



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 08:01 am GMT

PDB ID : 4ZKT  
Title : Crystal structure of the progenitor M complex of Clostridium botulinum type E neurotoxin  
Authors : Eswaramoorthy, S.; Swaminathan, S.  
Deposited on : 2015-04-30  
Resolution : 3.05 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

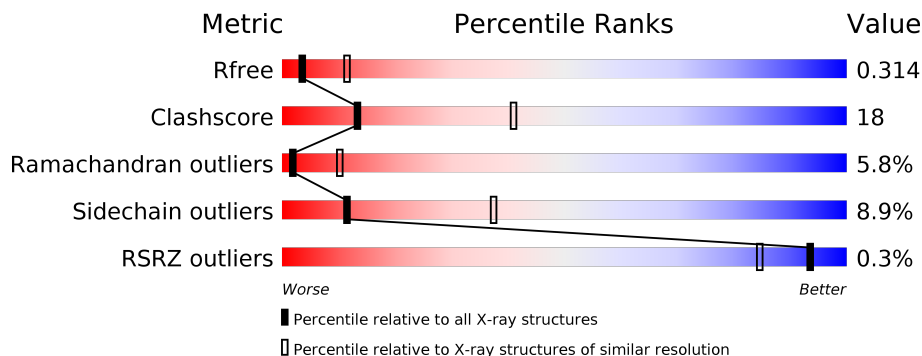
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.05 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))
$R_{free}$	100719	1348 (3.10-3.02)
Clashscore	112137	1462 (3.10-3.02)
Ramachandran outliers	110173	1410 (3.10-3.02)
Sidechain outliers	110143	1410 (3.10-3.02)
RSRZ outliers	101464	1355 (3.10-3.02)

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. 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	1252	<div> <div>59%</div> <div>32%</div> <div>7%</div> <div>.</div> </div>
1	C	1252	<div> <div>60%</div> <div>32%</div> <div>7%</div> <div>.</div> </div>
1	E	1252	<div> <div>59%</div> <div>32%</div> <div>7%</div> <div>.</div> </div>
2	B	1163	<div> <div>54%</div> <div>35%</div> <div>6%</div> <div>.</div> </div>
2	D	1163	<div> <div>55%</div> <div>34%</div> <div>6%</div> <div>.</div> </div>
2	F	1163	<div> <div>55%</div> <div>34%</div> <div>6%</div> <div>.</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 56952 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 Bontoxilysin A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1235	Total	C	N	O	S	0	0	0
			9978	6351	1665	1940	22			
1	C	1235	Total	C	N	O	S	0	0	0
			9978	6351	1665	1940	22			
1	E	1235	Total	C	N	O	S	0	0	0
			9978	6351	1665	1940	22			

- Molecule 2 is a protein called Botulinum neurotoxin type E, nontoxic-nonhemagglutinin component, NTNH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1114	Total	C	N	O	S	0	0	0
			9005	5804	1451	1726	24			
2	D	1114	Total	C	N	O	S	0	0	0
			9005	5804	1451	1726	24			
2	F	1114	Total	C	N	O	S	0	0	0
			9005	5804	1451	1726	24			

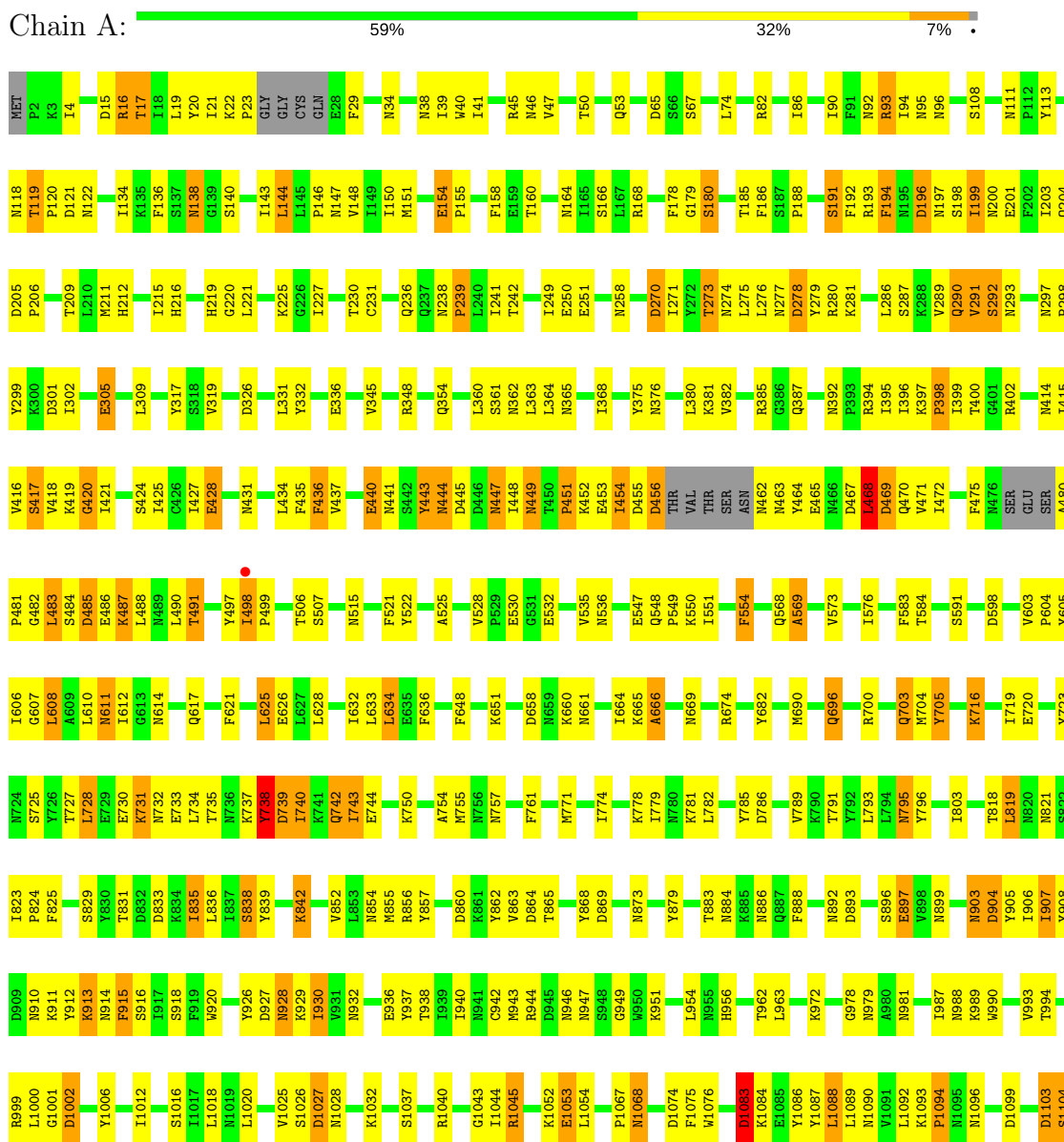
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		
3	C	1	Total	Zn	0	0
			1	1		
3	E	1	Total	Zn	0	0
			1	1		

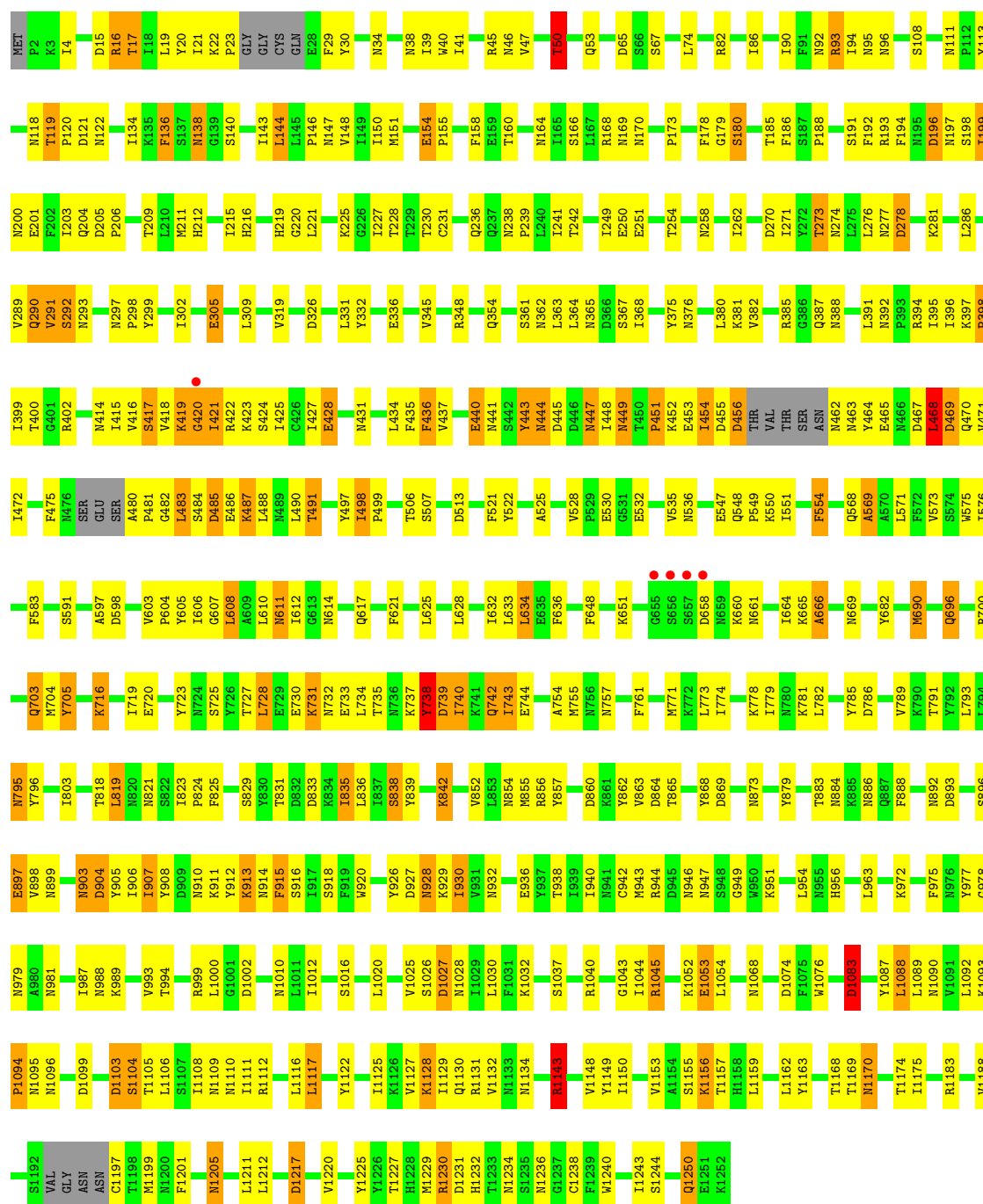
### 3 Residue-property plots

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.

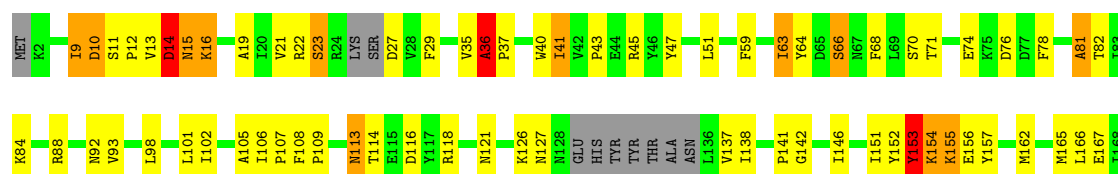
#### • Molecule 1: Bontoxilysin A

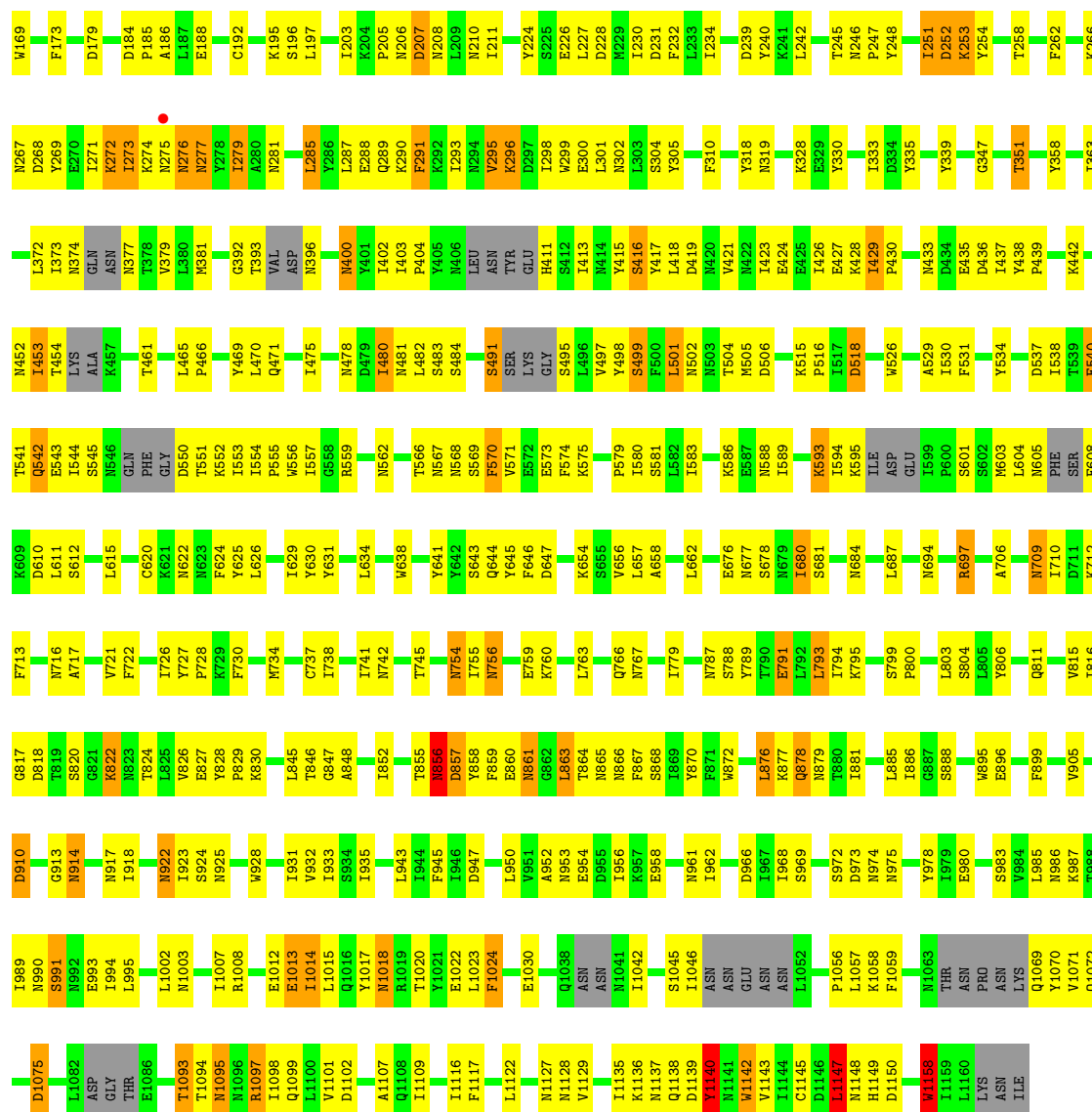




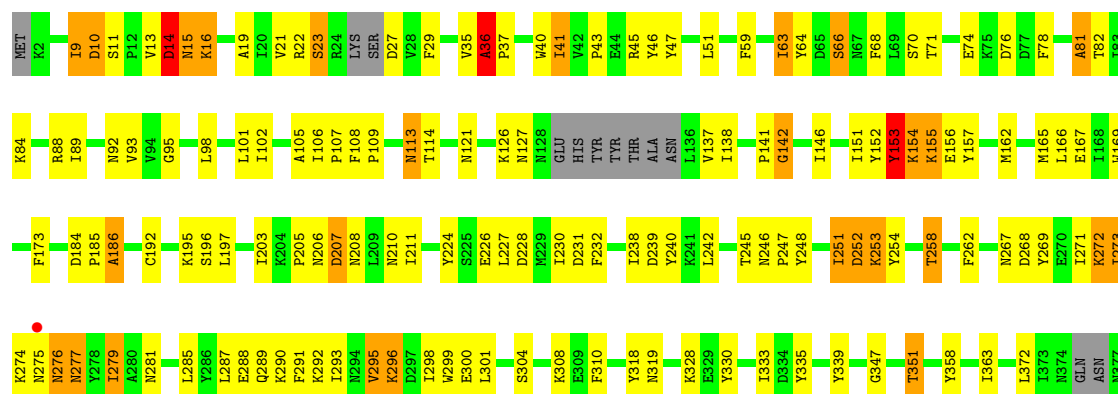
Chain E: 

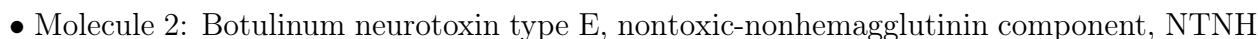
• Molecule 2: Botulinum neurotoxin type E, nontoxic-nonhemagglutinin component, NTNH

Chain B: 



- Molecule 2: Botulinum neurotoxin type E, nontoxic-nonhemagglutinin component, NTNH







ASN	LYS	Q1069	Y1070	V1071	Q1072	D1075	L1082	ASP	GLY	THR	E1086	T1093	T1094	N1095	M1096	R1097	I1098	Q1099	L1100	V1101	D1102	A1107	I1116	F1117	L1122	M1127	M1128	V1129	I1135	K1136	M1137	Q1138	D1139	Y1140	N1141	W1142	V1143	I1144	C1145	D1146	L1147	N1148	H1149	D1150	W1153	I1159	L1160	LYS	ASN	ILE				
L985	N986	K987	T988	I989	N990	S991	R992	E993	I994	L995	L1002	N1003	I1007	R1008	E1012	E1013	I1014	L1015	Q1016	Y1017	N1018	R1019	Y1021	E1022	L1023	F1024	E1030	I1036	E1037	Q1038	ASN	ASN	N1041	I1042	S1045	I1046	ASN	ASN	GLU	ASN	C1145	D1146	L1147	N1148	H1149	D1150	W1153	I1159	L1160	LYS	ASN	ILE		
F899	E900	N901	D910	N914	N922	I923	S924	N925	W928	I931	V932	I933	S934	I935	K939	L943	I944	F945	I946	D947	N948	I949	L950	V951	A952	N953	E954	I956	K957	E958	N961	I962	D966	I967	I968	S969	L970	N971	S972	D973	N974	N975	Y978	I979	E980	S983	V984							
T819	S820	G821	K822	H823	T824	L825	V826	E827	Y828	P829	K830	I831	T843	H844	L845	T846	G847	A848	I852	K853	F854	T855	D857	N858	F859	E860	N861	G862	L863	T864	N865	N866	F867	S868	Y870	F871	L792	W872	L873	L876	K877	Q878	N879	T880	I881	L885	I886	G887	S888	W895	G896	E897		
F608	K609	D610	L611	S612	L615	O620	F624	Y625	L626	I629	Y630	Y631	L634	W638	Y641	S643	Q644	Y645	V656	L657	A658	L662	E676	N677	S678	W679	I680	N684	L687	N694	M709	I710	F713	N716	A717	V721	F722																	
I726	Y727	P728	K729	F730	T731	W734	E735	Q736	C737	I738	K739	N740	I741	N742	T745	N754	I755	N756	E759	K760	L763	Q766	L657	N677	L772	I779	Q780	S788	Y789	T790	E791	L792	L793	I794	K795	S799	P800	L803	S804	L805	Y806	Q811	V815	I816	D818									
T539	E540	T541	O542	E543	I544	S545	H546	GLN	PHE	GLY	D550	T551	K552	L553	I554	P555	W556	I557	G558	R559	N562	T566	N567	N568	S569	F570	V571	E572	E573	F574	K575	P579	T580	S581	I582	L583	K586	E587	N588	K593	I594	K595	ILE	ASP	GLU	L599	P600	S601	G602	N603	L604	N605	PHE	SER

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	192.60Å 192.60Å 286.54Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.30 – 3.05 49.25 – 3.05	Depositor EDS
% Data completeness (in resolution range)	59.7 (49.30-3.05) 59.8 (49.25-3.05)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 3.07Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.243 , 0.321 0.247 , 0.314	Depositor DCC
$R_{free}$ test set	1334 reflections (1.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	110.6	Xtriage
Anisotropy	0.154	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 110.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.36$ , $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	0.135 for -h,-k,l 0.196 for h,-h-k,-l 0.137 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	56952	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	134.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.18% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.58	0/10176	0.74	2/13800 (0.0%)
1	C	0.56	0/10176	0.73	1/13800 (0.0%)
1	E	0.54	0/10176	0.72	2/13800 (0.0%)
2	B	0.64	2/9182 (0.0%)	0.75	0/12454
2	D	0.60	1/9182 (0.0%)	0.74	0/12454
2	F	0.58	1/9182 (0.0%)	0.73	0/12454
All	All	0.58	4/58074 (0.0%)	0.73	5/78762 (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
2	B	0	1
2	D	0	1
2	F	0	2
All	All	0	4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1158	TRP	CB-CG	-7.00	1.37	1.50
2	D	1158	TRP	CB-CG	-5.84	1.39	1.50
2	F	1158	TRP	CB-CG	-5.69	1.40	1.50
2	B	116	ASP	C-O	5.60	1.33	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1143	ARG	NE-CZ-NH2	7.34	123.97	120.30
1	E	1143	ARG	NE-CZ-NH2	7.21	123.90	120.30
1	A	394	ARG	NE-CZ-NH2	-6.68	116.96	120.30
1	C	1143	ARG	NE-CZ-NH2	6.28	123.44	120.30
1	E	394	ARG	NE-CZ-NH2	-5.63	117.48	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	36	ALA	Peptide
2	D	36	ALA	Peptide
2	F	1138	GLN	Peptide
2	F	36	ALA	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	9978	0	9735	338	1
1	C	9978	0	9735	348	1
1	E	9978	0	9735	352	0
2	B	9005	0	8666	361	0
2	D	9005	0	8666	344	0
2	F	9005	0	8666	356	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
3	E	1	0	0	0	0
All	All	56952	0	55203	2034	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 18.

The worst 5 of 2034 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)
2:D:885:LEU:HG	2:D:886:ILE:HD12	1.33	1.09

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:885:LEU:HG	2:F:886:ILE:HD12	1.32	1.09
2:B:436:ASP:HA	2:B:437:ILE:HB	1.34	1.08
2:B:885:LEU:HG	2:B:886:ILE:HD12	1.29	1.08
2:D:436:ASP:HA	2:D:437:ILE:HB	1.35	1.05

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:A:280:ARG:NH2	1:C:420:GLY:O[3_564]	2.16	0.04

## 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	1225/1252 (98%)	999 (82%)	159 (13%)	67 (6%)	2	11
1	C	1225/1252 (98%)	995 (81%)	166 (14%)	64 (5%)	2	12
1	E	1225/1252 (98%)	997 (81%)	162 (13%)	66 (5%)	2	12
2	B	1084/1163 (93%)	832 (77%)	183 (17%)	69 (6%)	1	8
2	D	1084/1163 (93%)	835 (77%)	179 (16%)	70 (6%)	1	8
2	F	1084/1163 (93%)	838 (77%)	177 (16%)	69 (6%)	1	8
All	All	6927/7245 (96%)	5496 (79%)	1026 (15%)	405 (6%)	2	10

5 of 405 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	119	THR
1	A	199	ILE
1	A	291	VAL
1	A	419	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	449	ASN

### 5.3.2 Protein sidechains ⓘ

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	1125/1155 (97%)	1013 (90%)	112 (10%)	9	30
1	C	1125/1155 (97%)	1011 (90%)	114 (10%)	9	30
1	E	1125/1155 (97%)	1014 (90%)	111 (10%)	9	31
2	B	990/1103 (90%)	915 (92%)	75 (8%)	15	46
2	D	990/1103 (90%)	915 (92%)	75 (8%)	15	46
2	F	990/1103 (90%)	914 (92%)	76 (8%)	15	45
All	All	6345/6774 (94%)	5782 (91%)	563 (9%)	11	37

5 of 563 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	348	ARG
1	E	1027	ASP
2	F	570	PHE
1	E	440	GLU
1	E	696	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 110 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	956	HIS
2	B	636	GLN
2	F	677	ASN
1	E	1060	GLN
2	B	171	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	1235/1252 (98%)	-0.69	1 (0%) 95 90	53, 120, 189, 296	0
1	C	1235/1252 (98%)	-0.66	3 (0%) 94 88	46, 129, 207, 365	0
1	E	1235/1252 (98%)	-0.67	5 (0%) 92 81	62, 143, 214, 391	0
2	B	1114/1163 (95%)	-0.69	1 (0%) 95 90	37, 109, 179, 274	0
2	D	1114/1163 (95%)	-0.68	4 (0%) 92 81	48, 129, 194, 278	0
2	F	1114/1163 (95%)	-0.61	10 (0%) 84 66	51, 143, 220, 328	0
All	All	7047/7245 (97%)	-0.67	24 (0%) 93 84	37, 129, 203, 391	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	657	SER	9.4
1	E	655	GLY	7.5
1	E	657	SER	7.4
2	B	275	ASN	6.8
1	E	656	SER	4.7

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	E	1301	1/1	0.98	0.14	-	99,99,99,99	0
3	ZN	A	1301	1/1	0.99	0.18	-	86,86,86,86	0
3	ZN	C	1301	1/1	0.98	0.17	-	83,83,83,83	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.