WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 7, 2023 – 07:59 pm BST

PDB ID	:	8BUS
Title	:	Structure of DDB1 bound to DS59-engaged CDK12-cyclin K
Authors	:	Kozicka, Z.; Kempf, G.; Focht, V.; Thoma, N.H.
Deposited on	:	2022-11-30
Resolution	:	3.26 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

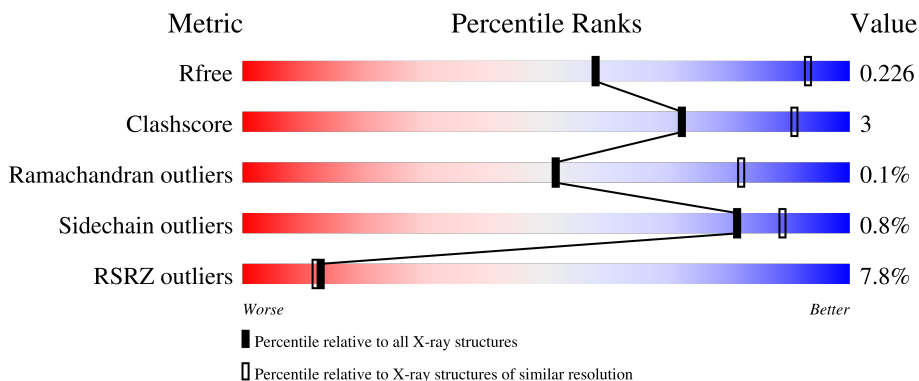|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 3.26 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	Percentile Ranks	Value
Rfree		0.226
Clashscore		3
Ramachandran outliers		0.1%
Sidechain outliers		0.8%
RSRZ outliers		7.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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 >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% 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	840	5% / 88% / 11% / •
1	D	840	9% / 88% / 10% / •
1	G	840	8% / 84% / 14% / •
2	B	344	12% / 84% / 9% / • / 6%
2	E	344	14% / 87% / 8% / •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	344	7% ... 89% ... 7% ·
3	C	271	% ... 88% ... · 8%
3	F	271	4% ... 87% ... · 8%
3	I	271	4% ... 86% ... 6% 8%

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
2	TPO	H	893	-	-	-	X

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 67538 atoms, of which 33668 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 damage-binding protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	D	826	Total 12938	C 4105	H 6451	N 1094	O 1252	S 36	6451	0	0
1	A	827	Total 12957	C 4111	H 6462	N 1095	O 1253	S 36	6462	0	0
1	G	826	Total 12939	C 4106	H 6453	N 1093	O 1251	S 36	6453	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	expression tag	UNP Q16531
D	-2	GLY	-	expression tag	UNP Q16531
D	-1	GLY	-	expression tag	UNP Q16531
D	0	ARG	-	expression tag	UNP Q16531
D	700	GLY	-	linker	UNP Q16531
D	701	ASN	-	linker	UNP Q16531
D	702	GLY	-	linker	UNP Q16531
D	703	ASN	-	linker	UNP Q16531
D	704	SER	-	linker	UNP Q16531
D	705	GLY	-	linker	UNP Q16531
A	-3	GLY	-	expression tag	UNP Q16531
A	-2	GLY	-	expression tag	UNP Q16531
A	-1	GLY	-	expression tag	UNP Q16531
A	0	ARG	-	expression tag	UNP Q16531
A	700	GLY	-	linker	UNP Q16531
A	701	ASN	-	linker	UNP Q16531
A	702	GLY	-	linker	UNP Q16531
A	703	ASN	-	linker	UNP Q16531
A	704	SER	-	linker	UNP Q16531
A	705	GLY	-	linker	UNP Q16531
G	-3	GLY	-	expression tag	UNP Q16531
G	-2	GLY	-	expression tag	UNP Q16531
G	-1	GLY	-	expression tag	UNP Q16531

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	0	ARG	-	expression tag	UNP Q16531
G	700	GLY	-	linker	UNP Q16531
G	701	ASN	-	linker	UNP Q16531
G	702	GLY	-	linker	UNP Q16531
G	703	ASN	-	linker	UNP Q16531
G	704	SER	-	linker	UNP Q16531
G	705	GLY	-	linker	UNP Q16531

- Molecule 2 is a protein called Cyclin-dependent kinase 12.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
2	E	329	Total 5385	C 1723	H 2698	N 454	O 492	P 1	S 17	2698	0	0
2	B	325	Total 5310	C 1695	H 2664	N 447	O 486	P 1	S 17	2664	0	0
2	H	330	Total 5396	C 1726	H 2703	N 455	O 494	P 1	S 17	2703	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	709	GLY	-	expression tag	UNP Q9NYV4
E	710	GLY	-	expression tag	UNP Q9NYV4
E	711	GLY	-	expression tag	UNP Q9NYV4
E	965	ARG	LYS	engineered mutation	UNP Q9NYV4
E	1052	GLN	-	expression tag	UNP Q9NYV4
B	709	GLY	-	expression tag	UNP Q9NYV4
B	710	GLY	-	expression tag	UNP Q9NYV4
B	711	GLY	-	expression tag	UNP Q9NYV4
B	965	ARG	LYS	engineered mutation	UNP Q9NYV4
B	1052	GLN	-	expression tag	UNP Q9NYV4
H	709	GLY	-	expression tag	UNP Q9NYV4
H	710	GLY	-	expression tag	UNP Q9NYV4
H	711	GLY	-	expression tag	UNP Q9NYV4
H	965	ARG	LYS	engineered mutation	UNP Q9NYV4
H	1052	GLN	-	expression tag	UNP Q9NYV4

- Molecule 3 is a protein called Cyclin-K.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	F	248	Total 4112	C 1341	H 2049	N 346	O 363	S 13	2049	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	248	Total	C	H	N	O	S	2049	0	0
			4112	1341	2049	346	363	13			
3	I	248	Total	C	H	N	O	S	2049	0	0
			4112	1341	2049	346	363	13			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-3	GLY	-	expression tag	UNP O75909
F	-2	GLY	-	expression tag	UNP O75909
F	-1	GLY	-	expression tag	UNP O75909
F	0	ARG	-	expression tag	UNP O75909
C	-3	GLY	-	expression tag	UNP O75909
C	-2	GLY	-	expression tag	UNP O75909
C	-1	GLY	-	expression tag	UNP O75909
C	0	ARG	-	expression tag	UNP O75909
I	-3	GLY	-	expression tag	UNP O75909
I	-2	GLY	-	expression tag	UNP O75909
I	-1	GLY	-	expression tag	UNP O75909
I	0	ARG	-	expression tag	UNP O75909

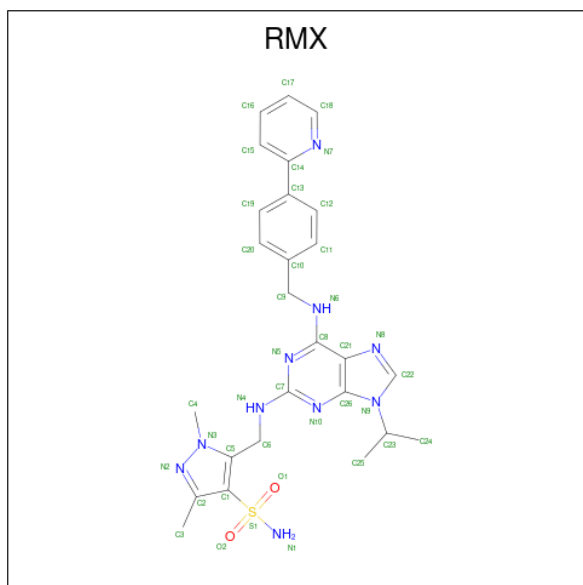- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
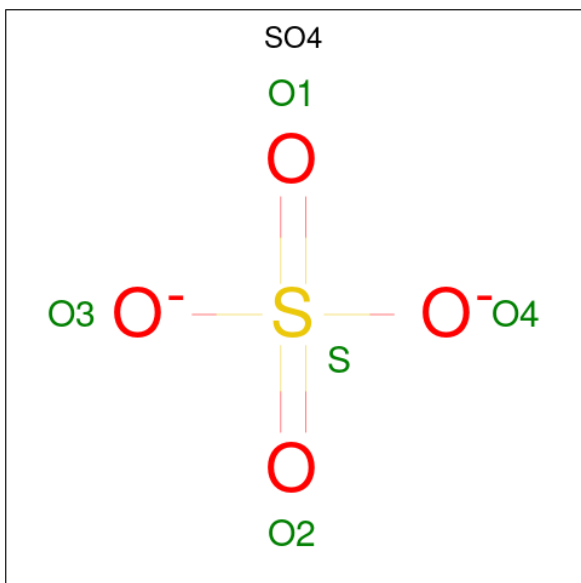
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total O S			0	0
			5	4	1		
4	D	1	Total O S			0	0
			5	4	1		
4	F	1	Total O S			0	0
			5	4	1		
4	F	1	Total O S			0	0
			5	4	1		
4	A	1	Total O S			0	0
			5	4	1		
4	C	1	Total O S			0	0
			5	4	1		
4	C	1	Total O S			0	0
			5	4	1		
4	G	1	Total O S			0	0
			5	4	1		
4	G	1	Total O S			0	0
			5	4	1		
4	H	1	Total O S			0	0
			5	4	1		
4	I	1	Total O S			0	0
			5	4	1		
4	I	1	Total O S			0	0
			5	4	1		
4	I	1	Total O S			0	0
			5	4	1		

- Molecule 5 is 1,3-dimethyl-5-[[[9-propan-2-yl-6-[(4-pyridin-2-ylphenyl)methylamino]purin-2-yl]amino]methyl]pyrazole-4-sulfonamide (three-letter code: RMX) (formula: $C_{26}H_{30}N_{10}O_2S$) (labeled as "Ligand of Interest" by depositor).
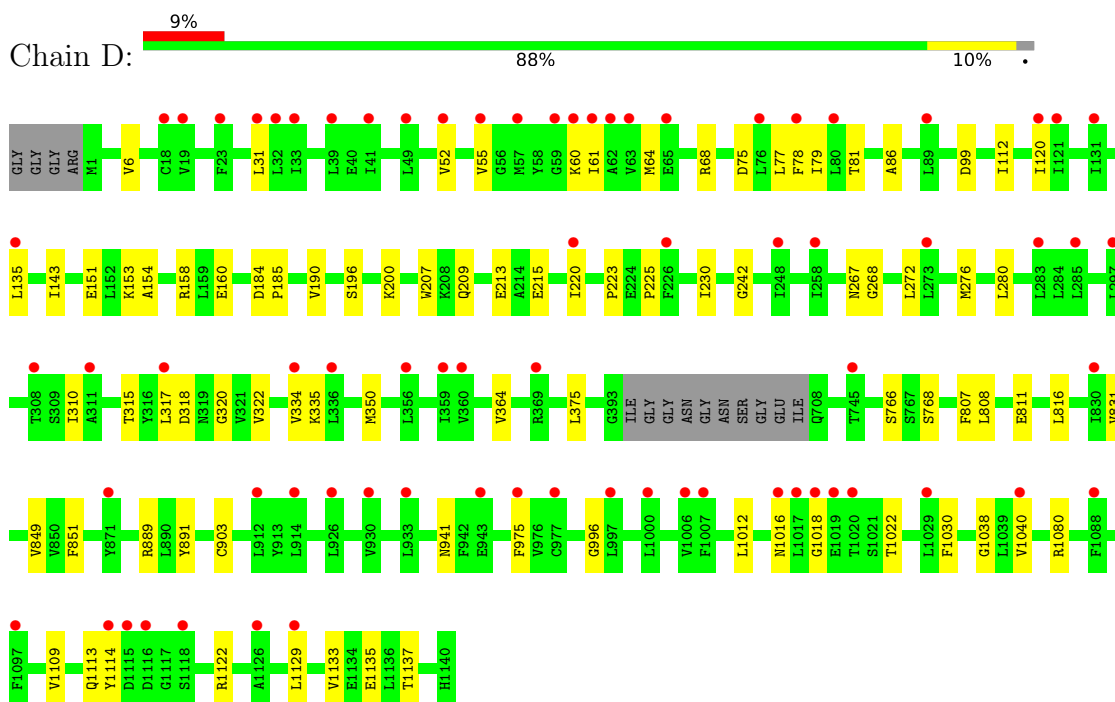
RMX

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	E	1	Total	C	H	N	O	S	30	0
			69	26	30	10	2	1		
5	B	1	Total	C	H	N	O	S	30	0
			69	26	30	10	2	1		
5	H	1	Total	C	H	N	O	S	30	0
			69	26	30	10	2	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 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.

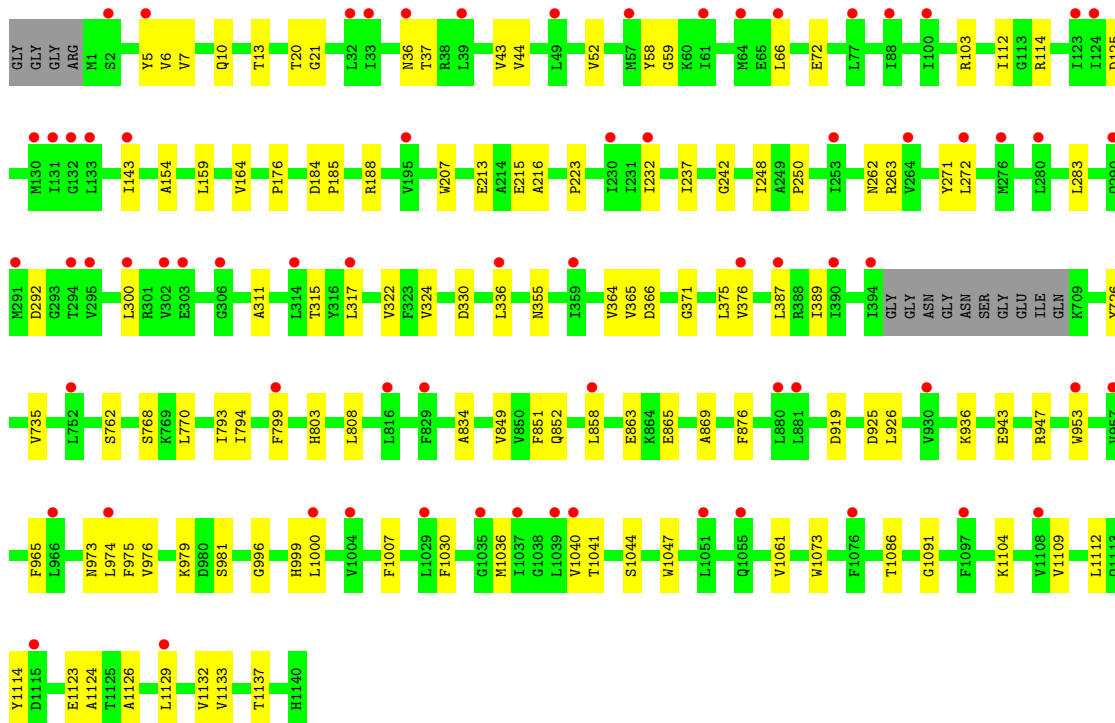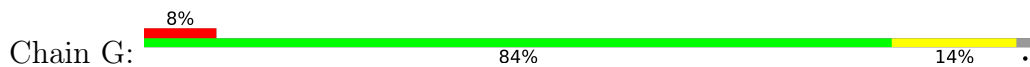• Molecule 1: DNA damage-binding protein 1
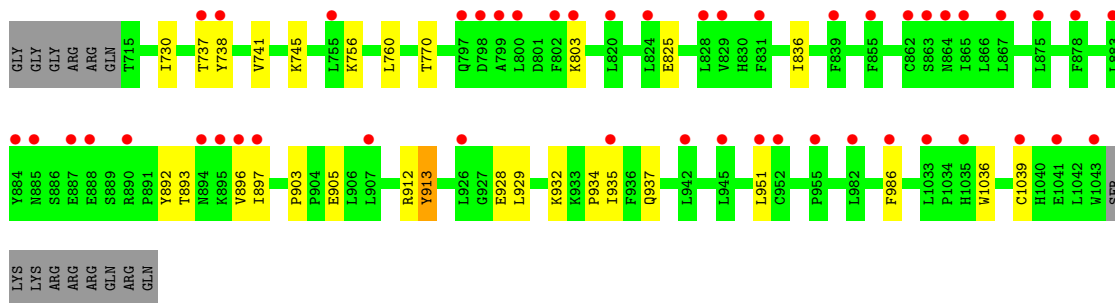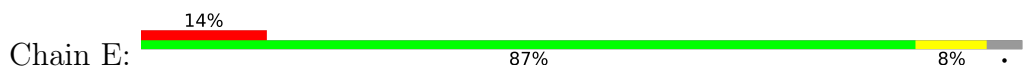
Chain D:

• Molecule 1: DNA damage-binding protein 1

Chain A:

I921 D925 L926 M927 V930 L933 F942 P951 V957 E958 I959 L966 E969 L974 T985 D986 E987 V995 G1001 N1005 L1012 E1019 T1024 W1047 V1065 R1080 G1091 F1097 M1105 V1109 L1129 I1130 K1131 V1132
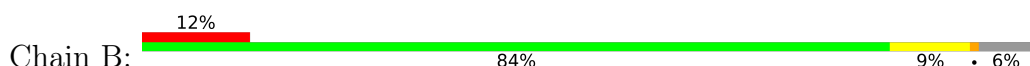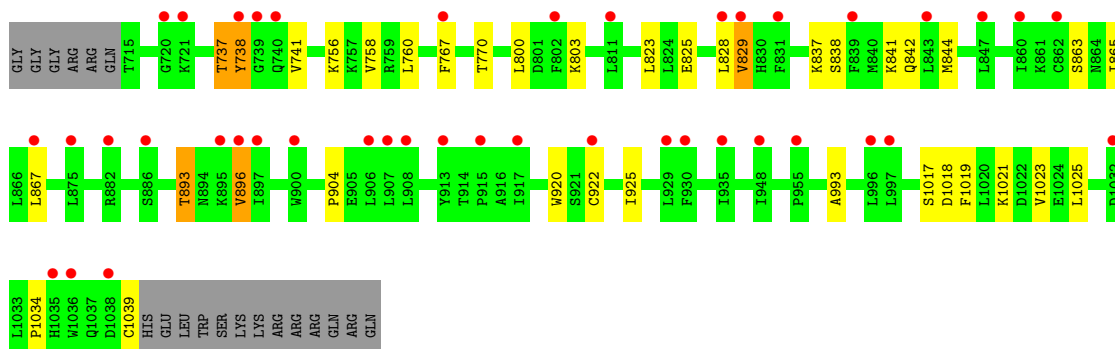
L1136 I1139 H1140

• Molecule 1: DNA damage-binding protein 1

Chain G:

8%

84%   14%

GLY GLY GLY ARG M1 S2 Y5 V6 V7 Q10 T13 T20 G21 L32 I33 N36 T37 R38 L39 V43 V44 L49 V52 M57 Y58 G59 K60 I61 M64 E65 L66 E72 L77 I88 I100 R103 I112 G113 R114 T123 I124 D125

M130 I131 G132 L133 I143 A154 L159 V164 P176 D184 P185 R188 V195 W207 E213 A214 E215 A216 P223 I230 I231 I232 I237 G242 I248 A249 P250 I253 N262 R263 V264 Y271 L272 M276 L280 L283 Q290

M291 D292 G293 T294 V295 L300 R301 V302 E303 G306 A311 L314 T315 Y316 L317 V322 F323 V324 D330 L336 N355 I359 V364 V365 D366 G371 L375 V376 L387 R388 I389 I390 I394 GLY ASN GLY ASN SER GLY GLU ILE GLN K709 Y726

V735 L752 S762 S768 K769 L770 I793 I794 F799 H803 L808 I816 F829 A834 V849 V850 F851 Q852 L858 E863 K864 E865 A869 F876 L880 L881 D919 D925 L926 V930 K936 E943 R947 W953 V957

F965 L966 N973 L974 F975 V976 K979 D980 S981 G996 H999 L1000 V1004 F1007 L1029 F1030 G1035 M1036 I1037 G1038 L1039 V1040 T1041 S1044 W1047 L1051 Q1055 V1061 W1073 F1076 T1086 G1091 F1097 K1104 V1108 V1109 L1112 Q1113
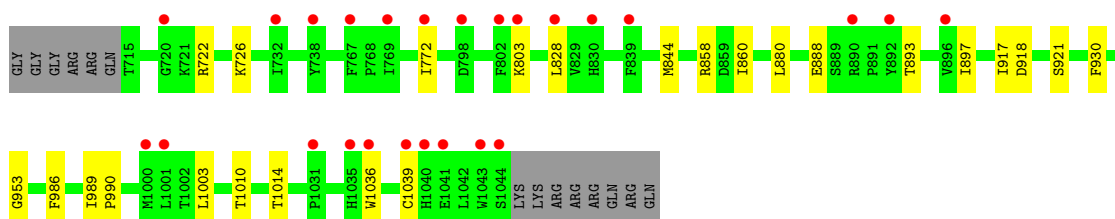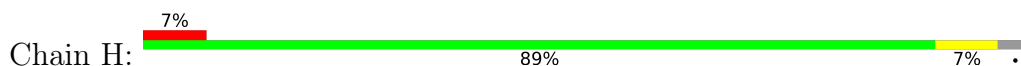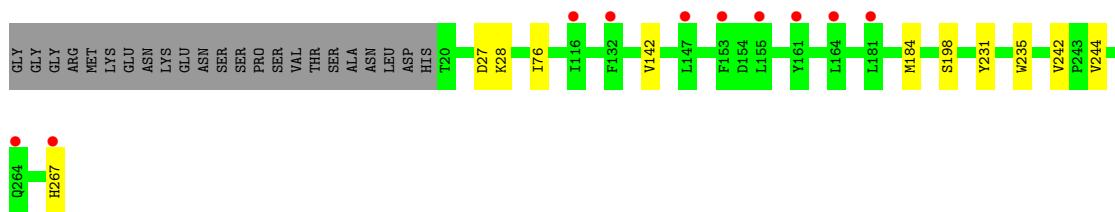
Y1114 D1115 E1123 A1124 T1125 A1126 L1129 V1132 V1133 T1137 H1140

• Molecule 2: Cyclin-dependent kinase 12

Chain E:

14%

87%   8%

GLY GLY GLY ARG ARG GLN T715 I730 T737 Y738 V741 K745 L755 K756 L760 T770 Q797 D798 A799 L800 D801 F802 K803 L820 L824 E825 L828 V829 H830 F831 I836 F839 F855 C862 S863 N864 I865 L866 L867 L875 F878 L883

Y884 N885 S886 E887 E888 S889 R890 P891 Y892 T893 N894 K895 V896 I897 P903 P904 E905 L906 L907 R912 Y913 L926 G927 E928 L929 K932 K933 P934 I935 F936 Q937 L942 L945 L951 C952 P955 L982 F986 L1033 P1034 H1035 W1036 C1039 H1040 E1041 L1042 W1043 SER

LYS LYS ARG ARG GLN ARG GLN

• Molecule 2: Cyclin-dependent kinase 12

Chain B:

12%

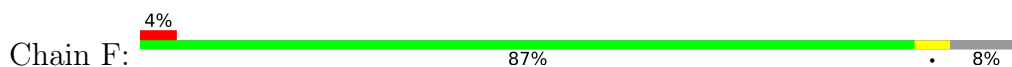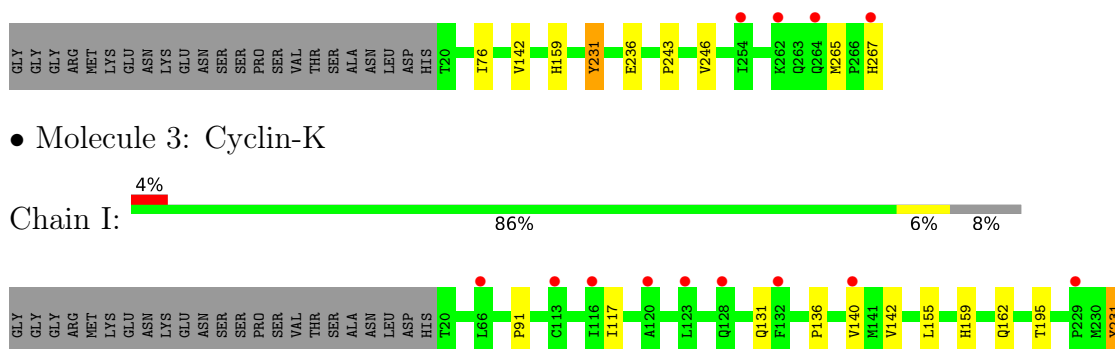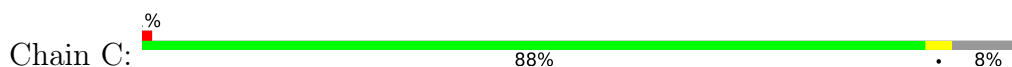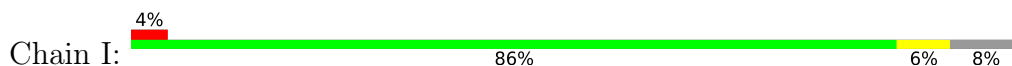84%   9%   6%

• Molecule 2: Cyclin-dependent kinase 12

Chain H:

• Molecule 3: Cyclin-K

Chain F:

• Molecule 3: Cyclin-K

Chain C:

• Molecule 3: Cyclin-K

Chain I:

R232 R233 P243 V246 L257 Q264 M265 P266 H267

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	249.52Å 249.52Å 218.55Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	60.40 – 3.26 216.09 – 3.26	Depositor EDS
% Data completeness (in resolution range)	83.2 (60.40-3.26) 83.2 (216.09-3.26)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 3.26Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.208 , 0.228 0.208 , 0.226	Depositor DCC
R_{free} test set	5017 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	119.1	Xtriage
Anisotropy	0.003	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 78.9	EDS
L-test for twinning[2]	$< |L| > = 0.45, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.034 for -h,-k,l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	67538	wwPDB-VP
Average B, all atoms (Å^2)	142.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.62% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, TPO, RMX

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	0.28	0/6612	0.54	0/8942
1	D	0.28	0/6604	0.55	0/8931
1	G	0.28	0/6603	0.54	0/8930
2	B	0.29	0/2693	0.57	0/3630
2	E	0.30	0/2737	0.58	0/3691
2	H	0.28	0/2743	0.52	0/3699
3	C	0.28	0/2120	0.48	0/2868
3	F	0.27	0/2120	0.47	0/2868
3	I	0.28	0/2120	0.50	0/2868
All	All	0.28	0/34352	0.54	0/46427

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	6495	6462	6464	49	0
1	D	6487	6451	6451	43	1
1	G	6486	6453	6454	66	1
2	B	2646	2664	2662	25	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	2687	2698	2696	14	0
2	H	2693	2703	2702	15	0
3	C	2063	2049	2048	5	0
3	F	2063	2049	2048	8	0
3	I	2063	2049	2048	13	1
4	A	5	0	0	0	0
4	C	10	0	0	0	0
4	D	15	0	0	0	0
4	F	10	0	0	0	0
4	G	10	0	0	0	0
4	H	5	0	0	0	0
4	I	15	0	0	0	0
5	B	39	30	0	0	0
5	E	39	30	0	0	0
5	H	39	30	0	0	0
All	All	33870	33668	33573	226	2

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 3.

The worst 5 of 226 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:G:953:TRP:CH2	2:H:828:LEU:HD11	2.25	0.71
2:E:741:VAL:HG22	2:E:756:LYS:HG2	1.70	0.71
1:D:225:PRO:HG2	1:D:267:ASN:HB2	1.76	0.67
1:G:322:VAL:HG21	1:G:336:LEU:HD11	1.77	0.66
1:G:7:VAL:HG22	1:G:1091:GLY:HA3	1.80	0.64

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:1122:ARG:HE	1:G:292:ASP:OD1[2_565]	1.49	0.11
2:B:1018:ASP:OD2	3:I:232:ARG:H[5_554]	1.58	0.02

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	823/840 (98%)	801 (97%)	22 (3%)	0	100	100
1	D	822/840 (98%)	797 (97%)	24 (3%)	1 (0%)	51	82
1	G	822/840 (98%)	800 (97%)	22 (3%)	0	100	100
2	B	322/344 (94%)	313 (97%)	8 (2%)	1 (0%)	41	72
2	E	326/344 (95%)	315 (97%)	10 (3%)	1 (0%)	41	72
2	H	327/344 (95%)	317 (97%)	10 (3%)	0	100	100
3	C	246/271 (91%)	245 (100%)	1 (0%)	0	100	100
3	F	246/271 (91%)	245 (100%)	1 (0%)	0	100	100
3	I	246/271 (91%)	244 (99%)	2 (1%)	0	100	100
All	All	4180/4365 (96%)	4077 (98%)	100 (2%)	3 (0%)	51	82

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	737	THR
2	E	737	THR
1	D	318	ASP

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	
1	A	722/728 (99%)	719 (100%)	3 (0%)	91	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	721/728 (99%)	716 (99%)	5 (1%)	84	90
1	G	721/728 (99%)	719 (100%)	2 (0%)	92	96
2	B	292/308 (95%)	286 (98%)	6 (2%)	53	75
2	E	296/308 (96%)	289 (98%)	7 (2%)	49	72
2	H	297/308 (96%)	294 (99%)	3 (1%)	76	85
3	C	223/242 (92%)	222 (100%)	1 (0%)	91	94
3	F	223/242 (92%)	222 (100%)	1 (0%)	91	94
3	I	223/242 (92%)	222 (100%)	1 (0%)	91	94
All	All	3718/3834 (97%)	3689 (99%)	29 (1%)	81	89

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	930	VAL
2	H	1039	CYS
2	B	829	VAL
1	G	1044	SER
2	B	738	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1015	GLN
3	I	179	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

3 non-standard protein/DNA/RNA residues 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	#$|Z| > 2$
2	TPO	E	893	2	8,10,11	1.65	1 (12%)	10,14,16	1.33	1 (10%)
2	TPO	H	893	2	8,10,11	1.59	1 (12%)	10,14,16	1.28	2 (20%)
2	TPO	B	893	2	8,10,11	1.71	1 (12%)	10,14,16	1.01	0

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
2	TPO	E	893	2	-	0/9/11/13	-
2	TPO	H	893	2	-	1/9/11/13	-
2	TPO	B	893	2	-	7/9/11/13	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	893	TPO	P-O1P	3.52	1.61	1.50
2	H	893	TPO	P-O1P	3.51	1.61	1.50
2	B	893	TPO	P-O1P	3.48	1.61	1.50

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	893	TPO	P-OG1-CB	-3.01	114.12	123.21
2	H	893	TPO	P-OG1-CB	-2.74	114.92	123.21
2	H	893	TPO	CG2-CB-CA	-2.18	108.87	113.16

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	893	TPO	N-CA-CB-CG2
2	B	893	TPO	N-CA-CB-OG1
2	B	893	TPO	C-CA-CB-CG2
2	B	893	TPO	O-C-CA-CB
2	B	893	TPO	CG2-CB-OG1-P

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	893	TPO	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

17 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	$\#|Z| > 2$
4	SO4	C	301	-	4,4,4	0.13	0	6,6,6	0.13	0
4	SO4	D	1203	-	4,4,4	0.13	0	6,6,6	0.07	0
4	SO4	F	302	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	D	1201	-	4,4,4	0.14	0	6,6,6	0.08	0
4	SO4	F	301	-	4,4,4	0.14	0	6,6,6	0.07	0
5	RMX	E	1101	-	38,43,43	0.92	2 (5%)	45,63,63	1.02	4 (8%)
5	RMX	H	1101	-	38,43,43	0.95	2 (5%)	45,63,63	1.04	4 (8%)
4	SO4	I	303	-	4,4,4	0.15	0	6,6,6	0.07	0
4	SO4	I	301	-	4,4,4	0.14	0	6,6,6	0.12	0
4	SO4	G	1201	-	4,4,4	0.14	0	6,6,6	0.06	0
5	RMX	B	1101	-	38,43,43	0.93	2 (5%)	45,63,63	0.97	4 (8%)
4	SO4	I	302	-	4,4,4	0.13	0	6,6,6	0.10	0
4	SO4	H	1102	-	4,4,4	0.14	0	6,6,6	0.06	0
4	SO4	A	1201	-	4,4,4	0.14	0	6,6,6	0.09	0
4	SO4	G	1202	-	4,4,4	0.15	0	6,6,6	0.06	0
4	SO4	D	1202	-	4,4,4	0.14	0	6,6,6	0.08	0
4	SO4	C	302	-	4,4,4	0.14	0	6,6,6	0.05	0

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
5	RMX	H	1101	-	-	2/16/24/24	0/5/5/5
5	RMX	B	1101	-	-	2/16/24/24	0/5/5/5
5	RMX	E	1101	-	-	2/16/24/24	0/5/5/5

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	H	1101	RMX	C1-S1	-3.47	1.73	1.79
5	B	1101	RMX	C1-S1	-3.40	1.73	1.79
5	E	1101	RMX	C1-S1	-3.33	1.73	1.79
5	B	1101	RMX	S1-N1	-2.45	1.55	1.60
5	H	1101	RMX	S1-N1	-2.44	1.55	1.60

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	E	1101	RMX	C22-N9-C23	3.52	128.87	125.42
5	H	1101	RMX	C22-N9-C23	3.37	128.72	125.42
5	B	1101	RMX	C22-N9-C23	2.89	128.26	125.42
5	H	1101	RMX	C21-C8-N5	-2.58	118.66	120.81
5	E	1101	RMX	C21-C8-N5	-2.52	118.71	120.81

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

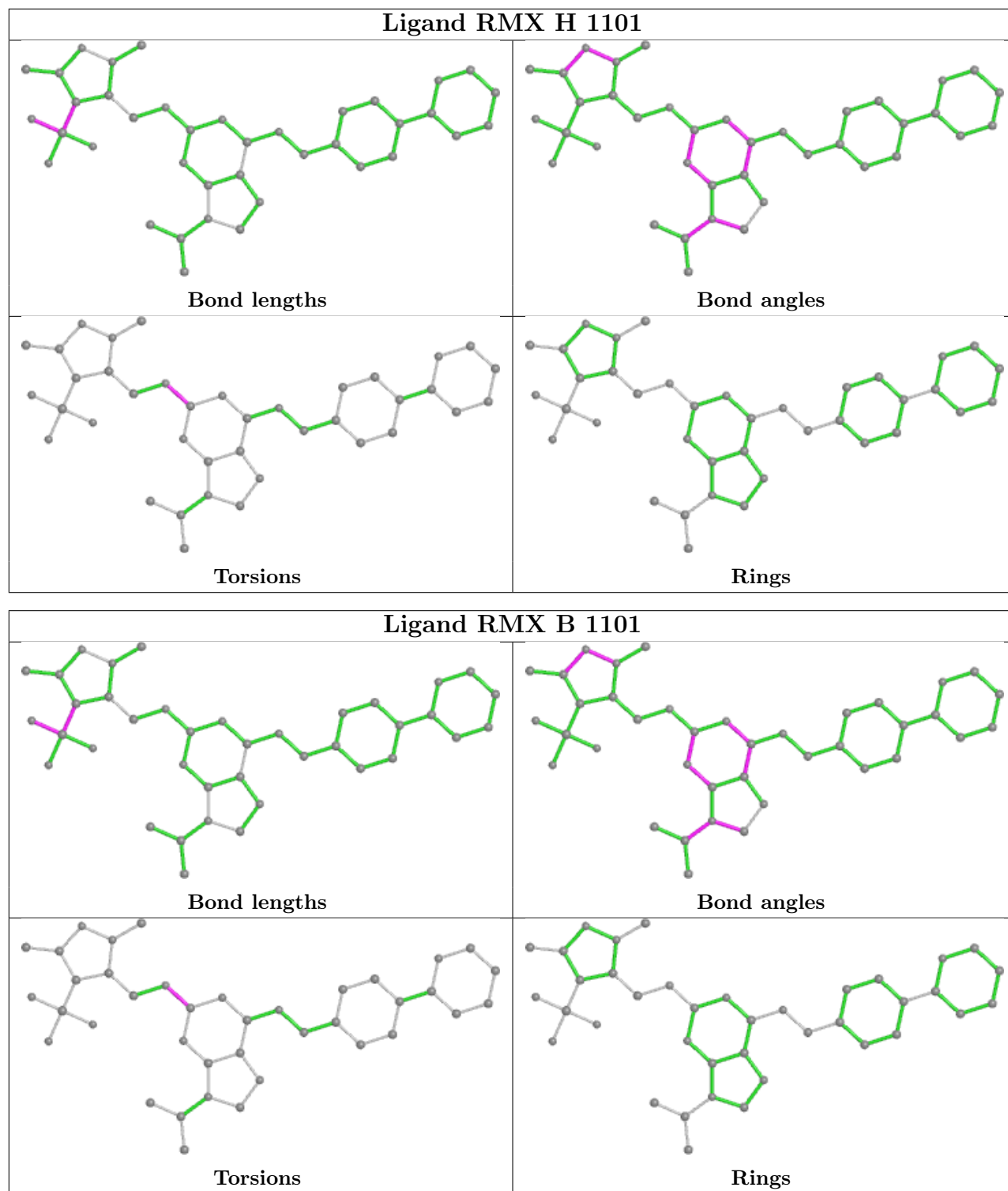| Mol	Chain	Res	Type	Atoms
5	E	1101	RMX	N10-C7-N4-C6
5	E	1101	RMX	N5-C7-N4-C6
5	B	1101	RMX	N10-C7-N4-C6
5	B	1101	RMX	N5-C7-N4-C6
5	H	1101	RMX	N10-C7-N4-C6

There are no ring outliers.

No monomer is involved in short contacts.

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 then 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.

Ligand RMX E 1101

Bond lengths

Bond angles

Torsions

Rings

Ligand RMX H 1101

Bond lengths

Bond angles

Torsions

Rings

Ligand RMX B 1101

Bond lengths

Bond angles

Torsions

Rings
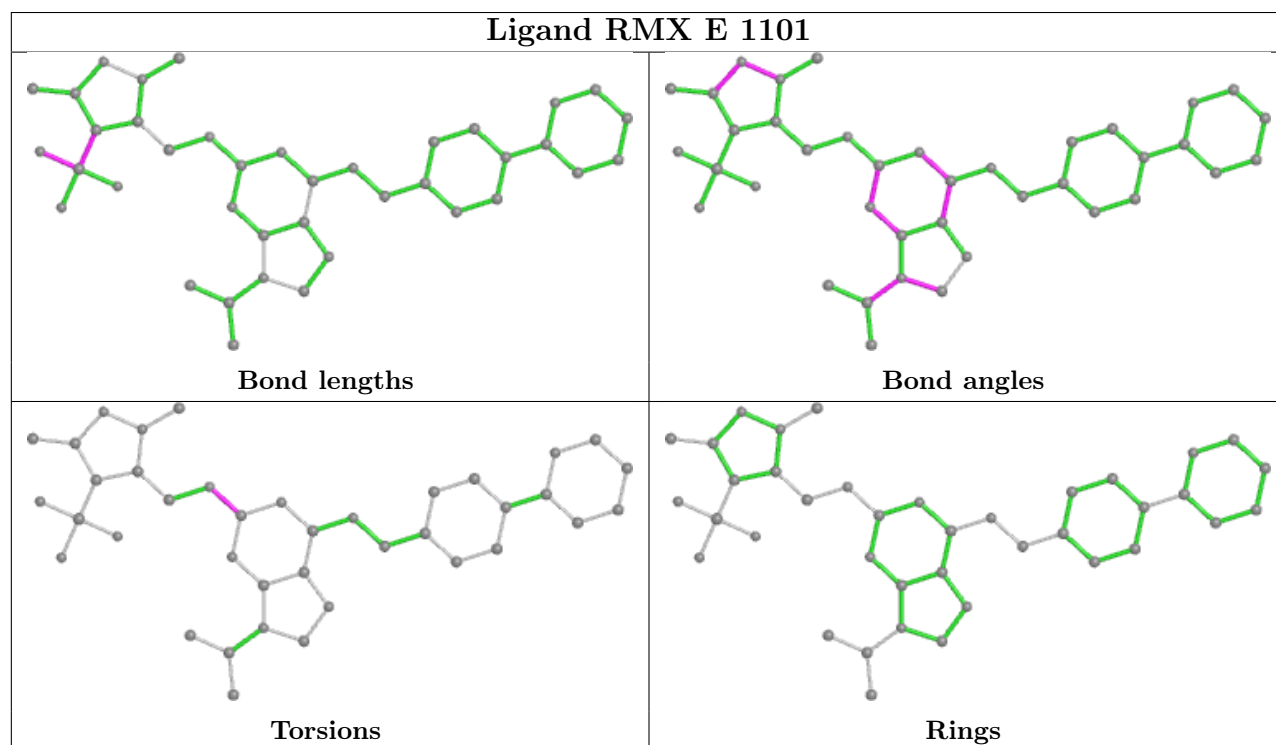
5.7 Other polymers ⓘ

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, 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	827/840 (98%)	0.55	41 (4%)	28	26	86, 122, 191, 254	0
1	D	826/840 (98%)	0.69	73 (8%)	10	10	90, 128, 201, 284	0
1	G	826/840 (98%)	0.60	71 (8%)	10	10	94, 128, 198, 287	0
2	B	324/344 (94%)	0.85	42 (12%)	3	3	95, 123, 192, 247	0
2	E	328/344 (95%)	0.93	48 (14%)	2	2	105, 133, 205, 257	0
2	H	329/344 (95%)	0.78	25 (7%)	13	12	84, 108, 196, 243	0
3	C	248/271 (91%)	0.67	4 (1%)	72	69	79, 101, 138, 215	0
3	F	248/271 (91%)	0.55	10 (4%)	38	35	97, 122, 159, 218	0
3	I	248/271 (91%)	0.66	12 (4%)	30	28	83, 104, 150, 219	0
All	All	4204/4365 (96%)	0.67	326 (7%)	13	12	79, 122, 195, 287	0

The worst 5 of 326 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	738	TYR	8.4
2	E	799	ALA	6.8
1	D	1114	TYR	6.2
1	D	1018	GLY	6.1
1	D	1016	ASN	5.7

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

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, 95th 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(Å²)	Q<0.9
2	TPO	H	893	11/12	0.65	0.41	154,155,189,189	6
2	TPO	E	893	11/12	0.74	0.21	185,207,251,251	6
2	TPO	B	893	11/12	0.76	0.20	198,214,257,257	6

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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, 95th 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(Å²)	Q<0.9
4	SO4	F	302	5/5	0.45	0.31	161,177,187,187	0
4	SO4	F	301	5/5	0.73	0.21	142,144,145,145	0
4	SO4	H	1102	5/5	0.74	0.29	153,161,167,177	0
4	SO4	C	302	5/5	0.78	0.27	138,149,156,171	0
4	SO4	D	1203	5/5	0.80	0.17	165,172,176,184	0
4	SO4	D	1202	5/5	0.83	0.20	119,128,137,143	0
4	SO4	I	302	5/5	0.84	0.22	157,167,171,176	0
5	RMX	H	1101	39/39	0.85	0.50	100,126,173,181	30
4	SO4	I	303	5/5	0.88	0.14	136,146,152,164	0
4	SO4	D	1201	5/5	0.88	0.15	136,137,141,141	0
5	RMX	B	1101	39/39	0.89	0.41	101,124,178,185	30
4	SO4	G	1202	5/5	0.90	0.24	132,134,137,138	0
5	RMX	E	1101	39/39	0.92	0.42	112,137,187,190	30
4	SO4	C	301	5/5	0.92	0.17	109,114,116,117	0
4	SO4	A	1201	5/5	0.92	0.12	133,138,140,143	0
4	SO4	I	301	5/5	0.93	0.17	108,110,114,115	0
4	SO4	G	1201	5/5	0.94	0.14	132,138,139,145	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.
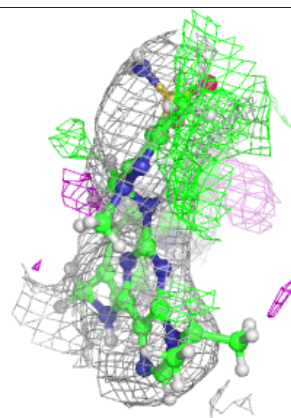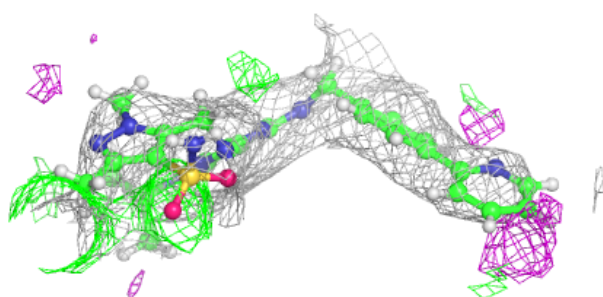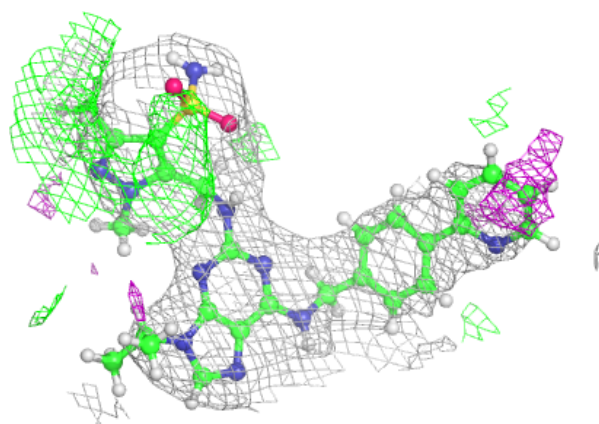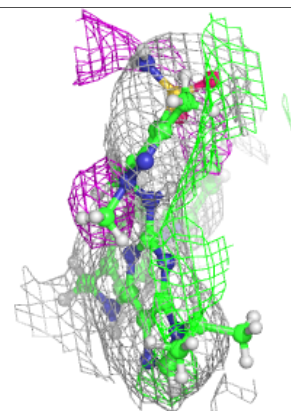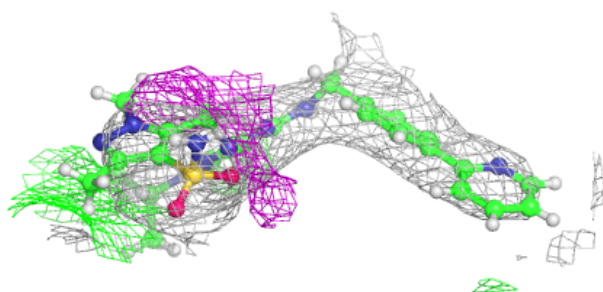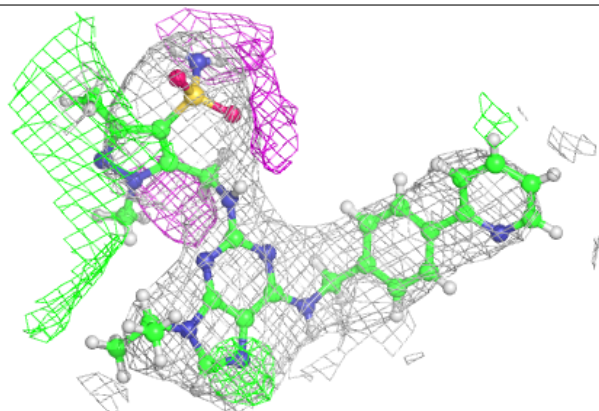
Electron density around RMX H 1101:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
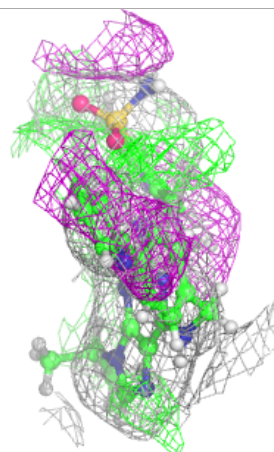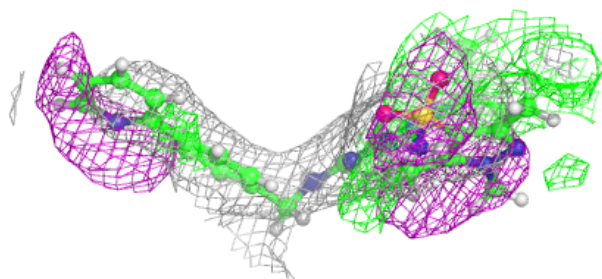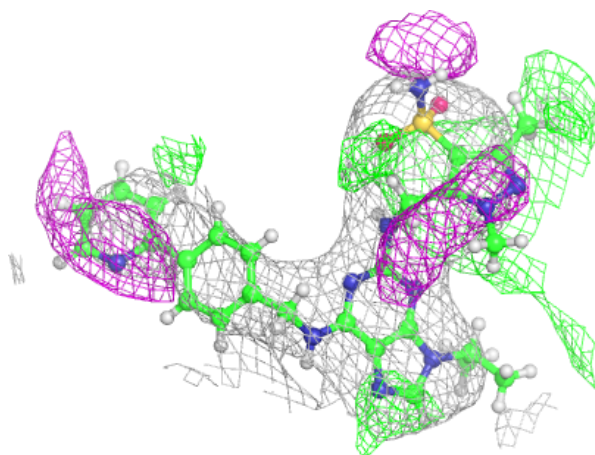
Electron density around RMX B 1101:

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

Electron density around RMX E 1101:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.