Full wwPDB X-ray Structure Validation Report ⓘ

Oct 17, 2024 – 06:24 PM EDT

PDB ID	:	9DZM
Title	:	Dimeric human OCT2 (POU2F2) POU domain bound to palindromic MORE DNA
Authors	:	Terrell, J.R.; Poon, G.M.K.
Deposited on	:	2024-10-16
Resolution	:	2.54 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 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:

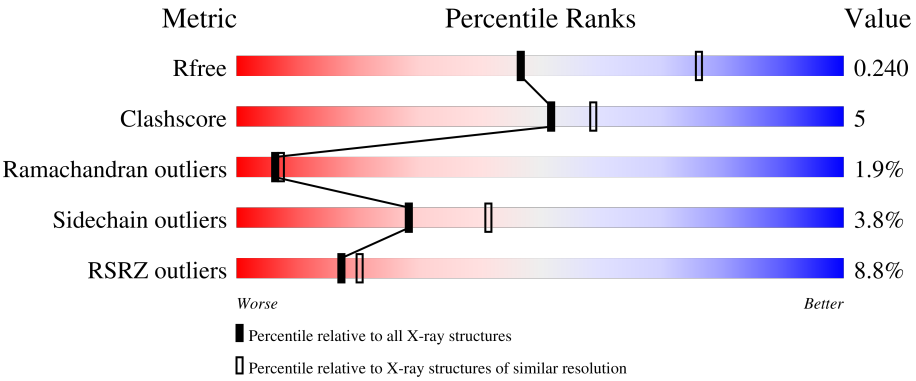
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.54 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.



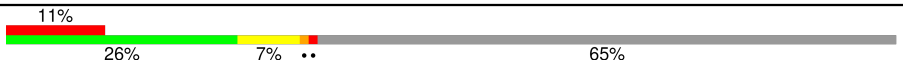
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1004 (2.54-2.54)
Clashscore	180529	1055 (2.54-2.54)
Ramachandran outliers	177936	1048 (2.54-2.54)
Sidechain outliers	177891	1048 (2.54-2.54)
RSRZ outliers	164620	1004 (2.54-2.54)

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	A	21	5% 81% 19%
2	B	22	5% 77% 18% 5%
3	C	167	3% 40% 7% 51%
3	D	167	% 44% 53%
3	E	167	% 31% 65%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	167	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5823 atoms, of which 2696 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 (5'-D(*TP*CP*CP*TP*CP*AP*TP*GP*CP*AP*TP*AP*TP*GP*CP*AP*TP*GP*AP*GP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
1	A	21	666	205	239	77	125	20	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*CP*CP*TP*CP*AP*TP*GP*CP*AP*TP*AP*TP*GP*CP*AP*TP*GP*AP*GP*GP*A)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	B	22	698	215	250	82	130	21	0	0	0

- Molecule 3 is a protein called POU domain, class 2, transcription factor 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	C	81	1302	414	646	111	127	4	0	1	0
3	D	78	1242	395	622	106	115	4	0	0	0
3	E	58	965	302	485	90	86	2	0	0	0
3	F	58	915	291	454	84	84	2	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	193	GLY	-	expression tag	UNP P09086
C	194	SER	-	expression tag	UNP P09086
C	195	HIS	-	expression tag	UNP P09086
C	196	MET	-	expression tag	UNP P09086
D	193	GLY	-	expression tag	UNP P09086
D	194	SER	-	expression tag	UNP P09086

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	195	HIS	-	expression tag	UNP P09086
D	196	MET	-	expression tag	UNP P09086
E	191	GLY	-	expression tag	UNP P09086
E	192	SER	-	expression tag	UNP P09086
E	193	HIS	-	expression tag	UNP P09086
E	194	MET	-	expression tag	UNP P09086
F	191	GLY	-	expression tag	UNP P09086
F	192	SER	-	expression tag	UNP P09086
F	193	HIS	-	expression tag	UNP P09086
F	194	MET	-	expression tag	UNP P09086

- Molecule 4 is BROMIDE ION (three-letter code: BR) (formula: Br).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	1	Total Br 1 1	0	0

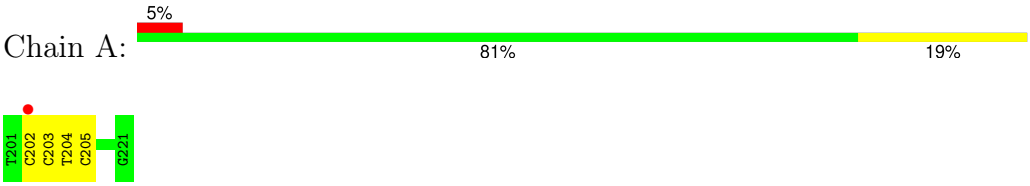
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	6	Total O 6 6	0	0
5	B	7	Total O 7 7	0	0
5	C	11	Total O 11 11	0	0
5	D	7	Total O 7 7	0	0
5	E	2	Total O 2 2	0	0
5	F	1	Total O 1 1	0	0

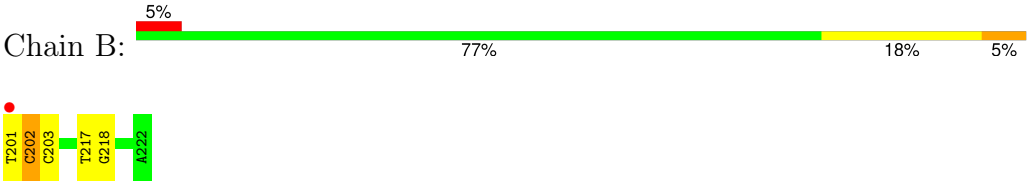
3 Residue-property plots (i)

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.

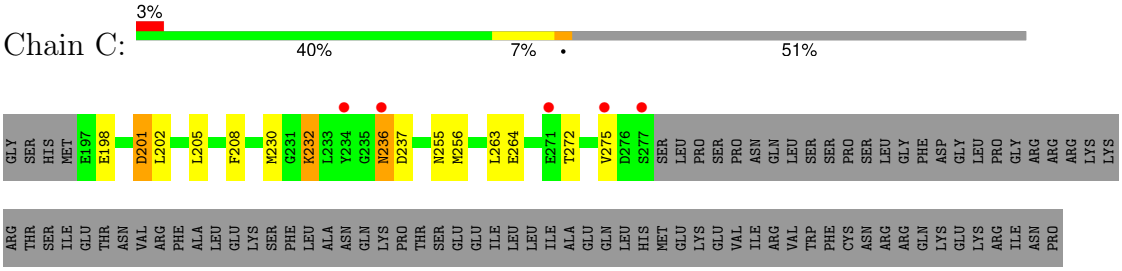
• Molecule 1: DNA (5'-D(*TP*CP*CP*TP*CP*AP*TP*GP*CP*AP*TP*AP*TP*GP*CP*AP *TP*GP*AP*GP*G)-3')



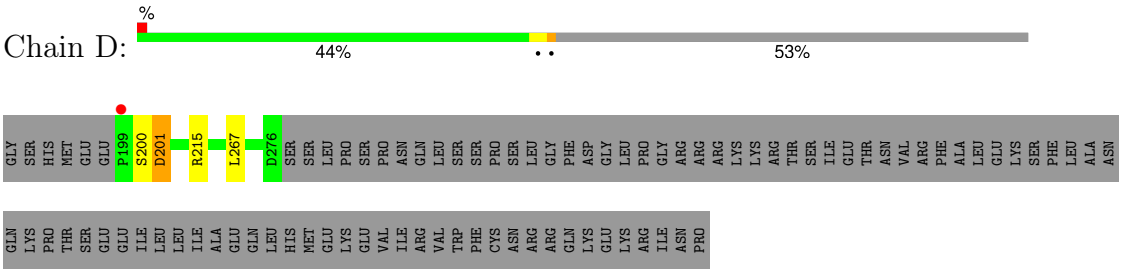
• Molecule 2: DNA (5'-D(*TP*CP*CP*TP*CP*AP*TP*GP*CP*AP*TP*AP*TP*GP*CP*AP *TP*GP*AP*GP*GP*A)-3')



• Molecule 3: POU domain, class 2, transcription factor 2

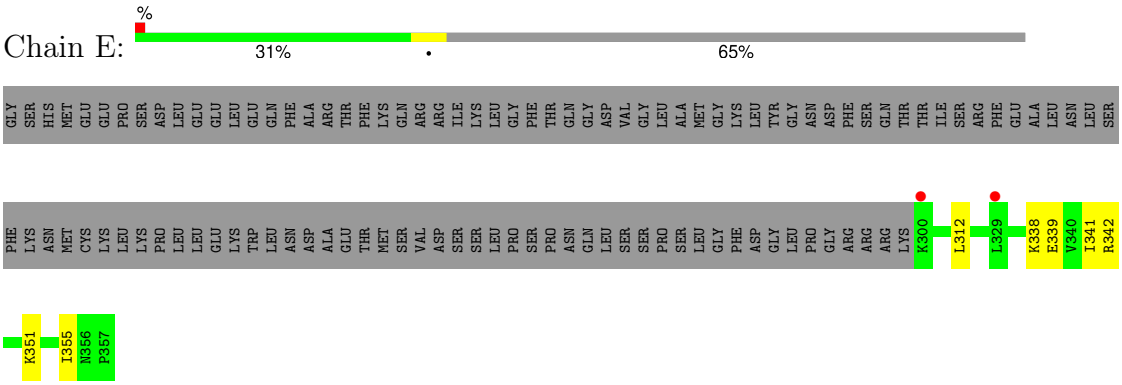


• Molecule 3: POU domain, class 2, transcription factor 2



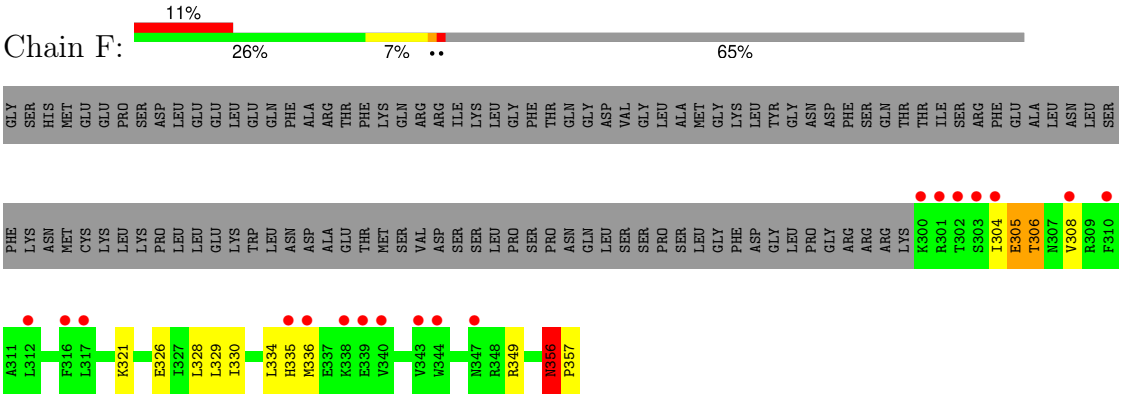
• Molecule 3: POU domain, class 2, transcription factor 2

Chain E:



• Molecule 3: POU domain, class 2, transcription factor 2

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	38.03Å 54.92Å 69.16Å 82.37° 79.66° 71.77°	Depositor
Resolution (Å)	33.91 – 2.54 33.91 – 2.54	Depositor EDS
% Data completeness (in resolution range)	95.5 (33.91-2.54) 95.4 (33.91-2.54)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.54Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.217 , 0.242 0.216 , 0.240	Depositor DCC
R_{free} test set	15517 reflections (10.01%)	wwPDB-VP
Wilson B-factor (Å^2)	51.5	Xtriage
Anisotropy	0.787	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 48.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5823	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.21% 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

Bond lengths and bond angles in the following residue types are not validated in this section: BR

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	0.62	0/478	0.97	0/736
2	B	0.61	0/502	0.97	1/773 (0.1%)
3	C	0.31	0/669	0.53	0/896
3	D	0.30	0/630	0.55	0/843
3	E	0.30	0/487	0.60	0/653
3	F	0.32	0/467	0.59	0/629
All	All	0.42	0/3233	0.73	1/4530 (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})
2	B	202	DC	O4'-C1'-N1	5.64	111.95	108.00

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	427	239	239	3	0
2	B	448	250	250	5	0
3	C	656	646	648	10	0
3	D	620	622	622	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	480	485	485	4	0
3	F	461	454	454	9	0
4	E	1	0	0	0	0
5	A	6	0	0	0	0
5	B	7	0	0	0	0
5	C	11	0	0	0	0
5	D	7	0	0	0	0
5	E	2	0	0	0	0
5	F	1	0	0	0	0
All	All	3127	2696	2698	28	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 5.

All (28) 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:A:205:DC:O2	2:B:218:DG:N2	2.17	0.77
3:F:305:GLU:O	3:F:306:THR:OG1	2.09	0.69
3:C:236:ASN:O	3:C:236:ASN:ND2	2.34	0.60
3:F:304:ILE:CD1	3:F:336:MET:SD	2.89	0.60
3:C:208:PHE:CE1	3:C:264:GLU:HG2	2.40	0.57
3:C:201:ASP:O	3:C:205:LEU:HD23	2.05	0.56
3:E:312:LEU:HD21	3:E:341:ILE:HD12	1.88	0.55
3:D:200:SER:O	3:D:201:ASP:HB2	2.07	0.54
1:A:202:DC:H2'	1:A:203:DC:C6	2.44	0.53
3:F:326:GLU:O	3:F:329:LEU:HG	2.11	0.51
2:B:203:DC:OP2	3:E:342:ARG:HD3	2.10	0.50
3:C:232:LYS:HE2	3:C:232:LYS:N	2.27	0.50
3:C:230:MET:SD	3:C:263:LEU:HD23	2.52	0.49
3:C:256:MET:HG3	3:E:355:ILE:HD12	1.94	0.49
1:A:204:DT:OP2	3:F:349:ARG:NE	2.36	0.49
3:F:329:LEU:HD12	3:F:330:ILE:N	2.28	0.48
3:F:304:ILE:HD13	3:F:336:MET:SD	2.55	0.47
3:F:308:VAL:HG13	3:F:334:LEU:HD22	1.97	0.46
3:D:215:ARG:CZ	3:D:267:LEU:HD13	2.45	0.46
2:B:201:DT:H2"	3:E:339:GLU:OE2	2.16	0.46
3:C:272:THR:HA	3:C:275:VAL:HG22	1.98	0.45
3:C:237:ASP:OD1	3:C:237:ASP:N	2.41	0.45
3:C:208:PHE:CZ	3:C:264:GLU:HG2	2.53	0.44
3:F:356:ASN:HB3	3:F:357:PRO:HD3	1.98	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:202:LEU:HD13	3:C:202:LEU:O	2.19	0.42
2:B:202:DC:H2"	2:B:203:DC:O5'	2.20	0.42
3:F:328:LEU:C	3:F:328:LEU:HD13	2.41	0.41
2:B:217:DT:H2"	2:B:218:DG:C8	2.56	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
3	C	80/167 (48%)	79 (99%)	1 (1%)	0	100 100
3	D	76/167 (46%)	73 (96%)	2 (3%)	1 (1%)	10 13
3	E	56/167 (34%)	55 (98%)	1 (2%)	0	100 100
3	F	56/167 (34%)	50 (89%)	2 (4%)	4 (7%)	1 0
All	All	268/668 (40%)	257 (96%)	6 (2%)	5 (2%)	6 7

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	201	ASP
3	F	306	THR
3	F	335	HIS
3	F	305	GLU
3	F	356	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	
3	C	72/151 (48%)	67 (93%)	5 (7%)	13	16
3	D	67/151 (44%)	67 (100%)	0	100	100
3	E	52/151 (34%)	50 (96%)	2 (4%)	28	41
3	F	48/151 (32%)	46 (96%)	2 (4%)	25	37
All	All	239/604 (40%)	230 (96%)	9 (4%)	28	41

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	198	GLU
3	C	201	ASP
3	C	232	LYS
3	C	236	ASN
3	C	255	ASN
3	E	338	LYS
3	E	351	LYS
3	F	321	LYS
3	F	356	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 oligosaccharides in this entry.

5.6 Ligand geometry

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	21/21 (100%)	-0.02	1 (4%) 36 42	38, 53, 82, 97	0
2	B	22/22 (100%)	0.00	1 (4%) 39 44	34, 62, 88, 121	0
3	C	81/167 (48%)	0.37	5 (6%) 28 31	22, 56, 111, 125	1 (1%)
3	D	78/167 (46%)	0.17	1 (1%) 74 77	32, 57, 94, 115	0
3	E	58/167 (34%)	0.45	2 (3%) 48 53	36, 71, 101, 104	0
3	F	58/167 (34%)	1.63	18 (31%) 1 1	84, 140, 158, 164	0
All	All	318/711 (44%)	0.51	28 (8%) 17 20	22, 66, 149, 164	1 (0%)

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	300	LYS	5.2
2	B	201	DT	4.3
3	C	236	ASN	4.0
3	F	343	VAL	3.7
3	F	338	LYS	3.5
3	F	312	LEU	3.4
3	F	301	ARG	3.1
3	F	310	PHE	3.0
3	C	277	SER	2.9
3	D	199	PRO	2.8
3	C	234	TYR	2.8
3	F	339	GLU	2.8
3	F	335	HIS	2.6
3	C	271	GLU	2.6
3	F	316	PHE	2.6
3	E	300	LYS	2.5
3	F	340	VAL	2.4
3	F	336	MET	2.3
3	C	275	VAL	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	202	DC	2.2
3	F	302	THR	2.2
3	F	304	ILE	2.2
3	F	317	LEU	2.1
3	F	308	VAL	2.1
3	E	329	LEU	2.1
3	F	344	TRP	2.1
3	F	303	SER	2.1
3	F	347	ASN	2.0

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	BR	E	401	1/1	0.84	0.11	138,138,138,138	0

6.5 Other polymers (i)

There are no such residues in this entry.