



# wwPDB X-ray Structure Validation Summary Report

Jun 16, 2024 – 03:29 PM EDT

PDB ID : 5DQT  
Title : Crystal Structure of Cas-DNA-22 complex  
Authors : Wang, J.; Li, J.; Zhao, H.; Sheng, G.; Wang, M.; Yin, M.; Wang, Y.  
Deposited on : 2015-09-15  
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 : 2.37.1  
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) : 2.37.1

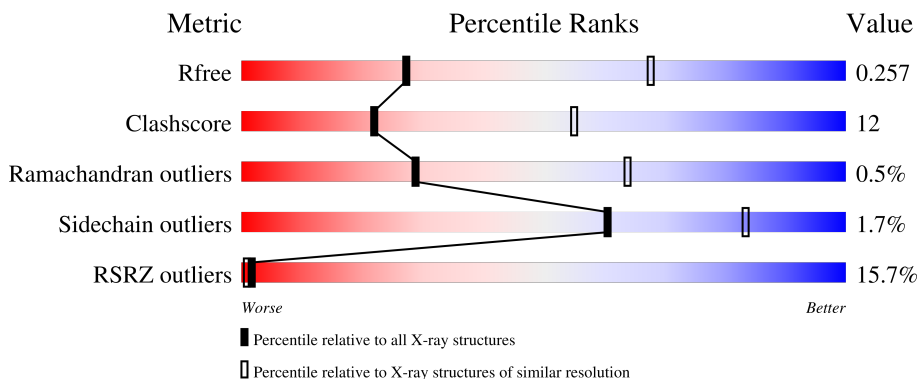
# 1 Overall quality at a glance

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))
$R_{free}$	130704	1094 (3.10-3.10)
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	305	 14% 65% 21% • 12%
1	B	305	 14% 72% 18% • 9%
1	C	305	 12% 70% 19% • 9%
1	D	305	 16% 60% 21% •• 16%
1	I	305	 20% 65% 22% 12%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	305	
1	K	305	
1	L	305	
2	E	94	
2	F	94	
2	M	94	
2	N	94	
3	G	34	
3	O	34	
4	H	33	
4	P	33	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 21956 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 CRISPR-associated endonuclease Cas1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	255	Total 1920	C 1230	N 336	O 347	S 7	0	0	0
1	C	278	Total 2076	C 1329	N 361	O 379	S 7	0	0	0
1	A	267	Total 2028	C 1298	N 356	O 367	S 7	0	0	0
1	B	279	Total 2128	C 1363	N 379	O 379	S 7	0	0	0
1	L	267	Total 1988	C 1272	N 351	O 358	S 7	0	0	0
1	K	278	Total 2111	C 1354	N 374	O 376	S 7	0	0	0
1	I	267	Total 2022	C 1296	N 355	O 364	S 7	0	0	0
1	J	280	Total 2121	C 1360	N 377	O 377	S 7	0	0	0

- Molecule 2 is a protein called CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	F	93	Total 732	C 470	N 127	O 131	S 4	0	0	0
2	E	93	Total 732	C 470	N 127	O 131	S 4	0	0	0
2	N	92	Total 722	C 464	N 126	O 128	S 4	0	0	0
2	M	92	Total 708	C 457	N 123	O 124	S 4	0	0	0

- Molecule 3 is a DNA chain called DNA (34-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	G	33	671	321	106	211	33	0	0	0
3	O	33	672	322	106	211	33	0	0	0

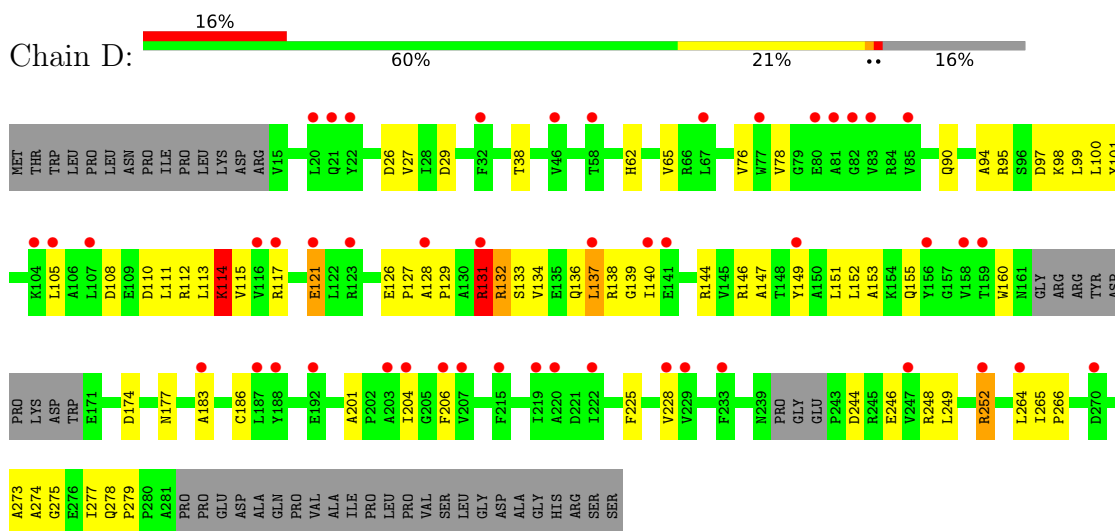
- Molecule 4 is a DNA chain called DNA (33-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	H	33	662	318	104	207	33	0	0	0
4	P	33	663	319	104	207	33	0	0	0

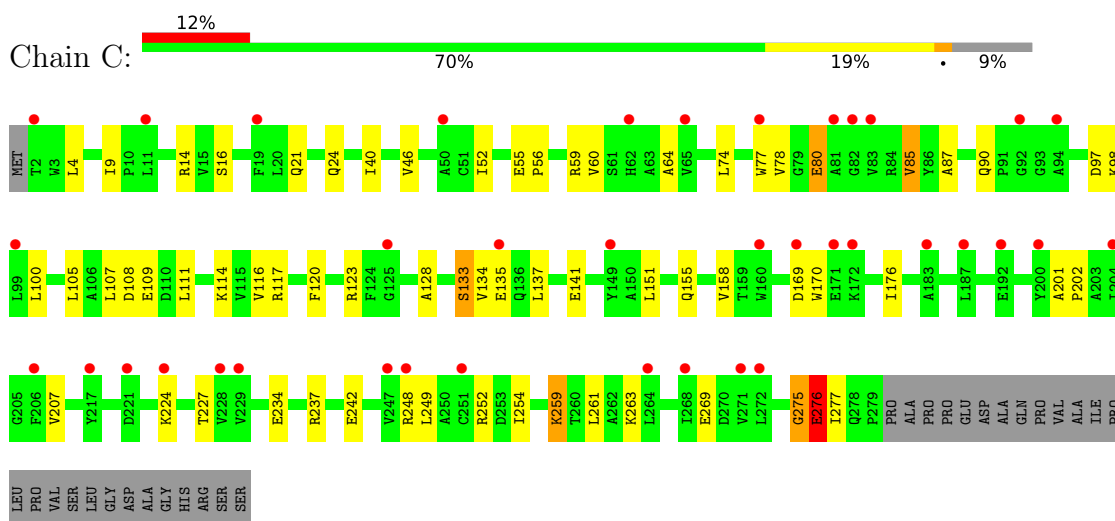
### 3 Residue-property plots [i](#)

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 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: CRISPR-associated endonuclease Cas1

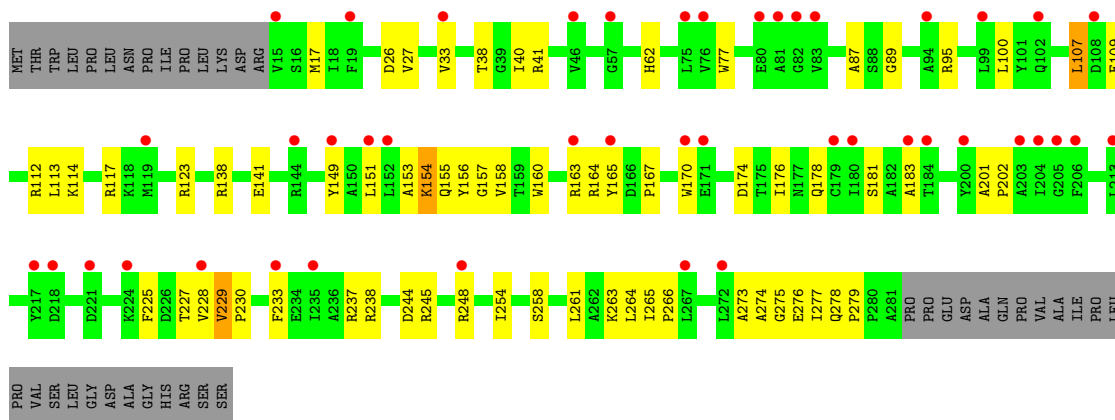


- Molecule 1: CRISPR-associated endonuclease Cas1

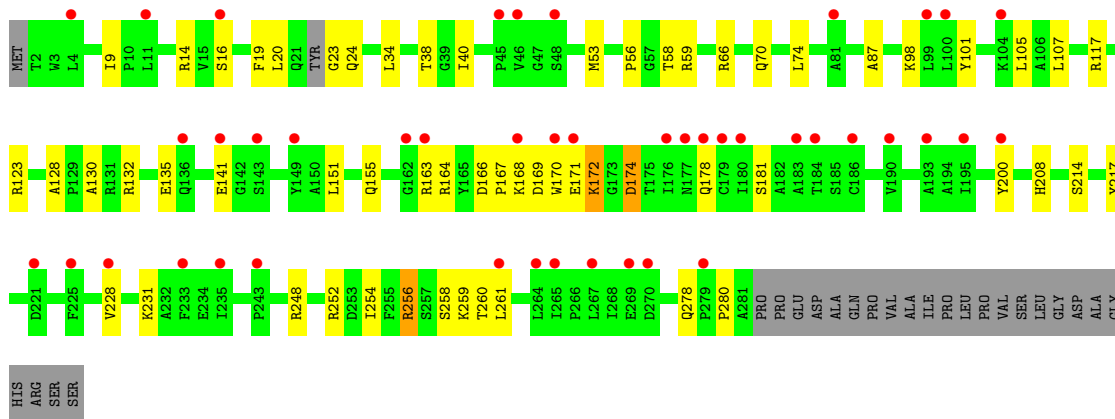
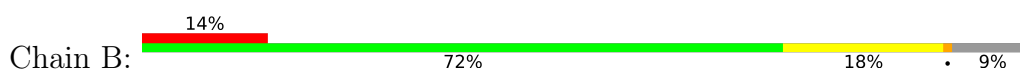


- Molecule 1: CRISPR-associated endonuclease Cas1

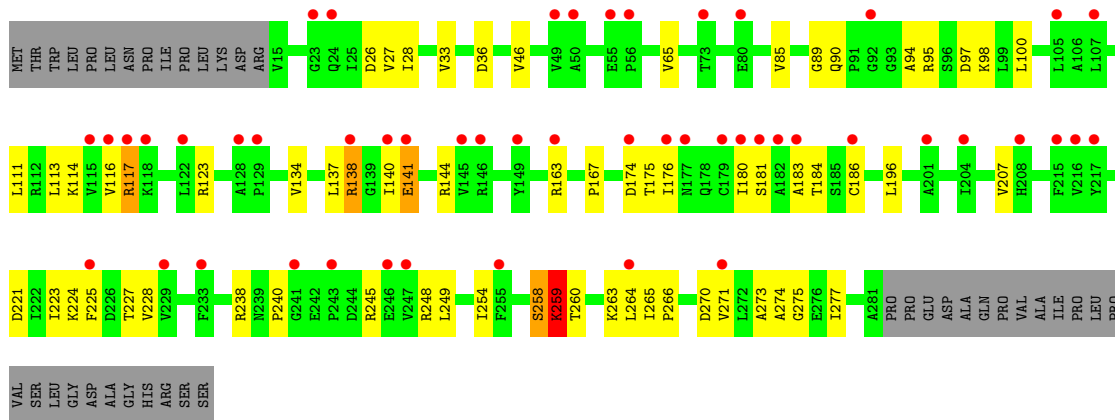




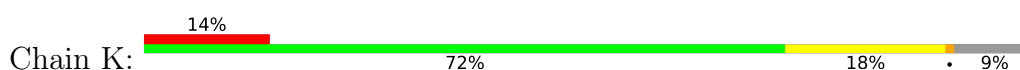
• Molecule 1: CRISPR-associated endonuclease Cas1

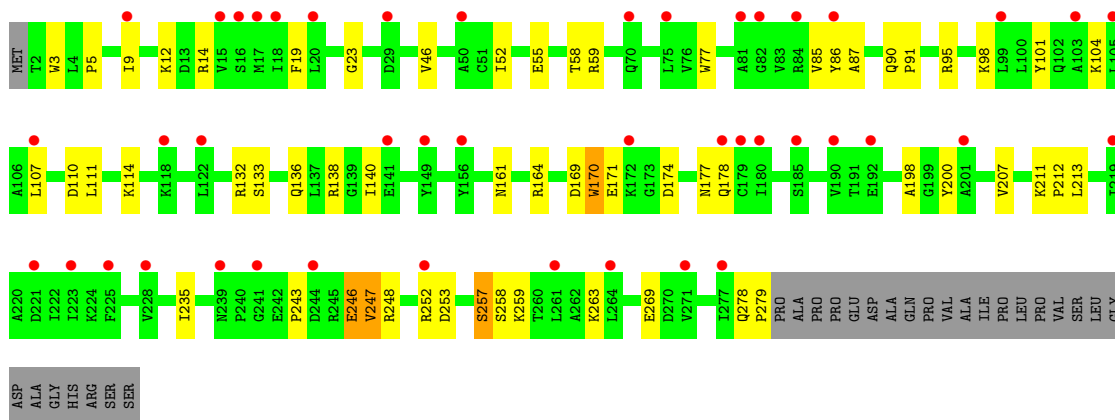


• Molecule 1: CRISPR-associated endonuclease Cas1

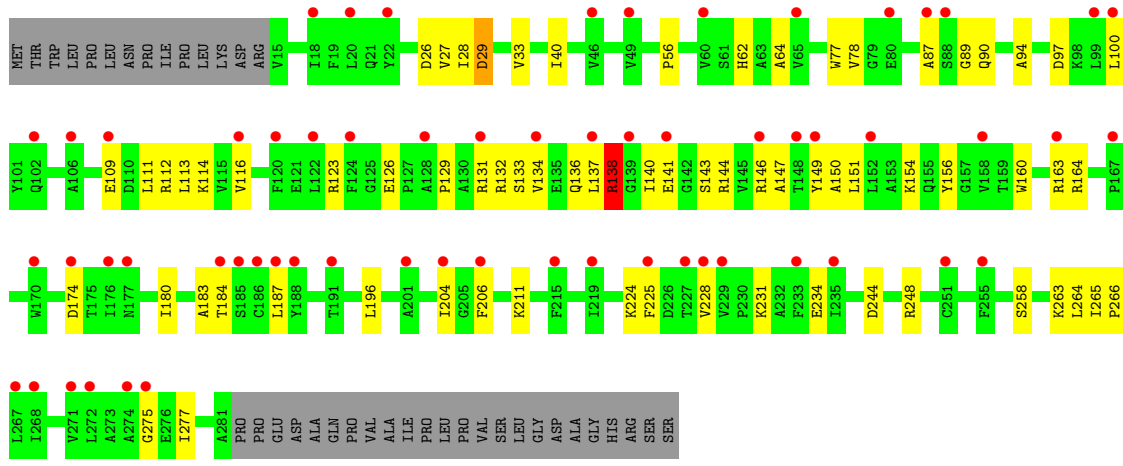


• Molecule 1: CRISPR-associated endonuclease Cas1

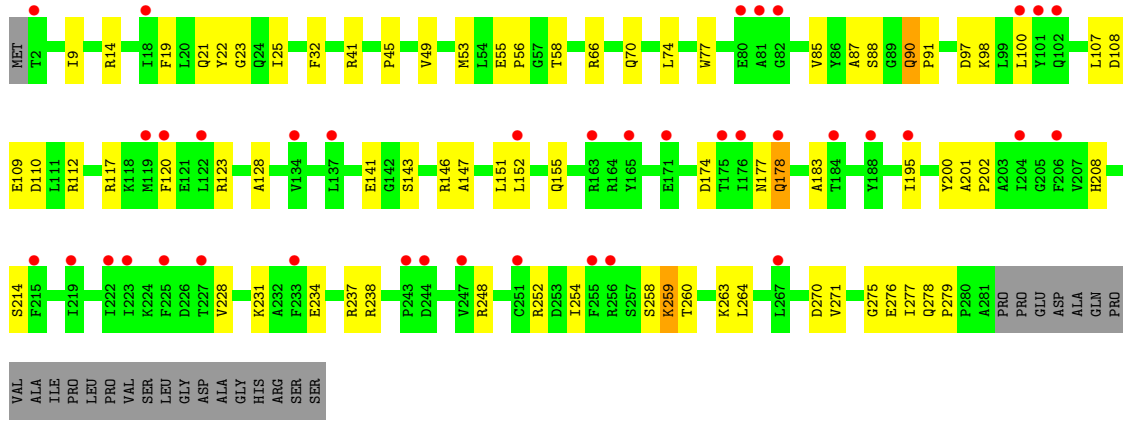




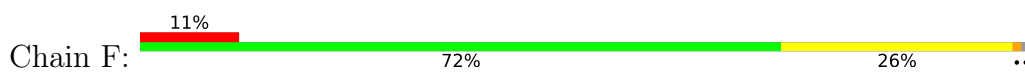
• Molecule 1: CRISPR-associated endonuclease Cas1



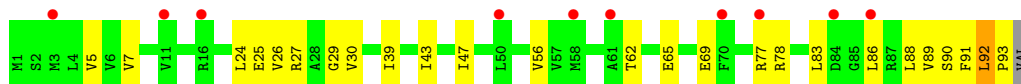
• Molecule 1: CRISPR-associated endonuclease Cas1



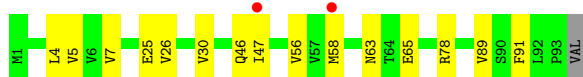
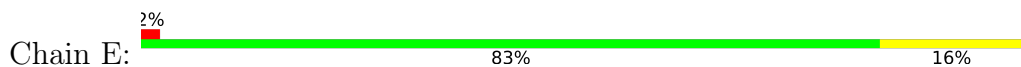
• Molecule 2: CRISPR-associated endoribonuclease Cas2



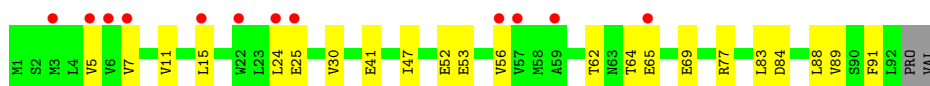
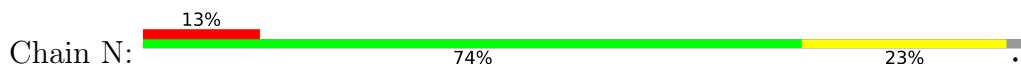




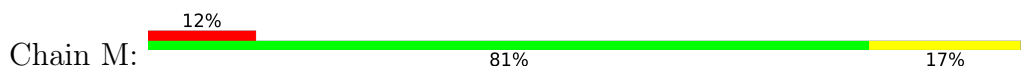
- Molecule 2: CRISPR-associated endoribonuclease Cas2



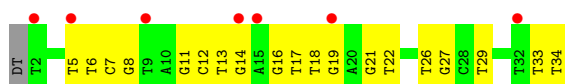
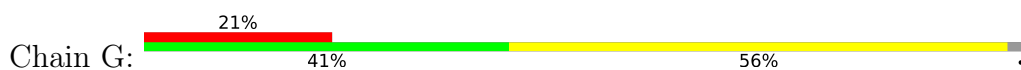
- Molecule 2: CRISPR-associated endoribonuclease Cas2



- Molecule 2: CRISPR-associated endoribonuclease Cas2



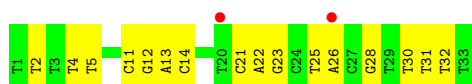
- Molecule 3: DNA (34-MER)



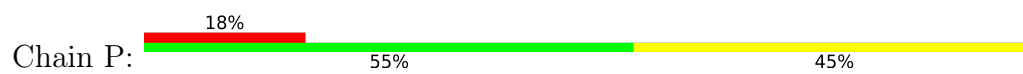
- Molecule 3: DNA (34-MER)



- Molecule 4: DNA (33-MER)



- Molecule 4: DNA (33-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.43Å 194.96Å 195.20Å 90.00° 90.22° 90.00°	Depositor
Resolution (Å)	48.80 – 3.10 49.68 – 3.07	Depositor EDS
% Data completeness (in resolution range)	93.2 (48.80-3.10) 92.2 (49.68-3.07)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.28 (at 3.07Å)	Xtrriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.211 , 0.258 0.208 , 0.257	Depositor DCC
$R_{free}$ test set	4578 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	98.4	Xtrriage
Anisotropy	0.167	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 84.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.368 for -h,l,k 0.387 for -h,-l,-k 0.408 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	21956	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	110.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.02% 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](#)

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	0.39	0/2068	0.66	1/2809 (0.0%)
1	B	0.42	0/2171	0.67	1/2950 (0.0%)
1	C	0.39	0/2115	0.65	4/2878 (0.1%)
1	D	0.47	0/1953	0.80	6/2650 (0.2%)
1	I	0.38	0/2062	0.70	2/2801 (0.1%)
1	J	0.37	0/2164	0.67	1/2943 (0.0%)
1	K	0.35	0/2154	0.60	0/2929
1	L	0.36	0/2025	0.68	2/2753 (0.1%)
2	E	0.37	0/746	0.62	0/1014
2	F	0.37	0/746	0.61	0/1014
2	M	0.34	0/721	0.59	0/981
2	N	0.35	0/735	0.60	0/998
3	G	0.69	0/748	1.10	0/1152
3	O	0.70	0/749	1.11	1/1154 (0.1%)
4	H	0.77	0/737	1.06	0/1132
4	P	0.70	0/738	1.04	0/1134
All	All	0.44	0/22632	0.74	18/31292 (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
1	D	0	1
1	L	0	1
All	All	0	3

There are no bond length outliers.

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	132	ARG	NE-CZ-NH2	-10.81	114.90	120.30
1	C	275	GLY	N-CA-C	7.98	133.06	113.10
1	D	114	LYS	CD-CE-NZ	7.62	129.23	111.70
1	B	164	ARG	NE-CZ-NH1	-7.04	116.78	120.30
1	D	137	LEU	CA-CB-CG	6.41	130.05	115.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	154	LYS	Peptide
1	D	131	ARG	Peptide
1	L	258	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	2028	0	2063	55	0
1	B	2128	0	2173	44	1
1	C	2076	0	2100	56	0
1	D	1920	0	1961	85	0
1	I	2022	0	2061	53	0
1	J	2121	0	2167	48	0
1	K	2111	0	2156	49	0
1	L	1988	0	2014	59	0
2	E	732	0	747	16	0
2	F	732	0	747	21	0
2	M	708	0	719	16	0
2	N	722	0	738	16	0
3	G	671	0	366	13	0
3	O	672	0	369	12	0
4	H	662	0	367	11	1
4	P	663	0	370	11	0
All	All	21956	0	21118	504	1

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 12.

The worst 5 of 504 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:K:211:LYS:HD2	1:K:212:PRO:HD2	1.33	1.09
1:L:138:ARG:HA	1:L:141:GLU:HG3	1.36	1.07
1:D:114:LYS:HG2	1:D:115:VAL:N	1.73	1.03
1:L:138:ARG:HE	1:L:207:VAL:HG11	1.25	0.96
1:B:166:ASP:HB3	1:B:172:LYS:HG3	1.48	0.93

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:217:TYR:OH	4:H:2:DT:OP1[2_846]	2.16	0.04

## 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	265/305 (87%)	256 (97%)	9 (3%)	0	100	100
1	B	275/305 (90%)	260 (94%)	14 (5%)	1 (0%)	34	69
1	C	276/305 (90%)	261 (95%)	11 (4%)	4 (1%)	11	40
1	D	249/305 (82%)	238 (96%)	10 (4%)	1 (0%)	34	69
1	I	265/305 (87%)	256 (97%)	9 (3%)	0	100	100
1	J	278/305 (91%)	268 (96%)	8 (3%)	2 (1%)	22	57
1	K	276/305 (90%)	267 (97%)	7 (2%)	2 (1%)	22	57
1	L	265/305 (87%)	256 (97%)	8 (3%)	1 (0%)	34	69
2	E	91/94 (97%)	89 (98%)	2 (2%)	0	100	100
2	F	91/94 (97%)	86 (94%)	4 (4%)	1 (1%)	14	46
2	M	90/94 (96%)	86 (96%)	4 (4%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	N	90/94 (96%)	87 (97%)	3 (3%)	0	100	100
All	All	2511/2816 (89%)	2410 (96%)	89 (4%)	12 (0%)	29	64

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	131	ARG
1	C	276	GLU
1	B	259	LYS
1	L	259	LYS
1	K	246	GLU

### 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	207/245 (84%)	204 (99%)	3 (1%)	67	86
1	B	218/245 (89%)	214 (98%)	4 (2%)	59	82
1	C	211/245 (86%)	205 (97%)	6 (3%)	43	73
1	D	196/245 (80%)	191 (97%)	5 (3%)	46	74
1	I	206/245 (84%)	202 (98%)	4 (2%)	57	81
1	J	216/245 (88%)	211 (98%)	5 (2%)	50	77
1	K	216/245 (88%)	213 (99%)	3 (1%)	67	86
1	L	199/245 (81%)	195 (98%)	4 (2%)	55	80
2	E	78/79 (99%)	78 (100%)	0	100	100
2	F	78/79 (99%)	78 (100%)	0	100	100
2	M	73/79 (92%)	73 (100%)	0	100	100
2	N	76/79 (96%)	76 (100%)	0	100	100
All	All	1974/2276 (87%)	1940 (98%)	34 (2%)	60	83

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	211	LYS
1	J	90	GLN
1	J	178	GLN
1	A	107	LEU
1	A	95	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	155	GLN
1	B	70	GLN
1	B	178	GLN
1	K	178	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](#)

There are no ligands in this entry.

## 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	267/305 (87%)	0.82	44 (16%) 1 1	63, 109, 167, 194	0
1	B	279/305 (91%)	0.87	44 (15%) 2 1	61, 98, 160, 201	0
1	C	278/305 (91%)	0.81	38 (13%) 3 1	67, 101, 159, 203	0
1	D	255/305 (83%)	0.93	48 (18%) 1 0	64, 111, 161, 208	0
1	I	267/305 (87%)	1.09	61 (22%) 0 0	64, 108, 167, 209	0
1	J	280/305 (91%)	0.84	39 (13%) 2 1	63, 100, 153, 176	0
1	K	278/305 (91%)	0.88	44 (15%) 2 1	66, 100, 157, 194	0
1	L	267/305 (87%)	0.95	50 (18%) 1 0	64, 113, 159, 187	0
2	E	93/94 (98%)	0.42	2 (2%) 62 41	70, 92, 120, 148	0
2	F	93/94 (98%)	0.74	10 (10%) 5 2	67, 91, 129, 168	0
2	M	92/94 (97%)	0.78	11 (11%) 4 2	62, 91, 127, 166	0
2	N	92/94 (97%)	0.88	12 (13%) 3 1	64, 92, 128, 135	0
3	G	33/34 (97%)	1.13	7 (21%) 0 0	88, 116, 153, 155	0
3	O	33/34 (97%)	0.49	1 (3%) 50 27	91, 118, 151, 166	0
4	H	33/33 (100%)	0.69	2 (6%) 21 9	81, 115, 149, 153	0
4	P	33/33 (100%)	1.17	6 (18%) 1 0	87, 116, 147, 161	0
All	All	2673/2950 (90%)	0.87	419 (15%) 2 1	61, 104, 159, 209	0

The worst 5 of 419 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	82	GLY	9.3
1	B	163	ARG	8.8
1	J	81	ALA	8.2
1	D	149	TYR	8.0
1	C	81	ALA	7.8

## 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.