



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 06:29 am BST

PDB ID : 4Y97
Title : Crystal Structure of human Pol alpha B-subunit in complex with C-terminal domain of catalytic subunit
Authors : Suwa, Y.; Gu, J.; Baranovskiy, A.G.; Babayeva, N.D.; Tahirov, T.H.
Deposited on : 2015-02-17
Resolution : 2.51 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

<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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

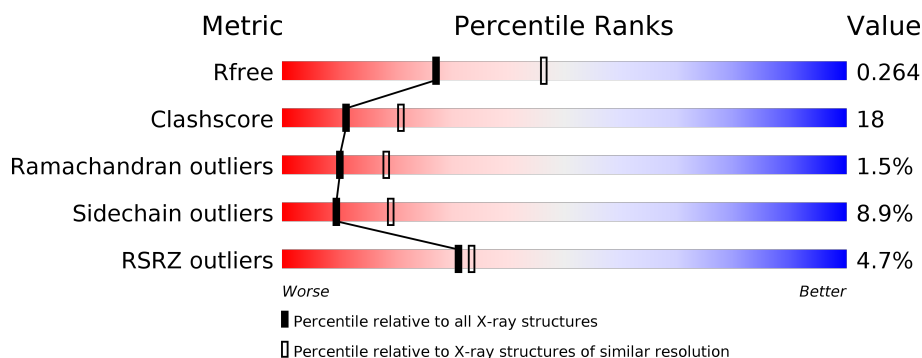
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.51 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}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 respectively. 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	598	<div> <div>2%</div> <div> <div></div> <div>50%</div> <div>21%</div> <div>•</div> <div>26%</div> </div> </div>
1	C	598	<div> <div>2%</div> <div> <div></div> <div>51%</div> <div>19%</div> <div>•</div> <div>26%</div> </div> </div>
1	E	598	<div> <div>0%</div> <div> <div></div> <div>53%</div> <div>17%</div> <div>•</div> <div>27%</div> </div> </div>
1	G	598	<div> <div>2%</div> <div> <div></div> <div>47%</div> <div>21%</div> <div>5%</div> <div>•</div> <div>27%</div> </div> </div>
2	B	180	<div> <div>10%</div> <div> <div></div> <div>56%</div> <div>35%</div> <div>6%</div> <div>•</div> </div> </div>
2	D	180	<div> <div>8%</div> <div> <div></div> <div>66%</div> <div>27%</div> <div>6%</div> <div>•</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	180	<div><div></div><div>4%</div><div>61%</div><div>33%</div><div></div><div></div></div>
2	H	180	<div><div></div><div>20%</div><div>57%</div><div>38%</div><div></div><div></div></div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 19724 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 polymerase alpha subunit B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	442	Total	C	N	O	S	0	0	0
			3439	2187	574	663	15			
1	C	441	Total	C	N	O	S	0	0	0
			3435	2187	573	660	15			
1	E	439	Total	C	N	O	S	0	0	0
			3422	2178	571	658	15			
1	G	436	Total	C	N	O	S	0	0	0
			3400	2164	568	653	15			

- Molecule 2 is a protein called DNA polymerase alpha catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	173	Total	C	N	O	S	0	0	0
			1426	902	240	269	15			
2	D	178	Total	C	N	O	S	0	0	0
			1466	926	248	277	15			
2	F	176	Total	C	N	O	S	0	0	0
			1451	919	244	273	15			
2	H	179	Total	C	N	O	S	0	0	0
			1475	931	250	279	15			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	2	Total	Zn	0	0
			2	2		
3	B	2	Total	Zn	0	0
			2	2		
3	D	2	Total	Zn	0	0
			2	2		
3	F	2	Total	Zn	0	0
			2	2		

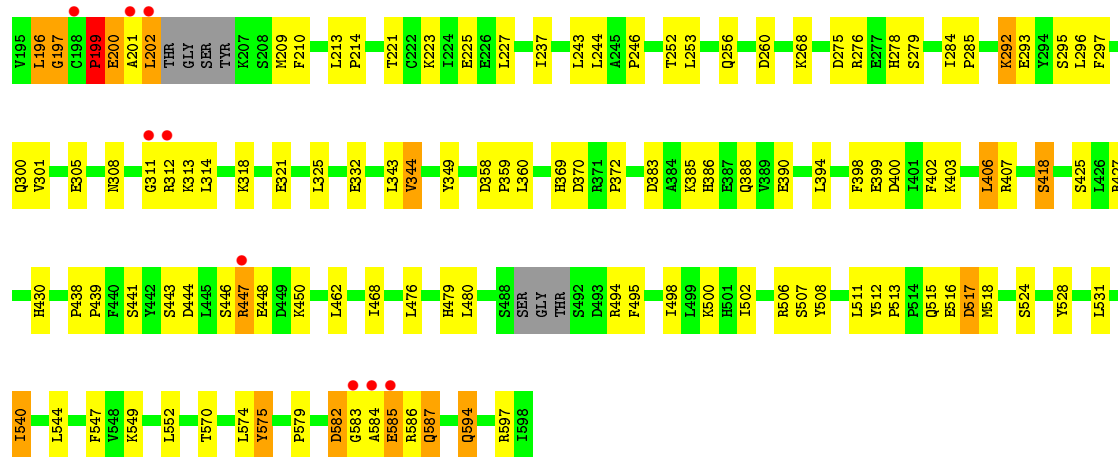
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	37	Total 37	O 37	0	0
4	B	8	Total 8	O 8	0	0
4	C	63	Total 63	O 63	0	0
4	D	24	Total 24	O 24	0	0
4	E	39	Total 39	O 39	0	0
4	F	14	Total 14	O 14	0	0
4	G	13	Total 13	O 13	0	0
4	H	4	Total 4	O 4	0	0

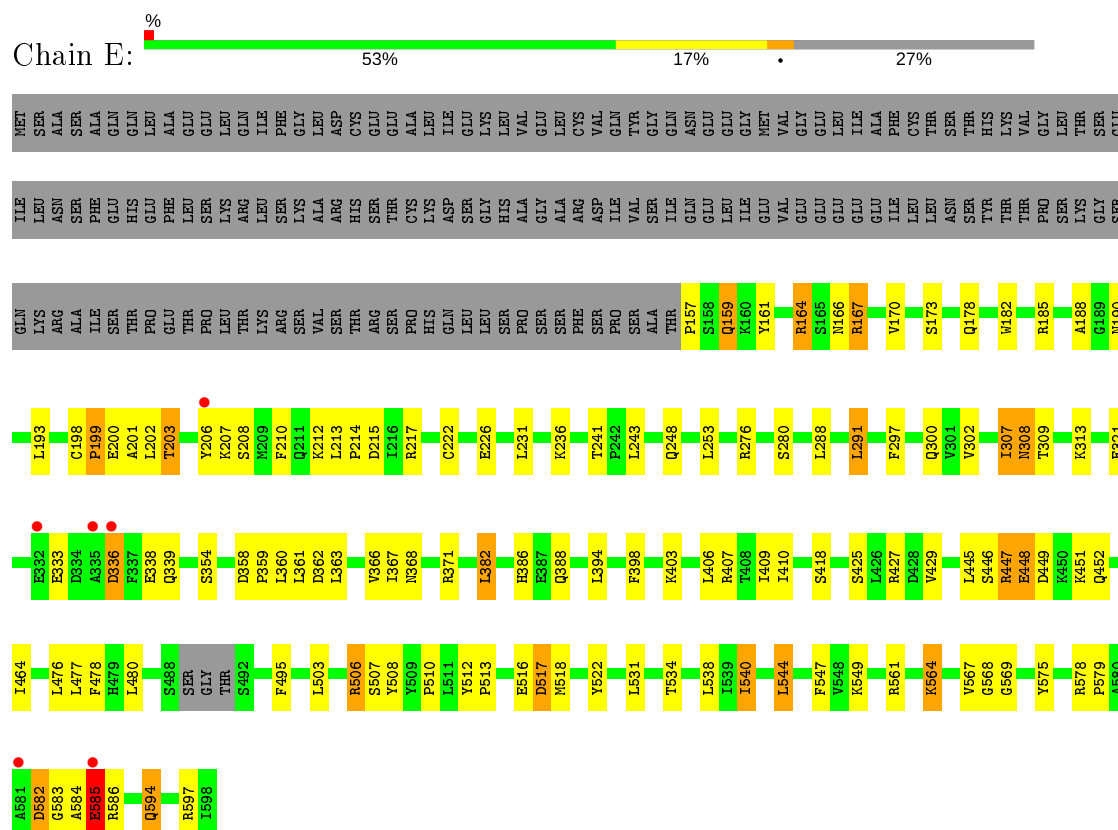
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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 ($\text{RSRZ} > 2$). 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.

- [illegible]

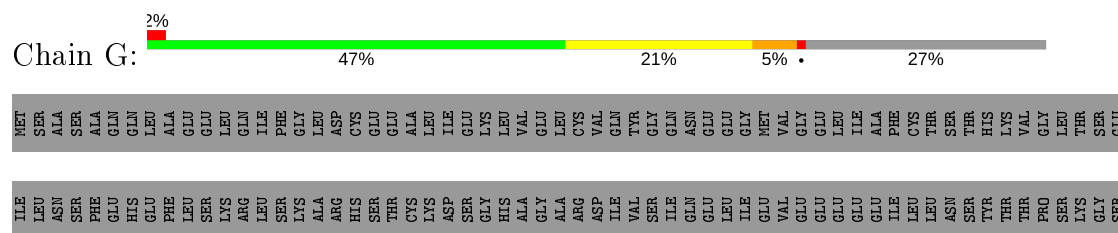
- [illegible]

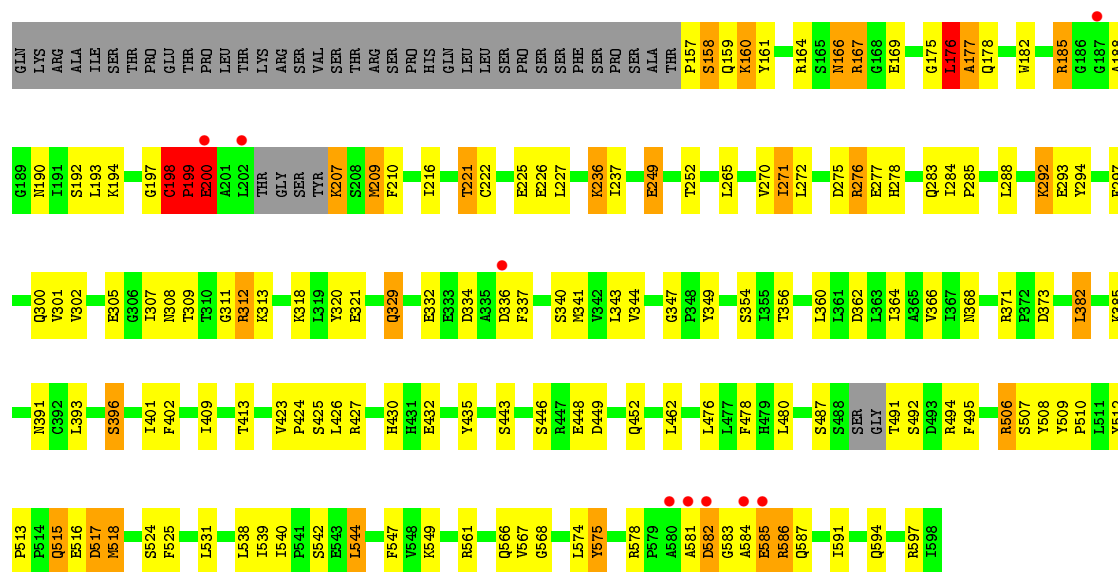


• Molecule 1: DNA polymerase alpha subunit B

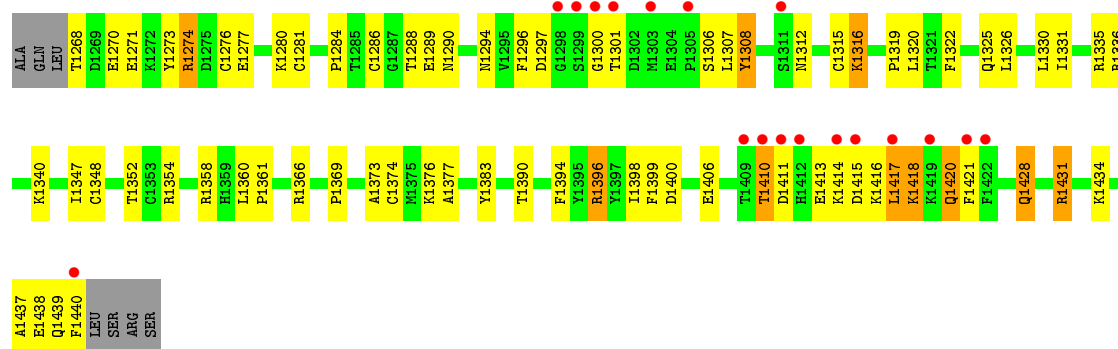


• Molecule 1: DNA polymerase alpha subunit B

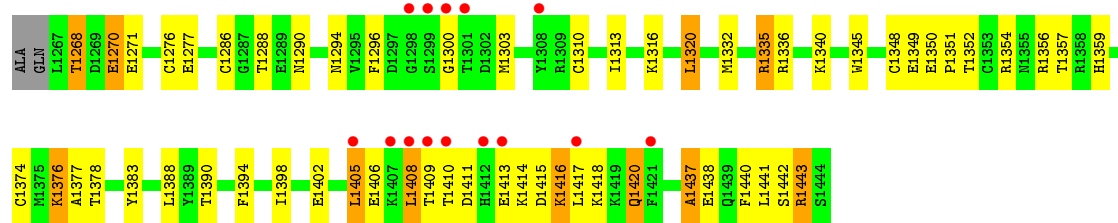




• Molecule 2: DNA polymerase alpha catalytic subunit

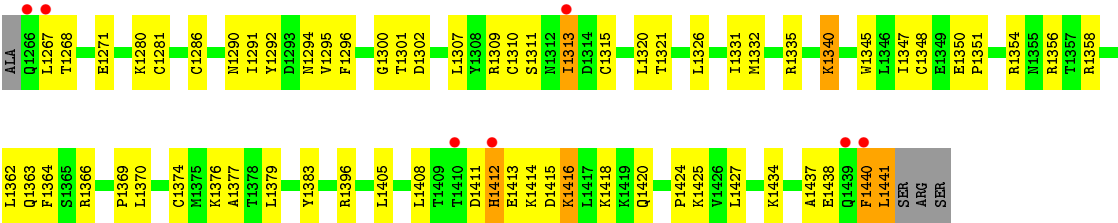


• Molecule 2: DNA polymerase alpha catalytic subunit

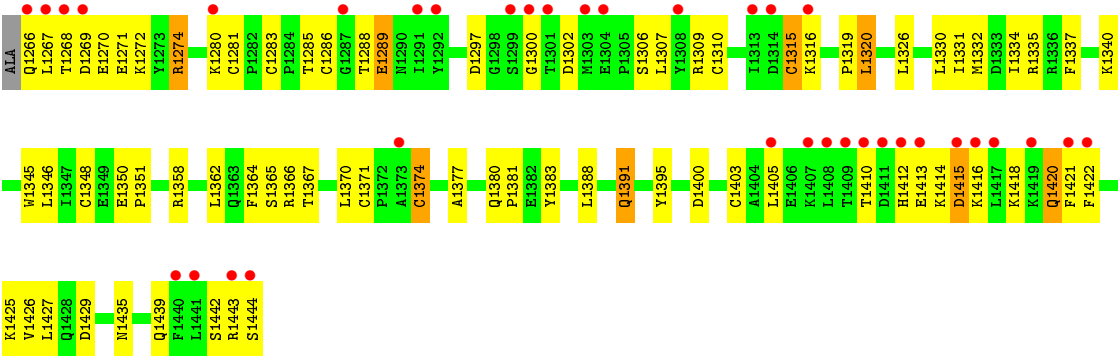


• Molecule 2: DNA polymerase alpha catalytic subunit





● Molecule 2: DNA polymerase alpha catalytic subunit



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	103.59Å 137.13Å 132.09Å 90.00° 100.97° 90.00°	Depositor
Resolution (Å)	47.68 – 2.51 47.68 – 2.51	Depositor EDS
% Data completeness (in resolution range)	94.2 (47.68-2.51) 94.3 (47.68-2.51)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.07 (at 2.51Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.218 , 0.257 0.228 , 0.264	Depositor DCC
R_{free} test set	5894 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	31.2	Xtriage
Anisotropy	0.639	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 53.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	19724	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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

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	0.43	0/3517	0.74	1/4777 (0.0%)
1	C	0.49	0/3512	0.84	7/4769 (0.1%)
1	E	0.46	0/3499	0.77	3/4751 (0.1%)
1	G	0.39	0/3475	0.78	8/4717 (0.2%)
2	B	0.43	0/1459	0.69	2/1967 (0.1%)
2	D	0.44	0/1499	0.67	1/2019 (0.0%)
2	F	0.42	0/1484	0.71	1/2001 (0.0%)
2	H	0.40	0/1508	0.68	1/2031 (0.0%)
All	All	0.44	0/19953	0.76	24/27032 (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	E	0	1

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	156	THR	C-N-CD	-20.28	75.98	120.60
1	G	198	CYS	N-CA-C	-10.39	82.94	111.00
1	C	156	THR	C-N-CA	9.90	163.59	122.00
1	G	198	CYS	C-N-CD	-9.72	99.21	120.60
1	G	197	GLY	N-CA-C	7.01	130.63	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	522	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	3439	0	3414	134	0
1	C	3435	0	3410	132	0
1	E	3422	0	3398	78	0
1	G	3400	0	3380	144	0
2	B	1426	0	1389	53	0
2	D	1466	0	1434	50	0
2	F	1451	0	1419	63	0
2	H	1475	0	1442	72	0
3	B	2	0	0	0	0
3	D	2	0	0	0	0
3	F	2	0	0	0	0
3	H	2	0	0	0	0
4	A	37	0	0	3	0
4	B	8	0	0	0	0
4	C	63	0	0	3	0
4	D	24	0	0	2	0
4	E	39	0	0	4	0
4	F	14	0	0	2	0
4	G	13	0	0	2	0
4	H	4	0	0	0	0
All	All	19724	0	19286	691	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 18.

The worst 5 of 691 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:E:161:TYR:O	1:E:164:ARG:HD2	1.42	1.16
1:G:209:MET:HB2	2:H:1443:ARG:HB3	1.30	1.11
1:C:185:ARG:HB3	1:C:185:ARG:HH11	1.06	1.08

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:166:ASN:H	1:C:166:ASN:ND2	1.46	1.08
1:G:169:GLU:HB3	1:G:597:ARG:HD2	1.40	1.03

There are no symmetry-related clashes.

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	440/598 (74%)	412 (94%)	22 (5%)	6 (1%)	11	20
1	C	435/598 (73%)	409 (94%)	18 (4%)	8 (2%)	8	14
1	E	435/598 (73%)	414 (95%)	16 (4%)	5 (1%)	14	26
1	G	430/598 (72%)	400 (93%)	19 (4%)	11 (3%)	5	8
2	B	171/180 (95%)	163 (95%)	7 (4%)	1 (1%)	25	43
2	D	176/180 (98%)	166 (94%)	8 (4%)	2 (1%)	14	26
2	F	174/180 (97%)	164 (94%)	7 (4%)	3 (2%)	9	16
2	H	177/180 (98%)	168 (95%)	8 (4%)	1 (1%)	25	43
All	All	2438/3112 (78%)	2296 (94%)	105 (4%)	37 (2%)	10	18

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	158	SER
1	A	517	ASP
1	A	585	GLU
1	C	157	PRO
1	C	199	PRO

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	389/527 (74%)	361 (93%)	28 (7%)	14	28
1	C	389/527 (74%)	356 (92%)	33 (8%)	10	21
1	E	387/527 (73%)	350 (90%)	37 (10%)	8	16
1	G	385/527 (73%)	347 (90%)	38 (10%)	8	15
2	B	162/168 (96%)	147 (91%)	15 (9%)	9	17
2	D	167/168 (99%)	152 (91%)	15 (9%)	9	19
2	F	165/168 (98%)	149 (90%)	16 (10%)	8	16
2	H	168/168 (100%)	154 (92%)	14 (8%)	11	22
All	All	2212/2780 (80%)	2016 (91%)	196 (9%)	9	19

5 of 196 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	1420	GLN
1	E	403	LYS
1	G	591	ILE
1	E	159	GLN
1	E	231	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	256	GLN
1	E	386	HIS
2	H	1294	ASN
1	E	283	GLN
1	E	329	GLN

5.3.3 RNA ⓘ

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 8 ligands modelled in this entry, 8 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](#)

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	442/598 (73%)	-0.07	12 (2%) 54 58	18, 42, 87, 110	0
1	C	441/598 (73%)	-0.12	14 (3%) 47 51	16, 35, 86, 106	0
1	E	439/598 (73%)	-0.14	6 (1%) 75 77	20, 38, 86, 105	0
1	G	436/598 (72%)	0.03	9 (2%) 63 66	31, 53, 91, 115	0
2	B	173/180 (96%)	0.45	18 (10%) 6 6	15, 55, 104, 116	0
2	D	178/180 (98%)	0.22	14 (7%) 12 12	17, 50, 96, 110	0
2	F	176/180 (97%)	0.03	7 (3%) 38 41	26, 48, 89, 104	0
2	H	179/180 (99%)	0.89	36 (20%) 1 0	39, 71, 105, 118	0
All	All	2464/3112 (79%)	0.06	116 (4%) 31 33	15, 46, 94, 118	0

The worst 5 of 116 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1300	GLY	8.4
1	C	152	SER	8.3
1	G	580	ALA	7.4
2	D	1300	GLY	6.5
2	H	1412	HIS	5.8

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. 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	B-factors(\AA^2)	Q<0.9
3	ZN	H	1501	1/1	0.93	0.05	88,88,88,88	0
3	ZN	B	1501	1/1	0.96	0.04	82,82,82,82	0
3	ZN	F	1502	1/1	0.96	0.10	62,62,62,62	0
3	ZN	H	1502	1/1	0.98	0.09	64,64,64,64	0
3	ZN	F	1501	1/1	0.99	0.07	53,53,53,53	0
3	ZN	D	1501	1/1	0.99	0.06	66,66,66,66	0
3	ZN	D	1502	1/1	1.00	0.10	43,43,43,43	0
3	ZN	B	1502	1/1	1.00	0.14	37,37,37,37	0

6.5 Other polymers [i](#)

There are no such residues in this entry.