wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:39 AM GMT

PDB ID	:	2NPH
Title	:	Crystal structure of HIV1 protease in situ product complex
Authors	:	Hosur, M.V.; Das, A.; Prashar, V.
Deposited on	:	2006-10-27
Resolution	:	1.65 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

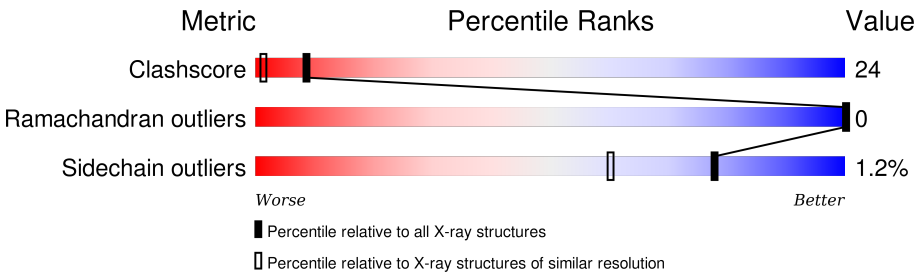
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 1.65 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	102246	1323 (1.66-1.66)
Ramachandran outliers	100387	1295 (1.66-1.66)
Sidechain outliers	100360	1295 (1.66-1.66)

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. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	99	68% 31% .
1	B	99	69% 31%
2	S	4	100%
3	T	5	40% 60%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1741 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	A	99	759	490	130	135	4	0	0	0
1	B	99	756	488	130	135	3	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	MET	CYS	ENGINEERED	UNP Q72874
B	1095	MET	CYS	ENGINEERED	UNP Q72874

- Molecule 2 is a protein called tetrapeptide fragment.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	S	4	33	21	4	8	0	0	0

- Molecule 3 is a protein called pentapeptide fragment.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	T	5	37	23	5	9	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	82	82	82	0	0
4	B	72	72	72	0	0
4	T	2	2	2	0	0

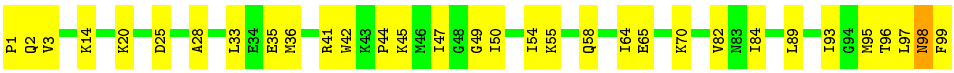
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

Note EDS was not executed.

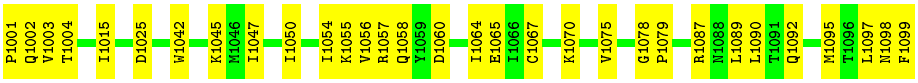
- Molecule 1: PROTEASE RETROPEPSIN

Chain A:



- Molecule 1: PROTEASE RETROPEPSIN

Chain B:

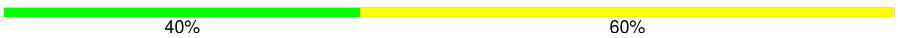


- Molecule 2: tetrapeptide fragment

Chain S:

- Molecule 3: pentapeptide fragment

Chain T:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	62.03Å 62.03Å 81.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	32.53 – 1.65	Depositor
% Data completeness (in resolution range)	100.0 (32.53-1.65)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
Refinement program	REFMAC 5.2	Depositor
R, R_{free}	0.214 , 0.259	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1741	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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	0.79	0/772	0.89	0/1044
1	B	0.77	0/769	0.90	0/1041
2	S	0.64	0/33	0.52	0/42
3	T	0.81	0/37	0.73	0/48
All	All	0.78	0/1611	0.89	0/2175

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	759	0	808	54	0
1	B	756	0	799	35	0
2	S	33	0	29	7	0
3	T	37	0	29	4	0
4	A	82	0	0	17	5
4	B	72	0	0	6	4
4	T	2	0	0	0	0
All	All	1741	0	1665	77	5

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 24.

The worst 5 of 77 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:20:LYS:NZ	4:A:399:HOH:O	1.56	1.29
1:B:1067:CYS:SG	4:B:488:HOH:O	1.98	1.17
1:A:55:LYS:HG3	4:A:470:HOH:O	1.59	1.01
1:A:95:MET:HE3	1:B:1099:PHE:HB3	1.50	0.91
1:A:35:GLU:O	4:A:406:HOH:O	1.96	0.83

All (5) 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:A:316:HOH:O	4:B:475:HOH:O[6_665]	1.16	1.04
4:A:345:HOH:O	4:B:395:HOH:O[2_665]	1.34	0.86
4:A:447:HOH:O	4:B:494:HOH:O[2_565]	1.44	0.76
4:A:344:HOH:O	4:B:393:HOH:O[2_665]	1.52	0.68
4:A:398:HOH:O	4:A:500:HOH:O[2_665]	1.69	0.51

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%)	96 (99%)	1 (1%)	0	100	100
1	B	97/99 (98%)	95 (98%)	2 (2%)	0	100	100
2	S	2/4 (50%)	1 (50%)	1 (50%)	0	100	100
3	T	3/5 (60%)	2 (67%)	1 (33%)	0	100	100
All	All	199/207 (96%)	194 (98%)	5 (2%)	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	A	83/83 (100%)	82 (99%)	1 (1%)	78	60
1	B	82/83 (99%)	81 (99%)	1 (1%)	78	60
2	S	3/3 (100%)	3 (100%)	0	100	100
3	T	3/3 (100%)	3 (100%)	0	100	100
All	All	171/172 (99%)	169 (99%)	2 (1%)	78	60

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	98	ASN
1	B	1060	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1002	GLN
1	B	1098	ASN
1	B	1007	GLN
1	A	7	GLN
1	B	1061	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 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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

6.3 Carbohydrates (i)

EDS was not executed - this section will therefore be empty.

6.4 Ligands (i)

EDS was not executed - this section will therefore be empty.

6.5 Other polymers (i)

EDS was not executed - this section will therefore be empty.