



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 10:34 pm BST

PDB ID : 2UY1  
Title : CRYSTAL STRUCTURE OF CSTF-77  
Authors : Legrand, P.; Pinaud, N.; Minvielle-Sebastia, L.; Fribourg, S.  
Deposited on : 2007-04-02  
Resolution : 2.00 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  
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

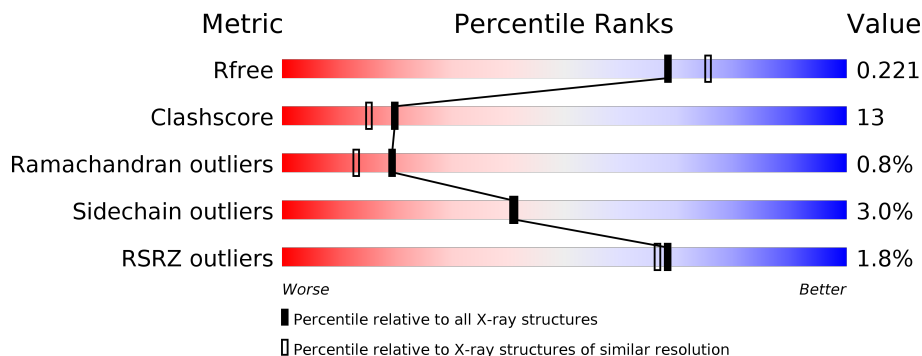
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	493	
2	B	493	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7223 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 CLEAVAGE STIMULATION FACTOR 77.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	438	3668	2374	601	672	21	0	2	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	GLU	LEU	conflict	UNP Q8SWC5

- Molecule 2 is a protein called CLEAVAGE STIMULATION FACTOR 77.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	384	3235	2112	530	574	19	0	1	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	42	TRP	LYS	conflict	UNP Q8SWC5
B	68	PHE	TYR	conflict	UNP Q8SWC5
B	88	PHE	TYR	conflict	UNP Q8SWC5
B	152	LEU	ILE	conflict	UNP Q8SWC5
B	210	PHE	TYR	conflict	UNP Q8SWC5

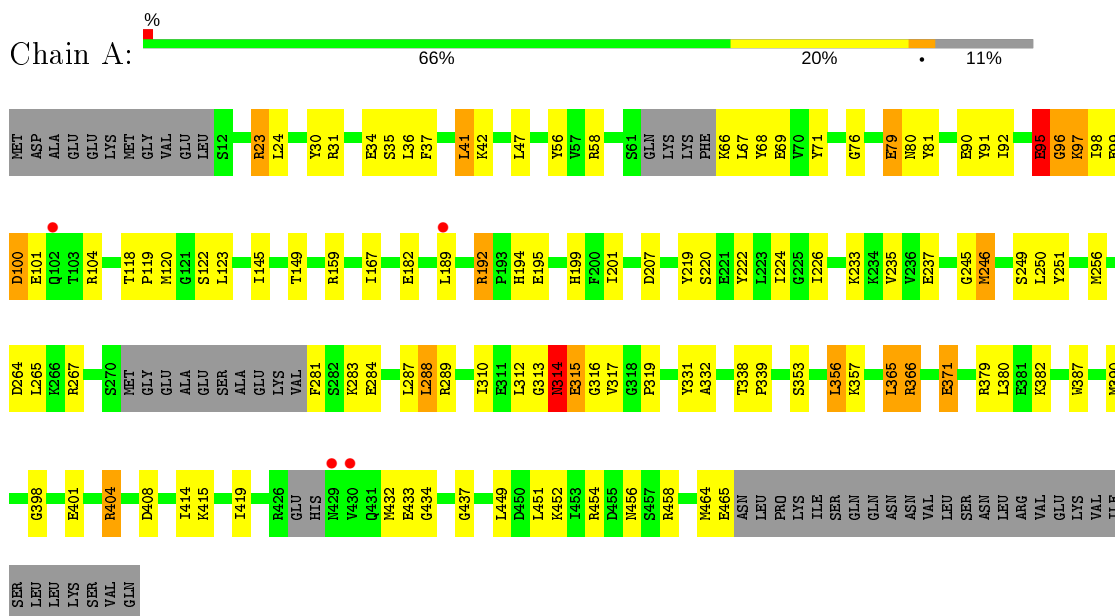
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	221	Total	O	0	0
			221	221		
3	B	99	Total	O	0	0
			99	99		

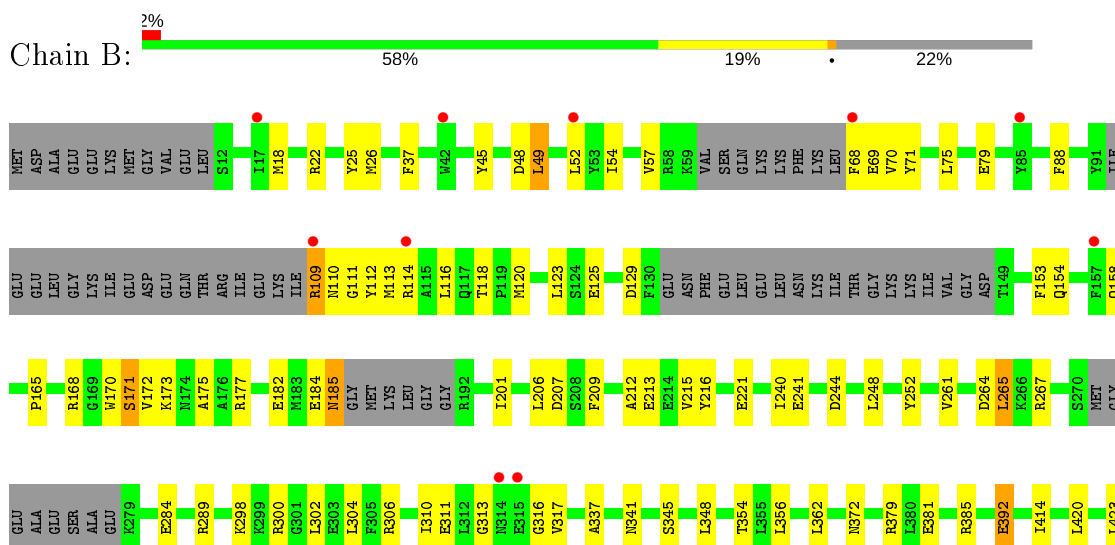
### 3 Residue-property plots [i](#)

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: CLEAVAGE STIMULATION FACTOR 77



- Molecule 2: CLEAVAGE STIMULATION FACTOR 77



E424	
P425	
R426	
GLU	
HIS	
ASN	
V430	
E433	
G434	
I435	
R438	
R454	
D455	
ASN	
SER	
ARG	
LEU	
LEU	
ASP	
GLU	
PHE	
MET	
GLU	
ASN	
LEU	
PRO	
LYS	
ILE	
SER	
GLN	
GLN	
ASN	
ASN	
VAL	
LEU	
SER	
ASN	
LEU	
LEU	
ARG	
VAL	
GLU	
LYS	
VAL	
ILE	
SER	
LEU	
LEU	
LYS	
SER	
VAL	
GLN	

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.44Å 148.28Å 91.44Å 90.00° 92.15° 90.00°	Depositor
Resolution (Å)	91.29 – 2.00 29.84 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.7 (91.29-2.00) 99.4 (29.84-2.00)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.225 , 0.280 0.227 , 0.221	Depositor DCC
$R_{free}$ test set	3954 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.8	Xtrriage
Anisotropy	0.238	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 30.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	0.075 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	7223	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.43% 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	1.16	7/3757 (0.2%)	0.99	11/5041 (0.2%)
2	B	0.98	3/3319 (0.1%)	0.91	3/4455 (0.1%)
All	All	1.08	10/7076 (0.1%)	0.95	14/9496 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	8
2	B	0	4
All	All	0	12

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	251	TYR	C-N	-13.68	1.02	1.34
1	A	331	TYR	CD2-CE2	6.55	1.49	1.39
1	A	95[A]	GLU	CA-C	6.08	1.68	1.52
1	A	95[B]	GLU	CA-C	6.08	1.68	1.52
2	B	372	ASN	CG-OD1	5.93	1.36	1.24

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	356	LEU	CA-CB-CG	9.93	138.13	115.30
1	A	404	ARG	NE-CZ-NH2	-8.03	116.28	120.30
1	A	95[A]	GLU	CA-C-O	-6.97	105.46	120.10
1	A	95[B]	GLU	CA-C-O	-6.97	105.46	120.10
1	A	404	ARG	NE-CZ-NH1	6.76	123.68	120.30

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	312	LEU	Peptide
1	A	95[A]	GLU	Mainchain,Peptide
1	A	95[B]	GLU	Mainchain,Peptide
1	A	96	GLY	Peptide
1	A	97	LYS	Peptide

## 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	3668	0	3621	83	0
2	B	3235	0	3180	98	0
3	A	221	0	0	12	0
3	B	99	0	0	11	0
All	All	7223	0	6801	173	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 13.

The worst 5 of 173 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)
2:B:109:ARG:HA	2:B:111:GLY:N	1.37	1.37
2:B:109:ARG:CA	2:B:111:GLY:H	1.37	1.35
2:B:45:TYR:HB2	2:B:207:ASP:OD2	1.42	1.19
1:A:192:ARG:HH11	1:A:192:ARG:HG2	1.09	1.12
2:B:171:SER:HB3	2:B:175:ALA:H	1.08	1.09

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	432/493 (88%)	412 (95%)	17 (4%)	3 (1%)	22	16
2	B	371/493 (75%)	358 (96%)	10 (3%)	3 (1%)	19	13
All	All	803/986 (81%)	770 (96%)	27 (3%)	6 (1%)	19	16

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	100	ASP
1	A	314	ASN
2	B	49	LEU
2	B	110	ASN
2	B	316	GLY

### 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	395/444 (89%)	379 (96%)	16 (4%)	30	28
2	B	346/444 (78%)	340 (98%)	6 (2%)	60	65
All	All	741/888 (83%)	719 (97%)	22 (3%)	41	41

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	315	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	371	GLU
2	B	244	ASP
1	A	365	LEU
1	A	366	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	409	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 carbohydrates 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	251:TYR	C	252:TYR	N	1.02

## 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	438/493 (88%)	-0.19	4 (0%) 84 83	11, 22, 44, 58	0
2	B	384/493 (77%)	0.01	11 (2%) 51 50	12, 31, 52, 61	0
All	All	822/986 (83%)	-0.10	15 (1%) 68 66	11, 26, 50, 61	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	430	VAL	5.5
2	B	42	TRP	5.3
2	B	315	GLU	3.6
2	B	68	PHE	3.3
1	A	430	VAL	3.2

### 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 [i](#)

There are no such residues in this entry.