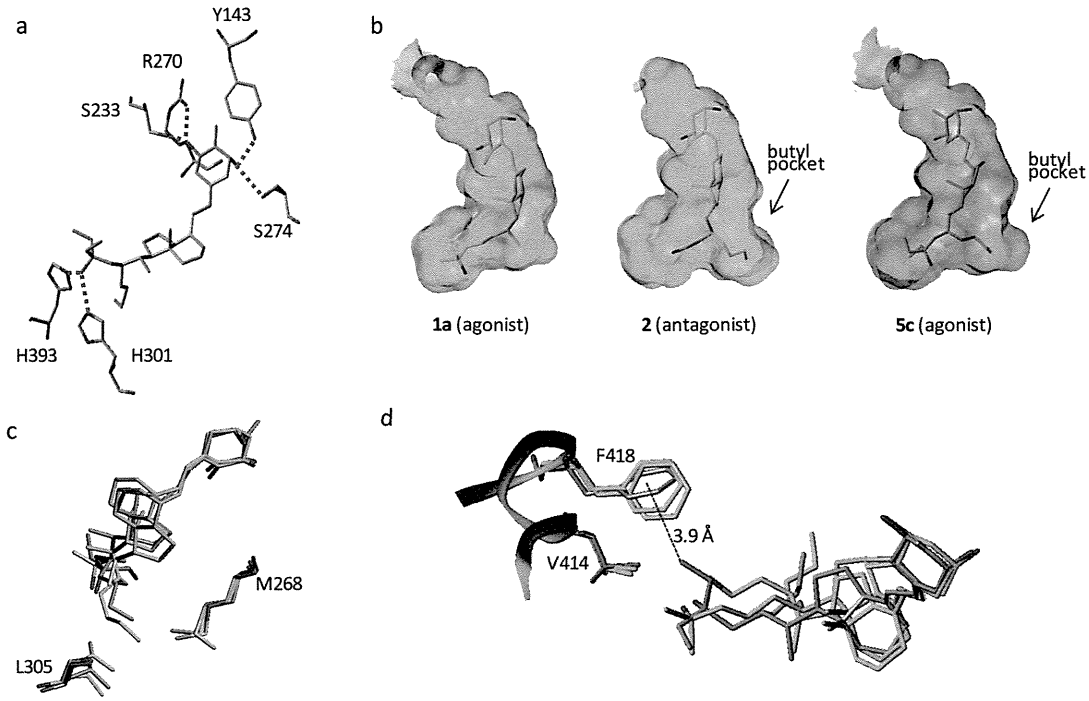


Figure 3. Crystal structures of VDR-LBD complexed with **5c**: (a) hydrogen bonds between VDR-LBD and **5c** (pink); (b) Conolly channel surface of the ligand-binding pocket of the VDR-LBD/**1a** complex (left, green), VDR-LBD/**2** complex (middle, cyan), and VDR-LBD/**5c** complex (right, pink); (c, d) superposition of VDR-LBD complexed with **1a** (green), **2** (cyan), or **5c** (pink). (a) The 1α-, 3β-, and 25-hydroxyl groups of **5c** form pincer-type hydrogen bonds with Ser233 and Arg270, Tyr143 and Ser274, and His301 and His393, respectively. (b) Formation of the butyl pocket to accommodate the 22-butyl group is clearly observed in the complex with **2** and **5c**. (c) The terminal *S*-methyl group of Met268 in the VDR/**2** complex (cyan) rotated in about 120° compared with that in VDR/**1a** complex (green), but that in VDR/**5c** (pink) did not rotate. Leu305 in VDR/**5c** (pink) adopted intermediate conformation between conformations of the complexes with **1a** and **2**. (d) **5c** (pink) and Phe418 at helix 12 make C–H···π interactions. The distance between the terminal carbon of one of 24-ethyl groups in **5c** and the center of the π-system (d_{C-X}) is 3.9 Å.

Table 2. Summary of Data Collection Statistics and Refinement

	ligand			
parameter	**3a**	**4a**	**5a**	**5c**
X-ray source	KEK-PF BL-6A	KEK-PFAR NW-12A	KEK-PFAR NW-12A	KEK-PFAR NW-12A
wavelength (Å)	0.97800	1.00000	1.00000	1.00000
space group	*C*2	*C*2	*C*2	*C*2
unit cell dimensions				
bond (Å)	a = 152.99, b = 43.87, c = 42.43	a = 154.34, b = 42.29, c = 42.21	a = 154.14, b = 42.77, c = 42.12	a = 154.21, b = 43.21, c = 42.38
angle (deg)	α = 90.00, β = 95.94, γ = 90.00	α = 90.00, β = 95.89, γ = 90.00	α = 90.00, β = 95.42, γ = 90.00	α = 90.00, β = 95.49, γ = 90.00
resolution range (Å)[a]	38.04–2.40 (2.49–2.40)	76.76–2.00 (2.11–2.00)	76.73–1.90 (2.00–1.90)	76.75–2.40 (2.53–2.40)
total no. of reflections	41156	56092	75735	34783
no. of unique reflections	11168	17728	20511	10065
% completeness[a]	99.5 (99.6)	96.4 (96.3)	94.6 (95.1)	91.9 (91.4)
R_{merge}[a,b]	0.045 (0.299)	0.074 (0.308)	0.047 (0.343)	0.081 (0.336)
Refinement Statistics				
resolution range (Å)[a]	38.04–2.40 (2.49–2.40)	26.98–2.00 (2.11–2.00)	29.74–1.90 (2.00–1.90)	33.01–2.40 (2.53–2.40)
R factor (R_{free}/R_{work})[a,c]	0.273/0.225	0.279/0.219	0.248/0.209	0.256/0.218

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{merge} = \sum|(I_{hkl} - \langle I_{hkl}\rangle)|/(\sum I_{hkl})$, where $\langle I_{hkl}\rangle$ is the mean intensity of all reflections equivalent to reflection *hkl*. [c]R_{work} (R_{free}) = $\sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, where 5% of randomly selected data were used for R_{free}.

than natural hormone **1a**. All three analogues **5a–c** showed potent transcriptional activity (Figure 1). 22-Ethyl analogue **5b** acts as a full agonist, whereas **3b** and **4b** are partial agonists. More surprisingly, 22-butyl analogue **5c** completely restored agonistic activity, although **2** and **3c** are antagonists and **4c** is a quite weak agonist.

This drastic change in transcriptional activity was explained by comparison of the crystal structures. As expected, **5c** induced the formation of a butyl pocket similar to **2** (Figure 3b), but in contrast to **2**, **5c** interacts intimately with helix 12 to form a stable complex with VDR. As shown in Table 3, the number of amino acid residues within 4.5-Å of **5c** is 27, whereas for **2** there are only 22 residues within that distance. Compound **5c** therefore forms sufficient interactions with VDR, and especially with helix 12, to form a stable complex, but **2** does not. This is a primary reason why **2** is an antagonist whereas **5c** is an agonist.

dx.doi.org/10.1021/jm300230a | *J. Med. Chem.* 2012, 55, 4373–4381

Table 3. Residues within 4.5 Å from ligand

compd	residues within 4.5 Å from ligand
1a (28 residues)[a]	Tyr143, Tyr147, Phe150, Leu223, Leu226, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, Ala299, His301, Leu305, Leu309, His393, Tyr397, Leu400, Leu410, Val414, Phe418
2 (22 residues)[a]	Tyr143, Tyr147, Phe150, Leu226, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, His301, Leu305, Leu309, His393, Leu398
3a (20 residues)[a]	Tyr143, Tyr147, Phe150, Leu226, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, His301, Leu309, His393
4a (24 residues)[a]	Tyr143, Tyr147, Phe150, Leu223, Leu226, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, Ala299, His301, Leu305, His393, Tyr397, Phe418
5a (27 residues)[a]	Tyr143, Tyr147, Phe150, Leu223, Leu226, Ala227, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, Ala299, His301, Leu305, Leu309, His393, Tyr397, Val414, Phe418
5c (27 residues)[a]	Tyr143, Tyr147, Phe150, Leu223, Leu226, Leu229, Val230, Ser233, Ile264, Ile267, Met268, Arg270, Ser271, Ser274, Trp282, Cys284, Tyr291, Val296, Ala299, His301, Leu305, Leu309, His393, Tyr397, Leu410, Val414, Phe418

[a]Total number of residues within 4.5 Å from ligand.

Interestingly, the butyl pocket of the VDR-LBD/**5c** complex is slightly smaller than that of VDR-LBD/**2**; this is mainly due to the conformation of Met268 and Leu305 (Figure 3c). The terminal *S*-methyl group of Met268 in the VDR-LBD/**2** complex rotated in about 120° compared with that in the VDR-LBD/**1a** complex, but that in VDR-LBD/**5c** did not rotate. In addition, Leu305 adopted a conformation intermediate between conformations of the complexes with **1a** and **2**. Furthermore, we found that **5c** and Phe418 in helix 12 make C−H···π interactions (Figure 3d). The distance between the terminal carbon of one of the 24-ethyl groups in **5c** and the center of the π-system (d_{C-X}) is 3.9 Å, indicating that this C−H···π interaction would contribute significantly to the overall stability of the protein.[32] The results agree well with our hypothesis that increased hydrophobic interactions overcome the strain caused by the formation of an extra cavity such as a butyl pocket.

In 22-H analogues, the EC_{50} of transcriptional activity decreased in the order **3a** > **4a** > **5a**. These results are explained by the crystal structures. Specific differences between the three complexes are observed in the interactions between the ligand side chain and the C-terminus of VDR-LBD. As shown in Figure 2b, among the C-terminal residues within 4.5 Å of hormone **1a**, five residues (Tyr397, Leu400, Leu410, Val414, and Phe418), three residues (Leu400, Leu410, and Val414), and two residues (Leu400 and Leu410) are not within 4.5 Å of **3a**, **4a**, and **5a**, respectively, indicating that interactions with helices 11 and 12 increase in the order **3a** < **4a** < **5a**. These results indicate that intimate interactions with VDR reinforce transcriptional activity.

Structural modification of the secondary structure of a protein by the binding of a small molecule such as a ligand, a substrate, or an inhibitor is difficult, but modification of a loop region would be possible. Our synthetic compounds with a 22-butyl substituent, **2** and **5c**, push out the loop region, as observed in the crystal structures. This study suggests that modification of the three-dimensional structure of a protein at a flexible loop region by a ligand binding is a novel strategy for the discovery of new drugs that have the desired selectivity.

CONCLUSIONS

We designed and synthesized analogues **5a**−**c** bearing a diethyl substituent at C(24) and evaluated their biological activities. The crystal structures of VDR-LBD complexed with **3a**, **4a**, **5a**, or **5c** were solved and the results confirmed the following. First, analogues with a 22*S*-butyl group, such as **2** and **5c**, induce butyl pocket formation to accommodate the butyl group. Second, among ligands that induce butyl pocket formation (**2**, **3c**, and **5c**), a ligand acts as an antagonist when it does not interact with the C-terminus of VDR, whereas a ligand acts as an agonist when it interacts intimately with the C-terminus of VDR. Third, ligands that do not induce an extra cavity like a butyl pocket (**3a**, **4a**, and **5a**) increase agonistic activity when the ligand increases interactions with the C-terminus of VDR. Thus, butyl pocket formation in VDR strongly affects the agonistic or antagonistic behavior of ligands. These results indicate that structural modification of a target protein at a flexible region such as loop region may be an important strategy for the discovery of new drugs for the treatment of various diseases.

EXPERIMENTAL SECTION

All reagents were purchased from commercial sources. Unless otherwise stated, NMR spectra were recorded at 300 MHz for ^{1}H NMR and 75 MHz for ^{13}C NMR in $CDCl_3$ solution with TMS as an internal standard, and the chemical shifts are given in δ values. High and low resolution mass spectra were obtained with JEOL JMS D-300, JEOL AccuTOF LC-plus JMS-T100LP, and JEOL JMS-HX110A spectrometers. Relative intensities are given in parentheses in low mass. IR spectra were recorded on a Shimadzu FTIR-8400S spectrophotometer, and data are given in cm^{-1}. UV spectra were recorded on a Beckman DU7500 spectrophotometer. All air and moisture sensitive reactions were carried out under argon or nitrogen atmosphere. Purity was determined by HPLC [PEGASIL silica SP100, 4.6 mm × 150 mm, hexane/$CHCl_3$/MeOH (100:25:8), flow rate 1.0 mL/min] and was >95% for all compounds tested.

6-[(1*R*,3*R*,7*E*,17β)-1,3-Bis{[*tert*-butyl(dimethyl)silyl]oxy}-2-methylidene-9,10-secoestra-5,7-dien-17-yl]-3-ethylheptan-3-ol (7). To a solution of ester **6** (54.6 mg, 0.0867 mmol) in THF (1 mL) at −78 °C was added EtLi (870 μL of 0.5 M benzene/cyclohexane (90/10) solution, 0.435 mmol), and the mixture was stirred for 1 h. The reaction was quenched with saturated NH_4Cl, and the mixture was extracted with AcOEt. The organic layer was washed with brine, dried over $MgSO_4$, and evaporated. The residue was chromatographed on silica gel (AcOEt/hexane = 5/95) to afford **7** (30.4 mg, 0.0462 mmol, 53%). ^{1}H NMR δ 0.02, 0.05, 0.06, 0.08 (each 3 H, s, SiMe), 0.55 (3 H, s, H-18), 0.80−0.96 (6H, m, $CH_2\underline{CH_3}$ × 2), 0.86, 0.89 (each 9 H, s, *t*-Bu), 0.94 (3 H, d, J = 6.1 Hz, H-21), 2.80 (1 H, m, H-9), 4.42 (2 H, m, H-1, 3), 4.92, 4.96 (each 1 H, s, −C=CH_2), 5.83 (1 H, d, J = 11.1 Hz, H-7), 6.21 (1 H, d, J = 11.1 Hz, H-6). ^{13}C NMR δ −5.1, −4.9, −4.8 (2 C), 7.7, 7.9, 12.0, 18.1, 18.2, 18.9, 22.2, 23.4, 25.7 (3 C), 25.8 (3 C), 27.7, 28.7, 29.2, 30.9, 31.1, 34.2, 36.5, 38.6, 40.6, 45.6, 47.6, 56.2, 56.3, 71.7, 72.5, 74.7, 106.2, 116.1, 122.4, 132.7, 141.1, 152.9. MS (EI) m/z (%): 658 (M^+, 2), 526 (26), 508 (8), 366 (20), 234 (14), 147 (15), 73 (100). HRMS (EI) calcd for $C_{40}H_{74}O_3Si_2$ 658.5177, found 658.5196. IR (neat) 2954, 2856, 1461, 1251, 1101, 1072, 835 cm^{-1}. UV (hexane) λ_{max} 246, 254, 264 nm.

dx.doi.org/10.1021/jm300230a | *J. Med. Chem.* 2012, 55, 4373−4381

2-Methylidene-26,27-dimethyl-19,24-dinor-1α,25-dihydroxyvitamin D_3 (5a). A solution of 7 (29.4 mg, 0.0447 mmol) and camphor sulfonic acid (39.0 mg, 0.168 mmol) in MeOH (1 mL) was stirred at room temperature for 0.5 h. Aqueous $NaHCO_3$ was added, and the mixture was extracted with AcOEt. The organic layer was washed with brine, dried over $MgSO_4$, and evaporated. The residue was chromatographed on silica gel (AcOEt/hexane = 1/1) to afford **5a** (17.0 mg, 89%). ^{1}H NMR δ 0.55 (3 H, s, H-18), 0.86 (6H, td, J = 2.0, 7.4 Hz, $CH_2CH_3 \times 2$), 0.95 (3 H, d, J = 6.0 Hz, H-21), 4.47 (2 H, m, H-1, 3), 5.09, 5.11 (each 1 H, s, $-C{=}CH_2$), 5.88 (1 H, d, J = 11.2 Hz, H-7), 6.36 (1 H, d, J = 11.2 Hz, H-6). ^{13}C NMR δ 7.7, 7.8, 12.1, 18.9, 22.3, 23.5, 27.6, 29.0, 29.1, 30.9, 31.1, 34.2, 36.4, 38.1, 40.4, 45.8 (2 C), 56.2, 56.3, 70.6, 71.8, 74.7, 107.7, 115.3, 124.2, 130.5, 143.4, 152.0. MS (EI) m/z (%): 430 (M^+, 40), 383 (36), 365 (25), 285 (32), 187 (20), 175 (30), 161 (46), 147 (57), 135 (75), 119 (55), 69 (100). HRMS (EI) calcd for $C_{28}H_{46}O_3$ 430.3447, found 430.3462. IR (neat) 3377, 2941, 2875, 1658, 1612, 1452, 1145, 1074, 1045, 977, 912, 756 cm^{-1}. UV (EtOH) λ_{max} 245, 254, 263 nm.

(5*S*,6*R*)-6-{1-[(1*R*,3*R*,7*E*,17β)-1,3-Bis{[*tert*-butyl(dimethyl)silyl]oxy}-2-methylidene-9,10-secoestra-5,7-dien-17-yl]ethyl}-3,5-diethylheptan-3-ol (9). To a solution of ester **8** (118.9 mg, 0.18 mmol) in THF (3 mL) at room temperature was added EtMgBr (530 μL of 1.0 M THF solution, 0.53 mmol), and the mixture was stirred for 3 h. The reaction was quenched with saturated NH_4Cl, and the mixture was extracted with AcOEt. The organic layer was washed with brine, dried over $MgSO_4$, and evaporated. The residue was chromatographed on silica gel (AcOEt/hexane = 3/97) to afford **9** (104.0 mg, 0.15 mmol) in 86% yield. ^{1}H NMR δ 0.03, 0.04, 0.06, 0.08 (each 3 H, s, SiMe), 0.53 (3 H, s, H-18), 0.83 (3 H, J = 6.3 Hz, d, H-21), 0.84–0.94 (9H, m, $CH_2CH_3 \times 3$), 0.86, 0.89 (each 9 H, s, t-Bu), 2.82 (1 H, m, H-9), 4.42 (2 H, m, H-1,3), 4.92, 4.96 (each 1 H, s, $-C{=}CH_2$), 5.84 (1 H, d, J = 11.1 Hz, H-7), 6.21 (1 H, d, J = 11.1 Hz, H-6). ^{13}C NMR δ −4.8, −4.69, −4.66, −4.63, 8.2, 8.3, 12.2, 13.5, 13.8, 18.4, 18.5, 22.3, 22.4, 23.7, 26.0 (3 C), 26.1 (3 C), 27.7, 29.0, 31.0, 32.0, 37.4, 38.9, 39.9, 40.3, 40.9, 45.9, 47.8, 54.8, 56.5, 71.9, 72.6, 75.8, 106.5, 116.3, 122.6, 133.0, 141.5, 153.2. MS (APCI) m/z (%):711 (M + Na^+, 1), 429 (45), 234 (35), 220 (25), 202 (100), 167 (95). HRMS (APCI) calcd for $C_{42}H_{79}O_3Si_2Na$ 710.5465, found 710.5462. IR (neat) 3489, 2955, 2930, 2883, 2856, 1730, 1655, 1618, 1472, 1462, 1256, 1101, 935, 897, 835, 775, 667 cm^{-1}. UV (hexane) λ_{max} 246, 255, 265 nm.

22*S*-Ethyl-2-methylidene-26,27-dimethyl-19,24-dinor-1α,25-dihydroxyvitamin D_3 (5b). In a manner similar to that for the synthesis of **5a** from **7**, target compound **5b** (55.0 mg, 0.120 mmol) was obtained from **9** (83.8 mg, 0.122 mmol) in 98% yield. ^{1}H NMR δ 0.53 (3 H, s, H-18), 0.81 (3 H, J = 6.3 Hz, d, H-21), 0.81–0.93 (9H, m, $CH_2CH_3 \times 3$), 2.81 (1 H, m, H-9), 4.44 (2 H, m, H-1,3), 5.06, 5.08 (each 1 H, s, $-C{=}CH_2$), 5.87 (1 H, d, J = 11.1 Hz, H-7), 6.33 (1 H, d, J = 11.1 Hz, H-6). ^{13}C NMR δ 7.9, 8.0, 12.1, 13.3, 13.6, 22.2 (2 C), 23.6, 27.4, 29.0, 30.8, 31.7, 37.2, 38.1, 39.5, 40.0, 40.6, 45.8 (2 C), 54.6, 56.3, 70.6, 71.8, 75.7, 107.7, 115.3, 124.1, 130.6, 143.3, 152.0. MS (APCI) m/z (%): 481 (M + Na^+, 100), 437 (52), 429 (23). HRMS (APCI) calcd for $C_{30}H_{50}O_3Na$ 481.3658 found 481.3664. IR (neat) 3390, 2961, 2939, 2874, 1715, 1651, 1614, 1462, 1456, 1381, 1074, 1045, 978, 912, 756, 667 cm^{-1}. UV (EtOH) λ_{max} 246, 254, 264 nm.

(5*S*)-4-{1-[(1*R*,3*R*,7*E*,17β)-1,3-Bis{[*tert*-butyl(dimethyl)silyl]oxy}-2-methylidene-9,10-secoestra-5,7-dien-17-yl]ethyl}-3-ethylnonan-3-ol (11). To a solution of EtMgI in Et_2O (2 mL) prepared from EtI (91 μL, 1.13 mmol) and magnesium turnings (27.3 mg, 1.13 mmol) was added ester **10** (37.8 mg, 0.054 mmol) in Et_2O (1 mL) at room temperature, and the mixture was stirred for 3.5 h. The reaction was quenched with 1 N HCl, and the mixture was extracted with AcOEt. The organic layer was washed with saturated $NaHCO_3$ and brine, dried over $MgSO_4$, and evaporated. The residue was chromatographed on silica gel (AcOEt/hexane = 5/95) to afford **11** (26.7 mg, 0.0374 mmol) in 69% yield. ^{1}H NMR δ 0.03, 0.04, 0.06, 0.08 (each 3 H, s, SiMe), 0.53 (3 H, s, H-18), 0.83 (3 H, d, J = 6.6 Hz, H-21), 0.80–0.94 (9H, m, $CH_2CH_3 \times 3$), 0.86, 0.89 (each 9 H, s, t-Bu), 2.82 (1 H, m, H-9), 4.42 (2 H, m, H-1, 3), 4.92, 4.97 (each 1 H, s, $-C{=}CH_2$), 5.83 (1 H, d, J = 11.1 Hz, H-7), 6.21 (1 H, d, J = 11.1 Hz, H-6); ^{13}C NMR δ −5.1, −4.9, −4.8 (2 C), 7.9, 8.1, 12.0, 13.5, 14.2, 18.1, 18.2, 22.1, 23.3, 23.5, 25.7 (3 C), 25.8 (3 C), 27.5, 28.8, 29.4, 30.7, 30.9, 31.7, 35.1, 38.6, 39.5, 40.6, 40.7, 45.6, 47.5, 54.5, 56.2, 71.7, 72.4, 75.6, 106.2, 116.1, 122.4, 132.7, 141.2, 152.9. MS (EI) m/z (%): 714 (M^+, 2), 582 (36), 564 (10), 366 (35), 351 (10), 257 (11), 234 (12), 197 (10), 147 (15), 73 (100). HRMS (EI) calcd for $C_{44}H_{82}O_3Si_2$ 714.5802, found 714.5804. IR (neat) 3487, 2954, 2927, 2856, 1726, 1658, 1620, 1461, 1251, 1101, 1072, 835, 775 cm^{-1}. UV (hexane) λ_{max} 246, 255, 264 nm.

22*S*-Butyl-2-methylidene-26,27-dimethyl-19,24-dinor-1α,25-dihydroxyvitamin D_3 (5c). In a manner similar to that for the synthesis of **5a** from **7**, target compound **5c** (16.6 mg, 0.0342 mmol) was obtained from **11** (26.7 mg, 0.0374 mmol) in 91% yield. ^{1}H NMR δ 0.54 (3 H, s, H-18), 0.82 (3 H, d, J = 6.3 Hz, H-21), 0.86 (6 H, t, J = 7.4 Hz, 2 × CH_3 of Et), 0.90 (3 H, t, J = 6.8 Hz, CH_3 of Bu), 4.47 (2 H, m, H-1, 3), 5.09, 5.11 (each 1 H, s, $-C{=}CH_2$), 5.88 (1 H, d, J = 11.2 Hz, H-7), 6.35 (1 H, d, J = 11.2 Hz, H-6). ^{13}C NMR δ 7.9, 8.0, 12.1, 13.5, 14.2, 22.1, 23.3, 23.5, 27.4, 29.0, 29.4, 30.8 (2 C), 31.6, 35.1, 38.1, 39.3, 40.5 (2 C), 45.8 (2 C), 54.5, 56.3, 70.6, 71.8, 75.6, 107.7, 115.2, 124.2, 130.4, 143.4, 152.0. MS (EI) m/z (%): 486 (M^+, 31), 439 (15), 421 (10), 384 (22), 315 (18), 285 (20), 269 (10), 251 (12), 231 (10), 173 (20), 153 (60), 135 (95), 119 (40), 105 (70), 91 (70), 81 (75), 69 (100). HRMS (EI) calcd for $C_{32}H_{54}O_3$ 486.4072, found 486.4073. IR (neat) 3388, 2933, 2871, 1712, 1658, 1612, 1456, 1380, 1076, 910, 754 cm^{-1}. UV (EtOH) λ_{max} 246, 254, 263 nm.

Competitive Binding Assay, Human VDR. The human recombinant VDR ligand-binding domain (LBD) was expressed as an N-terminal GST-tagged protein in *E. coli* BL21 (DE3) pLys S (Promega).[31] The cells were lysed by sonication. The supernatants were diluted approximately 500 times in 50 mM Tris buffer (100 mM KCl, 5 mM DTT, 0.5% CHAPS, pH 7.5) containing bovine serum albumin (100 μg/mL). Binding to GST-hVDR-LBD was evaluated according to the procedure reported.[33] The receptor solution (570 μL) in an assay tube was incubated with [^{3}H]-1α,25-dihydroxyvitamin D_3 (specific activity, 5.85 TBq/mmol, ~2000 cpm) together with graded amounts of each vitamin D analogue (0.001–100 nM) or vehicle for 16 h at 4 °C. The bound and free [^{3}H]-1α,25-dihydroxyvitamin D_3 molecules were separated by treating with dextran-coated charcoal for 30 min at 4 °C. The assay tubes were centrifuged at 1000*g* for 10 min. The radioactivity of the supernatant was counted. Nonspecific binding was subtracted. These experiments were done in duplicate.

Transfection and Transactivation Assay. COS-7 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal bovine serum (FBS). Cells were seeded on 24-well plates at a density of 2×10^4 per well. After 24 h, the cells were transfected with a reporter plasmid containing three copies of the mouse osteopontin VDRE (5′-GGTTCAcgaGGTTCA, SPPx3-TK-Luc), a wild-type or mutant hVDR expression plasmid (pCMX-hVDR), and the internal control plasmid containing sea pansy luciferase expression constructs (pRL-CMV) by the lipofection method as described previously.[12] After an 8 h incubation, the cells were treated with either the ligand or ethanol vehicle and cultured for 16 h. Cells in each well were harvested with a cell lysis buffer, and the luciferase activity was measured with a luciferase assay kit (Promega, WI, U.S.). Transactivation measured by the luciferase activity was normalized with the internal control. All experiments were done in triplicate.

Protein Expression. The rat VDR-LBD (residues 116–423, Δ165–211) was subcloned as an N-terminal His_6-tagged fusion protein into the pET-28a vector. *E. coli* Rosetta 2 (DE3) was freshly transformed with the plasmid and grown in four flasks containing 0.75 L of 2xTY medium with kanamycin, 34 μg/mL, and chloramphenicol, 50 μg/mL, at 37 °C until an OD of 0.8 was obtained. The cultures were then induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside and further incubated at 23 °C for 18 h. Cells were harvested and resuspended in 50 mL of lysis buffer (50 mM Na/K phosphate, pH 8.0, 10 mM imdazole, 500 mM NaCl, 5% glycerol, 1% Tween 20, 1 mM TCEP, 1 mM PMSF). Cells were lysed by sonication, and the

soluble fraction was isolated by centrifugation (18000*g* for 20 min). The supernatant was applied to Ni-NTA agarose (Qiagen, Santa Clarita, CA), and the resin was thoroughly washed in wash buffer (50 mM Na/K phosphate, pH 8.0, 20 mM imdazole, 500 mM NaCl, 5% glycerol, 1% Tween 20, 1 mM TCEP, 1 mM PMSF). The rat VDR-LBD was eluted with elution buffer (150 mM imidazole, 50 mM Na/K phosphate, pH 8.0, 500 mM NaCl, 5% glycerol, 1% Tween 20, 1 mM TCEP, 1 mM PMSF). The protein was dialyzed overnight against dialysis buffer A (20 mM Na/K phosphate, pH 7.0, 5% glycerol, 1 mM EDTA, 0.5 mM DTT) and then loaded onto a HiTrap SP HP (1 mL) column (GE Healthcare) equilibrated with buffer A. The elution was performed by NaCl gradient buffer from 0 to 1.0 M. His-tag of the protein in elution mixture (17 mL) was cleaved by addition of 70 units of thrombin and subsequent incubation at 23 °C for 18 h. Then NaCl (1.96 g) was added to the digested mixture (17 mL), and the resulting mixture was passed through a HiTrap benzamidine FF (1 mL) columun (GE healthcare). The flow-through was further purified by Superdex S75 gel filtration (25 mL) column (GE Healthcare) with a buffer (100 mM NaCl, Tris-HCl, pH 7.0). Purified rat VDR-LBD was concentrated in buffer B (10 mM Tris-HCl, pH 7.0, 2 mM NaN_3, 10 mM DTT, 0.5 mM PMSF) to 7.5 mg/mL, which was estimated by UV absorbance at 280 nm.

X-ray Crystallographic Analysis. A ligand (~10 equiv) was added to the rVDR-LBD, and then coactivator peptide (H_2N-KNHPMLMNLLKDN-$CONH_2$) derived from DRIP205 in buffer C (25 mM Tris-HCl, pH 8.0; 50 mM NaCl; 10 mM DTT; 2 mM NaN_3) was added. The mixture of VDR-LBD/ligand/peptide was allowed to crystallize by the vapor diffusion method using a series of precipitant solutions containing 0.1 M MOPS-Na (pH 7.0) or 0.1 M MES-Na (pH 7.0), 0.05–0.4 M sodium formate, 12–22% (w/v) PEG4000, and 5% ethylene glycol. Droplets for crystallization were prepared by mixing 1 μL of complex solution and 1 μL of precipitant solution, and droplets were equilibrated against 300 μL of precipitant solution at 20 °C. Prior to diffraction data collection, crystals were soaked in a cryoprotectant solution containing 0.1 M MOPS-Na, pH 7.0, or 0.1 M MES-Na, pH 7.0, 0.05–0.3 M sodium formate, 12–22% (w/v) PEG4000, and 5% ethylene glycol. Diffraction data sets were collected at 100 K in a stream of nitrogen gas at beamline BL-6A of KEK-PF or NW-12A of KEK-PFAR (Tsukuba, Japan). Reflections were recorded with an oscillation range per image of 1.0°. Difflections data were indexed, integrated, and scaled using the program iMOSFLM[34,35] The structures of ternary complex were solved by molecular replacement with the software Phaser[36] in the CCP4 program[37] using a rat VDR-LBD coordinates (PDB code 2ZLC), and finalized sets of atomic coordinates were obtained after iterative rounds of model modification with the program Coot[38] and refinement with refmac5.[39−43] The coordinate data for the structures were deposited in Protein Data Bank with accession numbers 3VRT (VDR-LBD/**3a** complex), 3VRU (VDR-LBD/**4a** complex), 3VRV (VDR-LBD/**5a** complex), and 3VRW (VDR-LBD/**5c** complex).

Graphical Manipulations and Ligand Docking. Graphical manipulations were performed using SYBYL 8.0 (Tripos, St. Louis). The atomic coordinates of the crystal structure of rVDR-LBD complexed with **1a** were retrieved from Protein Data Bank (PDB) (entry 2ZLC).[24]

AUTHOR INFORMATION

Corresponding Author

*Phone: +81 42 721 1580. Fax: +81 42 721 1580. E-mail: yamamoto@ac.shoyaku.ac.jp.

Author Contributions

⊥These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Part of this work was supported by a Grant-in-Aid for Scientific Research (Grant 23590135) and a Grant-in-Aid for High Technology Research Center Project (Grant 19-8) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The synchrotron-radiation experiment was performed at the Photon Factory (Proposal Grants 2008G169 and 2009G647). We acknowledge the help provided by the beamline scientists at the Photon Factory.

ABBREVIATIONS USED

VDR, vitamin D receptor; LBP, ligand binding pocket; RXR, retinoid X receptor; LBD, ligand binding domain

REFERENCES

(1) Feldman, D., Pike, J. W., Glorieux, F. H., Eds. *Vitamin D*, 2nd ed.; Elsevier Academic Press: Amsterdam, 2005.

(2) Jurutka, P. W.; Whitfield, G. K.; Hsieh, J. C.; Thompson, P. D.; Haussler, C. A.; Haussler, M. R. Molecular nature of the vitamin D receptor and its role in regulation of gene expression. *Rev. Endocr. Metab. Disord.* **2001**, *2*, 203–216.

(3) Yamada, S.; Shimizu, M.; Yamamoto, K. Vitamin D Receptor. In *Vitamin D and Rickets*; Hochberg, Z., Ed.; Endocrine Develpment, Vol. *6*; Karger: Basel, Switzerland, 2003; pp 50–68.

(4) Binderup, L.; Binderup, E.; Godtfredsen, W. O. Development of New Vitamin D Analogs. In *Vitamin D*, 2nd ed.; Feldman, D., Pike, J. W., Glorieux, F. H., Eds.; Elsevier Academic Press: Amsterdam, 2005; pp 1027–1043.

(5) Kubodera, N. Pharmaceutical studies on vitamin D derivatives and practical syntheses of six commercially available vitamin D derivatives that contribute to current clinical practice. *Heterocycles* **2010**, *80*, 83–98.

(6) Bury, Y.; Steinmeyer, A.; Carlberg, C. Structure activity relationship of carboxylic ester antagonists of the vitamin D3 receptor. *Mol. Pharmacol.* **2000**, *58*, 1067–1074.

(7) Igarashi, M.; Yoshimoto, N.; Yamamoto, K.; Shimizu, M.; Makishima, M.; DeLuca, H. F.; Yamada, S. Identification of a highly potent vitamin D receptor antagonist: (2*5S*)-26-adamantyl-25-hydroxy-2-methylene-22,23-didehydro-19,27-dinor-20-epi-vitamin D_3 (ADMI3). *Arch. Biochem. Biophys.* **2007**, *460*, 240–253.

(8) Miura, D.; Manabe, K.; Ozono, K.; Saito, M.; Gao, Q.; Norman, A. W.; Ishizuka, S. Antagonistic action of novel 1α,25-dihydroxyvitamin D_3-26,23-lactone analogs on differentiation of human leukemia cells (HL-60) induced by 1α,25-dihydroxyvitamin D_3. *J. Biol. Chem.* **1999**, *274*, 16392–16399.

(9) (a) Ishizuka, S.; Kurihara, N.; Miura, D.; Takenouchi, K.; Cornish, J.; Cundy, T.; Reddy, S. V.; Roodman, G. D. Vitamin D antagonist, TEI-9647, inhibits osteoclast formation induced by 1α,25-dihydroxyvitamin D_3 from pagetic bone marrow cells. *J. Steroid Biochem. Mol. Biol.* **2004**, *89–90*, 331–334. (b) Roodman, G. D.; Windle, J. J. Paget disease of bone. *J. Clin. Invest.* **2005**, *115*, 200–208.

(10) Saito, N.; Saito, H.; Anzai, M.; Yoshida, A.; Fujishima, T.; Takenouchi, K.; Miura, D.; Ishizuka, S.; Takayama, H.; Kittaka, A. Dramatic enhancement of antagonistic activity on vitamin D receptor: a double functionalization of 1α-hydroxyvitamin D_3 26,23-lactones. *Org. Lett.* **2003**, *5*, 4859–4862.

(11) Yoshimoto, N.; Inaba, Y.; Yamada, S.; Makishima, M.; Shimizu, M.; Yamamoto, K. 2-Methylene 19-nor-25-dehydro-1α-hydroxyvitamin D_3 26,23-lactones: synthesis, biological activities and molecular basis of passive antagonism. *Bioorg. Med. Chem.* **2008**, *16*, 457–473.

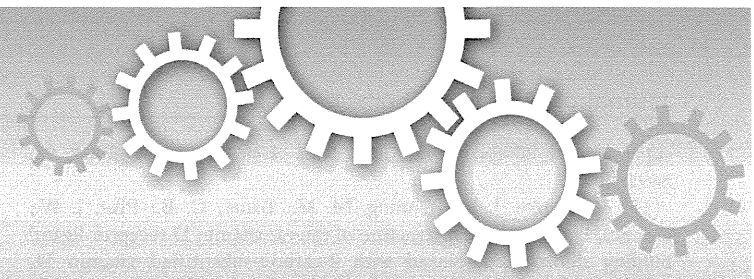
(12) Inaba, Y.; Yoshimoto, N.; Sakamaki, Y.; Nakabayashi, M.; Ikura, T.; Tamamura, H.; Ito, N.; Shimizu, M.; Yamamoto, K. A new class of vitamin D analogues that induce structural rearrangement of the ligand-binding pocket of the receptor. *J. Med. Chem.* **2009**, *52*, 1438–1449.

(13) Rochel, N.; Wurtz, J. M.; Mitschler, A.; Klaholz, B.; Moras, D. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol. Cell* **2000**, *5*, 173–179.

(14) Tocchini-Valentini, G.; Rochel, N.; Wurtz, J. M.; Mitschler, A.; Moras, D. Crystal structures of the vitamin D receptor complexed to

super agonist 20-epi ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5491−5496.
(15) Vanhooke, J. L.; Benning, M. M.; Bauer, C. B.; Pike, J. W.; DeLuca, H. F. Molecular structure of the rat vitamin D receptor ligand binding domain complexed with 2-carbon-substituted vitamin D_3 hormone analogues and a LXXLL-containing coactivator peptide. *Biochemistry* **2004**, *43*, 4101−4110.
(16) Ciesielski, F.; Rochel, N.; Mitschler, A.; Kouzmenko, A.; Moras, D. Structural investigation of the ligand binding domain of the zebrafish VDR in complexes with 1α,25$(OH)_2D_3$ and Gemini: purification, crystallization and preliminary X-ray diffraction analysis. *J. Steroid Biochem. Mol. Biol.* **2004**, *89*−*90*, 55−59.
(17) Tocchini-Valentini, G.; Rochel, N.; Wurtz, J. -M.; Moras, D. Crystal structures of the vitamin D nuclear receptor liganded with the vitamin D side chain analogues calcipotriol and seocalcitol, receptor agonists of clinical importance. Insights into a structural basis for the switching of calcipotriol to a receptor antagonist by further side chain modification. *J. Med. Chem.* **2004**, *47*, 1956−1961.
(18) Eelen, G.; Verlinden, L.; Rochel, N.; Claessens, F.; De Clercq, P.; Vandewalle, M.; Tocchini-Valentini, G.; Moras, D.; Bouillon, R.; Verstuyf, A. Superagonistic action of 14-epi-analogs of 1,25-dihydroxyvitamin D explained by vitamin D receptor−coactivator interaction. *Mol. Pharmacol.* **2005**, *67*, 1566−1573.
(19) Hourai, S.; Fujishima, T.; Kittaka, A.; Suhara, Y.; Takayama, H.; Rochel, N.; Moras, D. Probing a water channel near the A-ring of receptor-bound 1α,25-dihydroxyvitamin D_3 with selected 2α-substituted analogues. *J. Med. Chem.* **2006**, *49*, 5199−5205.
(20) Ciesielski, F.; Rochel, N.; Moras, D. Adaptability of the vitamin D nuclear receptor to the synthetic ligand Gemini: remodelling the LBP with one side chain rotation. *J. Steroid Biochem. Mol. Biol.* **2007**, *103*, 235−242.
(21) Rochel, N.; Hourai, S.; Perez-Garcia, X.; Rumbo, A.; Mourino, A.; Moras, D. Crystal structure of the vitamin D nuclear receptor ligand binding domain in complex with a locked side chain analog of calcitriol. *Arch. Biochem. Biophys.* **2007**, *460*, 172−176.
(22) Vanhooke, J. L.; Tadi, B. P.; Benning, M. M.; Plum, L. A.; DeLuca, H. F. New analogs of 2-methylene-19-nor-(20*S*)-1,25-dihydroxyvitamin D_3 with conformationally restricted side chains: evaluation of biological activity and structural determination of VDR-bound conformations. *Arch. Biochem. Biophys.* **2007**, *460*, 161−165.
(23) Hourai, S.; Rodrigues, L. C.; Antony, P.; Reina-San-Martin, B.; Ciesielski, F.; Magnier, B. C.; Schoonjans, K.; Mourino, A.; Rochel, N.; Moras, D. Structure-based design of a superagonist ligand for the vitamin D nuclear receptor. *Chem. Biol.* **2008**, *15*, 383−392.
(24) Shimizu, M.; Miyamoto, Y.; Takaku, H.; Matsuo, M.; Nakabayashi, M.; Masuno, H.; Udagawa, N.; DeLuca, H. F.; Ikura, T.; Ito, N. 2-Substituted-16-ene-22-thia-1α,25-dihydroxy-26,27-dimethyl-19-nor-vitamin D_3 analogs: synthesis, biological evaluation, and crystal structure. *Bioorg. Med. Chem.* **2008**, *16*, 6949−6964.
(25) Nakabayashi, M.; Yamada, S.; Yoshimoto, N.; Tanaka, T.; Igarashi, M.; Ikura, T.; Ito, N.; Makishima, M.; Tokiwa, H.; Deluca, H. F.; Shimizu, M. Crystal structures of rat vitamin D receptor bound to adamantyl vitamin D analogs: structural basis for vitamin D receptor antagonism and partial agonism. *J. Med. Chem.* **2008**, *51*, 5320−5329.
(26) Uskokovic, M. R.; Manchand, P. S.; Peleg, S.; Norman, A. W. Synthesis and Preliminary Evaluation of the Biological Properties of a 1α,25-Dihydroxyvitamin D_3 Analogue with Two Side-Chains. In *Vitamin D: Chemistry, Biology and Clinical Applications of the Steroid Hormone*; Norman, A. W., Bouillon, R., Thomasset, M., Eds.; University of California: Riverside, CA, 1997; pp19−21.
(27) Kurek-Tyrlik, A.; Makaev, F. Z.; Wicha, J.; Zhabinskii, V.; Calverley, M. J. Pivoting 20-Normal and 20-Epi Calcitriols: Synthesis and Crystal Structure of a "Double Side Chain" Analogue. In *Vitamin D: Chemistry, Biology and Clinical Applications of the Steroid Hormone*; Norman, A. W., Bouillon, R., Thomasset, M., Eds.; University of California: Riverside, CA, 1997; pp 30−31.
(28) Sakamaki, Y.; Inaba, Y.; Yoshimoto, N.; Yamamoto, K. Potent antagonist for the vitamin D receptor: vitamin D analogues with simple side chain structure. *J. Med. Chem.* **2010**, *53*, 5813−5826.
(29) Sicinski, R. R.; Prahl, J. M.; Smith, C. M.; DeLuca, H. F. New 1α,25-dihydroxy-19-norvitamin D_3 compounds of high biological activity: synthesis and biological evaluation of 2-hydroxymethyl, 2-methyl, and 2-methylene analogues. *J. Med. Chem.* **1998**, *41*, 4662−4674.
(30) Yamamoto, K.; Ogura, H.; Jukuta, J.; Inoue, H.; Hamada, K.; Sugiyama, Y.; Yamada, S. Stereochemical and mechanistic studies on conjugate addition of organocuprates to acyclic enones and enoates: simple rule for diastereofacial selectivity. *J. Org. Chem.* **1998**, *63*, 4449−4458.
(31) Inaba, Y.; Yamamoto, K.; Yoshimoto, N.; Matsunawa, M.; Uno, S.; Yamada, S.; Makishima, M. Vitamin D_3 derivatives with adamantane or lactone ring side chains are cell type-selective vitamin D receptor modulators. *Mol. Pharmacol.* **2007**, *71*, 1298−1311.
(32) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C−H···π-interactions in proteins. *J. Mol. Biol.* **2001**, *16*, 357−377.
(33) Yamamoto, K.; Sun, W.-Y.; Ohta, M.; Hamada, K.; DeLuca, H. F.; Yamada, S. Conformationally restricted analogs of 1α,25-dihydroxyvitamin D_3 and its 20-epimer: compounds for study of the three-dimensional structure of vitamin D responsible for binding to the receptor. *J. Med. Chem.* **1996**, *39*, 2727−2737.
(34) Leslie, A. G. W.; Powell, H. R. In *Evolving Methods for Macromolecular Crystallography*, ; Read, R. J., Sussman, J. L., Eds.; NATO Science Series II, Vol. *245*; Springer Verlag: Dordrecht, The Netherlands, 2007; pp 41−51.
(35) Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 271−281.
(36) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(37) Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760−763.
(38) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(39) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.
(40) Pannu, N. S.; Murshudov, G. N.; Dodson, E. J.; Read, R. J. Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 1285−1294.
(41) Winn, M. D.; Isupov, M. N.; Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 122−133.
(42) Steiner, R. A; Lebedev, A. A.; Murshudov, G. N. Fisher's information in maximum-likelihood macromolecular crystallographic refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 2114−2124.
(43) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−367.

SCIENTIFIC REPORTS



SUBJECT AREAS:
CANCER GENOMICS
HAEMATOLOGICAL CANCER
DNA DAMAGE AND REPAIR
ONCOGENES

Received
2 August 2012

Accepted
10 October 2012

Published
13 November 2012

Correspondence and requests for materials should be addressed to K.S. (shind009@kuhp.kyoto-u.ac.jp)

APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells

Masanobu Shinohara, Katsuhiro Io, Keisuke Shindo, Masashi Matsui, Takashi Sakamoto, Kohei Tada, Masayuki Kobayashi, Norimitsu Kadowaki & Akifumi Takaori-Kondo

Department of Hematology and Oncology, Graduate school of medicine, Kyoto University, Kyoto 606-8507, Japan.

Human APOBEC3 proteins play pivotal roles in intracellular defense against viral infection by catalyzing deamination of cytidine residues, leading to base substitutions in viral DNA. Activation-induced cytidine deaminase (AID), another member of the APOBEC family, is capable of editing immunoglobulin (Ig) and non-Ig genes, and aberrant expression of AID leads to tumorigenesis. However, it remains unclear whether APOBEC3 (A3) proteins affect stability of human genome. Here we demonstrate that both A3A and A3B can induce base substitutions into human genome as AID can. A3B is highly expressed in several lymphoma cells and somatic mutations occur in some oncogenes of the cells highly expressing A3B. Furthermore, transfection of *A3B* gene into lymphoma cells induces base substitutions in *cMYC* gene. These data suggest that aberrant expression of A3B can evoke genomic instability by inducing base substitutions into human genome, which might lead to tumorigenesis in human cells.

It is widely recognized that the accumulation of genetic changes in tumor-related genes is essential for cancer development[1]. With the innovation of high-throughput sequencing technology, genome-wide analyses on various types of cancer cells have revealed numerous somatic mutations in tumor-related genes[2]. Some of these mutations are caused by defects in DNA repair systems (e.g., DNA mismatch repair deficiencies give rise to hereditary non-polyposis colon cancer[3]), whereas mechanisms that account for the majority of genetic changes in cancer cells are poorly understood. Referring to somatic base substitution spectra in cancer cells, C/G to T/A transitions are most prevalent, especially in gastric cancer, colorectal cancer, glioma, and melanoma[2,4,5]. This strong bias in somatic mutations suggests the existence of active mechanisms that induce C/G to T/A transitions into genomic DNA. It is obviously attributable to ultraviolet irradiation and following repair process against pyrimidine dimer in case of melanoma, but not in others.

The human APOBEC family proteins can induce C to T (G to A, in complementary sequences) transitions into target DNA through cytidine deamination. The APOBEC family is comprised of a series of molecules with conserved cytidine deaminase domains (CDAs), including AID, APOBEC1, APOBEC2, APOBEC3A to H, and APOBEC4[6,7]. Among them, AID plays a crucial role in somatic hypermutation and class switch recombination of Ig genes, which enables diversification of immune system[8]. AID has been considered the only molecule that can induce C/G to T/A transitions into genomic DNA. The expression of AID is highly regulated and restricted in germinal center B-cells under physiological conditions, but with inflammatory stimulations, AID can be overexpressed in not only B-cells but also other types of cells (e.g., epithelial cells) via activation of NF-κB[9]. Aberrant expression of AID results in the accumulation of mutations in non-Ig genes[10], which leads to development of various cancers such as gastric and hepatic cancers as well as lymphomas[9,11–13].

A series of seven A3 genes are tandemly arrayed on human chromosome 22, and the main function of the resulting gene products is to protect the cells from retroviruses and endogenous mobile retroelements[14,15]. A3B, A3D, A3F, and A3G contain two CDAs, instead of one in A3A, A3C, and A3H. A3G is a powerful anti-retroviral molecule that induces cytidine deamination in viral genome and acts as a host defensive factor against viruses such as HIV-1[16]. A3A and A3B have been reported as potent inhibitors of retrotransposons[17]. Thus, A3 proteins act as sentinels in innate immunity against mobile DNA/RNA including viruses, while little is known about the effect of these proteins on nuclear DNA, in other words, host human genome. Recent studies have demonstrated that A3A impairs nuclear DNA under the condition of suppressing uracil DNA-glycosylase (UNG) which prevents base alterations by eliminating uracil from DNA and initiating the base-excision repair pathway[18,19].

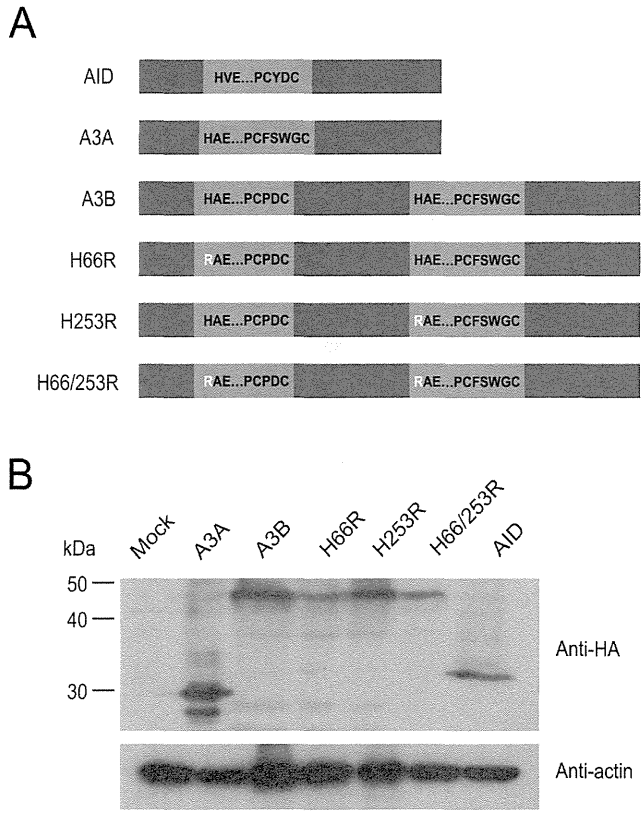


Figure 1 | Expression of A3A, A3B wild-type and mutants, and AID. (A) Schematic of expression vectors. The consensus amino acid residues for zinc-coordinating motifs are shown. Substituted residues are shown in white. (B) Expression of HA-tagged proteins. Expression vectors were transfected into HEK293 cells, and cell lysates were analyzed by immunoblotting with anti-HA antibody (top panel) and anti-ß-actin antibody (bottom panel) for loading control.

However, it is still unclear whether A3 proteins can induce somatic mutations into human genome with intact DNA repair systems. Here we first demonstrate that expression of A3B and A3A as well as AID can induce somatic mutations in genomic DNA in human cells even in the presence of UNG. We also find that high expression of A3B leads to somatic mutations in tumor-related genes. These data suggest that aberrant expression of A3B might be one of the active mechanisms that induce somatic mutations in cancer cells.

Results

A3 and AID induce hypermutations into foreign DNA. Besides A3A, we focused on A3B because it is localized predominantly in the nucleus[20,21] and highly expressed in many types of cancer cells[14] referring to microarray database (e.g., NextBio: http://www.nextbio.com). Previous studies have shown that A3B contains two enzymatically active CDAs in restricting HIV-1[22], whereas only carboxyl-terminal CDA is responsible for inhibiting HBV replication[23,24] and editing bacterial DNA[22]. A3B is also shown to restrict foreign DNA in mammalian cells[25], but it has not been tested which CDA is active in this context. First, to examine whether A3 and AID induce mutations in foreign DNA in human cells and which CDA is responsible for this DNA editing, we constructed amino- and/or carboxyl-terminal CDA mutants (H66R, H253R, and H66/253R) by site directed mutagenesis (Fig. 1a) and confirmed their expression in HEK293 cells by immunoblotting (Fig 1b). We transfected expression vectors for these together with EGFP expression vector into HEK293 cells and examined base substitutions in *EGFP* sequences. The expression vector for UNG inhibitor (UGI) was also co-transfected to avoid UNG-triggered degradation of uracil-containing foreign DNA as described previously[25]. We recovered total DNA from the cells 2 days after transfection and performed differential DNA denaturation PCR (3D-PCR) to efficiently recover edited DNA sequences[26]. 3D-PCR is based on the principle that DNA sequences with fewer interstrand hydrogen bonds dissociates easier. If cytidine deamination takes place frequently, resulting AT-rich *EGFP* gene can be amplified at lower denaturation temperatures. Although PCR products were obtained from all samples at 92°C of denaturation temperature (Td), we obtained robust PCR products at 83.8°C of Td only from A3A-, A3B wild-type (WT)-, and AID-expressing cells (Fig 2a). Amplification of *EGFP* at the lowest Td was impaired in H66R-expressing cells compared to A3B WT-expressing cells and undetectable in H253R- or H66/253R-expressing cells (Fig 2a, bottom). To ascertain whether *EGFP* gene was actually hyperedited, we cloned and sequenced the amplicons at 83.8°C of Td. As can be seen from the mutation matrices, high levels of C/G to T/A transitions were introduced into *EGFP* sequences (Fig 2b). To compare the extent of baseline mutations and that of A3B-induced mutations, we also cloned and sequenced the amplicons at 94.0°C of Td. Mutation frequency of A3B-expressing cells were about 6 times higher than that of mock-transfected cells (Supplementary Fig. S1 online). The mutation frequency in H66R-expressing cells was approximately a half compared to that in A3B WT-expressing cells in the amplicons at the lowest Td (Fig 2c). These data suggest that carboxyl-terminal CDA of A3B is mainly responsible for foreign DNA editing, but both domains are requisite for full editing activity. It is worth noting that AID is also capable of inducing cytidine deamination into foreign DNA.

Human A3 proteins have preferred target dinucleotide sequences in the substrate DNA; A3A and A3B prefer to deaminate cytosine residues flanked by 5′ thymine residue, 5′-TC, whereas A3G prefers to deaminate cytosine residues flanked by 5′ cytosine residue, 5′-CC[25,27–29]. We analyzed the context of C/G to T/A transitions in hyperedited *EGFP* sequences. We observed a strong bias toward deamination at 5′-TC dinucleotides in A3A-, A3B WT-, and H66R-expressing cells, but not in AID-expressing cells (Fig 2d). 5′-TC dinucleotide preference of A3B was also confirmed by sequencing amplicon at 94.0°C of Td which is supposed to be unbiased (Supplementary Fig. S1 online). These data suggest that the preference of editing sites in foreign DNA by A3s coincides with that seen in viral DNA.

A3A and A3B can edit genomic DNA in human cells. We next investigated whether A3 proteins induce C/G to T/A transitions into not only foreign DNA but also nuclear DNA in human cells. We first established a HEK293 cell line stably expressing EGFP (HEK293/EGFP) using retrovirus vector that carries *EGFP*. We transfected HEK293/EGFP cells with expression vectors for A3A, A3B WT or mutant (H66R, H253R, or H66/253R), or AID by lipofection, and then recovered total DNA from these cells after 7-day culture. We performed 3D-PCR of *EGFP* gene and obtained amplicons from A3A-, A3B WT-, H66R-, and AID-expressing cells at lower Td (Fig 3a). *EGFP* gene was recovered at Td as low as 86.3°C from A3B WT-expressing cells, while as low as 86.5°C from A3A-, H66R-, and AID-expressing cells. By contrast, *EGFP* gene was not amplified below Td of 87°C from cells transfected with mock, H253R or H66/253R. We repeated this procedure consisting of transfection, DNA extraction, and 3D-PCR three times and obtained similar results (Fig 3b). To unambiguously confirm the presence of C/G to T/A transitions, we cloned and sequenced amplicons obtained at the lowest Td in A3A-, A3B WT-, H66R-, and AID-expressing cells (Fig 3c). These analyses revealed approximately 2 to 5 C/G to T/A transitions per *EGFP* sequence from each sample (Fig 3d). The transitions were detected most frequently in A3A-expressing cells, and deaminase activity of H66R mutant was approximately a half

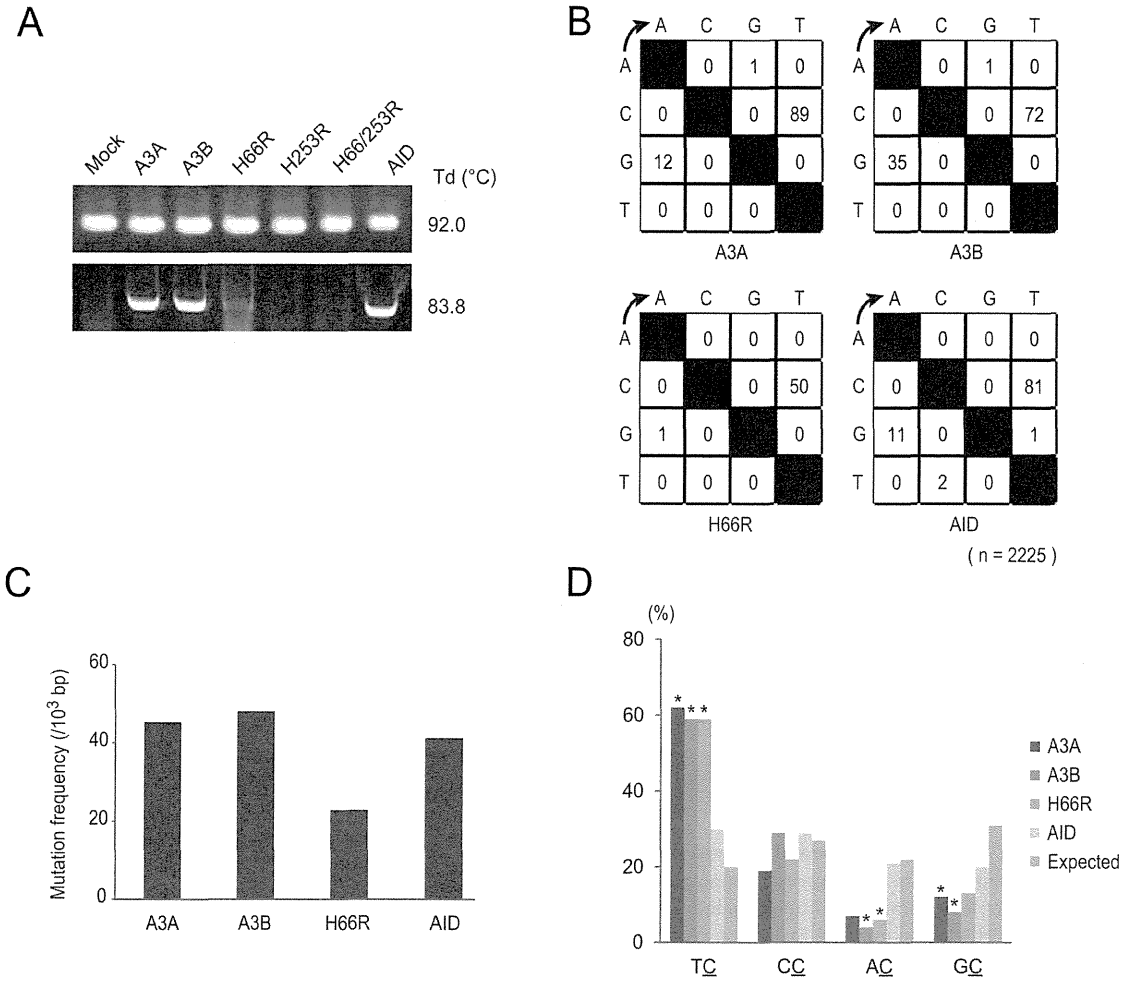


Figure 2 | Foreign DNA editing by A3A, A3B, and AID. (A) Agarose gel analyses of 3D-PCR products from HEK293 cells. Cells were transfected with expression vector for A3A, A3B wild-type or mutant, or AID together with pEGFP-N3 and pEF-UGI. Total DNA was recovered 2 days after transfection, and *EGFP* gene was amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Mutation matrices of hyperedited *EGFP* sequences derived from cloned amplicons at 83.8°C of Td. "n" indicates the number of bases sequenced. We sequenced 5 clones (2,225 base pairs in total) in each group. (C) Frequencies of C/G to T/A transitions in hyperedited *EGFP* genes. C/G to T/A transitions per 1,000 sequenced base pairs are shown. (D) Dinucleotide contexts in foreign DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Asterisks indicate statistical significance in a χ^2 test ($p < 0.01$).

compared to that of A3B WT as seen in foreign DNA assays. The contexts of C/G to T/A transitions detected from the lowest Td amplicons in genomic DNA editing in A3-expressing cells were distinct from those in foreign DNA editing (Fig 3e). A preference for 5′-TC dinucleotide was not apparently observed, alternatively, 5′-GC dinucleotides were preferred in all samples. However, this bias fails to reach statistical significance ($p < 0.01$) in a χ^2 test. The preferred target sequences of AID editing were 5′-GC and 5′-AC dinucleotides as described by many prior studies[27,30]. Mutation frequencies and preferred target sequence of A3B was also analyzed by using amplicons at 94.0°C of Td. Mutation frequency of A3B-expressing cells were about 3 times higher than that of mock-transfected cells (Supplementary Fig. S2 online). A preference for 5′-TC dinucleotide was impaired, compared to that in foreign DNA editing assays (Supplementary Fig. S2 online). Our results reveal that in addition to AID, A3A and A3B can induce C/G to T/A transitions into human nuclear DNA without repressing proofreading enzymes (e.g., UNG). Mutation frequencies were 6 to 9 per 1000 base pairs in A3A-, A3B WT-, and AID-expressing cells, and much less frequent compared to those in foreign DNA editing. As seen with foreign DNA editing, carboxyl-terminal CDA is mainly responsible for catalytic activity but not sufficient for full editing activity. The preference context of genomic DNA editing by A3A and A3B is different from that of viral or foreign DNA editing.

Deep sequencing reveals hyperediting of human genomic DNA by A3 proteins. Amplicon sequencing by next-generation sequencer has enabled to detect extremely low levels of mutations of targeted regions in genomic DNA. To verify more certainly that A3 proteins edit human nuclear DNA, we performed deep sequencing of A3-expressing cells. HEK293/EGFP cells were transfected with an empty vector or expression vectors for A3A, A3B WT, H66/253R, or AID by lipofection, and total DNA were extracted after 7-day culture. We amplified a portion of *EGFP* gene with 443 base pair length (from thymine 56 to cytosine 498) by conventional PCR protocol, not by 3D-PCR, and performed amplicon sequencing with the coverage of 1337 to 2654 reads per sample. This analysis revealed that extremely large numbers of nucleotides were substituted over the full length of amplicons in A3A-, A3B WT-, and AID-expressing cells, whereas very few mutations were detected in mock and H66/253R-expressing cells (Supplementary Table 1 online). C/G to T/A transitions were observed most frequently in A3A-expressing cells as variation rates reach approximately 7% at the maximum, while below 3% at most in A3B- and AID-expressing cells (Fig 4a

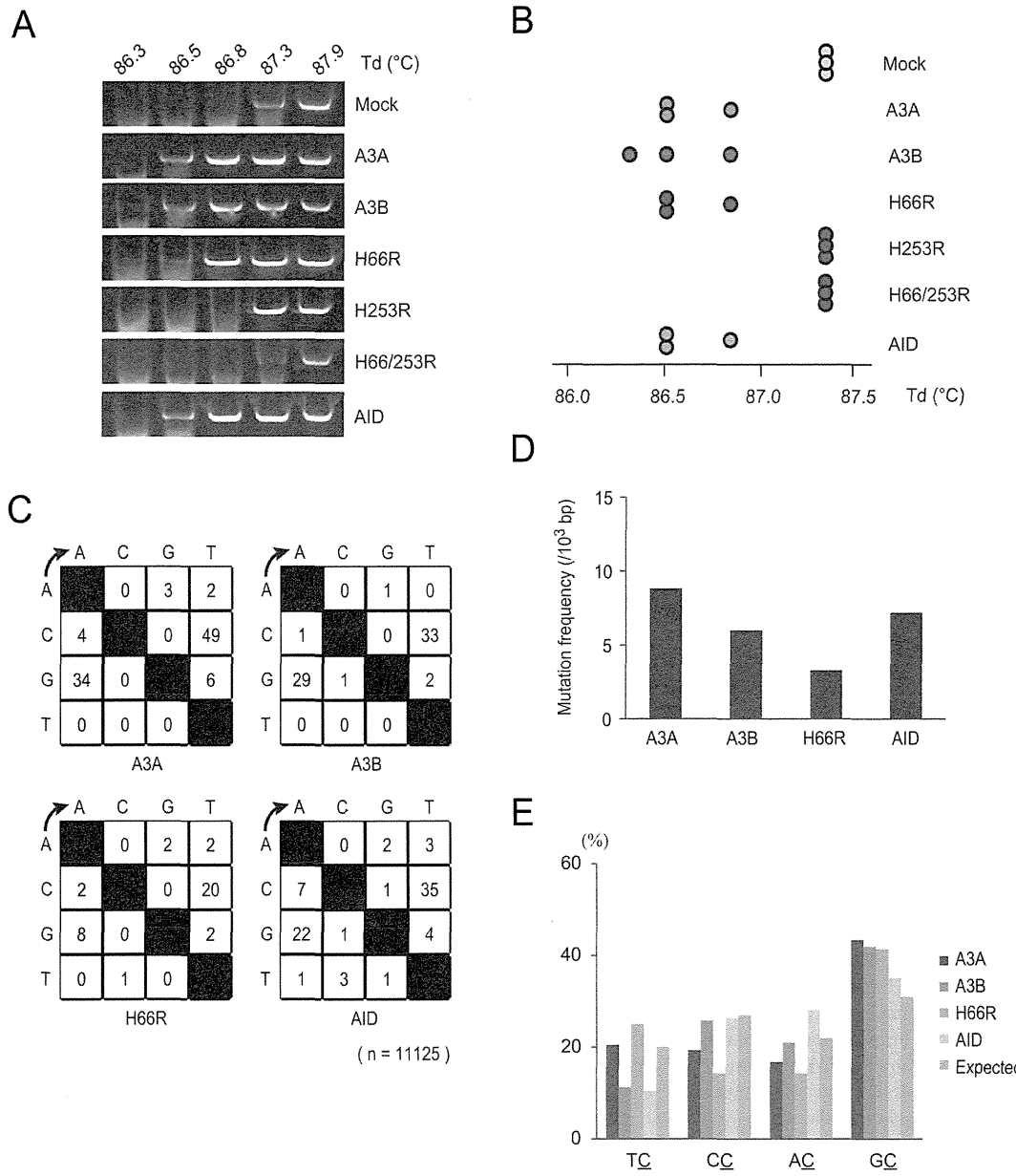


Figure 3 | Hypermutations in *EGFP* genes integrated in genomic DNA of HEK293 cells. (A) Agarose gel analyses of 3D-PCR products of *EGFP* genes extracted from HEK293/EGFP cells. Cells were transfected with expression vector for A3A, A3B wild-type or mutant, or AID. Total DNA was recovered 7 days after transfection and *EGFP* genes were amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Distributions of the lowest denaturation temperatures for positive PCR amplification in each sample. Each circle represents independent experiment consisting of transfection, DNA extraction, and 3D-PCR. (C) Mutation matrices of hyperedited *EGFP* sequences derived from cloned PCR products at Td lower than 87°C. "n" indicates the number of bases sequenced. We sequenced 25 clones (11,125 base pairs in total) in each group. (D) Frequencies of C/G to T/A transitions in hyperedited *EGFP* genes. C/G to T/A transitions per 1,000 sequenced base pairs are shown. (E) Dinucleotide contexts in genomic DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Deviations in the editing contexts do not reach statistical significance ($p < 0.01$) in a χ^2 test.

and Supplementary Table 1 online). The mutation frequency analysis revealed that large numbers of C/G to T/A substitutions were induced in A3A-, A3B-, and AID-expressing cells, whereas other types of base substitutions were very few (Fig 4b). These results are similar to the data obtained by 3D-PCR and clonal sequencing of A3- and AID-expressing cells, and further demonstrated that A3A and A3B as well as AID can induce C/G to T/A transitions into genomic DNA in human cells with intact DNA repair systems. Dinucleotide preference of target sequence for deamination by A3A, A3B and AID was also analyzed, however, we did not found any preference in this experiment (Fig. 4C), suggesting the difference between foreign DNA editing and genomic DNA editing.

Expression of A3B and somatic mutations in lymphoma cells. Although AID has been reported to play important roles in lymphomagenesis by inducing mutations in both Ig and non-Ig genes[11,12,31–34], AID-independent mechanisms are also suggested, because AID is not expressed in all types of B-cell lymphomas[31,35]. We hypothesized that A3 may contribute to somatic mutations in some lymphoma cells. To examine this hypothesis, we first determined expression levels of *A3A*, *A3B*, and *AID* by quantitative RT-PCR in several B-cell lymphoma cell lines using peripheral blood lymphocytes (PBL) as control (Fig 5a). Our analysis revealed that *A3B* was highly expressed in 3 of 4 cell lines, particularly, markedly high in KIS1 cells, whereas expression of *A3A* transcripts was not detected in any

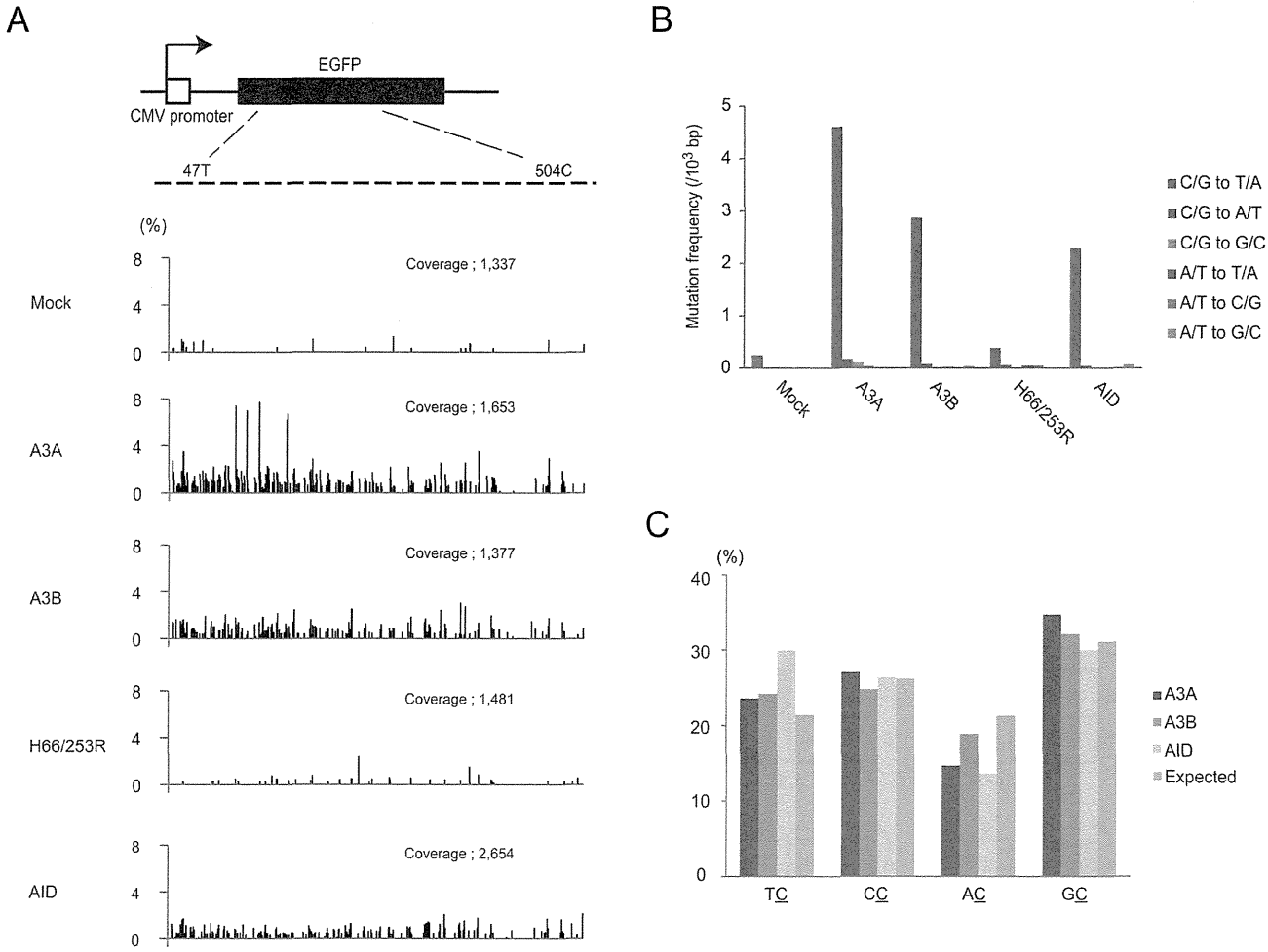


Figure 4 | Deep sequencing of *EGFP* genes in genomic DNA. (A) The distributions of C/G to T/A substitutions in the *EGFP* sequences. Total DNA was recovered form HEK293/EGFP cells 7 days after transfection with expression vector for A3A, A3B wild type or H66/253R or AID. We amplified a portion of *EGFP* sequence from thymine 47 to cytidine 504 (top schematic) by PCR with high-fidelity polymerase and sequenced the amplicons by GS-junior bench top system (Roche). Sequence data were analyzed with equipped software. "Coverage" indicates the total numbers of sequenced reads. (B) Frequencies of base substitutions in hyperedited *EGFP* genes. Base substitutions were classified to 6 groups and substituted base number of each group per 1,000 sequenced base pairs are show. (C) Dinucleotide contexts in genomic DNA editing. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Deviations in the editing contexts do not reach statistical significance ($p < 0.01$) in a χ^2 test.

lymphoma cell lines consistent with prior work suggesting myeloid specificity[25,36]. *AID* transcripts were detected in 2 of 4 cell lines, which is consistent with previous studies[31,32,37]. We also examined expression of *A3B* in two lymph node samples of diffuse large B-cell lymphoma, and found that A3B is actually expressed (supplementary Fig. 3 online).

To investigate the correlation between *A3B* expression and frequency of somatic mutations, we next performed direct sequencing of *cMYC*, *PAX5*, and *A20* genes which are exemplary genes mutated frequently in B-cell lymphoma[33,38]. We compared mutation frequencies of these genes in SUDHL-6 and KIS-1, because the expression of *A3B* was the lowest in the former and the highest in the latter, while AID was not expressed in either cell line. DNA sequences between exon 1 and intron 1 of these three genes were analyzed (899 base pairs of *cMYC*, 1550 base pairs of *Pax5*, and 1088 base pairs of *A20*), since it have been reported that somatic mutations induced by cytidine deaminases were concentrated within 2 kb downstream from transcription initiation sites[33,34]. We found nine mutations within investigated sequences of *cMYC* and *PAX5* in KIS-1, but not in SUDHL-6, in which five of nine mutations detected were C/G to T/A transitions. On the other hand, no mutation was detected within sequenced region of *A20* in either cells (Fig 5b). To analyze ongoing mutations in the genome in individual cells, we next sequenced the same region of *cMYC* sub-cloned from KIS-1 and SUDHL-6, and found several more C to T mutations in KIS-1 cells, but not in SUDHL-6 cells (Supplementary Fig. S4 online). We next determined expression of these tumor-related genes by quantitative RT-PCR and found that the transcripts of *cMYC* and *PAX5* were highly expressed in both SUDHL6 and KIS1 cells as compared to PBL, whereas *A20* was less transcribed in these lymphoma cells. These results suggest that high expression of *A3B* resulted in accumulation of base alterations, especially C/G to T/A transitions, in actively transcribed tumor-related genes in lymphoma cells.

To ascertain more definitely that A3B can edit tumor-related genes in lymphoma cells, we introduced *A3B* into a lymphoma cell line and analyzed somatic mutations in *cMYC*. SUDHL-6 cells were transfected with expression vector for A3B WT, H66/253R, or mock by electroporation and total DNA was extracted after 7-day culture. With 3D-PCR analysis of *cMYC*, we obtained the amplicon from only A3B WT-expressing cells at the lower Td (Fig 6a). Clonal sequencing of amplicons at 85.9°C revealed 2 to 7 nucleotide substitutions per strand, and more than 80% of these mutations were C/G to T/A transitions, with a preference for 5′-GC dinucleotide sites (Fig 6b and c). We also sequenced the amplicons at 94.0°C of Td and

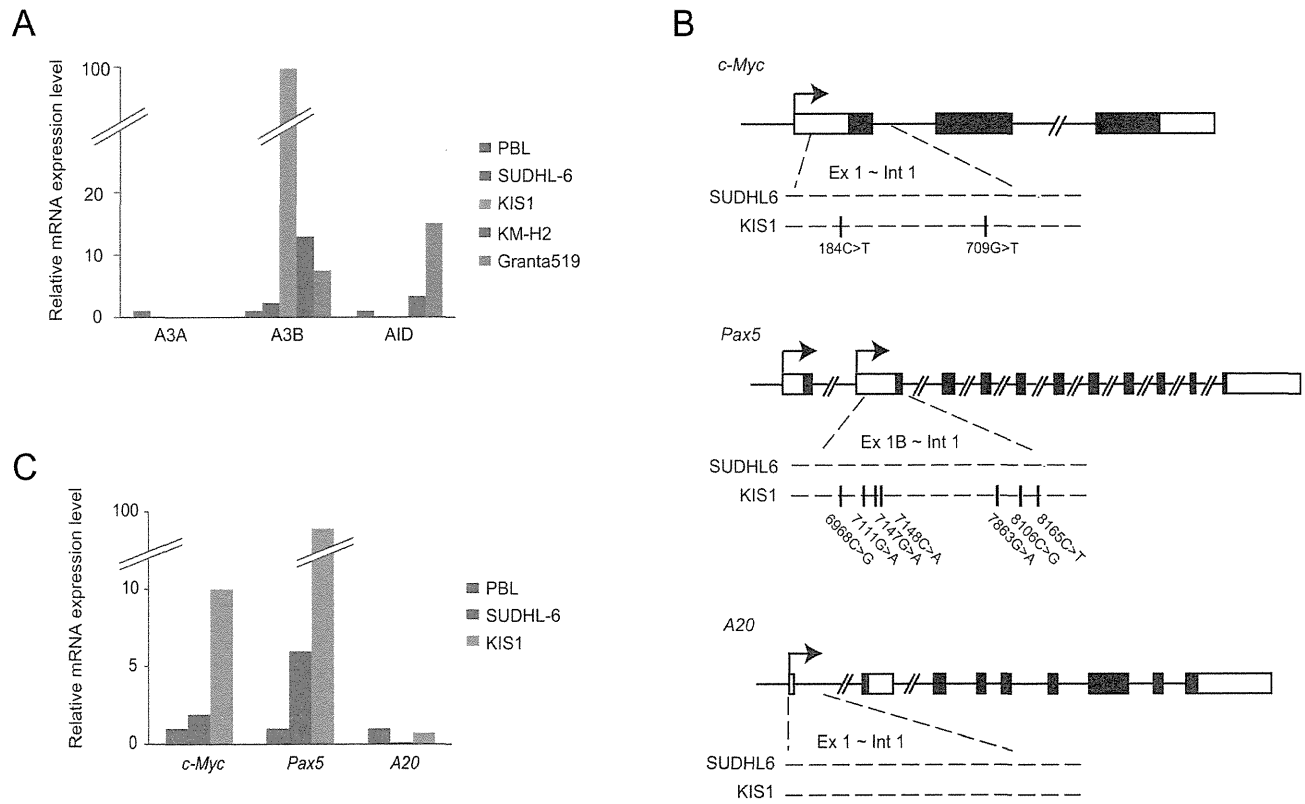


Figure 5 | Expression of *A3B* and somatic mutations in oncogenes in human lymphoma cell lines. (A) Quantitative RT-PCR for *A3A*, *A3B*, and *AID* in lymphoma cell lines. The levels of target cDNA were normalized to the endogenous hypoxanthine phosphoribosyl transferase 1 (*HPRT1*) and then compared to those in peripheral blood lymphocytes. (B) Mutational analyses of *C-myc*, *Pax5*, and *A20* in SUDH6 and KIS1 cells. We recovered total DNA from the cells and amplified the sequence between exon1 and intron1 of *C-myc*, *Pax5* and *A20* by PCR and performed direct sequencing of the amplicons. Locations of somatic mutations are shown below the loci with their positions. (C) The expression levels of transcripts of *C-myc*, *Pax5*, and *A20* in KIS1 and SUDHL6 cells. Quantitative RT-PCR was similarly performed with (a).

found A3B-induced C/G to T/A transitions without 5′-TC dinucleotide preference (supplementary Fig. 5 online). These data demonstrate that expression of A3B can induce somatic mutations into actively transcribed tumor-related genes in lymphoma cells.

Discussion

To date, most studies on A3 proteins have focused on their abilities as antiviral or antitransposon factors, whereas the capability of A3 proteins to induce mutations into genomic DNA in host cells has been scarcely verified. In contrast, many studies have elucidated that AID induces somatic mutations into not only Ig genes, but also tumor-related genes in human cells and that ubiquitous expression of AID in mice leads to cancers of various organs as well as lymphomas, with the accumulation of nucleotides alterations[9–13]. Thus, AID has been considered as the only DNA cytosine deaminase that can induce somatic mutations into human genome and has potential to cause cancers or hematologic malignancies.

Suspène *et al.* have recently reported that hyperediting of both mitochondrial and nuclear DNA was detected in human cells defective for UNG derived from hyper IgM syndrome patients and demonstrated that A3A-induced mutations in nuclear DNA are detectable under the condition of suppressing UNG in human cells[19]. In their report, deamination of nuclear DNA was not observed in cells expressing other A3 proteins or in cells expressing A3A without UNG suppression. Furthermore, Landry *et al.* have reported that expression of A3A together with UGI in mammalian cell lines resulted in breaking of DNA and activation of DNA damage response in a deaminase-dependent manner[18]. In these two reports, the effect on genomic integrity by A3A was dependent on the presence of UGI. Thus, there has been no direct evidence that A3 proteins induce mutations in genomic DNA in the cells with intact DNA repair systems.

In this study, we demonstrate that A3A and A3B as well as AID can induce C/G to T/A transitions into nuclear DNA without suppressing UNG by two different assays, 3D-PCR and deep sequencing. We assume that increased number of cytidine deamination catalyzed by highly expressed A3A or A3B exceeded the processivity of DNA repair enzymes such as UNG and resulted in leaving C/G to T/A transitions in nuclear DNA. Mutation frequencies were considerably lower compared to Suspène's report. However, Yoshikawa *et al.* reported that AID induced hypermutations into an actively transcribed gene in fibroblasts and that the mutation frequency was approximately 4 to 6 per 1000 base pairs[10], almost to the same extent as our results. Hence the frequency of mutations induced by cytidine deaminases into nuclear DNA is probably this extent or less in the cells with intact DNA repair systems.

We also find that *A3B* is highly expressed in several lymphoma cell lines and that the cells expressing high levels of *A3B* actually possesses somatic mutations, especially C/G to T/A transitions, in actively transcribed tumor-related genes. Furthermore, we reveal that introduction of *A3B* into lymphoma cells induces the accumulation of C/G to T/A transitions in *cMYC* gene. This is the first report that suggests the involvement of A3B in inducing somatic mutations of oncogenes in tumor cells. Together with the microarray database of A3B expression in miscellaneous cancer cell lines[14] (NextBio: http://www.nextbio.com), it is possible that A3B may induce somatic mutations into tumor related genes in various types of cancers.

Several questions remain open. First, it remains unclear what is preferred target sequences of A3 proteins in genomic DNA. Because

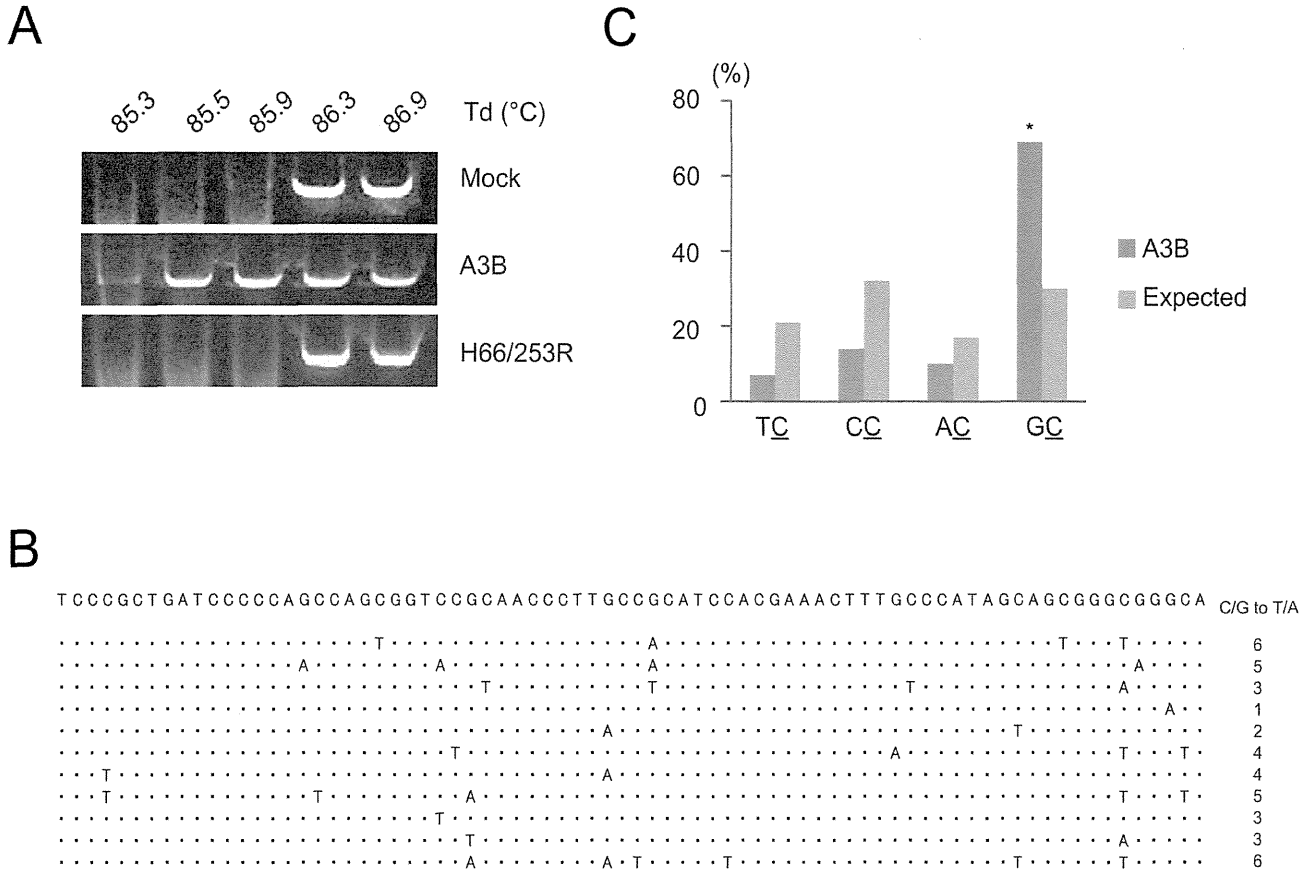


Figure 6 | A3B induced somatic mutations into *c-myc* gene in human lymphoma cells. (A) Agarose gel analyses of 3D-PCR products of *c-Myc* genes in SUDHL6. We transfected expression vector for A3B wild-type or H66/253R or empty vector and recovered total DNA 7 days after transfection. *C-myc* genes were amplified by 3D-PCR at the indicated denaturation temperatures (Td). (B) Clonal sequencing of amplicons from A3B-WT expressing SUDHL6 cells. We sequenced 11 clones (5104 base pairs in total). Seventy six bases from thymine 310 to adenine 385 in which mutations are concentrated among sequenced 464 base pairs are shown. The numbers of C/G to T/A substitutions in sequenced 464 base pair length are shown at the right end. (C) Dinucleotide contexts of somatic mutations in *c-Myc* gene by A3B. The rates of indicated dinucleotide sequence at the C to T transitions are shown. Asterisks indicate statistical significance in a χ^2 test ($p < 0.01$).

in several cancers such as breast cancer and melanoma, 5′-TC is the most prevalent target in C to T base substitutions, A3 is the most potential candidate to induce these mutations[25,28]. However, in our results, neither A3A nor A3B had a definite preference of editing site in nuclear DNA editing, whereas a preference for 5′-TC dinucleotide was observed in foreign DNA editing as previously reported. It is possible that A3A and A3B have no distinct favorite context in nuclear DNA editing, unlike viral and bacterial DNA editing, because human genomic DNA is more profoundly protected in transcription than viral or bacterial DNA and is under survey of DNA repair systems. However, Suspene *et al.* reported that target contexts of cytidine deamination in A3A+UGI-expressing cells were 5′-TC and 5′-CC dinucleotides, which were identical to the contexts of viral or bacterial DNA editing[19]. We assume that this discrepancy might be attributable to cell types or expression levels in cells. Hence, further analyses should be required to clarify the favorite target contexts of A3 proteins in nuclear DNA editing. The second question is how transcriptional control and post-translational modification of A3 proteins regulate A3 activity. Because the molecules that possess a capability of editing nuclear DNA threaten cell homeostasis, expression and activity of A3A and A3B must be strictly controlled. AID is known to be regulated at multiple steps[39], for example, transcriptional regulation[40–42], post-transcriptional regulation by micro-RNA[43,44], regulation of intracellular localization[45,46], and phosphorylation by PKA[47,48]. In contrast to AID, little is known about how A3 proteins are regulated. It has been reported that A3A is abundantly expressed in CD14+ monocytes and upregulated by interferon-α stimulation[25,49,50]. Meanwhile, it is not clear where A3B is expressed normally[25,42,49,50] and how it is regulated. As for post-translational modification, we previously reported that PKA-mediated phosphorylation of A3G regulates the interaction between A3G and HIV Vif[51]. To better understand the physiological roles of A3 proteins, it is important to elucidate how their expression and activity are regulated. The last question is whether A3 proteins can serve as an "initiator" of tumorigenesis. Our results suggest A3B indeed induces somatic mutations into genomic DNA in various human tumor cells, however, it is unclear whether A3B proteins impair genomic DNA from the early stage of oncogenesis. To address this question, hereafter, histopathological and genetic analyses of transgenic mouse constitutively expressing A3 proteins are necessary.

In conclusion, our findings provides the first evidence that A3A and A3B can induce C/G to T/G transitions into genomic DNA without suppressing DNA repair system. Our data also show that high expression of *A3B* is related to mutation frequencies of oncogenes in lymphoma cells. Our results suggest that *A3B* is an oncogene, like AID, which may have the capacity to evoke genomic instability through base substitutions in human cells. Further studies will be required to test whether endogenous A3B is capable of impairing genomic integrity as a DNA mutator and contributing to the development of human cancers and hematologic malignancies.

Methods

DNA constructs and cell lines. Plasmids containing coding sequence of human *A3A* and *A3B* were kindly provided by Dr. Kenzo Tokunaga[21]. Expression vectors for HA-tagged A3A, A3B and AID were generated by sub-cloning of coding sequences into pCAG-GS vector. A3B catalytic domain mutants (H66R, H253R, and H66/253R) were generated by KOD-plus mutagenesis Kit (Toyobo). Expression vector for Uracil-DNA glycosylase inhibitor, pEF-UGI was kindly provided by Dr. Ruben S Harris[25]. HEK293 and HEK293T cells were maintained with Dulbecco's modified Eagle's medium containing 10% of fetal bovine serum (FBS) and penicillin, streptomycin, and glutamine (PSG). All B-cell lymphoma cell lines were maintained with RPMI1640 containing 10% FBS and PSG. Retrovirus containing *EGFP* sequence was produced by co-transfection of pMLV gag-pol, VSV-G, and pDON-EGFP into HEK293T cells. HEK293/EGFP cells were generated by retroviral transduction of *EGFP* and selection of 1 mg/ml G418 for two months.

Immunoblotting. HEK293 cells were transfected with expression vector for A3A, A3B wild-type or mutant (H66R, H253R or H66/253R) or AID, and lysed with RIPA buffer (50 mM Tris-HCl pH7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton-X, 0.1% SDS, 0.1% DOC) after 2-day culture. After centrifugation at 20,000 x g for 15 min, supernatant was mixed with sample buffer (Biorad), boiled for 5 minutes, resolved on 12% (w/v) polyacrylamide gel, transferred to PVDF membrane (Immobilon, Millipore), and analyzed by standard immunoblotting procedure with anti-HA monoclonal antibody (12CA5, Roche) or anti-β-actin monoclonal antibody (AC-15, Sigma).

3D-PCR and clonal sequencing. For foreign DNA editing assay, HEK293 cells were transfected with pEGFP-N3, pEF-UGI, and expression vector for A3A, A3B WT or mutant, or AID by using Fugene HD (Roche). After two-day culture, total DNA was extracted by using Quick Gene DNA whole blood kit (Fuji Film). First round PCR was performed with primers listed in Supplementary Table S2 using rTaq DNA polymerase (Takara), with the following reaction profile; 30 s at 94°C, 25cycles of 30 s at 94°C, 40 s at 62°C, and 90 s at 72°C followed by 10 min at 72°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, and extracted from the gel using Qiaquick Gel Extraction kit (Qiagen). We used 25 ng of first-round PCR products as template for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen), with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 83–92°C, 60 s at 62°C, 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 83.8°C were cloned into pT7-blue vector (Novagen). For nuclear DNA editing assay, HEK293/EGFP cells were transfected with expression vector for A3A, A3B WT or mutants, or AID using Fugene HD (Roche). Seven days after transfection, we extracted total DNA from these cells with the same method of foreign DNA editing assay. First round PCR was performed using Advantage HF2 polymerase kit (Clontech), with the following reaction profile;1 min at 94°C, 30 cycles of 30 s at 94°C followed by 2 min at 68°C, followed by 3 min at 68°C. We used 25 ng of first-round PCR products for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen) with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 86–89°C, 60 s at 62°C, and 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 86.5°C and 83.8°C were cloned into pT7-blue vector (Novagen). For *c-myc* gene editing assay in lymphoma cells, we transfected SUDHL6 cells with expression vectors for A3B WT or H66/253R by electroporation using Nucleofector (Amaxa) and extracted total DNA from the cells 7 days after transfection. First round PCR and gel extractions of amplicons were performed with the same methods of nuclear DNA editing assay. We used 25 ng of first-round PCR products for nested PCR using Hotstar Hifidelity DNA polymerase (Qiagen), with the following reaction profile; 5 min at 95°C, 35 cycles of 15 s at 85–88°C, 60 s at 62°C, 80 s at 72°C, followed by 10 min at 72°C. The amplicons derived at 85.3°C were cloned into pT7-blue vector (Novagen) and sequenced using 3130xl Genetic Analyzer (Applied Biosystems).

Deep sequencing. Total DNA was extracted from HEK293/EGFP transfected with expression vectors for A3A, A3B WT or H66/253R or AID 7 days after transfection. A portion of *EGFP* gene with 443 base pair length, from thymine 56 to cytosine 498, was amplified with the primers listed in Supplementary Table 2 using Advantage HF2 polymerase kit (Clontech), with the following reaction profile; 1 min at 94°C, 30 cycles of 30 s at 94°C followed by 2 min at 68°C, and followed by 3 min at 68°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, and extracted from the gel using Quiaquick Gel Extraction Kit (Qiagen). Purified amplicons were sequenced using GS junior bench top system (Roche) according to the manufacturer's protocol and analyzed with equipped software, GS Amplicon Variant Analyzer.

Lymphoma cell lines and patient samples. Four B-cell lymphoma cell lines (SUDHL6, KIS1, KM-H2, and Granta519) were cultured in RPMI1640 containing 10% of fetal bovine serum (FBS) and penicillin, streptomycin, and glutamine (PSG). We extracted total DNA from these cells by using Quick Gene DNA whole blood kit (Fuji Film) and total RNA by using mir Vana miRNA isolation kit (Ambion). Tumor biopsy specimens prior to treatment were obtained from two patients with diffuse large B-cell lymphoma. The study was approved by the Kyoto University Institutional Review Board and written informed consent was obtained from each patient. Total RNA were extracted similarly to lymphoma cell lines. Naïve B-cells were isolated from healthy donor's peripheral blood by using MACS® naive B cell isolation kit (Miltenyi Biotec).

Quantitative RT-PCR. Complementary DNA was synthesized from 200 ng of total RNA using Revertra Ace qPCR RT Master Mix (Toyobo). Real-time PCR were performed with Thunderbird SYBR qPCR Mix (Toyobo) according to manufacturer's protocol. Target cDNAs were normalized to the endogenous expression level of the house keeping reference gene for hypoxanthine-guanine phosphoribosyl transferase 1 (*HPRT1*) or glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). All primers for real-time PCR are listed in Supplementary Table 2.

Sequencing of oncogenes from lymphoma cell lines. We amplified portions of *C-myc*, *Pax5*, and *A20* from with the primers listed in Supplementary Table 2 using Advantage HF2 polymerase kit (Clontech), with the following reaction profile; 1 min at 94°C, 30 cycles of 30 s at 94°C followed by 4 min at 68°C, and followed by 3 min at 68°C. The amplicons were separated by electrophoresis on 1% (w/v) agarose gel, extracted from the gel using Qiaquick Gel Extraction kit (Qiagen), and sequenced using 3130xl Genetic Analyzer (Applied Biosystems). In *c-Myc* clonal sequencing, the amplicon was subcloned into pTA2-vector (TOYOBO) and subsequently sequenced.

1. Hahn, W. C. & Weinberg, R. A. Rules for making human tumor cells. *N Engl J Med* **347**, 1593–603 (2002).
2. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–6 (2010).
3. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–61 (1994).
4. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–8 (2007).
5. Prickett, T. D. *et al.* Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nat Genet* **41**, 1127–32 (2009).
6. Macduff, D. A. & Harris, R. S. Directed DNA deamination by AID/APOBEC3 in immunity. *Curr Biol* **16**, R186–9 (2006).
7. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229 (2008).
8. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–63 (2000).
9. Matsumoto, Y. *et al.* Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* **13**, 470–6 (2007).
10. Yoshikawa, K. *et al.* AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**, 2033–6 (2002).
11. Okazaki, I. M. *et al.* Constitutive expression of AID leads to tumorigenesis. *J Exp Med* **197**, 1173–81 (2003).
12. Pasqualucci, L. *et al.* AID is required for germinal center-derived lymphomagenesis. *Nat Genet* **40**, 108–12 (2008).
13. Endo, Y. *et al.* Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* **135**, 889–98, 898 e1-3 (2008).
14. Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–96 (2002).
15. Goila-Gaur, R. & Strebel, K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* **5**, 51 (2008).
16. Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
17. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**, 8780–5 (2006).
18. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep* **12**, 444–50 (2011).
19. Suspène, R. *et al.* Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci U S A* **108**, 4858–63 (2011).
20. Lackey, L. *et al.* APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol* **419**, 301–14 (2012).
21. Kinomoto, M. *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* **35**, 2955–64 (2007).
22. Bogerd, H. P., Wiegand, H. L., Doehle, B. P. & Cullen, B. R. The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains. *Virology* **364**, 486–93 (2007).
23. Suspène, R. *et al.* Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. *Proc Natl Acad Sci U S A* **102**, 8321–6 (2005).
24. Bonvin, M. & Greeve, J. Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. *J Gen Virol* **88**, 3270–4 (2007).
25. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol* **17**, 222–9 (2010).
26. Suspène, R., Henry, M., Guillot, S., Wain-Hobson, S. & Vartanian, J. P. Recovery of APOBEC3-edited human immunodeficiency virus G->A hypermutants by differential DNA denaturation PCR. *J Gen Virol* **86**, 125–9 (2005).

27. Vartanian, J. P. *et al.* Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog* **6**, e1000928 (2010).
28. Bishop, K. N. *et al.* Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol* **14**, 1392–6 (2004).
29. Harris, R. S. *et al.* DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–9 (2003).
30. Beale, R. C. *et al.* Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol* **337**, 585–96 (2004).
31. Greeve, J. *et al.* Expression of activation-induced cytidine deaminase in human B-cell non-Hodgkin lymphomas. *Blood* **101**, 3574–80 (2003).
32. Deutsch, A. J. *et al.* MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation. *Blood* **109**, 3500–4 (2007).
33. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–6 (2001).
34. Kotani, A. *et al.* A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc Natl Acad Sci U S A* **102**, 4506–11 (2005).
35. Smit, L. A. *et al.* Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype. *Cancer Res* **63**, 3894–8 (2003).
36. Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr Biol* **16**, 480–5 (2006).
37. Mottok, A., Hansmann, M. L. & Bräuninger, A. Activation induced cytidine deaminase expression in lymphocyte predominant Hodgkin lymphoma. *J Clin Pathol* **58**, 1002–4 (2005).
38. Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712–6 (2009).
39. Delker, R. K., Fugmann, S. D. & Papavasiliou, F. N. A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat Immunol* **10**, 1147–53 (2009).
40. Crouch, E. E. *et al.* Regulation of AID expression in the immune response. *J Exp Med* **204**, 1145–56 (2007).
41. Gonda, H. *et al.* The balance between Pax5 and Id2 activities is the key to AID gene expression. *J Exp Med* **198**, 1427–37 (2003).
42. Pauklin, S., Sernández, I. V., Bachmann, G., Ramiro, A. R. & Petersen-Mahrt, S. K. Estrogen directly activates AID transcription and function. *J Exp Med* **206**, 99–111 (2009).
43. Teng, G. *et al.* MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* **28**, 621–9 (2008).
44. de Yébenes, V. G. *et al.* miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* **205**, 2199–206 (2008).
45. Ito, S. *et al.* Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* **101**, 1975–80 (2004).
46. Patenaude, A. M. *et al.* Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* **16**, 517–27 (2009).
47. Basu, U. *et al.* The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–11 (2005).
48. McBride, K. M. *et al.* Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J Exp Med* **205**, 2585–94 (2008).
49. Koning, F. A. *et al.* Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol* **83**, 9474–85 (2009).
50. Berger, G. *et al.* APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. *PLoS Pathog* **7**, e1002221 (2011).
51. Shirakawa, K. *et al.* Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat Struct Mol Biol* **15**, 1184–91 (2008).

Acknowledgments

We thank R. Harris for thoughtful feedback and for UGI vector, and K. Tokunaga for A3A and A3B vectors.

Author contributions

M.S. performed most of the experiments, analyzed the data and wrote the manuscript; I.K., M.M., T.S. and K.T. performed sequencing of lymphoma cell lines; K.S. and N.K. wrote the manuscript; M.K. established q-PCR; A.T.K. designed the experiments and wrote the manuscript.

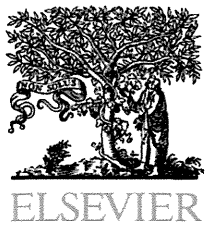
Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing interest statements: The authors declare no competing financial interests.

How to cite this article: Shinohara, M. *et al.* APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci. Rep.* **2**, 806; DOI:10.1038/srep00806 (2012).

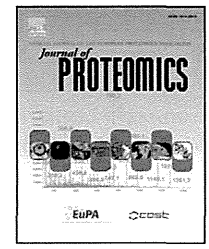
JOURNAL OF PROTEOMICS 75 (2012) 4863–4873



Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



Molecular and enzymatic characterization of XMRV protease by a cell-free proteolytic analysis

Satoko Matsunaga[a], *Tatsuya Sawasaki*[b], *Hirotaka Ode*[c], *Ryo Morishita*[a,d], *Ayako Furukawa*[e], *Ryuta Sakuma*[f], *Wataru Sugiura*[c], *Hironori Sato*[g], *Masato Katahira*[e], *Akifumi Takaori-Kondo*[h], *Naoki Yamamoto*[i], *Akihide Ryo*[a,*]

[a]*Department of Microbiology, Yokohama City University School of Medicine, Yokohama 236-0004, Japan*
[b]*Cell-Free Science and Technology Research Center, Ehime University, Matsuyama 790-8577, Japan*
[c]*Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya 460-0001, Japan*
[d]*CellFree Sciences Co., Ltd., Ehime Univ. Venture Business Laboratory, Matsuyama 790-8577, Japan*
[e]*Institute of Advanced Energy, Kyoto University, Kyoto 611-0011, Japan*
[f]*Department of Molecular Virology, Tokyo Medical and Dental University, Tokyo 113-8510, Japan*
[g]*Pathogen Genomics Center, National Institutes of Infectious Diseases, Tokyo 208-0011, Japan*
[h]*Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan*
[i]*Department of Microbiology, National University of Singapore, Singapore 117597, Singapore*

ARTICLE INFO

Article history:
Received 21 March 2012
Accepted 31 May 2012
Available online 9 June 2012

Keywords:
XMRV
Protease
Cell-free protein synthesis
AlphaSceen

ABSTRACT

Xenotropic murine leukemia virus-related virus (XMRV) is a virus generated under artificial conditions by the recombination of 2 murine leukemia virus (MLV) proviruses, PreXMRV-1 and PreXMRV-2, during the in vivo passage of human prostate cancer cells in athymic nude mice. The molecular etiology of XMRV infection has not been characterized and its implication in human prostate cancer progression remains equivocal. As a step toward resolving this issue we developed an in vitro enzymatic assay system to characterize XMRV protease (PR)-mediated cleavage of host-cell proteins. Enzymatically-active XMRV PR protein was synthesized using a wheat-germ cell-free system. By monitoring cleavage activity of XMRV PR by AlphaScreen and 2-color immunoblot analyses, we revealed that the catalytic activity of XMRV PR is selectively blocked by the HIV PR inhibitor, Amprenavir, and identified several human tumor suppressor proteins, including PTEN and BAX, to be substrates of XMRV PR. This system may provide an attractive means for analyzing the function of retrovirus proteases and provide a technology platform for drug screening.

1. Introduction

Xenotropic murine leukemia virus-related virus (XMRV) was originally isolated from a human prostate cancer (PC) in 2006 [1]. This virus is highly homologous to several endogenous Murine leukemia viruses (MLVs) found in mice [2]. Although previous reports suggested the involvement of XMRV in PC as well as chronic fatigue syndrome (CFS), as an etiological agent, no evidence of this etiological link between XMRV and human disease has been shown to date [3–5].

The nucleotide sequence of XMRV isolated from humans indicates that the virus is nearly identical with the XMRV isolated from the human prostate tumor cell line 22Rv1 [6]. This cell line was generated by serial passage of human prostate tumor tissue in nude mice. Sequence analysis revealed that the genomes of these mouse strains contain two different proviral

* *Corresponding author.* Tel.: +81 45 787 2602; fax: +81 45 787 2851.
E-mail address: aryo@yokohama-cu.ac.jp (A. Ryo).

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.05.047

DNAs related to XMRV (Pre-XMRV-1 and 2). It appears that these proviral genomes recombined to produce the XMRV isolated from 22Rv1 cells. It is plausible that the reported association of XMRV with human disease is due to contamination of human samples with virus originating from this recombination event in mice prior to the analysis.

While XMRV arose from an unusual recombination, several lines of study have indicated that XMRV can indeed infect, and proliferate in, several human prostate cancer cell lines including LNCaP and PC3 [7]. It is of significance that dihydrotestosterone (DHT) was shown to stimulate transcription and replication of XMRV through the transactivation of the XMRV-LTR via the hormone response element (HRE) [8]. Mutations in the HRE of XMRV impaired basal transcription and androgen responsiveness, suggesting a relationship between virus productivity and prostatic hypertrophy and neoplasia.

If XMRV is indeed an etiological agent in PCs, detection and elimination of XMRV infection could provide an effective strategy for early diagnosis and treatment of this tumor. To date, however, conflicting epidemiological data has precluded investigations into whether the virus is truly pathogenic or not and, moreover, whether this virus is oncogenic and associated with human PCs.

Retroviral protease (PR) is essential for virus replication and has been the major target for anti-retroviral therapy. Concomitant with particle release, the virally-encoded PR cleaves Gag into its four mature protein domains; MA, CA, NC and p12, in case of XMRV. Gag-Pol is also cleaved by PR, creating the viral enzymes RT, IN and PR itself. The maturation step coupled with PR activation is essential to confer viral infectivity. Therefore, for the effective inhibition of XMRV infection, should this virus be found to be pathogenic, development of anti-retroviral drugs targeting the XMRV protease would seem logical.

Several recent studies have indicated that viral-protease cleavage of host proteins that promotes viral replication and cytopathic effects [9–11]. It remains elusive whether this is the case for XMRV PR. One of the major bottlenecks in PR research has been the difficulty in producing recombinant protein with enzymatic activity in conventional cell-associated protein expression systems (e.g. *E. coli* or insect cell) due to host cell toxicity.

The wheat germ protein production system is a robust in vitro cell-free protein synthesis system comprising a crude wheat germ extract containing ions (buffer) and all the macromolecular components required for protein translation (e.g. ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination factors) [12,13]. The extract is further supplemented with amino acids, energy sources such as ATP, energy generating systems and cofactors for efficient and abundant protein production. This system is a powerful tool for the preparation of multiple proteins at one time, and large amounts of specific proteins for both biochemical and biomedical applications. It also enables synthesis of specific proteins that are difficult to express and/or purify in *E. coli* or other cell-based systems. The wheat germ system can produce a wide variety of proteins, including viral proteins, in sufficient amounts for functional assays. Furthermore, this system can yield enzymatically-active proteins in their naturally folded state owing to the constituent eukaryotic translation and folding machinery.

In our current study, we utilized the wheat germ cell-free system to synthesize catalytically active XMRV PR for the identification of potential inhibitors and substrates. By this approach we delineated a molecular link between XMRV PR and human tumor suppressor proteins pointing to a potential role in oncogenesis.

2. Materials and methods

2.1. *Construction of DNA template for transcription and the protein expression*

DNA templates were made using the Gateway and split-primer polymerase chain reaction (PCR) systems [12,14,15]. The tumor suppressor genes were amplified from MGC, a human cDNA resource purchased from Danaform (Tokyo, Japan).

The XMRV protease fragment was amplified by PCR from XMRV VP62 clone [1,16] using the following forward and reverse primers, respectively: (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAAGGACTGCCCAAAGAAGCC-3′) and (5′-GGGGACCACTTTGTACAAGAA AGCTGGGTCTTATAGAGGAACATCTGGCTC-3′). For HIV-1 protease, cDNA fragment was amplified by PCR from HIV-1 NL4-3 clone by using the following forward and reverse primers, respectively: (5′-CCACCCACCACCACCAATGTTTTTTTAGGGAAGATC-3′) and (5′-TCCAGCACTAGCTCCAGATTAGCCATCCATTCCTGGC-3′). Subsequently, attB-flanked fragment were amplified by PCR using attB1-S1 primer (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCACCCACCACCACCAATG-3′) and attB2-T1 primer (5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCCAGCACTAGCTCCAGA-3′). The single Capsid (CA) and Nucleocapsid (NC) fragment of XMRV Gag (control substrate), and the 24 tumor suppressor gene fragments (test substrates) were amplified by two-step PCR (without stop codon). The first round of PCR was performed on using 10 nM of S1 (forward) primer (5′- CCACCCACCACCACCAatg(n)19-3′) and T1 (reverse) primer (5′-TCCAGCACTAGCTCCAGA(n)19-3′). The second round of PCR was carried out using the first PCR product as template, with 100nM of attB1-FLAG-S1 (forward) primer (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGACTACAAGGATGACGATGACAAGCTCCACCCACCACCACCAATG-3′) and T1-biotin ligase site (bls)-stop-attB2-anti (reverse) primer (5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTCGTGCCACTCGATCTTCTGGGCCTCGAAGATGTCGTTCAGGCCGCTTCCAGCACTAGCTCCAGA-3′) [17].

The amplified attB-flanked fragments were each inserted into the pDONR221 vector using the Gateway BP Clonase II enzyme mix (Invitrogen, Carlsbad, CA, USA) to give pDONR-XMRV PR vector and pDONR-FLAG-gene-bls vectors, respectively.

pDONR-XMRV PR and pDONR-FLAG-CA/NC-bls vectors were subcloned into pEU-E01-GW [18] to generate the pEU-based-plasmids by LR reaction. BP and LR reactions were performed according to the manufacturer's instructions (Invitrogen). PCR reactions were performed using PrimeStar enzyme according to the manufacturer's instructions (Takara Bio, Otsu, Japan).

For HIV-1 PR substrate, the p2–p7 fragment in HIV-1 Gag precursor was amplified by PCR from HIV-1 NL4-3 clone by using the following forward and reverse primers, respectively: (5′-GAGACTCGAGGCCGAGGCCATGAGCCAGG-3′) and

(5′-GAGCGGTACCTTATTCGTGCCACTCGATCTTCTGGGCCTC-GAAGATGTCGTTCAGGCCATTAGCCTGCCTCTCGGTGCA-3′). The p2–p7-bls fragment was inserted into the pEU-E01-GST-MCS vector (Cellfree Sciences, Yokohama, Japan). The transcription template from pEU vector was amplified using the following forward and reverse primers, respectively: SPu primer (5′-GCGTAGCATTTAGGTGACACT-3′) and AODA2303 (5′-GTCA-GACCCCGTAGAAAAGA-3′). PCR was carried out using the TaKaRa Ex Taq (Takara Bio Inc, Shiga, Japan) according to the manufacturer's instructions.

DNA templates of human genes for transcription were constructed using split-primer PCR in two steps as described previously [18]. For the first step, S1 primers and pDONR221-1st_4080 (5′-ATCTTTTCTACGGGGTCTGA-3′) or AODA2306 (5′- AGCGTCAGACCCCGTAGAAA -3′) were used. For the second step, the primers SPu and pDONR221-2nd_4035 (5′-ACGTTAAGG-GATTTTGGTCA-3′) or AODA2303 were used to generate the final DNA template for transcription.

2.2. *Cell-free protein synthesis*

In vitro transcription and cell-free protein synthesis was performed as described previously [19]. Transcripts were made from each of the DNA templates mentioned above using SP6 RNA polymerase. The synthetic mRNAs were then precipitated with ethanol, collected by centrifugation and washed. Each mRNA (typically 30–35 μg) was added to the translation mixture and the translation reaction was performed in the bilayer mode [20] with slight modifications. The translation mixture that formed the bottom layer consisted of 60 A260 units of wheat germ extract (CellFree Sciences) and 2 μg creatine kinase (Roche Diagnostics K. K., Tokyo, Japan) in 25 μl SUB-AMIX solution (CellFree Sciences). SUB-AMIX contained (final concentrations) 30 mM Hepes/KOH at pH 8.0, 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 4 mM DTT, 0.4 mM spermidine, 0.3 mM each of the 20 amino acids, 2.7 mM magnesium acetate, and 100 mM potassium acetate. SUB-AMIX (125 μl) was placed on the top of the translation mixture, forming the upper layer. After incubation at 16 °C for 16 h, protein synthesis was confirmed by SDS-PAGE. For biotin labeling, 1 μl (50 ng) of crude biotin ligase (BirA) produced by the wheat germ cell-free expression system was added to the bottom layer, and 0.5 μM (final concentration) of D-biotin (Nacalai Tesque, Inc., Kyoto, Japan) was added to both upper and bottom layers, as described previously [21].

2.3. *Detection of cleavage activity of XMRV protease by luminometry*

In vitro cleavage activity assays of XMRV protease were carried out in a total volume of 15 μl consisting of 100 mM Tris–HCl pH 8.0, 0.01% Tween-20, 1 mg/ml BSA, 1 μl crude recombinant protease (~0.75 μM) and 0.5 μl crude recombinant FLAG-biotin-tagged CA/NC (~0.037 μM) at 37 °C for 1 h in a 384-well Optiplate (PerkinElmer, Boston, MA, USA). To assay the effects of HIV protease inhibitors on XMRV protease, after 3 μl recombinant viral protease and HIV protease inhibitor was incubated at 37 °C for 10 min, FLAG-biotin-tagged CA/NC or GST-biotin-tagged p2–p7 was added and the reaction further incubated at 37 °C for 1 h in a 384-well Optiplate. In accordance with the AlphaScreen IgG (Protein A) detection kit (PerkinElmer) instruction manual, 10 μl of detection mixture containing 100 mM Tris–HCl pH 8.0, 0.01% Tween-20, 1 mg/ml BSA, 5 μg/ml Anti-FLAG antibody (Sigma-Aldrich, St. Louis, MO, USA) or Anti-GST antibody (GE Healthcare, Buckinghamshire, UK), 0.1 μl streptavidin-coated donor beads and 0.1 μl anti-IgG (Protein A) acceptor beads were added to each well followed by incubation at 26 °C for 1 h. Luminescence was analyzed by the AlphaScreen detection program. Each assay was performed in triplicate, and the data represent the means and standard deviations of three independent experiments.

2.4. *Detection of cleavage activity by immunoblotting*

3 μl crude recombinant viral protease (~0.75 μM) and 7 μl crude FLAG-biotin-tagged recombinant proteins were incubated at 37 °C for 2 h. To assay the effect of HIV protease inhibitors, 3 μl crude recombinant XMRV protease and 1 μl of 10 μM HIV protease inhibitor were incubated at 37 °C for 10 min followed by addition of 6 μl crude FLAG-biotin-tagged recombinant proteins, and incubated at 37 °C for 120 min. Proteins were separated by SDS-PAGE and transferred to a PVDF membrane (Millipore Bedford, MA, USA) according to standard procedures. Immunoblot analysis was carried out with anti-FLAG (M2) antibodies (Sigma-Aldrich) or Streptavidin-HRP conjugate (GE Healthcare) according to the procedure described above. For fluorescent imaging, immunoblotted proteins were detected by Alexa592-anti-mouse antibodies (N-cleaved fragments), and Alexa488-streptavidin (C-cleaved fragments). The labeled proteins were visualized using a Typhoon Imager (GE Healthcare).

2.5. *Homology modeling of XMRV PR in complex with APV*

To predict interactions between XMRV PR and APV, we performed homology modeling [22] of the complex structure formed between XMRV PR and APV using the Molecular Operating Environment (MOE) software ver. 2008.10. (Chemical Computing Group, Canada). Firstly, the homologues of XMRV PR were searched for with the MOE-search PDB module from the MOE homology databank. Secondly, to minimize misalignments of the target sequence, multiple alignments were made using sequences of the homologues and those of HIV-1 PR (PDB code: 1HPV) and HTLV PR (PDB code: 3LIN) with the MOE-Align module. The aligned sequences showed that amino acids at the active site of HIV-1, HTLV PRs and those likely to be at the active site of XMRV PR were comparatively conserved, suggesting a structure of HIV-1 PR with APV (PDB code: 1HPV) would be appropriate for a template structure to predict interactions between XMRV PR and APV. Thirdly, homology modeling was performed with MOE-Homology, using the structure of HIV-1 PR in complex with APV (PDB code: 1HPV) as a template structure. During the modeling, the MMFF94x force field and the GB/VI implicit solvent function [23] were applied for energy calculation. In this study, we predicted ten structures of the complex, and selected the structure with the lowest energy as the model for the XMRV PR-APV complex.

3. Results

3.1. Synthesis of an enzymatically active XMRV PR using the wheat germ cell-free system

To synthesize enzymatically-active XMRV PR, we generated a transcription template of this enzyme derived from the XMRV VP62 clone. The template cDNA encodes the open reading frame of XMRV PR flanked by N-terminal 20 amino acid and a C-terminal 20 amino acid regions, as shown in Fig. 1A (PR; 20aa-PR-20aa). This PR differs from the synthesized inactive native PR template by an introduced substitution of the termination codon at the 3′-terminus of the Gag coding sequence for Glu-coding codon (CAG) to avoid translational termination at the end of Gag protein. As a catalytic-incompetent PR, we also designed a PR mutant harboring the catalytic active site substitution D32N (PR_D32N) (Fig. 1A). All 3 cDNA templates were subjected to cell-free transcription-translation. The protein yield of XMRV PR produced by this system was approximately 0.75 μM and the solubility was ~90%, as calculated by semi-quantitative CBB staining following SDS-PAGE (Fig. 1B). By immunoblotting (IB), two specific protein bands appeared at 14 kDa and 17 kDa, corresponding to the expected mobility of the full-length PR and the truncated form of PR by auto-cleavage of the flanking 20 a.a. at both ends (Fig. 1B). The auto-cleavage site of the XMRV PR was also confirmed by amino acid sequencing of the truncated protein band (Fig. 1C).

We next examined the enzymatic activity of the XMRV PR by monitoring its cleavage activity upon a native FLAG-pr55^{Gag} protein substrate. Wheat germ-synthesized XMRV PR was incubated with FLAG-pr55^{Gag} protein followed by immunoblotting analysis. p55^{Gag} was efficiently digested into the expected cleavage products (FLAG-MA-p12 and FLAG-MA) predicted from the known cleavage sites (Fig. 1D).

3.2. Evaluation of protease activity using AlphaScreen

For the quantitative and high-throughput measurement of XMRV PR activity using AlphaScreen technology, we designed a reporter substrate comprising a partial capsid (CA)-nucleo capsid (NC) junction peptide flanked by N-terminal FLAG and C-terminal biotin binding sequence (FLAG-CA/NC-biotin), as described in Materials and methods [14,15]. Fig. 2A shows a schematic representation of our assay system. Briefly, cell-free synthesized active XMRV PR, its D32N mutant or dihydrofolate reductase (DHFR) as a negative control, were incubated with the reporter substrate at 37 °C for 1 h, followed by the addition of AlphaScreen streptavidin donor and protein A acceptor beads as depicted in Fig. 2A. The cleavage of the reporter substrate was measured by level of luminescence (Fig. 2A). Wild-type XMRV PR, but not D32N_PR diminished the Alphascreen luminescent signal indicating proteolytic cleavage of the reporter polypeptide (Fig. 2B). The cleavage activity of PR was normalized relative to the luminescent activity of DHFR (Fig. 2C). Parallel immunoblot analysis with an anti-FLAG antibody demonstrated that the substrate protein was selectively cleaved by PR alone (Fig. 2D).

3.3. Screening of XMRV PR inhibitors by AlphaScreen

We next tested whether our assay system is applicable for drug screening targeting XMVR PR. As an initial approach, we examined the susceptibility of XMRV PR to six HIV-1 PIs: SQV (saquinavir), APV (amprenavir), IDV (indinavir), NFV (nelfinavir), DRV (darunavir) and LPV (lopinavir). Although all HIV-1 PIs tested showed marked inhibitory effects on HIV-1 PR, only two of them, APV and DRV, were found to block the activity of XMRV PR at the 1 μM concentration (Fig. 3A). This was also confirmed by IB of the blockade of cleavage of the reporter polypeptide containing the CA-NC junction (Fig. 3B). We next determined the IC_{50} value by titration of PIs (Fig. 3C). For XMRV PR, the IC_{50} values for APV, DRV, IDV and LPV were 0.2 μM, 1.0 μM, 60 μM and 17 μM, respectively. We next delineate the sensitivity of XMRV PR to APV in comparison with HIV-1 PR. Parallel experiment using recombinant HIV-1 PR and XMRV PR proteins revealed that IC_{50} values for APV was 34.7 nM in HIV PR and 200 nM in XMRV PR, respectively (Fig. 3D). These results indicate that this assay system can provide a tool to screen for selective PR inhibitors.

Retroviruses often exhibit drug resistant properties against anti-retrovirals due to their highly frequent genomic mutation. We next asked whether our assay system is useful for investigating drug-resistant properties of XMRV PR. To predict the sites of interaction between APV and XMRV PR, we modeled the three-dimensional (3D) complex of XMRV PR bound to APV. A recently published report on the crystal structure of XMRV PR shows that XMRV PR has a structural topology similar to that of HIV-1 PR [24]. Thus, we constructed our 3D structural model of the XMRV PR-APV complex by homology modeling, using the X-ray crystal structure of the HIV-1 PR-APV complex as a starting template (Fig. 4A). The constructed model indicates that APV interacts with aspartate Asp32 of the catalytic domain of XMRV PR, and also contacts the residues Val39, Lys61, Tyr90, and Leu92. Moreover, a water molecule would intermediate interactions between APV and Ala57 of the PR. A sequence alignment of PRs between XMRV and HIV-1 shows that the Val39, Ala57, Lys61, Tyr90, and Leu92 in XMRV PR are corresponding to the Val32, Ile50, Ile54, Val82, and Ile84 in HIV-1 PR, respectively (Fig.4B). These residues in HIV-1 PR are reported to be important for interactions between HIV-1 PR and APV and associated with viral resistance against APV [25].

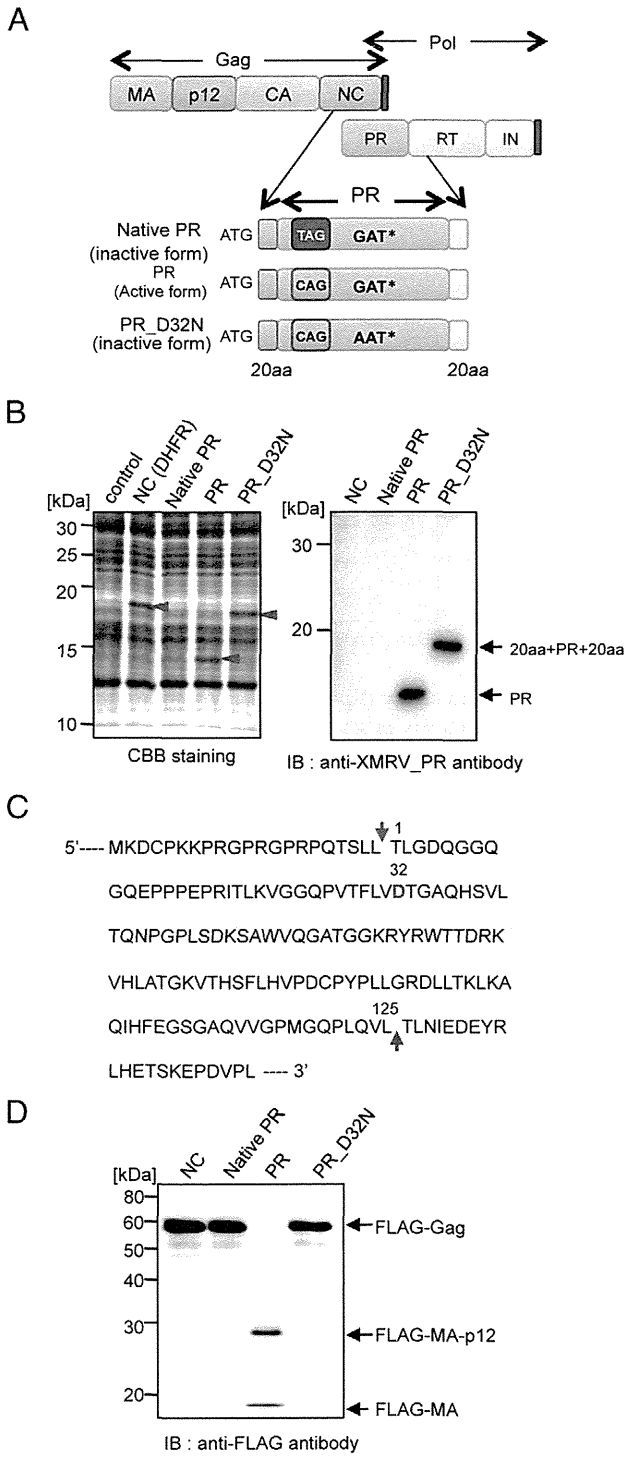
We then created selected site-directed mutants (V39I, K61L, A57V, V39I/A57V, Y90A/L92V) and investigated the catalytic activity and drug-resistant properties of these mutants to APV (Fig. 4C). As shown in Fig. 4C, V39I and A57V substitutions resulted in significantly ($P<0.01$) (Fig. 4C) higher drug resistance as compared with wild-type PR. This effect resulted in the 2.8-fold drug-resistance based on IC_{50} value (Fig.4D). These results indicate that our current assay system can predict drug-susceptibility of mutated proteases and may be useful for drug development targeting XMRV PR.

3.4. Identification of human proteins cleaved by XMRV PR

As it is known that viral proteases can cleave cellular proteins [26–28], we hypothesized that XMRV PR might be capable of digesting human proteins. As a representative demonstration

we selected twenty-four tumor suppressor proteins and synthesized them with N-terminal FLAG and C-terminal biotin tags by wheat cell-free system. These tester proteins were then incubated with XMRV PR followed by 2-color immunoblot analysis (Fig.5B). The result revealed that XMRV PR, but not DHFR as a negative control, can digest 4/24 tumor suppressor proteins examined: BAX, PTEN, DKK3 and ARL11 (Fig. 5C). Cleavage of the tumor suppressor proteins by XMRV PR was clearly inhibited by APV (Fig. 5D). For BAX and PTEN, the cleavage sites by XMRV PR were determined by peptide sequencing of the C-terminal cleavage products (Fig. 5E). The cleavage sites were found to be located in functional domains of both proteins, suggesting that proteolytic digestion by XMRV PR may diminish the native function of these tumor suppressor proteins by proteolytic digestion.

Since XMRV PR and HIV-1 PR have some similarity and can both be inhibited by APV, it is highly possible that XMRV PR can cleave the same substrates as HIV-1 PR. To this end, we tested whether XMRV PR could cleave two reported HIV-1 PR substrates, caspase-8 and NDR2 [28,29]. Interestingly, XMRV PR was found to digest caspase-8 although the cleavage site was distinct from that of HIV-1 PR. In contrast, XMRV PR was not able digest NDR2. Conversely, HIV-1 PR did not cleave Bax whereas XMRV PR can cleave it. Furthermore, both proteases could not cleave p53. These results indicate that there is certain substrate specificity of retroviral proteases toward host proteins (Fig. 6).



4. Discussion

In the current study, we developed a cell free protease assay with XMRV PR which can evaluate the cleavage activity via AlphaScreen or immunoblot analysis. We demonstrate the advantage of utilizing wheat cell-free system that was able to systematically produce catalytically active viral protease with a large amount for biochemical assays. Furthermore, our in vitro enzymatic assay revealed that APV is a potent inhibitor of XMRV PR. We have also delineated the physical interaction between APV and XMRV PR and identified the amino acid residues involved in the binding. Finally, we demonstrated the substrate specificity for XMRV PR as compared with HIV-1 PR. These results might reveal that our current assay system is a powerful tool to characterize viral proteases and to screen their specific inhibitors.

XMRV is a virus that was generated as the result of a unique recombination event between two endogenous MLV-like viruses in a nude mouse carrying the CWR22 prostate cancer xenograft [6]. Although XMRV is an unusual virus, XMRV has been associated with prostate cancer [2]. In fact, the human cell line 22Rv1, which was established from a human prostate tumor (CWR22), produces infectious XMRV particles [30]. While the absence of XMRV in non-prostatic tumors

Fig. 1 – Synthesis of enzymatically active XMRV protease (PR) by wheat cell-free protein production system. A. Construction of expression vector of XMRV PR for wheat cell-free synthesis. TAG (stop) codon between Gag and PR was substituted for CAG (Q) codon. Non-active form of PR was generated by the substitution of AAT (N) for catalytic center GAT (D). B. XMRV PRs (DHFR as a negative control) were separated by SDS-PAGE followed by CBB-stained (left panel) and immunoblotting using anti-XMRV PR antibody (right). The arrows depict protein products. C. Amino acid sequence of XMRV PR. The arrows indicate self-cleavage site in XMRV PR. D. Cleavage of XMRV Gag by XMRV PR produced by wheat cell-free system. XMRV PR was incubated with cell-free synthesized FLAG-tagged XMRV Gag (arrow), and the cleaved Gag was detected by immunoblot analysis with anti-FLAG antibody.

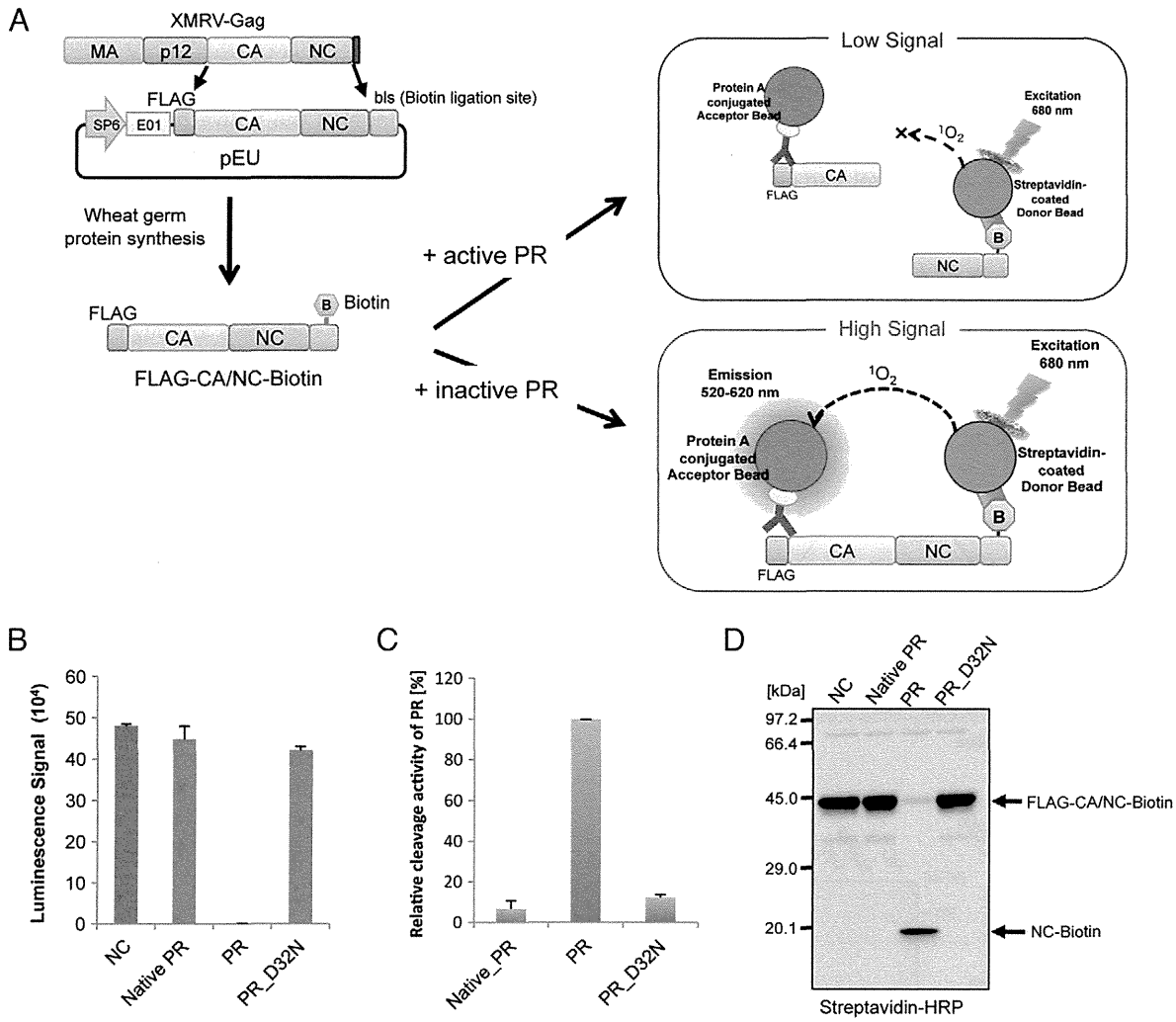


Fig. 2 – Development of a cleavage activity assay for XMRV PR using the luminescent assay AlphaScreen. A. Schematic diagram of the substrate construction of XMRV PR and detection system for the cleavage activity of XMRV PR by luminescent analysis. Substrate was designed as XMRV Gag capside (CA) and nucleocapside (NC) flanked by N-terminal FLAG and C-terminal biotin (FLAG_CA-NC_biotin). PR was incubated with the substrate, for 1 h at 37 °C. Subsequently, protein A-conjugated acceptor beads with anti-FLAG antibody and streptavidin coated donor beads were added and bound to the tagged substrate. Upon laser excitation, Donor beads convert ambient oxygen to a singlet oxygen. In the case of non-activity PR, singlet oxygen transfers across to activate Acceptor beads and subsequently emit light at 520–620 nm. In the case of active PR, no light is produced because the singlet oxygen can not transfer from Donor beads to Acceptor beads due to the distance (>200 nm). B,C,D. Cleavage activity of XMRV PR was quantitated by the luminescent assay (Fig.2B). Actual cleavage of XMRV Gag substrate was also confirmed by immunoblotting with streptavidin-HRP (Fig.2D). The arrow indicates the band for the non-cleaved substrates (FLAG_CA/NC_biotin).

remains controversial [31], XMRV can however proliferate in other human prostate cancer cells such as LNCaP or PC3 without severe cytopathic effects [32]. Such conditions of persistent infection without cell death could conceivably lead to prolonged exposure of host cell proteins to XMRV PR, increasing their susceptibility to cleavage with oncogenic consequences. The important question remains, however, as to whether this virus has indeed tumorigenic capability. Previous reports have indicated that XMRV integration is characterized by a strong preference for transcriptional start sites, CpG islands, and DNase-hypersensitive regions, all features that are frequently associated with structurally-open transcription regulatory regions of the chromosome in prostate cancer cells [33]. Integration of XMRV occurs preferentially in actively-transcribed genes and gene-dense regions within the chromosome [33]. Oncogenic properties of XMRV have been investigated in cell culture models. Although XMRV has been reported to lack direct transforming activity, the virus is able to induce low rates of transformation in cultured fibroblast cells [34]. Therefore, the molecular link between XMRV infection and cell transformation merits further investigation.

Our current data demonstrates that APV is a potent antagonist of XMRV PR. During the preparation of this manuscript, Li et al. reported the crystal structure of complexes