wwPDB X-ray Structure Validation Summary Report (i)

May 19, 2020 – 02:28 pm BST

PDB ID	:	2GHO
Title	:	Recombinant Thermus aquaticus RNA polymerase for Structural Studies
Authors	:	Lamour, V.; Darst, S.A.
Deposited on	:	2006-03-27
Resolution	:	5.00 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 following versions of software and data (see references (1)) were used in the production of this report:

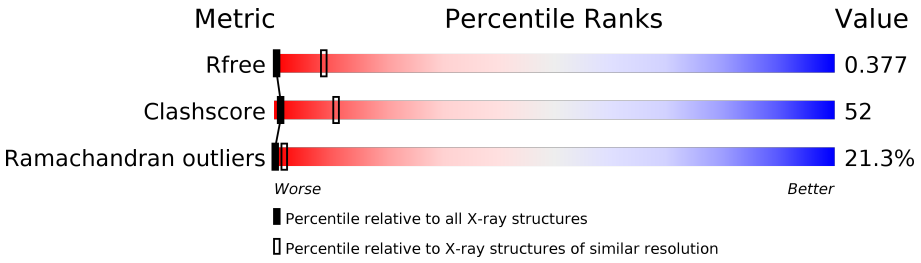
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 5.00 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	1140 (6.20-3.80)
Clashscore	141614	1000 (6.16-3.82)
Ramachandran outliers	138981	1146 (6.20-3.80)

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 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%

Mol	Chain	Length	Quality of chain
1	A	314	48% 19% 6% 27%
1	B	314	51% 17% • 28%
2	C	1119	63% 29% 7%
3	D	1233	48% 33% 15% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11060 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	0	0	0
			920	460	230	230			
1	B	225	Total	C	N	O	0	0	0
			900	450	225	225			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	1114	Total	C	N	O	0	0	0
			4456	2228	1114	1114			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta',DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	D	1196	Total	C	N	O	0	0	0
			4784	2392	1196	1196			

There are 2 discrepancies between the modelled and reference sequences:

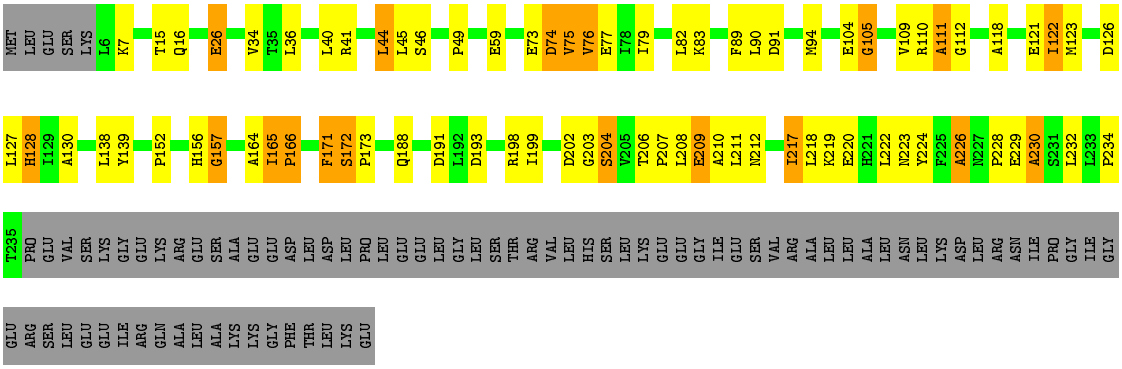
Chain	Residue	Modelled	Actual	Comment	Reference
D	159	GLY	-	linker	UNP Q9KWU6
D	160	GLY	-	linker	UNP Q9KWU6

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 the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

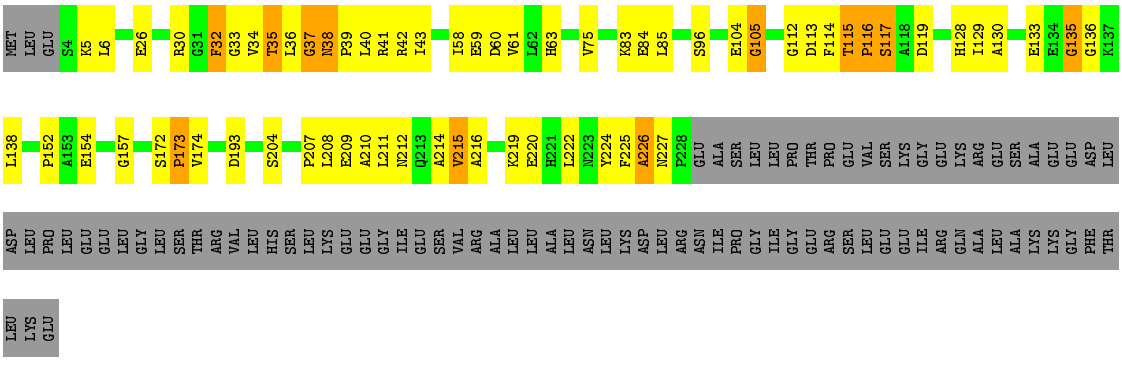
- Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



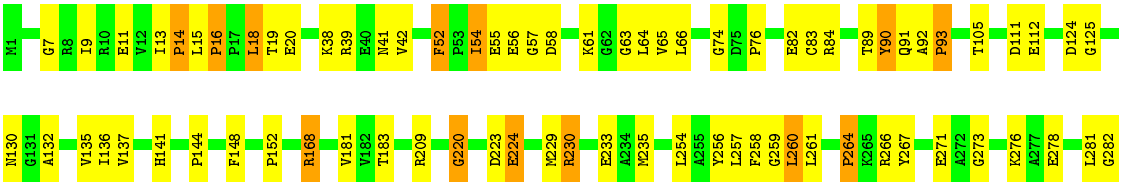
- Molecule 1: DNA-directed RNA polymerase subunit alpha

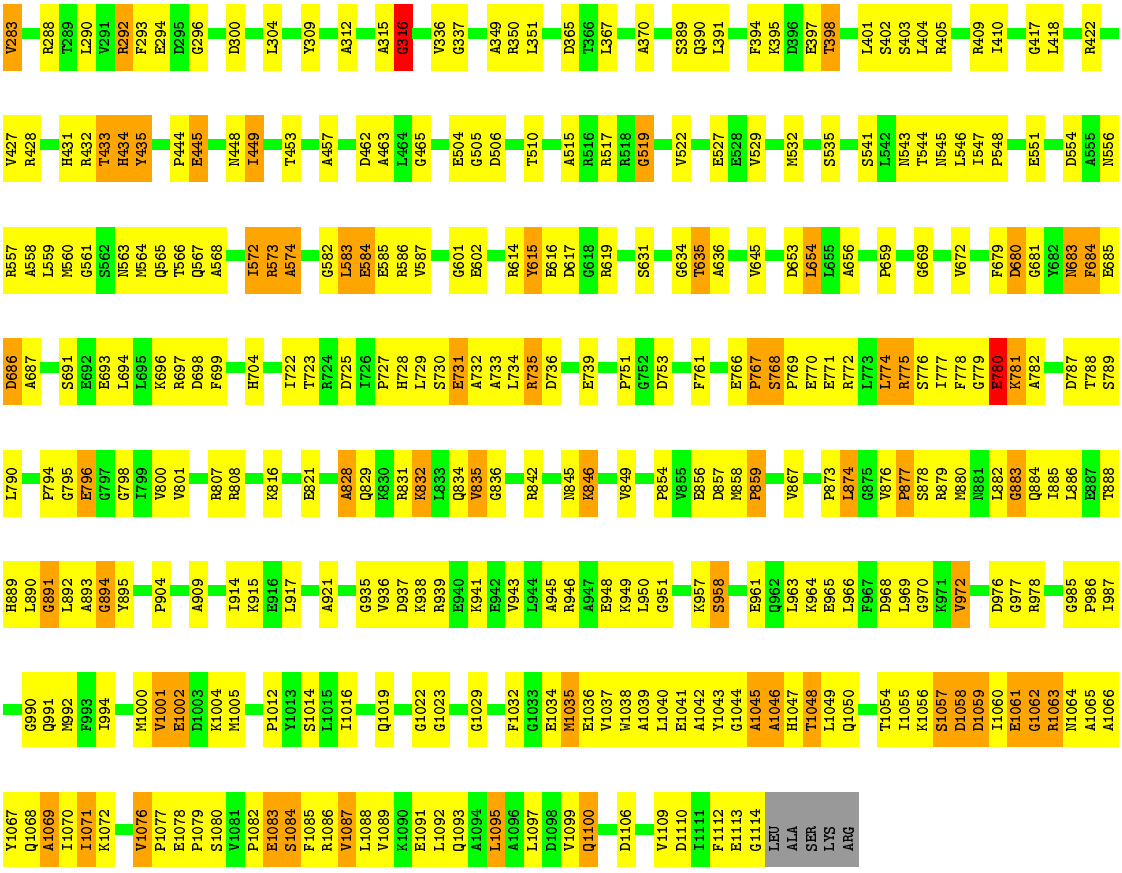
Chain B:



- Molecule 2: DNA-directed RNA polymerase subunit beta

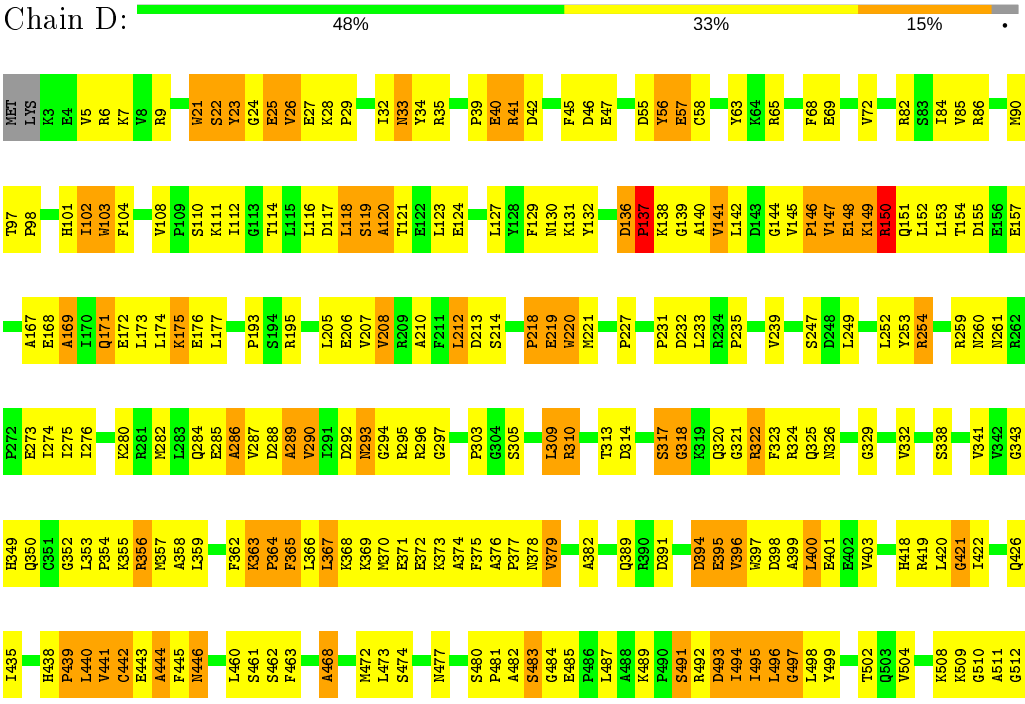
Chain C:

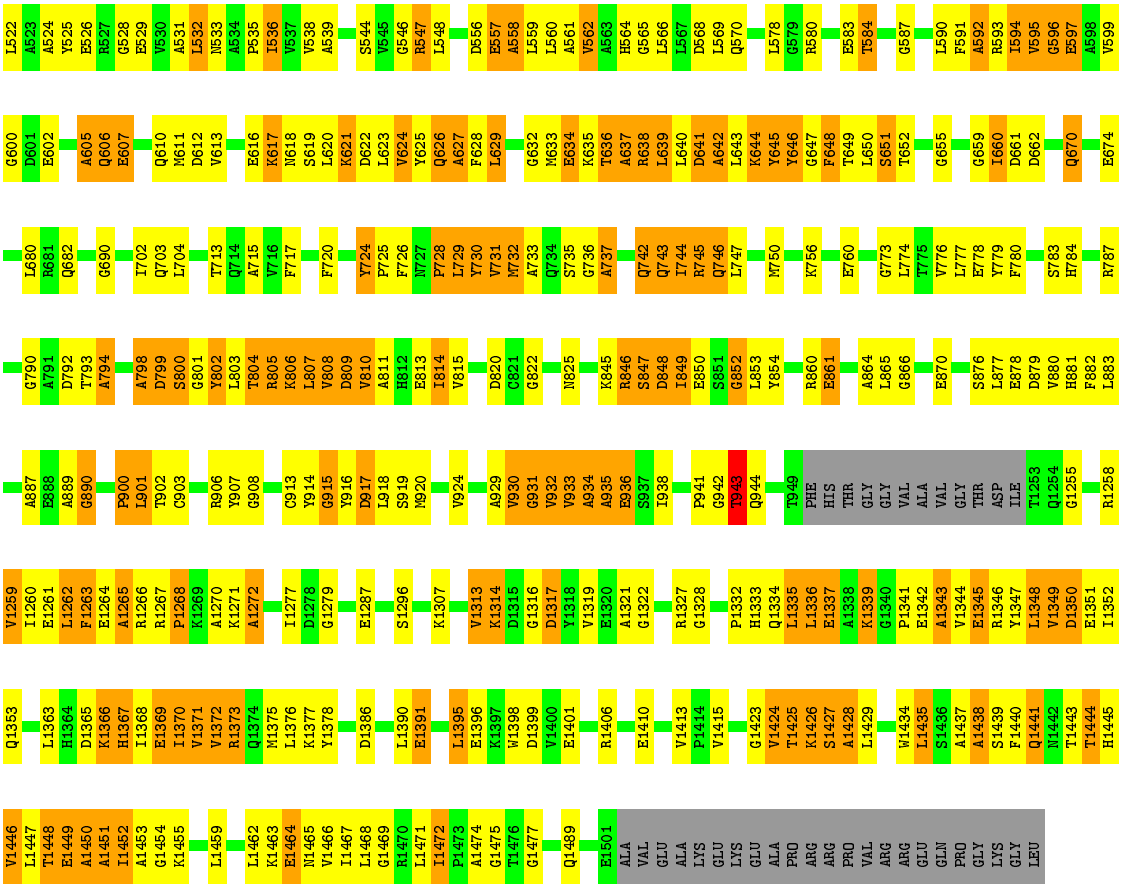




• Molecule 3: DNA-directed RNA polymerase subunit beta',DNA-directed RNA polymerase subunit beta'

Chain D:




L522 A523 A524 Y525 E526 R527 G528 E529 V530 A531 L532 N533 A534 P535 I536 V537 V538 A539 S544 V545 G546 R547 L548 D556 E557 A558 L559 L560 A561 V562 A563 H564 G565 L566 L567 D568 L569 Q570 L578 G579 R580 E583 T584 G587 L590 F591 A592 R593 I594 V595 G596 E597 A598 V599
G600 D601 E602 A605 Q606 E607 Q610 M611 D612 V613 E616 K617 N618 S619 L620 K621 D622 L623 V624 Y625 Q626 A627 F628 L629 G632 M633 E634 K635 T636 A637 R638 L639 L640 D641 A642 L643 K644 Y645 Y646 G647 F648 T649 L650 S651 T652 G655 G659 I660 D661 D662 Q670 E674
L680 R681 Q682 G690 I702 Q703 L704 T713 Q714 A715 V716 F717 F720 Y724 P725 F726 N727 P728 L729 Y730 V731 M732 A733 Q734 S735 G736 A737 Q742 Q743 I744 R745 Q746 L747 M750 K756 E760 G773 L774 T775 V776 L777 E778 Y779 F780 S783 H784 R787
G790 A791 D792 T793 A794 A798 D799 S800 G801 Y802 L803 T804 R805 K806 L807 V808 D809 V810 A811 H812 E813 I814 V815 D820 C821 G822 N825 K845 R846 S847 D848 I849 E850 S851 G852 L853 Y854 R860 E861 A864 L865 G866 E870 S876 L877 E878 D879 V880 H881 F882 L883
A887 E888 A889 G890 P900 L901 T902 C903 R906 Y907 G908 C913 Y914 G915 Y916 D917 I918 S919 N920 V924 A929 V930 G931 V932 V933 A934 A935 E936 S937 I938 P941 G942 I943 Q944 T949 PHE HIS THR GLY GLY VAL ALA VAL GLY THR ASP ILE T1253 Q1254 G1255 R1258
V1259 I1260 E1261 L1262 F1263 E1264 A1265 R1266 R1267 P1268 K1269 A1270 K1271 A1272 I1277 D1278 G1279 E1287 S1296 K1307 V1313 K1314 D1315 G1316 D1317 Y1318 V1319 E1320 A1321 G1322 R1327 G1328 P1332 H1333 Q1334 L1335 L1336 E1337 A1338 K1339 G1340 P1341 E1342 A1343 V1344 E1345 R1346 Y1347 L1348 V1349 D1350 E1351 I1352
Q1353 L1363 H1364 D1365 K1366 H1367 I1368 E1369 I1370 V1371 V1372 R1373 Q1374 M1375 L1376 K1377 Y1378 D1386 L1390 E1391 L1395 E1396 K1397 W1398 D1399 Y1400 E1401 R1406 E1410 V1413 P1414 V1415 G1423 V1424 T1425 K1426 S1427 A1428 L1429 W1434 L1435 S1436 A1437 A1438 S1439 F1440 Q1441 N1442 T1443 T1444 H1445
V1446 L1447 T1448 E1449 A1450 A1451 I1452 A1453 G1454 K1455 L1459 L1462 K1463 E1464 N1465 V1466 I1467 L1468 G1469 R1470 L1471 I1472 P1473 A1474 G1475 T1476 G1477 Q1489 E1501 ALA VAL GLU ALA LYS GLU LYS GLU ALA PRO ARG ARG PRO VAL ARG ARG GLU GLN PRO GLY LYS GLY LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	202.80Å 202.80Å 326.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 5.00 24.99 – 4.10	Depositor EDS
% Data completeness (in resolution range)	87.1 (25.00-5.00) 80.7 (24.99-4.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 4.10Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.336 , 0.337 0.383 , 0.377	Depositor DCC
R_{free} test set	3069 reflections (7.06%)	wwPDB-VP
Wilson B-factor (Å^2)	138.4	Xtriage
Anisotropy	0.324	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.04 , -9.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.79	EDS
Total number of atoms	11060	wwPDB-VP
Average B, all atoms (Å^2)	193.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% 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

5.1 Standard geometry

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	0.35	0/919	0.77	0/1147
1	B	0.40	0/899	0.78	0/1122
2	C	0.55	7/4455 (0.2%)	0.93	9/5567 (0.2%)
3	D	0.56	8/4782 (0.2%)	1.03	22/5974 (0.4%)
All	All	0.53	15/11055 (0.1%)	0.95	31/13810 (0.2%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	704	HIS	C-N	11.46	1.60	1.34
3	D	943	THR	C-N	-9.94	1.11	1.34
2	C	828	ALA	C-N	7.47	1.51	1.34
3	D	137	PRO	N-CA	-7.25	1.34	1.47
3	D	1435	LEU	C-N	-6.64	1.18	1.34

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	137	PRO	CA-C-N	-15.02	84.17	117.20
3	D	137	PRO	O-C-N	10.21	139.04	122.70
2	C	781	LYS	C-N-CA	-9.35	98.32	121.70
3	D	140	ALA	C-N-CA	9.26	144.84	121.70
3	D	151	GLN	CA-C-N	-9.15	97.06	117.20

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	920	0	246	39	0
1	B	900	0	242	34	0
2	C	4456	0	1247	227	0
3	D	4784	0	1309	435	0
All	All	11060	0	3044	732	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 52.

The worst 5 of 732 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)
3:D:145:VAL:O	3:D:146:PRO:O	1.56	1.19
2:C:775:ARG:O	2:C:778:PHE:N	1.88	1.05
3:D:1438:ALA:O	3:D:1440:PHE:N	1.98	0.94
2:C:775:ARG:O	2:C:779:GLY:N	2.00	0.94
3:D:917:ASP:C	3:D:919:SER:H	1.68	0.87

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	228/314 (73%)	138 (60%)	48 (21%)	42 (18%)	0 2
1	B	223/314 (71%)	133 (60%)	58 (26%)	32 (14%)	0 4
2	C	1112/1119 (99%)	640 (58%)	275 (25%)	197 (18%)	0 3
3	D	1192/1233 (97%)	536 (45%)	339 (28%)	317 (27%)	0 0
All	All	2755/2980 (92%)	1447 (52%)	720 (26%)	588 (21%)	0 2

5 of 588 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	7	LYS
1	A	45	LEU
1	A	59	GLU
1	A	74	ASP
1	A	75	VAL

5.3.2 Protein sidechains

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	D	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	1435:LEU	C	1436:SER	N	1.18
1	D	943:THR	C	944:GLN	N	1.11

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.