



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 08:49 PM GMT

PDB ID : 2OTO  
Title : N-terminal fragment of Streptococcus pyogenes M1 protein  
Authors : McNamara, C.W.; Ghosh, P.  
Deposited on : 2007-02-08  
Resolution : 3.04 Å(reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

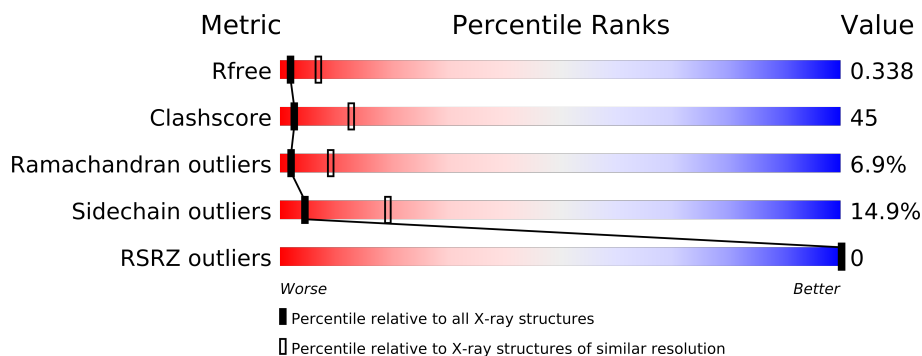
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.04 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))
$R_{free}$	66092	1460 (3.08-3.00)
Clashscore	79885	1881 (3.08-3.00)
Ramachandran outliers	78287	1814 (3.08-3.00)
Sidechain outliers	78261	1817 (3.08-3.00)
RSRZ outliers	66119	1462 (3.08-3.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	155	
1	B	155	
1	C	155	
1	D	155	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4521 atoms, of which 0 are hydrogen and 0 are deuterium.

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 M protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	137	Total	C	N	O	Se	0	0	0
			1131	691	208	231	1			
1	B	138	Total	C	N	O	Se	0	0	0
			1136	694	209	232	1			
1	C	138	Total	C	N	O	Se	0	0	0
			1136	694	209	232	1			
1	D	135	Total	C	N	O	Se	0	0	0
			1118	684	205	228	1			

There are 12 discrepancies between the modelled and reference sequences:

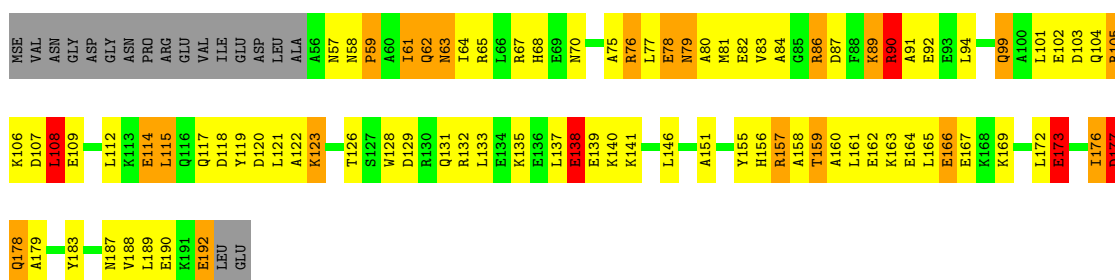
Chain	Residue	Modelled	Actual	Comment	Reference
A	40	MSE	-	SEE REMARK 999	UNP Q48WD8
A	41	VAL	-	EXPRESSION TAG	UNP Q48WD8
A	81	MSE	MET	MODIFIED RESIDUE	UNP Q48WD8
B	40	MSE	-	SEE REMARK 999	UNP Q48WD8
B	41	VAL	-	EXPRESSION TAG	UNP Q48WD8
B	81	MSE	MET	MODIFIED RESIDUE	UNP Q48WD8
C	40	MSE	-	SEE REMARK 999	UNP Q48WD8
C	41	VAL	-	EXPRESSION TAG	UNP Q48WD8
C	81	MSE	MET	MODIFIED RESIDUE	UNP Q48WD8
D	40	MSE	-	SEE REMARK 999	UNP Q48WD8
D	41	VAL	-	EXPRESSION TAG	UNP Q48WD8
D	81	MSE	MET	MODIFIED RESIDUE	UNP Q48WD8

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

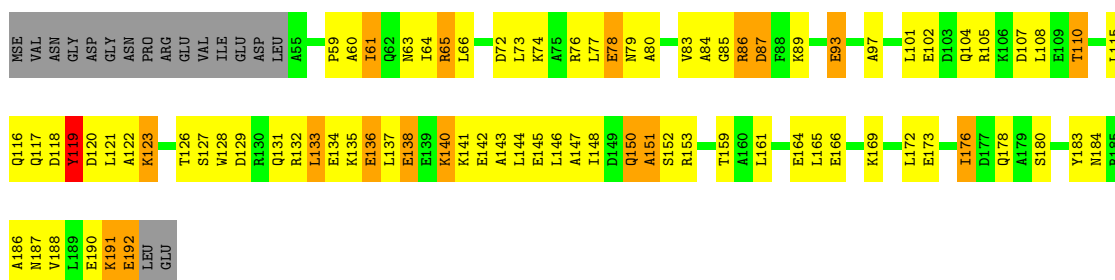
- Molecule 1: M protein

Chain A:



- Molecule 1: M protein

Chain B:



- Molecule 1: M protein

Chain C:



- Molecule 1: M protein

Chain D:





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	119.37Å 83.41Å 93.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.76 – 3.04 40.77 – 3.04	Depositor EDS
% Data completeness (in resolution range)	99.3 (40.76-3.04) 99.4 (40.77-3.04)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.12 (at 3.06Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.337 , 0.338 0.337 , 0.338	Depositor DCC
$R_{free}$ test set	963 reflections (5.51%)	DCC
Wilson B-factor (Å <sup>2</sup> )	81.9	Xtriage
Anisotropy	0.385	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 45.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	1 of 18433 reflections (0.005%)	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	4521	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.99% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	A	0.50	1/1138 (0.1%)	0.66	1/1517 (0.1%)
1	B	0.60	0/1143	0.74	1/1524 (0.1%)
1	C	0.58	1/1143 (0.1%)	0.68	1/1524 (0.1%)
1	D	0.48	1/1125 (0.1%)	0.58	0/1499
All	All	0.54	3/4549 (0.1%)	0.67	3/6064 (0.0%)

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	D	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	152	SER	CB-OG	6.75	1.51	1.42
1	C	111	LYS	CE-NZ	6.07	1.64	1.49
1	A	89	LYS	CE-NZ	5.41	1.62	1.49

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	108	LEU	CA-CB-CG	6.59	130.46	115.30
1	C	61	ILE	C-N-CA	-6.24	106.10	121.70
1	B	119	TYR	CA-C-N	-6.19	103.58	117.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	59	PRO	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1131	0	1129	114	0
1	B	1136	0	1134	151	0
1	C	1136	0	1134	170	3
1	D	1118	0	1118	110	3
All	All	4521	0	4515	406	3

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 45.

The worst 5 of 406 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:77:LEU:CA	1:D:77:LEU:HD21	1.36	1.54
1:C:58:ASN:HB2	1:C:61:ILE:CD1	1.42	1.49
1:A:63:ASN:CB	1:B:63:ASN:HD21	1.34	1.41
1:C:77:LEU:HA	1:D:77:LEU:CD2	1.57	1.33
1:C:58:ASN:HB2	1:C:61:ILE:CG1	1.63	1.26

All (3) 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	Distance(Å)	Clash(Å)
1:C:190:GLU:OE1	1:D:65:ARG:NH1[3_647]	1.87	0.33
1:C:190:GLU:OE2	1:D:65:ARG:CZ[3_647]	2.16	0.04
1:C:190:GLU:OE2	1:D:65:ARG:NE[3_647]	2.18	0.02



## 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	A	135/155 (87%)	84 (62%)	37 (27%)	14 (10%)	1	3
1	B	136/155 (88%)	91 (67%)	36 (26%)	9 (7%)	2	10
1	C	136/155 (88%)	93 (68%)	34 (25%)	9 (7%)	2	10
1	D	133/155 (86%)	98 (74%)	30 (23%)	5 (4%)	5	26
All	All	540/620 (87%)	366 (68%)	137 (25%)	37 (7%)	2	9

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	59	PRO
1	A	76	ARG
1	A	173	GLU
1	A	176	ILE
1	B	78	GLU

### 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	118/131 (90%)	95 (80%)	23 (20%)	2	10
1	B	118/131 (90%)	102 (86%)	16 (14%)	5	23
1	C	118/131 (90%)	104 (88%)	14 (12%)	8	29
1	D	117/131 (89%)	100 (86%)	17 (14%)	5	21
All	All	471/524 (90%)	401 (85%)	70 (15%)	4	19

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	133	LEU
1	C	73	LEU
1	D	136	GLU
1	B	136	GLU
1	B	150	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	58	ASN
1	C	104	GLN
1	D	178	GLN
1	B	116	GLN
1	D	58	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

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, 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	137/155 (88%)	-0.22	0 100 100	43, 84, 127, 154	0
1	B	138/155 (89%)	-0.19	0 100 100	40, 85, 122, 166	0
1	C	138/155 (89%)	-0.20	0 100 100	53, 98, 148, 175	0
1	D	135/155 (87%)	-0.26	0 100 100	44, 84, 126, 177	0
All	All	548/620 (88%)	-0.22	0 100 100	40, 89, 136, 177	0

There are no RSRZ outliers to report.

### 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 ⓘ

There are no ligands in this entry.

### 6.5 Other polymers ⓘ

There are no such residues in this entry.