wwPDB X-ray Structure Validation Summary Report (i)

Jan 2, 2024 – 06:22 pm GMT

PDB ID	:	4UOU
Title	:	Crystal Structure of Fucose binding lectin from Aspergillus Fumigatus (AFL) - apo-form
Authors	:	Houser, J.; Komarek, J.; Cioci, G.; Varrot, A.; Imberty, A.; Wimmerova, M.
Deposited on	:	2014-06-10
Resolution	:	2.40 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 (i)) were used in the production of this report:

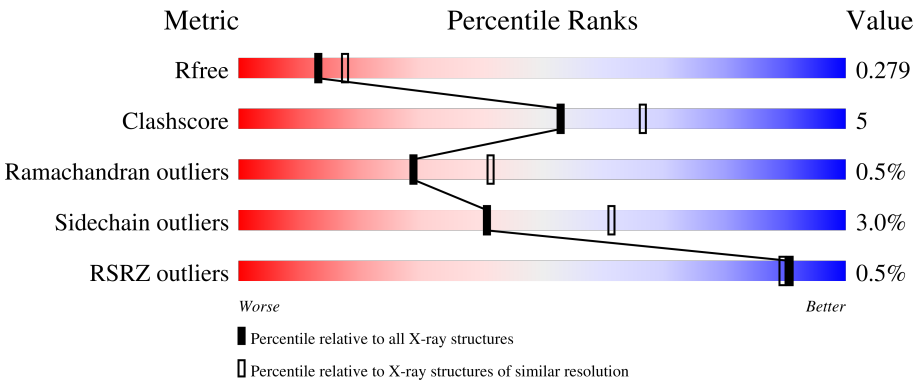
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	314	% 86% 13% •
1	B	314	% 88% 12% •
1	C	314	84% 14% •
1	D	314	85% 15% •

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
2	PEG	A	900	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10295 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.

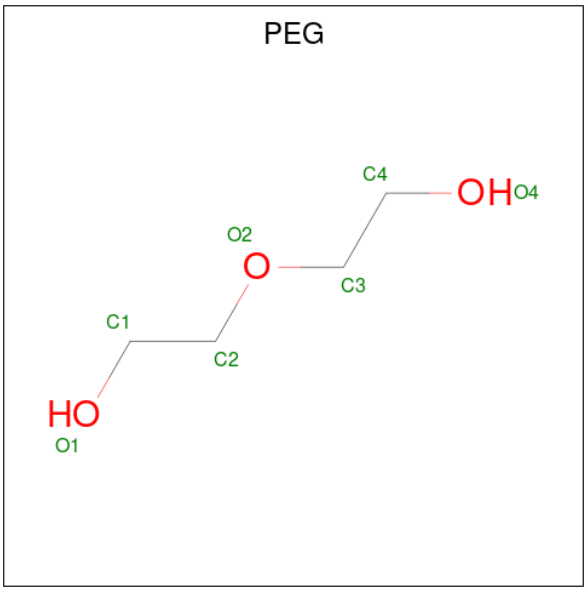
- Molecule 1 is a protein called FUCOSE-SPECIFIC LECTIN FLEA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	314	2437	1554	420	459	4	0	0	0
1	B	314	2439	1555	419	461	4	0	0	0
1	C	314	2437	1554	418	461	4	0	0	0
1	D	314	2442	1556	421	461	4	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	SER	LEU	variant	UNP Q4WW81
A	111	CSX	ARG	variant	UNP Q4WW81
B	20	SER	LEU	variant	UNP Q4WW81
B	111	CSX	ARG	variant	UNP Q4WW81
C	20	SER	LEU	variant	UNP Q4WW81
C	111	CSX	ARG	variant	UNP Q4WW81
D	20	SER	LEU	variant	UNP Q4WW81
D	111	CSX	ARG	variant	UNP Q4WW81

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	7	4	3	0	0
			Total	C	O		
2	D	1	7	4	3	0	0

- Molecule 3 is water.

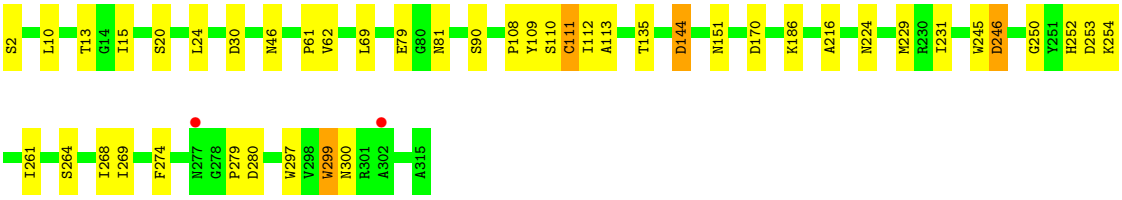
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	150	150	150	0	0
			Total	O		
3	B	138	138	138	0	0
			Total	O		
3	C	123	123	123	0	0
			Total	O		
3	D	115	115	115	0	0

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.

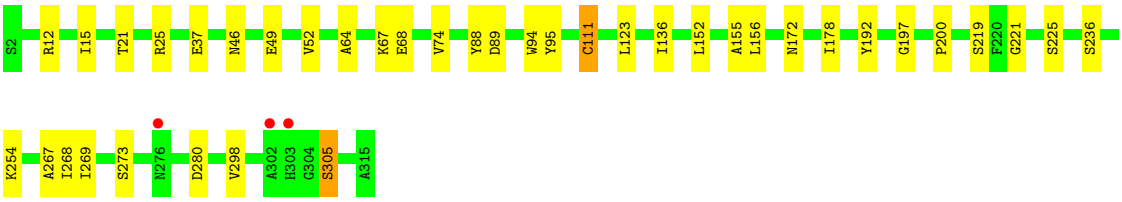
• Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain A:



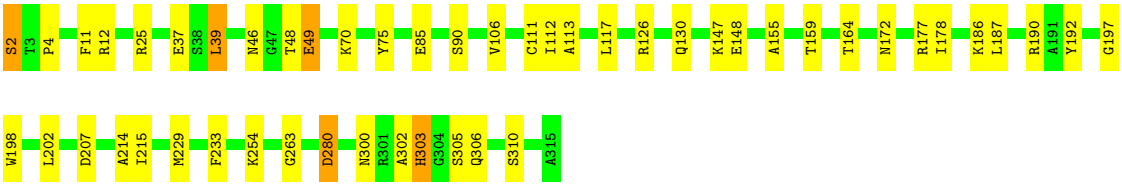
• Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain B:



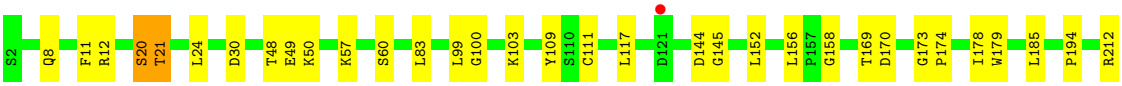
• Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain C:



• Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain D:



S225
M229
R230
I231
N235
W245
D246
H252
D253
K254
G255
T256
D280
T289
Y290
W297
V298
W299
S310
A315

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.63Å 140.22Å 78.79Å 90.00° 92.15° 90.00°	Depositor
Resolution (Å)	78.74 – 2.40 41.43 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.3 (78.74-2.40) 99.3 (41.43-2.40)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	8.92 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.8.0071	Depositor
R, R_{free}	0.217 , 0.279 0.224 , 0.279	Depositor DCC
R_{free} test set	2008 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	12.6	Xtriage
Anisotropy	0.175	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , -0.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.106 for h,-k,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	10295	wwPDB-VP
Average B, all atoms (Å^2)	11.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.86 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.6597e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CSX, 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	0.70	0/2500	0.80	2/3407 (0.1%)
1	B	0.71	0/2502	0.83	0/3410
1	C	0.71	0/2500	0.81	1/3408 (0.0%)
1	D	0.70	0/2505	0.79	0/3414
All	All	0.71	0/10007	0.81	3/13639 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	30	ASP	CB-CG-OD1	6.10	123.79	118.30
1	C	39	LEU	CA-CB-CG	5.86	128.77	115.30
1	A	246	ASP	CB-CG-OD1	5.83	123.55	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	20	SER	Peptide

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	2437	0	2319	31	0
1	B	2439	0	2317	22	0
1	C	2437	0	2312	27	0
1	D	2442	0	2321	24	0
2	A	7	0	10	4	0
2	D	7	0	10	1	0
3	A	150	0	0	7	0
3	B	138	0	0	5	0
3	C	123	0	0	2	0
3	D	115	0	0	1	0
All	All	10295	0	9289	102	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 102 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:79:GLU:HG2	3:A:2040:HOH:O	1.85	0.75
1:B:200:PRO:HD2	3:B:2097:HOH:O	1.87	0.72
1:A:280:ASP:OD1	1:A:300:ASN:HB3	1.94	0.67
1:C:148:GLU:OE1	3:C:2059:HOH:O	2.13	0.67
1:A:253:ASP:OD2	3:A:2117:HOH:O	2.14	0.66

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	311/314 (99%)	296 (95%)	14 (4%)	1 (0%)	41	55
1	B	311/314 (99%)	296 (95%)	13 (4%)	2 (1%)	25	36
1	C	311/314 (99%)	298 (96%)	12 (4%)	1 (0%)	41	55
1	D	311/314 (99%)	295 (95%)	14 (4%)	2 (1%)	25	36
All	All	1244/1256 (99%)	1185 (95%)	53 (4%)	6 (0%)	29	41

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	46	ASN
1	D	21	THR
1	C	49	GLU
1	A	46	ASN
1	D	290	TYR

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	246/248 (99%)	241 (98%)	5 (2%)	55	74
1	B	246/248 (99%)	238 (97%)	8 (3%)	38	57
1	C	246/248 (99%)	235 (96%)	11 (4%)	27	44
1	D	247/248 (100%)	241 (98%)	6 (2%)	49	68
All	All	985/992 (99%)	955 (97%)	30 (3%)	41	61

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	4	PRO
1	D	254	LYS
1	C	202	LEU
1	D	280	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	24	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	276	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
1	CSX	C	111	1	3,6,7	0.59	0	1,6,8	3.31	1 (100%)
1	CSX	A	111	1	3,6,7	1.46	0	1,6,8	4.00	1 (100%)
1	CSX	D	111	1	3,6,7	0.71	0	1,6,8	2.36	1 (100%)
1	CSX	B	111	1	3,6,7	0.91	0	1,6,8	2.28	1 (100%)

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
1	CSX	C	111	1	-	1/1/5/7	-
1	CSX	A	111	1	-	1/1/5/7	-
1	CSX	D	111	1	-	1/1/5/7	-
1	CSX	B	111	1	-	1/1/5/7	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	111	CSX	CA-CB-SG	-4.00	104.62	113.36
1	C	111	CSX	CA-CB-SG	-3.31	106.13	113.36
1	D	111	CSX	CA-CB-SG	-2.36	108.21	113.36
1	B	111	CSX	CA-CB-SG	-2.28	108.37	113.36

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	111	CSX	N-CA-CB-SG
1	B	111	CSX	N-CA-CB-SG
1	C	111	CSX	N-CA-CB-SG
1	D	111	CSX	N-CA-CB-SG

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	111	CSX	1	0
1	B	111	CSX	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

2 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	$\#\|Z\| > 2$
2	PEG	A	900	-	6,6,6	0.69	0	5,5,5	1.41	1 (20%)
2	PEG	D	900	-	6,6,6	0.67	0	5,5,5	0.47	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
2	PEG	A	900	-	-	2/4/4/4	-
2	PEG	D	900	-	-	3/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	900	PEG	O2-C3-C4	2.39	120.56	110.07

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	900	PEG	C4-C3-O2-C2
2	D	900	PEG	O1-C1-C2-O2
2	D	900	PEG	O2-C3-C4-O4
2	A	900	PEG	C1-C2-O2-C3
2	D	900	PEG	C4-C3-O2-C2

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	900	PEG	4	0
2	D	900	PEG	1	0

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	313/314 (99%)	-0.43	2 (0%)	89	88	3, 9, 25, 49	0
1	B	313/314 (99%)	-0.44	3 (0%)	82	80	3, 9, 27, 44	0
1	C	313/314 (99%)	-0.44	0	100	100	3, 9, 25, 38	0
1	D	313/314 (99%)	-0.44	1 (0%)	94	93	3, 9, 24, 35	0
All	All	1252/1256 (99%)	-0.44	6 (0%)	91	89	3, 9, 25, 49	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	302	ALA	2.7
1	A	302	ALA	2.5
1	B	276	ASN	2.3
1	A	277	ASN	2.2
1	D	121	ASP	2.1

6.2 Non-standard residues in protein, DNA, RNA chains (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
1	CSX	B	111	7/8	0.90	0.15	8,8,11,11	0
1	CSX	C	111	7/8	0.93	0.13	10,11,15,16	0
1	CSX	D	111	7/8	0.93	0.12	9,11,15,17	0
1	CSX	A	111	7/8	0.95	0.10	7,7,7,8	0

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
2	PEG	D	900	7/7	0.90	0.18	18,19,20,20	0
2	PEG	A	900	7/7	0.91	0.18	9,9,9,10	0

6.5 Other polymers (i)

There are no such residues in this entry.