Pergamon

Bioorganic & Medicinal Chemistry Letters 12 (2002) 955–958

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Construction of the Novel Conformationally-Restricted Peptide Library for Screening of Peptides that Control the Interaction Between Nucleobases

Mizuki Takahashi,[a] Akihiko Ueno[a] and Hisakazu Mihara[a,b,*]

[a]*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan*
[b]*Form and Function, PRESTO, Japan Science and Technology Corporation, Yokohama 226-8501, Japan*

Received 12 November 2001; accepted 19 January 2002

Abstract—A unique conformationally-restricted peptide library was constructed using a loop structure as a structural scaffold. This library was used for the screening of the amino acid sequences that control the interaction between nucleobase triplets. The peptides have PNAs at the C-terminus as the recognition site and the random amino acid sequence at the N-terminus as the effector for the interaction between PNA and its complementary DNA triplets. From the peptide libraries constructed by the positional scanning method, the sequences that affect the interaction between PNA and complementary DNA were selected. The difference in the characteristic results by using A–T and G–C pairs was presented. This study would also give us some useful information about interaction between peptides and nucleic acids, such as relevancies between these biomolecules in a prebiotic era.

The combinatorial peptide library has been widely used as a useful tool for screening of the bioactive ligands and/or recognition motifs against specific molecules.[1] The random peptide libraries, however, generally have considerable conformational flexibility; thus they do not always meet the consensus sequence nor produce an expected affinity compound. According to these aspects, kinds of conformationally restricted peptide libraries have been reported recently.[2–4] Examples include the strategies for presenting random sequences in a conformationally restricted manner, grafting randomized positions onto a protein scaffold,[2] cyclizing via flanking cysteins,[3] or using α-helical peptides as a structural scaffold by de novo peptide design.[4] In these libraries, however, screening methods still depend on the traditional method based on the intermolecular interaction between library peptides and target molecules. Thus, it is difficult to find a weak binder or effector-like molecule which controls the interaction between other molecules.

Here, we propose a unique conformationally-restricted peptide library including the intramolecular interaction between a library peptide and a target molecule by using a loop structure as a structural scaffold and peptide nucleic acids (PNAs) as a recognition site. So far, there have been a lot of studies on the folding of loop peptides derived from natural proteins as well as β-hairpins designed by the de novo method.[5] Hence, a well-designed loop peptide sequence is available to determine the structure to gather the two parts in both termini to interact in close proximity. On the other hand, PNA is a DNA mimic with the nucleobases on a pseudopeptide backbone composed of *N*-(2-aminoethyl)glycine units.[6] It has been revealed that a PNA molecule has the ability to bind DNA with high efficiency and sequence-specificity. Furthermore, it can be easily conjugated with the peptide structure since PNA has a pseudopeptide backbone. Therefore, PNA is a useful tool to introduce the recognition site against nucleic acids in the peptide structure. In our peptide design, 3mer PNAs were conjugated with the C-terminus of a peptide structure to introduce a defined recognition site to its complementary DNA. The library sequence consisting of 3mer random amino acids was placed at the N-terminus of peptide connected via a loop sequence so as to be in close to the PNA–DNA complex (Fig. 1). By using this structure, we carried out the effective screening of short amino acid sequences that control the interaction between nucleobase triplets with relatively weak interaction. Such a weak interaction may not be elucidated

*Corresponding author. Tel.: +81-45-924-5756; fax: +81-45-924-5833; e-mail: hmihara@bio.titech.ac.jp

0960-894X/02/$ - see front matter
PII: S0960-894X(02)00065-3

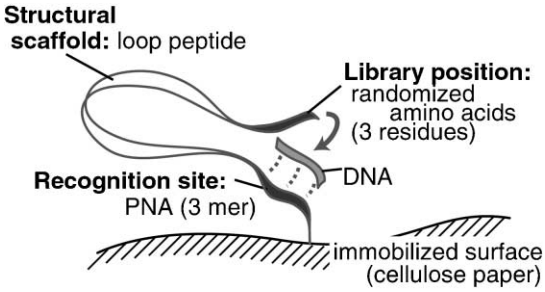

Figure 1. Schematic representation of this system.

when the library peptides, PNA and DNA were mixed separately.

Designed libraries were synthesized using the spot-synthesis method[7,8] as an individual spot on a cellulose membrane. To detect the binding of DNA that has a complementary sequence to PNA, we used a biotinylated DNA and peroxidase-labeled streptavidin. Amount of bound DNA was estimated by the depth of color with 3,3′,5,5′-tetramethylbenzidine as a substrate of peroxidase.[9]

Initially, we determined the sequence of the loop peptide as a scaffold of the β-hairpin structure using $(A_{PNA})_3$ as the C-terminus PNA (Fig. 2). To optimize the assay condition, we compared the binding affinity of biotinylated

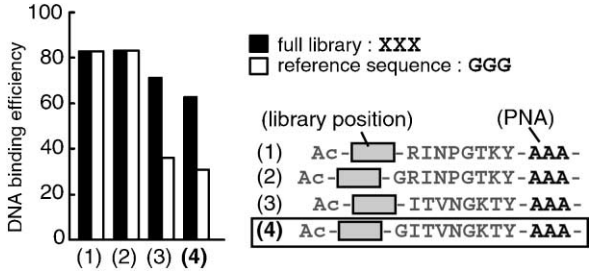

Figure 2. Determination of the loop sequences by the comparison between the full library (XXX) and the reference sequence (GGG). The peptide sequences were described by the usual one-letter code of amino acids except for the boldfaced **A** which represents the adenine PNA. The rectangular in the sequences (1)–(4) denote XXX or GGG. X represents an equal distribution of 19 amino acids (20 natural amino acids except for Cys).

$(dT)_3$ to the full library, XXX, with that to the reference sequence, GGG, where X represents an equal distribution of 19 amino acids (20 natural amino acids except for Cys). The results under the optimized conditions[10] indicated that the loop peptide sequences designed based on an antibody loop[11] showed a non-specific binding to DNA, possibly due to the existence of Arg residue in the loop sequence [Fig. 2(1) and (2)]. In contrast, the sequences based on the peptide designed to form a stable β-hairpin structure[5a] showed a good contrast between XXX and GGG [Fig. 2(3) and (4)]. Furthermore, the molecular modeling study suggested that the 9mer sequence including the Gly residue as a linker was more suitable to interact with PNA–DNA region. Therefore, we selected this 9mer sequence (4) as a scaffold of the loop structure.

By using the sequence selected as above for the loop structure, several sets of peptide libraries were constructed by the positional scanning method with $(A_{PNA})_3$ sequence at the C-terminus. At first, the library OXX, which has 18 defined amino acid (Cys and Trp were omitted because of the synthetic problems) at the O position, was synthesized. In this series, a stronger binding of biotinylated $(dT)_3$ was observed for positively charged amino acids such as Arg, Lys, His, as well as large hydrophobic amino acids such as Phe, Tyr at the O site [Fig. 3(a)]. However, there was no significant difference in the reference assay using biotinylated $(dC)_3$ for positively charged amino acids and Tyr. These results suggested that the peptides with these amino acids have a binding affinity non-specific to DNA. In

Table 1. A part of results of the positional scanning libraries with adenine PNA

Sequence	$(dT)_3$	$(dC)_3$	Sequence	$(dT)_3$	$(dC)_3$
FXX	54.9	49.0	(PNA/**TAA**)		
FGX	61.7	38.4	FGP	49.0	24.7
FGP	60.0	12.9	HGP	49.0	7.8
	25.9 (with $(dA)_3$)		IGP	37.6	3.9
FXP	25.5	19.2			

X represents an equal distribution of 19 amino acids. The values are the binding efficiency of each DNA represented as a modal value of gray% of each oval selection. [biotinylated $(dT)_3$ or $(dC)_3$] = 0.2 μM, non-labeled $(dA)_3$ = 1.0 μM.

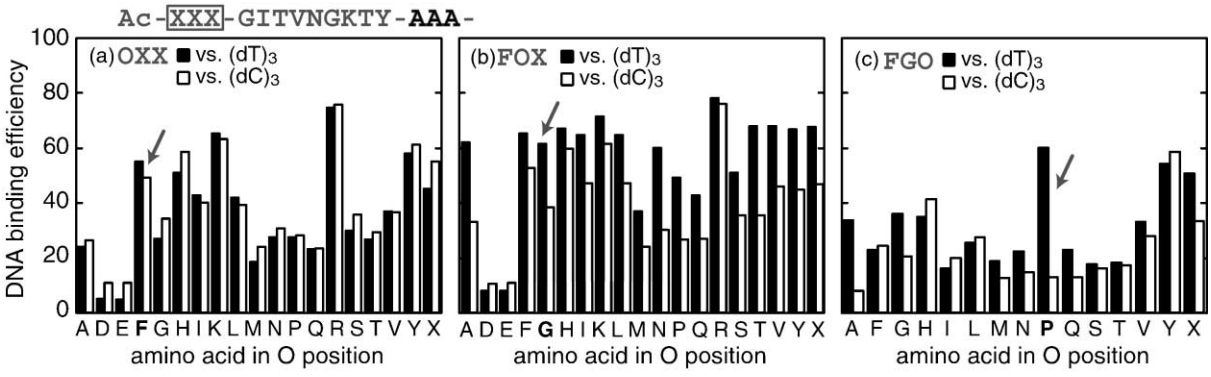

Figure 3. Results of the positional scanning libraries using adenine PNA: (a) first generation: OXX; (b) second generation: FOX; (c) third generation: FGO. O represents the defined amino acids (horizontal axis). Black bar and white bar represent the results with complementary target DNA [biotinylated $(dT)_3$] and non-complementary target DNA [biotinylated $(dC)_3$], respectively.

contrast, the peptide which has Phe at the O position (FXX) showed slightly weaker binding to biotinylated $(dC)_3$ (Table 1), indicating that this peptide has some specificity for $(dT)_3$, complementarily interacting to $(A_{PNA})_3$ in the peptide. On the other hand, no characteristic result was observed from the library XOX and XXO (data not shown). According to these results, the second library which has FOX at the library sequence was synthesized. In this O position, the introduction of amino acids with a relatively small side chain such as Ala, Gly, Asn, and Thr showed a good contrast in the results between biotinylated $(dT)_3$ and $(dC)_3$ as a ligand [Fig. 3b and Table 1]. To the contrary, positively charged residues, Arg and Lys, again caused a non-specific binding to $(dT)_3$ and $(dC)_3$. In this study, Gly was selected at this position to the library in the next generation.

From the third library, FGO, the FGP sequence was selected as the strongest binder of $(dT)_3$.[12] In contrast, this peptide showed a rather lower ability in the binding to $(dC)_3$ [Fig. 3c and Table 1]. Furthermore, the binding of biotinylated $(dT)_3$ to the FGP peptide was significantly inhibited by the coexistence of non-labeled $(dA)_3$ [5 equiv to biotinylated $(dT)_3$ (0.2 μM)], while no significant inhibition was observed for the peptides that have other amino acids at the O position (Table 1). These results suggest that the selected FGP sequence has a higher specificity to enhance the interaction of A_{PNA}–T_{DNA} base pair. In addition, we also synthesized another library, FXO, which has a defined amino acid at the third position and X at the second position. In this library, FXP sequence also gave some contrast between the binding of $(dT)_3$ and $(dC)_3$ (Table 1), indicating that the Pro residue at the third position is important to the selective binding of the complementary DNA. However, the contrast was smaller than the case of FGP, possibly due to the existence of the peptides which have larger amino acids at the X position. This result suggests that the Gly residue at the second position is also important to the effective interaction of complementary nucleobases. According to above results, it was possibly suggested that the Pro and Gly residues may contribute to structural suitability, while the aromatic ring in the side chain of Phe residue may interact to A_{PNA}–T_{DNA} base pair directly.

On the basis of the above results, we synthesized another library with T_{PNA}–A_{PNA}–A_{PNA} sequence at the C-terminus and OGP sequence at the library position. In the comparison between the binding of $(dT)_3$ and $(dC)_3$, the selective binding to $(dT)_3$ was observed with His and Ile residues at the O position as well as Phe (Table 1). In this case, the selected sequences still have some specificity to $(dT)_3$ despite a single base mismatch, and some different amino acids were selected as compared to $(A_{PNA})_3$.

We then constructed some sets of libraries with $(G_{PNA})_3$ sequence at the C-terminus.[13] In the case of the libraries with G_{PNA}, WXX, WRX, and WRR sequence at the library position were continuously selected as selective binders to $(dC)_3$ rather than $(dT)_3$ in the libraries of the first, second, and third generation, respectively (Fig. 4). In this case, however, the differences between the binding affinity to $(dC)_3$ and $(dT)_3$ are small even in the library of the third generation. In addition, we also synthesized other libraries, YOX, ROX, and RRO, but no selective result was observed because of too strong non-specific affinity. These results seem to be related to the known property of Arg and/or Tyr, which has a strong affinity to nucleic acids especially to the guanine base.[14] Therefore, it was suggested that relatively small differences between the binding of complementary and non-complementary DNA were caused by the too strong affinity of selected arginine-rich sequence to DNA. However, the differences in the binding of $(dC)_3$ and $(dT)_3$ indicated that the selected sequence may interact stronger to G_{PNA}–C_{DNA} base pair than added DNA alone.

In conclusion, we have demonstrated a unique conformationally-restricted peptide library using a loop structure as a structural scaffold. This library was used for the screening of the amino acid sequences that control the interaction between A_{PNA}–T_{DNA} and G_{PNA}–C_{DNA} triplets. In the results using A_{PNA}, it was suggested that the selected amino acid sequences did not simply bind to a DNA or PNA molecule, but enhanced the interaction of A_{PNA}–T_{DNA} base pairs. On the other hand, from the libraries using G_{PNA}, a very strong binder to G_{PNA}–C_{DNA} was obtained, although the

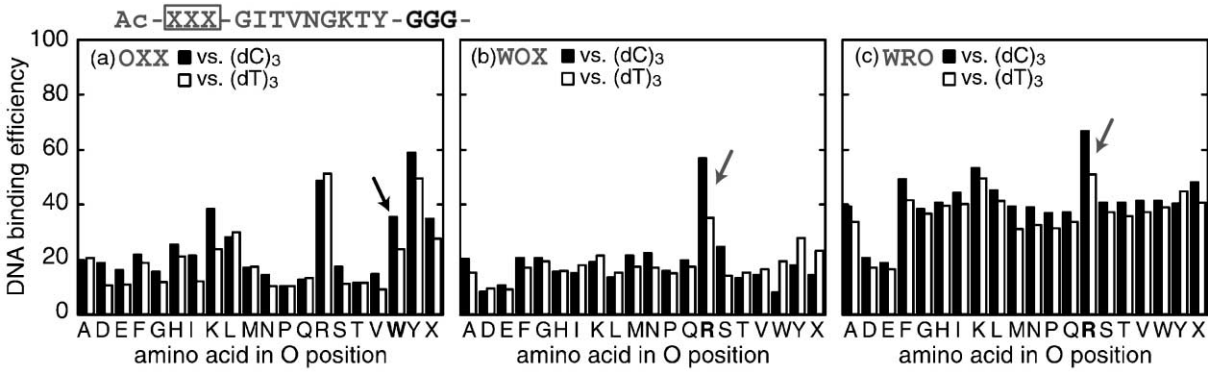

Figure 4. Results of the positional scanning libraries using guanine PNA: (a) first generation: OXX; (b) second generation: WOX; (c) third generation: WRO. Black bar and white bar represent the results with complementary target DNA [biotinylated $(dC)_3$] and non-complementary target DNA [biotinylated $(dT)_3$], respectively.

selectivity to the complementary DNA is rather low. Furthermore, this PNA–peptide conjugate system for the screening of peptides that interact specific nucleobases would give us some information about the interaction between amino acids and nucleobases, such as relevancies between these biomolecules in a prebiotic era. For example, it is known that the U bases in codon (A in anticodon) often represent the relatively hydrophobic amino acids.[15] Especially, UUU codon (AAA anticodon) is corresponded to Phe, which is a selected amino acid as the binder of $(A_{PNA})_3$–$(dT)_3$ by this study. In addition, Ile residue selected by the library using $(TAA)_{PNA}$ sequence is corresponded to TAA anticodon. On the other hand, however, the libraries with $(G_{PNA})_3$ did not give Gly or Pro as the selected amino acids, which represent GGG or CCC codon, respectively. Probably, such amino acids, which have no characteristic side chain, are hard to interact to guanine or cytosine base directly. Therefore, it was possibly suggested that the amino acids having affinities to DNA such as the arginine-rich sequence was selected from the libraries with G_{PNA}, and Gly or Pro residue was not selected as a correspondent of GGG or CCC codon directly. Throughout this study, it was indicated that the library system used here is useful not only for obtaining the binder to a specific base pair, but also for suggesting the unique feature of nucleobases and/or base pairs in the relevancies to amino acids, such as differences between A–T and G–C pair. Furthermore, this system using a loop scaffold can be utilized in the screenings of various molecules such as drugs affording relatively weak interaction.

References and Notes

1. (a) Lam, K. S. *Chem. Rev.* **1997**, *97*, 411. (b) Roberts, R. W.; Ja, W. W. *Curr. Opin. Struct. Biol.* **1999**, *9*, 521.
2. (a) Nygrenn, P. Å; Uhlén, M. *Curr. Opin. Struct. Biol.* **1997**, *7*, 463. (b) McConnell, S. J.; Hoess, R. H. *J. Mol. Biol.* **1995**, *250*, 460. (c) Smith, G. P.; Patel, S. U.; Windass, J. D.; Thornton, J. M.; Winter, G.; Griffiths, A. D. *J. Mol. Biol.* **1998**, *277*, 317.
3. (a) Eichler, J.; Lucka, A. W.; Houghten, R. A. *Pept. Res.* **1994**, *7*, 300. (b) McBride, J. D.; Freeman, H. N. M.; Leatherbarrow, R. J. *Eur. J. Biochem.* **1999**, *266*, 403.
4. (a) Pérez-Payá, E.; Houghten, R. A.; Blondelle, S. E. *J. Biol. Chem.* **1996**, *271*, 4120. (b) Fujii, I.; Takaoka, Y.; Suzuki, K.; Tanaka, T. *Tetrahedron Lett.* **2001**, *42*, 3323.
5. (a) Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nat. Struct. Biol.* **1996**, *3*, 604. (b) Ramírez-Alvarado, M.; Kortemme, T.; Blanco, F. J.; Serrano, L. *Bioorg. Med. Chem.* **1999**, *7*, 93. (c) Cochran, A. G.; Tong, R. T.; Starovasnik, M. A.; Park, E. J.; McDowell, R. S.; Theaker, J. E.; Skelton, N. J. *J. Am. Chem. Soc.* **2001**, *123*, 625.
6. Nielsen, P. E.; Haaima, G. *Chem. Soc. Rev.* **1997**, 73.
7. (a) Frank, R. *Tetrahedron* **1992**, *48*, 9217. (b) Kramer, A.; Schneider-Mergener, J. *Meth. Mol. Biol.* **1997**, *87*, 25. (c) Wenschuh, H.; Volkmer-Engert, R.; Schmidt, M.; Schulz, M.; Schneider-Mergener, J.; Reineke, U. *Biopolym. Pept. Sci. Sect.* **2000**, *55*, 188.
8. The spot-synthesis method using this work was generally followed the method described in ref 7 except for the peptide loading level. For the libraries with A_{PNA}, the loading level was ca. 20 nmol/spot.
9. The paper colored by the addition of H_2O_2 was taken by the photograph using a digital camera, and then the image was processed by using ImageJ (NIH; http://rsb.info.nih.gov/ij). The DNA binding efficiency was represented as a modal value of gray % of each oval selection.
10. The optimized assay conditions were as follows; washing buffer: 50 mM Tris–HCl (pH 7.0) containing 137 mM NaCl, 2.7 mM KCl, and 0.1% Tween20, biotinylated DNA as a first ligand: 0.2 μM for 30 min, HRP-SA as a second ligand: 1.6 μg/mL for 30 min, reaction buffer: 100 mM AcOH/AcONa (pH 5.5) containing 0.01% Tween20, 3,3′,5,5′-tetramethylbenzidine and H_2O_2 as a substrate: 0.5 and 0.8 mM, respectively, reaction time: 5 min (first and second generation libraries with A_{PNA}), 0.5 min (third library with A_{PNA}), 3 min (the libraries with G_{PNA}).
11. Takahashi, M.; Ueno, A.; Mihara, H. *Chem. Eur. J.* **2000**, *6*, 3196.
12. In the third generation library, negatively charged residues (Asp and Glu) and positively charged residues (Arg and Lys) were omitted, because the very weak binding or highly non-specific binding was easily expectable.
13. In this case, the peptide loading level was reduced to ca. 2 nmol because of high background signals due to the strong affinity of G_{PNA} to DNA. The synthetic problems with Trp residue could be cleared by using Trp(Boc).
14. Michael, Y. *J. Mol. Evol.* **1998**, *47*, 109.
15. D'Onofrio, G.; Jabbari, K.; Musto, H.; Bernardi, G. *Gene* **1999**, *238*, 3.