Supporting information

Interfacial cavity filling to optimize CD4-mimetic miniprotein interactions with the HIV-1 surface protein

Laurence Morellato-Castillo,[†,∞] Priyamvada Acharya,[§,∞] Olivier Combes,[†] Johan Michiels,[‡] Anne Descours,[†] Oscar H. P. Ramos,[†] Yongping Yang,[§] Guido Vanham,[‡] Kevin K. Ariën,[‡] Peter D. Kwong,[§] Loïc Martin,[,†] and Pascal Kessler*[,†]*

[†]CEA, iBiTec-S, Service d'Ingénierie Moléculaire des Protéines, F-91191 Gif-sur-Yvette, France
[§]Vaccine Research Center, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892, United States
[‡]Institute of Tropical Medicine of Antwerp, Department Microbiology, Unit Virology, Nationalestraat 155, Antwerp 2000, Belgium

Table of contents **Page**

Experimental section

Loaded resins **6'** to **13'**:

N-allyloxycarbonyl-O-(c-alkyl or c-aryl-alkyl chain)-L-tyrosine immobilized on Wang resin (6' to 13'). N-allyloxycarbonyl-O-(2-cyclopentylethyl)-L-tyrosine immobilized on Wang resin (**6'**) was synthesized according to the procedure G1 using 3.06 g of 2-cyclopentylethanol (4 equiv, 18 mmol) with 4 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(2-phenylethyl)-L-tyrosine immobilized on Wang resin (**7'**) was synthesized according to the procedure G1 using 4.3 mL of 2-phenylethanol (8 equiv, 36 mmol) with 8 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(2-cyclohexylethyl)-L-tyrosine immobilized on Wang resin (**8'**) was synthesized according to the procedure G1 using 5 mL of 2-cyclohexylethanol (8 equiv, 36 mmol) with 8 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(3-cyclohexyl-1-propyl)-L-tyrosine immobilized on Wang resin (**9'**) was synthesized according to the procedure G1 using 3.42 mL of 3-cyclohexyl-1-propanol (5 equiv, 22.5 mmol) with 5 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(4-benzyloxy-1-butyl)-L-tyrosine immobilized on Wang resin (**10'**) was synthesized according to the procedure G1 using 6.44 g of 4-benzyloxy-1-butanol (**41'**) (8 equiv, 36 mmol) with 8 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(5-benzyloxy-1-pentyl)-L-tyrosine immobilized on Wang resin (**11'**) was synthesized according to the procedure G1 using 6.98 g of 5-benzyloxy-1-pentanol (**42'**) (8 equiv, 36 mmol) with 8 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-(((2-(benzyloxy)ethyl)thio)ethyl)-L-tyrosine immobilized on Wang resin (**12'**) was synthesized according to the procedure G1 using 4.01 g of ((2-(benzyloxy)ethyl)thio)ethanol (**43'**) (4.2 equiv, 18.9 mmol) with 4.2 equiv of PPh_3 and DIAD.

N-allyloxycarbonyl-O-((N-Cbz)-(S)-(+)-2-pyrrolidinemethyl)-L-tyrosine immobilized on Wang resin (**13'**) was synthesized according to the procedure G1 using 4.98 g of (N-Cbz)-(S)-(+)-2-pyrrolidinemethanol (**48'**) (4.7 equiv, 21.2 mmol) with 4.7 equiv of PPh_3 and DIAD.

N-(9-fluorenylmethoxycarbonyl)-O-(c-alkyl or c-aryl-alkyl chain)-L-tyrosine (33' to 40').

Loaded resins **15'** to **22'**:

O-(2-cyclopentylethyl)-L-tyrosine immobilized on Wang resin (**15'**) was synthesized according to the procedure G2 using resin **6'**.

O-(2-phenylethyl)-L-tyrosine immobilized on Wang resin (**16'**) was synthesized according to the procedure G2 using resin **7'**.

O-(2-cyclohexylethyl)-L-tyrosine immobilized on Wang resin (**17'**) was synthesized according to the procedure G2 using resin **8'**.

O-(3-cyclohexyl-1-propyl)-L-tyrosine immobilized on Wang resin (**18'**) was synthesized according to the procedure G2 using resin **9'**.

O-(4-benzyloxy-1-butyl)-L-tyrosine immobilized on Wang resin (**19'**) was synthesized according to the procedure G2 using resin **10'**.

O-(5-benzyloxy-1-pentyl)-L-tyrosine immobilized on Wang resin (**20'**) was synthesized according to the procedure G2 using resin **11'**.

O-(2'-benzyloxyethyl-sulfid-2-ethyl)-L-tyrosine immobilized on Wang resin (**21'**) was synthesized according to the procedure G2 using resin **12'**.

O-[(N-benzyloxycarbonyl)-(S)-(+)-2-pyrrolidinemethyl]-L-tyrosine immobilized on Wang resin (**22'**) was synthesized according to the procedure G2 using resin **13'**.

Loaded resins **24'** to **31'**:

N-(9-fluorenylmethoxycarbonyl)-O-(2-cyclopentylethyl)-L-tyrosine immobilized on Wang resin (**24'**) was synthesized according to the procedure G3 using resin **15'**.

N-(9-fluorenylmethoxycarbonyl)-O-(2-phenylethyl)-L-tyrosine immobilized on Wang resin (**25'**) was synthesized according to the procedure G3 using resin **16'**.

N-(9-fluorenylmethoxycarbonyl)-O-(2-cyclohexylethyl)-L-tyrosine immobilized on Wang resin (**26'**) was synthesized according to the procedure G3 using resin **17'**.

N-(9-fluorenylmethoxycarbonyl)-O-(3-cyclohexyl-1-propyl)-L-tyrosine immobilized on Wang resin (**27'**) was synthesized according to the procedure G3 using resin **18'**.

N-(9-fluorenylmethoxycarbonyl)-O-(4-benzyloxy-1-butyl)-L-tyrosine immobilized on Wang resin (**28'**) was synthesized according to the procedure G3 using resin **19'**.

N-(9-fluorenylmethoxycarbonyl)-O-(5-benzyloxy-1-pentyl)-L-tyrosine immobilized on Wang resin (**29'**) was synthesized according to the procedure G3 using resin **20'**.

N-(9-fluorenylmethoxycarbonyl)-O-(2'-benzyloxyethyl-sulfid-2-ethyl)-L-tyrosine immobilized on Wang resin (**30'**) was synthesized according to the procedure G3 using resin **21'**.

N-(9-fluorenylmethoxycarbonyl)-O-[(N-benzyloxycarbonyl)-(S)-(+)-2-pyrrolidinemethyl]-L-tyrosine immobilized on Wang resin (**31'**) was synthesized according to the procedure G3 using resin **22'**.

Compounds **33'** to **40'**:

N-(9-fluorenylmethoxycarbonyl)-O-(2-cyclopentylethyl)-L-tyrosine (33'). Procedure G4 was applied on **24'** and the crude residue was purified according to FC1 to yield 1.17 g of **33'** (52%) as a white solid: Rf (DCM/MeOH: 96/4) = 0.29; mp 133 °C; ^{1}H NMR ($CDCl_3$) δ 1.11-1.19 (m, 2H), 1.52-1.65 (m, 4H), 1.76-1.81 (m, 4H), 1.90-1.99 (m, 1H), 3.11 (ddt, *J* = 13.8, 18.8, 5.3 Hz, 2H), 3.93 (t, *J* = 6.6 Hz, 2H), 4.21 (t, *J* = 6.8 Hz, 1H), 4.37 (dd, *J* = 10.2, 6.8 Hz, 1H), 4.45 (dd, *J* = 10.0, 7.2 Hz, 1H), 4.67 (dt, *J* = 7.3, 5.5 Hz, 1H), 5.20 (d, *J* = 8.1 Hz, 1H), 6.82 (d, *J* = 8.2 Hz, 2H), 7.05 (d, *J* = 8.2 Hz, 2H), 7.31 (t, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.3 Hz, 2H), 7.56 (t, *J* = 6.5 Hz, 2H), 7.77 (d, *J* = 7.5 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 25.03, 32.67, 35.38, 36.85, 47.10, 54.61, 67.02, 67.38, 114.63, 119.98, 125.03, 127.05, 127.72, 130.32, 141.29, 143.64, 155.72, 158.36, 175.71; ES/MS for $C_{31}H_{33}NO_5$ (negative ionization): Mol.wt calcd: 499.2, found: 499.3; Rt (HPLC-A): 4.99 (100%).

N-(9-fluorenylmethoxycarbonyl)-O-(2-phenylethyl)-L-tyrosine (34'). Procedure G4 was applied on **25'** and the crude residue was purified according to FC1 to yield 1.44 g of **34'** (63%) as a white solid: Rf (DCM/MeOH: 95/5) = 0.26; mp 110-111 °C; ^{1}H NMR ($CDCl_3$) δ 2.99-3.17

(m, 4H), 4.12 (t, *J* = 7.1 Hz, 2H), 4.20 (t, *J* = 6.9 Hz, 1H), 4.36 (dd, *J* = 10.7, 7.0 Hz, 1H), 4.44 (dd, *J* = 10.4, 7.3 Hz, 1H), 4.64 (dt, *J* = 7.8, 3.4 Hz, 1H), 5.20 (d, *J* = 8.5 Hz, 1H), 6.81 (d, *J* = 8.3 Hz, 2H), 7.04 (d, *J* = 8.3 Hz, 2H), 7.25-7.35 (m, 2H), 7.28 (s, 5H), 7.39 (t, *J* = 7.3 Hz, 2H), 7.53-7.55 (m, 2H), 7.76 (d, *J* = 7.6 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 35.75, 36.86, 47.11, 54.54, 67.04, 68.55, 114.71, 119.99, 126.49, 127.06, 127.39, 127.73, 128.47, 128.98, 130.36, 138.14, 141.30, 143.64, 155.68, 158.02, 175.59; ES/MS for $C_{32}H_{29}NO_5$ (negative ionization): Mol.wt calcd: 507.2, found: 507.1; Rt (HPLC-A): 3.74 (85%).

N-(9-fluorenylmethoxycarbonyl)-O-(2-cyclohexylethyl)-L-tyrosine (35'). Procedure G4 was applied on **26'** and the crude residue was purified according to FC1 to yield 600 mg of **35'** (26%) as a beige solid: Rf (DCM/MeOH: 97/3) = 0.33; mp 88 °C; ^{1}H NMR ($CDCl_3$) δ 0.96 (dt, *J* = 11.7, 9.7 Hz, 2H), 1.15-1.30 (m, 6H), 1.63-1.77 (m, 5H), 3.11 (ddt, *J* = 14.0, 18.5, 5.5 Hz, 2H), 3.94 (t, *J* = 6.8 Hz, 2H), 4.21 (t, *J* = 7.0 Hz, 1H), 4.36 (dd, *J* = 10.6, 7.0 Hz, 1H), 4.45 (dd, *J* = 10.2, 7.3 Hz, 1H), 4.67 (dt, *J* = 6.9, 6.2 Hz, 1H), 5.21 (d, *J* = 8.5 Hz, 1H), 6.81 (d, *J* = 8.2 Hz, 2H), 7.04 (d, *J* = 8.1 Hz, 2H), 7.31 (t, *J* = 7.5 Hz, 2H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.55 (t, *J* = 6.6 Hz, 2H), 7.76 (d, *J* = 7.5 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 26.23, 26.52, 33.28, 34.49, 36.64, 36.88, 47.11, 54.61, 65.88, 67.04, 114.65, 119.97, 125.04, 127.06, 127.72, 130.32, 141.30, 143.62, 155.72, 158.37, 175.66; ES/MS for $C_{32}H_{35}NO_5$ (negative ionization): Mol.wt calcd: 513.3, found: 513.3; Rt (HPLC-A): 5.35 (86%).

N-(9-fluorenylmethoxycarbonyl)-O-(3-cyclohexyl-1-propyl)-L-tyrosine (36'). Procedure G4 was applied on **27'** and the crude residue was purified according to FC1 to yield 1.30 g of **36'** (55%) as a yellow solid: Rf (DCM/MeOH: 96/4) = 0.23; mp 124 °C; ^{1}H NMR ($CDCl_3$) δ 0.90 (dt, *J* = 12.0, 9.1 Hz, 2H), 1.13-1.35 (m, 8H), 1.64-1.80 (m, 5H), 3.11 (ddt, *J* = 14.3, 17.6, 5.6 Hz, 2H), 3.89 (t, *J* = 6.7 Hz, 2H), 4.21 (t, *J* = 6.9 Hz, 1H), 4.38 (dd, *J* = 10.4, 6.9 Hz, 1H), 4.45 (dd, *J* = 10.4, 7.1 Hz, 1H), 4.67 (dt, *J* = 7.0, 5.8 Hz, 1H), 5.19 (d, *J* = 7.2 Hz, 1H), 6.81 (d, *J* = 7.8 Hz, 2H), 7.04 (d, *J* = 8.0 Hz, 2H), 7.29 (t, *J* = 7.4 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.55 (t, *J* = 6.3 Hz, 2H), 7.76 (d, *J* = 7.5 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 26.35, 26.63, 26.64, 33.31, 33.67, 36.83, 37.44, 47.09, 54.61, 67.13, 68.33, 114.68, 119.99, 125.02, 127.07, 127.75, 130.31, 141.30, 143.61, 155.82, 158.40, 175.65; ES/MS for $C_{33}H_{37}NO_5$ (negative ionization): Mol.wt calcd: 527.3, found: 527.3; Rt (HPLC-A): 5.49 (88%).

N-(9-fluorenylmethoxycarbonyl)-O-(4-benzyloxy-1-butyl)-L-tyrosine (37'). Procedure G4 was applied on **28'** and the crude residue was purified according to FC1 to yield 1.34 g of **37'** (52.5%) as a pale yellow solid: Rf (DCM/MeOH: 97/3) = 0.48; mp 88 °C; ^{1}H NMR ($CDCl_3$) δ 1.76-1.88 (m, 4H), 3.11 (ddt, *J* = 14.1, 19.2, 5.8 Hz, 2H), 3.54 (t, *J* = 6.2 Hz, 2H), 3.93 (t, *J* = 6.2 Hz, 2H), 4.21 (t, *J* = 7.0 Hz, 1H), 4.36 (dd, *J* = 10.8, 7.2 Hz, 1H), 4.45 (dd, *J* = 10.5, 7.1 Hz, 1H), 4.52 (s, 2H), 4.67 (dt, *J* = 7.9, 5.9 Hz, 1H), 5.23 (d, *J* = 8.1 Hz, 1H), 6.80 (d, *J* = 8.6 Hz, 2H), 7.04 (d, *J* = 8.4 Hz, 2H), 7.26-7.35 (m, 7H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.56 (t, *J* = 6.7 Hz, 2H), 7.76 (d, *J* = 7.5 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 26.06, 26.25, 36.83, 47.06, 54.59, 66.99, 67.51, 69.83, 72.82, 114.58, 119.92, 124.99, 127.00, 127.17, 127.54, 127.68, 128.33, 130.30, 138.31, 141.24, 143.60, 155.69, 158.19, 175.57; ES/MS for $C_{35}H_{35}NO_6$ (negative ionization): Mol.wt calcd: 565.2, found: 565.1; Rt (HPLC-A): 4.66 (90%).

N-(9-fluorenylmethoxycarbonyl)-O-(5-benzyloxy-1-pentyl)-L-tyrosine (38'). Procedure G4 was applied on **29'** and the crude residue was purified according to FC2 to yield 0.91 g of **38'** (35%) as a yellow solid: Rf (DCM/AcOH: 9.8/0.2) = 0.45; mp 66 °C; ^{1}H NMR ($CDCl_3$) δ 1.42-1.51 (m, 2H), 1.56-1.73 (m, 4H), 3.03 (ddt, *J* = 14.3, 7.7, 5.7 Hz, 2H), 3.42 (t, *J* = 6.4 Hz, 2H),

3.83 (t, *J* = 6.3 Hz, 2H), 4.13 (t, *J* = 6.9 Hz, 1H), 4.27 (dd, *J* = 10.6, 6.8 Hz, 1H), 4.37 (dd, *J* = 10.1, 7.3 Hz, 1H), 4.44 (s, 2H), 4.58 (dt, *J* = 8.0, 5.6 Hz, 1H), 5.14 (d, *J* = 8.2 Hz, 1H), 6.72 (d, *J* = 8.5 Hz, 2H), 6.96 (d, *J* = 8.4 Hz, 2H), 7.18-7.35 (m, 9H), 7.46-7.50 (m, 2H), 7.69 (d, *J* = 7.4 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 23.34, 29.64, 30.03, 37.43, 47.67, 55.20, 67.58, 68.29, 70.76, 73.47, 115.18, 120.56, 125.61, 125.67, 127.63, 127.72, 128.13, 128.24, 128.29, 128.94, 130.91, 138.99, 141.86, 144.23, 155.69, 158.84, 177.49; Rt (HPLC-A): 4.77 (93%).

N-(9-fluorenylmethoxycarbonyl)-O-((2(2-benzyloxyethyl)thio)ethyl)-L-tyrosine (39'). Procedure G4 was applied on **30'** and the crude residue was purified according to FC2 to yield 0.79 g of **39'** (29%) as a yellow solid: Rf (DCM/AcOH: 99/1) = 0.17; mp 68 °C; ^{1}H NMR ($CDCl_3$) δ 2.75 (t, *J* = 6.4 Hz, 2H), 2.84 (t, *J* = 5.2 Hz, 2H), 3.02 (ddt, *J* = 14.2, 9.3, 5.5 Hz, 2H), 3.60 (t, *J* = 6.6 Hz, 2H), 4.01 (t, *J* = 6.7 Hz, 2H), 4.12 (t, *J* = 6.8 Hz, 1H), 4.25-4.41 (m, 2H), 4.47 (s, 2H), 4.57 (dt, *J* = 7.5, 5.9 Hz, 1H), 5.15 (d, *J* = 8.3 Hz, 1H), 6.71 (d, *J* = 8.4 Hz, 2H), 6.96 (d, *J* = 8.3 Hz, 2H), 7.18-7.27 (m, 7H), 7.32 (t, *J* = 7.3 Hz, 2H), 7.47 (t, *J* = 7.1 Hz, 2H), 7.69 (d, *J* = 7.4 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 32.04, 32.79, 37.45, 47.67, 55.19, 67.58, 68.35, 70.46, 73.66, 115.26, 120.57, 125.59, 125.66, 127.63, 128.30, 128.31, 129.01, 131.00, 138.47, 141.86, 144.21, 156.29, 158.22, 177.45; Rt (HPLC-A): 4.61 (93%).

N-(9-fluorenylmethoxycarbonyl)-O-[(N-benzyloxycarbonyl)-(S)-(+)-2-pyrrolidinemethyl]-L-tyrosine (40'). Procedure G4 was applied on **31'** and the crude residue was purified according to FC2 to yield 0.89 g of **40'** (32%) as a pale yellow powder: Rf (DCM/AcOH: 99/1) = 0.14; mp 99 °C; ^{1}H NMR ($CDCl_3$) δ 1.77-1.95 (m, 4H), 2.90-3.06 (m, 2H), 3.29-3.38 (m, 2H), 3.59-3.96 (m, 1H), 4.00-4.15 (m, 3H), 4.24-4.41 (m, 2H), 4.57 (m, 1H), 5.07 (s, 2H), 5.16 (d, *J* = 8.4 Hz, 1H), 6.63 (d, *J* = 8.0 Hz, 1H), 6.78 (d, *J* = 8.0 Hz, 1H), 6.88 (d, *J* = 7.8 Hz, 1H), 6.95 (d, *J* = 7.8 Hz, 1H), 7.22-7.35 (m, 9H), 7.47-7.49 (m, 2H), 7.69 (d, *J* = 7.4 Hz, 2H); ^{13}C NMR ($CDCl_3$) δ 24.34, 28.55, 37.49, 47.39, 47.67, 55.25, 57.12, 67.60, 67.86, 68.04, 115.19, 120.56, 125.64, 125.70, 127.64, 128.30, 128.43, 128.58, 128.75, 128.80, 129.06, 129.16, 131.00, 136.94, 137.16, 141.85, 144.26, 155.86, 156.36, 158.41, 177.45; Rt (HPLC-A): 4.59 (100%).

5-benzyloxy-1-pentanol (45'). It was synthesized according to the same procedure than **44'**, with 3.84 g (96 mmol) of sodium hydride (60% in mineral oil), 50 g of 1,5-pentanediol (**42'**) (5 equiv, 480 mmol) and 11.42 mL of benzyl bromide (96 mmol). After purification by column chromatography on silica gel (eluted with DCM and DCM/MeOH: 97/3), 9.5 g (51%) of **45'** were obtained as a slightly yellow oil: Rf (DCM/MeOH: 97/3) = 0.33; ^{1}H NMR ($CDCl_3$) δ 1.29-1.60 (m, 6H), 2.37 (s, 1H), 3.39 (t, *J* = 6.4 Hz, 2H), 3.49 (t, *J* = 6.4 Hz, 2H), 4.41 (s, 2H), 7.16-7.24 (m, 5H); ^{13}C NMR ($CDCl_3$) δ 22.97, 29.98, 32.97, 63.04, 70.90, 73.49, 128.15, 128.27, 128.95, 138.99.

2-((2-benzyloxyethyl)thio)ethanol (46'). It was synthesized according to the same procedure than **44'**, with 3.27 g (82 mmol) of sodium hydride (60% in mineral oil), 50 g of 2,2'-thiodiethanol (**43'**) (5 equiv, 410 mmol) and 9.73 mL of benzyl bromide (82 mmol). After purification by column chromatography on silica gel (eluted with DCM and DCM/MeOH: 98/2), 6.2 g (36%) of **46'** were obtained as a colorless oil: Rf (DCM/MeOH: 98/2) = 0.41; ^{1}H NMR ($CDCl_3$) δ 2.70 (t, *J* = 6.0 Hz, 2H), 2.73 (t, *J* = 6.9 Hz, 2H), 2.87 (s, 1H), 3.61 (t, *J* = 6.4 Hz, 2H), 3.67 (t, *J* = 6.0 Hz, 2H), 4.52 (s, 2H), 7.24-7.32 (m, 5H); ^{13}C NMR ($CDCl_3$) δ 32.11, 36.31, 61.48, 70.43, 73.69, 128.58, 128.99, 138.67.

Table S1. ELISA IC_{50} values, obtained by competition against sCD4 on gp120SF162 and gp120CN54, of the miniCD4s belonging to the M33 (**16**) series

No		R	gp120SF162 IC_{50} (nM ± SEM)	gp120CN54 IC_{50} (nM ± SEM)	A: gp120SF162 $\frac{IC_{50}\ M33x}{IC_{50}\ M33AT}$	B: gp120CN54 $\frac{IC_{50}\ M33x}{IC_{50}\ M33AT}$	$\frac{IC50_{CN54}}{IC50_{SF162}}$	B/A
16	**M33**[b]		3 ± 0.4	41 ± 4	0.9	0.6	14	0.7
17	**M33A1**[b]		3.2 ± 0.3	66 ± 1	1	1	21	1
18	**M33A2***		146 ± 18	12600 ± 1300[a]	46	192	87	4.2
19	**M33A3***		44 ± 4	3882 ± 735	14	59	88	4.2
20	**M33A4***		5.6 ± 0.5	1321 ± 133	1.8	20	236	11
21	**M33A5***		78 ± 8	1866 ± 202	24	28	24	1.2
22	**M33A6***		14 ± 1	1946 ± 188	4.4	29	139	6.6
23	**M33A7***		13 ± 2	1730 ± 331	4.1	26	133	6.3
24	**M33A8***		471 ± 98	21400 ± 1000[a]	147	325	46	2.2

[a]Extrapolated values (maximum concentration tested was 10 mM). [b]Mentioned in reference (13); MiniCD4 **17** was previously called [Phe23]M33.[16] *New compounds. IC_{50} values are presented as the mean ± SEM values from three independent experiments.

Table S2: Characterization of the M33 (**16**) miniCD4 series

MiniCD4	Rt (min)	Formula	MW calc.	MW found	Purity % (214 nm)[a]
16	22.48	$C_{128}H_{198}N_{36}O_{31}S_6$	2929.6	2929.6 ± 0.2	98.4
17	18.08	$C_{122}H_{194}N_{36}O_{31}S_6$	2853.5	2853.5 ± 0.1	97.0
18	19.35	$C_{123}H_{196}N_{36}O_{31}S_6$	2867.5	2867.1 ± 0.4	98.7
19	19.45	$C_{124}H_{196}N_{36}O_{31}S_6$	2879.5	2879.5 ± 0.3	99.8
20	20.54	$C_{126}H_{196}N_{36}O_{31}S_6$	2903.5	2903.5 ± 0.3	99.4
21	19.70	$C_{124}H_{195}N_{37}O_{31}S_6$	2892.5	2892.5 ± 0.1	96.3
22	18.84	$C_{123}H_{196}N_{36}O_{31}S_6$	2867.5	2867.6 ± 0.2	98.3
23	19.24	$C_{123}H_{196}N_{36}O_{32}S_6$	2883.5	2883.5 ± 0.2	98.2
24	22.80	$C_{126}H_{202}N_{36}O_{31}S_6$	2909.6	2909.5 ± 0.2	95.1

[a]Purity was checked by measurements of peak area at 214 nm on reverse-phase HPLC.

Table S3: Characterization of the M48 (**1**) miniCD4 series

MiniCD4	Rt (min)	Formula	MW calc.	MW found	Purity % (214 nm)[b]
1	18.73	$C_{126}H_{200}N_{38}O_{31}S_6$	2935.6	2935.7 ± 0.1	99.7
2	28.80	$C_{133}H_{212}N_{38}O_{32}S_6$	3047.7	3047.6 ± 0.2	97.2
3	28.92	$C_{133}H_{212}N_{38}O_{32}S_6$	3047.7	3047.8 ± 0.1	99.2
4	26.50	$C_{134}H_{208}N_{38}O_{32}S_6$	3055.7	3055.5 ± 0.2	98.5
5	30.45	$C_{134}H_{214}N_{38}O_{32}S_6$	3061.8	3061.7 ± 0.3	100
6	32.13	$C_{135}H_{216}N_{38}O_{32}S_6$	3075.8	3075.6 ± 0.2	96.0
7	17.54	$C_{130}H_{208}N_{38}O_{33}S_6$	3023.7	3023.7 ± 0.1	97.4
8	19.08	$C_{131}H_{210}N_{38}O_{33}S_6$	3037.7	3037.5 ± 0.2	97.2
9	18.55	$C_{130}H_{208}N_{38}O_{33}S_7$	3055.7	3055.6 ± 0.2	95.8
10	15.50	$C_{131}H_{209}N_{39}O_{32}S_6$	3034.7	3034.8 ± 0.1	99.8
11	23.45	$C_{129}H_{206}N_{38}O_{31}S_6$	2977.7	2977.7 ± 0.1	97.1
12	24.80	$C_{130}H_{208}N_{38}O_{31}S_6$	2991.7	2991.7 ± 0.2	97.6
13	20.11	$C_{133}H_{213}N_{39}O_{31}S_6$	3046.7	3046.5 ± 0.2	99.4
14	24.84	$C_{132}H_{204}N_{38}O_{31}S_6$	3011.7	3011.8 ± 0.2	99.4
15	15.72/15.92[a]	$C_{130}H_{208}N_{38}O_{34}S_7$	3071.7	3071.6 ± 0.3	96.3

[a] **15** was purified as a racemic product engendered by the sulfoxyde group. [b] Purity was checked by measurements of peak area at 214 nm on reverse-phase HPLC.

Table S4: Crystallographic data collection and refinement statistics

	YU2gp120core$_e$ + miniCD4 **13**
Data collection	
Space group	$C222_1$
Cell constants	
a, b, c (Å)	64.4, 163.4, 78.2
α, β, γ (°)	90.0, 90.0, 90.0
Wavelength (Å)	1.00
Resolution (Å)	50.0-1.90 (1.93-1.90)
R_{sym}	12.8 (55.8)
$I / \sigma I$	12.4 (1.3)
Completeness (%)	92.3 (51.9)
Redundancy	5.5 (1.9)
Refinement	
Resolution (Å)	17.5-1.79 (1.84-1.79)
Unique reflections	31,931 (517)
R_{free} / R_{work} (%)	22.4/16.8
No. atoms	
gp120	3,208
MiniCD4 **13**	245
Ligand/ion	0
Water	236
B-factors (Å^2)	
Overall	43.27
gp120	41.96
MiniCD4 **13**	59.41
Residue $23_{\mathbf{13}}$	31.94
Water	44.38
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.3
Ramachandran	
Favored (%)	98.0
Outliers (%)	0

Values in parentheses are for highest-resolution shell.
$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$, where I is the observed intensity, and <I> is the average intensity of multiple observations of symmetry-related reflections.
$R = \Sigma hkl||Fobs| - |Fcalc||/\Sigma hkl||Fobs|$.
R_{free} is calculated from 5% of the reflections excluded from refinement.

Figure S1. Schematic diagram of the folding pattern of miniprotein **1**

Tpa = thiopropionyl. Position 23 to be optimized in the M48 (**1**) series is marked in red. Disulfide bridges are shown by connecting lines.

Figure S2. Surface plasmon resonance (SPR) analysis of the binding of miniCD4s **2** and **13** to HIV-1 gp120. gp120 was immobilized at a level of 1000-2000 RU on CM5 chips. The black lines indicate independent injections of the miniCD4s. The red lines show the global fit of the data to a Langmuir 1:1 binding model. The following concentrations were sampled for **2**: YU2 gp120- 3 nM, 1.5 nM, 0.75 nM and 0.38 nM; TRO.11 gp120- 32 nM to 0.25 nM at 2-fold dilution; ZM135- 32 nM to 0.06 nM at 2-fold dilution; and for **13**: YU2 gp120- 3.6 nM, 1.8 nM, 0.45 nM, 0.11 nM, 0.056 nM and 0.028 nM, TRO.11 gp120- 3.6 nM to 0.028 nM at 2-fold dilution, ZM135 gp120- 3.6 nM to 0.028 nM at 2-fold dilution.

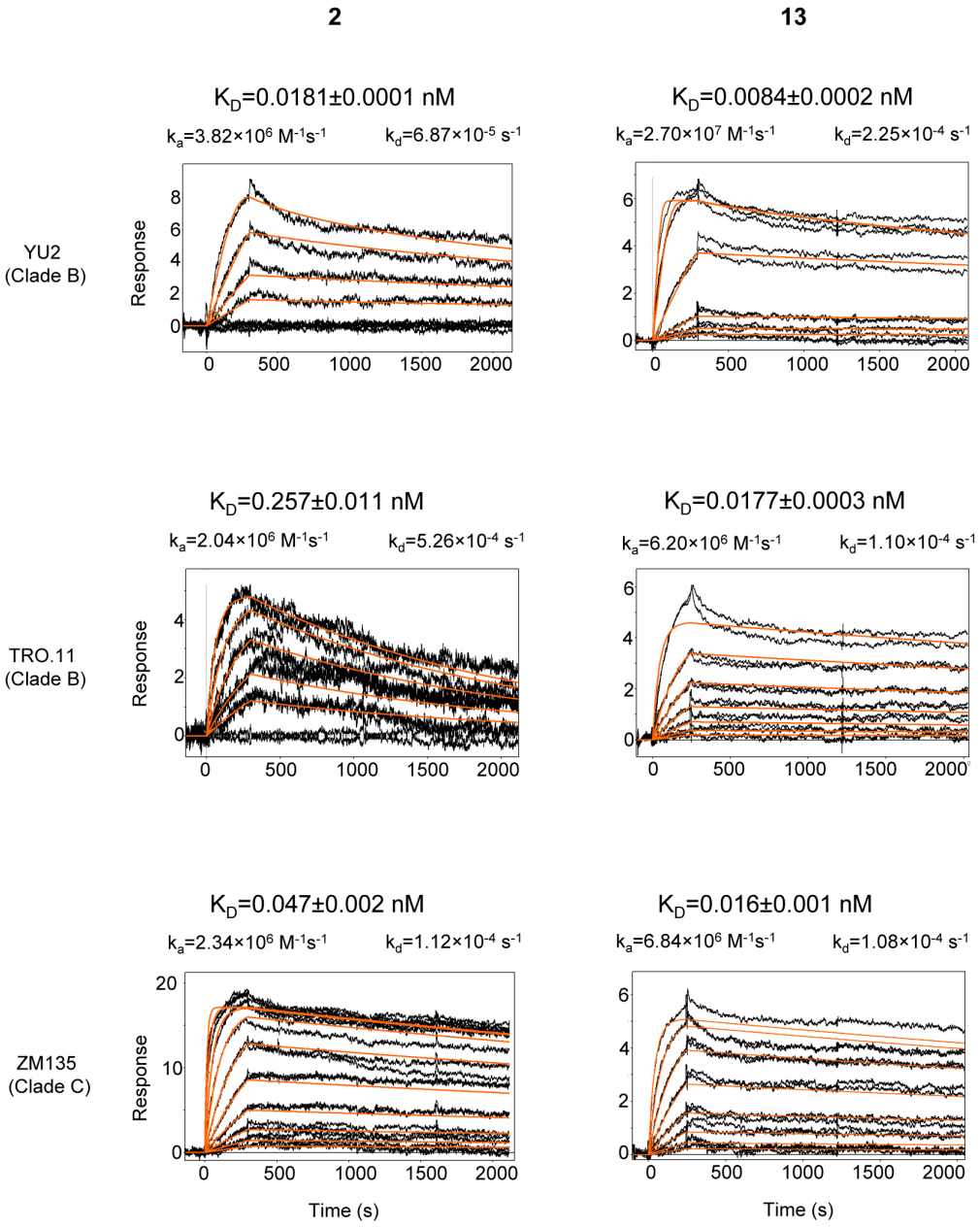

Figure S3. Comparison of **2** and **13** bound to the gp120 Phe43 cavity. (A and B) 90° rotated views of superposition of gp120 molecules in the **2**-gp120 and **13**-gp120 complex structures. gp120 chains are shown as grey cartoon, miniCD4 backbone is shown as ribbon, miniCD4 residue 23 side chain is shown in stick representation with **2** in green (with the ether oxygen colored red) and **13** in orange (with the aniline nitrogen colored blue). The water molecules occupying the conserved gp120 solvent channel are shown as blue spheres with the water molecules in the **2**-gp120 complex shown in a lighter shade of blue than the waters from the **13**-gp120 complex. Side chains of (C) **2** and (D) **13** bound to the Phe43 cavity. The blue mesh represents the 2Fo-Fc map electron density map contoured at 2.5 σ for the **2**-gp120 structure and at 2.7 σ for the **13**-gp120 structure.

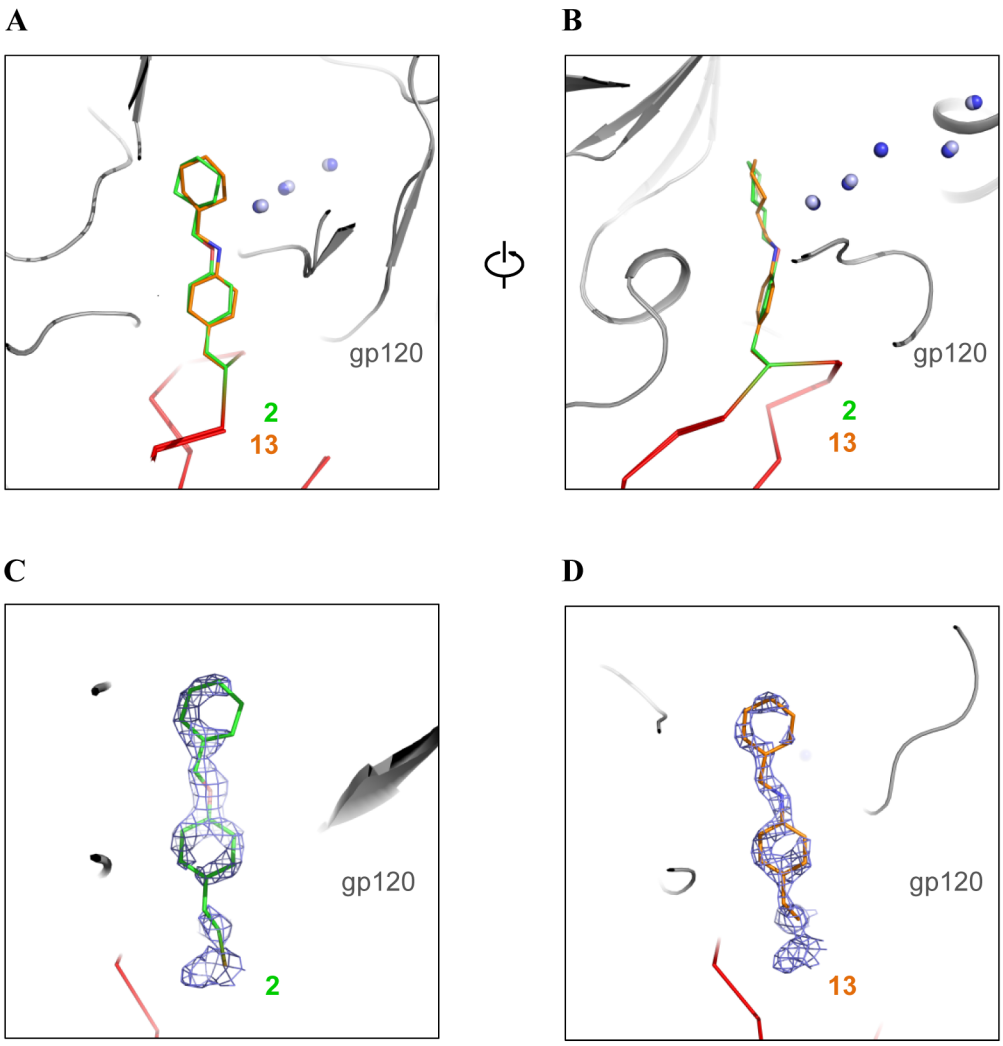

Figure S4. Definition of the insert side chain of **13** bound to the gp120 Phe43 cavity. Zoomed-in view of **13** bound to gp120 shown in cartoon representation. MiniCD4 **13** is colored red, gp120 is colored by domain with the inner domain colored cyan, outerdomain green and the bridging sheet pink. The grey mesh represents the 2Fo-Fc electron density map defining the location of the side chain in the gp120 Phe43 cavity contoured at (A) 2.0 σ, (B) 2.5 σ and (C) 3.0 σ .

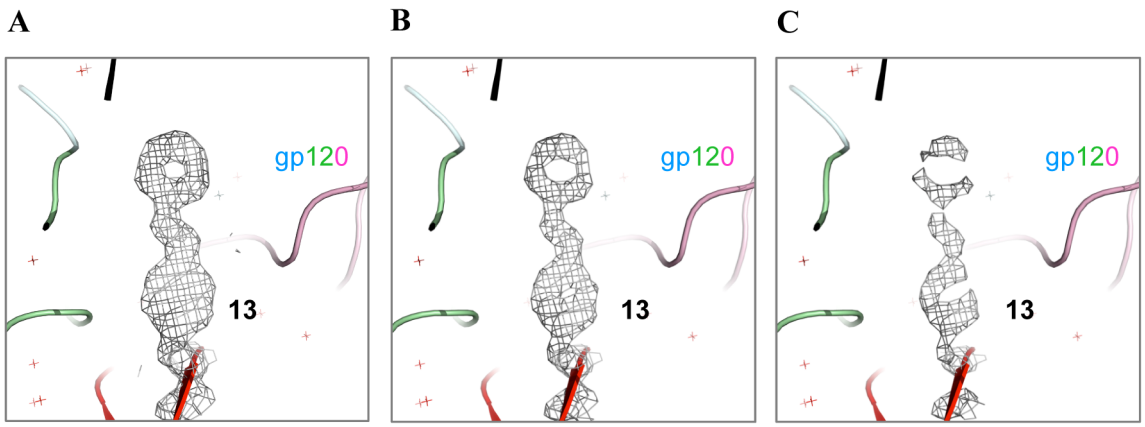

Figure S5. Extent of Phe43 cavity filling by **13**. A cross-section of gp120 in grey surface representation shows (A-C) a view of **13** bound to the Phe43 cavity. MiniCD4 **13** is shown as red cartoon with the side chain shown in stick representation with the carbon atoms colored orange, nitrogen blue and hydrogens white. Pocket-Finder was used to define the binding cavities. Ligand volume was calculated in the ChemBio3D module of ChemBioOffice 2010. (B) The side chain binding cavity (grid points shown as blue spheres) was defined by using gp120 coordinates from the **13**-gp120 complex structure as input into the pocket detection algorithm. The volume of the ligand was determined for atoms spanning the phenyl ring and the cyclohexane ring. (C) The cavity that holds the cyclohexane ring was defined by using cyclohexane ring deleted **13**-gp120 complex coordinates as Pocket-Finder input. Packing Coefficient (%) = (Volume of ligand/Volume of binding pocket) × 100.

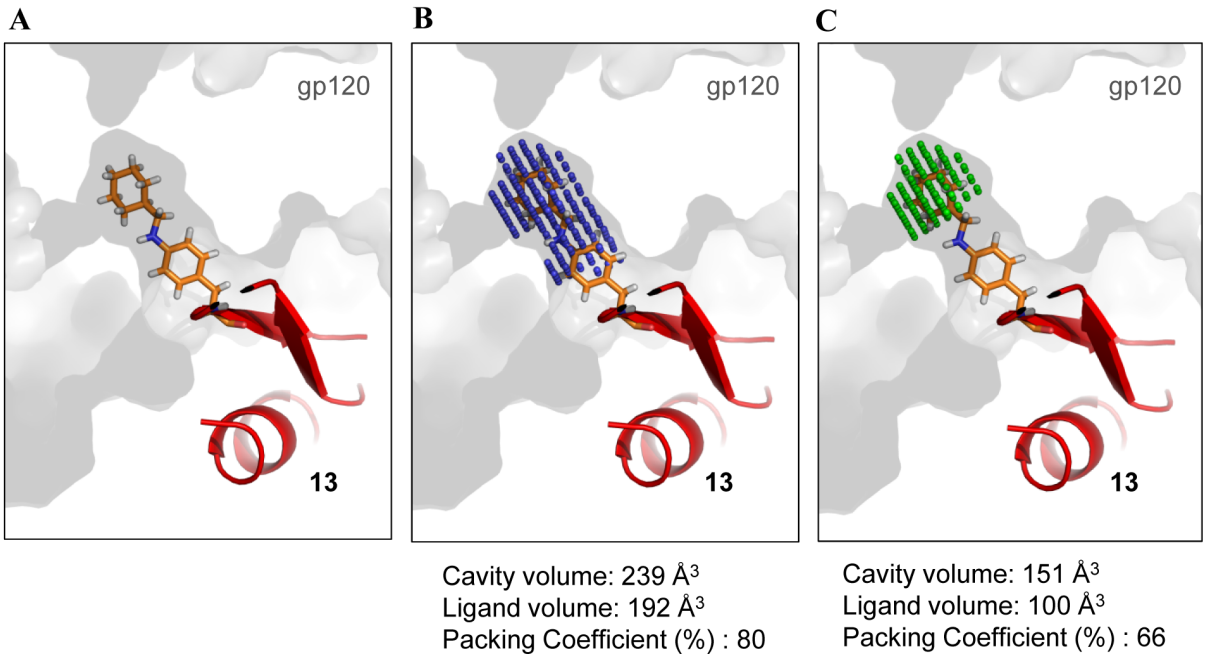

Figure S6. Sequence alignment of gp120HXBc2, gp120SF162 and gp120CN54. The alignment was made with ClustalX software and uses therefore its three characters code to mark conserved residues. Asterisk stands for fully conserved; colon for strong groups conservation; dot for weaker groups conservation. The absence of character means no conservation. Residues are colored red when the side chain lines the cavity, and green when the main chain lines the cavity. An arrow points position 426 (HXBc2 numbering).