Identification of Sequence-Selective Receptors for Peptides with a Carboxylic Acid Terminus**

Thilo Fessmann and Jeremy D. Kilburn*

The ability to prepare synthetic receptors for specific ligands is a highly desirable goal. In particular selective receptors for specific peptide sequences would have potential applications for separation of peptide mixtures, biosensors, and new therapeutics, as well as providing model systems for biological protein–peptide complexes. Much recent work in the area of peptide receptors has focused on "tweezer" receptors,[1, 2] which, despite their inherent flexibility, have proved to be highly selective for certain peptide sequences in both nonpolar[3] and aqueous solvent systems.[4] Many of these tweezer receptors have been screened against combinatorial libraries of resin-bound peptides and have allowed a rapid evaluation of the binding properties of the receptor.[3, 4a] The reverse process, that is, screening of a library of receptors with a chosen peptide substrate, has been less well developed, apart from the pioneering work of Still et al. They have screened libraries of receptors with a steroid core for selective peptide recognition,[5] and more recently have screened libraries of receptors for the kinetic resolution of cyclic amino acid derivatives.[6]

The basic design of tweezer receptors incorporates a "head group" or "hinge" bearing two side arms that incorporate appropriate functionality for binding with the backbone of suitable substrates (Figure 1). In contrast to many of the

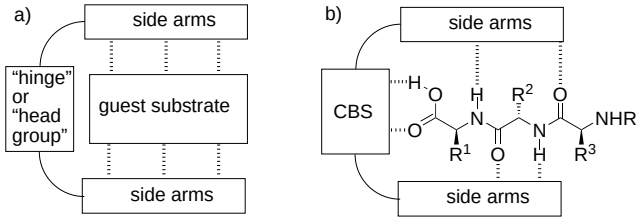
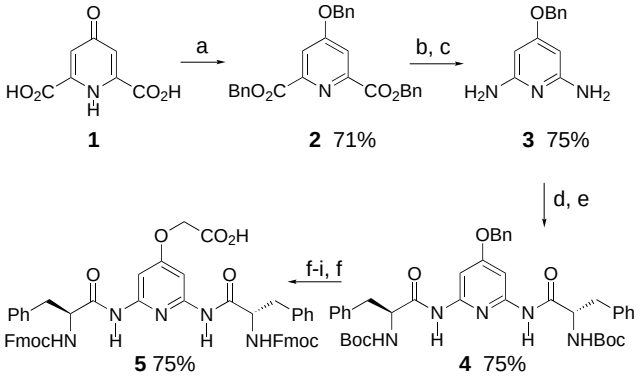


Figure 1. a) Schematic representation of a tweezer receptor with a hinge or head group and side arms that provide binding interactions with a suitable guest substrate. b) Schematic representation of a tweezer receptor with a carboxylic acid binding site (CBS) as the head group.

tweezer receptors reported to date, we are interested in incorporating a head group with a specific recognition site for the terminal functional group of the peptide guests.[3a] Incorporation of such a binding site or "anchor point", in addition to binding interactions from the tweezer side arms, should greatly increase the binding affinity of such tweezer receptors with suitable substrates, and should ultimately lead to receptors for the C-terminal sequence of larger peptides. Here we describe a novel tweezer structure that incorporates a diamidopyridine unit[7] as the head group to provide a binding site for the carboxylic acid terminus of peptide guests. The arms of the tweezer are themselves simple peptides which can potentially provide selective interactions with the backbone of the peptide guests. A small (2197 membered) resin-bound library of such tweezer receptors has been prepared using the "split and mix" strategy,[8] and has been used to demonstrate the potential of such libraries to identify selective receptors for selected tripeptide guests with a carboxylic acid terminus.

In order to prepare libraries of tweezer structures we synthesised a diamidopyridine derivative suitably functionalized with a carboxylic acid moiety to allow attachment to resin beads. Such a diamidopyridine derivative **5** was prepared by first converting chelidamic acid **1** into the tribenzyl derivative **2**, followed by aminolysis of the resulting benzyl esters and a Hofmann degradation to give diaminopyridine **3** in 53% overall yield (Scheme 1).[9] Treatment of **5** with *N*,*O*-bis(trimethylsilyl)acetamide,[10] followed by Boc-protected phenylalanine acid fluoride[11] gave **4** in 75% yield. Hydrogenolysis of the benzyl ether, alkylation with benzyl bromoacetate, exchange of the Boc protecting group for an Fmoc group, and finally further hydrogenolysis of the benzyl ester gave **5** in 75% overall yield from **4**.



Scheme 1. a) K_2CO_3, BnBr, acetone, reflux; b) NH_3, MeOH; c) KOH, Br_2, 90 °C; d) *N*,*O*-bis(trimethylsilyl)acetamide; e) BocPheF; f) 10% Pd/C, NH_4CO_2H, MeOH; g) $BrCH_2CO_2CH_2Ph$, K_2CO_3, DMF; h) 20% CF_3CO_2H, CH_2Cl_2; i) FmocONSu, Na_2CO_3, H_2O, dioxan. Boc = *tert*-butyloxycarbonyl; Fmoc = (9*H*-fluoren-9-ylmethoxy)carbonyl.

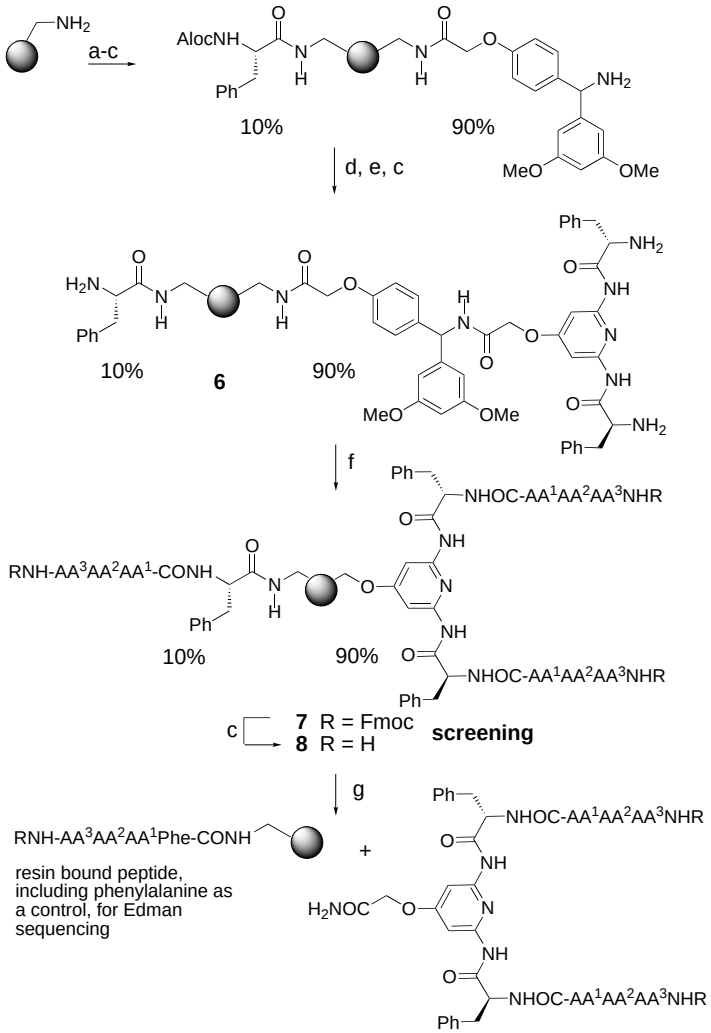
In initial studies the Boc-protected diamidopyridine derivative analogous to **5** was successfully coupled directly to TentaGel resin (without a linker). After removal of the Boc protecting groups a standard solid-phase split and mix synthesis with Fmoc-protected amino acids gave a library of tweezer receptors with tetrapetide side arms. Attempted sequencing of individual beads by Edman degradation was, however, only partially successful (typically only the first two amino acids could be unambiguously identified) and we concluded that under the conditions of the Edman degradation the amide bonds to the diamidopyridine were being cleaved, leading to the loss of the peptide side arm.[12] Thus we adopted a strategy using a peptide-coding strand and with the tweezer attached through the Rink-amide linker,[13] which

[*] Dr. J. D. Kilburn, T. Fessmann
Department of Chemistry
University of Southampton
Highfield, Southampton SO17 1BJ (UK)
Fax: (+44) 1703-593781
E-mail: jdk1@soton.ac.uk

[**] We would like to thank Prof. Mark Bradley for helpful advice concerning this work and Paul Skipp for carrying out the peptide sequencing.

  1433-7851/99/3813-1993 $ 17.50+.50/0

would be cleaved and washed away under the conditions for the Edman sequencing of the coding strand. The peptide coding strand was conveniently incorporated by first coupling Aloc-protected phenylalanine (Aloc = allyloxycarbonyl) onto 10% of the amine sites on TentaGel-NH_2 resin beads (130 μm beads, 0.29 mmol g^{-1}) followed by coupling the Fmoc-Rink amide linker[13] to the remaining 90% of amine sites (Scheme 2). Removal of the Fmoc group from the Rink linker was followed by the coupling of diamidopyridine **5** and further deprotection of both the Fmoc and Aloc protecting groups to yield resin **6** ready for library generation. A 2197 member library of Fmoc-protected tweezers **7** was prepared by a threefold coupling of thirteen Fmoc-protected amino acids (Gly, L-Ala, L-Val, L-Phe, L-Leu, L-Lys(Boc), L-Pro, L-Glu(O*t*Bu), L-Ser(O*t*Bu), L-Met, L-Trp, L-Asn, L-Gln) to the free amine groups by using the split and mix strategy. A second library of tweezers **8**, with free terminal amino groups, was prepared by treatment of a portion of library **7** with piperidine.



Scheme 2. a) AlocPhe, HOBT, DIC, DMAP, CH_2Cl_2 (10% of amine sites); b) Fmoc-Rink amide linker, HOBT, DIC, DMAP, CH_2Cl_2; c) 20% piperidine, DMF; d) **5**, HOBT, DIC, DMAP, CH_2Cl_2; e) $[Pd(PPh_3)_4]$, dimedone, CH_2Cl_2, THF; f) 3-fold split and mix Fmoc-peptide synthesis using Gly, L-Ala, L-Val, L-Phe, L-Leu, L-Lys(Boc), L-Pro, L-Glu(O*t*Bu), L-Ser(O*t*Bu), L-Met, L-Trp, L-Asn, L-Gln; g) Edman sequencing. HOBT = 1-hydroxy 1*H*-benzotriazole; DIC = 1,3-diisopropylcarbodiimide; DMAP = 4-dimethylaminopyridine; DMF = *N*,*N*-dimethylformamide.

The dansyl-labeled tripeptide with a free carboxylic acid terminus, DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH (DNS = dansyl = 5-(dimethylamino)-1-naphthalenesulfonyl), was used as the guest in screening experiments.[14] A sample of typically 4–5 mg (4000–5000 beads) of the library was equilibrated in a chosen solvent system (160 μL) for 24 h, followed by the addition of the dansylated tripeptide as a solution in the same solvent system (160 μL), and incubation for a further 24 h. The beads were analyzed in flat-bottomed glass pots under a Leica inverted DML microscope (magnification × 40) with a filter cube that contained a suppression filter at 425 nm and an excitation filter at 340–380 nm.

Control tests were carried out for each screening experiment. Thus the dansylated tripeptide guest was incubated with TentaGel resin alone, and samples of the tweezer libraries **7** and **8** were equilibrated in the solvent system in the absence of guest. In these experiments no selective fluorescence was observed, which confirms that all observed fluorescence in the actual screening experiments was derived from the interaction of the guest with selected library members. In addition, in order to show that the observed selectivity was not a consequence of an interaction of the tripeptide guest simply with the peptide side arm of the tweezer receptor, or with the coding strand on the library beads, a simple 2197 membered peptide library directly attached to TentaGel resin (analogous to the coding strand) was prepared. Incubation of this library with the dansylated peptide guests again showed no selective fluorescence.

The screening experiments are simple and allow the rapid evaluation of the binding potential and selectivity of the libraries with the guest in a range of solvent systems. In practice we observed no selective binding with any of the guest-library combinations in a range of buffered aqueous solvents. With chloroform as the solvent, however, high selectivity was observed for DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH with both the protected and deprotected libraries (approximately 1% of the beads were highly fluorescent in both cases against a background of low fluorescence beads).

Highly fluorescent beads were selected from each of the successful screening experiments and sequenced by Edman degradation.[15] (Edman sequencing was carried out in all cases for a fourth cycle and always showed the presence of phenylalanine for all beads, which was expected as it was introduced by the synthesis, and provided a useful check that the coding strand was operating correctly). The results of the sequencing experiments had the highest level of consensus for screening of DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH against the protected library **7** (Table 1) with the sequence Val-xxx-Trp found for six of the beads analyzed, and more specifically Val-Leu-Trp was found for three of the beads. A strong consensus for valine or leucine at the second and third positions was also found with the deprotected library **8** (Table 2), but the terminal amino acid residue was less well defined, with phenylalanine detected for four of the beads. Significantly tryptophan was not detected at the terminal position for any of the beads from the deprotected library **8**.

  1433-7851/99/3813-1994 $ 17.50+.50/0

Table 1. Sequencing data for ten fluorescent beads selected from the screening experiment of DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH with the protected library **7** in $CHCl_3$.

Bead	AA[1]	AA[2]	AA[3]
1	Val	Leu	Trp
2	Val	Leu	Trp
3	Val	Leu	Trp
4	Val	Ala	Trp
5	Val	Val	Trp
6	Val	Met	Trp
7	Val	Gly	Phe
8	Val	Ala	Val
9	Val	Gly	Val
10	Gln	Val	Gln

Table 2. Sequencing data for ten fluorescent beads selected from the screening experiment of DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH with the deprotected library **8** in $CHCl_3$.

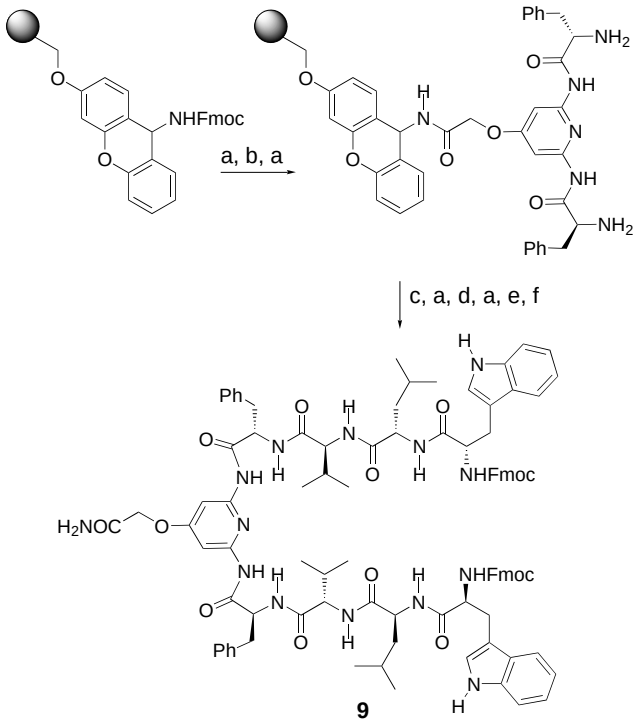
Bead	AA[1]	AA[2]	AA[3]
1	Val	Leu	Met
2	Val	Leu	Met
3	Val	Leu	Leu
4	Val	Leu	Leu
5	Val	Gly	Ala
6	Val	Met	Met
7	Leu	Val	Phe
8	Leu	Val	Phe
9	Phe	Val	Phe
10	Ala	Ala	Phe

The effect of solvent polarity on the binding selectivity of these libraries could be readily established by the simple addition of more-polar solvents to the screening experiments set up in $CHCl_3$ solution. Thus addition of DMSO led to a rapid loss of selective fluorescence, with both protected and deprotected libraries **7** and **8**, when the solvent composition reached 10:90 DMSO:$CHCl_3$. Likewise, addition of 10% MeOH effectively destroyed the selectivity. Addition of CH_3CN, however, did not have any noticeable effect on the selectivity observed for the screening of DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH with the Fmoc-protected library **7**, although addition of 50% CH_3CN did destroy the selectivity for the screening with the deprotected library **8**. Indeed, screening DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH against the Fmoc-protected library **7** in neat CH_3CN showed high selectivity (again, approximately 1% of beads were highly fluorescent). The sequencing of five highly fluorescent beads selected from this screening experiment by Edman degradation gave the identical consensus sequence Val-Leu-Trp for two of the five beads, as found for the same screen in $CHCl_3$, and Val-Met-Trp for a third bead (Table 3). The strong consensus for tryptophan at the terminal position when screening with the Fmoc protected library **7** in both $CHCl_3$ and CH_3CN, the absence of tryptophan at the terminal position when screening with the deprotected library **8** in $CHCl_3$, and absence of any selectivity for the latter screen in CH_3CN suggests that the combination of the Fmoc protecting group and the tryptophan residue play an important role in the binding of the peptide guest.

Table 3. Sequencing data for five fluorescent beads selected from the screening experiment of DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH with the protected library **7** in CH_3CN.

Bead	AA[1]	AA[2]	AA[3]
1	Val	Leu	Trp
2	Val	Leu	Trp
3	Val	Met	Trp
4	Phe	Val	Ala
5	Phe	Met/Val	Met/Trp

To establish that the observed binding with resin-bound tweezers was also operating in free solution, tweezer **9** was prepared by solid-phase synthesis as a single compound with tetrapeptide side arms Phe-Val-Leu-Trp, which corresponds to the most commonly found sequence in the screening experiments. The synthesis of **9** was most easily achieved with the Sieber amide resin,[16] which ultimately allowed cleavage of the tweezer from the resin with 1% CF_3CO_2H in CH_2Cl_2 (Scheme 3). Unfortunately tweezer **9** was essentially insoluble



Scheme 3. a) 20% piperidine, DMF; b) **5**, HOBT, DIC, DMAP, CH_2Cl_2; c) Fmoc-L-Val-OH, HOBT, DIC, DMAP, CH_2Cl_2; d) Fmoc-L-Leu-OH, HOBT, DIC, DMAP, CH_2Cl_2; e) Fmoc-L-Trp-OH, HOBT, DIC, DMAP, CH_2Cl_2; f) 1% CF_3CO_2H, CH_2Cl_2.

in neat $CDCl_3$ or neat CD_3CN and therefore did not allow NMR studies on the formation of a complex with DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH. A 500 μM solution of **9** in DMSO:$CHCl_3$ (2:98) could, however, be prepared and allowed us to study the complexation of **9** with DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH by fluorescence spectroscopy. The intensity of the fluorescence emission maximum (at 494 nm) for the dansyl group of the peptide guest decreased as aliquots of the tweezer receptor **9** were added, with the drop in intensity exhibiting typical saturation. The data from this experiment showed a good fit for the presumed 1:1 binding and allowed an estimation of the binding constant

  1433-7851/99/3813-1995 $ 17.50+.50/0

as $2.6 \times 10^5\,\text{M}^{-1}$ ($-\Delta G_a = 30.4\,\text{kJ mol}^{-1}$).[17] Thus, as anticipated the incorporation of a specific binding site for the carboxylic acid terminus of peptide guests into a tweezer structure provides binders with considerably higher affinity than tweezers with a nonspecific head group.

The selectivity of the library was further probed by carrying out identical screening experiments to those described above, but with Ac-L-Lys(DNS)-D-Ala-D-Ala-OH as the guest tripeptide.[18] No selectivity was observed for this guest when it was screened with the Fmoc-protected library **7**, but high selectivity was again observed when screened against the deprotected library **8** (approximately 1–2% of beads were highly fluorescent, although there were also a number of beads with intermediate levels of fluorescence, which suggested that the library was less selective for this peptide guest than for DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH). Indeed Edman sequencing of nine selected beads from this latter screen showed a lower overall consensus compared to screening with DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH, but there was considerable consensus for the second position, with seven of the nine sequences having Asn or Gln (Table 4). Significantly, the observed sequences were essentially completely different to those found for the screening with DNS-L-Glu(O*t*Bu)-L-Ser(O*t*Bu)-L-Val-OH, which confirms that the tweezer receptors are selective for different peptide substrates.

Table 4. Sequencing data for nine fluorescent beads selected from the screening experiment of Ac-L-Lys(DNS)-D-Ala-D-Ala-OH with the deprotected library **8** in $CHCl_3$.

Bead	AA[1]	AA[2]	AA[3]
1	Trp	Asn	Gly
2	Trp	Asn	Gly
3	Trp	Val	Phe
4	Leu	Val	Phe
5	Leu	Gln	Leu
6	Val	Gln	Leu
7	Phe	Gln	Val
8	Gln	Gln	Gln
9	Met	Gln	Lys

Received: December 30, 1998 [Z12854IE]
German version: *Angew. Chem.* **1999**, *111*, 2170–2174

Keywords: combinatorial chemistry · peptides · receptors · tweezers

[1] For original work on tweezer receptors, see S. C. Zimmerman, W. Wu, Z. Zeng, *J. Am. Chem. Soc.* **1991**, *113*, 196–201.

[2] For other recent examples of sequence-selective peptide receptors, see a) J. Dowden, P. D. Edwards, S. S. Flack, J. D. Kilburn, *Chem. Eur. J.* **1999**, *5*, 79–89; b) M. A. Hossain, H.-J. Schneider, *J. Am. Chem. Soc.* **1998**, *120*, 11208–11209; c) R. Breslow, Z. Yang, R. Ching, G. Trojandt, F. Odobel, *J. Am. Chem. Soc.* **1998**, *120*, 3536–3537; d) W. C. Still, *Acc. Chem. Res.* **1996**, *29*, 155–163; e) B. Hinzen, P. Seiler, F. Diederich, *Helv. Chim. Acta* **1996**, *79*, 942–960; f) G. Jung, H. Hofstetter, S. Feiertag, D. Stoll, O. Hofstetter, K.-H. Wiesmuller, V. Schurig, *Angew. Chem.* **1996**, *108*, 2261–2263; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2148–2150, and references therein.

[3] a) D. W. P. M. Löwik, M. D. Weingarten, M. Broekema, A. J. Brouwer, W. C. Still, R. M. J. Liskamp, *Angew. Chem.* **1998**, *110*, 1947–1950; *Angew. Chem. Int. Ed.* **1998**, *37*, 1846–1850; b) D. P. W. M. Löwik, S. J. E. Mulders, Y. Cheng, Y. Shao, R. M. J. Liskamp, *Tetrahedron Lett.* **1996**, *37*, 8253–8256; c) C. Gennari, H. P. Nestler, B. Salom, W. C. Still, *Angew. Chem.* **1995**, *107*, 1894–1896; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1765–1768; d) H. Wennemers, S. S. Yoon, W. C. Still, *J. Org. Chem.* **1995**, *60*, 1108–1109.

[4] a) M. Davies, M. Bonnat, F. Guillier, J. D. Kilburn, M. Bradley, *J. Org. Chem.* **1998**, *63*, 8696–8703; b) M. Torneiro, W. C. Still, *Tetrahedron* **1997**, *53*, 8739–8750; c) S. R. LaBrenz, J. W. Kelly, *J. Am. Chem. Soc.* **1995**, *117*, 1655–1656.

[5] a) Y. Cheng, T. Suenaga, W. C. Still, *J. Am. Chem. Soc.* **1996**, *118*, 1813–1814; b) R. Boyce, G. Li, H. P. Nestler, T. Suenaga, W. C. Still, *J. Am. Chem. Soc.* **1994**, *116*, 7955–7956; c) H. P. Nestler, *Mol. Diversity* **1996**, *2*, 35–40.

[6] M. D. Weingarten, K. Sekanina, W. C. Still, *J. Am. Chem. Soc.* **1998**, *120*, 9112–9113.

[7] a) F. Garcia-Tellado, S. Goswami, S.-K. Chang, S. J. Geib, A. D. Hamilton, *J. Am. Chem. Soc.* **1990**, *112*, 7393–7394; b) F. Garcia-Tellado, J. Albert, A. D. Hamilton, *J. Chem. Soc. Chem. Commun.* **1991**, 1761–1762.

[8] a) A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein Res.* **1991**, *36*, 487–493; b) K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82–84.

[9] For related approaches to 4-alkoxy-2,5-diaminopyridines, see a) D. G. Markees, V. C. Dewey, G. W. Kidder, *J. Med. Chem.* **1968**, *11*, 126; b) B. Feibush, A. Figueroa, R. Charles, K. D. Onan, P. Feibush, B. L. Karger, *J. Am. Chem. Soc.* **1986**, *108*, 3310–3318; c) M. Kotera, J.-M. Lehn, J.-P. Vigneron, *J. Chem. Soc. Chem. Commun.* **1994**, 197–199.

[10] S. Rajeswari, R. J. Jones, M. P. Cava, *Tetrahedron Lett.* **1987**, *28*, 5099–5102.

[11] L. A. Carpino, D. Sadat-Aalaee, H. G. Chao, R. H. DeSelms, *J. Am. Chem. Soc.* **1990**, *112*, 9651–9652.

[12] We have previously observed that the amide bonds of the diamidopyridine are susceptible to the weakly basic hydrolysis conditions that are normally suitable for methyl ester hydrolysis, see C. P. Waymark, J. D. Kilburn, I. Gillies, *Tetrahedron Lett.* **1995**, *36*, 3051–3054.

[13] a) M. S. Bernatowicz, S. B. Daniels, H. Köster, *Tetrahedron Lett.* **1989**, *30*, 4645–4648; b) H. Rink, *Tetrahedron Lett.* **1987**, *28*, 3787–3790.

[14] DNS-Glu(O*t*Bu)-Ser(O*t*Bu)-Val-OH was chosen for initial screening experiments as it was already available in the laboratory.

[15] P. Edman, G. Begg, *Eur. J. Biochem.* **1967**, *1*, 80–91. Sequencing was performed using an Applied Biosystems 477A pulsed liquid phase sequencer coupled on-line to an Applied Biosystems 120A Phenylthiohydantoin (PTH)-derivative analyzer.

[16] P. Sieber, *Tetrahedron Lett.* **1987**, *28*, 2107–2110.

[17] The binding constant was calculated by fitting the data to a 1:1 binding isotherm by using *NMRTit HG* software, kindly provided by Prof C. A. Hunter, University of Sheffield. See A. P. Bisson, C. A. Hunter, J. C. Morales, K. Young, *Chem. Eur. J.* **1998**, *4*, 845–851.

[18] Ac-L-Lys(DNS)-D-Ala-D-Ala-OH was chosen because of its biological relevance as a bacterial cell wall precursor. See H. C. Neu, *Science* **1992**, *257*, 1064–1073.

  1433-7851/99/3813-1996 $ 17.50+.50/0