Translating instructions into function by nucleic acid programmed self-assembly

Nicolas Winssinger

Résumé **Auto-assemblage d'acides nucléiques programmés pour la traduction d'instructions en fonctions**
La prédictibilité de l'hybridation d'acides nucléiques offre une plate-forme simple pour programmer des assemblages moléculaires avec des fonctions émergentes. Cet article illustre quelques exemples issus de notre laboratoire appliquant ce concept en chimie biologique. Ces développements s'appuient sur les PNA (« peptide nucleic acids »), un analogue fonctionnel d'oligonucléotides naturels, pour programmer ces assemblages. Dans un premier exemple, l'hybridation est utilisée pour organiser des mélanges de molécules étiquetées par des PNA sur une surface de type puce à ADN ; dans un deuxième exemple, les étiquettes PNA sont utilisées pour programmer la dimérisation de ligands se traduisant par une affinité augmentée pour une biomolécule ; dans un troisième exemple, l'hybridation est utilisée pour programmer une réaction en alignant des groupes fonctionnels, qui se traduit par un signal fluorescent. Bien que ces exemples ne soient qu'un prélude dans la traduction d'instructions en une fonction, il est espéré qu'ils serviront de bases à des réseaux plus complexes et inspireront de nouveaux développements dans le contexte plus large de la chimie des systèmes.
Mots-clés **Acide nucléique peptide, hybridation, fonction, oligonucléotide, réseaux.**

Abstract The predictable nature of nucleic acid hybridization offers a simple platform to program assemblies with emerging function. This review will highlight some examples from our laboratory towards the use of this concept in chemical biology. These developments capitalize on peptide nucleic acids (PNAs), a functional analogue of natural oligonucleotides to program assemblies. In the first example, hybridization is used to organize mixtures of molecules tagged with PNAs into a microarray format; in the second example, the PNA tags are used to program the dimerization of ligands onto a nucleic acid template which translates into enhanced affinity for a biomolecule; in the third one, hybridization is used to program a reaction by aligning reactive functional groups and translated into a fluorescent signal. While these examples are only a prelude in translating instructions into a function, we hope that they can be the basis of more complex network and inspire further developments in the broader context of system chemistry.
Keywords **Peptid nucleic acid, hybridization, function, oligonucleotide, network.**

The seminal publication of the DNA double helix by J.D. Watson and F.H. Crick offered an immediate solution to the storage and amplification of cellular information [1]. Since then, the predictable nature of nucleic acid hybridization has inspired researchers to harness its properties for new applications [2]. From a supramolecular perspective, the formation of oligonucleotide complexes can be conveniently tuned by varying the length of the oligomer and choice of nucleic acids. For example, a DNA oligonucleotide containing 8 residues (8-mer) will be in a dynamic equilibrium with its complimentary strand at room temperature while a 25-mer will form a complex with such slow dissociation kinetics at room temperature that it can be practically considered irreversible within the time frame of a day. A number of modifications have been reported which stabilize or destabilize a hybridized duplex and allow further fine tuning of the desired affinity or dissociation kinetics. The fact that duplex formation is sequence specific enables one to program association within complex mixtures. From this stand point, if one considers strictly the four natural base pairs (a number of designed base pairs have extended the repertoire beyong the C:G and A:T/U pairs), an 11-mer offers over one million unique sequences that can be used to tag or barcode small or macromolecules. The capacity of natural or unnatural oligonucleotides to program assemblies has important applications at the frontiers of chemistry, biology and material science. As will be discussed in the subsequent section, some applications can benefit from the use of unnatural nucleic acids, particularly peptide nucleic acids (PNAs).

Peptide nucleic acids (PNAs)

PNAs are functional analogues of the natural oligonucleotides where the ribose-phosphate backbone has been replaced by an *N*-(2-aminoethyl)glycine unit (*figure 1*) [3-4]. Despite this fairly dramatic change and the additional conformational flexibility of the backbone, P.E. Nielsen and co-workers demonstrated in the early 90's that PNA hybridize to DNA or RNA following the same Watson and

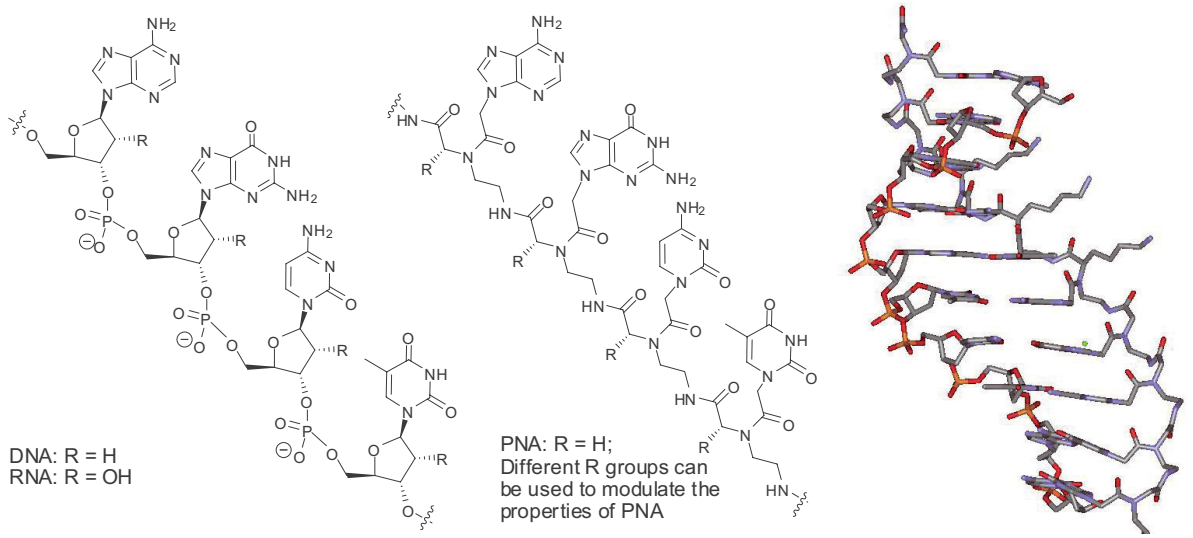

Figure 1 - Structure of DNA or RNA (left) compared to PNA (middle), crystal structure of DNA:PNA duplex (right, PDB id: 1NR8).

Crick base pairing rules as for natural oligonucleotides and that PNAs hybridize preferentially in the antiparallel mode (amino terminus of the PNA facing the 3'-end of the oligonucleotide).

Since then, a number of backbone analogues have been explored and one substitution which is well tolerated is the replacement of the glycine unit from the *N*-(2-aminoethyl)glycine backbone by other amino acids, preferably in the *R* stereochemistry (*figure 1*). A particularly significant one is the substitution of glycine for arginine as it confers cellular permeability [5]. Impressively, PNA-DNA complexes were found to be more stable than the corresponding DNA duplexes and the deleterious impact of a base-pairing mismatch on the complex's stability was higher for PNA-DNA than for DNA-DNA (a single mismatch within the sequence typically reduces the thermal stability (T_m) of a PNA-DNA complex by 8-20 °C). The increased thermal stability of PNA-DNA or PNA-RNA complex compared to DNA or RNA duplex is attributed to the lack of negative charges on the PNA backbone, thus avoiding the inherent electrostatic repulsive interaction of natural oligonucleotide duplexes. Accordingly, PNA-DNA interaction is less sensitive to presence of salts which can dissipate the charge of the phosphate groups and attenuate the repulsive interaction in natural oligonucleotide (the thermal stability of DNA duplexes increases significantly in a 1 M NaCl solution relatively to a 0.1 M NaCl).

Hence, PNA hybridization properties can be exploited in a broader array of buffers or even using mixtures of organic solvent. The enhanced recognition properties have brought tremendous attention to PNAs and their application in diagnosis and as biochemical reagents. From a chemistry perspective, while the automated synthesis of DNA is exquisitely well optimized, the sensitivity of the purine-glycoside bond to acidic conditions limits tremendously the protecting group combinations which may be used for DNA or RNA synthesis. On the other hand, the oligomerization of PNA relies on a simple peptide bond formation (*figure 2*) and the chemical stability of the PNAs to strong acids and bases allows a diversity of protecting groups for the terminal nitrogen (including of course Fmoc and Boc) and nucleobase

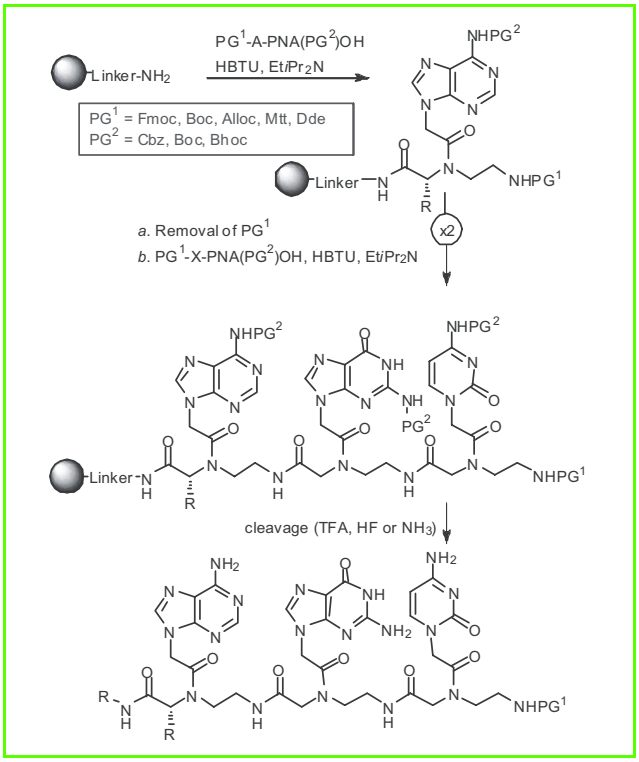

Figure 2 - Solid phase synthesis of PNA.
PNA monomers with different protecting group combination on the terminal nitrogen (PG^1) and nucleobase nitrogen (PG^2) are oligomerized under standard peptide synthesis protocol using HATU or other condensating agents and released from the resin under acidic or basic treatment (the nucleobase protecting groups are typically removed simultaneously).

(amides, Cbz or acid labile protecting groups) to be used. Thus PNA synthesis is more prone to be compatible with other chemistries for the tagging or co-synthesis of small or macromolecules.

Last but not least, despite the peptidic structure of the PNA backbone, PNAs are not prone to enzymatic degradation by proteases and, contrarily to DNA or RNA, have been found to be stable *in vivo* and to cell extracts. On

the other hand, a limitation of PNAs relatively to natural oligonucleotides is the fact that they are not substrates for polymerases and thus cannot be amplified by the polymerase chain reaction (PCR). A more practical limitation is that PNA oligomers are not as readily accessible from commercial sources as DNA which has limited their use predominantly to laboratories with PNA (or peptide) synthesis capabilities.

Nucleic acid tagging for self-organization into microarray

The 90s witnessed the development of a new analytical format to analyze complex mixtures of nucleic acids: DNA microarrays, wherein different DNA probes are immobilized on a surface (typically glass) at high density (10 000 to 100 000 sequences per cm^2). This format emerged as a highly attractive solution in response to the increasing pressures to boost screening throughput and reduce sample requirement following advances in genomics. The microarray format enables the parallel detection of thousands of analytes in a few microliters and as such, represents at significant miniaturization relatively to microtiter plate-based format. It was widely embraced in the late 90s to measure the quantity of cellular extract of *m*RNA and its success enticed researchers from other disciplines to embrace it for small molecules, carbohydrates, antibodies and other proteins [6].

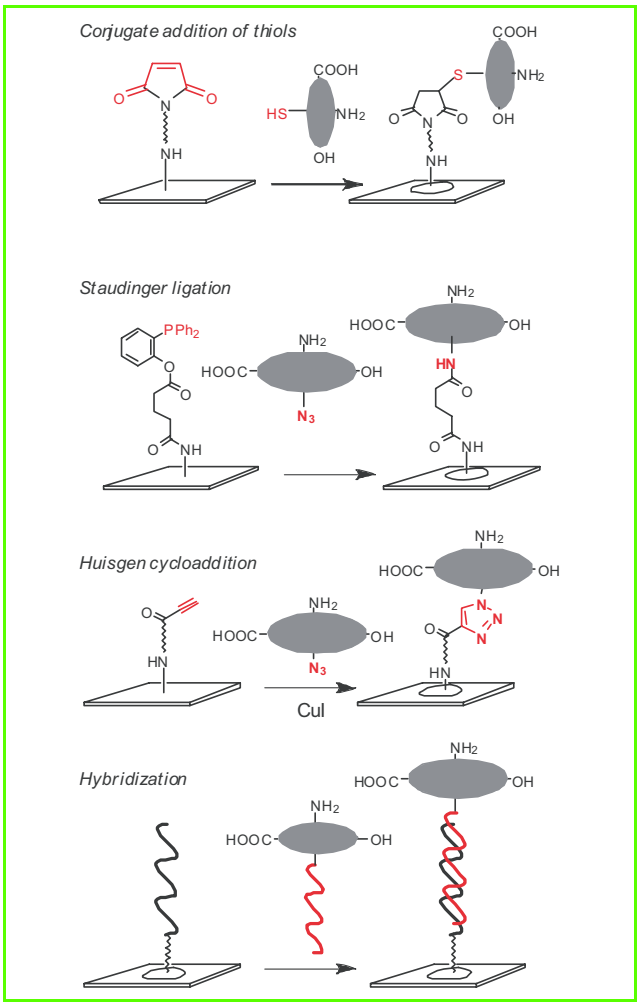

Figure 3 - Immobilization of molecules to a microarray by chemoselective coupling or hybridization.

The homogeneous physicochemical characteristics of oligonucleotides are well suited for a generic immobilization method however the diversity amongst small molecules and proteins required development of novel strategies for immobilization. Several elegant solutions exploiting bio-orthogonal reactions (thiol conjugate addition, Staudinger ligation, Huisgen cycloaddition, *figure 3*) were developed to immobilize molecules on the array surface. Alternatively, molecules can be tagged with oligonucleotides and microarrayed by simple hybridization to readily available DNA microarrays (*figure 3*).

For this purpose, we reasoned that PNA tagging would offer significant advantages [7]. The flexibility of the PNA synthesis can accommodate most reactions used in combinatorial synthesis. The library can thus be synthesized using the powerful split and mix technique as illustrated in *figure 4*. At each step where a synthon of diversity is introduced, its structure is encoded using a unique three to four letter codon on the PNA. At the end of the synthesis, a fluorophore is added on the PNA such that it can be detected upon hybridization onto a microarray, and the library is cleaved from the polymeric support affording a mixture in solution wherein every library member is covalently linked to a PNA tag which encodes its synthetic history, hence, its structure. Based on the higher affinity of

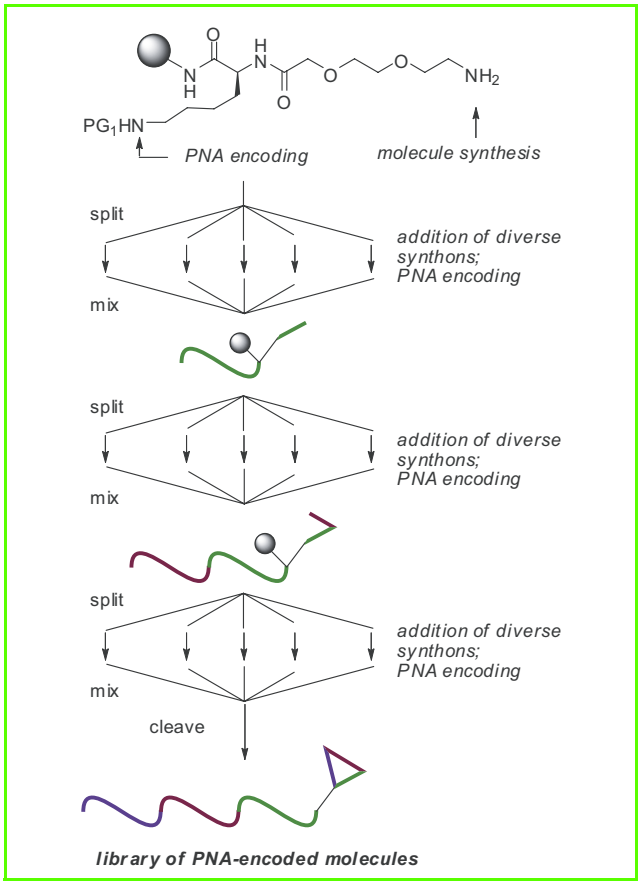

Figure 4 - Combinatorial synthesis of a PNA-encoded library by split and mix synthesis.
Starting from a resin bearing a linker with two orthogonal functionalities, the library can be synthesized on one arm of the linker while the other arm is used for the PNA synthesis. At each step the resin is split into different reaction vessel for the chemical coupling of the synthon (represented as a colour line on the right of the linker) and its corresponding PNA codon (represented as a wave on the left of the linker). The resin is then mixed and the cycle is reiterated to obtain a library of all possible combinations. Cleavage of the library from the resin affords the library as a mixture in solution.

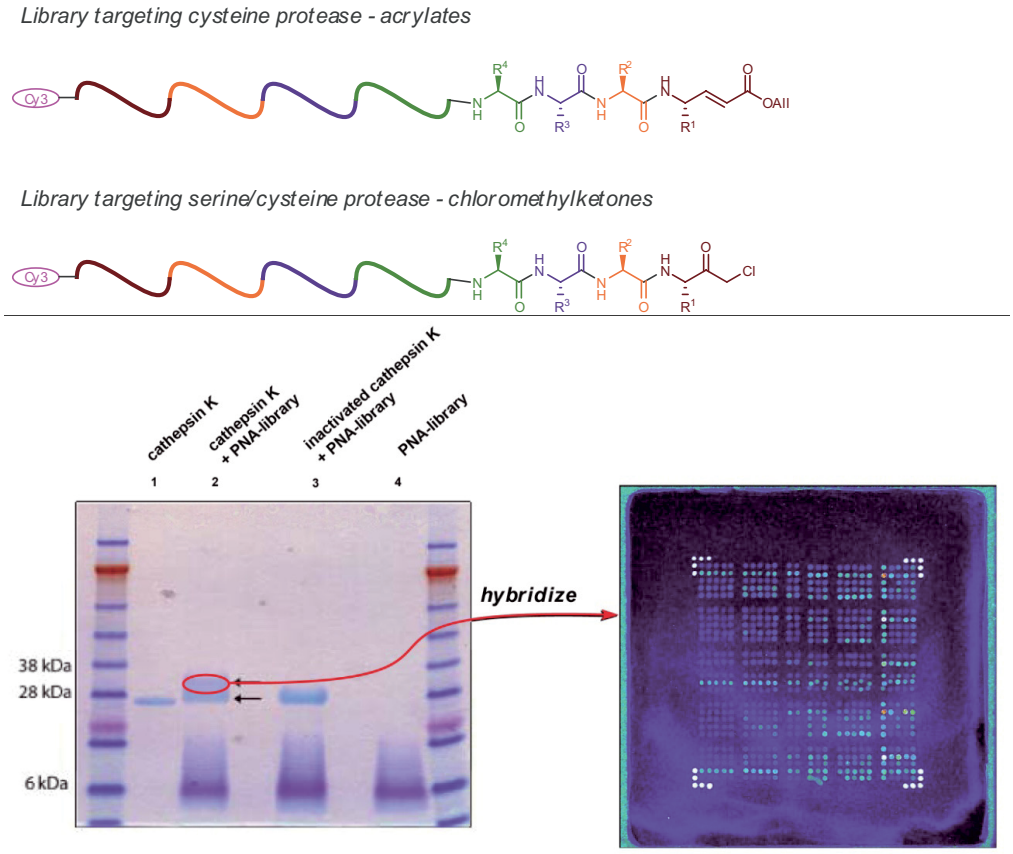

Figure 5 - Top: representative examples of PNA-encoded libraries designed to target proteases with a reactive functional group (shown in red) which can form a covalent bond with the proteases. The PNA is labeled with a fluorophore (Cy3) for microarray detection. Bottom: screening of PNA-encoded libraries by selection of the fittest ligand. The library is incubated with a protease of interest and the selected ligand is isolated by gel electrophoresis. The band corresponding to the protein-inhibitor-PNA conjugate is cut from the gel, and the product hybridized to a microarray. The most intense spot corresponds to the fittest inhibitor.

PNA relatively to natural oligonucleotides, a 14-mer PNA tag of is sufficient to get very sensitive detection. Attesting to the miniaturization of this approach, 1 fmol of PNA is sufficient to get reliable detection. An advantage of this supramolecular arraying strategy over direct covalent immobilization on the microarray is that the library can be used in solution prior to its reformatting into a microarray. This latter point opens the possibility to select for desired properties amongst library members.

An obvious application which we have actively pursued is to select for the fittest ligand in a library of small molecules targeting a given protein. As shown in *figure 5*, libraries can be designed to target specific classes of proteases based on a reactive functional group which will form a covalent bond between the protease and the small molecule-PNA adduct. When incubated with a protease of interest, the fittest ligand will react with the protease which can be isolated from the rest of the library by size exclusion filtration or by gel electrophoresis. Hybridization on a microarray will then reveal the structure of the selected inhibitor. These techniques have been used to discover a selective inhibitor against a dust mite protease from crude dust mite extract [8] and to identify selective inhibitors against closely related proteases [9]. The utility of nucleic acid tagged small molecules was further extended by C.M. Niemeyer and co-workers to immobilize live cells in a microarray format based on the interaction of cell surface receptors to small molecule immobilized on the microarray through hybridization [10].

Combinatorial assembly of PNA-tagged ligands onto a DNA template

Specific recognition and signal transduction processes occurring in biological systems are often leveraged on multimeric interactions which provide additional level of regulation by means of distance between the molecules and their respective arrangement in space, in addition to their chemical nature. Receptor oligomerization provides a well characterized mechanism for signal transduction across a cell membranes and heterodimerization has been harnessed by nature to provide a combinatorial output of signals. While a number of landmark studies have already highlighted [11-12], the benefit of ligand oligomerization, accessing multivalent ligands with controlled distance and geometry pose new synthetic challenges. The programmability of hybridization with nucleic acid-ligand conjugates offers an attractive solution towards this goal and could allow recapitulating complex biological interactions with simpler small molecule ligands which are dimerized through hybridization as illustrate in *figure 6*.

As a proof of principle, we investigated the gain of affinity conferred by dimerizing ligands of DR5, a prototypical member of the TNF receptor super family which relies on oligomerization for signal transduction. Using peptide macrocycles tagged at different positions with a PNA (*figure 6*), a large number of different dimer geometries could be rapidly tested for their gain in affinity towards the dimeric

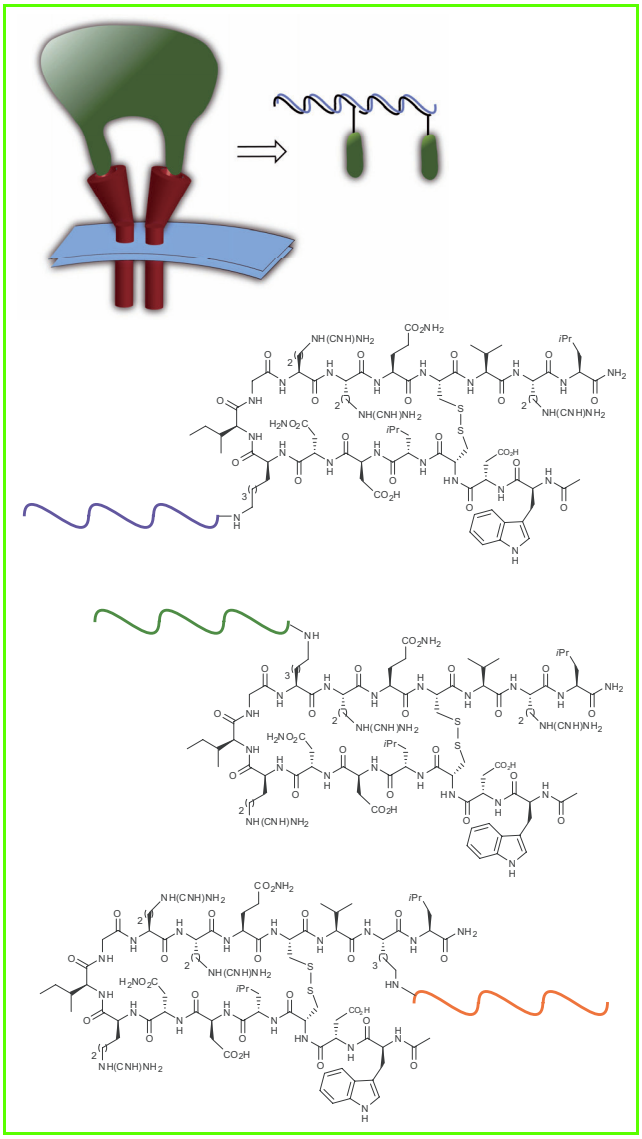

Figure 6 - Top: schematic representation of a cell-surface receptor dimer (red) interacting with a divalent protein (green) which could be emulated a nucleic acid assembly of dimerized ligands. Bottom: examples of ligands targeting DR5 with different attachment points to a PNA. Based on the sequence of the PNA, the ligands can be dimerized in different combinations and with different geometries.

receptor. This led to the identification of dimers which indeed had substantially improved affinity for the receptor and perhaps more importantly, slower dissociation kinetics, consistent with cooperative effect associated with divalent interactions [13].

Another example of biological interactions which capitalizes on the cooperativity between multiple ligands to achieve high binding avidity is carbohydrate-protein interaction. For instance, the cholera toxin, which forms a pentamer, interacts with cell surface carbohydrates with high affinity while the interaction between a single unit of the toxin and the carbohydrate ligand is rather weak. Another therapeutically significant example is the interaction of 2G12, a broadly neutralizing antibody against HIV, which interacts with the carbohydrates on the surface of the virus. The carbohydrate is a nonasaccharide of glucose amine and branching mannoses (*figure 7*). It was shown that the antibody actually interacts only with the terminal portion of this complex carbohydrate and has weak affinity for a single

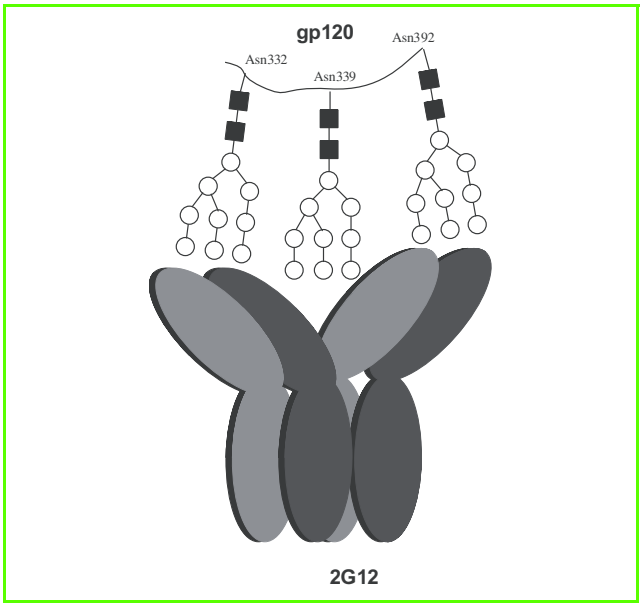

Figure 7 - Schematic representation of 2G12, a dimeric antibody which broadly neutralizes HIV, interacting with the carbohydrate motifs coating the HIV capsid protein (gp120).

unit of this nonasaccharide. But the antibody forms a dimer which achieves high affinity to HIV by interacting with multiple copies of the nonasaccharide [14]. We wondered if the geometry of such complex interaction could perhaps be mimicked with simpler carbohydrates (disaccharide or trisaccharides) tagged with PNA which would be oligomerized with controled topology on a DNA template. As illustrated in *figure 8*, the distance between the ligands can easily be varied depending on the choice of DNA template. It is also important to note that double strand DNA has a very long persistent length and as such, should effectively work as a crude molecular ruler. This strategy was used to evaluate over thirty different geometries of carbohydrate

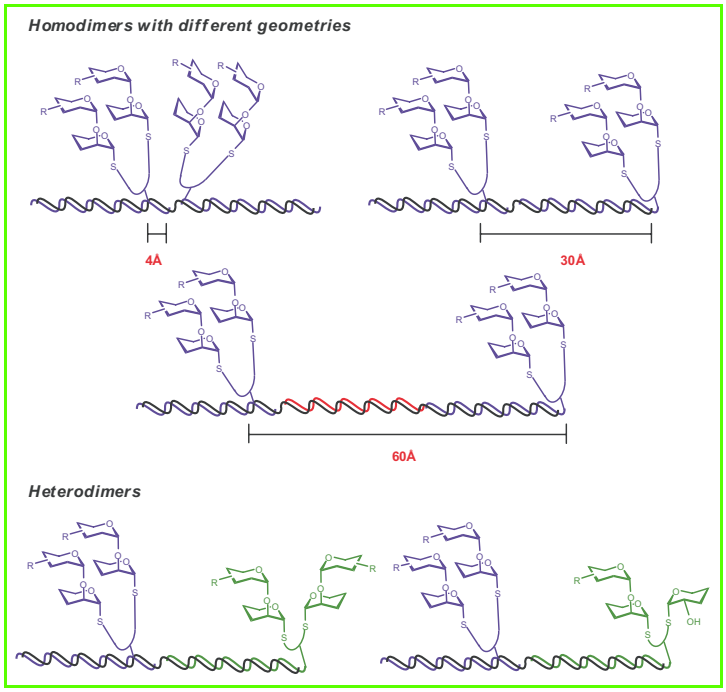

Figure 8 - Schematic representation of nucleic acid programmed assemblies of carbohydrates with controlled geometries.

oligomers against 2G12 and led to the identification of an assembly with appropriately spaced mannose disaccharides which could recapitulate the cooperativity between the different binding sites to achieve μM affinity [15]. It should be noted that the complex nonasaccharide has no measurable affinity in this assay attesting to the level of cooperativity achieved with the DNA template assembly.

In a broader context, these two examples demonstrate that nucleic acid templated assembly of PNA-tagged molecules can be translated into a binding event to a macromolecule and as such represents a gain of function. The combinatorial assembly of larger libraries could be achieved on a microarray which would be interrogated with proteins. Conversely, as the instruction for the assembly of the ligand fragments are DNA, if one selects an assembly from a combinatorial library by affinity using an immobilized protein, the identity of that assembly can be decoded using DNA sequencing technologies.

Templated reactions

Reactions accelerated by the preorganization of two reactive species on a nucleic acid template have a long history in the context of prebiotic chemistry [16]. However, applications of this concept beyond the questions related to nucleic acid amplification are more recent [17].

A useful and interesting application is the design of reactions that yield a functional product such as a fluorophore or a bioactive small molecule in response to a specific nucleic acid sequence (*figure 9*). A practical application of such a system would be to detect by fluorescence the presence of a nucleic acid sequence for nucleic acid sensing (detection of a mutation, etc.) or imaging (the presence of a particular nucleic acid sequence in a cell or organism) [18]. From an imaging standpoint, while several technologies have had a tremendous impact on our understanding of protein function and dynamics [19], there does not exist a general solution to image messenger, ribosomal or micro RNA in intact non-engineered cells.

A particularly attractive feature of nucleic acid templated reactions is the fact that the template could act catalytically (if the product of the reaction doesn't compete unfavorably with the starting materials), thus providing a highly sensitive detection method. In an elegant demonstration of this concept, J.-S. Taylor and co-workers showed that short DNA fragments bearing an imidazole on the 5'-end of a 15-mer could hydrolyze a nitrophenol ester on a second DNA probe (9-mer) upon exposure to a complementary template [20]. It was indeed shown that the reaction turned over

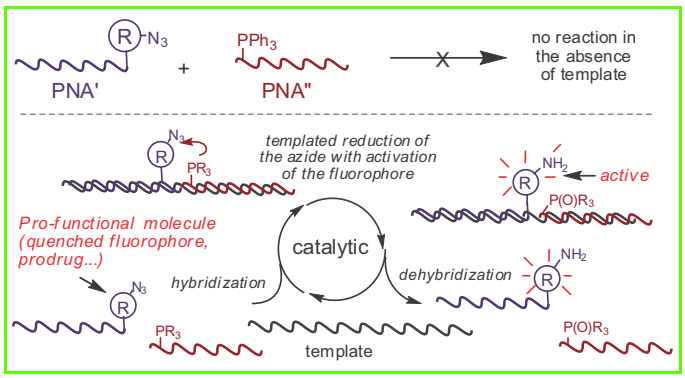

Figure 9 - Schematic representation of nucleic acid template Staudinger reaction.

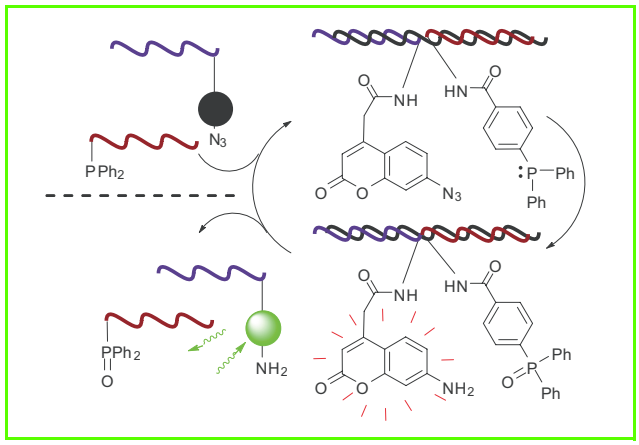

Figure 10 - Nucleic acid catalyzed reduction of azidocoumarin.

leading to a higher concentration of nitrophenol than template, thus demonstrating that the template could turn over catalytically.

Our interest in this area has focused on the Staudinger reaction (the reduction of an azide by phosphine) based on the fact that this reaction is known to be compatible and orthogonal with cellular chemistry. We reasoned that a fluorophore which depended on an amino group (or aniline) for high quantum yield of fluorescence would be quenched if that aniline was converted to an azide. Upon Staudinger reduction of this azide, the fluorophore would thus be revealed. As illustrated in *figure 10*, using a PNA labelled with a quenched coumarin fluorophore and a second PNA with a phosphine, upon hybridization to the template, the phosphine reduces the azide leading to a highly fluorescent coumarin product. For the purpose of diagnosis, we noted that the reaction performance could be enhanced with the use of formamide as a co-solvent thus resulting in extremely fast conversion. The reaction affords over 20% conversion with 1% template within 30 min. While this level of turn-over is sluggish from a catalysis perspective, it is significant from a signal amplification perspective. Importantly, a single nucleotide mismatch dramatically reduced the rate of reaction.

However, two obstacles precluded the use of this approach in cellular imaging. First, unmodified PNA have poor cellular permeability. Second, coumarin, which is excited with blue light, is poorly suited for cellular imaging because of the high level of background fluorescence from proteins and their cofactors in this part of the visible spectrum. However, the same strategy based on azide quenching can be applied to other fluorophores such as rhodamine which is excited with green light. As discussed in the above section on PNA, incorporation arginie residues within the PNA backbone does confer cellular permeability.

Using these modified PNA we were able to achieve delivery of the PNA in live cells and showed that using PNA sequences designed to selectively target an *m*RNA of a highly expressed protein resulted in higher fluorescence than control sequences (*figure 11*) [21].

Outlook

Organization is an intrinsic property of living systems. While the examples discussed herein are dwarfed by the complexity and intricacies of living systems, the development of approaches which can translate a signal or an instruction

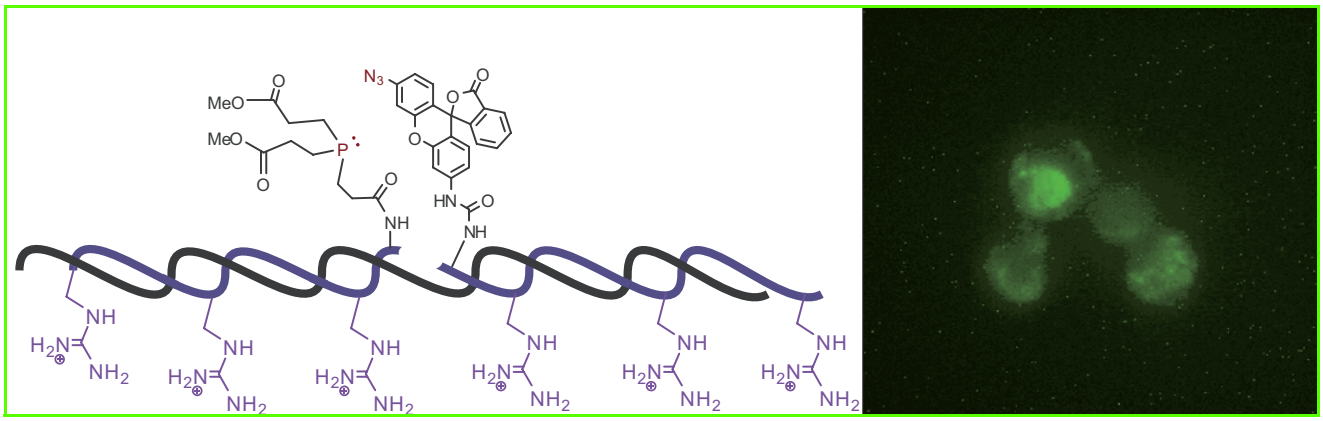

Figure 11 - Nucleic acid imaging in live cells.
Top: schematic representation of the nucleic acid catalyzed reduction of azidorhodamine. Bottom: imagining of a transcript mRNA in live cells.

into a function are a first step in the design of more complex networks. It is safe to predict that nucleic acid-based programmed supramolecular assemblies will continue to provide a reliable framework to achieve new biological functions whether by controlling ligand dimerization, conformation or positioning, or by promoting a chemical transformation. PNAs are an attractive platform in this endeavour for their biological stability, flexible chemistry and high affinity.

Acknowledgment

It is with pleasure that numerous collaborators whose contributions made the described work possible and enjoyable are acknowledged.

References

[1] Watson J.D., Crick F.H., The structure of DNA, *Cold Spring Harb. Symp. Quant. Biol.*, **1953**, *18*, p. 123.
[2] Feldkamp U., Niemeyer C.M., Rational design of DNA nanoarchitectures, *Angew. Chem. Int. Ed. Engl.*, **2006**, *45*, p. 1856.
[3] Egholm M., Buchardt O., Christensen L., Behrens C., Freier S.M., Driver D.A., Berg R.H., Kim S.K., Nordén B., Nielsen P.E., PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules, *Nature*, **1993**, *365*, p. 566.
[4] Nielsen P.E., PNA technology, *Mol. Biotech.*, **2004**, *26*, p. 233.
[5] Zhou P,. Wang M., Du L., Fisher G.W., Waggoner A., Ly D.H., Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA), *J. Am. Chem. Soc.*, **2003**, *125*, p. 6878.
[6] Winssinger N., Pianowski, Z., Debaene F., Probing biology with small molecule microarrays (SMM), *Top. Curr. Chem.*, **2007**, *311*, p. 311.
[7] Harris J.L., Winssinger N, PNA encoding (PNA = peptide nucleic acid): From solution-based libraries to organized microarrays, *Chemistry-A European Journal*, **2005**, *11*, p. 6792.
[8] Harris J., Mason D.E., Li J., Burdick K.W., Backes B.J., Chen T., Shipway A., Van Heeke G., Gough L., Ghaemmaghami A., Shakig F., Debaene F., Winssinger N., Activity profile of dust mite allergen extract using substrate libraries and functional proteomic microarrays, *Chemistry & Biology*, **2004**, *11*, p. 1361.
[9] Urbina H.D., Debaene F., Jost B., Bole-Feysot C., Mason D.E., Kuzmic P., Harris J.L., Winssinger N., Self-assembled small-molecule microarrays for protease screening and profiling, *ChemBioChem*, **2006**, *7*, p. 1790.
[10] Schroeder H., Ellinger B., Becker C.F., Waldmann H., Niemeyer C.M., Generation of live-cell microarrays by means of DNA-Directed immobilization of specific cell-surface ligands, *Angew. Chem. Int. Ed. Engl.*, **2007**, *46*, p. 4180.
[11] Mammen M., Choi S.K., Whitesides G.M., Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors, *Angew. Chem. Int. Ed. Engl.*, **1998**, *37*, p. 2755.
[12] Kiessling L.L., Gestwicki J.E., Strong L.E., Synthetic ligands as probes of signal transduction, *Angew. Chem. Int. Ed. Engl.*, **2006**, *45*, p. 2348.
[13] Mammen M., Colton I.J., Carbeck J.D., Bradley R., Whitesides G.M., Representing primary electrophoretic data in the 1/time domain: comparison to representations in the time domain, *Anal. Chem.*, **1997**, *69*, p. 2165.
[14] Calarese D.A., Scanlan C.N., Zwick M.B., Deechongkit S., Mimura Y., Kunert R., Zhu P., Wormald M.R., Stanfield R.L., Roux K.H., Kelly J.W., Rudd P.M., Dwek R.A., Katinger H., Burton D.R., Wilson I.A., Antibody domain exchange is an immunological solution to carbohydrate cluster recognition, *Science*, **2003**, *300*, p. 2065.
[15] Mammen M., Helmerson K., Kishore R., Choi S.K., Phillips W.D., Whitesides G.M., Optically controlled collisions of biological objects to evaluate potent polyvalent inhibitors of virus-cell adhesion, *Chem. Biol.*, **1996**, *3*, p. 757.
[16] Orgel L.E., Unnatural selection in chemical systems, *Acc. Chem. Res.*, **1995**, *28*, p. 109.
[17] Li X., Liu D.R., DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules, *Angew. Chem. Int. Ed. Engl.*, **2004**, *43*, p. 4848.
[18] Silverman A.P., Kool E.T., Detecting RNA and DNA with templated chemical reactions, *Chemical Reviews*, **2006**, *106*, p. 3775.
[19] Giepmans B.N., Adams S.R., Ellisman M.H., Tsien R.Y., The fluorescent toolbox for assessing protein location and function, *Science*, **2006**, *312*, p. 217.
[20] Ma Z., Taylor J.-S., Nucleic acid-triggered catalytic drug release, *Proceedings of the National Academy of Sciences of the United States of America*, **2000**, *97*, p. 11159.
[21] Gao J., Mammen M., Whitesides G.M., Evaluating electrostatic contributions to binding with the use of protein charge ladders, *Science*, **1996**, *272*, p. 535.

Nicolas Winssinger
est professeur à l'Université de Strasbourg*.

* Institut de Science et Ingénierie Supramoléculaires (ISIS – UMR 7006 CNRS), Université de Strasbourg, 8 allée Gaspard Monge, F-67000 Strasbourg.
E-mail: winssinger@unistra.fr