wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 11:59 PM EDT

PDB ID	:	7SEM
Title	:	Structure-based design of prefusion-stabilized human metapneumovirus fusion proteins
Authors	:	Rush, S.A.; Hsieh, C.-L.; McLellan, J.S.
Deposited on	:	2021-09-30
Resolution	:	2.20 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 (1)) were used in the production of this report:

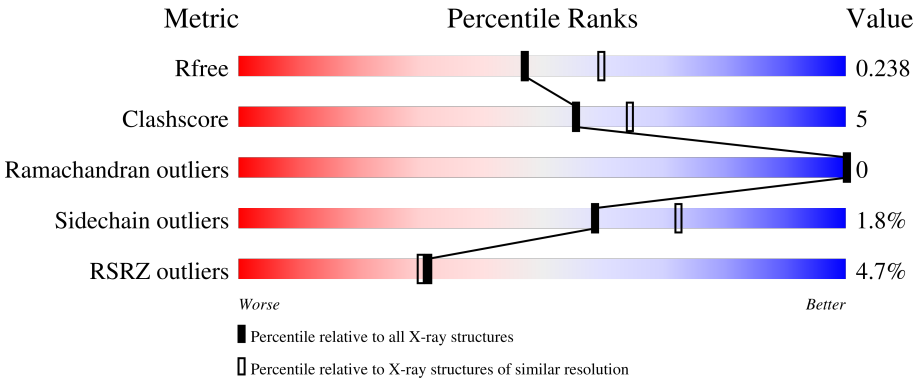
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	B	228	9% 74% 19% 6%
2	F	551	% 73% 5% 22%
3	C	216	7% 83% 12% 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6595 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 MPE8 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	214	Total	C	N	O	S	0	0	0
			1595	1010	266	313	6			

- Molecule 2 is a protein called Fusion glycoprotein F0.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	430	Total	C	N	O	S	0	0	0
			3270	2044	562	639	25			

There are 69 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	100	ARG	GLN	conflict	UNP H6X1Z0
F	101	ARG	SER	conflict	UNP H6X1Z0
F	127	CYS	THR	conflict	UNP H6X1Z0
F	153	CYS	ASN	conflict	UNP H6X1Z0
F	185	PRO	ALA	conflict	UNP H6X1Z0
F	365	CYS	THR	conflict	UNP H6X1Z0
F	368	ASN	HIS	conflict	UNP H6X1Z0
F	463	CYS	VAL	conflict	UNP H6X1Z0
F	491	GLY	-	expression tag	UNP H6X1Z0
F	492	GLY	-	expression tag	UNP H6X1Z0
F	493	GLY	-	expression tag	UNP H6X1Z0
F	494	SER	-	expression tag	UNP H6X1Z0
F	495	GLY	-	expression tag	UNP H6X1Z0
F	496	TYR	-	expression tag	UNP H6X1Z0
F	497	ILE	-	expression tag	UNP H6X1Z0
F	498	PRO	-	expression tag	UNP H6X1Z0
F	499	GLU	-	expression tag	UNP H6X1Z0
F	500	ALA	-	expression tag	UNP H6X1Z0
F	501	PRO	-	expression tag	UNP H6X1Z0
F	502	ARG	-	expression tag	UNP H6X1Z0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	503	ASP	-	expression tag	UNP H6X1Z0
F	504	GLY	-	expression tag	UNP H6X1Z0
F	505	GLN	-	expression tag	UNP H6X1Z0
F	506	ALA	-	expression tag	UNP H6X1Z0
F	507	TYR	-	expression tag	UNP H6X1Z0
F	508	VAL	-	expression tag	UNP H6X1Z0
F	509	ARG	-	expression tag	UNP H6X1Z0
F	510	LYS	-	expression tag	UNP H6X1Z0
F	511	ASP	-	expression tag	UNP H6X1Z0
F	512	GLY	-	expression tag	UNP H6X1Z0
F	513	GLU	-	expression tag	UNP H6X1Z0
F	514	TRP	-	expression tag	UNP H6X1Z0
F	515	VAL	-	expression tag	UNP H6X1Z0
F	516	LEU	-	expression tag	UNP H6X1Z0
F	517	LEU	-	expression tag	UNP H6X1Z0
F	518	SER	-	expression tag	UNP H6X1Z0
F	519	THR	-	expression tag	UNP H6X1Z0
F	520	PHE	-	expression tag	UNP H6X1Z0
F	521	LEU	-	expression tag	UNP H6X1Z0
F	522	GLY	-	expression tag	UNP H6X1Z0
F	523	ARG	-	expression tag	UNP H6X1Z0
F	524	SER	-	expression tag	UNP H6X1Z0
F	525	LEU	-	expression tag	UNP H6X1Z0
F	526	GLU	-	expression tag	UNP H6X1Z0
F	527	VAL	-	expression tag	UNP H6X1Z0
F	528	LEU	-	expression tag	UNP H6X1Z0
F	529	PHE	-	expression tag	UNP H6X1Z0
F	530	GLN	-	expression tag	UNP H6X1Z0
F	531	GLY	-	expression tag	UNP H6X1Z0
F	532	PRO	-	expression tag	UNP H6X1Z0
F	533	GLY	-	expression tag	UNP H6X1Z0
F	534	HIS	-	expression tag	UNP H6X1Z0
F	535	HIS	-	expression tag	UNP H6X1Z0
F	536	HIS	-	expression tag	UNP H6X1Z0
F	537	HIS	-	expression tag	UNP H6X1Z0
F	538	HIS	-	expression tag	UNP H6X1Z0
F	539	HIS	-	expression tag	UNP H6X1Z0
F	540	HIS	-	expression tag	UNP H6X1Z0
F	541	HIS	-	expression tag	UNP H6X1Z0
F	542	SER	-	expression tag	UNP H6X1Z0
F	543	ALA	-	expression tag	UNP H6X1Z0
F	544	TRP	-	expression tag	UNP H6X1Z0

Continued on next page...

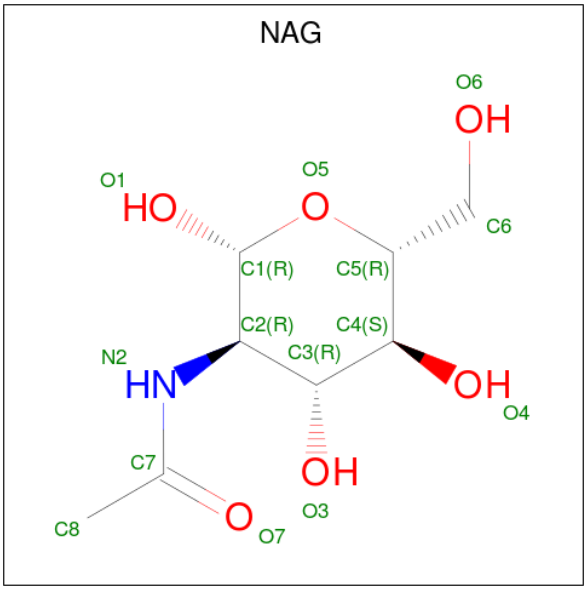
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	545	SER	-	expression tag	UNP H6X1Z0
F	546	HIS	-	expression tag	UNP H6X1Z0
F	547	PRO	-	expression tag	UNP H6X1Z0
F	548	GLN	-	expression tag	UNP H6X1Z0
F	549	PHE	-	expression tag	UNP H6X1Z0
F	550	GLU	-	expression tag	UNP H6X1Z0
F	551	LYS	-	expression tag	UNP H6X1Z0

- Molecule 3 is a protein called MPE8 Fab light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	204	Total	C	N	O	S	0	0	0
			1515	947	254	310	4			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	F	1	Total	C	N	O	0	0
			14	8	1	5		
4	F	1	Total	C	N	O	0	0
			14	8	1	5		
4	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is water.

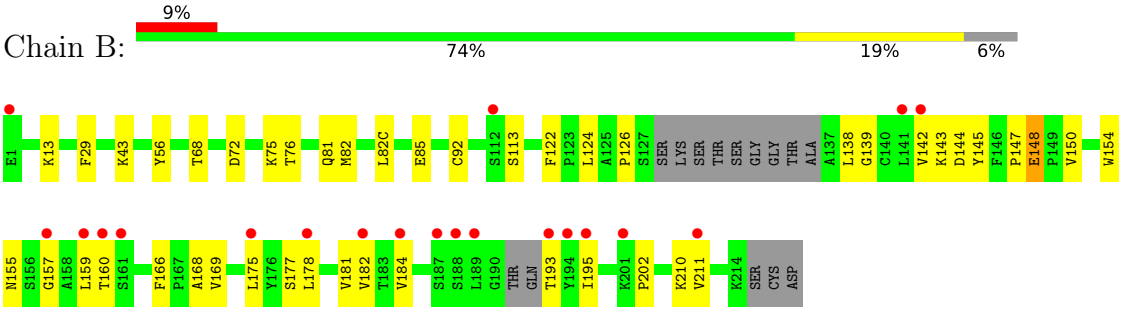
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	47	Total O 47 47	0	0
5	F	89	Total O 89 89	0	0
5	C	37	Total O 37 37	0	0

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.

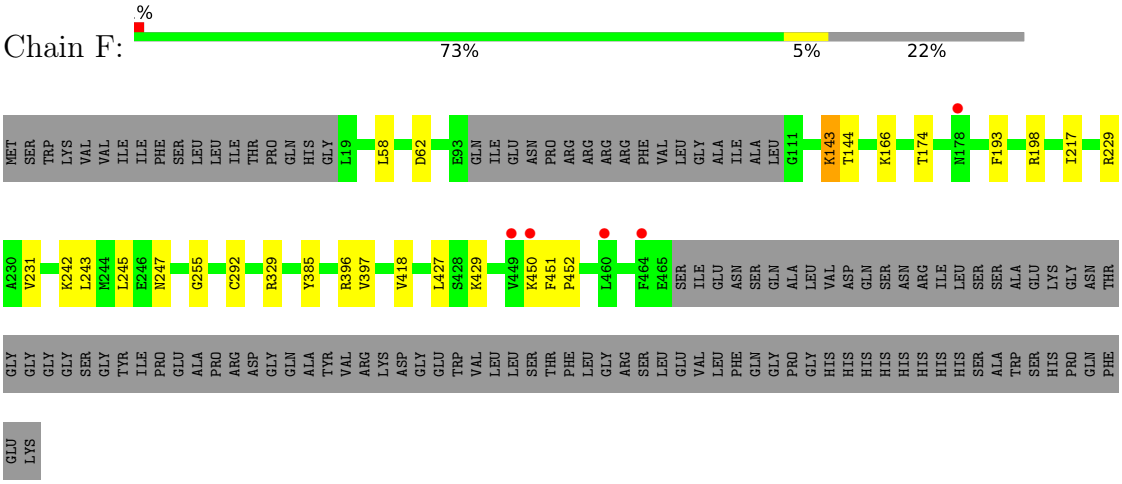
• Molecule 1: MPE8 Fab heavy chain

Chain B:



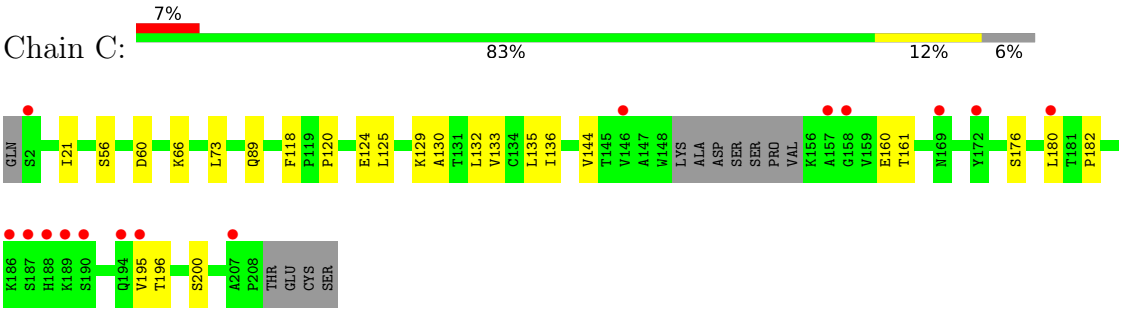
• Molecule 2: Fusion glycoprotein F0

Chain F:



• Molecule 3: MPE8 Fab light chain

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	68.17Å 45.56Å 176.27Å 90.00° 94.21° 90.00°	Depositor
Resolution (Å)	43.95 – 2.20 43.95 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.2 (43.95-2.20) 97.2 (43.95-2.20)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.216 , 0.239 0.215 , 0.238	Depositor DCC
R_{free} test set	2545 reflections (4.71%)	wwPDB-VP
Wilson B-factor (Å^2)	39.4	Xtriage
Anisotropy	0.610	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 37.1	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	0.94	EDS
Total number of atoms	6595	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.21% 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

Bond lengths and bond angles in the following residue types are not validated in this section: NAG

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	B	0.53	0/1632	0.64	0/2218
2	F	0.42	0/3316	0.60	0/4491
3	C	0.42	0/1551	0.57	0/2119
All	All	0.45	0/6499	0.60	0/8828

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	B	1595	0	1567	35	0
2	F	3270	0	3270	14	0
3	C	1515	0	1464	22	0
4	F	42	0	39	0	0
5	B	47	0	0	0	0
5	C	37	0	0	0	0
5	F	89	0	0	0	0
All	All	6595	0	6340	65	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 5.

The worst 5 of 65 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:142:VAL:HG22	1:B:178:LEU:H	1.32	0.95
1:B:169:VAL:HG11	3:C:160:GLU:HB3	1.63	0.81
1:B:157:GLY:O	1:B:160:THR:HG23	1.80	0.81
1:B:138:LEU:HD13	1:B:211:VAL:HG21	1.68	0.76
1:B:157:GLY:O	1:B:160:THR:CG2	2.35	0.75

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	B	208/228 (91%)	199 (96%)	9 (4%)	0	100	100
2	F	426/551 (77%)	406 (95%)	20 (5%)	0	100	100
3	C	200/216 (93%)	188 (94%)	12 (6%)	0	100	100
All	All	834/995 (84%)	793 (95%)	41 (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	B	179/190 (94%)	176 (98%)	3 (2%)	60	74

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	F	365/466 (78%)	358 (98%)	7 (2%)	57 71
3	C	170/181 (94%)	167 (98%)	3 (2%)	59 72
All	All	714/837 (85%)	701 (98%)	13 (2%)	59 72

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	229	ARG
2	F	429	LYS
3	C	89	GLN
3	C	56	SER
3	C	60	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	164	HIS
2	F	455	GLN
3	C	126	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)

3 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	$\#\|Z\| > 2$
4	NAG	F	603	2	14,14,15	0.36	0	17,19,21	0.76	0
4	NAG	F	601	2	14,14,15	0.37	0	17,19,21	0.67	0
4	NAG	F	602	2	14,14,15	0.38	0	17,19,21	1.06	2 (11%)

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
4	NAG	F	603	2	-	3/6/23/26	0/1/1/1
4	NAG	F	601	2	-	2/6/23/26	0/1/1/1
4	NAG	F	602	2	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	F	602	NAG	C2-N2-C7	-3.15	118.42	122.90
4	F	602	NAG	C4-C3-C2	-2.25	107.72	111.02

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	F	601	NAG	C8-C7-N2-C2
4	F	601	NAG	O7-C7-N2-C2
4	F	603	NAG	C8-C7-N2-C2
4	F	603	NAG	O7-C7-N2-C2
4	F	603	NAG	C1-C2-N2-C7

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

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	B	214/228 (93%)	0.52	20 (9%) 8 7	28, 66, 119, 136	0
2	F	430/551 (78%)	0.11	5 (1%) 79 77	27, 50, 80, 122	0
3	C	204/216 (94%)	0.39	15 (7%) 14 13	26, 56, 117, 143	0
All	All	848/995 (85%)	0.28	40 (4%) 31 30	26, 53, 112, 143	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	449	VAL	7.0
1	B	189	LEU	5.8
3	C	207	ALA	5.2
1	B	160	THR	5.0
2	F	450	LYS	5.0

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

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
4	NAG	F	603	14/15	0.76	0.26	73,78,83,83	0
4	NAG	F	601	14/15	0.77	0.36	69,76,79,79	0
4	NAG	F	602	14/15	0.80	0.42	67,75,76,76	0

6.5 Other polymers (i)

There are no such residues in this entry.