



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 07:03 AM 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 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/ValidationPDFNotes.html>

---

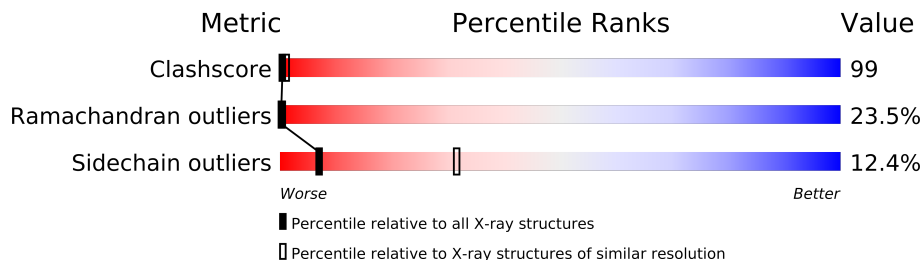
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	21963
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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	79885	1696 (3.40-3.20)
Ramachandran outliers	78287	1664 (3.40-3.20)
Sidechain outliers	78261	1662 (3.40-3.20)

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.

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 hydrogen 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 DNA-DIRECTED RNA POLYMERASE.

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.

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

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2	LYS	GLU	CONFLICT	GB 7573273
C	?	-	GLU	DELETION	GB 7573273

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE.

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	GB 10803423
D	863	THR	VAL	CONFLICT	GB 10803423
D	866	THR	VAL	CONFLICT	GB 10803423
D	876	ASN	SER	CONFLICT	GB 10803423
D	947	ILE	-	INSERTION	GB 10803423
D	1010	ASN	LYS	CONFLICT	GB 10803423
D	1117	LYS	ASN	CONFLICT	GB 10803423
D	1389	PRO	ARG	CONFLICT	GB 10803423

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE.

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		

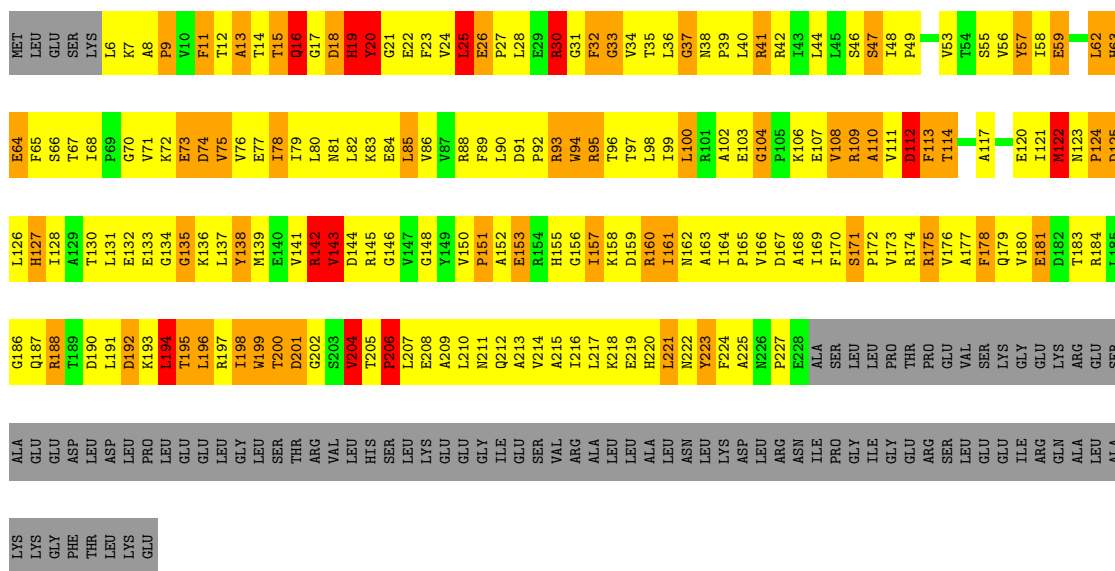
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

Note EDS was not executed.

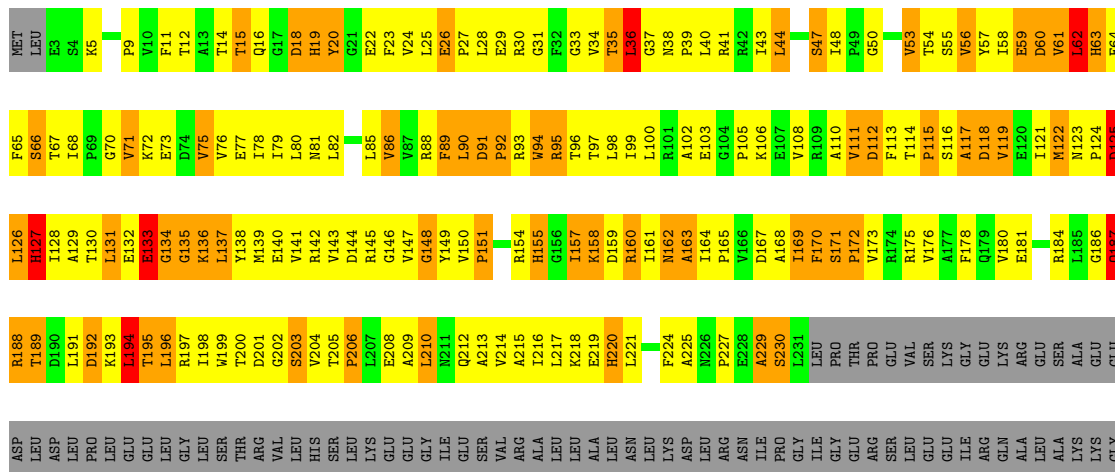
#### • Molecule 1: DNA-DIRECTED RNA POLYMERASE

Chain A:



#### • Molecule 1: DNA-DIRECTED RNA POLYMERASE

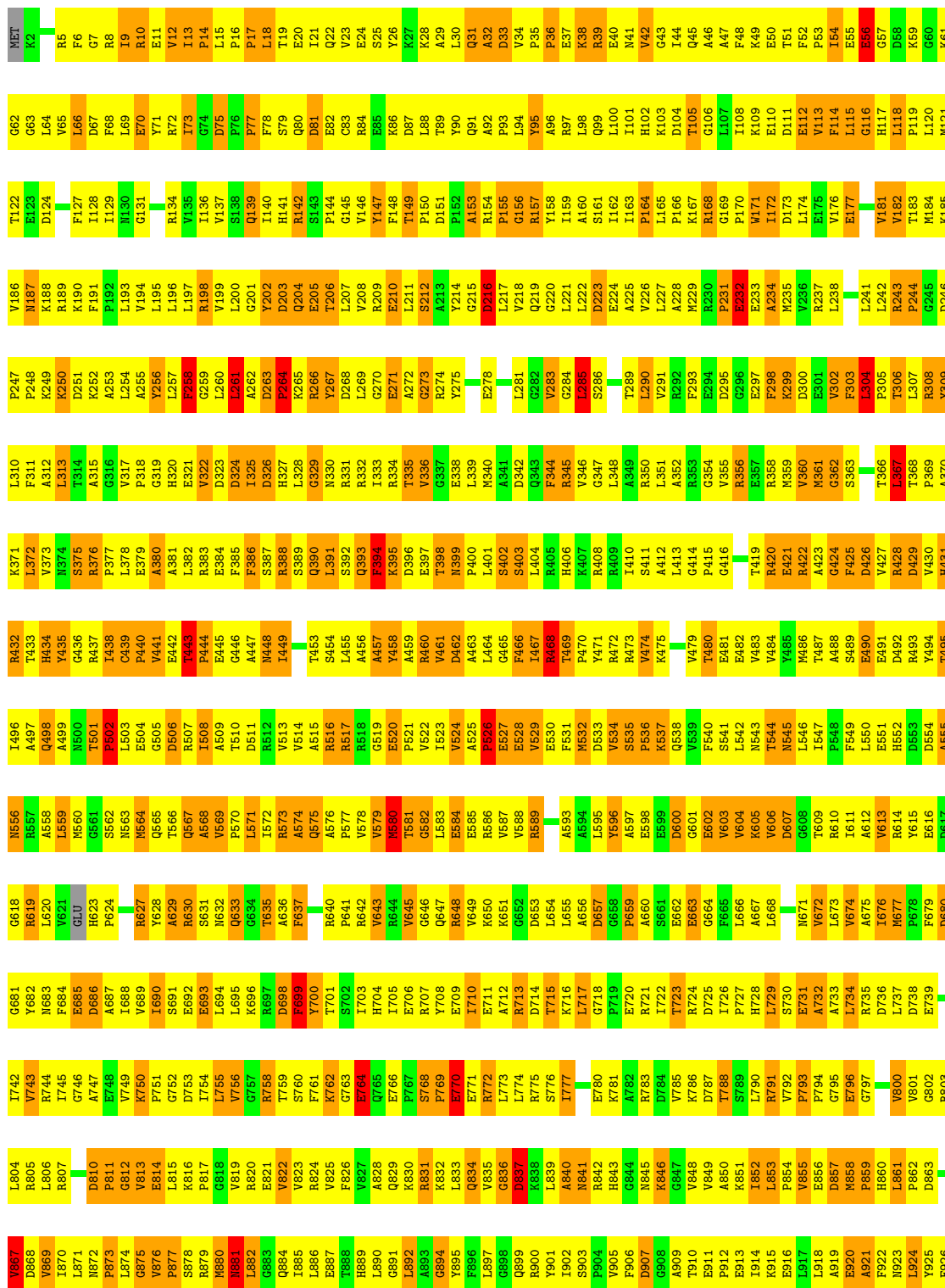
Chain B:



PHE  
THR  
LEU  
LYS  
GLU

- Molecule 2: DNA-DIRECTED RNA POLYMERASE

Chain C:

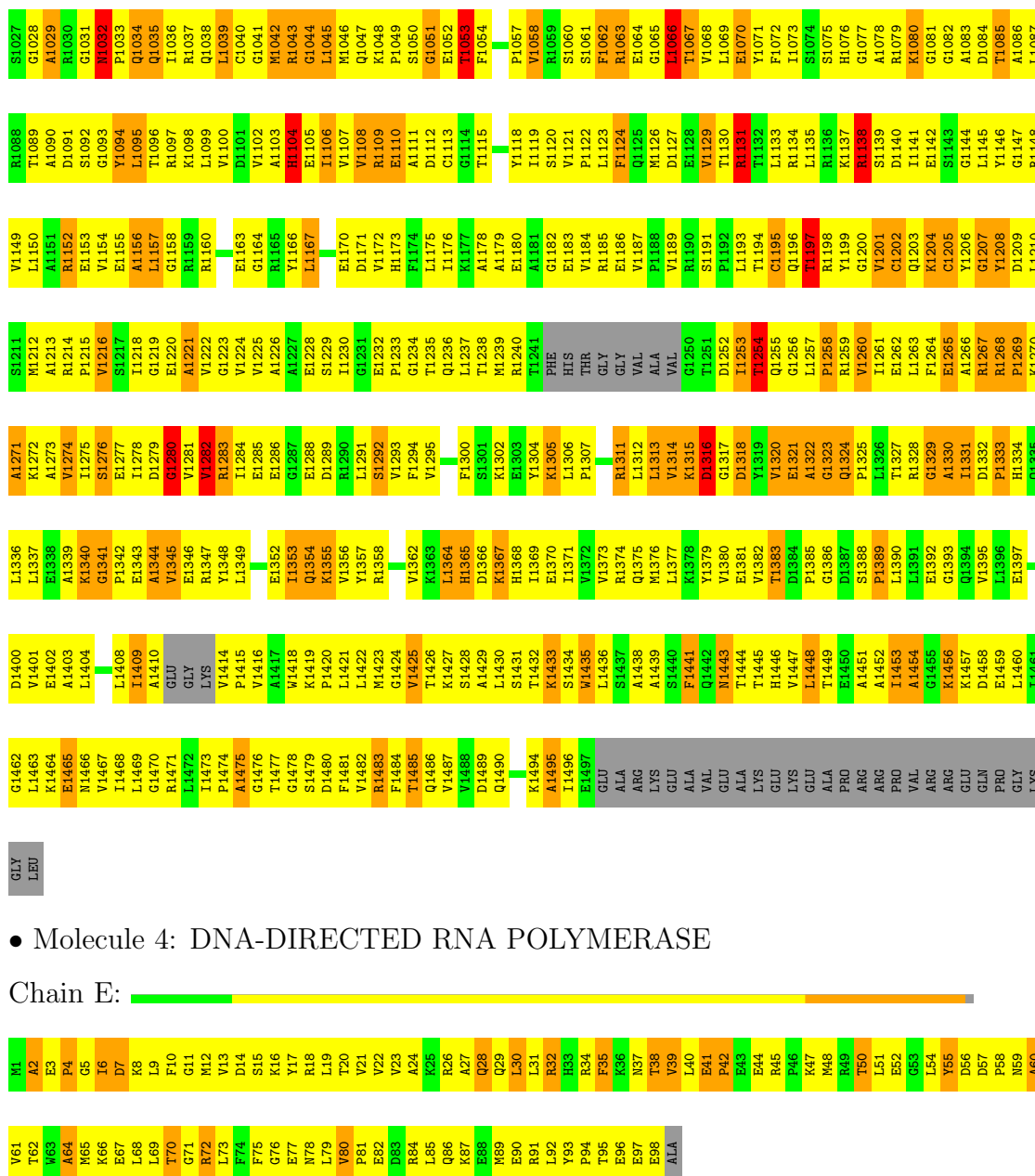


- Molecule 3: DNA-DIRECTED RNA POLYMERASE

Chain D:



P005	L339	E776	A715	K654	R537	PRO	K466	Y128	MET
Q906	R340	P777	F716	P655	G589	MET	L467	F129	LYS
Q907	R341	P778	F717	P656	S589	VAL	L468	N130	ARG
K908	R342	A779	F718	L657	P590	GLN	D469	K131	PHE
N909	R343	K780	F719	L658		VAL	L470	Y132	V8
S910	R344		F720	K659	E597	ASP	E471	I133	R9
L911	R345	R783	F721	K660	E598	GLY	K472	V134	I10
K912	P846	D784	E722	M661	P599	GLY	L473	L135	A11
D913			G723	E662	L600	ARG	E474	D136	
L914	A849	L787	G724	E663	R601	PHE	L475	P137	
V915	A850	G788	S725	K664	S602	ASP	E476	K138	S14
V916	R553	L789	F726	A665	L603	THR	L477	G139	P15
	A852	Y790	G727	P666	T604	SSS8	L478	A140	F16
F919	R854	Y791	L728	A667	D605	E479	E480	F141	K17
L920	R855	F792	R729	P668	L606	L540	E481	L142	R18
R921	R856	T793	P730	N669	L607	N541	N483	D143	R19
L922	L857	Q794	L731	V670	L607	D542	K482	G144	S20
G923	L858	F795	F732	K671	S608	L543	K483	V145	W21
M924	D859	R796	C733	A672		Y544	P484	F146	S22
E925	L860	Q797	E734	A673	R613	R545	S485	P147	Y23
R926	Q861	E798	F735	R674	F614	R546	R486	E148	G24
T927	D862	F799	F736	R675	R615	L547	A487	K149	E25
R928	R863	K800	N737	M676	D616	I548	R488	R150	E27
Q929	R864	G801	A738	L677	R617	N549	R489	Q151	K28
L930	R865	A802	D739	E678	L618	R550	A490	L152	P29
L931	R866	R803	F740	R679	L619	N551	K491	L153	E30
D932		M804	Q680	Q680	G620	N552	A492	T154	G31
L934	L869	F806	D743	R681	K621	R553	R493	X9U	ILE
K935	R872	A807	W745	L683	G623	L554	K494	X10U	ASN
Y936	L873	T808	A746	K694	D624	K556	L496		TYR
Y937	E874	P809	F747	D685	Y625	L557	E497	X23U	ARG
Q938	R875	E810	W748	E686	S626	L558	V498	X24U	LEU
F939	R876	E811	F749	V687	G627	A559	V499	X25U	LYS
T1001	P877	A812	T750	W688	R628	Q660	R500		PRO
L941	C878	L813	L751	D689	S629		A501	X28U	GLU
K1003	R879	A814	S752	A690	G630	P563	F502	X29U	ASP
T1004	L880	A815	S753	L691	L631	E564	F104	X30U	GLY
T944	L881	R816	F754	E592	G632	L565	D504		LEU
S945	R882	E817	A755	E593	G633	L566	S505	X35U	PHE
G946	A883	R818	Q756	V694	G634	L567	G506	K36U	ASP
L947	R884		A757	L695	P635	R568	N507	X37U	GLU
T948	L885	V821	E758	H696	Q636	R569	R508	X38U	ARG
T949	V886	A822	A759	G697	L637	E570	P509	X39U	ILE
F1010		L823	R760	K698	L638	K571	E510	X40U	PHE
G951	R890	N824	F761	V699	L639	N511	X41U	X41U	ILE
L952	G891	A825	Q762	V700	H640	R572	G13	X41U	GLY
E953	D892	P826	W763	L701	G641	M573	T114	X42U	PRO
D954	R893	L827	L764	G642	G641	L574	L115	X43U	ILE
L955	K894	R828	S765	R703	G643	Q575	L116		LYS
P1017	R895	E829	A766	R704	L644	E577	D117	I452	ASP
F1018	A896	R830	Q767	A705	P645	V517	A454	A454	TYR
P958	Q897	G831	N768	P706	K646	D579	R455	A120	GLU
E959	R898	R832	L769		R647	A580	T121	T121	CYS
E960	L899	E833	F770	R709	R648	V581	E122		ALA
T1022	K901	T834	S771	F7		L582	L123		CYS
G962	R902	R835	R772	L711	L651	D583	E124	I461	GLY
M1023	R963	R836	F773	G712	R651	N584	Q125	E463	LYS
A1025	Y964	D903	S774	T713	L652	L585	L464	L464	TYR
	L965	T904	G775	T714	L552	R586	L465	L465	LYS



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section will therefore be 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 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	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

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 99.

The worst 5 of 4161 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(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%)	5	24
1	B	191/271 (70%)	171 (90%)	20 (10%)	10	40
2	C	869/936 (93%)	747 (86%)	122 (14%)	5	25
3	D	782/1036 (76%)	693 (89%)	89 (11%)	8	36
4	E	67/88 (76%)	64 (96%)	3 (4%)	38	81
All	All	2099/2602 (81%)	1838 (88%)	261 (12%)	7	31

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

Mol	Chain	Res	Type
2	C	672	VAL
2	C	873	PRO

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
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 ⓘ

There are no RNA chains 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 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.

## 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 ⓘ

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.