wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:26 PM GMT

PDB ID	:	1TKZ
Title	:	CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE IN COMPLEX WITH GW429576
Authors	:	Hopkins, A.L.; Ren, J.; Stuart, D.I.; Stammers, D.K.
Deposited on	:	2004-06-09
Resolution	:	2.81 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:

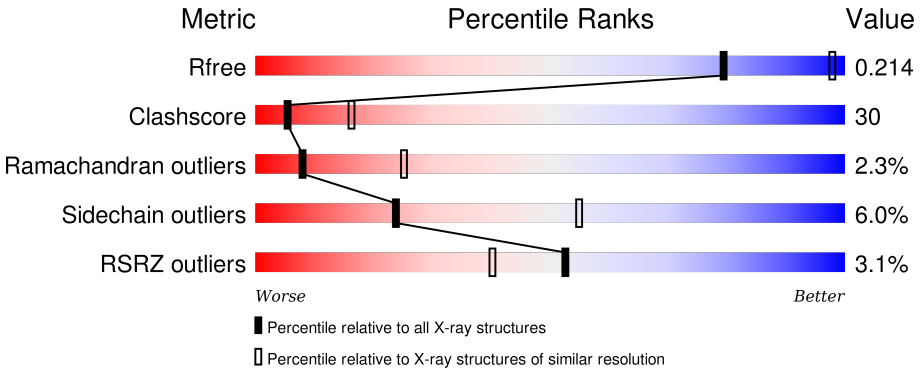
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.81 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	2676 (2.84-2.80)
Clashscore	102246	3124 (2.84-2.80)
Ramachandran outliers	100387	3072 (2.84-2.80)
Sidechain outliers	100360	3074 (2.84-2.80)
RSRZ outliers	91569	2690 (2.84-2.80)

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	3%, 43%, 43%, 5%, 8%
2	B	440	3%, 42%, 47%, 9%

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	PO4	A	1301	-	-	-	X
3	PO4	A	1302	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7612 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 Pol polyprotein, Reverse transcriptase, Chain A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	517	4247	2754	701	784	8	0	0	0

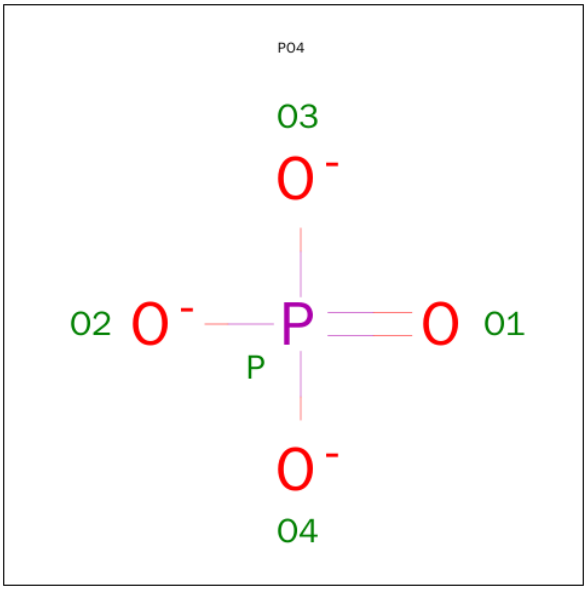
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	CSD	CYS	OXIDIZED CY	UNP P04585

- Molecule 2 is a protein called Pol polyprotein, Reverse transcriptase, Chain B.

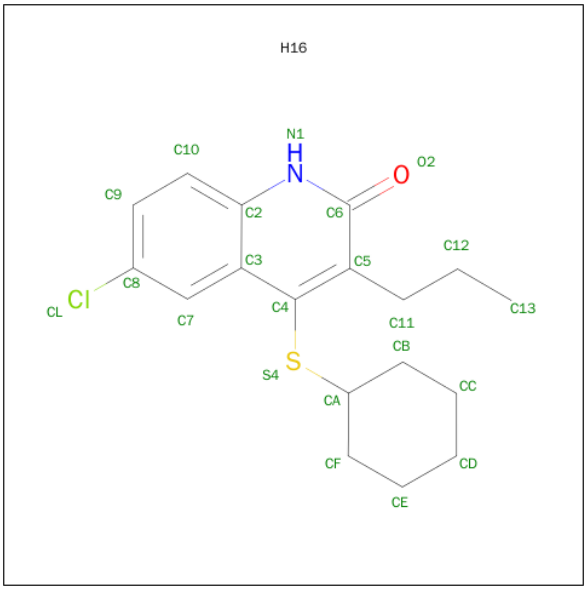
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	401	3328	2173	548	600	7	0	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

- Molecule 4 is 6-CHLORO-4-(CYCLOHEXYLSULFANYL)-3-PROPYLQUINOLIN-2(1H)-ONE (three-letter code: H16) (formula: $C_{18}H_{22}ClNOS$).



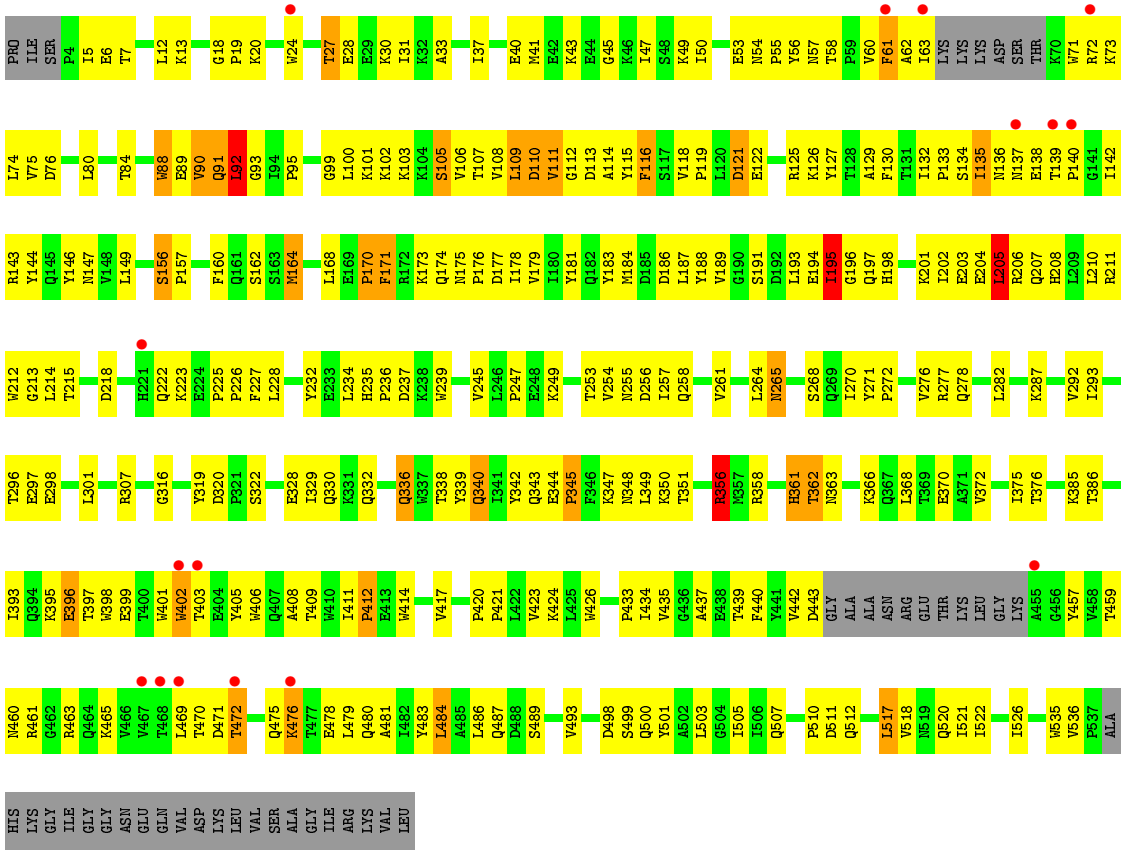
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C Cl N O S 22 18 1 1 1 1	0	0

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.

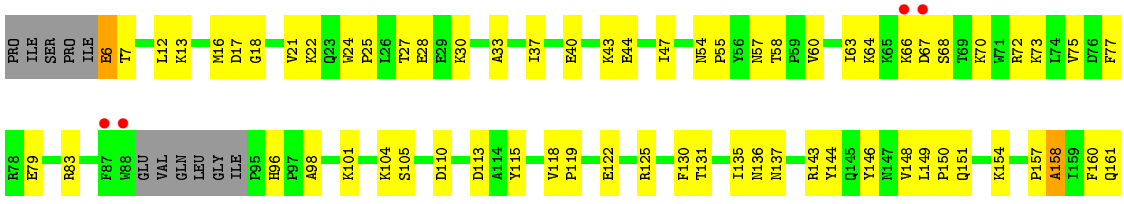
- Molecule 1: Pol polyprotein, Reverse transcriptase, Chain A

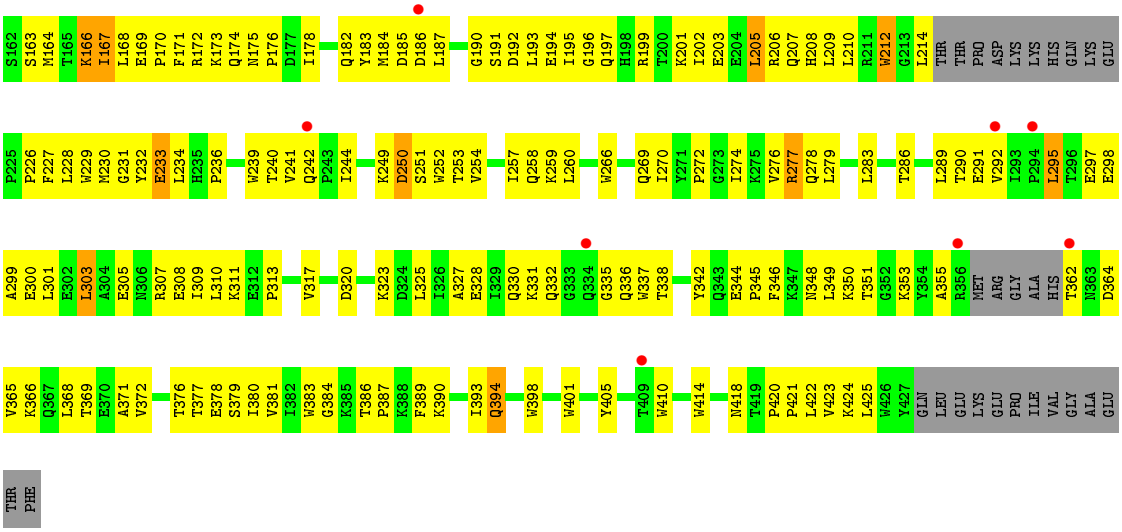
Chain A:



- Molecule 2: Pol polyprotein, Reverse transcriptase, Chain B

Chain B:




S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 P176 D177 I178 Q182 Y183 M184 D185 D186 L187 G190 S191 D192 L193 E194 I195 G196 Q197 H198 R199 T200 K201 I202 E203 E204 L205 R206 Q207 H208 L209 L210 R211 W212 G213 L214 THR THR PRO ASP LYS LYS HIS GLN LYS GLU
P225 P226 F227 L228 W229 M230 G231 Y232 E233 L234 H235 P236 W239 T240 V241 Q242 P243 I244 K249 D250 S251 W252 T253 V254 I257 Q258 K259 L260 W266 Q269 I270 Y271 P272 G273 I274 K275 V276 R277 Q278 L279 L283 T286 L289 T290 E291 V292 I293 P294 L295 T296 E297 E298
A299 E300 L301 E302 L303 A304 E305 N306 R307 E308 I309 L310 K311 E312 P313 V317 D320 K323 D324 L325 I326 A327 E328 I329 Q330 K331 Q332 G333 Q334 G335 Q336 W337 T338 Y342 Q343 E344 P345 F346 K347 N348 L349 K350 T351 G352 K353 Y354 A355 R356 MET ARG GLY ALA HIS T362 N363 D364
V365 K366 Q367 L368 T369 E370 A371 V372 T376 T377 E378 S379 I380 V381 I382 W383 G384 K385 T386 P387 K388 F389 K390 I393 Q394 W398 W401 Y405 T409 W410 W414 N418 T419 P420 P421 L422 V423 K424 L425 W426 Y427 GLN LEU GLU LYS GLU PRO ILE VAL GLY ALA GLU
THR PHE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	137.80Å 115.20Å 65.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.81 – 2.81 29.81 – 2.81	Depositor EDS
% Data completeness (in resolution range)	93.7 (29.81-2.81) 93.9 (29.81-2.81)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.206 , 0.283 0.199 , 0.214	Depositor DCC
R_{free} test set	1165 reflections (4.79%)	DCC
Wilson B-factor (Å^2)	60.1	Xtriage
Anisotropy	0.210	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 84.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 24363 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7612	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.43% 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: CSD, PO4, H16

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	0.50	0/4352	0.72	3/5918 (0.1%)
2	B	0.50	0/3425	0.70	0/4652
All	All	0.50	0/7777	0.71	3/10570 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	88	TRP	N-CA-C	-6.47	93.54	111.00
1	A	92	LEU	CA-CB-CG	5.63	128.25	115.30
1	A	205	LEU	CA-CB-CG	5.28	127.44	115.30

There are no chirality outliers.

There are no planarity outliers.

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	4247	0	4275	269	0
2	B	3328	0	3353	209	0
3	A	15	0	0	3	0
4	A	22	0	22	2	0
All	All	7612	0	7650	455	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 30.

The worst 5 of 455 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:206:ARG:HH12	1:A:218:ASP:HA	1.28	0.98
2:B:227:PHE:HB3	2:B:231:GLY:HA2	1.49	0.95
1:A:195:ILE:H	1:A:195:ILE:HD13	1.33	0.94
1:A:343:GLN:HG3	1:A:349:LEU:HD11	1.50	0.93
2:B:66:LYS:HE3	2:B:230:MET:HG3	1.50	0.91

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	510/560 (91%)	453 (89%)	42 (8%)	15 (3%)	6 19
2	B	393/440 (89%)	338 (86%)	49 (12%)	6 (2%)	13 38
All	All	903/1000 (90%)	791 (88%)	91 (10%)	21 (2%)	8 25

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	90	VAL
1	A	91	GLN
1	A	92	LEU
1	A	137	ASN
1	A	156	SER

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	466/499 (93%)	433 (93%)	33 (7%)	18 45
2	B	366/400 (92%)	349 (95%)	17 (5%)	33 67
All	All	832/899 (92%)	782 (94%)	50 (6%)	24 55

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	345	PRO
1	A	396	GLU
2	B	303	LEU
1	A	348	ASN
1	A	361	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	480	GLN
1	A	509	GLN
2	B	348	ASN
1	A	500	GLN
1	A	507	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
1	CSD	A	280	1	3,7,8	0.83	0	3,8,10	4.18	1 (33%)

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
1	CSD	A	280	1	-	1/2/6/8	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	280	CSD	OD1-SG-CB	7.00	117.06	105.40

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	280	CSD	CA-CB-SG-OD1

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
3	PO4	A	1300	-	4,4,4	1.04	0	6,6,6	0.27	0
3	PO4	A	1301	-	4,4,4	1.11	0	6,6,6	0.27	0
3	PO4	A	1302	-	4,4,4	1.15	0	6,6,6	0.27	0
4	H16	A	999	-	22,24,24	2.01	7 (31%)	25,33,33	1.74	3 (12%)

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	PO4	A	1300	-	-	0/0/0/0	0/0/0/0
3	PO4	A	1301	-	-	0/0/0/0	0/0/0/0
3	PO4	A	1302	-	-	0/0/0/0	0/0/0/0
4	H16	A	999	-	-	0/7/15/15	0/3/3/3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	999	H16	C4-C3	-2.86	1.39	1.43
4	A	999	H16	C7-C8	2.01	1.40	1.36
4	A	999	H16	CF-CA	2.42	1.60	1.51
4	A	999	H16	C2-N1	2.45	1.39	1.35
4	A	999	H16	C10-C9	3.04	1.43	1.36

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	999	H16	C3-C2-N1	-3.31	120.42	123.45
4	A	999	H16	C4-C3-C2	2.30	120.20	117.59
4	A	999	H16	C6-N1-C2	6.96	120.93	116.23

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1302	PO4	3	0
4	A	999	H16	2	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

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	516/560 (92%)	-0.21	16 (3%)	52	40	25, 63, 107, 131	0
2	B	401/440 (91%)	-0.07	12 (2%)	54	41	25, 62, 110, 135	0
All	All	917/1000 (91%)	-0.15	28 (3%)	52	40	25, 62, 108, 135	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	469	LEU	6.4
2	B	88	TRP	4.5
1	A	455	ALA	4.2
1	A	63	ILE	4.2
2	B	356	ARG	3.9

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

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
1	CSD	A	280	8/9	0.97	0.13	-	48,54,61,78	0

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	PO4	A	1301	5/5	0.87	0.19	2.35	121,133,134,136	0
3	PO4	A	1302	5/5	0.83	0.26	1.77	145,147,149,149	0
3	PO4	A	1300	5/5	0.81	0.23	1.62	123,128,130,133	0
4	H16	A	999	22/22	0.97	0.19	0.86	26,48,56,59	0

6.5 Other polymers

There are no such residues in this entry.