wwPDB X-ray Structure Validation Summary Report (i)

Dec 10, 2022 – 11:23 PM EST

PDB ID	:	1ERJ
Title	:	CRYSTAL STRUCTURE OF THE C-TERMINAL WD40 DOMAIN OF TUP1
Authors	:	Sprague, E.R.; Redd, M.J.; Johnson, A.D.; Wolberger, C.
Deposited on	:	2000-04-06
Resolution	:	2.30 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 (1)) were used in the production of this report:

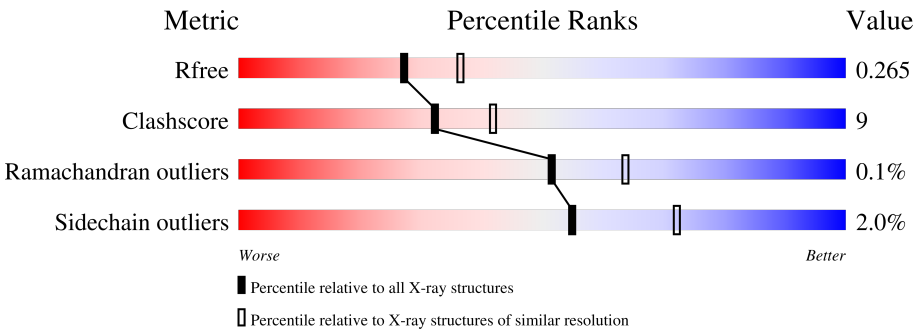
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)

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%

Mol	Chain	Length	Quality of chain
1	A	393	77% 12% • 11%
1	B	393	71% 16% • 11%
1	C	393	81% 9% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8244 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 TRANSCRIPTIONAL REPRESSOR TUP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	350	2628	1659	454	506	9	0	0	0
1	B	348	2655	1676	456	514	9	0	0	0
1	C	357	2696	1701	465	521	9	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	281	MET	-	initiating methionine	UNP P16649
A	388A	LYS	-	linker	UNP P16649
A	388B	ASP	-	linker	UNP P16649
A	388C	PRO	-	linker	UNP P16649
B	281	MET	-	initiating methionine	UNP P16649
B	388A	LYS	-	linker	UNP P16649
B	388B	ASP	-	linker	UNP P16649
B	388C	PRO	-	linker	UNP P16649
C	281	MET	-	initiating methionine	UNP P16649
C	388A	LYS	-	linker	UNP P16649
C	388B	ASP	-	linker	UNP P16649
C	388C	PRO	-	linker	UNP P16649

- Molecule 2 is water.

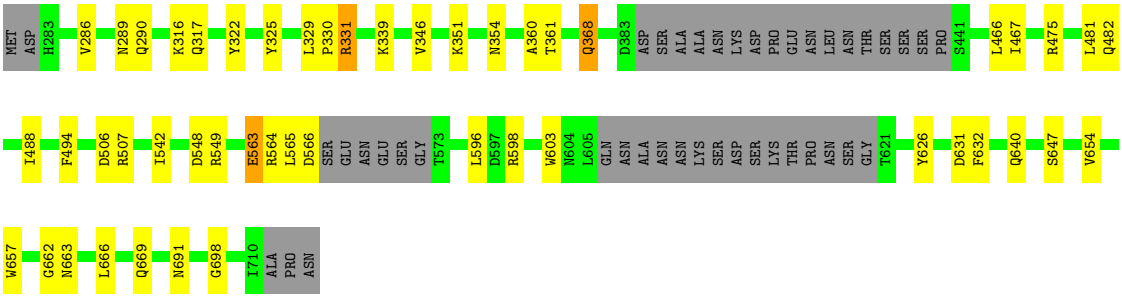
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	77	77	77	0	0
2	B	100	100	100	0	0
2	C	88	88	88	0	0

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. 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. 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.

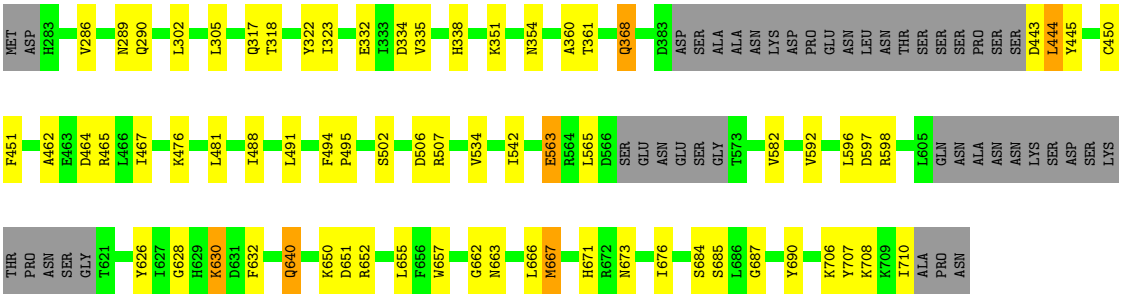
- Molecule 1: TRANSCRIPTIONAL REPRESSOR TUP1

Chain A:



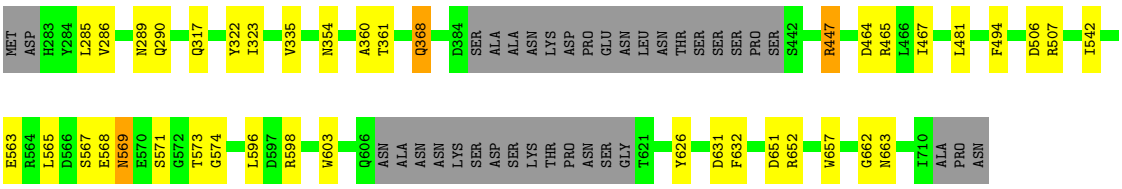
- Molecule 1: TRANSCRIPTIONAL REPRESSOR TUP1

Chain B:



- Molecule 1: TRANSCRIPTIONAL REPRESSOR TUP1

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	119.28Å 119.28Å 77.07Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.30 29.82 – 2.30	Depositor EDS
% Data completeness (in resolution range)	98.8 (30.00-2.30) 98.6 (29.82-2.30)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.32 (at 2.29Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.228 , 0.266 0.228 , 0.265	Depositor DCC
R_{free} test set	2724 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	27.1	Xtriage
Anisotropy	0.075	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 18.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l 0.447 for h,-h-k,-l 0.000 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8244	wwPDB-VP
Average B, all atoms (Å^2)	28.0	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 38.24 % 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 3.7999e-04. 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)

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	0.45	0/2682	0.74	0/3655
1	B	0.49	0/2709	0.76	0/3686
1	C	0.46	0/2751	0.75	0/3746
All	All	0.47	0/8142	0.75	0/11087

There are no bond length outliers.

There are no bond angle outliers.

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	2628	0	2466	45	0
1	B	2655	0	2543	60	0
1	C	2696	0	2552	31	0
2	A	77	0	0	3	0
2	B	100	0	0	8	0
2	C	88	0	0	1	0
All	All	8244	0	7561	136	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 9.

The worst 5 of 136 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:B:351:LYS:HZ3	1:B:451:PHE:H	1.11	0.91
1:B:318:THR:HG21	2:B:771:HOH:O	1.72	0.89
1:B:351:LYS:HD2	1:B:450:CYS:HA	1.56	0.87
1:B:351:LYS:NZ	1:B:451:PHE:H	1.72	0.87
1:B:317:GLN:HG3	1:B:322:TYR:HB3	1.59	0.84

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	342/393 (87%)	328 (96%)	14 (4%)	0	100	100
1	B	340/393 (86%)	327 (96%)	13 (4%)	0	100	100
1	C	351/393 (89%)	335 (95%)	15 (4%)	1 (0%)	41	50
All	All	1033/1179 (88%)	990 (96%)	42 (4%)	1 (0%)	51	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	569	ASN

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	274/344 (80%)	269 (98%)	5 (2%)	59	75
1	B	287/344 (83%)	279 (97%)	8 (3%)	43	60
1	C	284/344 (83%)	280 (99%)	4 (1%)	67	81
All	All	845/1032 (82%)	828 (98%)	17 (2%)	55	72

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	368	GLN
1	C	563	GLU
1	B	443	ASP
1	B	444	LEU
1	B	563	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	663	ASN
1	C	691	ASN
1	B	663	ASN
1	B	669	GLN
1	B	691	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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.