



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 14, 2017 – 03:00 AM EDT

PDB ID : 2PG6  
Title : Crystal Structure of Human Microsomal P450 2A6 L240C/N297Q  
Authors : Sansen, S.; Hsu, M.H.; Stout, C.D.; Johnson, E.F.  
Deposited on : unknown  
Resolution : 2.53 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-20030345  
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-20030345

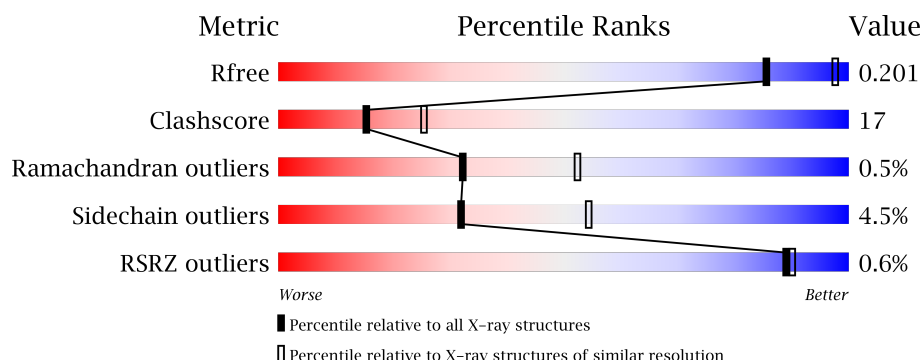
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.53 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	4636 (2.54-2.50)
Clashscore	112137	5382 (2.54-2.50)
Ramachandran outliers	110173	5282 (2.54-2.50)
Sidechain outliers	110143	5284 (2.54-2.50)
RSRZ outliers	101464	4669 (2.54-2.50)

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	
1	B	476	
1	C	476	
1	D	476	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 15496 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 2A6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	465	Total	C	N	O	S	0	0	0
			3759	2412	650	678	19			
1	B	464	Total	C	N	O	S	0	0	0
			3756	2410	650	677	19			
1	C	464	Total	C	N	O	S	0	0	0
			3750	2406	648	677	19			
1	D	464	Total	C	N	O	S	0	0	0
			3750	2406	648	677	19			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	MET	-	CLONING ARTIFACT	UNP P11509
A	24	ALA	-	CLONING ARTIFACT	UNP P11509
A	25	LYS	-	CLONING ARTIFACT	UNP P11509
A	26	LYS	-	CLONING ARTIFACT	UNP P11509
A	27	THR	-	CLONING ARTIFACT	UNP P11509
A	28	SER	-	CLONING ARTIFACT	UNP P11509
A	160	LEU	HIS	VARIANT	UNP P11509
A	240	CYS	LEU	ENGINEERED	UNP P11509
A	297	GLN	ASN	ENGINEERED	UNP P11509
A	495	HIS	-	EXPRESSION TAG	UNP P11509
A	496	HIS	-	EXPRESSION TAG	UNP P11509
A	497	HIS	-	EXPRESSION TAG	UNP P11509
A	498	HIS	-	EXPRESSION TAG	UNP P11509
B	23	MET	-	CLONING ARTIFACT	UNP P11509
B	24	ALA	-	CLONING ARTIFACT	UNP P11509
B	25	LYS	-	CLONING ARTIFACT	UNP P11509
B	26	LYS	-	CLONING ARTIFACT	UNP P11509
B	27	THR	-	CLONING ARTIFACT	UNP P11509
B	28	SER	-	CLONING ARTIFACT	UNP P11509
B	160	LEU	HIS	VARIANT	UNP P11509
B	240	CYS	LEU	ENGINEERED	UNP P11509

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	297	GLN	ASN	ENGINEERED	UNP P11509
B	495	HIS	-	EXPRESSION TAG	UNP P11509
B	496	HIS	-	EXPRESSION TAG	UNP P11509
B	497	HIS	-	EXPRESSION TAG	UNP P11509
B	498	HIS	-	EXPRESSION TAG	UNP P11509
C	23	MET	-	CLONING ARTIFACT	UNP P11509
C	24	ALA	-	CLONING ARTIFACT	UNP P11509
C	25	LYS	-	CLONING ARTIFACT	UNP P11509
C	26	LYS	-	CLONING ARTIFACT	UNP P11509
C	27	THR	-	CLONING ARTIFACT	UNP P11509
C	28	SER	-	CLONING ARTIFACT	UNP P11509
C	160	LEU	HIS	VARIANT	UNP P11509
C	240	CYS	LEU	ENGINEERED	UNP P11509
C	297	GLN	ASN	ENGINEERED	UNP P11509
C	495	HIS	-	EXPRESSION TAG	UNP P11509
C	496	HIS	-	EXPRESSION TAG	UNP P11509
C	497	HIS	-	EXPRESSION TAG	UNP P11509
C	498	HIS	-	EXPRESSION TAG	UNP P11509
D	23	MET	-	CLONING ARTIFACT	UNP P11509
D	24	ALA	-	CLONING ARTIFACT	UNP P11509
D	25	LYS	-	CLONING ARTIFACT	UNP P11509
D	26	LYS	-	CLONING ARTIFACT	UNP P11509
D	27	THR	-	CLONING ARTIFACT	UNP P11509
D	28	SER	-	CLONING ARTIFACT	UNP P11509
D	160	LEU	HIS	VARIANT	UNP P11509
D	240	CYS	LEU	ENGINEERED	UNP P11509
D	297	GLN	ASN	ENGINEERED	UNP P11509
D	495	HIS	-	EXPRESSION TAG	UNP P11509
D	496	HIS	-	EXPRESSION TAG	UNP P11509
D	497	HIS	-	EXPRESSION TAG	UNP P11509
D	498	HIS	-	EXPRESSION TAG	UNP P11509

- 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 43	C 34	Fe 1	N 4	O 4	0	0
2	B	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	C	1	Total 43	C 34	Fe 1	N 4	O 4	0	0
2	D	1	Total 43	C 34	Fe 1	N 4	O 4	0	0

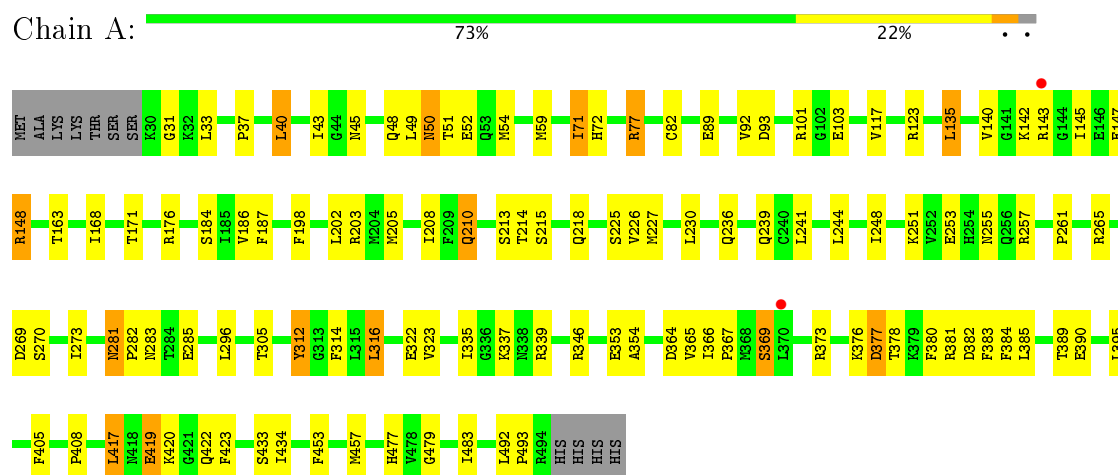
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	76	Total	O	0	0
			76	76		
3	B	65	Total	O	0	0
			65	65		
3	C	82	Total	O	0	0
			82	82		
3	D	86	Total	O	0	0
			86	86		

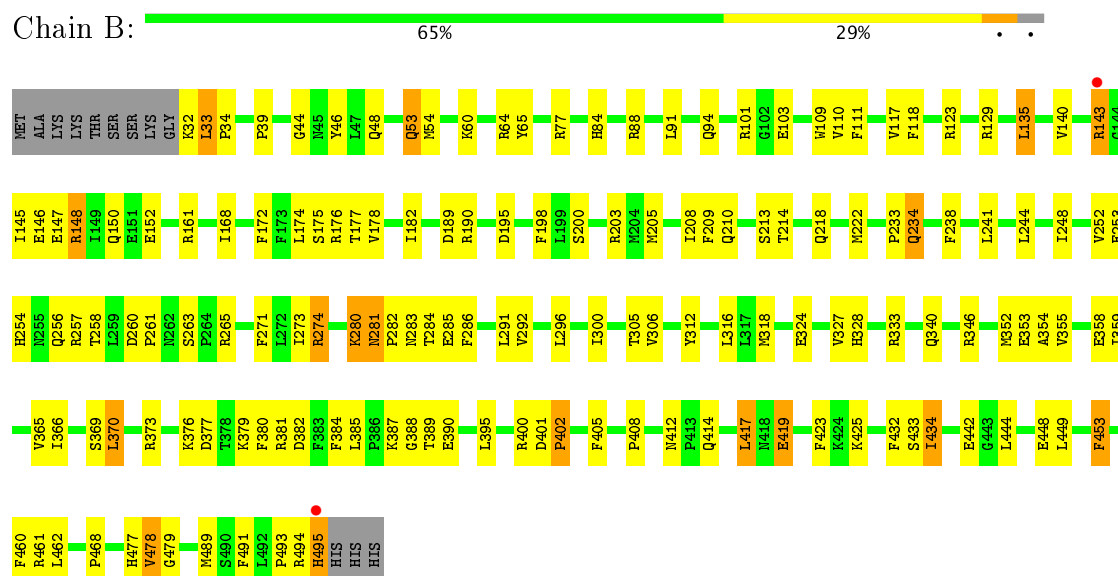
### 3 Residue-property plots

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: Cytochrome P450 2A6

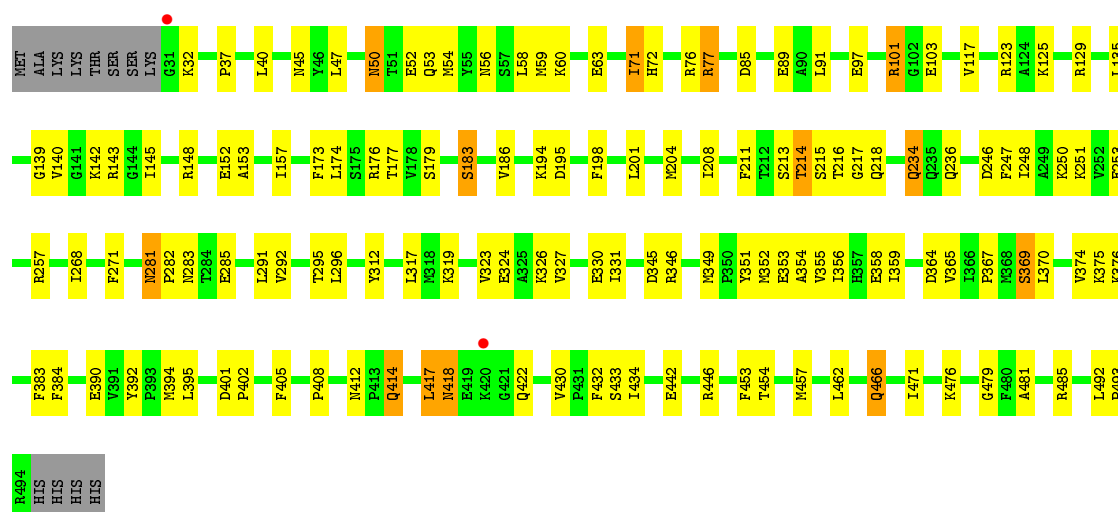


#### • Molecule 1: Cytochrome P450 2A6

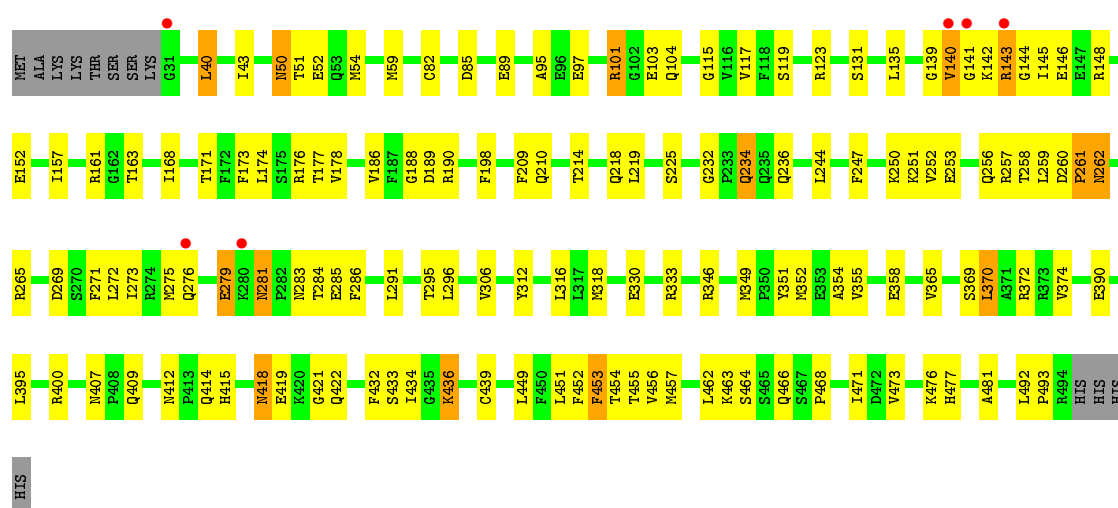


#### • Molecule 1: Cytochrome P450 2A6





• Molecule 1: Cytochrome P450 2A6



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.70 Å 158.25 Å 103.76 Å 90.00° 92.09° 90.00°	Depositor
Resolution (Å)	40.00 – 2.53 47.51 – 2.53	Depositor EDS
% Data completeness (in resolution range)	90.1 (40.00-2.53) 93.0 (47.51-2.53)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.59 (at 2.51 Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.209 , 0.261 0.207 , 0.201	Depositor DCC
$R_{free}$ test set	3536 reflections (5.25%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.0	Xtriage
Anisotropy	0.443	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 38.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	15496	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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

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	0.44	0/3850	0.61	0/5184
1	B	0.44	0/3848	0.62	0/5183
1	C	0.43	0/3841	0.62	0/5173
1	D	0.43	0/3841	0.63	1/5173 (0.0%)
All	All	0.44	0/15380	0.62	1/20713 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	104	GLN	N-CA-C	-5.13	97.16	111.00

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	3759	0	3743	105	0
1	B	3756	0	3734	135	0
1	C	3750	0	3730	129	0
1	D	3750	0	3730	136	0
2	A	43	0	30	2	0
2	B	43	0	30	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	43	0	30	3	0
2	D	43	0	30	3	0
3	A	76	0	0	2	0
3	B	65	0	0	1	1
3	C	82	0	0	3	1
3	D	86	0	0	9	0
All	All	15496	0	15057	501	1

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 17.

The worst 5 of 501 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:D:54:MET:HG3	1:D:218:GLN:HE21	1.16	1.10
1:D:414:GLN:HE21	1:D:414:GLN:HA	1.27	0.99
1:C:54:MET:HG3	1:C:218:GLN:HE21	1.25	0.98
1:B:54:MET:HG3	1:B:218:GLN:HE21	1.29	0.96
1:C:32:LYS:HB3	1:C:384:PHE:HB3	1.48	0.95

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:547:HOH:O	3:C:572:HOH:O[1_554]	1.98	0.22

## 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	463/476 (97%)	437 (94%)	23 (5%)	3 (1%)	28 47

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	462/476 (97%)	427 (92%)	33 (7%)	2 (0%)	38	58
1	C	462/476 (97%)	440 (95%)	21 (4%)	1 (0%)	51	72
1	D	462/476 (97%)	433 (94%)	26 (6%)	3 (1%)	28	47
All	All	1849/1904 (97%)	1737 (94%)	103 (6%)	9 (0%)	32	52

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	140	VAL
1	D	418	ASN
1	D	261	PRO
1	A	369	SER
1	C	369	SER

### 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	412/422 (98%)	398 (97%)	14 (3%)	42	68
1	B	412/422 (98%)	388 (94%)	24 (6%)	23	41
1	C	411/422 (97%)	391 (95%)	20 (5%)	29	50
1	D	411/422 (97%)	395 (96%)	16 (4%)	37	62
All	All	1646/1688 (98%)	1572 (96%)	74 (4%)	32	54

5 of 74 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	419	GLU
1	C	77	ARG
1	D	281	ASN
1	B	434	ILE
1	B	495	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	409	GLN
1	C	94	GLN
1	D	283	ASN
1	B	414	GLN
1	C	50	ASN

### 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](#)

4 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	500	1	28,50,50	2.10	13 (46%)	17,82,82	1.65	4 (23%)
2	HEM	B	500	1	28,50,50	2.26	15 (53%)	17,82,82	1.73	5 (29%)
2	HEM	C	500	1	28,50,50	2.20	15 (53%)	17,82,82	1.74	5 (29%)
2	HEM	D	500	1	28,50,50	2.25	12 (42%)	17,82,82	1.74	5 (29%)

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	500	1	-	0/6/54/54	0/0/8/8
2	HEM	B	500	1	-	0/6/54/54	0/0/8/8
2	HEM	C	500	1	-	0/6/54/54	0/0/8/8
2	HEM	D	500	1	-	0/6/54/54	0/0/8/8

The worst 5 of 55 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	500	HEM	C3C-CAC	-5.39	1.37	1.47
2	A	500	HEM	C3C-CAC	-5.06	1.37	1.47
2	B	500	HEM	C3C-CAC	-5.03	1.37	1.47
2	C	500	HEM	C3C-CAC	-4.72	1.38	1.47
2	C	500	HEM	C1C-NC	-3.82	1.32	1.36

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	500	HEM	C4C-C3C-C2C	-3.33	104.58	106.90
2	B	500	HEM	C4C-C3C-C2C	-3.11	104.73	106.90
2	A	500	HEM	C4C-C3C-C2C	-3.01	104.80	106.90
2	C	500	HEM	C4C-C3C-C2C	-2.95	104.83	106.90
2	B	500	HEM	CMA-C3A-C4A	-2.11	125.21	128.46

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	500	HEM	2	0
2	B	500	HEM	1	0
2	C	500	HEM	3	0
2	D	500	HEM	3	0

## 5.7 Other polymers ⓘ

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	465/476 (97%)	-0.26	2 (0%) 92 93	19, 31, 44, 53	1 (0%)
1	B	464/476 (97%)	-0.17	2 (0%) 92 93	23, 36, 49, 63	0
1	C	464/476 (97%)	-0.39	2 (0%) 92 93	21, 32, 44, 54	0
1	D	464/476 (97%)	-0.30	6 (1%) 77 79	20, 33, 49, 60	0
All	All	1857/1904 (97%)	-0.28	12 (0%) 89 90	19, 33, 47, 63	1 (0%)

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	31	GLY	5.0
1	D	140	VAL	3.4
1	A	370	LEU	3.3
1	B	495	HIS	3.3
1	D	141	GLY	2.7

### 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
2	HEM	C	500	43/43	0.97	0.15	1.21	21,24,29,33	0
2	HEM	D	500	43/43	0.98	0.14	0.17	22,26,27,28	0
2	HEM	B	500	43/43	0.98	0.14	-0.31	23,27,29,31	0
2	HEM	A	500	43/43	0.98	0.13	-0.34	19,21,23,27	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.