



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 16, 2023 – 07:18 AM JST

PDB ID : 6L6M  
Title : HSP18.5 from *E. histolytica*  
Authors : Kurre, D.; Suguna, K.  
Deposited on : 2019-10-29  
Resolution : 3.28 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

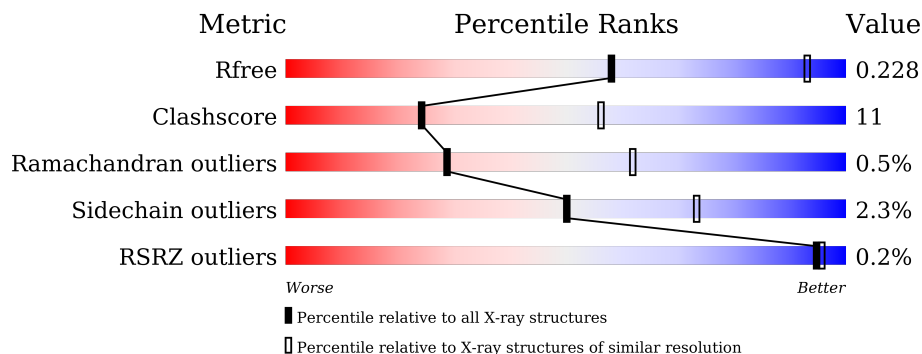
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.28 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	1177 (3.32-3.24)
Clashscore	141614	1044 (3.30-3.26)
Ramachandran outliers	138981	1026 (3.30-3.26)
Sidechain outliers	138945	1025 (3.30-3.26)
RSRZ outliers	127900	1141 (3.32-3.24)

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  $\geq 3$ , 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	167	 47% 14% 38%
1	B	167	 46% 16% 37%
1	C	167	 49% 14% 37%
1	D	167	 44% 19% 37%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3269 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 Heat shock protein hsp20 family putative.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	104	830	542	131	153	4	0	1	0
1	B	105	806	519	128	155	4	0	0	0
1	C	105	822	532	130	156	4	0	0	0
1	D	105	811	523	127	157	4	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	expression tag	UNP C4M4U3
A	-1	ALA	-	expression tag	UNP C4M4U3
A	0	SER	-	expression tag	UNP C4M4U3
B	-2	MET	-	expression tag	UNP C4M4U3
B	-1	ALA	-	expression tag	UNP C4M4U3
B	0	SER	-	expression tag	UNP C4M4U3
C	-2	MET	-	expression tag	UNP C4M4U3
C	-1	ALA	-	expression tag	UNP C4M4U3
C	0	SER	-	expression tag	UNP C4M4U3
D	-2	MET	-	expression tag	UNP C4M4U3
D	-1	ALA	-	expression tag	UNP C4M4U3
D	0	SER	-	expression tag	UNP C4M4U3

### 3 Residue-property plots [i](#)

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: Heat shock protein hsp20 family putative

Chain A:  47% 14% 38%

MET ALA SER MET SER SER SER GLU ALA ALA PRO PRO ILE VAL GLN SER LEU LEU LEU ALA ILE ILE PRO PRO SER SER GLN ASN ASN ASN ASN ASN GLN GLN LEU LEU ALA ALA LYS LYS PRO PRO GLU PRO LYS TRP TRP ILE HIS HIS LEU SER ARG ARG TYR LEU LEU SER SER LYS THR SER GLN ASN ASN ARG ARG VAL PHE VAL VAL ASP ASP PRO PRO SER SER SER GLY VAL GLY HIS HIS PHE ASN

SER MET T60 C65 E66 L67 L68 G71 T72 N73 K77 V80 I83 L88 S93 N94 N95 K103 E116 I117 L118 Y119 V128 P129 V130 D131 K138 A139 Y140 K149 K152 E162 I163 VAL

- Molecule 1: Heat shock protein hsp20 family putative

Chain B:  46% 16% 37%

MET ALA SER MET SER SER SER GLU ALA ALA PRO PRO ILE VAL GLN SER LEU LEU LEU ALA ILE ILE PRO PRO SER SER GLN ASN ASN ASN ASN ASN GLN GLN LEU LEU ALA ALA LYS LYS PRO PRO GLU PRO LYS TRP TRP ILE HIS HIS LEU SER ARG ARG TYR LEU LEU SER SER LYS THR SER GLN ASN ASN ARG ARG VAL PHE VAL VAL ASP ASP PRO PRO SER SER SER GLY VAL GLY HIS HIS PHE ASN

SER MET T60 W61 C65 L75 L76 K77 E79 W80 P81 I83 L88 S93 W96 G101 N102 K103 N104 M105 P106 E109 C113 F114 Q121 F122 P127 S133 K134 I145 L146 Y147 V148 K149 V164

- Molecule 1: Heat shock protein hsp20 family putative

Chain C:  49% 14% 37%

MET ALA SER MET SER SER SER GLU ALA ALA PRO PRO ILE VAL GLN SER LEU LEU LEU ALA ILE ILE PRO PRO SER SER GLN ASN ASN ASN ASN ASN GLN GLN LEU LEU ALA ALA LYS LYS PRO PRO GLU PRO LYS TRP TRP ILE HIS HIS LEU SER ARG ARG TYR LEU LEU SER SER LYS THR SER GLN ASN ASN ARG ARG VAL PHE VAL VAL ASP ASP PRO PRO SER SER SER GLY VAL GLY HIS HIS PHE ASN

SER MET T60 P63 P64 C65 E66 N73 Y74 K77 F78 E79 I83 D84 L88 W96 P106 E109 F112 C113 F114 T115 R124 P127 D131 A139 Y147 K149 L150 L151 K152 V164

- Molecule 1: Heat shock protein hsp20 family putative

Chain D:  44% 19% 37%

MET ALA SER MET SER SER SER GLU ALA ALA PRO PRO ILE VAL GLN SER LEU LEU LEU ALA ILE ILE PRO PRO SER SER GLN ASN ASN ASN ASN ASN GLN GLN LEU LEU ALA ALA LYS LYS PRO PRO GLU PRO LYS TRP TRP ILE HIS HIS LEU SER ARG ARG TYR LEU LEU SER SER LYS THR SER GLN ASN ASN ARG ARG VAL PHE VAL VAL ASP ASP PRO PRO SER SER SER GLY VAL GLY HIS HIS PHE ASN

SER MET T60 W61 E62 P63 P64 L67 C70 G71 T72 N73 K77 F78 E79 V80 P81 G82 I83 L88 V97 I98 K103 P106 I107 D108 E109 F112 Q121 F122 R123 R124 E125 D131 A132 S133 Y141 Y147 L151 K152 V164

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.24Å 173.99Å 159.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.53 – 3.28 79.53 – 3.28	Depositor EDS
% Data completeness (in resolution range)	99.9 (79.53-3.28) 100.0 (79.53-3.28)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.99 (at 3.26Å)	Xtrriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.187 , 0.228 0.187 , 0.228	Depositor DCC
$R_{free}$ test set	769 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	112.4	Xtrriage
Anisotropy	0.039	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 77.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3269	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	99.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.62% 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	0.53	0/855	0.63	0/1165
1	B	0.47	0/824	0.59	0/1125
1	C	0.54	0/843	0.60	0/1152
1	D	0.48	0/831	0.61	0/1138
All	All	0.51	0/3353	0.60	0/4580

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 [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	830	0	806	17	0
1	B	806	0	764	15	0
1	C	822	0	782	20	0
1	D	811	0	756	25	0
All	All	3269	0	3108	70	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 11.

The worst 5 of 70 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:73:ASN:HD22	1:A:149:LYS:HG3	1.49	0.77
1:D:98:ILE:HG13	1:D:125:GLU:HG2	1.69	0.73
1:B:106:PRO:HB2	1:B:109:GLU:HB2	1.75	0.68
1:C:66:GLU:OE1	1:C:77:LYS:HE2	1.96	0.65
1:C:63:PRO:HD2	1:C:124:ARG:HD3	1.80	0.63

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	103/167 (62%)	99 (96%)	3 (3%)	1 (1%)	15	48
1	B	103/167 (62%)	98 (95%)	5 (5%)	0	100	100
1	C	103/167 (62%)	98 (95%)	5 (5%)	0	100	100
1	D	103/167 (62%)	98 (95%)	4 (4%)	1 (1%)	15	48
All	All	412/668 (62%)	393 (95%)	17 (4%)	2 (0%)	29	62

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	71	GLY
1	A	71	GLY

### 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	91/152 (60%)	90 (99%)	1 (1%)	73	85
1	B	87/152 (57%)	82 (94%)	5 (6%)	20	51
1	C	90/152 (59%)	89 (99%)	1 (1%)	73	85
1	D	87/152 (57%)	86 (99%)	1 (1%)	73	85
All	All	355/608 (58%)	347 (98%)	8 (2%)	50	73

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	123	ARG
1	C	84	ASP
1	B	133	SER
1	B	121	GLN
1	B	149	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides 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

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	104/167 (62%)	0.05	0 <a href="#">100</a>   <a href="#">100</a>	72, 92, 114, 131	0
1	B	105/167 (62%)	-0.09	1 (0%) <a href="#">82</a>   <a href="#">82</a>	79, 98, 127, 150	0
1	C	105/167 (62%)	0.12	0 <a href="#">100</a>   <a href="#">100</a>	72, 94, 132, 169	0
1	D	105/167 (62%)	-0.06	0 <a href="#">100</a>   <a href="#">100</a>	80, 105, 124, 143	0
All	All	419/668 (62%)	0.00	1 (0%) <a href="#">95</a>   <a href="#">96</a>	72, 98, 126, 169	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	78	PHE	2.2

### 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 monosaccharides 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.