W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 16, 2023 – 09:47 AM EDT

PDB ID	:	2IIE
Title	:	single chain Integration Host Factor protein (scIHF2) in complex with DNA
Authors	:	Bao, Q.; Droege, P.; Davey, C.A.
Deposited on	:	2006-09-28
Resolution	:	2.41 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

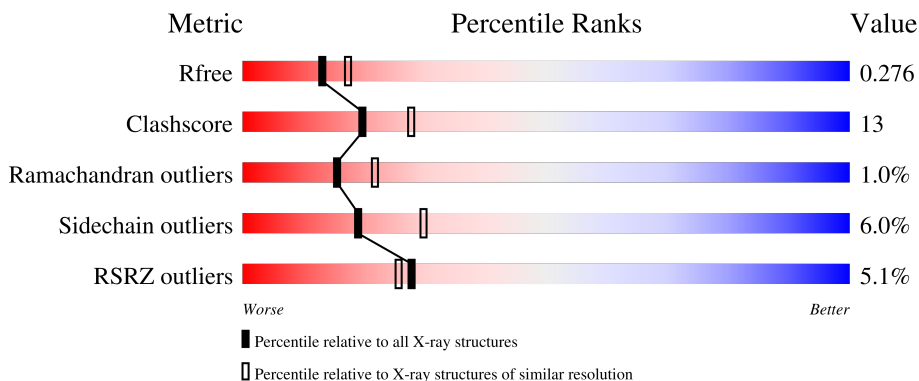|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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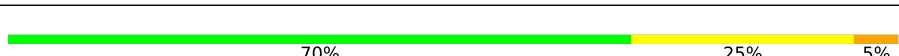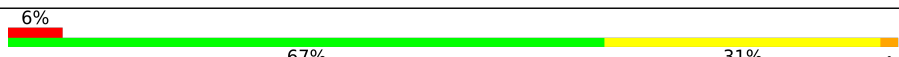*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.41 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		0.276
Clashscore		13
Ramachandran outliers		1.0%
Sidechain outliers		6.0%
RSRZ outliers		5.1%

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$_{free}$	130704	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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	35	60% 37% •
2	D	15	7% 60% 40%
3	E	20	70% 25% 5%
4	A	204	6% 67% 31% •

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 3318 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 Phage P H' site.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	35	Total	C	N	O	P	0	0	0
			717	344	130	209	34			

- Molecule 2 is a DNA chain called DNA (5'-D(*DGP*DGP*DCP*DCP*DAP*DAP*DAP* DAP*DAP*DAP*DGP*DCP*DAP*DTP*DT)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	15	Total	C	N	O	P	0	0	0
			307	147	63	83	14			

- Molecule 3 is a DNA chain called DNA (5'-D(*DGP*DCP*DTP*DTP*DAP*DTP*DCP*D AP*DAP*DTP*DTP*DTP*DGP*DTP*DTP*DGP*DCP*DAP*DCP*DC)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	20	Total	C	N	O	P	0	0	0
			402	195	66	122	19			

- Molecule 4 is a protein called Integration host factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	204	Total	C	N	O	S	0	0	0
			1610	998	308	300	4			

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP P0A6Y1
A	2	ALA	-	expression tag	UNP P0A6Y1
A	3	SER	-	expression tag	UNP P0A6Y1
A	42	GLY	-	linker	UNP P0A6Y1
A	43	GLY	-	linker	UNP P0A6Y1
A	44	SER	-	linker	UNP P0A6Y1

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	45	GLY	-	linker	UNP P0A6Y1
A	46	GLY	-	linker	UNP P0A6Y1
A	139	GLY	-	linker	UNP P0A6Y1
A	140	GLY	-	linker	UNP P0A6Y1
A	196	GLY	-	expression tag	UNP P0A6Y1
A	197	SER	-	expression tag	UNP P0A6Y1
A	198	GLY	-	expression tag	UNP P0A6Y1
A	199	HIS	-	expression tag	UNP P0A6Y1
A	200	HIS	-	expression tag	UNP P0A6Y1
A	201	HIS	-	expression tag	UNP P0A6Y1
A	202	HIS	-	expression tag	UNP P0A6Y1
A	203	HIS	-	expression tag	UNP P0A6Y1
A	204	HIS	-	expression tag	UNP P0A6Y1

- Molecule 5 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total 1	Mn 1	0	0
5	D	2	Total 2	Mn 2	0	0
5	E	1	Total 1	Mn 1	0	0
5	A	3	Total 3	Mn 3	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	58	Total 58	O 58	0	0
6	D	23	Total 23	O 23	0	0
6	E	45	Total 45	O 45	0	0
6	A	149	Total 149	O 149	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.
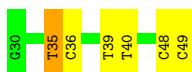
• Molecule 1: Phage P H' site

Chain C:

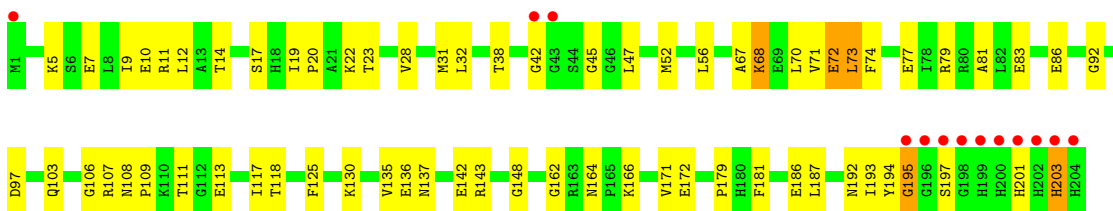• Molecule 2: DNA (5'-D(*DGP*DGP*DCP*DCP*DAP*DAP*DAP*DAP*DAP*DAP*DGP*DCP*DAP*DTP*DT)-3')

Chain D:

• Molecule 3: DNA (5'-D(*DGP*DCP*DTP*DTP*DAP*DTP*DCP*DAP*DAP*DTP*DTP*DTP*DGP*DTP*DTP*DGP*DCP*DAP*DCP*DC)-3')

Chain E:

• Molecule 4: Integration host factor

Chain A:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	47.37Å 55.31Å 177.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.41 40.44 – 2.41	Depositor EDS
% Data completeness (in resolution range)	98.7 (40.00-2.41) 98.7 (40.44-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 2.42Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.233 , 0.272 0.233 , 0.276	Depositor DCC
R_{free} test set	587 reflections (3.16%)	wwPDB-VP
Wilson B-factor (Å^2)	47.6	Xtriage
Anisotropy	0.677	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 47.9	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3318	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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 (i)

5.1 Standard geometry (i)

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

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	0.39	0/804	0.78	0/1240
2	D	0.36	0/346	0.71	0/532
3	E	0.42	0/448	0.77	0/689
4	A	0.35	0/1641	0.63	2/2194 (0.1%)
All	All	0.37	0/3239	0.70	2/4655 (0.0%)

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
3	E	0	1
All	All	0	3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	164	ASN	N-CA-C	-5.62	95.82	111.00
4	A	203	HIS	N-CA-C	5.62	126.17	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	-31	DG	Sidechain
1	C	-33	DA	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
3	E	35	DT	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	717	0	398	9	0
2	D	307	0	169	10	0
3	E	402	0	230	4	0
4	A	1610	0	1605	47	0
5	A	3	0	0	0	0
5	C	1	0	0	0	0
5	D	2	0	0	0	0
5	E	1	0	0	0	0
6	A	149	0	0	8	0
6	C	58	0	0	3	0
6	D	23	0	0	1	0
6	E	45	0	0	1	0
All	All	3318	0	2402	69	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 13.

The worst 5 of 69 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:D:17:DC:H2"	2:D:18:DC:H5'	1.51	0.93
4:A:79:ARG:HG2	4:A:187:LEU:HD13	1.64	0.80
4:A:5:LYS:HE3	4:A:32:LEU:HD12	1.62	0.79
1:C:-18:DG:H2"	1:C:-17:DG:OP2	1.88	0.74
4:A:136:GLU:HA	6:A:801:HOH:O	1.88	0.72

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	
4	A	202/204 (99%)	187 (93%)	13 (6%)	2 (1%)	15	22

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	193	ILE
4	A	195	GLY

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	
4	A	168/168 (100%)	158 (94%)	10 (6%)	19	30

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	166	LYS
4	A	201	HIS
4	A	203	HIS
4	A	72	GLU
4	A	73	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
4	A	129	GLN
4	A	137	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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 ⓘ

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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	35/35 (100%)	-0.05	0	100	100	35, 46, 78, 84	0
2	D	15/15 (100%)	0.04	1 (6%)	17	15	34, 45, 77, 77	0
3	E	20/20 (100%)	-0.27	0	100	100	29, 43, 58, 68	0
4	A	204/204 (100%)	0.46	13 (6%)	19	17	27, 45, 78, 93	0
All	All	274/274 (100%)	0.32	14 (5%)	28	26	27, 46, 78, 93	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	198	GLY	10.1
4	A	197	SER	8.1
4	A	202	HIS	8.1
4	A	203	HIS	7.1
4	A	199	HIS	6.8

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 ⓘ

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($\text{\AA}^2$)	Q<0.9
5	MN	A	306	1/1	0.77	0.21	87,87,87,87	1
5	MN	D	304	1/1	0.84	0.12	71,71,71,71	1
5	MN	C	307	1/1	0.89	0.10	60,60,60,60	1
5	MN	A	305	1/1	0.93	0.06	89,89,89,89	0
5	MN	E	302	1/1	0.97	0.14	50,50,50,50	0
5	MN	D	303	1/1	0.98	0.07	52,52,52,52	0
5	MN	A	301	1/1	0.99	0.10	37,37,37,37	0

6.5 Other polymers ⓘ

There are no such residues in this entry.