wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 10:14 AM GMT

PDB ID	:	3LAL
Title	:	Crystal structure of HIV-1 reverse transcriptase in complex with N1-ethyl pyrimidinedione non-nucleoside inhibitor
Authors	:	Lansdon, E.B.; Mitchell, M.L.
Deposited on	:	2010-01-06
Resolution	:	2.51 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:

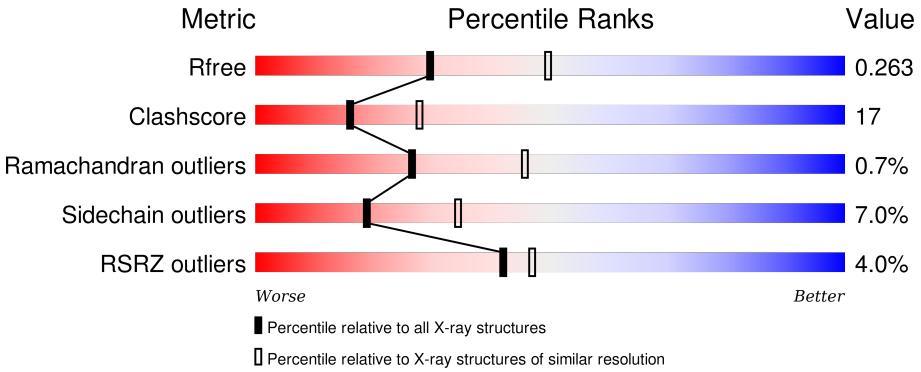
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.51 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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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 >=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	560	4% 61% 34%
1	B	560	3% 47% 23% 28%

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
3	SO4	A	564	-	-	-	X

2 Entry composition

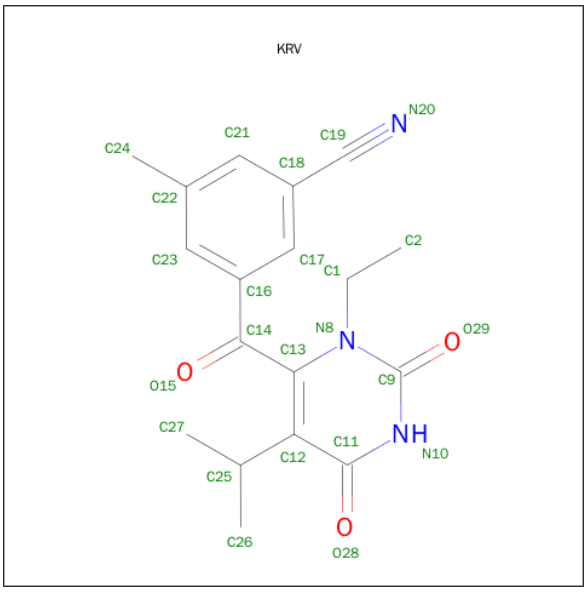
There are 4 unique types of molecules in this entry. The entry contains 8013 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 HIV Reverse transcriptase.

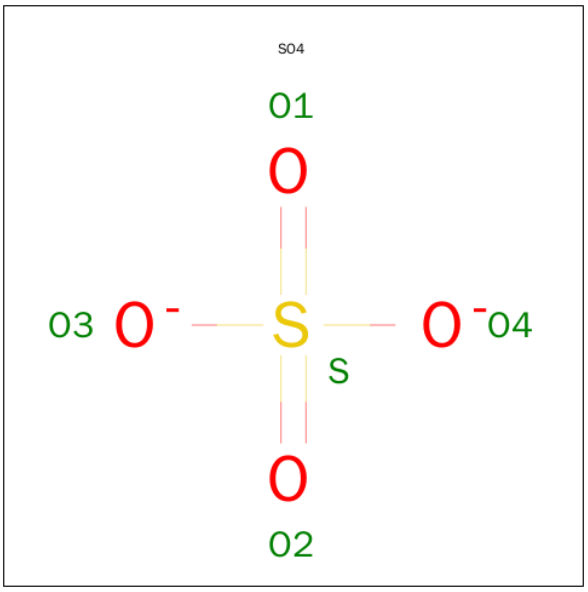
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	553	Total	C	N	O	S	0	0	0
			4508	2914	752	834	8			
1	B	403	Total	C	N	O	S	0	0	0
			3334	2173	550	605	6			

- Molecule 2 is 3-{[3-ETHYL-5-(1-METHYLETHYL)-2,6-DIOXO-1,2,3,6-TETRAHYDRO PYRIMIDIN-4-YL]CARBONYL}-5-METHYLBENZONITRILE (three-letter code: KRV) (formula: $C_{18}H_{19}N_3O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			24	18	3	3		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

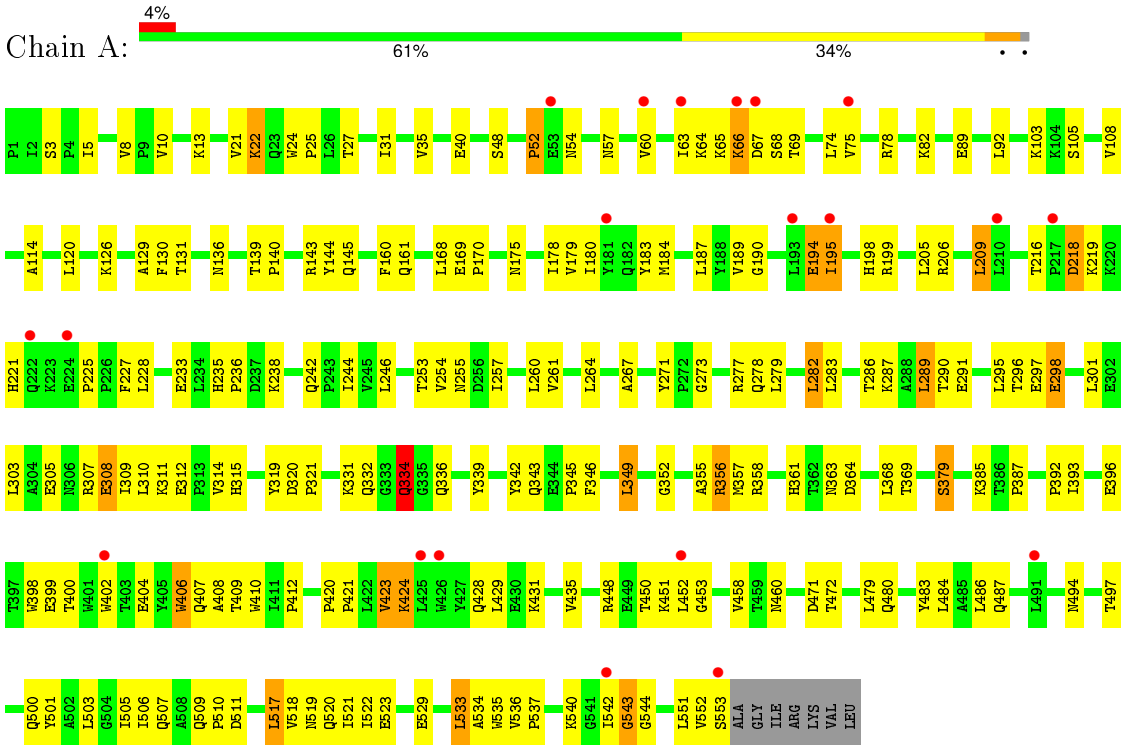
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	69	Total O 69 69	0	0
4	B	53	Total O 53 53	0	0

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.

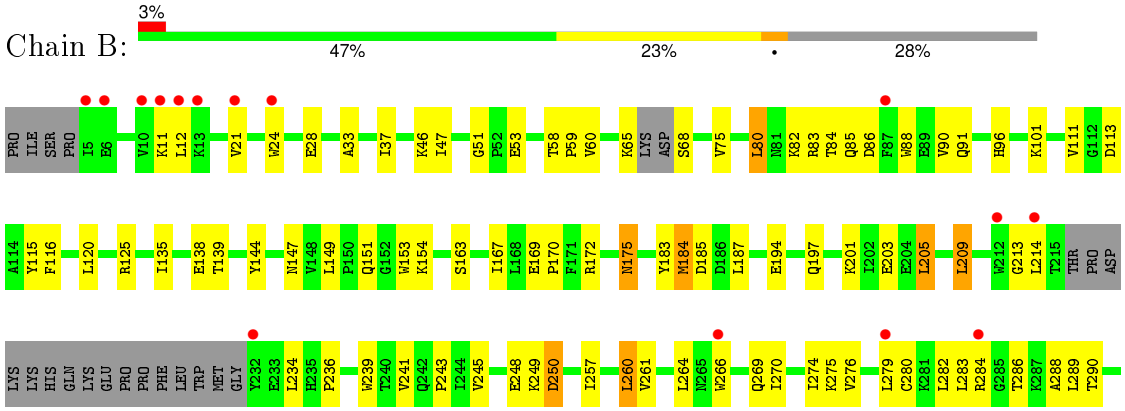
- Molecule 1: HIV Reverse transcriptase

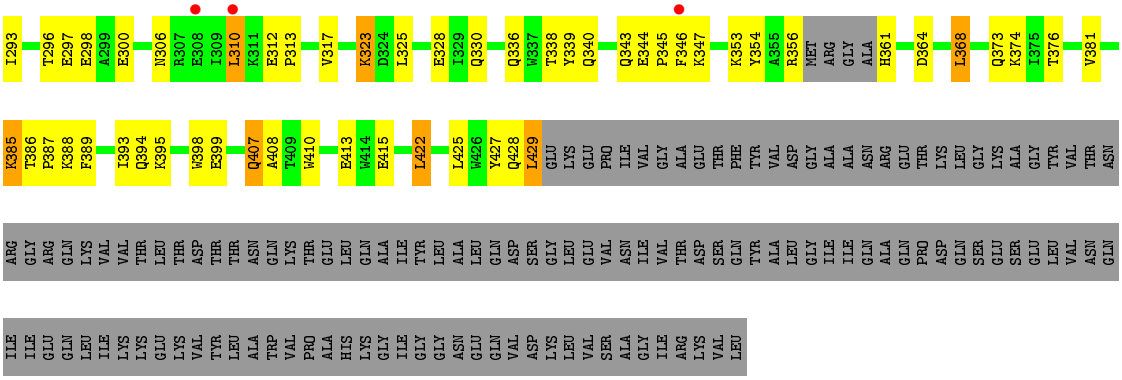
Chain A:



- Molecule 1: HIV Reverse transcriptase

Chain B:




T293 T296 E297 E298 A299 E300 N306 R307 E308 I309 L310 K311 E312 P313 V317 K323 D324 L325 E328 I329 Q330 Q336 W337 T338 Y339 Q340 Q343 E344 P345 F346 K347 K353 Y354 A355 R356 MET ARG GLY ALA H361 D364 L368 Q373 K374 I375 T376 V381
K385 T386 P387 K388 F389 I393 Q394 K395 W398 E399 Q407 A408 T409 W410 E413 W414 E415 L422 L425 W426 Y427 Q428 L429 GLU LYS GLU PRO ILE VAL GLY ALA GLU THR PHE TYR VAL ASP GLY ALA ALA ASN ARG GLU THR LYS LEU GLY LYS ALA GLY TYR VAL THR ASN
ARG GLY ARG GLN LYS VAL VAL THR LEU THR ASP THR THR ASN GLN LYS THR GLU LEU GLN ALA ILE TYR LEU ALA LEU GLN ASP SER GLY LEU GLU VAL ASN ILE VAL THR ASP SER GLN TYR ALA LEU GLY ILE ILE GLN ALA GLN PRO ASP GLN SER GLU SER GLU LEU VAL ASN GLN
ILE ILE GLU GLN LEU ILE LYS LYS GLU LYS VAL TYR LEU ALA TRP VAL PRO ALA HIS LYS GLY ILE GLY GLY ASN GLU GLN VAL ASP LYS LEU VAL SER ALA GLY ILE ARG LYS VAL LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	118.15Å 154.12Å 153.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.83 – 2.51 47.11 – 2.51	Depositor EDS
% Data completeness (in resolution range)	87.8 (29.83-2.51) 87.9 (47.11-2.51)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.19 (at 2.51Å)	Xtriage
Refinement program	CNX 2005	Depositor
R, R_{free}	0.218 , 0.267 0.209 , 0.263	Depositor DCC
R_{free} test set	2112 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	58.8	Xtriage
Anisotropy	0.359	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 45761 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8013	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.74% 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: KRV, SO4

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	0.37	0/4625	0.52	0/6284
1	B	0.38	0/3426	0.52	0/4655
All	All	0.38	0/8051	0.52	0/10939

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	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	52	PRO	Peptide
1	B	361	HIS	Peptide

5.2 Too-close contacts

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	4508	0	4558	167	0
1	B	3334	0	3365	113	0
2	A	24	0	19	0	0
3	A	20	0	0	0	0
3	B	5	0	0	0	0
4	A	69	0	0	3	0
4	B	53	0	0	4	0
All	All	8013	0	7942	268	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 17.

The worst 5 of 268 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:131:THR:HG22	1:A:143:ARG:HG2	1.24	1.10
1:B:296:THR:HG22	1:B:298:GLU:H	1.28	0.95
1:A:510:PRO:HG3	4:A:568:HOH:O	1.68	0.91
1:B:373:GLN:HE22	1:B:407:GLN:H	1.17	0.90
1:A:393:ILE:HB	1:A:423:VAL:HG13	1.54	0.89

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	551/560 (98%)	522 (95%)	22 (4%)	7 (1%)	15 26
1	B	395/560 (70%)	384 (97%)	11 (3%)	0	100 100
All	All	946/1120 (84%)	906 (96%)	33 (4%)	7 (1%)	26 46

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	195	ILE
1	A	543	GLY
1	A	334	GLN
1	A	410	TRP
1	A	136	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	495/500 (99%)	459 (93%)	36 (7%)	17	32
1	B	367/500 (73%)	343 (94%)	24 (6%)	21	39
All	All	862/1000 (86%)	802 (93%)	60 (7%)	19	34

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	428	GLN
1	A	529	GLU
1	B	385	LYS
1	A	507	GLN
1	A	551	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	524	GLN
1	B	85	GLN
1	B	407	GLN
1	A	545	ASN
1	B	175	ASN

5.3.3 RNA

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)

6 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	KRV	A	561	-	20,25,25	2.06	6 (30%)	24,36,36	3.13	6 (25%)
3	SO4	A	562	-	4,4,4	0.18	0	6,6,6	0.15	0
3	SO4	A	563	-	4,4,4	0.22	0	6,6,6	0.14	0
3	SO4	A	564	-	4,4,4	0.32	0	6,6,6	0.11	0
3	SO4	A	565	-	4,4,4	0.25	0	6,6,6	0.14	0
3	SO4	B	561	-	4,4,4	0.07	0	6,6,6	0.11	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	KRV	A	561	-	-	0/16/16/16	0/2/2/2
3	SO4	A	562	-	-	0/0/0/0	0/0/0/0
3	SO4	A	563	-	-	0/0/0/0	0/0/0/0
3	SO4	A	564	-	-	0/0/0/0	0/0/0/0
3	SO4	A	565	-	-	0/0/0/0	0/0/0/0
3	SO4	B	561	-	-	0/0/0/0	0/0/0/0

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	561	KRV	C17-C16	2.74	1.43	1.39
2	A	561	KRV	C23-C22	3.03	1.44	1.38
2	A	561	KRV	C23-C16	3.20	1.44	1.39
2	A	561	KRV	C13-N8	3.39	1.42	1.35
2	A	561	KRV	C21-C18	3.51	1.46	1.39

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	561	KRV	O15-C14-C13	-5.46	113.93	120.21
2	A	561	KRV	C12-C11-N10	-4.23	118.43	125.18
2	A	561	KRV	C21-C18-C19	2.57	122.76	119.51
2	A	561	KRV	C13-C12-C25	2.64	125.25	122.34
2	A	561	KRV	C16-C14-C13	6.82	127.39	119.53

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	553/560 (98%)	0.37	20 (3%)	46	51	40, 62, 93, 117	0
1	B	403/560 (71%)	0.32	18 (4%)	37	42	36, 57, 93, 108	0
All	All	956/1120 (85%)	0.35	38 (3%)	42	47	36, 60, 93, 117	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	346	PHE	4.1
1	A	193	LEU	4.0
1	B	12	LEU	3.9
1	B	279	LEU	3.9
1	B	232	TYR	3.7

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	SO4	A	564	5/5	0.85	0.45	12.61	121,121,121,121	0
2	KRV	A	561	24/24	0.93	0.23	0.87	55,59,66,67	0
3	SO4	A	562	5/5	0.94	0.16	0.39	77,78,80,82	0
3	SO4	B	561	5/5	0.94	0.13	-1.08	82,83,84,85	0
3	SO4	A	563	5/5	0.97	0.12	-1.15	76,77,78,79	0
3	SO4	A	565	5/5	0.84	0.14	-	120,120,121,121	0

6.5 Other polymers (i)

There are no such residues in this entry.