Full wwPDB X-ray Structure Validation Report ⓘ

Nov 7, 2024 – 03:54 pm GMT

PDB ID	:	9EX5
Title	:	Crystal structure of Yeast Clathrin Heavy Chain N-terminal domain bound to Epsin-5 peptide (LIDL)
Authors	:	Defelipe, L.A.; Bento, I.; Garcia Alai, M.M.
Deposited on	:	2024-04-05
Resolution	:	2.01 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 ⓘ 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 ⓘ) were used in the production of this report:

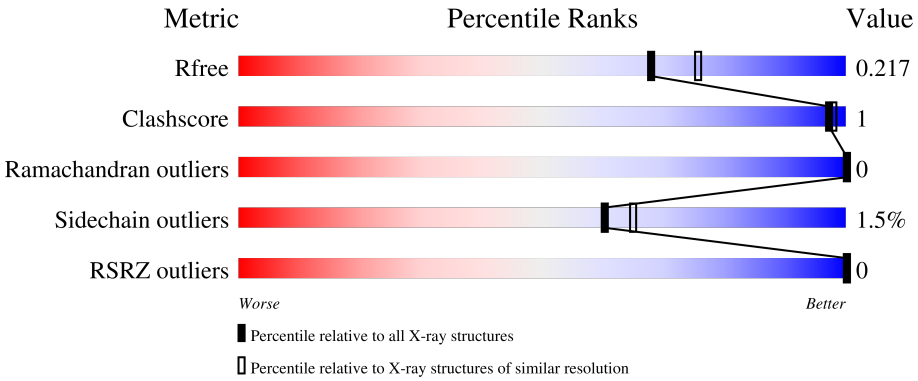
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.01 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}	164625	12358 (2.04-2.00)
Clashscore	180529	13897 (2.04-2.00)
Ramachandran outliers	177936	13770 (2.04-2.00)
Sidechain outliers	177891	13769 (2.04-2.00)
RSRZ outliers	164620	12358 (2.04-2.00)

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	A	373	93%
1	B	373	92%
1	C	373	95%
2	D	9	67% 33%
2	E	9	44% 11% 44%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	9	56% 11% 33%
2	G	9	78% 22%
2	H	9	78% 22%
2	I	9	56% 44%
2	J	9	67% 11% 22%
2	L	9	56% 22% 22%
2	M	9	78% 22%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 18509 atoms, of which 9007 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 Clathrin heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	363	Total	C	H	N	O	S	76	0	0
			5663	1799	2844	480	535	5			
1	B	362	Total	C	H	N	O	S	76	0	0
			5645	1793	2834	479	534	5			
1	C	363	Total	C	H	N	O	S	76	0	0
			5663	1799	2844	480	535	5			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P22137
A	-2	ALA	-	expression tag	UNP P22137
A	-1	MET	-	expression tag	UNP P22137
A	0	ALA	-	expression tag	UNP P22137
B	-3	GLY	-	expression tag	UNP P22137
B	-2	ALA	-	expression tag	UNP P22137
B	-1	MET	-	expression tag	UNP P22137
B	0	ALA	-	expression tag	UNP P22137
C	-3	GLY	-	expression tag	UNP P22137
C	-2	ALA	-	expression tag	UNP P22137
C	-1	MET	-	expression tag	UNP P22137
C	0	ALA	-	expression tag	UNP P22137

- Molecule 2 is a protein called Epsin-5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	6	Total	C	H	N	O	0	0	0
			96	31	49	6	10			
2	E	5	Total	C	H	N	O	0	0	0
			81	26	41	5	9			
2	F	6	Total	C	H	N	O	0	0	0
			96	31	49	6	10			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	7	Total	C	H	N	O	0	0	0
			116	37	61	7	11			
2	H	7	Total	C	H	N	O	0	0	0
			116	37	61	7	11			
2	I	5	Total	C	H	N	O	0	0	0
			81	26	41	5	9			
2	J	7	Total	C	H	N	O	0	0	0
			116	37	61	7	11			
2	L	7	Total	C	H	N	O	0	0	0
			116	37	61	7	11			
2	M	7	Total	C	H	N	O	0	0	0
			116	37	61	7	11			

- Molecule 3 is water.

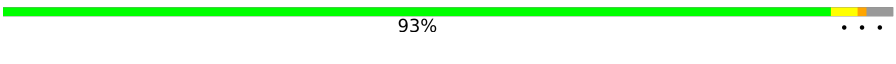
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	190	Total	O	0	0
			190	190		
3	B	200	Total	O	0	0
			200	200		
3	C	193	Total	O	0	0
			193	193		
3	D	2	Total	O	0	0
			2	2		
3	E	2	Total	O	0	0
			2	2		
3	F	2	Total	O	0	0
			2	2		
3	I	5	Total	O	0	0
			5	5		
3	J	4	Total	O	0	0
			4	4		
3	L	5	Total	O	0	0
			5	5		
3	M	1	Total	O	0	0
			1	1		

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.

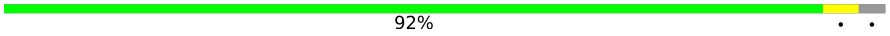
• Molecule 1: Clathrin heavy chain

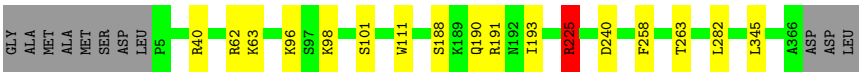
Chain A:



• Molecule 1: Clathrin heavy chain

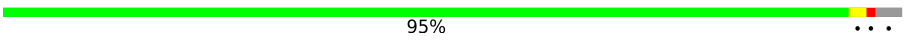
Chain B:





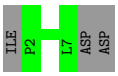
• Molecule 1: Clathrin heavy chain

Chain C:



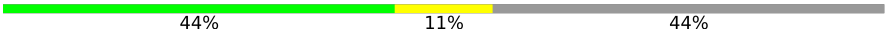
• Molecule 2: Epsin-5

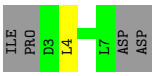
Chain D:



• Molecule 2: Epsin-5

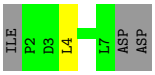
Chain E:





• Molecule 2: Epsin-5

Chain F:



• Molecule 2: Epsin-5

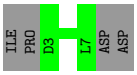
Chain G:

• Molecule 2: Epsin-5

Chain H:

• Molecule 2: Epsin-5

Chain I:

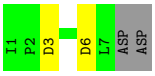


• Molecule 2: Epsin-5

Chain J:

• Molecule 2: Epsin-5

Chain L:



• Molecule 2: Epsin-5

Chain M:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	77.00Å 133.45Å 285.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	71.48 – 2.01 71.48 – 2.01	Depositor EDS
% Data completeness (in resolution range)	95.9 (71.48-2.01) 95.9 (71.48-2.01)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 2.02Å)	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
R, R_{free}	0.182 , 0.215 0.183 , 0.217	Depositor DCC
R_{free} test set	4607 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	33.6	Xtriage
Anisotropy	0.549	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 28.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.480 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.479 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	18509	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.32% 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	0.42	0/2872	0.77	3/3907 (0.1%)
1	B	0.41	0/2864	0.76	3/3895 (0.1%)
1	C	0.41	0/2872	0.84	7/3907 (0.2%)
2	D	0.44	0/47	0.80	0/63
2	E	0.46	0/39	1.36	0/52
2	F	0.39	0/47	1.22	0/63
2	G	0.47	0/55	0.78	0/75
2	H	0.46	0/55	0.73	0/75
2	I	0.47	0/39	1.21	0/52
2	J	0.61	0/55	0.65	0/75
2	L	0.52	0/55	0.68	0/75
2	M	0.56	0/55	0.61	0/75
All	All	0.42	0/9055	0.80	13/12314 (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	A	0	4
1	B	0	1
1	C	0	4
All	All	0	9

There are no bond length outliers.

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	225	ARG	NE-CZ-NH1	-12.99	113.80	120.30
1	C	225	ARG	NE-CZ-NH1	-12.38	114.11	120.30
1	C	191	ARG	NE-CZ-NH2	-11.52	114.54	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	360	ARG	NE-CZ-NH2	10.99	125.80	120.30
1	C	360	ARG	NE-CZ-NH1	-10.47	115.07	120.30
1	A	225	ARG	NE-CZ-NH2	-7.85	116.37	120.30
1	B	225	ARG	NE-CZ-NH2	7.31	123.95	120.30
1	A	82	ARG	NE-CZ-NH2	-6.34	117.13	120.30
1	C	191	ARG	CG-CD-NE	-5.88	99.46	111.80
1	A	40	ARG	NE-CZ-NH1	-5.45	117.58	120.30
1	C	225	ARG	NE-CZ-NH2	5.43	123.02	120.30
1	C	191	ARG	NE-CZ-NH1	5.02	122.81	120.30
1	B	282	LEU	CB-CG-CD1	-5.02	102.47	111.00

There are no chirality outliers.

All (9) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	179	ARG	Sidechain
1	A	225	ARG	Sidechain
1	A	40	ARG	Sidechain
1	A	82	ARG	Sidechain
1	B	225	ARG	Sidechain
1	C	191	ARG	Sidechain
1	C	225	ARG	Sidechain
1	C	360	ARG	Sidechain
1	C	40	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	A	2819	2844	2836	5	0
1	B	2811	2834	2826	5	0
1	C	2819	2844	2836	3	0
2	D	47	49	48	0	0
2	E	40	41	40	1	0
2	F	47	49	48	1	0
2	G	55	61	61	0	0
2	H	55	61	61	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	I	40	41	40	0	0
2	J	55	61	61	1	0
2	L	55	61	61	1	0
2	M	55	61	61	0	0
3	A	190	0	0	0	0
3	B	200	0	0	0	0
3	C	193	0	0	2	0
3	D	2	0	0	0	0
3	E	2	0	0	0	0
3	F	2	0	0	0	0
3	I	5	0	0	0	0
3	J	4	0	0	0	0
3	L	5	0	0	0	0
3	M	1	0	0	0	0
All	All	9502	9007	8979	15	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 1.

All (15) 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:226:ASN:OD1	1:A:228:THR:HG22	1.97	0.63
1:B:191:ARG:HB3	1:B:193:ILE:HD12	1.87	0.56
2:J:1:ILE:HG13	2:J:1:ILE:O	2.06	0.55
1:A:191:ARG:HB3	1:A:193:ILE:HD12	1.87	0.54
1:C:360:ARG:HD2	3:C:531:HOH:O	2.09	0.53
1:B:62:ARG:O	1:B:63:LYS:HE2	2.10	0.51
1:A:228:THR:HG23	1:A:229:THR:HG23	1.94	0.49
1:B:188:SER:HB3	2:F:4:LEU:HD11	1.95	0.49
2:L:3:ASP:OD2	2:L:6:ASP:OD2	2.34	0.45
1:C:63:LYS:HG2	3:C:552:HOH:O	2.17	0.44
1:A:188:SER:HB3	2:E:4:LEU:HD11	1.99	0.44
1:B:225:ARG:HD2	1:B:258:PHE:CD1	2.54	0.42
1:B:98:LYS:HD3	1:B:101:SER:HB2	1.99	0.42
1:A:4:LEU:HD13	1:A:5:PRO:N	2.35	0.41
1:C:225:ARG:HD2	1:C:258:PHE:CD1	2.55	0.41

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	361/373 (97%)	355 (98%)	6 (2%)	0	100	100
1	B	360/373 (96%)	354 (98%)	6 (2%)	0	100	100
1	C	361/373 (97%)	354 (98%)	7 (2%)	0	100	100
2	D	4/9 (44%)	4 (100%)	0	0	100	100
2	E	3/9 (33%)	3 (100%)	0	0	100	100
2	F	4/9 (44%)	4 (100%)	0	0	100	100
2	G	5/9 (56%)	5 (100%)	0	0	100	100
2	H	5/9 (56%)	5 (100%)	0	0	100	100
2	I	3/9 (33%)	3 (100%)	0	0	100	100
2	J	5/9 (56%)	5 (100%)	0	0	100	100
2	L	5/9 (56%)	5 (100%)	0	0	100	100
2	M	5/9 (56%)	5 (100%)	0	0	100	100
All	All	1121/1200 (93%)	1102 (98%)	19 (2%)	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	312/319 (98%)	308 (99%)	4 (1%)	65	70
1	B	311/319 (98%)	304 (98%)	7 (2%)	45	47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	312/319 (98%)	308 (99%)	4 (1%)	65 70
2	D	6/9 (67%)	6 (100%)	0	100 100
2	E	5/9 (56%)	5 (100%)	0	100 100
2	F	6/9 (67%)	6 (100%)	0	100 100
2	G	7/9 (78%)	7 (100%)	0	100 100
2	H	7/9 (78%)	7 (100%)	0	100 100
2	I	5/9 (56%)	5 (100%)	0	100 100
2	J	7/9 (78%)	7 (100%)	0	100 100
2	L	7/9 (78%)	7 (100%)	0	100 100
2	M	7/9 (78%)	7 (100%)	0	100 100
All	All	992/1038 (96%)	977 (98%)	15 (2%)	60 65

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

Mol	Chain	Res	Type
1	A	4	LEU
1	A	96	LYS
1	A	111	TRP
1	A	345	LEU
1	B	40	ARG
1	B	96	LYS
1	B	111	TRP
1	B	190	GLN
1	B	240	ASP
1	B	263	THR
1	B	345	LEU
1	C	96	LYS
1	C	111	TRP
1	C	190	GLN
1	C	360	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	58	ASN
1	A	289	HIS
1	A	319	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	289	HIS
1	B	319	ASN
1	C	58	ASN
1	C	241	HIS
1	C	319	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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

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	363/373 (97%)	-1.42	0	100	100	28, 40, 59, 81	0
1	B	362/373 (97%)	-1.41	0	100	100	28, 40, 57, 82	0
1	C	363/373 (97%)	-1.41	0	100	100	28, 40, 59, 80	0
2	D	6/9 (66%)	-0.87	0	100	100	40, 50, 56, 63	0
2	E	5/9 (55%)	-0.87	0	100	100	43, 43, 46, 54	0
2	F	6/9 (66%)	-0.54	0	100	100	44, 47, 63, 98	0
2	G	7/9 (77%)	-1.18	0	100	100	41, 51, 62, 74	0
2	H	7/9 (77%)	-1.11	0	100	100	41, 51, 65, 81	0
2	I	5/9 (55%)	-0.91	0	100	100	43, 44, 47, 59	0
2	J	7/9 (77%)	-1.22	0	100	100	42, 44, 54, 66	0
2	L	7/9 (77%)	-1.33	0	100	100	43, 45, 54, 60	0
2	M	7/9 (77%)	-1.27	0	100	100	43, 44, 54, 59	0
All	All	1145/1200 (95%)	-1.40	0	100	100	28, 41, 60, 98	0

There are no RSRZ outliers to report.

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 monosaccharides 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.