wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 02:38 pm BST

PDB ID	:	8S1R
Title	:	Crystal structure of SHANK1 PDZ in complex with a SLiM internal ligand
Authors	:	Li, Y.; Trinh, C.H.; Wilson, A.J.
Deposited on	:	2024-02-15
Resolution	:	1.98 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:

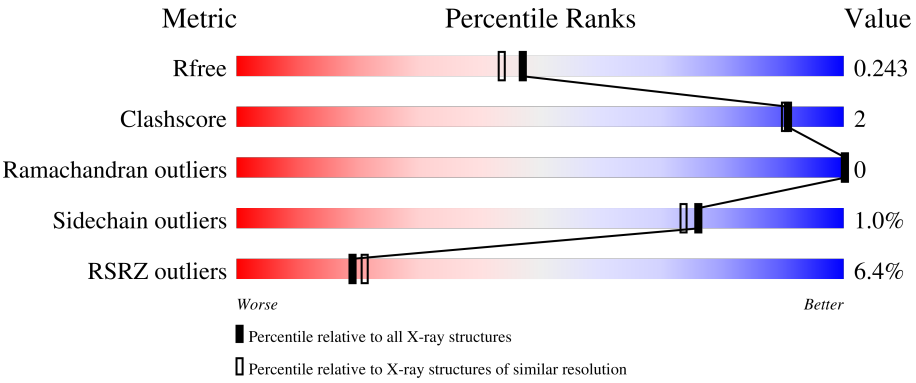
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.98 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))
R_{free}	130704	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	AAA	112	4% 90% 7%
1	BBB	112	2% 90% 7%
1	CCC	112	8% 94% 5%
1	DDD	112	10% 90% 5%
2	EaE	9	100%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	FaF	9	100%
2	GaG	9	11% 100%
2	HaH	9	22% 100%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7380 atoms, of which 3656 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 SH3 and multiple ankyrin repeat domains protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	109	Total	C	H	N	O	S	35	0	0
			1709	537	867	149	152	4			
1	BBB	104	Total	C	H	N	O	S	33	0	0
			1628	512	828	143	141	4			
1	CCC	111	Total	C	H	N	O	S	35	0	0
			1735	546	878	151	156	4			
1	DDD	107	Total	C	H	N	O	S	33	1	0
			1669	526	847	142	150	4			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	651	GLY	-	expression tag	UNP Q9Y566
AAA	652	PRO	-	expression tag	UNP Q9Y566
AAA	653	LEU	-	expression tag	UNP Q9Y566
BBB	651	GLY	-	expression tag	UNP Q9Y566
BBB	652	PRO	-	expression tag	UNP Q9Y566
BBB	653	LEU	-	expression tag	UNP Q9Y566
CCC	651	GLY	-	expression tag	UNP Q9Y566
CCC	652	PRO	-	expression tag	UNP Q9Y566
CCC	653	LEU	-	expression tag	UNP Q9Y566
DDD	651	GLY	-	expression tag	UNP Q9Y566
DDD	652	PRO	-	expression tag	UNP Q9Y566
DDD	653	LEU	-	expression tag	UNP Q9Y566

- Molecule 2 is a protein called F-box only protein 41.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	EaE	9	Total	C	H	N	O	5	0	0
			125	41	56	10	18			
2	FaF	9	Total	C	H	N	O	5	0	0
			125	41	56	10	18			

Continued on next page...

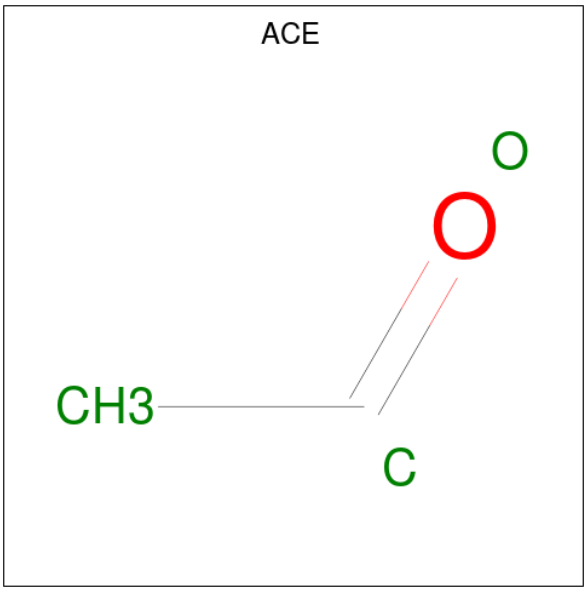
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	GaG	9	Total	C	H	N	O	5	0	0
			125	41	56	10	18			
2	HaH	9	Total	C	H	N	O	5	0	0
			125	41	56	10	18			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
EaE	1	GLU	-	expression tag	UNP Q8TF61
EaE	9	PRO	-	expression tag	UNP Q8TF61
FaF	1	GLU	-	expression tag	UNP Q8TF61
FaF	9	PRO	-	expression tag	UNP Q8TF61
GaG	1	GLU	-	expression tag	UNP Q8TF61
GaG	9	PRO	-	expression tag	UNP Q8TF61
HaH	1	GLU	-	expression tag	UNP Q8TF61
HaH	9	PRO	-	expression tag	UNP Q8TF61

- Molecule 3 is ACETYL GROUP (three-letter code: ACE) (formula: C_2H_4O) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	EaE	1	Total	C	H	O	0	0
			6	2	3	1		
3	FaF	1	Total	C	H	O	0	0
			6	2	3	1		
3	GaG	1	Total	C	H	O	0	0
			6	2	3	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	HaH	1	Total C H O 6 2 3 1	0	0

- Molecule 4 is water.

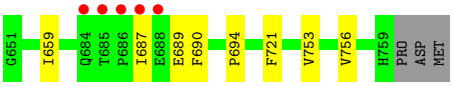
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	31	Total O 31 31	0	0
4	BBB	26	Total O 26 26	0	0
4	CCC	27	Total O 27 27	0	0
4	DDD	9	Total O 9 9	0	0
4	EaE	9	Total O 9 9	0	0
4	FaF	6	Total O 6 6	0	0
4	GaG	6	Total O 6 6	0	0
4	HaH	1	Total O 1 1	0	0

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.

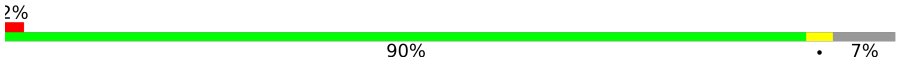
• Molecule 1: SH3 and multiple ankyrin repeat domains protein 1

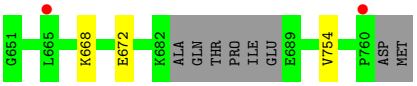
Chain AAA:



• Molecule 1: SH3 and multiple ankyrin repeat domains protein 1

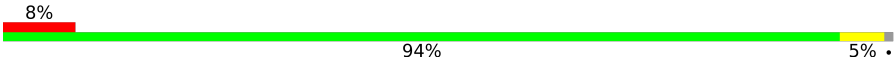
Chain BBB:

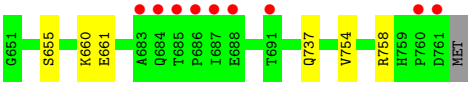




• Molecule 1: SH3 and multiple ankyrin repeat domains protein 1

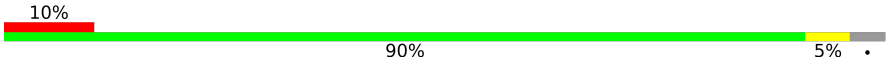
Chain CCC:

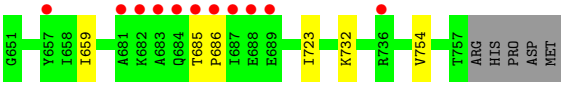




• Molecule 1: SH3 and multiple ankyrin repeat domains protein 1

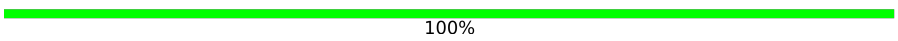
Chain DDD:





• Molecule 2: F-box only protein 41

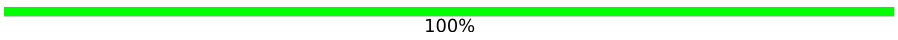
Chain EaE:



There are no outlier residues recorded for this chain.

• Molecule 2: F-box only protein 41

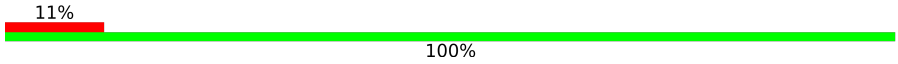
Chain FaF:



There are no outlier residues recorded for this chain.

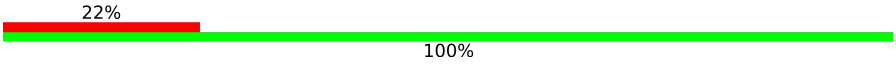
• Molecule 2: F-box only protein 41

Chain GaG:



• Molecule 2: F-box only protein 41

Chain HaH:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	149.10Å 149.10Å 64.07Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	129.12 – 1.98 129.13 – 1.98	Depositor EDS
% Data completeness (in resolution range)	100.0 (129.12-1.98) 100.0 (129.13-1.98)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 1.98Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.216 , 0.236 0.223 , 0.243	Depositor DCC
R_{free} test set	3012 reflections (5.27%)	wwPDB-VP
Wilson B-factor (Å^2)	37.6	Xtriage
Anisotropy	0.323	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 42.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.026 for -h,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7380	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.96% 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: ACE

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	AAA	0.66	0/857	0.77	0/1156
1	BBB	0.66	0/814	0.82	0/1096
1	CCC	0.65	0/873	0.80	0/1179
1	DDD	0.67	0/844	0.80	0/1139
2	EaE	0.66	0/70	0.77	0/92
2	FaF	0.74	0/70	0.78	0/92
2	GaG	0.74	0/70	0.76	0/92
2	HaH	0.78	0/70	0.76	0/92
All	All	0.67	0/3668	0.80	0/4938

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	AAA	842	867	862	5	0
1	BBB	800	828	820	2	0
1	CCC	857	878	873	2	0
1	DDD	822	847	842	2	0
2	EaE	69	56	56	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	FaF	69	56	56	0	0
2	GaG	69	56	56	0	0
2	HaH	69	56	56	0	0
3	EaE	3	3	3	0	0
3	FaF	3	3	3	0	0
3	GaG	3	3	3	0	0
3	HaH	3	3	3	0	0
4	AAA	31	0	0	0	0
4	BBB	26	0	0	0	0
4	CCC	27	0	0	0	0
4	DDD	9	0	0	0	0
4	EaE	9	0	0	0	0
4	FaF	6	0	0	0	0
4	GaG	6	0	0	0	0
4	HaH	1	0	0	0	0
All	All	3724	3656	3633	10	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 2.

The worst 5 of 10 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:AAA:687:ILE:HG23	1:AAA:689:GLU:CD	2.30	0.52
1:AAA:659:ILE:HD12	1:AAA:756:VAL:HG22	1.91	0.52
1:DDD:685:THR:HB	1:DDD:686:PRO:HD2	1.95	0.49
1:CCC:661:GLU:HG2	1:CCC:754:VAL:HG12	1.95	0.47
1:BBB:668:LYS:HB2	1:BBB:672:GLU:HG3	1.97	0.46

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	AAA	107/112 (96%)	105 (98%)	2 (2%)	0	100	100
1	BBB	100/112 (89%)	99 (99%)	1 (1%)	0	100	100
1	CCC	109/112 (97%)	107 (98%)	2 (2%)	0	100	100
1	DDD	106/112 (95%)	103 (97%)	3 (3%)	0	100	100
2	EaE	7/9 (78%)	7 (100%)	0	0	100	100
2	FaF	7/9 (78%)	7 (100%)	0	0	100	100
2	GaG	7/9 (78%)	7 (100%)	0	0	100	100
2	HaH	7/9 (78%)	7 (100%)	0	0	100	100
All	All	450/484 (93%)	442 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	90/93 (97%)	90 (100%)	0	100	100
1	BBB	85/93 (91%)	85 (100%)	0	100	100
1	CCC	92/93 (99%)	90 (98%)	2 (2%)	52	46
1	DDD	89/93 (96%)	87 (98%)	2 (2%)	52	46
2	EaE	8/8 (100%)	8 (100%)	0	100	100
2	FaF	8/8 (100%)	8 (100%)	0	100	100
2	GaG	8/8 (100%)	8 (100%)	0	100	100
2	HaH	8/8 (100%)	8 (100%)	0	100	100
All	All	388/404 (96%)	384 (99%)	4 (1%)	76	73

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	CCC	660	LYS
1	CCC	737	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	DDD	659	ILE
1	DDD	732	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
3	ACE	EaE	101	2	1,2,2	0.08	0	1,1,1	0.56	0
3	ACE	HaH	101	2	1,2,2	0.02	0	1,1,1	0.63	0
3	ACE	FaF	101	2	1,2,2	0.02	0	1,1,1	0.44	0
3	ACE	GaG	101	2	1,2,2	0.36	0	1,1,1	0.47	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

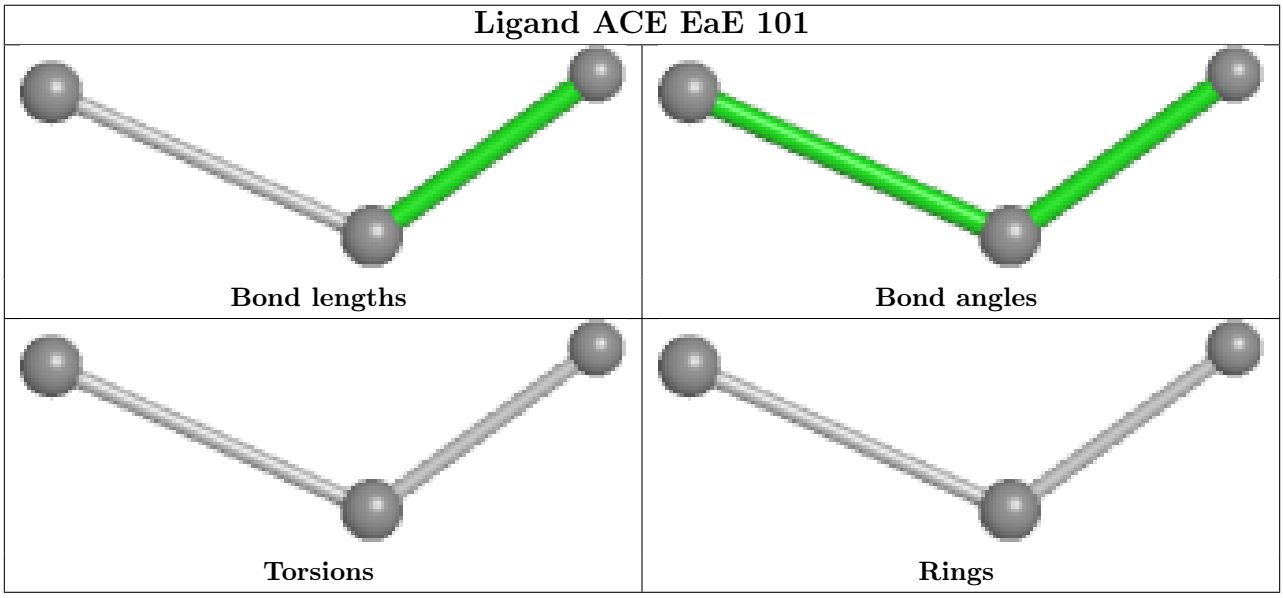
There are no torsion outliers.

There are no ring outliers.

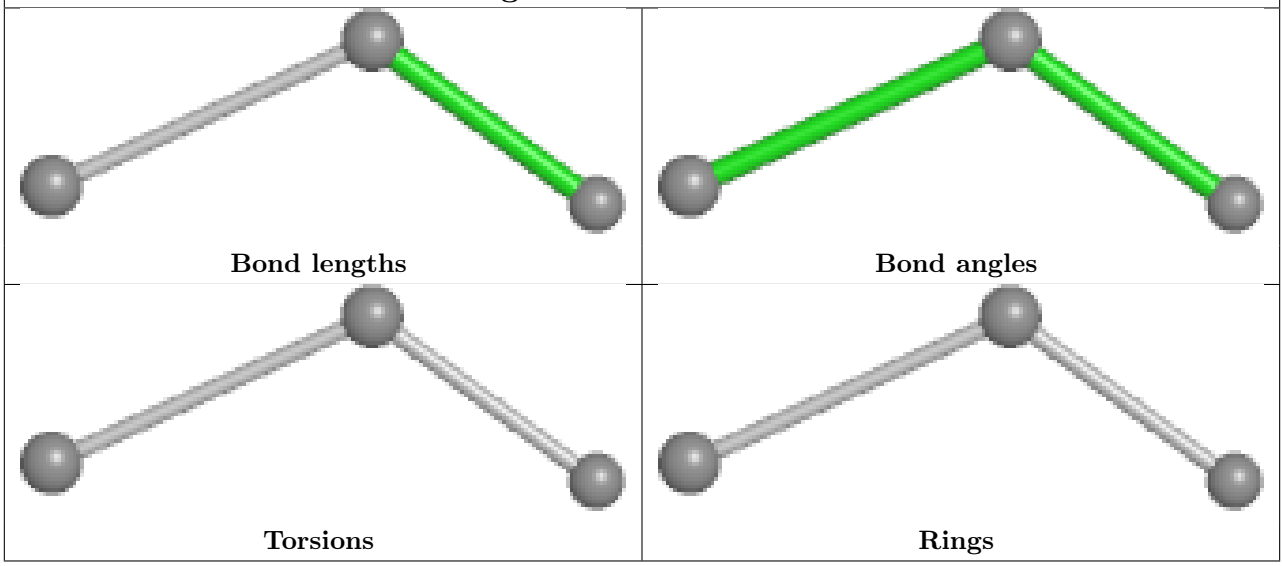
No monomer is involved in short contacts.

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.

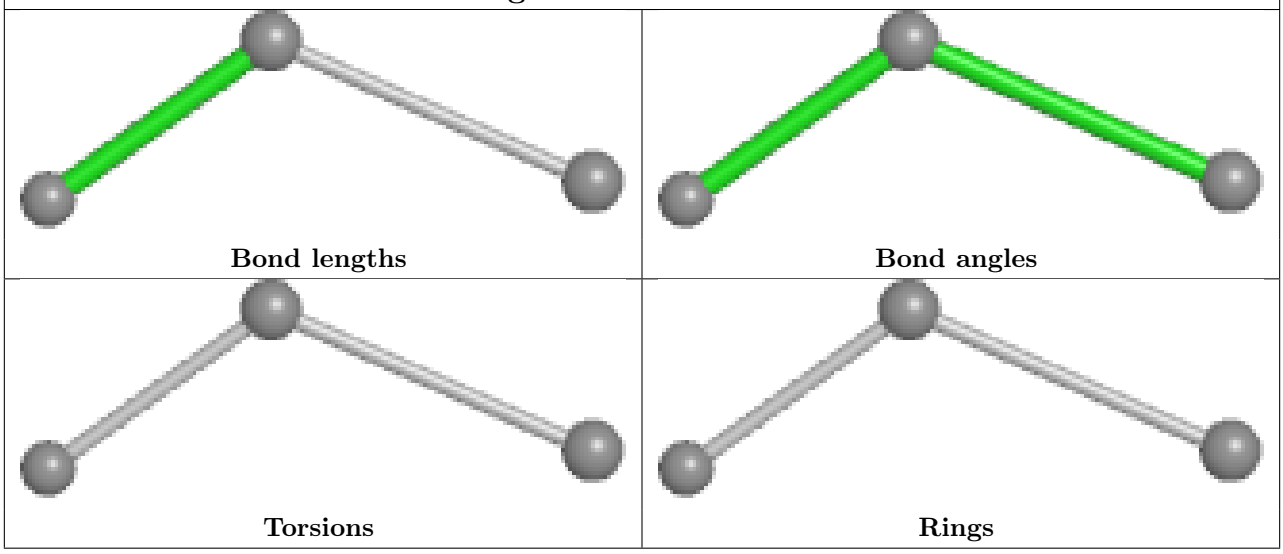
Ligand ACE EaE 101



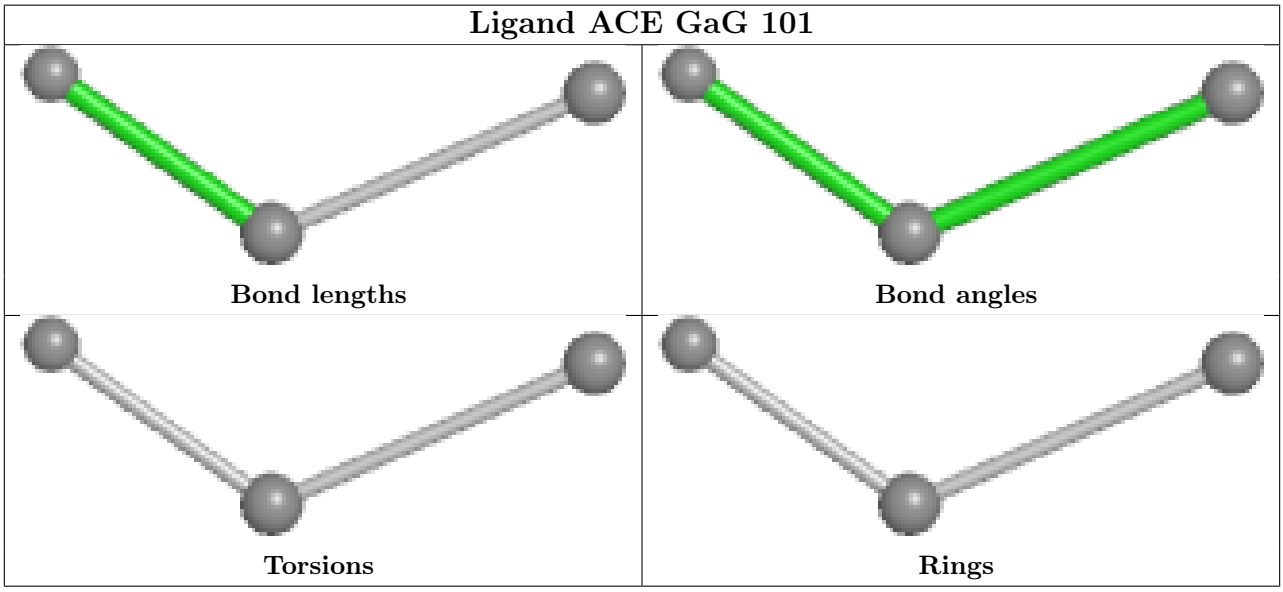
Ligand ACE HaH 101



Ligand ACE FaF 101



Ligand ACE GaG 101



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

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	AAA	109/112 (97%)	0.77	5 (4%) 32 34	31, 40, 87, 112	0
1	BBB	104/112 (92%)	0.67	2 (1%) 66 68	31, 43, 64, 94	0
1	CCC	111/112 (99%)	0.83	9 (8%) 12 13	30, 44, 99, 113	0
1	DDD	107/112 (95%)	1.05	11 (10%) 6 7	37, 54, 107, 130	0
2	EaE	9/9 (100%)	0.41	0 100 100	34, 39, 52, 55	0
2	FaF	9/9 (100%)	0.44	0 100 100	34, 37, 49, 60	0
2	GaG	9/9 (100%)	0.47	1 (11%) 5 6	35, 39, 55, 67	0
2	HaH	9/9 (100%)	1.16	2 (22%) 0 0	47, 55, 83, 85	0
All	All	467/484 (96%)	0.82	30 (6%) 19 21	30, 45, 90, 130	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	DDD	685	THR	6.9
1	DDD	686	PRO	5.9
1	DDD	684	GLN	5.4
1	AAA	684	GLN	5.2
1	CCC	685	THR	5.0

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 (i)

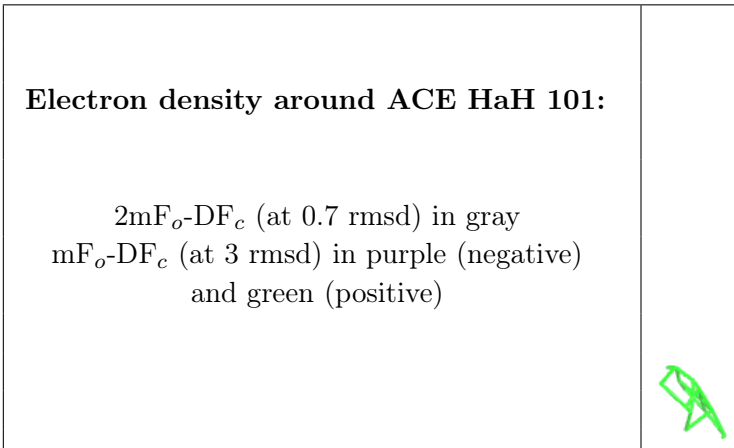
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

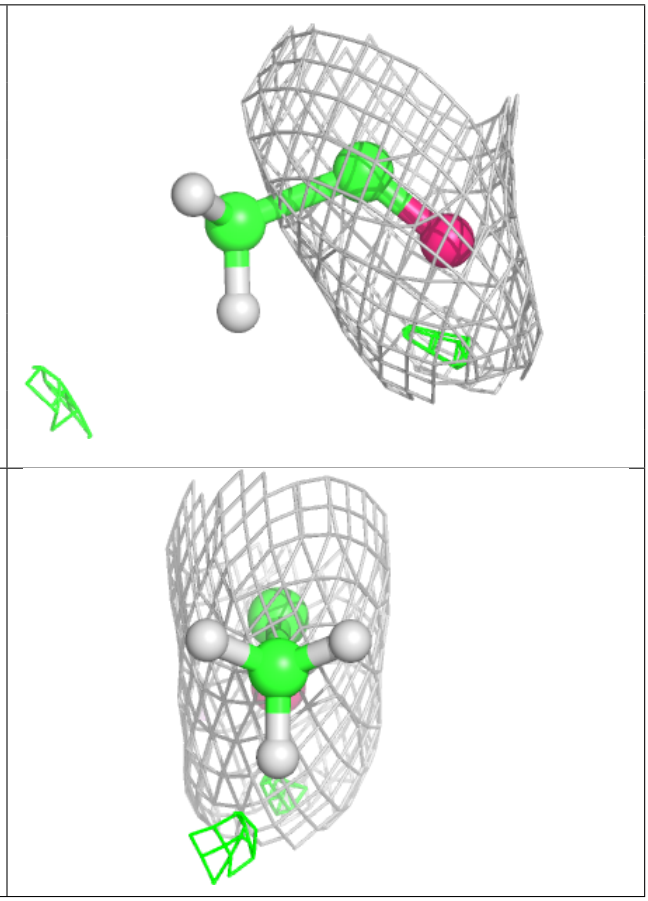
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	ACE	HaH	101	3/3	0.74	0.32	66,76,76,76	0
3	ACE	GaG	101	3/3	0.89	0.25	55,58,58,58	0
3	ACE	EaE	101	3/3	0.94	0.17	50,51,51,53	0
3	ACE	FaF	101	3/3	0.97	0.18	42,42,42,43	0

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 ACE HaH 101:

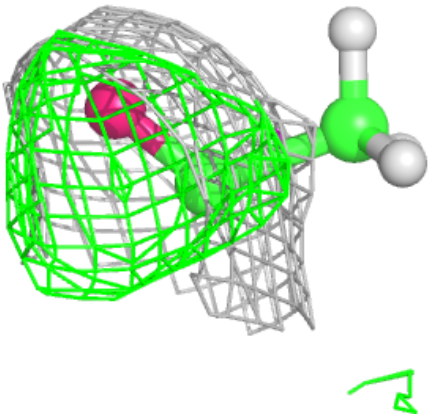
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

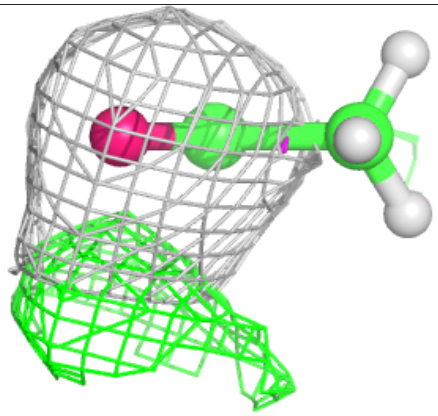


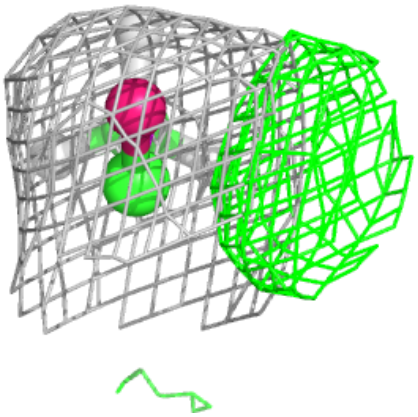


Electron density around ACE GaG 101:

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

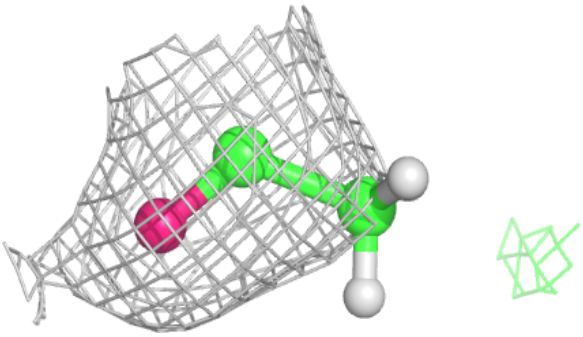


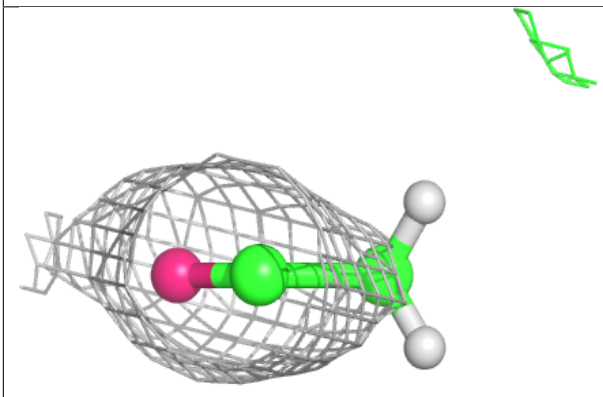


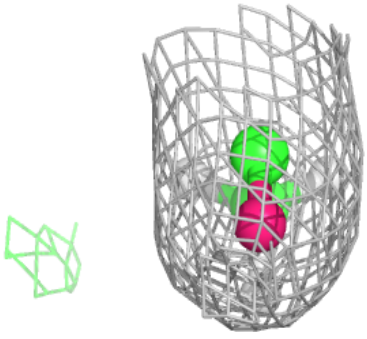


Electron density around ACE EaE 101:

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

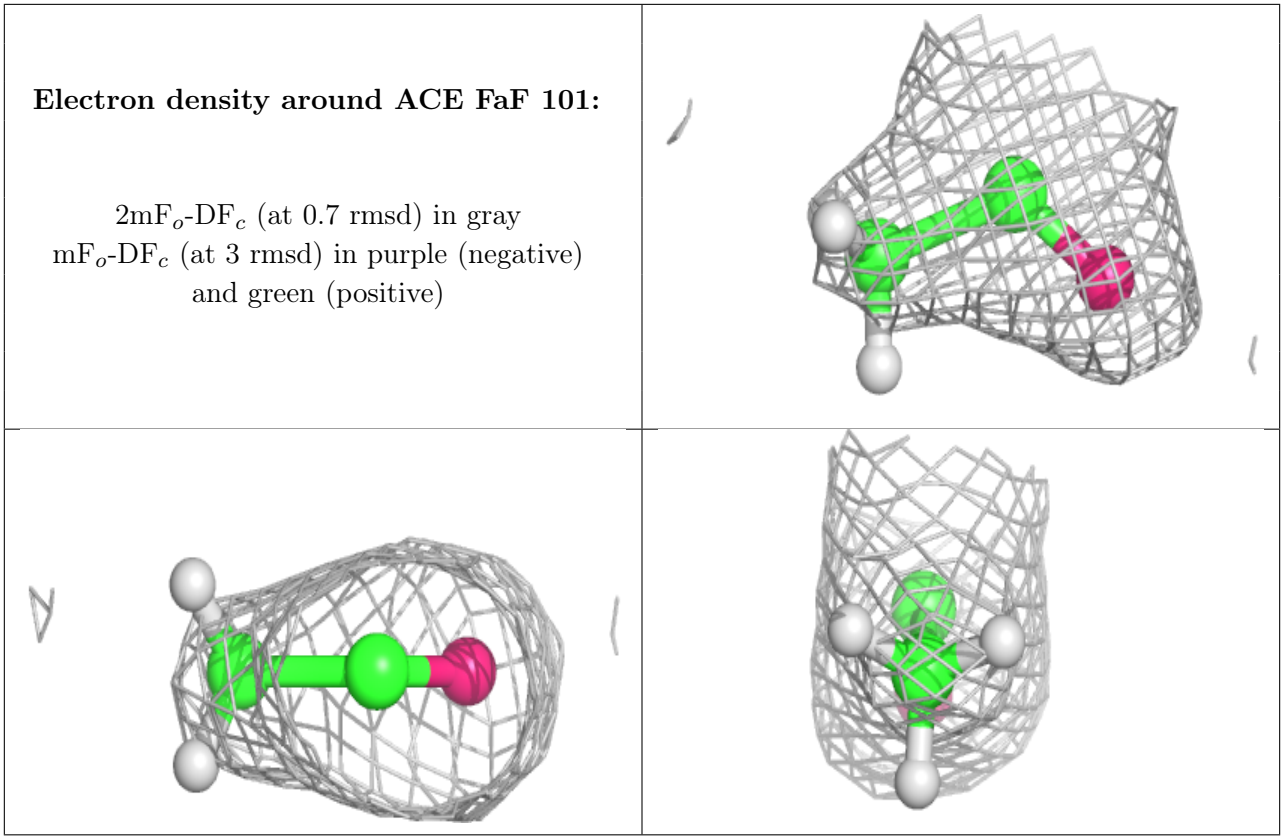






Electron density around ACE FaF 101:

$2mF_o$-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.