Comparing AceDRG vs CSD

LIGANDS Meeting Sept 2017

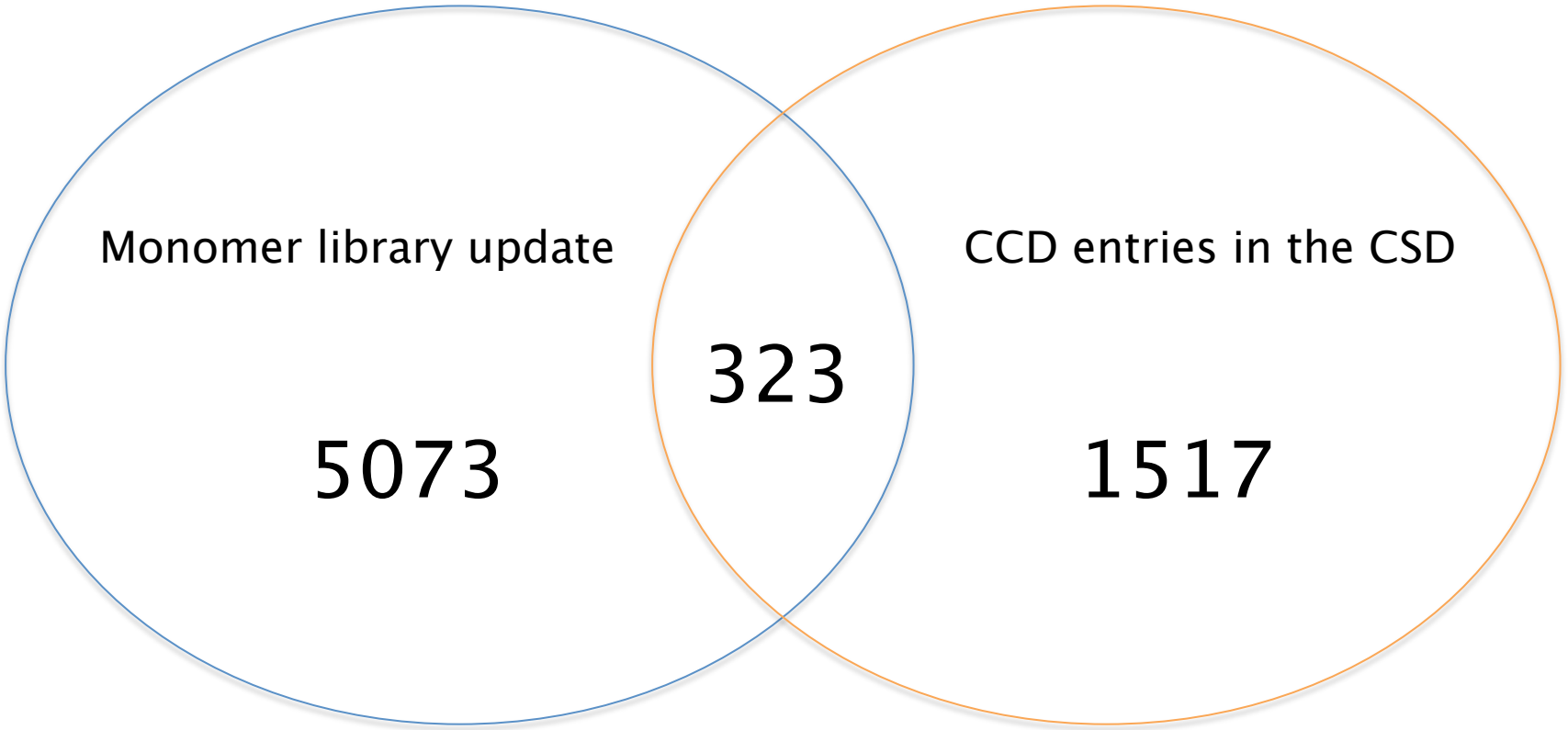
Rob Nicholls
nicholls@mrc-lmb.cam.ac.uk

MRC | Laboratory of Molecular Biology

CCP4

AceDRG vs CSD

Suggested by Oliver Smart



AceDRG vs CSD

323 Comp IDs from the CCD

Establish Baseline – original PDB files:

- Take PDBs
- Extract ligand coordinates
- Run REFMAC5 to assess geometry (idealisation mode)

AceDRG vs CSD

323 Comp IDs from the CCD

Establish Baseline – original PDB files:

- Take PDBs
- Extract ligand coordinates
- Run REFMAC5 to assess geometry (idealisation mode)

Re-refined PDB using AceDRG dictionary (monomer library update)

- Extract ligand coordinates
- Run REFMAC5 to assess geometry (idealisation mode)

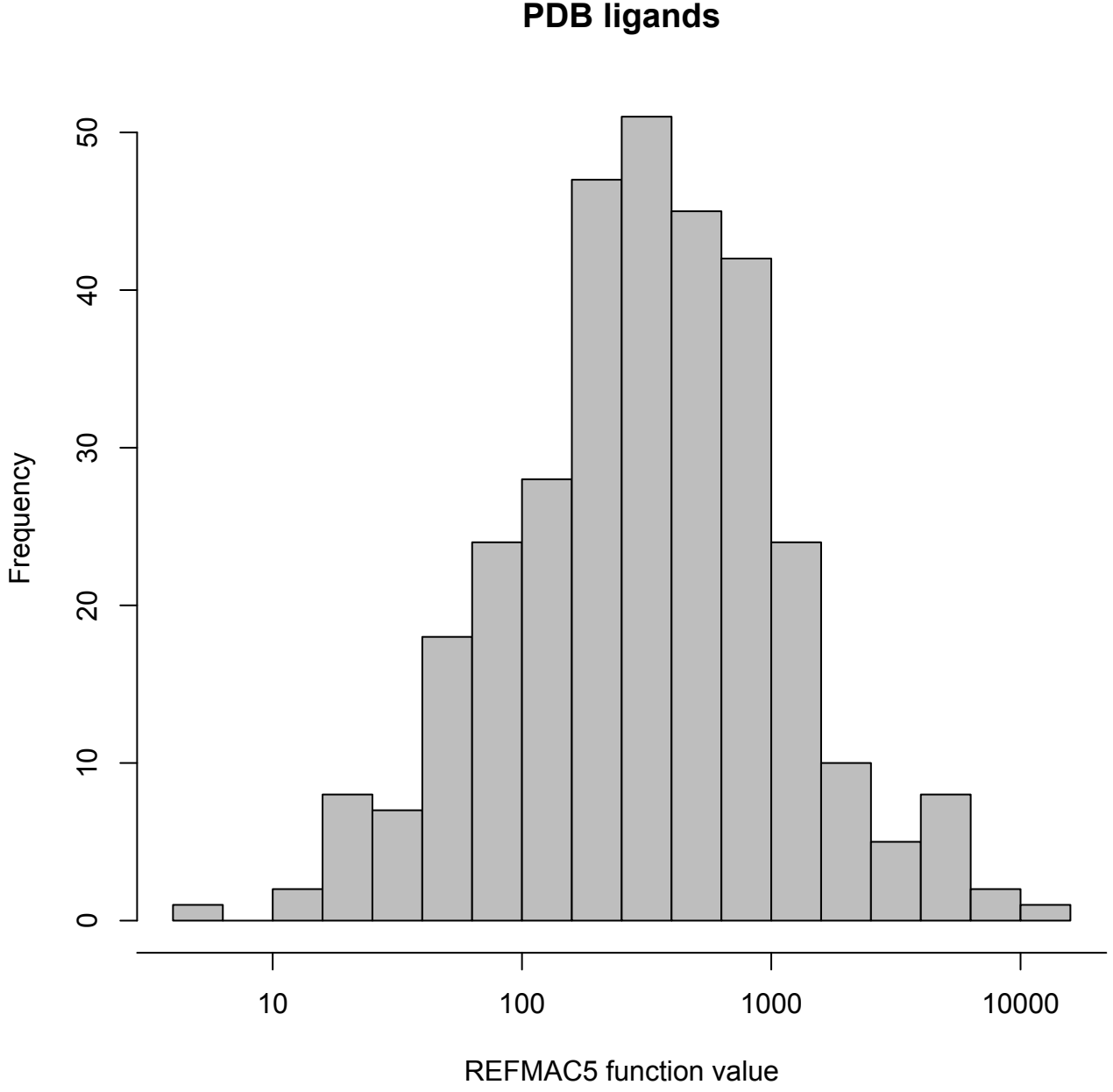
AceDRG vs CSD

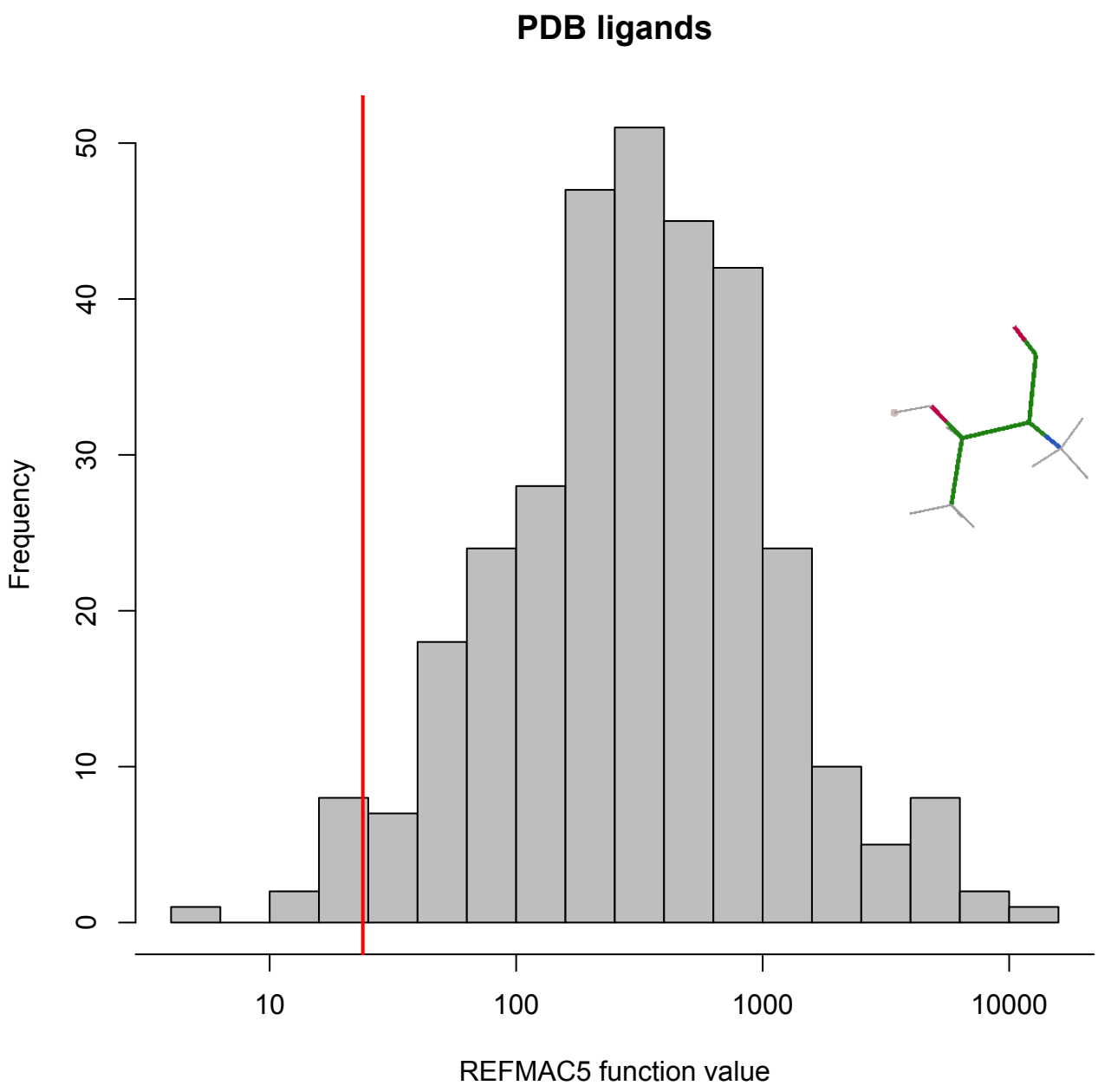
323 Comp IDs from the CCD

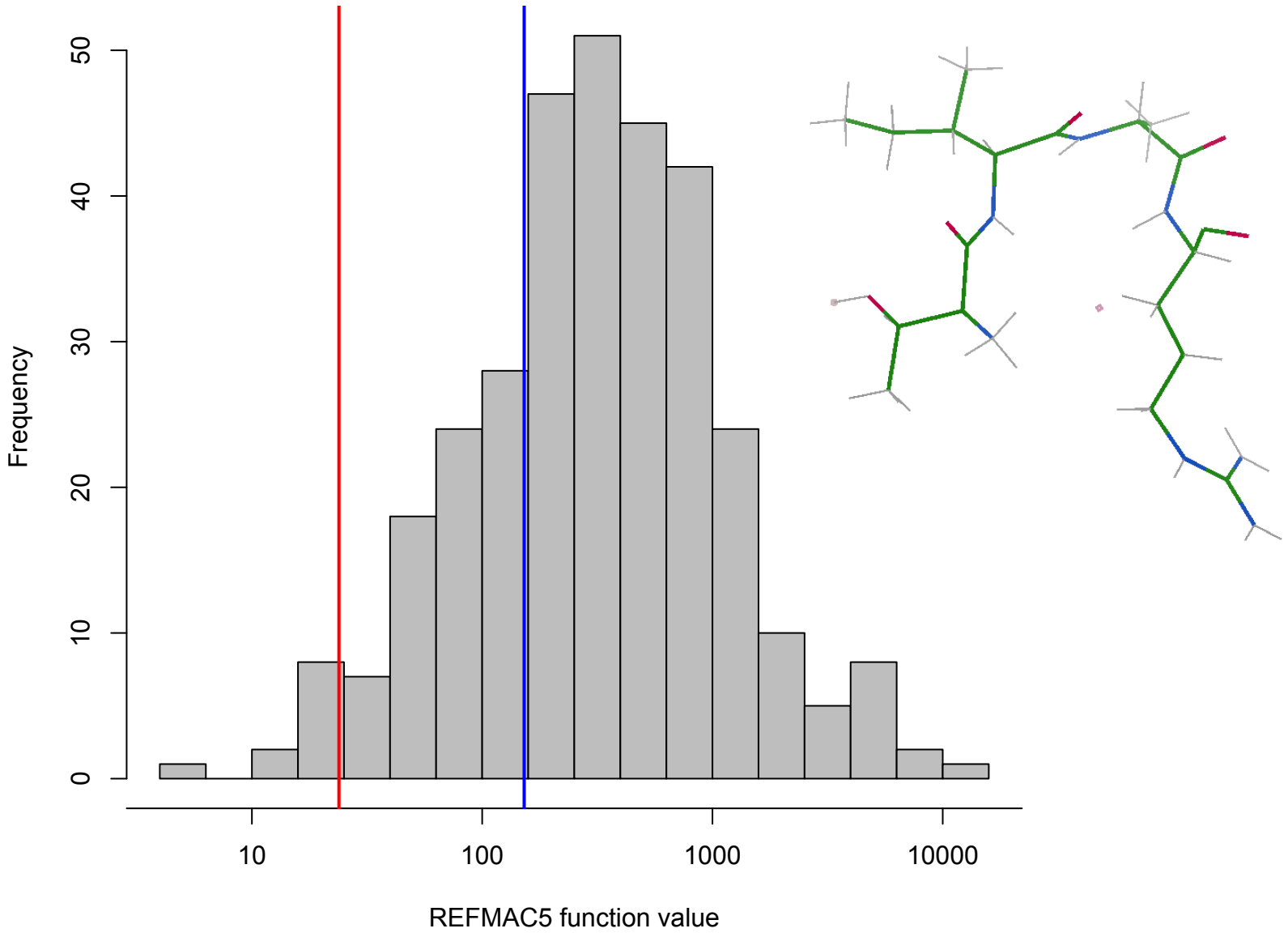
Establish Baseline – original PDB files:
- Take PDBs
- Extract ligand coordinates
- Run REFMAC5 to assess geometry (idealisation mode)

Re-refined PDB using AceDRG dictionary (monomer library update)
- Extract ligand coordinates
- Run REFMAC5 to assess geometry (idealisation mode)

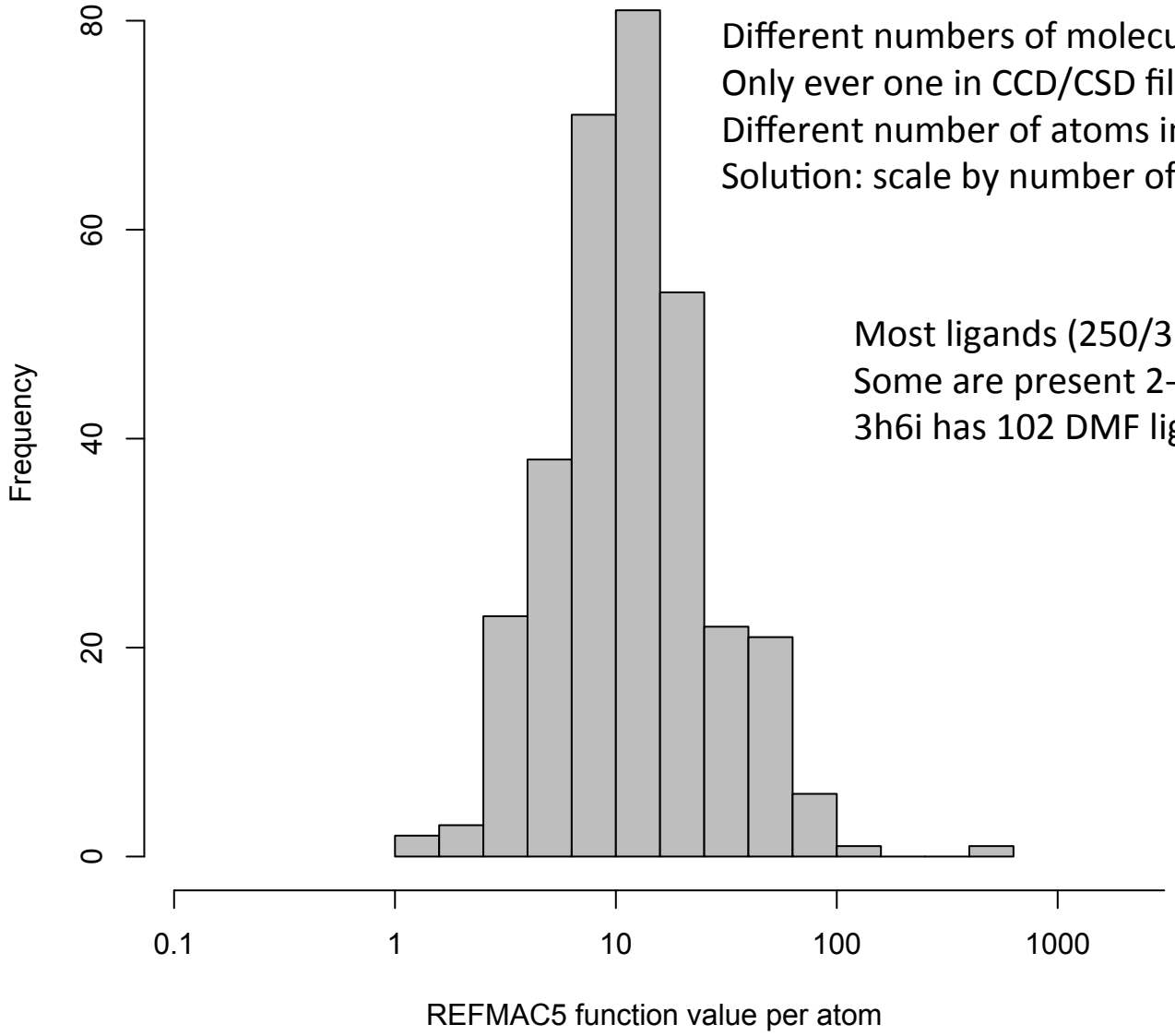
Compare with CSD model:
- Take coordinates
- Run REFMAC5 to assess geometry (idealisation mode)


PDB ligands
Frequency
0
10
20
30
40
50
10
100
1000
10000
REFMAC5 function value


PDB ligands
Frequency
0
10
20
30
40
50
10
100
1000
10000
REFMAC5 function value

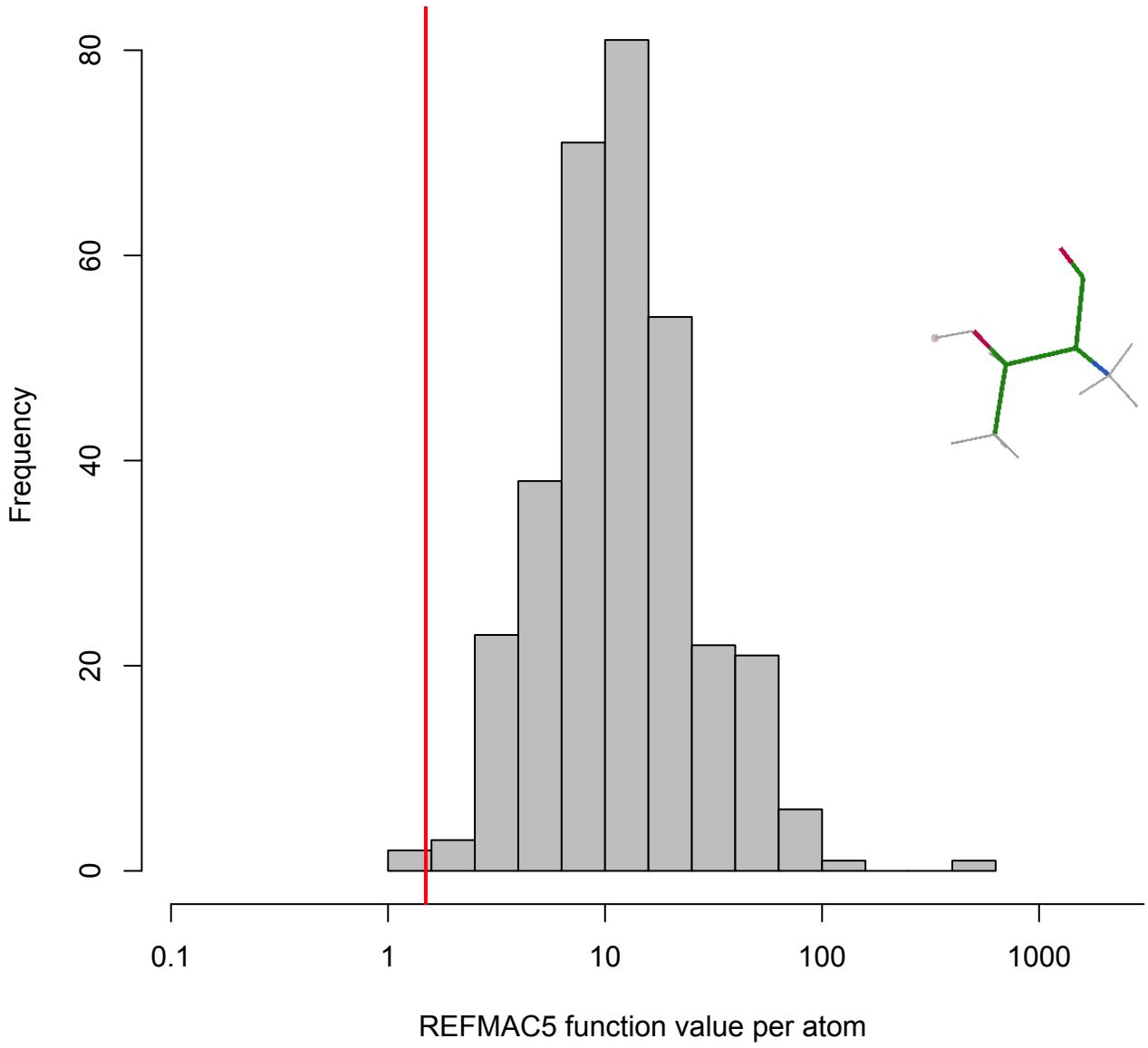

PDB ligands
Frequency
0
10
20
30
40
50
10
100
1000
10000
REFMAC5 function value

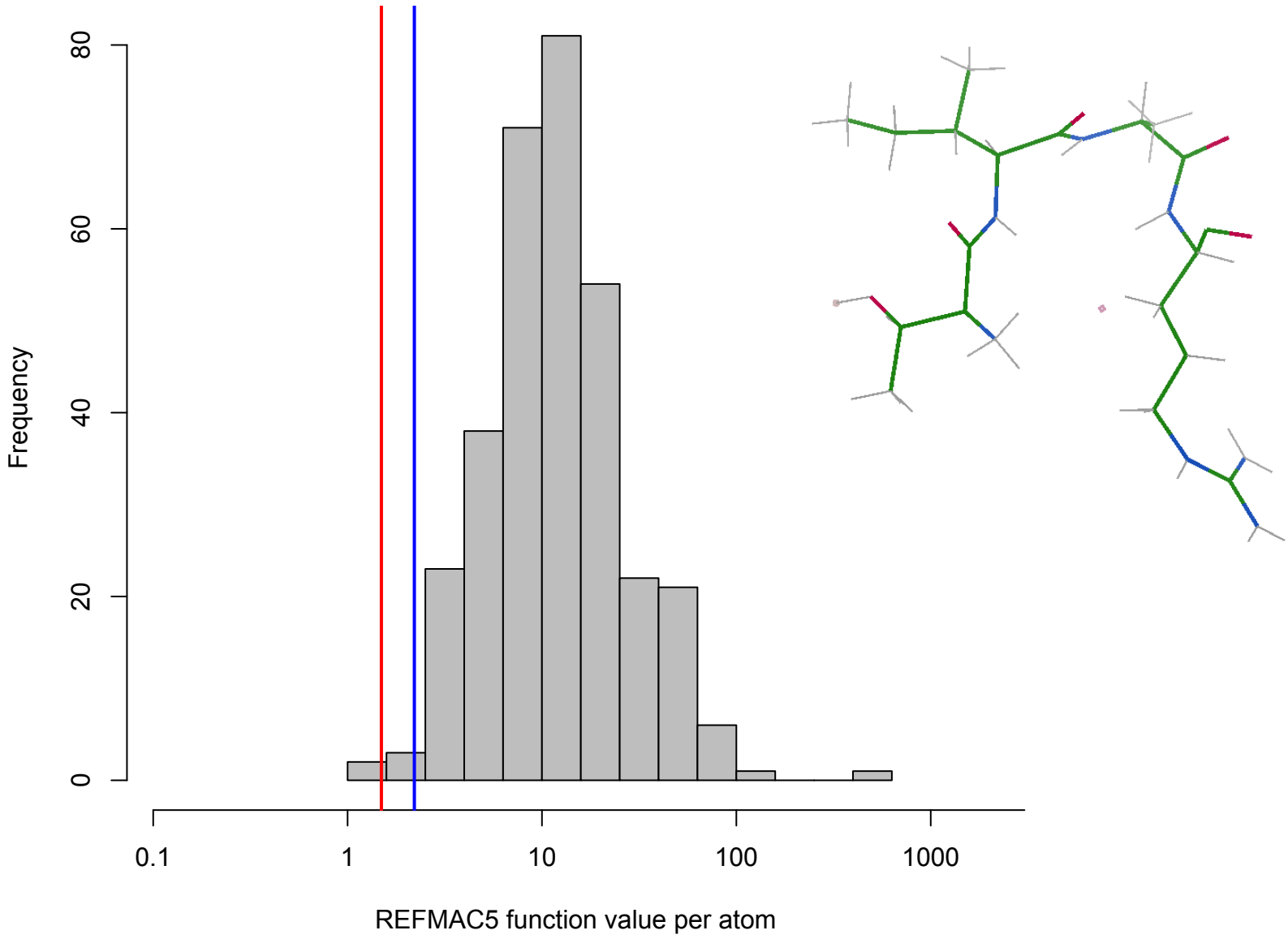
PDB ligands

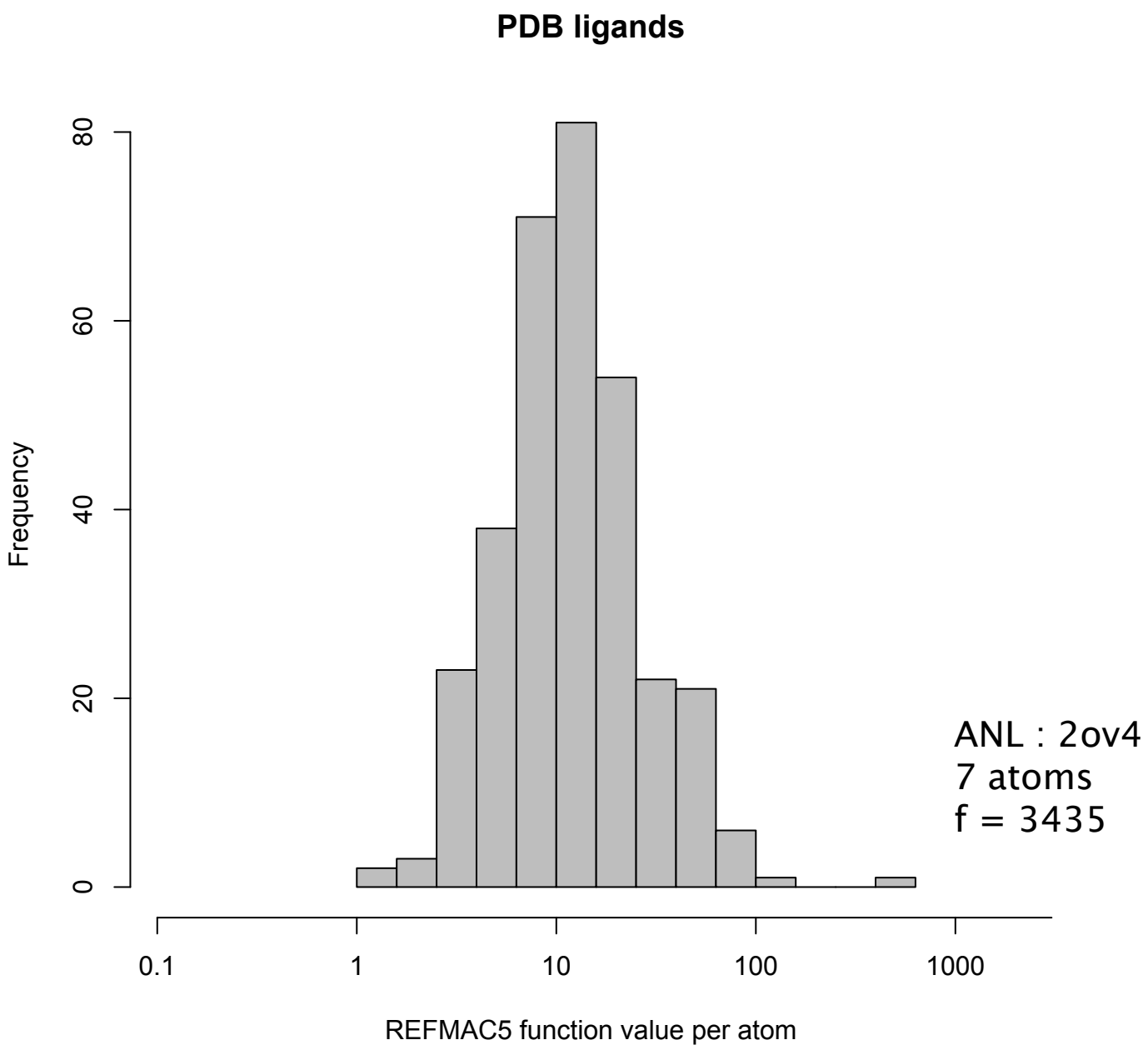


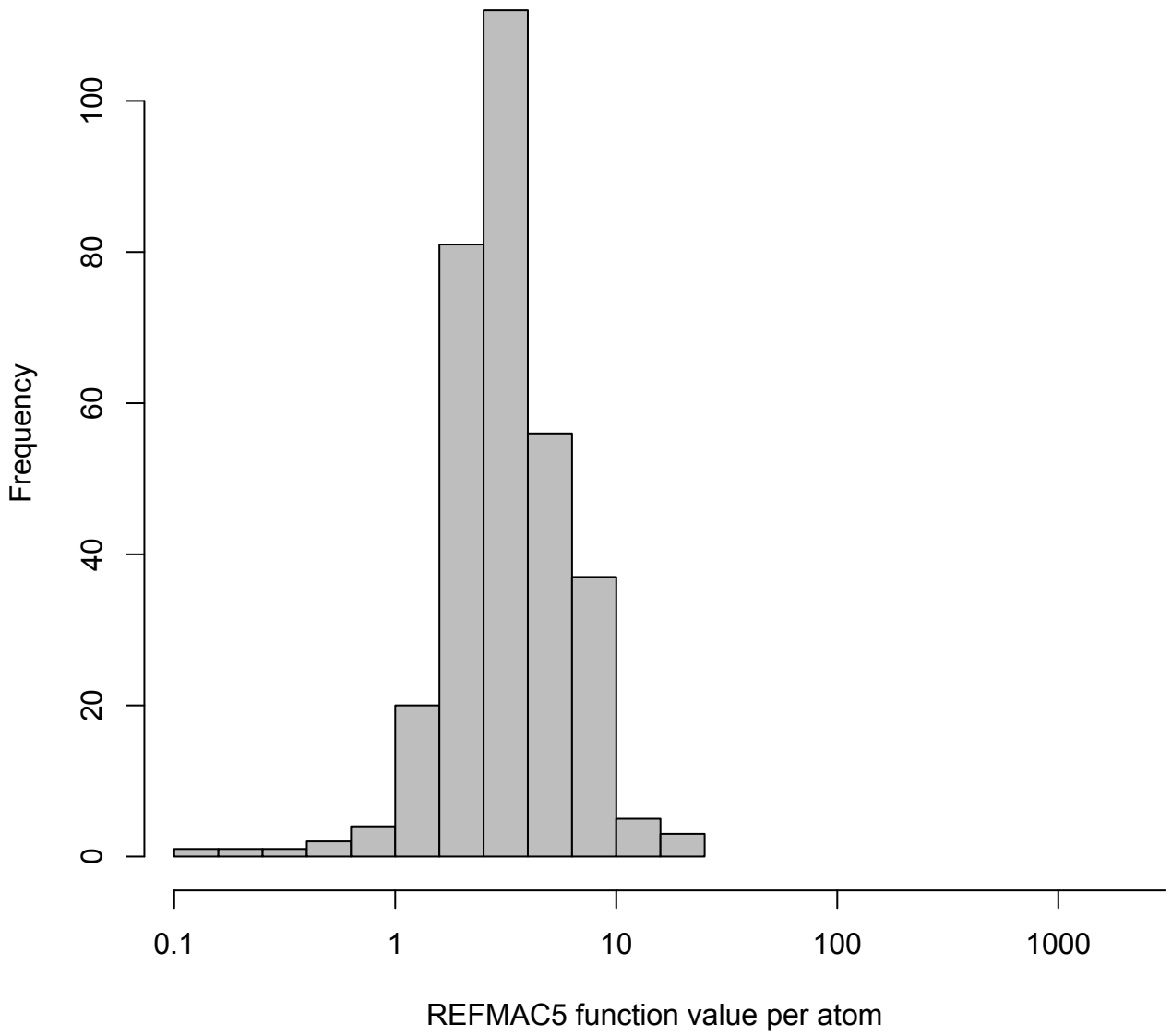
Different numbers of molecules in PDB file.
Only ever one in CCD/CSD file.
Different number of atoms in different ligands.
Solution: scale by number of atoms.

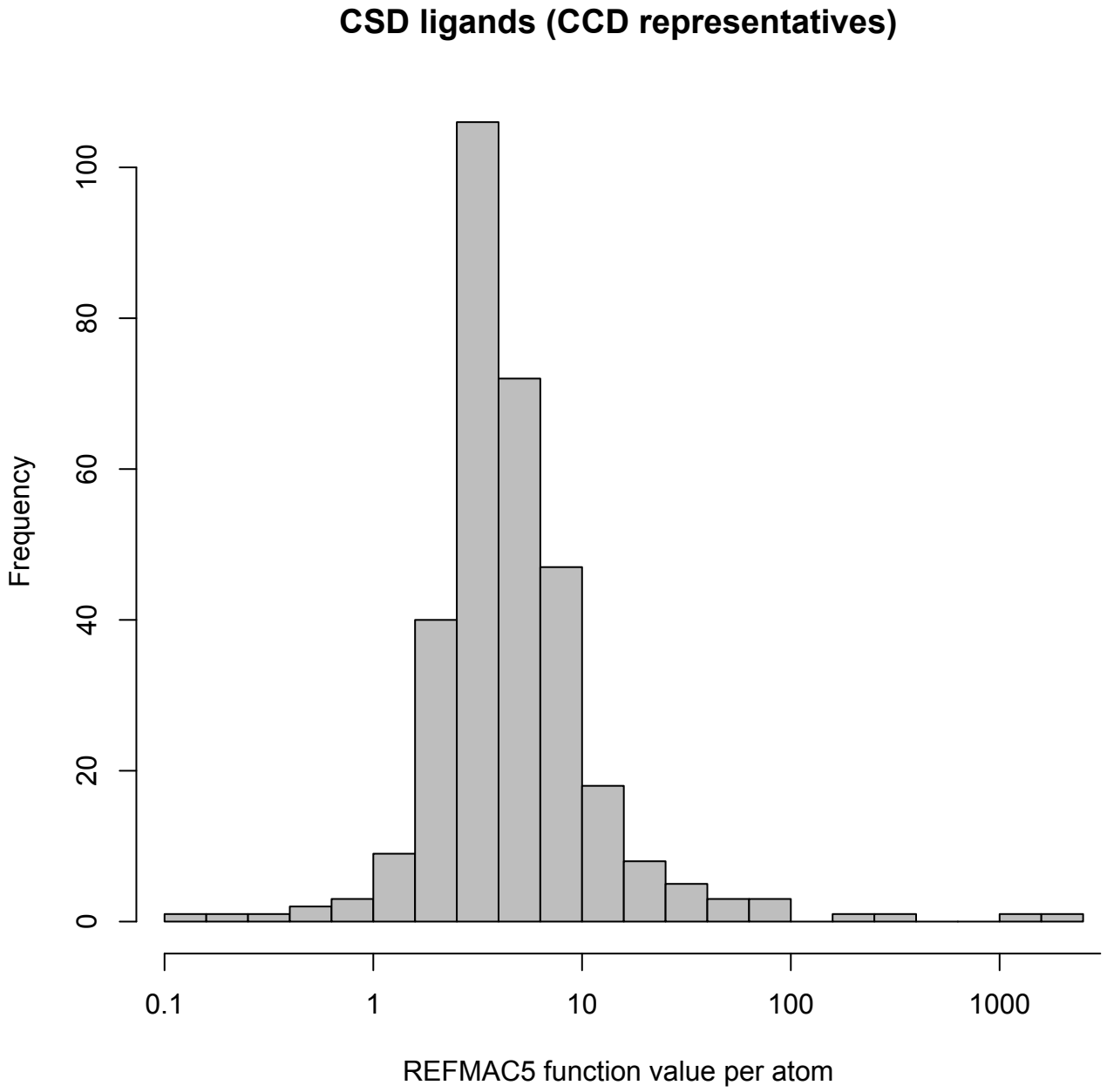
Most ligands (250/323) present once.
Some are present 2-11 times.
3h6i has 102 DMF ligands.

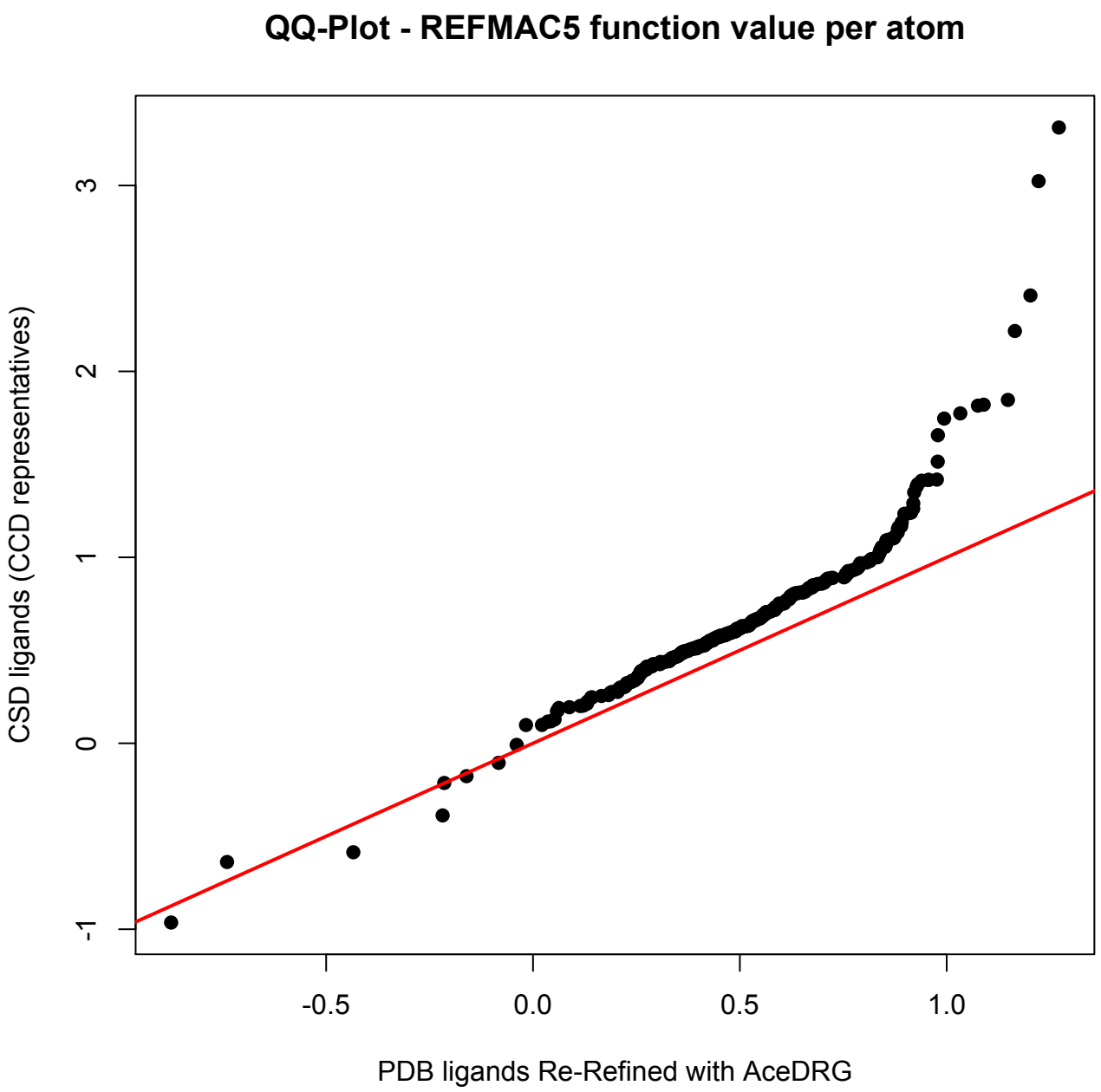

PDB ligands
Frequency
0
20
40
60
80
0.1
1
10
100
1000
REFMAC5 function value per atom

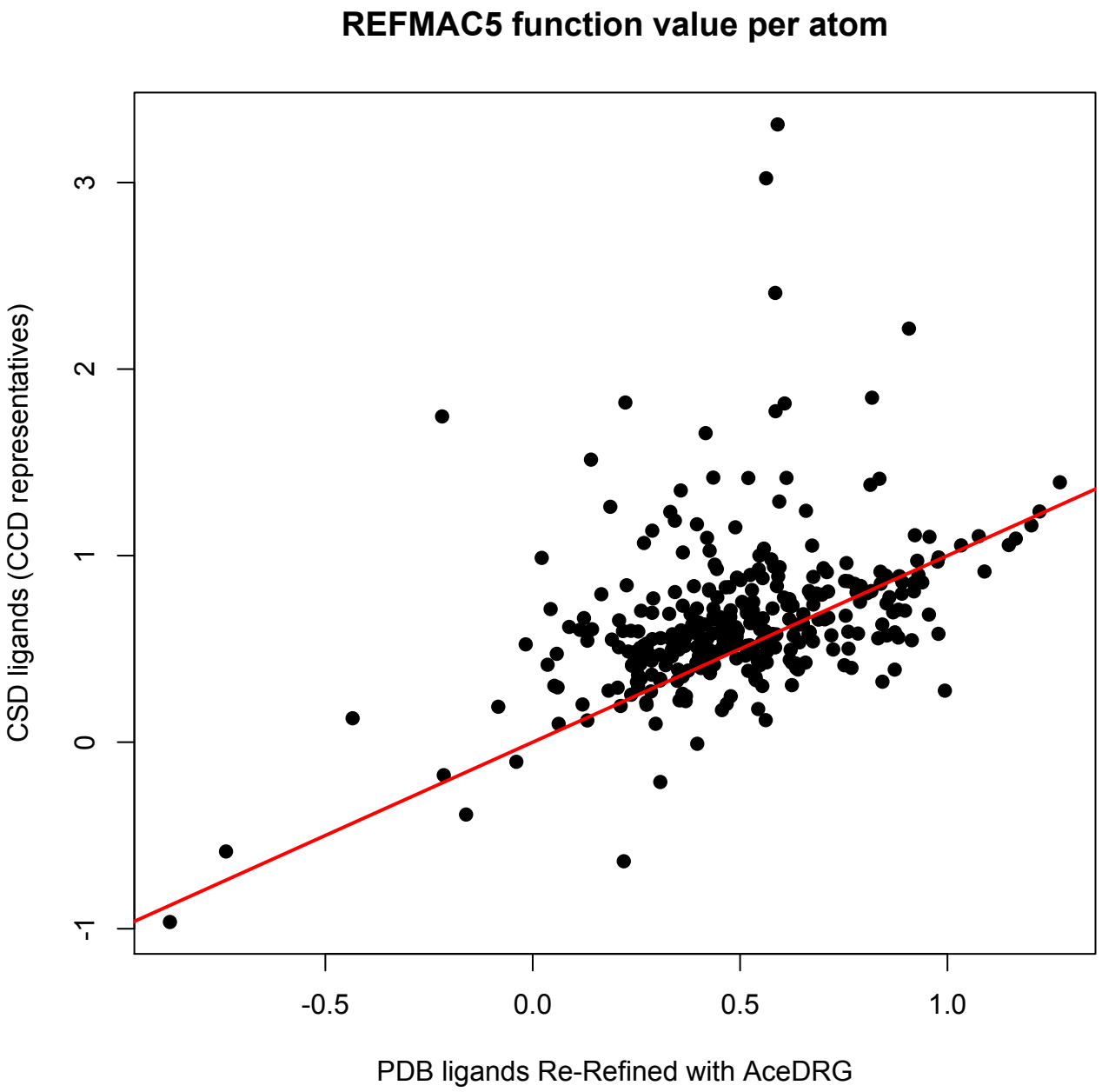

PDB ligands
Frequency
0
20
40
60
80
0.1
1
10
100
1000
REFMAC5 function value per atom

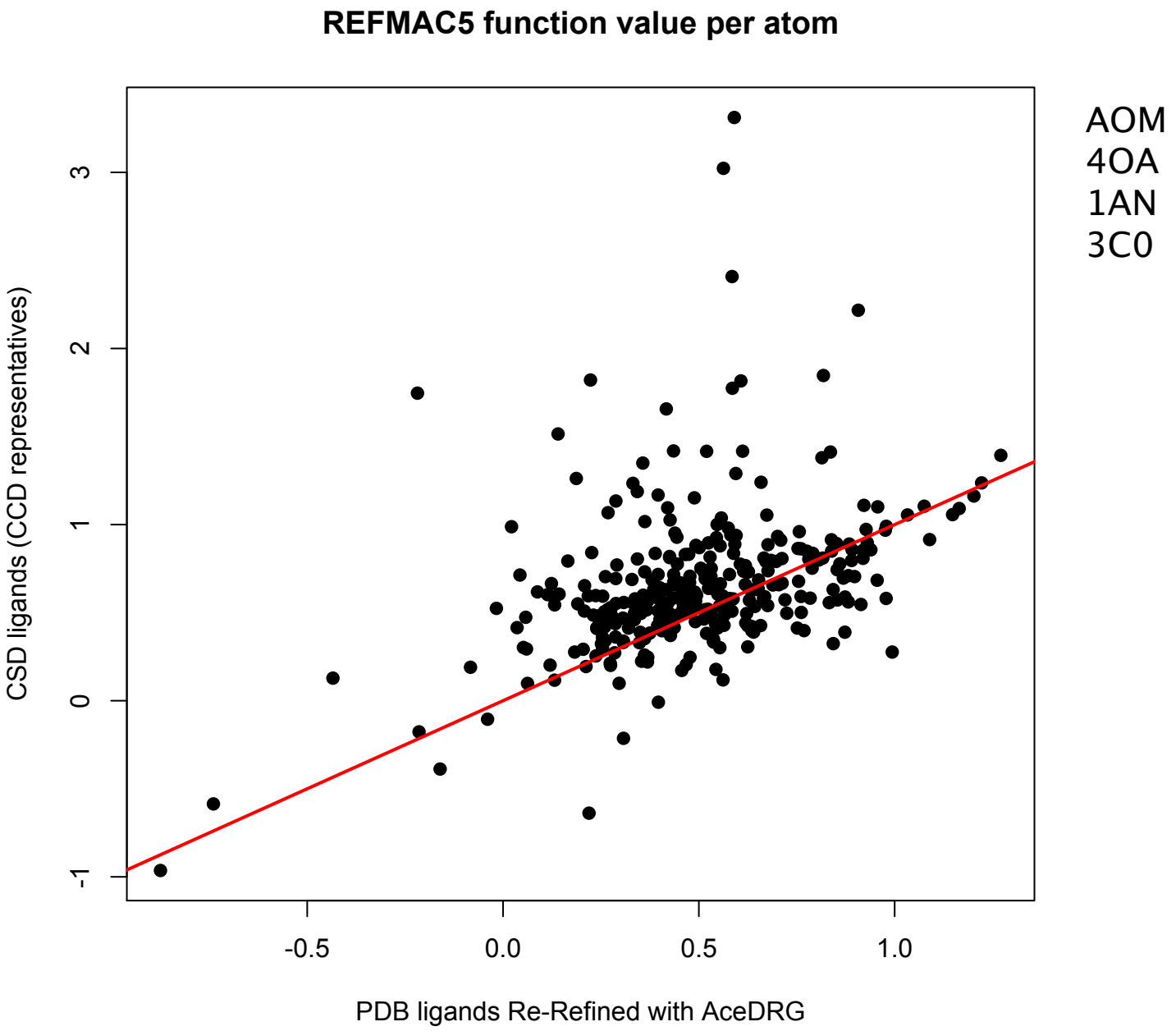

PDB ligands
Frequency
0
20
40
60
80
0.1
1
10
100
1000
REFMAC5 function value per atom
ANL : 2ov4
7 atoms
f = 3435


PDB ligands Re-Refined with AceDRG
Frequency
0
20
40
60
80
100
0.1
1
10
100
1000
REFMAC5 function value per atom

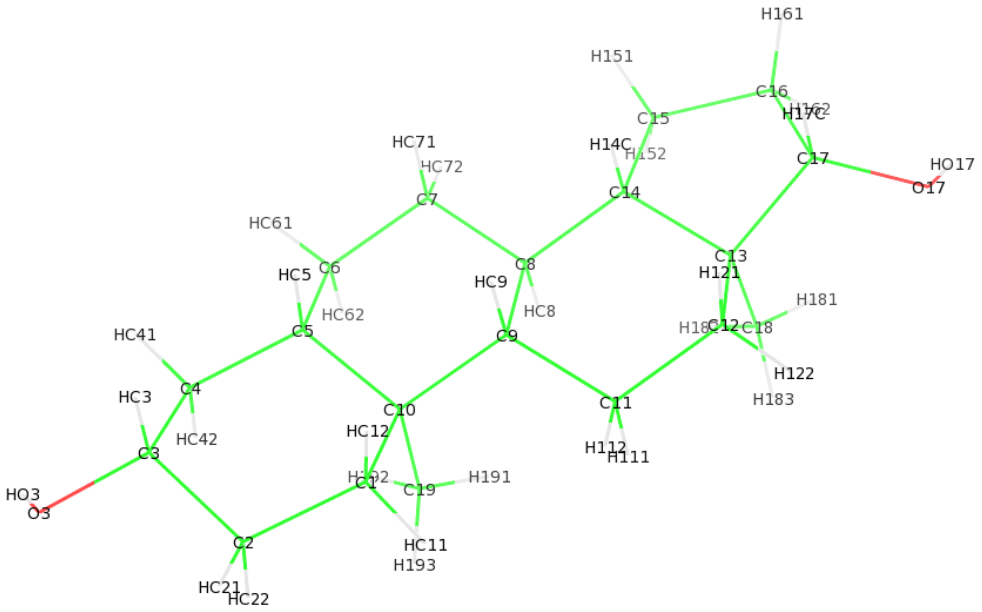

CSD ligands (CCD representatives)
Frequency
0
20
40
60
80
100
0.1
1
10
100
1000
REFMAC5 function value per atom


QQ-Plot - REFMAC5 function value per atom
CSD ligands (CCD representatives)
PDB ligands Re-Refined with AceDRG
-1
0
1
2
3
-0.5
0.0
0.5
1.0

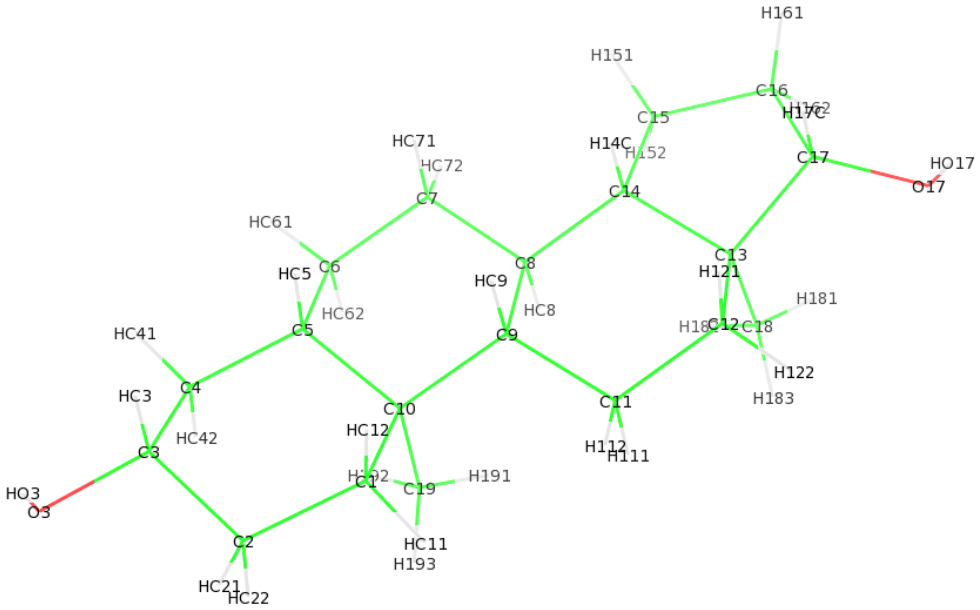

REFMAC5 function value per atom
CSD ligands (CCD representatives)
PDB ligands Re-Refined with AceDRG
-1
0
1
2
3
-0.5
0.0
0.5
1.0


REFMAC5 function value per atom
CSD ligands (CCD representatives)
PDB ligands Re-Refined with AceDRG
-1
0
1
2
3
-0.5
0.0
0.5
1.0
AOM
4OA
1AN
3C0

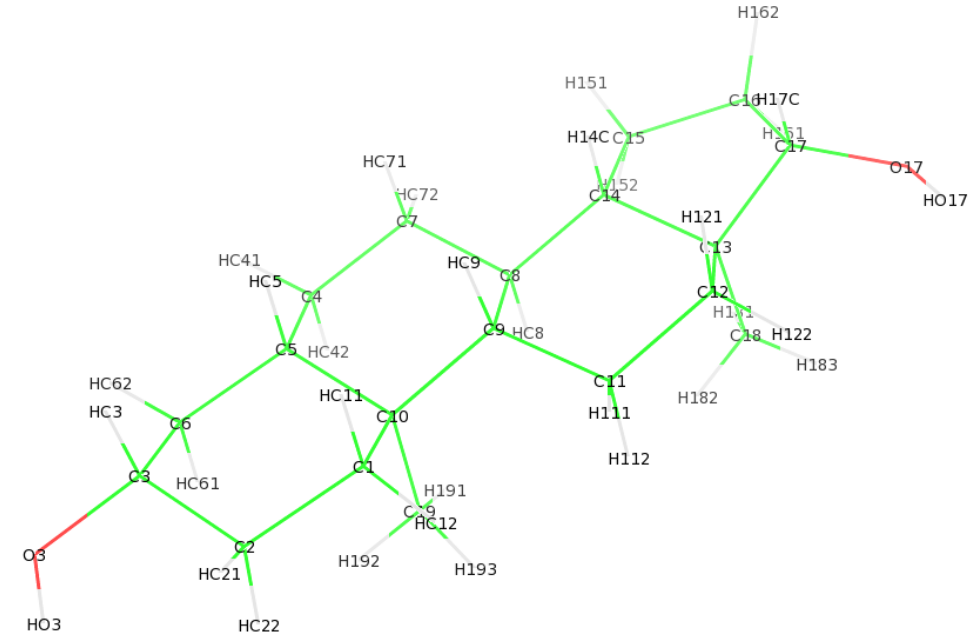
AOM : 1lho


H161
H151
C16
C15
H17C
H14C
H152
C17
HO17
O17
HC71
HC72
C7
C14
HC61
C6
C8
C13
H121
HC5
HC9
HC8
H181
HC62
C5
C9
C12
C18
HC41
H122
C4
HC3
C10
C11
H183
HC42
HC12
H112
H111
C3
C1
C19
H191
HO3
O3
C2
HC11
H193
HC21
HC22

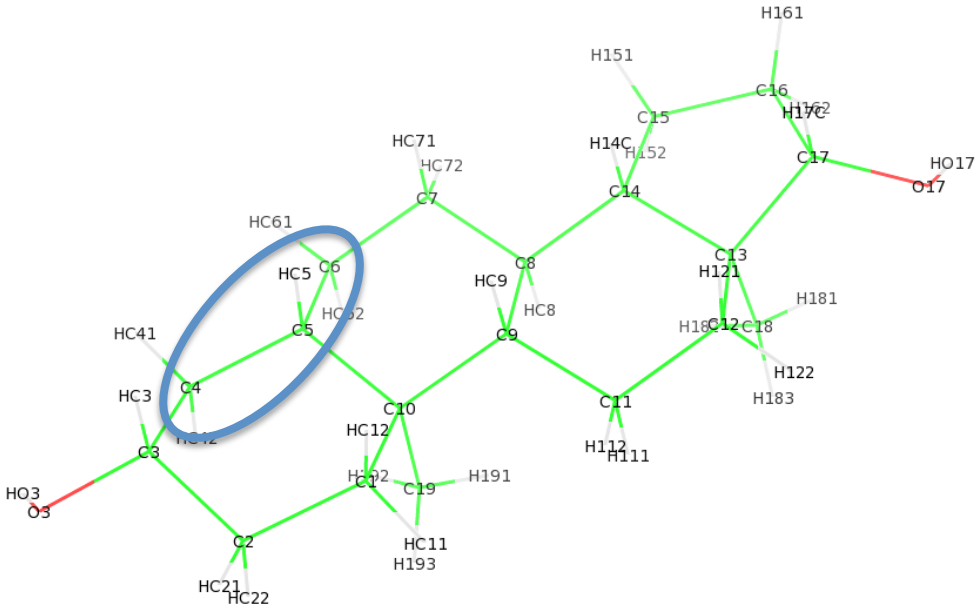
AOM : 1lho



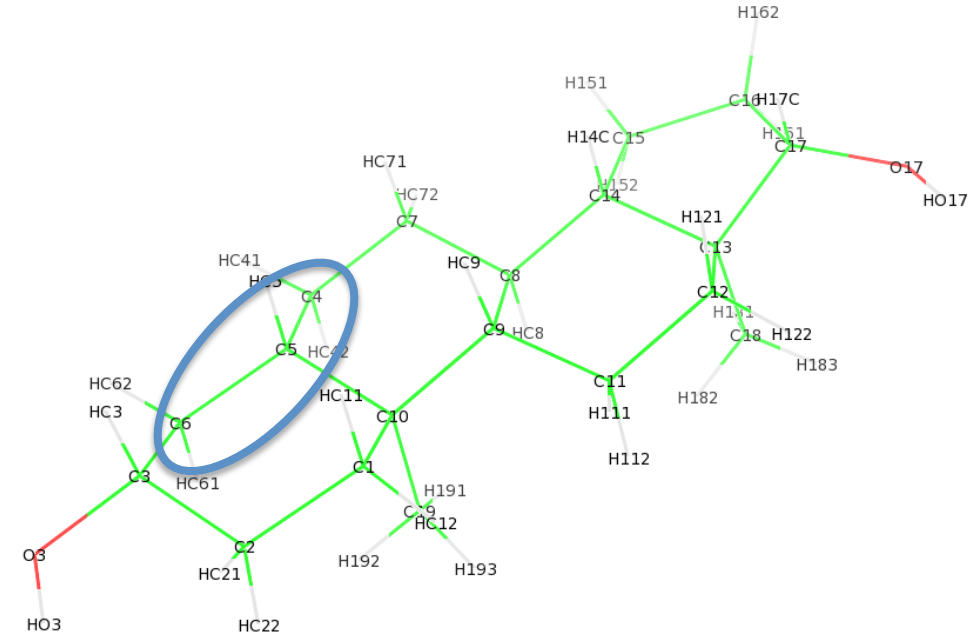
AOM (CSD)



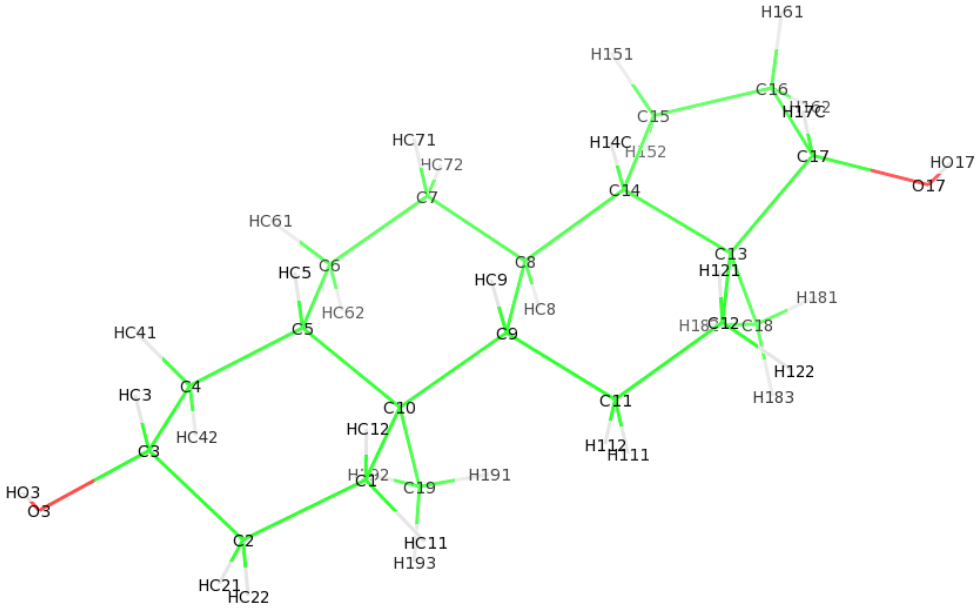
AOM : 1lho



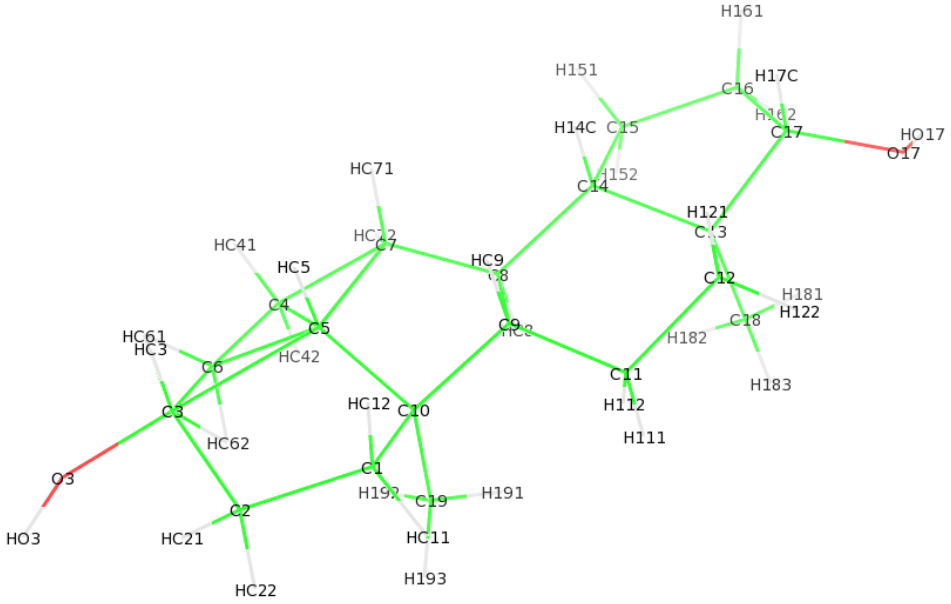
AOM (CSD)



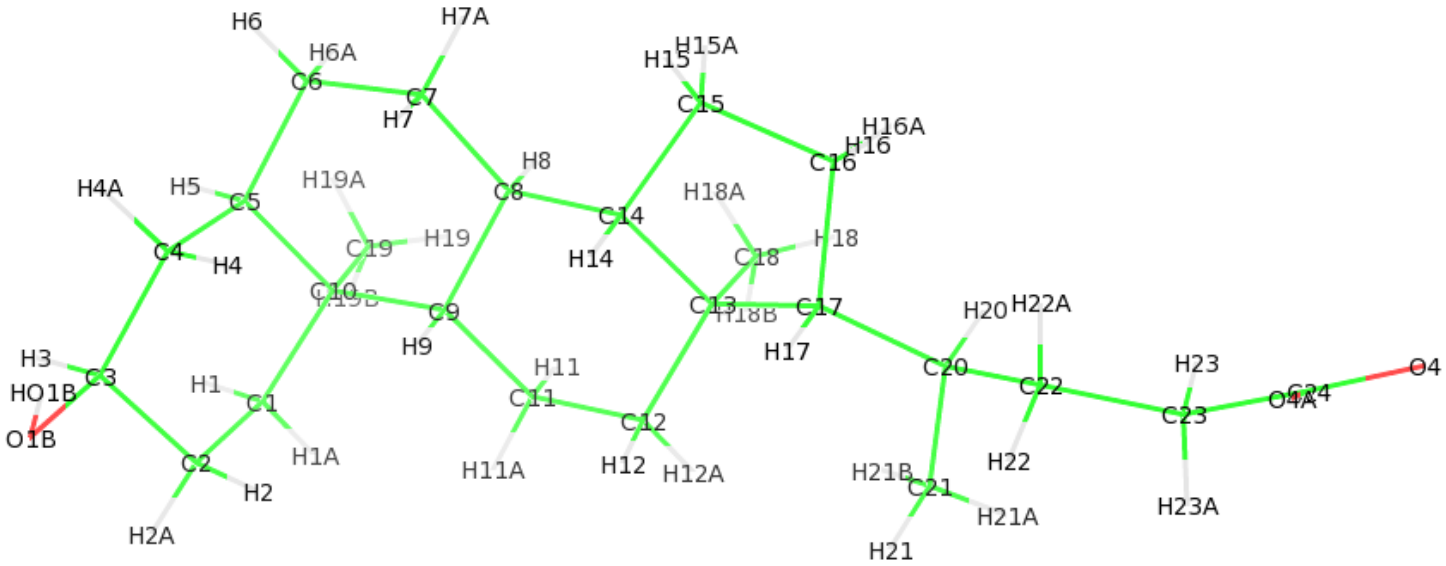
AOM : 1lho



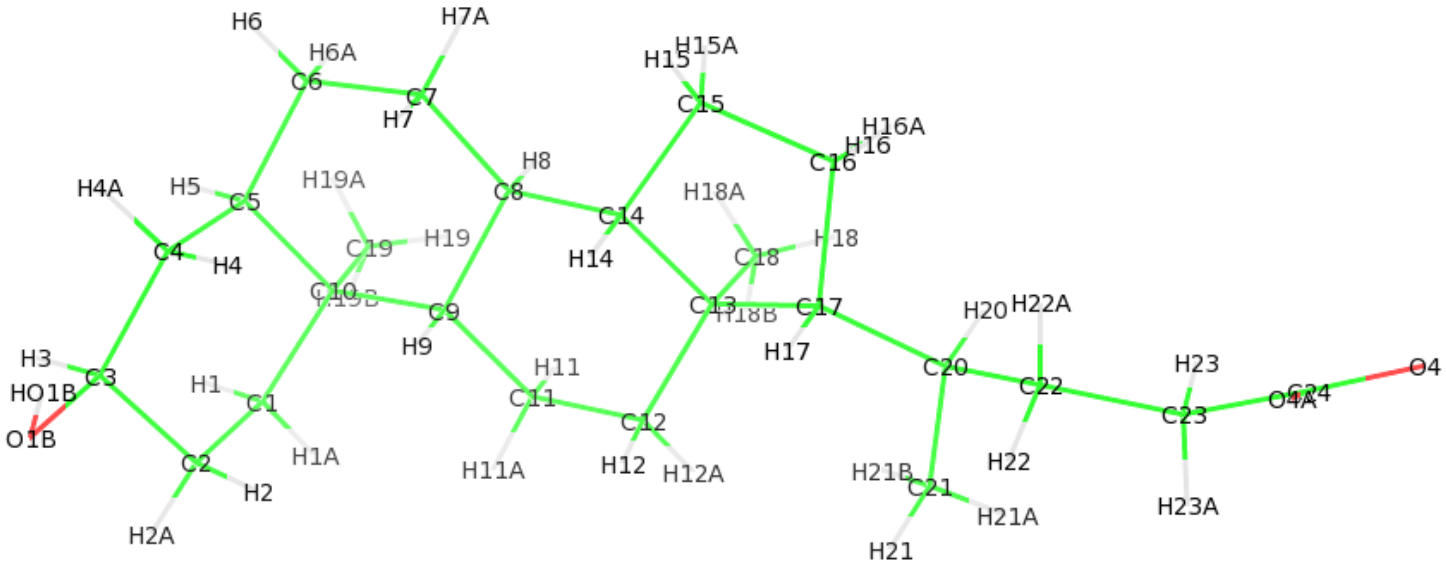
AOM (CSD)
Refined 1cyc



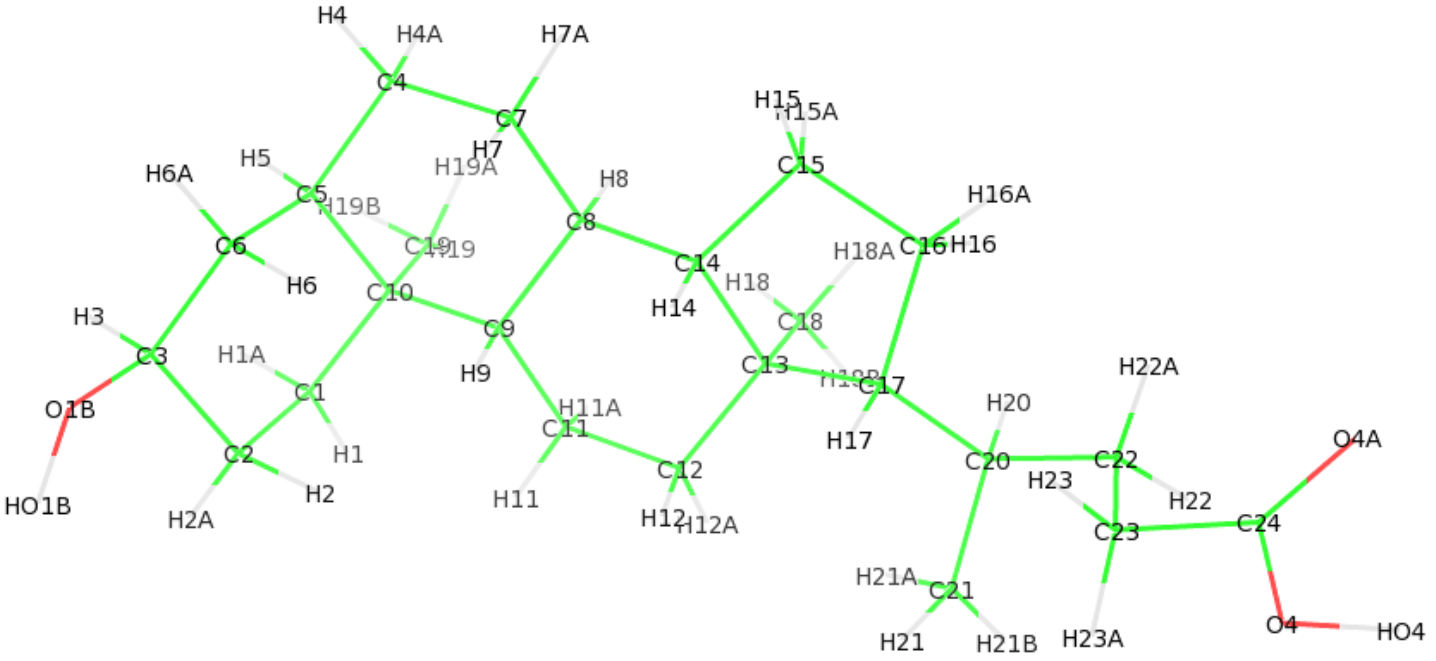
4OA: 4q0a


H6
H7A
H6A
C6
C7
H7
H15
H15A
C15
H16A
H16
C16
H8
C8
H4A
H5
C5
H19A
C14
H18A
C4
H4
C19
H19
H14
C18
H18
C10
C9
C13
C17
H20
H22A
H3
C3
H9
H11
H17
C20
C22
H23
HO1B
H1
C1
C11
C12
C24
O4
O4A
C23
O1B
H1A
C2
H11A
H12
H12A
H22
H21B
C21
H21A
H2
H23A
H2A
H21

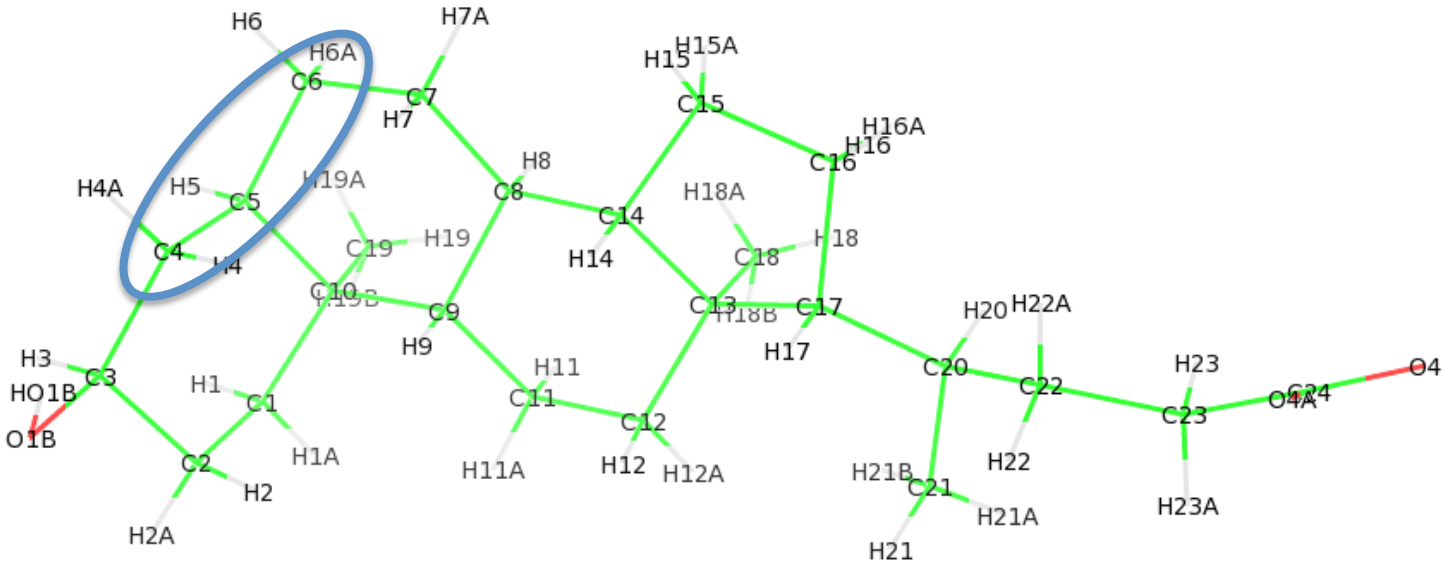
4OA: 4q0a



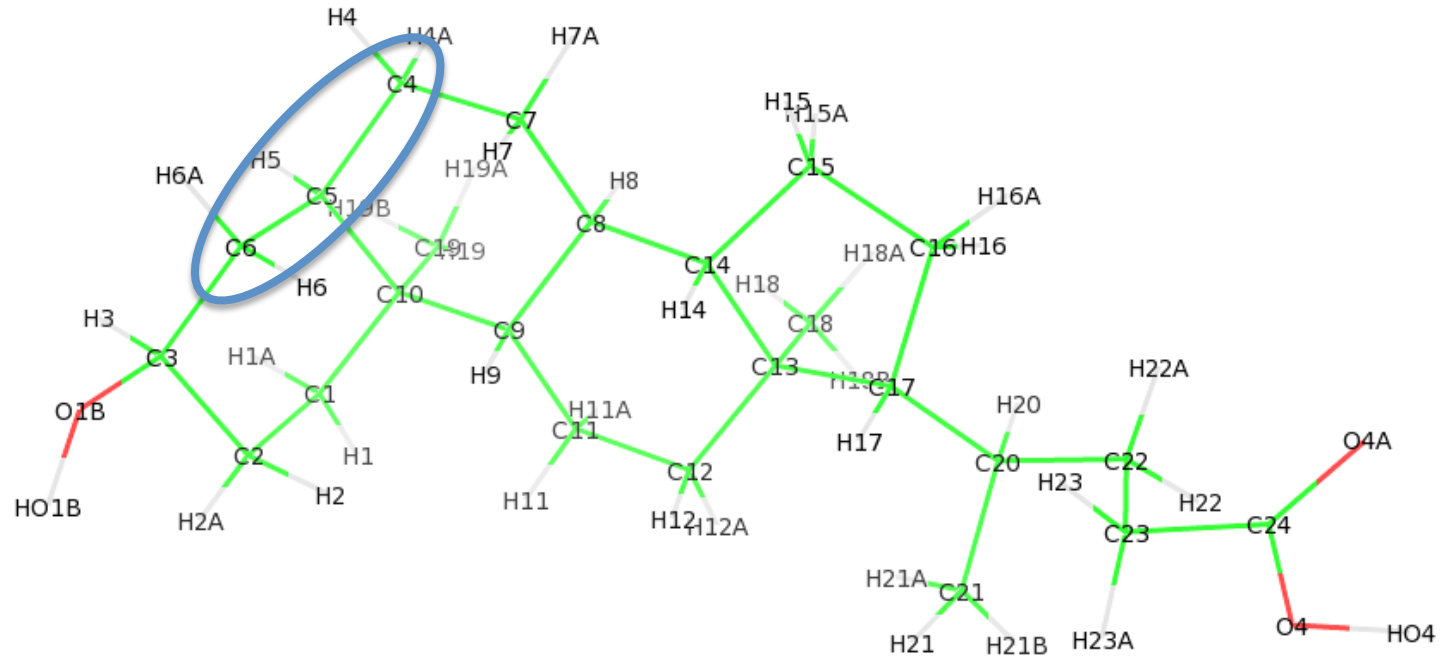
4OA (CSD)



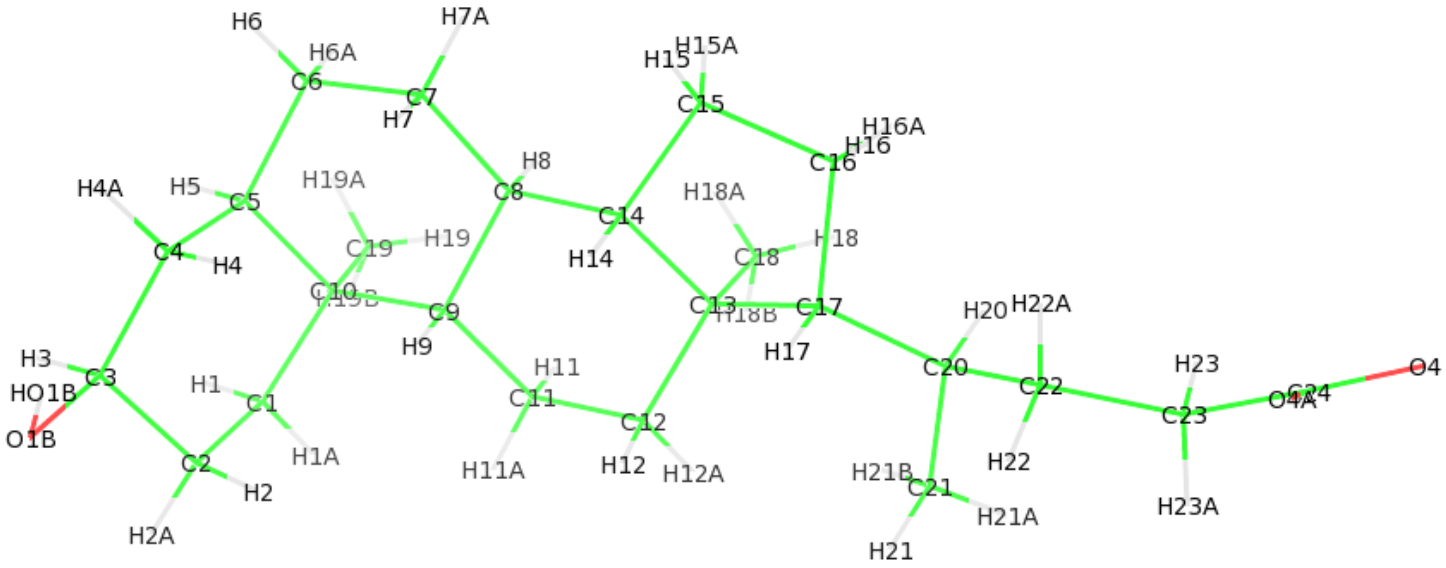
4OA: 4q0a



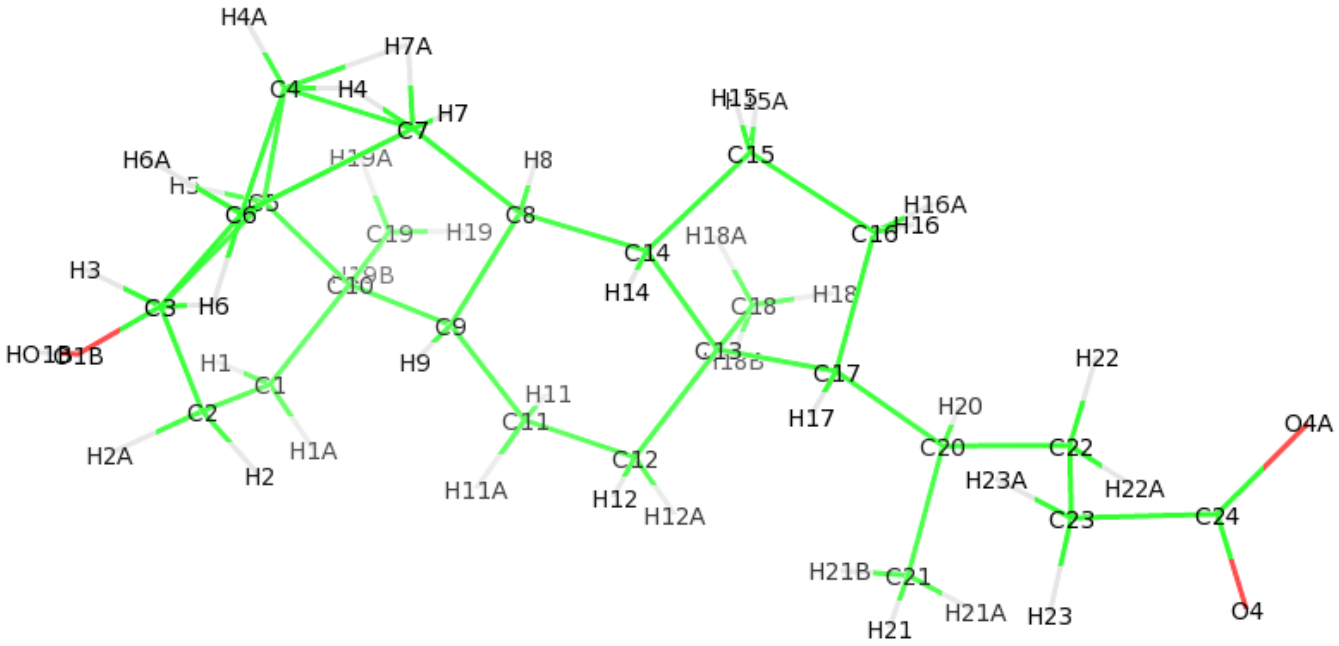
4OA (CSD)



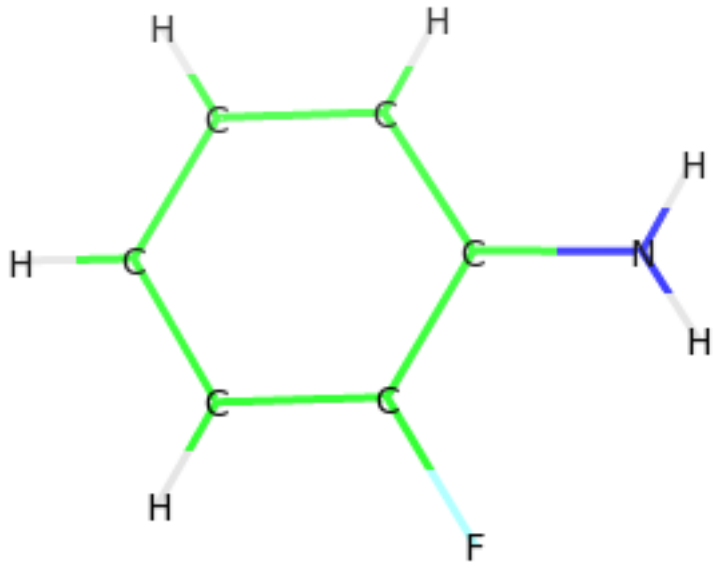
4OA: 4q0a



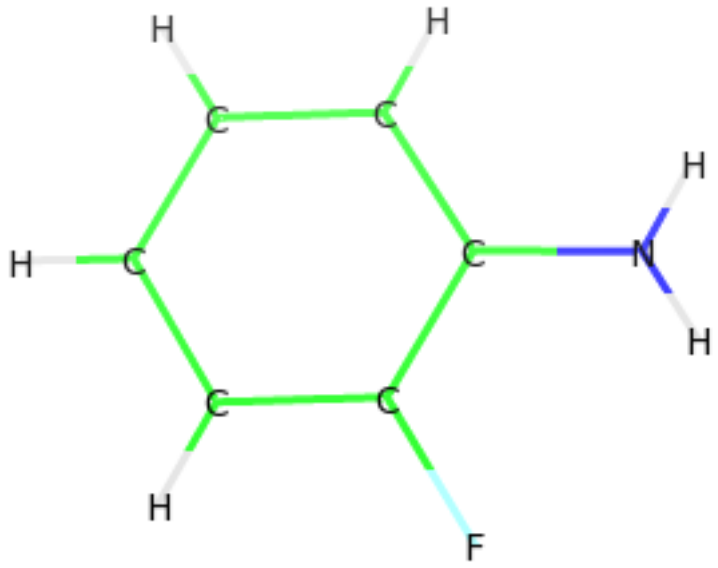
4OA (CSD)
Refined 1cyc



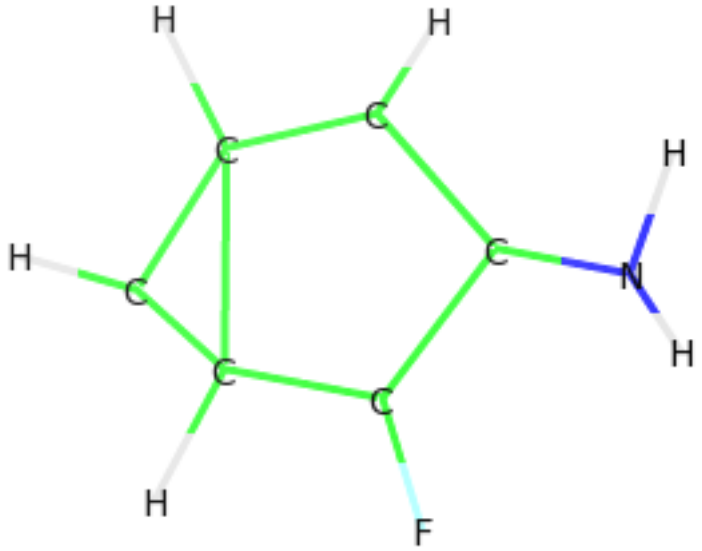
1AN: 1lgw



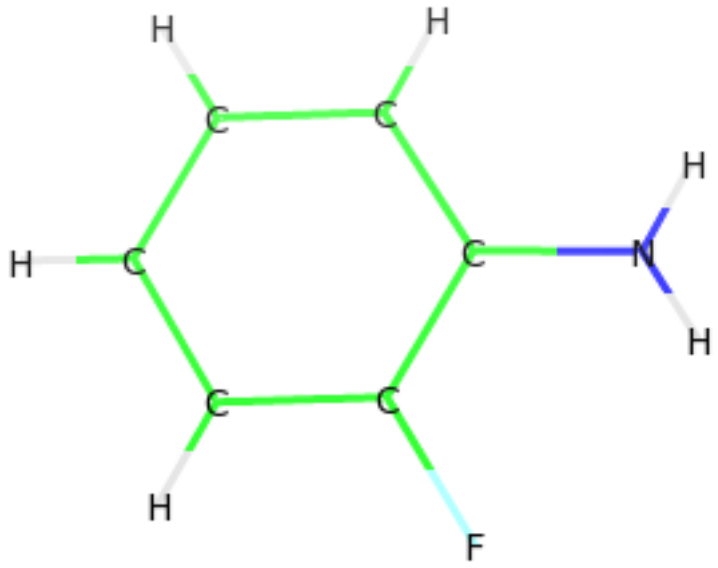
1AN: 1lgw



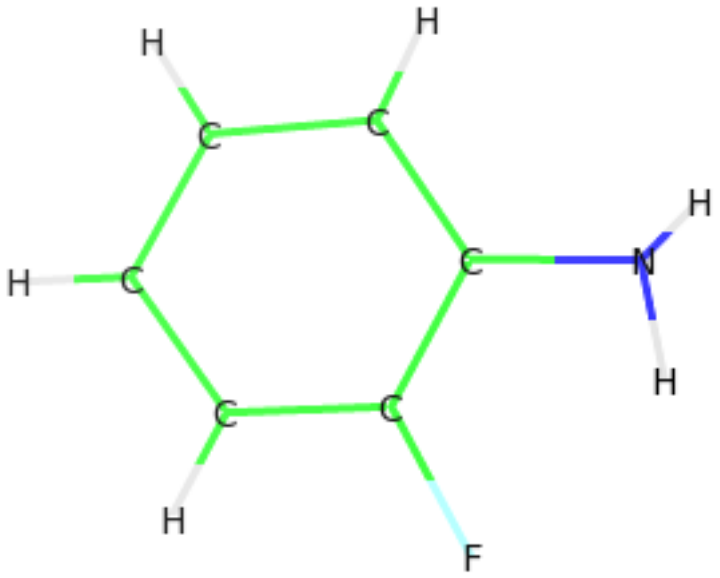
1AN (CSD)



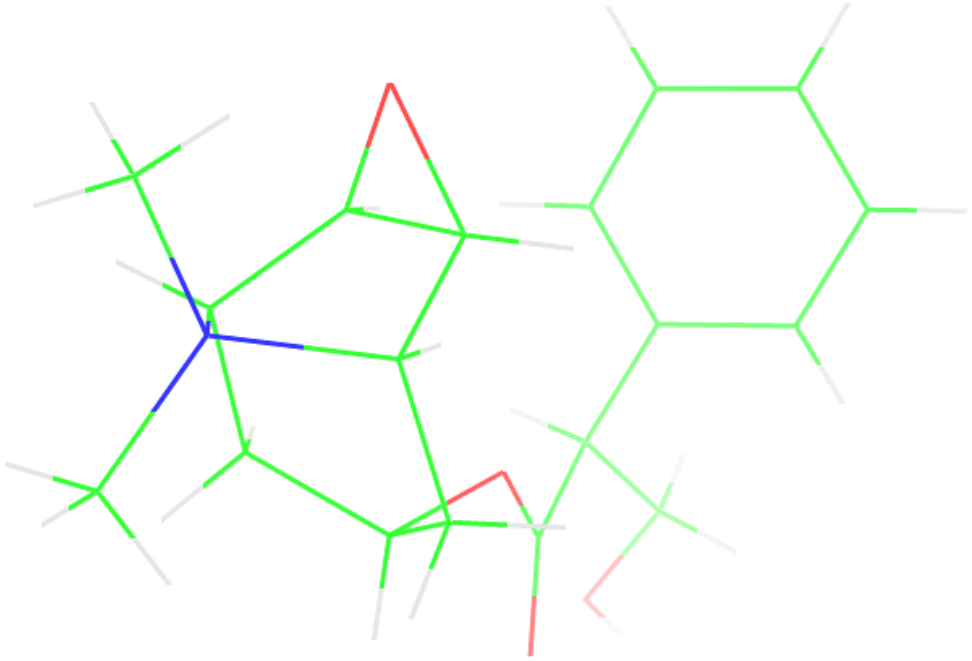
1AN: 1lgw



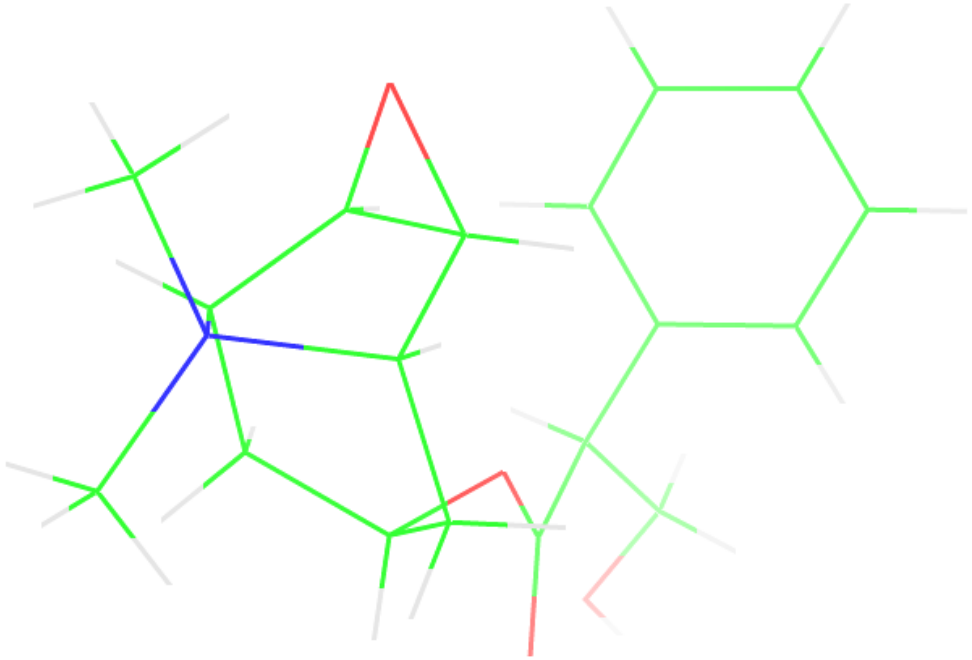
1AN (CSD)
Refined 1cyc



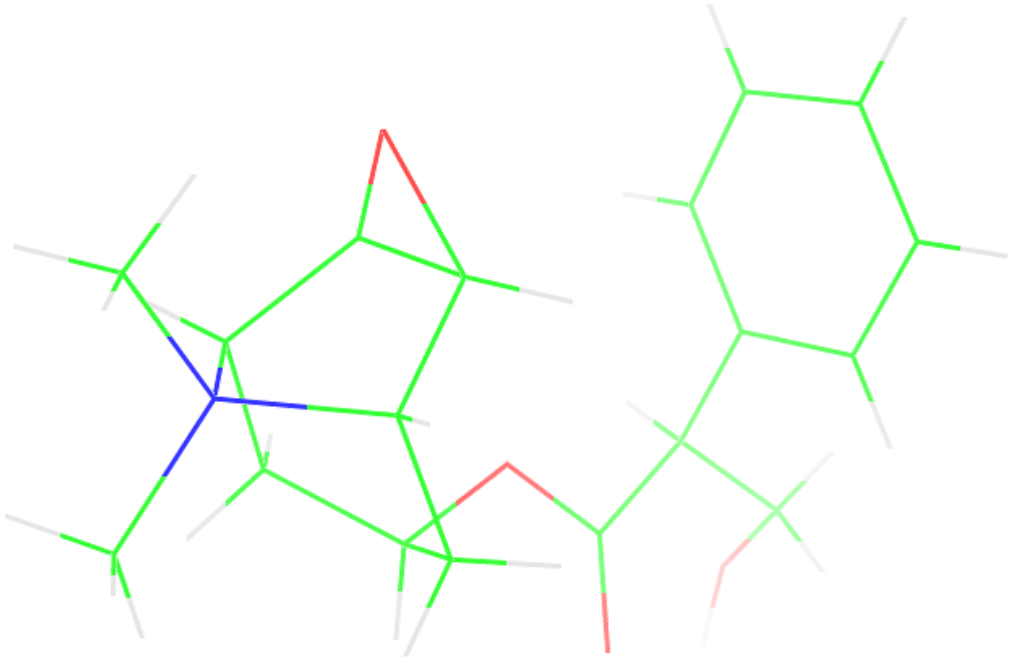
3C0: 4u16



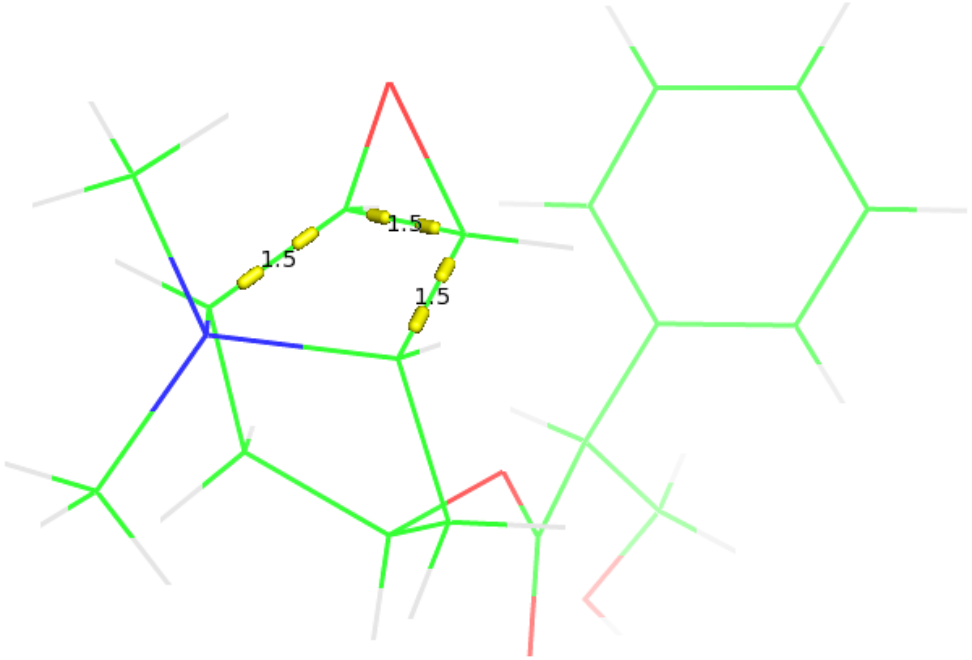
3C0: 4u16



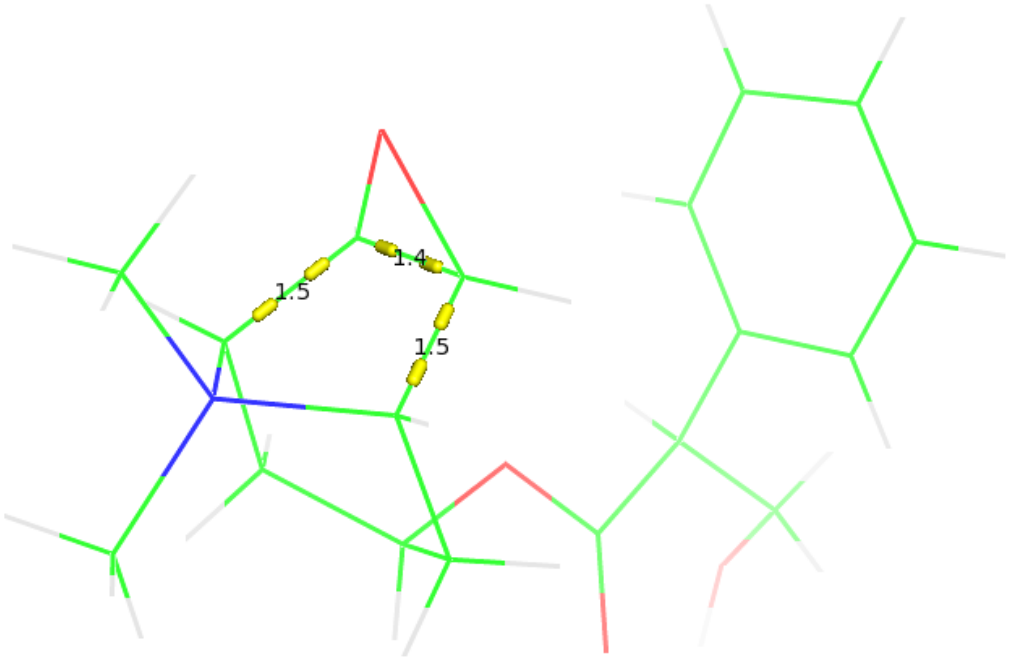
3C0 (CSD)



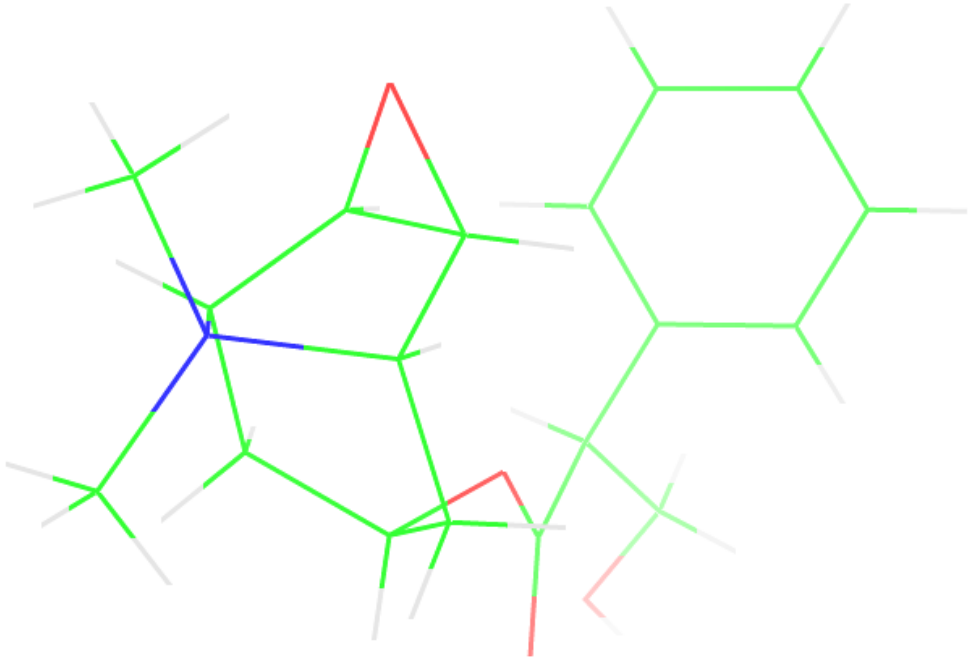
3C0: 4u16



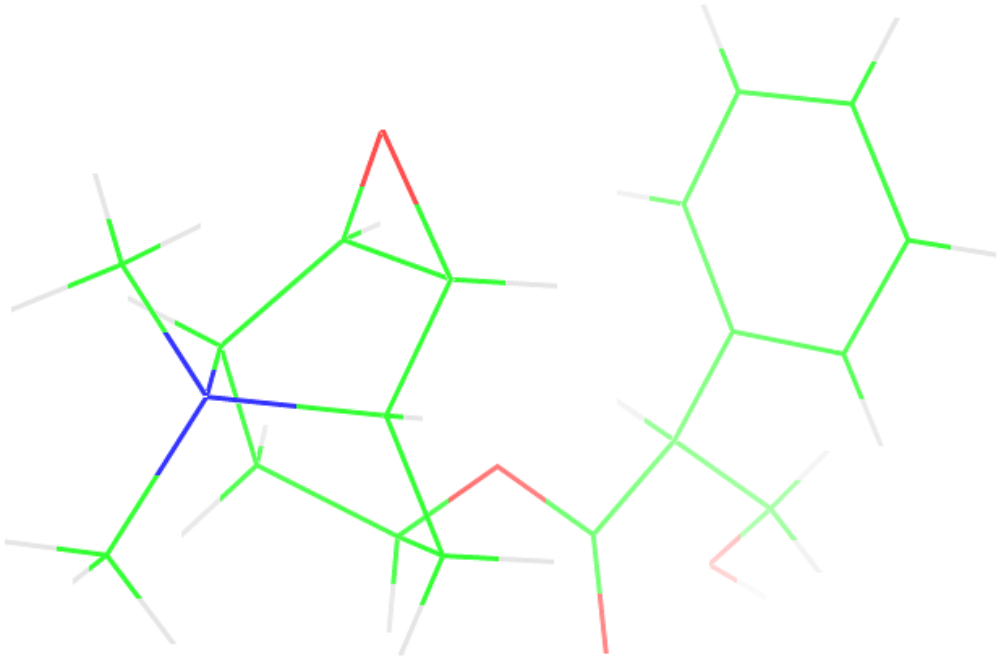
3C0 (CSD)



3C0: 4u16



3C0 (CSD)
Refined 1cyc



Summary

Notes:

- Ligand may be in a particular conformation due to environment – not scoring ideal conformations

- Riding hydrogens are added by REFMAC and used for scoring.
  - Made a difference for a few original PDB entries (3AQ in 4txs).
  - No difference in re-refined or CSD ligands.

Conclusions:

- There are problems in the CSD models.
  - Could investigate further, and report to CSD.
  - Note issues could be in intermediate steps – CCD, Coot, etc.

- Nomenclature inconsistencies – how do we deal with this?

- Nothing concerning identified that should make use worried about AceDRG.
  - No extreme outliers after re-refinement with AceDRG.
  - There are issues in the models deposited in the PDB (no surprise).