wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:51 AM GMT

PDB ID	:	2YKN
Title	:	CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE (RT) IN COMPLEX WITH A DIFLUOROMETHYLBENZOXAZOLE (DFMB) PYRIMIDINE THIOETHER DERIVATIVE, A NON-NUCLEOSIDE RT INHIBITOR (NNRTI)
Authors	:	Boyer, J.; Arnoult, E.; Medebielle, M.; Guillemont, J.; Unge, T.; Unge, J.; Jochmans, D.
Deposited on	:	2011-05-28
Resolution	:	2.12 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

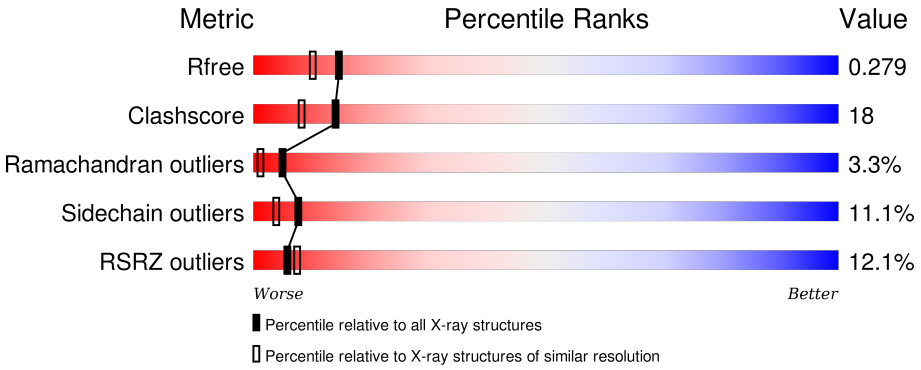
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.12 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	4587 (2.14-2.10)
Clashscore	102246	5132 (2.14-2.10)
Ramachandran outliers	100387	5080 (2.14-2.10)
Sidechain outliers	100360	5081 (2.14-2.10)
RSRZ outliers	91569	4597 (2.14-2.10)

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 <=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	562	10% 68% 23% 7%
2	B	428	14% 64% 23% 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8192 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 REVERSE TRANSCRIPTASE/RIBONUCLEASE H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	558	4506	2917	746	835	8	0	1	1

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	558	HIS	-	EXPRESSION TAG	UNP P03366
A	559	HIS	-	EXPRESSION TAG	UNP P03366
A	560	HIS	-	EXPRESSION TAG	UNP P03366
A	561	HIS	-	EXPRESSION TAG	UNP P03366
A	562	HIS	-	EXPRESSION TAG	UNP P03366
A	57	SER	ASN	CONFLICT	UNP P03366
A	227	LEU	PHE	ENGINEERED MUTATION	UNP P03366
A	478	GLN	GLU	ENGINEERED MUTATION	UNP P03366

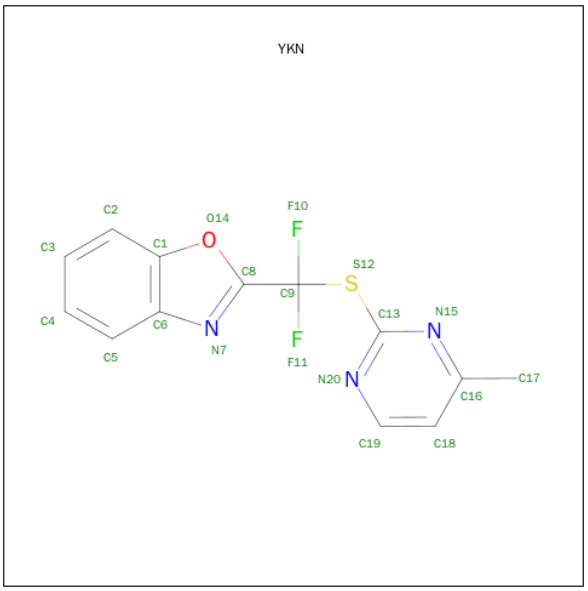
- Molecule 2 is a protein called REVERSE TRANSCRIPTASE/RIBONUCLEASE H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	411	3389	2211	557	615	6	0	1	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	CONFLICT	UNP P03366

- Molecule 3 is 2-[DIFLUORO-[(4-METHYL-PYRIMIDINYL)-THIO]METHYL]-BENZOXAZOLE (three-letter code: YKN) (formula: $C_{13}H_9F_2N_3OS$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C F N O S 20 13 2 3 1 1	0	0

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

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

- Molecule 5 is water.

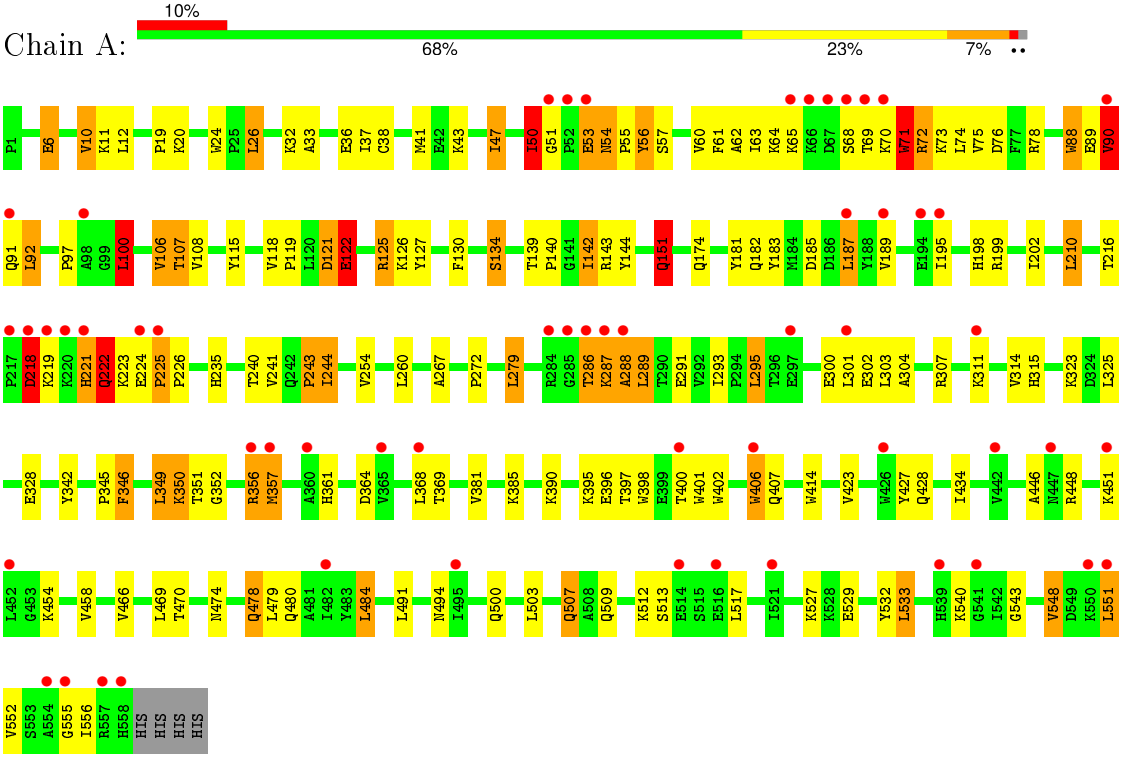
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	162	Total O 162 162	0	0
5	B	114	Total O 114 114	0	0

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 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 (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.

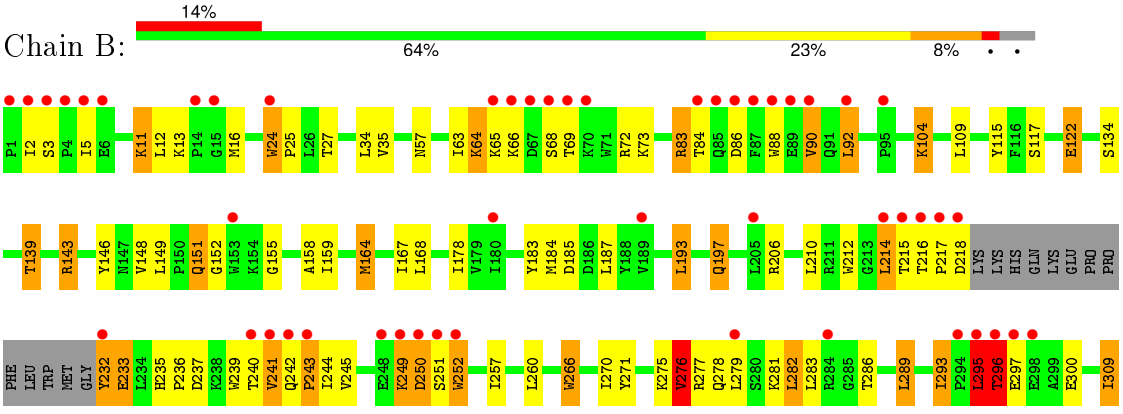
- Molecule 1: REVERSE TRANSCRIPTASE/RIBONUCLEASE H

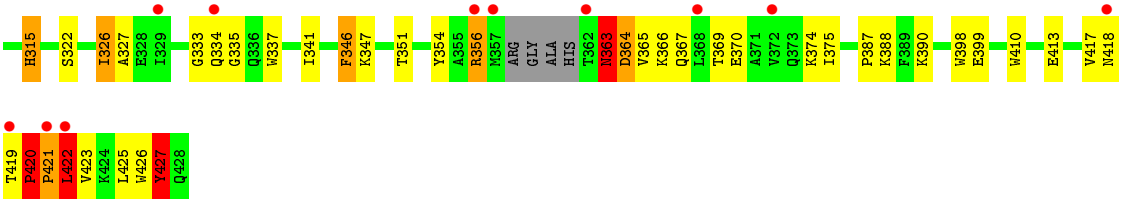
Chain A:



- Molecule 2: REVERSE TRANSCRIPTASE/RIBONUCLEASE H

Chain B:




H315
S322
I326
A327
E328
I329
G333
Q334
G335
Q336
W337
I341
F346
K347
T351
Y354
A355
R356
N357
ARG
GLY
ALA
HIS
T362
N363
D364
V365
K366
Q367
L368
T369
E370
A371
V372
Q373
K374
I375
P387
K388
F389
K390
W398
E399
W410
E413
V417
N418
T419
P420
P421
L422
V423
K424
L425
W426
Y427
Q428

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	119.05Å 155.58Å 152.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.12 19.45 – 2.12	Depositor EDS
% Data completeness (in resolution range)	92.6 (30.00-2.12) 92.6 (19.45-2.12)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 2.11Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.227 , 0.279 0.227 , 0.279	Depositor DCC
R_{free} test set	3761 reflections (5.30%)	DCC
Wilson B-factor (Å^2)	39.9	Xtriage
Anisotropy	0.040	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 54.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 74727 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8192	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.26% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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

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	1.07	10/4626 (0.2%)	1.00	18/6289 (0.3%)
2	B	1.05	5/3490 (0.1%)	1.14	8/4747 (0.2%)
All	All	1.06	15/8116 (0.2%)	1.06	26/11036 (0.2%)

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	1	1
2	B	0	3
All	All	1	4

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	90	VAL	CB-CG2	7.41	1.68	1.52
2	B	143	ARG	CG-CD	-6.98	1.34	1.51
1	A	10	VAL	CB-CG1	-6.42	1.39	1.52
2	B	148	VAL	CB-CG1	6.10	1.65	1.52
1	A	57	SER	CB-OG	6.01	1.50	1.42

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	427	TYR	O-C-N	-32.69	70.40	122.70
2	B	427	TYR	CA-C-O	-18.30	81.67	120.10
2	B	427	TYR	CA-C-N	15.84	152.04	117.20
1	A	185	ASP	CB-CG-OD1	-10.45	108.90	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	345	PRO	C-N-CA	7.60	140.70	121.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	346	PHE	CA

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	88	TRP	Peptide
2	B	295	LEU	Peptide
2	B	417	VAL	Peptide
2	B	427	TYR	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	4506	0	4554	168	0
2	B	3389	0	3416	134	0
3	A	20	0	9	0	0
4	A	1	0	0	0	0
5	A	162	0	0	9	0
5	B	114	0	0	6	0
All	All	8192	0	7979	291	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 18.

The worst 5 of 291 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:89:GLU:CG	1:A:90:VAL:H	1.28	1.30
2:B:266:TRP:HH2	2:B:427:TYR:OH	1.20	1.20
1:A:107:THR:HG21	1:A:202:ILE:CD1	1.73	1.17

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:89:GLU:HG2	1:A:90:VAL:N	1.49	1.12
2:B:271:TYR:HB3	2:B:309:ILE:HD11	1.25	1.12

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	557/562 (99%)	507 (91%)	32 (6%)	18 (3%)	5 1
2	B	406/428 (95%)	375 (92%)	17 (4%)	14 (3%)	5 1
All	All	963/990 (97%)	882 (92%)	49 (5%)	32 (3%)	5 1

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	90	VAL
1	A	122	GLU
1	A	218	ASP
1	A	222	GLN
1	A	243	PRO

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	493/502 (98%)	443 (90%)	50 (10%)	9 5
2	B	373/390 (96%)	326 (87%)	47 (13%)	5 3
All	All	866/892 (97%)	769 (89%)	97 (11%)	8 4

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	512	LYS
2	B	24[B]	TRP
2	B	347	LYS
1	A	527	LYS
1	A	551	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	407	GLN
1	A	494	ASN
2	B	363	ASN
1	A	474	ASN
1	A	507	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 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	YKN	A	1559	-	15,22,22	0.81	0	16,32,32	2.78	5 (31%)

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	YKN	A	1559	-	-	0/0/11/11	0/2/3/3

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1559	YKN	N20-C13-N15	-5.96	119.67	127.66
3	A	1559	YKN	C18-C16-N15	-2.10	118.52	121.44
3	A	1559	YKN	C4-C3-C2	2.83	124.58	120.45
3	A	1559	YKN	C19-N20-C13	2.84	119.47	114.93
3	A	1559	YKN	C13-N15-C16	6.99	123.54	115.91

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

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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	558/562 (99%)	0.52	56 (10%) 9 13	22, 38, 65, 83	0
2	B	411/428 (96%)	0.75	61 (14%) 3 5	21, 38, 73, 93	0
All	All	969/990 (97%)	0.62	117 (12%) 6 7	21, 38, 69, 93	0

The worst 5 of 117 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	294	PRO	12.8
2	B	216	THR	9.6
2	B	2	ILE	9.1
2	B	295	LEU	9.0
2	B	357	MET	9.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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^{th} 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(Å^2)	Q<0.9
3	YKN	A	1559	20/20	0.89	0.15	0.21	28,35,46,50	0
4	CA	A	1560	1/1	0.92	0.12	-	73,73,73,73	0

6.5 Other polymers (i)

There are no such residues in this entry.