



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 18, 2018 – 06:56 pm GMT

PDB ID : 1HQM  
Title : CRYSTAL STRUCTURE OF THERMUS AQUATICUS CORE RNA POLYMERASE-INCLUDES COMPLETE STRUCTURE WITH SIDE-CHAINS (EXCEPT FOR DISORDERED REGIONS)-FURTHER REFINED FROM ORIGINAL DEPOSITION-CONTAINS ADDITIONAL SEQUENCE INFORMATION  
Authors : Minakhin, L.; Bhagat, S.; Brunning, A.; Campbell, E.A.; Darst, S.A.; Ebright, R.H.; Severinov, K.  
Deposited on : 2000-12-18  
Resolution : 3.30 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30686

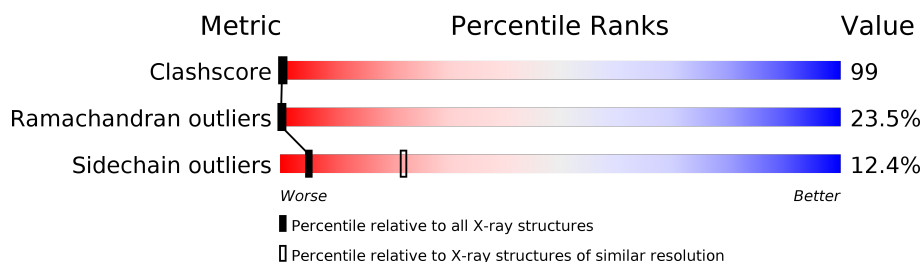
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 Å.



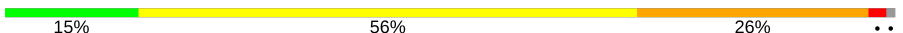

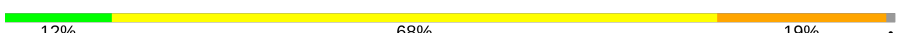
Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	122078	1021 (3.34-3.26)
Ramachandran outliers	120005	1003 (3.34-3.26)
Sidechain outliers	119972	1002 (3.34-3.26)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	313	
1	B	313	
2	C	1119	
3	D	1265	
4	E	99	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 21254 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	223	Total	C	N	O	S	0	0	0
			1750	1118	302	328	2			
1	B	229	Total	C	N	O	S	0	0	0
			1776	1135	305	334	2			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	deletion	UNP Q9KWU8
A	93	ARG	MET	conflict	UNP Q9KWU8
A	94	TRP	ALA	conflict	UNP Q9KWU8
A	95	ARG	SER	conflict	UNP Q9KWU8
A	111	VAL	GLY	conflict	UNP Q9KWU8
B	?	-	LYS	deletion	UNP Q9KWU8
B	93	ARG	MET	conflict	UNP Q9KWU8
B	94	TRP	ALA	conflict	UNP Q9KWU8
B	95	ARG	SER	conflict	UNP Q9KWU8
B	111	VAL	GLY	conflict	UNP Q9KWU8

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1113	Total	C	N	O	S	12	0	0
			8508	5386	1514	1585	23			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2	LYS	GLU	conflict	UNP Q9KWU7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1175	Total	C	N	O	S	17	0	0
			8499	5328	1549	1595	27			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	119	PHE	SER	conflict	UNP Q9KWU6
D	863	THR	VAL	conflict	UNP Q9KWU6
D	866	THR	VAL	conflict	UNP Q9KWU6
D	876	ASN	SER	conflict	UNP Q9KWU6
D	947	ILE	-	insertion	UNP Q9KWU6
D	1010	ASN	LYS	conflict	UNP Q9KWU6
D	1117	LYS	ASN	conflict	UNP Q9KWU6
D	1389	PRO	ARG	conflict	UNP Q9KWU6

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	98	Total	C	N	O	S	0	0	0
			719	453	132	130	4			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	1	Total	Mg	0	0
			1	1		

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

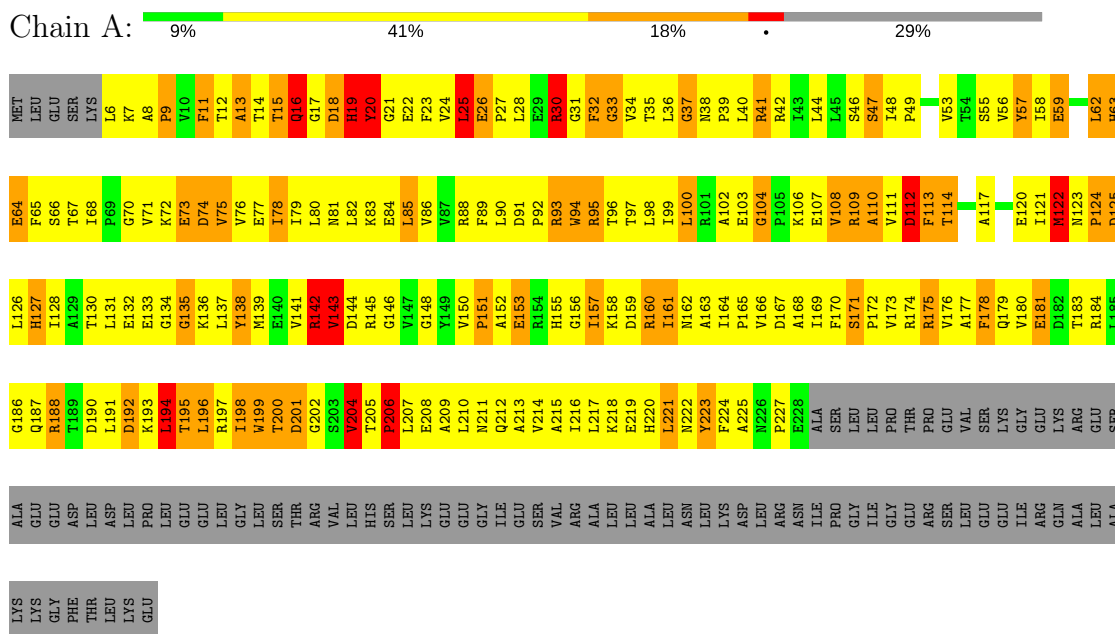
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	1	Total	Zn	0	0
			1	1		

### 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 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.

Note EDS was not executed.

- Molecule 1: DNA-directed RNA polymerase subunit alpha

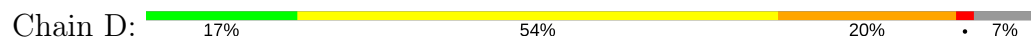


ASP	PHE
LEU	THR
ASP	LEU
LEU	LYS
PRO	GLU
LEU	
GLU	
LEU	
GLY	
LEU	
SER	
THR	
ARG	
VAL	
LEU	
SER	
HIS	
SER	
LEU	
LYS	
GLU	
GLY	
ILE	
GLU	
SER	
VAL	
ARG	
ALA	
LEU	
LEU	
ALA	
LEU	
ASN	
LYS	
ASP	
LEU	
ARG	
ASN	
ILE	
PRO	
GLY	
ILE	
GLY	
ILE	
GLU	
ILE	
ARG	
GLN	
ALA	
LEU	
ALA	
LYS	
LYS	
GLY	

• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C: 15% 56% 26% ..

L804	R805	L806	R807	D810	P811	G812	V813	E814	L815	G816	P817	G818	V819	R820	E821	R822	A823	R824	E825	R826	G827	E828	Q829	R830	K831	R832	K833	Q834	R835	G836	D837	K838	L839	A840	N841	R842	H843	G844	N845	K846	G847	V848	V849	A850	K851	I852	L853	P854	V855	E856	D857	M858	P859	H860	L861	P862	D863					
G681	Y682	M683	F684	E685	D686	A687	I688	K689	E690	S691	D692	E693	L694	L695	K696	R697	D698	F699	Y700	T701	S702	I703	H704	I705	E706	S707	E708	E709	I710	E711	A712	R713	D714	T715	K716	L717	G718	E719	E720	R721	I722	S661	E662	E663	G664	F665	L666	A667	L668	M671	V672	L673	V674	A675	A676	L677	P678	F679	R680	G681	D682	R683
I742	V743	R744	L745	G746	A747	E748	V749	K750	P751	G752	D753	L754	L755	V756	G757	R758	T759	S760	F761	K762	G763	E764	Q765	E766	P767	S768	P769	E770	E771	R772	L773	L774	R775	S776	L777	E780	K781	A782	R783	D784	V785	K786	D787	T788	S789	L790	R791	L792	P793	E794	G795	E796	G797	V800	V801	P802	D803					
G681	Y682	M683	F684	E685	D686	A687	I688	K689	E690	S691	D692	E693	L694	L695	K696	R697	D698	F699	Y700	T701	S702	I703	H704	I705	E706	S707	E708	E709	I710	E711	A712	R713	D714	T715	K716	L717	G718	E719	E720	R721	I722	S661	E662	E663	G664	F665	L666	A667	L668	M671	V672	L673	V674	A675	A676	L677	P678	F679	R680	G681	D682	R683
N556	R557	A558	L559	M560	G561	S562	N563	M564	Q565	T566	Q567	A568	V569	P570	L571	I572	R573	A574	Q575	A576	P577	E578	V579	N580	T581	G582	L583	E584	E585	R586	V587	E588	R589	A593	A594	L595	T596	A597	E598	E599	D600	G601	E602	V603	V604	K605	V606	D607	G608	T609	L610	R611	A612	A613	R614	Y615	D616	D617				
I496	A497	Q498	A499	H500	T501	P502	L503	E504	G505	D506	R507	I508	A509	T510	L511	R512	V513	V514	A515	R516	P517	E518	G519	E520	P521	V522	L523	V524	A525	E526	E527	E528	V529	F530	M531	M532	D533	V534	S535	P536	K537	Q538	V539	F540	S541	L542	V543	M544	M545	L546	T547	P548	R549	E550	A551	H552	P553	D554	A555			
R432	T433	H434	A435	G436	R437	I438	C439	P440	V441	E442	T443	P444	E445	G446	A447	I448	T449	S453	S454	L455	E456	A457	Y458	A459	F460	L461	D462	A463	L464	G465	F466	E467	T468	P469	Y470	Y471	R472	R473	V474	K475	V479	T480	E481	E482	V483	V484	Y485	M486	T487	A488	F489	E490	A491	D492	R493	D494	A495					
K371	L372	V373	K374	R375	R376	P377	L378	E379	A380	A381	L382	R383	E384	F385	R386	S387	R388	S389	Q390	L391	S392	Q393	F394	K395	D396	E397	T398	N399	P400	L401	S402	F403	L404	R405	T406	P407	Y408	R409	L410	S411	A412	L413	G414	P415	G416	T419	E420	E421	R422	A423	G424	F425	A426	V427	R428	D429	V430	H431				
L310	F311	A312	L313	T314	G315	V316	P317	G318	G319	H320	L321	V322	D323	D324	I325	D326	H327	L328	G329	N330	R331	R332	I333	R334	T335	V336	G337	E338	A339	G340	L341	D342	Q343	F344	R345	V346	G347	L348	A349	R350	L351	A352	R353	G354	V355	E356	E357	R358	M359	V360	M361	G362	S363	T364	L365	R366	P367	A368	A369	A370		
P247	F248	K249	R250	D251	K252	A253	L254	A255	L256	Y257	F258	E259	L260	L261	A262	D263	P264	K265	E266	Y267	D268	L269	G270	E271	A272	G273	R274	Y275	E276	L281	G282	Q283	G284	L285	S286	T289	L290	V291	E292	F293	E294	D295	G296	E297	F298	K299	D300	E301	V302	F303	S304	L305	T306	L307	R308	Y309						
V186	M187	K188	R189	K190	F191	P192	L193	V194	L195	L196	L197	R198	V199	L200	G201	Y202	D203	Q204	E205	T206	L207	V208	R209	E210	L211	S212	A213	Y214	G215	D216	L217	V218	Q219	G220	L221	L222	D223	E224	A225	P226	L227	V228	E229	D230	P231	E232	E233	A234	M235	V236	L237	L238	L239	L240	L241	L242	R243	P244	M184	G245	D246	
T122	E123	D124	F127	I128	I129	R130	G131	R134	I135	V136	G137	S138	P139	I140	H141	R142	S143	P144	G145	V146	E147	F148	T149	P150	D151	T89	A96	P96	I97	A98	Q99	E100	H101	H102	K103	D104	T105	G106	L107	T108	K109	E110	D111	E112	V113	F114	L115	E116	G117	H118	L119	P120	M121									
G62	G63	L64	V65	L66	D67	F68	L69	E70	Y71	I72	V73	I74	D75	P76	P77	F78	H79	Q80	D81	E82	C83	R84	E85	E86	E87	L88	T89	A96	P96	I97	A98	Q99	E100	H101	H102	K103	D104	T105	G106	L107	T108	K109	E110	D111	E112	V113	F114	L115	E116	G117	H118	L119	P120	M121								
MET	K2	R5	G7	F6	R8	I9	R10	E11	V12	I13	G14	I15	P16	P17	L18	H19	S79	E20	I21	Q22	V23	E24	S25	E26	K27	K28	A29	P35	E37	K38	R39	E40	N41	V42	G43	I44	Q45	A46	A47	F48	K49	E50	T51	F52	P53	E55	E56	G57	D58	K59	G60	K61										



WORLDWIDE  
**PDB**  
PROTEIN DATA BANK



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	200.76Å 200.76Å 292.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 3.30	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-3.30)	Depositor
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 0.9	Depositor
R, $R_{free}$	0.300 , 0.360	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	21254	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

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

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	0.42	0/1786	0.77	0/2434
1	B	0.39	0/1812	0.74	0/2471
2	C	0.42	0/8672	0.78	5/11752 (0.0%)
3	D	0.42	0/8437	0.78	14/11443 (0.1%)
4	E	0.35	0/730	0.65	0/991
All	All	0.41	0/21437	0.77	19/29091 (0.1%)

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
1	A	0	1
2	C	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	834	THR	N-CA-C	-6.40	93.72	111.00
3	D	137	PRO	N-CA-CB	6.22	110.77	103.30
2	C	580	MET	N-CA-C	6.19	127.70	111.00
2	C	836	GLY	N-CA-C	-6.04	97.99	113.10
2	C	329	GLY	N-CA-C	-5.97	98.17	113.10

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	138	TYR	Sidechain
2	C	975	TYR	Sidechain

## 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	1750	0	1759	403	0
1	B	1776	0	1776	323	0
2	C	8508	0	8418	1886	0
3	D	8499	0	7993	1651	0
4	E	719	0	685	125	0
5	D	1	0	0	0	0
6	D	1	0	0	0	0
All	All	21254	0	20631	4161	0

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 99.

The worst 5 of 4161 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:1020:PRO:HB2	3:D:1023:VAL:HB	1.20	1.18
2:C:508:ILE:H	2:C:508:ILE:HD13	1.10	1.15
2:C:438:ILE:HG21	2:C:470:PRO:HB3	1.22	1.15
2:C:605:LYS:HG2	2:C:606:VAL:H	1.05	1.14
2:C:262:ALA:HB1	2:C:266:ARG:HD2	1.23	1.14

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	221/313 (71%)	98 (44%)	68 (31%)	55 (25%)	0	0
1	B	227/313 (72%)	109 (48%)	61 (27%)	57 (25%)	0	0
2	C	1111/1119 (99%)	559 (50%)	300 (27%)	252 (23%)	0	0
3	D	1127/1265 (89%)	543 (48%)	319 (28%)	265 (24%)	0	0
4	E	96/99 (97%)	49 (51%)	22 (23%)	25 (26%)	0	0
All	All	2782/3109 (90%)	1358 (49%)	770 (28%)	654 (24%)	0	0

5 of 654 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	26	GLU
1	A	59	GLU
1	A	64	GLU
1	A	73	GLU
1	A	75	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	190/271 (70%)	163 (86%)	27 (14%)	3	17
1	B	191/271 (70%)	171 (90%)	20 (10%)	7	28
2	C	869/936 (93%)	747 (86%)	122 (14%)	4	18
3	D	782/1036 (76%)	693 (89%)	89 (11%)	6	25
4	E	67/88 (76%)	64 (96%)	3 (4%)	30	63
All	All	2099/2602 (81%)	1838 (88%)	261 (12%)	5	22

5 of 261 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	672	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	C	873	PRO
3	D	1197	THR
2	C	680	ASP
2	C	758	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 68 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	845	ASN
2	C	1026	GLN
3	D	1368	HIS
2	C	860	HIS
2	C	889	HIS

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	D	3

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	155:ASP	C	2(U):UNK	N	55.17
1	D	46(U):UNK	C	452:ILE	N	46.65
1	D	10(U):UNK	C	20(U):UNK	N	14.79

## 6 Fit of model and data [i](#)

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

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates [i](#)

EDS was not executed - this section is therefore empty.

### 6.4 Ligands [i](#)

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers [i](#)

EDS was not executed - this section is therefore empty.