wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2023 – 11:33 PM EDT

PDB ID	:	3ITH
Title	:	Crystal structure of the HIV-1 reverse transcriptase bound to a 6-vinylpyrimidine inhibitor
Authors	:	Freisz, S.; Bec, G.; Wolff, P.; Dumas, P.; Radi, M.; Botta, M.
Deposited on	:	2009-08-28
Resolution	:	2.80 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

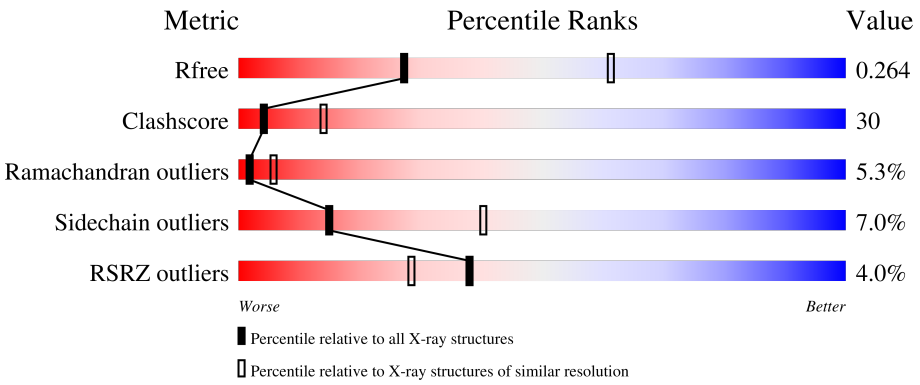
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	5% 45% 47% 7%
1	C	560	5% 49% 43% 8%
2	B	427	2% 52% 38% 6%
2	D	427	3% 47% 44% 6%

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	EDM	A	561	-	-	X	X
3	EDM	C	561	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 15912 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	556	4521	2927	753	833	8	0	0	0
1	C	556	4521	2927	753	833	8	0	0	0

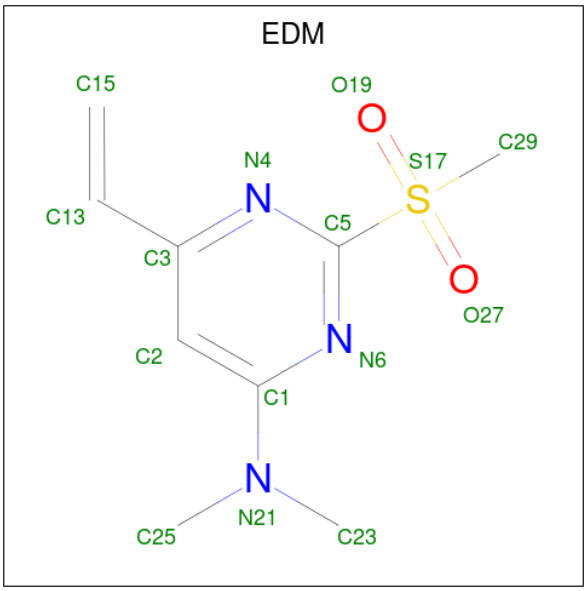
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	478	GLN	GLU	SEE REMARK 999	UNP P03366
C	478	GLN	GLU	SEE REMARK 999	UNP P03366

- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	415	3420	2225	565	623	7	0	0	0
2	D	415	3420	2225	565	623	7	0	0	0

- Molecule 3 is 6-ethenyl-N,N-dimethyl-2-(methylsulfonyl)pyrimidin-4-amine (three-letter code: EDM) (formula: $C_9H_{13}N_3O_2S$).



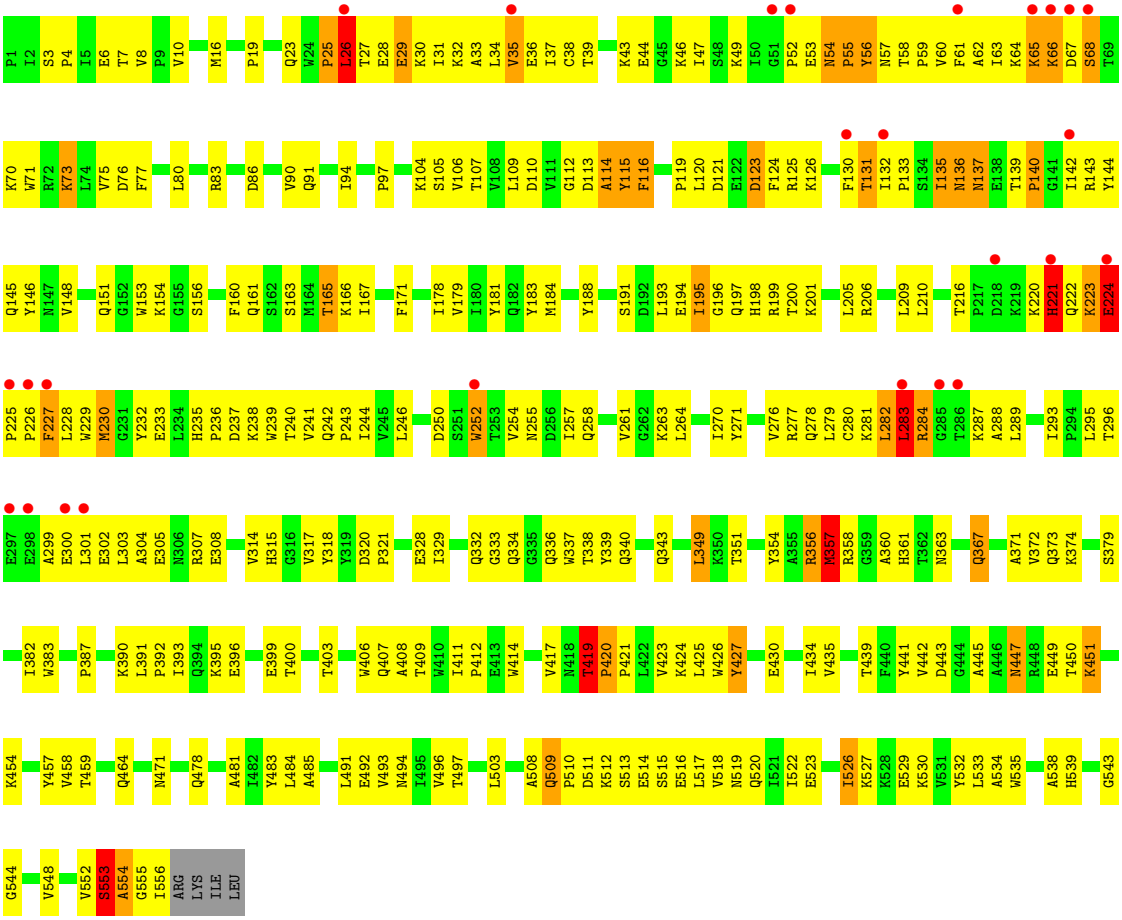
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	15	9	3	2	1	0	0
			Total	C	N	O	S		
3	C	1	15	9	3	2	1	0	0

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.

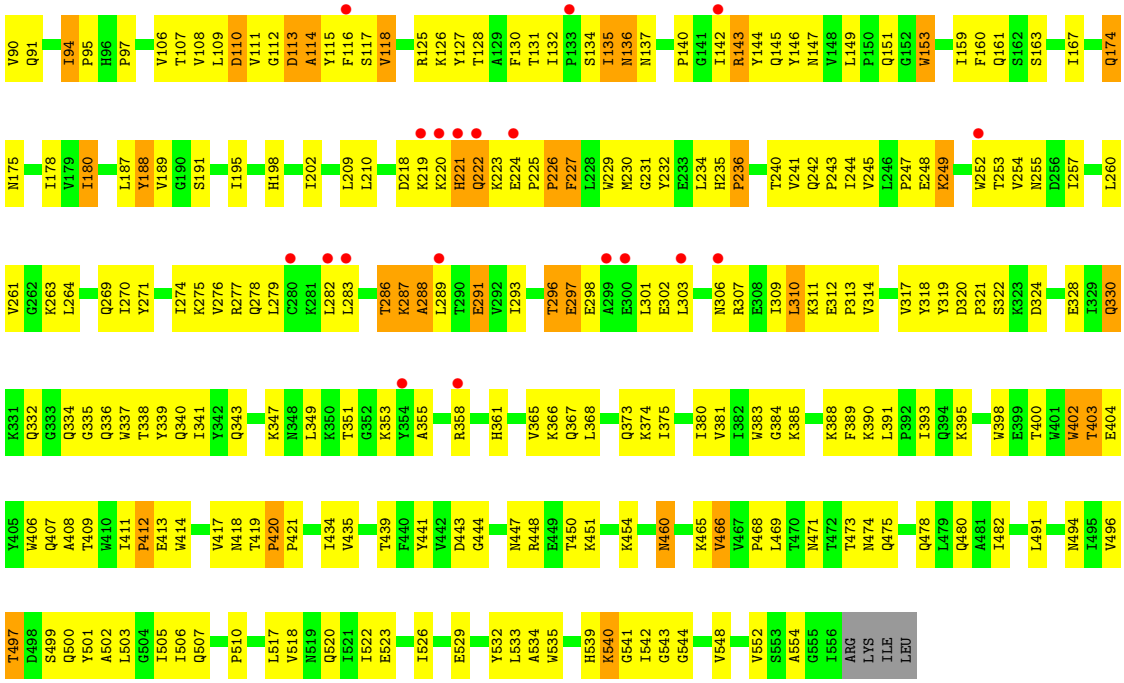
• Molecule 1: Reverse transcriptase/ribonuclease H

Chain A:



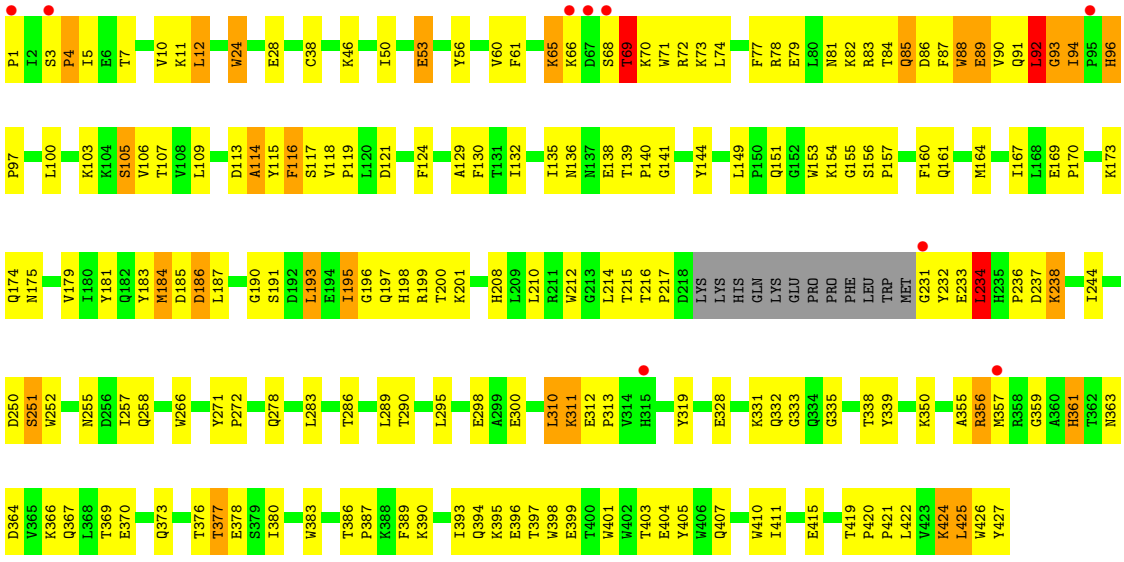
• Molecule 1: Reverse transcriptase/ribonuclease H

Chain C:



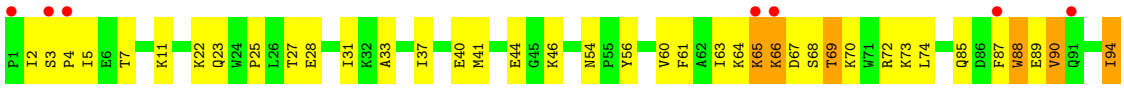
- Molecule 2: p51 RT

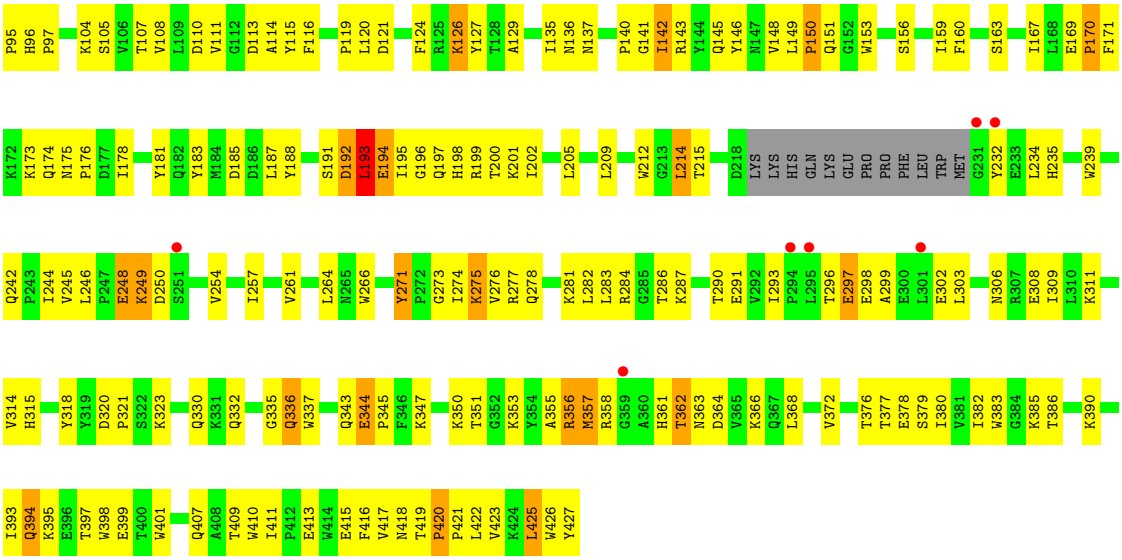
Chain B:



- Molecule 2: p51 RT

Chain D:




P95 H96 P97 K104 S105 V106 T107 V108 L109 D110 V111 G112 D113 A114 Y115 F116 P119 L120 D121 F124 R125 K126 Y127 T128 A129 I135 N136 N137 P140 G141 I142 R143 Y144 Q145 Y146 N147 V148 L149 P150 Q151 G152 W153 S156 I159 F160 S163 I167 L168 E169 P170 F171
K172 K173 Q174 N175 P176 D177 I178 Y181 Q182 Y183 M184 D185 D186 L187 Y188 S191 D192 L193 E194 I195 G196 Q197 H198 R199 T200 K201 I202 L205 L209 W212 G213 L214 T215 D218 LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET G231 Y232 E233 L234 H235 W239
Q242 P243 I244 V245 L246 P247 E248 K249 D250 S251 V254 I257 V261 L264 N265 W266 Y271 P272 G273 I274 K275 V276 R277 Q278 K281 L282 L283 R284 G285 T286 K287 T290 E291 V292 I293 P294 L295 T296 E297 E298 A299 E300 L301 E302 L303 N306 R307 E308 I309 L310 K311
V314 H315 Y318 Y319 D320 P321 S322 K323 Q330 K331 Q332 G335 Q336 W337 Q343 E344 P345 F346 K347 K350 T351 G352 K353 Y354 A355 R356 M357 R358 G359 A360 H361 T362 N363 D364 V365 K366 Q367 L368 V372 T376 T377 E378 S379 I380 V381 I382 W383 G384 K385 T386 K390
I393 Q394 K395 E396 T397 W398 E399 T400 W401 Q407 A408 T409 W410 I411 P412 E413 W414 E415 F416 V417 N418 T419 P420 P421 L422 V423 K424 L425 W426 Y427

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	92.10Å 72.49Å 216.29Å 90.00° 95.89° 90.00°	Depositor
Resolution (Å)	43.90 – 2.80 48.61 – 2.80	Depositor EDS
% Data completeness (in resolution range)	85.3 (43.90-2.80) 85.3 (48.61-2.80)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.4_57)	Depositor
R, R_{free}	0.212 , 0.270 0.204 , 0.264	Depositor DCC
R_{free} test set	5835 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	79.9	Xtriage
Anisotropy	0.230	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 53.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	15912	wwPDB-VP
Average B, all atoms (Å^2)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.97% 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, 0.333 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: EDM

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	0.49	0/4639	0.64	1/6302 (0.0%)
1	C	0.45	0/4639	0.62	0/6302
2	B	0.52	0/3518	0.69	1/4781 (0.0%)
2	D	0.46	0/3518	0.62	0/4781
All	All	0.48	0/16314	0.64	2/22166 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	187	LEU	CA-CB-CG	5.29	127.46	115.30
1	A	156	SER	CB-CA-C	-5.18	100.26	110.10

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	4521	0	4586	316	0
1	C	4521	0	4586	285	0
2	B	3420	0	3454	194	0
2	D	3420	0	3454	195	0
3	A	15	0	13	11	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	15	0	13	8	0
All	All	15912	0	16106	949	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 949 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:C:419:THR:HB	1:C:420:PRO:HD2	1.14	1.11
2:B:216:THR:HB	2:B:217:PRO:HD2	1.31	1.08
1:C:91:GLN:HE22	2:D:140:PRO:HA	1.20	1.06
1:A:435:VAL:HG13	2:B:290:THR:HG21	1.38	1.01
1:A:454:LYS:HB2	1:A:552:VAL:HG23	1.41	1.01

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	554/560 (99%)	456 (82%)	61 (11%)	37 (7%)	1 3
1	C	554/560 (99%)	437 (79%)	85 (15%)	32 (6%)	1 4
2	B	411/427 (96%)	332 (81%)	57 (14%)	22 (5%)	2 5
2	D	411/427 (96%)	356 (87%)	44 (11%)	11 (3%)	5 17
All	All	1930/1974 (98%)	1581 (82%)	247 (13%)	102 (5%)	2 6

5 of 102 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	35	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	52	PRO
1	A	55	PRO
1	A	65	LYS
1	A	140	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	496/500 (99%)	464 (94%)	32 (6%)	17	44
1	C	496/500 (99%)	460 (93%)	36 (7%)	14	38
2	B	377/389 (97%)	351 (93%)	26 (7%)	15	41
2	D	377/389 (97%)	348 (92%)	29 (8%)	13	35
All	All	1746/1778 (98%)	1623 (93%)	123 (7%)	15	40

5 of 123 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	425	LEU
2	D	275	LYS
1	C	180	ILE
2	D	271	TYR
2	D	394	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	175	ASN
1	C	373	GLN
2	D	306	ASN
1	C	198	HIS
1	C	330	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 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	EDM	A	561	-	15,15,15	2.24	2 (13%)	18,22,22	2.95	7 (38%)
3	EDM	C	561	-	15,15,15	2.33	4 (26%)	18,22,22	3.18	8 (44%)

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	EDM	A	561	-	-	5/12/12/12	0/1/1/1
3	EDM	C	561	-	-	6/12/12/12	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	561	EDM	C5-S17	-7.49	1.75	1.80

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	561	EDM	C5-S17	-7.43	1.75	1.80
3	A	561	EDM	O19-S17	2.45	1.50	1.44
3	C	561	EDM	O19-S17	2.23	1.50	1.44
3	C	561	EDM	C5-N4	2.07	1.34	1.32

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	561	EDM	C29-S17-C5	7.41	110.03	104.39
3	C	561	EDM	N6-C1-N21	7.23	120.89	116.91
3	A	561	EDM	C29-S17-C5	7.17	109.85	104.39
3	A	561	EDM	N6-C1-N21	5.95	120.19	116.91
3	A	561	EDM	O19-S17-O27	-5.00	109.30	117.92

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	561	EDM	N6-C5-S17-C29
3	C	561	EDM	N6-C1-N21-C23
3	A	561	EDM	C15-C13-C3-C2
3	C	561	EDM	C15-C13-C3-C2
3	C	561	EDM	N6-C1-N21-C25

There are no ring outliers.

2 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	561	EDM	11	0
3	C	561	EDM	8	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, 95th 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	556/560 (99%)	0.02	26 (4%)	31	22	56, 84, 151, 171	0
1	C	556/560 (99%)	0.08	29 (5%)	27	18	55, 91, 156, 177	0
2	B	415/427 (97%)	-0.09	9 (2%)	62	52	49, 75, 119, 145	0
2	D	415/427 (97%)	0.07	14 (3%)	45	35	55, 86, 129, 157	0
All	All	1942/1974 (98%)	0.03	78 (4%)	38	28	49, 84, 144, 177	0

The worst 5 of 78 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	299	ALA	7.9
2	D	3	SER	6.2
2	B	3	SER	5.5
2	B	1	PRO	5.4
2	D	1	PRO	5.2

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, 95th 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^2)	Q<0.9
3	EDM	A	561	15/15	0.78	0.73	118,132,136,150	0
3	EDM	C	561	15/15	0.82	0.27	114,131,137,145	0

6.5 Other polymers (i)

There are no such residues in this entry.