



# Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 07:43 am BST

PDB ID : 1GSM  
Title : A reassessment of the MAdCAM-1 structure and its role in integrin recognition.  
Authors : Dando, J.; Wilkinson, K.W.; Ortlepp, S.; King, D.J.; Brady, R.L.  
Deposited on : 2002-01-08  
Resolution : 1.90 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
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

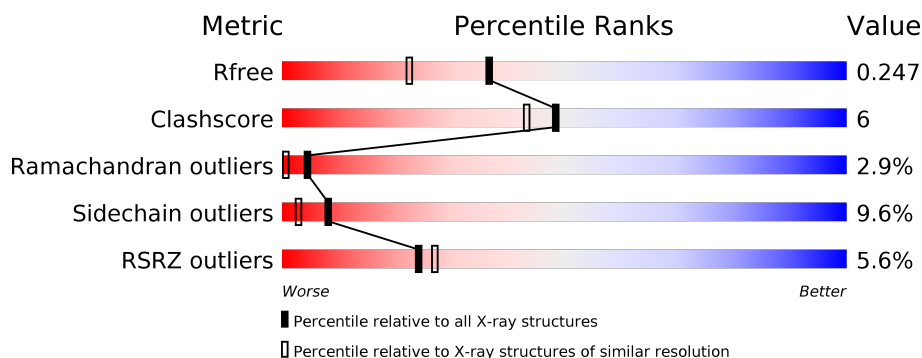
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	210	<div> <div>5%</div> <div> <div></div> <div>76%</div> <div>16%</div> <div>5% ..</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1715 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 MUCOSAL ADDRESSIN CELL ADHESION MOLECULE-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	206	Total	C	N	O	S	109	0	0
			1531	957	274	293	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	ASN	ASP	conflict	UNP Q13477

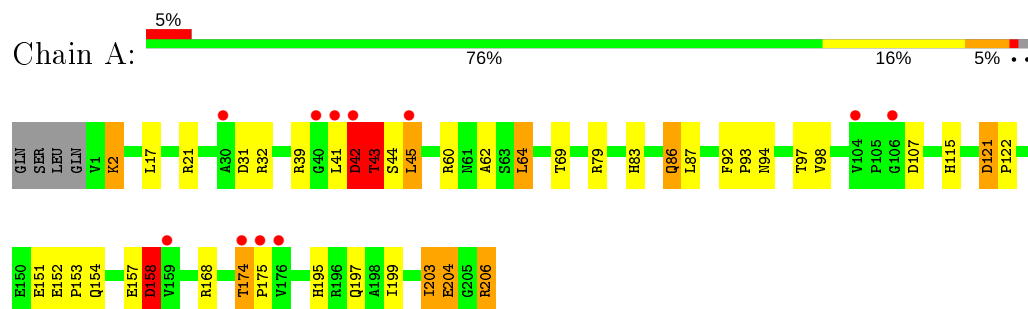
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	184	Total	O	0	0
			184	184		

### 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 ( $\text{RSRZ} > 2$ ). 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: MUCOSAL ADDRESSIN CELL ADHESION MOLECULE-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.00Å 99.20Å 70.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.49 – 1.90 29.48 – 1.78	Depositor EDS
% Data completeness (in resolution range)	97.9 (29.49-1.90) 96.1 (29.48-1.78)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.76 (at 1.78Å)	Xtriage
Refinement program	REFMAC 5.0	Depositor
R, $R_{free}$	0.217 , 0.250 0.217 , 0.247	Depositor DCC
$R_{free}$ test set	1088 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	24.5	Xtriage
Anisotropy	0.578	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 41.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1715	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% 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

### 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	1.30	7/1562 (0.4%)	1.37	19/2138 (0.9%)

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	1	3

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	44	SER	C-N	23.74	1.88	1.34
1	A	32	ARG	CB-CG	-18.30	1.03	1.52
1	A	45	LEU	CA-CB	-17.17	1.14	1.53
1	A	42	ASP	C-N	-16.18	0.96	1.34
1	A	204	GLU	CA-CB	-14.17	1.22	1.53
1	A	203	ILE	CB-CG1	10.14	1.82	1.54
1	A	158	ASP	CA-CB	9.95	1.75	1.53

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	42	ASP	CA-C-N	-19.02	75.35	117.20
1	A	44	SER	O-C-N	-16.09	96.96	122.70
1	A	42	ASP	C-N-CA	-15.95	81.83	121.70
1	A	149	GLU	CA-C-N	-15.52	83.06	117.20
1	A	152	GLU	C-N-CD	-14.33	89.08	120.60
1	A	149	GLU	O-C-N	14.12	145.29	122.70
1	A	206	ARG	CB-CA-C	10.80	132.00	110.40
1	A	42	ASP	O-C-N	-9.24	107.92	122.70
1	A	44	SER	CA-C-N	9.08	137.17	117.20

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	2	LYS	CG-CD-CE	7.48	134.34	111.90
1	A	203	ILE	CG1-CB-CG2	-6.55	96.99	111.40
1	A	107	ASP	CB-CG-OD2	6.18	123.86	118.30
1	A	45	LEU	CB-CA-C	5.74	121.11	110.20
1	A	158	ASP	CB-CG-OD2	5.62	123.36	118.30
1	A	42	ASP	CB-CG-OD2	5.45	123.20	118.30
1	A	203	ILE	CA-CB-CG1	-5.35	100.83	111.00
1	A	158	ASP	CA-CB-CG	-5.29	101.76	113.40
1	A	31	ASP	CB-CG-OD2	5.20	122.98	118.30
1	A	132	GLY	N-CA-C	-5.16	100.20	113.10

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	206	ARG	CA

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	149	GLU	Mainchain,Peptide
1	A	42	ASP	Mainchain

## 5.2 Too-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 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	1531	0	1530	17	3
2	A	184	0	0	3	0
All	All	1715	0	1530	17	3

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 6.

All (17) 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:69:THR:OG1	1:A:86:GLN:NE2	2.11	0.83
1:A:21:ARG:HD3	2:A:2030:HOH:O	1.86	0.75
1:A:174:THR:HA	1:A:175:PRO:C	2.11	0.70
1:A:39:ARG:HD2	2:A:2062:HOH:O	1.90	0.69
1:A:98:VAL:HG11	1:A:199:ILE:HG12	1.83	0.60
1:A:83:HIS:HE1	2:A:2019:HOH:O	1.86	0.56
1:A:64:LEU:HD22	1:A:158:ASP:HB3	1.90	0.54
1:A:41:LEU:HD13	1:A:69:THR:O	2.08	0.52
1:A:97:THR:HB	1:A:115:HIS:HE1	1.73	0.52
1:A:92:PHE:HE2	1:A:195:HIS:HB2	1.77	0.49
1:A:195:HIS:HE1	1:A:197:GLN:HE21	1.62	0.48
1:A:17:LEU:HD12	1:A:62:ALA:O	2.13	0.47
1:A:92:PHE:CD1	1:A:93:PRO:HD2	2.51	0.45
1:A:97:THR:CB	1:A:115:HIS:HE1	2.30	0.44
1:A:121:ASP:HA	1:A:122:PRO:HD3	1.71	0.43
1:A:92:PHE:CG	1:A:93:PRO:HD2	2.54	0.42
1:A:92:PHE:CE2	1:A:195:HIS:HB2	2.54	0.41

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	Interatomic distance (Å)	Clash overlap (Å)
1:A:43:THR:CG2	1:A:60:ARG:O[4_566]	1.12	1.08
1:A:43:THR:CG2	1:A:60:ARG:C[4_566]	2.12	0.08
1:A:43:THR:CB	1:A:60:ARG:O[4_566]	2.17	0.03

## 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	204/210 (97%)	189 (93%)	9 (4%)	6 (3%)	<b>4</b> <b>1</b>

All (6) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	A	43	THR
1	A	45	LEU
1	A	151	GLU
1	A	158	ASP
1	A	153	PRO
1	A	154	GLN

### 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	167/171 (98%)	151 (90%)	16 (10%)	8 3

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	LYS
1	A	42	ASP
1	A	43	THR
1	A	64	LEU
1	A	79	ARG
1	A	86	GLN
1	A	87	LEU
1	A	94	ASN
1	A	121	ASP
1	A	142	LEU
1	A	157	GLU
1	A	168	ARG
1	A	174	THR
1	A	203	ILE
1	A	204	GLU
1	A	206	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	83	HIS
1	A	86	GLN
1	A	95	GLN
1	A	115	HIS
1	A	197	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	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	44:SER	C	45:LEU	N	1.88
1	A	42:ASP	C	43:THR	N	0.96

## 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	196/210 (93%)	0.33	11 (5%) 24 27	18, 28, 48, 55	10 (5%)

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	45	LEU	7.9
1	A	41	LEU	5.7
1	A	30	ALA	5.5
1	A	42	ASP	4.7
1	A	174	THR	4.0
1	A	176	VAL	3.5
1	A	159	VAL	3.3
1	A	175	PRO	3.1
1	A	104	VAL	2.7
1	A	106	GLY	2.7
1	A	40	GLY	2.2

### 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

There are no such residues in this entry.