wwPDB X-ray Structure Validation Summary Report (i)

Mar 4, 2024 – 12:33 AM EST

PDB ID	:	1FEJ
Title	:	STRUCTURAL IMPLICATIONS OF DRUG RESISTANT MUTANTS OF HIV-1 PROTEASE: HIGH RESOLUTION CRYSTAL STRUCTURES OF THE MUTANT PROTEASE/SUBSTRATE ANALOG COMPLEXES
Authors	:	Mahalingam, B.; Louis, J.M.; Harrison, R.W.; Weber, I.T.
Deposited on	:	2000-07-21
Resolution	:	1.78 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 (1)) were used in the production of this report:

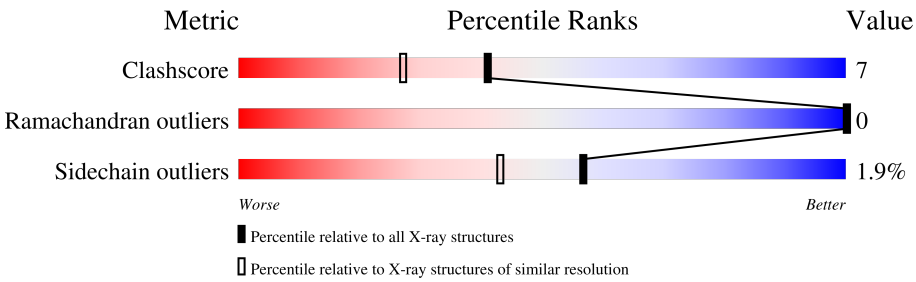
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.78 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))
Clashscore	141614	10184 (1.80-1.76)
Ramachandran outliers	138981	10051 (1.80-1.76)
Sidechain outliers	138945	10050 (1.80-1.76)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	99	84% 15% •
1	D	99	87% 12% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1714 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.

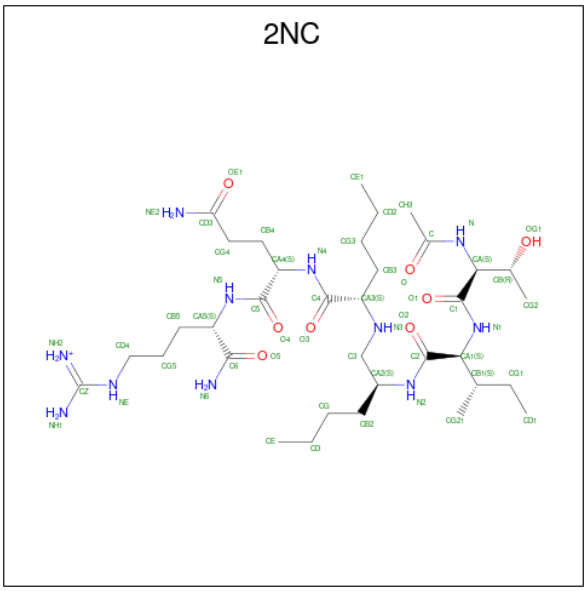
- Molecule 1 is a protein called PROTEASE RETROPEPSIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	C	99	756	489	130	134	3	0	0	0
1	D	99	756	489	130	134	3	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	7	LYS	GLN	engineered mutation	UNP P04587
C	33	ILE	LEU	engineered mutation	UNP P04587
C	63	ILE	LEU	engineered mutation	UNP P04587
C	67	ALA	CYS	engineered mutation	UNP P04587
C	90	MET	LEU	engineered mutation	UNP P04587
C	95	ALA	CYS	engineered mutation	UNP P04587
D	107	LYS	GLN	engineered mutation	UNP P04587
D	133	ILE	LEU	engineered mutation	UNP P04587
D	163	ILE	LEU	engineered mutation	UNP P04587
D	167	ALA	CYS	engineered mutation	UNP P04587
D	190	MET	LEU	engineered mutation	UNP P04587
D	195	ALA	CYS	engineered mutation	UNP P04587

- Molecule 2 is N-{(2S)-2-[(N-acetyl-L-threonyl-L-isoleucyl)amino]hexyl}-L-norleucyl-L-glutaminyl-N 5 -[amino(iminio)methyl]-L-ornithinamide (three-letter code: 2NC) (formula: $C_{35}H_{68}N_{11}O_8$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	D	1	54	35	11	8	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	C	64	64	64	0	0
			Total	O		
3	D	84	84	84	0	0

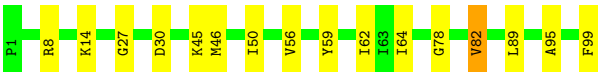
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

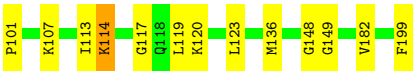
• Molecule 1: PROTEASE RETROPEPSIN

Chain C:



• Molecule 1: PROTEASE RETROPEPSIN

Chain D:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	51.23Å 58.05Å 61.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 1.78	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-1.78)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.843	Depositor
R, R_{free}	0.215 , 0.285	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1714	wwPDB-VP
Average B, all atoms (Å^2)	13.0	wwPDB-VP

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: 2NC

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	C	0.35	0/769	0.64	0/1039
1	D	0.38	0/769	0.67	0/1039
All	All	0.36	0/1538	0.66	0/2078

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	C	756	0	809	13	0
1	D	756	0	807	11	0
2	D	54	0	68	9	0
3	C	64	0	0	1	0
3	D	84	0	0	1	0
All	All	1714	0	1684	23	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 7.

The worst 5 of 23 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:C:8:ARG:NH2	2:D:400:2NC:HD22	1.90	0.85
1:C:82:VAL:HG11	2:D:400:2NC:HE31	1.67	0.76
1:C:27:GLY:HA2	1:D:123:LEU:HD21	1.81	0.61
2:D:400:2NC:HE11	2:D:400:2NC:HG29	1.83	0.59
1:C:59:TYR:HB3	1:C:62:ILE:HD11	1.84	0.58

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	C	97/99 (98%)	97 (100%)	0	0	100	100
1	D	97/99 (98%)	96 (99%)	1 (1%)	0	100	100
All	All	194/198 (98%)	193 (100%)	1 (0%)	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	C	81/81 (100%)	80 (99%)	1 (1%)	71	62
1	D	81/81 (100%)	79 (98%)	2 (2%)	47	31
All	All	162/162 (100%)	159 (98%)	3 (2%)	57	43

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	82	VAL
1	D	107	LYS
1	D	114	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	18	GLN

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	2NC	D	400	-	53,53,53	1.03	2 (3%)	64,68,68	1.11	2 (3%)

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	2NC	D	400	-	-	13/75/75/75	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	400	2NC	C3-N3	-5.95	1.32	1.47
2	D	400	2NC	C6-N6	-3.05	1.25	1.32

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	400	2NC	C3-CA2-N2	5.97	122.22	109.05
2	D	400	2NC	C4-CA3-N3	-3.12	104.17	112.73

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	400	2NC	N3-CA3-CB3-CG3
2	D	400	2NC	C4-CA3-CB3-CG3
2	D	400	2NC	O2-C2-CA1-N1
2	D	400	2NC	N2-C2-CA1-N1
2	D	400	2NC	O2-C2-CA1-CB1

There are no ring outliers.

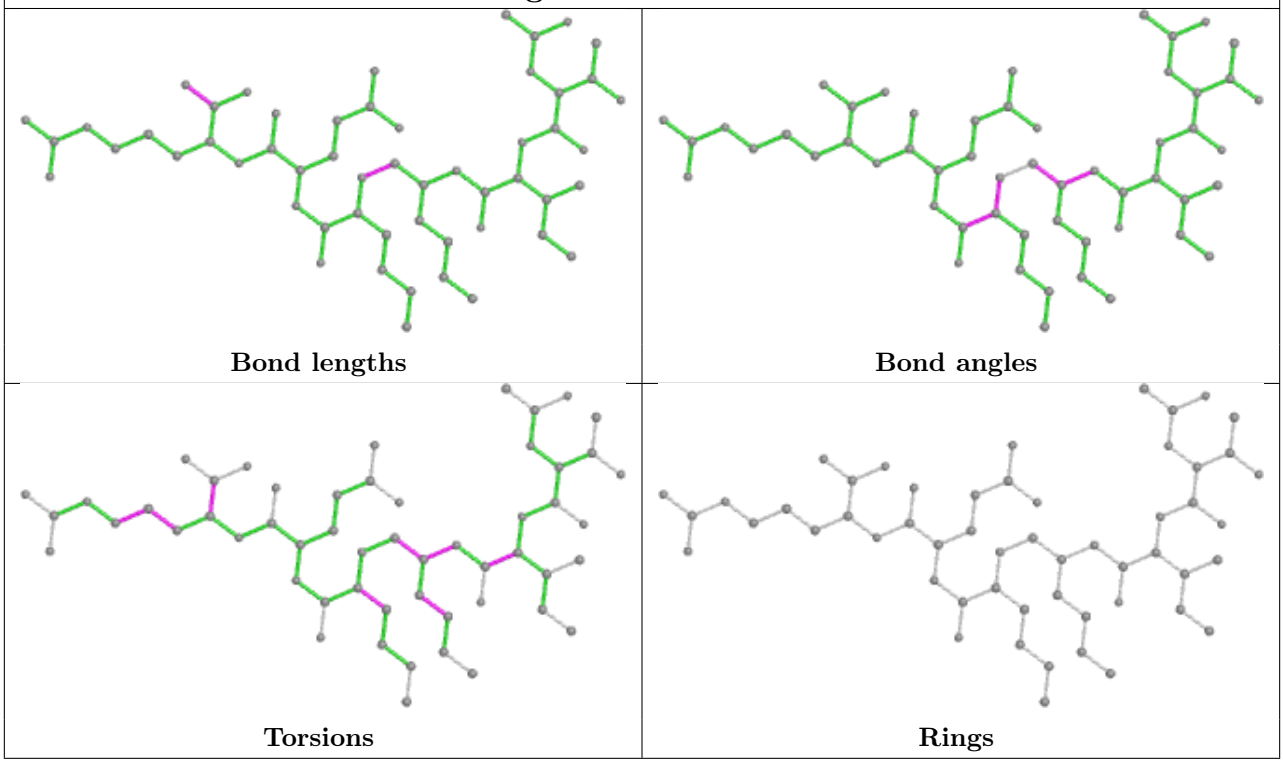
1 monomer is involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	400	2NC	9	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 2NC D 400



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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.