wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2021 – 07:52 AM EDT

PDB ID	:	3BTU
Title	:	Crystal structure of the super-repressor mutant of Gal80p from Saccharomyces cerevisiae; Gal80(S2) [E351K]
Authors	:	Kumar, P.R.; Joshua-Tor, L.
Deposited on	:	2007-12-30
Resolution	:	2.85 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 following versions of software and data (see references (1)) were used in the production of this report:

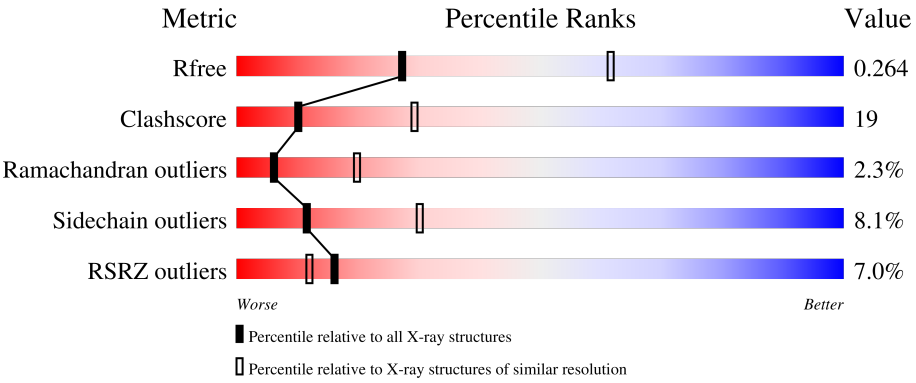
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.85 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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% 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	438	2% 57% 28% • 11%
1	B	438	2% 58% 26% • 11%
1	C	438	3% 57% 27% • 11%
1	D	438	% 59% 25% 5% 11%
1	E	438	11% 46% 37% 5% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	438	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 18389 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 Galactose/lactose metabolism regulatory protein GAL80.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	392	Total	C	N	O	S	0	0	0
			3088	1987	518	572	11			
1	B	389	Total	C	N	O	S	0	0	0
			3065	1975	513	566	11			
1	C	388	Total	C	N	O	S	0	0	0
			3057	1969	511	566	11			
1	D	392	Total	C	N	O	S	0	0	0
			3087	1987	518	571	11			
1	E	387	Total	C	N	O	S	0	0	0
			3046	1960	512	563	11			
1	F	387	Total	C	N	O	S	0	0	0
			3046	1960	512	563	11			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P04387
A	-1	SER	-	expression tag	UNP P04387
A	0	HIS	-	expression tag	UNP P04387
A	351	LYS	GLU	engineered mutation	UNP P04387
B	-2	GLY	-	expression tag	UNP P04387
B	-1	SER	-	expression tag	UNP P04387
B	0	HIS	-	expression tag	UNP P04387
B	351	LYS	GLU	engineered mutation	UNP P04387
C	-2	GLY	-	expression tag	UNP P04387
C	-1	SER	-	expression tag	UNP P04387
C	0	HIS	-	expression tag	UNP P04387
C	351	LYS	GLU	engineered mutation	UNP P04387
D	-2	GLY	-	expression tag	UNP P04387
D	-1	SER	-	expression tag	UNP P04387
D	0	HIS	-	expression tag	UNP P04387
D	351	LYS	GLU	engineered mutation	UNP P04387
E	-2	GLY	-	expression tag	UNP P04387

Continued on next page...

Continued from previous page...

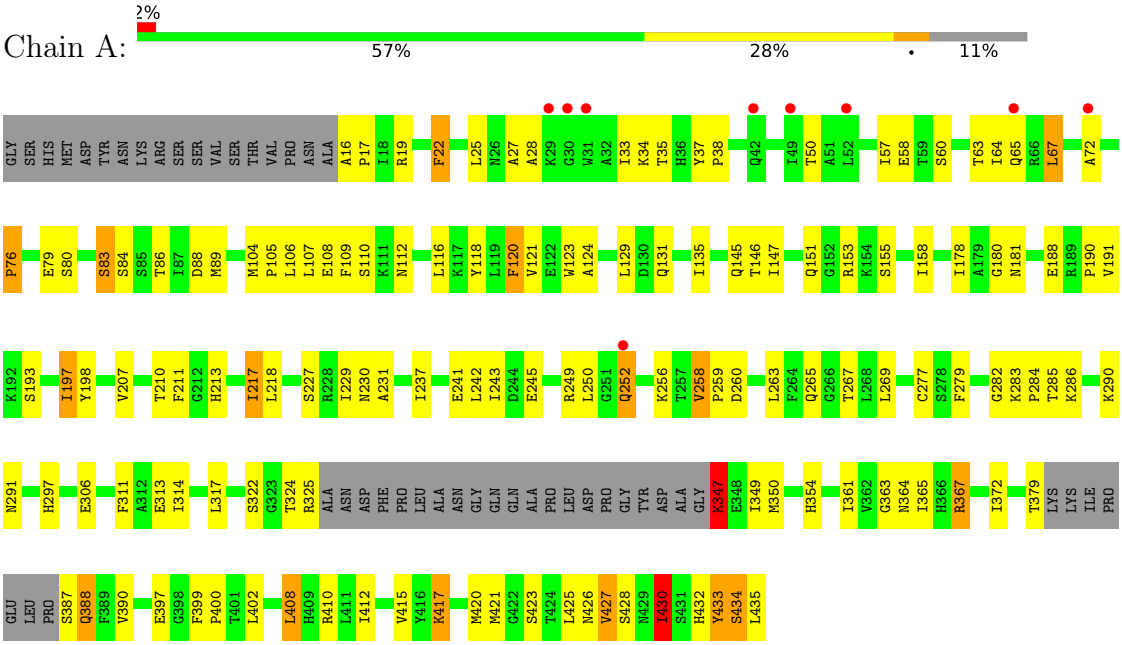
Chain	Residue	Modelled	Actual	Comment	Reference
E	-1	SER	-	expression tag	UNP P04387
E	0	HIS	-	expression tag	UNP P04387
E	351	LYS	GLU	engineered mutation	UNP P04387
F	-2	GLY	-	expression tag	UNP P04387
F	-1	SER	-	expression tag	UNP P04387
F	0	HIS	-	expression tag	UNP P04387
F	351	LYS	GLU	engineered mutation	UNP P04387

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.

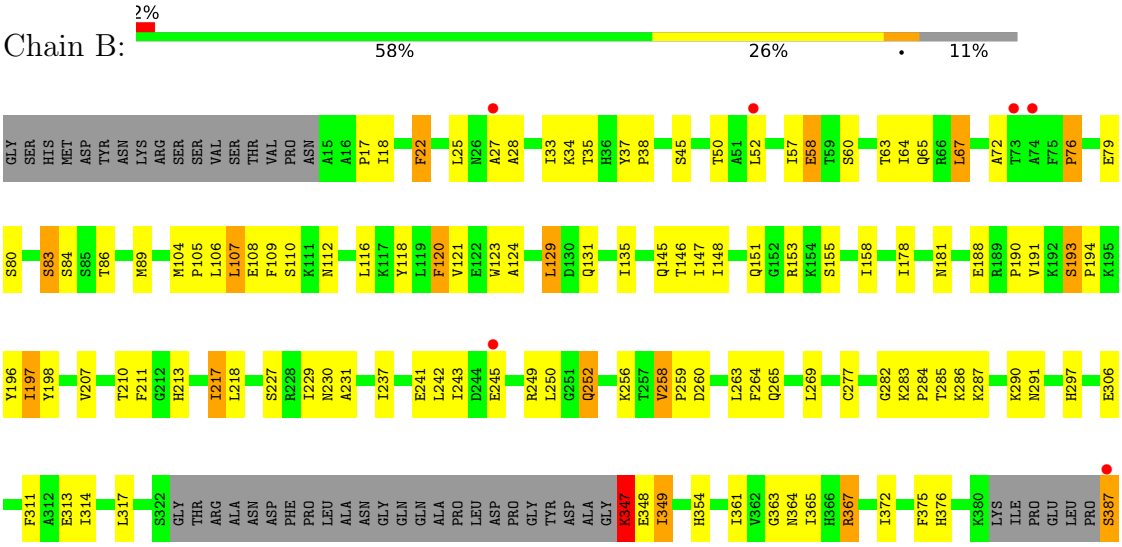
• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

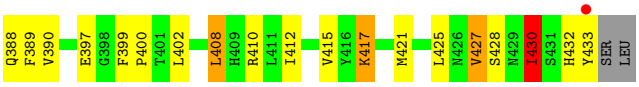
Chain A:



• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

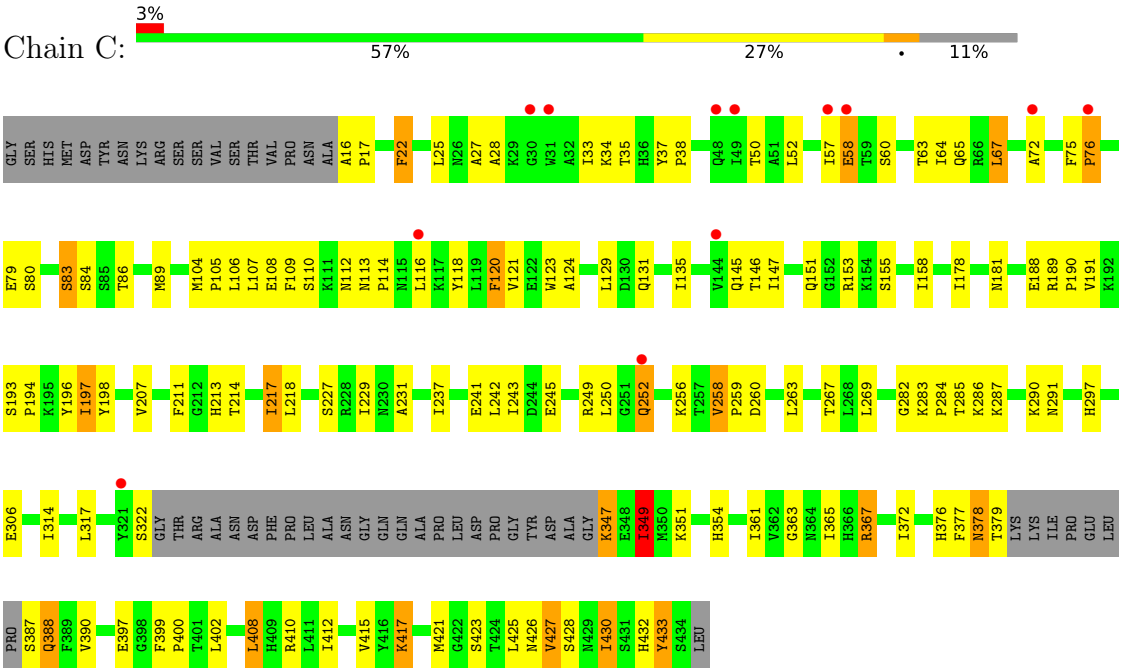
Chain B:





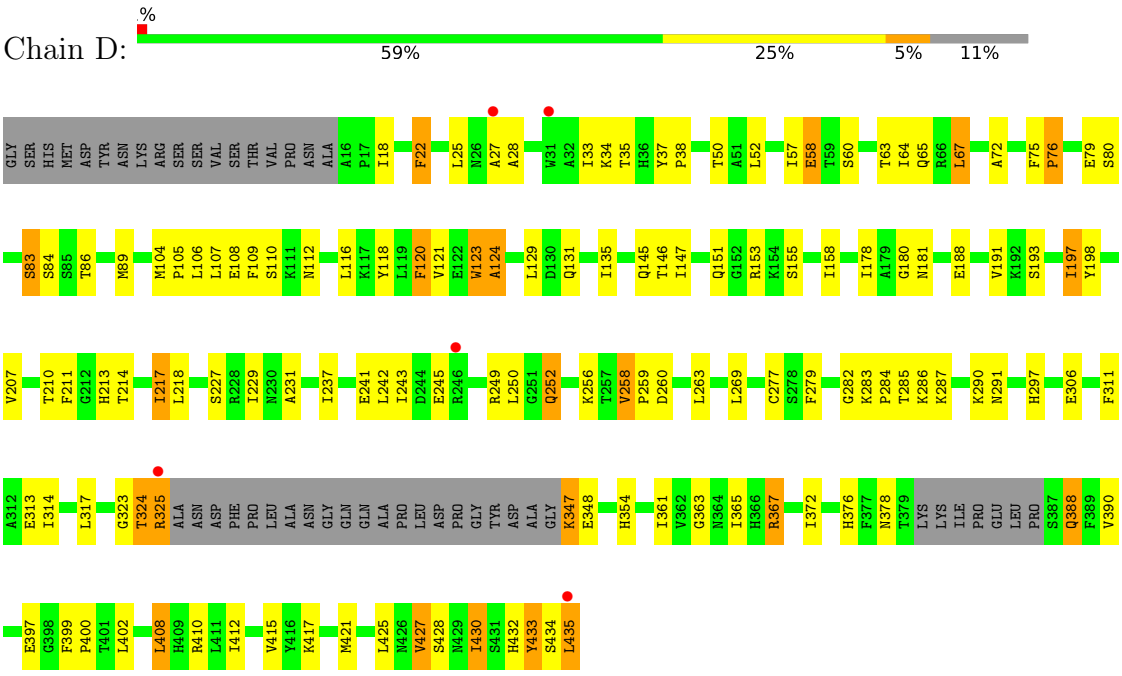
• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

Chain C:



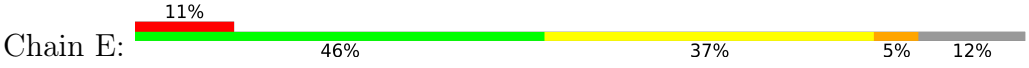
• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

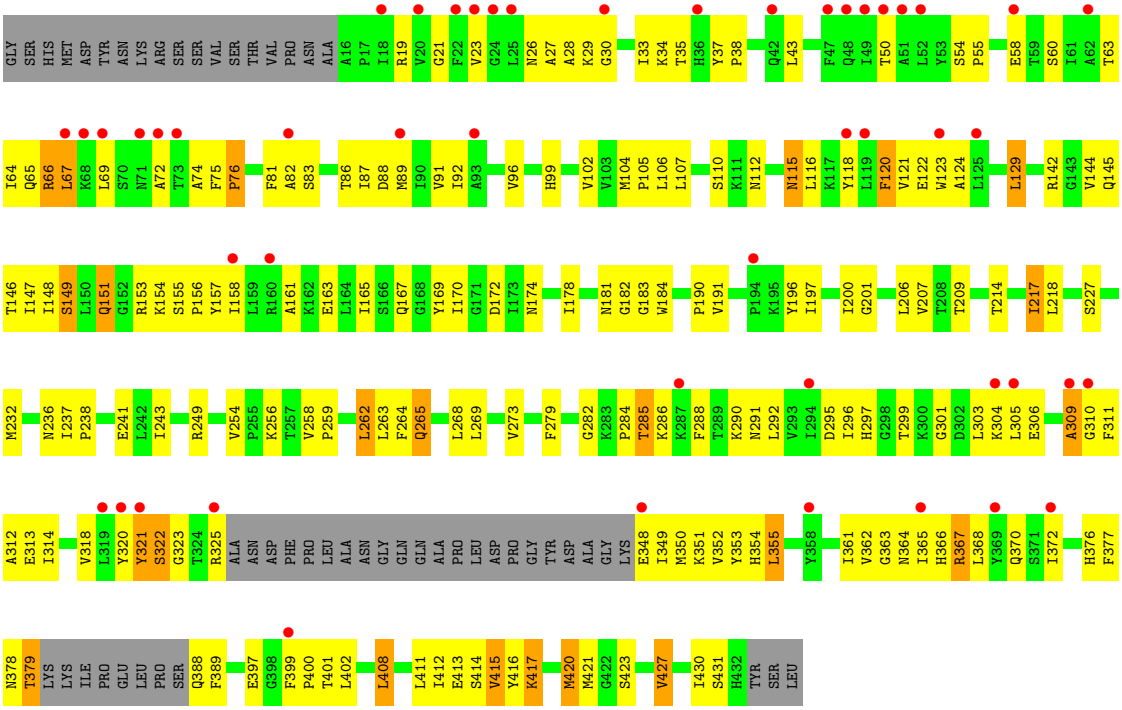
Chain D:



• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

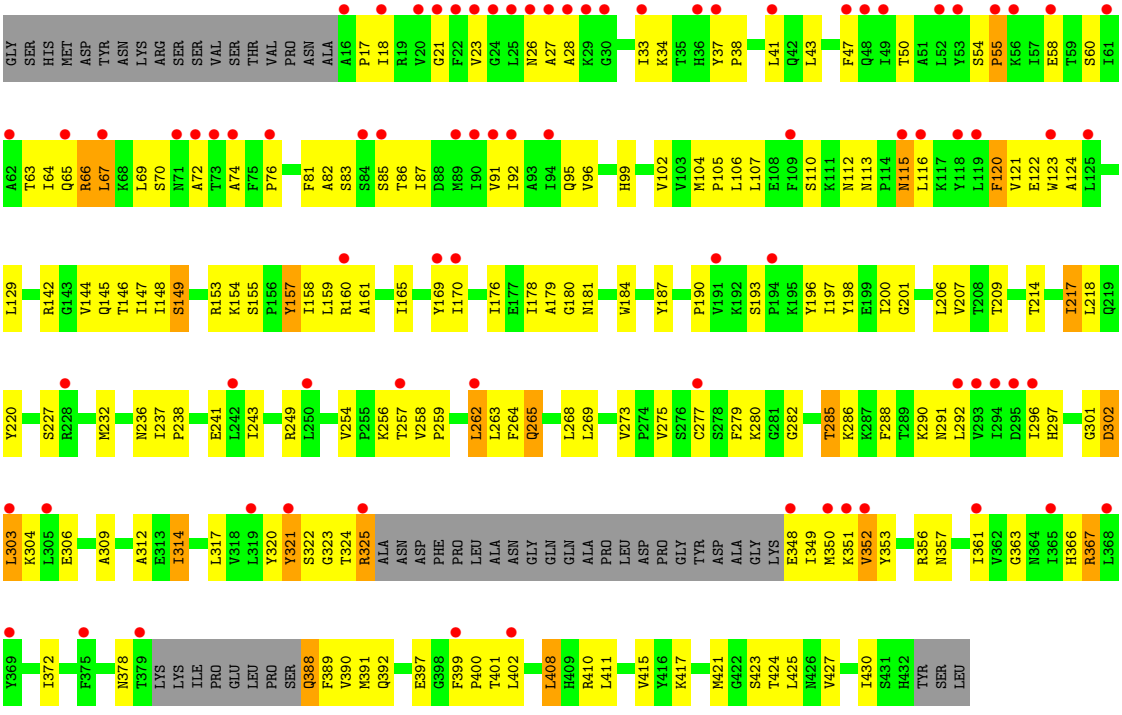
Chain E:





• Molecule 1: Galactose/lactose metabolism regulatory protein GAL80

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	495.32Å 84.86Å 66.46Å 90.00° 98.90° 90.00°	Depositor
Resolution (Å)	50.00 – 2.85 48.39 – 2.85	Depositor EDS
% Data completeness (in resolution range)	85.6 (50.00-2.85) 85.6 (48.39-2.85)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.2.0019, PHENIX	Depositor
R, R_{free}	0.228 , 0.278 0.242 , 0.264	Depositor DCC
R_{free} test set	5498 reflections (10.08%)	wwPDB-VP
Wilson B-factor (Å^2)	66.4	Xtriage
Anisotropy	0.629	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 122.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.040 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	18389	wwPDB-VP
Average B, all atoms (Å^2)	109.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.76% 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

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	0.34	0/3155	0.51	1/4270 (0.0%)
1	B	0.35	0/3132	0.52	1/4240 (0.0%)
1	C	0.34	0/3124	0.53	2/4230 (0.0%)
1	D	0.36	0/3154	0.51	0/4270
1	E	0.28	0/3112	0.47	0/4214
1	F	0.28	0/3112	0.47	0/4214
All	All	0.33	0/18789	0.50	4/25438 (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	D	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	347	LYS	N-CA-C	-10.61	82.37	111.00
1	B	347	LYS	N-CA-C	-10.35	83.05	111.00
1	C	379	THR	N-CA-C	-6.64	93.06	111.00
1	A	347	LYS	N-CA-C	-5.91	95.06	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	123	TRP	Peptide

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	3088	0	3101	108	0
1	B	3065	0	3080	106	0
1	C	3057	0	3067	111	0
1	D	3087	0	3101	107	0
1	E	3046	0	3058	160	0
1	F	3046	0	3058	151	0
All	All	18389	0	18465	698	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 19.

The worst 5 of 698 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:347:LYS:N	1:D:348:GLU:CA	1.70	1.34
1:E:155:SER:HB3	1:E:158:ILE:HD13	1.41	1.03
1:B:347:LYS:HD2	1:B:347:LYS:O	1.59	0.99
1:C:349:ILE:HD11	1:C:351:LYS:HE3	1.48	0.94
1:D:347:LYS:N	1:D:348:GLU:HA	0.77	0.92

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	386/438 (88%)	356 (92%)	24 (6%)	6 (2%)	9 28

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	383/438 (87%)	353 (92%)	26 (7%)	4 (1%)	15 40
1	C	382/438 (87%)	350 (92%)	25 (6%)	7 (2%)	8 25
1	D	386/438 (88%)	352 (91%)	27 (7%)	7 (2%)	8 25
1	E	381/438 (87%)	304 (80%)	65 (17%)	12 (3%)	4 14
1	F	381/438 (87%)	298 (78%)	67 (18%)	16 (4%)	3 8
All	All	2299/2628 (88%)	2013 (88%)	234 (10%)	52 (2%)	6 20

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	433	TYR
1	C	349	ILE
1	C	378	ASN
1	C	433	TYR
1	D	124	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	338/375 (90%)	309 (91%)	29 (9%)	10 27
1	B	335/375 (89%)	305 (91%)	30 (9%)	9 25
1	C	335/375 (89%)	307 (92%)	28 (8%)	11 28
1	D	338/375 (90%)	309 (91%)	29 (9%)	10 27
1	E	333/375 (89%)	308 (92%)	25 (8%)	13 34
1	F	333/375 (89%)	311 (93%)	22 (7%)	16 40
All	All	2012/2250 (89%)	1849 (92%)	163 (8%)	11 30

5 of 163 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	435	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	67	LEU
1	E	107	LEU
1	E	348	GLU
1	F	184	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 85 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	145	GLN
1	F	99	HIS
1	E	174	ASN
1	E	272	ASN
1	F	202	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 95^{th} 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	392/438 (89%)	0.10	9 (2%) 60 57	51, 88, 138, 216	0
1	B	389/438 (88%)	0.03	7 (1%) 68 66	53, 89, 131, 169	0
1	C	388/438 (88%)	0.11	12 (3%) 49 44	54, 93, 163, 296	0
1	D	392/438 (89%)	-0.11	5 (1%) 77 76	20, 86, 131, 183	0
1	E	387/438 (88%)	0.74	49 (12%) 3 2	80, 129, 203, 281	0
1	F	387/438 (88%)	1.08	81 (20%) 1 0	74, 136, 215, 350	0
All	All	2335/2628 (88%)	0.32	163 (6%) 16 12	20, 102, 181, 350	0

The worst 5 of 163 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	27	ALA	10.0
1	F	47	PHE	8.7
1	F	30	GLY	8.7
1	E	325	ARG	7.7
1	F	48	GLN	7.0

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.