



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2016 – 06:10 AM EDT

PDB ID : 5KS8  
Title : Crystal structure of two-subunit pyruvate carboxylase from *Methylobacillus flagellatus*  
Authors : Choi, P.H.; Tong, L.  
Deposited on : 2016-07-07  
Resolution : 3.01 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
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) : rb-20027939

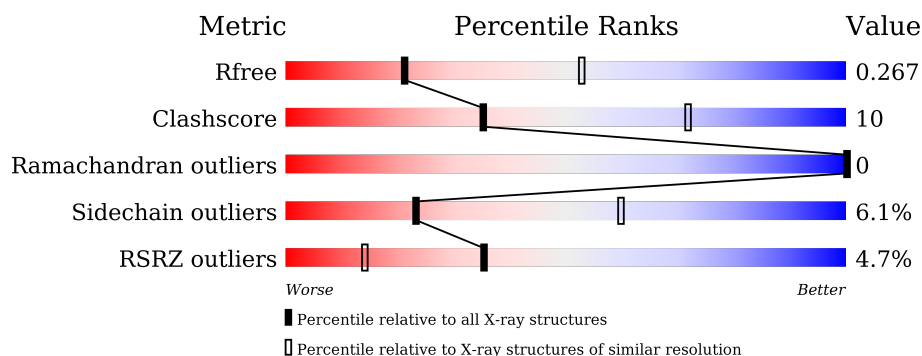
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*





The reported resolution of this entry is 3.01 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	1773 (3.04-3.00)
Clashscore	102246	2117 (3.04-3.00)
Ramachandran outliers	100387	2050 (3.04-3.00)
Sidechain outliers	100360	2053 (3.04-3.00)
RSRZ outliers	91569	1788 (3.04-3.00)

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	405	 79% 19% •
1	B	405	 79% 20% •
2	C	617	 72% 24% • •
2	D	617	 72% 20% • 6%
2	E	617	 10% 57% 11% • 30%
2	F	617	 11% 63% 14% • 22%

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
4	PYR	C	2001	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 21684 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 Pyruvate carboxylase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	404	Total	C	N	O	S	0	0	0
			3139	1985	557	582	15			
1	B	404	Total	C	N	O	S	0	0	0
			3149	1992	558	583	16			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	131	GLY	-	linker	UNP Q1H158
A	132	SER	-	linker	UNP Q1H158
A	200	SER	-	linker	UNP Q1H158
A	201	GLY	-	linker	UNP Q1H158
B	131	GLY	-	linker	UNP Q1H158
B	132	SER	-	linker	UNP Q1H158
B	200	SER	-	linker	UNP Q1H158
B	201	GLY	-	linker	UNP Q1H158

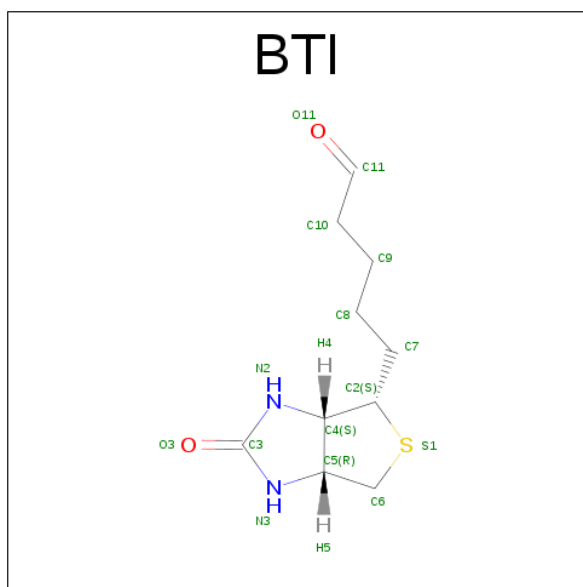
- Molecule 2 is a protein called Pyruvate carboxylase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	603	Total	C	N	O	S	0	0	0
			4554	2862	787	885	20			
2	D	580	Total	C	N	O	S	0	0	0
			4410	2777	760	853	20			
2	E	430	Total	C	N	O	S	0	0	0
			2871	1768	514	581	8			
2	F	482	Total	C	N	O	S	0	0	0
			3531	2211	606	698	16			

There are 12 discrepancies between the modelled and reference sequences:

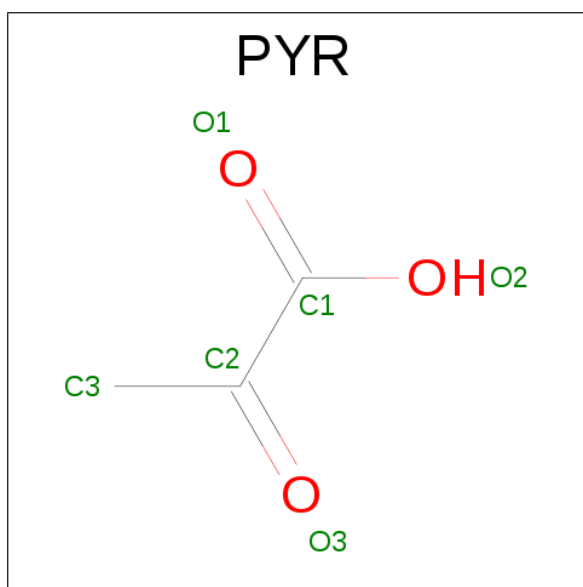
Chain	Residue	Modelled	Actual	Comment	Reference
C	419	ALA	LYS	conflict	UNP Q1H157
C	421	ALA	GLU	conflict	UNP Q1H157
C	422	ALA	GLU	conflict	UNP Q1H157
D	419	ALA	LYS	conflict	UNP Q1H157
D	421	ALA	GLU	conflict	UNP Q1H157
D	422	ALA	GLU	conflict	UNP Q1H157
E	419	ALA	LYS	conflict	UNP Q1H157
E	421	ALA	GLU	conflict	UNP Q1H157
E	422	ALA	GLU	conflict	UNP Q1H157
F	419	ALA	LYS	conflict	UNP Q1H157
F	421	ALA	GLU	conflict	UNP Q1H157
F	422	ALA	GLU	conflict	UNP Q1H157

- Molecule 3 is 5-(HEXAHYDRO-2-OXO-1H-THIENO[3,4-D]IMIDAZOL-6-YL)PENTANAL (three-letter code: BTI) (formula:  $C_{10}H_{16}N_2O_2S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	C	1	Total	C	N	O	S	0	0
			15	10	2	2	1		

- Molecule 4 is PYRUVIC ACID (three-letter code: PYR) (formula:  $C_3H_4O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			6	3	3		
4	D	1	Total	C	O	0	0
			6	3	3		

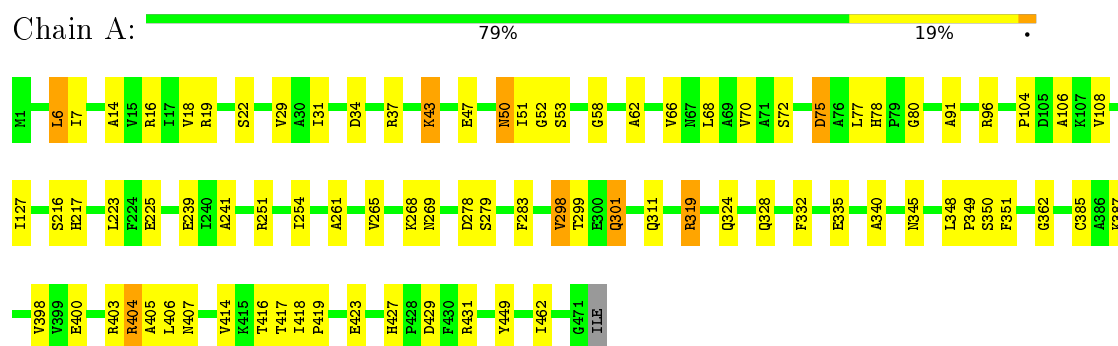
- Molecule 5 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	1	Total	Mn	0	0
			1	1		
5	C	1	Total	Mn	0	0
			1	1		
5	F	1	Total	Mn	0	0
			1	1		

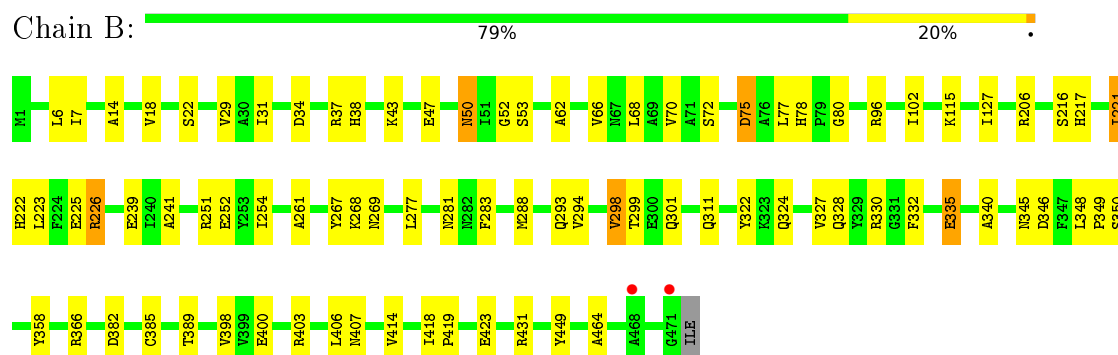
### 3 Residue-property plots

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

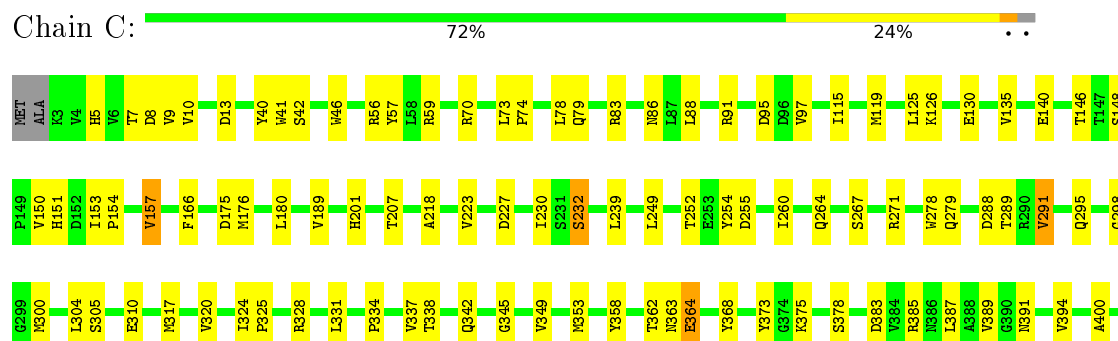
- Molecule 1: Pyruvate carboxylase subunit alpha

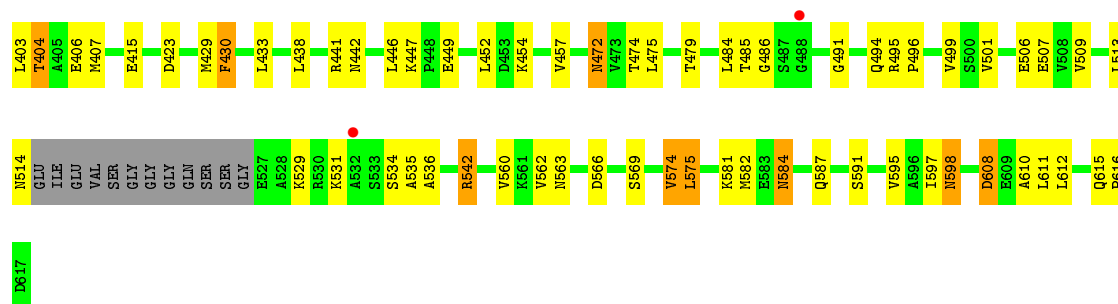


- Molecule 1: Pyruvate carboxylase subunit alpha



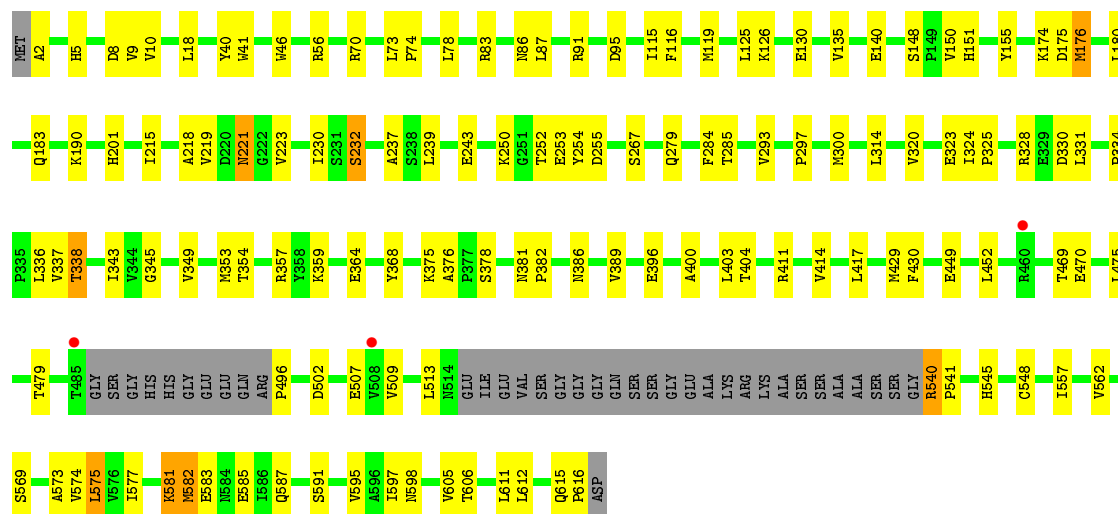
- Molecule 2: Pyruvate carboxylase subunit beta





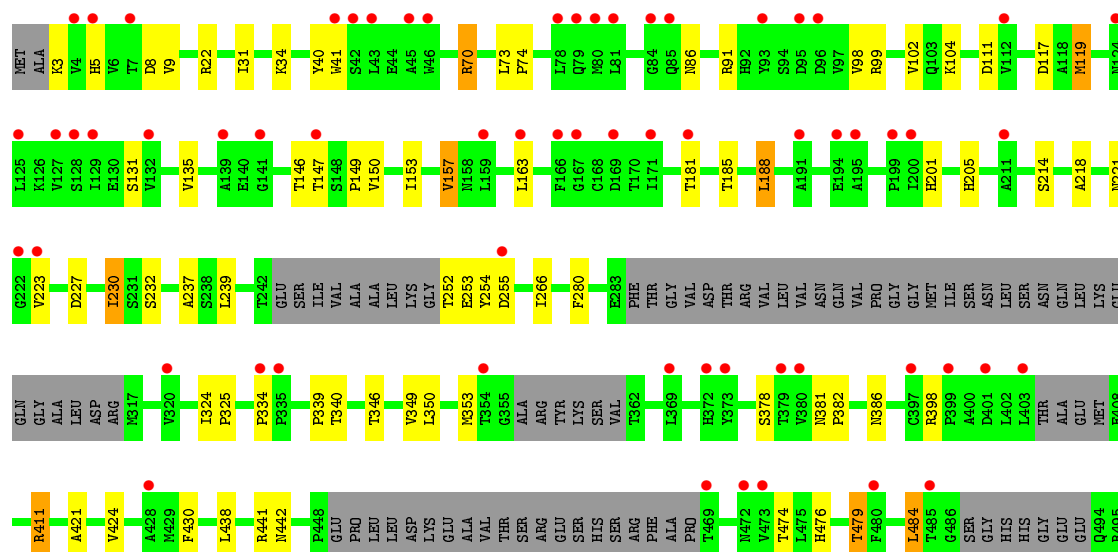
• Molecule 2: Pyruvate carboxylase subunit beta

Chain D: 72% 20% 6%

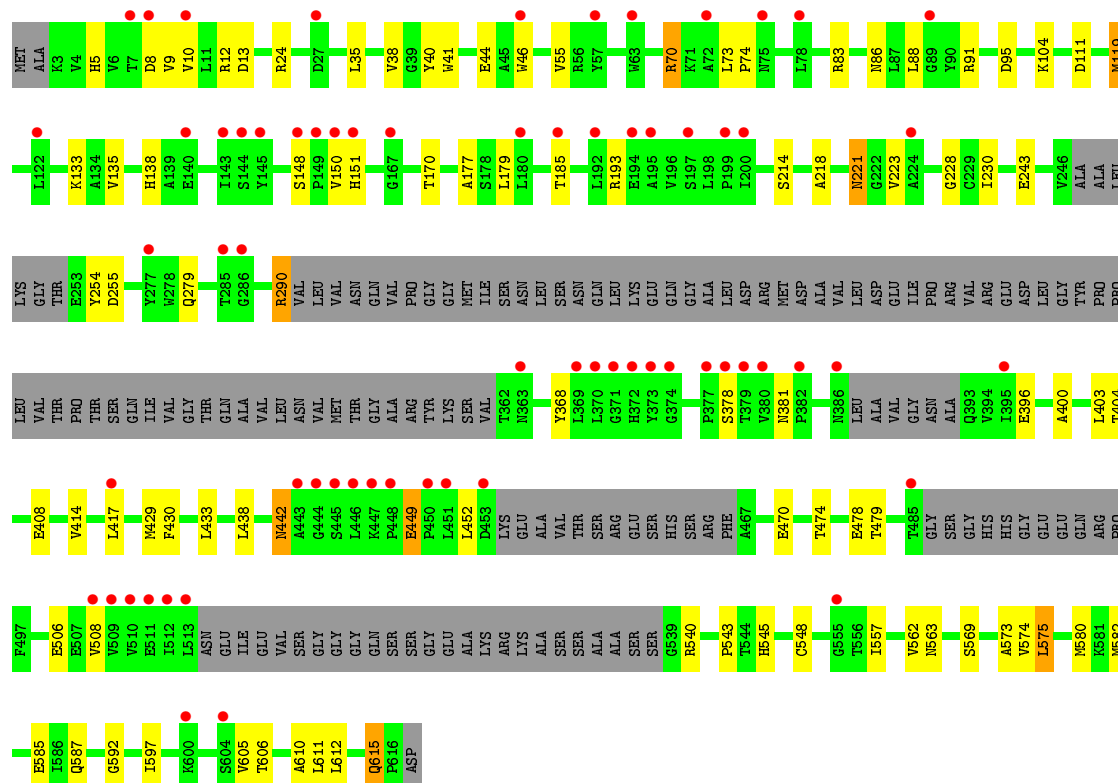


• Molecule 2: Pyruvate carboxylase subunit beta

Chain E: 10% 57% 11% 30%







## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	285.76 Å   285.76 Å   274.87 Å 90.00°   90.00°   120.00°	Depositor
Resolution (Å)	47.44 – 3.01 47.39 – 3.01	Depositor EDS
% Data completeness (in resolution range)	95.5 (47.44-3.01) 95.6 (47.39-3.01)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.99 (at 3.01 Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.225   ,   0.272 0.224   ,   0.267	Depositor DCC
$R_{free}$ test set	4070 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	87.8	Xtriage
Anisotropy	0.004	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 79.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	21684	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PYR, MN, BTI

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	0.53	0/3203	0.78	3/4340 (0.1%)
1	B	0.58	0/3214	0.79	0/4353
2	C	0.54	0/4629	0.76	3/6287 (0.0%)
2	D	0.56	0/4484	0.75	3/6093 (0.0%)
2	E	0.43	0/2910	0.72	4/3989 (0.1%)
2	F	0.42	0/3586	0.67	1/4884 (0.0%)
All	All	0.52	0/22026	0.75	14/29946 (0.0%)

There are no bond length outliers.

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	398	ARG	NE-CZ-NH1	6.95	123.77	120.30
2	D	612	LEU	CA-CB-CG	6.94	131.26	115.30
2	E	398	ARG	NE-CZ-NH2	-6.68	116.96	120.30
2	C	291	VAL	CB-CA-C	-6.12	99.78	111.40
2	D	502	ASP	CB-CG-OD2	5.84	123.56	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	3139	0	3117	58	0
1	B	3149	0	3128	59	0
2	C	4554	0	4540	117	0
2	D	4410	0	4410	101	0
2	E	2871	0	2421	55	1
2	F	3531	0	3343	60	1
3	C	15	0	15	1	0
4	C	6	0	3	0	0
4	D	6	0	3	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
5	F	1	0	0	0	0
All	All	21684	0	20980	418	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 10.

The worst 5 of 418 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)
2:F:548:CYS:SG	2:F:610:ALA:HB1	1.86	1.15
2:F:548:CYS:SG	2:F:610:ALA:CB	2.49	1.00
2:E:22:ARG:HG3	2:E:280:PHE:CE1	2.08	0.88
2:D:2:ALA:O	2:D:254:TYR:HA	1.72	0.88
2:F:474:THR:HG22	2:F:479:THR:HG23	1.54	0.87

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)
2:E:149:PRO:O	2:F:193:ARG:NH2[4_555]	2.07	0.13

## 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	402/405 (99%)	375 (93%)	27 (7%)	0	100	100
1	B	402/405 (99%)	372 (92%)	30 (8%)	0	100	100
2	C	599/617 (97%)	572 (96%)	27 (4%)	0	100	100
2	D	574/617 (93%)	553 (96%)	21 (4%)	0	100	100
2	E	416/617 (67%)	397 (95%)	19 (5%)	0	100	100
2	F	468/617 (76%)	445 (95%)	23 (5%)	0	100	100
All	All	2861/3278 (87%)	2714 (95%)	147 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 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	327/332 (98%)	314 (96%)	13 (4%)	38	76
1	B	329/332 (99%)	312 (95%)	17 (5%)	29	67
2	C	491/513 (96%)	455 (93%)	36 (7%)	17	51
2	D	479/513 (93%)	455 (95%)	24 (5%)	30	69
2	E	242/513 (47%)	224 (93%)	18 (7%)	17	50
2	F	360/513 (70%)	333 (92%)	27 (8%)	17	50
All	All	2228/2716 (82%)	2093 (94%)	135 (6%)	23	60

5 of 135 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	575	LEU
2	D	338	THR
2	F	470	GLU
2	C	584	ASN
2	D	95	ASP

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	324	GLN
2	C	92	HIS
2	F	138	HIS
1	B	328	GLN
1	B	407	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](#)

Of 6 ligands modelled in this entry, 3 are monoatomic - leaving 3 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	$\# Z  > 2$
3	BTI	C	2000	2	14,16,16	1.93	1 (7%)	14,21,21	2.05	6 (42%)
4	PYR	C	2001	-	2,5,5	0.53	0	2,6,6	1.18	0
4	PYR	D	2001	-	2,5,5	1.20	0	2,6,6	0.78	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
3	BTI	C	2000	2	-	0/5/27/27	0/2/2/2
4	PYR	C	2001	-	-	0/0/4/4	0/0/0/0
4	PYR	D	2001	-	-	0/0/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	2000	BTI	C2-S1	-6.39	1.72	1.82

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	2000	BTI	O3-C3-N3	-3.20	122.16	125.89
3	C	2000	BTI	C9-C10-C11	-2.26	107.07	113.35
3	C	2000	BTI	C5-N3-C3	-2.21	110.81	112.49
3	C	2000	BTI	C4-C5-N3	2.02	104.65	102.37
3	C	2000	BTI	N2-C3-N3	2.70	110.71	108.85

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	2000	BTI	1	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	404/405 (99%)	-0.16	0 100 100	55, 94, 126, 158	0
1	B	404/405 (99%)	-0.33	2 (0%) 91 76	54, 78, 132, 198	0
2	C	603/617 (97%)	-0.33	2 (0%) 94 84	52, 94, 137, 214	0
2	D	580/617 (94%)	-0.38	3 (0%) 91 76	47, 74, 138, 194	0
2	E	430/617 (69%)	0.63	62 (14%) 3 1	95, 169, 222, 253	0
2	F	482/617 (78%)	0.58	67 (13%) 4 1	97, 156, 196, 229	0
All	All	2903/3278 (88%)	-0.02	136 (4%) 35 14	47, 101, 191, 253	0

The worst 5 of 136 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	510	VAL	5.8
2	F	8	ASP	5.7
2	F	372	HIS	5.7
2	F	374	GLY	5.6
2	F	195	ALA	5.4

### 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(Å <sup>2</sup> )	Q<0.9
4	PYR	C	2001	6/6	0.87	0.24	2.16	84,98,101,101	0
4	PYR	D	2001	6/6	0.89	0.26	1.61	64,71,72,73	0
5	MN	C	2002	1/1	0.98	0.21	0.51	92,92,92,92	0
3	BTI	C	2000	15/15	0.98	0.21	0.35	54,60,81,91	0
5	MN	F	701	1/1	0.87	0.08	-1.71	164,164,164,164	0
5	MN	D	2002	1/1	0.99	0.28	-	64,64,64,64	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.