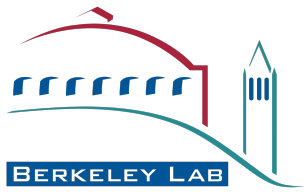


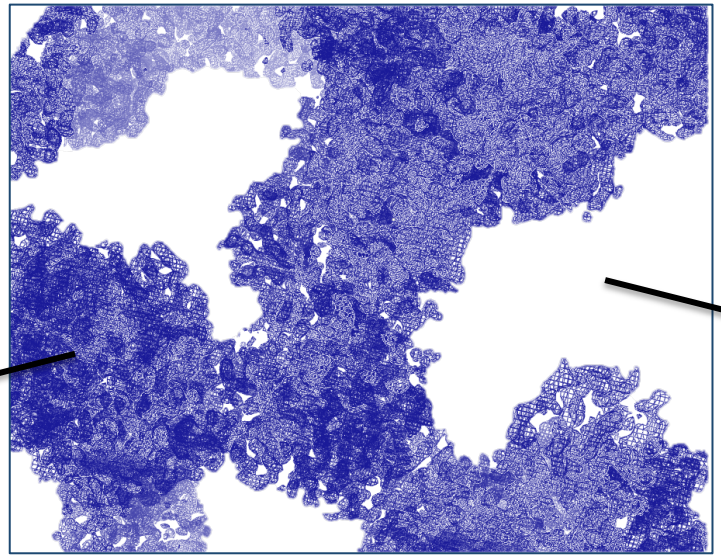

# phenix.polder

A tool for calculating difference maps around atom selections by excluding the bulk solvent mask

Dorothee Liebschner

# How to model disordered solvent in protein crystals?

Protein crystals have a large solvent content:
→ Substantial amount of scattering matter



Macromolecule

Bulk solvent:
constant density,
~0.2-0.6 e/Å$^3$

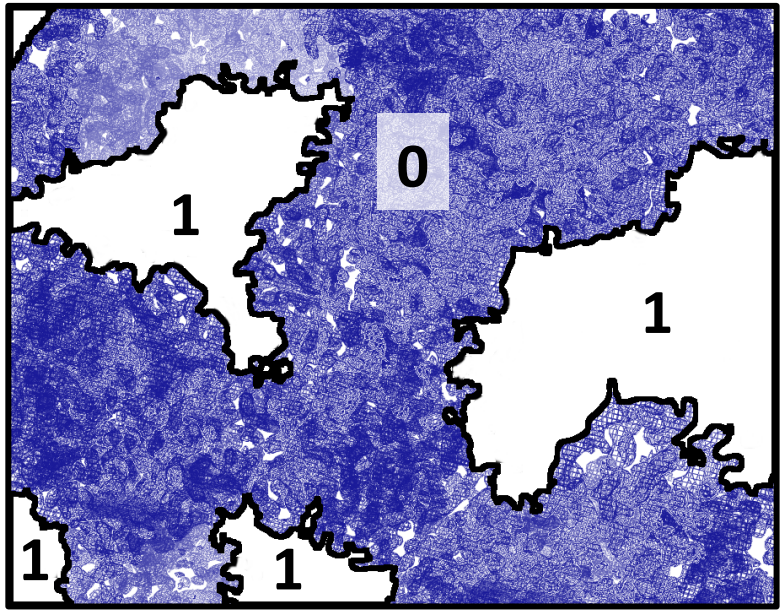
→ The bulk solvent needs to be accounted for in calculated structure factors

Largely applied method (Phenix, CNS, REFMAC): **Flat bulk solvent model**[1,2]

$$\mathbf{F}_{\text{model}} = k_{\text{overall}} \left(\mathbf{F}_{\text{calc(atoms)}} + \mathbf{F}_{\text{bulk solvent}}\right)$$

[1]Jiang, J. & Brünger, A. (1994). *J. Mol. Biol*. **243**, 100–115.
[2]Phillips, S. E. (1980). *J. Mol. Biol.* **142**, 531–554.

# Flat bulk solvent model: defining the solvent contribution[1]




1. Compute solvent mask M:

   0 – inside protein,
   1 – outside (and in pockets)

2. Calculate structure factors from mask:

$$M \xrightarrow{FT} F_{Mask}$$

3. Define contribution to total structure factor:

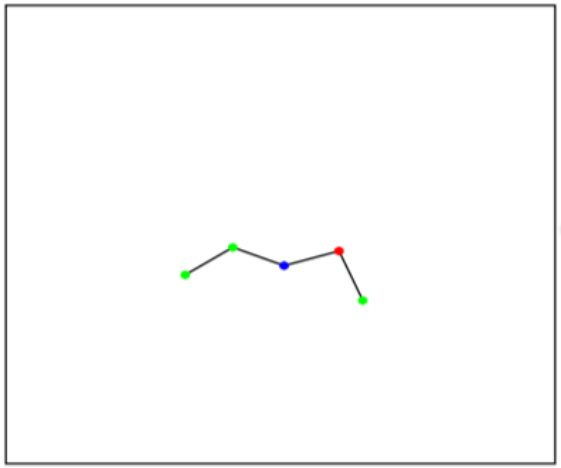
$$F_{bulk\_solvent} = k \cdot F_{Mask}$$

*k* represents mean solvent density in eÅ$^{-3}$, modulated by a smearing factor

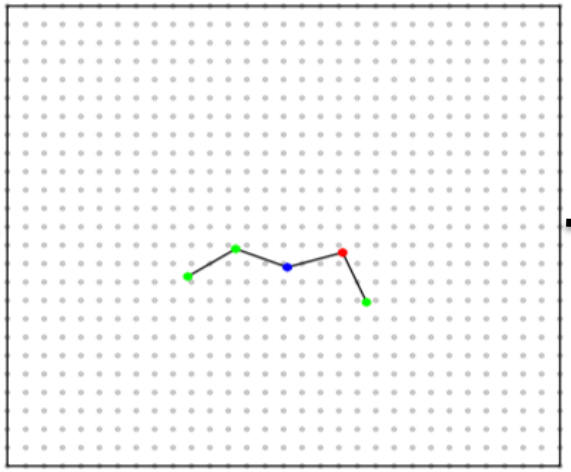
[1]Jiang, J. & Brünger, A. (1994). *J. Mol. Biol*. **243**, 100–115.

# Flat bulk solvent model: how to define the solvent mask[1]

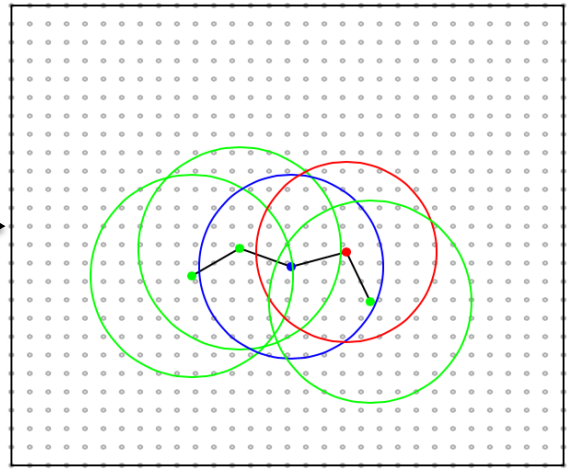
Starting point: atomic model



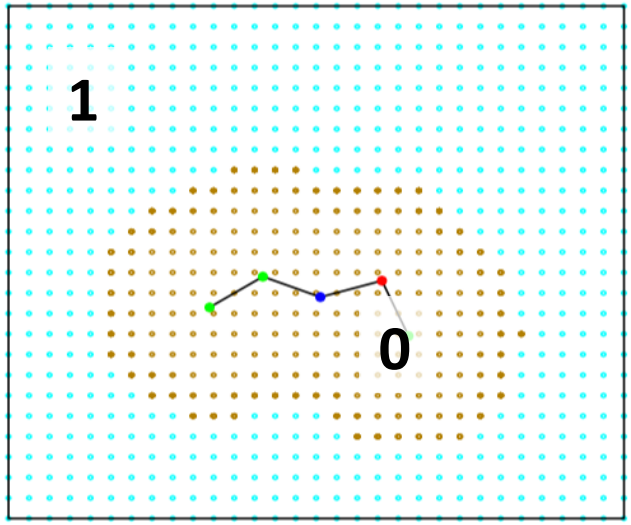
Define grid



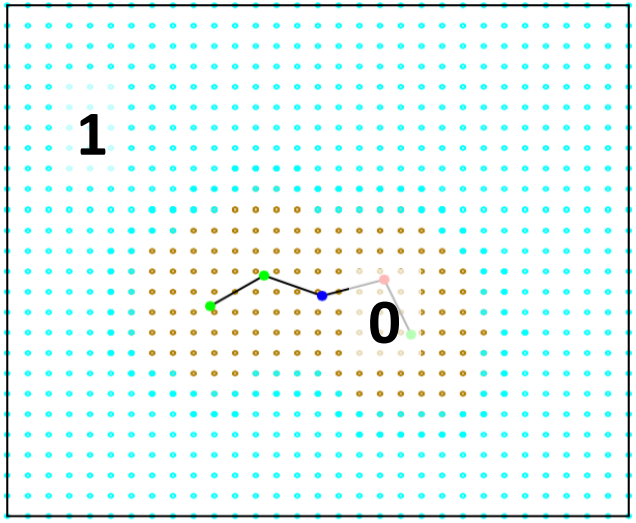
Define atom radii
($r_{vdW} + r_{solv}$)



Set mask to 1 outside the circles, 0 inside




Shrinking: Reset some grid points inside the circles to 1




$F_{Mask}$

[1]Jiang, J. & Brünger, A. (1994). *J. Mol. Biol*. **243**, 100–115.

# The solvent mask can influence omit maps

**Omit-maps can be computed in two ways:**

1. Delete atoms* from the pdb file (refine model) and calculate map ("**omit map**")

   → solvent mask will flood into the pocket. If the density is weak, bulk solvent can obscure it

   Atoms are "hidden" by solvent

2. Keep atoms in pdb file and exclude them from bulk solvent mask ("**biased omit map**")

   → This map will be biased because the difference map has the shape of the atom selection, but may be only bulk solvent.

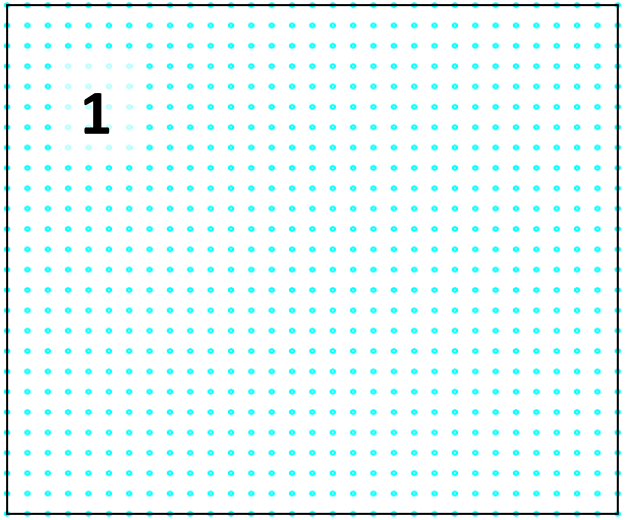
   

   Map is biased

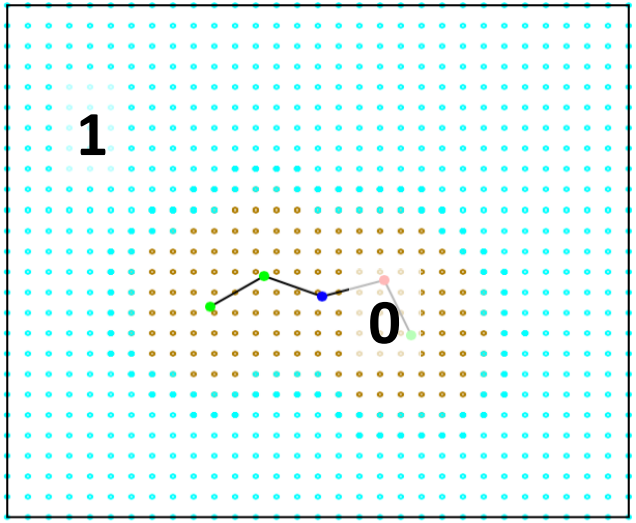
*Ligand, alternate conformations, side-chain orientations, loops, …

# How can omit maps be biased by solvent?

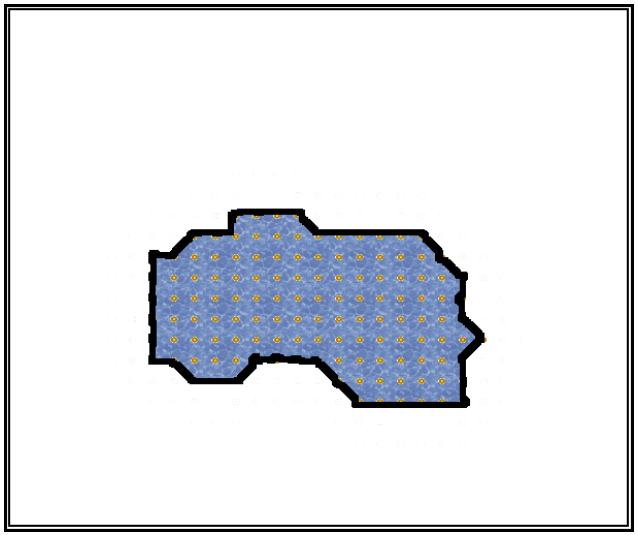
No molecule is present




If a molecule is placed with occupancy 0 and its coordinates are used for mask calculation…




A difference peak with the shape of the molecule will appear in the difference map - and “confirm” its presence

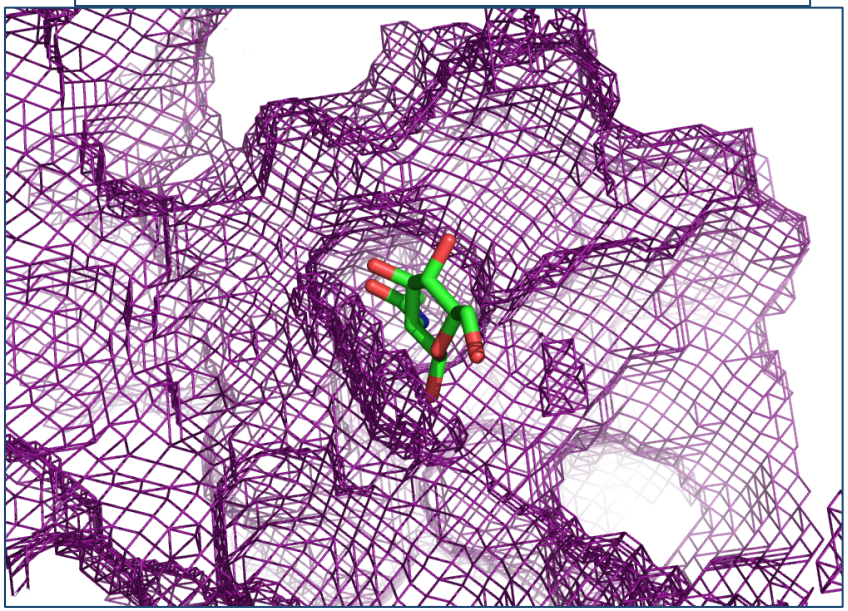


# How to avoid bias and make omit density visible

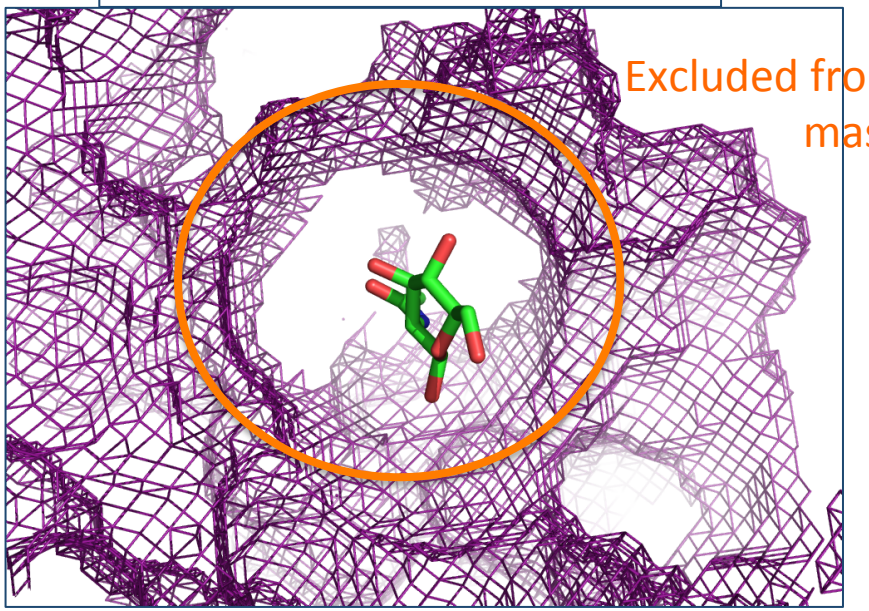
**Solution: Polder map**

The area around the atom selection (e.g. radius of 5 Å) is excluded from the solvent mask.

Solvent mask when ligand is taken into account



Solvent mask used by polder




→ If the atoms are present, their features will appear in the difference density map

→ otherwise, all features will have similar level

# When are polder maps useful?

**Everywhere where density is weak and features can be masked by solvent:**

- Ligands
- solvent molecules
- Alternate conformations
- side-chain orientations
- Loops
- N- or C-terminal

} **Can be used for model completion**

**Implemented in:**

/cctbx_project/mmtbx/command_line/polder.py

# Why call the tool phenix.polder?

**Polder** = “low-lying tract of land enclosed by (…) dikes that forms an artificial hydrological entity, meaning it has no connection with outside water (…). ”*

→ Land is gained by keeping water from penetrating the area

**phenix.polder** = weak features become visible in electron density maps by keeping bulk solvent mask out of the area

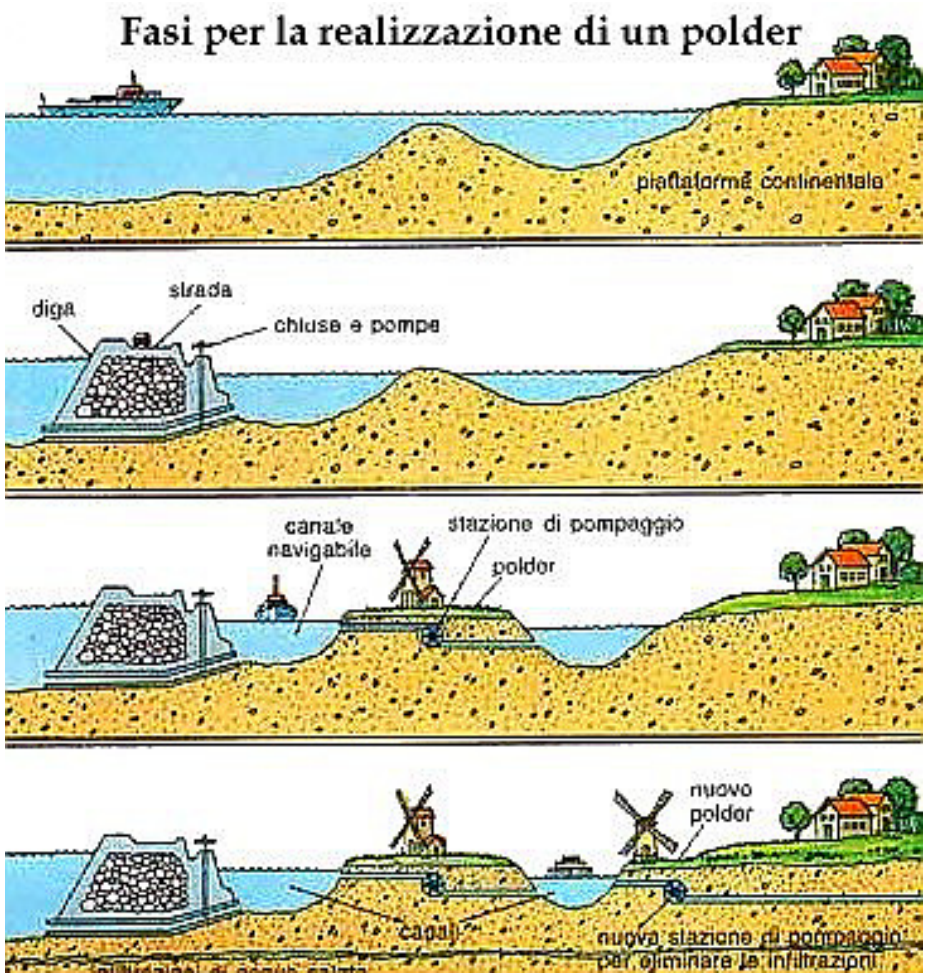



Image: http://cmapspublic.ihmc.us/rid=1M8P9V5BL-1CSMGQC-3TFB/immagini%20olanda
*https://en.wikipedia.org/wiki/Polder

# Workflow of phenix.polder

**Input:**

- Reflection file
- Model file **containing the atoms in the area desired to be masked out**. They do not participate in Fcalc calculation but are used for defining the region excluded from the mask
- Atom selection string
- Optional: radius for bulk solvent mask exclusion (default: 5 Å)

**Program flow:**

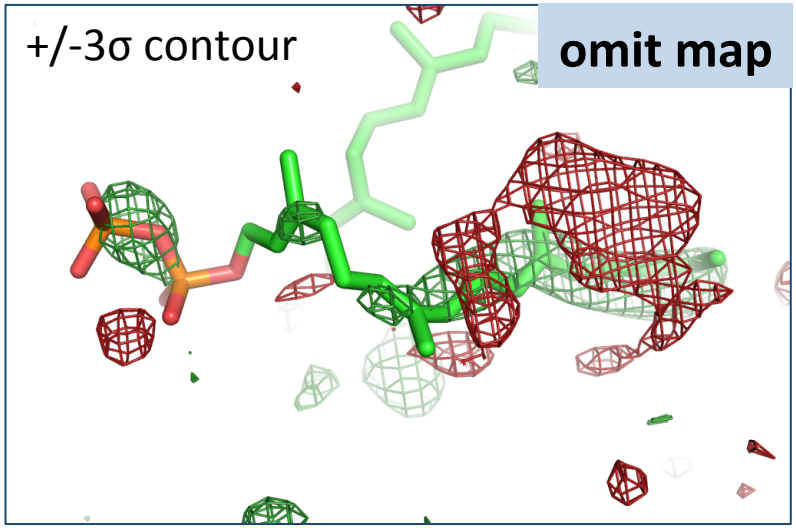
- Determine atom coordinates from selection string
- Calculate solvent mask excluding atom coordinates
- Reset grid points with bulk solvent mask exclusion radius around atom coordinates to zero
- Calculate maps, *R*-factors

**Output:**

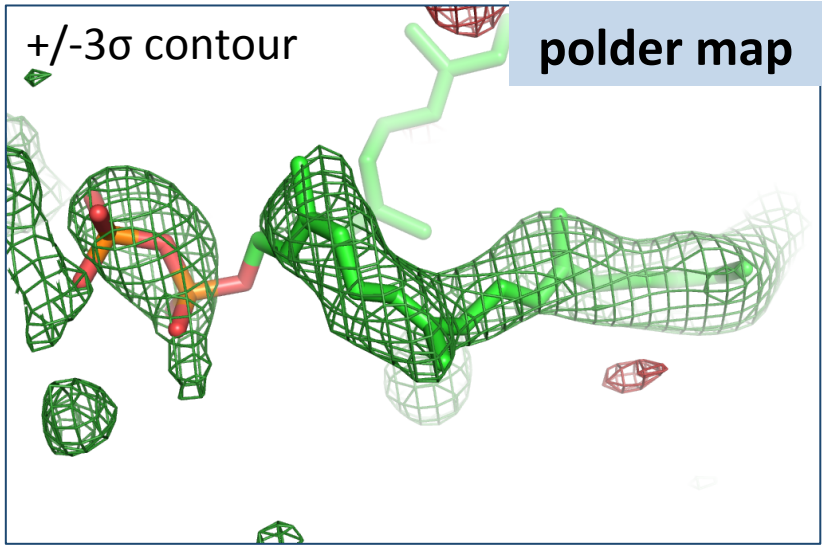
- map coefficients for polder map and omit map
- maps for bulk solvent mask of model and ligand, polder procedure and omit map

# Example: improved ligand density for GRG 503

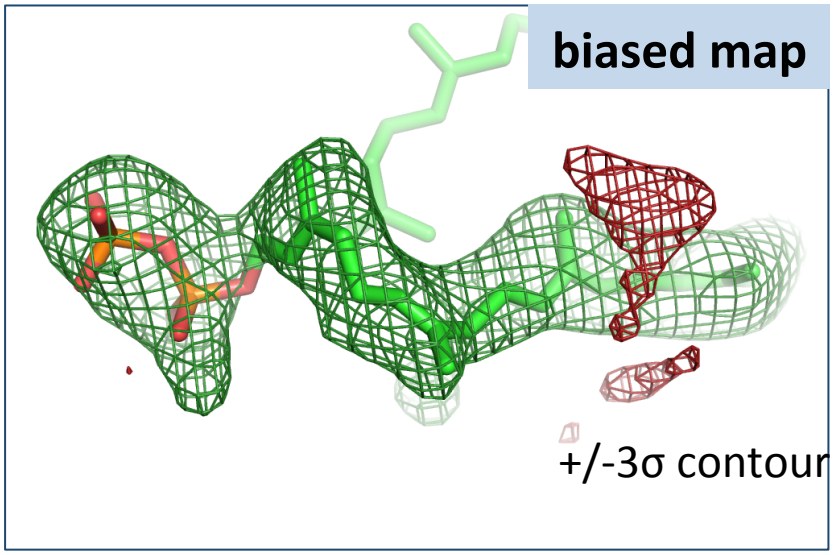
PDB code: 4opi
GRG = geranyl geranyl diphosphate
Resolution: 2.24 Å
$R_{free}$: 0.231
$R_{work}$: 0.190
Residues: 453




No clear density in omit map



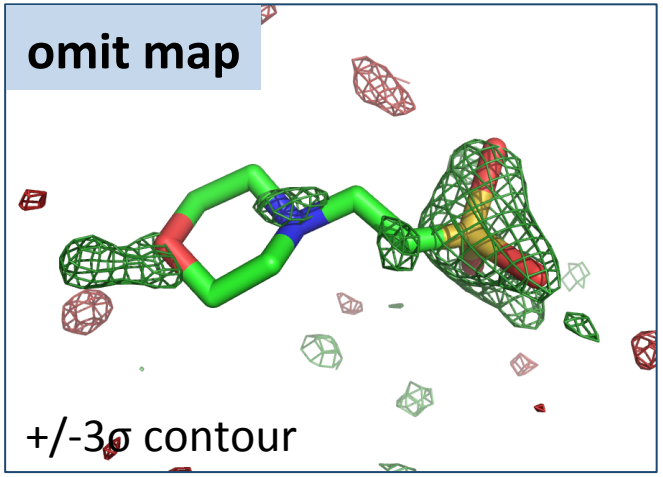

Density for (almost) entire ligand



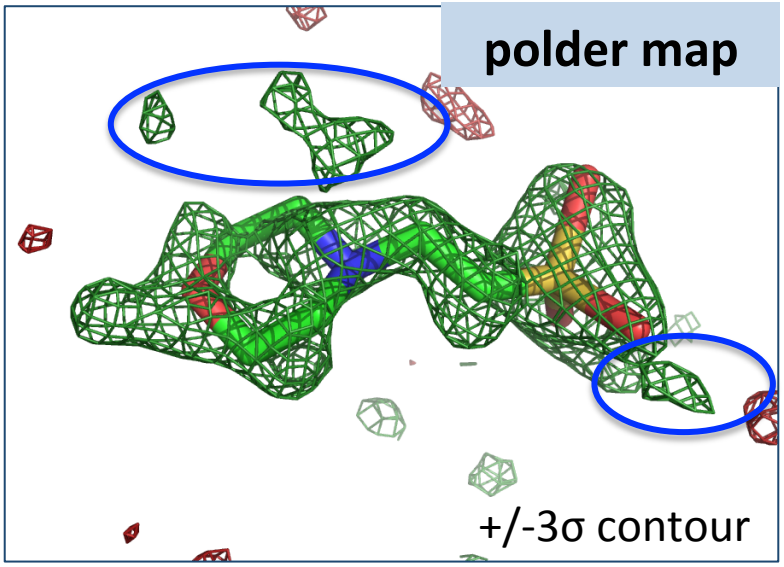

Density has exactly the shape of the ligand → bias

# Example: improved density for solvent molecule MES A 88

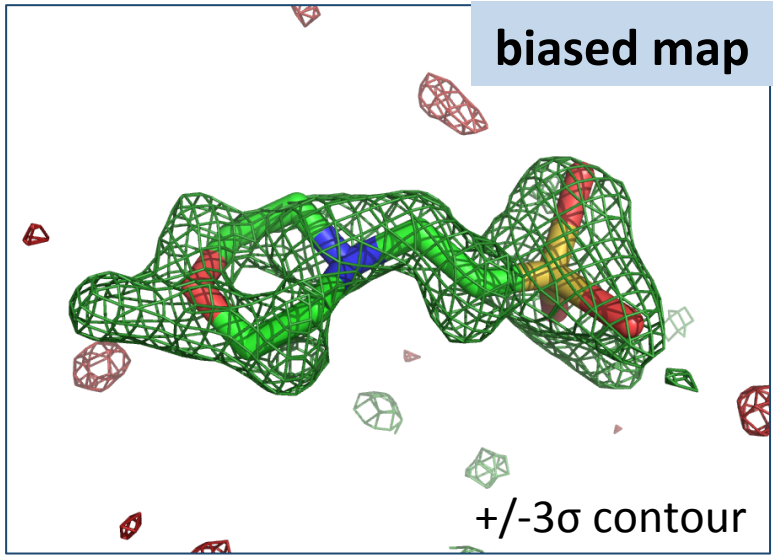
PDB code 1aba
The structure of oxidized bacteriophage T4 Glutaredoxin
Resolution: 1.45 Å
$R_{work}$: 0.175
Residues: 87




Omit map does not show clear density




Density surrounds ligand atoms
Some additional peaks representing bulk solvent



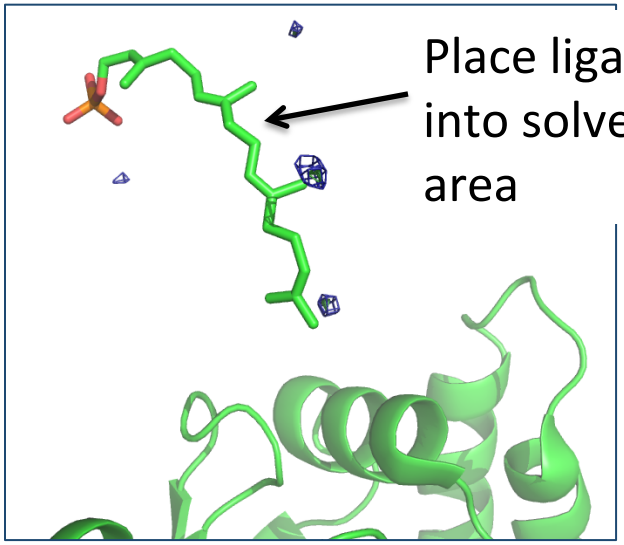

Map is clean and easy to interpret. But density can be MES or bulk solvent.

# Example: Biased omit map has shape of ligand

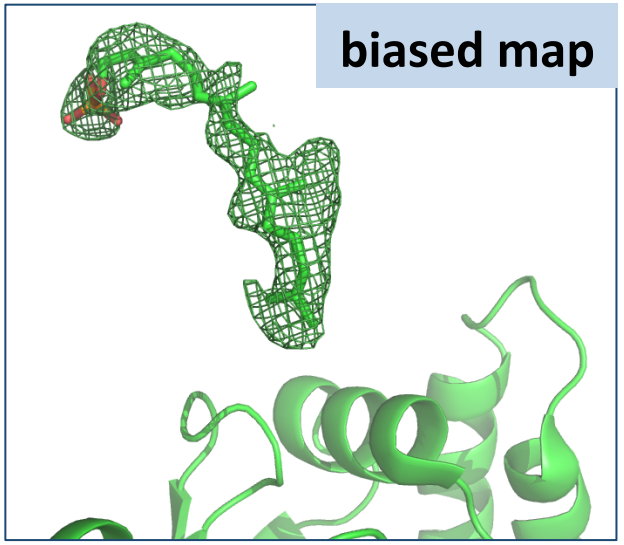
PDB code: 4opi
Resolution: 2.24 Å

**Blue**:
+1σ 2Fobs-Fcalc
**Green**, **red**:
+/-3σ Fobs-Fcalc

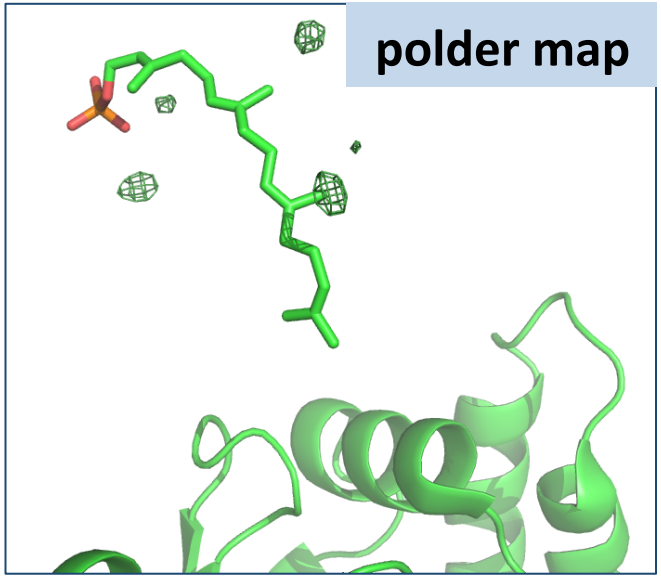
GRG =
geranyl geranyl diphosphate

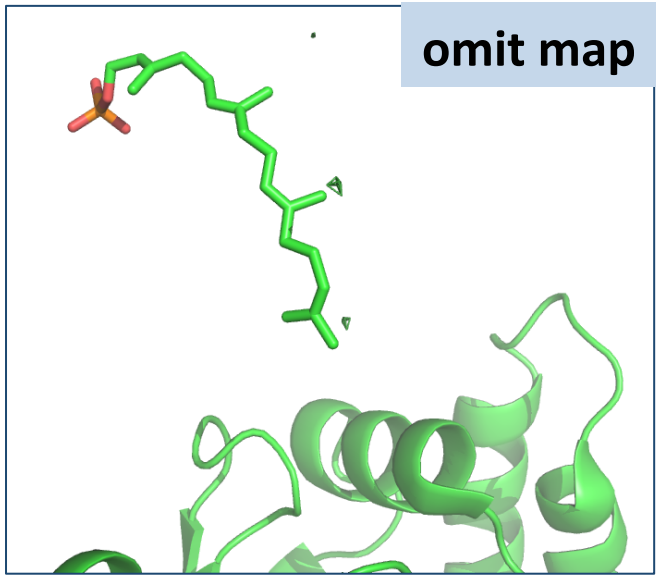



No supporting density




Bulk solvent has shape of ligand



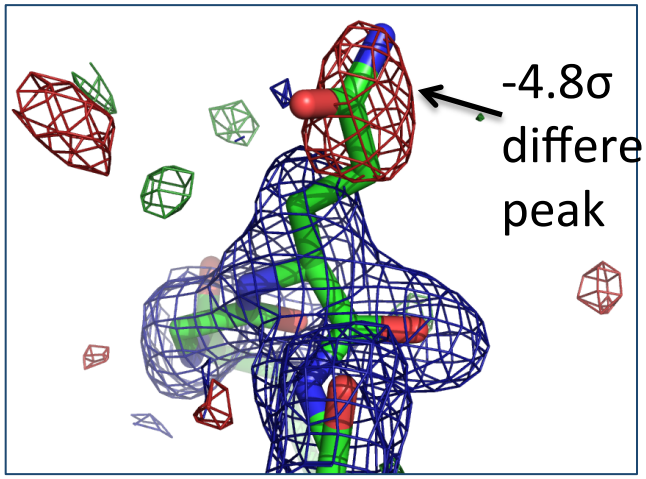




No density in omit and polder map

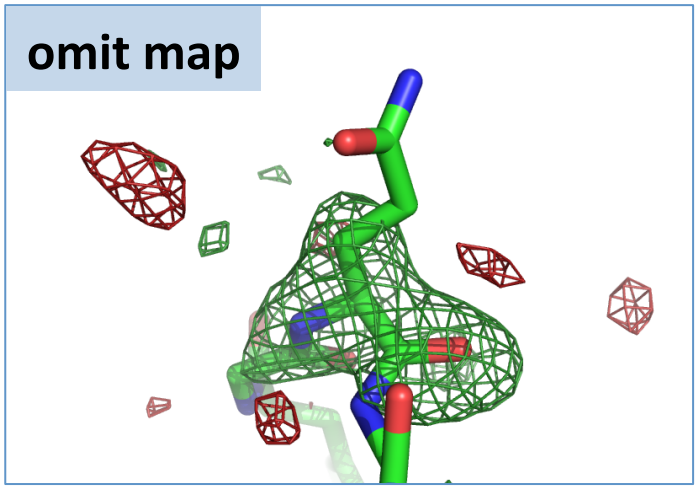
# Example: improved side chain density for Gln H105

PDB code: 1f8t
Resolution: 2.2 Å

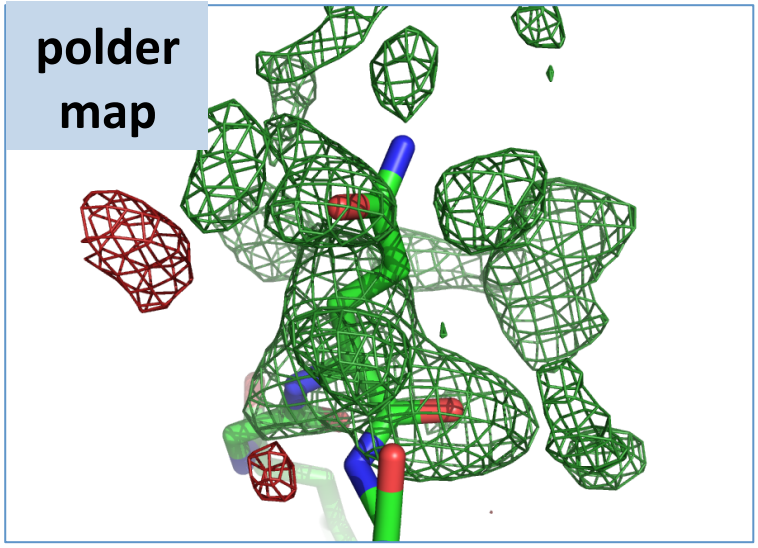
**Blue**:
+1σ 2Fobs-Fcalc
**Green**, **red**:
+/-3σ Fobs-Fcalc



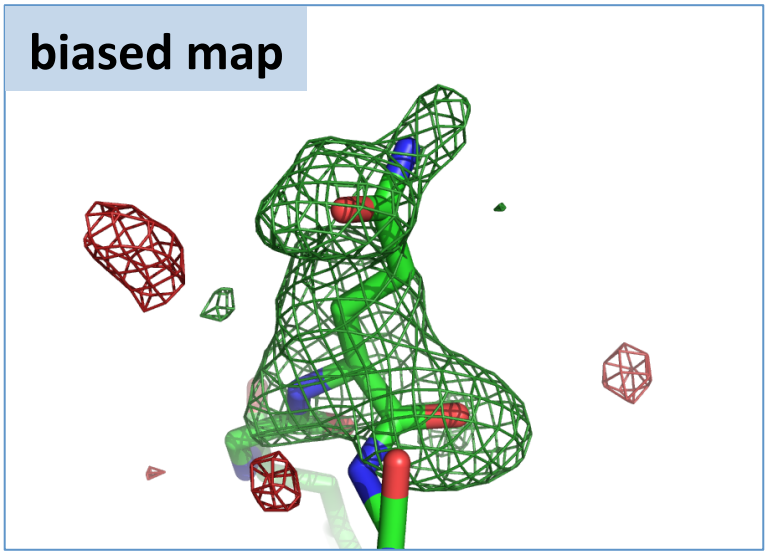

No density for side chain




No density for side chain



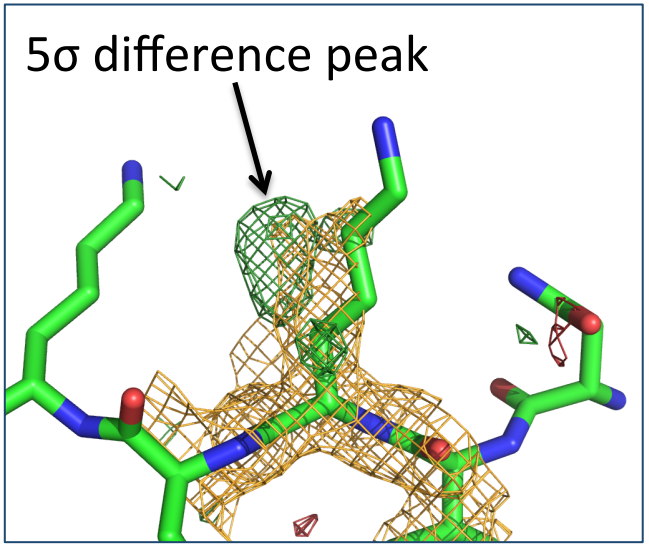

Continuous density showing slightly different orientation for side chain



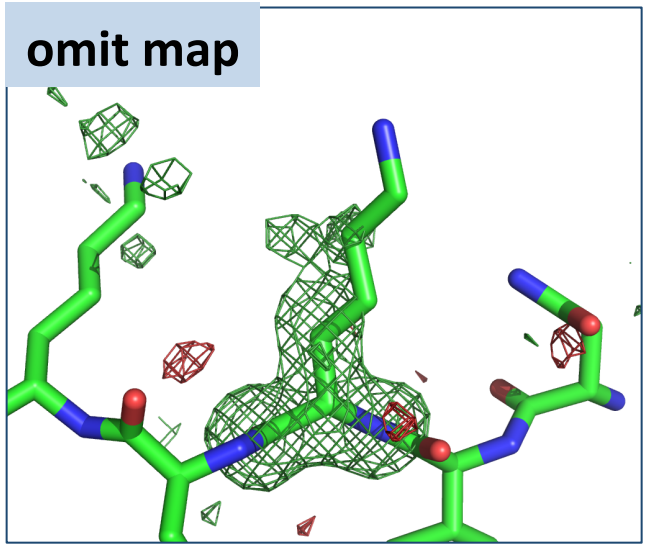

Contour of side chain orientation as it is in the model

# Example: double conformation Lys L147 (1f8t)

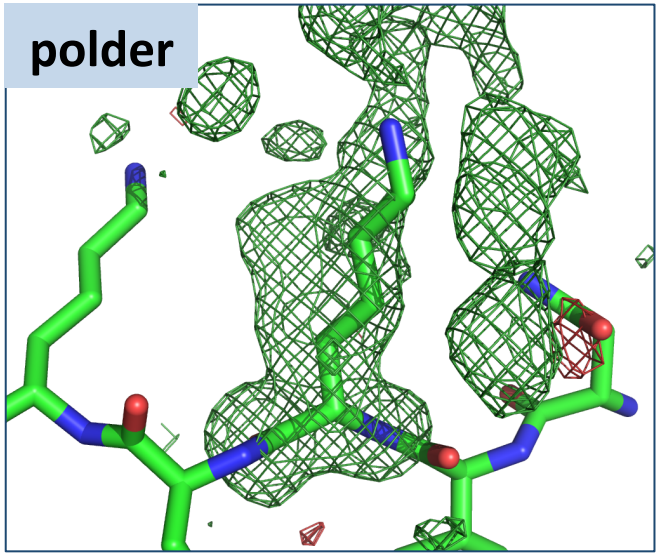
**Orange**:
+1σ 2Fobs-Fcalc
**Green**, **red**:
+/-3σ Fobs-Fcalc




No clear density for side chain

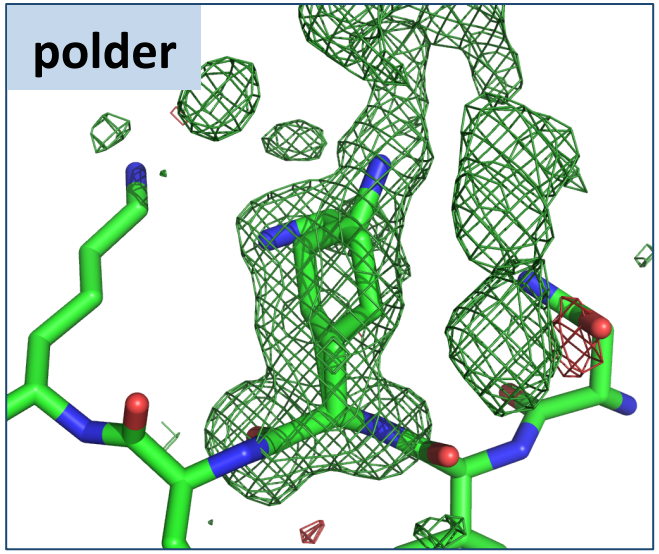



Side chain outside density

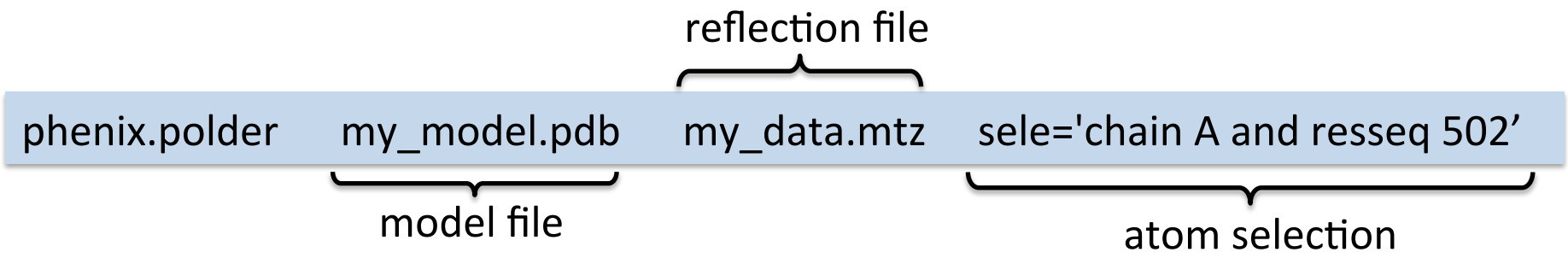



Initial side chain orientation not confirmed by polder map

Add alternate conformation



# How to run phenix.polder




**Example:**

→ Calculates polder map for residue 3 of chain A

```
> phenix.polder   3eka.pdb   3eka.mtz   selection='chain A and resseq 3'
```

```
phenix.polder is running...
selecting atoms...
Selection: chain A and resseq 3
Atoms selected: 12
HETATM 2540  C1  ASC A   3      16.014  23.278  61.896  0.50 47.98           C
HETATM 2541  C2  ASC A   3      16.636  22.943  60.650  0.50 44.98           C
HETATM 2542  C3  ASC A   3      15.793  22.221  59.939  0.50 44.86           C
HETATM 2543  C4  ASC A   3      14.540  21.934  60.721  0.50 47.27           C
HETATM 2544  C5  ASC A   3      14.352  20.433  60.978  0.50 49.24           C
HETATM 2545  C6  ASC A   3      12.948  20.119  61.566  0.50 48.19           C
...
```

The selected atoms are listed → it is good to check if the selection is correct

*https://www.phenix-online.org/documentation/reference/atom_selections.html

# How to run phenix.polder

```
Calculating solvent mask...
**************************************************

R factors for unmodified input model and data:
r_work=0.2114 r_free=0.2880
**************************************************

R factor for polder map
r_work=0.2153 r_free=0.2934
**************************************************

R factor when ligand is excluded for mask calculation:
r_work=0.2119 r_free=0.2903
**************************************************

Finished.
Time: 4.82
```

→ *R*-factors
(they should not change much, and not differ significantly between polder- and omit-map)

**Output files**

- polder_map_coeffs.mtz with coefficients:
  - mFo-DFc_polder → polder map
  - mFo-DFc_omit → omit map

→ ccp4 maps can be opened with PyMol and Coot

- mask_all.ccp4 → ccp4 map of the mask for model containing ligand
- mask_polder.ccp4 → mask where bulk solvent is excluded around ligand
- mask_omit.ccp4 → mask where bulk solvent floods into ligand pocket

# Acknowledgements


- **Lawrence Berkeley Laboratory**
  - Paul Adams, Pavel Afonine, Billy Poon, Oleg Sobolev, Nigel Moriarty
- **Los Alamos National Laboratory**
  - Tom Terwilliger, Li-Wei Hung
- **Cambridge University**
  - Randy Read, Airlie McCoy, Gabor Bunkoczi, Robert Oeffner
- **Duke University**
  - Jane Richardson & David Richardson, Chris Williams, Bradley Hintze
- **OpenEye**
  - Pawel Janowski, Greg Warren
- **Rutgers University**
  - Pawel Janowski, David Case
- **Other Collaborators**
  - Dale Tronrud, Donnie Berholz, Andy Karplus
  - Alexandre Urzhumtsev & Vladimir Lunin
  - Garib Murshudov & Alexi Vagin
  - Kevin Cowtan, Paul Emsley, Bernhard Lohkamp
  - David Abrahams
  - Phenix Testers & Users
- **Funding:**
  - NIH/NIGMS:
  - Phenix Industrial Consortium
  - Lawrence Berkeley Laboratory

# The Phenix Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Nigel Moriarty, Billy Poon, Oleg Sobolev, Dorothee Liebschner

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner, Richard Mifsud

Cambridge University

Duke University

Jane & David Richardson, Chris Williams, Bradley Hintze

*An NIH/NIGMS funded Program Project*