



# Full wwPDB X-ray Structure Validation Report ⓘ

May 25, 2020 – 02:55 pm BST

PDB ID : 2GL7  
Title : Crystal Structure of a beta-catenin/BCL9/Tcf4 complex  
Authors : Sampietro, J.  
Deposited on : 2006-04-04  
Resolution : 2.60 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](mailto: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 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 : 2.11  
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) : 2.11

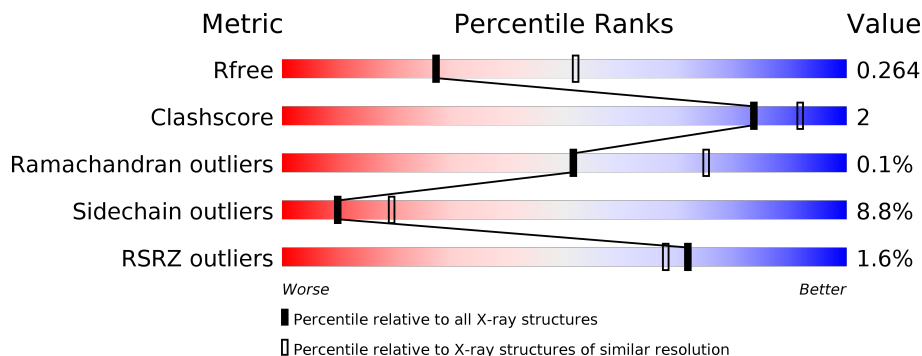
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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  $\geq 3$ , 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	550	
1	D	550	
2	B	53	
2	E	53	
3	C	46	
3	F	46	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8429 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 Beta-catenin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	512	3830	2409	688	708	25	0	0	0
1	D	506	3753	2363	679	686	25	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	137	SER	-	CLONING ARTIFACT	UNP P35222
A	142	GLU	TYR	ENGINEERED	UNP P35222
D	137	SER	-	CLONING ARTIFACT	UNP P35222
D	142	GLU	TYR	ENGINEERED	UNP P35222

- Molecule 2 is a protein called Transcription factor 7-like 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	30	218	134	35	49	0	0	0
2	E	28	201	123	32	46	0	0	0

- Molecule 3 is a protein called B-cell lymphoma 9 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	23	167	104	32	30	1	0	0	0
3	F	17	132	80	27	25		0	0	0


- Molecule 4 is water.

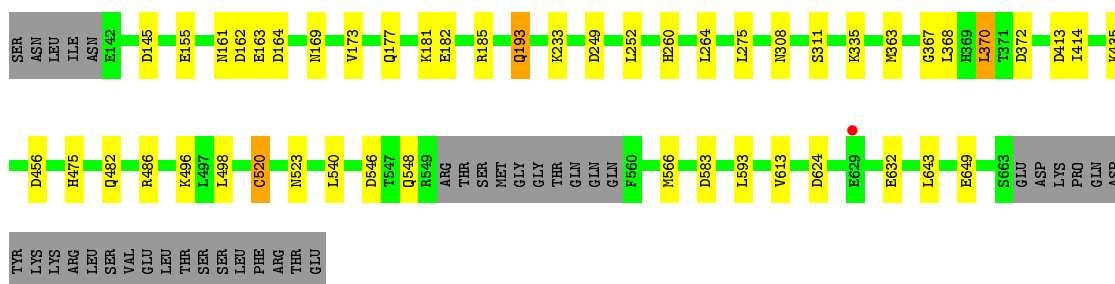
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	70	Total 70	O 70	0	0
4	B	1	Total 1	O 1	0	0
4	D	56	Total 56	O 56	0	0
4	E	1	Total 1	O 1	0	0

### 3 Residue-property plots [i](#)


These plots are drawn for all protein, RNA and DNA 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.

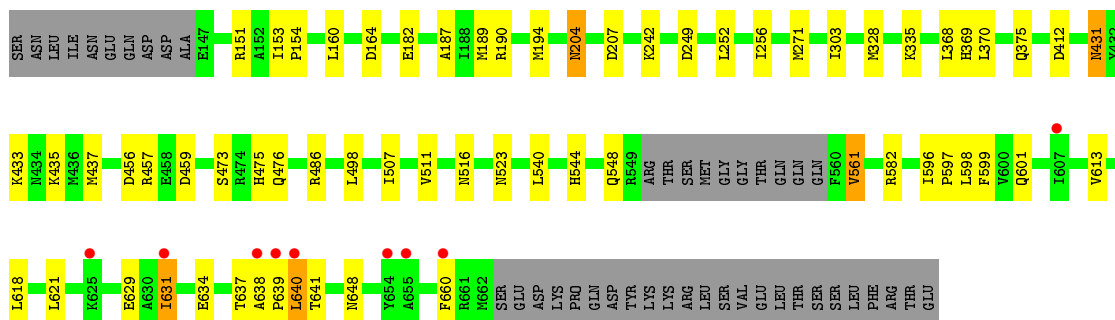
- Molecule 1: Beta-catenin

Chain A: 



- Molecule 1: Beta-catenin

Chain D: 



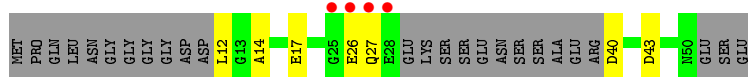
- Molecule 2: Transcription factor 7-like 2

Chain B: 

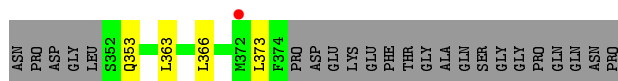
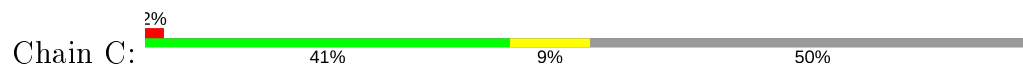


- Molecule 2: Transcription factor 7-like 2

Chain E: 



- Molecule 3: B-cell lymphoma 9 protein



- Molecule 3: B-cell lymphoma 9 protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.76Å 119.44Å 130.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.60 48.38 – 2.39	Depositor EDS
% Data completeness (in resolution range)	99.1 (30.00-2.60) 91.1 (48.38-2.39)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 2.39Å)	Xtrriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.221 , 0.267 0.217 , 0.264	Depositor DCC
$R_{free}$ test set	2787 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.8	Xtrriage
Anisotropy	0.750	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 45.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8429	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.99% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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 [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	0.40	0/3883	0.63	8/5284 (0.2%)
1	D	0.40	0/3806	0.63	6/5180 (0.1%)
2	B	0.45	0/217	0.87	3/290 (1.0%)
2	E	0.40	0/200	0.73	1/268 (0.4%)
3	C	0.40	0/168	0.48	0/226
3	F	0.39	0/132	0.76	0/177
All	All	0.40	0/8406	0.64	18/11425 (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
2	B	1	0

There are no bond length outliers.

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	412	ASP	CB-CG-OD2	7.27	124.84	118.30
1	A	456	ASP	CB-CG-OD2	7.07	124.67	118.30
1	D	459	ASP	CB-CG-OD2	7.00	124.60	118.30
1	A	249	ASP	CB-CG-OD2	6.12	123.81	118.30
1	D	456	ASP	CB-CG-OD2	5.89	123.60	118.30
1	A	583	ASP	CB-CG-OD2	5.71	123.44	118.30
1	A	162	ASP	CB-CG-OD2	5.60	123.34	118.30
1	A	624	ASP	CB-CG-OD2	5.52	123.27	118.30
1	D	164	ASP	CB-CG-OD2	5.32	123.09	118.30
1	A	164	ASP	CB-CG-OD2	5.27	123.04	118.30
2	B	43	ASP	CB-CG-OD2	5.21	122.99	118.30
1	A	413	ASP	CB-CG-OD2	5.21	122.98	118.30

*Continued on next page...*



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	43	ASP	CB-CG-OD2	5.20	122.98	118.30
1	D	207	ASP	CB-CG-OD2	5.20	122.98	118.30
2	B	16	ASP	CB-CG-OD2	5.17	122.95	118.30
1	D	249	ASP	CB-CG-OD2	5.14	122.92	118.30
1	A	145	ASP	CB-CG-OD2	5.11	122.90	118.30
2	B	23	ASP	CB-CG-OD2	5.01	122.81	118.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	39	ARG	CA

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	A	3830	0	3909	10	0
1	D	3753	0	3819	21	0
2	B	218	0	199	0	0
2	E	201	0	180	4	0
3	C	167	0	146	0	0
3	F	132	0	122	9	0
4	A	70	0	0	1	1
4	B	1	0	0	0	0
4	D	56	0	0	0	1
4	E	1	0	0	0	0
All	All	8429	0	8375	40	1

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 2.

All (40) 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:F:365:THR:O	3:F:369:ILE:CG1	1.88	1.21
3:F:365:THR:O	3:F:369:ILE:HG12	1.40	1.20
3:F:365:THR:O	3:F:369:ILE:HG13	1.85	0.77
3:F:365:THR:HG22	3:F:369:ILE:HD11	1.74	0.70
3:F:365:THR:O	3:F:369:ILE:CD1	2.46	0.63
1:A:308:ASN:HD22	1:A:311:SER:H	1.50	0.59
1:D:204:ASN:C	1:D:204:ASN:HD22	2.05	0.59
3:F:365:THR:HG22	3:F:369:ILE:CD1	2.34	0.57
1:D:516:ASN:HD21	2:E:14:ALA:H	1.53	0.56
3:F:366:LEU:HA	3:F:369:ILE:HG13	1.87	0.56
1:A:435:LYS:HB3	1:A:475:HIS:CD2	2.42	0.55
1:A:367:GLY:HA2	1:A:370:LEU:HD22	1.90	0.54
1:D:431:ASN:HD22	1:D:431:ASN:C	2.10	0.54
1:D:486:ARG:HH21	1:D:523:ASN:ND2	2.08	0.51
1:D:303:ILE:HA	2:E:27:GLN:HE22	1.76	0.51
1:D:637:THR:HA	1:D:640:LEU:HD22	1.93	0.49
3:F:356:LEU:HD13	3:F:356:LEU:H	1.77	0.49
1:D:153:ILE:N	1:D:154:PRO:HD2	2.28	0.49
1:A:193:GLN:HE21	1:A:193:GLN:H	1.62	0.47
1:D:160:LEU:HD12	1:D:194:MET:HE1	1.97	0.46
1:D:328:MET:O	1:D:369:HIS:HE1	1.98	0.45
1:D:507:ILE:O	1:D:511:VAL:HG23	2.17	0.45
1:A:486:ARG:HH21	1:A:523:ASN:ND2	2.14	0.45
1:A:169:ASN:O	1:A:173:VAL:HG23	2.17	0.45
1:D:618:LEU:HD22	1:D:631:ILE:HD11	2.00	0.44
1:D:638:ALA:HB3	1:D:639:PRO:HD3	1.99	0.44
1:D:640:LEU:HD23	1:D:641:THR:N	2.33	0.44
3:F:365:THR:O	3:F:369:ILE:HD11	2.16	0.44
1:D:160:LEU:HD12	1:D:194:MET:CE	2.48	0.44
1:D:582:ARG:HH22	2:E:12:LEU:HD11	1.83	0.44
1:D:435:LYS:HD3	1:D:475:HIS:HD2	1.83	0.43
1:D:486:ARG:HB2	1:D:523:ASN:HD21	1.84	0.43
1:A:482:GLN:HE21	1:A:520:CYS:HB2	1.84	0.42
1:A:486:ARG:HB2	1:A:523:ASN:HD21	1.84	0.42
1:D:596:ILE:N	1:D:597:PRO:CD	2.83	0.42
1:A:260:HIS:CE1	1:A:264:LEU:HD11	2.55	0.42
1:D:516:ASN:HD21	2:E:14:ALA:N	2.18	0.42
1:D:153:ILE:HD12	1:D:187:ALA:HB1	2.02	0.41
1:A:363:MET:HB2	4:A:3:HOH:O	2.20	0.41
1:D:599:PHE:HB2	1:D:618:LEU:HD21	2.02	0.41

All (1) 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)
4:A:4:HOH:O	4:D:1:HOH:O[2_655]	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	
1	A	508/550 (92%)	500 (98%)	8 (2%)	0	100	100
1	D	502/550 (91%)	496 (99%)	5 (1%)	1 (0%)	47	71
2	B	26/53 (49%)	25 (96%)	1 (4%)	0	100	100
2	E	24/53 (45%)	21 (88%)	3 (12%)	0	100	100
3	C	21/46 (46%)	21 (100%)	0	0	100	100
3	F	15/46 (33%)	15 (100%)	0	0	100	100
All	All	1096/1298 (84%)	1078 (98%)	17 (2%)	1 (0%)	51	75

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	561	VAL

#### 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	407/460 (88%)	379 (93%)	28 (7%)	15	31
1	D	393/460 (85%)	359 (91%)	34 (9%)	10	20

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	23/44 (52%)	19 (83%)	4 (17%)	2	3
2	E	21/44 (48%)	18 (86%)	3 (14%)	3	5
3	C	15/41 (37%)	11 (73%)	4 (27%)	0	1
3	F	13/41 (32%)	9 (69%)	4 (31%)	0	0
All	All	872/1090 (80%)	795 (91%)	77 (9%)	10	19

All (77) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	155	GLU
1	A	161	ASN
1	A	163	GLU
1	A	177	GLN
1	A	181	LYS
1	A	182	GLU
1	A	185	ARG
1	A	193	GLN
1	A	233	LYS
1	A	252	LEU
1	A	275	LEU
1	A	335	LYS
1	A	368	LEU
1	A	370	LEU
1	A	372	ASP
1	A	414	ILE
1	A	496	LYS
1	A	498	LEU
1	A	520	CYS
1	A	540	LEU
1	A	546	ASP
1	A	548	GLN
1	A	566	MET
1	A	593	LEU
1	A	613	VAL
1	A	632	GLU
1	A	643	LEU
1	A	649	GLU
2	B	22	LYS
2	B	23	ASP
2	B	28	GLU
2	B	41	LEU

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>
3	C	353	GLN
3	C	363	LEU
3	C	366	LEU
3	C	373	LEU
1	D	151	ARG
1	D	182	GLU
1	D	189	MET
1	D	190	ARG
1	D	204	ASN
1	D	242	LYS
1	D	252	LEU
1	D	256	ILE
1	D	271	MET
1	D	335	LYS
1	D	368	LEU
1	D	370	LEU
1	D	375	GLN
1	D	431	ASN
1	D	433	LYS
1	D	437	MET
1	D	457	ARG
1	D	473	SER
1	D	476	GLN
1	D	498	LEU
1	D	540	LEU
1	D	544	HIS
1	D	548	GLN
1	D	561	VAL
1	D	598	LEU
1	D	601	GLN
1	D	613	VAL
1	D	621	LEU
1	D	629	GLU
1	D	631	ILE
1	D	634	GLU
1	D	640	LEU
1	D	648	ASN
1	D	660	PHE
2	E	17	GLU
2	E	26	GLU
2	E	40	ASP
3	F	356	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	F	363	LEU
3	F	368	ASP
3	F	369	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (33) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	169	ASN
1	A	193	GLN
1	A	308	ASN
1	A	322	GLN
1	A	326	ASN
1	A	369	HIS
1	A	380	ASN
1	A	430	ASN
1	A	434	ASN
1	A	482	GLN
1	A	523	ASN
1	A	545	GLN
1	A	586	ASN
1	A	648	ASN
1	D	169	ASN
1	D	204	ASN
1	D	260	HIS
1	D	326	ASN
1	D	369	HIS
1	D	380	ASN
1	D	395	GLN
1	D	430	ASN
1	D	431	ASN
1	D	434	ASN
1	D	470	HIS
1	D	482	GLN
1	D	516	ASN
1	D	523	ASN
1	D	538	GLN
1	D	548	GLN
1	D	578	HIS
1	D	601	GLN
2	E	27	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

### 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	512/550 (93%)	-0.41	1 (0%) 95 95	8, 21, 43, 56	0
1	D	506/550 (92%)	-0.31	9 (1%) 68 64	9, 21, 47, 59	0
2	B	30/53 (56%)	-0.31	0 100 100	22, 29, 46, 50	0
2	E	28/53 (52%)	0.23	4 (14%) 2 1	24, 30, 40, 41	0
3	C	23/46 (50%)	0.05	1 (4%) 35 28	18, 36, 40, 41	0
3	F	17/46 (36%)	0.67	3 (17%) 1 0	35, 43, 48, 48	0
All	All	1116/1298 (85%)	-0.32	18 (1%) 72 68	8, 22, 46, 59	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	655	ALA	5.0
1	D	638	ALA	4.9
1	D	654	TYR	3.8
1	D	639	PRO	3.1
1	D	631	ILE	2.8
1	D	660	PHE	2.8
1	D	625	LYS	2.7
2	E	28	GLU	2.7
1	D	640	LEU	2.6
3	F	368	ASP	2.5
2	E	26	GLU	2.4
3	F	369	ILE	2.4
1	D	607	ILE	2.3
2	E	27	GLN	2.2
3	C	372	MET	2.2
2	E	25	GLY	2.2
1	A	629	GLU	2.1
3	F	364	GLN	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 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.