



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2022 – 04:03 PM EST

PDB ID : 7SG1
Title : XPA5 TCR in complex with HLA-DQ2-alpha1
Authors : Ciacchi, L.; Farenc, C.; Petersen, J.; Reid, H.H.; Rossjohn, J.
Deposited on : 2021-10-04
Resolution : 3.10 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

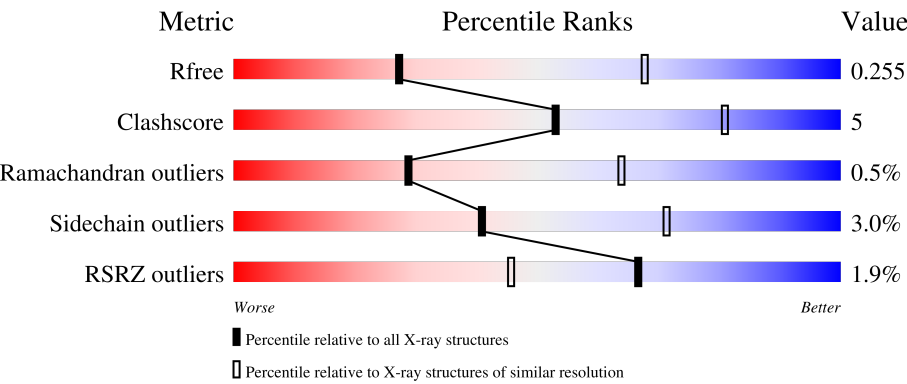
MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.26
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.26

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 3.10 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.



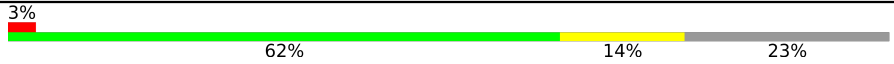

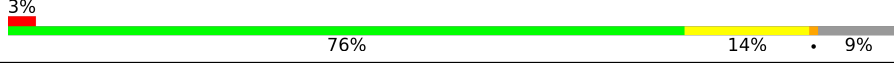


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R _{free}	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	183	
1	F	183	
2	B	202	
2	G	202	
3	D	203	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	203	
4	E	259	
4	J	259	
5	C	16	
5	H	16	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
7	EDO	F	202	-	-	-	X

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12544 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 HLA class II histocompatibility antigen, DQ alpha 1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	180	Total	C	N	O	S	0	0	0
			1436	926	235	272	3			
1	F	181	Total	C	N	O	S	0	0	0
			1444	930	236	275	3			

- Molecule 2 is a protein called MHC class II HLA-DQ-beta-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	174	Total	C	N	O	S	0	0	0
			1434	908	254	265	7			
2	G	179	Total	C	N	O	S	0	0	0
			1473	931	263	272	7			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-9	ILE	-	expression tag	UNP O19712
B	-8	GLU	-	expression tag	UNP O19712
B	-7	GLY	-	expression tag	UNP O19712
B	-6	ARG	-	expression tag	UNP O19712
B	-5	GLY	-	expression tag	UNP O19712
B	-4	GLY	-	expression tag	UNP O19712
B	-3	SER	-	expression tag	UNP O19712
B	-2	GLY	-	expression tag	UNP O19712
B	-1	ALA	-	expression tag	UNP O19712
B	0	SER	-	expression tag	UNP O19712
G	-9	ILE	-	expression tag	UNP O19712
G	-8	GLU	-	expression tag	UNP O19712
G	-7	GLY	-	expression tag	UNP O19712
G	-6	ARG	-	expression tag	UNP O19712
G	-5	GLY	-	expression tag	UNP O19712
G	-4	GLY	-	expression tag	UNP O19712

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	-3	SER	-	expression tag	UNP O19712
G	-2	GLY	-	expression tag	UNP O19712
G	-1	ALA	-	expression tag	UNP O19712
G	0	SER	-	expression tag	UNP O19712

- Molecule 3 is a protein called T-cell receptor, xpa5, alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	193	Total	C	N	O	S	0	0	0
			1512	947	253	302	10			
3	I	156	Total	C	N	O	S	0	0	0
			1221	767	204	242	8			

- Molecule 4 is a protein called T-cell receptor, xpa5, beta chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	244	Total	C	N	O	S	0	0	0
			1916	1205	327	375	9			
4	J	235	Total	C	N	O	S	0	0	0
			1846	1165	315	357	9			

- Molecule 5 is a protein called DQ2-glia-alpha1a peptide.

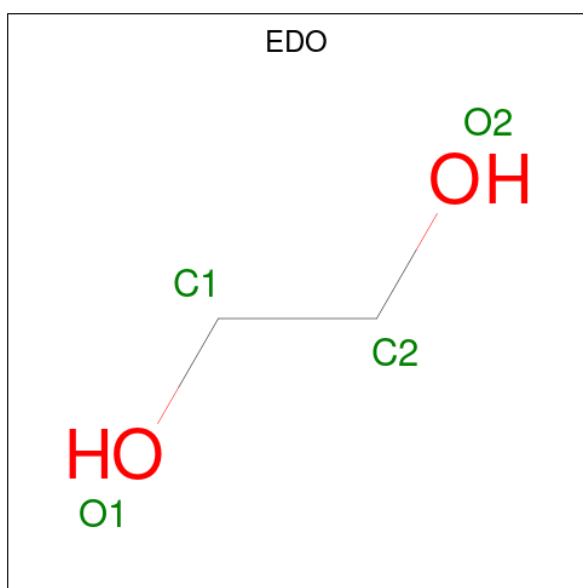
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
5	C	15	Total	C	N	O	0	0	0
			110	71	17	22			
5	H	15	Total	C	N	O	0	0	0
			110	71	17	22			

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆).



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

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			4	2	2		
7	F	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	F	1	Total	C	O	0	0
			4	2	2		


- Molecule 8 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	1	Total	Ca	0	0
			1	1		
8	E	1	Total	Ca	0	0
			1	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.

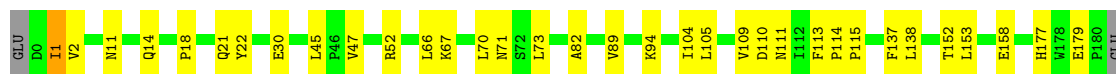
- Molecule 1: HLA class II histocompatibility antigen, DQ alpha 1 chain

Chain A: 



- Molecule 1: HLA class II histocompatibility antigen, DQ alpha 1 chain

Chain F: 



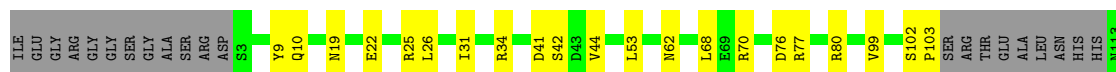
- Molecule 2: MHC class II HLA-DQ-beta-1

Chain B: 




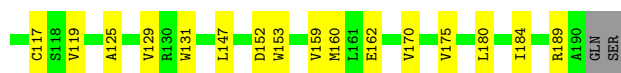
- Molecule 2: MHC class II HLA-DQ-beta-1

Chain G: 



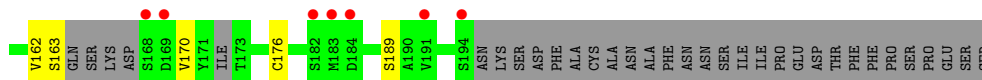
- Molecule 3: T-cell receptor, xpa5, alpha chain

Chain D: 

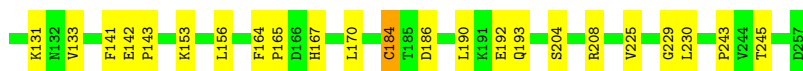




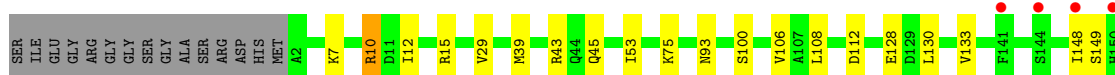
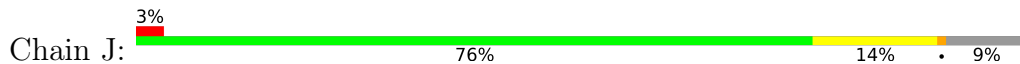
- Molecule 3: T-cell receptor, xpa5, alpha chain



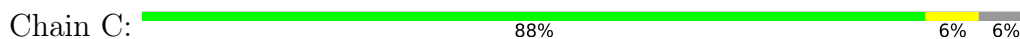
- Molecule 4: T-cell receptor, xpa5, beta chain



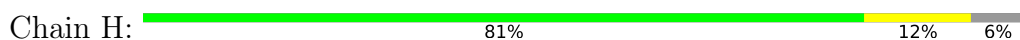
- Molecule 4: T-cell receptor, xpa5, beta chain



- Molecule 5: DQ2-glia-alpha1a peptide



- Molecule 5: DQ2-glia-alpha1a peptide



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	137.60Å 75.56Å 216.80Å 90.00° 103.76° 90.00°	Depositor
Resolution (Å)	45.42 – 3.10 45.42 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (45.42-3.10) 99.9 (45.42-3.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.58 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.216 , 0.258 0.220 , 0.255	Depositor DCC
R_{free} test set	1990 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	62.0	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 27.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.010 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-l$ 0.014 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-l$	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	12544	wwPDB-VP
Average B, all atoms (Å ²)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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: EDO, CA, 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	A	0.44	0/1478	0.65	0/2019
1	F	0.31	0/1486	0.50	0/2030
2	B	0.27	0/1465	0.44	0/1988
2	G	0.27	0/1506	0.46	0/2046
3	D	0.30	0/1544	0.47	0/2097
3	I	0.31	0/1245	0.48	0/1691
4	E	0.31	0/1963	0.53	0/2674
4	J	0.32	0/1890	0.53	0/2570
5	C	0.55	0/115	1.00	1/157 (0.6%)
5	H	0.46	0/115	0.51	0/157
All	All	0.32	0/12807	0.52	1/17429 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	3	PRO	N-CA-CB	-5.47	96.58	102.60

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	1436	0	1391	12	0
1	F	1444	0	1395	23	0
2	B	1434	0	1401	20	0
2	G	1473	0	1442	23	0
3	D	1512	0	1447	11	0
3	I	1221	0	1173	14	0
4	E	1916	0	1833	27	0
4	J	1846	0	1775	20	0
5	C	110	0	100	0	0
5	H	110	0	100	1	0
6	A	14	0	13	0	0
6	F	14	0	13	0	0
7	A	4	0	6	0	0
7	F	8	0	12	0	0
8	A	1	0	0	0	0
8	E	1	0	0	0	0
All	All	12544	0	12101	133	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 133 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)
3:I:68:ASN:OD1	3:I:74:GLU:N	2.04	0.89
1:F:11:ASN:HB3	1:F:66:LEU:HD11	1.61	0.81
3:D:54:ILE:HG13	3:D:68:ASN:ND2	2.00	0.77
4:J:229:GLY:H	4:J:245:THR:HG22	1.50	0.75
1:A:11:ASN:HB3	1:A:66:LEU:HD11	1.72	0.70

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	178/183 (97%)	170 (96%)	7 (4%)	1 (1%)	25	59
1	F	179/183 (98%)	171 (96%)	7 (4%)	1 (1%)	25	59
2	B	168/202 (83%)	161 (96%)	7 (4%)	0	100	100
2	G	175/202 (87%)	169 (97%)	6 (3%)	0	100	100
3	D	189/203 (93%)	182 (96%)	5 (3%)	2 (1%)	14	46
3	I	148/203 (73%)	144 (97%)	3 (2%)	1 (1%)	22	57
4	E	242/259 (93%)	232 (96%)	8 (3%)	2 (1%)	19	54
4	J	229/259 (88%)	222 (97%)	7 (3%)	0	100	100
5	C	13/16 (81%)	10 (77%)	3 (23%)	0	100	100
5	H	13/16 (81%)	11 (85%)	2 (15%)	0	100	100
All	All	1534/1726 (89%)	1472 (96%)	55 (4%)	7 (0%)	29	64

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	217	PRO
4	E	9	SER
1	F	18	PRO
1	A	18	PRO
3	D	92	PRO

5.3.2 Protein sidechains ⓘ

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	164/167 (98%)	159 (97%)	5 (3%)	41	71
1	F	165/167 (99%)	162 (98%)	3 (2%)	59	82
2	B	160/181 (88%)	155 (97%)	5 (3%)	40	70
2	G	164/181 (91%)	163 (99%)	1 (1%)	86	94
3	D	174/184 (95%)	168 (97%)	6 (3%)	37	69
3	I	140/184 (76%)	135 (96%)	5 (4%)	35	67

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	E	215/225 (96%)	208 (97%)	7 (3%)	38	69
4	J	207/225 (92%)	198 (96%)	9 (4%)	29	62
5	C	12/13 (92%)	12 (100%)	0	100	100
5	H	12/13 (92%)	11 (92%)	1 (8%)	11	38
All	All	1413/1540 (92%)	1371 (97%)	42 (3%)	41	71

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	I	30	THR
4	J	149	SER
3	I	38	TYR
3	I	170	VAL
4	J	230	LEU

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

Mol	Chain	Res	Type
3	D	68	ASN
4	E	167	HIS
4	J	44	GLN
3	I	44	GLN
1	F	177	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	$\# Z > 2$
7	EDO	F	202	-	3,3,3	0.10	0	2,2,2	0.14	0
6	NAG	F	201	1	14,14,15	0.64	1 (7%)	17,19,21	0.45	0
6	NAG	A	201	1	14,14,15	0.57	0	17,19,21	0.81	0
7	EDO	F	203	-	3,3,3	0.46	0	2,2,2	0.37	0
7	EDO	A	202	-	3,3,3	0.46	0	2,2,2	0.32	0

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
7	EDO	F	202	-	-	0/1/1/1	-
6	NAG	F	201	1	-	4/6/23/26	0/1/1/1
6	NAG	A	201	1	-	2/6/23/26	0/1/1/1
7	EDO	F	203	-	-	0/1/1/1	-
7	EDO	A	202	-	-	0/1/1/1	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	F	201	NAG	O5-C1	-2.10	1.40	1.43

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	F	201	NAG	O5-C5-C6-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
6	F	201	NAG	C4-C5-C6-O6
6	F	201	NAG	C8-C7-N2-C2
6	F	201	NAG	O7-C7-N2-C2
6	A	201	NAG	O5-C5-C6-O6

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	180/183 (98%)	-0.10	7 (3%) 39 20	34, 58, 105, 131	0
1	F	181/183 (98%)	-0.26	0 100 100	30, 46, 80, 137	0
2	B	174/202 (86%)	-0.01	6 (3%) 45 24	34, 62, 119, 149	0
2	G	179/202 (88%)	-0.23	0 100 100	34, 52, 92, 111	0
3	D	193/203 (95%)	-0.04	1 (0%) 91 81	38, 67, 108, 144	1 (0%)
3	I	156/203 (76%)	0.05	7 (4%) 33 16	50, 72, 121, 143	0
4	E	244/259 (94%)	-0.20	0 100 100	34, 48, 82, 121	0
4	J	235/259 (90%)	0.17	9 (3%) 40 20	40, 72, 120, 154	1 (0%)
5	C	15/16 (93%)	-0.26	0 100 100	37, 51, 106, 113	0
5	H	15/16 (93%)	-0.10	0 100 100	39, 48, 110, 125	0
All	All	1572/1726 (91%)	-0.08	30 (1%) 66 46	30, 58, 111, 154	2 (0%)

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	J	148	ILE	4.6
2	B	170	VAL	3.9
2	B	171	TYR	3.6
1	A	126	HIS	3.6
2	B	169	ASP	3.5

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
8	CA	A	203	1/1	0.53	0.18	30,30,30,30	0
7	EDO	F	202	4/4	0.76	0.40	49,51,56,71	0
7	EDO	A	202	4/4	0.81	0.32	45,45,60,65	0
7	EDO	F	203	4/4	0.82	0.24	47,49,56,62	0
6	NAG	F	201	14/15	0.83	0.35	66,94,110,113	0
6	NAG	A	201	14/15	0.85	0.20	73,97,116,142	0
8	CA	E	301	1/1	0.96	0.06	30,30,30,30	0

6.5 Other polymers [i](#)

There are no such residues in this entry.