WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:06 PM GMT

PDB ID	:	4H63
Title	:	Structure of the Schizosaccharomyces pombe Mediator head module
Authors	:	Lariviere, L.; Plaschka, C.; Seizl, M.; Wenzeck, L.; Kurth, F.; Cramer, P.
Deposited on	:	2012-09-19
Resolution	:	3.40 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

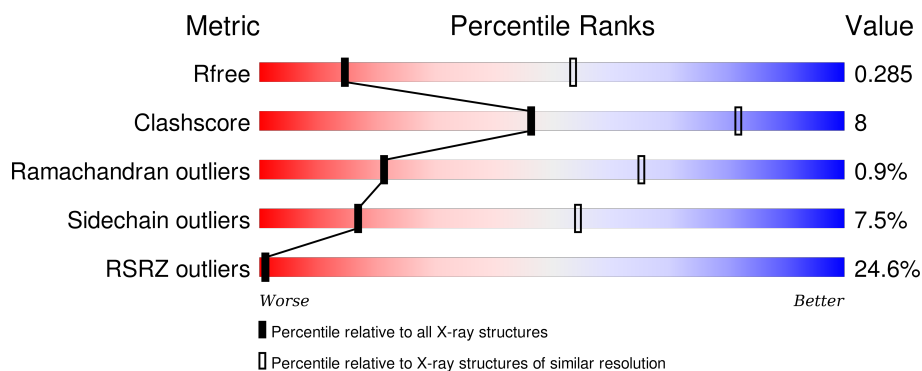|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.40 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	Percentile Ranks	Value
Rfree		0.285
Clashscore		8
Ramachandran outliers		0.9%
Sidechain outliers		7.5%
RSRZ outliers		24.6%

Worse ◼ *Better*

◼ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	1476 (3.50-3.30)
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

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 <=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	F	183	16% / 57% 16% • 26%
2	H	200	19% / 69% 16% • 13%
3	K	112	27% / 64% 21% • 13%
4	Q	469	14% / 61% 23% • 13%
5	R	209	45% / 78% 17% • •

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	V	135	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 9254 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 Mediator of RNA polymerase II transcription subunit 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	136	Total	C	N	O	S	0	0	0
			1138	736	190	205	7			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-2	GLY	-	EXPRESSION TAG	UNP Q9US45
F	-1	SER	-	EXPRESSION TAG	UNP Q9US45
F	0	HIS	-	EXPRESSION TAG	UNP Q9US45

- Molecule 2 is a protein called Mediator of RNA polymerase II transcription subunit 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	174	Total	C	N	O	S	0	0	0
			1429	902	245	279	3			

- Molecule 3 is a protein called Mediator of RNA polymerase II transcription subunit 11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	K	98	Total	C	N	O	S	0	0	0
			777	490	129	156	2			

- Molecule 4 is a protein called Mediator of RNA polymerase II transcription subunit 17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	Q	410	Total	C	N	O	S	0	0	0
			3316	2111	554	633	18			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Q	77	GLY	-	EXPRESSION TAG	UNP P87306

- Molecule 5 is a protein called Mediator of RNA polymerase II transcription subunit 18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	R	201	1649	1053	281	307	8	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
R	-1	ALA	-	EXPRESSION TAG	UNP O14198
R	0	SER	-	EXPRESSION TAG	UNP O14198

- Molecule 6 is a protein called Mediator of RNA polymerase II transcription subunit 22.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	V	117	945	595	160	188	2	0	0	0

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.
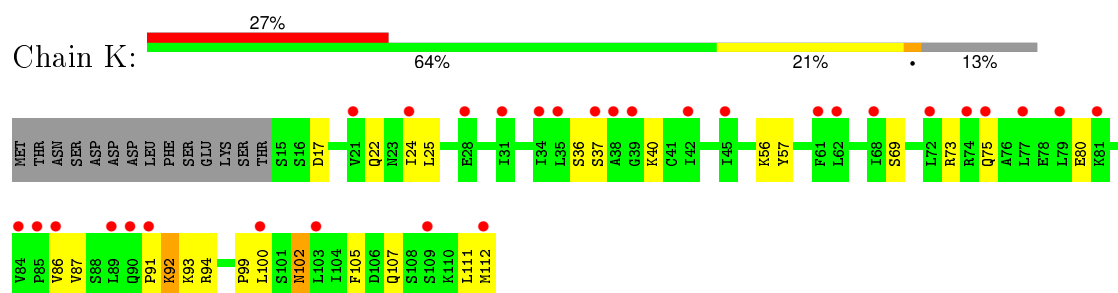
• Molecule 1: Mediator of RNA polymerase II transcription subunit 6
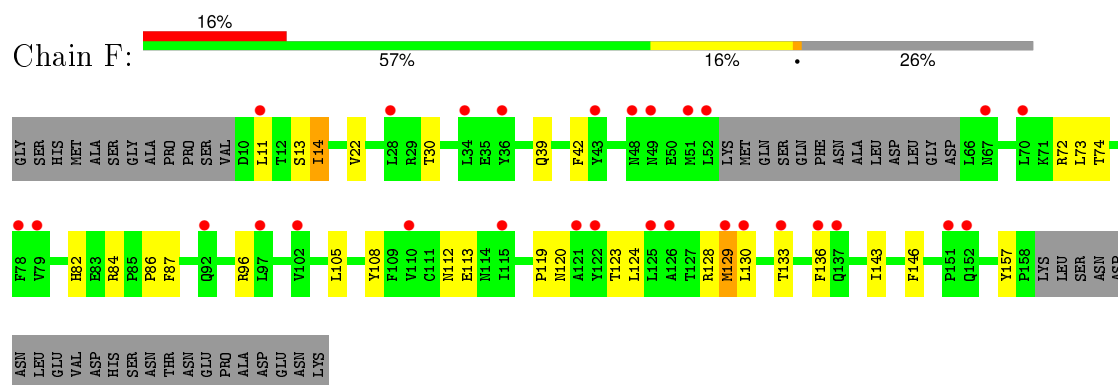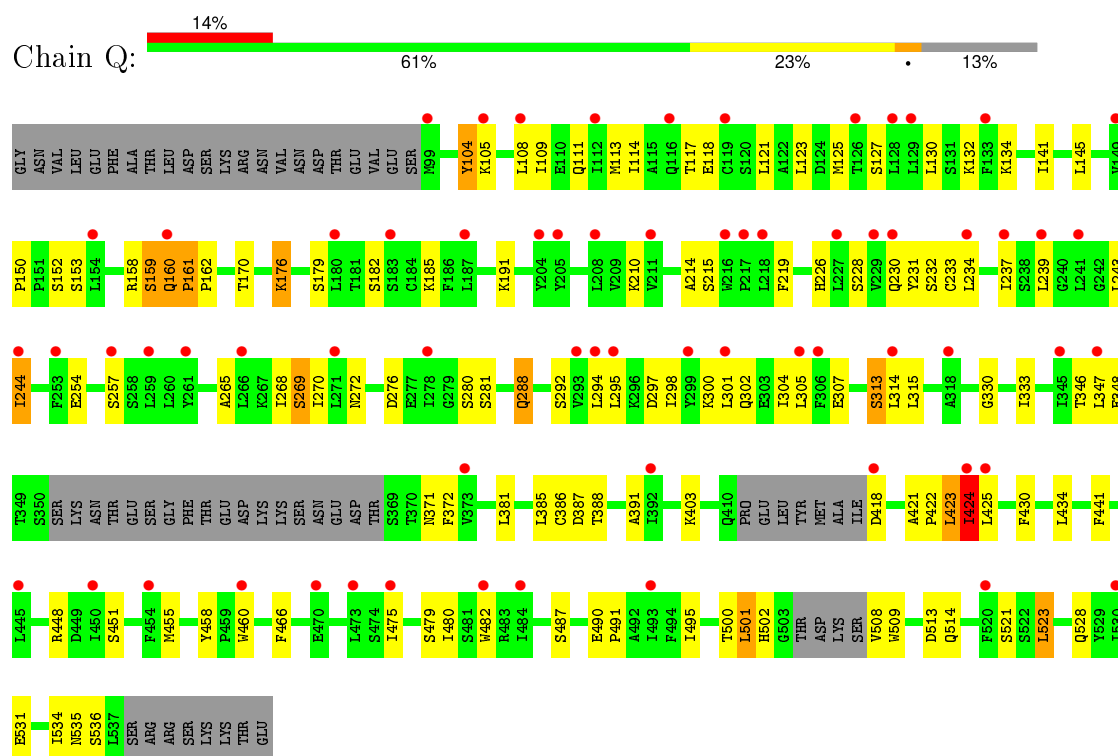

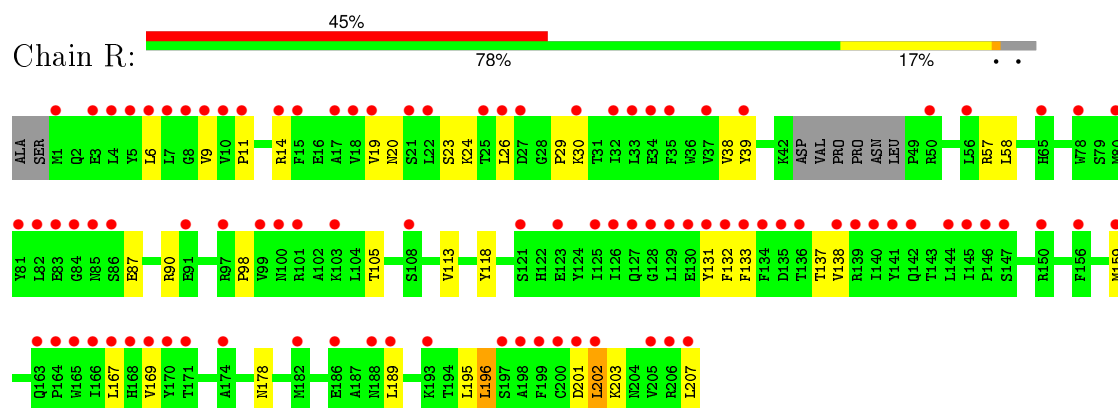• Molecule 2: Mediator of RNA polymerase II transcription subunit 8
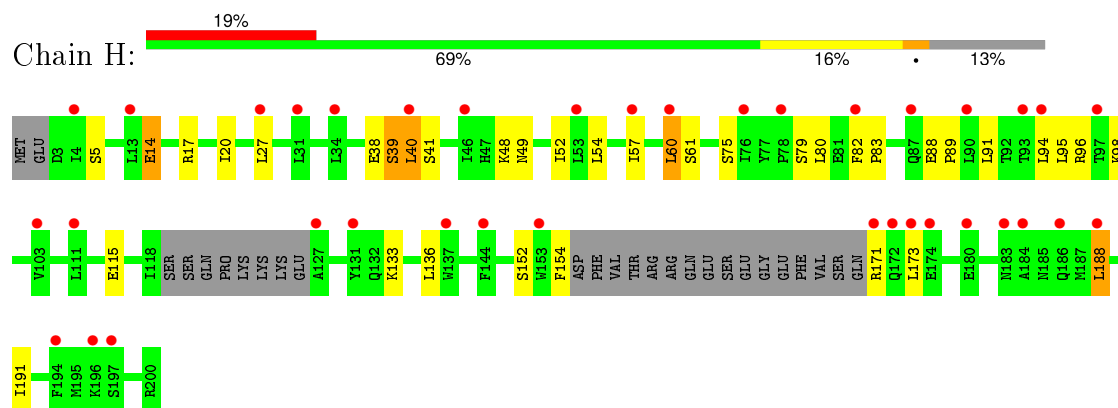
• Molecule 3: Mediator of RNA polymerase II transcription subunit 11
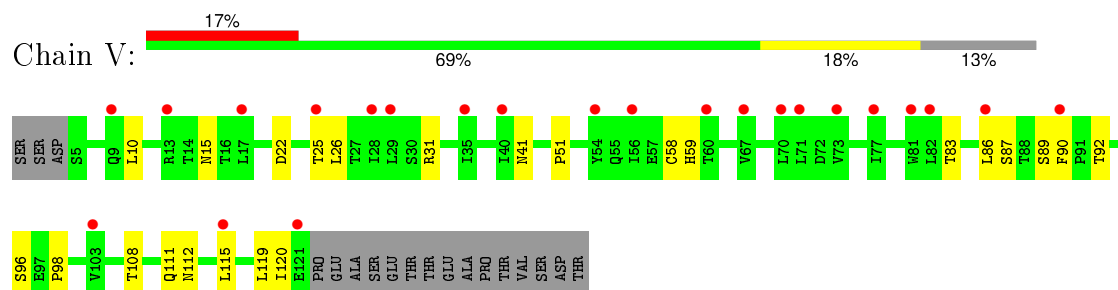

• Molecule 4: Mediator of RNA polymerase II transcription subunit 17

Chain Q:

• Molecule 5: Mediator of RNA polymerase II transcription subunit 18

Chain R:

• Molecule 6: Mediator of RNA polymerase II transcription subunit 22

Chain V:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	145.60Å　　145.60Å　　241.62Å 90.00°　　　90.00°　　　120.00°	Depositor
Resolution (Å)	126.09　–　3.40 126.09　–　3.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (126.09-3.40) 100.0 (126.09-3.40)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$<I/\sigma(I)>$ [1]	1.47 (at 3.41Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.232　,　0.258 0.250　,　0.285	Depositor DCC
R_{free} test set	2101 reflections (5.07%)	DCC
Wilson B-factor (Å^2)	132.8	Xtriage
Anisotropy	0.145	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 150.7	EDS
Estimated twinning fraction	0.038 for -h,-k,l	Xtriage
L-test for twinning[2]	$<|L|> = 0.46$, $<L^2> = 0.29$	Xtriage
Outliers	0 of 41454 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9254	wwPDB-VP
Average B, all atoms (Å^2)	165.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.84% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	F	0.47	0/1169	0.67	0/1587
2	H	0.48	0/1457	0.65	0/1975
3	K	0.48	0/786	0.66	0/1057
4	Q	0.48	0/3384	0.71	1/4570 (0.0%)
5	R	0.47	0/1691	0.61	0/2291
6	V	0.49	0/958	0.73	0/1301
All	All	0.48	0/9445	0.68	1/12781 (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})
4	Q	159	SER	C-N-CA	5.46	135.34	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	F	1138	0	1118	23	0
2	H	1429	0	1396	28	0
3	K	777	0	796	27	0
4	Q	3316	0	3302	78	0
5	R	1649	0	1626	16	0
6	V	945	0	956	22	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	9254	0	9194	140	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 8.

The worst 5 of 140 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)
4:Q:214:ALA:O	6:V:90:PHE:HB3	1.60	1.00
3:K:40:LYS:HB3	3:K:57:TYR:HD2	1.31	0.96
3:K:94:ARG:HH12	6:V:89:SER:HB2	1.39	0.85
3:K:40:LYS:HB3	3:K:57:TYR:CD2	2.18	0.77
4:Q:234:LEU:HD12	4:Q:302:GLN:OE1	1.84	0.77

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	F	132/183 (72%)	125 (95%)	7 (5%)	0	100	100
2	H	168/200 (84%)	159 (95%)	7 (4%)	2 (1%)	16	59
3	K	96/112 (86%)	96 (100%)	0	0	100	100
4	Q	402/469 (86%)	381 (95%)	14 (4%)	7 (2%)	11	51
5	R	197/209 (94%)	189 (96%)	8 (4%)	0	100	100
6	V	115/135 (85%)	109 (95%)	5 (4%)	1 (1%)	21	65
All	All	1110/1308 (85%)	1059 (95%)	41 (4%)	10 (1%)	21	65

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	Q	424	ILE
2	H	39	SER
4	Q	160	GLN
4	Q	535	ASN
6	V	96	SER

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	F	127/167 (76%)	119 (94%)	8 (6%)	22	62
2	H	160/185 (86%)	147 (92%)	13 (8%)	15	51
3	K	90/104 (86%)	86 (96%)	4 (4%)	35	73
4	Q	384/439 (88%)	348 (91%)	36 (9%)	11	42
5	R	185/192 (96%)	172 (93%)	13 (7%)	19	58
6	V	112/128 (88%)	107 (96%)	5 (4%)	34	73
All	All	1058/1215 (87%)	979 (92%)	79 (8%)	17	54

5 of 79 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	Q	268	ILE
4	Q	381	LEU
5	R	207	LEU
4	Q	269	SER
4	Q	346	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
3	K	102	ASN
4	Q	272	ASN
5	R	204	ASN
6	V	59	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 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	F	136/183 (74%)	1.19	29 (21%)	1	1	99, 149, 196, 212	0
2	H	174/200 (87%)	1.13	37 (21%)	1	1	107, 145, 256, 264	0
3	K	98/112 (87%)	1.32	30 (30%)	1	1	118, 146, 184, 205	0
4	Q	410/469 (87%)	0.98	66 (16%)	3	2	101, 140, 192, 231	0
5	R	201/209 (96%)	2.35	95 (47%)	0	0	204, 246, 260, 266	0
6	V	117/135 (86%)	1.04	23 (19%)	1	2	108, 134, 193, 217	0
All	All	1136/1308 (86%)	1.30	280 (24%)	1	1	99, 149, 254, 266	0

The worst 5 of 280 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	R	199	PHE	11.4
5	R	133	PHE	9.2
5	R	144	LEU	8.5
5	R	202	LEU	8.3
5	R	6	LEU	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 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.