wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2024 – 11:24 AM EST

PDB ID	:	9E83
Title	:	TMPRSS2 crystal structure following acylation by UCSF_157
Authors	:	Fraser, B.J.; Dong, A.; Seitova, A.; Li, Y.; Hutchinson, A.; Young, N.; Bender, B.; Gahbauer, S.; Edwards, A.; Shoichet, B.; Craik, C.; Arrowsmith, C.H.
Deposited on	:	2024-11-04
Resolution	:	2.07 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

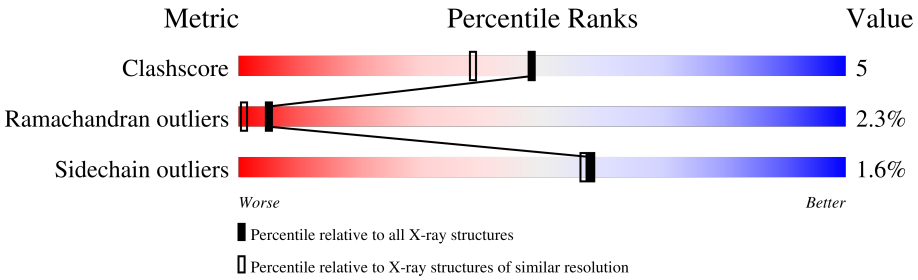
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.07 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	180529	3661 (2.08-2.04)
Ramachandran outliers	177936	3649 (2.08-2.04)
Sidechain outliers	177891	3649 (2.08-2.04)

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.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	B	249	84% 12% 5%
2	A	110	65% 11% • 22%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 2708 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 Transmembrane protease serine 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	237	1846	1189	305	336	16	0	7	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	493	GLU	-	expression tag	UNP O15393
B	494	PHE	-	expression tag	UNP O15393
B	495	VAL	-	expression tag	UNP O15393
B	496	GLU	-	expression tag	UNP O15393
B	497	HIS	-	expression tag	UNP O15393
B	498	HIS	-	expression tag	UNP O15393
B	499	HIS	-	expression tag	UNP O15393
B	500	HIS	-	expression tag	UNP O15393
B	501	HIS	-	expression tag	UNP O15393
B	502	HIS	-	expression tag	UNP O15393
B	503	HIS	-	expression tag	UNP O15393
B	504	HIS	-	expression tag	UNP O15393

- Molecule 2 is a protein called Transmembrane protease serine 2 non-catalytic chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	86	621	395	110	109	7	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

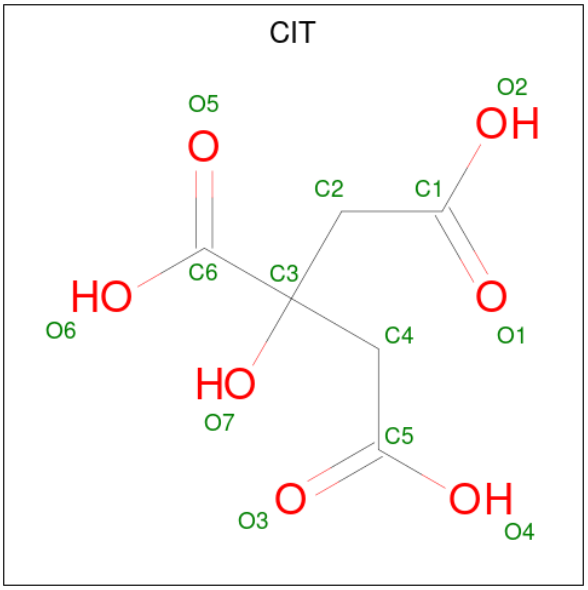
Chain	Residue	Modelled	Actual	Comment	Reference
A	146	ALA	-	expression tag	UNP O15393
A	147	ALA	-	expression tag	UNP O15393
A	250	ASP	SER	engineered mutation	UNP O15393
A	251	ASP	SER	engineered mutation	UNP O15393

Continued on next page...

Continued from previous page...

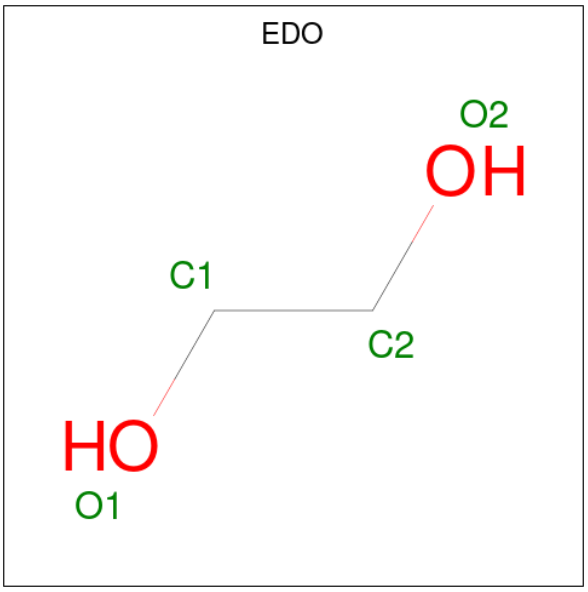
Chain	Residue	Modelled	Actual	Comment	Reference
A	252	ASP	ARG	engineered mutation	UNP O15393
A	253	ASP	GLN	engineered mutation	UNP O15393
A	254	ASP	SER	engineered mutation	UNP O15393
A	255	LYS	ARG	engineered mutation	UNP O15393

- Molecule 3 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



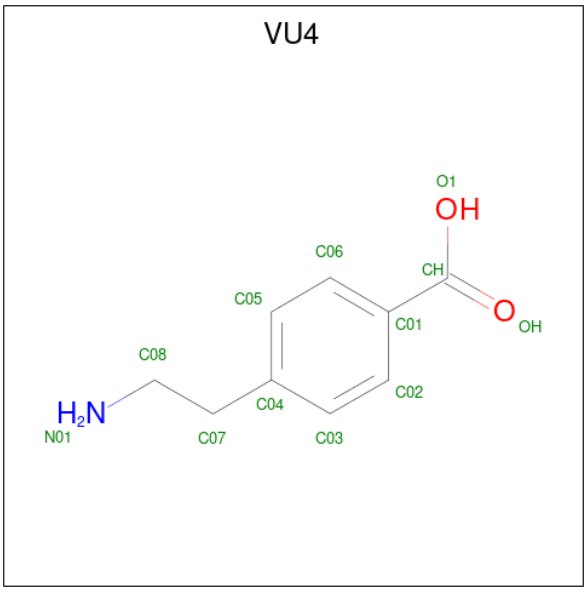
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 13 6 7	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

- Molecule 5 is 4-(2-aminoethyl)benzoic acid (three-letter code: VU4) (formula: $C_9H_{11}NO_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C N O 11 9 1 1	0	0

- Molecule 6 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total X 1 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	B	183	Total O 184 184	0	1
7	A	20	Total O 20 20	0	0

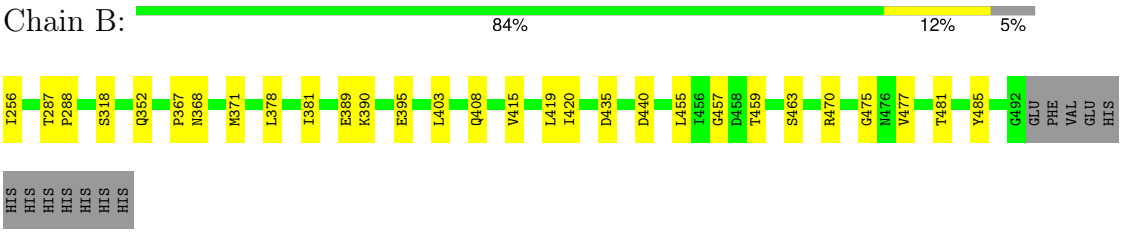
3 Residue-property plots (i)

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.

Note EDS failed to run properly.

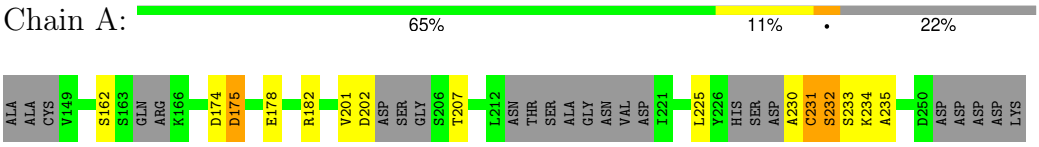
• Molecule 1: Transmembrane protease serine 2

Chain B:



• Molecule 2: Transmembrane protease serine 2 non-catalytic chain

Chain A:



4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	60.85Å 51.39Å 64.62Å 90.00° 91.01° 90.00°	Depositor
Resolution (Å)	39.29 – 2.07	Depositor
% Data completeness (in resolution range)	96.2 (39.29-2.07)	Depositor
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 2.06Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
R, R_{free}	0.200 , 0.260	Depositor
Wilson B-factor (Å^2)	26.4	Xtriage
Anisotropy	0.854	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.041 for h,-k,-l	Xtriage
Total number of atoms	2708	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.29% 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: CIT, VU4, UNX, EDO

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	B	0.50	0/1922	0.82	0/2624
2	A	0.45	0/631	0.66	0/853
All	All	0.48	0/2553	0.78	0/3477

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	B	1846	0	1790	20	0
2	A	621	0	541	6	0
3	B	13	0	5	0	0
4	B	12	0	18	3	0
5	B	11	0	0	1	0
6	A	1	0	0	0	0
7	A	20	0	0	0	0
7	B	184	0	0	3	0
All	All	2708	0	2354	26	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 5.

The worst 5 of 26 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:B:419:LEU:O	4:B:601:EDO:H21	1.84	0.77
1:B:420:ILE:O	7:B:701:HOH:O	2.14	0.64
1:B:367:PRO:HD3	1:B:455:LEU:O	2.05	0.57
1:B:457:GLY:HA2	1:B:477:VAL:HG23	1.88	0.55
1:B:435:ASP:OD2	5:B:602:VU4:N01	2.40	0.53

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	B	242/249 (97%)	233 (96%)	8 (3%)	1 (0%)	30 23
2	A	76/110 (69%)	63 (83%)	7 (9%)	6 (8%)	1 0
All	All	318/359 (89%)	296 (93%)	15 (5%)	7 (2%)	5 1

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	232	SER
2	A	231	CYS
2	A	233	SER
2	A	175	ASP
2	A	234	LYS

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	B	199/210 (95%)	197 (99%)	2 (1%)	73 73
2	A	57/95 (60%)	54 (95%)	3 (5%)	19 12
All	All	256/305 (84%)	251 (98%)	5 (2%)	58 47

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	352[A]	GLN
1	B	352[B]	GLN
2	A	162	SER
2	A	207	THR
2	A	225	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	277	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 1 is unknown - leaving 5 for Mogul analysis.

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	$\#\|Z\| > 2$
4	EDO	B	604	-	3,3,3	0.28	0	2,2,2	0.39	0
4	EDO	B	603	-	3,3,3	0.21	0	2,2,2	0.40	0
5	VU4	B	602	1	11,11,12	0.61	0	13,13,15	1.15	2 (15%)
4	EDO	B	601	-	3,3,3	0.12	0	2,2,2	0.17	0
3	CIT	B	600	-	12,12,12	1.38	1 (8%)	17,17,17	1.36	2 (11%)

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
4	EDO	B	604	-	-	1/1/1/1	-
4	EDO	B	603	-	-	1/1/1/1	-
5	VU4	B	602	1	-	2/5/5/7	0/1/1/1
4	EDO	B	601	-	-	1/1/1/1	-
3	CIT	B	600	-	-	5/16/16/16	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	600	CIT	C3-C6	3.19	1.56	1.53

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	600	CIT	O6-C6-C3	3.19	119.26	113.14
3	B	600	CIT	O5-C6-C3	-2.78	116.70	122.09
5	B	602	VU4	C07-C08-N01	-2.61	105.56	112.71
5	B	602	VU4	C08-C07-C04	2.21	117.77	112.82

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	600	CIT	C2-C3-C6-O5
3	B	600	CIT	C2-C3-C6-O6
3	B	600	CIT	O7-C3-C6-O5
3	B	600	CIT	O7-C3-C6-O6
5	B	602	VU4	C02-C01-CH-OH

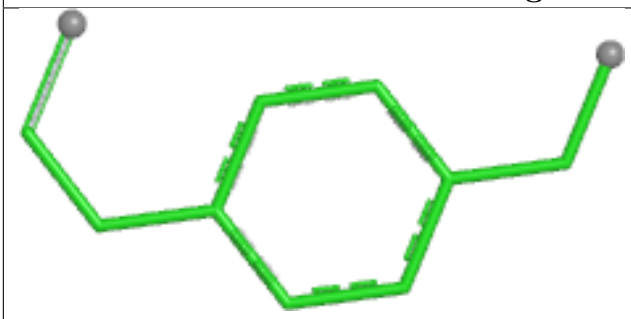
There are no ring outliers.

2 monomers are involved in 4 short contacts:

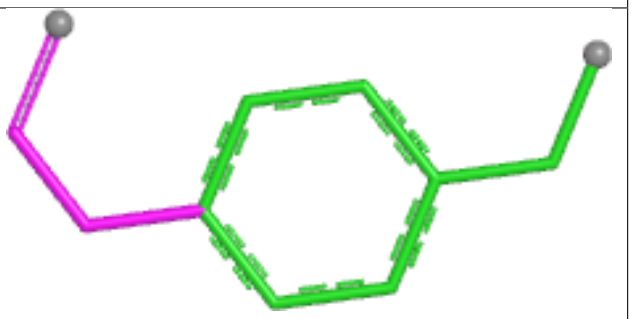
Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	602	VU4	1	0
4	B	601	EDO	3	0

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.

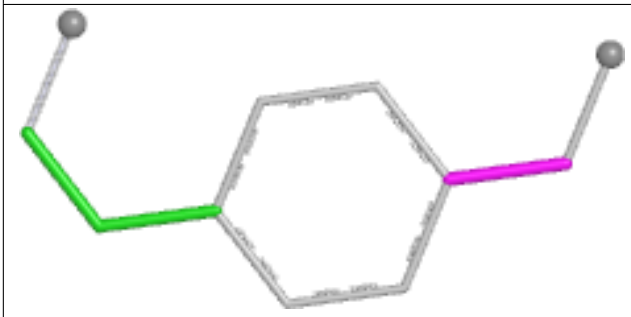
Ligand VU4 B 602



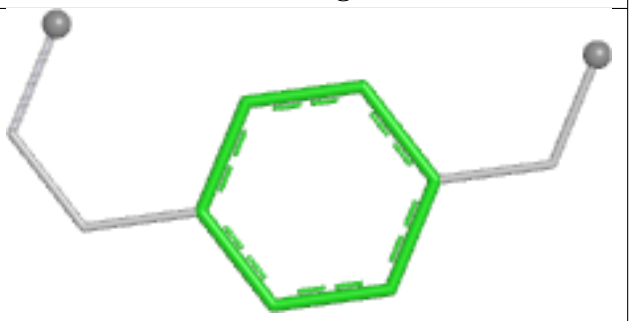
Bond lengths



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

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

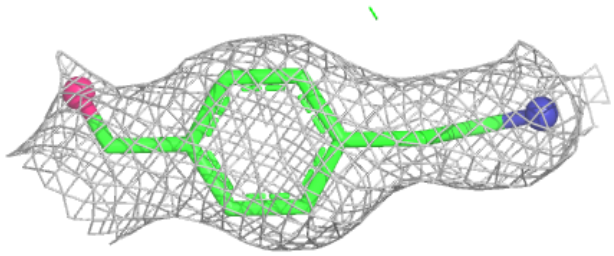
6.4 Ligands

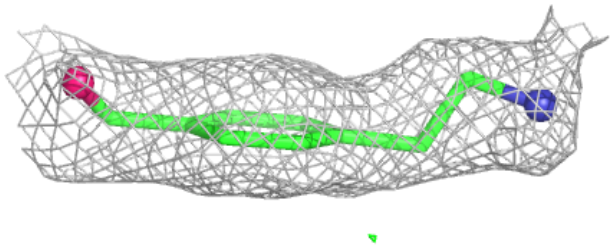
EDS failed to run properly - this section is therefore empty.

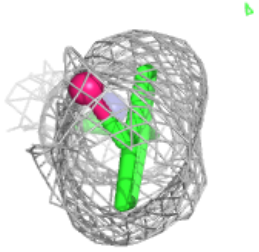
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around VU4 B 602:

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

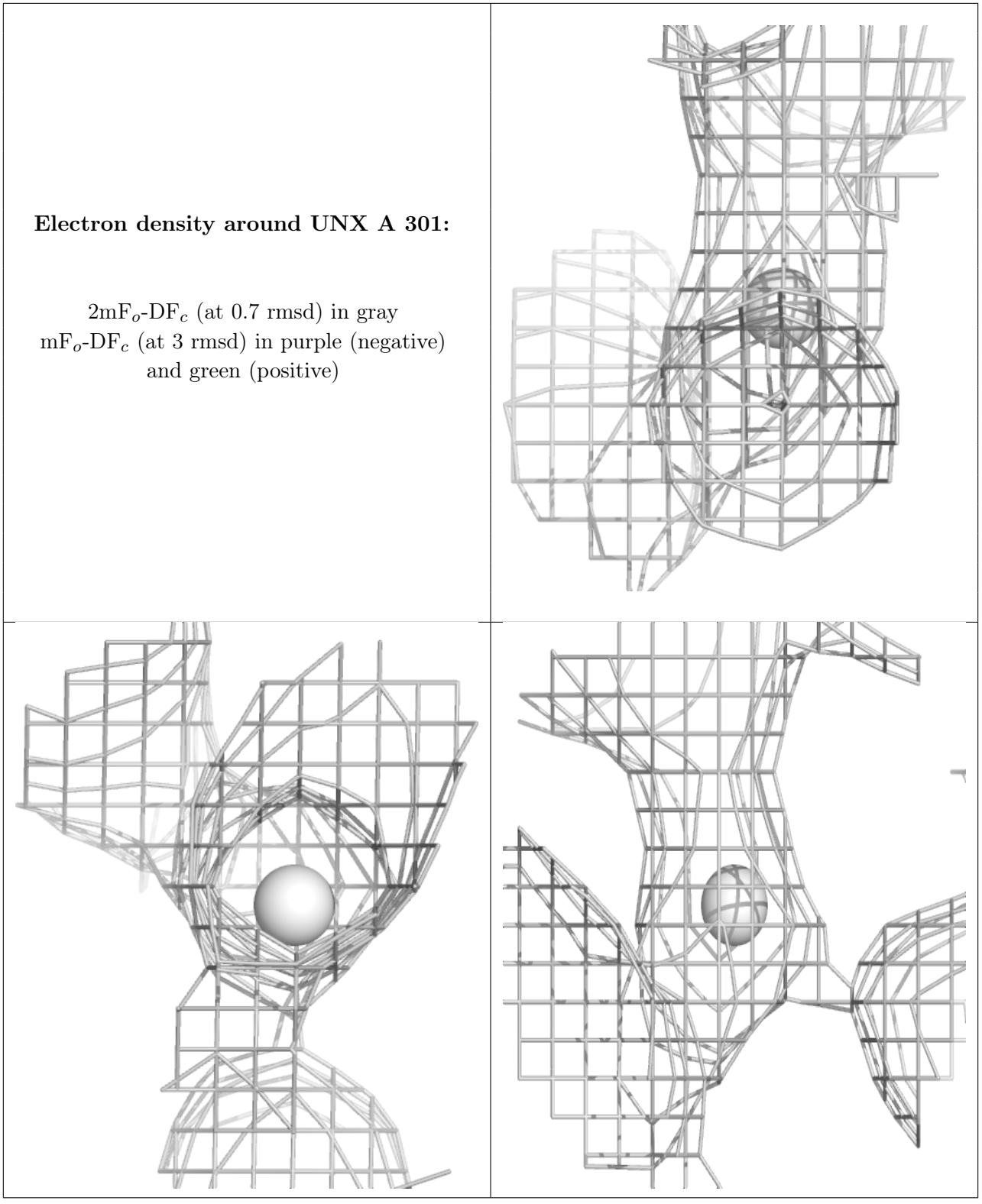






Electron density around UNX A 301:

$2mF_o$-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)

EDS failed to run properly - this section is therefore empty.