Full wwPDB X-ray Structure Validation Report (i)

Jan 15, 2024 – 02:36 pm GMT

PDB ID	:	6NVY
Title	:	Crystal structure of penicillin G acylase from Bacillus thermotolerans
Authors	:	Blankenfeldt, W.
Deposited on	:	2019-02-05
Resolution	:	2.27 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 (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 (1)) were used in the production of this report:

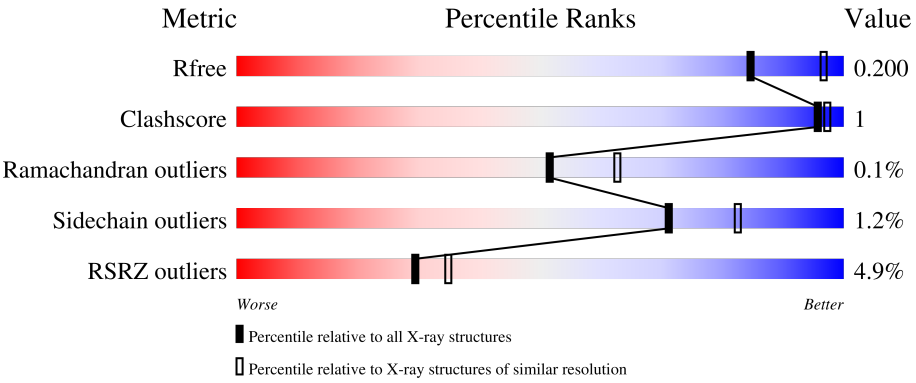
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.27 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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	212	2% 88% 10%
1	C	212	6% 85% 10%
2	B	538	6% 95%
2	D	538	4% 97%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 23672 atoms, of which 11095 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 Penicillin G acylase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	191	Total	C	H	N	O	S	0	0	0
			3064	996	1489	258	315	6			
1	C	191	Total	C	H	N	O	S	0	0	0
			3057	995	1482	257	317	6			

- Molecule 2 is a protein called Penicillin G acylase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	536	Total	C	H	N	O	S	0	1	0
			8405	2758	4062	724	847	14			
2	D	536	Total	C	H	N	O	S	0	2	0
			8399	2759	4054	722	850	14			

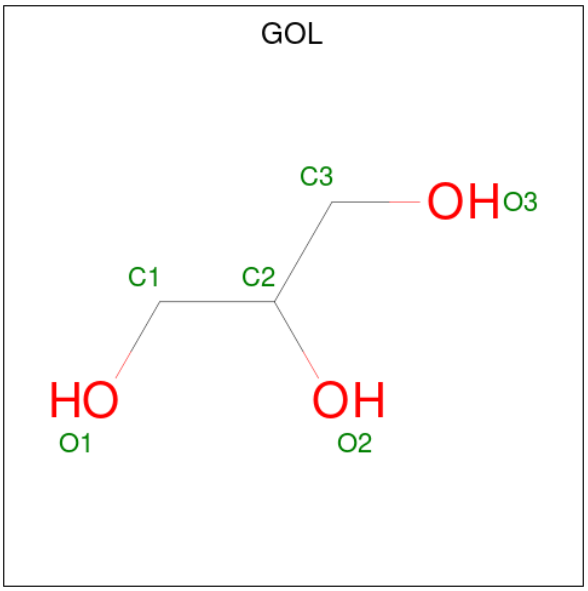
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	40	PRO	SER	conflict	UNP A0A0F5I5V4
D	40	PRO	SER	conflict	UNP A0A0F5I5V4

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		
3	B	5	Total	Ca	0	0
			5	5		
3	C	1	Total	Ca	0	0
			1	1		
3	D	5	Total	Ca	0	0
			5	5		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 5 is water.

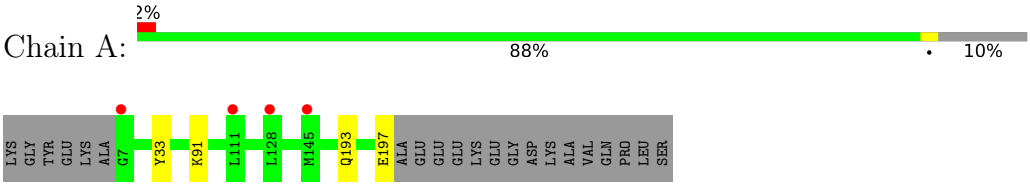
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	75	Total	O	0	0
			75	75		
5	B	292	Total	O	0	0
			292	292		
5	C	79	Total	O	0	0
			79	79		
5	D	275	Total	O	0	0
			275	275		

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.

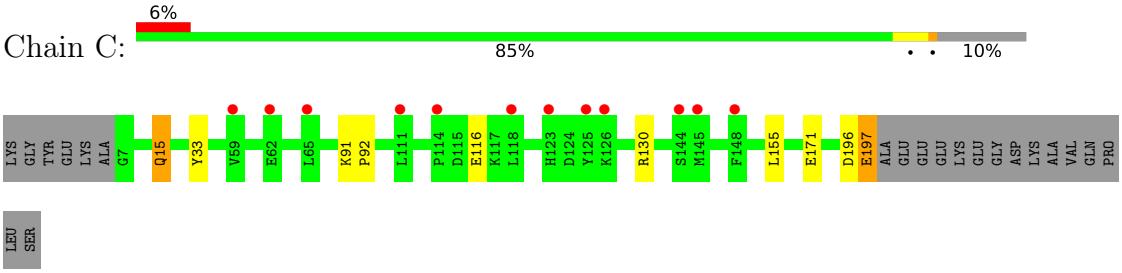
• Molecule 1: Penicillin G acylase

Chain A:



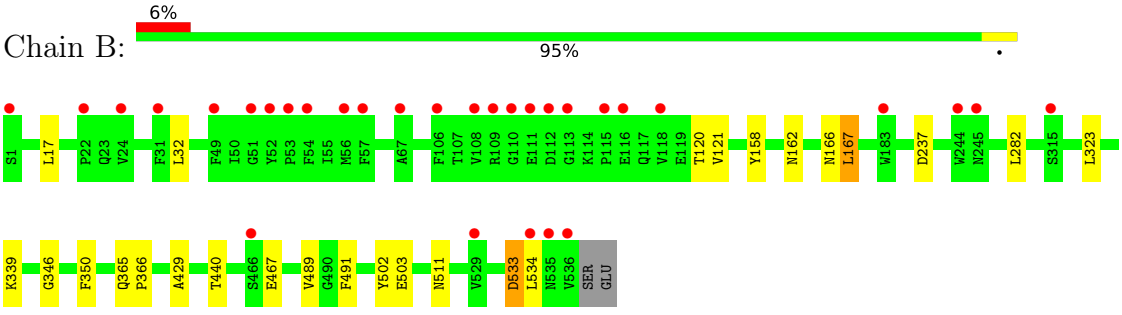
• Molecule 1: Penicillin G acylase

Chain C:



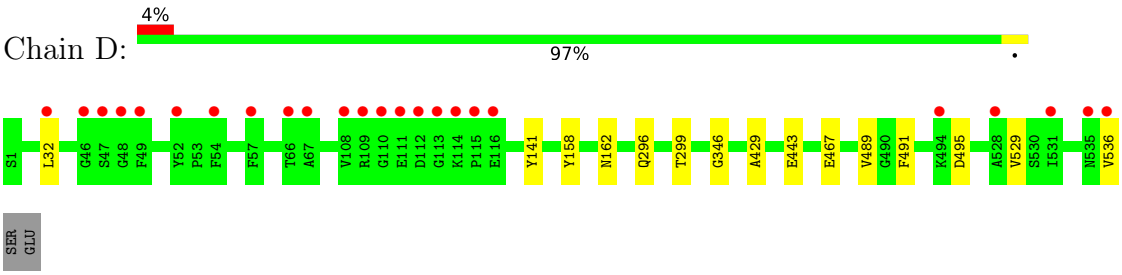
• Molecule 2: Penicillin G acylase

Chain B:



• Molecule 2: Penicillin G acylase

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	173.00Å 173.00Å 209.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.25 – 2.27 43.25 – 2.27	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.25-2.27) 99.9 (43.25-2.27)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.92 (at 2.27Å)	Xtriage
Refinement program	PHENIX dev_3386	Depositor
R, R_{free}	0.172 , 0.200 0.172 , 0.200	Depositor DCC
R_{free} test set	5148 reflections (4.77%)	wwPDB-VP
Wilson B-factor (Å^2)	43.8	Xtriage
Anisotropy	0.088	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 52.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -2/3*h-1/3*k+2/3*l,-1/3*h-2/3*k-2/3*l,2/3*h-2/3*k+1/3*l 0.000 for -h,1/3*h-1/3*k+2/3*l,2/3*h+4/3*k+1/3*l 0.000 for -1/3*h+1/3*k-2/3*l,-k,-4/3*h-2/3*k+1/3*l 0.000 for -h,2/3*h+1/3*k-2/3*l,-2/3*h-4/3*k-1/3*l 0.000 for 1/3*h+2/3*k+2/3*l,-k,4/3*h+2/3*k-1/3*l 0.000 for -1/3*h-2/3*k-2/3*l,-2/3*h-1/3*k+2/3*l,-2/3*h+2/3*k-1/3*l 0.020 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	23672	wwPDB-VP
Average B, all atoms (Å^2)	60.0	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 47.23 % 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 1.0220e-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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, GOL

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	0.25	0/1608	0.40	0/2169
1	C	0.25	0/1608	0.41	0/2170
2	B	0.27	0/4468	0.46	0/6059
2	D	0.27	0/4473	0.46	0/6067
All	All	0.26	0/12157	0.45	0/16465

There are no bond length outliers.

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	1575	1489	1488	1	0
1	C	1575	1482	1481	7	0
2	B	4343	4062	4062	14	0
2	D	4345	4054	4055	7	0
3	A	1	0	0	0	0
3	B	5	0	0	0	0
3	C	1	0	0	0	0
3	D	5	0	0	0	0
4	B	6	8	8	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	75	0	0	1	0
5	B	292	0	0	2	0
5	C	79	0	0	1	0
5	D	275	0	0	0	1
All	All	12577	11095	11094	26	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 1.

All (26) 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)
2:B:533:ASP:OD1	5:B:701:HOH:O	1.97	0.83
2:B:166:ASN:O	5:B:702:HOH:O	1.99	0.81
1:A:193:GLN:NE2	5:A:401:HOH:O	2.28	0.66
2:B:440:THR:OG1	1:C:171:GLU:OE1	2.10	0.59
1:C:116:GLU:OE1	1:C:116:GLU:N	2.36	0.59
2:B:166:ASN:O	2:B:167:LEU:CB	2.55	0.54
2:B:32:LEU:HD13	2:B:467:GLU:HB3	1.88	0.54
2:B:17:LEU:HD22	2:B:282:LEU:CD1	2.40	0.52
2:B:323:LEU:HD21	2:B:350:PHE:CZ	2.46	0.51
2:D:32:LEU:HD13	2:D:467:GLU:HB3	1.93	0.51
2:D:443:GLU:N	2:D:443:GLU:OE1	2.43	0.50
2:B:365:GLN:HB3	2:B:366:PRO:HD3	1.95	0.48
2:B:489:VAL:HG12	2:B:491:PHE:H	1.77	0.48
1:C:130:ARG:NH2	5:C:403:HOH:O	2.46	0.48
2:D:489:VAL:HG12	2:D:491:PHE:H	1.79	0.47
2:B:489:VAL:O	2:B:502:TYR:HA	2.15	0.46
1:C:91:LYS:HB2	1:C:92:PRO:HD3	1.99	0.44
1:C:196:ASP:O	1:C:197:GLU:C	2.56	0.44
1:C:15:GLN:HA	2:D:529:VAL:HG21	2.00	0.43
2:B:339:LYS:NZ	2:B:511:ASN:OD1	2.44	0.42
2:B:346:GLY:HA3	2:B:429:ALA:O	2.18	0.42
1:C:155:LEU:HD21	2:D:141:TYR:CD1	2.54	0.42
2:D:296:GLN:HB3	2:D:299:THR:HG23	2.01	0.41
2:D:346:GLY:HA3	2:D:429:ALA:O	2.19	0.41
2:B:503:GLU:CG	2:B:503:GLU:O	2.69	0.41
2:B:120:THR:HG22	2:B:121:VAL:N	2.36	0.41

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)
5:D:702:HOH:O	5:D:784:HOH:O[2_665]	1.91	0.29

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	189/212 (89%)	184 (97%)	5 (3%)	0	100 100
1	C	189/212 (89%)	186 (98%)	3 (2%)	0	100 100
2	B	535/538 (99%)	518 (97%)	16 (3%)	1 (0%)	47 57
2	D	536/538 (100%)	524 (98%)	12 (2%)	0	100 100
All	All	1449/1500 (97%)	1412 (97%)	36 (2%)	1 (0%)	51 63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	167	LEU

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	166/184 (90%)	163 (98%)	3 (2%)	59 72
1	C	166/184 (90%)	163 (98%)	3 (2%)	59 72
2	B	456/458 (100%)	451 (99%)	5 (1%)	73 84
2	D	456/458 (100%)	452 (99%)	4 (1%)	78 88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1244/1284 (97%)	1229 (99%)	15 (1%)	71 82

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	33	TYR
1	A	91	LYS
1	A	197	GLU
2	B	158	TYR
2	B	162	ASN
2	B	237	ASP
2	B	533	ASP
2	B	534	LEU
1	C	15	GLN
1	C	33	TYR
1	C	197	GLU
2	D	158	TYR
2	D	162	ASN
2	D	495	ASP
2	D	536	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 13 ligands modelled in this entry, 12 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
4	GOL	B	606	-	5,5,5	0.85	0	5,5,5	0.99	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
4	GOL	B	606	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	606	GOL	O1-C1-C2-C3
4	B	606	GOL	O1-C1-C2-O2

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	191/212 (90%)	-0.03	4 (2%)	63	69	35, 61, 91, 114	0
1	C	191/212 (90%)	-0.01	12 (6%)	20	24	37, 62, 88, 115	0
2	B	536/538 (99%)	0.14	31 (5%)	23	28	31, 49, 87, 145	0
2	D	536/538 (99%)	0.06	24 (4%)	33	39	30, 49, 87, 151	0
All	All	1454/1500 (96%)	0.07	71 (4%)	29	35	30, 52, 89, 151	0

All (71) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	115	PRO	5.1
2	B	536	VAL	4.8
2	D	110	GLY	4.6
2	D	113	GLY	4.1
2	B	534	LEU	3.6
2	D	111	GLU	3.6
2	B	111	GLU	3.5
2	B	115	PRO	3.4
2	B	24	VAL	3.3
2	B	116	GLU	3.3
2	B	110	GLY	3.3
2	B	56	MET	3.3
1	C	114	PRO	3.2
1	C	62	GLU	3.2
2	B	315	SER	3.1
1	A	7	GLY	3.1
2	B	52	TYR	3.0
1	A	111	LEU	2.9
2	B	118	VAL	2.9
2	B	22	PRO	2.9
2	B	57	PHE	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	D	54	PHE	2.8
1	C	65	LEU	2.8
2	B	535	ASN	2.8
2	B	1	SER	2.7
2	D	112	ASP	2.7
2	D	116	GLU	2.7
2	B	106	PHE	2.7
1	A	128	LEU	2.7
2	B	49	PHE	2.7
2	D	108	VAL	2.7
2	D	49	PHE	2.6
2	D	48	GLY	2.6
2	B	112	ASP	2.6
2	D	47	SER	2.6
2	D	57	PHE	2.6
1	C	144	SER	2.5
1	C	123	HIS	2.5
2	B	113	GLY	2.5
1	C	118	LEU	2.5
1	A	145	MET	2.5
1	C	125	TYR	2.4
2	B	67	ALA	2.4
1	C	126	LYS	2.4
2	D	114	LYS	2.4
2	D	67	ALA	2.4
1	C	145	MET	2.4
1	C	59	VAL	2.3
2	B	51	GLY	2.3
2	B	109	ARG	2.3
2	D	109	ARG	2.3
2	D	32	LEU	2.3
2	D	494	LYS	2.3
1	C	148	PHE	2.3
2	B	108	VAL	2.2
2	D	528	ALA	2.2
2	B	54	PHE	2.2
2	D	46	GLY	2.2
2	B	529	VAL	2.2
2	D	531	ILE	2.2
2	B	245	ASN	2.1
2	D	52	TYR	2.1
2	D	536	VAL	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	111	LEU	2.1
2	B	31	PHE	2.1
2	B	53	PRO	2.1
2	B	244	TRP	2.1
2	D	535	ASN	2.1
2	B	466	SER	2.1
2	D	66	THR	2.0
2	B	183	TRP	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
3	CA	B	605	1/1	0.14	0.22	140,140,140,140	0
3	CA	D	605	1/1	0.60	0.14	105,105,105,105	0
4	GOL	B	606	6/6	0.76	0.18	79,95,103,104	0
3	CA	A	301	1/1	0.94	0.09	76,76,76,76	0
3	CA	C	301	1/1	0.95	0.09	70,70,70,70	0
3	CA	D	604	1/1	0.95	0.08	74,74,74,74	0
3	CA	B	604	1/1	0.96	0.08	70,70,70,70	0
3	CA	D	603	1/1	0.97	0.06	54,54,54,54	0
3	CA	B	602	1/1	0.98	0.08	47,47,47,47	0
3	CA	D	601	1/1	0.98	0.10	37,37,37,37	0
3	CA	B	603	1/1	0.99	0.04	61,61,61,61	0
3	CA	D	602	1/1	0.99	0.07	47,47,47,47	0
3	CA	B	601	1/1	1.00	0.10	36,36,36,36	0

6.5 Other polymers (i)

There are no such residues in this entry.