WORLDWIDE wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 19, 2023 – 07:23 PM EDT

PDB ID	:	2HHL
Title	:	Crystal structure of the human small CTD phosphatase 3 isoform 1
Authors	:	Malashkevich, V.N.; Toro, R.; Ramagopal, U.; Sauder, J.M.; Schwinn, K.D.; Thompson, D.A.; Rutter, M.E.; Dickey, M.; Groshong, C.; Bain, K.T.; Adams, J.M.; Reyes, C.; Rooney, I.; Powell, A.; Boice, A.; Gheyi, T.; Ozyurt, S.; Atwell, S.; Wasserman, S.R.; Emtage, S.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2006-06-28
Resolution	:	2.10 Å(reported)

––––––––––––––––––––––

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

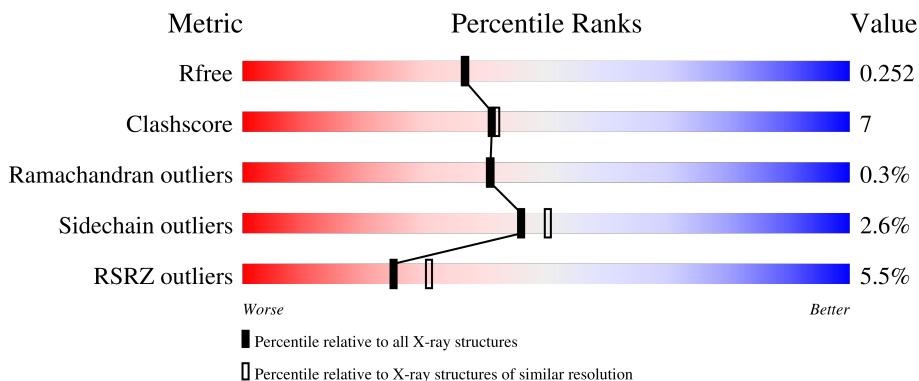|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)

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 2.10 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		0.252
Clashscore		7
Ramachandran outliers		0.3%
Sidechain outliers		2.6%
RSRZ outliers		5.5%

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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	195	7% 83% 9% 8%
1	B	195	5% 80% 11% · 8%
1	C	195	5% 77% 14% · 6%

Continued on next page...

Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)	
Validation Pipeline (wwPDB-VP)	:	2.35	

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	D	195	3% ... 80% ... 8% · 12%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	KEG	A	902	-	-	X	-
2	KEG	D	901	-	-	X	-

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 6472 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 CTD small phosphatase-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	180	Total	C	N	O	S	0	0	0
			1451	944	239	263	5			
1	B	179	Total	C	N	O	S	0	0	0
			1442	939	237	261	5			
1	C	184	Total	C	N	O	S	0	0	0
			1487	965	246	271	5			
1	D	172	Total	C	N	O	S	0	0	0
			1394	904	230	255	5			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP O15194
A	2	SER	-	cloning artifact	UNP O15194
A	3	LEU	-	cloning artifact	UNP O15194
A	188	GLU	-	cloning artifact	UNP O15194
A	189	GLY	-	cloning artifact	UNP O15194
A	190	HIS	-	expression tag	UNP O15194
A	191	HIS	-	expression tag	UNP O15194
A	192	HIS	-	expression tag	UNP O15194
A	193	HIS	-	expression tag	UNP O15194
A	194	HIS	-	expression tag	UNP O15194
A	195	HIS	-	expression tag	UNP O15194
B	1	MET	-	initiating methionine	UNP O15194
B	2	SER	-	cloning artifact	UNP O15194
B	3	LEU	-	cloning artifact	UNP O15194
B	188	GLU	-	cloning artifact	UNP O15194
B	189	GLY	-	cloning artifact	UNP O15194
B	190	HIS	-	expression tag	UNP O15194
B	191	HIS	-	expression tag	UNP O15194
B	192	HIS	-	expression tag	UNP O15194
B	193	HIS	-	expression tag	UNP O15194
B	194	HIS	-	expression tag	UNP O15194

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	195	HIS	-	expression tag	UNP O15194
C	1	MET	-	initiating methionine	UNP O15194
C	2	SER	-	cloning artifact	UNP O15194
C	3	LEU	-	cloning artifact	UNP O15194
C	188	GLU	-	cloning artifact	UNP O15194
C	189	GLY	-	cloning artifact	UNP O15194
C	190	HIS	-	expression tag	UNP O15194
C	191	HIS	-	expression tag	UNP O15194
C	192	HIS	-	expression tag	UNP O15194
C	193	HIS	-	expression tag	UNP O15194
C	194	HIS	-	expression tag	UNP O15194
C	195	HIS	-	expression tag	UNP O15194
D	1	MET	-	initiating methionine	UNP O15194
D	2	SER	-	cloning artifact	UNP O15194
D	3	LEU	-	cloning artifact	UNP O15194
D	188	GLU	-	cloning artifact	UNP O15194
D	189	GLY	-	cloning artifact	UNP O15194
D	190	HIS	-	expression tag	UNP O15194
D	191	HIS	-	expression tag	UNP O15194
D	192	HIS	-	expression tag	UNP O15194
D	193	HIS	-	expression tag	UNP O15194
D	194	HIS	-	expression tag	UNP O15194
D	195	HIS	-	expression tag	UNP O15194

- Molecule 2 is 12-TUNGSTOPHOSPHATE (three-letter code: KEG) (formula: $O_{40}PW_{12}$).
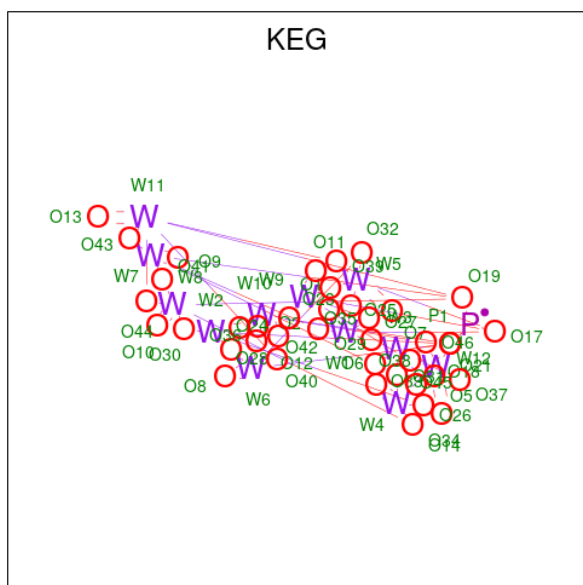
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 53	O 40	P 1	W 12	0	0
2	C	1	Total 52	O 40		W 12	0	0
2	D	1	Total 53	O 40	P 1	W 12	0	0

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	102	Total 102	O 102	0	0
3	B	141	Total 141	O 141	0	0
3	C	138	Total 138	O 138	0	0
3	D	159	Total 159	O 159	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.
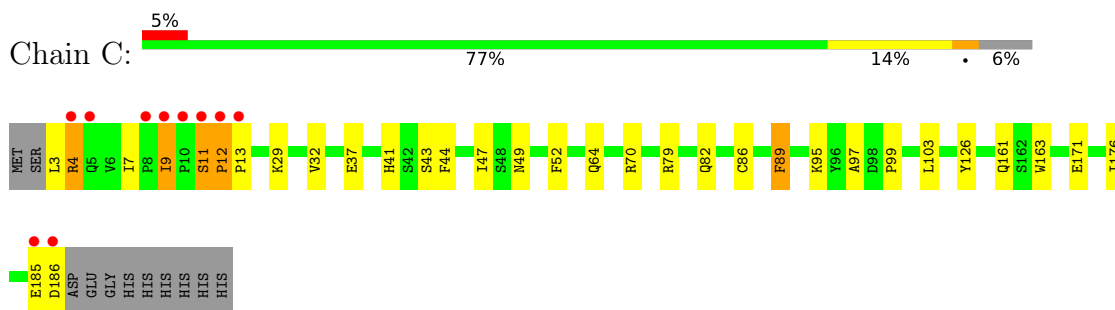
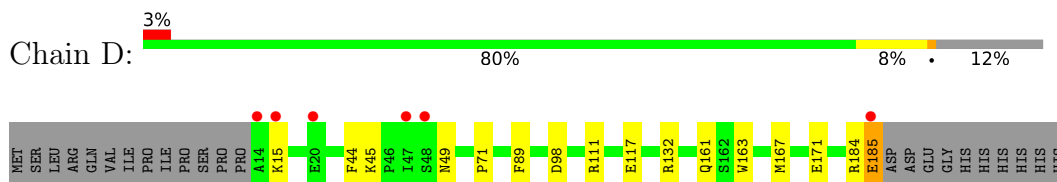• Molecule 1: CTD small phosphatase-like protein
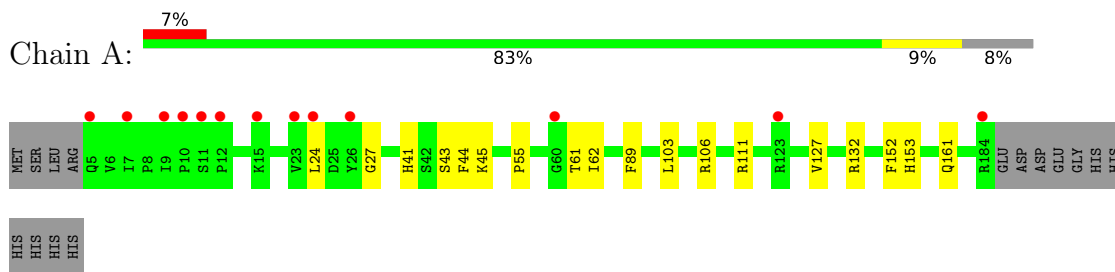
Chain A:

• Molecule 1: CTD small phosphatase-like protein

Chain B:

• Molecule 1: CTD small phosphatase-like protein

Chain C:

• Molecule 1: CTD small phosphatase-like protein

Chain D:
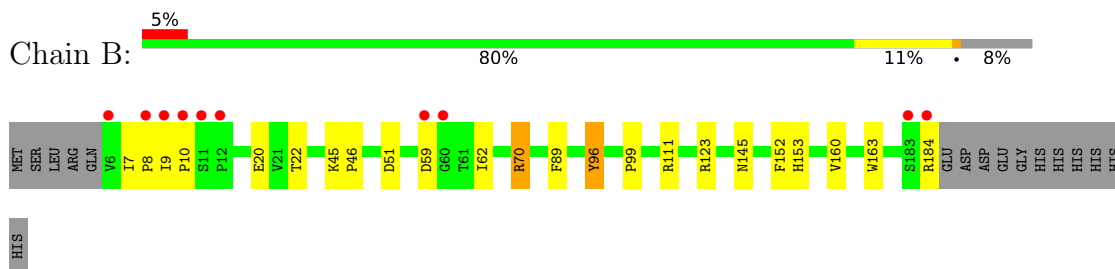
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.42Å 49.86Å 179.75Å 90.00° 96.90° 90.00°	Depositor
Resolution (Å)	19.86 – 2.10 19.86 – 2.10	Depositor EDS
% Data completeness (in resolution range)	96.4 (19.86-2.10) 96.4 (19.86-2.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.49 (at 2.09Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.188 , 0.252 0.189 , 0.252	Depositor DCC
R_{free} test set	2386 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	26.6	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 43.8	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6472	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.14 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.5944e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

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

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.96	0/1490	0.86	0/2030
1	B	1.00	2/1481 (0.1%)	0.88	3/2018 (0.1%)
1	C	1.00	3/1526 (0.2%)	0.86	1/2078 (0.0%)
1	D	1.03	1/1429 (0.1%)	0.88	1/1942 (0.1%)
All	All	1.00	6/5926 (0.1%)	0.87	5/8068 (0.1%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	96	TYR	CE2-CZ	5.96	1.46	1.38
1	C	126	TYR	CD2-CE2	5.69	1.47	1.39
1	D	117	GLU	CG-CD	5.65	1.60	1.51
1	C	97	ALA	CA-CB	5.40	1.63	1.52
1	C	86	CYS	CB-SG	-5.22	1.73	1.81

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	111	ARG	NE-CZ-NH2	-5.94	117.33	120.30
1	D	111	ARG	NE-CZ-NH2	-5.79	117.41	120.30
1	C	95	LYS	CD-CE-NZ	-5.40	99.27	111.70
1	B	70	ARG	NE-CZ-NH2	-5.38	117.61	120.30
1	B	51	ASP	CB-CG-OD2	5.09	122.88	118.30

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	1451	0	1447	15	1
1	B	1442	0	1439	10	0
1	C	1487	0	1481	26	0
1	D	1394	0	1381	19	0
2	A	53	0	0	12	1
2	C	52	0	0	8	0
2	D	53	0	0	9	0
3	A	102	0	0	4	0
3	B	141	0	0	3	0
3	C	138	0	0	2	0
3	D	159	0	0	0	0
All	All	6472	0	5748	83	1

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 7.

The worst 5 of 83 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:62:ILE:HG13	2:C:900:KEG:O41	1.51	1.09
2:A:902:KEG:O33	2:A:902:KEG:O32	1.71	1.09
1:D:163:TRP:HZ2	1:D:167:MET:CE	1.72	1.03
1:D:163:TRP:HZ2	1:D:167:MET:HE3	1.24	1.00
2:A:902:KEG:O13	2:A:902:KEG:P1	2.23	0.97

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:27:GLY:CA	2:A:902:KEG:O27[1_655]	2.06	0.14

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	178/195 (91%)	173 (97%)	5 (3%)	0	100	100
1	B	177/195 (91%)	169 (96%)	8 (4%)	0	100	100
1	C	182/195 (93%)	176 (97%)	4 (2%)	2 (1%)	14	9
1	D	170/195 (87%)	162 (95%)	8 (5%)	0	100	100
All	All	707/780 (91%)	680 (96%)	25 (4%)	2 (0%)	41	41

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	12	PRO
1	C	11	SER

5.3.2 Protein sidechains ⓘ

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	164/178 (92%)	160 (98%)	4 (2%)	49	53
1	B	163/178 (92%)	158 (97%)	5 (3%)	40	43
1	C	168/178 (94%)	163 (97%)	5 (3%)	41	44
1	D	156/178 (88%)	153 (98%)	3 (2%)	57	63
All	All	651/712 (91%)	634 (97%)	17 (3%)	46	50

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	89	PHE
1	D	185	GLU
1	B	123	ARG
1	B	184	ARG
1	C	4	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	145	ASN
1	C	153	HIS
1	C	41	HIS
1	B	153	HIS
1	C	145	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

3 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	KEG	C	900	-	72,72,76	95.04	70 (97%)	-		
2	KEG	A	902	-	76,76,76	90.54	75 (98%)	6,234,234	20.64	6 (100%)
2	KEG	D	901	-	76,76,76	95.25	75 (98%)	6,234,234	22.47	5 (83%)

The worst 5 of 220 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	900	KEG	W3-O37	203.04	9.63	1.93
2	C	900	KEG	W8-O41	182.12	8.84	1.93
2	C	900	KEG	W10-O12	171.78	6.19	1.71
2	C	900	KEG	W4-O6	166.38	6.05	1.71
2	D	901	KEG	W10-O12	164.44	6.18	1.71

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	901	KEG	O21-P1-O17	-41.92	28.70	108.92
2	A	902	KEG	O21-P1-O17	-38.49	35.28	108.92
2	D	901	KEG	O21-P1-O19	-22.61	65.67	108.92
2	A	902	KEG	O21-P1-O19	-21.96	66.90	108.92
2	A	902	KEG	O21-P1-O18	21.33	149.74	108.92

There are no chirality outliers.

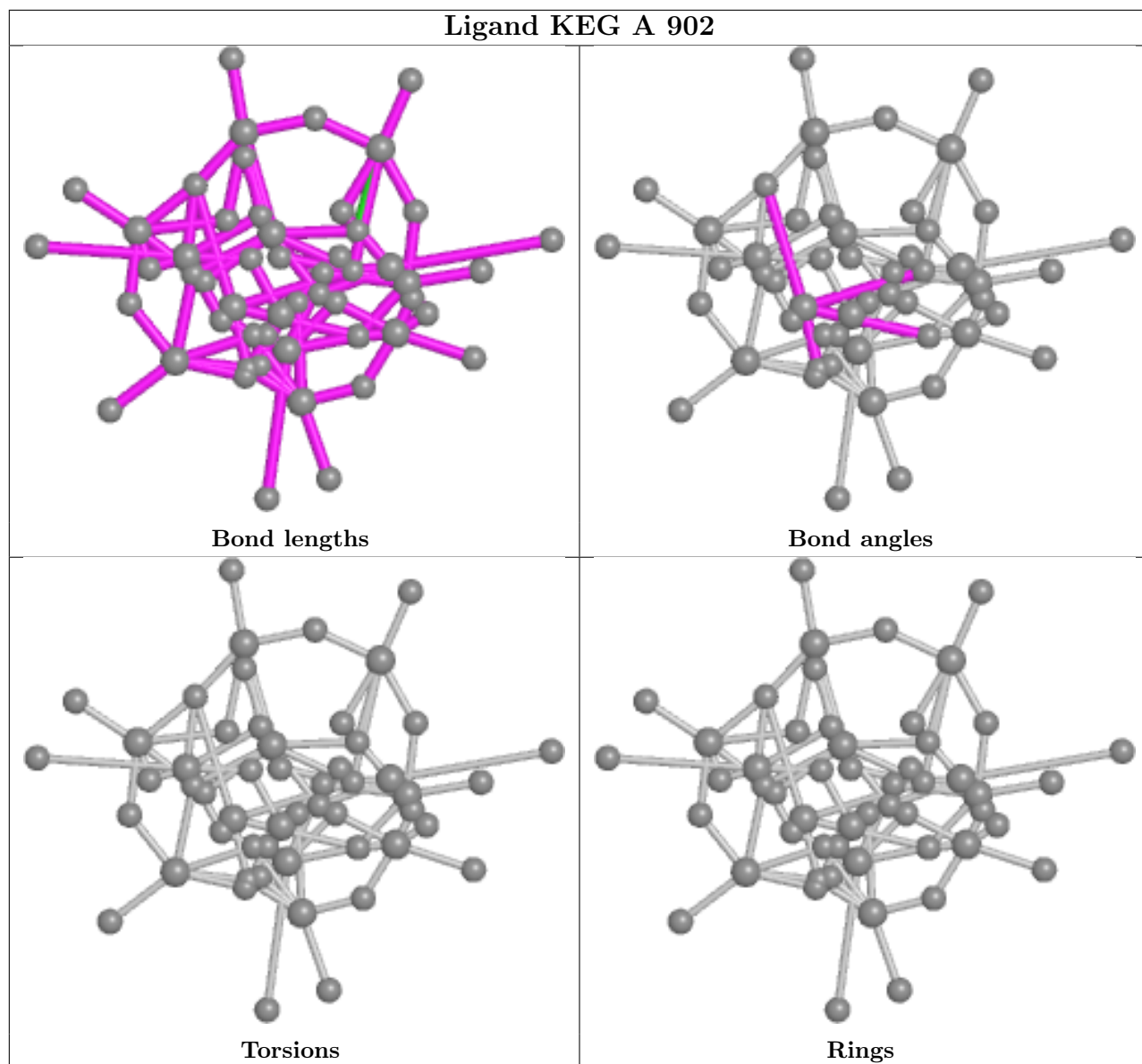There are no torsion outliers.

There are no ring outliers.
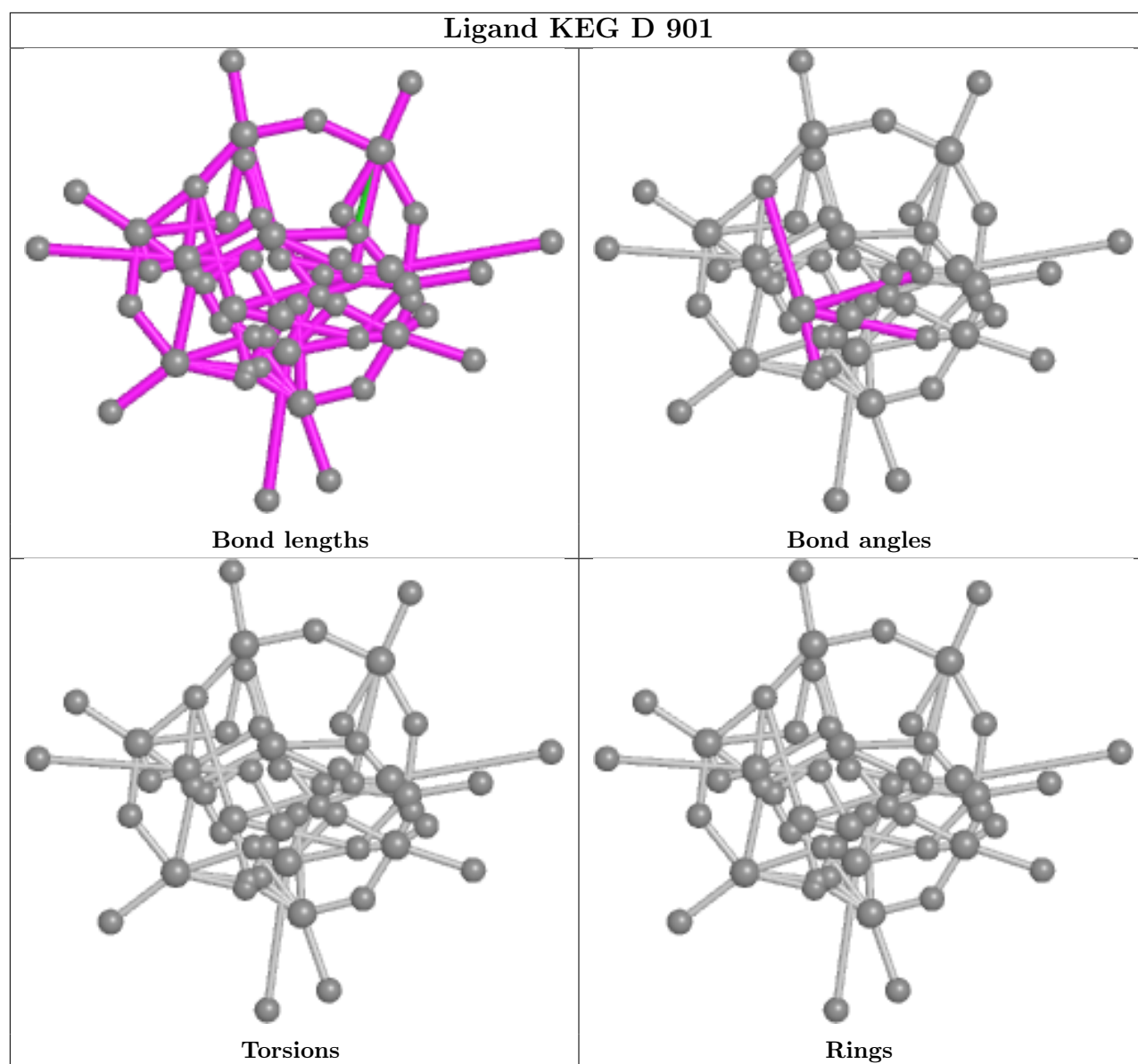
3 monomers are involved in 30 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	900	KEG	8	0
2	A	902	KEG	12	1
2	D	901	KEG	9	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 KEG C 900

Bond lengths

Bond angles

Torsions

Rings

Ligand KEG A 902

Bond lengths

Bond angles

Torsions

Rings

Ligand KEG D 901

Bond lengths

Bond angles

Torsions

Rings
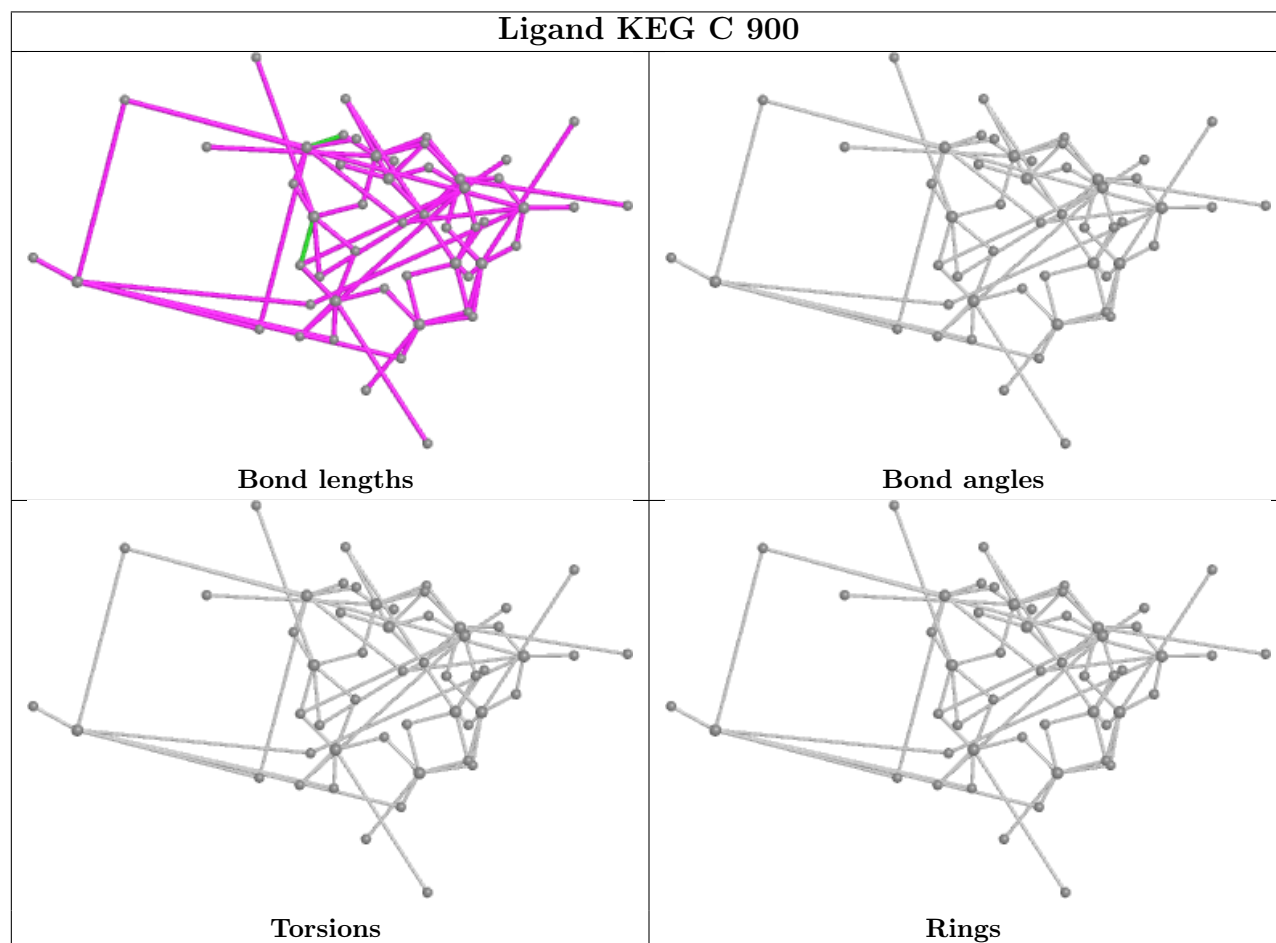
5.7 Other polymers ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	180/195 (92%)	0.07	13 (7%)	15	19	14, 23, 49, 68	0
1	B	179/195 (91%)	-0.05	10 (5%)	24	29	14, 21, 45, 65	0
1	C	184/195 (94%)	0.02	10 (5%)	25	31	13, 21, 51, 66	0
1	D	172/195 (88%)	-0.29	6 (3%)	44	50	13, 20, 36, 52	0
All	All	715/780 (91%)	-0.06	39 (5%)	25	31	13, 21, 45, 68	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	11	SER	9.7
1	C	9	ILE	7.6
1	C	12	PRO	7.5
1	C	4	ARG	6.4
1	A	9	ILE	5.6

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

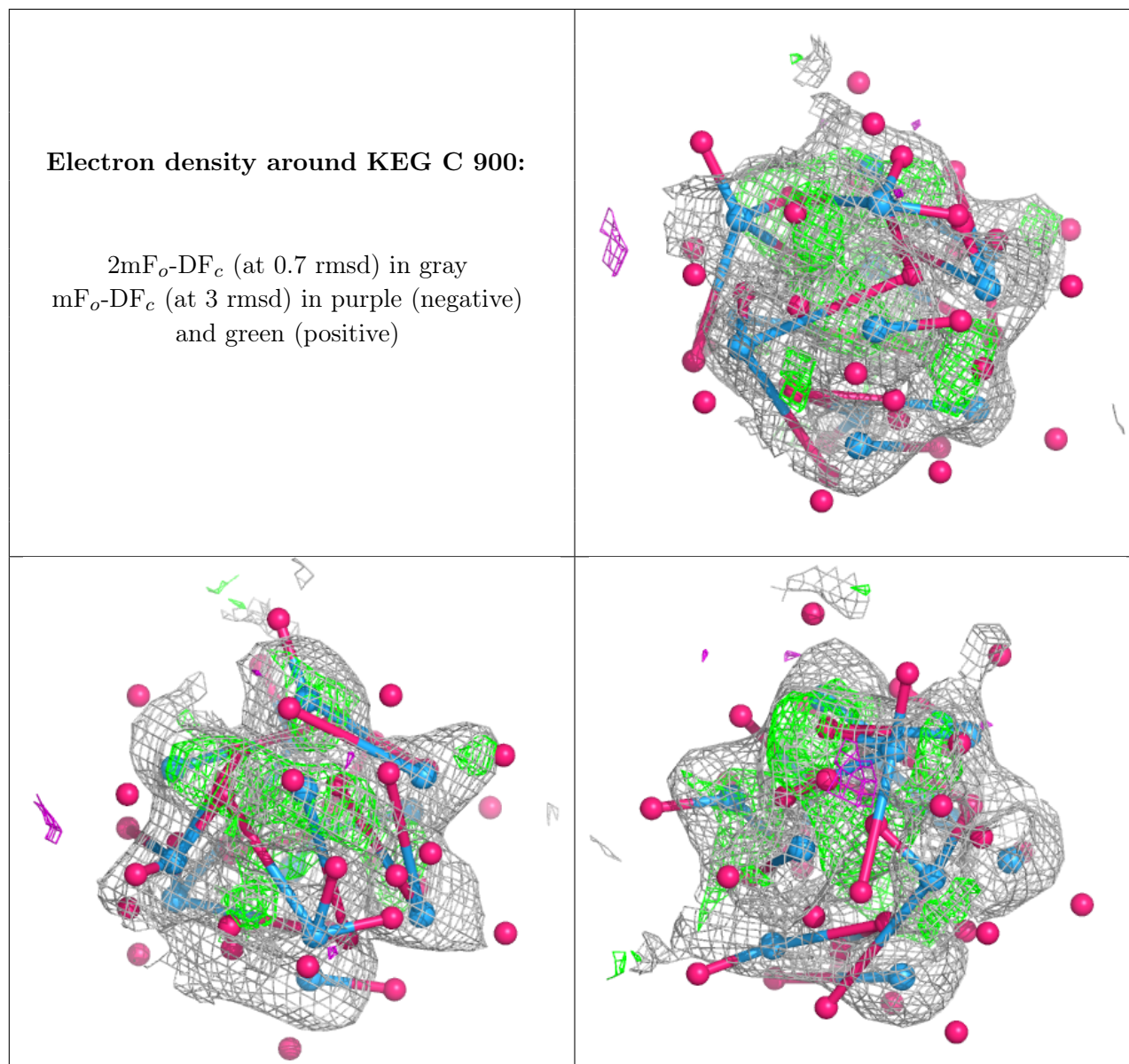There are no non-standard protein/DNA/RNA residues in this entry.

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, 95th 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.
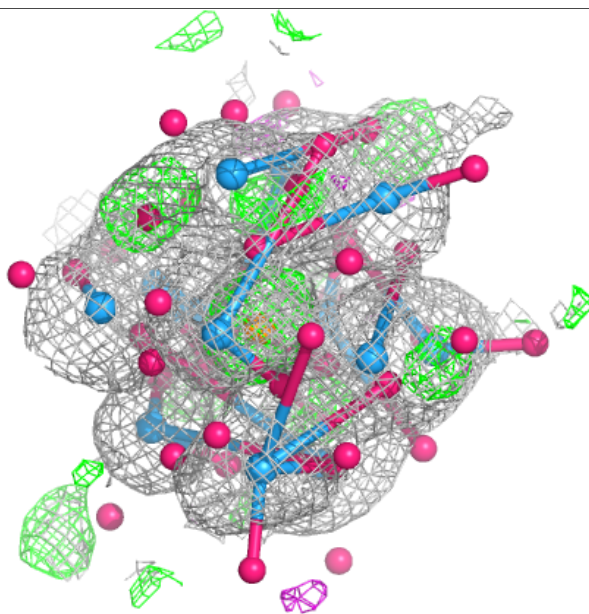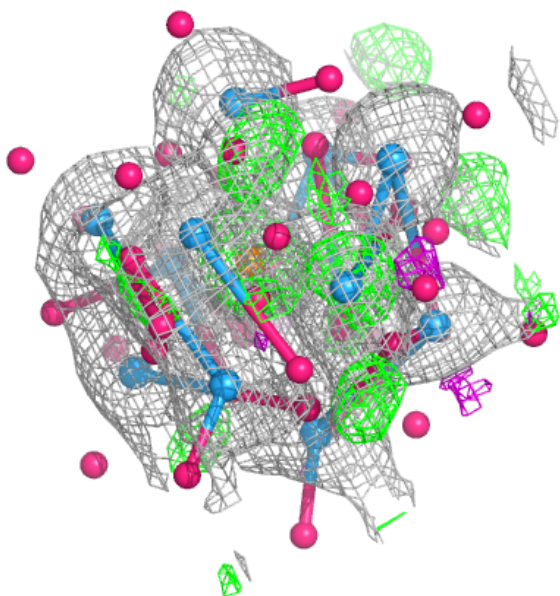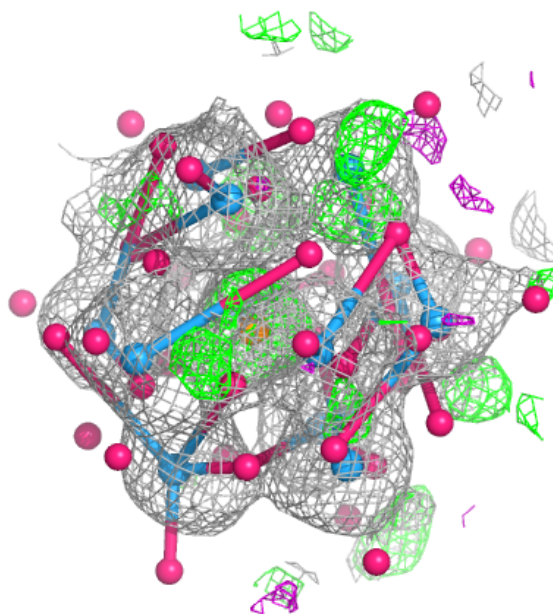
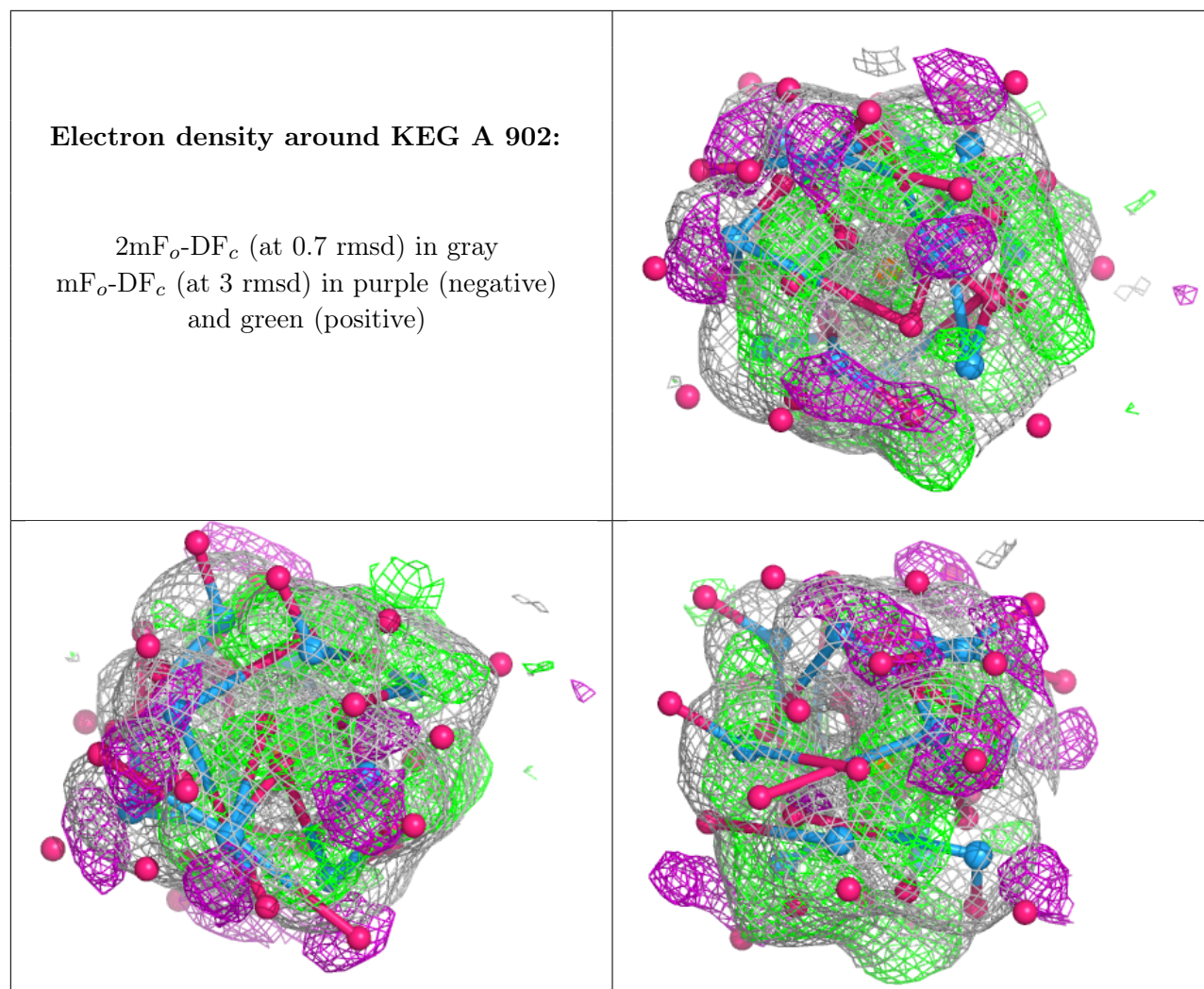| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	KEG	C	900	52/53	0.72	0.35	80,83,84,85	52
2	KEG	D	901	53/53	0.83	0.29	51,64,68,69	53
2	KEG	A	902	53/53	0.92	0.19	41,53,60,63	53

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 KEG C 900:

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

Electron density around KEG D 901:

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

Electron density around KEG A 902:

$2mF_o$-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 ⓘ

There are no such residues in this entry.