Full wwPDB NMR Structure Validation Report (i)

Mar 7, 2022 – 08:11 AM EST

PDB ID	:	2YW5
Title	:	Solution structure of the bromodomain from human bromodomain containing protein 3
Authors	:	Furue, M.; Suzuki, S.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-04-19

This is a Full wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

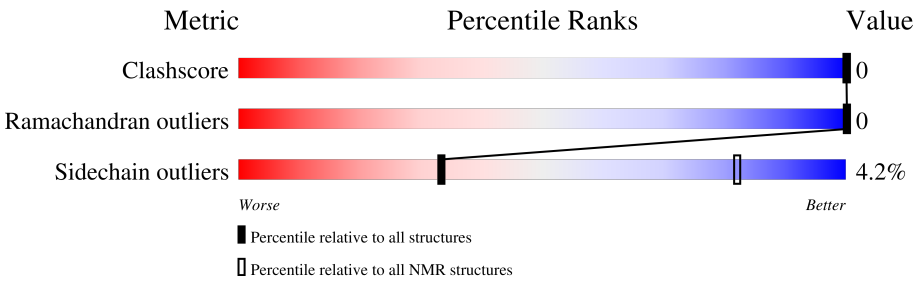
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	138	83% • 17%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 14 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:33-A:147 (115)	0.26	14

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 17, 18
2	12, 13, 16, 19, 20

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2229 atoms, of which 1112 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Bromodomain-containing protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	138	2229	717	1112	185	206	9	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	18	GLY	-	expression tag	UNP Q15059
A	19	SER	-	expression tag	UNP Q15059
A	20	SER	-	expression tag	UNP Q15059
A	21	GLY	-	expression tag	UNP Q15059
A	22	SER	-	expression tag	UNP Q15059
A	23	SER	-	expression tag	UNP Q15059
A	24	GLY	-	expression tag	UNP Q15059

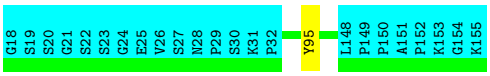
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Bromodomain-containing protein 3

Chain A:



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

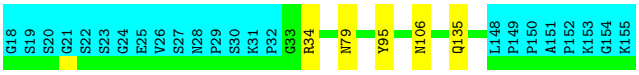
• Molecule 1: Bromodomain-containing protein 3

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Bromodomain-containing protein 3

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Bromodomain-containing protein 3

Chain A: 80% · 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 K78 R89 Y95 N116 Q135 L148 P149 P150 A151 P152 K153 G154 K155

4.2.4 Score per residue for model 4

- Molecule 1: Bromodomain-containing protein 3

Chain A: 77% 7% 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 G33 R34 Y60 K78 N79 K87 K88 Y95 Q103 N106 L148 P149 P150 A151 P152 K153 G154 K155

4.2.5 Score per residue for model 5

- Molecule 1: Bromodomain-containing protein 3

Chain A: 81% · 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 K75 Y95 M125 L148 P149 P150 A151 P152 K153 G154 K155

4.2.6 Score per residue for model 6

- Molecule 1: Bromodomain-containing protein 3

Chain A: 78% 6% 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 G33 R34 Q54 D64 M81 K88 R89 Y95 Q103 L148 P149 P150 A151 P152 K153 G154 K155

4.2.7 Score per residue for model 7

- Molecule 1: Bromodomain-containing protein 3

Chain A: 77% 7% 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 I77 K78 K87 N92 Y95 N106 N111 Y115 Q135 L148 P149 P150 A151 P152 K153 G154 K155

4.2.8 Score per residue for model 8

- Molecule 1: Bromodomain-containing protein 3

Chain A: 76% 7% 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Q54 Y60 K78 N79 P80 M81 R89 L90 E91 Y95 N111 E144 L148 P149 P150 A151 P152 K153 G154 K155

4.2.9 Score per residue for model 9

- Molecule 1: Bromodomain-containing protein 3

Chain A: 79% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Y60 K75 N79 K87 Y95 M125 L148 P149 P150 A151 P152 K153 G154 K155

4.2.10 Score per residue for model 10

- Molecule 1: Bromodomain-containing protein 3

Chain A: 79% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Y60 K75 R89 Y95 N106 N111 L148 P149 P150 A151 P152 K153 G154 K155

4.2.11 Score per residue for model 11

- Molecule 1: Bromodomain-containing protein 3

Chain A: 80% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 K48 N79 R89 Y95 L148 P149 P150 A151 P152 K153 G154 K155

4.2.12 Score per residue for model 12

- Molecule 1: Bromodomain-containing protein 3

Chain A: 83% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Y95 L148 P149 P150 A151 P152 K153 G154 K155

4.2.13 Score per residue for model 13

- Molecule 1: Bromodomain-containing protein 3

Chain A: 80% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 N37 N79 R89 Y95 K117 L148 P149 P150 A151 P152 K153 G154 K155

4.2.14 Score per residue for model 14 (medoid)

- Molecule 1: Bromodomain-containing protein 3

Chain A: 79% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 G33 R34 Y60 K88 R89 Y95 K117 L148 P149 P150 A151 P152 K153 G154 K155

4.2.15 Score per residue for model 15

- Molecule 1: Bromodomain-containing protein 3

Chain A: 80% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Y60 K78 K87 Y95 L148 P149 P150 A151 P152 K153 G154 K155

4.2.16 Score per residue for model 16

- Molecule 1: Bromodomain-containing protein 3

Chain A: 80% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 Y60 K87 Y95 C101 L148 P149 P150 A151 P152 K153 G154 K155

4.2.17 Score per residue for model 17

- Molecule 1: Bromodomain-containing protein 3

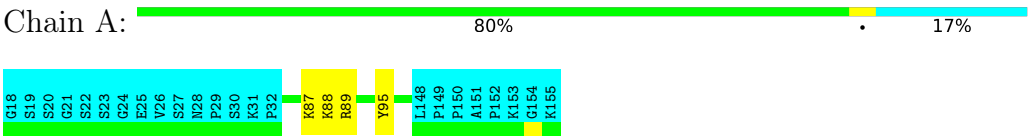
Chain A: 80% • 17%

G18 S19 S20 G21 S22 S23 G24 E25 V26 S27 N28 P29 S30 K31 P32 N79 K87 Y95 K117 L148 P149 P150 A151 P152 K153 G154 K155

4.2.18 Score per residue for model 18

- Molecule 1: Bromodomain-containing protein 3

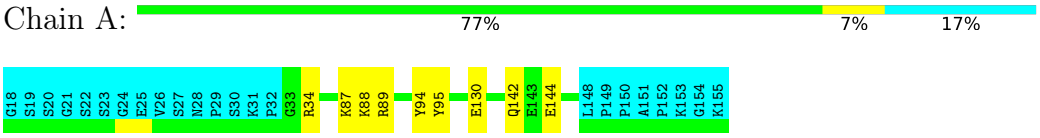
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Bromodomain-containing protein 3

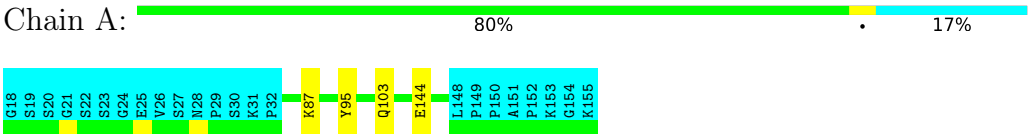
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Bromodomain-containing protein 3

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dyanamics, simulated annealing, energy minimization.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations,structures with the lowest energy,target function.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure solution	2.0.17
Amber	refinement	9

No chemical shift data was provided.

6 Model quality (i)

6.1 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 (average) 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	0.71±0.00	0±0/991 (0.0± 0.0%)	0.92±0.01	1±1/1343 (0.1± 0.1%)
All	All	0.71	0/19820 (0.0%)	0.92	23/26860 (0.1%)

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	Planarity
1	A	0.0±0.0	0.1±0.2
All	All	0	1

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	34	ARG	NE-CZ-NH1	6.54	123.57	120.30	14	6
1	A	89	ARG	NE-CZ-NH1	6.52	123.56	120.30	8	9
1	A	60	TYR	CB-CG-CD2	-5.63	117.62	121.00	8	8

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	94	TYR	Sidechain	1

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	965	957	957	0±0
All	All	19300	19140	19140	1

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:77:ILE:HD11	1:A:115:TYR:HB2	0.40	1.94	7	1

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	115/138 (83%)	111±1 (97±1%)	4±1 (3±1%)	0±0 (0±0%)	100	100
All	All	2300/2760 (83%)	2223 (97%)	77 (3%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	107/125 (86%)	103±2 (96±2%)	4±2 (4±2%)	33 82
All	All	2140/2500 (86%)	2051 (96%)	89 (4%)	33 82

All 26 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	95	TYR	20
1	A	87	LYS	9
1	A	79	ASN	7
1	A	78	LYS	6
1	A	88	LYS	6
1	A	75	LYS	4
1	A	111	ASN	4
1	A	106	ASN	4
1	A	135	GLN	3
1	A	103	GLN	3
1	A	144	GLU	3
1	A	117	LYS	3
1	A	125	MET	2
1	A	54	GLN	2
1	A	81	MET	2
1	A	35	LYS	1
1	A	116	ASN	1
1	A	64	ASP	1
1	A	92	ASN	1
1	A	91	GLU	1
1	A	48	LYS	1
1	A	89	ARG	1
1	A	37	ASN	1
1	A	101	CYS	1
1	A	130	GLU	1
1	A	142	GLN	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided