



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 19, 2024 – 09:03 PM EDT

PDB ID : 4EDB  
Title : Structures of monomeric hemagglutinin and its complex with an Fab fragment of a neutralizing antibody that binds to H1 subtype influenza viruses: molecular basis of infectivity of 2009 pandemic H1N1 influenza A viruses  
Authors : Kim, K.H.; Cho, K.J.; Lee, J.H.; Park, Y.H.; Khan, T.G.; Lee, J.Y.; Kang, S.H.; Alam, I.  
Deposited on : 2012-03-27  
Resolution : 2.50 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](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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

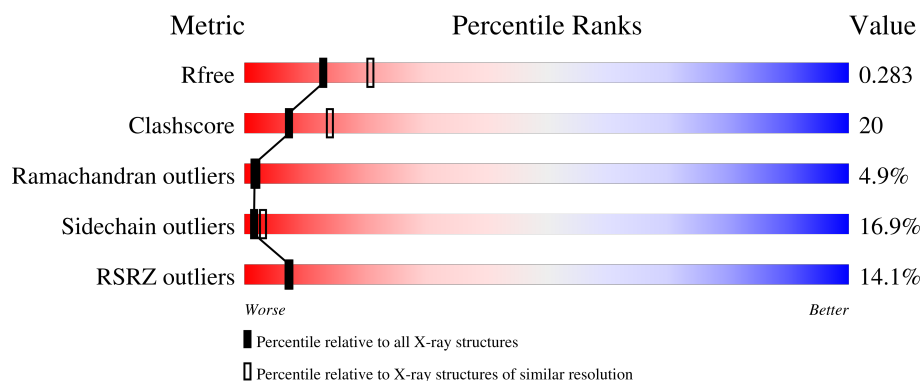
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	164625	5504 (2.50-2.50)
Clashscore	180529	6282 (2.50-2.50)
Ramachandran outliers	177936	6191 (2.50-2.50)
Sidechain outliers	177891	6193 (2.50-2.50)
RSRZ outliers	164620	5504 (2.50-2.50)

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 of 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	330	
1	C	330	
1	E	330	
2	B	182	
2	D	182	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	182	<div><div></div><div>5%</div><div>54%</div><div>27%</div><div>7%</div><div>12%</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11449 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	323	Total	C	N	O	S	0	0	0
			2530	1598	438	482	12			
1	C	323	Total	C	N	O	S	0	0	0
			2530	1598	438	482	12			
1	E	323	Total	C	N	O	S	0	0	0
			2530	1598	438	482	12			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	ALA	-	expression tag	UNP A7LI25
A	-2	ASP	-	expression tag	UNP A7LI25
A	-1	PRO	-	expression tag	UNP A7LI25
A	0	GLY	-	expression tag	UNP A7LI25
C	-3	ALA	-	expression tag	UNP A7LI25
C	-2	ASP	-	expression tag	UNP A7LI25
C	-1	PRO	-	expression tag	UNP A7LI25
C	0	GLY	-	expression tag	UNP A7LI25
E	-3	ALA	-	expression tag	UNP A7LI25
E	-2	ASP	-	expression tag	UNP A7LI25
E	-1	PRO	-	expression tag	UNP A7LI25
E	0	GLY	-	expression tag	UNP A7LI25

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	160	Total	C	N	O	S	0	0	0
			1281	804	218	252	7			
2	D	160	Total	C	N	O	S	0	0	0
			1281	804	218	252	7			
2	F	160	Total	C	N	O	S	0	0	0
			1281	804	218	252	7			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	177	ARG	-	expression tag	UNP A7LI25
B	178	SER	-	expression tag	UNP A7LI25
B	179	LEU	-	expression tag	UNP A7LI25
B	180	VAL	-	expression tag	UNP A7LI25
B	181	PRO	-	expression tag	UNP A7LI25
B	182	ARG	-	expression tag	UNP A7LI25
D	177	ARG	-	expression tag	UNP A7LI25
D	178	SER	-	expression tag	UNP A7LI25
D	179	LEU	-	expression tag	UNP A7LI25
D	180	VAL	-	expression tag	UNP A7LI25
D	181	PRO	-	expression tag	UNP A7LI25
D	182	ARG	-	expression tag	UNP A7LI25
F	177	ARG	-	expression tag	UNP A7LI25
F	178	SER	-	expression tag	UNP A7LI25
F	179	LEU	-	expression tag	UNP A7LI25
F	180	VAL	-	expression tag	UNP A7LI25
F	181	PRO	-	expression tag	UNP A7LI25
F	182	ARG	-	expression tag	UNP A7LI25

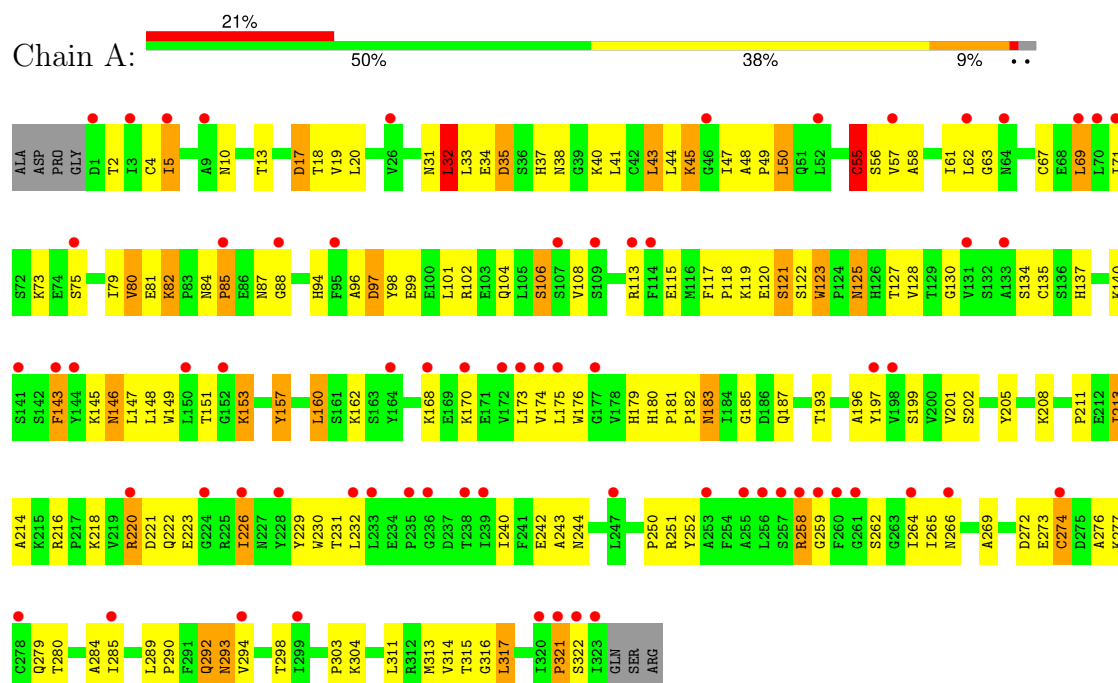
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	8	Total O 8 8	0	0
3	B	2	Total O 2 2	0	0
3	C	1	Total O 1 1	0	0
3	D	2	Total O 2 2	0	0
3	E	2	Total O 2 2	0	0
3	F	1	Total O 1 1	0	0

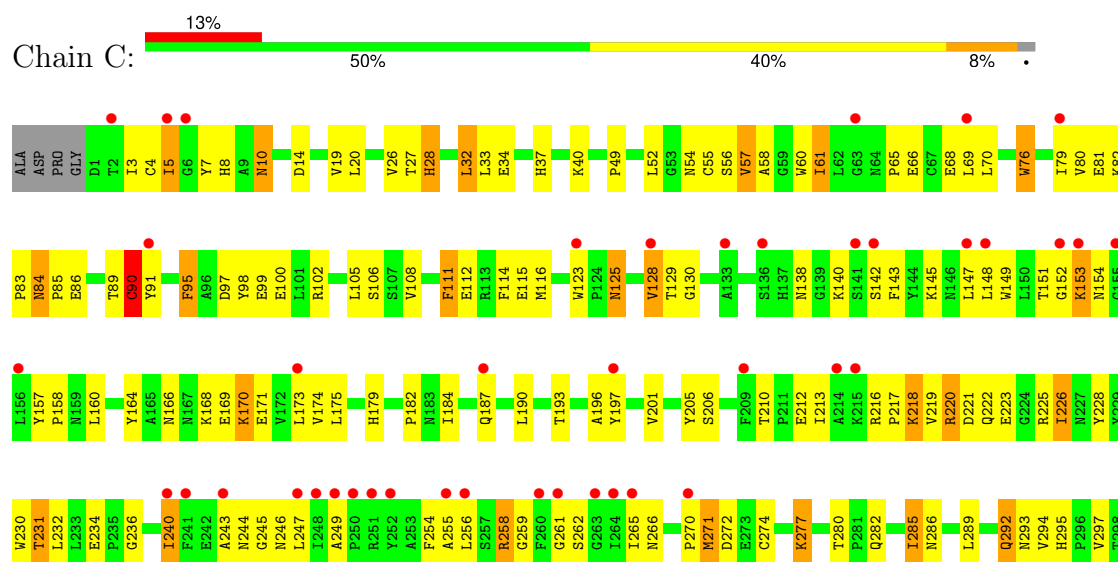
### 3 Residue-property plots [i](#)

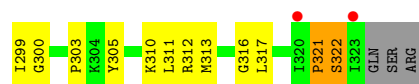
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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

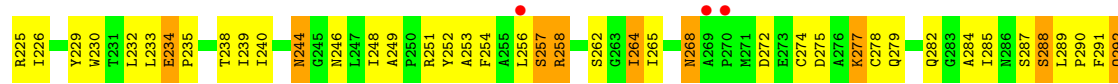
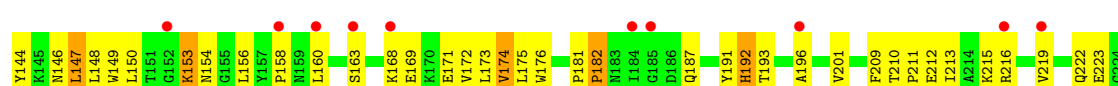
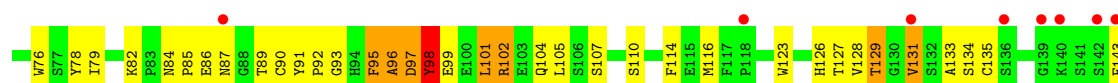


#### • Molecule 1: Hemagglutinin

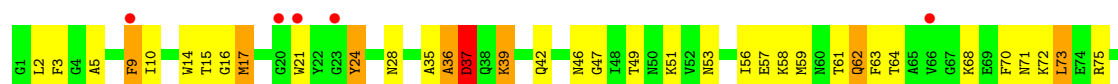


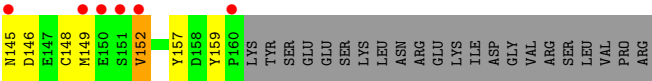


• Molecule 1: Hemagglutinin

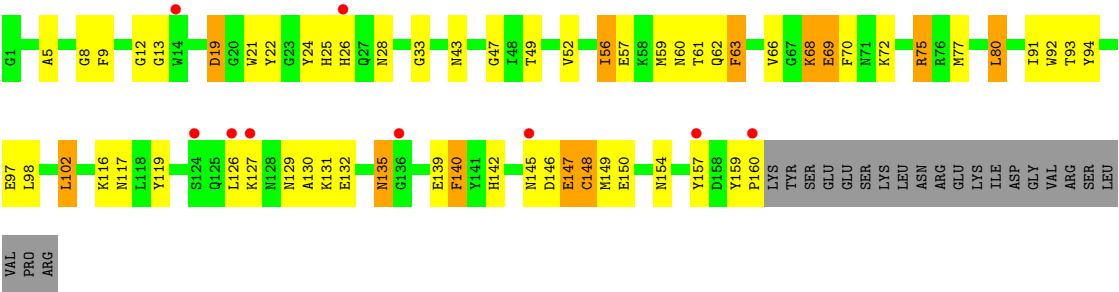


• Molecule 2: Hemagglutinin





● Molecule 2: Hemagglutinin





## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	217.12Å 217.12Å 266.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.34 – 2.50 44.34 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.8 (44.34-2.50) 99.5 (44.34-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.239 , 0.274 0.252 , 0.283	Depositor DCC
$R_{free}$ test set	4105 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.3	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 40.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.085 for -2/3*h-1/3*k+2/3*l,-1/3*h-2/3*k-2/3*l,2/3*h-2/3*k+1/3*l 0.085 for -h,1/3*h-1/3*k+2/3*l,2/3*h+4/3*k+1/3*l 0.077 for -1/3*h+1/3*k-2/3*l,-k,-4/3*h-2/3*k+1/3*l	Xtriage
Reported twinning fraction	0.249 for H, K, L 0.250 for -2/3H-1/3K+2/3L, -1/3H-2/3K-2/3L, 2/3H-2/3K+1/ 0.247 for -1/3H+1/3K-2/3L, -K, -4/3H-2/3K+1/3L 0.254 for -H, 1/3H-1/3K+2/3L, 2/3H+4/3K+1/3L	Depositor
Outliers	1 of 83091 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	11449	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% 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	0.49	0/2596	0.60	0/3532
1	C	0.49	2/2596 (0.1%)	0.58	0/3532
1	E	0.49	2/2596 (0.1%)	0.62	0/3532
2	B	0.47	0/1307	0.56	0/1759
2	D	0.49	1/1307 (0.1%)	0.56	0/1759
2	F	0.47	0/1307	0.54	0/1759
All	All	0.49	5/11709 (0.0%)	0.59	0/15873

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	E	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	14	TRP	CD2-CE2	5.21	1.47	1.41
1	C	149	TRP	CD2-CE2	5.17	1.47	1.41
1	C	76	TRP	CD2-CE2	5.11	1.47	1.41
1	E	123	TRP	CD2-CE2	5.10	1.47	1.41
1	E	149	TRP	CD2-CE2	5.04	1.47	1.41

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	98	TYR	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	2530	0	2455	100	0
1	C	2530	0	2455	115	0
1	E	2530	0	2455	119	0
2	B	1281	0	1209	63	0
2	D	1281	0	1209	72	0
2	F	1281	0	1209	49	0
3	A	8	0	0	0	0
3	B	2	0	0	0	0
3	C	1	0	0	0	0
3	D	2	0	0	0	0
3	E	2	0	0	0	0
3	F	1	0	0	0	0
All	All	11449	0	10992	452	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 20.

The worst 5 of 452 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:59:MET:HB2	2:F:94:TYR:CD1	2.03	0.93
1:E:37:HIS:HD2	1:E:294:VAL:HG21	1.36	0.89
1:E:101:LEU:HA	1:E:104:GLN:OE1	1.75	0.85
2:B:59:MET:HB2	2:F:94:TYR:HD1	1.39	0.85
1:C:3:ILE:O	2:D:137:CYS:HA	1.74	0.85

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	A	321/330 (97%)	262 (82%)	44 (14%)	15 (5%)	2	2
1	C	321/330 (97%)	256 (80%)	45 (14%)	20 (6%)	1	1
1	E	321/330 (97%)	251 (78%)	55 (17%)	15 (5%)	2	2
2	B	158/182 (87%)	124 (78%)	31 (20%)	3 (2%)	6	12
2	D	158/182 (87%)	120 (76%)	27 (17%)	11 (7%)	1	1
2	F	158/182 (87%)	128 (81%)	23 (15%)	7 (4%)	2	2
All	All	1437/1536 (94%)	1141 (79%)	225 (16%)	71 (5%)	2	2

5 of 71 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	32	LEU
1	A	258	ARG
2	B	36	ALA
2	B	37	ASP
1	C	169	GLU

### 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	283/288 (98%)	236 (83%)	47 (17%)	2	3
1	C	283/288 (98%)	242 (86%)	41 (14%)	2	5
1	E	283/288 (98%)	231 (82%)	52 (18%)	1	2
2	B	136/157 (87%)	111 (82%)	25 (18%)	1	2
2	D	136/157 (87%)	109 (80%)	27 (20%)	1	2
2	F	136/157 (87%)	116 (85%)	20 (15%)	2	5
All	All	1257/1335 (94%)	1045 (83%)	212 (17%)	1	3

5 of 212 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	15	THR
1	E	20	LEU
2	F	68	LYS
2	D	34	TYR
2	D	106	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	292	GLN
2	F	28	ASN
2	D	53	ASN
2	F	135	ASN
1	E	286	ASN

### 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 oligosaccharides 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	323/330 (97%)	1.32	68 (21%) 3 3	37, 51, 65, 80	0
1	C	323/330 (97%)	0.95	44 (13%) 8 7	27, 48, 58, 67	0
1	E	323/330 (97%)	0.86	31 (9%) 15 14	28, 45, 59, 64	0
2	B	160/182 (87%)	0.83	14 (8%) 17 16	29, 48, 63, 70	0
2	D	160/182 (87%)	1.14	38 (23%) 2 3	20, 56, 79, 87	0
2	F	160/182 (87%)	0.69	9 (5%) 31 29	27, 45, 58, 71	0
All	All	1449/1536 (94%)	0.99	204 (14%) 7 7	20, 49, 64, 87	0

The worst 5 of 204 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	160	PRO	6.2
2	D	30	GLN	5.9
1	C	197	TYR	5.1
1	A	69	LEU	4.8
1	A	198	VAL	4.5

### 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 monosaccharides 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.