



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 16, 2017 – 04:12 PM EDT

PDB ID : 5UDA  
Title : Crystal structure of CYP2B6 (Y226H/K262R) in complex with a monoterpene bornane  
Authors : Shah, M.B.; Halpert, J.R.  
Deposited on : unknown  
Resolution : 1.93 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 : 1.7.2 (RC1), CSD as538be (2017)  
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

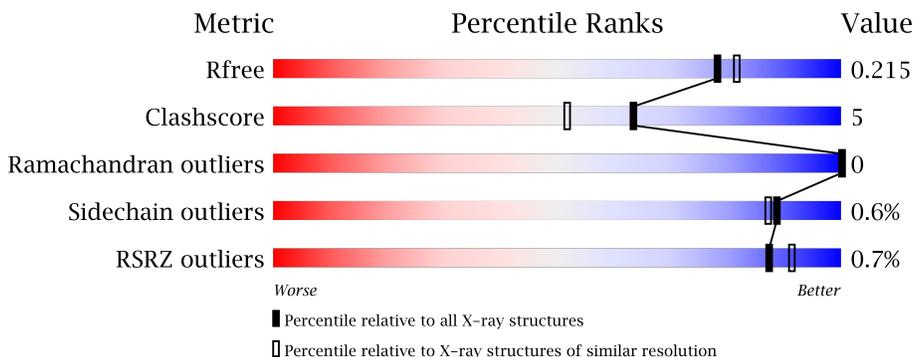
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.93 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	3233 (1.96-1.92)
Clashscore	112137	3430 (1.96-1.92)
Ramachandran outliers	110173	3395 (1.96-1.92)
Sidechain outliers	110143	3395 (1.96-1.92)
RSRZ outliers	101464	3250 (1.96-1.92)

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	476	 86% 10% ..
1	B	476	 88% 8% ..

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
3	CAE	A	502[A]	-	-	-	X
3	CAE	A	502[B]	-	-	-	X
3	CAE	B	502[A]	-	-	-	X
3	CAE	B	502[B]	-	-	-	X
4	CM5	A	504	-	-	-	X
4	CM5	A	505	-	-	-	X
4	CM5	B	504	-	-	-	X
4	CM5	B	505	-	-	-	X

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 8139 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 Cytochrome P450 2B6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	461	3703	2404	628	655	16	0	0	0
1	B	460	3713	2406	635	656	16	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

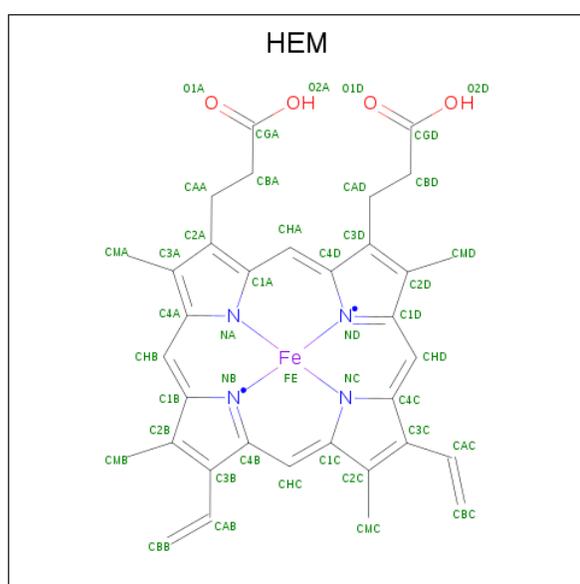
Chain	Residue	Modelled	Actual	Comment	Reference
A	20	MET	-	initiating methionine	UNP P20813
A	21	ALA	GLN	engineered mutation	UNP P20813
A	22	LYS	ARG	engineered mutation	UNP P20813
A	23	LYS	HIS	engineered mutation	UNP P20813
A	24	THR	PRO	engineered mutation	UNP P20813
A	25	SER	ASN	engineered mutation	UNP P20813
A	26	SER	THR	engineered mutation	UNP P20813
A	27	LYS	HIS	engineered mutation	UNP P20813
A	28	GLY	ASP	engineered mutation	UNP P20813
A	29	LYS	ARG	engineered mutation	UNP P20813
A	226	HIS	TYR	engineered mutation	UNP P20813
A	262	ARG	LYS	engineered mutation	UNP P20813
A	492	HIS	-	expression tag	UNP P20813
A	493	HIS	-	expression tag	UNP P20813
A	494	HIS	-	expression tag	UNP P20813
A	495	HIS	-	expression tag	UNP P20813
B	20	MET	-	initiating methionine	UNP P20813
B	21	ALA	GLN	engineered mutation	UNP P20813
B	22	LYS	ARG	engineered mutation	UNP P20813
B	23	LYS	HIS	engineered mutation	UNP P20813
B	24	THR	PRO	engineered mutation	UNP P20813
B	25	SER	ASN	engineered mutation	UNP P20813
B	26	SER	THR	engineered mutation	UNP P20813
B	27	LYS	HIS	engineered mutation	UNP P20813
B	28	GLY	ASP	engineered mutation	UNP P20813

*Continued on next page...*

Continued from previous page...

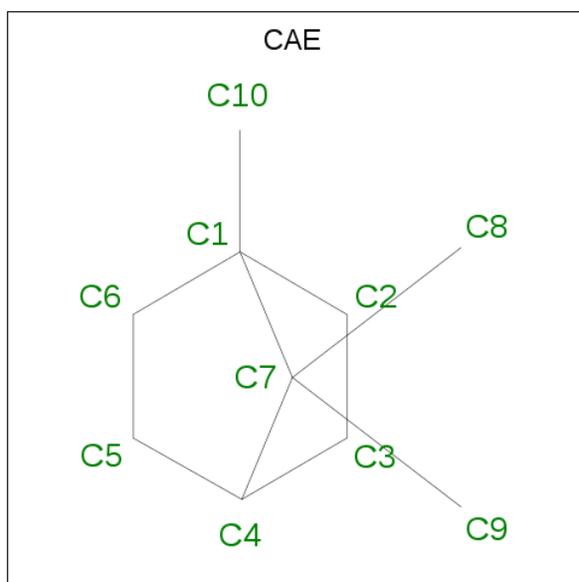
Chain	Residue	Modelled	Actual	Comment	Reference
B	29	LYS	ARG	engineered mutation	UNP P20813
B	226	HIS	TYR	engineered mutation	UNP P20813
B	262	ARG	LYS	engineered mutation	UNP P20813
B	492	HIS	-	expression tag	UNP P20813
B	493	HIS	-	expression tag	UNP P20813
B	494	HIS	-	expression tag	UNP P20813
B	495	HIS	-	expression tag	UNP P20813

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula:  $C_{34}H_{32}FeN_4O_4$ ).



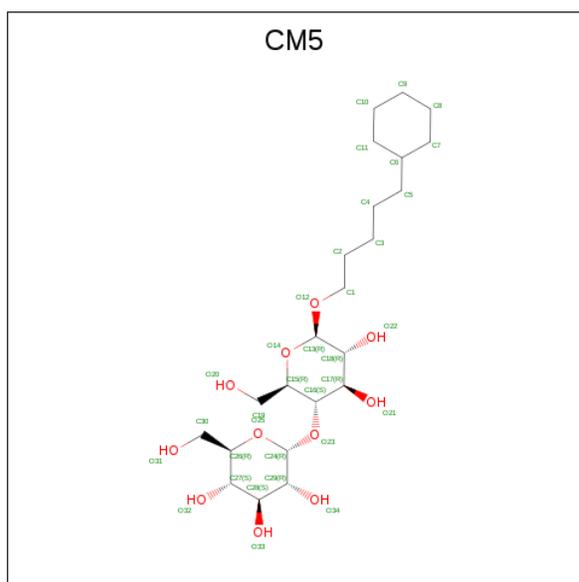
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
2	A	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	B	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		

- Molecule 3 is CAMPHANE (three-letter code: CAE) (formula:  $C_{10}H_{18}$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C 20 20	0	1
3	B	1	Total C 20 20	0	1

- Molecule 4 is 5-CYCLOHEXYL-1-PENTYL-BETA-D-MALTOSE (three-letter code: CM5) (formula: C<sub>23</sub>H<sub>42</sub>O<sub>11</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 12 11 1	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			34	23	11		
4	A	1	Total	C	O	0	0
			12	11	1		
4	B	1	Total	C	O	0	0
			12	11	1		
4	B	1	Total	C	O	0	0
			34	23	11		
4	B	1	Total	C	O	0	0
			12	11	1		

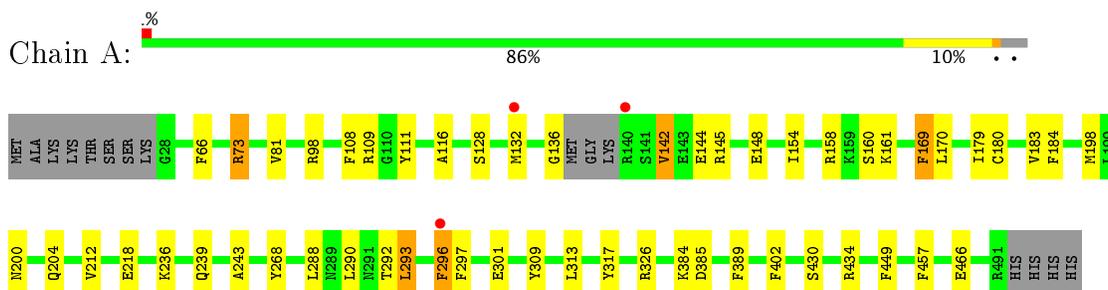
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	238	Total	O	0	0
			238	238		
5	B	243	Total	O	0	0
			243	243		

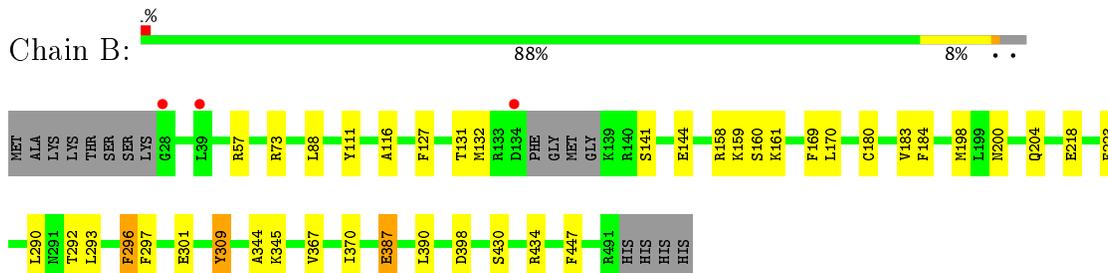
### 3 Residue-property plots [i](#)

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: Cytochrome P450 2B6



- Molecule 1: Cytochrome P450 2B6



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.49Å 77.49Å 202.15Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.43 – 1.93 40.43 – 1.93	Depositor EDS
% Data completeness (in resolution range)	96.2 (40.43-1.93) 96.2 (40.43-1.93)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.05	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.72 (at 1.92Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, $R_{free}$	0.173 , 0.210 0.180 , 0.215	Depositor DCC
$R_{free}$ test set	4817 reflections (5.16%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.2	Xtriage
Anisotropy	0.023	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l 0.477 for h,-h-k,-l 0.026 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	8139	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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: HEM, CM5, CAE

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	1.32	9/3803 (0.2%)	0.99	8/5147 (0.2%)
1	B	1.32	7/3812 (0.2%)	0.97	2/5156 (0.0%)
All	All	1.32	16/7615 (0.2%)	0.98	10/10303 (0.1%)

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	296	PHE	CB-CG	-8.62	1.36	1.51
1	B	296	PHE	CB-CG	-8.54	1.36	1.51
1	B	309	TYR	CD2-CE2	-5.94	1.30	1.39
1	B	447	PHE	CD1-CE1	5.73	1.50	1.39
1	A	169	PHE	CG-CD2	5.54	1.47	1.38

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	158	ARG	CG-CD-NE	-6.36	98.44	111.80
1	B	398	ASP	CB-CG-OD1	5.91	123.62	118.30
1	A	313	LEU	CB-CG-CD2	-5.36	101.89	111.00
1	A	293	LEU	CA-CB-CG	5.34	127.58	115.30
1	A	73	ARG	NE-CZ-NH1	5.31	122.96	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	3703	0	3670	38	0
1	B	3713	0	3694	33	0
2	A	43	0	30	5	0
2	B	43	0	30	5	0
3	A	20	0	36	0	0
3	B	20	0	36	0	0
4	A	58	0	83	6	0
4	B	58	0	76	5	0
5	A	238	0	0	3	0
5	B	243	0	0	0	0
All	All	8139	0	7655	75	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 5.

The worst 5 of 75 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:184:PHE:CE2	1:B:296:PHE:CZ	2.52	0.96
1:A:184:PHE:CE2	1:A:296:PHE:CZ	2.63	0.86
1:B:198:MET:HG3	4:B:504:CM5:H52	1.62	0.81
1:A:326:ARG:NH2	5:A:601:HOH:O	1.88	0.80
1:B:184:PHE:CE2	1:B:296:PHE:CE2	2.71	0.78

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	457/476 (96%)	447 (98%)	10 (2%)	0	<a href="#">100</a> <a href="#">100</a>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	456/476 (96%)	446 (98%)	10 (2%)	0	100	100
All	All	913/952 (96%)	893 (98%)	20 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	397/418 (95%)	395 (100%)	2 (0%)	91	90
1	B	399/418 (96%)	396 (99%)	3 (1%)	85	82
All	All	796/836 (95%)	791 (99%)	5 (1%)	89	87

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	297	PHE
1	A	309	TYR
1	B	57	ARG
1	B	297	PHE
1	B	309	TYR

Some 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

12 ligands 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	HEM	A	501	1	28,50,50	2.35	11 (39%)	17,82,82	2.54	9 (52%)
3	CAE	A	502[A]	-	11,11,11	1.35	1 (9%)	16,19,19	1.24	0
3	CAE	A	502[B]	-	11,11,11	1.34	1 (9%)	16,19,19	1.24	0
4	CM5	A	503	-	12,12,36	0.66	0	13,13,49	0.99	1 (7%)
4	CM5	A	504	-	36,36,36	3.18	19 (52%)	49,49,49	4.50	23 (46%)
4	CM5	A	505	-	12,12,36	0.73	0	13,13,49	0.70	0
2	HEM	B	501	1	28,50,50	2.16	10 (35%)	17,82,82	2.63	9 (52%)
3	CAE	B	502[A]	-	11,11,11	1.34	1 (9%)	16,19,19	1.24	0
3	CAE	B	502[B]	-	11,11,11	1.33	1 (9%)	16,19,19	1.24	0
4	CM5	B	503	-	12,12,36	0.56	0	13,13,49	0.88	0
4	CM5	B	504	-	36,36,36	3.62	17 (47%)	49,49,49	4.64	27 (55%)
4	CM5	B	505	-	12,12,36	0.92	0	13,13,49	0.89	0

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	HEM	A	501	1	-	0/6/54/54	0/0/8/8
3	CAE	A	502[A]	-	-	0/0/27/27	0/0/2/2
3	CAE	A	502[B]	-	-	0/0/27/27	0/0/2/2
4	CM5	A	503	-	-	0/6/14/65	0/1/1/3
4	CM5	A	504	-	-	0/17/65/65	0/3/3/3
4	CM5	A	505	-	-	0/6/14/65	0/1/1/3

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HEM	B	501	1	-	0/6/54/54	0/0/8/8
3	CAE	B	502[A]	-	-	0/0/27/27	0/0/2/2
3	CAE	B	502[B]	-	-	0/0/27/27	0/0/2/2
4	CM5	B	503	-	-	0/6/14/65	0/1/1/3
4	CM5	B	504	-	-	0/17/65/65	0/3/3/3
4	CM5	B	505	-	-	0/6/14/65	0/1/1/3

The worst 5 of 61 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	504	CM5	C17-C18	-11.50	1.23	1.52
4	B	504	CM5	O22-C18	-9.07	1.22	1.43
4	A	504	CM5	O34-C29	-8.10	1.24	1.43
4	B	504	CM5	C13-C18	-5.76	1.36	1.52
4	A	504	CM5	O23-C16	-5.72	1.30	1.43

The worst 5 of 69 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	504	CM5	O23-C24-C29	-13.63	77.40	108.11
4	A	504	CM5	O33-C28-C27	-13.17	81.69	110.36
4	A	504	CM5	O23-C24-C29	-13.01	78.80	108.11
4	B	504	CM5	O34-C29-C28	-7.63	93.76	110.36
4	A	504	CM5	O23-C16-C15	-7.42	91.09	109.34

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HEM	5	0
4	A	504	CM5	5	0
4	A	505	CM5	1	0
2	B	501	HEM	5	0
4	B	503	CM5	1	0
4	B	504	CM5	4	0

## 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	461/476 (96%)	-0.15	3 (0%) 87 91	14, 23, 40, 56	8 (1%)
1	B	460/476 (96%)	-0.17	3 (0%) 87 91	14, 24, 42, 55	8 (1%)
All	All	921/952 (96%)	-0.16	6 (0%) 87 91	14, 23, 42, 56	16 (1%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	39	LEU	2.6
1	A	132	MET	2.5
1	A	140	ARG	2.3
1	B	134	ASP	2.1
1	B	28	GLY	2.1

### 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	CAE	A	502[A]	10/10	0.73	0.51	23.26	20,29,47,59	10
3	CAE	A	502[B]	10/10	0.73	0.51	22.74	21,27,39,40	10
3	CAE	B	502[B]	10/10	0.67	0.49	19.84	21,29,43,50	10
3	CAE	B	502[A]	10/10	0.67	0.49	19.76	22,30,56,61	10
4	CM5	B	504	34/34	0.60	0.33	7.07	40,75,83,83	0
4	CM5	A	504	34/34	0.67	0.30	5.00	36,75,86,86	0
4	CM5	B	505	12/34	0.79	0.24	4.90	55,57,64,66	0
4	CM5	A	505	12/34	0.80	0.23	4.03	55,59,65,67	0
4	CM5	A	503	12/34	0.84	0.17	1.74	56,58,60,60	0
4	CM5	B	503	12/34	0.91	0.16	0.68	57,60,61,62	0
2	HEM	B	501	43/43	0.99	0.10	-0.07	9,14,19,29	0
2	HEM	A	501	43/43	0.99	0.09	-0.38	9,14,19,29	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.