Full wwPDB X-ray Structure Validation Report ⓘ

Sep 9, 2023 – 01:21 PM EDT

PDB ID	:	8SST
Title	:	ZnFs 1-7 of CCCTC-binding factor (CTCF) K365T Mutant Complexed with 23mer
Authors	:	Horton, J.R.; Yang, J.; Cheng, X.
Deposited on	:	2023-05-08
Resolution	:	2.19 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:

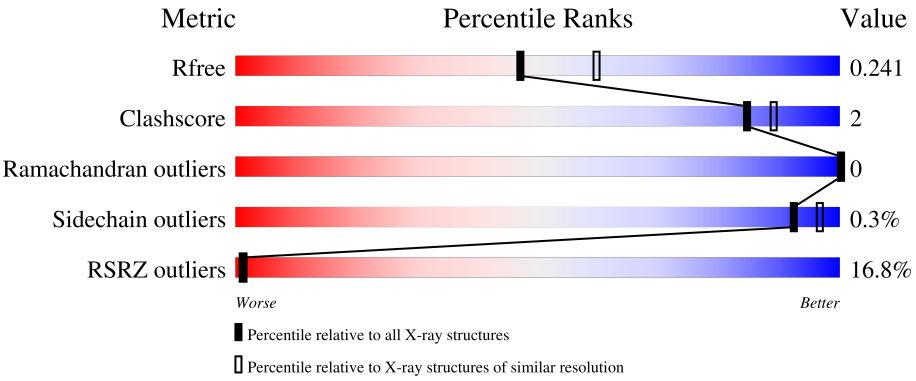
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.19 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.



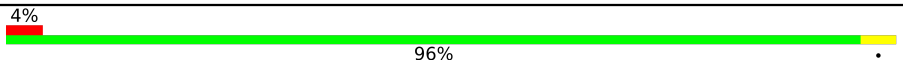
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	203	17% 93%
1	D	203	22% 92% 6%
2	B	23	9% 100%
2	E	23	78% 22%
3	C	23	4% 74% 26%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	23	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 5149 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 Transcriptional repressor CTCF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	197	Total	C	N	O	S	0	0	0
			1559	959	313	266	21			
1	D	198	Total	C	N	O	S	0	0	0
			1581	973	317	270	21			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	365	THR	LYS	engineered mutation	UNP P49711
D	365	THR	LYS	engineered mutation	UNP P49711

- Molecule 2 is a DNA chain called DNA Strand (23mer) I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	23	Total	C	N	O	P	0	0	0
			457	218	79	138	22			
2	E	23	Total	C	N	O	P	0	0	0
			457	218	79	138	22			

- Molecule 3 is a DNA chain called DNA Strand (23mer) II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	23	Total	C	N	O	P	0	0	0
			480	225	99	134	22			
3	F	23	Total	C	N	O	P	0	0	0
			480	225	99	134	22			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

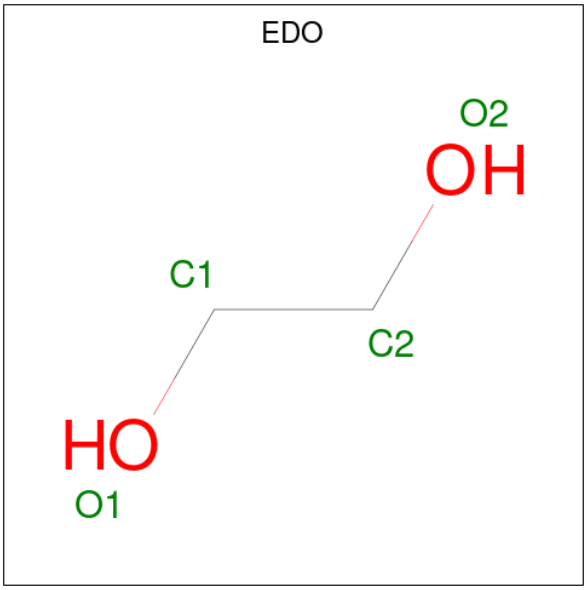
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	7	Total	Zn	0	0
			7	7		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	7	Total Zn 7 7	0	0

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C O 4 2 2	0	0
5	B	1	Total C O 4 2 2	0	0
5	C	1	Total C O 4 2 2	0	0
5	E	1	Total C O 4 2 2	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	31	Total O 31 31	0	0
6	B	12	Total O 12 12	0	0
6	C	12	Total O 12 12	0	0
6	D	22	Total O 22 22	0	0

Continued on next page...

Continued from previous page...

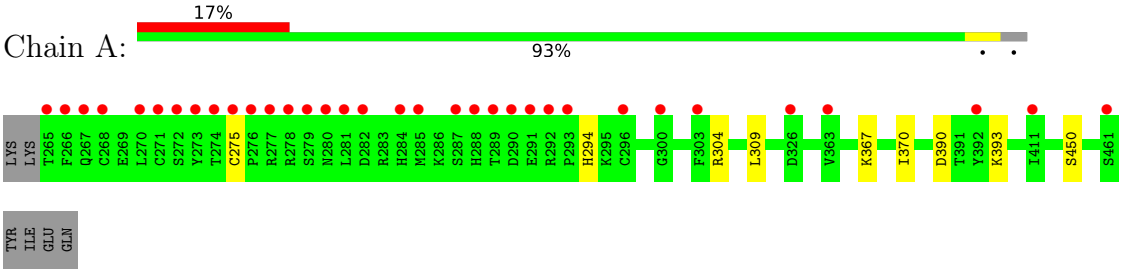
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	E	10	Total O 10 10	0	0
6	F	18	Total O 18 18	0	0

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.

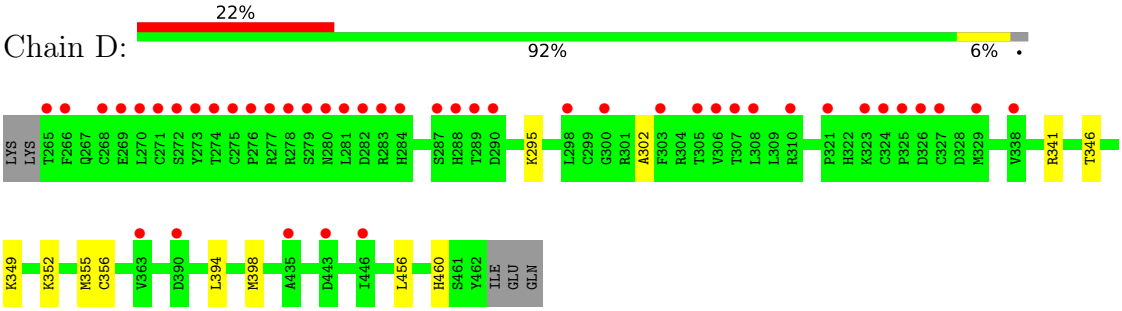
• Molecule 1: Transcriptional repressor CTCF

Chain A:



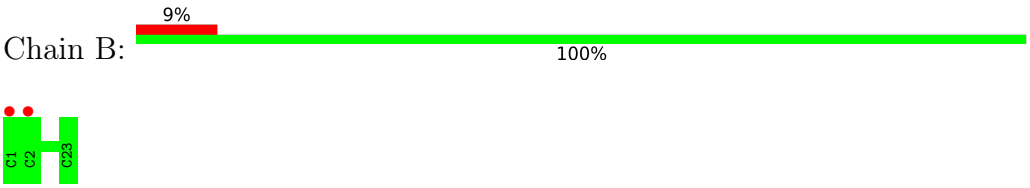
• Molecule 1: Transcriptional repressor CTCF

Chain D:



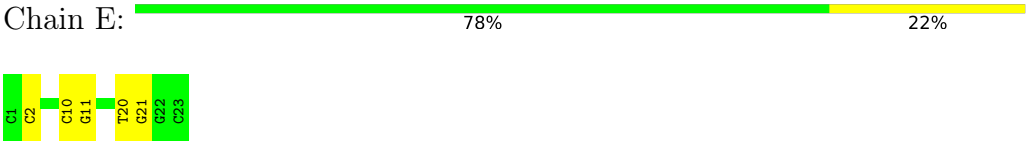
• Molecule 2: DNA Strand (23mer) I

Chain B:



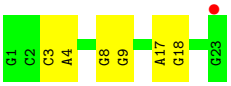
• Molecule 2: DNA Strand (23mer) I

Chain E:



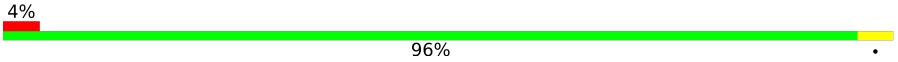
• Molecule 3: DNA Strand (23mer) II

Chain C:



• Molecule 3: DNA Strand (23mer) II

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	160.56Å 41.60Å 135.22Å 90.00° 105.61° 90.00°	Depositor
Resolution (Å)	40.17 – 2.19 40.17 – 2.19	Depositor EDS
% Data completeness (in resolution range)	96.7 (40.17-2.19) 96.7 (40.17-2.19)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.220 , 0.243 0.218 , 0.241	Depositor DCC
R_{free} test set	1991 reflections (4.60%)	wwPDB-VP
Wilson B-factor (Å^2)	50.7	Xtriage
Anisotropy	0.219	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 42.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5149	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.59% 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: EDO, 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	0.23	0/1601	0.50	0/2157
1	D	0.23	0/1624	0.50	0/2187
2	B	0.53	0/509	0.86	0/781
2	E	0.54	0/509	0.85	0/781
3	C	0.53	0/541	0.78	0/836
3	F	0.52	0/541	0.78	0/836
All	All	0.38	0/5325	0.65	0/7578

There are no bond length outliers.

There are no bond angle outliers.

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	1559	0	1471	6	0
1	D	1581	0	1495	7	0
2	B	457	0	259	0	0
2	E	457	0	259	3	0
3	C	480	0	257	4	0
3	F	480	0	257	1	0
4	A	7	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	7	0	0	0	0
5	B	8	0	12	1	0
5	C	4	0	6	0	0
5	E	4	0	6	0	0
6	A	31	0	0	0	0
6	B	12	0	0	0	0
6	C	12	0	0	0	0
6	D	22	0	0	0	0
6	E	10	0	0	0	0
6	F	18	0	0	0	0
All	All	5149	0	4022	17	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 2.

All (17) 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:D:341:ARG:O	1:D:346:THR:N	2.38	0.55
1:A:390:ASP:HB3	1:A:393:LYS:HG2	1.91	0.53
2:E:20:DT:H2"	2:E:21:DG:C8	2.44	0.53
1:A:450:SER:HB3	5:B:601:EDO:H12	1.90	0.52
1:D:349:LYS:O	1:D:352:LYS:NZ	2.34	0.52
2:E:2:DC:H42	3:F:22:DG:H1	1.57	0.51
1:A:370:ILE:HD13	1:D:356:CYS:HB3	1.94	0.50
3:C:17:DA:H2"	3:C:18:DG:H5"	1.94	0.49
1:D:295:LYS:HA	1:D:302:ALA:HA	1.96	0.47
3:C:8:DG:H2'	3:C:9:DG:C8	2.51	0.46
1:D:456:LEU:HD23	1:D:460:HIS:HD2	1.82	0.45
2:E:10:DC:H2'	2:E:11:DG:C8	2.53	0.43
1:A:367:LYS:HE2	1:D:355:MET:HA	2.00	0.43
3:C:3:DC:H2"	3:C:4:DA:C8	2.55	0.41
1:D:394:LEU:O	1:D:398:MET:HG2	2.21	0.41
1:A:304:ARG:NH1	3:C:18:DG:OP2	2.54	0.40
1:A:294:HIS:HB3	1:A:309:LEU:HD22	2.03	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	
1	A	195/203 (96%)	188 (96%)	7 (4%)	0	100	100
1	D	196/203 (97%)	192 (98%)	4 (2%)	0	100	100
All	All	391/406 (96%)	380 (97%)	11 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	169/189 (89%)	168 (99%)	1 (1%)	86	93
1	D	172/189 (91%)	172 (100%)	0	100	100
All	All	341/378 (90%)	340 (100%)	1 (0%)	92	97

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	275	CYS

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

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 18 ligands modelled in this entry, 14 are monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
5	EDO	B	602	-	3,3,3	0.46	0	2,2,2	0.31	0
5	EDO	B	601	-	3,3,3	0.47	0	2,2,2	0.31	0
5	EDO	E	101	-	3,3,3	0.47	0	2,2,2	0.34	0
5	EDO	C	201	-	3,3,3	0.46	0	2,2,2	0.32	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	EDO	B	602	-	-	0/1/1/1	-
5	EDO	B	601	-	-	0/1/1/1	-
5	EDO	E	101	-	-	0/1/1/1	-
5	EDO	C	201	-	-	0/1/1/1	-

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	601	EDO	1	0

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	197/203 (97%)	1.60	34 (17%) 1 1	42, 68, 194, 228	0
1	D	198/203 (97%)	1.94	44 (22%) 0 0	38, 76, 212, 234	0
2	B	23/23 (100%)	0.64	2 (8%) 10 8	46, 73, 184, 195	0
2	E	23/23 (100%)	0.50	0 100 100	49, 62, 147, 174	0
3	C	23/23 (100%)	0.66	1 (4%) 35 33	41, 55, 164, 237	0
3	F	23/23 (100%)	0.53	1 (4%) 35 33	44, 56, 172, 185	0
All	All	487/498 (97%)	1.55	82 (16%) 1 1	38, 70, 195, 237	0

All (82) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	276	PRO	19.0
1	A	271	CYS	17.2
1	D	265	THR	17.1
1	D	275	CYS	17.0
1	D	280	ASN	16.7
1	D	281	LEU	16.4
1	D	277	ARG	16.3
1	A	276	PRO	15.5
1	D	270	LEU	13.0
1	A	281	LEU	12.4
1	A	275	CYS	11.6
1	D	266	PHE	11.1
1	D	278	ARG	10.0
1	A	274	THR	10.0
1	A	277	ARG	9.9
1	A	266	PHE	9.8
1	A	268	CYS	8.8
1	D	269	GLU	8.8
1	D	273	TYR	8.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	285	MET	8.6
1	A	279	SER	7.7
1	A	280	ASN	7.4
1	A	270	LEU	7.4
1	D	271	CYS	7.3
1	D	274	THR	7.1
1	D	284	HIS	6.6
3	C	23	DG	6.0
1	A	289	THR	5.8
1	A	300	GLY	5.4
1	A	288	HIS	5.1
1	A	284	HIS	4.6
1	D	268	CYS	4.6
1	D	289	THR	4.5
1	D	325	PRO	4.5
1	D	279	SER	4.5
1	D	308	LEU	4.3
1	A	293	PRO	4.2
1	A	461	SER	4.2
1	D	310	ARG	4.2
1	D	287	SER	4.1
1	A	265	THR	4.0
1	A	287	SER	4.0
1	D	283	ARG	4.0
1	A	292	ARG	3.9
1	A	273	TYR	3.7
2	B	2	DC	3.6
1	A	303	PHE	3.6
1	D	282	ASP	3.5
1	D	323	LYS	3.4
1	D	306	VAL	3.4
1	A	278	ARG	3.3
1	A	272	SER	3.2
1	A	282	ASP	3.2
1	D	327	CYS	3.2
1	A	290	ASP	3.1
1	D	321	PRO	2.9
1	D	290	ASP	2.9
1	A	267	GLN	2.9
3	F	23	DG	2.8
1	D	363	VAL	2.6
1	D	435	ALA	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	291	GLU	2.6
1	D	272	SER	2.5
1	D	307	THR	2.5
1	D	324	CYS	2.4
1	D	303	PHE	2.4
1	D	288	HIS	2.4
1	A	326	ASP	2.4
1	A	392	TYR	2.3
1	D	326	ASP	2.3
1	D	300	GLY	2.3
1	D	443	ASP	2.3
1	D	305	THR	2.3
1	A	296	CYS	2.2
1	D	338	VAL	2.2
1	D	446	ILE	2.2
2	B	1	DC	2.2
1	A	363	VAL	2.2
1	D	329	MET	2.1
1	D	390	ASP	2.1
1	D	298	LEU	2.1
1	A	411	ILE	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	ZN	A	501	1/1	0.54	0.23	207,207,207,207	0
5	EDO	C	201	4/4	0.68	0.23	59,71,73,73	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	EDO	B	601	4/4	0.74	0.28	56,64,69,70	0
5	EDO	B	602	4/4	0.75	0.35	70,73,73,80	0
5	EDO	E	101	4/4	0.88	0.21	55,62,66,67	0
4	ZN	D	503	1/1	0.89	0.13	122,122,122,122	0
4	ZN	D	502	1/1	0.90	0.14	80,80,80,80	0
4	ZN	D	501	1/1	0.92	0.27	212,212,212,212	0
4	ZN	A	505	1/1	0.97	0.17	52,52,52,52	0
4	ZN	A	507	1/1	0.97	0.14	61,61,61,61	0
4	ZN	A	502	1/1	0.97	0.13	78,78,78,78	0
4	ZN	A	503	1/1	0.98	0.10	66,66,66,66	0
4	ZN	A	506	1/1	0.98	0.12	55,55,55,55	0
4	ZN	D	505	1/1	0.98	0.14	56,56,56,56	0
4	ZN	D	507	1/1	0.98	0.17	50,50,50,50	0
4	ZN	D	506	1/1	0.99	0.11	57,57,57,57	0
4	ZN	D	504	1/1	0.99	0.16	62,62,62,62	0
4	ZN	A	504	1/1	0.99	0.18	49,49,49,49	0

6.5 Other polymers (i)

There are no such residues in this entry.