WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 3, 2023 – 11:24 AM EDT

PDB ID	:	6PY8
Title	:	Crystal structure of the RBPJ-NOTCH1-NRARP ternary complex bound to DNA
Authors	:	Jarrett, S.M.; Seegar, T.C.M.; Blacklow, S.C.
Deposited on	:	2019-07-29
Resolution	:	3.75 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	:	2.35.1
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)	:	2.35.1

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 3.75 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.
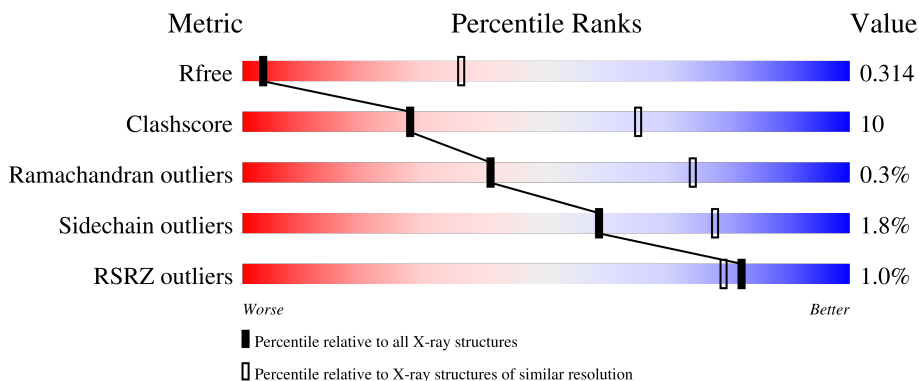
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1039 (3.94-3.58)
Clashscore	141614	1051 (3.92-3.60)
Ramachandran outliers	138981	1015 (3.92-3.60)
Sidechain outliers	138945	1011 (3.92-3.60)
RSRZ outliers	127900	1050 (3.96-3.56)

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	B	114	2% / 69% 13% · 17%
1	G	114	% / 63% 10% · 26%
2	A	16	38% 62%
2	X	16	38% 62%
3	D	16	31% 69%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Y	16	19% · 81%
4	C	444	74% · 22% ·
4	E	444	.% · 69% · 27% · ·
5	F	369	.% · 56% · 12% · · 31%
5	K	369	52% · 18% · 30%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 13305 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 Notch-regulated ankyrin repeat-containing protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	B	95	Total 642	C 406	N 117	O 119	0	0	0
1	G	84	Total 590	C 375	N 107	O 108	0	0	0

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	16	Total 333	C 159	N 66	O 93	P 15	0	0	0
2	X	16	Total 333	C 159	N 66	O 93	P 15	0	0	0

- Molecule 3 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	16	Total 317	C 154	N 53	O 95	P 15	0	0	0
3	Y	16	Total 317	C 154	N 53	O 95	P 15	0	0	0

- Molecule 4 is a protein called Recombining binding protein suppressor of hairless.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	429	Total 3408	C 2157	N 590	O 636	S 25	0	0	0
4	E	431	Total 3420	C 2163	N 592	O 639	S 26	0	0	0

- Molecule 5 is a protein called Neurogenic locus notch homolog protein 1.

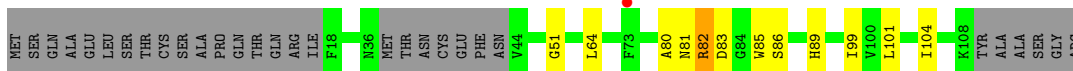| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	254	Total	C	N	O	S	0	0	0
			1957	1202	369	378	8			
5	K	258	Total	C	N	O	S	0	0	0
			1988	1220	373	387	8			

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: Notch-regulated ankyrin repeat-containing protein

Chain B:

- Molecule 1: Notch-regulated ankyrin repeat-containing protein

Chain G:

- Molecule 2: DNA

Chain A:

- Molecule 2: DNA

Chain X:

- Molecule 3: DNA

Chain D:

- Molecule 3: DNA

Chain Y:

A101 A102 T105 T106 T107 C108 C109 C110 A111 C112 A113 G114 T115 C116

• Molecule 4: Recombining binding protein suppressor of hairless

Chain C:

• Molecule 4: Recombining binding protein suppressor of hairless

Chain E:

• Molecule 5: Neurogenic locus notch homolog protein 1

Chain F:

THR GLY ASN SER GLU **E1904 E1905** N1923 D1943 D1959 R1963 H1967 S1971 F1978 R1992 D1995 G1996 M2011 L2012 E2013 V2025 D2026 D2027 L2028 G2029 K2030 A2037 V2047 N2054 M2057 E2062 E2063 T2064 F2067 L2068 A2069 E2076 T2077 N2087 R2088

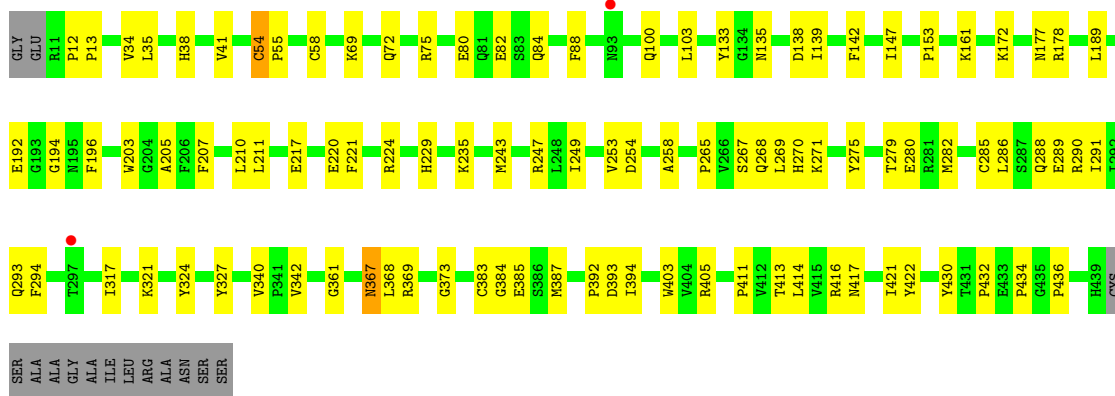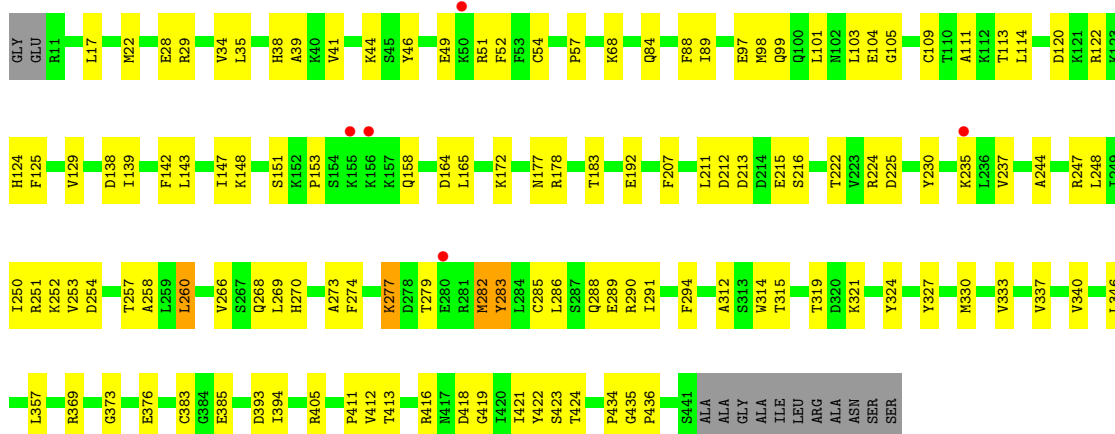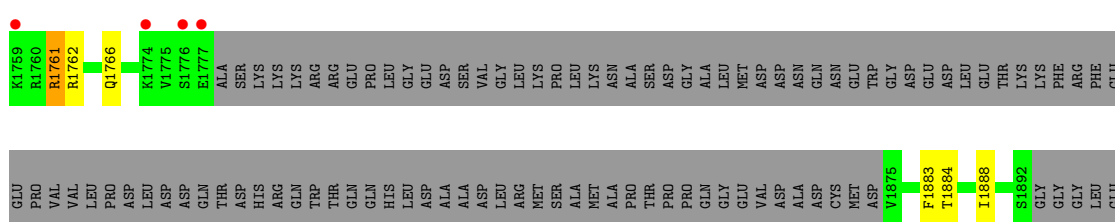D2089 I2090 T2091 D2092 H2093 M2094 D2095 R2096 H2107 H2108 D2109 I2110 V2111 R2112 Y2117 N2118 L2119 V2120 ARG SER PRO GLN LEU HIS GLY ALA

• Molecule 5: Neurogenic locus notch homolog protein 1

Chain K: 52%   18%   30%

K1759 R1760 R1761 R1762 P1770 V1775 S1776 E1777 ALA SER LYS LYS LYS ARG ARG GLU PRO LEU GLY ASP SER VAL GLY LEU LYS PRO LEU LYS ASN ALA SER ASP GLY ALA LEU MET ASP ASP ASN GLN ASN GLU TRP GLY ASP ASP GLU ASP LEU GLU THR LYS LYS PHE ARG PHE GLU GLU

PRO VAL VAL LEU PRO ASP LEU ASP ASP GLN THR ASP ASP HIS ARG GLN GLN TRP THR GLN GLN HIS LEU ASP ALA ALA ASP LEU ARG MET SER ALA MET ALA PRO THR PRO PRO GLN GLY GLU VAL ASP ALA ASP CYS MET ASP V1875 N1876 V1877 R1878 T1884 P1885 L1886 M1887 G1893 GLY GLY LEU

GLU THR GLY ASN S1901 E1905 D1906 A1907 P1908 A1909 E1930 H1934 L1935 A1936 A1937 S1942 D1943 A1944 R1947 L1948 L1949 Q1958 D1959 G1962 R1963 L1966 V1977 F1978 Q1979 I1980 L1981 I1982 R1992 D1995 M2011 V2025 D2026 D2027 K2030 A2037 N2054 M2057 Q2058

N2059 N2060 R2061 E2062 L2068 R2071 E2072 Y2075 E2076 T2077 R2088 T2091 D2092 H2093 M2094 D2095 R2096 H2107 H2108 D2109 I2110 V2111 R2112 L2113 L2114 D2115 E2116 V2120 ARG SER PRO GLN LEU HIS GLY ALA

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	79.88Å 103.65Å 301.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.46 – 3.75 49.01 – 3.75	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.46-3.75) 93.5 (49.01-3.75)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 3.77Å)	Xtriage
Refinement program	PHENIX (1.15.2_3472: ???)	Depositor
R, R_{free}	0.267 , 0.315 0.266 , 0.314	Depositor DCC
R_{free} test set	2000 reflections (7.61%)	wwPDB-VP
Wilson B-factor (Å^2)	103.6	Xtriage
Anisotropy	0.692	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 39.9	EDS
L-test for twinning [2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	13305	wwPDB-VP
Average B, all atoms (Å^2)	119.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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 ⓘ

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	B	0.25	0/651	0.43	0/888
1	G	0.24	0/598	0.41	0/810
2	A	0.48	0/375	0.86	0/579
2	X	0.56	0/375	0.88	0/579
3	D	0.52	0/353	0.94	0/541
3	Y	0.56	0/353	0.98	0/541
4	C	0.25	0/3486	0.46	0/4713
4	E	0.25	0/3498	0.47	0/4729
5	F	0.23	0/1987	0.41	0/2691
5	K	0.23	0/2018	0.42	0/2732
All	All	0.29	0/13694	0.52	0/18803

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 ⓘ

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	B	642	0	558	13	0
1	G	590	0	545	11	0
2	A	333	0	182	10	0
2	X	333	0	182	9	0
3	D	317	0	183	7	0
3	Y	317	0	183	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	3408	0	3385	67	0
4	E	3420	0	3395	78	0
5	F	1957	0	1886	33	0
5	K	1988	0	1915	45	0
All	All	13305	0	12414	257	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 10.

The worst 5 of 257 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:F:2091:THR:HB	5:F:2095:ASP:HA	1.65	0.79
4:C:282:MET:HA	4:C:294:PHE:O	1.85	0.76
4:E:279:THR:HG22	4:E:282:MET:HB2	1.71	0.72
5:K:2037:ALA:HA	5:K:2077:THR:HG21	1.70	0.72
5:K:1958:GLN:HB3	5:K:1962:GLY:HA2	1.73	0.70

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	B	93/114 (82%)	86 (92%)	6 (6%)	1 (1%)	14	51
1	G	80/114 (70%)	72 (90%)	8 (10%)	0	100	100
4	C	427/444 (96%)	393 (92%)	34 (8%)	0	100	100
4	E	429/444 (97%)	391 (91%)	34 (8%)	4 (1%)	17	54
5	F	248/369 (67%)	233 (94%)	15 (6%)	0	100	100
5	K	252/369 (68%)	237 (94%)	15 (6%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1529/1854 (82%)	1412 (92%)	112 (7%)	5 (0%)	41 74

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	97	GLU
4	E	436	PRO
1	B	44	VAL
4	E	282	MET
4	E	288	GLN

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	
1	B	50/93 (54%)	50 (100%)	0	100	100
1	G	51/93 (55%)	50 (98%)	1 (2%)	55	75
4	C	378/387 (98%)	372 (98%)	6 (2%)	62	80
4	E	380/387 (98%)	374 (98%)	6 (2%)	62	80
5	F	200/302 (66%)	194 (97%)	6 (3%)	41	66
5	K	204/302 (68%)	200 (98%)	4 (2%)	55	75
All	All	1263/1564 (81%)	1240 (98%)	23 (2%)	59	78

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	1762	ARG
5	F	2028	LEU
5	F	2027	ASP
5	F	2112	ARG
4	C	369	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
4	C	124	HIS
4	C	367	ASN
4	E	349	ASN
5	K	2093	HIS

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 ⓘ

There are no ligands in this entry.

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(Å²)	Q<0.9
1	B	95/114 (83%)	-0.01	2 (2%)	63	58	92, 129, 228, 254	0
1	G	84/114 (73%)	0.15	1 (1%)	79	75	98, 131, 151, 159	0
2	A	16/16 (100%)	-0.57	0	100	100	125, 148, 165, 165	0
2	X	16/16 (100%)	-0.48	0	100	100	109, 146, 165, 166	0
3	D	16/16 (100%)	-0.65	0	100	100	116, 145, 162, 163	0
3	Y	16/16 (100%)	-0.78	0	100	100	107, 134, 168, 177	0
4	C	429/444 (96%)	-0.09	2 (0%)	91	89	69, 105, 148, 175	1 (0%)
4	E	431/444 (97%)	0.01	5 (1%)	79	75	65, 117, 163, 192	1 (0%)
5	F	254/369 (68%)	-0.04	5 (1%)	65	60	79, 107, 160, 204	0
5	K	258/369 (69%)	-0.12	1 (0%)	92	91	80, 111, 141, 162	0
All	All	1615/1918 (84%)	-0.06	16 (0%)	82	79	65, 114, 162, 254	2 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	E	156	LYS	3.8
5	F	1776	SER	3.5
5	F	1759	LYS	2.7
4	E	155	LYS	2.5
1	G	73	PHE	2.5

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.