



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 16, 2016 – 11:40 AM EDT

PDB ID : 5K98
Title : Structure of HipA-HipB-O2-O3 complex
Authors : Schumacher, M.
Deposited on : 2016-05-31
Resolution : 3.99 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
<http://wwpdb.org/validation/2016/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
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027790
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027790

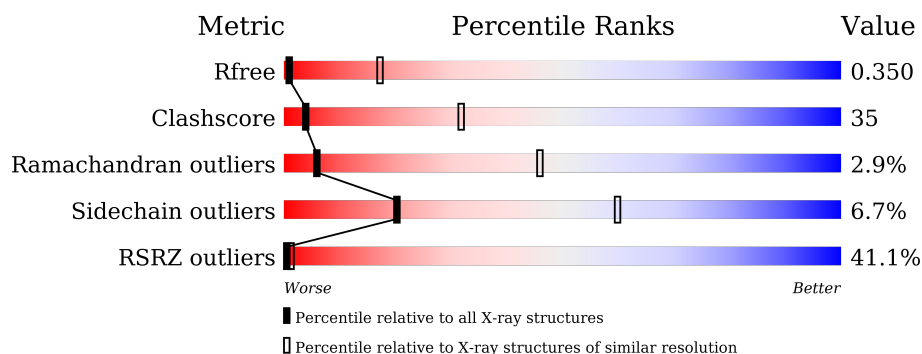
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.99 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}	91344	1010 (4.42-3.56)
Clashscore	102246	1052 (4.40-3.60)
Ramachandran outliers	100387	1005 (4.40-3.60)
Sidechain outliers	100360	1013 (4.42-3.58)
RSRZ outliers	91569	1013 (4.42-3.56)

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. 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	448	<div> <div>38%</div> <div>48%</div> <div>40%</div> <div>8%</div> </div>
1	D	448	<div> <div>39%</div> <div>49%</div> <div>38%</div> <div>8%</div> </div>
2	B	91	<div> <div>25%</div> <div>34%</div> <div>37%</div> <div>7%</div> <div>22%</div> </div>
2	P	91	<div> <div>25%</div> <div>36%</div> <div>33%</div> <div>9%</div> <div>22%</div> </div>
3	T	23	<div> <div>57%</div> <div>13%</div> <div>65%</div> <div>22%</div> </div>
4	E	23	<div> <div>52%</div> <div>22%</div> <div>35%</div> <div>43%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8605 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 Serine/threonine-protein kinase HipA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	414	Total	C	N	O	S	0	0	0
			3270	2098	570	590	12			
1	D	414	Total	C	N	O	S	0	0	0
			3270	2098	570	590	12			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	initiating methionine	UNP P23874
A	-6	HIS	-	expression tag	UNP P23874
A	-5	HIS	-	expression tag	UNP P23874
A	-4	HIS	-	expression tag	UNP P23874
A	-3	HIS	-	expression tag	UNP P23874
A	-2	HIS	-	expression tag	UNP P23874
A	-1	HIS	-	expression tag	UNP P23874
A	0	SER	-	expression tag	UNP P23874
A	1	ARG	-	expression tag	UNP P23874
A	309	GLN	ASP	engineered mutation	UNP P23874
D	-7	MET	-	initiating methionine	UNP P23874
D	-6	HIS	-	expression tag	UNP P23874
D	-5	HIS	-	expression tag	UNP P23874
D	-4	HIS	-	expression tag	UNP P23874
D	-3	HIS	-	expression tag	UNP P23874
D	-2	HIS	-	expression tag	UNP P23874
D	-1	HIS	-	expression tag	UNP P23874
D	0	SER	-	expression tag	UNP P23874
D	1	ARG	-	expression tag	UNP P23874
D	309	GLN	ASP	engineered mutation	UNP P23874

- Molecule 2 is a protein called Antitoxin HipB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	71	Total	C	N	O	S	0	0	0
			564	358	95	108	3			
2	P	71	Total	C	N	O	S	0	0	0
			564	358	95	108	3			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP M9IJX7
B	-1	SER	-	expression tag	UNP M9IJX7
B	0	HIS	-	expression tag	UNP M9IJX7
P	-2	GLY	-	expression tag	UNP M9IJX7
P	-1	SER	-	expression tag	UNP M9IJX7
P	0	HIS	-	expression tag	UNP M9IJX7

- Molecule 3 is a DNA chain called DNA (5'-D(*TP*CP*CP*CP*TP*AP*TP*CP*CP*CP*CP*TP*TP*AP*AP*GP*GP*GP*GP*AP*TP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	23	Total	C	N	O	P	0	0	0
			465	223	83	137	22			

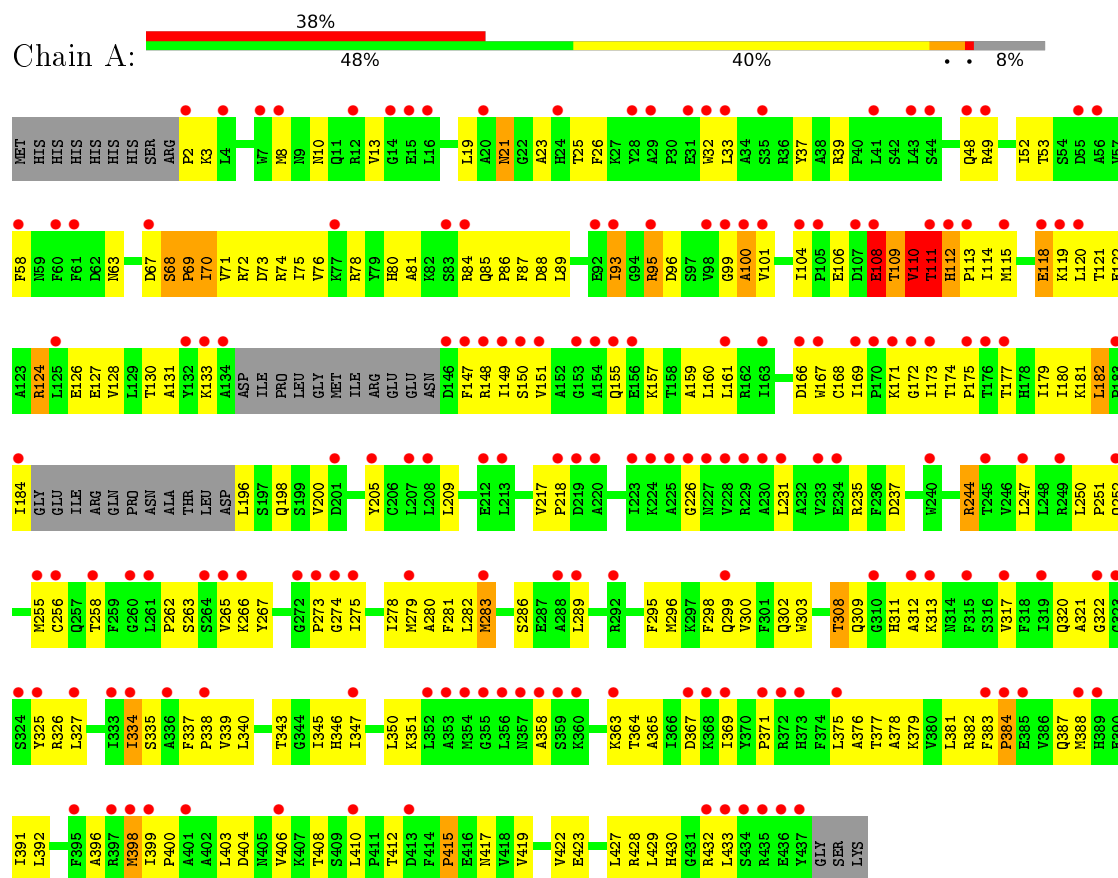
- Molecule 4 is a DNA chain called DNA (5'-D(*CP*TP*AP*TP*CP*CP*CP*CP*TP*TP*AP*AP*GP*GP*GP*GP*AP*TP*AP*GP*GP*GP*A)-3').

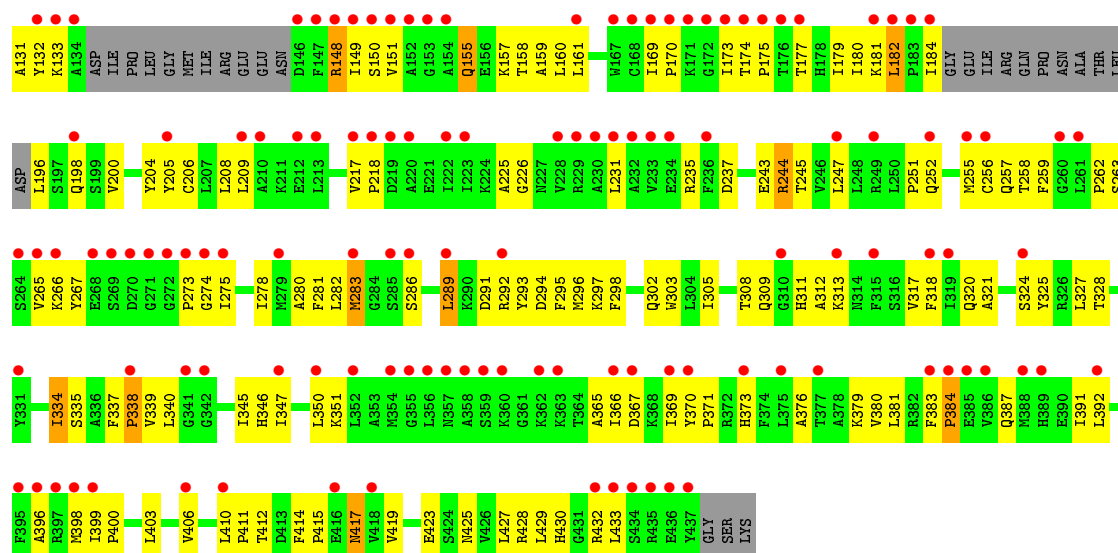
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	23	Total	C	N	O	P	0	0	0
			472	225	90	135	22			

3 Residue-property plots

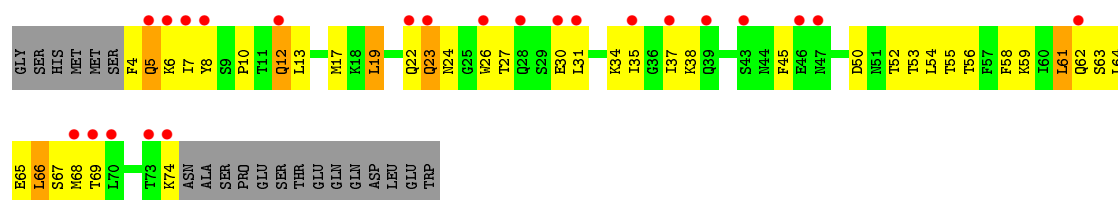
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Serine/threonine-protein kinase HipA

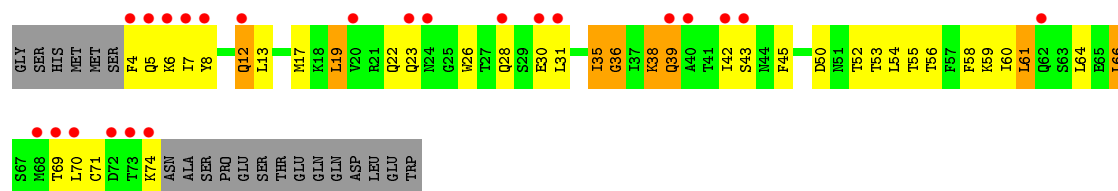




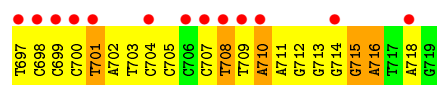
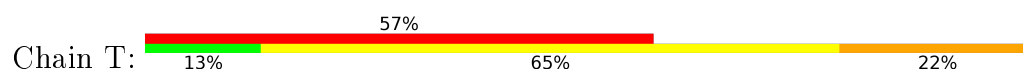
● Molecule 2: Antitoxin HipB



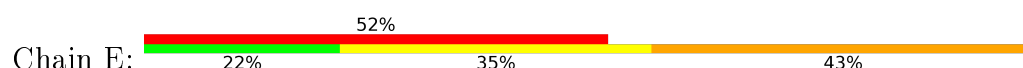
● Molecule 2: Antitoxin HipB



● Molecule 3: DNA (5'-D(*TP*CP*CP*CP*TP*AP*TP*CP*CP*CP*CP*TP*TP*AP*AP*GP*GP*GP*GP*AP*TP*AP*G)-3')



● Molecule 4: DNA (5'-D(*CP*TP*AP*TP*CP*CP*CP*CP*TP*TP*AP*AP*GP*GP*GP*GP*AP*TP*AP*GP*GP*GP*A)-3')



C700	T701	A702	T703	C704	C705	C706	C707	T708	T709	A710	A711	G712	G713	G714	G715	A716	T717	A718	G719	G720	G721	A722

4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	214.04Å 146.83Å 53.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	121.08 – 3.99 121.08 – 3.99	Depositor EDS
% Data completeness (in resolution range)	92.2 (121.08-3.99) 92.4 (121.08-3.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.97 (at 4.01Å)	Xtriage
Refinement program	PHENIX 1.6.4 _486	Depositor
R, R_{free}	0.350 , 0.375 0.323 , 0.350	Depositor DCC
R_{free} test set	1398 reflections (10.02%)	DCC
Wilson B-factor (Å ²)	120.1	Xtriage
Anisotropy	0.930	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 108.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.36$, $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.82	EDS
Total number of atoms	8605	wwPDB-VP
Average B, all atoms (Å ²)	179.0	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 24.73 % 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 3.6243e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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 ⓘ

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	0.91	8/3343 (0.2%)	0.73	7/4530 (0.2%)
1	D	0.67	9/3343 (0.3%)	0.75	11/4530 (0.2%)
2	B	0.54	0/572	0.65	0/771
2	P	0.48	0/572	0.65	1/771 (0.1%)
3	T	1.02	1/520 (0.2%)	1.56	8/800 (1.0%)
4	E	1.00	3/530 (0.6%)	1.98	21/817 (2.6%)
All	All	0.79	21/8880 (0.2%)	0.94	48/12219 (0.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	110	VAL	CA-CB	30.12	2.18	1.54
1	A	110	VAL	CB-CG1	28.62	2.12	1.52
1	D	109	THR	CA-CB	15.28	1.93	1.53
1	D	107	ASP	CB-CG	11.98	1.76	1.51
1	D	109	THR	CA-C	10.09	1.79	1.52

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	713	DG	O4'-C1'-N9	23.20	124.24	108.00
4	E	715	DG	O4'-C1'-N9	16.66	119.66	108.00
1	A	111	THR	N-CA-CB	15.95	140.61	110.30
3	T	708	DT	O4'-C1'-N1	11.79	116.25	108.00
1	D	109	THR	CA-CB-CG2	10.08	126.51	112.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	108	GLU	Peptide

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	3270	0	3313	236	10
1	D	3270	0	3313	214	10
2	B	564	0	574	52	0
2	P	564	0	574	48	0
3	T	465	0	261	37	0
4	E	472	0	260	55	0
All	All	8605	0	8295	582	10

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 35.

The worst 5 of 582 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:D:109:THR:C	1:D:109:THR:CA	1.79	1.51
1:D:107:ASP:CB	1:D:107:ASP:CG	1.76	1.50
1:D:109:THR:CA	1:D:109:THR:CB	1.93	1.42
1:A:110:VAL:CB	1:A:110:VAL:CG1	2.12	1.27
1:A:68:SER:OG	1:A:70:ILE:HD12	1.31	1.25

The worst 5 of 10 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:110:VAL:CB	1:D:109:THR:C[3_456]	1.68	0.52
1:A:110:VAL:CA	1:D:110:VAL:N[3_456]	1.94	0.26
1:A:110:VAL:CG1	1:D:109:THR:O[3_456]	1.95	0.25

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:110:VAL:O	1:D:110:VAL:N[3_456]	2.00	0.20
1:A:110:VAL:CB	1:D:110:VAL:N[3_456]	2.09	0.11

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	408/448 (91%)	353 (86%)	45 (11%)	10 (2%)	7 49
1	D	408/448 (91%)	353 (86%)	41 (10%)	14 (3%)	5 43
2	B	69/91 (76%)	63 (91%)	5 (7%)	1 (1%)	14 59
2	P	69/91 (76%)	63 (91%)	3 (4%)	3 (4%)	3 35
All	All	954/1078 (88%)	832 (87%)	94 (10%)	28 (3%)	6 46

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	95	ARG
1	A	108	GLU
1	A	110	VAL
1	D	20	ALA
1	D	21	ASN

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	351/382 (92%)	328 (93%)	23 (7%)	21	60
1	D	351/382 (92%)	332 (95%)	19 (5%)	27	67
2	B	64/83 (77%)	57 (89%)	7 (11%)	8	38
2	P	64/83 (77%)	57 (89%)	7 (11%)	8	38
All	All	830/930 (89%)	774 (93%)	56 (7%)	20	60

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	12	GLN
1	D	21	ASN
2	P	13	LEU
2	B	13	LEU
2	B	61	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	39	GLN
1	D	21	ASN
2	P	22	GLN
1	D	10	ASN
1	D	85	GLN

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 carbohydrates 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	414/448 (92%)	2.13	171 (41%) 0 1	155, 179, 191, 199	0
1	D	414/448 (92%)	2.09	176 (42%) 0 1	157, 179, 191, 206	0
2	B	71/91 (78%)	1.57	23 (32%) 1 1	162, 176, 183, 185	0
2	P	71/91 (78%)	1.69	23 (32%) 1 1	158, 175, 187, 190	0
3	T	23/23 (100%)	2.31	13 (56%) 0 1	168, 186, 195, 198	0
4	E	23/23 (100%)	2.93	12 (52%) 0 1	166, 191, 196, 204	0
All	All	1016/1124 (90%)	2.06	418 (41%) 0 1	155, 179, 191, 206	0

The worst 5 of 418 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	99	GLY	20.9
1	D	273	PRO	13.1
1	A	229	ARG	11.5
1	D	230	ALA	11.2
1	A	100	ALA	11.1

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 carbohydrates 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.