2α-*N*-Pivaloyl-1α-25(OH)$_2$ vitamin D$_3$ (IIIb). $[\alpha]_D^{23}$ = +37.5 (*c* 1.8, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.48–6.37 (m, 2H), 5.94 (d, *J* = 11.4 Hz, 1H), 5.35 (s, 1H), 5.08 (s, 1H), 4.33 (d, *J* = 3.2 Hz, 1H), 4.05–3.93 (m, 1H), 3.93–3.84 (m, 1H), 2.82 (d, *J* = 12.3 Hz, 1H), 2.65 (dd, *J* = 13.5, 4.4 Hz, 1H), 2.30 (t, *J* = 11.4 Hz, 1H), 2.05–0.80 (m, 39H), 0.52 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 180.8, 144.2, 132.1, 128.5, 125.8, 116.7, 116.4, 74.2, 71.1, 70.8, 57.8, 56.4, 56.3, 45.9, 44.3, 43.2, 40.4, 38.9, 36.3, 36.1, 29.5, 29.3, 29.2, 27.7, 27.6, 23.5, 22.2, 20.8, 18.8, 12.0; HRMS: calcd for $C_{32}H_{53}NO_4Na$, 538.3872; found, 538.3834.

2α-*N*-Benzoyl-1α-25(OH)$_2$ vitamin D$_3$ (IIIc). $[\alpha]_D^{23}$ = +26.2 (*c* 1.3, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.89–7.31 (m, 5H), 6.98 (d, *J* = 8.0 Hz, 1H), 6.46 (d, *J* = 11.4 Hz, 1H), 5.97 (d, *J* = 11.4 Hz, 1H), 5.38 (s, 1H), 5.12 (s, 1H), 4.45 (d, *J* = 3.4 Hz, 1H), 4.25–4.17 (m, 1H), 4.09–4.00 (m, 1H), 2.83 (t, *J* = 13.1 Hz, 1H), 2.71 (dd, *J* = 13.8, 4.6 Hz, 1H), 2.36 (t, *J* = 11.8 Hz, 1H), 2.23–0.80 (m, 30H), 0.53 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 169.1, 144.1, 132.5, 132.0, 131.8, 131.0, 128.5, 127.2, 125.8, 116.8, 74.3, 71.1, 70.4, 58.5, 56.4, 56.3, 45.9, 44.3, 43.2, 40.4, 36.3, 36.1, 29.4, 29.3, 29.1, 27.6, 23.5, 22.2, 20.8, 18.8, 12.1; HRMS: calcd for $C_{34}H_{49}NO_4Na$, 558.3559; found, 558.3540.

2α-*N*-*tert*-Butoxycarbonyl-1α-25(OH)$_2$ vitamin D$_3$ (IVd). $[\alpha]_D^{31}$ = +44.7 (*c* 0.8, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.44 (d, *J* = 11.0 Hz, 1H), 5.94 (d, *J* = 11.0 Hz, 1H), 5.35 (s, 1H), 5.07 (s, 1H), 4.35 (d, *J* = 2.7 Hz, 1H), 3.92–3.84 (m, 1H), 3.73–3.64 (m, 1H), 2.82 (d, *J* = 12.8 Hz, 1H), 2.67 (dd, *J* = 13.5, 4.8 Hz, 1H), 2.27 (t, *J* = 11.5 Hz, 1H), 2.06–0.72 (m, 30H), 1.46 (s, 9H), 0.52 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 157.2, 144.3, 144.1, 131.1, 128.0, 125.8, 116.7, 80.4, 74.7, 71.1, 70.5, 59.1, 56.4, 56.3, 45.9, 44.3, 43.2, 40.4, 36.3, 36.1, 29.7, 29.3, 29.2, 28.3, 27.6, 23.4, 22.2, 20.8, 18.8, 12.1; HRMS: calcd for $C_{32}H_{53}NO_5Na$, 554.3821; found, 554.3814.

2α-*N*-Methanesulfonyl-1α-25(OH)$_2$ vitamin D$_3$ (Va). $[\alpha]_D^{14}$ = +29.4 (*c* 0.5, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.45 (d, *J* = 11.5 Hz, 1H), 5.93 (d, *J* = 11.4 Hz, 1H), 5.39 (s, 1H), 5.12 (s, 1H), 5.01 (d, *J* = 8.3 Hz, 1H), 4.47 (s, 1H), 4.00–3.87 (m, 1H), 3.51–3.38 (m, 1H), 3.01 (s, 3H), 2.82 (d, *J* = 11.9 Hz, 1H), 2.70 (dd, *J* = 13.5, 4.3 Hz, 1H), 2.35–2.20 (m, 1H), 2.10–0.75 (m, 30H), 0.53 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 144.6, 143.9, 130.3, 126.0, 116.9, 116.5, 74.2, 71.1, 69.3, 61.9, 56.5, 56.3, 46.0, 44.4, 42.1, 41.4, 40.4, 36.3, 36.1, 29.7, 29.4, 29.3, 27.6, 23.5, 22.2, 20.8, 18.8, 12.1; HRMS: calcd for $C_{28}H_{47}NO_5SNa$, 532.3073; found, 532.3076.

2α-*N*-Benzenesulfonyl-1α-25(OH)$_2$ vitamin D$_3$ (Vb). $[\alpha]_D^{17}$ = +29.2 (*c* 0.8, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.94 (d, *J* = 7.8 Hz, 1H), 7.62 (t, *J* = 7.4 Hz, 1H), 7.55 (t, *J* = 7.6 Hz, 2H), 6.42 (d, *J* = 11.4 Hz, 1H), 5.87 (d, *J* = 11.5 Hz, 1H), 5.31 (d, *J* = 8.3 Hz, 1H), 5.15 (s, 1H), 5.02 (d, *J* = 1.3 Hz, 1H), 3.94 (s, 1H), 3.91–3.81 (m, 1H), 3.23 (ddd, *J* = 8.7, 8.7, 3.2 Hz, 1H), 2.79 (d, *J* = 12.8 Hz, 1H), 2.67 (dd, *J* = 13.6, 4.9 Hz, 1H), 2.32 (s, 1H), 2.20 (t, *J* = 11.7 Hz, 1H), 2.05–0.82 (m, 30H), 0.50 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 144.6, 143.7, 140.3, 133.0, 130.2, 129.3, 127.1, 126.1, 116.9, 116.5, 73.4, 71.1, 68.7, 62.0, 56.4, 56.3, 45.9, 44.3, 41.6, 40.4, 36.3, 36.0, 29.3, 29.2, 29.1, 27.6, 23.4, 22.1, 20.8, 18.8, 12.1; HRMS: calcd for $C_{33}H_{49}NO_5SNa$, 594.3229; found, 594.3271.

2α-*N*-(4-*tert*-Butyl)-benzenesulfonyl-1α-25(OH)$_2$ vitamin D$_3$ (Vc). $[\alpha]_D^{18}$ = +12.0 (*c* 0.4, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.85 (d, *J* = 8.6 Hz, 2H), 7.55 (t, *J* = 8.6 Hz, 2H), 6.42 (d, *J* = 11.5 Hz, 1H), 5.87 (d, *J* = 10.9 Hz, 1H), 5.24 (d, *J* = 8.6 Hz, 1H), 5.11 (d, *J* = 1.7 Hz, 1H), 5.01 (d, *J* = 1.7 Hz, 1H), 3.90 (d, *J* = 3.5 Hz, 1H), 3.86 (ddd, *J* = 13.7, 9.1, 4.6 Hz, 1H), 3.22 (dt, *J* = 8.6, 3.4 Hz, 1H), 2.79 (d, *J* = 12.6 Hz, 1H), 2.68 (dd, *J* = 13.8, 4.6 Hz, 1H), 2.21 (t, *J* = 11.8 Hz, 1H), 2.10–0.60 (m, 39H), 0.50 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 156.9, 144.6, 143.8, 137.1, 130.2, 127.0, 126.3, 126.2, 116.8, 116.5, 73.4, 71.1, 68.7, 61.9, 56.4, 56.3, 45.9, 44.4, 41.6, 40.4, 36.3, 36.1, 31.6, 31.1, 29.4, 29.2, 29.1, 27.6, 23.5, 22.1, 20.8, 18.8, 12.0; HRMS: calcd for $C_{37}H_{57}NO_5SNa$, 650.3855; found, 650.3858.

2α-*N*-(4-Methoxy)-benzenesulfonyl-1α-25(OH)$_2$ vitamin D$_3$ (Vd). $[\alpha]_D^{20}$ = +12.2 (*c* 0.5, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.86 (d, *J* = 9.1 Hz, 2H), 7.00 (d, *J* = 8.6 Hz, 2H), 6.42 (d, *J* = 11.5 Hz, 1H), 5.87 (d, *J* = 11.5 Hz, 1H), 5.22 (d, *J* = 8.6 Hz, 1H), 5.18 (s, 1H), 5.03 (d, *J* = 1.1 Hz, 1H), 3.99 (d, *J* = 3.5 Hz, 1H), 3.88 (s, 3H), 3.88–3.81 (m, 1H), 3.19 (dt, *J* = 8.6, 3.5 Hz, 1H), 2.79 (d, *J* = 13.2 Hz, 1H), 2.67 (dd, *J* = 13.2, 4.6 Hz, 1H), 2.20 (t, *J* = 11.8 Hz, 1H), 2.10–0.60 (m, 39H), 0.50 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 163.1, 144.6, 143.8, 131.7, 130.3, 129.3, 126.1, 116.9, 116.5, 114.4, 73.4, 71.1, 68.7, 61.9, 56.4, 56.3, 55.6, 45.9, 44.3, 41.6, 40.4, 36.3, 36.0, 29.3, 29.2, 29.1, 27.6, 23.5, 22.1, 20.8, 18.8, 12.0; HRMS: calcd for $C_{34}H_{51}NO_6SNa$, 624.3335; found, 624.3337.

2β-*N*,*N*′-Dibenzyl-1α-25(OH)$_2$ vitamin D$_3$ (VIIe). To a solution of **5** (40 mg, 0.0346 mmol) and **32** (42 mg, 0.0737 mmol) in toluene and triethylamine (1 : 1, 2 mL) was added $Pd(PPh_3)_4$ (about 50 mg) at room temperature. The resulting mixture was stirred at room temperature for 15 min, then 100 °C for 3 h. The resulting mixture was cooled to room temperature, and diluted with ethyl acetate, filtered through a pad of Celite, and concentrated *in vacuo*. The residue was purified by silica gel chromatography (hexane–ethyl acetate = 20 : 1) to give bis silylether. To a solution of bis silylether in THF (1.0 mL) was added HF–Et_3N (1.0 mL), and the mixture was stirred for 15 h. To the reaction mixture was added saturated $NaHCO_3$, and the mixture was stirred for 30 min. The resulting mixture was extracted with ethyl acetate, and the organic layer was washed with brine. The extracts were dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The residue was purified by silica gel chromatography ($CHCl_3$–MeOH = 9 : 1) to give **VIIe** (8.7 mg, 0.0142 mmol, 21%). $[\alpha]_D^{18}$ = −35.3 (*c* 0.1, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.33–7.24 (m, 10H), 6.34 (d, *J* = 10.8 Hz, 1H), 6.08 (d, *J* = 10.8 Hz, 1H), 5.46 (s, 1H), 4.94 (s, 1H), 4.52 (s, 1H), 4.44 (d, *J* = 10.8 Hz, 1H), 4.16 (d, *J* = 13.7 Hz, 2H), 3.66 (d, *J* = 13.7 Hz, 2H), 2.80 (dd, *J* = 3.4, 12.6 Hz, 1H), 2.56 (d, *J* = 10.8 Hz, 1H), 2.38–0.83 (m, 26H), 0.93 (d, *J* = 5.7 Hz, 3H), 0.53 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 143.7, 139.4, 131.7, 131.0, 128.8, 128.5, 127.1, 126.2, 116.9, 110.4, 71.0, 66.6, 66.0, 65.8, 56.4, 56.3, 54.2, 50.9, 45.9, 44.3, 40.3, 36.3,

36.0, 31.9, 29.6, 29.3, 29.2, 27.6, 23.7, 22.2, 20.7, 18.8, 11.8; HRMS: (ESI, M + H^+) calcd for $C_{41}H_{58}NO_3$, 612.4416; found, 612.4422.

2β-*N*-Substituted vitamin D derivatives VIIIa and Xa. As described for **VIIe**, **VIIIa** and **Xa** were obtained from the corresponding A-ring synthons (**33** and **34**), respectively.

2β-*N*-Acetyl-1α-25(OH)$_2$ vitamin D$_3$ (VIIIa). $[\alpha]_D^{20}$ = −15.6 (*c* 0.3, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.39–6.33 (m, 2H), 6.03 (d, *J* = 11.4 Hz, 1H), 5.59 (s, 1H), 5.09 (s, 1H), 4.13–4.04 (m, 2H), 3.96–3.89 (m, 1H), 2.80 (d, *J* = 10.8 Hz, 1H), 2.61 (d, *J* = 13.1 Hz, 1H), 2.15 (d, *J* = 14.3 Hz, 1H), 2.08 (s, 3H), 2.20–0.84 (m, 24H), 0.93 (d, *J* = 5.1 Hz, 3H), 0.54 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 172.4, 145.1, 144.1, 131.0, 125.9, 116.7, 112.2, 73.7, 71.0, 69.4, 58.4, 56.5, 56.3, 46.0, 44.3, 42.4, 40.3, 36.3, 36.0, 29.6, 29.3, 29.2, 27.5, 23.7, 23.3, 22.1, 20.7, 18.8, 11.8; HRMS: (ESI, M + Na^+) calcd for $C_{29}H_{47}NO_4Na$, 496.3402; found, 496.3405.

2β-*N*-Methanesulfonyl-1α-25(OH)$_2$ vitamin D$_3$ (Xa). $[\alpha]_D^{21}$ = −41.4 (*c* 0.3, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.35 (d, *J* = 10.9 Hz, 1H), 6.00 (d, *J* = 11.4 Hz, 1H), 5.52 (s, 1H), 5.24 (d, *J* = 8.24 Hz, 1H), 5.12 (s, 1H), 4.23–4.14 (m, 2H), 3.39–3.32 (m, 1H), 3.09 (s, 3H), 2.80 (d, *J* = 10.9 Hz, 1H), 2.57 (d, *J* = 14.6 Hz, 1H), 2.44 (d, *J* = 14.2 Hz, 1H), 2.09–0.84 (m, 24H), 0.93 (d, *J* = 5.9 Hz, 3H), 0.54 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 144.5, 144.1, 130.9, 125.7, 116.6, 112.3, 71.6, 71.1, 69.8, 62.6, 56.5, 56.3, 46.0, 44.3, 42.0, 41.4, 40.3, 36.3, 36.0, 29.6, 29.3, 29.2, 27.5, 23.7, 22.3, 20.7, 18.8, 11.8; HRMS: (ESI, M + Na^+) calcd for $C_{28}H_{47}NO_5SNa$, 532.3072; found, 532.3089.

Acknowledgements

This study was supported in part by the Advanced research for medical products Mining Programme of the National Institute of Biomedical Innovation (NIBIO) and Grant-in Aid for Scientific Research on Innovative Areas.

Notes and references

1 (*a*) S. A. Kliewer, K. Umesono, D. J. Mangelsdorf and R. M. Evans, *Nature*, 1992, **355**, 446; (*b*) D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon and R. M. Evans, *Cell*, 1995, **83**, 835; (*c*) S. Björklund, G. Almouzni, I. Davidson and K. P. Nightingale, *Cell*, 1999, **96**, 759; (*d*) K. Takeyama, Y. Masuhiro, H. Fuse, H. Endoh, A. Murayama, S. Kitanaka, M. Suzuwa, J. Yanagisawa and S. Kato, *Mol. Cell. Biol.*, 1999, **19**, 1049; (*e*) M. R. Haussler, C. A. Haussler, L. Bartik, G. K. Whitfield, J.-C. Hsieh, S. Slater and P. W. Jurutka, *Nutr. Rev.*, 2008, **66**, s98; (*f*) R. Bouillon, G. Carmeliet, L. Verlinden, E. V. Etten, A. Verstuyf, H. F. Luderer, L. Lieben, C. Mathieu and M. Demay, *Endocrine Rev.*, 2008, **29**, 726.

2 (*a*) T. Suda, T. Shinki and N. Takahashi, *Annu. Rev. Nutr.*, 1990, **10**, 195; (*b*) R. Bouillon, W. H. Okamura and A. W. Norman, *Endocrine Rev.*, 1995, **16**, 200; (*c*) H. F. DeLuca, *Am. J. Clin. Nutr.*, 2004, **80**, 1689S.

3 (*a*) E. Abe, C. Miyaura, H. Sakagami, M. Takeda, K. Konno, T. Yamazaki, S. Yoshiki and T. Suda, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 4990; (*b*) X. Wang and G. P. Studzinski, *J. Steroid Biochem. Mol. Biol.*, 2010, **97**, 395.

4 (*a*) M. W. Tilyard, G. F. S. Spears, J. Thomson and S. Dovey, *N. Engl. J. Med.*, 1992, **326**, 357; (*b*) M. F. Holick, *Am. J. Clin. Nutr.*, 2004, **79**, 362; (*c*) E. A. Slatopolsky, C. Weerts, J. Thielan, R. Horst, H. Harter and K. J. Marti, *J. Clin. Invest.*, 1984, **74**, 2136; (*d*) K. E. Lowe and A. W. Norman, *Nutr. Rev.*, 1992, **50**, 138.

5 (*a*) S. Ishizuka, K. Bannai, T. Naruuchi, Y. Hashimoto, T. Noguchi and N. Hosoya, *J. Biochem.*, 1980, **88**, 87; (*b*) E. Slatopolsky, J. Finch, C. Ritter, M. Denda, J. Morrissey, A. Brown and H. DeLuca, *Am. J. Kidney Dis.*, 1995, **26**, 852; (*c*) N. Saito, Y. Suhara, M. Kurihara, T. Fujishima, S. Honzawa, H. Takayanagi, T. Kozono, M. Matsumoto, M. Ohmori, N. Miyata, H. Takayama and A. Kittaka, *J. Org. Chem.*, 2004, **69**, 7463; (*d*) N. Kubodera, *Curr. Bioact. Compd.*, 2006, **2**, 301; (*e*) N. Kubodera, *Heterocycles*, 2010, **80**, 83.

6 M. Shimizu, Y. Miyamoto, H. Takaku, M. Matsuo, M. Nakabayashi, H. Masuno, N. Udagawa, H. F. DeLuca, T. Ikura and N. Ito, *Bioorg. Med. Chem.*, 2008, **16**, 6949.

7 A. Kittaka, Y. Suhara, H. Takayanagi, T. Fujishima, M. Kurihara and H. Takayama, *Org. Lett.*, 2000, **2**, 2619.

8 (*a*) Y. Suhara, K. Nihei, H. Tanigawa, T. Fujishima, K. Konno, K. Nakagawa, T. Okano and H. Takayama, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 1129; (*b*) Y. Suhara, K. Nihe, M. Kurihara, A. Kittaka, K. Yamaguchi, T. Fujishima, K. Konno, N. Miyata and H. Takayama, *J. Org. Chem.*, 2001, **66**, 8760; (*c*) S. Honzawa, Y. Suhara, K. Nihei, N. Saito, S. Kishimoto, T. Fujishima, M. Kurihara, T. Sugiura, K. Waku, H. Takayama and A. Kittaka, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 3503.

9 (*a*) K. Konno, S. Maki, T. Fujishima, Z. P. Liu, D. Miura, M. Chokki and H. Takayama, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 151; (*b*) K. Konno, T. Fujishima, S. Maki, Z. P. Liu, D. Miura, M. Chokki, S. Ishizuka, K. Yamaguchi, Y. Kan, M. Kurihara, N. Miyata, C. Smith, H. F. DeLuca and H. Takayama, *J. Med. Chem.*, 2000, **43**, 4247.

10 N. Rochel, J. M. Wurtz, A. Mitschler, B. B. Klaholz and D. Moras, *Mol. Cell*, 2000, **5**, 173.

11 (*a*) K. Miyamoto, E. Murayama, K. Ochi, H. Watanabe and N. Kubodera, *Biol. Pharm. Bull.*, 1993, **43**, 1111; (*b*) N. Tsugawa, K. Nakagawa, M. Kurobe, Y. Ono, N. Kubodera, K. Ozono and T. Okano, *Biol. Pharm. Bull.*, 2000, **23**, 66; (*c*) N. Kubodera and S. Hatakeyama, *Heterocycles*, 2009, **79**, 145.

12 (*a*) R. R. Sicinski, J. M. Prahl, C. M. Smith and H. F. DeLuca, *J. Med. Chem.*, 1998, **41**, 4662; (*b*) H. Yamamoto, N. K. Shevde, A. Warrier, L. A. Plum and H. F. DeLuca, *J. Biol. Chem.*, 2003, **278**, 31756; (*c*) H. F. DeLuca, *J. Steroid Biochem. Mol. Biol.*, 2004, **89–90**, 67; (*d*) H. Z. Ke, H. Qi, D. T. Crawford, H. A. Simmons, G. Xu, M. Li, L. Plum, M. Clagett-Dame, H. F. DeLuca, D. D. Thompson and T. A. Brown, *J. Bone Miner. Res.*, 2005, **20**, 1742.

13 (*a*) Y. Kato, Y. Hashimoto and K. Nagasawa, *Molecules*, 2003, **8**, 488; (*b*) Y. Kato, Y. Nakano, H. Sano, A. Tanatani, H. Kobayashi, R. Shimazawa, H. Koshino, H. Hashimoto and K. Nagasawa, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 2579; (*c*) Y. Nakano, Y. Kato, K. Imai, E. Ochiai, J. Namekawa, S. Ishizuka, K. Takenouchi, A. Tanatani, Y. Hashimoto and K. Nagasawa, *J. Med. Chem.*, 2006, **49**, 2398; (*d*) K. Cho, F. Uneuchi, Y. Kato-Nakamura, J. Namekawa, S. Ishizuka, K. Takenouchi and K. Nagasawa, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 4278.

14 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739.

15 Vitamin D$_3$ derivatives were docked into the X-ray structure of VDR (PDB code: 1DB1) using Glide 5.0 (Schrödinger, LLC). Hydrogen-bonding interaction between compounds and VDR and their visualization were generated by PyMOL (Schrödinger, LLC).

16 Lower values of docking score indicate more favorable interaction with VDR.

17 B. M. Trost, J. Dumas and M. Villa, *J. Am. Chem. Soc.*, 1992, **114**, 9836.

18 T. Yamamoto, H. Hasegawa, T. Hakogi and S. Katsumura, *Org. Lett.*, 2006, **8**, 5569.

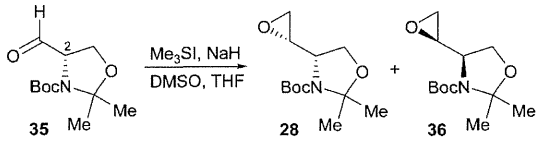
19 The diastereomer ratio of the alcohol was determined by ^{1}H NMR.

20 G. Ageno, L. Banfi, G. Cascio, G. Guanti, E. Manghisi, R. Riva and V. Rocca, *Tetrahedron*, 1995, **51**, 8121.

21 (*a*) W. J. Gensler, F. Johnson and A. D. B. Sloan, *J. Am. Chem. Soc.*, 1960, **82**, 6074; (*b*) H. H. Wasserman, M. Xia, J. Wang, A. K. Petersen and M. Jorgensen, *Tetrahedron Lett.*, 1999, **40**, 6163; (*c*) H. H. Wasserman, A. K. Peterson and M. Xia, *Tetrahedron*, 2003, **59**, 6771.

22 Synthesis of epoxide **28** *via* Corey–Chaikovsky reaction with Garner aldehyde has been reported.[23] We examined the reported procedure, but found that epimerization at C-2 took place under the reaction conditions, affording **28** and **36** as a 1 : 1 mixture. $[\alpha]_D^{26}$ = +1.18 (*c* 1.2, $CHCl_3$); reported $[\alpha]_D^{20}$ = +1.46 (*c* 1, $CHCl_3$). The method employed here

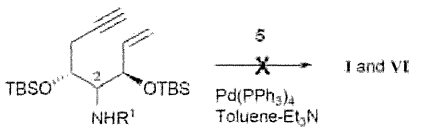
(Scheme 3) afforded **28** as a sole product. $[\alpha]_D^{22} = +12.5$ (*c* 1.1, $CHCl_3$).



23 (*a*) W. J. Moore and F. A. Luzzio, *Tetrahedron Lett.*, 1995, **36**, 6599; (*b*) C. Bedia, J. Casas, V. Garcia, T. Levade and G. Fabrias, *ChemBioChem*, 2007, **8**, 642.

24 X. Cong, F. Fu, K.-G. Liu, Q.-J. Liao and Z.-J. Yao, *J. Org. Chem.*, 2005, **70**, 4514.

25 Unfortunately coupling reaction with **5** did not take place in the case of *N*-monoalkyl A-ring synthons, and we could not obtain 2-*N*-monoalkyl 1,25-VD_3 derivatives **I** and **VI**.



R^1 = Me, Et, *n*-Pr, *n*-Bu, Bn

26 F. Jehan and H. F. Deluca, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 10138.

Virus Research 170 (2012) 109–117

Contents lists available at SciVerse ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Genetics and infectivity of H5N1 highly pathogenic avian influenza viruses isolated from chickens and wild birds in Japan during 2010–11

Yuko Uchida[a,b], Yasushi Suzuki[a], Masayuki Shirakura[c], Akira Kawaguchi[c], Eri Nobusawa[c], Taichiro Tanikawa[a], Hirokazu Hikono[a], Nobuhiro Takemae[a], Masaji Mase[a], Katsushi Kanehira[a], Tsuyoshi Hayashi[a], Yuichi Tagawa[a], Masato Tashiro[c], Takehiko Saito[a,b,*]

[a] *Viral Disease and Epidemiology Research Division, National Institute of Animal Health, National Agriculture and Food Research Organization (NARO), Kannondai, Tsukuba, Ibaraki 305-0856, Japan*

[b] *Zoonotic Diseases Collaboration Center (ZDCC), Kasetklang, Chatuchak, Bangkok 10900, Thailand*

[c] *Laboratory of Pandemic Vaccine Development, Influenza Virus Research Center, National Institute of Infectious Diseases, Gakuen 4-7-1, Musashimurayama-shi, Tokyo 208-0011, Japan*

ARTICLE INFO

Article history:
Received 28 June 2012
Received in revised form
11 September 2012
Accepted 12 September 2012
Available online 19 September 2012

Keywords:
H5N1 subtype highly pathogenic avian influenza
Chicken
Wild bird
Clade 2.3.2.1
PA gene segment
Infectivity

ABSTRACT

Outbreaks of H5N1 subtype highly pathogenic avian influenza virus (HPAIV) were recorded in chickens, domesticated birds and wild birds throughout Japan from November 2010 to March 2011. Genetic analysis of the Japanese isolates indicated that all gene segments, except the PA gene, were closely related to Japanese wild bird isolates in 2008 and belonged to clade 2.3.2.1 classified by the WHO/OIE/FAO H5N1 Evolution Working Group. Direct ancestors of the PA gene segment of all Japanese viruses analyzed in this study can be found in wild bird strains of several subtypes other than H5N1 isolated between 2007 and 2009. The PA gene of these wild bird isolates share a common ancestor with H5N1 HPAIVs belonging to clades 2.5, 7 and 9, indicating that wild birds were involved in the emergence of the current reassortant 2.3.2.1 viruses. To determine how viruses were maintained in the wild bird population, two isolates derived from chickens (A/chicken/Shimane/1/2010, Ck10 and A/chicken/Miyazaki/S4/2011, CkS411) and one from a wild bird (A/mandarin duck/Miyazaki/22M-765/2011, MandarinD11) were compared in their ability to infect and be transmitted to chickens. There was a significant difference in the survival of chickens that were infected with 10^6 EID_{50} of CkS411 compared to those with MandarinD11 and the transmission efficiency of CkS411 was greater than the other viruses. The increased titer of CkS411 excreted from infected chickens contributed to the improved transmission rates. It was considered that reduced virus excretion and transmission of MandarinD11 could have been due to adaptation of the virus in wild birds.

1. Introduction

H5N1 subtype highly pathogenic avian influenza viruses (HPAIVs) have spread to many Asian, European and African countries by natural and artificial means since an outbreak in geese in Guangdong province, China, in 1996 (World organization for animal health, Web portal on Avian Influenza: http://www.oie.int/animal-health-in-the-world/web-portal-on-avian-influenza/; Xu et al., 1999). The spread of HPAI infection can be attributed to several factors such as increased infection efficiency of the virus in several hosts through genetic reassortment, movement of poultry and poultry products, and migration of wild birds (Neumann et al., 2010; Vijaykrishna et al., 2008). The spread of HPAI in poultry has caused economic losses in affected countries and increased the threat of the emergence of a pandemic virus (Coker et al., 2011; Pongcharoensuk et al., 2011).

An FAO-OIE-WHO collaboration has been working towards reducing the risks associated with zoonotic infections of influenza viruses through animal, food and human sectors in preparation for a pandemic (The FAO-OIE-WHO Collaboration, http://www.oie.int/fileadmin/Home/eng/Current_Scientific_Issues/docs/pdf/FINAL_CONCEPT_NOTE_Hanoi.pdf#). The collaboration collects and updates information on HPAIVs and viruses are

* Corresponding author at: National Institute of Animal Health, National Agriculture and Food Research Organization (NARO), Kannondai, Tsukuba, Ibaraki 305-0856, Japan. Tel.: +81 29 838 7914; fax: +81 29 838 7914.
E-mail addresses: uchiyu@affrc.go.jp (Y. Uchida), yashisuzuki@affrc.go.jp (Y. Suzuki), masayuki@nih.go.jp (M. Shirakura), a-kawa@nih.go.jp (A. Kawaguchi), nobusawa@nih.go.jp (E. Nobusawa), ttanikawa@affrc.go.jp (T. Tanikawa), hhikono@affrc.go.jp (H. Hikono), ntakemae@affrc.go.jp (N. Takemae), masema@affrc.go.jp (M. Mase), kanehira@affrc.go.jp (K. Kanehira), tsuyoh@affrc.go.jp (T. Hayashi), ytagawa@affrc.go.jp (Y. Tagawa), mtashiro@nih.go.jp (M. Tashiro), taksaito@affrc.go.jp (T. Saito).

0168-1702/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2012.09.004

classified into twelve clades based on phylogeny of their HA genes (WHO, 2011b). These clades are further classified into several small subclades because HPAIVs have evolved through maintenance in the poultry and wild bird populations, in part, and possibly due to vaccination of domestic birds in particular regions (Cattoli et al., 2011; Li et al., 2010; WHO/OIE/FAO H5N1 Evolution Working Group, 2012).

A virus of clade 2.2 caused mortalities in wild birds at Qinghai Lake in May 2005 and has spread to Europe and Africa (Chen et al., 2005; Liu et al., 2005). Another virus of the same clade was confirmed at Qinghai Lake between 2006 and 2008, but there was no genetic diversity in the isolates at that time. Subsequently, a virus classified as clade 2.3.2 was isolated from dead birds at the lake between May and June of 2009. The emergence of this virus indicated a possible adaptation to wild birds and the virus spread widely, as the clade 2.2 virus had previously (Li et al., 2011). Intensive surveillance of wild birds in Hong Kong has been performed following outbreaks in wild birds and poultry in 2002–03. Surveillance of dead wild birds revealed the two antigenically distinct groups of viruses that belong to either clade 2.3.4, which was established in Asian poultry, or clade 2.3.2, which is a group of viruses that have been circulating in wild birds and poultry since 2008. Clade 2.3.2 viruses have been isolated from poultry and wild birds since 2007 (Smith et al., 2009). A human case in 2010 in Hong Kong was also caused by a clade 2.3.2 virus (WHO, 2011a). HPAIVs of clade 2.3.2 were isolated from wild birds in Mongolia in 2008 and 2009. The HPAIVs in 2009 were distinguished from those in 2008 by the origin of the PA gene segment (Kang et al., 2011).

During the winter of 2010–11 in Japan, there were 24 HPAI outbreaks in chickens, 60 cases in wild birds and 3 cases in domesticated birds (Sakoda et al., 2012). Before these outbreaks in October 2010, an HPAIV belonging to clade 2.3.2.1 was first isolated from wild duck feces in Hokkaido in the northern part of Japan. The incidence of massive viral infection in birds all over Japan suggested that waterfowl, a natural reservoir of type A influenza virus, could play a role as virus disseminators after maintaining HPAIVs during nesting (Kajihara et al., 2011).

The virulence of an HPAIV derived from wild birds decreased in chickens and the mortality of chickens experimentally infected with HPAIV isolate from tree sparrows in Thailand in 2005 was 70% (Hayashi et al., 2011). These results suggest that the maintenance of HPAIVs within wild birds could affect the pathogenicity of a virus in chickens. In this study, characteristics such as genetics, antigenicity, infectivity and transmissibility to chickens of HPAIVs isolated from outbreaks in either wild birds or chickens between 2010 and 2011 in Japan were examined to understand the possible implications of their maintenance in wild birds. When the infectivity and transmissibility of these HPAIVs were compared in chickens, it appeared that a wild bird isolate used in this study had become attenuated. This study will contribute to the available information on HPAI infection and transmission from wild birds to domestic chickens.

2. Materials and methods

2.1. Virus isolation and characterization

Tracheal swabs, cloacal swabs or internal organ homogenates from chickens suspected of HPAIV infection or dead chickens, domesticated or wild birds were used for virus isolation. Specimens were inoculated into the allantoic cavities of 10-day-old embryonated chicken eggs and incubated at 37 °C for about 30 h at each municipal animal health laboratory. After the allantoic fluid was collected, the specimens were sent to the National Institute of Animal Health for subtype identification by the hemagglutination inhibition (HI) test and the neuraminidase inhibition (NI) test with a panel of antisera (Table 1).

Two chicken viruses, A/chicken/Shimane/1/2010 (Ck10), an index strain from the 2010–11 poultry outbreaks and A/chicken/Miyazaki/S4/2011 (CkS411) and a wild bird strain, A/mandarin duck/Miyazaki/22M-765/2011 (MandarinD11) were used in animal experiments.

2.2. Genetic and phylogenetic characterization of HPAIVs from 2010–11

The DNA sequences of the HA genes of A/chicken/Miyazaki/M6/2011, A/chicken/Aichi/T1/2011 and A/chicken/Mie/1/2011 have been published previously (GenBank: AB675739–AB675741) (Sakoda et al., 2012). In this study, the HA gene of the remaining 24 strains (21 chicken, 2 domesticated bird and 1 wild bird), the NA gene of all strains (24 chicken, 2 domesticated and 1 wild bird) and the internal genes of 22 strains (19 chicken, 2 domesticated and 1 wild bird) were sequenced (GenBank: AB684081–AB684263) as described previously (Uchida et al., 2008). Phylogenetic analysis of the amino acid sequences obtained in this study was carried out with BioEdit software (Hall, 1999) and MEGA 4.0 (Tamura et al., 2007) using the neighbor-joining method. Further phylogenetic analysis with other algorithms, such as minimum evolution method, unweighted pair-group method with an arithmetic mean and the maximum parsimony method were performed on the PA gene segment. Bootstrap value was calculated at 1000 replicates.

Putative amino acid substitutions found in Ck10, CkS411 and MandarinD11 relative to A/common magpie/HK/5052/2007 (CmHK07) were illustrated on a three-dimensional structure of the HA protein (Protein Data Bank ID: 2IBX) by PyMOLver. 0.99 (PyMOL, http://pymol.sourceforge.net/).

2.3. Antigenic analysis

The antigenic characteristics of viruses isolated in 2010–11 in Japan were analyzed by the HI test described in the WHO manual on animal influenza diagnosis and surveillance using antiserum against HPAIVs. Antisera against Ck10, A/whooper swan/Akita/1/2008 (WsAkita08) and A/chicken/Yamaguchi/7/2004 (Ck04) were produced by immunizing chickens with formalin-inactivated virus, and ferret antiserum against CmHK07 was kindly provided by Dr. R. Webby, St. Jude Children's Research Hospital, Memphis, TN, USA.

2.4. Animal experiments

Four-week-old specific pathogen-free (SPF) white leghorn chickens, L-M-6 strain, were obtained from Nisseiken Co., Ltd. (Tokyo, Japan). All animal experiments were carried out in biosafety level 3 facilities at the National Institute of Animal Health, Japan and were approved by the Ethics Committee of the institute.

To investigate the infectivity and pathogenicity of each virus, 10^6, 10^4 or 10^2 EID_{50}/100 μl of each virus diluted in PBS were inoculated intranasally into five chickens. Animals were observed for 10 days post inoculation and survival rate and period of infection were recorded for the survival analysis. Tracheal and cloacal swabs taken at 3, 5, 7 and 10 days post inoculation (DPI) or at death were collected and dipped into MEM containing 0.5% BSA, 25 μg/ml of Fungizone, 1000 units/ml of penicillin and 1000 μg/ml of streptomycin, 0.01 M HEPES and 8.8 mg/ml of $NaHCO_3$. Swabs were removed from the MEM and stored at −80 °C until titration. Frozen samples were thawed and centrifuged at 3000 rpm for 5 min

Table 1
List of H5N1 subtype HPAIVs isolated from poultry, domesticated bird and wild bird in Japan during November 2010 to March 2011.

Strain name	Place of isolation	Virus isolated host	Date of specimen collection (mm/dd/yy)
Ck/Shimane/1/2010	Shimane	Layer	11/29/2010
Mute swan/Toyama/1/2010	Toyama	Domesticated bird	12/16/2010
Ck/Miyazaki/M6/2011	Miyazaki	Breeding chicken	1/21/2011
Ck/Miyazaki/S4/2011	Miyazaki	Layer	1/23/2011
Ck/Kagoshima/I-2/2011	Kagoshima	Layer	1/25/2011
Ck/Aichi/T1/11	Aichi	Layer	1/26/2011
Ck/Miyazaki/T10/2011	Miyazaki	Broilar	1/27/2011
Ck/Miyazaki/K3/2011	Miyazaki	Broilar	1/28/2011
Ck/Miyazaki/N7/2011	Miyazaki	Broilar	1/28/2011
Ck/Miyazaki/TA3/2011	Miyazaki	Broilar	1/30/2011
Ck/Miyazaki/MT2/2011	Miyazaki	Broilar	2/1/2011
Ck/Oita/1/2011	Oita	Layer	2/2/2011
Ck/Miyazaki/8/2011	Miyazaki	Broilar	2/4/2011
Ck/Miyazaki/9/2011	Miyazaki	Broilar	2/5/2011
Ck/Miyazaki/10/2011	Miyazaki	Broilar	2/5/2011
Ck/Miyazaki/11/2011	Miyazaki	Broilar	2/6/2011
Black swan/Yamaguchi/1/2011	Yamaguchi	Domesticated bird	2/9/2011
Ck/Aichi/2/2011	Aichi	Broilar	2/14/2011
Ck/Wakayama/1/2011	Wakayama	Layer	2/15/2011
Ck/Mie/1/2011	Mie	Broilar	2/15/2011
Ck/Miyazaki/12/2011	Miyazaki	Broilar	2/16/2011[a]
Mandarin duck/Miyazaki/22M-765/2011	Miyazaki	Wild bird	2/17/2011
Ck/Mie/2/2011	Mie	Layer	2/26/2011
Ck/Nara/1/2011	Nara	Layer	2/28/2011
Ck/Miyazaki/13/2011	Miyazaki	Broilar	3/5/2011
Ck/Chiba/1/2011	Chiba	Layer	3/12/2011
Ck/Chiba/2/2011	Chiba	Broilar	3/14/2011

[a] Date of virus isolation (mm/dd/yy).

at 4 °C. The supernatant was subjected to viral titration by EID_{50}. To confirm viral infection in the surviving chickens, blood samples taken at the end of the observation period were used to detect antibodies against the influenza A virus by FLOCKTYPE recAIV Screening ELISA (VERITAS, Tokyo, Japan).

Viral transmission was assessed by intranasal inoculation of one chicken with 10^6 EID_{50}/100 μl of each virus housed together with four naïve chickens in the same isolator. The observation period was 10 days from the date of the first inoculation and feed and water were shared between the inoculated and naïve chickens. The survival period and rate were recorded to draw a survival curve. Tracheal and cloacal swabs were collected and virus titers were determined as described above.

2.5. Survival analysis

Survival analysis was carried out using the results of survivability in the 10^6, 10^4 and 10^2 EID_{50} inoculation groups and the naïve chickens in the transmission test. A Kaplan–Meier survival curve (Kaplan and Meier, 1958) was constructed using the survival rate and infection period for chickens in each titer group. Differences in the Kaplan–Meier survival curves were analyzed by log-rank test under Bonferroni correction.

3. Results

3.1. Genetic characterization of the Japanese isolates

Twenty four HPAIVs isolated in Japan from chickens from individual outbreaks during 2010–11 and viruses from domesticated birds (2 strains) and a dead wild bird (1 strain) were sequenced and the HA and NA gene segments were phylogenetically analyzed. The deduced amino acid sequence at the HA cleavage site was PQRERRRKR for all isolates except A/chicken/Miyazaki/T10/2011, for which the sequence was PQREKRRKR. All of the Japanese isolates from this period formed a cluster (Fig. 1A cluster A) within clade 2.3.2.1 (Sakoda et al., 2012) at 87% bootstrap values in the HA gene phylogenetic tree and were clearly distinguishable from clade 2.3.2.1 strains isolated in Japan during 2008 from wild birds (Fig. 1A cluster B). Cluster A also included a Mongolian virus isolated in 2010 (A/whooper swan/Mongolia/21/2010) and wild bird isolates from both Japan and South Korea isolated during 2010–11. An isolate, A/hooded crane/Kagoshima/4612J008/2010, from an outbreak in cranes which occurred within 5 km of the farm where A/chicken/Kagoshima/I-2/2011 was isolated, formed a branch with it at 85% bootstrap values. A similar topology was observed for the NA and internal genes except the PA gene of cluster A strains.

The genetic origin of the cluster A strain PA genes was scrutinized by phylogenetic analysis (Fig. 1B). The direct ancestors of the PA gene segment of cluster A strains were found in wild bird strains of several subtypes isolated between 2007 and 2009. H3N8, H4N6, H4N9, H5N2, H5N5, H6N2, H6N8, H10N8 and H11N2 avian influenza viruses formed a cluster in the PA tree as the genetic origin of the PA gene of cluster A strains. These avian influenza viruses and the HPAIVs of clade 2.3.2.1 isolated in 2010–11 shared a common ancestor to HPAIVs of clades 2.5, 7 and 9.

3.2. Antigenic analysis of the Japanese isolates

Inter-clade and intra-clade antigenic variation was demonstrated by HI tests using Ck10, A/chicken/Miyazaki/11/2011 (Ck1111), CmHK07, WsAkita08 isolated in 2008 and Ck04 (clade 5) as antigens (Table 2). Although anti-Ck10 serum reacted well with Ck10, Ck1111 and WsAkita08, the reaction with CmHK07 was 16 fold lower compared to Ck10, although both viruses belong to clade 2.3.2.1. In addition, anti-CmHK07 showed a titer that was 8 fold lower than Ck10 compared to CmHK07. One of the pandemic vaccine candidates, SJRG-166615, from clade 2.3.2.1 was derived from CmHK07 and the HI data indicates that the 2010–2011 Japanese HPAIVs of clade 2.3.2.1 were antigenically distinguishable from this.

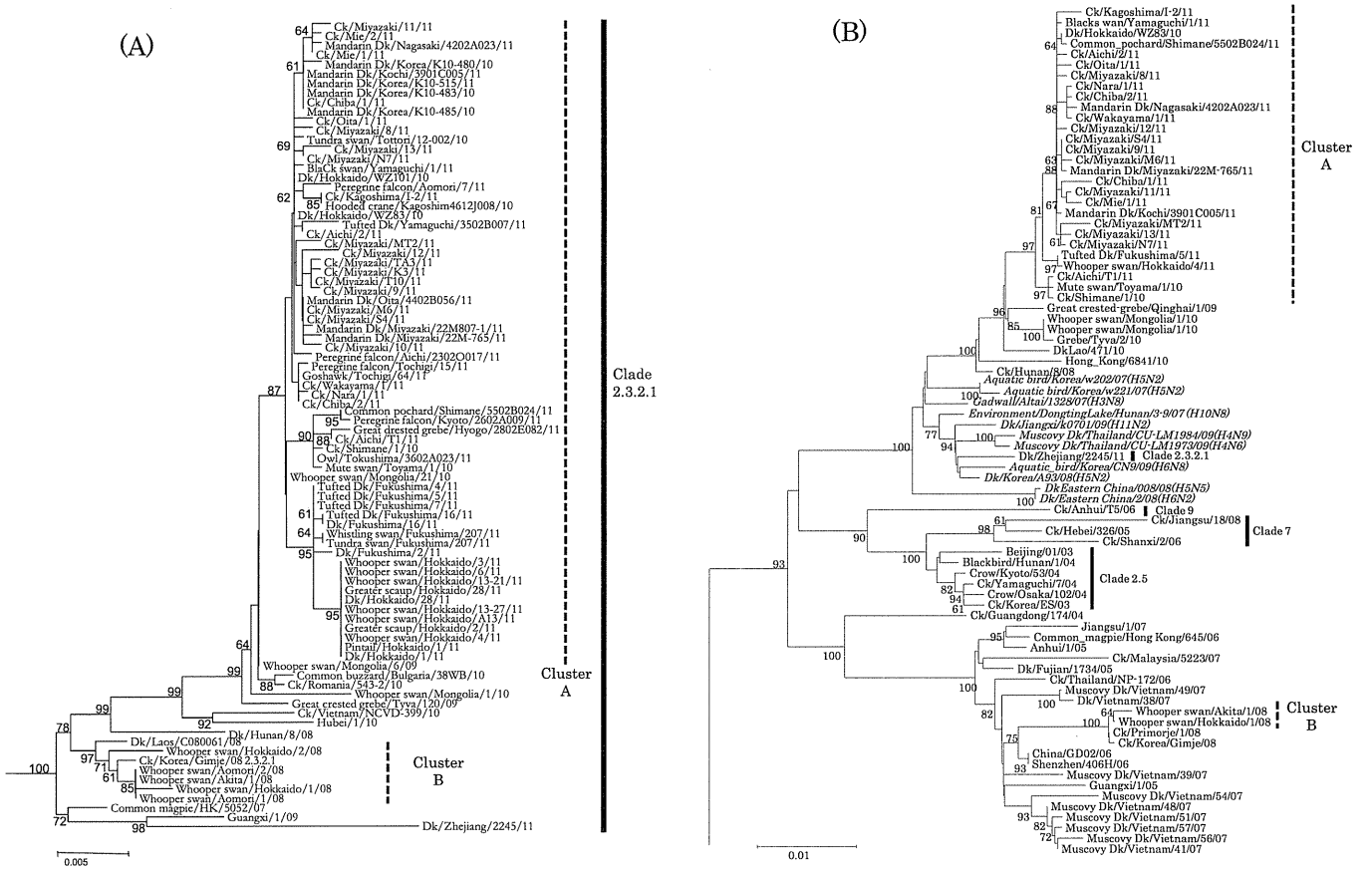


Fig. 1. Phylogenetic tree of HA and PA gene. (A) *Phylogenetic tree of the HA gene*: Phylogenetic tree of the HA gene of the 2010–11 Japanese isolates was constructed by the neighbor-joining method with the sequences published in GenBank. The nucleotide length of the HA gene used for the analysis was 952 base pairs. The strain used to root this tree was A/goose/Guangdong/1/1996. Bootstrap values of more than 60 are shown at each branch. The length of branches represents the distance between sequences and the bar and numbers at the bottom of the tree indicate the number of substitutions. Isolated host and year in the strain names were abbreviated as follows: chicken; Ck, duck; Dk, quail; Qa, 2011; 11. (B) *Phylogenetic tree of PA gene*: Phylogenetic tree of PA gene was constructed by the neighbor-joining methods. PA gene sequences of the H5N1 subtype HPAIVs and other subtype strains published in GenBank are shown as a representative of phylogenetic analysis of the PA gene. Topology of the tree was similar when other calculating methods, such as the minimum evolution method, the unweighted pair-group method with arithmetic mean and the maximum parsimony method, were applied. The nucleotide length of the PA gene used for the analysis was 1975 base pairs. Root strain, drawn bootstrap values and host and year of abbreviation of the isolated strain in the tree were the same as the HA tree. Viruses written in italics mean other subtypes of avian influenza viruses except H5N1 subtype HPAIVs.

3.3. Comparison of infectivity and pathogenicity of isolates derived from chickens and a wild bird

CkS411, a chicken isolate from 2010/11 outbreaks, and a wild bird isolate, MandarinD11, differed in their lethality to chickens when five chickens of each group were inoculated with 10^6 EID_{50} viruses (Fig. 2A). Survival analysis showed a significant difference between the viruses. The mean death time (MDT) was 51.6 hpi for CkS411 and 75.6 hpi for MandarinD11 infected chickens, although all of the infected chickens eventually died. All of the chickens inoculated with 10^4 EID_{50} of Ck10 or MandarinD11 died (Fig. 2B), while 60% of those infected with CkS411 survived and were negative for anti-influenza A virus antibodies by ELISA (data not shown). The MDTs of this infectious dose of CkS411, Ck10 and MandarinD11 were 170.4, 144 and 148.8 hpi, respectively. These values were more than twice the MDTs obtained with 10^6 EID_{50} of each virus. All five chickens infected with 10^2 EID_{50}/100 μl of Ck10 or MandarinD11 virus survived (data not shown). One of five chickens infected with 10^2 EID_{50}/100 μl CkS411 virus died, although no virus was detected in the tracheal or cloacal swabs collected at the time of death. Since the cause of death did not appear to be due to the viral infection, the data for this chicken was excluded from

Table 2
Antigenic characterization of viruses isolated in Japan in 2010–11 by the HI test using hyper immune antisera.

	Hyper immune antisera			
	Chicken		Ferret	Chicken
	Clade 2.3.2.1 Ck/Shimane/1/2010	Clade 2.3.2.1 Ws/Akita/1/2008	Clade 2.3.2.1 Common magpie/HK/5052/2007	Clade 2.5 Ck/Yamaguchi/7/2004
A/chicken/Shimane/1/2010	**1280**	640	20	160
A/chicken/Miyazaki/11/2011	640	640	20	160
A/whooper swan/Akita/1/2008	640	**1280**	40	160
A/common magpie/HK/5052/2007	80	2560	**160**	160
A/chicken/Yamaguchi/7/2004	160	1280	20	**1280**

Boldface shows HI titer used with homogenous antigen and antisera.

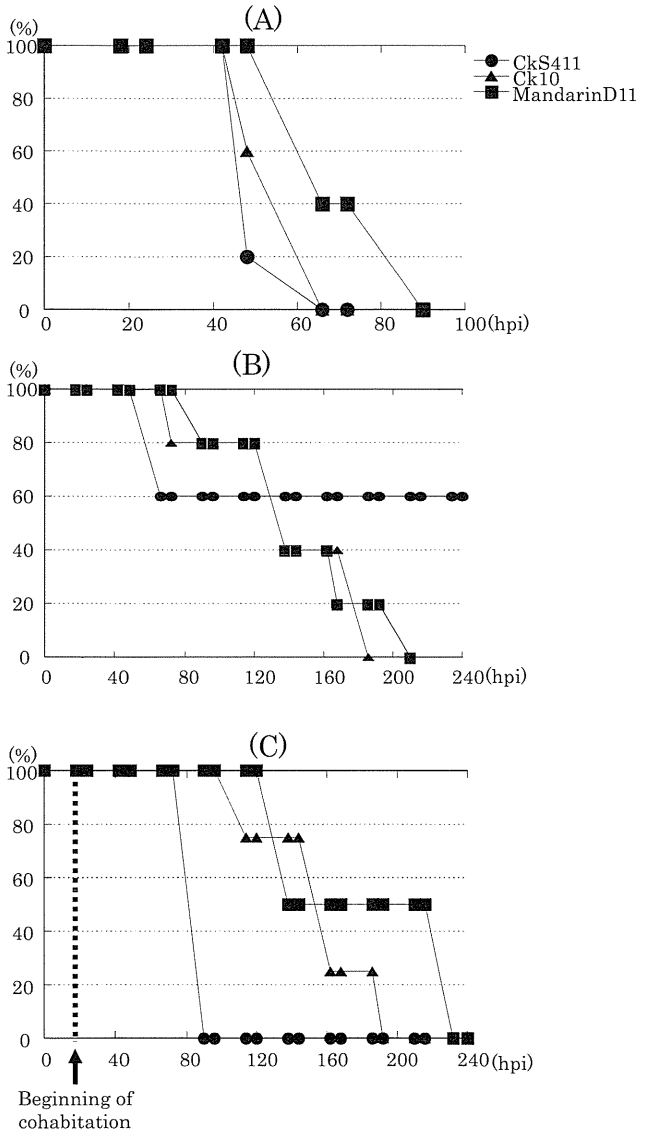


Fig. 2. *The survival curves of the 10^6 (A) and 10^4 (B) viral infection and transmission test (C).* In each figure, circle (●), triangle (▲) and square (■) represent a data set obtained following infection with CkS411, Ck10 and MandarinD11, respectively. The *x*-axis is hours post inoculation (hpi) and the *y*-axis is the survival rate (%). (C) The survival curves of the chickens cohabiting with the chicken infected with either CkS411 (●), Ck10 (▲) or MandarinD11 (■), respectively, are shown by broken lines. Cohabitation with the infected chicken started at 18 hpi.

all of the analyses in this study. There were no significant differences in the survival rates observed for either 10^4 or 10^2 EID_{50} infectious doses. Fifty percent chicken lethal doses (CLD_{50}/ml) were calculated as $10^{4.3}$ EID_{50} for CkS411 and 10^3 EID_{50} for the others.

Viral titers in the tracheal and cloacal swabs from chickens infected with 10^6 EID_{50} of MandarinD11 were lower than in chickens infected with chicken isolates. Significant differences were observed in the viral titers in the tracheal swabs from chickens infected with 10^6 EID_{50} Ck10 compared to MandarinD11 ($p = 0.0098$) (Fig. 3A). Viral load in cloacal swabs from birds infected with CkS411 and Ck10 was significantly higher ($p = 0.0013$ and 0.019) than in MandarinD11 swabs (Fig. 3B). Viral titers in chickens infected with CkS411 and Ck10 reached $>10^{5.1}$ EID_{50}/ml in the tracheal swabs and $10^{3.1}$ EID_{50}/ml in the cloacal swabs shortly before death. The maximum viral titers of chickens infected with MandarinD11 were $10^{4.05}$ EID_{50}/ml and $10^{1.13}$ EID_{50}/ml in the trachea and cloaca, respectively.

There was no difference in the maximum viral titers in the tracheal or cloacal swabs collected on 3, 5, 7 and 10 dpi or at death in the chickens infected with 10^4 EID_{50}. Viral excretion into the trachea and cloaca occurred later in these groups compared to that observed in the 10^6 EID_{50} inoculation groups (Fig. 3C and D).

No virus was isolated from tracheal or cloacal swabs from any of the chickens inoculated with 10^2 of any of the three viruses during the observation period. Anti-influenza A virus antibody was not detected by ELISA in any of the surviving chickens indicating that they were not infected with the influenza virus (data not shown).

3.4. Transmission of virus isolates to chickens

The transmission study demonstrated that CkS411 was more readily transmitted to uninfected chickens compared to the other two isolates. Chickens infected with 10^6 EID_{50} of CkS411 or Ck10 died at 48 hpi whilst the chicken infected with the same dose of MandarinD11 died at 66 hpi. The MDT of the naïve chickens was 72 h after cohabitation (hac) for CkS411, 139.5 hac for Ck10 and 168 hac for MandarinD11 (Fig. 2C). Survival analysis of the groups of cohabiting chickens revealed significant differences between CkS411 compared to Ck10 and MandarinD11 groups. CkS411 replicated well in both inoculated and naïve chickens and was transmitted with greater efficiency (Fig. 4A and B). In the inoculated chicken in the CkS411 group, tracheal and cloacal titers were $10^{6.53}$ and $10^{5.02}$ EID_{50}/ml, and mean max titers of the cohabiting chickens were $10^{6.24}$ and $10^{4.73}$ EID_{50}/ml, respectively. These values in the MandarinD11 infected chickens were low. The efficiency of viral transmission is correlated with the titer of virus excreted from an inoculated chicken. The mean time of the first viral isolation in the trachea or cloaca of cohabiting chickens was 48, 108 and 108 hac in CkS411, Ck10 and MandarinD11 groups, respectively. The time of viral isolation in CkS411 group was significantly shorter than that of Ck10 and MandarinD11.

3.5. Comparison of amino acid substitutions between the isolates

Animal experiments indicated that the replication and transmission characteristics of Ck10 and MandarinD11 viruses were different. Deduced amino acid sequences were compared to explore whether any signature substitutions relating to such phenotypical differences could be found. Only three amino acid substitutions in the HA amino acid sequences were identified between the three viruses. When compared with the sequence of MandarinD11, one substitution at 83 aa (87 aa in H3 numbering) in CkS411 and two substitutions at 84 and 259 aa (88 and 263 aa in H3 numbering) in Ck10 were found. As a result, the amino acid sequences in Ck10 tended to be different from those of the other two viruses. Several amino acid substitutions that were seen in other segments of the viruses did not coincide with amino acids changes that have previously been reported to correlate with pathogenicity (Table 3). Amino acid substitutions in the HA gene of the three viruses compared to CmHK07 are shown in the HA three-dimensional structures (Fig. 5). Five amino acid substitutions from CmHK07were commonly found among the three viruses at antigenic sites A, B, D and E. They were at 66 aa (75 aa in H3 numbering) for site E, 133 and 136 aa (137 and 140 aa in H3 numbering) for site A, 154 aa (158 aa in H3 numbering) for site B and 240 aa (244 aa in H3 numbering) for site D. These substitutions appeared to correlate with antigenic differences between Ck10 and CmHK07. No amino acid substitutions were found within the antigenic sites of CkS411, Ck10 or MandarinD11.

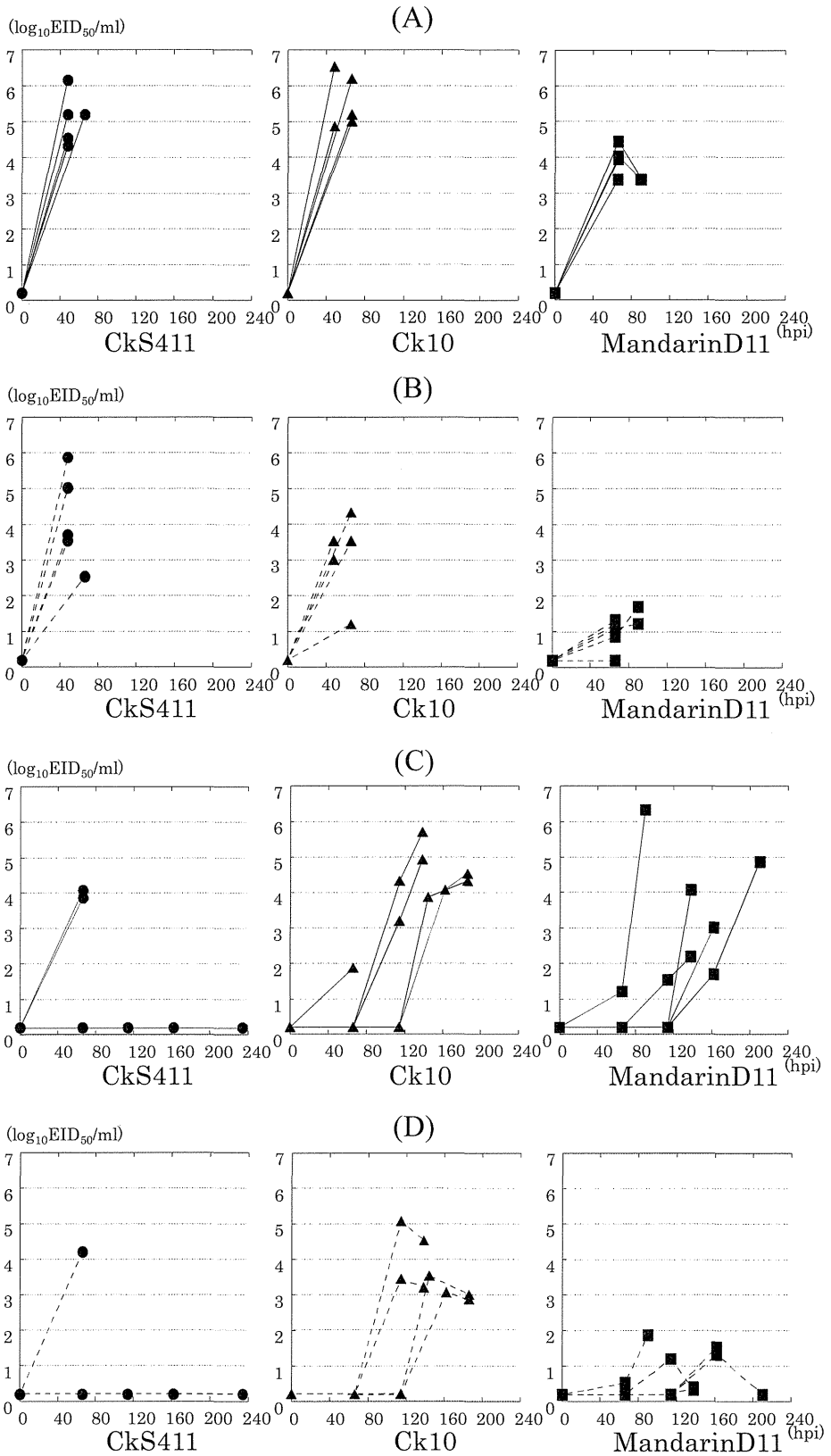


Fig. 3. *The kinetics of viral multiplication in the trachea (A, C) and cloaca (B, D) of infected chickens.* Circle (●), triangle (▲) and square (■) represent a data set obtained during the animal experiment with CkS411, Ck10 and MandarinD11, respectively. The solid or broken lines indicate the results in the trachea or cloaca. The *x*-axis is hours post inoculation (hpi) and the *y*-axis is 50% egg infective dose ($\log_{10}$ EID$_{50}$/ml). Data obtained following infection with 10^6 EID$_{50}$ (A, B) and 10^4 EID$_{50}$ (C, D) are shown.

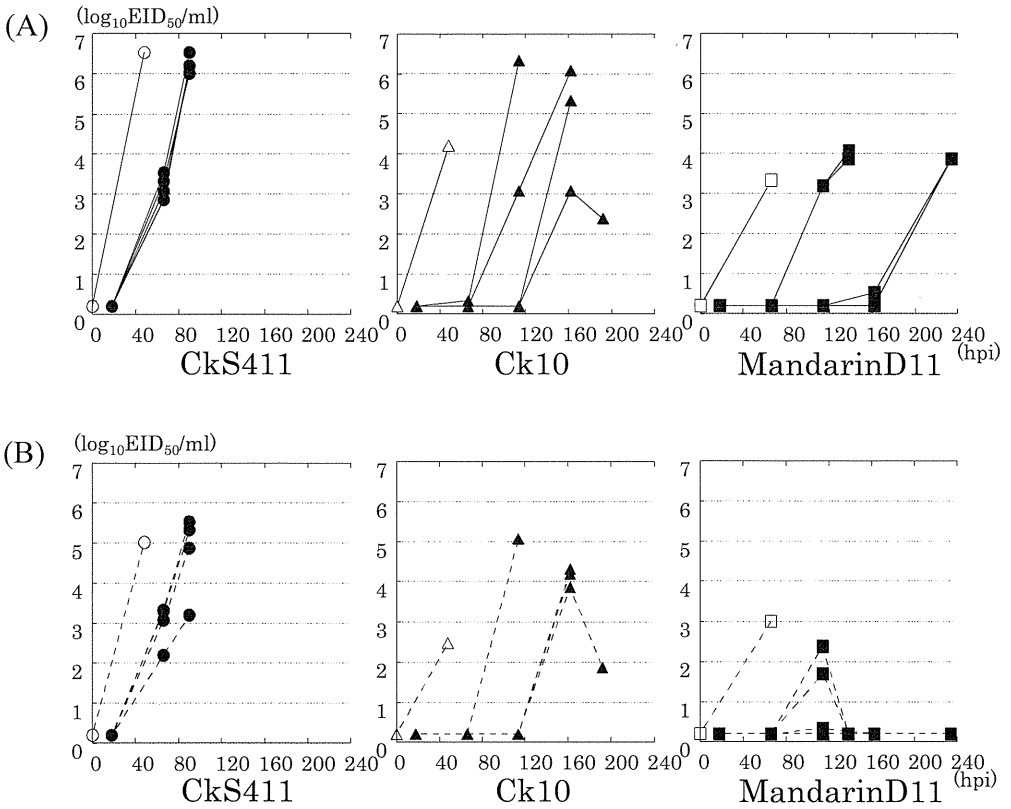


Fig. 4. *The kinetics of viral multiplication in the trachea (A) and cloaca (B) of chickens in the transmission test.* Circle (●), triangle (▲) and square (■) represent a data set obtained in the animal experiments with CkS411, Ck10 and MandarinD11, respectively. The white and black shapes indicate the results of the inoculated and cohabiting chickens, respectively. The *x*-axis is hours post inoculation (hpi) and the *y*-axis is 50% egg infective dose ($\log_{10}$ EID_{50}/ml).

Table 3
Comparison of deduced amino acid sequences among three viruses used for the animal experiment.

	PB2	PB1				PB1-F2	PA					HA[b]			NA							NS1		NS2
	491	175	566	621	746	88	344	347	400	552	578	83	84	259	48	58	60	175	189	346	414	3	79	3
MandarinD11	T	E	T	Q	T	W	K	D	P	A	G	T	K	K	P	N	V	V	K	S	G	S	M	S
CkS411	T	E	T	R	T	W	E	D	P	A	G	A	K	K	P	N	A	V	K	S	G	S	L	S
Ck10	I	D	A	Q	I	[a]	E	N	S	T	S	A	N	R	S	S	V	I	N	N	S	L	M	L

[a] There is a termination codon at 88 aa in PB1-F2 of Ck10, and it is shorter than others.
[b] Numbering starts from the beginning of HA1.

4. Discussion

The PA gene of recent Japanese HPAI isolates form a phylogenetic cluster (A) which also includes avian influenza viruses with low pathogenicity of various subtypes other than H5N1, isolated from wild birds as well as Ck/Hunan/8/2008 (H5N1). The non-H5N1 isolates in this cluster originated from wild birds, poultry and the environment in Russia, Korea, Thailand and China. There are several possible explanations for the gene constellation of cluster A viruses which has occurred. Firstly, cluster A viruses could have arisen from the reassortment between cluster B viruses and wild bird strains with PA gene segments originating from clades 2.5, 7 and 9. A descendant virus from a common ancestor of clades 2.5, 7 and 9 may have entered the wild bird population prior to 2007 and reassortment events may have occurred in the wild birds to generate viruses such as Gadwall/Altai/1328/2007 (H3N8) and Aquatic bird/Korea/w202/2007 (H5N2). Then, after the introduction of an HPAIV related to cluster B viruses into the wild bird population prior to 2008, a wild bird strain with a PA gene derived from the HPAIV reassorted with it to generate an ancestral strain of cluster A viruses. Another possibility is that cluster A viruses could have arisen due to reassortment in poultry between a virus related to cluster B and a descendant HPAIV from a common ancestor of clades 2.5, 7 and 9. The non-H5N1 isolates with PA gene segments originating from clades 2.5, 7 and 9 in the cluster resulted from reassortment between cluster A viruses and low pathogenic avian influenza viruses. However, the host in which the reassortant viruses were generated remains unknown. Which of these pathways actually caused the generation of cluster A viruses remains to be solved. However, there is little doubt that genetic material from HPAIVs of H5N1 subtype have mixed with those of avian influenza viruses in wild birds.

Using molecular epidemiology, Uchida et al., demonstrated that Thai HPAIVs isolated from wild birds could have been maintained in the wild bird population for a certain period of time (Uchida et al., 2008). It is hypothesized that the viruses were brought to Japan after being perpetuated in the wild bird population during the summer nesting season. MandarinD11 replicated exclusively in wild birds for a time while CkS411entered and replicated in chickens before the outbreak was recognized. Passage in wild birds

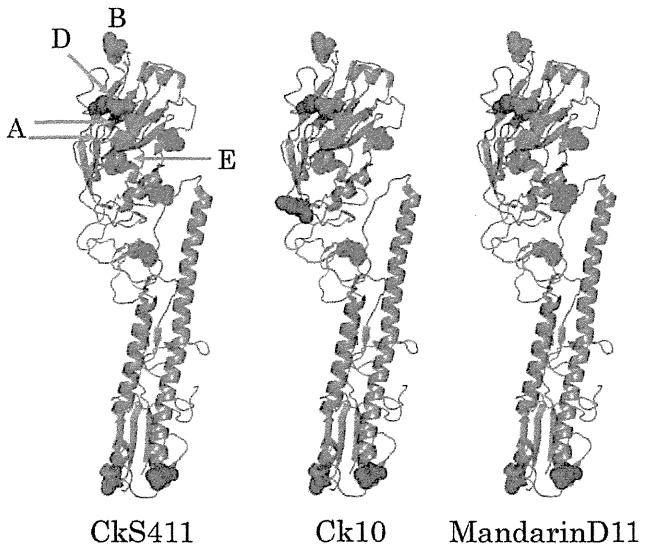


Fig. 5. Illustration of putative amino acid substitutions found in the HA of Ck10, CkS411 and MandarinD11 compared to CmHK07. Substitutions located within the predicted antigenic sites are shown in red, whereas those found elsewhere are colored in blue.

might have resulted in attenuation of MandarinD11 in chickens. An HPAIV isolated from a tree sparrow in Thailand in 2005 (TsThai) killed 70% of chickens infected with 10^6 EID_{50} and the MDT of chickens infected with it was longer than that of chickens infected with a pigeon isolate (Hayashi et al., 2011). It could be considered that passage of HPAIV in different bird species resulted in attenuated virus pathogenicity in chickens. However, in the same study, a pigeon isolate did not appear to be attenuated in chickens, suggesting that passage in wild birds does not necessarily lead to attenuation in chickens. Although we were unable to determine the amino acid substitution(s) responsible for attenuation of MandarinD11 in chickens, there were only four amino acid differences between MandarinD11 and CkS411. The generation of mutants of CkS411 with such substitutions and examination of their pathogenicity in chickens would elucidate the possible changes that may lead to attenuation of HPAIV during passage in wild birds.

It was enigmatic to note that CLD_{50} of CkS411 was higher than those of Ck10 and MandarinD11, although it appeared to be the most pathogenic isolate in chickens in terms of MDT, virus shedding and transmission. CLD_{50} might not be a factor that determines pathogenicity *in vivo* when sufficient amounts of virus are given in an experimental setting. Hayashi et al. (2011) demonstrated that when two HPAIVs with differing LD_{50} in mice, 0.8 log EID_{50} and 2.2 log EID_{50}, respectively, were inoculated at the same dose of 5 log EID_{50} into mice, lethality and death time did not significantly differ. In another study (Spekreijse et al., 2011), it was demonstrated that when a chicken was given $\geq$1 CLD_{50}, the rate of successful transmission did not differ significantly between 3.0 and 4.0 log EID_{50}. The reason for the dissociation between CLD_{50} and other pathogenic parameters remains to be determined.

In the virus transmission study, CkS411 was found to have the highest transmission speed followed by Ck10 and MandarinD11, determined by the timing of virus isolation from cohabiting chickens. It appeared that greater levels of excreted virus from infected chickens correlated with a higher transmission speed. Suzuki et al., reported that transmissibility of a virus to chickens correlated with the viral titer in oropharyngeal and cloacal swabs and the time taken for the virus to replicate to titers corresponding to the 50% chicken infectious dose (Suzuki et al., 2010). Infection of chickens with 10^6 EID_{50} MandarinD11 resulted in significantly lower viral titers from the tracheal and cloacal swabs compared to those of CkS411, coinciding with the slower transmission speed of this strain observed in the transmission study. When the kinetics of viral replication in chickens cohabiting with the chicken inoculated with MandarinD11 were looked at more closely, it was possible to identify two phases of transmission. Viral replication was detected in two chickens in both tracheal and cloacal swabs at 120 hpi, whereas the rest did not show evidence of viral replication in tracheal swabs until 160 hpi or later and not at all in cloacal swabs. It could be that the first two cohabiting chickens were infected from the inoculated chicken and that the virus was subsequently transmitted from these two, not from the inoculated chicken. If this is the case, lower virus excretion could affect the mode of transmission of the virus, although it is possible that lower exposure to the virus resulted in slower replication in affected chickens.

A wild bird isolate in this study, MandarinD11, appeared to become attenuated in chickens which could have been caused by passage and adaptation to wild birds. Under natural conditions, when a wild bird adapted virus first enters a farm, it is likely that a longer time lag would exist before the virus invasion could be detected. Early virus detection is an essential factor for successful eradication programs. It is therefore important to elucidate the molecular basis of the attenuation of virus in chickens following passage in wild birds.

Acknowledgements

We would like to thank the prefectural livestock hygiene service centers in Japan for isolation of the viruses and for providing information on the occurrences. This study was supported by a Health Labour Sciences Research Grant by the Ministry of Health Labour and Welfare of Japan and a grant-in-aid for scientific research from the Zoonoses Control Project commissioned by the Ministry of Agriculture, Forestry and Fisheries of Japan. We thank Dr. R. Webby of St. Jude Children's Research Hospital, Memphis, Tennessee, USA for providing, SJRG166615, used for the HI test antigen and ferret antiserum against A/common magpie/HK/5052/2007 (CmHK07).

References

Cattoli, G., Fusaro, A., Monne, I., Coven, F., Joannis, T., El-Hamid, H.S., Hussein, A.A., Cornelius, C., Amarin, N.M., Mancin, M., Holmes, E.C., Capua, I., 2011. Evidence for differing evolutionary dynamics of A/H5N1 viruses among countries applying or not applying avian influenza vaccination in poultry. Vaccine 29 (50), 9368–9375.

Chen, H., Smith, G.J., Zhang, S.Y., Qin, K., Wang, J., Li, K.S., Webster, R.G., Peiris, J.S., Guan, Y., 2005. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature 436 (7048), 191–192.

Coker, R.J., Hunter, B.M., Rudge, J.W., Liverani, M., Hanvoravongchai, P., 2011. Emerging infectious diseases in southeast Asia: regional challenges to control. Lancet 377 (9765), 599–609.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

Hayashi, T., Chaichoune, K., Patchimasiri, T., Hiromoto, Y., Kawasaki, Y., Wiriyarat, W., Chakritbudsabong, W., Prayoonwong, N., Chaisilp, N., Parchariyanon, S., Ratanakorn, P., Uchida, Y., Tsuda, T., Saito, T., 2011. Differential host gene responses in mice infected with two highly pathogenic avian influenza viruses of subtype H5N1 isolated from wild birds in Thailand. Virology 412 (1), 9–18.

Kajihara, M., Matsuno, K., Simulundu, E., Muramatsu, M., Noyori, O., Manzoor, R., Nakayama, E., Igarashi, M., Tomabechi, D., Yoshida, R., Okamatsu, M., Sakoda, Y., Ito, K., Kida, H., Takada, A., 2011. An H5N1 highly pathogenic avian influenza virus that invaded Japan through waterfowl migration. Japanese Journal of Veterinary Research 59 (2–3), 89–100.

Kang, H.M., Batchuluun, D., Kim, M.C., Choi, J.G., Erdene-Ochir, T.O., Paek, M.R., Sugir, T., Sodnomdarjaa, R., Kwon, J.H., Lee, Y.J., 2011. Genetic analyses of H5N1 avian influenza virus in Mongolia, 2009 and its relationship with those of eastern Asia. Veterinary Microbiology 147 (1–2), 170–175.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53 (282), 457–481.

Li, Y., Liu, L., Zhang, Y., Duan, Z., Tian, G., Zeng, X., Shi, J., Zhang, L., Chen, H., 2011. New avian influenza virus (H5N1) in wild birds, Qinghai, China. Emerging Infectious Diseases 17 (2), 265–267.

Li, Y., Shi, J., Zhong, G., Deng, G., Tian, G., Ge, J., Zeng, X., Song, J., Zhao, D., Liu, L., Jiang, Y., Guan, Y., Bu, Z., Chen, H., 2010. Continued evolution of H5N1 influenza viruses in wild birds, domestic poultry, and humans in China from 2004 to 2009. Journal of Virology 84 (17), 8389–8397.

Liu, J., Xiao, H., Lei, F., Zhu, Q., Qin, K., Zhang, X.W., Zhang, X.L., Zhao, D., Wang, G., Feng, Y., Ma, J., Liu, W., Wang, J., Gao, G.F., 2005. Highly pathogenic H5N1 influenza virus infection in migratory birds. Science 309 (5738), 1206.

Neumann, G., Green, M.A., Macken, C.A., 2010. Evolution of highly pathogenic avian H5N1 influenza viruses and the emergence of dominant variants. Journal of General Virology 91 (Pt 8), 1984–1995.

Pongcharoensuk, P., Adisasmito, W., Sat, L.M., Silkavute, P., Muchlisoh, L., Cong Hoat, P., Coker, R., 2011. Avian and pandemic human influenza policy in South-East Asia: the interface between economic and public health imperatives. Health Policy and Planning.

Sakoda, Y., Ito, H., Uchida, Y., Okamatsu, M., Yamamoto, N., Soda, K., Nomura, N., Kuribayashi, S., Shichinohe, S., Sunden, Y., Umemura, T., Usui, T., Ozaki, H., Yamaguchi, T., Murase, T., Ito, T., Saito, T., Takada, A., Kida, H., 2012. Reintroduction of H5N1 highly pathogenic avian influenza virus by migratory water birds, causing poultry outbreaks in the 2010–2011 winter season in Japan. Journal of General Virology 93 (Pt 3), 541–550.

Smith, G.J., Vijaykrishna, D., Ellis, T.M., Dyrting, K.C., Leung, Y.H., Bahl, J., Wong, C.W., Kai, H., Chow, M.K., Duan, L., Chan, A.S., Zhang, L.J., Chen, H., Luk, G.S., Peiris, J.S., Guan, Y., 2009. Characterization of avian influenza viruses A (H5N1) from wild birds, Hong Kong, 2004–2008. Emerging Infectious Diseases 15 (3), 402–407.

Spekreijse, D., Bouma, A., Stegeman, J.A., Koch, G., de Jong, M.C., 2011. The effect of inoculation dose of a highly pathogenic avian influenza virus strain H5N1 on the infectiousness of chickens. Veterinary Microbiology 147 (1–2), 59–66.

Suzuki, K., Okada, H., Itoh, T., Tada, T., Tsukamoto, K., 2010. Phenotypes influencing the transmissibility of highly pathogenic avian influenza viruses in chickens. Journal of General Virology 91 (Pt 9), 2302–2306.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24 (8), 1596–1599.

Uchida, Y., Chaichoune, K., Wiriyarat, W., Watanabe, C., Hayashi, T., Patchimasiri, T., Nuansrichay, B., Parchariyanon, S., Okamatsu, M., Tsukamoto, K., Takemae, N., Ratanakorn, P., Yamaguchi, S., Saito, T., 2008. Molecular epidemiological analysis of highly pathogenic avian influenza H5N1 subtype isolated from poultry and wild bird in Thailand. Virus Research 138 (1–2), 70–80.

Vijaykrishna, D., Bahl, J., Riley, S., Duan, L., Zhang, J.X., Chen, H., Peiris, J.S., Smith, G.J., Guan, Y., 2008. Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. PLOS Pathogens 4 (9), e1000161.

WHO/OIE/FAO H5N1 Evolution Working Group, 2012. Continued evolution of highly pathogenic avian influenza A (H5N1): updated nomenclature. Journal of Influenza and Other Respiratory Viruses 6 (1), 1–5.

World Health Organization, 2011a. Antigenic and genetic characteristics of influenza A(H5N1) and influenza A(H9N2) viruses for the development of candidate vaccine viruses for pandemic preparedness., http://www.who.int/influenza/resources/documents/2011_02_h5_h9_vaccinevirusupdate.pdf

World Health Organization, 2011b. Updated unified nomenclature system for the highly pathogenic H5N1 avian influenza viruses., http://www.who.int/influenza/gisrs_laboratory/h5n1_nomenclature/en/

Xu, X., Subbarao, Cox, N.J., Guo, Y., 1999. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. Virology 261 (1), 15–19.

J Biol Inorg Chem (2013) 18:309–321
DOI 10.1007/s00775-013-0975-7

ORIGINAL PAPER

The role of the N-terminal loop in the function of the colicin E7 nuclease domain

Anikó Czene · Eszter Németh · István G. Zóka · Noémi I. Jakab-Simon · Tamás Körtvélyesi · Kyosuke Nagata · Hans E. M. Christensen · Béla Gyurcsik

Received: 18 September 2012 / Accepted: 31 December 2012 / Published online: 19 January 2013

Abstract Colicin E7 (ColE7) is a metallonuclease toxin of *Escherichia coli* belonging to the HNH superfamily of nucleases. It contains highly conserved amino acids in its $HHX_{14}NX_8HX_3H$ ββα-type metal ion binding C-terminal active centre. However, the proximity of the arginine at the N-terminus of the nuclease domain of ColE7 (NColE7, 446–576) is necessary for the hydrolytic activity. This poses a possibility of allosteric activation control in this protein. To obtain more information on this phenomenon, two protein mutants were expressed, i.e. four and 25 N-terminal amino acids were removed from NColE7. The effect of the N-terminal truncation on the Zn^{2+} ion and DNA binding as well as on the activity was investigated in this study by mass spectrometry, synchrotron-radiation circular dichroism and fluorescence spectroscopy and agarose gel mobility shift assays. The dynamics of protein backbone movement was simulated by molecular dynamics. Semiempirical quantum chemical calculations were performed to obtain better insight into the structure of the active centre. The longer protein interacted with both Zn^{2+} ion and DNA more strongly than its shorter counterpart. The results were explained by the structural stabilization effect of the N-terminal amino acids on the catalytic centre. In agreement with this, the absence of the N-terminal sequences resulted in significantly increased movement of the backbone atoms compared with that in the native NColE7: in ΔN25-NColE7 the amino acid strings between residues 485–487, 511–515 and 570–571, and in ΔN4-NColE7 those between residues 467–468, 530–535 and 570–571.

Keywords Metallonuclease · Colicin E7 · N-terminally truncated mutants · Zinc(II) binding

Introduction

Colicin E7 (ColE7) is a metallonuclease toxin of *Escherichia coli* [1]. Its role is to protect the host cell from other

Electronic supplementary material The online version of this article (doi:10.1007/s00775-013-0975-7) contains supplementary material, which is available to authorized users.

A. Czene · E. Németh · I. G. Zóka · N. I. Jakab-Simon · B. Gyurcsik (✉)
Department of Inorganic and Analytical Chemistry,
University of Szeged,
Dóm tér 7, Szeged 6720, Hungary
e-mail: gyurcsik@chem.u-szeged.hu

A. Czene · B. Gyurcsik
Bioinorganic Chemistry Research Group of Hungarian Academy of Sciences,
Dóm tér 7, Szeged 6720, Hungary

E. Németh · T. Körtvélyesi
Department of Physical Chemistry and Material Sciences,
University of Szeged,
Aradi Vértanuk tere 1, Szeged 6720, Hungary

N. I. Jakab-Simon · H. E. M. Christensen
Department of Chemistry,
Technical University of Denmark,
Kemitorvet, Building 207,
2800 Kongens Lyngby, Denmark

K. Nagata
Department of Infection Biology,
Graduate School of Comprehensive Human Sciences and Faculty of Medicine,
University of Tsukuba,
1-1-1 Tennodai, Tsukuba 305-8575, Japan

related bacteria and bacteriophages [2] by degradation of their chromosomal DNA during environmental stress. To exert cell-killing activity, ColE7 has to get across both the outer and the inner cell membrane, facilitated by the receptor-binding and translocation domains [3, 4]. The host cell itself is protected by the simultaneously expressed immunity protein Im7 blocking the DNA binding site [5, 6] of the nuclease domain of ColE7 (NColE7) owing to tight interactions based on charge complementarity [7–12].

ColE7 belongs to the HNH superfamily of nucleases [13–15] possessing a 30–40 amino acid long ββα-type metal ion binding motif in their active centre. The amino acids histidine and asparagine are highly conserved within the sequence $HHX_{14}NX_8HX_3H$ corresponding to this motif at the C-terminal region of bacterial colicins and pyocins [16, 17]. At the same time, the HNH motif is found in various regions of a wide range of enzymes, including group I homing endonucleases (e.g. I-Hmu-I [18]), procaryotic extracellular nucleases (nuclease A [19]) and also in an increasing number of restriction endonucleases (e.g. *Mnl*I [20], *Kpn*I [21, 22], *Hph*I [23], *Eco*31I [24], *Hpy*99I [25]). Sequences are collected in the HNH [26] and Pfam [27] databases. The first histidine (H) in the name-giving HNH amino acids acts as the general base in DNA hydrolysis. The asparagine (N) residue plays a structural role constraining the HNH loop by extensive hydrogen-bonding interactions [14, 28]. The third conserved residue in the HNH string is a metal-binding histidine. The HNH motif of ColE7 binds to the 3′ site of the scissile phosphate in the minor groove of the DNA, whereas the other parts of the nuclease domain provide strong, non-specific binding within the major groove [16, 29], similarly to colicin E9 (ColE9) [30, 31]. As such, NColE7 catalyses the non-specific hydrolysis of nucleic acids.

There is still debate about the role of the different divalent metal ions in colicin nucleases [30–38]. In NColE7, three histidine side chains bind a metal cofactor, which is most probably Zn^{2+} ion under physiological conditions [32], but the apoprotein can be reactivated to a different extent by other divalent metal ions such as Mn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}, Mg^{2+}, Ca^{2+} and Sr^{2+} [5, 39]. The metal ion, having a free coordination site, has essential multiple roles in DNA cleavage: it binds to the scissile phosphodiester, polarizes the P–O bond for nucleophilic attack and stabilizes the phosphoanion transition state and the leaving group. As mentioned above, the attacking nucleophilic OH^- is supposed to be generated by the most conserved histidine residue of the HNH motif—which does coordinate to the Zn^{2+} ion. The hydrolytic reaction is also facilitated by the 19° bending of the DNA due to the protein binding [32]. The Zn^{2+} ion is not required for DNA binding, but it is essential for DNA hydrolysis [39].

In a recent article [40] it was demonstrated that during the membrane translocation process the periplasmic extracts cleave ColE7 between K446 and R447 and only the nuclease domain (R447–K576) enters the cell. The R447E ColE7 mutant lost its cell-killing activity owing to failed inner membrane translocation, but the K446E and N448A mutants retained it. However, it was shown in an in vitro assay that the R447E mutant of NColE7 (444–576) has only approximately 15 % of the endonuclease activity of the wild-type NColE7. This difference was assumed to be the consequence of lower affinity for DNA and not a consequence of the decrease in catalytic activity. On the basis of the crystal structure of Vvn endonuclease with DNA, it was proposed that the role of such a spatially close arginine residue might also be to stabilize the enzyme–product complex [41].

The necessity of the N-terminal amino acids in NColE7 for the function of the C-terminal catalytic centre poses the possibility of an allosteric activation within the enzyme that would be a desired property for use in an artificial nuclease [42]. The N-terminal end of NColE7 forms a loop leaning near to the active centre, and the interactions between them might be decisive in control of the function. In this work, two N-terminally truncated derivatives of NColE7 (446–576)—glutathione *S*-transferase (GST)–ΔN25-NColE7 and GST-ΔN4-NColE7-C* (instead of the GST-ΔN4-NColE7 protein we studied its C-terminal mutant GST-ΔN4-NColE7-C* selected by bacterial cells; the sequences are defined in Fig. 1a)—were expressed in *E. coli*. The proteins with and without the GST tag were purified for the studies of DNA- and Zn^{2+}-binding activities. Gel mobility shift assay, synchrotron-radiation circular dichroism (SRCD) spectroscopy, fluorescence spectroscopy and mass spectrometry experiments were performed and were complemented by bioinformatics, molecular dynamics and semiempirical quantum chemical calculations. The results will lead us to a better understanding of the role of the N-terminal loop in the catalysed reaction as well as its structural effects.

Materials and methods

Cloning, protein expression and purification

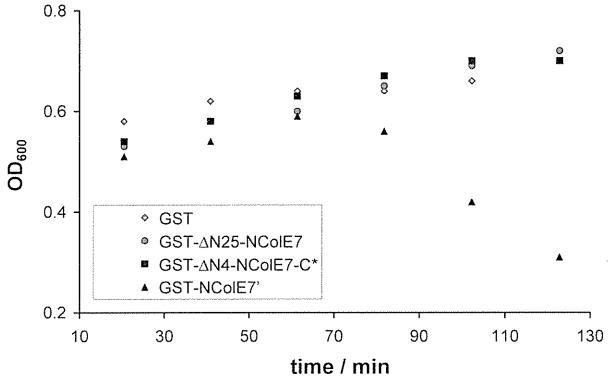
The genes of the mutant proteins were amplified by PCR from the pQE70 plasmid (a generous gift from K.-F. Chak, Institute of Biochemistry and Molecular Biology, National Yang Ming University, Taipei, Taiwan) by using the oligonucleotides ΔN4-NColE7-F: 5′-ggaattcccagggaaggcaacaggta-3′ and ΔN25-NColE7-F: 5′-ggaattcgacttaggttctcctgttcca-3′ as forward primers and NColE7-R: 5′-gccgctcgag*cta* tttacctcggtgaatatcaatatgc-3′ as the reverse primer and

a

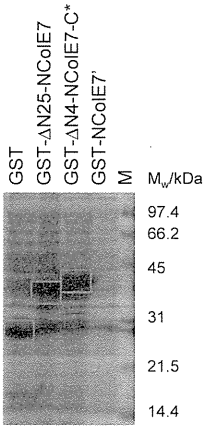
```
NColE7          GPLGSPEFKRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
ΔN4-NColE7-C*   GPLGSPEF----PGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
ΔN4-NColE7      GPLGSPEF----PGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL
ΔN25-NColE7     GPLGSPEF------------------------DLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPEL

NColE7          SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKStop-----
ΔN4-NColE7-C*   SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHQVNSSSGRIVTDStop
ΔN4-NColE7      SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKStop-----
ΔN25-NColE7     SKQFSRNNNDRMKVGKAPKTRTQDVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKStop-----
```

b



c



d

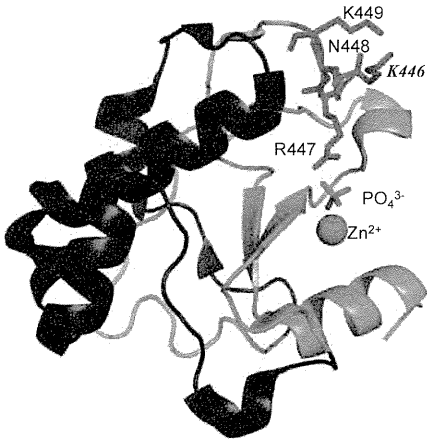


Fig. 1 **a** The sequences of the nuclease domain of colicin E7 (NColE7) [from K446 to K576 according to the original colicin E7 (ColE7) numbering] and the deletion mutants ΔN4-NColE7-C*, ΔN4-NColE7 and ΔN25-NColE7. The amino acids in *red* are fused to the N-terminus as a consequence of expression and purification from the pGEX-6-P1 vector and cleavage by PreScission protease. The amino acids in *green* indicate the result of the random mutation of the C-terminus in ΔN4-NColE7-C*. All the residues in *blue* are cut from ΔN25-NColE7 and only the residues in *dark blue* are from ΔN4-NColE7(-C*), and the HNH motif is in *orange*, similarly to **d**. **b** Growth of the *Escherichia coli* cells expressing different glutathione *S*-transferase (*GST*) protein variants after induction with isopropyl β-thiogalactoside as monitored by measurements of the optical density at 600 nm (OD_{600}). GST itself was applied as a control protein without any nuclease activity. The uncertainty of the measurements has not been plotted to simplify the diagram The average error was considered to be ±0.03 OD_{600} units. **c** The sodium dodecyl sulfate–polyacrylamide gel electrophoresis of the expressed proteins. (GST-NColE7′ is a toxic variant of NColE7—not detailed here) **d** The structure of NColE7 [Protein Data Bank (PDB) ID 1MZ8] in complex with a phosphate ion. Among the N-terminal amino acids, R447 is the closest to the phosphate ion that is bridging it with the Zn^{2+} ion. The diagram was created with PyMOL [58]

inserted into the pGEX-6-P1 (GE Healthcare) vector between the *Eco*RI and *Xho*I restriction enzyme sites (underlined sequences). The inserted DNA sequences contained a C-terminal stop codon (in italic in the primer sequence). The plasmids encoding the mutant proteins with a GST affinity tag at the N-terminus were cloned in *E. coli* DH10B or Mach1 (Invitrogen) cells and then transformed into *E. coli* BL21 (DE3), spread on Luria–Bertani medium supplemented with 100 μg/ml ampicillin (LB/Amp) plates and colonies were grown overnight at 37 °C. A small amount (4 ml) of LB/Amp medium was inoculated with a single colony and incubated overnight at 37 °C with shaking at 300 rpm. For large-scale protein production, the small-scale overnight cultures were transferred to 250 ml LB/Amp medium and bacteria were grown at 37 °C. The protein expression was induced by adding isopropyl β-thiogalactoside (200 mg/ml) to a final concentration of 0.42 mM to the cultures at an optical density at 600 nm (OD_{600}) of 0.5–0.6. The shaking at 140 rpm was continued for 2–3 h, until OD_{600} increased to 0.9–1.0. Cells were sedimented by centrifugation at 4 °C and 5,000*g* for 10 min, then resuspended in phosphate-buffered saline (PBS; 1.4 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4, pH 7.3). Pellets were disrupted by sonication and the debris was removed by centrifugation at 5,000*g* and 4 °C for 10 min. The supernatant and the resuspended aggregates were analysed by 12.5 % sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). A significant amount of the desired protein was present in inclusion bodies. The GST-based affinity purification step was done only with the soluble fractions. The protein solutions were loaded onto a 4-ml glutathione Sepharose 4B affinity column (Amersham Biosciences). The column was washed with ice-cold PBS to remove unbound material and then the bound fusion proteins were eluted with 15 mM reduced glutathione dissolved in PBS, containing 0.1 % Triton X-100 non-ionic detergent. Following an SDS-PAGE analysis, the fractions containing the target protein were pooled. Before the cleavage of the GST fusion tag, the excess of reduced glutathione was removed by dialysis against PBS, applying 150 times dilution. For digestion 1 μl (2 U/μl) of PreScission protease (GE Healthcare) was added for each 100 μg of fusion protein. The solution was incubated at 5 °C for 4 h. Following cleavage, batch purification was applied to remove the GST moiety and the PreScission protease (siliconized tubes were used to prevent the resin from sticking to the walls of the tubes): the glutathione Sepharose 4B resin was incubated with the protein solution for 30 min at 4 °C with gentle rotation. The medium was sedimented by centrifugation at 500*g* for 5 min and the supernatant was carefully transferred to new tubes. The efficiency of cleavage was checked by 17.5 % SDS-PAGE. The concentration was estimated from the gel, by comparing the intensity of the protein band with the intensity of the standard low-range (14–97 kDa) protein marker (Bio-Rad) applied in known concentrations (i.e. 5, 10, 15 and 20 μl from a 100 ng/μl stock solution). The large-scale protein expression and purification procedure has been detailed elsewhere [43].

Electrospray ionization mass spectrometry

Mass spectrometry measurements were obtained with an LCT Premier (Waters) instrument equipped with a Nanoflow electrospray ionization source and a time-of-flight analyser. The instrument was operated in positive ion mode and was calibrated using 100 mg/ml CsI in 50 % 2-propanol in the *m/z* range from 600 to 12,000. Samples were sprayed from medium-sized Au/Pd-coated borosilicate glass capillary needles (Proxeon) loaded with 3 μl protein solution. The protein concentration was 10–20 μM in 100 mM ammonium acetate (Sigma) buffer. The desalting of the protein solution and buffer exchange to the volatile buffer was done using a Micro BioSpin chromatography column (Bio-Rad). The needle voltage was typically around 1,200 V, and a cone voltage of 50 V was applied, with the cone gas flow rate maintained at 20 l/h and the source temperature maintained at 50 °C. A stock solution of 100 μM zinc(acetate)$_2$ (Sigma) was used to titrate the 20 μM ΔN25-NColE7 protein solution. The recorded *m/z* data were deconvoluted using MassLynx^{TM} version 4.1 (Waters) equipped with the MaxEnt1 algorithm. The high charge states of the multiply charged spectrum, ranging from +10 to +17, were used to calculate the apparent mass.

Gel mobility shift assay

An approximately 400 bp double-stranded DNA (dsDNA) was used as the substrate for gel mobility shift assays, and 50–100 ng of this DNA was used for each assay. The protein-to-dsDNA molar ratio ranged between 0.5:1 and 50:1, the protein-to-Zn^{2+} molar ratio was 1:1, 50 mM tris(hydroxymethyl)aminomethane/HCl buffer, pH 7.4 or PBS was used, and the final concentration of NaCl was adjusted to 50 mM. The solutions also contained 5 % glycerol. The reaction mixtures were incubated for 20–40 min at room temperature and then loaded onto 1 % agarose gel and run at 30–50 V in a buffer of 44.5 mM tris(hydroxymethyl)aminomethane, 1 mM EDTA, 44.5 mM boric acid, pH 7.3. Finally, the gel was stained with ethidium bromide solution with a concentration of 0.5 μg/ml for 30 min, followed by washing several times with distilled water. The documentation was done using UV light (Printgraph—ATTO,

Tokyo, Japan—or UVIdoc—UVItec, Cambridge, UK—gel documentation systems) at 312 nm.

SRCD spectroscopy measurements

The SRCD spectra of the proteins were recorded at the SRCD facility at the CD1 beamline [44] on the storage ring ASTRID at the Institute for Storage Ring Facilities (ISA), University of Aarhus, Denmark. The instrument was calibrated against camphorsulfonic acid. All spectra were recorded with 1 nm steps and a dwell time of 3 s per step, using 100-μm quartz cells (SUPRASIL, Hellma, Germany) for the wavelength range of 175–350 nm. The proteins were dissolved in distilled water and the pH was adjusted with HCl and NaOH solutions. The protein concentrations ranged from 10 to 50 μM. The water baseline was subtracted from raw spectra.

Fluorimetry

The most popular fluorescent probes for Zn^{2+} ions in living organisms contain quinoline units, such as 6-methoxy-(8-*p*-toluenesulfonamido)quinoline and its derivatives. These coordinate to the Zn^{2+} ion by four nitrogen donor atoms in the bis complex as illustrated in Fig. S1, but the protonation state of the complex changes with pH [45]. The fluorofor used in this study was an acid derivative of 6-methoxy-(8-*p*-toluenesulfonamido)quinoline: 4-[([6-methoxy-2-methyl-8-quinolinyl]amino)sulfonyl]benzoic acid (TFLZn) test (Sigma-Aldrich). It is soluble in water and shows high Zn^{2+} ion selectivity compared with other cations. Its affinity for Zn^{2+} ions is high enough to bind the free biological metal ion ($K_D \approx 20$ μM), but is not high enough to extract Zn^{2+} coordinated in proteins. TFLZn shows little fluorescence in the absence of Zn^{2+}. Upon formation of the bis complex, a 100-fold increase in intensity is experienced. The maxima of the excitation and emission spectra are at 360 and 498 nm, respectively.

The measurements were performed with a Hitachi F-4500 fluorescence spectrometer in a 1 cm × 1 cm path length quartz cell. The solutions were irradiated in the wavelength range between 340 and 420 nm, and the emission spectrum was recorded between 400 and 600 nm. The TFLZn concentration was adjusted to 5 μM in all cases, and mutant proteins were added in a concentration range between 0.37 and 4.8 μM. The Zn^{2+} concentration ranged from 1.2 to 27.5 μM, the EDTA concentration ranged from 1.2 to 30 μM and the DNA concentration (a 10 bp DNA string was considered to be a binding unit) ranged from 0.5 to 1.5 μM in the titration experiments.

Calculations

The initial conformation of NColE7 and the shortened mutants was taken from the Protein Data Bank (PDB) structure 1M08 [6]. This structure has a methionine (M446) instead of K446 at the N-terminus. The wild-type NColE7 calculations were done from the M446K mutant of the original PDB structure. All proteins had unprotected termini (i.e. NH_3^+ and COO^- groups).

Molecular dynamics calculations were performed with GROMACS 4.05 [46, 47], with the Gromos 53a6 [48] force field. The ionizable residues were charged according to the default pK_a values at pH 7.2, as no reason was found to alter the default pK_a values with use of propKa 3.0 [49]. The protein was placed in a cubic box with an edge size of approximately 8 nm, and was solvated by the explicit extended simple point charge (SPC/E) water model. About 16,000 equilibrated water molecules were needed to fill the box, which was neutralized owing to the positive charge of the protein by Cl^- ions replacing water molecules at the most positive parts of the box using GENION. Energy minimization was conducted using the steepest descent method. We performed 200 ps position-restrained dynamics simulations in the NVT ensemble to equilibrate the system (solvate and generate initial velocities with a Maxwell distribution) including explicit water molecules. We performed 20 ns productive molecular dynamics simulations in the NPT ensemble with periodic box conditions with integration steps of 2 fs. The temperature was set to 300 K and isotropic Berendsen P coupling and T coupling was used. For Coulomb interactions, the particle mesh Ewald method was applied with 0.9 nm electrostatic and 1.6 nm van der Waals interaction cut-offs, and the dielectric constant was set to 1.0. The constraint algorithm LINCS was used. Trajectories were analysed starting at 500 ps.

Semiempirical quantum chemical computations were performed using MOPAC2009 [50] with the PM6 method [51, 52]. Localized molecular orbitals were applied using the MOZYME [53] model implemented in MOPAC2009. The solvation was considered by the conductor-like screening model method [54] with a dielectric constant of 78.4. The geometry optimization was done by the quasi-Newton limited-memory Broyden–Fletcher–Goldfarb–Shanno method after the initial minimization of the hydrogen positions. A gradient norm of approximately 2–3 kcal/mol was achieved, and the heat of formation became essentially stationary.

Results and discussion

Cytotoxic effect of the mutant proteins

To study the necessity of the N-terminal end of the NColE7 protein for its activity, first the genes of the truncated mutants were constructed and inserted into a pGEX-6P-1 vector. The ΔN4-NColE7 and ΔN25-NColE7 proteins were designed to delete four and 25 N-terminal amino acids

from the original NColE7 sequence (starting with K446 according to the original numbering of ColE7 protein). It is known that the native ColE7 gene is toxic for the cells, owing to the unwanted minor expression level of the protein during the cloning process [55]. Thus, the success of the cloning and expression of a ColE7 mutant indicates that the given protein is not toxic for the cells. According to the PCR followed by agarose gel electrophoresis, the genes were successfully inserted into the vector and cloned in either DH10B or Mach1 cells. The transformed BL21 (DE3) cells based on the change in OD_{600} after the induction of the protein expression were grown in a manner similar to that for the cells expressing GST itself, unlike those of the toxic variant of NColE7, where the cells started to die shortly after induction (Fig. 1b). The SDS-PAGE of the expressed proteins showed intense bands around the expected molecular mass (Fig. 1c). However, the DNA sequencing and the mass spectra of the proteins showed that although the GST-ΔN25-NColE7 sequence was correct, instead of GST-ΔN4-NColE7 we had expressed a new mutant, GST-ΔN4-NColE7-C*, with a modified C-terminus—but containing all the amino acids necessary for binding of the Zn^{2+} ion. This strongly suggests that GST-ΔN4-NColE7 was cytotoxic. The sequences of the proteins after the GST cleavage are depicted in Fig. 1a.

To check the effect of the C-terminal modification in GST-ΔN4-NColE7-C*, we have also shown that the ΔN4-NColE7 mutant expressed from the pET21a plasmid without an N-terminal tag but having the correct sequence at the C-terminus is non-toxic for the cells [43]. These results together proved that the N-terminal basic amino acids are necessary for the cell-killing activity of the enzyme if it is overexpressed in bacterial cells. Looking at the available crystal structures of NColE7 (Table 1), we see that these amino acid residues with special emphasis on residue R447 were observed in only a few of them. In those few including this amino acid, however, the R447 side chain is situated close to the Zn^{2+} ion in the active centre [5, 6, 10]. The two positively charged residues are bridged by a phosphate ion (Fig. 1d), which is most probably replaced by the scissile phosphodiester group of the DNA in the catalytically active complex [16]. In the NColE7–DNA crystals, R447 is mostly missing from the solved structure [16, 29, 32], but it is close to the phosphate backbone in the structure of a metal ion deficient mutant [57].

In view of the possible allosteric control, it is important to know what the function of the N-terminal residues with positively charged side chains (one arginine and two lysines) is and the role of the whole N-terminal chain—a component without autonomous secondary structure—regarding the activity of the NColE7 protein. We attempted to get closer to the solution of this problem by means of the investigation of the Zn^{2+}- and DNA-binding abilities of the expressed mutant proteins.

Table 1 Crystal structures of the nuclease domain of colicin E7 (NColE7) containing amino acid sequences of different lengths

PDB ID	Mutation	Complex	Sequence in PDB file[a]	Reason for inactivity
1M08 [6]	K446M	Protein–Zn–PO_4[b]	446 MRNK-HRGK 576	–
1MZ8 [5]	–	Protein–Zn–PO_4–Im7	447 RNKP-IDIH 573	–
1PT3 [16]	–	Protein–8 bp DNA	449 KPGK-HRGK 576	No metal ion
1ZNS [32]	K443M/H545E	Protein–Zn–12 bp DNA	450 PGKA-DIHR 574	Mutation
1ZNV [32]	K443M/H545E	Protein–Ni–PO_4–Im7	450 PGKA-HRGK 576	–
7CEI [10]	–	Protein–Zn–Im7	447 RNKP-IDIH 573	–
2IVH [29]	H545Q	Protein–Zn–18 bp DNA	449 KPGK-IDIH 573	Mutation
2JAZ [28]	N560D	Protein–Zn–PO_4–Im7	450 PGKA-HRGK 576	–
2JB0 [28]	H573A	Protein–Zn–Im7	449 KPGK-HIDI 572	–
2JBG [28]	N560A	Protein–Zn–SO_4–Im7	448 NKPG-HRGK 576	–
3GJN [56]	H545A	Protein–Zn–Im[c]	450 PGKA-HRGK 576	–
3GKL [56]	H545A	Protein–Zn–Im[c]	450 PGKA-HRGK 576	–
3FBD [57]	D493Q	Protein–18 bp DNA	445 SKRN-HRGK 576	No metal ion

PDB Protein Data Bank

[a] All the proteins were expressed in the presence of the immunity protein. A general sequence of NColE7 was MLDKES+446–576, with the exception of the one with PDB ID 7CEI, where an N-terminal hexahistidine tag in a form of MRGSHHHHHHGSES was attached to the 446–576 sequence

[b] Charges are omitted for simplicity

[c] Mutant immunity proteins were applied in these experiments. The expression of NColE7 was not described in detail in the original article [56]

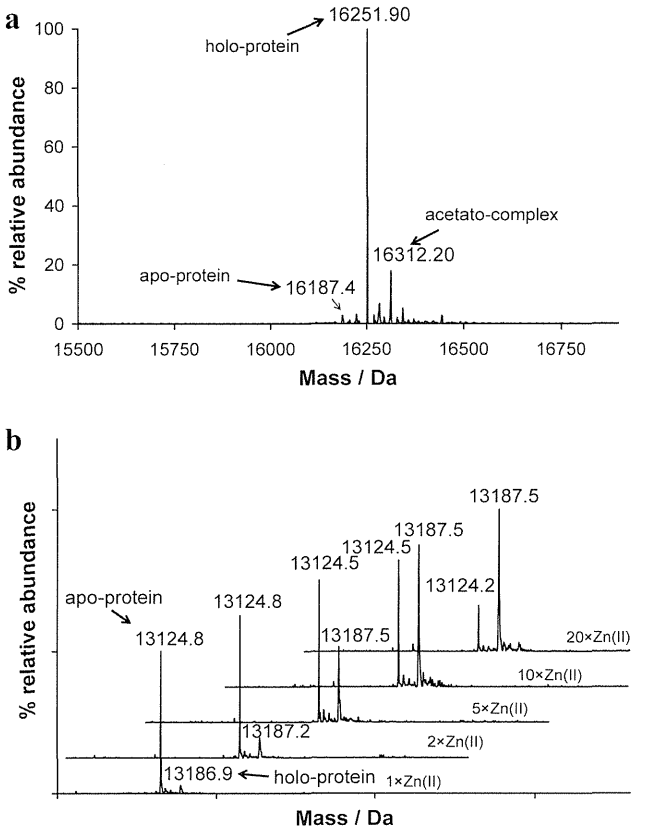


Fig. 2 **a** Mass spectrum of the ΔN4-NColE7-C* mutant. The main peak corresponds to the mass of the holoprotein. The theoretical average mass of the apoprotein is calculated to be 16,188.1 Da, whereas the mass of the Zn^{2+} complex is 16,253.5 Da. **b** Mass spectra of the ΔN25-NColE7 mutant in the presence of onefold to 20-fold molar excess of Zn^{2+} ions. The theoretical average mass of the apoprotein is 13,123.7 Da, whereas the mass of the Zn^{2+} complex is calculated to be 13,189.1 Da

Protein–Zn^{2+} interaction

Mass spectrometry investigations

Intact protein mass spectrometry was used to identify the truncated proteins and to further investigate their Zn^{2+} ion binding abilities. Figure 2a shows the mass spectrum of the purified ΔN4-NColE7-C* mutant recorded in the volatile ammonium acetate buffer without addition of Zn^{2+} ions. The apparent mass of the main peak, i.e.,16,251.9 Da, corresponded to the mass of the holoprotein. This clearly demonstrates that the ΔN4-NColE7-C* mutant was purified in its Zn^{2+}-bound form. The multiply charged spectrum showed at almost all charge states the presence of a significant amount of acetato complex as a result of a non-covalent interaction. Its amount increased with the decrease in the protein's charge state (see Fig. S2 for the *m/z* spectrum). Since the metal binding site consists of three histidine imidazole nitrogen ligands from the HNH motif, the presence of the acetate ligand, completing the tetrahedral coordination around the Zn^{2+} ion, is expected here instead of that of the phosphate ion, which usually occurs in the crystal structures.

In contrast to ΔN4-NColE7-C*, Fig. 2b shows that the purified ΔN25-NColE7 mutant did not contain Zn^{2+} ions, and that it was not able to complete the metallation of the apo form even in the presence of 20-fold molar excess of Zn^{2+} ions at pH 6.7 (pH of the ammonium acetate buffer used for mass spectrometry measurements). The apparent mass of the apoprotein (13,124.8 Da) fits very well with the calculated theoretical mass. As a result of increasing amounts of Zn^{2+} ions, the molar ratio of the holoprotein increased. In the presence of tenfold metal ion excess, the molar ratio of the apoprotein and the holoprotein is approximately 1:1, whereas 20-fold excess of Zn^{2+} ions is required to achieve approximately 85 % metallation. The K_D value calculated from the ratio of the ion signal intensities of the apoprotein and the holoprotein in the *m/z* spectra of the 11 times charged ion (see the series of spectra in Fig. S3) was 74 ± 18 μM, assuming that no dissociation occurs during the transmission through the mass spectrometer and the metal ion binding to the protein does not alter the ionization efficiency of the non-covalent complex [59]. It should, however, be noted that the estimated constant mentioned above would largely depend on the protonation state of the protein molecule. This means that the stability of the metal ion complex is lower than that reported for the Zn^{2+} binding of the nuclease domain of ColE9 (nanomolar K_D) and is similar to that for Ni^{2+} binding of the same protein [33]. These data unambiguously indicate that the 21 N-terminal amino acids of the ΔN4-NColE7-C* mutant play an important role in the metal binding in the HNH motif at the C-terminus of the protein. The amino acids of the N-terminal loop may, e.g., affect the dynamics of the protein folding and promote the formation of the proper structure of the protein.

Fluorimetry

Fluorimetry can also be applied to monitor the Zn^{2+} ion binding of proteins by probes that are fluorescent in their zinc(II) complexes [45], such as the TFLZn probe used by us. The maximal fluorescence intensity at 490 nm was monitored (Fig. S4). In the solutions containing the TFLZn probe and the truncated NColE7 proteins in 2:1 molar ratio, different behaviour was observed for ΔN25-NColE7 and ΔN4-NColE7-C*. There was no significant change in the fluorescence intensity of TFLZn in the presence of ΔN25-NColE7, whereas the addition of 1 equiv [$c(Zn^{2+}) = c$(protein)] of Zn^{2+} ions caused a large increase in the intensity. At the same time, an increase of the fluorescence was observed upon addition of ΔN4-NColE7-C* to the TFLZn solution. The

resulting fluorescence intensity was in both cases significantly higher in the presence of the proteins (and metal ion) than in the Zn^{2+}–TFLZn binary system. This suggests that in agreement with the mass spectrometry result, ΔN25-NColE7 does not contain Zn^{2+} ions, and that the proteins cannot completely replace the TFLZn probe in the coordination sphere of the Zn^{2+} ion. The latter can be explained supposing that the proteins do not fill all the coordination sites around the metal ion (coordination occurs through the three histidine side chains). Therefore, according to the thermodynamics of the system, the formation of Zn^{2+}(protein)(TFLZn) ternary complexes is also possible, in which an enhancement of the fluorescence can be observed. The addition of DNA to protein-containing solutions slightly decreased the fluorescence, probably owing to the replacement of the dye in the ternary complex. The slightly larger extent of the change for the ΔN4-NColE7-C* protein points to its stronger DNA binding (see later).

SRCD spectroscopy results

In a chiral environment there is a difference between the absorption of the left and right circularly polarized light, and a plot of the difference in their absorption coefficients ($\Delta\varepsilon = \varepsilon_{left} - \varepsilon_{right}$) versus wavelength yields a characteristic circular dichroism spectrum of the sample. The relative position of chiral amide chromophores in proteins, i.e. the secondary structure, and its changes are responsible for this effect in the wavelength region of UV light (180–250 nm). Since SRCD spectroscopy provides an optimal and even flux of UV light in a highly controlled manner, it can be applied for accurate study of the solution structure and interactions of proteins [60].

The effect of metal ion binding on the structure of the mutant NColE7 proteins was investigated by monitoring the changes in their SRCD spectra on addition of Zn^{2+} ions and/or EDTA to their solutions, as described in "Materials and methods". Figure 3a shows the spectra obtained for ΔN4-NColE7-C* protein. As can be seen, the addition of Zn^{2+} ions did not affect the SRCD spectrum (not even at fivefold Zn^{2+} excess—data not shown). This again suggests that the ΔN4-NColE7-C* protein already includes a bonded metal ion. At the same time, an excess of EDTA caused a slight decrease in the intensity. This suggests that the removal of the Zn^{2+} ions from the protein by EDTA caused only a negligible change in the secondary structure composition of the protein, suggesting that the structure is also stable without the metal ion.

In similar experiments with ΔN25-NColE7, both the intensity and the shape of the SRCD spectra changed continuously (the spectral pattern becoming more similar to that of the ΔN4-NColE7-C* spectrum) upon gradual addition of up to 10 equiv of Zn^{2+} ions, and the extent of this change became negligible at higher metal ion excess. The addition of an excess of EDTA resulted in a spectrum similar to that recorded in the absence of metal ions (Fig. 3b). These results further show that the shorter protein binds Zn^{2+} ions more weakly than ΔN4-NColE7-C*, which could result from the more extensive distortion of the metal ion binding site upon the deletion of the further 21 amino acids.

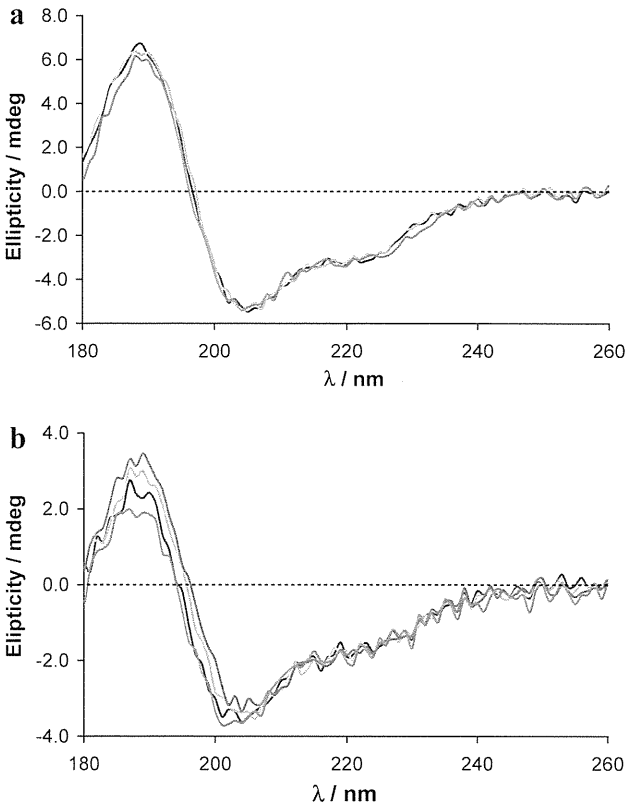


Fig. 3 Comparison of the synchrotron-radiation circular dichroism (SRCD) spectra recorded for **a** ΔN4-NColE7-C* (c = 36 μM) and **b** ΔN25-NColE7 (c = 18 μM) under various conditions. The spectra of the aqueous solutions of the proteins are in *blue*. The *yellow curves* belong to the systems where 1 equiv of Zn^{2+} ions has been added to the protein solutions. For ΔN25-NColE7, the SRCD spectrum recorded in the presence of 10 equiv of Zn^{2+} ions has also been plotted (*orange curve*), since the change here is more expressed than in the case of ΔN4-NColE7-C*. Finally, an excess of EDTA was added to the previous solutions and the spectra were recorded (*green*). In all cases the average of three measurements was plotted

Protein–DNA interactions

SRCD spectroscopy results

SRCD spectroscopy was also applied to study the dsDNA binding of the mutant proteins. The spectra recorded in the presence of dsDNA are presented in Fig. S5. Although the gel mobility shift experiments (see later) proved there was DNA binding, the recorded spectra showed that the