



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 05:05 pm GMT

PDB ID : 5W0R  
Title : Crystal structure of MBP fused activation-induced cytidine deaminase (AID)  
in complex with cacodylic acid  
Authors : Qiao, Q.; Wang, L.; Wu, H.  
Deposited on : 2017-05-31  
Resolution : 2.40 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

<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 : trunk30967  
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) : trunk30967

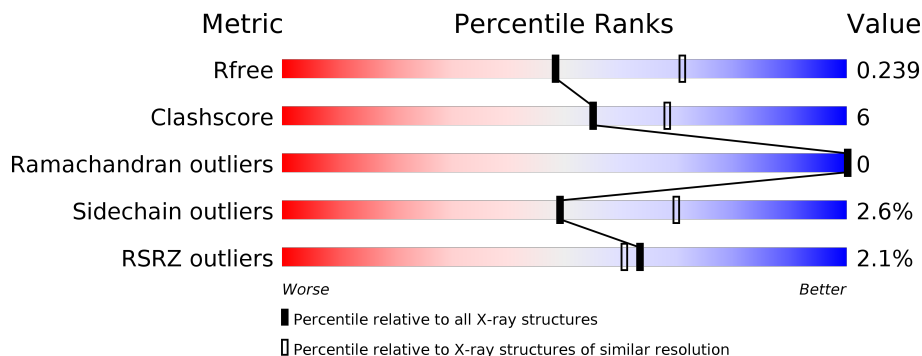
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3481 (2.40-2.40)
Clashscore	122126	3956 (2.40-2.40)
Ramachandran outliers	120053	3897 (2.40-2.40)
Sidechain outliers	120020	3898 (2.40-2.40)
RSRZ outliers	108989	3386 (2.40-2.40)

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	549	<div> <div>3%</div> <div> <div></div> <div>88%</div> <div>11%</div> <div>.</div> </div> </div>
1	B	549	<div> <div>%</div> <div> <div></div> <div>85%</div> <div>12%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8856 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 MBP fused activation-induced cytidine deaminase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	542	Total	C	N	O	S	0	0	0
			4298	2762	726	795	15			
1	A	548	Total	C	N	O	S	0	3	0
			4365	2801	737	812	15			

There are 58 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP P0AEY0
B	83	ALA	ASP	engineered mutation	UNP P0AEY0
B	84	ALA	LYS	engineered mutation	UNP P0AEY0
B	173	ALA	GLU	engineered mutation	UNP P0AEY0
B	174	ALA	ASN	engineered mutation	UNP P0AEY0
B	240	ALA	LYS	engineered mutation	UNP P0AEY0
B	360	ALA	GLU	engineered mutation	UNP P0AEY0
B	363	ALA	LYS	engineered mutation	UNP P0AEY0
B	364	ALA	ASP	engineered mutation	UNP P0AEY0
B	368	ASN	-	linker	UNP P0AEY0
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0
B	371	ALA	-	linker	UNP P0AEY0
B	372	GLU	-	linker	UNP P0AEY0
B	373	PHE	-	linker	UNP P0AEY0
B	1006	MET	-	linker	UNP P0AEY0
B	1007	ASP	-	linker	UNP P0AEY0
B	1008	PRO	-	linker	UNP P0AEY0
B	1009	ALA	-	linker	UNP P0AEY0
B	1010	THR	-	linker	UNP P0AEY0
B	1011	PHE	-	linker	UNP P0AEY0
B	1012	THR	-	linker	UNP P0AEY0
B	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1058	ALA	GLU	engineered mutation	UNP Q9GZX7
B	1130	ALA	HIS	engineered mutation	UNP Q9GZX7

*Continued on next page...*

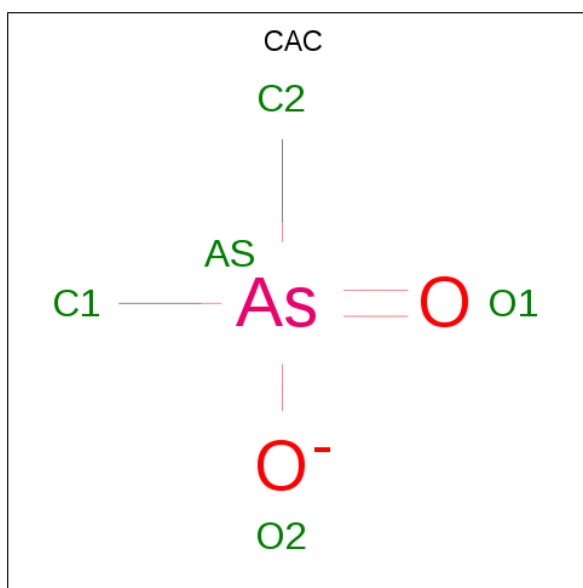
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	1131	GLU	ARG	engineered mutation	UNP Q9GZX7
B	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
B	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1181	GLN	LEU	engineered mutation	UNP Q9GZX7
A	1	MET	-	initiating methionine	UNP P0AEY0
A	83	ALA	ASP	engineered mutation	UNP P0AEY0
A	84	ALA	LYS	engineered mutation	UNP P0AEY0
A	173	ALA	GLU	engineered mutation	UNP P0AEY0
A	174	ALA	ASN	engineered mutation	UNP P0AEY0
A	240	ALA	LYS	engineered mutation	UNP P0AEY0
A	360	ALA	GLU	engineered mutation	UNP P0AEY0
A	363	ALA	LYS	engineered mutation	UNP P0AEY0
A	364	ALA	ASP	engineered mutation	UNP P0AEY0
A	368	ASN	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
A	371	ALA	-	linker	UNP P0AEY0
A	372	GLU	-	linker	UNP P0AEY0
A	373	PHE	-	linker	UNP P0AEY0
A	1006	MET	-	linker	UNP P0AEY0
A	1007	ASP	-	linker	UNP P0AEY0
A	1008	PRO	-	linker	UNP P0AEY0
A	1009	ALA	-	linker	UNP P0AEY0
A	1010	THR	-	linker	UNP P0AEY0
A	1011	PHE	-	linker	UNP P0AEY0
A	1012	THR	-	linker	UNP P0AEY0
A	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1058	ALA	GLU	engineered mutation	UNP Q9GZX7
A	1130	ALA	HIS	engineered mutation	UNP Q9GZX7
A	1131	GLU	ARG	engineered mutation	UNP Q9GZX7
A	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
A	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1181	GLN	LEU	engineered mutation	UNP Q9GZX7

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

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

- Molecule 3 is CACODYLATE ION (three-letter code: CAC) (formula:  $C_2H_6AsO_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	As	C	O	0	0
			5	1	2	2		
3	A	1	Total	As	C	O	0	0
			5	1	2	2		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Ca	0	0
			1	1		
4	A	1	Total	Ca	0	0
			1	1		

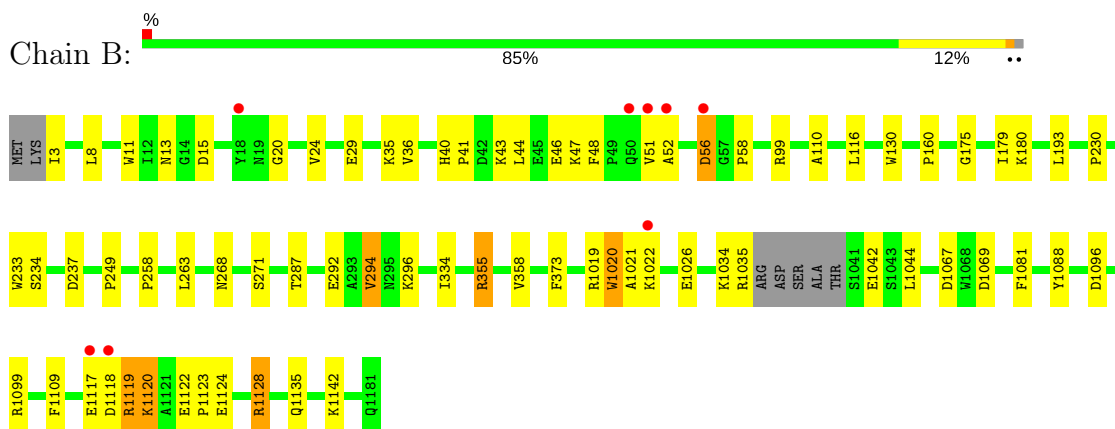
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	105	Total	O	0	0
			105	105		
5	A	74	Total	O	0	0
			74	74		

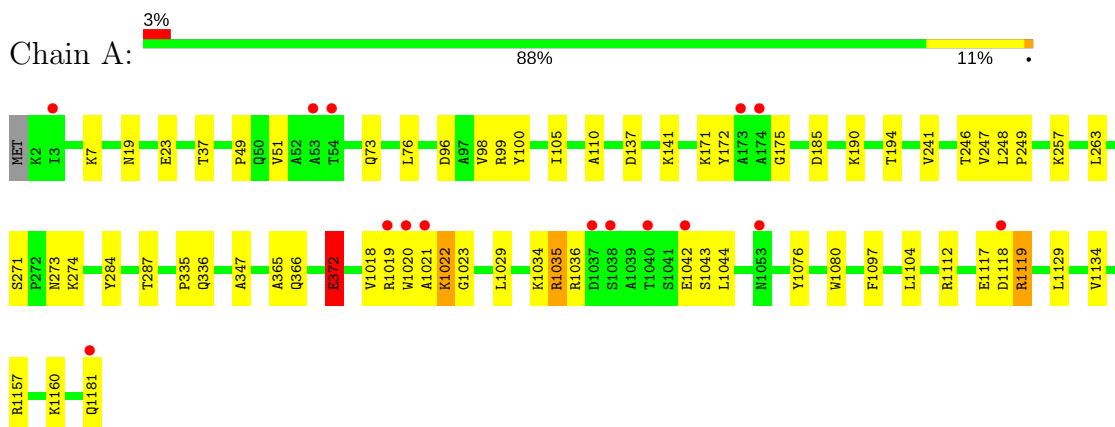
### 3 Residue-property plots [i](#)

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: MBP fused activation-induced cytidine deaminase



- Molecule 1: MBP fused activation-induced cytidine deaminase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	36.96Å 179.07Å 188.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.81 – 2.40 129.69 – 2.29	Depositor EDS
% Data completeness (in resolution range)	97.4 (24.81-2.40) 88.0 (129.69-2.29)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.18 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.219 , 0.239 0.224 , 0.239	Depositor DCC
$R_{free}$ test set	2572 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.6	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 31.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.007 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	8856	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.63% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CAC, 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	0.36	0/4483	0.53	2/6083 (0.0%)
1	B	0.43	0/4409	0.52	1/5983 (0.0%)
All	All	0.39	0/8892	0.52	3/12066 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1122	GLU	C-N-CD	5.98	140.95	128.40
1	A	372[A]	GLU	CA-C-O	5.16	130.94	120.10
1	A	372[B]	GLU	CA-C-O	5.16	130.94	120.10

There are no chirality outliers.

There are no planarity outliers.

### 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	4365	0	4275	44	0
1	B	4298	0	4210	52	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	74	0	0	1	0
5	B	105	0	0	0	0
All	All	8856	0	8485	96	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 6.

The worst 5 of 96 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:1118:ASP:O	1:A:1119:ARG:HG2	1.19	1.34
1:B:355:ARG:NH1	1:B:1069:ASP:OD2	1.64	1.28
1:B:1035:ARG:HD2	1:B:1042:GLU:OE2	1.10	1.25
1:A:1118:ASP:O	1:A:1119:ARG:CG	1.91	1.17
1:B:1035:ARG:CD	1:B:1042:GLU:OE2	1.95	1.15

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	549/549 (100%)	530 (96%)	19 (4%)	0	100	100
1	B	538/549 (98%)	522 (97%)	16 (3%)	0	100	100
All	All	1087/1098 (99%)	1052 (97%)	35 (3%)	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	447/445 (100%)	437 (98%)	10 (2%)	55	74
1	B	439/445 (99%)	425 (97%)	14 (3%)	42	62
All	All	886/890 (100%)	862 (97%)	24 (3%)	49	68

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1124	GLU
1	B	1142	LYS
1	A	1119	ARG
1	B	1128	ARG
1	B	1135	GLN

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

Mol	Chain	Res	Type
1	A	73	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 [i](#)

Of 6 ligands modelled in this entry, 4 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	$\# Z  > 2$
3	CAC	A	2002	2	0,4,4	0.00	-	0,6,6	0.00	-
3	CAC	B	2002	2	0,4,4	0.00	-	0,6,6	0.00	-

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	CAC	A	2002	2	-	0/0/0/0	0/0/0/0
3	CAC	B	2002	2	-	0/0/0/0	0/0/0/0

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 [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	548/549 (99%)	0.19	15 (2%) 54 52	26, 48, 81, 115	0
1	B	542/549 (98%)	-0.01	8 (1%) 73 71	22, 41, 73, 101	0
All	All	1090/1098 (99%)	0.09	23 (2%) 63 60	22, 46, 77, 115	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	173	ALA	7.2
1	A	174	ALA	5.4
1	A	53	ALA	5.2
1	A	1038	SER	4.9
1	A	1037	ASP	4.8

### 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	CA	A	2003	1/1	0.88	0.17	77,77,77,77	0
4	CA	B	2003	1/1	0.95	0.10	58,58,58,58	0
2	ZN	A	2001	1/1	0.98	0.12	42,42,42,42	0
2	ZN	B	2001	1/1	0.98	0.13	36,36,36,36	0
3	CAC	A	2002	5/5	0.99	0.15	33,33,53,58	0
3	CAC	B	2002	5/5	0.99	0.17	26,29,39,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.