



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 26, 2023 – 06:22 PM EDT

PDB ID : 3GE8  
Title : Toluene 4-monooxygenase HD T201A diferric, resting state complex  
Authors : Elsen, N.L.; Bailey, L.J.; Hauser, A.D.; Fox, B.G.  
Deposited on : 2009-02-25  
Resolution : 2.19 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

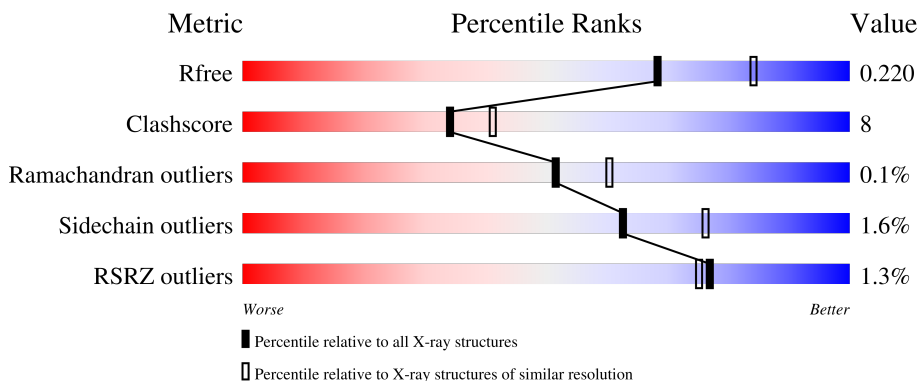
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.19 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.



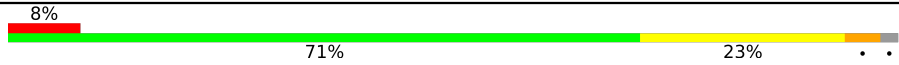


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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  $\geq 3$ , 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	500	82% 15% ..
1	D	500	82% 15% ..
2	B	327	77% 15% • 6%
2	F	327	81% 13% 7%
3	C	84	85% 11% ..

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	G	84	 <p>8% 71% 23% . .</p>
4	E	103	 <p>% 79% 19% ..</p>
4	H	103	 <p>81% 16% ..</p>

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 17621 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 Toluene-4-monooxygenase system protein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	491	Total	C	N	O	S	0	13	0
			4085	2619	690	751	25			
1	D	490	Total	C	N	O	S	0	13	0
			4055	2601	683	747	24			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	ALA	THR	engineered mutation	UNP Q6Q8Q7
D	201	ALA	THR	engineered mutation	UNP Q6Q8Q7

- Molecule 2 is a protein called Toluene-4-monooxygenase system protein E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	306	Total	C	N	O	S	0	8	0
			2564	1624	446	478	16			
2	F	305	Total	C	N	O	S	0	7	0
			2534	1604	438	476	16			

- Molecule 3 is a protein called Toluene-4-monooxygenase system protein B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	82	Total	C	N	O	S	0	1	0
			654	412	117	121	4			
3	G	82	Total	C	N	O	S	0	2	0
			667	419	119	124	5			

- Molecule 4 is a protein called Toluene-4-monooxygenase system protein D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	102	Total	C	N	O	S	0	4	0
			806	503	140	161	2			

*Continued on next page...*

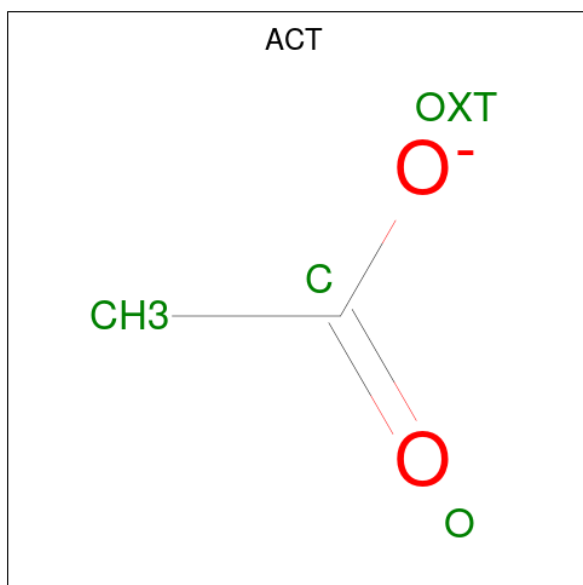
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	101	Total	C	N	O	S	0	4	0
			801	500	139	160	2			

- Molecule 5 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Fe	0	0
			2	2		
5	D	2	Total	Fe	0	0
			2	2		

- Molecule 6 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			4	2	2		
6	D	1	Total	C	O	0	0
			4	2	2		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	344	Total	O	0	0
			344	344		
7	B	229	Total	O	0	0
			229	229		

Continued on next page...

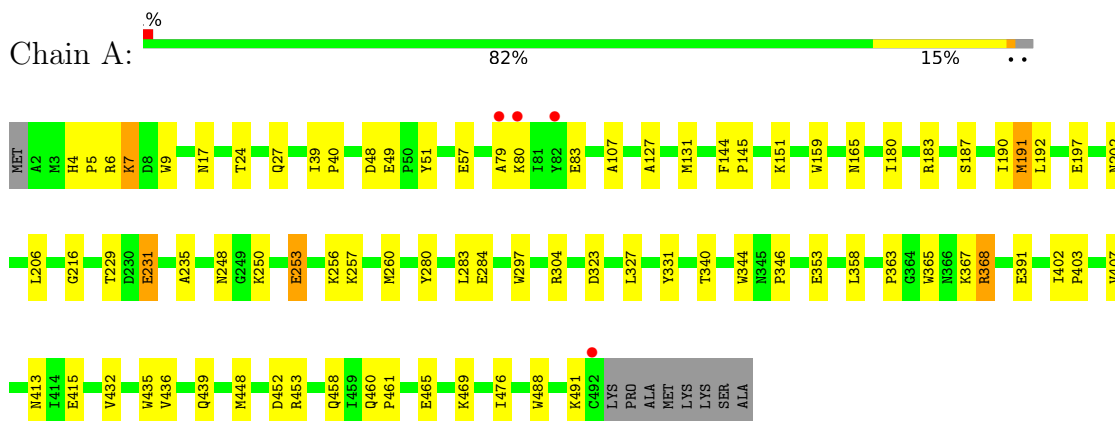
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
7	C	48	Total 48	O 48	0	0
7	E	103	Total 103	O 103	0	0
7	D	337	Total 337	O 337	0	0
7	F	243	Total 243	O 243	0	0
7	G	46	Total 46	O 46	0	0
7	H	93	Total 93	O 93	0	0

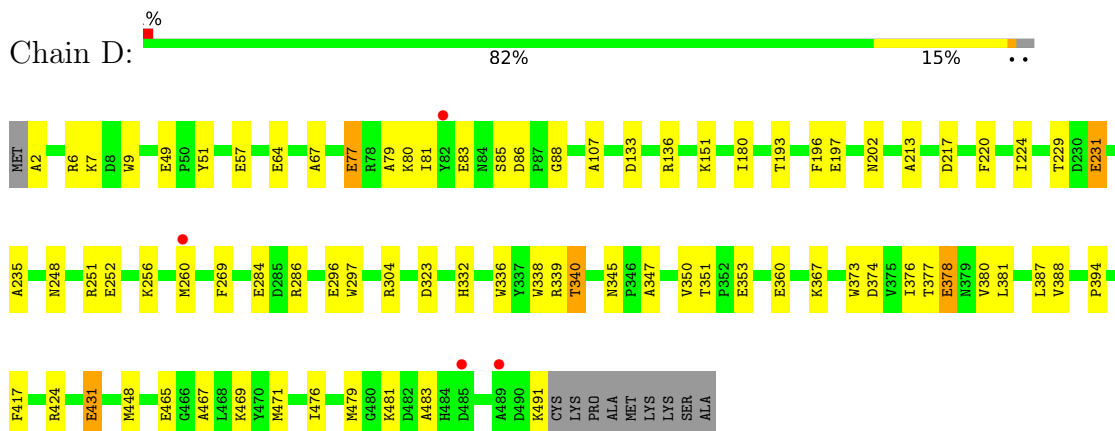
### 3 Residue-property plots

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.

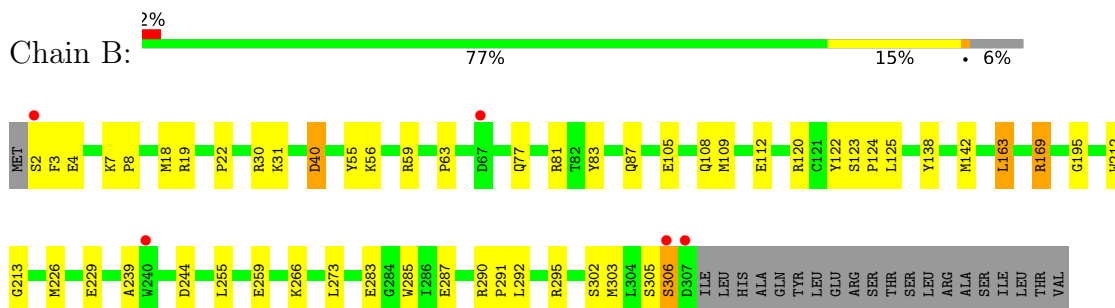
- Molecule 1: Toluene-4-monooxygenase system protein A



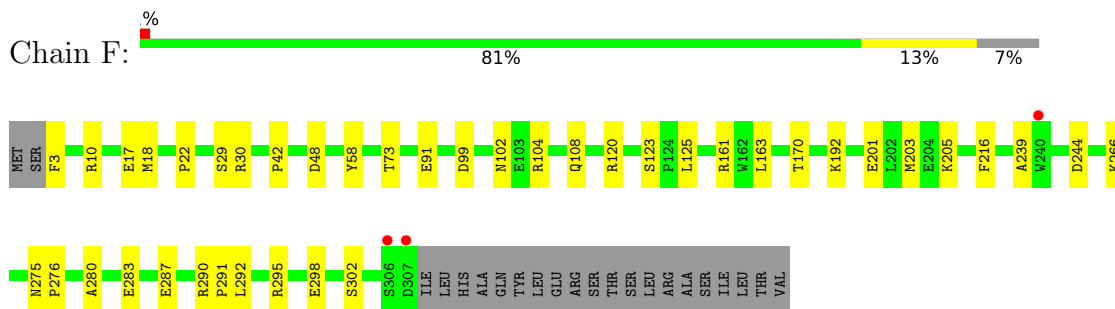
- Molecule 1: Toluene-4-monooxygenase system protein A



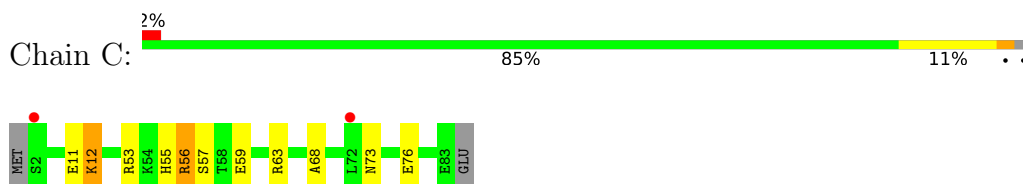
- Molecule 2: Toluene-4-monooxygenase system protein E



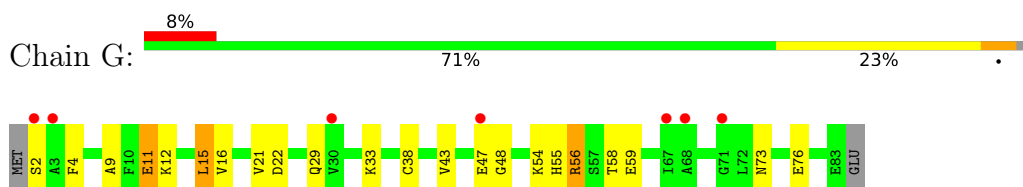
- Molecule 2: Toluene-4-monooxygenase system protein E



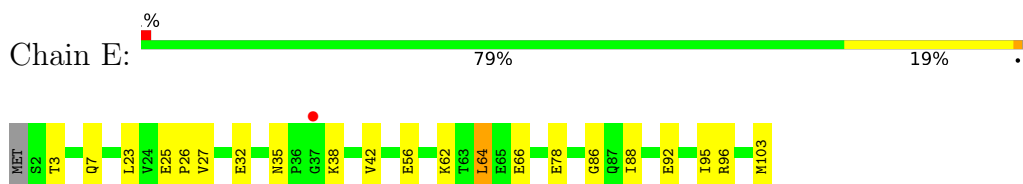
- Molecule 3: Toluene-4-monooxygenase system protein B



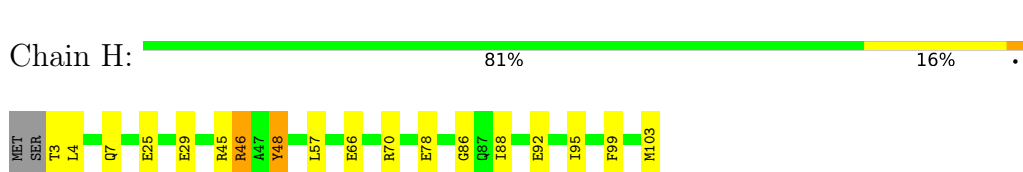
- Molecule 3: Toluene-4-monooxygenase system protein B



- Molecule 4: Toluene-4-monooxygenase system protein D



- Molecule 4: Toluene-4-monooxygenase system protein D





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.29Å 115.39Å 180.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.17 – 2.19 47.17 – 2.19	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.17-2.19) 99.6 (47.17-2.19)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.42 (at 2.18Å)	Xtrriage
Refinement program	REFMAC 5.5.0066	Depositor
R, $R_{free}$	0.161 , 0.229 0.155 , 0.220	Depositor DCC
$R_{free}$ test set	5414 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.3	Xtrriage
Anisotropy	0.090	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 48.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	17621	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 74.76 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.4380e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: ACT, 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	1.06	1/4217 (0.0%)	0.90	3/5722 (0.1%)
1	D	1.12	5/4190 (0.1%)	0.94	8/5690 (0.1%)
2	B	1.02	2/2646 (0.1%)	0.85	1/3594 (0.0%)
2	F	1.08	2/2609 (0.1%)	0.87	2/3546 (0.1%)
3	C	1.00	0/666	0.85	1/902 (0.1%)
3	G	1.13	1/679 (0.1%)	0.84	0/920
4	E	1.06	1/816 (0.1%)	1.00	1/1104 (0.1%)
4	H	1.05	1/811 (0.1%)	0.90	1/1097 (0.1%)
All	All	1.07	13/16634 (0.1%)	0.90	17/22575 (0.1%)

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	378	GLU	CG-CD	7.26	1.62	1.51
2	F	216	PHE	CE1-CZ	6.18	1.49	1.37
1	D	296	GLU	CG-CD	6.16	1.61	1.51
2	B	55	TYR	CD2-CE2	5.86	1.48	1.39
4	H	48	TYR	CD2-CE2	5.82	1.48	1.39

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	424	ARG	NE-CZ-NH2	-9.61	115.49	120.30
4	E	64	LEU	CA-CB-CG	8.84	135.62	115.30
1	D	286	ARG	NE-CZ-NH1	-7.16	116.72	120.30
1	A	368	ARG	NE-CZ-NH2	-6.77	116.92	120.30
2	F	99	ASP	CB-CG-OD1	6.55	124.19	118.30

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

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	4085	0	3848	66	0
1	D	4055	0	3811	67	0
2	B	2564	0	2440	47	0
2	F	2534	0	2404	30	0
3	C	654	0	649	9	0
3	G	667	0	659	21	0
4	E	806	0	782	20	1
4	H	801	0	787	27	0
5	A	2	0	0	0	0
5	D	2	0	0	0	0
6	A	4	0	3	0	0
6	D	4	0	3	0	0
7	A	344	0	0	14	0
7	B	229	0	0	10	0
7	C	48	0	0	4	0
7	D	337	0	0	23	1
7	E	103	0	0	7	0
7	F	243	0	0	12	0
7	G	46	0	0	3	0
7	H	93	0	0	7	0
All	All	17621	0	15386	264	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 8.

The worst 5 of 264 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:80:LYS:HD3	4:H:46:ARG:NH1	1.39	1.32
1:D:378:GLU:HB3	7:D:1172:HOH:O	1.35	1.25
2:B:109[B]:MET:HE2	7:B:1673:HOH:O	1.37	1.20
2:B:108:GLN:HG3	7:B:986:HOH:O	1.48	1.13
1:A:83:GLU:OE1	1:A:151:LYS:HE2	1.47	1.12

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:E:62:LYS:NZ	7:D:619:HOH:O[4_545]	1.39	0.81

## 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	496/500 (99%)	476 (96%)	20 (4%)	0	100	100
1	D	493/500 (99%)	475 (96%)	17 (3%)	1 (0%)	47	55
2	B	308/327 (94%)	303 (98%)	5 (2%)	0	100	100
2	F	304/327 (93%)	300 (99%)	4 (1%)	0	100	100
3	C	80/84 (95%)	74 (92%)	6 (8%)	0	100	100
3	G	82/84 (98%)	75 (92%)	7 (8%)	0	100	100
4	E	100/103 (97%)	99 (99%)	1 (1%)	0	100	100
4	H	99/103 (96%)	97 (98%)	2 (2%)	0	100	100
All	All	1962/2028 (97%)	1899 (97%)	62 (3%)	1 (0%)	51	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	79	ALA

### 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	422/422 (100%)	415 (98%)	7 (2%)	60	74
1	D	418/422 (99%)	412 (99%)	6 (1%)	67	80
2	B	279/296 (94%)	274 (98%)	5 (2%)	59	72
2	F	276/296 (93%)	271 (98%)	5 (2%)	59	72
3	C	73/75 (97%)	71 (97%)	2 (3%)	44	57
3	G	75/75 (100%)	73 (97%)	2 (3%)	44	57
4	E	85/87 (98%)	84 (99%)	1 (1%)	71	83
4	H	85/87 (98%)	85 (100%)	0	100	100
All	All	1713/1760 (97%)	1685 (98%)	28 (2%)	62	76

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	E	64	LEU
3	G	56	ARG
1	D	284	GLU
2	F	120	ARG
1	D	231	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	123	ASN
1	A	382	ASN
4	E	89	GLN
1	D	123	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 for Mogul analysis.

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
6	ACT	D	1428	5	3,3,3	0.78	0	3,3,3	1.79	1 (33%)
6	ACT	A	1428	5	3,3,3	0.75	0	3,3,3	2.18	2 (66%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	1428	ACT	OXT-C-CH3	2.78	126.68	115.18
6	A	1428	ACT	OXT-C-O	-2.49	112.86	122.05
6	D	1428	ACT	OXT-C-CH3	2.36	124.92	115.18

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](#)

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	491/500 (98%)	-0.45	4 (0%) 86 85	3, 13, 26, 48	0
1	D	490/500 (98%)	-0.44	4 (0%) 86 85	4, 13, 26, 44	0
2	B	306/327 (93%)	-0.31	5 (1%) 72 70	5, 18, 34, 49	0
2	F	305/327 (93%)	-0.42	3 (0%) 82 81	4, 14, 29, 52	0
3	C	82/84 (97%)	0.28	2 (2%) 59 56	12, 24, 35, 44	0
3	G	82/84 (97%)	0.57	7 (8%) 10 9	13, 25, 36, 42	0
4	E	102/103 (99%)	-0.61	1 (0%) 82 81	6, 13, 28, 34	0
4	H	101/103 (98%)	-0.65	0 100 100	7, 15, 31, 36	0
All	All	1959/2028 (96%)	-0.37	26 (1%) 77 75	3, 15, 32, 52	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	307	ASP	5.2
3	G	2	SER	4.4
1	D	82	TYR	3.7
2	F	240	TRP	3.0
2	B	307	ASP	3.0

### 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](#)

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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
6	ACT	A	1428	4/4	0.97	0.17	14,16,16,17	0
6	ACT	D	1428	4/4	0.99	0.11	13,13,14,15	0
5	FE	D	501	1/1	1.00	0.06	13,13,13,13	0
5	FE	D	502	1/1	1.00	0.09	8,8,8,8	0
5	FE	A	501	1/1	1.00	0.10	8,8,8,8	0
5	FE	A	502	1/1	1.00	0.05	12,12,12,12	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.