



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 02:54 pm BST

PDB ID : 5YVG
Title : Crystal structure of Karyopherin beta2 in complex with FUS(full length)
Authors : Yoshizawa, T.; Fung, H.Y.J.; Chook, Y.M.
Deposited on : 2017-11-25
Resolution : 4.05 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

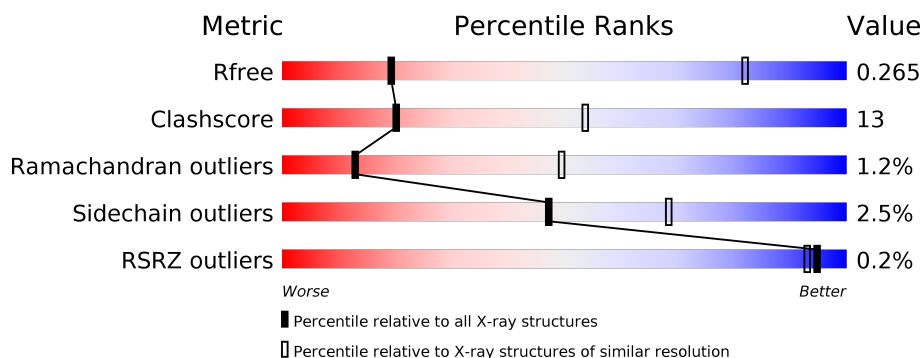
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.05 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	1127 (4.42-3.70)
Clashscore	141614	1033 (4.40-3.72)
Ramachandran outliers	138981	1145 (4.42-3.70)
Sidechain outliers	138945	1133 (4.42-3.70)
RSRZ outliers	127900	1005 (4.44-3.68)

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	868	
1	B	868	
2	X	528	
2	Y	528	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 12966 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 Transportin-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	830	Total	C	N	O	S	0	0	0
			6620	4248	1100	1220	52			
1	B	764	Total	C	N	O	S	0	0	0
			6057	3904	1001	1104	48			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q92973
A	0	SER	-	expression tag	UNP Q92973
A	361	GLY	-	linker	UNP Q92973
A	362	GLY	-	linker	UNP Q92973
A	363	SER	-	linker	UNP Q92973
A	364	GLY	-	linker	UNP Q92973
A	365	GLY	-	linker	UNP Q92973
A	366	SER	-	linker	UNP Q92973
A	367	GLY	-	linker	UNP Q92973
B	-1	GLY	-	expression tag	UNP Q92973
B	0	SER	-	expression tag	UNP Q92973
B	361	GLY	-	linker	UNP Q92973
B	362	GLY	-	linker	UNP Q92973
B	363	SER	-	linker	UNP Q92973
B	364	GLY	-	linker	UNP Q92973
B	365	GLY	-	linker	UNP Q92973
B	366	SER	-	linker	UNP Q92973
B	367	GLY	-	linker	UNP Q92973

- Molecule 2 is a protein called RNA-binding protein FUS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	18	Total	C	N	O	S	0	0	0
			159	91	37	30	1			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	Y	14	Total	C	N	O	0	0	0
			130	74	32	24			

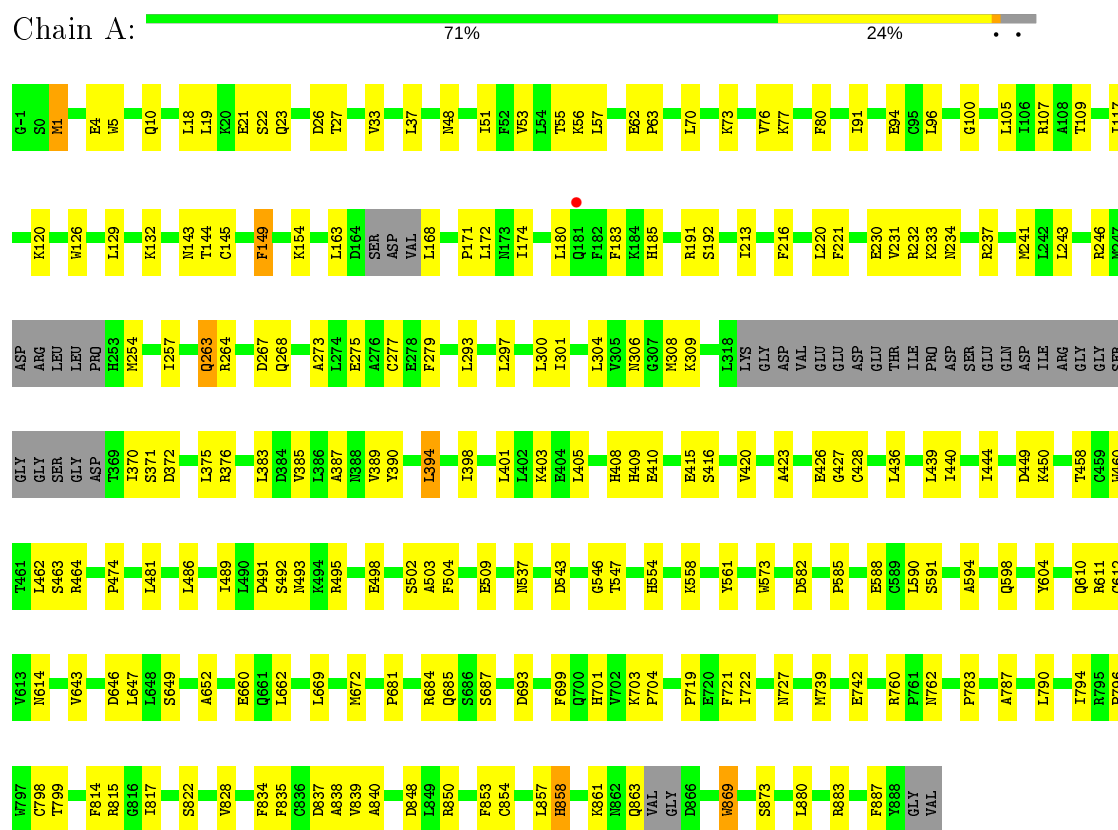
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	-1	GLY	-	expression tag	UNP P35637
X	0	SER	-	expression tag	UNP P35637
Y	-1	GLY	-	expression tag	UNP P35637
Y	0	SER	-	expression tag	UNP P35637

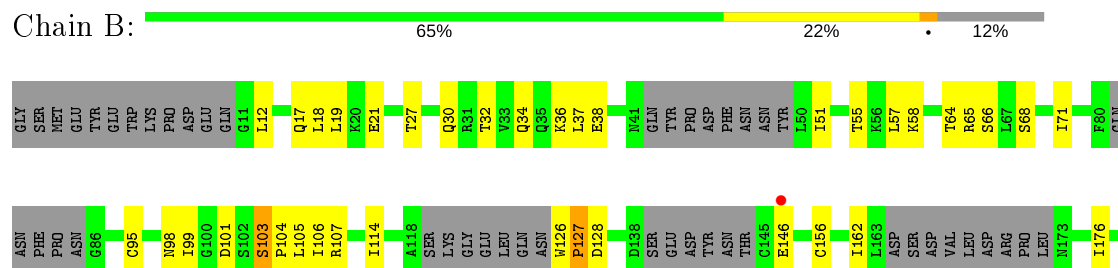
3 Residue-property plots

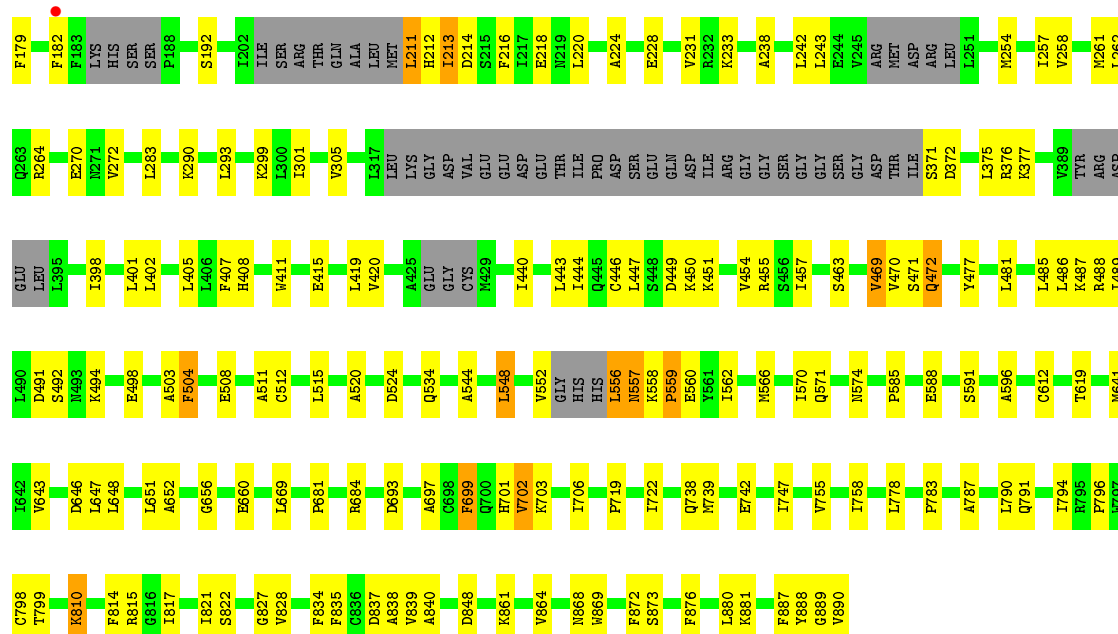
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Transportin-1



• Molecule 1: Transportin-1

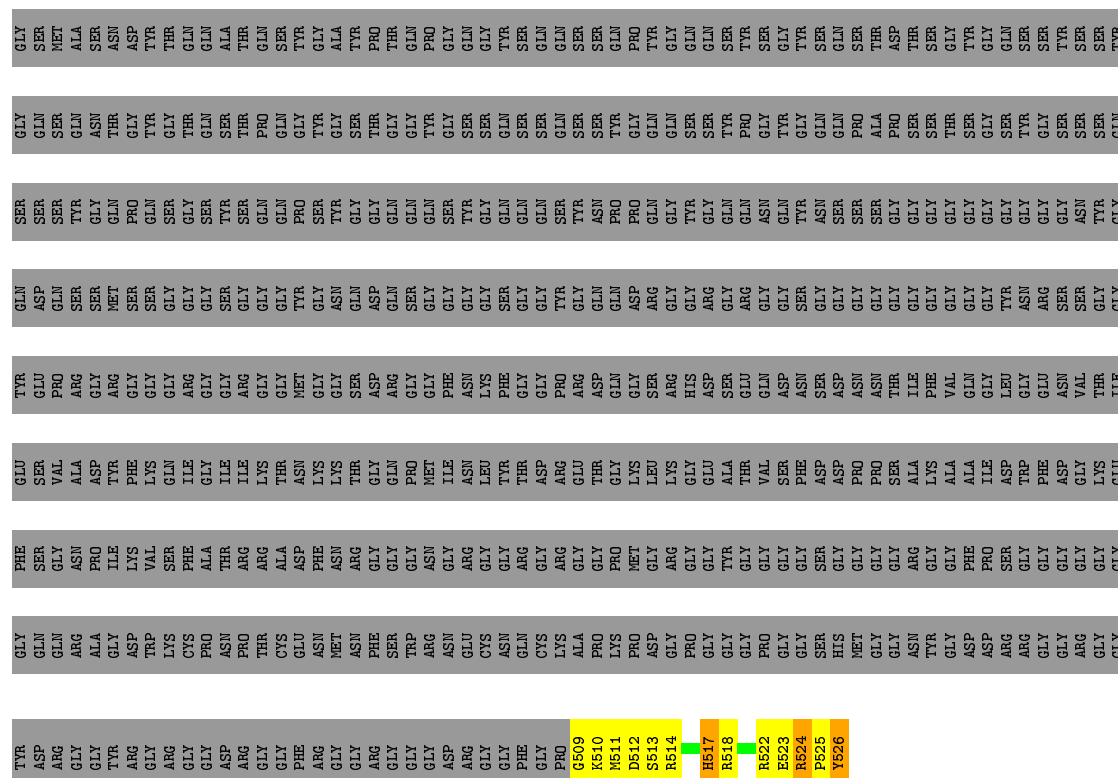




• Molecule 2: RNA-binding protein FUS

Chain X: ...

97%



• Molecule 2: RNA-binding protein FUS

Chain Y: ..

97%

Tyr	Gly	Phe	Glu	Tyr	Gln	Ser	Gly	Ser	Gln	Ser	Gly	Gly
Asp	Gln	Ser	Ser	Glu	Asp	Ser	Ser	Ser	Ser	Ser	Ser	Ser
Arg	Gln	Gly	Val	Pro	Gln	Ser	Ser	Tyr	Gln	Gln	Ala	Ala
Gly	Ala	Asn	Ala	Arg	Ser	Ser	Ser	Gly	Asn	Thr	Ser	Ser
Gly	Ala	Pro	Asp	Gly	Met	Ser	Met	Gln	Gln	Thr	Asn	Asn
Tyr	Gly	Ile	Tyr	Arg	Arg	Arg	Arg	Gly	Gln	Ser	Thr	Asp
Arg	Asp	Lys	Phe	Lys	Lys	Gly	Ser	Pro	Gly	Tyr	Thr	Thr
Gly	Trp	Val	Val	Lys	Lys	Gly	Ser	Gly	Tyr	Thr	Gly	Gly
Arg	Lys	Ser	Ser	Gly	Gly	Gly	Gly	Gly	Gly	Gln	Gln	Gln
Gly	Cys	Phe	Ile	Arg	Gly	Gly	Ser	Ser	Ser	Ser	Ser	Thr
Gly	Pro	Ala	Gly	Gly	Gly	Gly	Gly	Tyr	Gln	Gln	Gln	Gln
Asp	Asn	Thr	Ile	Gly	Ser	Tyr	Tyr	Tyr	Ser	Ser	Ala	Ala
Pro	Pro	Arg	Ile	Arg	Gly	Gly	Gly	Ser	Ser	Ser	Thr	Thr
Gly	Thr	Arg	Lys	Gly	Gly	Gly	Gly	Gln	Gln	Gln	Pro	Gln
Gly	Cys	Ala	Thr	Gly	Asn	Met	Tyr	Pro	Gly	Gln	Thr	Ser
Phe	Glu	Asp	Asn	Met	Lys	Lys	Gly	Gly	Ser	Ser	Ala	Asp
Arg	Asn	Phe	Lys	Lys	Lys	Gly	Ser	Gly	Gly	Gly	Pro	Thr
Gly	Met	Asn	Lys	Gly	Asn	Tyr	Tyr	Tyr	Gln	Gln	Thr	Asp
Gly	Asn	Arg	Thr	Ile	Gly	Gly	Gly	Gln	Gln	Gln	Ala	Ala
Gly	Asn	Arg	Ile	Arg	Ser	Ser	Gln	Ser	Ser	Ser	Thr	Thr
Arg	Phe	Gly	Gly	Asp	Asp	Arg	Gly	Gly	Gln	Gln	Pro	Pro
Gly	Ser	Gly	Gln	Arg	Gln	Gln	Gln	Gln	Gln	Gln	Ser	Ser
Gly	Trp	Gly	Pro	Pro	Ser	Gly	Gly	Pro	Pro	Pro	Thr	Thr
Gly	Arg	Asn	Met	Lys	Asp	Arg	Arg	Pro	Pro	Gln	Thr	Asp
Asp	Asn	Gly	Lys	Lys	Lys	Leu	Ser	Gly	Asp	Gln	Gly	Gly
Arg	Glu	Arg	Arg	Arg	Arg	Leu	Arg	Gly	Gly	Gln	Thr	Thr
Gly	Glu	Arg	Glu	Asp	Gly	Leu	His	Tyr	Tyr	Ser	Tyr	Tyr
Gly	Cys	Gly	Glu	Lys	Gly	Lys	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly	Gly	Gln	Gln	Gln
Gly	Asn	Gly	Leu	Asp	Gly	Leu	Arg	Gly				

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	108.22Å 145.01Å 108.68Å 90.00° 94.55° 90.00°	Depositor
Resolution (Å)	41.33 – 4.05 41.33 – 4.01	Depositor EDS
% Data completeness (in resolution range)	87.4 (41.33-4.05) 85.4 (41.33-4.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.11 (at 4.00Å)	Xtriage
Refinement program	phenix.refine 1.13_2998, PHENIX 1.13_2998	Depositor
R, R_{free}	0.251 , 0.260 0.262 , 0.265	Depositor DCC
R_{free} test set	1189 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	38.6	Xtriage
Anisotropy	0.593	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , -6.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.278 for l,-k,h	Xtriage
Reported twinning fraction	0.360 for l,-k,h	Depositor
Outliers	0 of 23952 reflections	Xtriage
F_o, F_c correlation	0.79	EDS
Total number of atoms	12966	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.87% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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	0.27	0/6760	0.42	0/9176
1	B	0.27	0/6174	0.43	0/8370
2	X	0.28	0/161	0.50	0/209
2	Y	0.23	0/132	0.38	0/172
All	All	0.27	0/13227	0.42	0/17927

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 [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	6620	0	6675	146	0
1	B	6057	0	6164	170	0
2	X	159	0	148	33	0
2	Y	130	0	119	7	0
All	All	12966	0	13106	327	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 13.

The worst 5 of 327 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:B:103:SER:HB2	1:B:104:PRO:HD2	1.19	1.16
1:A:854:CYS:O	1:A:858:HIS:HB2	1.50	1.11
1:B:508:GLU:HG3	1:B:548:LEU:HD12	1.25	1.10
2:Y:516:GLU:O	2:Y:520:ASP:HB2	1.57	1.05
1:B:512:CYS:O	1:B:552:VAL:HG22	1.56	1.04

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	820/868 (94%)	753 (92%)	64 (8%)	3 (0%)	34	71
1	B	738/868 (85%)	675 (92%)	50 (7%)	13 (2%)	8	41
2	X	16/528 (3%)	10 (62%)	4 (25%)	2 (12%)	0	5
2	Y	12/528 (2%)	9 (75%)	2 (17%)	1 (8%)	1	13
All	All	1586/2792 (57%)	1447 (91%)	120 (8%)	19 (1%)	13	49

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	470	VAL
1	B	557	ASN
2	X	511	MET
1	B	127	PRO
1	B	162	ILE

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	746/776 (96%)	731 (98%)	15 (2%)	55	73
1	B	683/776 (88%)	669 (98%)	14 (2%)	55	73
2	X	16/362 (4%)	13 (81%)	3 (19%)	1	10
2	Y	13/362 (4%)	8 (62%)	5 (38%)	0	0
All	All	1458/2276 (64%)	1421 (98%)	37 (2%)	47	68

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	156	CYS
1	B	504	PHE
2	Y	520	ASP
1	B	211	LEU
1	B	213	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	701	HIS
1	B	557	ASN
1	A	727	ASN
1	A	685	GLN
1	B	296	HIS

5.3.3 RNA ⓘ

There are no RNA molecules 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 [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, 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	830/868 (95%)	-0.37	1 (0%) 95 95	4, 33, 70, 97	0
1	B	764/868 (88%)	-0.21	2 (0%) 94 90	6, 59, 109, 137	0
2	X	18/528 (3%)	-0.31	0 100 100	14, 24, 32, 32	0
2	Y	14/528 (2%)	0.14	0 100 100	40, 59, 77, 78	0
All	All	1626/2792 (58%)	-0.29	3 (0%) 95 93	4, 38, 101, 137	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	146	GLU	2.4
1	A	181	GLN	2.3
1	B	182	PHE	2.2

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 carbohydrates 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.