



# Full wwPDB X-ray Structure Validation Report

Jun 24, 2014 – 05:03 PM EDT

PDB ID : 4OGR  
Title : crystal structure of P-TEFb complex with AFF4 and Tat  
Authors : Schulze-Gahmen, U.; Alber, T.  
Deposited on : 2014-01-16  
Resolution : 3.00 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

---

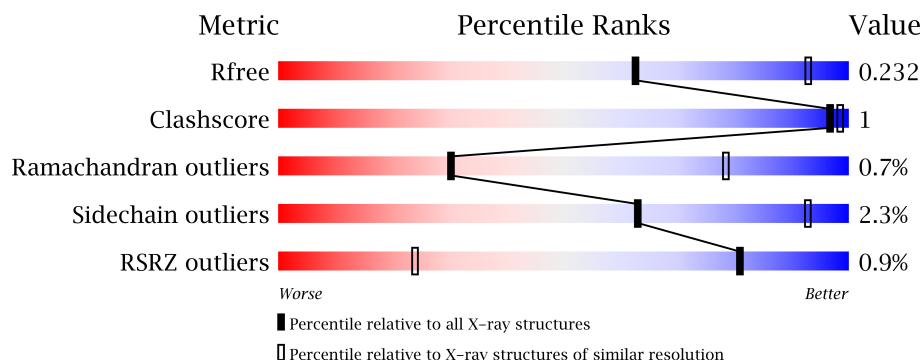
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable23161  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23161

# 1 Overall quality at a glance

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))
$R_{free}$	66092	1216 (3.00-3.00)
Clashscore	79885	1594 (3.00-3.00)
Ramachandran outliers	78287	1537 (3.00-3.00)
Sidechain outliers	78261	1540 (3.00-3.00)
RSRZ outliers	66119	1217 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	332	
1	E	332	
1	I	332	
2	B	264	
2	F	264	
2	K	264	
3	C	75	
3	G	75	
3	L	75	
4	D	58	
4	H	58	
4	M	58	

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 32026 atoms, of which 15872 are hydrogens and 0 are deuterium.

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 Cyclin-dependent kinase 9.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
1	A	315	Total	C	H	N	O	P	S		0	0	0
			5137	1628	2596	438	459	1	15				
1	E	311	Total	C	H	N	O	P	S		0	0	0
			5052	1604	2550	429	453	1	15				
1	I	319	Total	C	H	N	O	P	S		0	0	0
			5141	1636	2588	439	462	1	15				

- Molecule 2 is a protein called Cyclin-T1.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
2	B	255	Total	C	H	N	O	S			0	0	0
			4081	1320	2020	351	380	10					
2	F	255	Total	C	H	N	O	S			0	0	0
			4067	1312	2014	352	379	10					
2	K	255	Total	C	H	N	O	S			0	0	0
			4065	1315	2007	353	380	10					

- Molecule 3 is a protein called AF4/FMR2 family member 4.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
3	C	55	Total	C	H	N	O	S			0	0	0
			829	265	391	83	87	3					
3	G	36	Total	C	H	N	O	S			0	0	0
			488	167	220	42	57	2					
3	L	54	Total	C	H	N	O	S			0	0	0
			785	254	363	79	86	3					

- Molecule 4 is a protein called Protein Tat.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
4	D	50	Total	C	H	N	O	S			0	0	0
			766	246	378	68	66	8					

*Continued on next page...*

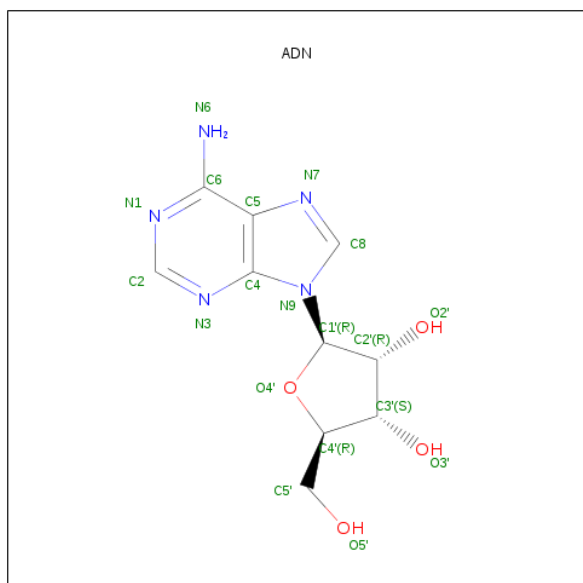
Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	H	50	Total	C	H	N	O	S	0	0	0
			749	243	367	65	66	8			
4	M	50	Total	C	H	N	O	S	0	0	0
			766	246	378	68	66	8			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	ACE	-	EXPRESSION TAG	UNP P69698
H	0	ACE	-	EXPRESSION TAG	UNP P69698
M	0	ACE	-	EXPRESSION TAG	UNP P69698

- Molecule 5 is ADENOSINE (three-letter code: ADN) (formula:  $C_{10}H_{13}N_5O_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			19	10	5	4		
5	E	1	Total	C	N	O	0	0
			19	10	5	4		
5	I	1	Total	C	N	O	0	0
			19	10	5	4		

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	H	2	Total 2	Zn 2	0	0
6	D	2	Total 2	Zn 2	0	0
6	M	2	Total 2	Zn 2	0	0

- Molecule 7 is water.

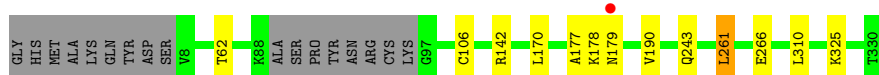
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	9	Total 9	O 9	0	0
7	B	5	Total 5	O 5	0	0
7	C	1	Total 1	O 1	0	0
7	E	2	Total 2	O 2	0	0
7	F	6	Total 6	O 6	0	0
7	I	4	Total 4	O 4	0	0
7	K	6	Total 6	O 6	0	0
7	L	1	Total 1	O 1	0	0
7	D	1	Total 1	O 1	0	0
7	M	2	Total 2	O 2	0	0

### 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.

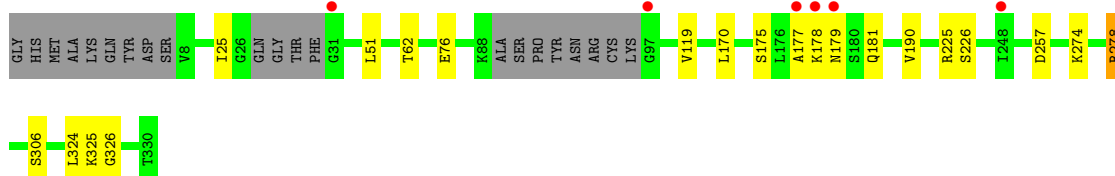
- Molecule 1: Cyclin-dependent kinase 9

Chain A: 



- Molecule 1: Cyclin-dependent kinase 9

Chain E: 



- Molecule 1: Cyclin-dependent kinase 9

Chain I: 



- Molecule 2: Cyclin-T1

Chain B: 



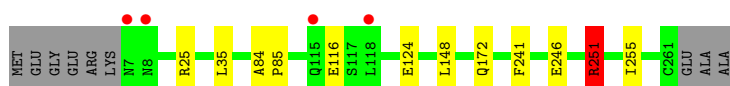
- Molecule 2: Cyclin-T1

Chain F: 



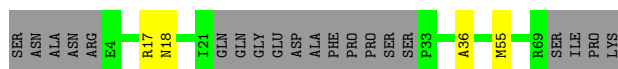
- Molecule 2: Cyclin-T1

Chain K: 



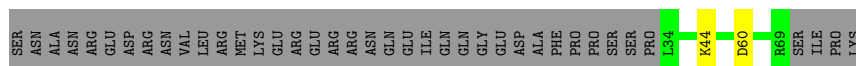
- Molecule 3: AF4/FMR2 family member 4

Chain C:



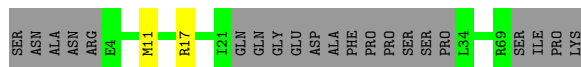
- Molecule 3: AF4/FMR2 family member 4

Chain G:



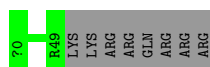
- Molecule 3: AF4/FMR2 family member 4

Chain L:



- Molecule 4: Protein Tat

Chain D:



- Molecule 4: Protein Tat

Chain H:



- Molecule 4: Protein Tat

Chain M:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	184.91Å 184.91Å 360.40Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.02 – 3.00 49.01 – 3.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.02-3.00) 100.0 (49.01-3.00)	Depositor EDS
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.41 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1419)	Depositor
R, $R_{free}$	0.205 , 0.232 0.205 , 0.232	Depositor DCC
$R_{free}$ test set	2404 reflections (3.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	77.9	Xtriage
Anisotropy	0.339	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 56.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 73322 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	32026	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.43% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.



## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: TPO, ADN, ZN, ACE

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	0/2580	0.43	0/3482
1	E	0.23	0/2539	0.43	0/3427
1	I	0.23	0/2593	0.42	0/3504
2	B	0.23	0/2115	0.42	0/2888
2	F	0.23	0/2105	0.43	0/2873
2	K	0.22	0/2112	0.40	0/2885
3	C	0.23	0/442	0.42	0/589
3	G	0.23	0/272	0.36	0/367
3	L	0.23	0/425	0.42	0/567
4	D	0.24	0/398	0.39	0/538
4	H	0.24	0/392	0.40	0/531
4	M	0.24	0/398	0.39	0/538
All	All	0.23	0/16371	0.42	0/22189

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	F	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	250	ASN	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2541	2596	0	7	0
1	E	2502	2550	0	7	0
1	I	2553	2588	0	6	0
2	B	2061	2020	0	2	0
2	F	2053	2014	1	4	0
2	K	2058	2007	0	4	0
3	C	438	391	0	4	0
3	G	268	220	0	0	0
3	L	422	363	0	2	0
4	D	388	378	0	0	0
4	H	382	367	0	1	0
4	M	388	378	0	3	0
5	A	19	0	13	1	0
5	E	19	0	13	0	0
5	I	19	0	13	0	0
6	D	2	0	0	0	0
6	H	2	0	0	0	0
6	M	2	0	0	0	0
7	A	9	0	0	0	0
7	B	5	0	0	0	0
7	C	1	0	0	0	0
7	D	1	0	0	0	0
7	E	2	0	0	0	0
7	F	6	0	0	0	0
7	I	4	0	0	0	0
7	K	6	0	0	1	0
7	L	1	0	0	0	0
7	M	2	0	0	0	0
All	All	16154	15872	40	33	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

All (33) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:M:22:CYS:SG	4:M:33:HIS:CE1	2.93	0.61
1:E:119:VAL:O	1:E:225:ARG:NH2	2.36	0.57
2:F:251:ARG:NH2	4:H:26:TYR:O	2.37	0.57
1:E:324:LEU:O	1:E:326:GLY:N	2.40	0.55
1:I:114:LEU:O	1:I:225:ARG:NH1	2.39	0.54
2:K:246:GLU:O	2:K:251:ARG:NH1	2.41	0.53
1:I:119:VAL:O	1:I:225:ARG:NH2	2.42	0.51
1:I:48:LYS:NZ	1:I:167:ASP:OD1	2.43	0.51
2:B:242:LEU:HB3	3:C:55:MET:HE1	1.93	0.51
1:A:142:ARG:NH2	3:C:18:ASN:OD1	2.44	0.50
1:A:106:CYS:H	5:A:401:ADN:HN62	1.61	0.49
1:E:177:ALA:HB1	1:E:178:LYS:CG	2.45	0.47
1:I:181:GLN:H	4:M:12:LYS:HZ2	1.63	0.47
2:F:250:ASN:HB3	2:F:251:ARG:HB2	1.97	0.46
3:C:36:ALA:HA	3:L:11:MET:HG3	1.98	0.46
2:F:84:ALA:HB3	2:F:85:PRO:HD3	1.98	0.45
2:K:84:ALA:HB3	2:K:85:PRO:HD3	1.99	0.45
1:I:144:LYS:HG2	1:I:176:LEU:HD11	1.99	0.44
2:B:84:ALA:HB3	2:B:85:PRO:HD3	1.99	0.44
2:F:250:ASN:HB3	2:F:251:ARG:CB	2.46	0.44
1:A:177:ALA:HB1	1:A:178:LYS:CG	2.47	0.44
1:E:177:ALA:HB1	1:E:178:LYS:HG3	1.99	0.43
1:A:261:LEU:HD23	1:A:261:LEU:H	1.82	0.43
1:I:310:LEU:HD13	3:L:17:ARG:HD2	2.01	0.43
1:A:310:LEU:HD13	3:C:17:ARG:HD2	2.01	0.43
1:A:177:ALA:HB1	1:A:178:LYS:HG2	2.01	0.42
1:E:274:LYS:O	1:E:278:ARG:HG2	2.19	0.41
1:E:175:SER:OG	1:E:177:ALA:HB2	2.20	0.41
2:K:172:GLN:OE1	7:K:306:HOH:O	2.22	0.41
4:M:25:CYS:HB3	4:M:30:CYS:SG	2.60	0.41
1:E:178:LYS:O	1:E:179:ASN:HB2	2.22	0.40
1:A:178:LYS:O	1:A:179:ASN:HB2	2.22	0.40
2:K:255:ILE:HG22	2:K:255:ILE:O	2.21	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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

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

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	310/332 (93%)	292 (94%)	15 (5%)	3 (1%)	22	70
1	E	304/332 (92%)	288 (95%)	14 (5%)	2 (1%)	30	78
1	I	314/332 (95%)	299 (95%)	12 (4%)	3 (1%)	22	70
2	B	253/264 (96%)	244 (96%)	7 (3%)	2 (1%)	27	76
2	F	253/264 (96%)	247 (98%)	5 (2%)	1 (0%)	43	87
2	K	253/264 (96%)	247 (98%)	5 (2%)	1 (0%)	43	87
3	C	51/75 (68%)	49 (96%)	2 (4%)	0	100	100
3	G	34/75 (45%)	31 (91%)	2 (6%)	1 (3%)	7	35
3	L	50/75 (67%)	48 (96%)	2 (4%)	0	100	100
4	D	48/58 (83%)	47 (98%)	1 (2%)	0	100	100
4	H	48/58 (83%)	48 (100%)	0	0	100	100
4	M	48/58 (83%)	47 (98%)	1 (2%)	0	100	100
All	All	1966/2187 (90%)	1887 (96%)	66 (3%)	13 (1%)	30	78

All (13) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	251	ARG
1	E	325	LYS
1	A	190	VAL
2	B	260	ALA
1	E	190	VAL
3	G	44	LYS
1	I	190	VAL
1	I	325	LYS
1	A	266	GLU
2	K	251	ARG
1	A	325	LYS
2	F	259	ARG
1	I	149	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	277/292 (95%)	273 (99%)	4 (1%)	78	96
1	E	272/292 (93%)	262 (96%)	10 (4%)	45	86
1	I	275/292 (94%)	270 (98%)	5 (2%)	71	95
2	B	228/239 (95%)	220 (96%)	8 (4%)	48	87
2	F	227/239 (95%)	222 (98%)	5 (2%)	64	93
2	K	227/239 (95%)	220 (97%)	7 (3%)	52	89
3	C	43/69 (62%)	43 (100%)	0	100	100
3	G	25/69 (36%)	24 (96%)	1 (4%)	42	84
3	L	40/69 (58%)	40 (100%)	0	100	100
4	D	44/52 (85%)	44 (100%)	0	100	100
4	H	43/52 (83%)	43 (100%)	0	100	100
4	M	44/52 (85%)	43 (98%)	1 (2%)	63	93
All	All	1745/1956 (89%)	1704 (98%)	41 (2%)	63	93

All (41) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	THR
1	A	170	LEU
1	A	243	GLN
1	A	261	LEU
2	B	8	ASN
2	B	25	ARG
2	B	35	LEU
2	B	116	GLU
2	B	136	LEU
2	B	241	PHE
2	B	257	ASN
2	B	261	CYS
1	E	25	ILE
1	E	51	LEU
1	E	62	THR
1	E	76	GLU
1	E	170	LEU
1	E	181	GLN
1	E	226	SER
1	E	257	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	278	ARG
1	E	306	SER
2	F	25	ARG
2	F	35	LEU
2	F	116	GLU
2	F	241	PHE
2	F	251	ARG
3	G	60	ASP
1	I	62	THR
1	I	170	LEU
1	I	176	LEU
1	I	255	ASN
1	I	306	SER
2	K	25	ARG
2	K	35	LEU
2	K	116	GLU
2	K	124	GLU
2	K	148	LEU
2	K	241	PHE
2	K	251	ARG
4	M	24	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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
1	TPO	A	186	1	10,10,11	5.70	1 (10%)	12,14,16	1.15	1 (8%)
1	TPO	E	186	1	10,10,11	5.67	1 (10%)	12,14,16	1.04	1 (8%)
1	TPO	I	186	1	10,10,11	5.76	1 (10%)	12,14,16	1.15	1 (8%)

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
1	TPO	A	186	1	-	0/9/11/13	0/0/0/0
1	TPO	E	186	1	-	0/9/11/13	0/0/0/0
1	TPO	I	186	1	-	0/9/11/13	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	I	186	TPO	O-C	17.78	1.23	1.11
1	A	186	TPO	O-C	17.60	1.23	1.11
1	E	186	TPO	O-C	17.49	1.23	1.11

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	I	186	TPO	P-OG1-CB	-2.44	109.18	120.44
1	A	186	TPO	P-OG1-CB	-2.37	109.51	120.44
1	E	186	TPO	P-OG1-CB	-2.35	109.58	120.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 9 ligands modelled in this entry, 6 are monoatomic - leaving 3 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	$\# Z  > 2$
5	ADN	A	401	-	21,21,21	1.92	3 (14%)	31,31,31	3.10	8 (25%)
5	ADN	E	401	-	21,21,21	1.90	3 (14%)	31,31,31	3.08	8 (25%)
5	ADN	I	401	-	21,21,21	1.91	3 (14%)	31,31,31	3.11	8 (25%)

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	ADN	A	401	-	-	0/6/22/22	0/3/3/3
5	ADN	E	401	-	-	0/6/22/22	0/3/3/3
5	ADN	I	401	-	-	0/6/22/22	0/3/3/3

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	401	ADN	C8-N9	-5.66	1.28	1.36
5	E	401	ADN	C8-N9	-5.63	1.28	1.36
5	I	401	ADN	C8-N9	-5.58	1.28	1.36
5	A	401	ADN	C4-N9	-5.03	1.30	1.37
5	I	401	ADN	C4-N9	-5.03	1.30	1.37
5	E	401	ADN	C4-N9	-4.91	1.30	1.37
5	I	401	ADN	C6-N6	3.64	1.46	1.34
5	E	401	ADN	C6-N6	3.59	1.46	1.34
5	A	401	ADN	C6-N6	3.57	1.46	1.34

All (24) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	401	ADN	C8-N9-C4	9.95	115.05	106.96
5	A	401	ADN	C8-N9-C4	9.81	114.93	106.96
5	E	401	ADN	C8-N9-C4	9.69	114.83	106.96
5	A	401	ADN	N3-C2-N1	-7.81	122.02	128.89

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	I	401	ADN	N3-C2-N1	-7.66	122.15	128.89
5	E	401	ADN	N3-C2-N1	-7.45	122.34	128.89
5	E	401	ADN	C5-C4-N3	-7.09	119.06	125.98
5	I	401	ADN	C5-C4-N3	-6.92	119.24	125.98
5	A	401	ADN	C5-C4-N3	-6.88	119.27	125.98
5	E	401	ADN	N3-C4-N9	6.72	136.92	125.39
5	A	401	ADN	N3-C4-N9	6.65	136.80	125.39
5	I	401	ADN	N3-C4-N9	6.64	136.78	125.39
5	E	401	ADN	C4-C5-N7	-4.05	105.50	109.41
5	I	401	ADN	C4-C5-N7	-4.05	105.50	109.41
5	A	401	ADN	C4-C5-N7	-3.90	105.64	109.41
5	I	401	ADN	C1'-N9-C4	-2.65	122.05	126.64
5	E	401	ADN	C2-N3-C4	2.50	120.47	113.27
5	A	401	ADN	C2-N3-C4	2.47	120.39	113.27
5	I	401	ADN	C2-N3-C4	2.46	120.36	113.27
5	A	401	ADN	C1'-N9-C4	-2.44	122.41	126.64
5	E	401	ADN	C1'-N9-C4	-2.42	122.45	126.64
5	A	401	ADN	C5-C4-N9	-2.28	103.93	107.09
5	I	401	ADN	C5-C4-N9	-2.24	103.99	107.09
5	E	401	ADN	C5-C4-N9	-2.21	104.02	107.09

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	315/332 (94%)	0.05	1 (0%) 91 48	46, 66, 135, 210	0
1	E	311/332 (93%)	0.16	6 (1%) 64 13	50, 85, 142, 228	0
1	I	319/332 (96%)	0.10	3 (0%) 81 24	57, 78, 151, 215	0
2	B	255/264 (96%)	0.06	1 (0%) 90 41	48, 69, 134, 212	0
2	F	255/264 (96%)	0.14	0 100 100	51, 75, 136, 185	0
2	K	255/264 (96%)	0.10	4 (1%) 68 16	61, 87, 165, 241	0
3	C	55/75 (73%)	-0.05	0 100 100	55, 82, 123, 149	0
3	G	36/75 (48%)	0.42	0 100 100	77, 106, 157, 179	0
3	L	54/75 (72%)	-0.06	0 100 100	53, 88, 144, 155	0
4	D	50/58 (86%)	-0.06	0 100 100	40, 75, 104, 144	0
4	H	50/58 (86%)	0.11	0 100 100	48, 76, 102, 118	0
4	M	50/58 (86%)	-0.02	1 (2%) 62 12	43, 98, 132, 154	0
All	All	2005/2187 (91%)	0.09	16 (0%) 81 26	40, 78, 143, 241	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	K	8	ASN	4.8
2	K	7	ASN	3.9
1	I	56	LYS	2.9
1	A	179	ASN	2.9
1	I	96	LYS	2.9
1	E	177	ALA	2.8
1	E	178	LYS	2.6
1	E	179	ASN	2.5
2	K	115	GLN	2.4
1	I	181	GLN	2.4
1	E	97	GLY	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
4	M	49	ARG	2.2
1	E	248	ILE	2.1
1	E	31	GLY	2.1
2	K	118	LEU	2.0
2	B	8	ASN	2.0

## 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. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
1	TPO	E	186	11/12	0.19	-0.61	56,71,85,86	0
1	TPO	A	186	11/12	0.21	-0.95	47,62,74,75	0
1	TPO	I	186	11/12	0.18	-1.54	72,82,99,99	0

## 6.3 Carbohydrates ⓘ

There are no carbohydrates 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. 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	ADN	I	401	19/19	0.31	1.04	96,98,102,102	0
6	ZN	H	101	1/1	0.21	0.70	71,71,71,71	0
5	ADN	E	401	19/19	0.29	0.66	113,114,115,115	0
6	ZN	M	101	1/1	0.18	0.50	106,106,106,106	0
6	ZN	D	101	1/1	0.18	0.47	71,71,71,71	0
5	ADN	A	401	19/19	0.27	0.44	87,91,95,95	0
6	ZN	H	102	1/1	0.17	-0.45	95,95,95,95	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
6	ZN	D	102	1/1	0.16	-0.72	115,115,115,115	0
6	ZN	M	102	1/1	0.10	-1.57	159,159,159,159	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.