Supporting Information

Nucleophilic Catalysis of Hydrazone Ligation and Transimination: Implications for Dynamic Covalent Chemistry.

Anouk Dirksen,[a,b] Sjoerd Dirksen,[b] Tilman M. Hackeng,[b] Philip E. Dawson[a,*]

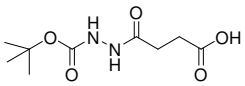
[a] *Departments of Cell Biology and Chemistry, Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA;* [b] *Cardiovascular Research Institute Maastricht, University Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands.*

Solvents and Starting Materials. Unless stated otherwise, all reagents and solvents were purchased from commercial sources and used without purification.

Instrumentation. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Varian Mercury 300 MHz spectrometer at 298 K. Reversed phase high pressure liquid chromatography (RP HPLC) was performed on a Varian Pro Star HPLC system (214 nm) (reactions at pH 5.7) or a HP1050 HPLC System (214 nm) (reactions at pH 4.5). Electrospray ionization mass spectrometry (ESI-MS) was performed on a SCIEX API-150 EX single quadruple mass spectrometer or on a SCIEX API-III triple quadruple mass spectrometer.

Synthesis. The peptides were obtained *via* manual solid phase peptide synthesis (SPPS) using the *in situ* neutralization/ 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation procedure for *t*Boc chemistry on a *p*-methylbenzhydrylamine (MBHA) resin.[1] Glyoxylyl-LYRAG **2**[2] and Ac-GRGDSGG-MPAL[3] (MPAL = mercaptopropionic acid-leucine) were synthesized as previously reported.

Ac-GRGDSGG-hydrazide (1). A solution of 10 µL of hydrazine in 5 mL of H_2O was added to 105 mg (0.12 mmol) Ac-GRGDSGG-MPAL. Within 1.5 h the reaction went to completion. Purification by preparative RP HPLC over a C18 column (gradient: 0–10% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 60 min; flow: 20 mL/min) and subsequent lyophilization gave 47 mg (0.071 mmol, 59%) of **1** as white powder. ESI-MS calcd. for $C_{23}H_{41}N_{12}O_{11}$ ($[M+H]^+$): 661.6, found 661.6.

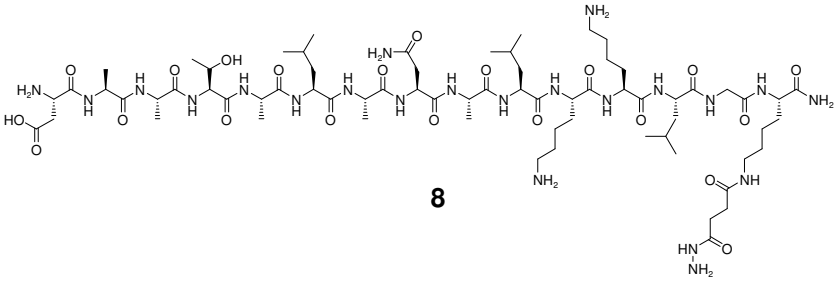


12

Synthesis of 4-(*N*'-*tert*-butoxycarbonyl-hydrazino)-4-oxo-butyric acid (12). 5.0 g (50 mmol) of succinic anhydride and 6.608 g (50 mmol) of *tert*-butyl carbazate were dissolved in 100 mL water. Within 10 min the solution became clear. The reaction proceeded clean, without any side product formation and full conversion to **12** was

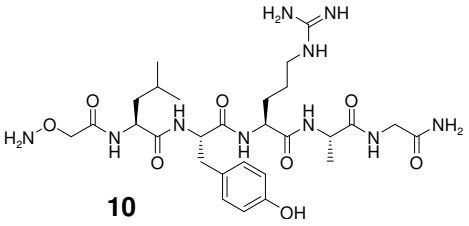
confirmed by ESI-MS. The reaction mixture was lyophilized and **12** was used in SPPS without any further purification. ^{1}H NMR (dmso-d_6, 300.08 MHz): δ (ppm) = 1.38 (s, 9H, C(CH_3)$_3$), 2.29 (t, 2H, CH_2COOH), 2.41 (t, 2H, CH_2C=ONH), 8.68 (s, 1H, N*H*-*t*Boc), 9.52 (s, 1H, N*H*C=OCH$_2$), 11.0 (bs, 1H, COO*H*); ^{13}C NMR (dmso-d_6, 75.46 MHz): δ (ppm) = 28.1, 28.8, 29.0, 79.1, 155.3, 170.8, 173.6; ESI-MS calcd. for $C_5H_{11}N_2O_2$ ([M-H]$^-$): 231.2, found 231.1.

H$_2$N-DAATALANALKKLGK(hydrazide)-amide (8). *t*Boc-DAATALANALKKLGK(Fmoc) was synthesized on a MBHA resin using *t*Boc chemistry on a 0.3 mmol scale. The Fmoc group was removed by a 4 × 3 min treatment with 20 *v*-% piperidine in DMF. 261 mg (1.12 mmol) of **12** was dissolved in 0.5 mL of DMF and a separate solution of 572 mg (1.1 mmol) of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP®) and 167 mL of DIEA in 1 mL DMF was prepared. Both solutions were added directly to the resin (no preactivation of **12**). The reaction was continued for 1.5 h and the resin was subsequently washed with DMF. The *t*Boc groups were removed with a 2 × 1 min treatment with TFA. After a DMF, DCM, and a 1:1 *v/v* MeOH/DCM flow wash, the resin was dried under vacuum. HF cleavage (4 *v*-% of *p*-cresol added as a scavenger) and subsequent lyophilization gave **8** as crude product, which was purified by preparative RP HPLC over a C18 column (gradient: 30–50% 9:1 *v/v* MeCN/H$_2$O in H$_2$O, 0.1% TFA in 90 min; flow: 10 mL/min). ESI-MS calcd. for $C_{69}H_{125}N_{22}O_{21}$ ([M+H]$^+$): 1598.9, found 1598.1.



8

Aminooxyacetyl-LYRAG (10). H_2N-LYRAG was synthesized on a MBHA resin using *t*Boc chemistry on a 0.3 mmol scale. The resin was neutralized with 10 *v*-% DIEA in DMF followed by a flow wash with DMF. (*t*Boc-aminooxy)acetic acid (385 mg, 2.0 mmol) was activated with 310 µL of *N*,*N*'-diisopropylcarbodiimide (d = 0.815 kgL^{-1}, 2.0 mmol) and 227 mg (2.0 mmol) of *N*-hydroxysuccinimide in 4 mL of DMF during 10 min and subsequently added to the resin. Coupling was continued for 1 h. After a DMF flow wash, the *t*Boc groups were removed with a 2 × 1 min TFA treatment. After a DMF, DCM, and a 1:1 *v/v* MeOH/DCM flow wash, the resin was dried under vacuum. HF cleavage (4 *v*-% of *p*-cresol added as a scavenger) and subsequent lyophilization gave **10** as crude product, which was purified by preparative RP HPLC over a C18 column (gradient: 5–25% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 90 min; flow: 10 mL/min). ESI-MS calcd. for $C_{28}H_{47}N_{10}O_8$ ($[M+H]^+$): 651.7, found 651.4.



10

Covalent capture of the aniline Schiff base 4 with $NaBH_3CN$. To a ligation mixture of equimolar amounts of hydrazide **1** and glyoxylyl **2** (1 mM each) and 10 mM aniline in a 0.1 M NH_4OAc buffer (pH 4.5) in equilibrium with hydrazone **3**, a 10-fold excess of $NaBH_3CN$ was added. The reaction was followed by HPLC (214 nm) (Phenomenex Prodigy™ C18 column; gradient: 5–20% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 15 min; flow: 3 mL/min). After 25 h the aniline Schiff base **4**, invisible under dynamic conditions, was for > 90 % reduced to the corresponding *N*-acetamide **6**. The conversion was calculated from the integrals of the HPLC signals (214 nm). Neither the reduced hydrazone **7**, nor the reduced glyoxylyl **5** was observed by HPLC or ESI-MS after 25 h, only a small amount (~ 10 %) of hydrazone **3**. **4**: ESI-MS calcd. for $C_{34}H_{51}N_{10}O_7$ ($[M+H^+]$): 711.8, found 711.4.

Hydrazone reactions. A 4 mM stock solution of Ac-GRGDSGG-hydrazide **1** (**I**) and a 4 mM stock solution of glyoxylyl-LYRAG **2** (**II**) were prepared in a 0.1 M NH_4OAc buffer (either pH 4.5 or pH 5.7).

1 mM peptide concentration (pH 4.5 or pH 5.7): for the reactions performed in the absence of aniline, 250 µL of **I** and 250 µL of **II** were added to 500 µL 0.1 M NH_4OAc buffer. For the reactions performed in the presence of 10 mM aniline, 250 µL of **I** and 250 µL of **II** were sequentially added to 500 µL 0.1 M NH_4OAc buffer containing 20 mM of aniline.

0.1 mM peptide concentration (pH 5.7): for the reaction performed in the absence of aniline 25 µL of **I** and 25 µL of **II** were added to 950 µL 0.1 M NH_4OAc buffer. For the reaction performed in the presence of 10 mM aniline, 450 µL 0.1 M NH_4OAc buffer, 25 µL of **I**, and 25 µL of **II** were sequentially added to 500 µL 0.1 M NH_4OAc buffer containing 20 mM of aniline.

The reactions were performed at room temperature (RT) and followed by HPLC (214 nm); reactions at pH 4.5: Phenomenex Prodigy™ C18 column; gradient: 5–15% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 10 min; flow: 3 mL/min; reactions at pH 5.7: Vydac™ C18 column; gradient: 0–67% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 30 min; flow: 1 mL/min. The concentrations and conversions were calculated from the integrals of the HPLC signals (214 nm). The formation of hydrazone **3** was confirmed by ESI-MS. ESI-MS calcd. for $C_{51}H_{82}N_{21}O_{18}$ ($[M+H^+]$): 1277.3, found 1276.5.

Hydrazone reactions at pH 4.5.

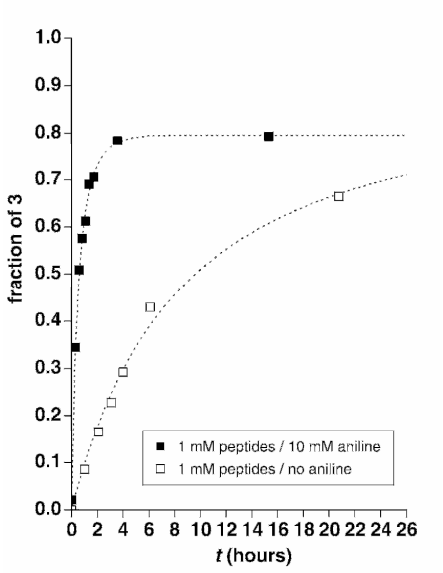


Figure 1. Formation of hydrazone **3** over time at 1 mM reactant concentration at RT in a 0.1 M NH_4OAc buffer (pH 4.5) in the absence (□) and in the presence (■) of 10 mM aniline.

Fitting of the data of the hydrazone reactions. The reaction between Ac-GRGDSGG-hydrazide **1** and glyoxylyl-LYRAG **2** is a bimolecular reversible reaction following 2nd order reaction kinetics. Since [**1**] = [**2**] and [**3**] = $[\mathbf{2}]_0$-[**2**] at any time, the rate equation for the disappearance of **2** is given by:

$$\frac{d[\mathbf{2}]}{dt} = -k_1[\mathbf{2}]^2 + k_{-1}([\mathbf{2}]_0-[\mathbf{2}])$$

The reaction rates, k_1 and k_{-1}, were calculated from the disappearance of **2** in time, by fitting [**2**] (M) against *t* (s) to the solution of the rate equation (see below for details):

$$x(t) = \frac{a_+(x_0 - a_-) - a_-(x_0 - a_+)e^{-k_1(a_+ - a_-)t}}{(x_0 - a_-) - (x_0 - a_+)e^{-k_1(a_+ - a_-)t}} \qquad (4)$$

In which

$$a_+ = \frac{-k_{-1} + \sqrt{k_{-1}^2 + 4k_1 k_{-1} x_0}}{2k_1}$$

$$a_- = \frac{-k_{-1} - \sqrt{k_{-1}^2 + 4k_1 k_{-1} x_0}}{2k_1}$$

and $x(t) = [\mathbf{2}]$ and $x_0 = [\mathbf{2}]_0$ in this particular case. The errors calculated for the rate constants are standard errors that take into account the number of data points. Equation (4) was programmed into Origin version 7.5 (OriginLab®), which was used as the software for the data analysis.

The equilibrium constants (K_{eq}) of the hydrazone reactions were calculated from the rate constants, since $K_{eq} = k_1/k_{-1}$:

1 mM peptides, no aniline, pH 5.7: $K_{eq} = (2.8 \pm 1.4) \times 10^4\ M^{-1}$

1 mM peptides, 10 mM aniline, pH 5.7: $K_{eq} = (5.3 \pm 0.8) \times 10^4\ M^{-1}$

0.1 mM peptides, 10 mM aniline, pH 5.7: $K_{eq} = (5.7 \pm 1.4) \times 10^4\ M^{-1}$

0.1 mM peptides, no aniline, pH 5.7: K_{eq} could not be calculated accurately, because the back rate of the reaction (k_{-1}) could not be derived due to insufficient data points and due to the slow rate of equilibration.

1 mM peptides, no aniline, pH 4.5: $K_{eq} = (3.0 \pm 3.0) \times 10^4\ M^{-1}$

1 mM peptides, 10 mM aniline, pH 4.5: $K_{eq} = (1.9 \pm 0.3) \times 10^4\ M^{-1}$

General solution of the rate equation. The rate equation for a reversible bimolecular reaction was derived in the following manner. k_1 ($M^{-1}s^{-1}$) is the forward rate of the hydrazone reaction (hydrazone formation), k_{-1} (s^{-1}) the back rate (hydrolysis), t the time in s, $x(t)$ the concentration of peptides at time t in M, and x_0 the concentration of peptides at $t = 0$ in M, *i.e.* $x(0) = x_0$. The rate of change in the concentration of peptides is modelled by the differential equation

$$\frac{dx}{dt} = -k_1 x^2 + k_{-1}(x_0 - x) \qquad (1)$$

In the initial condition $x(0)$ equals x_0 and (1) can be written as

$$\frac{dx}{dt} = -k_1 x^2 - k_{-1} x + k_{-1} x_0$$

The right hand side of this equation is a second degree polynomial which, by the discriminant formula, has two roots

$$a_+ = \frac{-k_{-1} + \sqrt{k_{-1}^2 + 4k_1 k_{-1} x_0}}{2k_1}$$

$$a_- = \frac{-k_{-1} - \sqrt{k_{-1}^2 + 4k_1 k_{-1} x_0}}{2k_1}$$

Since k_1, k_{-1}, $x_0 > 0$, both a_+ and a_- are real. Also, $a_+ > 0$ and $a_- < 0$. Hence, if $x(0) = a_+$, then $dx/dt = 0$ and we obtain the constant solution $x(t) \equiv a_+$. Recognizing (1) as a separable differential equation in one variable, we proceed to a general solution by the method of separation of variables. If $I \subset \mathbb{R}$ is an open interval such that $0 \in I$ and $-k_1 x^2 - k_{-1} x + k_{-1} x_0 \neq 0$, for all $t \in I$, then, for $t \in I$ we have

$$\int_{x_0}^{x} \frac{1}{-k_1 \xi^2 - k_{-1} \xi + k_{-1} x_0} d\xi = \int_0^t 1 d\nu = t$$

Multiplying both sides by $-k_1$ and factoring the polynomial in the divisor of the integrand on the left hand side, we obtain

$$\int_{x_0}^{x} \frac{1}{(\xi - a_+)(\xi - a_-)} d\xi = -k_1 t \qquad (2)$$

Now,

$$\frac{1}{\xi - a_+} - \frac{1}{\xi - a_-} = \frac{(\xi - a_-) - (\xi - a_+)}{(\xi - a_+)(\xi - a_-)} = \frac{(a_+ - a_-)}{(\xi - a_+)(\xi - a_-)}$$

So (2) is equivalent with

$$\frac{1}{a_+ - a_-} \int_{x_0}^{x} \frac{1}{(\xi - a_+)} - \frac{1}{(\xi - a_-)} d\xi = -k_1 t$$

This leads to

$$\frac{1}{a_+ - a_-}(\log|x - a_+| - \log|x_0 - a_+| - \log|x - a_-| + \log|x_0 - a_-|) = -k_1 t$$

and we finally obtain by rearranging and taking the exponential on both sides

$$\frac{|x - a_+|}{|x - a_-|}\frac{|x_0 - a_-|}{|x_0 - a_+|} = e^{-k_1(a_+ - a_-)t} \qquad (3)$$

We can focus now on three separate cases: $x_0 > a_+$, $a_+ > x_0 > a_-$, and $a_- > x_0$. Since we require that $x_0 > 0$, we dismiss the third possibility $a_- > x_0$. When $x_0 > a_+$, the absolute value signs in (3) can be dropped and rewriting gives the solution

$$x(t) = \frac{a_+(x_0 - a_-) - a_-(x_0 - a_+)e^{-k_1(a_+ - a_-)t}}{(x_0 - a_-) - (x_0 - a_+)e^{-k_1(a_+ - a_-)t}} \qquad (4)$$

Note that for $t \geq 0$, the denominator of x is strictly larger than 0, since $x_0 - a_- > x_0 - a_+$ and

$$e^{-k_1(a_+ - a_-)t} \leq 1$$

Hence, $x(t)$ exists for $t \geq 0$, *i.e.* $[0,\infty) \subset I$. Taking the limit for $t \to \infty$ in (4) the equilibrium concentration is given by

$$\lim_{t \to \infty} x(t) = a_+$$

Finally, we take the derivative of x and arrive at

$$x'(t) = \frac{k_1(a_+ - a_-)(a_- - a_+)(\frac{x_0 - a_-}{x_0 - a_+})e^{-k_1(a_+ - a_-)t}}{((\frac{x_0 - a_-}{x_0 - a_+}) - e^{-k_1(a_+ - a_-)t})^2}$$

Since $x'(t) < 0$ for all $t \geq 0$, x is monotonically decreasing in t.

Analogously, the second case has the solution

$$x(t) = \frac{a_+(x_0 - a_-) + a_-(a_+ - x_0)e^{-k_1(a_+ - a_-)t}}{(x_0 - a_-) + (a_+ - x_0)e^{-k_1(a_+ - a_-)t}}$$

Similar to the above, it can be shown that x is monotonically increasing in t and

$$\lim_{t\to\infty} x(t) = a_+$$

To complete the mathematical analysis of the kinetics, suppose we have two different reaction samples with two different pairs of forward and back rates of the hydrazone reaction, denoted as k_1, k_{-1} and k^*_1, k^*_{-1}, respectively. By dividing the numerator and denominator of a_+ by k_1, we get

$$a_+ = \frac{-\frac{k_{-1}}{k_1} + \sqrt{(\frac{k_{-1}}{k_1})^2 + 4\frac{k_{-1}}{k_1}x_0}}{2} = \frac{-K_{eq}^{-1} + \sqrt{K_{eq}^{-2} + 4K_{eq}^{-1}x_0}}{2}$$

Where $K_{eq} = k_1/k_{-1}$ is the equilibrium constant of the reaction. Hence, if the equilibrium constants are equal, *i.e.* $K_{eq} = K^*_{eq}$, and the initial concentrations are equal ($x_0 = x^*_0$), then it follows that the equilibrium concentrations of the two reactions are equal, *i.e.* $a_+ = a^*_+$.

Transimination experiment (Figure 2 in the main text). A 4 mM stock solution of Ac-GRGDSGG-hydrazide **1** (**I**) and a 4 mM stock solution of glyoxylyl-LYRAG **2** (**II**) were prepared in a 0.1 M NH_4OAc buffer (pH 4.5).

Ad Figure 2A: 500 µL of **I** and 500 µL of **II** were added to 1 mL 0.1 M NH_4OAc buffer containing 20 mM aniline (pH 4.5). The reaction was performed at RT and followed by HPLC (214 nm) (Phenomenex Prodigy™ C18 column; gradient: 5–15% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 10 minutes; flow: 3 mL/min) until equilibrium was reached.

Ad Figure 2B: 2.85 mg (1.78 µmol) of hydrazide **8** was dissolved in 1.8 mL of the ligation mixture of *Figure 2A*. An additional 1.64 µL of aniline were added (final concentration: 20 mM). The reaction was continued at RT and followed by analytical RP HPLC (214 nm) (Phenomenex

Prodigy™ C18 column; gradient: 5–15% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 10 minutes; flow: 3 mL/min) until a new equilibrium was reached. The formation of hydrazone **9** was confirmed by ESI-MS. ESI-MS calcd. for $C_{97}H_{166}N_{31}O_{28}$ ([M+H^+]): 2214.5, found 2214.0.

Ad Figure 2C: 1.0 mg (1.5 μmol) of aminooxyacetyl-LYRAG **10** was dissolved in 1.4 mL of the ligation mixture of *Figure 2B*. An additional 10.2 μL of aniline were added (final concentration: 100 mM). The reaction was continued at RT and followed by HPLC (214 nm) (Phenomenex Prodigy™ C18 column; gradient: 5–15% 9:1 *v/v* MeCN/H_2O in H_2O, 0.1% TFA in 10 minutes; flow: 3 mL/min) until a new equilibrium was reached. The formation of oxime **11** was confirmed by ESI-MS. ESI-MS calcd. for $C_{56}H_{88}N_{19}O_{15}$ ([M+H^+]): 1267.4, found 1266.5.

The concentrations and conversions were calculated from the integrals of the HPLC signals (214 nm). Increasing amounts of aniline were used in the transimination experiment to compensate for the varying equilibration rates. The higher the concentration of aniline in solution the greater the rate enhancements.

[1] Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Peptide Protein Res.*, 1992, **40**, 180-193.

[2] Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2006**, in press.

[3] Dirksen, A.; Meijer, E. W.; Adriaens, W.; Hackeng, T. M. *Chem. Commun.* **2006**, (15), 1667-1669.