



# Full wwPDB X-ray Structure Validation Report ⓘ

Dec 12, 2023 – 04:26 pm GMT

PDB ID : 4A2S  
Title : Structure of the engineered retro-aldolase RA95.5  
Authors : Giger, L.; Caner, S.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D.  
Deposited on : 2011-09-28  
Resolution : 1.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

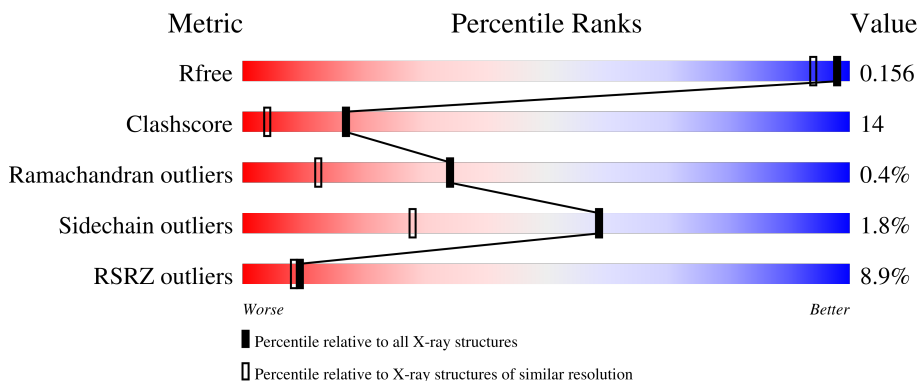
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.40 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))
$R_{free}$	130704	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	258	

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	3NK	A	1083[B]	-	-	X	-

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4730 atoms, of which 2285 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.

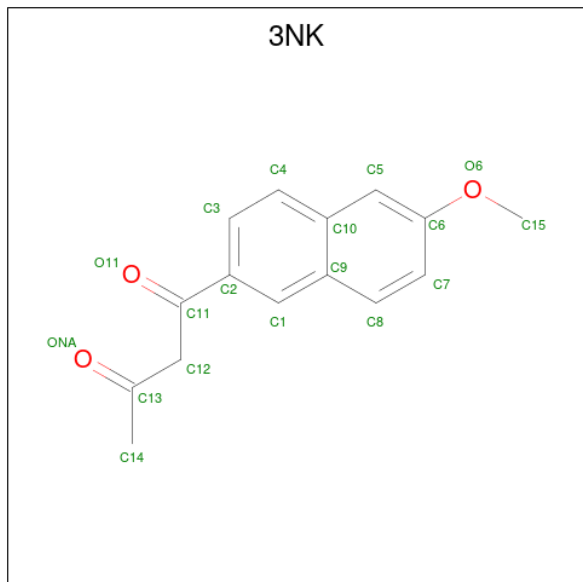
- Molecule 1 is a protein called INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	247	4429	1394	2243	370	413	9	0	31	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	GLU	LYS	engineered mutation	UNP Q06121
A	22	VAL	PHE	engineered mutation	UNP Q06121
A	51	TYR	GLU	engineered mutation	UNP Q06121
A	53	SER	LYS	engineered mutation	UNP Q06121
A	83	LYS	LEU	engineered mutation	UNP Q06121
A	110	SER	LYS	engineered mutation	UNP Q06121
A	159	LEU	GLU	engineered mutation	UNP Q06121
A	180	PHE	ASN	engineered mutation	UNP Q06121
A	182	MET	ARG	engineered mutation	UNP Q06121
A	183	ASN	ASP	engineered mutation	UNP Q06121
A	184	PHE	LEU	engineered mutation	UNP Q06121
A	210	LYS	GLU	engineered mutation	UNP Q06121
A	211	LEU	SER	engineered mutation	UNP Q06121
A	233	SER	GLY	engineered mutation	UNP Q06121
A	246	LEU	-	expression tag	UNP Q06121
A	247	ILE	-	expression tag	UNP Q06121
A	248	GLU	-	expression tag	UNP Q06121
A	249	GLY	-	expression tag	UNP Q06121
A	250	SER	-	expression tag	UNP Q06121
A	251	LEU	-	expression tag	UNP Q06121
A	252	GLU	-	expression tag	UNP Q06121
A	253	HIS	-	expression tag	UNP Q06121
A	254	HIS	-	expression tag	UNP Q06121
A	255	HIS	-	expression tag	UNP Q06121
A	256	HIS	-	expression tag	UNP Q06121
A	257	HIS	-	expression tag	UNP Q06121
A	258	HIS	-	expression tag	UNP Q06121

- Molecule 2 is 1-(6-METHOXYNAPHTHALEN-2-YL)BUTANE-1,3-DIONE (three-letter code: 3NK) (formula: C<sub>15</sub>H<sub>14</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
2	A	1	62	30	28	4	0	1
2	A	1	31	15	14	2	0	0

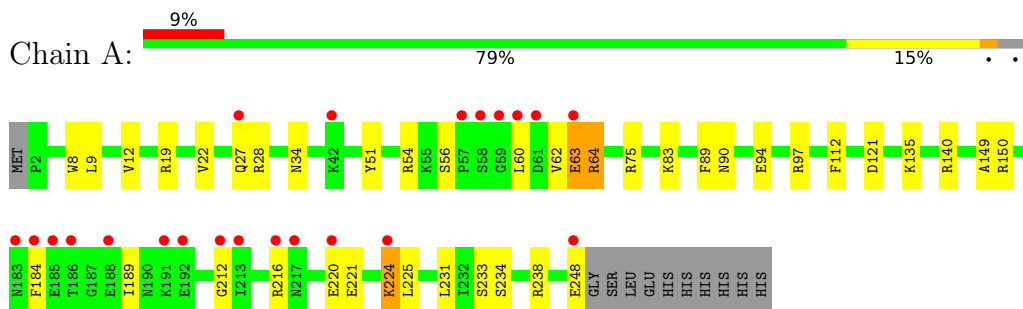
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	208	Total	O	0	0
			208	208		

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

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: INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	53.81Å 62.81Å 93.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 1.40 8.00 – 1.40	Depositor EDS
% Data completeness (in resolution range)	99.5 (8.00-1.40) 99.5 (8.00-1.40)	Depositor EDS
$R_{merge}$	0.02	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.92 (at 1.40Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.132 , 0.162 0.129 , 0.156	Depositor DCC
$R_{free}$ test set	1867 reflections (3.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.8	Xtrriage
Anisotropy	0.345	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.64 , 83.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	4730	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.75% 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: 3NK

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	0.68	0/2250	0.81	5/3028 (0.2%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	64	ARG	NE-CZ-NH1	11.41	126.01	120.30
1	A	64	ARG	NE-CZ-NH2	-10.22	115.19	120.30
1	A	140	ARG	NE-CZ-NH1	8.15	124.37	120.30
1	A	140	ARG	NE-CZ-NH2	-7.28	116.66	120.30
1	A	150	ARG	NE-CZ-NH2	-5.06	117.77	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	2186	2243	2222	63	0
2	A	51	42	42	17	0
3	A	208	0	0	12	0
All	All	2445	2285	2264	64	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.

All (64) 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:83[B]:LYS:HG3	3:A:2108:HOH:O	1.37	1.20
1:A:83[B]:LYS:HE3	3:A:2108:HOH:O	1.49	1.11
1:A:83[B]:LYS:CD	3:A:2108:HOH:O	2.06	1.04
1:A:83[B]:LYS:CE	3:A:2108:HOH:O	2.00	1.04
1:A:83[B]:LYS:HE2	1:A:89:PHE:CD2	2.00	0.97
1:A:83[B]:LYS:CG	3:A:2108:HOH:O	1.99	0.89
1:A:216:ARG:NH1	1:A:220:GLU:OE2	2.14	0.81
1:A:64:ARG:HD3	3:A:2090:HOH:O	1.85	0.76
1:A:51:TYR:OH	1:A:231[B]:LEU:HD21	1.88	0.73
1:A:51:TYR:CE2	1:A:231[B]:LEU:HG	2.24	0.72
1:A:83[B]:LYS:HE2	1:A:89:PHE:CE2	2.26	0.71
1:A:28:ARG:N	3:A:2047:HOH:O	2.25	0.70
1:A:83[B]:LYS:HE2	1:A:89:PHE:CG	2.27	0.70
1:A:27:GLN:C	3:A:2047:HOH:O	2.31	0.69
1:A:22[B]:VAL:HG12	1:A:121:ASP:OD1	1.95	0.66
1:A:51:TYR:HE2	1:A:231[B]:LEU:HG	1.61	0.66
1:A:184:PHE:CZ	2:A:1083[B]:3NK:C5	2.79	0.65
1:A:63:GLU:OE1	3:A:2082:HOH:O	2.14	0.65
1:A:12[B]:VAL:HG11	1:A:112:PHE:CD1	2.32	0.64
1:A:184:PHE:CE2	2:A:1083[B]:3NK:H5	2.33	0.64
1:A:221:GLU:HA	1:A:224:LYS:HD3	1.81	0.62
2:A:1083[A]:3NK:H121	2:A:1210:3NK:H3	1.81	0.61
1:A:34:ASN:OD1	3:A:2056:HOH:O	2.16	0.60
1:A:8:TRP:CH2	2:A:1083[B]:3NK:H3	2.36	0.60
1:A:184:PHE:HZ	2:A:1083[B]:3NK:C10	2.19	0.56
1:A:184:PHE:HZ	2:A:1083[B]:3NK:C5	2.17	0.56
1:A:22[B]:VAL:CG1	1:A:121:ASP:HA	2.36	0.56
1:A:54:ARG:NH1	3:A:2087:HOH:O	2.38	0.55
1:A:75:ARG:NH1	3:A:2103:HOH:O	2.36	0.55
1:A:224:LYS:HE2	1:A:225:LEU:HG	1.89	0.55
1:A:180:PHE:O	1:A:181[B]:SER:HB2	2.07	0.54
1:A:224:LYS:HG2	1:A:225:LEU:N	2.22	0.53
1:A:83[B]:LYS:HG2	1:A:89:PHE:CG	2.44	0.52
1:A:8:TRP:HH2	2:A:1083[B]:3NK:H3	1.75	0.52
1:A:234[B]:SER:O	1:A:238:ARG:HG2	2.09	0.52
1:A:83[B]:LYS:CE	1:A:89:PHE:CE2	2.95	0.49
1:A:56:SER:HB3	1:A:60:LEU:HB3	1.94	0.49
1:A:12[B]:VAL:HG11	1:A:112:PHE:CG	2.48	0.48

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:51:TYR:CE1	1:A:83[A]:LYS:HE2	2.49	0.48
1:A:51:TYR:CE2	1:A:231[B]:LEU:CG	2.95	0.47
1:A:184:PHE:CZ	2:A:1083[B]:3NK:C4	2.97	0.47
1:A:8:TRP:CZ2	2:A:1083[B]:3NK:H3	2.51	0.46
1:A:51:TYR:CD2	1:A:233[B]:SER:HB2	2.50	0.46
1:A:149:ALA:CB	1:A:156[B]:PRO:HG3	2.45	0.46
1:A:184:PHE:CE2	2:A:1083[B]:3NK:H4	2.51	0.46
1:A:189[A]:ILE:HD12	1:A:212:GLY:N	2.32	0.45
1:A:135:LYS:HG3	2:A:1083[A]:3NK:H151	1.98	0.45
1:A:180:PHE:O	1:A:181[B]:SER:CB	2.64	0.45
1:A:184:PHE:CZ	2:A:1083[B]:3NK:C10	3.00	0.44
1:A:54:ARG:CZ	1:A:90[A]:ASN:ND2	2.81	0.43
1:A:51:TYR:CD1	1:A:83[A]:LYS:HE2	2.54	0.43
1:A:184:PHE:CE2	2:A:1083[B]:3NK:C5	3.02	0.43
1:A:83[B]:LYS:NZ	1:A:89:PHE:CE2	2.85	0.43
1:A:51:TYR:HB2	1:A:233[B]:SER:HA	2.01	0.43
1:A:83[A]:LYS:NZ	2:A:1083[A]:3NK:O11	2.52	0.43
1:A:9[B]:LEU:HD11	2:A:1083[B]:3NK:H5	2.01	0.42
1:A:8:TRP:NE1	2:A:1210:3NK:H153	2.34	0.42
1:A:51:TYR:CD2	1:A:233[A]:SER:HB3	2.54	0.42
1:A:83[B]:LYS:HG2	1:A:89:PHE:CB	2.50	0.42
1:A:60:LEU:HD11	1:A:62:VAL:CG2	2.50	0.42
1:A:22[B]:VAL:HG12	1:A:121:ASP:CG	2.40	0.41
1:A:224:LYS:HG2	1:A:225:LEU:HG	2.02	0.41
1:A:184:PHE:HE2	2:A:1083[B]:3NK:H5	1.82	0.41
1:A:94:GLU:OE2	1:A:97:ARG:NE	2.32	0.40

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	276/258 (107%)	264 (96%)	10 (4%)	2 (1%)	22 5

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	181[A]	SER
1	A	181[B]	SER

### 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	254/233 (109%)	250 (98%)	4 (2%)	62 33

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	19	ARG
1	A	63	GLU
1	A	224	LYS
1	A	248	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

3 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	3NK	A	1210	1	18,18,19	0.94	1 (5%)	24,24,26	0.88	0
2	3NK	A	1083[A]	1	18,18,19	0.91	1 (5%)	24,24,26	0.76	0
2	3NK	A	1083[B]	1	18,18,19	0.62	0	24,24,26	0.88	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	3NK	A	1210	1	-	2/9/9/10	0/2/2/2
2	3NK	A	1083[A]	1	-	0/9/9/10	0/2/2/2
2	3NK	A	1083[B]	1	-	2/9/9/10	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1083[A]	3NK	C7-C6	3.35	1.45	1.38
2	A	1210	3NK	C3-C2	2.20	1.43	1.39

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1210	3NK	C7-C6-O6-C15
2	A	1210	3NK	C5-C6-O6-C15
2	A	1083[B]	3NK	C7-C6-O6-C15
2	A	1083[B]	3NK	C5-C6-O6-C15

There are no ring outliers.

3 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1210	3NK	2	0
2	A	1083[A]	3NK	3	0
2	A	1083[B]	3NK	13	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	247/258 (95%)	-0.17	22 (8%) <b>9</b> <b>8</b>	18, 30, 65, 84	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	184	PHE	9.6
1	A	185	GLU	8.5
1	A	213	ILE	7.2
1	A	212	GLY	6.9
1	A	58	SER	6.8
1	A	60	LEU	6.5
1	A	248	GLU	5.9
1	A	217	ASN	4.7
1	A	61	ASP	4.1
1	A	57	PRO	4.1
1	A	216	ARG	3.7
1	A	186	THR	3.5
1	A	224	LYS	3.5
1	A	63	GLU	3.1
1	A	220	GLU	2.8
1	A	42	LYS	2.6
1	A	59	GLY	2.4
1	A	183	ASN	2.2
1	A	192	GLU	2.2
1	A	188	GLU	2.2
1	A	191	LYS	2.2
1	A	27	GLN	2.0

### 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 monosaccharides 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(Å <sup>2</sup> )	Q<0.9
2	3NK	A	1210	17/18	0.90	0.11	28,36,50,50	31
2	3NK	A	1083[B]	17/18	0.92	0.11	28,36,49,49	31
2	3NK	A	1083[A]	17/18	0.92	0.11	22,27,37,37	31

### 6.5 Other polymers [i](#)

There are no such residues in this entry.