



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 4, 2023 – 11:09 AM EST

PDB ID : 8FZK
Title : Dimeric human importin alpha subunit
Authors : Donnelly, C.M.; Stewart, M.; Forwood, J.K.
Deposited on : 2023-01-29
Resolution : 2.10 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 : 2.32.1
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) : 2.32.1

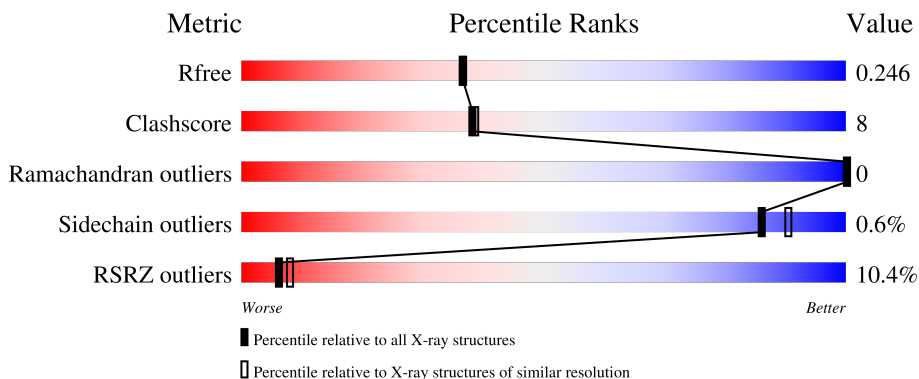
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	460	
1	B	460	
1	C	460	
1	D	460	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 26583 atoms, of which 13208 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 Importin subunit alpha-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	419	6495	2044	3282	539	617	13	0	7	0
1	B	422	6494	2050	3276	542	615	11	0	2	0
1	C	422	6567	2068	3321	546	621	11	0	9	0
1	D	424	6563	2068	3316	545	622	12	0	5	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	70	SER	-	expression tag	UNP P52292
B	70	SER	-	expression tag	UNP P52292
C	70	SER	-	expression tag	UNP P52292
D	70	SER	-	expression tag	UNP P52292

- Molecule 2 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C₆H₁₃NO₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	H	N	O			S
2	A	1	25	6	13	1	4	1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	C	1	1	1	0	0

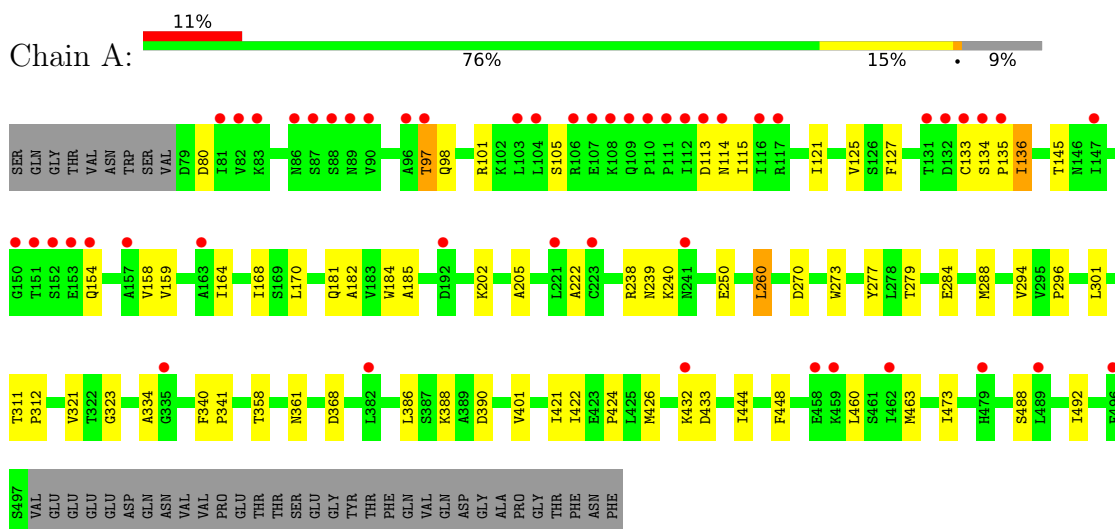
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	110	110	110	0	0
4	B	49	49	49	0	0
4	C	171	171	171	0	0
4	D	108	108	108	0	0

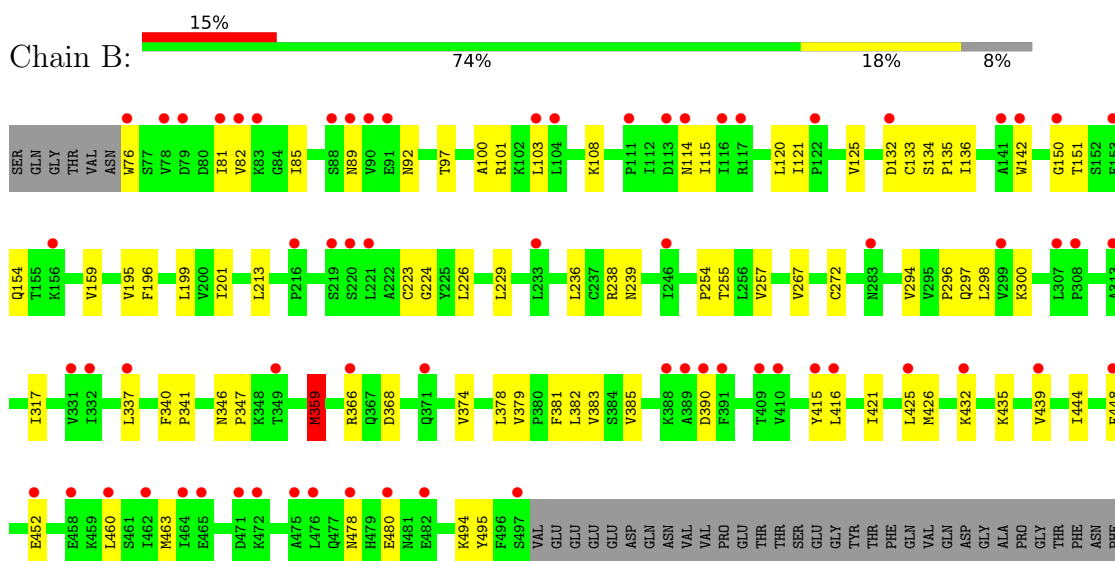
3 Residue-property plots [i](#)

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.

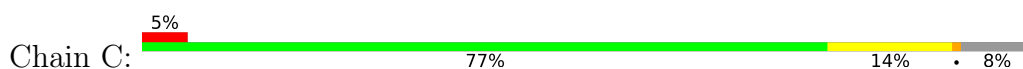
- Molecule 1: Importin subunit alpha-1

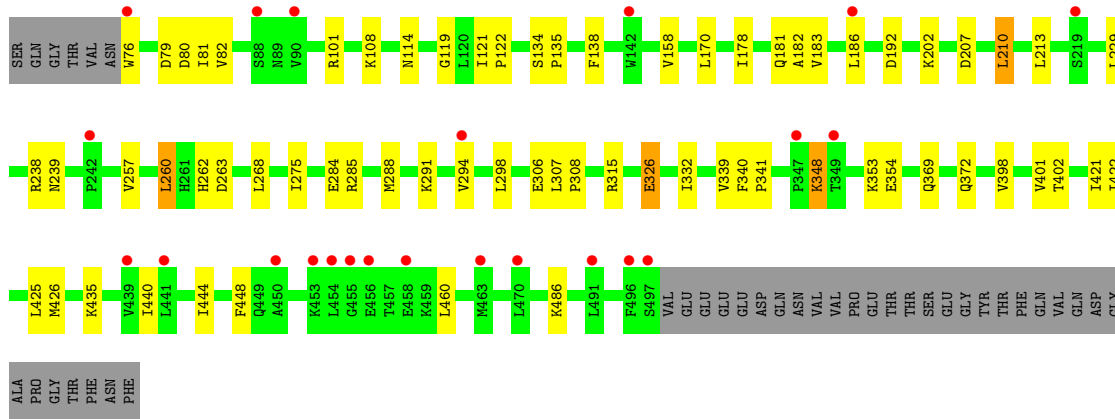


- Molecule 1: Importin subunit alpha-1

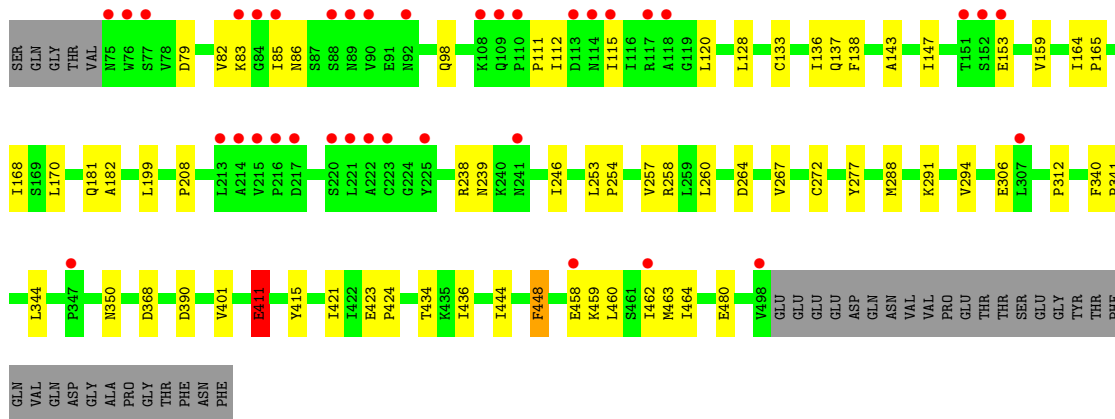
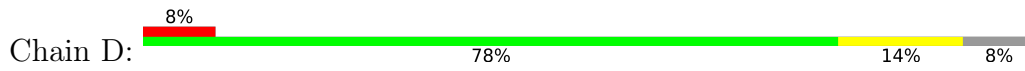


- Molecule 1: Importin subunit alpha-1





● Molecule 1: Importin subunit alpha-1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	50.11Å 90.84Å 94.60Å 103.19° 90.37° 97.39°	Depositor
Resolution (Å)	29.91 – 2.10 29.91 – 2.10	Depositor EDS
% Data completeness (in resolution range)	98.0 (29.91-2.10) 98.0 (29.91-2.10)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.07 (at 2.10Å)	Xtrriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.201 , 0.247 0.201 , 0.246	Depositor DCC
R_{free} test set	4586 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å ²)	34.7	Xtrriage
Anisotropy	0.140	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 46.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	26583	wwPDB-VP
Average B, all atoms (Å ²)	58.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.18% 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: MES, MG

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	0.71	1/3288 (0.0%)	0.83	6/4484 (0.1%)
1	B	0.57	1/3281 (0.0%)	0.68	1/4476 (0.0%)
1	C	0.76	2/3341 (0.1%)	0.89	8/4558 (0.2%)
1	D	0.76	5/3322 (0.2%)	0.83	4/4532 (0.1%)
All	All	0.70	9/13232 (0.1%)	0.81	19/18050 (0.1%)

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	411	GLU	CD-OE1	-13.55	1.10	1.25
1	D	411	GLU	CD-OE2	-8.08	1.16	1.25
1	D	415	TYR	CD1-CE1	6.25	1.48	1.39
1	C	339	VAL	CB-CG1	-6.03	1.40	1.52
1	C	326	GLU	CG-CD	5.68	1.60	1.51

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	411	GLU	OE1-CD-OE2	17.83	144.69	123.30
1	D	411	GLU	CG-CD-OE1	-12.79	92.71	118.30
1	A	388	LYS	CD-CE-NZ	-8.01	93.27	111.70
1	A	368	ASP	CB-CG-OD1	7.67	125.21	118.30
1	C	229[A]	LEU	CB-CG-CD2	-7.67	97.96	111.00

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	3213	3282	3263	54	0
1	B	3218	3276	3278	58	0
1	C	3246	3321	3275	49	1
1	D	3247	3316	3290	44	1
2	A	12	13	12	3	0
3	C	1	0	0	0	0
4	A	110	0	0	5	4
4	B	49	0	0	0	0
4	C	171	0	0	8	3
4	D	108	0	0	2	1
All	All	13375	13208	13118	198	5

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 8.

The worst 5 of 198 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:B:81:ILE:HD11	1:B:100:ALA:HB2	1.37	1.04
1:C:121[B]:ILE:HG22	1:C:122:PRO:HD3	1.42	1.02
1:B:416:LEU:HG	1:B:421:ILE:HG21	1.41	0.99
1:D:411:GLU:N	1:D:411:GLU:OE1	2.03	0.90
1:C:207:ASP:HA	4:C:946:HOH:O	1.74	0.86

All (5) 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)
4:A:805:HOH:O	4:C:909:HOH:O[1_655]	1.97	0.23
4:A:904:HOH:O	4:C:1054:HOH:O[1_666]	1.98	0.22
4:A:814:HOH:O	4:D:683:HOH:O[1_566]	2.00	0.20
4:A:821:HOH:O	4:C:989:HOH:O[1_655]	2.06	0.14
1:C:202:LYS:HZ1	1:D:344:LEU:O[1_455]	1.58	0.02

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	424/460 (92%)	421 (99%)	3 (1%)	0	100	100
1	B	422/460 (92%)	418 (99%)	4 (1%)	0	100	100
1	C	429/460 (93%)	423 (99%)	6 (1%)	0	100	100
1	D	427/460 (93%)	423 (99%)	4 (1%)	0	100	100
All	All	1702/1840 (92%)	1685 (99%)	17 (1%)	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	359/389 (92%)	358 (100%)	1 (0%)	92	95
1	B	356/389 (92%)	353 (99%)	3 (1%)	81	86
1	C	365/389 (94%)	362 (99%)	3 (1%)	81	86
1	D	363/389 (93%)	362 (100%)	1 (0%)	92	95
All	All	1443/1556 (93%)	1435 (99%)	8 (1%)	86	90

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	411	GLU
1	C	348	LYS
1	C	114	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	390	ASP
1	C	284	GLU

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	C	251	GLN
1	D	239	ASN
1	D	477	GLN
1	D	449	GLN
1	B	95	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](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	MES	A	701	-	12,12,12	2.13	1 (8%)	14,16,16	2.78	6 (42%)

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	MES	A	701	-	-	3/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	MES	C8-S	-6.81	1.67	1.77

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	MES	O2S-S-C8	6.69	114.97	106.92
2	A	701	MES	C5-N4-C3	4.16	118.19	108.83
2	A	701	MES	C2-C3-N4	3.86	115.95	110.10
2	A	701	MES	O3S-S-C8	-3.57	100.00	105.77
2	A	701	MES	C6-O1-C2	2.65	118.74	109.89

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	701	MES	C7-C8-S-O3S
2	A	701	MES	C7-C8-S-O1S
2	A	701	MES	C7-C8-S-O2S

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	701	MES	3	0

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	419/460 (91%)	0.64	49 (11%) 4 6	24, 45, 99, 125	0
1	B	422/460 (91%)	1.00	67 (15%) 1 2	44, 66, 90, 103	0
1	C	422/460 (91%)	0.34	23 (5%) 25 31	25, 38, 72, 90	0
1	D	424/460 (92%)	0.49	37 (8%) 10 13	22, 52, 80, 102	0
All	All	1687/1840 (91%)	0.62	176 (10%) 6 8	22, 52, 87, 125	0

The worst 5 of 176 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	90	VAL	8.7
1	B	332	ILE	7.9
1	D	221	LEU	7.5
1	B	78	VAL	7.0
1	A	117	ARG	6.8

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(\AA^2)	Q<0.9
3	MG	C	801	1/1	0.89	0.40	60,60,60,60	0
2	MES	A	701	12/12	0.97	0.15	33,42,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.