W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Nov 9, 2024 – 09:13 AM EST

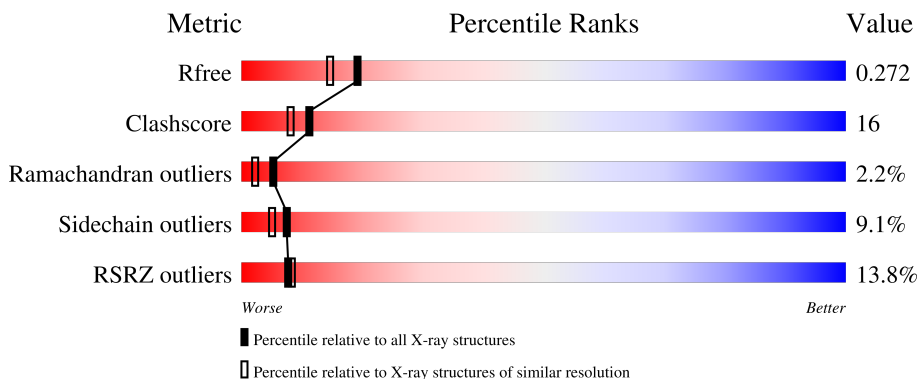|		
PDB ID	:	1AIS
Title	:	TATA-BINDING PROTEIN/TRANSCRIPTION FACTOR (II)B/TATA-BOX COMPLEX FROM PYROCOCCUS WOESEI
Authors	:	Kosa, P.F.; Ghosh, G.; Dedecker, B.S.; Sigler, P.B.
Deposited on	:	1997-04-24
Resolution	:	2.10 Å(reported)

---

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	Percentile Ranks	Value
Rfree		0.272
Clashscore		16
Ramachandran outliers		2.2%
Sidechain outliers		9.1%
RSRZ outliers		13.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	6234 (2.10-2.10)
Clashscore	180529	6893 (2.10-2.10)
Ramachandran outliers	177936	6839 (2.10-2.10)
Sidechain outliers	177891	6840 (2.10-2.10)
RSRZ outliers	164620	6234 (2.10-2.10)

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	C	17	35% 35% 29%
2	E	17	29% 65% 6%
3	A	182	2% / 81% 12% 5% ••
4	B	200	26% / 60% 31% • • •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3929 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(*AP*AP*CP*TP*TP*AP*CP*TP*TP*TP*(5IU)P*(5IU)P*AP*AP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	C	17	Total 342	C 165	I 2	N 58	O 101	P 16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	17	Total 349	C 169	N 65	O 99	P 16	0	0	0

- Molecule 3 is a protein called PROTEIN (TATA-BINDING PROTEIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	181	Total 1414	C 910	N 235	O 262	S 7	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	182	LEU	-	insertion	UNP P62001

- Molecule 4 is a protein called PROTEIN (TRANSCRIPTION INITIATION FACTOR IIB).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	193	Total 1536	C 972	N 285	O 276	S 3	0	0	0
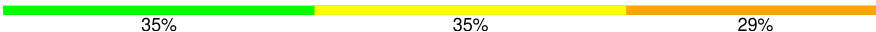
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	36	Total 36	O 36	1	0
5	E	45	Total 45	O 45	0	0
5	A	146	Total 146	O 146	0	0
5	B	61	Total 61	O 61	0	0

3 Residue-property plots ⓘ

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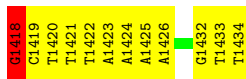
• Molecule 1: DNA (5'-D(*AP*AP*CP*TP*TP*AP*CP*TP*TP*TP*(5IU)P*(5IU)P*AP*AP* AP*GP*C)-3')
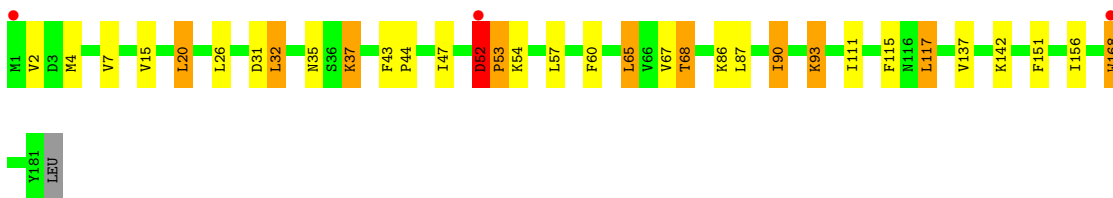
Chain C:

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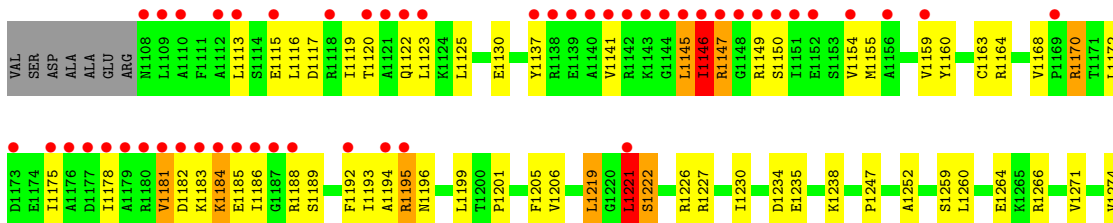
Chain E:

• Molecule 3: PROTEIN (TATA-BINDING PROTEIN)

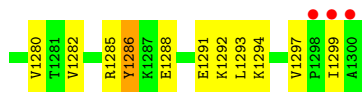Chain A:

• Molecule 4: PROTEIN (TRANSCRIPTION INITIATION FACTOR IIB)

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	125.70Å 91.20Å 74.20Å 90.00° 122.70° 90.00°	Depositor
Resolution (Å)	20.00 – 2.10 20.00 – 2.10	Depositor EDS
% Data completeness (in resolution range)	92.0 (20.00-2.10) 97.0 (20.00-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	4.10 (at 2.08Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, R_{free}	0.212 , 0.268 0.218 , 0.272	Depositor DCC
R_{free} test set	1986 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	31.4	Xtriage
Anisotropy	0.286	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 75.7	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3929	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.83% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 5IU

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	0.59	0/337	1.04	1/515 (0.2%)
2	E	0.60	0/392	1.03	1/604 (0.2%)
3	A	0.50	0/1437	0.75	2/1939 (0.1%)
4	B	0.43	0/1553	0.65	1/2086 (0.0%)
All	All	0.50	0/3719	0.79	5/5144 (0.1%)

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	2
2	E	0	3
All	All	0	5

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	52	ASP	C-N-CD	6.26	141.54	128.40
2	E	1418	DG	N9-C1'-C2'	-5.46	102.23	112.60
4	B	1221	LEU	N-CA-C	5.36	125.47	111.00
1	C	1401	DA	N9-C1'-C2'	-5.22	102.68	112.60
3	A	53	PRO	N-CA-C	-5.15	98.70	112.10

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	1407	DC	Sidechain
1	C	1414	DA	Sidechain
2	E	1418	DG	Sidechain
2	E	1424	DA	Sidechain
2	E	1425	DA	Sidechain

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	C	342	0	190	22	0
2	E	349	0	195	18	0
3	A	1414	0	1473	22	0
4	B	1536	0	1639	55	0
5	A	146	0	0	0	0
5	B	61	0	0	2	0
5	C	36	0	0	2	0
5	E	45	0	0	0	0
All	All	3929	0	3497	113	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 16.

The worst 5 of 113 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)
2:E:1418:DG:H2"	2:E:1419:DC:C5'	1.72	1.19
2:E:1418:DG:H2"	2:E:1419:DC:H5"	1.29	1.08
1:C:1403:DC:H2"	1:C:1404:DT:H5'	1.14	1.08
2:E:1418:DG:H5'	2:E:1418:DG:H8	1.29	0.97
2:E:1419:DC:H5'	2:E:1419:DC:H6	1.27	0.96

There are no symmetry-related clashes.

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

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	179/182 (98%)	176 (98%)	2 (1%)	1 (1%)	22	19
4	B	191/200 (96%)	173 (91%)	11 (6%)	7 (4%)	2	1
All	All	370/382 (97%)	349 (94%)	13 (4%)	8 (2%)	5	2

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	52	ASP
4	B	1146	ILE
4	B	1181	VAL
4	B	1221	LEU
4	B	1299	ILE

5.3.2   Protein sidechains ⓘ

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	159/160 (99%)	144 (91%)	15 (9%)	7	5
4	B	161/166 (97%)	147 (91%)	14 (9%)	8	6
All	All	320/326 (98%)	291 (91%)	29 (9%)	7	5

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	168	TRP

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type
4	B	1285	ARG
4	B	1146	ILE
4	B	1219	LEU
4	B	1145	LEU

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

Mol	Chain	Res	Type
4	B	1198	ASN
4	B	1213	ASN
4	B	1268	GLN
3	A	103	GLN
3	A	85	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 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	5IU	C	1411	1,2	18,21,22	0.56	0	25,30,33	1.05	1 (4%)
1	5IU	C	1412	1,2	18,21,22	0.79	1 (5%)	25,30,33	0.87	1 (4%)

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	5IU	C	1411	1,2	-	0/7/21/22	0/2/2/2
1	5IU	C	1412	1,2	-	<mark>3/7/21/22</mark>	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1412	5IU	C5-I5	2.88	2.16	2.08

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	1411	5IU	C4-C5-I5	2.90	123.30	118.54
1	C	1412	5IU	C4-C5-I5	2.30	122.31	118.54

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	1412	5IU	C2'-C1'-N1-C6
1	C	1412	5IU	C3'-C4'-C5'-O5'
1	C	1412	5IU	O4'-C1'-N1-C6

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	1411	5IU	2	0
1	C	1412	5IU	1	0

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

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, 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	C	15/17 (88%)	-0.56	0	100	100	16, 21, 42, 47	0
2	E	17/17 (100%)	-0.79	0	100	100	13, 21, 42, 44	0
3	A	181/182 (99%)	-0.29	3 (1%)	69	70	10, 22, 47, 70	0
4	B	193/200 (96%)	1.04	53 (27%)	2	2	14, 44, 95, 100	0
All	All	406/416 (97%)	0.31	56 (13%)	8	8	10, 29, 91, 100	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	1300	ALA	6.8
4	B	1299	ILE	5.7
4	B	1147	ARG	5.4
4	B	1141	VAL	5.3
4	B	1148	GLY	5.0

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

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	5IU	C	1411	20/21	0.97	0.05	11,17,28,42	1
1	5IU	C	1412	20/21	0.97	0.06	15,20,30,55	1

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.