wwPDB X-ray Structure Validation Summary Report (i)

Nov 7, 2023 – 06:44 PM JST

PDB ID	:	8H2N
Title	:	gp96 RNA polymerase from P23-45 phage (crystal 2)
Authors	:	Chaban, A.; Sokolova, M.L.; Tagami, S.
Deposited on	:	2022-10-06
Resolution	:	4.41 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

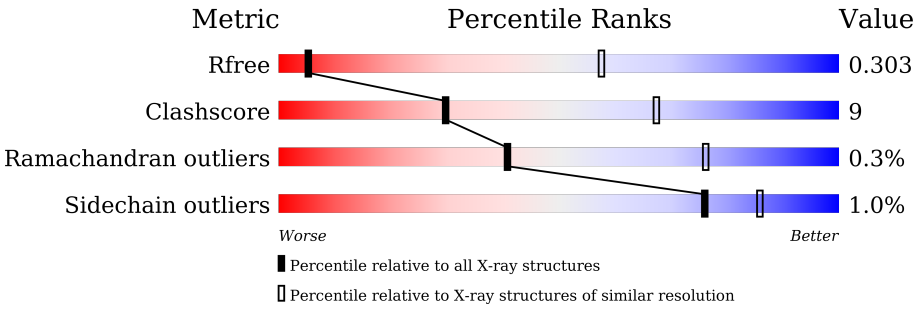
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 4.41 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	1046 (5.04-3.80)
Clashscore	141614	1114 (5.04-3.80)
Ramachandran outliers	138981	1061 (5.04-3.80)
Sidechain outliers	138945	1043 (5.04-3.80)

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 of 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	821	64% 21% • 14%
1	B	821	66% 19% 14%
1	C	821	67% 19% • 14%
1	D	821	68% 18% 14%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 22384 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 Tape tail measure protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	709	Total	C	N	O	Se	0	0	0
			5595	3533	979	1077	6			
1	B	709	Total	C	N	O	Se	0	0	0
			5595	3533	979	1077	6			
1	C	709	Total	C	N	O	Se	0	0	0
			5595	3533	979	1077	6			
1	D	709	Total	C	N	O	Se	0	0	0
			5595	3533	979	1077	6			

There are 92 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	304	MSE	-	initiating methionine	UNP A7XXD0
A	305	ALA	-	expression tag	UNP A7XXD0
A	306	HIS	-	expression tag	UNP A7XXD0
A	307	HIS	-	expression tag	UNP A7XXD0
A	308	HIS	-	expression tag	UNP A7XXD0
A	309	HIS	-	expression tag	UNP A7XXD0
A	310	HIS	-	expression tag	UNP A7XXD0
A	311	HIS	-	expression tag	UNP A7XXD0
A	312	SER	-	expression tag	UNP A7XXD0
A	313	ALA	-	expression tag	UNP A7XXD0
A	314	ALA	-	expression tag	UNP A7XXD0
A	315	LEU	-	expression tag	UNP A7XXD0
A	316	GLU	-	expression tag	UNP A7XXD0
A	317	VAL	-	expression tag	UNP A7XXD0
A	318	LEU	-	expression tag	UNP A7XXD0
A	319	PHE	-	expression tag	UNP A7XXD0
A	320	GLN	-	expression tag	UNP A7XXD0
A	321	GLY	-	expression tag	UNP A7XXD0
A	322	PRO	-	expression tag	UNP A7XXD0
A	323	GLY	-	expression tag	UNP A7XXD0
A	324	ARG	-	expression tag	UNP A7XXD0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	325	ASN	-	expression tag	UNP A7XXD0
A	326	ASN	-	expression tag	UNP A7XXD0
B	304	MSE	-	initiating methionine	UNP A7XXD0
B	305	ALA	-	expression tag	UNP A7XXD0
B	306	HIS	-	expression tag	UNP A7XXD0
B	307	HIS	-	expression tag	UNP A7XXD0
B	308	HIS	-	expression tag	UNP A7XXD0
B	309	HIS	-	expression tag	UNP A7XXD0
B	310	HIS	-	expression tag	UNP A7XXD0
B	311	HIS	-	expression tag	UNP A7XXD0
B	312	SER	-	expression tag	UNP A7XXD0
B	313	ALA	-	expression tag	UNP A7XXD0
B	314	ALA	-	expression tag	UNP A7XXD0
B	315	LEU	-	expression tag	UNP A7XXD0
B	316	GLU	-	expression tag	UNP A7XXD0
B	317	VAL	-	expression tag	UNP A7XXD0
B	318	LEU	-	expression tag	UNP A7XXD0
B	319	PHE	-	expression tag	UNP A7XXD0
B	320	GLN	-	expression tag	UNP A7XXD0
B	321	GLY	-	expression tag	UNP A7XXD0
B	322	PRO	-	expression tag	UNP A7XXD0
B	323	GLY	-	expression tag	UNP A7XXD0
B	324	ARG	-	expression tag	UNP A7XXD0
B	325	ASN	-	expression tag	UNP A7XXD0
B	326	ASN	-	expression tag	UNP A7XXD0
C	304	MSE	-	initiating methionine	UNP A7XXD0
C	305	ALA	-	expression tag	UNP A7XXD0
C	306	HIS	-	expression tag	UNP A7XXD0
C	307	HIS	-	expression tag	UNP A7XXD0
C	308	HIS	-	expression tag	UNP A7XXD0
C	309	HIS	-	expression tag	UNP A7XXD0
C	310	HIS	-	expression tag	UNP A7XXD0
C	311	HIS	-	expression tag	UNP A7XXD0
C	312	SER	-	expression tag	UNP A7XXD0
C	313	ALA	-	expression tag	UNP A7XXD0
C	314	ALA	-	expression tag	UNP A7XXD0
C	315	LEU	-	expression tag	UNP A7XXD0
C	316	GLU	-	expression tag	UNP A7XXD0
C	317	VAL	-	expression tag	UNP A7XXD0
C	318	LEU	-	expression tag	UNP A7XXD0
C	319	PHE	-	expression tag	UNP A7XXD0
C	320	GLN	-	expression tag	UNP A7XXD0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	321	GLY	-	expression tag	UNP A7XXD0
C	322	PRO	-	expression tag	UNP A7XXD0
C	323	GLY	-	expression tag	UNP A7XXD0
C	324	ARG	-	expression tag	UNP A7XXD0
C	325	ASN	-	expression tag	UNP A7XXD0
C	326	ASN	-	expression tag	UNP A7XXD0
D	304	MSE	-	initiating methionine	UNP A7XXD0
D	305	ALA	-	expression tag	UNP A7XXD0
D	306	HIS	-	expression tag	UNP A7XXD0
D	307	HIS	-	expression tag	UNP A7XXD0
D	308	HIS	-	expression tag	UNP A7XXD0
D	309	HIS	-	expression tag	UNP A7XXD0
D	310	HIS	-	expression tag	UNP A7XXD0
D	311	HIS	-	expression tag	UNP A7XXD0
D	312	SER	-	expression tag	UNP A7XXD0
D	313	ALA	-	expression tag	UNP A7XXD0
D	314	ALA	-	expression tag	UNP A7XXD0
D	315	LEU	-	expression tag	UNP A7XXD0
D	316	GLU	-	expression tag	UNP A7XXD0
D	317	VAL	-	expression tag	UNP A7XXD0
D	318	LEU	-	expression tag	UNP A7XXD0
D	319	PHE	-	expression tag	UNP A7XXD0
D	320	GLN	-	expression tag	UNP A7XXD0
D	321	GLY	-	expression tag	UNP A7XXD0
D	322	PRO	-	expression tag	UNP A7XXD0
D	323	GLY	-	expression tag	UNP A7XXD0
D	324	ARG	-	expression tag	UNP A7XXD0
D	325	ASN	-	expression tag	UNP A7XXD0
D	326	ASN	-	expression tag	UNP A7XXD0

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

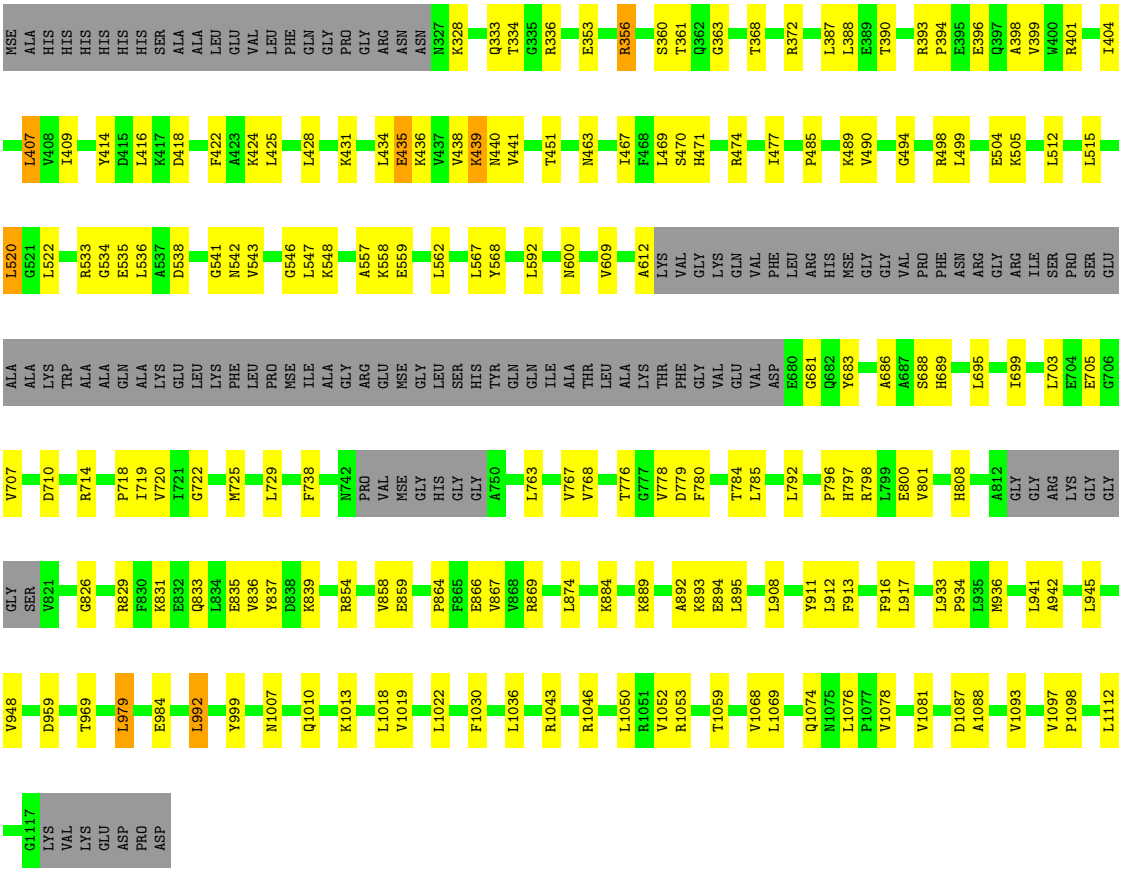
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Mg 1 1	0	0
2	B	1	Total Mg 1 1	0	0
2	C	1	Total Mg 1 1	0	0
2	D	1	Total Mg 1 1	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

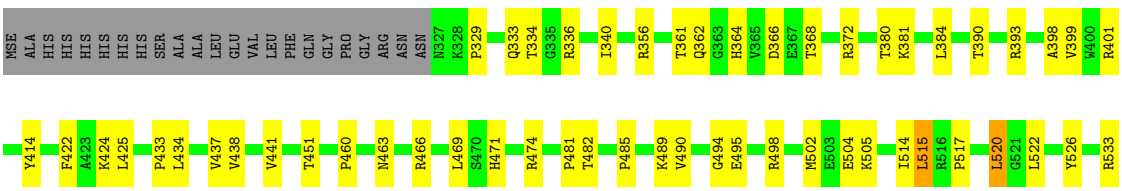
• Molecule 1: Tape tail measure protein

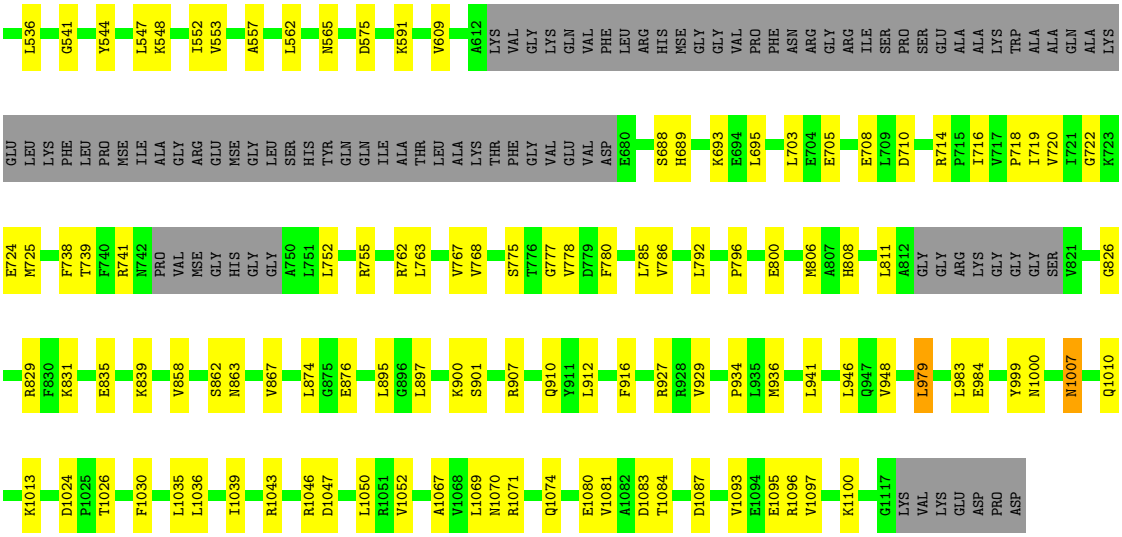
Chain A:



• Molecule 1: Tape tail measure protein

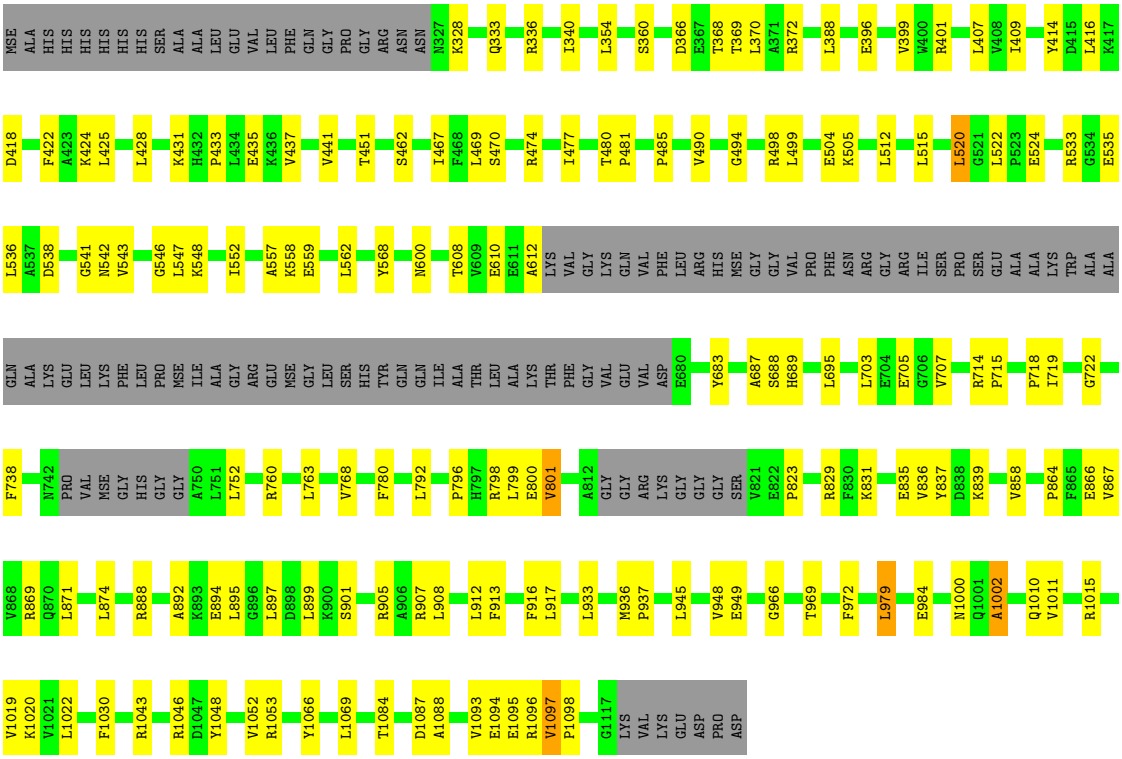
Chain B:





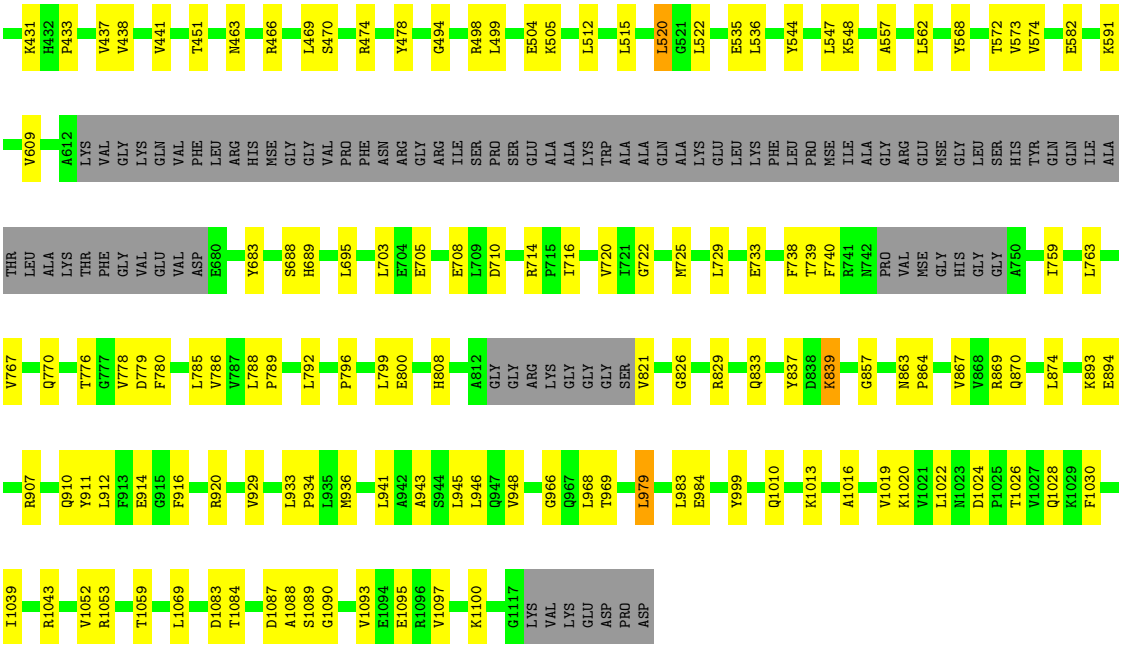
• Molecule 1: Tape tail measure protein

Chain C:



• Molecule 1: Tape tail measure protein

Chain D:


K431 H432 P433 V437 V438 V441 T451 N463 R466 L469 S470 R474 Y478 G494 R498 L499 E504 K505 L512 L515 L520 G521 L522 E535 L536 Y544 L547 K548 A557 L562 Y568 T572 V573 V574 E582 K591
V609 A612 LYS VAL GLY LYS GLN VAL PHE LEU ARG HIS MSE GLY GLY VAL PRO PHE ASN ARG GLY ARG ILE SER PRO SER GLU ALA ALA LYS TRP ALA ALA GLN ALA LYS GLU LEU LYS PHE LEU PRO MSE ILE ALA GLY ARG GLU MSE GLY LEU SER HIS TYR GLN GLN ILE ALA
THR LEU ALA LYS THR PHE GLY VAL GLU VAL ASP E680 Y683 S688 H689 L695 L703 E704 E705 E708 L709 D710 R714 P715 I716 V720 I721 G722 M725 L729 E733 F738 T739 F740 R741 N742 PRO VAL MSE GLY HIS GLY GLY A750 I759 L763
V767 Q770 T776 G777 V778 D779 F780 L785 V786 V787 L788 P789 L792 P796 L799 E800 H808 A812 GLY GLY ARG LYS GLY GLY GLY SER V821 G826 R829 Q833 Y837 D838 K839 G857 N863 P864 V867 V868 R869 Q870 L874 K893 E894
R907 Q910 Y911 L912 F913 E914 G915 F916 R920 V929 L933 P934 L935 M936 L941 A942 A943 S944 L945 L946 Q947 V948 G966 Q967 L968 T969 L979 L983 E984 Y999 Q1010 K1013 A1016 V1019 K1020 V1021 L1022 N1023 D1024 P1025 T1026 V1027 Q1028 K1029 F1030
I1039 R1043 V1052 R1053 T1059 L1069 D1083 T1084 D1087 A1088 S1089 G1090 V1093 E1094 E1095 R1096 V1097 K1100 G1117 LYS VAL LYS GLU ASP PRO ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	118.16Å 128.91Å 136.76Å 90.00° 96.84° 90.00°	Depositor
Resolution (Å)	44.37 – 4.41 47.95 – 4.41	Depositor EDS
% Data completeness (in resolution range)	99.2 (44.37-4.41) 82.0 (47.95-4.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.58 (at 4.45Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.245 , 0.303 0.245 , 0.303	Depositor DCC
R_{free} test set	1998 reflections (7.74%)	wwPDB-VP
Wilson B-factor (Å^2)	146.8	Xtriage
Anisotropy	0.140	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 115.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	22384	wwPDB-VP
Average B, all atoms (Å^2)	203.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.55% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: 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	0.27	0/5672	0.47	0/7644
1	B	0.27	0/5672	0.47	0/7644
1	C	0.27	0/5672	0.47	0/7644
1	D	0.26	0/5672	0.46	0/7644
All	All	0.26	0/22688	0.47	0/30576

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	5595	0	5684	120	0
1	B	5595	0	5684	121	0
1	C	5595	0	5684	103	0
1	D	5595	0	5684	94	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	22384	0	22736	426	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 9.

The worst 5 of 426 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:D:1010:GLN:HE22	1:D:1013:LYS:HE3	1.47	0.79
1:C:1002:ALA:HB1	1:C:1010:GLN:HG3	1.66	0.78
1:C:520:LEU:HD13	1:C:522:LEU:HD13	1.66	0.77
1:A:435:GLU:OE2	1:B:1007:ASN:ND2	2.17	0.77
1:A:547:LEU:HB3	1:A:567:LEU:HG	1.67	0.76

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	701/821 (85%)	645 (92%)	53 (8%)	3 (0%)	34 72
1	B	701/821 (85%)	650 (93%)	51 (7%)	0	100 100
1	C	701/821 (85%)	651 (93%)	46 (7%)	4 (1%)	25 65
1	D	701/821 (85%)	663 (95%)	36 (5%)	2 (0%)	41 76
All	All	2804/3284 (85%)	2609 (93%)	186 (7%)	9 (0%)	41 76

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	435	GLU
1	C	1002	ALA
1	C	1088	ALA
1	C	1097	VAL
1	D	1088	ALA

5.3.2 Protein sidechains (i)

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	597/670 (89%)	589 (99%)	8 (1%)	69 82
1	B	597/670 (89%)	593 (99%)	4 (1%)	84 90
1	C	597/670 (89%)	589 (99%)	8 (1%)	69 82
1	D	597/670 (89%)	592 (99%)	5 (1%)	81 89
All	All	2388/2680 (89%)	2363 (99%)	25 (1%)	76 86

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	418	ASP
1	C	520	LEU
1	D	979	LEU
1	C	515	LEU
1	C	907	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	967	GLN
1	B	910	GLN
1	C	1010	GLN
1	D	1010	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.