wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2024 – 10:34 PM EST

PDB ID	:	3FPV
Title	:	Crystal Structure of HbpS
Authors	:	Ortiz de Orue Lucana, D.; Bogel, G.; Zou, P.; Groves, M.R.
Deposited on	:	2009-01-06
Resolution	:	2.20 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

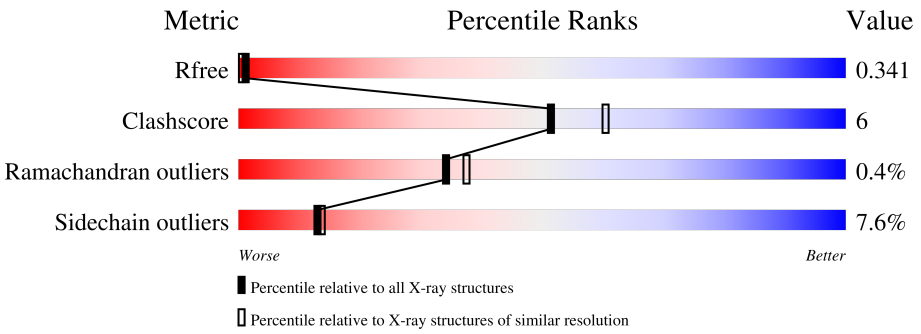
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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 >=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%

Mol	Chain	Length	Quality of chain
1	A	192	60% 11% • 27%
1	B	192	64% 7% • • 27%
1	C	192	61% 10% • 27%
1	D	192	60% 11% • 27%
1	E	192	61% 10% • • 27%
1	F	192	59% 11% • • 27%
1	G	192	62% 9% • 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	192	61% 9% • • 27%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8497 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 Extracellular haem-binding protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	141	997	614	185	198	0	0	0
1	B	141	997	614	185	198	0	0	0
1	C	141	997	614	185	198	0	0	0
1	D	141	997	614	185	198	0	0	0
1	E	141	997	614	185	198	0	0	0
1	F	141	997	614	185	198	0	0	0
1	G	141	997	614	185	198	0	0	0
1	H	141	997	614	185	198	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-35	GLY	-	expression tag	UNP Q9RIM2
A	-34	ALA	-	expression tag	UNP Q9RIM2
A	-33	GLY	-	expression tag	UNP Q9RIM2
A	-32	ALA	-	expression tag	UNP Q9RIM2
B	-35	GLY	-	expression tag	UNP Q9RIM2
B	-34	ALA	-	expression tag	UNP Q9RIM2
B	-33	GLY	-	expression tag	UNP Q9RIM2
B	-32	ALA	-	expression tag	UNP Q9RIM2
C	-35	GLY	-	expression tag	UNP Q9RIM2
C	-34	ALA	-	expression tag	UNP Q9RIM2
C	-33	GLY	-	expression tag	UNP Q9RIM2
C	-32	ALA	-	expression tag	UNP Q9RIM2
D	-35	GLY	-	expression tag	UNP Q9RIM2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-34	ALA	-	expression tag	UNP Q9RIM2
D	-33	GLY	-	expression tag	UNP Q9RIM2
D	-32	ALA	-	expression tag	UNP Q9RIM2
E	-35	GLY	-	expression tag	UNP Q9RIM2
E	-34	ALA	-	expression tag	UNP Q9RIM2
E	-33	GLY	-	expression tag	UNP Q9RIM2
E	-32	ALA	-	expression tag	UNP Q9RIM2
F	-35	GLY	-	expression tag	UNP Q9RIM2
F	-34	ALA	-	expression tag	UNP Q9RIM2
F	-33	GLY	-	expression tag	UNP Q9RIM2
F	-32	ALA	-	expression tag	UNP Q9RIM2
G	-35	GLY	-	expression tag	UNP Q9RIM2
G	-34	ALA	-	expression tag	UNP Q9RIM2
G	-33	GLY	-	expression tag	UNP Q9RIM2
G	-32	ALA	-	expression tag	UNP Q9RIM2
H	-35	GLY	-	expression tag	UNP Q9RIM2
H	-34	ALA	-	expression tag	UNP Q9RIM2
H	-33	GLY	-	expression tag	UNP Q9RIM2
H	-32	ALA	-	expression tag	UNP Q9RIM2

- Molecule 2 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Fe		
2	A	1	1	1	0	0
			Total	Fe		
2	B	1	1	1	0	0
			Total	Fe		
2	C	1	1	1	0	0
			Total	Fe		
2	D	1	1	1	0	0
			Total	Fe		
2	E	2	2	2	0	0
			Total	Fe		
2	F	1	1	1	0	0
			Total	Fe		
2	H	1	1	1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	81	81	81	0	0

Continued on next page...

Continued from previous page...

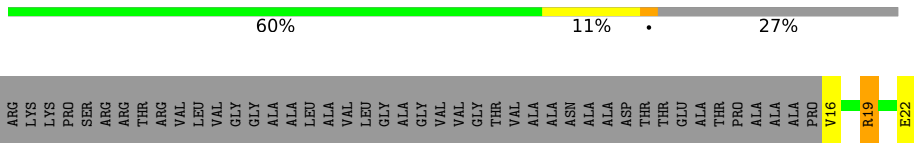
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	B	101	101	101	0	0
			Total	O		
3	C	57	57	57	0	0
			Total	O		
3	D	50	50	50	0	0
			Total	O		
3	E	35	35	35	0	0
			Total	O		
3	F	84	84	84	0	0
			Total	O		
3	G	46	46	46	0	0
			Total	O		
3	H	59	59	59	0	0

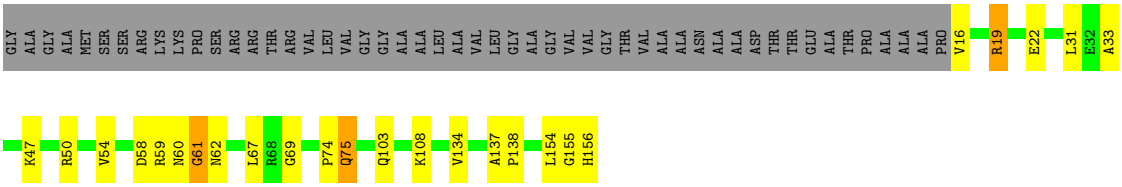
3 Residue-property plots (i)

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. 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. 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: Extracellular haem-binding protein

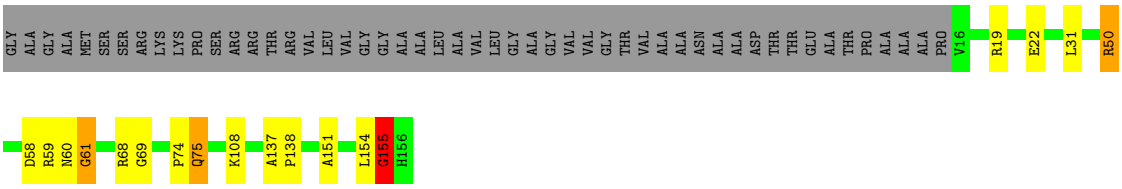
Chain A:





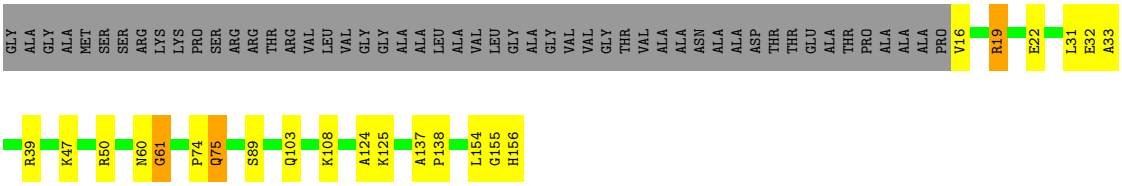
• Molecule 1: Extracellular haem-binding protein

Chain B:



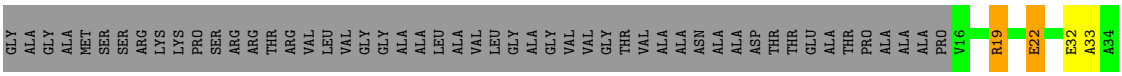
• Molecule 1: Extracellular haem-binding protein

Chain C:



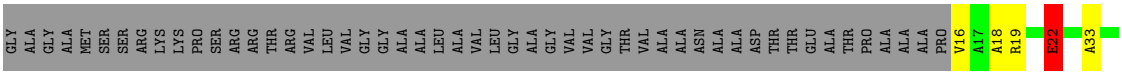
• Molecule 1: Extracellular haem-binding protein

Chain D:



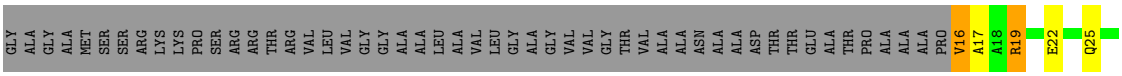
• Molecule 1: Extracellular haem-binding protein

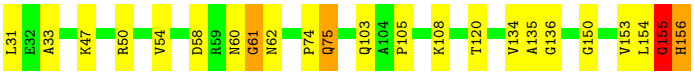
Chain E:



• Molecule 1: Extracellular haem-binding protein

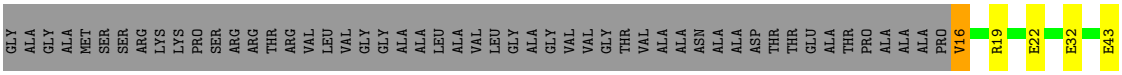
Chain F:





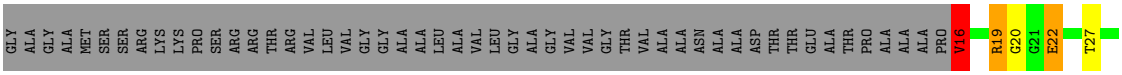
• Molecule 1: Extracellular haem-binding protein

Chain G:



• Molecule 1: Extracellular haem-binding protein

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	152.52Å 152.52Å 152.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.86 – 2.20 19.86 – 2.21	Depositor EDS
% Data completeness (in resolution range)	99.1 (19.86-2.20) 99.1 (19.86-2.21)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 2.21Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.201 , 0.264 0.295 , 0.341	Depositor DCC
R_{free} test set	2915 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	44.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 43.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.033 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8497	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.84% 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 (i)

5.1 Standard geometry (i)

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

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	0.21	0/1012	0.36	0/1376
1	B	0.21	0/1012	0.36	0/1376
1	C	0.20	0/1012	0.35	0/1376
1	D	0.20	0/1012	0.36	0/1376
1	E	0.21	0/1012	0.36	0/1376
1	F	0.21	0/1012	0.36	0/1376
1	G	0.20	0/1012	0.35	0/1376
1	H	0.20	0/1012	0.35	0/1376
All	All	0.21	0/8096	0.36	0/11008

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	A	0	4
1	B	0	5
1	C	0	6
1	D	0	8
1	E	0	10
1	F	0	8
1	G	0	4
1	H	0	6
All	All	0	51

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 51 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	154	LEU	Peptide
1	A	19	ARG	Peptide
1	A	60	ASN	Peptide
1	A	61	GLY	Peptide
1	B	19	ARG	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	997	0	996	20	0
1	B	997	0	996	17	0
1	C	997	0	996	11	0
1	D	997	0	996	14	0
1	E	997	0	996	15	0
1	F	997	0	996	22	0
1	G	997	0	996	12	0
1	H	997	0	996	16	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	2	0	0	0	0
2	F	1	0	0	0	0
2	H	1	0	0	0	0
3	A	81	0	0	0	0
3	B	101	0	0	1	0
3	C	57	0	0	0	0
3	D	50	0	0	0	0
3	E	35	0	0	0	0
3	F	84	0	0	2	0
3	G	46	0	0	0	0
3	H	59	0	0	1	0
All	All	8497	0	7968	90	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 6.

The worst 5 of 90 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:C:22:GLU:HB2	1:H:33:ALA:HB2	1.56	0.85
1:A:155:GLY:HA2	1:F:22:GLU:OE1	1.80	0.82
1:A:33:ALA:HB2	1:F:22:GLU:HB3	1.65	0.77
1:C:33:ALA:HB2	1:H:22:GLU:HB2	1.69	0.75
1:A:155:GLY:N	1:A:156:HIS:HA	2.02	0.74

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	139/192 (72%)	136 (98%)	3 (2%)	0	100	100
1	B	139/192 (72%)	137 (99%)	1 (1%)	1 (1%)	22	22
1	C	139/192 (72%)	134 (96%)	4 (3%)	1 (1%)	22	22
1	D	139/192 (72%)	132 (95%)	6 (4%)	1 (1%)	22	22
1	E	139/192 (72%)	134 (96%)	4 (3%)	1 (1%)	22	22
1	F	139/192 (72%)	136 (98%)	2 (1%)	1 (1%)	22	22
1	G	139/192 (72%)	138 (99%)	1 (1%)	0	100	100
1	H	139/192 (72%)	135 (97%)	4 (3%)	0	100	100
All	All	1112/1536 (72%)	1082 (97%)	25 (2%)	5 (0%)	34	37

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	125	LYS
1	D	125	LYS
1	E	125	LYS
1	F	155	GLY
1	B	155	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	95/125 (76%)	88 (93%)	7 (7%)	13	14
1	B	95/125 (76%)	92 (97%)	3 (3%)	39	50
1	C	95/125 (76%)	86 (90%)	9 (10%)	8	8
1	D	95/125 (76%)	86 (90%)	9 (10%)	8	8
1	E	95/125 (76%)	89 (94%)	6 (6%)	18	20
1	F	95/125 (76%)	87 (92%)	8 (8%)	11	11
1	G	95/125 (76%)	86 (90%)	9 (10%)	8	8
1	H	95/125 (76%)	88 (93%)	7 (7%)	13	14
All	All	760/1000 (76%)	702 (92%)	58 (8%)	13	14

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

Mol	Chain	Res	Type
1	E	22	GLU
1	H	75	GLN
1	F	19	ARG
1	H	50	ARG
1	G	154	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	75	GLN
1	G	75	GLN
1	E	25	GLN
1	G	25	GLN
1	D	145	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 monosaccharides 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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.