



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 07:26 am BST

PDB ID : 4A11  
Title : Structure of the hsDDB1-hsCSA complex  
Authors : Bohm, K.; Scrima, A.; Fischer, E.S.; Gut, H.; Thomae, N.H.  
Deposited on : 2011-09-13  
Resolution : 3.31 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
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

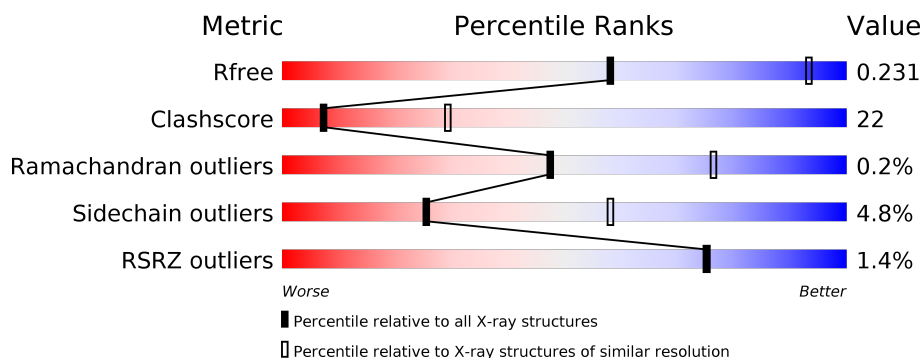
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.31 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	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

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 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  $\leq 5\%$ . 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	1159	<div> <div>2%</div> <div> <div></div> <div>55%</div> <div>37%</div> <div>6%</div> </div> </div>
2	B	408	<div> <div>51%</div> <div>36%</div> <div>11%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11081 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 DNA DAMAGE-BINDING PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1091	8262	5265	1367	1585	45	0	0	0

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MET	-	expression tag	UNP Q16531
A	-17	HIS	-	expression tag	UNP Q16531
A	-16	HIS	-	expression tag	UNP Q16531
A	-15	HIS	-	expression tag	UNP Q16531
A	-14	HIS	-	expression tag	UNP Q16531
A	-13	HIS	-	expression tag	UNP Q16531
A	-12	HIS	-	expression tag	UNP Q16531
A	-11	VAL	-	expression tag	UNP Q16531
A	-10	ASP	-	expression tag	UNP Q16531
A	-9	GLU	-	expression tag	UNP Q16531
A	-8	ASN	-	expression tag	UNP Q16531
A	-7	LEU	-	expression tag	UNP Q16531
A	-6	TYR	-	expression tag	UNP Q16531
A	-5	PHE	-	expression tag	UNP Q16531
A	-4	GLN	-	expression tag	UNP Q16531
A	-3	GLY	-	expression tag	UNP Q16531
A	-2	GLY	-	expression tag	UNP Q16531
A	-1	GLY	-	expression tag	UNP Q16531
A	0	ARG	-	expression tag	UNP Q16531

- Molecule 2 is a protein called DNA EXCISION REPAIR PROTEIN ERCC-8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	365	2819	1754	498	548	19	0	0	0

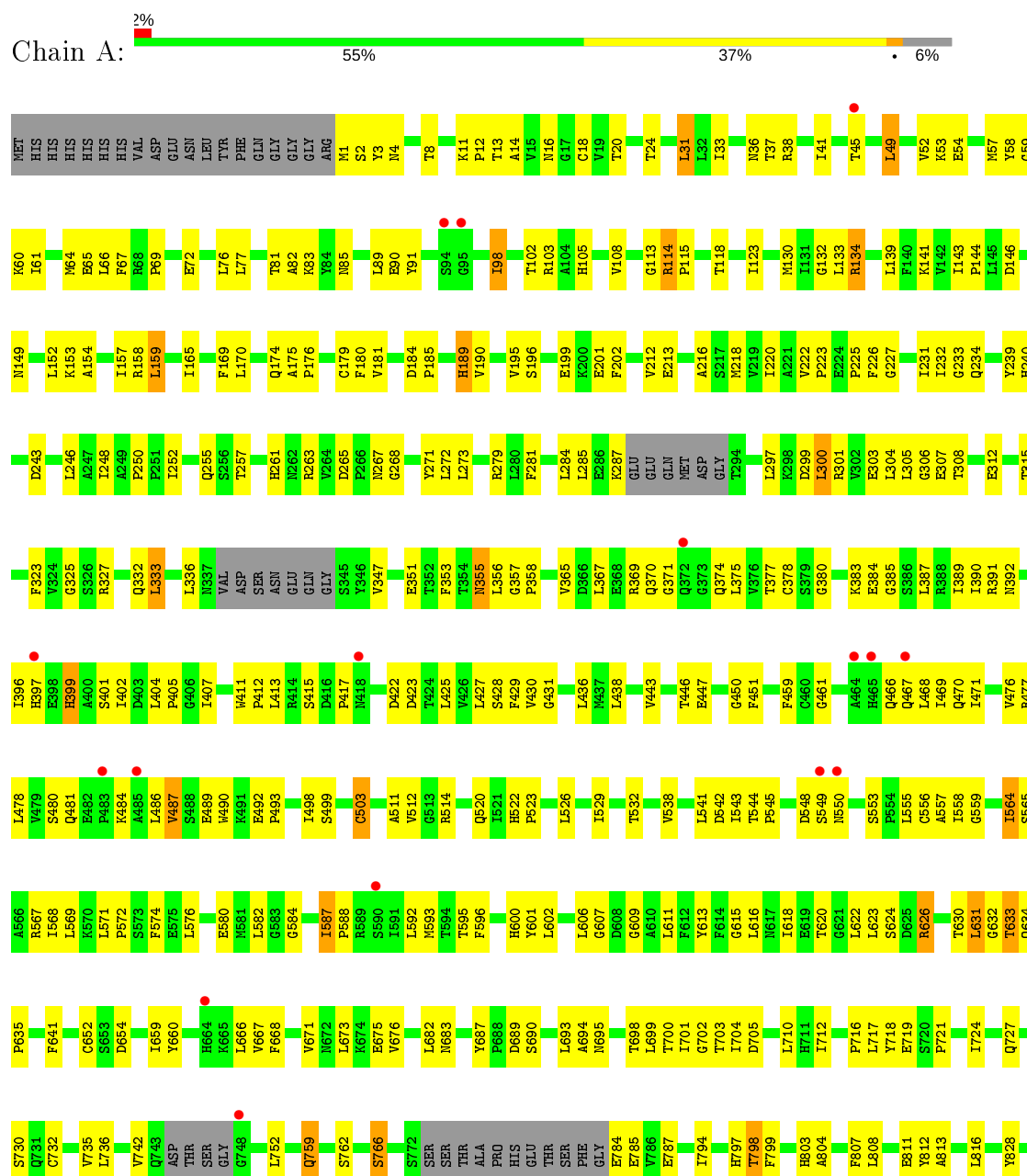
There are 12 discrepancies between the modelled and reference sequences:

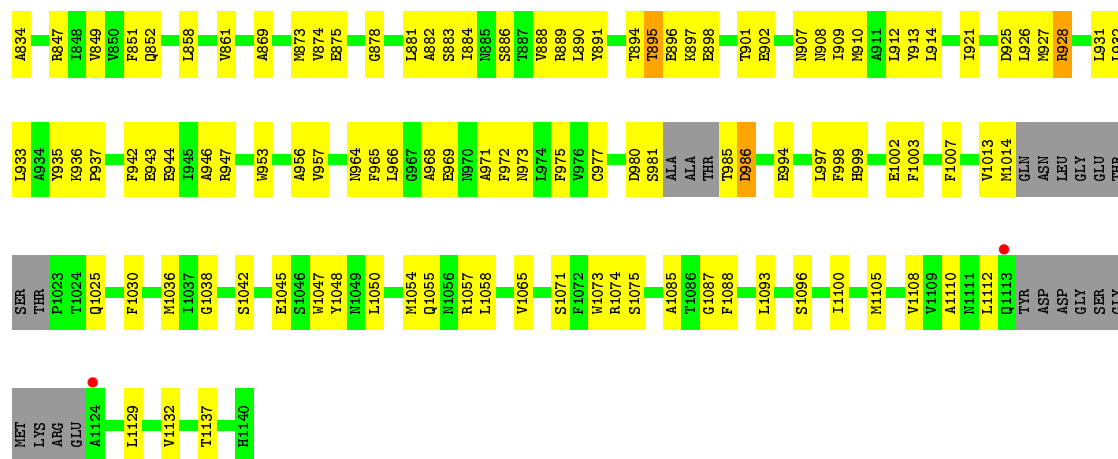
Chain	Residue	Modelled	Actual	Comment	Reference
B	397	GLY	-	expression tag	UNP Q13216
B	398	THR	-	expression tag	UNP Q13216
B	399	SER	-	expression tag	UNP Q13216
B	400	ALA	-	expression tag	UNP Q13216
B	401	TRP	-	expression tag	UNP Q13216
B	402	SER	-	expression tag	UNP Q13216
B	403	HIS	-	expression tag	UNP Q13216
B	404	PRO	-	expression tag	UNP Q13216
B	405	GLN	-	expression tag	UNP Q13216
B	406	PHE	-	expression tag	UNP Q13216
B	407	GLU	-	expression tag	UNP Q13216
B	408	LYS	-	expression tag	UNP Q13216

### 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.

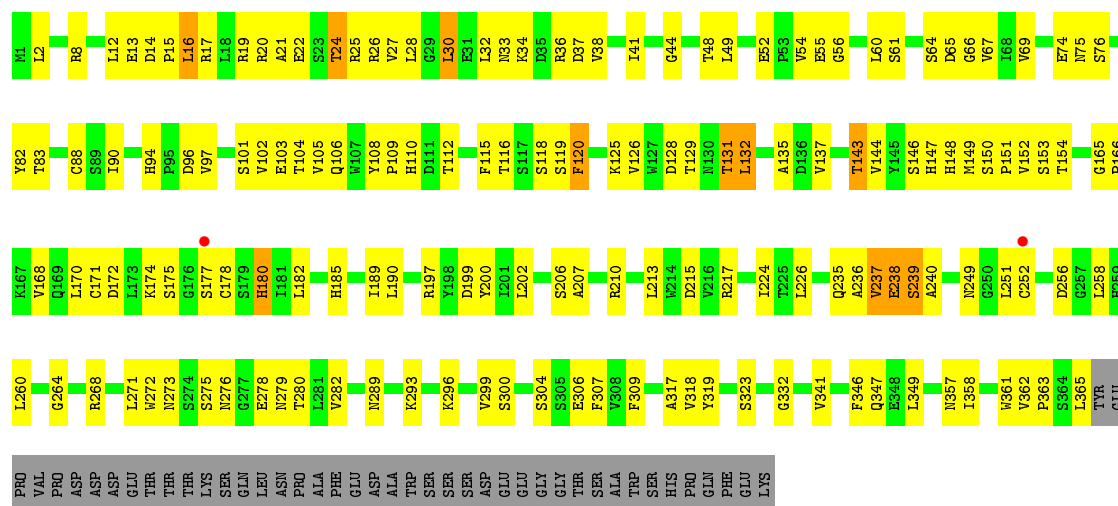
#### • Molecule 1: DNA DAMAGE-BINDING PROTEIN 1





• Molecule 2: DNA EXCISION REPAIR PROTEIN ERCC-8

Chain B: 51% 36% 11%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	138.33Å 138.33Å 244.97Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	32.07 – 3.31 32.07 – 3.31	Depositor EDS
% Data completeness (in resolution range)	92.7 (32.07-3.31) 92.7 (32.07-3.31)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.33 (at 3.32Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.176 , 0.233 0.173 , 0.231	Depositor DCC
$R_{free}$ test set	1908 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.8	Xtriage
Anisotropy	0.512	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 76.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.037 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11081	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% 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](#)

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	1/8416 (0.0%)	0.64	1/11456 (0.0%)
2	B	0.47	0/2878	0.70	0/3908
All	All	0.44	1/11294 (0.0%)	0.66	1/15364 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	977	CYS	CB-SG	-5.58	1.72	1.81

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	49	LEU	CA-CB-CG	5.46	127.85	115.30

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	8262	0	8011	362	0
2	B	2819	0	2701	131	0
All	All	11081	0	10712	488	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 22.



The worst 5 of 488 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:158:ARG:NH2	2:B:258:LEU:O	1.89	1.04
1:A:134:ARG:HH11	1:A:134:ARG:HG2	0.87	1.02
1:A:520:GLN:HG3	1:A:529:ILE:HD11	1.42	1.01
1:A:57:MET:HG3	1:A:61:ILE:HD11	1.45	0.99
1:A:199:GLU:HG2	1:A:201:GLU:HB3	1.43	0.98

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	1075/1159 (93%)	1019 (95%)	54 (5%)	2 (0%)	47 76
2	B	363/408 (89%)	347 (96%)	15 (4%)	1 (0%)	41 71
All	All	1438/1567 (92%)	1366 (95%)	69 (5%)	3 (0%)	47 76

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	928	ARG
2	B	237	VAL
1	A	564	ILE

### 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	891/1015 (88%)	855 (96%)	36 (4%)	31	62
2	B	313/358 (87%)	291 (93%)	22 (7%)	15	44
All	All	1204/1373 (88%)	1146 (95%)	58 (5%)	25	58

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

Mol	Chain	Res	Type
1	A	752	LEU
1	A	957	VAL
2	B	238	GLU
1	A	759	GLN
1	A	798	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	964	ASN
2	B	148	HIS
2	B	133	GLN
1	A	695	ASN
1	A	1077	HIS

### 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, 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	1091/1159 (94%)	-0.16	18 (1%) 72 70	44, 84, 140, 243	0
2	B	365/408 (89%)	-0.38	2 (0%) 91 91	43, 69, 103, 173	0
All	All	1456/1567 (92%)	-0.21	20 (1%) 75 75	43, 80, 136, 243	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	485	ALA	3.9
1	A	483	PRO	3.6
1	A	550	ASN	3.6
1	A	549	SER	3.6
1	A	1124	ALA	3.6

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.