**h2o-overlap User's Guide**

**Installation instructions:**

After downloading h2o-overlap.pl, type in a terminal:

```
chmod +x h2o-overlap.pl
```

The program can be executed by typing:

```
./h2o-overlap.pl
```

**Purpose:**

The determination of viral structural proteins via x-ray crystallography often involves working with a crystallographic asymmetric unit that consists of a multimer composed of non-crystallographically related monomers. Unless the structure is solved at a sufficiently high enough resolution, the density around each monomer is assumed to be identical. Thus, the final coordinates are only of a single monomer which are then deposited in the protein data bank.

At a resolution better than 2.8Å it becomes possible to place water molecules in the crystal structure. A conventional approach is to mask the density around a single monomer and pick waters accordingly. However, a complication arises when NCS is applied to generate the crystallographic asymmetric unit. Overlapping, or near overlapping waters are generated at the interface between monomers. In respect to the capsid symmetry these overlapping waters are likely equal contributors to the putative water density. Thus, to perform refinement of the crystallographic asymmetric unit coordinates, for example in REFMAC, it becomes necessary to examine overlapping waters and reduce the occupancy of these waters accordingly. For example, a water molecule placed closed to a three-fold symmetry axis in an icosahedral virus would require that its occupancy be set to 0.33 to account for the presence of the two additional symmetry related water molecules.

The program h2o-overlap was designed to easily adjust the occupancy of these overlapping waters without requiring the structural biologist to manually inspect the distance between waters. The user defines a tolerance in angstroms with which two waters will be considered overlapping. The program then examines the distance between all waters in an inputted PDB reducing the occupancy as necessary for waters whose distance is less than or equal to the previously set tolerance. The output is written to a separate PDB file.

**Example case:**

The structure of the main capsid protein of AAV8 was previously solved to 2.60 angstroms (PDB ID 2QA0). The deposited structure consists of a single monomer with coordinates for 89 waters. Application of BIOMT matrices 2, 3, 6, 10, 22, 23, 41, and 50 in the 2QA0 pdb will generate the crystallographic asymmetric unit, a 10-mer, with the two-fold, three-fold, and most of the five-fold related monomers. These matrices can applied by using the LSQKAB program in CCP4, for example.

Users who prefer to avoid applying the matrices themselves can download the 10-mer below:

[2QA0-10-mer.pdb]

Inspection of water 739 in chain A shows two additional waters related by C3 symmetry. These waters
are separated by 1.77 angstroms – too close for hydrogen bonding. The occupancy of these waters must therefore be set to 0.33.

Run h2o-overlap in the directory containing the above 10-mer by typing in a terminal:
./h2o-overlap.pl

Set the tolerance to 1.8.

The input pdb is 2QA0-10-mer.pdb

The output pdb can be called, for example, 2QA0-10-mer-h2o-overlap.pdb

Viewing the output PDB indicates that the occupancy for waters 729 in chains A, C, B, D, E, F, G, H, and I are all set to 0.33.

A REMARK is written at the top of the PDB indicating what tolerance was used at the time h2o-overlap was run.