



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:50 AM GMT

PDB ID : 3N2E  
Title : Crystal structure of Helicobacter pylori shikimate kinase in complex with NSC162535  
Authors : Cheng, W.C.; Chen, T.J.; Lin, S.C.; Wang, W.C.  
Deposited on : 2010-05-18  
Resolution : 2.53 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

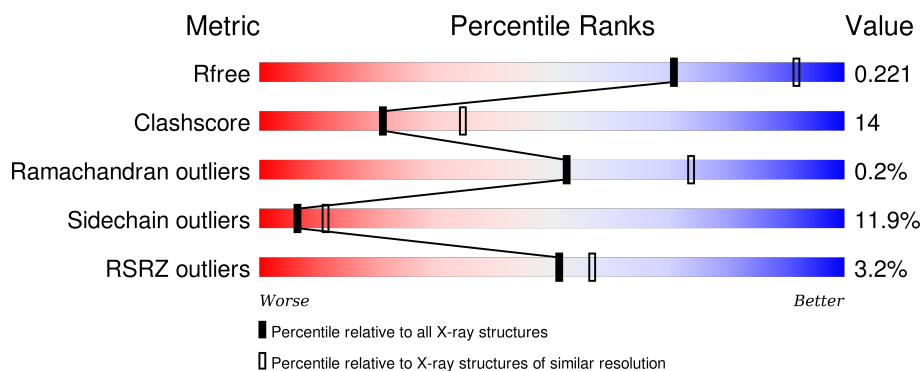
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.53 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))
$R_{free}$	91344	4241 (2.54-2.50)
Clashscore	102246	4968 (2.54-2.50)
Ramachandran outliers	100387	4873 (2.54-2.50)
Sidechain outliers	100360	4875 (2.54-2.50)
RSRZ outliers	91569	4253 (2.54-2.50)

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	168	<div> <div>73%</div> <div>23%</div> <div>.</div> </div>
1	B	168	<div> <div>68%</div> <div>22%</div> <div>5%</div> <div>.</div> </div>
1	C	168	<div> <div>7%</div> <div>52%</div> <div>24%</div> <div>.</div> <div>20%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	OSA	A	1001	-	-	X	-
3	TLA	A	2002	-	-	-	X
3	TLA	B	2003	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3860 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 Shikimate kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	168	Total	C	N	O	S	0	0	0
			1332	850	229	247	6			
1	B	162	Total	C	N	O	S	0	0	0
			1288	822	222	238	6			
1	C	134	Total	C	N	O	S	0	0	0
			1041	667	172	196	6			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	114	ALA	GLU	ENGINEERED MUTATION	UNP P56073
A	163	LEU	-	EXPRESSION TAG	UNP P56073
A	164	GLN	-	EXPRESSION TAG	UNP P56073
A	165	PRO	-	EXPRESSION TAG	UNP P56073
A	166	SER	-	EXPRESSION TAG	UNP P56073
A	167	LEU	-	EXPRESSION TAG	UNP P56073
A	168	SER	-	EXPRESSION TAG	UNP P56073
B	114	ALA	GLU	ENGINEERED MUTATION	UNP P56073
B	163	LEU	-	EXPRESSION TAG	UNP P56073
B	164	GLN	-	EXPRESSION TAG	UNP P56073
B	165	PRO	-	EXPRESSION TAG	UNP P56073
B	166	SER	-	EXPRESSION TAG	UNP P56073
B	167	LEU	-	EXPRESSION TAG	UNP P56073
B	168	SER	-	EXPRESSION TAG	UNP P56073
C	114	ALA	GLU	ENGINEERED MUTATION	UNP P56073
C	163	LEU	-	EXPRESSION TAG	UNP P56073
C	164	GLN	-	EXPRESSION TAG	UNP P56073
C	165	PRO	-	EXPRESSION TAG	UNP P56073
C	166	SER	-	EXPRESSION TAG	UNP P56073
C	167	LEU	-	EXPRESSION TAG	UNP P56073
C	168	SER	-	EXPRESSION TAG	UNP P56073

- Molecule 2 is 7-AMINO-4-HYDROXY-3-[(E)-(5-HYDROXY-7-SULFONAPHTHALEN-2-

The chemical structure is a complex molecule with 25 atoms labeled A through Z and 10 bonds labeled OBA through OCG. The structure features a central benzene ring (atoms A, B, C, D, E, F) with several substituents. A hydroxyl group (OH) is attached to atom A (bond OBA). A sulfonamide group is attached to atom B (bond OBB), consisting of a sulfur atom (S) double-bonded to an oxygen atom (O) and single-bonded to a nitrogen atom (N) and an oxygen atom (O). The nitrogen atom is further substituted with a hydroxyl group (OH) and a hydroxymethyl group (CH2OH). A hydroxyl group (OH) is attached to atom C (bond OBC). A hydroxyl group (OH) is attached to atom D (bond OBE). A hydroxyl group (OH) is attached to atom E (bond OBF). A hydroxyl group (OH) is attached to atom F (bond OBG). A hydroxyl group (OH) is attached to atom G (bond OCA). A hydroxyl group (OH) is attached to atom H (bond OCB). A hydroxyl group (OH) is attached to atom I (bond OCF). A hydroxyl group (OH) is attached to atom J (bond OCG).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 33	C 20	N 3	O 8	S 2	0	0
2	B	1	Total 33	C 20	N 3	O 8	S 2	0	0

- TLA
- 
- Chemical structure of 2,3,4-trihydroxybutanoic acid (D-glucose) in a Fischer projection. The structure shows a four-carbon chain. C1 is at the top, double-bonded to an oxygen (O1) and single-bonded to a hydroxyl group (OH, O11). C2 has a hydroxyl group (OH, O2) on the right. C3 has a hydroxyl group (OH, O3) on the left. C4 is at the bottom, double-bonded to an oxygen (O4) and single-bonded to a hydroxyl group (OH, O41).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			10	4	6		
3	A	1	Total	C	O	0	0
			10	4	6		
3	B	1	Total	C	O	0	0
			10	4	6		
3	B	1	Total	C	O	0	0
			10	4	6		
3	B	1	Total	C	O	0	0
			10	4	6		

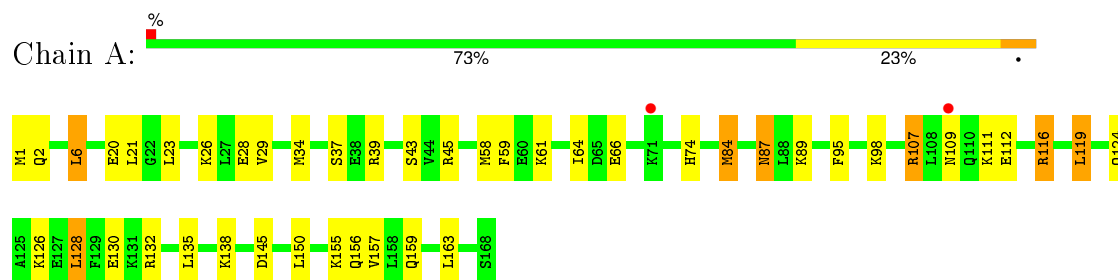
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	35	Total	O	0	0
			35	35		
4	B	32	Total	O	0	0
			32	32		
4	C	16	Total	O	0	0
			16	16		

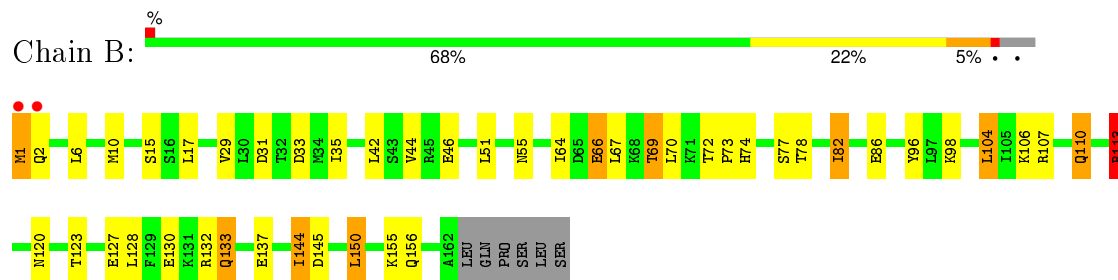
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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.

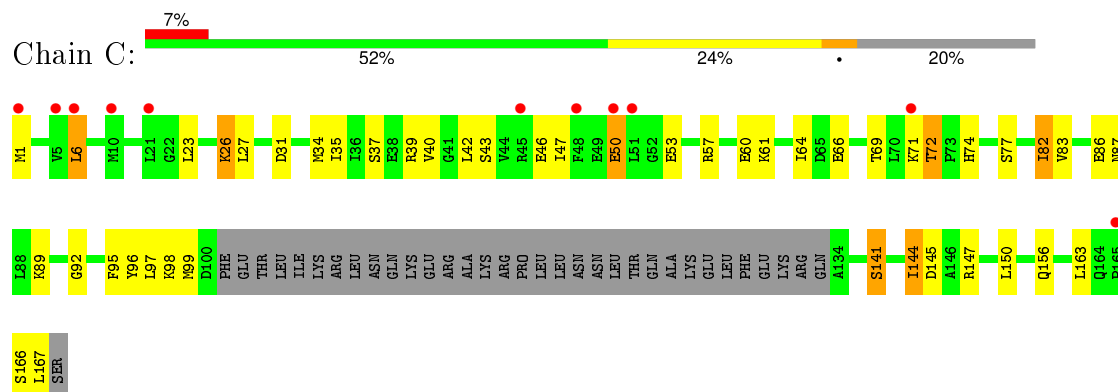
- Molecule 1: Shikimate kinase



- Molecule 1: Shikimate kinase



- Molecule 1: Shikimate kinase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	193.91Å 71.79Å 47.63Å 90.00° 91.96° 90.00°	Depositor
Resolution (Å)	30.00 – 2.53 28.84 – 2.53	Depositor EDS
% Data completeness (in resolution range)	99.5 (30.00-2.53) 99.5 (28.84-2.53)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	0.05	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.14 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.217 , 0.263 0.215 , 0.221	Depositor DCC
$R_{free}$ test set	1122 reflections (5.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	44.8	Xtriage
Anisotropy	0.587	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 39.4	EDS
Estimated twinning fraction	0.025 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 21908 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3860	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.44% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: OSA, TLA

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	1.09	0/1348	1.01	2/1805 (0.1%)
1	B	1.01	1/1303 (0.1%)	1.00	1/1743 (0.1%)
1	C	0.91	0/1053	0.98	1/1411 (0.1%)
All	All	1.01	1/3704 (0.0%)	1.00	4/4959 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	130	GLU	CG-CD	5.66	1.60	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	113	ARG	CG-CD-NE	5.93	124.26	111.80
1	C	163	LEU	CA-CB-CG	5.49	127.92	115.30
1	A	107	ARG	NE-CZ-NH1	5.32	122.96	120.30
1	A	116	ARG	NE-CZ-NH2	-5.22	117.69	120.30

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	1332	0	1392	29	0
1	B	1288	0	1345	40	0
1	C	1041	0	1077	32	0
2	A	33	0	14	12	0
2	B	33	0	14	4	0
3	A	20	0	8	5	0
3	B	30	0	12	5	0
4	A	35	0	0	2	0
4	B	32	0	0	4	0
4	C	16	0	0	5	0
All	All	3860	0	3862	109	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 14.

The worst 5 of 109 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:110:GLN:HE22	1:B:113:ARG:NH1	1.40	1.18
1:A:135:LEU:HD23	2:A:1001:OSA:HNAG	1.13	1.12
1:B:110:GLN:HE22	1:B:113:ARG:HH12	1.01	1.00
1:B:64:ILE:HD11	1:B:82:ILE:CD1	2.05	0.86
1:C:66:GLU:O	1:C:69:THR:HB	1.75	0.85

There are no symmetry-related clashes.

## 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	166/168 (99%)	159 (96%)	6 (4%)	1 (1%)	30 49
1	B	160/168 (95%)	155 (97%)	5 (3%)	0	100 100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	130/168 (77%)	127 (98%)	3 (2%)	0	100	100
All	All	456/504 (90%)	441 (97%)	14 (3%)	1 (0%)	52	74

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	87	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	147/147 (100%)	130 (88%)	17 (12%)	7	12
1	B	141/147 (96%)	123 (87%)	18 (13%)	5	9
1	C	115/147 (78%)	102 (89%)	13 (11%)	7	13
All	All	403/441 (91%)	355 (88%)	48 (12%)	6	11

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	66	GLU
1	B	106	LYS
1	C	89	LYS
1	B	69	THR
1	B	86	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	74	HIS
1	B	87	ASN
1	B	159	GLN
1	A	159	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	156	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 ⓘ

7 ligands 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$
2	OSA	A	1001	-	35,36,36	1.46	4 (11%)	49,56,56	1.84	7 (14%)
3	TLA	A	2001	-	3,9,9	1.56	0	6,12,12	0.81	0
3	TLA	A	2002	-	3,9,9	0.97	0	6,12,12	1.83	2 (33%)
2	OSA	B	1001	-	35,36,36	1.81	6 (17%)	49,56,56	1.85	9 (18%)
3	TLA	B	2001	-	3,9,9	0.80	0	6,12,12	1.20	1 (16%)
3	TLA	B	2002	-	3,9,9	0.56	0	6,12,12	2.09	2 (33%)
3	TLA	B	2003	-	3,9,9	0.62	0	6,12,12	0.91	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
2	OSA	A	1001	-	-	0/17/17/17	0/4/4/4
3	TLA	A	2001	-	-	0/4/12/12	0/0/0/0
3	TLA	A	2002	-	-	0/4/12/12	0/0/0/0
2	OSA	B	1001	-	-	0/17/17/17	0/4/4/4
3	TLA	B	2001	-	-	0/4/12/12	0/0/0/0
3	TLA	B	2002	-	-	0/4/12/12	0/0/0/0
3	TLA	B	2003	-	-	0/4/12/12	0/0/0/0

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1001	OSA	CAI-SBD	-5.02	1.70	1.78
2	A	1001	OSA	CAI-SBD	-3.60	1.73	1.78
2	B	1001	OSA	CAK-CAC	-2.93	1.37	1.43
2	A	1001	OSA	CAO-NAN	-2.53	1.33	1.44
2	B	1001	OSA	CAO-NAN	-2.29	1.34	1.44

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2002	TLA	C4-C3-C2	-3.90	105.35	113.35
2	B	1001	OSA	CAQ-CAR-CAX	-3.19	117.33	122.29
3	B	2002	TLA	C1-C2-C3	-3.12	106.95	113.35
3	A	2002	TLA	O2-C2-C1	-2.94	103.79	111.21
3	A	2002	TLA	C4-C3-C2	-2.87	107.46	113.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1001	OSA	12	0
3	A	2001	TLA	3	0
3	A	2002	TLA	2	0
2	B	1001	OSA	4	0
3	B	2001	TLA	1	0
3	B	2003	TLA	4	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	168/168 (100%)	-0.01	2 (1%) 81 84	25, 40, 57, 64	0
1	B	162/168 (96%)	0.02	2 (1%) 81 84	31, 41, 54, 64	0
1	C	134/168 (79%)	0.44	11 (8%) 14 16	30, 47, 68, 77	0
All	All	464/504 (92%)	0.13	15 (3%) 51 57	25, 42, 62, 77	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	MET	6.0
1	C	45	ARG	5.0
1	C	48	PHE	4.2
1	C	165	PRO	4.1
1	C	10	MET	3.7

### 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	TLA	A	2002	10/10	0.90	0.20	2.26	54,55,59,62	0
3	TLA	A	2001	10/10	0.78	0.19	1.65	63,65,68,68	0
3	TLA	B	2002	10/10	0.86	0.20	0.80	73,78,81,81	0
3	TLA	B	2003	10/10	0.97	0.14	0.14	53,55,57,60	0
2	OSA	A	1001	33/33	0.93	0.16	-0.09	44,49,82,84	0
3	TLA	B	2001	10/10	0.91	0.11	-0.21	54,57,63,64	0
2	OSA	B	1001	33/33	0.96	0.12	-0.65	29,36,58,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.