



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 03:45 am GMT

PDB ID : 4QD2  
Title : Molecular basis for disruption of E-cadherin adhesion by botulinum neurotoxin A complex  
Authors : Lee, K.; Zhong, X.; Gu, S.; Krueel, A.; Dorner, M.B.; Perry, K.; Rummel, A.; Dong, M.; Jin, R.  
Deposited on : 2014-05-13  
Resolution : 2.40 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](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

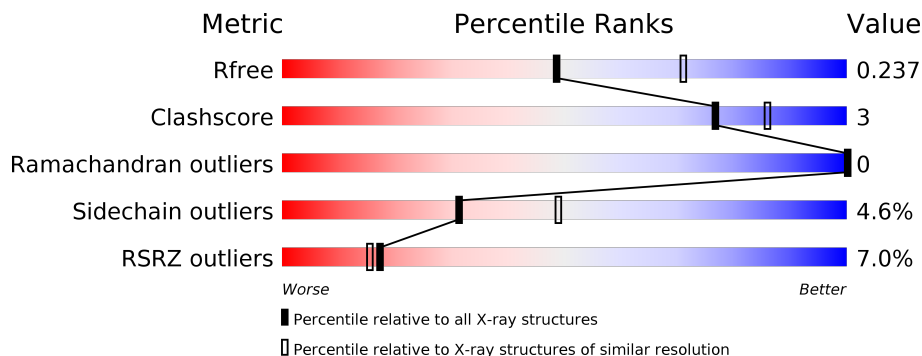
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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))
$R_{free}$	100719	3166 (2.40-2.40)
Clashscore	112137	3674 (2.40-2.40)
Ramachandran outliers	110173	3616 (2.40-2.40)
Sidechain outliers	110143	3617 (2.40-2.40)
RSRZ outliers	101464	3195 (2.40-2.40)

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  $\geq 3$ , 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	C	296	<div> <div>7%</div> <div>85%</div> <div>11%</div> <div>..</div> </div>
1	D	296	<div> <div>21%</div> <div>80%</div> <div>16%</div> <div>..</div> </div>
1	H	296	<div> <div>7%</div> <div>84%</div> <div>12%</div> <div>..</div> </div>
1	I	296	<div> <div>2%</div> <div>86%</div> <div>9%</div> <div>..</div> </div>
2	B	147	<div> <div>%</div> <div>88%</div> <div>10%</div> <div>.</div> </div>
2	G	147	<div> <div>%</div> <div>86%</div> <div>11%</div> <div>.</div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	A	254	
3	F	254	
4	E	213	
4	J	213	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	CA	E	301	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 19012 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 Hemagglutinin component HA33.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	286	Total	C	N	O	S	0	0	0
			2333	1476	399	455	3			
1	C	286	Total	C	N	O	S	0	0	0
			2333	1476	399	455	3			
1	I	286	Total	C	N	O	S	0	0	0
			2333	1476	399	455	3			
1	H	286	Total	C	N	O	S	0	0	0
			2333	1476	399	455	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	294	PRO	-	EXPRESSION TAG	UNP A5HZZ6
D	295	GLY	-	EXPRESSION TAG	UNP A5HZZ6
D	296	SER	-	EXPRESSION TAG	UNP A5HZZ6
D	297	ALA	-	EXPRESSION TAG	UNP A5HZZ6
C	294	PRO	-	EXPRESSION TAG	UNP A5HZZ6
C	295	GLY	-	EXPRESSION TAG	UNP A5HZZ6
C	296	SER	-	EXPRESSION TAG	UNP A5HZZ6
C	297	ALA	-	EXPRESSION TAG	UNP A5HZZ6
I	294	PRO	-	EXPRESSION TAG	UNP A5HZZ6
I	295	GLY	-	EXPRESSION TAG	UNP A5HZZ6
I	296	SER	-	EXPRESSION TAG	UNP A5HZZ6
I	297	ALA	-	EXPRESSION TAG	UNP A5HZZ6
H	294	PRO	-	EXPRESSION TAG	UNP A5HZZ6
H	295	GLY	-	EXPRESSION TAG	UNP A5HZZ6
H	296	SER	-	EXPRESSION TAG	UNP A5HZZ6
H	297	ALA	-	EXPRESSION TAG	UNP A5HZZ6

- Molecule 2 is a protein called Hemagglutinin component HA17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	143	Total	C	N	O	S	0	0	0
			1180	762	188	225	5			
2	G	143	Total	C	N	O	S	0	0	0
			1180	762	188	225	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	EXPRESSION TAG	UNP A5HZZ5
B	1	PRO	-	EXPRESSION TAG	UNP A5HZZ5
G	0	GLY	-	EXPRESSION TAG	UNP A5HZZ5
G	1	PRO	-	EXPRESSION TAG	UNP A5HZZ5

- Molecule 3 is a protein called Hemagglutinin component HA70.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	224	Total	C	N	O	S	0	0	0
			1825	1154	310	360	1			
3	F	225	Total	C	N	O	S	0	0	0
			1833	1158	312	362	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	373	GLY	-	EXPRESSION TAG	UNP A5HZZ4
A	374	PRO	-	EXPRESSION TAG	UNP A5HZZ4
A	375	LEU	-	EXPRESSION TAG	UNP A5HZZ4
A	376	GLY	-	EXPRESSION TAG	UNP A5HZZ4
A	377	SER	-	EXPRESSION TAG	UNP A5HZZ4
F	373	GLY	-	EXPRESSION TAG	UNP A5HZZ4
F	374	PRO	-	EXPRESSION TAG	UNP A5HZZ4
F	375	LEU	-	EXPRESSION TAG	UNP A5HZZ4
F	376	GLY	-	EXPRESSION TAG	UNP A5HZZ4
F	377	SER	-	EXPRESSION TAG	UNP A5HZZ4

- Molecule 4 is a protein called Cadherin-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	212	Total	C	N	O	S	0	0	0
			1627	1023	266	334	4			
4	J	212	Total	C	N	O	S	0	0	0
			1627	1023	266	334	4			

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	J	3	Total 3	Ca 3	0	0
5	E	3	Total 3	Ca 3	0	0

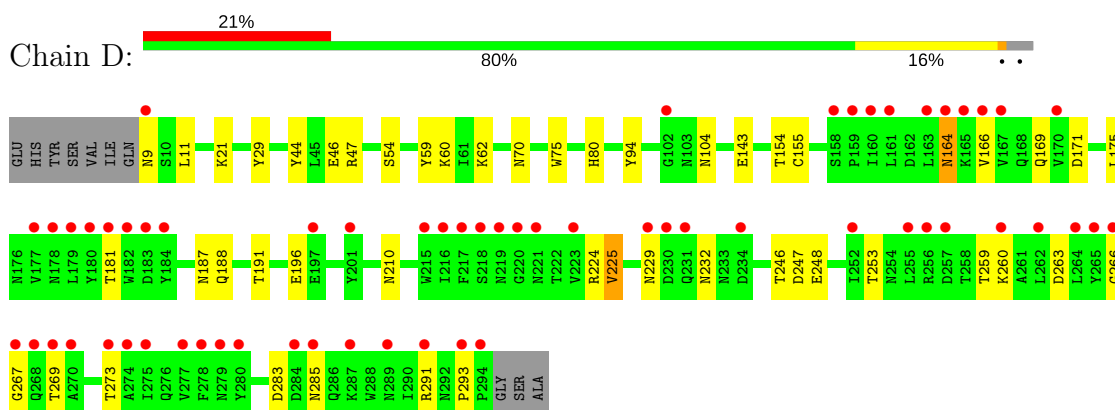
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	37	Total 37	O 37	0	0
6	B	44	Total 44	O 44	0	0
6	C	47	Total 47	O 47	0	0
6	A	31	Total 31	O 31	0	0
6	E	46	Total 46	O 46	0	0
6	I	50	Total 50	O 50	0	0
6	G	50	Total 50	O 50	0	0
6	H	43	Total 43	O 43	0	0
6	F	28	Total 28	O 28	0	0
6	J	26	Total 26	O 26	0	0

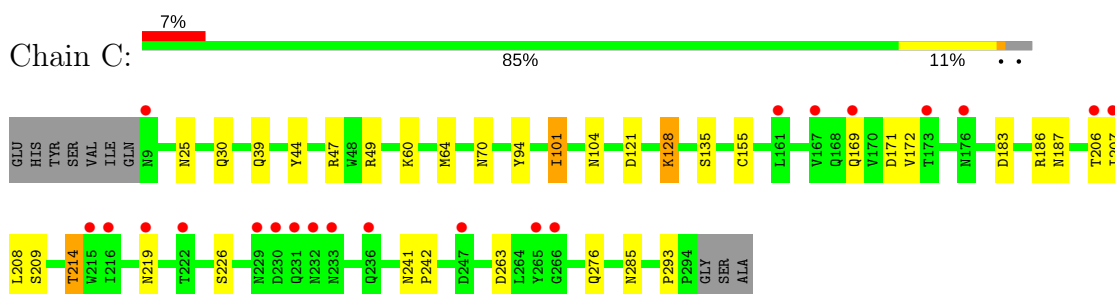
### 3 Residue-property plots [i](#)

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.

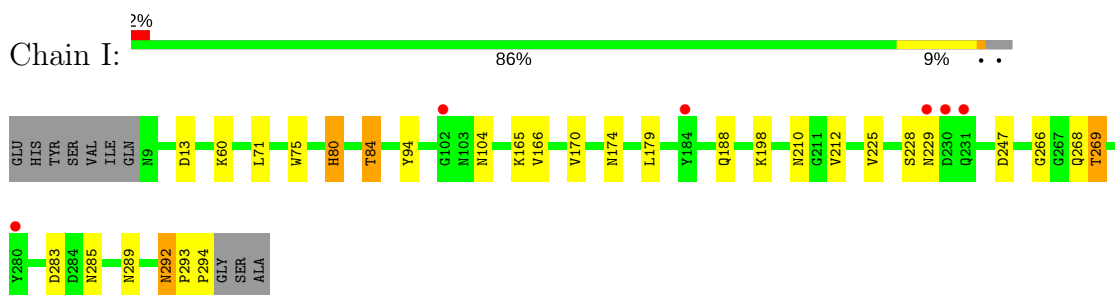
- Molecule 1: Hemagglutinin component HA33



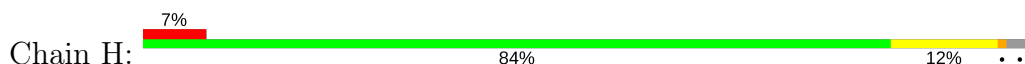
- Molecule 1: Hemagglutinin component HA33

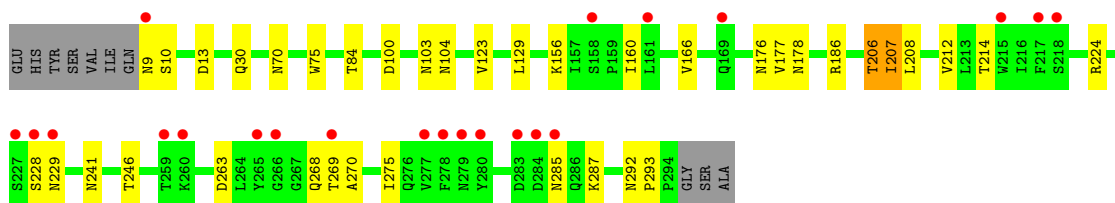


- Molecule 1: Hemagglutinin component HA33

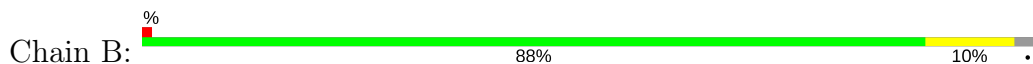


- Molecule 1: Hemagglutinin component HA33

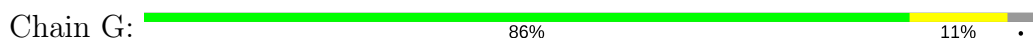




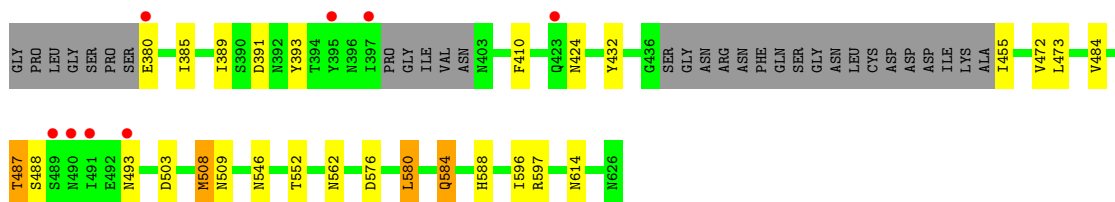
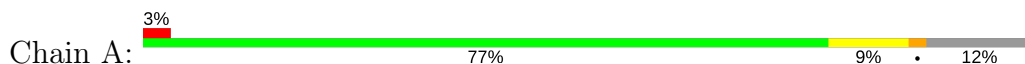
• Molecule 2: Hemagglutinin component HA17



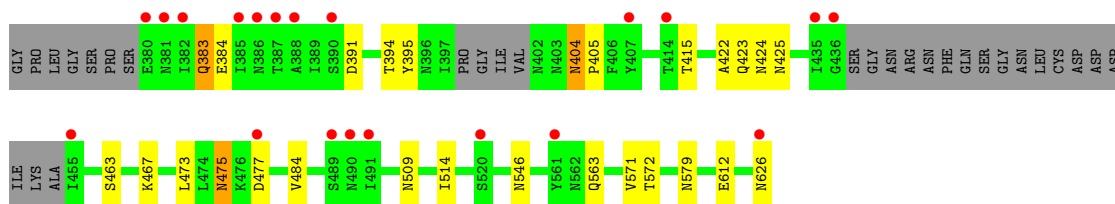
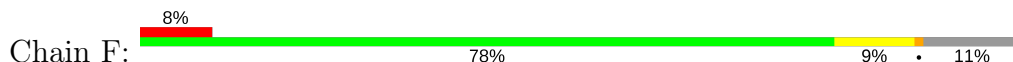
• Molecule 2: Hemagglutinin component HA17



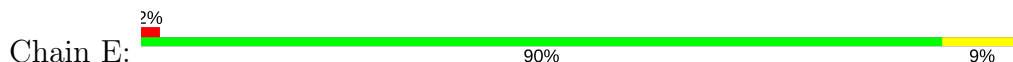
• Molecule 3: Hemagglutinin component HA70



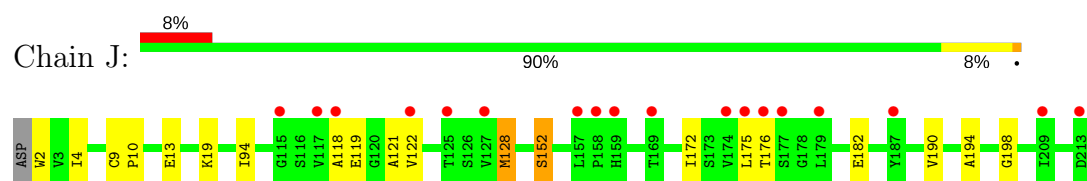
• Molecule 3: Hemagglutinin component HA70



• Molecule 4: Cadherin-1



• Molecule 4: Cadherin-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.99Å 95.73Å 96.14Å 116.89° 96.34° 93.70°	Depositor
Resolution (Å)	47.73 – 2.40 47.73 – 2.40	Depositor EDS
% Data completeness (in resolution range)	87.7 (47.73-2.40) 81.9 (47.73-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.96 (at 2.39Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.2_869)	Depositor
R, $R_{free}$	0.208 , 0.247 0.197 , 0.237	Depositor DCC
$R_{free}$ test set	5230 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	47.5	Xtriage
Anisotropy	0.273	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 41.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.009 for -h,-l,-k	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	19012	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.12% 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:  
CA

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	C	0.23	0/2384	0.41	0/3255
1	D	0.23	0/2384	0.42	0/3255
1	H	0.23	0/2384	0.42	0/3255
1	I	0.24	0/2384	0.42	0/3255
2	B	0.25	0/1210	0.41	0/1643
2	G	0.25	0/1210	0.42	0/1643
3	A	0.24	0/1857	0.45	1/2527 (0.0%)
3	F	0.23	0/1865	0.44	1/2538 (0.0%)
4	E	0.23	0/1658	0.42	0/2263
4	J	0.22	0/1658	0.42	0/2263
All	All	0.23	0/18994	0.42	2/25897 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	391	ASP	CB-CG-OD2	5.23	123.01	118.30
3	A	391	ASP	CB-CG-OD2	5.12	122.91	118.30

There are no chirality outliers.

There are no planarity outliers.

### 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	C	2333	0	2250	16	0
1	D	2333	0	2250	17	0
1	H	2333	0	2250	15	0
1	I	2333	0	2250	13	0
2	B	1180	0	1141	16	0
2	G	1180	0	1141	14	0
3	A	1825	0	1779	14	0
3	F	1833	0	1785	10	0
4	E	1627	0	1595	10	0
4	J	1627	0	1595	11	0
5	E	3	0	0	0	0
5	J	3	0	0	0	0
6	A	31	0	0	0	0
6	B	44	0	0	0	0
6	C	47	0	0	3	0
6	D	37	0	0	1	0
6	E	46	0	0	1	0
6	F	28	0	0	0	0
6	G	50	0	0	0	0
6	H	43	0	0	2	0
6	I	50	0	0	0	0
6	J	26	0	0	1	0
All	All	19012	0	18036	127	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 3.

The worst 5 of 127 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:B:145:LYS:CG	2:B:146:ILE:HG22	1.80	1.10
2:B:145:LYS:HG2	2:B:146:ILE:CG2	1.93	0.98
2:B:145:LYS:HG2	2:B:146:ILE:HG22	0.97	0.95
2:G:145:LYS:HB3	2:G:146:ILE:CG2	2.10	0.82
1:C:64:MET:SD	6:C:333:HOH:O	2.43	0.75

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	C	284/296 (96%)	270 (95%)	14 (5%)	0	100	100
1	D	284/296 (96%)	270 (95%)	14 (5%)	0	100	100
1	H	284/296 (96%)	266 (94%)	18 (6%)	0	100	100
1	I	284/296 (96%)	270 (95%)	14 (5%)	0	100	100
2	B	141/147 (96%)	137 (97%)	4 (3%)	0	100	100
2	G	141/147 (96%)	134 (95%)	7 (5%)	0	100	100
3	A	218/254 (86%)	207 (95%)	11 (5%)	0	100	100
3	F	219/254 (86%)	202 (92%)	17 (8%)	0	100	100
4	E	210/213 (99%)	204 (97%)	6 (3%)	0	100	100
4	J	210/213 (99%)	205 (98%)	5 (2%)	0	100	100
All	All	2275/2412 (94%)	2165 (95%)	110 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 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	C	262/270 (97%)	250 (95%)	12 (5%)	31	49
1	D	262/270 (97%)	239 (91%)	23 (9%)	12	17
1	H	262/270 (97%)	247 (94%)	15 (6%)	24	38
1	I	262/270 (97%)	247 (94%)	15 (6%)	24	38

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	135/138 (98%)	131 (97%)	4 (3%)	46	67
2	G	135/138 (98%)	132 (98%)	3 (2%)	57	76
3	A	208/232 (90%)	199 (96%)	9 (4%)	33	52
3	F	209/232 (90%)	199 (95%)	10 (5%)	30	47
4	E	183/184 (100%)	181 (99%)	2 (1%)	78	90
4	J	183/184 (100%)	179 (98%)	4 (2%)	57	76
All	All	2101/2188 (96%)	2004 (95%)	97 (5%)	31	49

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	493	ASN
1	I	165	LYS
3	F	477	ASP
3	A	580	LEU
4	E	156	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	579	ASN
3	A	584	GLN
1	H	187	ASN
3	A	509	ASN
1	I	229	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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 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

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 [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	C	286/296 (96%)	0.35	21 (7%) 16 14	34, 66, 105, 155	0
1	D	286/296 (96%)	0.98	61 (21%) 1 1	34, 72, 150, 180	0
1	H	286/296 (96%)	0.39	22 (7%) 14 12	28, 67, 125, 158	0
1	I	286/296 (96%)	0.08	6 (2%) 64 61	31, 55, 91, 127	0
2	B	143/147 (97%)	-0.04	1 (0%) 87 86	31, 44, 67, 95	0
2	G	143/147 (97%)	0.06	0 100 100	29, 44, 67, 92	0
3	A	224/254 (88%)	0.25	8 (3%) 43 42	34, 59, 111, 149	0
3	F	225/254 (88%)	0.38	20 (8%) 10 9	32, 60, 118, 141	0
4	E	212/213 (99%)	0.01	5 (2%) 59 56	34, 57, 98, 116	0
4	J	212/213 (99%)	0.33	18 (8%) 11 10	37, 59, 126, 138	0
All	All	2303/2412 (95%)	0.32	162 (7%) 17 15	28, 57, 120, 180	0

The worst 5 of 162 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	215	TRP	8.3
1	H	228	SER	7.8
1	D	160	ILE	7.7
1	C	231	GLN	7.2
3	F	387	THR	7.1

### 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](#)

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(Å <sup>2</sup> )	Q<0.9
5	CA	E	301	1/1	0.95	0.17	4.99	44,44,44,44	0
5	CA	E	302	1/1	0.96	0.14	1.48	36,36,36,36	0
5	CA	E	303	1/1	0.99	0.14	0.28	42,42,42,42	0
5	CA	J	301	1/1	0.97	0.12	-0.65	45,45,45,45	0
5	CA	J	302	1/1	0.99	0.12	-0.73	38,38,38,38	0
5	CA	J	303	1/1	0.91	0.07	-3.51	45,45,45,45	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.