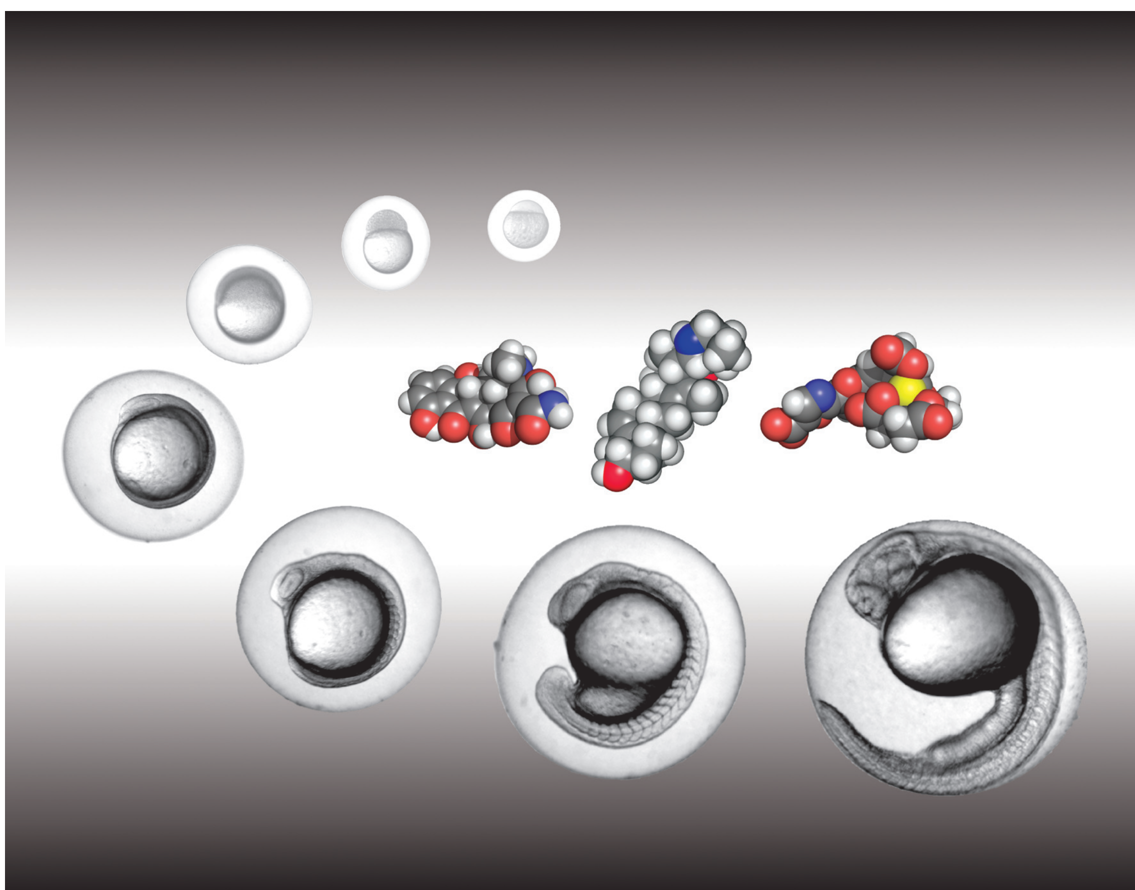
Chem Soc Rev

This article was published as part of the

2008 Chemistry–Biology Interface Issue

Reviewing research at the interface where chemistry meets biology

Please take a look at the full table of contents to access the other papers in this issue



Nucleic acid encoding to program self-assembly in chemical biology†

Zbigniew L. Pianowski and Nicolas Winssinger*

Received 2nd February 2008
First published as an Advance Article on the web 17th April 2008
DOI: 10.1039/b706610b

This *tutorial review* serves as an introduction to the use of oligonucleotides and in particular peptide nucleic acids (PNAs) to encode function beyond heredity. Applications in chemical biology are reviewed starting with the use of nucleic acid tags to program self-assembled microarrays of small and macromolecules, followed by the use of nucleic acid templated reactions for the purpose of DNA or RNA sensing and finally, the use of nucleic acid templates to display ligands.

1 Introduction

Following the landmark publication of the structure of DNA by Watson and Crick, generations of scientists have been inspired by the programmable nature of nucleic acid hybridization and have sought to apply these principles beyond their natural function. From a supramolecular perspective, oligonucleotides of just 11 residues assembled with the basic four letter alphabet offer over 1 million unique sequences that can be used to tag or barcode small- or macro-molecules. The tags can be used to organize mixtures of labeled entities based on hybridization with applications ranging from material science and nanotechnology to chemical biology.[1] The binding strength of oligonucleotide complexes can be conveniently tuned by varying the type of nucleic acids and the lengths of the oligomers. The tagged components can then be positioned on surfaces, assembled as dimers or higher order structures based on the complementarity of their tags with nucleic acid templates. Conversely, the supramolecular organization can be used to bring reactive species within the proper alignment to induce a reaction. The capacity of natural or unnatural oligonucleotides to program assemblies has important applications at the frontiers of chemistry and biology which will be reviewed herein.

Organic and Bioorganic Chemistry Laboratory, Institut de Science et Ingénierie Supramoléculaires, Université Louis Pasteur, 8 allée Gaspard Monge, Strasbourg, 67000. E-mail: winssinger@isis-u.strasbg.fr; Fax: +33 (0)3 89 24 51 12; Tel: +33 (0)3 89 24 51 13
† Part of a thematic issue examining the interface of chemistry with biology.

2 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 (Fig. 1).[2,3] Despite the additional conformational flexibility of the backbone, Nielsen and co-workers demonstrated in the early 1990s that PNAs hybridize similarly to DNA or RNA following the same Watson and 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). 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 complexes compared to DNA or RNA duplexes is

Zbigniew L. Pianowski was born in 1982 in Kraków, Poland. After winning the 47th Polish Chemistry Olympiad and being awarded a prize in the 13th EU Contest for Young Scientists, he received his MSc summa cum laude from the University of Warsaw, Poland. Currently he is carrying out his PhD in Strasbourg, France, under the supervision of Professor Nicolas Winssinger on templated reactions for oligonucleotide detection.

Nicolas Winssinger was born in Belgium. He received his BSc from Tufts university and then joined the group of K. C. Nicolaou at the Scripps Research Institute for his PhD. He then remained at Scripps as an NIH postdoctoral fellow in the group of Peter G. Schultz. In 2002 he moved to the Institut de Science et Ingénierie Supramoléculaires, Université Louis Pasteur in Strasbourg as an assistant professor and was promoted to full professor in 2005. His current interests lie in chemical biology, total and combinatorial synthesis, oligosaccharide chemistry and templated reactions.

 This journal is © The Royal Society of Chemistry 2008

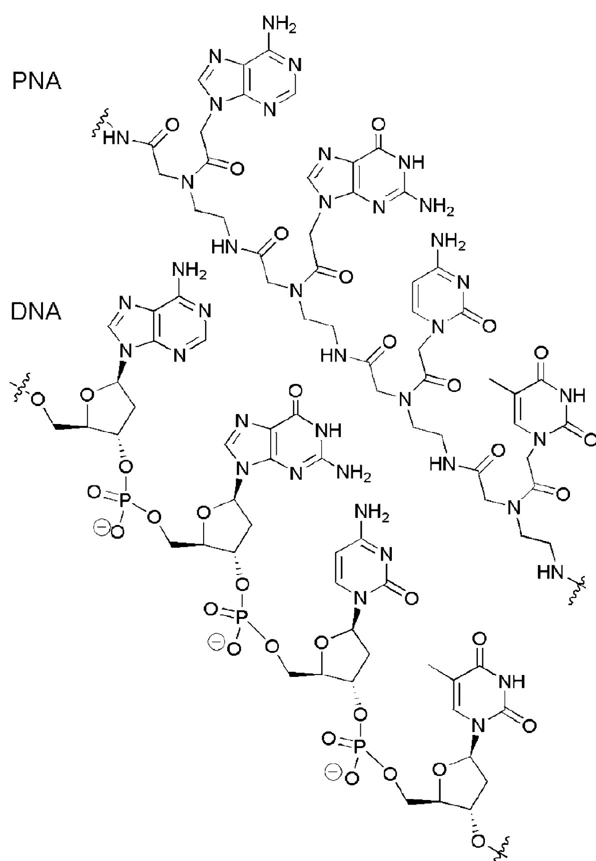


Fig. 1 Structure of PNA and DNA oligomers.

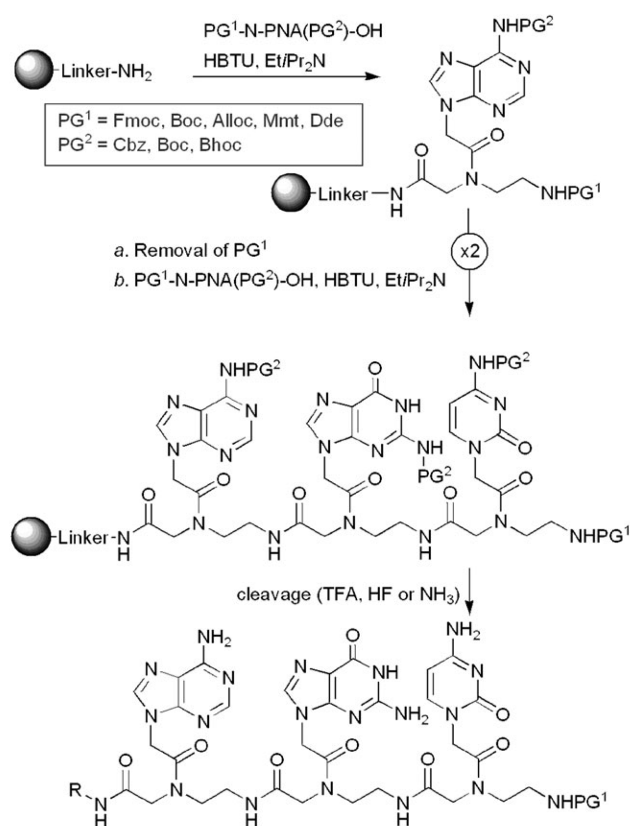


Fig. 2 Solid-phase synthesis of PNA: PNA monomers with different protecting group combinations on the terminal nitrogen (PG^1) and nucleobase nitrogen (PG^2) are oligomerized under a standard peptide synthesis protocol using HBTU or other condensing agents and released from the resin under acidic or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

attributed to the lack of negative charges on the PNA backbone thus avoiding the inherent electrostatic repulsive interaction of natural oligonucleotide duplexes. Accordingly, the PNA–DNA interaction is less sensitive to the 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 0.1 M NaCl). The enhanced recognition properties have brought tremendous attention to PNAs and their application in diagnostics 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 (Fig. 2) and the chemical stability of the PNAs to strong acids and bases allows a diversity of protecting group for the terminal nitrogen (including Fmoc and Boc) and nucleobase (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 macro-molecules. 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 in cell extracts. However, cellular permeability of oligonucleotides including PNA remains a significant hurdle for any cellular application. A solution is to temporarily disrupt the cellular membrane by electroporation or pore forming agents to allow penetration of the oligonucleotides. While these techniques have enabled important experiments, their utility is limited to cells in cultures and results in significant cellular stress which may bias experiments. For natural oligonucleotides, cationic polymers can complex to the phosphoribose backbone and enhance their cellular permeability. This technique has been found effective to introduce the DNA or RNA inside cells[4] however, it does not work for PNAs due to their uncharged backbone. Much effort has been devoted to improve the cellular permeability of PNAs by appending cell penetrating peptides or small molecules which are actively taken up by cells with moderate success. Alternatively, a number of modifications in PNA structure were investigated, of which GPNA (a PNA analogue containing an arginine residue rather than a glycine residue—Fig. 3) was shown to be effectively internalized by cells and able to repress the transcription of *m*RNA.[5] Most importantly, GPNAs were found not to be toxic at concentrations up to 10 μM thus raising the possibility that this modification could offer a viable platform for other applications requiring cellular penetration without the use of additional reagents. On the other hand, a limitation of PNAs relative to natural oligonucleotide is the fact that they are not substrates for polymerases and thus cannot be

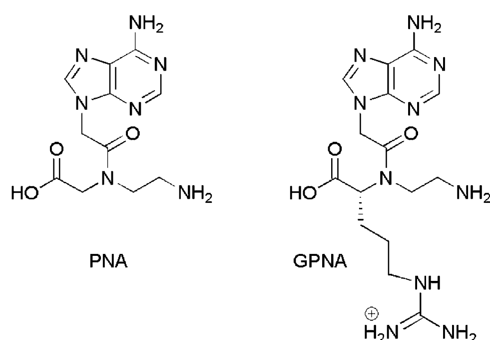


Fig. 3 Structure of GPNA (adenine monomer shown).

amplified by the polymerase chain reaction (PCR). A more practical limitation is that PNA oligomers are not commercially available and their synthesis is not as streamlined as DNA, thus currently limiting their use to laboratories with PNA (or peptide) synthesis capabilities.

3 Tagging small- and macro-molecules for self-microarraying

Microarrays are typically prepared by immobilizing small amounts of sample on a glass surface at a density of 10 000 analytes cm^{-2} for contact printing to $>$100 000 cm^{-2} for photolithographic preparations which can be screened with 50 to 300 μL of sample.[6] This format emerged as a highly attractive solution in response to the increasing pressure to boost screening throughput and reduce sample requirement following advances in genomics. The microarray enables the parallel detection of thousands of analytes in a few microliters and as such represents at least a thousand-fold miniaturization relatively to the classical 384-well microtiter plate format. It was widely embraced in the late 1990s 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] 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 immobilization strategies. Several elegant solutions exploiting bio-orthogonal reactions (oxime formation, Staudinger ligation, Huisgen cycloaddition, Fig. 4) and more indiscriminate reactions based on carbene formation were reported. Aside from the additional immobilization challenges, the detection of interactions involving proteins on microarrays revealed itself to be more complicated than with oligonucleotides, again in part due to the more diverse properties of proteins and the tendency of certain proteins to be unspecifically adsorbed onto surfaces.

Alternatively, small molecules or proteins can be tagged with oligonucleotides and microarrayed by simple hybridization to readily available DNA microarrays (Fig. 4). This strategy has been successfully implemented for antibodies and proteins, carbohydrates and small molecules. The possibility of arraying protein libraries tagged with DNA by hybridization was first demonstrated by Niemeyer *et al.*[7] This strategy was later refined with the use of PNAs which could be

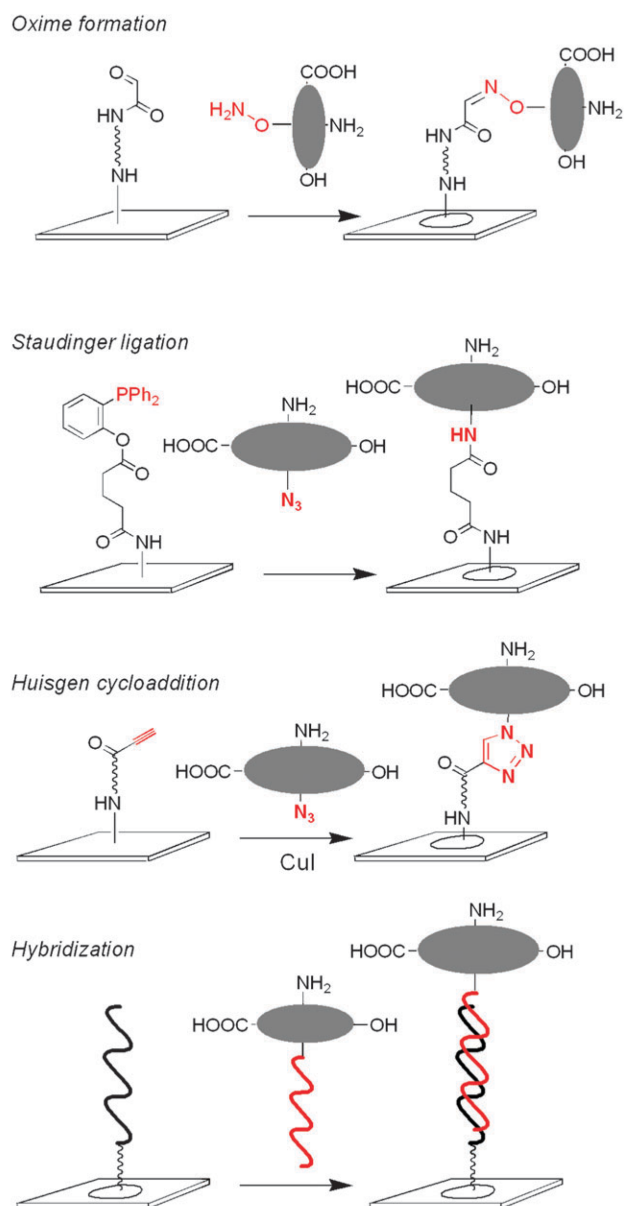


Fig. 4 Immobilization to a microarray by chemoselective coupling or hybridization.

readily derivatized with a terminal cysteine residue. This cysteine enabled the use of the native chemical ligation, which relies on a fast transthioesterification of the terminal cysteine with the protein thioester followed by a S to N migration, thus yielding a native peptide bond. Using intein-fusion proteins (proteins that spontaneously convert to thioesters) which can be readily expressed, the PNA could be chemoselectively coupled to the C-terminus of the proteins and effectively purified in the hybridization step.[8] The utility of oligonucleotide-tagged antibodies was recently extended by Heath and coworkers who showed that a mixture of different cell types could be resolved on a microarray by incubating the cell culture with oligonucleotide-tag antibodies specific to each cell type.[9] As a proof of principle, two cell lines expressing a green fluorescent protein and a red fluorescent protein respectively were separated on the microarray (Fig. 5).

This journal is © The Royal Society of Chemistry 2008

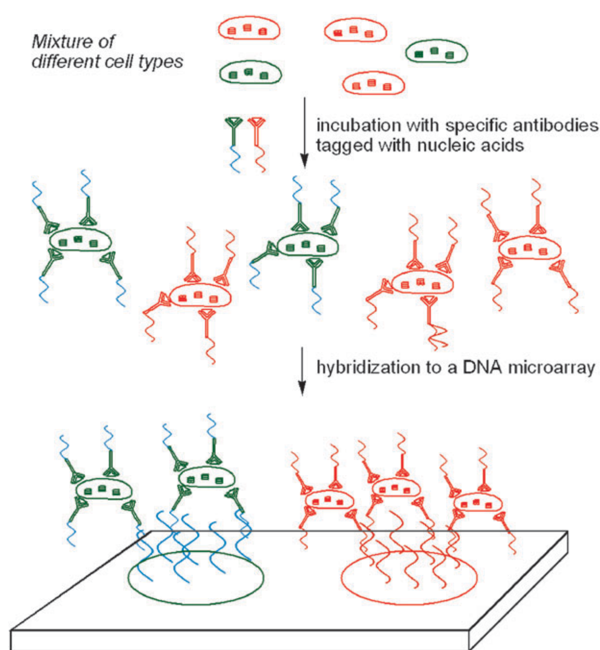


Fig. 5 Two different cell populations expressing a red and a green fluorophore, respectively are incubated with specific antibodies tagged with oligonucleotides. Upon hybridization of a mixture to a DNA microarray, the cells are separated in the respective populations.

For the purpose of small molecule microarrays, we have shown that PNA-encoding offers significant advantages.[10] The flexibility of the PNA synthesis can accommodate most reactions used in combinatorial synthesis. The library can thus be synthesized using the powerful mix-and-split technique with a unique three to four letter codon for every building block in the library. At the end of the synthesis, the library is cleaved from the polymeric support affording a mixture in solution where every library member is covalently linked to a PNA tag which encodes its synthetic history, hence, its structure (Fig. 6). By hybridizing this library to readily available DNA microarrays, every compound is immobilized at a programmed location based on the PNA tag. An advantage of this immobilization strategy is that the library can be screened in solution prior to the immobilization thus avoiding potential problems associated with the display of ligands on surfaces (non-specific interaction of proteins with the surface and high local concentration of a given ligand or substrate at its site of immobilization). A second advantage is that it offers a detection method which is simply not possible with small molecules covalently immobilized on the microarray. As shown in Fig. 6, the library of PNA-encoded small molecules can be separated from macromolecules either by size exclusion filtration (filtration through membranes with defined pore size) or by gel electrophoresis. Thus, inhibitors or ligands to proteins of interest can be separated from the inactive compounds prior to hybridization. This technique has been used to discover a selective inhibitor against a dust mite protease from crude dust mite extract[11] and to identify selective inhibitors against closely related proteases.[12] This technique was also used to measure the activity of proteases in complex mixtures such as crude cell lysates[13] or to measure kinase specificity.[14]

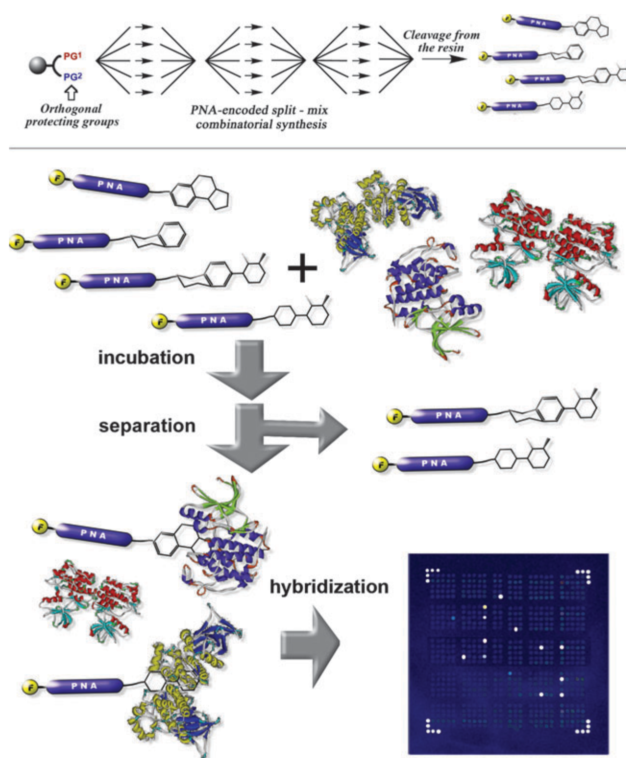


Fig. 6 Synthesis and screening of PNA-encoded libraries. Starting from a resin bearing a linker with two orthogonal protecting groups, the library can be synthesized on one arm of the linker while the other arm is used for the PNA synthesis. Cleavage of the library from the resin affords the library as a mixture in solution (top). The library can be incubated with proteins as a mixture and the compounds that are not bound to proteins are separated by size exclusion or gel electrophoresis. Hybridization then reveals the structure of the ligand (bottom).

The utility of nucleic acid tagged small molecules was further extended by Niemeyer and co-workers to immobilize live cells in a microarray format based on the interaction of cell surface receptors with small molecules immobilized on the microarray through hybridization.[15]

4 Templated reactions and nucleic acid sensing

Reactions accelerated by the preorganization of two reactive species on a nucleic acid template have a long history in 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 fluorescent products for the detection of specific nucleic acid sequences.[18] While imaging technologies have had a tremendous impact on our understanding of protein function and dynamics,[19] there does not exists 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 does not compete unfavorably with the starting materials) thus providing a highly sensitive detection method. In an elegant demonstration of this concept, Ma and Taylor showed that short

This journal is © The Royal Society of Chemistry 2008

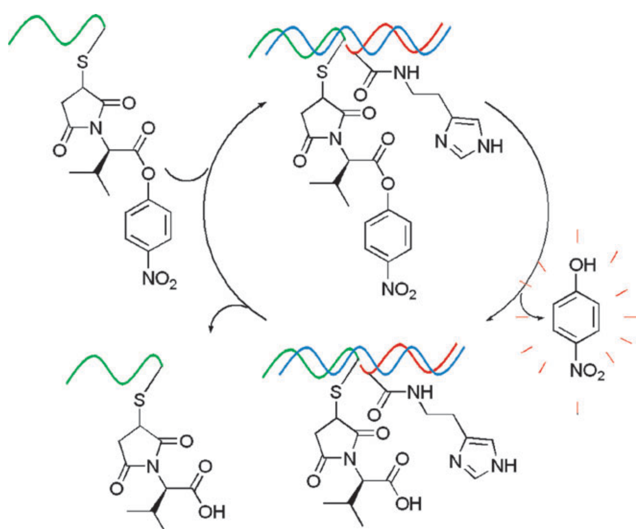


Fig. 7 DNA-templated hydrolysis of nitrophenol.

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 (Fig. 7).[20] The authors showed that the template turned over leading to a higher concentration of nitrophenol than template and suggested that this concept could be used to release chemotherapeutic agents in response to specific genetic information. While the labile ester required for imidazole-catalyzed release may not be compatible with cellular chemistry, the Taylor group reported a second generation leveraged on the reactivity of the aza-ylide generated in a Staudinger ligation between two PNA strands to hydrolyze a quenched fluorescein-ester (Fig. 8).[21] The reaction proceeded under physiological conditions and with good mismatch selectivity (over 30-fold difference in reaction rate for a single mismatch) using 1 μM template. However, partial phosphine oxidation was noted to limit the yield and applicability of this particular system. Kool and co-workers developed an alternative strategy based on nucleophilic thiophosphonate from one DNA strand displacing a fluorophore quencher (Dabsyl) from a second DNA strand upon hybridization to a nucleic acid template (Fig. 9).[22] While their early efforts yielded ligated products with higher affinity for the template than either starting material resulting in product inhibition, this issue has been resolved by introducing a three-atom spacer in the ligation site.[23] This additional spacer affords a distorted product which favours dissociation and template turnover thus enabling 17–100 fold signal amplification. Most impressively, this system was successfully used to detect ribosomal RNA in human cells. For this purpose, modified RNA were used to enhance cellular stability and pore-forming peptides added to the cell culture to enable cellular delivery of the probes.[24] Seitz and co-workers have also recently developed two systems capable of high turnover, both being inspired by the native chemical ligation. In the first one, two PNA segments are ligated through a transthioesterification to yield a product which rapidly rearranges thus yielding a less stable duplex which facilitates dissociation of the template and turnover (not shown).[25a] The second and more recent system is based on the transfer of a fluorophore quencher from one PNA strand to another, each strand bearing a different fluorophore (Fig. 10).[25b] This system is also based on a templated transthioesterification however it leads to the transfer a quencher rather than a ligation. The subsequent rapid S to N acyl migration yields a stable amide bond thus making the transfer irreversible. The progress of the reaction can be measured by both the increase in intensity of one colour and decrease of the second colour thus providing a particularly sensitive detection (albeit partial thioester hydrolysis is difficult to avoid).

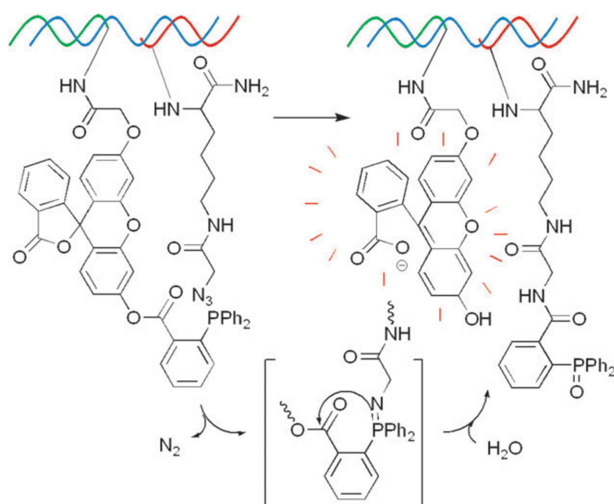


Fig. 8 DNA-templated unmasking of fluorescein *via* a Staudinger reaction.

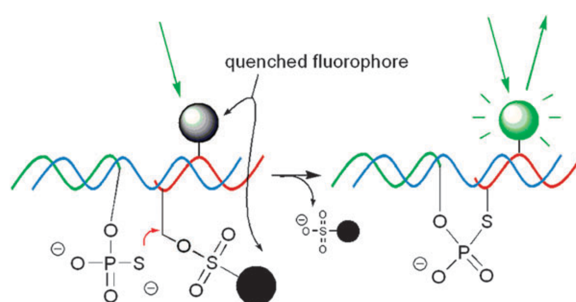


Fig. 9 DNA-templated displacement of a fluorophore quencher (Dabsyl) from a fluorescently labeled strand.

We have recently introduced a system combining the virtues of the bioorthogonal Staudinger reaction with catalytic template turnover based on an azide-quenched coumarin fluorophore.[26] Upon hybridization to the template the phosphine reduces the azide leading to a highly fluorescent coumarin product. For the purpose of diagnostics, we noted that the reaction performance could be increased with the use of formamide as a co-solvent thus resulting in extremely fast conversion (over 20% conversion using low nM concentration of the template within 30 min) and single nucleotide resolution (Fig. 11).

5 Ligand presentation and conformation control

The prospect of using supramolecular assemblies programmed by hybridization to display multiple ligands or controlling a

 This journal is © The Royal Society of Chemistry 2008

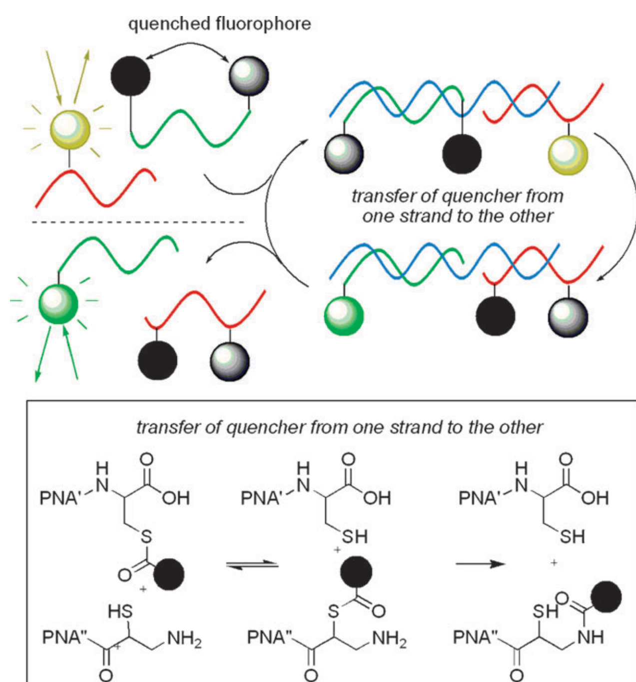


Fig. 10 DNA-templated transfer of fluorophore quencher (Dabsyl, black ball) from one PNA strand labeled with a yellow fluorophore (fluorescein) to a green fluorophore (rhodamine).

ligand conformation has also been explored. In the first case, the work was carried out with DNA tags to capitalize on tag amplification by PCR.[27] As a proof of principle, Neri and co-workers demonstrated that starting with a known small pharmacophore, the affinity for its target could be increased by a factor >40 fold in an analogous process to phage display maturation. The two proteins used in this first pilot experiment were human serum albumin and bovine carbonic anhydrase. Thus, starting with a library of a sulfonamides coupled to the 3′-end of one strand, which was hybridized to 137 oligonucleotides bearing a small molecule at the 5′-end as well as a 6-mer overhang on the 3′-end encoding the structure of the small molecule (Fig. 12) led in each case to the identification of a small molecule which, once covalently linked to the pharmacophore with the appropriate linker (three lengths were tested), resulted in a 40-fold increase in affinity (from 146 μM to 4 μM for human serum albumin and from 1 μM to 25 nM for bovine carbonic anhydrase) (Fig. 13).

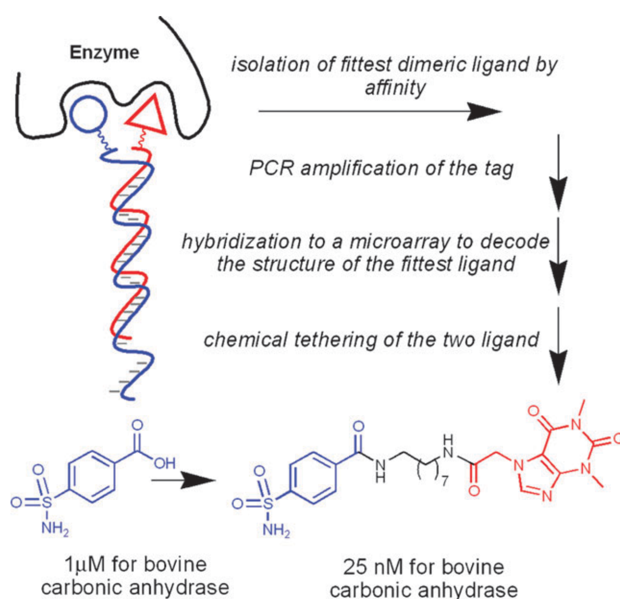


Fig. 12 Affinity maturation for small molecule by DNA-mediated dimerization with a small molecule library.

To control the conformation of a flexible peptide, Seitz and co-workers used a known ligand for the SH2 domain flanked by two PNA sequences. They demonstrated that they could tune the affinity of this peptide for its target by biasing its conformation through hybridization to nine different oligonucleotides bearing inserts from 0 to 9 nucleotides between the hybridization sites thus leading to a restricted conformational landscape for each complex.[28] It was shown that the conformationally fittest complex had at least 10-fold higher activity than the least fit complex. The authors noted that in this case, the use of PNA was deemed important as their cellular stability raises the possibility of controlling a ligand conformation in response to specific genomic information within a cell. Additionally, the conformation could be switched in a reversible manner by the addition of competing oligonucleotide strands.

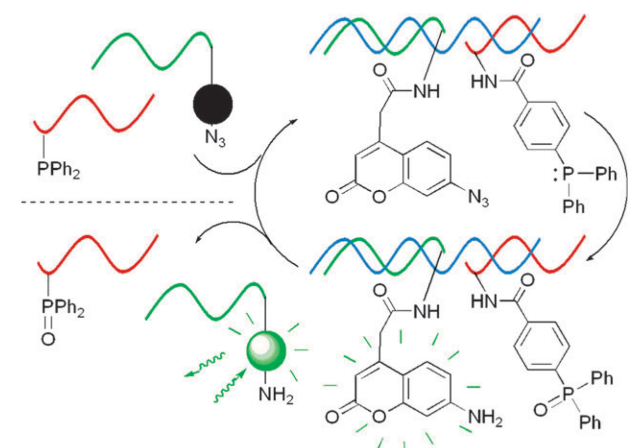


Fig. 11 DNA-templated reduction of azidocoumarin.

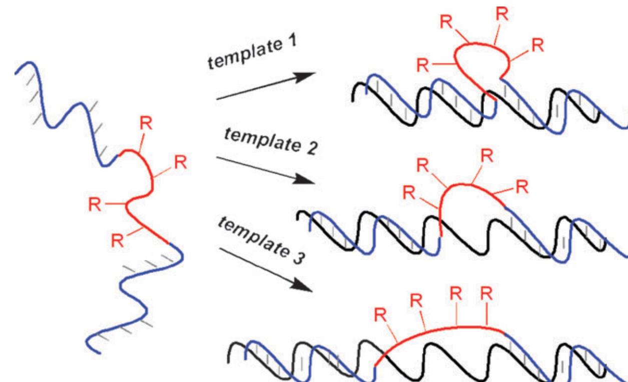


Fig. 13 Control of ligand conformation and topology by hybridization to different templates.

6 Outlook

It is safe to predict that nucleic acid-based programmed supramolecular assemblies to achieve new biological functions whether by controlling dimerization, conformation or the position of a ligand on a surface or nanostructure will continue to find new applications. PNAs are an attractive platform in this endeavour owing to their biological stability, flexible chemistry and high affinity. Furthermore, the ability of PNA to form triplexes or quadruplexes under some circumstances could be beneficial. While several important proofs of concept have been reported regarding templated reactions and ligand display, the ability to carry out these experiments in cells or whole organisms without recourse to membrane permeabilization techniques will open new horizons.

Abbreviations

Alloc	Allyloxycarbonyl
Boc	*tert*-Butyloxycarbonyl
Cbz	Benzyloxycarbonyl
Fmoc	9-Fluorenyloxymethyl carbonyl
PNA	Peptide Nucleic Acid
PCR	Polymerase chain reaction

Acknowledgements

We thank our collaborators who have contributed to our work in the area of nucleic-acid encoded projects and the funding agencies (Human Frontier Science Program - HFSP, Association pour la Recherche sur le Cancer - ARC, Agence Nationale de la Recherche - ANR and Marie Curie). This review was not intended to be comprehensive but rather a highlight of some recent work and we apologize for arbitrary omissions.

References

1. U. Feldkamp and C. M. Niemeyer, *Angew. Chem., Int. Ed.*, 2006, **45**, 1856–1876.
2. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, 1993, **365**, 566–568; P. E. Nielsen, *Acc. Chem. Res.*, 1999, **32**, 624–630.
3. P. E. Nielsen, *Mol. Biotechnol.*, 2004, **26**, 233–248.
4. J. Behr, *Chimia*, 1997, **51**, 34–36.
5. A. Dragulescu-Andrasi, S. Rapireddy, G. He, B. Bhattacharya, J. J. Hyldig-Nielsen, G. Zon and D. H. Ly, *J. Am. Chem. Soc.*, 2006, **128**, 16104–16112.
6. N. Winssinger, Z. Pianowski and F. Debaene, *Top. Curr. Chem.*, 2007, **278**, 311–342.
7. C. M. Niemeyer, L. Boldt, B. Ceyhan and D. Blohm, *Anal. Biochem.*, 1999, **268**, 54–63.
8. M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. S. Goody and C. M. Niemeyer, *Chem. Commun.*, 2003, 822–823.
9. R. C. Bailey, G. A. Kwong, C. G. Radu, O. N. Witte and J. R. Heath, *J. Am. Chem. Soc.*, 2007, **129**, 1959–1967.
10. J. L. Harris and N. Winssinger, *Chem.–Eur. J.*, 2005, **11**, 6792–6801.
11. 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 and N. Winssinger, *Chem. Biol.*, 2004, **11**, 1361–1372.
12. H. D. Urbina, F. Debaene, B. Jost, C. Bole-Feysot, D. E. Mason, P. Kuzmic, J. L. Harris and N. Winssinger, *ChemBioChem*, 2006, **7**, 1790–1797.
13. N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick and J. L. Harris, *Chem. Biol.*, 2004, **11**, 1351–1360.
14. N. Winssinger and J. L. Harris, *Exp. Rev. Proteom.*, 2005, **2**, 937–947; D. Pouchain, J. J. Diaz-Mochon, L. Bialy and M. Bradley, *ACS Chem. Biol.*, 2007, **2**, 810–818.
15. H. Schroeder, B. Ellinger, C. F. Becker, H. Waldmann and C. M. Niemeyer, *Angew. Chem., Int. Ed.*, 2007, **46**, 4180–4183.
16. L. E. Orgel, *Acc. Chem. Res.*, 1995, **28**, 109–118.
17. X. Li and D. R. Liu, *Angew. Chem., Int. Ed.*, 2004, **43**, 4848–4870.
18. A. P. Silverman and E. T. Kool, *Chem. Rev.*, 2006, **106**, 3775–3789.
19. B. N. Giepmans, S. R. Adams, M. H. Ellisman and R. Y. Tsien, *Science*, 2006, **312**, 217–224.
20. Z. Ma and J.-S. Taylor, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11159–11163.
21. J. Cai, X. Li, X. Yue and J. S. Taylor, *J. Am. Chem. Soc.*, 2004, **126**, 16324–16325.
22. Y. Xu, N. B. Karalkar and E. T. Kool, *Nat. Biotechnol.*, 2001, **19**, 148–152.
23. H. Abe and E. T. Kool, *J. Am. Chem. Soc.*, 2004, **126**, 13980–13986.
24. H. Abe and E. T. Kool, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 263–268.
25. (*a*) C. Dose, S. Ficht and O. Seitz, *Angew. Chem., Int. Ed.*, 2006, **45**, 5369–5373; (*b*) T. N. Grossmann and O. Seitz, *J. Am. Chem. Soc.*, 2006, **128**, 15596–15597.
26. Z. Pianowski and N. Winssinger, *Chem. Commun.*, 2007, 3820–3822.
27. S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat. Biotechnol.*, 2004, **22**, 568–574.
28. L. Roeglin, M. R. Ahmadian and O. Seitz, *Angew. Chem., Int. Ed.*, 2007, **46**, 2704–2707.

 This journal is © The Royal Society of Chemistry 2008