



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 03:19 PM EST

PDB ID : 5MSN  
Title : Structure of the Dcc1 Protein  
Authors : Wade, B.O.; Singleton, M.R.  
Deposited on : 2017-01-05  
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

<http://wwpdb.org/validation/2016/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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

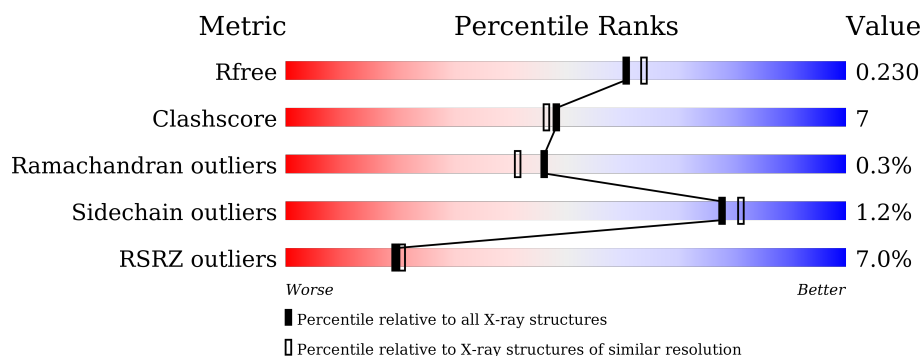
# 1 Overall quality at a glance

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))
$R_{free}$	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (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. 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>4%</div> <div> <div></div> <div>73%</div> <div>14%</div> <div>•</div> <div>12%</div> </div> </div>
1	B	293	<div> <div>4%</div> <div> <div></div> <div>71%</div> <div>16%</div> <div>•</div> <div>13%</div> </div> </div>
1	C	293	<div> <div>8%</div> <div> <div></div> <div>74%</div> <div>12%</div> <div>•</div> <div>13%</div> </div> </div>
1	D	293	<div> <div>9%</div> <div> <div></div> <div>67%</div> <div>18%</div> <div>•</div> <div>13%</div> </div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 8931 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 DCC1 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	259	Total	C	N	O	S	0	0	0
			2114	1365	355	384	10			
1	B	256	Total	C	N	O	S	0	0	0
			2089	1349	351	379	10			
1	C	255	Total	C	N	O	S	0	0	0
			2077	1343	347	377	10			
1	D	255	Total	C	N	O	S	0	0	0
			2078	1343	348	377	10			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	88	GLY	-	expression tag	UNP P25559
A	89	SER	-	expression tag	UNP P25559
B	88	GLY	-	expression tag	UNP P25559
B	89	SER	-	expression tag	UNP P25559
C	88	GLY	-	expression tag	UNP P25559
C	89	SER	-	expression tag	UNP P25559
D	88	GLY	-	expression tag	UNP P25559
D	89	SER	-	expression tag	UNP P25559

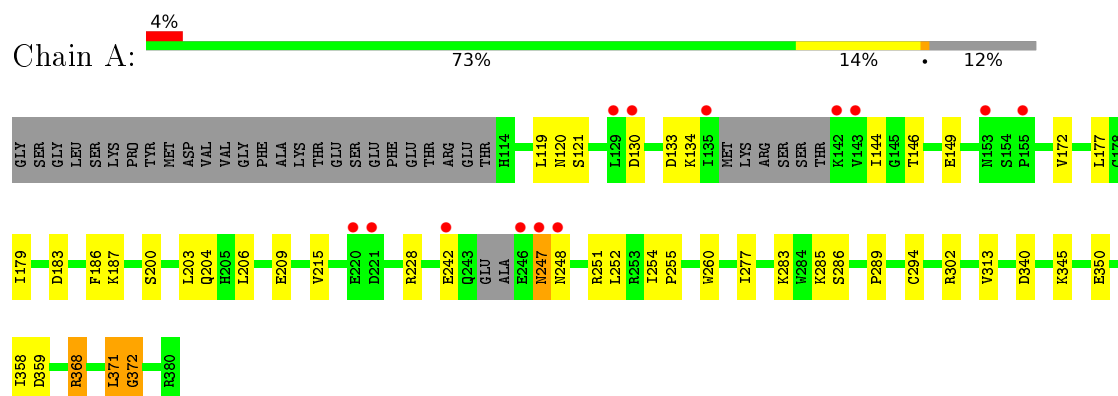
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	176	Total	O	0	0
			176	176		
2	B	177	Total	O	0	0
			177	177		
2	C	109	Total	O	0	0
			109	109		
2	D	111	Total	O	0	0
			111	111		

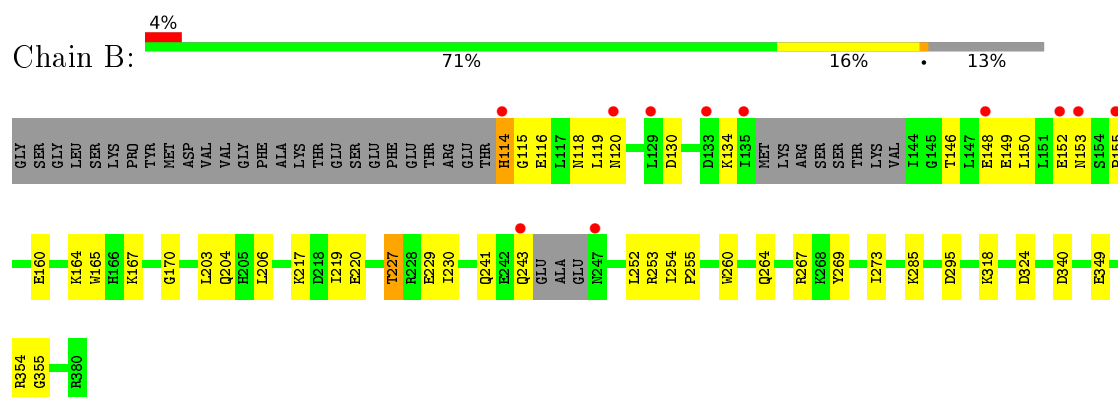
### 3 Residue-property plots [i](#)

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

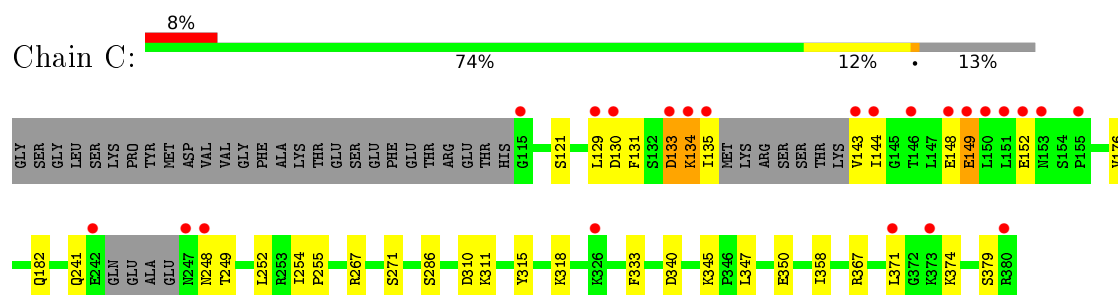
#### • Molecule 1: DCC1 protein



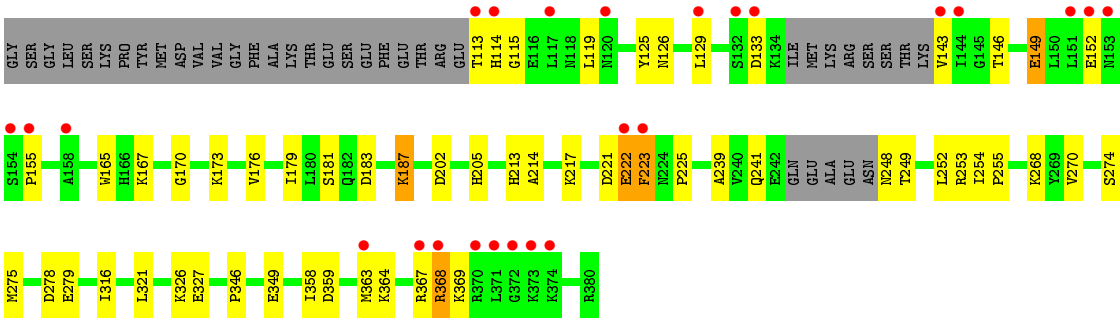
#### • Molecule 1: DCC1 protein



#### • Molecule 1: DCC1 protein



#### • Molecule 1: DCC1 protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.68 Å 100.70 Å 83.50 Å 90.00° 100.67° 90.00°	Depositor
Resolution (Å)	29.36 – 2.00 29.36 – 2.00	Depositor EDS
% Data completeness (in resolution range)	95.0 (29.36-2.00) 95.0 (29.36-2.00)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.98 (at 2.00 Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.189 , 0.230 0.188 , 0.230	Depositor DCC
$R_{free}$ test set	3558 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtriage
Anisotropy	0.014	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 57.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8931	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.85% 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	0.58	0/2161	0.77	5/2914 (0.2%)
1	B	0.61	0/2136	0.82	7/2881 (0.2%)
1	C	0.60	3/2123 (0.1%)	0.91	11/2864 (0.4%)
1	D	0.67	3/2125 (0.1%)	1.09	19/2867 (0.7%)
All	All	0.61	6/8545 (0.1%)	0.90	42/11526 (0.4%)

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
1	D	0	1
All	All	0	2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	223	PHE	CE1-CZ	8.43	1.53	1.37
1	C	152	GLU	CD-OE2	6.13	1.32	1.25
1	C	149	GLU	CB-CG	-5.91	1.41	1.52
1	D	223	PHE	CG-CD2	5.83	1.47	1.38
1	C	149	GLU	CD-OE2	5.50	1.31	1.25

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	129	LEU	CB-CG-CD1	-17.63	81.03	111.00
1	D	368	ARG	NE-CZ-NH1	14.07	127.34	120.30
1	C	182	GLN	CA-CB-CG	-12.98	84.84	113.40
1	D	368	ARG	NE-CZ-NH2	-12.43	114.08	120.30
1	D	367	ARG	CG-CD-NE	-11.16	88.36	111.80

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	247	ASN	Peptide
1	D	152	GLU	Sidechain

## 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	2114	0	2134	34	0
1	B	2089	0	2106	36	0
1	C	2077	0	2100	28	0
1	D	2078	0	2097	37	0
2	A	176	0	0	6	0
2	B	177	0	0	6	1
2	C	109	0	0	5	1
2	D	111	0	0	5	0
All	All	8931	0	8437	123	1

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 7.

The worst 5 of 123 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:C:143:VAL:HG22	1:C:144:ILE:HG12	1.57	0.84
1:B:130:ASP:H	1:B:134:LYS:HZ2	1.28	0.81
1:C:271:SER:OG	1:D:113:THR:HG21	1.82	0.80
1:A:294:CYS:SG	2:A:551:HOH:O	2.45	0.75
1:B:267:ARG:NH1	2:B:403:HOH:O	2.20	0.74

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	Interatomic distance (Å)	Clash overlap (Å)
2:B:422:HOH:O	2:C:481:HOH:O[1_455]	2.11	0.09

## 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	253/293 (86%)	244 (96%)	7 (3%)	2 (1%)	24	15
1	B	250/293 (85%)	244 (98%)	6 (2%)	0	100	100
1	C	249/293 (85%)	241 (97%)	8 (3%)	0	100	100
1	D	249/293 (85%)	241 (97%)	7 (3%)	1 (0%)	39	33
All	All	1001/1172 (85%)	970 (97%)	28 (3%)	3 (0%)	46	41

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	248	ASN
1	A	372	GLY
1	D	222	GLU

### 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	236/265 (89%)	232 (98%)	4 (2%)	68	71
1	B	233/265 (88%)	229 (98%)	4 (2%)	68	71

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	232/265 (88%)	231 (100%)	1 (0%)	93	95
1	D	232/265 (88%)	230 (99%)	2 (1%)	84	88
All	All	933/1060 (88%)	922 (99%)	11 (1%)	78	81

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	114	HIS
1	B	119	LEU
1	C	252	LEU
1	A	371	LEU
1	B	227	THR

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	205	HIS
1	A	247	ASN
1	B	247	ASN
1	C	241	GLN
1	D	126	ASN

### 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](#)

There are no chain breaks in this entry.

## 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	259/293 (88%)	-0.02	13 (5%) 32 34	18, 33, 66, 87	0
1	B	256/293 (87%)	-0.16	11 (4%) 39 40	17, 29, 57, 75	0
1	C	255/293 (87%)	0.16	23 (9%) 12 12	21, 37, 71, 91	0
1	D	255/293 (87%)	0.22	25 (9%) 10 10	19, 38, 72, 81	0
All	All	1025/1172 (87%)	0.05	72 (7%) 19 21	17, 34, 70, 91	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	113	THR	9.3
1	D	372	GLY	6.7
1	C	143	VAL	5.9
1	D	371	LEU	5.7
1	C	135	ILE	5.6

### 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.