



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 11:18 am BST

PDB ID : 5CTD  
Title : Crystal structure of the type IX collagen NC2 hetero-trimerization domain with a guest fragment a2a1a1 of type I collagen  
Authors : Boudko, S.P.; Bachinger, H.P.  
Deposited on : 2015-07-23  
Resolution : 1.60 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](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 following versions of software and data (see [references ⓘ](#)) 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

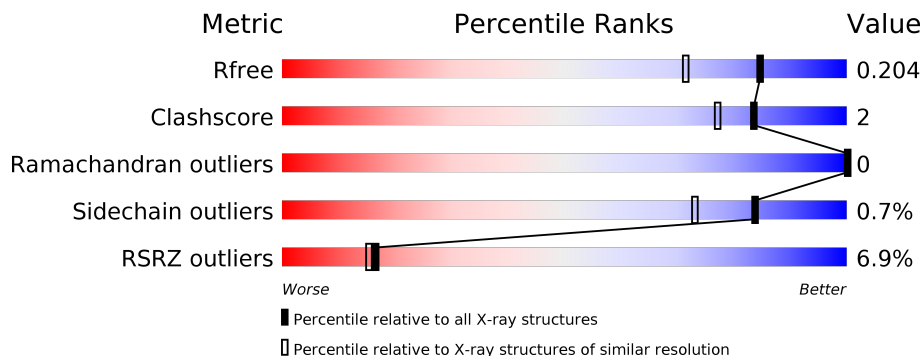
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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  $\geq 3$ , 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	71	
2	B	71	
3	C	72	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3061 atoms, of which 1429 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 Collagen alpha-1(I) chain, Collagen alpha-1(IX) chain.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S	Se			
1	A	64	898	284	447	86	77	1	3	0	1	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P02452
A	2	SER	-	expression tag	UNP P02452
A	3	GLY	-	expression tag	UNP P02452
A	4	PRO	-	expression tag	UNP P02452
A	5	PRO	-	expression tag	UNP P02452
A	6	GLY	-	expression tag	UNP P02452
A	7	PRO	-	expression tag	UNP P02452
A	8	PRO	-	expression tag	UNP P02452
A	9	GLY	-	expression tag	UNP P02452
A	10	PRO	-	expression tag	UNP P02452
A	11	PRO	-	expression tag	UNP P02452
A	12	GLY	-	expression tag	UNP P02452
A	13	PRO	-	expression tag	UNP P02452
A	14	PRO	-	expression tag	UNP P02452
A	27	GLY	-	linker	UNP P02452
A	28	PRO	-	linker	UNP P02452
A	29	PRO	-	linker	UNP P02452
A	30	GLY	-	linker	UNP P02452
A	31	PRO	-	linker	UNP P02452
A	32	PRO	-	linker	UNP P02452
A	33	GLY	-	linker	UNP P02452
A	34	PRO	-	linker	UNP P02452
A	35	PRO	-	linker	UNP P02452

- Molecule 2 is a protein called Collagen alpha-2(I) chain, Collagen alpha-2(IX) chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	71	993	313	496	89	94	1	0	2	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	GLY	-	expression tag	UNP P08123
B	2	SER	-	expression tag	UNP P08123
B	3	GLY	-	expression tag	UNP P08123
B	4	PRO	-	expression tag	UNP P08123
B	5	PRO	-	expression tag	UNP P08123
B	6	GLY	-	expression tag	UNP P08123
B	7	PRO	-	expression tag	UNP P08123
B	8	PRO	-	expression tag	UNP P08123
B	9	GLY	-	expression tag	UNP P08123
B	10	PRO	-	expression tag	UNP P08123
B	11	PRO	-	expression tag	UNP P08123
B	12	GLY	-	expression tag	UNP P08123
B	13	PRO	-	expression tag	UNP P08123
B	14	PRO	-	expression tag	UNP P08123
B	27	GLY	-	linker	UNP P08123
B	28	PRO	-	linker	UNP P08123
B	29	PRO	-	linker	UNP P08123
B	30	GLY	-	linker	UNP P08123
B	31	PRO	-	linker	UNP P08123
B	32	PRO	-	linker	UNP P08123
B	33	GLY	-	linker	UNP P08123
B	34	PRO	-	linker	UNP P08123
B	35	PRO	-	linker	UNP P08123

- Molecule 3 is a protein called Collagen alpha-1(I) chain, Collagen alpha-3(IX) chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	C	72	976	306	486	91	90	3	0	1	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	GLY	-	expression tag	UNP P02452
C	2	SER	-	expression tag	UNP P02452
C	3	GLY	-	expression tag	UNP P02452
C	4	PRO	-	expression tag	UNP P02452

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	5	PRO	-	expression tag	UNP P02452
C	6	GLY	-	expression tag	UNP P02452
C	7	PRO	-	expression tag	UNP P02452
C	8	PRO	-	expression tag	UNP P02452
C	9	GLY	-	expression tag	UNP P02452
C	10	PRO	-	expression tag	UNP P02452
C	11	PRO	-	expression tag	UNP P02452
C	12	GLY	-	expression tag	UNP P02452
C	13	PRO	-	expression tag	UNP P02452
C	14	PRO	-	expression tag	UNP P02452
C	27	GLY	-	linker	UNP P02452
C	28	PRO	-	linker	UNP P02452
C	29	PRO	-	linker	UNP P02452
C	30	GLY	-	linker	UNP P02452
C	31	PRO	-	linker	UNP P02452
C	32	PRO	-	linker	UNP P02452
C	33	GLY	-	linker	UNP P02452
C	34	PRO	-	linker	UNP P02452
C	35	PRO	-	linker	UNP P02452

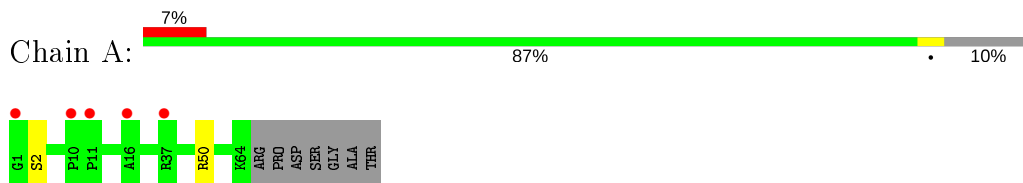
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	48	Total O 48 48	0	0
4	B	66	Total O 66 66	0	0
4	C	80	Total O 80 80	0	0

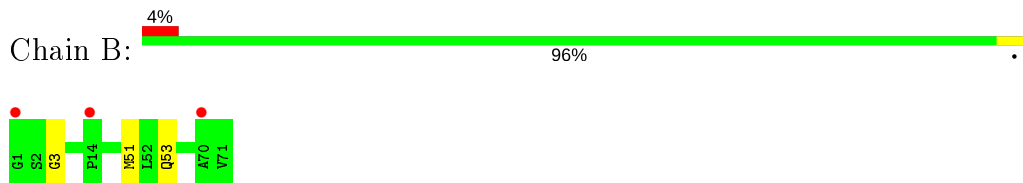
### 3 Residue-property plots [i](#)

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.

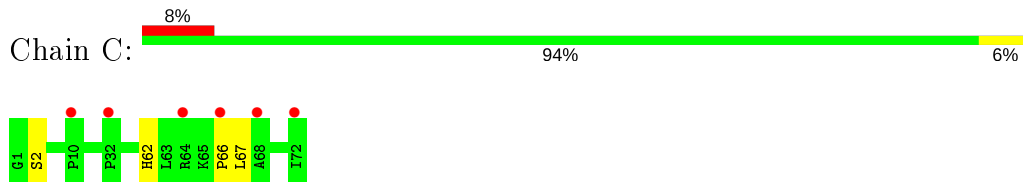
- Molecule 1: Collagen alpha-1(I) chain, Collagen alpha-1(IX) chain



- Molecule 2: Collagen alpha-2(I) chain, Collagen alpha-2(IX) chain



- Molecule 3: Collagen alpha-1(I) chain, Collagen alpha-3(IX) chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	27.89Å 55.92Å 62.10Å 90.00° 92.86° 90.00°	Depositor
Resolution (Å)	41.53 – 1.60 41.53 – 1.60	Depositor EDS
% Data completeness (in resolution range)	94.4 (41.53-1.60) 94.2 (41.53-1.60)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.27 (at 1.60Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.177 , 0.203 0.183 , 0.204	Depositor DCC
$R_{free}$ test set	2002 reflections (8.21%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.3	Xtriage
Anisotropy	0.642	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.46 , 57.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.048 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3061	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.56% 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	0.58	0/470	0.65	0/637
2	B	0.47	0/520	0.63	1/713 (0.1%)
3	C	0.47	0/511	0.58	0/697
All	All	0.51	0/1501	0.62	1/2047 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	51	MET	CG-SD-CE	-5.41	91.55	100.20

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	451	447	449	2	0
2	B	497	496	499	3	0
3	C	490	486	487	4	0
4	A	48	0	0	2	3
4	B	66	0	0	1	3
4	C	80	0	0	2	6
All	All	1632	1429	1435	7	6

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 2.

The worst 5 of 7 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)
3:C:66:PRO:O	4:C:101:HOH:O	1.90	0.89
3:C:67:LEU:HA	4:C:101:HOH:O	1.94	0.67
1:A:2:SER:C	4:A:101:HOH:O	2.48	0.51
1:A:50[A]:ARG:NH2	4:A:103:HOH:O	2.40	0.49
2:B:53[B]:GLN:NE2	4:B:104:HOH:O	2.47	0.46

The worst 5 of 6 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:136:HOH:O	4:C:161:HOH:O 2_655	1.82	0.38
4:A:130:HOH:O	4:C:164:HOH:O 1_554	1.87	0.33
4:A:109:HOH:O	4:C:149:HOH:O 1_656	1.87	0.33
4:B:152:HOH:O	4:C:154:HOH:O 2_654	1.92	0.28
4:B:150:HOH:O	4:C:145:HOH:O 2_654	1.95	0.25

## 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	63/71 (89%)	62 (98%)	1 (2%)	0	100	100
2	B	71/71 (100%)	70 (99%)	1 (1%)	0	100	100
3	C	71/72 (99%)	70 (99%)	1 (1%)	0	100	100
All	All	205/214 (96%)	202 (98%)	3 (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	46/47 (98%)	46 (100%)	0	100	100
2	B	51/49 (104%)	51 (100%)	0	100	100
3	C	49/49 (100%)	48 (98%)	1 (2%)	55	31
All	All	146/145 (101%)	145 (99%)	1 (1%)	84	73

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	62	HIS

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

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 [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, 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	61/71 (85%)	0.42	5 (8%) 11 10	11, 20, 38, 54	0
2	B	71/71 (100%)	0.37	3 (4%) 36 33	9, 23, 38, 44	0
3	C	72/72 (100%)	0.45	6 (8%) 11 10	11, 23, 40, 47	0
All	All	204/214 (95%)	0.41	14 (6%) 16 15	9, 23, 40, 54	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1	GLY	4.1
1	A	37	ARG	3.2
3	C	66	PRO	2.8
3	C	64	ARG	2.7
3	C	32	PRO	2.7

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.