wwPDB X-ray Structure Validation Summary Report (i)

Apr 27, 2024 – 04:42 pm BST

PDB ID	:	2WVR
Title	:	Human Cdt1:Geminin complex
Authors	:	De Marco, V.; Perrakis, A.
Deposited on	:	2009-10-19
Resolution	:	3.30 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

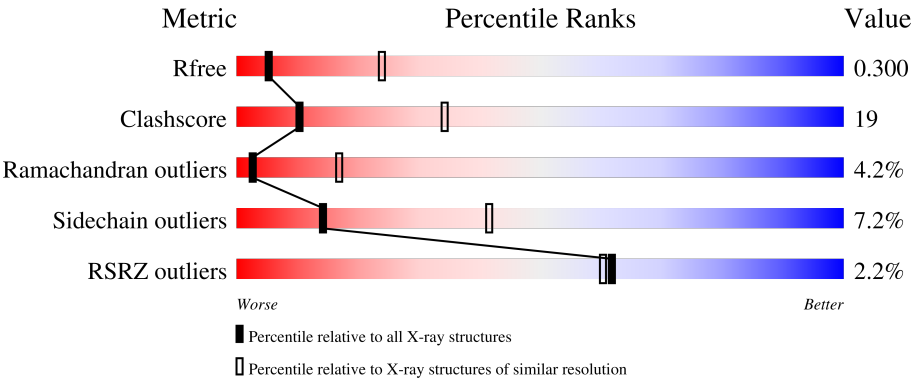
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	209	23% 9% • 67%
1	B	209	23% 10% • 65%
2	C	546	% 17% 11% • 68%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2611 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 GEMININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	70	Total	C	N	O	S	0	0	0
			590	368	106	115	1			
1	B	74	Total	C	N	O	S	0	0	0
			629	396	110	121	2			

- Molecule 2 is a protein called DNA REPLICATION FACTOR CDT1.

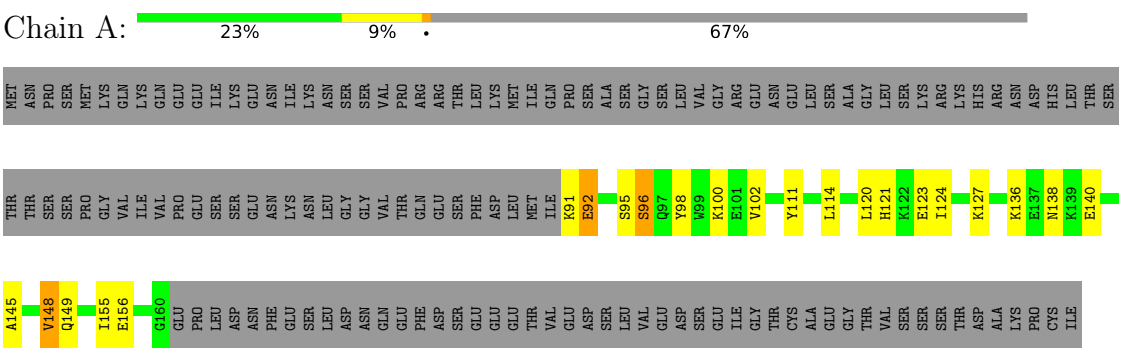
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	172	Total	C	N	O	S	0	0	0
			1392	885	254	247	6			

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

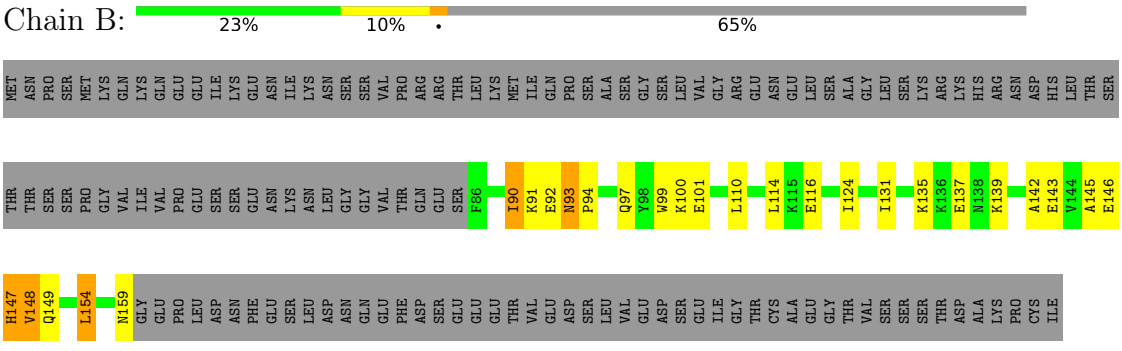
• Molecule 1: GEMININ

Chain A:



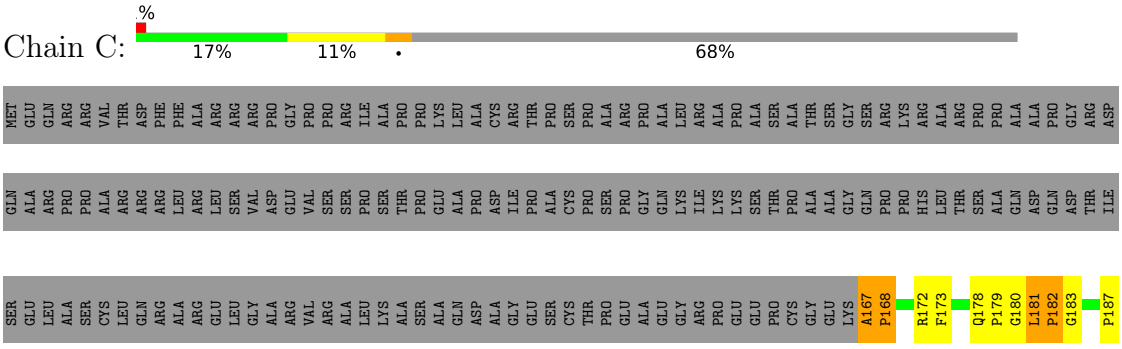
• Molecule 1: GEMININ

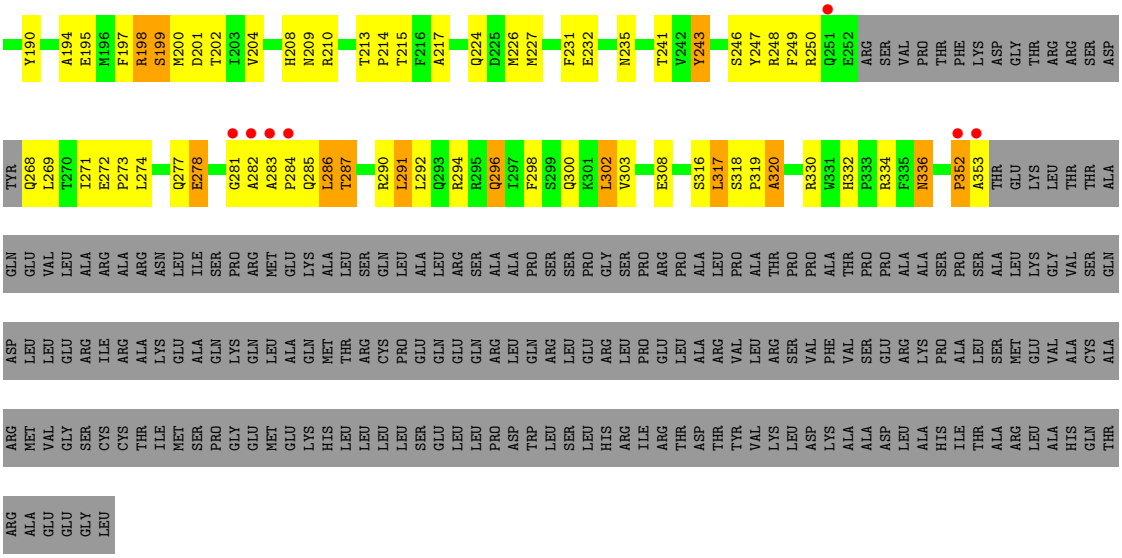
Chain B:



• Molecule 2: DNA REPLICATION FACTOR CDT1

Chain C:




Y190 A194 E195 M196 F197 R198 S199 M200 D201 T202 I203 V204 H208 N209 R210 T213 P214 T215 F216 A217 Q224 D225 M226 M227 F231 E232 N235 T241 V242 Y243 S246 Y247 R248 F249 R250 Q251 E252 ARG SER VAL PRO THR PHE LYS ASP GLY THR ARG ARG SER ASP
TYR Q268 L269 T270 I271 E272 P273 L274 Q277 E278 G281 A282 A283 P284 Q285 L286 T287 R290 L291 L292 Q293 R294 R295 Q296 I297 F298 S299 Q300 K301 L302 V303 E308 S316 L317 S318 P319 A320 R330 W331 H332 P333 R334 F335 N336 P352 A353 THR GLU LYS LEU THR THR ALA
GLN GLU VAL LEU ALA ARG ALA ARG ASN LEU ILE SER PRO ARG MET GLU LYS ALA LEU SER GLN LEU ALA LEU ARG SER ALA ALA PRO SER SER PRO GLY SER PRO ARG PRO ALA LEU PRO ALA THR PRO PRO ALA THR PRO PRO ALA ALA SER PRO SER ALA LEU LYS GLY VAL SER GLN
ASP LEU LEU GLU ARG ILE ARG ALA LYS GLU ALA GLN LYS GLN LEU ALA GLN MET THR ARG CYS PRO GLU SER GLN GLU GLN GLN ARG LEU GLN ARG LEU GLU LEU ARG LEU PRO GLU LEU ALA ARG VAL LEU ARG SER VAL PHE VAL SER GLU ARG LYS PRO ALA LEU SER MET GLU VAL ALA GLN THR
ARG MET VAL GLY SER CYS CYS THR ILE MET SER PRO GLY GLU MET GLU LYS HIS LEU LEU LEU LEU SER GLU LEU LEU PRO ASP TRP LEU SER LEU HIS ARG ILE ARG THR ASP THR TYR VAL LYS LEU ASP LYS ALA ALA ASP LEU ALA HIS ILE THR ALA ARG LEU ALA HIS GLN THR
ARG ALA GLU GLU GLY LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	92.80Å 92.80Å 164.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.41 – 3.30 19.41 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.41-3.30) 99.7 (19.41-3.30)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.60 (at 3.29Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.240 , 0.301 0.239 , 0.300	Depositor DCC
R_{free} test set	529 reflections (4.68%)	wwPDB-VP
Wilson B-factor (Å^2)	136.5	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 100.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2611	wwPDB-VP
Average B, all atoms (Å^2)	146.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.97% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.25	0/597	0.39	0/795
1	B	0.25	0/637	0.39	0/849
2	C	0.27	0/1427	0.48	0/1935
All	All	0.26	0/2661	0.44	0/3579

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
2	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	167	ALA	Peptide

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	590	0	594	17	0
1	B	629	0	635	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	1392	0	1392	71	0
All	All	2611	0	2621	98	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 19.

The worst 5 of 98 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:C:283:ALA:HB3	2:C:284:PRO:HD3	1.55	0.87
2:C:298:PHE:O	2:C:302:LEU:HD12	1.75	0.86
2:C:215:THR:HG22	2:C:217:ALA:H	1.44	0.83
1:B:148:VAL:HG13	1:B:149:GLN:H	1.46	0.81
2:C:290:ARG:O	2:C:294:ARG:HG3	1.94	0.68

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	68/209 (32%)	62 (91%)	4 (6%)	2 (3%)	4 24
1	B	72/209 (34%)	66 (92%)	4 (6%)	2 (3%)	5 25
2	C	168/546 (31%)	141 (84%)	18 (11%)	9 (5%)	2 12
All	All	308/964 (32%)	269 (87%)	26 (8%)	13 (4%)	3 17

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	96	SER
1	B	147	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	148	VAL
2	C	168	PRO
2	C	180	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	62/189 (33%)	59 (95%)	3 (5%)	25	56
1	B	67/189 (35%)	62 (92%)	5 (8%)	13	39
2	C	149/456 (33%)	137 (92%)	12 (8%)	11	36
All	All	278/834 (33%)	258 (93%)	20 (7%)	14	41

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	291	LEU
2	C	302	LEU
2	C	336	ASN
2	C	317	LEU
1	B	154	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	285	GLN
2	C	310	HIS
2	C	336	ASN
2	C	332	HIS
1	B	159	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 monosaccharides 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

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^{th} 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^2)	Q<0.9
1	A	70/209 (33%)	-0.40	0 100 100	119, 149, 166, 172	0
1	B	74/209 (35%)	-0.37	0 100 100	116, 153, 178, 183	0
2	C	172/546 (31%)	-0.08	7 (4%) 37 35	106, 137, 208, 243	0
All	All	316/964 (32%)	-0.22	7 (2%) 62 60	106, 144, 181, 243	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	283	ALA	6.6
2	C	282	ALA	4.7
2	C	284	PRO	3.5
2	C	281	GLY	2.7
2	C	251	GLN	2.4

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.