



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 31, 2017 – 09:45 PM EDT

PDB ID : 5KIS  
Title : YenB/RHS2 complex  
Authors : Busby, J.N.; Hurst, M.R.H.; Lott, J.S.  
Deposited on : unknown  
Resolution : 2.40 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20029824

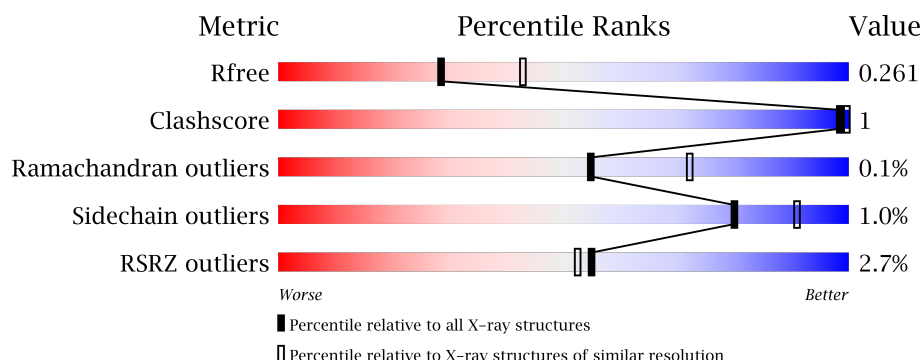
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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))
$R_{free}$	100719	3166 (2.40-2.40)
Clashscore	112137	3674 (2.40-2.40)
Ramachandran outliers	110173	3616 (2.40-2.40)
Sidechain outliers	110143	3617 (2.40-2.40)
RSRZ outliers	101464	3195 (2.40-2.40)

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	A	1491	
2	B	965	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 16842 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 YenB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1446	Total	C	N	O	S	0	7	0
			11438	7236	1966	2199	37			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP B6A880
A	-2	SER	-	expression tag	UNP B6A880
A	-1	GLY	-	expression tag	UNP B6A880
A	0	ALA	-	expression tag	UNP B6A880

- Molecule 2 is a protein called RHS2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	675	Total	C	N	O	S	0	0	0
			5177	3211	913	1048	5			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Ca	0	0
			2	2		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Na	0	0
			2	2		

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total 2	Cl 2	0	0
5	A	2	Total 2	Cl 2	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	161	Total 161	O 161	0	0
6	B	58	Total 58	O 58	0	0

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.

Chain A:

3%

94%

• Good

• Bad

Residue	Category
GLY	Good
SER	Good
GLY	Good
ALA	Good
MET	Good
GLN	Good
ASN	Good
SER	Good
GLN	Good
GLU	Good
M7	Good
A8	Good
G18	Good
GLY	Good
A20	Good
I21	Good
G25	Good
E26	Good
S27	Good
N109	Good
D110	Good
S111	Good
K146	Good
Y165	Good
S193	Good
R238	Good
I239	Good
D240	Good
W263	Good
G287	Good
W291	Good
D344	Good
A356	Good
T364	Good
G384	Good
N399	Good
P454	Good
R455	Good
L466	Good
Q457	Good
GLY	Good
A460	Good
W461	Good
L471	Good
A478	Good
GLY	Good
VAL	Good
ARG	Good
GLY	Good
PHE	Good
HIS	Good
SER	Good
ILE	Good
HIS	Good
SER	Good
THR	Good
GLY	Good
E491	Good
F506	Good
G517	Good
A518	Good
N536	Good
GLN	Good
GLN	Good
GLY	Good
ASN	Good
TRP	Good
ALA	Good
PRO	Good
ALA	Good
GLN	Good
ASP	Good
VAL	Good
THR	Good
GLN	Good
ALA	Good
GLU	Good
N552	Good
V553	Good
S554	Good
V557	Good
G558	Good
I559	Good
G599	Good
I560	Good
D561	Good
Q564	Good
T612	Good
L640	Good
L659	Good
H690	Good
T692	Good
T805	Good
V806	Good
D855	Good
S856	Good
GLU	Good
ALA	Good
G859	Good
G889	Good
R936	Good
L1008	Good
D1030	Good
S1047	Good
D1072	Good
L1089	Good
D1090	Good
E1091	Good
E1125	Good
S1126	Good
W1131	Good
R1148	Good
V1172	Good
G1226	Good
G1230	Good
I1262	Good
H1275	Good
E1279	Good
R1280	Good
M1281	Good
A1282	Good
D1283	Good
N1284	Good
G1285	Good
T1286	Good
L1292	Good
R1295	Good
E1299	Good
R1308	Good
I1324	Good

Chain B:

68% 30% 2%

Chain B: MET SER THR S4 S91 T109 ASP GLU T113 E114 T164 A217 W220 T234 T243 I262 N357 N364 T412 A429 P491 V492 MET ARG ASP GLY SER ALA D499 D600 K520 T521 GLY ASN S524 A525 Q526 T527 G544 D545 T546 S560 R561

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	148.61Å 132.76Å 155.36Å 90.00° 103.85° 90.00°	Depositor
Resolution (Å)	32.51 – 2.40 32.51 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.0 (32.51-2.40) 99.1 (32.51-2.40)	Depositor EDS
$R_{merge}$	0.30	Depositor
$R_{sym}$	0.30	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.8.0151	Depositor
R, $R_{free}$	0.216 , 0.260 0.219 , 0.261	Depositor DCC
$R_{free}$ test set	5641 reflections (5.24%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.8	Xtriage
Anisotropy	0.149	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 27.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	16842	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CA, CL

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	0.45	0/11761	0.68	3/16036 (0.0%)
2	B	0.46	0/5276	0.69	1/7200 (0.0%)
All	All	0.45	0/17037	0.68	4/23236 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1407	ARG	NE-CZ-NH1	5.54	123.07	120.30
1	A	238	ARG	NE-CZ-NH2	5.37	122.98	120.30
2	B	262	LEU	CA-CB-CG	5.23	127.32	115.30
1	A	1148	ARG	NE-CZ-NH2	5.14	122.87	120.30

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	A	11438	0	10838	19	0
2	B	5177	0	4917	5	0
3	A	2	0	0	0	0
4	A	2	0	0	0	0
5	A	2	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	2	0	0	0	0
6	A	161	0	0	0	0
6	B	58	0	0	0	0
All	All	16842	0	15755	23	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 1.

The worst 5 of 23 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:A:1295[B]:ARG:HH21	1:A:1295[B]:ARG:HB2	1.34	0.92
1:A:640:LEU:HD13	1:A:659:LEU:HD21	1.72	0.71
1:A:553:VAL:HG12	1:A:554:SER:N	2.26	0.50
1:A:1292:LEU:HD11	1:A:1299:GLU:HA	1.94	0.49
1:A:1230:GLY:HA2	1:A:1343:TYR:CE1	2.49	0.47

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	1441/1491 (97%)	1402 (97%)	38 (3%)	1 (0%)	55	72
2	B	667/965 (69%)	645 (97%)	21 (3%)	1 (0%)	55	72
All	All	2108/2456 (86%)	2047 (97%)	59 (3%)	2 (0%)	55	72

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1125	GLU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
2	B	364	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	A	1214/1267 (96%)	1201 (99%)	13 (1%)	78	90
2	B	539/810 (66%)	535 (99%)	4 (1%)	87	94
All	All	1753/2077 (84%)	1736 (99%)	17 (1%)	80	91

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1030	ASP
1	A	1091	GLU
2	B	91	SER
1	A	1008	LEU
2	B	412	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1001	GLN
1	A	1455	GLN
1	A	1022	GLN
1	A	994	GLN
1	A	1024	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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, 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	A	1446/1491 (96%)	-0.17	39 (2%) 55 52	25, 41, 73, 106	0
2	B	675/965 (69%)	-0.20	19 (2%) 53 51	27, 44, 76, 104	0
All	All	2121/2456 (86%)	-0.18	58 (2%) 55 52	25, 42, 73, 106	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	560	ILE	6.6
1	A	806	VAL	5.3
1	A	561	ASP	4.7
1	A	856	SER	4.6
2	B	492	VAL	4.6

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CA	A	1506	1/1	0.96	0.12	-0.06	50,50,50,50	0
3	CA	A	1501	1/1	0.97	0.13	-0.13	48,48,48,48	0
4	NA	A	1502	1/1	0.98	0.12	-0.36	28,28,28,28	0
5	CL	B	1002	1/1	0.99	0.10	-0.90	46,46,46,46	0
4	NA	A	1503	1/1	0.81	0.11	-1.42	65,65,65,65	0
5	CL	A	1504	1/1	0.97	0.06	-3.42	39,39,39,39	0
5	CL	B	1001	1/1	0.96	0.07	-3.77	70,70,70,70	0
5	CL	A	1505	1/1	0.94	0.05	-	58,58,58,58	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.