wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 09:13 pm GMT

|		
PDB ID	:	2XQ2
Title	:	Structure of the K294A mutant of vSGLT
Authors	:	Watanabe, A.; Choe, S.; Chaptal, V.; Rosenberg, J.M.; Wright, E.M.; Grabe, M.; Abramson, J.
Deposited on	:	2010-09-01
Resolution	:	2.73 Å(reported)

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 2.73 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.
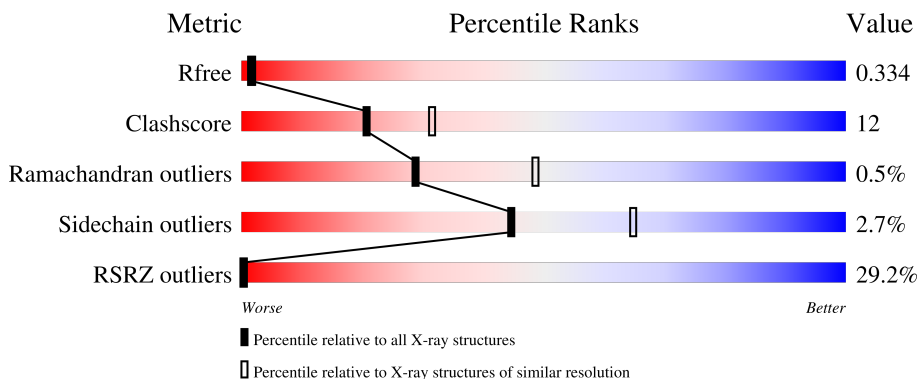
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

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	593	18% / 69% 21% · 9%
2	B	593	33% / 72% 16% · 12%

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
3	PEG	A	1574	-	-	-	X
3	PEG	A	1575	-	-	-	X
3	PEG	A	1576	-	-	-	X

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 8090 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 SODIUM/GLUCOSE COTRANSPORTER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	538	4107	2762	612	711	22	0	0	0

There are 53 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	294	ALA	LYS	engineered mutation	UNP P96169
A	411	ALA	CYS	engineered mutation	UNP P96169
A	423	CYS	ALA	engineered mutation	UNP P96169
A	544	VAL	-	expression tag	UNP P96169
A	545	ASN	-	expression tag	UNP P96169
A	546	ALA	-	expression tag	UNP P96169
A	547	ASP	-	expression tag	UNP P96169
A	548	ALA	-	expression tag	UNP P96169
A	549	GLU	-	expression tag	UNP P96169
A	550	ILE	-	expression tag	UNP P96169
A	551	THR	-	expression tag	UNP P96169
A	552	LEU	-	expression tag	UNP P96169
A	553	ILE	-	expression tag	UNP P96169
A	554	ILE	-	expression tag	UNP P96169
A	555	PHE	-	expression tag	UNP P96169
A	556	GLY	-	expression tag	UNP P96169
A	557	VAL	-	expression tag	UNP P96169
A	558	MET	-	expression tag	UNP P96169
A	559	ALA	-	expression tag	UNP P96169
A	560	GLY	-	expression tag	UNP P96169
A	561	VAL	-	expression tag	UNP P96169
A	562	ILE	-	expression tag	UNP P96169
A	563	GLY	-	expression tag	UNP P96169
A	564	THR	-	expression tag	UNP P96169
A	565	ILE	-	expression tag	UNP P96169
A	566	LEU	-	expression tag	UNP P96169
A	567	LEU	-	expression tag	UNP P96169

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	568	ILE	-	expression tag	UNP P96169
A	569	SER	-	expression tag	UNP P96169
A	570	TYR	-	expression tag	UNP P96169
A	571	GLY	-	expression tag	UNP P96169
A	572	ILE	-	expression tag	UNP P96169
A	573	LYS	-	expression tag	UNP P96169
A	574	LYS	-	expression tag	UNP P96169
A	575	LEU	-	expression tag	UNP P96169
A	576	ILE	-	expression tag	UNP P96169
A	577	LYS	-	expression tag	UNP P96169
A	578	ALA	-	expression tag	UNP P96169
A	579	SER	-	expression tag	UNP P96169
A	580	TYR	-	expression tag	UNP P96169
A	581	LYS	-	expression tag	UNP P96169
A	582	SER	-	expression tag	UNP P96169
A	583	GLY	-	expression tag	UNP P96169
A	584	GLY	-	expression tag	UNP P96169
A	585	SER	-	expression tag	UNP P96169
A	586	PRO	-	expression tag	UNP P96169
A	587	GLY	-	expression tag	UNP P96169
A	588	HIS	-	expression tag	UNP P96169
A	589	HIS	-	expression tag	UNP P96169
A	590	HIS	-	expression tag	UNP P96169
A	591	HIS	-	expression tag	UNP P96169
A	592	HIS	-	expression tag	UNP P96169
A	593	HIS	-	expression tag	UNP P96169

- Molecule 2 is a protein called SODIUM/GLUCOSE COTRANSPORTER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	522	3957	2651	594	691	21	0	0	0

There are 53 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	294	ALA	LYS	engineered mutation	UNP P96169
B	411	ALA	CYS	engineered mutation	UNP P96169
B	423	CYS	ALA	engineered mutation	UNP P96169
B	544	VAL	-	expression tag	UNP P96169
B	545	ASN	-	expression tag	UNP P96169
B	546	ALA	-	expression tag	UNP P96169

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	547	ASP	-	expression tag	UNP P96169
B	548	ALA	-	expression tag	UNP P96169
B	549	GLU	-	expression tag	UNP P96169
B	550	ILE	-	expression tag	UNP P96169
B	551	THR	-	expression tag	UNP P96169
B	552	LEU	-	expression tag	UNP P96169
B	553	ILE	-	expression tag	UNP P96169
B	554	ILE	-	expression tag	UNP P96169
B	555	PHE	-	expression tag	UNP P96169
B	556	GLY	-	expression tag	UNP P96169
B	557	VAL	-	expression tag	UNP P96169
B	558	MET	-	expression tag	UNP P96169
B	559	ALA	-	expression tag	UNP P96169
B	560	GLY	-	expression tag	UNP P96169
B	561	VAL	-	expression tag	UNP P96169
B	562	ILE	-	expression tag	UNP P96169
B	563	GLY	-	expression tag	UNP P96169
B	564	THR	-	expression tag	UNP P96169
B	565	ILE	-	expression tag	UNP P96169
B	566	LEU	-	expression tag	UNP P96169
B	567	LEU	-	expression tag	UNP P96169
B	568	ILE	-	expression tag	UNP P96169
B	569	SER	-	expression tag	UNP P96169
B	570	TYR	-	expression tag	UNP P96169
B	571	GLY	-	expression tag	UNP P96169
B	572	ILE	-	expression tag	UNP P96169
B	573	LYS	-	expression tag	UNP P96169
B	574	LYS	-	expression tag	UNP P96169
B	575	LEU	-	expression tag	UNP P96169
B	576	ILE	-	expression tag	UNP P96169
B	577	LYS	-	expression tag	UNP P96169
B	578	ALA	-	expression tag	UNP P96169
B	579	SER	-	expression tag	UNP P96169
B	580	TYR	-	expression tag	UNP P96169
B	581	LYS	-	expression tag	UNP P96169
B	582	SER	-	expression tag	UNP P96169
B	583	GLY	-	expression tag	UNP P96169
B	584	GLY	-	expression tag	UNP P96169
B	585	SER	-	expression tag	UNP P96169
B	586	PRO	-	expression tag	UNP P96169
B	587	GLY	-	expression tag	UNP P96169
B	588	HIS	-	expression tag	UNP P96169

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	589	HIS	-	expression tag	UNP P96169
B	590	HIS	-	expression tag	UNP P96169
B	591	HIS	-	expression tag	UNP P96169
B	592	HIS	-	expression tag	UNP P96169
B	593	HIS	-	expression tag	UNP P96169

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).
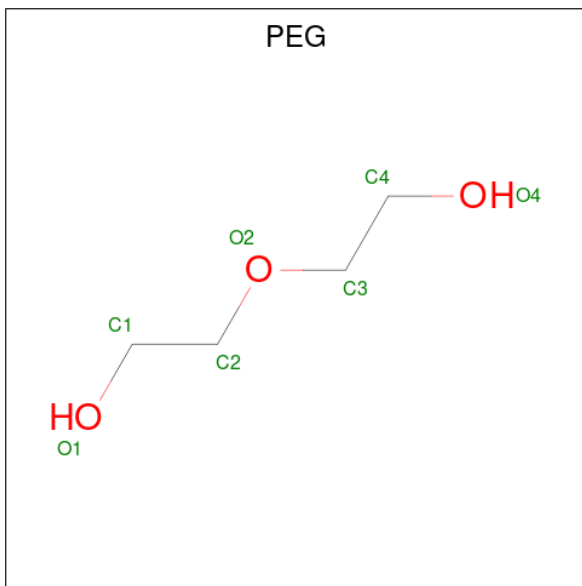
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 7	C 4	O 3	0	0
3	A	1	Total 7	C 4	O 3	0	0
3	A	1	Total 7	C 4	O 3	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total 5	O 5	0	0

3 Residue-property plots ⓘ

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: SODIUM/GLUCOSE COTRANSPORTER
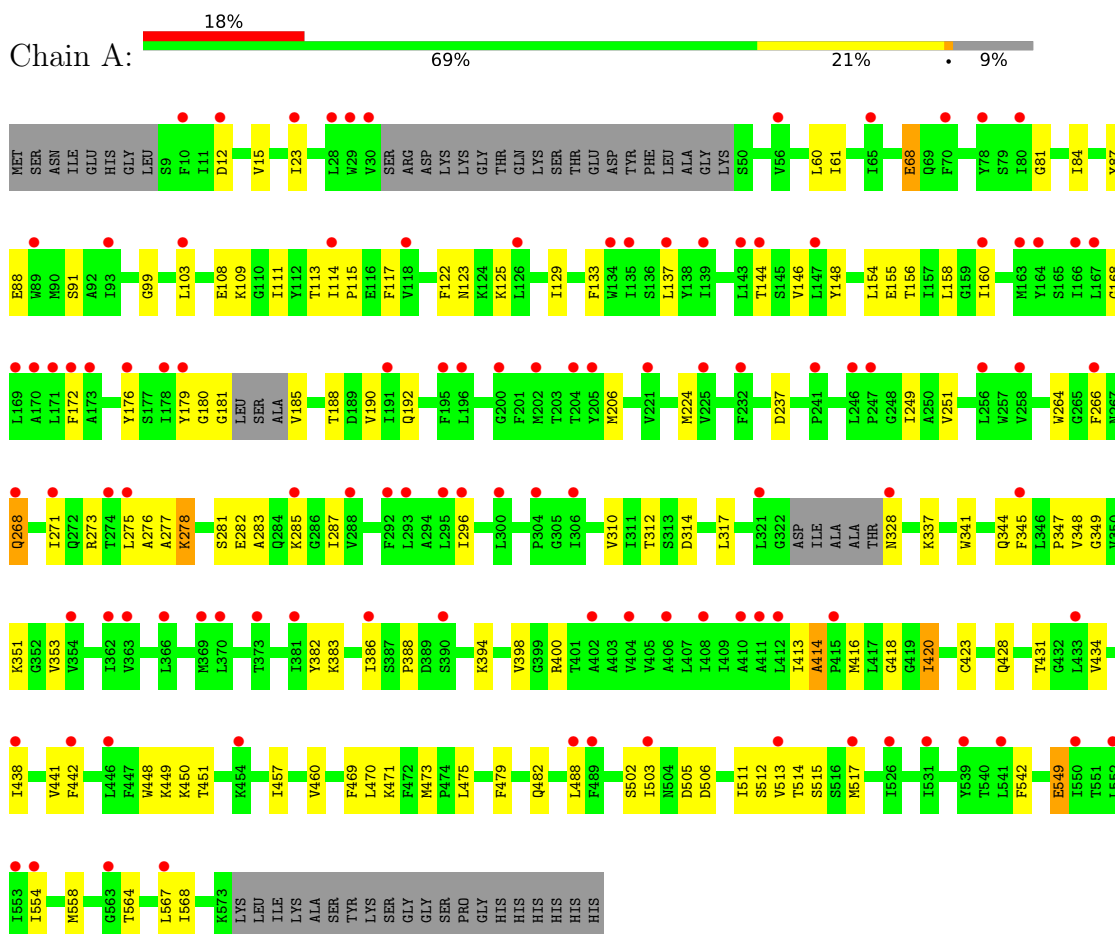
• Molecule 2: SODIUM/GLUCOSE COTRANSPORTER
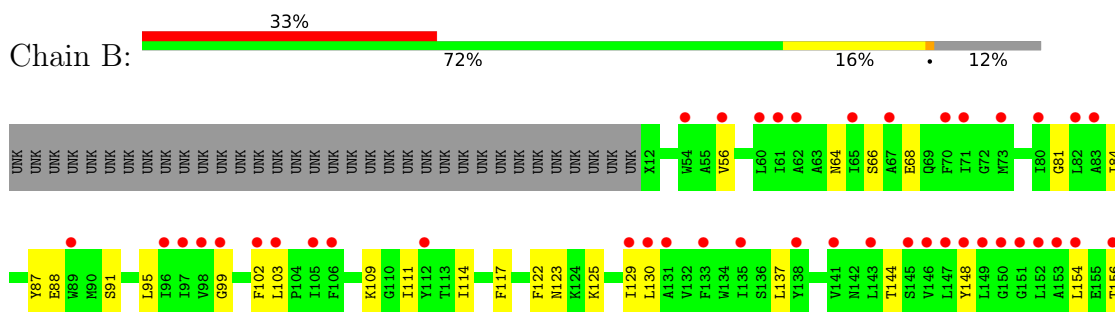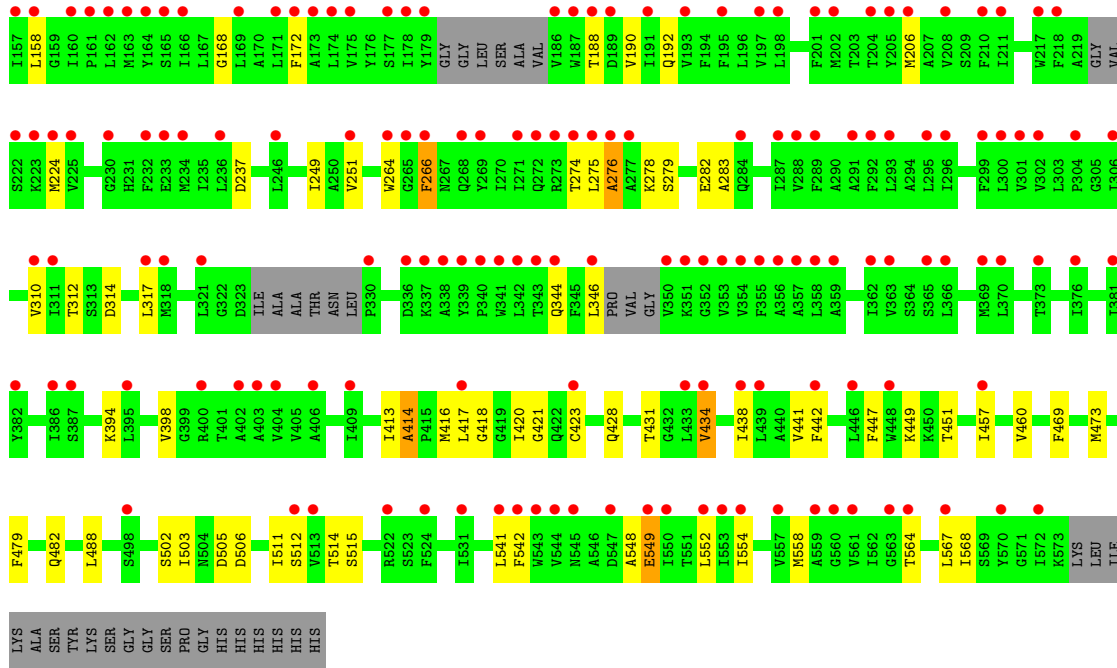
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	85.30Å 112.68Å 238.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.01 – 2.73 68.01 – 2.69	Depositor EDS
% Data completeness (in resolution range)	(Not available) (68.01-2.73) 96.5 (68.01-2.69)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 2.69Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, R_{free}	0.251 , 0.274 0.287 , 0.334	Depositor DCC
R_{free} test set	3193 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	86.6	Xtriage
Anisotropy	0.316	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 53.6	EDS
L-test for twinning[2]	$< |L| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	8090	wwPDB-VP
Average B, all atoms (Å^2)	112.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.08% 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: PEG

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	0.51	0/4210	0.72	2/5741 (0.0%)
2	B	0.46	0/3973	0.70	2/5413 (0.0%)
All	All	0.48	0/8183	0.71	4/11154 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	276	ALA	CB-CA-C	7.57	121.45	110.10
2	B	266	PHE	N-CA-C	6.79	129.35	111.00
1	A	268	GLN	CB-CA-C	5.43	121.26	110.40
1	A	348	VAL	N-CA-C	5.06	124.67	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	4107	0	4252	122	0
2	B	3957	0	4022	89	0
3	A	21	0	30	0	0
4	A	5	0	0	0	0
All	All	8090	0	8304	201	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 12.

The worst 5 of 201 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:A:114:ILE:CG2	1:A:115:PRO:HD3	1.82	1.09
1:A:502:SER:O	2:B:515:SER:HB2	1.64	0.98
2:B:148:TYR:CD2	2:B:417:LEU:HD22	2.00	0.96
1:A:81:GLY:O	1:A:84:ILE:HG22	1.65	0.96
2:B:81:GLY:O	2:B:84:ILE:HG22	1.65	0.95

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	530/593 (89%)	505 (95%)	22 (4%)	3 (1%)	25	44
2	B	496/593 (84%)	469 (95%)	25 (5%)	2 (0%)	34	55
All	All	1026/1186 (86%)	974 (95%)	47 (5%)	5 (0%)	29	48

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	278	LYS
2	B	276	ALA
1	A	276	ALA
1	A	414	ALA
2	B	414	ALA

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	434/478 (91%)	421 (97%)	13 (3%)	41	61
2	B	410/436 (94%)	400 (98%)	10 (2%)	49	68
All	All	844/914 (92%)	821 (97%)	23 (3%)	44	65

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	249	ILE
2	B	312	THR
2	B	279	SER
2	B	434	VAL
1	A	296	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	428	GLN
2	B	525	ASN
1	A	428	GLN
1	A	525	ASN
2	B	245	ASN

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 ⓘ

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	#$|Z| > 2$
3	PEG	A	1576	-	6,6,6	0.52	0	5,5,5	0.94	0
3	PEG	A	1575	-	6,6,6	0.56	0	5,5,5	0.73	0
3	PEG	A	1574	-	6,6,6	0.64	0	5,5,5	0.69	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
3	PEG	A	1576	-	-	2/4/4/4	-
3	PEG	A	1575	-	-	2/4/4/4	-
3	PEG	A	1574	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1575	PEG	O1-C1-C2-O2
3	A	1576	PEG	O1-C1-C2-O2
3	A	1575	PEG	C1-C2-O2-C3
3	A	1574	PEG	O2-C3-C4-O4
3	A	1576	PEG	C1-C2-O2-C3

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	27:UNK	C	51:LEU	N	23.43

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, 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	538/593 (90%)	1.22	107 (19%)	1	1	44, 88, 129, 164	0
2	B	506/593 (85%)	2.02	198 (39%)	0	0	67, 131, 188, 241	0
All	All	1044/1186 (88%)	1.61	305 (29%)	0	0	44, 104, 174, 241	0

The worst 5 of 305 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	353	VAL	13.3
2	B	354	VAL	9.7
2	B	218	PHE	9.7
2	B	154	LEU	8.9
2	B	179	TYR	8.9

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, 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	PEG	A	1574	7/7	0.29	0.50	121,207,245,253	0
3	PEG	A	1575	7/7	0.42	0.46	55,128,205,249	0
3	PEG	A	1576	7/7	0.65	0.59	45,63,205,221	0

6.5 Other polymers ⓘ

There are no such residues in this entry.