



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 05:42 am BST

PDB ID : 3WIN
Title : Clostridium botulinum Hemagglutinin
Authors : Amatsu, S.; Sugawara, Y.; Matsumura, T.; Fujinaga, Y.; Kitadokoro, K.
Deposited on : 2013-09-19
Resolution : 3.50 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

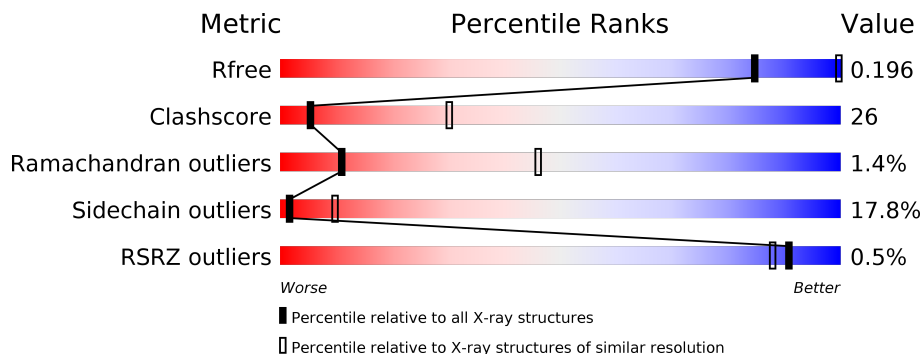
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	D	194	
2	E	431	
3	C	168	
4	A	316	
4	B	316	

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 10522 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 HA3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	167	1361	883	218	258	2	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-5	MET	-	EXPRESSION TAG	UNP Q33CP8
D	-4	ALA	-	EXPRESSION TAG	UNP Q33CP8
D	-3	SER	-	EXPRESSION TAG	UNP Q33CP8
D	-2	TRP	-	EXPRESSION TAG	UNP Q33CP8
D	-1	SER	-	EXPRESSION TAG	UNP Q33CP8
D	0	HIS	-	EXPRESSION TAG	UNP Q33CP8
D	1	PRO	-	EXPRESSION TAG	UNP Q33CP8
D	2	GLN	-	EXPRESSION TAG	UNP Q33CP8
D	3	PHE	-	EXPRESSION TAG	UNP Q33CP8
D	4	GLU	-	EXPRESSION TAG	UNP Q33CP8
D	5	LYS	-	EXPRESSION TAG	UNP Q33CP8
D	6	GLY	-	EXPRESSION TAG	UNP Q33CP8
D	7	ALA	-	EXPRESSION TAG	UNP Q33CP8
D	8	LEU	-	EXPRESSION TAG	UNP Q33CP8
D	9	GLU	-	EXPRESSION TAG	UNP Q33CP8
D	10	VAL	-	EXPRESSION TAG	UNP Q33CP8
D	11	LEU	-	EXPRESSION TAG	UNP Q33CP8
D	12	PHE	-	EXPRESSION TAG	UNP Q33CP8
D	13	GLN	-	EXPRESSION TAG	UNP Q33CP8
D	14	GLY	-	EXPRESSION TAG	UNP Q33CP8
D	15	PRO	-	EXPRESSION TAG	UNP Q33CP8
D	16	GLY	-	EXPRESSION TAG	UNP Q33CP8
D	17	TYR	-	EXPRESSION TAG	UNP Q33CP8
D	18	GLN	-	EXPRESSION TAG	UNP Q33CP8

- Molecule 2 is a protein called HA3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	419	3347	2113	559	671	4	0	0	0

- Molecule 3 is a protein called 17 kD hemagglutinin component.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	141	1156	746	184	221	5	0	0	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-21	MET	-	EXPRESSION TAG	UNP Q45841
C	-20	GLY	-	EXPRESSION TAG	UNP Q45841
C	-19	SER	-	EXPRESSION TAG	UNP Q45841
C	-18	SER	-	EXPRESSION TAG	UNP Q45841
C	-17	HIS	-	EXPRESSION TAG	UNP Q45841
C	-16	HIS	-	EXPRESSION TAG	UNP Q45841
C	-15	HIS	-	EXPRESSION TAG	UNP Q45841
C	-14	HIS	-	EXPRESSION TAG	UNP Q45841
C	-13	HIS	-	EXPRESSION TAG	UNP Q45841
C	-12	HIS	-	EXPRESSION TAG	UNP Q45841
C	-11	SER	-	EXPRESSION TAG	UNP Q45841
C	-10	SER	-	EXPRESSION TAG	UNP Q45841
C	-9	GLY	-	EXPRESSION TAG	UNP Q45841
C	-8	LEU	-	EXPRESSION TAG	UNP Q45841
C	-7	VAL	-	EXPRESSION TAG	UNP Q45841
C	-6	PRO	-	EXPRESSION TAG	UNP Q45841
C	-5	ARG	-	EXPRESSION TAG	UNP Q45841
C	-4	GLY	-	EXPRESSION TAG	UNP Q45841
C	-3	SER	-	EXPRESSION TAG	UNP Q45841
C	-2	HIS	-	EXPRESSION TAG	UNP Q45841
C	-1	MET	-	EXPRESSION TAG	UNP Q45841
C	0	ALA	-	EXPRESSION TAG	UNP Q45841
C	1	SER	-	EXPRESSION TAG	UNP Q45841

- Molecule 4 is a protein called HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	286	2304	1465	390	444	5	0	0	0
4	B	286	2304	1465	390	444	5	0	0	0

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	EXPRESSION TAG	UNP Q33CP6
A	-20	ALA	-	EXPRESSION TAG	UNP Q33CP6
A	-19	SER	-	EXPRESSION TAG	UNP Q33CP6
A	-18	TRP	-	EXPRESSION TAG	UNP Q33CP6
A	-17	SER	-	EXPRESSION TAG	UNP Q33CP6
A	-16	HIS	-	EXPRESSION TAG	UNP Q33CP6
A	-15	PRO	-	EXPRESSION TAG	UNP Q33CP6
A	-14	GLN	-	EXPRESSION TAG	UNP Q33CP6
A	-13	PHE	-	EXPRESSION TAG	UNP Q33CP6
A	-12	GLU	-	EXPRESSION TAG	UNP Q33CP6
A	-11	LYS	-	EXPRESSION TAG	UNP Q33CP6
A	-10	GLY	-	EXPRESSION TAG	UNP Q33CP6
A	-9	ALA	-	EXPRESSION TAG	UNP Q33CP6
A	-8	LEU	-	EXPRESSION TAG	UNP Q33CP6
A	-7	GLU	-	EXPRESSION TAG	UNP Q33CP6
A	-6	VAL	-	EXPRESSION TAG	UNP Q33CP6
A	-5	LEU	-	EXPRESSION TAG	UNP Q33CP6
A	-4	PHE	-	EXPRESSION TAG	UNP Q33CP6
A	-3	GLN	-	EXPRESSION TAG	UNP Q33CP6
A	-2	GLY	-	EXPRESSION TAG	UNP Q33CP6
A	-1	PRO	-	EXPRESSION TAG	UNP Q33CP6
A	0	GLY	-	EXPRESSION TAG	UNP Q33CP6
A	1	TYR	-	EXPRESSION TAG	UNP Q33CP6
A	2	PRO	-	EXPRESSION TAG	UNP Q33CP6
A	3	ASP	-	EXPRESSION TAG	UNP Q33CP6
A	4	ASP	-	EXPRESSION TAG	UNP Q33CP6
A	5	ASP	-	EXPRESSION TAG	UNP Q33CP6
A	6	ASP	-	EXPRESSION TAG	UNP Q33CP6
B	-21	MET	-	EXPRESSION TAG	UNP Q33CP6
B	-20	ALA	-	EXPRESSION TAG	UNP Q33CP6
B	-19	SER	-	EXPRESSION TAG	UNP Q33CP6
B	-18	TRP	-	EXPRESSION TAG	UNP Q33CP6
B	-17	SER	-	EXPRESSION TAG	UNP Q33CP6
B	-16	HIS	-	EXPRESSION TAG	UNP Q33CP6
B	-15	PRO	-	EXPRESSION TAG	UNP Q33CP6
B	-14	GLN	-	EXPRESSION TAG	UNP Q33CP6
B	-13	PHE	-	EXPRESSION TAG	UNP Q33CP6
B	-12	GLU	-	EXPRESSION TAG	UNP Q33CP6
B	-11	LYS	-	EXPRESSION TAG	UNP Q33CP6
B	-10	GLY	-	EXPRESSION TAG	UNP Q33CP6
B	-9	ALA	-	EXPRESSION TAG	UNP Q33CP6
B	-8	LEU	-	EXPRESSION TAG	UNP Q33CP6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-7	GLU	-	EXPRESSION TAG	UNP Q33CP6
B	-6	VAL	-	EXPRESSION TAG	UNP Q33CP6
B	-5	LEU	-	EXPRESSION TAG	UNP Q33CP6
B	-4	PHE	-	EXPRESSION TAG	UNP Q33CP6
B	-3	GLN	-	EXPRESSION TAG	UNP Q33CP6
B	-2	GLY	-	EXPRESSION TAG	UNP Q33CP6
B	-1	PRO	-	EXPRESSION TAG	UNP Q33CP6
B	0	GLY	-	EXPRESSION TAG	UNP Q33CP6
B	1	TYR	-	EXPRESSION TAG	UNP Q33CP6
B	2	PRO	-	EXPRESSION TAG	UNP Q33CP6
B	3	ASP	-	EXPRESSION TAG	UNP Q33CP6
B	4	ASP	-	EXPRESSION TAG	UNP Q33CP6
B	5	ASP	-	EXPRESSION TAG	UNP Q33CP6
B	6	ASP	-	EXPRESSION TAG	UNP Q33CP6

- Molecule 5 is water.

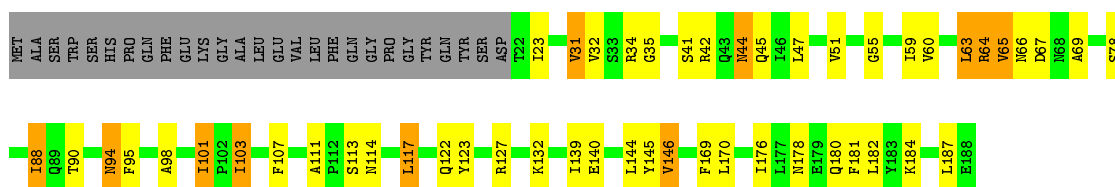
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	2	Total O 2 2	0	0
5	E	19	Total O 19 19	0	0
5	C	7	Total O 7 7	0	0
5	A	11	Total O 11 11	0	0
5	B	11	Total O 11 11	0	0

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.

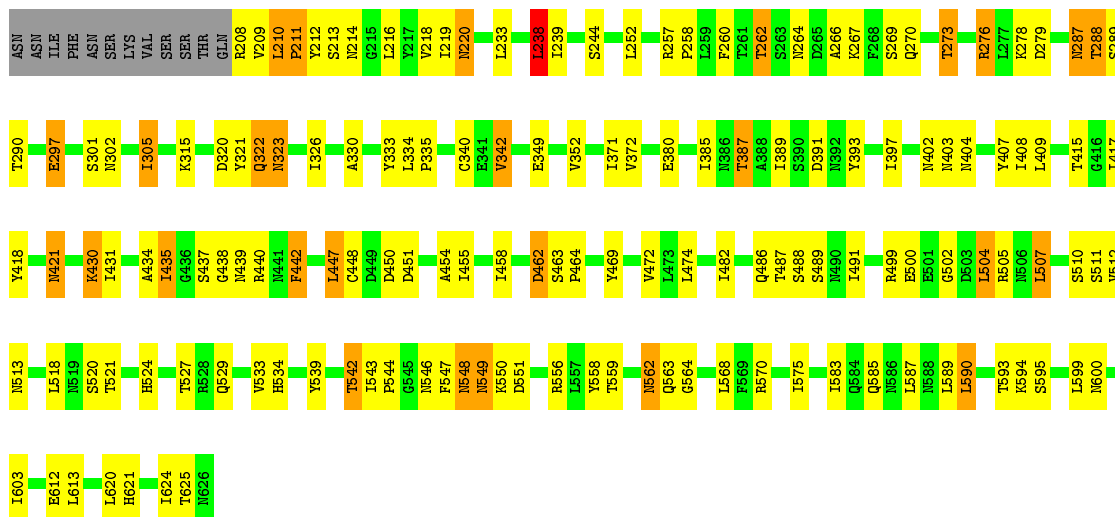
- Molecule 1: HA3

Chain D: 



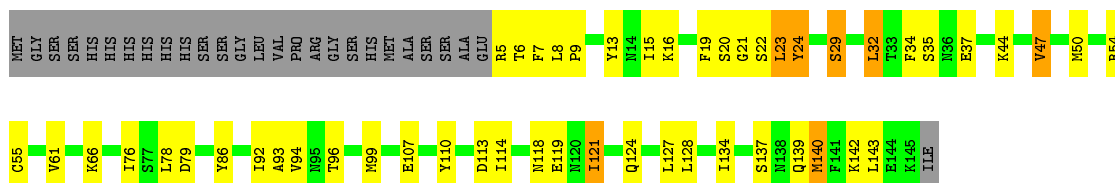
- Molecule 2: HA3

Chain E: 

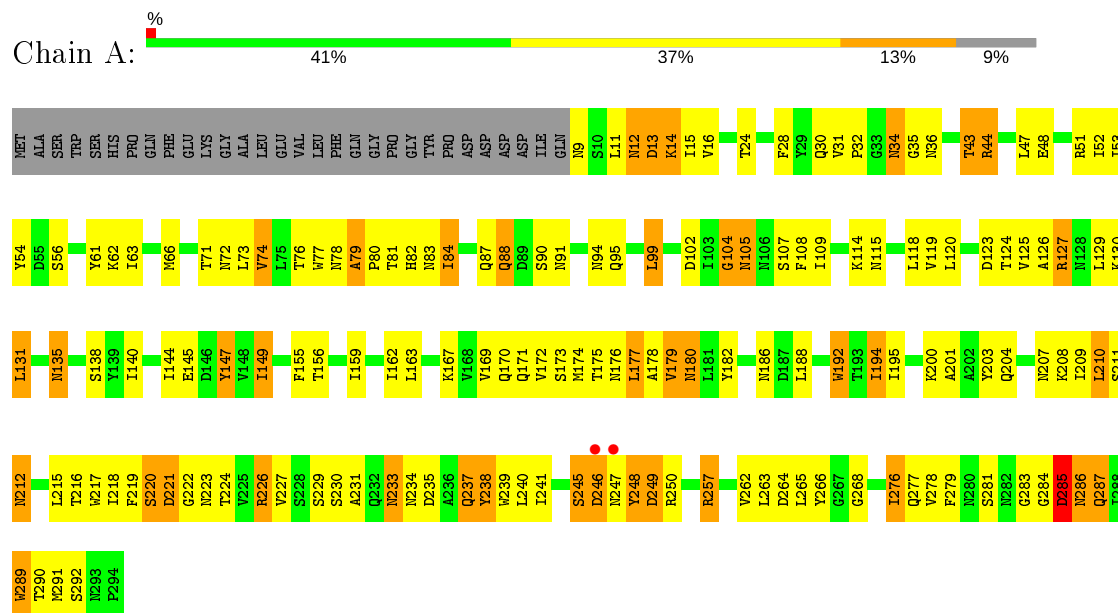


- Molecule 3: 17 kD hemagglutinin component

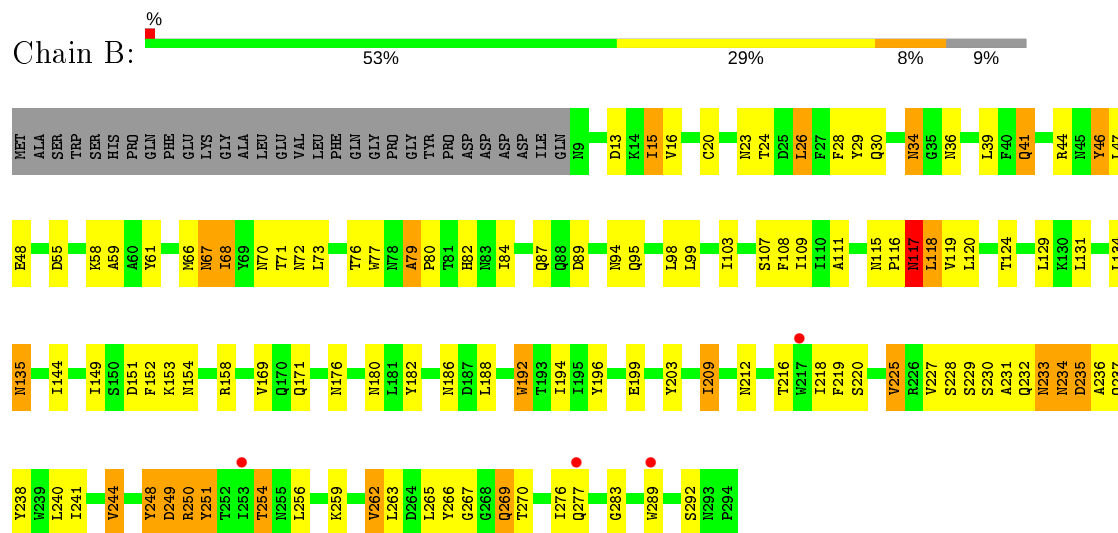
Chain C: 



- Molecule 4: HA1



- Molecule 4: HA1



4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, α , β , γ	324.73Å 324.73Å 117.59Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.43 – 3.50 49.16 – 3.50	Depositor EDS
% Data completeness (in resolution range)	98.2 (48.43-3.50) 98.3 (49.16-3.50)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.19 (at 3.48Å)	Xtrriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.200 , 0.251 0.199 , 0.196	Depositor DCC
R_{free} test set	2268 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	99.8	Xtrriage
Anisotropy	0.160	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 74.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	10522	wwPDB-VP
Average B, all atoms (Å ²)	113.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.98% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	D	0.58	0/1393	0.81	0/1891
2	E	0.57	2/3409 (0.1%)	0.81	4/4636 (0.1%)
3	C	0.58	1/1186 (0.1%)	0.78	0/1611
4	A	0.49	0/2354	0.80	2/3208 (0.1%)
4	B	0.53	0/2354	0.79	0/3208
All	All	0.55	3/10696 (0.0%)	0.80	6/14554 (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
2	E	0	1
4	B	0	2
All	All	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	464	PRO	N-CD	5.32	1.55	1.47
3	C	9	PRO	N-CD	5.04	1.54	1.47
2	E	211	PRO	N-CD	5.00	1.54	1.47

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	285	ASP	N-CA-CB	16.97	141.14	110.60
4	A	285	ASP	CB-CA-C	-9.43	91.55	110.40
2	E	238	LEU	CA-CB-CG	7.30	132.09	115.30
2	E	210	LEU	C-N-CD	5.74	140.45	128.40
2	E	463	SER	C-N-CD	5.65	140.26	128.40

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	15	ILE	Peptide
4	B	244	VAL	Peptide
1	D	35	GLY	Peptide
2	E	438	GLY	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	D	1361	0	1338	37	0
2	E	3347	0	3274	107	0
3	C	1156	0	1113	47	0
4	A	2304	0	2240	256	0
4	B	2304	0	2241	102	0
5	A	11	0	0	1	0
5	B	11	0	0	0	0
5	C	7	0	0	0	0
5	D	2	0	0	0	0
5	E	19	0	0	2	0
All	All	10522	0	10206	542	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 26.

The worst 5 of 542 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:B:46:TYR:CZ	4:B:68:ILE:HD11	1.40	1.56
1:D:65:VAL:CG2	1:D:66:ASN:H	1.11	1.41
4:A:162:ILE:HD11	4:A:250:ARG:NH2	1.22	1.40
4:A:162:ILE:CD1	4:A:250:ARG:NH2	1.90	1.34
4:A:9:ASN:N	4:A:61:TYR:HH	1.36	1.21

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	D	165/194 (85%)	144 (87%)	19 (12%)	2 (1%)	13	50
2	E	417/431 (97%)	392 (94%)	22 (5%)	3 (1%)	22	61
3	C	139/168 (83%)	120 (86%)	18 (13%)	1 (1%)	22	61
4	A	284/316 (90%)	249 (88%)	28 (10%)	7 (2%)	5	34
4	B	284/316 (90%)	250 (88%)	29 (10%)	5 (2%)	8	41
All	All	1289/1425 (90%)	1155 (90%)	116 (9%)	18 (1%)	11	46

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	64	ARG
1	D	65	VAL
4	A	79	ALA
4	A	246	ASP
2	E	288	THR

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	D	150/172 (87%)	128 (85%)	22 (15%)	3	18
2	E	380/392 (97%)	315 (83%)	65 (17%)	2	12
3	C	132/154 (86%)	116 (88%)	16 (12%)	5	24
4	A	257/283 (91%)	199 (77%)	58 (23%)	1	4

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	B	257/283 (91%)	209 (81%)	48 (19%)	1 8
All	All	1176/1284 (92%)	967 (82%)	209 (18%)	2 10

5 of 209 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	47	VAL
4	A	90	SER
4	B	228	SER
3	C	107	GLU
4	A	24	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	60	ASN
4	A	135	ASN
4	B	234	ASN
4	A	82	HIS
4	A	88	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](#)

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	D	167/194 (86%)	-0.35	0 100 100	71, 100, 146, 182	0
2	E	419/431 (97%)	-0.53	0 100 100	55, 100, 138, 186	0
3	C	141/168 (83%)	-0.41	0 100 100	73, 99, 146, 180	0
4	A	286/316 (90%)	-0.24	2 (0%) 87 83	81, 129, 200, 221	0
4	B	286/316 (90%)	-0.12	4 (1%) 75 69	80, 111, 166, 203	0
All	All	1299/1425 (91%)	-0.34	6 (0%) 91 88	55, 106, 175, 221	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	217	TRP	3.0
4	B	253	ILE	2.5
4	A	246	ASP	2.4
4	B	277	GLN	2.3
4	B	289	TRP	2.1

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

There are no such residues in this entry.