



# wwPDB X-ray Structure Validation Summary Report

May 15, 2020 – 09:21 pm BST

PDB ID : 5O7O  
Title : The crystal structure of DfoC, the desferrioxamine biosynthetic pathway acetyltransferase/Non-Ribosomal Peptide Synthetase (NRPS)-Independent Siderophore (NIS) from the fire blight disease pathogen *Erwinia amylovora*  
Authors : Salomone-Stagni, M.; Bartho, J.D.; Polsinelli, I.; Bellini, D.; Walsh, M.A.; Demitri, N.; Benini, S.  
Deposited on : 2017-06-09  
Resolution : 2.11 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 following versions of software and data (see [references](#) ) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

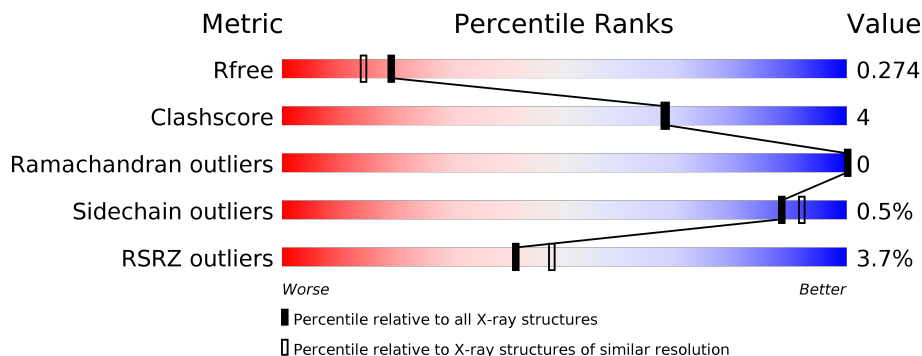
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.11 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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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 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	787	
1	B	787	
1	C	787	
1	D	787	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 12009 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 Desferrioxamine siderophore biosynthesis protein dfoC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	146	Total 1179	C 754	N 206	O 210	S 9	0	0	0
1	B	145	Total 1173	C 752	N 204	O 207	S 10	0	0	0
1	C	575	Total 4616	C 2950	N 803	O 840	S 23	0	0	0
1	D	575	Total 4616	C 2950	N 802	O 841	S 23	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP D4I247
A	-1	ALA	-	expression tag	UNP D4I247
A	0	MET	-	expression tag	UNP D4I247
A	1	SER	-	expression tag	UNP D4I247
B	-2	GLY	-	expression tag	UNP D4I247
B	-1	ALA	-	expression tag	UNP D4I247
B	0	MET	-	expression tag	UNP D4I247
B	1	SER	-	expression tag	UNP D4I247
C	-2	GLY	-	expression tag	UNP D4I247
C	-1	ALA	-	expression tag	UNP D4I247
C	0	MET	-	expression tag	UNP D4I247
C	1	SER	-	expression tag	UNP D4I247
D	-2	GLY	-	expression tag	UNP D4I247
D	-1	ALA	-	expression tag	UNP D4I247
D	0	MET	-	expression tag	UNP D4I247
D	1	SER	-	expression tag	UNP D4I247

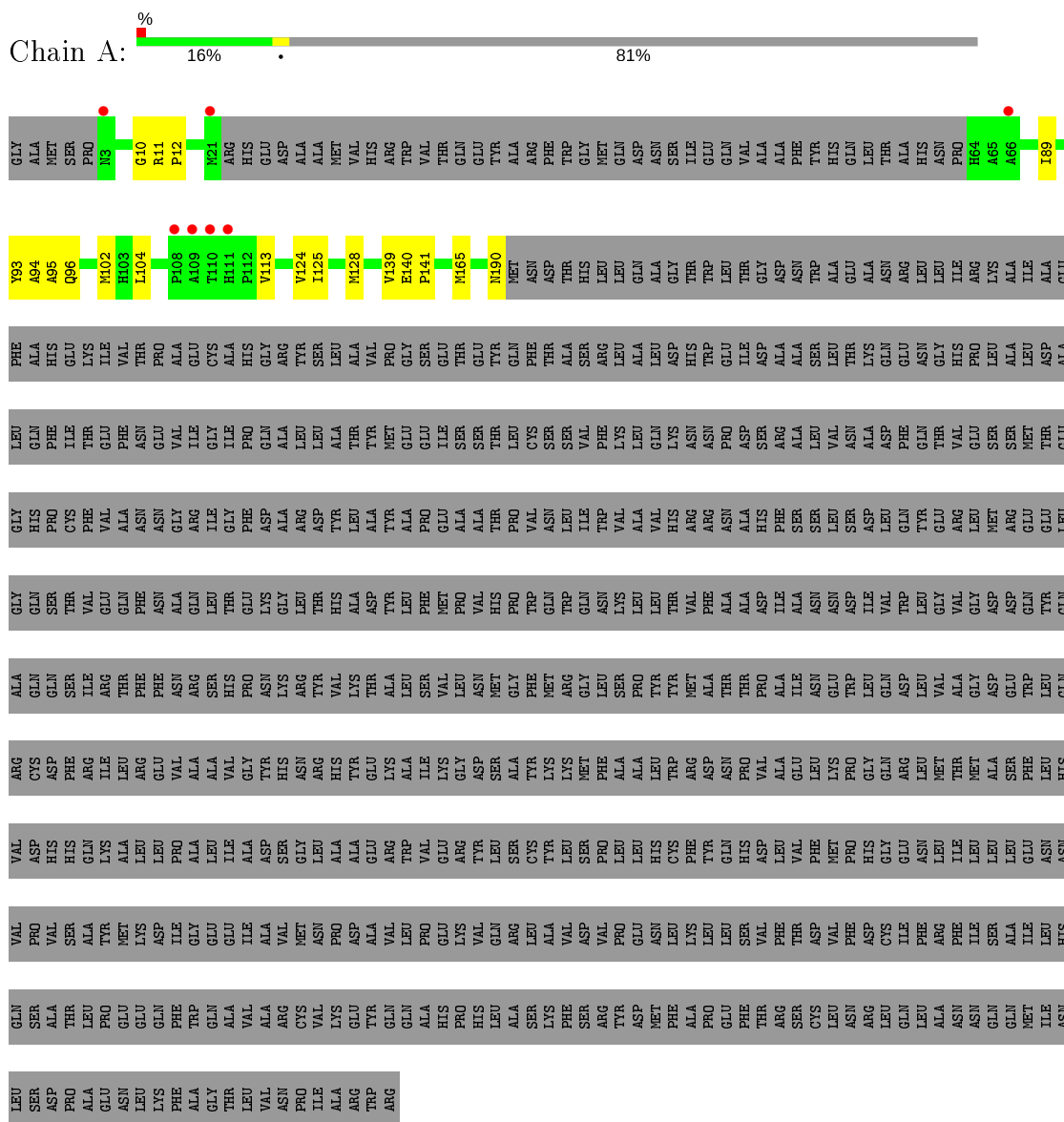
- Molecule 2 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	A	30	Total 30	O 30	0	0
2	B	32	Total 32	O 32	0	0
2	C	195	Total 195	O 195	0	0
2	D	168	Total 168	O 168	0	0

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

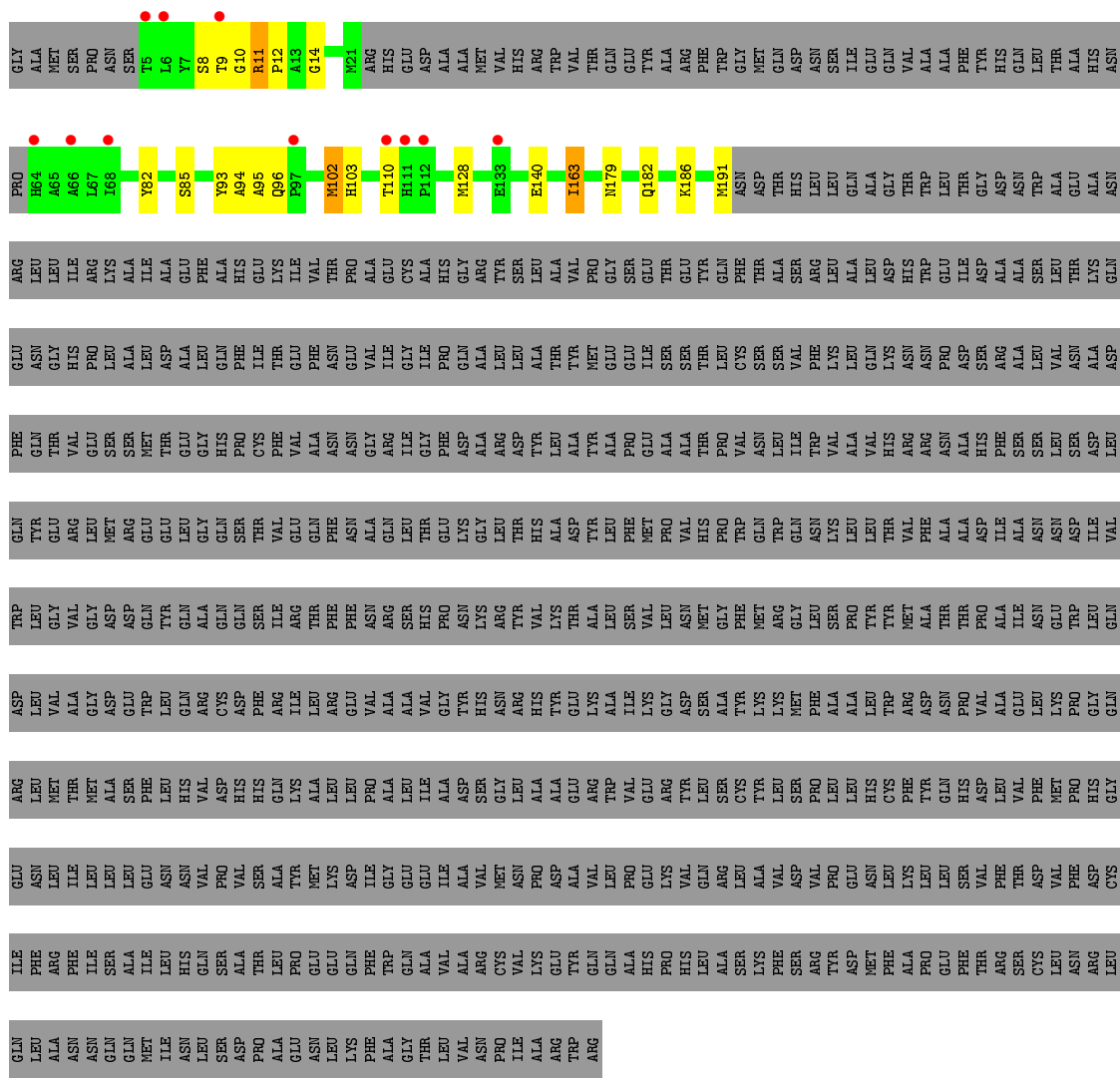
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Desferrioxamine siderophore biosynthesis protein dfoC

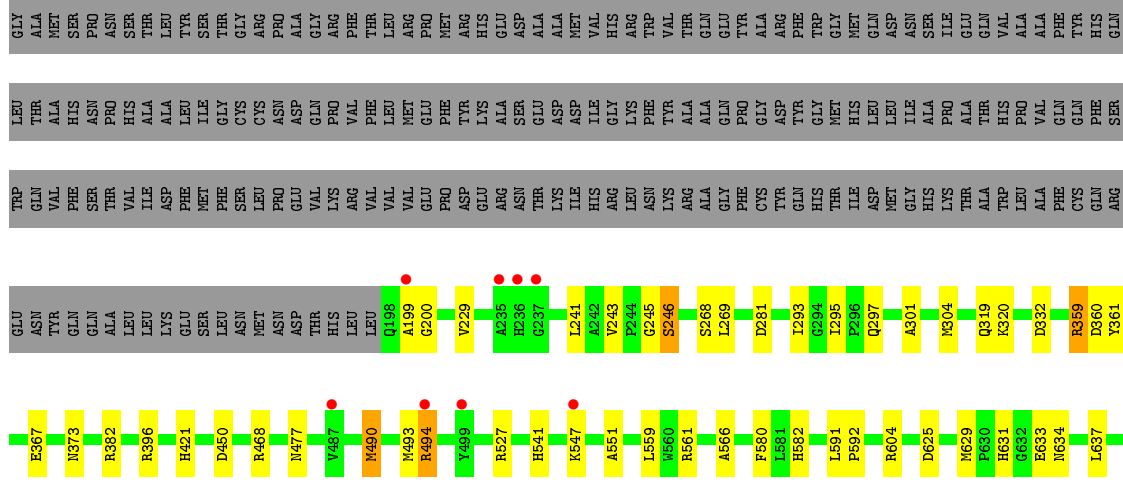


- Molecule 1: Desferrioxamine siderophore biosynthesis protein dfoC





• Molecule 1: Desferrioxamine siderophore biosynthesis protein dfoC





• Molecule 1: Desferrioxamine siderophore biosynthesis protein dfoC



GLY  
ALA  
MET  
SER  
PRO  
ASN  
SER  
THR  
LEU  
TYR  
SER  
ILE  
THR  
GLY  
ARG  
PRO  
ASN  
ALA  
ALA  
GLY  
ARG  
PHE  
THR  
LEU  
ARG  
PRO  
MET  
MET  
GLN  
HIS  
GLU  
ASP  
ASP  
ALA  
ALA  
SER  
MET  
VAL  
HIS  
ARG  
LYS  
TRP  
VAL  
ALA  
GLN  
GLU  
TYR

THR  
ALA  
HIS  
ASN  
PRO  
HIS  
ALA  
LEU  
ILE  
LEU  
GLY  
CYS  
CYS  
ASP  
GLU  
GLN  
PRO  
VAL  
PHE  
LEU  
MET  
GLU  
PRO  
PHE  
TYR  
LYS  
ALA  
VAL  
ILE  
LEU  
GLY  
PHE  
ALA  
GLN  
PHE  
ALA  
ALA

TRP  
GLN  
VAL  
PHE  
SER  
THR  
VAL  
ILE  
ASP  
PHE  
MET  
PHE  
SER  
LEU  
PRO  
GLU  
VAL  
VAL  
LYS  
ARG  
VAL  
VAL  
VAL  
GLU  
GLU  
PRO  
ASP  
GLU  
ARG  
ASN  
THR  
LYS  
LYS  
ILE  
HIS  
ARG  
LYS  
ASN  
TRP  
ARG  
ALA  
GLY  
PHE  
CYS  
TYR  
TYR  
HIS  
THR  
HIS  
ILE  
ASP  
MET  
GLY  
HIS  
LYS  
THR  
ALA  
TRP  
LEU  
ALA  
PHE  
CYS  
GLN  
PHE  
HIS  
SER  
ARG

GLU  
ASN  
TYR  
GLN  
GLN  
ALA  
LEU  
LEU  
LYS  
GLU  
SER  
LEU  
ASN  
MET  
ASN  
ASP  
THR  
HIS  
LEU  
LEU  
GLN  
A199  
T204  
H207  
R213  
I228  
V229  
C234  
A235  
H236  
G237  
H238  
L241  
A242  
V243  
P244  
G245  
S246  
E247  
E263  
H274  
G275  
H276  
D281  
E291  
Y292  
I293  
Q297  
A298

L299  
D324  
T341  
M351  
D360  
Y361  
E367  
R382  
R396  
W434  
V441  
D450  
R468  
G495  
L496  
Y499  
R527  
H541  
K547  
G548  
D549  
M555  
R561  
F580  
L591  
P592  
Y622  
H631  
M634  
L637  
M649  
K650  
D651  
V664

V669  
L683  
F686  
T687  
F690  
I693  
Q703  
V720  
M739  
R746  
S747  
L754  
A755  
N756  
ASN  
GLN  
MET  
ILE  
ASN  
LEU  
SER  
ASP  
ALA  
E768  
N769  
R782  
W783  
R784

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.17Å 156.35Å 93.81Å 90.00° 95.57° 90.00°	Depositor
Resolution (Å)	78.18 – 2.11 78.17 – 2.11	Depositor EDS
% Data completeness (in resolution range)	99.8 (78.18-2.11) 99.8 (78.17-2.11)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.221 , 0.264 0.235 , 0.274	Depositor DCC
$R_{free}$ test set	6119 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	44.9	Xtrriage
Anisotropy	0.106	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 42.2	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	0.95	EDS
Total number of atoms	12009	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.11% 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](#)

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	0.94	1/1211 (0.1%)	0.93	4/1638 (0.2%)
1	B	0.90	1/1205 (0.1%)	1.02	8/1629 (0.5%)
1	C	0.87	4/4736 (0.1%)	0.93	19/6440 (0.3%)
1	D	0.83	2/4736 (0.0%)	0.90	14/6440 (0.2%)
All	All	0.86	8/11888 (0.1%)	0.92	45/16147 (0.3%)

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	B	0	1
1	C	0	1
1	D	0	1
All	All	0	3

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	649	MET	C-N	11.38	1.60	1.34
1	C	268	SER	C-N	-11.35	1.07	1.34
1	B	163	ILE	C-N	-11.08	1.08	1.34
1	C	245	GLY	C-N	10.91	1.59	1.34
1	C	246	SER	C-N	-9.44	1.12	1.34

The worst 5 of 45 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	490	MET	CB-CA-C	-15.29	79.83	110.40
1	B	163	ILE	O-C-N	-13.47	101.15	122.70
1	B	10	GLY	C-N-CA	10.81	148.73	121.70
1	B	10	GLY	O-C-N	-9.72	107.15	122.70

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	10	GLY	C-N-CA	9.54	145.55	121.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	163	ILE	Mainchain
1	C	200	GLY	Peptide
1	D	650	LYS	Mainchain

## 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	1179	0	1142	18	0
1	B	1173	0	1139	14	0
1	C	4616	0	4499	36	0
1	D	4616	0	4499	30	0
2	A	30	0	0	2	0
2	B	32	0	0	0	0
2	C	195	0	0	4	0
2	D	168	0	0	2	0
All	All	12009	0	11279	87	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 4.

The worst 5 of 87 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:139:VAL:HG12	1:A:141:PRO:HD3	1.46	0.93
1:D:361:TYR:OH	1:D:367:GLU:OE2	1.91	0.88
1:A:94:ALA:O	1:A:96:GLN:NE2	2.17	0.77
1:A:140:GLU:OE1	1:A:165:MET:CE	2.35	0.75
1:A:89:ILE:HD12	1:A:140:GLU:HB3	1.73	0.69

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	142/787 (18%)	136 (96%)	6 (4%)	0	100	100
1	B	141/787 (18%)	136 (96%)	5 (4%)	0	100	100
1	C	571/787 (73%)	551 (96%)	20 (4%)	0	100	100
1	D	571/787 (73%)	552 (97%)	19 (3%)	0	100	100
All	All	1425/3148 (45%)	1375 (96%)	50 (4%)	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	126/670 (19%)	126 (100%)	0	100	100
1	B	125/670 (19%)	123 (98%)	2 (2%)	62	68
1	C	488/670 (73%)	487 (100%)	1 (0%)	93	96
1	D	488/670 (73%)	485 (99%)	3 (1%)	86	90
All	All	1227/2680 (46%)	1221 (100%)	6 (0%)	88	92

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	490	MET
1	D	351	ASN
1	D	204	THR
1	B	110	THR
1	D	341	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	522	GLN
1	C	624	HIS
1	D	631	HIS
1	C	541	HIS
1	C	582	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](#)

There are no ligands in this entry.

### 5.7 Other polymers [i](#)

There are no such residues in this entry.

### 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	C	3
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	269:LEU	C	270:THR	N	1.18
1	C	246:SER	C	247:GLU	N	1.12
1	B	163:ILE	C	164:ASP	N	1.08
1	C	268:SER	C	269:LEU	N	1.07

## 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	146/787 (18%)	0.30	7 (4%) 30 35	36, 51, 70, 95	0
1	B	145/787 (18%)	0.40	11 (7%) 13 17	37, 55, 74, 84	0
1	C	575/787 (73%)	0.20	9 (1%) 72 76	34, 49, 73, 104	0
1	D	575/787 (73%)	0.25	27 (4%) 31 36	33, 50, 75, 115	0
All	All	1441/3148 (45%)	0.25	54 (3%) 41 48	33, 50, 74, 115	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	499	TYR	7.8
1	C	494	ARG	5.6
1	D	236	HIS	5.3
1	D	495	GLY	5.1
1	A	110	THR	5.0

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.