wwPDB X-ray Structure Validation Summary Report (i)

Jun 19, 2024 – 04:06 AM EDT

PDB ID	:	4AU3
Title	:	Crystal Structure of a Hsp47-collagen complex
Authors	:	Widmer, C.; Gebauer, J.M.; Brunstein, E.; Rodenbaum, S.; Zaucke, F.; Drogemuller, C.; Leeb, T.; Baumann, U.
Deposited on	:	2012-05-14
Resolution	:	2.78 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i)) were used in the production of this report:

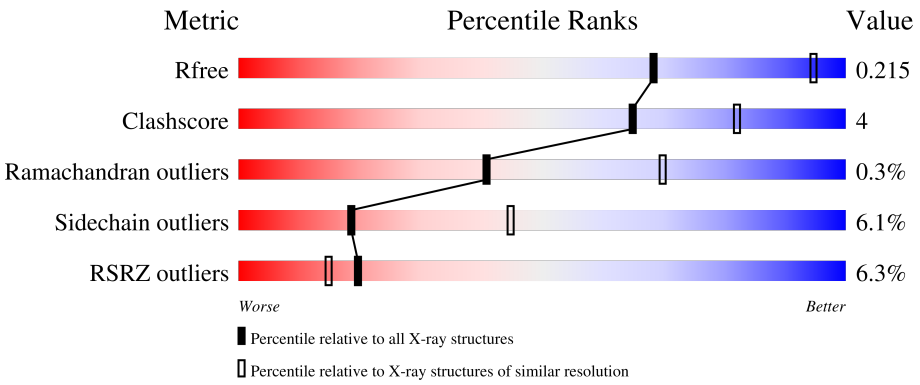
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.78 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	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (2.80-2.76)

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	% 84% 9% • 6%
1	C	392	14% 79% 15% • 6%
1	D	392	3% 81% 11% • 5%
2	B	392	6% 78% 12% • 7%
3	E	20	70% 5% 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	20	80% 20%
3	G	20	70% 5% 25%
3	H	20	80% 5% 15%
3	I	20	70% 10% 20%
3	J	20	65% 10% 25%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 12171 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 SERPIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	367	2886	1839	501	532	1	13	0	0	0
1	C	370	2911	1854	505	538	1	13	0	0	0
1	D	372	2926	1861	513	538	1	13	0	0	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	35	MSE	-	expression tag	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
C	35	MSE	-	expression tag	UNP E2RHY7
C	419	LEU	-	expression tag	UNP E2RHY7
C	420	GLU	-	expression tag	UNP E2RHY7
C	421	HIS	-	expression tag	UNP E2RHY7
C	422	HIS	-	expression tag	UNP E2RHY7
C	423	HIS	-	expression tag	UNP E2RHY7
C	424	HIS	-	expression tag	UNP E2RHY7
C	425	HIS	-	expression tag	UNP E2RHY7
C	426	HIS	-	expression tag	UNP E2RHY7
D	35	MSE	-	expression tag	UNP E2RHY7
D	419	LEU	-	expression tag	UNP E2RHY7
D	420	GLU	-	expression tag	UNP E2RHY7
D	421	HIS	-	expression tag	UNP E2RHY7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	422	HIS	-	expression tag	UNP E2RHY7
D	423	HIS	-	expression tag	UNP E2RHY7
D	424	HIS	-	expression tag	UNP E2RHY7
D	425	HIS	-	expression tag	UNP E2RHY7
D	426	HIS	-	expression tag	UNP E2RHY7

- Molecule 2 is a protein called SERPYIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	B	364	2856	1817	498	527	1	13	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	35	MSE	-	expression tag	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 3 is a protein called 18ER COLLAGEN MODEL PEPTIDE 15-R8.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	E	15	94	61	18	15	0	0	0
3	F	16	101	66	19	16	0	0	0
3	G	15	94	61	18	15	0	0	0
3	H	17	108	71	20	17	0	0	0
3	I	16	101	66	19	16	0	0	0
3	J	15	94	61	18	15	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

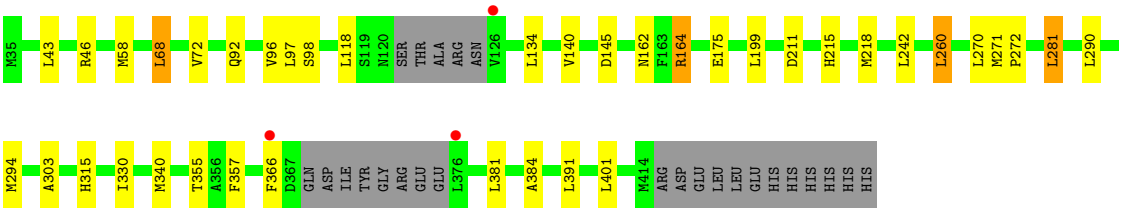
Chain	Residue	Modelled	Actual	Comment	Reference
E	0	ACE	-	expression tag	UNP Q96A83
E	19	NH2	-	expression tag	UNP Q96A83
F	0	ACE	-	expression tag	UNP Q96A83
F	19	NH2	-	expression tag	UNP Q96A83
G	0	ACE	-	expression tag	UNP Q96A83
G	19	NH2	-	expression tag	UNP Q96A83
H	0	ACE	-	expression tag	UNP Q96A83
H	19	NH2	-	expression tag	UNP Q96A83
I	0	ACE	-	expression tag	UNP Q96A83
I	19	NH2	-	expression tag	UNP Q96A83
J	0	ACE	-	expression tag	UNP Q96A83
J	19	NH2	-	expression tag	UNP Q96A83

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.

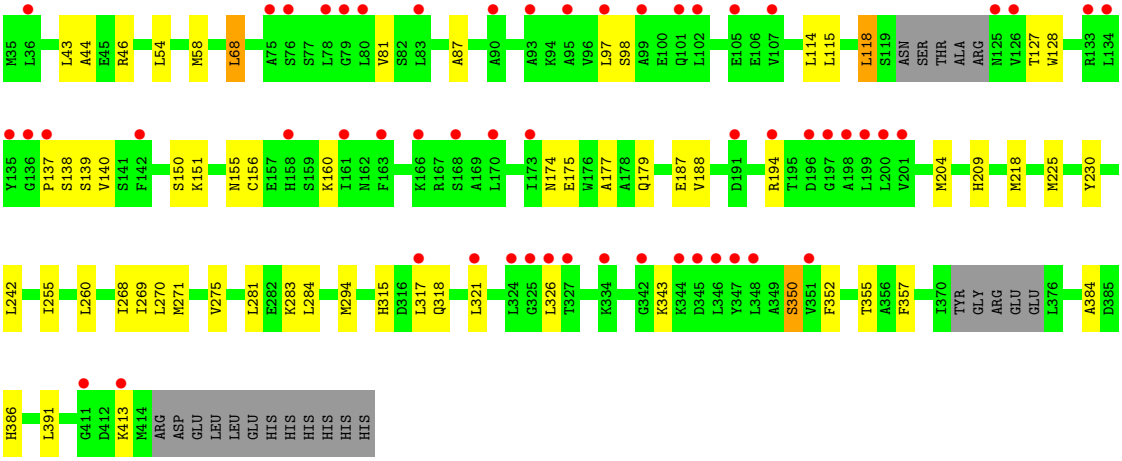
• Molecule 1: SERPIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1)

Chain A:



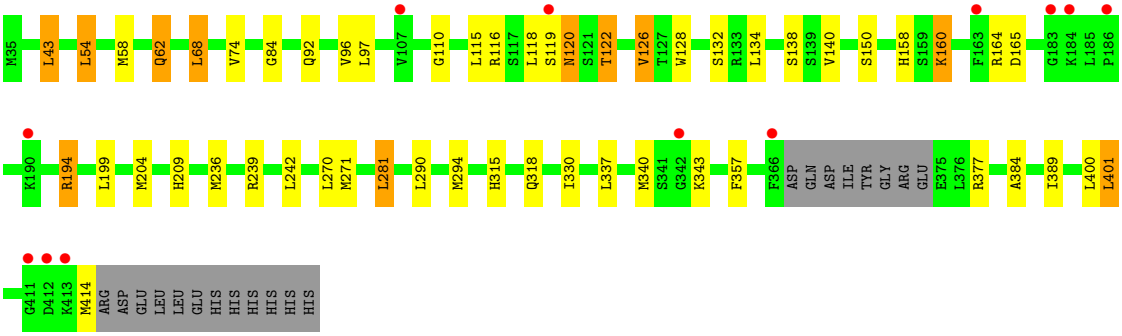
• Molecule 1: SERPIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1)

Chain C:



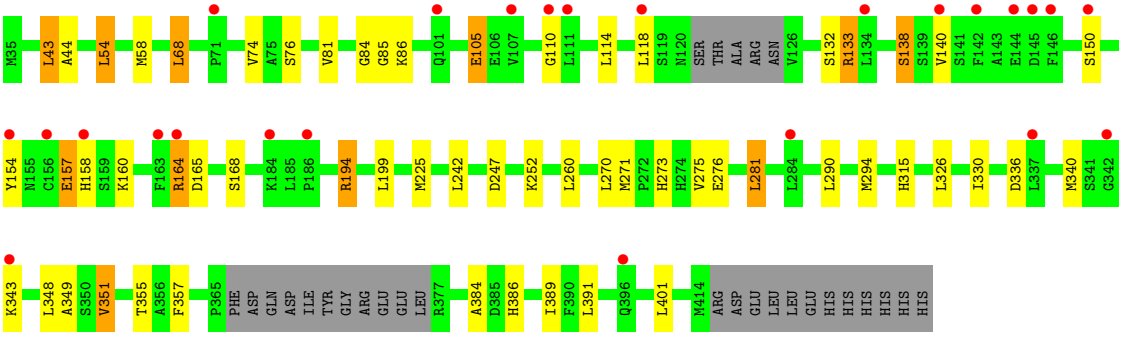
• Molecule 1: SERPIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1)

Chain D:



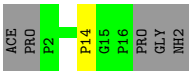
• Molecule 2: SERPYIN PEPTIDASE INHIBITOR, CLADE H (HEAT SHOCK PROTEIN 47), MEMBER 1, (COLLAGEN BINDING PROTEIN 1)

Chain B:



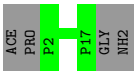
• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain E:



• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain F:

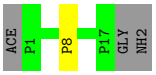


• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain G:

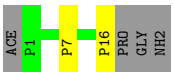
• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain H:



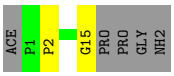
• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain I:



• Molecule 3: 18ER COLLAGEN MODEL PEPTIDE 15-R8

Chain J:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	101.26Å 101.26Å 366.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.74 – 2.78 91.75 – 2.68	Depositor EDS
% Data completeness (in resolution range)	98.7 (91.74-2.78) 97.8 (91.75-2.68)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 2.69Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.203 , 0.234 0.216 , 0.215	Depositor DCC
R_{free} test set	1463 reflections (2.72%)	wwPDB-VP
Wilson B-factor (Å^2)	62.8	Xtriage
Anisotropy	0.131	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 54.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12171	wwPDB-VP
Average B, all atoms (Å^2)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.18% 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 (i)

5.1 Standard geometry (i)

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	0.54	0/2932	0.73	0/3932
1	C	0.50	0/2957	0.72	0/3966
1	D	0.49	0/2973	0.73	1/3988 (0.0%)
2	B	0.49	0/2901	0.73	0/3889
3	E	0.44	0/102	0.43	0/143
3	F	0.45	0/110	0.48	0/155
3	G	0.52	0/102	0.44	0/143
3	H	0.55	0/118	0.44	0/167
3	I	0.44	0/110	0.48	0/155
3	J	0.54	0/102	0.46	0/143
All	All	0.51	0/12407	0.72	1/16681 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	164	ARG	C-N-CA	5.16	134.60	121.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	2886	0	2919	15	0
1	C	2911	0	2942	30	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	2926	0	2965	22	0
2	B	2856	0	2883	31	0
3	E	94	0	91	1	0
3	F	101	0	98	0	0
3	G	94	0	93	1	0
3	H	108	0	107	1	0
3	I	101	0	100	2	0
3	J	94	0	93	2	0
All	All	12171	0	12291	99	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.

The worst 5 of 99 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:230:TYR:HE1	1:C:413:LYS:HB3	1.47	0.77
1:D:43:LEU:HD23	1:D:110:GLY:HA3	1.68	0.74
1:C:58:MSE:HE1	1:C:315:HIS:HB3	1.70	0.73
1:A:72:VAL:HG21	1:A:118:LEU:HD21	1.71	0.73
1:A:58:MSE:HE1	1:A:315:HIS:HB3	1.76	0.68

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	361/392 (92%)	350 (97%)	10 (3%)	1 (0%)	41 70
1	C	364/392 (93%)	346 (95%)	16 (4%)	2 (0%)	29 58
1	D	368/392 (94%)	349 (95%)	17 (5%)	2 (0%)	29 58

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	358/392 (91%)	348 (97%)	10 (3%)	0	100	100
3	E	13/20 (65%)	13 (100%)	0	0	100	100
3	F	14/20 (70%)	14 (100%)	0	0	100	100
3	G	13/20 (65%)	13 (100%)	0	0	100	100
3	H	15/20 (75%)	15 (100%)	0	0	100	100
3	I	14/20 (70%)	14 (100%)	0	0	100	100
3	J	13/20 (65%)	13 (100%)	0	0	100	100
All	All	1533/1688 (91%)	1475 (96%)	53 (4%)	5 (0%)	41	70

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	165	ASP
1	A	366	PHE
1	C	138	SER
1	C	87	ALA
1	D	377	ARG

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	313/324 (97%)	298 (95%)	15 (5%)	25	55
1	C	316/324 (98%)	299 (95%)	17 (5%)	22	50
1	D	317/324 (98%)	292 (92%)	25 (8%)	12	31
2	B	309/325 (95%)	285 (92%)	24 (8%)	12	32
3	E	10/12 (83%)	10 (100%)	0	100	100
3	F	11/12 (92%)	11 (100%)	0	100	100
3	G	10/12 (83%)	10 (100%)	0	100	100
3	H	12/12 (100%)	12 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	I	11/12 (92%)	11 (100%)	0	100	100
3	J	10/12 (83%)	10 (100%)	0	100	100
All	All	1319/1369 (96%)	1238 (94%)	81 (6%)	18	45

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	54	LEU
1	D	239	ARG
1	D	68	LEU
1	D	134	LEU
1	D	318	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	174	ASN
1	D	62	GLN
1	D	262	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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	354/392 (90%)	0.04	3 (0%) 86 84	35, 55, 89, 130	0
1	C	357/392 (91%)	0.66	55 (15%) 2 1	33, 82, 153, 179	0
1	D	359/392 (91%)	0.27	12 (3%) 46 41	41, 79, 139, 166	0
2	B	351/392 (89%)	0.48	25 (7%) 16 11	42, 82, 127, 178	0
3	E	15/20 (75%)	0.17	0 100 100	47, 60, 89, 98	0
3	F	16/20 (80%)	0.18	0 100 100	49, 60, 91, 95	0
3	G	15/20 (75%)	0.11	0 100 100	39, 51, 86, 97	0
3	H	17/20 (85%)	-0.08	0 100 100	37, 46, 75, 83	0
3	I	16/20 (80%)	-0.03	0 100 100	40, 51, 86, 94	0
3	J	15/20 (75%)	0.00	0 100 100	33, 48, 78, 98	0
All	All	1515/1688 (89%)	0.34	95 (6%) 20 15	33, 70, 138, 179	0

The worst 5 of 95 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	413	LYS	8.8
2	B	342	GLY	8.0
1	A	366	PHE	7.6
2	B	343	LYS	6.6
1	C	348	LEU	6.1

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.