



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2022 – 07:07 pm BST

PDB ID : 7ZKZ  
Title : Crystal structure of cystinosin from Arabidopsis thaliana bound to two nanobodies  
Authors : Loebel, M.; Newstead, S.  
Deposited on : 2022-04-13  
Resolution : 2.33 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 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 : 2.29  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.29

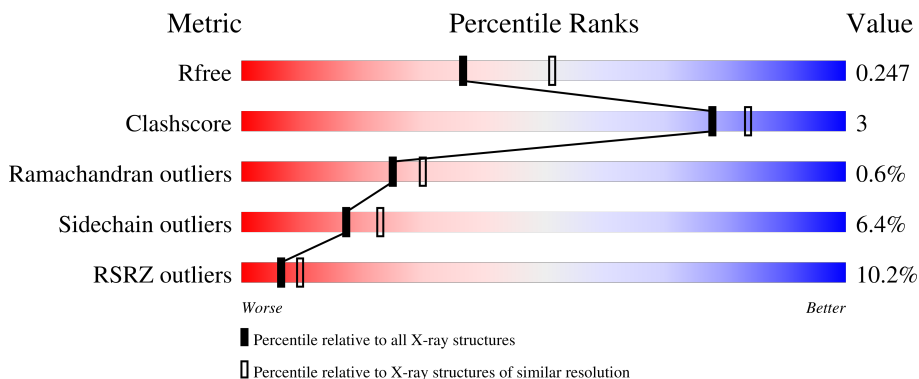
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.33 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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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  $\geq 3$ , 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	277	 7% 71% 11% 17%
2	B	128	 16% 90% 8% ..
3	C	123	 7% 85% 12% ..

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3891 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 Cystinosin homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	231	1882	1263	291	317	11	0	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	271	GLY	-	expression tag	UNP P57758
A	272	GLU	-	expression tag	UNP P57758
A	273	ASN	-	expression tag	UNP P57758
A	274	LEU	-	expression tag	UNP P57758
A	275	TYR	-	expression tag	UNP P57758
A	276	PHE	-	expression tag	UNP P57758
A	277	GLN	-	expression tag	UNP P57758

- Molecule 2 is a protein called Llama nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	127	982	613	176	189	4	0	2	0

- Molecule 3 is a protein called nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	122	920	568	164	185	3	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	38	Total	O	0	0
			38	38		

*Continued on next page...*

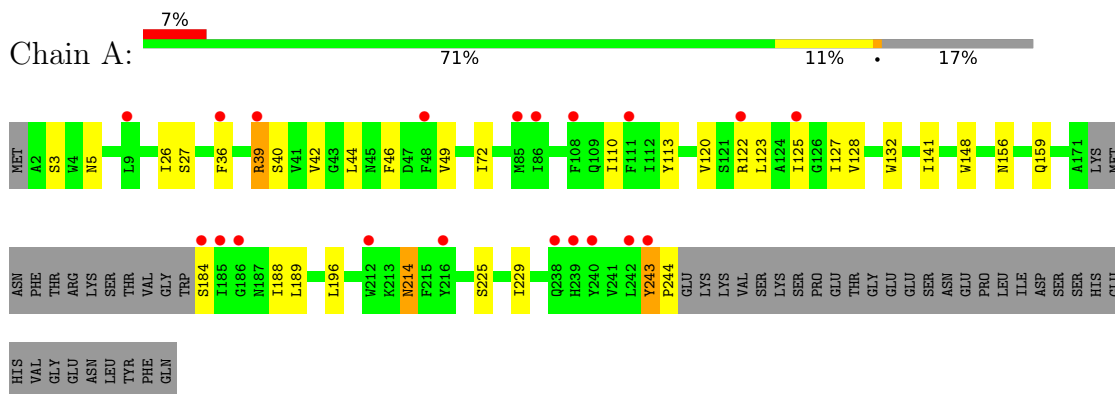
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	B	30	Total	O	0	0
			30	30		
4	C	39	Total	O	0	0
			39	39		

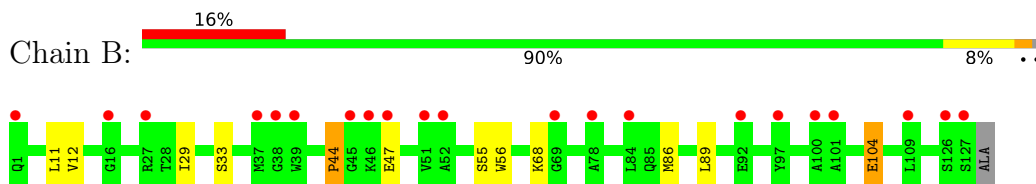
### 3 Residue-property plots [i](#)

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.

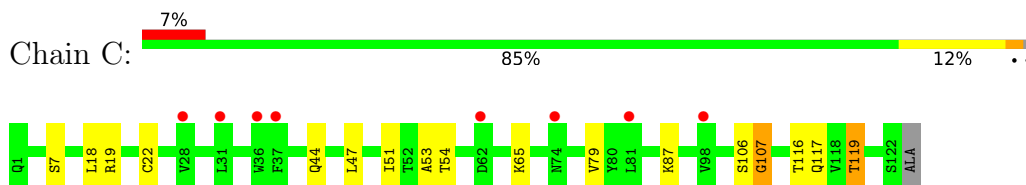
- Molecule 1: Cystinosin homolog



- Molecule 2: Llama nanobody



- Molecule 3: nanobody



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	210.28Å 77.46Å 46.30Å 90.00° 93.25° 90.00°	Depositor
Resolution (Å)	20.79 – 2.33 20.79 – 2.33	Depositor EDS
% Data completeness (in resolution range)	99.3 (20.79-2.33) 99.3 (20.79-2.33)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.16 (at 2.33Å)	Xtriage
Refinement program	BUSTER 2.10.4 (20-OCT-2021)	Depositor
R, $R_{free}$	0.235 , 0.265 0.223 , 0.247	Depositor DCC
$R_{free}$ test set	1566 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.0	Xtriage
Anisotropy	0.455	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3891	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.65% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0/1940	0.55	0/2642
2	B	0.44	0/1005	0.63	0/1363
3	C	0.44	0/939	0.70	0/1275
All	All	0.43	0/3884	0.61	0/5280

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 [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	1882	0	1880	17	0
2	B	982	0	957	4	0
3	C	920	0	877	4	0
4	A	38	0	0	0	0
4	B	30	0	0	0	0
4	C	39	0	0	0	0
All	All	3891	0	3714	22	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 (22) 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:243:TYR:HB2	1:A:244:PRO:HD3	1.78	0.66
1:A:40:SER:HA	1:A:113:TYR:HB3	1.89	0.53
1:A:36:PHE:CE1	1:A:39:ARG:NH2	2.76	0.53
3:C:106:SER:O	3:C:107:GLY:O	2.28	0.51
3:C:22:CYS:HB3	3:C:79:VAL:HG13	1.94	0.50
1:A:128:VAL:HG13	1:A:132:TRP:CD1	2.48	0.49
1:A:49:VAL:HG21	1:A:110:ILE:HD11	1.96	0.47
2:B:86:MET:HE2	2:B:89:LEU:HD21	1.96	0.47
1:A:214:ASN:HD22	1:A:214:ASN:H	1.62	0.47
1:A:123:LEU:O	1:A:127:ILE:HG12	2.14	0.47
1:A:225:SER:O	1:A:229:ILE:HG13	2.15	0.46
1:A:156:ASN:O	1:A:159:GLN:HG3	2.15	0.46
1:A:42:VAL:HG12	3:C:53:ALA:HB1	1.97	0.45
2:B:33:SER:O	2:B:56:TRP:HB2	2.17	0.45
1:A:122:ARG:HA	1:A:125:ILE:HG12	1.99	0.45
1:A:36:PHE:HE1	1:A:39:ARG:HH21	1.61	0.42
1:A:3:SER:HA	2:B:104:GLU:HA	2.01	0.42
1:A:72:ILE:HG12	1:A:148:TRP:HB2	2.02	0.41
1:A:5:ASN:HA	2:B:55:SER:HB2	2.02	0.40
3:C:117:GLN:HE21	3:C:119:THR:HG22	1.86	0.40
1:A:184:SER:O	1:A:188:ILE:HG12	2.21	0.40
1:A:46:PHE:HB3	1:A:120:VAL:HG22	2.04	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	227/277 (82%)	223 (98%)	3 (1%)	1 (0%)	34	41
2	B	127/128 (99%)	123 (97%)	3 (2%)	1 (1%)	19	23
3	C	120/123 (98%)	118 (98%)	1 (1%)	1 (1%)	19	23
All	All	474/528 (90%)	464 (98%)	7 (2%)	3 (1%)	25	30



All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	44	PRO
3	C	107	GLY
1	A	243	TYR

### 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	206/249 (83%)	198 (96%)	8 (4%)	32	45
2	B	103/102 (101%)	96 (93%)	7 (7%)	16	20
3	C	99/99 (100%)	88 (89%)	11 (11%)	6	6
All	All	408/450 (91%)	382 (94%)	26 (6%)	17	23

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	26	ILE
1	A	27	SER
1	A	39	ARG
1	A	44	LEU
1	A	141	ILE
1	A	189	LEU
1	A	196	LEU
1	A	214	ASN
2	B	11	LEU
2	B	12	VAL
2	B	29	ILE
2	B	44	PRO
2	B	47	GLU
2	B	68	LYS
2	B	104	GLU
3	C	7	SER
3	C	18	LEU
3	C	19	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	C	44	GLN
3	C	47	LEU
3	C	51	ILE
3	C	54	THR
3	C	65	LYS
3	C	87	LYS
3	C	116	THR
3	C	119	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	31	GLN
1	A	35	ASN
1	A	214	ASN
2	B	42	GLN
3	C	39	GLN
3	C	117	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 monosaccharides 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<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	231/277 (83%)	0.58	20 (8%) <b>10</b> <b>14</b>	49, 65, 99, 130	0
2	B	127/128 (99%)	0.73	21 (16%) <b>1</b> <b>2</b>	50, 61, 79, 90	0
3	C	122/123 (99%)	0.39	8 (6%) <b>18</b> <b>24</b>	48, 65, 78, 87	0
All	All	480/528 (90%)	0.57	49 (10%) <b>6</b> <b>10</b>	48, 65, 91, 130	0

All (49) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	111	PHE	8.2
1	A	242	LEU	7.2
1	A	240	TYR	6.1
1	A	36	PHE	6.0
1	A	184	SER	5.8
1	A	243	TYR	5.7
1	A	122	ARG	5.2
2	B	39	TRP	4.6
2	B	78	ALA	4.2
1	A	185	ILE	4.2
2	B	84	LEU	4.0
1	A	85	MET	3.9
3	C	31	LEU	3.8
3	C	62	ASP	3.8
3	C	98	VAL	3.7
2	B	100[A]	ALA	3.7
2	B	52	ALA	3.5
2	B	37	MET	3.5
2	B	127	SER	3.3
3	C	28	VAL	3.3
2	B	45	GLY	3.1
1	A	86	ILE	3.1
2	B	101	ALA	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	212	TRP	3.0
1	A	108	PHE	2.8
2	B	27	ARG	2.8
1	A	216	TYR	2.8
3	C	37	PHE	2.8
1	A	186	GLY	2.8
1	A	39	ARG	2.7
3	C	81	LEU	2.7
2	B	46	LYS	2.6
1	A	9	LEU	2.5
3	C	74	ASN	2.5
2	B	47	GLU	2.5
1	A	48	PHE	2.5
2	B	126	SER	2.5
2	B	51	VAL	2.4
2	B	69	GLY	2.3
2	B	97	TYR	2.3
3	C	36	TRP	2.3
2	B	92	GLU	2.2
2	B	38	GLY	2.2
1	A	125	ILE	2.2
1	A	239	HIS	2.2
2	B	16	GLY	2.1
2	B	109	LEU	2.0
1	A	238	GLN	2.0
2	B	1	GLN	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 [\(i\)](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.