wwPDB X-ray Structure Validation Summary Report (i)

Nov 6, 2023 – 01:49 AM EST

PDB ID	:	6O4G
Title	:	Structure of ALDH7A1 mutant P169S complexed with alpha-aminoadipate
Authors	:	Tanner, J.J.; Korasick, D.A.; Laciak, A.R.
Deposited on	:	2019-02-28
Resolution	:	2.05 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:

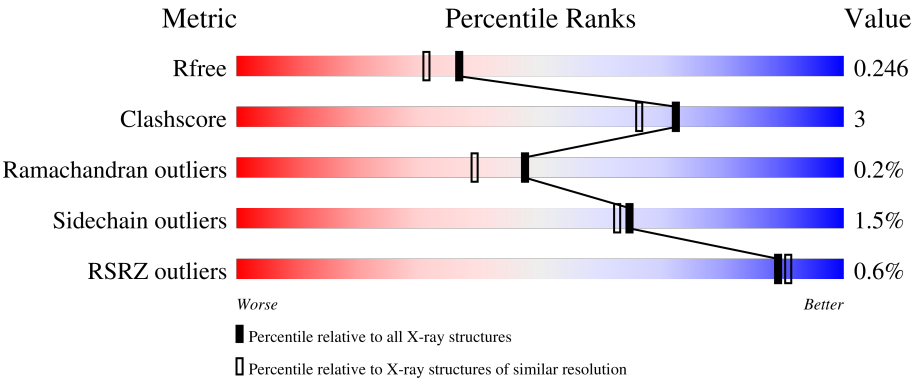
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.05 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.



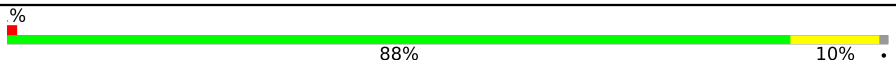
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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.

Mol	Chain	Length	Quality of chain
1	A	513	93% 6% ·
1	B	513	89% 10% ·
1	C	513	% 88% 11% ·
1	D	513	89% 10% ·
1	E	513	2% 88% 11% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	513	
1	G	513	
1	H	513	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 31763 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 Alpha-aminoadipic semialdehyde dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	509	Total	C	N	O	S	0	0	0
			3839	2436	665	721	17			
1	B	508	Total	C	N	O	S	0	1	0
			3823	2429	661	715	18			
1	C	508	Total	C	N	O	S	0	0	0
			3823	2427	665	714	17			
1	D	508	Total	C	N	O	S	0	0	0
			3826	2430	661	718	17			
1	E	508	Total	C	N	O	S	0	15	0
			3936	2500	682	736	18			
1	F	508	Total	C	N	O	S	0	0	0
			3821	2422	664	718	17			
1	G	508	Total	C	N	O	S	0	0	0
			3845	2444	671	713	17			
1	H	509	Total	C	N	O	S	0	1	0
			3854	2444	669	723	18			

There are 24 discrepancies between the modelled and reference sequences:

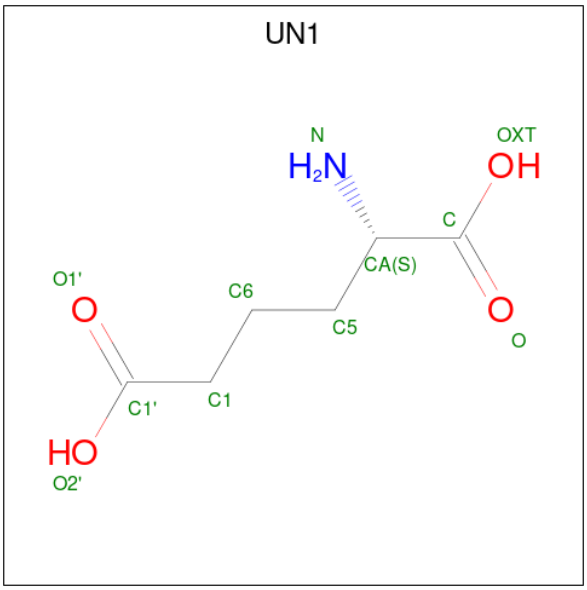
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P49419
A	0	HIS	-	expression tag	UNP P49419
A	169	SER	PRO	engineered mutation	UNP P49419
B	-1	GLY	-	expression tag	UNP P49419
B	0	HIS	-	expression tag	UNP P49419
B	169	SER	PRO	engineered mutation	UNP P49419
C	-1	GLY	-	expression tag	UNP P49419
C	0	HIS	-	expression tag	UNP P49419
C	169	SER	PRO	engineered mutation	UNP P49419
D	-1	GLY	-	expression tag	UNP P49419
D	0	HIS	-	expression tag	UNP P49419
D	169	SER	PRO	engineered mutation	UNP P49419
E	-1	GLY	-	expression tag	UNP P49419

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	0	HIS	-	expression tag	UNP P49419
E	169	SER	PRO	engineered mutation	UNP P49419
F	-1	GLY	-	expression tag	UNP P49419
F	0	HIS	-	expression tag	UNP P49419
F	169	SER	PRO	engineered mutation	UNP P49419
G	-1	GLY	-	expression tag	UNP P49419
G	0	HIS	-	expression tag	UNP P49419
G	169	SER	PRO	engineered mutation	UNP P49419
H	-1	GLY	-	expression tag	UNP P49419
H	0	HIS	-	expression tag	UNP P49419
H	169	SER	PRO	engineered mutation	UNP P49419

- Molecule 2 is 2-AMINOHEXANEDIOIC ACID (three-letter code: UN1) (formula: $C_6H_{11}NO_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	N	O	0	0
			11	6	1	4		
2	E	1	Total	C	N	O	0	0
			11	6	1	4		
2	G	1	Total	C	N	O	0	0
			11	6	1	4		
2	H	1	Total	C	N	O	0	0
			11	6	1	4		

- Molecule 3 is water.

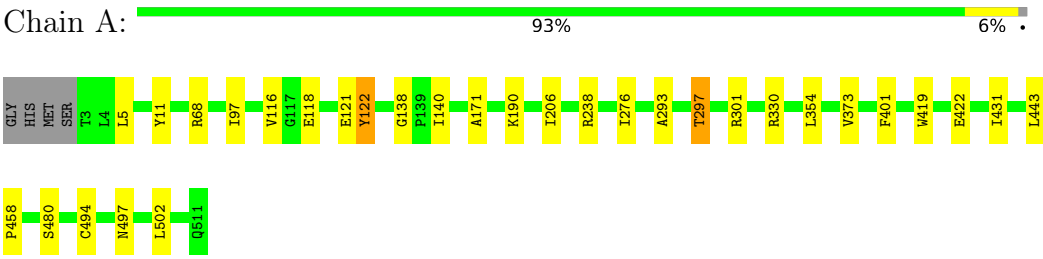
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	145	Total O 145 145	0	0
3	B	120	Total O 120 120	0	0
3	C	118	Total O 118 118	0	0
3	D	106	Total O 106 106	0	0
3	E	121	Total O 121 121	0	0
3	F	111	Total O 111 111	0	0
3	G	109	Total O 109 109	0	0
3	H	121	Total O 122 122	0	1

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.

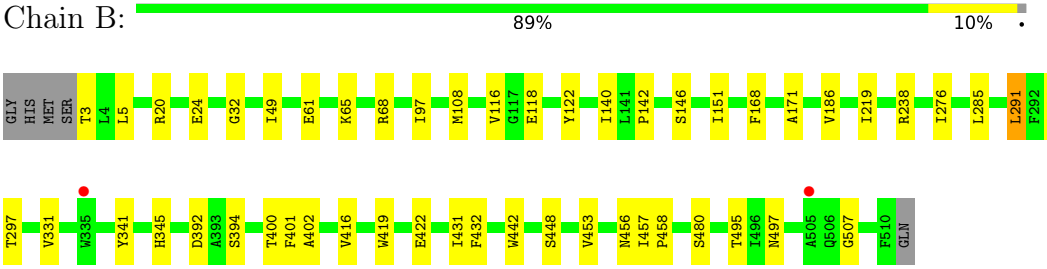
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain A:



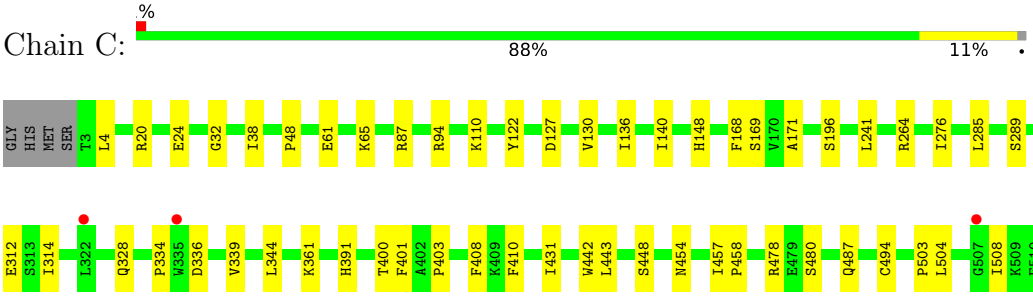
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain B:



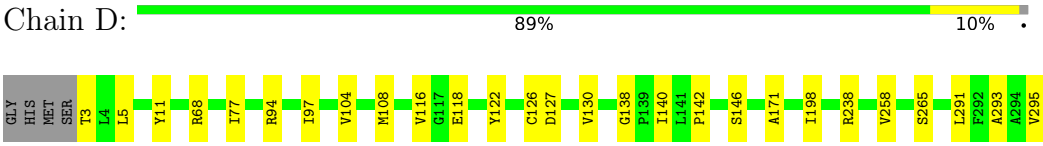
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

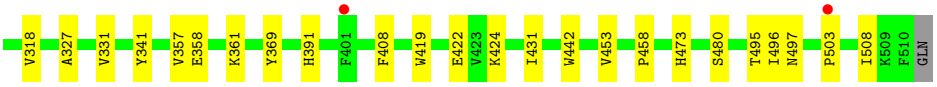
Chain C:



• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

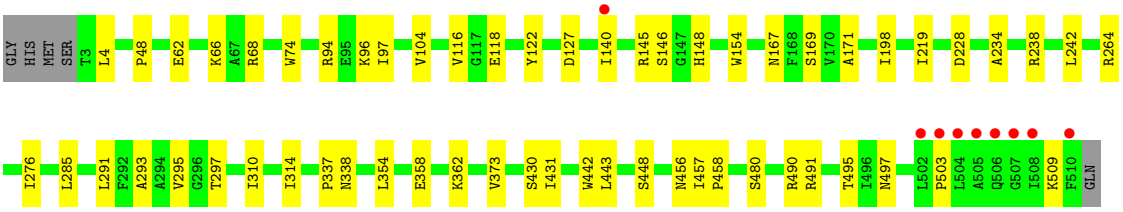
Chain D:





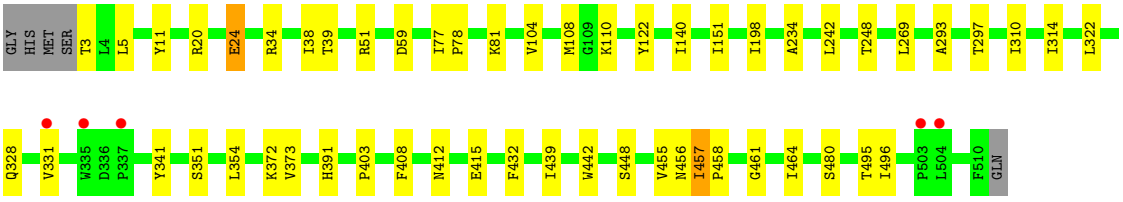
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain E:



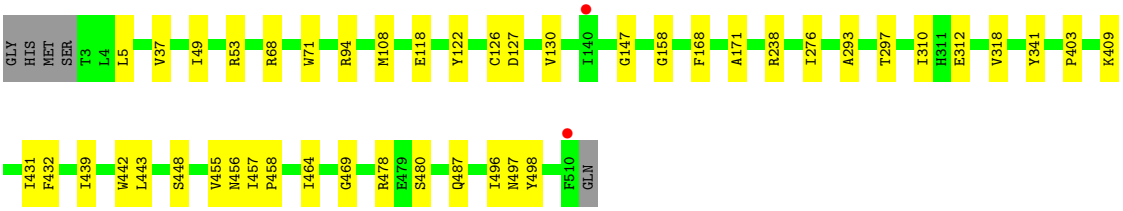
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain F:



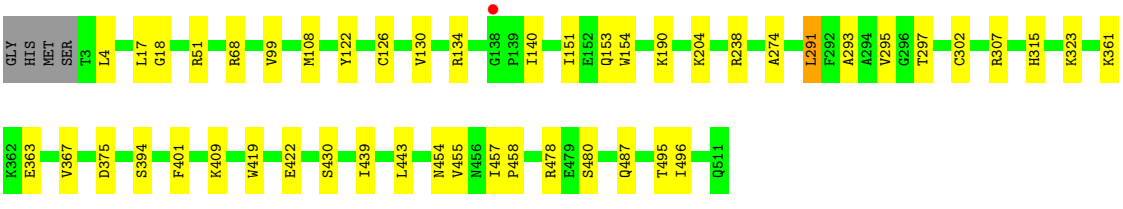
• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain G:



• Molecule 1: Alpha-aminoadipic semialdehyde dehydrogenase

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	155.30Å 160.03Å 158.26Å 90.00° 95.14° 90.00°	Depositor
Resolution (Å)	49.07 – 2.05 49.07 – 2.05	Depositor EDS
% Data completeness (in resolution range)	97.0 (49.07-2.05) 99.0 (49.07-2.05)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 2.05Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.192 , 0.250 0.187 , 0.246	Depositor DCC
R_{free} test set	12030 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.469	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	31763	wwPDB-VP
Average B, all atoms (Å^2)	37.0	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 21.04 % 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 7.6281e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: UN1

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	0.39	0/3917	0.55	0/5320
1	B	0.38	0/3905	0.54	0/5302
1	C	0.40	0/3901	0.55	0/5298
1	D	0.37	0/3904	0.52	0/5301
1	E	0.38	0/4023	0.54	0/5461
1	F	0.39	0/3898	0.54	0/5292
1	G	0.39	0/3923	0.53	0/5321
1	H	0.38	0/3936	0.53	0/5342
All	All	0.38	0/31407	0.54	0/42637

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	3839	0	3797	22	0
1	B	3823	0	3781	33	0
1	C	3823	0	3781	37	0
1	D	3826	0	3787	31	0
1	E	3936	0	3889	34	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	3821	0	3775	32	0
1	G	3845	0	3845	25	0
1	H	3854	0	3819	26	0
2	B	11	0	4	1	0
2	E	11	0	4	0	0
2	G	11	0	4	1	0
2	H	11	0	4	1	0
3	A	145	0	0	0	0
3	B	120	0	0	0	0
3	C	118	0	0	1	0
3	D	106	0	0	0	0
3	E	121	0	0	0	0
3	F	111	0	0	0	0
3	G	109	0	0	0	0
3	H	122	0	0	0	0
All	All	31763	0	30490	207	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 3.

The worst 5 of 207 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:E:293:ALA:HB2	1:E:458:PRO:HB3	1.53	0.90
1:B:118:GLU:HG3	1:B:171:ALA:HB2	1.69	0.72
1:E:94:ARG:NH2	1:E:127:ASP:OD2	2.20	0.72
1:C:293:ALA:HB2	1:C:458:PRO:HB3	1.72	0.71
1:C:289:SER:HB3	1:C:458:PRO:HG3	1.70	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	507/513 (99%)	486 (96%)	20 (4%)	1 (0%)	47	39
1	B	507/513 (99%)	490 (97%)	15 (3%)	2 (0%)	34	24
1	C	506/513 (99%)	487 (96%)	18 (4%)	1 (0%)	47	39
1	D	506/513 (99%)	490 (97%)	15 (3%)	1 (0%)	47	39
1	E	520/513 (101%)	499 (96%)	20 (4%)	1 (0%)	47	39
1	F	506/513 (99%)	488 (96%)	17 (3%)	1 (0%)	47	39
1	G	506/513 (99%)	488 (96%)	17 (3%)	1 (0%)	47	39
1	H	508/513 (99%)	487 (96%)	20 (4%)	1 (0%)	47	39
All	All	4066/4104 (99%)	3915 (96%)	142 (4%)	9 (0%)	47	39

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	480	SER
1	G	480	SER
1	A	480	SER
1	B	401	PHE
1	B	480	SER

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	397/410 (97%)	392 (99%)	5 (1%)	69	67
1	B	394/410 (96%)	388 (98%)	6 (2%)	65	62
1	C	393/410 (96%)	389 (99%)	4 (1%)	76	75
1	D	395/410 (96%)	388 (98%)	7 (2%)	59	55
1	E	406/410 (99%)	401 (99%)	5 (1%)	71	70
1	F	393/410 (96%)	385 (98%)	8 (2%)	55	50
1	G	399/410 (97%)	395 (99%)	4 (1%)	76	75
1	H	400/410 (98%)	392 (98%)	8 (2%)	55	50

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3177/3280 (97%)	3130 (98%)	47 (2%)	65 62

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	24	GLU
1	G	108	MET
1	F	38	ILE
1	F	122	TYR
1	G	297	THR

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	$\#\|Z\|>2$
2	UN1	E	601	-	9,10,10	0.97	0	11,12,12	1.48	4 (36%)
2	UN1	B	601	-	9,10,10	0.97	0	11,12,12	1.30	0
2	UN1	G	601	-	9,10,10	1.05	0	11,12,12	1.49	3 (27%)
2	UN1	H	601	-	9,10,10	1.03	0	11,12,12	1.31	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	UN1	E	601	-	-	0/10/10/10	-
2	UN1	B	601	-	-	0/10/10/10	-
2	UN1	G	601	-	-	2/10/10/10	-
2	UN1	H	601	-	-	2/10/10/10	-

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	601	UN1	OXT-C-CA	2.32	121.27	113.38
2	G	601	UN1	OXT-C-O	-2.27	118.92	124.09
2	E	601	UN1	C6-C1-C1'	-2.25	108.79	114.47
2	G	601	UN1	O2'-C1'-C1	2.09	120.74	114.03
2	E	601	UN1	O2'-C1'-C1	2.07	120.70	114.03

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	G	601	UN1	OXT-C-CA-N
2	G	601	UN1	O-C-CA-N
2	H	601	UN1	C6-C1-C1'-O2'
2	H	601	UN1	C6-C1-C1'-O1'

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	UN1	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	G	601	UN1	1	0
2	H	601	UN1	1	0

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	A	509/513 (99%)	-0.27	0 100 100	23, 34, 49, 72	0
1	B	508/513 (99%)	-0.28	2 (0%) 92 93	24, 34, 52, 79	0
1	C	508/513 (99%)	-0.19	3 (0%) 89 91	21, 35, 62, 79	0
1	D	508/513 (99%)	-0.16	2 (0%) 92 93	24, 38, 59, 78	0
1	E	508/513 (99%)	-0.27	9 (1%) 68 71	26, 36, 51, 66	0
1	F	508/513 (99%)	-0.16	5 (0%) 82 84	27, 36, 56, 79	0
1	G	508/513 (99%)	-0.28	2 (0%) 92 93	24, 36, 52, 89	0
1	H	509/513 (99%)	-0.18	1 (0%) 95 95	24, 36, 56, 73	0
All	All	4066/4104 (99%)	-0.22	24 (0%) 89 91	21, 36, 55, 89	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	504	LEU	4.9
1	E	505[A]	ALA	4.6
1	E	508[A]	ILE	4.2
1	E	503[A]	PRO	3.9
1	E	504[A]	LEU	3.7

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
2	UN1	E	601	11/11	0.91	0.19	53,56,62,63	0
2	UN1	G	601	11/11	0.91	0.16	45,51,55,56	0
2	UN1	H	601	11/11	0.93	0.12	42,51,61,62	0
2	UN1	B	601	11/11	0.96	0.11	37,44,49,49	0

6.5 Other polymers (i)

There are no such residues in this entry.