wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2020 – 10:17 pm BST

PDB ID	:	1PEK
Title	:	STRUCTURE OF THE COMPLEX OF PROTEINASE K WITH A SUBSTRATE-ANALOGUE HEXA-PEPTIDE INHIBITOR AT 2.2 ANGSTROMS RESOLUTION
Authors	:	Betzel, C.; Singh, T.P.; Visanji, M.; Peters, K.; Fittkau, S.; Saenger, W.; Wilson, K.S.
Deposited on	:	1993-01-19
Resolution	:	2.20 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 following versions of software and data (see references (1)) were used in the production of this report:

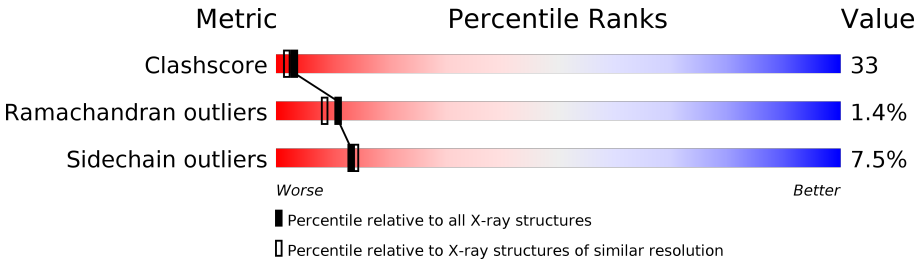
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.20 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	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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 >=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	E	279	38% 42% 17%
2	C	4	75% 25%
3	D	3	67% 33%

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
3	DAL	D	5	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2246 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 PROTEINASE K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	279	Total	C	N	O	S	0	0	0
			2018	1243	352	413	10			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	85	VAL	ALA	CONFLICT	UNP P06873

- Molecule 2 is a protein called PEPTIDE PRO-ALA-PRO-PHE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	4	Total	C	N	O	1	0	0
			30	22	4	4			

- Molecule 3 is a protein called D-DAL-ALA-NH2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	D	2	Total	C	N	O	0	0	0
			10	6	2	2			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	183	Total	O	0	0
			183	183		
4	C	1	Total	O	0	0
			1	1		
4	D	4	Total	O	0	0
			4	4		

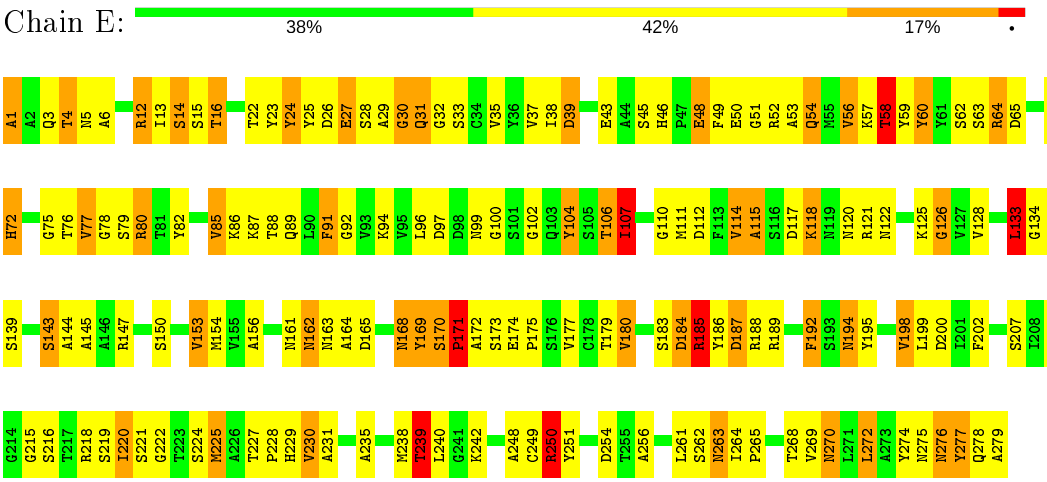
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

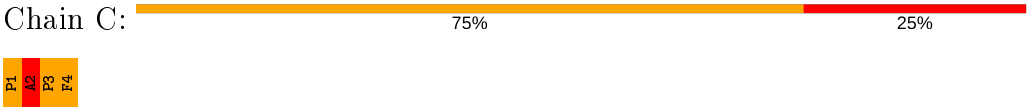
- Molecule 1: PROTEINASE K

Chain E:



- Molecule 2: PEPTIDE PRO-ALA-PRO-PHE

Chain C:



- Molecule 3: D-DAL-ALA-NH2

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 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	68.28Å 68.28Å 107.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 2.20	Depositor
% Data completeness (in resolution range)	95.0 (8.00-2.20)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.165 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2246	wwPDB-VP
Average B, all atoms (Å^2)	11.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: DAL

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	E	1.53	11/2057 (0.5%)	3.20	228/2797 (8.2%)
2	C	1.72	0/32	6.50	19/43 (44.2%)
3	D	31.88	4/4 (100.0%)	19.31	4/4 (100.0%)
All	All	2.07	15/2093 (0.7%)	3.35	251/2844 (8.8%)

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	E	1	4
2	C	1	2
3	D	0	1
All	All	2	7

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	6	ALA	C-O	54.54	2.27	1.23
3	D	6	ALA	N-CA	26.88	2.00	1.46
3	D	6	ALA	CA-C	16.43	1.95	1.52
1	E	168	ASN	C-O	10.27	1.42	1.23
3	D	6	ALA	CA-CB	9.83	1.73	1.52

The worst 5 of 251 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	80	ARG	CD-NE-CZ	37.31	175.84	123.60
1	E	185	ARG	NE-CZ-NH2	-23.10	108.75	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	168	ASN	O-C-N	-23.04	85.83	122.70
3	D	6	ALA	CB-CA-C	22.91	144.46	110.10
2	C	3	PRO	C-N-CA	20.05	171.84	121.70

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	E	169	TYR	CA
2	C	2	ALA	CA

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	1	PRO	Mainchain
2	C	2	ALA	Mainchain
1	E	168	ASN	Mainchain
1	E	170	SER	Mainchain,Peptide
1	E	185	ARG	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	E	2018	0	1915	122	2
2	C	30	0	29	10	0
3	D	10	0	8	32	0
4	C	1	0	0	0	0
4	D	4	0	0	4	0
4	E	183	0	0	11	4
All	All	2246	0	1952	133	4

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 33.

The worst 5 of 133 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:E:220:ILE:HB	3:D:6:ALA:CB	1.51	1.39
3:D:6:ALA:C	3:D:6:ALA:CA	1.95	1.32
1:E:220:ILE:CD1	3:D:6:ALA:HB3	1.60	1.30
3:D:6:ALA:N	3:D:6:ALA:CA	2.00	1.25
3:D:5:DAL:O	3:D:5:DAL:C	1.88	1.22

All (4) 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)
4:E:457:HOH:O	4:E:457:HOH:O[7_465]	0.95	1.25
1:E:150:SER:CB	4:E:423:HOH:O[7_465]	1.41	0.79
1:E:150:SER:OG	4:E:423:HOH:O[7_465]	1.54	0.66
4:E:342:HOH:O	4:E:422:HOH:O[5_444]	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	E	277/279 (99%)	256 (92%)	18 (6%)	3 (1%)	14 12
2	C	2/4 (50%)	1 (50%)	0	1 (50%)	0 0
All	All	279/283 (99%)	257 (92%)	18 (6%)	4 (1%)	11 8

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	169	TYR
1	E	171	PRO
2	C	2	ALA
1	E	39	ASP

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	E	211/214 (99%)	195 (92%)	16 (8%)	13 14
2	C	3/3 (100%)	3 (100%)	0	100 100
All	All	214/217 (99%)	198 (92%)	16 (8%)	13 14

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	133	LEU
1	E	162	ASN
1	E	239	THR
1	E	107	ILE
1	E	240	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	162	ASN
1	E	168	ASN
1	E	229	HIS
1	E	119	ASN
1	E	194	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

1 non-standard protein/DNA/RNA residue is modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.