

# Full wwPDB X-ray Structure Validation Report (i)

#### Oct 29, 2024 – 01:55 pm GMT

PDB ID	:	9FIS
Title	:	Structure-guided discovery of selective USP7 inhibitors with in vivo activity
Authors	:	Baker, L.M.; Murray, J.; Hubbard, R.E.; Whitehead, N.
Deposited on	:	2024-05-29
Resolution	:	2.77  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	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
buster-report	:	1.1.7 (2018)
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 (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 2.77 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R <sub>free</sub>	164625	4924 (2.80-2.76)
Clashscore	180529	5458 (2.80-2.76)
Ramachandran outliers	177936	5386 (2.80-2.76)
Sidechain outliers	177891	5388 (2.80-2.76)
RSRZ outliers	164620	4926 (2.80-2.76)

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	А	360	77%	14%	•	7%
1	В	360	74%	16%	·	8%



## 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10952 atoms, of which 5388 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
1	1 A	335	Total	С	Η	Ν	0	$\mathbf{S}$	75	0	0
1			5384	1719	2668	459	522	16	15	0	0
1	1 D	222	Total	С	Η	Ν	0	S	77	0	0
I B	ამა	5343	1708	2644	457	518	16	11	0	0	

• Molecule 1 is a protein called Ubiquitin carboxyl-terminal hydrolase 7.

Chain	Residue	Modelled	Actual	Actual Comment	
А	207	MET	-	initiating methionine	UNP Q93009
А	561	HIS	-	expression tag	UNP Q93009
А	562	HIS	-	expression tag	UNP Q93009
А	563	HIS	-	expression tag	UNP Q93009
А	564	HIS	-	expression tag	UNP Q93009
А	565	HIS	-	expression tag	UNP Q93009
А	566	HIS	-	expression tag	UNP Q93009
В	207	MET	-	initiating methionine	UNP Q93009
В	561	HIS	-	expression tag	UNP Q93009
В	562	HIS	-	expression tag	UNP Q93009
В	563	HIS	-	expression tag	UNP Q93009
В	564	HIS	-	expression tag	UNP Q93009
В	565	HIS	-	expression tag	UNP Q93009
В	566	HIS	-	expression tag	UNP Q93009

There are 14 discrepancies between the modelled and reference sequences:

• Molecule 2 is 3-[[4-oxidanyl-1-[(3 {R},4 {R})-3-phenyl-1-(2-phenylethyl)piperidin-4-y l]carbonyl-piperidin-4-yl]methyl]quinazolin-4-one (three-letter code: A1ICW) (formula:  $C_{34}H_{38}N_4O_3$ ) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf		
9	2 1	1	Total	С	Η	Ν	Ο	1	0	
	Л	1	79	34	38	4	3	T		
0	0 D	1	Total	С	Η	Ν	Ο	1	0	
2	D	D	79	34	38	4	3	L	0	

• Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	28	TotalO2828	0	0
3	В	39	Total         O           39         39	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: Ubiquitin carboxyl-terminal hydrolase 7



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	74.84Å 70.58Å 78.59Å	Deperitor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $92.96^{\circ}$ $90.00^{\circ}$	Depositor
$\mathbf{P}_{\text{assolution}}(\hat{\mathbf{A}})$	52.78 - 2.77	Depositor
Resolution (A)	52.78 - 2.77	EDS
% Data completeness	97.6 (52.78-2.77)	Depositor
(in resolution range)	97.6 (52.78-2.77)	EDS
R <sub>merge</sub>	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.92 (at 2.77 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
B B.	0.185 , $0.243$	Depositor
$\mathbf{n}, \mathbf{n}_{free}$	0.186 , $0.244$	DCC
$R_{free}$ test set	1052 reflections $(5.11%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	70.7	Xtriage
Anisotropy	0.190	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.35 , $44.0$	EDS
L-test for $twinning^2$	$< L >=0.49, < L^2>=0.32$	Xtriage
	0.000 for l,k,-h	
Estimated twinning fraction	0.024 for h,-k,-l	Xtriage
	0.018 for l,-k,h	
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10952	wwPDB-VP
Average B, all atoms $(Å^2)$	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 15.48% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 5 Model quality (i)

### 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: A1ICW

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		
	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.41	0/2772	0.83	1/3739~(0.0%)	
1	В	0.41	0/2755	0.86	6/3717~(0.2%)	
All	All	0.41	0/5527	0.85	7/7456~(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 outliers	#Planarity outliers
1	А	0	1
1	В	0	2
All	All	0	3

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	В	343	ARG	NE-CZ-NH1	-8.80	115.90	120.30
1	В	454	LEU	CB-CG-CD2	7.95	124.52	111.00
1	В	225	MET	CG-SD-CE	-6.49	89.81	100.20
1	В	343	ARG	NE-CZ-NH2	6.22	123.41	120.30
1	В	454	LEU	CB-CG-CD1	-5.82	101.11	111.00
1	А	454	LEU	CB-CG-CD2	5.56	120.46	111.00
1	В	455	VAL	N-CA-CB	-5.01	100.47	111.50

There are no chirality outliers.

All (3) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	А	344	ARG	Sidechain
1	В	343	ARG	Sidechain
1	В	549	ARG	Sidechain

#### 5.2 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 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	А	2716	2668	2656	16	0
1	В	2699	2644	2630	19	0
2	А	41	38	0	0	0
2	В	41	38	0	0	0
3	А	28	0	0	0	0
3	В	39	0	0	1	0
All	All	5564	5388	5286	32	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 3.

All (32) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom 1	Atom 2	Interatomic	Clash
Atom-1	Atom-2	distance (Å)	overlap (Å)
1:B:455:VAL:HG13	1:B:511:THR:HB	1.58	0.84
1:A:227:SER:HB3	1:A:454:LEU:HD13	1.80	0.64
1:A:320:ILE:HB	1:A:321:PRO:HD3	1.79	0.63
1:A:553:GLN:O	1:A:554:LYS:HB3	2.02	0.59
1:A:256:VAL:HB	1:A:257:PRO:HD3	1.86	0.57
1:B:289:ASP:HB2	1:B:293:GLN:NE2	2.21	0.56
1:A:372:GLN:HA	1:A:372:GLN:OE1	2.05	0.56
1:B:340:ARG:HG2	1:B:342:ASP:OD1	2.06	0.55
1:A:245:MET:SD	1:A:307:VAL:HG22	2.48	0.54
1:B:334:CYS:HA	1:B:389:ALA:HB2	1.89	0.53
1:B:370:VAL:HG13	1:B:390:GLU:HB2	1.92	0.52
1:A:292:MET:CE	1:B:220:GLY:HA2	2.40	0.51
1:B:441:ASP:OD1	1:B:443:LYS:HE2	2.13	0.49
1:B:248:GLU:HG3	1:B:539:GLN:HG2	1.94	0.48
1:A:457:SER:HB3	1:A:511:THR:HG23	1.96	0.47



Atom-1	Atom-2	Interatomic	Clash
Atom-1	At0111-2	distance $(Å)$	overlap (Å)
1:A:336:GLU:H	1:A:336:GLU:HG2	1.47	0.46
1:A:331:TYR:CE1	1:A:392:GLY:HA3	2.53	0.44
1:A:329:VAL:HG23	1:A:396:LEU:HD11	1.99	0.43
1:B:546:GLU:HG2	1:B:549:ARG:HH12	1.83	0.43
1:A:288:LEU:HA	1:B:288:LEU:HD23	2.01	0.43
1:A:337:VAL:HG12	1:A:338:ASP:N	2.34	0.43
1:A:429:GLU:OE2	1:A:490:LYS:HD3	2.19	0.42
1:B:334:CYS:HA	1:B:389:ALA:CB	2.50	0.42
1:B:308:GLU:HG3	3:B:723:HOH:O	2.19	0.42
1:B:459:ASP:OD2	1:B:461:HIS:ND1	2.53	0.41
1:B:288:LEU:HD23	1:B:288:LEU:HA	1.86	0.41
1:A:292:MET:HE1	1:B:220:GLY:HA2	2.02	0.41
1:B:223:CYS:HA	1:B:465:TYR:CE1	2.55	0.41
1:A:329:VAL:CG2	1:A:396:LEU:HD11	2.50	0.41
1:B:379:TYR:O	1:B:386:LEU:HA	2.21	0.41
1:B:459:ASP:OD1	1:B:463:GLY:N	2.46	0.41
1:B:256:VAL:N	1:B:257:PRO:CD	2.85	0.40

Continued from previous page...

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	А	331/360~(92%)	314~(95%)	17~(5%)	0	100 100		
1	В	327/360~(91%)	315~(96%)	12~(4%)	0	100 100		
All	All	658/720~(91%)	629 (96%)	29~(4%)	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 side chain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Rotameric Outliers	
1	А	302/327~(92%)	265~(88%)	37~(12%)	4 11
1	В	300/327~(92%)	256~(85%)	44 (15%)	2 7
All	All	602/654~(92%)	521 (86%)	81 (14%)	3 9

All (81) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	214	VAL
1	А	248	GLU
1	А	254	LYS
1	А	260	LEU
1	А	272	LYS
1	А	277	LYS
1	А	279	LEU
1	А	292	MET
1	А	296	VAL
1	А	299	LEU
1	А	301	ARG
1	А	305	ASP
1	А	312	LYS
1	А	316	VAL
1	А	325	ARG
1	А	327	LYS
1	А	336	GLU
1	А	343	ARG
1	А	355	LYS
1	А	402	LEU
1	А	403	HIS
1	A	417	GLN
1	A	418	ASN
1	А	439	LYS
1	А	454	LEU
1	А	460	ASN
1	А	470	ASN



Mol	Chain	Res	Type
1	А	476	LYS
1	А	487	ARG
1	А	511	THR
1	А	512	ASN
1	А	524	LEU
1	А	525	SER
1	А	540	LEU
1	А	549	ARG
1	А	551	GLU
1	А	554	LYS
1	В	210	HIS
1	В	219	GLN
1	В	228	LEU
1	В	229	LEU
1	В	238	LEU
1	В	252	SER
1	В	258	LEU
1	В	260	LEU
1	В	279	LEU
1	В	280	THR
1	В	287	THR
1	В	288	LEU
1	В	299	LEU
1	В	301	ARG
1	В	312	LYS
1	В	316	VAL
1	В	330	SER
1	В	343	ARG
1	В	345	GLU
1	В	358	LYS
1	В	370	VAL
1	В	376	ASP
1	В	383	GLU
1	В	387	GLN
1	В	402	LEU
1	В	410	MET
1	В	413	PRO
1	В	417	GLN
1	В	419	ILE
1	В	439	LYS
1	В	440	THR
1	В	443	LYS

Continued from previous page...



Mol	Chain	Res	Type
1	В	454	LEU
1	В	457	SER
1	В	459	ASP
1	В	460	ASN
1	В	461	HIS
1	В	470	ASN
1	В	487	ARG
1	В	510	CYS
1	В	525	SER
1	В	546	GLU
1	В	551	GLU
1	В	553	GLN

Continued from previous page...

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	218	ASN
1	А	384	HIS
1	А	512	ASN
1	В	237	GLN
1	В	418	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)

2 ligands are 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).

Mal	Truno	Chain	Dec	Bog Link Bond lengths		$_{\rm ths}$	B	ond ang	les	
	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	A1ICW	В	601	-	46,46,46	0.53	0	$58,\!65,\!65$	1.04	5 (8%)
2	A1ICW	А	601	-	46,46,46	0.48	0	58,65,65	0.80	2 (3%)

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
2	A1ICW	В	601	-	-	4/22/47/47	0/6/6/6
2	A1ICW	А	601	-	-	4/22/47/47	0/6/6/6

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
2	В	601	A1ICW	O1-C1-N2	-3.95	117.04	121.67
2	В	601	A1ICW	C14-C13-N1	3.21	115.65	110.63
2	А	601	A1ICW	C14-C13-N1	-2.85	106.17	110.63
2	В	601	A1ICW	C5-N1-C13	-2.75	105.20	111.66
2	А	601	A1ICW	C31-C32-N3	2.68	115.36	113.80
2	В	601	A1ICW	C31-C32-N3	2.38	115.19	113.80
2	В	601	A1ICW	C2-C1-N2	2.21	123.19	118.99

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	В	601	A1ICW	O1-C1-N2-C34
2	А	601	A1ICW	C6-C5-N1-C4
2	А	601	A1ICW	N1-C5-C6-C7
2	В	601	A1ICW	N1-C5-C6-C7



Mol	Chain	Res	Type	Atoms
2	В	601	A1ICW	O1-C1-N2-C21
2	А	601	A1ICW	C5-C6-C7-C12
2	В	601	A1ICW	C2-C1-N2-C34
2	A	601	A1ICW	C5-C6-C7-C8

Continued from previous page...

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.









### 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	$\langle RSRZ \rangle$	#RSRZ>2	$OWAB(Å^2)$	Q<0.9
1	А	335/360~(93%)	-0.69	1 (0%) 90 88	39, 64, 110, 189	0
1	В	333/360~(92%)	-0.71	1 (0%) 90 88	35, 63, 108, 141	0
All	All	668/720~(92%)	-0.70	2 (0%) 90 88	35, 64, 109, 189	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	500	GLY	2.9
1	В	500	GLY	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)

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
2	A1ICW	А	601	41/41	0.86	0.16	$60,\!113,\!184,\!191$	1
2	A1ICW	В	601	41/41	0.94	0.09	30,61,80,83	1



The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.







### 6.5 Other polymers (i)

There are no such residues in this entry.

