



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 15, 2014 – 12:26 PM EDT

PDB ID : 3WSR  
Title : Crystal structure of CLEC-2 in complex with O-glycosylated podoplanin  
Authors : Nagae, M.; Morita-Matsumoto, K.; Kato, M.; Kato-Kaneko, M.; Kato, Y.; Yamaguchi, Y.  
Deposited on : 2014-03-20  
Resolution : 1.91 Å(reported)

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

---

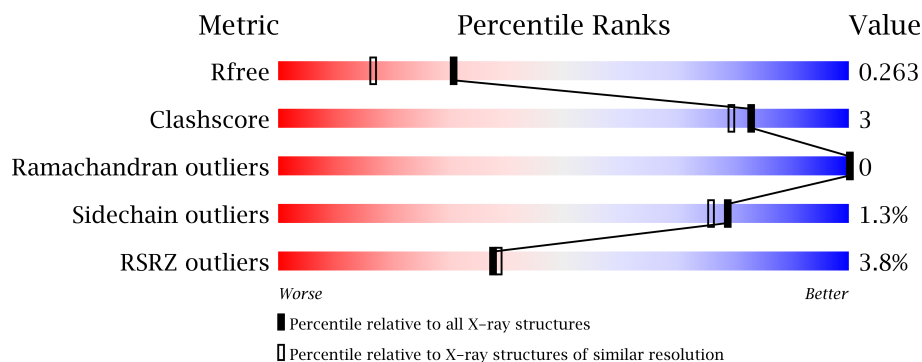
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : trunk24003  
Percentile statistics : 23426  
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) : trunk24003

# 1 Overall quality at a glance

The reported resolution of this entry is 1.91 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))
$R_{free}$	77520	5062 (1.94-1.90)
Clashscore	88313	5754 (1.94-1.90)
Ramachandran outliers	86584	5685 (1.94-1.90)
C $\alpha$ geometry	86677	5678 (1.94-1.90)
Sidechain outliers	86556	5686 (1.94-1.90)
RSRZ outliers	77580	5064 (1.94-1.90)

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. 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	128	
1	B	128	
2	C	18	
2	D	18	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2292 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 C-type lectin domain family 1 member B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	120	Total	C	N	O	S	0	0	0
			1012	639	178	184	11			
1	B	120	Total	C	N	O	S	0	0	0
			1012	639	178	184	11			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	GLY	-	EXPRESSION TAG	UNP Q9P126
A	95	SER	-	EXPRESSION TAG	UNP Q9P126
A	99	SER	CYS	ENGINEERED MUTATION	UNP Q9P126
B	94	GLY	-	EXPRESSION TAG	UNP Q9P126
B	95	SER	-	EXPRESSION TAG	UNP Q9P126
B	99	SER	CYS	ENGINEERED MUTATION	UNP Q9P126

- Molecule 2 is a protein called Peptide from Podoplanin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	S	0	0	0
			69	40	10	18	1			
2	D	11	Total	C	N	O	S	0	0	0
			73	42	11	19	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	55	CYS	-	EXPRESSION TAG	UNP Q86YL7
D	55	CYS	-	EXPRESSION TAG	UNP Q86YL7

- Molecule 3 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	C	3	Total	C	N	O	0	0
			45	25	2	18		
3	D	3	Total	C	N	O	0	0
			45	25	2	18		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	21	Total	O	0	0
			21	21		
4	B	12	Total	O	0	0
			12	12		
4	C	1	Total	O	0	0
			1	1		
4	D	2	Total	O	0	0
			2	2		

### 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: C-type lectin domain family 1 member B

Chain A: 



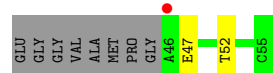
- Molecule 1: C-type lectin domain family 1 member B

Chain B: 



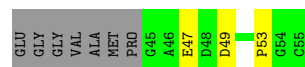
- Molecule 2: Peptide from Podoplanin

Chain C: 



- Molecule 2: Peptide from Podoplanin

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.37Å 51.19Å 54.10Å 90.00° 112.77° 90.00°	Depositor
Resolution (Å)	50.13 – 1.91 35.82 – 1.91	Depositor EDS
% Data completeness (in resolution range)	99.3 (50.13-1.91) 99.4 (35.82-1.91)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.01 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.217 , 0.259 0.225 , 0.263	Depositor DCC
$R_{free}$ test set	1022 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 29.9	EDS
Estimated twinning fraction	0.025 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 21361 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	2292	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	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.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SIA, GAL, A2G

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	0.51	0/1041	0.71	1/1401 (0.1%)
1	B	0.47	0/1041	0.70	1/1401 (0.1%)
2	C	0.53	0/69	0.74	0/93
2	D	0.58	0/73	0.73	0/98
All	All	0.50	0/2224	0.70	2/2993 (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	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	214	LEU	CA-CB-CG	6.84	131.04	115.30
1	A	214	LEU	CA-CB-CG	5.27	127.42	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	111	ASP	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	1012	0	940	9	0
1	B	1012	0	940	5	0
2	C	69	0	56	3	0
2	D	73	0	59	3	0
3	C	45	0	38	0	0
3	D	45	0	38	3	0
4	A	21	0	0	0	1
4	B	12	0	0	0	0
4	C	1	0	0	0	0
4	D	2	0	0	0	0
All	All	2292	0	2071	15	1

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 3.

All (15) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:163:ARG:HE	1:B:194:ASN:HD22	1.42	0.68
1:A:163:ARG:HE	1:A:194:ASN:HD22	1.44	0.66
1:A:104:THR:HG21	3:D:1003:GAL:O2	2.00	0.61
2:C:52:THR:HG22	2:D:53:PRO:HG2	1.82	0.60
1:B:142:ASN:OD1	1:B:144:ASN:HB2	2.06	0.55
1:B:152:ARG:HG2	2:D:47:GLU:HG3	1.91	0.53
1:B:152:ARG:HG2	2:D:47:GLU:CG	2.41	0.51
1:A:163:ARG:HE	1:A:194:ASN:ND2	2.08	0.50
1:A:143:ARG:NH1	1:A:147:GLU:HG3	2.28	0.48
1:A:104:THR:HG23	3:D:1001:A2G:H8B	1.96	0.48
1:A:152:ARG:HG3	2:C:47:GLU:HG2	1.96	0.47
1:A:140:ILE:HG23	1:A:145:ILE:HG23	1.96	0.46
1:B:117:PHE:CZ	3:D:1002:SIA:H112	2.51	0.46
1:A:152:ARG:CG	2:C:47:GLU:HG2	2.46	0.45
1:A:126:SER:HG	1:A:158:TRP:HD1	1.68	0.42

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	Distance(Å)	Clash(Å)
4:A:315:HOH:O	4:A:316:HOH:O[2_645]	1.98	0.22

## 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	118/128 (92%)	116 (98%)	2 (2%)	0	100	100
1	B	118/128 (92%)	115 (98%)	3 (2%)	0	100	100
2	C	8/18 (44%)	7 (88%)	1 (12%)	0	100	100
2	D	9/18 (50%)	9 (100%)	0	0	100	100
All	All	253/292 (87%)	247 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	109/114 (96%)	108 (99%)	1 (1%)	86	85
1	B	109/114 (96%)	108 (99%)	1 (1%)	86	85
2	C	8/12 (67%)	8 (100%)	0	100	100
2	D	8/12 (67%)	7 (88%)	1 (12%)	6	2
All	All	234/252 (93%)	231 (99%)	3 (1%)	79	76

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	214	LEU
1	B	214	LEU
2	D	49	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	194	ASN
1	B	194	ASN
1	B	204	HIS

### 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 ⓘ

6 carbohydrates 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	A2G	C	1001	3,2	14,14,15	0.38	0	19,19,21	1.05	1 (5%)
3	SIA	C	1002	3	17,20,21	0.70	0	25,28,31	1.53	3 (12%)
3	GAL	C	1003	3	11,11,12	0.75	0	15,15,17	1.20	2 (13%)
3	A2G	D	1001	3,2	14,14,15	0.48	0	19,19,21	1.08	0
3	SIA	D	1002	3	17,20,21	0.46	0	25,28,31	1.71	2 (8%)
3	GAL	D	1003	3	11,11,12	0.71	0	15,15,17	1.31	3 (20%)

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
3	A2G	C	1001	3,2	-	0/6/23/26	0/1/1/1
3	SIA	C	1002	3	-	0/14/34/38	0/1/1/1
3	GAL	C	1003	3	-	0/2/19/22	0/1/1/1
3	A2G	D	1001	3,2	-	0/6/23/26	0/1/1/1
3	SIA	D	1002	3	-	0/14/34/38	0/1/1/1
3	GAL	D	1003	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	1002	SIA	O6-C2-C1	7.60	112.91	108.23
3	C	1002	SIA	O6-C2-C1	5.67	111.72	108.23
3	D	1003	GAL	C1-O5-C5	3.16	116.19	112.13
3	C	1002	SIA	O9-C9-C8	-2.96	104.57	111.06
3	D	1003	GAL	O5-C1-C2	-2.75	106.53	110.84
3	C	1003	GAL	O5-C1-C2	-2.54	106.86	110.84
3	C	1001	A2G	C3-C2-N2	-2.37	104.62	110.50
3	C	1002	SIA	C8-C7-C6	-2.21	108.70	113.02
3	C	1003	GAL	O3-C3-C2	2.13	113.88	110.00
3	D	1003	GAL	O5-C5-C6	2.08	110.56	106.97
3	D	1002	SIA	O9-C9-C8	-2.00	106.67	111.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	120/128 (93%)	0.10	4 (3%) 43 45	15, 25, 48, 53	0
1	B	120/128 (93%)	0.03	5 (4%) 34 35	14, 26, 50, 68	0
2	C	10/18 (55%)	0.27	1 (10%) 7 7	25, 32, 39, 41	0
2	D	11/18 (61%)	-0.05	0 100 100	23, 26, 36, 39	0
All	All	261/292 (89%)	0.07	10 (3%) 37 38	14, 26, 48, 68	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	111	ASP	3.4
1	B	188	ASP	3.0
1	B	220	ALA	2.8
1	A	143	ARG	2.3
1	B	111	ASP	2.3
2	C	46	ALA	2.3
1	A	220	ALA	2.2
1	A	183	PHE	2.2
1	B	192	ASN	2.1
1	B	104	THR	2.0

### 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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains.

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	A2G	C	1001	14/15	0.11	2.36	23,27,31,33	0
3	GAL	D	1003	11/12	0.17	0.62	29,37,40,41	0
3	GAL	C	1003	11/12	0.20	0.47	40,43,46,51	0
3	SIA	D	1002	20/21	0.10	-0.05	22,23,29,33	0
3	SIA	C	1002	20/21	0.10	-0.17	18,21,26,31	0
3	A2G	D	1001	14/15	0.09	-1.02	22,22,24,24	0

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.