W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 11:19 am BST

PDB ID	:	1JX7
Title	:	Crystal structure of ychN protein from E.coli
Authors	:	Shin, D.H.; Yokota, H.; Kim, R.; Kim, S.-H.; Berkeley Structural Genomics Center (BSGC)
Deposited on	:	2001-09-05
Resolution	:	2.80 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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 2.80 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.
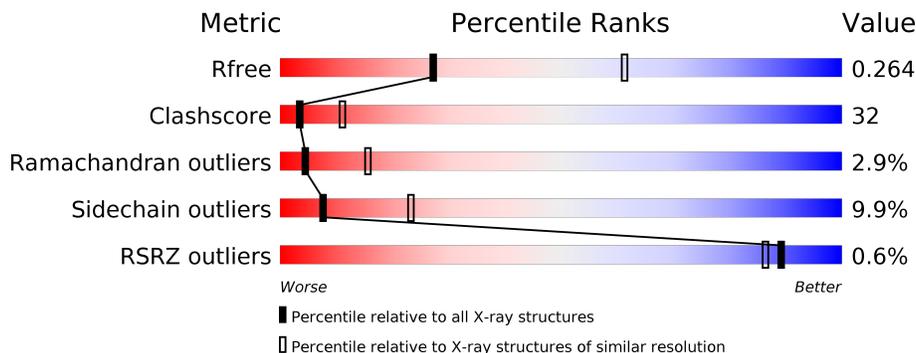
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 <=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	A	117	44% 43% 14%
1	B	117	2% 47% 46% 7%
1	C	117	49% 44% 8%
1	D	117	% 52% 42% 5% ·
1	E	117	50% 42% 9%
1	F	117	% 54% 39% 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5420 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 HYPOTHETICAL PROTEIN YCHN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			
1	B	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			
1	C	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			
1	D	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			
1	E	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			
1	F	117	Total	C	N	O	S	Se	0	0	0
			889	563	150	171	2	3			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	CLONING ARTIFACT	UNP P0AB52
A	42	MSE	MET	CLONING ARTIFACT	UNP P0AB52
A	64	MSE	MET	CLONING ARTIFACT	UNP P0AB52
B	201	MSE	MET	CLONING ARTIFACT	UNP P0AB52
B	242	MSE	MET	CLONING ARTIFACT	UNP P0AB52
B	264	MSE	MET	CLONING ARTIFACT	UNP P0AB52
C	401	MSE	MET	CLONING ARTIFACT	UNP P0AB52
C	442	MSE	MET	CLONING ARTIFACT	UNP P0AB52
C	464	MSE	MET	CLONING ARTIFACT	UNP P0AB52
D	601	MSE	MET	CLONING ARTIFACT	UNP P0AB52
D	642	MSE	MET	CLONING ARTIFACT	UNP P0AB52
D	664	MSE	MET	CLONING ARTIFACT	UNP P0AB52
E	801	MSE	MET	CLONING ARTIFACT	UNP P0AB52
E	842	MSE	MET	CLONING ARTIFACT	UNP P0AB52
E	864	MSE	MET	CLONING ARTIFACT	UNP P0AB52
F	1001	MSE	MET	CLONING ARTIFACT	UNP P0AB52
F	1042	MSE	MET	CLONING ARTIFACT	UNP P0AB52

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	1064	MSE	MET	CLONING ARTIFACT	UNP P0AB52

- Molecule 2 is water.

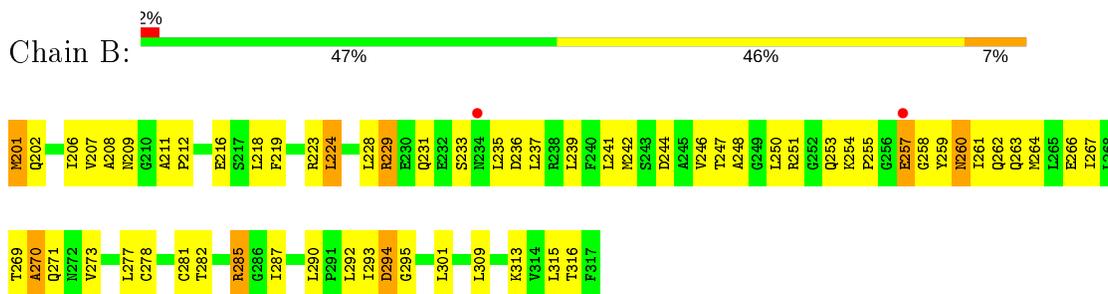| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	16	Total 16	O 16	0	0
2	B	14	Total 14	O 14	0	0
2	C	10	Total 10	O 10	0	0
2	D	13	Total 13	O 13	0	0
2	E	21	Total 21	O 21	0	0
2	F	12	Total 12	O 12	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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.
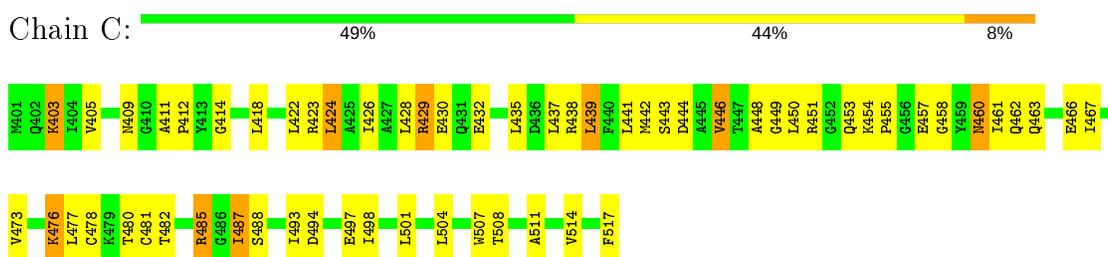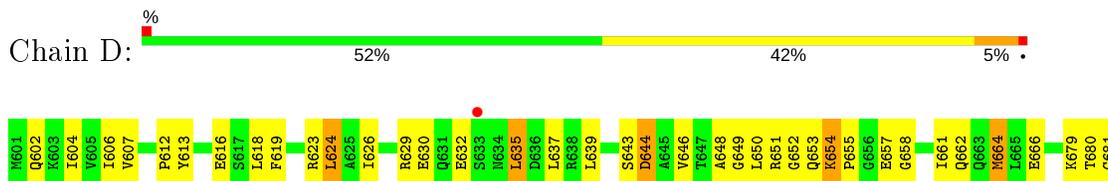
- Molecule 1: HYPOTHETICAL PROTEIN YCHN

- Molecule 1: HYPOTHETICAL PROTEIN YCHN

- Molecule 1: HYPOTHETICAL PROTEIN YCHN

- Molecule 1: HYPOTHETICAL PROTEIN YCHN

T682 D683 G684 R685 G686 I687 S688 I693 D694 I698 G699 V702 E703 L704 A705 Q706 W707 T708 A711 D712 K713 V714 L715 T716 F717

● Molecule 1: HYPOTHETICAL PROTEIN YCHN

Chain E:

50% 42% 9%

M801 Q802 K803 I806 N809 G810 A811 P812 L818 F819 L822 R823 L824 A825 I826 R829 E830 N834 R838 L839 F840 L841 M842 S843 D844 A845 V846 T847 A848 G849 L850 R851 G852 Q853 E857 G858 Y859 N860 I861 Q862 Q863 M864 L865 E866 I867 L868 T869 A870 Q871 N872 V873

K876 L877 C878 K879 T882 D883 G884 R885 G886 I887 L890 I893 D894 L904 W907 T908 K913 V914 L915 T916 F917

● Molecule 1: HYPOTHETICAL PROTEIN YCHN

Chain F:

% 54% 39% 7%

M1001 Q1002 K1003 N1009 P1012 L1018 F1019 R1023 R1029 L1037 R1038 L1039 F1040 S1043 D1044 A1045 V1046 T1047 A1048 G1049 Q1053 K1054 P1055 G1056 E1057 G1058 Y1059 N1060 I1061 Q1062 Q1063 M1064 Q1071 N1072 V1073 P1074 V1075 K1076 C1081 T1082 L1090 P1091 L1092 I1093 D1094 G1095 I1098

L1101 V1102 E1103 L1104 A1105 Q1106 W1107 T1108 L1109 S1110 A1111 D1112 K1113 V1114 L1115 T1116 F1117

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	66.21Å 80.46Å 140.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 2.80 69.78 – 2.69	Depositor EDS
% Data completeness (in resolution range)	95.4 (19.98-2.80) 94.8 (69.78-2.69)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.68 (at 2.69Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.229 , 0.265 0.229 , 0.264	Depositor DCC
R_{free} test set	2009 reflections (9.83%)	wwPDB-VP
Wilson B-factor (Å^2)	29.1	Xtriage
Anisotropy	0.085	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 36.0	EDS
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	5420	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.31% 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	A	0.53	1/897 (0.1%)	0.70	0/1210
1	B	0.52	1/897 (0.1%)	0.66	0/1210
1	C	0.49	0/897	0.63	0/1210
1	D	0.52	1/897 (0.1%)	0.65	0/1210
1	E	0.50	0/897	0.71	0/1210
1	F	0.49	0/897	0.74	1/1210 (0.1%)
All	All	0.51	3/5382 (0.1%)	0.68	1/7260 (0.0%)

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
1	F	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	664	MSE	SE-CE	-5.32	1.64	1.95
1	B	201	MSE	CG-SE	-5.21	1.77	1.95
1	A	64	MSE	SE-CE	-5.09	1.65	1.95

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	1095	GLY	N-CA-C	-9.34	89.74	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	1095	GLY	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	A	889	0	924	77	0
1	B	889	0	921	70	0
1	C	889	0	921	63	0
1	D	889	0	921	58	0
1	E	889	0	921	59	0
1	F	889	0	921	60	0
2	A	16	0	0	1	0
2	B	14	0	0	1	0
2	C	10	0	0	1	0
2	D	13	0	0	3	0
2	E	21	0	0	2	0
2	F	12	0	0	2	0
All	All	5420	0	5529	349	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 32.

The worst 5 of 349 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)
1:F:1117:PHE:OXT	2:F:2068:HOH:O	1.68	1.12
1:F:1060:ASN:ND2	1:F:1063:GLN:H	1.50	1.07
1:B:246:VAL:HG11	1:B:282:THR:HG22	1.37	1.05
1:B:294:ASP:O	2:B:2067:HOH:O	1.77	1.01
1:B:246:VAL:CG1	1:B:282:THR:HG22	1.99	0.92

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	A	115/117 (98%)	105 (91%)	4 (4%)	6 (5%)	2	6
1	B	115/117 (98%)	106 (92%)	5 (4%)	4 (4%)	3	12
1	C	115/117 (98%)	106 (92%)	8 (7%)	1 (1%)	17	46
1	D	115/117 (98%)	109 (95%)	4 (4%)	2 (2%)	9	29
1	E	115/117 (98%)	101 (88%)	11 (10%)	3 (3%)	5	18
1	F	115/117 (98%)	105 (91%)	6 (5%)	4 (4%)	3	12
All	All	690/702 (98%)	632 (92%)	38 (6%)	20 (3%)	4	15

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	94	ASP
1	B	257	GLU
1	E	802	GLN
1	E	870	ALA
1	E	894	ASP

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	A	98/95 (103%)	85 (87%)	13 (13%)	4	12
1	B	98/95 (103%)	91 (93%)	7 (7%)	14	39
1	C	98/95 (103%)	87 (89%)	11 (11%)	6	18

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	98/95 (103%)	90 (92%)	8 (8%)	11	33
1	E	98/95 (103%)	86 (88%)	12 (12%)	5	15
1	F	98/95 (103%)	91 (93%)	7 (7%)	14	39
All	All	588/570 (103%)	530 (90%)	58 (10%)	8	23

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	460	ASN
1	D	629	ARG
1	F	1039	LEU
1	C	476	LYS
1	C	487	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	463	GLN
1	D	662	GLN
1	F	1020	ASN
1	C	460	ASN
1	F	1031	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 (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	A	114/117 (97%)	-0.23	0	100	100	9, 24, 40, 64	0
1	B	114/117 (97%)	-0.01	2 (1%)	68	61	10, 25, 52, 86	0
1	C	114/117 (97%)	-0.32	0	100	100	12, 24, 55, 71	0
1	D	114/117 (97%)	-0.30	1 (0%)	84	80	9, 24, 48, 85	0
1	E	114/117 (97%)	-0.17	0	100	100	9, 24, 49, 58	0
1	F	114/117 (97%)	-0.02	1 (0%)	84	80	13, 25, 45, 68	0
All	All	684/702 (97%)	-0.17	4 (0%)	89	86	9, 24, 49, 86	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	633	SER	4.7
1	B	257	GLU	4.3
1	F	1094	ASP	3.2
1	B	234	ASN	2.0

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.