



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 06:40 am BST

PDB ID : 2H8R
Title : Hepatocyte Nuclear Factor 1b bound to DNA: MODY5 Gene Product
Authors : Lu, P.; Rha, G.B.; Chi, Y.I.
Deposited on : 2006-06-07
Resolution : 3.20 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

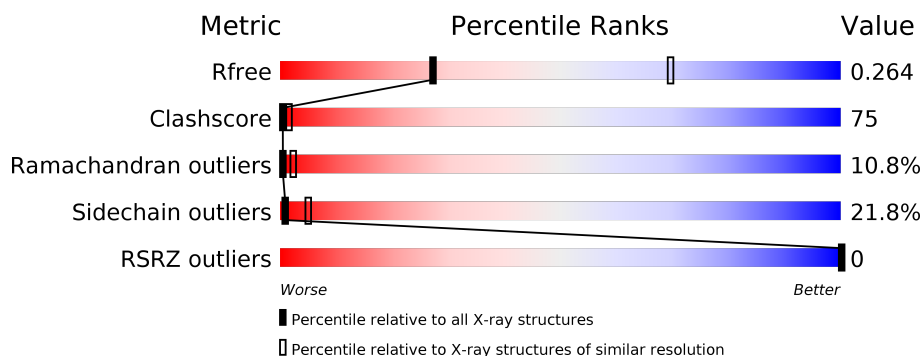
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 $\leq 5\%$. 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	E	20	<div> <div style="width: 15%; background-color: red;"></div> <div style="width: 85%; background-color: orange;"></div> </div> <div> <div style="width: 15%; background-color: yellow;"></div> <div style="width: 85%; background-color: orange;"></div> </div>
2	F	20	<div> <div style="width: 10%; background-color: yellow;"></div> <div style="width: 90%; background-color: orange;"></div> </div> <div> <div style="width: 10%; background-color: yellow;"></div> <div style="width: 90%; background-color: orange;"></div> </div>
3	A	221	<div> <div style="width: 8%; background-color: green;"></div> <div style="width: 20%; background-color: yellow;"></div> <div style="width: 38%; background-color: orange;"></div> <div style="width: 14%; background-color: red;"></div> <div style="width: 20%; background-color: grey;"></div> </div> <div> <div style="width: 8%; background-color: green;"></div> <div style="width: 20%; background-color: yellow;"></div> <div style="width: 38%; background-color: orange;"></div> <div style="width: 14%; background-color: red;"></div> <div style="width: 20%; background-color: grey;"></div> </div>
3	B	221	<div> <div style="width: 6%; background-color: green;"></div> <div style="width: 22%; background-color: yellow;"></div> <div style="width: 39%; background-color: orange;"></div> <div style="width: 13%; background-color: red;"></div> <div style="width: 20%; background-color: grey;"></div> </div> <div> <div style="width: 6%; background-color: green;"></div> <div style="width: 22%; background-color: yellow;"></div> <div style="width: 39%; background-color: orange;"></div> <div style="width: 13%; background-color: red;"></div> <div style="width: 20%; background-color: grey;"></div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3736 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 5'-D(*CP*TP*TP*GP*GP*TP*TP*AP*AP*TP*AP*A P*TP*TP*CP*AP*CP*CP*AP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	20	Total	C	N	O	P	0	0	0
			405	196	71	119	19			

- Molecule 2 is a DNA chain called 5'-D(*GP*CP*TP*GP*GP*TP*GP*AP*AP*TP*TP*A P*TP*TP*AP*AP*CP*CP*AP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	20	Total	C	N	O	P	0	0	0
			409	197	76	117	19			

- Molecule 3 is a protein called Hepatocyte nuclear factor 1-beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	176	Total	C	N	O	S	0	0	0
			1453	898	281	267	7			
3	B	176	Total	C	N	O	S	0	0	0
			1462	904	284	267	7			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	90	SER	-	CLONING ARTIFACT	UNP P35680
B	90	SER	-	CLONING ARTIFACT	UNP P35680

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	O	0	0
			1	1		
4	F	2	Total	O	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	O	0	0
			1	1		
4	B	3	Total	O	0	0
			3	3		

- Molecule 1: 5'-D(*CP*TP*TP*GP*GP*TP*TP*AP*AP*TP*AP*AP*TP*TP*CP*AP*CP*CP*AP*AP*G)-3'

C1	T2	T3	G4	G5	T6	T7	A8	A9	T10	A11	A12	T13	T14	C15	A16	C17	C18	A19	G20
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- Molecule 2: 5'-D(*GP*CP*TP*GP*GP*TP*GP*AP*AP*TP*TP*AP*TP*TP*AP*AP*CP*CP*AP*A)-3'

G2	C3	T4	G5	G6	T7	G8	A9	A10	T11	T12	A13	T14	T15	A16	A17	C18	C19	A20	A21
----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

- Molecule 3: Hepatocyte nuclear factor 1-beta

SER	R270	L450	S90
	A271	S151	I91
	E272	L451	L92
	C273	H153	K93
	L274	L154	E94
	Q275	L155	L95
	R276	K156	A96
	G277	G157	G97
	V278	L158	L98
	S279	GLN	S99
P280	ASP	T100	
S281	ASP	T100	
K282	ALA	K161	
A283	CYS	K162	
E284	SER	K163	
H285	SER	K164	
G286	GLU	R165	
L286	PRO	A166	
G287	THR	A167	
S288	ASN	L168	
N289	LYS	V169	
L290	LYS	T170	
V291	M231	M171	
T292	R232	M172	
E293	R233	M113	
V294	M234	L114	
R295	R235	K175	
V296	F236	K176	
V297	K237	R177	
L298	M238	E178	
N299	R239	I179	
F300	P240	L180	
A301	A241	R181	
R302	S242	R182	
R303	Q243	F183	
R304	Q244	M184	
K305	L245	K185	
E306	L246	G186	
E307	Y247	V187	
A308	Q248	GLN	
PHE	A249	SER	
ARG	Y250	SER	
	D251	GLY	
	R252	ASN	
	Q253	MTT	
	K254	THR	
	P255	ASP	
	P256	LYS	
	S257	SER	
	K258	SER	
	E259	GLN	
	E260	ASP	
	R261	GLN	
	E262	LEU	
	A263	LEU	
	L264	PHE	
	V265	LEU	
	E266	PHE	
	E267	PRO	
	C268	GLU	

- Molecule 3: Hepatocyte nuclear factor 1-beta

L450	S90
S151	S191
S152	S192
H153	S93
L154	S94
S155	
S156	S96
S157	A97
L158	S98
S159	S99
M160	T100
K161	E101
T162	E102
Q163	A103
K164	A104
S165	E105
A166	Q106
A167	A107
L168	A108
Y169	E109
T170	T110
M171	D111
Y172	R112
V173	M113
R174	L114
K175	S115
Q176	E116
R177	D117
E178	P118
I179	M119
L180	R120
L181	A121
Q182	A122
F183	K123
M184	M124
Q185	T125
THR	K126
VAL	G127
GLN	Y128
SER	M129
SER	Q130
GLY	Q131
ASN	H132
MET	M133
THR	L134
ASP	P135
LYS	Q136
SER	R137
SER	E138
GLN	V139
ASP	V140
GLN	D141
LEU	Y142
LEU	T143
PHE	G144
LEU	L145
PHE	M146
PRO	Q147
GLU	S148



4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	174.69 Å 174.69 Å 72.43 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.20 27.92 – 3.01	Depositor EDS
% Data completeness (in resolution range)	93.3 (20.00-3.20) 83.8 (27.92-3.01)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.09 (at 3.00 Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.223 , 0.290 0.195 , 0.264	Depositor DCC
R_{free} test set	691 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	79.5	Xtriage
Anisotropy	0.573	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.24 , 20.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.478 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3736	wwPDB-VP
Average B, all atoms (Å ²)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.25% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	E	6.56	188/453 (41.5%)	7.51	278/697 (39.9%)
2	F	6.32	183/459 (39.9%)	7.05	270/707 (38.2%)
3	A	3.51	211/1478 (14.3%)	2.43	81/1986 (4.1%)
3	B	3.73	232/1488 (15.6%)	2.41	90/1996 (4.5%)
All	All	4.45	814/3878 (21.0%)	4.26	719/5386 (13.3%)

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
3	A	0	2

The worst 5 of 814 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	19	DA	N7-C5	23.41	1.53	1.39
2	F	2	DG	C5-C4	22.38	1.54	1.38
2	F	2	DG	N1-C2	21.66	1.55	1.37
1	E	19	DA	C5-C4	21.43	1.53	1.38
1	E	19	DA	N9-C4	20.63	1.50	1.37

The worst 5 of 719 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	9	DA	N9-C4-C5	31.43	118.37	105.80
2	F	10	DA	C5-N7-C8	30.70	119.25	103.90
1	E	9	DA	C6-N1-C2	-27.61	102.04	118.60
1	E	17	DC	N3-C4-C5	-26.25	111.40	121.90
1	E	4	DG	C2-N3-C4	-25.37	99.22	111.90

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	183	PHE	Peptide
3	A	286	LEU	Peptide

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	E	405	0	228	41	0
2	F	409	0	224	43	0
3	A	1453	0	1444	217	0
3	B	1462	0	1450	243	0
4	A	1	0	0	0	0
4	B	3	0	0	0	0
4	E	1	0	0	0	0
4	F	2	0	0	0	0
All	All	3736	0	3346	526	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 75.

The worst 5 of 526 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:120:ARG:CB	3:B:120:ARG:CG	1.74	1.65
3:B:185:GLN:CB	3:B:185:GLN:CG	1.77	1.63
3:B:237:LYS:CD	3:B:237:LYS:CG	1.76	1.62
3:A:254:LYS:CG	3:A:254:LYS:CB	1.76	1.62
3:A:98:LEU:CD1	3:A:98:LEU:CG	1.75	1.61

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	172/221 (78%)	124 (72%)	33 (19%)	15 (9%)	1	4
3	B	172/221 (78%)	125 (73%)	25 (14%)	22 (13%)	0	1
All	All	344/442 (78%)	249 (72%)	58 (17%)	37 (11%)	0	2

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	141	ASP
3	A	142	VAL
3	A	173	VAL
3	A	234	ASN
3	A	262	GLU

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	154/195 (79%)	116 (75%)	38 (25%)	0	2
3	B	154/195 (79%)	125 (81%)	29 (19%)	1	8
All	All	308/390 (79%)	241 (78%)	67 (22%)	1	5

5 of 67 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	279	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	A	305	LYS
3	B	282	LYS
3	A	281	SER
3	A	302	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	243	GLN
3	A	253	GLN
3	B	176	GLN
3	A	182	GLN
3	A	185	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 carbohydrates 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	E	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	E	2:DT	O3'	3:DT	P	1.81
1	E	1:DC	O3'	2:DT	P	1.75

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	E	20/20 (100%)	-0.91	0 100 100	43, 65, 108, 112	0
2	F	20/20 (100%)	-1.00	0 100 100	48, 66, 109, 116	0
3	A	176/221 (79%)	-0.67	0 100 100	29, 65, 100, 108	0
3	B	176/221 (79%)	-0.73	0 100 100	36, 65, 99, 108	0
All	All	392/482 (81%)	-0.73	0 100 100	29, 65, 103, 116	0

There are no RSRZ outliers to report.

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.