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/Å³

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

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

\[ F_{\text{model}} = k_{\text{overall}} \left( F_{\text{calc}(\text{atoms})} + F_{\text{bulk solvent}} \right) \]

Flat bulk solvent model: defining the solvent contribution

1. Compute solvent mask M:
   0 – inside protein,
   1 – outside (and in pockets)

2. Calculate structure factors from mask:
   \[ M \xrightarrow{FT} F_{\text{Mask}} \]

3. Define contribution to total structure factor:
   \[ F_{\text{bulk solvent}} = k \cdot F_{\text{Mask}} \]
   
   \( k \) represents mean solvent density in e\( Å^{-3} \), modulated by a smearing factor

---

Flat bulk solvent model: how to define the solvent mask

Starting point: atomic model

Define grid

Set mask to 1 outside the circles, 0 inside

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

Shrinking: Reset some grid points inside the circles to 1

$F_{\text{Mask}}$

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

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.

*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.

- 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
**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

*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
$R_{\text{free}}$: 0.231
$R_{\text{work}}$: 0.190
Residues: 453

Density for (almost) entire ligand
Density has exactly the shape of the ligand \( \rightarrow \) bias
Example: improved density for solvent molecule MES A 88

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

Place ligand into solvent area

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

-4.8σ difference peak

Continuous density showing slightly different orientation for side chain

Biased map

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

5σ difference peak

omit map
# 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

<table>
<thead>
<tr>
<th>HETATM 2540 C1 ASC A 3</th>
<th>16.014</th>
<th>23.278</th>
<th>61.896</th>
<th>0.50</th>
<th>47.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>HETATM 2541 C2 ASC A 3</td>
<td>16.636</td>
<td>22.943</td>
<td>60.650</td>
<td>0.50</td>
<td>44.98</td>
</tr>
<tr>
<td>HETATM 2542 C3 ASC A 3</td>
<td>15.793</td>
<td>22.221</td>
<td>59.939</td>
<td>0.50</td>
<td>44.86</td>
</tr>
<tr>
<td>HETATM 2543 C4 ASC A 3</td>
<td>14.540</td>
<td>21.934</td>
<td>60.721</td>
<td>0.50</td>
<td>47.27</td>
</tr>
<tr>
<td>HETATM 2544 C5 ASC A 3</td>
<td>14.352</td>
<td>20.433</td>
<td>60.978</td>
<td>0.50</td>
<td>49.24</td>
</tr>
<tr>
<td>HETATM 2545 C6 ASC A 3</td>
<td>12.948</td>
<td>20.119</td>
<td>61.566</td>
<td>0.50</td>
<td>48.19</td>
</tr>
</tbody>
</table>

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*
Output files

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

- 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

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)
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

Duke University
Jane & David Richardson, Chris Williams, Bradley Hintze

An NIH/NIGMS funded Program Project