



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2023 – 05:22 PM EDT

PDB ID : 7JOS
Title : Crystal structure of BbKI complexed with Human Kallikrein 4
Authors : Li, M.; Wlodawer, A.; Gustchina, A.
Deposited on : 2020-08-07
Resolution : 2.10 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 ⓘ 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 : 2.35.1
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) : 2.35.1

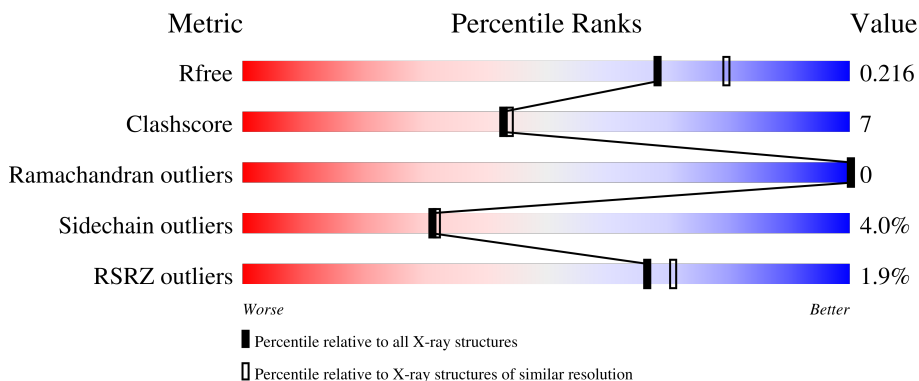
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 $\leq 5\%$. 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	223	87% 12% .
1	C	223	89% 10% .
1	E	223	91% 7% .
1	G	223	95% 5% .
1	I	223	89% 9% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	223	<p>% 89% 10% .</p>
1	M	223	<p>91% 8% .</p>
2	B	166	<p>3% 86% 11% ..</p>
2	D	166	<p>2% 90% 6% ..</p>
2	F	166	<p>2% 86% 10% ..</p>
2	H	166	<p>2% 89% 8% ..</p>
2	J	166	<p>2% 86% 11% ..</p>
2	L	166	<p>2% 91% 7% ..</p>
2	N	166	<p>4% 86% 11% ..</p>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 22065 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 Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	223	Total	C	N	O	S	0	3	0
			1687	1057	284	329	17			
1	C	223	Total	C	N	O	S	0	2	0
			1679	1052	281	329	17			
1	E	223	Total	C	N	O	S	0	2	0
			1679	1052	281	329	17			
1	G	223	Total	C	N	O	S	0	0	0
			1670	1043	281	329	17			
1	I	223	Total	C	N	O	S	0	2	0
			1679	1052	281	329	17			
1	K	223	Total	C	N	O	S	0	2	0
			1680	1051	281	331	17			
1	M	223	Total	C	N	O	S	0	0	0
			1670	1043	281	329	17			

- Molecule 2 is a protein called Kunitz-type inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	165	Total	C	N	O	S	0	1	0
			1285	821	222	241	1			
2	D	165	Total	C	N	O	S	0	1	0
			1280	818	219	242	1			
2	F	165	Total	C	N	O	S	0	0	0
			1277	816	219	241	1			
2	H	165	Total	C	N	O	S	0	1	0
			1285	821	222	241	1			
2	J	165	Total	C	N	O	S	0	1	0
			1285	821	222	241	1			
2	L	165	Total	C	N	O	S	0	1	0
			1285	821	222	241	1			
2	N	165	Total	C	N	O	S	0	1	0
			1285	821	222	241	1			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP Q6VEQ7
D	0	GLY	-	expression tag	UNP Q6VEQ7
F	0	GLY	-	expression tag	UNP Q6VEQ7
H	0	GLY	-	expression tag	UNP Q6VEQ7
J	0	GLY	-	expression tag	UNP Q6VEQ7
L	0	GLY	-	expression tag	UNP Q6VEQ7
N	0	GLY	-	expression tag	UNP Q6VEQ7

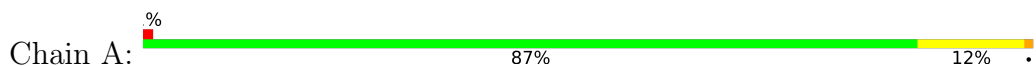
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	153	Total 153	O 153	0	0
3	B	109	Total 109	O 109	0	0
3	C	121	Total 121	O 121	0	0
3	D	74	Total 74	O 74	0	0
3	E	91	Total 91	O 91	0	0
3	F	100	Total 100	O 100	0	0
3	G	59	Total 59	O 59	0	0
3	H	68	Total 68	O 68	0	0
3	I	131	Total 131	O 131	0	0
3	J	70	Total 70	O 70	0	0
3	K	84	Total 84	O 84	0	0
3	L	80	Total 80	O 80	0	0
3	M	115	Total 115	O 115	0	0
3	N	84	Total 84	O 84	0	0

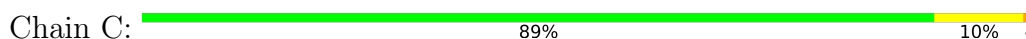
3 Residue-property plots i

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.

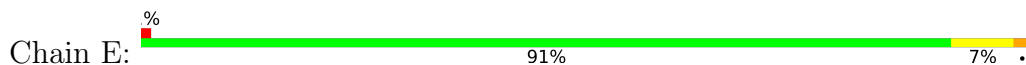
- Molecule 1: Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a



- Molecule 1: Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a



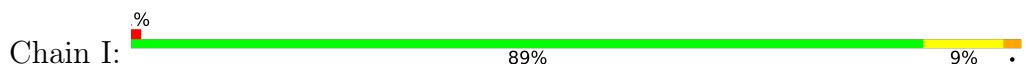
- Molecule 1: Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a



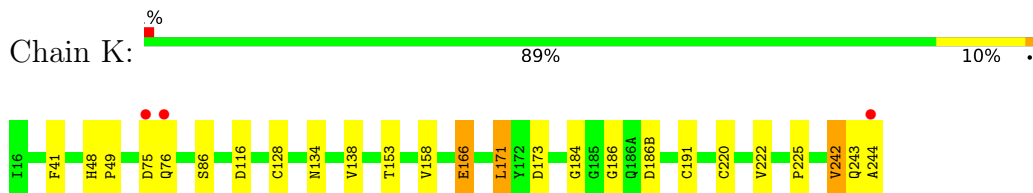
- Molecule 1: Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a



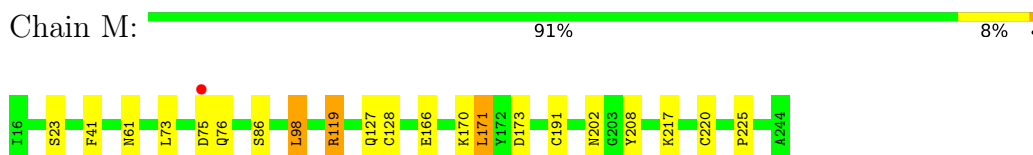
- Molecule 1: Kallikrein 4 (Protease, enamel matrix, prostate), isoform CRA_a



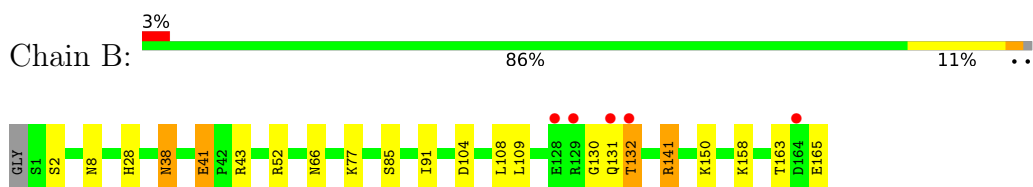
- Molecule 1: Kallikrein 4 (Prostase, enamel matrix, prostate), isoform CRA_a



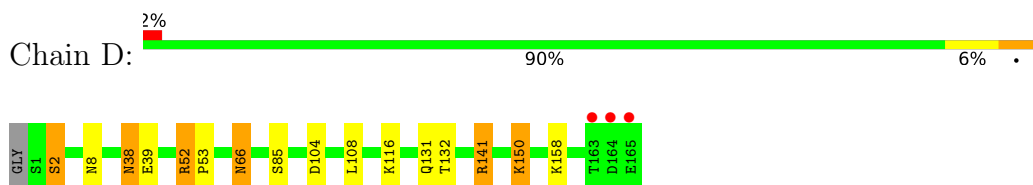
- Molecule 1: Kallikrein 4 (Prostase, enamel matrix, prostate), isoform CRA_a



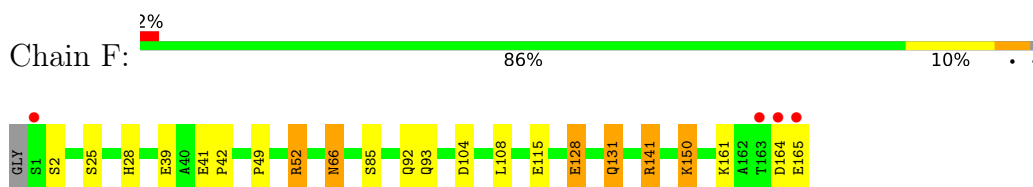
- Molecule 2: Kunitz-type inhibitor



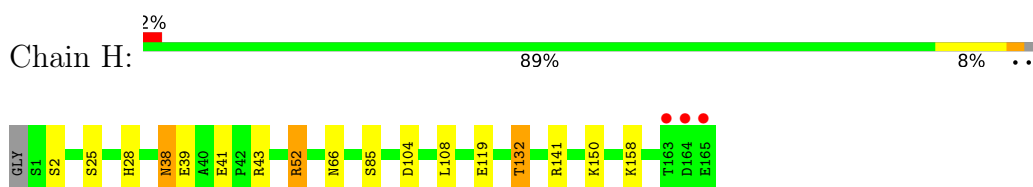
- Molecule 2: Kunitz-type inhibitor



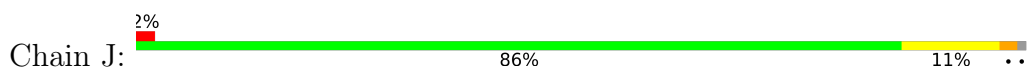
- Molecule 2: Kunitz-type inhibitor



- Molecule 2: Kunitz-type inhibitor

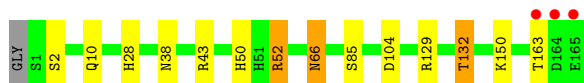
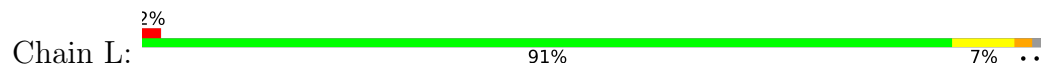


- Molecule 2: Kunitz-type inhibitor

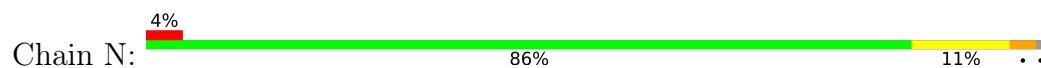




- Molecule 2: Kunitz-type inhibitor



- Molecule 2: Kunitz-type inhibitor



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	92.14Å 117.19Å 159.42Å 90.00° 89.57° 90.00°	Depositor
Resolution (Å)	49.49 – 2.10 49.45 – 2.10	Depositor EDS
% Data completeness (in resolution range)	86.4 (49.49-2.10) 85.9 (49.45-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.84 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.184 , 0.212 0.191 , 0.216	Depositor DCC
R_{free} test set	1652 reflections (0.97%)	wwPDB-VP
Wilson B-factor (Å ²)	27.2	Xtrriage
Anisotropy	0.064	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 41.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.022 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	22065	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.48% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.74	0/1733	0.85	2/2358 (0.1%)
1	C	0.68	0/1722	0.81	0/2344
1	E	0.68	0/1722	0.80	0/2344
1	G	0.69	0/1707	0.81	0/2323
1	I	0.69	0/1722	0.82	2/2344 (0.1%)
1	K	0.68	0/1723	0.78	0/2345
1	M	0.72	0/1707	0.87	3/2323 (0.1%)
2	B	0.76	0/1319	0.87	0/1791
2	D	0.70	0/1314	0.81	0/1785
2	F	0.77	0/1308	0.89	0/1777
2	H	0.73	0/1319	0.85	0/1791
2	J	0.71	0/1319	0.83	0/1791
2	L	0.71	0/1319	0.84	0/1791
2	N	0.74	0/1319	0.86	1/1791 (0.1%)
All	All	0.71	0/21253	0.83	8/28898 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	M	119	ARG	CG-CD-NE	7.95	128.50	111.80
1	M	208	TYR	CB-CG-CD1	7.00	125.20	121.00
1	A	119[A]	ARG	NE-CZ-NH1	-6.50	117.05	120.30
1	A	119[B]	ARG	NE-CZ-NH1	-6.50	117.05	120.30
1	I	119	ARG	NE-CZ-NH1	-5.59	117.51	120.30

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	1687	0	1633	33	1
1	C	1679	0	1620	20	0
1	E	1679	0	1620	20	0
1	G	1670	0	1600	12	0
1	I	1679	0	1620	18	0
1	K	1680	0	1615	22	0
1	M	1670	0	1600	24	0
2	B	1285	0	1288	37	3
2	D	1280	0	1280	18	0
2	F	1277	0	1275	27	0
2	H	1285	0	1288	13	0
2	J	1285	0	1288	23	0
2	L	1285	0	1288	14	0
2	N	1285	0	1288	31	4
3	A	153	0	0	18	2
3	B	109	0	0	20	2
3	C	121	0	0	8	0
3	D	74	0	0	10	0
3	E	91	0	0	10	0
3	F	100	0	0	8	0
3	G	59	0	0	8	0
3	H	68	0	0	5	0
3	I	131	0	0	5	0
3	J	70	0	0	10	0
3	K	84	0	0	13	0
3	L	80	0	0	8	0
3	M	115	0	0	9	0
3	N	84	0	0	11	0
All	All	22065	0	20303	273	6

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 7.

The worst 5 of 273 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:E:153:THR:HB	3:E:384:HOH:O	1.24	1.30
2:B:38:ASN:ND2	3:B:203:HOH:O	1.68	1.27
1:A:153:THR:HB	3:A:412:HOH:O	1.43	1.18
2:L:163:THR:HA	3:L:268:HOH:O	1.42	1.17
1:C:93:GLU:HG3	3:C:409:HOH:O	1.46	1.15

The worst 5 of 6 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)
2:B:165:GLU:OE1	2:N:52:ARG:CZ[1_655]	1.95	0.25
1:A:243:GLN:OE1	2:N:141:ARG:NH1[2_645]	1.99	0.21
3:A:301:HOH:O	3:B:203:HOH:O[2_745]	2.00	0.20
2:B:165:GLU:OE2	2:N:52:ARG:NE[1_655]	2.01	0.19
2:B:165:GLU:OE1	2:N:52:ARG:NH2[1_655]	2.01	0.19

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	224/223 (100%)	221 (99%)	3 (1%)	0	100	100
1	C	223/223 (100%)	220 (99%)	3 (1%)	0	100	100
1	E	223/223 (100%)	220 (99%)	3 (1%)	0	100	100
1	G	221/223 (99%)	218 (99%)	3 (1%)	0	100	100
1	I	223/223 (100%)	220 (99%)	3 (1%)	0	100	100
1	K	223/223 (100%)	220 (99%)	3 (1%)	0	100	100
1	M	221/223 (99%)	218 (99%)	3 (1%)	0	100	100
2	B	164/166 (99%)	156 (95%)	8 (5%)	0	100	100
2	D	164/166 (99%)	157 (96%)	7 (4%)	0	100	100
2	F	163/166 (98%)	157 (96%)	6 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	H	164/166 (99%)	157 (96%)	7 (4%)	0	100	100
2	J	164/166 (99%)	157 (96%)	7 (4%)	0	100	100
2	L	164/166 (99%)	157 (96%)	7 (4%)	0	100	100
2	N	164/166 (99%)	159 (97%)	5 (3%)	0	100	100
All	All	2705/2723 (99%)	2637 (98%)	68 (2%)	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	192/189 (102%)	190 (99%)	2 (1%)	76	82
1	C	191/189 (101%)	185 (97%)	6 (3%)	40	43
1	E	191/189 (101%)	186 (97%)	5 (3%)	46	50
1	G	189/189 (100%)	185 (98%)	4 (2%)	53	59
1	I	191/189 (101%)	183 (96%)	8 (4%)	30	30
1	K	191/189 (101%)	185 (97%)	6 (3%)	40	43
1	M	189/189 (100%)	184 (97%)	5 (3%)	46	50
2	B	140/139 (101%)	132 (94%)	8 (6%)	20	18
2	D	140/139 (101%)	130 (93%)	10 (7%)	14	11
2	F	139/139 (100%)	132 (95%)	7 (5%)	24	23
2	H	140/139 (101%)	132 (94%)	8 (6%)	20	18
2	J	140/139 (101%)	132 (94%)	8 (6%)	20	18
2	L	140/139 (101%)	134 (96%)	6 (4%)	29	29
2	N	140/139 (101%)	132 (94%)	8 (6%)	20	18
All	All	2313/2296 (101%)	2222 (96%)	91 (4%)	31	33

5 of 91 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	242	VAL
1	K	242	VAL
2	J	38	ASN
2	J	150	LYS
2	L	66	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
2	L	66	ASN
2	N	66	ASN
2	L	92	GLN
2	N	10	GLN
2	N	140	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	223/223 (100%)	-0.34	2 (0%) 84 86	10, 24, 57, 93	1 (0%)
1	C	223/223 (100%)	-0.45	1 (0%) 92 93	13, 29, 65, 106	1 (0%)
1	E	223/223 (100%)	-0.42	3 (1%) 77 80	16, 33, 67, 102	1 (0%)
1	G	223/223 (100%)	0.10	11 (4%) 29 35	24, 47, 84, 109	1 (0%)
1	I	223/223 (100%)	-0.51	2 (0%) 84 86	13, 24, 58, 95	1 (0%)
1	K	223/223 (100%)	-0.34	3 (1%) 77 80	17, 34, 68, 104	1 (0%)
1	M	223/223 (100%)	-0.34	1 (0%) 92 93	18, 32, 71, 100	1 (0%)
2	B	165/166 (99%)	-0.17	5 (3%) 50 56	13, 26, 58, 133	0
2	D	165/166 (99%)	-0.29	3 (1%) 68 72	16, 33, 71, 153	0
2	F	165/166 (99%)	-0.33	4 (2%) 59 64	14, 29, 65, 136	0
2	H	165/166 (99%)	-0.32	3 (1%) 68 72	19, 34, 66, 155	0
2	J	165/166 (99%)	-0.24	3 (1%) 68 72	19, 37, 74, 141	0
2	L	165/166 (99%)	-0.33	3 (1%) 68 72	12, 24, 60, 144	0
2	N	165/166 (99%)	0.02	7 (4%) 36 42	24, 37, 86, 140	0
All	All	2716/2723 (99%)	-0.29	51 (1%) 66 71	10, 31, 73, 155	7 (0%)

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	L	165	GLU	9.9
2	H	164	ASP	9.4
2	N	165	GLU	9.2
2	L	164	ASP	8.9
2	J	165	GLU	8.0

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.