wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 12:33 pm BST

PDB ID	:	6FNP
Title	:	Crystal structure of ECF-CbrT, a cobalamin transporter
Authors	:	Santos, J.A.; Rempel, S.; Guskov, A.; Slotboom, D.J.
Deposited on	:	2018-02-05
Resolution	:	3.40 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:

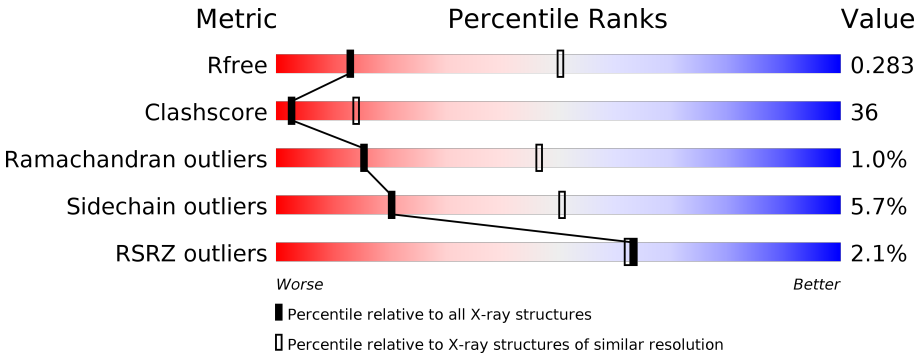
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.40 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.



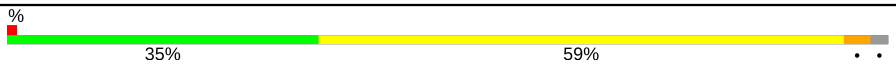
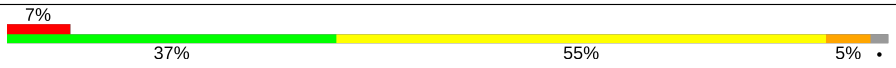
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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 <=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	182	% 34% 46% 7% • 13%
1	E	182	% 21% 47% • 27%
2	B	300	33% 54% 5% • 7%
2	F	300	3% 40% 51% • 7%
3	C	287	43% 52% • •
3	G	287	% 41% 54% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	265	
4	H	265	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15083 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 Membrane protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	159	Total	C	N	O	S	0	0	0
			1242	847	190	199	6			
1	E	132	Total	C	N	O	S	0	0	0
			999	676	152	166	5			

- Molecule 2 is a protein called Energy-coupling factor transporter ATP-binding protein EcfA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	280	Total	C	N	O	S	0	0	0
			2147	1354	359	430	4			
2	F	280	Total	C	N	O	S	0	0	0
			2147	1354	359	430	4			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-17	MET	-	initiating methionine	UNP Q1GBJ0
B	-16	HIS	-	expression tag	UNP Q1GBJ0
B	-15	HIS	-	expression tag	UNP Q1GBJ0
B	-14	HIS	-	expression tag	UNP Q1GBJ0
B	-13	HIS	-	expression tag	UNP Q1GBJ0
B	-12	HIS	-	expression tag	UNP Q1GBJ0
B	-11	HIS	-	expression tag	UNP Q1GBJ0
B	-10	HIS	-	expression tag	UNP Q1GBJ0
B	-9	HIS	-	expression tag	UNP Q1GBJ0
B	-8	HIS	-	expression tag	UNP Q1GBJ0
B	-7	HIS	-	expression tag	UNP Q1GBJ0
B	-6	GLY	-	expression tag	UNP Q1GBJ0
B	-5	GLU	-	expression tag	UNP Q1GBJ0
B	-4	ASN	-	expression tag	UNP Q1GBJ0
B	-3	LEU	-	expression tag	UNP Q1GBJ0
B	-2	TYR	-	expression tag	UNP Q1GBJ0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	PHE	-	expression tag	UNP Q1GBJ0
B	0	GLN	-	expression tag	UNP Q1GBJ0
B	1	GLY	-	expression tag	UNP Q1GBJ0
F	-17	MET	-	initiating methionine	UNP Q1GBJ0
F	-16	HIS	-	expression tag	UNP Q1GBJ0
F	-15	HIS	-	expression tag	UNP Q1GBJ0
F	-14	HIS	-	expression tag	UNP Q1GBJ0
F	-13	HIS	-	expression tag	UNP Q1GBJ0
F	-12	HIS	-	expression tag	UNP Q1GBJ0
F	-11	HIS	-	expression tag	UNP Q1GBJ0
F	-10	HIS	-	expression tag	UNP Q1GBJ0
F	-9	HIS	-	expression tag	UNP Q1GBJ0
F	-8	HIS	-	expression tag	UNP Q1GBJ0
F	-7	HIS	-	expression tag	UNP Q1GBJ0
F	-6	GLY	-	expression tag	UNP Q1GBJ0
F	-5	GLU	-	expression tag	UNP Q1GBJ0
F	-4	ASN	-	expression tag	UNP Q1GBJ0
F	-3	LEU	-	expression tag	UNP Q1GBJ0
F	-2	TYR	-	expression tag	UNP Q1GBJ0
F	-1	PHE	-	expression tag	UNP Q1GBJ0
F	0	GLN	-	expression tag	UNP Q1GBJ0
F	1	GLY	-	expression tag	UNP Q1GBJ0

- Molecule 3 is a protein called Energy-coupling factor transporter ATP-binding protein EcfA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	282	Total	C	N	O	S	0	0	0
			2191	1399	371	412	9			
3	G	282	Total	C	N	O	S	0	0	0
			2191	1399	371	412	9			

- Molecule 4 is a protein called Energy-coupling factor transporter transmembrane protein EcfT.

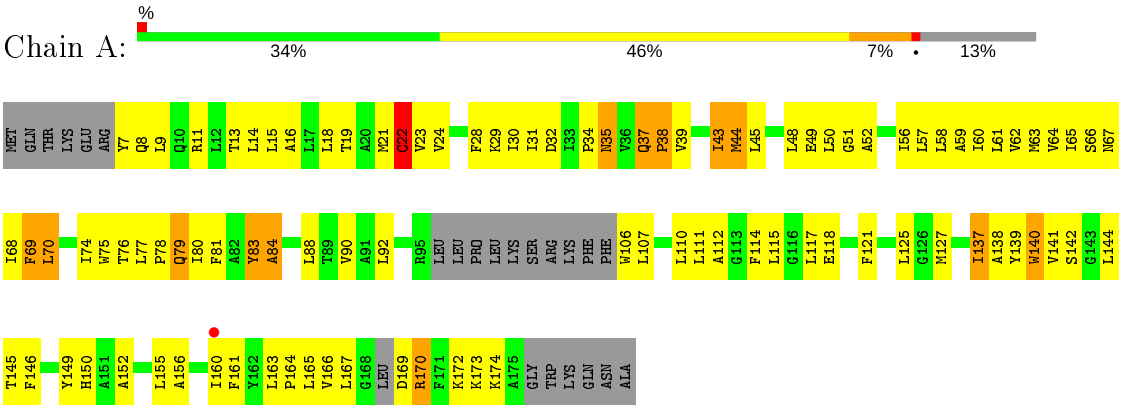
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	259	Total	C	N	O	S	0	0	0
			2083	1391	333	345	14			
4	H	259	Total	C	N	O	S	0	0	0
			2083	1391	333	345	14			

3 Residue-property plots (i)

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.

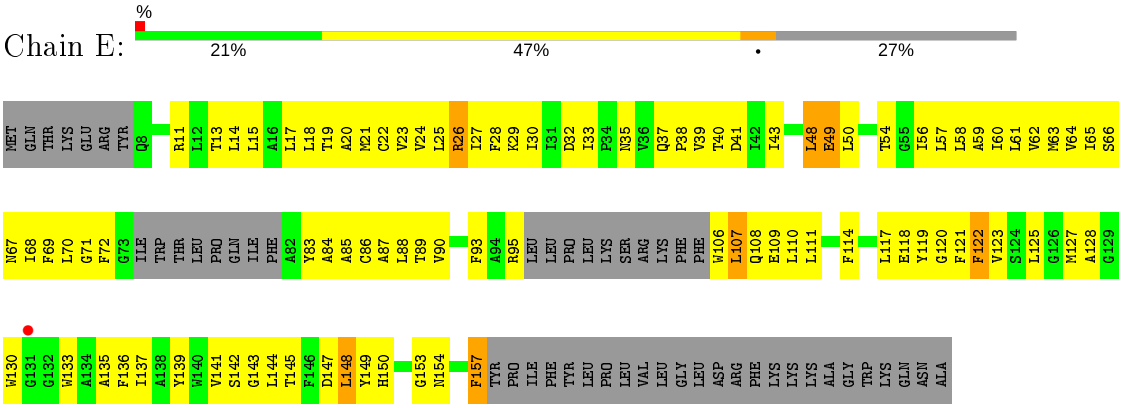
- Molecule 1: Membrane protein

Chain A:



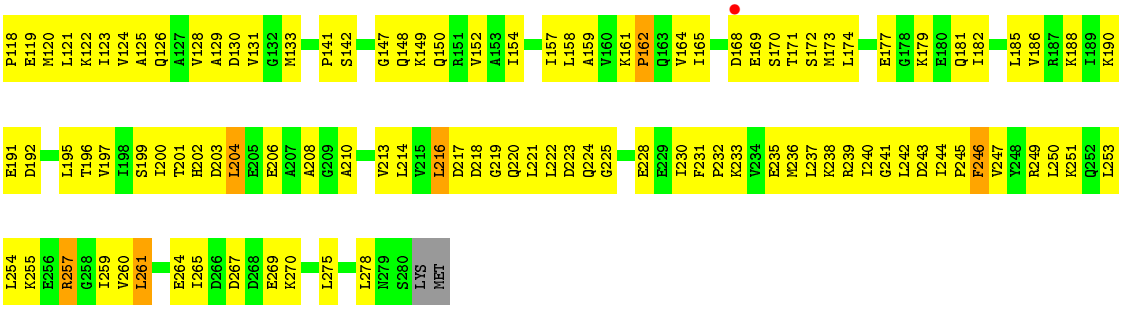
- Molecule 1: Membrane protein

Chain E:



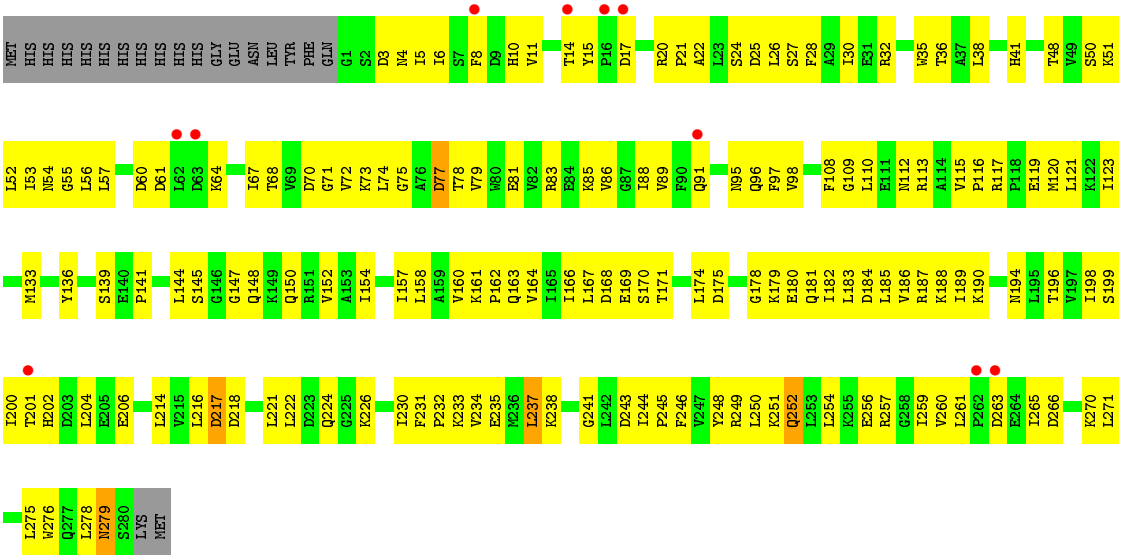
- Molecule 2: Energy-coupling factor transporter ATP-binding protein EcfA1

Chain B:



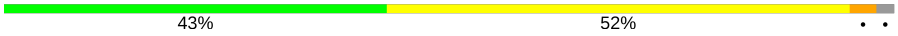
• Molecule 2: Energy-coupling factor transporter ATP-binding protein EcfA1

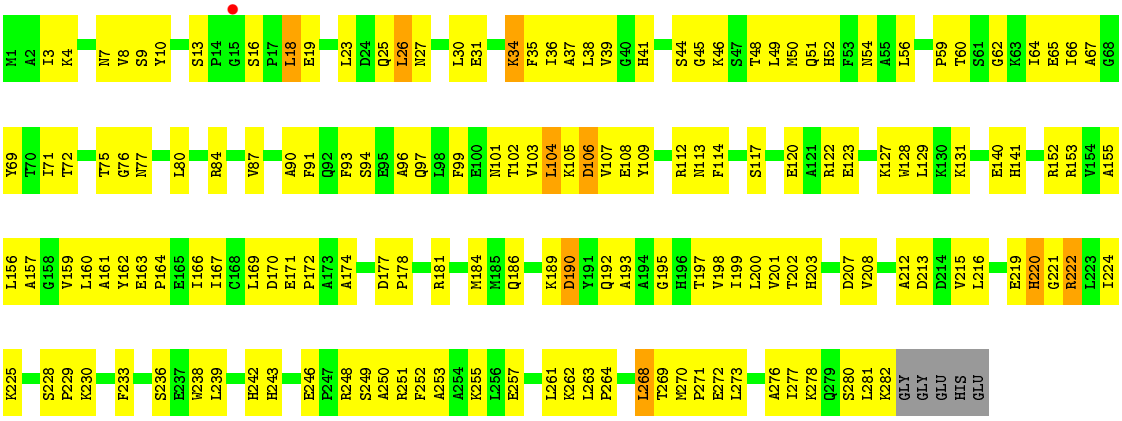
Chain F:



• Molecule 3: Energy-coupling factor transporter ATP-binding protein EcfA2

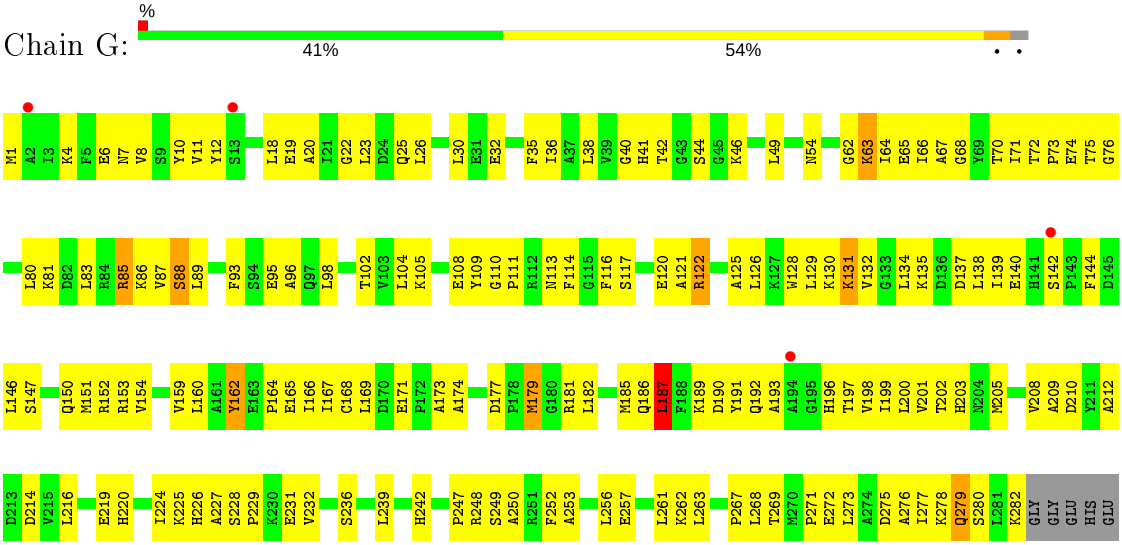
Chain C:





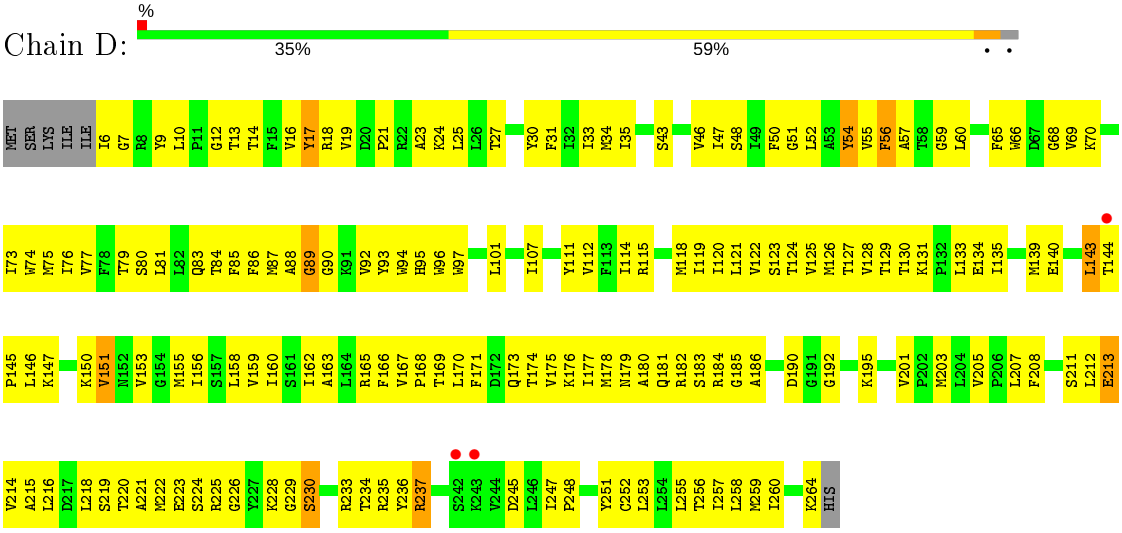
• Molecule 3: Energy-coupling factor transporter ATP-binding protein EcfA2

Chain G:



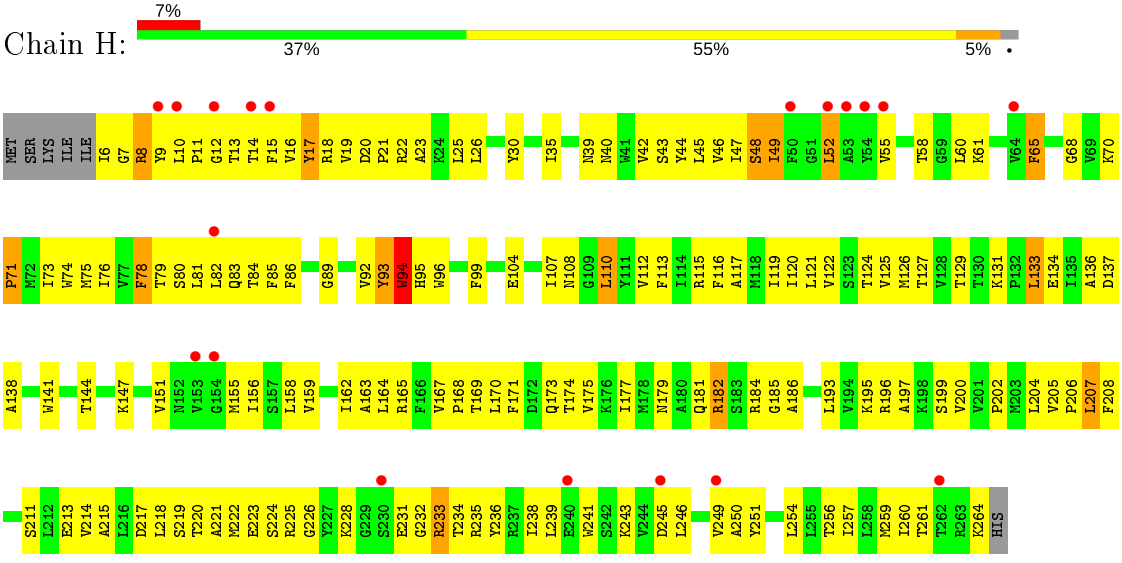
• Molecule 4: Energy-coupling factor transporter transmembrane protein EcfT

Chain D:



• Molecule 4: Energy-coupling factor transporter transmembrane protein EcfT

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	85.47Å 92.86Å 105.51Å 72.57° 66.27° 62.89°	Depositor
Resolution (Å)	47.80 – 3.40 47.80 – 3.40	Depositor EDS
% Data completeness (in resolution range)	88.3 (47.80-3.40) 86.8 (47.80-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.98 (at 3.40Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.238 , 0.293 0.241 , 0.283	Depositor DCC
R_{free} test set	1587 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	34.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.19 , -11.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.256 for h,h-k,h-l	Xtriage
F_o,F_c correlation	0.79	EDS
Total number of atoms	15083	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.64% 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	0.54	0/1276	1.07	5/1738 (0.3%)
1	E	0.40	0/1021	0.72	0/1391
2	B	0.55	0/2180	0.82	2/2959 (0.1%)
2	F	0.45	0/2180	0.68	1/2959 (0.0%)
3	C	0.54	0/2237	0.83	1/3019 (0.0%)
3	G	0.43	0/2237	0.71	1/3019 (0.0%)
4	D	0.50	0/2136	0.77	1/2901 (0.0%)
4	H	0.41	0/2136	0.66	1/2901 (0.0%)
All	All	0.48	0/15403	0.78	12/20887 (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	6
2	B	0	1
3	C	0	1
3	G	0	1
4	D	0	1
4	H	0	2
All	All	0	12

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	35	ASN	CB-CG-OD1	18.98	159.55	121.60
1	A	35	ASN	CB-CG-ND2	-17.75	74.09	116.70
3	G	187	LEU	CA-CB-CG	7.96	133.61	115.30
1	A	35	ASN	OD1-CG-ND2	-7.88	103.77	121.90
3	C	268	LEU	CA-CB-CG	6.69	130.68	115.30

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	35	ASN	Sidechain
1	A	37	GLN	Peptide
1	A	38	PRO	Peptide
1	A	79	GLN	Peptide
1	A	84	ALA	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	1242	0	1289	116	1
1	E	999	0	1028	83	0
2	B	2147	0	2159	172	1
2	F	2147	0	2159	167	2
3	C	2191	0	2199	157	0
3	G	2191	0	2199	152	0
4	D	2083	0	2185	183	0
4	H	2083	0	2185	168	0
All	All	15083	0	15403	1091	2

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 36.

The worst 5 of 1091 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)
4:D:107:ILE:HB	4:D:111:TYR:CE2	1.37	1.56
2:B:83:ARG:NH2	4:D:223:GLU:OE2	1.67	1.27
4:D:107:ILE:CB	4:D:111:TYR:CE2	2.24	1.21
4:D:107:ILE:HB	4:D:111:TYR:CD2	1.92	1.03
2:B:105:ASP:O	4:D:225:ARG:NH2	1.94	0.99

All (2) 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:A:170:ARG:NH2	2:F:235:GLU:OE2[1_654]	2.12	0.08
2:B:122:LYS:NZ	2:F:160:VAL:O[1_564]	2.15	0.05

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	155/182 (85%)	107 (69%)	46 (30%)	2 (1%)	12 39
1	E	126/182 (69%)	96 (76%)	28 (22%)	2 (2%)	9 34
2	B	278/300 (93%)	215 (77%)	60 (22%)	3 (1%)	14 44
2	F	278/300 (93%)	206 (74%)	71 (26%)	1 (0%)	34 67
3	C	280/287 (98%)	212 (76%)	65 (23%)	3 (1%)	14 44
3	G	280/287 (98%)	203 (72%)	75 (27%)	2 (1%)	22 55
4	D	257/265 (97%)	179 (70%)	75 (29%)	3 (1%)	13 41
4	H	257/265 (97%)	164 (64%)	89 (35%)	4 (2%)	9 34
All	All	1911/2068 (92%)	1382 (72%)	509 (27%)	20 (1%)	15 46

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	94	TRP
1	E	38	PRO
1	E	39	VAL
3	G	279	GLN
3	C	26	LEU

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	125/146 (86%)	118 (94%)	7 (6%)	21 51
1	E	98/146 (67%)	85 (87%)	13 (13%)	4 15
2	B	240/259 (93%)	223 (93%)	17 (7%)	14 44
2	F	240/259 (93%)	233 (97%)	7 (3%)	42 69
3	C	231/234 (99%)	225 (97%)	6 (3%)	46 72
3	G	231/234 (99%)	215 (93%)	16 (7%)	15 45
4	D	227/233 (97%)	217 (96%)	10 (4%)	28 58
4	H	227/233 (97%)	210 (92%)	17 (8%)	13 41
All	All	1619/1744 (93%)	1526 (94%)	93 (6%)	20 50

5 of 93 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	35	ASN
1	E	157	PHE
4	H	93	TYR
1	E	37	GLN
1	E	110	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	G	41	HIS
3	G	113	ASN
4	H	83	GLN
2	F	279	ASN
3	G	242	HIS

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

There are no carbohydrates 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	159/182 (87%)	-0.57	1 (0%)	89	89	2, 14, 41, 57	0
1	E	132/182 (72%)	-0.47	1 (0%)	86	85	3, 28, 58, 67	0
2	B	280/300 (93%)	-0.62	1 (0%)	92	92	2, 11, 32, 44	0
2	F	280/300 (93%)	-0.34	10 (3%)	42	42	3, 23, 52, 76	0
3	C	282/287 (98%)	-0.63	1 (0%)	92	92	2, 10, 27, 55	0
3	G	282/287 (98%)	-0.38	4 (1%)	75	74	6, 31, 56, 80	0
4	D	259/265 (97%)	-0.59	3 (1%)	79	77	2, 13, 39, 66	0
4	H	259/265 (97%)	-0.16	19 (7%)	15	17	3, 31, 59, 69	0
All	All	1933/2068 (93%)	-0.46	40 (2%)	63	62	2, 19, 50, 80	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	2	ALA	5.1
4	H	52	LEU	5.0
3	G	13	SER	4.8
1	E	131	GLY	4.7
3	G	142	SER	4.1

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.