Full wwPDB X-ray Structure Validation Report (i)

Aug 19, 2023 – 06:16 PM EDT

PDB ID	:	2GPE
Title	:	Structure of the DNA-binding domain of E. Coli Proline Utilization A (PUTA)
Authors	:	Jenkins, J.L.; Larson, J.; Tanner, J.J.
Deposited on	:	2006-04-17
Resolution	:	1.90 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

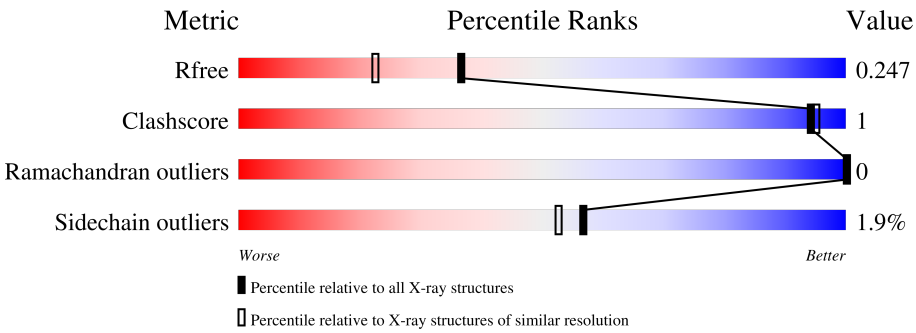
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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 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	52	87% • • 10%
1	B	52	83% • • 12%
1	C	52	83% 6% 12%
1	D	52	88% • • 8%

2 Entry composition (i)

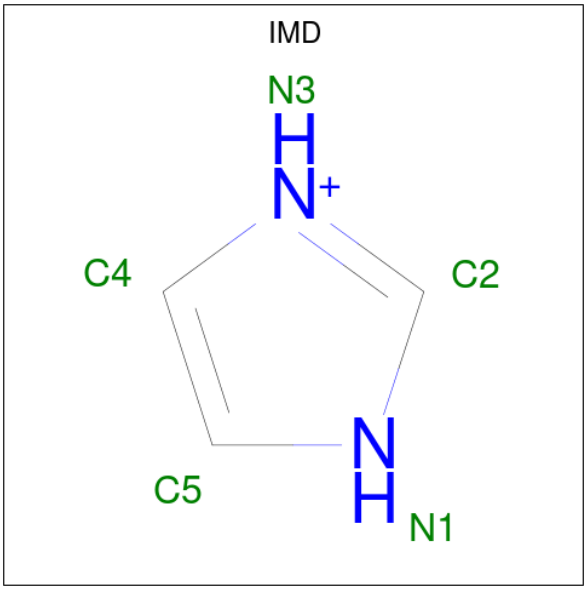
There are 3 unique types of molecules in this entry. The entry contains 1581 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Bifunctional protein putA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	47	373	232	66	74	1	0	1	0
1	B	46	365	228	65	71	1	0	1	0
1	C	46	376	236	67	72	1	0	2	0
1	D	48	379	239	64	75	1	0	2	0

- Molecule 2 is IMIDAZOLE (three-letter code: IMD) (formula: $C_3H_5N_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	N		
2	A	1	5	3	2	0	0
2	B	1	5	3	2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	N		
2	C	1	5	3	2	0	0
			Total	C	N		
2	D	1	5	3	2	0	0

- Molecule 3 is water.

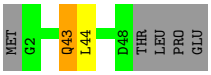
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	17	17	17	0	0
			Total	O		
3	B	19	19	19	0	0
			Total	O		
3	C	14	14	14	0	0
			Total	O		
3	D	18	18	18	0	0

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.

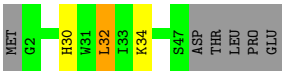
- Molecule 1: Bifunctional protein putA

Chain A:



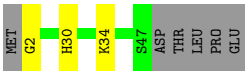
- Molecule 1: Bifunctional protein putA

Chain B:



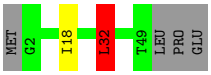
- Molecule 1: Bifunctional protein putA

Chain C:



- Molecule 1: Bifunctional protein putA

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	55.73Å 55.73Å 125.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.90 – 1.90 25.45 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.6 (50.90-1.90) 99.7 (25.45-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.54 (at 1.91Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.203 , 0.249 0.203 , 0.247	Depositor DCC
R_{free} test set	815 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	32.5	Xtriage
Anisotropy	0.423	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 55.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1581	wwPDB-VP
Average B, all atoms (Å^2)	36.0	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 19.86 % 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 9.8755e-03. 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: IMD

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	0.44	0/381	0.52	0/515
1	B	0.46	0/373	0.60	1/504 (0.2%)
1	C	0.54	1/384 (0.3%)	0.52	0/518
1	D	0.47	0/390	0.65	1/529 (0.2%)
All	All	0.48	1/1528 (0.1%)	0.57	2/2066 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	2	GLY	N-CA	6.45	1.55	1.46

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	32	LEU	CA-CB-CG	7.41	132.34	115.30
1	B	32	LEU	CA-CB-CG	5.23	127.33	115.30

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	373	0	367	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	365	0	363	1	0
1	C	376	0	382	1	0
1	D	379	0	373	1	0
2	A	5	0	5	0	0
2	B	5	0	5	0	0
2	C	5	0	5	0	0
2	D	5	0	5	0	0
3	A	17	0	0	0	0
3	B	19	0	0	0	0
3	C	14	0	0	0	0
3	D	18	0	0	0	0
All	All	1581	0	1505	4	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 1.

All (4) 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:30:HIS:CE1	1:B:34:LYS:HD2	2.39	0.57
1:C:30:HIS:CE1	1:C:34:LYS:HD2	2.49	0.47
1:D:18:ILE:HG23	1:D:32:LEU:HG	1.98	0.46
1:A:43:GLN:HG3	1:A:44:LEU:N	2.34	0.42

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	46/52 (88%)	46 (100%)	0	0	100	100
1	B	45/52 (86%)	45 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	46/52 (88%)	45 (98%)	1 (2%)	0	100	100
1	D	48/52 (92%)	48 (100%)	0	0	100	100
All	All	185/208 (89%)	184 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

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	40/46 (87%)	39 (98%)	1 (2%)	47	41
1	B	39/46 (85%)	38 (97%)	1 (3%)	46	39
1	C	41/46 (89%)	41 (100%)	0	100	100
1	D	40/46 (87%)	39 (98%)	1 (2%)	47	41
All	All	160/184 (87%)	157 (98%)	3 (2%)	57	53

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

Mol	Chain	Res	Type
1	A	43	GLN
1	B	32	LEU
1	D	32	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

4 ligands 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	$\#\|Z\| > 2$
2	IMD	B	991	-	3,5,5	0.40	0	4,5,5	0.55	0
2	IMD	A	990	-	3,5,5	0.40	0	4,5,5	0.62	0
2	IMD	C	992	-	3,5,5	0.43	0	4,5,5	0.43	0
2	IMD	D	993	-	3,5,5	0.45	0	4,5,5	0.50	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
2	IMD	B	991	-	-	-	0/1/1/1
2	IMD	A	990	-	-	-	0/1/1/1
2	IMD	C	992	-	-	-	0/1/1/1
2	IMD	D	993	-	-	-	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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.