wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 11:20 pm BST

PDB ID	:	4TY9
Title	:	An Ligand-observed Mass Spectrometry-based Approach Integrated into the Fragment Based Lead Discovery Pipeline
Authors	:	Shui, W.; Yang, C.; Lin, J.; Chen, X.; Qin, S.; Chen, S.
Deposited on	:	2014-07-08
Resolution	:	2.78 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 following versions of software and data (see references (1)) were used in the production of this report:

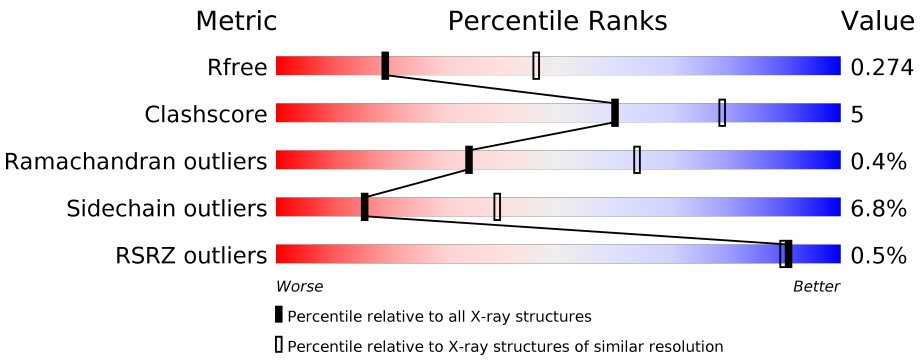
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.78 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	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (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 on 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	566	84% 10% • •
1	B	566	% 82% 12% • •
1	C	566	81% 12% • •
1	D	566	82% 12% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 16883 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.

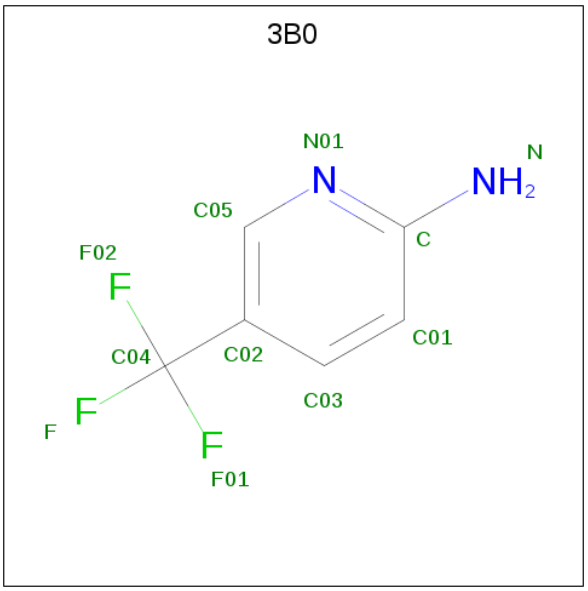
- Molecule 1 is a protein called Polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	544	4218	2659	743	784	32	0	0	0
1	B	544	4218	2659	743	784	32	0	0	0
1	C	544	4218	2659	743	784	32	0	0	0
1	D	544	4218	2659	743	784	32	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	539	LEU	ILE	engineered mutation	UNP D0PY27
A	549	SER	GLY	engineered mutation	UNP D0PY27
A	550	MET	TRP	engineered mutation	UNP D0PY27
B	539	LEU	ILE	engineered mutation	UNP D0PY27
B	549	SER	GLY	engineered mutation	UNP D0PY27
B	550	MET	TRP	engineered mutation	UNP D0PY27
C	539	LEU	ILE	engineered mutation	UNP D0PY27
C	549	SER	GLY	engineered mutation	UNP D0PY27
C	550	MET	TRP	engineered mutation	UNP D0PY27
D	539	LEU	ILE	engineered mutation	UNP D0PY27
D	549	SER	GLY	engineered mutation	UNP D0PY27
D	550	MET	TRP	engineered mutation	UNP D0PY27

- Molecule 2 is 5-(trifluoromethyl)pyridin-2-amine (three-letter code: 3B0) (formula: $C_6H_5F_3N_2$).



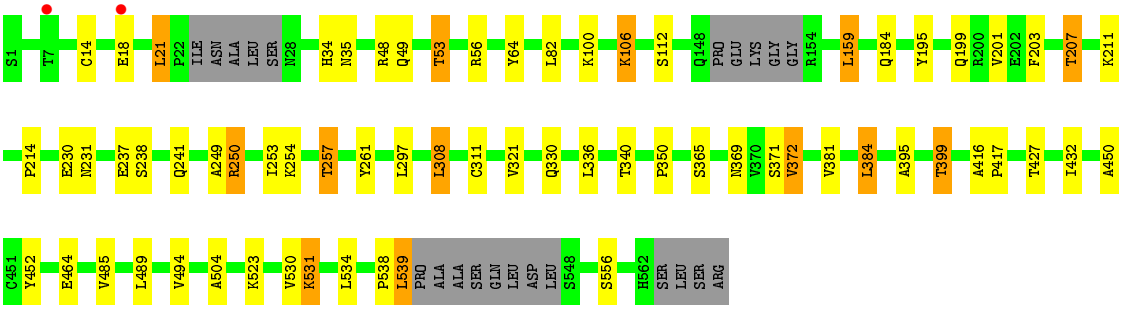
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	F	N		
2	A	1	11	6	3	2	0	0

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.

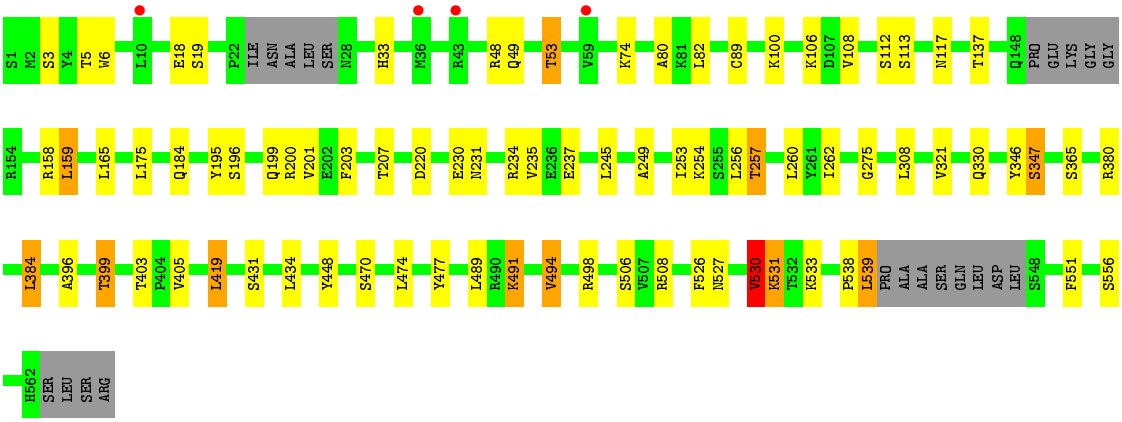
- Molecule 1: Polyprotein

Chain A:



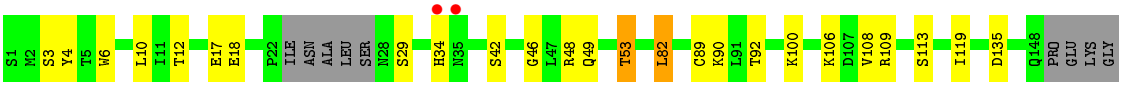
- Molecule 1: Polyprotein

Chain B:



- Molecule 1: Polyprotein

Chain C:



GLY R154 L159 C170 M173 D177 T181 Q184 Y195 S196 Q199 R200 V201 E202 F203 T207 F217 E230 N231 E237 A249 I253 T257 T267 C274 L308 C311 T312 M313 V321 V322 Q330 L336 A348 P349 D352

S367 L384 T385 R386 T403 L419 F430 L434 E437 D444 C445 Q446 I447 Y448 Y452 S453 I454 S478 S497 C468 L489 L492 G493 V494 R508 V530 K531 T532 L539 PRO ALA ALA SER GLN LEU ASP LEU S548 S549 M550 F551 S556 H562

SER LEU SER ARG

- Molecule 1: Polyprotein

Chain D:

S1 L21 P22 ILE ASN ALA LEU SER N28 N35 T40 R43 R48 Q49 T53 R56 K79 L82 K100 K106 E128 D129 T130 V131 V147 Q148 PRO GLU LYS GLY GLY R154 A157 R158 L159 K172 Q184 M187 Y191 Y195 S196

Q199 F203 T207 F217 R222 S226 E230 I233 R234 E237 A249 I253 K254 S255 L256 T257 E258 R259 L260 Y261 I262 G263 G264 Y276 R277 C303 L308 V321 Q330 L336 Y346 S347 S365 S368 T385 R386 A395

T399 T403 A416 P417 T418 L419 M423 T427 L434 E440 Y448 Y477 L489 V494 R498 R503 A504 V507 W528 A529 V530 K531 L539 PRO ALA ALA SER GLN LEU ASP LEU S548 F551 I560 Y561 H562 SER LEU SER ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	101.77Å 101.64Å 251.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.79 – 2.78 42.79 – 2.78	Depositor EDS
% Data completeness (in resolution range)	99.5 (42.79-2.78) 99.5 (42.79-2.78)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.38 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.208 , 0.273 0.211 , 0.274	Depositor DCC
R_{free} test set	3368 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	37.9	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 18.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.085 for k,h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	16883	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.90% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 3B0

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	0.57	0/4307	0.76	0/5842
1	B	0.55	0/4307	0.77	2/5842 (0.0%)
1	C	0.55	0/4307	0.77	1/5842 (0.0%)
1	D	0.56	0/4307	0.76	2/5842 (0.0%)
All	All	0.56	0/17228	0.76	5/23368 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	234	ARG	NE-CZ-NH1	7.43	124.02	120.30
1	B	419	LEU	CA-CB-CG	6.05	129.22	115.30
1	D	234	ARG	NE-CZ-NH2	-5.60	117.50	120.30
1	C	444	ASP	CB-CG-OD1	5.32	123.09	118.30
1	B	158	ARG	NE-CZ-NH1	5.18	122.89	120.30

There are no chirality outliers.

There are no planarity outliers.

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	A	4218	0	4213	33	0
1	B	4218	0	4213	39	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	4218	0	4213	40	0
1	D	4218	0	4213	43	0
2	A	11	0	5	1	0
All	All	16883	0	16857	155	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 5.

The worst 5 of 155 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:B:49:GLN:O	1:B:53:THR:CG2	2.21	0.89
1:D:49:GLN:O	1:D:53:THR:HG22	1.73	0.87
1:D:253:ILE:O	1:D:257:THR:HG23	1.76	0.86
1:C:203:PHE:O	1:C:207:THR:HG23	1.77	0.85
1:C:237:GLU:HG3	1:C:257:THR:HG21	1.60	0.83

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	536/566 (95%)	509 (95%)	26 (5%)	1 (0%)	47 76
1	B	536/566 (95%)	506 (94%)	27 (5%)	3 (1%)	25 54
1	C	536/566 (95%)	509 (95%)	25 (5%)	2 (0%)	34 64
1	D	536/566 (95%)	511 (95%)	23 (4%)	2 (0%)	34 64
All	All	2144/2264 (95%)	2035 (95%)	101 (5%)	8 (0%)	34 64

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	531	LYS
1	B	530	VAL
1	C	531	LYS
1	D	531	LYS
1	B	531	LYS

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	461/483 (95%)	432 (94%)	29 (6%)	18	43
1	B	461/483 (95%)	429 (93%)	32 (7%)	15	38
1	C	461/483 (95%)	425 (92%)	36 (8%)	12	32
1	D	461/483 (95%)	432 (94%)	29 (6%)	18	43
All	All	1844/1932 (95%)	1718 (93%)	126 (7%)	16	39

5 of 126 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	498	ARG
1	C	159	LEU
1	D	368	SER
1	B	530	VAL
1	C	34	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	406	ASN
1	C	33	HIS
1	D	199	GLN
1	C	28	ASN
1	C	58	GLN

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)

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	$\#\|Z\| > 2$
2	3B0	A	601	-	11,11,11	0.95	1 (9%)	16,16,16	1.42	3 (18%)

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	3B0	A	601	-	-	0/6/6/6	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	3B0	C-N	2.71	1.43	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	601	3B0	F02-C04-C02	-2.35	107.77	112.93
2	A	601	3B0	F-C04-C02	-2.10	108.31	112.93
2	A	601	3B0	F-C04-F02	2.06	113.27	105.72

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	3B0	1	0

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, 95th 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	544/566 (96%)	-0.18	2 (0%)	92	92	20, 32, 52, 81	0
1	B	544/566 (96%)	-0.13	4 (0%)	87	86	19, 34, 55, 84	0
1	C	544/566 (96%)	-0.16	2 (0%)	92	92	18, 32, 53, 77	0
1	D	544/566 (96%)	-0.18	2 (0%)	92	92	19, 32, 54, 87	0
All	All	2176/2264 (96%)	-0.16	10 (0%)	91	90	18, 33, 54, 87	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	10	LEU	4.2
1	D	43	ARG	3.0
1	A	7	THR	2.8
1	B	43	ARG	2.8
1	C	34	HIS	2.8

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, 95th 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	3B0	A	601	11/11	0.90	0.30	77,81,83,88	0

6.5 Other polymers (i)

There are no such residues in this entry.