



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 27, 2016 – 11:41 AM EDT

PDB ID : 5D4D  
Title : Crystal structure of Thermus thermophilus product complex for transcription initiation with NAD and CTP  
Authors : Zhang, Y.; Ebright, R.H.  
Deposited on : 2015-08-07  
Resolution : 3.00 Å(reported)

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

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

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027939

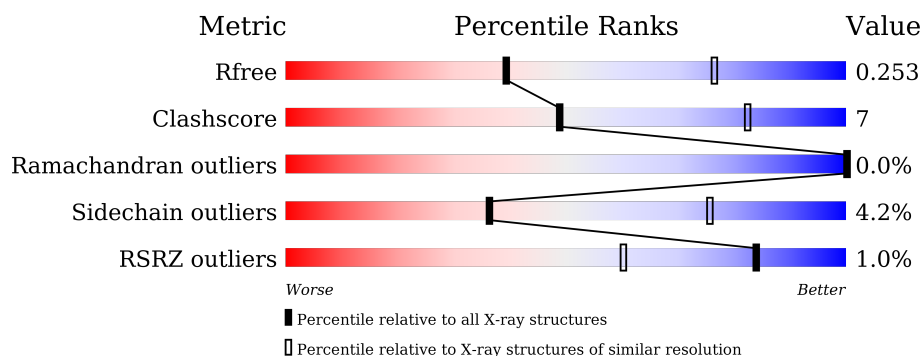
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.










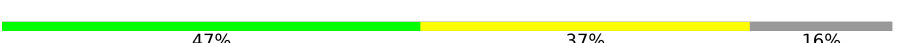


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	315	<div> <div>56%</div> <div>17%</div> <div>27%</div> </div>
1	B	315	<div> <div>56%</div> <div>14%</div> <div>30%</div> </div>
1	K	315	<div> <div>54%</div> <div>18%</div> <div>28%</div> </div>
1	L	315	<div> <div>52%</div> <div>17%</div> <div>30%</div> </div>
2	C	1119	<div> <div>3%</div> <div>80%</div> <div>18%</div> <div>••</div> </div>
2	M	1119	<div> <div>3%</div> <div>72%</div> <div>24%</div> <div>••</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	D	1524	
3	N	1524	
4	E	99	
4	O	99	
5	F	443	
5	P	443	
6	G	19	
6	R	19	
7	H	27	
7	S	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
13	A	R	101	-	-	-	X

## 2 Entry composition

There are 14 unique types of molecules in this entry. The entry contains 57349 atoms, of which 0 are hydrogens and 0 are deuteriums.

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	229	Total	C	N	O	S	0	0	0
			1797	1147	313	335	2			
1	B	222	Total	C	N	O	S	0	0	0
			1750	1120	303	325	2			
1	K	228	Total	C	N	O	S	0	0	0
			1792	1144	312	334	2			
1	L	222	Total	C	N	O	S	0	0	0
			1750	1120	303	325	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1108	Total	C	N	O	S	0	0	0
			8747	5536	1561	1626	24			
2	M	1091	Total	C	N	O	S	0	0	0
			8611	5449	1539	1600	23			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1485	Total	C	N	O	S	0	0	0
			11730	7435	2066	2194	35			
3	N	1483	Total	C	N	O	S	0	0	0
			11716	7427	2064	2190	35			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			758	483	132	139	4			
4	O	94	Total	C	N	O	S	0	0	0
			758	483	132	139	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total	C	N	O	S	0	0	0
			2807	1770	509	524	4			
5	P	316	Total	C	N	O	S	0	0	0
			2574	1624	466	480	4			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	initiating methionine	UNP Q72L95
F	-18	GLY	-	expression tag	UNP Q72L95
F	-17	SER	-	expression tag	UNP Q72L95
F	-16	SER	-	expression tag	UNP Q72L95
F	-15	HIS	-	expression tag	UNP Q72L95
F	-14	HIS	-	expression tag	UNP Q72L95
F	-13	HIS	-	expression tag	UNP Q72L95
F	-12	HIS	-	expression tag	UNP Q72L95
F	-11	HIS	-	expression tag	UNP Q72L95
F	-10	HIS	-	expression tag	UNP Q72L95
F	-9	SER	-	expression tag	UNP Q72L95
F	-8	SER	-	expression tag	UNP Q72L95
F	-7	GLY	-	expression tag	UNP Q72L95
F	-6	LEU	-	expression tag	UNP Q72L95
F	-5	VAL	-	expression tag	UNP Q72L95
F	-4	PRO	-	expression tag	UNP Q72L95
F	-3	ARG	-	expression tag	UNP Q72L95
F	-2	GLY	-	expression tag	UNP Q72L95
F	-1	SER	-	expression tag	UNP Q72L95
F	0	HIS	-	expression tag	UNP Q72L95
F	46	THR	ALA	conflict	UNP Q72L95
P	-19	MET	-	initiating methionine	UNP Q72L95
P	-18	GLY	-	expression tag	UNP Q72L95
P	-17	SER	-	expression tag	UNP Q72L95
P	-16	SER	-	expression tag	UNP Q72L95
P	-15	HIS	-	expression tag	UNP Q72L95
P	-14	HIS	-	expression tag	UNP Q72L95
P	-13	HIS	-	expression tag	UNP Q72L95
P	-12	HIS	-	expression tag	UNP Q72L95
P	-11	HIS	-	expression tag	UNP Q72L95
P	-10	HIS	-	expression tag	UNP Q72L95
P	-9	SER	-	expression tag	UNP Q72L95
P	-8	SER	-	expression tag	UNP Q72L95
P	-7	GLY	-	expression tag	UNP Q72L95

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
P	-6	LEU	-	expression tag	UNP Q72L95
P	-5	VAL	-	expression tag	UNP Q72L95
P	-4	PRO	-	expression tag	UNP Q72L95
P	-3	ARG	-	expression tag	UNP Q72L95
P	-2	GLY	-	expression tag	UNP Q72L95
P	-1	SER	-	expression tag	UNP Q72L95
P	0	HIS	-	expression tag	UNP Q72L95
P	46	THR	ALA	conflict	UNP Q72L95

- Molecule 6 is a DNA chain called DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*T  
P\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	16	Total	C	N	O	P	0	0	0
			328	157	62	94	15			
6	R	16	Total	C	N	O	P	0	0	0
			328	157	62	94	15			

- Molecule 7 is a DNA chain called DNA (27-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	21	Total	C	N	O	P	0	0	0
			435	207	87	121	20			
7	S	21	Total	C	N	O	P	0	0	0
			434	206	87	121	20			

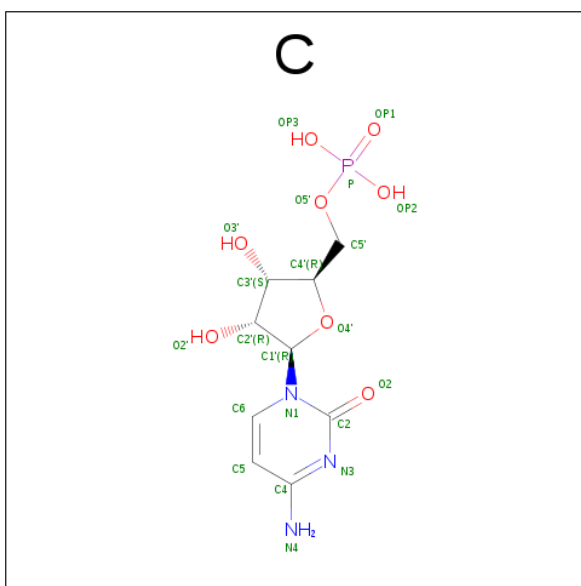
- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	P	1	Total	Mg	0	0
			1	1		
8	D	3	Total	Mg	0	0
			3	3		
8	B	1	Total	Mg	0	0
			1	1		
8	N	3	Total	Mg	0	0
			3	3		
8	L	1	Total	Mg	0	0
			1	1		
8	F	1	Total	Mg	0	0
			1	1		

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		
9	N	2	Total	Zn	0	0
			2	2		

- Molecule 10 is CYTIDINE-5'-MONOPHOSPHATE (three-letter code: C) (formula: C<sub>9</sub>H<sub>14</sub>N<sub>3</sub>O<sub>8</sub>P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
10	D	1	Total	C	N	O	P	0	0
			20	9	3	7	1		
10	N	1	Total	C	N	O	P	0	0
			20	9	3	7	1		

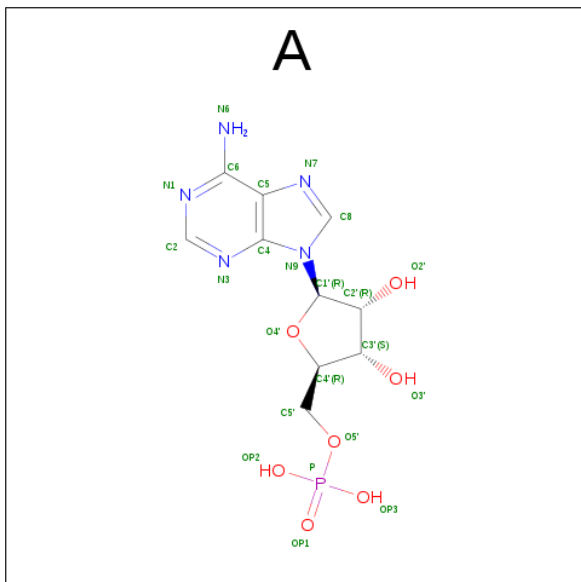
- Molecule 11 is CYTIDINE-5'-TRIPHOSPHATE (three-letter code: CTP) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>3</sub>O<sub>14</sub>P<sub>3</sub>).



- Molecule 12 is NICOTINAMIDE-ADENINE-DINUCLEOTIDE (three-letter code: NAD) (formula:  $\text{C}_{21}\text{H}_{27}\text{N}_7\text{O}_{14}\text{P}_2$ ).



- Molecule 13 is ADENOSINE-5'-MONOPHOSPHATE (three-letter code: A) (formula:  $C_{10}H_{14}N_5O_7P$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
13	R	1	Total	C	N	O	P	0	0
			23	10	5	7	1		

- Molecule 14 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	A	23	Total	O	0	0
			23	23		
14	B	16	Total	O	0	0
			16	16		
14	C	196	Total	O	0	0
			196	196		
14	D	244	Total	O	0	0
			244	244		
14	E	16	Total	O	0	0
			16	16		
14	F	45	Total	O	0	0
			45	45		
14	G	8	Total	O	0	0
			8	8		
14	H	4	Total	O	0	0
			4	4		
14	K	14	Total	O	0	0
			14	14		

*Continued on next page...*

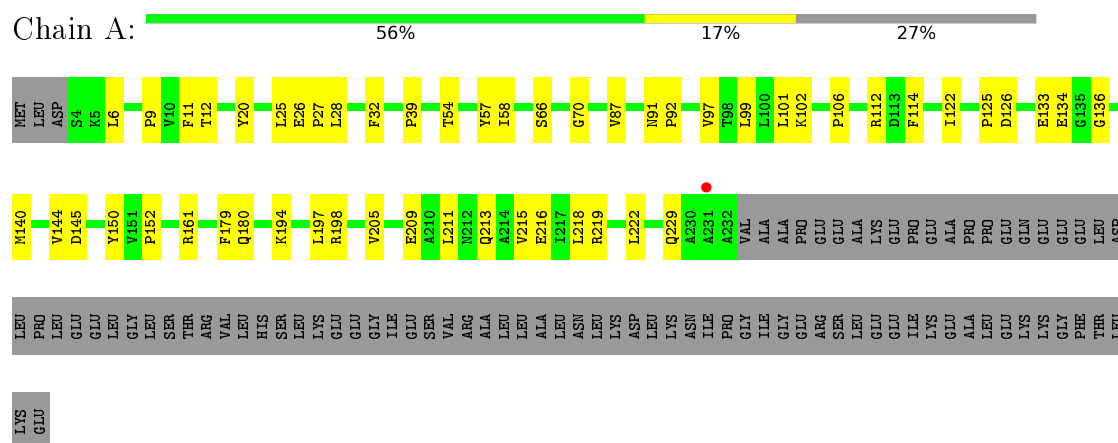
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	L	13	Total 13	O 13	0	0
14	M	97	Total 97	O 97	0	0
14	N	187	Total 187	O 187	0	0
14	O	12	Total 12	O 12	0	0
14	P	14	Total 14	O 14	0	0
14	R	3	Total 3	O 3	0	0
14	S	3	Total 3	O 3	0	0

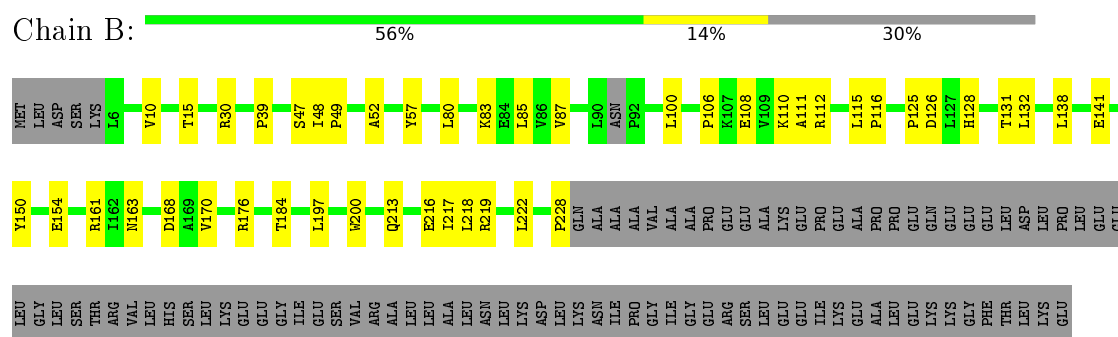
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). 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.

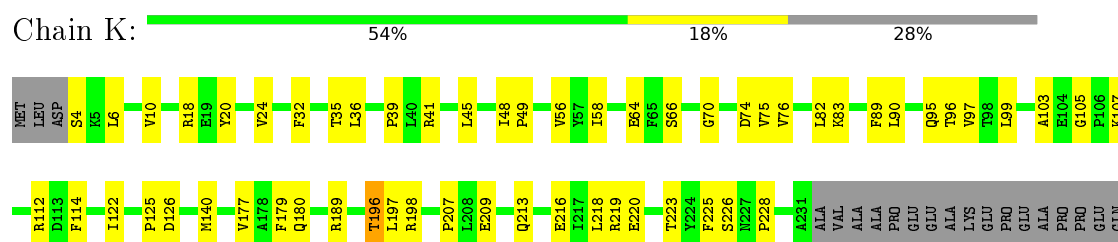
#### • Molecule 1: DNA-directed RNA polymerase subunit alpha



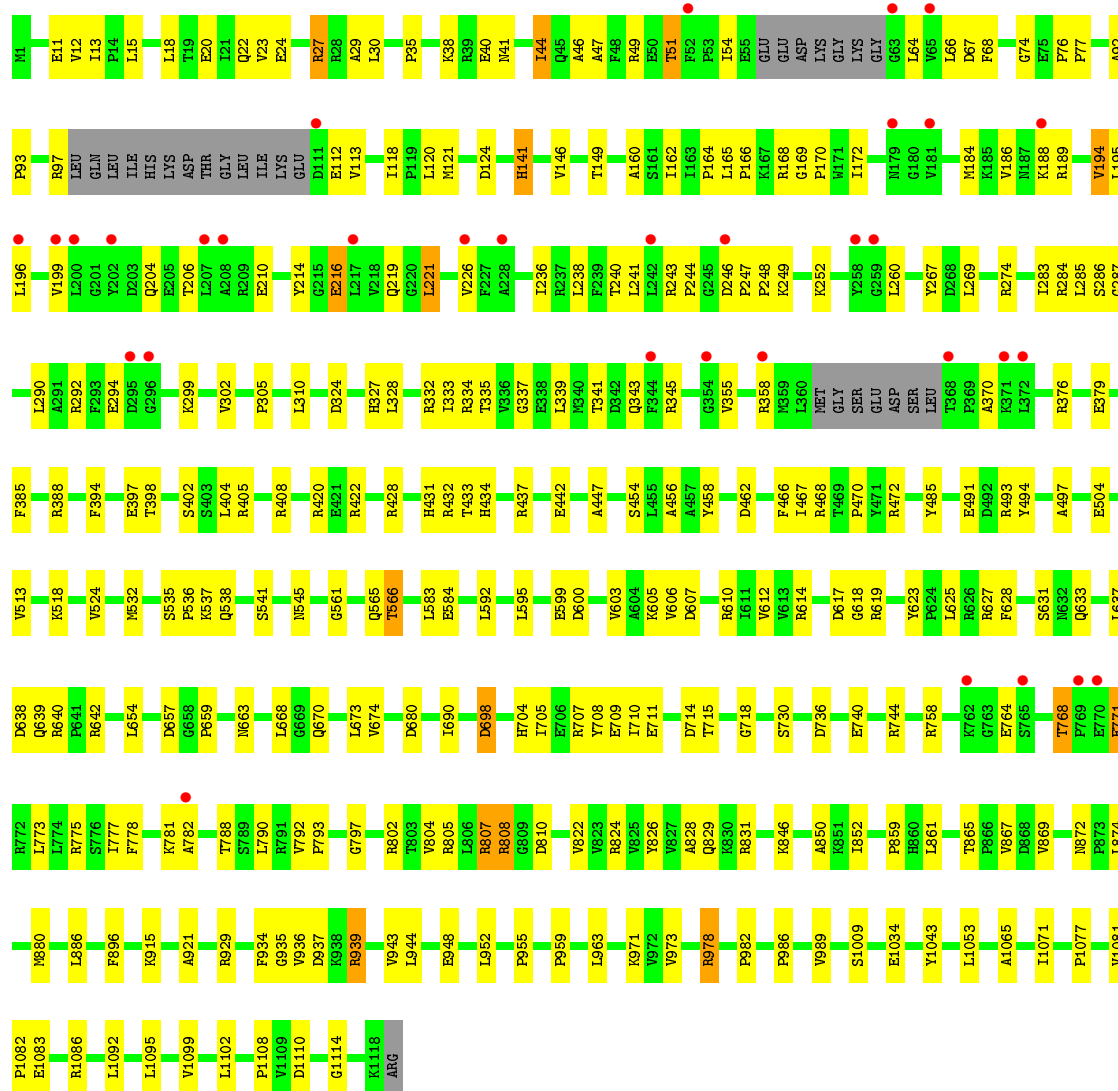
#### • Molecule 1: DNA-directed RNA polymerase subunit alpha



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

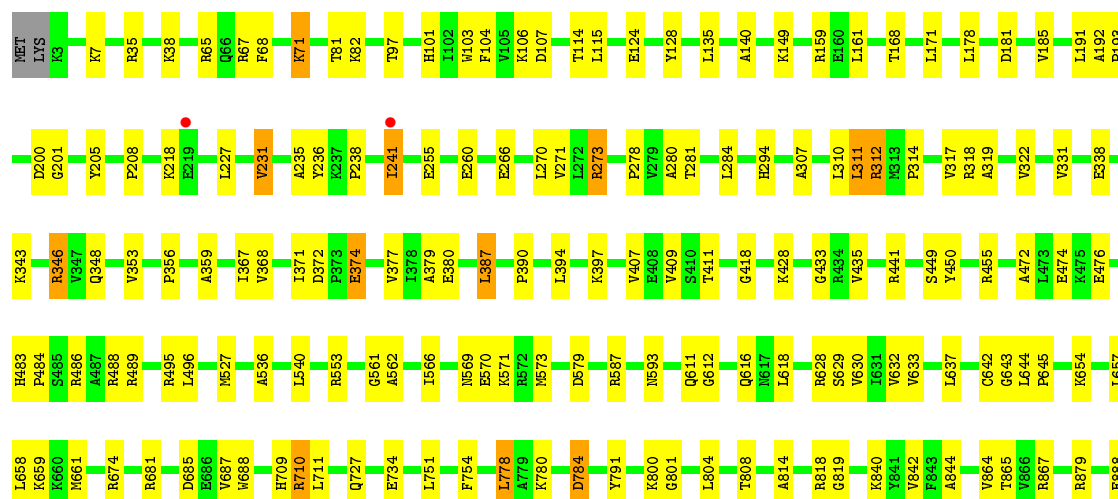


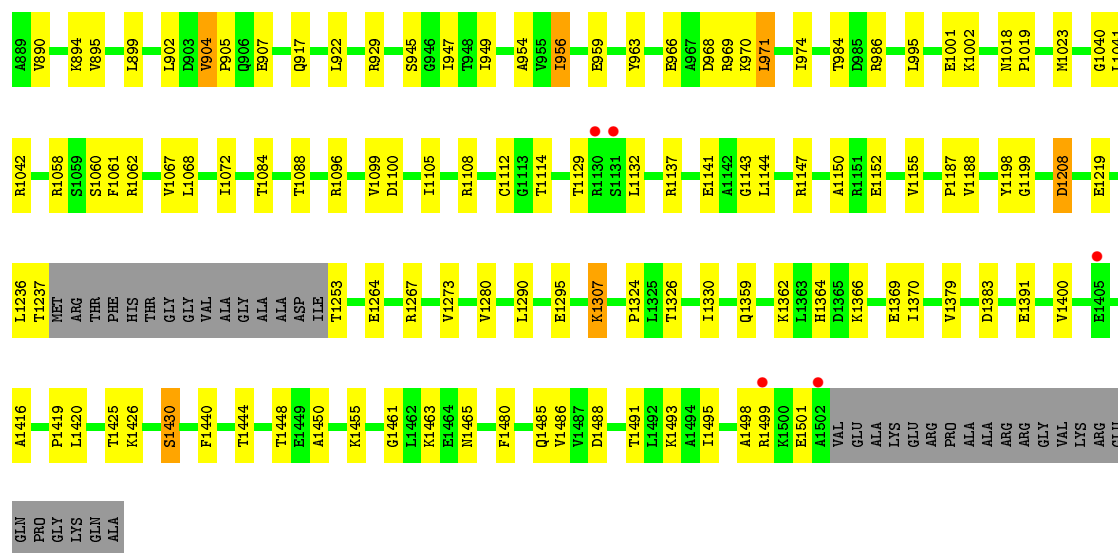




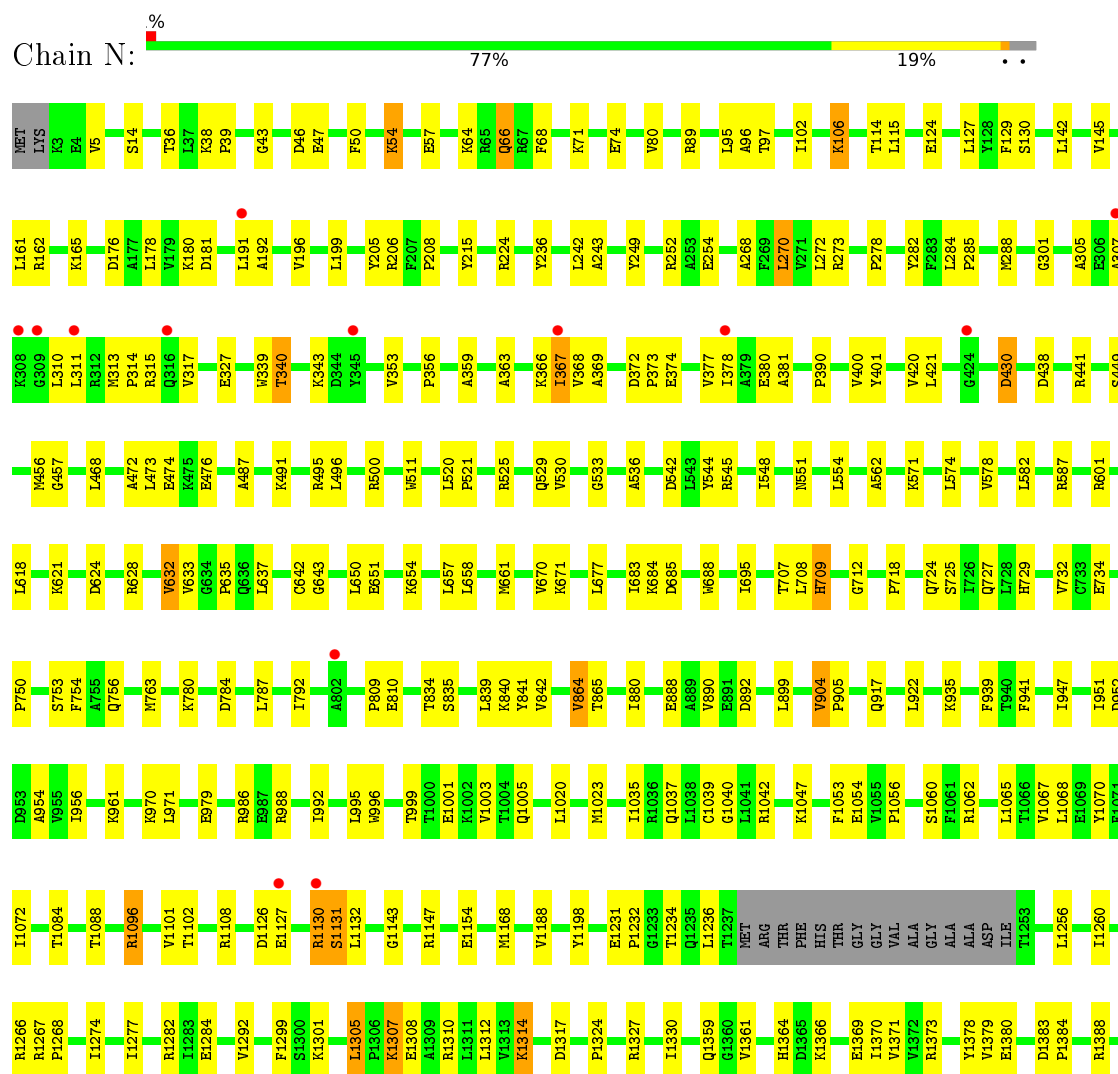
● Molecule 3: DNA-directed RNA polymerase subunit beta'

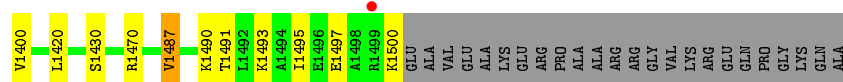
Chain D: 79% 17%



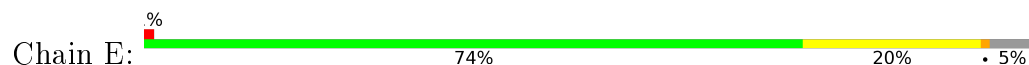


• Molecule 3: DNA-directed RNA polymerase subunit beta'

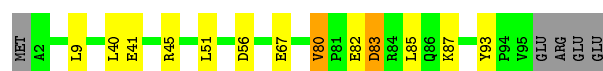
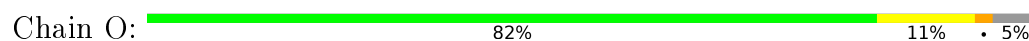




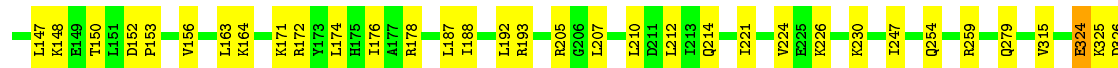
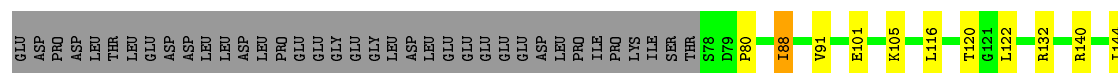
- Molecule 4: DNA-directed RNA polymerase subunit omega



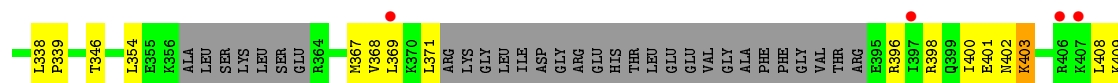
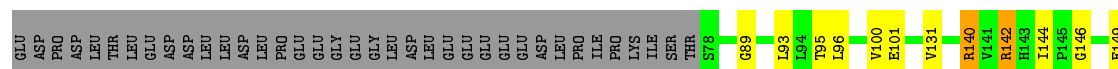
- Molecule 4: DNA-directed RNA polymerase subunit omega

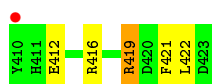


- Molecule 5: RNA polymerase sigma factor SigA

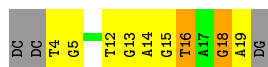
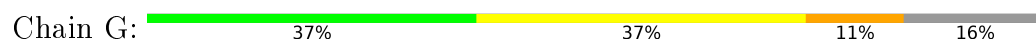


- Molecule 5: RNA polymerase sigma factor SigA





- Molecule 6: DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3')



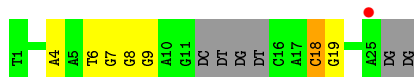
- Molecule 6: DNA (5'-D(\*CP\*C\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*AP\*GP\*AP\*G)-3')



- Molecule 7: DNA (27-MER)



- Molecule 7: DNA (27-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	185.00Å 103.51Å 296.30Å 90.00° 98.27° 90.00°	Depositor
Resolution (Å)	42.17 – 3.00 49.42 – 2.99	Depositor EDS
% Data completeness (in resolution range)	97.5 (42.17-3.00) 97.7 (49.42-2.99)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.78 (at 3.01Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.201 , 0.254 0.201 , 0.253	Depositor DCC
$R_{free}$ test set	10877 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	48.8	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 31.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	57349	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 42.42 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.0439e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.29	0/1829	0.53	0/2487
1	B	0.28	0/1781	0.50	0/2420
1	K	0.27	0/1824	0.49	0/2480
1	L	0.28	0/1781	0.50	0/2420
2	C	0.32	0/8913	0.51	0/12053
2	M	0.29	0/8775	0.49	0/11867
3	D	0.31	0/11936	0.50	1/16138 (0.0%)
3	N	0.30	0/11922	0.49	0/16119
4	E	0.31	0/772	0.50	0/1040
4	O	0.28	0/772	0.47	0/1040
5	F	0.29	0/2852	0.46	0/3837
5	P	0.28	0/2614	0.46	0/3516
6	G	0.66	1/368 (0.3%)	1.15	2/567 (0.4%)
6	R	0.53	0/368	1.08	2/567 (0.4%)
7	H	0.57	0/489	1.14	1/752 (0.1%)
7	S	0.55	0/488	1.15	2/750 (0.3%)
All	All	0.31	1/57484 (0.0%)	0.53	8/78053 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	G	18	DG	O3'-P	-7.08	1.52	1.61

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	G	16	DT	O4'-C4'-C3'	-7.75	101.35	106.00
7	S	18	DC	O4'-C1'-N1	6.36	112.45	108.00
6	R	16	DT	O4'-C4'-C3'	-5.75	102.20	104.50
6	G	14	DA	O4'-C4'-C3'	-5.62	102.25	104.50
6	R	14	DA	O4'-C4'-C3'	-5.49	102.30	104.50

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	1797	0	1849	33	0
1	B	1750	0	1802	29	0
1	K	1792	0	1844	39	0
1	L	1750	0	1802	41	0
2	C	8747	0	8858	121	0
2	M	8611	0	8710	169	0
3	D	11730	0	11960	158	0
3	N	11716	0	11949	174	1
4	E	758	0	770	14	0
4	O	758	0	770	7	0
5	F	2807	0	2882	41	0
5	P	2574	0	2643	45	1
6	G	328	0	182	7	0
6	R	328	0	182	3	0
7	H	435	0	238	2	0
7	S	434	0	235	10	0
8	B	1	0	0	0	0
8	D	3	0	0	0	0
8	F	1	0	0	0	0
8	L	1	0	0	0	0
8	N	3	0	0	0	0
8	P	1	0	0	0	0
9	D	2	0	0	0	0
9	N	2	0	0	0	0
10	D	20	0	11	0	0
10	N	20	0	11	2	0
11	D	9	0	0	0	0
11	M	9	0	0	0	0
12	D	44	0	24	1	0
13	R	23	0	11	0	0
14	A	23	0	0	0	0
14	B	16	0	0	0	0
14	C	196	0	0	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
14	D	244	0	0	7	0
14	E	16	0	0	0	0
14	F	45	0	0	1	0
14	G	8	0	0	0	0
14	H	4	0	0	0	0
14	K	14	0	0	0	0
14	L	13	0	0	1	0
14	M	97	0	0	4	0
14	N	187	0	0	2	0
14	O	12	0	0	0	0
14	P	14	0	0	1	0
14	R	3	0	0	0	0
14	S	3	0	0	0	0
All	All	57349	0	56733	807	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 7.

The worst 5 of 807 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)
2:M:428:ARG:NH2	2:M:447:ALA:O	2.07	0.88
3:N:562:ALA:O	5:P:140:ARG:NH1	2.10	0.84
3:D:124:GLU:OE2	3:D:587:ARG:NH2	2.11	0.84
2:M:674:VAL:HG12	2:M:869:VAL:HB	1.63	0.81
2:C:787:ASP:OD2	2:C:791:ARG:NH2	2.15	0.78

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)
3:N:327:GLU:OE2	5:P:263:HIS:NE2[2_545]	2.16	0.04

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries

of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	227/315 (72%)	224 (99%)	3 (1%)	0	100	100
1	B	218/315 (69%)	214 (98%)	4 (2%)	0	100	100
1	K	226/315 (72%)	221 (98%)	5 (2%)	0	100	100
1	L	218/315 (69%)	213 (98%)	5 (2%)	0	100	100
2	C	1102/1119 (98%)	1076 (98%)	26 (2%)	0	100	100
2	M	1083/1119 (97%)	1050 (97%)	33 (3%)	0	100	100
3	D	1481/1524 (97%)	1452 (98%)	29 (2%)	0	100	100
3	N	1479/1524 (97%)	1449 (98%)	28 (2%)	2 (0%)	56	90
4	E	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
4	O	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
5	F	344/443 (78%)	339 (98%)	5 (2%)	0	100	100
5	P	310/443 (70%)	304 (98%)	6 (2%)	0	100	100
All	All	6872/7630 (90%)	6722 (98%)	148 (2%)	2 (0%)	100	100

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	N	1131	SER
3	N	530	VAL

### 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	199/273 (73%)	192 (96%)	7 (4%)	43	80
1	B	195/273 (71%)	189 (97%)	6 (3%)	47	83
1	K	199/273 (73%)	193 (97%)	6 (3%)	48	83

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	L	195/273 (71%)	189 (97%)	6 (3%)	47	83
2	C	933/941 (99%)	899 (96%)	34 (4%)	42	79
2	M	917/941 (97%)	876 (96%)	41 (4%)	34	74
3	D	1252/1279 (98%)	1196 (96%)	56 (4%)	34	74
3	N	1251/1279 (98%)	1195 (96%)	56 (4%)	34	74
4	E	82/88 (93%)	79 (96%)	3 (4%)	41	79
4	O	82/88 (93%)	77 (94%)	5 (6%)	23	61
5	F	301/388 (78%)	293 (97%)	8 (3%)	52	85
5	P	277/388 (71%)	258 (93%)	19 (7%)	19	56
All	All	5883/6484 (91%)	5636 (96%)	247 (4%)	36	76

5 of 247 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	416	ARG
2	M	204	GLN
5	P	95	THR
1	K	10	VAL
1	L	146	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	83	GLN
1	K	213	GLN
5	P	269	ASN
1	K	63	HIS
3	D	66	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 20 ligands modelled in this entry, 14 are monoatomic - leaving 6 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	C	D	2005	8,12	14,21,22	0.72	0	18,30,33	0.55	0
11	CTP	D	2006	8	6,8,30	1.73	1 (16%)	6,13,47	0.71	0
12	NAD	D	2008	10	42,48,48	2.05	11 (26%)	46,73,73	2.32	9 (19%)
11	CTP	M	1201	8	6,8,30	1.19	1 (16%)	6,13,47	0.52	0
10	C	N	2005	8	14,21,22	0.74	0	18,30,33	0.53	0
13	A	R	101	-	22,25,25	2.08	7 (31%)	22,38,38	2.50	2 (9%)

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
10	C	D	2005	8,12	-	0/3/25/26	0/2/2/2
11	CTP	D	2006	8	-	0/6/6/38	0/0/0/2
12	NAD	D	2008	10	-	0/22/62/62	0/5/5/5
11	CTP	M	1201	8	-	0/6/6/38	0/0/0/2
10	C	N	2005	8	-	0/3/25/26	0/2/2/2
13	A	R	101	-	-	0/6/26/26	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	D	2008	NAD	O2D-C2D	-2.84	1.36	1.43
12	D	2008	NAD	O2B-C2B	-2.50	1.37	1.43
13	R	101	A	O2'-C2'	-2.44	1.37	1.43

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	D	2008	NAD	C2B-C3B	-2.18	1.47	1.53
12	D	2008	NAD	O3D-C3D	-2.17	1.37	1.43

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
13	R	101	A	N3-C2-N1	-9.49	121.41	128.87
12	D	2008	NAD	N3A-C2A-N1A	-9.42	121.47	128.87
12	D	2008	NAD	C4D-O4D-C1D	-3.21	106.24	109.64
12	D	2008	NAD	O7N-C7N-N7N	-2.50	119.02	122.58
12	D	2008	NAD	C4B-O4B-C1B	-2.20	107.32	109.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
12	D	2008	NAD	1	0
10	N	2005	C	2	0

## 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	229/315 (72%)	-0.51	1 (0%) 93 80	7, 21, 48, 79	1 (0%)
1	B	222/315 (70%)	-0.44	0 100 100	6, 31, 64, 81	0
1	K	228/315 (72%)	-0.38	0 100 100	10, 31, 57, 71	1 (0%)
1	L	222/315 (70%)	-0.40	0 100 100	12, 36, 73, 103	0
2	C	1108/1119 (99%)	-0.46	7 (0%) 90 73	0, 14, 62, 89	3 (0%)
2	M	1091/1119 (97%)	-0.12	33 (3%) 54 25	1, 39, 97, 116	2 (0%)
3	D	1485/1524 (97%)	-0.36	7 (0%) 91 76	0, 18, 68, 103	4 (0%)
3	N	1483/1524 (97%)	-0.30	14 (0%) 85 64	0, 24, 77, 106	4 (0%)
4	E	94/99 (94%)	-0.56	1 (1%) 82 58	1, 13, 49, 67	0
4	O	94/99 (94%)	-0.42	0 100 100	4, 23, 62, 71	0
5	F	346/443 (78%)	-0.43	1 (0%) 94 84	3, 26, 68, 84	0
5	P	316/443 (71%)	-0.20	5 (1%) 74 47	17, 46, 91, 109	0
6	G	16/19 (84%)	-0.20	0 100 100	16, 48, 112, 117	0
6	R	16/19 (84%)	-0.12	0 100 100	39, 65, 115, 122	0
7	H	21/27 (77%)	-0.20	1 (4%) 34 14	19, 56, 108, 124	0
7	S	21/27 (77%)	0.03	1 (4%) 34 14	33, 84, 121, 144	0
All	All	6992/7722 (90%)	-0.33	71 (1%) 84 60	0, 26, 78, 144	15 (0%)

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	M	188	LYS	5.1
2	M	200	LEU	4.5
2	M	196	LEU	4.5
2	M	63	GLY	3.8
2	M	368	THR	3.8

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

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

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
13	A	R	101	23/23	0.83	0.25	2.88	37,54,63,83	23
10	C	N	2005	20/21	0.88	0.23	1.69	22,46,53,59	20
12	NAD	D	2008	44/44	0.89	0.24	1.64	15,34,53,67	44
11	CTP	M	1201	9/29	0.89	0.20	1.43	25,52,69,73	0
11	CTP	D	2006	9/29	0.91	0.21	1.30	21,41,62,80	0
10	C	D	2005	20/21	0.94	0.18	0.17	6,19,38,47	20
8	MG	P	2001	1/1	0.92	0.24	-0.22	51,51,51,51	0
9	ZN	D	2001	1/1	1.00	0.14	-0.36	5,5,5,5	0
8	MG	D	2004	1/1	0.99	0.17	-0.36	15,15,15,15	0
9	ZN	N	2001	1/1	1.00	0.15	-1.11	7,7,7,7	0
9	ZN	N	2002	1/1	0.96	0.07	-1.71	89,89,89,89	0
9	ZN	D	2002	1/1	0.99	0.05	-2.28	40,40,40,40	0
8	MG	L	2001	1/1	0.93	0.10	-3.06	41,41,41,41	0
8	MG	N	2004	1/1	0.98	0.06	-3.16	27,27,27,27	0
8	MG	F	2001	1/1	0.93	0.05	-3.17	25,25,25,25	0
8	MG	D	2007	1/1	0.94	0.33	-	28,28,28,28	0
8	MG	N	2006	1/1	0.94	0.28	-	32,32,32,32	0
8	MG	N	2003	1/1	0.93	0.26	-	6,6,6,6	0
8	MG	D	2003	1/1	0.88	0.23	-	1,1,1,1	0
8	MG	B	2001	1/1	0.92	0.20	-	17,17,17,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.