

# Full wwPDB X-ray Structure Validation Report (i)

#### Nov 22, 2023 – 09:14 PM JST

PDB ID 7WWR

> Title : Structure of a triple-helix region of human collagen type III from Trautec

Authors Qian, S.; Li, H.; Fan, X.; Tian, X.; Li, J.; Wang, L.; Chu, Y.

2022-02-14 Deposited on

1.30 Å(reported) Resolution

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 (i) 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 (1)) were used in the production of this report:

MolProbity 4.02b-467

> 1.8.5 (274361), CSD as541be (2020) Mogul

Xtriage (Phenix) 1.13

EDS 2.36

20191225.v01 (using entries in the PDB archive December 25th 2019) Percentile statistics

> Refmac 5.8.0158

CCP4 7.0.044 (Gargrove)

Ideal geometry (proteins) Engh & Huber (2001) Ideal geometry (DNA, RNA) Parkinson et al. (1996)

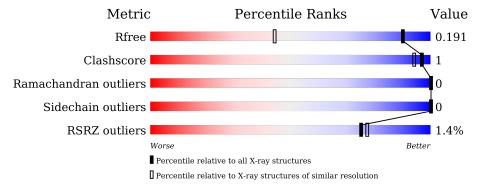
Validation Pipeline (wwPDB-VP) 2.36

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY\ DIFFRACTION$ 

The reported resolution of this entry is 1.30 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{ resolution range}( ext{Å}))$
$R_{free}$	130704	1058 (1.30-1.30)
Clashscore	141614	1101 (1.30-1.30)
Ramachandran outliers	138981	1058 (1.30-1.30)
Sidechain outliers	138945	1058 (1.30-1.30)
RSRZ outliers	127900	1029 (1.30-1.30)

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 >=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	30	80%	20%
1	В	30	90%	10%
1	С	30	97%	•



# 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 780 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 Collagen alpha-1(III) chain.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace			
1	Λ	30	Total C	N	О	0	1	0		
1	Α	30	200   122	35	43	U	1	U		
1	D	30	Total C	N	О	0	1	0		
1	Б	30	197   120	35	42	U	1	U		
1	С	30	Total C	N	О	0	0	0		
1			C	30	194 118	35	41	0	U	

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	HYP	GLN	conflict	UNP P02461
A	4	PRO	VAL	conflict	UNP P02461
A	5	HYP	LYS	conflict	UNP P02461
A	7	PRO	GLU	conflict	UNP P02461
A	8	HYP	SER	conflict	UNP P02461
A	26	HYP	GLN	conflict	UNP P02461
A	28	PRO	LEU	conflict	UNP P02461
В	2	HYP	GLN	conflict	UNP P02461
В	4	PRO	VAL	conflict	UNP P02461
В	5	HYP	LYS	conflict	UNP P02461
В	7	PRO	GLU	conflict	UNP P02461
В	8	HYP	SER	conflict	UNP P02461
В	26	HYP	GLN	conflict	UNP P02461
В	28	PRO	LEU	conflict	UNP P02461
С	2	HYP	GLN	conflict	UNP P02461
С	4	PRO	VAL	conflict	UNP P02461
С	5	HYP	LYS	conflict	UNP P02461
С	7	PRO	GLU	conflict	UNP P02461
С	8	HYP	SER	conflict	UNP P02461
С	26	HYP	GLN	conflict	UNP P02461
С	28	PRO	LEU	conflict	UNP P02461

• Molecule 2 is water.



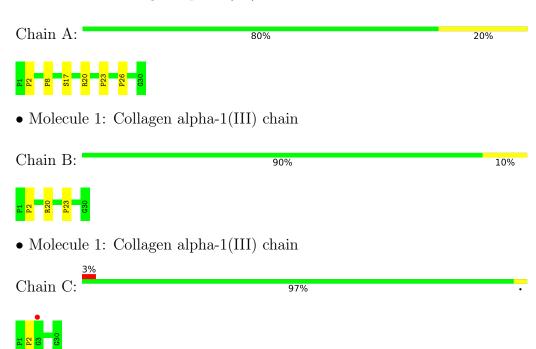
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	62	Total O 62 62	0	0
2	В	66	Total O 66 66	0	0
2	С	61	Total O 61 61	0	0



# 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: Collagen alpha-1(III) chain





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	23.81Å 21.32Å 57.64Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $93.65^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	23.77 - 1.30	Depositor
resolution (A)	23.76 - 1.30	EDS
% Data completeness	99.5 (23.77-1.30)	Depositor
(in resolution range)	99.6 (23.76-1.30)	EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	20.47 (at 1.30Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
$R, R_{free}$	0.137 , 0.194	Depositor
it, it free	0.143 , $0.191$	DCC
$R_{free}$ test set	764 reflections $(5.27\%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	5.8	Xtriage
Anisotropy	1.120	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.17, 32.1	EDS
L-test for twinning <sup>2</sup>	$ < L > = 0.46, < L^2> = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	780	wwPDB-VP
Average B, all atoms $(Å^2)$	10.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The analyses of the Patterson function reveals a significant off-origin peak that is 35.66 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.6349e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 5 Model quality (i)

### 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: HYP

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	
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	A	0.88	0/158	1.08	1/202~(0.5%)
1	В	0.95	0/152	1.02	1/194~(0.5%)
1	С	0.76	0/146	0.90	0/186
All	All	0.87	0/456	1.00	$2/582 \ (0.3\%)$

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	В	20	ARG	NE-CZ-NH1	5.63	123.11	120.30
1	A	20	ARG	NE-CZ-NH1	5.04	122.82	120.30

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	200	0	192	1	0
1	В	197	0	187	0	0
1	С	194	0	182	0	0
2	A	62	0	0	1	0
2	В	66	0	0	0	1
2	С	61	0	0	0	1

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	780	0	561	1	1

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

Atom-1	Atom-2	$egin{aligned} &  ext{Interatomic} \ &  ext{distance} \ &  ext{(Å)} \end{aligned}$	$\begin{array}{c} \text{Clash} \\ \text{overlap } (\text{\AA}) \end{array}$
1:A:17[B]:SER:OG	2:A:101:HOH:O	2.00	0.79

All (1) 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)	$egin{aligned}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{aligned}$
2:B:103:HOH:O	2:C:150:HOH:O[1_655]	1.93	0.27

### 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	Percei	Percentiles	
1	A	24/30 (80%)	24 (100%)	0	0	100	100	
1	В	23/30 (77%)	23 (100%)	0	0	100	100	
1	С	22/30 (73%)	22 (100%)	0	0	100	100	
All	All	69/90 (77%)	69 (100%)	0	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	15/13~(115%)	15 (100%)	0	100 100
1	В	14/13 (108%)	14 (100%)	0	100 100
1	С	13/13 (100%)	13 (100%)	0	100 100
All	All	42/39 (108%)	42 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes 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)

18 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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| > 2 is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Mol Type	Chain Res Link Bond lengths Bond				ond ang	gles			
Moi Type	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2 $ 1 (20%) $ $0$ $0$ $2 (40%)$
1	HYP	A	26	1	6,8,9	0.63	0	5,10,12	1.37	1 (20%)
1	HYP	С	29	1	6,8,9	0.82	0	5,10,12	1.02	0
1	HYP	В	29	1	6,8,9	0.62	0	5,10,12	0.86	0
1	HYP	С	2	1	6,8,9	0.80	0	5,10,12	1.69	2 (40%)
1	HYP	A	8	1	6,8,9	0.52	0	5,10,12	1.29	1 (20%)



Mol	Trino	Chain	Res	Link	В	ond leng	$\operatorname{gths}$	Е	ond ang	gles
MIOI	Type	Chain	nes	Lilik	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
1	HYP	В	23	1	6,8,9	0.46	0	5,10,12	1.68	2 (40%)
1	HYP	A	5	1	6,8,9	0.58	0	5,10,12	0.80	0
1	HYP	В	8	1	6,8,9	0.54	0	5,10,12	0.91	0
1	HYP	С	8	1	6,8,9	0.41	0	5,10,12	0.95	0
1	HYP	В	26	1	6,8,9	1.02	0	5,10,12	0.81	0
1	HYP	A	29	1	6,8,9	0.54	0	5,10,12	0.92	0
1	HYP	A	23	1	6,8,9	0.62	0	5,10,12	1.28	1 (20%)
1	HYP	В	2	1	6,8,9	0.88	0	5,10,12	1.16	1 (20%)
1	HYP	В	5	1	6,8,9	0.91	0	5,10,12	0.74	0
1	HYP	С	5	1	6,8,9	0.65	0	5,10,12	1.22	0
1	HYP	С	23	1	6,8,9	0.46	0	5,10,12	0.50	0
1	HYP	A	2	1	6,8,9	0.93	0	5,10,12	1.56	1 (20%)
1	HYP	С	26	1	6,8,9	0.40	0	5,10,12	0.97	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	HYP	A	26	1	-	0/0/11/13	0/1/1/1
1	HYP	С	29	1	-	0/0/11/13	0/1/1/1
1	HYP	В	29	1	-	0/0/11/13	0/1/1/1
1	HYP	С	2	1	-	0/0/11/13	0/1/1/1
1	HYP	A	8	1	-	0/0/11/13	0/1/1/1
1	HYP	В	23	1	-	0/0/11/13	0/1/1/1
1	HYP	A	5	1	-	0/0/11/13	0/1/1/1
1	HYP	В	8	1	-	0/0/11/13	0/1/1/1
1	HYP	С	8	1	-	0/0/11/13	0/1/1/1
1	HYP	В	26	1	-	0/0/11/13	0/1/1/1
1	HYP	A	29	1	-	0/0/11/13	0/1/1/1
1	HYP	A	23	1	-	0/0/11/13	0/1/1/1
1	HYP	В	2	1	-	0/0/11/13	0/1/1/1
1	HYP	В	5	1	-	0/0/11/13	0/1/1/1
1	HYP	С	5	1	-	0/0/11/13	0/1/1/1
1	HYP	С	23	1	-	0/0/11/13	0/1/1/1
1	HYP	A	2	1	-	0/0/11/13	0/1/1/1
1	HYP	С	26	1	-	0/0/11/13	0/1/1/1

There are no bond length outliers.

All (9) bond angle outliers are listed below:



Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$Ideal(^{o})$
1	A	2	HYP	OD1-CG-CD	3.09	117.09	110.35
1	A	26	HYP	OD1-CG-CD	2.52	115.86	110.35
1	С	2	HYP	O-C-CA	-2.39	118.53	124.78
1	В	23	HYP	CG-CB-CA	2.22	106.77	103.96
1	С	2	HYP	CB-CG-CD	-2.20	100.57	103.27
1	A	23	HYP	CB-CG-CD	-2.19	100.58	103.27
1	В	2	HYP	O-C-CA	-2.10	119.26	124.78
1	В	23	HYP	OD1-CG-CD	-2.09	105.78	110.35
1	A	8	HYP	OD1-CG-CB	-2.05	104.96	110.03

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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 (i)

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,  $95^{th}$  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 $>$	$\#\mathrm{RSRZ}{>}2$	$\mathbf{OWAB}(\mathbf{\mathring{A}}^2)$	Q < 0.9
1	A	24/30 (80%)	-0.05	0 100 100	3, 7, 13, 18	0
1	В	24/30 (80%)	-0.14	0 100 100	3, 6, 11, 20	0
1	С	24/30 (80%)	-0.16	1 (4%) 36 33	3, 5, 12, 15	0
All	All	72/90 (80%)	-0.12	1 (1%) 75 77	3, 6, 13, 20	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	С	3	GLY	2.4

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median,  $95^{th}$  percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\operatorname{B-factors}(\mathring{\mathbf{A}}^2)$	Q < 0.9
1	HYP	В	29	8/9	0.89	0.13	9,12,15,17	0
1	HYP	A	5	8/9	0.90	0.12	10,12,14,14	0
1	HYP	С	2	8/9	0.91	0.21	14,15,18,20	0
1	HYP	В	2	8/9	0.91	0.15	10,12,14,16	0
1	HYP	В	5	8/9	0.93	0.09	8,9,10,11	0
1	HYP	С	29	8/9	0.94	0.10	9,10,11,11	0
1	HYP	A	2	8/9	0.94	0.17	10,13,14,18	0
1	HYP	В	23	8/9	0.95	0.09	5,7,11,16	0
1	HYP	A	29	8/9	0.95	0.08	8,10,12,15	0
1	HYP	A	23	8/9	0.96	0.08	6,8,10,15	0
1	HYP	С	5	8/9	0.96	0.10	7,9,11,14	0
1	HYP	В	8	8/9	0.96	0.09	6,7,7,10	0

Continued on next page...



Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
1	HYP	С	26	8/9	0.96	0.08	6,7,7,7	0
1	HYP	A	26	8/9	0.97	0.07	8,12,16,18	0
1	HYP	В	26	8/9	0.97	0.06	7,7,8,9	0
1	HYP	A	8	8/9	0.97	0.10	7,7,11,12	0
1	HYP	С	8	8/9	0.98	0.06	6,6,7,7	0
1	HYP	С	23	8/9	0.98	0.06	3,4,4,5	0

### 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.

