Full wwPDB X-ray Structure Validation Report (i)

Nov 1, 2021 – 12:55 pm GMT

PDB ID	:	7BDU
Title	:	Crystal structure of a Hsp47-collagen peptide complex
Authors	:	Abraham, E.T.; Gebauer, J.M.; Baumann, U.
Deposited on	:	2020-12-22
Resolution	:	2.49 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 following versions of software and data (see references (1)) were used in the production of this report:

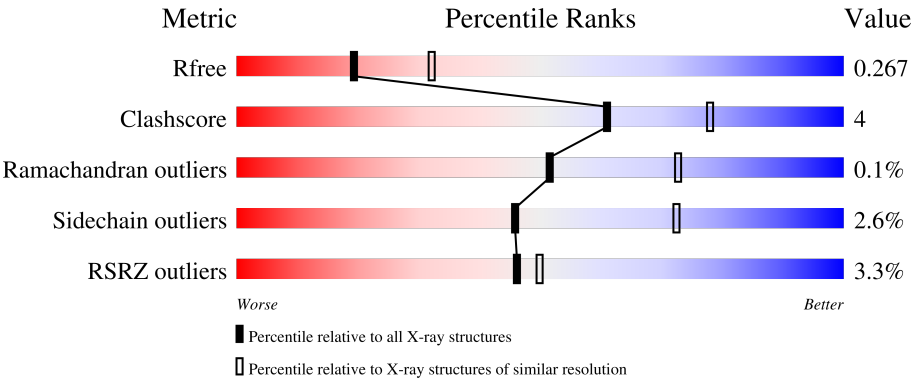
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.49 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.



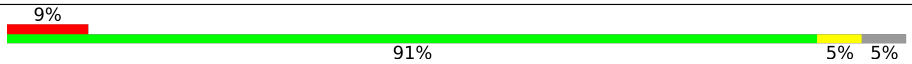
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	392	3% 84% 11% • 5%
1	B	392	4% 78% 16% 5%
2	C	22	5% 91% 9%
2	D	22	91% 5% 5%
2	E	22	5% 91% 5% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	22	
2	G	22	
2	H	22	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6726 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 Collagen-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	374	Total	C	N	O	S	0	0	0
			2944	1873	511	547	13			
1	B	371	Total	C	N	O	S	0	0	0
			2932	1865	511	543	13			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	35	MET	-	initiating methionine	UNP E2RHY7
A	419	LEU	-	expression tag	UNP E2RHY7
A	420	GLU	-	expression tag	UNP E2RHY7
A	421	HIS	-	expression tag	UNP E2RHY7
A	422	HIS	-	expression tag	UNP E2RHY7
A	423	HIS	-	expression tag	UNP E2RHY7
A	424	HIS	-	expression tag	UNP E2RHY7
A	425	HIS	-	expression tag	UNP E2RHY7
A	426	HIS	-	expression tag	UNP E2RHY7
B	35	MET	-	initiating methionine	UNP E2RHY7
B	419	LEU	-	expression tag	UNP E2RHY7
B	420	GLU	-	expression tag	UNP E2RHY7
B	421	HIS	-	expression tag	UNP E2RHY7
B	422	HIS	-	expression tag	UNP E2RHY7
B	423	HIS	-	expression tag	UNP E2RHY7
B	424	HIS	-	expression tag	UNP E2RHY7
B	425	HIS	-	expression tag	UNP E2RHY7
B	426	HIS	-	expression tag	UNP E2RHY7

- Molecule 2 is a protein called 21er collagen model peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	20	Total	C	N	O	0	0	0
			124	81	23	20			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	D	21	Total C N O 131 86 24 21	0	0	0
2	E	21	Total C N O 130 86 23 21	0	0	0
2	F	21	Total C N O 131 86 24 21	0	0	0
2	G	21	Total C N O 131 86 24 21	0	0	0
2	H	21	Total C N O 130 86 23 21	0	0	0

- Molecule 3 is water.

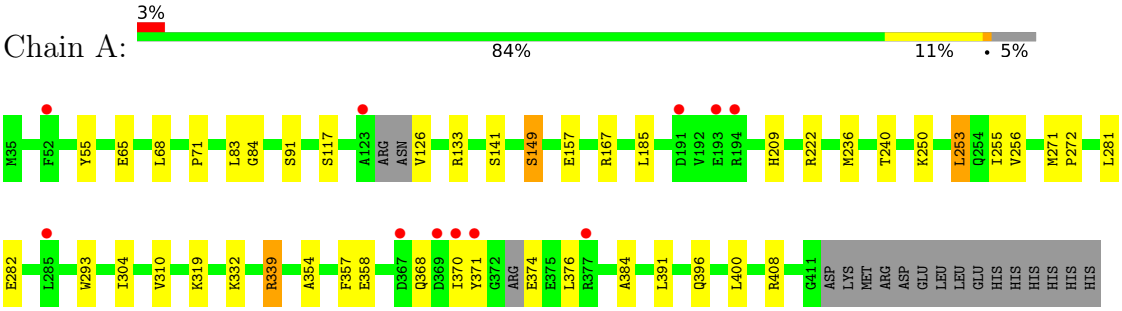
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	31	Total O 31 31	0	0
3	B	31	Total O 31 31	0	0
3	D	3	Total O 3 3	0	0
3	E	2	Total O 2 2	0	0
3	F	1	Total O 1 1	0	0
3	G	2	Total O 2 2	0	0
3	H	3	Total O 3 3	0	0

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.

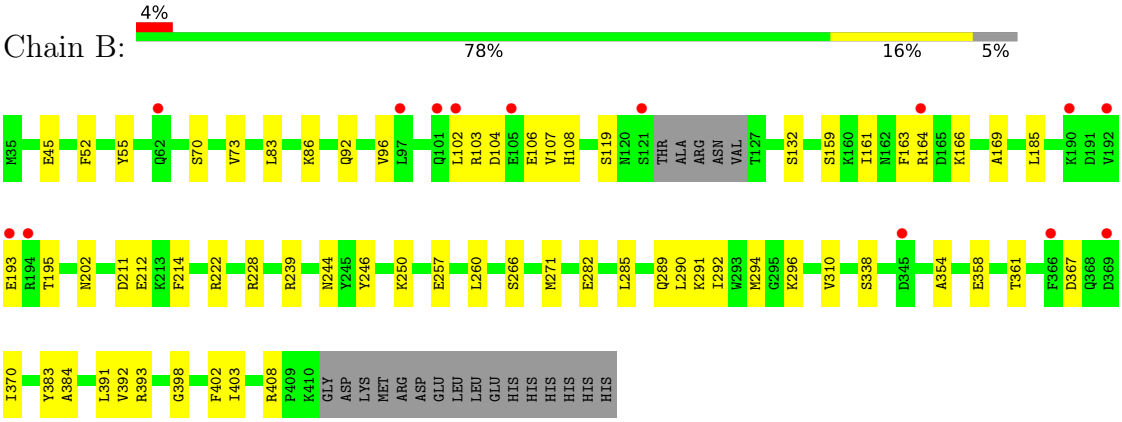
• Molecule 1: Collagen-binding protein

Chain A:



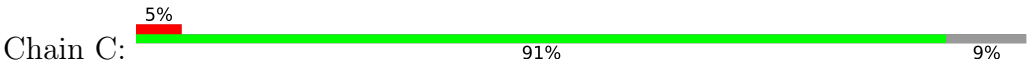
• Molecule 1: Collagen-binding protein

Chain B:



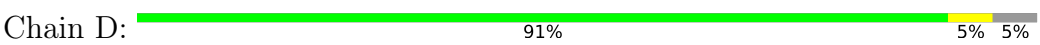
• Molecule 2: 21er collagen model peptide

Chain C:



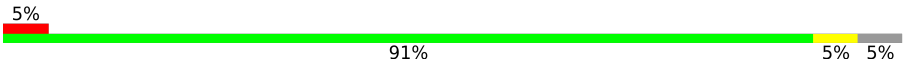
• Molecule 2: 21er collagen model peptide

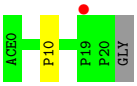
Chain D:



• Molecule 2: 21er collagen model peptide

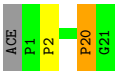
Chain E:





• Molecule 2: 21er collagen model peptide

Chain F:

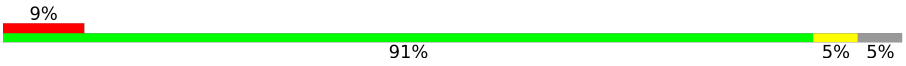


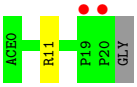
• Molecule 2: 21er collagen model peptide

Chain G:

• Molecule 2: 21er collagen model peptide

Chain H:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	91.82Å 129.80Å 173.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.91 – 2.49 45.91 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.0 (45.91-2.49) 98.4 (45.91-2.49)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.01 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.217 , 0.267 0.217 , 0.267	Depositor DCC
R_{free} test set	1843 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	50.9	Xtriage
Anisotropy	0.201	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6726	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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: ACE

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	0.24	0/3003	0.42	0/4049
1	B	0.24	0/2992	0.43	0/4034
2	C	0.30	0/134	0.35	0/188
2	D	0.32	0/142	0.39	0/200
2	E	0.32	0/139	0.46	0/198
2	F	0.29	0/142	0.36	0/200
2	G	0.32	0/142	0.38	0/200
2	H	0.33	0/139	0.41	0/198
All	All	0.25	0/6833	0.42	0/9267

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	2944	0	2968	25	0
1	B	2932	0	2958	31	0
2	C	124	0	122	0	0
2	D	131	0	131	1	0
2	E	130	0	129	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	131	0	131	2	0
2	G	131	0	131	4	0
2	H	130	0	129	1	0
3	A	31	0	0	1	0
3	B	31	0	0	0	0
3	D	3	0	0	0	0
3	E	2	0	0	0	0
3	F	1	0	0	0	0
3	G	2	0	0	0	0
3	H	3	0	0	0	0
All	All	6726	0	6699	60	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 4.

All (60) 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:368:GLN:HB2	1:A:370:ILE:HG12	1.60	0.83
1:A:368:GLN:OE1	1:A:370:ILE:HB	1.86	0.75
1:B:102:LEU:HD12	1:B:107:VAL:HG13	1.73	0.68
1:A:126:VAL:HG12	1:A:209:HIS:H	1.59	0.68
1:A:222:ARG:HG3	1:A:236:MET:HE2	1.77	0.67
1:B:228:ARG:NH2	2:G:14:PRO:O	2.30	0.65
1:A:374:GLU:N	3:A:503:HOH:O	2.29	0.64
1:A:71:PRO:HG2	1:A:400:LEU:HB3	1.82	0.61
2:F:20:PRO:HB3	2:G:19:PRO:HG2	1.85	0.59
2:F:2:PRO:HA	2:G:1:PRO:HG2	1.87	0.57
1:B:102:LEU:HD11	1:B:106:GLU:HB2	1.88	0.56
1:B:163:PHE:HB3	1:B:195:THR:HB	1.86	0.56
2:G:11:ARG:NH1	2:H:11:ARG:O	2.38	0.56
1:A:83:LEU:HD21	1:A:149:SER:HB3	1.88	0.54
1:A:253:LEU:HD13	1:A:281:LEU:HD21	1.90	0.53
1:B:55:TYR:HE1	1:B:282:GLU:HG2	1.74	0.53
1:B:193:GLU:N	1:B:193:GLU:OE2	2.41	0.53
1:B:244:ASN:HB2	1:B:257:GLU:HB3	1.91	0.52
1:B:246:TYR:CZ	1:B:296:LYS:HD3	2.45	0.52
1:A:55:TYR:HE1	1:A:282:GLU:HG2	1.74	0.52
1:A:310:VAL:HG22	1:A:358:GLU:HB2	1.91	0.52
1:B:367:ASP:HB3	1:B:370:ILE:HG13	1.93	0.51
1:B:393:ARG:NH1	1:B:398:GLY:O	2.44	0.50

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:65:GLU:OE2	1:A:408:ARG:NE	2.42	0.50
1:B:211:ASP:HB2	1:B:260:LEU:O	2.12	0.49
1:B:102:LEU:CD1	1:B:106:GLU:HB2	2.43	0.48
1:B:214:PHE:HB2	1:B:361:THR:HB	1.95	0.48
1:A:240:THR:HG22	1:A:376:LEU:HB3	1.97	0.47
1:B:86:LYS:HE3	1:B:338:SER:HB2	1.96	0.47
1:B:290:LEU:HD13	1:B:294:MET:HG2	1.97	0.46
1:A:84:GLY:HA2	1:A:339:ARG:HB2	1.97	0.46
2:D:11:ARG:HB2	2:E:10:PRO:HG2	1.97	0.46
1:A:319:LYS:HA	1:A:319:LYS:HD2	1.79	0.45
1:B:212:GLU:HB3	1:B:239:ARG:CZ	2.46	0.45
1:A:272:PRO:HD3	1:A:281:LEU:HD12	1.98	0.45
1:B:222:ARG:NH1	1:B:383:TYR:OH	2.47	0.45
1:A:256:VAL:HG21	1:A:304:ILE:HD11	1.99	0.44
1:A:400:LEU:HD12	1:A:400:LEU:HA	1.80	0.44
1:B:310:VAL:HG22	1:B:358:GLU:HB2	1.99	0.44
1:B:161:ILE:HD11	1:B:169:ALA:HA	1.98	0.44
1:B:119:SER:O	1:B:119:SER:OG	2.25	0.44
1:B:391:LEU:HB3	1:B:403:ILE:HG22	1.99	0.44
1:A:185:LEU:HD21	1:A:354:ALA:HB1	2.00	0.44
1:A:250:LYS:HE2	1:A:250:LYS:HB2	1.76	0.44
1:A:68:LEU:HD11	1:A:357:PHE:HB2	1.99	0.43
1:B:132:SER:OG	1:B:202:ASN:OD1	2.29	0.43
1:B:92:GLN:O	1:B:96:VAL:HG23	2.19	0.42
1:B:392:VAL:HB	1:B:402:PHE:HB2	2.02	0.42
1:B:185:LEU:HD21	1:B:354:ALA:HB1	2.01	0.42
1:A:133:ARG:HE	1:A:157:GLU:HB2	1.84	0.42
1:A:271:MET:SD	1:A:384:ALA:HA	2.60	0.42
1:B:70:SER:HB3	1:B:73:VAL:HB	2.02	0.42
1:B:52:PHE:HZ	1:B:285:LEU:O	2.03	0.41
1:B:83:LEU:HB2	1:B:108:HIS:CE1	2.55	0.41
1:A:255:ILE:HD12	1:A:293:TRP:HB3	2.02	0.41
1:A:332:LYS:HB3	1:A:332:LYS:HE2	1.81	0.41
1:B:163:PHE:H	1:B:164:ARG:NH2	2.18	0.41
1:B:271:MET:SD	1:B:384:ALA:HA	2.61	0.41
1:A:368:GLN:O	1:A:368:GLN:NE2	2.54	0.40
1:B:289:GLN:HA	1:B:292:ILE:HD12	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	368/392 (94%)	349 (95%)	19 (5%)	0	100 100
1	B	367/392 (94%)	351 (96%)	16 (4%)	0	100 100
2	C	18/22 (82%)	18 (100%)	0	0	100 100
2	D	19/22 (86%)	19 (100%)	0	0	100 100
2	E	19/22 (86%)	19 (100%)	0	0	100 100
2	F	19/22 (86%)	18 (95%)	0	1 (5%)	2 2
2	G	19/22 (86%)	19 (100%)	0	0	100 100
2	H	19/22 (86%)	19 (100%)	0	0	100 100
All	All	848/916 (93%)	812 (96%)	35 (4%)	1 (0%)	51 73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	20	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	319/337 (95%)	309 (97%)	10 (3%)	40 67
1	B	318/337 (94%)	309 (97%)	9 (3%)	43 70
2	C	13/14 (93%)	13 (100%)	0	100 100
2	D	14/14 (100%)	14 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	E	14/14 (100%)	14 (100%)	0	100	100
2	F	14/14 (100%)	14 (100%)	0	100	100
2	G	14/14 (100%)	14 (100%)	0	100	100
2	H	14/14 (100%)	14 (100%)	0	100	100
All	All	720/758 (95%)	701 (97%)	19 (3%)	46	72

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	91	SER
1	A	117	SER
1	A	141	SER
1	A	149	SER
1	A	167	ARG
1	A	253	LEU
1	A	339	ARG
1	A	371	TYR
1	A	391	LEU
1	A	396	GLN
1	B	45	GLU
1	B	103	ARG
1	B	104	ASP
1	B	159	SER
1	B	166	LYS
1	B	250	LYS
1	B	266	SER
1	B	291	LYS
1	B	408	ARG

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

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	374/392 (95%)	0.22	11 (2%) 51 55	34, 49, 76, 105	0
1	B	371/392 (94%)	0.29	14 (3%) 40 43	36, 57, 86, 112	0
2	C	20/22 (90%)	0.16	1 (5%) 28 30	39, 48, 65, 72	0
2	D	21/22 (95%)	-0.05	0 100 100	37, 45, 60, 62	0
2	E	20/22 (90%)	0.13	1 (5%) 28 30	40, 47, 72, 80	0
2	F	21/22 (95%)	-0.15	0 100 100	39, 45, 76, 86	0
2	G	21/22 (95%)	0.15	0 100 100	39, 49, 80, 81	0
2	H	20/22 (90%)	0.27	2 (10%) 7 6	40, 48, 93, 96	0
All	All	868/916 (94%)	0.23	29 (3%) 46 50	34, 52, 82, 112	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	370	ILE	7.4
1	B	101	GLN	7.0
1	A	371	TYR	5.5
1	B	193	GLU	4.2
1	A	193	GLU	4.1
1	A	377	ARG	4.1
1	A	123	ALA	3.8
1	B	194	ARG	2.8
1	A	367	ASP	2.7
1	B	164	ARG	2.6
1	A	52	PHE	2.5
2	H	20	PRO	2.5
1	A	194	ARG	2.5
1	B	345	ASP	2.4
1	B	105	GLU	2.4
2	C	2	PRO	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	366	PHE	2.3
1	B	192	VAL	2.3
2	E	19	PRO	2.3
1	A	369	ASP	2.3
1	B	190	LYS	2.2
1	B	97	LEU	2.2
2	H	19	PRO	2.2
1	B	102	LEU	2.1
1	B	369	ASP	2.1
1	A	285	LEU	2.1
1	B	62	GLN	2.1
1	A	191	ASP	2.0
1	B	121	SER	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.