STRUCTURAL
BIOLOGY

Volume 75 (2019)

Supporting information for article:

Comparison of a retroviral protease in monomeric and dimeric states

Stanislaw Wosicki, Miroslaw Gilski, Helena Zabranska, Iva Pichova and Mariusz Jaskolski

Supporting information

S1. PDB screening for water molecules neighboring the active site of retroviral proteases

Screening of the Protein Data Bank holdings using the built-in search engine on its website (www.rcsb.org) with the following key-words: *Taxonomy Tree = Retroviridae, Experimental Method = X-RAY, Molecule Name = protease*; resulted in 708 unique entries (see S1.1.). All results were then screened using PyMol environment for water molecules within 3.5 Å of the catalytic aspartate or asparagine (in cases of mutation). The output contains 65 structures (see S1.2.), all of which were checked manually to verify the location of water molecules using Coot. The search was done on 25th of January 2019.

S1.1. List of tested entries:

1A30, 1A8G, 1A8K, 1A94, 1A9M, 1AAQ, 1AID, 1AJV, 1AJX, 1AXA, 1AZ5, 1B11, 1BAI, 1BDL, 1BDQ, 1BDR, 1BV7, 1BV9, 1BWA, 1BWB, 1C6X, 1C6Y, 1C6Z, 1C70, 1CPI, 1D4H, 1D4I, 1D4J, 1D4K, 1D4L, 1D4S, 1D4Y, 1DAZ, 1DIF, 1DMP, 1DW6, 1EBK, 1EBW, 1EBY, 1EBZ, 1EC0, 1EC1, 1EC2, 1EC3, 1FB7, 1FEJ, 1FF0, 1FFF, 1FFI, 1FG6, 1FG8, 1FGC, 1FIV, 1FMB, 1FQX, 1G2K, 1G35, 1G6L, 1GNM, 1GNN, 1GNO, 1HBV, 1HEF, 1HEG, 1HHP, 1HIH, 1HII, 1HIV, 1HOS, 1HPO, 1HPS, 1HPV, 1HPX, 1HSG, 1HSH, 1HSI, 1HTE, 1HTF, 1HTG, 1HVC, 1HVH, 1HVI, 1HVJ, 1HVK, 1HVL, 1HVR, 1HVS, 1HWR, 1HXB, 1HXW, 1IDA, 1IDB, 1IIQ, 1IVP, 1IVQ, 1IZH, 1IZI, 1JLD, 1K1T, 1K1U, 1K2B, 1K2C, 1KZK, 1LV1, 1LZQ, 1M0B, 1MER, 1MES, 1MET, 1MEU, 1MT7, 1MT8, 1MT9, 1MTB, 1MTR, 1MUI, 1MVP, 1N49, 1NH0, 1ODW, 1ODX, 1ODY, 1OHR, 1PRO, 1QBR, 1QBS, 1QBT, 1QBU, 1RL8, 1RPI, 1RQ9, 1RV7, 1SBG, 1SDT, 1SDU, 1SDV, 1SIP, 1SIV, 1SP5, 1T3R, 1T7K, 1TCW, 1TCX, 1TW7, 1U8G, 1UPJ, 1VIJ, 1VIK, 1W5V, 1W5W, 1W5X, 1W5Y, 1WBK, 1WBM, 1XL2, 1XL5, 1YTG, 1YTH, 1YTI, 1YTJ, 1Z8C, 1ZBG, 1ZJ7, 1ZLF, 1ZP8, 1ZPK, 1ZSF, 1ZSR, 1ZTZ, 2A1E, 2A4F, 2AID, 2AOC, 2AOD, 2AOG, 2AQU, 2AVM, 2AVO, 2AVQ, 2AVS, 2AVV, 2AZ8, 2AZ9, 2AZB, 2AZC, 2B7F, 2B7Z, 2BB9, 2BBB, 2BPV, 2BPW, 2BPX, 2BPY, 2BPZ, 2BQV, 2CEJ, 2CEM, 2CEN, 2F3K, 2FGU, 2FGV, 2FIV, 2FLE, 2FMB, 2FNS, 2FNT, 2G69, 2HAH, 2HB2, 2HB3, 2HB4, 2HC0, 2HPE, 2HPF, 2HS1, 2HS2, 2HVP, 2I0A, 2I0D, 2I4D, 2I4U, 2I4V, 2I4W, 2I4X, 2IDW, 2IEN, 2IEO, 2J9J, 2J9K, 2JE4, 2MIP, 2NMY, 2NMZ, 2NNK, 2NNP, 2NPH, 2NXD, 2NXL, 2NXM, 2O40, 2O4K, 2O4L, 2O4N, 2O4P, 2O4S, 2P3A, 2P3B, 2P3C, 2P3D, 2PC0, 2PK5, 2PK6, 2PSU, 2PSV, 2PYM, 2PYN, 2Q3K, 2Q54, 2Q55, 2Q5K, 2Q63, 2Q64, 2QAK, 2QCI, 2QD6, 2QD7, 2QD8, 2QHC, 2QHY, 2QHZ, 2QI0, 2QI1, 2QI3, 2QI4, 2QI5, 2QI6, 2QI7, 2R38, 2R3T, 2R3W, 2R43, 2R5P, 2R5Q, 2RKF, 2RKG, 2RSP, 2SAM, 2UPJ, 2UXZ, 2UY0, 2WHH, 2WKZ, 2WL0, 2XYE, 2XYF, 2Z4O, 2Z54, 2ZGA, 3A2O, 3AID, 3B7V, 3B80, 3BC4, 3BGB, 3BGC, 3BHE, 3BVA, 3BVB, 3BXR, 3BXS, 3CKT, 3CYW, 3CYX, 3D1X, 3D1Y, 3D1Z, 3D20, 3D3T, 3DCK, 3DCR, 3DJK, 3DK1, 3DOX, 3EBZ, 3EC0, 3ECG, 3EKP, 3EKQ, 3EKT, 3EKV, 3EKW, 3EKX, 3EKY, 3EL0, 3EL1, 3EL4, 3EL5, 3EL9, 3EM3, 3EM4, 3EM6, 3FIV, 3FSM, 3FX5, 3GGA, 3GGU, 3GGV, 3GGX, 3GI4, 3GI5, 3GI6, 3H5B, 3HAU, 3HAW, 3HBO, 3HDK, 3HLO, 3HVP, 3HZC, 3I2L, 3I6O, 3I7E, 3I8W, 3IA9, 3IAW, 3IXO, 3JVW, 3JVY, 3JW2, 3K4V, 3KA2, 3KDB, 3KDC, 3KDD, 3KF0, 3KFN, 3KFP, 3KFR, 3KFS, 3KT2, 3KT5, 3LIN, 3LIQ, 3LIT, 3LIV, 3LIX, 3LIY, 3LZS, 3LZU, 3LZV, 3M9F, 3MWS, 3MXD, 3MXE, 3N3I, 3NDT, 3NDU, 3NDW, 3NDX, 3NLS, 3NR6, 3NU3, 3NU4, 3NU5, 3NU6, 3NU9, 3NUJ, 3NUO, 3NWQ, 3NWX, 3NXE, 3NXN, 3NYG, 3O9F, 3O9G, 3O9H, 3O9I, 3OGP, 3OGQ, 3OK9, 3OQ7, 3OQA, 3OQD, 3OTS, 3OTY, 3OU1, 3OU3, 3OU4, 3OUA, 3OUB, 3OUC, 3OUD, 3OXC, 3OXV, 3OXW, 3OXX, 3OY4, 3PHV, 3PJ6, 3PSU, 3PWM, 3PWR, 3QAA, 3QBF, 3QIH, 3QN8, 3QP0, 3QPJ, 3QRM, 3QRO, 3QRS, 3R0W, 3R0Y, 3R4B, 3S43, 3S45, 3S53, 3S54, 3S56, 3S85, 3SA3, 3SA4, 3SA5, 3SA6, 3SA7, 3SA8, 3SA9, 3SAA, 3SAB, 3SAC, 3SO9, 3SPK, 3SQF, 3ST5, 3T11, 3T3C, 3TH9, 3TKG, 3TKW, 3TL9, 3TLH, 3TOF, 3TOG, 3TOH, 3TTP, 3U71, 3U7S, 3UCB, 3UF3, 3UHL, 3UPJ, 3VF5, 3VF7, 3VFA, 3VFB, 3WSJ, 3ZPS, 3ZPT, 3ZPU, 4A4Q, 4A6B, 4A6C, 4COE, 4CP7, 4CPQ, 4CPR, 4CPS, 4CPT, 4CPU, 4CPW, 4CPX, 4DFG, 4DJO, 4DJP, 4DJQ, 4DJR, 4DQB, 4DQC, 4DQE, 4DQF, 4DQG, 4DQH, 4E43, 4EJ8, 4EJD, 4EJK, 4EJL, 4EP2, 4EP3, 4EPJ, 4EQ0, 4EQJ, 4EYR, 4F73, 4F74, 4F75, 4F76, 4FAE, 4FAF, 4FE6, 4FIV, 4FL8, 4FLG, 4FM6, 4GB2, 4GYE, 4GZF, 4HDB, 4HDF, 4HDP, 4HE9, 4HEG, 4HLA, 4HVP, 4I8W, 4I8Z, 4J54, 4J55, 4J5J, 4JEC, 4K4P, 4K4Q, 4KB9, 4L1A, 4LL3, 4M8X, 4M8Y, 4MC1, 4MC2, 4MC6, 4MC9, 4NJS, 4NJT, 4NJU, 4NJV, 4NKK, 4NPT, 4NPU, 4OBD, 4OBF, 4OBG, 4OBH,

4OBJ, 4OBK, 4PHV, 4Q1W, 4Q1X, 4Q1Y, 4Q5M, 4QGI, 4QJ2, 4QJ6, 4QJ7, 4QJ8, 4QJ9, 4QJA, 4QLH, 4RVI, 4RVJ, 4RVX, 4TVG, 4TVH, 4U7Q, 4U7V, 4U8W, 4UPJ, 4YDF, 4YDG, 4YE3, 4YHQ, 4YOA, 4YOB, 4Z4X, 4Z50, 4ZIP, 4ZLS, 5AGZ, 5AH6, 5AH7, 5AH8, 5AH9, 5AHA, 5AHB, 5AHC, 5B18, 5BRY, 5BS4, 5COK, 5CON, 5COO, 5COP, 5E5J, 5E5K, 5HVP, 5IVQ, 5IVR, 5IVS, 5IVT, 5JFP, 5JFU, 5JG1, 5KAO, 5KQX, 5KQY, 5KQZ, 5KR0, 5KR1, 5KR2, 5T2E, 5T2Z, 5T84, 5T8H, 5TYR, 5TYS, 5UFZ, 5ULT, 5UOV, 5UPJ, 5UPZ, 5V4Y, 5VCK, 5VEA, 5VJ3, 5W5W, 5WLO, 5YOJ, 5YOK, 5YRS, 6B36, 6B38, 6B3C, 6B3F, 6B3G, 6B3H, 6B4N, 6BRA, 6BZ2, 6C8X, 6C8Y, 6CDJ, 6CDL, 6DGX, 6DGY, 6DGZ, 6DH0, 6DH1, 6DH2, 6DH3, 6DH4, 6DH5, 6DH6, 6DH7, 6DH8, 6DIF, 6DIL, 6DJ1, 6DJ2, 6DJ5, 6DJ7, 6DV0, 6DV4, 6E7J, 6E9A, 6UPJ, 7HVP, 7UPJ, 8HVP, 9HVP

S1.2. List of entries containing water molecules within the active site that were checked manually:

1AID, 1AXA, 1AZ5, 1BDL, 1G6L, 1HPX, 1HTE, 1LV1, 1MVP, 1N49, 1RPI, 1SIP, 1TCW, 1TW7, 1XL2, 2G69, 2HB4, 2PC0, 2RSP, 2WHH, 3BC4, 3DOX, 3GGA, 3I8W, 3IXO, 3KFN, 3KT2, 3KT5, 3MIM, 3MIM, 3NR6, 3OQ7, 3OQD, 3OTS, 3OU1, 3OU1, 3OU4, 3OU4, 3OUD, 3OUD, 3QN8, 3QP0, 3QRS, 3R0Y, 3SQF, 3T11, 3TKG, 3UF3, 4EJ8, 4EJL, 4EYR, 4GYE, 4L1A, 4L1A, 4NKK, 4NPT, 4NPU, 4QLH, 4YOB, 4Z4X, 4Z50, 5B18, 5KR1, 5T2E, 5T84

Fig. S1

Summary of modeled residues in structures 1M, 1MI and 3MI, by protein chains. Wild type sequence is denoted as wt; 3sqf stands for the previously reported monomeric structure of M-PMV PR (Gilski *et al*., 2011). Lower-case symbols indicate residues for which only main chain was modeled (truncated at Cα). Hyphens are used for residues that were not modeled.

```
       1       10        20        30        40        50        60        70        80        90       100       110 114
       |        |         |         |         |         |         |         |         |         |         |         |   |
   wt: WVQPITCQKPSLTLWLDDKMFTGLIDTGADVTIIKLEDWPPNWPITDTLTNLRGIGQSNNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSPNDIVTA

 1M_A: WVQPITCQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDT----------nPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSP------
 1M_B: WVQPITCQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNLRGIGQSNNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSP------

1MI_A: WVQPITCQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNLr-----NNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSP------
1MI_B: WVQPITCQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNLr---QSNNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSP------

3MI_A: WVQPITAQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNlr-----NNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMASP------
3MI_B: WVQPITAQKPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNLrG---SNNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMASPn-----

 3sqf: --------KPSLTLWLDDKMFTGLINTGADVTIIKLEDWPPNWPITDTLTNlRGIGQSNNPKQSSKYLTWRDKENNSGLIKPFVIPNLPVNLWGRDLLSQMKI-----------
```

Table S1

List of hydrogen bonds at the catalytic asparagine (N26) residues in 1MI, 3MI and 1M structures. In the 1M structure there are four interactions with water molecules instead of contacts with the inhibitor molecule.

atom /residue (chain)	atom or molecule or group /residue (chain)	label in text	distance [Å]		
			1MI	3MI	1M
O /Asn26 (A)	N /Ala29 (A)	-	3.18	3.08	3.15
N /Asn26 (A)	O /Trp93 (A)	-	2.78	2.82	2.76
Oδ1 /Asn26 (A)	H_2O /HOH201/202/207[a]	Wat1	2.28	2.50	2.58
Oδ1 /Asn26 (A)	H_2O /HOH216	-	-	-	2.70
Nδ2 /Asn26 (A)	H_2O /HOH223/218/210[a]	Wat2A	2.94	3.03	3.15
Nδ2 /Asn26 (A)	H_2O /HOH227	-	-	-	2.97
O /Asn26 (B)	N /Ala29 (B)	-	3.26	3.15	3.17
N /Asn26 (B)	O /Trp93 (B)	-	2.78	2.85	2.79
Oδ1 /Asn26 (B)	H_2O /HOH201/202/207[a]	Wat1	2.70	2.41	2.64
Oδ1 /Asn26 (B)	OH /PSA (I)	-	2.53	2.80	-
Oδ1 /Asn26 (B)	H_2O /HOH231	-	-	-	2.77
Nδ2 /Asn26 (B)	H_2O /HOH235/232/221[a]	Wat2B	2.92	3.06	3.01
Nδ2 /Asn26 (B)	O /PSA (I)	-	3.25	3.27	-
Nδ2 /Asn26 (B)	H_2O /HOH247	-	-	-	3.09

[a]Numbers correspond to 1MI/3MI/1M, respectively.