W O R L D W I D E
wwPDB
P R O T E I N D A T A B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 31, 2016 – 04:25 PM EDT

PDB ID	:	5DO7
Title	:	Crystal Structure of the Human Sterol Transporter ABCG5/ABCG8
Authors	:	Lee, J.-Y.; Kinch, L.N.; Borek, D.M.; Urbatsch, I.L.; Xie, X.-S.; Grishin, N.V.; Cohen, J.C.; Otwinowski, Z.; Hobbs, H.H.; Rosenbaum, D.M.
Deposited on	:	2015-09-10
Resolution	:	3.93 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

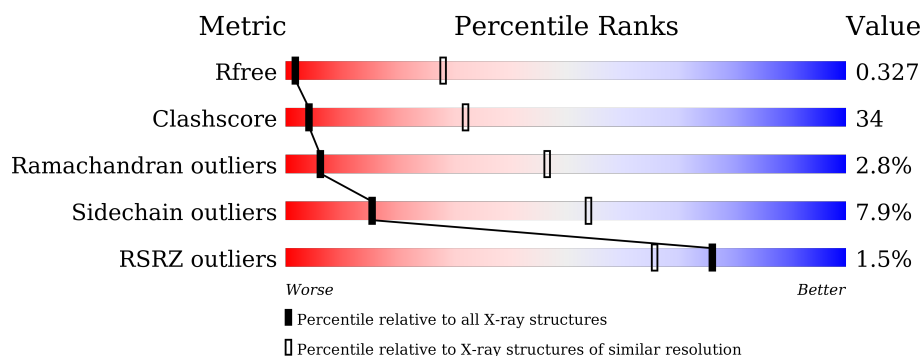|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027674
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

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 3.93 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		0.327
Clashscore		34
Ramachandran outliers		2.8%
Sidechain outliers		7.9%
RSRZ outliers		1.5%

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}	91344	1007 (4.34-3.54)
Clashscore	102246	1042 (4.30-3.58)
Ramachandran outliers	100387	1000 (4.30-3.58)
Sidechain outliers	100360	1021 (4.32-3.56)
RSRZ outliers	91569	1011 (4.34-3.54)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	666	.% / 39% 44% 5% • 12%
1	C	666	.% / 42% 41% • 13%
2	B	685	2% / 40% 36% 5% • 18%
2	D	685	.% / 38% 39% 6% • 17%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 18150 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 ATP-binding cassette sub-family G member 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	585	4606	2982	780	816	28	0	0	0
1	C	579	Total	C	N	O	S	0	0	0
			4551	2945	770	807	29			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLU	GLY	engineered mutation	UNP Q9H222
A	652	GLY	-	expression tag	UNP Q9H222
A	653	SER	-	expression tag	UNP Q9H222
A	654	HIS	-	expression tag	UNP Q9H222
A	655	HIS	-	expression tag	UNP Q9H222
A	656	HIS	-	expression tag	UNP Q9H222
A	657	HIS	-	expression tag	UNP Q9H222
A	658	HIS	-	expression tag	UNP Q9H222
A	659	HIS	-	expression tag	UNP Q9H222
A	660	GLY	-	expression tag	UNP Q9H222
A	661	HIS	-	expression tag	UNP Q9H222
A	662	HIS	-	expression tag	UNP Q9H222
A	663	HIS	-	expression tag	UNP Q9H222
A	664	HIS	-	expression tag	UNP Q9H222
A	665	HIS	-	expression tag	UNP Q9H222
A	666	HIS	-	expression tag	UNP Q9H222
C	2	GLU	GLY	engineered mutation	UNP Q9H222
C	652	GLY	-	expression tag	UNP Q9H222
C	653	SER	-	expression tag	UNP Q9H222
C	654	HIS	-	expression tag	UNP Q9H222
C	655	HIS	-	expression tag	UNP Q9H222
C	656	HIS	-	expression tag	UNP Q9H222
C	657	HIS	-	expression tag	UNP Q9H222
C	658	HIS	-	expression tag	UNP Q9H222
C	659	HIS	-	expression tag	UNP Q9H222

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	660	GLY	-	expression tag	UNP Q9H222
C	661	HIS	-	expression tag	UNP Q9H222
C	662	HIS	-	expression tag	UNP Q9H222
C	663	HIS	-	expression tag	UNP Q9H222
C	664	HIS	-	expression tag	UNP Q9H222
C	665	HIS	-	expression tag	UNP Q9H222
C	666	HIS	-	expression tag	UNP Q9H222

- Molecule 2 is a protein called ATP-binding cassette sub-family G member 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	562	Total 4464	C 2905	N 756	O 775	S 28	0	0	0
2	D	571	Total 4529	C 2936	N 775	O 789	S 29	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	MET	-	initiating methionine	UNP Q9H221
B	0	GLY	-	expression tag	UNP Q9H221
B	1	SER	-	expression tag	UNP Q9H221
B	674	ALA	-	expression tag	UNP Q9H221
B	675	SER	-	expression tag	UNP Q9H221
B	676	ASN	-	expression tag	UNP Q9H221
B	677	SER	-	expression tag	UNP Q9H221
B	678	LEU	-	expression tag	UNP Q9H221
B	679	GLU	-	expression tag	UNP Q9H221
B	680	VAL	-	expression tag	UNP Q9H221
B	681	LEU	-	expression tag	UNP Q9H221
B	682	PHE	-	expression tag	UNP Q9H221
B	683	GLN	-	expression tag	UNP Q9H221
D	-1	MET	-	initiating methionine	UNP Q9H221
D	0	GLY	-	expression tag	UNP Q9H221
D	1	SER	-	expression tag	UNP Q9H221
D	674	ALA	-	expression tag	UNP Q9H221
D	675	SER	-	expression tag	UNP Q9H221
D	676	ASN	-	expression tag	UNP Q9H221
D	677	SER	-	expression tag	UNP Q9H221
D	678	LEU	-	expression tag	UNP Q9H221
D	679	GLU	-	expression tag	UNP Q9H221
D	680	VAL	-	expression tag	UNP Q9H221

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
D	681	LEU	-	expression tag	UNP Q9H221
D	682	PHE	-	expression tag	UNP Q9H221
D	683	GLN	-	expression tag	UNP Q9H221

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

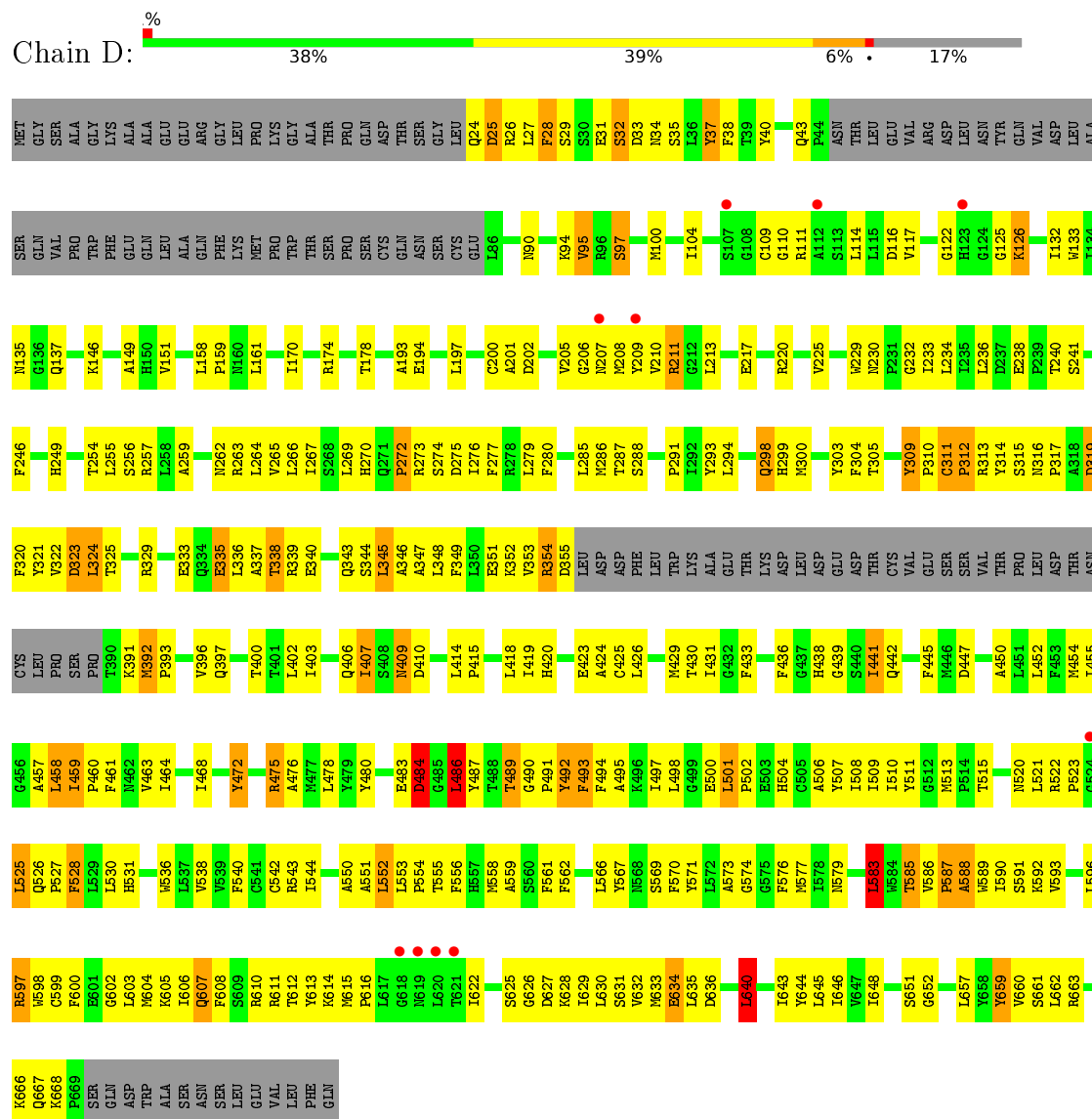- Molecule 1: ATP-binding cassette sub-family G member 5

- Molecule 1: ATP-binding cassette sub-family G member 5

● Molecule 2: ATP-binding cassette sub-family G member 8

Chain B:

• Molecule 2: ATP-binding cassette sub-family G member 8

Chain D:
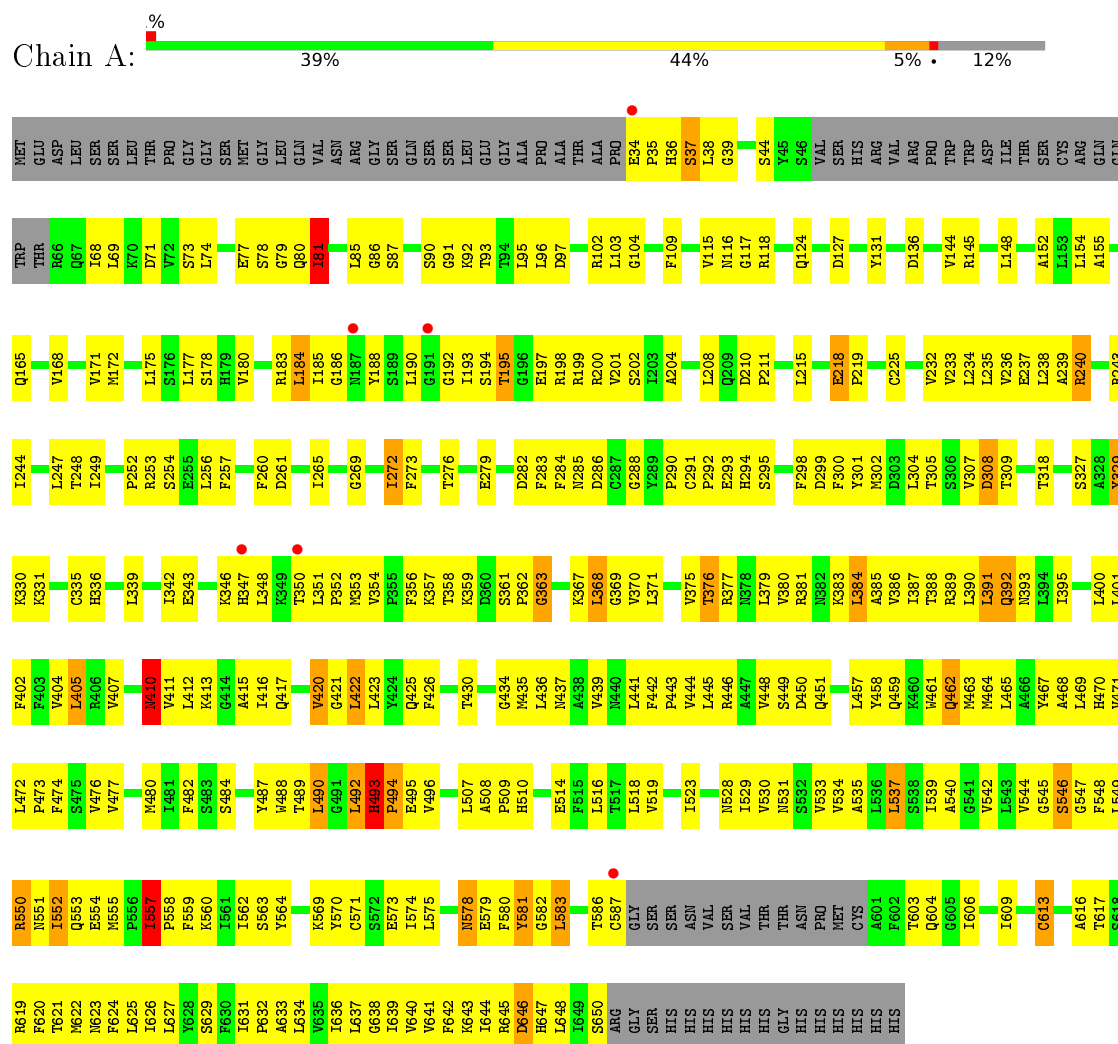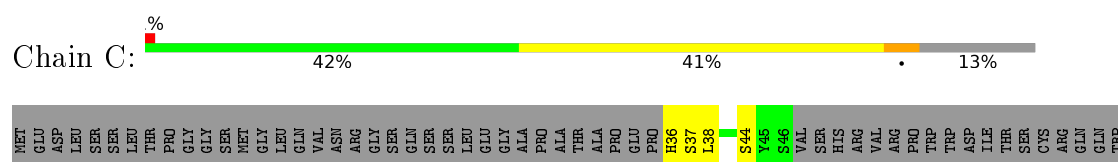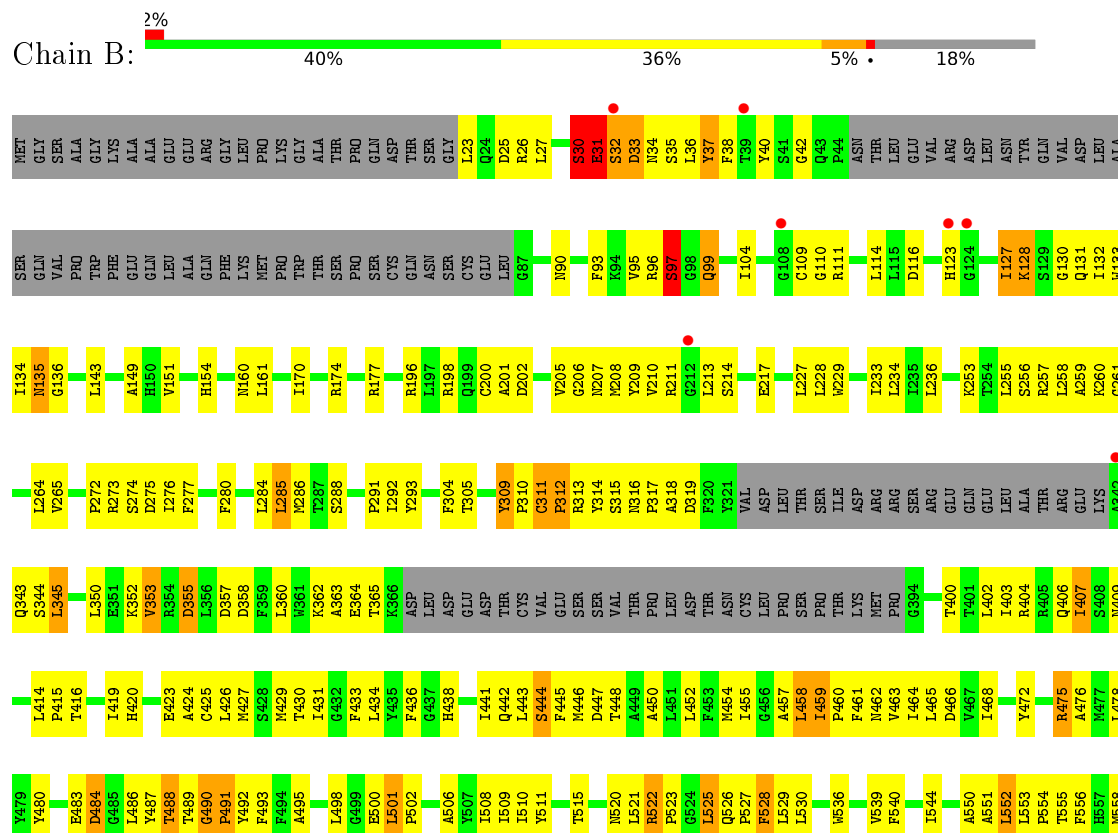
4 Data and refinement statistics $\;(\text{i})$

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	173.55Å 224.80Å 253.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 3.93 24.96 – 3.94	Depositor EDS
% Data completeness (in resolution range)	82.5 (25.00-3.93) 83.5 (24.96-3.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.79 (at 3.97Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.245 , 0.329 0.250 , 0.327	Depositor DCC
R_{free} test set	1845 reflections (5.29%)	DCC
Wilson B-factor (Å^2)	53.7	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 59.0	EDS
L-test for twinning[2]	$<|L|> = 0.44$, $<L^2> = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	18150	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.32% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	0.54	0/4697	0.82	4/6355 (0.1%)
1	C	0.52	0/4639	0.79	2/6272 (0.0%)
2	B	0.54	0/4572	0.84	4/6192 (0.1%)
2	D	0.55	0/4634	0.84	2/6273 (0.0%)
All	All	0.54	0/18542	0.82	12/25092 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	8
1	C	0	3
2	B	0	3
2	D	0	6
All	All	0	20

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	37	TYR	CA-CB-CG	8.16	128.91	113.40
2	B	571	TYR	CA-CB-CG	7.90	128.41	113.40
1	A	81	ILE	CB-CA-C	-6.54	98.52	111.60
2	D	486	LEU	CA-CB-CG	5.79	128.61	115.30
1	A	190	LEU	CA-CB-CG	5.56	128.09	115.30

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	104	GLY	Peptide
1	A	186	GLY	Peptide
1	A	192	GLY	Peptide
1	A	308	ASP	Peptide
1	A	34	GLU	Peptide

5.2 Too-close contacts (i)

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	4606	0	4747	323	0
1	C	4551	0	4682	287	0
2	B	4464	0	4528	321	1
2	D	4529	0	4615	327	1
All	All	18150	0	18572	1233	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 34.

The worst 5 of 1233 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:C:75:TYR:CZ	1:C:77:GLU:OE2	1.85	1.30
1:A:81:ILE:CD1	1:A:260:PHE:HA	1.60	1.29
2:B:36:LEU:CB	2:B:95:VAL:HG21	1.63	1.26
1:A:81:ILE:HD11	1:A:261:ASP:N	1.53	1.20
1:C:140:SER:HA	1:C:185:ILE:HD11	1.17	1.12

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)
2:B:357:ASP:OD1	2:D:339:ARG:NH2[6_454]	2.09	0.11

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	579/666 (87%)	468 (81%)	101 (17%)	10 (2%)	11	56
1	C	571/666 (86%)	465 (81%)	92 (16%)	14 (2%)	7	48
2	B	554/685 (81%)	445 (80%)	91 (16%)	18 (3%)	5	44
2	D	565/685 (82%)	443 (78%)	101 (18%)	21 (4%)	4	40
All	All	2269/2702 (84%)	1821 (80%)	385 (17%)	63 (3%)	6	47

5 of 63 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	557	ILE
2	B	127	ILE
2	B	135	ASN
2	B	484	ASP
2	B	491	PRO

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	512/582 (88%)	474 (93%)	38 (7%)	17	56
1	C	505/582 (87%)	467 (92%)	38 (8%)	17	56
2	B	481/591 (81%)	441 (92%)	40 (8%)	14	51
2	D	490/591 (83%)	449 (92%)	41 (8%)	14	51
All	All	1988/2346 (85%)	1831 (92%)	157 (8%)	15	54

5 of 157 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	579	ASN
2	D	202	ASP
1	C	444	VAL
2	B	597	ARG
2	D	28	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	667	GLN
2	D	420	HIS
1	C	470	HIS
2	B	438	HIS
1	C	528	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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, 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	585/666 (87%)	-0.53	6 (1%)	84	77	20, 64, 145, 252	0
1	C	579/666 (86%)	-0.48	6 (1%)	84	77	20, 65, 143, 220	0
2	B	562/685 (82%)	-0.39	12 (2%)	67	55	20, 77, 157, 247	0
2	D	571/685 (83%)	-0.52	10 (1%)	71	61	20, 49, 139, 234	0
All	All	2297/2702 (85%)	-0.48	34 (1%)	76	66	20, 63, 147, 252	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	621	THR	4.8
2	D	620	LEU	4.4
2	D	619	ASN	4.3
1	A	587	CYS	3.9
1	C	589	SER	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 carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.