Full wwPDB X-ray Structure Validation Report ⓘ

Nov 6, 2023 – 04:06 PM EST

PDB ID	:	8EGK
Title	:	Re-refinement of Crystal Structure of NosGet3d, the All4481 protein from Nostoc sp. PCC 7120
Authors	:	Barlow, A.N.; Manu, M.S.; Ramasamy, S.; Clemons Jr., W.M.
Deposited on	:	2022-09-12
Resolution	:	1.98 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

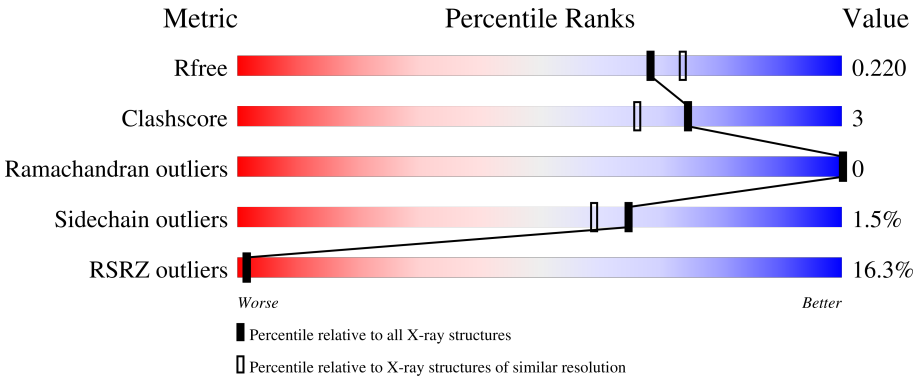
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.98 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	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	374	13% 89% 6% 5%
1	B	374	17% 84% 10% • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11536 atoms, of which 5663 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.

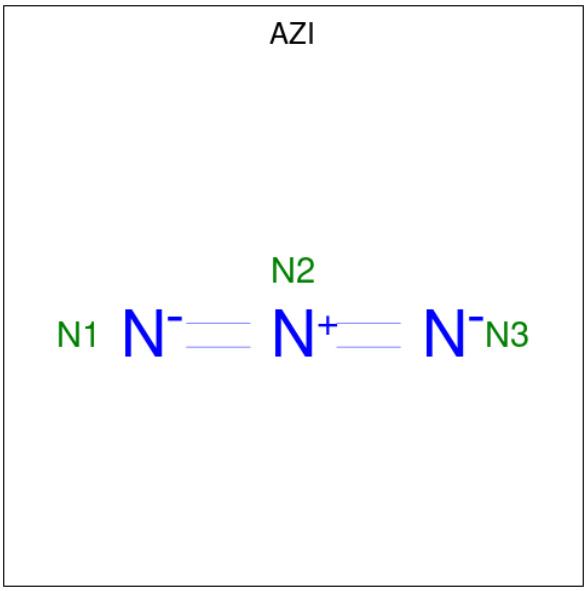
- Molecule 1 is a protein called All4481 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	355	Total	C	H	N	O	Se	0	1	0
			5599	1784	2833	464	516	2			
1	B	355	Total	C	H	N	O	Se	0	0	0
			5592	1781	2830	463	516	2			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	127	THR	ALA	engineered mutation	UNP Q8YNT0
A	367	LEU	-	expression tag	UNP Q8YNT0
A	368	GLU	-	expression tag	UNP Q8YNT0
A	369	HIS	-	expression tag	UNP Q8YNT0
A	370	HIS	-	expression tag	UNP Q8YNT0
A	371	HIS	-	expression tag	UNP Q8YNT0
A	372	HIS	-	expression tag	UNP Q8YNT0
A	373	HIS	-	expression tag	UNP Q8YNT0
A	374	HIS	-	expression tag	UNP Q8YNT0
B	127	THR	ALA	engineered mutation	UNP Q8YNT0
B	367	LEU	-	expression tag	UNP Q8YNT0
B	368	GLU	-	expression tag	UNP Q8YNT0
B	369	HIS	-	expression tag	UNP Q8YNT0
B	370	HIS	-	expression tag	UNP Q8YNT0
B	371	HIS	-	expression tag	UNP Q8YNT0
B	372	HIS	-	expression tag	UNP Q8YNT0
B	373	HIS	-	expression tag	UNP Q8YNT0
B	374	HIS	-	expression tag	UNP Q8YNT0

- Molecule 2 is AZIDE ION (three-letter code: AZI) (formula: N_3).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total N 3 3	0	0
2	B	1	Total N 3 3	0	0

- Molecule 3 is water.

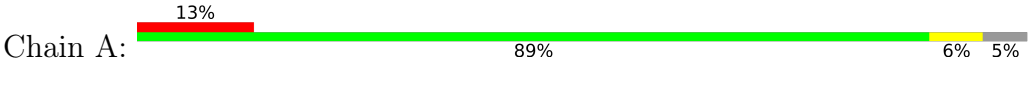
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	199	Total O 199 199	0	0
3	B	140	Total O 140 140	0	0

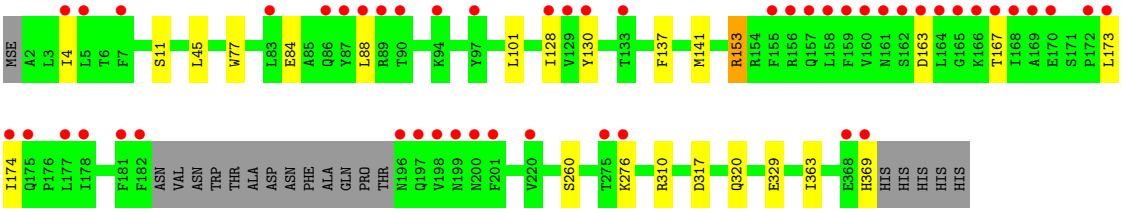
3 Residue-property plots (i)

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: All4481 protein

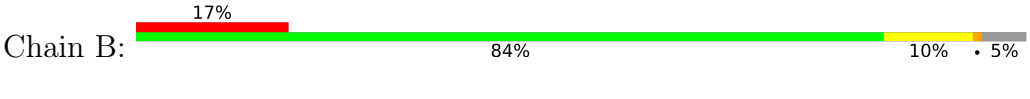
Chain A:

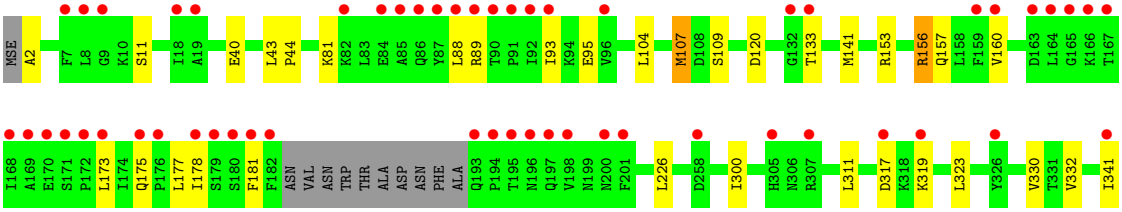


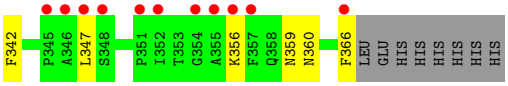


• Molecule 1: All4481 protein

Chain B:







4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	124.17Å 55.55Å 122.36Å 90.00° 98.87° 90.00°	Depositor
Resolution (Å)	27.52 – 1.98 27.52 – 1.98	Depositor EDS
% Data completeness (in resolution range)	94.3 (27.52-1.98) 94.4 (27.52-1.98)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.80 (at 1.98Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487, REFMAC Unknown, CNS	Depositor
R, R_{free}	0.189 , 0.221 0.189 , 0.220	Depositor DCC
R_{free} test set	2745 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	28.1	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 53.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11536	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.85% 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: AZI

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	0.40	0/2826	0.57	0/3844
1	B	0.36	0/2816	0.55	0/3832
All	All	0.38	0/5642	0.56	0/7676

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	A	0	1
1	B	0	2
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	153	ARG	Sidechain
1	B	153	ARG	Sidechain
1	B	156	ARG	Sidechain

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	2766	2833	2827	15	0
1	B	2762	2830	2830	25	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	199	0	0	0	0
3	B	140	0	0	0	0
All	All	5873	5663	5657	36	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 3.

All (36) 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:163:ASP:O	1:A:167:THR:HG23	1.96	0.66
1:B:323:LEU:HD12	1:B:332:VAL:HG22	1.87	0.56
1:B:330:VAL:HG13	1:B:341:ILE:HD12	1.89	0.55
1:B:330:VAL:CG1	1:B:341:ILE:HD12	2.36	0.55
1:A:153:ARG:HG2	1:A:153:ARG:HH11	1.72	0.53
1:A:130:TYR:CE2	1:A:141:MSE:HE1	2.45	0.52
1:B:88:LEU:HD12	1:B:89:ARG:N	2.25	0.52
1:A:137:PHE:CE2	1:B:133:THR:HG21	2.46	0.51
1:A:84:GLU:CG	1:A:88:LEU:HD12	2.42	0.50
1:B:323:LEU:CD1	1:B:332:VAL:HG22	2.41	0.49
1:A:174:ILE:HG22	1:B:177:LEU:HD11	1.96	0.48
1:B:88:LEU:HD12	1:B:89:ARG:H	1.78	0.47
1:A:153:ARG:NH2	1:A:329:GLU:OE2	2.48	0.47
1:B:109:SER:HB3	1:B:141:MSE:HG3	1.97	0.47
1:B:175:GLN:OE1	1:B:178:ILE:HD11	2.13	0.47
1:A:84:GLU:HG3	1:A:88:LEU:HD12	1.98	0.46
1:B:157:GLN:HA	1:B:157:GLN:OE1	2.16	0.46
1:B:156:ARG:O	1:B:160:VAL:HG13	2.17	0.45
1:A:45:LEU:HD23	1:B:226:LEU:HD23	1.99	0.45
1:B:95:GLU:HA	1:B:95:GLU:OE1	2.17	0.44
1:B:2:ALA:N	1:B:120:ASP:OD1	2.52	0.43
1:B:317:ASP:OD1	1:B:319:LYS:N	2.49	0.43
1:B:347:LEU:HD22	1:B:366:PHE:CG	2.53	0.43
1:A:310:ARG:NH1	1:A:363:ILE:HD11	2.34	0.43
1:B:89:ARG:HG2	1:B:181:PHE:HA	2.01	0.43
1:B:359:ASN:O	1:B:360:ASN:HB2	2.19	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:84:GLU:HG2	1:A:88:LEU:HD12	2.01	0.42
1:B:81:LYS:HG2	1:B:93:ILE:HD12	2.01	0.42
1:B:317:ASP:OD1	1:B:317:ASP:C	2.58	0.41
1:B:104:LEU:O	1:B:107:MSE:HB3	2.21	0.41
1:B:43:LEU:HB3	1:B:44:PRO:HD3	2.02	0.41
1:A:4:ILE:HD13	1:A:128:ILE:HB	2.03	0.41
1:A:173:LEU:HD23	1:B:173:LEU:HD23	2.03	0.41
1:A:77:TRP:CD2	1:A:101:LEU:HD11	2.56	0.41
1:A:317:ASP:OD1	1:A:320:GLN:HG3	2.21	0.41
1:B:300:ILE:HD12	1:B:311:LEU:HD21	2.02	0.40

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	352/374 (94%)	344 (98%)	8 (2%)	0	100	100
1	B	351/374 (94%)	344 (98%)	7 (2%)	0	100	100
All	All	703/748 (94%)	688 (98%)	15 (2%)	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	303/316 (96%)	299 (99%)	4 (1%)	69	64
1	B	302/316 (96%)	297 (98%)	5 (2%)	60	53
All	All	605/632 (96%)	596 (98%)	9 (2%)	65	59

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	11	SER
1	A	260	SER
1	A	276	LYS
1	A	369	HIS
1	B	11	SER
1	B	40	GLU
1	B	107	MSE
1	B	342	PHE
1	B	356	LYS

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	306	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)

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)

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	AZI	B	1101	-	0,2,2	-	-	0,1,1	-	-
2	AZI	A	1101	-	0,2,2	-	-	0,1,1	-	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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	A	353/374 (94%)	0.55	50 (14%)	2 2	18, 30, 89, 113	0
1	B	353/374 (94%)	0.91	65 (18%)	1 1	24, 38, 94, 107	0
All	All	706/748 (94%)	0.73	115 (16%)	1 1	18, 34, 93, 113	0

All (115) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	85	ALA	9.9
1	A	169	ALA	9.7
1	B	176	PRO	9.1
1	B	88	LEU	8.3
1	B	178	ILE	8.2
1	B	87	TYR	7.8
1	A	159	PHE	7.5
1	B	182	PHE	7.3
1	A	87	TYR	7.0
1	B	164	LEU	6.8
1	B	168	ILE	6.6
1	A	173	LEU	6.6
1	B	172	PRO	6.4
1	A	163	ASP	6.1
1	B	194	PRO	5.7
1	B	169	ALA	5.6
1	B	180	SER	5.5
1	B	163	ASP	5.4
1	A	174	ILE	5.4
1	B	181	PHE	5.3
1	A	161	ASN	5.1
1	A	160	VAL	5.1
1	B	173	LEU	5.0
1	A	165	GLY	4.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	93	ILE	4.8
1	B	90	THR	4.8
1	A	167	THR	4.7
1	B	366	PHE	4.5
1	B	193	GLN	4.5
1	B	159	PHE	4.5
1	A	197	GLN	4.3
1	A	166	LYS	4.2
1	B	197	GLN	4.2
1	A	198	VAL	4.1
1	B	170	GLU	4.0
1	B	89	ARG	3.9
1	B	348	SER	3.9
1	B	195	THR	3.8
1	B	91	PRO	3.8
1	B	166	LYS	3.8
1	A	156	ARG	3.7
1	A	155	PHE	3.7
1	A	181	PHE	3.7
1	B	326	TYR	3.7
1	A	196	ASN	3.7
1	B	345	PRO	3.7
1	B	175	GLN	3.6
1	B	196	ASN	3.6
1	A	182	PHE	3.5
1	B	356	LYS	3.5
1	A	164	LEU	3.5
1	B	7	PHE	3.5
1	B	346	ALA	3.3
1	B	167	THR	3.3
1	B	165	GLY	3.2
1	A	168	ILE	3.2
1	B	160	VAL	3.1
1	A	200	ASN	3.1
1	A	369	HIS	3.1
1	A	177	LEU	3.1
1	A	83	LEU	3.1
1	B	8	LEU	3.1
1	B	171	SER	3.0
1	B	198	VAL	3.0
1	A	201	PHE	3.0
1	A	170	GLU	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	158	LEU	2.9
1	B	86	GLN	2.9
1	A	178	ILE	2.8
1	A	172	PRO	2.8
1	A	89	ARG	2.8
1	B	305	HIS	2.7
1	B	84	GLU	2.7
1	B	200	ASN	2.7
1	B	179	SER	2.7
1	A	130	TYR	2.6
1	A	276	LYS	2.5
1	B	319	LYS	2.5
1	B	347	LEU	2.5
1	A	199	ASN	2.5
1	A	128	ILE	2.5
1	B	92	ILE	2.5
1	A	157	GLN	2.5
1	B	258	ASP	2.4
1	A	97	TYR	2.4
1	A	5	LEU	2.3
1	B	132	GLY	2.3
1	A	162	SER	2.3
1	A	220	VAL	2.3
1	A	275	THR	2.2
1	B	96	VAL	2.2
1	B	352	ILE	2.2
1	B	317	ASP	2.2
1	B	9	GLY	2.2
1	A	129	VAL	2.2
1	B	82	LYS	2.2
1	A	90	THR	2.2
1	B	201	PHE	2.2
1	B	357	PHE	2.2
1	A	368	GLU	2.2
1	B	18	ILE	2.2
1	A	86	GLN	2.1
1	B	307	ARG	2.1
1	A	133	THR	2.1
1	A	175	GLN	2.1
1	A	88	LEU	2.1
1	B	19	ALA	2.1
1	A	94	LYS	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	351	PRO	2.1
1	B	354	GLY	2.1
1	A	7	PHE	2.1
1	B	355	ALA	2.1
1	A	4	ILE	2.0
1	B	341	ILE	2.0
1	B	133	THR	2.0

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, 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
2	AZI	B	1101	3/3	0.82	0.25	32,32,39,48	0
2	AZI	A	1101	3/3	0.91	0.27	37,37,39,51	0

6.5 Other polymers (i)

There are no such residues in this entry.