



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 5, 2026 – 12:54 PM UTC

PDB ID : 5EH2 / pdb_00005eh2
Title : Human PRDM9 allele-A ZnF Domain with Associated Recombination Hotspot DNA Sequence III
Authors : Patel, A.; Horton, J.R.; Wilson, G.G.; Zhang, X.; Cheng, X.
Deposited on : 2015-10-27
Resolution : 2.05 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 : 4-5-2 with Phenix2.0
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

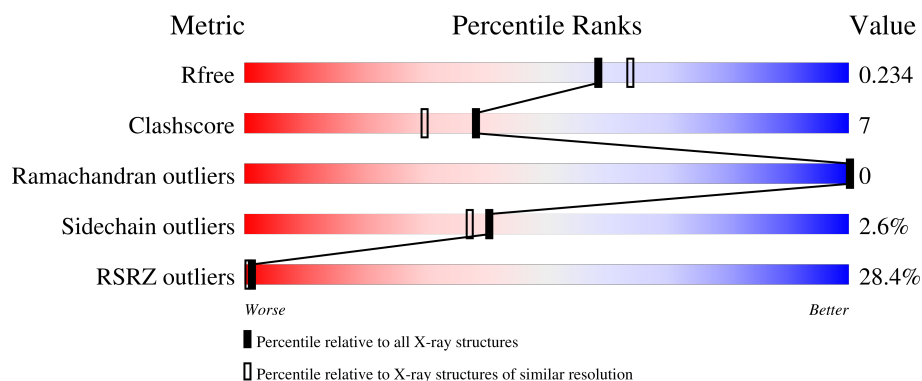
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.05 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}	180053	2260 (2.04-2.04)
Clashscore	190562	2333 (2.04-2.04)
Ramachandran outliers	187476	2318 (2.04-2.04)
Sidechain outliers	187428	2318 (2.04-2.04)
RSRZ outliers	180081	2260 (2.04-2.04)

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 $\leq 5\%$. 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	21	43% 62% 38%
1	C	21	43% 52% 43% 5%
2	B	21	52% 57% 43%
2	D	21	38% 67% 33%
3	E	144	19% 66% 10% 24%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	144	 A horizontal bar chart showing the quality of chain F. The bar is divided into four segments: a red segment at the beginning labeled '15%', a large green segment labeled '69%', a small yellow segment labeled '7%', and a grey segment at the end labeled '24%'.

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3671 atoms, of which 16 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(*AP*CP*AP*CP*GP*TP*GP*GP*CP*TP*AP*GP*GP*GP*AP*GP*GP*CP*CP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	21	Total	C	N	O	P	0	0	0
			431	204	84	123	20			
1	A	21	Total	C	N	O	P	0	0	0
			431	204	84	123	20			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	21	Total	C	N	O	P	0	0	0
			424	202	77	125	20			
2	B	21	Total	C	N	O	P	0	0	0
			424	202	77	125	20			

- Molecule 3 is a protein called Histone-lysine N-methyltransferase PRDM9.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	110	Total	C	H	N	O	S	0	2	0
			927	556	8	202	153	8			
3	F	109	Total	C	H	N	O	S	0	1	0
			904	539	8	198	151	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	714	GLY	-	expression tag	UNP Q9NQV7
E	715	SER	-	expression tag	UNP Q9NQV7
F	714	GLY	-	expression tag	UNP Q9NQV7
F	715	SER	-	expression tag	UNP Q9NQV7

- Molecule 4 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	4	Total 4	Zn 4	0	0
4	F	4	Total 4	Zn 4	0	0

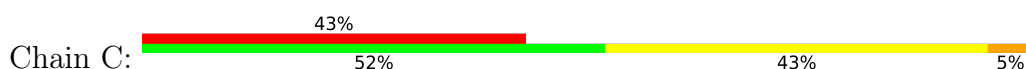
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	21	Total 21	O 21	0	0
5	D	17	Total 17	O 17	0	0
5	A	14	Total 14	O 14	0	0
5	B	11	Total 11	O 11	0	0
5	E	16	Total 16	O 16	0	0
5	F	43	Total 43	O 43	0	0

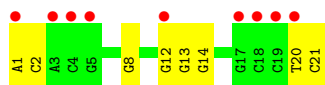
3 Residue-property plots [i](#)

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.

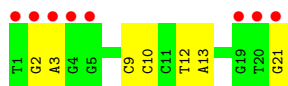
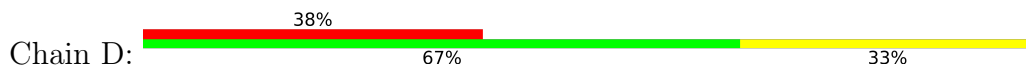
- Molecule 1: DNA (5'-D(*AP*CP*AP*CP*GP*TP*GP*GP*CP*TP*AP*GP*GP*GP*AP*GP*GP*CP*CP*TP*C)-3')



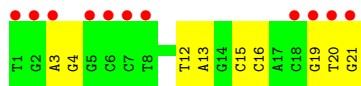
- Molecule 1: DNA (5'-D(*AP*CP*AP*CP*GP*TP*GP*GP*CP*TP*AP*GP*GP*GP*AP*GP*GP*CP*CP*TP*C)-3')



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

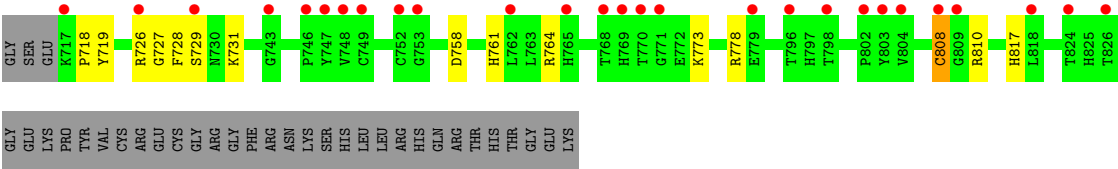


- Molecule 2: DNA (5'-D(*TP*GP*AP*GP*GP*CP*CP*TP*CP*CP*CP*TP*AP*GP*CP*CP*AP*CP*GP*TP*G)-3')

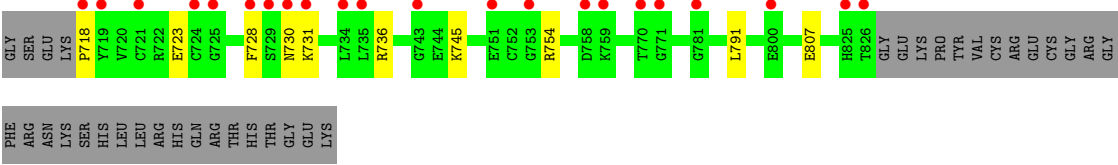


- Molecule 3: Histone-lysine N-methyltransferase PRDM9





● Molecule 3: Histone-lysine N-methyltransferase PRDM9



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	45.21Å 57.76Å 76.21Å 90.15° 75.30° 83.30°	Depositor
Resolution (Å)	32.75 – 2.05 32.75 – 2.05	Depositor EDS
% Data completeness (in resolution range)	89.7 (32.75-2.05) 84.1 (32.75-2.05)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.37 (at 2.05Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.201 , 0.231 0.203 , 0.234	Depositor DCC
R_{free} test set	1990 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å ²)	44.5	Xtriage
Anisotropy	0.477	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 77.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	3671	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.67% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: ZN

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	0.37	0/484	0.59	0/746
1	C	0.57	1/484 (0.2%)	0.75	0/746
2	B	0.29	0/474	0.48	0/729
2	D	0.38	0/474	0.60	0/729
3	E	0.52	0/948	0.92	2/1264 (0.2%)
3	F	0.60	0/920	0.85	2/1226 (0.2%)
All	All	0.49	1/3784 (0.0%)	0.75	4/5440 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	12	DG	O5'-C5'	-5.62	1.25	1.42

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	745	LYS	CA-C-N	6.54	126.13	119.19
3	F	745	LYS	C-N-CA	6.54	126.13	119.19
3	E	773	LYS	CA-C-N	5.32	125.02	119.64
3	E	773	LYS	C-N-CA	5.32	125.02	119.64

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	431	0	236	7	0
1	C	431	0	236	10	0
2	B	424	0	237	9	0
2	D	424	0	237	7	0
3	E	919	8	885	10	0
3	F	896	8	859	4	0
4	E	4	0	0	0	0
4	F	4	0	0	0	0
5	A	14	0	0	0	0
5	B	11	0	0	0	0
5	C	21	0	0	3	0
5	D	17	0	0	0	0
5	E	16	0	0	0	0
5	F	43	0	0	0	0
All	All	3655	16	2690	43	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 7.

All (43) 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:E:758:ASP:HB3	3:E:761[A]:HIS:HD2	1.35	0.90
1:C:2:DC:O2	2:D:21:DG:N2	2.17	0.76
1:A:1:DA:H2''	1:A:2:DC:H5'	1.78	0.64
2:D:2:DG:H2'	2:D:3:DA:C8	2.35	0.61
3:E:758:ASP:HB3	3:E:761[A]:HIS:CD2	2.26	0.61
2:D:9:DC:H2''	2:D:10:DC:H5'	1.82	0.60
1:C:12:DG:H5''	1:C:12:DG:C8	2.37	0.59
1:C:12:DG:H5''	1:C:12:DG:H8	1.68	0.59
1:C:13:DG:OP1	5:C:101:HOH:O	2.17	0.57
2:B:3:DA:H2''	2:B:4:DG:O5'	2.05	0.56
3:E:808:CYS:HB3	3:E:810:ARG:H	1.70	0.56
1:C:7:DG:OP2	5:C:102:HOH:O	2.18	0.54
1:A:8:DG:N7	3:E:817:HIS:NE2	2.52	0.51
2:D:2:DG:H5''	2:D:2:DG:H8	1.76	0.51
1:C:18:DC:H2'	1:C:19:DC:C6	2.47	0.49
2:B:3:DA:H3'	3:E:731:LYS:HD3	1.94	0.49
1:A:20:DT:H2'	1:A:21:DC:C6	2.48	0.48
3:F:718:PRO:HG2	3:F:728:PHE:O	2.13	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:E:726[A]:ARG:HD3	3:E:728:PHE:CZ	2.49	0.48
1:C:20:DT:H2'	1:C:21:DC:C6	2.49	0.47
2:B:3:DA:H4'	2:B:4:DG:OP1	2.16	0.46
2:B:15:DC:H2'	2:B:16:DC:C6	2.51	0.46
2:B:19:DG:H1'	2:B:20:DT:H5'	1.98	0.46
1:A:12:DG:H5''	1:A:12:DG:H8	1.81	0.45
3:E:726[A]:ARG:NH1	3:E:727:GLY:O	2.49	0.45
1:C:16:DG:O6	3:F:736:ARG:NH2	2.35	0.45
3:E:718:PRO:HG2	3:E:719:TYR:CD2	2.52	0.45
1:A:12:DG:H5''	1:A:12:DG:C8	2.52	0.45
1:A:13:DG:H2'	1:A:14:DG:C8	2.51	0.45
2:D:12:DT:H2'	2:D:13:DA:C8	2.51	0.45
3:F:730:ASN:OD1	3:F:731:LYS:N	2.50	0.45
2:B:20:DT:H2''	2:B:21:DG:C8	2.51	0.44
3:E:761[A]:HIS:ND1	3:E:764:ARG:NH2	2.67	0.43
2:B:3:DA:H2'	2:B:4:DG:H8	1.83	0.42
1:C:15:DA:N3	5:C:103:HOH:O	2.37	0.42
2:D:9:DC:H2''	2:D:10:DC:C5'	2.49	0.42
2:B:3:DA:H2'	2:B:4:DG:C8	2.55	0.42
3:E:778:ARG:HE	3:E:778:ARG:HB3	1.73	0.42
3:F:791:LEU:HD23	3:F:791:LEU:HA	1.92	0.41
2:D:12:DT:H2''	2:D:13:DA:H5'	2.02	0.41
2:B:12:DT:H2'	2:B:13:DA:C8	2.56	0.40
1:C:19:DC:H2'	1:C:20:DT:C6	2.57	0.40
1:A:1:DA:C2'	1:A:2:DC:H5'	2.51	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	E	110/144 (76%)	106 (96%)	4 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	F	108/144 (75%)	107 (99%)	1 (1%)	0	100	100
All	All	218/288 (76%)	213 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	E	99/128 (77%)	97 (98%)	2 (2%)	48	48
3	F	96/128 (75%)	92 (96%)	4 (4%)	26	20
All	All	195/256 (76%)	189 (97%)	6 (3%)	40	30

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	729	SER
3	E	808	CYS
3	F	723	GLU
3	F	754[A]	ARG
3	F	754[B]	ARG
3	F	807	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	E	738	GLN

5.3.3 RNA ⓘ

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 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 [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, 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	A	21/21 (100%)	1.85	9 (42%) 0 0	49, 64, 95, 95	0
1	C	21/21 (100%)	1.97	9 (42%) 0 0	39, 57, 91, 92	0
2	B	21/21 (100%)	2.15	11 (52%) 0 0	54, 78, 92, 94	0
2	D	21/21 (100%)	1.85	8 (38%) 1 0	44, 60, 94, 98	0
3	E	110/144 (76%)	1.53	27 (24%) 2 1	30, 60, 83, 95	2 (1%)
3	F	109/144 (75%)	1.30	22 (20%) 3 2	30, 51, 76, 91	1 (0%)
All	All	303/372 (81%)	1.56	86 (28%) 1 1	30, 58, 92, 98	3 (0%)

All (86) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	DA	5.3
3	F	718	PRO	5.0
3	E	752	CYS	4.5
2	D	1	DT	4.3
3	F	729	SER	4.2
3	F	735	LEU	4.1
2	D	2	DG	4.0
1	C	19	DC	4.0
2	B	1	DT	3.9
3	E	753	GLY	3.9
1	C	2	DC	3.8
3	F	721	CYS	3.8
3	F	719	TYR	3.7
1	A	4	DC	3.6
1	C	18	DC	3.6
2	D	3	DA	3.5
1	A	3	DA	3.4
3	E	768	THR	3.4
3	E	729	SER	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	21	DG	3.3
3	F	724	CYS	3.2
2	D	4	DG	3.2
1	C	20	DT	3.2
3	F	826	THR	3.2
3	E	746	PRO	3.2
1	C	3	DA	3.2
2	B	7	DC	3.1
2	B	8	DT	3.1
2	D	19	DG	3.0
2	B	2	DG	3.0
3	F	734	LEU	2.9
3	F	730	ASN	2.9
2	B	19	DG	2.9
1	A	5	DG	2.9
2	D	21	DG	2.9
3	E	808	CYS	2.8
1	A	19	DC	2.8
1	C	17	DG	2.8
2	B	20	DT	2.7
3	E	749	CYS	2.7
1	A	1	DA	2.7
1	A	18	DC	2.7
3	E	824	THR	2.7
3	E	747	TYR	2.7
3	F	731	LYS	2.7
3	E	826	THR	2.6
1	C	4	DC	2.6
3	E	771	GLY	2.6
2	D	5	DG	2.6
3	E	762	LEU	2.6
3	E	748	VAL	2.5
1	C	21	DC	2.5
3	E	779	GLU	2.5
3	E	743	GLY	2.5
2	B	6	DC	2.5
2	B	18	DC	2.4
3	F	800	GLU	2.4
3	F	753	GLY	2.4
3	E	769	HIS	2.4
1	A	20	DT	2.3
3	E	765	HIS	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	E	796	THR	2.3
3	F	770	THR	2.3
3	F	725	GLY	2.3
3	F	728	PHE	2.3
1	A	12	DG	2.3
3	F	825	HIS	2.3
3	E	802	PRO	2.3
3	F	771	GLY	2.3
3	E	717	LYS	2.2
2	B	5	DG	2.2
2	B	3	DA	2.2
3	E	818	LEU	2.2
3	E	803	TYR	2.2
3	E	798	THR	2.2
3	F	758	ASP	2.2
3	F	743	GLY	2.1
3	F	781	GLY	2.1
1	A	17	DG	2.1
2	D	20	DT	2.1
3	E	770	THR	2.1
3	E	726[A]	ARG	2.1
3	F	759	LYS	2.1
3	E	804	VAL	2.1
3	F	751	GLU	2.0
3	E	809	GLY	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 oligosaccharides 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, 95th 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($\text{\AA}^2$)	Q<0.9
4	ZN	E	1002	1/1	0.95	0.09	79,79,79,79	0
4	ZN	E	1003	1/1	0.98	0.05	51,51,51,51	0
4	ZN	F	1001	1/1	0.98	0.04	54,54,54,54	0
4	ZN	F	1003	1/1	0.98	0.04	46,46,46,46	0
4	ZN	E	1001	1/1	0.99	0.03	47,47,47,47	0
4	ZN	F	1002	1/1	0.99	0.04	48,48,48,48	0
4	ZN	E	1004	1/1	0.99	0.03	68,68,68,68	0
4	ZN	F	1004	1/1	0.99	0.03	50,50,50,50	0

6.5 Other polymers [i](#)

There are no such residues in this entry.