wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 01:51 am BST

PDB ID	:	5NPV
Title	:	Structure of human ATG5-ATG16L1(ATG5BD) complex (I4)
Authors	:	Archna, A.; Scrima, A.
Deposited on	:	2017-04-19
Resolution	:	3.10 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:

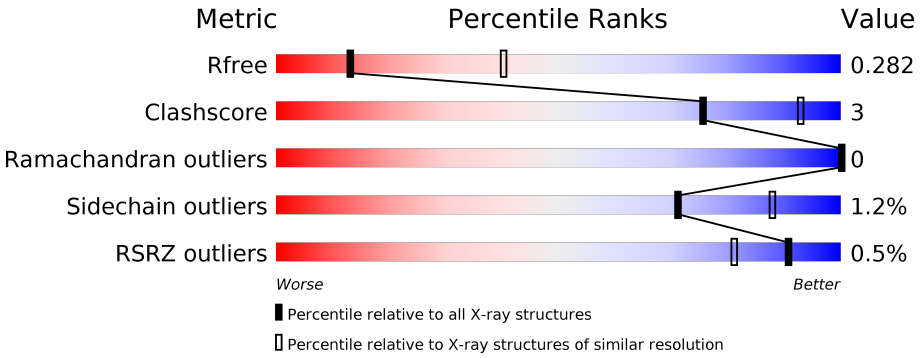
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	282	81% 9% 10%
1	C	282	% 78% 6% 16%
2	B	301	12% 87%
2	D	301	12% 87%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9055 atoms, of which 4421 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 Autophagy protein 5.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	255	Total	C	H	N	O	S	0	0	0
			4026	1347	1968	336	363	12			
1	C	237	Total	C	H	N	O	S	0	0	0
			3781	1263	1851	317	339	11			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLY	-	expression tag	UNP Q9H1Y0
A	-5	ALA	-	expression tag	UNP Q9H1Y0
A	-4	HIS	-	expression tag	UNP Q9H1Y0
A	-3	MET	-	expression tag	UNP Q9H1Y0
A	-2	SER	-	expression tag	UNP Q9H1Y0
A	-1	GLY	-	expression tag	UNP Q9H1Y0
A	0	ARG	-	expression tag	UNP Q9H1Y0
C	-6	GLY	-	expression tag	UNP Q9H1Y0
C	-5	ALA	-	expression tag	UNP Q9H1Y0
C	-4	HIS	-	expression tag	UNP Q9H1Y0
C	-3	MET	-	expression tag	UNP Q9H1Y0
C	-2	SER	-	expression tag	UNP Q9H1Y0
C	-1	GLY	-	expression tag	UNP Q9H1Y0
C	0	ARG	-	expression tag	UNP Q9H1Y0

- Molecule 2 is a protein called Autophagy-related protein 16-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	38	Total	C	H	N	O	0	0	0
			610	198	298	60	54			
2	D	39	Total	C	H	N	O	0	0	0
			627	203	304	61	59			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	7	GLY	-	expression tag	UNP Q676U5
B	8	GLY	-	expression tag	UNP Q676U5
B	9	GLY	-	expression tag	UNP Q676U5
B	10	ARG	-	expression tag	UNP Q676U5
D	7	GLY	-	expression tag	UNP Q676U5
D	8	GLY	-	expression tag	UNP Q676U5
D	9	GLY	-	expression tag	UNP Q676U5
D	10	ARG	-	expression tag	UNP Q676U5

- Molecule 3 is water.

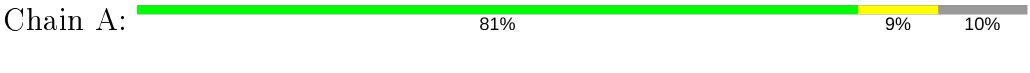
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	6	Total O 6 6	0	0
3	C	5	Total O 5 5	0	0

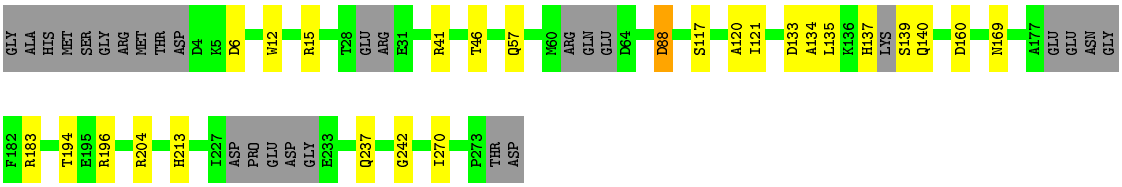
3 Residue-property plots (i)

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: Autophagy protein 5

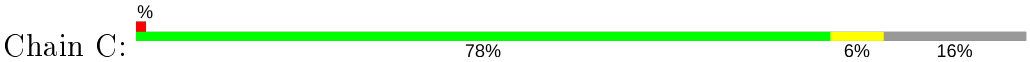
Chain A:

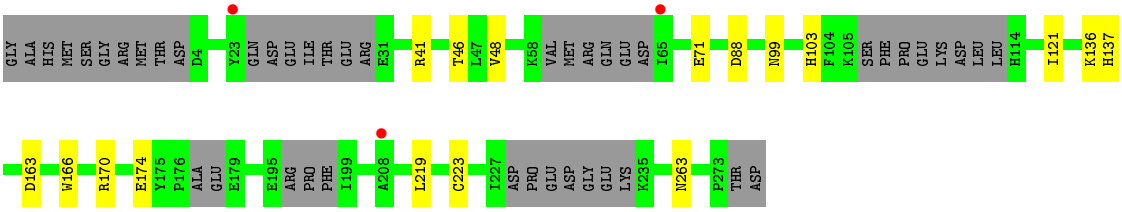




- Molecule 1: Autophagy protein 5

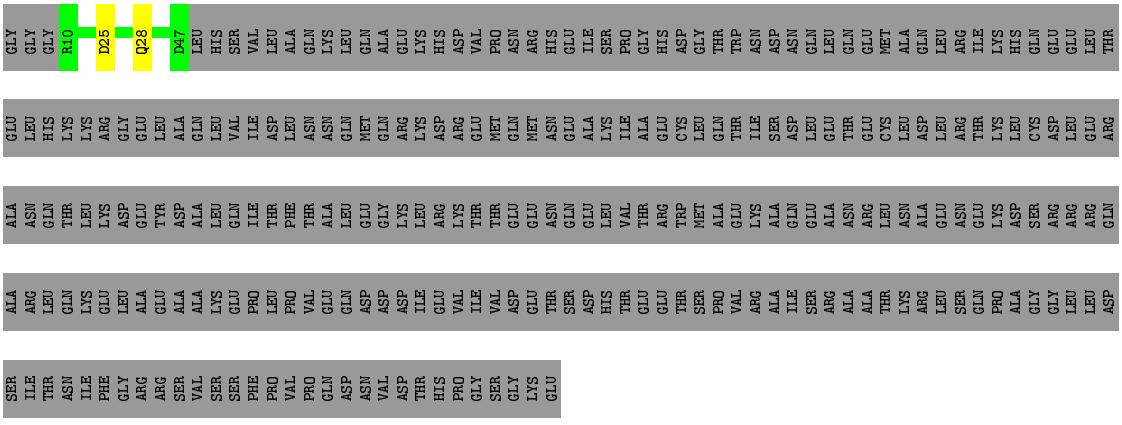
Chain C:





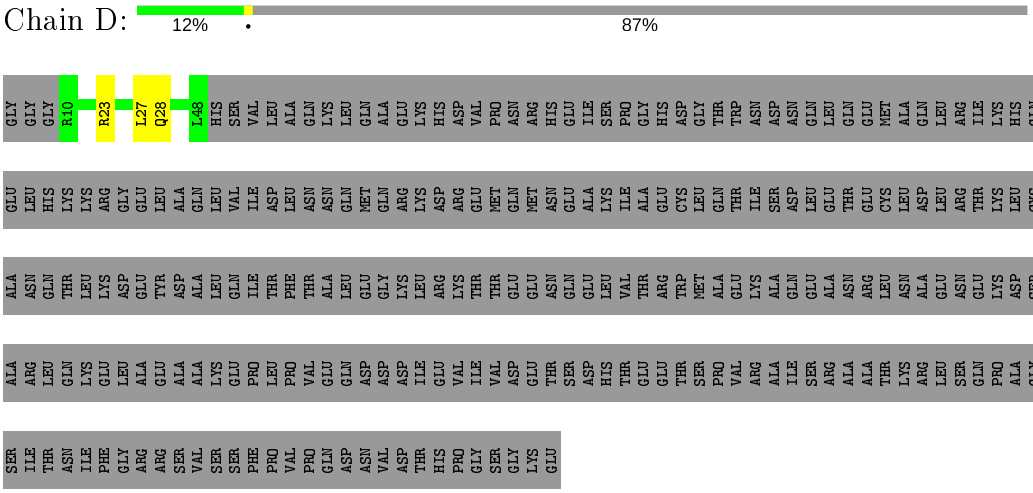
- Molecule 2: Autophagy-related protein 16-1

Chain B:



- Molecule 2: Autophagy-related protein 16-1

Chain D:


12%
87%
GLY GLY GLY R10 R23 L27 Q28 L48 HIS SER VAL LEU ALA GLN LYS LEU GLN ALA GLU LYS HIS ASP VAL PRO ASN ARG HIS GLU ILE SER PRO GLY HIS ASP GLY THR TRP ASN ASP ASN GLN LEU GLN GLU MET ALA GLN LEU ARG ILE LYS HIS GLN GLU GLU LEU THR
GLU LEU HIS LYS LYS ARG GLY GLU LEU ALA GLN LEU VAL ILE ASP LEU ASN ASN GLN MET GLN ARG LYS ASP ARG GLU MET GLN MET ASN GLU ALA LYS ILE ALA GLU CYS LEU GLN THR ILE SER ASP LEU GLU THR GLU CYS LEU ASP LEU ARG THR LYS LEU CYS ASP LEU GLU ARG
ALA ASN GLN THR LEU LYS ASP GLU TYR ASP ALA LEU GLN ILE THR PHE THR ALA LEU GLU GLY LYS LEU ARG LYS THR THR GLU GLU ASN GLN GLU LEU VAL THR ARG TRP MET ALA GLU LYS ALA GLN GLU ALA ASN ARG LEU ASN ALA GLU ASN GLU LYS ASP SER ARG ARG ARG GLN
ALA ARG LEU GLN LYS GLU LEU ALA GLU ALA LYS GLU PRO LEU PRO VAL GLU GLN ASP ASP ASP ILE GLU VAL ILE VAL ASP GLU THR SER ASP HIS THR GLU GLU THR SER PRO VAL ARG ALA ILE SER ARG ALA ALA THR LYS ARG LEU SER GLN PRO ALA GLY GLY LEU LEU ASP
SER ILE THR ASN ILE PHE GLY ARG ARG SER VAL SER SER PHE PRO VAL PRO GLN ASP ASN VAL ASP THR HIS PRO GLY SER GLY LYS GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α, β, γ	143.58Å 143.58Å 62.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.40 – 3.10 45.40 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.40-3.10) 99.6 (45.40-3.10)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.82 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.10_2155	Depositor
R, R_{free}	0.253 , 0.278 0.257 , 0.282	Depositor DCC
R_{free} test set	584 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	49.2	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 20.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.022 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	9055	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.87% 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	0.24	0/2119	0.39	0/2880
1	C	0.25	0/1987	0.40	0/2700
2	B	0.23	0/317	0.38	0/426
2	D	0.23	0/328	0.36	0/441
All	All	0.24	0/4751	0.39	0/6447

There are no bond length outliers.

There are no bond angle outliers.

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	2058	1968	1967	16	0
1	C	1930	1851	1850	11	0
2	B	312	298	298	1	0
2	D	323	304	304	2	0
3	A	6	0	0	0	0
3	C	5	0	0	1	0
All	All	4634	4421	4419	28	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.

The worst 5 of 28 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:A:204:ARG:O	1:A:213:HIS:ND1	2.20	0.74
1:C:99:ASN:OD1	3:C:301:HOH:O	2.05	0.73
1:C:41:ARG:NH2	1:C:88:ASP:OD1	2.24	0.70
1:C:136:LYS:NZ	1:C:263:ASN:OD1	2.25	0.70
1:A:41:ARG:NH1	2:B:28:GLN:OE1	2.29	0.66

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	243/282 (86%)	229 (94%)	14 (6%)	0	100	100
1	C	223/282 (79%)	216 (97%)	7 (3%)	0	100	100
2	B	36/301 (12%)	33 (92%)	3 (8%)	0	100	100
2	D	37/301 (12%)	34 (92%)	3 (8%)	0	100	100
All	All	539/1166 (46%)	512 (95%)	27 (5%)	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 sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	218/258 (84%)	213 (98%)	5 (2%)	50	77
1	C	206/258 (80%)	206 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	30/267 (11%)	29 (97%)	1 (3%)	38 69
2	D	32/267 (12%)	32 (100%)	0	100 100
All	All	486/1050 (46%)	480 (99%)	6 (1%)	71 88

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	88	ASP
2	B	25	ASP
1	A	135	LEU
1	A	57	GLN
1	A	169	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	255/282 (90%)	0.05	0 100 100	21, 49, 78, 90	0
1	C	237/282 (84%)	0.06	3 (1%) 77 59	24, 42, 71, 96	0
2	B	38/301 (12%)	-0.29	0 100 100	25, 47, 69, 78	0
2	D	39/301 (12%)	-0.17	0 100 100	22, 37, 85, 94	0
All	All	569/1166 (48%)	0.02	3 (0%) 91 81	21, 45, 75, 96	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	208	ALA	2.7
1	C	65	ILE	2.4
1	C	23	TYR	2.4

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.