



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:20 PM GMT

PDB ID : 4OIO
Title : Crystal structure of Thermus thermophilus pre-insertion substrate complex for de novo transcription initiation
Authors : Zhang, Y.; Ebright, R.H.; Arnold, E.
Deposited on : 2014-01-20
Resolution : 3.10 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
A user guide is available at
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

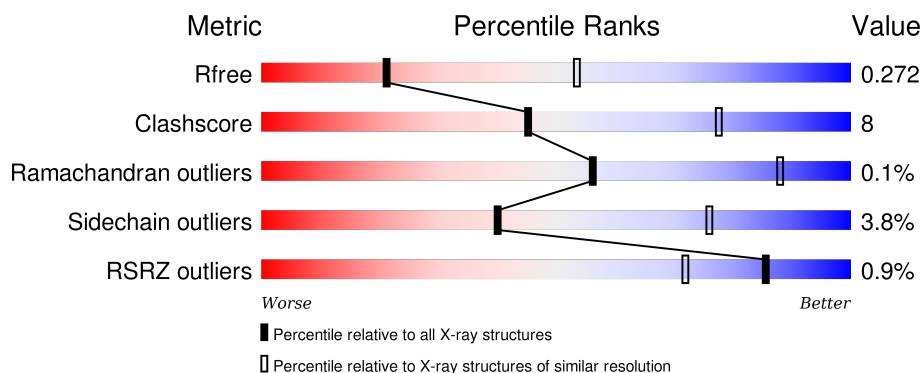
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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 ≥ 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>58%</div> <div>14%</div> <div>28%</div> </div>
1	B	315	<div> <div>57%</div> <div>12%</div> <div>30%</div> </div>
2	C	1119	<div> <div>79%</div> <div>19%</div> <div>2%</div> </div>
3	D	1524	<div> <div>74%</div> <div>21%</div> <div>5%</div> </div>
4	E	99	<div> <div>71%</div> <div>23%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	443	
6	G	21	
7	H	27	

2 Entry composition

There are 12 unique types of molecules in this entry. The entry contains 28649 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	226	Total	C	N	O	S	0	0	0
			1782	1138	310	332	2			
1	B	222	Total	C	N	O	S	0	0	0
			1750	1118	304	326	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1111	Total	C	N	O	S	0	1	0
			8779	5556	1564	1635	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1479	Total	C	N	O	S	0	0	0
			11684	7408	2059	2181	36			

- 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			

- Molecule 5 is a protein called DNA directed RNA polymerase sigma factor A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	344	Total	C	N	O	S	0	0	0
			2791	1760	507	520	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	EXPRESSION TAG	UNP Q5SKW1
F	-18	GLY	-	EXPRESSION TAG	UNP Q5SKW1
F	-17	SER	-	EXPRESSION TAG	UNP Q5SKW1
F	-16	SER	-	EXPRESSION TAG	UNP Q5SKW1
F	-15	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-14	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-13	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-12	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-11	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-10	HIS	-	EXPRESSION TAG	UNP Q5SKW1
F	-9	SER	-	EXPRESSION TAG	UNP Q5SKW1
F	-8	SER	-	EXPRESSION TAG	UNP Q5SKW1
F	-7	GLY	-	EXPRESSION TAG	UNP Q5SKW1
F	-6	LEU	-	EXPRESSION TAG	UNP Q5SKW1
F	-5	VAL	-	EXPRESSION TAG	UNP Q5SKW1
F	-4	PRO	-	EXPRESSION TAG	UNP Q5SKW1
F	-3	ARG	-	EXPRESSION TAG	UNP Q5SKW1
F	-2	GLY	-	EXPRESSION TAG	UNP Q5SKW1
F	-1	SER	-	EXPRESSION TAG	UNP Q5SKW1
F	0	HIS	-	EXPRESSION TAG	UNP Q5SKW1

- Molecule 6 is a DNA chain called 5'-D(*CP*CP*TP*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*AP*GP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	16	Total	C	N	O	P	0	0	0
			331	156	63	96	16			

- Molecule 7 is a DNA chain called 5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*AP*GP*CP*TP*GP*TP*CP*AP*CP*GP*GP*AP*TP*GP*CP*AP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	24	Total	C	N	O	P	0	0	0
			495	236	94	142	23			

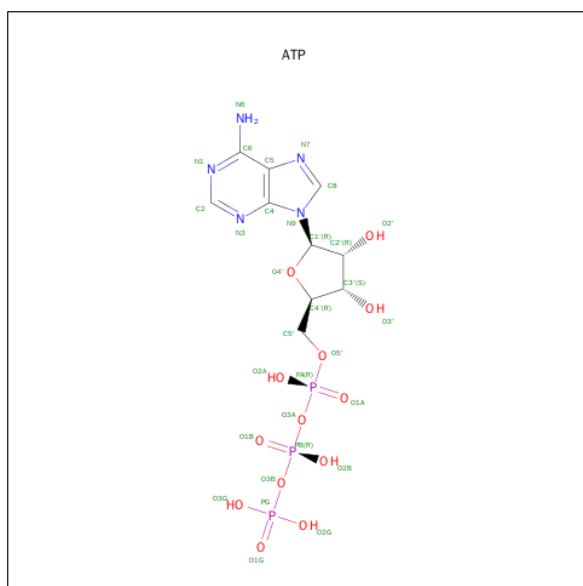
- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

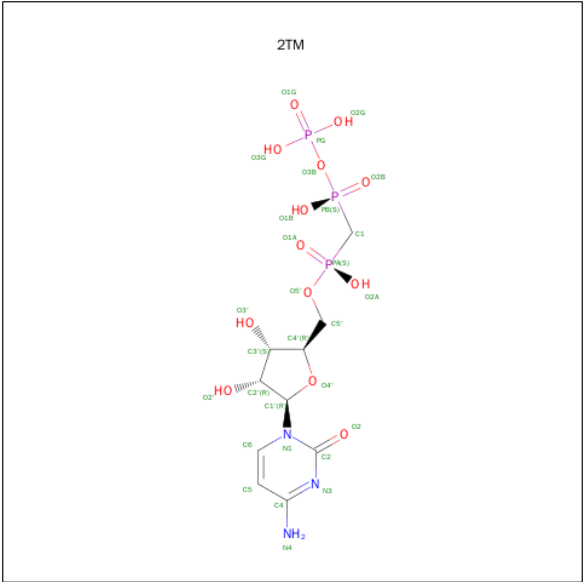
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	D	2	Total	Zn	0	0
			2	2		

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Mg	0	0
			2	2		

- Molecule 10 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$).

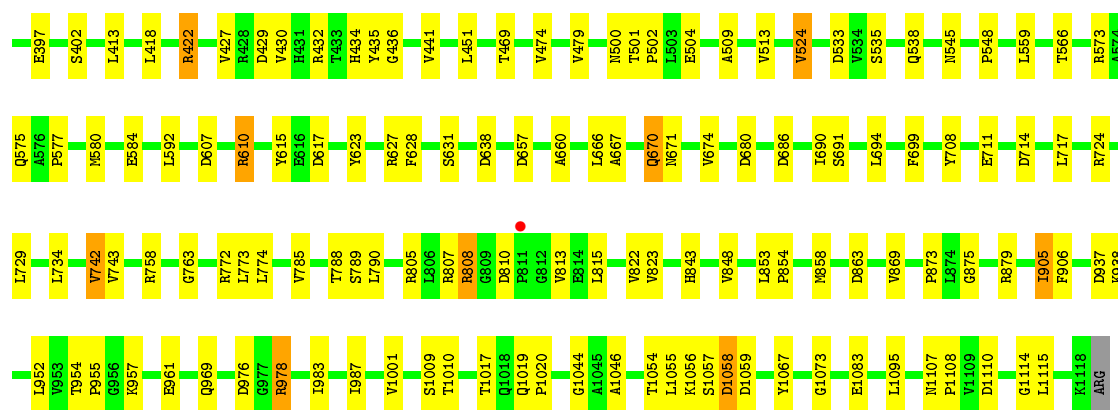




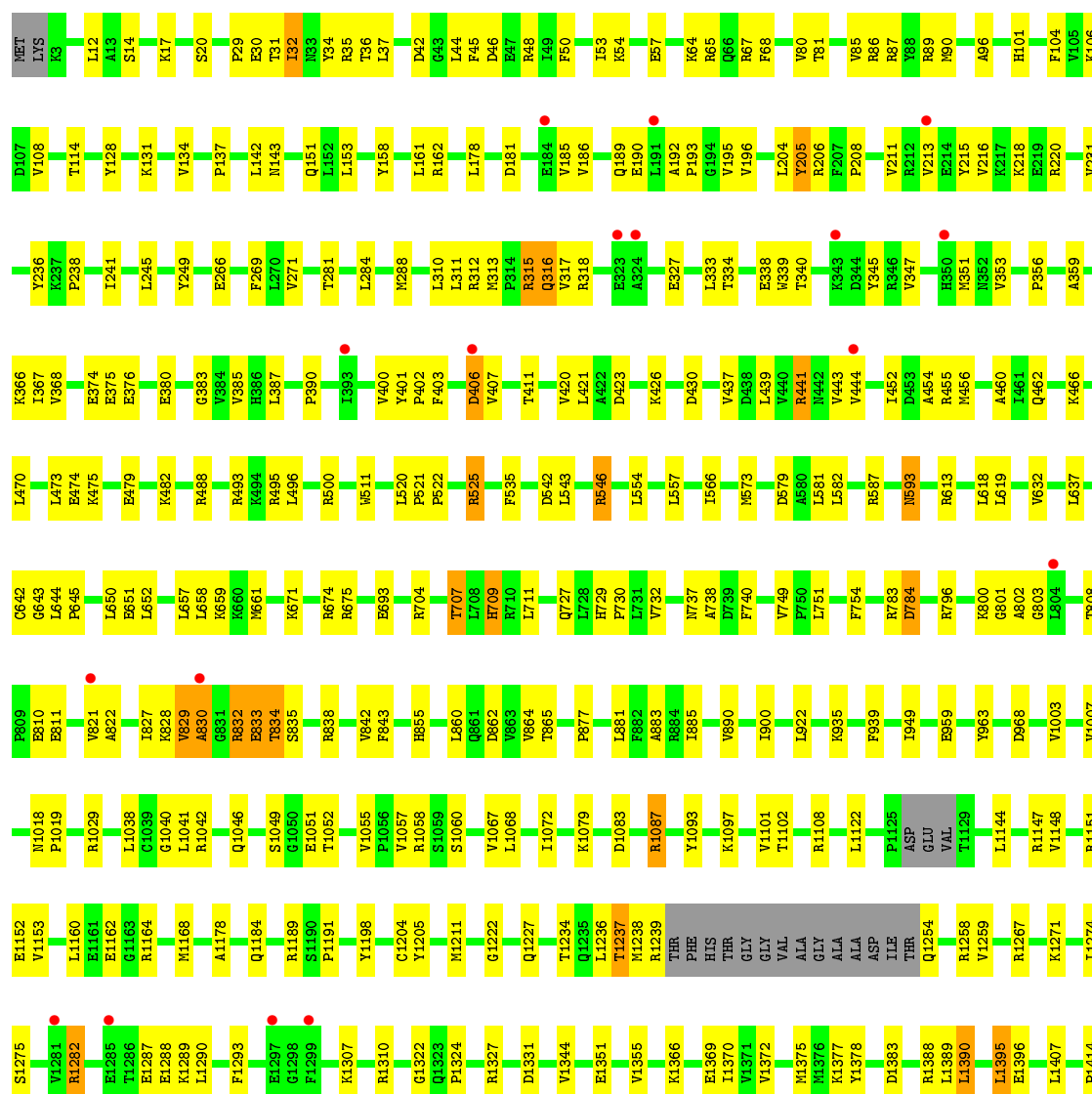
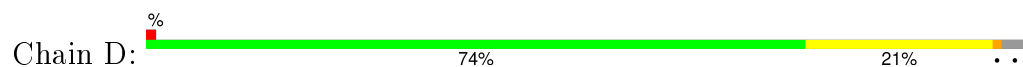
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
11	D	1	Total	C	N	O	P	0	0
			29	10	3	13	3		

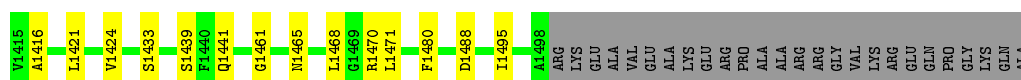
- Molecule 12 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
12	A	16	Total	O	0	0
			16	16		
12	B	7	Total	O	0	0
			7	7		
12	C	68	Total	O	0	0
			68	68		
12	D	89	Total	O	0	0
			89	89		
12	E	13	Total	O	0	0
			13	13		
12	F	15	Total	O	0	0
			15	15		
12	G	5	Total	O	0	0
			5	5		
12	H	2	Total	O	0	0
			2	2		



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

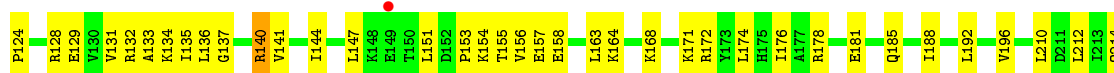
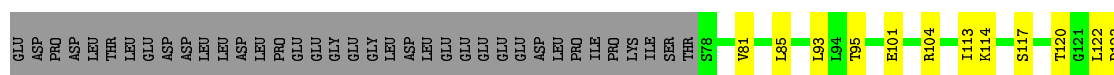
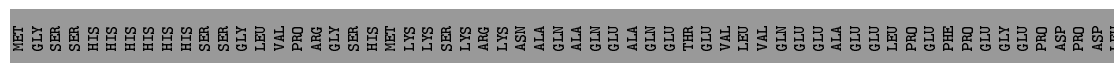




- Molecule 4: DNA-directed RNA polymerase subunit omega



- Molecule 5: DNA directed RNA polymerase sigma factor A



- Molecule 6: 5'-D(*CP*CP*TP*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*A P*GP*GP*G)-3'



- Molecule 7: 5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*AP*GP*CP*TP*GP*TP*CP*AP*C P*GP*GP*AP*TP*GP*CP*AP*GP*G)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	184.11Å 103.78Å 296.09Å 90.00° 98.86° 90.00°	Depositor
Resolution (Å)	48.92 – 3.10 48.92 – 3.10	Depositor EDS
% Data completeness (in resolution range)	96.1 (48.92-3.10) 94.8 (48.92-3.10)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	0.15	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.48 (at 3.12Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.205 , 0.256 0.224 , 0.272	Depositor DCC
R_{free} test set	2110 reflections (2.27%)	DCC
Wilson B-factor (Å ²)	59.2	Xtriage
Anisotropy	0.022	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 21.7	EDS
Estimated twinning fraction	0.022 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-l$ 0.019 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-l$	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 95184 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	28649	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 2TM, MG, ZN, ATP

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	0/1814	0.43	0/2466
1	B	0.24	0/1782	0.44	0/2424
2	C	0.25	0/8950	0.46	0/12105
3	D	0.24	0/11889	0.45	0/16072
4	E	0.24	0/772	0.42	0/1040
5	F	0.24	0/2836	0.42	0/3815
6	G	0.45	0/371	1.05	1/571 (0.2%)
7	H	0.48	0/556	1.13	4/858 (0.5%)
All	All	0.25	0/28970	0.49	5/39351 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	H	17	DA	O4'-C4'-C3'	-7.00	101.70	104.50
6	G	16	DC	O4'-C4'-C3'	-6.27	101.99	104.50
7	H	17	DA	O4'-C1'-N9	6.11	112.28	108.00
7	H	17	DA	C1'-O4'-C4'	-6.05	104.05	110.10
7	H	23	DG	O4'-C1'-N9	5.17	111.62	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	1782	0	1834	23	0
1	B	1750	0	1797	27	0
2	C	8779	0	8883	143	0
3	D	11684	0	11918	212	1
4	E	758	0	770	18	0
5	F	2791	0	2867	56	0
6	G	331	0	180	3	0
7	H	495	0	272	6	0
8	D	2	0	0	0	0
9	D	2	0	0	0	0
10	D	31	0	11	0	0
11	D	29	0	14	1	0
12	A	16	0	0	0	0
12	B	7	0	0	0	0
12	C	68	0	0	4	0
12	D	89	0	0	2	0
12	E	13	0	0	0	0
12	F	15	0	0	0	0
12	G	5	0	0	0	0
12	H	2	0	0	0	0
All	All	28649	0	28546	440	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 8.

The worst 5 of 440 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:58:ILE:HB	1:A:61:VAL:HB	1.69	0.74
3:D:245:LEU:HD21	3:D:249:TYR:HB3	1.68	0.73
1:B:206:THR:HG22	1:B:209:GLU:H	1.53	0.72
3:D:829:VAL:HG12	3:D:830:ALA:H	1.57	0.69
3:D:208:PRO:HA	3:D:390:PRO:HA	1.73	0.69

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:D:34:TYR:OH	3:D:327:GLU:OE2[4_555]	2.15	0.05

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	224/315 (71%)	219 (98%)	5 (2%)	0	100	100
1	B	220/315 (70%)	216 (98%)	4 (2%)	0	100	100
2	C	1108/1119 (99%)	1083 (98%)	24 (2%)	1 (0%)	56	88
3	D	1473/1524 (97%)	1445 (98%)	25 (2%)	3 (0%)	52	84
4	E	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
5	F	342/443 (77%)	335 (98%)	7 (2%)	0	100	100
All	All	3459/3815 (91%)	3388 (98%)	67 (2%)	4 (0%)	56	88

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	834	THR
3	D	830	ALA
2	C	905	ILE
3	D	829	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%)	193 (97%)	6 (3%)	48	81
1	B	195/273 (71%)	189 (97%)	6 (3%)	47	80
2	C	937/941 (100%)	893 (95%)	44 (5%)	32	70

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	D	1247/1279 (98%)	1202 (96%)	45 (4%)	42	77
4	E	82/88 (93%)	80 (98%)	2 (2%)	57	84
5	F	299/388 (77%)	289 (97%)	10 (3%)	45	79
All	All	2959/3242 (91%)	2846 (96%)	113 (4%)	40	76

5 of 113 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1001	VAL
3	D	311	LEU
5	F	95	THR
2	C	1017	THR
3	D	35	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	316	GLN
3	D	348	GLN
3	D	1195	GLN
3	D	66	GLN
3	D	189	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 ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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	ATP	D	2005	9	24,33,33	1.47	3 (12%)	31,52,52	2.47	8 (25%)
11	2TM	D	2006	9	22,30,30	1.57	5 (22%)	30,47,47	1.55	4 (13%)

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	ATP	D	2005	9	-	0/18/38/38	0/3/3/3
11	2TM	D	2006	9	-	0/15/38/38	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	D	2006	2TM	C3'-C2'	-2.59	1.46	1.53
11	D	2006	2TM	PA-O2A	-2.45	1.50	1.56
11	D	2006	2TM	C3'-C4'	-2.36	1.46	1.53
10	D	2005	ATP	C2-N1	2.32	1.38	1.33
11	D	2006	2TM	PA-O1A	2.72	1.58	1.51

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	D	2005	ATP	N3-C2-N1	-10.29	121.01	128.89
10	D	2005	ATP	PA-O3A-PB	-4.32	120.58	132.73
10	D	2005	ATP	C2'-C1'-N9	-4.05	108.11	114.29
11	D	2006	2TM	C4'-O4'-C1'	-3.75	105.60	109.72
10	D	2005	ATP	PB-O3B-PG	-3.68	120.33	132.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
11	D	2006	2TM	1	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, 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	226/315 (71%)	-0.52	0 100 100	24, 42, 69, 91	0
1	B	222/315 (70%)	-0.30	1 (0%) 91 83	25, 56, 90, 106	0
2	C	1111/1119 (99%)	-0.35	11 (0%) 84 69	4, 35, 107, 145	0
3	D	1479/1524 (97%)	-0.23	17 (1%) 82 66	5, 40, 110, 149	1 (0%)
4	E	94/99 (94%)	-0.58	0 100 100	13, 35, 80, 89	0
5	F	344/443 (77%)	-0.36	1 (0%) 94 88	18, 57, 97, 125	0
6	G	16/21 (76%)	-0.71	0 100 100	34, 56, 146, 151	0
7	H	24/27 (88%)	-0.49	0 100 100	56, 90, 138, 164	0
All	All	3516/3863 (91%)	-0.32	30 (0%) 85 72	4, 42, 106, 164	1 (0%)

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	191	LEU	3.2
3	D	1281	VAL	2.9
3	D	323	GLU	2.8
2	C	221	LEU	2.8
2	C	365	ASP	2.7

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

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

6.3 Carbohydrates ⓘ

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, 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	LLDF	B-factors(Å ²)	Q<0.9
11	2TM	D	2006	29/29	0.89	0.17	-0.12	31,48,75,85	0
10	ATP	D	2005	31/31	0.93	0.17	-0.24	26,37,52,60	31
8	ZN	D	2001	1/1	0.99	0.15	-0.66	20,20,20,20	0
8	ZN	D	2002	1/1	0.99	0.08	-1.13	57,57,57,57	0
9	MG	D	2004	1/1	0.95	0.21	-	31,31,31,31	0
9	MG	D	2003	1/1	0.99	0.28	-	12,12,12,12	0

6.5 Other polymers [i](#)

There are no such residues in this entry.