Full wwPDB X-ray Structure Validation Report (i)

Mar 5, 2024 – 12:37 PM EST

PDB ID	:	2ZS6
Title	:	HA3 subcomponent of botulinum type C progenitor toxin
Authors	:	Nakamura, T.; Tonozuka, T.; Kotani, M.; Oguma, K.; Nishikawa, A.
Deposited on	:	2008-09-02
Resolution	:	2.60 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 (i)) were used in the production of this report:

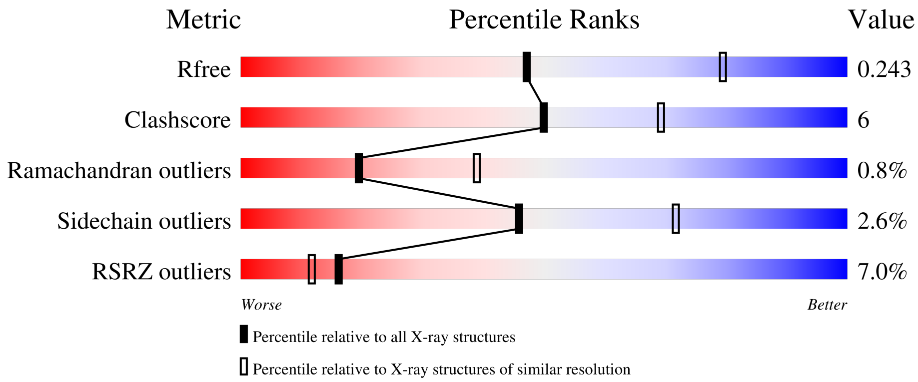
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

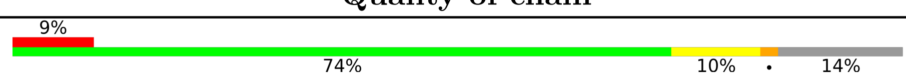
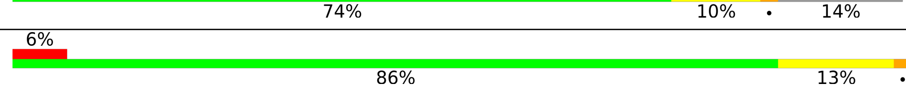
The reported resolution of this entry is 2.60 Å.

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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	A	205	 9% 74% 10% • 14%
2	B	420	 6% 86% 13% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5013 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 components HA3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	177	1447	926	235	283	3	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	ILE	-	expression tag	UNP P46085
A	-19	SER	-	expression tag	UNP P46085
A	-18	GLU	-	expression tag	UNP P46085
A	-17	PHE	-	expression tag	UNP P46085
A	-16	ASP	-	expression tag	UNP P46085
A	-15	TYR	-	expression tag	UNP P46085
A	-14	LYS	-	expression tag	UNP P46085
A	-13	ASP	-	expression tag	UNP P46085
A	-12	HIS	-	expression tag	UNP P46085
A	-11	GLU	-	expression tag	UNP P46085
A	-10	ILE	-	expression tag	UNP P46085
A	-9	ASP	-	expression tag	UNP P46085
A	-8	TYR	-	expression tag	UNP P46085
A	-7	LYS	-	expression tag	UNP P46085
A	-6	ASP	-	expression tag	UNP P46085
A	-5	ASP	-	expression tag	UNP P46085
A	-4	ASP	-	expression tag	UNP P46085
A	-3	ASP	-	expression tag	UNP P46085
A	-2	LYS	-	expression tag	UNP P46085
A	-1	TRP	-	expression tag	UNP P46085
A	0	ILE	-	expression tag	UNP P46085

- Molecule 2 is a protein called Hemagglutinin components HA3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	420	Total 3338	C 2109	N 553	O 671	S 5	0	0	0

- Molecule 3 is water.

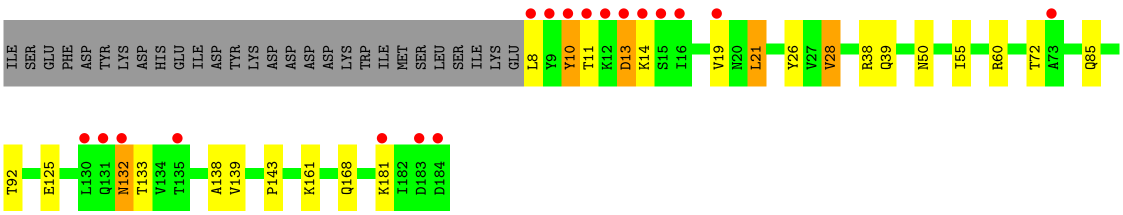
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	69	Total 69	O 69	0	0
3	B	159	Total 159	O 159	0	0

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.

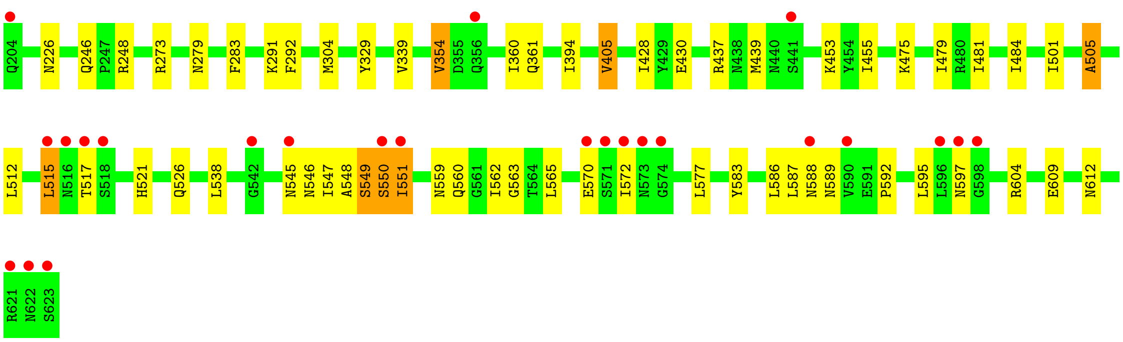
- Molecule 1: Hemagglutinin components HA3

Chain A:



- Molecule 2: Hemagglutinin components HA3

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	176.40Å 176.40Å 80.91Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.00 – 2.60 47.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.00-2.60) 100.0 (47.00-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	3.90 (at 2.61Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.222 , 0.243 0.221 , 0.243	Depositor DCC
R_{free} test set	4449 reflections (10.06%)	wwPDB-VP
Wilson B-factor (Å^2)	52.5	Xtriage
Anisotropy	0.211	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 41.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.029 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5013	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.07% 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 ⓘ

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	0.38	0/1478	0.60	2/2004 (0.1%)
2	B	0.34	0/3404	0.54	0/4642
All	All	0.35	0/4882	0.56	2/6646 (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	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	13	ASP	O-C-N	-5.89	113.27	122.70
1	A	14	LYS	N-CA-C	5.44	125.70	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	13	ASP	Mainchain

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	1447	0	1402	18	0
2	B	3338	0	3275	44	0
3	A	69	0	0	0	0
3	B	159	0	0	2	0
All	All	5013	0	4677	55	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 6.

All (55) 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:430:GLU:HG3	2:B:455:ILE:HD11	1.75	0.69
1:A:26:TYR:CE2	1:A:181:LYS:HG2	2.29	0.67
2:B:545:ASN:O	2:B:546:ASN:ND2	2.20	0.67
2:B:246:GLN:HE21	2:B:248:ARG:H	1.42	0.66
2:B:551:ILE:HD13	2:B:592:PRO:HB2	1.76	0.66
2:B:563:GLY:HA2	2:B:604:ARG:O	1.97	0.65
2:B:559:ASN:HD21	2:B:612:ASN:H	1.52	0.57
1:A:19:VAL:HG12	1:A:21:LEU:HG	1.85	0.56
1:A:8:LEU:HD12	2:B:453:LYS:HG3	1.87	0.56
2:B:512:LEU:HD12	2:B:538:LEU:HD21	1.87	0.56
1:A:21:LEU:HD22	1:A:26:TYR:CE2	2.41	0.56
1:A:50:ASN:HB2	2:B:339:VAL:HG23	1.89	0.54
2:B:562:ILE:O	2:B:583:TYR:O	2.24	0.54
2:B:226:ASN:CG	2:B:354:VAL:HG22	2.28	0.53
1:A:39:GLN:HE21	2:B:279:ASN:HD22	1.56	0.52
2:B:549:SER:O	2:B:550:SER:C	2.46	0.52
2:B:515:LEU:HD21	2:B:538:LEU:HD22	1.91	0.51
1:A:39:GLN:NE2	2:B:279:ASN:HD22	2.08	0.51
2:B:246:GLN:NE2	2:B:248:ARG:H	2.06	0.51
1:A:8:LEU:HD23	1:A:161:LYS:HZ1	1.75	0.51
1:A:8:LEU:HD23	1:A:161:LYS:NZ	2.25	0.50
2:B:549:SER:HB2	2:B:597:ASN:ND2	2.26	0.50
2:B:517:THR:HG23	2:B:521:HIS:NE2	2.27	0.49
1:A:28:VAL:HG13	1:A:125:GLU:HB2	1.94	0.49
2:B:512:LEU:HD12	2:B:538:LEU:CD2	2.42	0.49
2:B:394:ILE:HD13	2:B:481:ILE:HD13	1.95	0.49
1:A:11:THR:HG22	1:A:168:GLN:OE1	2.13	0.48
2:B:545:ASN:OD1	2:B:547:ILE:HG22	2.13	0.48
2:B:512:LEU:HB2	2:B:538:LEU:HD23	1.94	0.48
2:B:551:ILE:HD13	2:B:592:PRO:CB	2.40	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:72:THR:HG22	1:A:138:ALA:HB2	1.95	0.48
1:A:132:ASN:HD22	1:A:133:THR:N	2.11	0.48
2:B:501:ILE:HD12	2:B:577:LEU:HD11	1.95	0.48
2:B:526:GLN:NE2	3:B:690:HOH:O	2.45	0.47
1:A:139:VAL:HG21	2:B:437:ARG:HA	1.97	0.47
2:B:549:SER:CB	2:B:597:ASN:HD21	2.26	0.47
2:B:405:VAL:HG13	2:B:479:ILE:HB	1.96	0.46
2:B:428:ILE:HG23	2:B:455:ILE:HB	1.97	0.45
1:A:55:ILE:HB	2:B:291:LYS:HB3	2.00	0.44
2:B:360:ILE:HD12	2:B:361:GLN:H	1.82	0.44
2:B:226:ASN:ND2	2:B:354:VAL:HG22	2.33	0.43
2:B:283:PHE:HB3	2:B:329:TYR:OH	2.18	0.43
2:B:501:ILE:CD1	2:B:577:LEU:HD11	2.48	0.43
2:B:505:ALA:HB3	3:B:743:HOH:O	2.17	0.43
1:A:60:ARG:NH1	1:A:143:PRO:HB3	2.34	0.43
2:B:394:ILE:CD1	2:B:481:ILE:HD13	2.49	0.43
2:B:549:SER:CB	2:B:597:ASN:ND2	2.82	0.42
2:B:587:LEU:O	2:B:589:ASN:N	2.51	0.42
2:B:292:PHE:HB2	2:B:304:MET:HE3	2.01	0.42
1:A:10:TYR:CD1	2:B:475:LYS:HE2	2.55	0.41
1:A:85:GLN:NE2	1:A:92:THR:H	2.18	0.41
2:B:549:SER:CA	2:B:597:ASN:HD21	2.33	0.41
2:B:547:ILE:HD12	2:B:548:ALA:H	1.85	0.41
2:B:560:GLN:HB3	2:B:609:GLU:HG2	2.03	0.40
2:B:565:LEU:HD21	2:B:595:LEU:HD21	2.03	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	A	175/205 (85%)	164 (94%)	11 (6%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	418/420 (100%)	396 (95%)	17 (4%)	5 (1%)	13	27
All	All	593/625 (95%)	560 (94%)	28 (5%)	5 (1%)	19	39

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	551	ILE
2	B	505	ALA
2	B	550	SER
2	B	588	ASN
2	B	572	ILE

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	A	160/188 (85%)	155 (97%)	5 (3%)	40	66
2	B	385/385 (100%)	376 (98%)	9 (2%)	50	75
All	All	545/573 (95%)	531 (97%)	14 (3%)	46	72

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	10	TYR
1	A	21	LEU
1	A	28	VAL
1	A	38	ARG
1	A	132	ASN
2	B	273	ARG
2	B	354	VAL
2	B	405	VAL
2	B	439	MET
2	B	484	ILE
2	B	515	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	549	SER
2	B	570	GLU
2	B	586	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (21) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	25	ASN
1	A	39	GLN
1	A	41	GLN
1	A	42	ASN
1	A	83	ASN
1	A	85	GLN
1	A	132	ASN
2	B	226	ASN
2	B	246	GLN
2	B	361	GLN
2	B	367	ASN
2	B	393	ASN
2	B	401	ASN
2	B	422	ASN
2	B	436	ASN
2	B	526	GLN
2	B	558	ASN
2	B	559	ASN
2	B	582	ASN
2	B	588	ASN
2	B	597	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 monosaccharides 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^{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	A	177/205 (86%)	0.41	18 (10%)	6	4	29, 52, 105, 129	0
2	B	420/420 (100%)	0.13	24 (5%)	23	18	35, 55, 88, 110	0
All	All	597/625 (95%)	0.21	42 (7%)	16	12	29, 54, 93, 129	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	13	ASP	10.5
1	A	8	LEU	6.7
1	A	11	THR	6.4
1	A	184	ASP	6.1
1	A	15	SER	5.5
2	B	572	ILE	5.2
2	B	573	ASN	4.9
1	A	12	LYS	4.8
2	B	570	GLU	4.8
2	B	623	SER	4.8
2	B	550	SER	4.4
1	A	9	TYR	4.4
1	A	131	GLN	4.2
1	A	181	LYS	4.0
2	B	571	SER	4.0
1	A	10	TYR	3.9
2	B	516	ASN	3.3
2	B	588	ASN	3.3
2	B	542	GLY	3.1
2	B	590	VAL	3.1
1	A	183	ASP	3.0
2	B	622	ASN	2.9
1	A	73	ALA	2.9
1	A	132	ASN	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	441	SER	2.8
2	B	574	GLY	2.7
2	B	356	GLN	2.7
1	A	14	LYS	2.7
2	B	515	LEU	2.6
2	B	517	THR	2.5
2	B	204	GLN	2.4
1	A	19	VAL	2.4
2	B	518	SER	2.3
2	B	596	LEU	2.3
2	B	545	ASN	2.2
2	B	621	ARG	2.2
1	A	135	THR	2.2
2	B	597	ASN	2.2
2	B	598	GLY	2.1
1	A	130	LEU	2.1
2	B	551	ILE	2.0
1	A	16	ILE	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 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.