wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2023 – 12:58 AM JST

PDB ID	:	5YP3
Title	:	Crystal structure of dipeptidyl peptidase IV (DPP IV) with Ile-Pro from Pseudoxanthomonas mexicana
Authors	:	Roppongi, S.; Suzuki, Y.; Tateoka, C.; Fuimoto, M.; Morisawa, S.; Iizuka, I.; Nakamura, A.; Honma, N.; Shida, Y.; Ogasawara, W.; Tanaka, N.; Sakamoto, Y.; Nonaka, T.
Deposited on	:	2017-11-01
Resolution	:	2.44 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

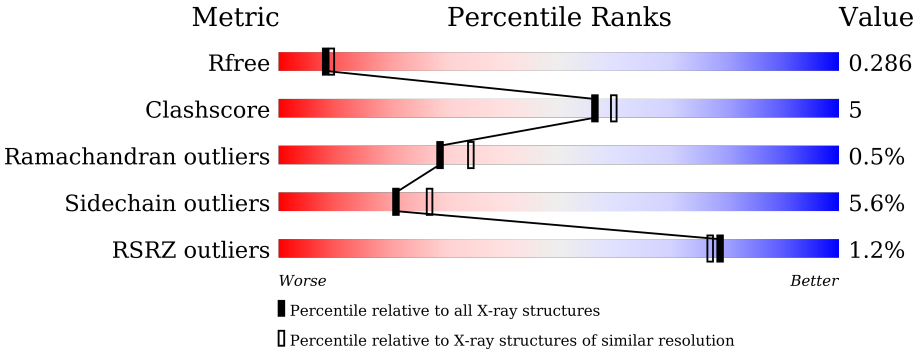
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.44 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	1564 (2.46-2.42)
Clashscore	141614	1631 (2.46-2.42)
Ramachandran outliers	138981	1617 (2.46-2.42)
Sidechain outliers	138945	1617 (2.46-2.42)
RSRZ outliers	127900	1547 (2.46-2.42)

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 >=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 <=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	745	.% 78% 17% • •
1	B	745	.% 82% 14% • •
1	C	745	2% 86% 10% • •
1	D	745	81% 14% • •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 23158 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.

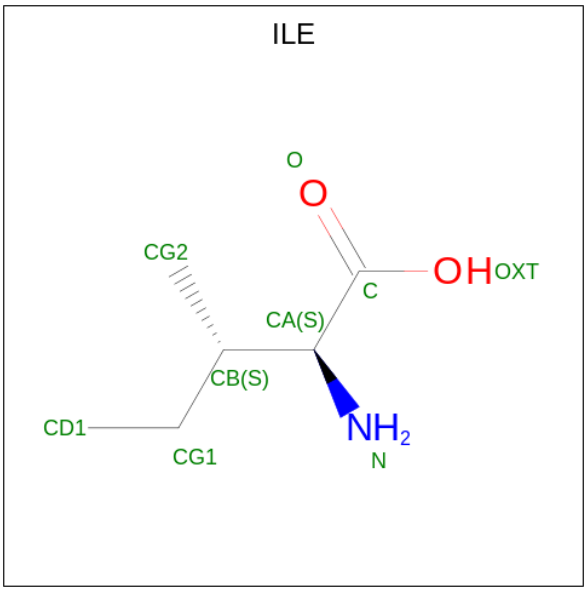
- Molecule 1 is a protein called Dipeptidyl aminopeptidase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	724	5660	3580	994	1076	10	0	0	0
1	B	724	5660	3580	994	1076	10	0	0	0
1	C	724	5660	3580	994	1076	10	0	0	0
1	D	724	5660	3580	994	1076	10	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

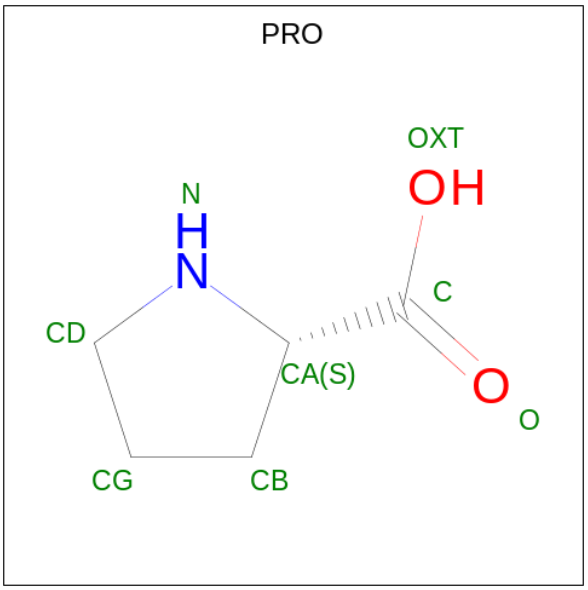
Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ILE	MET	see sequence details	UNP Q6F3I7
B	12	ILE	MET	see sequence details	UNP Q6F3I7
C	12	ILE	MET	see sequence details	UNP Q6F3I7
D	12	ILE	MET	see sequence details	UNP Q6F3I7

- Molecule 2 is ISOLEUCINE (three-letter code: ILE) (formula: $C_6H_{13}NO_2$).



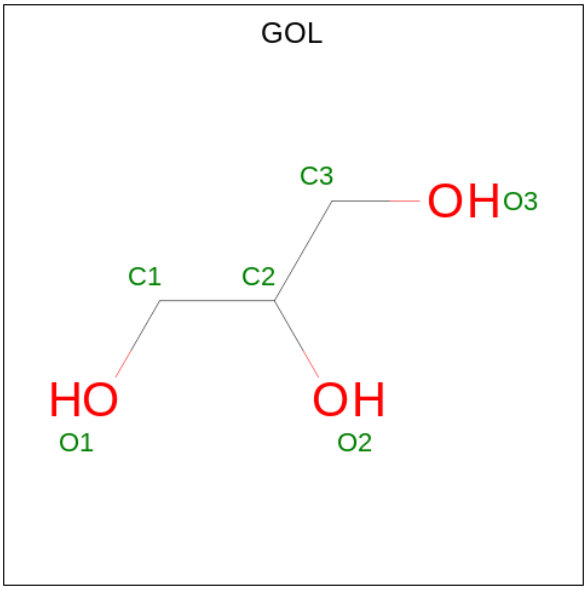
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	8	6	1	1	0	0
2	B	1	8	6	1	1	0	0
2	C	1	8	6	1	1	0	0
2	D	1	8	6	1	1	0	0

- Molecule 3 is PROLINE (three-letter code: PRO) (formula: $C_5H_9NO_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	7	5	1	1	0	0
3	B	1	7	5	1	1	0	0
3	C	1	7	5	1	1	0	0
3	D	1	7	5	1	1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 5 is water.

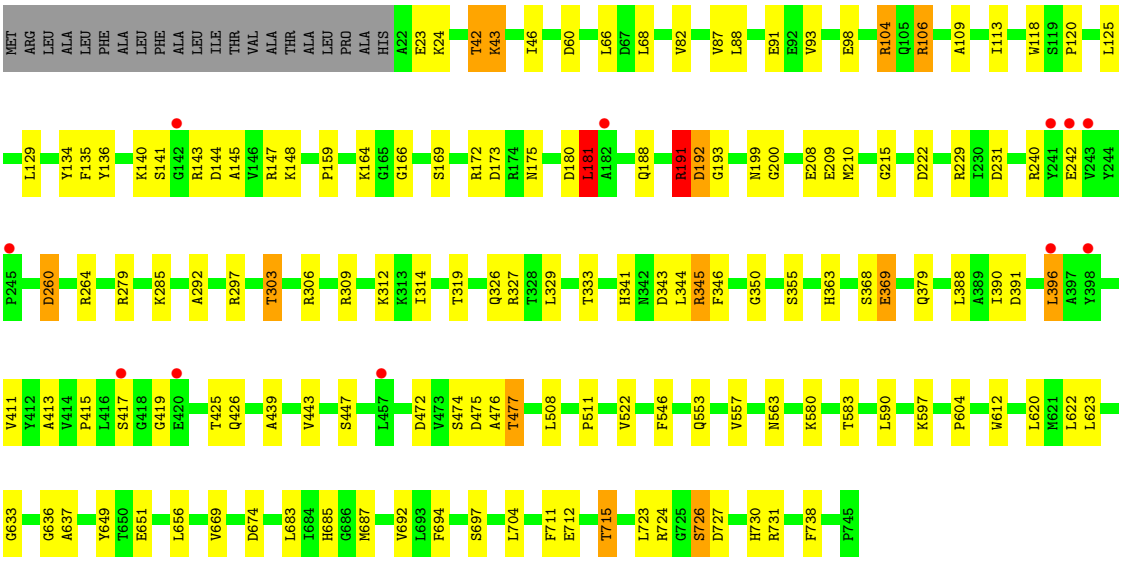
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	143	143	143	0	0
5	B	93	93	93	0	0
5	C	99	99	99	0	0
5	D	117	117	117	0	0

3 Residue-property plots (i)

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.

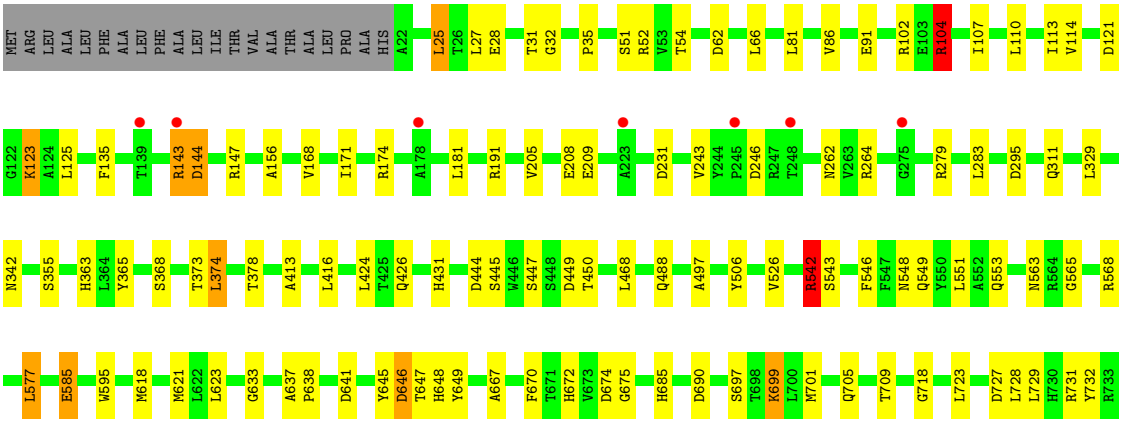
• Molecule 1: Dipeptidyl aminopeptidase 4

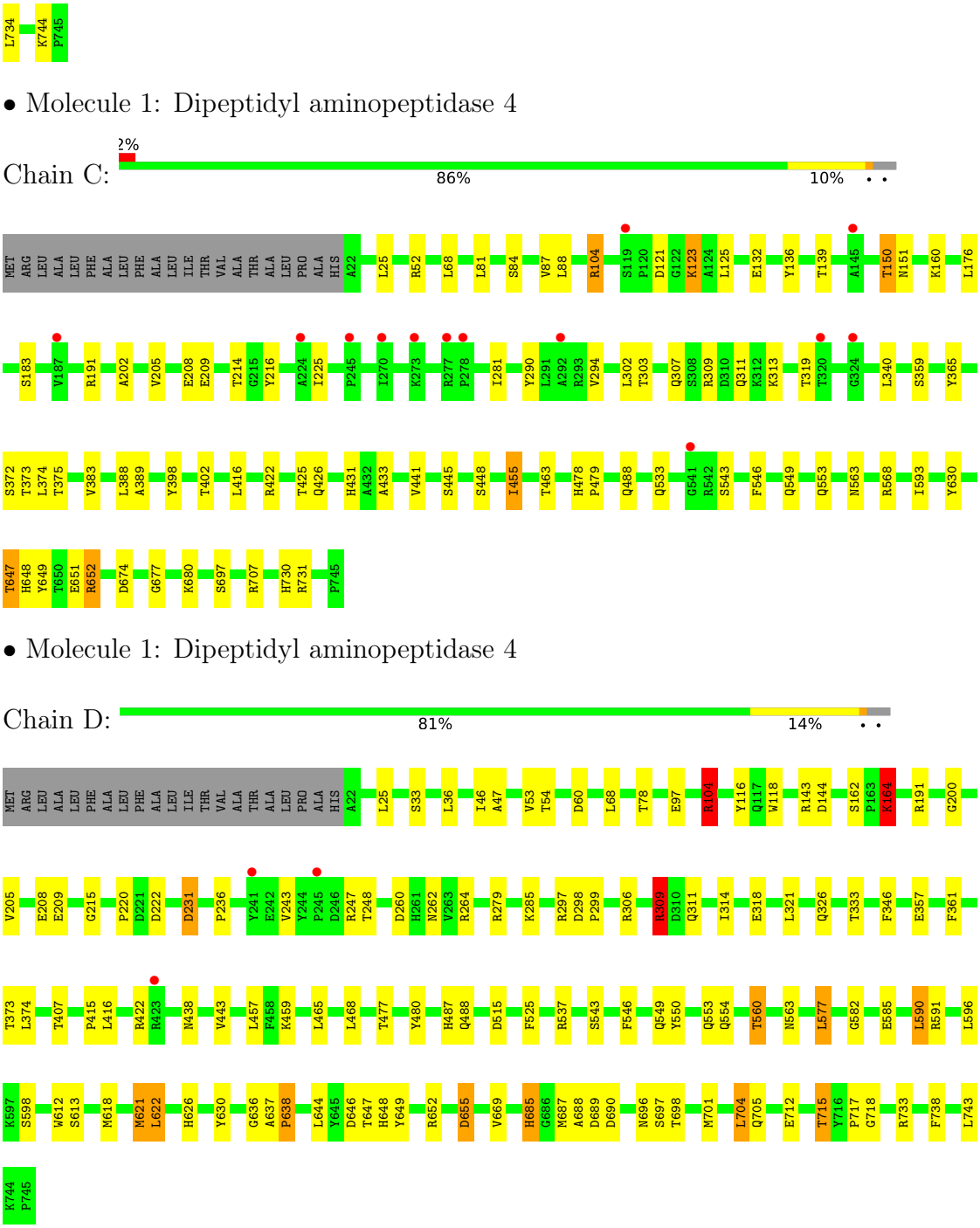
Chain A:



• Molecule 1: Dipeptidyl aminopeptidase 4

Chain B:




L734 K744 P745
• Molecule 1: Dipeptidyl aminopeptidase 4
Chain C:
2%
86%
10%
MET ARG LEU ALA LEU PHE ALA LEU PHE ALA LEU ILE THR VAL ALA THR ALA LEU PRO ALA HIS A22 L25 R52 L68 L81 S84 V87 L88 R104 S119 P120 D121 G122 K123 A124 L125 E132 Y136 T139 A145 T150 N151 K160 L176
S183 V187 R191 A202 V205 E208 E209 T214 G215 Y216 A224 I225 P245 I270 K273 R277 P278 I281 Y290 L291 A292 R293 V294 L302 T303 Q307 S308 R309 D310 Q311 K312 K313 T319 T320 G324 L340 S359 Y365
S372 T373 L374 T375 V383 L388 A389 Y398 T402 L416 R422 T425 Q426 H431 A432 A433 V441 S445 S448 I455 T463 H478 P479 Q488 Q533 G541 R542 S543 F546 Q549 Q553 N563 R568 I593 Y630
T647 H648 Y649 T650 E651 R652 D674 G677 K680 S697 R707 H730 R731 P745
• Molecule 1: Dipeptidyl aminopeptidase 4
Chain D:
81%
14%
MET ARG LEU ALA LEU PHE ALA LEU PHE ALA LEU ILE THR VAL ALA THR ALA LEU PRO ALA HIS A22 L25 S33 L36 I46 A47 V53 T54 D60 L68 T78 E97 R104 Y116 Q117 W118 R143 D144 S162 P163 K164 R191 G200
V205 E208 E209 G215 P220 D221 D222 D231 P236 Y241 E242 V243 Y244 P245 D246 R247 T248 D260 H261 N262 V263 R264 R279 K285 R297 D298 P299 R306 R309 D310 Q311 I314 E318 L321 Q326 T333 F346 E357 F361
T373 L374 T407 P415 L416 R422 R423 N438 V443 L457 F458 K459 L465 L468 T477 Y480 H487 Q488 D515 F525 R537 S543 F546 Q549 Y550 Q553 Q554 T560 N563 L577 Q582 E585 L590 R591 L596
K597 S598 W612 S613 M618 M621 L622 H626 Y630 G636 A637 P638 L644 Y645 D646 T647 H648 Y649 R652 D655 V669 H685 G686 M687 A688 D689 D690 N696 S697 T698 M701 L704 Q705 E712 T715 Y716 P717 G718 R733 F738 L743
K744 P745

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	119.88Å 120.12Å 262.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.44 60.06 – 2.44	Depositor EDS
% Data completeness (in resolution range)	97.4 (40.00-2.44) 97.4 (60.06-2.44)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.22 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
R, R_{free}	0.238 , 0.286 0.243 , 0.286	Depositor DCC
R_{free} test set	6835 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	39.3	Xtriage
Anisotropy	0.342	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 41.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	0.004 for k,h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	23158	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 39.35 % 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 3.2151e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL

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	0.77	0/5799	0.95	10/7887 (0.1%)
1	B	0.70	0/5799	0.90	8/7887 (0.1%)
1	C	0.68	0/5799	0.87	7/7887 (0.1%)
1	D	0.75	0/5799	0.96	13/7887 (0.2%)
All	All	0.72	0/23196	0.92	38/31548 (0.1%)

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	A	0	1
1	B	0	1
1	D	0	2
All	All	0	4

There are no bond length outliers.

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	104	ARG	NE-CZ-NH2	-14.85	112.87	120.30
1	D	104	ARG	NE-CZ-NH1	10.91	125.76	120.30
1	B	104	ARG	NE-CZ-NH2	9.40	125.00	120.30
1	B	104	ARG	NE-CZ-NH1	-8.78	115.91	120.30
1	D	309	ARG	NE-CZ-NH1	7.36	123.98	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	192	ASP	Peptide
1	B	638	PRO	Peptide
1	D	164	LYS	Peptide
1	D	638	PRO	Peptide

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	5660	0	5521	70	0
1	B	5660	0	5521	63	0
1	C	5660	0	5521	35	0
1	D	5660	0	5521	63	0
2	A	8	0	10	1	0
2	B	8	0	10	1	0
2	C	8	0	10	1	0
2	D	8	0	10	0	0
3	A	7	0	7	0	0
3	B	7	0	7	0	0
3	C	7	0	7	0	0
3	D	7	0	7	0	0
4	D	6	0	8	0	0
5	A	143	0	0	3	0
5	B	93	0	0	3	0
5	C	99	0	0	0	0
5	D	117	0	0	1	0
All	All	23158	0	22160	226	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 5.

The worst 5 of 226 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:262:ASN:ND2	1:D:647:THR:HG21	1.59	1.16
1:D:262:ASN:HD21	1:D:647:THR:HG21	0.96	1.11
1:B:262:ASN:HD21	1:B:647:THR:HG21	1.00	1.09

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:262:ASN:ND2	1:B:647:THR:HG21	1.75	1.01
1:D:262:ASN:HD21	1:D:647:THR:CG2	1.79	0.96

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	722/745 (97%)	669 (93%)	49 (7%)	4 (1%)	25	29
1	B	722/745 (97%)	661 (92%)	58 (8%)	3 (0%)	34	41
1	C	722/745 (97%)	664 (92%)	55 (8%)	3 (0%)	34	41
1	D	722/745 (97%)	666 (92%)	53 (7%)	3 (0%)	34	41
All	All	2888/2980 (97%)	2660 (92%)	215 (7%)	13 (0%)	29	34

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	542	ARG
1	A	222	ASP
1	C	543	SER
1	A	181	LEU
1	A	369	GLU

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	587/602 (98%)	546 (93%)	41 (7%)	15	18
1	B	587/602 (98%)	555 (94%)	32 (6%)	21	28
1	C	587/602 (98%)	562 (96%)	25 (4%)	29	38
1	D	587/602 (98%)	552 (94%)	35 (6%)	19	25
All	All	2348/2408 (98%)	2215 (94%)	133 (6%)	21	26

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	374	LEU
1	D	546	PHE
1	D	704	LEU
1	B	143	ARG
1	B	135	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	431	HIS
1	D	311	GLN
1	C	487	HIS
1	C	685	HIS
1	D	487	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

9 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	$\#\|Z\| > 2$
4	GOL	D	803	-	5,5,5	1.01	0	5,5,5	1.27	0
3	PRO	D	802	1,2	5,7,8	0.54	0	7,8,10	1.34	1 (14%)
3	PRO	C	802	1,2	5,7,8	0.50	0	7,8,10	2.64	3 (42%)
2	ILE	B	801	3	6,7,8	0.66	0	5,8,10	1.75	2 (40%)
3	PRO	A	802	1,2	5,7,8	0.61	0	7,8,10	1.95	2 (28%)
2	ILE	A	801	3	6,7,8	0.66	0	5,8,10	2.10	1 (20%)
2	ILE	D	801	3	6,7,8	0.85	0	5,8,10	2.23	2 (40%)
2	ILE	C	801	3	6,7,8	0.73	0	5,8,10	2.57	2 (40%)
3	PRO	B	802	1,2	5,7,8	0.78	0	7,8,10	1.29	1 (14%)

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
4	GOL	D	803	-	-	4/4/4/4	-
3	PRO	D	802	1,2	-	0/0/9/11	0/1/1/1
3	PRO	C	802	1,2	-	0/0/9/11	0/1/1/1
2	ILE	B	801	3	-	2/7/8/10	-
3	PRO	A	802	1,2	-	0/0/9/11	0/1/1/1
2	ILE	A	801	3	-	3/7/8/10	-
2	ILE	D	801	3	-	2/7/8/10	-
2	ILE	C	801	3	-	1/7/8/10	-
3	PRO	B	802	1,2	-	0/0/9/11	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	801	ILE	O-C-CA	-4.54	112.88	124.78
3	C	802	PRO	CB-CA-C	-4.50	106.51	112.70
2	A	801	ILE	O-C-CA	-4.17	113.84	124.78
3	C	802	PRO	O-C-CA	-4.08	114.08	124.78
3	A	802	PRO	CB-CA-C	-3.85	107.40	112.70

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	801	ILE	O-C-CA-CB
2	A	801	ILE	CG2-CB-CG1-CD1
2	A	801	ILE	CA-CB-CG1-CD1
4	D	803	GOL	O1-C1-C2-C3
4	D	803	GOL	C1-C2-C3-O3

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	801	ILE	1	0
2	A	801	ILE	1	0
2	C	801	ILE	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, 95th 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^2)	Q<0.9
1	A	724/745 (97%)	-0.01	11 (1%)	73	71	21, 42, 68, 105	0
1	B	724/745 (97%)	0.06	7 (0%)	82	81	24, 45, 68, 106	0
1	C	724/745 (97%)	0.09	13 (1%)	68	64	22, 44, 73, 94	0
1	D	724/745 (97%)	-0.09	3 (0%)	92	92	18, 39, 61, 99	0
All	All	2896/2980 (97%)	0.01	34 (1%)	79	77	18, 42, 68, 106	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	241	TYR	4.4
1	C	245	PRO	4.3
1	A	241	TYR	4.2
1	A	245	PRO	4.2
1	A	142	GLY	3.8

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, 95th 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(Å^2)	Q<0.9
4	GOL	D	803	6/6	0.79	0.27	42,53,57,58	0
3	PRO	A	802	7/8	0.92	0.20	23,24,24,24	0
3	PRO	C	802	7/8	0.93	0.15	27,28,29,29	0
2	ILE	D	801	8/9	0.94	0.18	21,24,25,27	0
3	PRO	B	802	7/8	0.95	0.13	24,25,28,30	0
2	ILE	B	801	8/9	0.95	0.20	23,25,25,27	0
2	ILE	C	801	8/9	0.95	0.15	24,26,26,27	0
2	ILE	A	801	8/9	0.96	0.17	24,24,26,26	0
3	PRO	D	802	7/8	0.97	0.10	24,24,26,26	0

6.5 Other polymers (i)

There are no such residues in this entry.