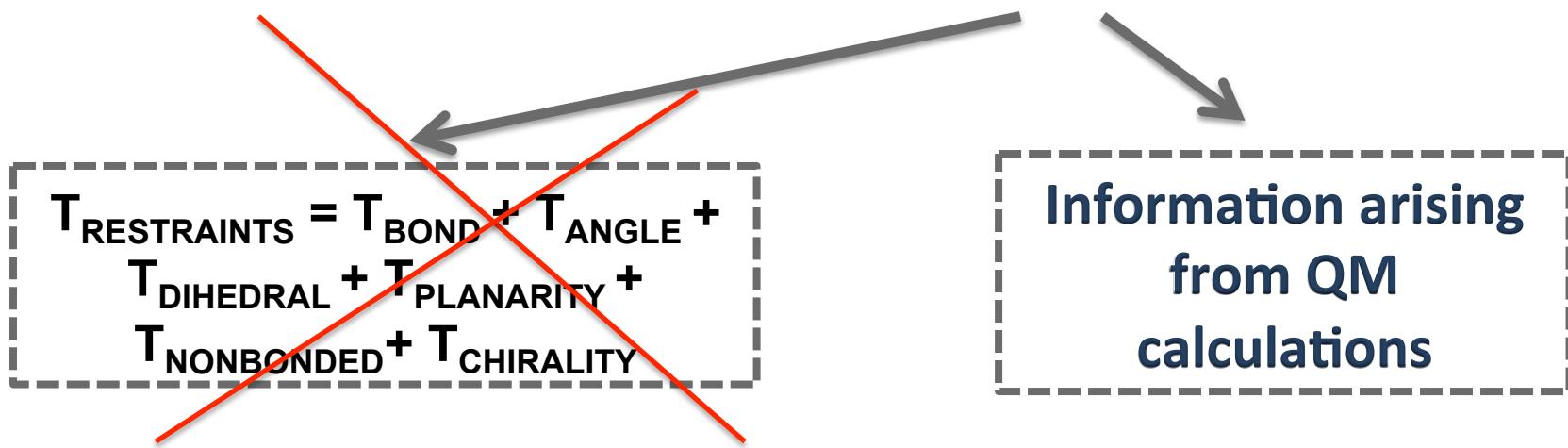


- More information (than standard restraints) is needed to make refinement meaningful

QR (Quantum Refinement)

- Refinement - optimization of T w.r.t. model parameters:

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$



Replace standard geometry restraints with those arising
from QM calculations

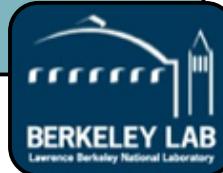
Approach: what's available

- *Phenix* software suite
- Jeff's suite of programs
- 3rd party QM packages
- Vast expertise in:
 - Theoretical chemistry (ICQMS)
 - Professional software development (*Phenix*)
 - Crystallography for protein structure solution (*Phenix*)

Phenix – large scale collaboration

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Youval Dar, Nat Echols, Nigel Moriarty, Nader Morshed, Ian Rees, Oleg Sobolev



Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Cambridge University



An NIH/NIGMS funded
Program Project

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung



Duke University

Jane & David Richardson, Chris Williams, Bryan Arendall, Bradley Hintze

Phenix – quick facts

Phenix publications:

157

Registered Users:

20008

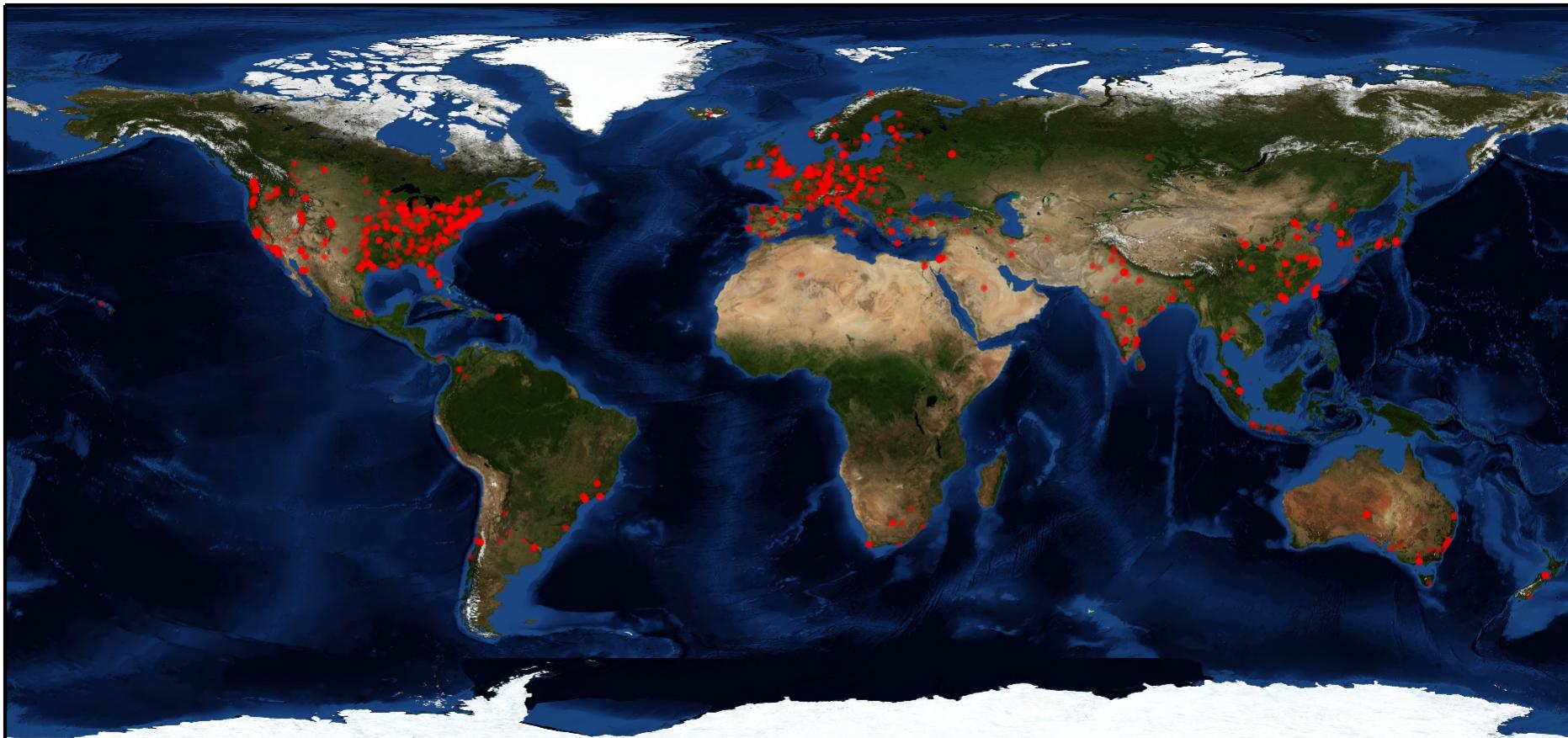
Publications citing *Phenix*:

7263

Registered Locations:

2142

Registered Locations: 2142



Mixing all together

$$T = T_{\text{DATA}}(F_{\text{OBS}}, F_{\text{MODEL}}) + wT_{\text{RESTRAINTS}}$$

Phenix software suite



QM resources

Theoretical chemistry (ICQMS)

Professional software development (*Phenix*)

Crystallography for protein structure solution (*Phenix*)

Mixing all together

Phenix software suite



QM resources

- Modify *Phenix* to incorporate all necessary changes?

Mixing all together

Phenix software suite



QM resources

- Modify *Phenix* to incorporate all necessary changes?
 - Great global goal
 - Requires a lot of team work for a long period of time

Mixing all together

- **Modify *Phenix* to incorporate all necessary changes?** 
- **Impractical as a short-term goal (to get some results now):**
 - *Phenix* is...
 - ...too large (7.5Gb)
 - ...too agile (new version every night)
 - Not all required software is...
 - ...plug-and-play:
 - 3rd party (license restrictions)
 - Language barriers
 - *Phenix*: C++/Python, nothing else allowed
 - ...readily available:
 - New developments may require years of team work
 - Can't push back ad hoc changes in local copy

Mixing all together

- Create new project based on *Phenix* libraries:

- Light-weight initial implementation
- Idea and proof-of-concept testing platform
- Easy to maintain
- No strict policies that may hinder experimenting
- Open to anyone

Phenix software suite



QM resources

Q|R – quantum refinement

Refinement tools
using CCTBX
library

Bridge between
Q|R and QM
software

QM software (3rd
party at present)

State-of-the-art: Q|R project at GitHub

The screenshot shows a GitHub repository page for the project "qrefine / qrefine". The repository is described as the "Quantum Refinement Project". Key statistics displayed include 126 commits, 2 branches, 0 releases, and 4 contributors. The most recent commit was made an hour ago. The repository's README.md file is also visible.

Key statistics:

- 126 commits
- 2 branches
- 0 releases
- 4 contributors

Latest commit: ebf0d72 an hour ago

Commits (partial list):

- pafonine Minor code cleanup, avoid infinite loop
- ase updated mopac keywords
- helix updated helix result
- log cctbx and qm(mopac) refinements of pdb 2o19, refined pdb structures a...
- mtz real PDB data, model with H added and refined with phenix.refine
- na_fragment forgot to add H
- p26 statistics analysis of p26
- pdb misc change
- qrefine Minor code cleanup, avoid infinite loop
- .travis.yml changed travis script
- README.md edited our readme
- setup.py added some folders for new modules

README.md content:

qrefine

Quantum Refinement Project

Status

build passing

Q|R project

- Major code contributors so far:
 - Mark Waller
 - Min Zheng
 - Pavel Afonine
 - ... others are warmly encouraged to participate!

Constructing test system

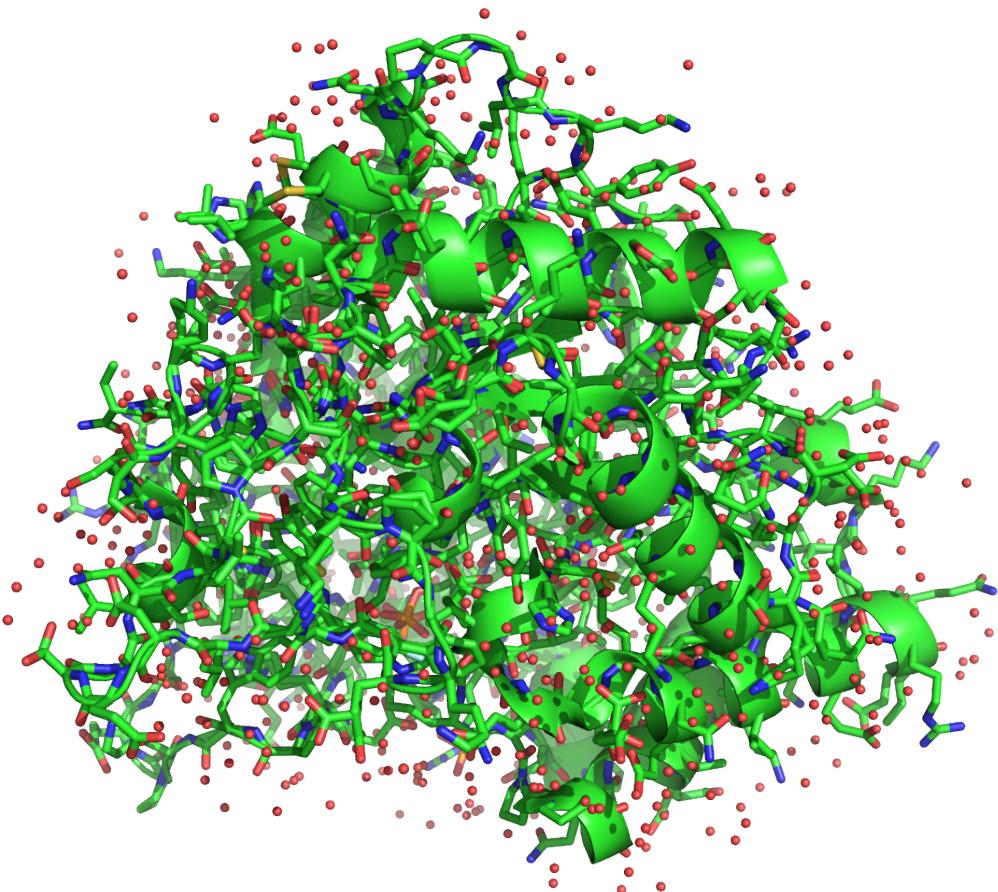
- **Self-consistent system with known answer:**
 - Choose reference model (THE answer)
 - Simulate (compute) “experimental” data from it
 - Perturb reference model in many different ways
 - Refine perturbed models
 - See how refined model match the answer model
 - Use this to rank and optimize refinement tools

Constructing test system - Goals

- Test software and tools:
 - Find bugs, corner cases, inefficiencies
 - Get ideas about:
 - Convergence radius
 - Runtimes
 - Overall feasibility

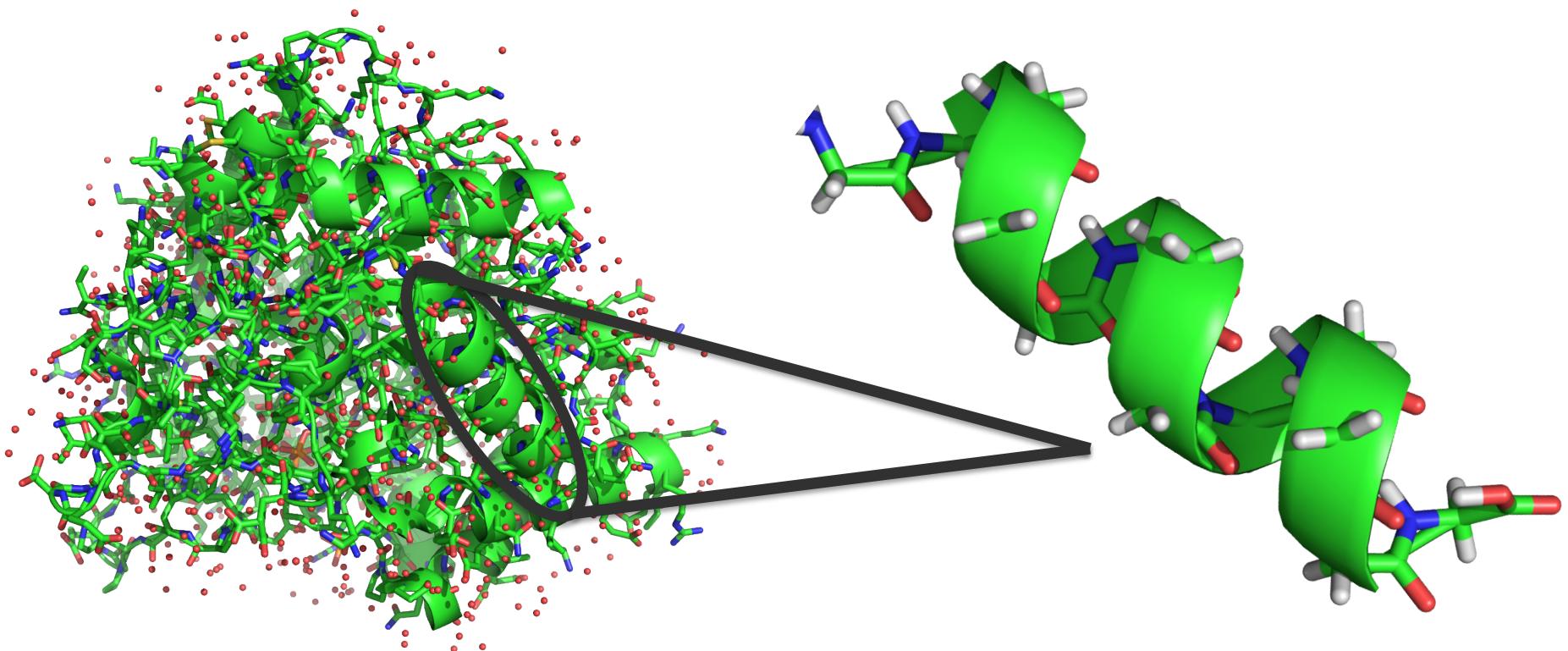
Constructing test system

- Aldose Reductase model:
 - PDB code: 1US0
 - Resolution: 0.66 Å



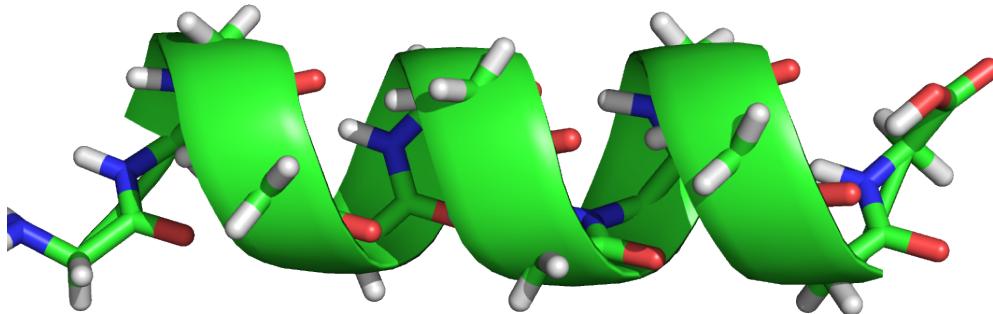
Constructing test system

- Carve out well resolved in electron density map helix:
 - A87-99
 - Convert to poly-GLY to speedup calculations
 - Add H (“neutron” longer X-H distances)



Constructing test system

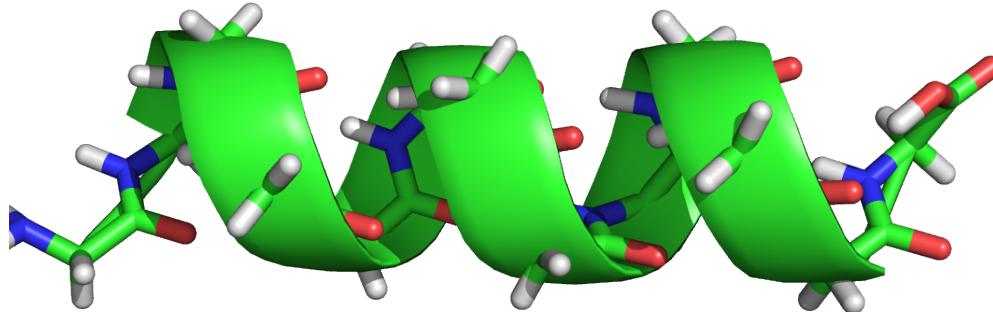
- Consider...



... as a reference and simulate (calculate) structure factors from it (F_{obs})

Constructing test system

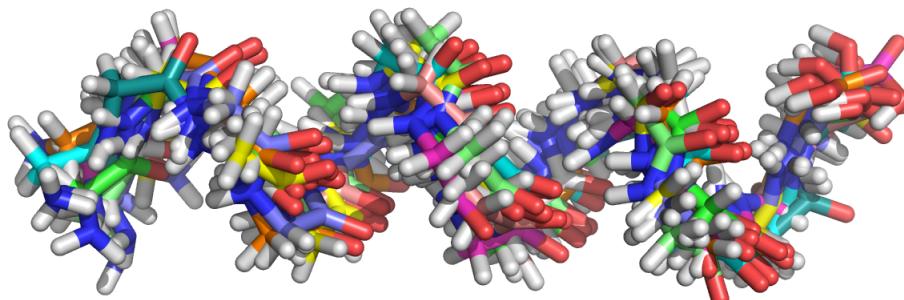
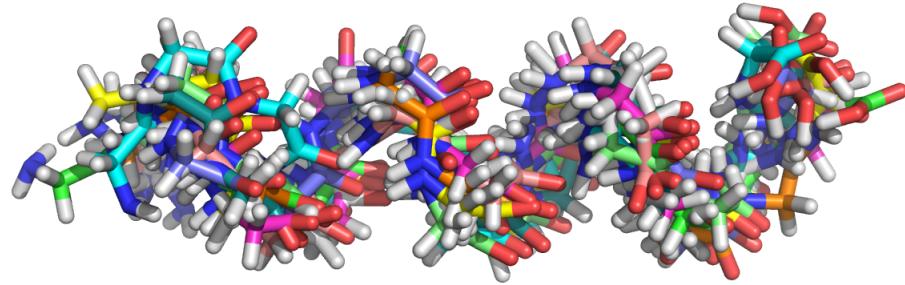
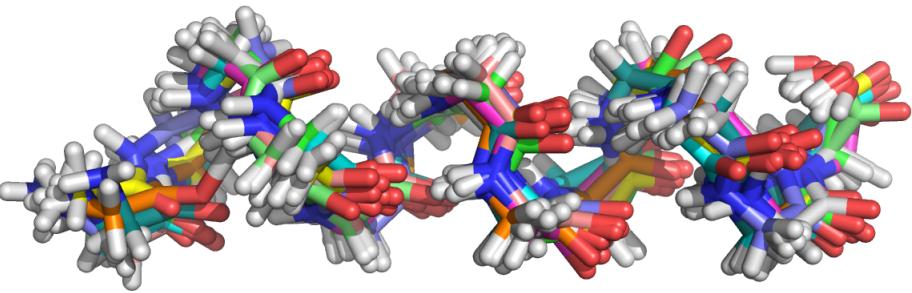
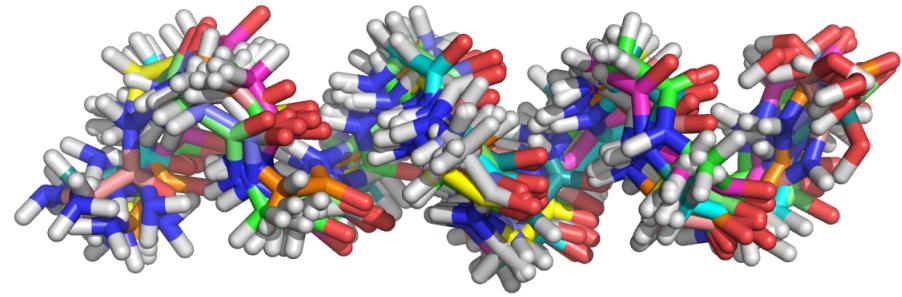
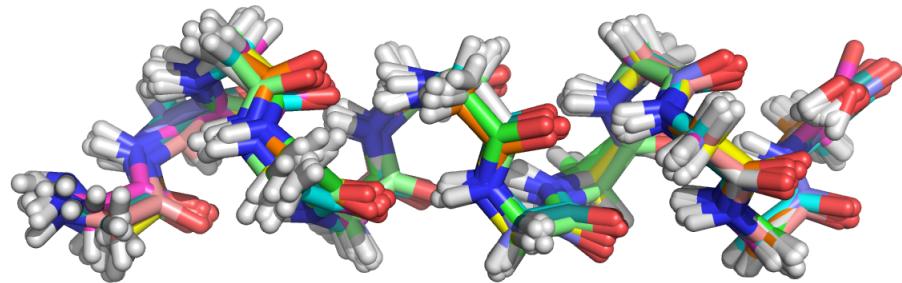
- Consider...



... as a reference and calculate a set of perturbed models with different perturbation dose introduced into coordinates

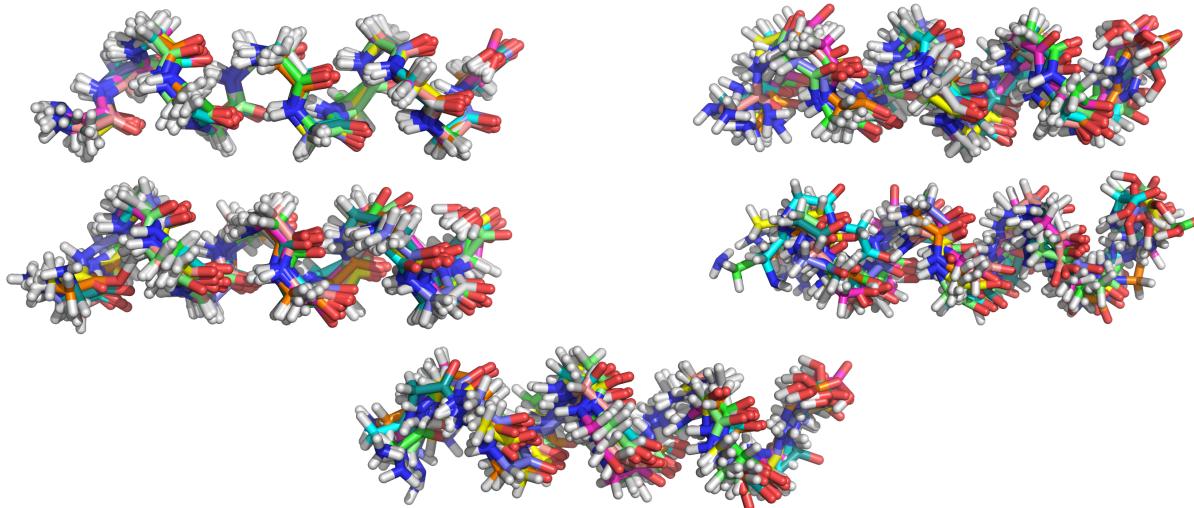
Constructing test system

- 5 sets of perturbed models: 0.3, 0.6, 0.9, 1.2, 1.5 Å
 - 10 models in each set, 50 models total



Refinement and pure optimization

- For each model...

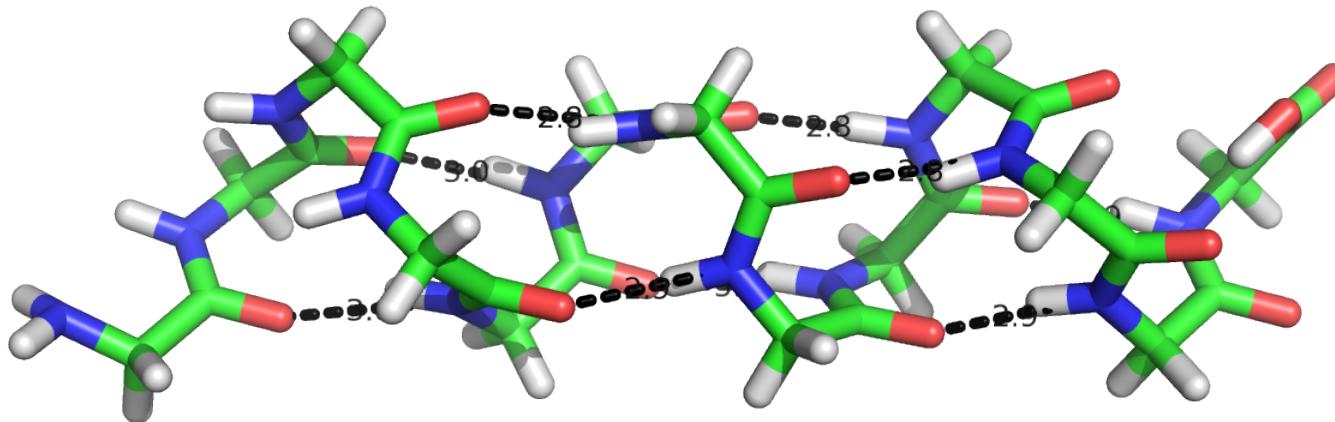


...Do:

- Optimization
 - QM (Mopac)
 - *Phenix*
- Refinement
 - QM (Mopac)
 - *Phenix*

Assessing results

- We know how many H-bonds are in reference model ...



... we can calculate how many H-bonds survived perturbation:

Perturbation amount (\AA)	Percentage of H conserved
0.3	60
0.6	34
0.9	14
1.2	10
1.5	3

Results

- ***Phenix optimization:***

Perturbation amount (Å)	Percentage of H conserved
0.3	54
0.6	32
0.9	20
1.2	9
1.5	3

- ***QM optimization:***

Perturbation amount (Å)	Percentage of H conserved
0.3	90
0.6	73
0.9	54
1.2	38
1.5	26

Results

- ***Phenix refinement:***

Perturbation amount (Å)	Percentage of H conserved
0.3	8
0.6	7
0.9	10
1.2	5
1.5	N/A

- ***QM refinement:***

Perturbation amount (Å)	Percentage of H conserved
0.3	19
0.6	13
0.9	6
1.2	9
1.5	N/A

Results – Question

- QM optimization:

Perturbation amount (Å)	Percentage of H conserved
0.3	90
0.6	73
0.9	54
1.2	38
1.5	26

- QM refinement:

Perturbation amount (Å)	Percentage of H conserved
0.3	19
0.6	13
0.9	6
1.2	9
1.5	N/A

- Why refinement loses so many H-bonds?