Full wwPDB X-ray Structure Validation Report ⓘ

Oct 12, 2020 – 10:46 AM EDT

PDB ID	:	4DGJ
Title	:	Structure of a human enteropeptidase light chain variant
Authors	:	Zahn, M.; Simeonov, P.; Straeter, N.
Deposited on	:	2012-01-26
Resolution	:	1.90 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:

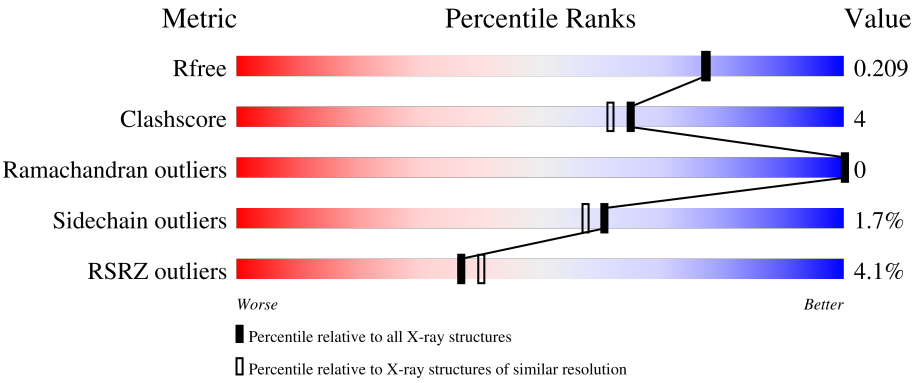
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.14.6
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	235	3% 86% 13%
1	B	235	5% 84% 14%
1	C	235	3% 92% 7%
1	D	235	6% 89% 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8393 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 Enteropeptidase catalytic light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	235	Total	C	N	O	S	0	8	0
			1888	1195	322	357	14			
1	B	235	Total	C	N	O	S	0	6	0
			1878	1186	319	359	14			
1	C	234	Total	C	N	O	S	0	4	0
			1863	1176	319	354	14			
1	D	234	Total	C	N	O	S	0	0	0
			1841	1162	315	350	14			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	ASP	ASN	engineered mutation	UNP P98073
A	21	ASP	GLY	engineered mutation	UNP P98073
A	22	ASP	GLY	engineered mutation	UNP P98073
A	112	SER	CYS	engineered mutation	UNP P98073
A	142	ASP	ASN	engineered mutation	UNP P98073
A	210	GLU	LYS	engineered mutation	UNP P98073
B	6	ASP	ASN	engineered mutation	UNP P98073
B	21	ASP	GLY	engineered mutation	UNP P98073
B	22	ASP	GLY	engineered mutation	UNP P98073
B	112	SER	CYS	engineered mutation	UNP P98073
B	142	ASP	ASN	engineered mutation	UNP P98073
B	210	GLU	LYS	engineered mutation	UNP P98073
C	6	ASP	ASN	engineered mutation	UNP P98073
C	21	ASP	GLY	engineered mutation	UNP P98073
C	22	ASP	GLY	engineered mutation	UNP P98073
C	112	SER	CYS	engineered mutation	UNP P98073
C	142	ASP	ASN	engineered mutation	UNP P98073
C	210	GLU	LYS	engineered mutation	UNP P98073
D	6	ASP	ASN	engineered mutation	UNP P98073
D	21	ASP	GLY	engineered mutation	UNP P98073
D	22	ASP	GLY	engineered mutation	UNP P98073

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	112	SER	CYS	engineered mutation	UNP P98073
D	142	ASP	ASN	engineered mutation	UNP P98073
D	210	GLU	LYS	engineered mutation	UNP P98073

- Molecule 2 is water.

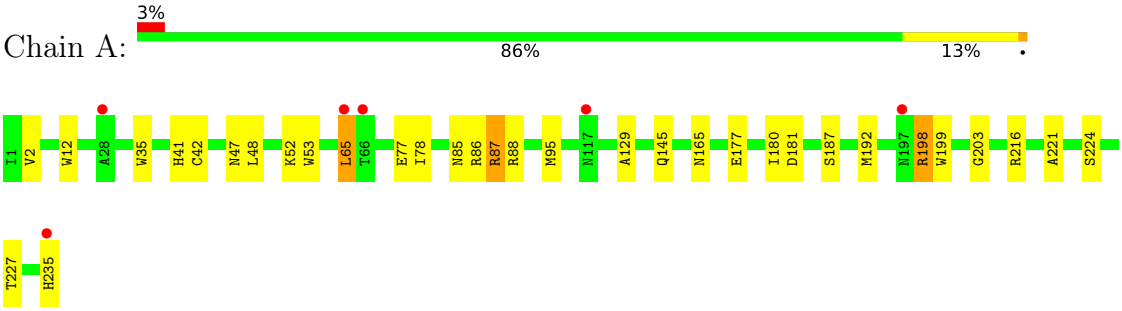
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	291	Total O 291 291	0	0
2	B	257	Total O 257 257	0	0
2	C	188	Total O 188 188	0	0
2	D	187	Total O 187 187	0	0

3 Residue-property plots

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.

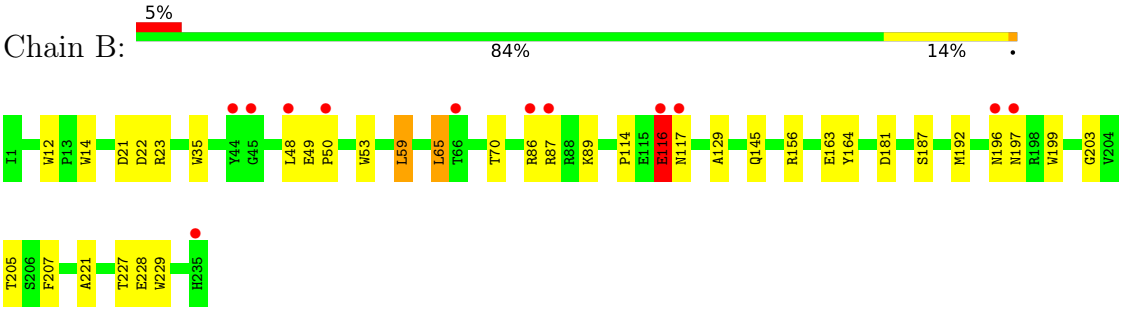
• Molecule 1: Enteropeptidase catalytic light chain

Chain A:



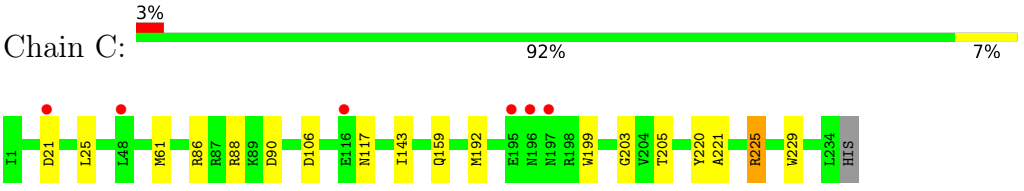
• Molecule 1: Enteropeptidase catalytic light chain

Chain B:



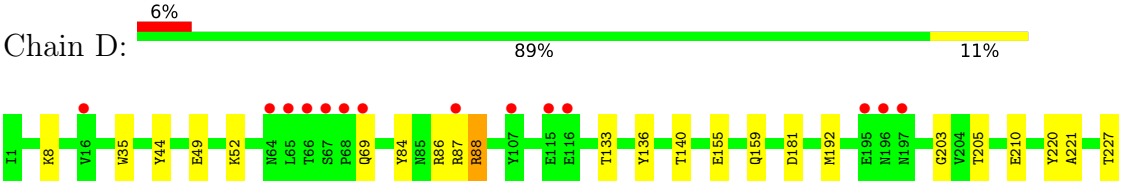
• Molecule 1: Enteropeptidase catalytic light chain

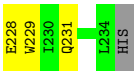
Chain C:



• Molecule 1: Enteropeptidase catalytic light chain

Chain D:




E228
W229
I230
Q231
L234
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	53.18Å 147.70Å 148.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 1.90 24.78 – 1.90	Depositor EDS
% Data completeness (in resolution range)	97.4 (25.00-1.90) 97.5 (24.78-1.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.37 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.0	Depositor
R, R_{free}	0.169 , 0.210 0.168 , 0.209	Depositor DCC
R_{free} test set	1834 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å^2)	16.2	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 51.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.018 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8393	wwPDB-VP
Average B, all atoms (Å^2)	20.0	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 23.39 % 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 4.7072e-03.*

[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	1.19	6/1961 (0.3%)	1.02	3/2671 (0.1%)
1	B	1.21	7/1945 (0.4%)	1.03	6/2649 (0.2%)
1	C	1.15	4/1923 (0.2%)	1.10	8/2619 (0.3%)
1	D	1.03	3/1889 (0.2%)	0.95	2/2574 (0.1%)
All	All	1.15	20/7718 (0.3%)	1.03	19/10513 (0.2%)

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	35	TRP	CD2-CE2	7.01	1.49	1.41
1	B	229	TRP	CD2-CE2	6.71	1.49	1.41
1	A	35	TRP	CD2-CE2	6.51	1.49	1.41
1	A	199	TRP	CD2-CE2	6.31	1.49	1.41
1	B	14	TRP	CD2-CE2	6.26	1.48	1.41
1	B	35	TRP	CD2-CE2	6.14	1.48	1.41
1	C	229	TRP	CD2-CE2	6.04	1.48	1.41
1	B	116	GLU	CD-OE1	6.02	1.32	1.25
1	D	229	TRP	CD2-CE2	5.99	1.48	1.41
1	A	53	TRP	CD2-CE2	5.97	1.48	1.41
1	D	155	GLU	CD-OE1	5.89	1.32	1.25
1	B	199	TRP	CD2-CE2	5.79	1.48	1.41
1	B	12	TRP	CD2-CE2	5.67	1.48	1.41
1	C	225	ARG	CD-NE	-5.66	1.36	1.46
1	C	199	TRP	CD2-CE2	5.59	1.48	1.41
1	A	41	HIS	CG-CD2	5.35	1.44	1.35
1	A	177	GLU	CG-CD	5.24	1.59	1.51
1	A	12	TRP	CD2-CE2	5.15	1.47	1.41
1	C	229	TRP	CG-CD1	5.11	1.44	1.36
1	B	53	TRP	CD2-CE2	5.02	1.47	1.41

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	225	ARG	NE-CZ-NH2	-17.40	111.60	120.30
1	C	225	ARG	NE-CZ-NH1	9.12	124.86	120.30
1	C	86	ARG	NE-CZ-NH2	-7.87	116.36	120.30
1	C	86	ARG	NE-CZ-NH1	6.91	123.75	120.30
1	B	156	ARG	NE-CZ-NH2	6.85	123.73	120.30
1	D	88	ARG	NE-CZ-NH2	-6.76	116.92	120.30
1	B	181	ASP	CB-CG-OD2	-6.14	112.78	118.30
1	B	181	ASP	CB-CG-OD1	6.03	123.73	118.30
1	B	65	LEU	CA-CB-CG	5.73	128.47	115.30
1	B	156	ARG	NE-CZ-NH1	-5.56	117.52	120.30
1	C	106	ASP	CB-CG-OD1	5.54	123.29	118.30
1	C	90	ASP	CB-CG-OD1	5.52	123.27	118.30
1	B	59	LEU	CB-CG-CD2	-5.50	101.64	111.00
1	A	95	MET	CG-SD-CE	5.37	108.79	100.20
1	A	181	ASP	CB-CG-OD2	-5.27	113.56	118.30
1	C	21	ASP	CB-CG-OD1	5.24	123.02	118.30
1	D	181	ASP	CB-CG-OD2	-5.17	113.65	118.30
1	A	216	ARG	NE-CZ-NH1	5.08	122.84	120.30
1	C	25	LEU	CB-CG-CD1	-5.04	102.42	111.00

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	1888	0	1825	23	0
1	B	1878	0	1800	22	0
1	C	1863	0	1785	6	0
1	D	1841	0	1755	17	0
2	A	291	0	0	1	0
2	B	257	0	0	3	0
2	C	188	0	0	0	0
2	D	187	0	0	3	0
All	All	8393	0	7165	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:B:21:ASP:HB3	2:B:307:HOH:O	1.50	1.11
1:A:86:ARG:H	1:A:87:ARG:HH11	1.29	0.79
1:A:85:ASN:OD1	1:A:87:ARG:HG2	1.89	0.73
1:A:87:ARG:H	1:A:87:ARG:HE	1.38	0.71
1:A:198:ARG:HH11	1:A:198:ARG:CG	2.06	0.68
1:B:48:LEU:HD21	1:D:84:TYR:HD2	1.59	0.66
1:A:198:ARG:HH11	1:A:198:ARG:HB3	1.61	0.64
1:B:114:PRO:O	1:B:227[B]:THR:HG21	1.99	0.63
1:A:198:ARG:HH11	1:A:198:ARG:HG2	1.63	0.62
1:C:61:MET:HG2	1:C:143:ILE:HG12	1.81	0.62
1:B:164:TYR:O	1:C:225:ARG:NH2	2.33	0.60
1:B:48:LEU:HG	1:D:44:TYR:HE1	1.65	0.60
1:D:227:THR:O	1:D:231:GLN:HG3	2.01	0.59
1:A:88[B]:ARG:NH1	1:D:228:GLU:OE2	2.32	0.58
1:A:198:ARG:HH11	1:A:198:ARG:CB	2.17	0.58
1:B:48:LEU:HD21	1:D:84:TYR:CD2	2.40	0.56
1:B:116:GLU:OE2	1:B:228[A]:GLU:HG3	2.06	0.55
1:B:163:GLU:O	1:C:117:ASN:HA	2.06	0.55
1:A:86:ARG:H	1:A:87:ARG:NH1	2.03	0.55
1:C:203:GLY:HA2	1:C:221:ALA:O	2.08	0.53
1:A:87:ARG:NE	1:A:87:ARG:H	2.04	0.53
1:A:65:LEU:HD12	1:A:65:LEU:H	1.73	0.53
1:A:42:CYS:SG	1:A:187[B]:SER:OG	2.68	0.52
1:A:198:ARG:NH1	1:A:198:ARG:HB3	2.26	0.51
1:D:203:GLY:HA2	1:D:221:ALA:O	2.11	0.51
1:D:159:GLN:HG3	2:D:440:HOH:O	2.10	0.51
1:B:187[B]:SER:HA	1:B:205:THR:HB	1.93	0.50
1:B:48:LEU:CD2	1:D:84:TYR:HD2	2.25	0.49
1:B:203:GLY:HA2	1:B:221:ALA:O	2.13	0.48
1:B:65:LEU:O	1:B:65:LEU:HD13	2.13	0.48
1:D:87:ARG:HG2	1:D:87:ARG:HH11	1.79	0.47
1:D:133:THR:HA	1:D:140:THR:HA	1.97	0.47
1:D:136:TYR:CZ	1:D:210:GLU:HG2	2.50	0.47
1:B:87:ARG:CZ	1:B:87:ARG:HB2	2.45	0.46
1:A:203:GLY:HA2	1:A:221:ALA:O	2.16	0.46
1:A:187[A]:SER:HB2	2:A:487:HOH:O	2.15	0.46
1:B:89:LYS:HD3	1:B:207:PHE:CG	2.50	0.45
1:D:205:THR:HA	1:D:220:TYR:CD2	2.52	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:49:GLU:HA	1:B:50:PRO:HD3	1.84	0.44
1:A:198:ARG:NH1	1:A:198:ARG:HG2	2.32	0.44
1:D:69:GLN:HG2	1:D:69:GLN:H	1.66	0.44
1:A:88[A]:ARG:NH2	1:A:165:ASN:OD1	2.49	0.44
1:A:129:ALA:HA	1:A:145:GLN:O	2.19	0.43
1:A:224:SER:HA	1:A:227[B]:THR:HG23	2.00	0.43
1:C:205:THR:HA	1:C:220:TYR:CD2	2.54	0.43
1:A:77:GLU:OE2	1:C:88[B]:ARG:NH2	2.52	0.43
1:B:23:ARG:HD3	2:B:346:HOH:O	2.19	0.43
1:D:86:ARG:HG3	2:D:449:HOH:O	2.18	0.42
1:D:49:GLU:OE1	1:D:52:LYS:HE3	2.18	0.42
1:A:52:LYS:HE2	1:A:52:LYS:HB3	1.58	0.42
1:B:129:ALA:HA	1:B:145:GLN:O	2.19	0.42
1:A:47:ASN:HB2	1:A:78:ILE:HB	2.01	0.42
1:B:116:GLU:OE1	1:B:228[A]:GLU:HG2	2.19	0.42
1:B:21:ASP:CB	2:B:307:HOH:O	2.34	0.42
1:D:87:ARG:HH22	1:D:88:ARG:HH21	1.68	0.41
1:B:59:LEU:HA	1:B:70[B]:THR:HG22	2.02	0.41
1:B:196:ASN:O	1:B:197:ASN:HB2	2.21	0.41
1:B:228[A]:GLU:H	1:B:228[A]:GLU:HG2	1.59	0.41
1:A:2:VAL:O	1:A:180:ILE:HA	2.20	0.41
1:D:8:LYS:HG2	2:D:397:HOH:O	2.21	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	241/235 (103%)	233 (97%)	8 (3%)	0	100	100
1	B	239/235 (102%)	229 (96%)	10 (4%)	0	100	100
1	C	236/235 (100%)	231 (98%)	5 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	232/235 (99%)	225 (97%)	7 (3%)	0	100	100
All	All	948/940 (101%)	918 (97%)	30 (3%)	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	209/201 (104%)	203 (97%)	6 (3%)	42	35
1	B	207/201 (103%)	202 (98%)	5 (2%)	49	43
1	C	204/201 (102%)	201 (98%)	3 (2%)	65	62
1	D	200/201 (100%)	199 (100%)	1 (0%)	88	89
All	All	820/804 (102%)	805 (98%)	15 (2%)	60	55

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

Mol	Chain	Res	Type
1	A	48	LEU
1	A	65	LEU
1	A	87	ARG
1	A	192	MET
1	A	198	ARG
1	A	235	HIS
1	B	22	ASP
1	B	86	ARG
1	B	116	GLU
1	B	117	ASN
1	B	192	MET
1	C	159[A]	GLN
1	C	159[B]	GLN
1	C	192	MET
1	D	192	MET

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

Mol	Chain	Res	Type
1	B	194	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

There are no ligands in this entry.

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	235/235 (100%)	-0.11	6 (2%)	56	58	7, 14, 31, 49	0
1	B	235/235 (100%)	0.05	12 (5%)	28	31	8, 14, 39, 53	0
1	C	234/235 (99%)	-0.10	6 (2%)	56	58	9, 15, 31, 52	0
1	D	234/235 (99%)	0.11	14 (5%)	21	24	10, 23, 43, 56	0
All	All	938/940 (99%)	-0.01	38 (4%)	37	40	7, 16, 37, 56	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	117	ASN	4.3
1	C	196	ASN	4.0
1	C	116	GLU	3.9
1	A	117	ASN	3.9
1	C	48	LEU	3.8
1	B	197	ASN	3.6
1	A	65	LEU	3.6
1	B	48	LEU	3.6
1	D	66	THR	3.6
1	A	235	HIS	3.6
1	B	66	THR	3.4
1	D	87	ARG	3.3
1	D	196	ASN	3.0
1	A	66	THR	2.9
1	B	87	ARG	2.8
1	D	67	SER	2.8
1	A	197	ASN	2.8
1	D	116	GLU	2.8
1	B	45	GLY	2.8
1	D	65	LEU	2.7
1	D	69	GLN	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	197	ASN	2.6
1	B	50	PRO	2.6
1	B	44	TYR	2.5
1	B	235	HIS	2.5
1	D	68	PRO	2.3
1	C	195	GLU	2.3
1	C	21	ASP	2.3
1	D	115	GLU	2.2
1	B	196	ASN	2.2
1	B	116	GLU	2.2
1	D	16	VAL	2.2
1	A	28	ALA	2.1
1	B	86	ARG	2.1
1	D	195	GLU	2.1
1	D	107	TYR	2.0
1	D	64	ASN	2.0
1	D	197	ASN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.