



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 30, 2024 – 06:59 PM EST

PDB ID : 1HQM  
Title : CRYSTAL STRUCTURE OF THERMUS AQUATICUS CORE RNA POLYMERASE-INCLUDES COMPLETE STRUCTURE WITH SIDE-CHAINS (EXCEPT FOR DISORDERED REGIONS)-FURTHER REFINED FROM ORIGINAL DEPOSITION-CONTAINS ADDITIONAL SEQUENCE INFORMATION  
Authors : Minakhin, L.; Bhagat, S.; Brunning, A.; Campbell, E.A.; Darst, S.A.; Ebright, R.H.; Severinov, K.  
Deposited on : 2000-12-18  
Resolution : 3.30 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](mailto: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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

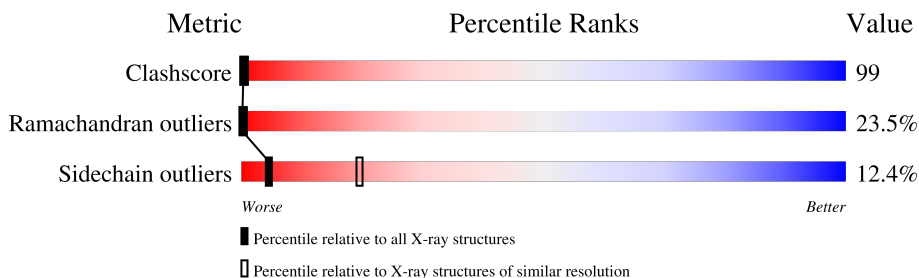
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)

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  $\geq 3$ , 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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	313	9% 41% 18% . 29%
1	B	313	14% 38% 19% . 27%
2	C	1119	15% 56% 26% ..
3	D	1265	17% 54% 20% . 7%
4	E	99	12% 68% 19% .

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 21254 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
			Total	C	N	O	S			
1	A	223	1750	1118	302	328	2	0	0	0
1	B	229	1776	1135	305	334	2	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	deletion	UNP Q9KWU8
A	93	ARG	MET	conflict	UNP Q9KWU8
A	94	TRP	ALA	conflict	UNP Q9KWU8
A	95	ARG	SER	conflict	UNP Q9KWU8
A	111	VAL	GLY	conflict	UNP Q9KWU8
B	?	-	LYS	deletion	UNP Q9KWU8
B	93	ARG	MET	conflict	UNP Q9KWU8
B	94	TRP	ALA	conflict	UNP Q9KWU8
B	95	ARG	SER	conflict	UNP Q9KWU8
B	111	VAL	GLY	conflict	UNP Q9KWU8

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1113	8508	5386	1514	1585	23	12	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2	LYS	GLU	conflict	UNP Q9KWU7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1175	8499	5328	1549	1595	27	17	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	119	PHE	SER	conflict	UNP Q9KWU6
D	863	THR	VAL	conflict	UNP Q9KWU6
D	866	THR	VAL	conflict	UNP Q9KWU6
D	876	ASN	SER	conflict	UNP Q9KWU6
D	947	ILE	-	insertion	UNP Q9KWU6
D	1010	ASN	LYS	conflict	UNP Q9KWU6
D	1117	LYS	ASN	conflict	UNP Q9KWU6
D	1389	PRO	ARG	conflict	UNP Q9KWU6

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	98	719	453	132	130	4	0	0	0

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	1	Total	Mg	0	0
			1	1		

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

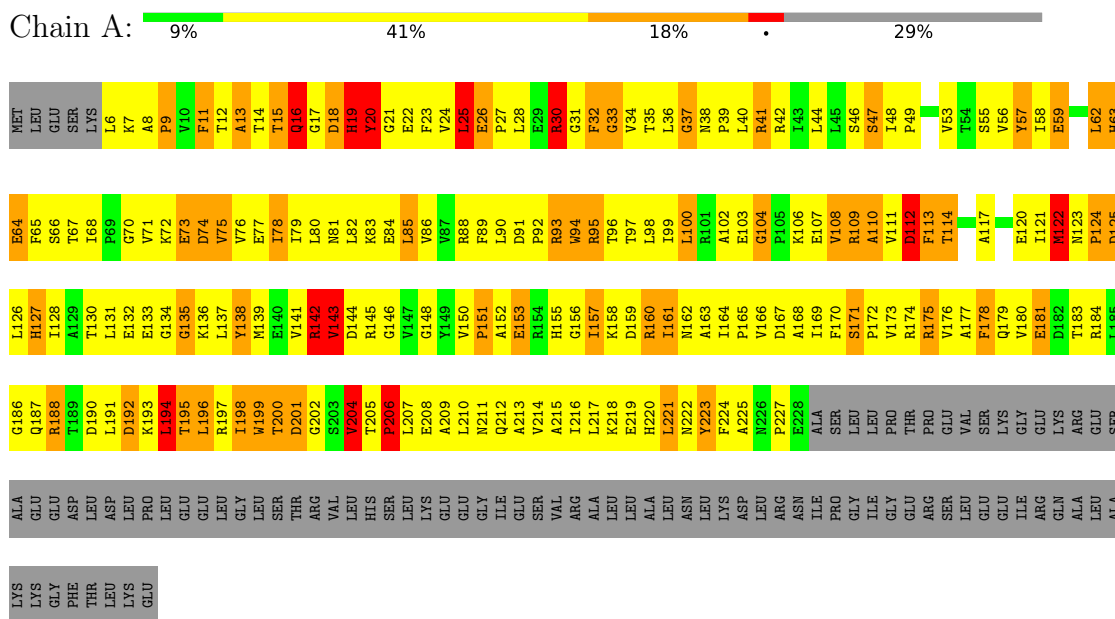
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	1	Total	Zn	0	0
			1	1		

### 3 Residue-property plots

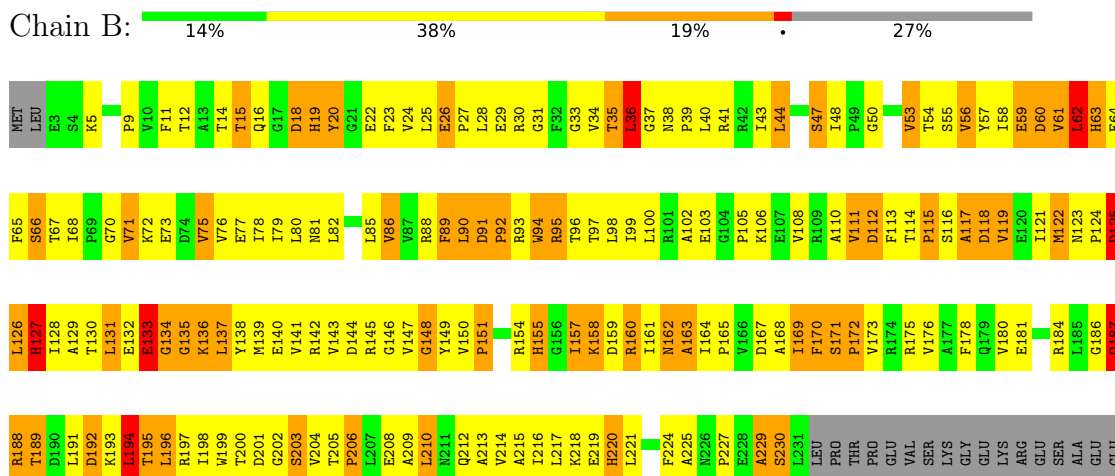
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded 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.

Note EDS was not executed.

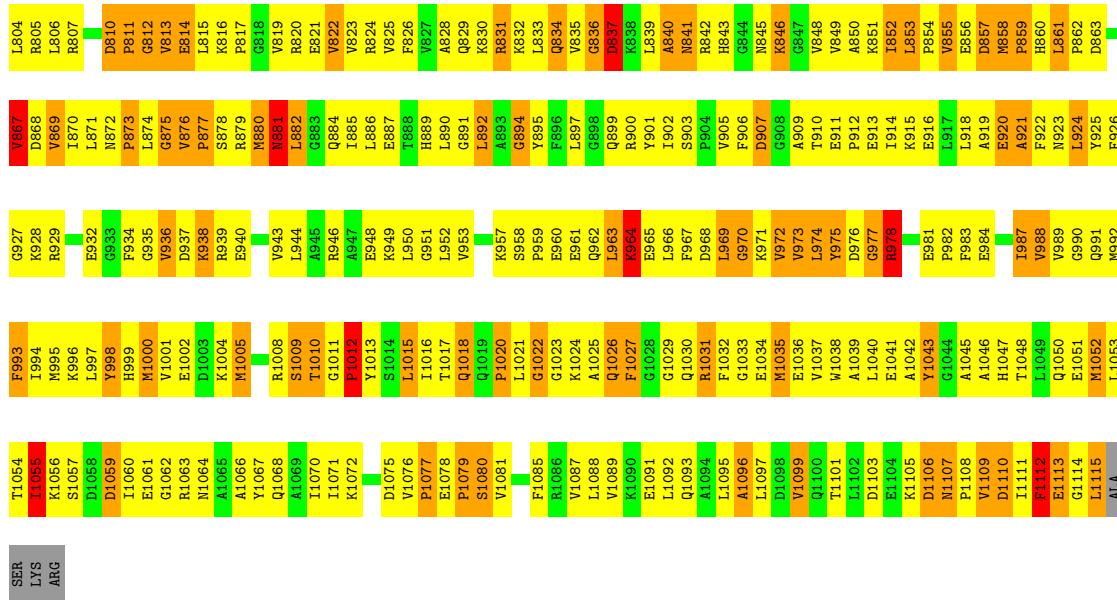
- Molecule 1: DNA-directed RNA polymerase subunit alpha



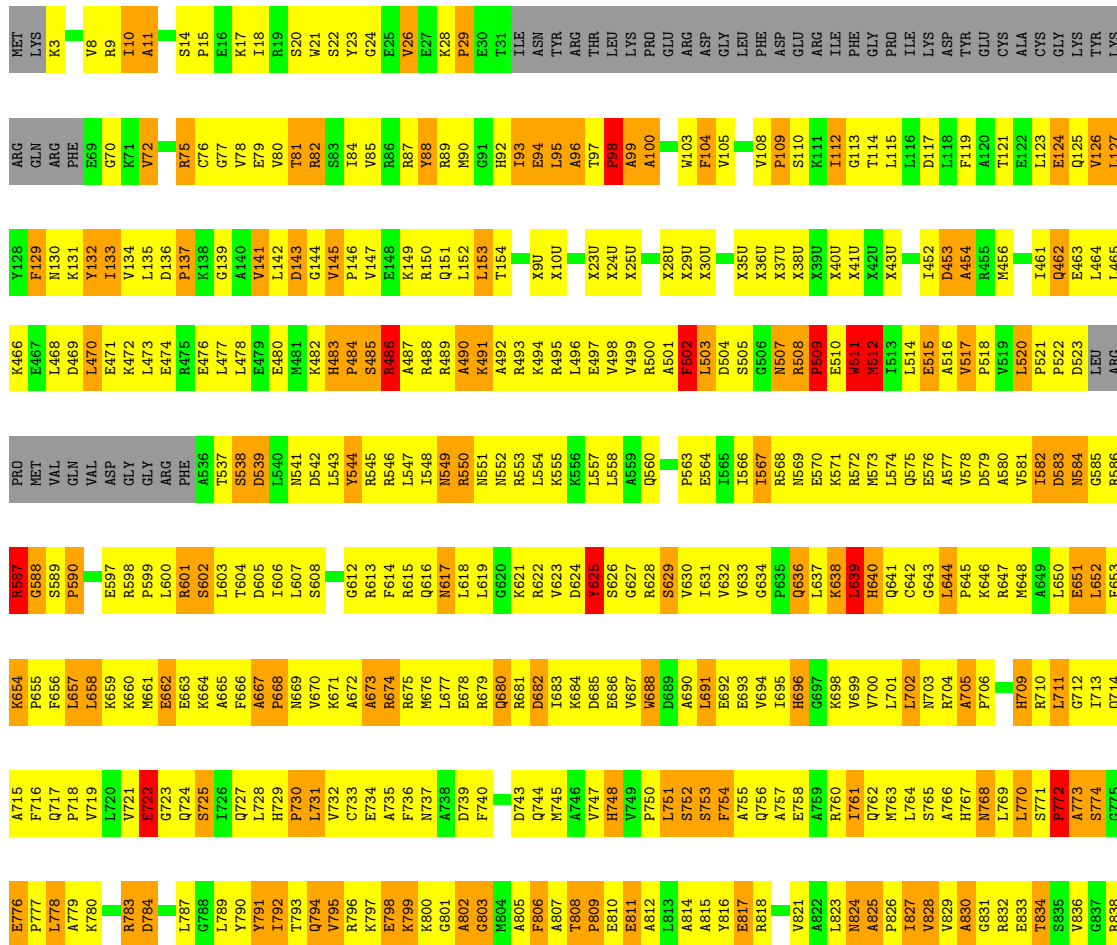
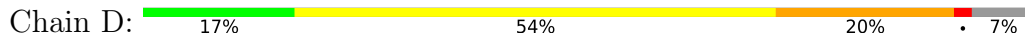
- Molecule 1: DNA-directed RNA polymerase subunit alpha





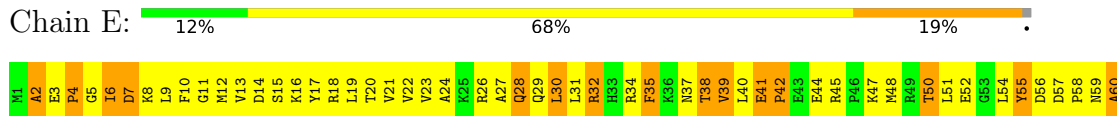


• Molecule 3: DNA-directed RNA polymerase subunit beta'



L839	P905	E966	G1027	E1088	V1149	S1231	A1271	L1336	D1400	G1462	GLY
K840	Q906	E967	G1028	T1089	L1150	M1212	K1272	L1337	V1401	L1463	LEU
F841	E907	A968	R1030	A1090	A1151	A1213	A1273	E1338	E1402	L1464	
K908	K908	D969	R1030	D1091	R1152	R1214	V1274	A1339	A1403	E1465	
N909	N909	R970	S1092	G1093	K1340	I1275	I1276	G1341	N1466	M65	
S910	S910	K971	N1032	G1093	V1154	V1216	S1276	G1341	V1467	V1467	
L911	L911	L972	P1033	Y1094	E1155	S1217	E1277	P1342	L1468	L1468	
K912	K912	R973	Q1034	L1095	A1156	I1218	I1278	E1343	L1469	L1469	
D913	D913	Q974	Q1035	T1096	L1157	I1219	D1279	A1344	A1410	G1470	
L914	L914	R975	I1036	R1097	E1158	E1220	G1280	V1345	GLY	L1471	
V915	V915	E976	R1037	K1098	R1159	A1221	V1281	E1346	GLY	L1472	
Y916	Y916	Q977	Q1038	L1099	R1160	V1222	V1282	R1347	LYS	I1473	
		A978	L1039	V1100	E1163	G1223	V1283	V1348	V1414	P1474	
		Y979	C1040	E1101	G1164	V1224	L1284	L1349	P1415	A1475	
		E980	G1041	V1102	G1164	A1225	E1285	E1352	V1416	G1476	
		E981	M1042	A1103	R1165	V1226	E1286	A1417	T1477	T1477	
		N981	R1043	H1104	V1166	A1227	G1287	I1353	V1418	G1478	
		L982	G1044	E1105	L1167	E1228	G1287	Q1354	K1419	S1479	
		T985	L1045	I1106	L1167	S1229	D1289	K1355	P1420	D1480	
		D986	M1046	V1107	E1170	I1230	R1290	V1356	L1421	F1481	
		K926	Q1047	V1108	D1171	G1231	L1291	Y1357	V1425	T1485	
		T927	K1048	R1109	H1172	E1232	F1300	L1364	K1427	Y1487	
		A928	P1049	E1110	V1173	P1233	S1301	H1365	S1428	V1488	
		R929	S1050	Y990	H1173	G1234	S1292	K1367	A1429	D1489	
		L930	G1051	A1111	F1174	G1235	V1293	H1368	L1430	Q1490	
		L931	G1051	D1112	F1174	G1236	F1294	I1369	S1431	K1494	
		L932	E1052	C1113	L1175	Q1236	V1295	E1370	T1432	A1495	
		D932	T1053	G1114	I1176	L1237	F1300	I1371	K1433	I1496	
		A933	F1054	V1115	A1177	L1238	F1301	V1372	S1434	E1497	
		L934	Q955	Y1118	A1178	T1238	S1301	G1373	V1436	GLY	
		K935	L996	Y1119	A1179	M1239	A1178	Q1375	S1437	ALA	
		Y936	L996	I1119	A1181	R1241	K1302	R1376	A1438	ARG	
		G937	R1058	S1120	E1182	PHE	L1306	M1376	A1439	LYS	
		G938	S1060	E1121	E1183	HIS	P1307	L1377	S1440	GLY	
		F939	S1061	P1122	E1184	THR	L1312	K1378	F1441	ALA	
		T940	F1062	L1123	R1185	GLY	D1315	V1379	M1443	VAL	
		L941	R1063	F1124	R1185	VAL	G1317	V1380	N1443	GLY	
		S942	E1064	Q1125	E1186	ALA	D1318	E1381	T1444	ALA	
		T943	K1003	M1126	V1187	ALA	Y1319	V1382	T1445	LYS	
		T944	V1004	M1127	F1188	VAL	V1320	T1383	H1446	GLY	
		S945	L1006	E1128	V1189	VAL	K1315	D1384	L1448	LYS	
		F946	T1067	V1129	R1190	G1250	D1316	P1385	V1447	VAL	
		I947	L1068	T1130	S1191	T1251	G1317	P1385	L1448	GLY	
		R948	V1008	R1131	P1192	D1252	D1318	G1386	T1449	ALA	
		V885	E1070	R1131	L1193	I1253	G1318	D1387	E1450	PRO	
		T949	M1010	T1132	T1194	T1254	Y1319	V1388	A1451	ARG	
		I950	F1071	L1133	G1195	Q1255	V1320	P1388	A1452	ARG	
		G951	F1072	L1134	Q1196	G1256	E1321	P1389	L1453	ARG	
		I952	I1073	L1135	Q1197	L1257	A1322	L1390	T1453	PRO	
		D953	S1074	R1136	R1198	L1263	A1322	L1391	L1454	VAL	
		D954	H1076	K1137	R1198	P1264	G1323	E1392	G1455	ARG	
		A955	G1077	E1138	Y1199	A1265	Q1324	A1393	L1456	ARG	
		V956	A1078	R1139	G1200	E1266	P1325	Q1394	K1457	GLY	
		I957	F1017	S1139	V1201	I1261	L1326	P1389	A1452	GLN	
		P958	R1079	D1140	C1202	I1262	T1327	L1390	T1453	PRO	
		E959	K1080	I1141	Q1203	L1263	R1328	L1391	L1454	VAL	
		E960	G1081	I1141	K1204	F1264	G1329	E1392	G1455	ARG	
		K961	G1082	S1143	C1205	A1266	I1331	Q1394	L1456	ARG	
		Q962	A1083	G1144	G1207	R1267	D1332	V1395	D1458	GLN	
		R963	D1084	L1145	G1207	R1268	P1333	L1396	D1458	PRO	
		Y964	T1085	Y1146	Y1208	R1288	P1333	E1459	E1459	PRO	
		M1025	A1024	G1147	G1147	R1288	P1333	L1460	E1459	PRO	
		R1026	L1087	R1148	L1210	K1270	Q1335	I1461	T1461	LYS	

Molecule 4: DNA-directed RNA polymerase subunit omega



V61	V62	V63	A64	M65	K66	E67	L68	L69	T70	G71	R72	L73	F74	F75	G76	E77	N78	L79	V80	P81	R82	R83	R84	L85	Q86	K87	E88	M89	E90	L91	L92	Y93	P94	T95	E96	E97	E98	ALA
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Xtrriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	200.76Å 200.76Å 292.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 3.30	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-3.30)	Depositor
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 0.9	Depositor
R, $R_{free}$	0.300 , 0.360	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	21254	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

## 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: ZN, MG

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	0.42	0/1786	0.77	0/2434
1	B	0.39	0/1812	0.74	0/2471
2	C	0.42	0/8672	0.78	5/11752 (0.0%)
3	D	0.42	0/8437	0.78	14/11443 (0.1%)
4	E	0.35	0/730	0.65	0/991
All	All	0.41	0/21437	0.77	19/29091 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
2	C	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	834	THR	N-CA-C	-6.40	93.72	111.00
3	D	137	PRO	N-CA-CB	6.22	110.77	103.30
2	C	580	MET	N-CA-C	6.19	127.70	111.00
2	C	836	GLY	N-CA-C	-6.04	97.99	113.10
2	C	329	GLY	N-CA-C	-5.97	98.17	113.10

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	138	TYR	Sidechain
2	C	975	TYR	Sidechain

## 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	1750	0	1759	403	0
1	B	1776	0	1776	323	0
2	C	8508	0	8418	1886	0
3	D	8499	0	7993	1651	0
4	E	719	0	685	125	0
5	D	1	0	0	0	0
6	D	1	0	0	0	0
All	All	21254	0	20631	4161	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 99.

The worst 5 of 4161 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:1020:PRO:HB2	3:D:1023:VAL:HB	1.20	1.18
2:C:508:ILE:H	2:C:508:ILE:HD13	1.10	1.15
2:C:438:ILE:HG21	2:C:470:PRO:HB3	1.22	1.15
2:C:605:LYS:HG2	2:C:606:VAL:H	1.05	1.14
2:C:262:ALA:HB1	2:C:266:ARG:HD2	1.23	1.14

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	221/313 (71%)	98 (44%)	68 (31%)	55 (25%)	0	0
1	B	227/313 (72%)	109 (48%)	61 (27%)	57 (25%)	0	0
2	C	1111/1119 (99%)	559 (50%)	300 (27%)	252 (23%)	0	0
3	D	1127/1265 (89%)	543 (48%)	319 (28%)	265 (24%)	0	0
4	E	96/99 (97%)	49 (51%)	22 (23%)	25 (26%)	0	0
All	All	2782/3109 (90%)	1358 (49%)	770 (28%)	654 (24%)	0	0

5 of 654 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	26	GLU
1	A	59	GLU
1	A	64	GLU
1	A	73	GLU
1	A	75	VAL

### 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	190/271 (70%)	163 (86%)	27 (14%)	3	15
1	B	191/271 (70%)	171 (90%)	20 (10%)	7	25
2	C	869/936 (93%)	747 (86%)	122 (14%)	3	16
3	D	782/1036 (76%)	693 (89%)	89 (11%)	5	22
4	E	67/88 (76%)	64 (96%)	3 (4%)	27	58
All	All	2099/2602 (81%)	1838 (88%)	261 (12%)	4	19

5 of 261 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1091	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	1138	ARG
4	E	32	ARG
2	C	441	VAL
2	C	429	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 68 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	917	GLN
3	D	1038	GLN
3	D	1446	HIS
2	C	632	ASN
2	C	623	HIS

### 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	D	3

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	155:ASP	C	2(U):UNK	N	55.17
1	D	46(U):UNK	C	452:ILE	N	46.65
1	D	10(U):UNK	C	20(U):UNK	N	14.79

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

EDS was not executed - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates [i](#)

EDS was not executed - this section is therefore empty.

### 6.4 Ligands [i](#)

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers [i](#)

EDS was not executed - this section is therefore empty.