wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 07:29 PM EDT

PDB ID	:	2G63
Title	:	Crystal structure of human dipeptidyl peptidase IV (DPPIV) complexed with cyanopyrrolidine (C5-pro-pro) inhibitor 24b
Authors	:	Longenecker, K.L.; Fry, E.H.; Lake, M.R.; Solomon, L.R.; Pei, Z.; Li, X.
Deposited on	:	2006-02-24
Resolution	:	2.00 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:

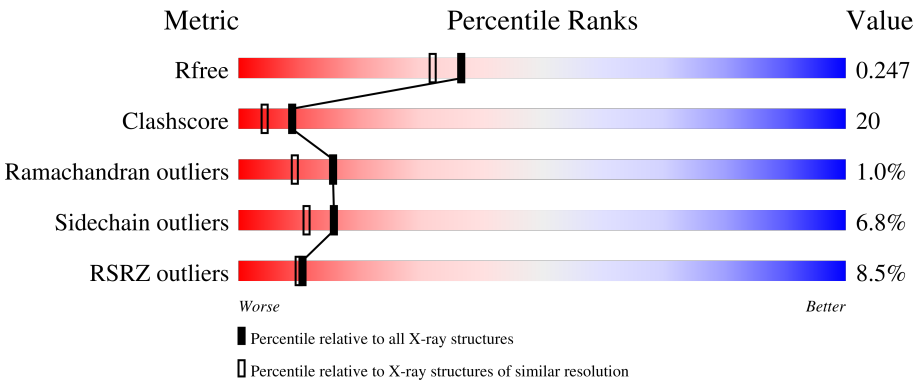
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	726	7% 70% 25% •
1	B	726	9% 69% 26% •
1	C	726	10% 66% 30% •
1	D	726	8% 68% 28% •

2 Entry composition (i)

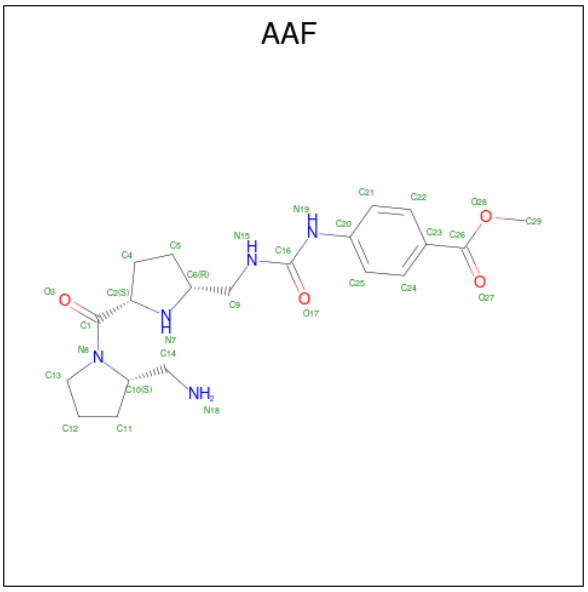
There are 3 unique types of molecules in this entry. The entry contains 28111 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 peptidase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	726	5949	3816	980	1127	26	0	0	0
1	B	726	5949	3816	980	1127	26	0	0	0
1	C	726	5949	3816	980	1127	26	0	0	0
1	D	726	5949	3816	980	1127	26	0	0	0

- Molecule 2 is METHYL 4-{[({[(2R,5S)-5-{[(2S)-2-(AMINOMETHYL)PYRROLIDIN-1-YL]CARBONYL}PYRROLIDIN-2-YL]METHYL}AMINO)CARBONYL]AMINO}BENZOATE (three-letter code: AAF) (formula: $C_{20}H_{29}N_5O_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	B	1	29	20	5	4	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1195	Total O 1195 1195	0	0
3	B	1121	Total O 1121 1121	0	0
3	C	938	Total O 938 938	0	0
3	D	1032	Total O 1032 1032	0	0

3 Residue-property plots (i)

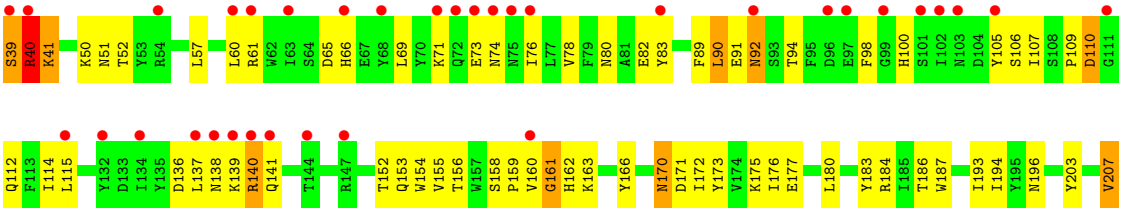
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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

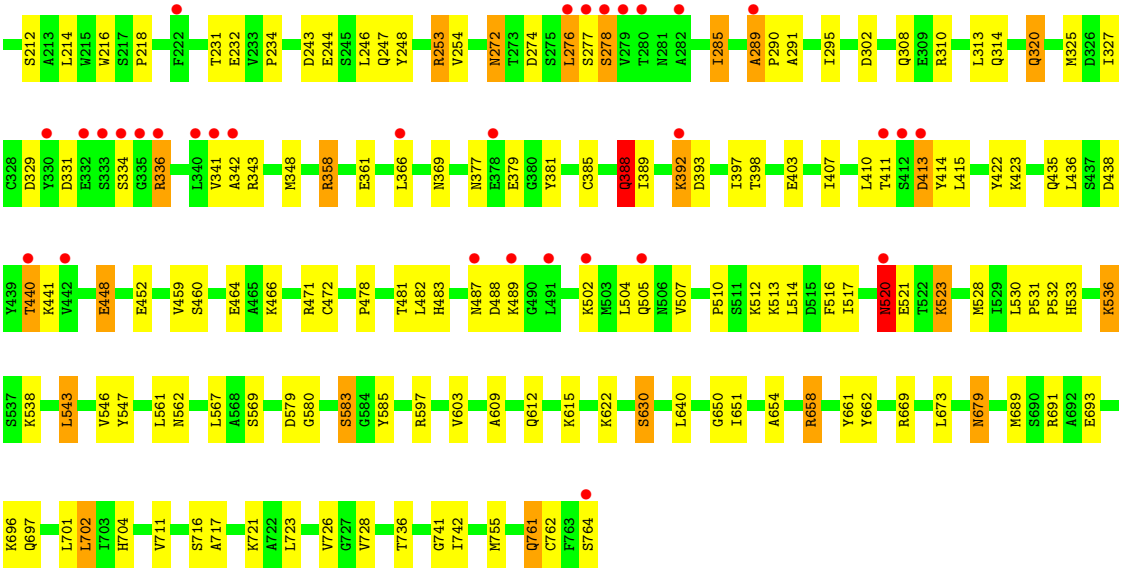
- Molecule 1: Dipeptidyl peptidase 4

Chain A:

- Molecule 1: Dipeptidyl peptidase 4

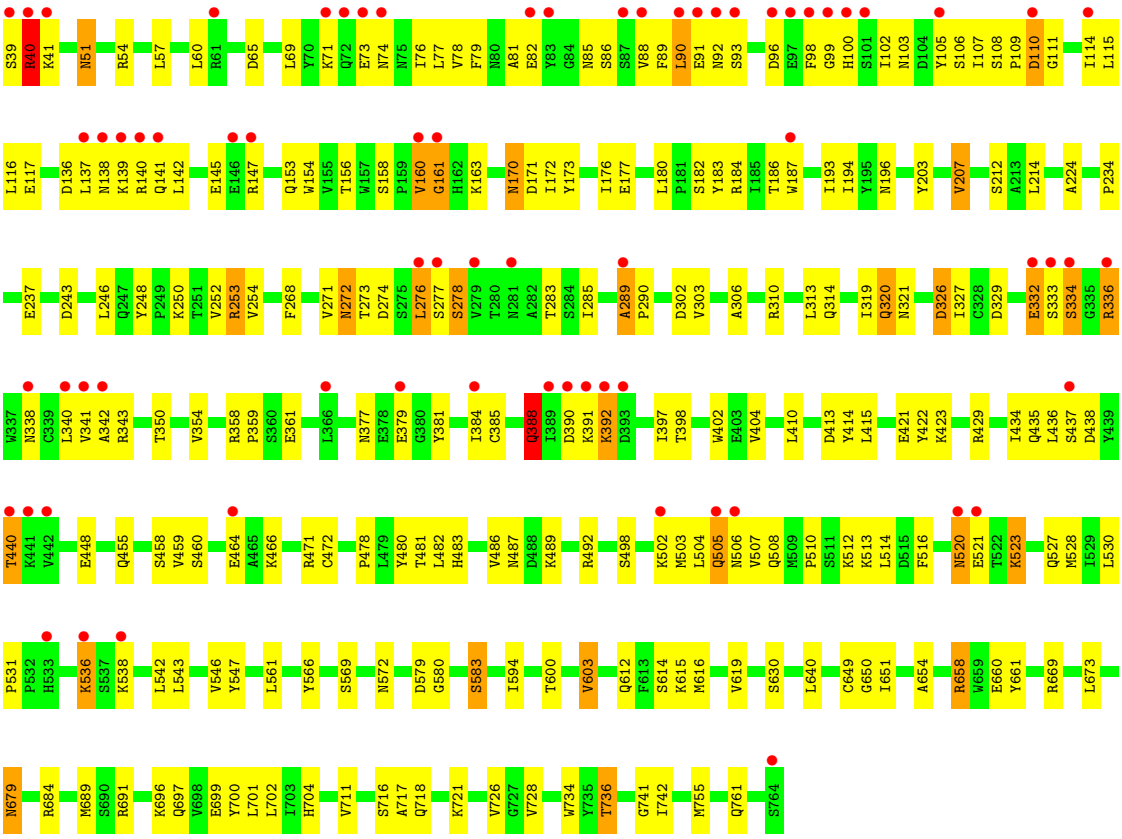
Chain B:





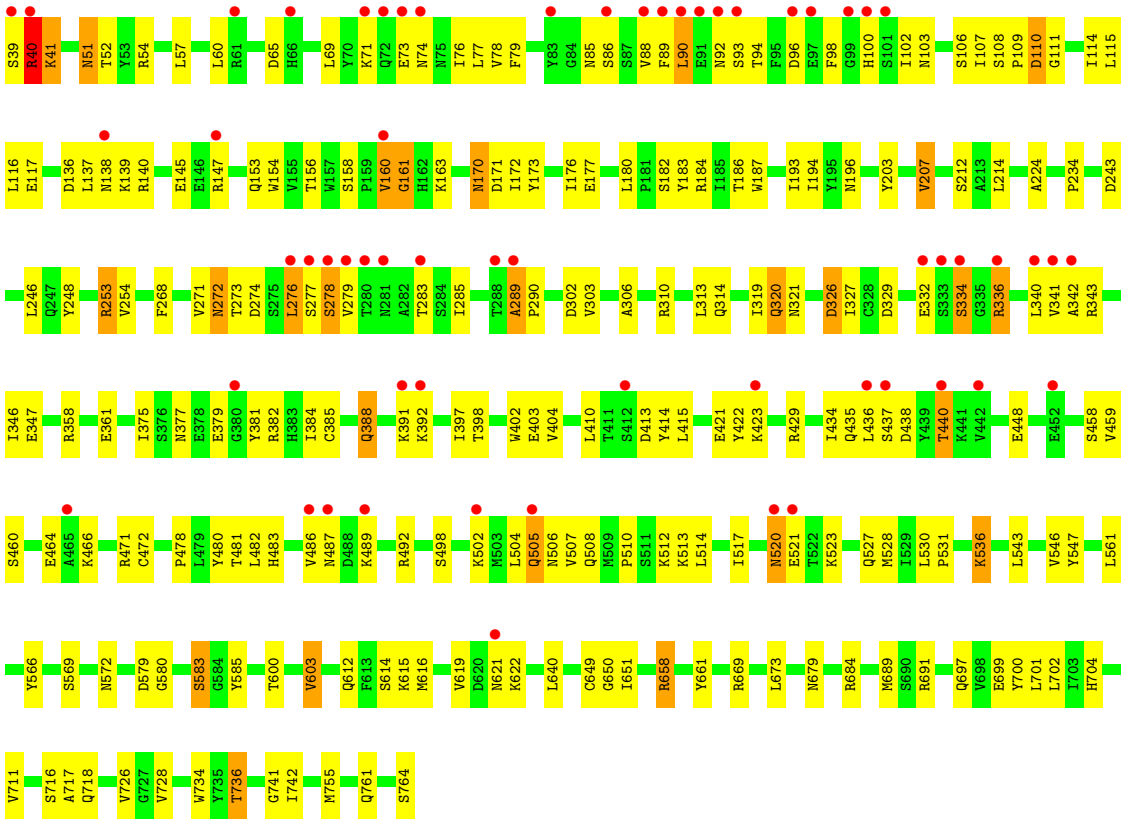
- Molecule 1: Dipeptidyl peptidase 4

Chain C:



- Molecule 1: Dipeptidyl peptidase 4

Chain D:


S39 R40 K41 N51 T52 Y53 R54 L57 L60 R61 D65 H66 L69 Y70 K71 Q72 E73 N74 N75 I76 L77 V78 F79 Y83 G84 N85 S86 S87 V88 F89 L90 E91 N92 S93 T94 F95 D96 E97 F98 G99 H100 S101 I102 N103 S106 I107 S108 P109 D110 G111 I114 L115
L116 E117 D136 L137 N138 K139 R140 E145 E146 R147 Q153 W154 V155 T156 W157 S158 P159 V160 G161 H162 K163 N170 D171 I172 Y173 I176 E177 L180 P181 S182 Y183 R184 I185 T186 W187 I193 I194 Y195 N196 Y203 V207 S212 A213 L214 A224 P234 D243
L246 Q247 Y248 R253 V254 F268 V271 N272 T273 D274 S275 L276 S277 S278 V279 T280 N281 A282 T283 S284 I285 T288 A289 P290 D302 V303 A306 R310 L313 Q314 I319 Q320 N321 D326 I327 C328 D329 E332 S333 S334 G335 R336 L340 V341 A342 R343
I346 E347 R358 E361 I375 S376 N377 E378 E379 G380 Y381 R382 H383 I384 C385 Q388 K391 K392 I397 T398 W402 E403 V404 L410 T411 S412 D413 Y414 L415 E421 Y422 K423 R429 I434 Q435 L436 S437 D438 Y439 T440 K441 V442 E448 E452 S458 V459
S460 E464 A465 K466 R471 C472 P478 L479 Y480 T481 L482 H483 V486 N487 D488 K489 R492 S498 K502 M503 L504 Q505 N506 V507 Q508 M509 P510 S511 K512 K513 L514 I517 N520 E521 T522 K523 Q527 M528 I529 L530 P531 K536 L543 V546 Y547 L561
V566 S569 N572 D579 G580 S583 G584 Y585 T600 V603 Q612 F613 S614 K615 M616 V619 D620 N621 K622 L640 C649 G650 I651 R658 Y661 R669 L673 N679 R684 M689 S690 R691 Q697 V698 E699 Y700 L701 L702 I703 H704
V711 S716 A717 Q718 V726 G727 V728 W734 Y735 T736 G741 I742 M755 Q761 S764

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	120.13Å 126.50Å 127.37Å 90.00° 96.66° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 19.96 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (20.00-2.00) 99.8 (19.96-2.00)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 2.01Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.215 , 0.251 0.210 , 0.247	Depositor DCC
R_{free} test set	12847 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 78.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	28111	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.54% of the height of the origin peak. No significant pseudotranslation is detected.*

[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: AAF

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	0.34	0/6120	0.67	4/8321 (0.0%)
1	B	0.36	1/6120 (0.0%)	0.71	4/8321 (0.0%)
1	C	0.32	0/6120	0.62	1/8321 (0.0%)
1	D	0.33	0/6120	0.62	1/8321 (0.0%)
All	All	0.34	1/24480 (0.0%)	0.66	10/33284 (0.0%)

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	C	0	1
1	D	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	630	SER	C-O	7.92	1.38	1.23

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	358	ARG	NE-CZ-NH2	18.50	129.55	120.30
1	B	358	ARG	NE-CZ-NH1	-17.66	111.47	120.30
1	A	358	ARG	NE-CZ-NH2	-10.76	114.92	120.30
1	A	358	ARG	NE-CZ-NH1	10.00	125.30	120.30
1	B	358	ARG	CD-NE-CZ	6.52	132.73	123.60

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	700	TYR	Sidechain
1	D	700	TYR	Sidechain

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	5949	0	5667	236	0
1	B	5949	0	5666	233	0
1	C	5949	0	5667	253	0
1	D	5949	0	5667	243	0
2	B	29	0	27	1	0
3	A	1195	0	0	63	0
3	B	1121	0	0	59	0
3	C	938	0	0	72	0
3	D	1032	0	0	67	0
All	All	28111	0	22694	949	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 20.

The worst 5 of 949 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:736:THR:HB	3:B:1870:HOH:O	1.37	1.24
1:A:289:ALA:HB1	1:A:290:PRO:HA	1.25	1.13
1:B:289:ALA:HB1	1:B:290:PRO:HA	1.28	1.12
1:A:736:THR:HG21	1:B:721:LYS:HB2	1.31	1.06
1:C:114:ILE:HD11	1:C:137:LEU:HD21	1.37	1.06

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	724/726 (100%)	678 (94%)	39 (5%)	7 (1%)	15	9
1	B	724/726 (100%)	678 (94%)	39 (5%)	7 (1%)	15	9
1	C	724/726 (100%)	677 (94%)	40 (6%)	7 (1%)	15	9
1	D	724/726 (100%)	676 (93%)	41 (6%)	7 (1%)	15	9
All	All	2896/2904 (100%)	2709 (94%)	159 (6%)	28 (1%)	15	9

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	40	ARG
1	A	278	SER
1	B	40	ARG
1	B	278	SER
1	C	40	ARG

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	651/651 (100%)	603 (93%)	48 (7%)	13	9
1	B	651/651 (100%)	605 (93%)	46 (7%)	14	10
1	C	651/651 (100%)	609 (94%)	42 (6%)	17	12
1	D	651/651 (100%)	610 (94%)	41 (6%)	18	13
All	All	2604/2604 (100%)	2427 (93%)	177 (7%)	16	11

5 of 177 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	392	LYS
1	D	160	VAL
1	C	482	LEU
1	C	679	ASN
1	D	276	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 83 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	572	ASN
1	D	377	ASN
1	C	679	ASN
1	D	123	GLN
1	D	533	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)

1 ligand is 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	AAF	B	800	1	31,31,31	1.48	7 (22%)	37,42,42	1.08	4 (10%)

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	AAF	B	800	1	-	2/25/44/44	0/3/3/3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	800	AAF	C1-N8	3.05	1.41	1.34
2	B	800	AAF	C14-C10	2.85	1.56	1.53
2	B	800	AAF	C25-C20	2.83	1.44	1.39
2	B	800	AAF	C14-N18	-2.78	1.34	1.48
2	B	800	AAF	C22-C23	2.58	1.43	1.39

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	800	AAF	C12-C13-N8	2.40	107.36	103.24
2	B	800	AAF	C29-O28-C26	2.39	120.40	115.81
2	B	800	AAF	C11-C10-C14	-2.29	103.61	112.17
2	B	800	AAF	C14-C10-N8	2.25	118.39	110.64

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	800	AAF	N8-C10-C14-N18
2	B	800	AAF	C11-C10-C14-N18

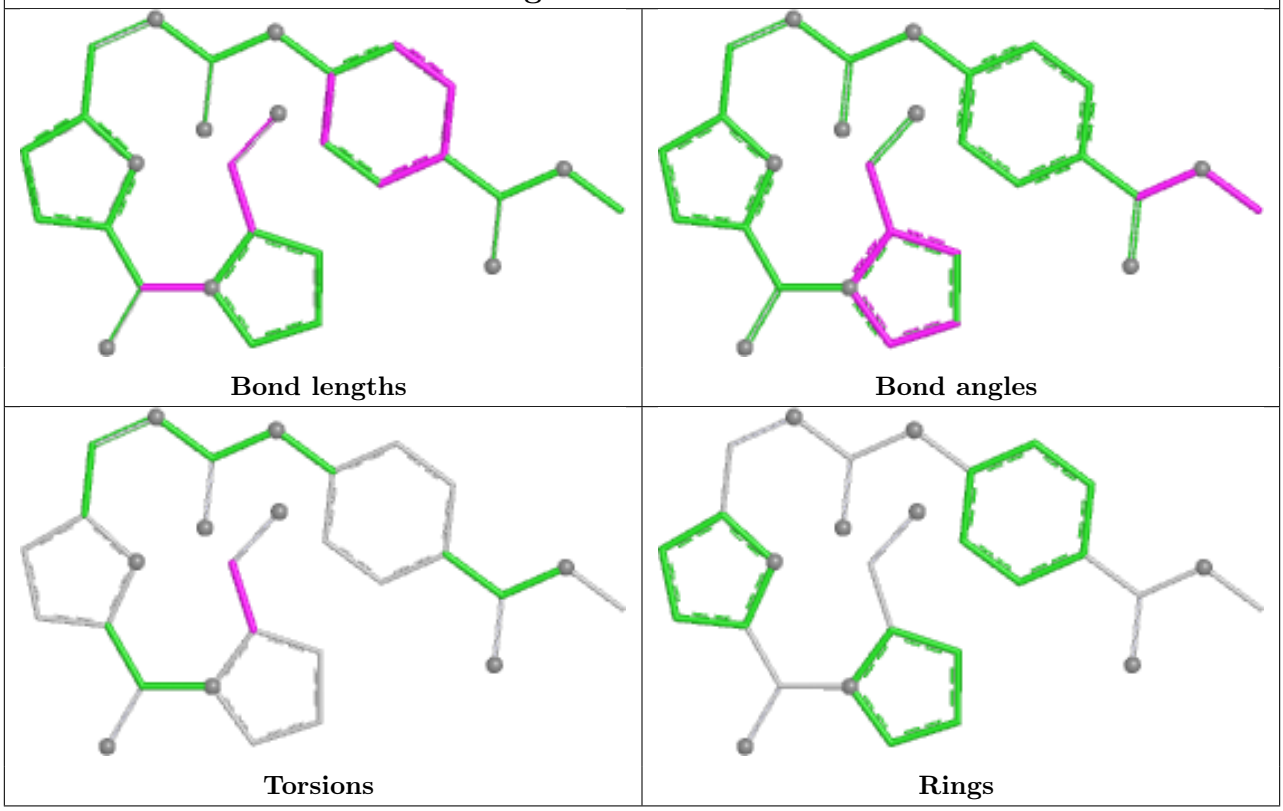
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	800	AAF	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand AAF B 800



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	726/726 (100%)	0.39	50 (6%)	16	16	16, 27, 56, 88	0
1	B	726/726 (100%)	0.43	67 (9%)	9	8	15, 28, 63, 85	0
1	C	726/726 (100%)	0.41	70 (9%)	8	7	18, 34, 61, 79	0
1	D	726/726 (100%)	0.35	59 (8%)	12	11	18, 32, 60, 79	0
All	All	2904/2904 (100%)	0.39	246 (8%)	10	10	15, 30, 60, 88	0

The worst 5 of 246 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	279	VAL	22.1
1	A	282	ALA	13.6
1	A	278	SER	12.0
1	C	39	SER	11.6
1	A	277	SER	10.7

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)

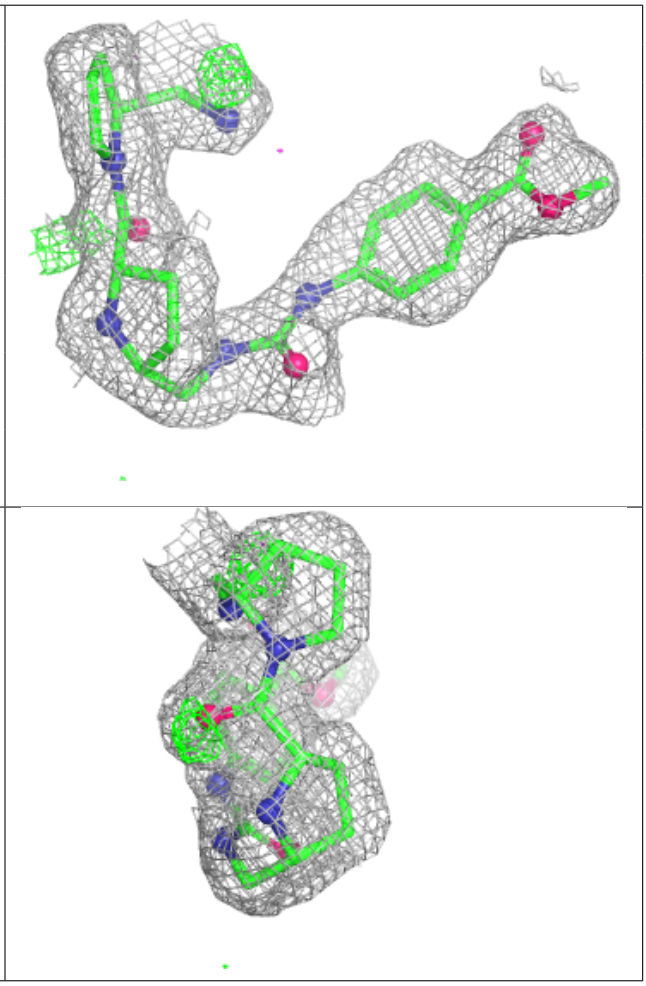
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	AAF	B	800	29/29	0.89	0.16	26,31,48,50	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around AAF B 800:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.