



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 12:03 am GMT

PDB ID : 2Q58  
Title : Cryptosporidium parvum putative polyprenyl pyrophosphate synthase (cgd4\_2550) in complex with zoledronate  
Authors : Chruszcz, M.; Artz, J.; Zheng, H.; Dong, A.; Dunford, J.; Lew, J.; Zhao, Y.; Kozieradski, I.; Kavanaugh, K.L.; Opperman, U.; Sundstrom, M.; Weigelt, J.; Edwards, A.; Arrowsmith, C.; Bochkarev, A.; Hui, R.; Minor, W.; Structural Genomics Consortium (SGC)  
Deposited on : 2007-05-31  
Resolution : 2.37 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

<http://wwpdb.org/validation/2016/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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

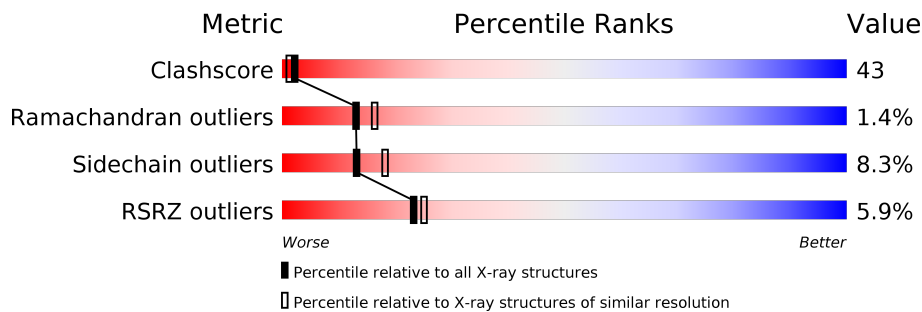
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.37 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))
Clashscore	112137	4984 (2.40-2.36)
Ramachandran outliers	110173	4907 (2.40-2.36)
Sidechain outliers	110143	4909 (2.40-2.36)
RSRZ outliers	101464	4423 (2.40-2.36)

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	368	<div> <div>6%</div> <div>38% 45% 8% 9%</div> </div>
1	B	368	<div> <div>4%</div> <div>31% 52% 6% 11%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5370 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 Farnesyl pyrophosphate synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	335	Total	C	N	O	S	0	0	0
			2655	1733	424	487	11			
1	B	329	Total	C	N	O	S	0	0	0
			2648	1723	422	492	11			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1001	MET	-	CLONING ARTIFACT	UNP Q5CR09
A	1002	GLY	-	CLONING ARTIFACT	UNP Q5CR09
A	1003	SER	-	CLONING ARTIFACT	UNP Q5CR09
A	1004	SER	-	CLONING ARTIFACT	UNP Q5CR09
A	1005	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1006	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1007	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1008	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1009	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1010	HIS	-	EXPRESSION TAG	UNP Q5CR09
A	1011	SER	-	CLONING ARTIFACT	UNP Q5CR09
A	1012	SER	-	CLONING ARTIFACT	UNP Q5CR09
A	1013	GLY	-	CLONING ARTIFACT	UNP Q5CR09
A	1014	ARG	-	CLONING ARTIFACT	UNP Q5CR09
A	1015	GLU	-	CLONING ARTIFACT	UNP Q5CR09
A	1016	ASN	-	CLONING ARTIFACT	UNP Q5CR09
A	1017	LEU	-	CLONING ARTIFACT	UNP Q5CR09
A	1018	TYR	-	CLONING ARTIFACT	UNP Q5CR09
A	1019	PHE	-	CLONING ARTIFACT	UNP Q5CR09
A	1020	GLN	-	CLONING ARTIFACT	UNP Q5CR09
A	1021	GLY	-	CLONING ARTIFACT	UNP Q5CR09
B	2001	MET	-	CLONING ARTIFACT	UNP Q5CR09
B	2002	GLY	-	CLONING ARTIFACT	UNP Q5CR09
B	2003	SER	-	CLONING ARTIFACT	UNP Q5CR09
B	2004	SER	-	CLONING ARTIFACT	UNP Q5CR09

*Continued on next page...*

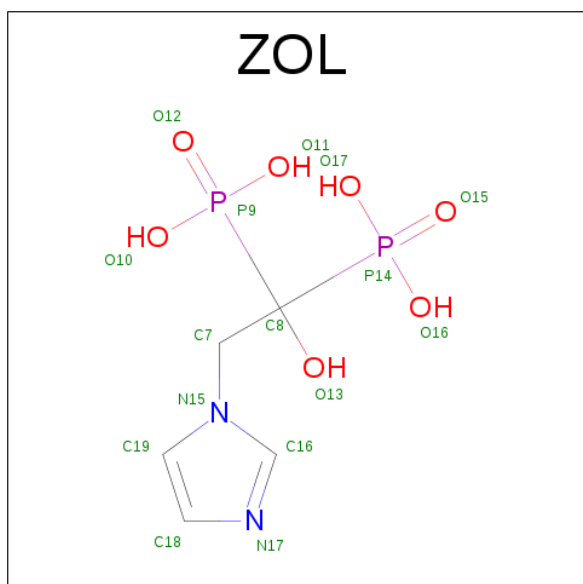
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	2005	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2006	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2007	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2008	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2009	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2010	HIS	-	EXPRESSION TAG	UNP Q5CR09
B	2011	SER	-	CLONING ARTIFACT	UNP Q5CR09
B	2012	SER	-	CLONING ARTIFACT	UNP Q5CR09
B	2013	GLY	-	CLONING ARTIFACT	UNP Q5CR09
B	2014	ARG	-	CLONING ARTIFACT	UNP Q5CR09
B	2015	GLU	-	CLONING ARTIFACT	UNP Q5CR09
B	2016	ASN	-	CLONING ARTIFACT	UNP Q5CR09
B	2017	LEU	-	CLONING ARTIFACT	UNP Q5CR09
B	2018	TYR	-	CLONING ARTIFACT	UNP Q5CR09
B	2019	PHE	-	CLONING ARTIFACT	UNP Q5CR09
B	2020	GLN	-	CLONING ARTIFACT	UNP Q5CR09
B	2021	GLY	-	CLONING ARTIFACT	UNP Q5CR09

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	3	Total Mg 3 3	0	0
2	A	3	Total Mg 3 3	0	0

- Molecule 3 is ZOLEDRONIC ACID (three-letter code: ZOL) (formula: C<sub>5</sub>H<sub>10</sub>N<sub>2</sub>O<sub>7</sub>P<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			16	5	2	7	2		
3	B	1	Total	C	N	O	P	0	0
			16	5	2	7	2		

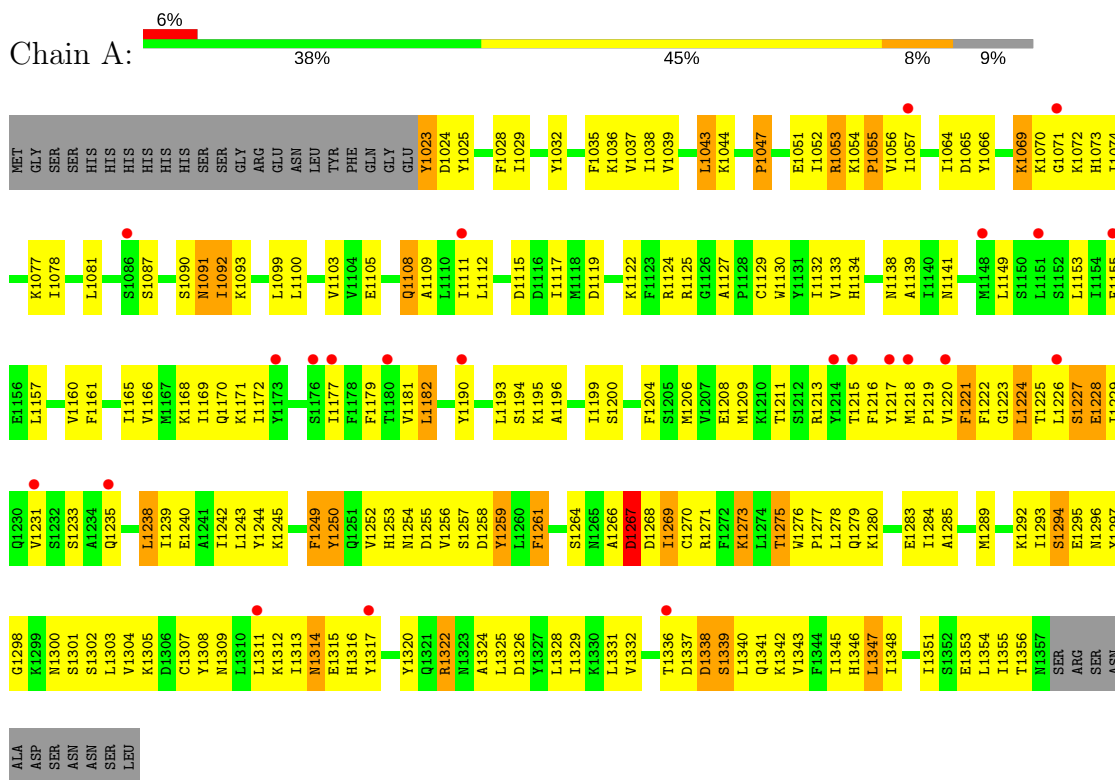
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	14	Total	O	0	0
			14	14		
4	B	15	Total	O	0	0
			15	15		

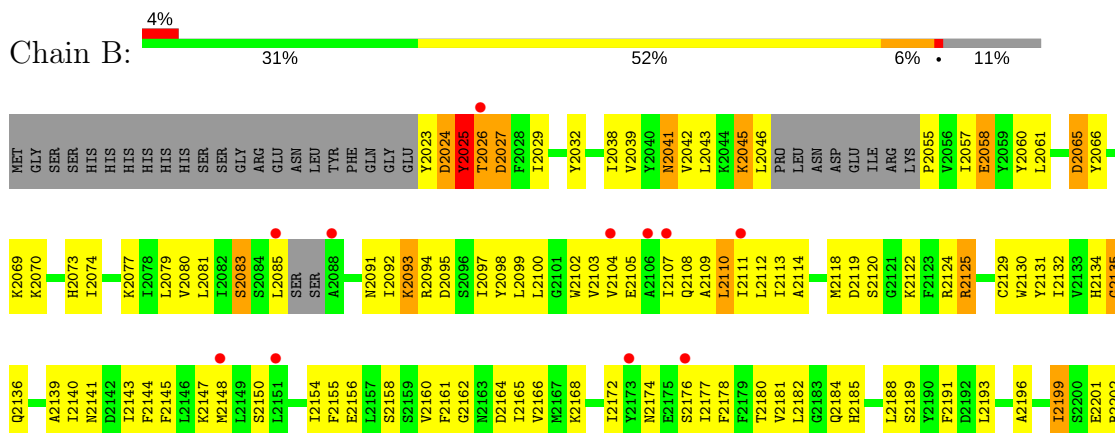
### 3 Residue-property plots

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: Farnesyl pyrophosphate synthase



#### • Molecule 1: Farnesyl pyrophosphate synthase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.78Å 101.78Å 75.28Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.46 – 2.37 30.46 – 2.37	Depositor EDS
% Data completeness (in resolution range)	94.8 (30.46-2.37) 94.8 (30.46-2.37)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.27 (at 2.36Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.216 , 0.238 0.209 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	50.7	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 118.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.009 for -h,-k,l 0.032 for h,-h-k,-l 0.013 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5370	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.87% 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: MG, ZOL

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	0.27	0/2713	0.57	0/3678
1	B	0.28	0/2701	0.61	1/3649 (0.0%)
All	All	0.27	0/5414	0.59	1/7327 (0.0%)

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	9
1	B	0	6
All	All	0	15

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	2065	ASP	CB-CG-OD1	5.26	123.03	118.30

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1047	PRO	Peptide
1	A	1053	ARG	Peptide
1	A	1055	PRO	Peptide
1	A	1091	ASN	Peptide
1	A	1221	PHE	Peptide

## 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	2655	0	2561	219	0
1	B	2648	0	2590	235	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	16	0	6	3	0
3	B	16	0	7	1	0
4	A	14	0	0	0	0
4	B	15	0	0	2	0
All	All	5370	0	5164	449	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 43.

The worst 5 of 449 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:2110:LEU:C	1:B:2110:LEU:HD23	1.54	1.26
1:A:1270:CYS:O	1:A:1298:GLY:HA2	1.37	1.18
1:B:2110:LEU:HD23	1:B:2111:ILE:N	1.60	1.16
1:B:2024:ASP:CA	1:B:2025:TYR:HB2	1.78	1.12
1:B:2024:ASP:HA	1:B:2025:TYR:HB2	1.37	1.07

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	333/368 (90%)	270 (81%)	58 (17%)	5 (2%)	12	15
1	B	323/368 (88%)	273 (84%)	46 (14%)	4 (1%)	15	20
All	All	656/736 (89%)	543 (83%)	104 (16%)	9 (1%)	13	16

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1047	PRO
1	A	1092	ILE
1	B	2025	TYR
1	A	1338	ASP
1	B	2237	ASN

### 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	279/341 (82%)	254 (91%)	25 (9%)	11	15
1	B	286/341 (84%)	264 (92%)	22 (8%)	15	21
All	All	565/682 (83%)	518 (92%)	47 (8%)	13	18

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1314	ASN
1	B	2027	ASP
1	B	2322	ARG
1	A	1339	SER
1	B	2041	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	2136	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	2341	GLN
1	B	2185	HIS
1	A	1262	ASN
1	B	2323	ASN

### 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, 6 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	ZOL	A	1	2	14,16,16	2.78	9 (64%)	19,26,26	1.49	4 (21%)
3	ZOL	B	2	2	14,16,16	2.81	9 (64%)	19,26,26	1.58	4 (21%)

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	ZOL	A	1	2	-	0/23/23/23	0/1/1/1
3	ZOL	B	2	2	-	0/23/23/23	0/1/1/1

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2	ZOL	P9-C8	-4.10	1.82	1.85
3	B	2	ZOL	C16-N17	-3.80	1.25	1.34
3	A	1	ZOL	C16-N17	-3.72	1.25	1.34
3	A	1	ZOL	P9-C8	-3.11	1.82	1.85
3	A	1	ZOL	C19-N15	-2.62	1.32	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1	ZOL	C19-C18-N17	-3.62	101.13	107.64
3	B	2	ZOL	C19-C18-N17	-3.59	101.19	107.64
3	A	1	ZOL	O16-P14-O15	-2.66	107.10	113.05
3	B	2	ZOL	P9-C8-P14	-2.58	107.92	112.70
3	B	2	ZOL	O16-P14-O15	-2.38	107.72	113.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1	ZOL	3	0
3	B	2	ZOL	1	0

## 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	335/368 (91%)	0.17	23 (6%) 18 18	24, 64, 113, 136	0
1	B	329/368 (89%)	0.04	16 (4%) 30 33	30, 59, 99, 157	0
All	All	664/736 (90%)	0.10	39 (5%) 23 25	24, 62, 107, 157	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1086	SER	5.2
1	B	2229	ILE	4.6
1	A	1311	LEU	4.2
1	B	2304	VAL	3.8
1	A	1217	TYR	3.5

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZOL	B	2	16/16	0.98	0.19	1.69	40,53,65,66	0
2	MG	B	7	1/1	0.95	0.14	0.19	33,33,33,33	0
3	ZOL	A	1	16/16	0.97	0.13	-0.18	40,54,70,86	0
2	MG	B	6	1/1	0.97	0.12	-0.60	33,33,33,33	0
2	MG	B	8	1/1	0.91	0.06	-3.72	30,30,30,30	0
2	MG	A	5	1/1	0.82	0.06	-	30,30,30,30	0
2	MG	A	3	1/1	0.97	0.05	-	33,33,33,33	0
2	MG	A	4	1/1	0.99	0.07	-	35,35,35,35	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.