Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2025 – 04:43 pm BST

PDB ID	:	3ZPA / pdb_00003zpa
Title	:	INFLUENZA VIRUS (VN1194) H5 I155F mutant HA
Authors	:	Liu, J.; Chen, Z.; Stevens, D.J.; Gamblin, S.J.; Skehel, J.J.
Deposited on	:	2013-02-27
Resolution	:	2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i) 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 (1)) were used in the production of this report:

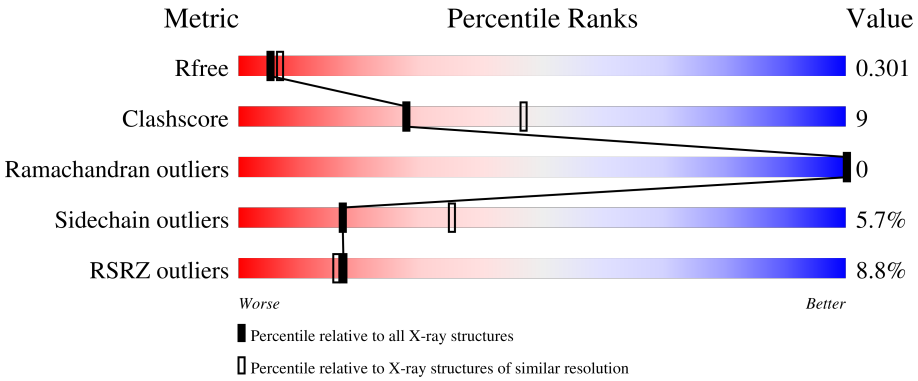
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.44

1 Overall quality at a glance (i)

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 >=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 <=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	E	340	6% 77% 14% · 6%
2	F	160	12% 83% 13% ··

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
3	NAG	E	1326	X	-	-	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3908 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	E	321	Total	C	N	O	S	0	0	0
			2552	1614	440	483	15			

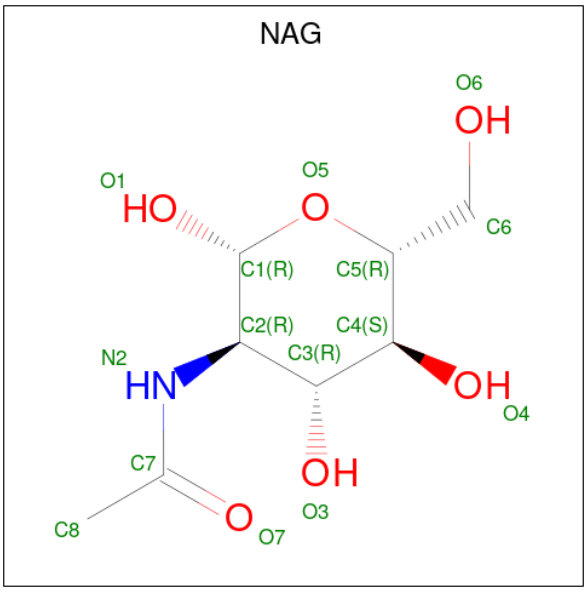
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	40	LYS	THR	conflict	UNP Q6DQ34
E	138	VAL	ALA	conflict	UNP Q6DQ34
E	155	PHE	ILE	engineered mutation	UNP Q6DQ34

- Molecule 2 is a protein called HEMAGGLUTININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	158	Total	C	N	O	S	0	0	0
			1272	791	221	252	8			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (CCD ID: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	1	Total C N O 14 8 1 5	0	0

- Molecule 4 is water.

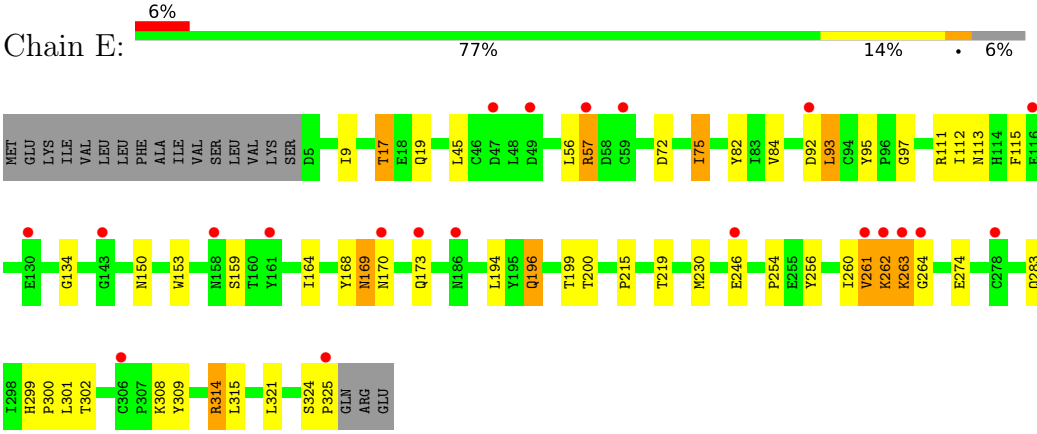
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	39	Total O 39 39	0	0
4	F	31	Total O 31 31	0	0

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.

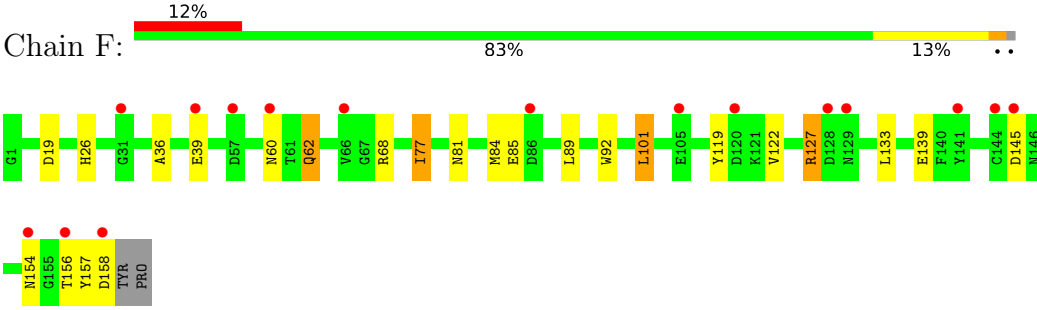
• Molecule 1: HEMAGGLUTININ

Chain E:



• Molecule 2: HEMAGGLUTININ

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	101.00Å 101.00Å 448.24Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.50 30.00 – 2.50	Depositor EDS
% Data completeness (in resolution range)	93.9 (30.00-2.50) 93.8 (30.00-2.50)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	9.03 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.244 , 0.286 0.269 , 0.301	Depositor DCC
R_{free} test set	1474 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	58.7	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 24.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	3908	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: NAG

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	E	0.53	0/2615	0.81	2/3551 (0.1%)
2	F	0.54	0/1296	0.87	1/1742 (0.1%)
All	All	0.53	0/3911	0.83	3/5293 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	194	LEU	N-CA-C	7.27	118.99	111.14
2	F	127	ARG	CB-CA-C	-6.62	108.95	116.63
1	E	57	ARG	CB-CA-C	-5.67	109.51	117.23

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	E	2552	0	2487	49	0
2	F	1272	0	1186	23	0
3	E	14	0	13	0	0
4	E	39	0	0	2	0
4	F	31	0	0	4	0
All	All	3908	0	3686	67	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 9.

All (67) 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:E:113:ASN:OD1	1:E:263:LYS:CB	1.80	1.29
1:E:113:ASN:OD1	1:E:263:LYS:HB3	0.94	1.11
1:E:262:LYS:O	1:E:262:LYS:HG2	1.50	1.05
1:E:159:SER:O	1:E:196:GLN:OE1	1.76	1.02
1:E:261:VAL:HG22	1:E:261:VAL:O	1.63	0.99
2:F:77:ILE:HD13	4:F:2011:HOH:O	1.70	0.91
1:E:299:HIS:HD2	1:E:301:LEU:H	1.23	0.86
1:E:261:VAL:O	1:E:261:VAL:CG2	2.30	0.79
1:E:159:SER:C	1:E:196:GLN:OE1	2.27	0.78
2:F:157:TYR:O	2:F:158:ASP:HB2	1.89	0.71
1:E:72:ASP:O	1:E:75:ILE:HG12	1.91	0.68
1:E:168:TYR:CE1	1:E:169:ASN:O	2.49	0.65
2:F:68:ARG:NH1	2:F:81:ASN:OD1	2.29	0.65
1:E:299:HIS:CD2	1:E:301:LEU:H	2.11	0.65
1:E:309:TYR:HD2	2:F:89:LEU:HD22	1.63	0.64
1:E:262:LYS:O	1:E:262:LYS:CG	2.36	0.64
1:E:111:ARG:O	1:E:264:GLY:HA2	1.98	0.63
2:F:133:LEU:HD11	2:F:139:GLU:HB2	1.80	0.63
2:F:157:TYR:O	2:F:158:ASP:CB	2.46	0.63
1:E:168:TYR:C	1:E:168:TYR:CD1	2.78	0.62
2:F:145:ASP:O	2:F:148:CYS:N	2.35	0.60
1:E:164:ILE:O	1:E:246:GLU:HA	2.05	0.56
1:E:9:ILE:HD11	2:F:122:VAL:HG21	1.87	0.56
2:F:19:ASP:HB2	4:F:2004:HOH:O	2.06	0.56
1:E:97:GLY:HA3	1:E:230:MET:O	2.06	0.55
1:E:120:ILE:HG22	1:E:168:TYR:CE2	2.42	0.55
1:E:168:TYR:CD1	1:E:169:ASN:N	2.75	0.54
1:E:115:PHE:HE1	1:E:260:ILE:HG12	1.74	0.52
2:F:26:HIS:HD2	4:F:2005:HOH:O	1.92	0.52
1:E:168:TYR:HE1	1:E:170:ASN:HA	1.75	0.52
1:E:92:ASP:O	1:E:93:LEU:C	2.52	0.52
1:E:56:LEU:O	1:E:57:ARG:HB2	2.10	0.51
1:E:113:ASN:CG	1:E:263:LYS:CB	2.74	0.51
1:E:9:ILE:HG13	2:F:119:TYR:HA	1.92	0.50
1:E:95:TYR:CD1	1:E:230:MET:HG2	2.46	0.50
1:E:112:ILE:HA	1:E:262:LYS:HB2	1.94	0.49
1:E:113:ASN:HB2	1:E:263:LYS:HB2	1.94	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:325:PRO:HD3	4:E:2036:HOH:O	2.13	0.48
2:F:156:THR:O	2:F:156:THR:OG1	2.30	0.47
1:E:169:ASN:OD1	1:E:169:ASN:C	2.58	0.47
1:E:113:ASN:OD1	1:E:263:LYS:HB2	2.02	0.47
1:E:17:THR:HG22	4:E:2007:HOH:O	2.14	0.46
2:F:84:MET:HE2	2:F:85:GLU:HG2	1.98	0.46
1:E:200:THR:OG1	1:E:215:PRO:HG2	2.16	0.46
1:E:95:TYR:CD1	1:E:230:MET:CG	3.00	0.45
1:E:168:TYR:CE1	1:E:169:ASN:C	2.95	0.45
1:E:308:LYS:HB3	2:F:62:GLN:NE2	2.32	0.45
1:E:309:TYR:CD2	2:F:89:LEU:HD22	2.48	0.44
2:F:145:ASP:O	2:F:148:CYS:HB3	2.17	0.44
2:F:62:GLN:HG2	2:F:92:TRP:CD2	2.52	0.44
2:F:101:LEU:HD13	4:F:2019:HOH:O	2.17	0.44
2:F:145:ASP:H	2:F:148:CYS:HB3	1.81	0.44
1:E:120:ILE:HD11	1:E:254:PRO:HB2	1.99	0.43
2:F:157:TYR:CD1	2:F:157:TYR:C	2.96	0.43
1:E:263:LYS:HA	1:E:263:LYS:HD2	1.83	0.43
2:F:145:ASP:OD1	2:F:145:ASP:C	2.60	0.43
1:E:45:LEU:HD13	1:E:84:VAL:HG21	2.00	0.43
2:F:133:LEU:HD11	2:F:139:GLU:CB	2.48	0.43
1:E:284:THR:HG22	1:E:302:THR:HG22	2.01	0.43
1:E:285:PRO:HD3	1:E:301:LEU:O	2.19	0.43
1:E:134:GLY:HA3	1:E:153:TRP:HB3	2.01	0.42
2:F:19:ASP:HB3	2:F:36:ALA:HB2	2.02	0.42
1:E:82:TYR:CZ	1:E:283:GLN:HG2	2.54	0.42
1:E:93:LEU:N	1:E:93:LEU:HD23	2.35	0.41
1:E:19:GLN:O	1:E:314:ARG:NH2	2.53	0.41
1:E:150:ASN:HA	1:E:256:TYR:HD2	1.85	0.41
1:E:299:HIS:CD2	1:E:300:PRO:HD2	2.55	0.40

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	E	319/340 (94%)	308 (97%)	11 (3%)	0	100	100
2	F	156/160 (98%)	151 (97%)	5 (3%)	0	100	100
All	All	475/500 (95%)	459 (97%)	16 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	E	289/308 (94%)	272 (94%)	17 (6%)	16	33
2	F	134/136 (98%)	127 (95%)	7 (5%)	19	39
All	All	423/444 (95%)	399 (94%)	24 (6%)	17	35

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	17	THR
1	E	75	ILE
1	E	93	LEU
1	E	169	ASN
1	E	173	GLN
1	E	196	GLN
1	E	199	THR
1	E	219	THR
1	E	261	VAL
1	E	262	LYS
1	E	263	LYS
1	E	274	GLU
1	E	297	ASN
1	E	314	ARG
1	E	315	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	321	LEU
1	E	324	SER
2	F	39	GLU
2	F	60	ASN
2	F	62	GLN
2	F	77	ILE
2	F	101	LEU
2	F	127	ARG
2	F	154	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	76	ASN
1	E	91	ASN
1	E	119	GLN
1	E	172	ASN
1	E	186	ASN
1	E	297	ASN
1	E	299	HIS
2	F	25	HIS
2	F	26	HIS
2	F	42	GLN
2	F	62	GLN
2	F	146	ASN
2	F	154	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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	NAG	E	1326	1	14,14,15	0.40	0	17,19,21	0.82	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	E	1326	1	1/1/5/7	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	E	1326	NAG	C1

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	E	1326	NAG	C8-C7-N2-C2
3	E	1326	NAG	O7-C7-N2-C2
3	E	1326	NAG	C4-C5-C6-O6
3	E	1326	NAG	O5-C5-C6-O6

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

6.1 Protein, DNA and RNA chains

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^{th} 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	E	321/340 (94%)	0.68	22 (6%) 24 22	18, 43, 54, 59	0
2	F	158/160 (98%)	0.63	20 (12%) 9 8	18, 34, 57, 67	0
All	All	479/500 (95%)	0.66	42 (8%) 17 16	18, 42, 55, 67	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	325	PRO	4.7
1	E	263	LYS	4.1
2	F	60	ASN	4.1
1	E	278	CYS	4.0
1	E	261	VAL	3.8
2	F	148	CYS	3.6
1	E	173	GLN	3.6
1	E	161	TYR	3.5
2	F	158	ASP	3.3
2	F	128	ASP	3.2
1	E	143	GLY	3.0
2	F	150	GLU	2.9
1	E	49	ASP	2.9
1	E	306	CYS	2.8
1	E	264	GLY	2.7
1	E	59	CYS	2.7
1	E	47	ASP	2.7
2	F	141	TYR	2.5
2	F	57	ASP	2.5
2	F	156	THR	2.4
2	F	129	ASN	2.4
2	F	86	ASP	2.4
1	E	116	GLU	2.4
1	E	170	ASN	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	E	130	GLU	2.4
1	E	246	GLU	2.4
1	E	57	ARG	2.3
1	E	158	ASN	2.2
2	F	145	ASP	2.2
2	F	105	GLU	2.2
2	F	144	CYS	2.1
1	E	186	ASN	2.1
1	E	262	LYS	2.1
2	F	66	VAL	2.1
2	F	149	MET	2.1
2	F	39	GLU	2.1
1	E	124	SER	2.1
1	E	92	ASP	2.1
2	F	31	GLY	2.0
2	F	147	GLU	2.0
2	F	120	ASP	2.0
2	F	154	ASN	2.0

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 oligosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
3	NAG	E	1326	14/15	0.73	0.16	92,92,93,94	0

6.5 Other polymers (i)

There are no such residues in this entry.