wwPDB X-ray Structure Validation Summary Report (i)

Dec 20, 2023 – 12:09 pm GMT

PDB ID	:	8BYX
Title	:	HOXB13-homodimer bound to DNA
Authors	:	Morgunova, E.; Popov, A.; Yin, Y.; Taipale, J.
Deposited on	:	2022-12-14
Resolution	:	3.00 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:

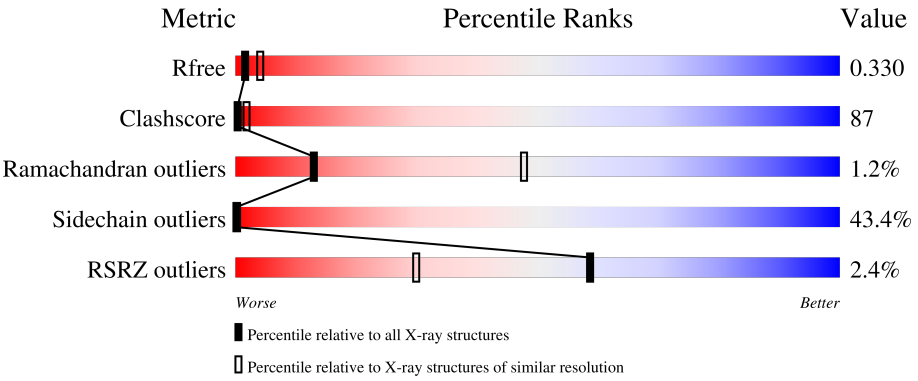
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.00 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.



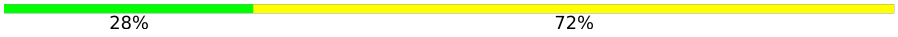
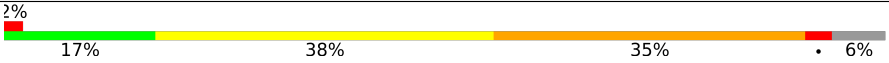
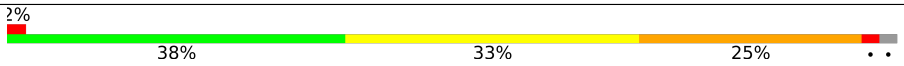
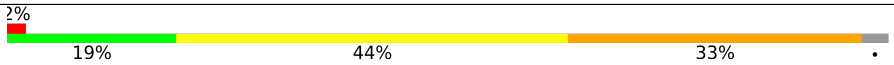
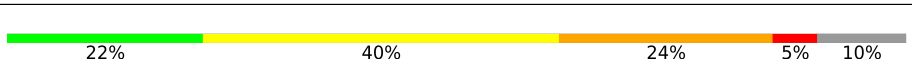
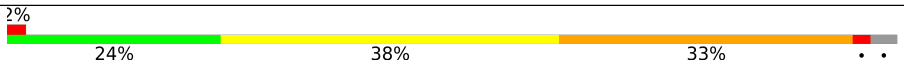
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	C	18	39% 61%
1	D	18	44% 50% 6%
1	H	18	33% 67%
2	E	18	28% 67% 6%
2	F	18	39% 61%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	18	
3	A	63	
3	B	63	
3	G	63	
3	J	63	
3	K	63	
3	L	63	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5225 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 DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	17	Total	C	N	O	P	0	0	0
			347	168	51	111	17			
1	C	18	Total	C	N	O	P	0	0	0
			369	178	56	117	18			
1	H	18	Total	C	N	O	P	0	0	0
			369	178	56	117	18			

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	18	Total	C	N	O	P	0	0	0
			371	176	76	101	18			
2	F	18	Total	C	N	O	P	0	0	0
			371	176	76	101	18			
2	I	18	Total	C	N	O	P	0	0	0
			371	176	76	101	18			

- Molecule 3 is a protein called Homeobox protein Hox-B13.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	B	62	Total	C	N	O	0	0	0
			525	331	106	88			
3	A	59	Total	C	N	O	0	0	0
			501	317	99	85			
3	G	61	Total	C	N	O	0	0	0
			521	329	105	87			
3	J	57	Total	C	N	O	0	0	0
			481	305	93	83			
3	K	61	Total	C	N	O	0	0	0
			521	329	105	87			
3	L	54	Total	C	N	O	0	0	0
			459	290	89	80			

- Molecule 4 is water.

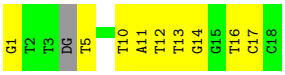
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	4	Total O 4 4	0	0
4	E	1	Total O 1 1	0	0
4	B	2	Total O 2 2	0	0
4	C	2	Total O 2 2	0	0
4	F	3	Total O 3 3	0	0
4	A	1	Total O 1 1	0	0
4	H	2	Total O 2 2	0	0
4	I	2	Total O 2 2	0	0
4	G	2	Total O 2 2	0	0

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.

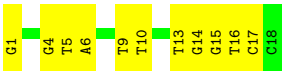
- Molecule 1: DNA

Chain D:



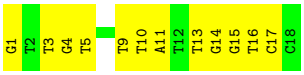
- Molecule 1: DNA

Chain C:



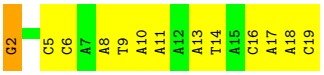
- Molecule 1: DNA

Chain H:



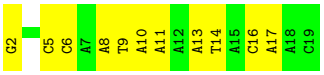
- Molecule 2: DNA

Chain E:



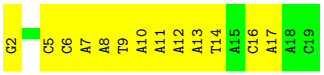
- Molecule 2: DNA

Chain F:



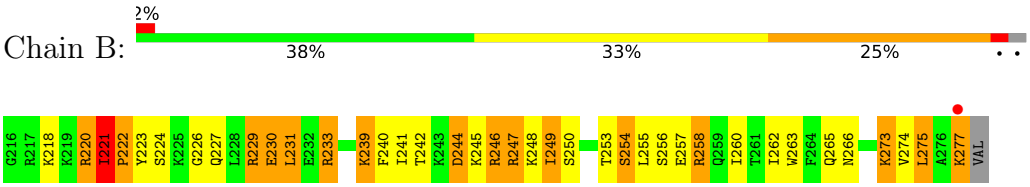
- Molecule 2: DNA

Chain I:



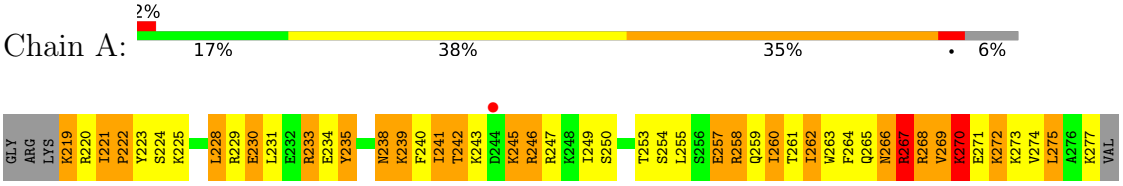
• Molecule 3: Homeobox protein Hox-B13

Chain B:



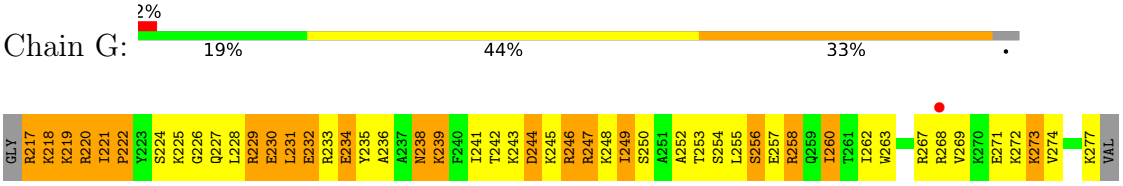
• Molecule 3: Homeobox protein Hox-B13

Chain A:



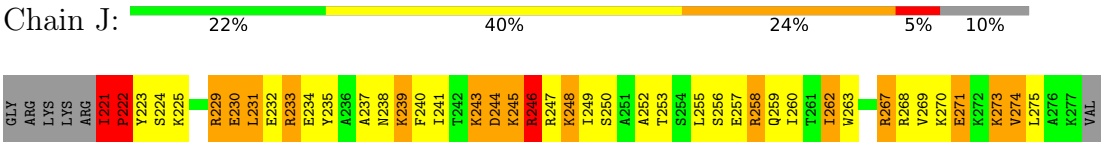
• Molecule 3: Homeobox protein Hox-B13

Chain G:



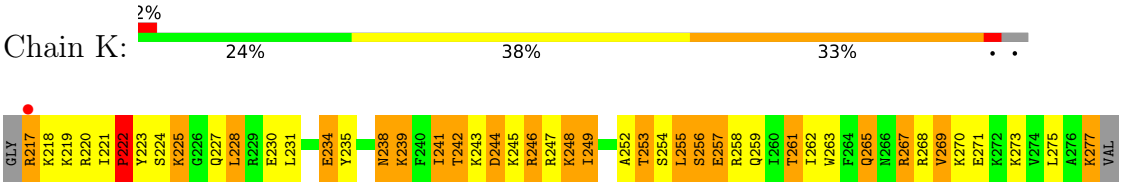
• Molecule 3: Homeobox protein Hox-B13

Chain J:



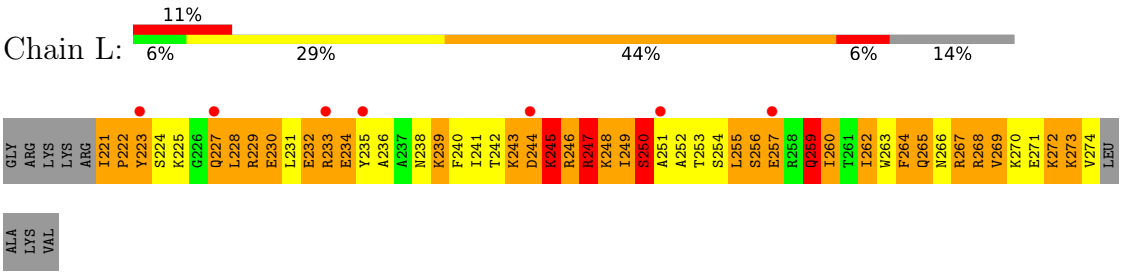
• Molecule 3: Homeobox protein Hox-B13

Chain K:



• Molecule 3: Homeobox protein Hox-B13

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	116.78Å 116.78Å 118.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 58.39 – 2.91	Depositor EDS
% Data completeness (in resolution range)	99.8 (30.00-3.00) 99.5 (58.39-2.91)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.41 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5	Depositor
R, R_{free}	0.254 , 0.304 0.282 , 0.330	Depositor DCC
R_{free} test set	941 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å^2)	72.3	Xtriage
Anisotropy	0.625	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.20 , 59.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.039 for -h,l,k 0.010 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5225	wwPDB-VP
Average B, all atoms (Å^2)	119.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.97% 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

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	C	0.72	1/410 (0.2%)	0.87	0/630
1	D	0.76	1/384 (0.3%)	0.89	0/587
1	H	0.74	1/410 (0.2%)	0.90	0/630
2	E	0.79	1/418 (0.2%)	0.94	0/640
2	F	0.68	1/418 (0.2%)	0.86	0/640
2	I	0.68	1/418 (0.2%)	0.87	0/640
3	A	0.84	2/507 (0.4%)	1.12	5/672 (0.7%)
3	B	0.85	2/531 (0.4%)	0.87	1/702 (0.1%)
3	G	0.84	2/527 (0.4%)	0.88	1/697 (0.1%)
3	J	0.86	2/487 (0.4%)	1.01	1/647 (0.2%)
3	K	0.85	2/527 (0.4%)	0.90	1/697 (0.1%)
3	L	0.92	2/465 (0.4%)	1.26	7/618 (1.1%)
All	All	0.80	18/5502 (0.3%)	0.95	16/7800 (0.2%)

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
3	A	0	1
3	J	0	1
3	K	0	1
3	L	0	3
All	All	0	6

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	L	222	PRO	N-CA	13.60	1.70	1.47
3	A	222	PRO	N-CA	13.30	1.69	1.47
3	G	222	PRO	N-CA	13.26	1.69	1.47
3	K	222	PRO	N-CA	13.13	1.69	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	222	PRO	N-CA	13.06	1.69	1.47

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	J	222	PRO	CA-N-CD	-10.85	96.32	111.50
3	A	267	ARG	CB-CA-C	-9.26	91.88	110.40
3	L	268	ARG	CG-CD-NE	7.27	127.06	111.80
3	A	222	PRO	CA-N-CD	-7.13	101.52	111.50
3	L	222	PRO	CA-N-CD	-6.90	101.84	111.50

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	266	ASN	Mainchain
3	J	246	ARG	Mainchain
3	K	253	THR	Mainchain
3	L	245	LYS	Mainchain
3	L	250	SER	Mainchain

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	C	369	0	209	30	0
1	D	347	0	199	31	0
1	H	369	0	209	17	0
2	E	371	0	201	19	4
2	F	371	0	201	27	0
2	I	371	0	201	23	0
3	A	501	0	541	139	0
3	B	525	0	571	89	0
3	G	521	0	567	118	0
3	J	481	0	515	100	0
3	K	521	0	566	127	4
3	L	459	0	484	208	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	1	0
4	B	2	0	0	0	0
4	C	2	0	0	1	0
4	D	4	0	0	0	0
4	E	1	0	0	0	0
4	F	3	0	0	0	0
4	G	2	0	0	0	0
4	H	2	0	0	0	0
4	I	2	0	0	1	0
All	All	5225	0	4464	842	4

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 87.

The worst 5 of 842 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)
3:L:223:TYR:CE2	3:L:227:GLN:HB3	1.28	1.64
3:K:242:THR:HG23	3:K:245:LYS:CD	1.20	1.60
3:K:241:ILE:HD12	3:K:245:LYS:CD	1.28	1.59
3:L:223:TYR:CZ	3:L:227:GLN:HB3	1.31	1.58
3:G:228:LEU:HA	3:G:231:LEU:CD1	1.12	1.55

All (4) 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:E:19:DC:OP2	3:K:243:LYS:CE[4_665]	1.68	0.52
2:E:2:DG:C4'	3:K:243:LYS:CD[7_645]	1.81	0.39
2:E:2:DG:O3'	3:K:243:LYS:CD[7_645]	1.92	0.28
2:E:2:DG:C3'	3:K:243:LYS:CD[7_645]	2.12	0.08

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
3	A	57/63 (90%)	52 (91%)	4 (7%)	1 (2%)	8 37
3	B	60/63 (95%)	58 (97%)	2 (3%)	0	100 100
3	G	59/63 (94%)	53 (90%)	5 (8%)	1 (2%)	9 39
3	J	55/63 (87%)	49 (89%)	5 (9%)	1 (2%)	8 37
3	K	59/63 (94%)	55 (93%)	4 (7%)	0	100 100
3	L	52/63 (82%)	43 (83%)	8 (15%)	1 (2%)	8 36
All	All	342/378 (90%)	310 (91%)	28 (8%)	4 (1%)	13 48

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	G	231	LEU
3	L	233	ARG
3	A	270	LYS
3	J	222	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
3	A	53/56 (95%)	31 (58%)	22 (42%)	0 0
3	B	55/56 (98%)	36 (66%)	19 (34%)	0 1
3	G	55/56 (98%)	34 (62%)	21 (38%)	0 0
3	J	51/56 (91%)	30 (59%)	21 (41%)	0 0
3	K	55/56 (98%)	29 (53%)	26 (47%)	0 0
3	L	49/56 (88%)	20 (41%)	29 (59%)	0 0
All	All	318/336 (95%)	180 (57%)	138 (43%)	0 0

5 of 138 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	L	234	GLU
3	L	244	ASP
3	L	262	ILE
3	G	232	GLU
3	G	230	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	K	265	GLN
3	L	227	GLN
3	L	266	ASN
3	L	238	ASN
3	K	259	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 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	C	18/18 (100%)	-0.77	0 100 100	76, 96, 120, 145	0
1	D	17/18 (94%)	-0.81	0 100 100	49, 76, 104, 124	0
1	H	18/18 (100%)	-0.81	0 100 100	81, 109, 148, 167	0
2	E	18/18 (100%)	-0.89	0 100 100	51, 78, 99, 116	0
2	F	18/18 (100%)	-0.72	0 100 100	89, 101, 126, 152	0
2	I	18/18 (100%)	-0.77	0 100 100	90, 104, 144, 149	0
3	A	59/63 (93%)	-0.05	1 (1%) 70 41	89, 140, 171, 200	0
3	B	62/63 (98%)	-0.40	1 (1%) 72 44	59, 92, 132, 159	0
3	G	61/63 (96%)	0.07	1 (1%) 72 44	82, 147, 184, 194	0
3	J	57/63 (90%)	-0.10	0 100 100	83, 132, 175, 196	0
3	K	61/63 (96%)	0.05	1 (1%) 72 44	94, 139, 184, 221	0
3	L	54/63 (85%)	0.55	7 (12%) 3 1	115, 178, 233, 251	0
All	All	461/486 (94%)	-0.18	11 (2%) 59 30	49, 124, 193, 251	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	223	TYR	6.1
3	B	277	LYS	4.9
3	L	244	ASP	4.6
3	L	251	ALA	3.5
3	L	235	TYR	3.4

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 (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.