



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 06:05 am BST

PDB ID : 4U30  
Title : Human mesotrypsin complexed with bikunin Kunitz domain 2  
Authors : Wang, R.; Soares, A.S.; Radisky, E.S.  
Deposited on : 2014-07-18  
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](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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

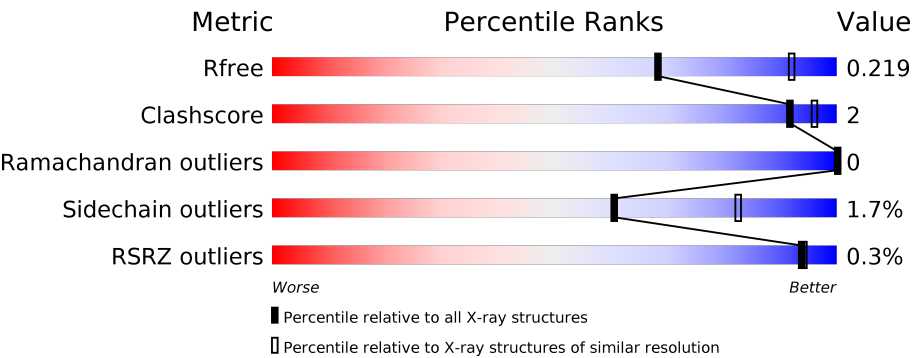
# 1 Overall quality at a glance i

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))
$R_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (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 on 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	224	<div><div></div><div>92%8%</div></div>
1	B	224	<div><div></div><div>92%8%</div></div>
1	C	224	<div><div></div><div>92%8%</div></div>
1	D	224	<div><div></div><div>91%9%</div></div>
2	W	59	<div><div></div><div>86%5%8%</div></div>
2	X	59	<div><div>2%</div><div>90%8%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	Y	59	<div><div></div><div>2%</div><div>92%</div><div>8%</div></div>
2	Z	59	<div><div></div><div>85%</div><div>7%</div><div>8%</div></div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8766 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 Trypsin-3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	224	Total	C	N	O	S	0	0	0
			1701	1070	294	325	12			
1	B	224	Total	C	N	O	S	0	0	0
			1701	1070	294	325	12			
1	C	224	Total	C	N	O	S	0	0	0
			1701	1070	294	325	12			
1	D	224	Total	C	N	O	S	0	0	0
			1701	1070	294	325	12			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	127	ALA	THR	variant	UNP P35030
A	195	ALA	SER	engineered mutation	UNP P35030
B	127	ALA	THR	variant	UNP P35030
B	195	ALA	SER	engineered mutation	UNP P35030
C	127	ALA	THR	variant	UNP P35030
C	195	ALA	SER	engineered mutation	UNP P35030
D	127	ALA	THR	variant	UNP P35030
D	195	ALA	SER	engineered mutation	UNP P35030

- Molecule 2 is a protein called Trypstatin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	54	Total	C	N	O	S	0	0	0
			417	267	73	71	6			
2	Y	54	Total	C	N	O	S	0	0	0
			417	267	73	71	6			
2	Z	54	Total	C	N	O	S	0	0	0
			417	267	73	71	6			
2	W	54	Total	C	N	O	S	0	0	0
			416	267	72	71	6			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	5A	ALA	-	INSERTION	UNP P02760
Y	5A	ALA	-	INSERTION	UNP P02760
Z	5A	ALA	-	INSERTION	UNP P02760
W	5A	ALA	-	INSERTION	UNP P02760

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Ca 1 1	0	0
3	A	1	Total Ca 1 1	0	0
3	D	1	Total Ca 1 1	0	0
3	C	1	Total Ca 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	65	Total O 65 65	0	0
4	B	60	Total O 60 60	0	0
4	C	60	Total O 60 60	0	0
4	D	59	Total O 59 59	0	0
4	X	12	Total O 12 12	0	0
4	Y	13	Total O 13 13	0	0
4	Z	10	Total O 10 10	0	0
4	W	12	Total O 12 12	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: Trypsin-3

Chain A: 



- Molecule 1: Trypsin-3

Chain B: 



- Molecule 1: Trypsin-3

Chain C: 




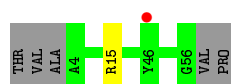
- Molecule 1: Trypsin-3

Chain D: 

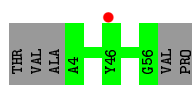
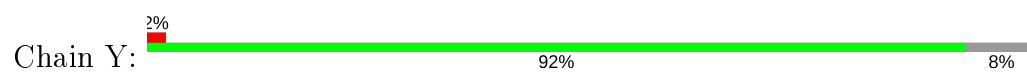


- Molecule 2: Trypstatin

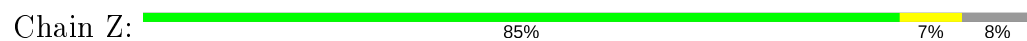
Chain X: 



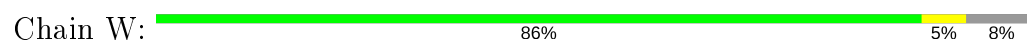
- Molecule 2: Trypstatin



- Molecule 2: Trypstatin



- Molecule 2: Trypstatin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	164.00 Å   164.00 Å   81.02 Å 90.00°   90.00°   120.00°	Depositor
Resolution (Å)	44.75 – 2.50 47.34 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (44.75-2.50) 100.0 (47.34-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.17 (at 2.51 Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.188 , 0.223 0.195 , 0.219	Depositor DCC
$R_{free}$ test set	4207 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.6	Xtriage
Anisotropy	0.750	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 6.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.480 for -h,-k,l 0.487 for h,-h-k,-l 0.487 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8766	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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	A	1.04	2/1740 (0.1%)	0.99	2/2365 (0.1%)
1	B	1.05	3/1740 (0.2%)	0.98	4/2365 (0.2%)
1	C	1.00	2/1740 (0.1%)	0.97	2/2365 (0.1%)
1	D	1.03	2/1740 (0.1%)	0.97	1/2365 (0.0%)
2	W	0.95	0/427	0.98	1/573 (0.2%)
2	X	0.92	0/428	0.91	0/575
2	Y	0.92	0/428	0.92	0/575
2	Z	0.96	0/428	0.98	1/575 (0.2%)
All	All	1.01	9/8671 (0.1%)	0.97	11/11758 (0.1%)

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	77	GLU	CD-OE1	10.76	1.37	1.25
1	D	77	GLU	CD-OE1	9.96	1.36	1.25
1	A	77	GLU	CD-OE1	9.76	1.36	1.25
1	C	77	GLU	CD-OE1	9.71	1.36	1.25
1	B	186	GLU	CD-OE1	6.46	1.32	1.25
1	A	186	GLU	CD-OE1	5.25	1.31	1.25
1	D	186	GLU	CD-OE1	5.25	1.31	1.25
1	B	221	TRP	CE3-CZ3	5.23	1.47	1.38
1	C	221	TRP	CE3-CZ3	5.02	1.47	1.38

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	96	ARG	NE-CZ-NH1	6.19	123.39	120.30
1	B	224	ARG	NE-CZ-NH1	5.92	123.26	120.30
1	B	224	ARG	NE-CZ-NH2	-5.75	117.42	120.30
1	A	153	ASP	CB-CG-OD1	5.64	123.38	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	236	ASP	CB-CG-OD1	5.46	123.21	118.30
1	A	157	CYS	CB-CA-C	-5.44	99.51	110.40
1	C	157	CYS	CB-CA-C	-5.32	99.76	110.40
1	D	90	ARG	NE-CZ-NH2	-5.27	117.67	120.30
1	C	62	ARG	NE-CZ-NH1	5.11	122.85	120.30
2	W	24	ASP	CB-CG-OD1	5.10	122.89	118.30
2	Z	24	ASP	CB-CG-OD1	5.09	122.88	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	A	1701	0	1652	7	0
1	B	1701	0	1652	5	0
1	C	1701	0	1652	5	0
1	D	1701	0	1652	7	0
2	W	416	0	393	1	0
2	X	417	0	397	1	0
2	Y	417	0	397	0	0
2	Z	417	0	397	2	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	65	0	0	0	0
4	B	60	0	0	0	0
4	C	60	0	0	0	0
4	D	59	0	0	0	0
4	W	12	0	0	0	0
4	X	12	0	0	0	0
4	Y	13	0	0	0	0
4	Z	10	0	0	0	0
All	All	8766	0	8192	26	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 2.

All (26) 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:C:45:SER:OG	1:C:198:PRO:HB3	2.03	0.58
1:B:45:SER:OG	1:B:198:PRO:HB3	2.04	0.57
1:D:142:GLY:HA2	1:D:193:ARG:HB2	1.90	0.54
1:A:142:GLY:HA2	1:A:193:ARG:HB2	1.90	0.53
1:C:142:GLY:HA2	1:C:193:ARG:HB2	1.92	0.50
1:D:45:SER:OG	1:D:198:PRO:HB3	2.12	0.50
1:D:17:VAL:O	1:D:188:LYS:HA	2.14	0.48
1:A:45:SER:OG	1:A:198:PRO:HB3	2.15	0.47
2:W:21:TRP:O	2:W:44:LYS:HA	2.15	0.47
1:B:17:VAL:O	1:B:188:LYS:HA	2.14	0.47
2:Z:21:TRP:O	2:Z:44:LYS:HA	2.15	0.46
1:C:217:HIS:HB2	1:C:224:ARG:HD2	1.97	0.46
1:C:72:ASN:HA	1:C:153:ASP:O	2.16	0.45
1:B:192:GLN:O	1:B:193:ARG:HB2	2.17	0.45
1:C:17:VAL:O	1:C:188:LYS:HA	2.17	0.45
1:D:217:HIS:HB2	1:D:224:ARG:HD2	1.98	0.44
1:A:17:VAL:O	1:A:188:LYS:HA	2.16	0.44
1:A:72:ASN:HA	1:A:153:ASP:O	2.17	0.44
1:A:117:ARG:HH11	1:A:117:ARG:HG2	1.82	0.44
1:D:67:LEU:N	1:D:67:LEU:HD12	2.33	0.44
1:D:72:ASN:HA	1:D:153:ASP:O	2.18	0.42
1:B:195:ALA:HB2	2:Z:15:ARG:C	2.39	0.42
1:B:72:ASN:HA	1:B:153:ASP:O	2.21	0.41
1:A:195:ALA:HB2	2:X:15:ARG:C	2.42	0.41
1:D:56:ALA:HA	1:D:104:MET:HB2	2.03	0.40
1:A:26:SER:C	1:A:27:LEU:HD12	2.42	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	222/224 (99%)	217 (98%)	5 (2%)	0	100	100
1	B	222/224 (99%)	212 (96%)	10 (4%)	0	100	100
1	C	222/224 (99%)	214 (96%)	8 (4%)	0	100	100
1	D	222/224 (99%)	215 (97%)	7 (3%)	0	100	100
2	W	52/59 (88%)	49 (94%)	3 (6%)	0	100	100
2	X	52/59 (88%)	50 (96%)	2 (4%)	0	100	100
2	Y	52/59 (88%)	50 (96%)	2 (4%)	0	100	100
2	Z	52/59 (88%)	49 (94%)	3 (6%)	0	100	100
All	All	1096/1132 (97%)	1056 (96%)	40 (4%)	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	A	185/185 (100%)	181 (98%)	4 (2%)	52	77
1	B	185/185 (100%)	183 (99%)	2 (1%)	73	89
1	C	185/185 (100%)	181 (98%)	4 (2%)	52	77
1	D	185/185 (100%)	180 (97%)	5 (3%)	44	71
2	W	42/46 (91%)	42 (100%)	0	100	100
2	X	42/46 (91%)	42 (100%)	0	100	100
2	Y	42/46 (91%)	42 (100%)	0	100	100
2	Z	42/46 (91%)	42 (100%)	0	100	100
All	All	908/924 (98%)	893 (98%)	15 (2%)	60	82

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	76	LEU
1	A	146	SER
1	A	178	ASN
1	A	185	LEU
1	B	146	SER
1	B	178	ASN
1	C	76	LEU
1	C	93	LYS
1	C	135	GLU
1	C	146	SER
1	D	96	ARG
1	D	135	GLU
1	D	146	SER
1	D	178	ASN
1	D	185	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	84	ASN
1	C	84	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 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	A	224/224 (100%)	-0.30	0 100 100	15, 26, 45, 68	0
1	B	224/224 (100%)	-0.28	1 (0%) 92 93	16, 26, 46, 71	0
1	C	224/224 (100%)	-0.29	0 100 100	16, 26, 45, 69	0
1	D	224/224 (100%)	-0.30	0 100 100	15, 26, 45, 65	0
2	W	54/59 (91%)	-0.11	0 100 100	17, 30, 55, 67	0
2	X	54/59 (91%)	-0.14	1 (1%) 66 69	17, 29, 55, 68	0
2	Y	54/59 (91%)	-0.08	1 (1%) 66 69	17, 29, 55, 68	0
2	Z	54/59 (91%)	-0.16	0 100 100	17, 30, 55, 67	0
All	All	1112/1132 (98%)	-0.26	3 (0%) 94 94	15, 27, 50, 71	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	246	SER	2.8
2	X	46	TYR	2.3
2	Y	46	TYR	2.1

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CA	B	301	1/1	0.89	0.10	52,52,52,52	0
3	CA	D	301	1/1	0.95	0.07	54,54,54,54	0
3	CA	C	301	1/1	0.96	0.12	54,54,54,54	0
3	CA	A	301	1/1	0.99	0.08	48,48,48,48	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.