# WORLDWIDE PDB PROTEIN DATA BANK

# wwPDB X-ray Structure Validation Summary Report (i)

Sep 17, 2024 – 02:10 PM EDT

|  |  |  |
|---|---|---|
| PDB ID | : | 8TE4 |
| Title | : | Crystal structure of the methyltransferase domain of R882H/N879A DNMT3A homotetramer |
| Authors | : | Lu, J.W.; Song, J.K. |
| Deposited on | : | 2023-07-05 |
| Resolution | : | 2.65 Å(reported) |

The following versions of software and data (see references (i)) were used in the production of this report:

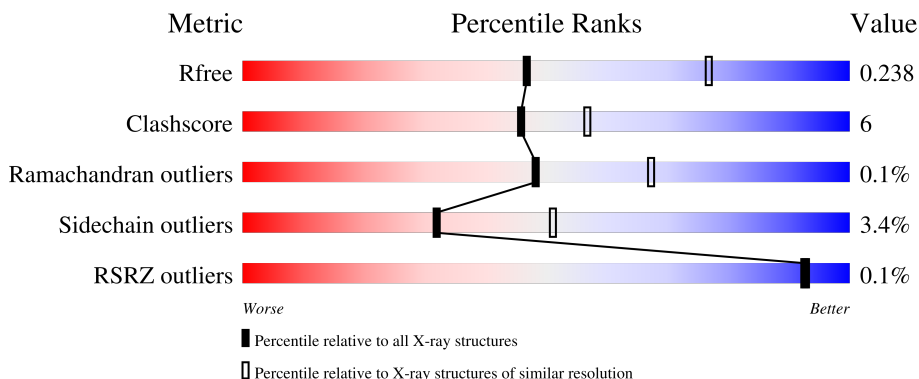|  |  |  |
|---|---|---|
| MolProbity | : | 4.02b-467 |
| Mogul | : | 2022.3.0, CSD as543be (2022) |
| Xtriage (Phenix) | : | 1.20.1 |
| EDS | : | 3.0 |
| buster-report | : | 1.1.7 (2018) |
| Percentile statistics | : | 20231227.v01 (using entries in the PDB archive December 27th 2023) |
| CCP4 | : | 9.0.002 (Gargrove) |
| Density-Fitness | : | 1.0.11 |
| Ideal geometry (proteins) | : | Engh & Huber (2001) |
| Ideal geometry (DNA, RNA) | : | Parkinson et al. (1996) |
| Validation Pipeline (wwPDB-VP) | : | 2.38.3 |

# 1　Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.
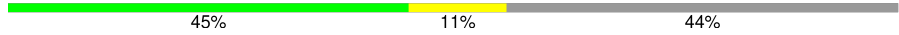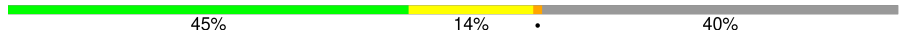
| Metric | Percentile Ranks | Value |
|--------|------------------|-------|
| Rfree | | 0.238 |
| Clashscore | | 6 |
| Ramachandran outliers | | 0.1% |
| Sidechain outliers | | 3.4% |
| RSRZ outliers | | 0.1% |

*Worse* ... *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

| Metric | Whole archive (#Entries) | Similar resolution (#Entries, resolution range(Å)) |
|--------|--------------------------|----------------------------------------------------|
| $R_{free}$ | 164625 | 1003 (2.66-2.66) |
| Clashscore | 180529 | 1063 (2.66-2.66) |
| Ramachandran outliers | 177936 | 1052 (2.66-2.66) |
| Sidechain outliers | 177891 | 1052 (2.66-2.66) |
| RSRZ outliers | 164620 | 1003 (2.66-2.66) |

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 1 | A | 287 | 81% 14% 5% |
| 1 | B | 287 | 83% 12% • |
| 1 | C | 287 | 84% 11% • • |
| 1 | D | 287 | 80% 14% • • |
| 1 | E | 287 | 46% 9% 45% |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Length | Quality of chain |
|-----|-------|--------|------------------|
| 1 | F | 287 | 45% 11% 44% |
| 1 | G | 287 | 48% 12% 40% |
| 1 | H | 287 | 45% 14% 40% |

# 2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 14869 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA (cytosine-5)-methyltransferase 3A.

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf | Trace |
|-----|-------|----------|-------|-----|-----|-----|-----|---------|---------|-------|
| 1 | A | 273 | Total | C | N | O | S | 0 | 3 | 0 |
| | | | 2199 | 1409 | 391 | 385 | 14 | | | |
| 1 | B | 275 | Total | C | N | O | S | 0 | 3 | 0 |
| | | | 2206 | 1414 | 391 | 387 | 14 | | | |
| 1 | C | 275 | Total | C | N | O | S | 0 | 3 | 0 |
| | | | 2205 | 1414 | 391 | 386 | 14 | | | |
| 1 | E | 158 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1254 | 821 | 217 | 211 | 5 | | | |
| 1 | H | 171 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1361 | 885 | 238 | 232 | 6 | | | |
| 1 | F | 161 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1265 | 829 | 215 | 215 | 6 | | | |
| 1 | D | 275 | Total | C | N | O | S | 0 | 4 | 0 |
| | | | 2217 | 1421 | 394 | 388 | 14 | | | |
| 1 | G | 173 | Total | C | N | O | S | 0 | 0 | 0 |
| | | | 1354 | 882 | 235 | 231 | 6 | | | |

There are 32 discrepancies between the modelled and reference sequences:

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| A | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| A | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| A | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| A | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| B | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| B | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| B | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| B | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| C | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| C | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| C | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| C | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| E | 626 | GLY | - | expression tag | UNP Q9Y6K1 |

*Continued on next page...*

*Continued from previous page...*

| Chain | Residue | Modelled | Actual | Comment | Reference |
|-------|---------|----------|--------|---------|-----------|
| E | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| E | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| E | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| H | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| H | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| H | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| H | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| F | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| F | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| F | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| F | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| D | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| D | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| D | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| D | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |
| G | 626 | GLY | - | expression tag | UNP Q9Y6K1 |
| G | 627 | SER | - | expression tag | UNP Q9Y6K1 |
| G | 879 | ALA | ASN | engineered mutation | UNP Q9Y6K1 |
| G | 882 | HIS | ARG | engineered mutation | UNP Q9Y6K1 |

- Molecule 2 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$) (labeled as "Ligand of Interest" by depositor).
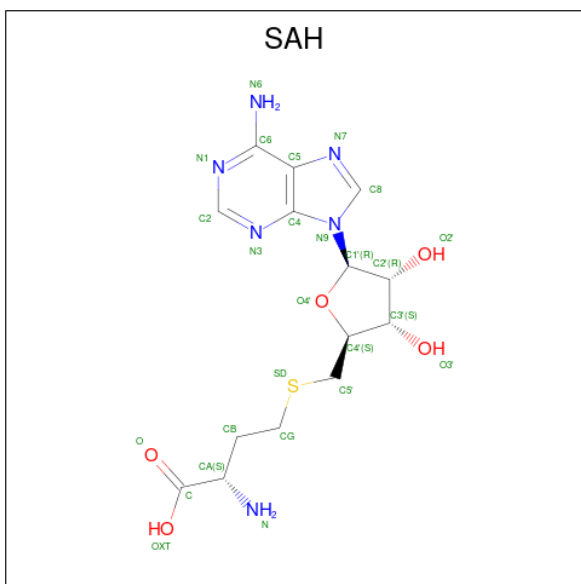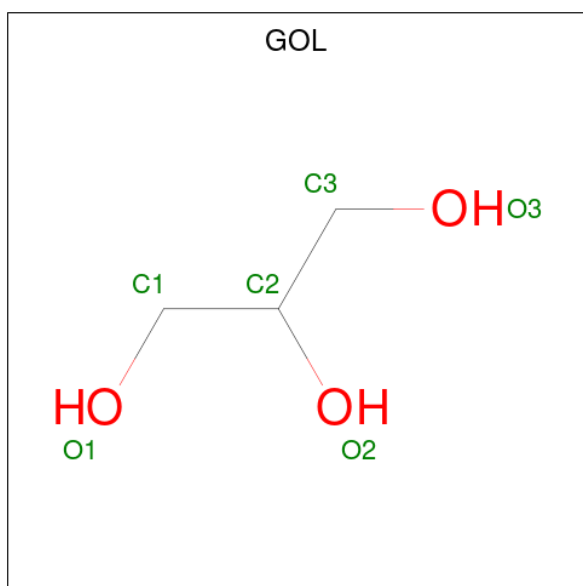


| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|-----|-------|----------|-------|---|---|---|---|---------|---------|
| | | | Total | C | N | O | S | | |
| 2 | A | 1 | 26 | 14 | 6 | 5 | 1 | 0 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Residues | Atoms | | | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|---|---|
| 2 | B | 1 | Total 26 | C 14 | N 6 | O 5 | S 1 | 0 | 0 |
| 2 | C | 1 | Total 26 | C 14 | N 6 | O 5 | S 1 | 0 | 0 |
| 2 | D | 1 | Total 26 | C 14 | N 6 | O 5 | S 1 | 0 | 0 |

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$) (labeled as "Ligand of Interest" by depositor).



| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| 3 | A | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | A | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | A | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | A | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | A | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | B | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | B | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | B | 1 | Total 6 | C 3 | O 3 | 0 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Residues | Atoms | | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|---|
| 3 | B | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | B | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | C | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | C | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | D | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | D | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | D | 1 | Total 6 | C 3 | O 3 | 0 | 0 |
| 3 | D | 1 | Total 6 | C 3 | O 3 | 0 | 0 |

- Molecule 4 is water.

| Mol | Chain | Residues | Atoms | | ZeroOcc | AltConf |
|---|---|---|---|---|---|---|
| 4 | A | 107 | Total 107 | O 107 | 0 | 0 |
| 4 | B | 124 | Total 124 | O 124 | 0 | 0 |
| 4 | C | 122 | Total 122 | O 122 | 0 | 0 |
| 4 | E | 36 | Total 36 | O 36 | 0 | 0 |
| 4 | H | 29 | Total 29 | O 29 | 0 | 0 |
| 4 | F | 34 | Total 34 | O 34 | 0 | 0 |
| 4 | D | 119 | Total 119 | O 119 | 0 | 0 |
| 4 | G | 37 | Total 37 | O 37 | 0 | 0 |

# 3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA (cytosine-5)-methyltransferase 3A
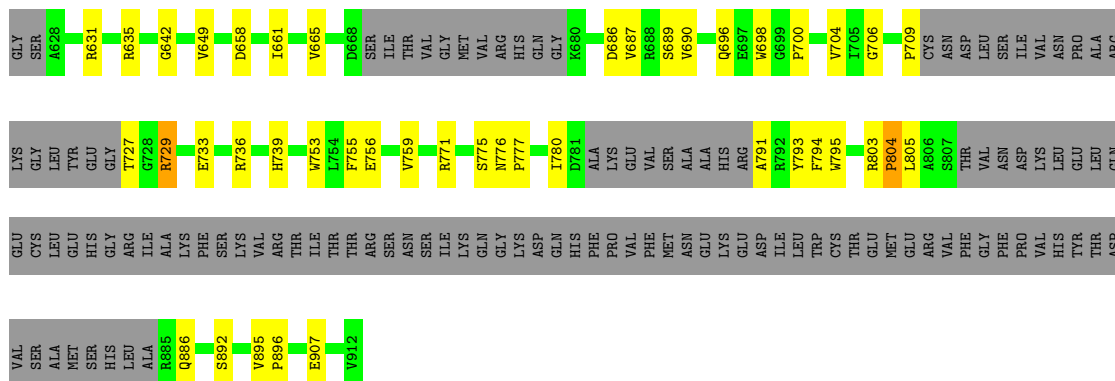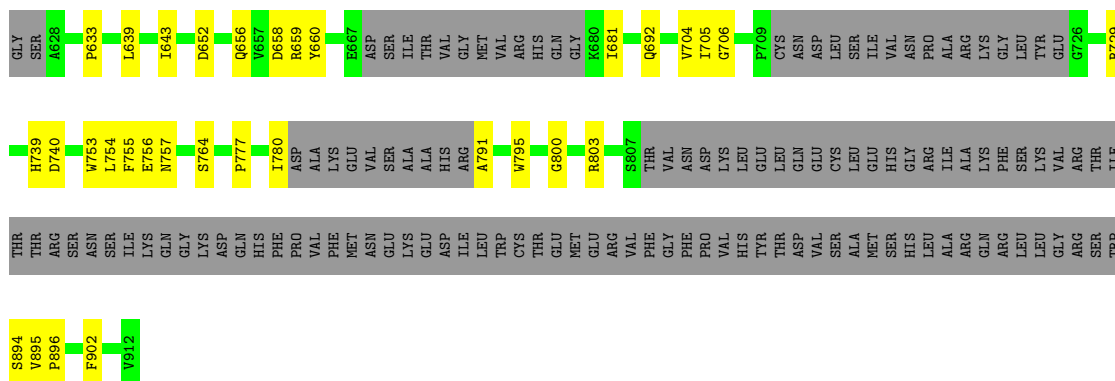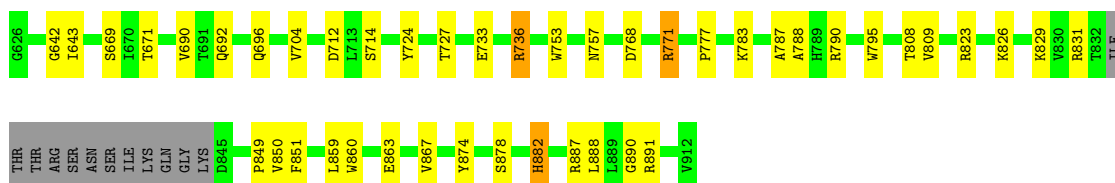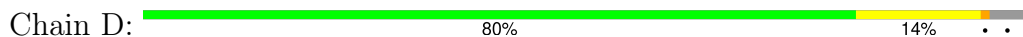
Chain A:



• Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain B:



• Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain C:



• Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain E:

R736 R742 E745 W753 L754 D768 R771 S775 N776 P777 I780 ASP ALA LYS GLU VAL SER ALA ALA HIS ARG A791 W795 L805 ALA SER THR VAL ASN ASP LYS LEU GLU LEU GLN GLU CYS LEU GLU HIS GLY ARG ILE ALA LYS PHE SER LYS VAL ARG THR

ILE THR THR ARG ASN SER SER ILE LYS GLN GLY LYS ASP LYS GLN HIS PHE PRO VAL PHE MET ASN GLU LYS GLU GLU ILE LEU TRP CYS THR GLU MET GLU ARG VAL PHE GLY PHE PRO VAL HIS TYR THR ASP VAL SER ALA MET SER HIS LEU ALA ARG ARG LEU LEU GLY ARG SER

W893 K906 E907 V912

● Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain H:

45%    14%    •    40%

GLY SER A628 R631 R635 G642 V649 D658 I661 V665 D668 SER ILE THR VAL GLY MET VAL ARG HIS GLN GLY K680 D686 V687 R688 S689 V690 Q696 E697 W698 G699 P700 V704 I705 G706 P709 CYS ASN ASP LEU SER ILE VAL ASN PRO ALA ARG

LYS GLY LEU TYR GLU GLY T727 G728 R729 E733 R736 H739 W753 L754 F755 E756 V759 R771 S775 N776 P777 I780 D781 ALA LYS GLU VAL SER ALA ALA HIS ARG A791 R792 Y793 F794 W795 R803 P804 L805 A806 S807 THR VAL ASN ASP LYS LEU GLU LEU GLN

GLU CYS LEU GLU HIS GLY ARG ILE ALA LYS PHE SER LYS VAL ARG THR ILE THR THR ARG SER ASN SER ILE LYS GLN GLY LYS ASP GLN HIS PHE PRO VAL PHE MET ASN GLU LYS GLU ASP ILE LEU TRP CYS THR GLU MET GLU ARG VAL PHE GLY PHE PRO VAL HIS TYR THR ASP

VAL SER ALA MET SER HIS LEU ALA R885 Q886 S892 V895 P896 E907 V912

● Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain F:

45%    11%    44%

GLY SER A628 P633 L639 I643 D652 Q656 V657 D658 R659 Y660 E667 ASP SER ILE THR VAL GLY MET VAL ARG HIS GLN GLY K680 I681 Q692 V704 I705 G706 P709 CYS ASN ASP LEU SER ILE VAL ASN PRO ALA ARG LYS GLY LEU TYR GLU G726 R729

H739 D740 W753 L754 F755 E756 N757 S764 P777 I780 ASP ALA LYS GLU VAL SER ALA ALA HIS ARG A791 W795 G800 R803 S807 THR VAL ASN ASP LYS LEU GLU LEU GLN GLU CYS LEU GLU HIS GLY ARG ILE ALA LYS PHE SER LYS VAL ARG THR ILE

THR THR ARG ASN SER ASN SER ILE LYS GLN GLY LYS ASP GLN HIS PHE PRO VAL PHE MET ASN GLU LYS GLU GLU ASP ILE LEU TRP CYS THR GLU MET GLU ARG VAL PHE GLY PHE PRO VAL HIS TYR THR ASP VAL SER ALA MET SER HIS LEU ALA ARG GLN LEU LEU GLY ARG SER TRP

S894 V895 P896 F902 V912

● Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain D:

80%    14%    •  •

G626 G642 I643 S669 I670 T671 V690 T691 Q692 Q696 V704 D712 L713 S714 Y724 T727 E733 R736 W753 N757 D768 R771 P777 K783 A787 A788 H789 R790 W795 T808 V809 R823 K826 K829 V830 R831 T832 ILE

THR THR ARG ASN SER ILE LYS GLN GLY LYS D845 P849 V850 F851 L859 W860 E863 V867 Y874 S878 H882 R887 L888 L889 G890 R891 V912

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

• Molecule 1: DNA (cytosine-5)-methyltransferase 3A

Chain G:

# 4   Data and refinement statistics   ⓘ

| Property | Value | Source |
|---|---|---|
| Space group | P 3 | Depositor |
| Cell constants<br>a, b, c, $\alpha$, $\beta$, $\gamma$ | 178.07Å    178.07Å    109.91Å<br>90.00°     90.00°     120.00° | Depositor |
| Resolution (Å) | 46.76   –   2.65<br>46.76   –   2.65 | Depositor<br>EDS |
| % Data completeness<br>(in resolution range) | 99.3 (46.76-2.65)<br>99.4 (46.76-2.65) | Depositor<br>EDS |
| $R_{merge}$ | 0.15 | Depositor |
| $R_{sym}$ | (Not available) | Depositor |
| $< I/\sigma(I) >$ [1] | 1.37 (at 2.65Å) | Xtriage |
| Refinement program | PHENIX 1.20.1_4487 | Depositor |
| R, $R_{free}$ | 0.203   ,   0.236<br>0.203   ,   0.238 | Depositor<br>DCC |
| $R_{free}$ test set | 110554 reflections (1.77%) | wwPDB-VP |
| Wilson B-factor (Å$^2$) | 53.1 | Xtriage |
| Anisotropy | 0.150 | Xtriage |
| Bulk solvent $k_{sol}$(e/Å$^3$), $B_{sol}$(Å$^2$) | 0.29 , 37.8 | EDS |
| L-test for twinning[2] | $< |L| > = 0.51, < L^2 > = 0.34$ | Xtriage |
| Estimated twinning fraction | 0.478 for -h,-k,l<br>0.479 for h,-h-k,-l<br>0.478 for -k,-h,-l | Xtriage |
| $F_o$,$F_c$ correlation | 0.96 | EDS |
| Total number of atoms | 14869 | wwPDB-VP |
| Average B, all atoms (Å$^2$) | 58.0 | wwPDB-VP |

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

# 5 Model quality ⓘ

## 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SAH, GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Chain | Bond lengths | | Bond angles | |
|-----|-------|------|--------|------|--------|
| | | RMSZ | #$|Z|$ >5 | RMSZ | #$|Z|$ >5 |
| 1 | A | 0.33 | 0/2262 | 0.53 | 0/3054 |
| 1 | B | 0.26 | 0/2269 | 0.52 | 0/3064 |
| 1 | C | 0.26 | 0/2268 | 0.52 | 0/3064 |
| 1 | D | 0.26 | 0/2283 | 0.53 | 0/3083 |
| 1 | E | 0.26 | 0/1287 | 0.50 | 0/1739 |
| 1 | F | 0.26 | 0/1296 | 0.53 | 0/1749 |
| 1 | G | 0.26 | 0/1386 | 0.54 | 0/1872 |
| 1 | H | 0.26 | 0/1394 | 0.53 | 0/1880 |
| All | All | 0.27 | 0/14445 | 0.52 | 0/19505 |

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|-----|-------|-------|----------|----------|---------|--------------|
| 1 | A | 2199 | 0 | 2170 | 23 | 0 |
| 1 | B | 2206 | 0 | 2172 | 26 | 0 |
| 1 | C | 2205 | 0 | 2168 | 24 | 0 |
| 1 | D | 2217 | 0 | 2185 | 29 | 0 |
| 1 | E | 1254 | 0 | 1215 | 15 | 0 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Non-H | H(model) | H(added) | Clashes | Symm-Clashes |
|-----|-------|-------|----------|----------|---------|--------------|
| 1 | F | 1265 | 0 | 1235 | 18 | 0 |
| 1 | G | 1354 | 0 | 1311 | 15 | 0 |
| 1 | H | 1361 | 0 | 1335 | 25 | 0 |
| 2 | A | 26 | 0 | 19 | 0 | 0 |
| 2 | B | 26 | 0 | 19 | 0 | 0 |
| 2 | C | 26 | 0 | 19 | 0 | 0 |
| 2 | D | 26 | 0 | 19 | 0 | 0 |
| 3 | A | 30 | 0 | 40 | 2 | 0 |
| 3 | B | 30 | 0 | 40 | 1 | 0 |
| 3 | C | 12 | 0 | 16 | 0 | 0 |
| 3 | D | 24 | 0 | 32 | 2 | 0 |
| 4 | A | 107 | 0 | 0 | 0 | 0 |
| 4 | B | 124 | 0 | 0 | 2 | 0 |
| 4 | C | 122 | 0 | 0 | 1 | 0 |
| 4 | D | 119 | 0 | 0 | 2 | 0 |
| 4 | E | 36 | 0 | 0 | 1 | 0 |
| 4 | F | 34 | 0 | 0 | 0 | 0 |
| 4 | G | 37 | 0 | 0 | 0 | 0 |
| 4 | H | 29 | 0 | 0 | 1 | 0 |
| All | All | 14869 | 0 | 13995 | 160 | 0 |

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

The worst 5 of 160 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

| Atom-1 | Atom-2 | Interatomic distance (Å) | Clash overlap (Å) |
|--------|--------|--------------------------|-------------------|
| 1:A:849:PRO:HD2 | 1:A:858:ILE:HA | 1.61 | 0.82 |
| 1:D:643:ILE:HD11 | 1:D:888:LEU:HD13 | 1.69 | 0.73 |
| 1:B:849:PRO:HG2 | 1:B:859:LEU:HD23 | 1.71 | 0.72 |
| 1:H:709:PRO:HG3 | 1:H:892:SER:HB3 | 1.72 | 0.72 |
| 1:H:780:ILE:O | 1:H:791:ALA:HA | 1.90 | 0.71 |

There are no symmetry-related clashes.

## 5.3 Torsion angles (i)

### 5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

| Mol | Chain | Analysed | Favoured | Allowed | Outliers | Percentiles | |
|-----|-------|----------|----------|---------|----------|-----|-----|
| 1 | A | 272/287 (95%) | 265 (97%) | 7 (3%) | 0 | 100 | 100 |
| 1 | B | 274/287 (96%) | 267 (97%) | 7 (3%) | 0 | 100 | 100 |
| 1 | C | 274/287 (96%) | 266 (97%) | 8 (3%) | 0 | 100 | 100 |
| 1 | D | 275/287 (96%) | 268 (98%) | 7 (2%) | 0 | 100 | 100 |
| 1 | E | 148/287 (52%) | 142 (96%) | 6 (4%) | 0 | 100 | 100 |
| 1 | F | 151/287 (53%) | 147 (97%) | 4 (3%) | 0 | 100 | 100 |
| 1 | G | 163/287 (57%) | 156 (96%) | 7 (4%) | 0 | 100 | 100 |
| 1 | H | 161/287 (56%) | 152 (94%) | 7 (4%) | 2 (1%) | 11 | 18 |
| All | All | 1718/2296 (75%) | 1663 (97%) | 53 (3%) | 2 (0%) | 48 | 67 |

All (2) Ramachandran outliers are listed below:

| Mol | Chain | Res | Type |
|-----|-------|-----|------|
| 1 | H | 804 | PRO |
| 1 | H | 642 | GLY |

### 5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles | |
|-----|-------|----------|-----------|----------|-----|-----|
| 1 | A | 232/250 (93%) | 226 (97%) | 6 (3%) | 41 | 63 |
| 1 | B | 232/250 (93%) | 228 (98%) | 4 (2%) | 56 | 75 |
| 1 | C | 231/250 (92%) | 227 (98%) | 4 (2%) | 56 | 75 |

*Continued on next page...*

*Continued from previous page...*

| Mol | Chain | Analysed | Rotameric | Outliers | Percentiles | |
|---|---|---|---|---|---|---|
| 1 | D | 233/250 (93%) | 227 (97%) | 6 (3%) | 41 | 63 |
| 1 | E | 126/250 (50%) | 121 (96%) | 5 (4%) | 27 | 44 |
| 1 | F | 127/250 (51%) | 121 (95%) | 6 (5%) | 22 | 37 |
| 1 | G | 136/250 (54%) | 124 (91%) | 12 (9%) | 8 | 12 |
| 1 | H | 139/250 (56%) | 133 (96%) | 6 (4%) | 25 | 41 |
| All | All | 1456/2000 (73%) | 1407 (97%) | 49 (3%) | 32 | 51 |

5 of 49 residues with a non-rotameric sidechain are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | F | 739 | HIS |
| 1 | D | 771 | ARG |
| 1 | F | 764 | SER |
| 1 | D | 714 | SER |
| 1 | G | 652 | ASP |

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

| Mol | Chain | Res | Type |
|---|---|---|---|
| 1 | A | 789 | HIS |
| 1 | B | 656 | GLN |
| 1 | D | 656 | GLN |

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry (i)

20 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

| Mol | Type | Chain | Res | Link | Bond lengths | | | Bond angles | | |
|-----|------|-------|-----|------|--------|------|---------|--------|------|---------|
| | | | | | Counts | RMSZ | $\#|Z| > 2$ | Counts | RMSZ | $\#|Z| > 2$ |
| 2 | SAH | A | 1001 | - | 23,28,28 | 1.26 | 3 (13%) | 22,40,40 | 1.77 | 3 (13%) |
| 3 | GOL | B | 1006 | - | 5,5,5 | 0.96 | 0 | 5,5,5 | 0.98 | 0 |
| 3 | GOL | D | 1003 | - | 5,5,5 | 0.91 | 0 | 5,5,5 | 1.04 | 0 |
| 2 | SAH | D | 1001 | - | 23,28,28 | 1.25 | 3 (13%) | 22,40,40 | 1.80 | 4 (18%) |
| 3 | GOL | A | 1002 | - | 5,5,5 | 0.93 | 0 | 5,5,5 | 1.04 | 0 |
| 3 | GOL | A | 1005 | - | 5,5,5 | 0.94 | 0 | 5,5,5 | 1.03 | 0 |
| 3 | GOL | D | 1005 | - | 5,5,5 | 0.93 | 0 | 5,5,5 | 1.11 | 0 |
| 3 | GOL | A | 1006 | - | 5,5,5 | 1.00 | 0 | 5,5,5 | 1.04 | 0 |
| 3 | GOL | C | 1003 | - | 5,5,5 | 0.93 | 0 | 5,5,5 | 1.10 | 0 |
| 3 | GOL | D | 1004 | - | 5,5,5 | 0.94 | 0 | 5,5,5 | 1.05 | 0 |
| 3 | GOL | A | 1004 | - | 5,5,5 | 0.96 | 0 | 5,5,5 | 1.03 | 0 |
| 3 | GOL | B | 1002 | - | 5,5,5 | 0.96 | 0 | 5,5,5 | 1.02 | 0 |
| 2 | SAH | C | 1001 | - | 23,28,28 | 1.25 | 3 (13%) | 22,40,40 | 1.77 | 3 (13%) |
| 3 | GOL | B | 1005 | - | 5,5,5 | 0.95 | 0 | 5,5,5 | 1.04 | 0 |
| 3 | GOL | B | 1004 | - | 5,5,5 | 0.96 | 0 | 5,5,5 | 1.02 | 0 |
| 3 | GOL | A | 1003 | - | 5,5,5 | 0.99 | 0 | 5,5,5 | 0.95 | 0 |
| 3 | GOL | B | 1003 | - | 5,5,5 | 0.93 | 0 | 5,5,5 | 1.08 | 0 |
| 3 | GOL | C | 1002 | - | 5,5,5 | 0.95 | 0 | 5,5,5 | 1.07 | 0 |
| 2 | SAH | B | 1001 | - | 23,28,28 | 1.26 | 3 (13%) | 22,40,40 | 1.81 | 4 (18%) |
| 3 | GOL | D | 1002 | - | 5,5,5 | 0.95 | 0 | 5,5,5 | 1.07 | 0 |

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|-----|------|-------|-----|------|---------|----------|-------|
| 2 | SAH | A | 1001 | - | - | 2/11/31/31 | 0/3/3/3 |
| 3 | GOL | B | 1006 | - | - | 0/4/4/4 | - |
| 3 | GOL | D | 1003 | - | - | 2/4/4/4 | - |

*Continued on next page...*

*Continued from previous page...*

| Mol | Type | Chain | Res | Link | Chirals | Torsions | Rings |
|-----|------|-------|-----|------|---------|----------|-------|
| 2 | SAH | D | 1001 | - | - | 3/11/31/31 | 0/3/3/3 |
| 3 | GOL | A | 1002 | - | - | 2/4/4/4 | - |
| 3 | GOL | A | 1005 | - | - | 0/4/4/4 | - |
| 3 | GOL | D | 1005 | - | - | 2/4/4/4 | - |
| 3 | GOL | A | 1006 | - | - | 2/4/4/4 | - |
| 3 | GOL | C | 1003 | - | - | 0/4/4/4 | - |
| 3 | GOL | D | 1004 | - | - | 2/4/4/4 | - |
| 3 | GOL | A | 1004 | - | - | 2/4/4/4 | - |
| 3 | GOL | B | 1002 | - | - | 0/4/4/4 | - |
| 2 | SAH | C | 1001 | - | - | 2/11/31/31 | 0/3/3/3 |
| 3 | GOL | B | 1005 | - | - | 2/4/4/4 | - |
| 3 | GOL | B | 1004 | - | - | 0/4/4/4 | - |
| 3 | GOL | A | 1003 | - | - | 2/4/4/4 | - |
| 3 | GOL | B | 1003 | - | - | 3/4/4/4 | - |
| 3 | GOL | C | 1002 | - | - | 0/4/4/4 | - |
| 2 | SAH | B | 1001 | - | - | 2/11/31/31 | 0/3/3/3 |
| 3 | GOL | D | 1002 | - | - | 3/4/4/4 | - |

The worst 5 of 12 bond length outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(Å) | Ideal(Å) |
|-----|-------|-----|------|-------|---|-------------|----------|
| 2 | A | 1001 | SAH | C2-N3 | 4.17 | 1.38 | 1.32 |
| 2 | B | 1001 | SAH | C2-N3 | 4.16 | 1.38 | 1.32 |
| 2 | C | 1001 | SAH | C2-N3 | 4.13 | 1.38 | 1.32 |
| 2 | D | 1001 | SAH | C2-N3 | 4.13 | 1.38 | 1.32 |
| 2 | A | 1001 | SAH | C2-N1 | 2.51 | 1.38 | 1.33 |

The worst 5 of 14 bond angle outliers are listed below:

| Mol | Chain | Res | Type | Atoms | Z | Observed(°) | Ideal(°) |
|-----|-------|-----|------|-------|---|-------------|----------|
| 2 | D | 1001 | SAH | N3-C2-N1 | -6.40 | 119.99 | 128.67 |
| 2 | B | 1001 | SAH | N3-C2-N1 | -6.39 | 120.00 | 128.67 |
| 2 | A | 1001 | SAH | N3-C2-N1 | -6.38 | 120.01 | 128.67 |
| 2 | C | 1001 | SAH | N3-C2-N1 | -6.31 | 120.11 | 128.67 |
| 2 | C | 1001 | SAH | C5'-SD-CG | -2.89 | 93.68 | 102.26 |

There are no chirality outliers.

5 of 31 torsion outliers are listed below:

| Mol | Chain | Res | Type | Atoms |
|-----|-------|------|------|-------------|
| 2 | A | 1001 | SAH | N-CA-CB-CG |
| 2 | B | 1001 | SAH | N-CA-CB-CG |
| 2 | C | 1001 | SAH | N-CA-CB-CG |
| 2 | D | 1001 | SAH | N-CA-CB-CG |
| 3 | A | 1003 | GOL | O1-C1-C2-O2 |

There are no ring outliers.

3 monomers are involved in 5 short contacts:

| Mol | Chain | Res | Type | Clashes | Symm-Clashes |
|-----|-------|------|------|---------|--------------|
| 3 | B | 1006 | GOL | 1 | 0 |
| 3 | D | 1005 | GOL | 2 | 0 |
| 3 | A | 1004 | GOL | 2 | 0 |

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

**Ligand SAH A 1001**



Bond lengths



Bond angles



Torsions



Rings

**Ligand GOL B 1006**



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL D 1003



Bond lengths



Bond angles



Torsions



Rings

## Ligand SAH D 1001



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL A 1002



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL A 1005



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL D 1005



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL A 1006



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL C 1003



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL D 1004



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL A 1004



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL B 1002



Bond lengths



Bond angles



Torsions



Rings

## Ligand SAH C 1001



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL B 1005



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL B 1004



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL A 1003



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL B 1003



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL C 1002



Bond lengths



Bond angles



Torsions



Rings

## Ligand SAH B 1001



Bond lengths



Bond angles



Torsions



Rings

## Ligand GOL D 1002



Bond lengths



Bond angles



Torsions



Rings

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

# 6 Fit of model and data ⓘ

## 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, $95^{th}$ percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

| Mol | Chain | Analysed | <RSRZ> | #RSRZ>2 | | | OWAB(Å$^2$) | Q<0.9 |
|-----|-------|----------|--------|---------|-----|-----|-------------|-------|
| 1 | A | 273/287 (95%) | -1.63 | 0 | 100 | 100 | 26, 48, 82, 108 | 3 (1%) |
| 1 | B | 275/287 (95%) | -1.66 | 0 | 100 | 100 | 25, 46, 83, 125 | 3 (1%) |
| 1 | C | 275/287 (95%) | -1.63 | 0 | 100 | 100 | 26, 47, 86, 122 | 3 (1%) |
| 1 | D | 275/287 (95%) | -1.61 | 0 | 100 | 100 | 25, 48, 82, 114 | 4 (1%) |
| 1 | E | 158/287 (55%) | -1.45 | 1 (0%) | 85 | 83 | 41, 63, 106, 133 | 0 |
| 1 | F | 161/287 (56%) | -1.48 | 0 | 100 | 100 | 45, 65, 102, 130 | 0 |
| 1 | G | 173/287 (60%) | -1.44 | 0 | 100 | 100 | 45, 67, 138, 167 | 0 |
| 1 | H | 171/287 (59%) | -1.46 | 0 | 100 | 100 | 41, 63, 129, 166 | 0 |
| All | All | 1761/2296 (76%) | -1.57 | 1 (0%) | 92 | 93 | 25, 53, 101, 167 | 13 (0%) |

All (1) RSRZ outliers are listed below:

| Mol | Chain | Res | Type | RSRZ |
|-----|-------|-----|------|------|
| 1 | E | 628 | ALA | 2.4 |

## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, $95^{th}$ percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

| Mol | Type | Chain | Res | Atoms | RSCC | RSR | B-factors($\text{Å}^2$) | Q<0.9 |
|---|---|---|---|---|---|---|---|---|
| 3 | GOL | D | 1003 | 6/6 | 0.97 | 0.13 | 43,82,95,99 | 0 |
| 3 | GOL | A | 1004 | 6/6 | 0.98 | 0.05 | 80,86,90,97 | 0 |
| 3 | GOL | A | 1005 | 6/6 | 0.98 | 0.07 | 56,93,97,105 | 0 |
| 3 | GOL | B | 1002 | 6/6 | 0.98 | 0.11 | 56,79,90,109 | 0 |
| 3 | GOL | B | 1005 | 6/6 | 0.98 | 0.08 | 60,79,89,91 | 0 |
| 3 | GOL | B | 1006 | 6/6 | 0.98 | 0.06 | 69,93,99,100 | 0 |
| 3 | GOL | C | 1002 | 6/6 | 0.98 | 0.07 | 72,85,96,104 | 0 |
| 3 | GOL | A | 1002 | 6/6 | 0.98 | 0.09 | 60,75,100,102 | 0 |
| 3 | GOL | D | 1004 | 6/6 | 0.98 | 0.09 | 44,80,88,93 | 0 |
| 2 | SAH | D | 1001 | 26/26 | 0.99 | 0.03 | 42,50,61,63 | 0 |
| 3 | GOL | B | 1003 | 6/6 | 0.99 | 0.05 | 64,71,79,89 | 0 |
| 3 | GOL | B | 1004 | 6/6 | 0.99 | 0.04 | 91,94,98,109 | 0 |
| 2 | SAH | A | 1001 | 26/26 | 0.99 | 0.03 | 42,51,61,65 | 0 |
| 3 | GOL | A | 1003 | 6/6 | 0.99 | 0.11 | 42,63,87,90 | 0 |
| 2 | SAH | B | 1001 | 26/26 | 0.99 | 0.03 | 40,48,57,61 | 0 |
| 3 | GOL | C | 1003 | 6/6 | 0.99 | 0.05 | 73,88,102,105 | 0 |
| 3 | GOL | D | 1002 | 6/6 | 0.99 | 0.05 | 68,78,98,111 | 0 |
| 2 | SAH | C | 1001 | 26/26 | 0.99 | 0.03 | 40,48,58,73 | 0 |
| 3 | GOL | A | 1006 | 6/6 | 0.99 | 0.08 | 46,72,81,97 | 0 |
| 3 | GOL | D | 1005 | 6/6 | 0.99 | 0.04 | 78,88,105,120 | 0 |

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around GOL D 1003:**

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
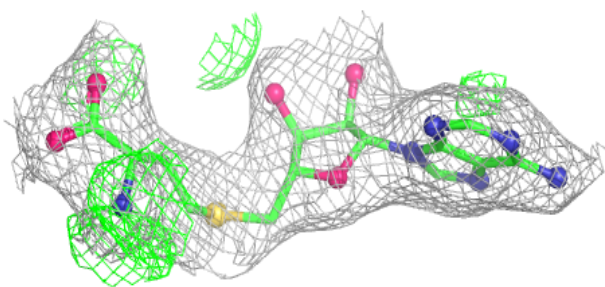mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)
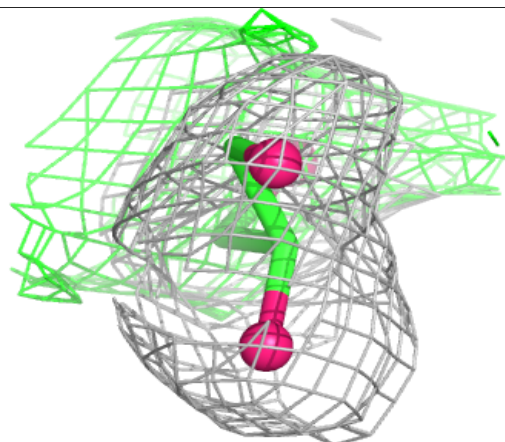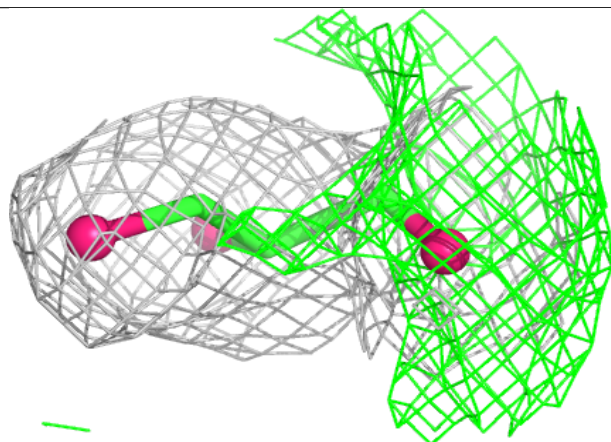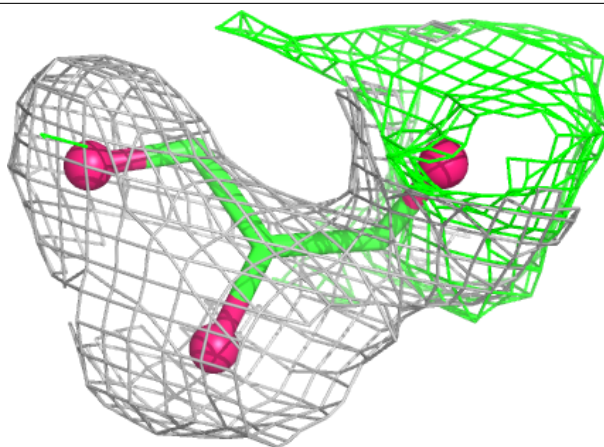
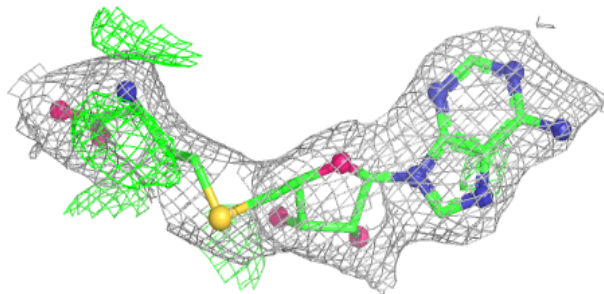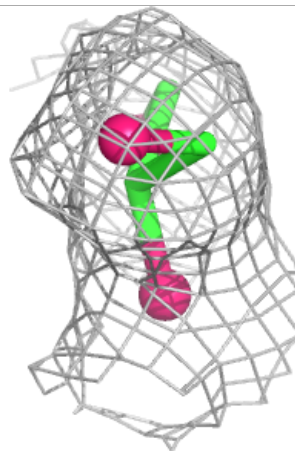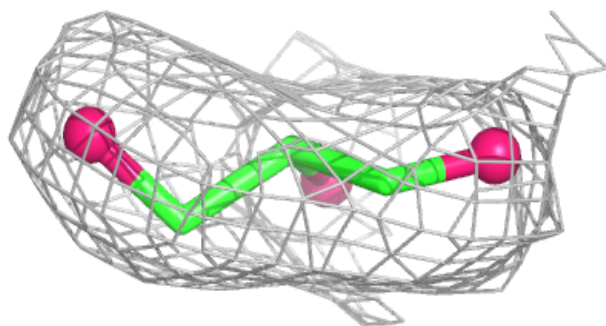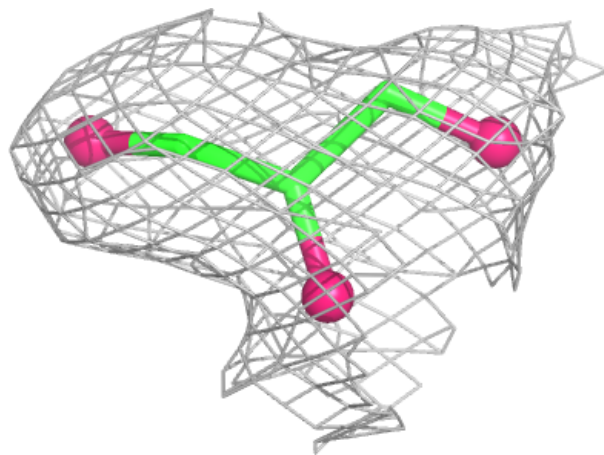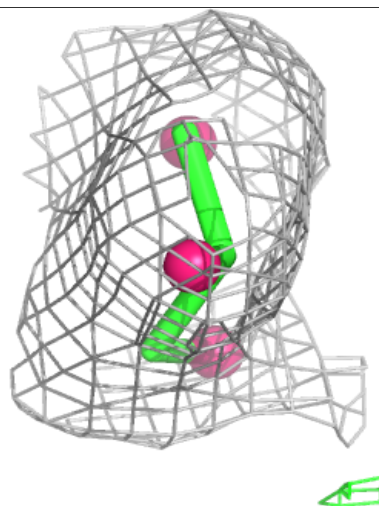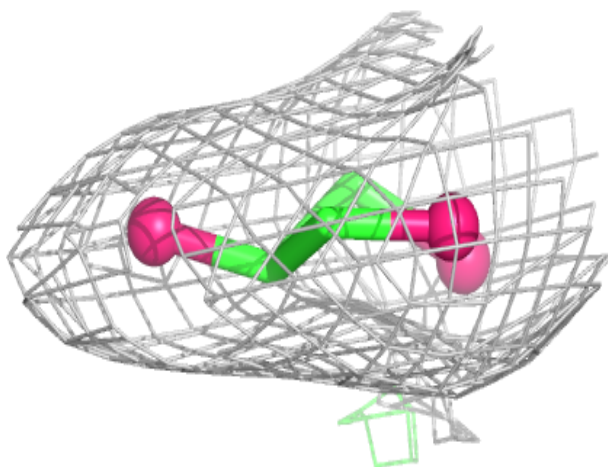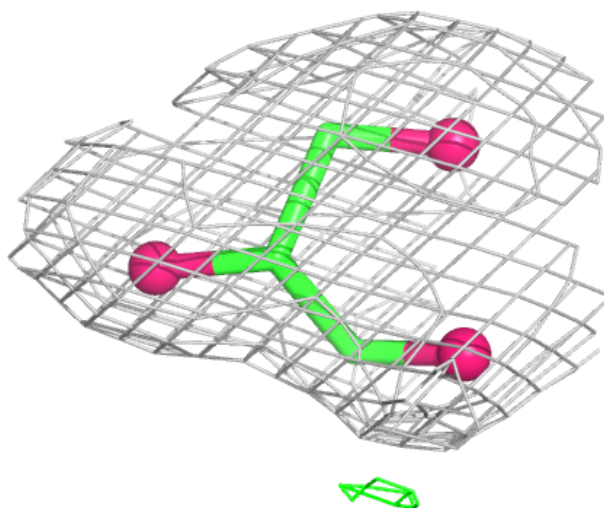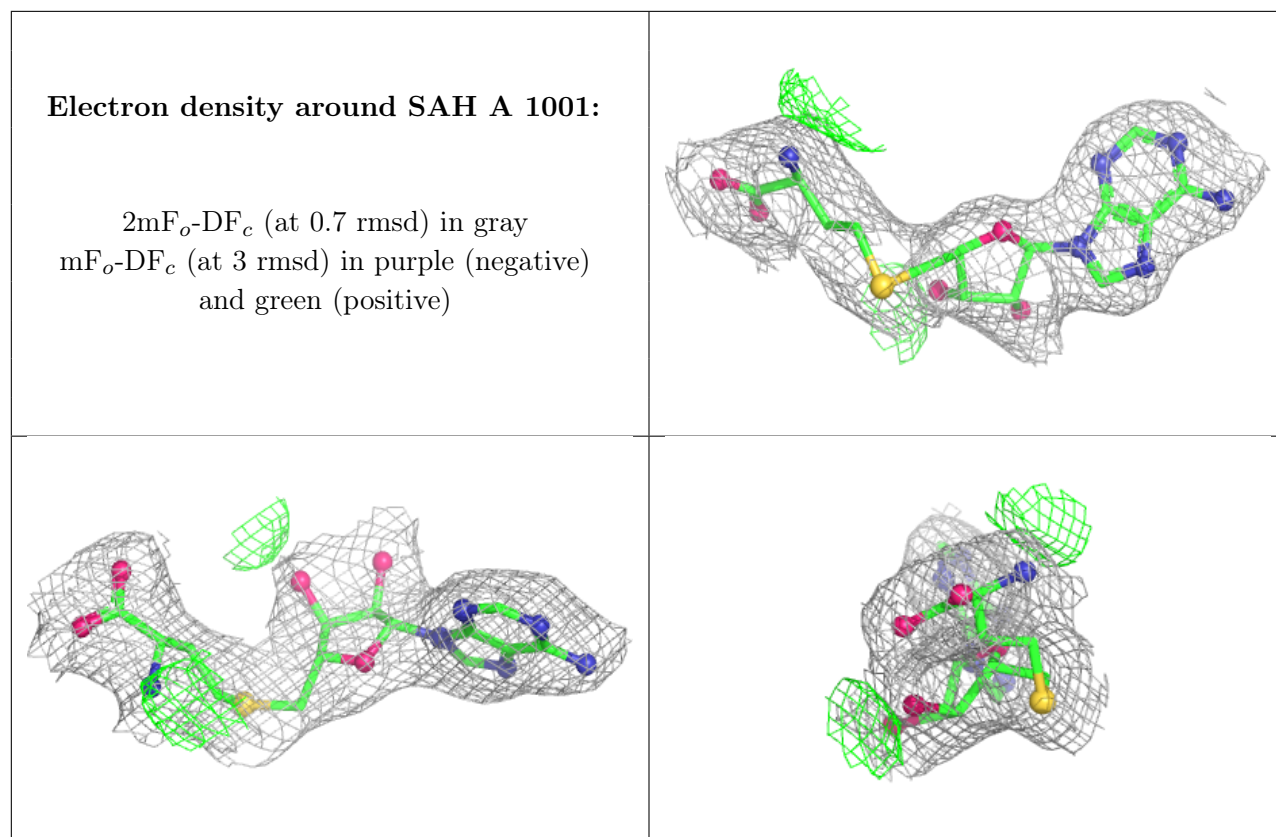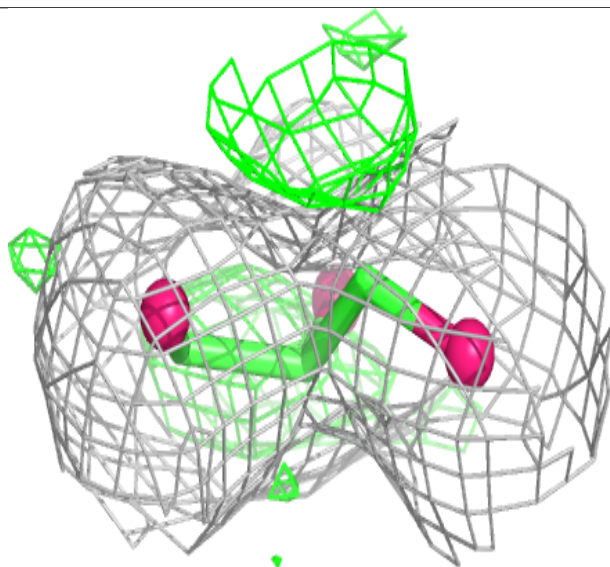**Electron density around GOL A 1004:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
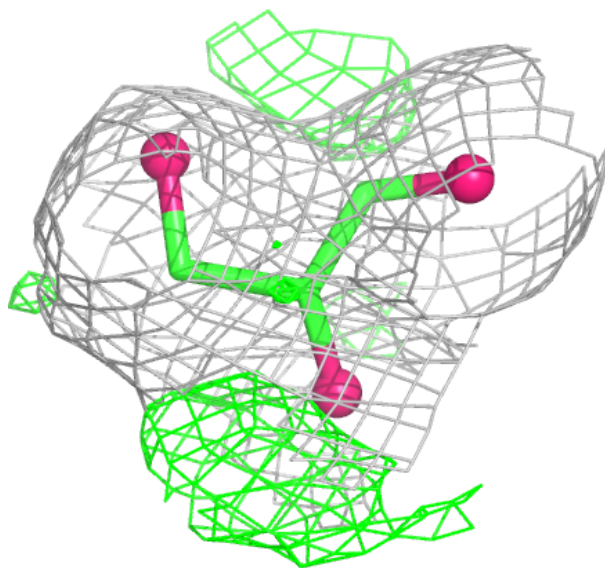and green (positive)
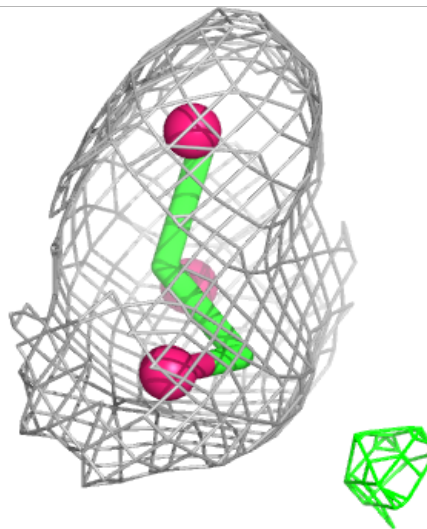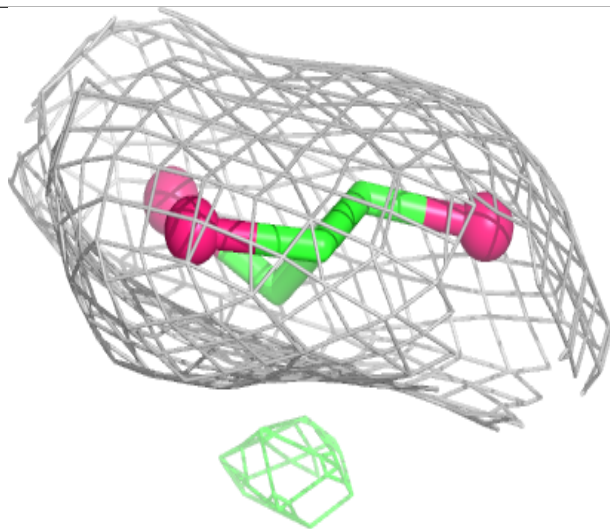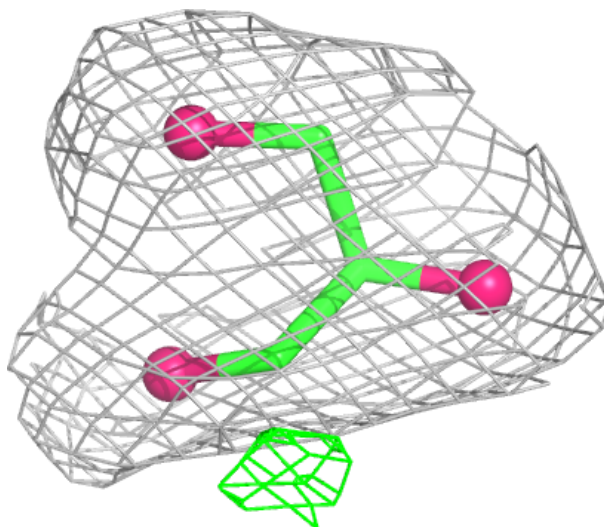
**Electron density around GOL A 1005:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL B 1002:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
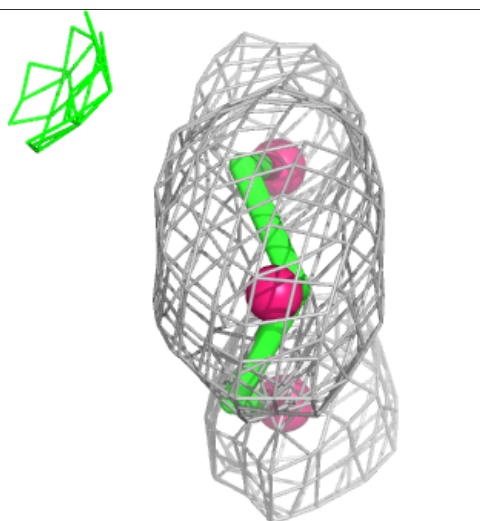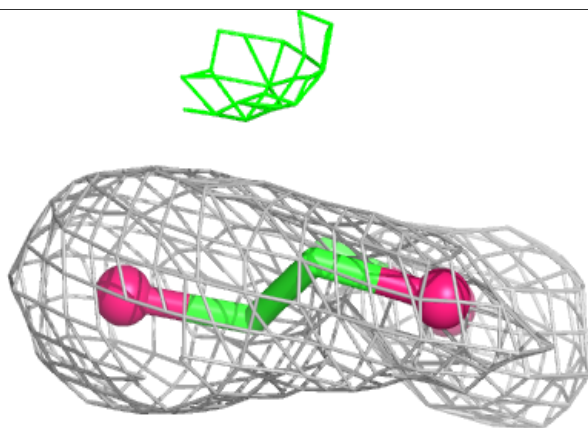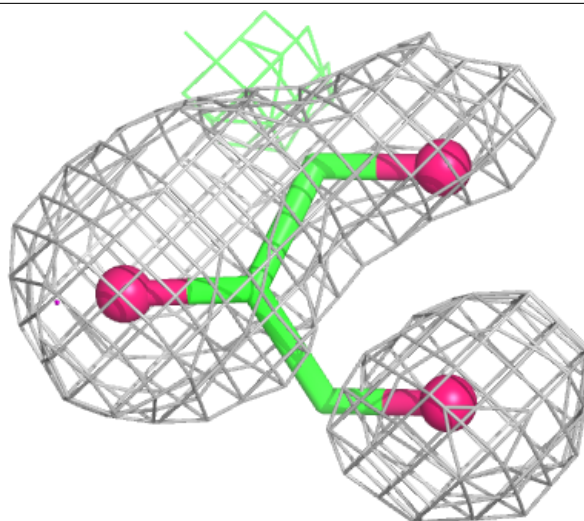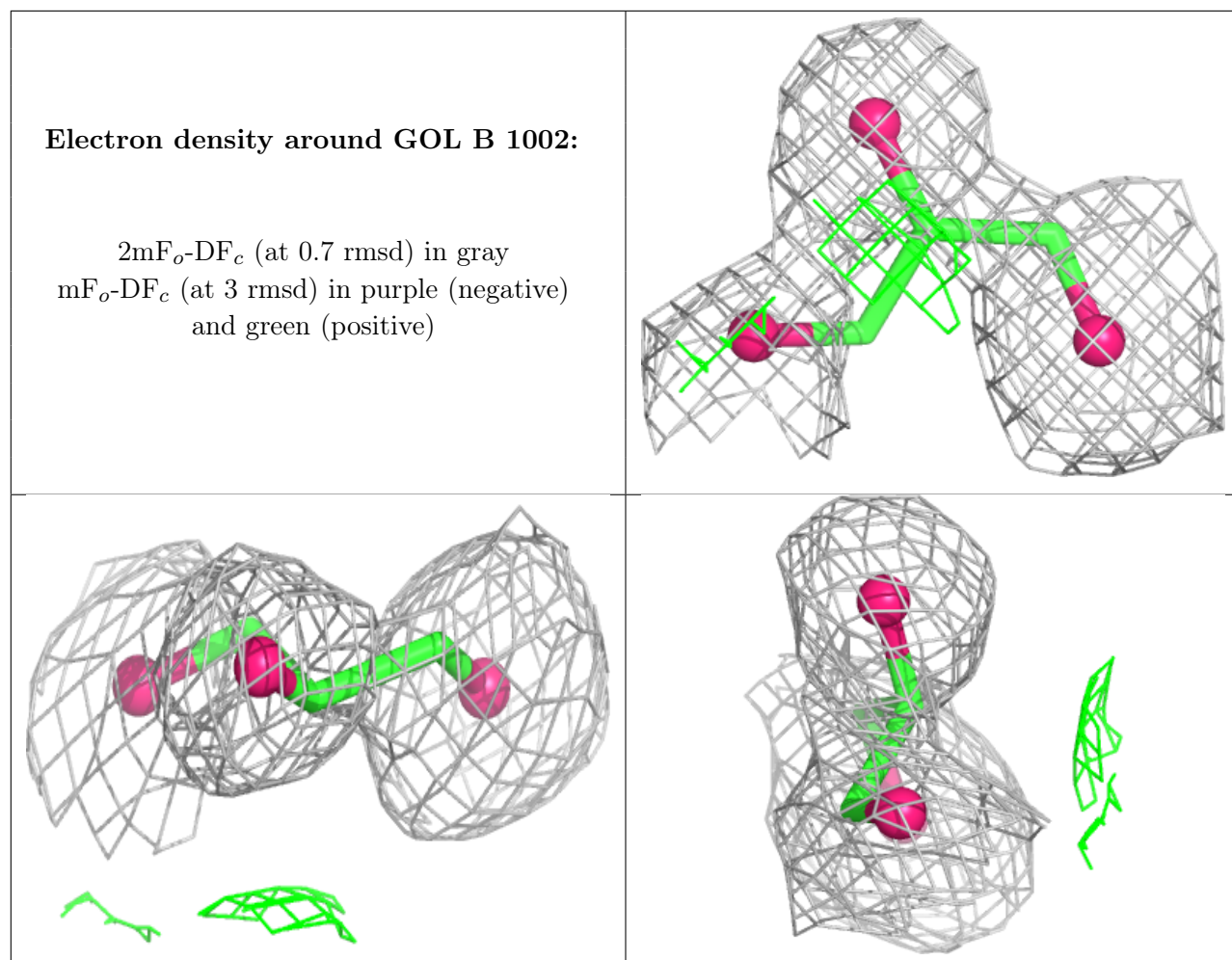
**Electron density around GOL B 1005:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
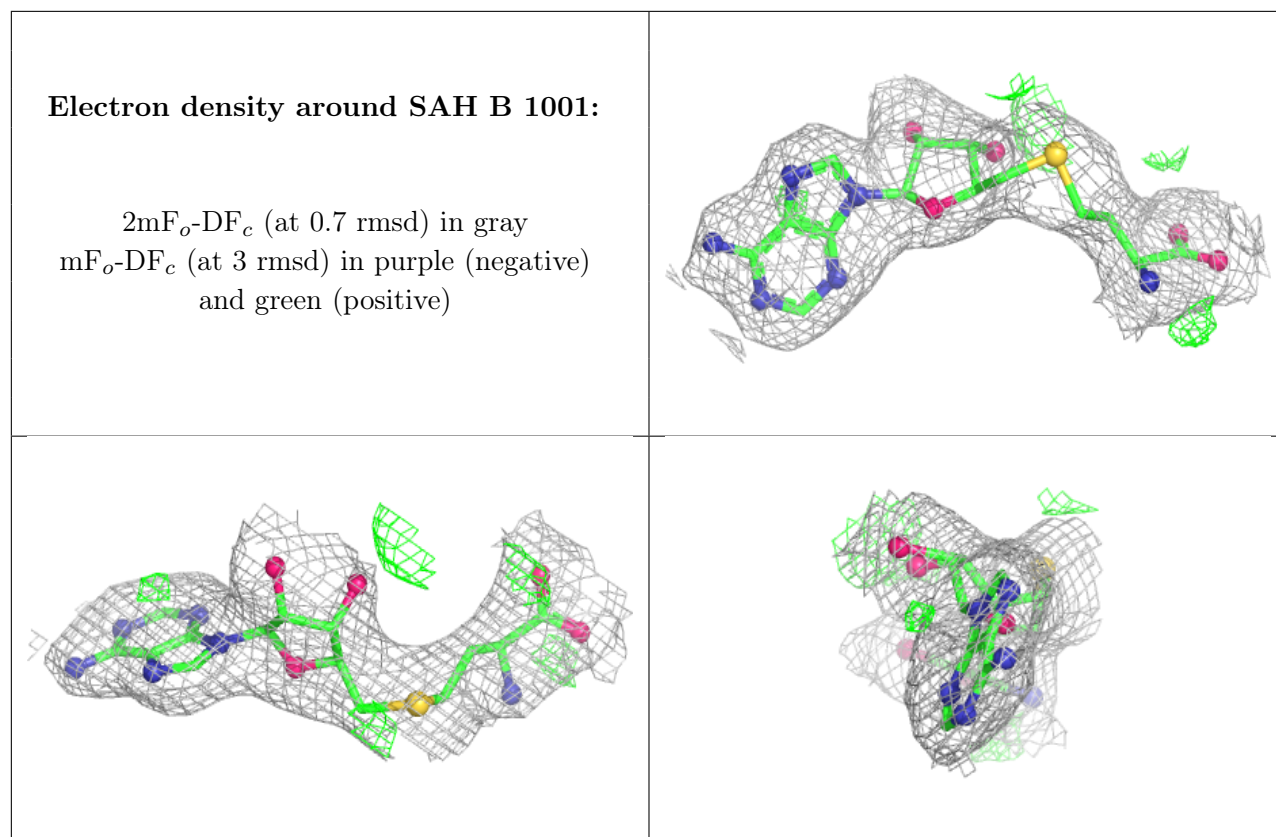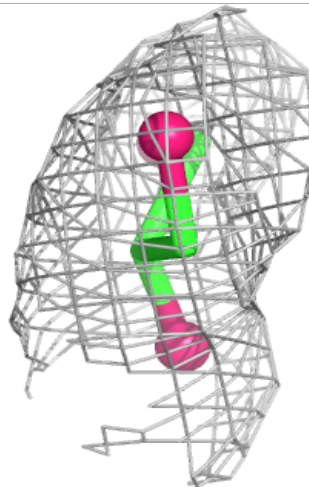and green (positive)
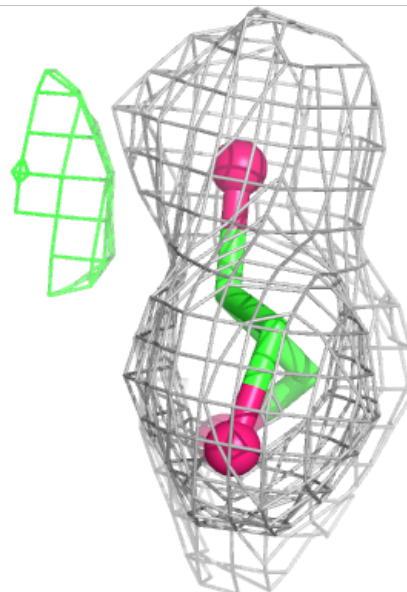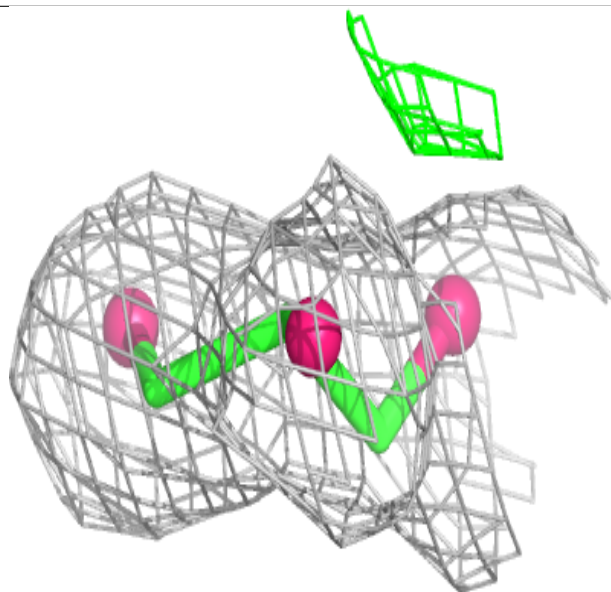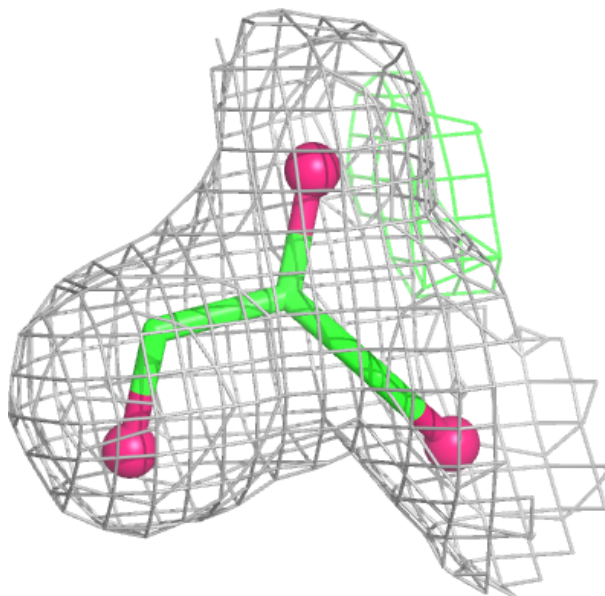
**Electron density around GOL B 1006:**
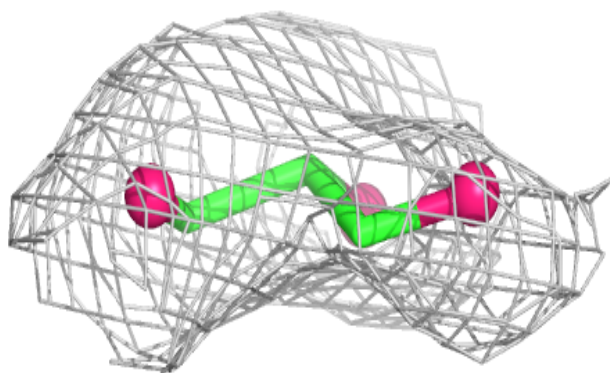
$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL C 1002:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
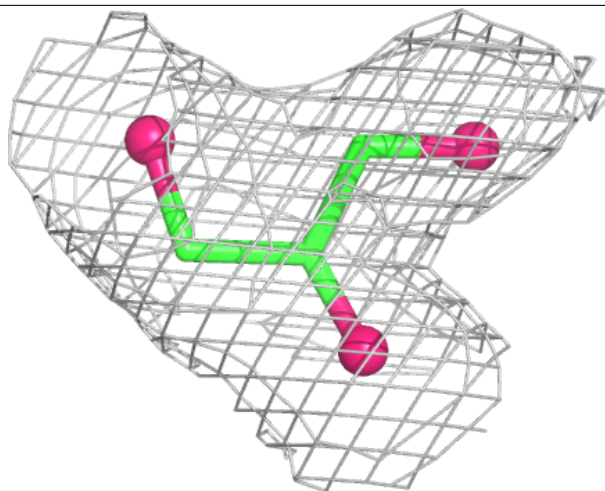


**Electron density around GOL A 1002:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
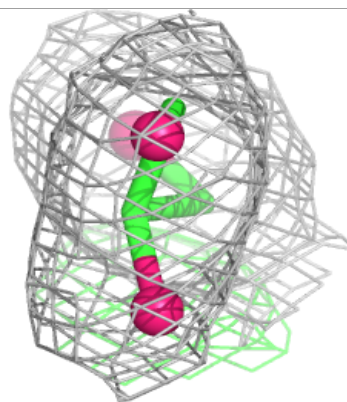and green (positive)

**Electron density around GOL D 1004:**

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)
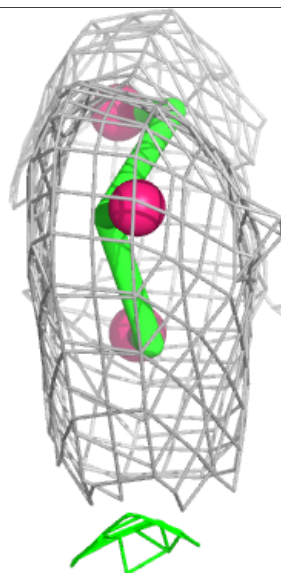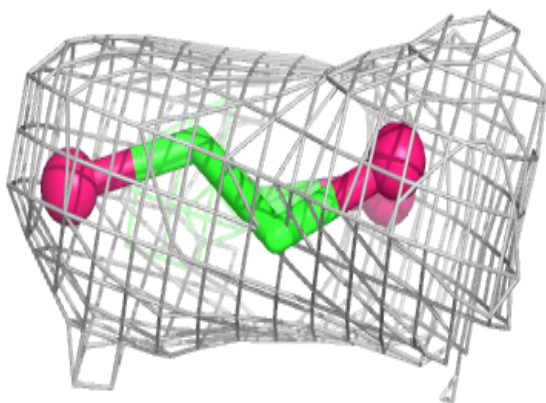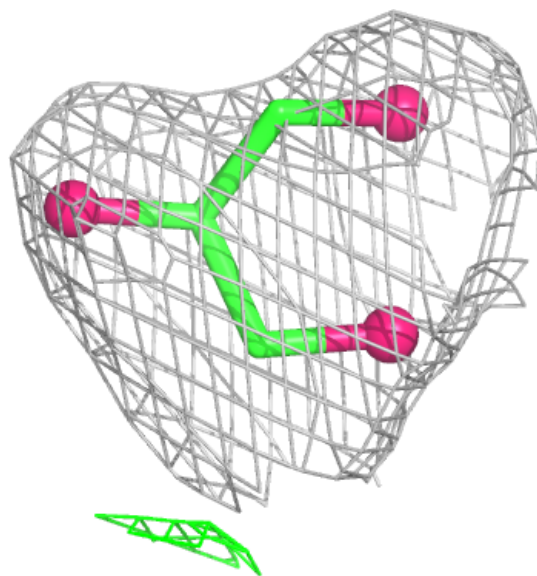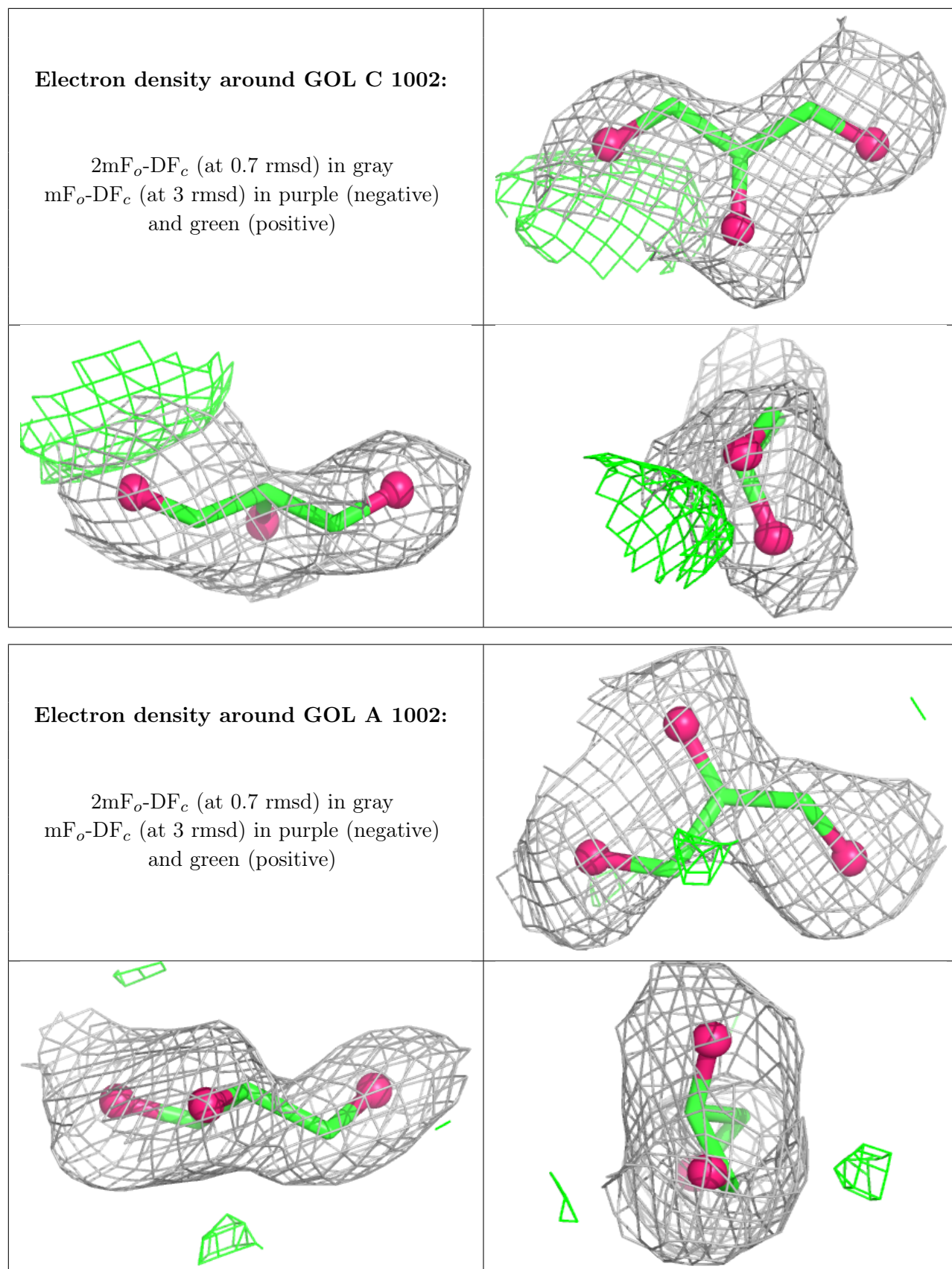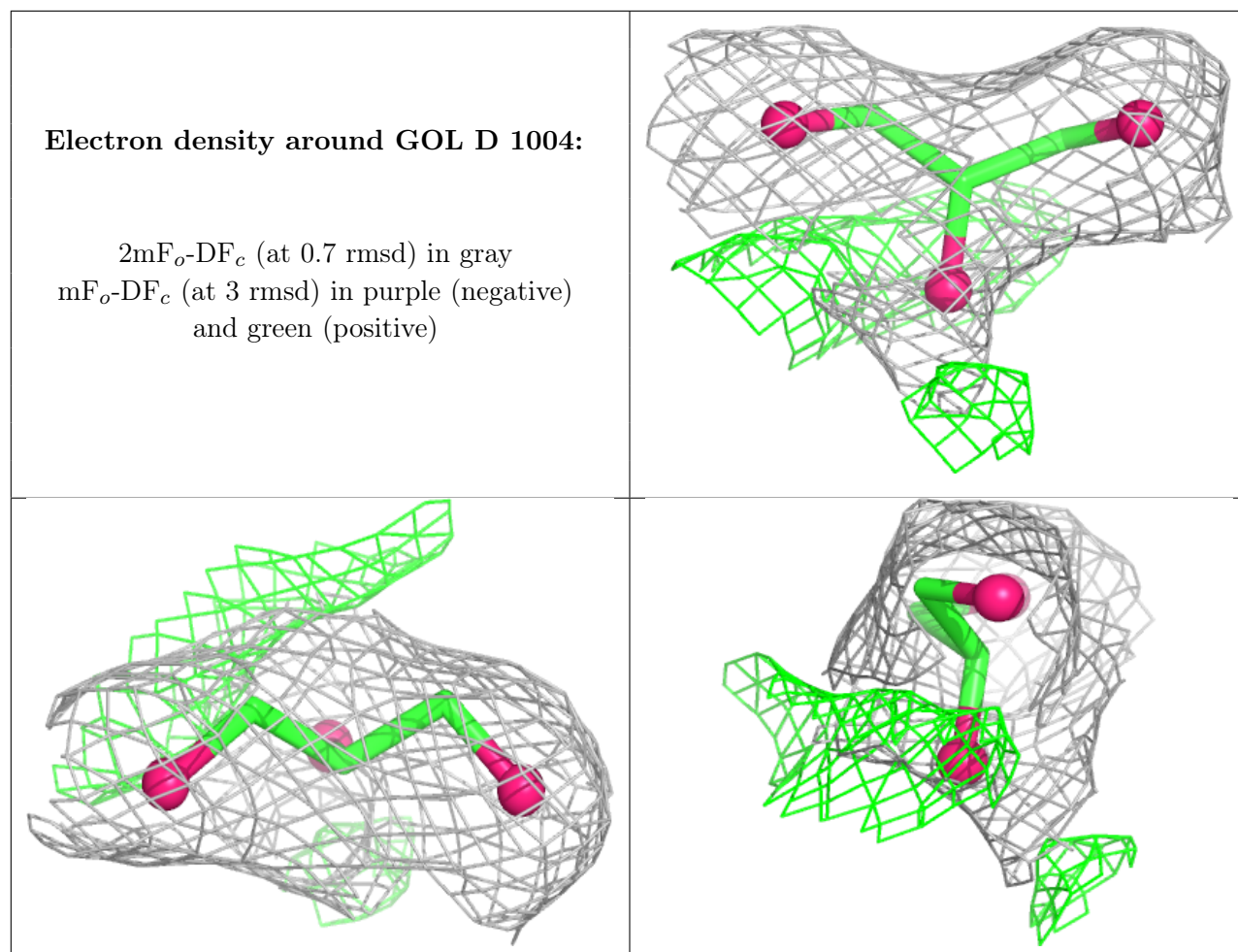
**Electron density around SAH D 1001:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
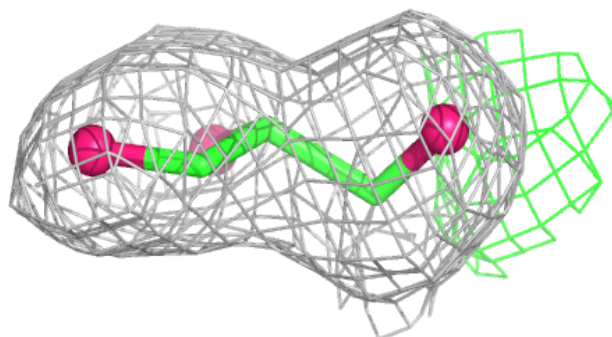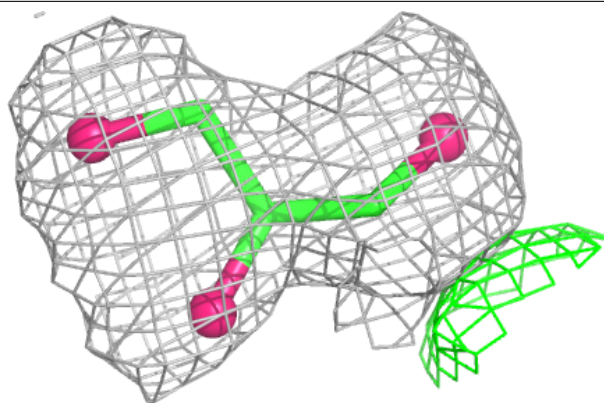and green (positive)
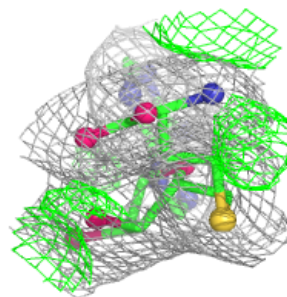
**Electron density around GOL B 1003:**

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL B 1004:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around SAH A 1001:**

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL A 1003:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around SAH B 1001:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL C 1003:**

$2mF_o$-$DF_c$ (at 0.7 rmsd) in gray
$mF_o$-$DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL D 1002:**

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
$mF_o-DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)



**Electron density around SAH C 1001:**

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
$mF_o-DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL A 1006:**

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Electron density around GOL D 1005:**

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)
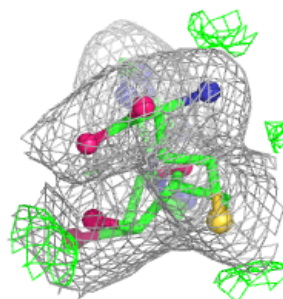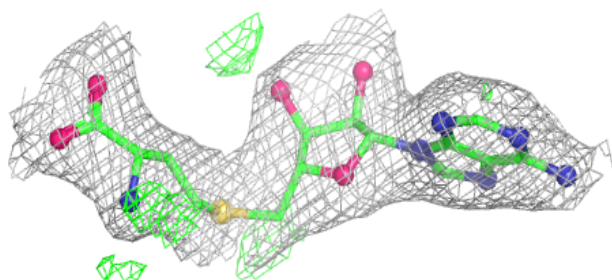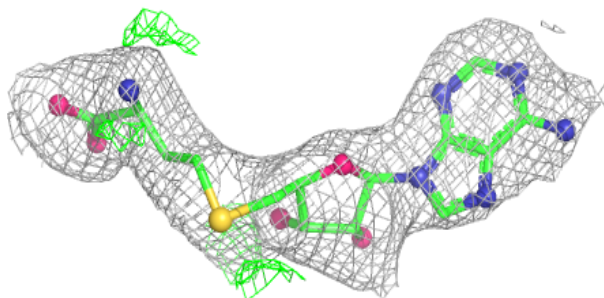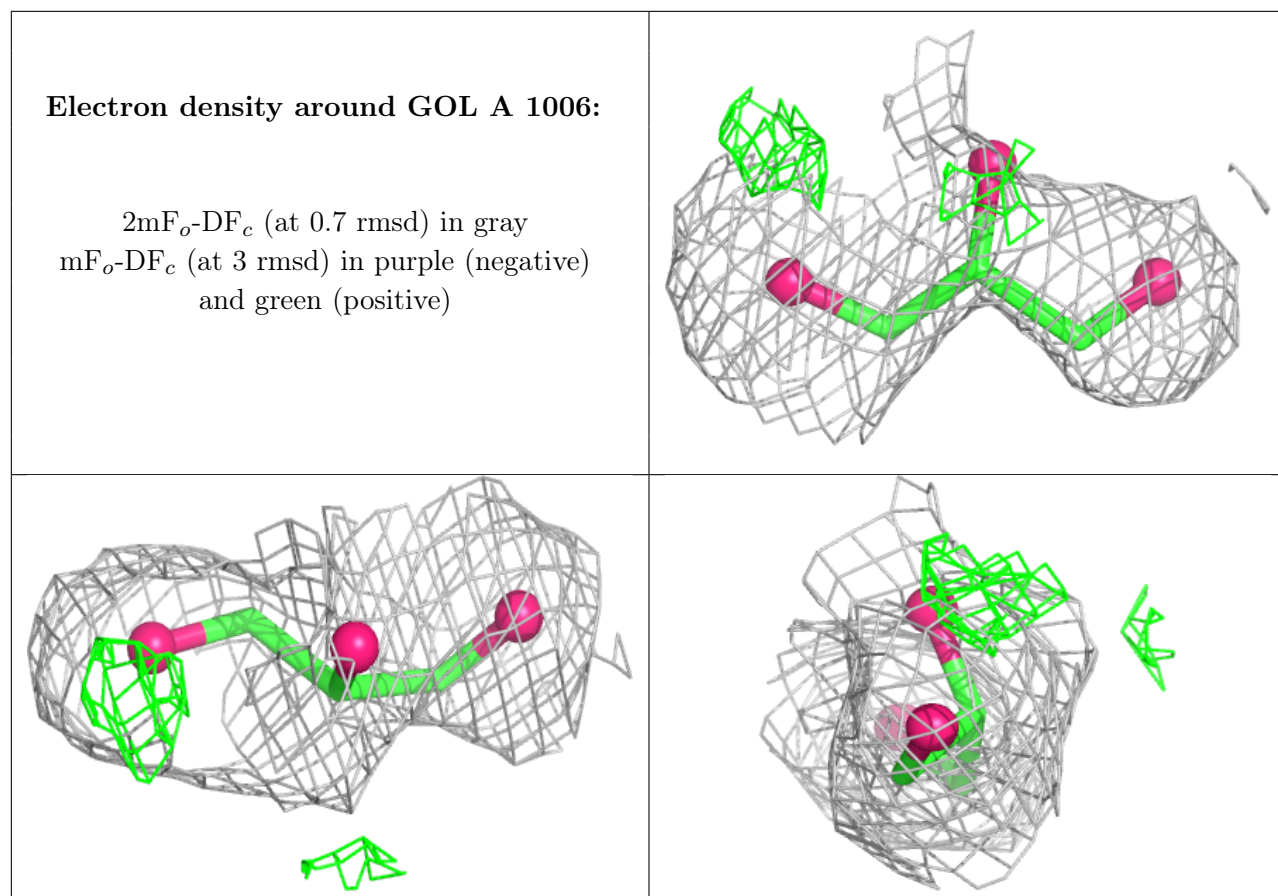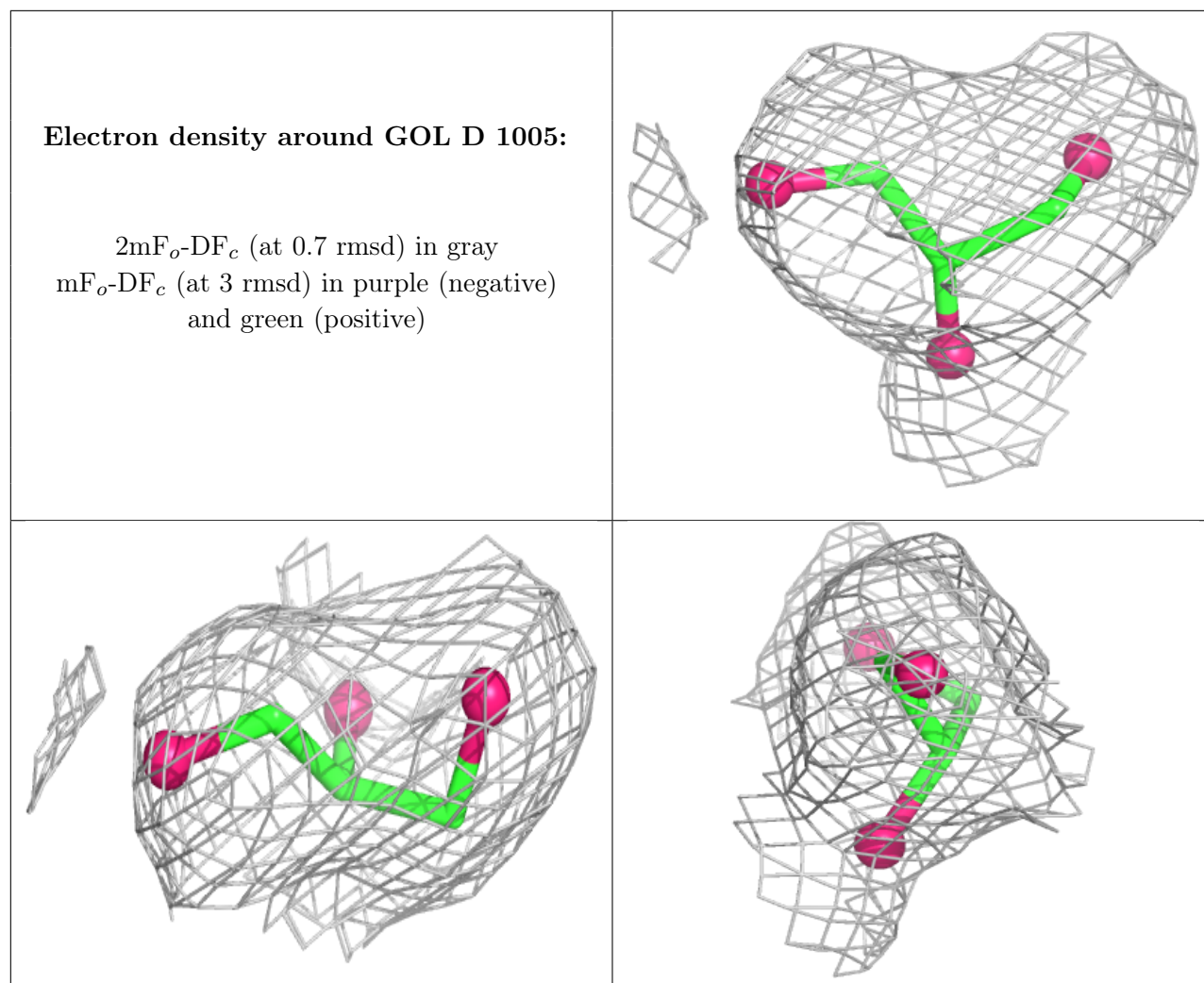


## 6.5 Other polymers (i)

There are no such residues in this entry.