wwPDB X-ray Structure Validation Summary Report (i)

Oct 26, 2021 – 10:12 AM EDT

PDB ID	:	7S6B
Title	:	Crystal structure of modular polyketide synthase apo-Lsd14 from the Lasalocid biosynthesis pathway, trapped in the transacylation step
Authors	:	Bagde, S.R.; Mathews, I.I.; Kim, C.-Y.
Deposited on	:	2021-09-13
Resolution	:	2.35 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 following versions of software and data (see references (1)) were used in the production of this report:

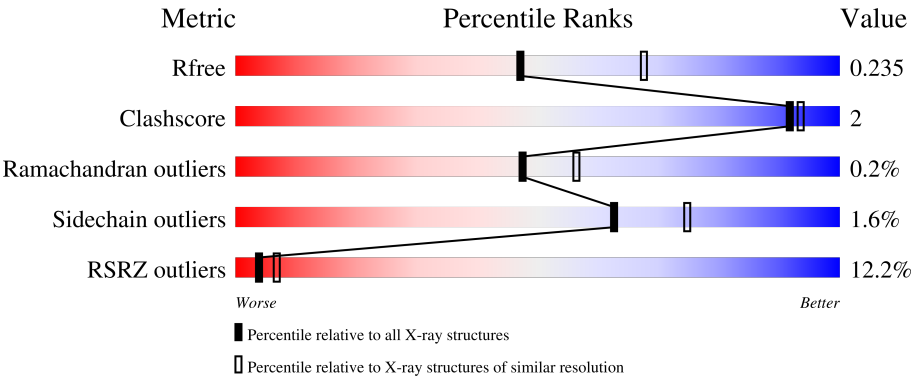
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	944	4% 91% · 5%
1	B	944	8% 88% · 7%
2	C	544	23% 87% 6% 6%
2	D	544	17% 91% · 6%
3	E	179	14% 46% · 52%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 42675 atoms, of which 20630 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 Polyketide synthase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	893	Total	C	H	N	O	S	6437	0	0
			13061	4128	6437	1209	1267	20			
1	B	879	Total	C	H	N	O	S	6313	4	0
			12832	4059	6313	1192	1249	19			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP B6ZK67
A	-18	GLY	-	expression tag	UNP B6ZK67
A	-17	SER	-	expression tag	UNP B6ZK67
A	-16	SER	-	expression tag	UNP B6ZK67
A	-15	HIS	-	expression tag	UNP B6ZK67
A	-14	HIS	-	expression tag	UNP B6ZK67
A	-13	HIS	-	expression tag	UNP B6ZK67
A	-12	HIS	-	expression tag	UNP B6ZK67
A	-11	HIS	-	expression tag	UNP B6ZK67
A	-10	HIS	-	expression tag	UNP B6ZK67
A	-9	SER	-	expression tag	UNP B6ZK67
A	-8	SER	-	expression tag	UNP B6ZK67
A	-7	GLY	-	expression tag	UNP B6ZK67
A	-6	LEU	-	expression tag	UNP B6ZK67
A	-5	VAL	-	expression tag	UNP B6ZK67
A	-4	PRO	-	expression tag	UNP B6ZK67
A	-3	ARG	-	expression tag	UNP B6ZK67
A	-2	GLY	-	expression tag	UNP B6ZK67
A	-1	SER	-	expression tag	UNP B6ZK67
A	0	HIS	-	expression tag	UNP B6ZK67
B	-19	MET	-	initiating methionine	UNP B6ZK67
B	-18	GLY	-	expression tag	UNP B6ZK67
B	-17	SER	-	expression tag	UNP B6ZK67
B	-16	SER	-	expression tag	UNP B6ZK67
B	-15	HIS	-	expression tag	UNP B6ZK67

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-14	HIS	-	expression tag	UNP B6ZK67
B	-13	HIS	-	expression tag	UNP B6ZK67
B	-12	HIS	-	expression tag	UNP B6ZK67
B	-11	HIS	-	expression tag	UNP B6ZK67
B	-10	HIS	-	expression tag	UNP B6ZK67
B	-9	SER	-	expression tag	UNP B6ZK67
B	-8	SER	-	expression tag	UNP B6ZK67
B	-7	GLY	-	expression tag	UNP B6ZK67
B	-6	LEU	-	expression tag	UNP B6ZK67
B	-5	VAL	-	expression tag	UNP B6ZK67
B	-4	PRO	-	expression tag	UNP B6ZK67
B	-3	ARG	-	expression tag	UNP B6ZK67
B	-2	GLY	-	expression tag	UNP B6ZK67
B	-1	SER	-	expression tag	UNP B6ZK67
B	0	HIS	-	expression tag	UNP B6ZK67

- Molecule 2 is a protein called Polyketide synthase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	510	Total	C	H	N	O	S	3578	0	0
			7269	2319	3578	663	706	3			
2	D	514	Total	C	H	N	O	S	3676	0	0
			7426	2355	3676	677	715	3			

- Molecule 3 is a protein called Polyketide synthase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	86	Total	C	H	N	O	S	626	0	0
			1256	391	626	121	115	3			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	433	Total	O	0	2
			435	435		
4	B	287	Total	O	0	0
			287	287		
4	C	40	Total	O	0	0
			40	40		
4	D	56	Total	O	0	0
			56	56		

Continued on next page...

Continued from previous page...

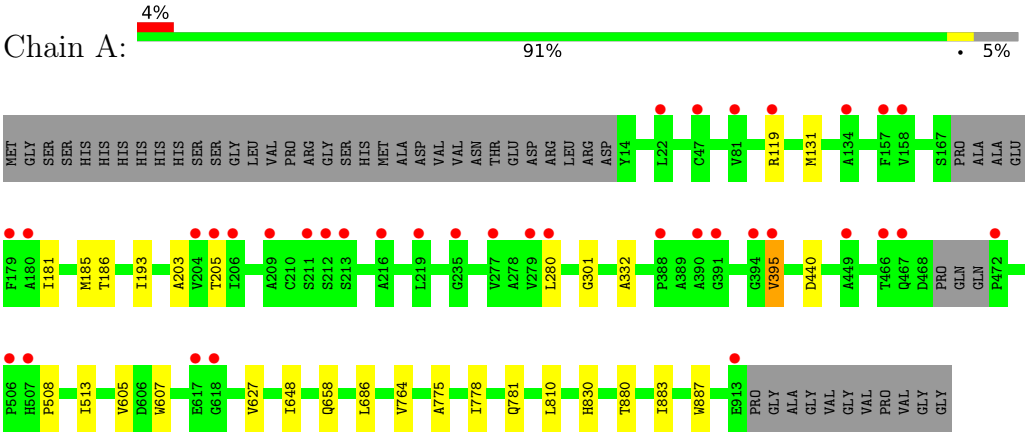
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	13	Total O 13 13	0	0

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.

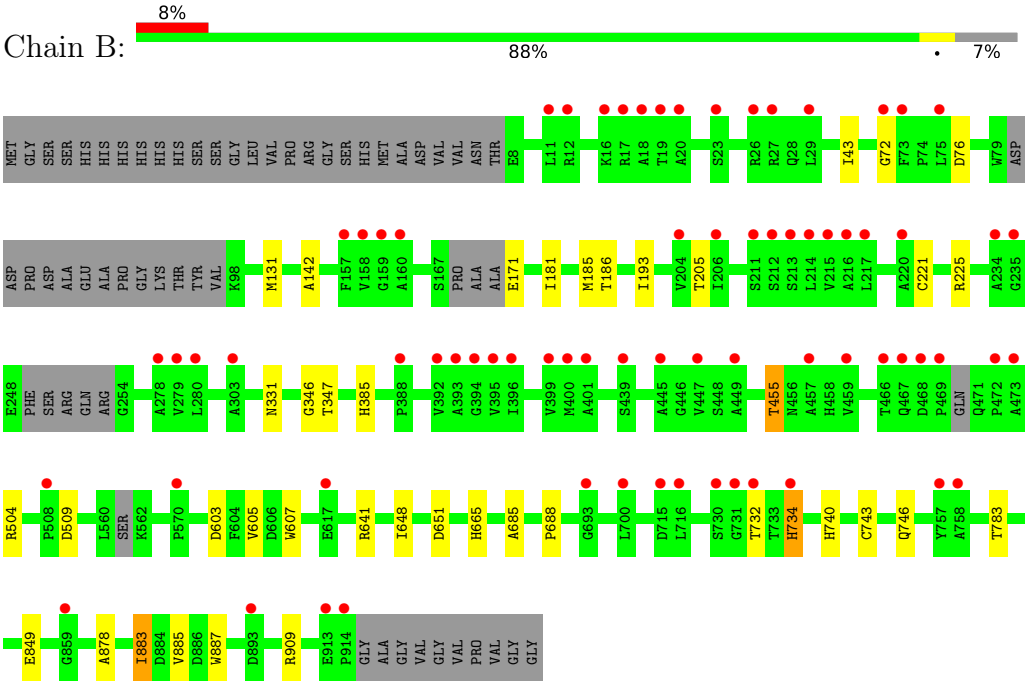
• Molecule 1: Polyketide synthase

Chain A:



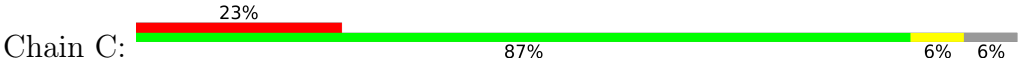
• Molecule 1: Polyketide synthase

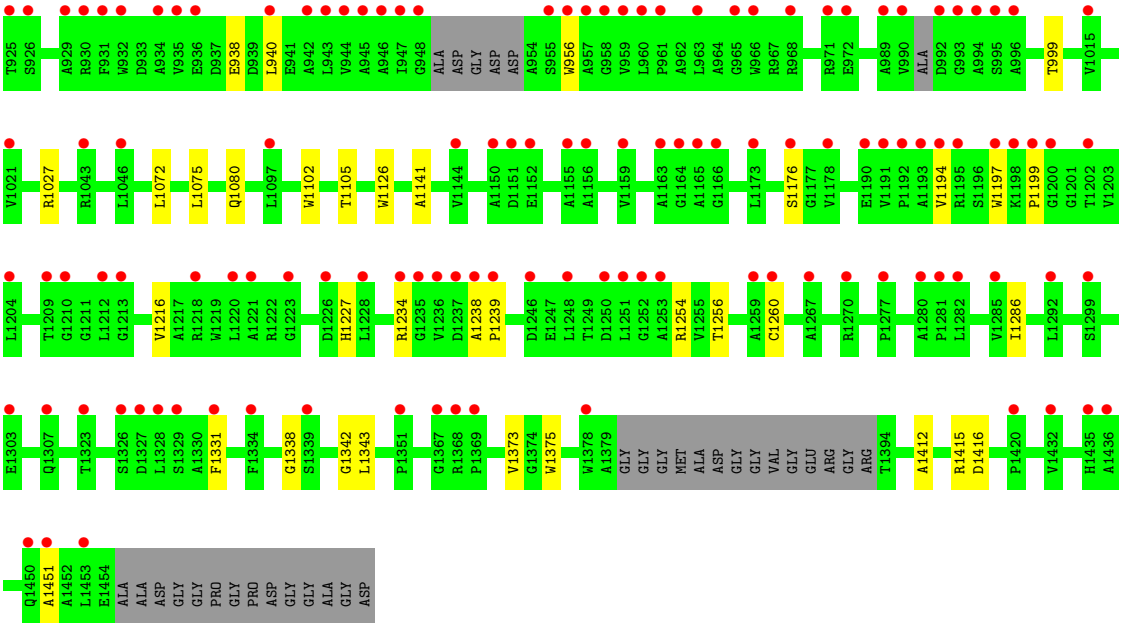
Chain B:



• Molecule 2: Polyketide synthase

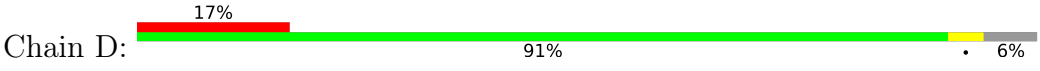
Chain C:

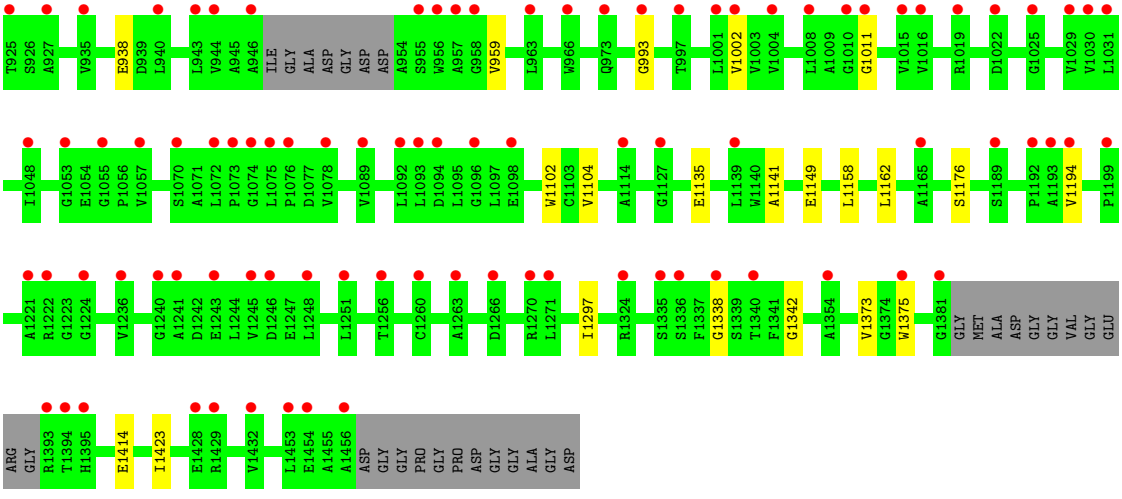




• Molecule 2: Polyketide synthase

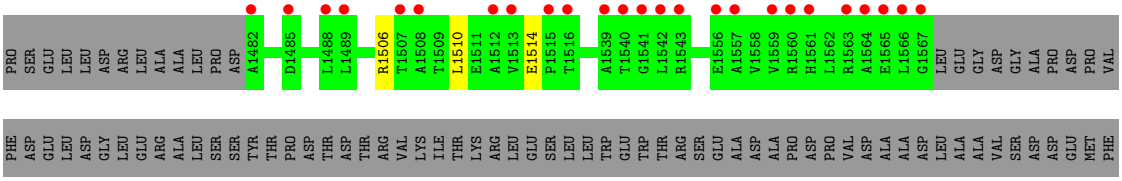
Chain D:





• Molecule 3: Polyketide synthase

Chain E:



GLU
LEU
ILE
ASP
ARG
GLU
LEU
GLY
SER
ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	87.74Å 92.85Å 107.45Å 99.63° 94.93° 106.07°	Depositor
Resolution (Å)	39.20 – 2.35 39.21 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.1 (39.20-2.35) 98.1 (39.21-2.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.34Å)	Xtriage
Refinement program	BUSTER 2.10.4	Depositor
R, R_{free}	0.206 , 0.241 0.200 , 0.235	Depositor DCC
R_{free} test set	6501 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	49.2	Xtriage
Anisotropy	0.487	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 51.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	42675	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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

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	0.41	0/6775	0.59	0/9260
1	B	0.39	0/6681	0.56	0/9135
2	C	0.33	0/3769	0.48	0/5162
2	D	0.34	0/3829	0.49	0/5242
3	E	0.38	0/639	0.55	0/868
All	All	0.38	0/21693	0.54	0/29667

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	6624	6437	6433	19	0
1	B	6519	6313	6288	24	0
2	C	3691	3578	3576	16	0
2	D	3750	3676	3673	11	0
3	E	630	626	625	0	0
4	A	435	0	0	1	0
4	B	287	0	0	1	0
4	C	40	0	0	0	0
4	D	56	0	0	1	0
4	E	13	0	0	0	0
All	All	22045	20630	20595	67	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 67 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)
2:C:1105:THR:HG21	2:C:1126:TRP:HE1	1.51	0.74
1:B:186:THR:HG22	1:B:205:THR:HG21	1.75	0.69
1:B:181:ILE:O	1:B:185:MET:HG3	1.95	0.66
2:C:1105:THR:HG21	2:C:1126:TRP:NE1	2.14	0.63
2:C:1227:HIS:NE2	2:C:1256:THR:OG1	2.31	0.62

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	887/944 (94%)	864 (97%)	22 (2%)	1 (0%)	51	63
1	B	871/944 (92%)	848 (97%)	21 (2%)	2 (0%)	47	56
2	C	502/544 (92%)	470 (94%)	32 (6%)	0	100	100
2	D	508/544 (93%)	488 (96%)	18 (4%)	2 (0%)	34	38
3	E	84/179 (47%)	81 (96%)	3 (4%)	0	100	100
All	All	2852/3155 (90%)	2751 (96%)	96 (3%)	5 (0%)	47	56

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	993	GLY
2	D	1011	GLY
1	A	440	ASP
1	B	346	GLY
1	B	72	GLY

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	671/724 (93%)	665 (99%)	6 (1%)	78 87
1	B	661/724 (91%)	651 (98%)	10 (2%)	65 76
2	C	356/388 (92%)	346 (97%)	10 (3%)	43 53
2	D	366/388 (94%)	362 (99%)	4 (1%)	73 84
3	E	61/144 (42%)	58 (95%)	3 (5%)	25 29
All	All	2115/2368 (89%)	2082 (98%)	33 (2%)	62 75

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	1149	GLU
2	D	1176	SER
3	E	1514	GLU
1	B	734	HIS
1	B	603	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	331	ASN
1	B	665	HIS
2	D	1136	HIS
2	C	973	GLN
1	B	226	GLN

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)

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, 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	893/944 (94%)	0.26	38 (4%) 35 47	37, 53, 81, 117	0
1	B	879/944 (93%)	0.55	77 (8%) 10 15	42, 67, 113, 149	0
2	C	510/544 (93%)	1.30	123 (24%) 0 1	73, 107, 160, 175	0
2	D	514/544 (94%)	0.92	90 (17%) 1 2	53, 96, 134, 146	0
3	E	86/179 (48%)	1.26	25 (29%) 0 0	57, 88, 128, 132	0
All	All	2882/3155 (91%)	0.68	353 (12%) 4 7	37, 73, 130, 175	0

The worst 5 of 353 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	1194	VAL	9.3
2	C	956	TRP	7.9
2	C	1259	ALA	7.4
2	C	1236	VAL	7.4
1	B	11	LEU	6.8

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.