wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 20, 2023 – 06:18 AM EDT

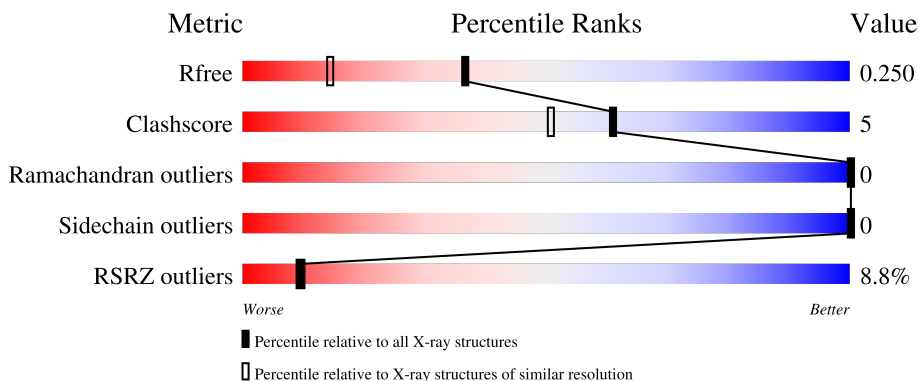|		
PDB ID	:	5KL4
Title	:	Wilms Tumor Protein (WT1) ZnF2-4 Q369H in complex with formylated DNA
Authors	:	Hashimoto, H.; Cheng, X.
Deposited on	:	2016-06-23
Resolution	:	1.78 Å(reported)

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 1.78 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	Percentile Ranks	Value
Rfree		0.250
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		8.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

❘ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	9185 (1.80-1.76)
Clashscore	141614	10184 (1.80-1.76)
Ramachandran outliers	138981	10051 (1.80-1.76)
Sidechain outliers	138945	10050 (1.80-1.76)
RSRZ outliers	127900	9032 (1.80-1.76)

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	93	12% — 88% / 9% ·
1	D	93	9% — 84% / 9% / 8%
2	B	11	73% / 27%
2	E	11	45% / 45% / 9%
3	C	11	27% / 64% / 9%

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	11	45% 45% 9%

2   Entry composition ⓘ

There are 8 unique types of molecules in this entry. The entry contains 4611 atoms, of which 1989 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 Wilms tumor protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	90	Total	C	H	N	O	S	0	0	0
			1510	468	740	165	130	7			
1	D	86	Total	C	H	N	O	S	0	0	0
			1454	452	714	158	123	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	345	GLY	-	expression tag	UNP P19544
A	346	PRO	-	expression tag	UNP P19544
A	347	LEU	-	expression tag	UNP P19544
A	348	GLY	-	expression tag	UNP P19544
A	349	SER	-	expression tag	UNP P19544
A	369	HIS	GLN	engineered mutation	UNP P19544
D	345	GLY	-	expression tag	UNP P19544
D	346	PRO	-	expression tag	UNP P19544
D	347	LEU	-	expression tag	UNP P19544
D	348	GLY	-	expression tag	UNP P19544
D	349	SER	-	expression tag	UNP P19544
D	369	HIS	GLN	engineered mutation	UNP P19544

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*GP*CP*GP*TP*GP*GP*GP*(5FC)P*GP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	11	Total	C	H	N	O	P	0	1	0
			368	113	132	45	68	10			
2	E	11	Total	C	H	N	O	P	0	1	0
			368	113	132	45	68	10			

- Molecule 3 is a DNA chain called DNA (5'-D(*TP*AP*(5CM)P*GP*CP*CP*CP*AP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	11	Total 344	C 105	H 126	N 40	O 63	P 10	0	0	0
3	F	11	Total 362	C 110	H 133	N 40	O 68	P 11	0	1	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total 3	Zn 3	0	0
4	D	3	Total 3	Zn 3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total 1	Ca 1	0	0
5	D	1	Total 1	Ca 1	0	0

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

O2
OH

C1

C2

HO

O1
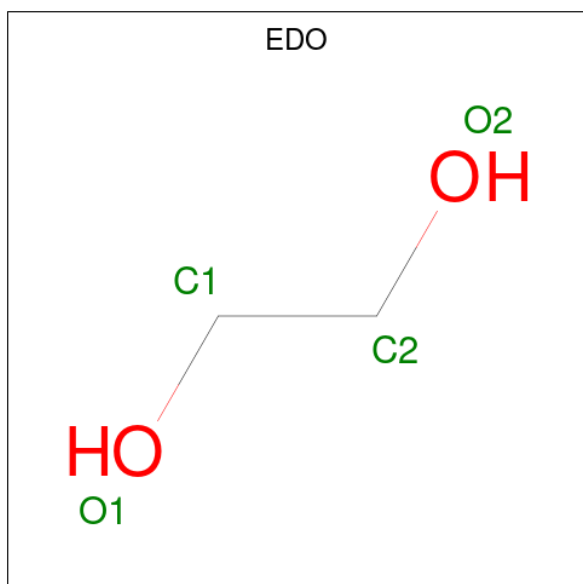
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total 10	C 2	H 6	O 2	0	0

Continued on next page...

W O R L D W I D E
PDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	D	1	Total 10	C 2	H 6	O 2	0	0

- Molecule 7 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	F	1	Total 1	Na 1	0	0

- Molecule 8 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	50	Total 50	O 50	0	0
8	B	18	Total 18	O 18	0	0
8	C	10	Total 10	O 10	0	0
8	D	54	Total 54	O 54	0	0
8	E	28	Total 29	O 29	0	1
8	F	15	Total 15	O 15	0	0

3 Residue-property plots ⓘ

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: Wilms tumor protein
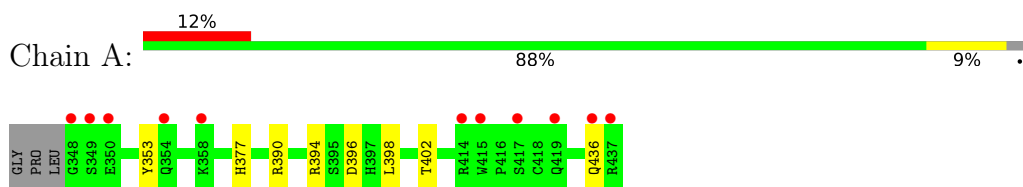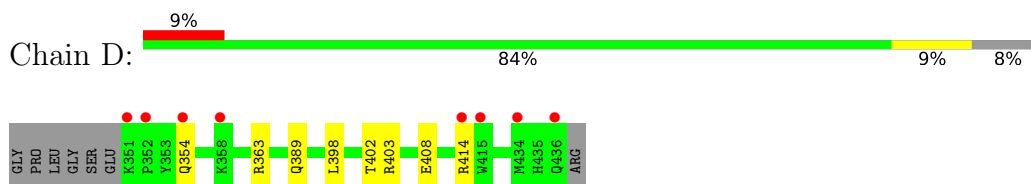
Chain A:

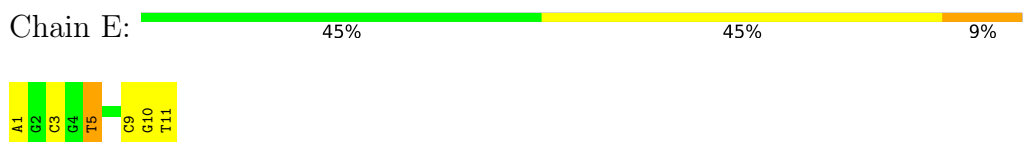- Molecule 1: Wilms tumor protein

Chain D:

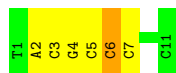- Molecule 2: DNA (5'-D(*AP*GP*CP*GP*TP*GP*GP*GP*(5FC)P*GP*T)-3')

Chain B:

- Molecule 2: DNA (5'-D(*AP*GP*CP*GP*TP*GP*GP*GP*(5FC)P*GP*T)-3')

Chain E:

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

Chain C:

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

Chain F:

45%　　　　　　　　　　　45%　　　　　　　9%

T1 A2 C3 G4 C5 C6 C7　C11
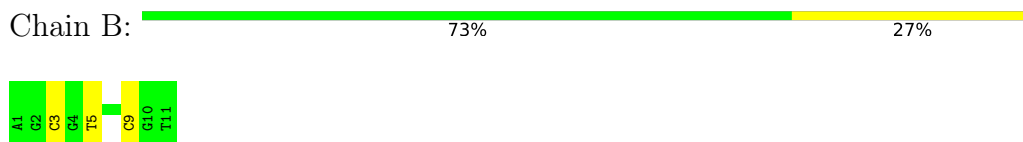
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	35.77Å 43.85Å 54.81Å 81.95° 88.57° 88.51°	Depositor
Resolution (Å)	29.56 – 1.78 29.56 – 1.78	Depositor EDS
% Data completeness (in resolution range)	92.2 (29.56-1.78) 86.7 (29.56-1.78)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 1.78Å)	Xtriage
Refinement program	PHENIX (dev_2257: ???)	Depositor
R, R_{free}	0.216 , 0.246 0.219 , 0.250	Depositor DCC
R_{free} test set	1458 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	23.8	Xtriage
Anisotropy	0.441	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 43.2	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.043 for h,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4611	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 22.96 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.1297e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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, NA, 5CM, 5FC, ZN, CA

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	0.50	0/791	0.65	0/1052
1	D	0.49	0/761	0.66	0/1013
2	B	1.37	0/259	1.13	2/398 (0.5%)
2	E	1.23	1/259 (0.4%)	1.10	2/398 (0.5%)
3	C	1.18	1/220 (0.5%)	1.17	2/334 (0.6%)
3	F	1.35	1/244 (0.4%)	1.30	2/371 (0.5%)
All	All	0.89	3/2534 (0.1%)	0.91	8/3566 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1	DT	C3'-O3'	-6.16	1.35	1.44
3	F	7	DC	C3'-O3'	-5.61	1.36	1.44
2	E	1	DA	N7-C5	-5.07	1.36	1.39

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	F	2	DA	O5'-P-OP2	-10.39	96.34	105.70
2	B	5	DT	O4'-C1'-N1	7.16	113.01	108.00
3	C	2	DA	O5'-P-OP2	-7.08	99.33	105.70
3	F	6	DC	O5'-P-OP2	-6.99	99.41	105.70
3	C	6	DC	O5'-P-OP2	-6.36	99.98	105.70

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	770	740	740	6	1
1	D	740	714	713	6	0
2	B	236	132	128	0	0
2	E	236	132	128	2	0
3	C	218	126	126	4	1
3	F	229	133	129	3	0
4	A	3	0	0	0	0
4	D	3	0	0	0	0
5	A	1	0	0	0	0
5	D	1	0	0	0	0
6	A	4	6	6	1	0
6	D	4	6	6	0	0
7	F	1	0	0	0	0
8	A	50	0	0	2	0
8	B	18	0	0	0	0
8	C	10	0	0	1	0
8	D	54	0	0	2	0
8	E	29	0	0	1	0
8	F	15	0	0	0	0
All	All	2622	1989	1976	20	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 5.

The worst 5 of 20 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:377:HIS:O	1:D:414:ARG:NH1	1.96	0.97
1:D:414:ARG:NH1	8:D:601:HOH:O	2.03	0.91
1:A:436:GLN:NE2	8:A:601:HOH:O	2.31	0.62
1:D:389:GLN:NE2	8:D:602:HOH:O	2.24	0.62
3:F:4:DG:H2'	3:F:5:DC:C6	2.39	0.57

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)
1:A:353:TYR:OH	3:C:11:DC:OP2[1_655]	2.17	0.03

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	
1	A	88/93 (95%)	87 (99%)	1 (1%)	0	100	100
1	D	84/93 (90%)	84 (100%)	0	0	100	100
All	All	172/186 (92%)	171 (99%)	1 (1%)	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	
1	A	86/88 (98%)	86 (100%)	0	100	100
1	D	83/88 (94%)	83 (100%)	0	100	100
All	All	169/176 (96%)	169 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

Mol	Chain	Res	Type
1	A	436	GLN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

5.4   Non-standard residues in protein, DNA, RNA chains ⓘ

4 non-standard protein/DNA/RNA residues are modelled in this entry.

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	#$|Z| > 2$
3	5CM	C	3	3,2	17,21,22	0.64	0	24,30,33	1.74	5 (20%)
2	5FC	B	9	3,2	19,22,23	1.26	2 (10%)	24,31,34	1.16	2 (8%)
3	5CM	F	3	3,2	17,21,22	0.58	0	24,30,33	1.63	3 (12%)
2	5FC	E	9	3,2	19,22,23	1.02	2 (10%)	24,31,34	1.35	2 (8%)

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
3	5CM	C	3	3,2	-	0/7/21/22	0/2/2/2
2	5FC	B	9	3,2	-	0/9/23/24	0/2/2/2
3	5CM	F	3	3,2	-	0/7/21/22	0/2/2/2
2	5FC	E	9	3,2	-	0/9/23/24	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	9	5FC	C6-N1	-2.91	1.33	1.38
2	B	9	5FC	O5'-C5'	-2.79	1.37	1.44
2	E	9	5FC	C6-N1	-2.76	1.33	1.38
2	E	9	5FC	C5-C4	2.12	1.49	1.44

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	3	5CM	O4'-C1'-N1	5.72	118.09	107.86
3	F	3	5CM	O4'-C1'-N1	5.51	117.71	107.86
2	E	9	5FC	C2'-C1'-N1	-3.88	104.84	113.77
3	F	3	5CM	C5A-C5-C6	-3.30	118.44	122.85
2	B	9	5FC	C2'-C1'-N1	-2.94	107.00	113.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 11 ligands modelled in this entry, 9 are monoatomic - leaving 2 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	$\#|Z| > 2$
6	EDO	A	505	-	3,3,3	0.28	0	2,2,2	0.60	0
6	EDO	D	505	-	3,3,3	0.25	0	2,2,2	0.60	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
6	EDO	A	505	-	-	1/1/1/1	-
6	EDO	D	505	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	505	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	505	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, 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(Å²)	Q<0.9
1	A	90/93 (96%)	0.83	11 (12%)	4	4	18, 29, 59, 79	0
1	D	86/93 (92%)	0.77	8 (9%)	8	8	18, 29, 51, 64	0
2	B	10/11 (90%)	-0.05	0	100	100	21, 27, 37, 43	0
2	E	10/11 (90%)	0.07	0	100	100	21, 27, 33, 37	0
3	C	10/11 (90%)	0.11	0	100	100	30, 35, 37, 39	0
3	F	10/11 (90%)	0.23	0	100	100	30, 32, 36, 36	0
All	All	216/230 (93%)	0.67	19 (8%)	10	9	18, 30, 54, 79	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	349	SER	5.1
1	A	348	GLY	5.1
1	A	437	ARG	4.5
1	A	417	SER	3.7
1	A	419	GLN	3.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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(Å²)	Q<0.9
2	5FC	B	9	21/22	0.89	0.13	36,39,48,49	0
3	5CM	F	3	20/21	0.89	0.11	23,33,47,48	0
3	5CM	C	3	20/21	0.92	0.11	29,39,53,55	0
2	5FC	E	9	21/22	0.92	0.12	32,37,44,45	0

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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(Å^2)	Q<0.9
6	EDO	A	505	4/4	0.92	0.16	33,39,41,42	0
6	EDO	D	505	4/4	0.93	0.19	25,30,33,35	0
5	CA	A	504	1/1	0.95	0.06	42,42,42,42	0
7	NA	F	101	1/1	0.96	0.08	54,54,54,54	0
5	CA	D	504	1/1	0.99	0.10	42,42,42,42	0
4	ZN	D	501	1/1	0.99	0.07	25,25,25,25	0
4	ZN	D	502	1/1	0.99	0.04	24,24,24,24	0
4	ZN	A	503	1/1	0.99	0.03	25,25,25,25	0
4	ZN	D	503	1/1	1.00	0.03	25,25,25,25	0
4	ZN	A	502	1/1	1.00	0.03	20,20,20,20	0
4	ZN	A	501	1/1	1.00	0.05	24,24,24,24	0

6.5 Other polymers ⓘ

There are no such residues in this entry.