



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 04:35 am BST

PDB ID : 1FOU  
Title : CONNECTOR PROTEIN FROM BACTERIOPHAGE PHI29  
Authors : Simpson, A.A.; Tao, Y.; Leiman, P.G.; Badasso, M.O.; He, Y.; Jardine, P.J.; Olson, N.H.; Morais, M.C.; Grimes, S.N.; Anderson, D.L.; Baker, T.S.; Rossmann, M.G.  
Deposited on : 2000-08-28  
Resolution : 3.20 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 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

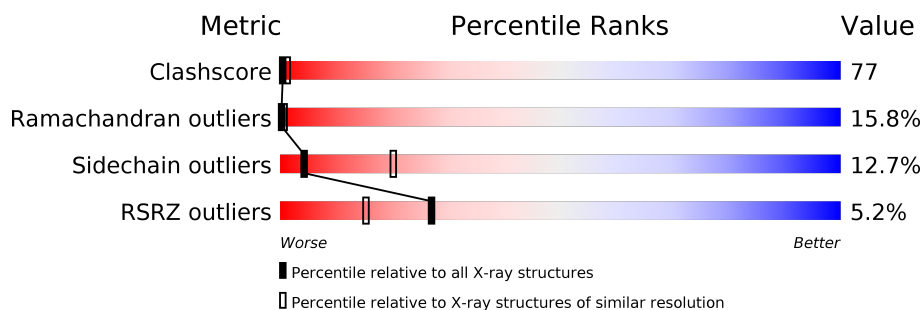
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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  $\geq 3$ , 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	309	<div> <div>3%</div> <div>20% 46% 15% 17%</div> </div>
1	B	309	<div> <div>4%</div> <div>17% 47% 16% 17%</div> </div>
1	C	309	<div> <div>5%</div> <div>15% 51% 15% 17%</div> </div>
1	D	309	<div> <div>2%</div> <div>13% 54% 14% 17%</div> </div>
1	E	309	<div> <div>11%</div> <div>16% 49% 16% 17%</div> </div>
1	F	309	<div> <div>6%</div> <div>20% 45% 14% 17%</div> </div>
1	G	309	<div> <div>5%</div> <div>17% 49% 16% 17%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	309	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>7%17%51%15%•17%</div></div>
1	I	309	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>2%19%48%14%•17%</div></div>
1	J	309	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>2%16%49%16%•17%</div></div>
1	K	309	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>4%17%49%14%•17%</div></div>
1	L	309	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>2%22%42%15%•17%</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 25272 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 UPPER COLLAR PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	B	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	C	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	D	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	E	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	F	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	G	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	H	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	I	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	J	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	K	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			
1	L	257	Total	C	N	O	S	0	0	0
			2106	1350	348	401	7			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	225	LYS	LEU	CONFLICT	UNP P04332
A	226	LEU	GLY	CONFLICT	UNP P04332
A	227	GLN	ILE	CONFLICT	UNP P04332
A	228	THR	LYS	CONFLICT	UNP P04332
A	251	ASP	GLU	CONFLICT	UNP P04332

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	225	LYS	LEU	CONFLICT	UNP P04332
B	226	LEU	GLY	CONFLICT	UNP P04332
B	227	GLN	ILE	CONFLICT	UNP P04332
B	228	THR	LYS	CONFLICT	UNP P04332
B	251	ASP	GLU	CONFLICT	UNP P04332
C	225	LYS	LEU	CONFLICT	UNP P04332
C	226	LEU	GLY	CONFLICT	UNP P04332
C	227	GLN	ILE	CONFLICT	UNP P04332
C	228	THR	LYS	CONFLICT	UNP P04332
C	251	ASP	GLU	CONFLICT	UNP P04332
D	225	LYS	LEU	CONFLICT	UNP P04332
D	226	LEU	GLY	CONFLICT	UNP P04332
D	227	GLN	ILE	CONFLICT	UNP P04332
D	228	THR	LYS	CONFLICT	UNP P04332
D	251	ASP	GLU	CONFLICT	UNP P04332
E	225	LYS	LEU	CONFLICT	UNP P04332
E	226	LEU	GLY	CONFLICT	UNP P04332
E	227	GLN	ILE	CONFLICT	UNP P04332
E	228	THR	LYS	CONFLICT	UNP P04332
E	251	ASP	GLU	CONFLICT	UNP P04332
F	225	LYS	LEU	CONFLICT	UNP P04332
F	226	LEU	GLY	CONFLICT	UNP P04332
F	227	GLN	ILE	CONFLICT	UNP P04332
F	228	THR	LYS	CONFLICT	UNP P04332
F	251	ASP	GLU	CONFLICT	UNP P04332
G	225	LYS	LEU	CONFLICT	UNP P04332
G	226	LEU	GLY	CONFLICT	UNP P04332
G	227	GLN	ILE	CONFLICT	UNP P04332
G	228	THR	LYS	CONFLICT	UNP P04332
G	251	ASP	GLU	CONFLICT	UNP P04332
H	225	LYS	LEU	CONFLICT	UNP P04332
H	226	LEU	GLY	CONFLICT	UNP P04332
H	227	GLN	ILE	CONFLICT	UNP P04332
H	228	THR	LYS	CONFLICT	UNP P04332
H	251	ASP	GLU	CONFLICT	UNP P04332
I	225	LYS	LEU	CONFLICT	UNP P04332
I	226	LEU	GLY	CONFLICT	UNP P04332
I	227	GLN	ILE	CONFLICT	UNP P04332
I	228	THR	LYS	CONFLICT	UNP P04332
I	251	ASP	GLU	CONFLICT	UNP P04332
J	225	LYS	LEU	CONFLICT	UNP P04332
J	226	LEU	GLY	CONFLICT	UNP P04332

*Continued on next page...*

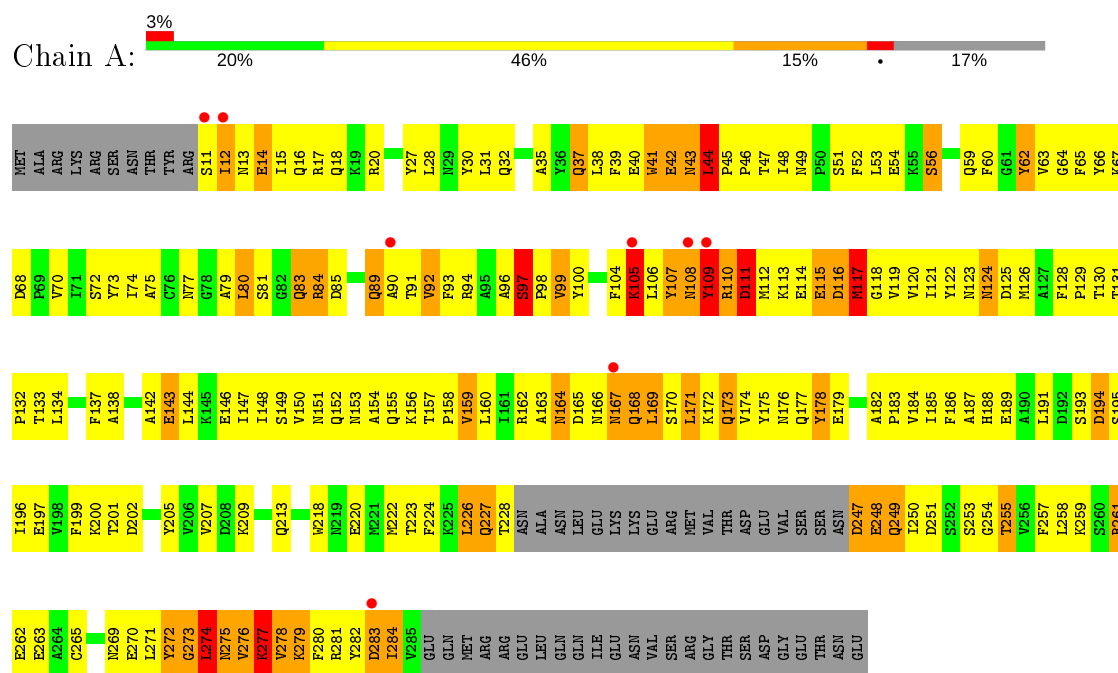
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
J	227	GLN	ILE	CONFLICT	UNP P04332
J	228	THR	LYS	CONFLICT	UNP P04332
J	251	ASP	GLU	CONFLICT	UNP P04332
K	225	LYS	LEU	CONFLICT	UNP P04332
K	226	LEU	GLY	CONFLICT	UNP P04332
K	227	GLN	ILE	CONFLICT	UNP P04332
K	228	THR	LYS	CONFLICT	UNP P04332
K	251	ASP	GLU	CONFLICT	UNP P04332
L	225	LYS	LEU	CONFLICT	UNP P04332
L	226	LEU	GLY	CONFLICT	UNP P04332
L	227	GLN	ILE	CONFLICT	UNP P04332
L	228	THR	LYS	CONFLICT	UNP P04332
L	251	ASP	GLU	CONFLICT	UNP P04332

### 3 Residue-property plots

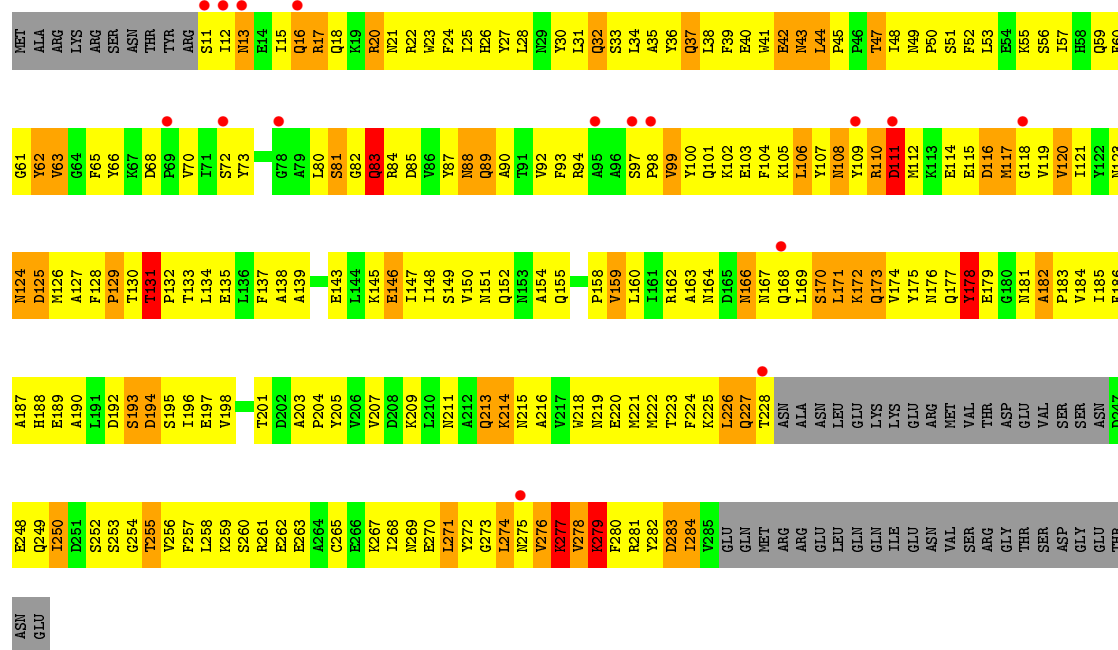
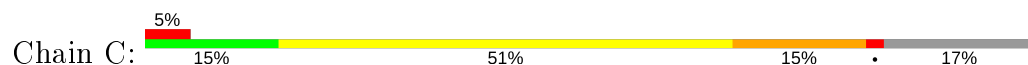
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

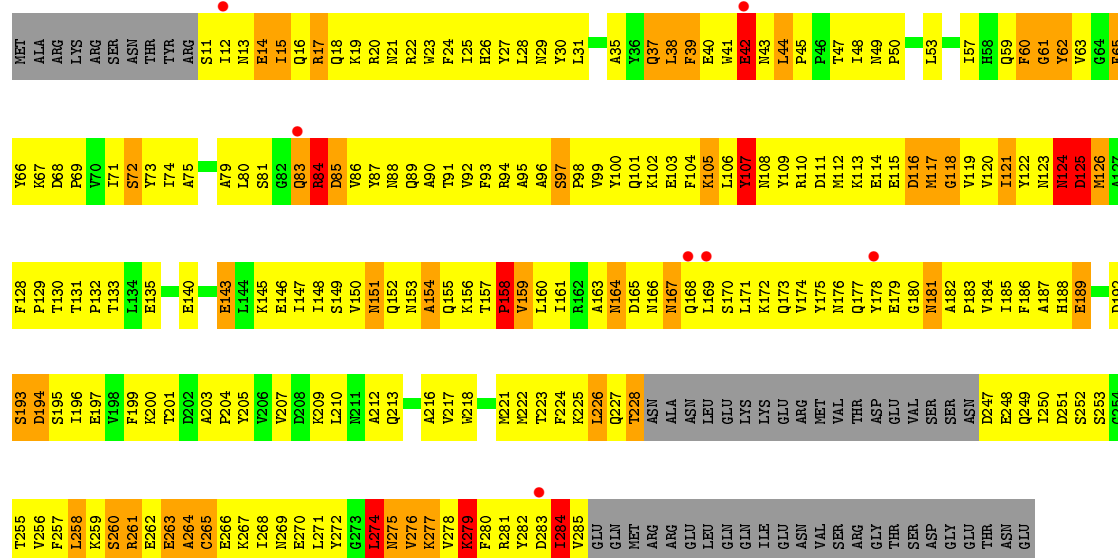
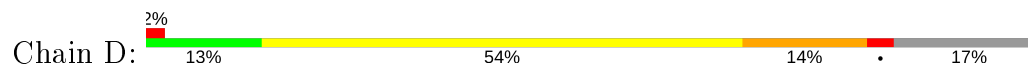




• Molecule 1: UPPER COLLAR PROTEIN

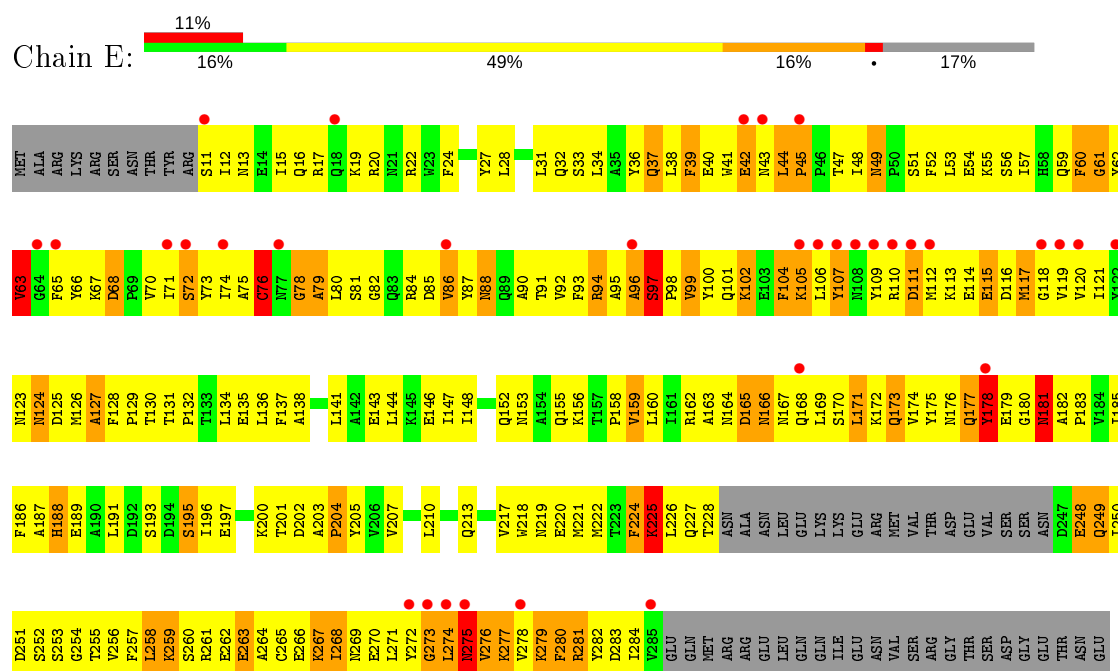


• Molecule 1: UPPER COLLAR PROTEIN

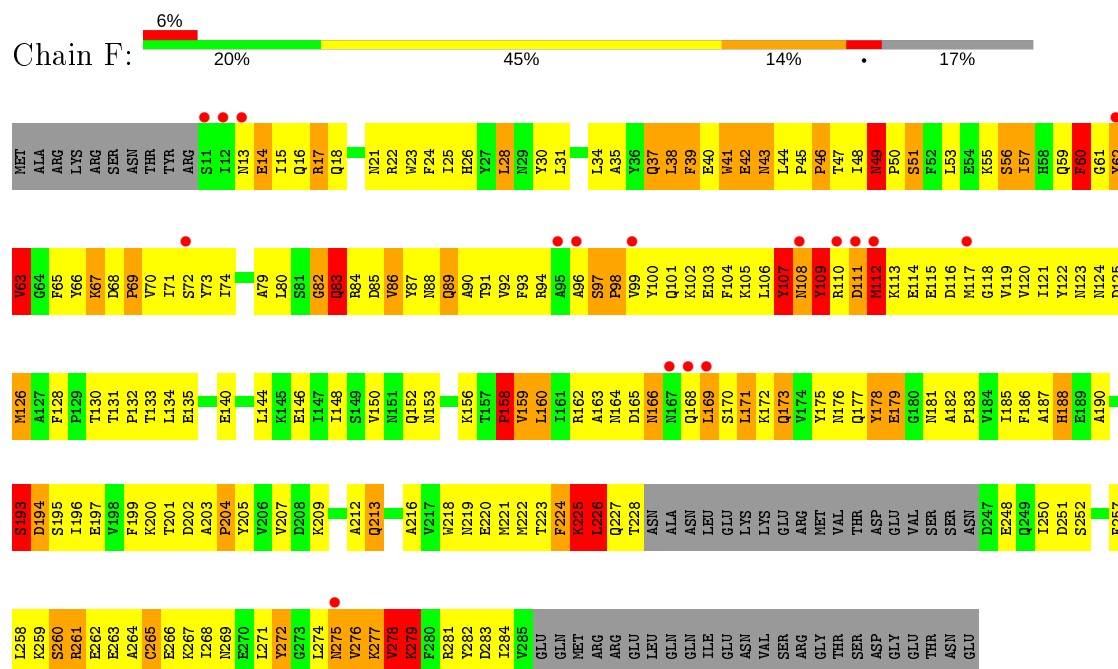


• Molecule 1: UPPER COLLAR PROTEIN

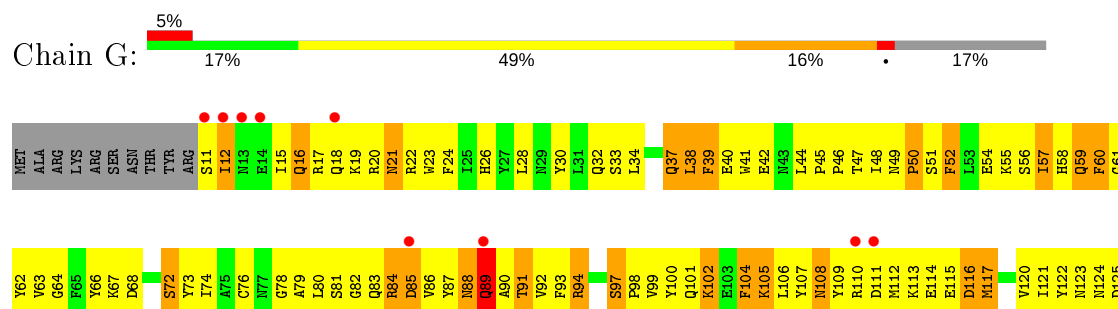


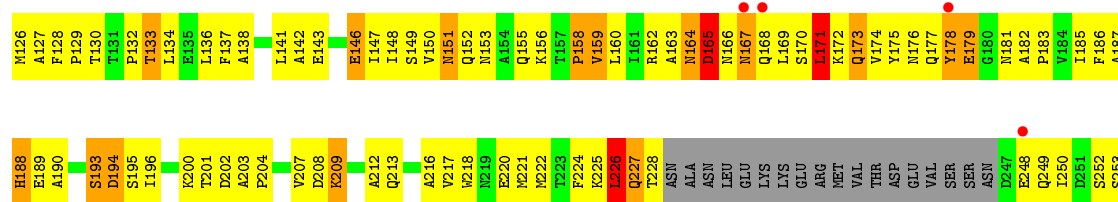


• Molecule 1: UPPER COLLAR PROTEIN

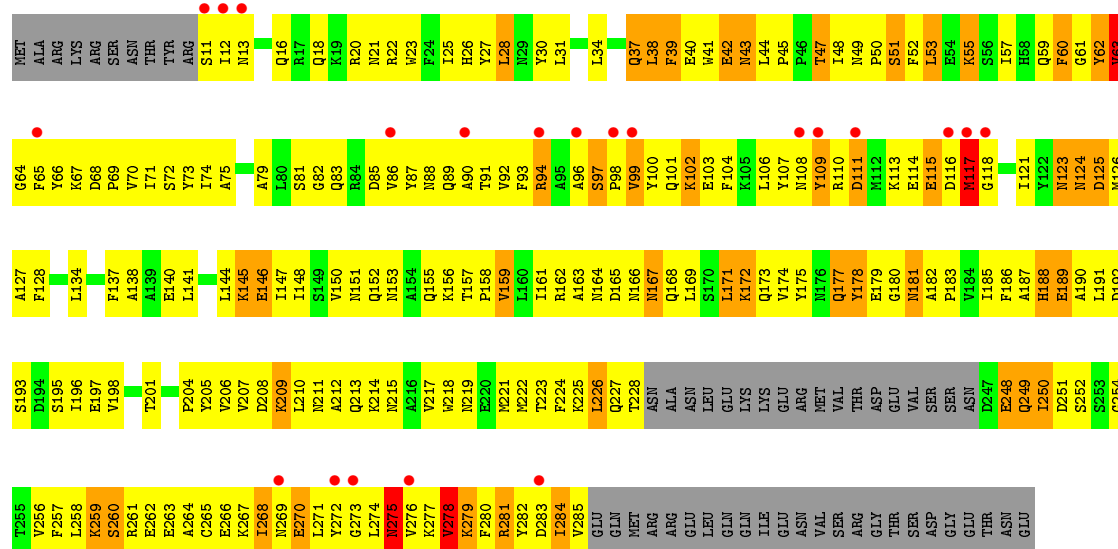
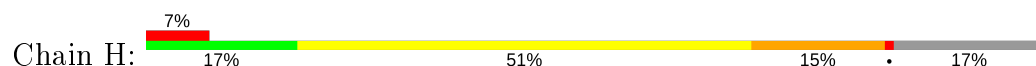


• Molecule 1: UPPER COLLAR PROTEIN

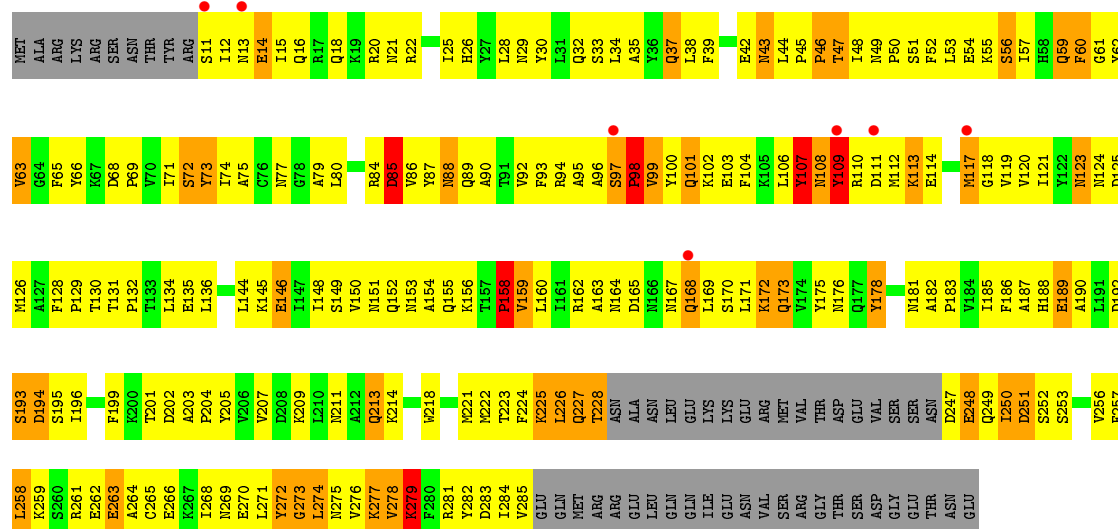
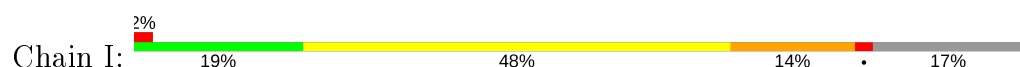




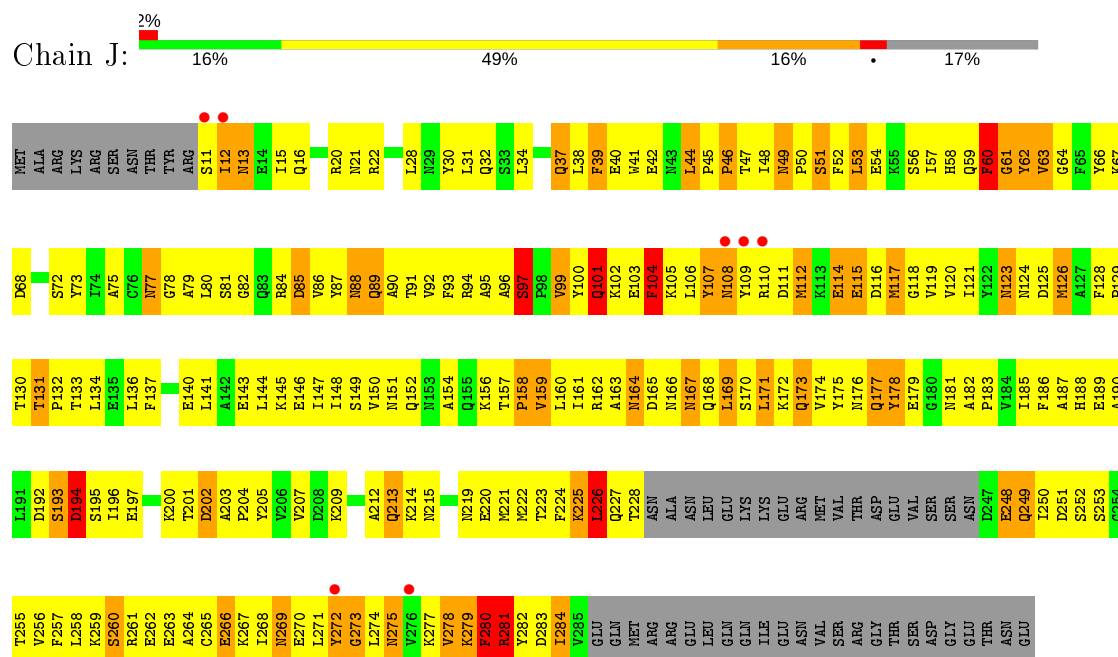
• Molecule 1: UPPER COLLAR PROTEIN



• Molecule 1: UPPER COLLAR PROTEIN



• Molecule 1: UPPER COLLAR PROTEIN





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	177.16Å 169.24Å 185.44Å 90.00° 114.10° 90.00°	Depositor
Resolution (Å)	9.00 – 3.20 48.53 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (9.00-3.20) 99.0 (48.53-3.50)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	14.94 (at 3.48Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.290 , 0.360 0.277 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.7	Xtriage
Anisotropy	0.434	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 42.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.80	EDS
Total number of atoms	25272	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.95% 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

### 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	0.47	0/2153	0.73	2/2918 (0.1%)
1	B	0.46	0/2153	0.74	2/2918 (0.1%)
1	C	0.46	0/2153	0.73	0/2918
1	D	0.44	0/2153	0.75	0/2918
1	E	0.43	0/2153	0.71	1/2918 (0.0%)
1	F	0.45	0/2153	0.70	1/2918 (0.0%)
1	G	0.44	0/2153	0.71	1/2918 (0.0%)
1	H	0.43	0/2153	0.70	2/2918 (0.1%)
1	I	0.49	0/2153	0.77	2/2918 (0.1%)
1	J	0.46	0/2153	0.78	4/2918 (0.1%)
1	K	0.45	0/2153	0.71	0/2918
1	L	0.49	0/2153	0.72	2/2918 (0.1%)
All	All	0.46	0/25836	0.73	17/35016 (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
1	L	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	44	LEU	CA-CB-CG	6.78	130.90	115.30
1	E	177	GLN	N-CA-C	-6.53	93.38	111.00
1	G	38	LEU	CA-CB-CG	6.40	130.03	115.30
1	H	38	LEU	CA-CB-CG	6.34	129.88	115.30
1	H	177	GLN	N-CA-C	-6.27	94.06	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	107	TYR	Sidechain
1	L	62	TYR	Sidechain

## 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	2106	0	2051	310	0
1	B	2106	0	2051	346	0
1	C	2106	0	2051	373	0
1	D	2106	0	2051	395	0
1	E	2106	0	2051	375	0
1	F	2106	0	2051	332	0
1	G	2106	0	2051	364	0
1	H	2106	0	2051	332	0
1	I	2106	0	2051	343	0
1	J	2106	0	2051	370	0
1	K	2106	0	2051	359	0
1	L	2106	0	2051	314	0
All	All	25272	0	24612	3838	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 77.

The worst 5 of 3838 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:F:119:VAL:HG11	1:F:268:ILE:HB	1.31	1.11
1:I:275:ASN:HB2	1:I:277:LYS:HE3	1.18	1.11
1:C:43:ASN:HB2	1:C:277:LYS:HD2	1.33	1.10
1:D:163:ALA:HB3	1:E:187:ALA:HB2	1.26	1.10
1:K:278:VAL:HG23	1:K:279:LYS:H	1.14	1.10

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	253/309 (82%)	170 (67%)	48 (19%)	35 (14%)	0	1
1	B	253/309 (82%)	160 (63%)	49 (19%)	44 (17%)	0	0
1	C	253/309 (82%)	172 (68%)	45 (18%)	36 (14%)	0	1
1	D	253/309 (82%)	150 (59%)	60 (24%)	43 (17%)	0	0
1	E	253/309 (82%)	151 (60%)	50 (20%)	52 (21%)	0	0
1	F	253/309 (82%)	166 (66%)	47 (19%)	40 (16%)	0	1
1	G	253/309 (82%)	160 (63%)	53 (21%)	40 (16%)	0	1
1	H	253/309 (82%)	176 (70%)	41 (16%)	36 (14%)	0	1
1	I	253/309 (82%)	187 (74%)	34 (13%)	32 (13%)	0	1
1	J	253/309 (82%)	175 (69%)	39 (15%)	39 (15%)	0	1
1	K	253/309 (82%)	169 (67%)	46 (18%)	38 (15%)	0	1
1	L	253/309 (82%)	163 (64%)	46 (18%)	44 (17%)	0	0
All	All	3036/3708 (82%)	1999 (66%)	558 (18%)	479 (16%)	0	1

5 of 479 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	12	ILE
1	A	42	GLU
1	A	92	VAL
1	A	97	SER
1	A	105	LYS

### 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	230/278 (83%)	198 (86%)	32 (14%)	3	16
1	B	230/278 (83%)	201 (87%)	29 (13%)	4	21
1	C	230/278 (83%)	199 (86%)	31 (14%)	4	18
1	D	230/278 (83%)	201 (87%)	29 (13%)	4	21
1	E	230/278 (83%)	209 (91%)	21 (9%)	9	34
1	F	230/278 (83%)	196 (85%)	34 (15%)	3	14
1	G	230/278 (83%)	200 (87%)	30 (13%)	4	19
1	H	230/278 (83%)	203 (88%)	27 (12%)	5	23
1	I	230/278 (83%)	202 (88%)	28 (12%)	5	22
1	J	230/278 (83%)	197 (86%)	33 (14%)	3	15
1	K	230/278 (83%)	200 (87%)	30 (13%)	4	19
1	L	230/278 (83%)	203 (88%)	27 (12%)	5	23
All	All	2760/3336 (83%)	2409 (87%)	351 (13%)	4	20

5 of 351 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	125	ASP
1	G	173	GLN
1	L	28	LEU
1	F	173	GLN
1	G	37	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 154 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	16	GLN
1	G	176	ASN
1	L	21	ASN
1	F	32	GLN
1	F	152	GLN

### 5.3.3 RNA ⓘ

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 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, 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	257/309 (83%)	0.01	8 (3%) 49 32	1, 7, 33, 38	0
1	B	257/309 (83%)	0.17	11 (4%) 35 22	1, 13, 32, 46	0
1	C	257/309 (83%)	0.27	16 (6%) 20 11	1, 13, 32, 39	0
1	D	257/309 (83%)	0.06	7 (2%) 54 39	1, 16, 35, 43	0
1	E	257/309 (83%)	0.56	33 (12%) 3 2	1, 21, 35, 42	0
1	F	257/309 (83%)	0.22	17 (6%) 18 11	1, 16, 35, 40	0
1	G	257/309 (83%)	0.22	15 (5%) 23 13	1, 14, 36, 51	0
1	H	257/309 (83%)	0.41	21 (8%) 11 6	1, 16, 34, 41	0
1	I	257/309 (83%)	0.00	7 (2%) 54 39	1, 9, 32, 43	0
1	J	257/309 (83%)	-0.01	7 (2%) 54 39	1, 9, 30, 43	0
1	K	257/309 (83%)	0.21	12 (4%) 31 19	1, 13, 33, 47	0
1	L	257/309 (83%)	0.01	7 (2%) 54 39	1, 11, 32, 41	0
All	All	3084/3708 (83%)	0.18	161 (5%) 27 15	1, 13, 34, 51	0

The worst 5 of 161 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	11	SER	7.4
1	B	108	ASN	6.1
1	A	11	SER	6.1
1	C	11	SER	5.6
1	H	12	ILE	5.5

### 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.