



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:50 PM GMT

PDB ID : 3V79  
Title : Structure of human Notch1 transcription complex including CSL, RAM, ANK, and MAML-1 on HES-1 promoter DNA sequence  
Authors : Nam, Y.; Sliz, P.; Blacklow, S.  
Deposited on : 2011-12-20  
Resolution : 3.85 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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  
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

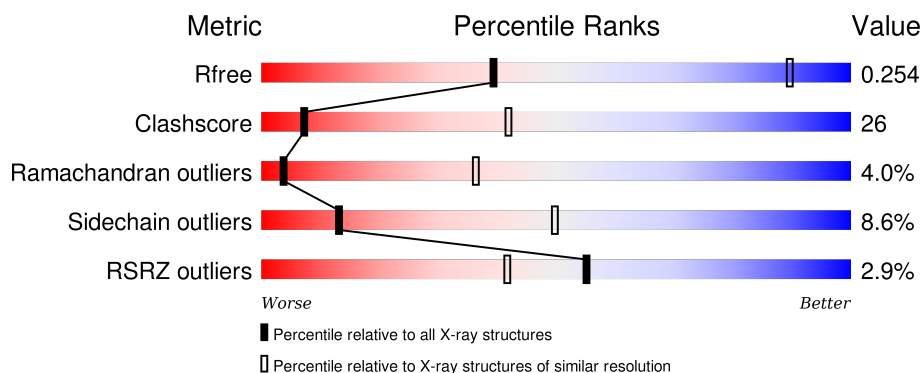
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.85 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	1000 (4.20-3.52)
Clashscore	102246	1090 (4.20-3.52)
Ramachandran outliers	100387	1046 (4.20-3.52)
Sidechain outliers	100360	1038 (4.20-3.52)
RSRZ outliers	91569	1004 (4.20-3.52)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	K	256	<div> <div>7%</div> <div> <div>45%</div> <div>37%</div> <div>•</div> <div>15%</div> </div> </div>
2	C	434	<div> <div>44%</div> <div>45%</div> <div>8%</div> <div>••</div> </div>
3	M	63	<div> <div>2%</div> <div> <div>54%</div> <div>33%</div> <div>13%</div> </div> </div>
4	X	18	<div> <div>39%</div> <div>61%</div> </div>
5	Y	18	<div> <div>22%</div> <div>78%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	R	19	<div><div></div><div></div><div></div><div></div></div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6421 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 protein called Neurogenic locus notch homolog protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	K	218	Total	C	N	O	S	0	0	0
			1683	1035	312	328	8			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	1872	GLY	-	EXPRESSION TAG	UNP P46531

- Molecule 2 is a protein called Recombining binding protein suppressor of hairless.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	424	Total	C	N	O	S	0	0	0
			3369	2133	580	631	25			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	8	MET	-	EXPRESSION TAG	UNP Q06330
C	436	HIS	-	EXPRESSION TAG	UNP Q06330
C	437	HIS	-	EXPRESSION TAG	UNP Q06330
C	438	HIS	-	EXPRESSION TAG	UNP Q06330
C	439	HIS	-	EXPRESSION TAG	UNP Q06330
C	440	HIS	-	EXPRESSION TAG	UNP Q06330
C	441	HIS	-	EXPRESSION TAG	UNP Q06330

- Molecule 3 is a protein called Mastermind-like protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	M	55	Total	C	N	O	S	0	0	0
			467	282	103	78	4			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	12	GLY	-	EXPRESSION TAG	UNP Q92585

- Molecule 4 is a DNA chain called DNA 5'-D(\*GP\*TP\*TP\*AP\*CP\*TP\*GP\*TP\*GP\*GP\*GP\*AP\*AP\*AP\*GP\*AP\*AP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	X	18	Total	C	N	O	P	0	0	0
			375	179	76	103	17			

- Molecule 5 is a DNA chain called DNA 5'-D(\*TP\*TP\*TP\*CP\*TP\*TP\*TP\*CP\*CP\*CP\*AP\*CP\*AP\*GP\*TP\*AP\*AP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	Y	18	Total	C	N	O	P	0	0	0
			357	174	57	109	17			

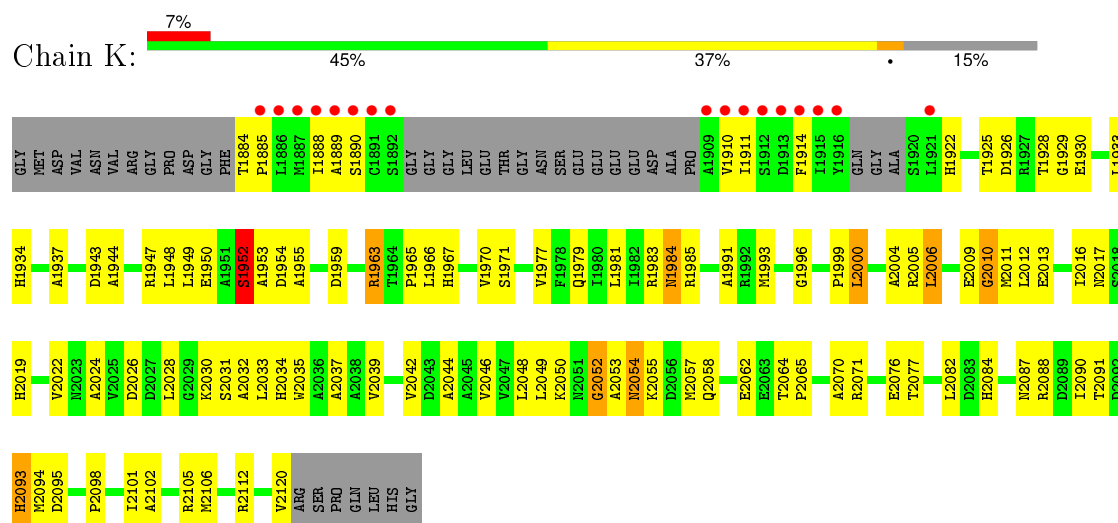
- Molecule 6 is a protein called RAM.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
6	R	19	Total	C	N	O	0	0	0
			170	108	35	27			

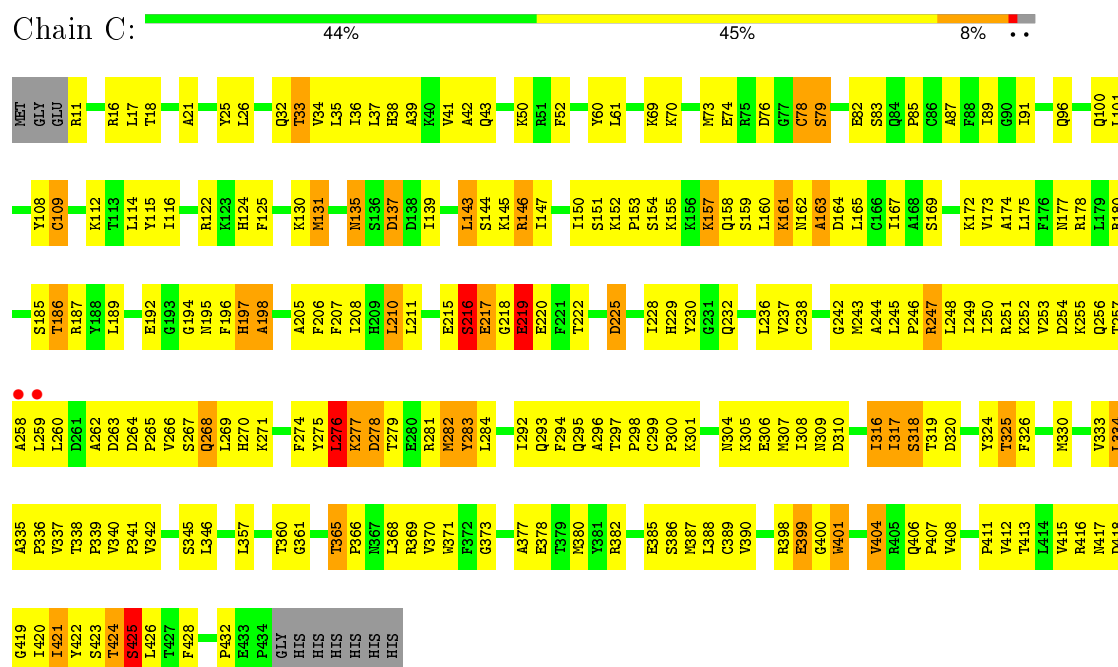
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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.

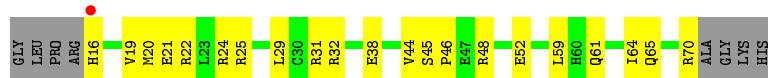
- Molecule 1: Neurogenic locus notch homolog protein 1



- Molecule 2: Recombining binding protein suppressor of hairless



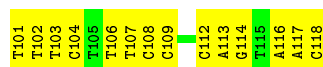
- Molecule 3: Mastermind-like protein 1



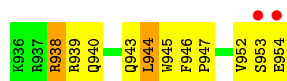
- Molecule 4: DNA 5'-D(\*GP\*TP\*TP\*AP\*CP\*TP\*GP\*TP\*GP\*GP\*GP\*AP\*AP\*AP\*GP\*AP\*AP\*A)-3'



- Molecule 5: DNA 5'-D(\*TP\*TP\*TP\*CP\*TP\*TP\*TP\*CP\*CP\*CP\*AP\*CP\*AP\*GP\*TP\*AP\*AP\*C)-3'



- Molecule 6: RAM



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	272.65Å 272.65Å 119.99Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 3.85 49.37 – 3.82	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-3.85) 92.3 (49.37-3.82)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.00 (at 3.77Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.213 , 0.245 0.222 , 0.254	Depositor DCC
$R_{free}$ test set	1204 reflections (5.14%)	DCC
Wilson B-factor (Å <sup>2</sup> )	96.0	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 69.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 43005 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	6421	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	124.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.61% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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	K	0.22	0/1705	0.43	0/2311
2	C	0.25	0/3444	0.47	0/4655
3	M	0.23	0/474	0.39	0/631
4	X	0.22	0/423	0.64	0/653
5	Y	0.21	0/397	0.67	0/609
6	R	0.33	0/175	0.48	0/230
All	All	0.24	0/6618	0.48	0/9089

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	K	1683	0	1650	85	0
2	C	3369	0	3348	206	0
3	M	467	0	472	17	0
4	X	375	0	204	11	0
5	Y	357	0	207	17	0
6	R	170	0	165	13	0
All	All	6421	0	6046	324	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 26.

The worst 5 of 324 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)
5:Y:112:DC:H2''	5:Y:113:DA:H5'	1.36	1.03
2:C:282:MET:HA	2:C:295:GLN:HA	1.42	0.97
2:C:333:VAL:HG21	2:C:420:ILE:HD11	1.54	0.89
4:X:3:DA:H2''	4:X:4:DC:H5''	1.54	0.87
5:Y:108:DC:H2''	5:Y:109:DC:H5'	1.59	0.84

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	
1	K	212/256 (83%)	173 (82%)	34 (16%)	5 (2%)	7	48
2	C	422/434 (97%)	324 (77%)	76 (18%)	22 (5%)	2	30
3	M	53/63 (84%)	53 (100%)	0	0	100	100
6	R	17/19 (90%)	13 (76%)	3 (18%)	1 (6%)	2	27
All	All	704/772 (91%)	563 (80%)	113 (16%)	28 (4%)	4	36

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	425	SER
6	R	947	PRO
1	K	2011	MET
1	K	2052	GLY
1	K	2054	ASN

### 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	
1	K	177/204 (87%)	169 (96%)	8 (4%)	34	71
2	C	374/382 (98%)	336 (90%)	38 (10%)	9	40
3	M	49/54 (91%)	46 (94%)	3 (6%)	23	62
6	R	17/17 (100%)	13 (76%)	4 (24%)	1	8
All	All	617/657 (94%)	564 (91%)	53 (9%)	13	50

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	216	SER
2	C	268	GLN
3	M	70	ARG
2	C	217	GLU
2	C	225	ASP

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

Mol	Chain	Res	Type
2	C	38	HIS
2	C	124	HIS
6	R	940	GLN
2	C	93	ASN
2	C	96	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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	K	218/256 (85%)	-0.17	17 (7%) 16 10	51, 89, 242, 266	0
2	C	424/434 (97%)	-0.21	2 (0%) 91 88	57, 118, 199, 237	1 (0%)
3	M	55/63 (87%)	-0.43	1 (1%) 71 61	56, 108, 171, 209	0
4	X	18/18 (100%)	-0.88	0 100 100	118, 145, 232, 235	0
5	Y	18/18 (100%)	-0.84	0 100 100	111, 146, 233, 245	0
6	R	19/19 (100%)	0.68	2 (10%) 8 6	163, 194, 243, 244	0
All	All	752/808 (93%)	-0.23	22 (2%) 55 42	51, 110, 211, 266	1 (0%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	1891	CYS	7.6
1	K	1890	SER	6.0
1	K	1912	SER	4.1
1	K	1909	ALA	3.9
2	C	258	ALA	3.8

### 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 carbohydrates in this entry.

### 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.