wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:27 AM GMT

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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

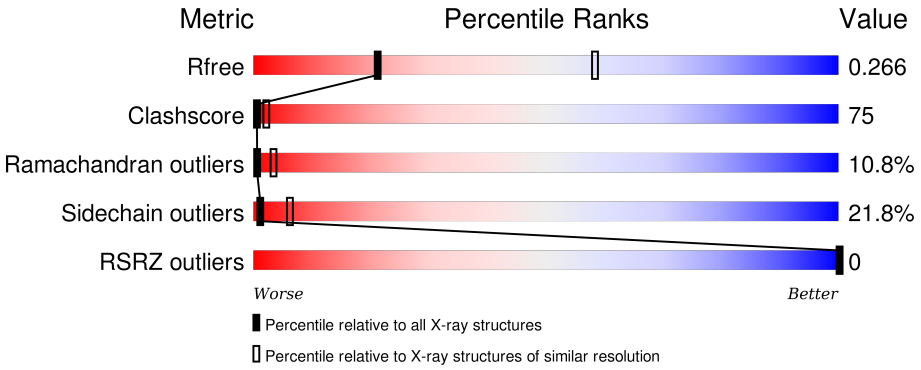
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 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}	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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. 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	E	20	15% 85%
2	F	20	10% 90%
3	A	221	8% 20% 38% 14% 20%
3	B	221	6% 22% 39% 13% 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*AP*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*AP*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	A	1	Total	O	0	0
			1	1		
4	B	3	Total	O	0	0
			3	3		

Continued on next page...

Continued from previous page...

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

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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'

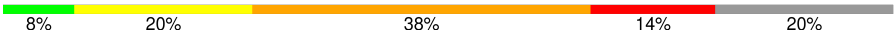
Chain E:

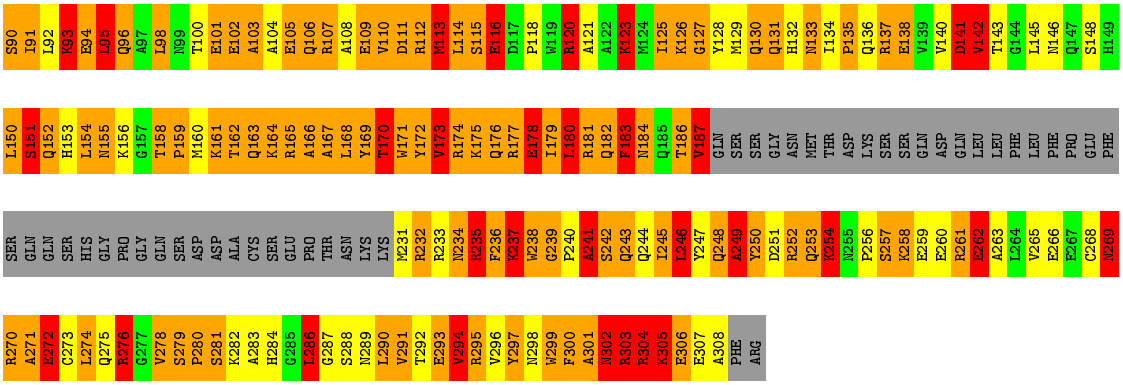
• Molecule 2: 5'-D(*GP*CP*TP*GP*GP*TP*GP*AP*AP*TP*TP*AP*TP*TP*AP*AP*CP*CP*AP*A)-3'

Chain F:

• Molecule 3: Hepatocyte nuclear factor 1-beta

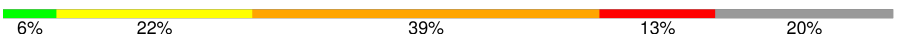
Chain A:

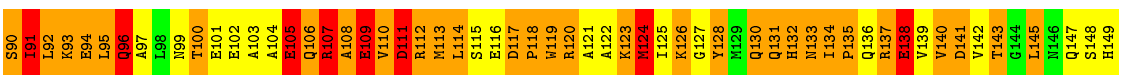


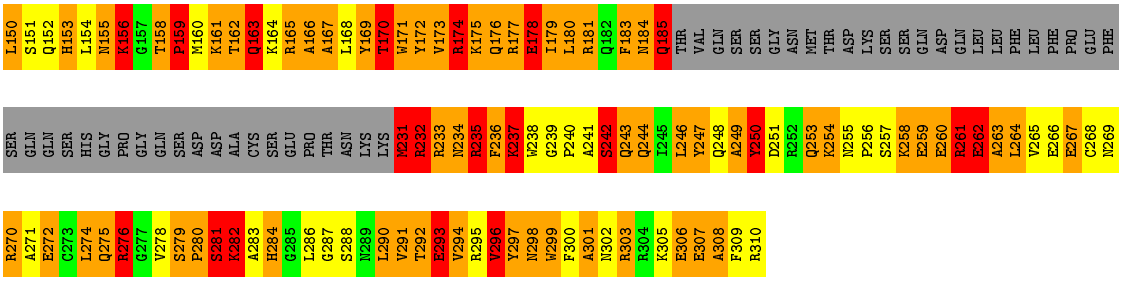


• Molecule 3: Hepatocyte nuclear factor 1-beta

Chain B:






L150 S151 Q152 H153 L154 N155 K156 G157 T158 P159 M160 K161 T162 Q163 K164 R165 A166 A167 L168 Y169 T170 W171 Y172 V173 R174 K175 Q176 R177 E178 I179 L180 R181 Q182 F183 N184 Q185 THR VAL GLN SER SER GLY ASN MET THR ASP LYS SER SER GLN ASP GLN LEU LEU PHE LEU PHE PRO GLU PHE
SER GLN GLN SER HIS GLY PRO GLY GLN SER ASP ASP ALA CYS SER GLU PRO THR ASN LYS LYS M231 R232 R233 N234 R235 F236 R237 W238 G239 P240 A241 S242 Q243 Q244 I245 L246 Y247 Q248 A249 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 G277 V278 S279 P280 S281 K282 A283 H284 G285 L286 G287 S288 N289 L290 V291 T292 E293 V294 R295 V296 Y297 N298 W299 F300 A301 N302 R303 R304 K305 E306 E307 A308 F309 R310

4 Data and refinement statistics (i)

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 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
$< I/\sigma(I) >$ [1]	2.09 (at 3.00Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.223 , 0.290 0.191 , 0.266	Depositor DCC
R_{free} test set	637 reflections (5.31%)	DCC
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
Estimated twinning fraction	0.478 for h,-h-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Outliers	0 of 13666 reflections	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.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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(^{o})	Ideal(^{o})
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 (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	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 5
3	B	172/221 (78%)	125 (73%)	25 (14%)	22 (13%)	0 2
All	All	344/442 (78%)	249 (72%)	58 (17%)	37 (11%)	0 3

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%)	1 3
3	B	154/195 (79%)	125 (81%)	29 (19%)	2 10
All	All	308/390 (79%)	241 (78%)	67 (22%)	1 6

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)

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, 95^{th} 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%)	-0.66	0	100	100	29, 65, 100, 108	0
3	B	176/221 (79%)	-0.72	0	100	100	36, 65, 99, 108	0
All	All	392/482 (81%)	-0.72	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

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

6.3 Carbohydrates

There are no carbohydrates 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.