doi:10.2533/chimia.2013.905

Chimia 67 (2013) 905–909 © Schweizerische Chemische Gesellschaft

Nucleic Acid-tagged Peptides: Encoding Libraries and Controlling Dimerization and Conformation

Kalyan K. Sadhu and Nicolas Winssinger*

Abstract: Recent progress in the synthesis, screening, and self-assembly of nucleic acid-tagged peptidic libraries is reviewed.

Keywords: Combinatorial peptide synthesis · DNA · Encoding · Multivalency · PNA · Self-assembly

Introduction

There is a long-standing interest in the discovery of peptidic ligands and inhibitors for biomedical research. Phage display^[1] offers an elegant solution for interrogating a large library of peptides (typically 10^{6} – 10^{9}) and has enabled the discovery of a number of peptidic ligands.^[2] Chemical modifications through chemoselective derivatization of cysteines has further broadened the scope of this powerful technology.^[3,4] Alternatively, ribosomal display technologies have also been used to interrogate large libraries of peptides $(>10^{12})$.^[5,6] The ability to include *t*RNA loaded with modified amino acids has also broadened this technology beyond the 20 canonical amino acids.^[7,8] Technologies based on cell-free translation of RNA into peptidic libraries have also yielded numerous ligands.^[9] In parallel, the development of combinatorial peptidic libraries synthesized by mix and split techniques^[10,11] have facilitated access to modified peptides or peptoids but analysis of the fittest ligand from a large synthetic library is comparatively complicated by the fact that it is not genetically encoded. Over the past decade, several technologies that facilitate the generation of nucleic acid-tagged peptidic libraries by chemical synthesis or templated synthesis have been reported. This review summarizes these recent developments and discusses additional use of nucleic acid tags for dimerizing (or oligomerizing) ligands to capitalize on the synergy of their interactions with a target or to control their conformation.

Synthesis of Nucleic Acid-tagged Libraries

Four different approaches have been described to chemically synthesize nucleic acid-encoded peptide or peptoid libraries, with the first libraries reported in 2004 (Fig. 1). The method developed by the group of D. Liu utilizes DNA-tagged reagents that are engaged in a reaction with appropriate DNA-tagged substrates based on their complementarity (Fig. 1A).^[12-14] This DNA-templated synthesis (DTS) allows a library to be synthesized in a single reaction vessel since the reagents sort themselves on the appropriate template based on their complementarity. This technology effectively enables researchers to translate a DNA sequence into an artificial oligomer. The method developed by our group utilizes peptide nucleic acids (PNAs)^[15] as the encoding element.^[16,17] PNA synthesis is essentially peptide chemistry that bypasses the requirement of an oxidation step present in DNA synthesis as well as the liability of depurination under acidic conditions typically employed in peptide synthesis. Thus, PNA encoding can be performed by chemical synthesis in parallel with library synthesis (Fig. 1B). Several protecting group strategies have been reported that enable co-synthesis of PNA with standard Fmoc-based strategies for the peptide, making PNA-encoded synthesis (PES)

compatible with standard solid-phase synthesis techniques.[18,19] The third strategy was reported by the group of Harbury and utilizes immobilized anti-codons to isolate and route the different members within a library of ssDNA (Fig. 1C).[20-22] Once separated, each strand is chemically derivatized using DNA-compatible chemistry. As for DTS, this technique effectively enables the translation of a DNA sequence into an artificial oligomer. The fourth strategy was pioneered by the group of Neri (Fig. 1D) and makes use of stepwise enzymatic DNA-tagging.^[23,24] With this method, pools of different oligonucleotides that contain an encoding segment flanked by primer-binding segments are chemically derivatized, mixed, and split (and deprotected if necessary) to introduce the second synthon of diversity. The encoding DNA for this second synthon of diversity is introduced by hybridization to the primer region of the oligonucleotide encoding the first step, followed by Klenow-assisted DNA polymerase.

While all four strategies have proven to be robust and enable the synthesis of peptide-based libraries, there are important technical differences in the chemistry that can be performed and in the nature of the tag. In the first case, DTS has been shown to be exceptionally versatile in the type of chemistry that can be performed.^[13] A 13,000-member library of peptidic macrocycles was prepared using three sequential DNA-templated peptide coupling (NHS activation of the acids) followed by stabilized Wittig macrocyclization.^[14] The synthesis was performed using a library of templates bearing primers on both ends with three 10-nucleobase sections used for each step. This library was used to identify an Src kinase inhibitor (Fig. 2).^[14,25] For PES, codons of 3 to 4 nucleobases are used at each encoding step, affording a PNA of 12–16mer. Several PNA-encoded peptidic

^{*}Correspondence: Prof. N. Winssinger Department of Organic Chemistry University of Geneva 30 quai Ernest Ansermet CH-1211 Geneva Tel.: +41 22 379 61 05 E-mail: nicolas.winssinger@unige.ch



Fig. 1. Synthesis of nucleic acid-tagged peptidic libraries. The methods are also applicable to other biologically relevant libraries (heterocycles, glycan).

libraries of up to 10,000 members have been reported, yielding protease inhibitors,^[26–28] GPCR ligands, and cell-specific targeting peptides.^[29–31] For DNA-routing, 20-nucleobase coding regions are used for each synthon of diversity flanked by 20-nucleobase constant regions.^[20–22] A peptoid library involving reiterative coupling of DMT-activated chloroacetic acid followed by nucleophilic displacement using primary amines was prepared. Remarkably, this procedure was found to be sufficiently robust to perform up to six iterations, thus making use of a 340-nucleotide DNA tag. This library was used to identify a ligand for the Crk-SH3 binding site (Fig. 2) out of 10⁸ peptoids.^[32] For Klenow-assisted DNA polymerase, a 6-nucelobase coding region was used with a 18-mer constant region serving a primer binding sites. This method was used to assemble two different 4,000 to 30,000-member peptidic libraries using an amide coupling to join the two elements of diversity.^[23,24] Other libraries based on cycloaddition chemistry have also been reported using this technology.^[33] A conceptually related strategy was used to encode a 10⁸ member heterocyclic library.^[34]

Screening and Decoding

Screening of nucleic acid-encoded libraries is generally performed by selecting the fittest member and then decoding its sequence (Fig. 3). Using a PNA-encoded library of mechanism-based inhibitors targeting cysteine proteases, it was shown



Fig. 2. Selected examples of peptidic inhibitors discovered from nucleic acid-encoded libraries.





Fig. 4. Substrate profiling for proteases and kinases.

that size-exclusion filtration (30 KDa spin filter) could be used to select for the best binders (out of a 4,000-member library) that remained covalently linked to their target.[26] Alternatively, gel electrophoresis has also been used to separate protease-inhibitor adducts from the rest of the library.^[27] Hybridization of the selected fraction onto a microarray revealed the identity of the best binders. This selection was possible because of the significant difference in molecular weight between the unbound PNA-inhibitors (4-6 KDa) and the adduct with a target protein (>30 KDa). A fluorophore was appended to the end of the PNA for detection on the microarray. This microarray-based decoding provides a rapid and practical decoding solution. However, because of the limited oligonucleotide length in most microarray formats, this decoding lends itself more readily to PNA-encoded libraries with shorter oligonucleotide encoding schemes. Commercial custom microarrays continue to improve the length and density of DNA

and are currently available with up to 1 million sequences and oligonucleotides of 60 nucleobases. However, parallel development of next-generation sequencing (discussed below) is providing a compelling alternative for large-scale decoding. The microarray decoding was also used to profile the preferred substrates for proteases and kinases.^[17,19,35,36] Ouantification of enzymatic activity from complex proteomes including crude cell lysates is one of the major targets of current research.^[37] For proteases, PNA-encoded libraries of substrate (either fluorogenic substrates^[35] or substrates flanked by a fluorophore and a quencher,^[19] Fig. 4) were treated with a protease, and the fittest substrate was identified following hybridization. For kinase substrates, the phosphorylated peptides were identified using a specific anti-phosphotyrosine antibody.^[17,36]

Aside from the aforementioned sizeexclusion separation, selections are more generally performed by immobilization of the target onto a matrix, followed by affinity purification of the best binder. For DNAencoded libraries, PCR amplification and sequencing of the tag afford the identity of the selected compounds.^[12,32,38,39] The power of next-generation sequencing to provide a quantitative readout (current technology provides a sequence output of >10⁷ sequence reads) was rapidly recognized.^[23,34]

While PNA tags cannot be amplified by PCR or directly sequenced, a complementary DNA strand can be used for this purpose, thus providing a way around this limitation. This has been implemented in two different formats. In the first, a DNA template is used to pair two sets of PNAencoded fragment (Fig. 5). Following selection, the encoding DNA can be amplified by PCR to decode the optimal fragment combination. It was further shown that the amplified DNA template recovered from a first selection could be used to reassemble the library of selected fragments (unselected fragments do not hybridize and are washed away). The selection and DNA amplification/reassembly can be reiterated to refine an optimal fragment combination. The DNA-encoded approach was used to identify ligands binding to carbonic anhydrase^[40] and DC-SIGN^[41] (see discussion below on DNA display). The second format using a complementary DNA sequence to amplify the result of PNA-encoded selection was applied to identify cell surface receptor ligands. Following affinity selection of the PNA-encoded peptide library on whole cells over-expressing a cell surface



Fig. 5. DNA display of paired ligands. Following selection, the DNA-templating strand can be amplified by PCR and sequenced or used to reassemble the library of selected fragment pairs.

and orientations between the two peptides

revealed an assembly (termed synbody)



receptor (integrins or CCR6), the recovered PNA tags were exposed to a library of complementary DNA and the mixture was subjected to S1 nuclease to degrade all ssDNA corresponding to unselected compounds.^[30] A subsequent PCR amplification revealed the sequence of DNA complementary to the selected PNA.

DNA Display: Hybridization to Control Ligand Pairing, Distance, and Valency

The programmability of oligonucleotide hybridization and the ensuing selforganization have also been harnessed to display nucleic acid-tagged ligands. In addition to the combinatorial pairing discussed above, this strategy can be used to control spacing between ligands, the order of ligands and linker flexibility as well as geometry (Fig. 6). Pioneering examples of this concept were reported by Kobayashi and coworkers using two half-sliding complementary DNA fragments derivatized with galactose to construct a glyco-cluster (Fig. 7A);^[42] Neri and coworkers used the self-assembly to pair drug fragments with complementary tags (Fig. 7B).^[38] Due to the fact that double-stranded nucleic acids are fairly rigid over short distances (<60 bp), hybridization can be used to control the inter-ligand distance. With the aim of emulating the epitope of HIV, which is made up of multiple branched oligosaccharides terminating with mannose, a library of PNA-encoded oligomannose was prepared using different peptide linkers between mannose oligosaccharides in combination with different hybridization distances (Fig. 7C).^[43] Evaluation of the affinity by SPR revealed a clear distance affinity correlation. Interestingly, a lysine branch mimicked the 3,6-manose branch on the natural epitope. In parallel work, Chaput and coworkers used hybridization to pair two DNA-tagged 12mer peptides^[44] (coupled to amine-modified DNA using standard amine coupling chemistry), each having moderate affinity (µM) for a selected protein, Gal89, a yeast regulatory protein (Fig. 7E). Screening 24 different assemblies with defined distances



Fig 7. Cartoon representation of DNA display for selected examples.

Fig. 6. DNA display of ligands with control over inter-ligand distance, ligand valency, linker rigidity, and geometry of presentation (relative angle between the ligands).

with 5 nM affinity which corresponds to ~1000-fold improvement in affinity over the individual peptides, comparable to the affinity of conventional antibodies. The same strategy was used to discover a high-affinity capture reagent for Grb2, a growth factor receptor.[45] Receptor dimerization or oligomerization is an important mechanism for signal transduction across membranes. In an effort to find the optimal geometry and distance to dimerize a macrocyclic peptide ligand of DR5 (Fig. 7D), a member of the TRAIL family, an assembly was discovered with a 20-fold longer dissociation half-life.^[46] In a further refinement of DNA display, Seitz and coworkers used DNA template to control the distance between the ligand, and compared the results with template having uncomplemented segments to introduce flexibility. This proved important for bridging binding sites that are on geometrically opposed faces of a protein.^[47] A DNA-tagged hexapeptide incorporating a phosphotyrosine was used to interrogate the optimal distance in a tandem SH2 domain interaction with Syk kinase (Fig. 7F). Unpaired spacer nucleotides between double-stranded segments in assemblies were found to relieve the torsion constraints imposed by dsDNA.[48]

More recently, Appella and coworkers demonstrated that DNA-displayed ligands can be functional *in vivo*.^[49] In this case, DNA-templated assemblies of cyclic RGD motif tagged with PNA were used to antagonize integrins, which mediate cell attachment through multivalent binding. The impact of multivalency was studied across a wide range of ligand numbers (up to 45), densities, and three-dimensional arrangements. The optimal assembly improved inhibitory activity by 100-fold and, when applied *in vivo*, reduced tumor colonies by 50%.

Aside from controlling oligomer order and geometry, hybridization of nucleic acid-peptide conjugates has also been used to control the conformation of peptides. In pioneering work, Seitz and coworkers used an SH2-binding peptide sequence that was flanked by PNA tags (Fig. 8A). Hybridization with different alignments of the PNA tags should have a strong impact on the conformation of the peptide.[50] Indeed, it was found that the affinity of the assemblies for the SH2 domain varied by 10-fold depending on the number of abasic sites (0-10 nucleotides tested) present between the hybridization sites. In an elegant extension of this concept, Ly and coworkers used complementary nucleobase hybridization within a peptidic macrocycle to recapitulate the conformational bias imposed by a series of disulfide brides.^[51] There are numerous examples of biologi-





Fig. 8. Schematic representation of hybridization-controlled conformation. (A) SH2 domain ligand and (B) the θ -defensin RTD-1 (in RTD-1M, the disulfide bonds of the cyclic RTD-1 peptides are replaced by C-G base pairs).

cally active macrocyclic peptides containing disulfide bonds that restrict the conformational flexibility of the macrocycle. As a proof of principle, a member of the θ -defensin family of antibiotic peptides was modified with three complementary PNA monomers in lieu of the disulfides (Fig. 8B). This supramolecularly constrained macrocycle showed comparable biological activities to the natural product (Fig. 8B). Moreover, it was efficient for killing both Gram-positive and Gramnegative bacterial strains.

Summary and Outlook

Impressive developments have been reported over the past 10 years in the synthesis and application of nucleic acid-tagged peptides and their derivatives. From a synthetic perspective, the repertoire of technologies available to conjugate or synthesize biologically relevant and functional molecules has significantly increased. Several landmark studies have demonstrated the selection of synthetic ligands from nucleic acid-encoded libraries. Beyond encoding, a series of studies have established that nucleic acid hybridization can be harnessed to control the geometry (distance and directionality) of multiple ligands and capitalize on the synergy of their binding to a macromolecule. Likewise, hybridization has also been exploited to constrain the conformation of peptides. It is well recognized that controlling the conformation of peptides is important not only to improve their affinity to a target (by reducing the entropic cost of binding) and their selectivity, but also to reduce their susceptibility to proteases. It can be anticipated that libraries combining the different assets of nucleic acid tags will continue to improve our ability to identify ligands for biomedical research.

Received: September 18, 2013

- [1] G. P. Smith, Science 1985, 228, 1315.
- [2] G. P. Smith, V. A. Petrenko, *Chem. Rev.* 1997, 97, 391.
- [3] C. Heinis, T. Rutherford, S. Freund, G. Winter, *Nat. Chem. Biol.* 2009, 5, 502.
- [4] I. R. Rebollo, C. Heinis, *Methods* 2013, 60, 46.
 [5] L. C. Mattheakis, R. R. Bhatt, W. J. Dower, *Proc. Natl. Acad. Sci. USA* 1994, 91, 9022.
- [6] R. W. Roberts, J. W. Szostak, Proc. Natl. Acad. Sci. USA 1997, 94, 12297.
- [7] C. J. Noren, S. J. Anthonycahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, 244, 182.
- [8] T. Kawakami, H. Murakami, H. Suga, Chem. Biol. 2008, 15, 32.
- [9] A. Wada, Frontiers Immun. 2013, 4, 224.
- [10] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, Int. J. Pept. Protein Res. 1991, 37, 487.
- [11] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, 354, 82.
- [12] Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* 2004, 305, 1601.
- [13] X. Li, D. R. Liu, Angew. Chem. Int. Ed. 2004, 43, 4848.
- [14] R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai, D. R. Liu, J. Am. Chem. Soc. 2010, 132, 11779.
- [15] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, P. E. Nielsen, *Nature* **1993**, *365*, 566.
- [16] F. Debaene, L. Mejias, J. L. Harris, N. Winssinger, *Tetrahedron* 2004, 60, 8677.
- [17] J. L. Harris, N. Winssinger, Chem. Eur. J. 2005, 11, 6792.
- [18] D. Chouikhi, M. Ciobanu, C. Zambaldo, V. Duplan, S. Barluenga, N. Winssinger, *Chem. Eur. J.* 2012, 18, 12698.
- [19] J. J. Diaz-Mochon, L. Biały, M. Bradley, Chem. Commun. 2006, 3984.
- [20] D. R. Halpin, P. B. Harbury, *PLoS Biol.* **2004**, *2*, E174.
- [21] D. R. Halpin, P. B. Harbury, *PLoS Biol.* **2004**, *2*, E173.
- [22] D. R. Halpin, J. A. Lee, S. J. Wrenn, P. B. Harbury, *PLoS Biol.* 2004, 2, 1031.
- [23] L. Mannocci, Y. X. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, *Proc. Natl. Acad. Sci. USA* 2008, 105, 17670.
- [24] M. Leimbacher, Y. X. Zhang, L. Mannocci, M. Stravs, T. Geppert, J. Scheuermann, G. Schneider, D. Neri, *Chem. Eur. J.* 2012, 18, 7729.
- [25] G. Georghiou, R. E. Kleiner, M. Pulkoski-Gross, D. R. Liu, M. A. Seeliger, *Nat. Chem. Biol.* 2012, 8, 366.
- [26] J. Harris, D. E. Mason, J. Li, K. W. Burdick, B. J. Backes, T. Chen, A. Shipway, G. Van Heeke, L. Gough, A. Ghaemmaghami, F. Shakib, F. Debaene, N. Winssinger, *Chem. Biol.* 2004, 11, 1361.

- [27] H. D. Urbina, F. Debaene, B. Jost, C. Bole-Feysot, D. E. Mason, P. Kuzmic, J. L. Harris, N. Winssinger, *ChemBioChem* **2006**, *7*, 1790.
- [28] F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran, N. Winssinger, *Tetrahedron* 2007, 63, 6577.
- [29] N. Svensen, J. J. Diaz-Mochon, M. Bradley, *Chem. Commun.* 2011, 47, 7638.
 [20] N. Swensen, J. L. Diaz Mashar, M. Bardlay, M. Bardley, M. Bar
- [30] N. Svensen, J. J. Diaz-Mochon, M. Bradley, *Chem. Biol.* 2011, 18, 1284.
- [31] N. Svensen, J. J. Diaz-Mochon, K. Dhaliwal, S. Planonth, M. Dewar, J. D. Armstrong, M. Bradley, Angew. Chem. Int. Ed. 2011, 50, 6133.
- [32] S. J. Wrenn, R. M. Weisinger, D. R. Halpin, P. B. Harbury, J. Am. Chem. Soc. 2007, 129, 13137.
- [33] F. Buller, Y. Zhang, J. Scheuermann, J. Schaefer, P. Buehlmann, D. Neri, *Chem. Biol.* 2009, 16, 1075.
- [34] M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao, B. A. Morgan, *Nat. Chem. Biol.* 2009, 5, 647.
- [35] N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick, J. L. Harris, *Chem. Biol.* 2004, 11, 1351.
- [36] D. Pouchain, J. J. Diaz-Mochon, L. Bialy, M. Bradley, ACS Chem. Biol. 2007, 2, 810.
- [37] N. Jessani, B. F. Cravatt, Curr. Opin. Chem. Biol. 2004, 8, 54.
- [38] S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, 22, 568.
- [39] S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Angew. Chem. Int. Ed.* 2007, 46, 4671.
- [40] J. P. Daguer, M. Ciobanu, S. Alvarez, S. Barluenga, N. Winssinger, *Chem. Sci.* 2011, 2, 625.
- [41] M. Ciobanu, K.-T. Huang, J.-P. Daguer, S. Barluenga, O. Chaloin, E. Schaeffer, C. G. Mueller, D. A. Mitchell, N. Winssinger, *Chem. Commun.* 2011, 47, 9321.
- [42] K. Matsuura, M. Hibino, Y. Yamada, K. Kobayashi, J. Am. Chem. Soc. 2001, 123, 357.
- [43] K. Gorska, K.-T. Huang, O. Chaloin, N. Winssinger, Angew. Chem. Int. Ed. 2009, 48, 7695.
- [44] B. A. R. Williams, C. W. Diehnelt, P. Belcher, M. Greving, N. W. Woodbury, S. A. Johnston, J. C. Chaput, J. Am. Chem. Soc. 2009, 131, 17233.
- [45] R. Liu, B. Jiang, H. Y. Yu, J. C. Chaput, *ChemBioChem* 2011, 12, 1813.
- [46] K. Gorska, J. Beyrath, S. Fournel, G. Guichard, N. Winssinger, *Chem. Commun.* 2010, 46, 7742.
- [47] C. Scheibe, A. Bujotzek, J. Dernedde, M. Weber, O. Seitz, *Chem. Sci.* 2011, 2, 770.
- [48] H. Eberhard, F. Diezmann, O. Seitz, Angew. Chem. Int. Ed. 2011, 50, 4146.
- [49] E. A. Englund, D. Wang, H. Fujigaki, H. Sakai, C. M. Micklitsch, R. Ghirlando, G. Martin-Manso, M. L. Pendrak, D. D. Roberts, S. R. Durell, D. H. Appella, *Nat. Commun.* 2012, *3*, 614.
- [50] L. Roglin, M. R. Ahmadian, O. Seitz, Angew. Chem. Int. Ed. 2007, 46, 2704.
- [51] S. Rapireddy, L. Nhon, R. E. Meehan, J. Franks, D. B. Stolz, D. Tran, M. E. Selsted, D. H. Ly, J. Am. Chem. Soc. 2012, 134, 4041.