Full wwPDB X-ray Structure Validation Report ⓘ

Aug 29, 2023 – 11:47 AM EDT

PDB ID	:	7UYX
Title	:	Structure of bacteriophage PA1c gp2
Authors	:	Enustun, E.; Deep, A.; Gu, Y.; Nguyen, K.; Chaikeeratisak, V.; Armbruster, E.; Ghassemian, M.; Pogliano, J.; Corbett, K.D.
Deposited on	:	2022-05-07
Resolution	:	2.63 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:

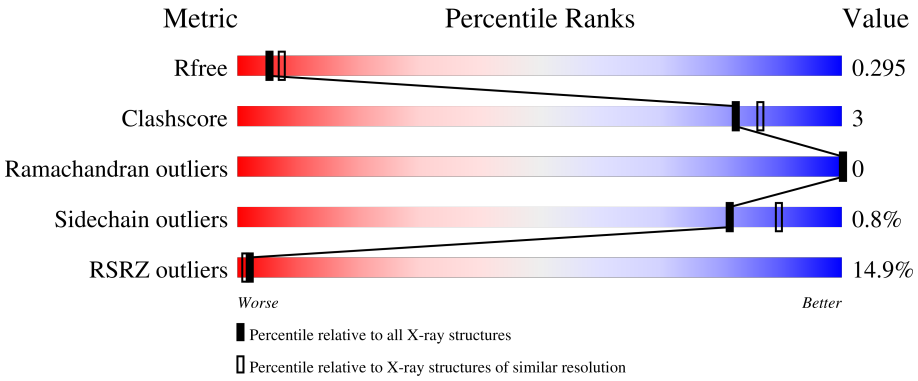
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.63 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	1426 (2.66-2.62)
Clashscore	141614	1472 (2.66-2.62)
Ramachandran outliers	138981	1446 (2.66-2.62)
Sidechain outliers	138945	1446 (2.66-2.62)
RSRZ outliers	127900	1408 (2.66-2.62)

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	196	10% 79% 6% 15%
1	B	196	14% 78% 8% 14%
1	C	196	14% 76% 8% 16%
1	D	196	12% 76% 9% 15%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 10607 atoms, of which 5251 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 Bacteriophage PA1C gp2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	167	Total	C	H	N	O	S	0	0	0
			2658	879	1315	212	248	4			
1	B	169	Total	C	H	N	O	S	0	0	0
			2699	891	1339	215	250	4			
1	C	164	Total	C	H	N	O	S	0	0	0
			2603	861	1287	209	243	3			
1	D	166	Total	C	H	N	O	S	0	0	0
			2647	876	1310	211	246	4			

There are 12 discrepancies between the modelled and reference sequences:

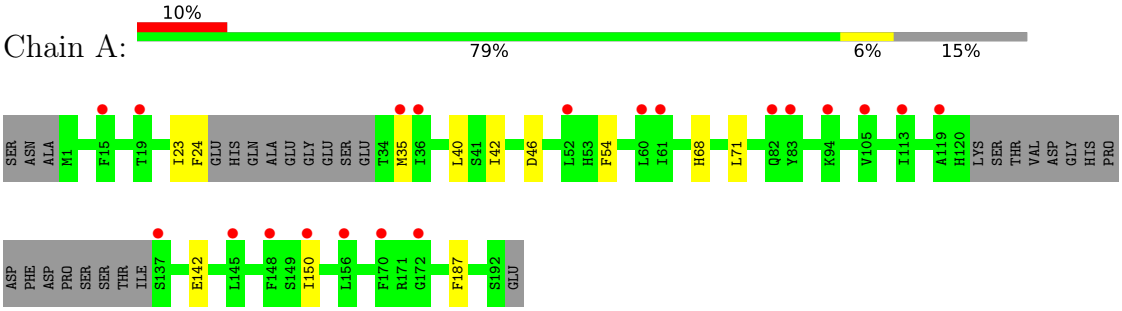
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP A0A4D6BFJ2
A	-1	ASN	-	expression tag	UNP A0A4D6BFJ2
A	0	ALA	-	expression tag	UNP A0A4D6BFJ2
B	-2	SER	-	expression tag	UNP A0A4D6BFJ2
B	-1	ASN	-	expression tag	UNP A0A4D6BFJ2
B	0	ALA	-	expression tag	UNP A0A4D6BFJ2
C	-2	SER	-	expression tag	UNP A0A4D6BFJ2
C	-1	ASN	-	expression tag	UNP A0A4D6BFJ2
C	0	ALA	-	expression tag	UNP A0A4D6BFJ2
D	-2	SER	-	expression tag	UNP A0A4D6BFJ2
D	-1	ASN	-	expression tag	UNP A0A4D6BFJ2
D	0	ALA	-	expression tag	UNP A0A4D6BFJ2

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.

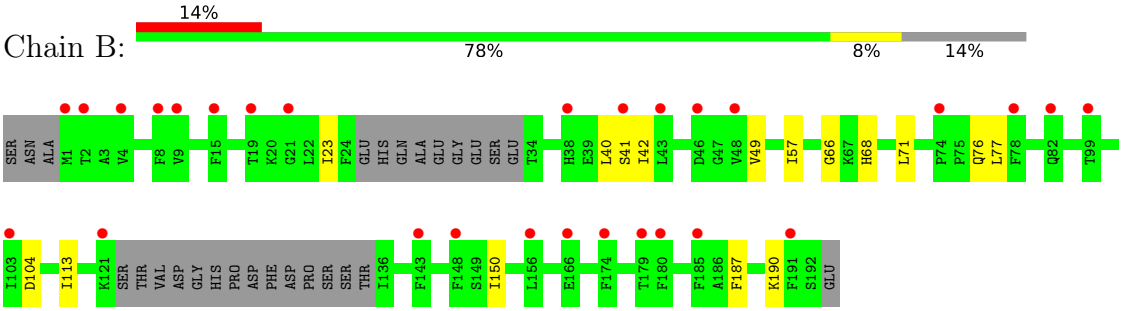
• Molecule 1: Bacteriophage PA1C gp2

Chain A:



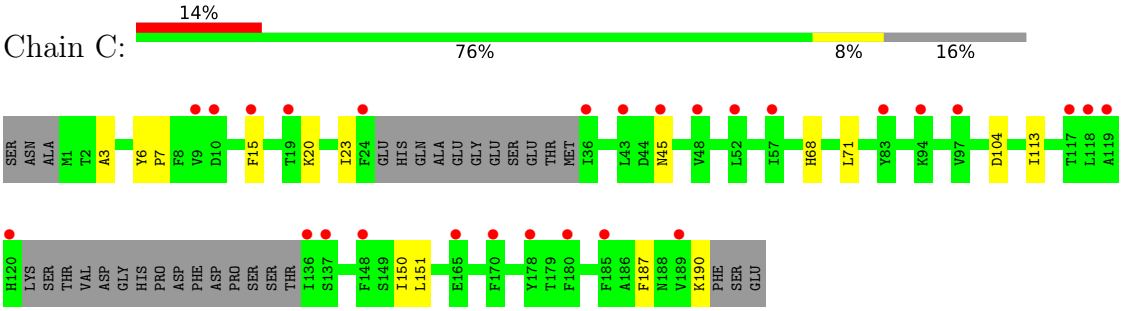
• Molecule 1: Bacteriophage PA1C gp2

Chain B:



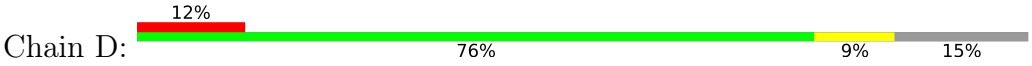
• Molecule 1: Bacteriophage PA1C gp2

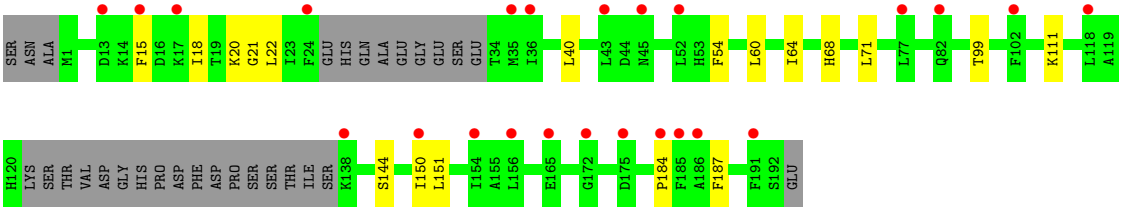
Chain C:



• Molecule 1: Bacteriophage PA1C gp2

Chain D:




SER ASN ALA M1 D13 K14 F15 D16 K17 I18 T19 K20 G21 L22 I23 F24 GLU HIS GLN ALA GLU GLY GLU SER GLU T34 M35 I36 L40 L43 D44 M45 L52 H53 F54 L60 I64 H68 L71 L77 Q82 T99 F102 K111 L118 A119
H120 LYS SER THR VAL ASP GLY HIS PRO ASP PHE ASP PRO SER SER THR ILE SER K138 S144 I150 L151 I154 A155 L156 E165 G172 D175 P184 F185 A186 F187 F191 S192 GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	60.95Å 86.56Å 71.88Å 90.00° 110.94° 90.00°	Depositor
Resolution (Å)	67.13 – 2.63 67.13 – 2.63	Depositor EDS
% Data completeness (in resolution range)	97.6 (67.13-2.63) 97.6 (67.13-2.63)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.01 (at 2.62Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.271 , 0.300 0.269 , 0.295	Depositor DCC
R_{free} test set	908 reflections (4.44%)	wwPDB-VP
Wilson B-factor (Å^2)	63.8	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 38.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	10607	wwPDB-VP
Average B, all atoms (Å^2)	81.0	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 24.43 % 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 3.8382e-03. 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

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	0.30	0/1377	0.47	0/1864
1	B	0.29	0/1394	0.47	0/1886
1	C	0.29	0/1349	0.47	0/1827
1	D	0.30	0/1371	0.47	0/1856
All	All	0.29	0/5491	0.47	0/7433

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	1343	1315	1317	6	4
1	B	1360	1339	1341	9	2
1	C	1316	1287	1289	10	0
1	D	1337	1310	1312	12	2
All	All	5356	5251	5259	27	4

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 (27) 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:C:15:PHE:HB3	1:D:151:LEU:HD22	1.79	0.65
1:C:23:ILE:HD11	1:D:150:ILE:HD13	1.85	0.58
1:C:45:ASN:ND2	1:D:21:GLY:O	2.39	0.56
1:D:99:THR:O	1:D:184:PRO:HD3	2.08	0.54
1:C:151:LEU:HD22	1:D:15:PHE:HB3	1.91	0.53
1:A:150:ILE:HD13	1:B:23:ILE:HD11	1.94	0.49
1:A:68:HIS:HB3	1:A:71:LEU:HG	1.93	0.49
1:A:23:ILE:HD11	1:B:150:ILE:HD13	1.95	0.49
1:B:68:HIS:HB3	1:B:71:LEU:HG	1.95	0.49
1:D:40:LEU:HG	1:D:54:PHE:HE1	1.79	0.47
1:C:150:ILE:HD12	1:D:20:LYS:HG2	1.95	0.47
1:C:3:ALA:HA	1:D:144:SER:HB3	1.96	0.47
1:C:104:ASP:HB2	1:C:113:ILE:HB	1.97	0.46
1:B:104:ASP:HB2	1:B:113:ILE:HB	1.98	0.46
1:B:76:GLN:HG2	1:B:77:LEU:N	2.31	0.45
1:A:42:ILE:HD13	1:B:42:ILE:HD13	1.98	0.45
1:C:68:HIS:HB3	1:C:71:LEU:HG	1.98	0.45
1:C:20:LYS:HG2	1:D:150:ILE:HD12	1.99	0.45
1:B:40:LEU:HB2	1:B:57:ILE:CD1	2.47	0.45
1:B:66:GLY:HA2	1:B:71:LEU:HD12	2.00	0.44
1:D:18:ILE:O	1:D:22:LEU:HG	2.18	0.44
1:A:24:PHE:CZ	1:A:35:MET:HB3	2.52	0.44
1:D:68:HIS:HB3	1:D:71:LEU:HG	2.00	0.43
1:B:41:SER:HB3	1:B:49:VAL:HG13	2.00	0.43
1:C:6:TYR:CD1	1:C:7:PRO:HD2	2.54	0.42
1:D:60:LEU:O	1:D:64:ILE:HG13	2.20	0.42
1:A:40:LEU:HG	1:A:54:PHE:HE1	1.87	0.40

All (4) 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)
1:A:46:ASP:OD2	1:D:111:LYS:NZ[1_554]	1.91	0.29
1:A:46:ASP:OD2	1:D:111:LYS:HZ3[1_554]	1.52	0.08
1:A:142:GLU:OE2	1:B:190:LYS:NZ[1_554]	2.14	0.06
1:A:142:GLU:OE2	1:B:190:LYS:HZ1[1_554]	1.58	0.02

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	161/196 (82%)	161 (100%)	0	0	100	100
1	B	163/196 (83%)	160 (98%)	3 (2%)	0	100	100
1	C	158/196 (81%)	154 (98%)	4 (2%)	0	100	100
1	D	160/196 (82%)	159 (99%)	1 (1%)	0	100	100
All	All	642/784 (82%)	634 (99%)	8 (1%)	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	150/176 (85%)	149 (99%)	1 (1%)	84	91
1	B	152/176 (86%)	151 (99%)	1 (1%)	84	91
1	C	146/176 (83%)	144 (99%)	2 (1%)	67	80
1	D	149/176 (85%)	148 (99%)	1 (1%)	84	91
All	All	597/704 (85%)	592 (99%)	5 (1%)	81	89

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	187	PHE
1	B	187	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	187	PHE
1	C	190	LYS
1	D	187	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	152	GLN
1	C	152	GLN
1	D	38	HIS
1	D	164	HIS

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)

There are no ligands in this entry.

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	167/196 (85%)	0.91	20 (11%)	4	3	48, 70, 95, 114	0
1	B	169/196 (86%)	1.03	28 (16%)	1	1	51, 68, 96, 108	0
1	C	164/196 (83%)	1.13	27 (16%)	1	1	49, 72, 103, 121	0
1	D	166/196 (84%)	1.13	24 (14%)	2	1	51, 76, 106, 153	0
All	All	666/784 (84%)	1.05	99 (14%)	2	1	48, 71, 100, 153	0

All (99) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	156	LEU	6.5
1	D	36	ILE	6.0
1	D	15	PHE	5.1
1	D	35	MET	4.9
1	A	52	LEU	4.7
1	C	36	ILE	4.7
1	C	178	TYR	4.6
1	D	43	LEU	4.5
1	D	186	ALA	4.5
1	C	94	LYS	4.4
1	C	118	LEU	4.2
1	A	83	TYR	4.1
1	C	148	PHE	4.1
1	B	148	PHE	4.0
1	D	102	PHE	4.0
1	B	180	PHE	4.0
1	B	191	PHE	3.9
1	A	119	ALA	3.9
1	B	82	GLN	3.8
1	C	136	ILE	3.7
1	C	43	LEU	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	191	PHE	3.7
1	C	119	ALA	3.7
1	B	156	LEU	3.6
1	C	120	HIS	3.6
1	B	174	PHE	3.6
1	A	148	PHE	3.5
1	A	35	MET	3.5
1	B	15	PHE	3.5
1	C	45	ASN	3.5
1	B	1	MET	3.5
1	D	165	GLU	3.5
1	A	36	ILE	3.5
1	C	52	LEU	3.5
1	C	185	PHE	3.4
1	C	10	ASP	3.3
1	C	9	VAL	3.2
1	A	15	PHE	3.2
1	D	13	ASP	3.2
1	B	19	THR	3.2
1	C	19	THR	3.1
1	A	137	SER	3.1
1	B	43	LEU	3.1
1	B	46	ASP	3.1
1	B	21	GLY	3.1
1	D	17	LYS	3.0
1	D	175	ASP	2.9
1	D	82	GLN	2.9
1	C	83	TYR	2.9
1	C	165	GLU	2.9
1	D	24	PHE	2.8
1	B	143	PHE	2.8
1	D	150	ILE	2.8
1	A	156	LEU	2.7
1	A	19	THR	2.7
1	A	94	LYS	2.7
1	C	24	PHE	2.6
1	B	179	THR	2.6
1	B	48	VAL	2.6
1	B	9	VAL	2.6
1	B	99	THR	2.5
1	D	77	LEU	2.5
1	D	172	GLY	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	180	PHE	2.5
1	B	185	PHE	2.5
1	A	60	LEU	2.5
1	D	52	LEU	2.5
1	B	8	PHE	2.5
1	B	38	HIS	2.4
1	C	170	PHE	2.4
1	C	117	THR	2.4
1	B	103	ILE	2.4
1	C	97	VAL	2.4
1	A	82	GLN	2.4
1	A	150	ILE	2.3
1	C	189	VAL	2.3
1	C	57	ILE	2.3
1	D	154	ILE	2.3
1	B	4	VAL	2.3
1	D	45	ASN	2.2
1	B	121	LYS	2.2
1	D	138	LYS	2.2
1	B	74	PRO	2.2
1	B	78	PHE	2.2
1	B	166	GLU	2.2
1	C	137	SER	2.2
1	A	61	ILE	2.1
1	B	41	SER	2.1
1	C	15	PHE	2.1
1	A	172	GLY	2.1
1	D	184	PRO	2.1
1	B	2	THR	2.1
1	A	105	VAL	2.0
1	C	48	VAL	2.0
1	D	185	PHE	2.0
1	D	118	LEU	2.0
1	A	113	ILE	2.0
1	A	170	PHE	2.0
1	A	145	LEU	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.