



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 09:03 PM GMT

PDB ID : 1N34  
Title : Structure of the *Thermus thermophilus* 30S ribosomal subunit in the presence of codon and crystallographically disordered near-cognate transfer rna anticodon stem-loop mismatched at the first codon position  
Authors : Ogle, J.M.; Murphy IV, F.V.; Tarry, M.J.; Ramakrishnan, V.  
Deposited on : 2002-10-25  
Resolution : 3.80 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](mailto: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

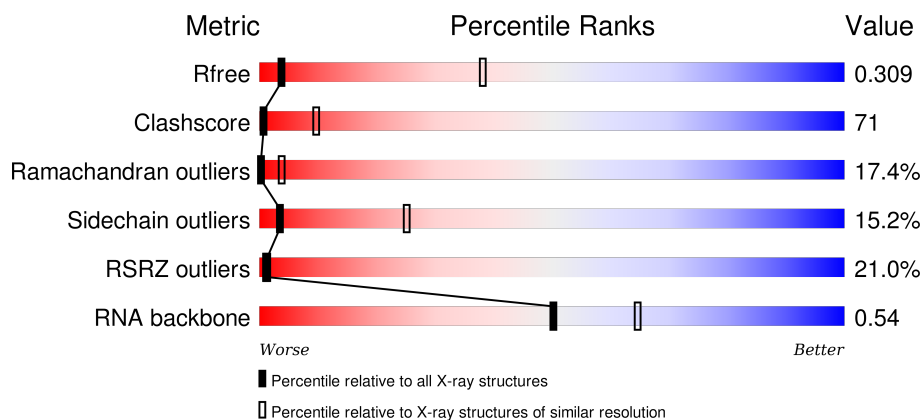
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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))
$R_{free}$	91344	1317 (4.10-3.50)
Clashscore	102246	1458 (4.10-3.50)
Ramachandran outliers	100387	1397 (4.10-3.50)
Sidechain outliers	100360	1392 (4.10-3.50)
RSRZ outliers	91569	1325 (4.10-3.50)
RNA backbone	2183	1070 (4.76-2.80)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	1522	<div> <div>37%</div> <div>10% 72% 17% ..</div> </div>
2	Z	6	<div> <div>17% 50% 33%</div> </div>
3	B	256	<div> <div>2% 8% 58% 24% . 9%</div> </div>
4	C	239	<div> <div>15% 11% 51% 23% . 14%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	D	208	
6	E	161	
7	F	101	
8	G	155	
9	H	138	
10	I	128	
11	J	104	
12	K	129	
13	L	135	
14	M	126	
15	N	60	
16	O	88	
17	P	88	
18	Q	104	
19	R	88	
20	S	92	
21	T	106	
22	V	26	

## 2 Entry composition [i](#)

There are 23 unique types of molecules in this entry. The entry contains 51757 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 RNA chain called 16S RIBOSOMAL RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1513	Total	C	N	O	P	42	0	0
			32508	14472	6016	10509	1511			

- Molecule 2 is a RNA chain called A-SITE MESSENGER RNA FRAGMENT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Z	4	Total	C	N	O	P	0	0	0
			77	36	8	30	3			

- Molecule 3 is a protein called 30S RIBOSOMAL PROTEIN S2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	234	Total	C	N	O	S	0	0	0
			1900	1213	341	341	5			

- Molecule 4 is a protein called 30S RIBOSOMAL PROTEIN S3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	206	Total	C	N	O	S	0	0	0
			1612	1016	314	281	1			

- Molecule 5 is a protein called 30S RIBOSOMAL PROTEIN S4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	D	208	Total	C	N	O	S	0	0	0
			1703	1066	339	291	7			

- Molecule 6 is a protein called 30S RIBOSOMAL PROTEIN S5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	E	150	Total	C	N	O	S	0	0	0
			1146	724	217	201	4			

- Molecule 7 is a protein called 30S RIBOSOMAL PROTEIN S6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	F	101	Total	C	N	O	S	0	0	0
			843	531	155	154	3			

- Molecule 8 is a protein called 30S RIBOSOMAL PROTEIN S7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	G	155	Total	C	N	O	S	0	0	0
			1257	781	252	218	6			

- Molecule 9 is a protein called 30S RIBOSOMAL PROTEIN S8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	H	138	Total	C	N	O	S	0	0	0
			1116	705	215	193	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	25	ASP	GLU	CONFLICT	UNP Q5SHQ2
H	37	ARG	LYS	CONFLICT	UNP Q5SHQ2
H	52	ASP	GLU	CONFLICT	UNP Q5SHQ2
H	61	VAL	ILE	CONFLICT	UNP Q5SHQ2
H	62	TYR	HIS	CONFLICT	UNP Q5SHQ2
H	81	HIS	LYS	CONFLICT	UNP Q5SHQ2
H	88	LYS	ARG	CONFLICT	UNP Q5SHQ2
H	115	SER	PRO	CONFLICT	UNP Q5SHQ2

- Molecule 10 is a protein called 30S RIBOSOMAL PROTEIN S9.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
10	I	127	Total	C	N	O	0	0	0
			1011	639	198	174			

- Molecule 11 is a protein called 30S RIBOSOMAL PROTEIN S10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	J	98	Total	C	N	O	S	0	0	0
			792	498	156	137	1			

- Molecule 12 is a protein called 30S RIBOSOMAL PROTEIN S11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	K	119	Total	C	N	O	S	0	0	0
			885	549	168	165	3			

- Molecule 13 is a protein called 30S RIBOSOMAL PROTEIN S12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	L	124	Total	C	N	O	S	0	0	0
			970	611	195	163	1			

- Molecule 14 is a protein called 30S RIBOSOMAL PROTEIN S13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	M	118	Total	C	N	O	S	0	0	0
			937	579	193	163	2			

- Molecule 15 is a protein called 30S RIBOSOMAL PROTEIN S14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	N	60	Total	C	N	O	S	0	0	0
			492	312	104	72	4			

- Molecule 16 is a protein called 30S RIBOSOMAL PROTEIN S15.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
16	O	88	Total	C	N	O	S	0	0	0
			734	459	147	126	2			

- Molecule 17 is a protein called 30S RIBOSOMAL PROTEIN S16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
17	P	83	Total	C	N	O	S	0	0	0
			700	443	139	117	1			

- Molecule 18 is a protein called 30S RIBOSOMAL PROTEIN S17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
18	Q	104	Total	C	N	O	S	0	0	0
			857	547	161	147	2			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Q	50	LYS	ARG	CONFLICT	UNP Q5SHP7
Q	53	LEU	VAL	CONFLICT	UNP Q5SHP7
Q	62	SER	ALA	CONFLICT	UNP Q5SHP7
Q	79	SER	GLU	CONFLICT	UNP Q5SHP7
Q	82	MET	LEU	CONFLICT	UNP Q5SHP7
Q	90	ILE	VAL	CONFLICT	UNP Q5SHP7
Q	96	GLN	ALA	CONFLICT	UNP Q5SHP7

- Molecule 19 is a protein called 30S RIBOSOMAL PROTEIN S18.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
19	R	73	Total	C	N	O	0	0	0
			597	380	118	99			

- Molecule 20 is a protein called 30S RIBOSOMAL PROTEIN S19.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
20	S	80	Total	C	N	O	S	0	0	0
			647	414	119	112	2			

- Molecule 21 is a protein called 30S RIBOSOMAL PROTEIN S20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
21	T	99	Total	C	N	O	S	0	0	0
			763	470	162	129	2			

- Molecule 22 is a protein called 30S RIBOSOMAL PROTEIN THX.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
22	V	24	Total	C	N	O	0	0	0
			208	128	50	30			

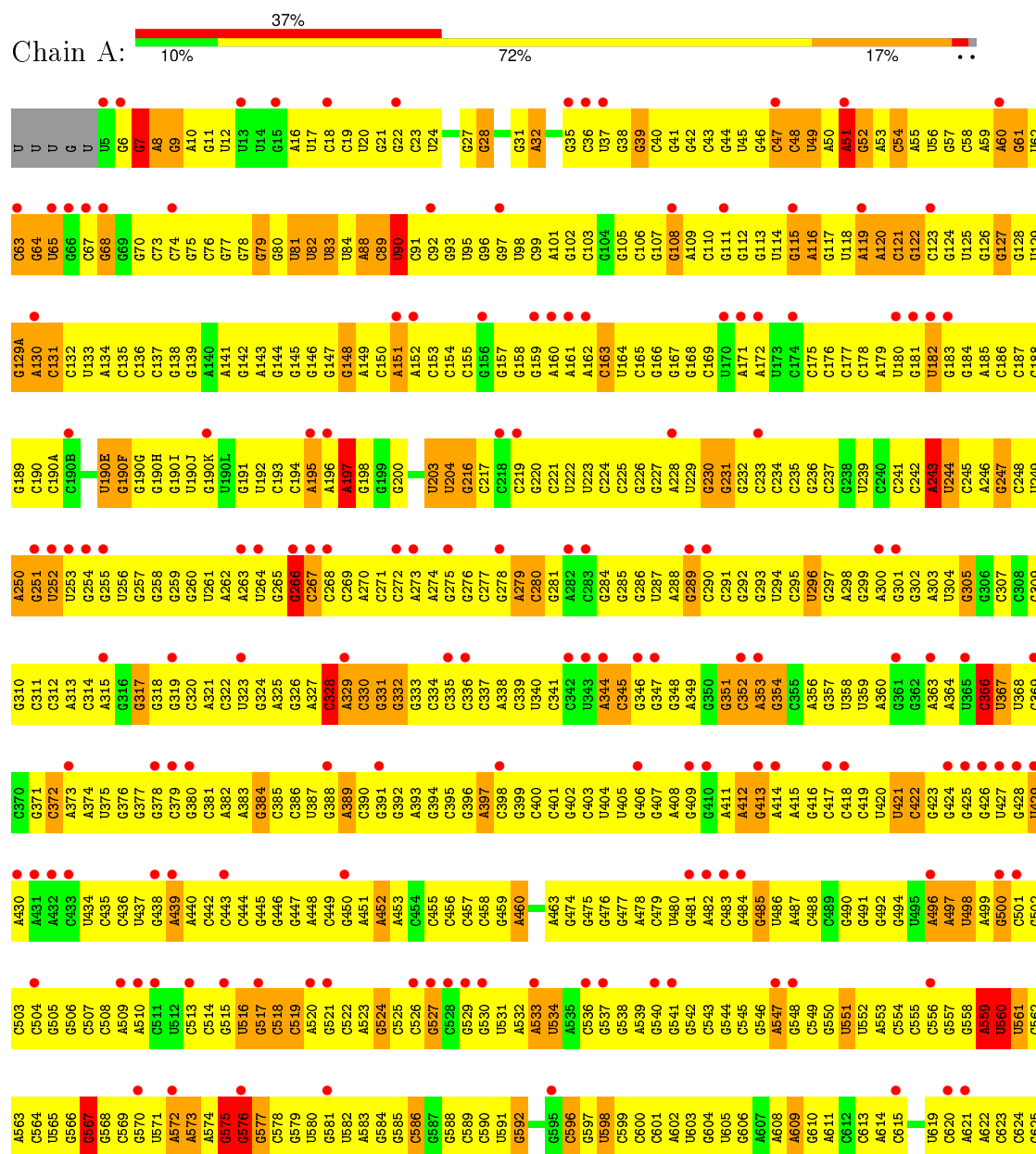
- Molecule 23 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
23	D	1	Total	Zn	0	0
			1	1		
23	N	1	Total	Zn	0	0
			1	1		

### 3 Residue-property plots

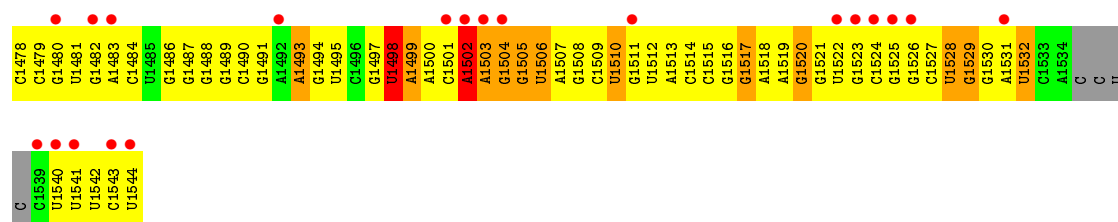
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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 ( $\text{RSRZ} > 2$ ). 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: 16S RIBOSOMAL RNA

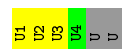
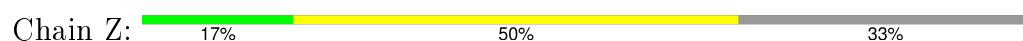




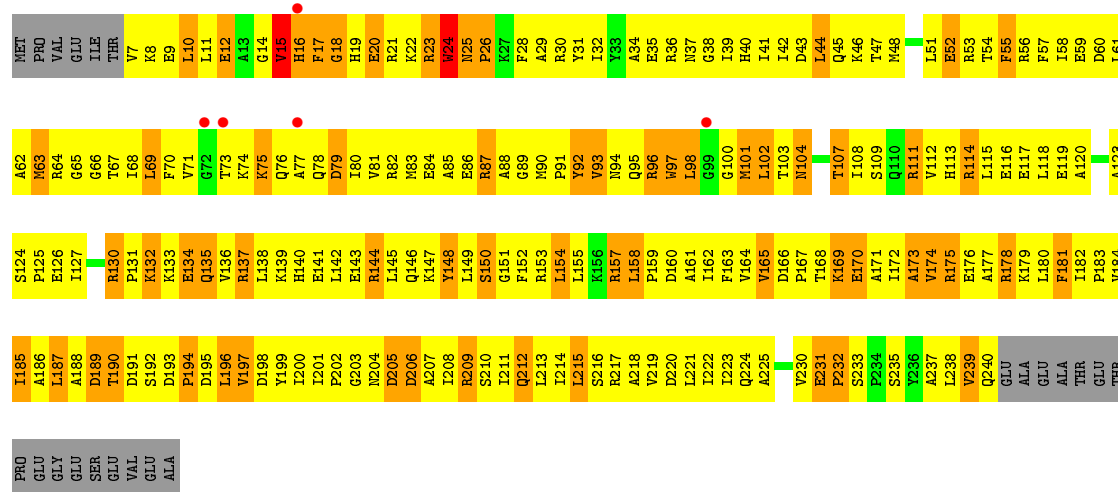
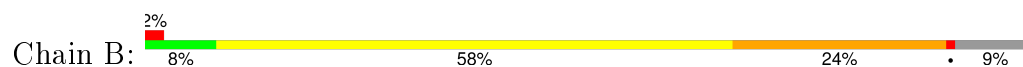
A1413	C1352	G1291	G1171	G1050	C995	A935	A873	C806	A746	U686	U626
U1414	G1353	U1292	C1172	C1051	A996	C996	G874	A807	C747	A687	G627
G1415	G1354	G1293	G1173	U1052	A997	A937	G875	C808	C748	G688	G628
G1416	G1355	G1294	G1174	G1053	G998	A938	G876	C809	C749	G689	G629
G1417	G1356	G1295	G1175	C1054	C999	G939	C877	C810	G750	G690	G630
A1418	A1357	G1296	G1176	U1055	U1000	C940	G878	C811	G751	G691	G631
G1419	U1358	G1297	G1177	U1056	A1001	G941	C879	C812	U682	A682	A632
C1420	A1359	A1298	G1178	G1057	G1002	G942	C880	U913	A752	G692	G633
G1421	A1360	A1299	A1179	G1058	G1003	U943	G881	A814	A753	G693	G634
G1422	G1361	U1300	U1180	C1059	G1003A	G944	C882	A815	C754	A694	C634
G1423	C1361A	U1301	G1119	C1060	A1004	G945	C883	A816	G755	A695	G635
C1424	G1362	G1303	U1121	U1061	C1006	A946	U884	C817	G756	A696	U636
U1425	A1363	G1304	U1122	U1062	G1007	G947	G885	G818	U757	U697	G637
A1426	U1364	G1305	A1123	C1063	C1008	C948	G886	A819	G758	G698	G638
U1427	G1365	A1306	U1124	G1064	C1009	U949	G887	U820	A759	C699	G639
A1428	G1366	U1307	U1125	U1065	G1009	U950	G888	G821	G760	G700	U640
C1429	G1367	U1308	U1126	C1066	G1010	G951	A889	C822	G761	C701	A641
C1430	G1368	A1248	A1187	A1067	G1011	U952	G890	G823	G762	A702	A642
C1431	C1369	G1310	C1189	G1068	U1012	G953	U891	C824	C764	A703	C643
G1432	G1370	G1311	G1190	C1069	G1013	G954	A892	G825	G765	A704	G644
A1433	G1371	A1251	A1191	U1070	A1014	U955	C893	G826	A766	U705	C645
G1434	U1372	A1252	G1192	C1071	A1015	U956	G894	U827	C767	A706	U646
G1435	G1373	G1313	G1193	G1072	A1016	U957	G895	A828	C768	C707	C647
U1436	A1374	C1254	U1194	U1073	G1017	A958	C896	G829	A769	G708	A648
C1437	A1375	U1315	U1195	G1074	C1018	A959	C897	G830	G770	G709	G649
G1438	U1376	C1317	U1196	C1075	C1019	U960	G898	U831	G771	G710	G650
C1439	A1377	A1318	G1197	C1076	U1020	U961	C899	G832	G772	G711	C651
C1440	G1378	A1319	G1198	G1077	G1021	C962	A900	U833	A773	A712	U652
G1441	G1379	C1320	U1199	U1078	G1022	G963	A901	C834	G774	G713	C653
G1442	U1380	G1321	G1200	G1079	G1023	A964	G902	U835	G775	A714	G654
C1443	C1381	C1322	C1140	A1080	U1024	A965	G903	G836	G776	A715	A655
A1444	C1382	G1323	C1141	G1081	U1025	G966	C904	G837	A777	A716	C656
G1445	C1383	A1324	G1142	G1082	G1026	C967	U905	G838	G778	C717	G657
G1446	G1384	C1325	G1143	U1083	C1027	A968	G906	U839	G779	G718	G658
C1447	G1385	C1326	G1144	G1084	C1028	A969	A907	C840	C780	C719	U659
U1448	G1386	G1267	C1145	U1085	G1029	C970	A908	U941	A781	C720	G660
C1449	C1387	C1327	A1146	U1086	U1030	G971	A909	C848	A782	G721	G661
A1450	G1388	A1328	C1147	G1087	G1030A	C972	C910	C849	C783	U723	A662
C1451	C1389	U1330	U1148	G1088	C1030B	G973	U911	U850	C784	G724	A663
C1452	U1390	G1331	C1149	U1089	G1030C	A974	C912	G851	G785	G725	A664
G1453	U1391	A1332	U1150	U1090	A1030D	A975	A913	G852	G786	C726	A665
C1454	G1392	A1333	A1151	U1091	G1031	G976	A914	G853	A787	G727	G666
A1455	U1393	G1334	A1152	A1092	G1032	A977	A915	G854	U788	A728	G667
G1456	C1394	C1335	C1153	A1093	G1033	A978	G916	G855	U789	A729	G668
C1457	A1395	C1336	G1154	U1094	G1034	C979	G917	G856	A790	G730	U669
G1458	A1396	G1337	G1155	U1095	A1035	C980	A918	G858	G791	G731	G670
C1459	G1397	G1338	G1156	C1096	G1036	U981	A919	A859	A792	C732	U672
A1460	A1398	A1339	A1157	C1097	C1037	U982	U920	A860	U793	A733	G673
C1461	C1399	A1340	U1218	G1098	C1038	A983	U921	G861	A794	G734	G674
G1462	U1400	U1341	U1159	C1099	C1039	C884	G922	C862	C795	C735	A675
A1463	G1401	A1280	G1160	U1100	U1040	C985	A923	U863	C796	C736	A676
C1464	C1402	U1281	C1161	A1101	A1041	A986	C924	A864	C797	A737	U677
G1465	U1403	G1343	G1162	A1102	G1042	G987	G925	A865	G798	C738	U678
C1466	G1404	U1344	C1163	C1103	C1043	G988	G926	C866	G799	C739	C679
U1467	G1405	A1285	G1164	A1044	U1044	C989	G927	G867	G800	U740	C680
A1468	U1406	G1347	G1165	A1105	C1045	C990	G928	C868	U801	G741	C681
C1469	C1407	A1287	G1166	G1106	A1046	U991	G929	G869	G742	G742	G682
G1470	U1408	U1288	A1167	C1107	G1047	U992	C932	U870	G803	U743	G683
A1471	G1409	A1350	G1168	G1108	U1048	G993	G933	U871	C744	A684	A684
U1472	C1410	G1351	A1169	C1109	U1049	A994	C934	A872	C805	C745	G685



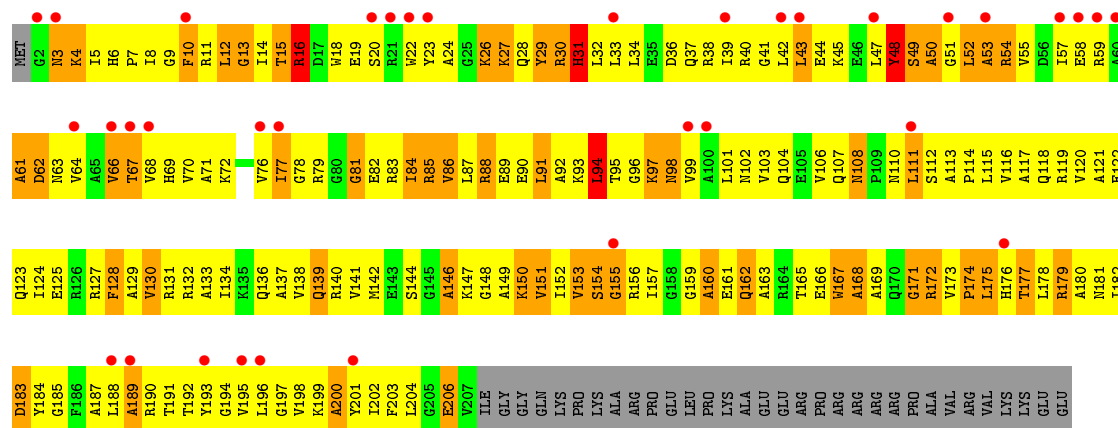
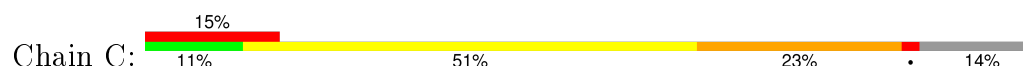
• Molecule 2: A-SITE MESSENGER RNA FRAGMENT



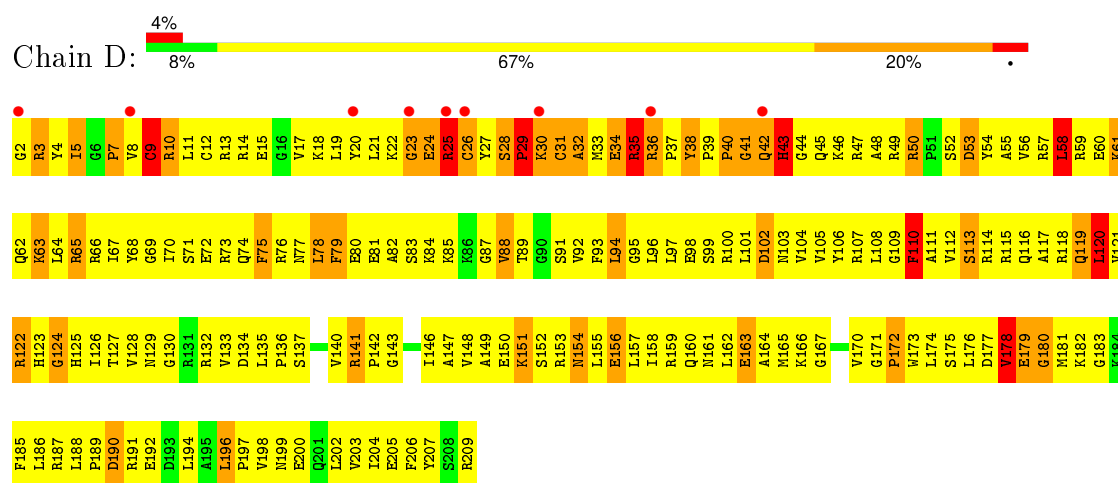
• Molecule 3: 30S RIBOSOMAL PROTEIN S2



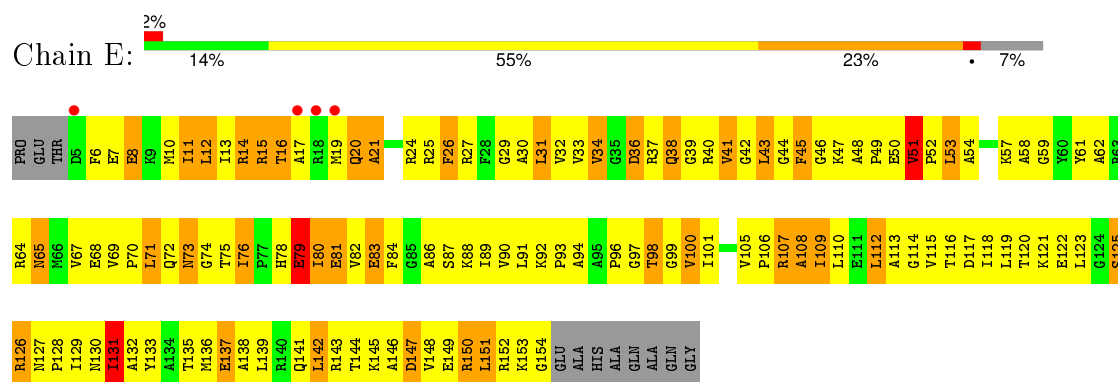
• Molecule 4: 30S RIBOSOMAL PROTEIN S3



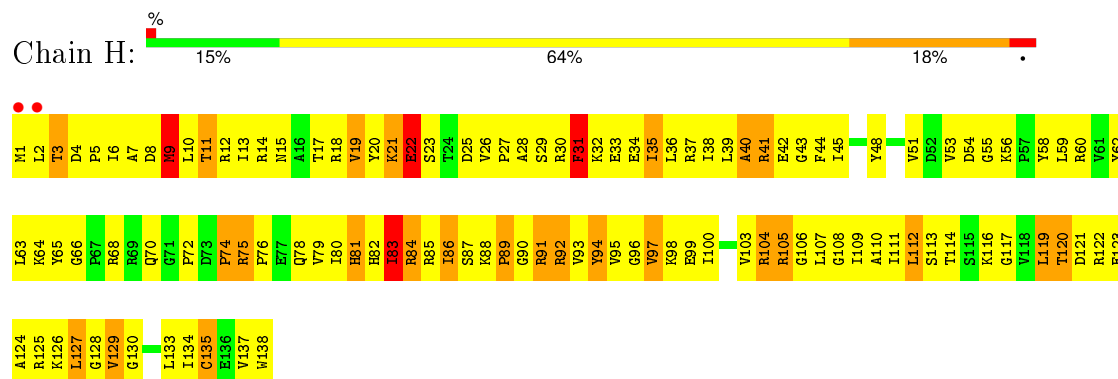
• Molecule 5: 30S RIBOSOMAL PROTEIN S4



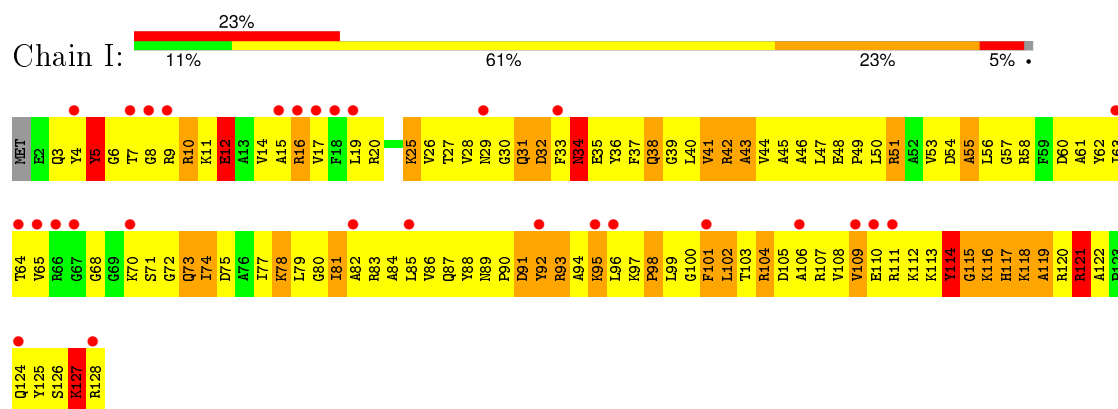
• Molecule 6: 30S RIBOSOMAL PROTEIN S5



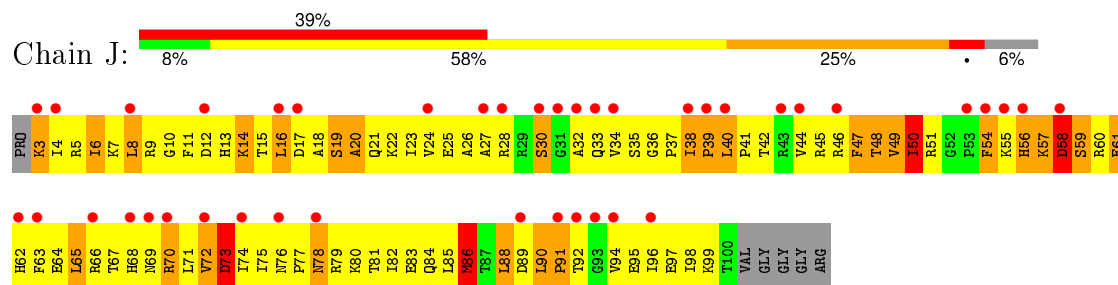
• Molecule 9: 30S RIBOSOMAL PROTEIN S8



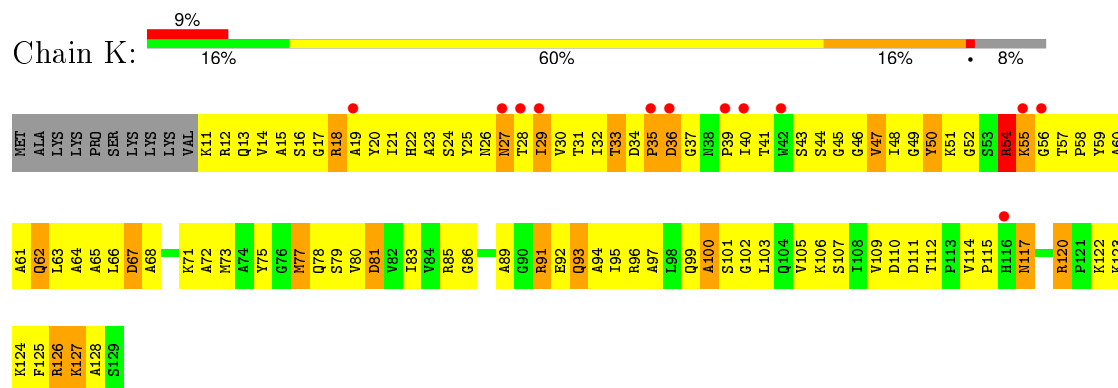
• Molecule 10: 30S RIBOSOMAL PROTEIN S9



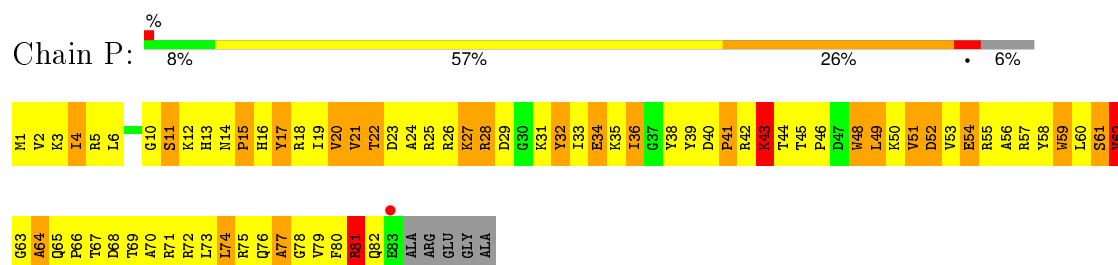
• Molecule 11: 30S RIBOSOMAL PROTEIN S10



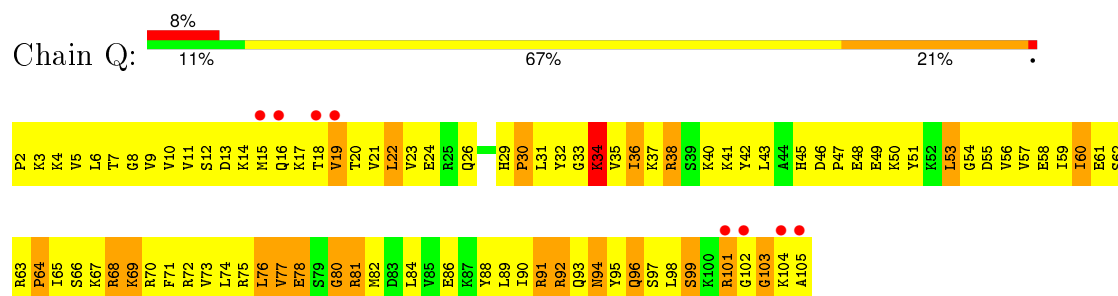
• Molecule 12: 30S RIBOSOMAL PROTEIN S11



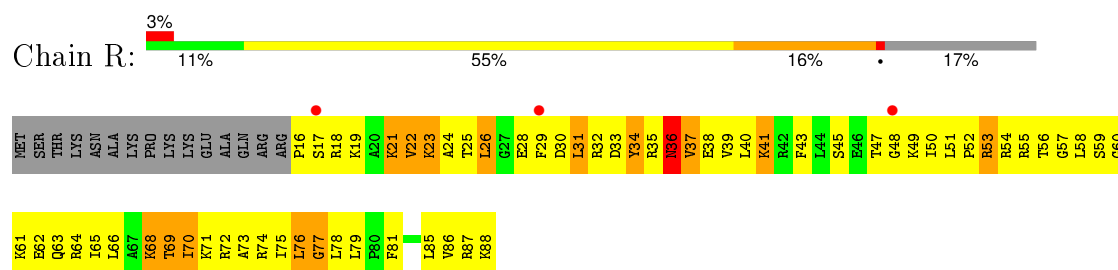
• Molecule 13: 30S RIBOSOMAL PROTEIN S12



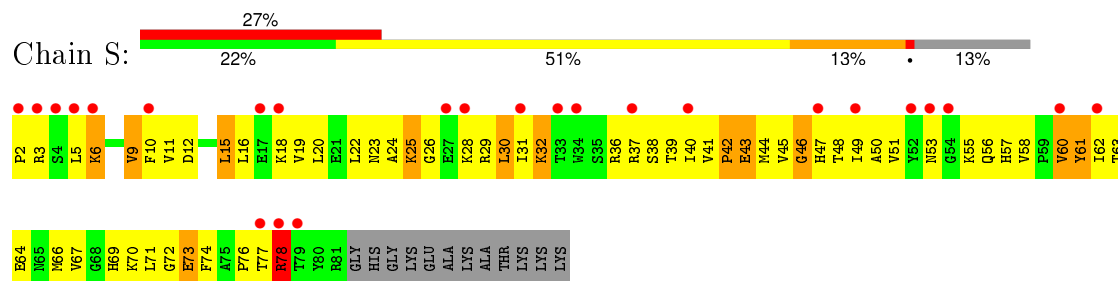
- Molecule 18: 30S RIBOSOMAL PROTEIN S17



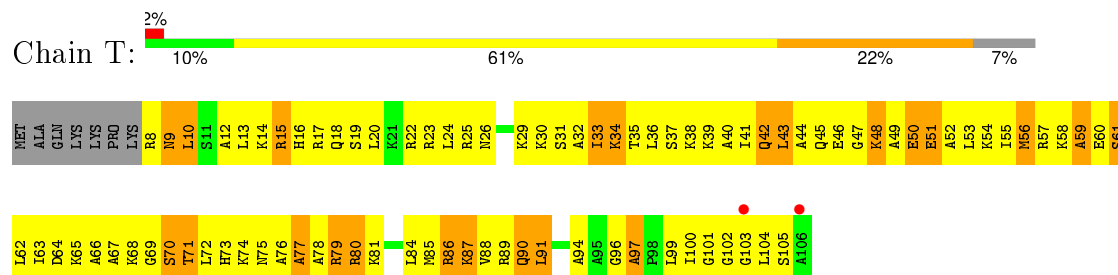
- Molecule 19: 30S RIBOSOMAL PROTEIN S18



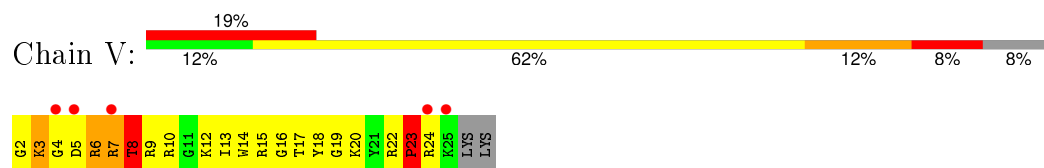
- Molecule 20: 30S RIBOSOMAL PROTEIN S19



- Molecule 21: 30S RIBOSOMAL PROTEIN S20



- Molecule 22: 30S RIBOSOMAL PROTEIN THX



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	401.84Å 401.84Å 173.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	141.42 – 3.80 148.30 – 3.78	Depositor EDS
% Data completeness (in resolution range)	92.6 (141.42-3.80) 92.4 (148.30-3.78)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.94 (at 3.78Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.241 , 0.312 0.242 , 0.309	Depositor DCC
$R_{free}$ test set	6382 reflections (5.20%)	DCC
Wilson B-factor (Å <sup>2</sup> )	122.1	Xtriage
Anisotropy	0.327	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 194.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 130759 reflections	Xtriage
$F_o, F_c$ correlation	0.84	EDS
Total number of atoms	51757	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	101.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.15% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section:  
ZN

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	0.60	1/36387 (0.0%)	0.76	22/56789 (0.0%)
2	Z	0.62	0/84	0.87	0/128
3	B	0.42	0/1935	0.73	0/2609
4	C	0.37	0/1636	0.70	0/2205
5	D	0.44	0/1733	0.73	0/2318
6	E	0.52	0/1162	0.83	0/1564
7	F	0.37	0/856	0.69	0/1154
8	G	0.35	0/1276	0.66	0/1709
9	H	0.57	0/1136	0.87	0/1527
10	I	0.36	0/1029	0.66	0/1378
11	J	0.36	0/805	0.68	0/1082
12	K	0.42	0/900	0.71	0/1213
13	L	0.41	0/986	0.76	0/1320
14	M	0.36	0/947	0.68	0/1270
15	N	0.38	0/501	0.74	0/664
16	O	0.44	0/745	0.67	0/992
17	P	0.47	0/716	0.71	0/963
18	Q	0.56	0/870	0.83	1/1159 (0.1%)
19	R	0.41	0/603	0.71	0/799
20	S	0.35	0/661	0.67	0/890
21	T	0.37	0/765	0.73	0/1007
22	V	0.40	0/212	0.71	0/277
All	All	0.54	1/55945 (0.0%)	0.75	23/83017 (0.0%)

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	2	42

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	#Chirality outliers	#Planarity outliers
17	P	0	1
All	All	2	43

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	858	G	C5-C6	-6.02	1.36	1.42

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	51	A	C2'-C3'-O3'	9.80	131.06	109.50
1	A	266	G	C2'-C3'-O3'	9.28	129.91	109.50
1	A	1498	U	C2'-C3'-O3'	9.05	129.41	109.50
1	A	1085	U	C2'-C3'-O3'	7.77	126.60	109.50
1	A	575	G	C2'-C3'-O3'	7.36	125.69	109.50

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	51	A	C3'
1	A	1498	U	C3'

5 of 43 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	127	G	Sidechain
1	A	148	G	Sidechain
1	A	197	A	Sidechain
1	A	28	G	Sidechain
1	A	90	U	Sidechain

## 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	32508	0	16414	2480	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	Z	77	0	42	5	0
3	B	1900	0	1951	427	0
4	C	1612	0	1677	396	0
5	D	1703	0	1765	380	0
6	E	1146	0	1207	261	0
7	F	843	0	857	159	0
8	G	1257	0	1296	257	0
9	H	1116	0	1177	235	0
10	I	1011	0	1043	246	0
11	J	792	0	835	245	0
12	K	885	0	904	142	0
13	L	970	0	1057	183	0
14	M	937	0	995	167	0
15	N	492	0	533	140	0
16	O	734	0	771	142	0
17	P	700	0	720	175	0
18	Q	857	0	930	180	0
19	R	597	0	668	143	0
20	S	647	0	673	114	0
21	T	763	0	861	174	0
22	V	208	0	221	53	0
23	D	1	0	0	0	0
23	N	1	0	0	0	0
All	All	51757	0	36597	6246	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 71.

The worst 5 of 6246 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:818:G:O2'	1:A:819:A:H5''	1.44	1.17
1:A:1064:G:H4'	1:A:1065:U:H5'	1.28	1.15
1:A:1443:G:H5''	1:A:1446:A:H5'	1.21	1.14
19:R:53:ARG:HH21	19:R:60:GLY:N	1.46	1.14
6:E:13:ILE:HG22	6:E:30:ALA:HA	1.22	1.13

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	
3	B	232/256 (91%)	134 (58%)	71 (31%)	27 (12%)	0	9
4	C	204/239 (85%)	95 (47%)	64 (31%)	45 (22%)	0	1
5	D	206/208 (99%)	109 (53%)	63 (31%)	34 (16%)	0	4
6	E	148/161 (92%)	98 (66%)	30 (20%)	20 (14%)	0	6
7	F	99/101 (98%)	56 (57%)	30 (30%)	13 (13%)	0	7
8	G	153/155 (99%)	68 (44%)	47 (31%)	38 (25%)	0	1
9	H	136/138 (99%)	91 (67%)	28 (21%)	17 (12%)	0	8
10	I	125/128 (98%)	70 (56%)	27 (22%)	28 (22%)	0	1
11	J	96/104 (92%)	51 (53%)	26 (27%)	19 (20%)	0	2
12	K	117/129 (91%)	72 (62%)	31 (26%)	14 (12%)	0	8
13	L	122/135 (90%)	81 (66%)	22 (18%)	19 (16%)	0	5
14	M	116/126 (92%)	66 (57%)	33 (28%)	17 (15%)	0	5
15	N	58/60 (97%)	25 (43%)	8 (14%)	25 (43%)	0	0
16	O	86/88 (98%)	42 (49%)	30 (35%)	14 (16%)	0	5
17	P	81/88 (92%)	44 (54%)	20 (25%)	17 (21%)	0	2
18	Q	102/104 (98%)	60 (59%)	27 (26%)	15 (15%)	0	5
19	R	71/88 (81%)	36 (51%)	24 (34%)	11 (16%)	0	5
20	S	78/92 (85%)	45 (58%)	24 (31%)	9 (12%)	0	9
21	T	97/106 (92%)	34 (35%)	42 (43%)	21 (22%)	0	2
22	V	22/26 (85%)	10 (46%)	7 (32%)	5 (23%)	0	1
All	All	2349/2532 (93%)	1287 (55%)	654 (28%)	408 (17%)	0	4

5 of 408 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	B	20	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	158	LEU
3	B	232	PRO
3	B	239	VAL
4	C	4	LYS

### 5.3.2 Protein sidechains ⓘ

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	
3	B	202/220 (92%)	156 (77%)	46 (23%)	1	8
4	C	160/188 (85%)	140 (88%)	20 (12%)	6	33
5	D	180/180 (100%)	150 (83%)	30 (17%)	3	20
6	E	115/122 (94%)	90 (78%)	25 (22%)	1	9
7	F	90/90 (100%)	81 (90%)	9 (10%)	9	43
8	G	126/126 (100%)	111 (88%)	15 (12%)	6	35
9	H	119/119 (100%)	102 (86%)	17 (14%)	4	28
10	I	98/99 (99%)	79 (81%)	19 (19%)	2	13
11	J	87/91 (96%)	70 (80%)	17 (20%)	2	13
12	K	90/99 (91%)	76 (84%)	14 (16%)	3	24
13	L	104/111 (94%)	94 (90%)	10 (10%)	10	46
14	M	94/101 (93%)	85 (90%)	9 (10%)	10	46
15	N	49/49 (100%)	35 (71%)	14 (29%)	0	4
16	O	79/79 (100%)	68 (86%)	11 (14%)	4	30
17	P	72/74 (97%)	59 (82%)	13 (18%)	2	16
18	Q	96/96 (100%)	83 (86%)	13 (14%)	5	30
19	R	64/77 (83%)	59 (92%)	5 (8%)	16	55
20	S	71/79 (90%)	63 (89%)	8 (11%)	7	38
21	T	76/82 (93%)	71 (93%)	5 (7%)	21	62
22	V	19/21 (90%)	17 (90%)	2 (10%)	8	41

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1991/2103 (95%)	1689 (85%)	302 (15%)	3 25

5 of 302 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	G	56	GLN
10	I	91	ASP
18	Q	93	GLN
8	G	94	ARG
9	H	84	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
7	F	94	GLN
8	G	122	HIS
19	R	36	ASN
8	G	37	ASN
8	G	86	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	1511/1522 (99%)	234 (15%)	51 (3%)
2	Z	3/6 (50%)	0	0
All	All	1514/1528 (99%)	234 (15%)	51 (3%)

5 of 234 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	8	A
1	A	9	G
1	A	31	G
1	A	32	A
1	A	39	G

5 of 51 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	575	G

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	960	U
1	A	1364	U
1	A	748	C
1	A	976	G

## 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](#)

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](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	1511/1522 (99%)	1.94	565 (37%) 0 1	15, 89, 187, 199	0
2	Z	4/6 (66%)	1.06	0 100 100	168, 184, 185, 199	0
3	B	234/256 (91%)	0.04	5 (2%) 67 51	23, 99, 178, 199	0
4	C	206/239 (86%)	0.93	35 (16%) 2 2	46, 134, 198, 199	0
5	D	208/208 (100%)	0.23	9 (4%) 39 25	14, 86, 165, 189	0
6	E	150/161 (93%)	0.14	4 (2%) 58 42	6, 56, 121, 160	0
7	F	101/101 (100%)	0.55	10 (9%) 9 6	44, 111, 171, 185	0
8	G	155/155 (100%)	0.17	7 (4%) 37 24	43, 141, 195, 199	0
9	H	138/138 (100%)	0.08	2 (1%) 78 63	3, 46, 115, 127	0
10	I	127/128 (99%)	0.93	29 (22%) 1 1	40, 146, 195, 199	0
11	J	98/104 (94%)	1.78	41 (41%) 0 1	63, 154, 199, 199	0
12	K	119/129 (92%)	0.41	12 (10%) 9 6	16, 93, 164, 199	0
13	L	124/135 (91%)	0.37	10 (8%) 15 9	19, 96, 159, 196	0
14	M	118/126 (93%)	0.60	17 (14%) 3 3	61, 125, 173, 199	0
15	N	60/60 (100%)	2.04	25 (41%) 0 1	42, 134, 185, 199	0
16	O	88/88 (100%)	0.19	6 (6%) 20 12	25, 70, 145, 189	0
17	P	83/88 (94%)	0.23	1 (1%) 81 67	12, 69, 141, 188	0
18	Q	104/104 (100%)	0.30	8 (7%) 16 10	2, 69, 140, 199	0
19	R	73/88 (82%)	0.34	3 (4%) 41 27	27, 87, 168, 194	0
20	S	80/92 (86%)	1.82	25 (31%) 1 1	55, 152, 198, 199	0
21	T	99/106 (93%)	0.00	2 (2%) 68 53	24, 82, 169, 199	0
22	V	24/26 (92%)	1.16	5 (20%) 1 1	39, 125, 181, 195	0
All	All	3904/4060 (96%)	1.06	821 (21%) 1 1	2, 98, 186, 199	0

The worst 5 of 821 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	C	2	GLY	10.1
1	A	1221	G	10.1
1	A	1053	G	9.1
20	S	2	PRO	8.9
5	D	42	GLN	8.5

## 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](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
23	ZN	D	306	1/1	0.88	0.39	0.16	78,78,78,78	0
23	ZN	N	307	1/1	0.99	0.11	-1.23	78,78,78,78	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.