



# Full wwPDB X-ray Structure Validation Report ⓘ

May 24, 2020 – 06:11 pm BST

PDB ID : 6FZW  
Title : Crystal structure of the metalloproteinase enhancer PCPE-1 bound to the procollagen C propeptide trimer (long)  
Authors : Hohenester, E.; Pulido, D.  
Deposited on : 2018-03-15  
Resolution : 2.78 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 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

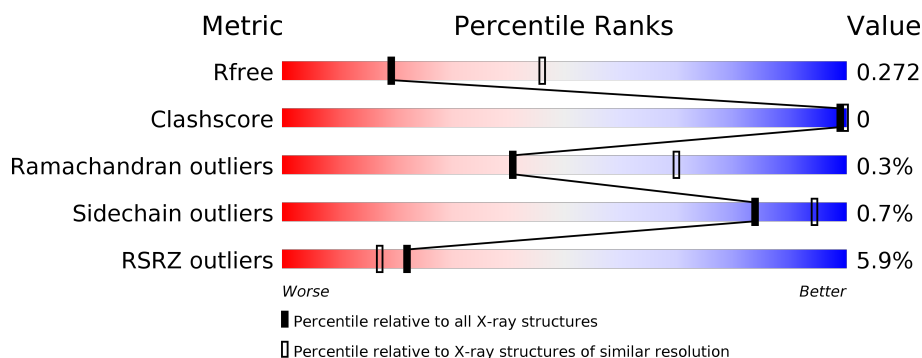
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.78 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	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (2.80-2.76)

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	293	<div> <div>3%</div> <div>80%</div> <div>19%</div> </div>
1	B	293	<div> <div>3%</div> <div>78%</div> <div>19%</div> </div>
1	C	293	<div> <div>2%</div> <div>79%</div> <div>21%</div> </div>
2	D	265	<div> <div>11%</div> <div>66%</div> <div>32%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13459 atoms, of which 6547 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	A	237	Total	C	H	N	O	S	0	0	0
			3645	1171	1789	319	353	13			
1	B	237	Total	C	H	N	O	S	0	0	0
			3630	1165	1779	318	355	13			
1	C	232	Total	C	H	N	O	S	0	0	0
			3550	1145	1734	310	348	13			

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-47	GLU	-	expression tag	UNP P02461
A	-46	THR	-	expression tag	UNP P02461
A	-45	GLY	-	expression tag	UNP P02461
A	-44	HIS	-	expression tag	UNP P02461
A	-43	HIS	-	expression tag	UNP P02461
A	-42	HIS	-	expression tag	UNP P02461
A	-41	HIS	-	expression tag	UNP P02461
A	-40	HIS	-	expression tag	UNP P02461
A	-39	HIS	-	expression tag	UNP P02461
A	-38	SER	-	expression tag	UNP P02461
A	-37	ALA	-	expression tag	UNP P02461
A	132	GLN	HIS	variant	UNP P02461
A	146	GLN	ASN	conflict	UNP P02461
B	-47	GLU	-	expression tag	UNP P02461
B	-46	THR	-	expression tag	UNP P02461
B	-45	GLY	-	expression tag	UNP P02461
B	-44	HIS	-	expression tag	UNP P02461
B	-43	HIS	-	expression tag	UNP P02461
B	-42	HIS	-	expression tag	UNP P02461
B	-41	HIS	-	expression tag	UNP P02461
B	-40	HIS	-	expression tag	UNP P02461
B	-39	HIS	-	expression tag	UNP P02461
B	-38	SER	-	expression tag	UNP P02461

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-37	ALA	-	expression tag	UNP P02461
B	132	GLN	HIS	variant	UNP P02461
B	146	GLN	ASN	conflict	UNP P02461
C	-47	GLU	-	expression tag	UNP P02461
C	-46	THR	-	expression tag	UNP P02461
C	-45	GLY	-	expression tag	UNP P02461
C	-44	HIS	-	expression tag	UNP P02461
C	-43	HIS	-	expression tag	UNP P02461
C	-42	HIS	-	expression tag	UNP P02461
C	-41	HIS	-	expression tag	UNP P02461
C	-40	HIS	-	expression tag	UNP P02461
C	-39	HIS	-	expression tag	UNP P02461
C	-38	SER	-	expression tag	UNP P02461
C	-37	ALA	-	expression tag	UNP P02461
C	132	GLN	HIS	variant	UNP P02461
C	146	GLN	ASN	conflict	UNP P02461

- Molecule 2 is a protein called Procollagen C-endopeptidase enhancer 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	180	Total	C	H	N	O	S	0	0	0
			2575	856	1230	219	262	8			

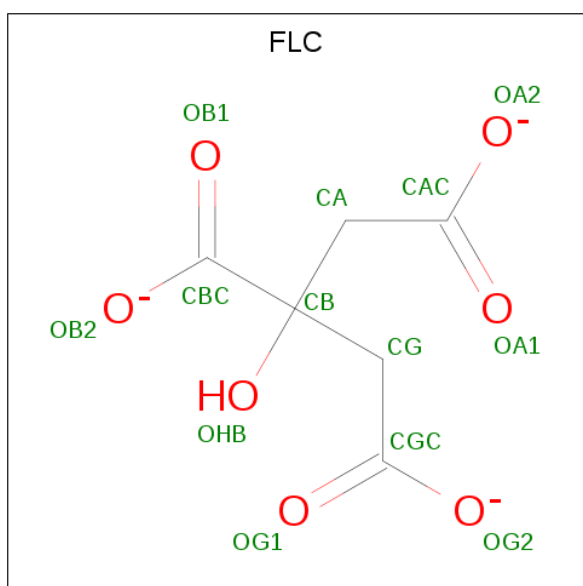
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	ALA	-	expression tag	UNP Q15113
D	-2	PRO	-	expression tag	UNP Q15113
D	-1	LEU	-	expression tag	UNP Q15113
D	0	ALA	-	expression tag	UNP Q15113
D	254	ALA	-	expression tag	UNP Q15113
D	255	ALA	-	expression tag	UNP Q15113
D	256	HIS	-	expression tag	UNP Q15113
D	257	HIS	-	expression tag	UNP Q15113
D	258	HIS	-	expression tag	UNP Q15113
D	259	HIS	-	expression tag	UNP Q15113
D	260	HIS	-	expression tag	UNP Q15113
D	261	HIS	-	expression tag	UNP Q15113

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Ca 1 1	0	0
3	A	1	Total Ca 1 1	0	0
3	D	2	Total Ca 2 2	0	0
3	C	1	Total Ca 1 1	0	0

- Molecule 4 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7$ ).

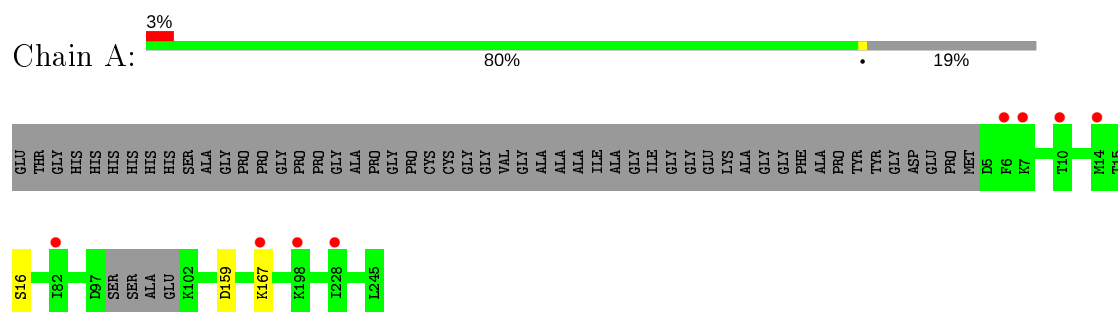


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C H O 18 6 5 7	0	0
4	B	1	Total C H O 18 6 5 7	0	0
4	C	1	Total C H O 18 6 5 7	0	0

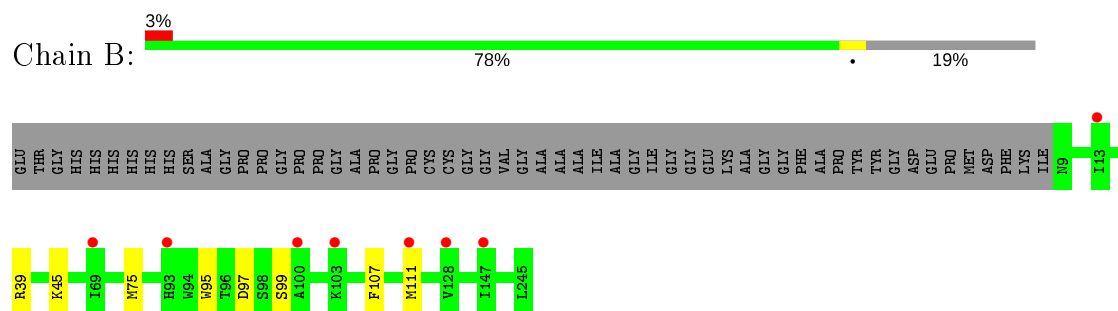
### 3 Residue-property plots [i](#)

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.

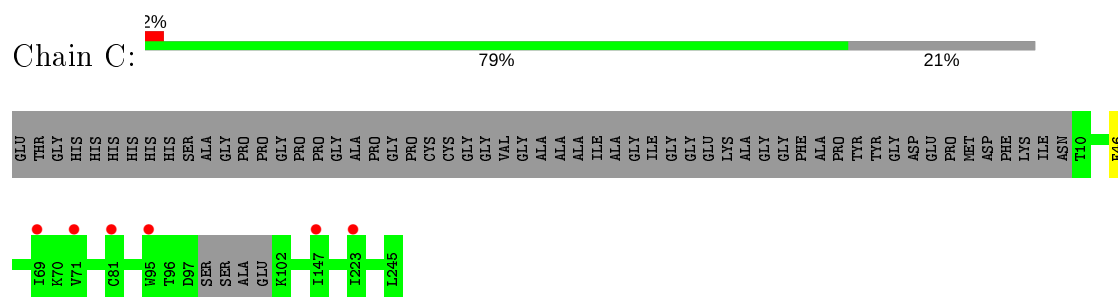
- Molecule 1: Collagen alpha-1(III) chain



- Molecule 1: Collagen alpha-1(III) chain

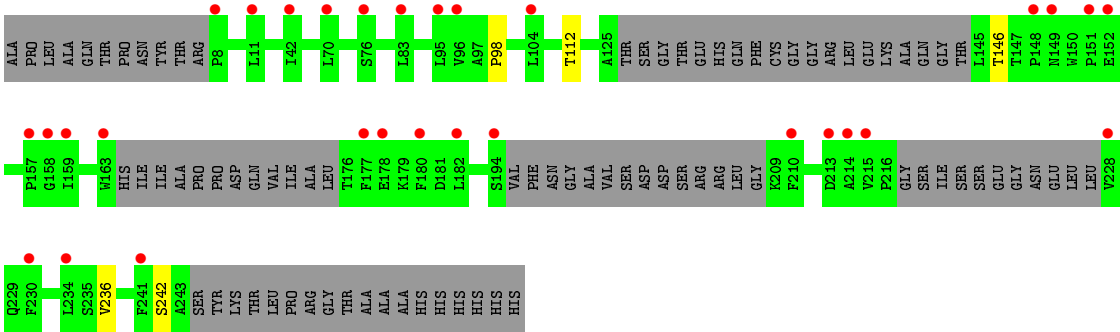


- Molecule 1: Collagen alpha-1(III) chain



- Molecule 2: Procollagen C-endopeptidase enhancer 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.88Å 143.65Å 156.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.79 – 2.78 68.79 – 2.78	Depositor EDS
% Data completeness (in resolution range)	90.5 (68.79-2.78) 90.6 (68.79-2.78)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.67 (at 2.77Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.247 , 0.271 0.252 , 0.272	Depositor DCC
$R_{free}$ test set	2252 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	81.2	Xtriage
Anisotropy	0.471	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 48.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13459	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% 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](#)

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

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	0.24	0/1899	0.38	0/2565
1	B	0.26	0/1895	0.40	0/2562
1	C	0.25	0/1859	0.40	0/2513
2	D	0.23	0/1382	0.38	0/1883
All	All	0.24	0/7035	0.39	0/9523

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	1856	1789	1788	1	0
1	B	1851	1779	1778	4	0
1	C	1816	1734	1733	0	0
2	D	1345	1230	1226	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	2	0	0	0	0
4	A	13	5	5	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	13	5	5	0	0
4	C	13	5	5	0	0
All	All	6912	6547	6540	5	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 0.

All (5) 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:159:ASP:OD2	1:A:167:LYS:NZ	2.40	0.55
1:B:95:TRP:NE1	1:B:97:ASP:OD1	2.43	0.52
1:B:45:LYS:NZ	1:B:75:MET:O	2.43	0.52
1:B:97:ASP:OD2	1:B:99:SER:OG	2.30	0.49
1:B:107:PHE:HA	1:B:111:MET:SD	2.60	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	233/293 (80%)	220 (94%)	13 (6%)	0	100	100
1	B	235/293 (80%)	223 (95%)	12 (5%)	0	100	100
1	C	228/293 (78%)	210 (92%)	18 (8%)	0	100	100
2	D	170/265 (64%)	153 (90%)	14 (8%)	3 (2%)	8	25
All	All	866/1144 (76%)	806 (93%)	57 (7%)	3 (0%)	41	70

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	112	THR
2	D	236	VAL
2	D	98	PRO

### 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	A	203/240 (85%)	202 (100%)	1 (0%)	88	95
1	B	203/240 (85%)	202 (100%)	1 (0%)	88	95
1	C	198/240 (82%)	197 (100%)	1 (0%)	88	95
2	D	142/217 (65%)	140 (99%)	2 (1%)	67	87
All	All	746/937 (80%)	741 (99%)	5 (1%)	84	94

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	16	SER
1	B	39	ARG
1	C	46	PHE
2	D	146	THR
2	D	242	SER

Some 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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 8 ligands modelled in this entry, 5 are monoatomic - leaving 3 for Mogul analysis.

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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	FLC	B	302	-	3,12,12	1.55	0	3,17,17	1.89	2 (66%)
4	FLC	A	302	-	3,12,12	1.85	1 (33%)	3,17,17	2.57	2 (66%)
4	FLC	C	302	-	3,12,12	1.69	0	3,17,17	2.07	2 (66%)

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
4	FLC	B	302	-	-	6/6/16/16	-
4	FLC	A	302	-	-	3/6/16/16	-
4	FLC	C	302	-	-	3/6/16/16	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	302	FLC	CG-CB	-2.24	1.51	1.54

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	302	FLC	CB-CG-CGC	-3.82	108.86	114.98
4	C	302	FLC	CB-CG-CGC	-2.88	110.37	114.98
4	B	302	FLC	CB-CA-CAC	-2.46	111.05	114.98

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	302	FLC	CB-CA-CAC	-2.27	111.35	114.98
4	C	302	FLC	CB-CA-CAC	-2.09	111.64	114.98
4	B	302	FLC	CB-CG-CGC	-2.07	111.67	114.98

There are no chirality outliers.

All (12) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	302	FLC	CAC-CA-CB-CBC
4	B	302	FLC	CA-CB-CG-CGC
4	B	302	FLC	CBC-CB-CG-CGC
4	B	302	FLC	OHB-CB-CG-CGC
4	A	302	FLC	CAC-CA-CB-CBC
4	A	302	FLC	CAC-CA-CB-CG
4	A	302	FLC	CAC-CA-CB-OHB
4	C	302	FLC	CBC-CB-CG-CGC
4	B	302	FLC	CAC-CA-CB-OHB
4	C	302	FLC	OHB-CB-CG-CGC
4	B	302	FLC	CAC-CA-CB-CG
4	C	302	FLC	CA-CB-CG-CGC

There are no ring outliers.

No monomer is involved in short contacts.

## 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 ⓘ

### 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	237/293 (80%)	0.61	8 (3%)	45	39	62, 86, 123, 159	0
1	B	237/293 (80%)	0.79	8 (3%)	45	39	61, 75, 127, 165	0
1	C	232/293 (79%)	0.56	6 (2%)	56	51	64, 82, 124, 147	0
2	D	180/265 (67%)	0.85	30 (16%)	1	1	76, 110, 157, 184	0
All	All	886/1144 (77%)	0.69	52 (5%)	22	17	61, 87, 137, 184	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	180	PHE	4.9
2	D	8	PRO	4.9
2	D	177	PHE	4.5
2	D	215	VAL	4.3
2	D	228	VAL	3.5
1	B	103	LYS	3.4
2	D	96	VAL	3.4
2	D	163	TRP	3.3
2	D	148	PRO	3.3
2	D	152	GLU	3.3
1	B	93	HIS	3.3
2	D	234	LEU	3.2
2	D	182	LEU	3.2
2	D	230	PHE	3.1
1	A	6	PHE	3.0
1	A	198	LYS	3.0
1	B	13	ILE	3.0
2	D	83	LEU	3.0
2	D	241	PHE	2.9
1	A	167	LYS	2.9
2	D	149	ASN	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	214	ALA	2.9
1	B	100	ALA	2.8
2	D	95	LEU	2.8
1	A	7	LYS	2.8
2	D	159	ILE	2.8
1	C	223	ILE	2.7
2	D	178	GLU	2.6
1	B	147	ILE	2.6
1	C	147	ILE	2.5
2	D	158	GLY	2.5
1	A	14	MET	2.5
2	D	70	LEU	2.5
2	D	194	SER	2.5
1	B	69	ILE	2.4
2	D	213	ASP	2.4
2	D	42	ILE	2.4
2	D	157	PRO	2.4
2	D	76	SER	2.4
1	A	228	ILE	2.4
1	C	71	VAL	2.3
1	C	81	CYS	2.3
2	D	104	LEU	2.3
1	B	111	MET	2.2
1	A	10	THR	2.2
1	B	128	VAL	2.2
2	D	11	LEU	2.2
1	C	69	ILE	2.1
1	C	95	TRP	2.1
1	A	82	ILE	2.1
2	D	151	PRO	2.1
2	D	210	PHE	2.1

## 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 carbohydrates in this entry.

## 6.4 Ligands

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<sup>th</sup> 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	B-factors(Å <sup>2</sup> )	Q<0.9
4	FLC	C	302	13/13	0.82	0.23	101,102,123,123	0
4	FLC	B	302	13/13	0.84	0.17	94,95,115,115	0
4	FLC	A	302	13/13	0.89	0.16	102,104,125,125	0
3	CA	D	301	1/1	0.94	0.15	79,79,79,79	0
3	CA	B	301	1/1	0.97	0.28	63,63,63,63	0
3	CA	C	301	1/1	0.97	0.22	71,71,71,71	0
3	CA	A	301	1/1	0.97	0.30	62,62,62,62	0
3	CA	D	302	1/1	0.99	0.12	89,89,89,89	0

## 6.5 Other polymers

There are no such residues in this entry.