W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

Full wwPDB X-ray Structure Validation Report (i)

May 23, 2022 – 01:08 pm BST

PDB ID	:	7R1Z
Title	:	C-terminal domain of hArc in complex with nanobodies H11 and C11, collapsed crystal form
Authors	:	Markusson, S.; Kursula, P.
Deposited on	:	2022-02-03
Resolution	:	1.94 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

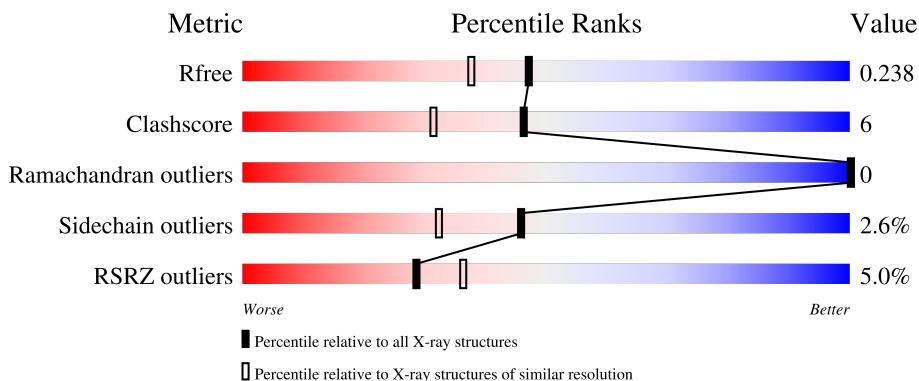|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 1.94 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	Percentile Ranks	Value
Rfree		0.238
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		2.6%
RSRZ outliers		5.0%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	E	128	86% 11% ··
2	A	181	2% 58% 7% · 34%
3	C	120	12% 81% 18% ·

WORLDWIDE PDB PROTEIN DATA BANK

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5909 atoms, of which 2815 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 NbArc-H11.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	E	125	Total 1936	C 618	H 949	N 174	O 189	S 6	0	3	0

- Molecule 2 is a protein called Activity-regulated cytoskeleton-associated protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	A	119	Total 2007	C 661	H 977	N 171	O 197	S 1	0	1	0

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	181	MET	-	initiating methionine	UNP Q7LC44
A	182	HIS	-	expression tag	UNP Q7LC44
A	183	HIS	-	expression tag	UNP Q7LC44
A	184	HIS	-	expression tag	UNP Q7LC44
A	185	HIS	-	expression tag	UNP Q7LC44
A	186	HIS	-	expression tag	UNP Q7LC44
A	187	HIS	-	expression tag	UNP Q7LC44
A	188	LEU	-	expression tag	UNP Q7LC44
A	189	GLU	-	expression tag	UNP Q7LC44
A	190	SER	-	expression tag	UNP Q7LC44
A	191	THR	-	expression tag	UNP Q7LC44
A	192	SER	-	expression tag	UNP Q7LC44
A	193	LEU	-	expression tag	UNP Q7LC44
A	194	TYR	-	expression tag	UNP Q7LC44
A	195	LYS	-	expression tag	UNP Q7LC44
A	196	LYS	-	expression tag	UNP Q7LC44
A	197	ALA	-	expression tag	UNP Q7LC44
A	198	GLY	-	expression tag	UNP Q7LC44
A	199	SER	-	expression tag	UNP Q7LC44
A	200	GLU	-	expression tag	UNP Q7LC44

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	ASN	-	expression tag	UNP Q7LC44
A	202	LEU	-	expression tag	UNP Q7LC44
A	203	TYR	-	expression tag	UNP Q7LC44
A	204	PHE	-	expression tag	UNP Q7LC44
A	205	GLN	-	expression tag	UNP Q7LC44

• Molecule 3 is a protein called NbArc-C11.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	C	118	1803	570	889	162	179	3	0	2	0

• Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	66	Total 66	O 66	0	0
4	A	70	Total 70	O 70	0	0
4	C	27	Total 27	O 27	0	0

3 Residue-property plots ⓘ

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: NbArc-H11

Chain E:

• Molecule 2: Activity-regulated cytoskeleton-associated protein

Chain A:
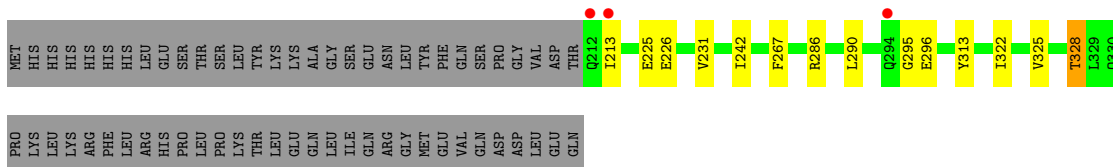
• Molecule 3: NbArc-C11

Chain C:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	66.26Å 139.29Å 43.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.00 – 1.94 48.00 – 1.94	Depositor EDS
% Data completeness (in resolution range)	86.5 (48.00-1.94) 80.1 (48.00-1.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.72 (at 1.94Å)	Xtriage
Refinement program	PHENIX dev_3958	Depositor
R, R_{free}	0.197 , 0.242 0.195 , 0.238	Depositor DCC
R_{free} test set	1999 reflections (7.57%)	wwPDB-VP
Wilson B-factor (Å^2)	31.0	Xtriage
Anisotropy	0.396	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5909	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.40% 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	E	0.29	0/1016	0.58	0/1372
2	A	0.28	0/1057	0.47	0/1427
3	C	0.27	0/949	0.55	0/1286
All	All	0.28	0/3022	0.54	0/4085

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	E	987	949	945	16	0
2	A	1030	977	976	18	0
3	C	914	889	870	11	0
4	A	70	0	0	2	0
4	C	27	0	0	2	0
4	E	66	0	0	3	0
All	All	3094	2815	2791	36	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 6.

All (36) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:109:ILE:HG12	2:A:213:ILE:HG23	1.60	0.81
1:E:110:SER:HB3	2:A:213:ILE:HG21	1.70	0.74
3:C:87:ARG:NH1	4:C:202:HOH:O	2.25	0.69
1:E:109:ILE:CG1	2:A:213:ILE:HG23	2.23	0.68
1:E:46:GLU:OE2	4:E:201:HOH:O	2.10	0.68
3:C:25:ALA:HA	3:C:80:ILE:HD13	1.78	0.66
3:C:85:MET:HE2	3:C:88:LEU:HD21	1.83	0.61
1:E:53:ILE:HD12	1:E:60:THR:HG22	1.84	0.60
1:E:109:ILE:HG12	2:A:213:ILE:CG2	2.33	0.58
2:A:325:VAL:O	2:A:328:THR:HG22	2.06	0.56
1:E:53:ILE:CD1	1:E:60:THR:HG22	2.38	0.54
3:C:115:GLN:OE1	4:C:201:HOH:O	2.18	0.54
3:C:24:CYS:HB3	3:C:81:VAL:HG22	1.88	0.53
1:E:53:ILE:HD11	1:E:57:GLY:HA2	1.92	0.50
2:A:225:GLU:OE1	4:A:401:HOH:O	2.20	0.50
1:E:3:GLU:N	4:E:206:HOH:O	2.44	0.49
2:A:213:ILE:HG22	2:A:213:ILE:O	2.13	0.49
1:E:106:ARG:NH1	4:E:207:HOH:O	2.45	0.49
1:E:24:CYS:SG	1:E:98[A]:CYS:HB3	2.52	0.48
1:E:108:ASP:OD1	2:A:213:ILE:HG22	2.14	0.47
2:A:242:ILE:HD13	2:A:267:PHE:CD1	2.49	0.47
3:C:33:ALA:HA	3:C:102:THR:HG22	1.96	0.47
1:E:110:SER:CB	2:A:213:ILE:HG21	2.42	0.47
2:A:286:ARG:O	2:A:290:LEU:HD13	2.15	0.47
2:A:322:ILE:HA	2:A:325:VAL:HG22	1.98	0.43
3:C:53:LEU:HD22	3:C:54:TRP:N	2.33	0.43
1:E:30:THR:HG21	2:A:231:VAL:HG12	2.01	0.42
1:E:55:TRP:CE3	2:A:226:GLU:HG2	2.55	0.42
3:C:53:LEU:HD23	3:C:60:THR:CG2	2.50	0.41
2:A:296:GLU:O	2:A:296:GLU:HG3	2.20	0.41
1:E:91:GLU:OE1	1:E:91:GLU:N	2.47	0.41
2:A:213:ILE:CG2	2:A:213:ILE:O	2.69	0.41
2:A:313:TYR:CD1	3:C:101:ARG:HD2	2.56	0.41
2:A:295:GLY:HA3	4:A:402:HOH:O	2.20	0.41
3:C:43:THR:HG23	3:C:44:GLY:N	2.36	0.40
3:C:53:LEU:HD23	3:C:60:THR:HG22	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	E	126/128 (98%)	123 (98%)	3 (2%)	0	100	100
2	A	118/181 (65%)	117 (99%)	1 (1%)	0	100	100
3	C	118/120 (98%)	115 (98%)	3 (2%)	0	100	100
All	All	362/429 (84%)	355 (98%)	7 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	E	102/102 (100%)	100 (98%)	2 (2%)	55	42
2	A	110/167 (66%)	109 (99%)	1 (1%)	78	75
3	C	98/97 (101%)	93 (95%)	5 (5%)	24	9
All	All	310/366 (85%)	302 (97%)	8 (3%)	46	32

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	106	ARG
1	E	123	GLN
2	A	328	THR
3	C	34	LEU
3	C	49	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	C	50	VAL
3	C	76	ASN
3	C	114	THR

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

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 monosaccharides 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, 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	E	125/128 (97%)	0.21	0	100	100	26, 36, 52, 65	0
2	A	119/181 (65%)	0.32	3 (2%)	57	64	26, 39, 56, 68	0
3	C	118/120 (98%)	0.76	15 (12%)	3	5	33, 51, 81, 98	0
All	All	362/429 (84%)	0.43	18 (4%)	28	36	26, 41, 68, 98	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	29	ARG	4.2
3	C	3	GLU	4.1
2	A	212	GLN	3.8
3	C	7	VAL	3.6
3	C	31	SER	3.6
3	C	55	TRP	3.4
3	C	5	GLN	3.1
3	C	26	ALA	2.9
3	C	4	VAL	2.9
3	C	80	ILE	2.6
3	C	75	ASP	2.6
3	C	43	THR	2.4
3	C	34	LEU	2.3
3	C	30	THR	2.2
3	C	6[A]	LEU	2.1
2	A	213	ILE	2.1
3	C	25	ALA	2.1
2	A	294	GLN	2.0

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

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

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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.