wwPDB X-ray Structure Validation Summary Report (i)

Sep 19, 2020 – 09:30 AM BST

PDB ID	:	6TUR
Title	:	human XPG, Apo1 form
Authors	:	Ruiz, F.M.; Fernandez-Tornero, C.
Deposited on	:	2020-01-08
Resolution	:	2.90 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 following versions of software and data (see references (1)) were used in the production of this report:

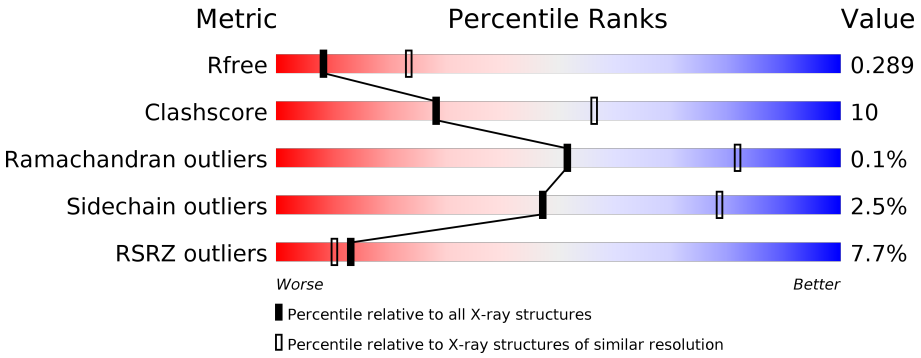
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.14.6
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 on 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	AAA	355	4% 71% 19% • 10%
1	BBB	355	5% 71% 16% • 11%
1	CCC	355	5% 71% 16% • 12%
1	DDD	355	13% 72% 17% • 10%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 10271 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 DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	320	Total	C	N	O	S	Se	0	0	0
			2600	1683	445	463	5	4			
1	BBB	317	Total	C	N	O	S	Se	0	0	0
			2561	1660	437	455	5	4			
1	CCC	311	Total	C	N	O	S	Se	0	0	0
			2516	1629	429	449	5	4			
1	DDD	319	Total	C	N	O	S	Se	0	0	0
			2594	1679	445	461	5	4			

There are 16 discrepancies between the modelled and reference sequences:

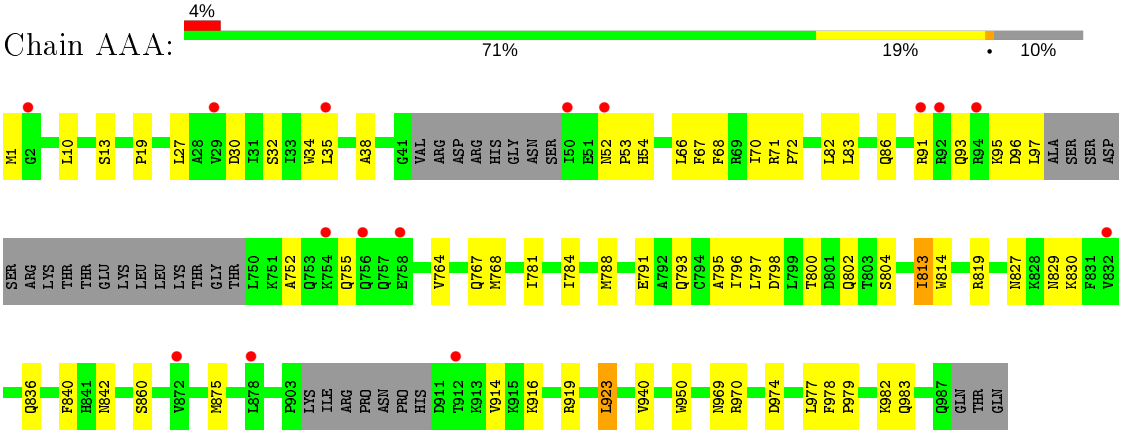
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	1	MSE	-	initiating methionine	UNP P28715
AAA	748	GLY	-	linker	UNP P28715
AAA	749	THR	-	linker	UNP P28715
AAA	812	ALA	ASP	conflict	UNP P28715
BBB	1	MSE	-	initiating methionine	UNP P28715
BBB	748	GLY	-	linker	UNP P28715
BBB	749	THR	-	linker	UNP P28715
BBB	812	ALA	ASP	conflict	UNP P28715
CCC	1	MSE	-	initiating methionine	UNP P28715
CCC	748	GLY	-	linker	UNP P28715
CCC	749	THR	-	linker	UNP P28715
CCC	812	ALA	ASP	conflict	UNP P28715
DDD	1	MSE	-	initiating methionine	UNP P28715
DDD	748	GLY	-	linker	UNP P28715
DDD	749	THR	-	linker	UNP P28715
DDD	812	ALA	ASP	conflict	UNP P28715

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.

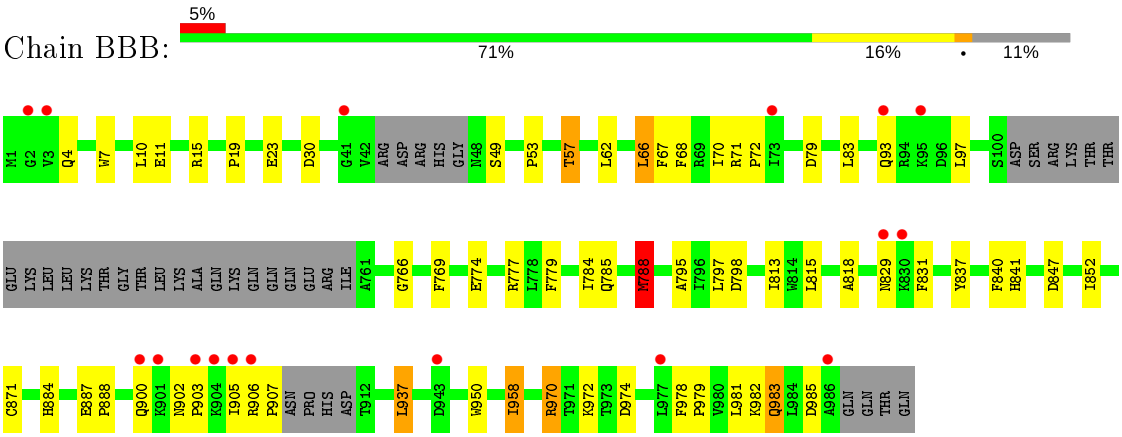
• Molecule 1: DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells

Chain AAA:



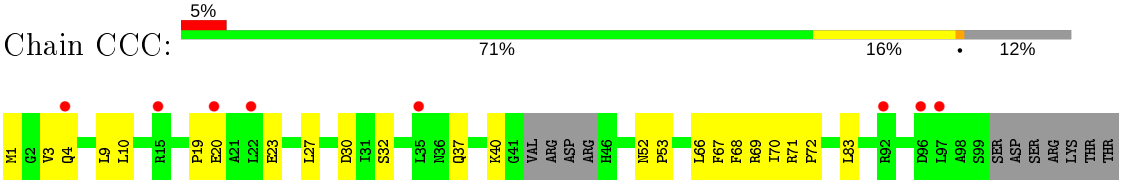
• Molecule 1: DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells

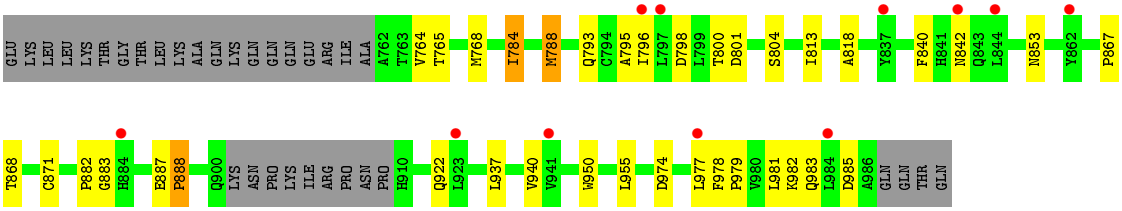
Chain BBB:



• Molecule 1: DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells

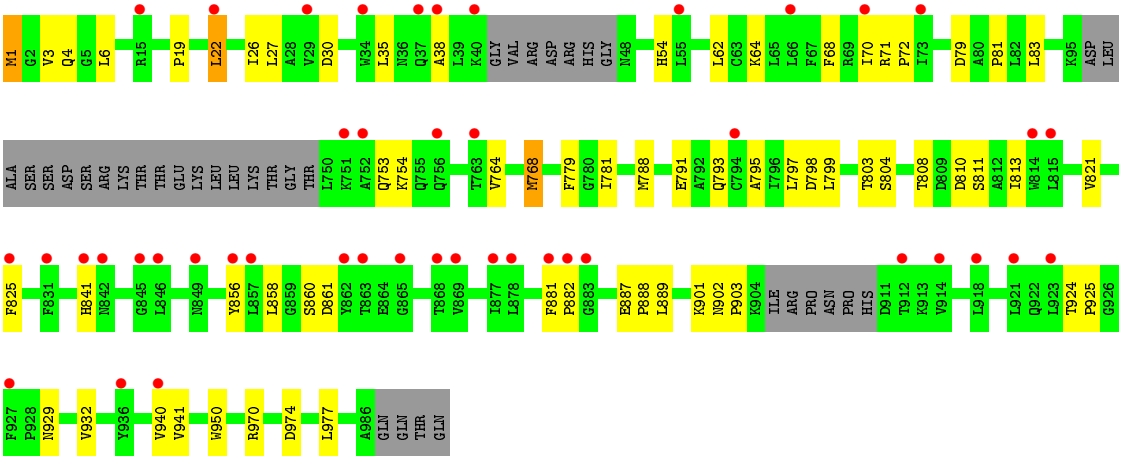
Chain CCC:





• Molecule 1: DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells

Chain DDD:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.24Å 89.56Å 268.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.60 – 2.90 134.47 – 2.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (67.60-2.90) 100.0 (134.47-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.55 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.237 , 0.281 0.245 , 0.289	Depositor DCC
R_{free} test set	1071 reflections (2.50%)	wwPDB-VP
Wilson B-factor (Å^2)	90.6	Xtriage
Anisotropy	0.202	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 89.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	10271	wwPDB-VP
Average B, all atoms (Å^2)	109.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.92% 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	AAA	0.71	0/2657	0.87	2/3585 (0.1%)
1	BBB	0.68	0/2619	0.84	1/3537 (0.0%)
1	CCC	0.68	0/2574	0.83	3/3476 (0.1%)
1	DDD	0.67	0/2651	0.84	2/3576 (0.1%)
All	All	0.69	0/10501	0.85	8/14174 (0.1%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	1	MSE	CG-SE-CE	11.01	123.13	98.90
1	BBB	788	MSE	CG-SE-CE	9.47	119.73	98.90
1	DDD	1	MSE	CG-SE-CE	8.13	116.79	98.90
1	CCC	768	MSE	CG-SE-CE	6.69	113.63	98.90
1	CCC	1	MSE	CG-SE-CE	6.03	112.17	98.90

There are no chirality outliers.

There are no planarity outliers.

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	AAA	2600	0	2625	43	0
1	BBB	2561	0	2588	46	1
1	CCC	2516	0	2520	40	1
1	DDD	2594	0	2623	80	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	10271	0	10356	202	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 10.

The worst 5 of 202 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:DDD:6:LEU:HD21	1:DDD:811:SER:OG	1.32	1.24
1:DDD:6:LEU:HD21	1:DDD:811:SER:CB	1.92	0.99
1:DDD:22:LEU:HD21	1:DDD:27:LEU:CD1	2.03	0.88
1:DDD:788:MSE:HG3	1:DDD:940:VAL:HB	1.58	0.85
1:DDD:26:ILE:HG13	1:DDD:71:ARG:HD2	1.60	0.83

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)
1:BBB:970:ARG:NH2	1:CCC:922:GLN:OE1[1_455]	2.14	0.06

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	AAA	312/355 (88%)	301 (96%)	11 (4%)	0	100 100
1	BBB	309/355 (87%)	298 (96%)	10 (3%)	1 (0%)	41 71
1	CCC	303/355 (85%)	290 (96%)	13 (4%)	0	100 100
1	DDD	311/355 (88%)	299 (96%)	12 (4%)	0	100 100
All	All	1235/1420 (87%)	1188 (96%)	46 (4%)	1 (0%)	51 82

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	766	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	AAA	280/308 (91%)	273 (98%)	7 (2%)	47	78
1	BBB	277/308 (90%)	267 (96%)	10 (4%)	35	69
1	CCC	271/308 (88%)	266 (98%)	5 (2%)	59	85
1	DDD	280/308 (91%)	274 (98%)	6 (2%)	53	81
All	All	1108/1232 (90%)	1080 (98%)	28 (2%)	47	78

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	871	CYS
1	BBB	970	ARG
1	DDD	804	SER
1	BBB	937	LEU
1	BBB	958	ILE

Some 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

There are no monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	AAA	316/355 (89%)	0.51	15 (4%)	31	28	25, 95, 162, 213	0
1	BBB	313/355 (88%)	0.64	17 (5%)	25	22	59, 92, 154, 234	0
1	CCC	307/355 (86%)	0.57	19 (6%)	20	16	66, 97, 154, 189	0
1	DDD	315/355 (88%)	0.76	45 (14%)	2	2	70, 127, 186, 215	0
All	All	1251/1420 (88%)	0.62	96 (7%)	13	10	25, 101, 169, 234	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	830	LYS	9.5
1	BBB	829	ASN	8.5
1	BBB	905	ILE	8.5
1	BBB	904	LYS	6.3
1	DDD	865	GLY	5.6

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.