



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:20 AM GMT

PDB ID : 2WRE  
Title : STRUCTURE OF H2 JAPAN HEMAGGLUTININ WITH HUMAN RECEPTOR  
Authors : Liu, J.; Stevens, D.J.; Haire, L.F.; Walker, P.A.; Coombs, P.J.; Russell, R.J.; Gamblin, S.J.; Skehel, J.J.  
Deposited on : 2009-09-01  
Resolution : 3.00 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

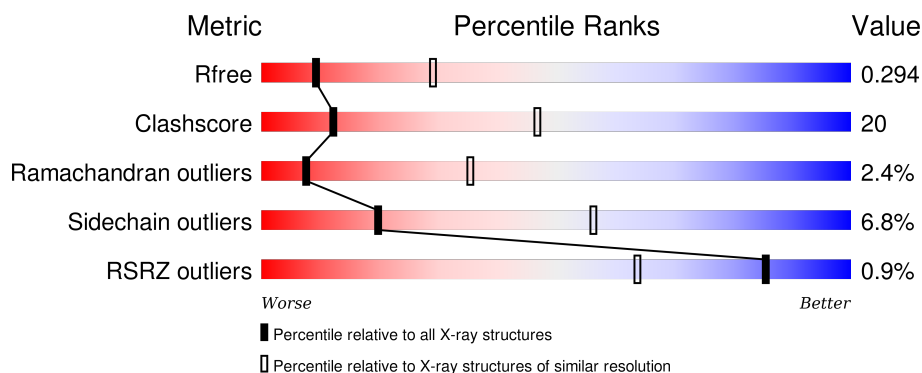
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	505	<div> <div>%</div> <div> <div></div> <div>55%</div> <div>35%</div> <div>7%</div> </div> </div>
1	B	505	<div> <div></div> <div>60%</div> <div>33%</div> <div></div> </div>
1	C	505	<div> <div>%</div> <div> <div></div> <div>57%</div> <div>34%</div> <div>5%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GAL	B	1496	-	-	-	X

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11443 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 HEMAGGLUTININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	471	Total	C	N	O	S	0	0	0
			3708	2335	633	716	24			
1	B	486	Total	C	N	O	S	0	0	0
			3841	2417	658	742	24			
1	C	481	Total	C	N	O	S	0	0	0
			3798	2392	652	730	24			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	107	HIS	TYR	CONFLICT	UNP Q67085
A	136	GLN	ARG	CONFLICT	UNP Q67085
A	159	ASN	ASP	CONFLICT	UNP Q67085
A	185	ILE	ASN	CONFLICT	UNP Q67085
A	188	LYS	THR	CONFLICT	UNP Q67085
A	307	LYS	ARG	CONFLICT	UNP Q67085
B	107	HIS	TYR	CONFLICT	UNP Q67085
B	136	GLN	ARG	CONFLICT	UNP Q67085
B	159	ASN	ASP	CONFLICT	UNP Q67085
B	185	ILE	ASN	CONFLICT	UNP Q67085
B	188	LYS	THR	CONFLICT	UNP Q67085
B	307	LYS	ARG	CONFLICT	UNP Q67085
C	107	HIS	TYR	CONFLICT	UNP Q67085
C	136	GLN	ARG	CONFLICT	UNP Q67085
C	159	ASN	ASP	CONFLICT	UNP Q67085
C	185	ILE	ASN	CONFLICT	UNP Q67085
C	188	LYS	THR	CONFLICT	UNP Q67085
C	307	LYS	ARG	CONFLICT	UNP Q67085

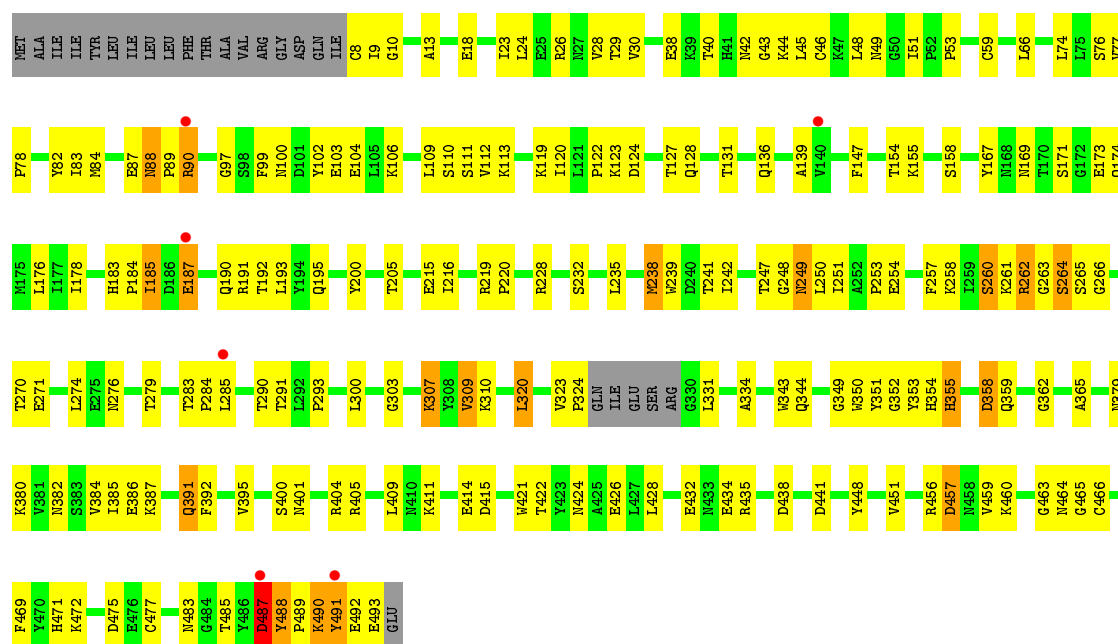
- Molecule 2 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	2	Total 32	C 17	N 1	O 14	0	0
2	B	2	Total 32	C 17	N 1	O 14	0	0
2	C	2	Total 32	C 17	N 1	O 14	0	0





• Molecule 1: HEMAGGLUTININ



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.97Å 117.46Å 225.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 3.00 46.27 – 3.00	Depositor EDS
% Data completeness (in resolution range)	96.8 (19.98-3.00) 96.8 (46.27-3.00)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.16 (at 3.01Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.239 , 0.295 0.234 , 0.294	Depositor DCC
$R_{free}$ test set	2005 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	68.9	Xtriage
Anisotropy	0.014	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 42.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Outliers	0 of 39874 reflections	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	11443	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.70% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SIA, GAL

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	0.43	0/3788	0.62	0/5121
1	B	0.46	0/3928	0.60	0/5313
1	C	0.46	0/3885	0.60	0/5255
All	All	0.45	0/11601	0.61	0/15689

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	B	0	1
1	C	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	490	LYS	Peptide
1	C	490	LYS	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	3708	0	3583	168	0
1	B	3841	0	3704	155	0
1	C	3798	0	3669	159	0
2	A	32	0	28	7	0
2	B	32	0	28	0	0
2	C	32	0	28	1	0
All	All	11443	0	11040	445	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 20.

The worst 5 of 445 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:B:487:ASP:HA	1:B:488:TYR:HB2	1.43	0.99
1:C:487:ASP:HA	1:C:488:TYR:HB2	1.46	0.95
1:C:488:TYR:O	1:C:488:TYR:CD2	2.23	0.91
1:A:140:VAL:HG21	1:A:148:ARG:HH22	1.34	0.91
1:C:492:GLU:HA	1:C:493:GLU:C	1.91	0.91

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	463/505 (92%)	389 (84%)	63 (14%)	11 (2%)	7	35
1	B	482/505 (95%)	433 (90%)	37 (8%)	12 (2%)	7	34
1	C	477/505 (94%)	429 (90%)	37 (8%)	11 (2%)	8	36
All	All	1422/1515 (94%)	1251 (88%)	137 (10%)	34 (2%)	7	35

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	262	ARG
1	B	262	ARG
1	B	487	ASP
1	B	488	TYR
1	C	262	ARG

### 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	A	409/442 (92%)	376 (92%)	33 (8%)	15	47
1	B	423/442 (96%)	396 (94%)	27 (6%)	22	59
1	C	418/442 (95%)	393 (94%)	25 (6%)	24	62
All	All	1250/1326 (94%)	1165 (93%)	85 (7%)	20	56

5 of 85 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	106	LYS
1	B	283	THR
1	C	355	HIS
1	B	109	LEU
1	B	238	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	129	HIS
1	B	243	ASN
1	C	424	ASN
1	B	190	GLN
1	B	249	ASN

### 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 ⓘ

6 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	SIA	A	1494	2	16,20,21	0.45	0	18,28,31	0.95	1 (5%)
2	GAL	A	1495	2	12,12,12	0.36	0	17,17,17	0.89	0
2	SIA	B	1495	2	16,20,21	0.78	0	18,28,31	1.18	2 (11%)
2	GAL	B	1496	2	12,12,12	0.69	0	17,17,17	0.54	0
2	SIA	C	1494	2	16,20,21	0.63	0	18,28,31	1.24	2 (11%)
2	GAL	C	1495	2	12,12,12	0.51	0	17,17,17	0.96	1 (5%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SIA	A	1494	2	-	0/14/34/38	0/1/1/1
2	GAL	A	1495	2	-	0/2/22/22	0/1/1/1
2	SIA	B	1495	2	-	0/14/34/38	0/1/1/1
2	GAL	B	1496	2	-	0/2/22/22	0/1/1/1
2	SIA	C	1494	2	-	0/14/34/38	0/1/1/1
2	GAL	C	1495	2	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1494	SIA	O9-C9-C8	-3.45	103.59	111.10
2	B	1495	SIA	C4-C5-N5	-3.04	103.80	110.41
2	B	1495	SIA	C8-C7-C6	-2.70	107.57	113.01
2	A	1494	SIA	C8-C7-C6	-2.25	108.48	113.01
2	C	1494	SIA	C9-C8-C7	-2.05	107.66	112.48

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1494	SIA	7	0
2	C	1494	SIA	1	0

## 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, 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	471/505 (93%)	-0.09	5 (1%) 82 58	34, 72, 122, 234	0
1	B	486/505 (96%)	-0.13	2 (0%) 93 80	34, 62, 103, 186	0
1	C	481/505 (95%)	-0.10	6 (1%) 81 55	35, 59, 106, 206	0
All	All	1438/1515 (94%)	-0.11	13 (0%) 85 64	34, 64, 112, 234	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	330	GLY	3.1
1	A	140	VAL	3.0
1	C	487	ASP	3.0
1	B	187	GLU	2.9
1	A	44	LYS	2.8

### 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](#)

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(Å <sup>2</sup> )	Q<0.9
2	GAL	B	1496	12/12	0.72	0.31	2.23	86,86,86,86	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	SIA	C	1494	20/21	0.91	0.24	1.11	78,78,78,78	0
2	SIA	A	1494	20/21	0.90	0.30	0.81	110,110,110,110	0
2	SIA	B	1495	20/21	0.93	0.18	-0.49	86,86,86,86	0
2	GAL	A	1495	12/12	0.85	0.28	-	110,110,110,110	0
2	GAL	C	1495	12/12	0.90	0.19	-	78,78,78,78	0

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.