



New dsDNA Binding Unnatural Oligopeptides with Pyrimidine Selectivity

Zhenyu Zhang, Patrick Chaltin, Arthur Van Aerschot, Jeff Lacey, Jef Rozenski, Roger Busson and Piet Herdewijn*

Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

Received 23 January 2002; accepted 3 July 2002

Abstract—Solid phase peptide library screening followed by extension of a lead recognition element for binding to a dsDNA sequence (NF binding site of IL6) using solution phase screening, delivered a new DNA binding peptide, Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂. In the present research, the contribution of the different amino acid side chains to the binding strength of the peptide to dsDNA was investigated using an ethidium bromide displacement test. Based on these results, the lead structure was optimized by deconvolution. Eight new unnatural amino acids were evaluated at two positions of the heptapeptide replacing the Ual-Sar fragment. The strongest dsDNA binding was observed using {[[(3-chlorophenyl)methyl]amino}acetic acid (Cbg) and β -cyclohexyl-L-alanine (Cha) respectively, at those two positions. A 10-fold increase in affinity compared to the Ual-Sar sequence was obtained. Further enhancement of dsDNA binding was obtained with hybrid molecules linking the newly developed peptide fragment to an acridine derivative with a flexible spacer. This resulted in ligands with affinities in the μ M range for the dsDNA target (K_d of 2.1×10^{-6} M). DNase I footprinting with the newly developed oligopeptide motifs showed the presence of a pronounced pyrimidine specificity, while conjugation to an intercalator seems to redirect the interaction to mixed sequences. This way, new unnatural oligopeptide motifs and hybrid molecules have been developed endowed with different sequence selectivities. The results demonstrate that the unnatural peptide library approach combined with subsequent modification of selected amino acid positions, is very suited for the discovery of novel sequence-specific dsDNA binding ligands.

© 2002 Elsevier Science Ltd. All rights reserved.

Introduction

Nucleic acids are important targets in the development of new compounds, able to control gene expression. The armamentarium of nucleic acids recognizing ligands is increasing exponentially.^{1–4} It is well known that most DNA-targeted drugs are small molecules of limited specificity and high toxicity.^{5–8} The specificity can be increased by using compounds with multiple interaction sites with the DNA target, and thus compounds of medium size. As DNA-protein interaction is the central event in the regulation of gene expression, it may be used as a model to design synthetic molecules that recognize DNA by groove binding.⁹ The significant progress in the discovery of new DNA binding agents within the past decade, particularly in the field of lexitropsins, combilexins and PNAs, has provided us with a

lot of supplementary information about DNA-ligand recognition motifs.^{1,4,9} The involvement of the amide backbone in these three strategies has supported our approach to synthesize peptide libraries based on modified amino acids with unique side chains or constrained backbones. This peptide-based approach offers, likewise, the advantage that combinatorial chemistry can be used to easily generate libraries, representing diverse recognition patterns.

In previous experiments, a first lead sequence (Arg-Ual-Sar) was generated using a solid phase screening method.¹⁰ This sequence was extended using solution phase screening, resulting in several heptapeptide lead compounds such as Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂, showing an affinity of 4×10^{-4} M for a dsDNA 14-mer target.¹¹ The structures of Arg, Ual, Sar, Chi and Tal are displayed in Figure 1. Investigation of the interaction capacities with dsDNA of the different amino acids in the peptide lead Ac-Arg-Ual-Sar-Chi-Chi-Chi-Arg-CONH₂, suggested that the uracil (Ual)

*Corresponding author. Tel.: +32-16-337387; fax: +32-16-337340; e-mail: piet.herdewijn@rega.kuleuven.ac.be

and sarcosine (Sar) side chains contribute less to the overall affinity of the peptide than quinazoline-dione (Chi). Therefore, starting from the lead peptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂, new libraries were prepared with the aim to select stronger binding residues for the two positions that initially were fixed during the solution phase selection round (Ual-Sar), thereby preserving the positively charged arginine (Arg). Eight amino acid building blocks (Fig. 1) were applied to optimize the two random positions X and Y in Ac-Arg-Y-X-Chi-Chi-Tal-Arg-CONH₂. Fmoc-strategy-based solid phase peptide synthesis was used to assemble the library. The dsDNA target used for screening in solution is the labeled 14-mer double stranded DNA, representing the binding site of the nuclear factor of Interleukine 6, as used in previous screening assays.^{10–12} Gel retardation experiments were performed to evaluate the binding strength of the ligands and to determine the dissociation constants of the selected peptides. Footprinting was used to evaluate the sequence specificity of the dsDNA interaction.

Results

Contribution of amino acid side chains to dsDNA binding

To determine the contribution to dsDNA binding of the individual amino acids of the oligopeptide lead structures selected in previous work, new oligopeptide analogues were prepared and investigated. The middle five amino acids of Ac-Arg-Ual-Sar-Chi-Chi-Chi-Arg-NH₂ were systematically replaced by glycine. This way the side chains of the different amino acids were removed, while the peptide backbone was retained. By measuring the affinity of the different analogues, the contribution of the amino acid side chains could be evaluated (although we are aware of the fact that the side chain

may also influence the binding indirectly by its importance for the conformation of the peptide). This could yield important information for further modifications of the oligopeptides to enhance the affinity or selectivity.

The glycine substituted peptides were synthesized on solid support as described previously.¹¹ Coupling of Fmoc-glycine was performed with an activation mixture containing 4 equiv of HOAt, DIC, DIEA and the amino acid in regard to the free amino groups. Activation mixtures for *N*-Fmoc-sarcosine, *N*α-Fmoc-β-(uracil-1-yl)-α-D-alanine, *N*α-Fmoc-*N*^G-2,2,5,7,8-pentamethylchroman-6-sulfonyl-L-arginine and *N*α-Fmoc-β-(1*H*-quinazolin-2,4-dion-3-yl)-α-D-alanine were as used before.¹¹ After cleavage from the solid support and reversed phase HPLC purification, the compounds were identified by mass spectrometry. All measured masses corresponded to the calculated ones. The binding strength of the glycine substituted analogues and the reference oligopeptide Ac-Arg-Ual-Sar-Chi-Chi-Chi-Arg-NH₂ was compared in an ethidium bromide displacement assay¹³ using the original 14-mer dsDNA target. The obtained percentages are relative to controls without DNA and without ligand addition. Results at different concentrations are averaged from two to three measurements and are shown in Table 1.

The ethidium bromide displacement experiment shows that replacement of each of the three quinazoline-dione amino acids with glycine leads to an equally reduced affinity, corresponding to a ±40% decrease of EtBr-displacement compared to the reference oligopeptide. Analogues with glycine substituted for sarcosine and the uracil amino acid retain, however, a similar affinity for the dsDNA target as the reference oligopeptide.

These results suggest that the uracil and sarcosine side chains contribute less to the overall affinity of the peptide Ac-Arg-Sar-Chi-Chi-Chi-Arg-NH₂, while the quinazoline-dione side chains actively participate in the stabilization of the DNA-peptide complex. Furthermore, this highlights again the importance of the selected quinazoline-dione amino acids (Chi) in the oligopeptide motif for interaction with dsDNA, as reported previously.¹¹

Synthesis of the amino acid building blocks

Based on the results obtained with the glycine substituted oligopeptides, the Ual-Sar positions of the peptide

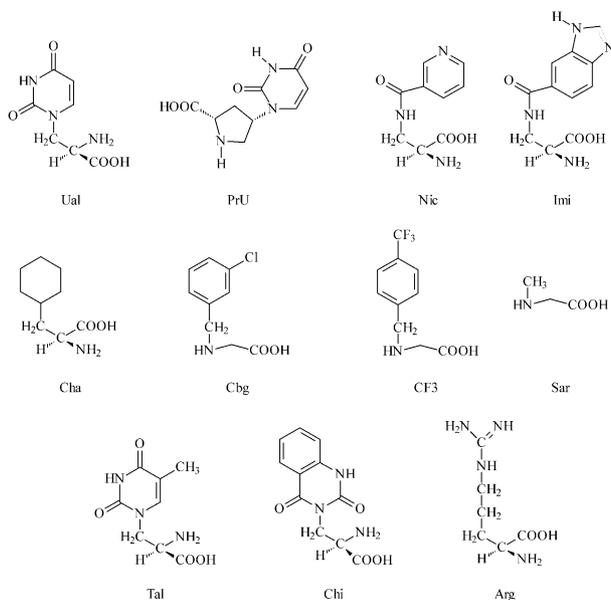


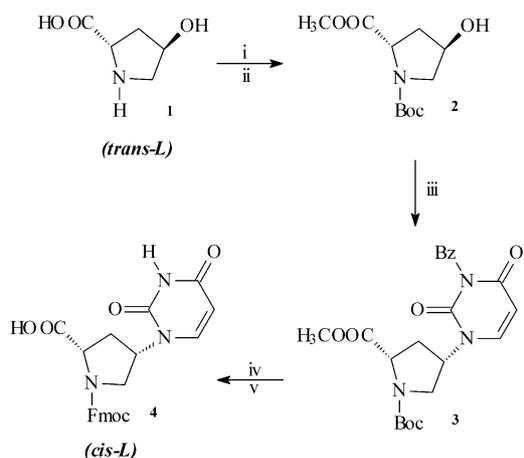
Figure 1. Structures of unnatural synthetic amino acids (Ual, PrU, Nic, Imi, Cha, Cbg, CF3 and Sar) used in the present library approach to select for position X and Y in Ac-Arg-Y-X-Chi-Chi-Tal-Arg-CONH₂. The structures of Chi, Tal and Arg are also shown.

Table 1. Fluorescence readings of glycine substituted oligopeptides and the reference peptide Ac-Arg-Ual-Sar-Chi-Chi-Chi-Arg-NH₂ at two different concentrations. The results are reported as % fluorescence relative to control wells

Oligopeptide sequences	250 μM	750 μM
Ac-Arg-Ual-Sar-Chi-Chi-Chi-Arg-CONH ₂	42%	27%
Ac-Arg-Ual-Sar-Chi-Chi-Gly-Arg-CONH ₂	87%	67%
Ac-Arg-Ual-Sar-Chi-Gly-Chi-Arg-CONH ₂	79%	66%
Ac-Arg-Ual-Sar-Gly-Chi-Chi-Arg-CONH ₂	77%	62%
Ac-Arg-Ual-Gly-Chi-Chi-Chi-Arg-CONH ₂	46%	30%
Ac-Arg-Gly-Sar-Chi-Chi-Chi-Arg-CONH ₂	45%	20%

Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂ were modified with several new amino acid building blocks to increase the contribution of these positions to the stability of the oligopeptide-DNA complexes. The peptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂ was chosen to be modified because of its mixed sequence and ease of synthesis. Looking at the recognition mechanisms between DNA and proteins, unnatural amino acids (Fig. 1) were selected for these positions in a way that they represent both functional and structural diversity. In previous libraries, the selection of new unnatural amino acids was predominantly based on the presence of hydrogen bonding functionalities and intercalating possibilities. In this research, three new building blocks (Cha, Cbg and CF3) were selected in order to incorporate the possibility for hydrophobic interactions, since these contacts are known to strongly stabilize DNA-ligand complexes.¹⁴ Furthermore, Nic and Imi were chosen as new amino acid residues to further enlarge the hydrogen bonding capacities of the peptides. Moreover, more rigidity was introduced by using the building block PrU to reduce the entropy loss during binding (Fig. 1).

In this way, using proline and alanine as scaffold, four new amino acids were introduced (PrU, Nic, Imi and Cha) and also two new peptoid monomers (CF3 and Cbg) were synthesized and used in the library. The Ual and Sar amino acids were already present in the previous library and they were added to serve as reference amino acids. Their synthesis was performed as described previously.¹¹ Following protection with a *N*³-benzoyl group, uracil was reacted with *N*-Boc-protected *trans*-L-hydroxyproline methyl ester by Mitsunobu reaction, similar to the method described by Lowe for the synthesis of a thymine derivative (Scheme 1).¹⁵ The obtained *cis*-L-proline derivative was fully deprotected using NaOH followed by trifluoroacetic acid. The secondary amino group of the pyrrolidine ring was protected with a Fmoc group under Schotten-Baumann conditions (Scheme 1).



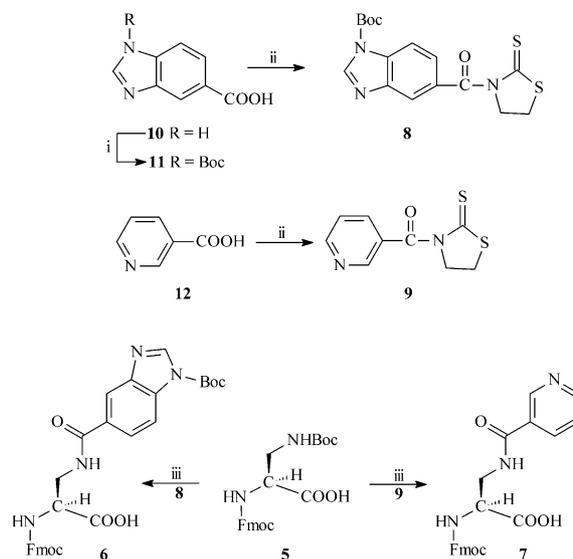
Scheme 1. Synthesis of *N*^α-Fmoc-*cis*-4-uracil-L-proline. (i) 1.2 equiv (Boc)₂O, 10% Na₂CO₃, 1,4-dioxane; 84%, (ii) CsCO₃, 1.2 equiv CH₃I in DMF; 85%, (iii) 1.5 equiv 3-benzoyluracil, PPh₃, DEAD in THF; 82%, (iv) (a) 0.1N NaOH, room temperature, (b) TFA/CH₂Cl₂ (1/1, v/v); 90%, (v) 1.2 equiv FmocCl, 10% Na₂CO₃, 1,4-dioxane; 80%.

Starting from *N*^α-(9-fluorenylmethoxycarbonyl)-*N*^β-*tert*-butyloxycarbonyl-L-α-2,3-diaminopropionic acid **5**,¹⁶ compound **6** and **7** were synthesized (Scheme 2). The Boc-group of **5** was removed with acid and heterocycles were introduced via amide bond formation using the activated amides **8** and **9** as reagent. The carboxylic function of **11** and **12** was activated with 1,3-thiazolidine-2-thione, affording a TT-amide.¹⁷ These activated amides occur as yellow crystals, facilitating monitoring of the reaction with the β-amino function of the diamino propionic acid by disappearance of the yellow color.

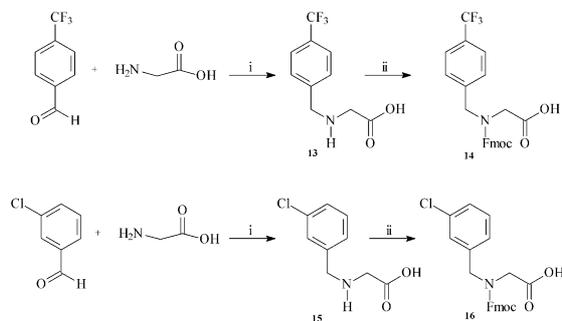
Apart from α-amino acids, we also introduced three peptoid monomers (CF3, Cbg, Sar) in the library. Preparations of CF3 and Cbg proceeded through a reductive amination by using NaCNBH₃ in a mixture of water and 1,4-dioxane (3:2, v/v)¹⁸ thereby reducing the in situ formed imine and retaining the unreacted aldehyde, followed by Fmoc protection (Scheme 3).

Synthesis and screening of the oligopeptide library and individual peptides

Utilizing a Fmoc-strategy-based solid phase peptide synthesis^{19,20} an oligopeptide library was assembled²¹ with the general structure Ac-Arg-Y-X-Chi-Chi-Tal-Arg-CONH₂, with X and Y representing the eight amino acid building blocks as shown in Figure 1. The remaining residues and their sequences were retained from our previous selection procedures.^{10,11} The coupling conditions between different residues were optimized by synthesizing several test peptides with random sequences. The absence of side reactions and racemization, was verified by performing HPLC analysis and mass spectrometry of these test peptides after cleavage from the solid support.²²



Scheme 2. Synthesis of (2*S*)-3-[(1-Boc-1*H*-benzimidazol-5-yl)carbamoyl]amino-2-Fmoc-amino-propanoic acid (**6**) and (2*S*)-2-Fmoc-amino-3-[(3-pyridinylcarbonyl)amino] propanoic acid (**7**). (i) 1.2 equiv (Boc)₂O, 10% Na₂CO₃, 1,4-dioxane; 85%, (ii) 1.2 equiv 1,3-thiazolidine-2-thione, DIC, CH₂Cl₂; 62% (iii) (a) TFA, CH₂Cl₂ (1/1, v/v), (b) 0.95 equiv **8** or **9** in saturated NaHCO₃, 1,4-dioxane (1/1, v/v); 80%.



Scheme 3. Synthesis of (Fmoc-4-(trifluoromethyl)phenylmethyl]amino) acetic acid (**14**) and [(3-chlorophenyl)methyl]-Fmoc-amino)acetic acid (**16**) (i) 1.5 equiv NaCNBH₃ in H₂O, 1,4-dioxane (3/2, v/v); 60%; (ii) 1.2 equiv FmocCl, 10% Na₂CO₃, 1,4-dioxane; 85%.

An iterative deconvolution procedure was used for the identification of X and Y. The first libraries were synthesized with the aim of optimizing position Y. The mixture of peptides released from the solid support, containing eight peptides in every batch, was subjected to a solution phase screening against the dsDNA target (5'-AGATTGTGCAATGT-3'/3'-TCTAACACGTTACA-5') by gel shift experiments.¹¹ Mass spectrometric investigations, performed as outlined previously,²³ proved the presence of all eight peptides in every sublibrary. These screening experiments led to determination of **Cbg** and **Cha** as the amino acids possessing the strongest binding affinity at position Y.

In the next step, 16 peptides with general structures Ac-Arg-**Cbg**-X-Chi-Chi-Tal-Arg-CONH₂ and Ac-Arg-**Cha**-X-Chi-Chi-Tal-Arg-CONH₂ were synthesized, with X representing the amino acids shown in Figure 1, except Tal, Chi and Arg. Following cleavage from the solid support, the peptides were purified by reversed phase HPLC, identified by mass spectrometry (Table 4) and individually screened for binding to dsDNA. From the eight peptides with a general structure corresponding to Ac-Arg-**Cbg**-X-Chi-Chi-Tal-Arg-CONH₂, those with Cha, Cbg or CF₃ at position X revealed as the strongest binding compounds. In the Ac-Arg-**Cha**-X-Chi-Chi-Tal-Arg-CONH₂ series, the highest affinity was observed with position X occupied by Imi, Cbg or CF₃.

It should be noted that the gel electrophoresis in many cases does not show one single band as the result of a gel shift. Often a long smear is seen and it is not clear yet if this is due to the association of several peptides to one dsDNA (or the opposite) or to a precipitation of the dsDNA-peptide complex (change of solubility due to charge neutralization and increase of size of the complex). Therefore in many of the figures it is difficult to observe the complex formed between dsDNA and peptide, and the disappearance of the free dsDNA band was taken as criterion for the selection of dsDNA binding peptides.¹¹

In order to have an idea of the strength of binding of these compounds, dissociation constants were determined for the six selected oligopeptides. Gel retardation experiments were performed at different concentrations of the peptides in 0.1 M NaCl PBS buffer, and the

apparent dissociation constants were determined as the concentration at which 50% of the target dsDNA was mobility shifted.²⁴ Although there was not a significant difference between the affinities of the peptides (Table 2), they did show a binding strength which was more than one order of a magnitude higher than the reference peptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂ with a K_d of 4×10^{-4} . The peptide Ac-Arg-**Cbg**-**Cha**-Chi-Chi-Tal-Arg-CONH₂ was endowed with the lowest dissociation constant of 2.0×10^{-5} .

Modification of the selected peptide with intercalators

It has been well documented from studies with the combilexin model that the introduction of intercalating heteropolyaromatic moieties to a ligand may increase the binding affinity to dsDNA as well as enhance the sequence specificity, particularly in the GC rich region.⁴ This strategy was already applied in previous experiments, where the peptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂ was conjugated to different intercalators by a variety of linker arms.²⁵ Therefore, the influence on the binding strength and the sequence selectivity of coupling intercalating agents to the peptide (Cbg-Cha) was investigated. As it is not clear yet at which position the peptides are bound to dsDNA, it is difficult to select an optimized spacer between peptide and intercalator based on modeling experiments or previous research. Therefore we evaluated different spacers to connect the intercalating agent to the peptide. This could be important for an optimal positioning of the intercalator with respect to the base sequences of dsDNA.^{26–28}

In the present study, two kinds of intercalators e.g., acridine-9-carboxylic acid and 9-fluorenone-2-carboxylic acid were evaluated and tethered to the terminal NH₂ function of H₂N-Arg-**Cbg**-**Cha**-Chi-Chi-Tal-Arg-CONH₂ via four sets of linkers composed of glycine and β -alanine (Figs 2 and 3). Moreover, two structurally more flexible linkers having three and five CH₂ groups between the amide functions e.g., γ -amino butyric acid (GABA) and 6-amino caproic acid (CAP) (Fig. 4) were evaluated by conjugation to acridine, so that we could compare the effect of the flexibility of the linkers on the binding strength.

The amino acid building blocks and linkers were introduced using standard procedures for amide formation. The carboxylic function of the intercalators were activated and coupled to the terminal NH₂ group of the peptide linker construct using 4 equiv HATU / DIEA in DMF. The peptides were purified by reversed phase HPLC and identified using mass spectrometry (Figs 2–4).

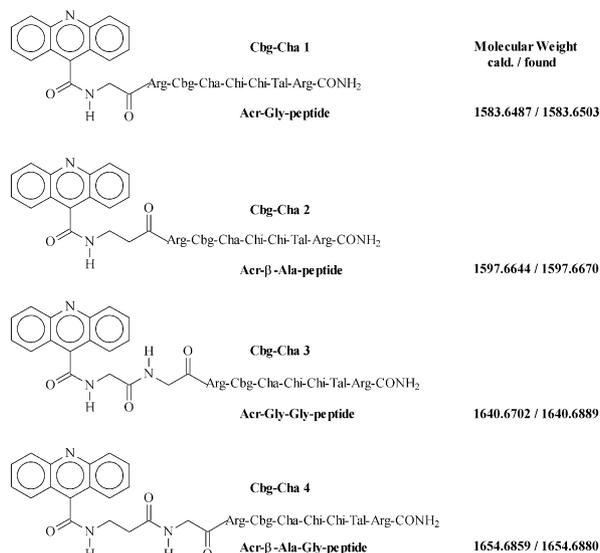
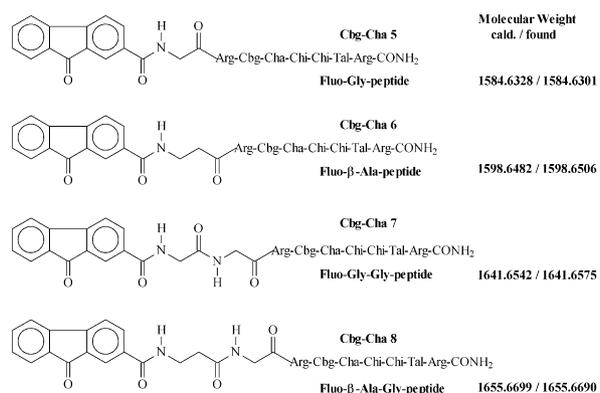
The solution screening of the ten oligopeptides with intercalators tethered by an α , β , γ or ϵ amino acid linker, revealed that the acridine-peptide with GABA as the linker shows a stronger binding than the other compounds and than the unconjugated peptide. (Figs 5 and 6). Comparison to the dipeptide linkers (Cbg-Cha3, Cbg-Cha4, Cbg-Cha7 or Cbg-Cha8), indicated that the removal of the amide bond and the increased flexibility

Table 2. Dissociation constants of the selected oligopeptide sequences compared to the peptide sequence obtained from previous work [(a) reference peptide]¹¹

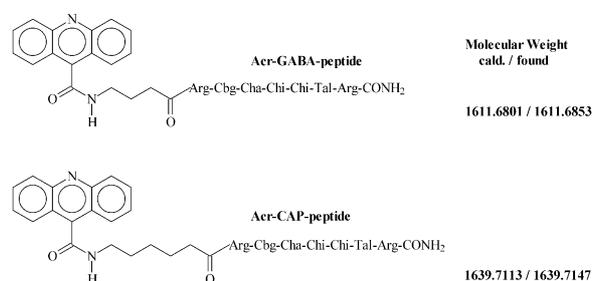
Oligopeptide sequence	Formula	K_D
Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH ₂ (a)	C ₅₄ H ₆₈ N ₂₂ O ₁₆	4.0 × 10 ⁻⁴
Ac-Arg-Cha-Imi-Chi-Chi-Tal-Arg-CoNH ₂	C ₆₄ H ₈₁ N ₂₃ O ₁₅	3.4 × 10 ⁻⁵
Ac-Arg-Cha-Cbg-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₉ N ₂₀ O ₁₄ Cl ₁	4.0 × 10 ⁻⁵
Ac-Arg-Cha-CF ₃ -Chi-Chi-Tal-Arg-CONH ₂	C ₆₃ H ₇₉ N ₂₀ O ₁₄ F ₃	2.6 × 10 ⁻⁵
Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₉ N ₂₀ O ₁₄ Cl ₁	2.0 × 10 ⁻⁵
Ac-Arg-Cbg-Cbg-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₂ N ₂₀ O ₁₄ Cl ₂	2.4 × 10 ⁻⁵
Ac-Arg-Cbg-CF ₃ -Chi-Chi-Tal-Arg-CONH ₂	C ₆₃ H ₇₂ N ₂₀ O ₁₄ Cl ₁ F ₃	4.0 × 10 ⁻⁵

of the linker has a beneficial effect on the binding and allows the intercalator to make a better contact with dsDNA.

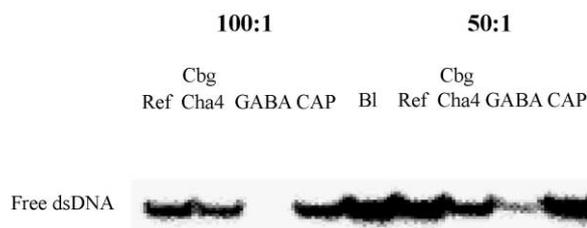
These results might be explained in a way that, in case of the α , β and ϵ linkers, there is an absence of a synergistic effect of binding by the peptide and the intercalating agent. Both parts of the molecule seem to be non-compatible for binding using these spacer groups. As the dipeptide linkers show limited flexibility due to the presence of two or more amide groups, it is indeed

**Figure 2.** Selected oligopeptide tethered to acridine by different linkers and their corresponding codes. Mass spectrometry confirmed the molecular weight of the isolated compounds.**Figure 3.** Oligopeptide Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ tethered with 9-fluorenone by different linkers.

not unrealistic to hypothesize that the lack of efficiency of the intercalators might be partially due to the difficulty of the polyaromatic rings to be positioned in the right way for optimal binding to dsDNA, or alternatively, optimal binding of the peptide is prohibited by positioning of the intercalating agent. Nevertheless, it was important to observe from these comparisons that the properties of the linkers did have an important impact on the outcome of the binding affinity.

**Figure 4.** Selected oligopeptide tethered with acridine by γ -amino butyric acid (GABA) and ϵ -amino caproic acid (CAP).

Cbg Cbg Cbg Cbg Cbg Cbg Cbg Cbg
Cha 1 Cha 2 Cha 3 Cha 4 Bl Ref Cha 5 Cha 6 Cha 7 Cha 8

**Figure 5.** Gel shift experiments of the selected oligopeptide modified with the introduction of different intercalators e.g., acridine for Cbg-Cha 1 to Cbg-Cha 4; fluorenone for Cbg-Cha 5 to Cbg-Cha 8 which were tethered by different linkers composed of glycine and β -alanine. The blank lane (Bl) contains no peptide, while the reference lane (Ref) contained the peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂. The same concentrations were used for all peptides.**Figure 6.** A comparison of binding affinity of oligopeptides either without intercalators (Ref) or with intercalators tethered by a dipeptide linker (Cbg-Cha 4) or a more flexible spacer (GABA, CAP) against dsDNA at different ligand/target ratios e.g., 100:1 and 50:1. The blank lane (Bl) contained no peptide and the reference lane the unconjugated peptide.

The apparent dissociation constant of the peptide tethered with the acridine derivative via a GABA linker was determined. In comparison with the peptide without intercalator ($K_d = 2.0 \times 10^{-5}$ M), a 10-fold enhancement of the binding capacity ($K_d = 2.1 \times 10^{-6}$ M) was observed (Table 3). When the binding constant was determined at higher salt concentration (1.0 M NaCl) the K_d value increased, which might point to the importance of ionic interactions between dsDNA and the peptides.

Sequence specificity of the selected oligopeptide and conjugates

To investigate the influence of the Ual-Sar modification and conjugation to intercalators on the sequence selectivity of the peptide, DNase I footprinting experiments were performed. A 271 bp fragment derived from the cloned plasmid pUC 19 and containing the 14-mer target used for screening, was applied as substrate, as described previously.¹¹ A typical autoradiograph of a sequencing gel used to fractionate the products of partial digestion of the DNA in the presence of the oligopeptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ and the Acridine-GABA conjugate is presented in Figure 7.

With the products bound, the DNase I cleavage pattern differs significantly from that seen in the control lane. Numerous bands in the product-containing lanes are weaker than the same bands in the product-free lanes, corresponding to attenuated cleavage, whereas others display relative enhancement of cutting. To obtain more detailed information, intensities from selected gel lanes were quantified by densitometry and converted into a set of differential cleavage plots (Fig. 8) indicating the extent to which cleavage at each internucleotide bond is

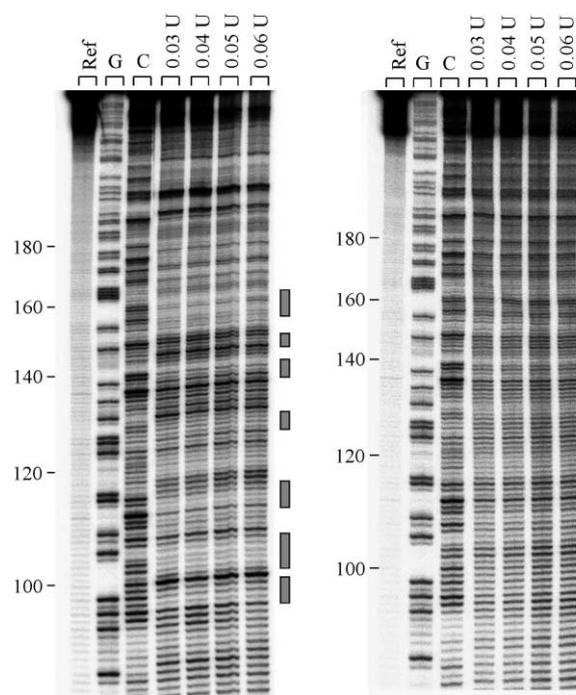


Figure 7. DNase I footprinting with the 271-mer Pvu II–Ase I restriction fragment of cloned plasmid pUC 19 in the presence of the oligopeptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ (left) and 10 μ M of the conjugate oligopeptide-GABA-Acridine (right). Reference track (Ref) contains dsDNA that was not digested. Control track (C) contained no product. Guanine-specific sequence markers were run in the lane marked G. Numbers at the side of the gel refer to the numbering scheme of the fragment. Preferential binding sites are indicated with gray boxes.

affected by complexation with the investigated ligands. The plots in Figure 8 compare the footprinting data obtained with 50 μ M Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-

Table 3. Comparison of the apparent dissociation constants (M) of the initial lead compound (a), the compound obtained from the peptide optimization procedure (b) and the compound developed during the intercalator optimization procedure (c)

Oligopeptide sequences	K_D
(a) Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH ₂	0.1 M NaCl 4.0×10^{-4}
(b) Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH ₂	2.0×10^{-5}
(c) Acr-GABA-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH ₂	2.1×10^{-6}

Table 4. Oligopeptides synthesized for the selection of amino acid X in the sequence Ac-Arg-Cha/Cbg-X-Chi-Chi-Tal-Arg-CONH₂

Oligopeptide sequences	Formula	Exact mass calculated for $[M + H]^+$	Exact mass found
Ac-Arg-Cha-PUal-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₈₀ N ₂₂ O ₁₆	1389.6201	1389.6200
Ac-Arg-Cha-Ual-Chi-Chi-Tal-Arg-CONH ₂	C ₆₀ H ₇₈ N ₂₂ O ₁₆	1363.6044	1363.6025
Ac-Arg-Cha-Sar-Chi-Chi-Tal-Arg-CONH ₂	C ₅₆ H ₇₆ N ₂₀ O ₁₄	1253.5928	1253.5927
Ac-Arg-Cha-Cha-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₈₆ N ₂₀ O ₁₄	1335.6711	1335.6705
Ac-Arg-Cha-Imi-Chi-Chi-Tal-Arg-CONH ₂	C ₆₄ H ₈₁ N ₂₃ O ₁₅	1412.6361	1412.6355
Ac-Arg-Cha-Cbg-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₉ N ₂₀ O ₁₄ Cl ₁	1363.5851	1363.5873
Ac-Arg-Cha-CF ₃ -Chi-Chi-Tal-Arg-CONH ₂	C ₆₃ H ₇₉ N ₂₀ O ₁₄ F ₃	1397.6114	1397.6106
Ac-Arg-Cha-Nic-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₈₀ N ₂₂ O ₁₅	1373.6251	1373.6273
Ac-Arg-Cbg-PUal-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₃ N ₂₂ O ₁₆ Cl ₁	1417.5341	1417.5343
Ac-Arg-Cbg-Ual-Chi-Chi-Tal-Arg-CONH ₂	C ₆₀ H ₇₁ N ₂₂ O ₁₆ Cl ₁	1391.5185	1391.5171
Ac-Arg-Cbg-Sar-Chi-Chi-Tal-Arg-CONH ₂	C ₅₆ H ₆₉ N ₂₀ O ₁₄ Cl ₁	1281.5068	1281.5082
Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₉ N ₂₀ O ₁₄ Cl ₁	1363.5851	1363.5853
Ac-Arg-Cbg-Imi-Chi-Chi-Tal-Arg-CONH ₂	C ₆₄ H ₇₄ N ₂₃ O ₁₅ Cl ₁	1440.5501	1440.5535
Ac-Arg-Cbg-Cbg-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₂ N ₂₀ O ₁₄ Cl ₂	1391.4992	1391.5082
Ac-Arg-Cbg-CF ₃ -Chi-Chi-Tal-Arg-CONH ₂	C ₆₃ H ₇₂ N ₂₀ O ₁₄ Cl ₁ F ₃	1425.5255	1425.5256
Ac-Arg-Cbg-Nic-Chi-Chi-Tal-Arg-CONH ₂	C ₆₂ H ₇₃ N ₂₂ O ₁₅ Cl ₁	1401.5392	1401.5387

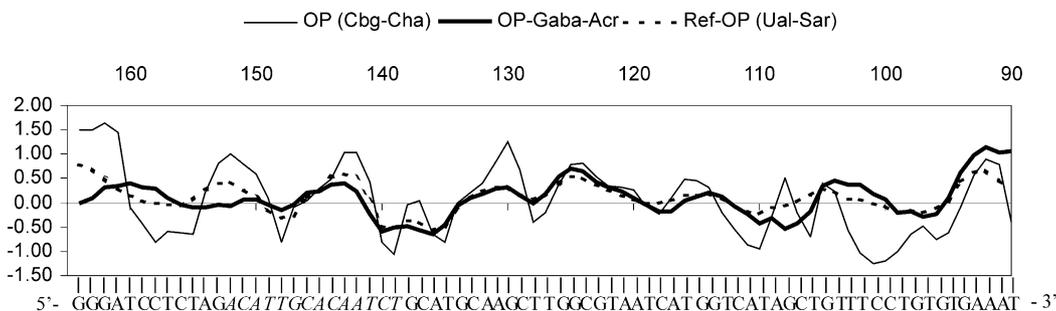


Figure 8. Differential cleavage plot comparing the susceptibility of the 3'-labeled 271 bp fragment to DNase I cutting in the absence and in the presence of 50 μM of the oligopeptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ [OP (Cbg-Cha)], 150 μM of Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-NH₂ [OP (Ual-Sar)] and 10 μM Acr-GABA-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ [OP-GABA-Acr]. Negative values correspond to a ligand-protected site and positive values represent enhanced cleavage. Vertical scale, units of $\ln(f_a) - \ln(f_c)$, where f_a is the fractional cleavage at any bond in the presence of the oligopeptide and f_c is the fractional cleavage of the same bond in the control. Data are derived from quantitative analysis of several sequencing gels like the one shown in Figure 7 and must be considered a set of average values.

NH₂ [OP (Cbg-Cha)], 10 μM Acr-GABA-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ [OP-GABA-Acr] and 150 μM of Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-NH₂ [Ref-OP (Ual-Sar)].

The differential cleavage plots show that the peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ binds to several sequences of varying length (from two to nine base pairs). The sites with pronounced inhibited cleavage are situated at nucleotide positions 95–103 (5'-TTCCTGTG-3') and 155–160 (5'-TCCTCT-3'). Other sites like 106–107 (5'-CT-3'), 113–117 (5'-TCATA-3'), 127–128 (5'-TT-3'), 135–136 (5'-AT-3'), 139–140 (5'-CT-3') and 147–149 (5'-TTG-3') have lesser base pairs involved. The sequences contain many sequential cytosines and thymines, indicating a relative pyrimidine specificity. To some extent, practically all multiple cytosine and thymine sequences are targeted. Comparison with the cleavage pattern of the reference oligopeptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-NH₂ shows that the pattern of the Cbg-Cha peptide is strongly pronounced showing multiple sites of increased protection from cleavage, in contrast to the single low intensity protected site of the reference peptide. Furthermore, pyrimidine specificity can be clearly identified for the modified oligopeptide in comparison with the Ual-Sar containing compound. The more hydrophobic side chains of Cbg and Cha and/or the different groove geometry and recognition pattern at the level of polypyrimidine tracts, could account for this pyrimidine selectivity, opening new possibilities to investigate the usefulness of analogous compounds in dsDNA sequence-specific targeting.

The fact that in some cases only two base pair sequences seem to be protected from cleavage, might suggest that in these cases only part of the oligopeptide is bound, while the other amino acids ensure a right positioning.

The acridine-GABA conjugate shows, however, not the same inhibited cleavage pattern as with the oligopeptide containing Cbg-Cha. Only two sites display pronounced protection against DNase I, namely at nucleotide positions 105–113 (5'-TCATAGCTG-3') and 134–141 (5'-TCTGCATG-3'). The latter sequence corresponds to the main protected site by the reference oligopeptide

Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂. The relative pyrimidine selectivity of the peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ is lost and the specificity of the reference peptide is restored.

Discussion

With the sequencing of the human genome accomplished, the functions of genes in cellular and in vivo models have to be determined. Once a target has been validated, drug discovery efforts will start. The demand for molecules recognizing sequence-specifically dsDNA or RNA and which can be used for gene functionalization and target validation will increase exponentially during coming years. Moreover, such molecules will provide a basis for new and more selective therapeutic approaches. The biochemical and structural studies in the past decade have already indicated that the details of ligand-DNA interaction are imposingly complex. The DNA sequence can be read by recognition of hydrogen bonding patterns along the floor of the grooves. Sequence dependent local deformability of the DNA duplex may be part of the recognition process. In addition, indirect readout elements and non-specific interactions may be involved in positioning the ligand in an entropy favored orientation for sequence-specific recognition.

The selection of high binding ligands can be accomplished by screening large numbers of molecules to discover an initial lead with modest binding activity followed by defining the key recognition elements for maximal activity and the optimization of the lead structure by synthesis of focussed libraries or of individual compounds. The different functionalities of the library might be placed on a template whose structure is based upon an important molecular recognition motif in a biological system. Therefore we started synthesis of peptide libraries, consisting of modified amino acids with unique side chains or constrained backbones.

In the first study we used solid phase screening to select an initial lead sequence.¹⁰ In the second round, solution phase screening was used to extend this sequence with functionalities involved in the recognition of the target,

yielding lead compounds as the peptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂.¹¹ In the present study we further optimized the binding making use of a library deconvolution technique. We selected eight new unnatural amino acid residues containing a diversity of functional groups and introduced them at the two random positions of Ac-Arg-Y-X-Chi-Chi-Tal-Arg-CONH₂, as Ual and Sar seemed to contribute to a lesser extent to the interaction capacity of the oligopeptide. Since in water, correct hydrogen bond complementarities may not add much to the stability of the complex in contrast to hydrophobic interactions, several of the selected amino acids had an apolar side chain.

Out of a pool of 64 peptides (8 sublibraries of 8 compounds) followed by the evaluation of 16 individual peptides, 6 peptides were selected and their binding strength was evaluated by measuring dissociation constants in 0.1 M NaCl at pH 7.4. The peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ demonstrated the lowest K_d value, which was one order of magnitude lower than the initial lead compound. The exchange of Ual and Sar at position X and Y by, respectively, Cbg and Cha enhanced the binding affinity of the peptide to dsDNA. The new residues are more likely involved in hydrophobic interactions due to the presence of cyclohexane and a substituted benzene ring, whilst the former residues possessing a uracil base are expected to be more involved in hydrogen bonding interaction. Although the importance of hydrogen bonding interactions in the recognition of dsDNA has been widely accepted, their significance for the determination of binding affinity might be overestimated. Geometrical and electrostatic factors might be the prime recognition system superimposed by hydrogen bonding and van der Waals interaction.²⁶ Recent studies within the combilexin model demonstrated that the introduction of an intercalating heteropolyaromatic moiety to molecules might increase the binding affinity to dsDNA.²⁹ Moreover, DNA binding molecules with a hybrid structure containing a peptide fragment and an intercalating fragment are well known e.g., actinomycin D, echinomycin, triostin and luzopeptin. Previous research showed already that conjugation of intercalators to the peptide lead structures has a beneficial effect on the binding strength.²⁵ With this in mind, an intercalating chromophore was tethered on our selected peptide. Because of its synthetic accessibility, acridine is extensively considered as the prototype of the intercalators.^{30,31} Due to the unique AT specificity of tilorone, different from most intercalators with GC specificity,³² its core structure, fluorenone was applied as second intercalator for evaluation. The spacer between the intercalator and the peptide is very important and should be designed so that it can deliver and optimally orient the intercalating agent (without being too long) so that the molecule can be bound to DNA with a minimal increase in entropy.^{33,34} Hence, four sets of linkers composed of glycine and β -alanine were connected to two intercalators and their influence on the binding affinity of the reference peptide to the same dsDNA was determined. Moreover, two flexible linkers with γ -amino butyric acid (GABA) and ϵ -amino caproic acid (CAP) were used as

well for tethering with the acridine intercalator. Based upon gel retardation experiments, out of these ten new peptides, a beneficial effect on the binding affinity was shown upon connecting acridine to the peptide with the GABA linker. The peptide Acr-GABA-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-CONH₂ demonstrated an improvement on the binding strength ($K_d = 2.1 \times 10^{-6}$ M) by one order of magnitude compared to the unconjugated peptide ($K_d = 2 \times 10^{-5}$ M). Hybrid molecules containing the lead peptide structure Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-CONH₂, were also endowed with affinities in the μ M range. These affinities are comparable to those described for known DNA binding ligands. 9-Aminoacridine-4-carboxamide, an intercalator with a limited extended aliphatic chain, demonstrates a K_d value of 4.5×10^{-6} M for poly (dA-dT) and of 2.0×10^{-6} M for poly (dG-dC).³⁵ A synthetic combilixin, NetGA, composed of netropsin and acridine, was reported to have a K_d value of 1.1×10^{-6} M.²⁶ Thus modification of the lead peptide by introducing acridine with GABA as the linker delivered a DNA binding agent with binding strength similar to that of known compounds.

The sequence specificity of the strongest binding peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ was investigated and compared with the reference compound Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-NH₂. The footprints showed a pronounced pyrimidine sequence selectivity, which was not observed with the Ual-Sar containing peptide. Probably the apolar side chains of Cbg and Cha are important for this interaction. Footprinting experiments with the newly developed conjugate acridine-GABA-OP revealed that the pyrimidine selectivity is lost, with the conjugate protecting analogous sites as the reference peptide containing Ual-Sar. Perhaps the intercalator is hampering interaction of the hydrophobic side chains of the Cbg and Cha amino acids, probably responsible for the CT-specificity, restoring the selectivity of the other part of the peptide, Chi-Chi-Tal-Arg. Further optimization of the structure of these peptides and conjugates and investigations of their binding mode are the subject of present research.

Conclusion

In our search for sequence-specific dsDNA binding ligands, the oligopeptide Ac-Arg-Ual-Sar-Chi-Chi-Tal-Arg-NH₂ with moderate affinity and selectivity was modified. The amino acids Ual and Sar, displaying less interaction capacities than the other side chains involved, were replaced with new unnatural building blocks. The oligopeptides with more hydrophobic side chains on these positions were endowed with a 10-fold increased affinity. The sequence specificity of the strongest binding peptide Ac-Arg-Cbg-Cha-Chi-Chi-Tal-Arg-NH₂ showed a pronounced pyrimidine sequence selectivity, which was not observed with the Ual-Sar containing peptide. Conjugation of an intercalator destroyed this specificity and restored interaction with the originally protected sequence by the Ual-Sar containing compound. As such, new dsDNA ligands, based on combinations of unnatural amino acids and inter-

calators and showing medium affinities and variable sequence selectivities have been developed. This approach of modifying dsDNA binding oligopeptides opens the possibility to develop, by combinatorial selection, a new class of dsDNA binding agents with a wide range of sequence selectivity.

Experimental

General methods

The ^1H and ^{13}C NMR spectra were recorded with a Varian Gemini 200 spectrometer in the solvent indicated. For ^1H spectra taken in CDCl_3 tetramethylsilane (TMS) was used as reference. For spectra taken in $\text{DMSO}-d_6$, the solvent signal at 2.5 ppm was used as reference. For the ^{13}C spectra, the solvent peaks were used as reference e.g., 77.00 ppm for CDCl_3 and 39.60 ppm for $\text{DMSO}-d_6$. Exact mass measurements were performed on a quadrupole/orthogonal-acceleration time-of-flight (Q/oaTOF) tandem mass spectrometer (qTof 2, Micromass, Manchester, UK) equipped with a standard electrospray ionization (ESI) interface. Samples were infused in a 2-propanol/water (1:1) mixture at 3 $\mu\text{L}/\text{min}$. In some cases, exact mass data were acquired on a Kratos Concept IH double focussing mass spectrometer (Kratos, Manchester, UK) with a MASPEC 2 data system (MSS Ltd., Manchester, UK). Samples were dissolved in a suitable matrix and ionized by liquid secondary ion mass spectrometry (LSIMS). Melting point was measured with a Büchi SMP-20. Precoated Machery-Nagel Alugram[®] SilG/UV 254 plates were used for TLC and spots were examined with UV light, sulfuric acid/anisaldehyde spray, or ninhydrine spray. Column chromatography was performed on Acros silica gel (0.06–200 nm). Anhydrous solvents were obtained as follows: pyridine and *N,N*-diisopropylethylamine (DIEA) were refluxed overnight over potassium hydroxide and distilled; *N,N*-dimethylformamide (DMF) was stored on molecular sieves for 3 days and was tested for the absence of dimethylamine by the bromophenol test prior to use. Tetrahydrofuran (THF) was refluxed over sodium/benzophenone under nitrogen and distilled. CH_3CN for HPLC was purchased from Rathburn (grade S) and water for HPLC purification was distilled twice. Rink amide MBHA resin was supplied by Novabiochem (Laufelfingen, Switzerland). Dichloromethane (DCM), *N,N*-dimethylformamide, acetic anhydride (Ac_2O) and pyridine were obtained from BDH (Poole, UK). 1-Hydroxy-7-azabenzotriazole (HOAt), Fmoc- β -alanine (β -Ala), and Fmoc-glycine (Gly) were purchased from Advanced ChemTech (Louisville, Kentucky). *trans*-4-Hydroxy-L-proline, *di-tert*-butyldicarbonate [$(\text{Boc})_2\text{O}$], iodomethane, uracil, benzoyl chloride, diethylazodicarboxylate (DEAD), triphenylphosphine (PPh_3), 9-fluorenylmethyloxycarbonyl chloride, 1,3-thiazolidine-2-thione (TT), 1H-benzimidazole-5-carboxylic acid, 3-pyridine carboxylic acid, 3-chlorobenzaldehyde, 4-(trifluoromethyl)benzaldehyde, sodium cyanoborohydride (NaCNBH_3), acridine-9-carboxylic acid, 9-fluorenone-2-carboxylic acid, γ -amino butyric acid (GABA) and 6-amino caproic acid (CAP),

piperidine, trifluoroacetic acid (TFA), diisopropylcarbodiimide (DIC) and 1-hydroxybenzotriazole (HOBT) were supplied by ACROS (Geel, Belgium). *N* ^{α} -(9-Fluorenylmethyloxycarbonyl)-*N* ^{β} -*tert*-butyloxycarbonyl-L- α -2,3-diaminopropionic acid was synthesized as described.¹⁶ Fmoc protection of the primary amine of both GABA and CAP was realized under standard Schotten-Baumann conditions. The two oligonucleotides which served as target were synthesized on an Applied Biosystem392 DNA synthesizer with phosphoramidites from Applied Biosystems and were worked up as described previously.³⁶ T4 polynucleotide kinase and DNase I were purchased from Gibco BRL and $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ was supplied by ICN. NAP-5[®] columns were obtained from Pharmacia.

Synthesis of the unnatural amino acids. *N-tert*-Butoxycarbonyl-*cis*-4-(*N*³-benzoyluracil-1-yl)-L-proline methyl ester (3). Protection of the α -amino function of *trans*-4-hydroxy-L-proline with a Boc group and protection of the carboxylic function as methyl ester was described previously.¹⁴ To a suspension of the alcohol **2** (0.245 g, 1.0 mmol), *N*³-benzoyluracil (0.216 g, 1.0 mmol) and triphenylphosphine (294 mg, 1.1 mmol) in dry THF (10 mL) was added DEAD (182 μL , 1.1 mmol) dropwise at -15°C . The reaction mixture was stirred at room temperature overnight. The clear solution was evaporated to dryness and the residue was purified by column chromatography (CH_2Cl_2 -acetone 95:5) to give *N-tert*-butoxycarbonyl-*cis*-4-(*N*³-benzoyluracil-1-yl)-L-proline methyl ester **3** as a foam (0.363 g, 0.82 mmol, 82%).

^1H NMR (CDCl_3) δ 1.41 (s, 9H, Boc- CH_3), 2.00–2.25, 2.65–2.95 (m, 2H, 3-H), 3.60–3.70, 3.85–4.10 (m, 2H, 5-H), 3.72 (s, 3H, Me ester), 4.25–4.45, 5.10–5.30 (m, 2H, 2-H, 4-H), 7.20–8.00 (m, 7H, aromatics) ppm.

^{13}C NMR (CDCl_3) δ 172.88 (ester-CO), 168.60 (Bz-CO), 161.71 (uracil C-4), 156.70 (Boc-CO), 149.93 (uracil C-2), 140.49 (uracil C-6), 135.15 (benzoyl *p*-CH), 131.39 (benzoyl-C), 130.45 (benzoyl *m*-CH), 129.17 (benzoyl *o*-CH), 102.80 (uracil C-5), 81.28 (Boc-C), 58.60 (C-2), 52.96 (C-5), 52.41 (ester CH_3), 49.86 (C-4), 36.90, 34.38 (C-3), 28.07 (Boc- CH_3) ppm. Exact mass (LSIMS, thioglycerol) calcd for $\text{C}_{22}\text{H}_{25}\text{N}_3\text{O}_7$ [$\text{M} + \text{H}$]⁺ 444.1770; found 444.1795.

***N* ^{α} -(9-Fluorenylmethyloxycarbonyl)-*cis*-4-uracil-L-proline (4).** To a solution of compound **3** (0.85 g, 1.92 mmol) in 100 mL of water and 1,4-dioxane (1/1, v/v) was added sodium hydroxide (0.38 g, 9.60 mmol). After 1.5 h stirring at room temperature, the solution was acidified with 0.1 N HCl to pH 3 and the white suspension was extracted three times with ethyl acetate. The combined organic layer was dried on anhydrous MgSO_4 , filtered and evaporated to dryness. The solid was treated with a mixture of TFA and CH_2Cl_2 (1/1, v/v) at room temperature over 2 h. After evaporation, the resulting foam was dissolved in 10% Na_2CO_3 (5 mL). To this solution was added a solution of 9-fluorenylmethyloxycarbonyl chloride (0.60 g, 2.30 mmol) in 5 mL 1,4-dioxane at 0°C . The reaction mixture was stirred overnight and then poured into 100 mL of H_2O and the solution was

extracted three times with diethyl ether. The aqueous layer was acidified with 2 N HCl and the white suspension was extracted three times with ethyl acetate. The combined organic layer was dried on anhydrous MgSO₄, filtered and evaporated to afford compound **4** (0.6 g, 1.34 mmol, 70%).

¹H NMR (CDCl₃) δ 1.90–2.45 (m, 2H, 3-H), 3.29–3.90 (m, 2H, 5-H), 4.10–4.60 (m, 5H, 2-H, 4-H, Fmoc-CH₂, Fmoc-9'-H), 7.00–7.90 (m, 10H, aromatics) ppm.

¹³C NMR (CDCl₃) δ 172.90 (COOH), 163.37 (uracil C-4), 154.15 (OCONH), 151.20 (uracil C-2), 143.89 (Fmoc-10'), 142.64 (uracil C-6), 140.91 (Fmoc-11'), 127.95 (Fmoc-3'), 127.41 (Fmoc-2'), 125.46 (Fmoc-1'), 120.33 (Fmoc-4'), 101.55 (uracil C-5), 67.10 (Fmoc-CH₂), 57.45 (C-2), 53.65 (C-5), 48.52 (C-4), 46.70 (Fmoc-9'), 34.08, 32.92 (C-3) ppm.

Exact mass (ESI-MS) calcd for C₂₄H₂₁N₃O₆ [M + H]⁺ 448.1508; found 448.1490; Elem. anal. for C₂₄H₂₁N₃O₆ C₄H₈O₂ H₂O calcd C 60.75, H 5.64, N 7.59; found C 60.77, H 5.36, N 7.33.

1-[[1,1-Dimethylethyl]oxy]carbonyl]-1H-benzimidazole-5-carboxylic acid (11). To a solution of 6.5 g (40 mmol) of 1H-benzimidazole-5-carboxylic acid **10** in 100 mL of 10% Na₂CO₃ were added 60 mL of 1,4-dioxane and 10.5 g (48 mmol) of *di-tert*-butyldicarbonate at 0 °C. The reaction mixture was stirred overnight at room temperature and was washed three times with diethyl ether. The aqueous layer was acidified with 2 N HCl and the white suspension was extracted three times with ethyl acetate. The combined organic layer was dried over anhydrous MgSO₄, filtered and evaporated to afford compound **11** (8.9 g, 34 mmol, 85%).

¹H NMR (DMSO-*d*₆) δ 1.65 (s, 9H, Boc-CH₃), 7.80–8.80 (m, 4H, aromatics) ppm.

¹³C NMR (DMSO-*d*₆) δ 167.44 (COOH), 147.53, 147.01 (OCONH), 144.56, 143.74 (C-5), 134.30 (C-8), 131.08 (C-9), 126.41, 125.40 (C-6), 121.67 (C-2), 120.15 (C-4), 115.99, 114.23 (C-7), 86.07 (Boc-C), 27.58 (Boc-CH₃) ppm.

Exact mass (ESI-MS) calcd for C₁₃H₁₄N₂O₄ [M + H]⁺ 263.1031; found 263.1027.

1,1-Dimethylethyl 5-[(2-thiono-1,3-thiazolidin-3-yl)carbonyl]-1H-benzimidazole-1-carboxylate (8). Diisopropylcarbodiimide (5 mL, 33 mmol) was added to a stirred solution of **11** (8.6 g, 33 mmol) and 1,3-thiazolidine-2-thione (4.7 g, 40 mmol) in CH₂Cl₂ (150 mL). The mixture was stirred overnight at room temperature and a precipitate (diisopropylurea) was filtered off. Evaporation of the filtrate under reduced pressure left an oily residue which was purified by column chromatography (CH₂Cl₂-acetone 98:2) to give compound **8** as yellow powder (7.2 g, 19.8 mmol, 60%).

¹H NMR (CDCl₃) δ 1.71 (s, 9H, Boc-CH₃), 3.47–3.55 (t, 2H, *J* = 7.4 Hz, TT-CH₂), 4.54–4.61 (t, 2H, *J* = 7.4 Hz, TT-CH₂), 7.80–8.60 (m, 4H, aromatics) ppm.

¹³C NMR (CDCl₃) δ 202.20 (TT-2'), 171.27 (CO), 147.60 (OCON), 143.83, 143.53 (C-5), 134.54 (C-8), 129.84 (C-9), 127.59, 127.10 (C-2), 126.85, 125.99 (C-6), 122.83 (C-2), 120.43 (C-4), 116.73, 114.33 (C-7), 86.35 (Boc-C), 56.57 (TT-4'), 29.71 (TT-5'), 27.89 (Boc-CH₃) ppm.

Exact mass (ESI-MS) calcd for C₁₆H₁₇N₃O₃S₂ [M + H]⁺ 364.0789; found 364.0797.

(2S)-3-[[1-[[1,1-Dimethylethyl]oxy]carbonyl]-1H-benzimidazol-5-yl]carbonyl]amino]-2-[[9H-fluoren-9-ylmethyl]oxy]carbonyl]amino]propanoic acid (6). *N*^α-(9-Fluorenylmethyl)oxycarbonyl-*N*^β-*tert*-butyloxycarbonyl-*λ*-*α*-2,3-diaminopropionic acid (2.0 g, 4.7 mmol) was treated with TFA in CH₂Cl₂ (1/1, v/v) at room temperature over 2 h. After evaporation, the resulting foam was dissolved in saturated NaHCO₃ (10 mL). To this solution was added a solution of compound **8** (1.62 g, 4.47 mmol) in 10 mL of 1,4-dioxane. The reaction mixture was stirred overnight at room temperature and the yellow color disappeared. The suspension was acidified with 2 N HCl to pH 3 and extracted three times with ethyl acetate. The combined organic layer was dried over anhydrous MgSO₄, filtered and evaporated. The crude product was purified by column chromatography (CH₂Cl₂-MeOH 95:5) and crystallized from diethyl ether to afford compound **6** (2.0 g, 3.58 mmol, 80%).

¹H NMR (DMSO-*d*₆) δ 1.65 (s, 9H, Boc-CH₃), 3.50–3.80 (br, 2H, β-H), 4.10–4.40 (br, 4H, α-H, Fmoc-CH₂, Fmoc-9'-H), 7.45–8.80 (m, 14H, aromatics, 2×NH, COOH) ppm.

¹³C NMR (DMSO-*d*₆) δ 172.45 (COOH), 166.68 (NHCO), 156.30 (Fmoc-OCONH), 147.65 (Boc-OCONH), 144.89–114.05 (aromatic-C), 86.01 (Boc-C), 65.88 (Fmoc-CH₂), 53.99 (C-α), 46.73 (Fmoc-9'), 40.88 (C-β), 27.61 (Boc-CH₃) ppm.

Exact mass (LSIMS, diethanolamine) calcd for C₃₁H₃₀N₄O₇ [M-H]⁻ 569.2036; found 569.2052; Elem. anal. for C₃₁H₃₀N₄O₇·C₄H₈O₂ calcd C 63.80, H 5.82, N 8.51; found C 63.76, H 6.11, N 8.58.

3-(3-Pyridinylcarbonyl)-1,3-thiazolidine-2-thione (9). Diisopropylcarbodiimide (1.9 mL, 12 mmol) was added to a stirred solution of 3-pyridine carboxylic acid **12** (1.23 g, 10 mmol) and 1,3-thiazolidine-2-thione (1.43 g, 12 mmol) in CH₂Cl₂ (50 mL). The mixture was stirred overnight at room temperature and a precipitate (diisopropylurea) was filtered off. Evaporation of the filtrate under reduced pressure left an oily residue which was purified by column chromatography (DCM-acetone 95:5) to give compound **9** as yellow powder (1.4 g, 6.2 mmol, 62%).

¹H NMR (CDCl₃) δ 3.47–3.55 (t, 2H, *J* = 7.4 Hz, TT-CH₂), 4.55–4.62 (t, 2H, *J* = 7.4 Hz, TT-CH₂), 7.29–8.90 (m, 4H, aromatics) ppm.

¹³C NMR (CDCl₃) δ 202.37 (TT-2'), 169.33 (CO), 152.57 (C-2), 149.84 (C-4), 136.55 (C-6), 129.99 (C-3), 123.07 (C-5), 56.09 (TT-4'), 29.56 (TT-5') ppm.

Exact mass (ESI-MS) calcd for $C_9H_8N_2O_1S_2 [M+H]^+$ 225.0156; found 225.0145.

(2S)-2-((9H-Fluoren-9-ylmethoxy)carbonyl)amino)-3-((3-pyridinylcarbonyl)amino) propanoic acid (7). N^{α} -(9-Fluorenylmethoxycarbonyl)- N^{β} -*tert*-butyloxycarbonyl-L- α -2,3-diaminopropionic acid **5** (1.84 g, 4.33 mmol) was treated with TFA in CH_2Cl_2 (1/1, v/v) at room temperature for 2 h. After evaporation, the resulting foam was dissolved in a saturated solution of $NaHCO_3$ (10 mL). Hereto was added a solution of compound **9** (0.92 g, 4.12 mmol) in 10 mL of 1,4-dioxane. The reaction mixture was stirred overnight at room temperature and the yellow color disappeared. The suspension was acidified with 2 N HCl to pH 3 and extracted three times with ethyl acetate. The combined organic layer was dried on anhydrous $MgSO_4$, filtered and evaporated. The crude product was purified by column chromatography (CH_2Cl_2 -MeOH 90:10) and crystallized from diethyl ether to afford compound **7** (1.42 g, 3.30 mmol, 80%).

1H NMR ($DMSO-d_6$) δ 3.50–3.80 (br, 3H, β -H, α -H), 4.10–4.40 (br, 3H, Fmoc- CH_2 , Fmoc-9'-H), 7.45–8.90 (m, 12H, aromatics) ppm.

^{13}C NMR ($DMSO-d_6$) δ 172.17 (COOH), 164.47 (NHCO), 156.33 (OCONH), 149.53–120.33 (aromatic-C), 65.91 (Fmoc- CH_2), 53.59 (C- α), 46.73 (Fmoc-9'), 40.87 (C- β) ppm.

Exact mass (LSIMS, diethanolamine) calcd for $C_{24}H_{21}N_3O_5 [M-H]^-$ 430.1402; found 430.1439; Elem. anal. for $C_{24}H_{21}N_3O_5 \cdot 0.5H_2O$ calcd C 65.45, H 5.03, N 9.54; found C 65.92, H 5.03, N 9.43.

[[4-(Trifluoromethyl)phenyl]methyl]amino)acetic acid (13). To a suspension of 4-(trifluoromethyl)benzaldehyde (4 mL, 30 mmol) and glycine (1.8 g, 24 mmol) in 50 mL of dioxane and H_2O (1/1, v/v) was added sodium cyanoborohydride (2.4 g, 36 mmol). The pH of the reaction mixture was adjusted to 5–6 with 2 N HCl. After overnight stirring at room temperature, the mixture was acidified to pH 3–4 and evaporated to dryness. Hot MeOH was added to the residue and the suspension was filtered. The filtrate was absorbed on silica gel and subjected to column chromatography (CH_2Cl_2 /MeOH/ H_2O 60:35:5) yielding compound **13** (2.8 g, 12 mmol, 50%).

1H NMR (D_2O) δ 3.95 (s, 2H, α - CH_2), 4.36 (s, 2H, Bn- CH_2), 7.50–7.90 (d, 2H, $J=8$ Hz; d, 2H, $J=8$ Hz, aromatics) ppm.

^{13}C NMR (D_2O) δ 171.68 (COOH), 137.04–129.03 (aromatics), 52.60 (Bn- CH_2), 49.57 (C- α) ppm.

Exact mass (ESI-MS) calcd for $C_{10}H_{10}N_1O_2F_3 [M+H]^+$ 234.0742; found 234.0744.

[[9H-Fluoren-9-ylmethoxy]carbonyl]-[[4-(trifluoromethyl)phenyl]methyl]amino) acetic acid (14). To a solution of 1.62 g (7.0 mmol) of **13** in 50 mL of 10% Na_2CO_3 was added 30 mL of 1,4-dioxane and 2.0 g (7.6 mmol) of

9-fluorenylmethoxycarbonyl chloride at 0°C. The reaction mixture was stirred overnight at room temperature and was poured into 120 mL of H_2O after which the mixture was washed three times with diethyl ether. The aqueous layer was acidified with 2 N HCl and the white suspension was extracted three times with ethyl acetate. The combined ethyl acetate layer was dried on anhydrous $MgSO_4$, filtered and evaporated to dryness. The crude product was crystallized from methanol to afford compound **14** (2.7 g, 5.95 mmol, 85%).

1H NMR ($DMSO-d_6$) δ 3.92 (s, 2H, α - CH_2), 4.10–4.60 (m, 5H, Fmoc- CH_2 , Fmoc-9'-H, Bn- CH_2), 7.15–7.90 (m, 12H, aromatics) ppm.

^{13}C NMR ($DMSO-d_6$) δ 171.08, 170.90 (COOH), 156.15, 155.91 (OCON), 143.95–120.24 (aromatic-C), 67.34, 67.01 (Fmoc- CH_2), 51.10, 50.69 (Bn- CH_2), 49.19, 48.55 (C- α), 46.85, 46.64 (Fmoc-9') ppm.

Exact mass (LSIMS, thioglycerol: NaOAc) calcd for $C_{25}H_{20}N_1O_4F_3 [M+2Na-H]^+$ 500.1062; found 500.1087; Elem. anal. for $C_{25}H_{20}N_1O_4F_3$ calcd C 65.93, H 4.43, N 3.08; found C 65.74, H 4.37, N 2.71.

[[3-Chlorophenyl]methyl]amino)acetic acid (15). To a suspension of 3-chlorobenzaldehyde (2.83 mL, 25 mmol) and glycine (1.5 g, 20 mmol) in 50 mL of dioxane and H_2O (1/1, v/v) was added sodium cyanoborohydride (2.0 g, 30 mmol). The pH of the reaction mixture was adjusted to 5–6 with 2 N HCl. After overnight stirring at room temperature, the mixture was acidified to pH 3–4 and evaporated to dryness. Hot MeOH was added to the residue and the suspension was filtered. The filtrate was absorbed on silica gel and purified by column chromatography (CH_2Cl_2 /MeOH/ H_2O 60:35:5) yielding compound **15** (2.4 g, 12 mmol, 60%).

1H NMR (D_2O) δ 3.86 (s, 2H, α - CH_2), 4.24 (s, 2H, Bn- CH_2), 7.30–7.55 (m, 4H, aromatics) ppm.

^{13}C NMR (D_2O) δ 169.78 (COOH), 134.95 (C-5), 132.75 (C-1), 131.42 (C-4), 130.53 (C-6), 130.48 (C-2), 128.99 (C-3), 50.72 (Bn- CH_2), 47.45 (C- α) ppm.

Exact mass (ESI-MS) calcd for $C_9H_{10}N_1O_2Cl_1 [M+H]^+$ 200.0478; found 200.0489.

[[3-Chlorophenyl]methyl]amino)acetic acid (16). To a solution of 0.42 g (2.1 mmol) of **15** in 15 mL of 10% Na_2CO_3 was added 10 mL of 1,4-dioxane and 0.65 g (2.5 mmol) of 9-fluorenylmethoxycarbonyl chloride at 0°C. The reaction mixture was stirred overnight at room temperature and was poured into 120 mL of H_2O after which the mixture was washed three times with diethyl ether. The aqueous layer was acidified with 2 N HCl and the white suspension was extracted three times with ethyl acetate. The combined ethyl acetate layer was dried on anhydrous $MgSO_4$, filtered and evaporated to dryness. The crude product was crystallized from methanol to afford compound **16** (0.75 g, 1.78 mmol, 85%).

^1H NMR (DMSO- d_6) δ 3.91 (s, 2H, α -CH $_2$), 4.15–4.55 (m, 5H, Fmoc-CH $_2$, Fmoc-9'-H, Bn-CH $_2$), 7.15–7.90 (m, 12H, aromatics) ppm.

^{13}C NMR (DMSO- d_6) δ 171.08, 170.87 (COOH), 156.09, 155.84 (OCON), 143.92–120.27 (aromatic-C), 67.37, 67.13 (Fmoc-CH $_2$), 50.92, 50.50 (Bn-CH $_2$), 49.01, 48.40 (C- α), 46.82, 46.64 (Fmoc-9') ppm.

Exact mass (ESI-MS) calcd for C $_{24}$ H $_{20}$ N $_1$ O $_4$ Cl $_1$ [M+Na] $^+$ 444.0979; found 444.0995; Elem. anal. for C $_{24}$ H $_{20}$ N $_1$ O $_4$ Cl $_1$ calcd C 68.33, H 4.78, N 3.32; found C 68.05, H 4.88, N 2.71.

Solid phase peptide synthesis of the library and individual oligopeptides

The glycine substituted oligopeptides were prepared on solid support as described previously.¹¹ Fmoc-glycine (4 equiv) was coupled with the use of 4 equiv HOAt, DIC and DIEA in DMF.

The library and individual peptides corresponding to the general structure Arg-Y-X-Chi-Chi-Tal-Arg-CONH $_2$ were likewise assembled on solid support and the mix and split method was applied.²¹ An amount of 10 mg (5.4 μmol) solid support (Rink amide MBHA resin) was applied per peptide for library synthesis and a cocktail of 4 equiv amino acid, 4 equiv HOAt, 4 equiv DIC and 4 equiv DIEA in DMF was used for couplings. For reactions of amino acids with the peptoid monomers Cbg and CF3, 2 equiv PyBop was added to the mixture. Procedures used during the peptide synthesis cycle were as described.¹¹ Before each coupling, the unreacted amino groups were capped using a mixture of pyridine/Ac $_2$ O/*N*-methylimidazole 4:1:0.5 over 10 min while the Fmoc-protecting group was removed by 15 min treatment with 20% piperidine in DMF. In the case of modification by, the coupling cocktail was composed of 4 equiv intercalator, 4 equiv HATU, 4 equiv DIEA and DMF. The linkers γ -amino butyric acid (GABA) and 6-amino caproic acid (CAP) were introduced using 4 equiv HOAt, 4 equiv DIC, 4 equiv DIEA and 2 equiv PyBop in DMF. After the final coupling cycle, the Fmoc-protecting group was removed and the terminal amino group was acetylated. The oligopeptides were released from the bead by treatment with TFA/H $_2$ O 95:5 in the presence of 5% thioanisole for two h at room temperature. The mixture was filtered and the filtrate was evaporated and co-evaporated with toluene, affording the peptides as solid material. In the case of individual oligopeptides, purification by reversed phase preparative HPLC (PLRP-S $^{\text{R}}$ column) was implemented. The peptides were eluted from the column using a linear gradient from 40 to 100% B in 30 min, with a flow rate of 3 mL/min [elute A: 0.1% TFA in CH $_3$ CN–H $_2$ O (5/95); elute B: 0.1% TFA in CH $_3$ CN–H $_2$ O (80/20)], with UV detection at 220 nm. The major peak was isolated and analyzed by high resolution mass spectrometry.

Ethidium bromide displacement test¹³

Wells of Costar black 96-well plates were loaded with 2 μL of a 50 μM dsDNA solution, 2 μL of a 0.35 mM

EtBr solution and a varying volume of oligopeptides (individual or mixtures) to obtain the necessary concentrations. The appropriate volume of a Tris/NaCl buffer (10 mM Tris–10 mM NaCl pH, 7.4) was added to obtain a total volume of 100 μL per well. Before adding the DNA in the wells, it was rendered double stranded by taking equal amounts of both complementary strands and heating them for 3 min at 80 $^{\circ}\text{C}$, at room temperature for 5 min and at 4 $^{\circ}\text{C}$ for 20 min. After incubation at room temperature for 30 min, each well was read on a FL600 Microplate Fluorescence reader, with 530/25 nm as excitation wavelength and 590/35 nm as the emission detection wavelength. Two control wells (no agent = 100% fluorescence, no DNA = 0% fluorescence) were used per 12 samples. Fluorescence readings are reported as % fluorescence relative to the control wells. Generally two to three sets of measurements were performed to calculate average values.

Solution phase screening of the library and individual oligopeptides

The solution screening process was performed by gel shift experiments as previously outlined.¹¹ Both ^{32}P -radiolabeled strands of the target oligonucleotide were separately dissolved in PBS-buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, pH 7.4) at a 2 μM concentration. Equal volumes of both solutions were mixed and treated to allow hybridization of the DNA-strands (heating and cooling). The oligopeptides (individual or mixtures) and the conjugates were dissolved in an appropriate volume of PBS. In a total reaction volume of 5 μL (PBS), 1 μL of the oligonucleotide solution (containing 1 pmol dsDNA) was mixed with the volume of conjugate or oligopeptide solution, necessary to acquire the desired concentrations. The mixture was stored at 4 $^{\circ}\text{C}$ for 2 h. To the 'blank', no peptide was added. The mixtures were resolved at 10 $^{\circ}\text{C}$ on a 15% native polyacrylamide gel with TBE pH 7.4 running-buffer at 2 W/gel over 3 h. The gels were quantitatively imaged using a Cyclone Phosphorimager (Packard). The degree of complex formation was quantified by measuring the residual amount of free DNA using Optiquant $^{\text{TM}}$.

Apparent dissociation constant determination by gel mobility-shift assay

The dissociation constants of peptides were determined by gel mobility-shift assay as described in the literature.¹¹ A series of concentrations of the investigated peptides were prepared. These solutions were incubated with dsDNA solutions and the mixtures were subjected to gel shift analysis. K_{D} values were simply determined as the concentration at which 50% of the target dsDNA was mobility shifted. At least 3 independent gel shift assays were performed for each compound under study to determine the apparent K_{d} . Concentration-response curves, obtained by analysis of the gel shifts, were fitted with the use of the equation $Y = E_{\text{max}}/[1 + (K_{\text{d}}/C)^{\text{nH}}]$ with Y standing for the response (% shift), E_{max} for the maximal response and nH for the Hill coefficient using Graphpad Prism (Graphpad Software Inc.).

DNase I footprinting

Cleavage reactions by DNase I were performed essentially according to the protocol outlined previously.¹¹ The 271 dsDNA fragment used in these experiments was derived from the vector pUC 19, which contained the 14-mer target sequence for the screening assays. The preparation of the vector and the 271 fragment was performed as previously described.¹¹ Reactions were conducted in a total volume of 10 μ L containing 2 μ L labeled DNA fragment, 5 μ L of the buffered solution with the investigated molecules in the appropriate concentrations. After a 30 min equilibration period, cleavage was initiated via addition of 3 μ L of a DNase I solution of which the concentration was adjusted to obtain an enzyme attack of \sim 30% of the starting material (final enzyme concentrations of 0.0005 to 1 U/10 μ L reaction mixture). After 3 min, the reaction was stopped by freeze drying. The DNA cleavage products were resolved by polyacrylamide gel electrophoresis under denaturing conditions. Following electrophoresis (\sim 2.5 h. at 65 W in TBE pH 8.0 buffered solution), gels were soaked in 10% acetic acid for 5 min, transferred to Whatmann 3MM paper, dried under vacuum at 80 °C and then exposed to a storage phosphor screen (Pharmacia). G-tracks were obtained by treatment with dimethylsulphate (DMS), followed by heating in piperidine.

Acknowledgements

The work was supported by a grant from K.U.Leuven (GOA 02/13) and from the Flemish Science Foundation (G.0269.98 and G.0089.02). We thank Chantal Biernaux for editorial help.

References and Notes

- Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
- Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.
- White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260.
- Dervan, P. B. *Science* **1986**, *232*, 464.
- Goodsell, D.; Dickerson, R. E. *J. Med. Chem.* **1986**, *29*, 727.
- Harshman, K. D.; Dervan, P. B. *Nucleic Acids Res.* **1985**, *13*, 4825.
- Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3260.
- Portugal, J.; Waring, M. J. *Biochim. Biophys. Acta* **1988**, *949*, 158.

- Hyrup, B.; Nielsen, P. E. *Bioorg. Med. Chem.* **1996**, *4*, 15.
- Lescrinier, T.; Hendrix, C.; Kerremans, L.; Rozenski, J.; Link, A.; Samyn, B.; Van Aerschot, A.; Lescrinier, E.; Eritja, R.; Van Beeumen, J.; Herdewijn, P. *Chem. Eur. J.* **1998**, *4*, 3425.
- Chaltin, P.; Lescrinier, E.; Lescrinier, T.; Rozenski, J.; Hendrix, C.; Rosemeyer, H.; Busson, R.; Van Aerschot, A.; Herdewijn, P. *Helv. Chim. Acta*: In press.
- Isshiki, H.; Akira, S.; Tanabe, O.; Nakajima, T.; Shimamoto, T.; Hirano, T.; Kishimoto, T. *Mol. Cell Biol.* **1990**, *10*, 2757.
- Boger, D. L.; Fink, B. E.; Brunette, S. R.; Winston, C. T.; Hedrick, M. P. *J. Am. Chem. Soc.* **2001**, *123*, 5878.
- Zhang, Z.; Herdewijn, P. *Curr. Med. Chem.* **2000**, *8*, 517.
- Lowe, G.; Vilaivan, T. *J. Chem. Soc., Perkin Trans. 1* **1997**, 539.
- Zhang, Z.; Van Aerschot, A.; Hendrix, C.; Busson, R.; David, F.; Sandra, P.; Herdewijn, P. *Tetrahedron* **2000**, *56*, 2513.
- Nagao, Y.; Miyasaka, T.; Seno, K.; Fujita, E.; Shibata, D.; Doi, E. *J. Chem. Soc., Perkin Trans. 1* **1984**, 2439.
- Borch, R. F.; Hassid, A. I. *J. Org. Chem.* **1972**, *37*, 101673.
- Carpino, L. A.; Han, G. Y. *J. Org. Chem.* **1972**, *37*, 223404.
- Chang, C.-D.; Felix, A. M.; Jimenez, M. H.; Meienhofer, J. *Int. J. Peptide Protein Res.* **1980**, *15*, 485.
- Lam, K. S.; Salmon, S. E.; Hersch, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. *Nature* **1991**, *82*, 354.
- Zhang, Z.; Van Aerschot, A.; Chaltin, P.; Busson, R.; Herdewijn, P. *Collect. Czech. Chem. Commun.* **2001**, *66*, 923.
- Rozenski, J.; Chaltin, P.; Van Aerschot, A.; Herdewijn, P. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 10982.
- Bruice, T. W.; Lima, W. F. *Biochemistry* **1997**, *36*, 5004.
- Chaltin, P.; Borgions, F.; Zhenyu, Z.; Rozenski, J.; Van Aerschot, A.; Herdewijn, P. *Bioconj. Chem.*: Submitted.
- Bailly, C.; Hénichart, J. P. *Bioconjugate Chem.* **1991**, *2*, 6.
- Bailly, C.; Pommery, N.; Houssin, R.; Hénichart, J. P. *J. Pharm. Sci.* **1989**, *78*, 910.
- Bailly, C.; Helbecque, N.; Hénichart, J. P.; Colson, P.; Houssier, C.; Rao, K. E.; Shea, R. G.; Lown, J. W. *J. Mol. Recognit.* **1990**, *3*, 26.
- Eliadis, A.; Phillips, D. R.; Reiss, J. A.; Skorobogaty, A. *J. Chem. Soc. Chem. Commun* **1988**, 1049.
- Denny, W. A. *Anti-Cancer Drug Res.* **1990**, *4*, 241.
- Baguley, B. C. *Anti-Cancer Drug Res.* **1991**, *6*, 1.
- Wilson, W. D.; Wang, Y.; Kusuma, S.; Chandrasekaran, S.; Yang, N. C.; Boykin, D. W. *J. Am. Chem. Soc.* **1985**, *107*, 4989.
- Bailly, C.; Pommery, N.; Houssin, R.; Hénichart, J. P. *J. Pharm. Sci.* **1989**, *78*, 910.
- Bailly, C.; Helbecque, N.; Hénichart, J. P.; Colson, P.; Houssier, C.; Rao, K. E.; Shea, R. G.; Lown, J. W. *J. Mol. Recognit.* **1990**, *3*, 26.
- Bailly, C.; Denny, W. A.; Mellor, L. E.; Wakelin, L. P. G.; Waring, M. J. *Biochemistry* **1992**, *31*, 13.
- Augustyns, K.; Vandendriessche, F.; Van Aerschot, A.; Busson, R.; Urbanke, C.; Herdewijn, P. *Nucleic Acids Res.* **1992**, *20*, 4711.