



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:17 AM GMT

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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

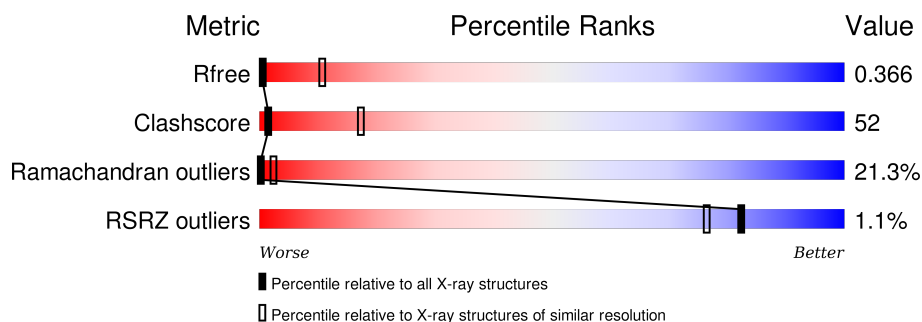
1 Overall quality at a glance

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}	91344	1119 (6.22-3.60)
Clashscore	102246	1019 (6.22-3.66)
Ramachandran outliers	100387	1158 (6.22-3.60)
RSRZ outliers	91569	1122 (6.22-3.60)

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 $\leq 5\%$. 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	314	<div> <div></div> <div>48%19%6%27%</div> </div>
1	B	314	<div> <div></div> <div>51%17%28%</div> </div>
2	C	1119	<div> <div></div> <div>63%29%7%</div> </div>
3	D	1233	<div> <div></div> <div>48%33%15%</div> </div>

2 Entry composition

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 alpha chain.

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 beta chain.

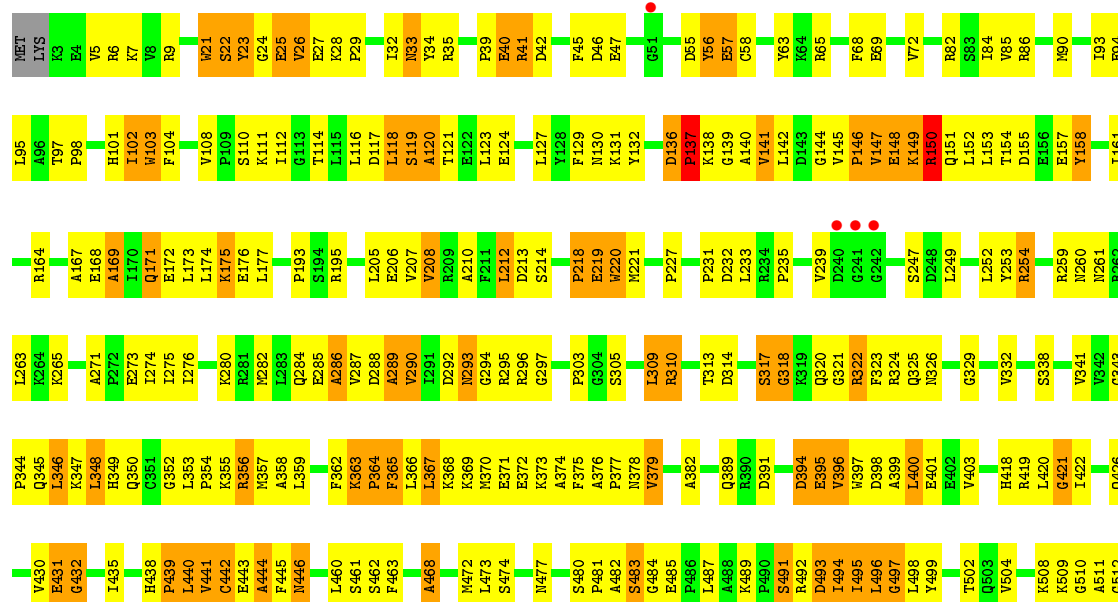
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	1114	Total	C	N	O	0	0	0
			4456	2228	1114	1114			

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

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



ARG	A1438	E1345	ILE	E378	F780	D662	V595	M513
GLU	S1439	R1346	T1253	D879	F780	D662	G596	A514
GLN	F1440	Y1347	Q1254	V880	S783	Q670	E597	
PRO	Q1441	L1348	G1255	H881	H784		A598	P518
GLY	N1442	V1349		F882		E674	V599	
LYS	T1443	D1350	R1258	L883	R787		G600	L522
GLY	T1444	E1351	V1259			L680	D601	A523
LEU	H1445	I1352	I1260	A887	G790		E802	A524
	V1446	Q1353	E1261	A889	A791	Q682		Y525
	L1447		L1262	T793	T793	G690	A605	E526
	T1448	L1363	F1263	G890	A794		Q606	A527
	E1449	H1364	E1264			I702	E607	G528
	A1450	D1365	A1265	P800	A798	Q703	Q610	E529
	A1451	K1366	R1266	L901	A798	Q703	M611	V530
	I1452	H1367	R1267	T902	D799	L704	M611	A531
	A1453	I1368	P1268	C903	S800		D612	L532
	G1454	E1369	K1269	R906	G801	T713	V613	N533
	K1455	I1370	A1270	Y907	L803	Q714	A534	
		V1371	K1271	G908	T804	A715	P535	
		R1372	A1272			F717	V537	
		R1373						
		Q1374	I1277	C913	K806		Q610	
		M1375	D1278	Y914	L807		M611	
		L1376	G1279	G915	V808		D612	
		K1377		Y916	D809		V613	
		Y1378	G1286	D917	V810		E616	
			E1287	L918	A811		K617	
				S919			M618	
			S1296	N920	E813		S619	
					I814		L620	
			K1307	Y924	V815		A627	
			D1308				F628	
			V1313	A929	D820		L629	
			K1314	V930	A733		G632	
			D1315	G931	Q734		M633	
			G1316	V932	S735		L560	
			D1317	V933	G736		E634	
			Y1318	A934			K635	
			E1319				T636	
			A1320	A935			A637	
			G1322	E936			R638	
				S937			G665	
				I938			L639	
							L640	
							D641	
							D568	
							A642	
				P941			L643	
				G942			K644	
				T943			Y645	
				Q944			Y646	
							G647	
				T949			F648	
				PHE			T649	
				HIS			L650	
				THR			E583	
				GLY			T584	
				VAL			T652	
				ALA			G587	
				VAL			G655	
				VAL				L590
				GLY				F591
				THR				A592
								R593
				ASP				I594

4 Data and refinement statistics

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
$\langle I/\sigma(I) \rangle$ ¹	2.05 (at 4.10Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.336 , 0.337 0.377 , 0.366	Depositor DCC
R_{free} test set	1580 reflections (3.64%)	DCC
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
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 43458 reflections	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.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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(°)	Ideal(°)
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 [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	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

5.3.1 Protein backbone

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	3
1	B	223/314 (71%)	133 (60%)	58 (26%)	32 (14%)	0	6
2	C	1112/1119 (99%)	640 (58%)	275 (25%)	197 (18%)	0	4
3	D	1192/1233 (97%)	536 (45%)	339 (28%)	317 (27%)	0	1
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 [i](#)

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

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](#)

There are no ligands in this entry.

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	230/314 (73%)	-0.58	2 (0%) 85 80	194, 194, 194, 194	0
1	B	225/314 (71%)	-0.60	1 (0%) 93 90	194, 194, 194, 194	0
2	C	1114/1119 (99%)	-0.44	11 (0%) 84 78	194, 194, 194, 194	0
3	D	1196/1233 (96%)	-0.60	16 (1%) 79 72	194, 194, 194, 194	0
All	All	2765/2980 (92%)	-0.53	30 (1%) 82 76	194, 194, 194, 194	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	241	GLY	7.9
3	D	242	GLY	6.7
3	D	1315	ASP	6.2
2	C	321	GLU	3.3
3	D	1501	GLU	3.2

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 carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.