



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 01:54 PM UTC

PDB ID : 2H8R / pdb_00002h8r
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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4-5-2 with Phenix2.0
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

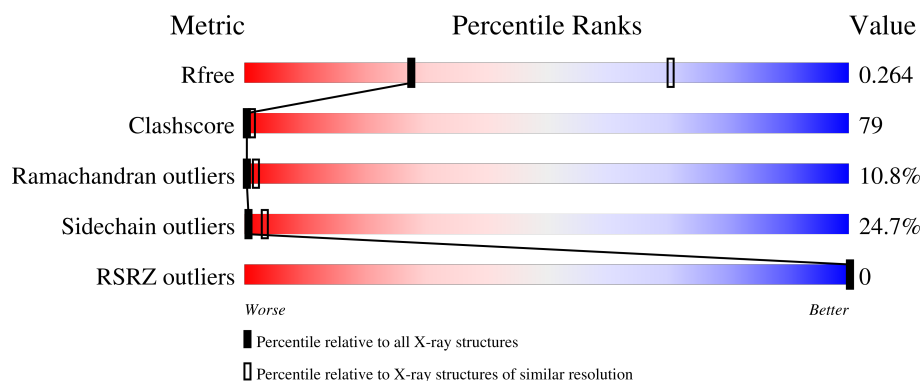
1 Overall quality at a glance

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}	180053	1466 (3.20-3.20)
Clashscore	190562	1573 (3.20-3.20)
Ramachandran outliers	187476	1548 (3.20-3.20)
Sidechain outliers	187428	1547 (3.20-3.20)
RSRZ outliers	180081	1466 (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 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 $\leq 5\%$. 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	15% 85%
2	F	20	10% 90%
3	A	221	19% 40% 17% 20%
3	B	221	17% 45% 16% 20%

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		

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.

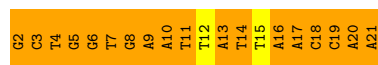
- Molecule 1: 5'-D(*CP*TP*TP*GP*GP*TP*TP*AP*AP*TP*AP*AP*TP*TP*CP*AP*CP*CP*AP*G)-3'

Chain E: 



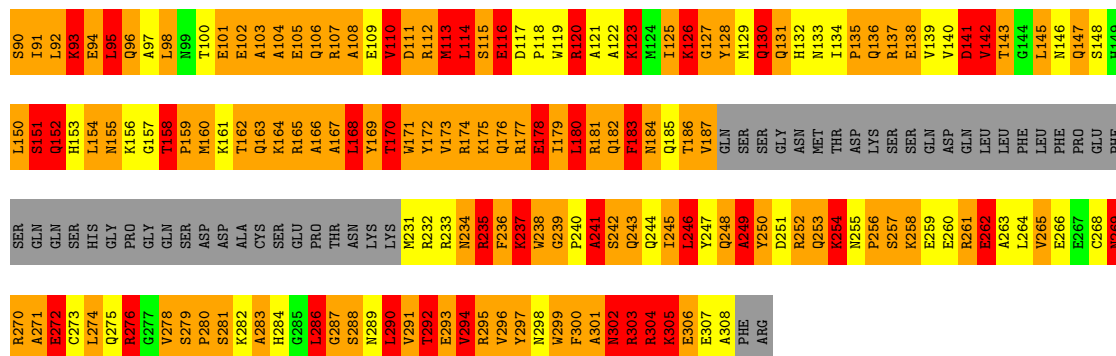
- Molecule 2: 5'-D(*GP*CP*TP*GP*GP*TP*GP*AP*AP*TP*TP*AP*TP*TP*AP*AP*CP*CP*AP*A)-3'

Chain F: 



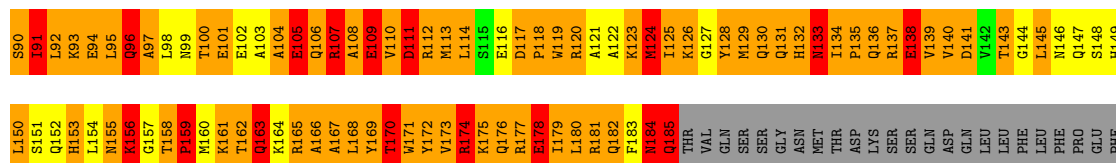
- Molecule 3: Hepatocyte nuclear factor 1-beta

Chain A: 



- Molecule 3: Hepatocyte nuclear factor 1-beta

Chain B: 



R270	R271	R272	R273	L274	Q275	R276	Q277	R278	S279	P280	ASP	ASP	ALA	CYS	SER	SER	GLU	PRO	THR	ASN	LVS	LVS	R231	R232	R233	R234	R235	R236	R237	R238	Q239	P240	A241	S242	Q243	Q244	L245	L246	Y247	Q248	Q249	A250	Y250	D251	R252	Q253	K254	N255	P256	S257	K258	E259	E260	R261	E262	A263	L264	V265	E266	E267	C268	N269
R270	A271	E272	C273	L274	Q275	R276	Q277	R278	S279	P280	ASP	ASP	ALA	CYS	SER	SER	GLU	PRO	THR	ASN	LVS	LVS	R231	R232	R233	R234	R235	R236	R237	R238	Q239	P240	A241	S242	Q243	Q244	L245	L246	Y247	Q248	Q249	A250	Y250	D251	R252	Q253	K254	N255	P256	S257	K258	E259	E260	R261	E262	A263	L264	V265	E266	E267	C268	N269

4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	174.69Å 174.69Å 72.43Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.20 20.00 – 3.20	Depositor EDS
% Data completeness (in resolution range)	93.3 (20.00-3.20) 93.2 (20.00-3.20)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.09 (at 3.00Å)	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 (4.23%)	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 , 60.5	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	0.98	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	3.28	56/453 (12.4%)	3.98	137/697 (19.7%)
2	F	3.14	48/459 (10.5%)	3.78	134/707 (19.0%)
3	A	3.95	286/1478 (19.4%)	2.98	181/1986 (9.1%)
3	B	4.20	303/1488 (20.4%)	2.88	155/1996 (7.8%)
All	All	3.89	693/3878 (17.9%)	3.20	607/5386 (11.3%)

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 693 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	134	ILE	CG1-CD1	22.58	2.39	1.51
3	A	249	ALA	CA-CB	-19.45	1.20	1.53
3	B	231	MET	SD-CE	17.49	2.23	1.79
3	B	108	ALA	CA-CB	-16.72	1.31	1.54
3	A	166	ALA	CA-CB	-16.72	1.22	1.53

The worst 5 of 607 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	281	SER	CA-C-N	-17.89	94.75	121.76
3	B	281	SER	C-N-CA	-17.89	94.75	121.76
1	E	16	DA	P-O5'-C5'	-15.30	97.05	120.00
3	A	168	LEU	N-CA-C	-14.98	94.56	113.43
3	A	181	ARG	N-CA-C	-14.29	95.25	113.12

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	230	0
3	B	1462	0	1450	258	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	554	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 79.

The worst 5 of 554 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:CG	3:B:120:ARG:CB	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%)	0	3
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%)	112 (73%)	42 (27%)	0	2
3	B	154/195 (79%)	120 (78%)	34 (22%)	1	5
All	All	308/390 (79%)	232 (75%)	76 (25%)	1	3

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	242	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	293	GLU
3	B	247	TYR
3	B	274	LEU
3	B	303	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	152	GLN
3	B	243	GLN
3	B	176	GLN
3	B	244	GLN
3	A	146	ASN

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 oligosaccharides 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%)	-1.78	0 100 100	43, 65, 108, 112	0
2	F	20/20 (100%)	-1.78	0 100 100	48, 66, 109, 116	0
3	A	176/221 (79%)	-1.53	0 100 100	29, 65, 100, 108	0
3	B	176/221 (79%)	-1.51	0 100 100	36, 65, 99, 108	0
All	All	392/482 (81%)	-1.55	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 oligosaccharides 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.