DOI: 10.1002/ejoc.200500567

Linking the 3′ Ends of Oligonucleotide Duplexes with Cystine Disulfide Bridges

Vicente Marchán,[a] Daniel Pulido,[a] Enrique Pedroso,[a] and Anna Grandas*[a]

Keywords: Conjugates / Peptides / Oligonucleotides / 3′-linked DNA duplexes / Cysteine

Oligonucleotide duplexes with cystine disulfide 3′-3′ interstrand cross-links were synthesized. Stepwise solid-phase procedures, followed by basic deprotection under controlled conditions and air oxidation, afforded various disulfide-linked duplexes differing in the peptide and oligonucleotide composition and length. Some of these nucleopeptides form unimolecular duplexes in which peptide–minor-groove interactions are established, while others may give rise to extended nanostructures.

Introduction

Peptide chains have been covalently attached to oligonucleotides for a variety of purposes, most often to facilitate the penetration of the oligonucleotide through cell membranes.[1] Peptide-oligonucleotide conjugates have been prepared using different synthetic approaches that may give rise to different types of covalent linkage.[2,3] Over the past years, our group has developed methodology for the synthesis of a particular type of peptide-oligonucleotide conjugate, often referred to as the nucleopeptide, in which a phosphodiester group links one end of the oligonucleotide moiety and the side chain of a hydroxylated amino acid.[4] We obtained two cysteine-containing nucleopeptides and their corresponding disulfide dimers,[5] which formed 3′-3′ interstrand cross-linked duplexes that were noticeably more stable than the corresponding parent duplexes.[6] Duplex melting was shown to be a unimolecular process in which the two strands separated from each other, leaving the disulfide bridge intact.[7] NMR spectroscopic studies on the structure of [Ac-Cys-Gly-Ala-Hse(p$^{3'}$dGCATGC)-Ala-OH]$_2$-[S-S] also showed that entropic factors were responsible for the increase in thermal stability. Although the peptide chains were oriented towards the minor groove of the duplex, no interactions between the peptide chains and the double helix were observed.

In this manuscript, we report on the syntheses of a new series of cysteine-containing nucleopeptides (Figure 1) and their corresponding disulfide dimers for use in structural and biophysical studies. Some of the nucleopeptides were designed to favor the interaction between the peptide and the minor groove of the duplex. Specific interactions with the minor groove can be established by the formation of hydrogen bonds between polyamides commonly containing *N*-methylpyrrole rings and hydrogen-acceptor groups in the nucleobases.[8] Binding to the minor groove is also favored by van der Waals and hydrophobic interactions with the walls of the groove, and by electrostatic interactions of cationic groups with the negative electrostatic potential in the groove. As a result, an *N*-methylpyrrole residue was incorporated into nucleopeptide **1**, and a positively charged ornithine residue was introduced into nucleopeptide **2**. The other nucleopeptides incorporate longer peptide (**3**) or oligonucleotide moieties (**4**, **5**), and were obtained to evaluate how these changes in length may lead to the formation of either unimolecular duplexes or nanostructures in which oligonucleotide duplexes are connected to each other through their 3′ ends by cystine disulfide bridges (Figure 2).[9]

Ac-Cys-pyrr-Hse-Ala-OH **1**
|
p$^{3'}$dGCATGC

Ac-Cys-Gly-Orn-Hse-Ala-OH **2**
|
p$^{3'}$dGCATGC

Ac-Cys-Gly-Ala-Gly-Ala-Hse-Ala-OH **3**
|
p$^{3'}$dGCATGC

Ac-Cys-Gly-Orn-Hse-Ala-OH **4**
|
p$^{3'}$dGCCATGGC

Ac-Cys-Gly-Orn-Hse-Ala-OH **5**
|
p$^{3'}$dGCCATATGGC

pyrr = *N*-methylpyrrole residue
Hse = homoserine

Figure 1. Nucleopeptides synthesized in this study.

[a] Departament de Química Orgànica, Facultat de Química, Universitat de Barcelona,
Martí i Franquès 1–11, 08028 Barcelona, Spain
E-mail: anna.grandas@ub.edu

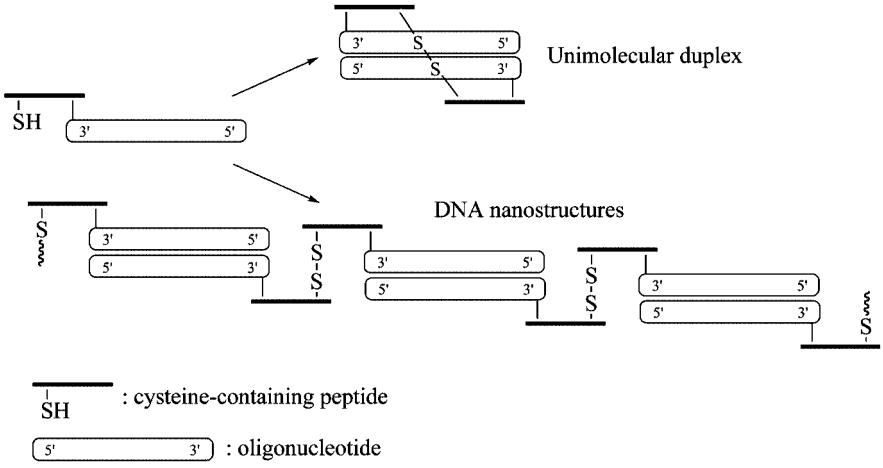


Figure 2. General structure for unimolecular and oligomeric duplexes with their 3′ ends linked by cystine disulfide bridges.

Results and Discussion

All nucleopeptide chains were assembled on a solid matrix with the *tert*-butylsulfenyl group for the permanent protection of the cysteine side chain. Previously developed methodology[4] was followed (Figure 3) with some modifications in the deprotection procedure, in particular in the case of nucleopeptide **1** (see below).

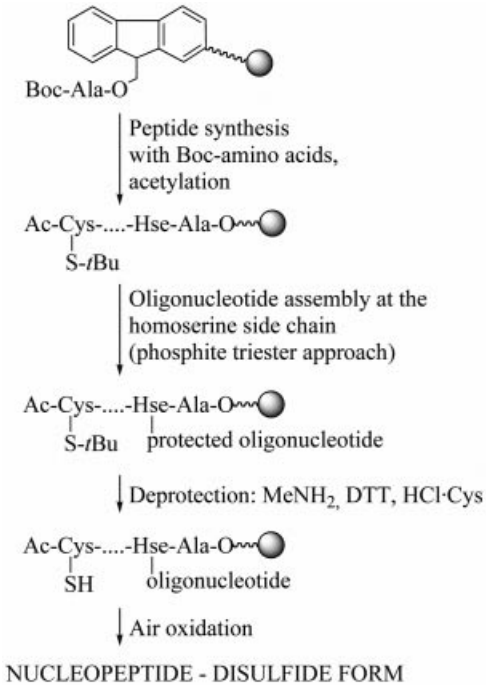


Figure 3. General synthesis scheme for the preparation of nucleopeptides **2**–**5**.

Earlier studies had shown that some side reactions can occur during the final deprotection treatment, namely alkylation of the thiol group of unprotected cysteine and cleavage of the peptide chain.[5] In spite of using the conditions that had previously yielded the best results, the crude products of nucleopeptides **2**–**5** showed complicated HPLC traces.

Alkylation of the cysteine side chain resulted from the reaction with the acrylonitrile generated upon removal of the cyanoethyl groups from the phosphates. The extent of this side reaction was reduced by carrying out the deprotection in a more dilute solution, although it depends on the nucleopeptide. The longer the oligonucleotide chain is, the higher the amount of acrylonitrile present in the reaction mixture and the higher the risk of alkylation. Since the unprotected cysteine residue reacts with acrylonitrile despite the large excess of ammonia and dithiothreitol (DTT) present in the reaction mixture, we decided to add cysteine hydrochloride to the deprotecting reagent to scavenge acrylonitrile, with optimal results.

The most difficult problem observed was the degradation of the peptide chain, which occurred in about a 50% extent regardless of the sequence of the nucleopeptide. The ammonia treatment did not cleave cysteine-containing peptide chains, and to date we have not observed this sort of proteolytic process with nucleopeptides containing any of the other trifunctional amino acids. The process thus seems to be associated with the presence, in the target molecule, of both an unprotected cysteine residue and a nucleotide or an oligonucleotide chain {peptide cleavage was also observed for a single nucleotide-containing nucleopeptide [Ac-Cys-Gly-Ala-Hse(p$^{3'}$dT)-Ala-OH], data not shown}. Since an increase in the distance between the cysteine residue and the oligonucleotide, as in nucleopeptide **3**, did not prevent degradation of the peptide chain, the only alternative to reduce the extent of this side reaction was to find deprotection conditions that allow a shorter exposure of the target product to bases.

Among the different conditions assayed, the best results were achieved with a final deprotection treatment with a mixture of methylamine (40%, aq.) and DTT (0.1 M in dioxane) [1:1 v/v] at room temperature for 1 h in the presence of cysteine hydrochloride (0.025 M) [20 mL/140 mg of nucleopeptide-resin]. Under these conditions ([DTT+Cys]/nucleopeptide molar ratio≈60), the level of nucleopeptide alkylation was found to be virtually nil. Although, unfortunately, some peptide degradation occurred, this treatment

allowed the isolation of nucleopeptides **4** and **5** in reasonable yields, whereas deprotection using the previously reported conditions had not.

It is important to point out that for the deprotection with methylamine, the exocyclic amine of cytosine must be protected with a labile group, such as an acetyl group, to prevent the transamination reaction that may yield 4-*N*-methylcytosine.[10] As in the case of ammonia treatments, cleavage of the nucleopeptide-resin bond occurs mostly by β-elimination[5] to yield the nucleopeptide acid (COOH). However, some aminolysis of the ester bond, which renders the C terminal *N*-methyl amide (CONHMe) of the nucleopeptide, can also occur. In this respect, the advantage of methylamine compared with ammonia is that the nucleopeptide acid and its *N*-methyl amide can be more easily separated by HPLC than the acid and the primary carboxamide ($CONH_2$).

After the deprotection treatment, the excess DTT was eliminated by gel filtration through Sephadex G-10 or, in our hands, slightly less effectively, by ethyl acetate extraction.[11] All nucleopeptides were subsequently purified and oxidized in air to the corresponding disulfides at pH 8. The extent of peptide degradation of the nucleopeptides under these mild conditions is very low, and the disulfide derivatives are totally stable.

As stated above, nucleopeptide **1** was designed to contain an *N*-methylpyrrole residue in its peptide sequence. Both peptide-[minor groove binder][12] and oligonucleotide-[minor groove binder][13] conjugates have been obtained, but, to the best of our knowledge, this is the first described conjugate in which the three types of molecules are combined. The synthesis of nucleopeptide **1** was particularly tough and it required specific adjustments (Figure 4). To allow for oligonucleotide elongation at the free hydroxy group of homoserine, peptide assembly had to be carried out by leaving the homoserine side chain untouched, thus preventing the use of strong carboxyl activators for peptide assembly. Therefore, it was not evident that this requirement could be made compatible with the low reactivity of *N*-methylpyrrole carboxylic acids.[14,15] In addition, we an-

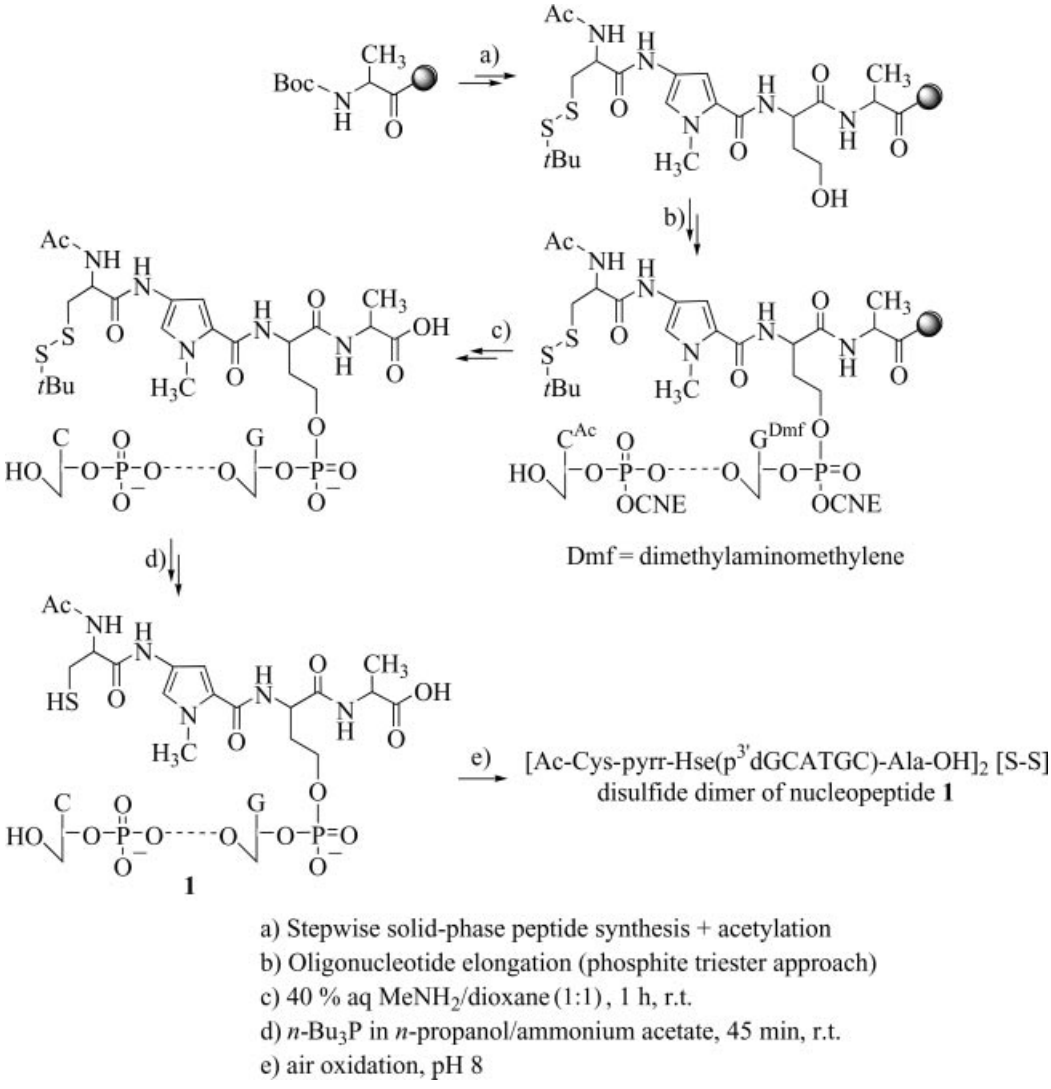
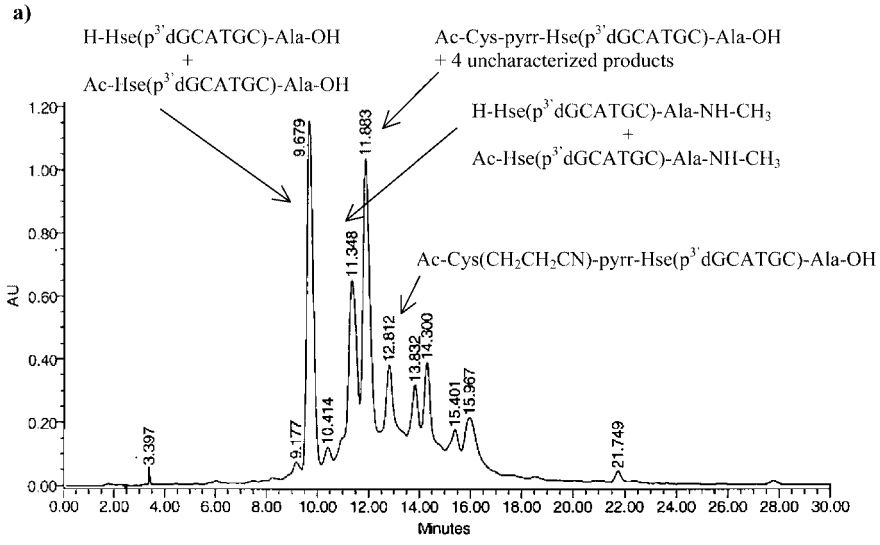


Figure 4. Main steps of the synthesis of nucleopeptide **1**.

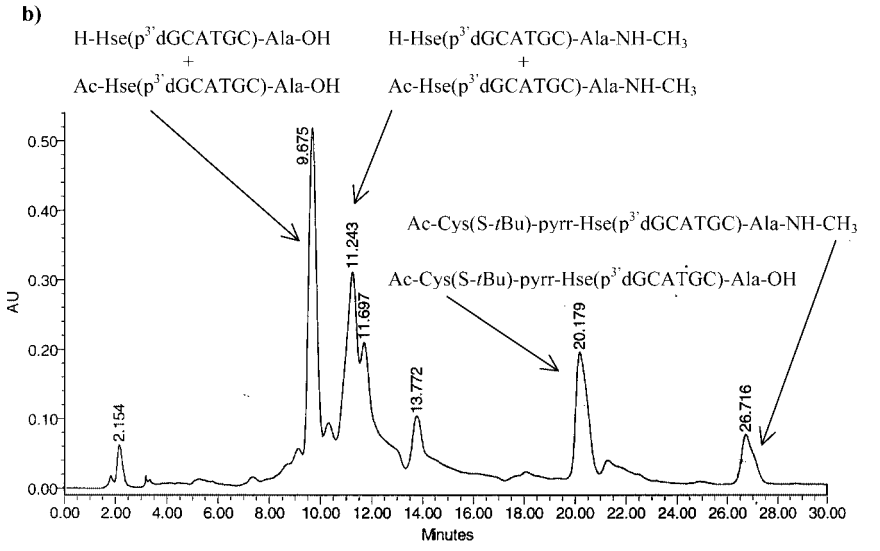
ticipated difficulties in the coupling of cysteine to the H-pyrr-Hse-Ala-resin, since the amine groups of aminopyrroles are less nucleophilic than those of the natural amino acids.

Carboxyl activation with DCC and HOBt gave the best results to couple the Boc-derivatives of both *N*-methylpyrrole and cysteine. Cysteine incorporation was quantitative, but the pyrrole residue could be coupled only in 80% yield. Unreacted homoserine amine groups at the dipeptide level and the N terminal of the peptide were acetylated by treating with acetic acid and DCC. No acylation of the free homoserine hydroxy group was detected, as inferred from quantification of the amount of dimethoxytrityl groups liberated after 5′-deprotection of the first nucleotide incorporated at the homoserine side chain.

In this case, the cleavage of the peptide chain under the basic final deprotection conditions was the most important problem. The main product in the nucleopeptide crude was H-Hse(p$^{3'}$dGCATGC)-Ala-OH, both when the deprotection was carried out with NH_3/DTT[5] and with methylamine/DTT, as described above, and no fraction containing only the target nucleopeptide could be isolated (Figure 5a). Although the presence of DTT in the deprotection mixture has the beneficial effect of reducing the amount of impurities derived from the oxidation of Cys(S-*t*Bu) during the elongation of the oligonucleotide chain,[5] it was clear that



Ac-Cys-pyrr-Hse(p3'dGCATGC)-Ala-OH crude (cleavage with $MeNH_2$ + DTT)



Ac-Cys-pyrr-Hse(p3'dGCATGC)-Ala-OH crude (cleavage with $MeNH_2$)

Figure 5. HPLC traces of the crudes obtained after treatment of [protected nucleopeptide **1**]-resin with either $MeNH_2$/DTT (a) or $MeNH_2$ (b).

the only way of obtaining the target product was by deprotecting all the functional groups, except the cysteine side chain. Although not the main product in the crude, the cysteine-protected nucleopeptide is less polar than the free nucleopeptide **1** and the accompanying side products, and it can be easily separated by reverse-phase liquid chromatography (Figure 5b).

Treatment of the nucleopeptide-resin with a mixture of methylamine (40%, aq.) and dioxane [1:1 v/v] at room temperature for 1 h without reducing agent, followed by medium-pressure reverse-phase liquid chromatography, allowed Ac-Cys(S-*t*Bu)-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH to be isolated in moderate yields. The cysteine side chain was deprotected by reacting with tris(*n*-butyl)phosphane (*n*-Bu_3P) in *n*-propanol and, after removal of the excess phosphane and phosphane oxide by gel filtration through Sephadex G-10, the disulfide dimer of nucleopeptide **1** was formed by oxidation in air at pH 8 and isolated by HPLC. No degradation of the nucleopeptide was observed after any of these treatments.

In summary, for the preparation of cysteine-containing nucleopeptides, as a general rule we suggest that stepwise solid-phase assembly with the most base-labile protecting groups be followed by deprotection with methylamine in the presence of cysteine and DTT. If the stability of the target nucleopeptide under these conditions is exceptionally low, as in the case of nucleopeptide **1**, deprotection with methylamine followed by the removal of the *tert*-butylsulfenyl group with *n*-Bu_3P allows the isolation of the target nucleopeptide.

Structural and biophysical studies of the disulfide derivatives of nucleopeptides **1**–**5** are in progress, and the results will be reported in due time. The information presently available indicates that nucleopeptides **1** and **2** form unimolecular duplexes, whereas longer oligonucleotides (nucleopeptides **4** and **5**) give rise to more complex duplexes with higher molecularity.

Experimental Section

General Procedures: Detailed experimental procedures for solid-phase nucleopeptide synthesis are described in ref.[4,5].

Amino acid derivatives used for peptide syntheses: N^{α}-Boc-L-Orn(Fmoc)-OH, N^{α}-Boc-Gly-OH, and N^{α}-Boc-L-Ala-OH were purchased from Bachem. N^{α}-Boc-L-Hse(DMT)-O^- $HTEA^+$ and N^{α}-Boc-L-Cys(*S*-*t*Bu)-OH were prepared as described in ref.[4] 2-(4-Boc-amino-*N*-methylpyrrole)carboxylic acid was obtained essentially as described in ref.[16] Oligonucleotide elongation at the side-chain hydroxy group of homoserine was carried out by using the 3′-*O*-(2-cyanoethyl)diisopropylphosphoramidites of 5′-*O*-DMT-dA^{Bz}, 5′-*O*-DMT-dC^{Ac}, 5′-*O*-DMT-dG^{Dmf}, and 5′-*O*-DMT-dT. HPLC was performed on Kromasyl C-18 columns by using ammonium acetate (0.05 M) and an acetonitrile/water (1:1 v/v) mixture as eluents A and B, respectively (flow = 1 mL/min for analytical conditions and 3 mL/min for purifications on the semi-preparative scale). Vydac C-18 was used for the medium-pressure liquid chromatography purifications. MALDI-TOF mass spectrometric analyses: trihydroxyacetophenone/ammonium citrate, negative mode, linear.

Nucleopeptides 2–5: Nucleopeptides **2** and **3** were deprotected by treatment with a mixture of ammonia (conc. aq.)/DTT (1 M in dioxane) [1:1 v/v, 7 h, room temperature]. Deprotection of nucleopeptides **4** and **5** was effected by treatment with cysteine hydrochloride (0.025 M) in a mixture of $MeNH_2$ (40% aq.)/DTT (0.1 M in dioxane) [1:1 v/v, 1 h, room temperature]. The nucleopeptide crudes were evaporated to dryness and redissolved in ammonium acetate (0.05 M), and DTT was separated by filtration through Sephadex G-10 by eluting with the same buffer. Alternatively, after evaporation to dryness, the nucleopeptide crude was dissolved in water, and DTT was eliminated by carrying out extractions with ethyl acetate. However, this procedure did not quantitatively remove DTT. Nucleopeptides were purified by reverse-phase medium- or high-pressure liquid chromatography,[4] and characterized by MALDI-TOF MS.

Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCATGC)-Ala-OH (2): Purification: MPLC, solvent A: ammonium acetate (0.05 M), solvent B: [ammonium acetate (0.05 M, 70%) + acetonitrile/H_2O (1:1 v/v, 30%)], gradient with 600 mL of each solvent; anal. HPLC: linear gradient from 5 to 35% of solvent B in 30 min; mass spectrometric characterization: m/z = 2361.2 $[M - H]^-$ (average calcd. mass: 2360.8).

Ac-Cys-Gly-Ala-Gly-Ala-Hse(p$^{3'}$dGCATGC)-Ala-OH (3): Purification and analysis: anal. HPLC: linear gradient from 5 to 35% of solvent B in 30 min; mass spectrometric characterization: m/z = 2444.7 $[M - H]^-$ (average calcd. mass: 2445.8).

Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCCATGGC)-Ala-OH (4): Purification: MPLC, solvent A: ammonium acetate (0.05 M), solvent B: [ammonium acetate (0.05 M, 75%) + acetonitrile/H_2O (1:1 v/v, 25%)], gradient with 600 mL of each solvent; anal. HPLC: linear gradient from 5 to 30% of solvent B in 30 min; mass spectrometric characterization: m/z = 2976.5 $[M - H]^-$ (average calcd. mass: 2979.2).

Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCCATATGGC)-Ala-OH (5): Purification and anal. HPLC analysis: same as for nucleopeptide **4**; mass spectrometric characterization: m/z = 3590.5 $[M - H]^-$ (average calcd. mass: 3596.6).

Disulfide Dimers of Nucleopeptides 2–5: The purified nucleopeptides were oxidized in air to the disulfides by stirring in triethylammonium hydrogencarbonate (100 mM, pH 8, 24 h, room temperature). The corresponding disulfides were purified by reverse-phase HPLC (more than one run was sometimes required) and characterized by MALDI-TOF MS.

[Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCATGC)-Ala-OH]$_2$[S-S]: Purification and analysis: anal. HPLC: linear gradient from 5 to 25% of solvent B in 40 min; mass spectrometric characterization: m/z = 4713.4 $[M - H]^-$, 2361.3 $[M - 2H]^{2-}$ (average calcd. mass: 4719.6).

[Ac-Cys-Gly-Ala-Gly-Ala-Hse(p$^{3'}$dGCATGC)-Ala-OH]$_2$[S-S]: Purification and anal. HPLC analysis: same as for disulfide dimer of nucleopeptide **2**; mass spectrometric characterization: m/z = 4876.7 $[M - H]^-$, 2442.3 $[M - 2H]^{2-}$ (average calcd. mass: 4889.7).

[Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCCATGGC)-Ala-OH]$_2$[S-S]: Purification and analysis: anal. HPLC: linear gradient from 5 to 20% of solvent B in 30 min; mass spectrometric characterization: m/z = 5940.7 $[M - H]^-$, 2975.3 $[M - 2H]^{2-}$ (average calcd. mass: 5956.3).

[Ac-Cys-Gly-Orn-Hse(p$^{3'}$dGCCATATGGC)-Ala-OH]$_2$[S-S]: Purification and anal. HPLC analysis: same as for disulfide dimer of nucleopeptide **4**; mass spectrometric characterization: m/z = 7167.8 $[M - H]^-$, 3587.5 $[M - 2H]^{2-}$ (average calcd. mass: 7191.2).

[Ac-Cys-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH]$_2$[S-S]: Nucleopeptide **1** was assembled following the standard procedures with some minor modifications. After incorporation of the first two amino acids,

peptide elongation on H-Hse-Ala-resin was accomplished with 5 equiv. amino acid, DCC, and HOBt (3 h, room temperature). A positive ninhydrin test was obtained after two couplings of the pyrrole derivative. The unreacted homoserine amine groups, as well as the N terminal of the peptide at the end of the synthesis, were acetylated by reaction with 3 equiv. acetic acid and DCC, (10 min, room temperature) [the side-chain hydroxy group of homoserine was not capped under these conditions]. Oligonucleotide elongation at the side chain of homoserine was carried out as usual (99% yield for incorporation of the first nucleotide). Ac-Cys(*S*-*t*Bu)-pyrr-Hse[prot.(p$^{3'}$dGCATGC)]-Ala-resin was treated with a mixture of $MeNH_2$ (40%, aq.)/dioxane [1:1 v/v, 1 h, room temperature], and, after filtration and washing, the crude was evaporated to dryness. Ac-Cys(*S*-*t*Bu)-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH was purified by HPLC (linear gradient from 5 to 25% of solvent B in 40 min). MALDI-TOF MS: *m*/*z* = 2397.7 [M – H]$^-$ (average calcd. mass: 2399.9).

Ac-Cys-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH (1): Compound **1** was obtained by addition of *n*-Bu_3P (0.6 M solution in *n*-propanol) to a solution of Ac-Cys(*S*-*t*Bu)-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH in ammonium acetate (0.05 M)/*n*-propanol [1:1 v/v], (relative molar ratio 100:1). After stirring under argon for 45 min at room temperature, the mixture was concentrated by evaporation under reduced pressure (the alcohol and most of the water were eliminated), and the phosphane and phosphane oxide were separated by gel filtration (Sephadex G-10, elution with 0.05 M ammonium acetate). MALDI-TOF MS: *m*/*z* = 2308.6 [M – H]$^-$ (average calcd. mass: 2311.7).

[Ac-Cys-pyrr-Hse(p$^{3'}$dGCATGC)-Ala-OH]$_2$[S-S]: Purified nucleopeptide **1** was oxidized as described above to the corresponding disulfide dimer, which was subsequently purified by HPLC. MALDI-TOF MS: *m*/*z* = 4611.1 [M – H]$^-$ (average calcd. mass: 4621.3).

Acknowledgments

This work was supported by funds from the Spanish Ministerio de Educación y Ciencia (grants BQU-2001-3693-C02-01 and CTQ2004-8275-C02-01) and the Generalitat de Catalunya (2001SGR-49 and Centre de Referència de Biotecnologia).

[1] See, for instance: M. J. Gait, *Cell. Mol. Life Sci.* **2003**, *60*, 1–10; A. Astriab-Fisher, D. Sergueev, M. Fisher, B. Ramsay-Shaw, R. L. Juliano, *Pharm. Res.* **2002**, *19*, 744–754.
[2] T. S. Zatsepin, D. A. Stetsenko, M. J. Gait, T. S. Oretskaya, *Bioconjugate Chem.* **2005**, *16*, 471–489.
[3] E. M. Zubin, E. A. Romanova, T. S. Oretskaya, *Russ. Chem. Rev.* **2002**, *71*, 239–264.
[4] A. Grandas, V. Marchán, L. Debéthune, E. Pedroso, *Current Protocols in Nucleic Acids Chemistry – Synthesis of Modified Oligonucleotides and Conjugates* (Eds.: S. L. Beaucage, D. E. Bergstrom, P. Herdewijn, A. Matsuda), series Ed.: E. W. Harkins, John Wiley & Sons, USA, **2004**, unit 4.22.
[5] L. Debéthune, V. Marchán, G. Fábregas, E. Pedroso, A. Grandas, *Tetrahedron* **2002**, *58*, 6965–6978.
[6] V. Marchán, L. Debéthune, E. Pedroso, A. Grandas, *Tetrahedron* **2004**, *60*, 5461–5469.
[7] I. Gómez-Pinto, V. Marchán, F. Gago, A. Grandas, C. González, *Chem. Commun.* **2003**, 2558–2559.
[8] a) P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–289; b) S. Neidle, *Nat. Prod. Rep.* **2001**, *18*, 291–309.
[9] As an example of a recent review on DNA as nanomaterial, see: Y. Ito, E. Fukusaki, *J. Mol. Catal. B: Enzym.* **2004**, *28*, 155–166.
[10] M. P. Reddy, N. B. Hanna, F. Farooqui, *Tetrahedron Lett.* **1994**, *35*, 4311–4314.
[11] V. G. Metelev, O. A. Borisova, E. M. Volkov, T. S. Oretskaya, N. G. Dolinnaya, *Nucleic Acids Res.* **2001**, *29*, 4062–4069.
[12] J. B. Blanco, O. Vázquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Chem. Eur. J.* **2005**, *11*, 4171–4178, and references therein.
[13] S. Ghosh, E. Defrancq, J. H. Lhomme, P. Dumy, S. Bhattacharya, *Bioconjugate Chem.* **2004**, *15*, 520–529, and references therein.
[14] T. Ernst, C. Richert, *Synlett* **2005**, 411–416.
[15] D. Pulido, unpublished results.
[16] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146.

Received: July 27, 2005
Published Online: December 5, 2005

