



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2023 – 05:57 AM EDT

PDB ID : 7UI1
Title : Pfs230 D1D2 domain in complex with 230AL-37
Authors : Tang, W.K.; Tolia, N.H.
Deposited on : 2022-03-28
Resolution : 3.30 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 ⓘ 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 ⓘ](#)) 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

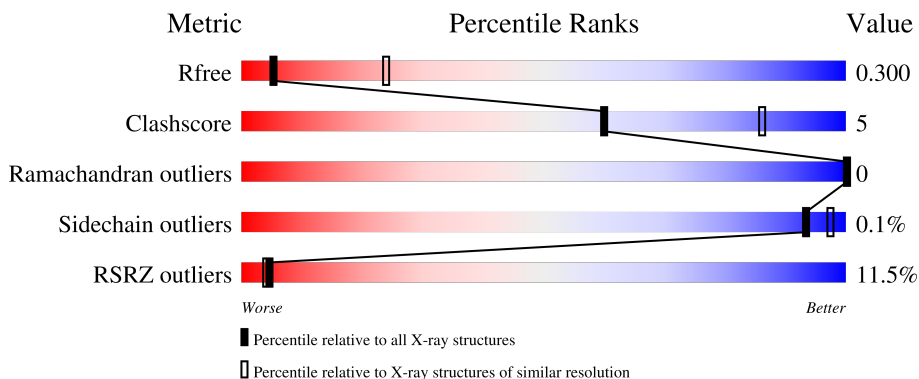
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 $\leq 5\%$. 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	355	
1	B	355	
1	C	355	
1	D	355	
2	H	257	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	257	 <p>% 77% 11% 12%</p>
2	J	257	 <p>75% 14% 12%</p>
2	K	257	 <p>% 77% 11% 12%</p>

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 34018 atoms, of which 16933 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 Gametocyte surface protein P230.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	308	4984	1604	2495	389	484	12	0	0	0
1	B	308	4982	1604	2493	389	484	12	0	0	0
1	C	308	4983	1604	2494	389	484	12	0	0	0
1	D	308	4983	1604	2494	389	484	12	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	540	THR	-	expression tag	UNP P68874
A	541	GLY	-	expression tag	UNP P68874
A	887	GLY	-	expression tag	UNP P68874
A	888	THR	-	expression tag	UNP P68874
A	889	HIS	-	expression tag	UNP P68874
A	890	HIS	-	expression tag	UNP P68874
A	891	HIS	-	expression tag	UNP P68874
A	892	HIS	-	expression tag	UNP P68874
A	893	HIS	-	expression tag	UNP P68874
A	894	HIS	-	expression tag	UNP P68874
B	540	THR	-	expression tag	UNP P68874
B	541	GLY	-	expression tag	UNP P68874
B	887	GLY	-	expression tag	UNP P68874
B	888	THR	-	expression tag	UNP P68874
B	889	HIS	-	expression tag	UNP P68874
B	890	HIS	-	expression tag	UNP P68874
B	891	HIS	-	expression tag	UNP P68874
B	892	HIS	-	expression tag	UNP P68874
B	893	HIS	-	expression tag	UNP P68874
B	894	HIS	-	expression tag	UNP P68874
C	540	THR	-	expression tag	UNP P68874

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	541	GLY	-	expression tag	UNP P68874
C	887	GLY	-	expression tag	UNP P68874
C	888	THR	-	expression tag	UNP P68874
C	889	HIS	-	expression tag	UNP P68874
C	890	HIS	-	expression tag	UNP P68874
C	891	HIS	-	expression tag	UNP P68874
C	892	HIS	-	expression tag	UNP P68874
C	893	HIS	-	expression tag	UNP P68874
C	894	HIS	-	expression tag	UNP P68874
D	540	THR	-	expression tag	UNP P68874
D	541	GLY	-	expression tag	UNP P68874
D	887	GLY	-	expression tag	UNP P68874
D	888	THR	-	expression tag	UNP P68874
D	889	HIS	-	expression tag	UNP P68874
D	890	HIS	-	expression tag	UNP P68874
D	891	HIS	-	expression tag	UNP P68874
D	892	HIS	-	expression tag	UNP P68874
D	893	HIS	-	expression tag	UNP P68874
D	894	HIS	-	expression tag	UNP P68874

- Molecule 2 is a protein called 230AL-37.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	H	227	3515	1128	1739	310	333	5	0	0	0
2	I	227	3514	1128	1738	310	333	5	0	0	0
2	J	227	3516	1128	1740	310	333	5	0	0	0
2	K	227	3516	1128	1740	310	333	5	0	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).

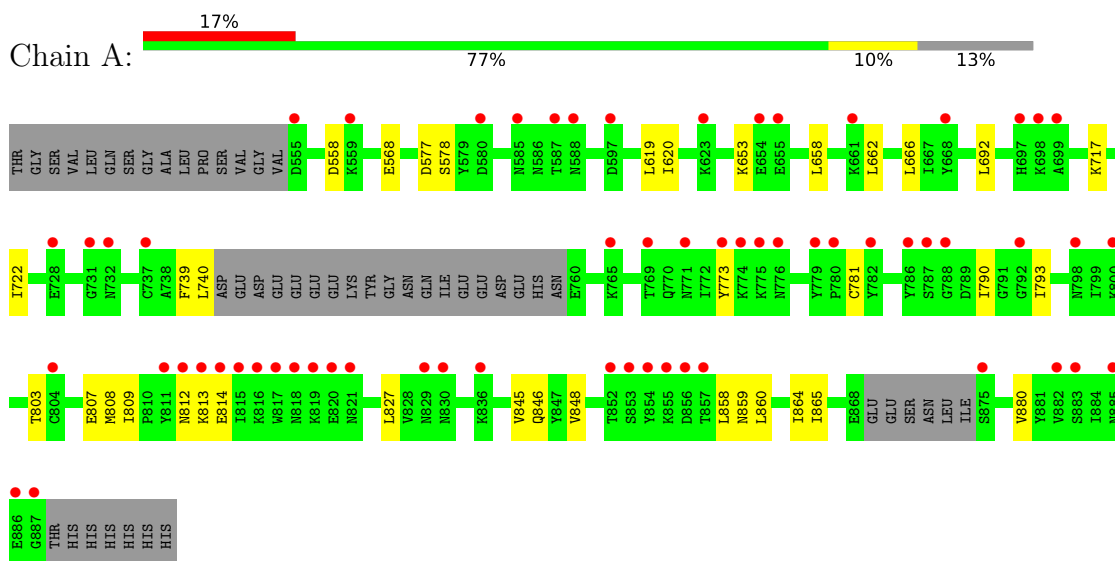


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	H	1	Total	O	S	0	0
			5	4	1		
3	H	1	Total	O	S	0	0
			5	4	1		
3	J	1	Total	O	S	0	0
			5	4	1		
3	K	1	Total	O	S	0	0
			5	4	1		

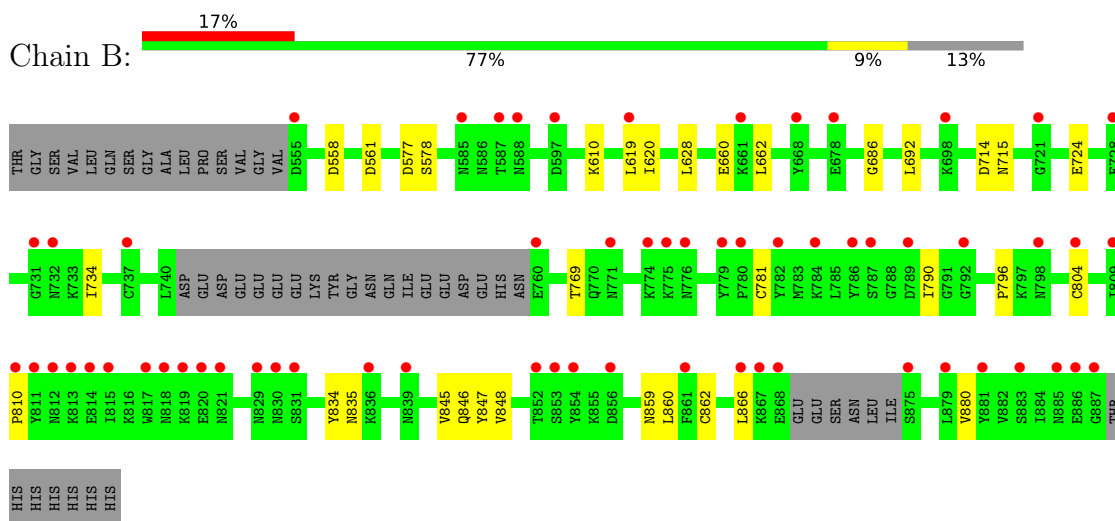
3 Residue-property plots [i](#)

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.

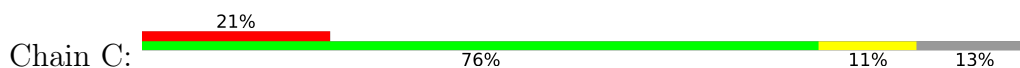
- Molecule 1: Gametocyte surface protein P230



- Molecule 1: Gametocyte surface protein P230

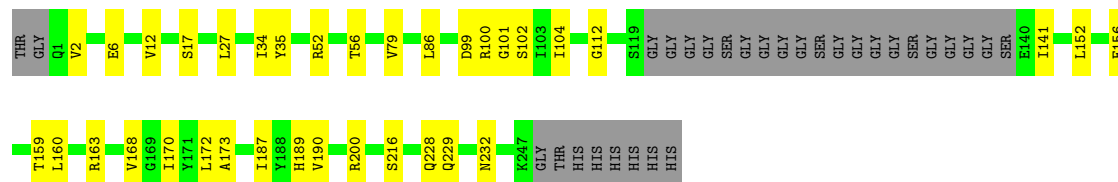


- Molecule 1: Gametocyte surface protein P230




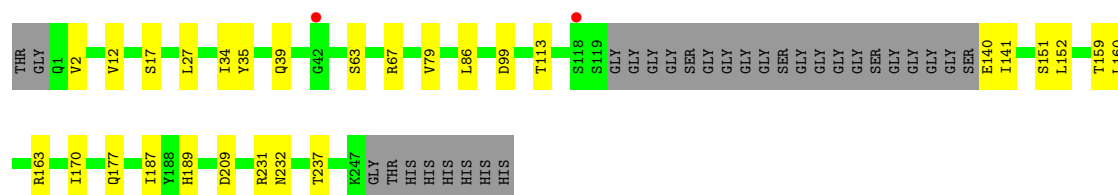
● Molecule 2: 230AL-37

Chain J:  75% 14% 12%



● Molecule 2: 230AL-37

Chain K:  77% 11% 12%



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α , β , γ	49.58Å 155.14Å 375.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.89 – 3.30 19.89 – 3.30	Depositor EDS
% Data completeness (in resolution range)	77.2 (19.89-3.30) 77.3 (19.89-3.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.54 (at 3.29Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.264 , 0.299 0.264 , 0.300	Depositor DCC
R_{free} test set	1717 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	65.3	Xtrriage
Anisotropy	0.055	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 28.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.84	EDS
Total number of atoms	34018	wwPDB-VP
Average B, all atoms (Å ²)	104.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.08% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: SO4

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	0.26	0/2538	0.46	0/3425
1	B	0.26	0/2538	0.46	0/3425
1	C	0.26	0/2538	0.47	0/3425
1	D	0.27	0/2538	0.47	0/3425
2	H	0.30	0/1818	0.52	0/2466
2	I	0.31	0/1818	0.53	0/2466
2	J	0.31	0/1818	0.54	0/2466
2	K	0.31	0/1818	0.53	0/2466
All	All	0.28	0/17424	0.49	0/23564

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 [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	2489	2495	2499	22	0
1	B	2489	2493	2499	21	0
1	C	2489	2494	2499	25	0
1	D	2489	2494	2499	18	0
2	H	1776	1739	1742	17	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	I	1776	1738	1742	20	1
2	J	1776	1740	1742	26	2
2	K	1776	1740	1742	18	2
3	A	5	0	0	0	0
3	H	10	0	0	0	0
3	J	5	0	0	0	0
3	K	5	0	0	0	0
All	All	17085	16933	16964	155	3

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 5.

The worst 5 of 155 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:J:12:VAL:HG21	2:J:86:LEU:HD13	1.58	0.84
2:J:101:GLY:O	2:J:102:SER:OG	2.02	0.78
2:I:34:ILE:HG21	2:I:79:VAL:HG11	1.69	0.74
2:K:34:ILE:HG21	2:K:79:VAL:HG11	1.70	0.74
1:A:814:GLU:HG2	1:A:858:LEU:HD12	1.73	0.70

All (3) 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:J:17:SER:H	2:K:17:SER:OG[3_555]	1.47	0.13
2:J:159:THR:N	2:K:209:ASP:OD2[1_655]	2.17	0.03
2:H:17:SER:OG	2:I:17:SER:O[3_555]	2.19	0.01

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	302/355 (85%)	283 (94%)	19 (6%)	0	100	100
1	B	302/355 (85%)	289 (96%)	13 (4%)	0	100	100
1	C	302/355 (85%)	285 (94%)	17 (6%)	0	100	100
1	D	302/355 (85%)	285 (94%)	17 (6%)	0	100	100
2	H	223/257 (87%)	216 (97%)	7 (3%)	0	100	100
2	I	223/257 (87%)	216 (97%)	7 (3%)	0	100	100
2	J	223/257 (87%)	213 (96%)	10 (4%)	0	100	100
2	K	223/257 (87%)	215 (96%)	8 (4%)	0	100	100
All	All	2100/2448 (86%)	2002 (95%)	98 (5%)	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	A	289/331 (87%)	289 (100%)	0	100	100
1	B	289/331 (87%)	289 (100%)	0	100	100
1	C	289/331 (87%)	288 (100%)	1 (0%)	92	96
1	D	289/331 (87%)	289 (100%)	0	100	100
2	H	191/203 (94%)	191 (100%)	0	100	100
2	I	191/203 (94%)	191 (100%)	0	100	100
2	J	191/203 (94%)	191 (100%)	0	100	100
2	K	191/203 (94%)	191 (100%)	0	100	100
All	All	1920/2136 (90%)	1919 (100%)	1 (0%)	93	97

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	782	TYR

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](#)

5 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	SO4	H	302	-	4,4,4	0.13	0	6,6,6	0.09	0
3	SO4	K	301	-	4,4,4	0.13	0	6,6,6	0.10	0
3	SO4	J	301	-	4,4,4	0.13	0	6,6,6	0.19	0
3	SO4	H	301	-	4,4,4	0.12	0	6,6,6	0.18	0
3	SO4	A	901	-	4,4,4	0.14	0	6,6,6	0.11	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.

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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(Å ²)	Q < 0.9
1	A	308/355 (86%)	1.27	62 (20%) 1 1	44, 127, 208, 239	0
1	B	308/355 (86%)	1.19	62 (20%) 1 1	47, 125, 204, 260	0
1	C	308/355 (86%)	1.33	75 (24%) 0 0	54, 125, 232, 277	0
1	D	308/355 (86%)	0.80	40 (12%) 3 3	46, 99, 178, 247	0
2	H	227/257 (88%)	0.07	3 (1%) 77 77	30, 51, 78, 109	0
2	I	227/257 (88%)	-0.06	2 (0%) 84 84	27, 43, 65, 87	0
2	J	227/257 (88%)	-0.02	0 100 100	31, 45, 63, 82	0
2	K	227/257 (88%)	0.06	2 (0%) 84 84	29, 50, 70, 122	0
All	All	2140/2448 (87%)	0.67	246 (11%) 4 4	27, 74, 201, 277	0

The worst 5 of 246 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	814	GLU	10.6
1	A	821	ASN	10.6
1	A	855	LYS	10.0
1	D	813	LYS	9.6
1	A	775	LYS	9.3

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

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, 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
3	SO4	H	302	5/5	0.91	0.18	68,68,74,74	0
3	SO4	K	301	5/5	0.93	0.24	61,66,68,73	0
3	SO4	J	301	5/5	0.96	0.13	57,58,59,61	0
3	SO4	H	301	5/5	0.96	0.17	54,56,59,61	0
3	SO4	A	901	5/5	0.97	0.12	56,56,59,63	0

6.5 Other polymers [i](#)

There are no such residues in this entry.