



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 15, 2024 – 09:45 PM EST

PDB ID : 6CCV  
Title : Crystal structure of a Mycobacterium smegmatis RNA polymerase transcription initiation complex with inhibitor Rifampicin  
Authors : Lilic, M.; Darst, S.A.; Campbell, E.A.  
Deposited on : 2018-02-07  
Resolution : 3.05 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

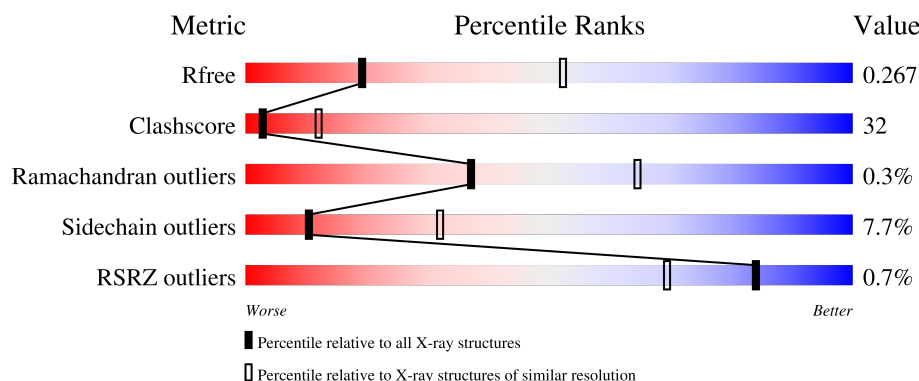
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.05 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))
$R_{free}$	164625	2258 (3.10-3.02)
Clashscore	180529	2399 (3.10-3.02)
Ramachandran outliers	177936	2269 (3.10-3.02)
Sidechain outliers	177891	2268 (3.10-3.02)
RSRZ outliers	164620	2258 (3.10-3.02)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . 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	350	
1	B	350	
1	T	350	
2	C	1169	
3	D	1317	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	E	107	
5	F	466	
6	G	17	
7	J	114	
8	O	31	
9	P	26	

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
10	SO4	D	2004	-	-	X	-
14	GLU	D	2012	-	-	X	-

## 2 Entry composition

There are 15 unique types of molecules in this entry. The entry contains 26582 atoms, of which 48 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	218	Total	C	N	O	S	0	0	0
			1605	1015	276	311	3			
1	B	233	Total	C	N	O	S	0	0	0
			1667	1054	289	322	2			
1	T	53	Total	C	N	O	S	0	0	0
			342	208	65	68	1			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1099	Total	C	N	O	S	0	0	0
			8262	5174	1450	1603	35			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1246	Total	C	N	O	S	0	0	0
			9555	5995	1720	1800	40			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	76	Total	C	N	O	0	0	0
			592	378	100	114			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	305	Total	C	N	O	S	0	0	0
			2414	1512	436	459	7			

- Molecule 6 is a protein called Unknown Peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
6	G	17	Total	C	N	O	0	0	0
			85	51	17	17			

- Molecule 7 is a protein called RNA polymerase-binding protein RbpA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	J	83	Total	C	N	O	S	0	0	0
			671	422	119	128	2			

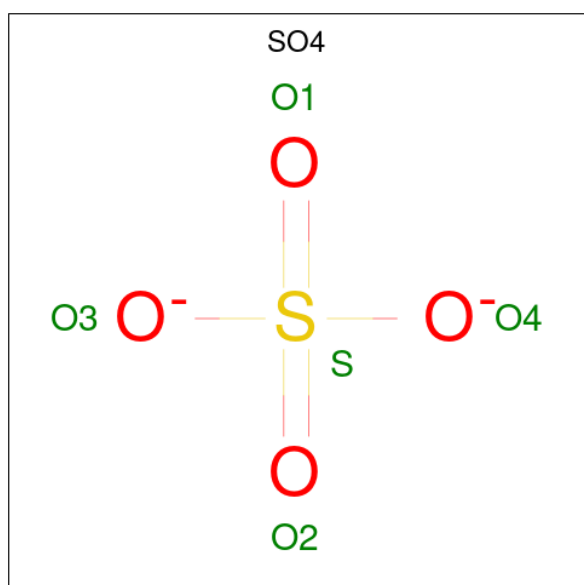
- Molecule 8 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	O	31	Total	C	N	O	P	0	0	0
			634	305	114	185	30			

- Molecule 9 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	P	26	Total	C	N	O	P	0	0	0
			526	254	94	153	25			

- Molecule 10 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



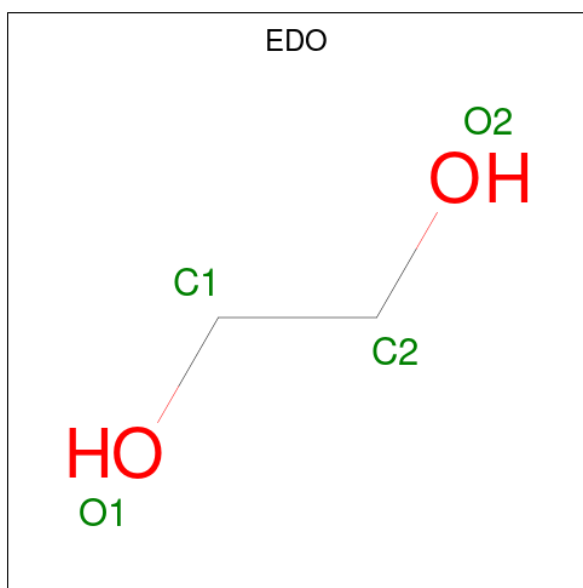
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	C	1	Total	O	S	0	0
			5	4	1		

*Continued on next page...*

*Continued from previous page...*

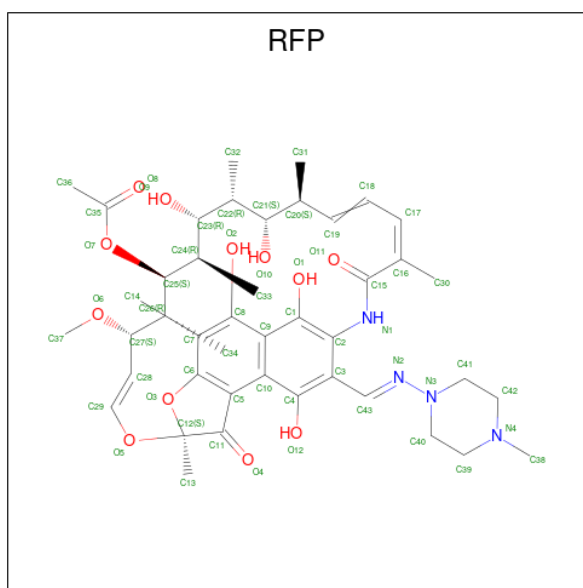
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	C	1	Total	O	S	0	0
			5	4	1		
10	C	1	Total	O	S	0	0
			5	4	1		
10	C	1	Total	O	S	0	0
			5	4	1		
10	D	1	Total	O	S	0	0
			5	4	1		
10	D	1	Total	O	S	0	0
			5	4	1		
10	D	1	Total	O	S	0	0
			5	4	1		
10	D	1	Total	O	S	0	0
			5	4	1		
10	F	1	Total	O	S	0	0
			5	4	1		
10	F	1	Total	O	S	0	0
			5	4	1		
10	F	1	Total	O	S	0	0
			5	4	1		
10	F	1	Total	O	S	0	0
			5	4	1		

- Molecule 11 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
11	C	1	Total	C	H	O	0	0
			10	2	6	2		
11	C	1	Total	C	H	O	0	0
			10	2	6	2		
11	D	1	Total	C	H	O	0	0
			10	2	6	2		
11	D	1	Total	C	H	O	0	0
			10	2	6	2		
11	D	1	Total	C	H	O	0	0
			10	2	6	2		
11	F	1	Total	C	H	O	0	0
			10	2	6	2		
11	F	1	Total	C	H	O	0	0
			10	2	6	2		

- Molecule 12 is RIFAMPICIN (three-letter code: RFP) (formula:  $C_{43}H_{58}N_4O_{12}$ ).

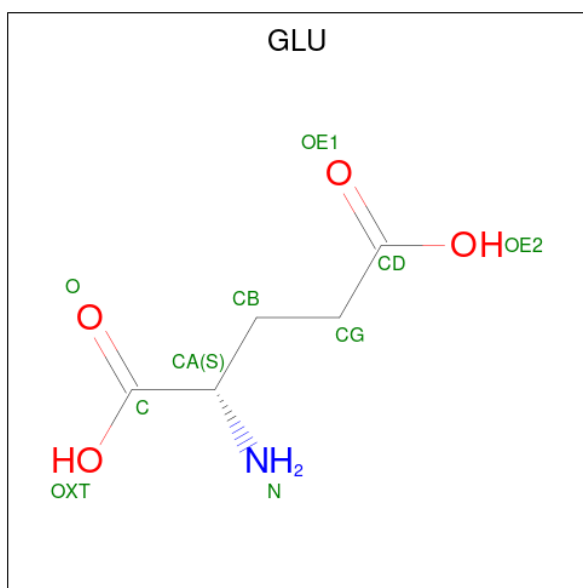


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
12	C	1	Total	C	N	O	0	0
			59	43	4	12		

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	D	2	Total	Zn	0	0
			2	2		

- Molecule 14 is GLUTAMIC ACID (three-letter code: GLU) (formula:  $C_5H_9NO_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
14	D	1	Total	C	N	O	0	0
			9	5	1	3		

- Molecule 15 is water.

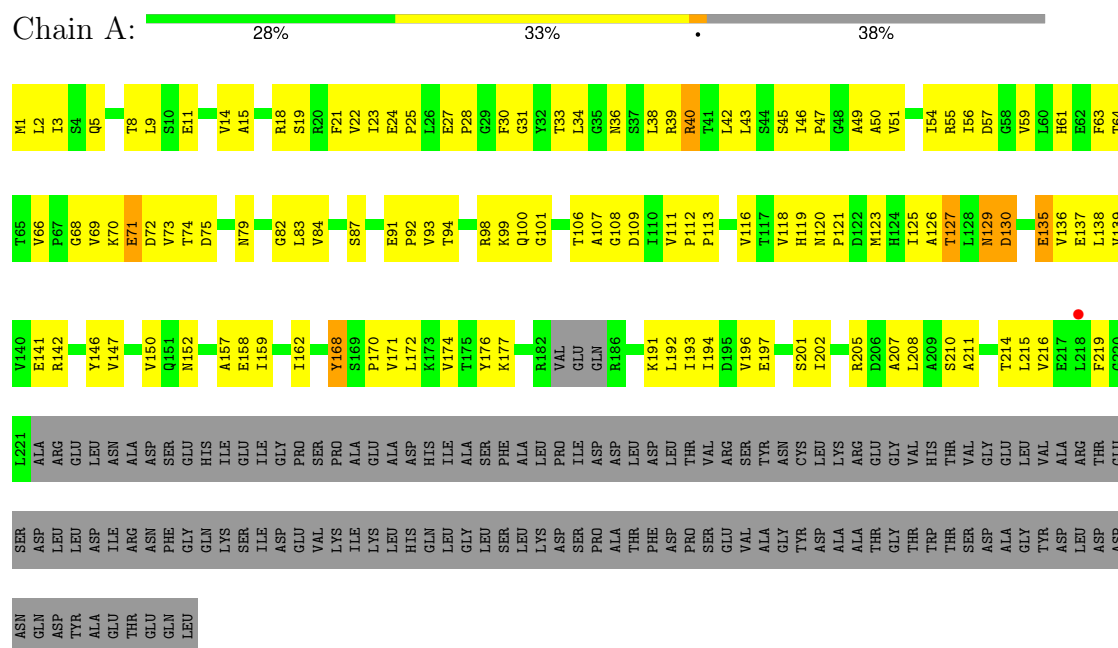
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
15	C	5	Total	O	0	0
			5	5		
15	D	8	Total	O	0	0
			8	8		
15	F	1	Total	O	0	0
			1	1		



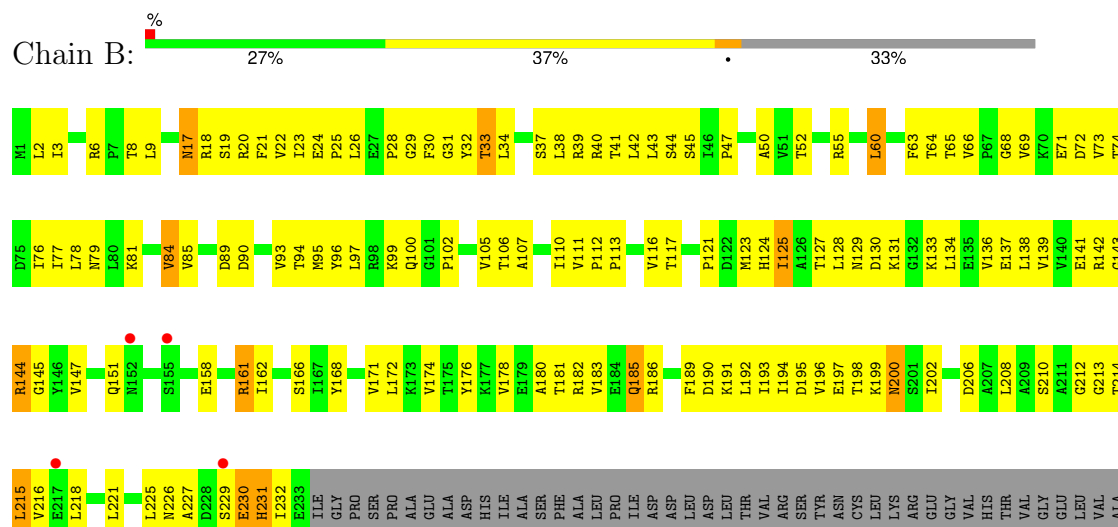
### 3 Residue-property plots

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



#### • Molecule 1: DNA-directed RNA polymerase subunit alpha



LEU	ASP	ASP	ASN	GLN	ASP	LEU	LEU	ILE	ARG	PHE	GLY	GLN	ASP	GLU	VAL	LYS	LYS	LEU	HIS	GLN	GLY	LEU	SER	LEU	LYS	ASP	SER	PRO	ALA	THR	PHE	ASP	ASP	PRO	SER	GLU	VAL	ALA	GLY	THR	TYR	ASP	ALA	ALA	GLY	ASP	THR	THR	TRP	THR	GLN	SER	ASP	ALA	GLY	THR	TYR	ASP	THR	ASP
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- Molecule 1: DNA-directed RNA polymerase subunit alpha

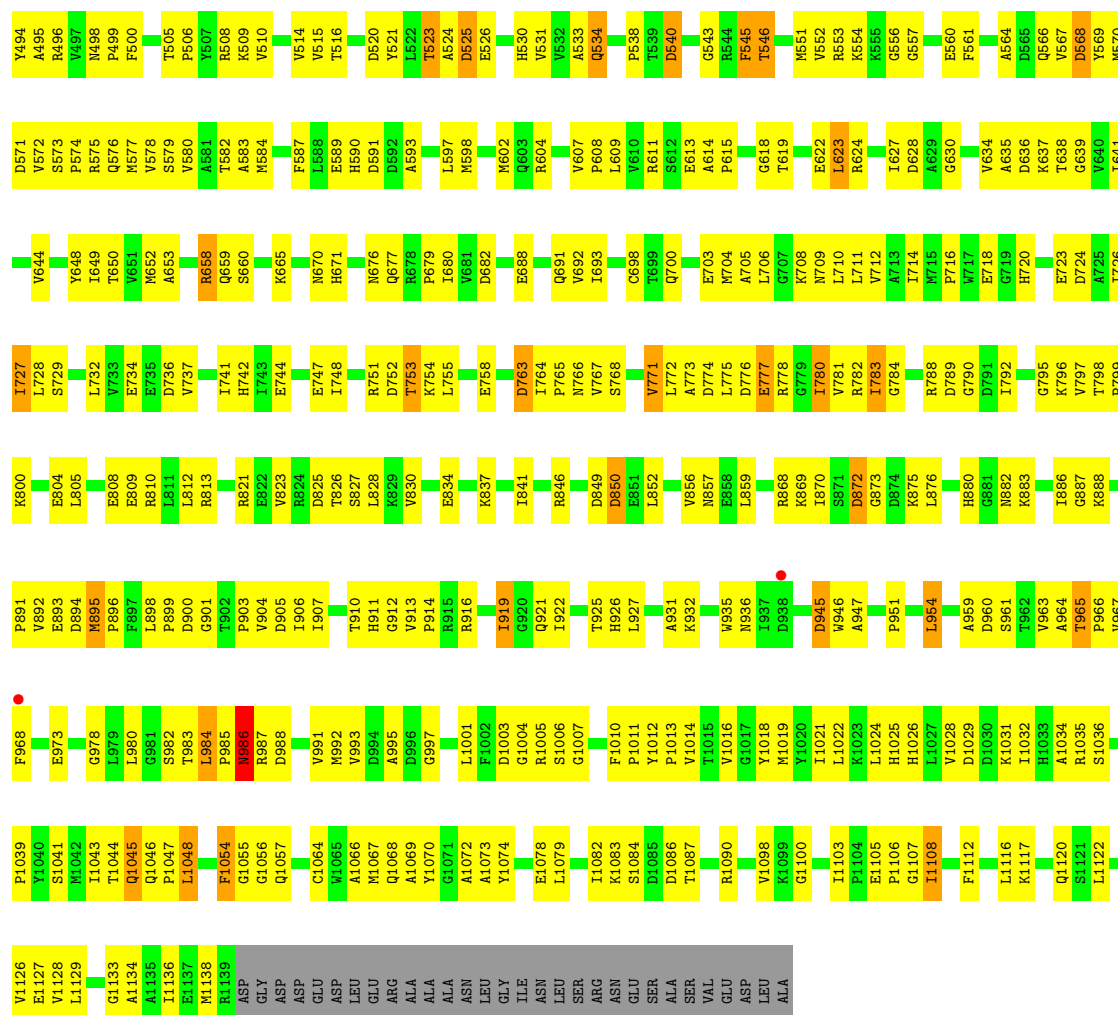
Chain T:  12% . 85%

THR	SER	THR	ALA	THR	PRO	MET
ASP	ASP	ARG	ASP	ARG	ASP	LEU
ALA	HIS	VAL	HIS	VAL	MET	ILE
GLY	ALA	GLN	ALA	GLN	ILE	THR
TYR	SER	ARG	SER	ARG	ALA	ARG
ASP	PHE	THR	PHE	THR	THR	PRO
LEU	ALA	ASP	ALA	ASP	LEU	THR
ASP	LEU	PHE	LEU	PHE	VAL	LEU
ASP	PRO	ASP	PRO	ASP	LYS	SER
ASN	ASN	LYS	1251	LYS	GLY	GLU
GLN	D252	LEU	D252	LEU	GLY	GLU
ASP	THR	ILE	THR	ILE	LYS	THR
ALA	THR	ILE	THR	ILE	LEU	VAL
ALA	2258	ASP	2258	ASP	GLU	THR
GLU	2259	VAL	2259	VAL	ASP	ALA
THR	2260	GLU	2260	GLU	VAL	ILE
THR	2261	GLU	2261	GLU	GLU	GLU
GLU	2264	THR	2264	THR	LEU	ASN
GLN	2269	LYS	2269	LYS	VAL	SER
LEU	2272	ASN	2272	ASN	VAL	ARG
	2279	SER	2279	SER	GLU	ARG
	2282	ILE	2282	ILE	GLU	LEU
	2283	ARG	2283	ARG	GLY	LYS
	2284	THR	2284	THR	GLY	PHE
	2290	ALA	2290	ALA	VAL	GLY
	2303	ALA	2303	ALA	GLN	GLU
	LEU	GLY	LEU	GLY	PRO	PRO
	LEU	LEU	LEU	LEU	ASN	THR
	SER	VAL	SER	VAL	TYR	GLY
	LEU	GLU	LEU	GLU	LEU	SER
	LEU	LEU	LEU	LEU	ARG	SER
	LYS	PHE	LYS	PHE	ILE	ARG
	ASP	ASP	ASP	ASP	GLY	ARG
	SER	LEU	SER	LEU	ARG	GLN
	PRO	ALA	PRO	ALA	GLY	THR
	ALA	ARG	ALA	ARG	PRO	LEU
	THR	GLU	THR	GLU	GLY	LEU
	PHE	LEU	PHE	LEU	VAL	SER
	ASP	ASN	ASP	ASN	THR	SER
	PRO	ALA	PRO	ALA	ILE	PRO
	SER	ASP	SER	ASP	TYR	GLY
	GLU	SER	GLU	SER	SER	ASP
	VAL	GLU	VAL	GLU	PRO	ALA
	ALA	HIS	ALA	HIS	VAL	ALA
	GLY	ILE	GLY	ILE	VAL	VAL
	TYR	GLU	TYR	GLU	PRO	THR
	ASP	ILE	ASP	ILE	LYS	SER
	ALA	THR	ALA	THR	VAL	ILE
	ALA	PRO	ALA	PRO	ALA	ARG
	THR	GLY	THR	GLY	GLY	THR
	GLY	PRO	GLY	PRO	THR	ASP
	THR	ALA	THR	ALA	VAL	GLY
	THR	GLU	THR	GLU	HIS	VAL
					ASN	THR

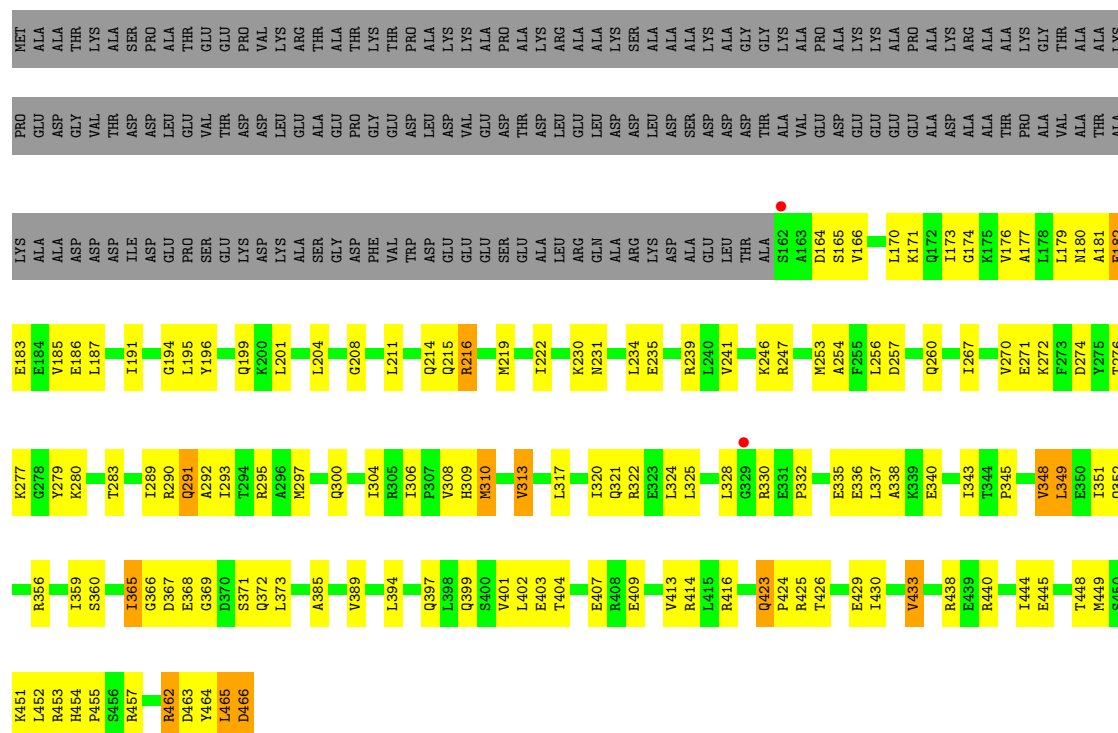
- Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:  46% 44% 6%

MET	S85	V166	T239	T325	R412	L440	L441	L442	L443	L444	L445	L446	L447	L448	L449	L450	L451	L452	L453	L454	L455	L456	L457	L458	L459	L460	L461	L462	L463	L464	L465	L466	L467	L468	L469	L470	L471	L472	L473	L474	L475	L476	L477	L478	L479	L480	L481	L482	L483	L484	L485	L486	L487	L488	L489	L490	L491	L492	L493	L494	L495	L496	L497	L498	L499	L500	L501	L502	L503	L504	L505	L506	L507	L508	L509	L510	L511	L512	L513	L514	L515	L516	L517	L518	L519	L520	L521	L522	L523	L524	L525	L526	L527	L528	L529	L530	L531	L532	L533	L534	L535	L536	L537	L538	L539	L540	L541	L542	L543	L544	L545	L546	L547	L548	L549	L550	L551	L552	L553	L554	L555	L556	L557	L558	L559	L560	L561	L562	L563	L564	L565	L566	L567	L568	L569	L570	L571	L572	L573	L574	L575	L576	L577	L578	L579	L580	L581	L582	L583	L584	L585	L586	L587	L588	L589	L590	L591	L592	L593	L594	L595	L596	L597	L598	L599	L600	L601	L602	L603	L604	L605	L606	L607	L608	L609	L610	L611	L612	L613	L614	L615	L616	L617	L618	L619	L620	L621	L622	L623	L624	L625	L626	L627	L628	L629	L630	L631	L632	L633	L634	L635	L636	L637	L638	L639	L640	L641	L642	L643	L644	L645	L646	L647	L648	L649	L650	L651	L652	L653	L654	L655	L656	L657	L658	L659	L660	L661	L662	L663	L664	L665	L666	L667	L668	L669	L670	L671	L672	L673	L674	L675	L676	L677	L678	L679	L680	L681	L682	L683	L684	L685	L686	L687	L688	L689	L690	L691	L692	L693	L694	L695	L696	L697	L698	L699	L700	L701	L702	L703	L704	L705	L706	L707	L708	L709	L710	L711	L712	L713	L714	L715	L716	L717	L718	L719	L720	L721	L722	L723	L724	L725	L726	L727	L728	L729	L730	L731	L732	L733	L734	L735	L736	L737	L738	L739	L740	L741	L742	L743	L744	L745	L746	L747	L748	L749	L750	L751	L752	L753	L754	L755	L756	L757	L758	L759	L760	L761	L762	L763	L764	L765	L766	L767	L768	L769	L770	L771	L772	L773	L774	L775	L776	L777	L778	L779	L780	L781	L782	L783	L784	L785	L786	L787	L788	L789	L790	L791	L792	L793	L794	L795	L796	L797	L798	L799	L800	L801	L802	L803	L804	L805	L806	L807	L808	L809	L810	L811	L812	L813	L814	L815	L816	L817	L818	L819	L820	L821	L822	L823	L824	L825	L826	L827	L828	L829	L830	L831	L832	L833	L834	L835	L836	L837	L838	L839	L840	L841	L842	L843	L844	L845	L846	L847	L848	L849	L850	L851	L852	L853	L854	L855	L856	L857	L858	L859	L860	L861	L862	L863	L864	L865	L866	L867	L868	L869	L870	L871	L872	L873	L874	L875	L876	L877	L878	L879	L880	L881	L882	L883	L884	L885	L886	L887	L888	L889	L890	L891	L892	L893	L894	L895	L896	L897	L898	L899	L900	L901	L902	L903	L904	L905	L906	L907	L908	L909	L910	L911	L912	L913	L914	L915	L916	L917	L918	L919	L920	L921	L922	L923	L924	L925	L926	L927	L928	L929	L930	L931	L932	L933	L934	L935	L936	L937	L938	L939	L940	L941	L942	L943	L944	L945	L946	L947	L948	L949	L950	L951	L952	L953	L954	L955	L956	L957	L958	L959	L960	L961	L962	L963	L964	L965	L966	L967	L968	L969	L970	L971	L972	L973	L974	L975	L976	L977	L978	L979	L980	L981	L982	L983	L984	L985	L986	L987	L988	L989	L990	L991	L992	L993	L994	L995	L996	L997	L998	L999	L1000
-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------







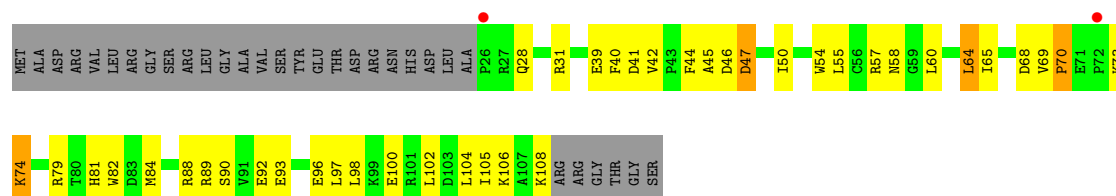
- Molecule 6: Unknown Peptide

Chain G:  100%

There are no outlier residues recorded for this chain.

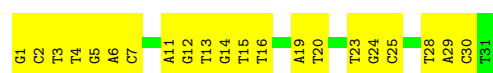
- Molecule 7: RNA polymerase-binding protein RbpA

Chain J: 



- Molecule 8: DNA (31-MER)

Chain 0:  32% 68%



- Molecule 9: DNA (26-MER)

Chain P:  62% 38%

A1	G2	C3	A4	C5	A11	A12	T17	G20	T21	G22	A23	A24	G25	C26
----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	132.35Å 162.32Å 139.40Å 90.00° 107.37° 90.00°	Depositor
Resolution (Å)	57.12 – 3.05 57.12 – 3.05	Depositor EDS
% Data completeness (in resolution range)	98.0 (57.12-3.05) 97.9 (57.12-3.05)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.22 (at 2.86Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, $R_{free}$	0.223 , 0.269 0.225 , 0.267	Depositor DCC
$R_{free}$ test set	105258 reflections (1.66%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	78.8	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 60.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	26582	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% 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: EDO, RFP, SO4, 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	0.23	0/1629	0.45	0/2220
1	B	0.23	0/1693	0.43	0/2316
1	T	0.22	0/343	0.37	0/468
2	C	0.24	0/8408	0.43	0/11428
3	D	0.24	0/9706	0.43	1/13140 (0.0%)
4	E	0.24	0/604	0.43	0/822
5	F	0.22	0/2445	0.39	0/3300
7	J	0.23	0/685	0.42	0/927
8	O	0.51	0/710	0.94	0/1095
9	P	0.57	0/589	0.93	0/906
All	All	0.26	0/26812	0.47	1/36622 (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
2	C	0	5
3	D	0	3
All	All	0	8

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	579	LEU	CA-CB-CG	5.98	129.06	115.30

There are no chirality outliers.

5 of 8 planarity outliers are listed below:



Mol	Chain	Res	Type	Group
2	C	1133	GLY	Peptide
2	C	324	LEU	Peptide
2	C	540	ASP	Peptide
2	C	985	PRO	Peptide
2	C	986	ASN	Peptide

## 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	1605	0	1623	141	0
1	B	1667	0	1636	166	0
1	T	342	0	275	15	0
2	C	8262	0	8009	627	1
3	D	9555	0	9509	600	1
4	E	592	0	583	40	0
5	F	2414	0	2434	151	0
6	G	85	0	19	0	0
7	J	671	0	660	48	0
8	O	634	0	350	40	0
9	P	526	0	296	15	0
10	C	20	0	0	3	0
10	D	25	0	0	3	0
10	F	20	0	0	1	0
11	C	8	12	12	0	0
11	D	16	24	24	3	0
11	F	8	12	12	3	0
12	C	59	0	58	9	0
13	D	2	0	0	0	0
14	D	9	0	5	10	0
15	C	5	0	0	0	0
15	D	8	0	0	1	0
15	F	1	0	0	0	0
All	All	26534	48	25505	1669	1

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 32.

The worst 5 of 1669 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:603:THR:HG22	3:D:604:LYS:HG3	1.19	1.15
1:A:197:GLU:OE1	2:C:987:ARG:NH1	1.87	1.07
2:C:203:LEU:HG	2:C:217:ILE:HG22	1.35	1.07
2:C:771:VAL:HG22	2:C:772:LEU:HD12	1.35	1.07
2:C:53:GLU:OE2	2:C:60:ARG:NH1	1.89	1.06

All (1) 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)
2:C:658:ARG:NH1	3:D:147:GLU:OE1[2_356]	1.99	0.21

## 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	214/350 (61%)	204 (95%)	9 (4%)	1 (0%)	25	54
1	B	231/350 (66%)	211 (91%)	20 (9%)	0	100	100
1	T	51/350 (15%)	51 (100%)	0	0	100	100
2	C	1093/1169 (94%)	1040 (95%)	51 (5%)	2 (0%)	44	71
3	D	1234/1317 (94%)	1185 (96%)	45 (4%)	4 (0%)	37	64
4	E	72/107 (67%)	66 (92%)	5 (7%)	1 (1%)	9	30
5	F	303/466 (65%)	299 (99%)	4 (1%)	0	100	100
7	J	81/114 (71%)	77 (95%)	3 (4%)	1 (1%)	11	34
All	All	3279/4223 (78%)	3133 (96%)	137 (4%)	9 (0%)	37	64

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	931	ALA

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	1010	THR
2	C	325	THR
3	D	1086	ARG
2	C	850	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	174/297 (59%)	161 (92%)	13 (8%)	11	33
1	B	171/297 (58%)	158 (92%)	13 (8%)	11	32
1	T	26/297 (9%)	25 (96%)	1 (4%)	28	55
2	C	860/984 (87%)	798 (93%)	62 (7%)	12	35
3	D	989/1095 (90%)	913 (92%)	76 (8%)	10	32
4	E	62/86 (72%)	51 (82%)	11 (18%)	1	5
5	F	253/379 (67%)	236 (93%)	17 (7%)	13	37
7	J	72/98 (74%)	65 (90%)	7 (10%)	6	22
All	All	2607/3533 (74%)	2407 (92%)	200 (8%)	10	32

5 of 200 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	456	VAL
3	D	876	LEU
1	T	284	LEU
3	D	558	LEU
3	D	736	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	684	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	888	HIS
7	J	58	ASN
3	D	692	GLN
3	D	778	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 25 ligands modelled in this entry, 2 are monoatomic - leaving 23 for Mogul analysis.

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	$\# Z  > 2$
10	SO4	C	1201	-	4,4,4	0.24	0	6,6,6	0.08	0
10	SO4	D	2006	-	4,4,4	0.24	0	6,6,6	0.07	0
11	EDO	D	2008	-	3,3,3	0.42	0	2,2,2	0.35	0
11	EDO	D	2011	-	3,3,3	0.42	0	2,2,2	0.31	0
12	RFP	C	1205	-	63,63,63	2.73	18 (28%)	94,94,94	1.89	20 (21%)
10	SO4	D	2003	-	4,4,4	0.23	0	6,6,6	0.10	0
10	SO4	F	503	-	4,4,4	0.24	0	6,6,6	0.08	0
11	EDO	C	1207	-	3,3,3	0.41	0	2,2,2	0.37	0
10	SO4	F	502	-	4,4,4	0.24	0	6,6,6	0.07	0
10	SO4	C	1203	-	4,4,4	0.23	0	6,6,6	0.07	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
10	SO4	D	2005	-	4,4,4	0.24	0	6,6,6	0.08	0
10	SO4	C	1202	-	4,4,4	0.24	0	6,6,6	0.08	0
10	SO4	F	504	-	4,4,4	0.24	0	6,6,6	0.08	0
10	SO4	D	2010	-	4,4,4	0.24	0	6,6,6	0.07	0
11	EDO	D	2007	-	3,3,3	0.42	0	2,2,2	0.35	0
10	SO4	F	501	-	4,4,4	0.23	0	6,6,6	0.07	0
11	EDO	F	506	-	3,3,3	0.42	0	2,2,2	0.31	0
11	EDO	D	2009	-	3,3,3	0.44	0	2,2,2	0.31	0
14	GLU	D	2012	-	7,8,9	0.92	0	4,9,11	1.10	0
10	SO4	D	2004	-	4,4,4	0.23	0	6,6,6	0.08	0
10	SO4	C	1206	-	4,4,4	0.24	0	6,6,6	0.06	0
11	EDO	C	1204	-	3,3,3	0.40	0	2,2,2	0.46	0
11	EDO	F	505	-	3,3,3	0.42	0	2,2,2	0.31	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
11	EDO	D	2009	-	-	1/1/1/1	-
12	RFP	C	1205	-	-	12/60/85/85	0/5/5/5
11	EDO	D	2011	-	-	1/1/1/1	-
14	GLU	D	2012	-	-	3/6/7/9	-
11	EDO	C	1207	-	-	0/1/1/1	-
11	EDO	D	2008	-	-	0/1/1/1	-
11	EDO	D	2007	-	-	0/1/1/1	-
11	EDO	C	1204	-	-	1/1/1/1	-
11	EDO	F	505	-	-	1/1/1/1	-
11	EDO	F	506	-	-	1/1/1/1	-

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	C	1205	RFP	C17-C16	8.76	1.58	1.34
12	C	1205	RFP	C15-N1	7.39	1.50	1.35
12	C	1205	RFP	C1-C9	7.04	1.63	1.43
12	C	1205	RFP	C18-C19	6.92	1.59	1.33
12	C	1205	RFP	C29-C28	5.89	1.60	1.30

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
12	C	1205	RFP	C2-C3-C4	6.72	123.46	119.19
12	C	1205	RFP	C2-C3-C43	-6.48	117.11	123.99
12	C	1205	RFP	O3-C6-C7	6.17	131.63	121.16
12	C	1205	RFP	O7-C35-C36	5.34	120.61	111.09
12	C	1205	RFP	C20-C21-C22	-3.54	107.80	114.97

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
12	C	1205	RFP	C4-C3-C43-N2
12	C	1205	RFP	C13-C12-O5-C29
12	C	1205	RFP	C26-C27-C28-C29
12	C	1205	RFP	O6-C27-C28-C29
12	C	1205	RFP	C43-N2-N3-C40

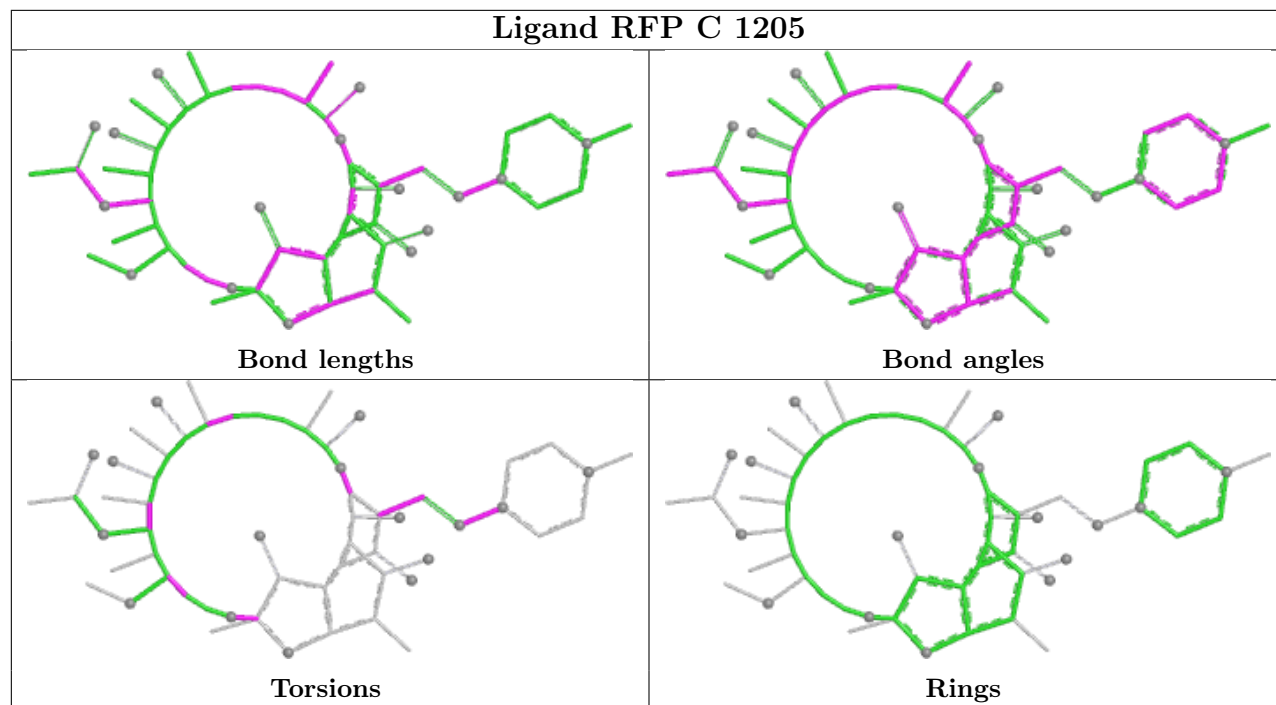
There are no ring outliers.

11 monomers are involved in 31 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	C	1201	SO4	1	0
11	D	2011	EDO	2	0
12	C	1205	RFP	9	0
10	F	502	SO4	1	0
10	C	1203	SO4	1	0
10	D	2010	SO4	1	0
11	D	2007	EDO	1	0
11	F	506	EDO	3	0
14	D	2012	GLU	10	0
10	D	2004	SO4	2	0
10	C	1206	SO4	1	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 [i](#)

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

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 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	218/350 (62%)	0.01	1 (0%) 87 74	68, 97, 128, 146	0
1	B	233/350 (66%)	0.17	4 (1%) 69 49	85, 121, 145, 161	0
1	T	53/350 (15%)	0.42	0 100 100	123, 159, 184, 195	0
2	C	1099/1169 (94%)	-0.01	10 (0%) 81 65	47, 93, 154, 176	0
3	D	1246/1317 (94%)	-0.23	3 (0%) 92 85	41, 84, 140, 168	0
4	E	76/107 (71%)	-0.23	2 (2%) 57 37	61, 89, 131, 144	0
5	F	305/466 (65%)	-0.28	2 (0%) 84 69	46, 86, 135, 170	0
6	G	0/17	-	-	-	-
7	J	83/114 (72%)	0.03	2 (2%) 59 40	70, 111, 158, 173	0
8	O	31/31 (100%)	-0.85	0 100 100	59, 73, 96, 100	0
9	P	26/26 (100%)	-0.63	0 100 100	66, 80, 94, 104	0
All	All	3370/4297 (78%)	-0.11	24 (0%) 84 69	41, 92, 150, 195	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	155	SER	3.7
3	D	1171	SER	3.6
4	E	24	SER	2.8
2	C	229	LEU	2.8
7	J	72	PRO	2.7

### 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 monosaccharides in this entry.

## 6.4 Ligands [i](#)

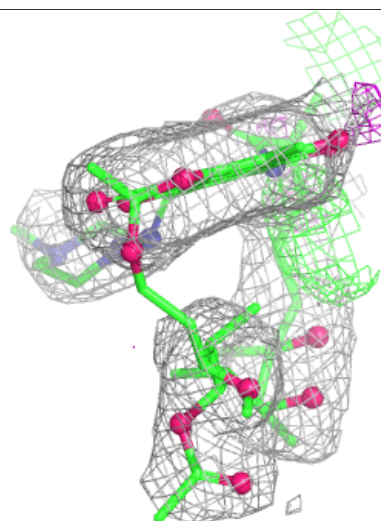
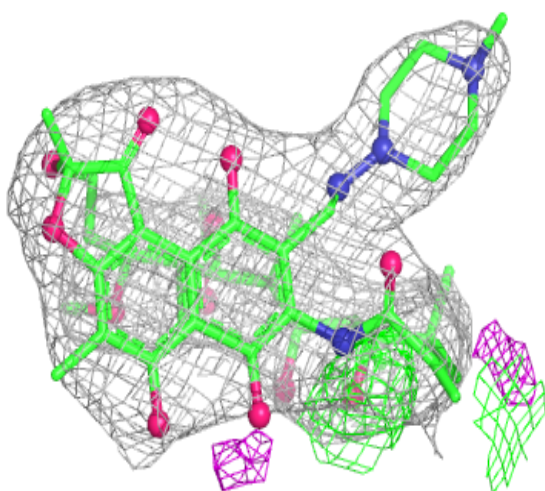
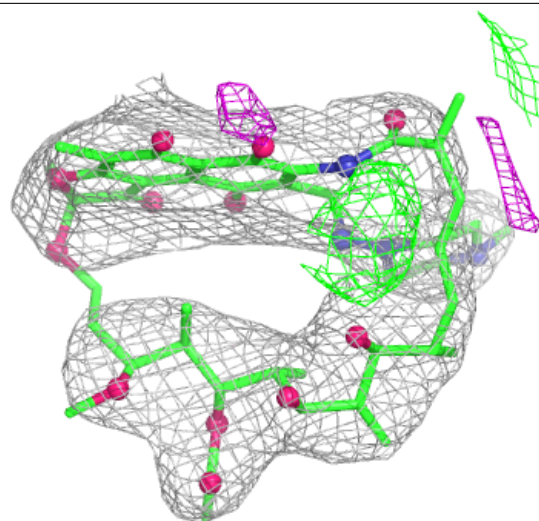
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
10	SO4	F	504	5/5	0.63	0.14	121,127,136,142	0
11	EDO	D	2007	4/4	0.69	0.15	75,94,113,123	0
11	EDO	D	2008	4/4	0.70	0.17	89,107,111,117	0
10	SO4	C	1203	5/5	0.71	0.09	114,127,160,168	0
10	SO4	C	1202	5/5	0.72	0.11	119,126,141,149	0
11	EDO	C	1207	4/4	0.73	0.25	94,113,134,134	0
10	SO4	F	503	5/5	0.73	0.12	104,110,129,130	0
10	SO4	C	1206	5/5	0.73	0.14	105,119,155,236	0
11	EDO	D	2011	4/4	0.77	0.23	73,87,105,105	0
11	EDO	F	505	4/4	0.78	0.15	81,97,121,121	0
10	SO4	D	2010	5/5	0.80	0.12	116,117,152,170	0
10	SO4	C	1201	5/5	0.81	0.11	101,114,125,135	0
10	SO4	D	2006	5/5	0.81	0.14	116,124,133,140	0
14	GLU	D	2012	9/10	0.83	0.15	90,97,100,110	0
10	SO4	D	2005	5/5	0.85	0.08	97,122,136,138	0
11	EDO	F	506	4/4	0.88	0.29	110,132,156,188	0
11	EDO	C	1204	4/4	0.89	0.13	61,74,92,92	0
10	SO4	F	501	5/5	0.90	0.06	104,105,128,134	0
10	SO4	F	502	5/5	0.90	0.07	79,101,117,120	0
12	RFP	C	1205	59/59	0.92	0.11	47,69,104,119	0
10	SO4	D	2003	5/5	0.93	0.08	64,76,98,102	0
10	SO4	D	2004	5/5	0.95	0.08	87,91,121,132	0
11	EDO	D	2009	4/4	0.96	0.12	65,87,104,104	0
13	ZN	D	2002	1/1	0.99	0.07	120,120,120,120	0
13	ZN	D	2001	1/1	0.99	0.11	129,129,129,129	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 RFP C 1205:**

$2mF_o-DF_c$  (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.