



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 23, 2018 – 07:59 PM EDT

PDB ID : 6FE1
Title : Three dimensional structure of human carbonic anhydrase IX in complex with benzenesulfonamide.
Authors : Leitans, J.; Tars, K.
Deposited on : 2017-12-28
Resolution : 1.95 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

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
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : rb-20031172
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031172

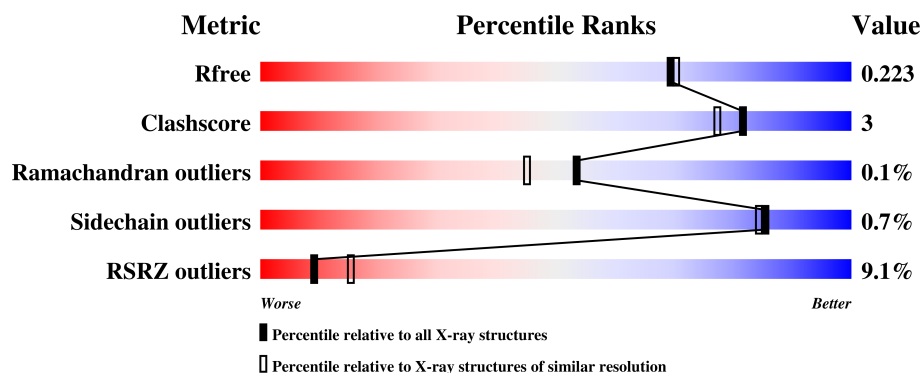
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.95 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}	111664	2220 (1.96-1.96)
Clashscore	122126	2333 (1.96-1.96)
Ramachandran outliers	120053	2314 (1.96-1.96)
Sidechain outliers	120020	2314 (1.96-1.96)
RSRZ outliers	108989	2174 (1.96-1.96)

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 ≥ 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	257	<div> <div>6%</div> <div> <div></div> <div>91%</div> <div>6% .</div> </div> </div>
1	B	257	<div> <div>8%</div> <div> <div></div> <div>88%</div> <div>5% . 5%</div> </div> </div>
1	C	257	<div> <div>10%</div> <div> <div></div> <div>93%</div> <div>. . .</div> </div> </div>
1	D	257	<div> <div>11%</div> <div> <div></div> <div>89%</div> <div>7% . .</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8588 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 Carbonic anhydrase 9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	251	Total	C	N	O	S	0	2	0
			1945	1241	343	357	4			
1	B	243	Total	C	N	O	S	0	5	0
			1904	1214	338	348	4			
1	C	251	Total	C	N	O	S	0	2	0
			1949	1243	346	356	4			
1	D	251	Total	C	N	O	S	0	2	0
			1938	1238	339	357	4			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q16790
A	0	PRO	-	expression tag	UNP Q16790
A	41	SER	CYS	engineered mutation	UNP Q16790
B	2	GLY	-	expression tag	UNP Q16790
B	3	PRO	-	expression tag	UNP Q16790
B	41	SER	CYS	engineered mutation	UNP Q16790
C	-1	GLY	-	expression tag	UNP Q16790
C	0	PRO	-	expression tag	UNP Q16790
C	41	SER	CYS	engineered mutation	UNP Q16790
D	-1	GLY	-	expression tag	UNP Q16790
D	0	PRO	-	expression tag	UNP Q16790
D	41	SER	CYS	engineered mutation	UNP Q16790

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

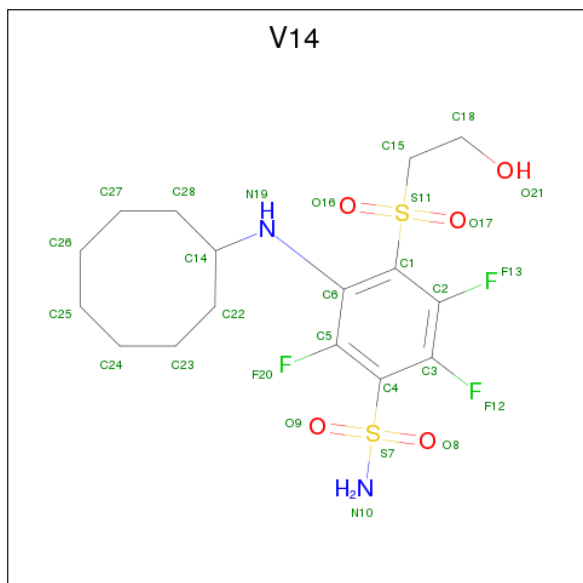
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Zn	0	0
			1	1		
2	C	1	Total	Zn	0	0
			1	1		

- Molecule 3 is 3-(cyclooctylamino)-2,5,6-trifluoro-4-[(2-hydroxyethyl)sulfonyl]benzenesulfonamide (three-letter code: V14) (formula: $C_{16}H_{23}F_3N_2O_5S_2$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total 28	C 16	F 3	N 2	O 5	S 2	0	0
3	B	1	Total 28	C 16	F 3	N 2	O 5	S 2	0	0
3	C	1	Total 28	C 16	F 3	N 2	O 5	S 2	0	0
3	D	1	Total 28	C 16	F 3	N 2	O 5	S 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	231	Total	O	0	0
			231	231		
4	B	180	Total	O	0	0
			180	180		
4	C	179	Total	O	0	0
			179	179		

Continued on next page...

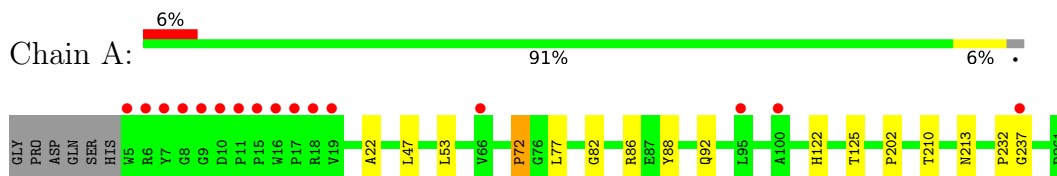
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	146	Total 146	O 146	0	0

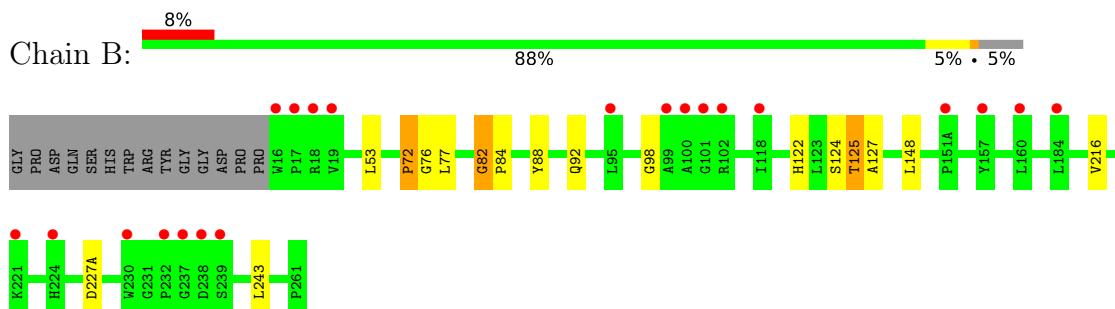
3 Residue-property plots [i](#)

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.

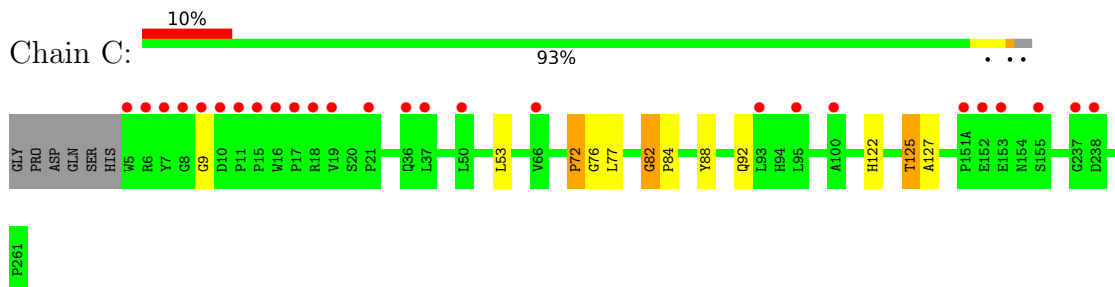
• Molecule 1: Carbonic anhydrase 9



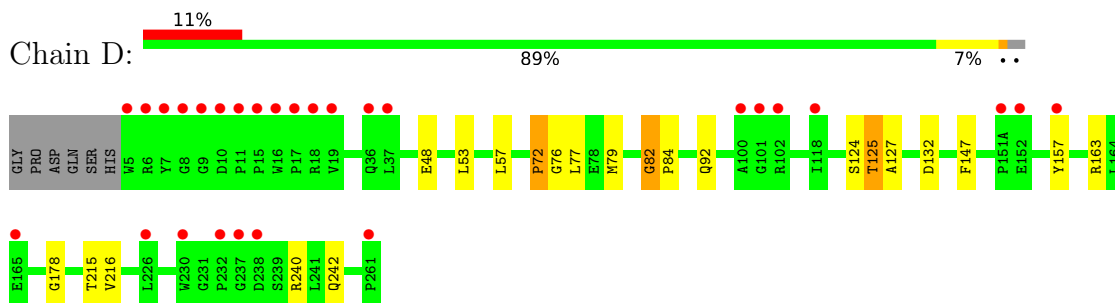
• Molecule 1: Carbonic anhydrase 9



• Molecule 1: Carbonic anhydrase 9



• Molecule 1: Carbonic anhydrase 9



4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	152.02Å 152.02Å 172.21Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.18 – 1.95 25.18 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.4 (25.18-1.95) 99.4 (25.18-1.95)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.00 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.182 , 0.216 0.190 , 0.223	Depositor DCC
R_{free} test set	5404 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	30.3	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 52.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.019 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	8588	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.75% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, V14

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	0.81	3/2007 (0.1%)	0.74	3/2739 (0.1%)
1	B	0.97	3/1973 (0.2%)	0.71	0/2691
1	C	1.07	3/2011 (0.1%)	0.73	3/2743 (0.1%)
1	D	0.98	3/2000 (0.1%)	0.71	2/2731 (0.1%)
All	All	0.96	12/7991 (0.2%)	0.72	8/10904 (0.1%)

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

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	72	PRO	C-N	27.45	1.82	1.33
1	C	82	GLY	C-N	26.66	1.84	1.34
1	D	72	PRO	C-N	23.68	1.75	1.33
1	D	82	GLY	C-N	23.11	1.78	1.34
1	B	82	GLY	C-N	22.96	1.77	1.34

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	72	PRO	O-C-N	-11.17	104.21	123.20
1	A	72	PRO	O-C-N	-8.82	108.21	123.20
1	C	72	PRO	CA-C-N	-8.14	99.91	116.20
1	C	125	THR	O-C-N	-8.09	109.76	122.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	132	ASP	CB-CG-OD1	5.50	123.25	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	72	PRO	Mainchain

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	1945	0	1894	5	0
1	B	1904	0	1867	13	0
1	C	1949	0	1898	9	0
1	D	1938	0	1879	14	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	28	0	23	1	0
3	B	28	0	23	1	0
3	C	28	0	23	0	0
3	D	28	0	23	1	0
4	A	231	0	0	0	0
4	B	180	0	0	0	2
4	C	179	0	0	0	0
4	D	146	0	0	0	0
All	All	8588	0	7630	44	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 3.

The worst 5 of 44 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:B:125:THR:C	1:B:127:ALA:N	1.71	1.44
1:B:72:PRO:C	1:B:76:GLY:N	1.72	1.41
1:D:72:PRO:C	1:D:76:GLY:N	1.75	1.40
1:B:82:GLY:C	1:B:84:PRO:N	1.77	1.36
1:C:125:THR:C	1:C:127:ALA:N	1.77	1.36

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)
4:B:443:HOH:O	4:B:443:HOH:O[2_555]	0.71	1.49
4:B:439:HOH:O	4:B:439:HOH:O[3_555]	2.01	0.19

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	247/257 (96%)	241 (98%)	6 (2%)	0	100	100
1	B	244/257 (95%)	238 (98%)	6 (2%)	0	100	100
1	C	247/257 (96%)	241 (98%)	5 (2%)	1 (0%)	36	24
1	D	247/257 (96%)	239 (97%)	8 (3%)	0	100	100
All	All	985/1028 (96%)	959 (97%)	25 (2%)	1 (0%)	53	45

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	9	GLY

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	204/210 (97%)	202 (99%)	2 (1%)	78	76
1	B	202/210 (96%)	200 (99%)	2 (1%)	78	76
1	C	204/210 (97%)	203 (100%)	1 (0%)	90	89
1	D	203/210 (97%)	202 (100%)	1 (0%)	90	89
All	All	813/840 (97%)	807 (99%)	6 (1%)	85	84

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	92	GLN
1	D	92	GLN
1	B	227(A)	ASP
1	A	213	ASN
1	C	92	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	213	ASN
1	C	213	ASN
1	C	64	HIS
1	A	64	HIS
1	C	159	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](#)

Of 8 ligands modelled in this entry, 4 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
3	V14	A	302	2	28,29,29	3.62	9 (32%)	34,43,43	2.28	9 (26%)
3	V14	B	302	2	28,29,29	4.10	9 (32%)	34,43,43	2.30	6 (17%)
3	V14	C	302	2	28,29,29	4.00	8 (28%)	34,43,43	2.65	11 (32%)
3	V14	D	302	2	28,29,29	4.04	8 (28%)	34,43,43	2.58	8 (23%)

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	V14	A	302	2	-	0/20/30/30	0/2/2/2
3	V14	B	302	2	-	0/20/30/30	0/2/2/2
3	V14	C	302	2	-	0/20/30/30	0/2/2/2
3	V14	D	302	2	-	0/20/30/30	0/2/2/2

The worst 5 of 34 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	302	V14	C4-S7	-10.37	1.61	1.79
3	C	302	V14	C4-S7	-10.00	1.61	1.79
3	B	302	V14	C4-S7	-9.58	1.62	1.79
3	A	302	V14	C4-S7	-8.85	1.63	1.79
3	B	302	V14	C15-S11	-4.95	1.67	1.78

The worst 5 of 34 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	302	V14	O16-S11-O17	-10.09	105.86	118.40
3	D	302	V14	O16-S11-O17	-9.92	106.07	118.40
3	B	302	V14	O16-S11-O17	-8.83	107.44	118.40
3	A	302	V14	O16-S11-O17	-8.03	108.42	118.40
3	D	302	V14	O8-S7-O9	-6.35	107.68	118.71

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	302	V14	1	0
3	B	302	V14	1	0
3	D	302	V14	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	D	5
1	C	5
1	B	4
1	A	4

The worst 5 of 18 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	11:PRO	C	15:PRO	N	3.28
1	D	232:PRO	C	237:GLY	N	3.18
1	C	11:PRO	C	15:PRO	N	3.07
1	C	232:PRO	C	237:GLY	N	3.02
1	D	11:PRO	C	15:PRO	N	2.97

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, 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	251/257 (97%)	0.28	16 (6%)	19 28	18, 28, 63, 147	0
1	B	243/257 (94%)	0.24	21 (8%)	10 17	22, 32, 62, 103	0
1	C	251/257 (97%)	0.56	26 (10%)	6 11	25, 38, 72, 146	0
1	D	251/257 (97%)	0.57	28 (11%)	5 8	24, 41, 84, 130	0
All	All	996/1028 (96%)	0.41	91 (9%)	9 15	18, 34, 72, 147	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	9	GLY	13.1
1	A	11	PRO	12.4
1	A	9	GLY	11.8
1	C	11	PRO	11.4
1	A	8	GLY	11.1

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. 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(\AA^2)	Q<0.9
3	V14	C	302	28/28	0.95	0.12	28,43,55,56	0
3	V14	D	302	28/28	0.95	0.12	37,62,69,73	0
3	V14	B	302	28/28	0.95	0.12	24,50,60,63	0
3	V14	A	302	28/28	0.97	0.10	21,36,46,52	0
2	ZN	D	301	1/1	0.99	0.07	32,32,32,32	0
2	ZN	C	301	1/1	0.99	0.08	28,28,28,28	0
2	ZN	A	301	1/1	1.00	0.07	21,21,21,21	0
2	ZN	B	301	1/1	1.00	0.09	28,28,28,28	0

6.5 Other polymers [i](#)

There are no such residues in this entry.