



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2023 – 02:26 PM EDT

PDB ID : 2PF4  
Title : Crystal structure of the full-length simian virus 40 small t antigen complexed with the protein phosphatase 2A Aalpha subunit  
Authors : Cho, U.; Morrone, S.; Xu, W.  
Deposited on : 2007-04-03  
Resolution : 3.10 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35  
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) : 2.35

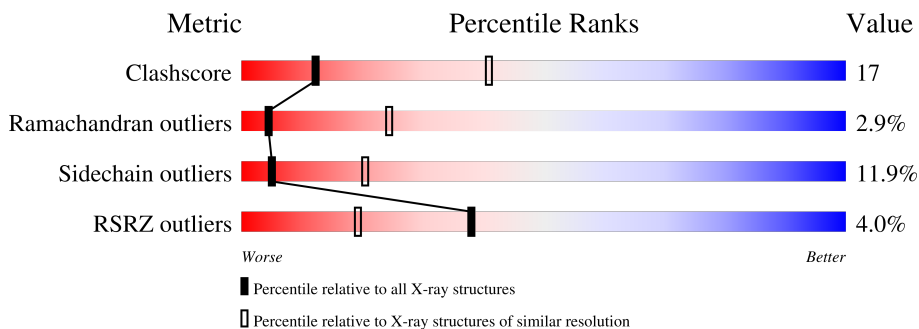
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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  $\geq 3$ , 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	589	
1	B	589	
1	C	589	
1	D	589	
2	E	174	
2	F	174	
2	G	174	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain												
2	H	174	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into segments of different colors: red (2%), green (41%), yellow (30%), orange (9%), and grey (18%). A small black dot is located on the orange segment.</p> <table border="1"><thead><tr><th>Color</th><th>Percentage</th></tr></thead><tbody><tr><td>Red</td><td>2%</td></tr><tr><td>Green</td><td>41%</td></tr><tr><td>Yellow</td><td>30%</td></tr><tr><td>Orange</td><td>9%</td></tr><tr><td>Grey</td><td>18%</td></tr></tbody></table>	Color	Percentage	Red	2%	Green	41%	Yellow	30%	Orange	9%	Grey	18%
Color	Percentage														
Red	2%														
Green	41%														
Yellow	30%														
Orange	9%														
Grey	18%														

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 22802 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 Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	575	Total 4472	C 2840	N 754	O 851	S 27	0	0	0
1	B	575	Total 4473	C 2840	N 754	O 852	S 27	0	0	0
1	C	577	Total 4486	C 2849	N 756	O 854	S 27	0	0	0
1	D	577	Total 4492	C 2852	N 759	O 854	S 27	0	0	0

- Molecule 2 is a protein called Small T antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	163	Total 1303	C 836	N 220	O 227	S 20	0	0	0
2	F	152	Total 1232	C 790	N 209	O 213	S 20	0	0	0
2	G	142	Total 1164	C 748	N 195	O 201	S 20	0	0	0
2	H	143	Total 1172	C 751	N 199	O 202	S 20	0	0	0

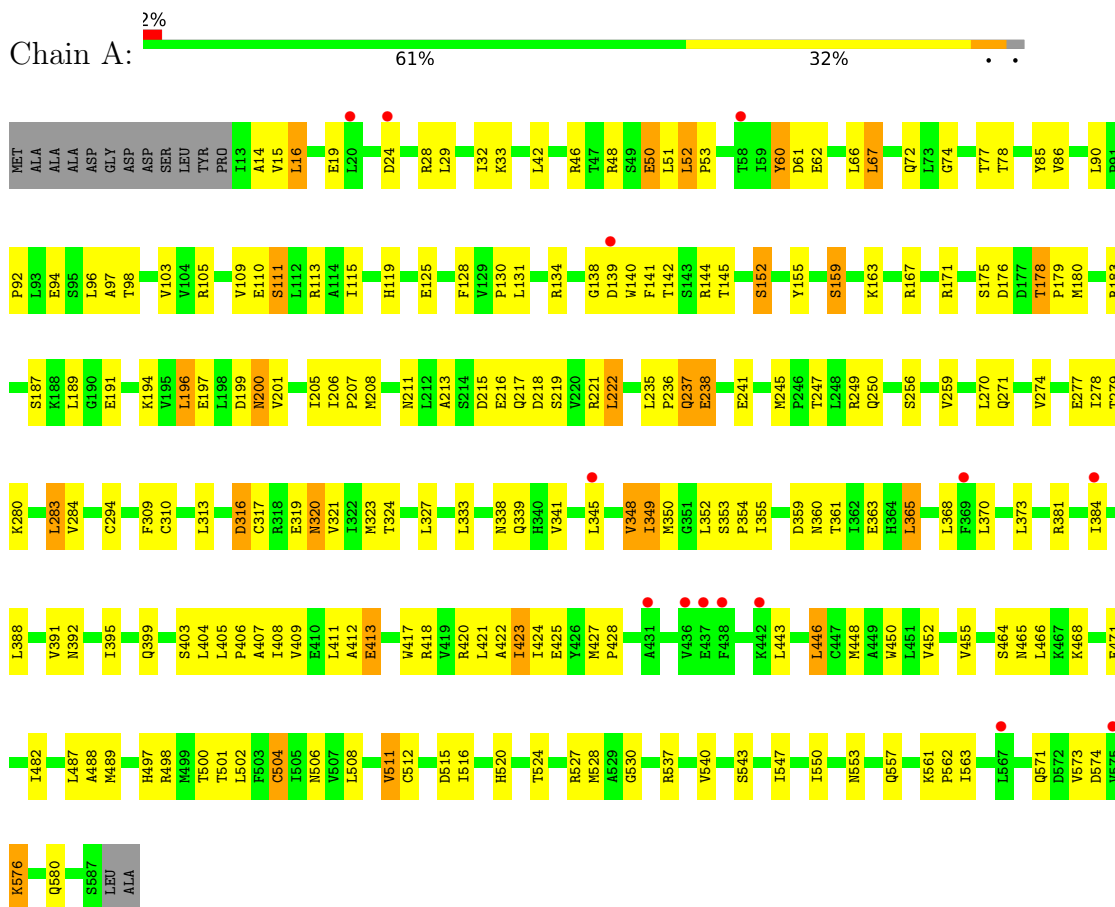
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	E	2	Total 2	Zn 2	0	0
3	F	2	Total 2	Zn 2	0	0
3	G	2	Total 2	Zn 2	0	0
3	H	2	Total 2	Zn 2	0	0

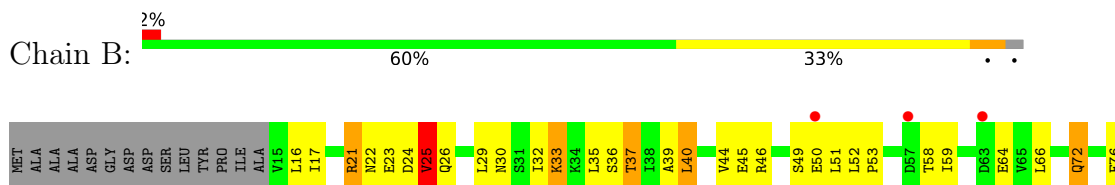
### 3 Residue-property plots

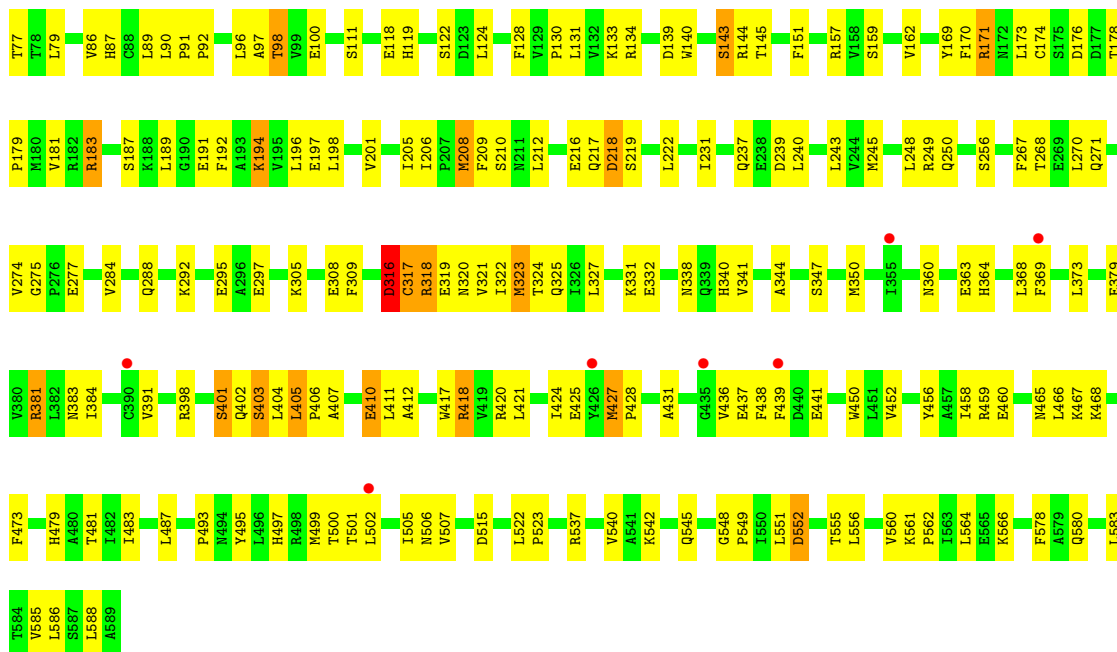
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform

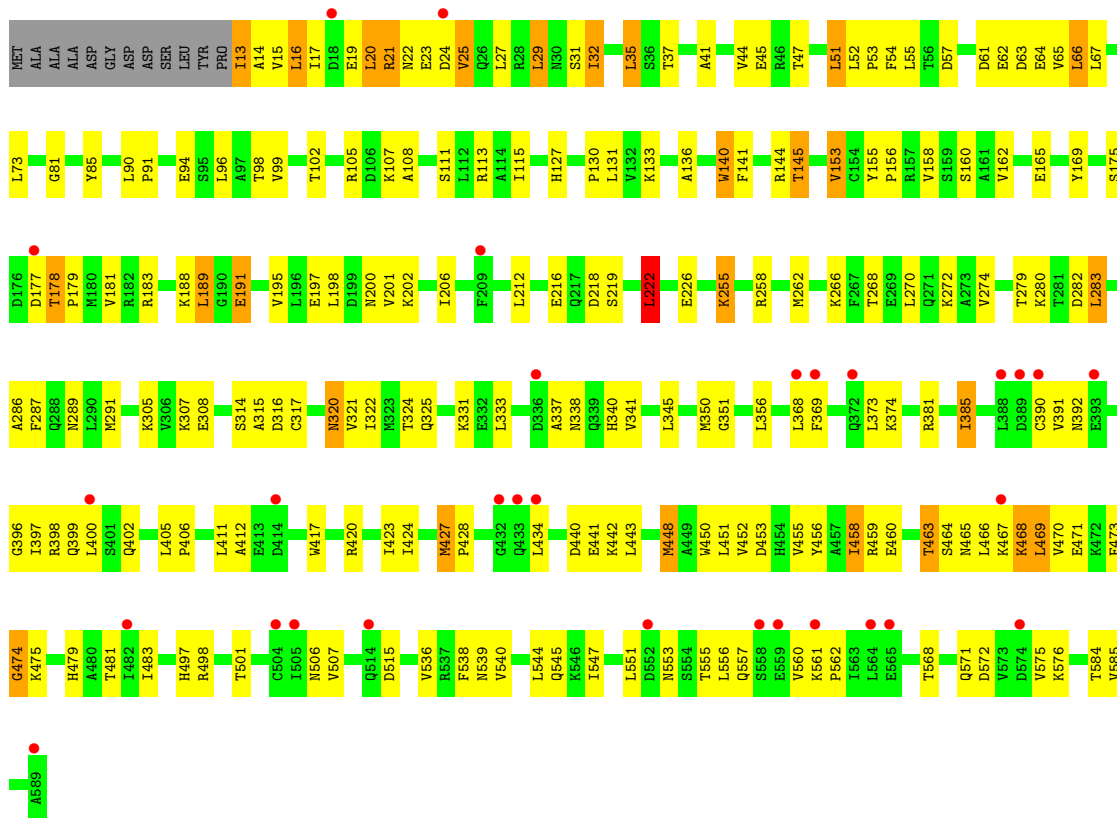


- Molecule 1: Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform

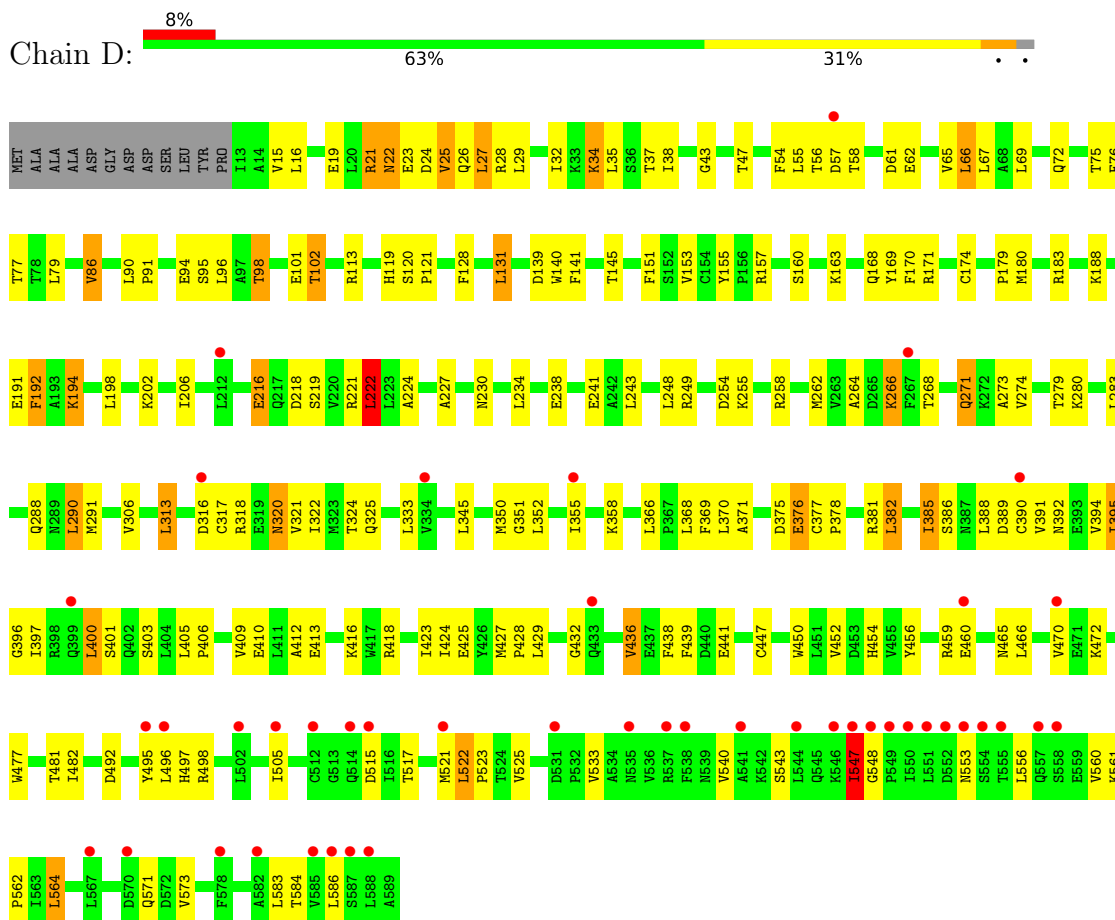




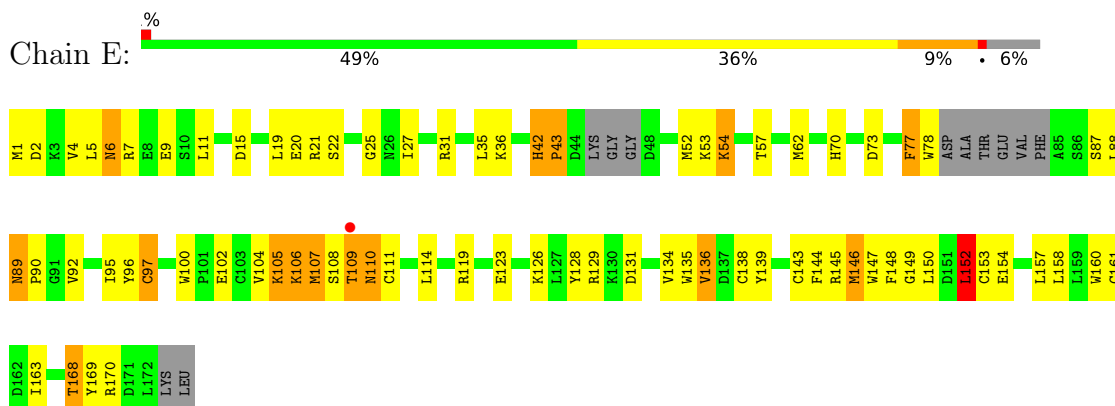
• Molecule 1: Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform



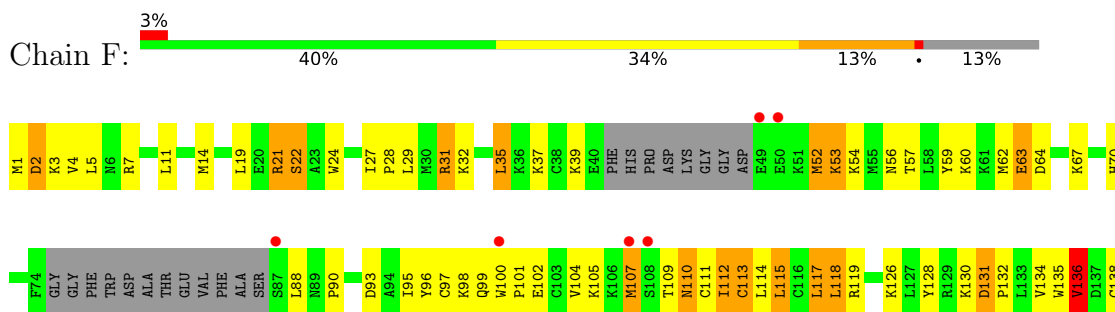
• Molecule 1: Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform



• Molecule 2: Small T antigen

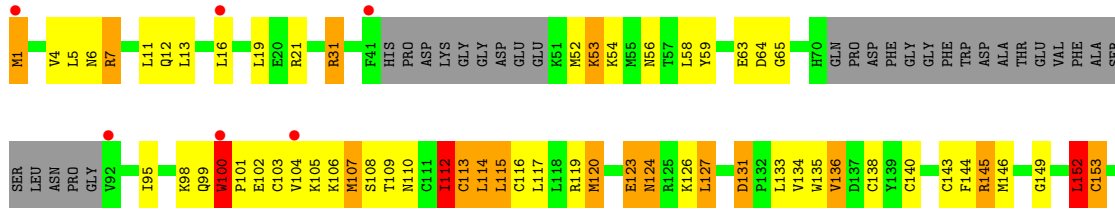


• Molecule 2: Small T antigen

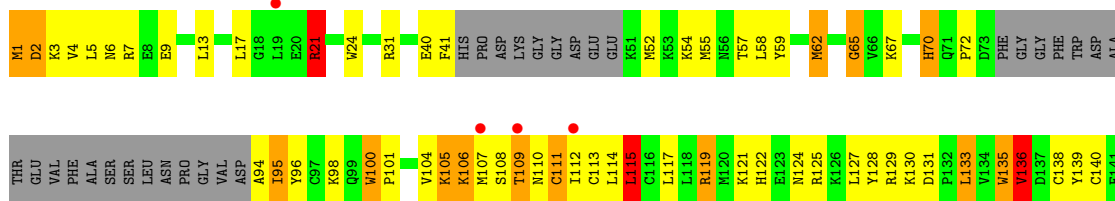




● Molecule 2: Small T antigen



● Molecule 2: Small T antigen





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	94.71Å 105.50Å 111.94Å 115.64° 109.55° 94.11°	Depositor
Resolution (Å)	20.00 – 3.10 47.13 – 3.11	Depositor EDS
% Data completeness (in resolution range)	97.4 (20.00-3.10) 87.3 (47.13-3.11)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.09 (at 3.12Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.244 , 0.304 0.237 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	99.6	Xtrriage
Anisotropy	0.155	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 118.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	22802	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	114.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.51% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical 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 i

### 5.1 Standard geometry i

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	0.46	0/4544	0.63	0/6170
1	B	0.50	0/4545	0.65	1/6170 (0.0%)
1	C	0.44	0/4558	0.61	2/6188 (0.0%)
1	D	0.44	0/4564	0.64	3/6195 (0.0%)
2	E	0.56	0/1332	0.73	0/1793
2	F	0.53	0/1257	0.77	0/1688
2	G	0.43	0/1186	0.71	2/1590 (0.1%)
2	H	0.49	0/1195	0.79	3/1602 (0.2%)
All	All	0.47	0/23181	0.66	11/31396 (0.0%)

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	B	0	2
1	D	0	2
2	E	0	1
2	H	0	1
All	All	0	6

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	115	LEU	CA-CB-CG	8.12	133.97	115.30
1	D	222	LEU	CA-CB-CG	7.28	132.04	115.30
1	D	313	LEU	CA-CB-CG	7.11	131.65	115.30
1	C	29	LEU	CA-CB-CG	6.63	130.55	115.30

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	G	133	LEU	CA-CB-CG	6.47	130.18	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	171	ARG	Sidechain
1	B	183	ARG	Sidechain
1	D	22	ASN	Peptide
1	D	436	VAL	Peptide
2	E	108	SER	Peptide

## 5.2 Too-close contacts [i](#)

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	4472	0	4567	140	0
1	B	4473	0	4567	178	0
1	C	4486	0	4583	145	0
1	D	4492	0	4594	133	0
2	E	1303	0	1265	60	0
2	F	1232	0	1222	94	0
2	G	1164	0	1168	43	0
2	H	1172	0	1170	60	0
3	E	2	0	0	0	0
3	F	2	0	0	0	0
3	G	2	0	0	0	0
3	H	2	0	0	0	0
All	All	22802	0	23136	779	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 17.

The worst 5 of 779 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)
2:F:31:ARG:NH1	2:F:31:ARG:HB3	1.70	1.07
1:D:24:ASP:HB2	1:D:28:ARG:HH21	1.23	1.03
1:D:24:ASP:HB2	1:D:28:ARG:NH2	1.77	0.99
1:B:405:LEU:HB2	1:B:406:PRO:HD3	1.45	0.99
1:B:456:TYR:HB2	2:H:128:TYR:O	1.64	0.97

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	573/589 (97%)	505 (88%)	53 (9%)	15 (3%)	5	26
1	B	573/589 (97%)	507 (88%)	57 (10%)	9 (2%)	9	37
1	C	575/589 (98%)	502 (87%)	65 (11%)	8 (1%)	11	40
1	D	575/589 (98%)	477 (83%)	85 (15%)	13 (2%)	6	28
2	E	157/174 (90%)	127 (81%)	20 (13%)	10 (6%)	1	8
2	F	146/174 (84%)	112 (77%)	24 (16%)	10 (7%)	1	7
2	G	136/174 (78%)	114 (84%)	17 (12%)	5 (4%)	3	19
2	H	137/174 (79%)	106 (77%)	18 (13%)	13 (10%)	0	3
All	All	2872/3052 (94%)	2450 (85%)	339 (12%)	83 (3%)	4	24

5 of 83 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	ALA
1	B	97	ALA
1	B	316	ASP
1	B	317	CYS
1	B	401	SER

### 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	501/512 (98%)	455 (91%)	46 (9%)	9	33
1	B	501/512 (98%)	448 (89%)	53 (11%)	6	26
1	C	502/512 (98%)	448 (89%)	54 (11%)	6	25
1	D	503/512 (98%)	454 (90%)	49 (10%)	8	30
2	E	138/157 (88%)	116 (84%)	22 (16%)	2	11
2	F	134/157 (85%)	110 (82%)	24 (18%)	2	8
2	G	128/157 (82%)	95 (74%)	33 (26%)	0	1
2	H	128/157 (82%)	107 (84%)	21 (16%)	2	10
All	All	2535/2676 (95%)	2233 (88%)	302 (12%)	5	20

5 of 302 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	105	LYS
2	H	58	LEU
2	F	118	LEU
2	G	107	MET
2	H	142	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 63 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	402	GLN
1	D	465	ASN
1	C	22	ASN
1	D	454	HIS
2	F	110	ASN

### 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 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 [i](#)

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

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	575/589 (97%)	0.03	14 (2%) 59 37	98, 114, 132, 146	0
1	B	575/589 (97%)	-0.01	10 (1%) 70 49	96, 112, 123, 132	0
1	C	577/589 (97%)	0.25	30 (5%) 27 12	98, 114, 131, 135	0
1	D	577/589 (97%)	0.32	45 (7%) 13 5	98, 113, 140, 146	0
2	E	163/174 (93%)	-0.21	1 (0%) 89 78	103, 113, 123, 133	0
2	F	152/174 (87%)	0.11	6 (3%) 39 20	103, 114, 127, 136	0
2	G	142/174 (81%)	0.02	6 (4%) 36 18	99, 114, 121, 125	0
2	H	143/174 (82%)	0.04	4 (2%) 53 30	96, 112, 121, 127	0
All	All	2904/3052 (95%)	0.11	116 (3%) 38 19	96, 113, 133, 146	0

The worst 5 of 116 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	589	ALA	12.7
2	F	49	GLU	10.8
1	D	586	LEU	8.3
1	D	557	GLN	7.5
1	D	550	ILE	6.8

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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, 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	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	ZN	E	601	1/1	0.87	0.16	103,103,103,103	0
3	ZN	E	602	1/1	0.92	0.20	99,99,99,99	0
3	ZN	H	901	1/1	0.94	0.13	101,101,101,101	0
3	ZN	G	801	1/1	0.96	0.16	108,108,108,108	0
3	ZN	F	702	1/1	0.96	0.16	100,100,100,100	0
3	ZN	G	802	1/1	0.98	0.16	98,98,98,98	0
3	ZN	F	701	1/1	0.98	0.13	108,108,108,108	0
3	ZN	H	902	1/1	0.98	0.24	98,98,98,98	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.