wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 10:57 AM EDT

PDB ID	:	1QVV
Title	:	Crystal structure of the S. cerevisiae YDR533c protein
Authors	:	Graille, M.; Leulliot, N.; Quevillon-Cheruel, S.; van Tilbeurgh, H.
Deposited on	:	2003-08-29
Resolution	:	2.35 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:

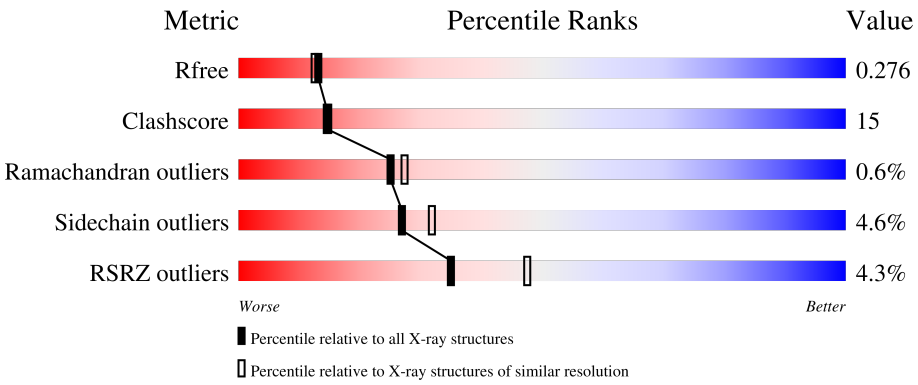
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	237	4% 69% 27% • •
1	B	237	4% 71% 26% •
1	C	237	5% 70% 27% • •
1	D	237	4% 69% 27% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7552 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 YDR533c protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	234	1800	1149	294	355	2	0	0	0
1	B	237	1820	1162	297	358	3	0	0	0
1	C	235	1807	1154	295	356	2	0	0	0
1	D	234	1800	1149	294	355	2	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	138	CME	CYS	modified residue	UNP Q04432
B	138	CME	CYS	modified residue	UNP Q04432
C	138	CME	CYS	modified residue	UNP Q04432
D	138	CME	CYS	modified residue	UNP Q04432

- Molecule 2 is water.

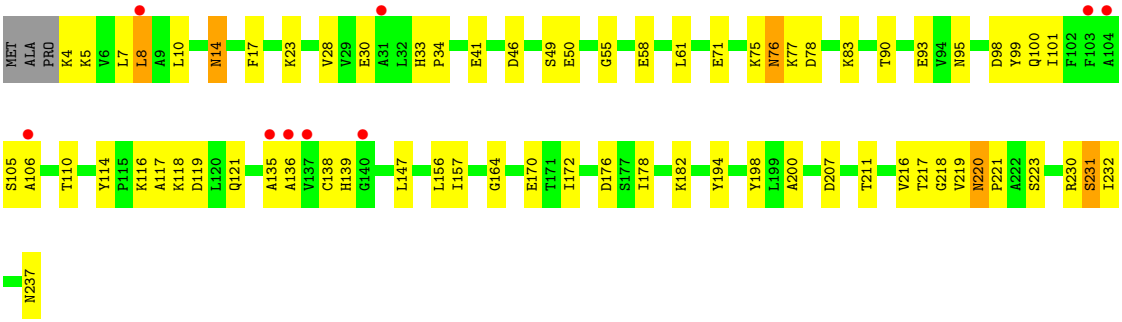
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	81	81	81	0	0
2	B	87	87	87	0	0
2	C	80	80	80	0	0
2	D	77	77	77	0	0

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.

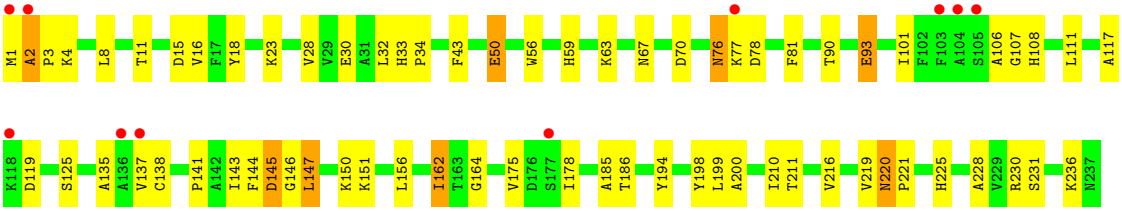
- Molecule 1: YDR533c protein

Chain A:



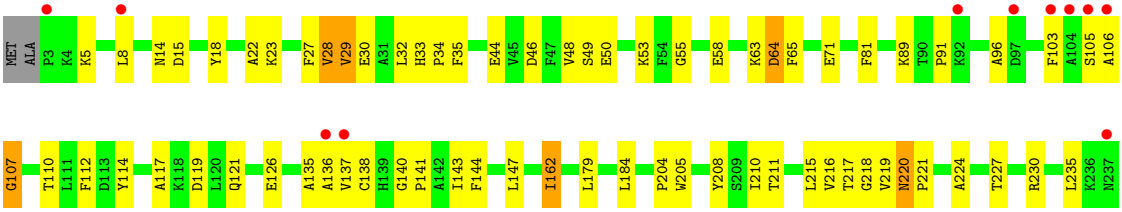
- Molecule 1: YDR533c protein

Chain B:



- Molecule 1: YDR533c protein

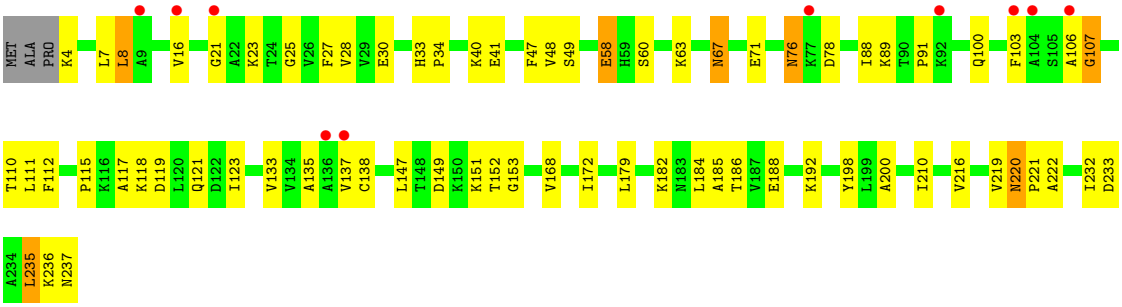
Chain C:



- Molecule 1: YDR533c protein

Chain D:

4%
69%
27%


MET ALA PRO K4 L7 L8 A9 V16 G21 A22 K23 T24 G25 V26 F27 V28 V29 E30 H33 P34 K40 E41 F47 V48 S49 E58 H59 S60 K63 N67 E71 N76 K77 D78 I88 K89 T90 P91 K92 Q100 F103 A104 S105 A106 G107
T110 L111 F112 P115 K116 A117 K118 D119 L120 Q121 D122 I123 V133 V134 A135 A136 V137 C138 L147 T148 D149 K150 K151 T152 G153 V168 I172 L179 K182 N183 L184 A185 T186 V187 E188 K192 Y198 L199 A200 I210 V216 V219 N220 P221 A222 I232 D233
A234 L235 K236 N237

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	67.02Å 93.64Å 151.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.47 – 2.35 29.47 – 2.20	Depositor EDS
% Data completeness (in resolution range)	77.3 (29.47-2.35) 73.4 (29.47-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	1.59 (at 2.20Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.228 , 0.283 0.220 , 0.276	Depositor DCC
R_{free} test set	2076 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	24.0	Xtriage
Anisotropy	0.429	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 44.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7552	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 58.01 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.1768e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CME

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	0.39	0/1828	0.62	0/2473
1	B	0.39	0/1849	0.61	1/2502 (0.0%)
1	C	0.39	0/1836	0.64	1/2484 (0.0%)
1	D	0.38	0/1828	0.63	1/2473 (0.0%)
All	All	0.38	0/7341	0.63	3/9932 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	107	GLY	N-CA-C	-6.77	96.18	113.10
1	D	107	GLY	N-CA-C	-6.48	96.89	113.10
1	B	107	GLY	N-CA-C	-5.39	99.61	113.10

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	1800	0	1759	53	0
1	B	1820	0	1783	54	0
1	C	1807	0	1767	55	0
1	D	1800	0	1759	60	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	81	0	0	2	0
2	B	87	0	0	4	0
2	C	80	0	0	2	0
2	D	77	0	0	5	0
All	All	7552	0	7068	220	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 15.

The worst 5 of 220 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:D:236:LYS:HD2	1:D:237:ASN:ND2	1.92	0.85
1:C:27:PHE:HB3	1:C:30:GLU:HG3	1.60	0.83
1:C:23:LYS:H	1:C:58:GLU:HG3	1.42	0.83
1:B:147:LEU:HD13	1:B:156:LEU:HD23	1.62	0.79
1:D:220:ASN:HB2	1:D:221:PRO:HD2	1.68	0.76

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	231/237 (98%)	220 (95%)	11 (5%)	0	100 100
1	B	234/237 (99%)	217 (93%)	14 (6%)	3 (1%)	12 10
1	C	232/237 (98%)	218 (94%)	13 (6%)	1 (0%)	34 38
1	D	231/237 (98%)	219 (95%)	10 (4%)	2 (1%)	17 17
All	All	928/948 (98%)	874 (94%)	48 (5%)	6 (1%)	25 27

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	2	ALA
1	C	28	VAL
1	B	145	ASP
1	B	108	HIS
1	D	235	LEU

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	189/191 (99%)	178 (94%)	11 (6%)	20 22
1	B	191/191 (100%)	181 (95%)	10 (5%)	23 27
1	C	190/191 (100%)	183 (96%)	7 (4%)	34 42
1	D	189/191 (99%)	182 (96%)	7 (4%)	34 42
All	All	759/764 (99%)	724 (95%)	35 (5%)	27 32

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	8	LEU
1	D	58	GLU
1	D	118	LYS
1	B	50	GLU
1	B	8	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	100	GLN
1	D	237	ASN
1	B	237	ASN
1	C	14	ASN
1	C	69	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)

4 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
1	CME	A	138	1	8,9,10	0.50	0	6,9,11	0.52	0
1	CME	B	138	1	8,9,10	0.74	0	6,9,11	0.55	0
1	CME	C	138	1	8,9,10	0.51	0	6,9,11	0.44	0
1	CME	D	138	1	8,9,10	0.73	0	6,9,11	0.53	0

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
1	CME	A	138	1	-	3/5/8/10	-
1	CME	B	138	1	-	4/5/8/10	-
1	CME	C	138	1	-	4/5/8/10	-
1	CME	D	138	1	-	3/5/8/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	138	CME	N-CA-CB-SG
1	B	138	CME	N-CA-CB-SG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
1	B	138	CME	CE-SD-SG-CB
1	B	138	CME	SD-CE-CZ-OH
1	C	138	CME	CE-SD-SG-CB

There are no ring outliers.

4 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	138	CME	5	0
1	B	138	CME	4	0
1	C	138	CME	4	0
1	D	138	CME	4	0

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 (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, 95th 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	233/237 (98%)	0.15	9 (3%)	39	52	18, 29, 42, 49	0
1	B	236/237 (99%)	0.21	10 (4%)	36	48	18, 32, 44, 54	0
1	C	234/237 (98%)	0.26	11 (4%)	31	44	18, 30, 45, 53	0
1	D	233/237 (98%)	0.29	10 (4%)	35	47	17, 30, 46, 53	0
All	All	936/948 (98%)	0.23	40 (4%)	35	47	17, 30, 45, 54	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	104	ALA	4.6
1	B	137	VAL	4.6
1	D	77	LYS	4.3
1	C	104	ALA	4.0
1	A	104	ALA	3.9

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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. The B-factors column lists the minimum, median, 95th 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	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	CME	B	138	10/11	0.84	0.25	32,37,45,50	0
1	CME	C	138	10/11	0.85	0.21	24,27,45,47	0
1	CME	D	138	10/11	0.85	0.25	26,32,50,50	0
1	CME	A	138	10/11	0.92	0.21	29,35,41,45	0

6.3 Carbohydrates (i)

There are no monosaccharides 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.