wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 11:18 AM EDT

PDB ID	:	1AN4
Title	:	STRUCTURE AND FUNCTION OF THE B/HLH/Z DOMAIN OF USF
Authors	:	Ferre-D'Amare, A.R.; Pognonec, P.; Roeder, R.G.; Burley, S.K.
Deposited on	:	1997-03-15
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 (i)) were used in the production of this report:

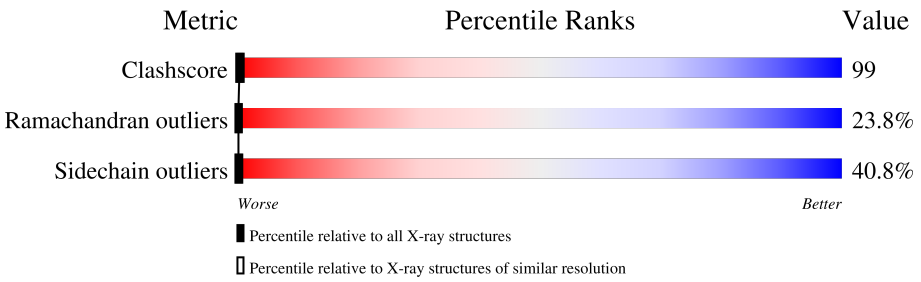
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.90 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))
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (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 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	21	62% 38%
2	D	21	19% 57% 24%
3	A	65	11% 43% 37% 9%
3	B	65	8% 37% 49% 6%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1923 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 DNA chain called DNA (5'-D(*CP*AP*CP*CP*CP*GP*GP*TP*CP*AP*CP*GP*TP*GP*GP*CP*C P*TP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	21	422	201	78	123	20	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*GP*TP*AP*GP*GP*CP*CP*AP*CP*GP*TP*GP*AP*CP*C P*GP*GP*GP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	21	433	205	83	125	20	0	0	0

- Molecule 3 is a protein called PROTEIN (UPSTREAM STIMULATORY FACTOR).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	65	534	320	109	103	2	0	0	0
3	B	65	534	320	109	103	2	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	196	MET	ARG	cloning artifact	UNP P22415
A	229	SER	CYS	engineered mutation	UNP P22415
A	248	SER	CYS	engineered mutation	UNP P22415
B	196	MET	ARG	cloning artifact	UNP P22415
B	229	SER	CYS	engineered mutation	UNP P22415
B	248	SER	CYS	engineered mutation	UNP P22415

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.

Note EDS was not executed.

• Molecule 1: DNA (5'-D(*CP*AP*CP*CP*CP*GP*GP*TP*CP*AP*CP*GP*TP*GP*GP*CP *C P*TP*AP*CP*A)-3')

Chain C: 62% 38%

C301 A302 C303 C304 C305 G306 G307 T308 G309 A310 C311 G312 T313 G314 G315 C316 C317 T318 A319 C320 A321

• Molecule 2: DNA (5'-D(*GP*TP*GP*TP*AP*GP*GP*CP*CP*AP*CP*GP*TP*GP*AP*CP *C P*GP*GP*GP*T)-3')

Chain D: 19% 57% 24%

G322 T323 G324 T325 A326 G327 G328 C329 C330 A331 C332 G333 T334 G335 A336 C337 C338 G339 G340 G341 T342

• Molecule 3: PROTEIN (UPSTREAM STIMULATORY FACTOR)

Chain A: 11% 43% 37% 9%

M196 D197 E198 K199 R200 R201 A202 Q203 H204 N205 E206 V207 E208 R209 R210 R211 R212 D213 K214 I215 N216 N217 W218 I219 V220 Q221 L222 S223 K224 I225 I226 P227 D228 S229 S230 M231 E232 S233 T234 K235 S236 G237 Q238 S239 K240 G241 G242 I243 L244 S245 K246 A247 S248 D249 Y250 I251 Q252 E253 L254 R255

Q256 S257 N258 H259 R260

• Molecule 3: PROTEIN (UPSTREAM STIMULATORY FACTOR)

Chain B: 8% 37% 49% 6%

M196 D197 E198 K199 R200 R201 A202 Q203 H204 N205 E206 V207 E208 R209 R210 R211 R212 D213 K214 I215 N216 N217 W218 I219 V220 Q221 L222 S223 K224 I225 I226 P227 D228 S229 S230 M231 E232 S233 T234 K235 S236 G237 Q238 S239 K240 G241 G242 I243 L244 S245 K246 A247 S248 D249 Y250 I251 Q252 E253 L254 R255

Q256 S257 N258 H259 R260

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	136.60Å 54.70Å 44.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	6.00 – 2.90	Depositor
% Data completeness (in resolution range)	(Not available) (6.00-2.90)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.236 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1923	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

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	C	2.46	32/472 (6.8%)	3.83	88/725 (12.1%)
2	D	2.42	25/486 (5.1%)	3.47	86/750 (11.5%)
3	A	0.89	1/539 (0.2%)	1.29	3/716 (0.4%)
3	B	0.95	1/539 (0.2%)	1.48	8/716 (1.1%)
All	All	1.80	59/2036 (2.9%)	2.78	185/2907 (6.4%)

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
1	C	0	9
2	D	0	6
3	B	0	1
All	All	0	16

The worst 5 of 59 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	318	DT	C5-C7	10.03	1.56	1.50
2	D	342	DT	C5-C7	9.54	1.55	1.50
2	D	331	DA	P-O5'	8.80	1.68	1.59
1	C	313	DT	C5-C7	8.66	1.55	1.50
1	C	308	DT	O3'-P	8.37	1.71	1.61

The worst 5 of 185 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	320	DC	O4'-C1'-N1	25.93	126.15	108.00
1	C	309	DC	O4'-C1'-N1	25.79	126.05	108.00
1	C	310	DA	O4'-C4'-C3'	-20.02	93.99	106.00
2	D	322	DG	O4'-C1'-N9	18.81	121.17	108.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	327	DG	C1'-O4'-C4'	-18.54	91.56	110.10

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	301	DC	Sidechain
1	C	302	DA	Sidechain
1	C	306	DG	Sidechain
1	C	307	DG	Sidechain
1	C	309	DC	Sidechain

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	C	422	0	236	62	0
2	D	433	0	237	50	0
3	A	534	0	537	106	0
3	B	534	0	537	166	0
All	All	1923	0	1547	335	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 99.

The worst 5 of 335 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)
3:B:219:ILE:CD1	3:B:240:LYS:HE3	1.55	1.36
3:B:227:PRO:HG3	3:B:250:TYR:HE2	0.99	1.12
3:B:220:VAL:CG2	3:B:221:GLN:OE1	1.99	1.10
3:B:227:PRO:HG3	3:B:250:TYR:CE2	1.84	1.10
3:B:219:ILE:HG21	3:B:240:LYS:HG3	1.33	1.10

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
3	A	63/65 (97%)	23 (36%)	25 (40%)	15 (24%)	0 0
3	B	63/65 (97%)	35 (56%)	13 (21%)	15 (24%)	0 0
All	All	126/130 (97%)	58 (46%)	38 (30%)	30 (24%)	0 0

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	201	ARG
3	A	231	MET
3	A	236	SER
3	A	240	LYS
3	A	247	ALA

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
3	A	60/60 (100%)	34 (57%)	26 (43%)	0 0
3	B	60/60 (100%)	37 (62%)	23 (38%)	0 0
All	All	120/120 (100%)	71 (59%)	49 (41%)	0 0

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	200	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	214	LYS
3	B	205	ASN
3	B	210	ARG
3	B	217	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	216	ASN
3	B	217	ASN
3	B	258	ASN
3	A	258	ASN
3	A	203	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)

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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.