Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2020 – 04:13 AM BST

PDB ID	:	6Q2F
Title	:	Structure of Rhamnosidase from Novosphingobium sp. PP1Y
Authors	:	Terry, B.; Ha, J.; Izzo, V.; Sazinsky, M.H.
Deposited on	:	2019-08-07
Resolution	:	2.20 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 following versions of software and data (see references ⓘ) were used in the production of this report:

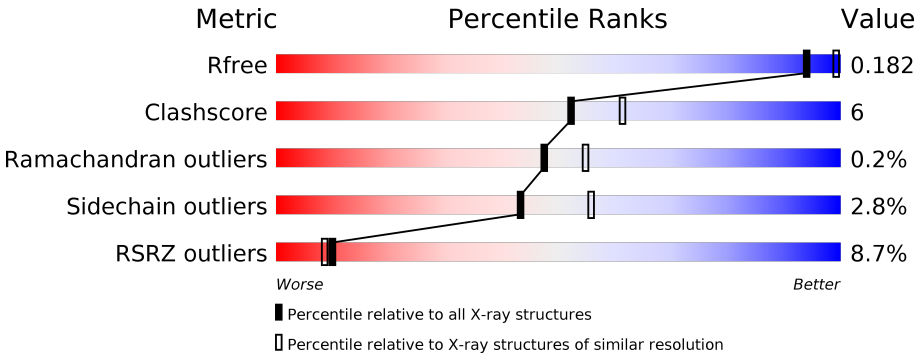
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 on 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	1151	8% 80% 11% • 8%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8812 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 Glycoside hydrolase family protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1055	Total	C	N	O	Se	0	0	0
			8079	5131	1410	1519	19			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1146	HIS	-	expression tag	UNP F6IEX3
A	1147	HIS	-	expression tag	UNP F6IEX3
A	1148	HIS	-	expression tag	UNP F6IEX3
A	1149	HIS	-	expression tag	UNP F6IEX3
A	1150	HIS	-	expression tag	UNP F6IEX3

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Na	0	0
			1	1		

- Molecule 3 is water.

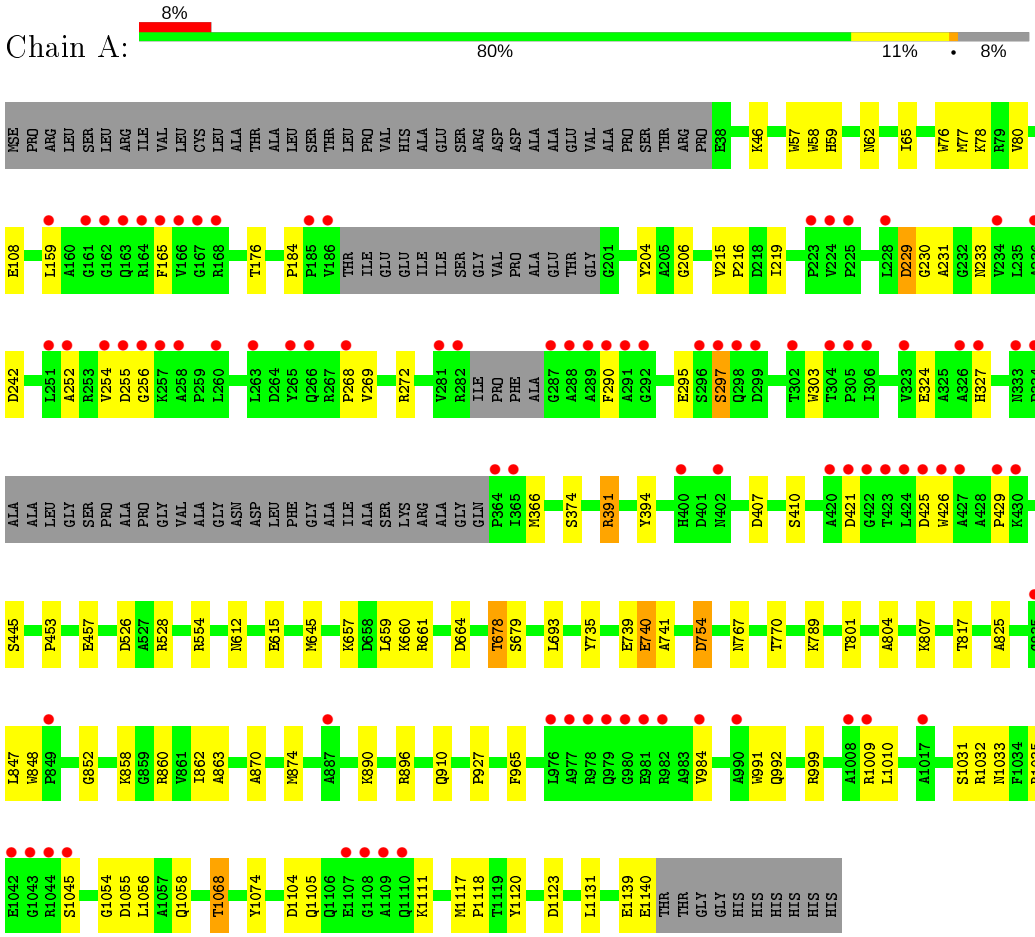
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	732	Total	O	0	0
			732	732		

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: Glycoside hydrolase family protein

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	80.81Å 114.57Å 158.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.70 – 2.20 39.70 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (39.70-2.20) 99.8 (39.70-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.35 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.10_2155	Depositor
R, R_{free}	0.163 , 0.177 0.164 , 0.182	Depositor DCC
R_{free} test set	2000 reflections (2.65%)	wwPDB-VP
Wilson B-factor (Å^2)	30.9	Xtriage
Anisotropy	0.812	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8812	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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: NA

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	0.50	1/8263 (0.0%)	0.61	0/11237

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	825	ALA	C-N	5.69	1.47	1.34

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	A	8079	0	7887	88	0
2	A	1	0	0	0	0
3	A	732	0	0	8	1
All	All	8812	0	7887	88	1

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 6.

All (88) 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:657:LYS:HE3	1:A:1074:TYR:OH	1.41	1.19
1:A:890:LYS:H	1:A:910:GLN:HE22	1.08	0.97
1:A:229:ASP:OD2	1:A:252:ALA:N	2.04	0.90
1:A:59:HIS:ND1	1:A:678:THR:HG21	1.94	0.82
1:A:241:VAL:O	1:A:241:VAL:HG12	1.79	0.81
1:A:65:ILE:HB	1:A:102:LEU:HD12	1.67	0.74
1:A:554:ARG:NH1	1:A:739:GLU:OE2	2.19	0.74
1:A:391:ARG:HG3	1:A:391:ARG:NH1	2.03	0.72
1:A:1032:ARG:NH1	1:A:1033:ASN:O	2.23	0.72
1:A:837:LEU:H	1:A:837:LEU:HD23	1.56	0.70
1:A:159:LEU:HD11	1:A:165:PHE:HD2	1.55	0.70
1:A:770:THR:HA	1:A:807:LYS:HG2	1.73	0.70
1:A:992:GLN:HB2	1:A:1031:SER:HB3	1.72	0.70
1:A:391:ARG:HG3	1:A:391:ARG:HH11	1.55	0.70
1:A:297:SER:HB2	1:A:327:HIS:HB2	1.74	0.70
1:A:59:HIS:CE1	1:A:678:THR:HG21	2.27	0.68
1:A:612:ASN:HB2	1:A:615:GLU:HG3	1.74	0.68
1:A:890:LYS:H	1:A:910:GLN:NE2	1.85	0.68
1:A:896:ARG:NH2	3:A:1301:HOH:O	2.14	0.67
1:A:984:VAL:HG21	1:A:1139:GLU:HG3	1.79	0.64
1:A:216:PRO:O	1:A:219:ILE:HG13	2.01	0.61
1:A:1117:MSE:HE3	1:A:1118:PRO:HD2	1.81	0.61
1:A:801:THR:HG23	1:A:804:ALA:H	1.67	0.59
1:A:241:VAL:O	1:A:241:VAL:CG1	2.49	0.59
1:A:391:ARG:CG	1:A:391:ARG:HH11	2.15	0.58
1:A:739:GLU:OE1	1:A:801:THR:HG22	2.04	0.58
1:A:268:PRO:HB2	1:A:324:GLU:HG3	1.87	0.57
1:A:554:ARG:HH12	1:A:739:GLU:CD	2.09	0.55
1:A:858:LYS:NZ	3:A:1313:HOH:O	2.38	0.54
1:A:1010:LEU:HD22	1:A:1131:LEU:O	2.08	0.54
1:A:789:LYS:NZ	3:A:1317:HOH:O	2.42	0.53
1:A:297:SER:OG	1:A:303:TRP:NE1	2.39	0.52
1:A:927:PRO:HB3	1:A:965:PHE:CE2	2.45	0.52
1:A:801:THR:CG2	1:A:804:ALA:H	2.23	0.51
1:A:837:LEU:H	1:A:837:LEU:CD2	2.18	0.51
1:A:862:ILE:HD12	1:A:874:MSE:HE2	1.91	0.51
1:A:159:LEU:HD11	1:A:165:PHE:CD2	2.43	0.51
1:A:421:ASP:O	1:A:421:ASP:OD1	2.29	0.51
1:A:1045:SER:H	1:A:1140:GLU:HG2	1.75	0.50
1:A:215:VAL:HG11	1:A:438:ARG:HD3	1.93	0.50
1:A:1104:ASP:HB3	1:A:1111:LYS:HD2	1.94	0.50
1:A:429:PRO:HG2	1:A:437:TRP:HZ2	1.76	0.50

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:272:ARG:HD2	1:A:374:SER:O	2.12	0.49
1:A:268:PRO:HB2	1:A:324:GLU:CG	2.43	0.49
1:A:429:PRO:HG2	1:A:437:TRP:CZ2	2.48	0.49
1:A:457:GLU:H	1:A:457:GLU:CD	2.16	0.49
1:A:229:ASP:OD1	1:A:231:ALA:N	2.46	0.48
1:A:254:VAL:HG12	1:A:255:ASP:N	2.29	0.47
1:A:77:MSE:SE	1:A:85:LEU:HD13	2.65	0.47
1:A:1105:GLN:HG3	1:A:1120:TYR:O	2.14	0.47
1:A:1058:GLN:HA	1:A:1068:THR:HB	1.95	0.47
1:A:76:TRP:O	1:A:80:VAL:HG22	2.14	0.47
1:A:870:ALA:C	1:A:874:MSE:HE3	2.36	0.46
1:A:102:LEU:CD2	1:A:108:GLU:HG2	2.46	0.46
1:A:102:LEU:HD22	1:A:108:GLU:HG2	1.98	0.46
1:A:407:ASP:HB3	1:A:410:SER:OG	2.16	0.46
1:A:807:LYS:HE2	3:A:1384:HOH:O	2.16	0.46
1:A:394:TYR:CE2	1:A:740:GLU:HB2	2.51	0.45
1:A:1054:GLY:HA2	1:A:1055:ASP:HA	1.78	0.44
1:A:255:ASP:OD1	1:A:256:GLY:N	2.50	0.44
1:A:767:ASN:HB2	3:A:1945:HOH:O	2.17	0.44
1:A:1068:THR:HG21	3:A:1933:HOH:O	2.18	0.44
1:A:65:ILE:CB	1:A:102:LEU:HD12	2.44	0.44
1:A:229:ASP:OD2	1:A:252:ALA:HB3	2.17	0.44
1:A:526:ASP:OD2	1:A:528:ARG:NH1	2.51	0.43
1:A:57:TRP:CB	1:A:678:THR:HG22	2.49	0.43
1:A:693:LEU:O	1:A:693:LEU:HG	2.17	0.43
1:A:297:SER:CB	1:A:327:HIS:HB2	2.46	0.42
1:A:58:TRP:HA	1:A:679:SER:O	2.19	0.42
1:A:426:TRP:CZ2	1:A:429:PRO:HD2	2.55	0.42
1:A:204:TYR:O	1:A:445:SER:HA	2.19	0.42
1:A:740:GLU:HB3	3:A:1347:HOH:O	2.19	0.42
1:A:645:MSE:SE	1:A:659:LEU:HB3	2.69	0.42
1:A:735:TYR:HE2	1:A:754:ASP:HB3	1.85	0.42
1:A:159:LEU:HD21	1:A:165:PHE:CD2	2.55	0.41
1:A:660:LYS:NZ	1:A:664:ASP:OD2	2.46	0.41
1:A:817:THR:HA	1:A:860:ARG:O	2.20	0.41
1:A:78:LYS:HD2	1:A:78:LYS:HA	1.83	0.41
1:A:838:THR:HB	3:A:1818:HOH:O	2.21	0.41
1:A:740:GLU:CD	1:A:741:ALA:N	2.73	0.41
1:A:46:LYS:HA	1:A:46:LYS:HD2	1.85	0.41
1:A:847:LEU:HB2	1:A:848:TRP:CE3	2.56	0.41
1:A:999:ARG:HD3	1:A:999:ARG:HA	1.87	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:852:GLY:HA2	1:A:863:ALA:HB3	2.02	0.41
1:A:176:THR:O	1:A:184:PRO:HD3	2.21	0.40
1:A:206:GLY:HA3	1:A:444:TYR:CE2	2.56	0.40
1:A:991:TRP:CE2	1:A:1032:ARG:HD2	2.56	0.40
1:A:229:ASP:OD1	1:A:230:GLY:N	2.55	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:1729:HOH:O	3:A:1756:HOH:O[3_545]	2.14	0.06

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	1047/1151 (91%)	1001 (96%)	44 (4%)	2 (0%)	47 55

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	837	LEU
1	A	453	PRO

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	825/887 (93%)	802 (97%)	23 (3%)	43 56

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	ASN
1	A	229	ASP
1	A	233	ASN
1	A	238	LYS
1	A	242	ASP
1	A	269	VAL
1	A	290	PHE
1	A	295	GLU
1	A	297	SER
1	A	366	MSE
1	A	391	ARG
1	A	425	ASP
1	A	661	ARG
1	A	678	THR
1	A	740	GLU
1	A	754	ASP
1	A	837	LEU
1	A	1009	ARG
1	A	1035	ARG
1	A	1039	LYS
1	A	1056	LEU
1	A	1068	THR
1	A	1123	ASP

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

Mol	Chain	Res	Type
1	A	910	GLN

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 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	1036/1151 (90%)	0.18	90 (8%) 10 8	21, 37, 84, 133	0

All (90) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	255	ASP	11.1
1	A	166	VAL	8.7
1	A	257	LYS	7.0
1	A	287	GLY	6.5
1	A	256	GLY	5.4
1	A	186	VAL	5.4
1	A	165	PHE	5.3
1	A	302	THR	5.3
1	A	167	GLY	5.1
1	A	288	ALA	5.1
1	A	168	ARG	4.9
1	A	421	ASP	4.8
1	A	1043	GLY	4.7
1	A	254	VAL	4.6
1	A	1041	GLY	4.5
1	A	1108	GLY	4.3
1	A	334	PRO	4.2
1	A	982	ARG	4.2
1	A	364	PRO	4.1
1	A	299	ASP	4.1
1	A	1045	SER	4.0
1	A	427	ALA	4.0
1	A	228	LEU	3.9
1	A	837	LEU	3.9
1	A	1109	ALA	3.9
1	A	224	VAL	3.9
1	A	1110	GLN	3.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	258	ALA	3.8
1	A	260	LEU	3.8
1	A	305	PRO	3.8
1	A	164	ARG	3.7
1	A	290	PHE	3.7
1	A	225	PRO	3.7
1	A	977	ALA	3.7
1	A	980	GLY	3.7
1	A	1042	GLU	3.7
1	A	163	GLN	3.7
1	A	185	PRO	3.6
1	A	326	ALA	3.6
1	A	327	HIS	3.5
1	A	420	ALA	3.5
1	A	423	THR	3.5
1	A	979	GLN	3.5
1	A	304	THR	3.4
1	A	990	ALA	3.4
1	A	1044	ARG	3.4
1	A	835	GLY	3.3
1	A	1009	ARG	3.2
1	A	297	SER	3.1
1	A	162	GLY	3.1
1	A	289	ALA	3.1
1	A	306	ILE	3.1
1	A	333	ASN	3.0
1	A	887	ALA	3.0
1	A	424	LEU	3.0
1	A	425	ASP	3.0
1	A	251	LEU	3.0
1	A	978	ARG	3.0
1	A	159	LEU	2.9
1	A	282	ARG	2.8
1	A	263	LEU	2.8
1	A	236	ALA	2.8
1	A	266	GLN	2.7
1	A	430	LYS	2.7
1	A	365	ILE	2.6
1	A	838	THR	2.6
1	A	429	PRO	2.5
1	A	298	GLN	2.5
1	A	422	GLY	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1008	ALA	2.5
1	A	234	VAL	2.5
1	A	981	GLU	2.4
1	A	252	ALA	2.4
1	A	161	GLY	2.4
1	A	323	VAL	2.4
1	A	223	PRO	2.4
1	A	1107	GLU	2.3
1	A	296	SER	2.3
1	A	291	ALA	2.3
1	A	268	PRO	2.2
1	A	984	VAL	2.2
1	A	426	TRP	2.2
1	A	402	ASN	2.2
1	A	849	PRO	2.2
1	A	265	TYR	2.2
1	A	400	HIS	2.1
1	A	976	LEU	2.1
1	A	292	GLY	2.0
1	A	281	VAL	2.0
1	A	1017	ALA	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	NA	A	1201	1/1	0.93	0.25	37,37,37,37	0

6.5 Other polymers (i)

There are no such residues in this entry.