



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 06:05 am BST

PDB ID : 2HHH  
Title : Crystal structure of kasugamycin bound to the 30S ribosomal subunit  
Authors : Schlutzen, F.  
Deposited on : 2006-06-28  
Resolution : 3.35 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
buster-report : 1.1.7 (2018)  
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

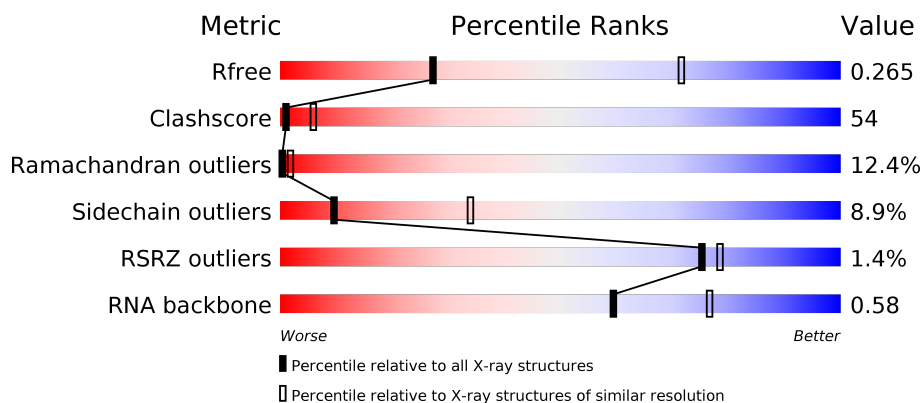
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.35 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	1558 (3.42-3.30)
Clashscore	141614	1627 (3.42-3.30)
Ramachandran outliers	138981	1599 (3.42-3.30)
Sidechain outliers	138945	1598 (3.42-3.30)
RSRZ outliers	127900	1507 (3.42-3.30)
RNA backbone	3102	1023 (3.80-2.92)

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 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  $\leq 5\%$ . 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>15%</div> <div>60%</div> <div>19%</div> <div>5%</div> </div>
2	B	256	<div> <div>5%</div> <div>16%</div> <div>55%</div> <div>17%</div> <div>9%</div> </div>
3	C	239	<div> <div>14%</div> <div>54%</div> <div>16%</div> <div>14%</div> </div>
4	D	209	<div> <div>19%</div> <div>64%</div> <div>15%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	E	162	
6	F	101	
7	G	156	
8	H	138	
9	I	128	
10	J	105	
11	K	129	
12	L	135	
13	M	126	
14	N	61	
15	O	89	
16	P	88	
17	Q	105	
18	R	88	
19	S	93	
20	T	106	
21	U	27	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
22	KSG	A	1524	-	-	-	X

## 2 Entry composition

There are 22 unique types of molecules in this entry. The entry contains 51632 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	1505	Total	C	N	O	P	44	0	0
			32349	14399	5994	10452	1504			

- Molecule 2 is a protein called 30S ribosomal protein S2.

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

- Molecule 3 is a protein called 30S ribosomal protein S3.

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

- Molecule 4 is a protein called 30S ribosomal protein S4.

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

- Molecule 5 is a protein called 30S ribosomal protein S5.

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

- Molecule 6 is a protein called 30S ribosomal protein S6.

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

- Molecule 7 is a protein called 30S ribosomal protein S7.

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

- Molecule 8 is a protein called 30S ribosomal protein S8.

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

- Molecule 9 is a protein called 30S ribosomal protein S9.

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

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	58	ARG	HIS	CONFLICT	UNP P80374

- Molecule 10 is a protein called 30S ribosomal protein S10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	98	Total	C	N	O	S	0	0	0
			794	499	156	138	1			

- Molecule 11 is a protein called 30S ribosomal protein S11.

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

- Molecule 12 is a protein called 30S ribosomal protein S12.

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

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	2	VAL	-	INSERTION	UNP P17293
L	3	ALA	-	INSERTION	UNP P17293
L	4	LEU	-	INSERTION	UNP P17293

- Molecule 13 is a protein called 30S ribosomal protein S13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	M	125	Total	C	N	O	S	0	0	0
			997	617	207	171	2			

- Molecule 14 is a protein called 30S ribosomal protein S14.

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

- Molecule 15 is a protein called 30S ribosomal protein S15.

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

- Molecule 16 is a protein called 30S ribosomal protein S16.

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

- Molecule 17 is a protein called 30S ribosomal protein S17.

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

- Molecule 18 is a protein called 30S ribosomal protein S18.

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

- Molecule 19 is a protein called 30S ribosomal protein S19.

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

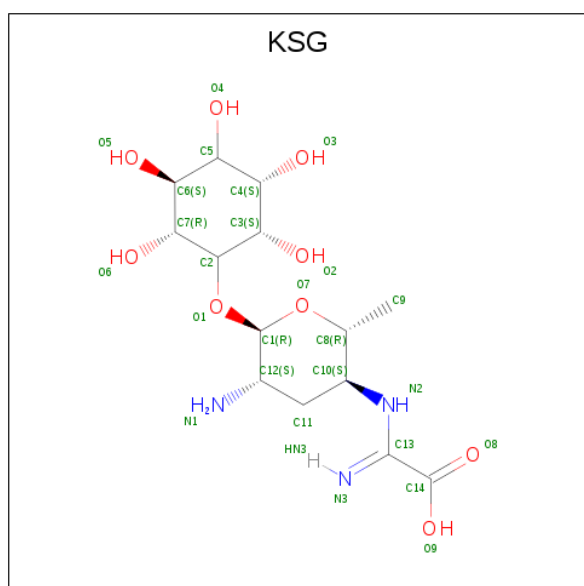
- Molecule 20 is a protein called 30S ribosomal protein S20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
20	T	99	Total	C	N	O	S	0	0	0
			762	469	162	129	2			

- Molecule 21 is a protein called 30S ribosomal protein Thx.

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

- Molecule 22 is (1S,2R,3S,4R,5S,6S)-2,3,4,5,6-PENTAHYDROXYCYCLOHEXYL 2-AMINO-4-{{CARBOXY(IMINO)METHYL}AMINO}-2,3,4,6-TETRADEOXY-ALPHA-D-ARABINO-HEXOPYRANOSIDE (three-letter code: KSG) (formula:  $C_{14}H_{25}N_3O_9$ ).

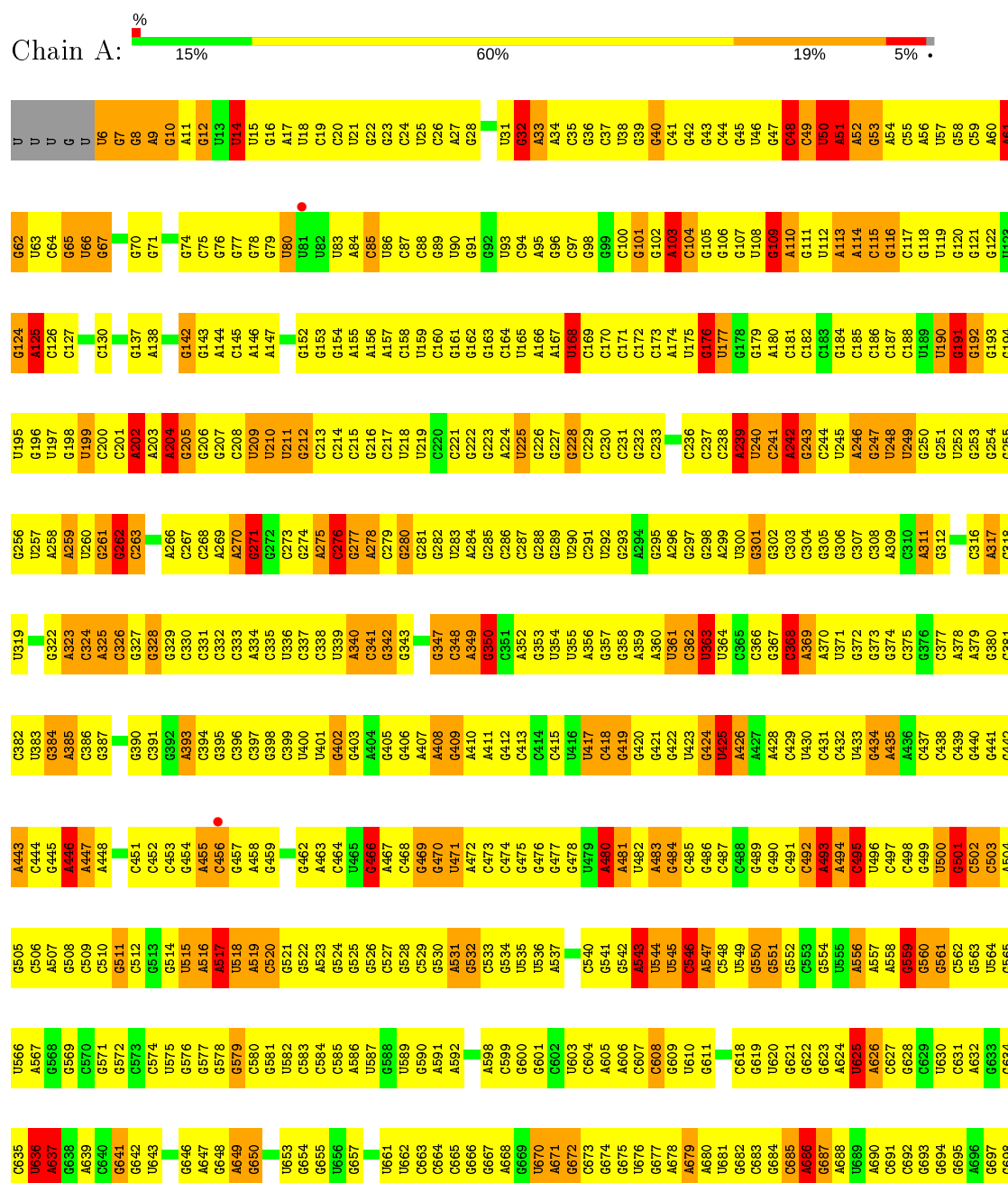


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
22	A	1	Total	C	N	O	0	0
			26	14	3	9		
22	A	1	Total	C	N	O	0	0
			26	14	3	9		

### 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 the various outlier classes 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.

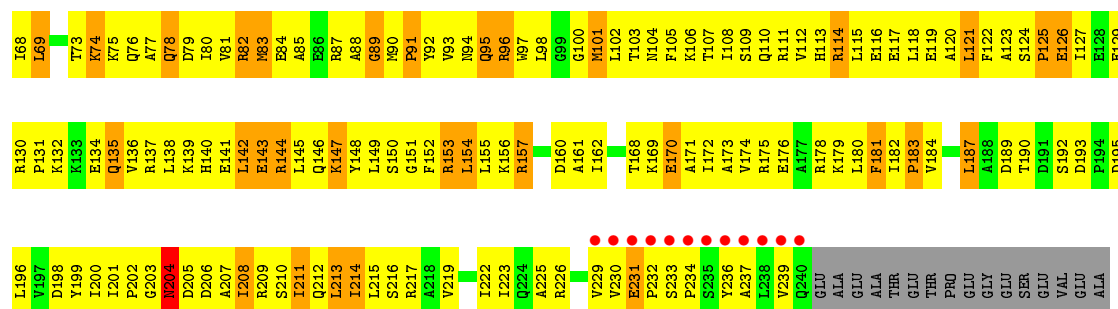
#### • Molecule 1: 16S ribosomal RNA



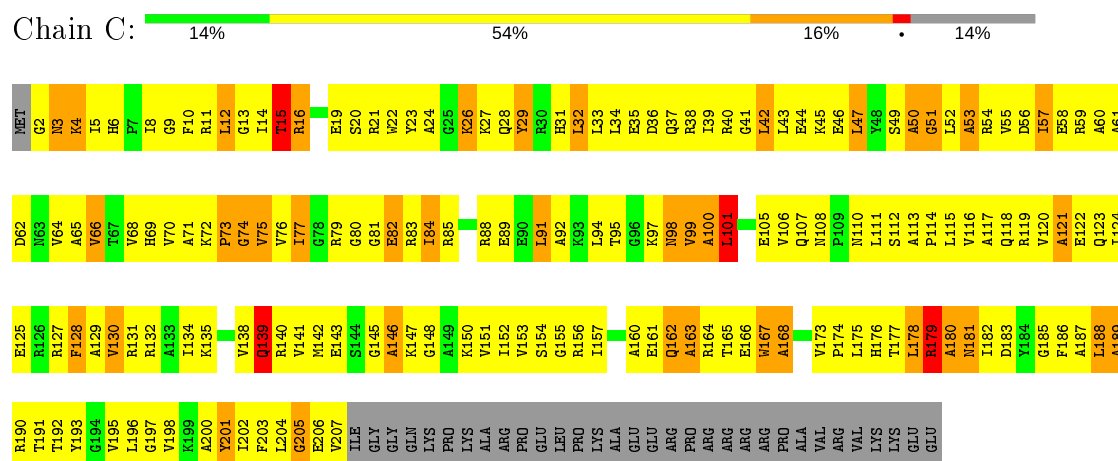


- Molecule 2: 30S ribosomal protein S2

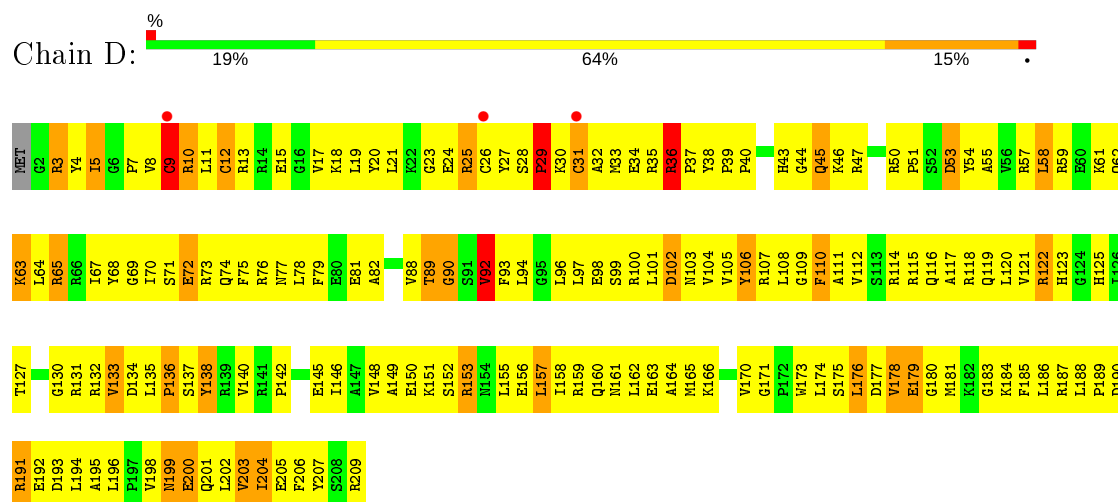
MET	PRO	VAL	GLU	ILE	THR	V7	K8	E9		F12	A13	G14	V15	H16	F17	G18	H19	E20	R21	R22	R23	M24	N25	P26	K27	F28	A29	R30	X31	I32	V33	A34	E35	R36		H40	T41	I42	D43	L44	Q45	K46	T47	M48	E49	F50	L51	E52	R53	T54	F55	R56	F57	E58	D60		G66	T67
-----	-----	-----	-----	-----	-----	----	----	----	--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--	-----	-----



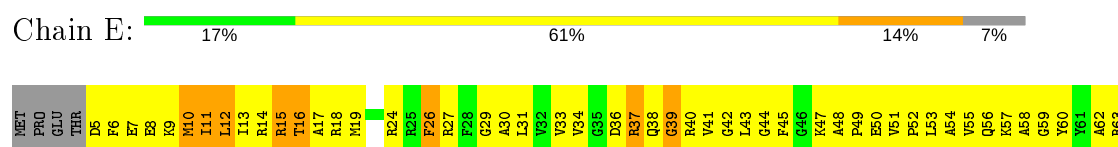
• Molecule 3: 30S ribosomal protein S3

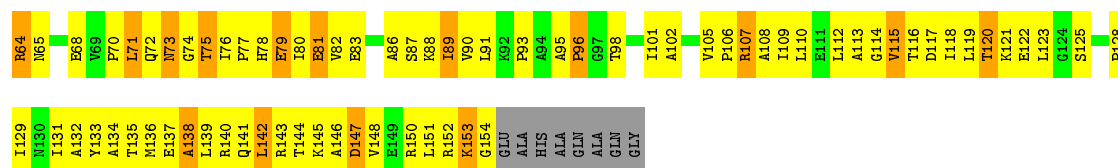


• Molecule 4: 30S ribosomal protein S4

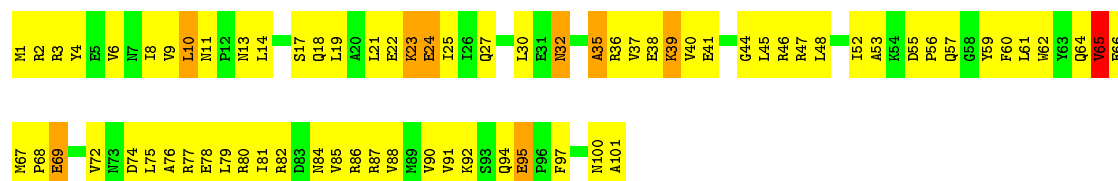


• Molecule 5: 30S ribosomal protein S5

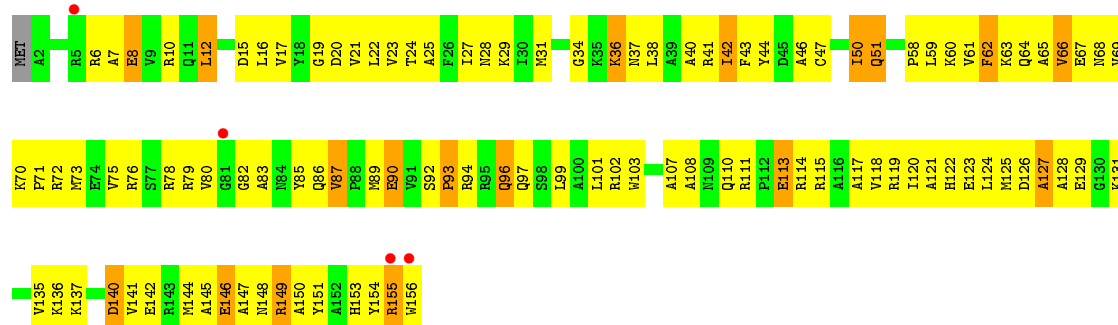




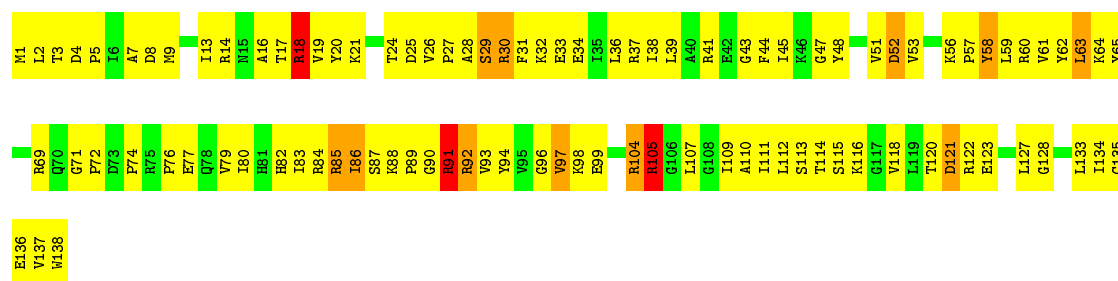
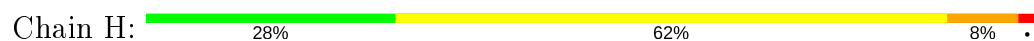
• Molecule 6: 30S ribosomal protein S6



• Molecule 7: 30S ribosomal protein S7

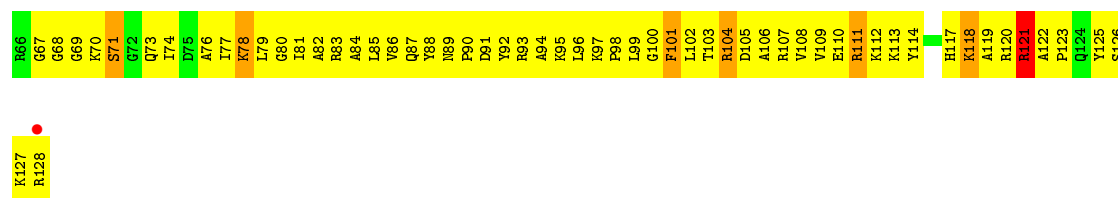


• Molecule 8: 30S ribosomal protein S8

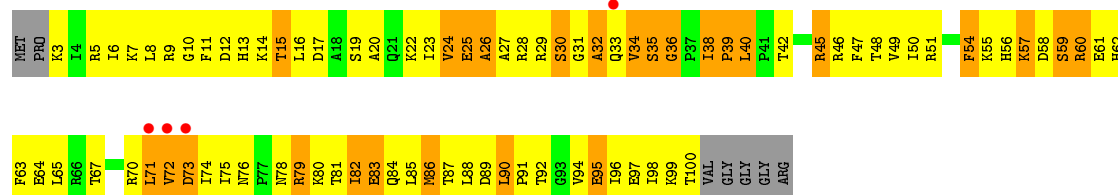
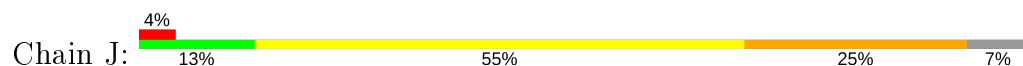


• Molecule 9: 30S ribosomal protein S9

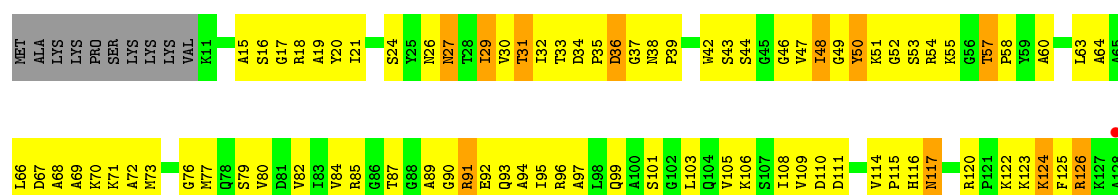




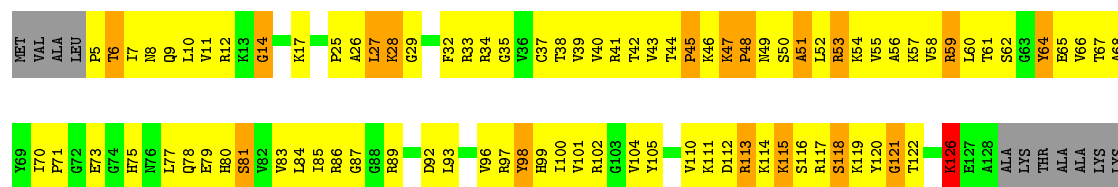
- Molecule 10: 30S ribosomal protein S10



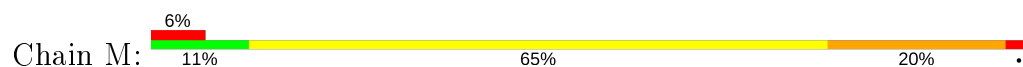
- Molecule 11: 30S ribosomal protein S11

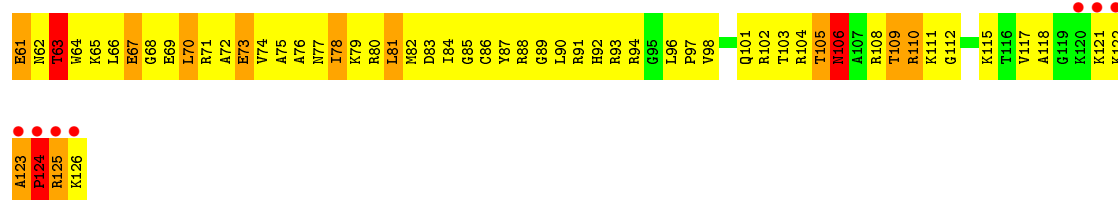


- Molecule 12: 30S ribosomal protein S12

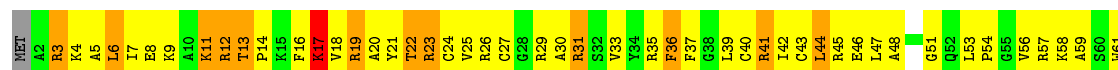
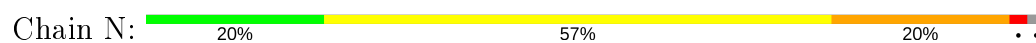


- Molecule 13: 30S ribosomal protein S13

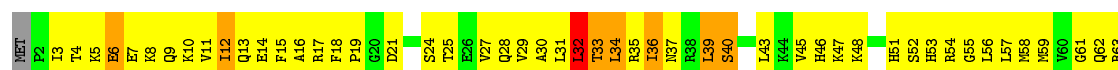




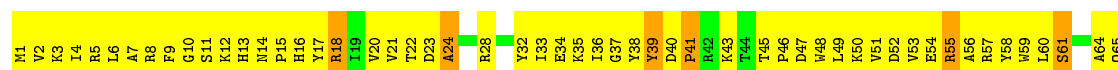
• Molecule 14: 30S ribosomal protein S14



• Molecule 15: 30S ribosomal protein S15



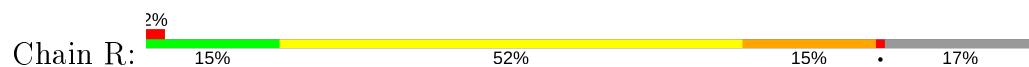
• Molecule 16: 30S ribosomal protein S16

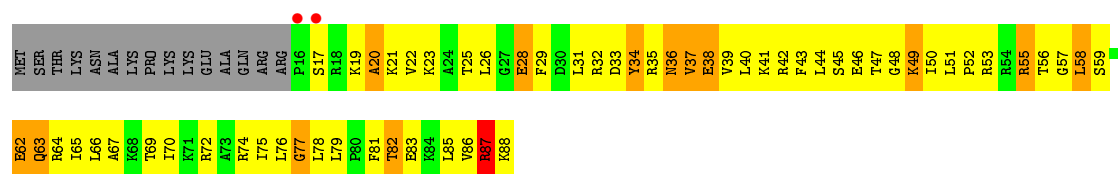


• Molecule 17: 30S ribosomal protein S17

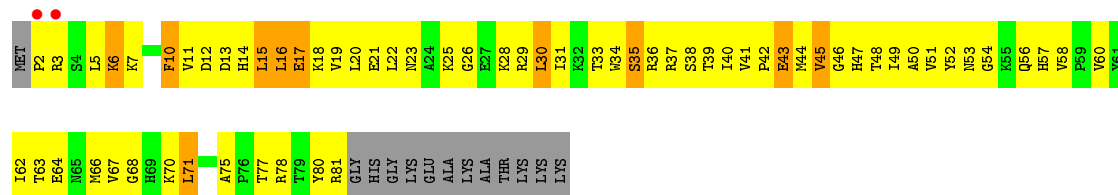
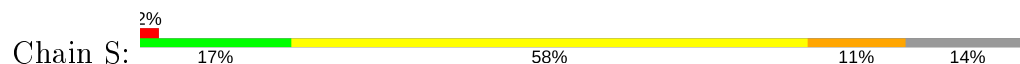


• Molecule 18: 30S ribosomal protein S18

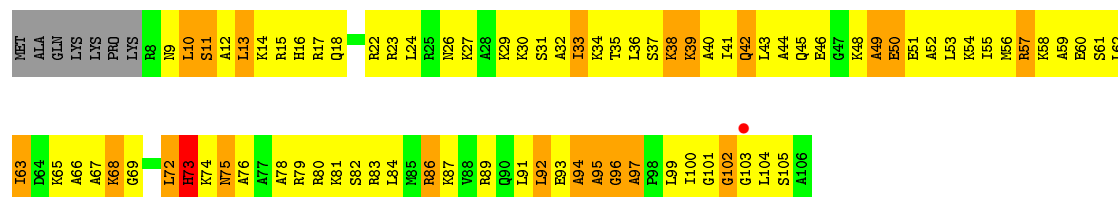
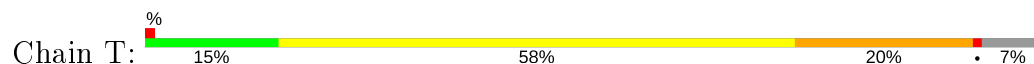




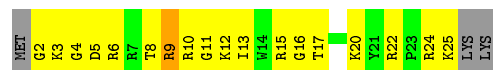
• Molecule 19: 30S ribosomal protein S19



• Molecule 20: 30S ribosomal protein S20



• Molecule 21: 30S ribosomal protein Thx



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	410.44Å 410.44Å 172.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.74 – 3.35 29.74 – 3.30	Depositor EDS
% Data completeness (in resolution range)	98.4 (29.74-3.35) 97.0 (29.74-3.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.55 (at 3.31Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.265 , 0.289 0.242 , 0.265	Depositor DCC
$R_{free}$ test set	10101 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.5	Xtriage
Anisotropy	0.534	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.16 , 19.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.36$ , $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	51632	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.65% 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, 0.333 respectively for untwinned datasets, and 0.375, 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: KSG

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	0.76	3/36212 (0.0%)	0.84	69/56520 (0.1%)
2	B	0.58	0/1935	0.77	1/2609 (0.0%)
3	C	0.56	0/1636	0.74	0/2205
4	D	0.64	1/1733 (0.1%)	0.79	2/2318 (0.1%)
5	E	0.76	0/1162	0.94	1/1564 (0.1%)
6	F	0.50	0/856	0.69	0/1154
7	G	0.48	0/1276	0.68	0/1709
8	H	0.69	0/1136	0.86	0/1527
9	I	0.51	0/1029	0.74	1/1378 (0.1%)
10	J	0.46	0/807	0.74	0/1085
11	K	0.53	0/900	0.76	0/1213
12	L	0.59	0/986	0.82	0/1320
13	M	0.48	0/1008	0.75	0/1347
14	N	0.59	0/501	0.84	0/664
15	O	0.58	0/745	0.77	0/992
16	P	0.59	0/716	0.81	0/963
17	Q	0.65	0/870	0.85	0/1159
18	R	0.50	0/603	0.74	0/799
19	S	0.43	0/661	0.70	0/890
20	T	0.52	0/764	0.76	0/1006
21	U	0.55	0/212	0.76	0/277
All	All	0.70	4/55748 (0.0%)	0.82	74/82699 (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	86
8	H	0	1

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	#Chirality outliers	#Planarity outliers
All	All	0	87

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	702	G	C5-C6	-6.16	1.36	1.42
4	D	12	CYS	CB-SG	6.00	1.92	1.82
1	A	1067	G	C5-C6	-5.41	1.36	1.42
1	A	1317	C	N1-C2	5.04	1.45	1.40

The worst 5 of 74 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	637	A	N9-C1'-C2'	9.39	126.21	114.00
1	A	242	A	N9-C1'-C2'	9.14	125.89	114.00
1	A	802	G	N9-C1'-C2'	9.02	125.72	114.00
1	A	938	U	N1-C1'-C2'	8.16	124.61	114.00
1	A	48	C	N1-C1'-C2'	7.85	124.20	114.00

There are no chirality outliers.

5 of 87 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	12	G	Sidechain
1	A	14	U	Sidechain
1	A	50	U	Sidechain
1	A	52	A	Sidechain
1	A	61	A	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	32349	0	16328	2022	0
2	B	1900	0	1951	273	0
3	C	1612	0	1677	302	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	1703	0	1767	271	0
5	E	1146	0	1207	206	0
6	F	843	0	857	101	0
7	G	1257	0	1296	158	0
8	H	1116	0	1177	132	0
9	I	1011	0	1043	181	0
10	J	794	0	840	144	0
11	K	885	0	904	116	0
12	L	970	0	1057	170	0
13	M	997	0	1072	185	0
14	N	492	0	533	93	0
15	O	734	0	771	101	0
16	P	700	0	720	102	0
17	Q	857	0	930	147	0
18	R	597	0	668	120	0
19	S	647	0	673	111	0
20	T	762	0	856	132	0
21	U	208	0	221	28	0
22	A	52	0	46	12	0
All	All	51632	0	36594	4702	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 54.

The worst 5 of 4702 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)
8:H:89:PRO:HA	8:H:92:ARG:NH1	1.49	1.27
12:L:41:ARG:HG2	12:L:42:THR:H	1.05	1.13
3:C:156:ARG:H	3:C:163:ALA:HA	1.04	1.13
19:S:28:LYS:HG2	19:S:29:ARG:H	1.07	1.12
10:J:45:ARG:HB3	10:J:45:ARG:HH11	1.15	1.11

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	
2	B	232/256 (91%)	134 (58%)	57 (25%)	41 (18%)	0	1
3	C	204/239 (85%)	114 (56%)	55 (27%)	35 (17%)	0	1
4	D	206/209 (99%)	131 (64%)	49 (24%)	26 (13%)	0	2
5	E	148/162 (91%)	99 (67%)	38 (26%)	11 (7%)	1	7
6	F	99/101 (98%)	73 (74%)	22 (22%)	4 (4%)	3	20
7	G	153/156 (98%)	103 (67%)	37 (24%)	13 (8%)	1	5
8	H	136/138 (99%)	108 (79%)	19 (14%)	9 (7%)	1	9
9	I	125/128 (98%)	80 (64%)	29 (23%)	16 (13%)	0	2
10	J	96/105 (91%)	56 (58%)	20 (21%)	20 (21%)	0	0
11	K	117/129 (91%)	78 (67%)	29 (25%)	10 (8%)	1	5
12	L	122/135 (90%)	89 (73%)	20 (16%)	13 (11%)	0	3
13	M	123/126 (98%)	69 (56%)	33 (27%)	21 (17%)	0	1
14	N	58/61 (95%)	35 (60%)	14 (24%)	9 (16%)	0	1
15	O	86/89 (97%)	47 (55%)	28 (33%)	11 (13%)	0	2
16	P	81/88 (92%)	57 (70%)	20 (25%)	4 (5%)	2	15
17	Q	102/105 (97%)	73 (72%)	15 (15%)	14 (14%)	0	1
18	R	71/88 (81%)	46 (65%)	16 (22%)	9 (13%)	0	2
19	S	78/93 (84%)	50 (64%)	18 (23%)	10 (13%)	0	2
20	T	97/106 (92%)	44 (45%)	38 (39%)	15 (16%)	0	1
21	U	22/27 (82%)	18 (82%)	2 (9%)	2 (9%)	1	4
All	All	2356/2541 (93%)	1504 (64%)	559 (24%)	293 (12%)	0	2

5 of 293 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	8	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	17	PHE
2	B	23	ARG
2	B	24	TRP
2	B	29	ALA

### 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	
2	B	202/220 (92%)	183 (91%)	19 (9%)	8	31
3	C	160/188 (85%)	142 (89%)	18 (11%)	6	23
4	D	180/181 (99%)	164 (91%)	16 (9%)	9	34
5	E	115/123 (94%)	102 (89%)	13 (11%)	6	23
6	F	90/90 (100%)	84 (93%)	6 (7%)	16	47
7	G	126/127 (99%)	120 (95%)	6 (5%)	25	57
8	H	119/119 (100%)	109 (92%)	10 (8%)	11	37
9	I	98/99 (99%)	88 (90%)	10 (10%)	7	28
10	J	88/92 (96%)	81 (92%)	7 (8%)	12	39
11	K	90/99 (91%)	85 (94%)	5 (6%)	21	53
12	L	104/111 (94%)	96 (92%)	8 (8%)	13	41
13	M	100/101 (99%)	84 (84%)	16 (16%)	2	10
14	N	49/50 (98%)	44 (90%)	5 (10%)	7	28
15	O	79/80 (99%)	72 (91%)	7 (9%)	9	34
16	P	72/74 (97%)	68 (94%)	4 (6%)	21	53
17	Q	96/97 (99%)	89 (93%)	7 (7%)	14	43
18	R	64/77 (83%)	58 (91%)	6 (9%)	8	31
19	S	71/80 (89%)	68 (96%)	3 (4%)	30	60
20	T	75/82 (92%)	64 (85%)	11 (15%)	3	13
21	U	19/22 (86%)	19 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1997/2112 (95%)	1820 (91%)	177 (9%)	9 34

5 of 177 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	H	52	ASP
10	J	38	ILE
19	S	10	PHE
8	H	63	LEU
9	I	3	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 66 such sidechains are listed below:

Mol	Chain	Res	Type
6	F	94	GLN
7	G	106	GLN
19	S	14	HIS
6	F	100	ASN
7	G	64	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	1505/1522 (98%)	286 (19%)	164 (10%)

5 of 286 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	7	G
1	A	8	G
1	A	9	A
1	A	10	G
1	A	14	U

5 of 164 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	637	A
1	A	848	U
1	A	1380	C

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	685	C
1	A	776	A

## 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](#)

2 ligands are 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
22	KSG	A	1524	-	24,27,27	1.47	1 (4%)	29,40,40	1.24	2 (6%)
22	KSG	A	1523	-	24,27,27	1.46	1 (4%)	29,40,40	1.23	2 (6%)

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
22	KSG	A	1524	-	-	1/8/52/52	0/2/2/2
22	KSG	A	1523	-	-	3/8/52/52	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
22	A	1524	KSG	C14-C13	-6.81	1.49	1.53
22	A	1523	KSG	C14-C13	-6.78	1.49	1.53

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
22	A	1524	KSG	C1-O1-C2	-4.51	106.80	117.96
22	A	1523	KSG	C1-O1-C2	-4.50	106.83	117.96
22	A	1524	KSG	C1-O7-C8	-3.56	107.55	113.67
22	A	1523	KSG	C1-O7-C8	-3.54	107.58	113.67

There are no chirality outliers.

All (4) torsion outliers are listed below:

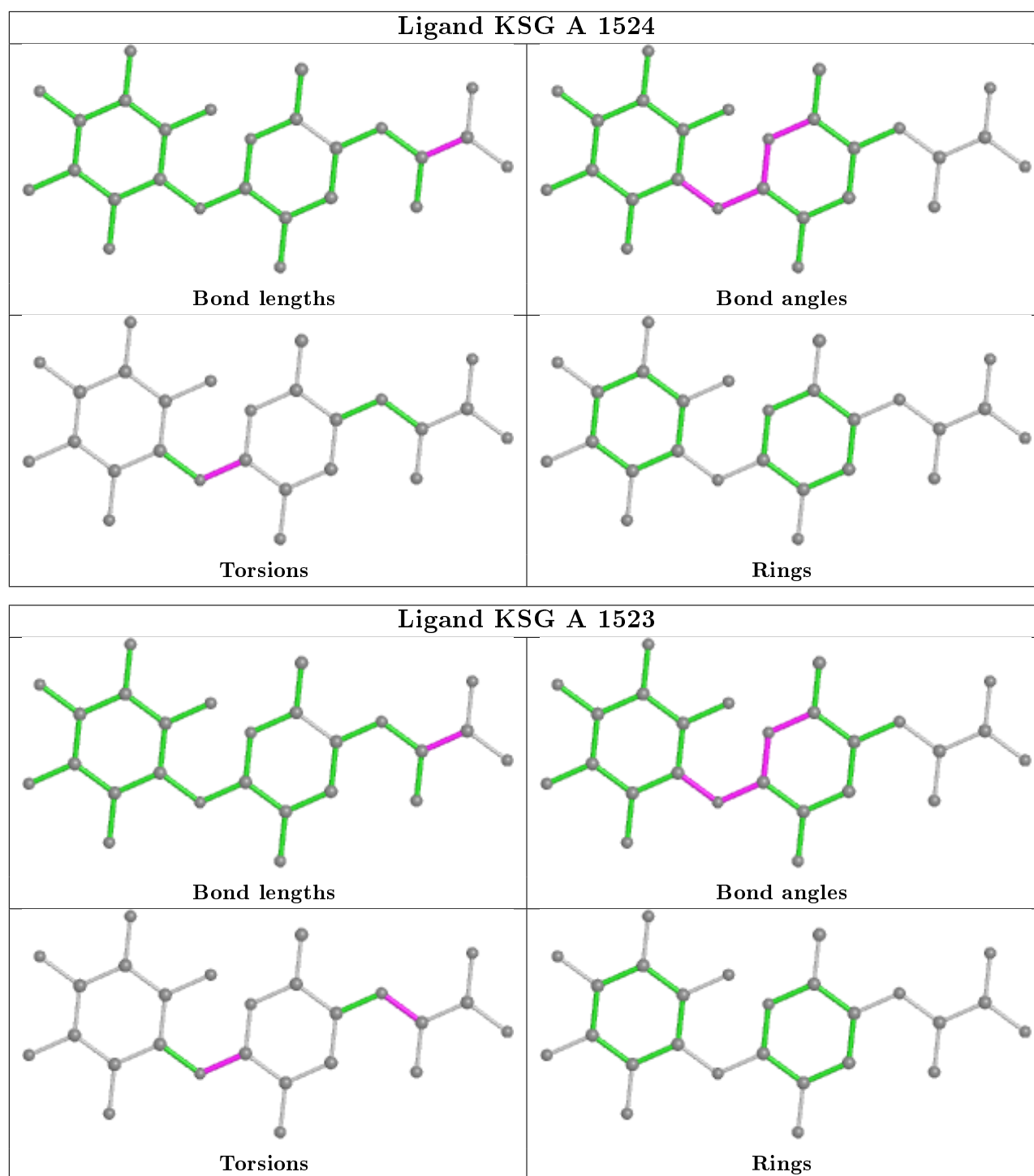
Mol	Chain	Res	Type	Atoms
22	A	1523	KSG	C14-C13-N2-C10
22	A	1523	KSG	N3-C13-N2-C10
22	A	1524	KSG	O7-C1-O1-C2
22	A	1523	KSG	O7-C1-O1-C2

There are no ring outliers.

2 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
22	A	1524	KSG	7	0
22	A	1523	KSG	5	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 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	1503/1522 (98%)	-0.55	12 (0%) 86 89	1, 29, 110, 161	0
2	B	234/256 (91%)	-0.25	12 (5%) 28 30	4, 41, 109, 134	0
3	C	206/239 (86%)	-0.60	0 100 100	7, 46, 83, 93	0
4	D	208/209 (99%)	-0.57	3 (1%) 75 78	1, 27, 57, 70	0
5	E	150/162 (92%)	-0.78	0 100 100	1, 10, 35, 60	0
6	F	101/101 (100%)	-0.57	0 100 100	25, 52, 71, 80	0
7	G	155/156 (99%)	-0.44	4 (2%) 56 58	27, 54, 85, 102	0
8	H	138/138 (100%)	-0.84	0 100 100	1, 6, 34, 41	0
9	I	127/128 (99%)	-0.36	1 (0%) 86 89	9, 59, 87, 96	0
10	J	98/105 (93%)	-0.09	4 (4%) 37 39	15, 78, 118, 124	0
11	K	119/129 (92%)	-0.42	2 (1%) 70 73	18, 39, 66, 110	0
12	L	124/135 (91%)	-0.59	0 100 100	1, 32, 55, 98	0
13	M	125/126 (99%)	-0.05	7 (5%) 24 26	32, 53, 94, 133	0
14	N	60/61 (98%)	-0.44	0 100 100	15, 37, 78, 93	0
15	O	88/89 (98%)	-0.78	0 100 100	2, 30, 61, 95	0
16	P	83/88 (94%)	-0.76	0 100 100	3, 22, 42, 75	0
17	Q	104/105 (99%)	-0.55	4 (3%) 40 42	1, 22, 93, 153	0
18	R	73/88 (82%)	-0.43	2 (2%) 54 57	10, 33, 95, 136	0
19	S	80/93 (86%)	-0.10	2 (2%) 57 59	47, 64, 106, 120	0
20	T	99/106 (93%)	-0.51	1 (1%) 82 86	22, 37, 80, 87	0
21	U	24/27 (88%)	-0.19	0 100 100	27, 38, 62, 78	0
All	All	3899/4063 (95%)	-0.51	54 (1%) 75 78	1, 35, 95, 161	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
13	M	123	ALA	10.6
11	K	129	SER	8.9
2	B	234	PRO	8.7
13	M	124	PRO	8.2
13	M	121	LYS	7.3

## 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](#)

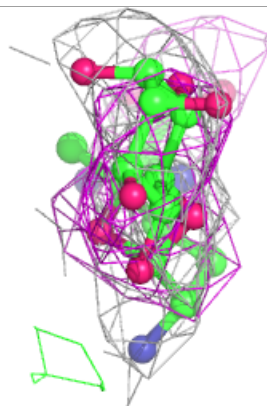
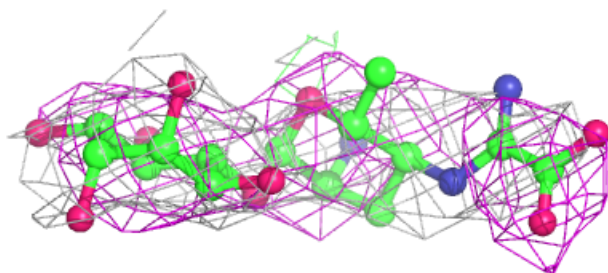
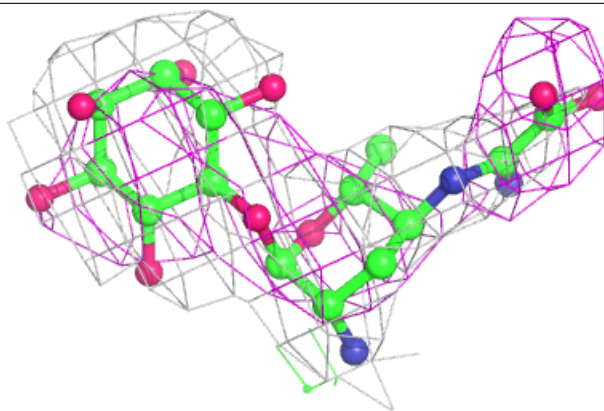
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
22	KSG	A	1524	26/26	0.80	0.62	28,32,34,36	0
22	KSG	A	1523	26/26	0.88	0.40	27,31,32,33	0

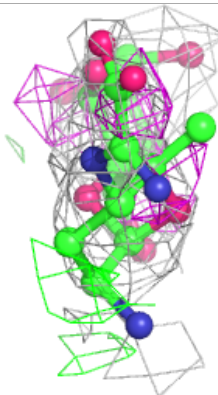
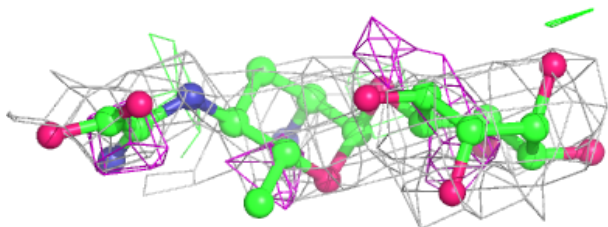
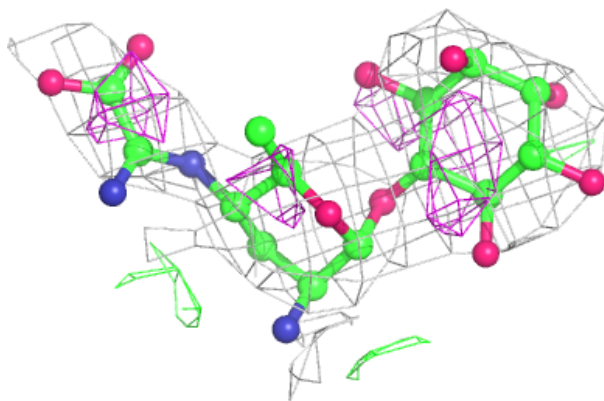
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around KSG A 1524:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)

**Electron density around KSG A 1523:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



## 6.5 Other polymers [i](#)

There are no such residues in this entry.