



# wwPDB X-ray Structure Validation Summary Report

Jan 31, 2016 – 09:30 PM GMT

PDB ID : 1P81  
Title : Crystal structure of the D181E variant of catalase HP11 from E. coli  
Authors : Chelikani, P.; Carpena, X.; Fita, I.; Loewen, P.C.  
Deposited on : 2003-05-06  
Resolution : 1.81 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

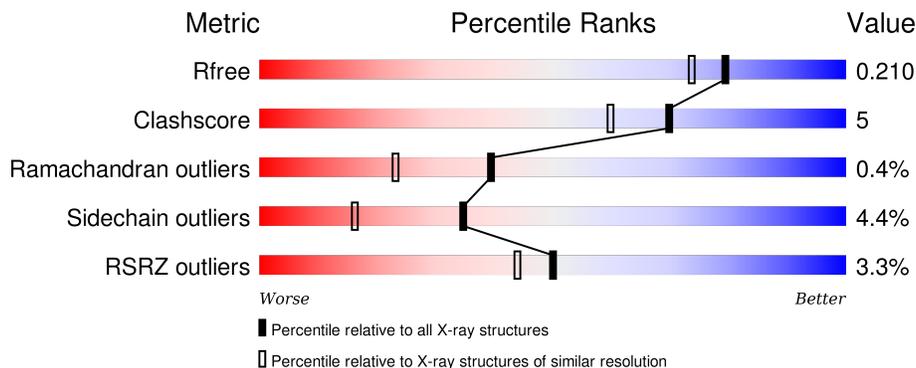
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.81 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))
$R_{free}$	91344	5422 (1.84-1.80)
Clashscore	102246	6347 (1.84-1.80)
Ramachandran outliers	100387	6276 (1.84-1.80)
Sidechain outliers	100360	6276 (1.84-1.80)
RSRZ outliers	91569	5439 (1.84-1.80)

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	753	 3% 86% 10% ..
1	B	753	 4% 81% 13% ..
1	C	753	 3% 84% 10% ..
1	D	753	 4% 81% 14% ...

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 26236 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.

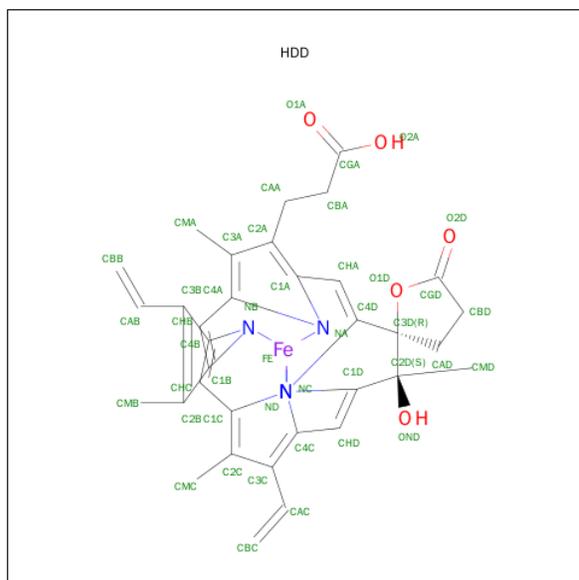
- Molecule 1 is a protein called Catalase HP11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	727	5747	3648	1005	1082	12	0	0	0
1	B	727	5747	3648	1005	1082	12	0	0	0
1	C	727	5747	3648	1005	1082	12	0	0	0
1	D	727	5747	3648	1005	1082	12	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	181	GLU	ASP	ENGINEERED	UNP P21179
B	181	GLU	ASP	ENGINEERED	UNP P21179
C	181	GLU	ASP	ENGINEERED	UNP P21179
D	181	GLU	ASP	ENGINEERED	UNP P21179

- Molecule 2 is CIS-HEME D HYDROXYCHLORIN GAMMA-SPIROLACTONE (three-letter code: HDD) (formula:  $C_{34}H_{32}FeN_4O_5$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
2	A	1	Total	C	Fe	N	O	0	0
			44	34	1	4	5		
2	B	1	Total	C	Fe	N	O	0	0
			44	34	1	4	5		
2	C	1	Total	C	Fe	N	O	0	0
			44	34	1	4	5		
2	D	1	Total	C	Fe	N	O	0	0
			44	34	1	4	5		

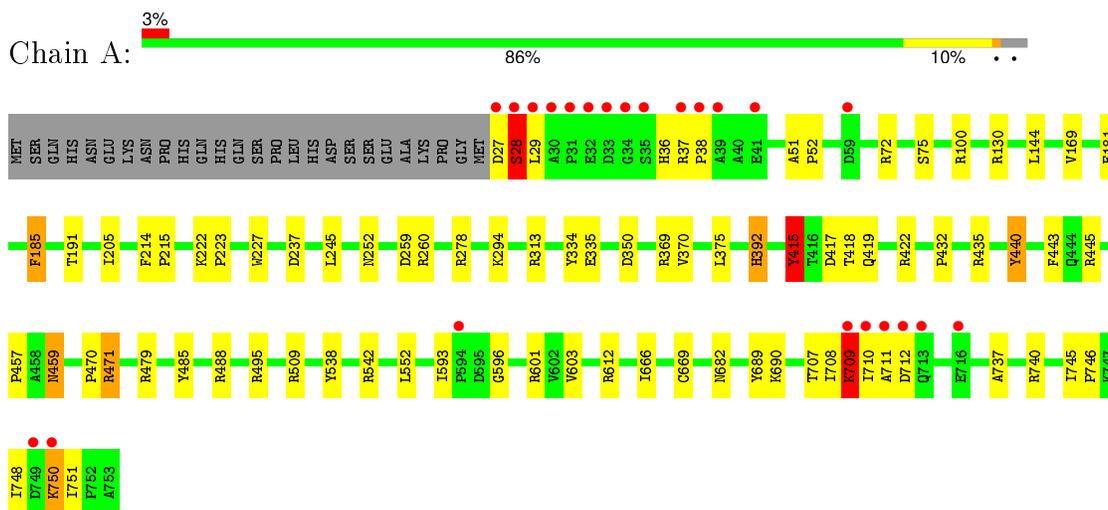
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	815	Total	O	0	0
			815	815		
3	B	702	Total	O	0	0
			702	702		
3	C	768	Total	O	0	0
			768	768		
3	D	787	Total	O	0	0
			787	787		

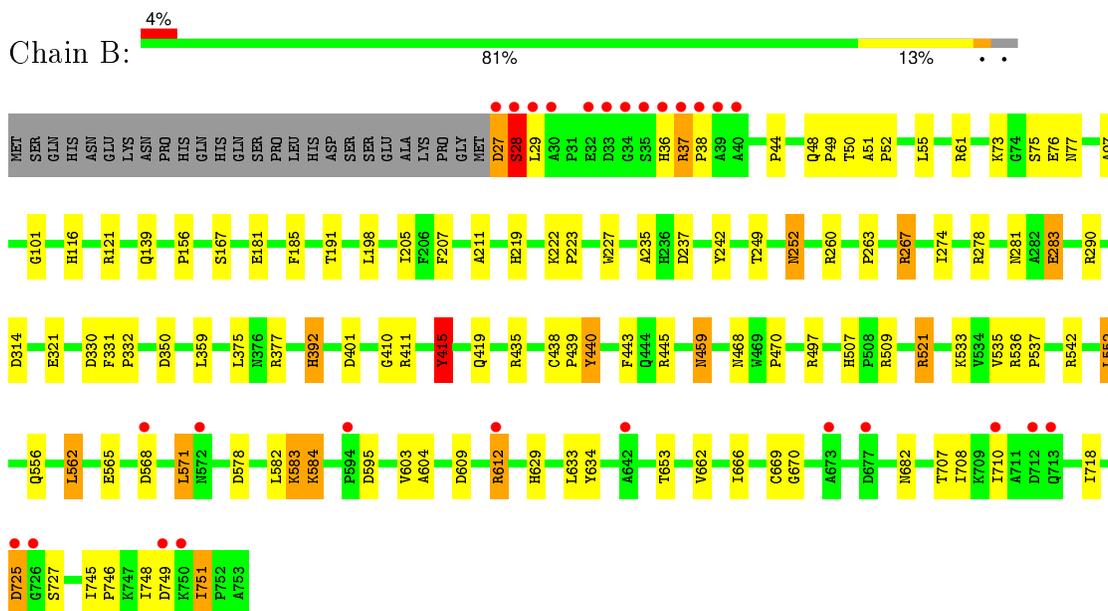
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Catalase HPII

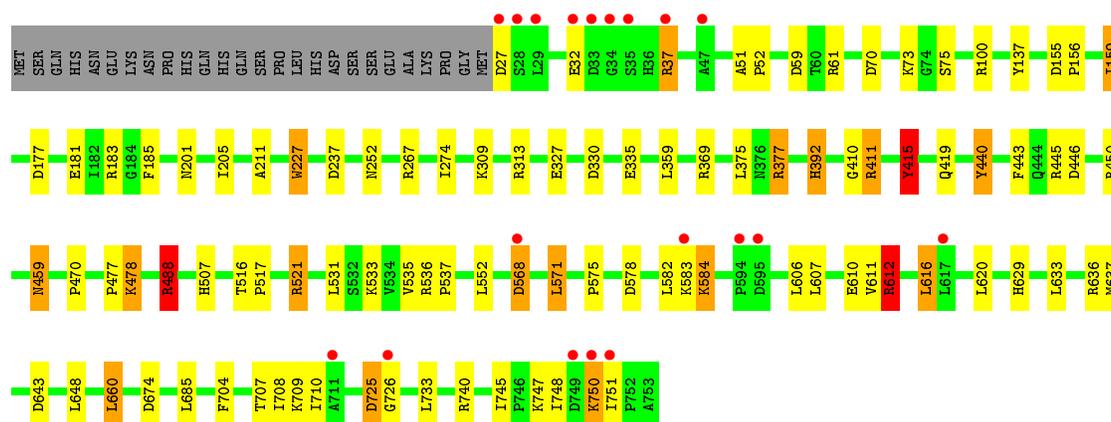


- Molecule 1: Catalase HPII



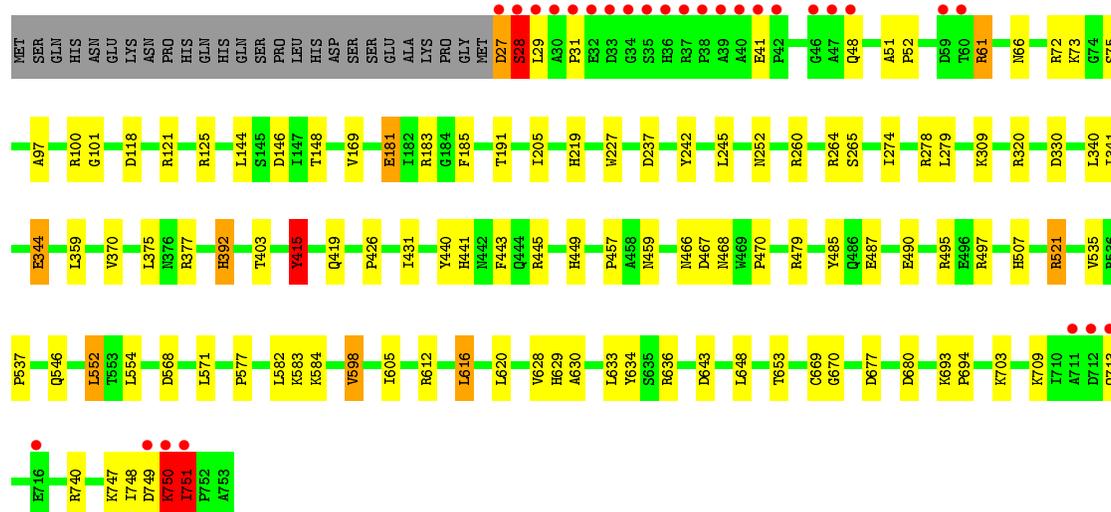
- Molecule 1: Catalase HPII

Chain C: 3% 84% 10%



• Molecule 1: Catalase HPII

Chain D: 4% 81% 14%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.38Å 132.86Å 122.04Å 90.00° 109.50° 90.00°	Depositor
Resolution (Å)	29.90 – 1.81 17.97 – 1.81	Depositor EDS
% Data completeness (in resolution range)	99.2 (29.90-1.81) 98.7 (17.97-1.81)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 1.81Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.172 , 0.216 0.171 , 0.210	Depositor DCC
$R_{free}$ test set	12575 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.9	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 61.1	EDS
Estimated twinning fraction	0.014 for h,-k,-h-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	5 of 251911 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	26236	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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, 0.375 respectively for untwinned datasets, and 0.333, 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:  
HDD

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	0.58	0/5903	1.24	31/8025 (0.4%)
1	B	0.56	0/5903	1.23	34/8025 (0.4%)
1	C	0.57	0/5903	1.23	28/8025 (0.3%)
1	D	0.58	0/5903	1.33	47/8025 (0.6%)
All	All	0.57	0/23612	1.26	140/32100 (0.4%)

There are no bond length outliers.

The worst 5 of 140 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	612	ARG	CD-NE-CZ	30.17	165.83	123.60
1	B	521	ARG	NE-CZ-NH1	14.58	127.59	120.30
1	D	740	ARG	NE-CZ-NH1	12.87	126.74	120.30
1	D	61	ARG	NE-CZ-NH1	-12.16	114.22	120.30
1	D	61	ARG	NE-CZ-NH2	11.55	126.08	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	5747	0	5580	43	0
1	B	5747	0	5580	66	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	5747	0	5580	52	0
1	D	5747	0	5580	52	0
2	A	44	0	31	5	0
2	B	44	0	31	4	0
2	C	44	0	31	6	0
2	D	44	0	31	2	0
3	A	815	0	0	6	1
3	B	702	0	0	6	1
3	C	768	0	0	9	0
3	D	787	0	0	7	0
All	All	26236	0	22444	206	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 5.

The worst 5 of 206 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:392:HIS:ND1	1:B:415:TYR:HB2	1.38	1.39
1:A:392:HIS:ND1	1:A:415:TYR:HB2	1.40	1.35
1:C:392:HIS:ND1	1:C:415:TYR:HB2	1.42	1.32
1:D:392:HIS:ND1	1:D:415:TYR:HB2	1.39	1.31
1:C:392:HIS:ND1	1:C:415:TYR:CB	2.33	0.89

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:A:1050:HOH:O	3:B:3904:HOH:O[2_747]	1.45	0.75

## 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	725/753 (96%)	699 (96%)	22 (3%)	4 (1%)	30	14
1	B	725/753 (96%)	699 (96%)	22 (3%)	4 (1%)	30	14
1	C	725/753 (96%)	703 (97%)	21 (3%)	1 (0%)	56	38
1	D	725/753 (96%)	700 (97%)	21 (3%)	4 (1%)	30	14
All	All	2900/3012 (96%)	2801 (97%)	86 (3%)	13 (0%)	39	23

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	28	SER
1	A	709	LYS
1	B	28	SER
1	D	28	SER
1	D	750	LYS

### 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	612/636 (96%)	595 (97%)	17 (3%)	51	34
1	B	612/636 (96%)	584 (95%)	28 (5%)	33	15
1	C	612/636 (96%)	579 (95%)	33 (5%)	27	10
1	D	612/636 (96%)	582 (95%)	30 (5%)	31	13
All	All	2448/2544 (96%)	2340 (96%)	108 (4%)	35	16

5 of 108 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	181	GLU
1	C	488	ARG
1	D	598	VAL
1	C	205	ILE
1	C	375	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	252	ASN
1	C	556	GLN
1	D	629	HIS
1	C	459	ASN
1	C	507	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](#)

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	HDD	A	760	1,3	30,52,52	3.65	13 (43%)	20,89,89	3.95	13 (65%)
2	HDD	B	760	1,3	30,52,52	3.43	13 (43%)	20,89,89	4.54	15 (75%)
2	HDD	C	760	1,3	30,52,52	3.63	13 (43%)	20,89,89	3.90	14 (70%)
2	HDD	D	760	1,3	30,52,52	3.68	11 (36%)	20,89,89	4.37	12 (60%)

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	HDD	A	760	1,3	-	0/3/89/89	0/1/9/9
2	HDD	B	760	1,3	-	0/3/89/89	0/1/9/9
2	HDD	C	760	1,3	-	0/3/89/89	0/1/9/9
2	HDD	D	760	1,3	-	0/3/89/89	0/1/9/9

The worst 5 of 50 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	760	HDD	C2D-C3D	-10.77	1.27	1.54
2	C	760	HDD	C2D-C3D	-10.66	1.28	1.54
2	A	760	HDD	C2D-C3D	-10.26	1.29	1.54
2	D	760	HDD	C2D-C3D	-10.24	1.29	1.54
2	C	760	HDD	CAD-C3D	-9.79	1.35	1.53

The worst 5 of 54 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	760	HDD	OND-C2D-CMD	-13.43	85.17	109.41
2	D	760	HDD	OND-C2D-CMD	-11.82	88.06	109.41
2	C	760	HDD	OND-C2D-CMD	-7.93	95.10	109.41
2	A	760	HDD	O1D-CGD-CBD	-7.14	101.83	110.20
2	B	760	HDD	C3B-C4B-NB	-7.03	97.68	110.94

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	760	HDD	5	0
2	B	760	HDD	4	0
2	C	760	HDD	6	0
2	D	760	HDD	2	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	727/753 (96%)	-0.41	23 (3%) 51 45	7, 13, 30, 55	1 (0%)
1	B	727/753 (96%)	-0.29	27 (3%) 45 39	7, 15, 36, 50	1 (0%)
1	C	727/753 (96%)	-0.31	19 (2%) 59 54	8, 15, 34, 47	1 (0%)
1	D	727/753 (96%)	-0.38	28 (3%) 43 37	7, 14, 32, 48	1 (0%)
All	All	2908/3012 (96%)	-0.35	97 (3%) 50 44	7, 14, 34, 55	4 (0%)

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	27	ASP	10.8
1	C	28	SER	9.5
1	A	28	SER	9.4
1	C	27	ASP	8.6
1	A	711	ALA	8.3

### 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	HDD	B	760	44/44	0.98	0.07	0.30	7,10,15,18	0
2	HDD	C	760	44/44	0.98	0.07	0.05	8,11,15,16	0
2	HDD	D	760	44/44	0.98	0.07	0.05	7,9,15,16	0
2	HDD	A	760	44/44	0.98	0.06	-0.29	7,9,14,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.