



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 02:38 am GMT

PDB ID : 2VBL  
Title : Molecular basis of human XPC gene recognition and cleavage by engineered homing endonuclease heterodimers  
Authors : Redondo, P.; Prieto, J.; Munoz, I.G.; Alibes, A.; Stricher, F.; Serrano, L.; Arnould, S.; Perez, C.; Cabaniols, J.P.; Duchateau, P.; Paques, F.; Blanco, F.J.; Montoya, G.  
Deposited on : 2007-09-14  
Resolution : 1.80 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](mailto: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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

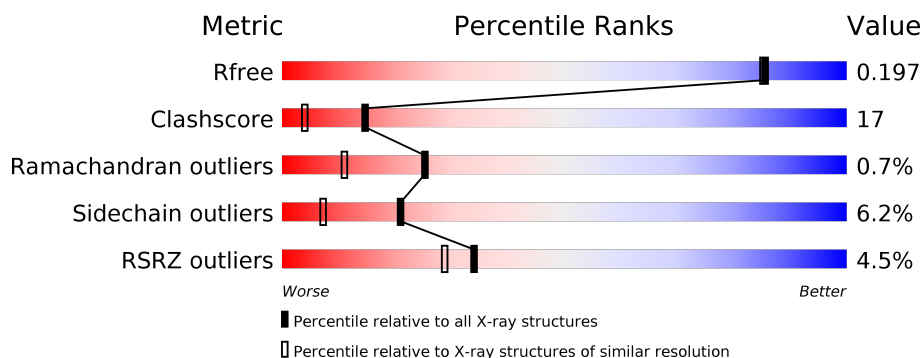
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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))
$R_{free}$	100719	4827 (1.80-1.80)
Clashscore	112137	5742 (1.80-1.80)
Ramachandran outliers	110173	5676 (1.80-1.80)
Sidechain outliers	110143	5675 (1.80-1.80)
RSRZ outliers	101464	4906 (1.80-1.80)

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. 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	153	<div> <div>6%</div> <div> <div></div> <div>70%</div> <div>25%</div> <div>• •</div> </div> </div>
2	B	153	<div> <div>5%</div> <div> <div></div> <div>82%</div> <div>13%</div> <div>• • •</div> </div> </div>
3	C	14	<div> <div>21%</div> <div>43%</div> <div>36%</div> </div>
4	E	14	<div> <div>14%</div> <div>57%</div> <div>29%</div> </div>
5	S	10	<div> <div>20%</div> <div>60%</div> <div>20%</div> </div>
6	T	10	<div> <div>10%</div> <div>40%</div> <div>50%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
7	MG	E	1025	-	-	-	X

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 3825 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 DNA ENDONUCLEASE I-CREI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	152	Total	C	N	O	S	0	0	0
			1224	788	209	226	1			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	GLU	LYS	CONFLICT	UNP P05725
A	31	ALA	GLN	CONFLICT	UNP P05725
A	38	ARG	GLN	CONFLICT	UNP P05725
A	40	LYS	SER	CONFLICT	UNP P05725
A	42	THR	ALA	CONFLICT	UNP P05725
A	44	LYS	GLN	CONFLICT	UNP P05725
A	70	GLU	ARG	CONFLICT	UNP P05725
A	75	ASN	ASP	CONFLICT	UNP P05725
A	85	ARG	HIS	CONFLICT	UNP P05725
A	109	THR	ILE	CONFLICT	UNP P05725
A	110	GLU	TRP	CONFLICT	UNP P05725
A	111	GLN	ARG	CONFLICT	UNP P05725

- Molecule 2 is a protein called DNA ENDONUCLEASE I-CREI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	152	Total	C	N	O	S	0	0	0
			1224	787	210	226	1			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	19	ALA	GLY	CONFLICT	UNP P05725
B	28	ALA	LYS	CONFLICT	UNP P05725
B	33	SER	TYR	CONFLICT	UNP P05725
B	38	ARG	GLN	CONFLICT	UNP P05725

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	40	LYS	SER	CONFLICT	UNP P05725
B	42	THR	ALA	CONFLICT	UNP P05725
B	69	GLY	ASP	CONFLICT	UNP P05725
B	70	SER	ARG	CONFLICT	UNP P05725
B	75	ASN	ASP	CONFLICT	UNP P05725
B	110	GLU	TRP	CONFLICT	UNP P05725
B	111	GLN	ARG	CONFLICT	UNP P05725

- Molecule 3 is a DNA chain called 5'-D(\*DT\*DC\*DT\*DG\*DC\*DC\*DT\*DT\*DT\*DT\*DT\*DT\*DGP\*DAP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	14	Total	C	N	O	P	0	0	0
			279	137	40	89	13			

- Molecule 4 is a DNA chain called 5'-D(\*DT\*DT\*DA\*DG\*DG\*DA\*DT\*DC\*DC\*DT\*DT\*DC\*DAP\*DAP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	14	Total	C	N	O	P	0	0	0
			282	137	49	83	13			

- Molecule 5 is a DNA chain called 5'-D(\*DA\*DA\*DA\*DA\*DG\*DG\*DC\*DA\*DG\*DAP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	S	10	Total	C	N	O	P	0	0	0
			212	99	48	55	10			

- Molecule 6 is a DNA chain called 5'-D(\*DA\*DG\*DG\*DA\*DT\*DC\*DC\*DT\*DA\*DAP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	T	10	Total	C	N	O	P	0	0	0
			207	98	40	59	10			

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	C	2	Total	Mg	0	0
			2	2		
7	E	1	Total	Mg	0	0
			1	1		

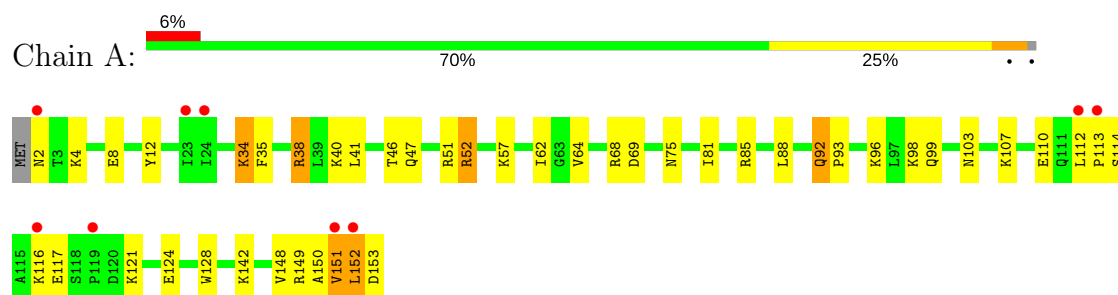
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	111	Total 111	O 111	0	0
8	B	151	Total 151	O 151	0	0
8	C	53	Total 53	O 53	0	0
8	E	34	Total 34	O 34	0	0
8	S	27	Total 27	O 27	0	0
8	T	18	Total 18	O 18	0	0

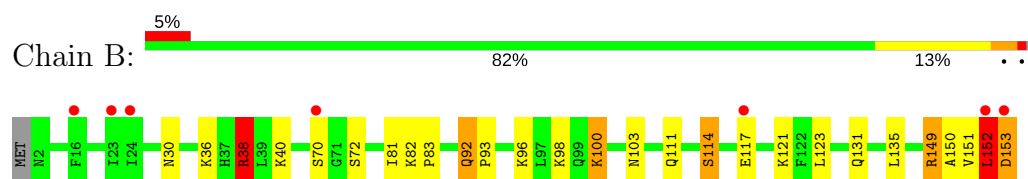
### 3 Residue-property plots [i](#)

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.

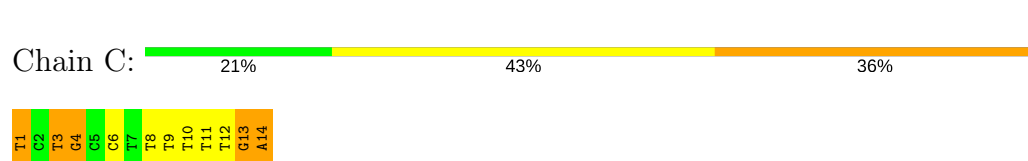
- Molecule 1: DNA ENDONUCLEASE I-CREI



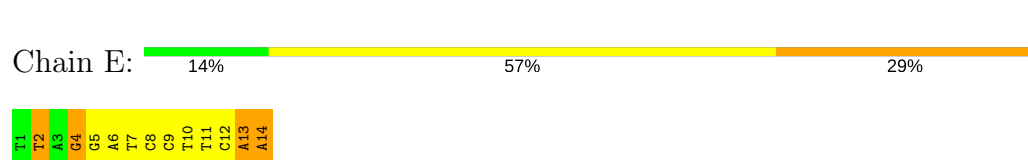
- Molecule 2: DNA ENDONUCLEASE I-CREI



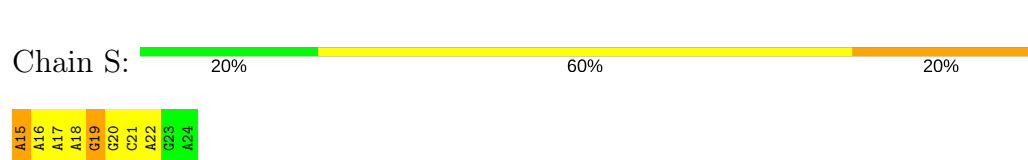
- Molecule 3: 5'-D(\*DT\*DC\*DT\*DG\*DC\*DC\*DT\*DT\*DT\*DT\*DT\*DT \*DGP\*DAP)-3'



- Molecule 4: 5'-D(\*DT\*DT\*DA\*DG\*DG\*DA\*DT\*DC\*DC\*DT\*DT\*DC \*DAP\*DAP)-3'



- Molecule 5: 5'-D(\*DA\*DA\*DA\*DA\*DG\*DG\*DC\*DA\*DG\*DAP)-3'



- Molecule 6: 5'-D(\*DA\*DG\*DG\*DA\*DT\*DC\*DC\*DT\*DA\*DAP)-3'





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.61Å 69.21Å 88.67Å 90.00° 95.46° 90.00°	Depositor
Resolution (Å)	32.92 – 1.80 32.91 – 1.80	Depositor EDS
% Data completeness (in resolution range)	95.5 (32.92-1.80) 95.5 (32.91-1.80)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.48 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.147 , 0.197 0.150 , 0.197	Depositor DCC
$R_{free}$ test set	2399 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.6	Xtriage
Anisotropy	0.300	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 49.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3825	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.10% 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

### 5.1 Standard geometry

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

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	0.56	0/1246	0.74	1/1682 (0.1%)
2	B	0.65	0/1246	0.79	2/1681 (0.1%)
3	C	1.25	0/309	2.38	22/475 (4.6%)
4	E	0.94	0/315	2.03	12/484 (2.5%)
5	S	1.20	1/240 (0.4%)	2.45	21/367 (5.7%)
6	T	1.03	1/232 (0.4%)	2.10	14/354 (4.0%)
All	All	0.80	2/3588 (0.1%)	1.44	72/5043 (1.4%)

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
2	B	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	T	15	DA	OP3-P	-10.09	1.49	1.61
5	S	15	DA	OP3-P	-8.13	1.51	1.61

The worst 5 of 72 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	S	15	DA	OP1-P-OP2	-11.95	101.68	119.60
4	E	13	DA	O4'-C1'-N9	-11.28	100.10	108.00
3	C	8	DT	C6-C5-C7	-10.66	116.50	122.90

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	10	DT	O4'-C1'-N1	-9.90	101.07	108.00
5	S	16	DA	O4'-C1'-N9	9.82	114.87	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	152	LEU	Peptide
2	B	152	LEU	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	1224	0	1256	54	0
2	B	1224	0	1266	36	0
3	C	279	0	163	8	0
4	E	282	0	160	12	0
5	S	212	0	111	5	0
6	T	207	0	113	7	0
7	C	2	0	0	0	0
7	E	1	0	0	0	0
8	A	111	0	0	15	0
8	B	151	0	0	16	2
8	C	53	0	0	7	2
8	E	34	0	0	6	0
8	S	27	0	0	0	0
8	T	18	0	0	3	0
All	All	3825	0	3069	111	2

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 17.

The worst 5 of 111 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:69:ASP:HB2	8:A:2057:HOH:O	1.32	1.24
4:E:14:DA:O3'	5:S:15:DA:P	1.97	1.22
3:C:1:DT:H5''	8:C:2001:HOH:O	1.05	1.21
6:T:15:DA:OP3	8:T:2005:HOH:O	1.56	1.20
2:B:149:ARG:HG2	8:B:2145:HOH:O	1.04	1.17

All (2) 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)
8:B:2019:HOH:O	8:C:2042:HOH:O[1_655]	1.57	0.63
8:B:2019:HOH:O	8:C:2036:HOH:O[1_655]	1.91	0.29

## 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	150/153 (98%)	143 (95%)	6 (4%)	1 (1%)	25	11
2	B	150/153 (98%)	142 (95%)	7 (5%)	1 (1%)	25	11
All	All	300/306 (98%)	285 (95%)	13 (4%)	2 (1%)	25	11

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	150	ALA
2	B	151	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	135/139 (97%)	126 (93%)	9 (7%)	19	6
2	B	137/138 (99%)	129 (94%)	8 (6%)	23	9
All	All	272/277 (98%)	255 (94%)	17 (6%)	21	7

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	124	GLU
1	A	151	VAL
2	B	114	SER
1	A	121	LYS
2	B	149	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	47	GLN
1	A	75	ASN
2	B	75	ASN
2	B	103	ASN
2	B	131	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 are 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 [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, 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	152/153 (99%)	0.15	9 (5%)	23 19	22, 36, 69, 80	0
2	B	152/153 (99%)	-0.04	7 (4%)	33 28	20, 29, 53, 75	0
3	C	14/14 (100%)	-0.78	0	100 100	25, 33, 37, 38	0
4	E	14/14 (100%)	-0.60	0	100 100	23, 41, 47, 49	0
5	S	10/10 (100%)	-0.33	0	100 100	23, 42, 43, 46	0
6	T	10/10 (100%)	-0.04	0	100 100	27, 49, 58, 59	0
All	All	352/354 (99%)	-0.02	16 (4%)	34 29	20, 33, 59, 80	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	152	LEU	3.9
1	A	112	LEU	3.1
1	A	23	ILE	2.7
1	A	24	ILE	2.6
1	A	113	PRO	2.5

### 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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
7	MG	E	1025	1/1	1.00	0.18	2.02	17,17,17,17	0
7	MG	C	1025	1/1	1.00	0.17	1.53	17,17,17,17	0
7	MG	C	1026	1/1	1.00	0.08	-1.88	23,23,23,23	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.