Automated ordered solvent (water) building into high-resolution cryo-EM maps

Pavel Afonine

LBNL, Berkeley, California, USA

March 31, 2021

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

- It isn't clear yet up to what resolution it is possible to build water reliably into cryo-EM maps

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

- Crystallography: fully occupied water appears at 1σ in 2mFo-DFc map. Partially occupied water appear at lower thresholds
- Cryo-EM: maps can be subject to many various manipulations (sharpening, flattening solvent region, …). Map scale isn't well defined.

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

- Difference map calculation isn't established yet (though there is *Phenix* tool to do it: *phenix.real_space_diff_map*)
  - Obtaining meaningful difference map requires accurately refined B-factors and occupancies (where applicable)

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.0 / 3.2	2.0 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

- Geometric (min/max H-bond distance) criteria are the same

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

- Crystallography:
  - Map is *a moving target*
    - Model phases are used in map calculation
    - Maps changes every time the model changes
  - Solvent is constantly updated during later stages of refinement

- Cryo-EM: Map does not change
  - Model does not affect the map
  - Solvent (water) can be built once into the final model

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

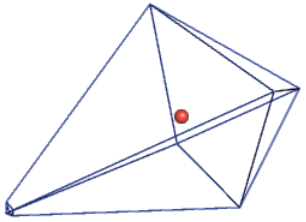
- Crystallography:
  - Maps calculated internally
    - Map grid steps are consistent, predictable and can be controlled internally
- Cryo-EM:
  - Map is supplied by the user
    - Grid step can be anything and cannot be enforced to be consistent

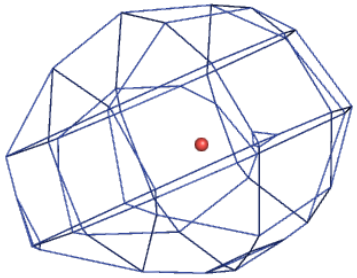
Why grid step is important?

Original map (0.71Å grid step)

Re-sampled map (0.25Å grid step)

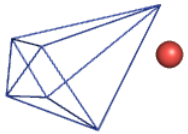
Example 1

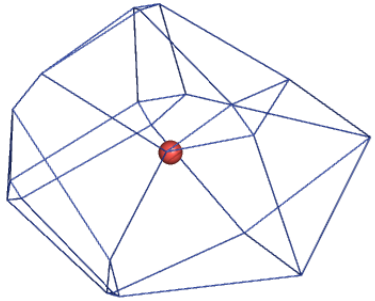




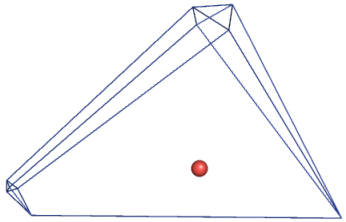
All maps contoured at identical thresholds

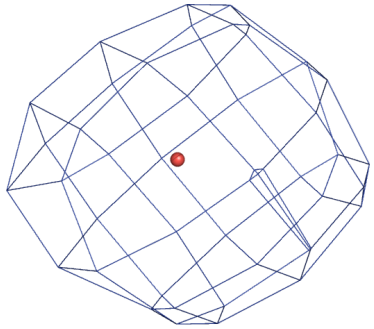
Example 2





Example 3





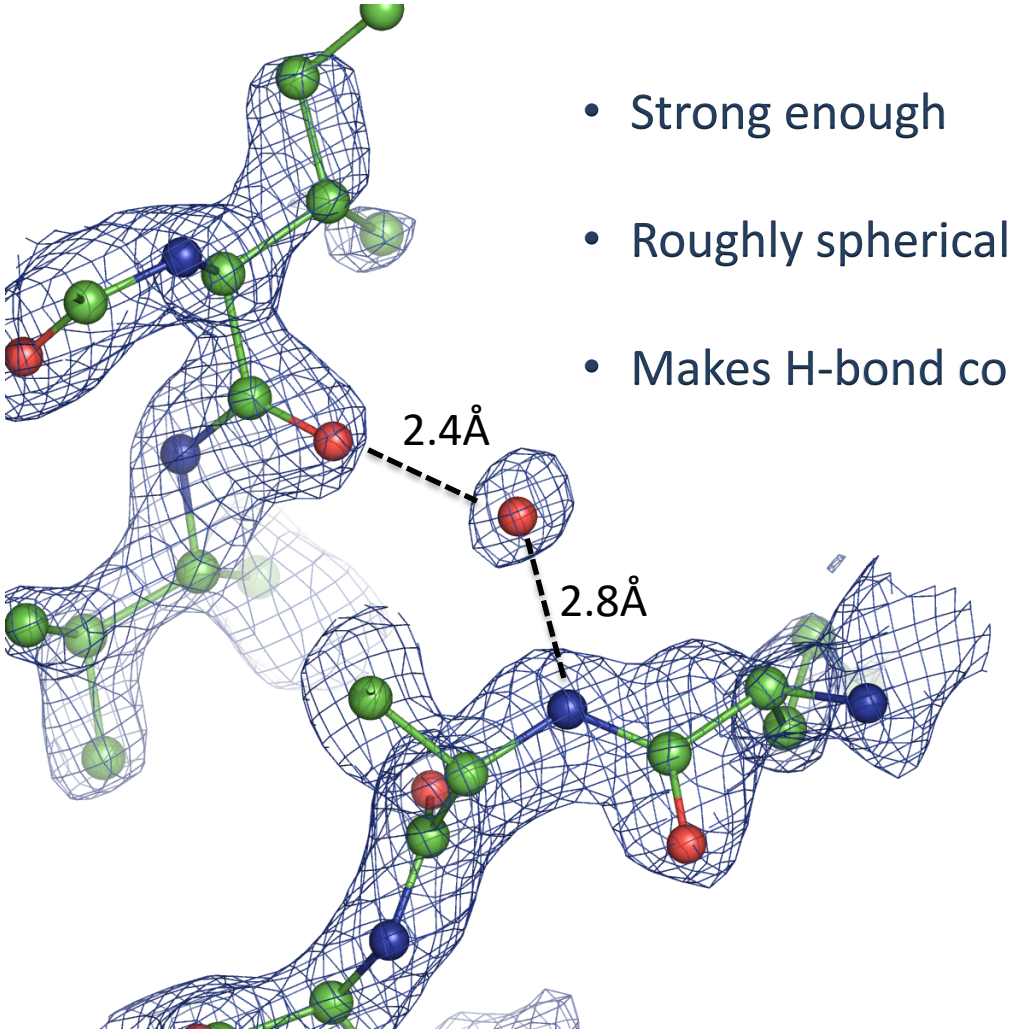
- Map re-sampling:
  - Easier to calculate and analyze peak properties
  - Maps become larger and computations become slower

Water picking criteria: crystallography *vs* cryo-EM

Criteria	Crystallography	Cryo-EM
Resolution	3Å or better	?
Map peak	1σ (2mFo-DFc)	?
Difference map peak	3σ (mFo-DFc)	?
Min/max peak-peak, peak-atom distance (Å)	2.2 / 3.2	2.2 / 3.2
Refinement	Yes	N/A
Map grid step (Å)	Resolution/3 or /4	?
Shape	Approx. spherical	Approx. spherical

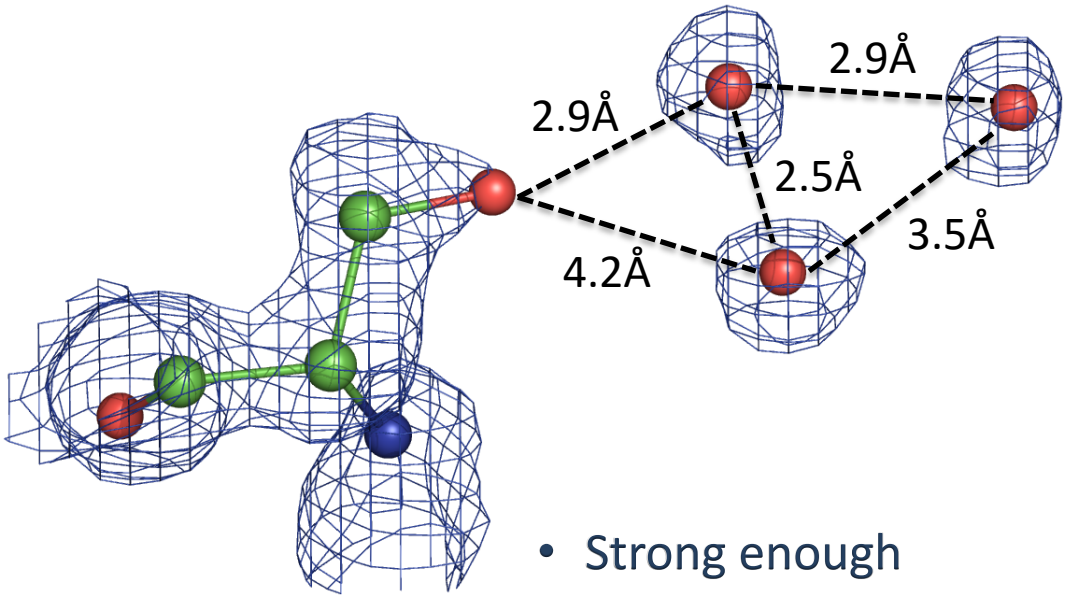
- Not used in major software packages (few exceptions, such as ARP/wARP)

Good water densities



- Strong enough
- Roughly spherical
- Makes H-bond contacts with macromolecule

Good water densities

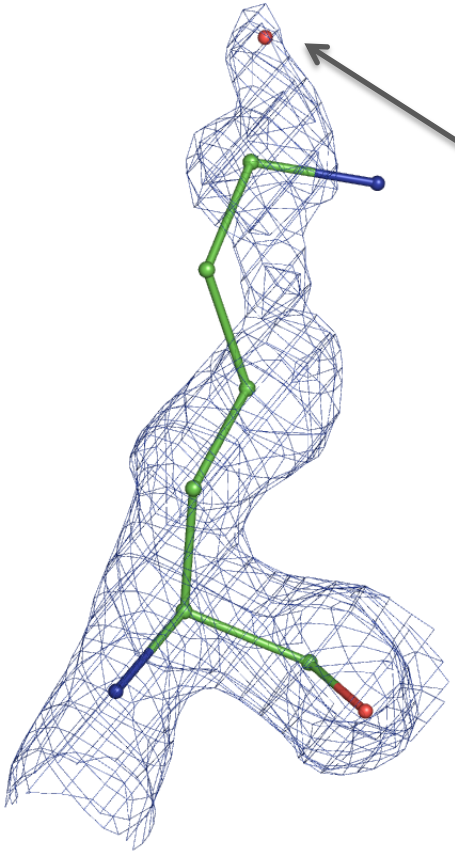


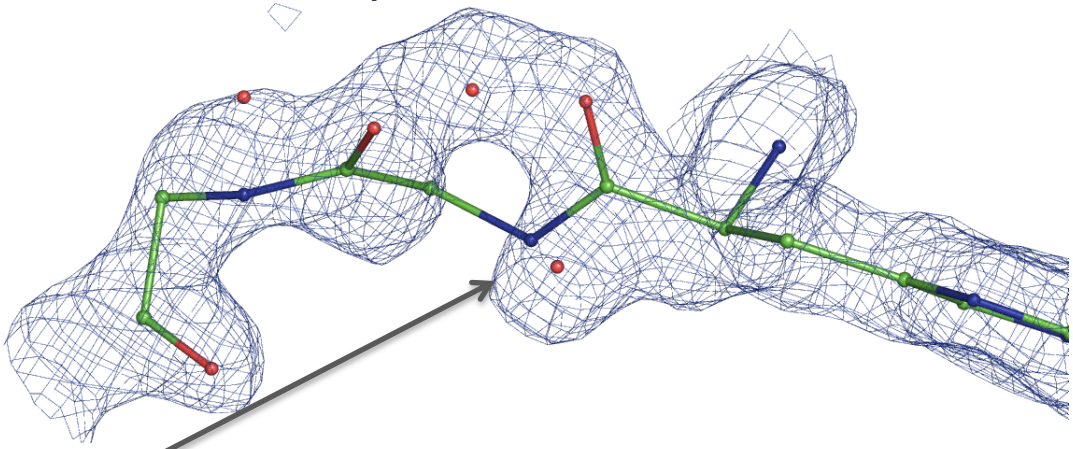
- Strong enough
- Roughly spherical
- Makes H-bond contacts with macromolecule
- Water not making contacts with macromolecule does make contacts with other water (that interfaces the macromolecule)

Tough cases

Poorly fit backbone

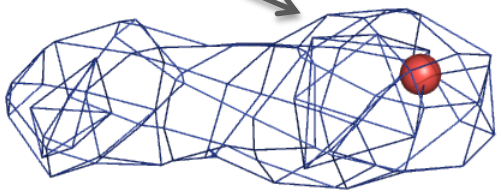
Poorly fit side chain





Bad water

Peak isn’t spherical and is too large

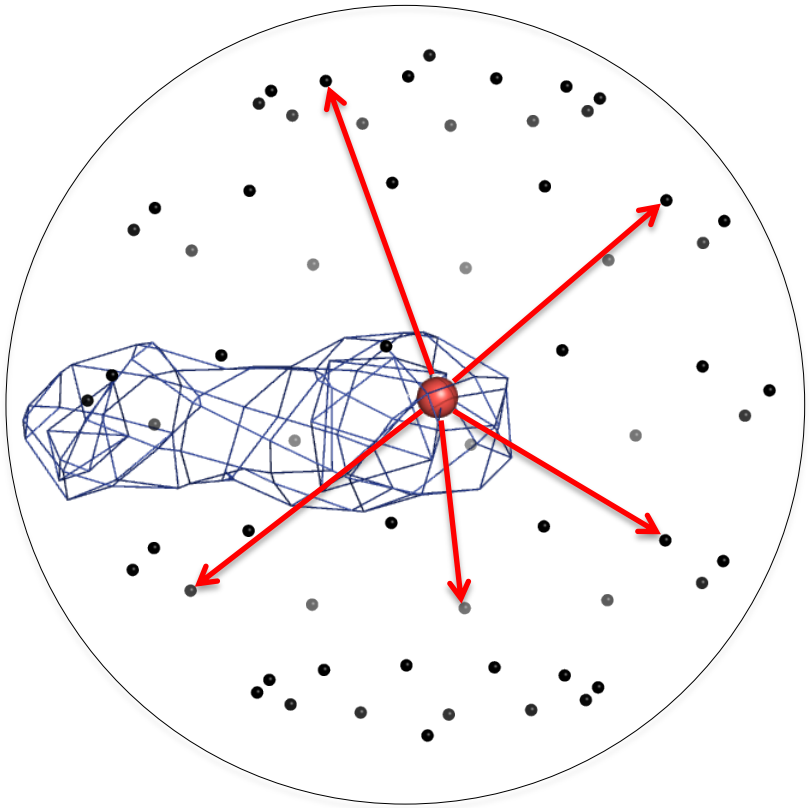


The protocol

1. **Map preparation and analysis**
   - Re-sample the map (0.3Å grid step)
   - Zero map inside molecular region (1Å masking radius)
   - Zero map outside molecular region + 6Å buffer (this is where shells of ordered solvent are expected)
   - Normalize map in the non-zero region (where solvent will be searched) to have mean=0 and standard deviation=1
2. **Find peaks in the non-zero region of the map that are higher than specified threshold**
3. **Filter peaks by peak-peak, atom-peak and distance, potential to form H-bonds and sphericity**

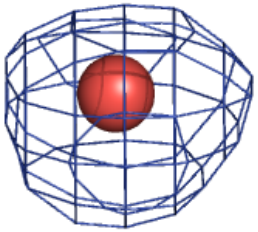
The procedure works on the whole model or per chain (default)

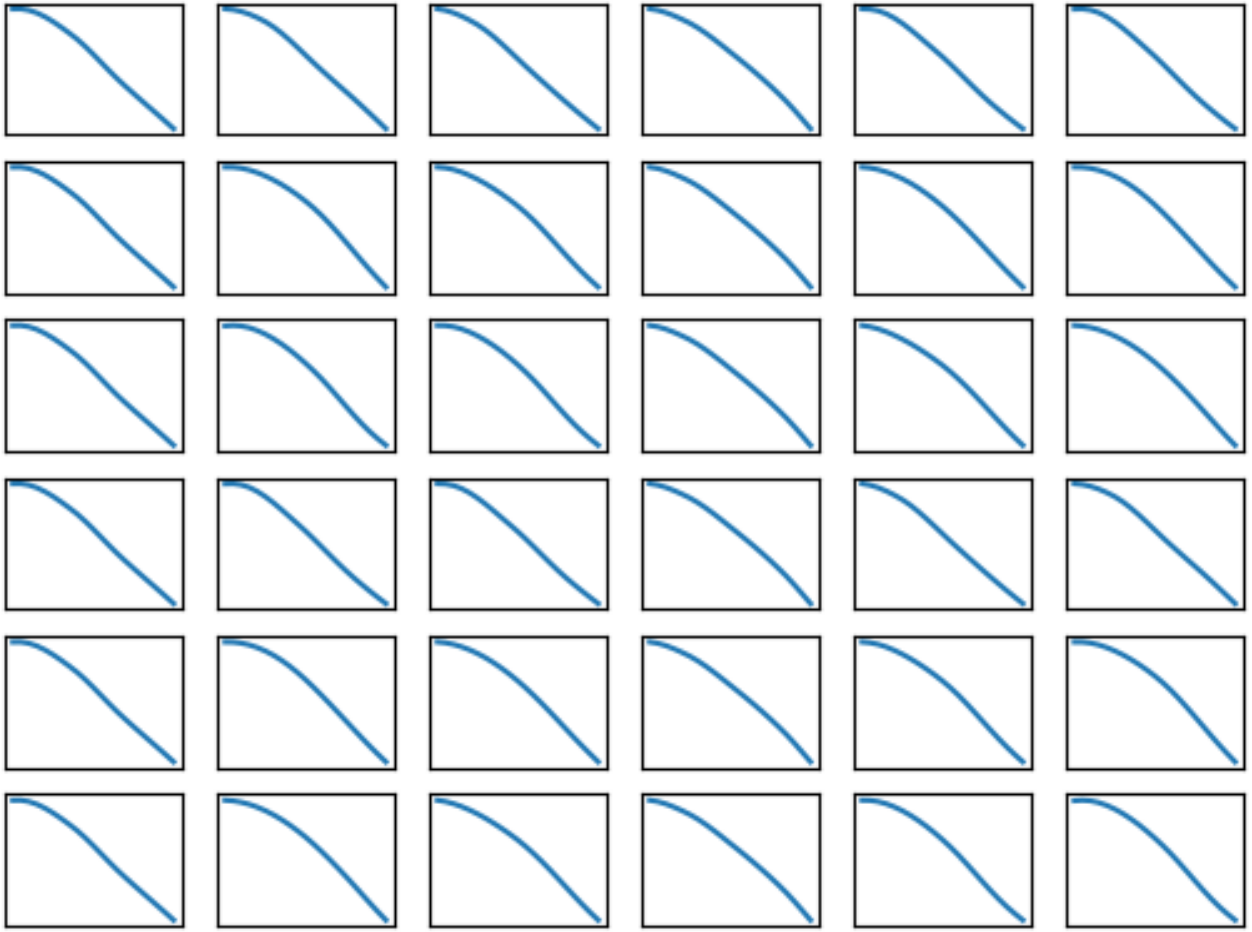
Sphericity analysis



- Surround the peak with a sphere (1Å radius)
- Spread points on the sphere surface
- Calculate map distribution along (“peak – dot on sphere” vector)
- Repeat for all “peak - dot on sphere” pairs
- Calculate correlation between all pairs of vectors

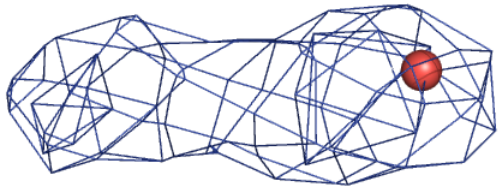
Density distributions for a good peak

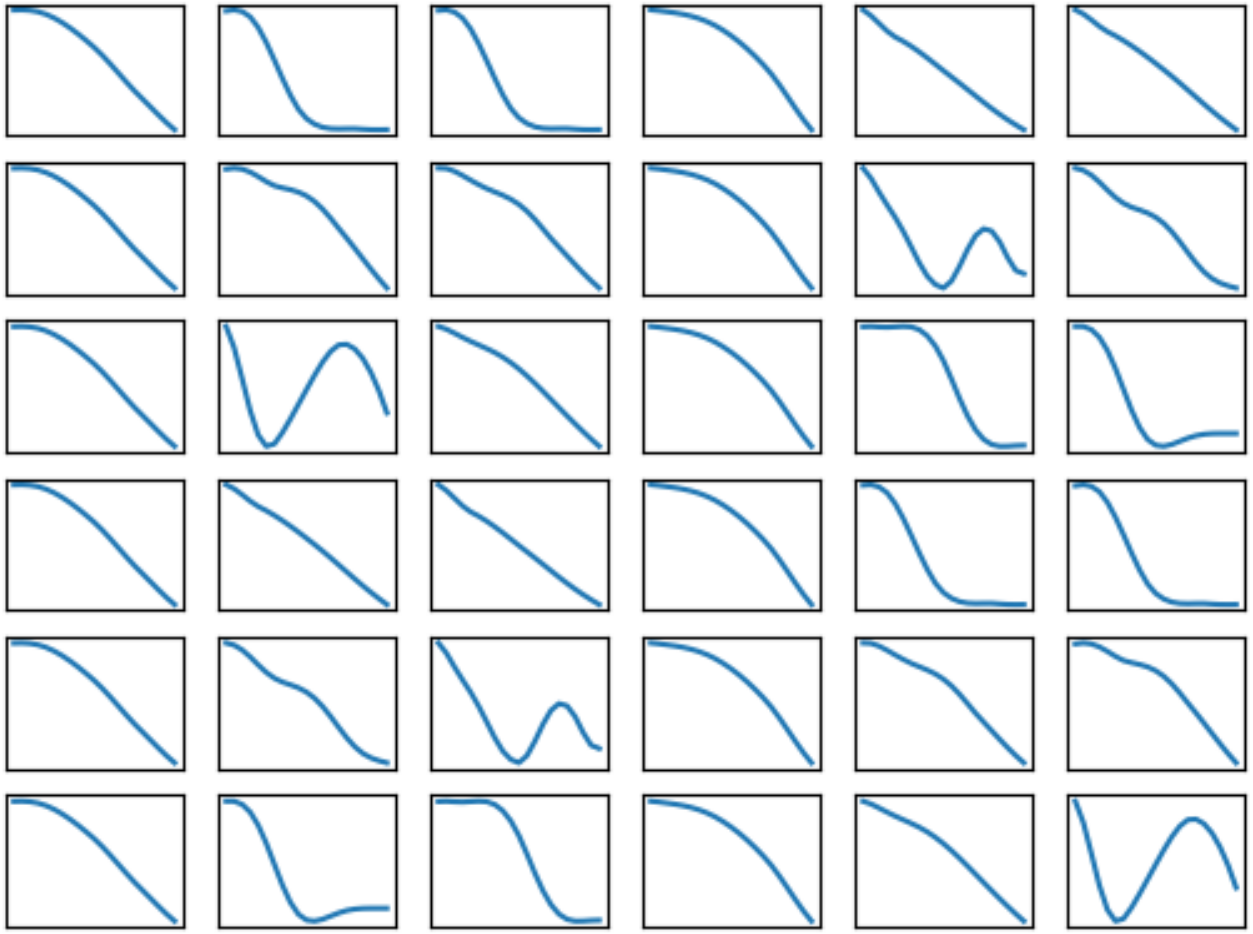




- All distributions are very similar (but not necessarily identical)

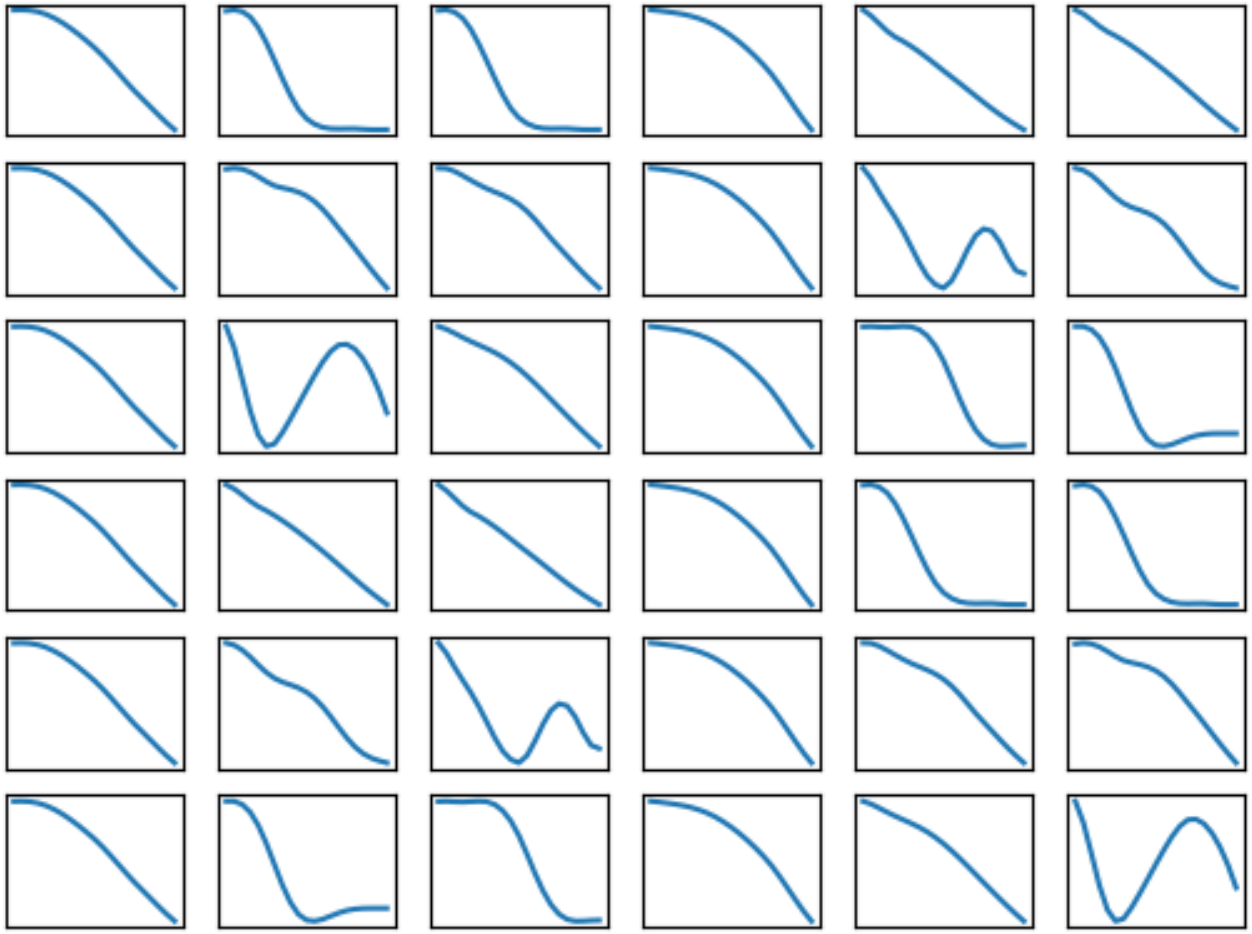
Density distributions for a poor peak





- Some distributions look different from the others

Selection by sphericity



- Calculate correlation between all pairs of plots for the given peak
  - Correlation less than 0.9-0.95 indicates bad peaks in most cases

Summary (I)

- Core code is implemented in *CCTBX*
- Available in *Phenix* now (command line and GUI)
- Performs best with final atom-complete models
- Operates on the whole model
- Use many empirical thresholds for decision-making
- Command line example:

```
phenix.douse model.pdb map.mrc
```

Summary (II)

- Key parameters and their default settings:
  - dist_min=2.0
    - Min peak-peak or peak-atom distance
  - dist_max=3.2
    - Max peak-peak or peak-atom distance
  - step=0.3
    - Grid step of re-sampled map (in Å)
  - map_threshold=1.0
    - Map threshold (in r.m.s.)
  - keep_input_water=false
    - Flag to keep or remove water in input file
  - sphericity_filter=true
    - Map threshold (in r.m.s.)
  - scc05=0.97 *and* scc1=0.9
    - Shpericity correlation thresholds for 0.5 and 1.0 Å spheres

- Example:

```
phenix.douse model.pdb map.mrc map_threshold=1.5
```

Summary (III)

- Add more water (loose selection criteria):
  - Decrease any or all: dist_min, map_threshold, scc05, scc1
- Add less water (strict selection criteria):
  - Increase any or all: dist_min, map_threshold, scc05, scc1
- Manual inspection is required. The procedure does not guarantee to interpret all peaks correctly
- Typical runtimes vary between a few seconds and a few minutes
- Map symmetry isn’t considered at the moment but will be used in future releases