wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2023 – 10:53 PM EDT

PDB ID	:	1KJ4
Title	:	SUBSTRATE SHAPE DETERMINES SPECIFICITY OF RECOGNITION RECOGNITION FOR HIV-1 PROTEASE: ANALYSIS OF CRYSTAL STRUCTURES OF SIX SUBSTRATE COMPLEXES
Authors	:	Schiffer, C.A.
Deposited on	:	2001-12-04
Resolution	:	2.90 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:

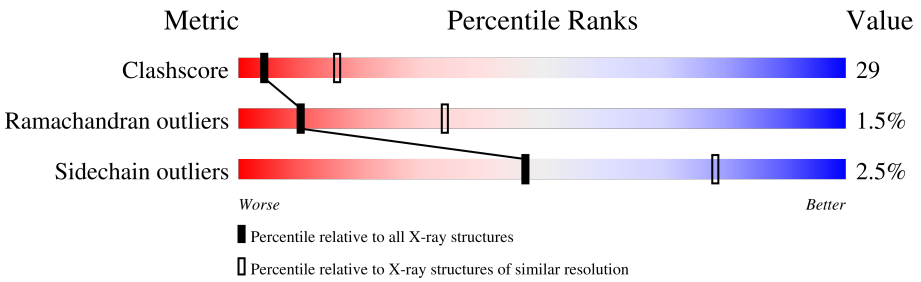
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.90 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	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)

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	A	99	57% 39% .
1	B	99	59% 39% .
1	C	99	48% 49% .
1	D	99	73% 25% .
2	P	10	30% 60% 10%
2	S	10	20% 60% 20%

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ACT	A	506	-	-	X	-
3	ACT	B	522	-	-	X	-
3	ACT	C	516	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3170 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 POL polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	99	734	473	127	130	4	0	0	0
1	B	99	738	475	127	132	4	0	0	0
1	C	99	729	472	125	128	4	0	0	0
1	D	99	731	470	125	132	4	0	0	0

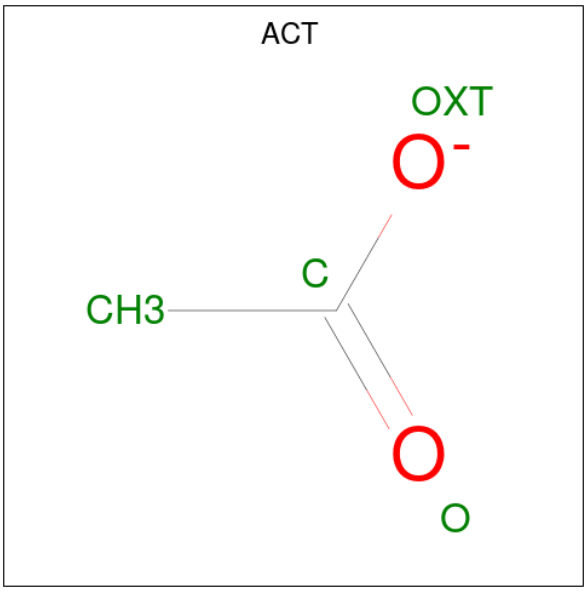
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	LYS	GLN	engineered mutation	UNP P03369
A	25	ASN	ASP	engineered mutation	UNP P03369
B	7	LYS	GLN	engineered mutation	UNP P03369
B	25	ASN	ASP	engineered mutation	UNP P03369
C	7	LYS	GLN	engineered mutation	UNP P03369
C	25	ASN	ASP	engineered mutation	UNP P03369
D	7	LYS	GLN	engineered mutation	UNP P03369
D	25	ASN	ASP	engineered mutation	UNP P03369

- Molecule 2 is a protein called gag polyprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	P	9	76	50	12	14	0	1	0
2	S	8	71	47	11	13	0	1	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	4	2	2	0	0
3	A	1	4	2	2	0	0
3	A	1	4	2	2	0	0
3	A	1	4	2	2	0	0
3	B	1	4	2	2	0	0
3	B	1	4	2	2	0	0
3	B	1	4	2	2	0	0
3	C	1	4	2	2	0	0
3	C	1	4	2	2	0	0
3	D	1	4	2	2	0	0
3	D	1	4	2	2	0	0
3	D	1	4	2	2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	10	10	10	0	0
			Total	O		
4	B	7	7	7	0	0
			Total	O		
4	C	11	11	11	0	0
			Total	O		
4	D	14	14	14	0	0
			Total	O		
4	S	1	1	1	0	0

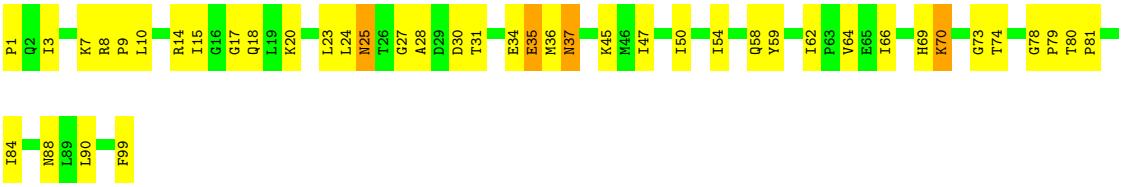
3 Residue-property plots (i)

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.

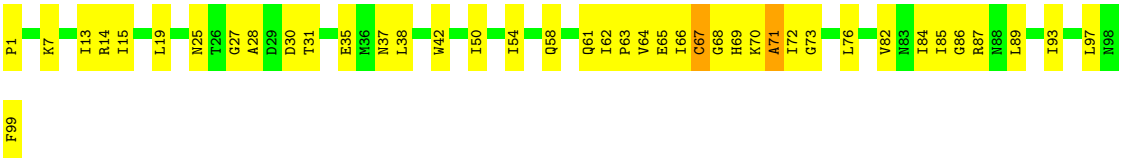
• Molecule 1: POL polyprotein

Chain A:



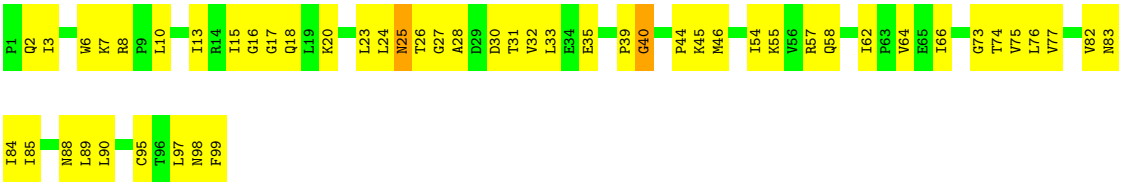
• Molecule 1: POL polyprotein

Chain B:



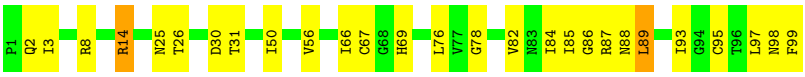
• Molecule 1: POL polyprotein

Chain C:



• Molecule 1: POL polyprotein

Chain D:



• Molecule 2: gag polyprotein

Chain P: 30% 60% 10%

V1 S2 Q3 N4 Y5 P6 I7 V8 Q9 ASN

- Molecule 2: gag polyprotein

Chain S: 20% 60% 20%

V1 S2 Q3 N4 Y5 P6 I7 V8 GLN ASN

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	91.65Å 93.81Å 118.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.88 – 2.90	Depositor
% Data completeness (in resolution range)	85.6 (38.88-2.90)	Depositor
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.197 , 0.248	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3170	wwPDB-VP
Average B, all atoms (Å^2)	36.0	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: ACT

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	0.38	0/748	0.66	0/1019
1	B	0.39	0/752	0.66	0/1024
1	C	0.37	0/743	0.69	0/1012
1	D	0.37	0/745	0.67	0/1017
2	P	0.45	0/63	0.82	0/86
2	S	0.49	0/58	0.93	0/79
All	All	0.38	0/3109	0.68	0/4237

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	A	734	0	754	54	0
1	B	738	0	758	47	0
1	C	729	0	752	70	0
1	D	731	0	738	33	0
2	P	76	0	78	19	0
2	S	71	0	76	21	0
3	A	16	0	12	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	12	0	9	4	0
3	C	8	0	6	10	0
3	D	12	0	9	0	0
4	A	10	0	0	0	0
4	B	7	0	0	1	0
4	C	11	0	0	1	0
4	D	14	0	0	0	0
4	S	1	0	0	0	0
All	All	3170	0	3192	186	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 29.

The worst 5 of 186 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:74:THR:H	3:C:516:ACT:CH3	1.57	1.16
1:C:74:THR:H	3:C:516:ACT:H2	1.17	1.08
1:A:70:LYS:H	1:A:70:LYS:HD3	1.22	1.01
1:B:82:VAL:HG11	2:S:1[2]:VAL:HG23	1.37	1.01
2:P:3:GLN:HE22	2:S:1[2]:VAL:HG22	1.28	0.98

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	97/99 (98%)	91 (94%)	5 (5%)	1 (1%)	15 45
1	B	97/99 (98%)	87 (90%)	8 (8%)	2 (2%)	7 26
1	C	97/99 (98%)	81 (84%)	13 (13%)	3 (3%)	4 16

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	97/99 (98%)	90 (93%)	7 (7%)	0	100	100
2	P	7/10 (70%)	6 (86%)	1 (14%)	0	100	100
2	S	6/10 (60%)	5 (83%)	1 (17%)	0	100	100
All	All	401/416 (96%)	360 (90%)	35 (9%)	6 (2%)	10	34

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	67	CYS
1	B	71	ALA
1	C	17	GLY
1	A	35	GLU
1	C	35	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	77/83 (93%)	74 (96%)	3 (4%)	32	66
1	B	78/83 (94%)	77 (99%)	1 (1%)	69	90
1	C	76/83 (92%)	75 (99%)	1 (1%)	69	90
1	D	76/83 (92%)	73 (96%)	3 (4%)	32	66
2	P	7/10 (70%)	7 (100%)	0	100	100
2	S	7/10 (70%)	7 (100%)	0	100	100
All	All	321/352 (91%)	313 (98%)	8 (2%)	47	78

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	89	LEU
1	D	67	CYS
1	C	25	ASN
1	B	37	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	14	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	69	HIS
1	D	88	ASN
2	S	3	GLN
2	P	3	GLN
1	C	2	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)

12 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	$\#\|Z\| > 2$
3	ACT	B	502	-	3,3,3	2.76	1 (33%)	3,3,3	1.29	0
3	ACT	A	512	-	3,3,3	2.71	1 (33%)	3,3,3	1.25	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	ACT	D	511	-	3,3,3	2.76	1 (33%)	3,3,3	1.11	0
3	ACT	C	516	-	3,3,3	2.78	1 (33%)	3,3,3	0.79	0
3	ACT	A	504	-	3,3,3	2.77	1 (33%)	3,3,3	1.69	1 (33%)
3	ACT	B	501	-	3,3,3	2.76	1 (33%)	3,3,3	1.07	0
3	ACT	D	518	-	3,3,3	2.71	1 (33%)	3,3,3	1.30	0
3	ACT	D	523	-	3,3,3	2.80	1 (33%)	3,3,3	1.12	0
3	ACT	A	506	-	3,3,3	2.75	1 (33%)	3,3,3	0.97	0
3	ACT	A	521	-	3,3,3	2.79	1 (33%)	3,3,3	1.26	0
3	ACT	C	524	-	3,3,3	2.78	1 (33%)	3,3,3	1.13	0
3	ACT	B	522	-	3,3,3	2.78	1 (33%)	3,3,3	0.99	0

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	516	ACT	O-C	4.76	1.44	1.22
3	D	523	ACT	O-C	4.73	1.43	1.22
3	C	524	ACT	O-C	4.73	1.43	1.22
3	B	522	ACT	O-C	4.72	1.43	1.22
3	A	506	ACT	O-C	4.71	1.43	1.22

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	504	ACT	OXT-C-O	-2.59	112.49	122.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	502	ACT	1	0
3	C	516	ACT	10	0
3	A	504	ACT	1	0
3	A	506	ACT	7	0
3	B	522	ACT	3	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)

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.