wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:59 AM GMT

PDB ID	:	2JLE
Title	:	NOVEL INDAZOLE NNRTIS CREATED USING MOLECULAR TEMPLATE HYBRIDIZATION BASED ON CRYSTALLOGRAPHIC OVERLAYS
Authors	:	Jones, L.H.; Allan, G.; Barba, O.; Burt, C.; Corbau, R.; Dupont, T.; Irving, S.; Mowbray, C.E.; Phillips, C.; Swain, N.A.; Webster, R.; Westby, M.
Deposited on	:	2008-09-08
Resolution	:	2.90 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:

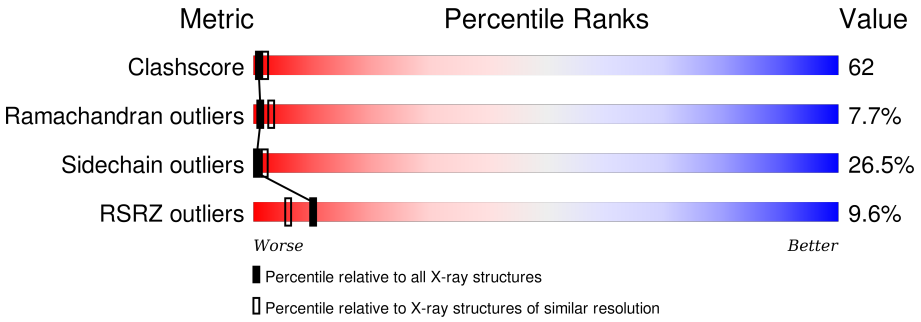
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.90 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))
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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	566	7%, 22%, 52%, 20%
1	B	566	9%, 18%, 38%, 16%, 27%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8120 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.

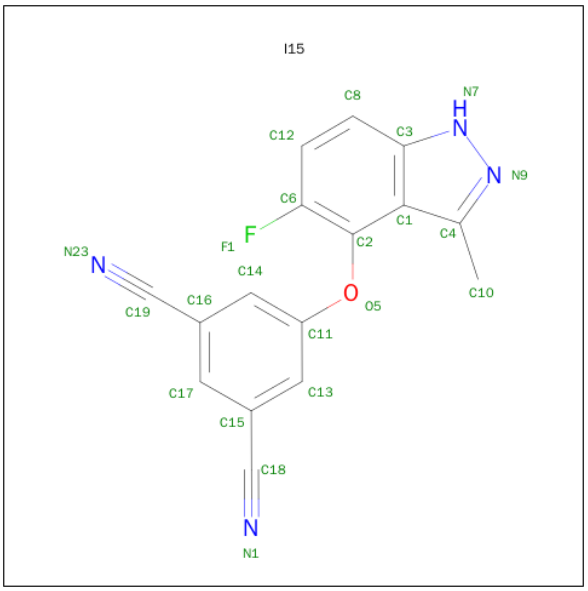
- Molecule 1 is a protein called REVERSE TRANSCRIPTASE/RNASEH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	545	Total	C	N	O	S	0	0	0
			4445	2875	742	820	8			
1	B	416	Total	C	N	O	S	0	0	1
			3414	2218	569	620	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	LYS	ARG	CONFLICT	UNP Q72547
B	103	LYS	ARG	CONFLICT	UNP Q72547
A	350	LYS	ARG	CONFLICT	UNP Q72547
B	350	LYS	ARG	CONFLICT	UNP Q72547

- Molecule 2 is 5-[(5-FLUORO-3-METHYL-1H-INDAZOL-4-YL)OXY]BENZENE-1,3-DICARBONITRILE (three-letter code: I15) (formula: $C_{16}H_9FN_4O$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	F	N	O	0	0
			22	16	1	4	1		

- Molecule 3 is water.

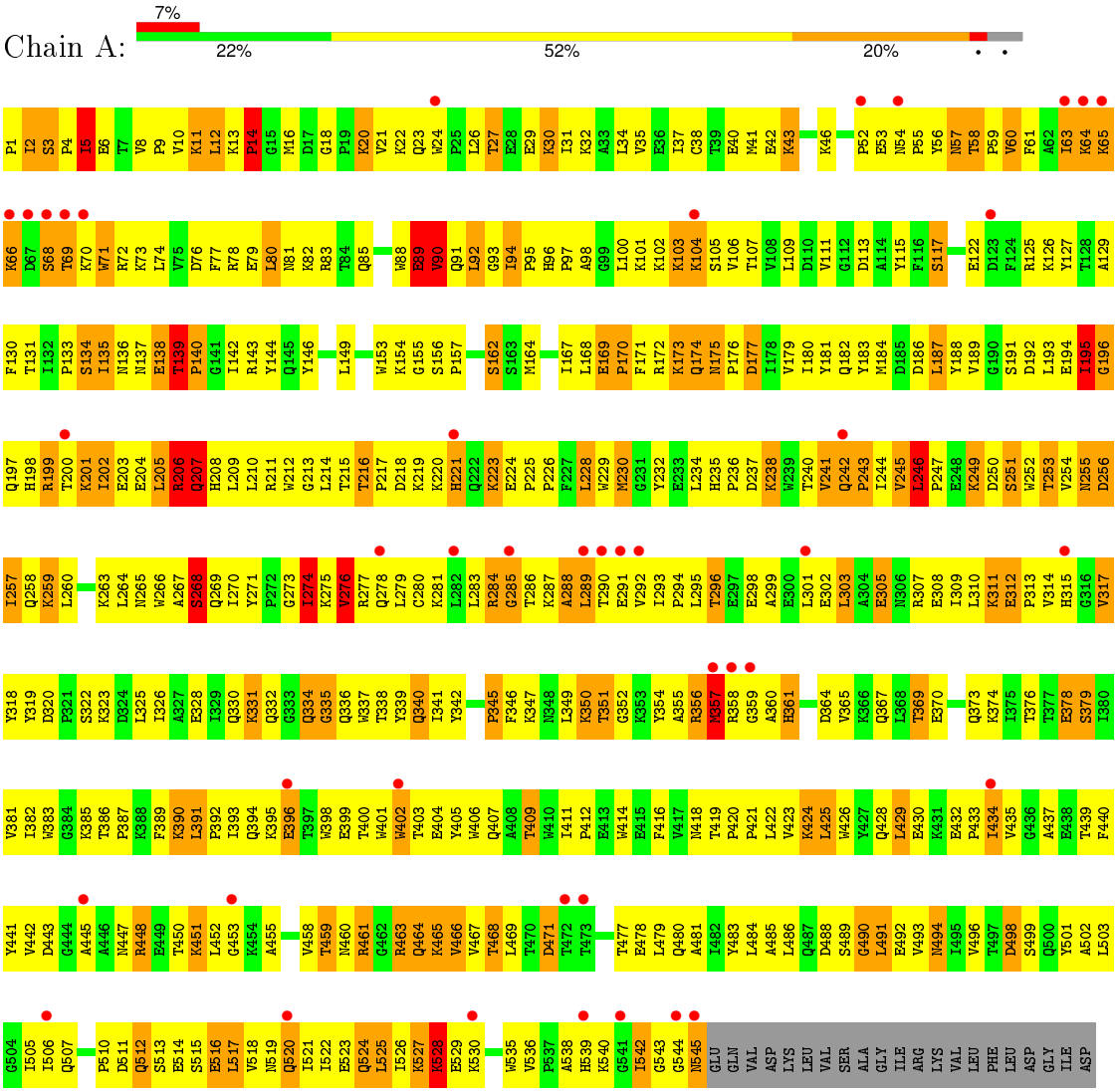
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	152	Total	O	0	0
			152	152		
3	B	87	Total	O	0	0
			87	87		

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.

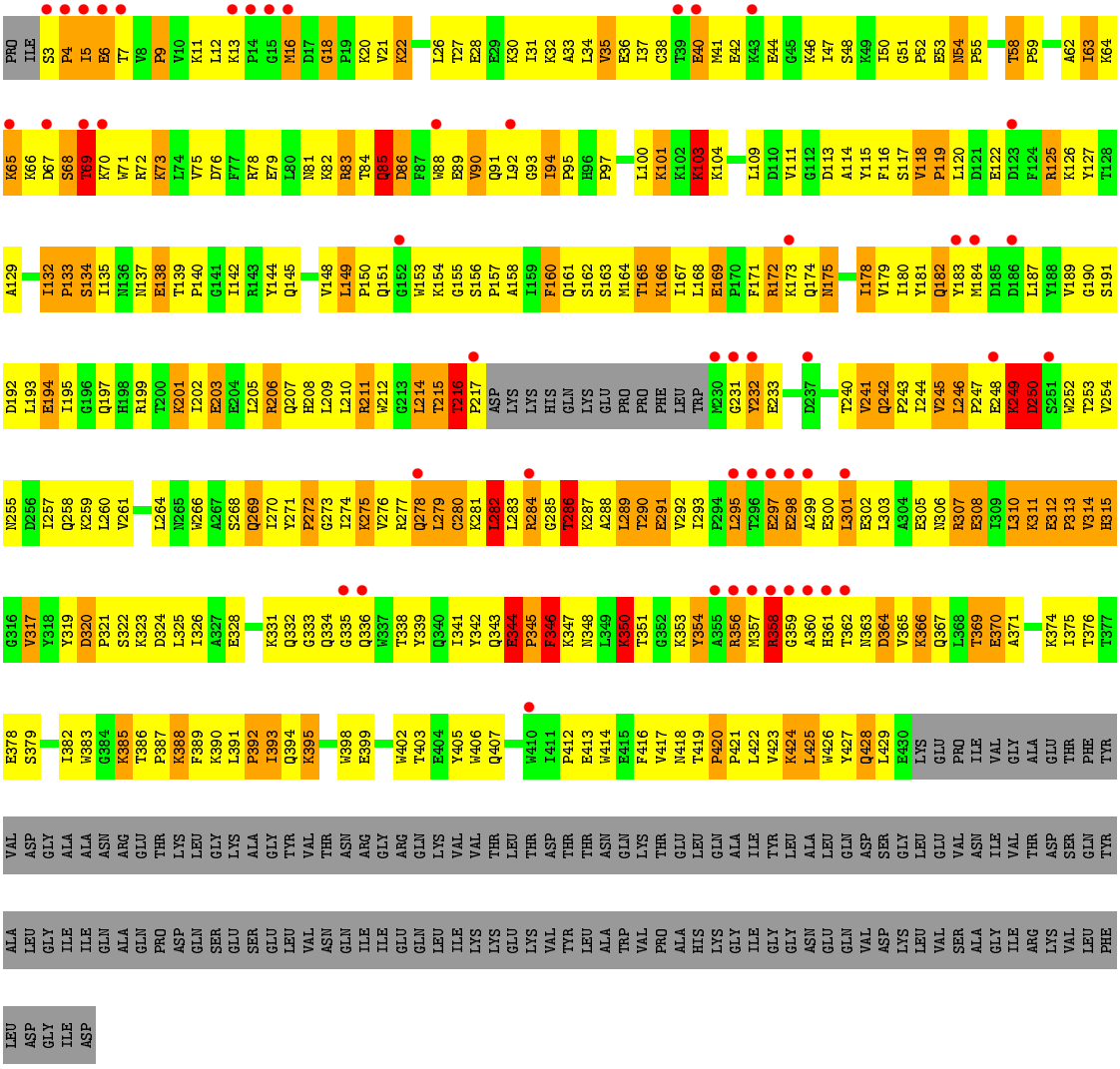
- Molecule 1: REVERSE TRANSCRIPTASE/RNASEH

Chain A:



- Molecule 1: REVERSE TRANSCRIPTASE/RNASEH

Chain B:


PRO ILE S3 P4 I5 E6 T7 V8 P9 V10 K11 L12 K13 P14 G15 M16 D17 G18 P19 K20 V21 K22 L26 T27 E28 E29 K30 I31 K32 A33 L34 V35 E36 I37 C38 T39 E40 M41 E42 K43 E44 G45 K46 I47 S48 K49 I50 G51 P52 E53 N54 P55 T58 P59 A62 I63 K64
K65 K66 D67 S68 T69 K70 W71 R72 K73 L74 V75 D76 F77 R78 E79 L80 N81 K82 R83 T84 Q85 D86 F87 W88 E89 W90 Q91 L92 G93 I94 P95 H96 P97 L100 K101 K102 K103 K104 L109 D110 V111 G112 D113 A114 Y115 F116 S117 V118 P119 L120 D121 E122 D123 F124 R125 K126 Y127 T128
A129 I132 P133 S134 I135 N136 N137 E138 T139 P140 G141 I142 R143 Y144 Q145 V148 L149 P150 Q151 G152 W153 K154 G155 S156 P157 A158 I159 F160 Q161 S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 I178 V179 I180 Y181 Q182 Y183 M184 D185 D186 L187 Y188 V189 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 T215 T216 P217 ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP M230 G231 Y232 E233 D237 T240 V241 Q242 P243 I244 V245 L246 P247 E248 K249 D250 S251 W252 T253 V254
N255 D256 I257 Q258 K259 L260 V261 L264 N265 W266 A267 S268 Q269 I270 Y271 P272 G273 I274 K275 V276 R277 Q278 L279 C280 K281 I282 I283 R284 G285 K286 K287 A288 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 V314 H315
G316 V317 Y318 Y319 D320 P321 S322 K323 D324 L325 I326 A327 E328 K331 Q332 G333 Q334 G335 Q336 W337 T338 Y339 Q340 I341 Y342 Q343 E344 P345 F346 K347 N348 L349 K350 T351 G352 K353 Y354 A355 R356 M357 R358 G359 A360 H361 T362 N363 D364 V365 K366 Q367 L368 T369 E370 A371 K374 I375 T376 T377
E378 S379 I382 W383 G384 K385 T386 P387 K388 F389 K390 L391 P392 I393 Q394 K395 W398 E399 W402 T403 E404 Y405 W406 Q407 W410 I411 P412 E413 W414 E415 F416 V417 N418 T419 P420 P421 L422 V423 K424 L425 W426 Y427 Q428 L429 E430 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 ALA 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 VAL THR ARG LYS VAL LEU PHE
LEU ASP GLY ILE ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	117.20Å 154.60Å 155.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	14.13 – 2.90 29.94 – 2.90	Depositor EDS
% Data completeness (in resolution range)	96.7 (14.13-2.90) 90.4 (29.94-2.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.90Å)	Xtriage
Refinement program	BUSTER/TNT	Depositor
R, R_{free}	0.261 , 0.351 0.270 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	66.3	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 84.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.58$, $< L^2 > = 0.44$	Xtriage
Outliers	3 of 30525 reflections (0.010%)	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	8120	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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: I15

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	0.90	3/4562 (0.1%)	1.08	8/6199 (0.1%)
1	B	0.90	0/3510	1.07	8/4772 (0.2%)
All	All	0.90	3/8072 (0.0%)	1.08	16/10971 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	516	GLU	CG-CD	6.18	1.61	1.51
1	A	516	GLU	CB-CG	5.39	1.62	1.52
1	A	432	GLU	CG-CD	5.20	1.59	1.51

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	420	PRO	C-N-CD	-9.69	99.27	120.60
1	A	139	THR	C-N-CD	-9.58	99.53	120.60
1	B	312	GLU	C-N-CD	-9.18	100.40	120.60
1	B	344	GLU	C-N-CD	-7.69	103.68	120.60
1	B	132	ILE	C-N-CD	-7.13	104.90	120.60

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	4445	0	4493	620	0
1	B	3414	0	3443	399	0
2	A	22	0	9	1	0
3	A	152	0	0	25	0
3	B	87	0	0	10	0
All	All	8120	0	7945	989	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 62.

The worst 5 of 989 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:104:LYS:HB3	1:A:192:ASP:HA	1.20	1.18
1:B:103:LYS:HE3	1:B:179:VAL:HG21	1.24	1.15
1:A:64:LYS:HE3	1:A:69:THR:HA	1.22	1.10
1:A:174:GLN:HA	1:A:174:GLN:HE21	1.14	1.07
1:A:288:ALA:HB3	1:A:291:GLU:HB2	1.33	1.06

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	543/566 (96%)	415 (76%)	86 (16%)	42 (8%)	1 3
1	B	412/566 (73%)	338 (82%)	42 (10%)	32 (8%)	1 3
All	All	955/1132 (84%)	753 (79%)	128 (13%)	74 (8%)	1 3

5 of 74 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	14	PRO
1	A	90	VAL
1	A	135	ILE
1	A	139	THR

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	487/505 (96%)	356 (73%)	131 (27%)	0	2
1	B	375/505 (74%)	278 (74%)	97 (26%)	0	2
All	All	862/1010 (85%)	634 (74%)	228 (26%)	0	2

5 of 228 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	428	GLN
1	A	520	GLN
1	B	353	LYS
1	A	439	THR
1	A	468	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	367	GLN
1	A	494	ASN
1	B	255	ASN
1	A	480	GLN
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)

1 ligand 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	$\#\|Z\| > 2$
2	I15	A	1546	-	21,24,24	2.42	7 (33%)	27,34,34	1.98	9 (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
2	I15	A	1546	-	-	0/8/8/8	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)
2	A	1546	I15	C10-C4	-7.50	1.45	1.50
2	A	1546	I15	C8-C3	-2.75	1.36	1.41
2	A	1546	I15	C13-C15	-2.67	1.34	1.39
2	A	1546	I15	C4-N9	2.50	1.37	1.33
2	A	1546	I15	C12-C6	3.01	1.44	1.38

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1546	I15	C12-C6-C2	-3.13	115.23	122.38
2	A	1546	I15	C16-C14-C11	-3.12	115.21	119.47
2	A	1546	I15	C13-C15-C18	-2.07	116.90	119.51
2	A	1546	I15	C11-O5-C2	2.32	122.50	118.47
2	A	1546	I15	C17-C15-C18	2.36	122.50	119.51

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1546	I15	1	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	545/566 (96%)	0.42	42 (7%) 16 11	34, 59, 86, 107	0
1	B	416/566 (73%)	0.60	50 (12%) 6 3	36, 58, 95, 119	0
All	All	961/1132 (84%)	0.50	92 (9%) 10 6	34, 59, 90, 119	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	357	MET	9.1
1	B	284	ARG	5.6
1	A	285	GLY	4.8
1	A	291	GLU	4.7
1	A	290	THR	4.6

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
2	I15	A	1546	22/22	0.95	0.21	-0.51	42,51,55,57	0

6.5 Other polymers (i)

There are no such residues in this entry.