

UNIVERSITE DE STRASBOURG
INSTITUT DE SCIENCES ET D'INGENIERIE
SUPRAMOLECULAIRES

***New Supramolecular and Fluorogenic Chemo-
Sensors: Proof-of-concept and Biological
Applications***

Kamel Meguellati

Une thèse soumise en vue de l'obtention du titre de Docteur en Sciences Chimiques de l'Université de
Strasbourg

Défendue publiquement le 24 Septembre 2010 en présence de la commission d'examen suivante:

Prof. Nicolas GIUSEPPONE
Dr. Marie-Paule TEULADE-FICHO
Prof. Kay SEVERIN
Prof. Olivia REINAUD
Prof. Jean-Marie LEHN
Dr Sylvain LADAME

Président du jury
Rapporteur externe
Rapporteur externe
Examineur externe
Co-Directeur de thèse
Directeur de thèse

UNIVERSITE DE STRASBOURG
INSTITUT DE SCIENCES ET D'INGENIERIE
SUPRAMOLECULAIRES

***Nouvelles Sondes Fluorogéniques
Supramoléculaires : Preuves de Concepts et
Applications Biologiques***

Kamel Meguellati

Une thèse soumise en vue de l'obtention du titre de Docteur en Sciences Chimiques de l'Université de
Strasbourg

Défendue publiquement le 24 Septembre 2010 en présence de la commission d'examen suivante:

Prof. Nicolas GIUSEPPONE
Dr. Marie-Paule TEULADE-FICHO
Prof. Kay SEVERIN
Prof. Olivia REINAUD
Prof. Jean-Marie LEHN
Dr Sylvain LADAME

Président du jury
Rapporteur externe
Rapporteur externe
Examineur externe
Co-Directeur de thèse
Directeur de thèse



TABLE OF CONTENTS

ABSTRACT

INTRODUCTION

<u>Introduction to fluorescence</u>	2
<i>Absorption</i>	2
<i>Fluorescence</i>	4
<u>Fluorophores</u>	7
<u><i>Fluorophores with emission wavelengths below 500 nm</i></u>	8
<i>Coumarins</i>	8
<i>Oligothiophenes</i>	9
<i>Naphthalenes</i>	10
<i>Dansyls</i>	11
<u><i>Fluorophores with emission wavelengths above 500 nm</i></u>	12
<i>Fluoresceins</i>	12
<i>Rhodamines</i>	14
<i>BODIPYs</i>	15
<i>Squaraines</i>	16
<u><i>Introduction to cyanine dyes</i></u>	16
<i>Generalities about cyanine dyes</i>	16
<i>Cyanine dyes synthesis</i>	20
<i>Cyanine dyes aggregation properties</i>	27
<i>Cyanine dyes biological applications</i>	29
<i>Cyanine dyes and surfactants</i>	34
<i>Stability of cyanine dyes</i>	36

CHAPTER 1: Sensing and targeting nucleic acid structures and sequences

<u>DNA: Structure and function</u>	42
<i>Watson-Crick base-pairing</i>	44
<i>DNA double helical conformation</i>	46
<i>From DNA to RNA to proteins</i>	47
<i>Alternative DNA secondary structures are also biologically relevant</i>	48
<u>Introduction to DNA hairpins and quadruplexes</u>	50
<i>Biological relevance of DNA hairpins</i>	50
<i>Biological relevance of DNA quadruplexes</i>	52
<u>DNA quadruplexes: Structures and Functions</u>	52
<i>Structure of DNA quadruplexes</i>	52
<i>Quadruplexes in the genome</i>	54
<i>Quadruplex topologies</i>	55
<i>Quadruplex sensors</i>	57
<u>DNA hairpins: Structures and Functions</u>	59
<u>Introduction to Peptide Nucleic Acids (PNAs)</u>	61
<u>Use of PNAs and DNAs in Oligonucleotide-Templated Reactions</u>	63
<u>Publications</u>	69

CHAPTER 2: Dynamic Combinatorial Chemistry and Cyanine dyes

<u>Principles of Dynamic Combinatorial Chemistry</u>	113
<u>Reversible reactions suitable for DCC</u>	116
<u>Publications</u>	123

CHAPTER 3: Improving the prediction of maximal absorption wavelengths of cyanine dyes: a computational study using TD-DFT

<u>Introduction to TD-DFT</u>	168
<u>Using TD-DFT to predict the max. absorption wavelength of cyanine dyes</u>	172

CHAPTER 4: Reversible formation of imino cyanine dyes in water-in-oil droplets

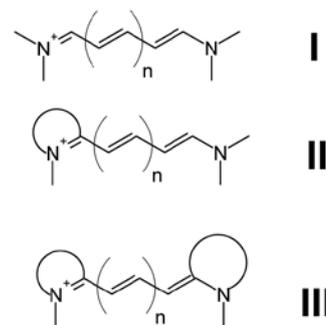
<u>Introduction to microfluidics</u>	198
<i>Basics principles of microfluidics</i>	199
<i>Devices design and preparation</i>	199
<i>Droplet production</i>	202
<u>Chemical Reactions in droplets</u>	203
<u>Publication</u>	204
CONCLUSION & PERSPECTIVES	225

Nouvelles Sondes Fluorogéniques Supramoléculaires : Preuves de Concepts et Applications Biologiques

Résumé en Français

Il existe actuellement un intérêt croissant pour la conception de nouvelles sondes chimiques basées sur des systèmes supramoléculaires qui ont la capacité de reconnaître un métabolite cible avec une grande affinité et spécificité. En réponse à cette interaction, le système subit une rééquilibration se traduisant par un changement des propriétés physiques du milieu. Parmi les sondes chimiques les plus intéressantes, les sondes fluorescentes pour lesquelles le phénomène de reconnaissance moléculaire du métabolite par le capteur s'accompagne d'un changement des propriétés de fluorescence de l'émetteur (longueurs d'onde d'excitation et d'émission) présentent l'avantage d'une très grande sensibilité et d'une grande simplicité de détection du signal. Nous proposons ici de développer un nouveau type de sonde moléculaire fluorescente inspirée des fluorophores de type cyanine dyes couramment utilisés en photographie, en stockage de données optiques ou encore en protéomique. La stratégie générale repose sur la capacité pour deux entités non-fluorescentes de réagir l'une avec l'autre de façon covalente (réversible ou irréversible) pour donner une molécule de type cyanine (ou analogue) fluorescente. Le système est conçu de telle sorte que la formation de la cyanine est thermodynamiquement fortement défavorisée mais peut être fortement amplifiée en présence d'un métabolite capable d'induire le rapprochement des deux entités réactives. Dans le cadre de cette thèse, nous nous sommes ainsi intéressés à la mise en évidence et à l'étude de nouvelles réactions fluorogéniques biocompatibles puis à leurs applications pour la détection de structures ou séquences d'acides nucléiques (ADN) *in vitro*.

Les cyanines sont des sels organiques possédant des coefficients d'extinction molaires élevés et des rendements quantiques de fluorescence modérés, et qui demeurent stables à l'air et à des températures élevées. Ces cyanines sont généralement constituées de deux monomères azotés (cycliques ou non) liés l'un à l'autre par une chaîne polyméthine dont la longueur détermine les longueurs d'onde d'absorption et d'émission, ces dernières augmentant d'environ 100 nm pour chaque double liaison ajoutée. Suivant la longueur de la chaîne polyméthine on parlera de cyanine monométhine (ou Cy1, $n=0$), triméthine (ou Cy3, $n=1$) ou encore pentaméthine (ou Cy5, $n=2$).

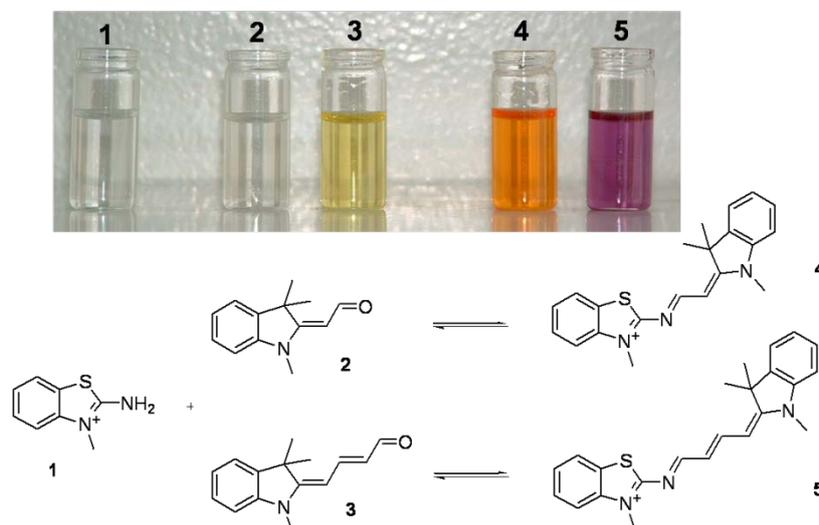


1) Développement de Sondes Dynamiques Fluorescentes.

a) Analogues mono-imines de cyanines dyes.

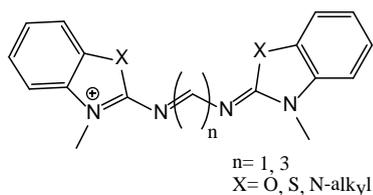
Dans un premier Chapitre, nous décrivons la conception d'un système dynamique basé sur une réaction réversible et fluorogénique de formation d'imine. Deux réactifs hétérocycliques de type amine et aldéhyde non-fluorescents vont réagir de façon covalente et réversible pour conduire à la formation d'une imine fluorescente. Pour cela, nous nous sommes inspirés des composés de la famille des cyanines. Nous avons démontré ici que le remplacement, dans la chaîne polyméthine d'une cyanine, d'un atome de carbone par un atome d'azote pouvait être obtenu par réaction entre un hétérocycle de type N-méthyl-2-

amino-benzothiazolium et un aldéhyde de base de Fischer. Nous avons montré que cette réaction était *réversible* et sous contrôle thermodynamique et que les analogues imines de cyanines présentaient des propriétés spectroscopiques (UV, fluorescence) comparables à celles de leurs analogues « tout-carbone », bien qu'absorbant et émettant à des longueurs d'onde significativement plus basses. Particulièrement intéressant est le dérivé imine 4 (schéma ci-dessous), analogue fluorescent d'une Cy3 obtenu de façon réversible par réaction entre deux monomères 1 et 2 incolores et non-fluorescents.

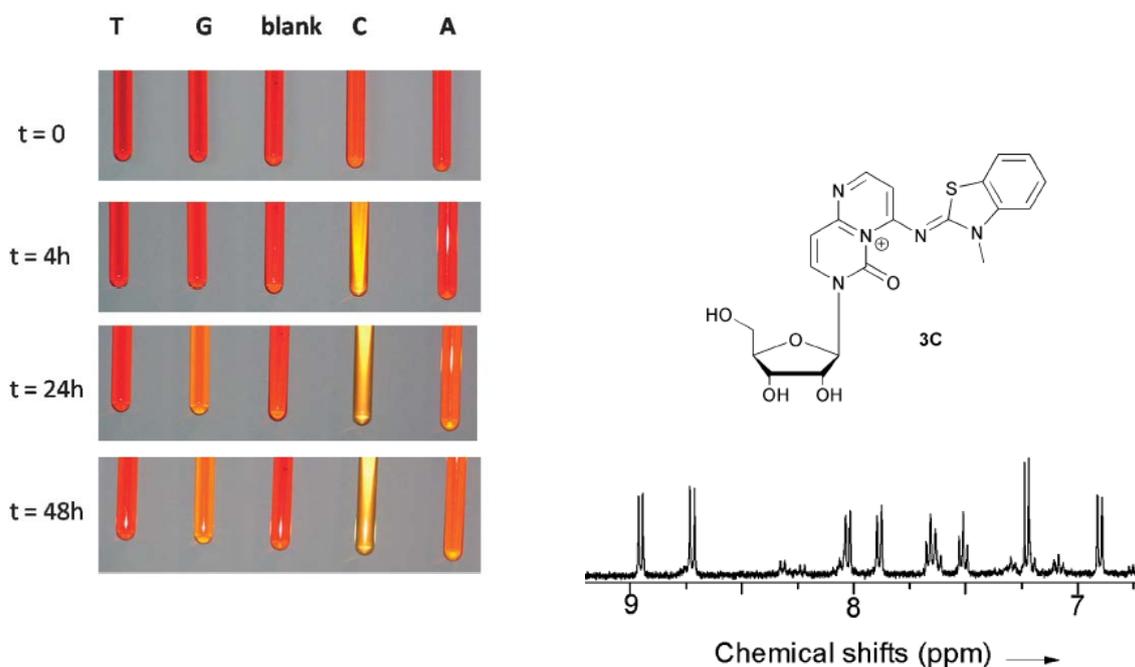


b) Analogues bis-imines de cyanines dyes.

Dans un second temps, nous nous sommes intéressés à la possible formation d'analogues « bis-imine » de cyanines et à leurs applications en tant que sondes fluorescentes ou agents alkylants d'ADN/ARN. Des cyanines symétriques bis-imines ont notamment été obtenues par réaction de deux monomères N-méthyl-2-amino-benzothiazolium avec un dialdéhyde connu comme intermédiaire du stress oxydant et comme puissant alkylant d'ADN, le malondialdéhyde (MDA). Une petite banque de dérivés bis-imines a été synthétisée en variant la nature de l'hétérocycle (benzothiazole, naphthothiazole, quinoline) et les propriétés spectroscopiques de l'ensemble de ces composés ont été caractérisées. L'introduction d'un second atome d'azote dans la chaîne polyméthine entraîne de nouveau un abaissement d'environ 70 nm des longueurs d'ondes maximales d'absorption et d'émission par rapport aux composés mono-imines (Section 1a). De plus ces composés sont fortement instables et présentent une durée de demi-vie d'environ 10-12 heures.



Néanmoins, particulièrement intéressant est le potentiel de ces adduits fluorescents de MDA en tant que pro-drogues pour l'alkylation des bases nucléiques de l'ADN. Nous avons notamment démontré que ces imino-cyanines pouvaient alkyler sélectivement les cytosines. Parmi les multiples adduits covalents formés par réaction des bis-imines avec la cytidine, un des adduits majoritaires a tout particulièrement attiré notre attention en raison de ses propriétés fluorescentes et semblerait correspondre au produit d'une bis-alkylation sur le noyau cytosine. (Schéma ci-dessous).



c) Modèle de calcul pour la prédiction des λ_{max} d'absorption des cyanines.

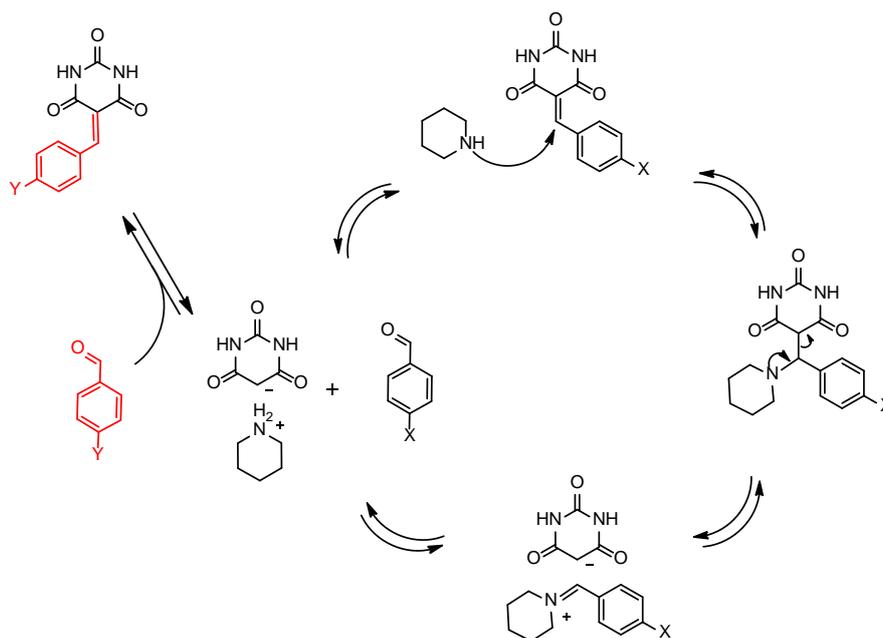
Les cyanines sont très couramment utilisées comme sondes ou marqueurs fluorescents pour détecter ou suivre la présence de biomacromolécules *in vitro* et *in vivo*. Nos cyanines réversibles ouvrent également la voie à de nouvelles applications biologiques, par exemple pour détecter la présence de MDA en milieux biologiques. En raison des multiples applications que ces molécules peuvent avoir, il serait important de disposer d'une méthode de calcul fiable permettant de prédire avec exactitude les propriétés d'absorption des cyanines (traditionnelles ou modifiées) avant de les synthétiser. Ceci a été réalisé dans le cadre d'une collaboration avec le Dr Martin Spichty et le Pr Martin Karplus en utilisant la « time-dependent density functional theory » (TD-DFT).

d) Premier exemple de réaction de Knoevenagel réversible.

Des barbiturates sont issus de la condensation d'un aldéhyde avec l'acide barbiturique ou un de ses dérivés par une réaction de KNOEVENAGEL dans le chloroforme en présence de base, dans l'eau ou dans le méthanol, etc. La synthèse des barbiturates peut être réalisée en présence d'une quantité catalytique dans un solvant organique. Ainsi une retro-KNOVENAGEL va consister en la dégradation du barbiturate en un aldéhyde et le sel de l'acide barbiturique. Ainsi, en présence d'un aldéhyde différent de celui utilisé pour la

synthèse du barbiturate, il est possible d'échanger la partie barbiturique vers l'aldéhyde cible comme indiqué sur la figure ci-dessous.

Cet échange a été démontré dans le DMSO à 80 °C, cependant l'échange est optimal en utilisant une quantité catalytique de pipéridine. La durée d'atteinte d'équilibre de l'échange dépend de la nature du substituant X et Y : la réaction étant sous contrôle thermodynamique l'échange se fait rapidement dans le sens d'obtention du produit thermodynamiquement le plus stable à savoir dans le sens où X est un groupement électro attracteur et Y un groupement électro donneur. Le but final sera d'effectuer de la chimie combinatoire dynamique asymétrique en utilisant de la L-proline pour faire des réactions d'échange et assurer une régio-sélectivité de l'échange.



2) Synthèse de Peptide Nucleic Acids (PNAs) fluorogéniques et applications pour la détection spécifique de structures et/ou séquences d'ADN in vitro.

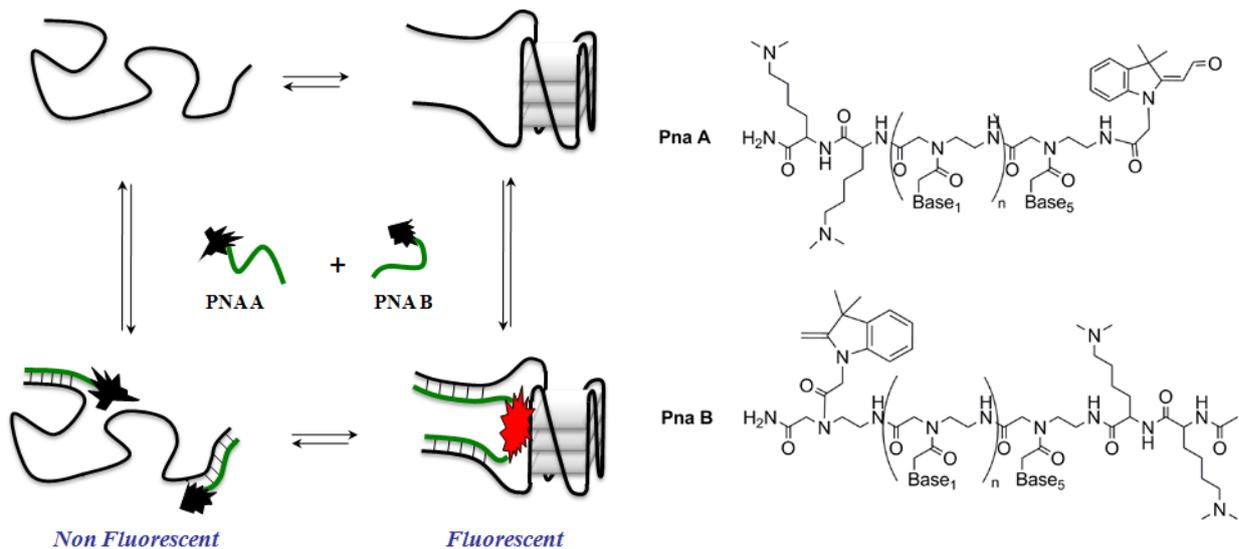
Les PNAs sont des analogues d'ADN dans lesquels le squelette ribose-phosphate a été remplacé par une chaîne peptidique neutre. Cette chaîne peptidique confère aux PNAs une plus grande stabilité chimique que l'ADN tout en conservant ses propriétés d'hybridation (avec un brin d'ADN -ou un autre brin de PNA- complémentaire). Un brin de PNA formera donc avec un brin d'ADN complémentaire un duplex hybride PNA:ADN particulièrement stable (et même plus stable que le duplex ADN:ADN correspondant).

Dans le cadre de ce projet, nous avons choisi de fonctionnaliser deux PNAs avec des précurseurs non-fluorescents d'une Cy3 : un aldéhyde de base de Fischer et un monomère 2-méthylène indolénine et de les utiliser comme sondes fluorescentes pour la détection de structures secondaires et de séquences d'ADN.

a) Détection de quadruplexes de guanines in vitro.

Il a été établi depuis plusieurs dizaines d'années que les brins d'acides nucléiques riches en guanines ont la capacité de former des structures extrêmement stables à quatre brins in vitro en présence de cations physiologiques monovalents (en particulier K^+ et Na^+). Il a également été démontré indirectement que ces structures communément appelées quadruplexes de guanines (ou G-quadruplexes ou G4) possédaient diverses fonctions biologiques et pouvaient être exploitées comme cible pour la conception de ligands possédant une activité anti-cancéreuse. Récemment, de nombreuses séquences capables de former des quadruplexes de guanines ont été identifiées dans le génome humain et en particulier dans la région des promoteurs d'oncogènes. Il a été proposé que ces structures pourraient agir comme éléments régulateurs de l'expression de ces proto-oncogènes. Parmi les gènes présentant dans leur promoteur une séquence pouvant adopter une structure de type G-quadruplex, l'oncogène c-kit a reçu une attention toute particulière. Toutefois, toutes les preuves suggérant la formation de quadruplexes d'ADN in vivo ne sont qu'indirectes et pour cette raison, l'existence de quadruplexes in vivo est encore sujet à débats.

Pour ce projet, nous avons exploité le fait que la formation au milieu d'un simple brin d'ADN d'une structure de type G4 s'accompagne d'une « contraction » de cet ADN. La stratégie générale de sonde dirigée contre les G4 est ainsi résumée sur le schéma ci-dessous.

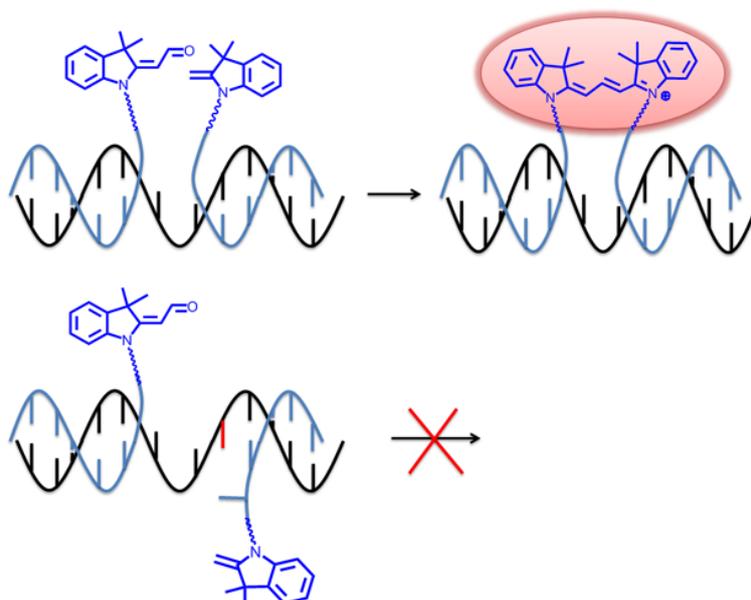


En utilisant des séquences de PNAs complémentaires aux brins d'ADN situés en amont et en aval du quadruplex d'ADN choisi comme cible (**PNA A** et **PNA B**, schéma ci-dessus), nous avons démontré que les groupements indolénine et aldéhyde étaient suffisamment proches pour réagir l'un avec l'autre (et former la cyanine fluorescente) uniquement quand le quadruplex est formé, entraînant ainsi l'émission d'un signal de fluorescence caractéristique. Cette approche originale a été appliquée avec succès à la détection in vitro d'un quadruplex intramoléculaire situé dans la région du promoteur de l'oncogène c-kit.

Nous avons démontré que ce système était également adapté à la détection d'autres structures secondaires d'ADN telles que les tiges boucles (structures obtenues par un repliement partiel d'un brin d'ADN par complémentarité des bases).

b) Détection de Single Nucleotide Polymorphism (SNP) in vitro.

Une stratégie similaire a été appliquée pour la détection de SNPs in vitro. Deux PNAs fonctionnalisés par les deux groupements fluorogéniques (indolenine et aldéhyde) ont été conçus de telle sorte qu'ils puissent s'hybrider de façon séquence-complémentaire à un même brin d'ADN. Suite à cette hybridation simultanée des deux PNAs à leur matrice ADN, les deux groupements fluorogéniques vont être localisés à proximité l'un de l'autre, permettant ainsi la formation de la cyanine fluorescente. En cas de présence de mutations sur le brin d'ADN, un des deux PNA ne pourra pas s'hybrider de façon optimale, entraînant une inhibition de la réaction de formation de cyanine. Ce système est détaillé sur le schéma ci-dessous.

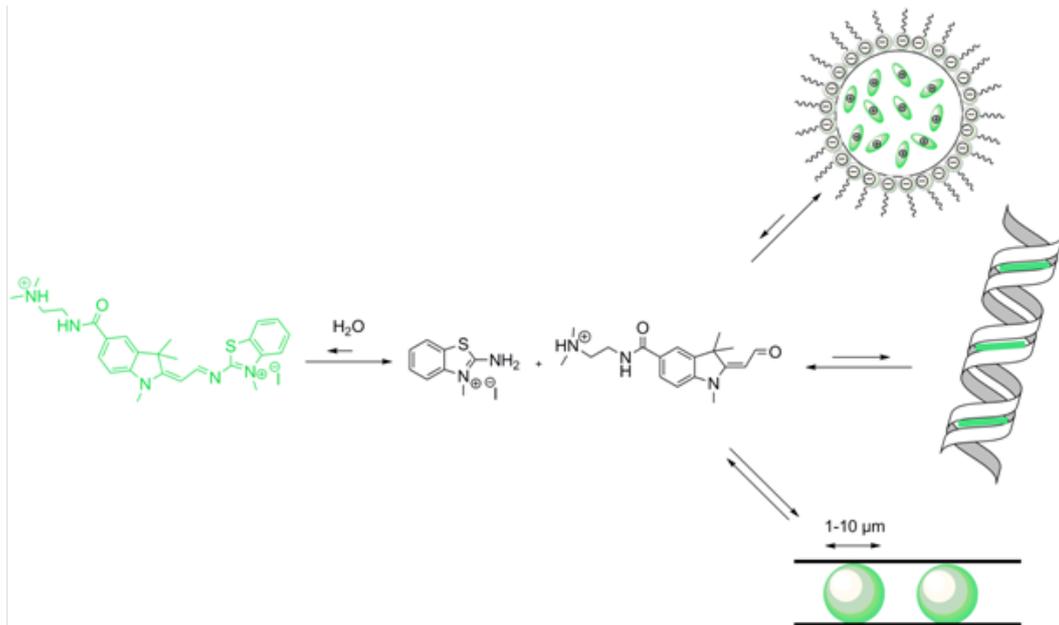


3) Etude d'une catalyse à l'interface au sein de microémulsions

La microfluidique est la science qui étudie l'écoulement des fluides dans des dispositifs. Le but est de produire une émulsion et contrôler la formation des gouttes. Les gouttes formées sont stabilisées par des surfactants. La nature du surfactant peut avoir une influence sur la taille des gouttes ainsi que sur leur polydispersité (ou non). Le surfactant peut également influencer une réaction chimique si celle-ci se déroule en surface.

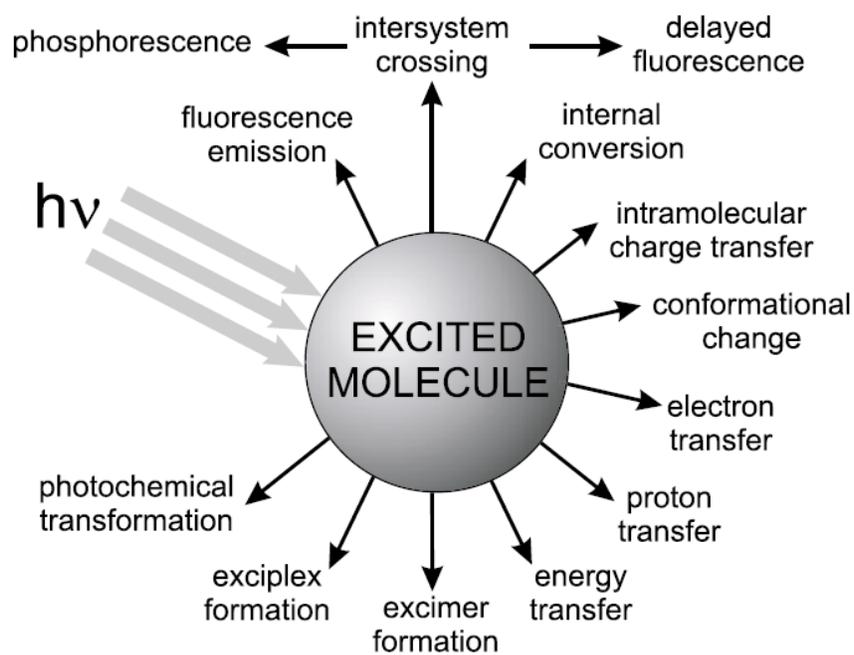
Dans le cadre de ce projet réalisé en collaboration avec le Pr A. Griffiths et le Dr A. El Harrak, nous avons utilisé notre réaction réversible et fluorogénique de formation de cyanine imine (Section 1a) pour étudier les effets de catalyse et/ou d'inhibition au sein de microémulsions. Nous avons ainsi successivement étudié les effets de compartimentalisation et de surfactant sur l'efficacité de la réaction de formation d'imine en milieu aqueux. Deux monomères amine et aldéhyde hydrosolubles ont été synthétisés (Schéma ci-dessous). De fait de leur très faible réactivité, et en l'absence de toute catalyse, la réaction de formation d'imine dans l'eau est très fortement déplacée vers les deux réactifs. Toutefois, nous avons réussi à démontrer qu'il était possible de déplacer significativement cet équilibre vers la formation de l'imine fluorescente en séquestrant ces deux réactifs dans des gouttelettes dont le volume n'excède pas quelques picolitres. Ce déplacement de la position d'équilibre s'accompagne également d'une forte accélération de la cinétique de la réaction. Enfin, ce déplacement de la

position d'équilibre est d'autant plus important que la taille de la goutte est petite, suggérant un effet de catalyse à l'interface de la goutte. Cette catalyse à l'interface a pu être confirmée en étudiant l'effet sur cette réaction de différents surfactants aux propriétés variables : ZONYL (chaîne perfluorée non chargé), SDS (anionique), CTAB (chargé positivement), PAA (anionique également), pluronic (partie éthylène glycol non perfluorée). Il est apparu que seul le SDS catalysait fortement la formation d'imine quand sa concentration est supérieure à sa concentration micellaire critique (CMC).



Ce phénomène de catalyse a été observée dans des gouttes c'est-à-dire dans des systèmes auto-organisés comme indiqués sur le schéma ci-dessus. La réaction est également accélérée par intercalation du produit dans les petits sillons d'ADN.

INTRODUCTION



I. Introduction to fluorescence

I.1 Absorption

Upon absorption, one electron reaches an upper vacant orbital of higher energy. Thus, the molecule is in an excited state. Transition from ground to upper excited state is accompanied by a reorganization of an electronic cloud within the molecular orbitals and this is the condition for a transition to occur.

In reference of the Franck–Condon principle, electronic transitions are so fast that they arrive without nuclei rearrangement: nuclei do not move during electronic transition. That is why electronic transitions are always drawn as vertical lines. The energy of a pair of atoms as a function of the distance is characterized by the Morse Curve as in figure 1 below:

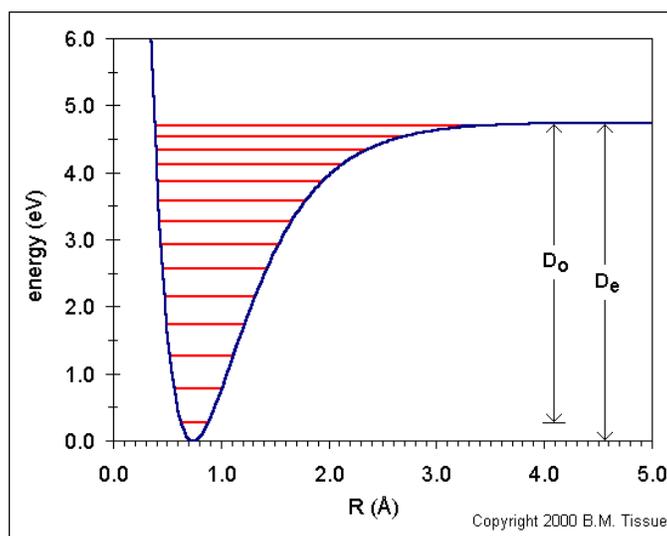


Figure 1: Representation of the energy depending of the distance between 2 atoms extracted from ref ¹.

R_e is called the equilibrium bond distance. At this distance, the molecule is in its most stable position and so its energy is called the molecular equilibrium energy, which is expressed as E_0 or E_e . Stretching or compressing the bond induces an energy increase. On the left-hand side of R_e , the two atoms become increasingly closer, inducing

repulsion forces. Thus, an energy increase will be observed as a consequence of these repulsion forces. On the right-hand side of R_e , the distance between the two atoms increases, and there will be attraction forces so that an equilibrium distance can be reached.

The attraction forces leads to an enhancement of the energy. The energy within a molecule is the sum of several distinct energies:

$$E(\text{total}) = E(\text{translation}) + E(\text{rotation}) + E(\text{vibration}) + E(\text{electronic}) + E(\text{electronic orientation of spin}) + E(\text{nuclear orientation of spin})$$

The rotational energy concerns molecular rotation around the gravity center, vibrational energy is the result of periodic displacement of atoms of the molecule away from the equilibrium position, and electronic energy is generated by electron movement within the molecular bonds. Rotational levels have lower energy than vibrational ones. Molecules which absorb photons as described in the Morse Curve are called chromophores.

Different compounds may have very different absorption maxima and molar absorptivities. Intensely absorbing compounds must be examined in dilute solution, so that significant light energy is received by the detector, and this requires the use of completely non-absorbing solvents (most commonly water, ethanol and hexane). Because the absorbance of a sample is proportional to its molar concentration in the sample cuvette, a corrected absorption value known as the molar absorptivity is used when comparing the spectra of different compounds. This is defined as:

$$\epsilon = \frac{A}{c l}$$

where ϵ is called Molar Absorptivity, A refers to the absorbance, c to the sample concentration in moles/liter and l to the length of light path through the cuvette in cm.

1.2 Fluorescence

A chromophore reaches an excited state after absorption of a photon. A fluorophore is a chromophore which emits light. Nevertheless many chromophores do not necessarily fluoresce. Figure 2 represents the electronic transitions diagram, also called Jablonski diagram.

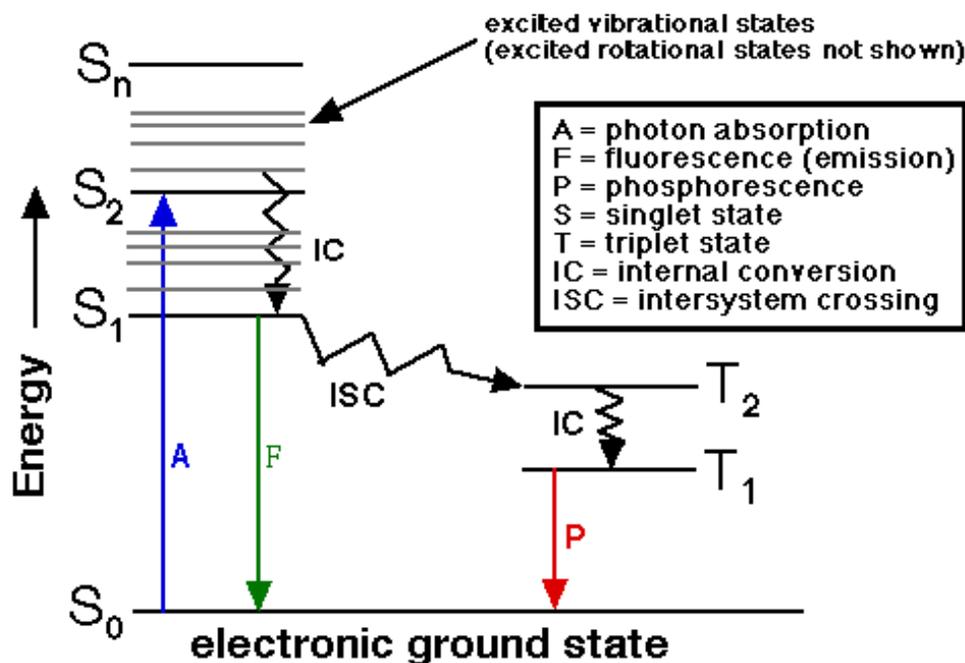


Figure 2: Representation of the Jablonski diagram extracted from ref².

The terminologies singlet and triplet state are the result of the spin quantum number s . Absorption of photons by a population of molecules induces the transition of electrons from the singlet ground electronic level S_0 to an excited state S_n ($n > 1$). An excited molecule will return to the ground state S_0 following two successive steps:

1. The molecule at S_n returns to the lowest excited state S_1 by dissipating a part of its energy in the surrounding environment. This phenomenon is usually called internal conversion (IC).

2. From the excited state S_1 , the molecule will reach the ground state S_0 via different competitive processes:
- Emission of a photon (fluorescence)
 - Part of the absorbed energy is dissipated in the medium as non radiative heat
 - Excited molecules can release some of their energy to molecules located nearby (fluorescence quenching)
 - A transient passage occurs to the excited triplet state T_1 of energy lower than S_1 called inter-system crossing which is responsible of the phosphorescence emission. For each excited state S , there is an excited state T of unstable lower energy.

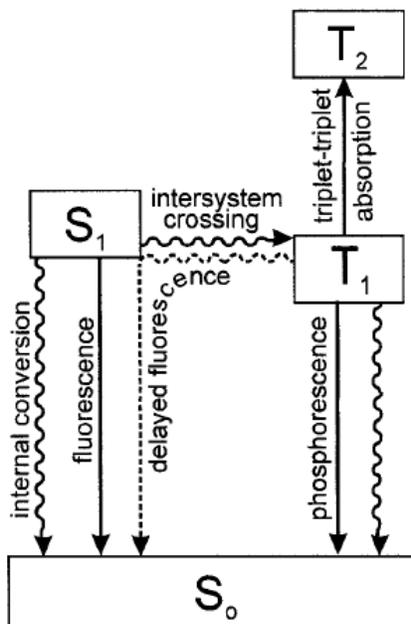


Figure 3: Difference between fluorescence and phosphorescence: the transition from the singlet state S_1 to S_0 is responsible of the fluorescence and the transition T_1 to S_0 is responsible of the phosphorescence. Adapted from ref³.

Emission occurs from the excited state S_1 , independently of the excitation wavelength. Therefore, the emission energy would be independent of the excitation wavelength. The fluorescence energy is higher than that of phosphorescence. The energy

of electronic transition is equal to the energy difference between the starting energy level and the final level. Therefore, the transition energy E is described by the Broglie law:

$E = \frac{hc}{\lambda}$ where h is the Planck constant ($h = 6.63 \times 10^{-34} \text{ J} \cdot \text{s}$), ν is the light frequency (s^{-1} or Hertz, Hz), c is the light velocity, and λ is the wavelength (nm).

Each transition occurs with a specific energy and so at a specific and single wavelength. However spectra (and no single lines) are observed, as a result of the contributions of rotational and vibrational levels to absorption and de-excitation energy. The fluorescence spectrum generated shows a maximum corresponding to the emission transition. The absorption spectrum characterizes the electronic distribution in this state because it appears from the ground state. Fluorescence and phosphorescence occur from excited states so they are the mirrors of electronic distribution within the excited states, S for fluorescence and T for phosphorescence. The emission lifetime is within the picosecond-to-nanosecond range. Thus, emission is a very fast process, and so in order to observe fluorescence emission, the fluorophore should be excited continuously.

The energy absorbed by a fluorophore is more important than the energy of an emitted photon. The absorption energy is $E_a = hc/\lambda_a$ and emission energy is $E_{em} = hc/\lambda_{em}$. Since $E_{em} < E_a$ we have $\lambda_{em} > \lambda_a$ where λ_a and λ_{em} are absorption and emission spectra peaks. Thus, the emission spectrum has its maximum shifted to longer wavelengths compared to the maximum of absorption spectrum. Commonly, the fluorescence spectrum is shifted with respect to the excitation spectrum to longer wavelengths and this shift is called the Stokes shift in memory of George Stokes who observed this phenomenon in 1852 (see figure 4).

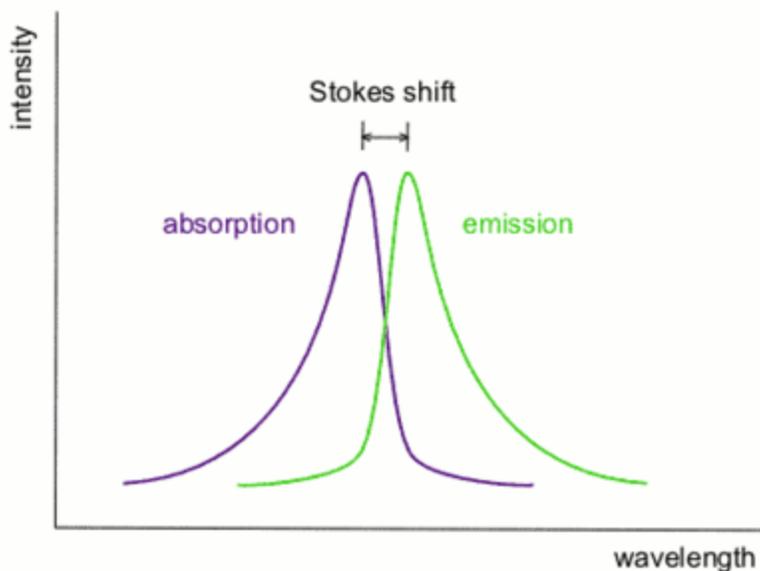


Figure 4: Stokes shift is defined as the difference between the wavelengths in fluorescence emission and absorption emission. Picture extracted from ref.⁴

The highest fluorescence intensity will be observed if the excitation and emission are provided at the wavelengths of the correspondent band maxima. In many cases it is relatively easy to provide the coupling of this enhancement/quenching response with a sensing event. Intensity change is the reflection of change of the fundamental parameter of emission, the quantum yield. The fluorescence quantum yield, Φ , is the ratio between the number of quanta emitted and the number of all absorbed quanta. Fluorescence intensity at any wavelength $F(\lambda)$ is proportional to the fluorescence quantum yield Φ , i.e., a high quantum yield leads automatically to a high fluorescence intensity.

In principle, Excitation fluorescence is the equivalent of absorption because the fluorescence excitation spectrum characterizes the electron distribution of the molecule in the ground state.

II. Fluorophores.

A large number of natural and synthetic fluorescent molecules are now available that cover a very broad spectral range. They have found valuable applications in modern bioresearch as labels or probes for sensing or tagging non-fluorescent molecules. Indeed,

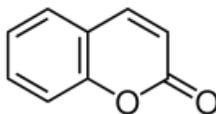
organic fluorophores can be covalently or non-covalently attached to the target sample to be analyzed (or sensed) in order to produce conjugates or complexes that emit fluorescence from short to very long wavelengths, depending on the marker used. More recently, the development of fluorophores absorbing and emitting at very long wavelengths (in the near-InfraRed region of the spectrum) proved highly valuable for biological applications.

Below, a non-exhaustive list of fluorophores (with a special emphasis for fluorophores used for biological applications) is reported. Compounds emitting fluorescence from the near-ultraviolet to the blue area of the spectrum are first presented. They include oxobenzopyrans, naphthofurans, oligothiophenes, dansyls, naphthalenes. Then selected compounds emitting in the blue to near-infrared (ca. near 900 nm) are also presented, which include fluoresceins (including biarsenical dyes), rhodamines, 4,4-difluoro-4-bora-3 α ,4 α -diazas-indacenes (BODIPY dyes), squaraines, and cyanine dyes.

II.1 Fluorophores with emission wavelengths below 500 nm

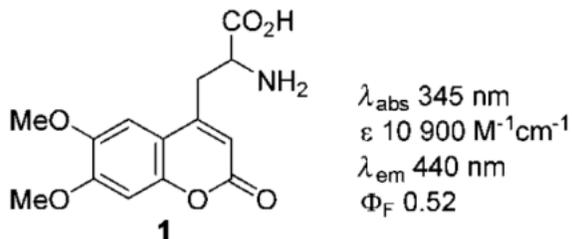
II.1-1 Coumarins

Coumarins, properly named 3-Oxo-3*H*-benzopyrans, are among the most famous families of fluorescent compounds⁵. Their general structure is depicted below:



Coumarin has appetite-suppressing properties, suggesting that its widespread occurrence in plants (especially grass) is because of its effect on reducing the impact of grazing animals. It has clinical medical value as the precursor for several anticoagulants, notably warfarin, and is used as a gain medium in some dye lasers. Modified fluorogenic amino acids with oxobenzopyrans appear to be interesting molecules due to their extended spectral range, high emission quantum yields, photostability, and good solubility in

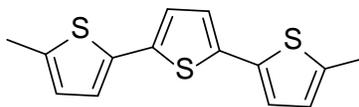
several solvents. For example, 2-Amino-3-(6,7-dimethoxy-3-oxo-3H-benzopyran) propanoic acid (Dmca, (compound **1** below) shows maximum absorption and emission wavelengths at 345 and 440 nm, respectively⁶.



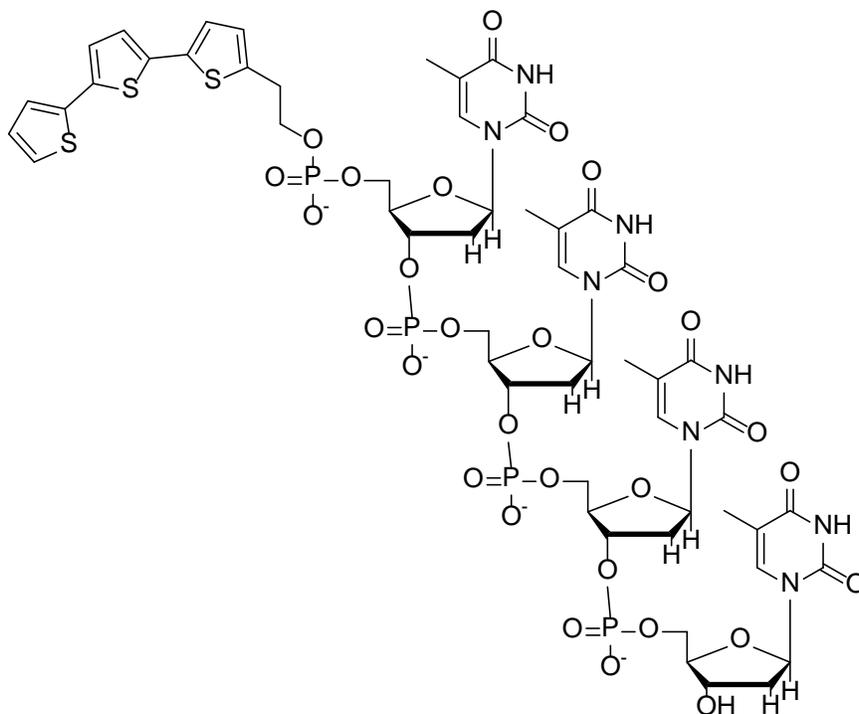
These properties enable selective determination of Dmca (1)-labeled peptides even when amino acid tryptophan (Trp, W) residues are present. Furthermore, compound **1** possesses a high fluorescent quantum yield (Φ_{F}) and large molar absorptivity (ϵ) (Φ_{F} = 0.52 and ϵ = 10 900 M⁻¹ cm⁻¹), which facilitate detection of the labeled peptides on a picomolar scale, with a sensitivity similar to that of radiolabeling.

II.1-2 Oligothiophenes

Molecules in which two or more thiophene rings are linked together are called oligothiophenes⁷.



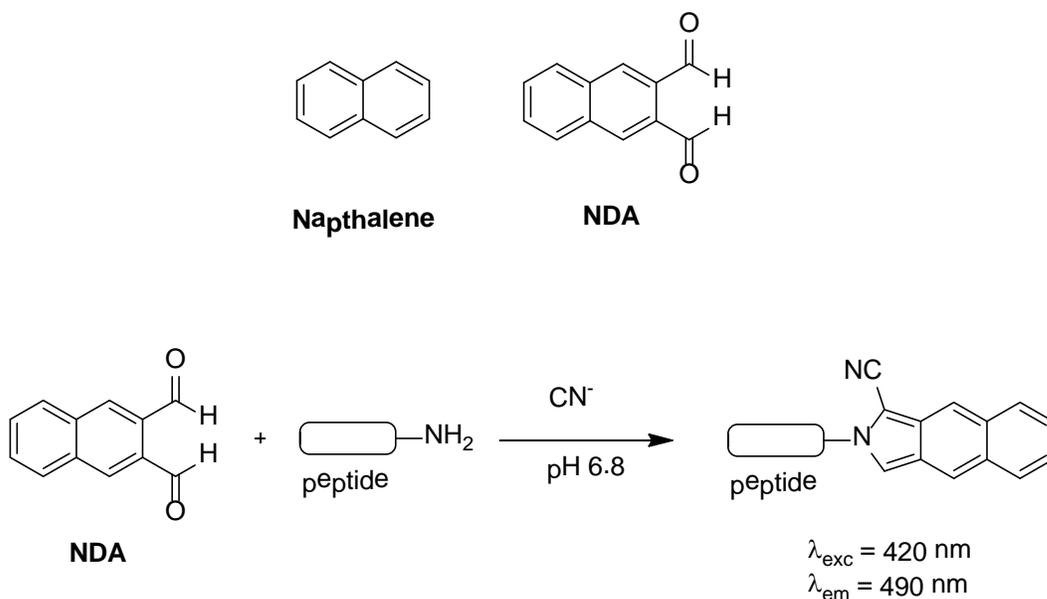
Such compounds possess interesting optical and electronic properties like: fluorescence, semi-conductance, and light emission if correctly stimulated. Thiophene oligomers are stable molecules and their optoelectronic properties can be reasonably well predicted. Their fluorescence properties can be also easily altered and tuned⁸. These oligoheteroaromatic compounds are often used as biomarkers. For instance, their suitability for labeling DNA was proven by Capobianco *et al.*⁹ who used terthiophene-T4 conjugate and coupled it to DNA as described in the scheme below.



The fluorescent tag is added to the DNA via a phosphoramidite coupling. The resulting conjugate (Figure above) showed absorption maxima at 265 and 360 nm and an emission maximum at 454 nm in water. The intensity of fluorescence of this conjugate allows for its easy detection at submicromolar concentrations such as those usually required in hybridization or microscopy studies.

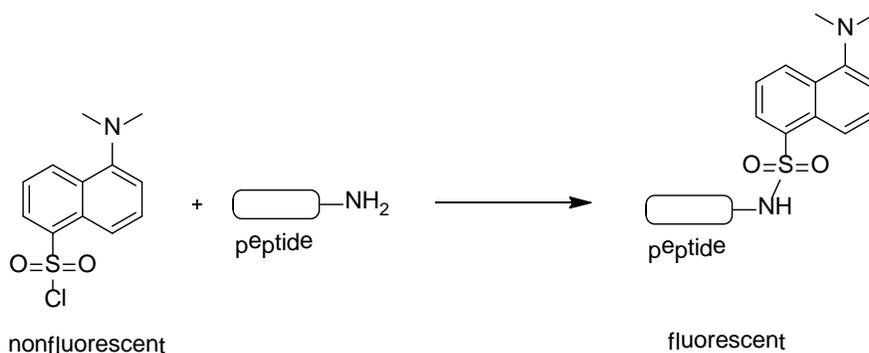
II.1-3 Naphthalenes

A naphthalene molecule is derived by the fusion of a pair of benzene rings. Naphthalene-based fluorophores such as the Naphthalene-2,3-dicarboxaldehyde (NDA, see scheme below for structure and reactivity towards amines) have been developed for labeling amino-acids, peptides, proteins, thus allowing their detection in HPLC and electrophoresis by fluorescence spectroscopy¹⁰.



II.1-4 Dansyls

Since its development by Weber in 1951, dansyl chloride has been used extensively to determine the N-terminal amino acid residue of proteins and to prepare fluorescent derivatives of drugs, amino acids, oligonucleotides and proteins for detection by numerous chromatographic methods. Non-fluorescent dansyl chloride reacts with amines to form fluorescent dansyl amides that exhibit large Stokes shifts, along with environment-sensitive fluorescence quantum yields and emission maxima.

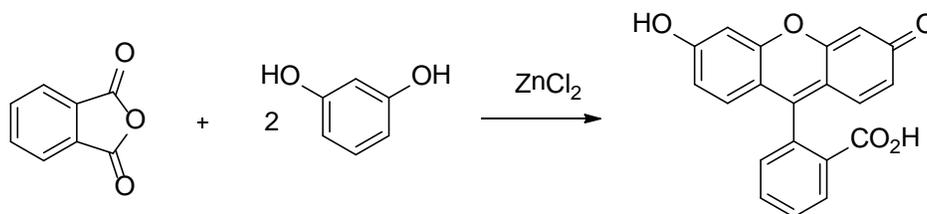


For instance, Chersie and collaborators recently used dansyl chloride for direct “in synthesis” labeling of peptides¹¹.

II.2 Fluorophores with emission wavelengths above 500 nm

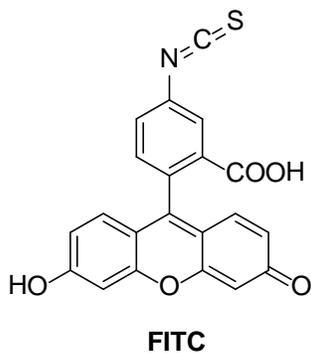
II.2-1 Fluorescein

Fluorescein is one of the most common labels used in biology with absorption and fluorescence maxima in the visible region of the spectrum ($\lambda_{\text{abs}} = 490 \text{ nm}$ and $\lambda_{\text{em}} = 512 \text{ nm}$, in water). It may be synthesized from phthalic anhydride and 1,3-dihydroxy-benzene (resorcinol) in the presence of zinc chloride via the Friedel-Crafts reaction¹². Alternatively methanesulfonic acid may be used as the catalyst¹³.

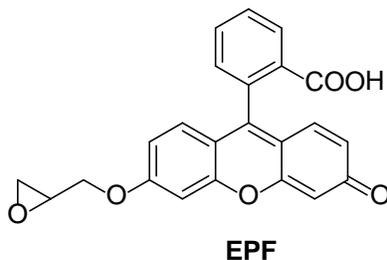


Thanks to its relatively high molar absorptivity, excellent fluorescence quantum yield, and good solubility in water, fluorescein possesses the advantage of an excitation maximum at 494 nm, which is close to the 488 nm spectral line of the argon laser, thus making it an important fluorophore for applications involving confocal laser-scanning microscopy and flow cytometry. Furthermore, protein conjugates based on fluorescein are not very susceptible to precipitation and can thus be obtained at high purity levels.

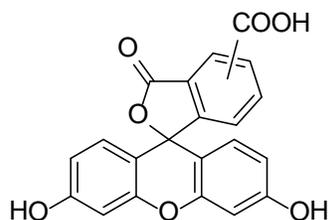
The most popular fluorophore used in the conjugation with proteins is fluorescein isothiocyanate (FITC) which reacts with primary amines to form the corresponding thioureas.



3-epoxypropoxy fluorescein (EPF), was can also be used for selective labeling of histidine residues¹⁴.

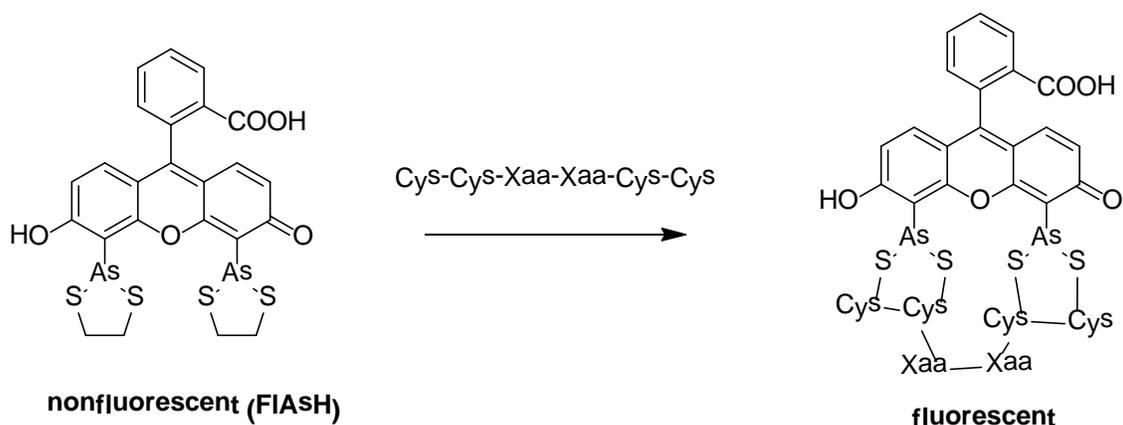


Fluorescein can also be coupled to peptide by solid-phase reaction by using 5(6)-Carboxyfluorescein (represented below)¹⁵. For instance Giralt and Fernandez-Carneado fluorescently labeled proline-based peptides (hexaproline P6, dodecaproline P12, and 18-proline P18) at their N terminus with 5(6)-carboxyfluorescein¹⁶.



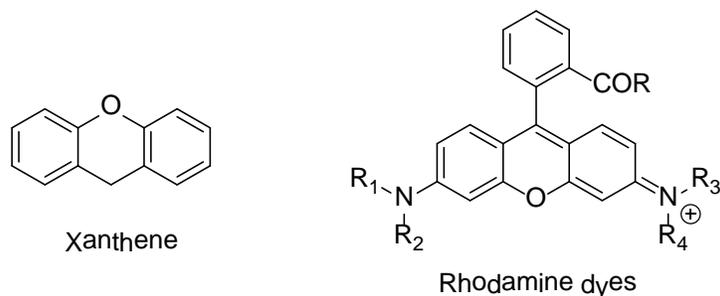
Study of the localization of proteins within live cells via fluorescence microscopy typically involves fusion of the required protein with a large fluorescent protein such as green fluorescent protein (GFP). Fluorescent biarsenical dye molecules (e.g., FAsH and ReAsH), were recently developed for the specific labeling of target peptides or proteins with small fluorescent dyes *in vivo*¹⁷. They represent an alternative to the use of large fusion proteins.

A motif composed by 4 cysteines Cys-Cys-Xaa-Xaa-Cys-Cys (where Xaa is any amino acid except cysteine) binds such biarsenical dyes with high affinity through its four cystein thiols (see scheme below).



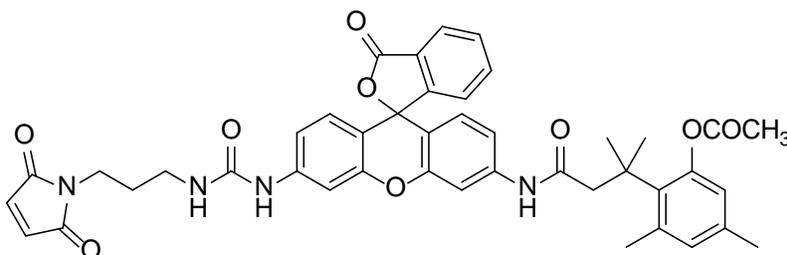
II.2-2 Rhodamine

Rhodamine dyes are fluorophores that belong to the family of xanthenes along with fluorescein and eosin dyes. The general structures of xanthene chromophore and rhodamine dyes are represented in the figure below.



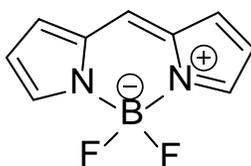
Due to their excellent photostability and photophysical properties, rhodamines are commonly used as laser dyes, fluorescence standards (for quantum yield and polarization), pigments and as fluorescent probes to characterize the surface of polymer nanoparticles, to assess the fluidity of lipid membranes, as well as for the detection of polymer-bioconjugates, single-molecule imaging and imaging in living cells. Although, for some of those applications, the dye is used in its free form, for most of them the probe must be attached to another molecule (polymer, oligonucleotide, biomolecule, etc.) or to a surface. Raines *et al.*¹⁸ described a versatile “latent” fluorophore (Scheme below), a

derivative of Rhodamine 110 in which one of the nitrogens has been modified as a urea and the other as a “trimethyl lock”. The first alteration produced suppression of half-fluorescence of rhodamine while facilitating conjugation with a target molecule. The second modification enabled fluorescence to be unmasked fully by a single user-designated chemical reaction.



II.2-3 BODIPY

Although a plethora of fluorophores exist, the fragility of organic molecules to sustained light bombardment, especially in the presence of molecular oxygen, has necessitated the search for new more robust molecules. In particular, low triplet yielding molecules are required since the pathway of decomposition via singlet oxygen is, at least, partially curtailed. One such dye, borondifluorodipyrromethene (Bodipy, see figure below), has been championed as such a molecule, and is based on the half-porphyrin motif but chelated by a difluoroboron moiety to maintain overall structural rigidity.

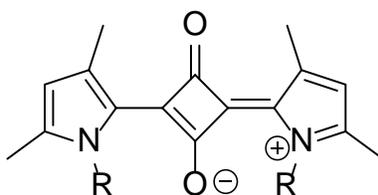


BODIPY exhibits optical properties that are often superior to fluorescein, tetramethylrhodamine, Texas Red, and other longer-wavelength dyes. They generally have high molar absorption coefficients (typically $\log \epsilon_{\max} > 8.8$), high fluorescence quantum yields ($\Phi_F \approx 1.0$, even in water), good photostability, and narrow emission band

widths (important in multicolor applications). BODIPY dyes proved to be extremely versatile and useful for a number of biological applications¹⁹.

II.2-4 Squaraines

Squaraine dyes are condensation products of electron rich aromatic and heterocyclic molecules such as *N,N*-dialkylanilines, benzothiazoles, phenols, azulenes and pyrroles with 3,4-dihydroxy-3-cyclobutene-1,2-dione (squaric acid). They are highly colored zwitterionic dyes, which possess intense absorption in the visible to the near-infrared region, high quantum yields and high photo-stability. A representative example is shown in the scheme below.



Additionally, some squaraine dyes show high sensitivity towards the polarity of the surrounding environment (e.g. of the solvent), and have their quantum yields increasing drastically in the presence of biomolecules such as proteins and antibodies, thus making squaraine dyes particularly well suited for bio-labeling purposes. Most of these dyes exhibit far-red fluorescence around 650 nm and recently, water-soluble squaraine dyes emitting in the NIR were also successfully developed for bio-labeling and bio-imaging.

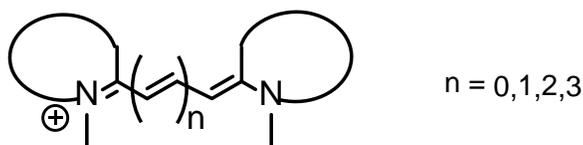
II.3 Introduction to Cyanine Dyes

II.3-1 Generalities about cyanine dyes

a) Definitions

Cyanine dyes are highly colored fluorescent compounds which generically consist of two nitrogen containing heterocycles linked by a conjugated chain of an odd number

of carbon atoms (Figure below). They are generally named based on the number of carbons in the polymethine linker with four main lengths of chain: monomethine ($n = 0$), trimethine ($n = 1$), pentamethine ($n = 2$) and heptamethine ($n = 3$).

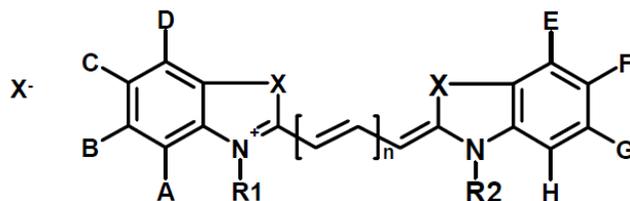


Generally, the nitrogen atoms in both heterocycles are alkylated with simple alkyl chains, however sometimes more complex functionalities are introduced. These modifications may be to convey physical properties on the dye, such as aqueous solubility. This position has also been used to introduce a functionality which allows attachment of dyes to biomolecules, such as *N*-hydroxy-succinimide. Alkylation gives the dye a net positive charge which is delocalized across the polymethine chain, thus leading to a push-pull system responsible for the fluorescence properties of these dyes. The polymethine chain is generally in an all-*trans* geometry, however photo-isomerization can occur switching the dye into a ‘dark’ or non-fluorescent state.

Polymethine cyanine dyes can cover a very broad spectral range from blue to red. Several classes of polymethine cyanine dyes that are excitable in the far-red (635-1100 nm) were also recently reported. These include heptamethine cyanine dyes²⁰, squaraine cyanine dyes²¹ and sulfoindocyanine dyes²². All of these dyes have a *trans* geometry in their stable form.

Although they cover a large variety of colors, cyanine dyes are not widely used for dyeing purpose, as they are decolorized by light and acid. They have, however, been employed extensively as spectral sensitizers for silver halide photography and other inorganic large band-gap semiconductor materials, in optical disks as recording media, in industrial paints, for trapping of solar energy, as laser materials, in light-harvesting systems of photosynthesis, as photorefractive materials, as antitumor agents and as probes for biological systems. They can also be utilized as optical switches, light emitting diodes, photovoltaic devices, dye lasers and are applied in reprographics and photodynamic therapy.

b) Nomenclature



The above figure represents the generic structure of a cationic streptopolymethine cyanine dye. Standard chemical names specify exactly the chemical structure of the molecule. The Cy3 and Cy5 nomenclature was first proposed by Ernst in 1989, and is non-standard, since it gives no hint of their chemical structures.

For $n=0$, these dyes are called Cyanine1, or Cy1;

For $n=1$, these dyes are so-called Cyanine3 or Cy3;

For $n=2$, these dyes are called Cyanine 5 or Cy5;

For $n=3$, these dyes are called Cyanine7 or Cy7.

In other words, the number after the contraction “Cy” refers to the number of atoms of the conjugated chain between the heteroaromatic moieties. The monomethine cyanines show absorption in the visible region, and each extension of the chromophore by one vinylene moiety (CH=CH) causes a bathochromic shift of about 100 nm²³. When this nomenclature was first introduced, the Cy number designated the count of the methines (as shown above), and the side chains were unspecified. Thus various structures are designated as Cy3 and Cy5 in the literature. Cy3 dyes are green (~550 nm excitation, ~570 nm emission), while Cy5 dyes are fluorescent in the red region of the spectrum (~650/670 nm).

Structure and Spectral Profiles of Cyanine Fluorochromes

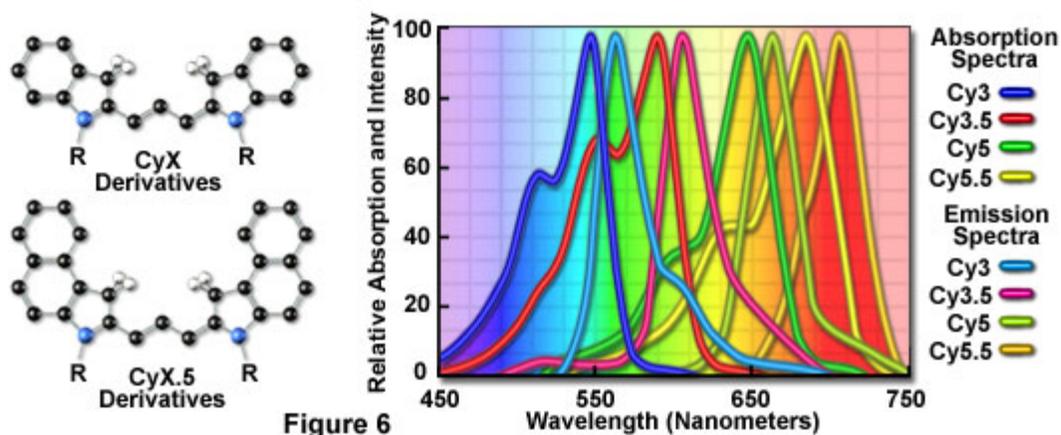
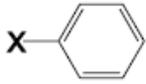


Figure 5: Structural and spectral profiles of cyanine fluorochromes extracted from ref²⁴

c) Fluorescence and Push-pull mechanism

A push-pull polymethine is a type of olefin characterized by an electron-withdrawing substituent on one side of the double bond and an electron-donating substituent on the other side. This makes the π bond very polarized. The rotational barrier for a push-pull olefin is lower than that of an ordinary olefin and this makes it an interesting candidate for a molecular switch for instance azobenzenes. A push-pull configuration also helps to stabilize the double bond because the carbon-carbon bond has considerably less double bond character.

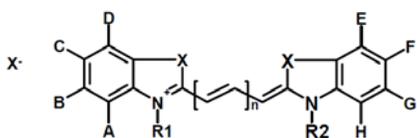
electron donors D*	electron acceptors A*
 <p>X = H-, (CH₃)₂N-, CH₃-O-, HS- naphthalene, anthracene, phenanthrene, pyrene, perylene</p>	 <p>Y = -C≡N, -C(=O)-R, -N₂O 9,10-dicyanoanthracene</p>

PUSH

PULL

Highly conjugated push-pull chromophores generally exhibit improved performance, lead to a significant bathochromic shift and increase the fluorescence quantum yield of a molecule. The push-pull activity of these chromophores is determined by the strength of the Donor/Acceptor pair (i.e., relative redox coupling) and by the electronic/structural alteration of the π -conjugating spacer.

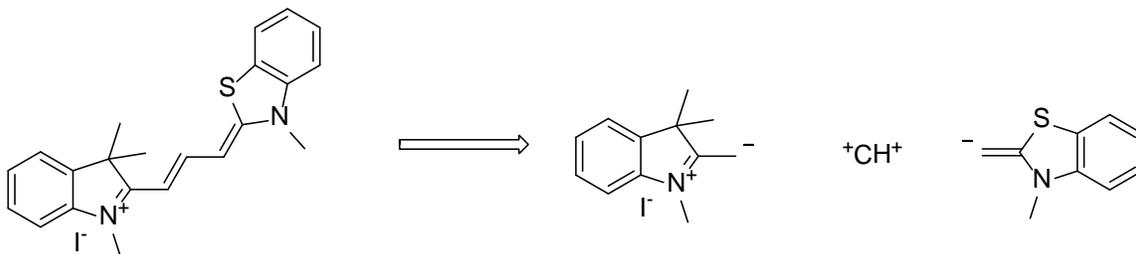
d) Structural modifications of Cyanine Dyes



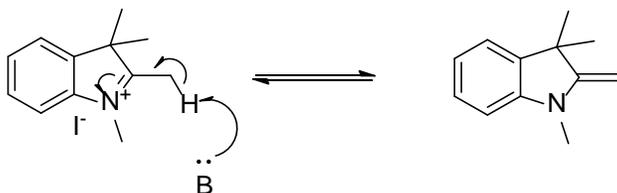
Different modifications on their generic structure can modify the spectroscopic properties of cyanine dyes (e.g. wavelength shift). In the structure above, X refers to an alkyl group or to a heteroatom (most commonly O, S, N or Se). The nature of the substituents (A-H) born by both heterocyclic moieties can also cause additional shifts in absorption. Typical modifications include replacement of indoles by benz[E]indoles to increase the aromaticity of the system or the functionalization of the benzene ring with carboxylic acids or halogens (resulting in changes in fluorescence quantum yield or fluorescence lifetime). Such modification can allow a convenient tuning of the fluorescent properties of cyanine dyes.

II.3-2 Cyanine dyes synthesis

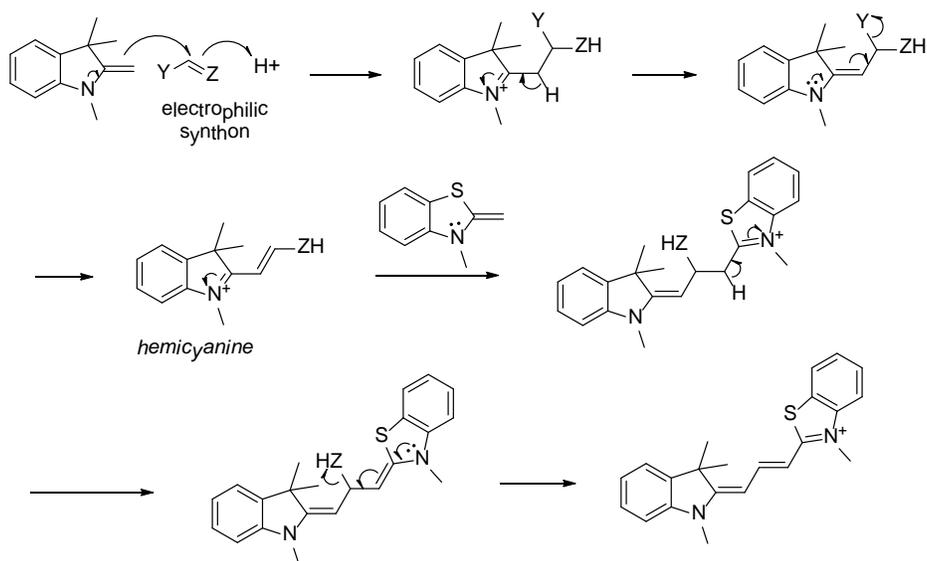
The retrosynthetic analysis for a Cy3 dye (scheme below) shows the synthons for the two heterocycles and the polyene chain. The same general route is used for the synthesis of Cy5 dyes, albeit with a longer linker synthon.



The heterocyclic synthons used are nucleophilic enamines. The heterocycle is normally added as the quaternary iminium salt and is deprotonated to form the methylene base in situ.



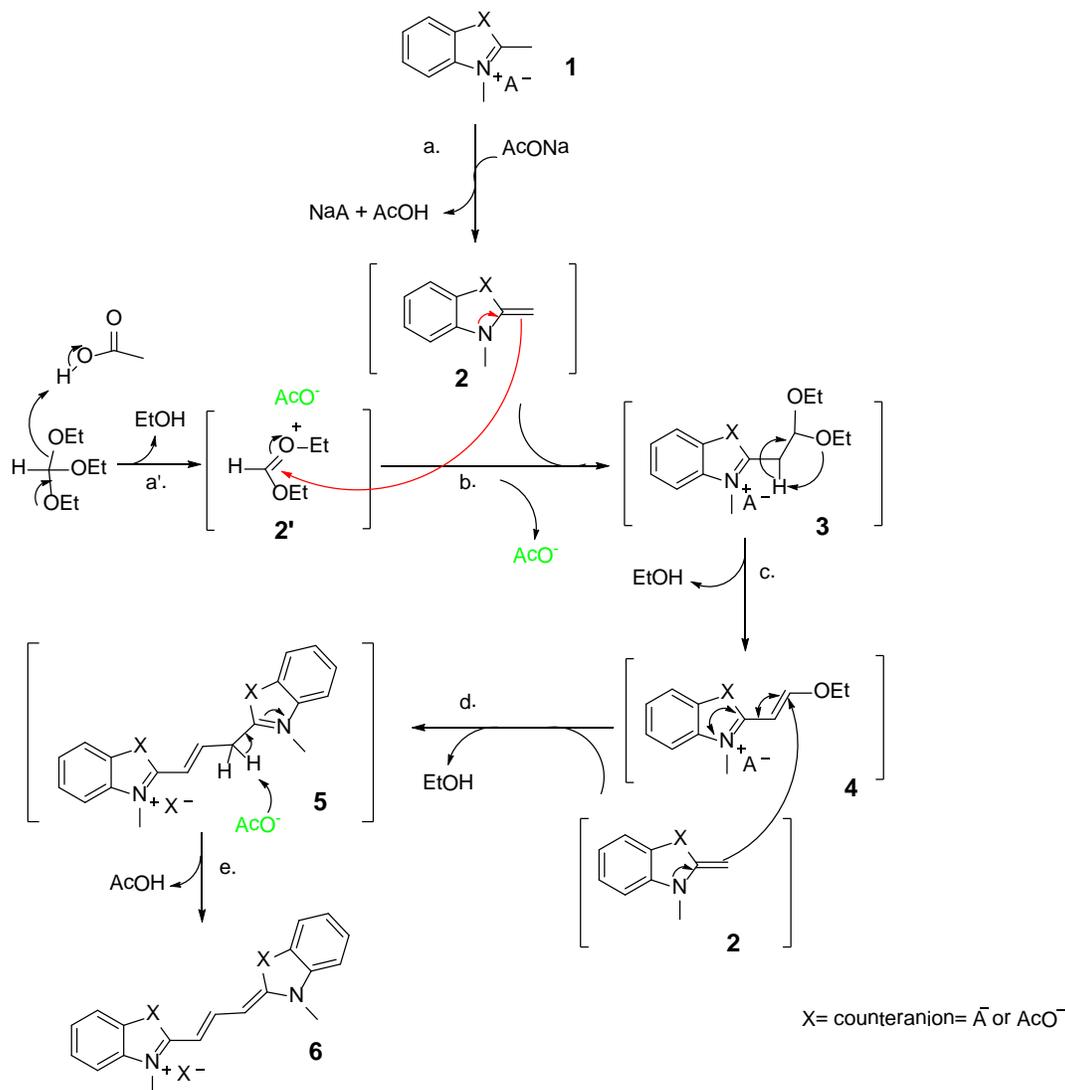
The synthon for the polyene bridge is provided by electrophilic compounds (scheme below) where Y and Z are leaving groups. Y and Z are replaced by stepwise attack of the methylene bases. The synthesis of symmetrical dyes (where both heterocycles are the same) is relatively straightforward and can be carried out simply by use of two equivalents of the heterocycle. Unsymmetrical dyes can be synthesized by the reaction of one equivalent of each of the two heterocycles. However, this is a wasteful route as the two corresponding symmetrical dyes are also formed simultaneously. Additionally, separation of symmetric and asymmetric dyes is non-trivial. For the synthesis of asymmetric dyes the intermediate “half-dye” (or hemicyanine) is isolated and purified before being reacted with a second (and different) heterocycle to yield the desired unsymmetrical product.



Different strategies have been reported in the literature for the synthesis of symmetrical and unsymmetrical cyanine dyes. Below, we are presenting a selection of the most commonly used methods for the synthesis of trimethine and pentamethine cyanine dyes.

- **Symmetrical trimethine cyanine dyes using $CH(OEt)_3$, AcOH, AcONa**

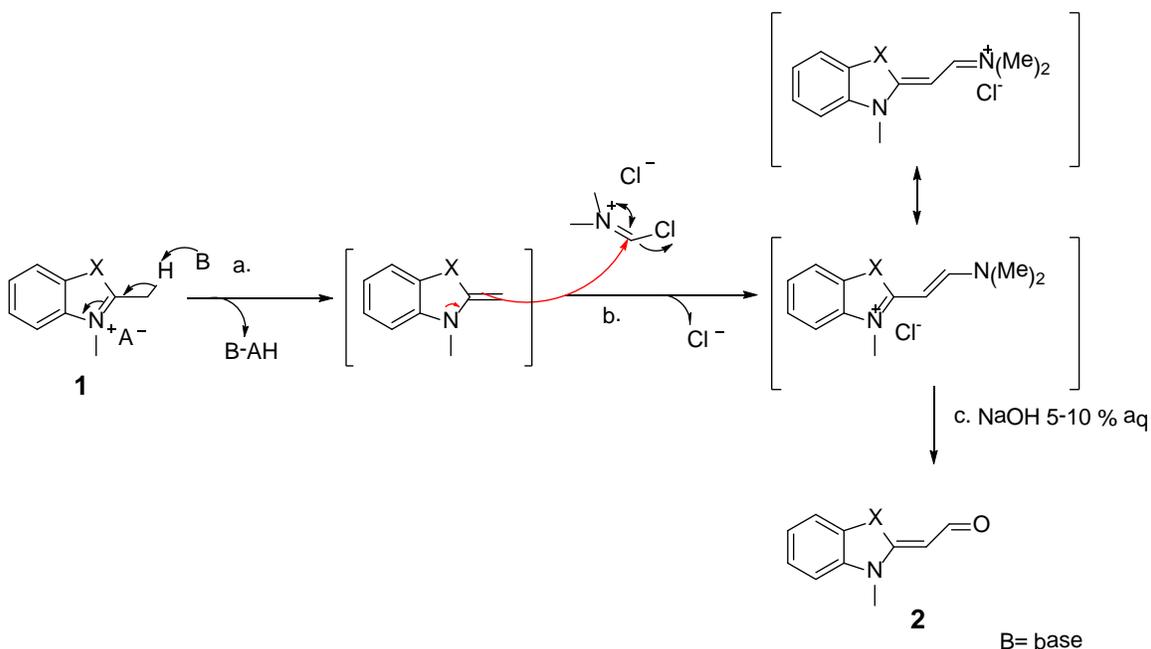
A useful route to obtain symmetrical trimethine cyanine dyes requires the use of triethyl orthoformate $HC(OEt)_3$ in acetic acid, in the presence of a base like AcONa. On the scheme below, we are presenting the detailed mechanism for this synthesis. The first step (a) is an isomerization in situ of the quarternized salt **1** to **2**. In the presence of acetic acid, triethylorthoformate undergoes an elimination reaction (with a loss of methanol) to generate the highly reactive electrophile **2'** which can then react with compound **2** to produce the intermediate **4**. Because salt **1** is added in excess, a second equivalent of **2** can react with **4** to generate, via an addition/elimination reaction the desired trimethine cyanine dye **6**.



- *Symmetrical and unsymmetrical trimethine cyanine dyes using (Chloromethylene)dimethylammonium chloride*

In order to synthesize unsymmetrical cyanine dyes, it is necessary to be able to isolate the activated hemicyanine (before it reacts with its heterocyclic precursor). One efficient route to achieve that is via the synthesis of a stable Fisher's base aldehyde intermediate (that can be isolated and purified before being reacted with a second heterocyclic nucleophile). This general strategy is detailed on the scheme below and requires the use of (Chloromethylene)dimethylammonium chloride. The first step is the

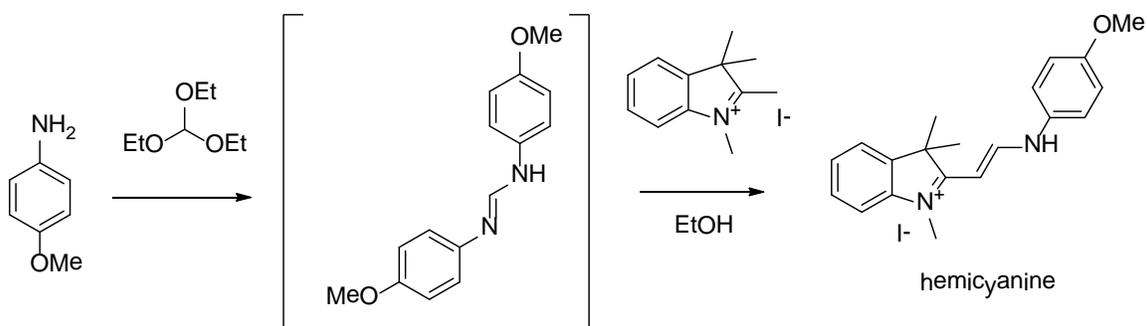
preparation of the aldehyde **2** via isomerization of salt **1** followed by a nucleophilic attack onto (Chloromethylene)dimethylammonium chloride, followed by a hydrolysis of the ammonium intermediate formed.



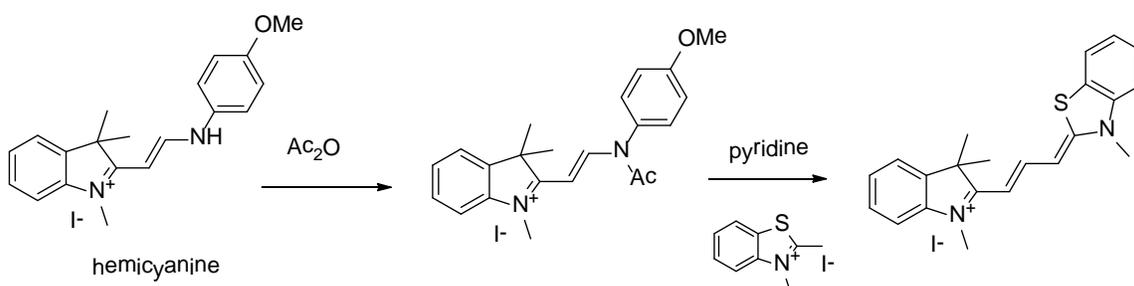
Symmetrical and unsymmetrical trimethine cyanine dyes can then be obtained by reaction of the isolated aldehyde with an identical or different heterocyclic salt, respectively.

- **Symmetrical and unsymmetrical trimethine cyanine dyes using amidines**

One of the most widely used routes for the synthesis of cyanine dyes involves reaction of heterocycles with an amidine (scheme below) to form a so-called hemicyanine intermediate. Hemicyanines of this type can be isolated and purified by recrystallisation in good yields. Amidines are commercially available but can also be formed in situ by reaction of *p*-methoxyaniline with triethylorthoformate.

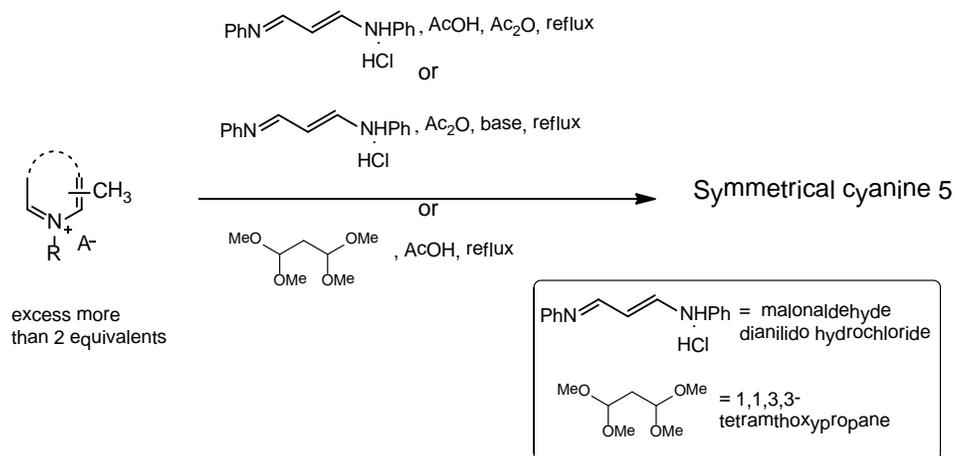


The hemicyanine can subsequently be activated in situ by acylation, then reacted with a second heterocycle to give the required symmetrical or unsymmetrical cyanine dye (Scheme below). Alternatively, the activated hemicyanine can be synthesized by inclusion of acetic anhydride in the reaction and this species can be isolated and purified.

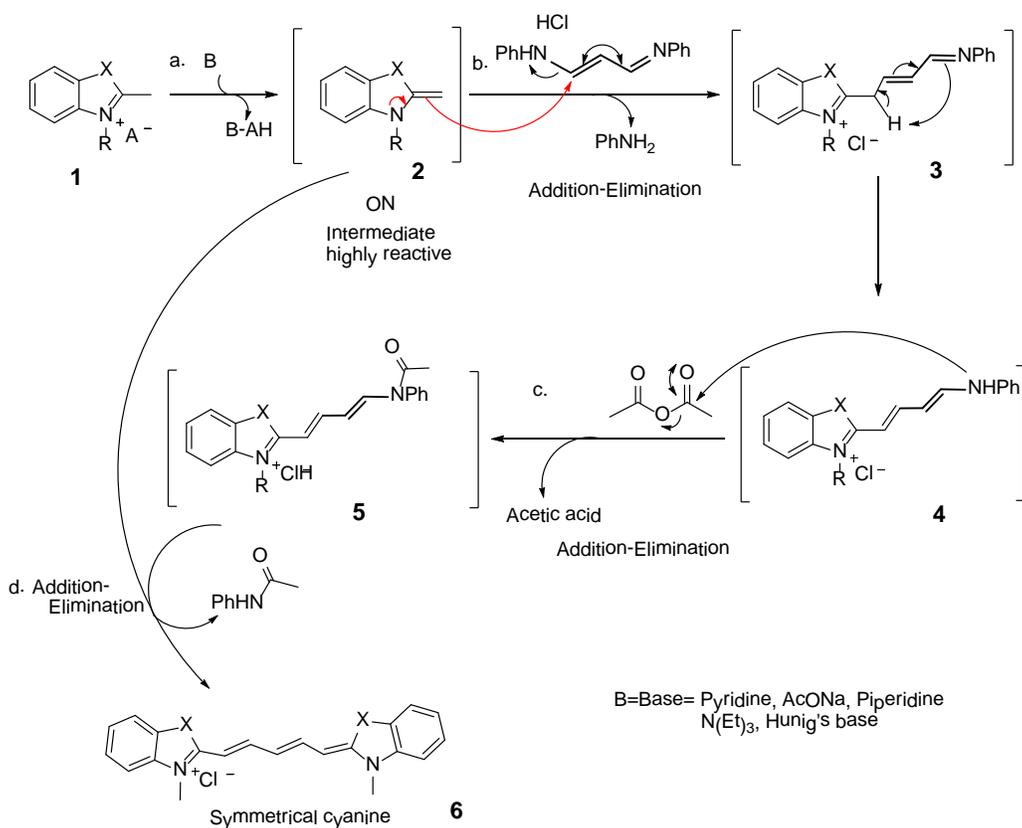


- **Symmetrical and unsymmetrical pentamethine cyanine dyes using amidines**

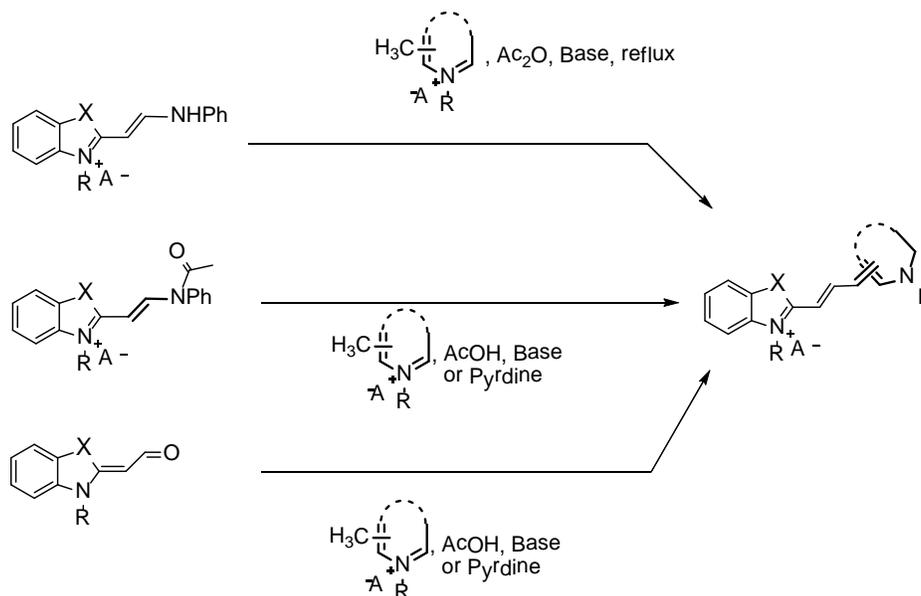
A similar strategy can be used to synthesize symmetrical and unsymmetrical pentamethine cyanine dyes, using as an electrophile either 1,1,3,3-tetramethoxy propane or an amidine formed from malondialdehyde (see general scheme below).



The mechanism of this synthesis is detailed on the scheme below. Once again, it starts with an isomerization of the salt **1** in the presence of base, followed by a condensation on malondialdehyde dianilido hydrochloride. The hemicyanine intermediate can then be isolated (as for the Cy3 synthesis), activated with acetic anhydride and then be reacted with a second equivalent of heterocyclic precursor **1** to generate the desired symmetrical pentamethine cyanine dyes.



As described above for pentamethine cyanine dyes, unsymmetrical pentamethine cyanine dyes can also be obtained if the hemicyanine intermediate can be isolated and subsequently reacted with a second heterocyclic precursor (scheme below).



II.3-3 Cyanine dyes aggregation properties

Cyanine dyes are well known to aggregate in solution. The self-association of dyes in solution or at the solid-liquid interface is a frequently encountered phenomenon in dye chemistry owing to strong intermolecular Van der Waals-like attractive forces between the molecules. π -stacking of the conjugated systems of the dyes is favored based on both the hydrophobicity and polarizability of the dye. Two limiting types of supramolecular structures are formed and are referred to as “H” and “J” aggregates. The dye molecules may aggregate in a parallel way (plane-to plane stacking) to form a sandwich-type arrangement (H-dimer) or in a head-to-tail arrangement (end-to end stacking) to form a J-dimer. The tendency of dye molecules to aggregate depends on the structure of the dye and also on the environment, such as micellar, microemulsion, pH, ionic strength, concentration, solvent polarity, electrolyte, temperature parameters²⁵, presence of metals (the probability to see aggregate with monovalent cation is highest than with divalent), presence of polymers such as (DNA, sugar, protein). Generally, dye

self-association in solution depends directly on dye concentration, added electrolyte and inversely on the temperature.

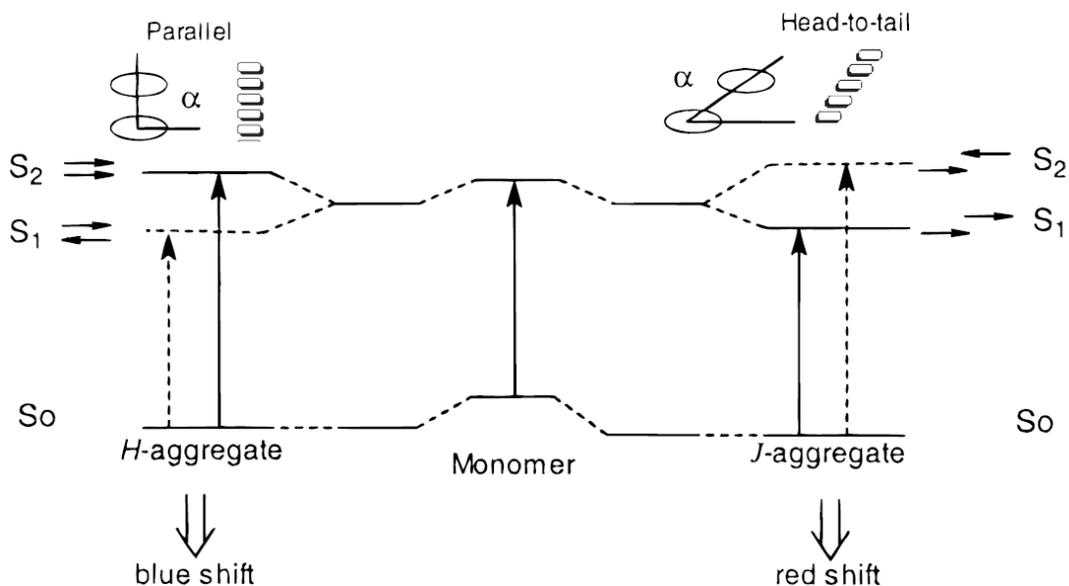


Figure 6: Representation of the H-aggregate and J-aggregate and their influence in terms of wavelengths

Unsubstituted dyes favor H-aggregation as this provides the greatest number of van der Waals interactions and minimizes exposure to water. However, substituents placed at various positions on the dye structure (for example the N-alkyl chain²⁶) can promote J-aggregation due to steric and/or electrostatic factors. J aggregates are also promoted in presence of surfactant²⁷, polyelectrolytes²⁸, polymers²⁹ and Langmuir-lodgett films³⁰.

H- and J-aggregates typically exhibit absorption maxima that are shifted to either shorter or longer wavelengths, respectively, compared to the monomeric form of the dye. Aggregation in solution also modifies the intensity of the absorption band. The bathochromically shifted J-bands³¹ (J for Jelly, one of the first workers who investigated these shifts) and hypsochromically shifted H-bands³² (H for hypsochromic) of the aggregates have been explained in terms of molecular exciton coupling theory, i.e., coupling of transition moments of the constituent dye molecules³³. The aggregation behavior of cyanine dyes has been studied extensively since these are the best known self-aggregating dyes. The red-shifted J-aggregate is the most commercially important

dye assemblage for photographic spectral sensitization. It is generally agreed that both H- and J-aggregates are composed of parallel dye molecules stacked plane-to-plane and end-to-end and form two-dimensional dye crystals.

II.3-4 Cyanine dyes biological applications

Cyanine dyes have found extremely valuable applications in modern bioresearch for instance for the labeling of biomolecules. Although they suffer from certain drawbacks (e.g., short fluorescence lifetimes; low fluorescence quantum yields; and, extensive aggregation in aqueous solution³⁴) cyanine dyes can be easily functionalized and anchored to biomacromolecules for labeling or sensing purposes. Structural modifications of the dye can also help improving the properties of these dyes (for example the introduction of sulfonate can improve water solubility, quantum yield and photochemical stability). Also, adding reactive groups, such as isothiocyanates, makes the dye suitable for use as covalent label. Indolium derivatives also exhibit good light stability that can be improved by incorporating a ring structure into the polymethine chain³⁵.

Below, we have listed a few representative examples of biological applications of polymethine cyanine dyes.

a) Cyanine dyes and DNA sensing:

In the 1980's Lee and co-workers discovered that an old photographic cyanine dye was an excellent fluorescent non-covalent label for nucleic acids [they called it Thiazole Orange (TO)]. In the following years many new monomethine and trimethine cyanine dyes have been designed, synthesized and commercialized. Today cyanine dyes are among the most widely used non-covalent fluorescent probes for nucleic acid detection and have hundreds of applications in bioinvestigations and for clinical and medical analysis. Cyanine dyes are known to bind to DNA through non-covalent interactions. Monomethine dyes intercalate between bases while trimethine and pentamethine dyes aggregate in the minor grooves.

An example of applications is the use of cyanine dyes as Light-up probes. Detection of specific DNA sequences or structures is important for genetic screening, clinical diagnostics and microarray analyses of gene expression. It is usually a two-step process with first hybridization of a complementary probe strand to the DNA target followed by a signaling that hybridization has occurred. Fluorescence remains the most common and most powerful method of detection that hybridization has indeed occurred. Hybridization is associated to a change in fluorescence wavelength, intensity or polarization that can therefore be measured and quantified. A well known example of this strategy involves molecular beacons probes in which the probe folds into a stable hairpin secondary structure, bringing together a fluorophore and a quencher. Hybridization to a complementary strand requires opening of the hairpin, thereby separating the quencher from the fluorophore and enhancing fluorescence.

The sensitivity of TO fluorescence based on whether it is free in solution or bound to double-stranded DNA also suggested its use as a reporter for hybridizations. TO was also recently used as a probe for nucleosome assembly and dynamics by Woodbury and coworkers. In order to study the protein-DNA interactions a reactive form of TO was covalently attached to a unique cysteine residue in one of the histone proteins. Reconstitution of the protein with DNA resulted in a substantial increase in fluorescence of the TO label.

b) Cyanine dyes and FRET pairs:

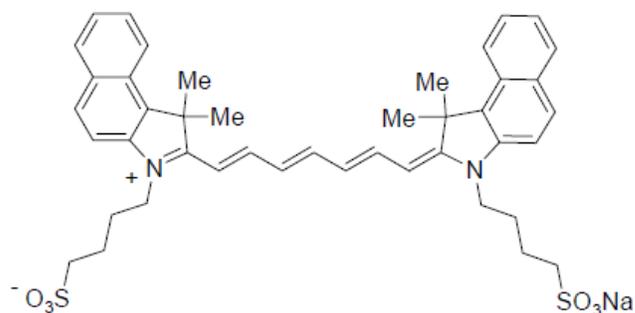
FRET is a versatile luminescent approach widely applied for analytical purposes. Usually a fluorescent donor transfers its electronic energy non-radiatively to an acceptor, which can be fluorescent or non-fluorescent. As a result, the fluorescence intensity of the donor decreases, while the fluorescence of the acceptor increases if the acceptor is fluorescent. For FRET-based applications to be successful there should be considerable overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor.

The organic cyanine dyes Cy3, Cy5, Cy5.5 and Cy7, which emit in the red range (>550 nm), offer a number of advantages for FRET applications. Their emission range is

such that background fluorescence is often reduced. Additionally, large distances (>100 Å) can be measured as a result of the high extinction coefficients and good quantum yields of this class of dyes. Even donor-acceptor pairs with separated emission spectra (i.e. low overlap integral) result in acceptable Förster distances. For example, Cy3, which emits maximally at 570 nm and Cy5, which emits at 670 nm, have a Förster distance >50 Å. Large separation between pairs allows the measurement of acceptor emission as a result of FRET without interference from donor emission. In addition these molecules can be linked directly to specific locations in synthetically produced nucleic acids, allowing FRET to be used to assess nucleic acid annealing.

c) Near-InfraRed absorbing cyanine dyes and in vivo imaging:

Indocyanine green (ICG, scheme below) is a negatively charged heptamethine cyanine dye that presents a strong absorption at 780 nm and a weak fluorescence at 820 nm in aqueous solution. Upon binding to a target protein a red-shift of the absorption band of the dye and an enhancement of the fluorescence intensity occurs. Proteins can therefore be resolved without any interference with unbound dye.



The indocyanine green was successfully used to detect proteins such as human serum albumin (HSA) by capillary electrophoresis with diode laser-induced fluorescence detection (CE-LIF)³⁶. Indocyanine green (ICG) was also approved for clinical retinal angiography and liver function testing³⁷. Additional applications of ICG include its use for determining cardiac output, hepatic function, and liver blood flow, and for ophthalmic

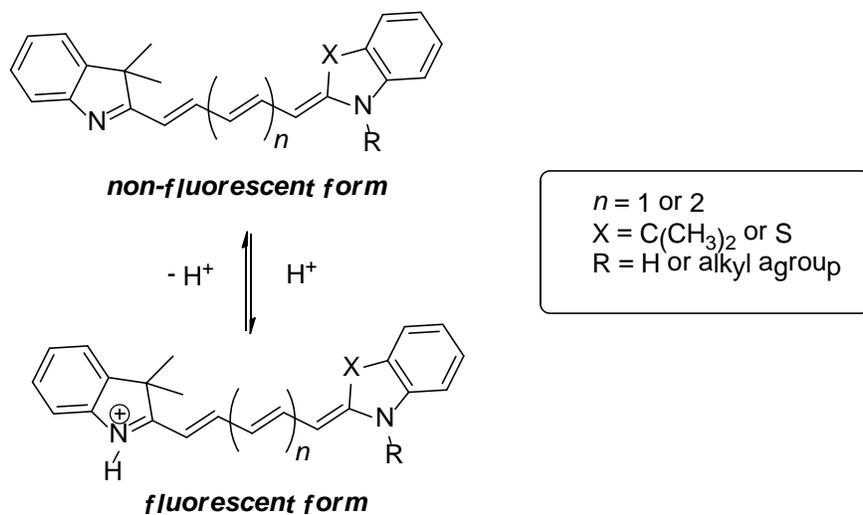
angiography. ICG binds tightly to plasma proteins and becomes confined to the vascular system. ICG has a half-life of 150 to 180 seconds and is removed from circulation exclusively by the liver to bile juice.

d) Cyanine dyes and pH Sensors:

Modulation of pH-responsive cyanine dye pKa values via heteroatom substitution allows the design of fluorescent reporters that are tuned for potential imaging of biologically relevant acidic environments. Two types of pH-sensitive dye exist:

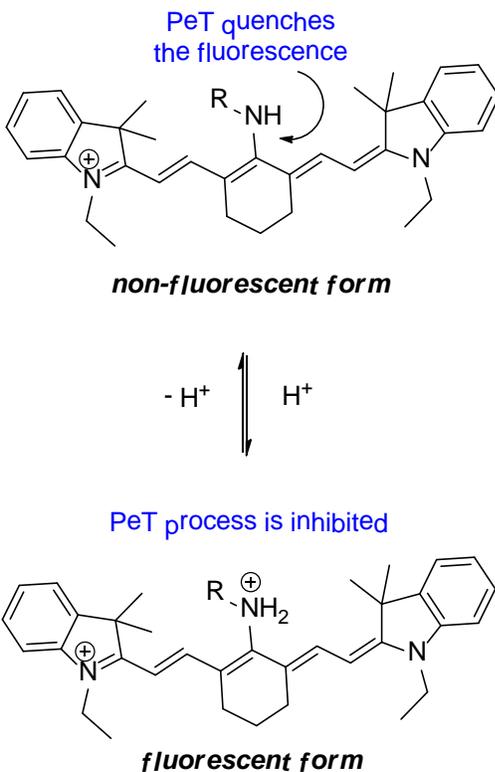
- *non-N-alkylated indolium structures:*

These dyes become fluorescent only if the non-quarternized intracyclic nitrogen is protonated (thus restoring the push-pull mechanism characteristic of cyanine dyes) whilst the non-N- protonated forms switch off the fluorescence brightness. These compounds have been used as pH probes, a new blue-shifted absorption peak appearing when increasing the pH of the solution. Subtle changes to the structure of these cyanine-based probes can change their pKa values.



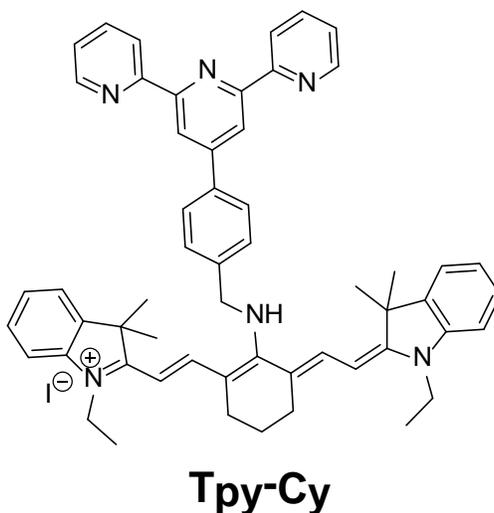
- *photoinduced electron transfer (PeT):*

This pH-sensitive dye consists of a heptamethine cyanine dye in which one central carbon of the polymethine chain is connected to the Nitrogen-containing modulator. Turning on and off the fluorescence is achieved by suppressing or allowing PeT processes by protonation/deprotonation of the modulator. In contrast, to the first type of pH indicators the maximum absorption peak has a large red shift when the pH is increased.



For instance, a tricyanin (Cy) fluorophore was coupled to a 4'-(aminomethylphenyl)-2,2'-6',2''-terpyridine (Tpy) and behaved as a receptor that responds sensitively to H⁺ near neutrality³⁸. This probe detects pH in biological systems through a fast photoinduced electron-transfer process (PET). The fluorescence of Tpy-Cy is quenched due to PET between the receptor and the fluorophore, whereas upon N atoms protonation the quenching process is rendered and the fluorescence emission is “switched

on”. The pH titration indicates that Tpy-Cy can monitor the minor pH fluctuations with a pKa of ~7.10 near physiological pH.



At pH 10, the brightness of **Tpy-Cy** was observed to be low ($\phi = 0.008$), presumably due to quenching via PeT, involving electrons of the terpyridine group. Protonation of the N atoms circumvents these PeT processes, and the dye fluoresces brightly ($\phi = 0.13$) at 750 nm with a pKa in aqueous buffer of ca. 7.1. The probe responds linearly and rapidly to minor pH fluctuations within the range of 6.70-7.90 and exhibits strong dependence on pH changes.

II.3-5 Cyanine dyes and surfactants

A surfactant is an amphiphilic derivative because it is formed by two parts with different affinities for the solvents. One of them has affinity for water (polar solvents) and the other for oil (non-polar solvents). A little quantity of surfactant molecules stays at the water-air interface and decreases the water surface tension value (the force per unit area needed to make available surface).

When water, oil and a surfactant are mixed, the surfactant rests at the water-oil interface. These systems, depending on their stability, are called emulsions or microemulsions. Although, the properties for an emulsion and a microemulsion are

different, both obey the same principle: they try to form enough interface for preventing the polar non-polar solvent contact. Microemulsions are very interesting systems, because the oil-surfactant-water interface forms a wide variety of structures to avoid the direct oil/water contact. The sizes of these structures are in the range of a few hundreds of nanometers, so the solutions are transparent. Micelles are the simplest structures: spherical or cylindrical objects formed by surfactant molecules, separating oil and water. Micelles are like drops of oil in water and reverse micelles are like drops of water in oil.

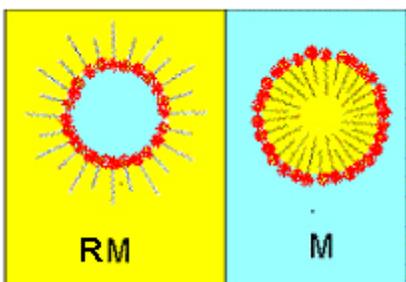
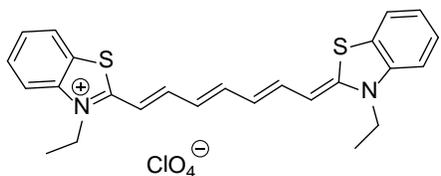


Figure 7: Spherical micelles (M) and reverse micelles (RM). The oil is represented in yellow and the water is in blue (<http://www.fisica.unam.mx/liquids/micelles.php>).

We previously mentioned that cyanine dyes were prone to aggregation. Addition of surfactant above the critical micelle concentration (cmc) leads to de-aggregation of cyanine dyes³⁹. The interaction of the 3,3'-diethylthiacarbocyanine (DTC), a cationic cyanine dye, with a molecule of surfactant was shown to depend on both electrostatic and hydrophobic forces⁴⁰. Nevertheless DTC is not always used to study aggregation-deaggregation because of the weak aggregation, A number of cyanine dye with two sulfonates groups have a tendency to form dimers and higher aggregates below the cmc and could be useful for these studies in aqueous solution⁴¹.



3,3'-diethylthiacarbocyanine (DTC)

II.3-6 Stability of Cyanine dyes

Cyanine dyes are usually stable at room temperature. Nevertheless, they are highly sensitive to nucleophiles (e.g. hydrazine, sulfide, phenolate), to reductive conditions and to radical species.

Hydrocyanines have been developed to be used as sensors for detecting the presence of oxygen radicals (superoxide and hydroxide radical) in living cells and tissue samples⁴². Reactive oxygen species are highly reactive metabolites of oxygen that have been implicated in a variety of inflammatory diseases, including cancer and atherosclerosis. Six hydrocyanine dyes have been developed to date (hydro-Cy3, hydro-Cy5, hydro-Cy7, hydro-IR-676, hydro-IR-783 and hydro-ICG) which vary by their different abilities to detect intracellular or extracellular reactive oxygen species and by their different emission wavelengths (from 560 to 830 nm). They were synthesized by reduction of a traditional cyanine dye with sodium borohydride. Reaction with a reactive oxygen species (ROS) oxidizes the hydrocyanine into the fluorescent cyanine dye.

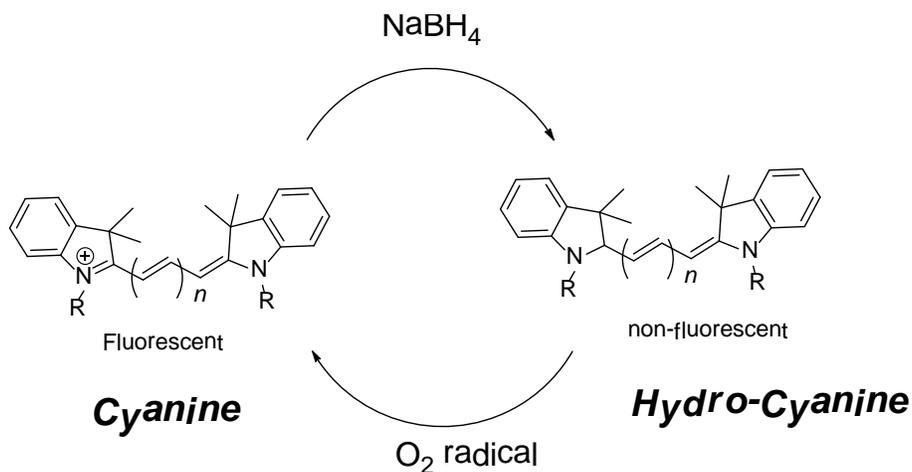
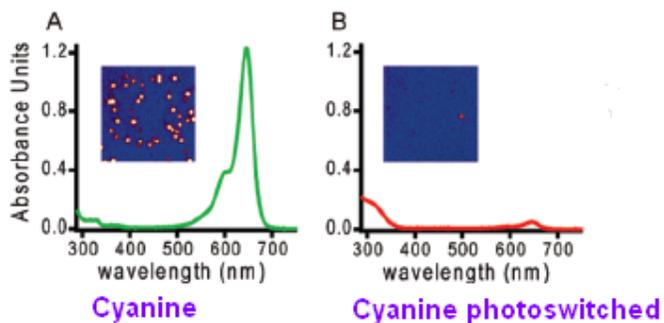
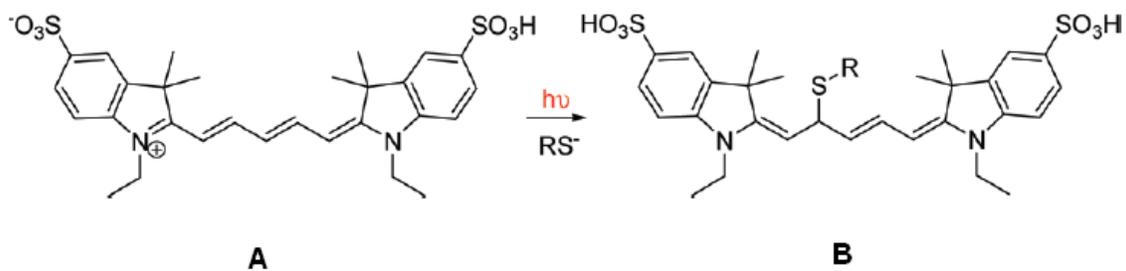


Figure 8: Hydrocyanines as sensors of O_2 radical

Cyanine photoswitches have also been recently developed by Tsien, Zhuang and collaborators⁴³ who reported the photoconversion of red cyanine dyes into their dark states upon illumination by red light. This phenomenon is facilitated by a primary thiol (e.g. β -mercaptoethanol, L-glutathione, or Cysteine) in solution as described in the figure below.



REFERENCES

- ¹ <http://www.chemicool.com/img1/graphics/morse-potential-curve.gif>
- ² <http://mounier.univ-tln.fr/projet/meef/image/JABLON.GIF> (Adapted from Jablonski, A. (1935). *Zeitschrift fur Physik*, 94, 38–64).
- ³ http://www.photobiology.info/Photochem_files/photochem13.gif
- ⁴ http://image.absoluteastronomy.com/images/encyclopediainages/s/st/stokes_shift.png
- ⁵ Gikas, E.; Parissi-Poulou, M.; Kazanis, M.; Vavagianis, A. *Anal. Chim. Acta* **2003**, 489, 153.
- ⁶ Wang, W.; Li, H. *Tetrahedron Lett.* **2004**, 45, 8479.
- ⁷ M. Sameiro T. Gonçalves *Chem. Rev.* **2009**, 109, 190.
- ⁸ (a) Doré, K.; Dubus, S.; Ho, H. A.; Lévesque, I.; Brunette, M.; Corbeil, G.; Boissinot, M.; Boivin, G.; Bergeron, M. G.; Boudreau, D.; Leclerc, M. *J. Am. Chem. Soc.* **2004**, 126, 4240 ; (b) Joussetme, B.; Blanchard, P.; Levillain, E.; Delaunay, J.; Allain, M.; Richomme, P.; Rondeau, D.; Gallego-Planas, N.; Roncali, J. *J. Am. Chem. Soc.* **2003**, 125, 1363.
- ⁹ Capobianco, M. L.; Naldi, M.; Zambianchi, M.; Barbarella, G. *Tetrahedron Lett.* **2005**, 46, 8181.
- ¹⁰ Solínová, V.; Kasicka, V.; Koval, D.; Barth, T.; Ciencialová, A.; Záková, L. *J. Chromatogr. B* **2004**, 808, 75.
- ¹¹ Muslin, A. J.; Tanner, J. W.; Allen, P. M.; Shaw, A. S. *Cell* **1996**, 84, 889.
- ¹² Ueno, Y.; Jiao, G.-S.; Burgess, K. *Pract. Synth. Proc.* **2004**, 31, 2591.
- ¹³ Confalone, P. N. *J. Heterocycl. Chem.* **1990**, 27, 31.
- ¹⁴ Li, X.; Ma, H.; Dong, S.; Duan, X.; Liang, S. *Talanta* **2004**, 62, 367
- ¹⁵ Fernandez-Carneado, J.; Giralt, E. *Tetrahedron Lett.* **2004**, 45, 6079
- ¹⁶ Fernandez-Carneado, J.; Kogan, M. J.; Castelo, S.; Giralt, E. *Angew. Chem. Int. Ed.* **2004**, 43, 1811.
- ¹⁷ Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, 281, 269.
- ¹⁸ Lavis, L. D.; Chao, T.-Y.; Raines, R. T. *ACS Chem. Biol.* **2006**, 1, 252.
- ¹⁹ M. Sameiro T. Gonçalves, *Chem. Rev.* **2009**, 109, 190.
- ²⁰ Mason, J. C.; Patonay, G. *J. Heterocyclic Chem.* **1996**, 33, 1685.

- ²¹ Oswald, B.; Lehmann, F.; Simon, L.; Terpetschnig, E.; Wolfbeis, O. *Anal. Biochem.* **2000**, *280*, 272.
- ²² Mujumdar, S. R.; Mujumdar, R. B.; Grant, C. M.; Waggoner, A. S. *Bioconjugate Chem.* **1996**, *7*, 356.
- ²³ Mishra, A.; Behera, R. K.; Behera, P. K.; Mishra, B. K.; Behera, G. B. *Chem. Rev.* **2000**, *100*, 1973.
- ²⁴ <http://www.olympusfluoview.com/theory/images/fluorophoresintrofigure6.jpg>
- ²⁵ A.H. Herz, *Adv. Colloid Interface Sci.* **1977**, *8*, 237
- ²⁶ (a) Tachibana, H.; Sato, F.; Terrettaz, S.; Azumi, R.; Nakamura, T.; Sakai, H.; Abe, M.; Matsumoto, M. *Thin Solid Films* **1998**, *813*, 327; (b) Rousseau, E.; Van der Auweraer, M.; De Schryver, F. C. *Langmuir* **2000**, *16*, 8865; (c) De Rossi, U.; Moll, J.; Spieles, M.; Bach, G.; Dähne, S. *J. Prakt. Chem.* **1995**, *337*, 203.
- ²⁷ (a) Chibisov, A. K.; Prokhorenko, V. I.; Gärner, H. *Chem. Phys.* **1999**, *250*, 47; (b) Tatikolov, A. S.; Costa, S. M. B. *Chem. Phys. Lett.* **2001**, *346*, 233.
- ²⁸ (a) Peyratout, C.; Donath, E.; Daehne, L. *Photochem. Photobiol. Sci.* **2002**, *1*, 87; (b) Liu, M.; Kira, A. *Thin Solid Films* **2000**, *359*, 104; (c) Fukumoto, H.; Yonezawa, Y. *Thin Solid Films* **1998**, *327*, 748; (d) Rousseau, E.; Koetse, M. M.; Van der Auweraer, M.; De Schryver, F. C. *Photochem. Photobiol. Sci.* **2002**, *1*, 395.
- ²⁹ (a) Higgins, D. A.; Barbara, P. F. *J. Phys. Chem.* **1995**, *99*, 3; (b) Misawa, K.; Ono, H.; Minoshima, K.; Kobayashi, T. *Appl. Phys. Lett.* **1993**, *63*, 577; (c) Misawa, K.; Ono, H.; Minoshima, K.; Kobayashi, T. *J. Lumin.* **1994**, *60*, 809; (d) Horng, M.-L.; Quitevis, E. L. *J. Phys. Chem.* **1993**, *97*, 12408; (e) Vacha, M.; Hashizume, K.-I.; Tani, T. *J. Lumin.* **2000**, *730*, 87.
- ³⁰ Ishizawa, H.; Sato, T.; Sluch, M. I.; Vitukhnovsky, A. G. *Thin Solid Films* **1996**, *285*, 134.
- ³¹ Jelly, E. E. *Nature* **1936**, *138*, 1009.
- ³² Brooker, L. G. S.; White, F. L.; Heseltine, D. W.; Keyes, G. H.; Dent S. G.; VanLare E. J. *J. Photogr. Sci.* **1953**, *1*, 173.
- ³³ McRae, E. G.; Kasha, M. *J. Chem. Phys.* **1958**, *28*, 721.
- ³⁴ E.M. McCorquodale, C.L. Colyer, *Electrophoresis* **2001**, *22*, 2403.

- ³⁵ L. Tarazi, A. George, G. Patonay, L. Strekowski, *Talanta*, **1998**, *46*, 413.
- ³⁶ (a) Moody, E. D.; Viskari, P. J.; Colyer, Ch. L., *J. Chromatogr. B.* **1999**, *729*, 55; (b) McCorquodale, E. M.; Colyer, Ch. L., *Electrophoresis* **2001**, *22*, 2403.
- ³⁷ Slakter, J.S., Yannuzzi, L.A., Guyer, D.R., Sorenson, J.A., Orlock, D. A. *Curr. Opin. Ophthalmol.* **1995**, *6*, 25.
- ³⁸ Tang, B.; Yu, F.; Li, P.; Tong, L.; Duan, X.; Xie, T.; Wang, X. *J. Am. Chem. Soc.* **2009**, *131*, 3016.
- ³⁹ (a) A. Herz., *Photogr. Sci. Eng.* **1974**, *18*, 23; (b) A.K Chibisov and colleagues, *Chem Phys.* **1999**, *250*, 47.
- ⁴⁰ A. S. Tatikolov, S. M. B. Costa *Chem. Phys. Letters* **2001**, *346*, 233.
- ⁴¹ (a) V.I. Avdeeva, B.I. Shapiro, *Sci Appl. Photo.* **1998**, *39*, 543.
- ⁴² Kousik Kundu, Sarah F. Knight, Nick Willett, Sungmun Lee, W. Robert Taylor, Niren Murthy, *Angew. Chem. Intl Ed.* **2009**, *121*, 305.
- ⁴³ Graham T. Dempsey, Mark Bates, Walter E. Kowtoniuk, David R. Liu, Roger Y. Tsien and Xiaowei Zhuang, *J. Am. Chem. Soc.* **2009**, *131*, 18192.

CHAPTER ONE

SENSING AND TARGETING NUCLEIC ACID SEQUENCES AND STRUCTURES

I. DNA: Structure and function

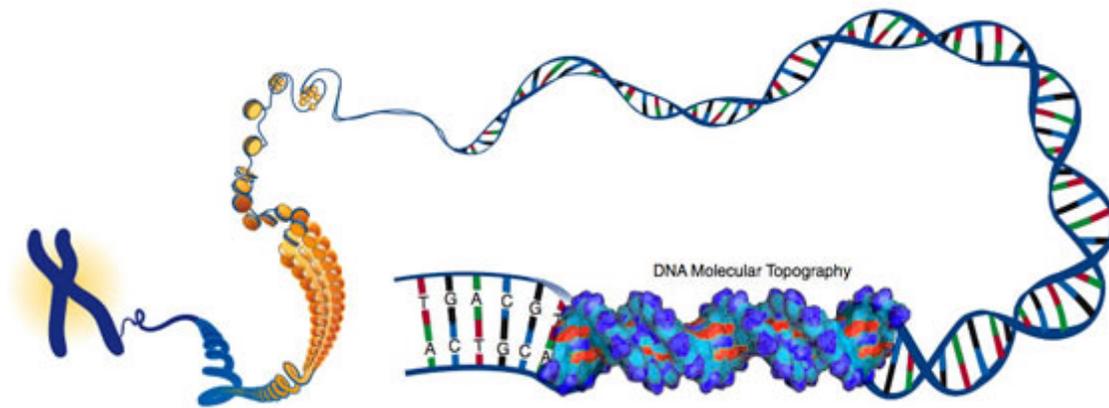


Figure 1: DNA Molecular topography extracted from ¹

DNA is a long polymer made from repeating units called nucleotides. A nucleotide is composed by a sugar, a phosphate backbone and a base:

- ✓ **Phosphate backbone**: It describes the association of individual nucleoside units joined together (within a nucleic acid strand) in a linear manner, through phosphate groups attached to the 3' and 5' positions of the sugars. The full repeating unit in a nucleic acid is a 3', 5'-nucleotide.
- ✓ **Sugar**: It is a 2'-desoxyribose sugar for DNA and a ribose for RNA. The glycosidic bond is that linking the sugar and the base and is always defined as the β -glycosidic bond in natural nucleic acids: the base is above the plane of the sugar when viewed onto the plane and therefore on the same face of the plane as the 5' hydroxyl substituent. The α -oligomers are more stable than the β -oligomers.

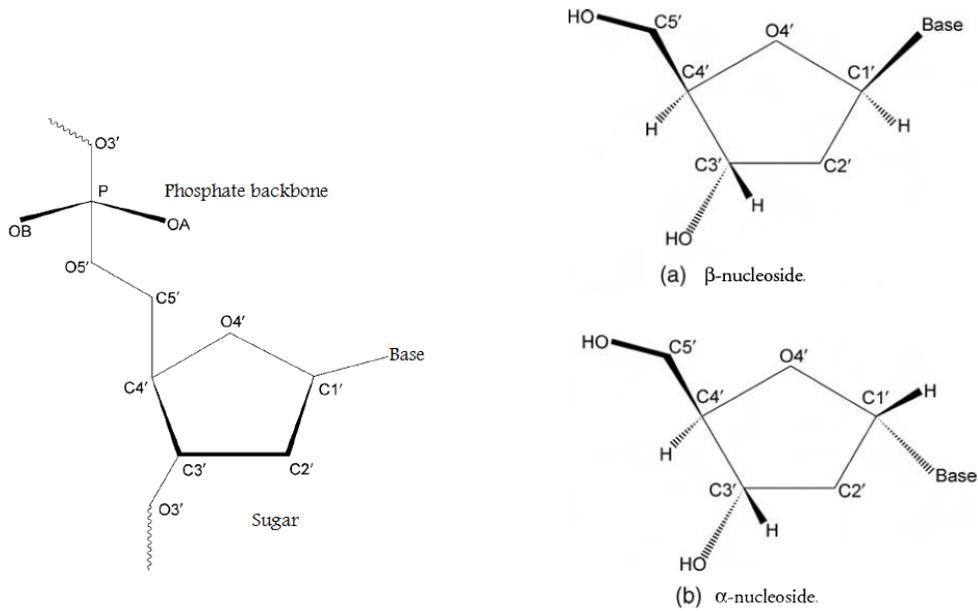


Figure 2: representation on the left of a nucleotide which is composed by a Phosphate backbone linked to a sugar (ribose for RNA and desoxyribose for DNA). On the left is represented the conformation of the sugar (α or β). Picture extracted from ²

- ✓ **Bases:** The bases are planar aromatic heterocyclic molecules and are divided into two groups: pyrimidines (i.e. Thymine and Cytosine) and purines (i.e. Adenine and Guanine).

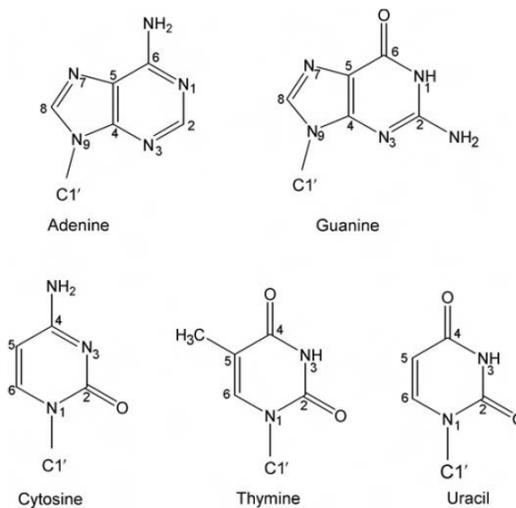


Figure 3: all base known for the nucleic acids (RNA and DNA)

Oligonucleotide sequences use single-letter codes for the five unit nucleotides: A, T, G, C and U standing for Adenosine, Thymidine, Guanosine, Cytidine and Uridine, respectively. A single oligonucleotide chain is conventionally oriented from the 5' to the 3'-end. The 5' terminal nucleoside has a free hydroxyl at its 5' position while the residue at the 3'-end has a free hydroxyl group at its 3' position.

1.1 Watson-Crick Base pairing

DNA is mainly present in cells under a double-helical conformation which was first proposed in 1953 by Watson and Crick. Within this DNA double-helix, purines associate with pyrimidines by specific hydrogen bonds, to form planar base pairs. While guanosines interact specifically with cytosines by three hydrogen bonds, thymidines interact specifically with adenosines via the formation of only two hydrogen bonds. This highly specific pairing rule of DNA nucleobases therefore occurs through specific hydrogen bonds which are commonly called Watson-Crick H-bonds.

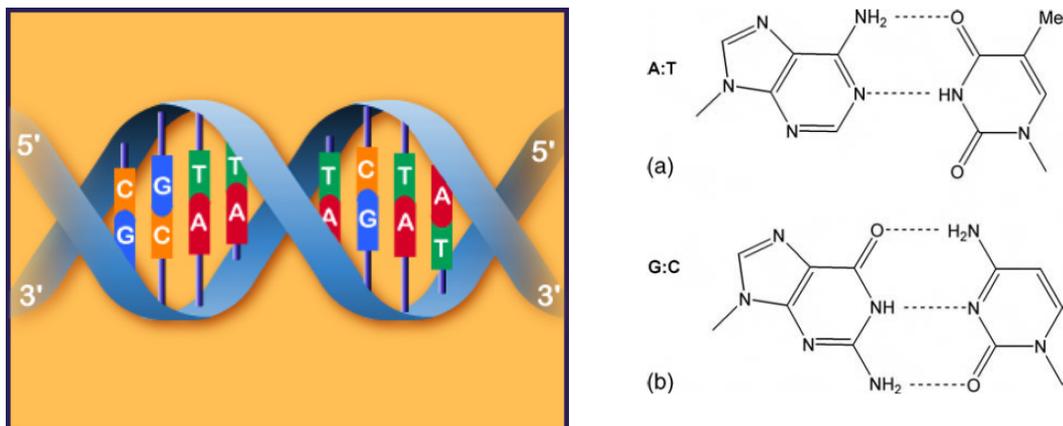


Figure 4: *schematic representation of Watson-Crick base pairs*

Hydrogen-bond distances (in Å) in Watson–Crick base pairs are summarized in the table below³:

U•A:	N3-H.....N1	2.835(8)
	O4.....H-N6	2.940(8)
C•G:	O2.....H-N2	2.86(1)
	N3.....H-N1	2.95(1)
	N4-H.....O6	2.91(1)

Based on molecular dynamics simulation studies free energies of -4.3 and -5.8 kcal/mol were found for A•T and G•C base pairs, respectively ⁴.

The intertwined strands make two grooves of different widths, referred to as the major groove and the minor groove. These minor and major grooves result from the asymmetry of the bases.

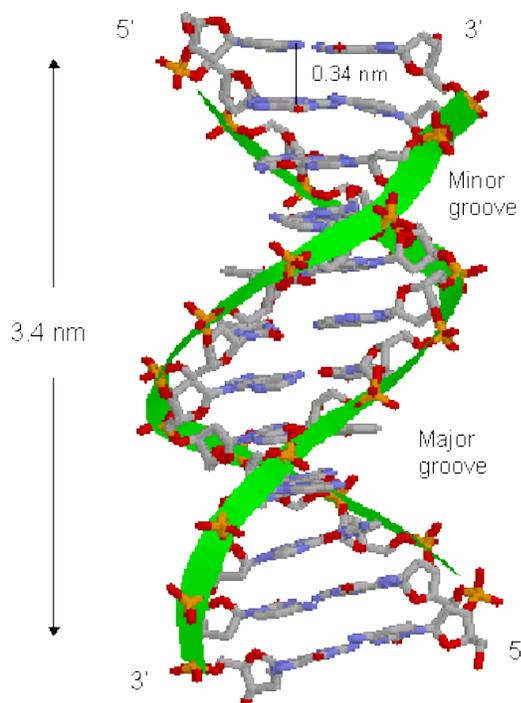


Figure 5: schematic representation of a DNA double helix

By convention, The C1'-N9 (purine) and C1'-N1 (pyrimidine) base sugar bonds are projected in the minor groove while the C6/N7 (purine) and C4 (pyrimidine) base atoms are in the major groove. The three-dimensional structure of the DNA depends of the individual base pair flexibility and the sugar flexibility.

1.2 DNA double-helical conformations

Three main types of DNA double-stranded structures have been elucidated: helices of type A, type B and type Z. The first structure of DNA described by Watson and Crick in 1953 was a B-form. The helix makes a turn every 3.4 nm, and the distance between two neighboring base pairs is 0.34 nm. Hence, there are about 10 pairs per turn. Double-stranded B-DNA displays systematic sequence-dependent modulations with changes in helical twist and roll angle, and alterations in sugar–phosphate torsion angles. The predominant furanose ring puckering mode is C2'-endo. B-form helix is the main biologically relevant secondary structure of DNA.

In a solution with higher salt concentrations or with alcohol added, the DNA structure may change to an A form, which is still right-handed, but every 2.3 nm makes a turn and there are 11 base pairs per turn. Another DNA structure is called the Z form, because its bases seem to zigzag. Z DNA is left-handed. One turn spans 4.6 nm, comprising 12 base pairs. The DNA molecule with alternating G-C sequences in alcohol or high salt solution tends to have such structure.

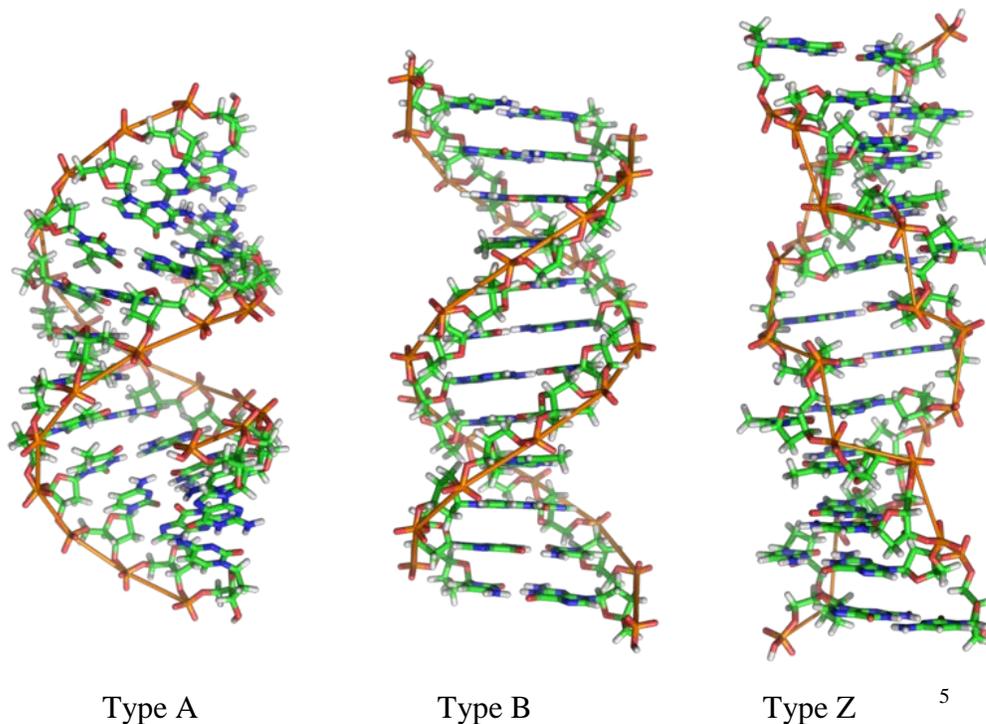


Figure 6: schematic representation of DNA double helices of Type A, B and Z

1.3 From DNA to RNA to proteins

Before a cell can divide, it must duplicate its entire DNA. This process is called DNA replication. In eukaryotes replication occurs during S phase of the cell cycle. DNA replication is semi-conservative: the double helix is unwound by helicases, and then DNA polymerase binds to one strand of the DNA and begins moving along it in the 3' to 5' direction, using it as a template for assembling a leading strand of nucleotides and reforming a double helix. In eukaryotes, this molecule is called DNA polymerase delta (δ). Because DNA synthesis can only occur from the 5' to 3' end, a molecule of a second type of DNA polymerase (epsilon, ϵ , in eukaryotes) binds to the other template strand as the double helix opens. This molecule must synthesize discontinuous segments of polynucleotides (called Okazaki fragments). Another enzyme, DNA ligase I then stitches these together into the lagging strand.

DNA is housed within the nucleus of our cells. It controls cellular activity by coding for the production of enzymes and proteins. The information contained in DNA is not directly converted into proteins, but must first be copied into messenger RNA. This step, so-called transcription, ensures that the information contained in the DNA does not become tainted. Since proteins are constructed in the cytoplasm of the cell by a process called translation, mRNA must cross the nuclear membrane to reach the cytoplasm. Once in the cytoplasm, mRNA along with ribosomes and another RNA molecule called transfer RNA (tRNA), work together to produce proteins. Proteins can be manufactured in large quantities because a single DNA sequence can be transcribed by many RNA polymerase molecules at once.

The processes of DNA transcription (DNA \rightarrow RNA) and RNA translation (RNA \rightarrow Protein) are schematically represented in the figure below⁶:

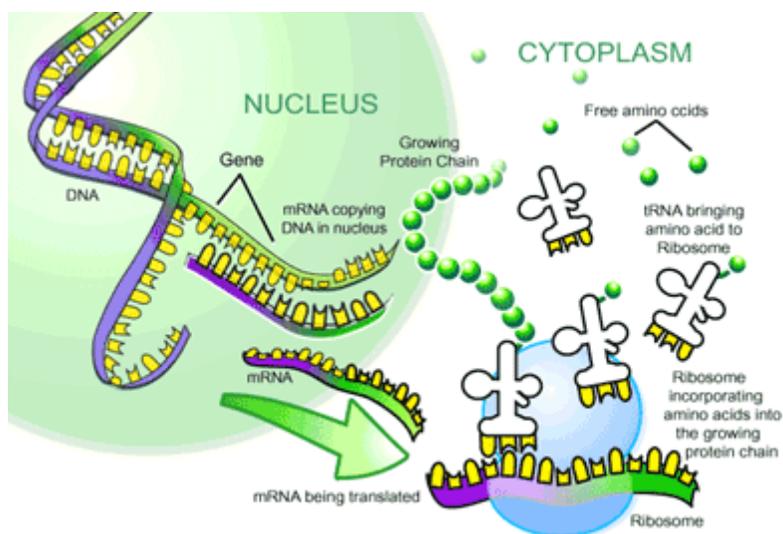


Figure 7: From DNA to proteins

The specificity of the A-T and G-C Watson–Crick hydrogen-bonding interaction also allows the convenient programming of artificial DNA receptor moieties through the simple four-letter alphabet. By automated methods, it is possible to synthesize all types of DNA⁷ or to amplify any DNA sequence (e.g. by the polymerase chain reaction (PCR) technique)⁸. A DNA double helix is highly flexible. Its flexibility depends on the strand length, DNA rigidity commonly increasing with its length/size. Below 150 base pairings, DNA can be considered as a rather rigid polymer⁹.

1.4 Alternative DNA secondary structures are also biologically relevant

Although existing mainly under its double-stranded form in the context of a cell, DNA can also adopt some alternative (non-B-DNA) structures, such as hairpins, cruciforms, three or four-way junctions or four-stranded G-quadruplexes¹⁰. The existence of such structures in vitro is now well-established and there is also increasing evidences supporting the existence in vivo.

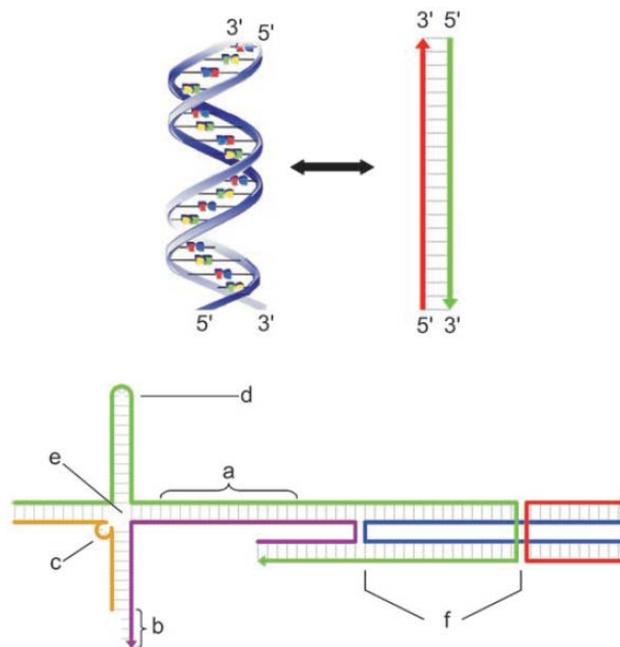


Figure 8: Selected examples of alternative DNA secondary structures: the choice of suitable DNA sequences allows the generation of complex motifs which contain a) double-helical regions, b) sticky ends, c) bulge loops, d) hairpin loops, e) junctions and f) crossovers. (Picture extracted from ref¹¹)

Within this manuscript we will focus on two types of structures commonly found in gene promoters and suspected to play a role in gene regulatory mechanisms: DNA G-quadruplexes and DNA hairpins. Indeed, the formation of stable DNA secondary structures in key regions of a gene can be considered as a potentially important component of regulation of gene expression that might be exploited for drug development.

Chromosomal DNA is composed of introns and exons. Exons are the sequence which mainly code for proteins. However upstream of the genes is a non-coding region which is critical for transcription regulation: the gene promoter. DNA transcription can therefore be activated or inhibited upon binding of specific proteins (e.g. transcription factors) to key sequences or structures contained in the promoter. Recent bioinformatics studies found an enrichment of the overlap between conserved DNA secondary structures

(CSSs) and the binding sites of transcription factors near transcription factor-encoding genes suggesting a role for the CSSs in transcriptional regulation networks¹².

In the section below, we will give a brief (and non-exhaustive) overview of the possible gene regulatory effects (at the transcriptional level) of hairpin and quadruplex DNAs when these structures are located in key areas of gene promoters.

II. Introduction to DNA Hairpins and Quadruplexes

II.1 Biological relevance of DNA hairpins

In eukaryotes, hairpins or cruciforms have also been implicated in control of transcription. One family of proteins that bind DNA secondary structures are the high mobility group (HMG) chromosomal proteins. These are a family of dsDNA- and ssDNA-binding proteins extractable from nuclei and chromatin, and, like histones, they are thought to have a role in chromosomal structure, function, and transcription. A unique feature of this family of proteins is that it includes proteins interacting with DNA in either a sequence-specific or a non-sequence-specific way. Although the proteins recognize DNA in two different ways, they have a homologous DNA binding domain called HMG-box. The structure of the domain was obtained in 1994 and presented a novel DNA binding motif. Another interesting feature of the HMG proteins is their increased affinity for distorted regions of DNA, which can be even larger than the natural affinity of the sequence-specific HMG proteins for their recognition sites. After binding to distorted DNA, HMG can recruit enhancer to regulate the transcription. For pig thymus HMG-1, it was shown that these HMG bind to the site sensible for nuclease (protection from S1 nuclease) and also remove a transcriptional block in the region of DNA containing a putative cruciform structure¹³. Both human and murine interferon- β have been proved to have a putative cruciform structure which have a high affinity for HMG-I and may regulate the transcription.

Beyond the HMG family, transcription regulatory roles have also been identified for an increasing number of eukaryotic ssDNA-binding proteins, and in several cases these proteins are thought to be binding to hairpin or cruciform structures. For instance,

II.2 Biological relevance of DNA quadruplexes

As discussed in the section above, gene promoters may contain features required to initiate and regulate the process of gene activation. The formation of a G-quadruplex structure in this location could therefore simply act as a steric block to transcription, by preventing the required transcription machinery from binding to the DNA target, or else by preventing it from proceeding along the DNA once bound. Other models can also be proposed for the effect of G-quadruplex formation in these regions, resulting in either increased or decreased transcription. A number of individual G-quadruplex motifs in such gene promoters have been studied, of which the best characterized is that found in the promoter of the oncogene c-myc. This oncogene has been implicated in a variety of cancers, including colorectal cancer, and is oncogenic when upregulated. It contains a nuclease hypersensitive site, NHE III1, which is responsible for approximately 80% of the total c-myc transcription¹⁹. The sequence in this region has been shown to form a number of possible G-quadruplexes, which have the effect of reducing transcription^{20, 21}. This has been the subject of much structural investigation^{22, 23} and it has been shown that stabilization of the G-quadruplex using the quadruplex ligand TMPyP4 further decreased transcription^{24, 25}. There is now increasing evidence that quadruplexes could also form in the promoter of a number of cancer-related genes such as K-ras²⁶, c-kit²⁷ and bcl2²⁸. Despite the high number of indirect evidences of the existence of G-quadruplexes in vivo (e.g. identification of quadruplex binding proteins, selection of specific antibodies raised against quadruplexes²⁹), there is still an urgent need for sensor capable of demonstrating the possible formation of such structures in the context of a cell.

In the section below, we will now define in more details the structural features of both Hairpin and Quadruplex DNAs.

III. DNA quadruplexes: structures and functions

III.1 Structure of DNA quadruplexes

The observation that guanosine 5'-monophosphate (GMP) forms helical aggregates at millimolar concentrations with four units per helical step dates back to the early work in the 1960s by Gellert and Davies ³⁰. Indeed, Gellert determined that the guanine (G) bases could form a tetrameric arrangement described as a G-quartet or G4-DNA arrangement. These consist of a square arrangement of four guanines, also called G-tetrad, stabilized by Hoogsteen hydrogen bonding. Each base is simultaneously a hydrogen bond acceptor and a hydrogen bond donor. For one quartet, the interaction of the atoms N1 and N2 from one side, O6 and N7 from the other side lead to the formation of 8 hydrogen bonds ³¹.

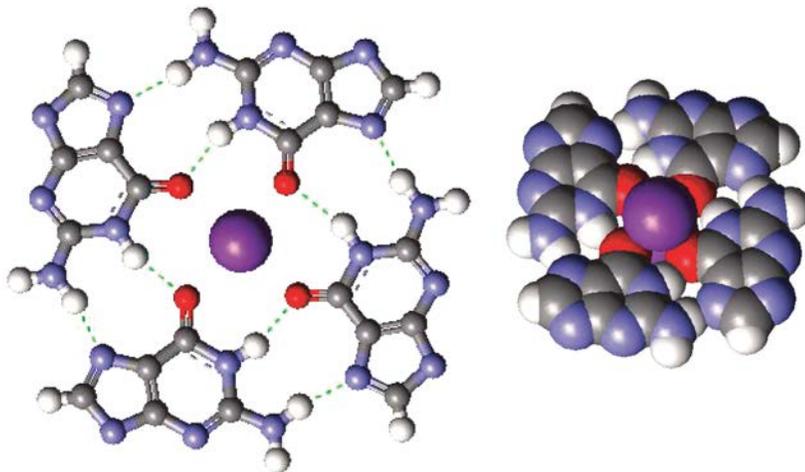


Figure 10: Structure of a tetrad of guanines or G-tetrad

While in double-stranded DNA, complementary nucleobases interact via formation of so-called Watson-Crick hydrogen-bonds, G-tetrads stability results from the strong interaction of 4 guanine residues via Hoogsteen hydrogen-bonds. This involves a total of eight hydrogen bonds between the “Watson–Crick” face of one guanine base and the “Hoogsteen” major-groove face of another, in a manner analogous to that observed in some G•G mismatched duplexes.

In the late 1980s, biochemists proved that oligonucleotides containing runs of three or four consecutive guanines (as found in telomeric DNA) could spontaneously form four-stranded DNA structures called G-quadruplex or G-tetraplexes ^{32,33}. The planar G-quartets stack on top of each other, giving rise to four-stranded helical structures

leading to a similar appearance of DNA duplex characterized by a regular rise and twist between the tetrad planes and generating a right-handed helical twist. Finally, the phosphate backbones, linking the nucleosides together, generate four grooves of variable width, instead of two. Since the early 1990s, formation of these structures was demonstrated using IR, X-Ray, NMR and CD spectroscopy.

The formation and stability of quadruplexes is strongly dependent on monovalent cations such as K^+ and Na^+ and hence, physiological conditions favor their formation. The cations play a key role by forming ionic interactions with the O6 atoms of the tetrad guanines. At one extreme an ion can be positioned symmetrically in the plane of four guanines in a tetrad. Potassium ions tend to be situated symmetrically between two consecutive tetrads so that they are coordinated to all eight O6 atoms from both tetrads. This arrangement is favored by potassium and ammonium ions. The smaller sodium ion is able to coordinate all four O6 atoms in a single tetrad while being in the plane of all four guanines, although it can also be accommodated midway between tetrads. A continuum between both extreme positions can be seen in several crystal and NMR structures. Numerous 3D structural analyses have revealed that G-quadruplex DNA structures are highly polymorphic and can be sub-grouped into various families, such as parallel or antiparallel, according to the orientation of the strands, or inter- or intramolecularly folded^{34,35}. Once formed, G-quadruplex DNA structures are much more stable than double-stranded DNA (20–30°C higher melting temperatures), suggesting that their unfolding into the linear form would require the aid of specific proteins.

III.2 Quadruplexes in the genome

Quadruplex formation was first proposed at 3' end of telomeric DNA. Telomeres protect DNA from nuclease damage and also from unwanted recombination in the chromosome. Telomeric DNA comprises a long double-stranded region (in the case of humans, it is of sequence $(TTAGGG)_n$ hybridized to its complementary sequence) but also a single-stranded 3'-end overhang (about 100–200 nucleotides long and of similar sequence $(TTAGGG)_n$) which is associated with various proteins (constituting the

sheltering complex) including the single-strand binding protein POT1³⁶. Single-stranded telomeric DNA in the absence of protein can fold back on itself and then dimerize to form four-stranded hairpin loops that can be stabilized^{37,38} by the formation of the guanine tetrad motif outlined above. It can also form intramolecular structures by repeated foldbacks. These G-tetrad-containing structures are termed human telomeric quadruplexes or tetraplexes.

More generally, intramolecular quadruplexes can potentially be formed by any short-length oligonucleotide containing four clusters of at least three consecutive guanines, each cluster being separated by a variable number of nucleotides (from 1 to 7). In recent years, combined computational and experimental approaches have allowed the identification of highly stable quadruplex forming sequences in the promoter regions of genes of clinical interests and it has been proposed that such structures could potential serve as gene regulatory elements both in vitro and in vivo. (see previous section).

III.3 Quadruplex topologies

The four guanosine nucleosides in an individual tetrad can in principle exist in either anti or syn glycosidic angle conformations, and thus there are 16 possible combinations. Antiparallel quadruplexes can have different orientation for the four backbones.

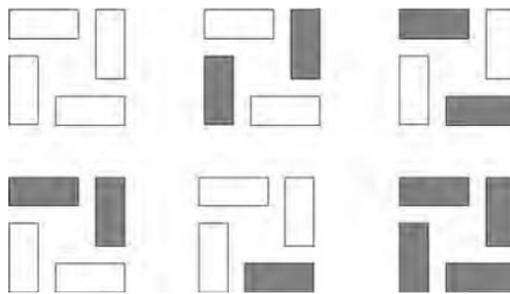


Figure 11: Schematic showing 6 of the 16 possible arrangements for the glycosidic angles of individual guanines in a G-tetrad. Shaded rectangles represent syn conformations, and unshaded rectangles represent anti arrangements.

Depending on the number of short G-tracts, there are different ways quadruplexes can form. Four single strands containing at least three consecutive guanines can assemble in a tetramolecular manner. Whether the strands are parallel or antiparallel to each other depends in part on the nature and length of the sequence, and sometimes as well as on the counter-ion used. An all-parallel orientation for all four strands, as in the tetramolecular quadruplex shown schematically in the above figure (a), requires all glycosidic angles to adopt an anti conformation.

Two single strands can also assemble to form dimeric or bimolecular quadruplexes. Different types of associations are also possible depending conformations adopted by the various glycosidic angles (syn or anti). More interestingly on a biological point of view is the fact that G-rich single DNA strands can also fold back on themselves to form a so-called intramolecular quadruplex. Intramolecular quadruplex can also adopt different topologies (parallel, anti-parallel, or mixed structures) but are more likely to exist in vivo than their tetra- and bi-molecular analogues since capable of forming even at very low DNA concentration. Few examples of structures are given below for illustration.

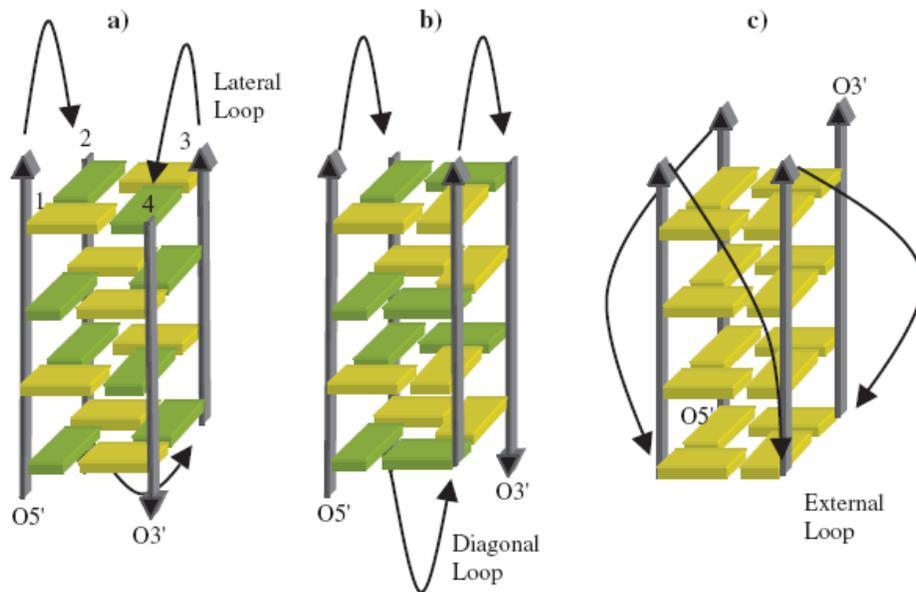


Figure 12: (a) Lateral loops connecting anti-parallel strands arranged on the same face; (b) mixed lateral and diagonal loops connecting anti-parallel strands. (c) External chain reversal loops connecting parallel strands together.

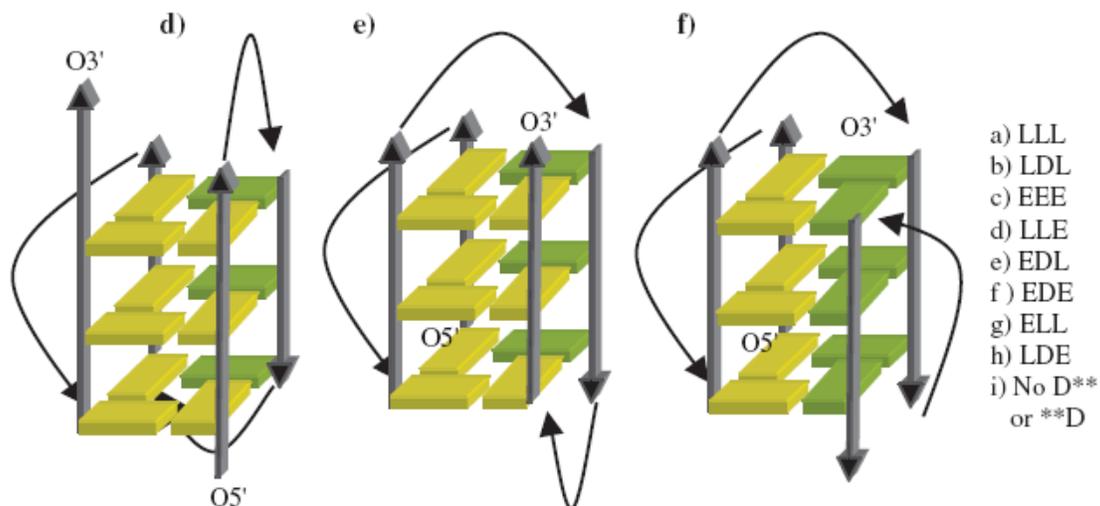


Figure 12bis: (d) Mixed external and lateral loops connecting strands; (e) Mixed external, diagonal, and lateral loops connecting strands together; (f) Mixed external and diagonal loops connecting strands together. (where L= Lateral loop, E= External (double chain reversal) loop, and D=Diagonal loop, *=any type of loop)

III.4 Quadruplex sensors

To date there only exist very few examples of chemical sensors or probes used for detection the formation of unique quadruplexes both in vitro and in vivo. This would however be very useful in order to validate the existence of such structures in vivo and in order to elucidate their potential cellular function(s). We are listing below few representative examples of quadruplex fluorescent sensors that have been recently reported in the literature.

In 2001, Stanton *et al.*³⁹ have developed a molecular beacon for the detection of thrombin binding, using a DNA aptamer labeled at its 5'- and 3'-ends with fluorescein (fluorophore) and Dabcyl (quencher), respectively. The addition of the thrombin to this DNA induces a transition from a stem-loop structure to a G-quadruplex structure, thus modifying the distance between the fluorophore and the quencher, which can be monitored by fluorescence spectroscopy.

Tanenaka *et al.*⁴¹ developed the “PSO-py oligonucleotide” for potassium sensing using an oligonucleotide of similar sequence than that used proposed by Stanton *et al* but functionalized at both ends with a molecule of pyrene. The addition of potassium cations to this specific sequence of DNA induces quadruplex folding thus allowing interaction between both molecules of pyrene. In the absence of K^+ a weak monomer emission at 390 nm is observed while after addition of K^+ pyrene forms an excimer by π -stacking which emits strongly at 480 nm. Formation of the excimer can be detected by fluorescence spectroscopy which correlates with the G-quadruplex formation (as proved by CD), which correlates also with potassium concentration.

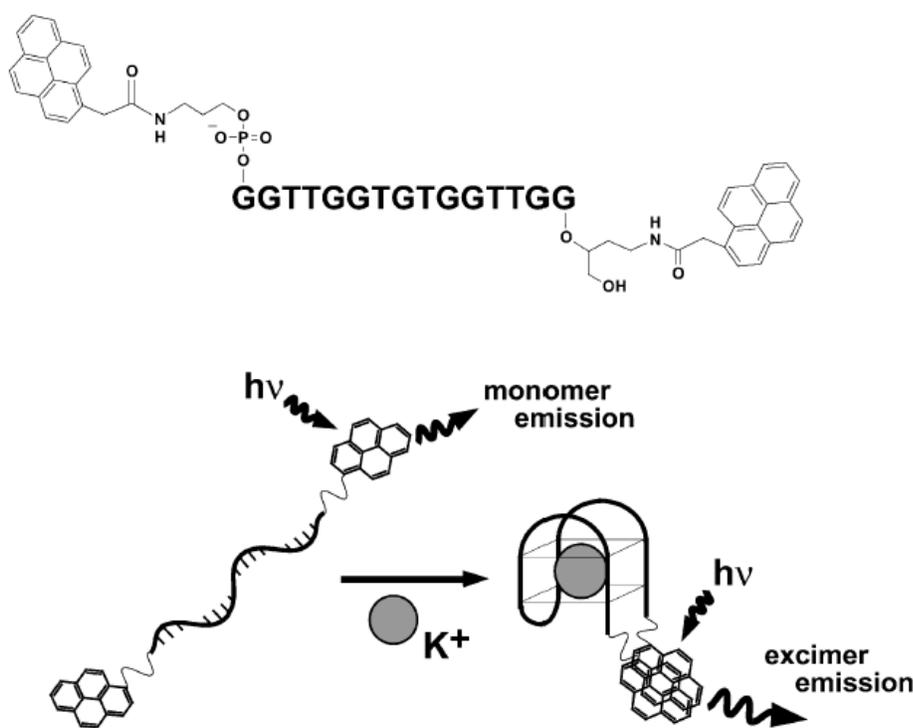


Figure 15: *G-quadruplex formation induced by K^+ binding and detected by pyrene excimer formation*

IV. DNA hairpins: structures and functions

Hairpin loop structures in nucleic acids consist of a base-paired stem structure and a loop sequence with unpaired or non-Watson-Crick-paired nucleotides. These common

structural motifs can be of functional importance as ligand recognition elements or folding initiation sites.

DNA replication, transcription, and/or DNA repair processing lead to unwinding of the duplex, affording the single-stranded repeat sequence the opportunity to base pair with itself in an intramolecular fashion, thereby forming a hairpin or cruciform structure. A 14 bp palindrome (7 bp per inverted repeat) is sufficient for formation of a stable hairpin *in vivo* with the negative supercoiling tension providing the energy for structure formation. Hairpin (or cruciform) structures are regarded as a source of genomic instability in both prokaryotic and eukaryotic cells.

Molecular beacons are oligonucleotides bearing a fluorophore and a proper quencher linked to their extremities. The sequence is designed to shape, at rest, a hairpin whose loop is available to bind a complementary single strand or duplex (target). Heating the hairpin or hybridizing it with its target, its conformation changes moving away the fluorophore and the quencher to re-enable the original fluorescence of the isolated fluorophore.

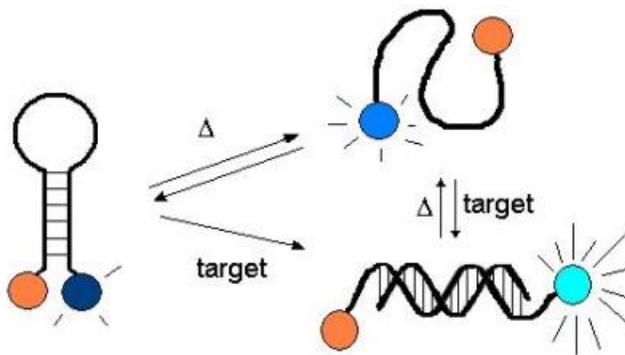


Figure 16: Schematic representation of a molecular beacon

Balasubramanian *et al.* for instance studied the non-Arrhenius kinetics for the loop closure of a DNA hairpin⁴². Donor fluorophore carboxytetramethylrhodamine (TMR) is attached at its 3'-end and Acceptor fluorophore indodicarbocyanine (Cy5) is attached at its 5'-end via a three-carbon linker. They got informations from the FRET about thermodynamical and kinetical parameters for the conformational transition.

MBs are now widely used in areas such as genetic screening, biosensor development, biochip construction, the detection of single-nucleotide polymorphisms (SNPs), and messenger-RNA (mRNA) monitoring in living cells. This broad spectrum of applications can be attributed to the particular way in which these probes interact with DNA, RNA, and protein molecules. The unique stem-loop structure and fluorophore-quencher pair enable the detection of target nucleic acids to be reported in real time with excellent sensitivity and selectivity. Nevertheless, there are still many challenges associated with the development of MB probes, for example, in the design and engineering of MBs for optimal use in the intracellular monitoring of gene expression. However, these aspects are beyond the scope of this manuscript and will not be developed here.

V. Introduction to Peptide Nucleic Acids PNAs)

The organic chemist Prof. Ole Buchardt's laboratory from Copenhagen and the Biochemist Peter Nielsen have developed new nucleic acid sequence-specific reagents: Peptide nucleic acids (PNAs). PNAs are DNA and RNA analogues in which the negatively charged ribose-phosphate backbone has been substituted by a neutral peptidic (i.e. N-(2-aminoethyl)glycine) backbone⁴³:

A methyl carbonyl linker connects natural (as well as unusual in some cases) nucleotide bases to this backbone. PNAs are non-ionic, achiral molecules and are not susceptible to hydrolytic (enzymatic) cleavage but with a very low cellular permeability (the disruption of the cellular membrane by electroporation or pore forming agents is an alternative for penetration of the oligonucleotides). PNA is capable of sequence-specific recognition of DNA and RNA obeying the Watson-Crick hydrogen bonding scheme, and the hybrid complexes exhibit extraordinary thermal stability and unique ionic strength effects. PNAs hybridize preferentially in the antiparallel mode (amino terminus of the PNAs facing the 3'-end of the oligonucleotide). It may also recognize duplex homopurine sequences of DNA to which it binds by strand invasion, forming a stable PNA-DNA-PNA triplex with a looped-out DNA strand.

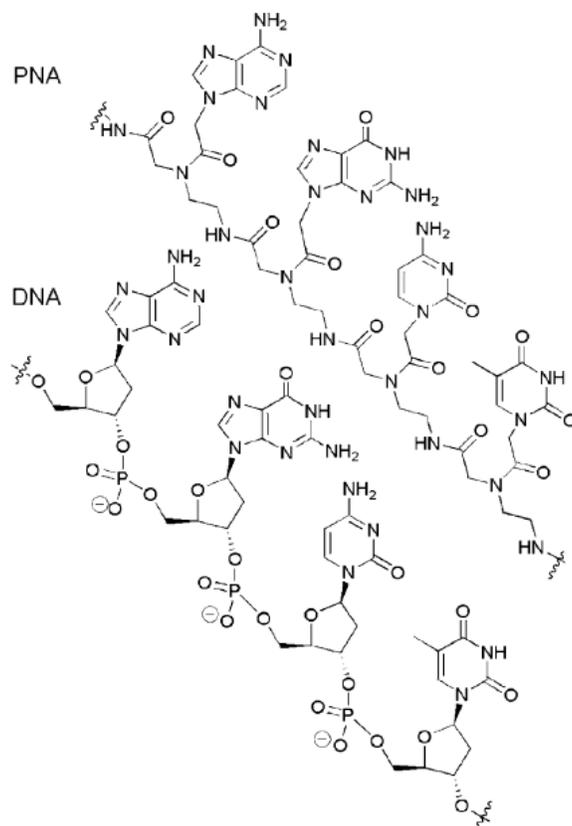


Figure 17: Structural comparison between DNA and PNA. Picture extracted from ref⁴⁴

PNA oligomers can be prepared following standard solid-phase synthesis protocols for peptides⁴⁵ using, for example, a (methylbenzhydryl) amine polystyrene resin as the solid support⁴⁶. The scheme for protecting the amino groups of PNA monomers is based on either Boc or Fmoc chemistry⁴⁷.

Amino acids can be coupled during solid-phase synthesis or compounds containing a carboxylic acid group can be attached to the exposed amino-terminal amine group to modify PNA oligomers. 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.

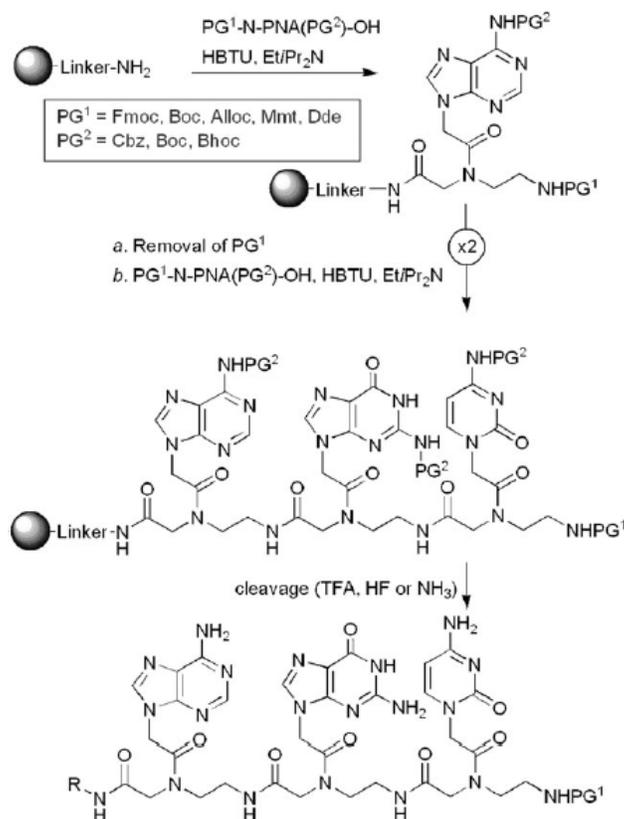


Figure 18: General strategy for the solid-phase synthesis of PNAs

In the figure above is summarized a typical example of solid-supported PNA synthesis. A N-(protected)-PNA-OH is coupled to the resin in presence of a coupling reagent such as HATU, EDC, PyBop and in the presence of base (piperidine, Hunig's base). After a cycle of orthogonal coupling deprotection, the PNA is released from the resin with TFA, HF or NH_3 depending of the nature of the resin.

VI. Use of PNAs and DNAs in Oligonucleotide-Templated Reactions

The biomedical importance of point mutations has generated a substantial demand for methods that detect nucleic acids with single nucleotide specificity. Oligonucleotide templated reactions are becoming an important tool for detecting oligonucleotide sequences⁴⁸. The binding strength of oligonucleotide complexes can be conveniently

tuned by varying the type of nucleic acids and the lengths of the oligomers. This strategy exploits the target strand as a template for a chemical reaction between two functionalized DNAs (or DNA analogues). These kinds of systems consist of two DNA probes that bind adjacently to the target oligonucleotide. Interestingly, this kind of approach can be developed in combination with fluorogenic probes to obtain a detection system based on fluorescence on/off switch-mechanisms. The probes hybridize to adjacent sites on the target sequence, increasing the effective concentration of the attached functionalities, and accelerating the reaction.

The supramolecular organization and reaction of two species templated by nucleic acids have a lot of applications. The design of reactions for which the yield of fluorescent products gives information about the detection of specific nucleic acid sequences⁴⁹ has received particular attention in recent years. In some cases, nucleic acid templated reactions can provide a highly sensitive detection method if acting catalytically. For instance, Ma and Taylor showed complementary DNA-templated hydrolysis of a para-nitrophenol ester fixed on one DNA probe by an imidazole beared by another DNA in 5'.⁵⁰

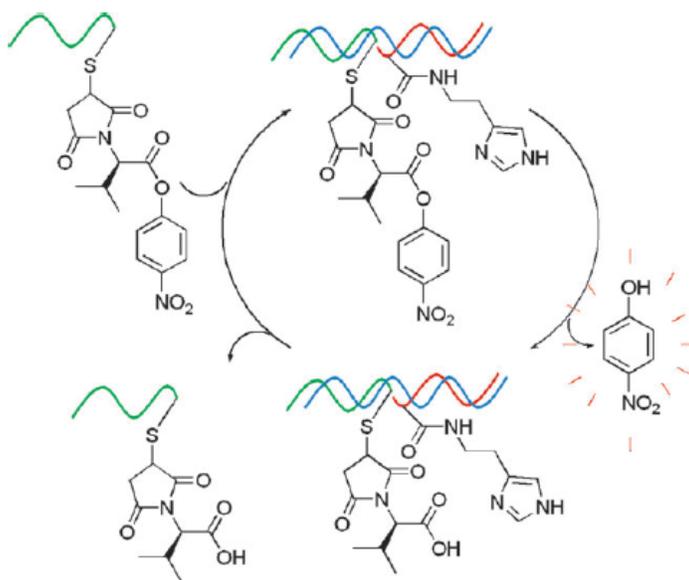


Figure 19: Mechanism of a DNA-templated hydrolysis of p-nitrophenol ester

The Taylor group reported a second generation of aza-ylide DNA-templated reaction using a Staudinger reaction leveraged on the reactivity of the aza-ylide between two PNA strands to hydrolyze a quenched fluorescein ester. They have developed a system with a good sensibility at 1 μM under physiological condition.

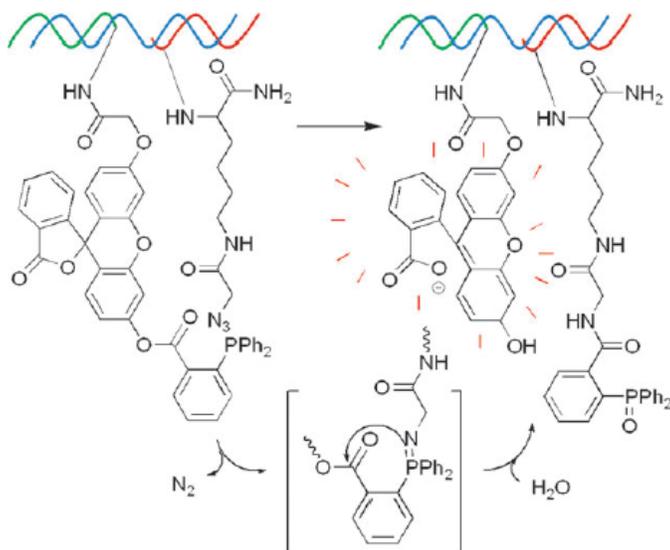


Figure 20: Mechanism of a DNA-templated Staudinger reaction

Winssinger *et al.* recently developed some PNA probes based on the bioorthogonal Staudinger reaction with catalytic template turnover based on an azide-quenched coumarin fluorophore⁵¹ using low nM concentration of DNA.

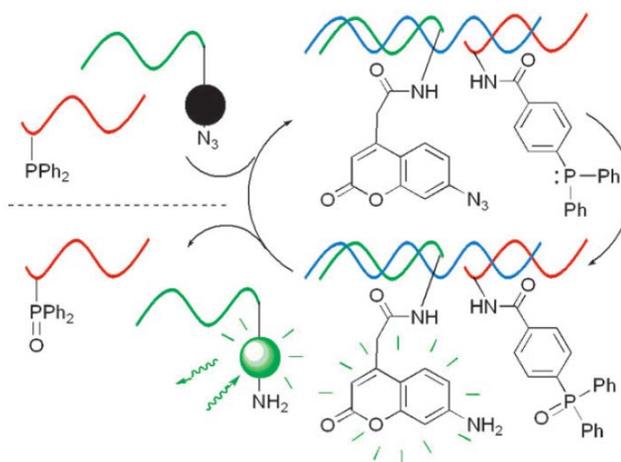


Figure 21: Mechanism of a Staudinger reaction with catalytic turnover.

This system was used for the screening of DNA mutations. A one nucleotide mutation located near to the site of the reaction could have a consequence on the yield of the reaction. Indeed, a mutation near to the site of the reaction (first base before the reactant of the reaction) leads to a partial hybridization of one of the two DNA strands (or of both) thus decreasing the efficiency of the reaction and the fluorescence yield.

Kool *et al.* have developed an alternative system based on a nucleophilic thiophosphonate reaction. After hybridization to a nucleic-acid template, the release of a fluorophore quencher (Dabsyl) which was fixed on the same DNA strand than the fluorophore group enables detection of the fluorescence emitted by the fluorophore⁵²:

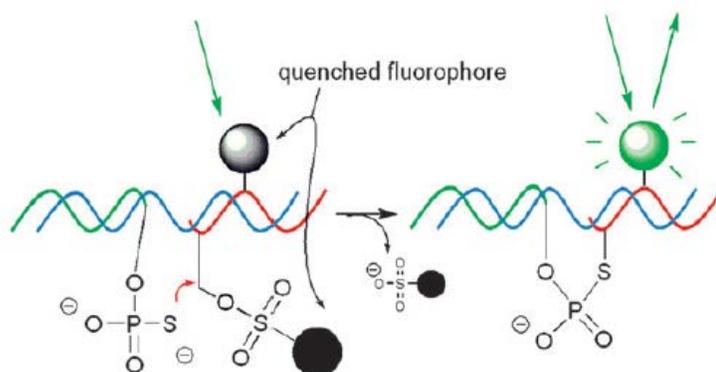


Figure 22: Mechanism of a DNA templated fluorogenic reaction based on fluorescence quencher release.

More recently, Ito *et al.* have developed fluorogenic probes having a dinitrobenzenesulfonyl (DNs) group protected by an aminocoumarin dye triggered by a nucleophilic aromatic substitution (S_NAr) reaction⁵³. One probe has an electrophilic DNs-protected 7-amino-4-methyl-3-coumarinylacetic acid (AMCA) dye at the 5'-terminal, while the other probe has a nucleophilic phosphorothioate group at the 3'-terminal. The chemistry of this strategy involves a S_NAr reaction between the DNs group of the nonfluorescent coumarin derivatives and the phosphorothioate group on the

oligonucleotide template. This reaction requires a nucleophilic addition of the sulfur from the phosphorothioate followed by an elimination of the aminocoumarin, thus releasing a free and highly fluorescent aminocoumarin with a Meisenheimer intermediate described in the Figure 23 below:

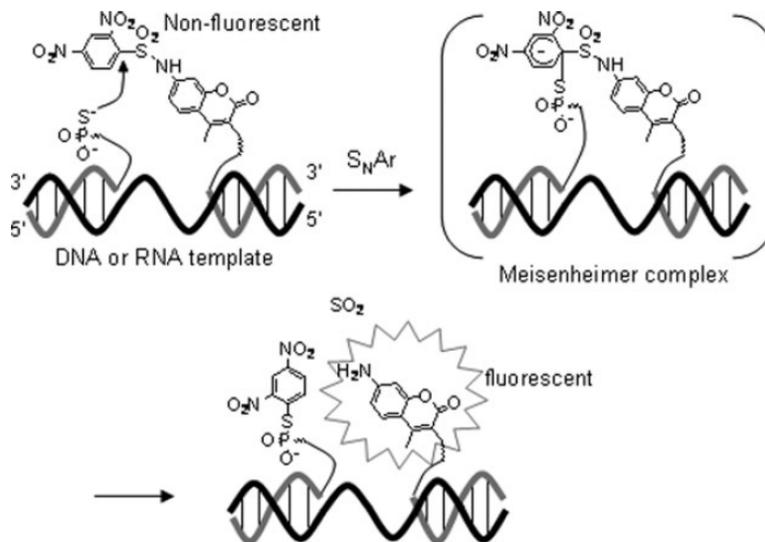


Figure 23: Example of DNA-templated fluorogenic reaction based on aminocoumarin release.

Seitz⁵⁴ *et al.* have also developed a system with a good turnover dealing with the transfer of one fluorophore quencher (the Dabsyl also) from one DNA strand having one fluorophore (Rhodamine) in the 5'-end to another probe containing a fluorescein in the 3'-end by using a trans-thioesterification reaction to another DNA strand on which a fluorescein is fixed upon hybridization to the template. A shift of wavelength is observed with this ping-pong DNA exchange.

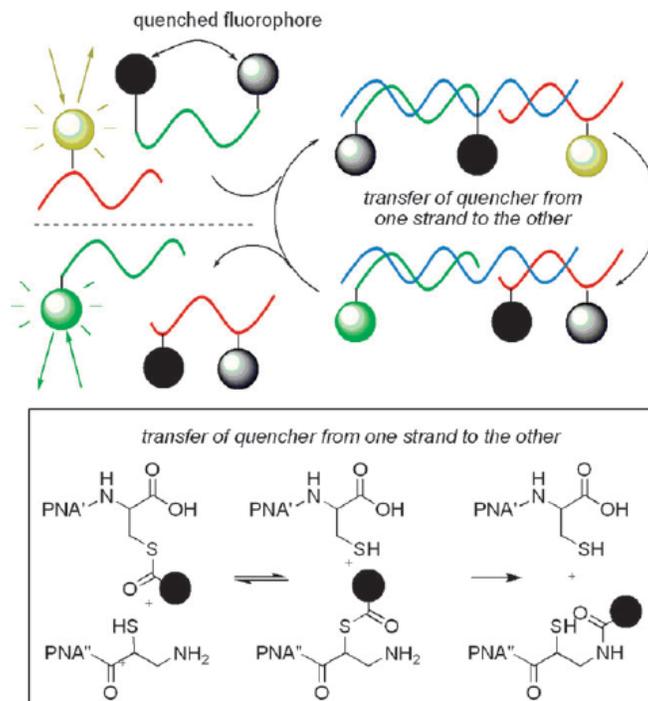


Figure 24: Mechanism of ping-pong DNA fluorophore exchange reported by Seitz and co-workers

Last example of this non-exhaustive list of reactions, DNA-mediated delivery of an organomercury activator to unmask a fluorophore was recently described by Kool *et al.*⁵⁵ The main drawback of this approach is that it uses highly toxic organomercury compounds and is therefore non-biocompatible.

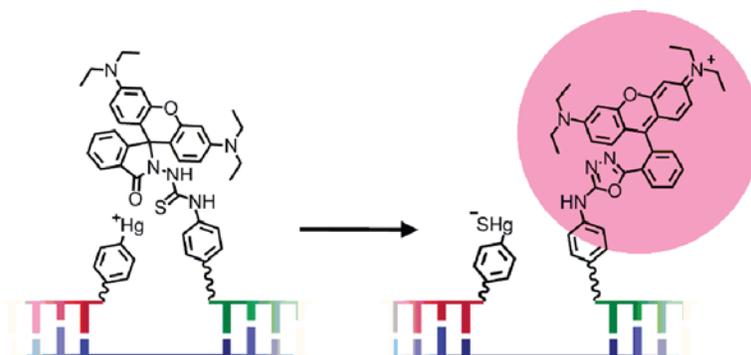


Figure 25: Mechanism of a DNA-templated organomercury-catalysed fluorogenic reaction as developed by Kool and coworkers.

VII. PUBLICATIONS

Herein, we have developed some original PNA-based fluorescent and fluorogenic⁵⁶ probes for sensing either oligonucleotide sequences (i.e. Single Nucleotide Polymorphism detection) or oligonucleotides structures (i.e. G-quadruplex and Hairpin DNAs). Our system relies on a fluorogenic reaction of trimethine cyanine dye formation from two non- or weakly-fluorescent precursors: a 2-methylene-indolenin and an indole Fisher's base aldehyde.

5-mer Peptide Nucleic Acids were synthesized using solid phase chemistry and were functionalized at their N- or C-terminus with one of the fluorogenic probes. The system was designed so that, at low PNA concentration, no reaction occurs and therefore no fluorescence can be detected. However, upon **simultaneous** binding of both PNAs to a carefully designed DNA target, the aldehyde and indolenin are positioned favorably with respect to each other to form the fluorescent cyanine dye. Appearance of the characteristic fluorescence signal can therefore be directly correlated to the presence/detection of the DNA target.

Compared to previous systems reported in the literature (see previous section), our approach offers the advantage of being biocompatible since it requires no catalysis and both reactants and products are stable in physiological conditions. Moreover, it is also highly versatile since it is possible to tune the fluorescence properties of the dye formed by modified the nature of the heterocyclic probe heads (e.g. replacement of the indolenine by a benzothiazole or a quinoline...).

This system was successively used for the first time for detecting nucleic acid secondary structures with high sensitivity and also high selectivity. Detection of unique parallel-stranded DNA quadruplexes and unique DNA-hairpins for instance were achieved with our PNA probes. Simultaneous sensing of both structures using a dual-colour system was also accomplished successfully in vitro. These probes could therefore find very valuable application for the detection of such structures in vivo and therefore for elucidating their biological function. Finally, the same probes were employed for detection of single nucleotide polymorphisms, although the efficiency of this reaction

will still require optimization. The three articles that describe in great details these three studies are given below.

REFERENCES

- ¹ <http://www.sciencedaily.com/images/2009/03/090312140852.jpg>
- ² S. Neidle, *Principles of Nucleic Acid Structure*, Acad. Press, Elsevier, Oct. 2007.
- ³ Seeman N.C., Rosenberg J.M., Rich A. *Proc Natl Acad Sci USA*. **1976**, 73, 804.
- ⁴ Stofer E., Lavery R. *Biopolymers* **1994**, 34, 337.
- ⁵ http://commons.wikimedia.org/wiki/File:A-DNA,_B-DNA_and_Z-DNA.png
- ⁶ http://www.scq.ubc.ca/wp-content/translation_01.gif
- ⁷ M. H. Caruthers, *Science* **1985**, 230, 281.
- ⁸ K. B. Mullis, *Angew. Chem. Int. Ed.* **1994**, 33, 1209.
- ⁹ J. F. Marko, S. Cocco, *Phys. World* **2003**, 16, 37.
- ¹⁰ U. Feldkamp and C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2006**, 45, 1856.
- ¹¹ See ref ix
- ¹² Xie, H.B., Irwin, D.M., Zhang, Y.P. *BMC Genomics*, **2008**, 9, 520.
- ¹³ Waga, S.; Mizuno, S.; Yoshida, M. *J. Biol. Chem.* **1990**, 265, 19424.
- ¹⁴ Potaman VN, Shlyakhtenko LS, Oussatcheva EA, Lyubchenko YL, Soldatenkov VA., *J. Mol. Biol.* **2005**, 348, 609.
- ¹⁵ Glucksmann, M. A.; Markiewicz, P.; Malone, C.; Rothman-Denes, L. B. *Cell* **1992**, 70, 491.
- ¹⁶ (a) Glucksmann-Kuis, M. A.; Dai, X.; Markiewicz, P.; Rothman-Denes, L. B. *Cell* **1996**, 84, 147; (b) Dai, X.; Kloster, M.; Rothman-Denes, L. B. *J. Mol. Biol.* **1998**, 283, 43.
- ¹⁷ Dai, X.; Rothman-Denes, L. B. *Genes Dev.* **1998**, 12, 2782.
- ¹⁸ South, T. L.; Summers, M. F. *Protein Sci.* **1993**, 2, 3.
- ¹⁹ Bossone SA, Asselin C, Patel AJ, Marcu KB. *Proc Natl Acad Sci USA*. **1992**, 89, 7452.
- ²⁰ Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. *Acta Cryst.* **1998**, 905.

- ²¹ A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. USA*. **2002**, *99*, 11593.
- ²² Phan A.T., Modi Y.S., Patel D.J., *J. Am. Chem. Soc.* **2004**, *126*, 8710.
- ²³ Ambrus A. *et al. Biochemistry* **2005**, *44*, 2048.
- ²⁴²⁴Seenisamy J, Rezler E.M, Powell T.J, Tye D, Gokhale V, Joshi C.S,Siddiqui-Jain A, Hurley L.H. *J. Am. Chem. Soc.* **2004**, *126*, 8702.
- ²⁵ Grand C.L, Han H, Muñoz R.M, Weitman S, Von Hoff D.D, Hurley L.H, Bearss D.J *Mol. Cancer Ther.* **2002**, *1*, 565.
- ²⁶ Cogo S., Xodo L. E., *Nucleic Acids Res.* **2006**, *34*, 2536.
- ²⁷ Rankin *et al.*, *J. Am. Chem. Soc.* **2005**, *127*, 10584.
- ²⁸ Dexheimer T. S., Sun D., Hurley L. H., *J. Am. Chem. Soc.* **2006**, *128*, 5404.
- ³⁰ Gellert, M. *et al. Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013.
- ³¹ Stephen Neidle and Shankar Balasubramanian, *Quadruplex Nucleic acids*, **2006**, Published by The Royal Society of Chemistry.
- ³² Sen, D. and Gilbert, W. *Nature* **1988**, *334*, 364.
- ³³ Sundquist, W.I. and Klug, A. *Nature* **1989**, *342*, 825.
- ³⁴ Patel, D.J. *et al. Nucleic Acids Res.* **2007**, *35*, 7429.
- ³⁵ Burge, S. *et al. Nucleic Acids Res.* **2006**, *34*, 5402.
- ³⁶ Lei, M.; Podell, E.R.; Cech, T.R. *Nature Struct. Mol. Biol.* **2004**, *11*, 1223.
- ³⁷ Sen D, Gilbert W, *Nature* **1988**, *334*, 364.
- ³⁸ Sundquist, W. I., & Klug, A. *Nature* **1989**, *342*, 825.
- ³⁹ N. Hamaguchi, A. Ellington and M. Stanton, *Anal. Biochem.* **2001**, *294*, 126.
- ⁴⁰ (a) J.W.J. Li, X.H. Fang and W.H. Tan, *Biochem. Biophys. Res. Commun.* **2002**, *292*, 31; (b) C.J. Yang, S. Jockusch, M. Vicens, N.J. Turro and W. Tan, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17278.
- ⁴¹ S. Nagatoishi, T. Nojima, B. Juskowiak and S. Takenaka, *Angew. Chem. Int. Ed.* **2005**, *44*, 5067.
- ⁴² M. I. Wallace, L. Ying, S. Balasubramanian and D. Klenerman, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5584.

- ⁴³ (a) 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; P. E. Nielsen, *Acc. Chem. Res.* **1999**, 32, 624; P. E. Nielsen, *Mol. Biotechnol.* **2004**, 26, 233.
- ⁴⁴ Z. L. Pianowski, N. Winssinger, *Chem. Soc. Rev.* **2008**, 37, 1330.
- ⁴⁵ (a) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, 85, 2149; (b) Merrifield, B. *Science* **1986**, 232, 341.
- ⁴⁶ (a) Dueholm, K., Egholm, M., Behrens, C., Christensen, L., Hansen, H. F., Vulpius, T., Petersen, K. H., Berg, R. H., Nielsen, P. E., and Buchardt, O. *J. Org. Chem.* **1994**, 59, 5767; (b) Haaima, G., Lohse, A., Buchardt, O., and Nielsen, P. E. *Angew. Chem. Int. Ed.* **1996**, 35, 1939.
- ⁴⁷ (a) Christensen, L., Fitzpatrick, R., Gildea, B., Petersen, K. H., Hansen, H. F., Koch, T., Egholm, M., Buchardt, O., Nielsen, P. E., Coull, J., and Berg, R. H. *J. Peptide Sci.* **1995**, 3, 175; (b) Norton, J. C., Waggenspack, J. H., Varnum, E., and Corey, D. R. *Bioorg. Med. Chem.* **1995**, 3, 437; (c) Thomson, S. A., Josey, J. A., Cadilla, R., Gaul, M. D., Hassman, C. F., Luzzio, M. J., Pipe, A. J., Reed, K. L., Ricca, D. J., Wiethe, R. W., and Noble, S. A. *Tetrahedron* **1995**, 51, 6179; (d) Norton, J. C., Waggenspack, J. H., Varnum, E., and Corey, D. R. *Bioorg. Med. Chem.* **1995**, 3, 437.
- ⁴⁸ (a) A. P. Silverman and E. T. Kool, *Chem. Rev.* **2006**, 106, 3775; (b) T. N. Grossmann, A. Strohbach and O. Seitz, *ChemBioChem* **2008**, 9, 2185; (c) X. Li and D. R. Liu, *Angew. Chem. Int. Ed.* **2004**, 43, 4848.
- ⁴⁹ A. P. Silverman and E. T. Kool, *Chem. Rev.* **2006**, 106, 3775.
- ⁵⁰ Z. Ma and J.-S. Taylor, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 11159.
- ⁵¹ Z. Pianowski and N. Winssinger, *Chem. Commun.* **2007**, 3820.
- ⁵² Y. Xu, N. B. Karalkar and E. T. Kool, *Nat. Biotechnol.* **2001**, 19, 148.
- ⁵³ Aya Shibata, Hiroshi Abe, Mika Ito, Yuko Kondo, Shigeru Shimizu, Kyoko Aikawa and Yoshihiro Ito, *Chem. Commun.* **2009**, 6586.
- ⁵⁴ T. N. Grossmann and O. Seitz, *J. Am. Chem. Soc.* **2006**, 128, 15596.
- ⁵⁵ R. M. Franzini and E. T. Kool, *Org. Lett.*, **2008**, 10, 2008.

Quadruplex Sensors

DNA-Templated Synthesis of Trimethine Cyanine Dyes: A Versatile Fluorogenic Reaction for Sensing G-Quadruplex Formation**

Kamel Meguellati, Girish Koripelly, and Sylvain Ladame*

It has been known for several decades that guanine-rich nucleic acid sequences have a propensity to fold into highly stable four-stranded structures in vitro in the presence of physiological cations, notably potassium and sodium.^[1] Such structures, termed quadruplexes, have had their biological significance demonstrated for a number of processes. For example, the single-stranded 3'-end of telomeric DNA could adopt a quadruplex conformation under near physiological conditions, which has implications on telomere maintenance mechanisms.^[2] More recently, a number of DNA G-quadruplex sequences have been identified in the promoter region of genes that have been proposed to act as regulatory elements for gene expression at the transcriptional level.^[3] Among the 43% of human genes that contain a putative quadruplex-forming sequence in their promoter, specific oncogenes have received particular attention. These include the *c-myc*,^[4] *bcl-2*,^[5] *K-ras*,^[6] and *c-kit*^[7] genes. Although there is an increasing amount of evidence for the formation of G-quadruplexes at telomere ends in vivo,^[8] the possible existence of promoter quadruplexes in vivo is still subject to debate. Recent studies using small-molecule approaches have demonstrated that quadruplex formation within the nuclease hypersensitive element of the *c-myc* gene or within the promoter of the *c-kit* gene were coupled to a significant inhibition of *c-myc*^[9] and *c-kit*^[10] expression at the transcriptional level in various cell lines. However, whilst the 3'-overhang of telomeric DNA is single-stranded, and therefore is free to adopt any stable secondary structure, quadruplex formation within a promoter would require at least a local and temporary opening of the DNA double helix, despite the high stability of Watson–Crick G–C base pairs. Recent studies using fluorescence resonance energy transfer (FRET)^[11] or fluorescent probes^[12] have demonstrated that quadruplexes could potentially form, even when in competition with a thermodynamically more stable duplex form. Moreover, it is well established that double-stranded DNA transiently becomes single-stranded during key biological processes, such as DNA replication, transcription or even recombina-

tion, thus allowing the folding of each DNA strand into alternative (that is, non-B-DNA) structures.^[13]

We are interested in designing sensitive fluorescent biosensors that would be highly specific for unique G-quadruplexes in the genome. The general strategy consists in simultaneously targeting the quadruplex structure itself and also its two flanking regions in a sequence specific manner. Briefly, two short peptide nucleic acids (PNAs)^[14] complementary to both quadruplex flanking regions are functionalized with two nonfluorescent components **A** and **B** of a fluorogenic reaction (that is, the reaction between nonfluorescent derivatives **A** with **B** leads to the formation of fluorescent entity **C**). The system can be designed in such a way that, upon hybridization of the PNA probes to their complementary DNA sequences by Watson–Crick base pairing, **A** and **B** will be in close enough proximity to react with each other when the DNA sequence between both PNAs is folded into a quadruplex structure only, whereas they will be kept separated if the DNA remains single-stranded (Figure 1).

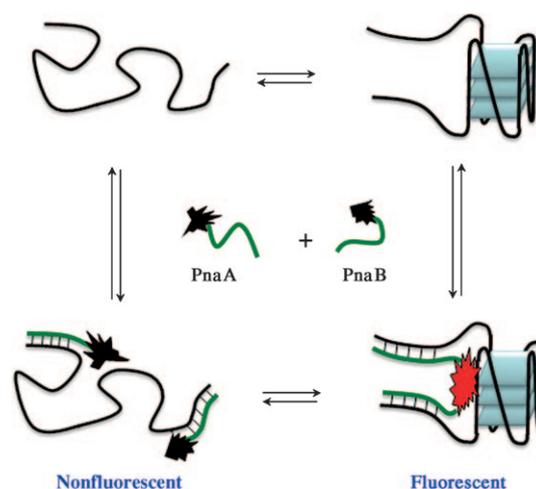


Figure 1. Quadruplex-templated fluorogenic reaction by hybridization of two labeled and nonfluorescent peptide nucleic acids PnaA and PnaB with the single-stranded flanking arms of a G-quadruplex.

Oligonucleotide-templated reactions that can be monitored with high sensitivity by the appearance or disappearance of a fluorescent signal upon binding to the oligonucleotide target have recently received particular attention. Representative examples of such technologies include the use of fluorogenic probes (e.g. molecular beacons), or rely on fluorogenic reactions of chemical ligation or primer exten-

[*] K. Meguellati,^[1] Dr. G. Koripelly,^[1] Dr. S. Ladame
ISIS Université de Strasbourg
8, Allée Gaspard Monge, BP 70028, 67083 Strasbourg (France)
Fax: (+33) 3-6885-5115
E-mail: s.ladame@isis.u-strasbg.fr

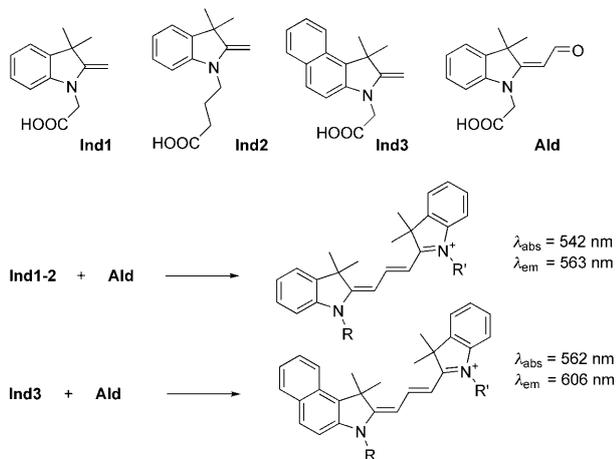
[†] These authors contributed equally to this paper.

[**] S.L. thanks the International Centre for Frontier Research in Chemistry (icFRC) for financial support.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201000291>.

sion.^[15] To date, although such systems offer the advantage of a very high signal-to-noise (S/N) ratio, there has been only few reports of DNA-templated fluorogenic reactions applied to oligonucleotide sensing. Most recent reports are based on the Staudinger reaction,^[16] aldol-type,^[17] organomercury-activated,^[18] or S_NAr ^[19] reactions. They were all developed for detecting oligonucleotide sequences with potential applications as single-nucleotide polymorphism (SNP) probes or RNA sensors in cells. Two modified oligonucleotides (or oligonucleotide analogues) are designed so that 1) they can hybridize specifically to a unique nucleic acid template through Watson–Crick base-pairing and 2) their hybridization to the complementary template only brings both reactive groups in close enough proximity to react with each other.

Herein, the fluorogenic synthesis of a symmetrical or unsymmetrical trimethine cyanine dye by an aldolization-elimination reaction between two nonfluorescent precursors was applied for sensing G-quadruplex formation in vitro. Two PNAs were designed that can each hybridize in a sequence-specific manner with five nucleobases upstream and five nucleobases downstream of the parallel-stranded ckit21T quadruplex^[7b,c] chosen as a model system. They were functionalized at their C-terminal or N-terminal end with either an *N*-alkyl-2-methyleneindoline (Ind1–3) or an *N*-alkyl-2-(3,3-dimethylindolin-2-ylidene)acetaldehyde (Ald; Scheme 1). Two ϵ -*N,N*-dimethyl lysine residues per PNA strand were also added to ensure solubility of both PNAs in water at near-physiological pH.^[20]



Scheme 1. Structures of nonfluorescent 2-methyleneindolines (Ind1, Ind2, and Ind3) and of an aldehyde derivative (Ald). Absorption and emission wavelengths of the cyanine dyes formed upon reaction of Ind1–3 with Ald are also given.

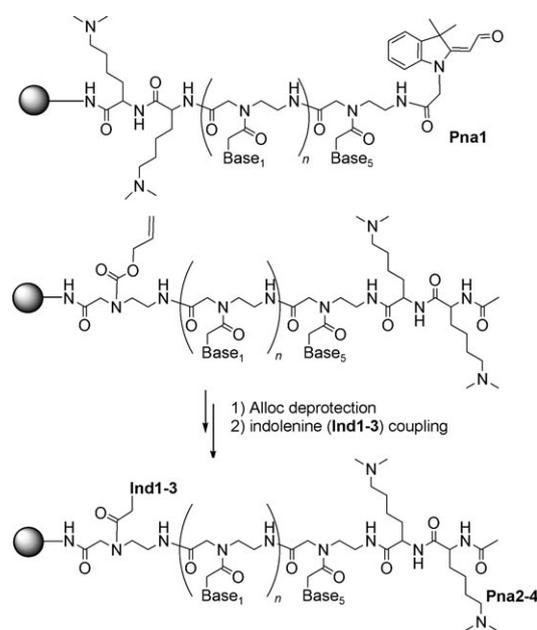
PNAs were synthesized on a rink amide resin (Merck Biosciences, loading 0.67 mmol g^{-1}) using standard solid-phase Fmoc chemistry (Fmoc = 9-fluoromethoxycarbonyl). Ald was introduced at the N-terminus of the PNA by amide coupling on solid support (Pna1; Table 1). For the introduction of an indoline moiety (Ind1, Ind2, or Ind3) at the C-terminus of the PNA, a versatile synthetic strategy was chosen that involves the use of an *N*-Fmoc-[2-(*N*-Alloc)aminoethyl]-glycine PNA monomer (Alloc = allyloxycarbonyl; Scheme 2).

Table 1: Oligonucleotide and PNA sequences.

	DNA ^[a] or PNA ^[b] sequences
Quad1	<u>GCATCCGGGCGGGCGCGAGGGAGGGTTCGGC</u> ^[a]
Quad2	<u>GCATCCGAGCGAGCGCGAGAGAGTTCGGC</u> ^[a]
Quad3	<u>TTCGTCCGGGCGGGCGCGAGGGAGGGTTAAGT</u> ^[a]
Pna1	Lys(NMe ₂)-Lys(NMe ₂)-CGTAG-Ald ^[b]
Pna2	Ind1-AGCCG-Lys(NMe ₂)-Lys(NMe ₂) ^[b]
Pna3	Ind2-AGCCG-Lys(NMe ₂)-Lys(NMe ₂) ^[b]
Pna4	Ind3-AGCCG-Lys(NMe ₂)-Lys(NMe ₂) ^[b]

[a] DNA sequences are given from the 5' to 3' end. [b] PNA sequences are given from the C-terminal to N-terminal end.

A PNA strand was synthesized on a solid support starting with the *N*-Fmoc-[2-(*N*-Alloc)aminoethyl]glycine monomer. After incorporation of the final residue, the resin was treated with



Scheme 2. Solid-phase synthesis of fluorogenic PNAs Pna1–4 on a rink amide resin (gray sphere). Base₁, base₅ = A, T, C, or G.

[Pd(PPh₃)₄] in the presence of dimethylamine–borane under strictly anhydrous and anaerobic conditions to remove selectively the Alloc protecting group.^[21] The indoline moiety (Ind1, Ind2, or Ind3) was finally coupled by amide bond formation using 2-(7-aza-1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) as coupling agent (Scheme 2). This original approach allows convenient functionalization of any immobilized PNA sequence at their C-terminal end just prior to cleavage from the solid support. Herein, this approach was used to introduce various indoline or benz[e]indoline moieties at the C-terminal end of a unique PNA sequence (Pna2–4, Table 1).

Functionalized PNAs were finally cleaved from their solid support by treatment with a solution of trifluoroacetic acid/triisopropylsilane/H₂O (95:2.5:2.5) and the desired PNA was isolated by HPLC and characterized by MALDI.

The interaction between Pna1 and Pna2–4 was next investigated by fluorescence spectroscopy in the presence and in the absence of various types of DNA. Three different DNA sequences (Quad1–3; Table 1) were tested for their capacity to template the fluorogenic reaction of cyanine dye formation. Quad1 corresponds to the previously reported ckit21T^[7b-c] quadruplex-forming sequence with two additional single-stranded flanking arms, located downstream and upstream of the quadruplex, and complementary to Pna1 and Pna2–4, respectively. Quad2 differs from Quad1 by only four G→A mutations to prevent G-quadruplex formation. Quad3 contains the same ckit21T sequence as Quad1 but the quadruplex-forming motif is now flanked with randomized single-stranded arms that are not complementary to the fluorogenic PNAs Pna1–4. Each DNA (200 μM) was folded in a potassium phosphate buffer (10 mM, pH 7.4) that also contained 100 mM KCl.^[22] Under such conditions, Quad1 and Quad3 formed a parallel-stranded quadruplex (see the Supporting Information, Figure S1) whereas Quad2 remained single-stranded. Briefly, a stoichiometric mixture of aldehyde (Pna1) and indoline (Pna2, Pna3, or Pna4) in potassium-containing buffer was incubated at room temperature in the presence or in the absence of an equimolar amount of folded DNA. The reaction of cyanine dye formation was then monitored by fluorescence spectroscopy at different time points. First, the reaction between the PNA aldehyde (Pna1) and the PNA indoline (Pna2) was investigated. Interestingly, only very moderate fluorescence was detectable when working at a PNA strand concentration up to 500 nM. However, when adding a stoichiometric amount (500 nM) of folded quadruplex Quad1 to the previous mixture, a strong fluorescence signal instantaneously appeared which increased up to after 2 h (Figure 2), at which time equilibrium was finally reached (Supporting Information, Figure S2). At equilibrium, a 45-fold increase in fluorescence intensity was observed compared to the quadruplex-free experiment.

A similar trend, although of weaker intensity, was also observed when decreasing the PNA and DNA concentrations down to 200 nM (Supporting Information, Figure S3). To demonstrate that the efficiency of the fluorogenic reaction was indeed linked to quadruplex formation, the same stoichiometric mixture of Pna1 and Pna2 (500 nM each) was reacted in potassium phosphate buffer and in the presence of either Quad2 or Quad3. Key mutations of the ckit21T sequence to prevent quadruplex formation resulted in a complete inhibition of the fluorogenic reaction. Randomization of the quadruplex flanking sequences to prevent PNA:DNA hybridization also led to a significant inhibition compared to the reaction templated by Quad1. These results are consistent with the proposed model suggesting that hybridization of the both aldehyde and indole PNAs to the quadruplex flanking regions associated with folding of the central DNA sequence into a quadruplex conformation are the only conditions that bring both reactive groups in close enough proximity to form the fluorescent cyanine dye. If only one of those requirements is satisfied, no or little reaction will take place.

The influence of the linker between the PNA and the indoline on the efficiency of the fluorogenic reaction was then

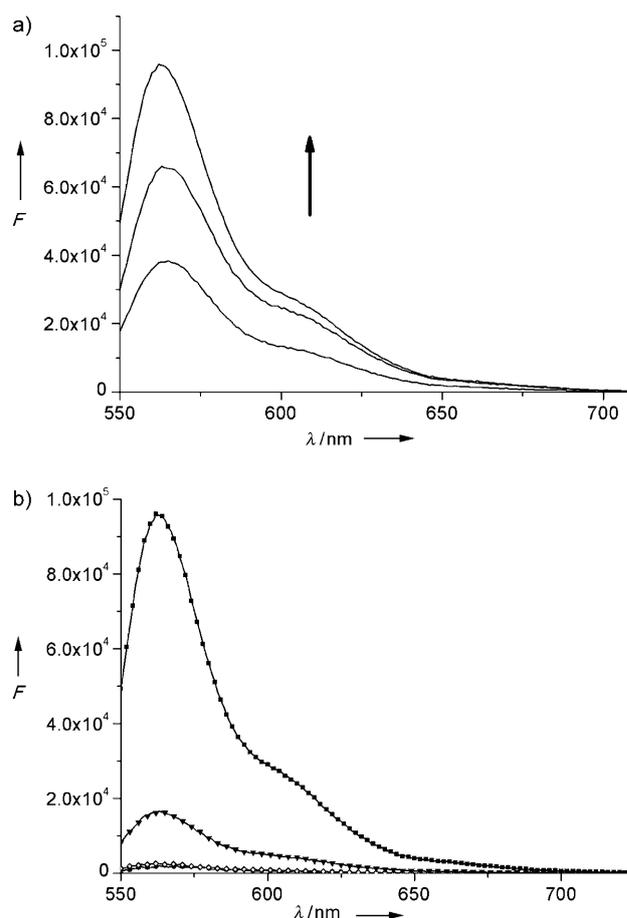


Figure 2. a) Fluorescence emission spectra ($\lambda_{exc} = 540$ nm) of a mixture of Pna1, Pna2, and Quad1 (500 nM each) in potassium phosphate buffer (10 mM, pH 7.4) and 100 mM KCl after 10 min, 1 h, and 2 h (bottom to top) at RT. b) Fluorescence emission spectra ($\lambda_{exc} = 540$ nm) of a mixture of Pna1 and Pna2 (500 nM each) in potassium phosphate buffer and in the absence (●) or in the presence of 500 nM of Quad1 (■), Quad2 (◇), or Quad3 (▼). Fluorescence spectra were recorded after 2 h.

investigated. Pna3 differs from Pna2 by two extra methylene groups between the heterocycle and the PNA scaffold. Although a specific quadruplex-templating effect was observed when mixing Pna1 and Pna3 in the presence of Quad1 which was similar to that obtained with Pna2, it was significantly weaker, thus suggesting the influence of the linker (for example flexibility) on the reaction efficiency (Supporting Information, Figure S4).

An interesting intrinsic property of cyanine dyes is the possibility to tune their spectroscopic properties by varying either the nature of the nitrogen-containing heterocycles or the length of the polymethine chain between them. To shift our quadruplex-specific fluorescent biosensor toward longer wavelengths, Pna4 was synthesized, which differs from Pna2 by the substitution of the indoline moiety by a benz[e]indoline (Table 1). Reaction of Pna4 with Pna1 was then expected to generate an unsymmetrical cyanine dye absorbing and emitting at significantly longer wavelengths than the symmetrical dye formed upon reaction between Pna1 and Pna2 (Scheme 1).^[23] Although no reaction was observed when

reacting Pna1 and Pna4 at 20 μM , a strong quadruplex-templating effect (circa 40-fold increase in fluorescence) was observed at this concentration, resulting in the time-dependent appearance of a characteristic fluorescence signal ($\lambda_{\text{em}} = 606 \text{ nm}$; see Figure 3). This study demonstrates the possibility

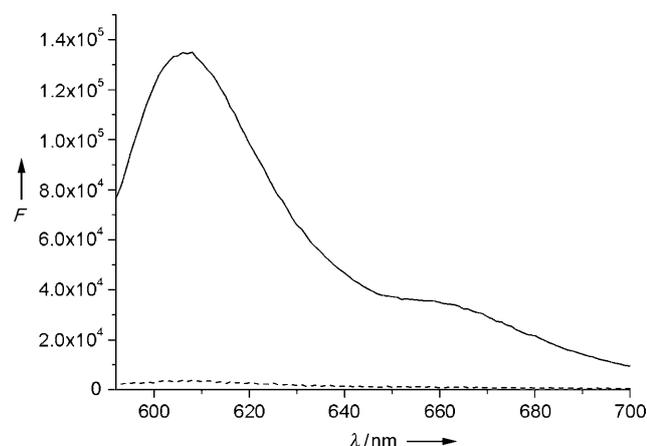


Figure 3. Fluorescence emission spectra ($\lambda_{\text{exc}} = 562 \text{ nm}$) of a mixture of Pna1 and Pna4 (20 μM each) in the absence (----) or presence (—) of Quad1 (20 μM) in potassium phosphate buffer (10 mM, pH 7.4) and 100 mM KCl.

of individual G-quadruplexes to template the formation of various trimethine cyanine dyes that absorb and emit at different wavelengths. However, it is noteworthy that structural modifications of the fluorogenic probes are also accompanied with changes in sensitivity, as a 15-fold loss of sensitivity was observed upon replacing Ind1 by Ind3. This is most likely due to a significantly lower reactivity of Ind3 when compared with Ind1. The specificity of this quadruplex fluorescent biosensor was finally assessed by reacting Pna1 with Pna2 in the presence of various amounts of double-stranded Calf Thymus DNA. Interestingly, no fluorescence was observed when working at low PNA concentration (500 nM each) and high CT concentration (10 $\mu\text{g mL}^{-1}$).

In conclusion, we reported the first example of fluorogenic synthesis of a trimethine cyanine dye that can be templated by a parallel-stranded G-quadruplex DNA in a “sequence + structure”-specific manner. By attaching two nonfluorescent aldehyde and indoline building blocks at the end of two PNA strands complementary to both single-stranded flanking regions of a DNA quadruplex, the fluorogenic reaction occurs only when a quadruplex is formed. Although a DNA-programmed synthesis of hemicyanine dyes that proceeds by a similar aldol-type reaction had already been reported by Huang and Coull,^[17] our system offers the advantage of being more biocompatible as it involves working at physiological pH and requiring no amine additive. This fluorescent biosensor enables the specific detection of a unique quadruplex *in vitro* that is located between both PNAs complementary sequences. Considering the versatility of the PNA functionalization and the broad spectral range covered by cyanine dyes, tunable quadruplex fluorosensors based on this principle can potentially be designed that emit at

different but specific wavelengths. Attempts for simultaneously sensing various quadruplex sequences and/or folds with different colors are currently underway in our laboratory.

Received: January 18, 2010

Published online: March 12, 2010

Keywords: biosensors · cyanines · fluorescent probes · G-quadruplexes · peptide nucleic acids

- [1] a) S. Neidle, S. Balasubramanian in *Quadruplex Nucleic Acids*, Royal Society of Chemistry, Cambridge, **2006**; b) S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- [2] a) E. H. Blackburn, *Nature* **1991**, *350*, 569–573; b) J. F. Riou, L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Mégnin-Chanet, C. Hélène, J. L. Mergny, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2672–2677; c) S. Neidle, G. N. Parkinson, *Nat. Rev. Drug Discovery* **2002**, *1*, 383–393.
- [3] For recent reviews or examples, see: a) Y. Qin, L. H. Hurley, *Biochimie* **2008**, *90*, 1149–1171; b) J. L. Huppert, *Biochimie* **2008**, *90*, 1140–1148; c) J. Eddy, N. Maizels, *Nucleic Acids Res.* **2008**, *36*, 1321–1333.
- [4] V. González, L. H. Hurley, *Annu. Rev. Pharmacol. Toxicol.* **2009**, *50*, 111–129.
- [5] J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones, D. Yang, *J. Am. Chem. Soc.* **2006**, *128*, 1096–1098.
- [6] S. Cogoi, L. E. Xodo, *Nucleic Acids Res.* **2006**, *34*, 2536–2549.
- [7] a) S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589; b) H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle, S. Balasubramanian, *Biochemistry* **2006**, *45*, 7854–7860; c) S. T. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle, S. Balasubramanian, *J. Am. Chem. Soc.* **2009**, *131*, 13399–13409.
- [8] a) K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes, H. J. Lipps, *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–854; b) J. Tang, Z. Y. Kan, Y. Yao, Q. Wang, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2008**, *36*, 1200–1208.
- [9] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598.
- [10] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid, S. Balasubramanian, *J. Am. Chem. Soc.* **2007**, *129*, 12926–12927.
- [11] a) P. S. Shirude, B. Okumus, L. Ying, T. Ha, S. Balasubramanian, *J. Am. Chem. Soc.* **2007**, *129*, 7484–7485; b) N. Kumar, B. Sahoo, K. A. Varun, S. Maiti, S. Maiti, *Nucleic Acids Res.* **2008**, *36*, 4433–4442.
- [12] J. Alzeer, B. R. Vummidi, P. J. Roth, N. W. Luedtke, *Angew. Chem.* **2009**, *121*, 9526–9529; *Angew. Chem. Int. Ed.* **2009**, *48*, 9362–9365.
- [13] For a recent review, see: H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
- [14] a) P. E. Nielsen, M. Egholm, R. H. Berg, B. Buchardt, *Science* **1991**, *254*, 1497–1500; b) 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*, 566–568.
- [15] For recent examples, see: a) K. Wang, Z. Tang, C. J. Yang, Y. Kim, X. Fang, W. Li, Y. Wu, C. D. Medley, Z. Cao, J. Li, P. Colon, H. Lin, W. Tan, *Angew. Chem.* **2009**, *121*, 870–885; *Angew. Chem. Int. Ed.* **2009**, *48*, 856–870; b) T. N. Grossmann, O. Seitz, *Chem. Eur. J.* **2009**, *15*, 6723–6730.
- [16] a) J. Cai, X. Li, X. Yue, J. S. Taylor, *J. Am. Chem. Soc.* **2004**, *126*, 16324–16325; b) Z. L. Pianowski, N. Winssinger, *Chem. Commun.* **2007**, 3820–3822; c) R. M. Franzini, E. T. Kool,

- ChemBioChem* **2008**, *9*, 2981–2988; d) Z. Pianowski, K. Gorska, L. Oswald, C. A. Merten, N. Winssinger, *J. Am. Chem. Soc.* **2009**, *131*, 6492–6497; e) K. Furukawa, H. Abe, K. Hibino, Y. Sako, S. Tsuneda, Y. Ito, *Bioconjugate Chem.* **2009**, *20*, 1026–1036.
- [17] Y. Huang, J. M. Coull, *J. Am. Chem. Soc.* **2008**, *130*, 3238–3239.
- [18] R. M. Franzini, E. T. Kool, *Org. Lett.* **2008**, *10*, 2935–2938.
- [19] A. Shibata, H. Abe, M. Ito, Y. Kondo, S. Shimizu, K. Aikawa, Y. Ito, *Chem. Commun.* **2009**, 6586–6588.
- [20] ϵ -*N,N*-dimethyl lysines were preferred to lysines to avoid the presence of nucleophilic primary amine within the PNA strand that could potentially react with Ald.
- [21] D. V. Jarikote, O. Köhler, E. Socher, O. Seitz, *Eur. J. Org. Chem.* **2005**, 3187–3195.
- [22] A DNA solution containing either Quad1, 2, or 3 (200 μ M) in potassium phosphate buffer (10 mM, pH 7.4) with KCl solution (100 mM) was heated at 95 °C for 5 minutes and slowly cooled to room temperature over 8 hours.
- [23] S. J. Mason, J. L. Hake, J. Nairne, W. J. Cummins, S. Balasubramanian, *J. Org. Chem.* **2005**, *70*, 2939–2949.
-

Supporting Information

© Wiley-VCH 2010

69451 Weinheim, Germany

DNA-Templated Synthesis of Trimethine Cyanine Dyes: A Versatile Fluorogenic Reaction for Sensing G-Quadruplex Formation**

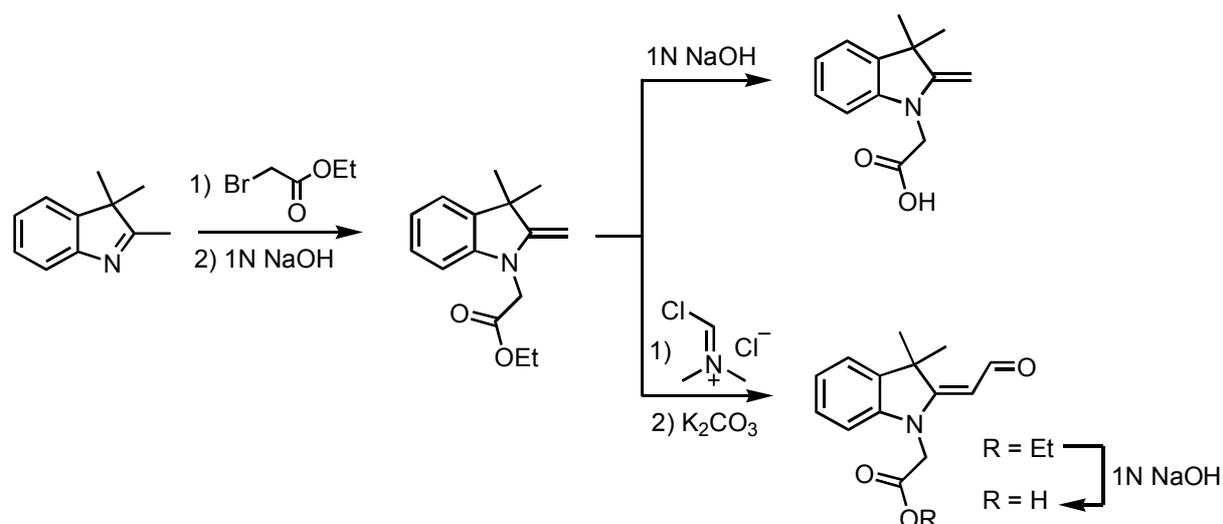
*Kamel Meguellati, Girish Koripelly, and Sylvain Ladame**

anie_201000291_sm_miscellaneous_information.pdf

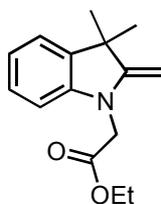
SUPPORTING INFORMATION

1	SYNTHESIS.....	2
2	SPECTROSCOPIC STUDIES.....	13
2.1	Circular Dichroism studies.....	14
2.2	Fluorescence studies.....	14

1 SYNTHESIS

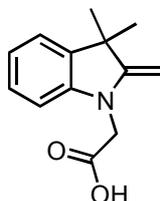


Ethyl 2-(3,3-dimethyl-2-methyleneindolin-1-yl)-acetate



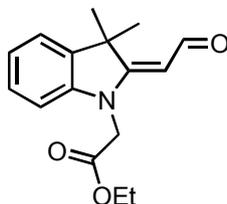
2,3,3-trimethyl-3H-indole (1.00 g, 6.28 mmol) and ethyl iodoacetate (1.61 g, 7.53 mmol) were heated together at 80°C for 5 h. After cooling to room temperature the red solid was triturated with ether, filtered and dried under vacuum. To the solid, 1 N NaOH (25 mL) was added and stirred for 1 h at RT. The reaction mixture was then extracted with CH₂Cl₂ (3 x 30 mL) and the combined organic layers were dried over Na₂SO₄, filtered and concentrated in *vacuo*. The residue was purified by column chromatography on SiO₂ (PE/Ether, 3:1; R_f = 0.48) to give the title compound (1.28 g, 83%) as a yellow oil. ¹H NMR (400 MHz, DMSO): δ = 7.16 (dd, *J* = 7.3, 0.9 Hz, 1H), 7.06 (td, *J* = 7.7, 1.2 Hz, 1H), 6.74 (td, *J* = 7.5, 0.6 Hz, 1H), 6.67 (d, *J* = 7.9 Hz, 1H), 4.39 (s, 2H), 4.10 (q, *J* = 7.2 Hz, 2H), 3.87 (d, *J* = 2.2 Hz, 1H), 3.81 (d, *J* = 2.0 Hz, 1H), 1.29 (s, 6H), 1.17 (t, *J* = 7.2 Hz, 3H) ppm. ¹³C NMR (100 MHz, DMSO): δ = 168.6, 161.0, 145.4, 136.7, 127.3, 121.8, 118.8, 105.7, 74.8, 60.4, 43.6, 43.5, 29.6 (x 2) 14.1 ppm. HRMS (ESI): *m/z*: calcd for C₁₅H₁₉NO₂ [M + H]⁺ 246.149; found 246.150.

2-(3,3-dimethyl-2-methyleneindolin-1-yl)acetic acid



Ethyl 2-(3,3-dimethyl-2-methyleneindolin-1-yl)acetate (1.30 g, 5.30 mmol) was dissolved in ethanol (40 mL) and 1N NaOH solution (30 mL). The mixture was stirred at RT for overnight and the solvent was removed in *vacuo*. The residue was dissolved in minimum amount of water (15 mL) and then acidified to pH=3 with 1N HCl solution. The solution was concentrated and the residue was redissolved in ethanol. The suspension was filtered and this process was repeated at least for two times. Finally the filtrate was concentrated and the residue was purified by column chromatography on SiO₂ (CH₂Cl₂/MeOH 8:2; R_f = 0.20) to give the title compound (860 mg, 75 %) as a thick brown oil. ¹H NMR (400 MHz, DMSO): δ = 7.15 (dd, *J* = 7.3, 0.7 Hz, 1H), 7.06 (td, *J* = 7.7, 1.2 Hz, 1H), 6.73 (t, *J* = 7.2 Hz, 1H), 6.65 (d, *J* = 7.9 Hz, 1H), 4.26 (s, 2H), 3.85 (d, *J* = 2.0 Hz, 1H), 3.81 (d, *J* = 2.0 Hz, 1H), 1.28 (s, 6H) ppm, CO₂H signal not visible. ¹³C NMR (100 MHz, DMSO): δ = 170.1, 160.8, 145.5, 136.7, 127.3, 121.7, 118.6, 105.6, 74.4, 43.6, 43.5, 29.7 (x 2) ppm. HRMS (ESI): *m/z*: calcd for C₁₃H₁₅NO₂ [M + H]⁺ 218.116; found 218.119.

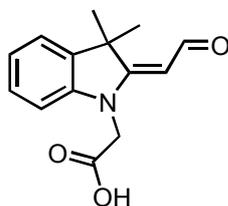
Ethyl 2-((E)-2-(formylmethylene)-3,3-dimethylindolin-1-yl)acetate



(Chloromethylene)dimethylammonium chloride (4.69 g, 36.6 mmol) was dissolved in CH₂Cl₂ (30 mL) and stirred at RT for 10 minutes. Then a solution of Ethyl 2-(3,3-dimethyl-2-methyleneindolin-1-yl)acetate (3.00 g, 12.2 mmol) in CH₂Cl₂ (20

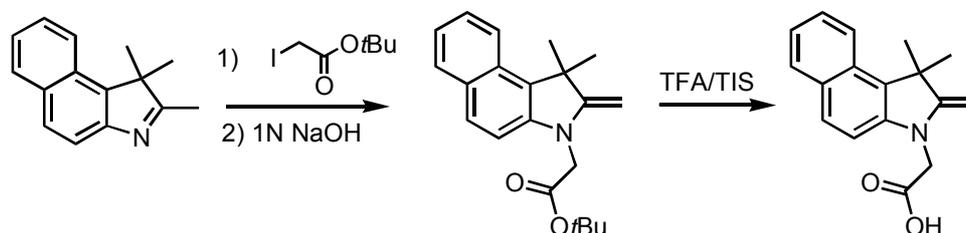
mL) was added dropwise. After stirring the reaction mixture for 1 h, the solvent was removed in *vacuo*. The residue was dissolved in THF (50) and a saturated sol. of K_2CO_3 (40 mL) was added carefully. After stirring the mixture for 5 h at RT, THF was removed under reduced pressure and the aqueous layer was extracted with CH_2Cl_2 (3 x 50 mL). The combined organic layers were dried over Na_2SO_4 , filtered and concentrated in *vacuo*. The residue was purified by column chromatography on SiO_2 (Ether; $R_f = 0.30$) to give the title compound (2.9 g, 87%) as a yellow oil. 1H NMR (400 MHz, DMSO): $\delta = 9.92$ (d, $J = 8.7$ Hz, 1H), 7.42 (d, $J = 7.3$ Hz, 1H), 7.34 (td, $J = 7.8, 1.0$ Hz, 1H), 7.07-7.02 (m, 2H), 5.22 (d, $J = 8.7$ Hz, 1H), 4.74 (s, 2H), 4.14 (q, $J = 7.2$ Hz, 2H), 1.61 (s, 6H), 1.18 (t, $J = 7.2$ Hz, 3H) ppm. ^{13}C NMR (100 MHz, DMSO): $\delta = 186.4, 173.1, 167.9, 143.4, 139.2, 128.4, 122.8, 122.5, 109.0, 99.5, 61.6, 47.5, 44.4, 29.6$ (x 2), 14.5 ppm. HRMS (ESI): m/z : calcd for $C_{16}H_{19}NO_3$ $[M + H]^+$ 274.144; found 274.144.

2-((E)-2-(formylmethylene)-3,3-dimethylindolin-1-yl)acetic acid

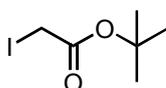


The Ethyl 2-((E)-2-(formylmethylene)-3,3-dimethylindolin-1-yl)acetate (2.8 g, 10.2 mmol) was dissolved in ethanol (60 ml) and NaOH solution (4%, 30 mL). The mixture was stirred at RT for overnight and the solvent was removed in *vacuo*. The residue was dissolved in minimum amount of water (25 mL) and then acidified to pH=3 with a 1N HCl solution. The solution was concentrated and the residue was redissolved in ethanol. The suspension was filtered and this process was repeated at least for two times. Finally the filtrate was concentrated and the residue was purified by column chromatography on SiO_2 ($CH_2Cl_2/MeOH$ 8:2; $R_f = 0.15$) to give the title compound (2.0 g, 80 %) as a thick yellow oil. 1H NMR (400 MHz, DMSO): $\delta = 9.85$ (d, $J = 8.9$ Hz, 1H), 7.38 (d, $J = 7.3$ Hz, 1H), 7.20 (t, $J = 7.6$ Hz, 1H), 6.99 (t, $J = 7.5$ Hz, 1H), 6.92 (d, $J = 7.8$ Hz, 1H), 5.18 (d, $J = 8.9$ Hz, 1H), 4.21 (s, 2H), 1.59 (s, 6H) ppm, CO_2H signal not visible. ^{13}C NMR (100 MHz,

DMSO): $\delta = 185.4, 172.9, 170.5, 143.7, 139.1, 127.7, 121.8, 121.7, 108.7, 98.5, 47.0, 46.2, 29.0$ (x 2) ppm. HRMS (ESI): m/z : calcd for $C_{14}H_{15}NO_3$ $[M + H]^+$ 246.112; found 246.111.

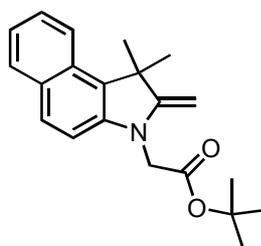


tert-butyl 2-iodoacetate



To a suspension of NaI (3.05 g, 20.4 mmol) in acetone (40 mL) was added *tert*-butyl 2-bromoacetate (2 g, 10.2 mmol) and the resulting mixture was stirred for 4 hours at 65 °C under nitrogen atmosphere. After cooling to room temperature the suspension was filtered and the filtrate was concentrated. The residue was dissolved in H_2O (50 mL) and the mixture was extracted with ether (3 x 50 mL). The combined organic layers were successively washed with water (50 mL), brine (50 mL), 10% aq. $Na_2S_2O_3$ soln. (50 mL) and dried over Na_2SO_4 , filtered and concentrated in vacuo to give the title compound (2.4 g, 98 %) as a light yellow oil. 1H NMR (400 MHz, DMSO): $\delta = 3.68$ (s, 2H), 1.40 (s, 9H) ppm. ^{13}C NMR (100 MHz, DMSO): $\delta = 168.1, 81.2, 27.3$ (x 3), - 1.3 ppm.

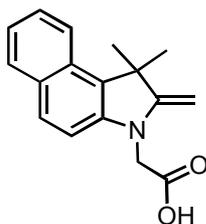
tert-butyl(1,1-dimethyl-2-methylene-1,2-dihydro-3H-benzo[e]indol-3-yl)acetate



1,1,2-trimethylbenzo[e]indole (1.55 g, 7.40 mmol) and *tert*-butyl 2-iodoacetate (2.04 g, 8.43 mmol) were heated together at 80°C for 6 h. After cooling to room temperature

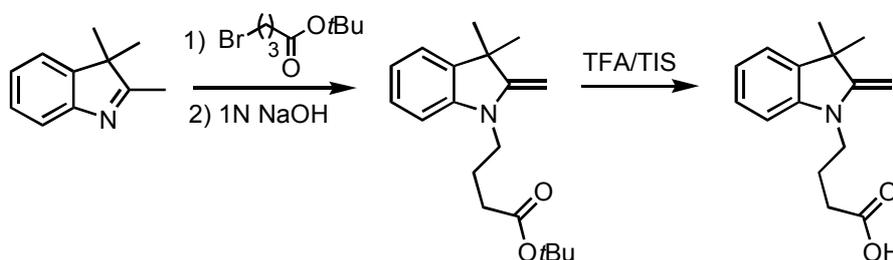
the green solid was triturated with ether, filtered and dried under vacuum. To the solid, CH₂Cl₂ (40 mL) and 1 N NaOH (30 mL) was added and stirred for 1 h at RT. The layers were separated and the aqueous layer was extracted with CH₂Cl₂ (3 x 30 mL). The combined organic layers were dried over Na₂SO₄, filtered and evaporated in *vacuo*. The residue was purified by column chromatography on SiO₂ (CH₂Cl₂/EtOAc 9:1; R_f = 0.40) to give the title compound (1.8 g, 73%) as a light brown oil. ¹H NMR (400 MHz, DMSO): δ = 7.96 (d, *J* = 8.8 Hz, 1H), 7.79 (d, *J* = 8.0 Hz, 1H), 7.73 (d, *J* = 8.8 Hz, 1H), 7.42-7.38 (m, 1H), 7.20-7.16 (m, 2H), 4.39 (s, 2H), 4.00 (d, *J* = 2.4 Hz, 1H), 3.88 (d, *J* = 2.4 Hz, 1H), 1.59 (s, 6H), 1.36 (s, 9H) ppm. ¹³C NMR (100 MHz, DMSO): δ = 167.9, 162.4, 143.2, 129.4, 129.3, 128.7, 128.6, 126.4, 125.7, 121.4, 121.2, 109.3, 80.9, 74.8, 45.5, 44.4, 29.3 (x 2), 27.6 (x 3) ppm. HRMS (ESI) : *m/z*: calcd for C₂₁H₂₅NO₂ [M + H]⁺ 324.196; found 324.193.

(1,1-dimethyl-2-methylene-1,2-dihydro-3*H*-benzo[*e*]indol-3-yl)acetic acid

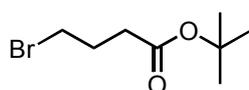


TFA (30 mL) was added dropwise to an ice cooled solution of *tert*-butyl (1,1-dimethyl-2-methylene-1,2-dihydro-3*H*-benzo[*e*]indol-3-yl)acetate (1.6 g, 4.9 mmol) and Et₃SiH (2.3 mL, 14.7 mmol) dissolved in CH₂Cl₂ (15mL). The reaction mixture was stirred for overnight by slowly warming to room temperature. Then the solvent was removed under reduced pressure and the remaining volatiles were co-evaporated with toluene (3 x 30 mL). The residue was purified by column chromatography on SiO₂ (CH₂Cl₂/MeOH 8:2; R_f = 0.30) to give the title compound (1.1 g, 85%) as a purple solid. ¹H NMR (400 MHz, DMSO): δ = 12.72 (br. s, 1H), 7.96 (d, *J* = 8.7 Hz, 1H), 7.79 (d, *J* = 8.0 Hz, 1H), 7.72 (d, *J* = 8.8 Hz, 1H), 7.40 (t, *J* = 7.6 Hz, 1H), 7.21-7.16 (m, 2H), 4.41 (s, 2H), 4.00 (d, *J* = 1.9 Hz, 1H), 3.90 (d, *J* = 1.9 Hz, 1H), 1.59 (s, 6H) ppm. ¹³C NMR (100 MHz, DMSO): δ = 170.2, 162.5, 143.3, 129.4, 129.3, 128.7, 128.6, 126.4, 126.0, 121.4, 121.2, 109.4, 74.8, 45.5,

43.7, 29.3 (x 2) ppm. HRMS (ESI): m/z : calcd for $C_{17}H_{17}NO_2$ $[M + H]^+$ 268.133; found 268.133.

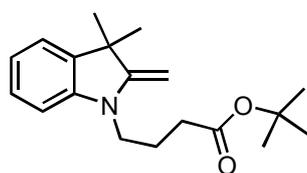


tert-butyl 4-bromobutanoate



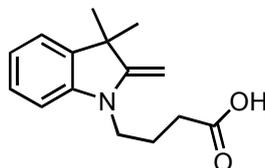
Conc. H_2SO_4 (1.1 mL, 20 mmol) was added to a vigorously stirred suspension of anhydrous magnesium sulfate (9.62 g, 80 mmol) in CH_2Cl_2 (100 mL). After stirring the mixture for 15 min, bromobutyric acid (3.35 g, 20 mmol) was added followed by *tert*-butanol (9.5 mL, 100 mmol) and the mixture was stirred at RT for 3 d under nitrogen atmosphere. The reaction mixture was then quenched with sat. aq. $NaHCO_3$ (200 ml) and stirred until all magnesium sulphate was dissolved. The organic layer was separated, washed with water, dried over Na_2SO_4 and the solvent evaporated in vacuo. The residue was purified by column chromatography on SiO_2 (PE/ CH_2Cl_2 9:1; $R_f = 0.20$) to give the title compound (3.2 g, 72%) as a colourless oil. 1H NMR (400 MHz, DMSO): $\delta = 3.53$ (t, $J = 6.9$ Hz, 2H), 2.34 (t, $J = 7.3$ Hz, 2H), 2.00 (quintet, $J = 7.0$ Hz, 2H), 1.40 (s, 9H) ppm. ^{13}C NMR (100 MHz, DMSO): $\delta = 171.8, 80.3, 34.4, 33.7, 28.3, 28.2$ (x 3) ppm. HRMS (ESI): m/z : calcd for $C_8H_{15}BrO_2$ $[M + Na]^+$ 245.015; found 245.016.

tert-butyl 4-(3,3-dimethyl-2-methyleneindolin-1-yl)butanoate



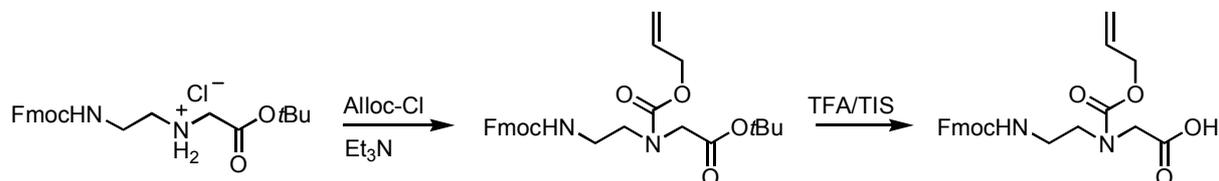
2,3,3-trimethyl-3H-indole (1.00 g, 6.28 mmol) and *tert*-butyl 4-bromobutanoate (1.68 g, 7.53 mmol) were heated together at 80°C for 8 h. After cooling to room temperature, the dark red solid was triturated with ether, filtered and dried under vacuum. To the solid 1 N NaOH (25 mL) was added and stirred for 1 h at RT. The reaction mixture was extracted with CH₂Cl₂ (3 x 30 mL) and the combined organic layers were dried over Na₂SO₄, filtered and concentrated *in vacuo*. The residue was purified by column chromatography on SiO₂ (PE/EtOAc, 20:1; R_f = 0.35) to give the title compound (1.4 g, 74%) as a light red oil. ¹H NMR (400 MHz, DMSO): δ = 7.13 (d, *J* = 7.0 Hz, 1H), 7.07 (td, *J* = 7.8, 1.1 Hz, 1H), 6.70 (t, *J* = 7.4 Hz, 1H), 6.64 (d, *J* = 7.8 Hz, 1H), 3.92 (d, *J* = 1.6 Hz, 1H), 3.84 (d, *J* = 1.6 Hz, 1H), 3.52 (t, *J* = 7.3 Hz, 2H), 2.26 (t, *J* = 7.2 Hz, 2H), 1.79-1.71 (m, 2H), 1.40 (s, 9H), 1.26 (s, 6H) ppm. ¹³C NMR (100 MHz, DMSO): δ = 171.9, 160.5, 145.4, 136.9, 127.5, 121.8, 118.2, 105.2, 79.7, 73.8, 43.6, 40.4, 31.7, 29.8 (x 2), 27.7 (x 3), 21.0 ppm. HRMS (ESI): *m/z*: calcd for C₁₉H₂₇NO₂ [M + H]⁺ 302.211; found 302.210.

4-(3,3-dimethyl-2-methyleneindolin-1-yl)butanoic acid

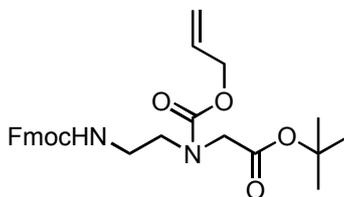


Trifluoroacetic acid (26 mL) was added dropwise to an ice cooled solution of *tert*-butyl (1,1-dimethyl-2-methylene-1,2-dihydro-3*H*-benzo[*e*]indol-3-yl)acetate (1.3 g, 4.3 mmol) and Et₃SiH (2.0 mL, 12.9 mmol) dissolved in CH₂Cl₂ (12mL). The reaction mixture was stirred for overnight by slowly warming to RT. Then the solvent was removed under reduced pressure and the remaining volatiles were co-evaporated with toluene (3 x 30 mL). The residue was purified by column chromatography on SiO₂ (CH₂Cl₂/MeOH 8:2; R_f = 0.35) to give the title compound (850 mg, 81%) as a red solid. ¹H NMR (400 MHz, DMSO): δ = 7.13 (d, *J* = 7.3 Hz, 1H), 7.07 (td, *J* = 7.5, 1.1 Hz, 1H), 6.70 (t, *J* = 7.3 Hz, 1H), 6.63 (d, *J* = 7.8 Hz, 1H), 3.92 (d, *J* = 1.9 Hz, 1H), 3.84 (d, *J* = 1.6 Hz, 1H), 3.51 (t, *J* = 7.0 Hz, 2H),

2.26 (t, $J = 7.3$ Hz, 2H), 1.76 (quintet, $J = 7.3$ Hz, 2H), 1.39 (s, 6H) ppm, CO₂H signal not visible. ¹³C NMR (100 MHz, DMSO): $\delta = 171.9, 160.5, 145.3, 136.8, 127.4, 121.7, 118.1, 105.1, 73.8, 43.5, 40.0, 31.7, 29.7$ (x 2), 21.0 ppm. HRMS (ESI): m/z : calcd for C₁₅H₁₉NO₂ [M + H]⁺ 246.149; found 246.147.



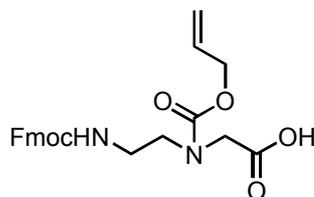
***tert*-butyl-*N*-[(allyloxy)carbonyl]-*N*-(2-[[*(9H*-fluoren-9-ylmethoxy)carbonyl]amino]ethyl)glycinate**



tert-butyl-*N*-(2-[[*(9H*-fluoren-9-ylmethoxy)carbonyl]amino]ethyl)glycinate hydrochloride (400mg, 0.92 mmol) was dissolved in CH₂Cl₂ (100 mL) and washed with a saturated aqueous solution of NaHCO₃ (3 x 50 mL). The organic layer was dried over Na₂SO₄ and filtered. To this organic layer was added Et₃N (128 μ L, 0.92 mmol) under nitrogen atmosphere and the solution was cooled to 0 °C in an ice bath. Allyl chloroformate (98 μ L, 0.92 mmol) was added dropwise and the resulting mixture was stirred for 2 h at the same temperature before it was diluted with H₂O (40 ml). The layers were separated and the aqueous layer was extracted with CH₂Cl₂ (3 x 50 mL). The combined organic layers were washed with brine (50 mL) dried over Na₂SO₄, filtered and concentrated *in vacuo*. The residue was purified by column chromatography on SiO₂ (PE/EtOAc, 7:3; $R_f = 0.25$) to give the title compound (420 mg, 95%) as a colourless oil. ¹H NMR (400 MHz, DMSO): $\delta = 7.88$ (d, $J = 7.5$ Hz, 2H), 7.67 (d, $J = 7.5$ Hz, 2H), 7.41 (t, $J = 7.4$ Hz, 2H), 7.33 (t, $J = 7.4$ Hz, 2H), 7.29-7.25 (m, 1H), 6.86-6.77 (m, minor rotamer, 0.1H), 5.93-5.72 (m, major rotamer, 0.9H), 5.30-5.21 (m, 1H), 5.17-5.12 (m, 1H), 4.51-4.47 (m, 2H), 4.39-4.34 (m, minor rotamer, 0.2H), 4.30-4.29 (m, major rotamer, 1.8H), 4.20 (t, $J = 6.3$ Hz, 1H), 3.88-3.66 (m, major rotamer, 1.8H), 3.75-

3.69 (m, minor rotamer, 0.2H), 3.32-3.29 (m, 2H), 3.17-3.12 (m, 2H), 1.40 (s, 9H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 168.8 (major) and 168.7 (minor rotamers), 157.2 (minor) and 156.1 (major rotamers), 155.4 (minor) and 155.2 (major rotamers), 143.9, 142.6, 140.7, 139.4, 137.4, 133.1 (minor) and 133.9 (major rotamers), 128.9, 127.6, 127.3 (minor) and 127 (major rotamers), 125.1 (major) and 125.0 (minor rotamers), 121.4, 120.0 (major) and 119.9 (minor rotamers), 116.7 (minor) and 116.4 (major rotamers), 109.7, 81.0 (major) and 80.9 (minor rotamers), 65.3 (major) and 65.2 (minor rotamers), 49.8 (major) and 49.7 (minor rotamers), 48.0 (minor) and 47.8 (major rotamers), 47.3 (minor) and 47.1 (major rotamers), 46.7, 38.5, 27.6 (x 3) ppm. HRMS (ESI): m/z : calcd for $\text{C}_{27}\text{H}_{32}\text{N}_2\text{O}_6$ [M + Na] $^+$ 503.215; found 503.214.

***tert*-butyl-*N*-[(allyloxy)carbonyl]-*N*-(2-[[*(9H*-fluoren-9-ylmethoxy)carbonyl]amino]ethyl)glycinate**



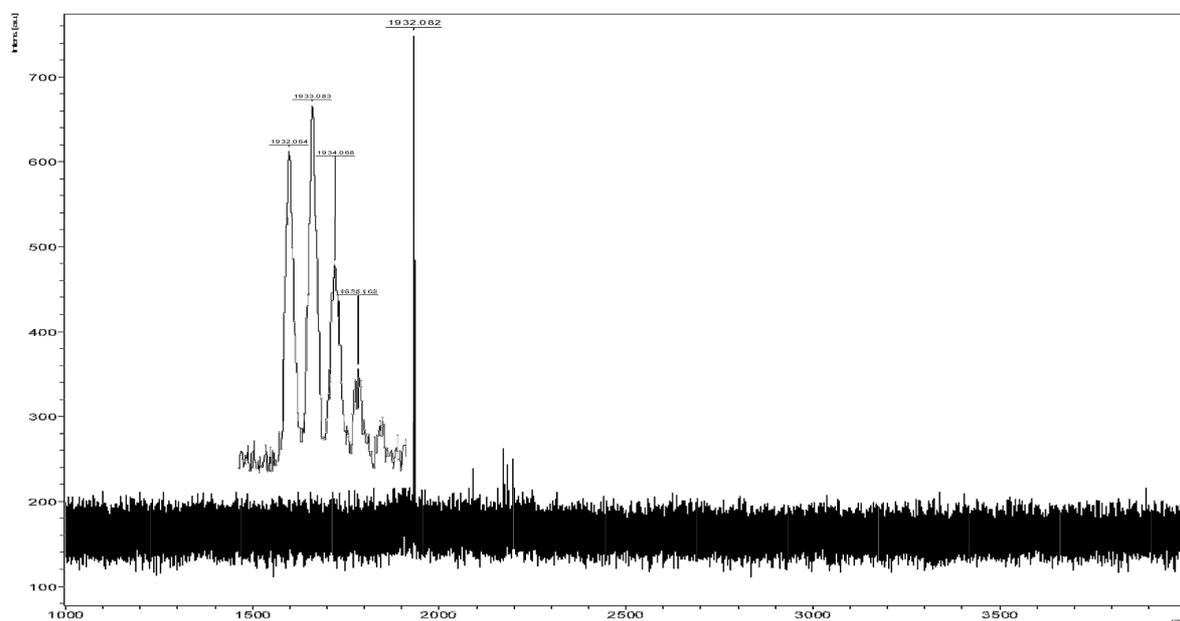
TFA (5 ml) was added dropwise to an ice cooled solution of *tert*-butyl *N*-[(allyloxy)carbonyl]-*N*-(2-[[*(9H*-fluoren-9-yl-methoxy)carbonyl]amino]ethyl)glycinate (400 mg, 0.83 mmol) and Et_3SiH (0.4 ml, 2.49 mmol) dissolved in CH_2Cl_2 (5 mL). The reaction mixture was stirred overnight by slowly warming to room temperature. Then the solvent was removed under reduced pressure and the remaining volatiles were co-evaporated with toluene (3 x 25 mL) to give the title compound (350 mg, 99%) as a colourless solid. The crude product was directly used for PNA coupling without further purification. ^1H NMR (400 MHz, DMSO): δ = 12.74 (br. s, 1H), 7.88 (d, J = 7.5 Hz, 2H), 7.67 (d, J = 7.3 Hz, 2H), 7.41 (t, J = 7.5 Hz, 2H), 7.33 (t, J = 7.5 Hz, 2H), 7.29-7.27 (m, 1H), 6.86-6.76 (m, minor rotamer, 0.1H), 5.93-5.80 (m, major rotamer, 0.9H), 5.30-5.21 (m, 1H), 5.16-5.11 (m, 1H), 4.50-4.48 (m, 2H), 4.37-4.33 (m, minor rotamer, 0.2H), 4.29-4.28 (m, major rotamer, 1.8H), 4.20 (t, J =7.0 Hz, 1H), 3.93-3.90 (m, major

rotamer, 1.8H), 3.82-3.74 (m, minor rotamer, 0.2H), 3.32-3.29 (m, 2H), 3.17-3.12 (m, 2H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 171.5, 171.3, 156.4, 155.8, 155.6, 144.0, 140.9, 133.3, 127.9, 127.3, 125.3, 125.2, 120.3, 116.9, 116.6, 65.6, 65.4, 49.2, 49.0, 47.9, 47.2, 46.9, 38.6 ppm HRMS (ESI): m/z : calcd for $\text{C}_{23}\text{H}_{24}\text{N}_2\text{O}_6$ $[\text{M} + \text{Na}]^+$ 447.153; found 447.152.

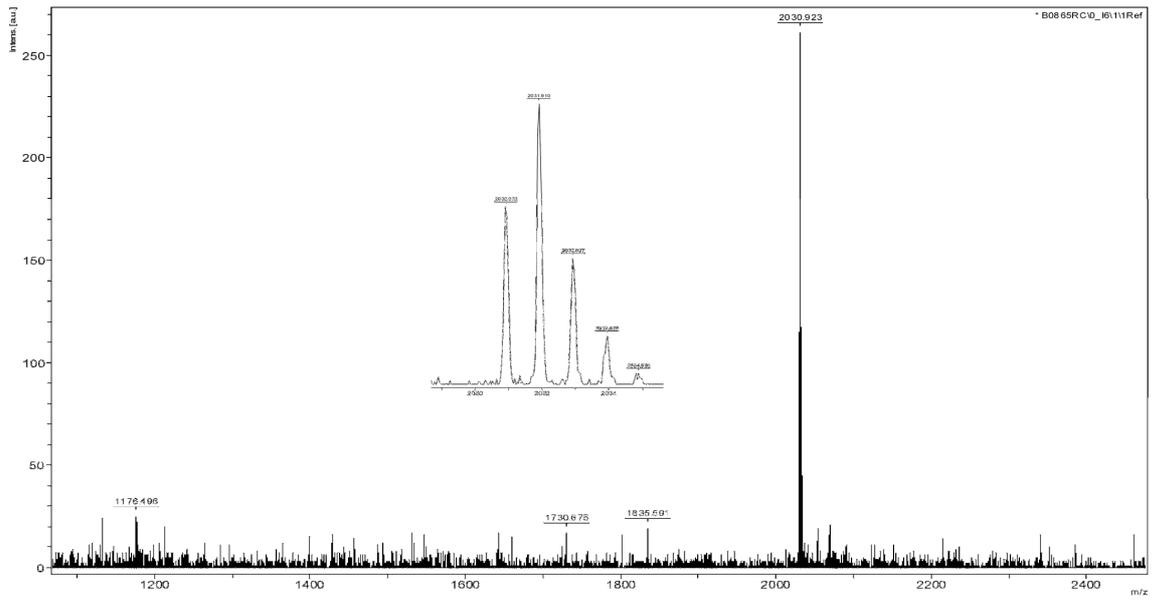
PNA synthesis and characterisation.

Pna1-4 were synthesized on a rink amide resin (Merck Biosciences, loading = 0.67 mmol/g) using standard solid-phase Fmoc chemistry and HATU as coupling agent. They were all purified by HPLC and characterized by mass spectrometry (Maldi-TOF). **Pna1** m/z : $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{84}\text{H}_{115}\text{N}_{36}\text{O}_{19}^+$ 1931.913, found 1932.082; **Pna2** m/z : $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{88}\text{H}_{124}\text{N}_{39}\text{O}_{19}^+$ 2030.993, found 2030.923; **Pna3** m/z : $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{90}\text{H}_{128}\text{N}_{39}\text{O}_{19}^+$ 2059.024, found 2059.067; **Pna4** m/z : $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{92}\text{H}_{126}\text{N}_{39}\text{O}_{19}^+$ 2081.009, found 2080.904.

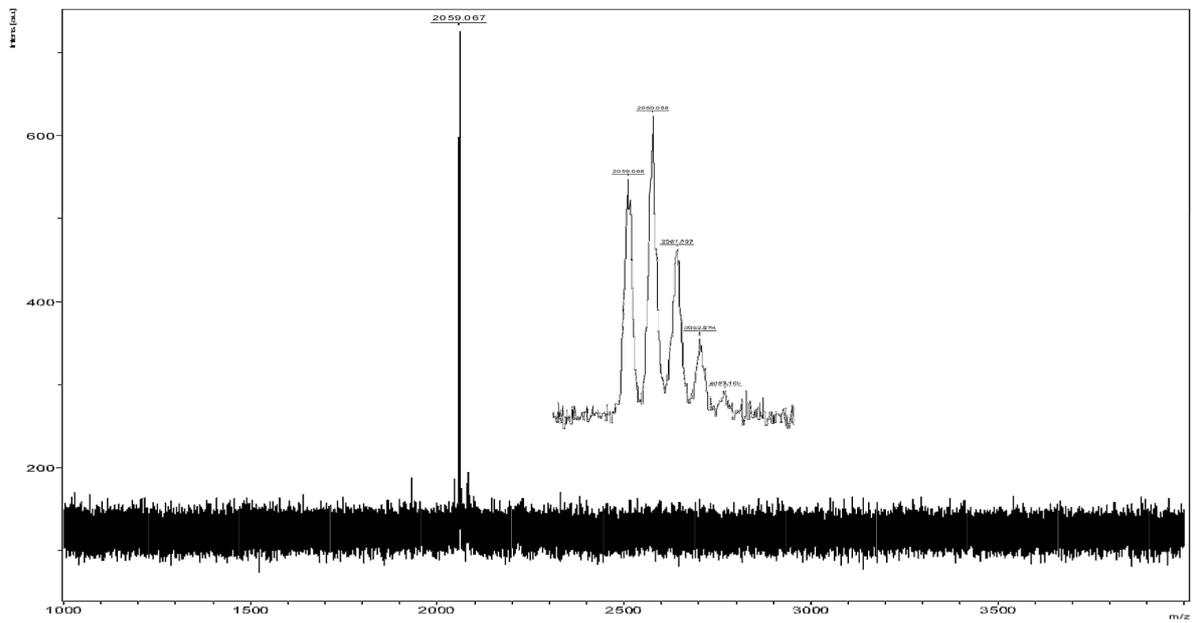
Pna1



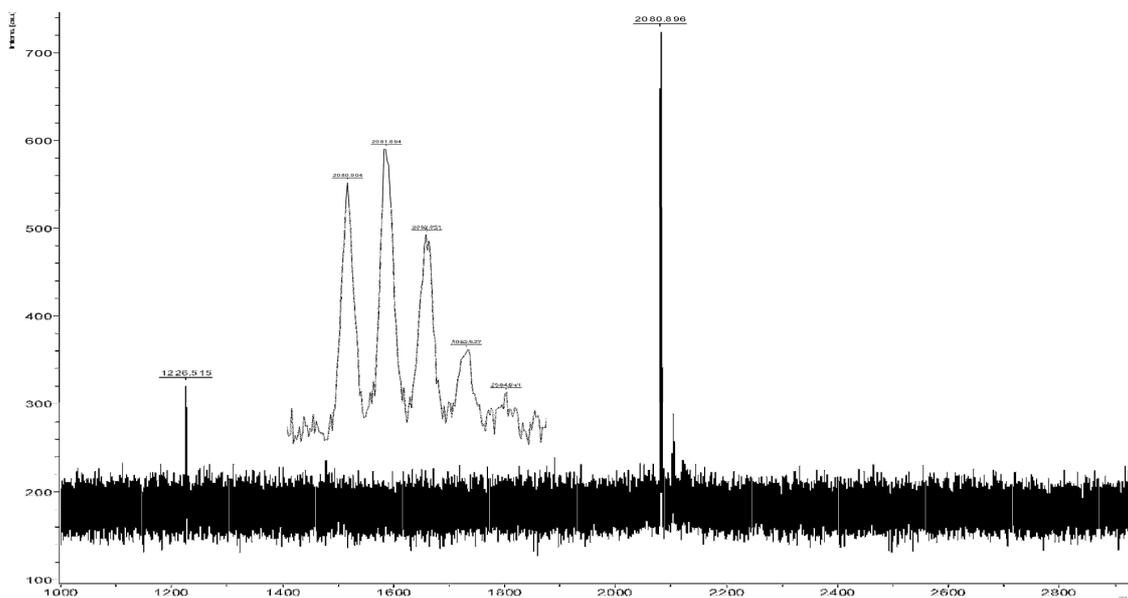
Pna2



Pna3



Pna4



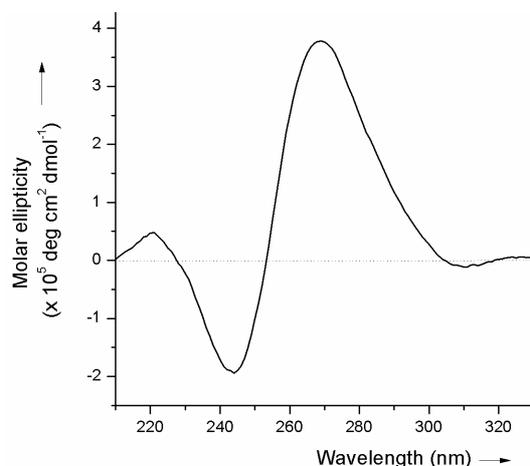
2 SPECTROSCOPIC STUDIES

Fluorescence emission spectra were recorded in quartz cells (with a 5 mm pathlength) at 20 °C on a Jobin Yvon Fluorolog 3.22 instrument. The excitation and emission bandwidths were fixed to 5 nm and 5 nm, respectively. Circular dichroism spectra were recorded on a Jobin-Yvon spectrophotometer in a 2 mm pathlength cuvette (300 μ L). Spectra were recorded at 25°C from 320 to 220 nm and are presented at an average of three successive scans. Finally, all the spectra were subtracted from a baseline corresponding to the buffer alone (10 mM potassium phosphate buffer pH 7.4 also containing 100 mM KCl).

The spectroscopic properties of the unconjugated cyanine dyes formed upon reaction of **Ind1-2** and **Ind3** with **Ald** have been previously reported in the literature. Molecular extinction coefficients at the maximum absorption wavelength for the corresponding symmetrical and unsymmetrical dyes were $7.8 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ and $7.5 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$, respectively (for instance see S.J. Mason *et al.* *J. Org. Chem.* **2005**, *70*, 2939-2949 and A. Fegan *et al.* *Chem. Commun.* **2008**, 2004-2006).

2.1 Circular dichroism (CD) studies

The ability of **Quad1** and **Quad3** to fold into a quadruplex conformation was assessed using circular dichroism (CD) spectroscopy. Both sequences were folded into a 10 mM potassium phosphate buffer (pH 7.4) solution containing 100 mM KCl. Interestingly, under those conditions, they both form a parallel-stranded intramolecular quadruplex (FigS1), as demonstrated previously for the ckit21T sequence (lacking the single stranded flanking arms).



FigS1. CD spectrum of a 4 μ M solution of **Quad1** DNA folded in potassium phosphate buffer (10 mM, pH 7.4) containing 100 mM KCl.

2.2 Fluorescence studies

2.2.1 The kinetics of the reaction between **Pna1** and **Pna2** (500 nM each) in the presence of a stoichiometric amount of quadruplex DNA (**Quad1**) was investigated by fluorescence spectroscopy. A time-dependent increase of fluorescence was observed and equilibrium was reached after 2 hours (Fig S2).

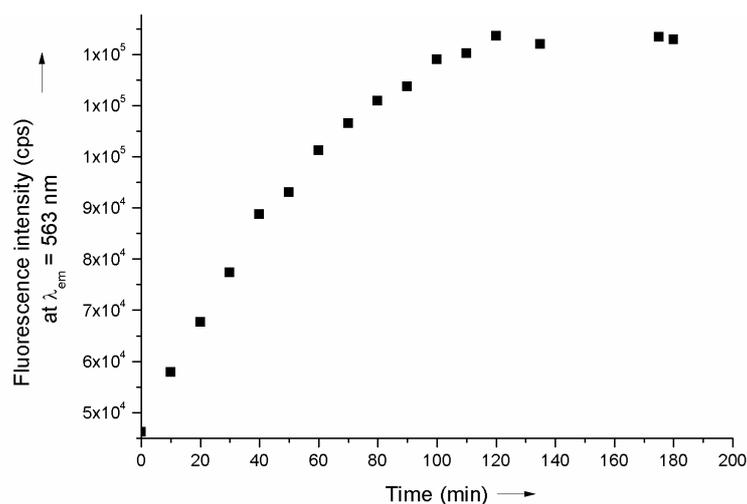


Fig S2. Plot of the fluorescence intensity ($\lambda_{\text{exc}} = 542 \text{ nm}$, $\lambda_{\text{em}} = 563 \text{ nm}$) vs time for a reaction between **Pna1** and **Pna2** (500 nM each) in the presence of a stoichiometric amount of quadruplex **Quad1** and in potassium phosphate buffer (10 mM, pH 7.4) also containing 100 mM KCl.

2.2.2 Fluorescence studies were carried out by reacting **Pna1** and **Pna2** (200 nM each) in the presence and absence of a stoichiometric amount of quadruplex DNA (**Quad1**). A time-dependent increase of fluorescence was observed and equilibrium was reached after 2 hours (Fig S3).

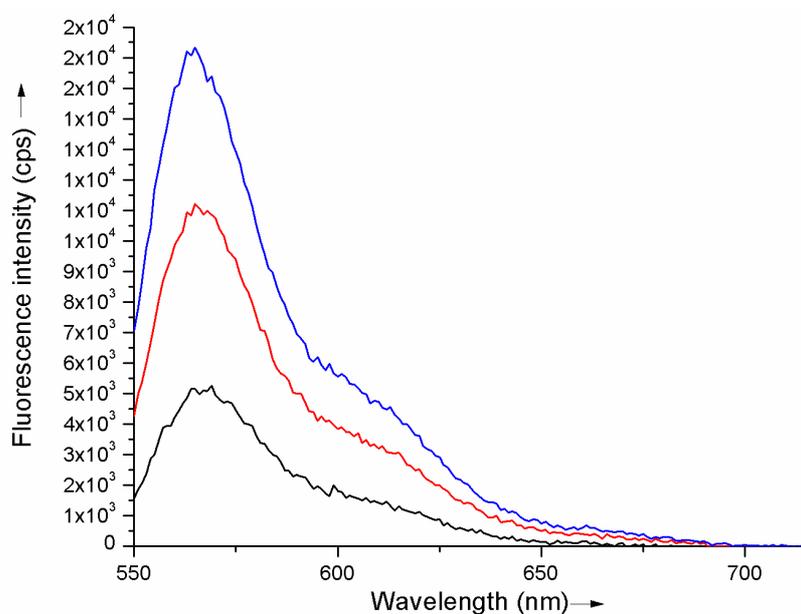


Fig S3. Fluorescence emission spectra ($\lambda_{\text{exc}} = 542 \text{ nm}$) of a mixture of **Pna1**, **Pna2** and **Quad1** (200 nM each) in potassium phosphate buffer (10 mM, pH 7.4) + 100 mM KCl after 10 min (black), 1h (red) and 2h (blue) at rt.

2.2.3 Fluorescence studies were carried out by reacting **Pna1** and **Pna3** (500 nM each) in the presence and absence of a stoichiometric amount of quadruplex DNA (**Quad1**). A time-dependent increase of fluorescence was observed and equilibrium was reached after 2 hours (Fig S4).

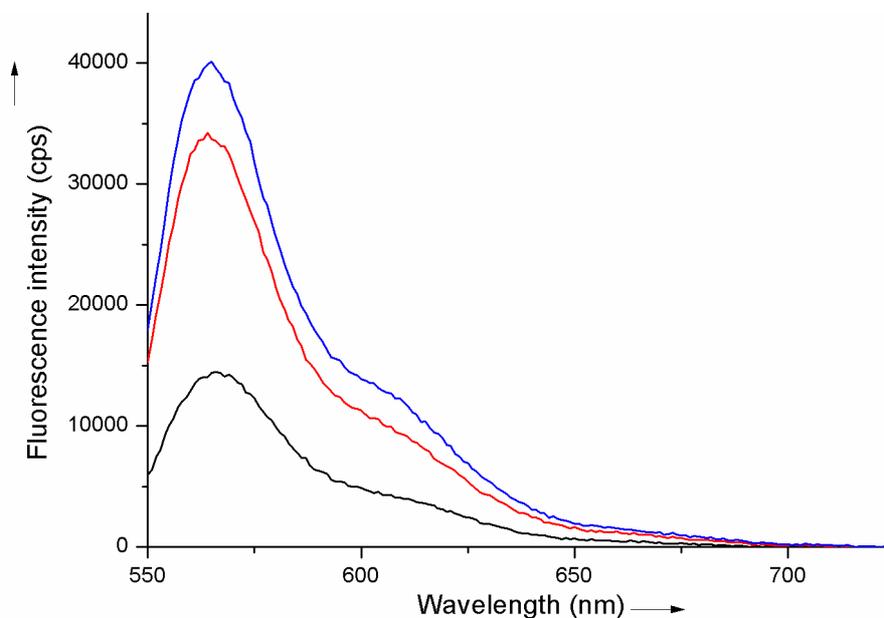


Fig S4. Fluorescence emission spectra ($\lambda_{exc} = 542$ nm) of a mixture of **Pna1**, **Pna3** and **Quad1** (200 nM each) in potassium phosphate buffer (10 mM, pH 7.4) + 100 mM KCl after 10 min (black), 1h (red) and 2h (blue) at rt.

Single nucleotide polymorphism detection using a biocompatible, fluorogenic and DNA-templated reaction of cyanine dye formation.

Girish Koripelly,^{‡a} Kamel Meguellati^{‡a} and Sylvain Ladame*^a **SUBMITTED MANUSCRIPT**

Received (in XXX, XXX) Xth XXXXXXXXX 200X, Accepted Xth XXXXXXXXX 200X

5 First published on the web Xth XXXXXXXXX 200X

DOI: 10.1039/b000000x

Water-soluble Peptide Nucleic Acids (PNAs) functionalised with fluorogenic cyanine dye precursors (i.e. a Fisher's base aldehyde and a 2-methylene indolenine) are employed for detecting single-nucleotide polymorphism in vitro under near physiological conditions of salt and pH. This quick and versatile fluorogenic system has been designed so that fluorescence is formed only upon simultaneous hybridisation of both PNAs to a fully complementary DNA strand. The reaction yield increased up to around 10% upon addition of the DNA template, even in the presence of a large excess of Calf-Thymus competitor DNA.

Introduction

15 Since the final sequencing of the human genome was completed, in 2003, considerable research effort has been devoted to the identification and analysis of variations among individual genomes.^{1,2} Although more than 99.9% of human DNA sequences are the same across the population, variations can have a major impact on how humans respond to diseases or drug therapies.³ Amongst the most common forms of genomic variation, single-nucleotide polymorphism (SNP) accounts for over 80% of sequence variants.⁴ Current estimates suggest that SNPs occur as frequently as every 100 to 300 bases and that they are stably inherited, making them a unique tool/marker for identifying numerous genetic and inherited complex diseases such as cancers and diabetes.⁴ Therefore, there is an increasing need for high throughput technologies suitable for rapid and sensitive SNP genotyping.⁵ So far, methods using short oligonucleotides or oligonucleotide mimics as probes to reveal the presence of complementary sequences have been successfully developed.⁶ Of particular interest are oligonucleotide-templated reactions that can be monitored with high sensitivity by the appearance/disappearance of a fluorescent signal upon binding to the oligonucleotide target.⁷ Representative examples of such technologies include the use of fluorogenic probes (e.g. molecular beacons⁸), or rely on fluorogenic reactions of chemical ligation⁹ or primer extension.¹⁰ Most recent reports are based on the Staudinger reaction,¹¹ diamine-catalysed aldol condensation,¹² organomercury-activated¹³ or S_NAr reactions.¹⁴ Among these examples, a small proportion only was successfully applied for the detection of SNPs in vitro. Briefly, two modified oligonucleotides (or oligonucleotide analogues) are designed so that they can hybridize in a sequence-specific manner to a unique nucleic acid template through Watson-Crick base pairing. Upon hybridization of the synthetic probes to the complementary DNA strand, both fluorogenic moieties are brought in close enough proximity to react with each other, thus generating the fluorescent product. In case of a partial or incomplete hybridization of one of the

modified oligonucleotides (e.g. as a consequence of a SNP), both probes are no longer positioned favorably to react with each other, thus leading to a significant decrease of reaction efficiency (i.e. a weaker fluorescence). Within such systems, genetic information can be directly linked to the appearance of a characteristic fluorescence signal.

Herein, we report a biocompatible and fluorogenic system that uses two water-soluble PNAs functionalised at their N- or C- terminus with non-fluorescent cyanine dye precursors for SNPs detection. Briefly, two 5-mer PNAs were synthesized on solid support that also contained two ϵ -N,N-dimethyl-Lysines to ensure water-solubility and were functionalised either at their N- or C-terminus with a Fischer's base aldehyde and 2-methylene-indolenine, respectively.¹⁵ Upon simultaneous hybridization of both PNAs to a complementary DNA template only, both fluorogenic probes will be orientated optimally to react with each other, thus producing irreversibly a highly fluorescent symmetrical trimethine cyanine dye (Figure 1).

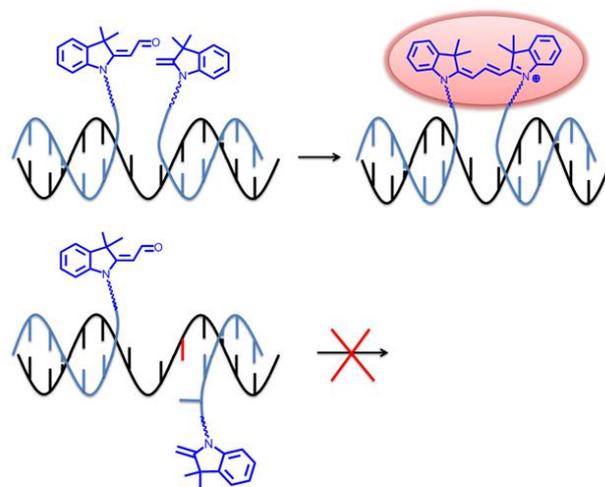


Fig 1. General strategy for SNPs detection using two PNAs functionalised with non-fluorescent trimethine cyanine dye precursors.

75

Results and discussion

The ability of **Pna1** and **Pna2** to react with each other in the absence and in the presence of various DNA sequences used as templates was investigated by fluorescence spectroscopy by monitoring the formation of a symmetrical trimethine cyanine dye ($\lambda_{\text{exc}} = 540 \text{ nm}$, $\lambda_{\text{em}} = 573 \text{ nm}$). Reactions were carried out in potassium phosphate buffer (40 mM, pH 7.4) at 37 °C using a stoichiometric mixture of **Pna1**, **Pna2** and DNA (10 μM each, total volume 50 μL) and the efficiency of the fluorogenic reaction was monitored over three hours in a fluorescence microplate reader. Importantly, no fluorescence was ever detected when incubating **Pna1** and **Pna2** under such conditions of concentration and pH and in the absence of any DNA template (Figure 3, grey curve).

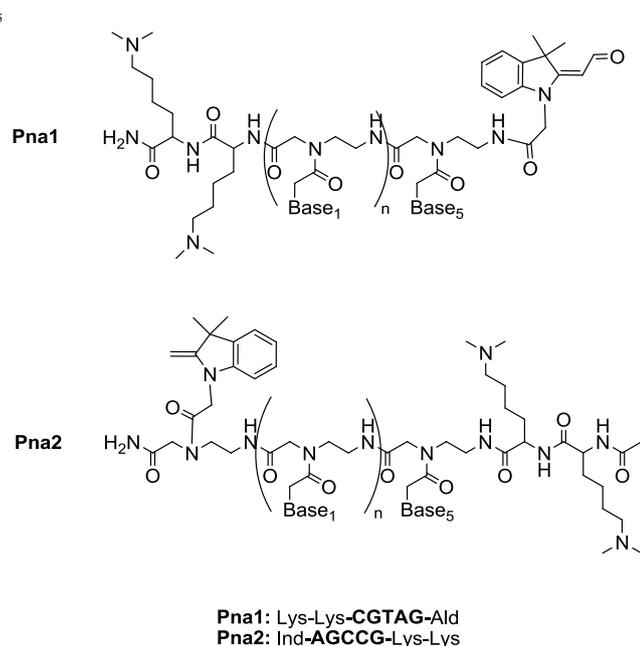


Fig 2. Structures and sequences of the fluorogenic PNA probes.

Table 1 Oligonucleotide sequences (given from the 5' to 3' end)^a

Name	Sequence	Name	Sequence
WT0	<u>GCATCTCGGC</u>	Mut5	<u>GCATCCTTCGGT</u>
WT1	<u>GCATCTTCGGC</u>	Mut6	<u>GCATCCTTCGGC</u>
WT2	<u>GCATCCTTCGGC</u>	Mut7	<u>GCATCCTT<u>X</u>CGGC^b</u>
WT3	<u>GCATCCTTTTCGGC</u>	Mut8	<u>GCAT<u>X</u>CCTTTCGGC^b</u>
WT4	<u>GCATCCTTTTCGGC</u>	Mut9	<u>GCATCCTT<u>AA</u>GGC</u>
Mut1	<u>GCATCCT<u>X</u>CGGC^b</u>	Mut10	<u>GCATCCTT<u>AA</u>AGC</u>
Mut2	<u>GCATCCTTTCGGC</u>	Mut11	<u>GCATACTT<u>AC</u>GGC</u>
Mut3	<u>GCATCCTTCTGC</u>	Mut12	<u>GCA<u>AA</u>CTT<u>AA</u>GGC</u>
Mut4	<u>GCATCCTTCGTC</u>		

^aUnderlined sequences are that complementary to **Pna1** and **Pna2**; ^b**X** represents A, T, C or G.

fluorogenic reaction. Five DNA strands (**WT0-4**, table 1), all containing both sequences complementary to **Pna1** and **Pna2** but separated by a variable number of nucleobases (from 0 to 4), were evaluated as potential candidates capable of templating the reaction of cyanine dye formation. While short distances between both PNAs (**WT0-2**) proved equally good at templating the fluorogenic reaction, slightly more flexible systems (e.g. **WT3** where PNA/DNA duplexes formed upon hybridisation of **Pna1** and **Pna2** with the DNA template are separated by three nucleotides) appeared more favorable (Figure 3). This tendency was confirmed when increasing the distance even further (up to 7 nucleotides, data not shown), although never exceeding the maximum fluorescence observed with a 3 nucleotide gap (**WT3**).

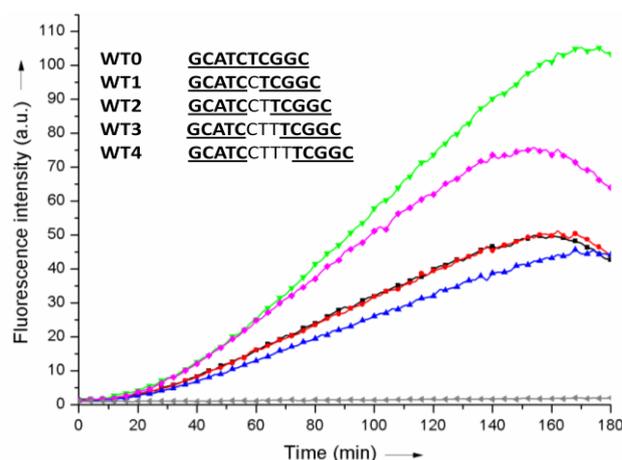


Fig 3. Reaction time-course of a stoichiometric mixture of **Pna1** and **Pna2** (10 μM each) in the absence (grey) and in the presence of a stoichiometric amount of **WT0** (blue), **WT1** (red), **WT2** (black), **WT3** (green) or **WT4** (pink). Reaction was monitored by fluorescence spectroscopy ($\lambda_{\text{exc}} = 540 \text{ nm}$, $\lambda_{\text{em}} = 573 \text{ nm}$).

Next, both **WT2** and **WT3** were selected for a more detailed study. The effect of systematic point mutations within the DNA template on the reaction efficiency was investigated. In the case of **WT3**, none of the single mutations tested led to a significant change in reaction efficiency, regardless of the nature and position of the mutation (Figure 4A). Significant inhibition of the fluorogenic reaction was observed only when mutating at least two consecutive residues in **WT3** while near-complete inhibition was obtained when mutating a minimum of three residues, hence preventing hybridization of at least one of the fluorogenic PNAs (Figure 4B).

Although **WT3** seemed optimal for templating the reaction of cyanine dye formation, it appeared that a 3-nucleotide gap between both PNA complementary sequences was probably making the PNA-DNA complex too flexible to be responsive to minimal changes such as single mutations.

We first investigated the influence of the distance (i.e. number of nucleotides) between both PNAs on the efficiency of the

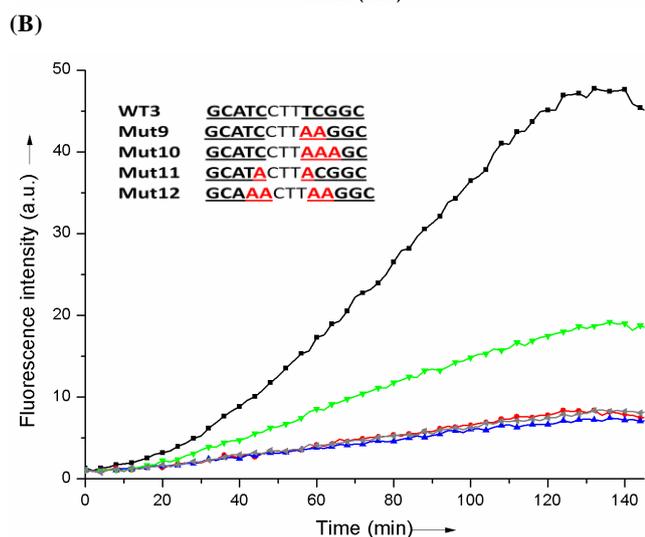
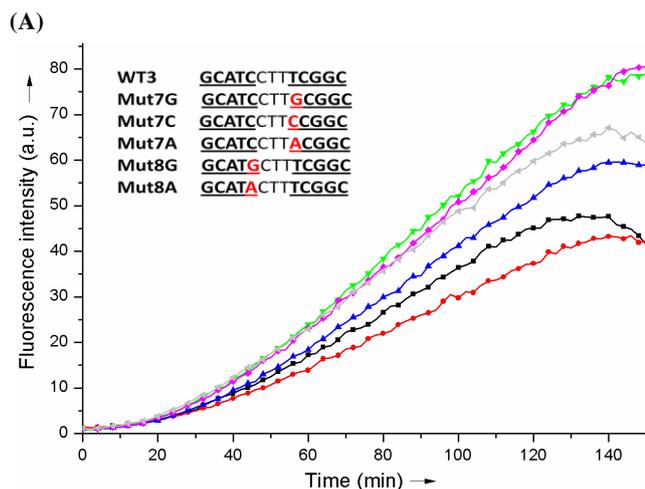


Fig 4. Reaction time-course of a stoichiometric mixture of **Pna1** and **Pna2** (10 μ M each) in the presence of a stoichiometric amount of **WT3** (black) or one of the mutated sequences **Mut7G** (red), **Mut7C** (blue), **Mut7A** (green), **Mut8G** (pink) or **Mut8A** (grey) (Fig. 4A), or **Mut9** (red), **Mut10** (blue), **Mut11** (green) or **Mut12** (grey) (Fig. 4B). Reaction was monitored by fluorescence spectroscopy ($\lambda_{exc} = 540$ nm, $\lambda_{em} = 573$ nm).

Therefore, and despite the fact that **WT2**, as a template, was not quite as efficient as **WT3**, we reasoned that a 2-nucleotide gap could provide a more rigid environment for the fluorogenic reaction so that the correct alignment of both non-fluorescent cyanine dye precursors would be more responsive to SNPs. DNA sequences derived from **WT2** but carrying single nucleotide mutations at various positions (**Mut1-6**, table 1) were tested for their ability to template the fluorogenic reaction. The reaction time-courses were determined by fluorescence spectroscopy on a microtiter plate format and compared to that obtained with the corresponding native sequence **WT2** (Figure 5).

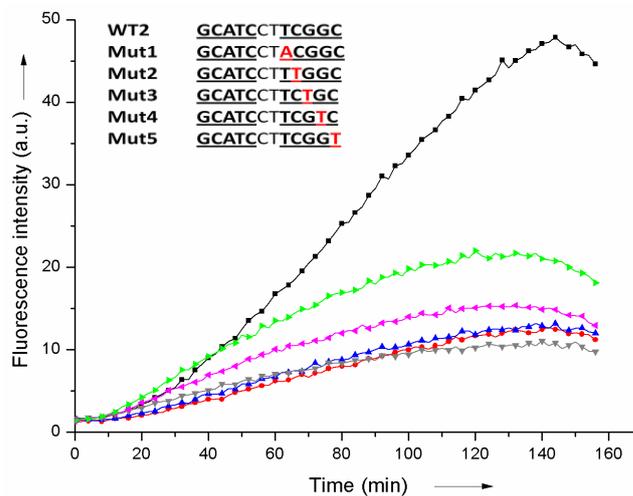


Fig 5. Reaction time-course of a stoichiometric mixture of **Pna1** and **Pna2** (10 μ M each) in the presence of a stoichiometric amount of **WT2** (black), **Mut1** (red), **Mut2** (blue), **Mut3** (grey), **Mut4** (pink), **Mut5** (green). Reaction was monitored by fluorescence spectroscopy ($\lambda_{exc} = 540$ nm, $\lambda_{em} = 573$ nm).

A significant decrease (up to 5-fold) in fluorescence intensity was observed when mutating one nucleotide within the sequence of **WT2** complementary to **Pna2** (**Mut1-5**, table 1). As previously observed by others, the strongest “inhibitory” effects were obtained with single mutations located near the indole fluorogenic probe (e.g. **Mut1** vs **Mut5**). However, the nature of the mutation had no influence on the reaction time-course. For instance, similar time courses were obtained for **Mut1**, regardless of the nature of X (A, C or G). Surprisingly, DNA strands carrying mutations within the sequence of **WT2** complementary to **Pna1** (**Mut6**, table 1) were found to be almost equally efficient templates when compared to the native sequence **WT2** (Figure S1, Supplementary Information). This is likely due to a greater conformational freedom of the Fischer’s base aldehyde allowing its reaction with the neighbouring indole even if **Pna1** is not fully hybridized to the DNA template. Altogether, these results suggest that a constrained system in which two 5-mer fluorogenic PNAs can hybridize, simultaneously and two nucleotides apart, to a unique DNA strand serving as a reaction template, can be used for SNPs detection. Most sensitive detection is achieved when the mutation is located opposite the indole’s neighbouring nucleobase (**Mut1**) although the system remains sensitive enough for detecting mutations located further away in the DNA template (**Mut2-4**).

Finally, in order to estimate the efficiency of the fluorogenic reaction upon addition of complementary DNA, a water-soluble analogue of the symmetrical trimethine cyanine dye reaction product was synthesized and used as a standard for quantifying the amount of dye formed in the absence or presence of various DNA templates (Figure S2, Supplementary Information). Using this external standard, the

yield of the reaction templated by **WT2** was found to be approximately $10 \pm 5\%$ while no fluorescent product was ever detectable under similar conditions but in the absence of template (Figure 6). Interestingly, addition of a large excess (20 $\mu\text{g/mL}$) of competitor Calf-Thymus DNA in the reaction mixture resulted in no significant loss of reaction efficiency (Figure S3, Supplementary Information), thus confirming the high specificity of our fluorogenic system.

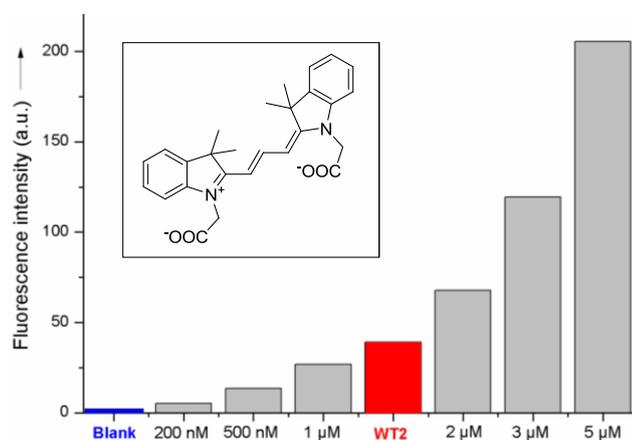


Fig 5. Fluorescence calibration curve (fluorescence intensity versus dye concentration, $\lambda_{\text{exc}} = 540\text{ nm}$, $\lambda_{\text{em}} = 573\text{ nm}$) obtained from an analytically pure sample of symmetrical trimethine cyanine dye similar to the product of the fluorogenic reaction (inset). The fluorescence intensities obtained at equilibrium (after 3 hours) for a stoichiometric mixture of **Pna1** and **Pna2** (10 μM each) in the absence and in the presence of a stoichiometric amount of **WT2** are highlighted in blue and red, respectively

Conclusion

In conclusion, we have reported the use of a fast, versatile, biocompatible and fluorogenic DNA-templated reaction of cyanine dye formation for the detection of SNPs in vitro. The system offers the advantage of a reasonably high sensitivity due to the high absorptivity and moderately high fluorescence quantum yield of the cyanine dye formed¹⁶ and requires minimal amounts (500 pmol) of both DNA and PNA probes when the reaction is carried out in a 384-well microtiter plate. The design of optimized (e.g. more rigid, multi-coloured) systems to further improve the sensitivity of the detection is currently underway in our laboratory.

Experimental

General

DNase- and RNase-free water, potassium phosphate buffer and Calf-Thymus DNA were purchased from Sigma-Aldrich. PNA monomers were purchased from ASM Research

Chemicals (Germany) and Rink amide resin for solid phase synthesis from Merck Biosciences (UK). DNA oligonucleotides were purchased from Sigma and were all HPLC purified (Table 1). Synthesis and Characterisation (HR-MALDI) of both purified fluorogenic PNAs (**Pna1** and **Pna2**) was recently reported by us.¹⁵

Sensing experiments

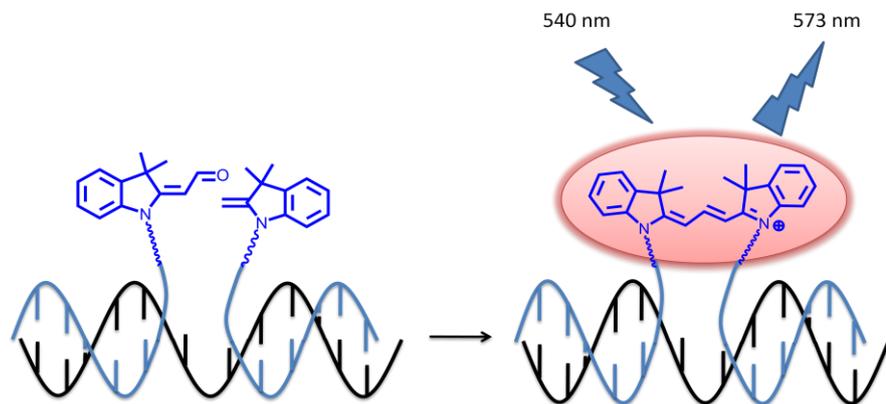
Stock solutions (50 μM) of **Pna1**, **Pna2** and DNA were prepared in water. In a typical experiment, 20 μL of a potassium phosphate buffer solution (100 mM, pH = 7.4) and 10 μL of each stock solution were transferred into one well to make the final concentrations of 10 μM **Pna1**, 10 μM **Pna2**, 10 μM DNA and 40 mM buffer. Reaction time-courses were determined in a 384-well plate using a SpectraMax M5 fluorescence plate reader (Molecular devices, UK). Reactions were carried out at 37°C with $\lambda_{\text{exc}} = 540\text{ nm}$ and $\lambda_{\text{em}} = 573\text{ nm}$.

Notes and references

- ^a ISIS Université de Strasbourg, 8 allée Gaspard Monge, BP 70028, 67083 Strasbourg, France. Fax: (+33) 368855115; Tel: (+33) 368855210; E-mail: s.ladame@isis.u-strasbg.fr
- † Electronic Supplementary Information (ESI) available: Reaction time-courses of a stoichiometric mixture of **Pna1** and **Pna2** in the presence of various DNA sequences as templates. See DOI: 10.1039/b000000x/
- ‡ These authors contributed equally to this work.
- F. S. Collins, E. D. Green, A. E. Gutmacher and M. S. Guyer, *Nature*, 2003, **422**, 835-847; J. C. Venter *et al.*, *Science*, 2001, **291**, 1304-1351.
 - For recent reviews, see P. Stankiewicz, and J. R. Lupski, *Annu. Rev. Med.*, 2010, **61**, 437-455; L. Feuk, A. R. Carson and S. W. Scherer, *Nat. Rev. Genet.*, 2006, **7**, 85-97.
 - D. Althuler, V. J. Pollara, C. R. Cowles, W. J. Van Etten, J. Baldwin, L. Linton and E. S. Lander, *Nature*, 2000, **407**, 513-516; M. K. Halushka, J. B. Fan, K. Bentley, L. Hsie, N. Shen, A. Weder, R. Cooper, R. Lipshutz and A. Chakravarti, *Nat. Genet.*, 1999, **22**, 239-247.
 - E. N. Imyanitov, *Human Genet.*, 2009, **125**, 239-246; K. A. Frazer *et al.*, *Nature*, 2007, **449**, 851-861; L. Kruglyak and D. A. Nickerson, *Nat. Genet.*, 2001, **27**, 234-236.
 - For recent reviews see C. Ding and S. Jin, *Methods Mol. Biol.*, 2009, **578**, 245-254; T. LaFramboise, *Nucleic Acids Res.*, 2009, **37**, 4181-4193; S. Kim and A. Misra, *Annu. Rev. Biomed. Eng.*, 2007, **9**, 289-320.
 - B. W. Kirk, M. Feinsod, R. Favis, R. M. Kliman and F. Barany, *Nucleic Acids Res.*, 2002, **30**, 3295-3311.
 - R. T. Ranasinghe and T. Brown, *Chem. Commun.*, 2005, 5487-5502.
 - N. Venkatesan, Y. J. Seo and B. H. Kim, *Chem. Soc. Rev.*, 2008, **37**, 648-663; S. A. Marras, F. R. Kramer and S. Tyagi, *Methods Mol. Biol.*, 2003, **212**, 111-128.
 - D. K. Toubanaki, T. K. Christopoulos, P. C. Ioannou and C. S. Flordellis, *Anal. Chem.*, 2009, **81**, 218-224.
 - A. Krieg, S. Laib, T. Ruckstuhl and S. Seeger, *ChemBioChem.*, 2004, **5**, 1680-1685.
 - J. Cai, X. Li, X. Yue and J. S. Taylor, *J. Am. Chem. Soc.*, 2004, **126**, 16324-16325; Z. L. Pianowski and N. Winsinger, *Chem. Commun.*, 2007, 3820-3822; R. M. Franzini and E. T. Kool, *ChemBioChem*, 2008, **9**, 2981-2988; Z. Pianowski, K. Gorska, L. Oswald, C. A. Merten and N. Winsinger, *J. Am. Chem. Soc.*, 2009, **131**, 6492-6497; K. Furukawa, H. Abe, K. Hibino, Y. Sako, S. Tsuneda and Y. Ito, *Bioconjugate Chem.*, 2009, **20**, 1026-1036.
 - Y. Huang and J. M. Coull, *J. Am. Chem. Soc.*, 2008, **130**, 3238-3239.

-
- 13 R. M. Franzini and E. T. Kool, *Org. Lett.*, 2008, **10**, 2935-2938.
- 14 A. Shibata, H. Abe, M. Ito, Y. Kondo, S. Shimizu, K. Aikawa and Y. Ito, *Chem. Commun.*, 2009, 6586-6588.
- 15 K. Meguellati, G. Koripelly and S. Ladame, *Angew. Chem. Int. Ed.*,
2010, **49**, 2738-2742.
- 5 16 R. C. Benson and H. A. Kues, *J. Chem. Eng. Data*, 1977, **22**, 379-383.

Graphical abstract



**Single nucleotide polymorphism detection using a biocompatible,
fluorogenic and DNA-templated reaction of cyanine dye
formation.**

*Girish Koripelly, Kamel Meguellati and Sylvain Ladame**

ISIS Université de Strasbourg, 8 Allée Gaspard Monge, BP 70028, 67083 Strasbourg, France.
Fax: (+33) 368855115; Tel: (+33) 368855210 E-mail: s.ladame@isis.u-strasbg.fr

SUPPORTING INFORMATION

Effect of the SNP position on the “templating effect” of WT2

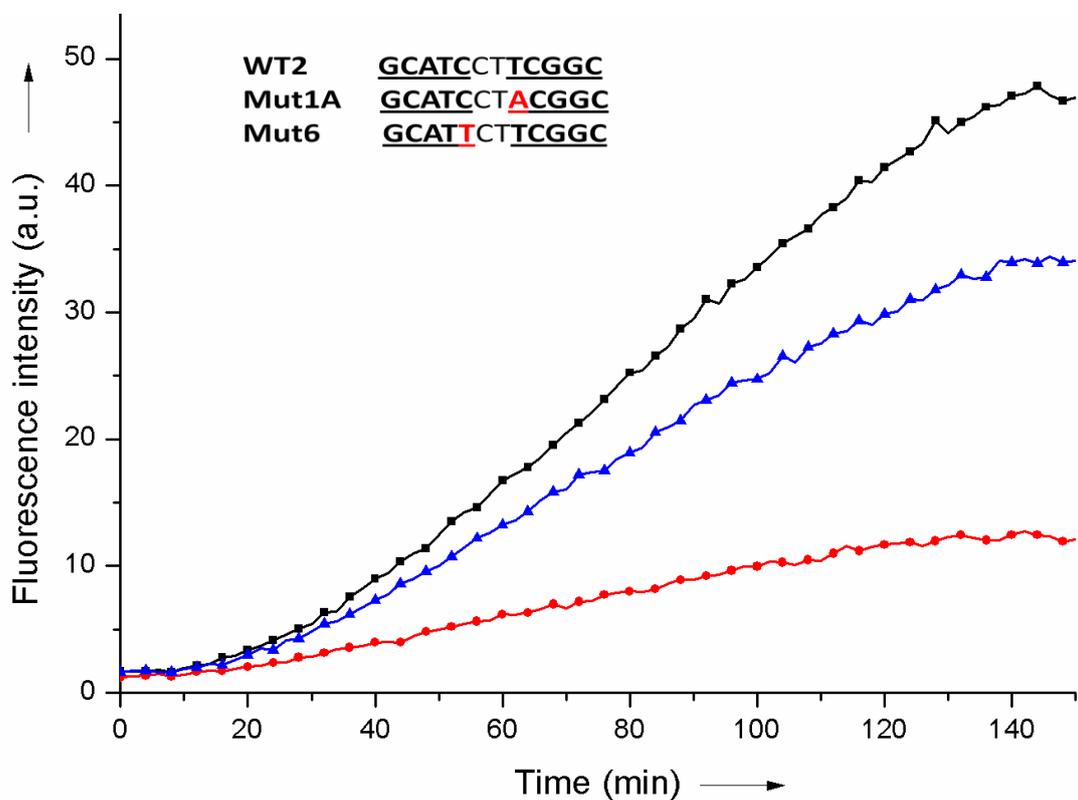


Figure S1. Reaction time-course of a stoichiometric mixture of **Pna1** and **Pna2** (10 μ M each) in the presence of a stoichiometric amount of **WT2** (black), **Mut1A** (red), or **Mut6** (blue). Reaction was monitored by fluorescence spectroscopy ($\lambda_{\text{exc}} = 540$ nm, $\lambda_{\text{em}} = 573$ nm).

Fluorescence Calibration Curve

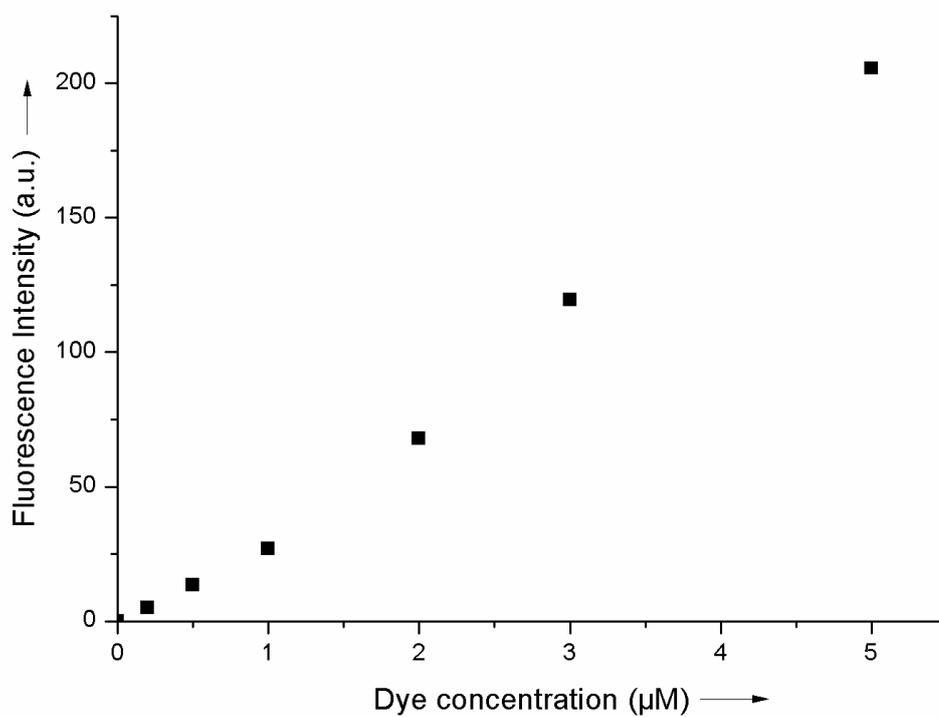
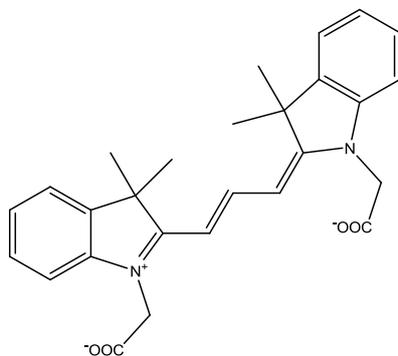


Figure S2. Structure of the water-soluble trimethine cyanine dye, analogue of the product of the fluorogenic reaction (top) and fluorescence calibration curve (bottom).

Effect of competitor CT DNA on the “templating effect” of WT2

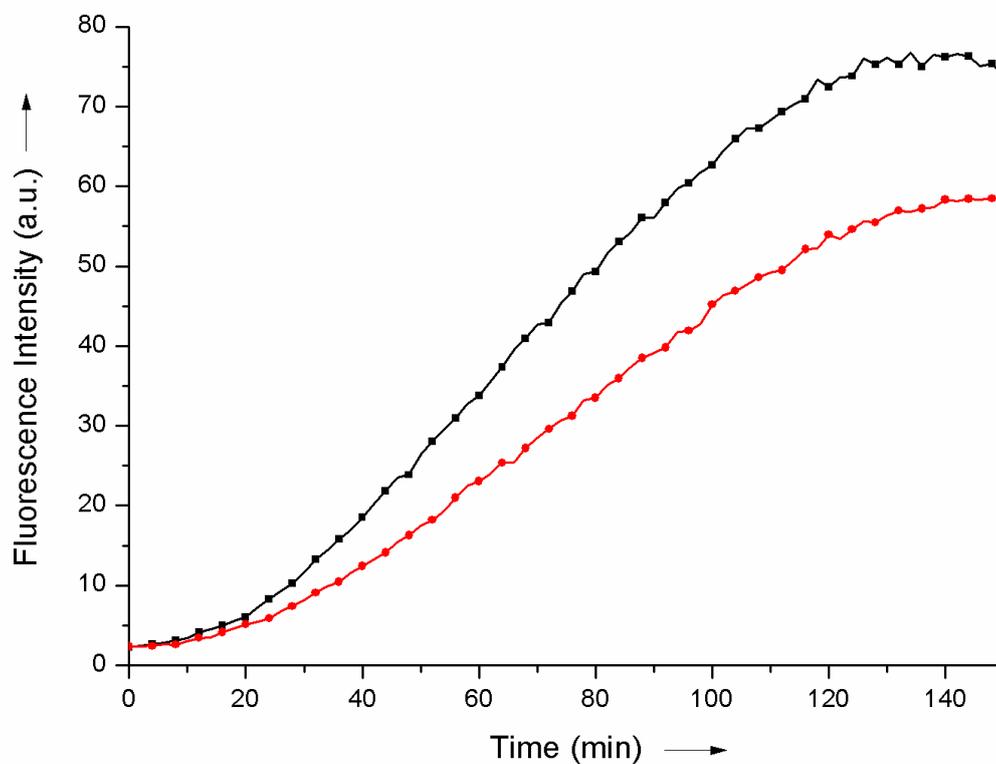


Figure S3. Reaction time-course of a stoichiometric mixture of **Pna1** and **Pna2** (10 μ M each) in the presence of a stoichiometric amount of **WT2** and in the absence (black) or presence (red) of Calf-Thymus competitor DNA (20 μ g/mL). Reaction was monitored by fluorescence spectroscopy ($\lambda_{\text{exc}} = 540$ nm, $\lambda_{\text{em}} = 573$ nm).

Dual Sensing of Hairpin and Quadruplex DNA Structures Using Multicolored Peptide Nucleic Acid Fluorescent Probes

Girish Koripelly,[†] Kamel Meguellati,[†] and Sylvain Ladame^{*,‡}

Institut de Science et d'Ingénierie Supramoléculaires (ISIS), CNRS UMR 7006, Université de Strasbourg, Strasbourg, France.
Received July 23, 2010; Revised Manuscript Received September 6, 2010

Synthesis of water-soluble 5-mer peptide nucleic acids (PNAs) functionalized at their 5'- and 3'-ends with two original precursors of pentamethine cyanine dye synthesis is reported. The successful use of these PNA probes for sensing DNA hairpin structures in vitro was also demonstrated where specific hairpin formation was associated with the appearance of a characteristic fluorescence signal at 660 nm. A comparative study between three different strategies where PNAs were targeting either the stem or the loop of the hairpin was carried out. Best sensitivity was obtained using PNA sequences complementary to the loop sequence and directing both functional moieties toward the base of loop. Unprecedented proof-of-concept for the simultaneous sensing of hairpin and quadruplex DNAs with a nonoverlapping two-color system (C3 and C5) is also demonstrated.

INTRODUCTION

Although the most abundant form of DNA in vivo is the double-stranded helical conformation reported for the first time in 1953 by Watson and Crick (1), it has been known for more than 50 years that nucleic acids can also adopt alternative (i.e., non B-DNA) secondary structures under physiological conditions. Such structures can be either single-stranded (e.g., hairpins, intramolecular G-quadruplexes, pseudoknots) or double-stranded (e.g., cruciforms, bimolecular G-quadruplexes) and have received increasing attention because of the role they were proposed to play in transcription regulatory mechanisms (2). For instance, it has been known for several decades that DNA sequences containing a high density of clustered guanines were able to adopt four-stranded secondary structures named guanine (G)-quadruplexes or tetraplexes in the presence of physiological cations, notably K^+ and Na^+ (Figure 1) (3–7). More recently, in silico studies have revealed the high prevalence of such G-rich DNA sequences throughout the human genome, particularly in gene promoter regions (8–10). Hairpin formation has also been demonstrated both in vitro and in vivo (11). Stem-loop hairpin structures are obtained by folding of a partially self-complementary or palindromic DNA sequence (Figure 1). Their presence in gene promoters (12) or near gene replication origins (13), for instance, strongly suggests that DNA hairpins could play a functional role in naturally occurring regulatory mechanisms of gene expression. For this reason, non-Watson–Crick DNA secondary structures represent attractive therapeutic targets to regulate the expression of specific genes via chemical or biochemical intervention. Successful examples of gene regulation with small molecules that bind to and stabilize hairpin or quadruplex DNA (or RNA) have been recently reported in the literature (14–18). Although these studies represent converging evidence that such structures do form in the context of the cell,

there is still an urgent need for chemical probes capable of sensing the formation of nucleic acid structures both in vitro and in vivo.

There is currently a growing interest in the development of new fluorescent chemosensors based on supramolecular systems which have the ability to recognize and bind to a specific metabolite with high affinity. As a result of this interaction, the system re-equilibrates, thus leading to a change of the physical properties (i.e., changes in fluorescence intensity and/or fluorescence emission wavelength) of the medium. Since the 1960s, engineered DNA molecules have been used extensively as highly specific fluorescent probes in order to explore various biological processes (e.g., DNA transcription, DNA recombination) and more recently for genotyping applications. Among them, molecular beacons (MBs) have received particular attention because of their broad applications in modern biochemistry (19). Typically, MBs are DNA hairpin structures functionalized at their 5'- and 3'-ends with a fluorescence donor and a fluorescence acceptor. Binding of the MB to a DNA or RNA target, through sequence-specific hybridization, results in the opening of the MB, thus separating the fluorescence donor from its neighboring fluorescence acceptor. Other examples include the use of modified (e.g., fluorescently labeled) oligonucleotides or oligonucleotide mimics in oligonucleotide-templated reactions. In such systems, a specific reaction of dye formation occurs only upon hybridization of the probes to a complementary nucleic acid target. A characteristic fluorescent signal will then be emitted which can be directly linked to the detection of the

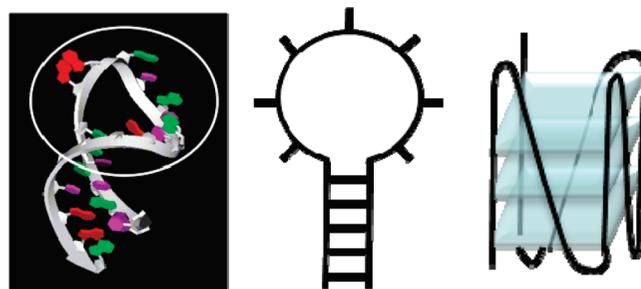


Figure 1. X-ray crystal structure and schematic representation of a DNA hairpin loop and schematic representation of a parallel-stranded intramolecular G-quadruplex (from left to right).

* To whom correspondence should be addressed. Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, CNRS UMR 7006, 8, allée Gaspard Monge, 67083 Strasbourg Cédex, France. Fax: (+33) 368855115, E-mail: s.ladame@isis.u-strasbg.fr.

[†] Both authors contributed equally to this work.

[‡] Current address: Imperial College London, Department of Bioengineering, Royal School of Mines Building, South Kensington Campus, London SW7 2AZ, U.K. E-mail: sladame@imperial.ac.uk, Phone: (+44) 2075945308.

targeted sequence and/or structure (20). Most recent reports of DNA-templated reactions applied to oligonucleotide sensing are based on the Staudinger reaction (21–25), aldol-type (26, 27), organomercury-activated (28), or S_NAr (29) reactions.

Because peptide nucleic acids (30, 31) offer the advantage of a higher stability, they have been recently used as an alternative to DNA for the design of highly specific nucleic acid probes (32–36). We recently reported the use of fluorogenic PNA probes for sensing G-quadruplex formation. Upon binding of both probes to the single-stranded flanking arms of a quadruplex, formation of a trimethine cyanine dye from two nonfluorescent precursors was templated (27). Herein, we describe the synthesis of a novel fluorescent Fisher's base aldehyde probe-head, its incorporation in a PNA strand, and the applications of the resulting fluorescently labeled probe for sensing hairpin DNA secondary structures. Reaction between this fluorescent Fisher's base aldehyde and a nonfluorescent indolenine molecule leads to the formation of a pentamethine cyanine dye (Cy5) absorbing and emitting at a significantly longer wavelength than its precursor. The system has been designed such that the Cy5 product forms solely upon simultaneous hybridization of both probes to either the loop or the stem of a unique hairpin DNA target. Therefore, hairpin formation will be directly linked to the appearance of a fluorescent signal at long wavelength ($\lambda_{em} > 650$ nm), well above the residual fluorescence of biological media, and above the fluorescence emission wavelength of most fluorescent products formed in oligonucleotide-templated reactions (21–29), including that previously reported by us (27). Simultaneous sensing of various DNA structures via a two-color (Cy3-Cy5) system will also be presented.

EXPERIMENTAL PROCEDURES

Ethyl 2-(3,3-Dimethyl-2-methyleneindolin-1-yl)-acetate (1). A mixture of 2,3,3-trimethyl-3*H*-indole (1.00 g, 6.28 mmol) and ethyl iodoacetate (1.61 g, 7.53 mmol) was heated at 80 °C for 5 h. After cooling the reaction mixture to room temperature, the red solid obtained was triturated with ether, filtered, and dried under vacuum. To the solid, 1 N NaOH (25 mL) was added and stirred for 1 h at RT. The reaction mixture was then extracted with CH_2Cl_2 (3 × 30 mL), and the combined organic layers were dried over Na_2SO_4 , filtered, and concentrated in vacuo. The residue was purified by column chromatography on SiO_2 (PE/ether, 3:1; Rf = 0.48) to give the title compound (1.28 g, 83%) as a yellow oil. 1H NMR (400 MHz, DMSO): δ = 7.16 (dd, J = 7.3, 0.9 Hz, 1H), 7.06 (td, J = 7.7, 1.2 Hz, 1H), 6.74 (td, J = 7.5, 0.6 Hz, 1H), 6.67 (d, J = 7.9 Hz, 1H), 4.39 (s, 2H), 4.10 (q, J = 7.2 Hz, 2H), 3.87 (d, J = 2.2 Hz, 1H), 3.81 (d, J = 2.0 Hz, 1H), 1.29 (s, 6H), 1.17 (t, J = 7.2 Hz, 3H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 168.6, 161.0, 145.4, 136.7, 127.3, 121.8, 118.8, 105.7, 74.8, 60.4, 43.6, 43.5, 29.6 (× 2) 14.1 ppm. HRMS (ESI) m/z : calcd for $C_{15}H_{19}NO_2$ [M+H]⁺ 246.149; found 246.150.

2-(3,3-Dimethyl-2-methyleneindolin-1-yl)acetic Acid (2). To a solution of ethyl 2-(3,3-dimethyl-2-methyleneindolin-1-yl)acetate (1.30 g, 5.30 mmol) in ethanol (40 mL) was added 1 N NaOH solution (30 mL) at RT, and the mixture was stirred at the same temperature overnight. Later, the solvent was removed in vacuo, and the residue was dissolved in a minimum amount of water (15 mL) and then acidified to pH = 3 with 1 N HCl solution. The solution was concentrated, and the residue was dissolved in ethanol. The suspension was filtered, and this process was repeated at least two times. Finally, the filtrate was concentrated, and the residue was purified by column chromatography on SiO_2 (CH_2Cl_2 /MeOH 8:2; Rf = 0.20) to give the title compound (860 mg, 75%) as a thick brown oil. 1H NMR (400 MHz, DMSO): δ = 7.15 (dd, J = 7.3, 0.7 Hz, 1H), 7.06

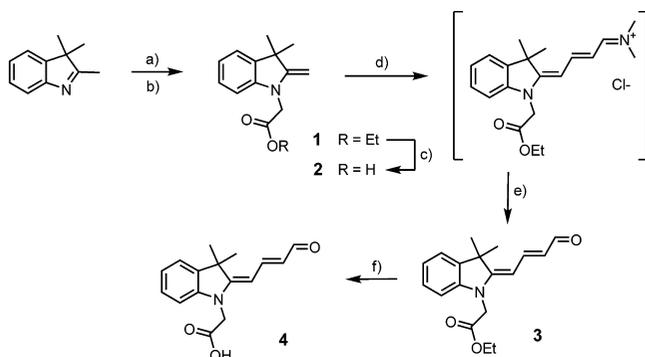
(td, J = 7.7, 1.2 Hz, 1H), 6.73 (t, J = 7.2 Hz, 1H), 6.65 (d, J = 7.9 Hz, 1H), 4.26 (s, 2H), 3.85 (d, J = 2.0 Hz, 1H), 3.81 (d, J = 2.0 Hz, 1H), 1.28 (s, 6H) ppm, CO_2H signal not visible. ^{13}C NMR (100 MHz, DMSO): δ = 170.1, 160.8, 145.5, 136.7, 127.3, 121.7, 118.6, 105.6, 74.4, 43.6, 43.5, 29.7 (× 2) ppm. HRMS (ESI) m/z : calcd for $C_{13}H_{15}NO_2$ [M+H]⁺ 218.116; found 218.119.

Ethyl 2-((2*E*)-2-((*E*)-3-Formylallylidene)-3,3-dimethylindolin-1-yl)acetate (3). Acetic anhydride (12.5 mL) was added dropwise to dimethylaminoacrolein (1.76 g, 1.78 mL, 17.8 mmol) and the mixture was stirred at RT for 15 min. Then, a solution of ethyl 2-(3,3-dimethyl-2-methyleneindolin-1-yl)acetate (1.25 g, 5.09 mmol) in CH_2Cl_2 (5 mL) was added dropwise. After stirring the reaction mixture for 2 d at RT, the solvent was removed in vacuo. The residue was dissolved in THF (30 mL), and a saturated solution of K_2CO_3 (60 mL) was added carefully. After stirring the mixture for 30 min at RT, the layers were separated and the aqueous layer was extracted with EtOAc (3 × 20 mL). The combined organic layers were dried over Na_2SO_4 , filtered, and concentrated in vacuo. The residue was purified by column chromatography on SiO_2 (CH_2Cl_2 /EtOAc 9.5:0.5; Rf = 0.20) to give the title compound (1.32 g, 86%) as a yellow oil. 1H NMR (400 MHz, DMSO): δ = 9.46 (d, J = 8.3 Hz, 1H), 7.88–7.82 (m, 1H), 7.37 (d, J = 7.3 Hz, 1H), 7.19 (td, J = 7.8, 1.1 Hz, 1H), 6.98–6.93 (m, 2H), 5.81 (dd, J = 14.2, 8.3 Hz, 1H), 5.62 (d, J = 12.6 Hz, 1H), 4.70 (s, 2H), 4.14 (q, J = 7.3 Hz, 2H), 1.59 (s, 6H), 1.19 (t, J = 7.3 Hz, 3H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 192.4, 167.8, 164.6, 150.2, 143.4, 138.5, 127.8, 123.6, 121.9, 121.4, 107.8, 95.3, 61.0, 46.3, 43.7, 27.9 (× 2), 14.0 ppm. HRMS (ESI, positive mode) m/z : calcd for $C_{18}H_{21}NO_3$ [M+H]⁺ 300.159; found 300.160.

2-((2*E*)-2-((*E*)-3-Formylallylidene)-3,3-dimethylindolin-1-yl)acetic Acid (4). To a solution of ethyl 2-((2*E*)-2-((*E*)-3-formylallylidene)-3,3-dimethylindolin-1-yl)acetate (300 mg, 1.00 mmol) in EtOH (2 mL) was added 1 N NaOH (1.5 mL) at RT and the mixture was stirred at the same temperature for 1.5 h. Later, the solvent was removed in vacuo, and the residue was dissolved in a minimum amount of water (3 mL) and then acidified to pH = 3 with 1 N HCl solution. The aqueous layer was extracted with EtOAc (4 × 10 mL). The combined organic layers were dried over Na_2SO_4 , filtered, and concentrated in vacuo to give the title compound (260 mg, 96%) as a blue–green solid. The crude product was directly used for PNA coupling without further purification. 1H NMR (400 MHz, DMSO): δ = 13.1 (br. s, 1H), 9.45 (d, J = 8.3 Hz, 1H), 7.88–7.81 (m, 1H), 7.36 (d, J = 7.0 Hz, 1H), 7.19 (td, J = 7.8, 1.1 Hz, 1H), 6.98–6.93 (m, 2H), 5.81 (dd, J = 14.2, 8.3 Hz, 1H), 5.62 (d, J = 12.6 Hz, 1H), 4.58 (s, 2H), 1.59 (s, 6H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 192.3, 169.1, 164.8, 150.4, 143.6, 138.6, 127.8, 123.3, 121.8, 121.3, 107.9, 95.1, 46.4, 43.7, 27.9 (× 2) ppm. HRMS (ESI, negative mode) m/z : calcd for $C_{16}H_{17}NO_3$ [M-H]⁻ 270.112; found 270.110.

PNA Synthesis. PNAs were synthesized on a solid support using Fmoc chemistry. The PNA strand was cleaved from the resin with trifluoroacetic acid/triisopropylsilane/ H_2O (95:2.5:2.5) solution and purified by HPLC. The final compounds were characterized by MALDI-TOF. **Pna1** m/z : [M+H]⁺ calcd for $C_{87}H_{116}N_{38}O_{19}^+$ 1997.928, found 1997.718. **Pna2** m/z : [M+H]⁺ calcd for $C_{88}H_{124}N_{39}O_{19}^+$ 2030.993, found 2030.923. **Pna3** m/z : [M+H]⁺ calcd for $C_{84}H_{115}N_{36}O_{19}^+$ 1931.913, found 1932.082.

DNA Folding. Hairpin and quadruplex DNA folding was carried out by heating a DNA solution (5 μ M) in potassium phosphate buffer (10 mM, pH 7.4 also containing 100 mM KCl) at 95 °C for 10 min and then letting it slowly cool down to room temperature over 6 h.

Scheme 1. Synthesis of 2-Methylene-Indolenin Derivative and Fischer's Base Aldehyde^a


^a Reagents and conditions: (a) ethyl iodoacetate, 80 °C, 5 h; (b) 1 N NaOH, EtOH, 25 °C, 1 h (83% over two steps); (c) 1 N NaOH, EtOH, 25 °C, 16 h (75%); (d) dimethylaminoacrolein, AcOH, CH₂Cl₂, 48 h; (e) sat. aq. K₂CO₃, THF, 25 °C, 1 h, (86% over two steps); (f) 1 N NaOH, EtOH, 25 °C, 1.5 h (96%).

Fluorescence Studies. Fluorescence emission spectra were recorded in quartz cells (500 μL with a 5 mm path length) at 20 °C on a Jobin Yvon Fluorolog 3.22 instrument. The excitation and emission bandwidths were fixed to 5 and 5 nm, respectively.

Ligation/Sensing Experiments. In a typical experiment, 500 μL of a mixture of **Pna1** (500 nM), **Pna2** (500 nM), and prefolded hairpin DNA (500 nM) in a potassium phosphate buffer (10 mM, pH 7.4 also containing 100 mM KCl) was prepared freshly in a fluorescence Quartz cuvette (5 mm path length) and kept sealed at 20 °C for the duration of experiment. Formation of the pentamethine cyanine dye product was then monitored by fluorescence spectroscopy (at *t*₀ and then every hour over 10 h).

Dual Sensing Experiments. After a screening of various concentrations of both DNA and PNA sequences, the following ratio 1:4:4:2:4 **Pna1/Pna2/Pna3/Hairp1a/Quad1** was found to give the optimal sensitivity and selectivity. Working with stoichiometric amounts of all PNA and DNA strands resulted in good sensitivity but lower selectivity (data not shown). Ligation experiments in the presence of both quadruplex and hairpin DNA were carried out in 500 μL quartz cuvettes by incubating a freshly prepared mixture of three PNA probes (**Pna1–3**) and two DNA targets (**Hairp1a** and **Quad1**) in potassium phosphate buffer.

RESULTS AND DISCUSSION

Design and Synthesis of the PNA Probes. Two 5-mer PNAs (**Pna1** and **Pna2**) were synthesized on solid support, which also contained two ε-*N,N*-dimethyl lysine residues to ensure solubility of both PNAs in water at near-physiological pH. These PNA probes were functionalized at their C- or N-terminus with two precursors for the synthesis of pentamethine cyanine dyes: *N*-alkyl-2,3,3-trimethylindolenine **2** and aldehyde **4**, respectively. Both precursors were synthesized from commercially available 2,3,3-trimethylindoline (Scheme 1).

Reaction of 2,3,3-trimethyl-3*H*-indole with ethyl iodoacetate gave the corresponding quarternized salt which was subsequently isomerized to 2-methylene-indolenin **1** via treatment with 1 N NaOH solution. Condensation of **1** with dimethylaminoacrolein in the presence of acetic anhydride provided the activated hemicyanine intermediate, which was then hydrolyzed into aldehyde **3** with a saturated K₂CO₃ solution. Final hydrolysis of the ethyl ester of compounds **1** and **3** using 1 N NaOH afforded both C5 precursors **2** and **4**, respectively. While indolenin **2** is nonfluorescent, aldehyde **4** emits maximally at 476 nm when excited at 410 nm. Reaction between compounds

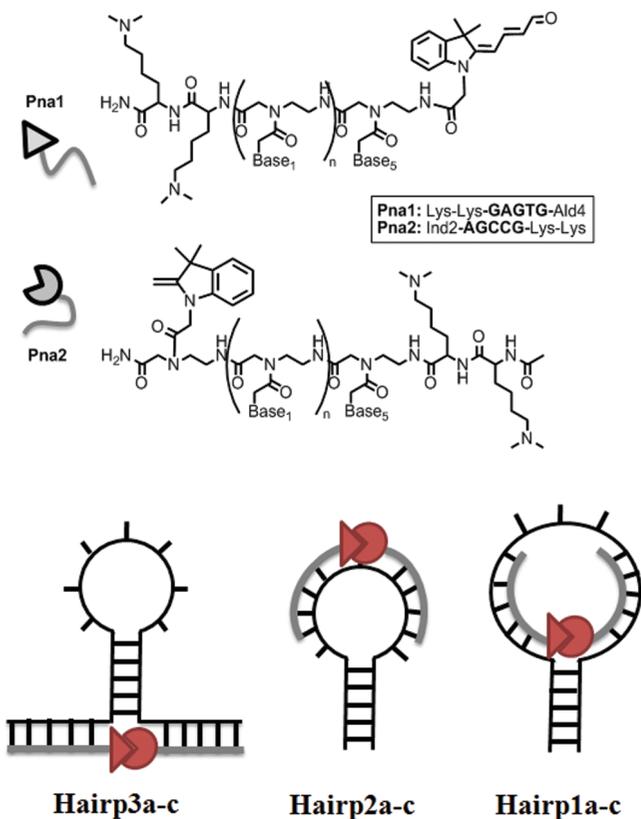


Figure 2. Structures of PNA probes **Pna1** and **Pna2** (top). Schematic representation of the hairpin DNA-templated synthesis of the pentamethine cyanine dye upon simultaneous hybridization of both PNA probes to either the hairpin flanking regions (**Hairp3a-c**) or the hairpin loop (**Hairp2a-c** and **Hairp1a-c**) (bottom).

Table 1. Oligonucleotide and PNA Sequences

	DNA ^a and PNA ^b sequences
Pna1	Lys(NMe ₂)-Lys(NMe ₂)-GAGTG-Ald4
Pna2	Ind2-AGCCG-Lys(NMe ₂)-Lys(NMe ₂)
Hairp1a	CCTAGCTCGGCTTTTCTCACGCTAGG
Hairp1b	CCTAGCTTCGGTTTCTCACTGCTAGG
Hairp1c	CCTAGCTTTCGGCTCTCACTGCTAGG
Hairp1d	CCTAGCTCGGCTTTTCTCAC
Hairp2a	CCTAGCTTTCCTCACTTTTCGGCTTTGCTAGG
Hairp2b	CCTAGCTTTCCTCACTTTTCGGCTTTGCTAGG
Hairp2c	CCTAGCTTTCCTCACTTTTCGGCTTTGCTAGG
Hairp3a	CTCACTTTCGATACTCATCGCATTCGGC
Hairp3b	CTCACTTTCGATACTCATCGCATTCGGC
Hairp3c	CTCACTTTCGATACTCATCGCATTCGGC
Quad1	GCATCCGGGGCGGGCGAGGGAGGGTTCGGC

^a DNA sequences are given from the 5'- to 3'-end. ^b PNA sequences are given from the C-terminal to N-terminal end.

2 and **4** leads to the formation of a symmetrical pentamethine cyanine dye (C5) characterized by its unique fluorescent properties ($\lambda_{exc} = 625$ nm, $\lambda_{em} = 660$ nm), very different from those of its two precursors.

Although highly unfavorable in water, this aldolization–elimination reaction can be templated when carried out in a favorable chemical environment. **Pna1** and **Pna2** (see Figure 2 and Table 1 for structures and sequences) were synthesized on rink amide MBHA resin (Merck Biosciences, loading 0.56 mmol/g) using standard solid-phase Fmoc chemistry. They were subsequently purified by reversed-phase HPLC and characterized by MALDI.

5-mers PNA was chosen as the minimal length necessary for a full and sequence-specific PNA–DNA hybridization under near-physiological conditions. On the basis of the abundant literature on PNAs (37), it is anticipated that our short PNA/

DNA duplexes will have melting temperatures above 40 °C, indicating that almost all the PNA probes will be hybridized to the DNA target in the conditions of our experiments. It is also noteworthy that the presence of two Lysine residues is likely to increase even further the stability of the heteroduplex formed. In the case of unusually unstable (AT-rich) sequences, however, stability of the PNA–DNA heteroduplex would need to be increased. For instance, this could be achieved by increasing the length of the PNA strands, without interfering with the reaction of cyanine dye formation which occurs at the other end of the PNA.

Hairpin DNA Sensing. Reaction between stoichiometric amounts of **Pna1** and **Pna2** (500 nM each) was investigated in a 10 mM potassium phosphate solution (pH = 7.4) also containing 100 mM potassium chloride, in the presence and in the absence of various nucleic acid templates. In each case, reaction efficiency was determined by monitoring the formation of the pentamethine cyanine (product of the reaction between indolenine **2** and aldehyde **4**) by fluorescence spectroscopy over a ten hour period. In the absence of any template, and at a 500 nM concentration of both PNAs, only traces of the fluorescent cyanine dye product were ever detectable. We then investigated the possibility to use our PNA probes for sensing hairpin DNA structures. Three different strategies were developed that consisted of targeting either the hairpin single-stranded flanking arms or the hairpin loop (Figure 2). In the former approach, a hairpin structure was designed which contained two single-stranded flanking arms at its 3'- and 5'-ends, the sequences of which were complementary to those of **Pna1** and **Pna2**. This system (sequences **Hairp3a-c**, table1) was designed so that (i) **Pna1** and **Pna2** could hybridize in a sequence-specific manner to these two single-stranded arms and (ii) the reactive functionalities of both PNAs would be brought in close enough proximity to react with each other only when the hairpin was formed, while remaining separated in the absence of any DNA folding.

In the latter approach, a hairpin DNA structure was chosen that also contained a large central loop, the sequence of which included both **Pna1** and **Pna2** complementary sequences (Figure 2). Once again, only upon simultaneous hybridization of both PNA probes to their cDNA sequences located in the loop region, will the aldehyde be able to react with the 2-methylene-indolenine. Two systems were designed in which the cyanine dye was expected to form either at the center of the loop (sequences **Hairp2a-c**, Table 1) or at the intercept between the loop and the stem (sequences **Hairp1a-d**, Table 1). It is noteworthy that (i) all DNA sequences **Hairp1–3** were shown to form stable hairpin structures under the conditions of our experiment and in the absence of PNA probes, and (ii) DNA hairpins **Hairp1** and **Hairp2** were designed which contained larger loops than **Hairp3** in order to allow complete hybridization of both 5-mer PNAs to the hairpin loop.

The potential of **Hairp3a-c**, **Hairp2a-c**, and **Hairp1a-c** to serve as a DNA template for the reaction of cyanine dye formation was determined. Briefly, a stoichiometric mixture of **Pna1**, **Pna2**, and one hairpin DNA was reacted in potassium phosphate buffer at physiological pH, and the appearance of the cyanine dye product was monitored by fluorescence spectroscopy ($\lambda_{\text{exc}} = 625 \text{ nm}$). As a control experiment, a reaction was carried out under the same conditions but lacking the DNA template. Interestingly, every hairpin DNA tested proved capable of significantly increasing the reaction efficiency, although at different levels depending on whether the PNA probes were targeting the loop or the stem of the hairpin (Figure 3). Moreover, as previously observed by us with other fluorogenic systems ((27) and unpublished data) and although no detectable fluorescence is observed in the absence of template ($F_0 < 350$),

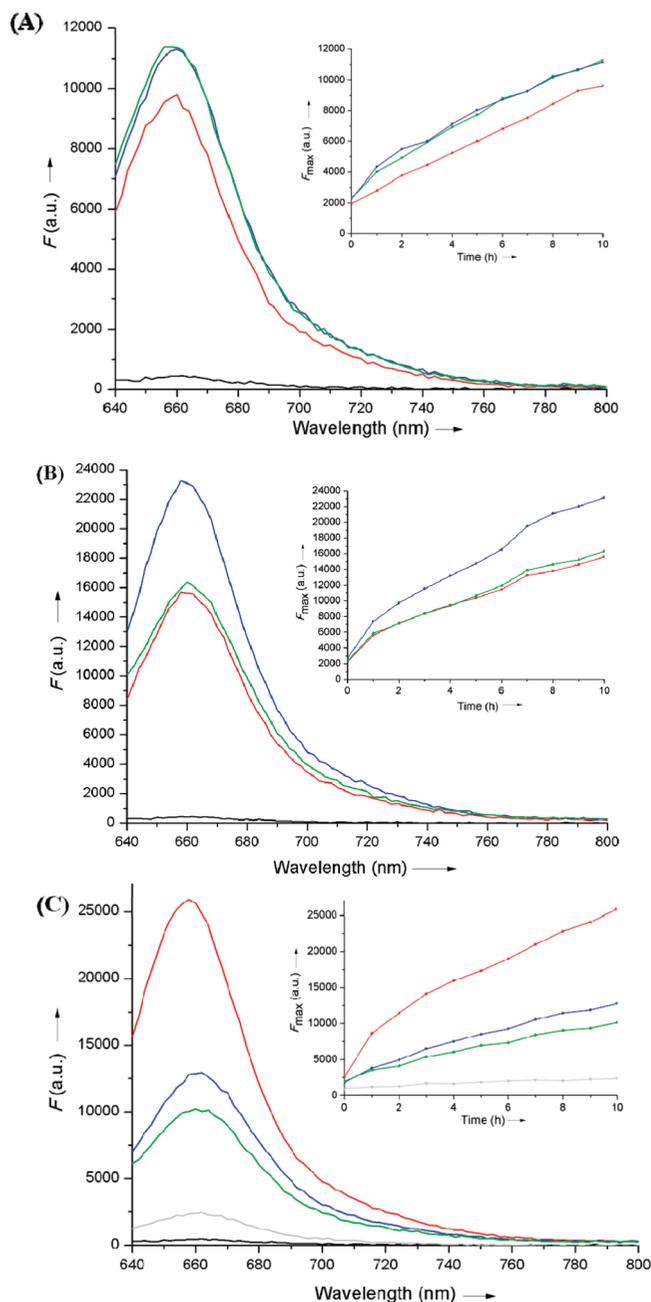


Figure 3. Fluorescence emission spectra of a stoichiometric mixture of **Pna1** and **Pna2** (500 nM each) measured after 10 h ($\lambda_{\text{exc}} = 625 \text{ nm}$) in the absence (black) or in the presence of various DNA templates: (A) **Hairp3a** (red), **Hairp3b** (blue), and **Hairp3c** (green); (B) **Hairp2a** (red), **Hairp2b** (blue), **Hairp2c** (green); (C) **Hairp1a** (red), **Hairp1b** (blue), **Hairp1c** (green), **Hairp1d** (gray). Reaction time courses (maximum fluorescence intensity vs time) are also given as insets for each DNA templated reaction.

DNA-catalyzed reactions are all characterized by a fast and almost instantaneous appearance of fluorescence ($F_0 > 6000$) followed by a slower increase over 10–12 h. The kinetics of cyanine dye formation proved generally slow, the fluorescence intensity still rising even after 10 h, even in the presence of a template (Figure 3). However, the facts that (i) only residual amounts of fluorescent product form in the absence of DNA template (i.e., very low background signal) and (ii) a very fast initial phase of cyanine dye formation is observed in the presence of template make this chemistry suitable for the detection of DNA secondary structures. Indeed, a signal-to-noise ratio of ≥ 17 was obtained for all hairpin DNA-templated reactions after less than 10 min reaction.

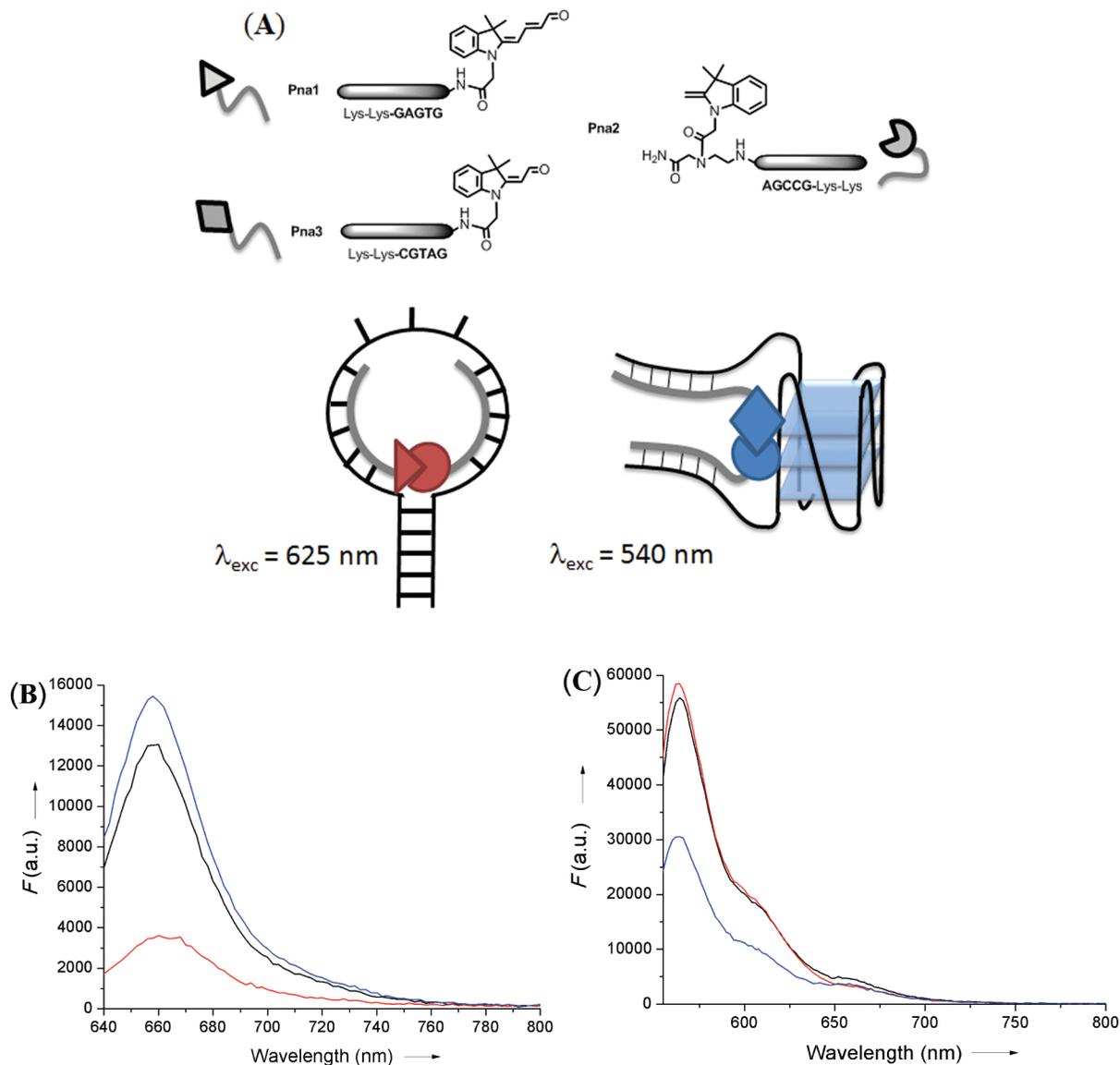


Figure 4. Schematic representation of the general system for sensing simultaneously hairpin and quadruplex DNA, based on a dual-color detection system (A). Fluorescence emission spectra ($\lambda_{\text{exc}} = 540 \text{ nm}$) of a mixture of **Pna1** (250 nM), **Pna2** (1 μM), and **Pna3** (1 μM) in potassium phosphate buffer (10 mM, pH = 7.4 also containing 100 mM KCl) in the presence of 500 nM **Hairp1a** (blue), 1 μM **Quad1** (red), or a mixture of 500 nM **Hairp1a** and 1 μM **Quad1** (black). Excitation wavelengths were either $\lambda_{\text{exc}} = 625 \text{ nm}$ (B) or $\lambda_{\text{exc}} = 540 \text{ nm}$ (C). Emission spectra were recorded after 12 h incubation at room temperature.

The least effective templating effect was obtained when targeting the 3' and 5' single-stranded flanking ends of the hairpin (**Hairp3a–c**). This effect also proved almost independent of the number of nucleotides located between the PNA complementary sequences and the first nucleotide involved in the stem formation. Indeed, comparable fluorescence intensities were obtained from **Hairp3a**, **Hairp3b**, and **Hairp3c**, which contain a one-, two-, and three-nucleotide gap between the PNA/DNA duplex regions and the hairpin stem, respectively (Figure 3a).

Sensing hairpin formation proved, however, more efficient when both PNAs were directed against the hairpin loop, the amount of cyanine dye formed after 10 h of reaction being, for instance, twice as high with **Hairp1a** as with **Hairp3a**. Two slightly different approaches were considered for targeting hairpin loops, the reaction of cyanine dye formation taking place either at the center of the loop or at its base, in the vicinity of the stem. Both approaches proved equally effective (same fluorescence intensity at the maximum emission wavelength obtained after 10 h with **Hairp1a** and **Hairp2b**) although being highly dependent on the overall design. While a four-nucleotide

gap between the functional ends of **Pna1** and **Pna2** led to the optimal reaction efficiency within the **Hairp2** series (**Hairp2b**, Figure 3b), hybridization of both PNAs starting from the first nucleotides constituting of the loop was found twice as effective as systems where **Pna1** and **Pna2** were separated further (**Hairp1a** versus **Hairp1b,c**, Figure 3c).

While **Hairp1a** and **Hairp2b** proved to be equally good templates for the aldolization–elimination reaction, they are based on two different strategies that can find distinct applications in modern bioresearch, in particular, for sensing hairpin DNA formation in vitro and maybe also in vivo. In order to demonstrate the strong link between hairpin formation and pentamethine cyanine dye formation, an analogue of **Hairp1a** was designed which was missing six nucleotides at its 3'-end in order to prevent hairpin formation. The ability of this new single-stranded DNA (so-called **Hairp1d**) to act as a template was investigated by fluorescence spectroscopy (gray curve, Figure 3c). Not surprisingly, the fluorescence intensity obtained with **Hairp1d** was ca. 10 times lower than that obtained with **Hairp1a**, thus demonstrating that the templating effect observed with the latter was indeed structure-specific.

Whether the hairpin remains “hairpin-like” or opens up upon hybridization of both PNAs to the DNA template remains to be demonstrated. **Hairp2** is likely to open after PNA ligation, as is commonly observed with MBs (thus forming a highly stable 10-mer PNA/DNA heteroduplex with single-stranded DNA flanking arms). In the case of **Hairp1** and **Hairp3**, however, the complex formed between the ligation product and the DNA template can be defined as a PNA/DNA hybrid three-way junction and could therefore be significantly more stable than the “open” structure. Additional structural studies will be required to further address this question.

Dual Hairpin + Quadruplex DNA Sensing Using a Dual-Color System. We recently demonstrated that two fluorogenic PNA probes (**Pna2** and **Pna3**, Figure 4a) could be used for sensing the formation of G-quadruplex DNA in vitro. The system was designed so that simultaneous and sequence-specific hybridization of both PNAs to a unique G-quadruplex-containing oligonucleotide were the only conditions inducing the formation of a fluorescent trimethine cyanine dye ($\lambda_{em} = 563$ nm) (27). Since the fluorescent properties of this trimethine cyanine dye were significantly different compared to those of **Pna1** aldehyde ($\lambda_{em} = 476$ nm) and to those of the pentamethine analogue formed when reacting **Pna1** with **Pna2** ($\lambda_{em} = 660$ nm), we reasoned that it should be possible to simultaneously detect the formation of both hairpin and quadruplex DNAs using a non-overlapping, two-color “Cy3-Cy5” system. The system was designed so that a unique nucleophilic PNA, **Pna2**, could potentially react with two electrophilic PNA aldehydes, **Pna1** and **Pna3**, to form a pentamethine ($\lambda_{em} = 660$ nm) and a trimethine ($\lambda_{em} = 563$ nm) cyanine dye, respectively. While **Pna1** and **Pna2** are directed against hairpin DNA **Hairp1a**, **Pna2** and **Pna3** are directed against quadruplex DNA **Quad1**. Hence, it was our expectation to be able to specifically correlate quadruplex and hairpin DNA formation with the appearance of characteristic fluorescence signals at $\lambda_{em} = 563$ nm and $\lambda_{em} = 660$ nm, respectively. In order to demonstrate the concept, a mixture of **Pna1** (250 nM), **Pna2** (1 μ M), and **Pna3** (1 μ M) was incubated under near-physiological conditions and in the presence of various DNA templates: (i) **Hairp1a** (500 nM), (ii) **Quad1** (1 μ M), and (iii) a mixture of **Hairp1a** and **Quad1** (500 nM and 1 μ M, respectively). In each case, the formation of the trimethine and pentamethine cyanine dyes were monitored by fluorescence spectroscopy ($\lambda_{exc} = 540$ and 625 nm, respectively) (Figure 4a). Under our experimental conditions, detection of hairpin DNA proved highly specific, even in the presence of an alternative DNA secondary structure. Indeed, comparable fluorescence intensities were obtained when using as a template either **Hairp1a** alone or a mixture of **Hairp1a** and **Quad1**, while only residual fluorescence was detectable when using the quadruplex **Quad1** alone (Figure 4b).

Detection of quadruplex DNA in the presence of hairpin DNA also proved possible, the amount of trimethine cyanine dye formed decreasing by only <5% upon addition of competitor hairpin DNA (Figure 4c). However, significant background fluorescence (ca. 50%) was also obtained under our conditions, which could be interpreted as the result of a nonspecific formation of trimethine cyanine dye templated by hairpin DNA. So far, attempts to combine on a single oligonucleotide the quadruplex and hairpin forming sequences (both structures being separated by a polyT linker) in order to simultaneously detect both structures were unsuccessful due to high background signal, which is likely due to the formation of complex tertiary structures resulting in nonspecific reaction of dye (C3 and C5) formation.

CONCLUSION

Detection of nucleic acid secondary structures remains a key challenge for chemists due to their large abundance and

structural diversity within the genome. Of particular interest are structures located in regions essential for DNA replication or transcription (e.g., in gene promoters) which could serve as therapeutic targets for regulating gene expression. Whether a DNA strand can potentially form into a hairpin or a quadruplex conformation based on the DNA primary sequence can now be easily predicted, and with reasonably good accuracy. However, whether these structures actually form (or not) depending on the context (in vitro or in vivo) or the experimental conditions remains much more challenging. Therefore, fluorescent probes capable of detecting the folding of specific nucleic acid sequences would represent an important way to validate the existence of such structures in vivo and also maybe to identify new targets. Herein, we reported three different strategies for sensing hairpin DNA formation by targeting either the loop or the stem with short PNAs functionalized at their 3'- or 5'-end with two precursors of a pentamethine cyanine dye. When working at low PNA concentrations (500 nM each), the aldolization-elimination reaction is highly unfavorable, thus resulting in no significant appearance of fluorescence. However, a significant increase (>50-fold) in fluorescence intensity can be obtained upon sequence-specific hybridization of both PNAs to either the loop or the stem of the hairpin target. Therefore, this PNA-based approach represents an original strategy for sensing hairpin DNA formation in vitro, the best sensitivity being obtained when targeting the hairpin loop with the cyanine dye forming in the vicinity of the hairpin stem (**Hairp1a**). Compared to other DNA-templated reactions described in the literature that also involve the formation of a fluorescent product, our system offers the significant advantages of (i) absorbing and emitting in the far-red part of the spectrum ($\lambda_{em} = 660$ nm), thus making it suitable for in vivo applications, and (ii) requiring no catalysis, thus making it biocompatible. It is also noteworthy that this unprecedented DNA-templated C5 synthesis can be orthogonal to the DNA-templated C3 synthesis recently reported by us (27), thus allowing simultaneous detection of at least two different structures within a two-color (C3-C5) system. Although such approach still requires optimization, we demonstrated that it was possible to use a combination of three PNA probes for sensing, within a mixture of hairpin and quadruplex DNA, either or both secondary structures. The development of PNA probes of increased specificity (e.g., longer PNAs) and higher sensitivity for in vivo applications is currently underway in our group.

ACKNOWLEDGMENT

The authors are very thankful to the Centre National de la Recherche Scientifique (CNRS) and the International Centre for Frontier Research in Chemistry (icFRC) for financial support.

LITERATURE CITED

- (1) Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- (2) Wadkins, R. M. (2000) Targeting DNA secondary structures. *Curr. Med. Chem.* 7, 1–15.
- (3) G.N. Parkinson *Quadruplex Nucleic Acids* (2006) Neidle, S., and Balasubramanian, S., Eds; pp 1–30, RSC Biomolecular Sciences Publishing, Cambridge, UK.
- (4) Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35, 406–413.
- (5) Qin, Y., and Hurley, L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* 90, 1149–1171.

- (6) Eddy, J., and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 36, 1321–1333.
- (7) Verma, A., Yadav, V. K., Basundra, R., Kumar, A., and Chowdhury, S. (2009) Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.* 37, 4194–4204.
- (8) Wong, H. M., and Huppert, J. L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* 5, 1713–1719.
- (9) Panayotatos, N., and Fontaine, A. (1987) A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. *J. Biol. Chem.* 262, 11364–11368.
- (10) Varani, G. (1995) Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* 24, 379–404.
- (11) Baker, E. S., Dupuis, N. F., and Bowers, M. T. (2009) DNA hairpin, pseudoknot, and cruciform stability in a solvent-free environment. *J. Phys. Chem. B* 113, 1722–1727.
- (12) Dai, X., Greizerstein, M. B., Nadas-Chinni, K., and Rothman-Denes, L. B. (1997) Supercoil-induced extrusion of a regulatory DNA hairpin. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2174–2179.
- (13) Willwand, K., Mumtsidu, E., Kuntz-Simon, G., and Rommelaere, J. (1998) Initiation of DNA replication at palindromic telomeres is mediated by a duplex-to-hairpin transition induced by the minute virus of mice nonstructural protein NS1. *J. Biol. Chem.* 273, 1165–1174.
- (14) Dervan, P. B., Poulin-Kerstien, A. T., Fechter, E. J., and Edelson, B. S. (2005) Regulation of gene expression by synthetic DNA-binding ligands. DNA binders and related subjects. *Top. Curr. Chem.* 253, 1–31.
- (15) Caldarelli, S. A., Mehiri, M., Di Giorgio, A., Martin, A., Hantz, O., Zoulim, F., Terreux, R., Condom, R., and Patino, N. (2005) A cyclic PNA-based compound targeting domain IV of HCV IRES RNA inhibits in vitro IRES-dependent translation. *Bioorg. Med. Chem.* 13, 5700–5709.
- (16) Bejugam, M., Sewitz, S., Shirude, P. S., Rodriguez, R., Shahid, R., and Balasubramanian, S. (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. *J. Am. Chem. Soc.* 129, 12926–12927.
- (17) Waller, Z. A., Sewitz, S. A., Hsu, S. T., and Balasubramanian, S. (2009) A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J. Am. Chem. Soc.* 131, 12628–12633.
- (18) Wu, P., Ma, D. L., Leung, C. H., Yan, S. C., Zhu, N., Abagyan, R., and Che, C. M. (2009) Stabilization of G-quadruplex DNA with platinum(II) Schiff base complexes: luminescent probe and down-regulation of c-myc oncogene expression. *Chem.—Eur. J.* 15, 13008–13021.
- (19) Wang, K., Tang, Z., Yang, C. J., Kim, Y., Fang, X., Li, W., Wu, Y., Medley, C. D., Cao, Z., Li, J., Colon, P., Lin, H., and Tan, W. (2009) Molecular engineering of DNA: molecular beacons. *Angew. Chem., Int. Ed.* 48, 856–870.
- (20) Ranasinghe, R. T., and Brown, T. (2005) Fluorescence based strategies for genetic analysis. *Chem. Commun.* 5487–5502.
- (21) Cai, J., Li, X., Yue, X., and Taylor, J. S. (2004) Nucleic acid-triggered fluorescent probe activation by the Staudinger reaction. *J. Am. Chem. Soc.* 126, 16324–16325.
- (22) Pianowski, Z. L., and Winssinger, N. (2007) Fluorescence-based detection of single nucleotide permutation in DNA via catalytically templated reaction. *Chem. Commun.* 3820–3822.
- (23) Franzini, R. M., and Kool, E. T. (2008) 7-Azidomethoxycoumarins as profluorophores for templated nucleic acid detection. *ChemBioChem* 9, 2981–2988.
- (24) Z. Pianowski, Z., Gorska, K., Oswald, L., Merten, C. A., and Winssinger, N. (2009) Imaging of mRNA in live cells using nucleic acid-templated reduction of azidorhodamine probes. *J. Am. Chem. Soc.* 131, 6492–6497.
- (25) Furukawa, K., Abe, H., Hibino, K., Sako, Y., Tsuneda, S., and Ito, Y. (2009) Reduction-triggered fluorescent amplification probe for the detection of endogenous RNAs in living human cells. *Bioconjugate Chem.* 20, 1026–1036.
- (26) Huang, Y., and Coull, J. M. (2008) Diamine catalyzed hemicyanine dye formation from nonfluorescent precursors through DNA programmed chemistry. *J. Am. Chem. Soc.* 130, 3238–3239.
- (27) Meguellati, K., Koripelly, G., and Ladame, S. (2010) DNA-templated synthesis of trimethine cyanine dyes: a versatile fluorogenic reaction for sensing G-quadruplex formation. *Angew. Chem., Int. Ed.* 49, 2738–2742.
- (28) Franzini, R. M., and Kool, E. T. (2008) Organometallic activation of a fluorogen for templated nucleic acid detection. *Org. Lett.* 10, 2935–2938.
- (29) Shibata, A., Abe, H., Ito, M., Kondo, Y., Shimizu, S., Aikawa, K., and Ito, Y. (2009) DNA templated nucleophilic aromatic substitution reactions for fluorogenic sensing of oligonucleotides. *Chem. Commun.* 6586–6588.
- (30) Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254, 1497–1500.
- (31) Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S. M., Driver, D. A., Berg, R. H., Kim, S. K., Norden, B., and Nielsen, P. E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* 365, 566–568.
- (32) Grossmann, T. N., and Seitz, O. (2009) Nucleic acid templated reactions: consequences of probe reactivity and readout strategy for amplified signaling and sequence selectivity. *Chem.—Eur. J.* 15, 6723–6730.
- (33) Hansen, M. E., Bentin, T., and Nielsen, P. E. (2009) High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers. *Nucleic Acids Res.* 37, 4498–4507.
- (34) Wojciechowski, F., and Hudson, R. H. (2009) Peptide nucleic acid containing a meta-substituted phenylpyrrolocytosine exhibits a fluorescence response and increased binding affinity toward RNA. *Org. Lett.* 11, 4878–4881.
- (35) Socher, E., Jarikote, D. V., Knoll, A., Röglin, L., Burmeister, J., and Seitz, O. (2008) FIT probes: peptide nucleic acid probes with a fluorescent base surrogate enable real-time DNA quantification and single nucleotide polymorphism discovery. *Anal. Biochem.* 375, 318–330.
- (36) Oquare, B. Y., and Taylor, J. S. (2008) Synthesis of peptide nucleic acid FRET probes via an orthogonally protected building block for post-synthetic labeling of peptide nucleic acids at the 5-position of uracil. *Bioconjugate Chem.* 19, 2196–2204.
- (37) Sugimoto, N., Satoh, N., Yasuda, K., and Nakano, S. (2001) Stabilization factors affecting duplex formation of peptide nucleic acid with DNA. *Biochemistry* 40, 8444–8451.

BC100335F

CHAPTER TWO

DYNAMIC COVALENT CHEMISTRY (DCC) AND CYANINE DYES

I. Principles of Dynamic Combinatorial Chemistry (DCC)

Darwin was the first to suggest the concept of evolution and selection for the development and optimization of biological systems. Stanley Miller then showed that when submitting a mixture of ammonia, methane, hydrogen and water to a source of energy such as ultraviolet radiation¹, 2 % of primary amino acid could be obtained. It is now well established that Nature has developed highly efficient biological systems through combination of many smaller entities/fragments. Although there exist many different ways of assembling small building blocks to create large and complex biological systems, Nature has been able to evolve and select only the most efficient (and best suited) ones. In Chemistry, there exists one efficient and elegant way to mimic this natural selection, or natural evolution: it is based on the principles of Dynamic Combinatorial Chemistry (DCC).

The concept of DCC first appeared in the mid 1990s. It is a powerful approach whereby a number of molecular elements each with binding potential can be reversibly combined via covalent or non-covalent linkages to generate a dynamic library of products under thermodynamic equilibrium. Once a target molecule has been added, the distribution of products can be shifted to favor products that bind to the target. Thus the approach can be employed to identify products that selectively recognize the target. This approach differs from the “traditional” combinatorial chemistry which consists in the intentional synthesis of a large mixture of compounds followed by the screening of each library member against one target.

Dynamic combinatorial chemistry (DCC) is based on the generation of a dynamic combinatorial library (DCL) of compounds under reversible conditions.² A DCL is generated by combining building blocks which are designed to react with each other through reversible covalent reactions or specific non-covalent interactions. At equilibrium, a mixture of compounds is obtained where each species is represented in proportion of its free energy. In other words, the formation of thermodynamically most stable compounds will be favored at the expense of thermodynamically less stable combinations of building blocks.

In 1999, Jean-Marie Lehn³ decrypted the concept of DCC by linking it to 5 essential requirements:

1. The components must possess features for binding to the target
2. Reversibility is the key. Components should connect via covalent or non-covalent reversible interactions
3. Self-assembled structures of equilibrating constituents must be “locked” to allow analyses
4. Virtual Combinatorial Libraries is not about screening and retrieval: it can be reduced to few components to identify a high-priority interaction
5. It must be possible to qualitatively and quantitatively characterize the content of a DCL using techniques like NMR or MS-HPLC.

DCC has since become a powerful method which relies on a simple principle of “adaptive behavior”. It now represents an interesting alternative to the classic combinatorial approach for drug discovery since it enables the identification, within a dynamic chemical library, of the best ligand (or receptor) without the need to identify and characterize each individual component of this library.

In enzymology, an enzyme **E** binds a substrate **S** forming a complex **ES** forming a complex **EP** releasing the product of the catalysis. The formation of such complexes involves the decrease of the transition state energy allowing the appearance of the product:

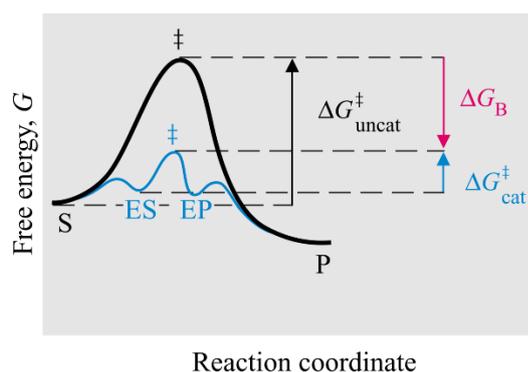


Figure 1: Free transition state

The enzyme recognizes specifically its substrate in a library of compounds in a similar way a lock (host) accepts the key (guest). This specificity arises from a template effect of the substrate(s) in the enzyme active site. In 1997, Brady and Sanders⁴ have introduced the concept of “thermodynamic templating” inspired from enzymatic catalysis. Indeed, they

demonstrated the combinatorial formation of a host in the presence of a template. Stabilization of a particular library member through non-covalent interactions with an added template was shown to alter the position of equilibrium of a DCL, governing the system as described in the bottom figure⁵. Due to stabilizing interactions with the target, the formation of the “best hosts” becomes thermodynamically favored, thus leading to an amplification of their production via a so-called “thermodynamic templating effect”.

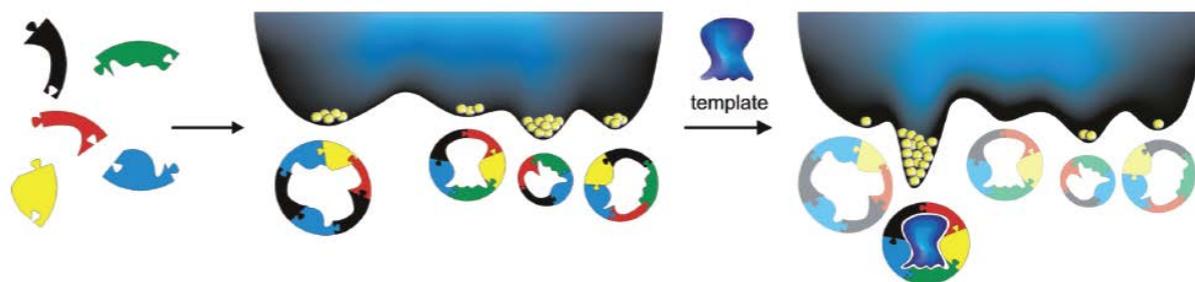


Figure 2: *Combinatorial formation of a host in the presence of a template*

During the same year, Huc and Lehn introduced the concept of “virtual combinatorial libraries” (VCLs)⁶ by using a library of imines (generated via a reversible reaction between aldehydes and amines) for selecting inhibitors of the enzyme carbonic anhydrase (CA). Upon addition of the target (i.e. CA) the formation of certain imines is favored, due to stabilizing interactions with the enzyme. A direct comparison of the libraries obtained in the absence and in the presence of the target allowed the fast identification of the species that had been amplified upon addition of CA (i.e. the fast identification of the best CA binders).

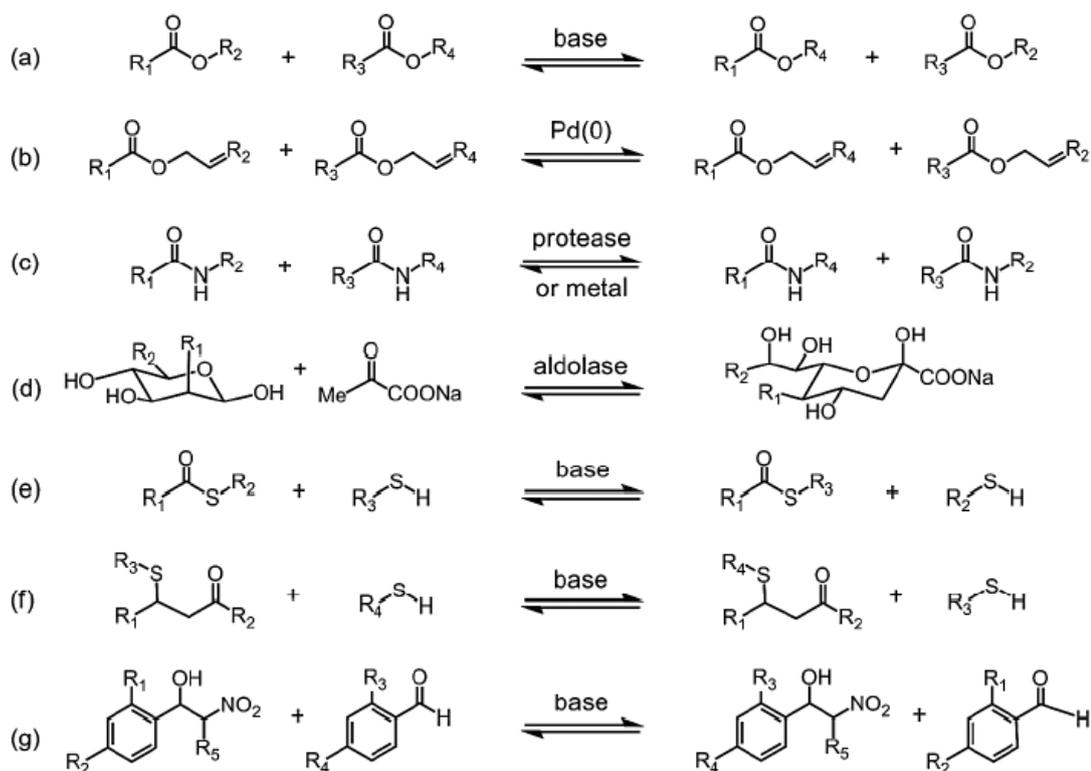
The reversible templating effect mentioned above is only possible due to the existence of covalent/non-covalent reversible interactions between the different library components. Reactions in DCC must be reversible to allow exchanges between library members. The rate of exchange needs to be fast enough so that equilibrium is reached within a convenient interval, but slower than binding to the target. Only for few examples, the reaction needs to be slow to follow the folding - driven oligomerization⁷. The equilibration must occur with the target conditions. For example to observe a biological exchange, it’s necessary to work at the right pH and to adjust it to a biocompatible buffer in aqueous solution and at the suitable temperature.

After reaching the equilibrium, the analysis of the mixture is a critical step and halting (freezing) the reversible reaction is an efficient way to do this. For halting, it's possible to play with few parameters such as the temperature, the pH, turning off the light for the cis-trans isomerisation, removal of one reactive or one catalyst⁸.

II. Reversible reactions suitable for DCC

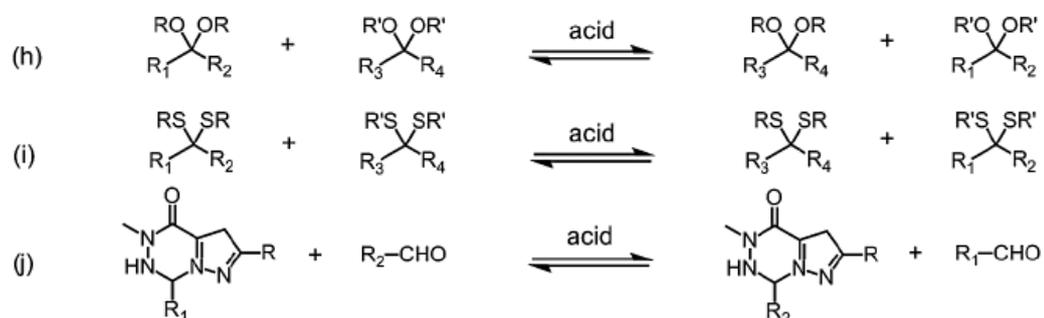
There exist only few reactions which satisfy these criteria and are therefore suitable for DCC. Below is a non-exhaustive list of reversible reactions which proved compatible with DCC. Lehn⁹ and Sanders¹⁰ proposed to classify reversible reactions used in DCLs as shown below:

✓ Acyl transfers and related reactions:



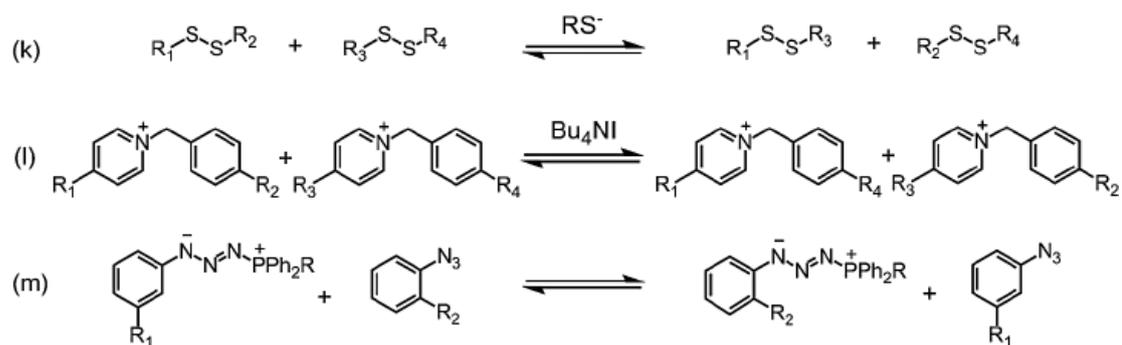
- (a) transesterification, (b) transallylesterification, (c) transamidation, (d) aldol exchange, (e) transthioesterification, (f) Michael/retro - Michael reactions, (g) nitroaldol exchange

✓ Acetal exchanges and related reactions:



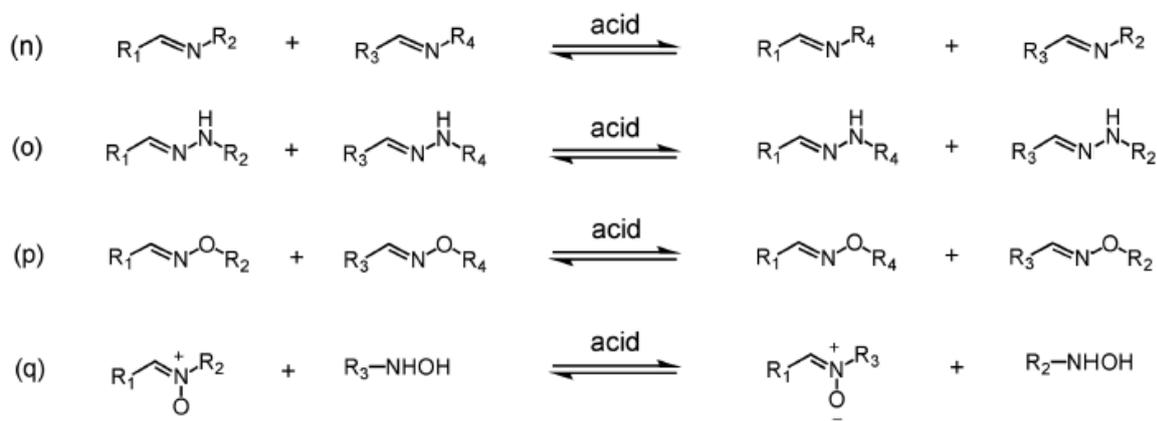
(h) acetal exchange, (i) thioacetal exchange, (j) pyrazolotriazone metathesis

✓ Nucleophilic substitutions:



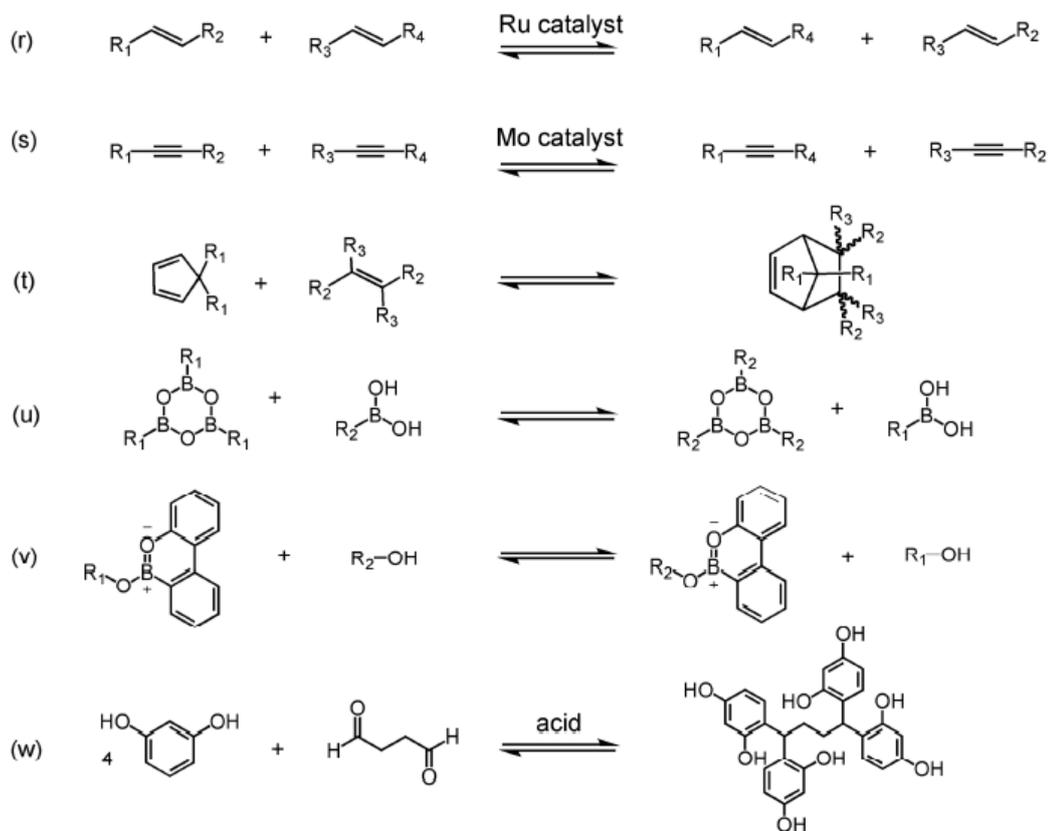
(k) disulfide exchange, (l) reversible benzylic nucleophilic substitution, (m) phosphazide exchange

✓ **Imine metathesis, C=N exchange:**



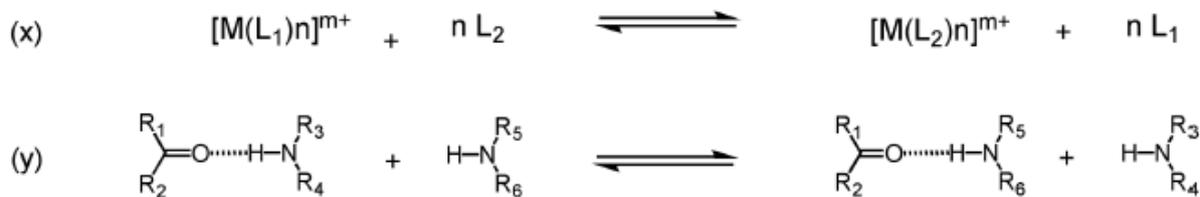
(n) imine exchange, (o) hydrazone exchange, (p) oxime exchange, (q) nitron exchange

✓ **Other reversible covalent bond:**



(r) alkene metathesis, (s) alkyne metathesis, (t) Diels-Alder/retro-Diels-Alder reactions, (u) reversible boroxine formation, (v) transboroxoaromatic esterification, (w) reversible resorcinol and 1,4-butanediol condensation

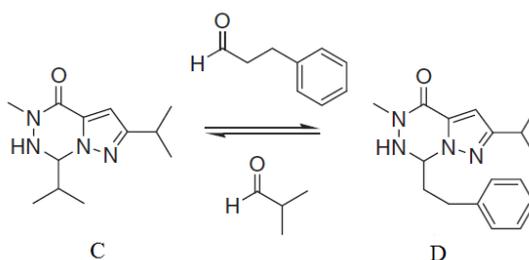
✓ Non-covalent bonds:



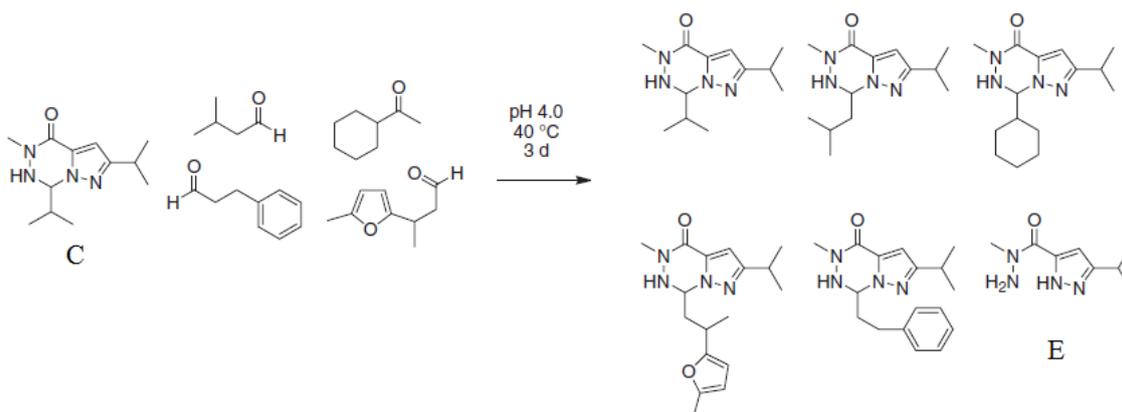
(x) metal –ligand exchange, and (y) hydrogen bond exchange.

✓ Processes related to imine metathesis:

In 2005, Wipf *et al.* proposed a dynamic exchange by using pyrazolotriazinones¹¹. When mixing isobutyraldehyde and hydrocinnamaldehyde in water with a pyrazolotriazinone, an aldehyde exchange occurred as described in the bottom figure:



This concept was next validated using more complex library of aldehyde as presented below.

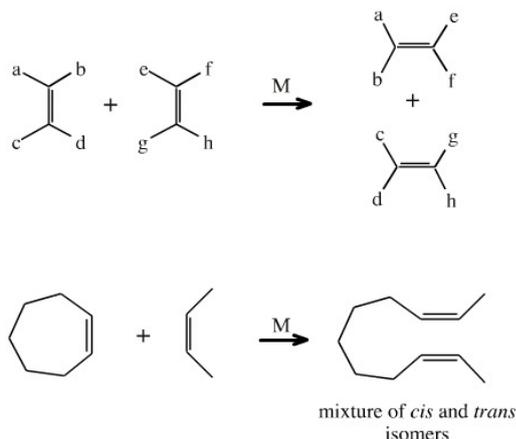


Heating at 40°C in water proved the optimal conditions for hydrolyzing the pyrazolotriazinone, thus forming the highly reactive intermediate E, which can finally react

with the different aldehydes. Direct metathesis of pyrazolotriazinones was also demonstrated. The exchange was halted by increasing the pH at 7.

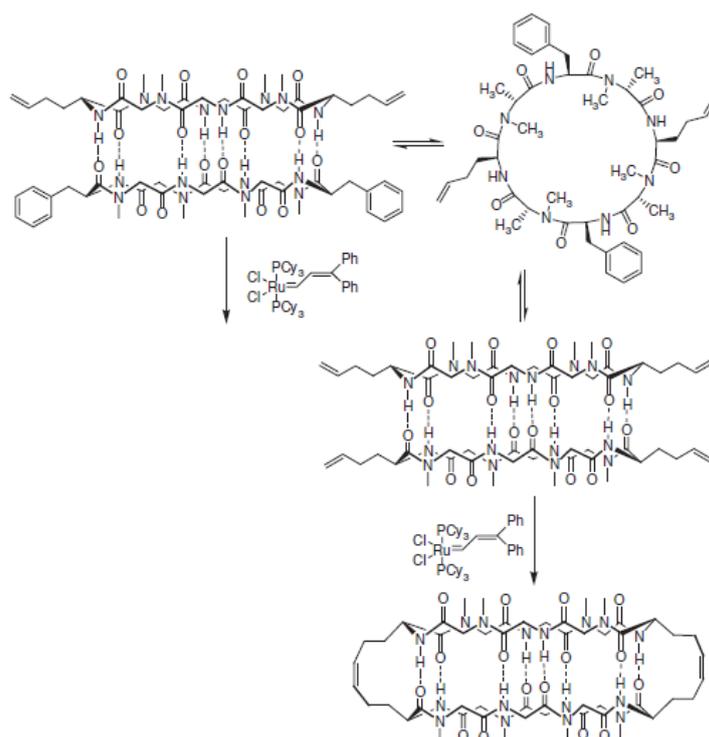
✓ Olefin metathesis:

Olefin Metathesis allows the exchange of substituents between different olefins via a transalkylation:



This reaction has a lot of applications in synthesis for example for ring closure between terminal vinyl groups. This is one of the rare carbon-carbon bond exchange reaction, after the Diels-Alder reaction, which has been successfully used for DCC. In the scheme below, the catalyst M is mainly a derivative of Ruthenium (such as the Grubbs catalyst) or Molybdenum (such as Schrock Catalyst). The Schrock catalysts are more active and are useful for the conversion of sterically demanding substrates, while the Grubbs catalysts tolerate a wide variety of functional groups.

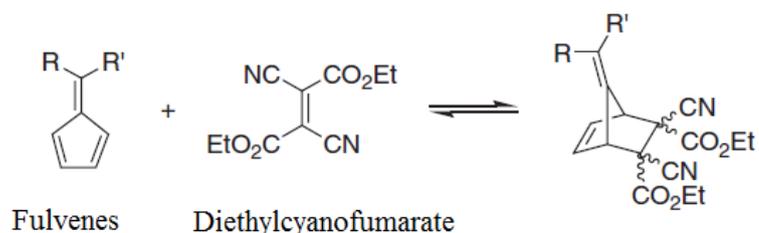
Ghadiri *et al.* developed cyclic peptide cyclo [-(L-Phe-D-(CH₃)NAla-L-Hag-D-(CH₃)NAla)₂-] with Hag= homoallylglycine for capturing an equilibrating mixture of self-assembled structures.¹² This cyclopeptide was designed to allow interconversion between diastereoisomer hydrogen-bonded isomers. The formation of dimeric compound of this tripeptide is possible only if both vinylic chains are in the same plane after a metathesis reaction by using Grubbs' first-generation catalyst.



This is one example of template-assisted cross olefin metathesis¹³ to develop metathesis libraries.

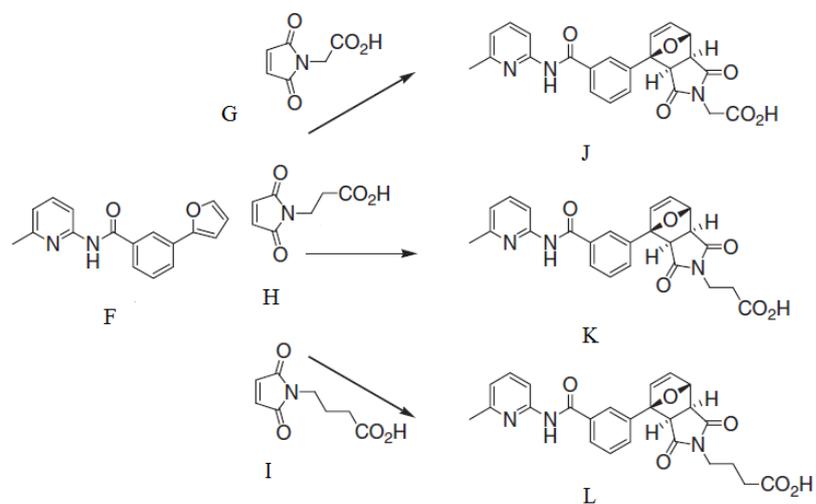
✓ **Diels Alder Reaction:**

Lehn *et al.* have recently proposed that the Diels-Alder reaction was an example of reversible carbon-carbon bond exchange¹⁴. The Diels Alder reaction works at room temperature in one direction while the retro-Diels Alder works at high temperature. Careful tuning of the diene and dienophile, however, can significantly alter this general trend. Reaction between fulvenes (diene) and diethylcyanofumarate (dienophile) was investigated using a library of different fulvenes and the exchange reaction proved reversible:



Bennes and Philip have studied kinetics versus thermodynamics selectivities by studying the Diels-Alder reaction¹⁵. The reaction between derivatives of furane (F) and one

dienophile which contains a one (G), two (H) or three (I) carbon-spacer between the nitrogen of the succinimide and the carboxylic acid was studied. The most thermodynamically stable product formed upon reaction between F and H (two-carbon spacer) due to optimal intramolecular H-bond between the carboxylic acid and the pyridine group or due to ionic consideration. Nevertheless, under conditions of efficient conversion, it is the kinetic product which is observed (J the product with one-carbon spacer). In the case of low conversion rates, it appears that this time the most stable compound (K) is preferentially formed because of the difference in concentration between the dienophiles.



III. PUBLICATIONS

Although it is non-exhaustive, this list of reversible reactions clearly shows that the number of reactions suitable for DCC remains rather limited and new reversible reactions need to be discovered in order to be able to create DCLs of increased chemical and structural diversity. Below, we are reporting new examples of reactions suitable for DCC that could also open the way to new applications.

- (1) Using libraries of benzilidene malononitriles and benzilidine barbiturates, we demonstrate the reversibility of the retro-Knoevenagel and retro-Michele reactions, thus adding two new examples of reactions of Carbon-Carbon bond exchange into the repertoire of reversible reactions suitable for DCC. We demonstrate that these reactions can be carried out in the absence or in the presence of secondary amines (e.g. piperidine, proline) acting as catalysts.
- (2) We then report the first example of reversible cyanine dye (Cy3 and Cy5) analogues. By replacing one carbon atom from the polymethine chain of cyanine dyes by a nitrogen atom, we demonstrate that stable and highly fluorescent analogues of trimethine cyanine dyes can be obtained by reaction between two non-fluorescent entities: a Fisher's base aldehyde and an N-alkyl-2-amino-benzothiazolium. We also demonstrate that this reaction is reversible and is responsive to external stimuli (e.g. temperature). This system could represent the basis for new dynamic fluorosensors based on the principles of DCC.

REFERENCES

- ¹ Johnson AP, Cleaves HJ, Dworkin JP, Glavin DP, Lazcano A, Bada JL *Science* **2008**, 322, 404.
- ² For recent reviews see (a) P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders and S. Otto *Chem. Rev.* **2006**, 106, 3652; (b) S. Ladame *Org. Biomol. Chem.* **2008**, 6, 219.
- ³ J. M. Lehn, *Chem. Eur. J.* **1999**, 5, 9.
- ⁴ P. A. Brady, R. P. Bonar Law, S. J. Rowan, C. J. Suckling and J. K. M. Sanders, *Chem. Commun.* **1996**, 319.

- ⁵ Joost N. H. Reek and Sijbren Otto, *Dynamic Combinatorial Chemistry*, 2010 WILEY-VCH
- ⁶ I. Huc and J. M. Lehn, *Proc. Natl Acad. Sci. USA* **1997**, *94*, 2106.
- ⁷ (a) Oh , K. ; Jeong , K. - S. ; Moore , J. S. *J. Org. Chem.* **2003**, *68* , 8397. (b) Zhao , D. ; Moore , J. S. *J. Am. Chem. Soc.* **2002**, *124*, 9996.
- ⁸ Benjamin L. Miller, “Dynamic Combinatorial Chemistry In Drug Discovery”, Bioorganic Chemistry, and Materials Science, Wiley edition.
- ⁹ O. Ramstrom and J. M. Lehn *Nature Reviews Drug Discovery* **2002**, *1*, 26.
- ¹⁰ Chapter of the following book: Joost N. H. Reek and Sijbren Otto, *Dynamic Combinatorial Chemistry*, 2010 WILEY-VCH edition.
- ¹¹ Wipf , P. ; Mahler , S. G. ; Okumura , K. *Org. Lett.* **2005**, *7*, 4483.
- ¹² Olsen, C.A.; Ghadiri, M.R. *J. Med. Chem.* **2009**, *52*, 7836.
- ¹³ Yang , X. W. ; Gong , B. *Angew. Chem. Int. Ed.* **2005**, *44*, 1352.
- ¹⁴ Boul , P. J. ; Reutenauer , P. ; Lehn , J.M. *Org. Lett.* **2005**, *7*, 15.
- ¹⁵ Bennes , R. M. ; Philp , D. *Org. Lett.* **2006**, *8*, 3651.

Reversible Synthesis and Characterization of Dynamic Imino Analogues of Trimethine and Pentamethine Cyanine Dyes

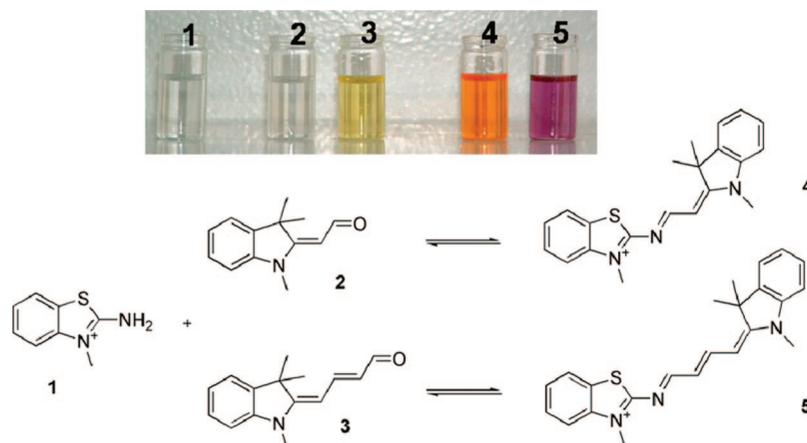
Kamel Meguellati, Martin Spichty, and Sylvain Ladame*

ISIS Université Louis Pasteur 8, Allée Gaspard Monge,
BP 70028, 67083 Strasbourg, France

s.ladame@isis.u-strasbg.fr

Received December 19, 2008

ABSTRACT



A new family of unsymmetrical imine-based trimethine and pentamethine cyanine dye analogues is reported that can form under reversible and thermodynamically controlled conditions from non- or weakly emissive amine and aldehyde building blocks. These dynamic fluorophores show spectroscopic properties comparable to those of their parent cyanine dyes and are responsive to external effectors.

Cyanine dyes generically consist of a conjugated system based on a polymethine chain linking two nitrogen-containing heterocycles (e.g., indoles, benzothiazoles).¹ They are generally named based on the number of carbon atoms in the polymethine chain. Trimethine and pentamethine dyes exhibit absorption maxima at 550 and 650 nm, respectively, and emission maxima around 570 and 670 nm, respectively, in the green or red part of the spectrum. Due to their relative chemical stability, high molar absorption coefficient, and high fluorescence quantum yield, cyanine dyes have been exten-

sively used in photography,² optical data storage,³ and more recently, proteomics⁴ and biomolecular labeling.⁵ Of particular interest are molecules that can reversibly switch between a nonemissive state and an emissive state. Most representative examples are fluorogenic unsymmetrical cyanine dyes which become fluorescent upon interaction with

(1) Hamer, F. M. *The Cyanine Dyes and Related Compounds*; Weissberger, A., Ed.; Interscience: London, U.K., 1964.

(2) James, T. H. *Adv. Photochem.* **1986**, *13*, 329.

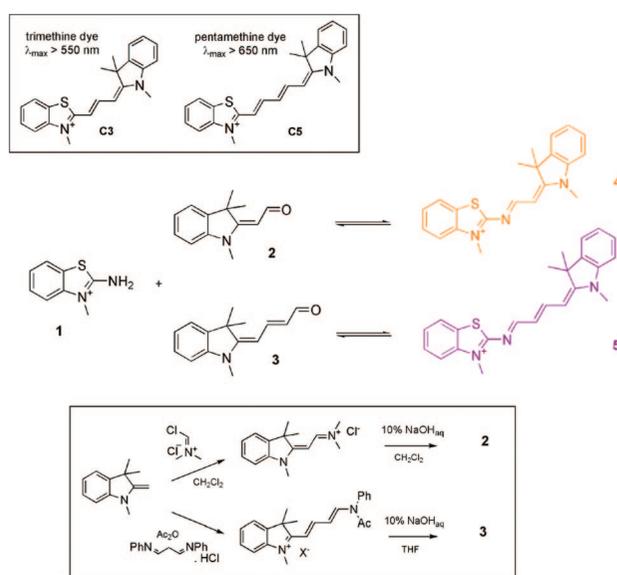
(3) (a) Heilemann, M.; Margeat, E.; Kasper, R.; Sauer, M.; Tinnefeld, P. M. *J. Am. Chem. Soc.* **2005**, *127*, 3801. (b) Spittler, T.; Ehret, A.; Kietzmann, R.; Willig, F. *J. Phys. Chem. B* **1997**, *101*, 2552.

(4) (a) Anaya, C.; Church, N.; Lewis, J. P. *Proteomics* **2007**, *7*, 215. (b) Karp, N. A.; Lilley, K. S. *Proteomics* **2005**, *5*, 3105.

(5) (a) Fegan, A.; Shirude, P. S.; Balasubramanian, S. *Chem. Commun.* **2008**, *17*, 2004. (b) Touthkine, A.; Nguyen, D. V.; Hahn, K. M. *Bioconjugate Chem.* **2007**, *18*, 1344.

specific nucleic acids⁶ or proteins⁷ and photochromic compounds⁸ which can undergo a reversible change of their optical properties under illumination. Such dynamic systems can find valuable applications as fluorescent supramolecular devices, in particular for the design of smart materials.⁹ Herein, we introduce a new family of dynamic fluorophore analogues of the well-known cyanine dyes (from the trimethine and pentamethine family) that can be obtained via a reversible and covalent reaction between non- and/or weakly fluorescent benzothiazolium amine and indolenin aldehyde building blocks (Scheme 1). These cyanine dye analogues

Scheme 1. General Structures of Unsymmetrical Trimethine and Pentamethine Cyanine Dyes (**C3** and **C5**). Proposed Synthetic Strategies for Their Imine Analogues **4** and **5** and Synthetic Route for Aldehydes **2** and **3**



differ from the original dyes solely by the introduction of an imine bond into the polymethine chain, thus making their formation reversible and adaptive to the pressure of external conditions.

As a model system, the reactions between an *N*-methyl-2-amino benzothiazole and two functionalized indolenin aldehydes were investigated. Amine **1** was obtained by quarternization of commercially available 2-aminobenzothia-

zole with methyl iodide. Aldehydes **2** and **3**¹⁰ were synthesized by reaction of 1,2,3,3-tetramethylindolenin with (chloromethylene)dimethyl ammonium chloride and malonaldehyde bisphenylimine hydrochloride, respectively, followed by alkaline hydrolysis of the activated hemicyanine intermediate. In order to monitor the reaction between amine **1** and Fisher's base aldehyde **2**, a 10 mM stoichiometric mixture of both building blocks was heated (50 °C) in $\text{DMSO-}d_6$, and ^1H NMR spectra were recorded every 24 h for 3 days. A new set of signals slowly appeared at the expense of the signals of monomers **1** and **2** along with a pale orange coloration of the solution. Particularly striking was the appearance of two doublets at 8.52 and 6.50 ppm that could be assigned to the imino proton and the adjacent methine proton, respectively, of the predicted **C3** analogue **4** (Figure 1a). The intensity of

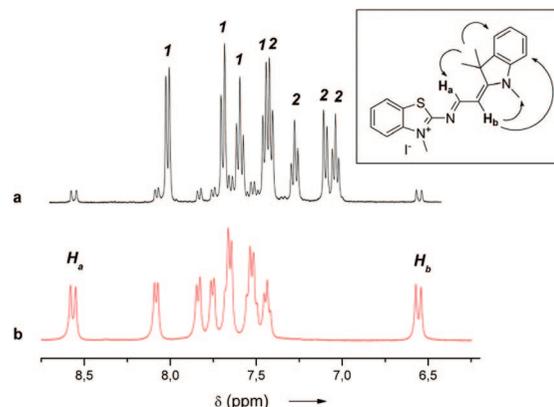


Figure 1. ^1H NMR spectra of (a) a stoichiometric mixture of amine **1** and aldehyde **2** after reacting for 72 h at 50 °C in $\text{DMSO-}d_6$ and (b) isolated **C3** imine analogue **4**. The all-trans structure of **4** is represented in the inset (arrows indicate the observed NOESY correlations).

these signals slowly increased for 3 days until saturation was reached.

In order to unambiguously characterize this new species, a mixture of amine **1** and aldehyde **2** in a mixture of toluene and DMF was heated at 95 °C in a Dean–Stark apparatus for 3 days. The imine derivative **4** was obtained pure as a bright orange solid and was subsequently fully characterized. The ^1H NMR of the isolated imine (Figure 1b) was consistent with the new set of signals appearing when mixing **1** and **2**. This allowed us to estimate around 15% the maximal conversion of aldehyde and amine into the corresponding imine at 50 °C. It is noteworthy that despite the possible cis/trans isomerization of the imine bond formed, only one isomer was always observed by ^1H and ^{13}C NMR. The all-trans conformation of the polymethine/imine chain was assigned by NOESY, thus confirming that this imino dye adopts the same thermodynamically stable conformation as a trimethine cyanine dye.

(6) (a) Constantin, T. P.; Silva, G. L.; Robertson, K. L.; Hamilton, T. P.; Fague, K.; Waggoner, A. S.; Armitage, B. A. *Org. Lett.* **2008**, *10*, 1561. (b) Monchaud, D.; Allain, C.; Bertrand, H.; Smargiasso, N.; Rosu, F.; Gabelica, V.; De Cian, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. *Biochimie* **2008**, *90*, 1207. (c) Socher, E.; Jarikote, D. V.; Knoll, A.; Röglin, L.; Burmeister, J.; Seitz, O. *Anal. Biochem.* **2008**, *375*, 318.

(7) Özhaliç-Ünal, H.; Lee Pow, C.; Marks, S. A.; Jesper, L. D.; Silva, G. L.; Shank, N. I.; Jones, E. W.; Burnette, J. M., III; Berget, P. B.; Armitage, B. A. *J. Am. Chem. Soc.* **2008**, *130*, 12620.

(8) (a) Yumoto, K.; Irie, M.; Matsuda, K. *Org. Lett.* **2008**, *10*, 2051. (b) Tomasulo, M.; Kaanumal, S. L.; Sortino, S.; Raymo, F. M. *J. Org. Chem.* **2007**, *72*, 595. (c) Raymo, F. M.; Tomasulo, M. *J. Phys. Chem. A* **2005**, *109*, 7343.

(9) (a) Mendes, P. M. *Chem. Soc. Rev.* **2008**, *37*, 2512. (b) Yagai, S.; Kitamura, A. *Chem. Soc. Rev.* **2008**, *37*, 1520.

(10) The formation of aldehyde **3** as a **C5** byproduct resulting from the hydrolysis of immobilized hemicyanines was recently proposed based on M.S. analysis in: Mason, S. J.; Hake, J. L.; Nairne, J.; Cummins, W. J.; Balasubramanian, S. *J. Org. Chem.* **2005**, *70*, 2939.

Next, ^1H NMR experiments were carried out from amine **1** and aldehyde **3** under similar conditions as with aldehyde **2**. As described above for imine **4**, the slow formation of imine **5** was evidenced by ^1H NMR. Compound **5** was isolated as a dark purple solid and was also present in solution as a unique all-trans isomer (Supporting Information). It is noteworthy that both imino dyes **4** and **5** proved highly stable in their solid state while slowly regenerating the amine and aldehyde precursors when stored in solution for more than 1 h. Fresh solutions of analytically pure samples of compounds **4** and **5** and of their aldehyde precursors **2** and **3** were prepared in DMSO, and their spectroscopic properties were investigated using UV–vis and fluorescence (Table 1).

Table 1. Spectral Properties of the Imino Dyes **4** and **5** and of Their Aldehyde Precursors **2** and **3**^a

compd	λ_{abs} (nm)	ϵ ($\text{M}^{-1} \text{cm}^{-1}$)	λ_{em} (nm)	QY^b
2	338	5.41×10^4	410	0.00017
3	410	5.47×10^4	460	0.006
4	469	7.42×10^4	522	0.010
5	579	4.83×10^4	608	0.033
C3 ¹¹	547	5.70×10^4	572	0.016
C5 ¹¹	648	8.82×10^4	674	0.038

^a All measurements were taken in DMSO, and reported values are the average of at least two independent measurements. ^b Fluorescence quantum yields were determined by comparative method using quinine sulfate, fluorescein, and sulforhodamine 101 as reference standards.

In comparison to their parent C3 and C5 cyanine dyes (i.e., lacking the imine bond), imines **4** and **5** have their maximum emission of fluorescence at slightly shorter wavelengths (–50 and –68 nm, respectively). However, the introduction of an imine bond into the conjugated polymethine chain has only a very limited effect on either the Stokes' shift or the molar absorption coefficient of the dye. Although notably low (0.01–0.033), fluorescence quantum yields of both imines proved significantly higher than those of their corresponding aldehyde precursors and only slightly lower than those of commonly used trimethine and pentamethine dyes. It is also noteworthy that the absorption spectra of aldehydes **2** and **3** do not overlap with those of their corresponding imines. Therefore, when forming imine **4** (or **5**) from amine **1** and aldehyde **2** (or **3**), excitation of the reaction mixture at 470 nm (or 580 nm) enables the specific detection of the fluorescence emitted by the imino dye only. Of particular interest is the case of bright orange and fluorescent C3 imino analogue **4** which results from the reaction between a dark and colorless amine **1** and a colorless and minimally fluorescent aldehyde **2** (fluorescence quantum yield 60-fold lower than that of imine **4**).

To elucidate the changes in electronic structure upon introduction of an imino group into the polymethine chain of cyanine dyes we performed a time-dependent density functional theory (TD-DFT)¹² study. Calculations were carried out for compounds **4** and **5** as well as for reference compounds C3 and C5 on the B3LYP/6-311G(d)//B3LYP/

6-31G(d) level of theory using an implicit treatment of the solvent (DMSO). The scaling scheme of Champagne et al.¹³ was used to correct for the slight, intrinsic overestimation of excitation energies by the TD-DFT method. The calculated energies for vertical excitation of 1.89, 2.06, 2.23, and 2.42 eV for dyes C5, **5**, C3, and **4** are in good agreement with the experimental values of 1.91, 2.14, 2.26, and 2.64 eV. The experimental shift to larger excitation energies, i.e., shorter wavelengths upon introduction of an imino group into the polymethine chain, is also seen in the calculations: 44 nm for **4** versus C3 and 54 nm for **5** versus C5. While the energy-minimized structures of C3 and C5 and of their respective imino analogues **4** and **5** all proved perfectly planar, the shift in wavelength can be understood in terms of the larger HOMO–LUMO splitting in the ground state for imino compounds **4** and **5** in comparison to the parent C3 and C5 molecules, respectively (Figure 2; note that the

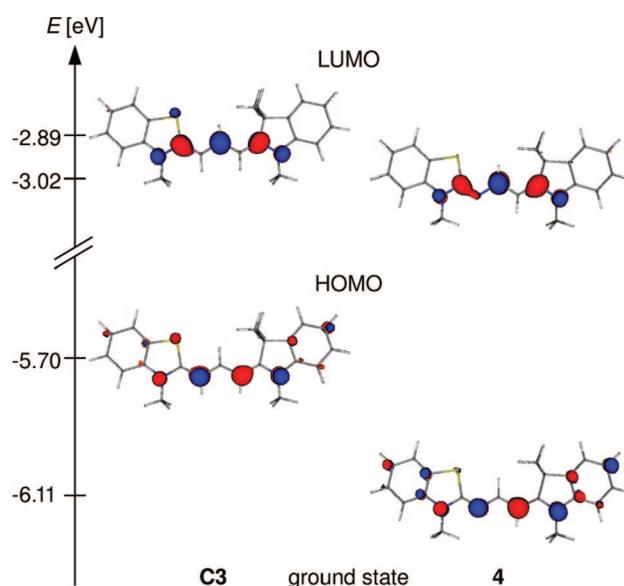


Figure 2. Highest occupied molecular orbital (HOMO) and lowest unoccupied orbital molecular orbital (LUMO) of C3 and **4**. The isodensity surfaces of the orbitals were drawn with the program MOLEKEL 5.3¹⁴ from GAUSSIAN 03¹⁵ cube-files with a contour value of ± 0.06 .

HOMO–LUMO gap in the ground state does not correspond to the reported values of the excitation energy as determined from TD-DFT). Due to the positive charge of the system, both HOMO and LUMO are largely negative in energy. It is not surprising that the orbitals of the imino dye **4** are shifted to lower energy with respect to the all-methine dye, the electrons being more strongly bound in **4** than in C3 due

(11) For direct comparison, analytically pure samples of the C3 and C5 direct analogues of imines **4** and **5** were prepared according to the literature procedure, and their spectroscopic properties in DMSO were determined under identical conditions. See the Supporting Information.

(12) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439.

(13) Champagne, B.; Guillaume, M.; Zutterman, F. *Chem. Phys. Lett.* **2006**, *425*, 105.

to the larger electronegativity of the nitrogen. The HOMO is also more strongly delocalized over the entire molecule in **4** than in C3, i.e., the atomic orbital coefficients of the heterocycles are larger in **4** than in C3. The replacement of a methine group by an imino group leads to a better overlap between the polymethine chain and the nitrogen-containing heterocycles, and thus, the HOMO is more stabilized. On the other hand, the LUMO is slightly less delocalized in **4** than in C3. The orbital coefficients of the sulfur atom are smaller in **4** than in C3, whereas for the imino nitrogen the opposite is true. As a result of the larger delocalization (stabilization) of the HOMO with respect to the LUMO in **4** than in C3 the gap between these orbitals is larger for the former. This explains qualitatively the larger excitation energy and the shorter absorbance wavelength for the imino dyes **4** and **5** in comparison to C3 and C5, respectively.

Finally, we investigated whether the reaction of formation of cyanine dye imino analogues from amine and aldehyde building blocks was truly reversible. The dynamic system could then undergo constitutional reorganization in response to an external stimulus.¹⁶ A stoichiometric mixture of amine **1** and aldehyde **2** (25 mM each) in DMSO was stirred at room temperature until thermodynamic equilibrium was reached (48 h). The perturbation of the system was monitored by HPLC, UV-vis, and fluorescence spectroscopy. At equilibrium, only a very small amount of C3 imine **4** was formed. Despite this low level of conversion, imine **4** was easily detected by fluorescence spectroscopy and could also be visualized by UV-vis (yellow color). The mixture was then heated to 60 °C and was again allowed to reach equilibrium. Equilibrium was reached after 24 h which was different from that observed at room temperature. The absolute amount of C3 analogue **4** present at equilibrium was increased by about 5-fold (as measured by both HPLC and fluorescence) (Figure 3). When the reaction mixture was being cooled to room temperature, the system reorganized

again until it reached an equilibrium position (after 72 h) that was comparable to that obtained initially before heating, thus proving true reversibility of the system. Similar results were obtained when reacting amine **1** (50 mM) with both aldehydes **2** and **3** (25 mM each) simultaneously.

It is also noteworthy that reorganization of the equilibrating mixture can be monitored in real time by fluorescence spectroscopy. While this dynamic system takes 48 h to equilibrate at room temperature, less than 24 h is required to reach equilibrium when the reaction is carried out at 60 °C (see the Supporting Information for kinetics of imino dye formation).

Synthesis of cyanine dyes is generally accomplished by the irreversible stepwise reaction between nucleophilic heterocycles (e.g., 1,2,3,3-tetramethylindolenin) and a polyene-chain precursor such as an amidine and proceeds via the formation of a hemicyanine intermediate. Herein, we reported the first family of cyanine dye (trimethine and pentamethine) imine analogues that are easily accessible via a reversible and thermodynamically controlled reaction from readily available amine and aldehyde building blocks. We provided proof-of-concept that this dynamic system could reorganize in response to an external stimulus, thus leading to a measurable perturbation of the global UV-vis and fluorescence spectra of the equilibrating mixture. The development of smart materials based on this concept is currently underway in our laboratory.

Acknowledgment. We thank Martin Karplus (ISIS, Strasbourg) for fruitful discussions, and we are grateful for the computational resources provided by the high-performance cluster of the Université Louis Pasteur. K.M. thanks the French Ministry of Research and Technology for a doctoral Fellowship. S.L. thanks the CNRS for funding. M.S. is grateful to ISIS for hosting him as a guest assistant professor.

Supporting Information Available: General experimental procedures and spectroscopic data for new compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

OL802913B

(14) MOLEKEL 5.3: Flükiger, P.; Lüthi, H. P.; Portmann, S.; Weber, J. Swiss National Supercomputing Centre CSCS, Manno, Switzerland, 2008.

(15) GAUSSIAN 03, Revision: Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc, Wallingford, CT, 2004.

(16) (a) Giuseppone, N.; Lehn, J.-M. *Chem.—Eur. J.* **2006**, *12*, 1715.
(b) Giuseppone, N.; Fuks, G.; Lehn, J.-M. *Chem.—Eur. J.* **2006**, *12*, 1723.

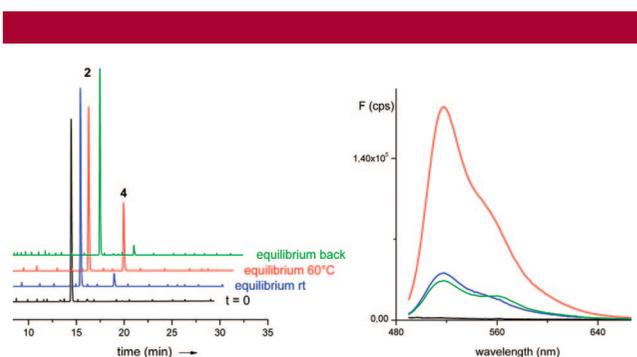


Figure 3. HPLC traces and fluorescence spectra showing the formation of imine **4** from a mixture of amine **1** and aldehyde **2** (25 mM each). Starting position (black) and positions of equilibrium after 48 h at rt (blue), then after heating 24 h at 60 °C (red), and after cooling back for 72 h at rt (green) are represented. HPLC traces show the absolute ratios of aldehyde **2** (retention time 14.5 min) and imine **4** (retention time 18.5 min). Fluorescence emission spectra were recorded when exciting at 480 nm.

Reversible Synthesis and Characterisation of Dynamic Imino Analogues of Trimethine and Pentamethine Cyanine Dyes.

*Kamel Meguellati, Martin Spichy and Sylvain Ladame**

ISIS Université Louis Pasteur, 8 Allée Gaspard Monge, BP 70028, 67083 Strasbourg, France.

Fax: (+33) 390245115; Tel: (+33) 390245210 E-mail: s.ladame@isis.u-strasbg.fr

SUPPORTING INFORMATION

1	SYNTHESIS.....	2
1.1	General Experimental.....	2
1.2	Synthesis of building blocks.....	2
1.3	Synthesis and isolation of Cyanine dye analogues.....	3
1.4	Synthesis of C3 and C5 cyanine dyes.....	5
1.5	Kinetic studies.....	5
2	SPECTROSCOPIC STUDIES.....	6
2.1	General Experimental.....	6
2.2	Fluorescence Quantum Yields (QY).....	6
3	DENSITY FUNCTIONAL THEORY CALCULATIONS.....	6
4	NMR SPECTRA.....	7

1 SYNTHESIS

1.1 General experimental

Solvents were of HPLC or reagent quality and purchased commercially. Starting material were purchased commercially and used without further purification. Compounds were characterized using ^1H and ^{13}C NMR that were recorded on a Bruker Avance DRX 400 spectrometer at 400 and 100.6 MHz, respectively. Chemical shifts are reported as δ values (ppm) with reference to the residual solvent peaks. Products were purified by flash column chromatography on silica gel. Melting points were measured on a Büchi B-540 apparatus. IR spectra were recorded on a Nicolet 6700 FT-IR spectrometer from Thermo scientific.

1.2 Synthesis of Building blocks

2-Amino-1-methylbenzothiazolium iodide (1)

2-Aminobenzothiazole (3.5 g, 23 mmol) was dissolved in acetonitrile (50 mL) then iodomethane (2 mL) was added. The solution was stirred at 45 °C for 12 hours. After cooling to room temperature, Et₂O (50 mL) was added and the white suspension was filtered. The precipitate was washed with diethylether and dried under vacuum. Compound **1** was obtained pure as a white solid (6.75 g, 100%, mp 227-228 °C).

^1H NMR (400 MHz, DMSO) δ 10.0 (s, 2 H), 8.0 (d, $J = 7.9$ Hz, 1 H), 7.69 (d, $J = 7.9$ Hz, 1 H), 7.6 (t, $J = 7.5$ Hz, 1H), 7.44 (t, $J = 7.5$ Hz, 1 H), 3.75 (s, 3 H).

^{13}C NMR (100 MHz, DMSO) δ 167.9, 138.9, 127.7, 125.1, 123.4, 122.2, 122.1, 113.3, 32.2

HRMS: m/z : calcd for C₈H₉N₂S⁺: 165.0481; found 165.0475.

(Z)-2-(1,3,3-Trimethylindolin-2-ylidene)acetaldehyde (2)

(Chloromethylene)dimethylammonium chloride (1.66 g, 13 mmol) was dissolved in CH₂Cl₂ (20 mL) and was stirred at room temperature for 10 minutes. Then a solution of 1,2,3,3-tetramethyl-3*H*-indole (2.3 g, 13 mmol) in CH₂Cl₂ (20 mL) was added dropwise. After 15 minutes under stirring at room temperature, an aqueous solution of 10 % NaOH (40 mL) was added carefully at 0 °C. After 1h, the organic solution was extracted and washed 3 times with water and then dried over Na₂SO₄, filtered and concentrated. The residue was purified by flash chromatography on silica gel (Et₂O, R_f = 0.3 in Et₂O). Compound **2** was obtained pure as a white powder (1 g, 40 %, mp 116-117 °C).

^1H NMR (400 MHz, DMSO) δ 9.9 (d, $J = 8.8$ Hz, 1 H), 7.41 (d, $J = 7.5$ Hz, 1 H), 7.28 (dt, $J = 7.5$, 1 Hz, 1 H), 7.1 (d, $J = 7.5$ Hz, 1 H), 7.02 (dt, $J = 7.5$, 0.6 Hz, 1 H), 5.3 (d, $J = 8.8$ Hz, 1H), 3.24 (s, 3H), 1.6 (s, 6 H)

^{13}C NMR (100 MHz, DMSO) δ 185.3, 172.8, 143.3, 139.1, 127.8, 121.9, 121.8, 108.4, 98.3, 46.8, 29.3, 28.8

IR (cm⁻¹) 1629, 1573, 1490, 1467, 1396, 1184.

HRMS: m/z : calcd for C₁₃H₁₆NO⁺: 202.1226; found 202.1221

(2E,4Z)-4-(1,3,3-Trimethylindolin-2-ylidene)but-2-enal (3)

1,2,3,3-Tetramethyl-3*H*-indolium iodide (1 g, 3.3 mmol), malonaldehydebis(phenylimine) monohydrochloride (859 mg, 3.3 mmol) and sodium acetate (600 mg) were suspended in acetic anhydride (40 mL). After 30 minutes under stirring at room temperature, Et₂O (200 mL) was added then the green dark precipitate was filtered off solution and washed with Et₂O. The solid was subsequently washed with CH₂Cl₂ and the filtrate was collected and concentrated off under vacuum. The crude residue was purified by flash chromatography on

silica gel (CH₂Cl₂/MeOH 98:2, R_f = 0.1) affording the desired activated hemicyanine pure as a brown foam (400 mg, 32 %).

¹H NMR (400 MHz, DMSO) δ 8.88 (d, *J* = 13.1 Hz, 1 H), 8.5 (dd, *J* = 15, 11.3 Hz, 1 H), 7.79 (d, *J* = 7.5 Hz), 7.72 (d, 7.5 Hz, 1 H), 7.44-7.69 (m, 7 H), 6.85 (d, *J* = 15 Hz, 1 H), 5.51 (dd, *J* = 13.1, 11.3 Hz, 1 H), 3.81 (s, 3 H), 2 (s, 3H), 1.7 (s, 6 H).

¹³C NMR (100 MHz, DMSO) δ 181, 170.1, 156.8, 147.7, 143.3, 142.3, 138.3, 130.9, 130, 129.2, 129.1, 128.9, 128.6, 123.2, 119.4, 114.6, 113.3, 112.3, 51.5, 33.6, 23.7.

This freshly prepared activated hemicyanine (400 mg, 1.05 mmol) was dissolved in THF (20 mL) then a solution of 10 % NaOH (60 mL) was added. The brown mixture was stirred at room temperature for 10 h then CH₂Cl₂ was added. The organic solution was extracted and washed with water, dried over Na₂SO₄ and concentrated under vacuum. The crude was purified by flash chromatography on silica gel (CH₂Cl₂/ Et₂O 1:1, R_f = 0.3 in Et₂O). Compound **3** was obtained pure as brown/orange oil (238 mg, 100 %).

¹H NMR (400 MHz, DMSO) δ 9.45 (d, *J* = 8.3 Hz, 1 H), 7.86 (dd, *J* = 13.9, 13 Hz, 1 H), 7.37 (d, *J* = 7.2 Hz, 1 H), 7.24 (t, *J* = 7.2 Hz, 1 H), 7 (d, *J* = 7.2 Hz, 1H), 6.96 (t, *J* = 7.2 Hz, 1H), 5.85 (dd, *J* = 13.9, 8.3 Hz, 1 H), 5.68 (d, *J* = 13 Hz, 1 H), 3.26 (s, 3 H), 1.59 (s, 3 H)

¹³C NMR (100 MHz, DMSO) δ 191.6, 164.9, 150, 143.4, 138.4, 127.4, 122.1, 121.2, 120.7, 107.3, 94.2, 45.8, 28.7, 27.4

IR (cm⁻¹) 1656, 1595, 1537, 1492, 1466, 1392, 1184, 1120.

HRMS: *m/z*: calcd for C₁₅H₁₈NO⁺: 228.1383; found 228.1391

1.3 Synthesis and isolation of cyanine dye analogues

3-Methyl-2-((*E*)-(*Z*)-2-(1,3,3-trimethylindolin-2-ylidene)ethylidene)amino)benzo[d]thiazol-3-ium iodide (**4**)

(*Z*)-2-(1,3,3-Trimethylindolin-2-ylidene)acetaldehyde **2** (845 mg, 4.2 mmol) and 2-amino-1-methylbenzothiazolium iodide **1** (1.23 g, 4.2 mmol) were added to a mixture of toluene (30 mL) and DMF (8 mL). The orange suspension was stirred at 95 °C for 16 h in a flask equipped with a Dean Starck apparatus. After cooling down to room temperature, Et₂O (200 mL) was added and the precipitate was isolated by filtration. The solid was washed with cold CHCl₃ and the filtrate was concentrated under vacuum. The crude residue was finally purified by a short column chromatography on silica gel (AcOEt then AcOEt/MeOH 200:5, AcOEt/MeOH 200:10, R_f = 0.1 in AcOEt/MeOH 8:2) affording the desired compound **4** pure as a bright orange/red solid (101 mg, 5 %, mp 223-224 °C).

¹H NMR (400 MHz, DMSO) δ 8.56 (d, *J* = 12.1 Hz, 1 H), 8.1 (d, *J* = 7.8 Hz, 1 H), 7.83 (d, *J* = 7.4 Hz, 1 H), 7.75 (d, *J* = 7.4 Hz, 1H), 7.65 (m, 2 H), 7.53 (t, *J* = 7.4 Hz, 1 H), 7.51 (t, *J* = 7.8 Hz, 1 H), 7.44 (t, *J* = 7.4 Hz, 1 H), 6.55 (d, *J* = 12.1 Hz), 3.92 (s, 3 H), 3.81 (s, 3 H), 1.73 (s, 6H)

¹³C NMR (100 MHz, DMSO) δ 179.4, 170.8, 161.6, 142.1, 141.4, 139.3, 128.5, 128.1, 126.5, 125.4, 124.3, 123.4, 123.1, 122.4, 113.9, 112.6, 102.6, 102, 50.1, 32.2, 27.6

IR (cm⁻¹) 1619, 1591, 1544, 1490, 1479, 1452, 1392, 1359, 1336, 1265, 1116.

HRMS: *m/z*: calcd for C₂₁H₂₂N₃S⁺: 348.1529; found 348.1512.

3-Methyl-2-((*E*)-((2*E*-4*Z*)-4-(1,3,3-trimethylindolin-2-ylidene)but-2-enylidene)amino)benzo[d]thiazol-3-ium iodide (**5**)

1,3,3-Trimethyl-2-((1*E*,3*E*)-4-(*N*-phenylacetamido)buta-1,3-dienyl)-3*H*-indolium chloride (280 mg, 0.73 mmol) and 2-amino-1-methylbenzothiazolium iodide (213 mg, 0.73 mmol) were dissolved in DMF (30 mL). The solution was stirred at 100 °C for 16 h. The green solution became red after few minutes. After cooling down to room temperature, Et₂O (200

mL) was added and the precipitate formed was washed with Et₂O. CHCl₃ (100 mL) was then added, the filtrate was collected and the solvents were evaporated off. The purple crude residue was purified by column chromatography on silica gel (CH₂Cl₂ then CH₂Cl₂/MeOH 200:5, 200:10, R_f= 0.2 in AcOEt/MeOH 8:2) affording the desired compound **5** pure as a purple solid (12.4 mg, 4 %, mp 225-226 °C)

¹H NMR (400 MHz, DMSO) δ 8.5 (dd, *J* = 14.2, 11.8 Hz, 1 H), 8.33 (d, *J* = 11.8 Hz, 1 H), 7.98 (d, *J* = 7.5 Hz, 1 H), 7.74 (d, *J* = 7.5 Hz, 1 H), 7.68 (d, 8 Hz, 1 H), 7.63 (d, *J* = 8 Hz, 1 H), 7.58 (t, *J* = 8 Hz, 1H), 7.52 (t, *J* = 8 Hz, 1H), 7.43 (t, *J* = 7.5 Hz, 1 H) 7.41 (t, *J* = 7.5 Hz, 1 H), 6.75 (d, *J* = 14.2 Hz, 1 H), 6.6 (t, *J* = 11.8 Hz, 1 H), 3.81 (s, 3H), 3.79 (s, 3H), 1.72 (s, 6 H)

¹³C NMR (100 MHz, DMSO) δ 177.8, 161.8, 142.1, 142, 139.4, 128.5, 128.4, 127.8, 126.7, 124.3, 123, 122.9, 122.5, 121.7, 121.4, 112.5, 107.7, 50.1, 32.2, 27.8

IR (cm⁻¹) 1617, 1595, 1542, 1494, 1478, 1452, 1366, 1358, 1332, 1310, 1264, 1117.

HRMS: *m/z*: calcd for C₂₃H₂₄N₃S⁺: 374.1685; found 374.1681.

¹H NMR experiments were carried out from amine **1** and aldehyde **3** (10mM each in d₆-DMSO). As for the formation of compound **4**, the reaction of **1** with **3** led to the slow formation of the imine analogue **5** of a Cy5 that can also be visualised by a colour change of the NMR solution from bright yellow to dark purple. After 72h, a plateau was reached that corresponded to the formation of c.a. 15% of compound **5**. To unambiguously characterise the generated imine, a sample was isolated by flash-chromatography and analysed by NMR (Figure S1). The signals corresponding to the four protons from the polymethine/imine chain were assigned using a COSY ¹H NMR. As for the Cy3 analogue **4**, the Cy5 analogue **5** is present in solution as a unique isomer. Coupling constants *J* of 11.8 Hz between *H_a* and *H_b* and *H_b* and *H_c* and 14.2 Hz between *H_c* and *H_d* also suggested an all-*trans* conformation for the imine/polymethine chain which was subsequently confirmed by a NOESY experiment.

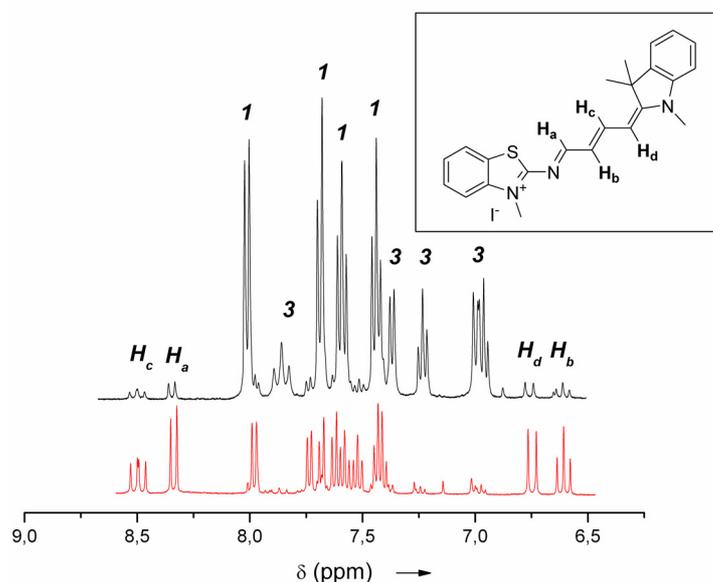


Figure S1. ¹H NMR spectrum of a mixture of amine **1** and aldehyde **3** after reacting for 72h at 50°C in d₆ DMSO (top) and ¹H NMR spectrum of isolated Cy5 imine analogue **5** (bottom). Structure of compound **5** is represented in the inset.

1.4 Synthesis of C3 and C5 cyanine dyes.

For comparison, unsymmetrical trimethine (C3) and pentamethine (C5) cyanine dyes, direct analogues of imines **4** and **5**, were prepared according to literature procedures (Hamer, F.M. *The Cyanine dyes and related compounds* (1964) in Interscience: New York, London). Absorption and emission spectra of analytically pure samples are reported in Figure S2.

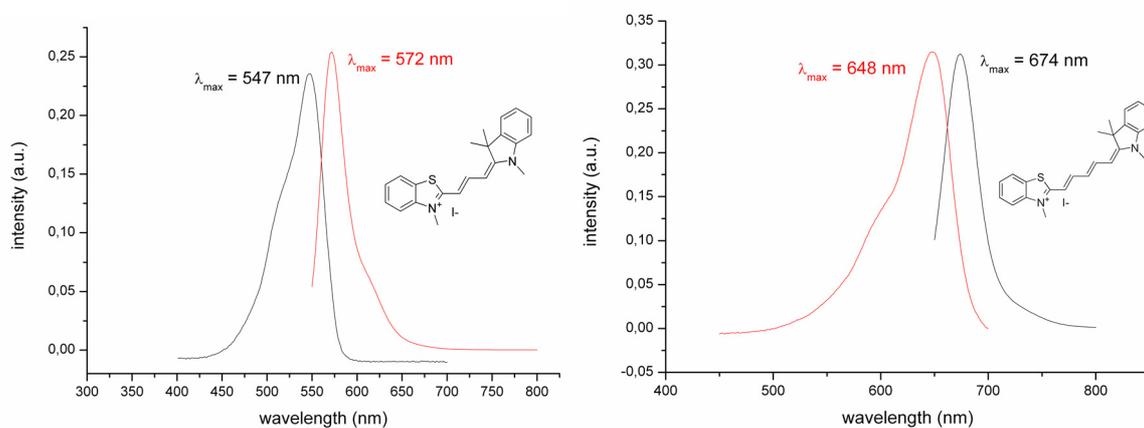


Figure S2. Absorption and fluorescence emission spectra of unsymmetrical trimethine (C3) and pentamethine (C5) cyanine dyes in DMSO solutions.

1.5 Kinetic studies.

In order to gain further insight into the kinetics of imine formation, we used fluorescence spectroscopy to monitor the formation of imine **4** from a mixture of amine **1** and Fisher's base aldehyde **2** either at room temperature or at 60 °C. Briefly, a mixture of compounds **1** and **2** (25 mM each) in DMSO was stirred at room temperature or at 60°C and the formation of imine **4** was quantitatively monitored by fluorescence spectroscopy ($\lambda_{\text{exc}} = 480 \text{ nm}$) until equilibrium was reached (see spectrum below). At room temperature, thermodynamic equilibrium was reached after 48h while increasing temperature to 60°C resulted not only in a different distribution at equilibrium (5-fold increase of the amount of imine **4** formed), equilibrium which was also reached much faster (less than 24h).

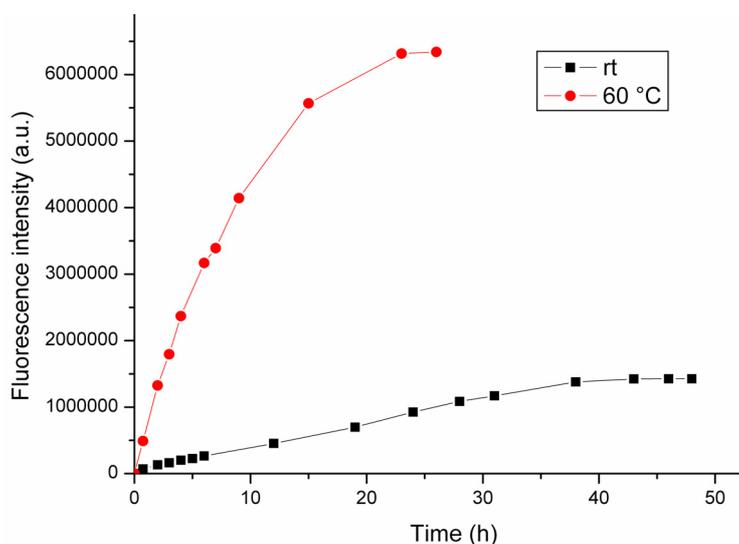


Figure S3. Kinetics of formation of imine **4** from a stoichiometric mixture of amine **1** and aldehyde **2** (25mM each) at room temperature (black squares) or 60°C (red circles)

2 SPECTROSCOPIC STUDIES

2.1 General Experimental

UV spectra were recorded on a UV-Visible Varian Cary 3 spectrophotometer in a 1 mm pathlength cuvette. Fluorescence emission spectra were recorded in quartz cells (with a 5 mm pathlength) at 20 °C on a Jobin Yvon Fluorolog 3.22 instrument. The excitation and emission bandwidths were fixed to 3 nm and 3 nm respectively.

2.2 Fluorescence Quantum Yields (QY)

Quantum yields of aldehydes and imines **2-5** in DMSO were determined relative to either quinine sulfate in 0.5M H₂SO₄, Fluorescein in 0.1M NaOH or Sulforhodamine 101 in ethanol assuming quantum yields of 0.55, 0.92 and 1.0, respectively (Standards purchased from *Invitrogen*). For fluorescence measurements, the total absorption of the samples never exceeded 0.06, making the inner filter effect negligible.

Fluorescence quantum yields (QY) were calculated according to the following equation :

$$\phi_X = \phi_{ST} (\text{Grad}_X / \text{Grad}_{ST}) (\eta_X / \eta_{ST})^2$$

Where the subscripts ST and X denote standard and sample respectively, ϕ is the fluorescence quantum yield, *Grad* the gradient from the plot of integrated fluorescence intensity vs absorbance, and η the refractive index of the solvent.

3 DENSITY FUNCTIONAL THEORY CALCULATIONS

All DFT calculations were carried out with the program GAUSSIAN 03. The charge and multiplicity of the system are both 1. The initial geometry of the dyes corresponds to that shown in Scheme 1 (main text), i.e. the configuration with all double bonds in a *trans* conformation with the sulphur atom being *cis* to the methine chain. We used Becke's three-parameter hybrid exchange functional (B3)ⁱ with the correlation functional of Lee, Yang, and Parr (LYP)ⁱⁱ. The geometry of the ground states was optimised with the basis set 6-31G(d); it is planar for all dyes (with the exception of the methyl groups, of course). The vertical excitation energies were calculated by single-point TD-DFT calculations using the basis set 6-311G(d) and the polarisable continuum model (PCM) of Tomasi and coworkersⁱⁱⁱ with a dielectric constant $\epsilon = 46.7$ (DMSO). The reported values for vertical excitation in the text are scaled values, ΔE_{scaled} , that have been obtained by applying the correction scheme of Champagne *et al.*: $\Delta E_{\text{scaled}} = -0.6017 \text{ eV} + 1.0941 \Delta E_{\text{calc}}$ (with ΔE_{calc} being the excitation energy extracted from the TD-DFT calculations).

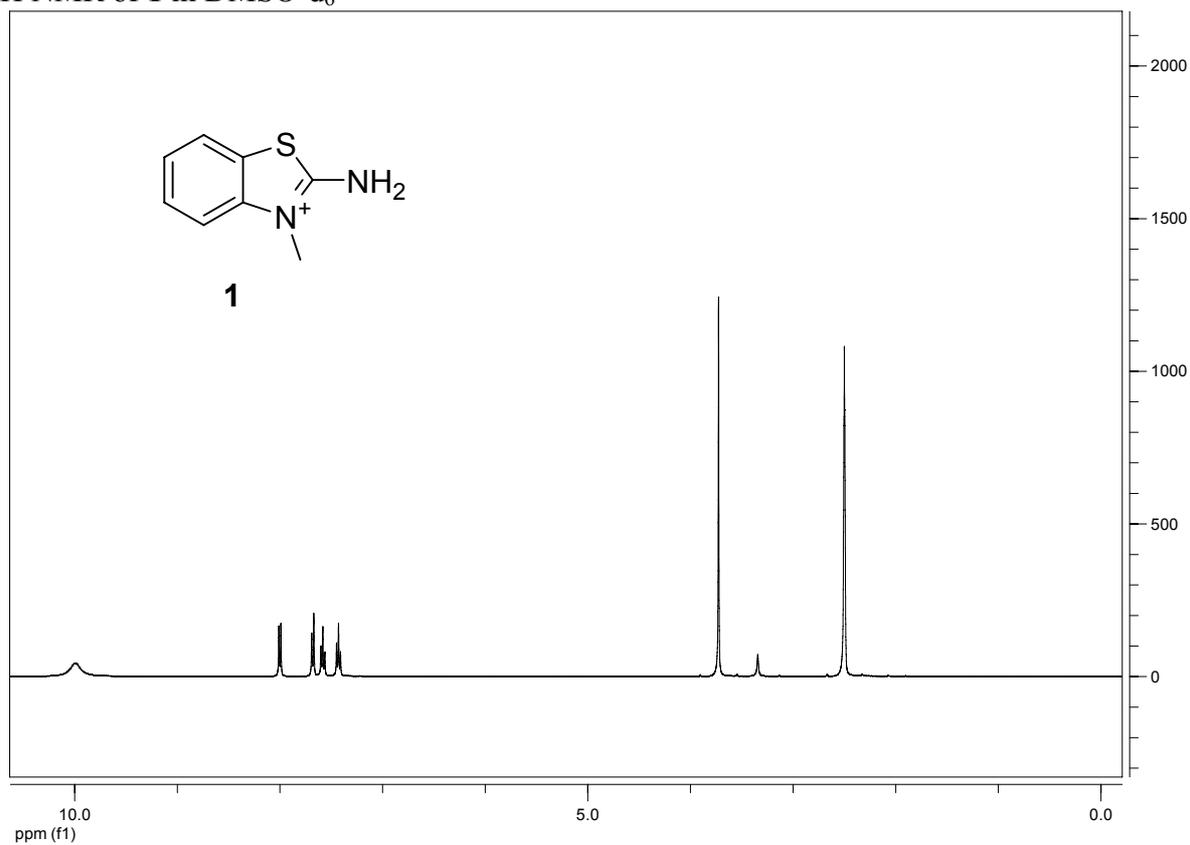
i) Becke, A. D.; *J. Chem. Phys.* **1993**, *98*, 5648.

ii) Lee, C.; Yang, W.; and Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

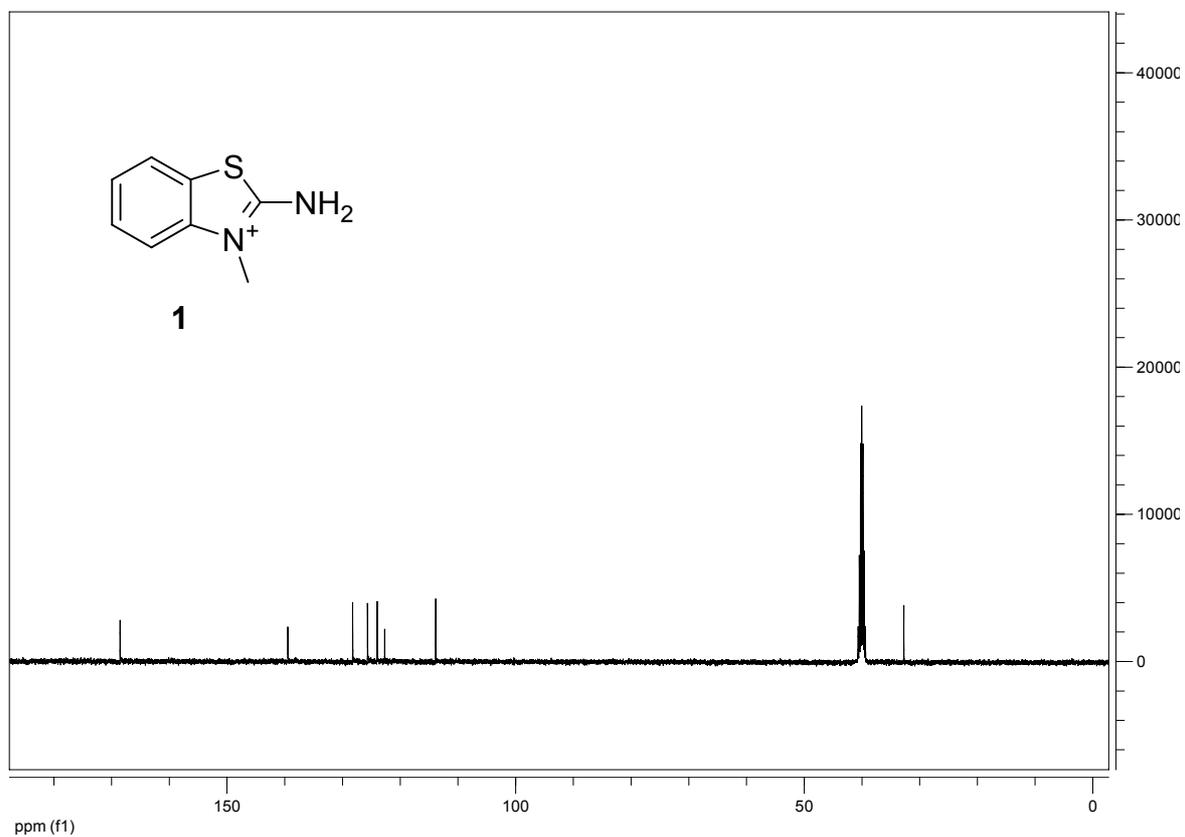
iii) Miertus, S.; Scrocco, E.; and J. Tomasi, *J. Chem. Phys.* **1981**, *55*, 117.

4 NMR SPECTRA

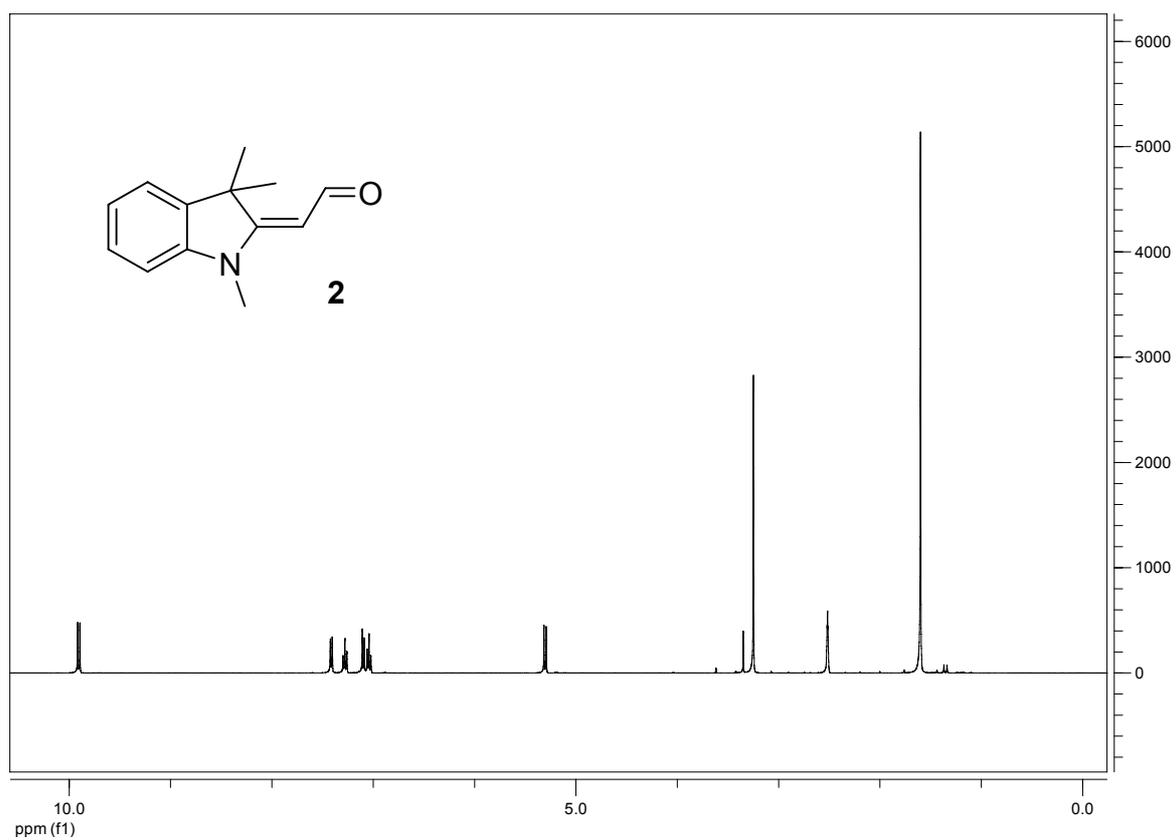
^1H NMR of **1** in DMSO-d_6



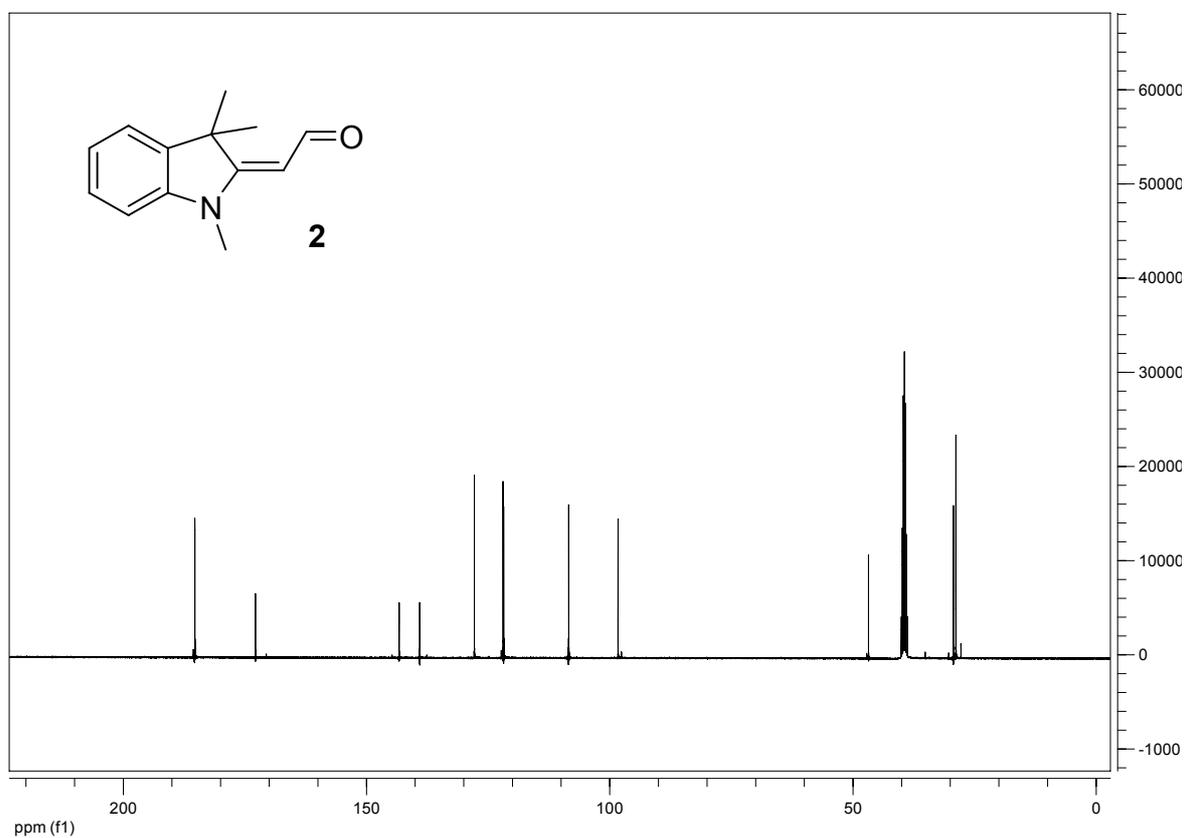
^{13}C NMR of **1** in DMSO-d_6



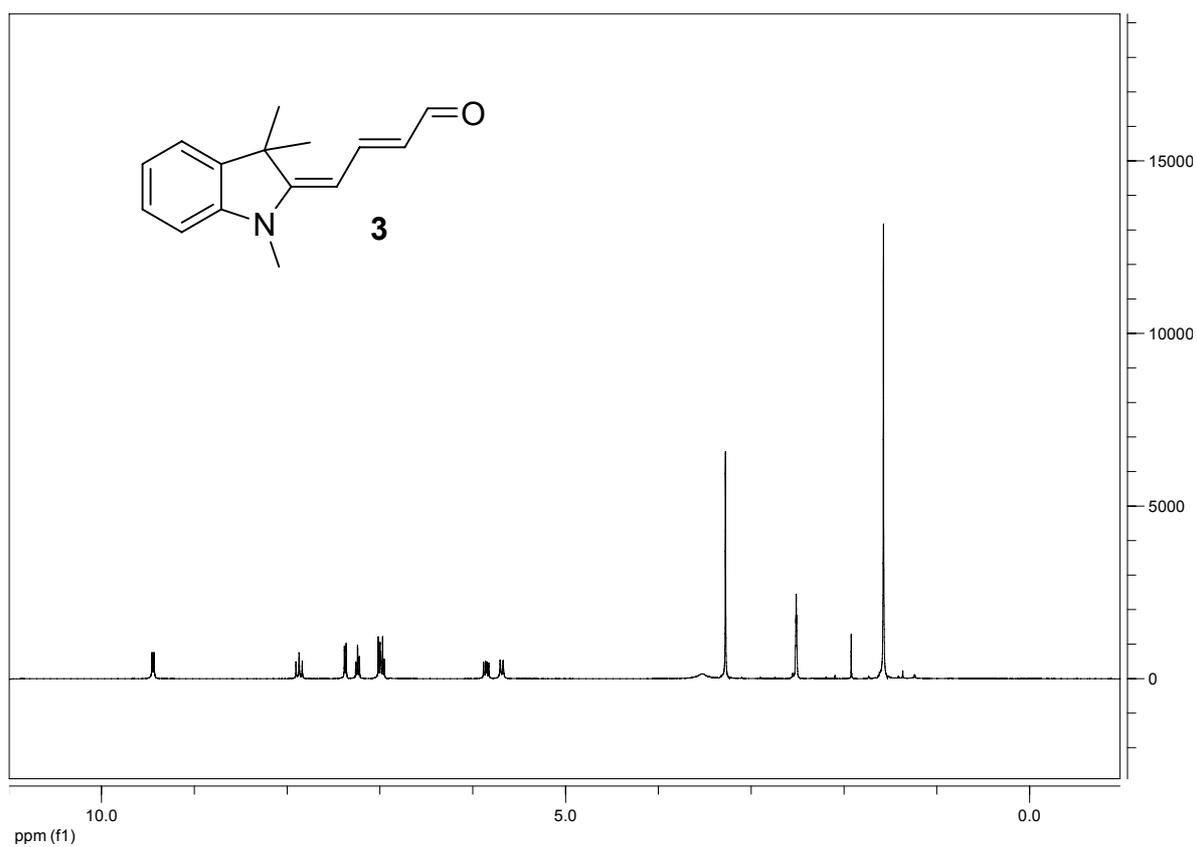
^1H NMR of aldehyde **2** in DMSO-d_6



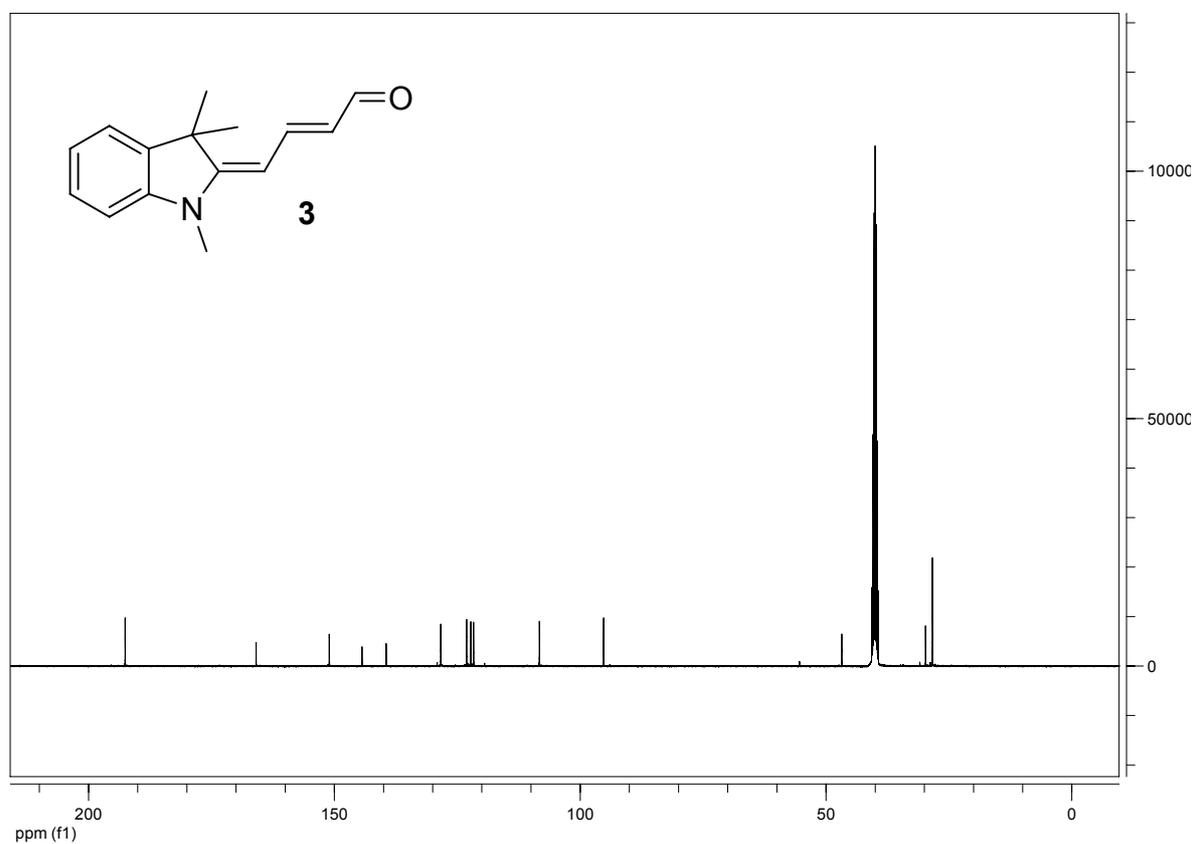
^{13}C NMR of aldehyde **2** in DMSO-d_6



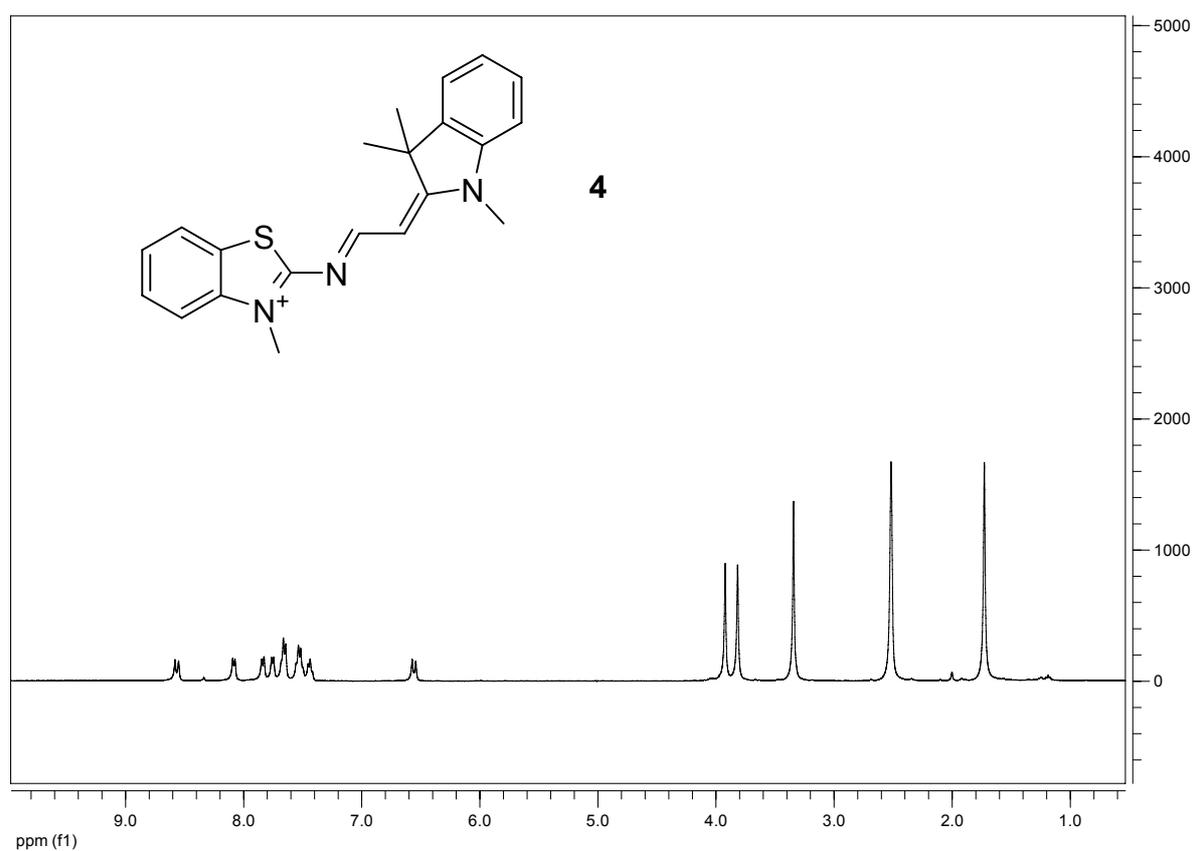
^1H NMR of aldehyde **3** in DMSO-d_6



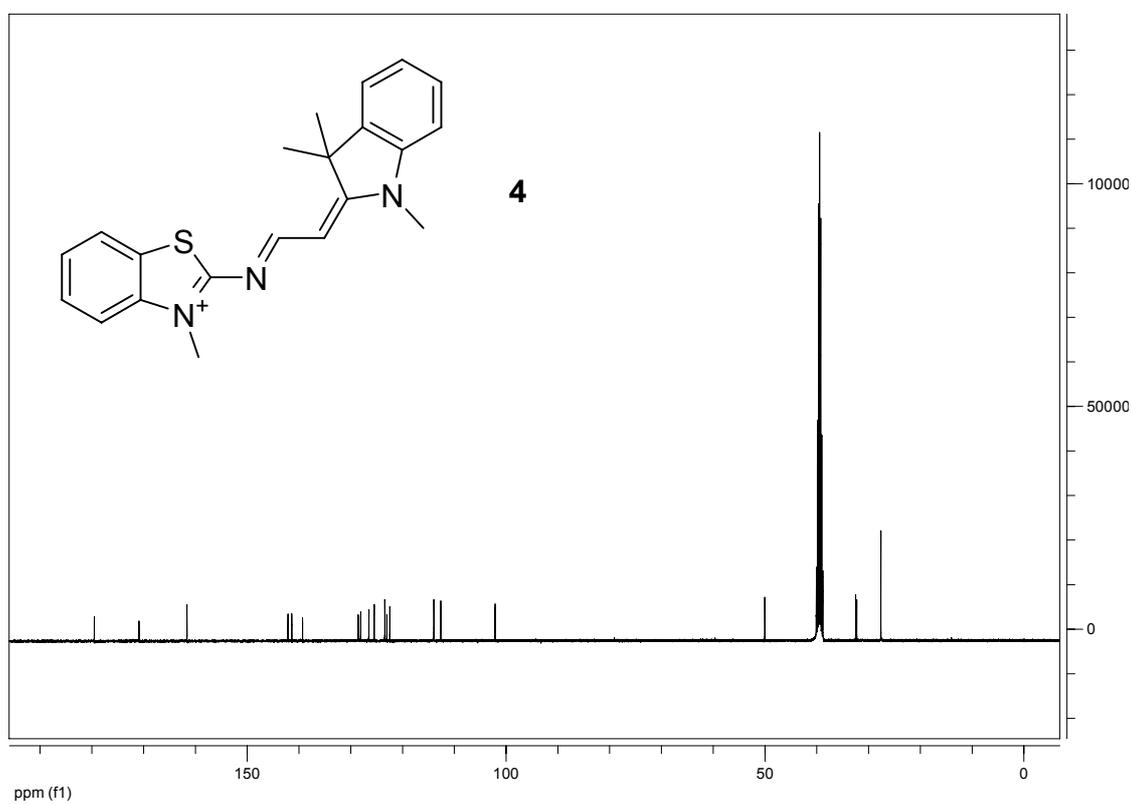
^{13}C NMR of aldehyde **3** in DMSO-d_6



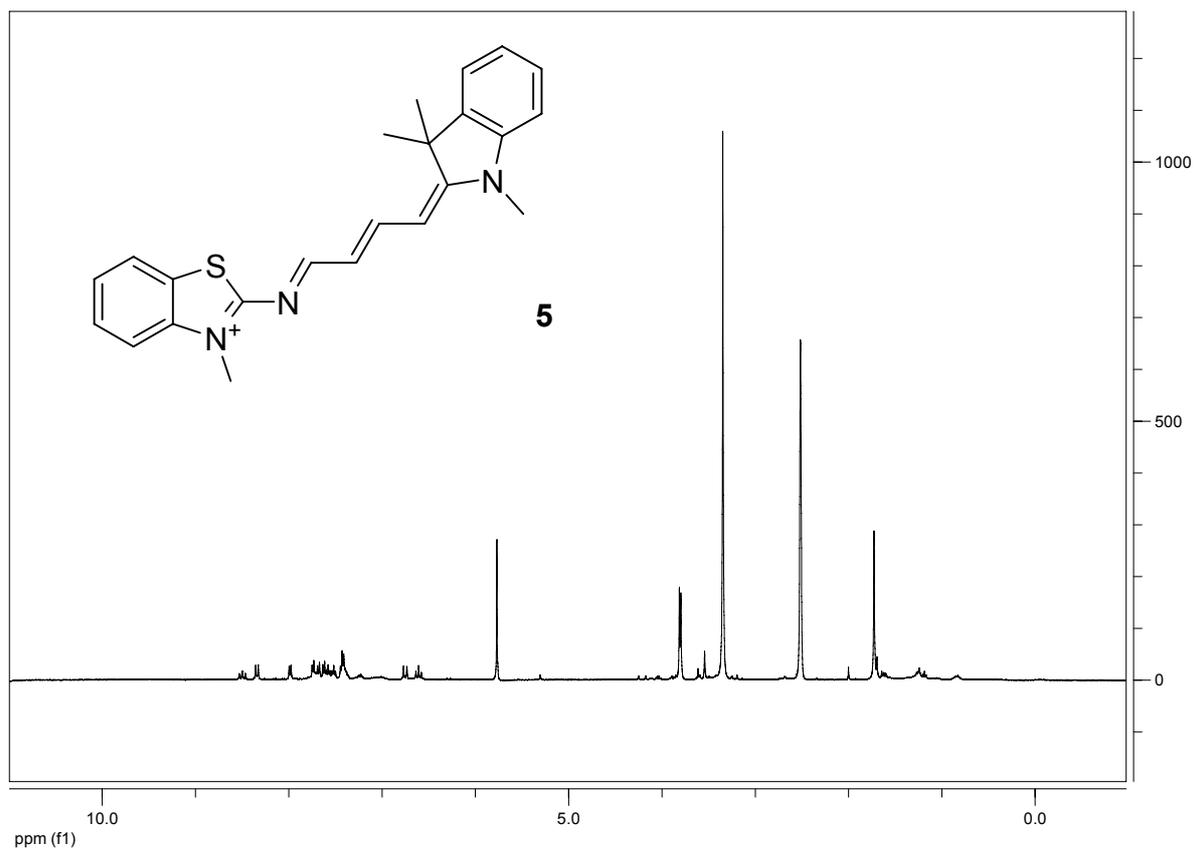
^1H NMR of cyanine **4** in DMSO-d_6



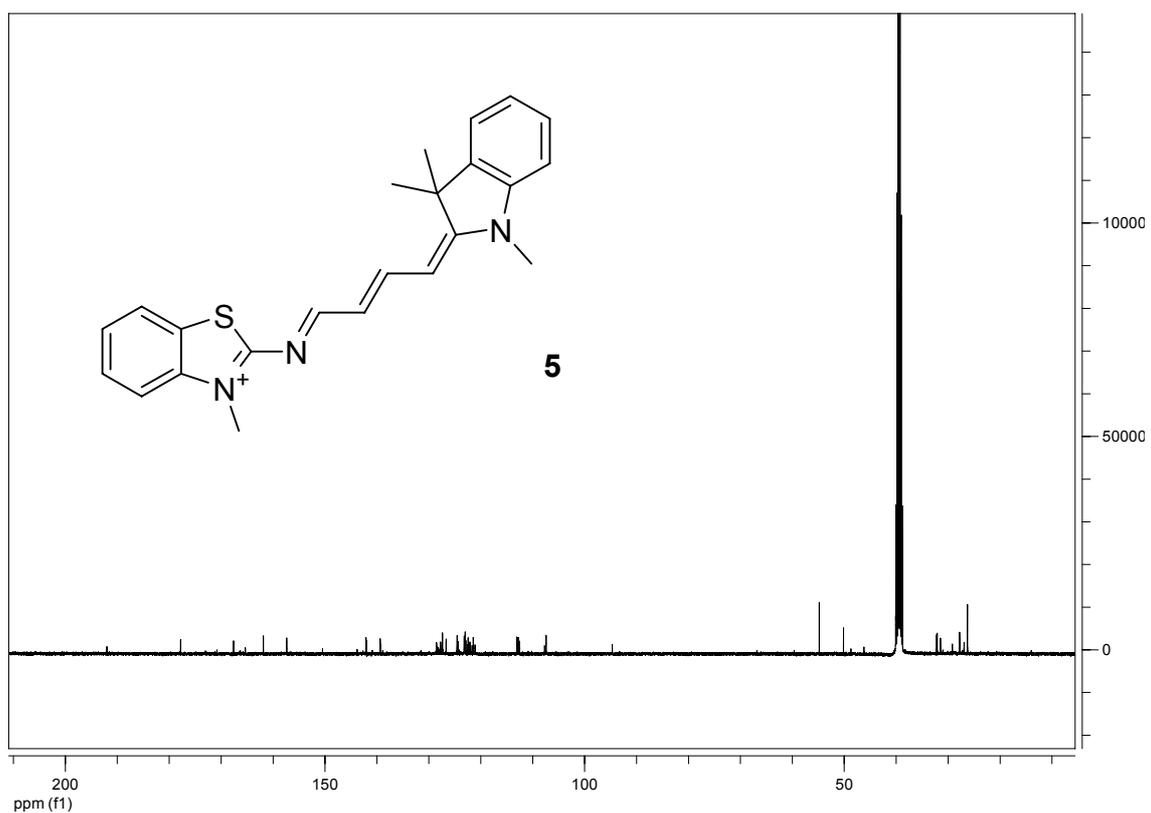
^{13}C NMR of cyanine **4** in DMSO-d_6



^1H NMR of cyanine **5** in DMSO-d_6 (additional signals in the ^1H and ^{13}C NMRs correspond to the amine **1** and aldehyde **3** and result from a partial hydrolysis of the imine bond during the NMR acquisitions)



^{13}C NMR of cyanine **5** in DMSO-d_6



Synthesis, spectroscopic and DNA alkylating properties of malondialdehyde (MDA) *bis*-imine fluorescent adducts†

Kamel Meguellati, Martin Spichty‡ and Sylvain Ladame*

Received 1st February 2010, Accepted 31st March 2010

DOI: 10.1039/c002157a

The synthesis of a series of malondialdehyde (MDA) fluorescent adducts that mimic the well-known pentamethine cyanine dyes is reported. This new subclass of *bis*-imino dyes shares some common spectroscopic properties with their polymethine analogues although absorbing at significantly shorter wavelengths. A small library of trimethine and pentamethine cyanine dye *bis*-imino analogues have been synthesised and characterised that cover a spectral range from blue to orange. Of particular interest is their capacity to act as mono- and *bis*-alkylating agents of nucleosides in general and of cytidine (and 2'-deoxycytidine) in particular.

Introduction

DNA alkylating agents, which include nitrogen mustards, have been used extensively in cancer chemotherapy.¹ These agents can be either mono- or bi-functional. The former have a single reactive group and therefore react with a single nucleophilic centre of a unique nucleobase (most commonly guanine or adenine).² The latter have two reactive groups and are therefore capable of reacting with two different nucleobases belonging to either a unique or two different DNA strands, thus resulting in intrastrand and interstrand cross-links, respectively.³ Interstrand cross-links prevent strand separation and hence constitute complete blocks to DNA replication and transcription.⁴ For this reason, a number of *bis*-alkylating agents have been designed that show promising anticancer activity. These include aziridines,⁵ nitrosoureas,⁶ alkane sulfonates⁷ or platinum derivatives.⁸ A major drawback of this class of compounds is their high cytotoxicity as a consequence of their high chemical reactivity. Reaction of the drug with intracellular nucleophiles (*e.g.* glutathione, proteins, ...) significantly lower their potential potency by preventing them from reaching their DNA target.⁹

There also exist a number of endogenous metabolites, products of enzymatic reactions, that are responsible for DNA damage and mutation. For instance, numerous studies have shown that formaldehyde (FA) and acetaldehyde (AA) were genotoxic and mutagenic to mammalian cells, mainly *via* formation of DNA–protein cross-links in target tissues.^{10,11} Malondialdehyde (MDA) is an endogenous product of lipid peroxidation and prostaglandin biosynthesis with mutagenic

properties in bacteria and human cells.¹² Although MDA can also form reversible Schiff adducts with deoxyadenosine and deoxycytosine, the most abundant MDA-DNA adduct is the so-called M₁dG that is formed by reaction of MDA with the exocyclic amine N(2) of deoxyguanosines followed by cyclization onto the N(1) position (Fig. 1).¹² This adduct, previously found in healthy human tissues recently proved mutagenic in human cells with MDA being suspected to be involved in the formation of DNA interstrand cross-links. MDA is also a widely used marker of oxidative lipid injury whose concentration varies in response to biotic and abiotic plant stress.¹³

Herein, we report on the synthesis and characterisation of an original family of fluorescent and reversible MDA adducts as *bis*-imino analogues of symmetrical pentamethine cyanine dyes. Analogues of trimethine cyanine dyes were also synthesised following a similar strategy. Although absorbing and emitting at slightly shorter wavelengths than their parent all-methine cyanine dyes, those reversible dyes have potential as MDA fluorescent sensors. More important are their potential applications as alkylating and/or cross-linking agents. The synthesis, stability and spectroscopic and DNA binding properties of this novel subclass of dyes are described.

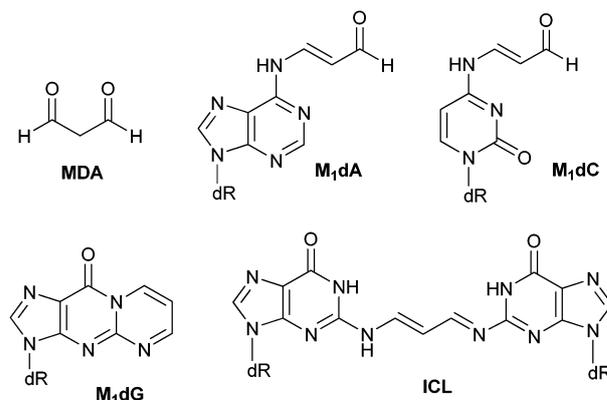


Fig. 1 Structures of MDA and of its mono- and *bis*-adducts with deoxyadenosine (M₁dA), deoxycytosine (M₁dC), deoxyguanosine (M₁dG). An example of interstrand cross-link (ICL) is also represented.

Institut de Science et d'Ingénierie Supramoléculaires (ISIS),
Université de Strasbourg, CNRS UMR 7006, 8,
allée Gaspard Monge, BP 70028, 67083 Strasbourg Cédex, France.
E-mail: s.ladame@isis.u-strasbg.fr; Fax: +33 (0)368855115;
Tel: +33 (0)368855210

† This article is part of the 2010 *Molecular BioSystems* 'Emerging Investigators' issue: highlighting the work of outstanding young scientists at the chemical- and systems-biology interfaces.

‡ Present address: Differentiation & Cell Cycle Group, Laboratoire de biologie moléculaire de la cellule, Ecole Normale Supérieure, 46 allée d'Italie, 69364 Lyon cedex 07, France.

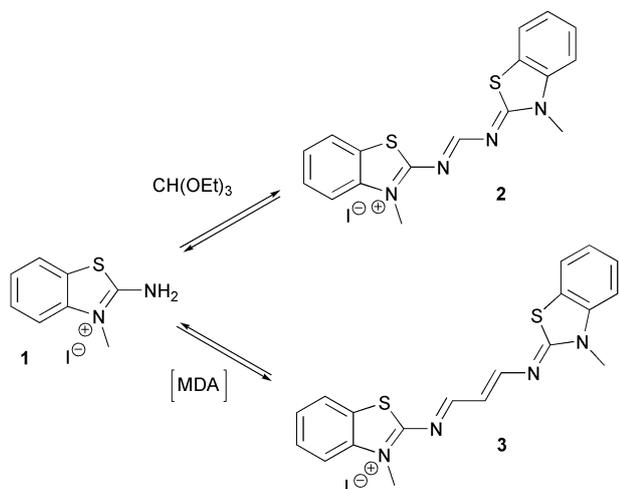
Results and discussion

Synthesis and spectroscopic properties of *bis*-imino analogues of symmetrical trimethine and pentamethine cyanine dyes

As a proof-of-concept experiment, firstly, analogues of trimethine and pentamethine cyanine dyes were synthesised by reaction of 1-methyl-2-aminobenzothiazolium iodide with triethyl orthoformate and 1,1,3,3-tetramethoxypropane (*i.e.* MDA precursor), respectively, using pyridine or acetic acid as a solvent (Scheme 1). Under such conditions, dyes **2** and **3** were isolated as pure solids by precipitation from the reaction mixture. They differ from commonly used symmetrical cyanine dyes by two C → N substitutions within the polymethine chain.

In order to demonstrate the versatility of our synthesis, a small number of *bis*-imino adducts was synthesized that cover a spectral range from blue to orange. They were obtained from reactions of triethylorthoformate or 1,1,3,3-tetramethoxypropane with either *N*-methyl-2-amino-quinolinium iodide¹⁴ or *N*-methyl-2-amino-naphthothiazolium iodide.¹⁵ The spectroscopic properties of the isolated dyes taken in DMSO are summarized in Table 1.

We recently reported the synthesis of unsymmetrical imino dyes from reaction between the aminobenzothiazolium derivative **1** and a Fisher's base aldehyde.¹⁶ From this example, it appeared that the introduction of one nitrogen atom within the polymethine bridge of a cyanine dye was responsible for a loss of around 60–70 nm of the maximum absorption and emission wavelengths of the dye. Interestingly, the substitution of a second carbon atom by a nitrogen atom accounts for an additional 70 nm drop of the maximum absorption wavelength. While *bis*-imino dyes **2** and **3** absorb maximally at 412 and 525 nm, respectively, (Table 1) their parent trimethine and pentamethine symmetrical cyanine dyes have their maximum absorption wavelengths at 562 and 660 nm respectively. It is also noteworthy that both imino dyes **2** and **3** are characterised by moderately large Stoke's shifts of 47 and 34 nm, respectively. Comparable effects are observed when replacing the benzothiazole heterocycles by either quinolines or naphthothiazoles. Although *bis*-quinoline and *bis*-benzothiazole dyes absorb and



Scheme 1 Synthetic route to fluorescent *bis*-imino adducts **2** and **3** as analogues of trimethine and pentamethine cyanine dye.

Table 1 Maximum absorption and emission wavelengths (taken in DMSO) of CH(OEt)₃ and MDA fluorescent adducts

Amine	Electrophile	<i>Bis</i> -imino dye	$\lambda_{\text{abs}}/\text{nm}$	$\lambda_{\text{em}}/\text{nm}$
	CH(OEt) ₃	2	412	459
	MDA	3	525	559
	CH(OEt) ₃	4	437	462
	MDA	5	530	560
	CH(OEt) ₃	6	465	506
	MDA	7	557	602

emit at nearly identical wavelengths, dyes from the naphthobenzothiazole family offer the advantage of absorbing (and emitting) at significantly longer wavelengths (maximum absorption wavelengths of $\lambda_{\text{max}} = 412$ and 465 nm for dyes **2** and **6**, respectively).

Due to the high molar absorptivity and tunable optical properties of these original MDA adducts, we believe that this dynamic and versatile system could represent the basis for the highly sensitive detection of MDA in biological media. MDA is commonly used as a biomarker of oxidative stress with respect to lipid peroxidation in body fluids or cells and is often quantified *via* detection of the fluorescent adduct it forms with thiobarbituric acid.¹⁷ Our strategy therefore represents a possible alternative to thiobarbituric acid that also offers the advantage of a much higher versatility with respect to the absorption and emission wavelengths of the adducts formed.

Stability of *bis*-imino cyanine dyes **2** and **3**

Due to their reversible nature, the reactions of imine formation from amines and carbonyl derivatives have been widely used in dynamic combinatorial chemistry.¹⁸ In order to assess the stability of these *bis*-imino dyes, the half-life values for compounds **2** and **3** in DMSO were determined by ¹H NMR and mass spectrometry. Because of their bright colour in solution, the degradation of both dyes could also be easily monitored by naked eye, looking at the slow disappearance of the bright yellow and bright red colours of solutions of **2** and **3**, respectively.

Bis-imino **2** proved significantly more stable than *bis*-imino **3** with a half-life of 36 h (*versus* 10 h for compound **3**) in DMSO. Interestingly, both dyes **2** and **3** were shown to undergo comparable mono-hydrolysis reactions in solution, leading to the formation of *N*-methyl-2-amino-benzothiazolium **1** and of imino-formamide **2a** and imino-aldehyde **3a**, respectively (Fig. 2 and 3). While appearance of formamide **2a** is characterised by a singlet at 8.96 ppm, formation of aldehyde **3a** is characterised by a doublet centred at 9.52 ppm. However, no traces of MDA¹⁹ (resulting from a complete hydrolysis of compound **3**) were ever detectable even after 72 h, thus

demonstrating the relative stability of mono-imine **3a**. It is also noteworthy that, in the case of compound **3**, a third degradation product (in addition to compounds **1** and **3a**) was also detectable by ^1H NMR although we were not able to unambiguously determine its structure.

Malondialdehyde (MDA) bis-imine fluorescent adducts as DNA alkylating prodrugs

Exposure of double-stranded DNA to MDA results in the formation of covalent adducts which are responsible for high levels of mutations when the DNA is replicated. Interestingly, more than 90% of the MDA-induced mutations occur at GC base pairs, consistent with the preferential reaction of MDA with deoxyguanosines to form M_1dG adducts (Fig. 1). There is also evidence that MDA can form DNA–DNA interstrand cross-links (ICL), a critical DNA damage that interferes with essential aspects of cellular metabolism (*e.g.* DNA replication and transcription). While endogenous alkylating molecules like MDA contribute to carcinogenesis through mutagenesis, an increasing number of antitumor agents have been designed that form covalent adducts with DNA and that can also form, in some cases, ICLs. A significant drawback of this class of drugs is a consequence of their extremely high reactivity, hence responsible for a serious lack of specificity. For instance, non-specific reactions with thiol and/or amine nucleophiles in the cytoplasm are common examples of drug detoxification.⁹

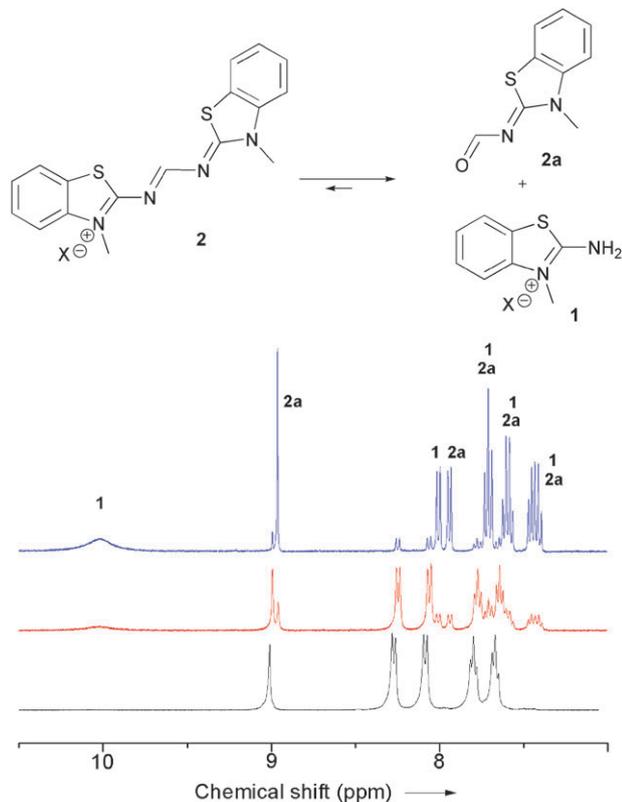


Fig. 2 ^1H NMR spectra of compound **2** (15 mM in d_6 -DMSO) recorded at $t = 0$ (black), 16 h (red) and 96 h (blue). After 4 days at rt, dye **2** is almost fully converted into amine **1** and aldehyde **2a**. Characteristic signals corresponding to imino aldehyde **2a** and amine **1** are highlighted. Half-life ($t_{1/2}$) of dye **2** was found to be 36 h.

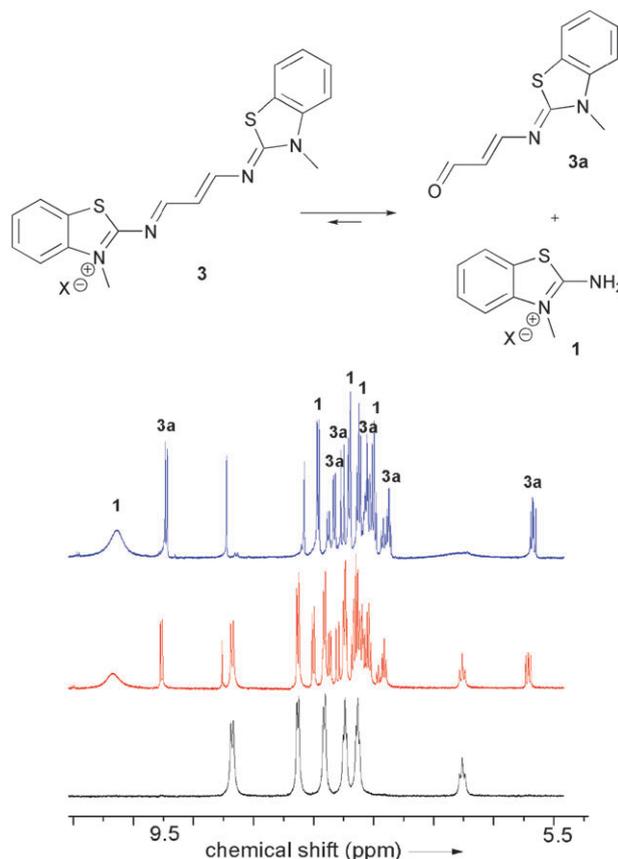


Fig. 3 ^1H NMR spectra of compound **3** (15 mM in d_6 -DMSO) recorded at rt at $t = 0$ (black), 12 h (red) and 36 h (blue). Characteristic signals corresponding to imino aldehyde **3a** and amine **1** are highlighted. Half-life ($t_{1/2}$) of dye **3** was found to be 10 h.

Successful strategies to render such molecules more selective for tumor cells include their conjugation with sequence-specific minor groove binders in order to direct the alkylating drug to the intended target site.²⁰ The use of prodrugs also proved to be an efficient alternative to increase either the lifetime or cell penetration of the DNA alkylator.²¹

Herein, we investigated whether our reversible cyanine dye analogues could act as DNA-alkylating prodrugs. The ability of pentamethine cyanine dyes to bind into the minor groove of double-stranded DNA is well documented in the literature.²² Therefore we reasoned that our *bis*-imino dyes could serve as vectors to deliver MDA within the minor groove of B-DNA. DNA alkylation could then proceed either (i) *via* hydrolysis of the dye once bound to DNA, thus liberating MDA at the vicinity of its target, or (ii) *via* direct *trans*-imination reaction between the nucleophilic nucleobases (G, A or C) and the imino dye. In order to demonstrate the potential of our *bis*-imino dyes as DNA alkylating agents, a reaction between compound **3** and each of the four natural nucleosides (adenosine, guanosine, cytidine and thymidine) was monitored by ^1H NMR, UV spectroscopy and mass spectrometry in DMSO. Briefly, to a solution of dye **3** (10 mM) in DMSO was added 5 equivalents of one nucleoside (A, T, C or G) and the four reactions were monitored individually by LC-MS over 72 h at room temperature. A nucleoside-free solution of dye was also used

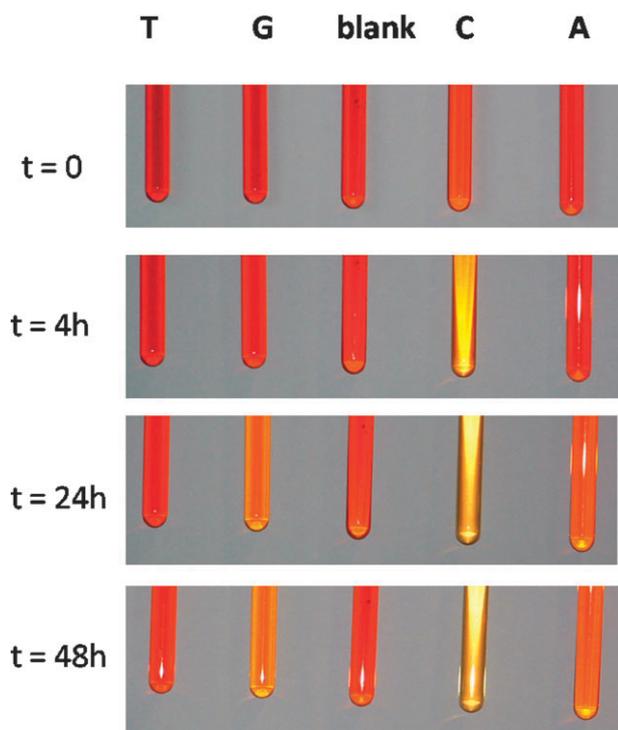


Fig. 4 Reaction between compound **3** (10 mM) and 5 equivalents of each of the four nucleosides (50 mM A, T, C or G) in DMSO. From left to right: thymidine, guanosine, no nucleoside, cytidine, adenosine. Reactions were monitored over 72 h at room temperature.

as a negative control. As anticipated, no reaction was ever observed with thymidine, the only nucleobase lacking an exocyclic primary amine. However, compound **3** proved capable of reacting covalently with the other three nucleosides, but with significantly different reactivities. Conveniently, the differences in reactivity of the four nucleobases could be easily monitored by naked eye, by looking at the solution's color change (Fig. 4).

Cytidine proved to be by far the most reactive nucleoside, inducing a red-to-yellow colour change within less than 4 h whilst a red-to-orange colour change appeared after 20–24 h only with either purine (A or G).

In order to further investigate the nature of the interaction between *bis*-imine **3** and nucleosides C, A and G, these three reactions were monitored by both NMR and LC-MS. This combined analysis allowed us to characterize the major covalent adducts formed upon reaction of dye **3** with each individual nucleoside.

Reaction of dye **3** with adenosine resulted in the slow and simultaneous appearance of the half-dye **3a** (resulting from the slow hydrolysis of *bis*-imine **3**) and of the adenosine-MDA covalent adduct (**M₁A**, Fig. 1). Both products were characterised by ¹H NMR, with the appearance of two new aldehyde signals, centred at 9.52 and 9.44 ppm, and corresponding to **3a** and **M₁A**, respectively. Formation of **M₁A** could potentially proceed *via* three different mechanisms: (i) complete hydrolysis of dye **3** into MDA and amine **1** followed by reaction of adenosine with the MDA generated *in situ*; (ii) *trans*-imination between dye **3** and adenosine followed by a mono-hydrolysis of either imine, leading either to the

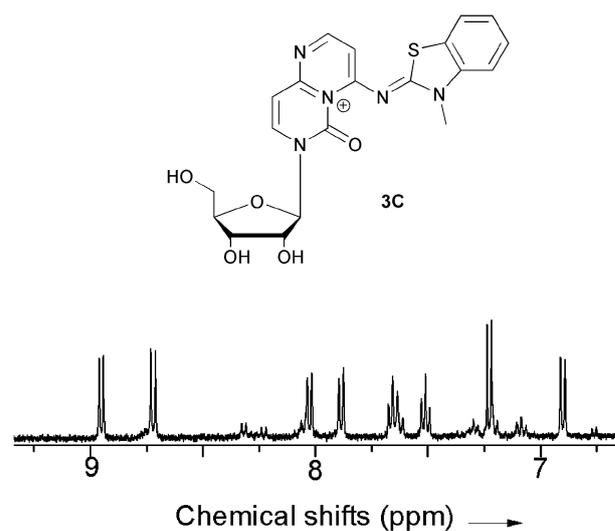


Fig. 5 Proposed structure of cytidine adduct **3C** ($m/z = 442$) and its ¹H NMR spectrum.

formation of **M₁A** or **3a**; (iii) partial hydrolysis of dye **3** into mono-imine **3a** and amine **1** followed by reaction between adenosine and **3a**. In a similar way, reaction of **3** with guanosine led to the slow formation of **M₁G** (Fig. 1).

Interestingly, reaction between **3** and cytidine led to the formation of multiple covalent adducts which proved significantly different to those obtained from adenosine and guanosine. Among them, one drew our attention because of its surprisingly high maximum absorption wavelength ($\lambda_{\text{max}} = 392$ nm) and because of its fluorescent properties ($\lambda_{\text{em}} = 472$ nm). This adduct **3C** was isolated by HPLC and a proposed structure based on tandem NMR and mass spectrometry studies is proposed in Fig. 5.

It is noteworthy that a similar chemical behaviour was observed when reacting dye **3** with 2'-deoxycytidine (dC). Again, the 2'-deoxy analogue of **3C** formed (**3dC**, $m/z = 426$) which showed identical UV absorption and fluorescence emission spectra as **3C**. The formation of such adducts would demonstrate the *bis*-alkylating reactivity of *bis*-imino dye **3** towards cytosines. A similar adduct was recently reported by Richter *et al.* which resulted from the reaction of clerocidin with unpaired cytosines.²³ Although the possible formation of such fluorescent adducts under biological conditions and in the context of an RNA (*e.g.* **3C**) or a DNA (*e.g.* **3dC**) strand remains to be demonstrated, imino dyes of general structure **3** could potentially serve as cytosine-specific alkylating agents, the formation of DNA adducts being detectable by fluorescence spectroscopy. Unfortunately, and due to the very poor water-solubility of the *bis*-imino dyes **3**, **5** and **7**, it was not yet possible to assess the stability and reactivity of this class of compounds under near-physiological aqueous conditions.

Conclusions

We report herein the first family of symmetrical cyanine dye *bis*-imino analogues that differ from the well-known trimethine and pentamethine cyanine dyes by two C → N substitutions within the polymethine chain. These changes account for

(i) a 120–140 nm decrease in the maximum absorption and emission wavelengths of the dyes and (ii) a relative instability of the dyes in solution due to the intrinsic reversible nature of imines. Of particular interest are the *bis*-imino dyes, analogues of pentamethine cyanine dyes, which are formed from reaction between MDA and two molecules of amino-substituted nitrogen-containing heterocycles. This fluorogenic reaction could serve as a basis for a versatile (*e.g.* tunable absorption and emission wavelengths by varying the nature of the heterocycle) alternative to thiobarbituric acid for sensing MDA *in vitro* and maybe also *in vivo*. More interesting is the reactivity of *bis*-imino dyes of general structure **3** towards natural nucleobases in general and cytosines in particular. We have demonstrated that dye **3** reacted selectively with cytidine and 2'-deoxycytidine to form a stable fluorescent adduct. The possible application of these dyes for alkylating cytosines with the context of a DNA or an RNA strand is currently underway in our group.

Experimental

^1H and ^{13}C NMR spectra were recorded on a Bruker Avance DRX 400 spectrometer at 400 and 100.6 MHz, respectively. Chemical shifts are reported as δ values (ppm) with reference to the residual solvent peaks. All reagents and solvents were obtained from commercial sources and used without further purification. Fluorescence emission spectra were recorded in quartz cells at 20 °C on a Jobin Yvon Fluorolog 3.22 instrument. The excitation and emission bandwidths were fixed to 3 and 3 nm, respectively. UV spectra were recorded on a V-670 UV-Visible spectrophotometer from JASCO in a 10 mm pathlength cuvette.

General procedure for the synthesis of *bis*-imino analogues of trimethine cyanine dyes

To a solution of *N*-methyl-2-amino-benzothiazolium, *N*-methyl-2-amino-quinolinium iodide or *N*-methyl-2-amino-naphthathiazolium iodide (3.5 mmol) in acetic acid (20 mL) was added triethylorthoacetate (2 mL) and the reaction mixture was stirred at 85 °C for 5 h. The solvent was then removed under reduced pressure and the residue was triturated with diethyl ether. The solid was collected by filtration under reduced pressure and washed with methanol (3 × 30 mL) and diethyl ether (3 × 30 mL). The desired *bis*-imino dyes were thus obtained pure as amorphous brown/yellow solids.

Compound 2 (yield 80%): ^1H NMR (400 MHz, DMSO): δ = 8.94 (s, 1H), 8.20 (d, J = 7.7 Hz, 2H), 8.00 (d, J = 8.2 Hz, 2H), 7.73 (t, J = 7.5 Hz, 2H), 7.60 (t, J = 7.5 Hz, 2H), 4.14 (s, 6H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 163.3, 138.0, 128.4 (× 3), 126.4 (× 3), 125.8, 123.8 (× 3), 115.0 (× 3), 34.0 (× 2) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{17}\text{H}_{15}\text{N}_4\text{S}_2^+$ 339.073; Found 339.071. **Compound 4** (yield 20%): ^1H NMR (400 MHz, DMSO): δ = 9.26 (s, 1H), 8.62 (d, J = 9.4 Hz, 2H), 8.16–8.12 (m, 4H), 8.00–7.96 (m, 4H), 7.69 (t, J = 7.2 Hz, 2H), 4.24 (s, 6H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 164.7, 161.1 (× 2), 142.6 (× 2), 140.0 (× 2), 133.8 (× 2), 130.0 (× 2), 126.5 (× 2), 124.5 (× 2), 118.1 (× 2), 116.9 (× 2), 34.8 (× 2) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{21}\text{H}_{19}\text{N}_4^+$ 327.160; found 327.158. **Compound 6** (yield 45%): ^1H NMR (400 MHz, DMSO): δ = 9.04 (s, 1H),

8.32 (d, J = 8.0 Hz, 2H), 8.27 (d, J = 8.0 Hz, 2H), 8.24–8.18 (m, 4H), 7.83 (t, J = 7.3 Hz, 2H), 7.74 (t, J = 7.5 Hz, 2H), 4.31 (s, 6H) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{25}\text{H}_{19}\text{N}_4\text{S}_2^+$ 439.105; found 439.101.

General procedure for the synthesis of *bis*-imino analogues of pentamethine cyanine dyes

To a solution of *N*-methyl-2-amino-benzothiazolium, *N*-methyl-2-amino-quinolinium iodide or *N*-methyl-2-amino-naphthathiazolium iodide (3.5 mmol) in acetic acid (20 mL) was added triethylorthoacetate (2 mL) and the reaction mixture was stirred at 85 °C for 5 h. The solvent was then removed under reduced pressure and the residue was triturated with diethyl ether. The solid was collected by filtration under reduced pressure and washed with methanol (3 × 30 mL) and diethyl ether (3 × 30 mL). The desired *bis*-imino dyes were thus obtained pure as amorphous dark red solids.

Compound 3 (yield 50%): ^1H NMR (400 MHz, DMSO): δ = 8.81 (d, J = 10.8 Hz, 2H), 8.15 (d, J = 8.3 Hz, 2H), 7.89 (d, J = 8.0 Hz, 2H), 7.68 (t, J = 7.8 Hz, 2H), 7.55 (t, J = 7.7 Hz, 2H), 6.51 (t, J = 10.9 Hz, 1H), 3.96 (s, 6H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 172.0, 171.9 (× 2), 139.3 (× 2), 128.3 (× 2), 126.1 (× 2), 123.9 (× 2), 123.7 (× 3), 118.4, 114.5 (× 3), 33.2 (× 2) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{19}\text{H}_{17}\text{N}_4\text{S}_2^+$ 365.089; found 365.085. **Compound 5** (yield 5%): ^1H NMR (400 MHz, DMSO): δ = 8.86 (d, J = 11.1 Hz, 2H), 8.58 (d, J = 7.9 Hz, 2H), 8.11–8.07 (m, 4H), 7.96 (t, J = 7.7 Hz, 2H), 7.72 (d, J = 9.2 Hz, 2H), 7.66 (t, J = 7.5 Hz, 2H), 6.43 (t, J = 10.8 Hz, 1H), 4.23 (s, 6H) ppm. ^{13}C NMR (100 MHz, DMSO): δ = 164.7, 159.2, 142.6 (× 2), 139.2 (× 2), 133.8 (× 3), 133.0 (× 3), 126.3 (× 2), 124.4 (× 2), 118.0 (× 3), 114.8 (× 2), 34.3 (× 2) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{23}\text{H}_{21}\text{N}_4^+$ 353.176; found 353.173. **Compound 7** (yield 40%): ^1H NMR (400 MHz, DMSO): δ = 8.85 (d, J = 11.1 Hz, 2H), 8.24 (d, J = 8.9 Hz, 2H), 8.15 (d, J = 8.0 Hz, 2H), 8.08–8.03 (m, 4H), 7.78 (t, J = 7.5 Hz, 2H), 7.66 (t, J = 7.5 Hz, 2H), 6.56 (t, J = 11.1 Hz, 1H), 4.07 (s, 6H) ppm. HR-MS ESI positive mode: m/z : calcd for $\text{C}_{27}\text{H}_{21}\text{N}_4\text{S}_2^+$ 465.120; found 465.121.

General procedure for the synthesis of and isolation of the covalent adduct **3C** (or **3dC**)

To a solution of *bis*-imino dye **3** (5 mg, 0.01 mmol) in DMSO (1 mL) was added cytidine or 2'-deoxycytidine (5 equivalents, 0.05 mmol from Sigma) and the reaction mixture was shaken at room temperature for 24 h. The crude mixture was finally purified by HPLC to afford the fluorescent adduct **3C** (or **3dC**) pure as a yellow/orange oil. **Compound 3C**: HR-MS ESI positive mode: m/z : calcd for $\text{C}_{20}\text{H}_{20}\text{N}_5\text{O}_5\text{S}^+$ 442.118; found 442.119. **Compound 3dC**: HR-MS ESI positive mode: m/z : calcd for $\text{C}_{20}\text{H}_{20}\text{N}_5\text{O}_4\text{S}^+$ 426.123; found 426.125.

Acknowledgements

S.L. thanks the Centre National de la Recherche Scientifique (CNRS) and the International Center for Frontier Research in Chemistry (FRC) for financial support. M.S. thanks the Institut de Science et d'Ingenierie Supramoleculaires (ISIS)

for hosting him as guest assistant professor. We also thank Dr Girish Koripelly for carefully proof-reading this manuscript.

Notes and references

- O. D. Scharer, *ChemBioChem*, 2005, **6**, 27; E. Izbiccka and A. W. Tolcher, *Curr. Opin. Invest. Drugs*, 2004, **5**, 587.
- K. Stone, M. B. Ksebati and L. J. Marnett, *Chem. Res. Toxicol.*, 1990, **3**, 33; A. K. Basu, S. M. O'Hara, P. Valladier, K. Stone, O. Mols and L. J. Marnett, *Chem. Res. Toxicol.*, 1988, **1**, 53.
- D. M. Noll, T. M. Mason and P. S. Miller, *Chem. Rev.*, 2006, **106**, 277.
- M. B. Smeaton, E. M. Hlavin, A. M. Noronha, S. P. Murphy, C. J. Wilds and P. S. Miller, *Chem. Res. Toxicol.*, 2009, **22**, 1285; Y. M. Akkari, R. L. Bateman, C. A. Reifsteck, S. B. Olson and M. Grompe, *Mol. Cell. Biol.*, 2000, **20**, 8283.
- S. T. Huang, H. D. Tsai, H. S. Kuo, Y. P. Yang, Y. C. Peng and Y. L. Lin, *ChemBioChem*, 2004, **5**, 797.
- P. Merle, D. Morvan, D. Caillaud and A. Demidem, *Anticancer Res.*, 2008, **28**, 21.
- S. D. Mertins, T. G. Myers, S. L. Holbeck, W. Medina-Perez, E. Wang, G. Kohlhagen, Y. Pommier and S. E. Bates, *Mol. Cancer Ther.*, 2004, **3**, 849.
- For a recent example see K. S. Lovejoy, R. C. Todd, S. Zhang, M. S. McCormick, J. A. D'aquino, J. T. Reardon, A. Sancar, K. M. Giacomini and S. J. Lippard, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8902.
- P. K. Smitherman, A. J. Townsend, T. E. Kute and C. S. Morrow, *J. Pharmacol. Exp. Ther.*, 2004, **308**, 260; C. S. Morrow, P. K. Smitherman, S. K. Diah, E. Schneider and A. Townsend, *J. Biol. Chem.*, 1998, **273**, 20114.
- G. Speit and O. Merk, *Mutagenesis*, 2002, **17**, 183; M. J. Solomon and A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 6470.
- B. Matter, R. Guzza, J. Zhao, Z. Z. Li, R. Jones and N. Tretyakova, *Chem. Res. Toxicol.*, 2007, **20**, 1379.
- H. Wang, L. J. Marnett, T. M. Harris and C. J. Rizzo, *Chem. Res. Toxicol.*, 2004, **17**, 144; X. Zhou, K. Taghizadeh and P. C. Dedon, *J. Biol. Chem.*, 2005, **280**, 25377; L. J. Niedernhofer, J. S. Daniels, C. A. Rouzer, R. E. Greene and L. J. Marnett, *J. Biol. Chem.*, 2003, **278**, 31426.
- M. W. Davey, E. Stals, B. Panis, J. Keulemans and R. L. Swennen, *Anal. Biochem.*, 2005, **347**, 201.
- 2-Aminoquinoline (1 g, 6.93 mmol, purchased from TCI Europe) was dissolved in a mixture of acetonitrile (40 mL) and iodomethane (2 mL). The solution was stirred for 14 h at 45 °C. The iodomethane was then evaporated off and the suspension was filtered under vacuum. The white precipitate was then washed with acetonitrile and diethyl ether, thus leading to the desired *N*-methyl-2-amino-quinolinium iodide as an amorphous white solid (1.68 g, 75%). NMR ¹H (DMSO) δ = 9.11 (s broad, 2H); 8.36 (d, *J* = 9.3 Hz, 1H); 8 (t, *J* = 7.7 Hz, 2H); 7.88 (t, *J* = 7.7 Hz, 1H); 7.58 (t, *J* = 7.7 Hz, 1H); 7.15 (d, *J* = 9.3 Hz, 1H); 3.89 (s, 3H). NMR ¹³C (DMSO) δ = 154.8, 141.5, 137.3, 132.6, 129.6, 125.0, 121.9, 116.3, 114.6, 34.2. HR-MS ESI positive mode: *m/z*: calcd for C₁₀H₁₁N₂⁺ 159.092; found 159.094.
- 2-Naphthylamine (1 g, 6.98 mmol) and sodium thiocyanate (1.13 g, 13.96 mmol) were dissolved in methanol (60 mL) at rt and the pale yellow/brown solution was cooled to -25 °C. Bromine (0.3 mL, 5.82 mmol) was added dropwise and the reaction mixture was stirred below -10 °C for 2 h. The pink suspension was then warmed up to rt and 400 mL of water was added. The solid was then collected by filtration under vacuum, washed with water and dried under vacuum. To this pale yellow solid were added ethanol (40 mL) and 1.5 M HCl (50 mL) and the suspension was refluxed for 2 h. The hot suspension was then filtered under vacuum and the filtrate was neutralized with K₂CO₃ at 0 °C until reaching pH 7-8. The aqueous solution was then extracted with CH₂Cl₂ (3 × 30 mL), the organic layers were dried over Na₂SO₄, filtered and concentrated under vacuum to giving the desired naphtha-[2,1-*d*]thiazol-2-amine as a brown powder (700 mg, 60%). To a solution of naphtha[2,1-*d*]thiazol-2-amine (700 mg, 3.49 mmol) in a mixture of acetonitrile (15 mL) and methanol (8 mL) was added iodomethane (2 mL) and the reaction mixture was stirred for 1 h at 45 °C. After evaporation under vacuum, the residue was triturated in acetonitrile. The solid was then filtered off solution, and was washed 3 times with hot acetonitrile and then with diethyl ether. The desired *N*-methyl-2-amino-naphthathiazolium iodide was thus obtained pure as a yellow solid (901 mg, 75%). NMR ¹H (DMSO) δ = 10.14 (s broad, 2H); 8.21 (d, *J* = 8.1 Hz, 1H); 8.15 (d, *J* = 8.1 Hz, 1H); 8.07 (d, *J* = 8.1 Hz, 1H); 7.93 (d, *J* = 8.9 Hz, 1H); 7.75 (t, *J* = 8.1 Hz, 1H); 7.66 (t, *J* = 7.9 Hz, 1H); 3.86 (s, 3H). NMR ¹³C (DMSO) δ = 167.4, 136.6, 130.2, 129.1, 128.9, 128.4, 126.4, 123.5, 123.0, 117.8, 113.4, 32.7. HR-MS ESI positive mode: *m/z*: calcd for C₁₂H₁₁N₂S⁺ 215.066; found 215.066.
- K. Meguellati, M. Spichy and S. Ladame, *Org. Lett.*, 2009, **11**, 1123.
- H. Esterbauer and K. H. Cheeseman, *Methods Enzymol.*, 1990, **186**, 407.
- For recent reviews see: S. Ladame, *Org. Biomol. Chem.*, 2008, **6**, 219; J. M. Lehn, *Chem. Soc. Rev.*, 2007, **36**, 151; P. T. Corbett, J. Leclair, L. Vial, K. R. West, J. L. Wietor, J. K. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652.
- S. H. Bertz and G. Dabbagh, *J. Org. Chem.*, 1990, **55**, 5161.
- For a recent review, see: T. Bando and H. Sugiyama, *Acc. Chem. Res.*, 2006, **39**, 935.
- For recent examples see: M. Tercel, G. J. Atwell, S. Yang, R. J. Stevenson, K. J. Botting, M. Boyd, E. Smith, R. F. Anderson, W. A. Denny, W. R. Wilson and F. B. Pruijn, *J. Med. Chem.*, 2009, **52**, 7258; K. Bielawski and A. Bielawska, *ChemMedChem*, 2008, **3**, 536.
- K. C. Hannah, R. R. Gil and B. A. Armitage, *Biochemistry*, 2005, **44**, 15924; H. J. Karlsson, M. H. Bergqvist, P. Lincoln and G. Westman, *Bioorg. Med. Chem.*, 2004, **12**, 2369.
- S. N. Richter, I. Menegazzo, D. Fabris and M. Palumbo, *Nucleic Acids Res.*, 2004, **32**, 5658.

Dynamic Knoevenagel reactions: organocatalysis of reversible carbon-carbon bond formation processes

Kamel MEGUELLATI, Jean-Marie LEHN (draft of the publication)

Constitutional dynamic Chemistry (CDC) ^[1] is a powerful approach whereby molecular components can be reversibly combined *via* either covalent or non-covalent linkages to generate a dynamic library of products. An efficient approach to create chemical diversity may be based on the dynamic recombination of fragments linked through non-covalent interactions or reversible covalent bonds. In recent years, the latter dynamic covalent chemistry (DCC) has emerged as a powerful way to create dynamic libraries of increasing structural diversity and screening of these Dynamic Combinatorial Libraries (DCLs) provides new perspectives in drug design ^[2] and material sciences ^{[3], [4]}. Representative examples of reversible reactions compatible with the DCC concept include amine/carbonyl condensations^[5], transesterifications ^[6], disulfide exchange ^[7], peptide exchange ^[8], boronic ester formation ^[9], olefin metathesis ^[10] and Diels-Alder condensation ^[11]. However, there is a constant need for new reversible reactions to create DCLs of increased structural and chemical diversity.

Aminocatalysis involving the reaction of primary and secondary amines with enamine and iminium intermediates is a very useful way to facilitate processes such as the Knoevenagel^[12] and aldolisation reactions. The Knoevenagel reaction consists in the condensation of an activated hydrogen compound such as β -ketoesters and malonates with aldehydes or ketones^[13] in the presence of primary and secondary amines, resulting in the formation of a C-C or C=C (after water eliminations bonds). Most commonly used secondary amine catalysts include piperidine, sarcosine, proline and its derivatives, the latter being extensively used for asymmetric organic catalysis. Recently, Wipf and *al.* reported an example of imine metathesis by reacting isobutyraldehyde and hydrocinnamaldehyde in water and demonstrated the reversible exchange between a library of aldehydes and a pyrazolotriazinone.^[14] Herein, we report examples of reversible Knoevenagel reactions and such processes amount to dynamic formation and cleavage of (carbon, carbon) bonds and allow for component exchange the ability of selected secondary amines to catalyze these reactions.

Results and discussions

A series of benzylidene derivatives were synthesized via a Knoevenagel reaction between barbituric acid as well as a malononitrile and the benzaldehydes **1a-c**: 4-dimethylaminobenzaldehyde, benzaldehyde, paranitrobenzaldehyde (Supporting information). The formation and the hydrolysis of these compounds were followed in presence of base (L-proline, L-proline methylester, piperidine). Such bases, especially L-proline, accelerated both the rate of formation and of hydrolysis of the compounds (see figure S1 and S3 for respectively the study of the formation and hydrolysis of barbituric acid benzylidene derivatives and figure S2 and S4 for respectively the study of the formation and hydrolysis of malononitrile benzylidene derivatives). Catalysis of the reverse reaction, the retro-Knoevenagel reaction, may be performed by the addition of a secondary amine on the benzylidene conjugate leading to an iminium intermediate by an addition-elimination process. The reversibility can be ascertained by the addition of another aldehyde component that would produce a new benzylidene conjugate via C-C bond interchange (figure 1). The capacity of various benzylidenes to undergo such exchange reactions in the presence of various aldehydes was investigated, in the absence or in the presence of different secondary amines such as proline, proline methylester or piperidine as potential catalysts.

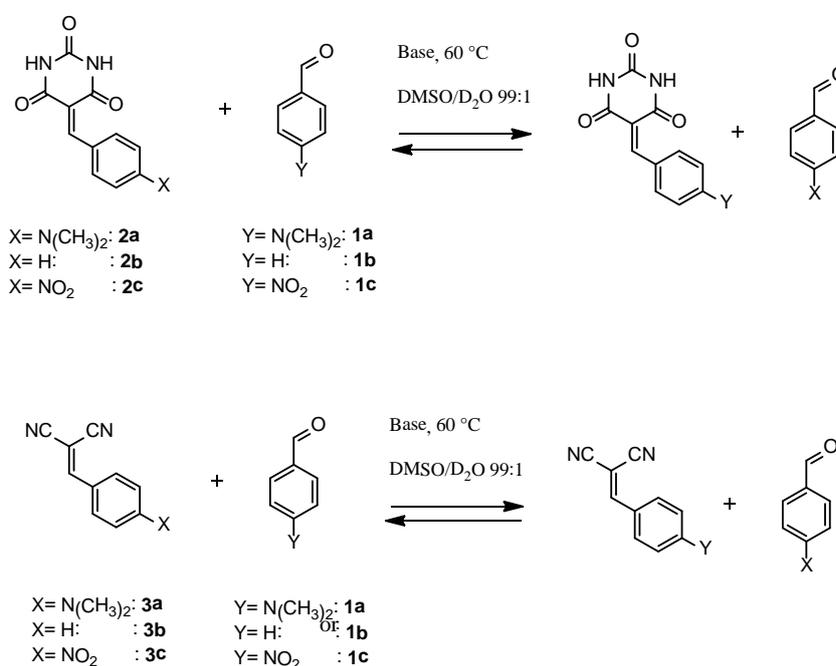


Figure 1: Examples of reversible Knoevenagel reactions involving exchange of the aldehyde moiety between a barbituric acid (top) and malononitrile (bottom) derived benzylidene derivative and a *para*-substituted benzaldehyde.

Synthesis and reactivity of the conjugated benzylidenes

The electrophilic behavior of benzylidenes is strongly dependent on the nature of the probe-head (i.e. malononitrile, barbituric acid or diethyl malonate). Most reactive are benzylidenes whose probe-head is the strongest electron-withdrawing group and has the lowest pKa. Thus, diethylmalonate benzylidenes (pKa diethylmalonate = 16.4) proved to be less reactive than the malononitrile benzylidenes (pKa malononitrile = 11.1) and much less reactive than the barbituric acid analogues (pKa barbituric acid = 4.5), consistent with a recent study^[15]. Because of the very low reactivity of diethylmalonate benzylidenes, we decided to focus our study on the malononitrile and barbiturate derivatives only. The stability of the benzylidenes toward hydrolysis also proved dependent on the nature of the para-substituent on the benzaldehyde component. Compounds bearing electron-donating substituents such as dimethylamine (NMe₂) were less reactive than those bearing an electron-attracting substituent.

The effects of the addition of a base during the synthesis or hydrolysis of each conjugated benzylidene were also examined (see supporting information). The reactions between one equivalent of aldehyde **1a-c** and one equivalent of either barbituric acid or malononitrile (12.8 mM each) were carried out in a mixture of DMSO/D₂O 99:1 (adjusted to an apparent pH 7) and heated at 60 °C, in the presence and in the absence of base as potential catalyst (proline, proline methyl ester or piperidine) and the formation of the corresponding conjugated benzylidene was monitored by ¹H NMR. The reverse reactions were carried out in parallel and the hydrolysis of each isolated conjugated benzylidene (12.8 mM in a mixture of DMSO/D₂O 99:1 at 60 °C, in the presence and in the absence of base) was monitored by ¹H NMR. After 36 h, the same distribution of all reaction components was observed by NMR for both forward and backward processes, thus demonstrating that the reactions were reversible and under thermodynamic control.

The effects of the addition of base on the kinetics and thermodynamic of the reaction were also investigated. Generally, all three bases tested were found to accelerate both the synthesis and the (reverse) hydrolysis of the conjugated benzylidene (see supporting information). Of all three bases tested, L-proline proved the most efficient catalyst, followed by L-proline methyl ester. Piperidine came last despite its higher basicity, possibly because of extensive protonation. These results demonstrate the likely participation of L-proline carboxylic acid into the catalytic mechanism. The most pronounced effect was observed for the formation of compound **2b** with a dimethylamine substituent for which L-proline

accelerated the reaction almost 10-fold ($t_{1/2}$ reduced from 19 to 2 hours) when compared to the reaction carried out in the absence of base. It is also noteworthy that the catalytic effect of L-proline was lower for benzylidenes substituted with electron-withdrawing group. This is mainly due to the fact that such compounds hydrolyze quickly, already in the absence of base.

Study of the Retro-Knoevenagel reaction

In order to confirm the reversibility of the Knoevenagel reaction, exchange of the aldehyde component was investigated, using all possible combinations of one benzylidene (**2a-c** and **3a-c**) with one aldehyde (**1a-c**) (Figure 1). Each conjugated benzylidene (12.8 mM) was solubilized in a mixture of DMSO/D₂O 99:1 (adjusted to apparent pH=7, see experimental), heated at 60 °C with one equivalent of a different substituted aldehyde and in the absence or in the presence of one of the three bases: proline, proline methyl ester or piperidine. For each reaction, the percentage of each component of the reaction mixtures (aldehydes and arylidenes) as function of time and the initial rates of disappearance of the starting benzylidene compound were determined by ¹H NMR spectroscopy.

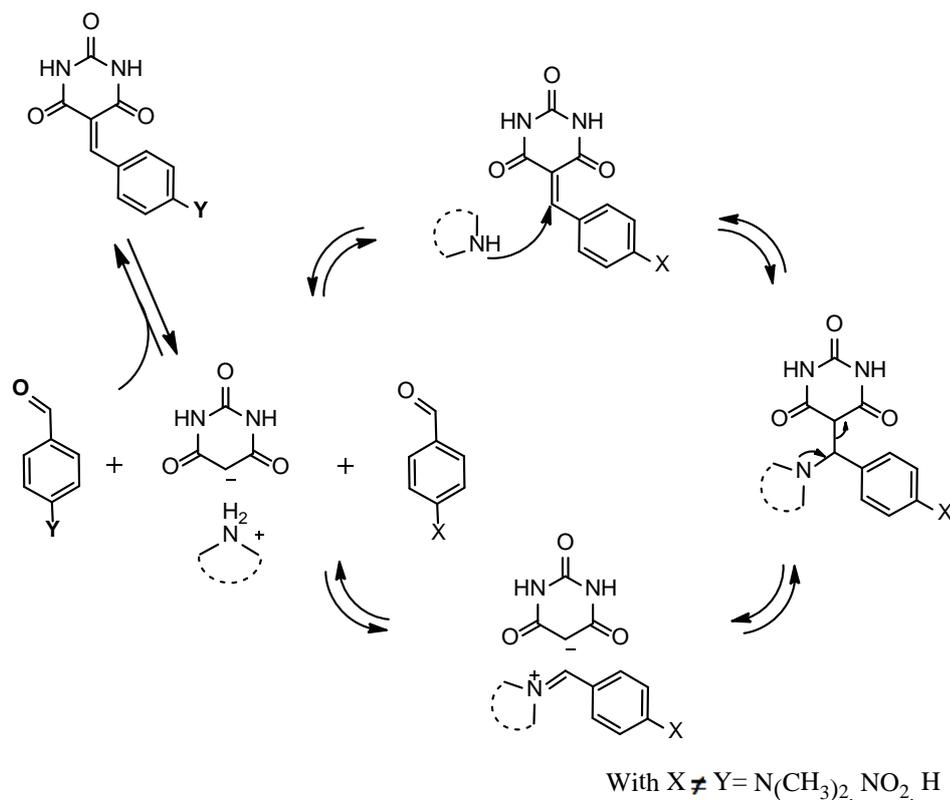


Figure 2: Possible mechanism of the amine-catalyzed retro-Knoevenagel and aldehyde exchange reaction.

	2a vs 1b				2b vs 1a			
<i>catalyst</i>	Blank	Proline	Pro.ester	Piperidine	Blank	Proline	Pro.ester	Piperidine
$t_{1/2}$ (h)	55	1	17	25	80	10	13	40
V_i (% per h)	0.15	4	1.25	0.3	2	0.64	0.64	1.72
	2a vs 1c				2c vs 1a			
<i>catalyst</i>	Blank	Proline	Pro. ester	Piperidine	Blank	Proline	Pro.ester	Piperidine
$t_{1/2}$ (h)	36	20	42	41	39	36	15	41
V_i (% per h)	0.33	1.2	0.4	0.7	0.33	1.13	5	0.34
	2c vs 1b				2b vs 1c			
<i>catalyst</i>	Blank	Proline	Pro. ester	Piperidine	Blank	Proline	Pro.ester	Piperidine
$t_{1/2}$ (h)	14	3	4	15	30	3	5	6
V_i (% per h)	0.8	5.83	2.5	1.05	1.14	14	5.2	4.5
33	3c vs 1b				3b vs 1a			
<i>catalyst</i>	Blank	Proline	Pro. ester	Piperidine	Blank	Proline	Pro.ester	Piperidine
$t_{1/2}$ (h)	5	0.6	2	1	--	2.5	--	--
V_i (% per h)	8.1	16.1	11	15.05	0.38	11.8	10	0.8
	3c vs 1a				3b vs 1c			
<i>catalyst</i>	Blank	Proline	Pro. ester	Piperidine	Blank	Proline	Pro.ester	Piperidine
$t_{1/2}$ (h)	75	34	49	50	5	2	3	5
V_i (% per h)	0.3	1	0.4	0.33	7.3	7.3	8	6

Table 1: Kinetic and thermodynamic parameters for the Retro-Knoevenagel reaction. Half-life $t_{1/2}$ and initial rate V_i . Conditions: experience majored at 60 °C in a mixture of DMSO/D2O with 10 % molar of base at pH=7 (adjusted with a solution of 0.1mM of DCl and/ or 0.1 mM of NaOD).

First, reactions were carried out starting from a benzylidene derivative obtained from barbituric acid and a given aldehyde. In all cases, aldehyde exchange was observed even in the absence of base. However, addition of 10 % molar of a base increased significantly the kinetics of the exchange. For instance, the half-time for exchange was reached up to 55 times faster upon addition of L-proline when compared to the control experiment without added base (reaction between **2a** and **1b**, Table 1). For the same reaction, piperidine reduced the reaction half-time by a factor of 2 only. There was almost no effect with the benzylidene from

the *p*-nitrobenzaldehyde because the hydrolysis was very fast, already in the absence of base. The benzylidenes formed from the benzaldehyde and the *p*-dimethylaminobenzaldehyde however, proved much more sensitive to base-catalysis.

Next, similar reactions were carried out starting from benzylidene of malononitrile. Table 1 indicates that the efficiency of the catalysis by L-proline was higher with malononitrile benzylidenes than with the barbituric acid derivatives, making proline the best catalyst for this particular reaction. Proline accelerates the exchange but also improves the reaction efficiency. For the exchange reaction between **3b** and **1a**, proline accelerates the reaction 10 times but also increases the conversion yield by 6-fold when compared with the base-free reaction. Therefore, L-proline has a dual effect on the kinetics of the reaction. Piperidine also improves the efficiency of the reaction, notably for the reaction between **3c** and **1b** (5-fold compared to the blank). It is noteworthy that no exchange reaction was ever observed when starting from compound **3a** which can be explained by its very low reactivity towards hydrolysis even in presence of catalytic amounts of L-proline.

Generally, proline methyl ester proved to be a less efficient catalyst compared to the free proline, thus suggesting participation of the carboxylic acid in the catalytic process. Although the secondary amine is the key moiety that is involved in the mechanism of the retro-Knoevenagel reaction, the iminium formed with the benzylidene is likely to be stabilized by the neighbouring carboxylate group, which cannot happen with the methyl ester analogue (Figure 2).

Metathesis reaction

Finally, the possibility of metathesis (crossing) reactions between barbiturate-benzylidenes and differently substituted malononitrile-benzylidenes was also investigated by ¹H NMR spectroscopy. Reactions were carried out using stoichiometric mixtures of both benzylidene derivatives (12.8 mM each) in a mixture of DMSO/D₂O 99:1 at 60°C (or 100°C) with 10 % L-proline as catalyst (Figure 3).

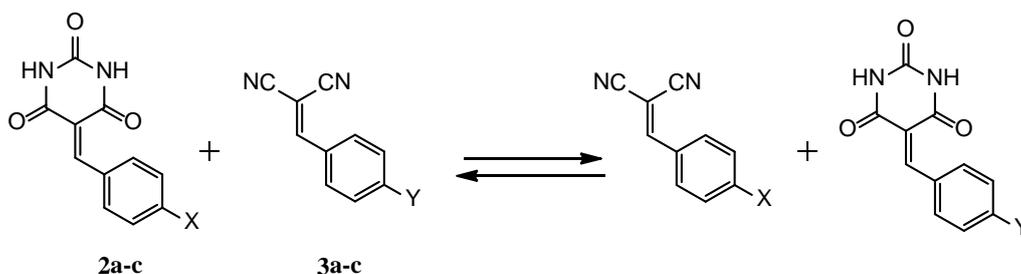


Figure 3: Reversible metathesis reaction with component exchange.

With compounds **2b** and **3c**, although traces of cross-metathesis derivatives (**2c** and **3b**) could be observed after less than one hour, equilibrium was only reached after 3 days. Interestingly, the same position of equilibrium (i.e. same composition of each species in solution) was obtained when starting from compounds **2c** and **3b**, thus demonstrating that the reaction is truly reversible and under thermodynamic control.

In order to generalize these observations, similar studies were carried out starting from derivatives **2a** and **3b** (or **2b** and **3a** for the reverse reaction). In this case, heating up to 100°C was required in order to reach equilibrium. However, once again the same position of equilibrium was obtained in both directions (after 4 days), thus confirming the results obtained in the previous experiment.

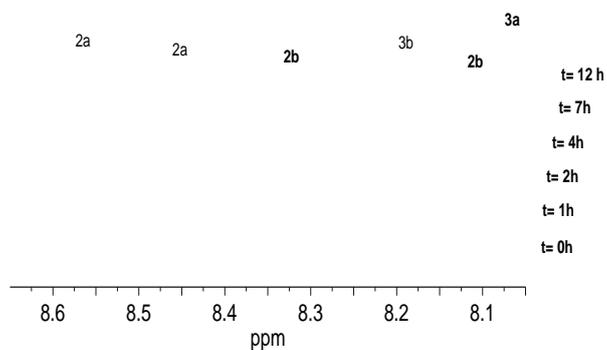


Figure 4: Cross metathesis: **2a** was mixed with **3b** in a mixture of DMSO/D₂O 99:1 at 60 °C, 12.8 mM. After few hours, products from cross metathesis were observed.

Conclusion

The results obtained here describe dynamic C=C bond processes based on Knoevenagel and cross-metathesis reactions. They demonstrate that these reactions are

reversible, under thermodynamic control and can occur even in the absence of a base catalyst. However, secondary amines such as L-proline, L-proline methyl ester and piperidine act as exchange catalysts accelerate the reactions. The fact that L-proline is a significantly better catalyst than its methyl ester analogue and favors formation of the most stable intermediate. We may also note that the reactions studied represent the C=C carbon analogues of the extensively studied C=N imine exchange processes. The present results will allow the creation of DCLs of higher chemical diversity, thus allowing the selection by DCC of receptors, ligands or biomaterials of increased complexity.

SUPPORTING INFORMATION

The synthesis of the benzylidene derivated from the barbituric acid was performed by condensation of barbituric acids with the aldehyde in hot water^[16] and the benzylidene derivated from the malononitrile were synthesized in emulsion^[17] at room temperature.

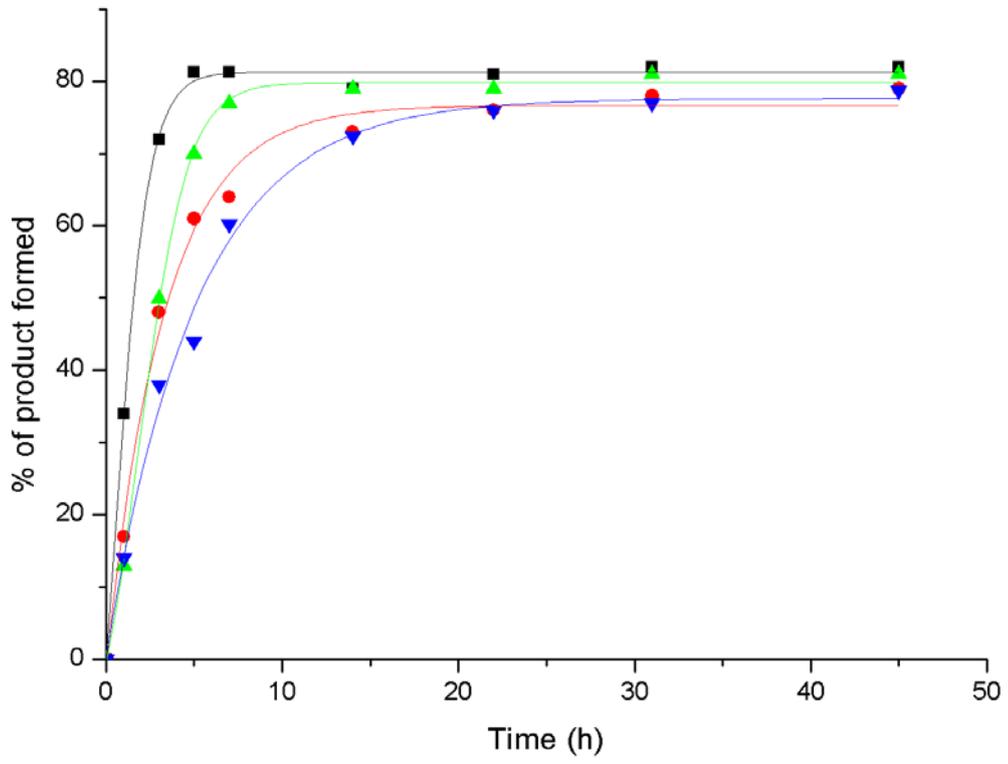
All compounds synthesized had properties in agreement with literature data: (m.p.

Synthesis and stability of barbituric and malononitrile benzylidene conjugates

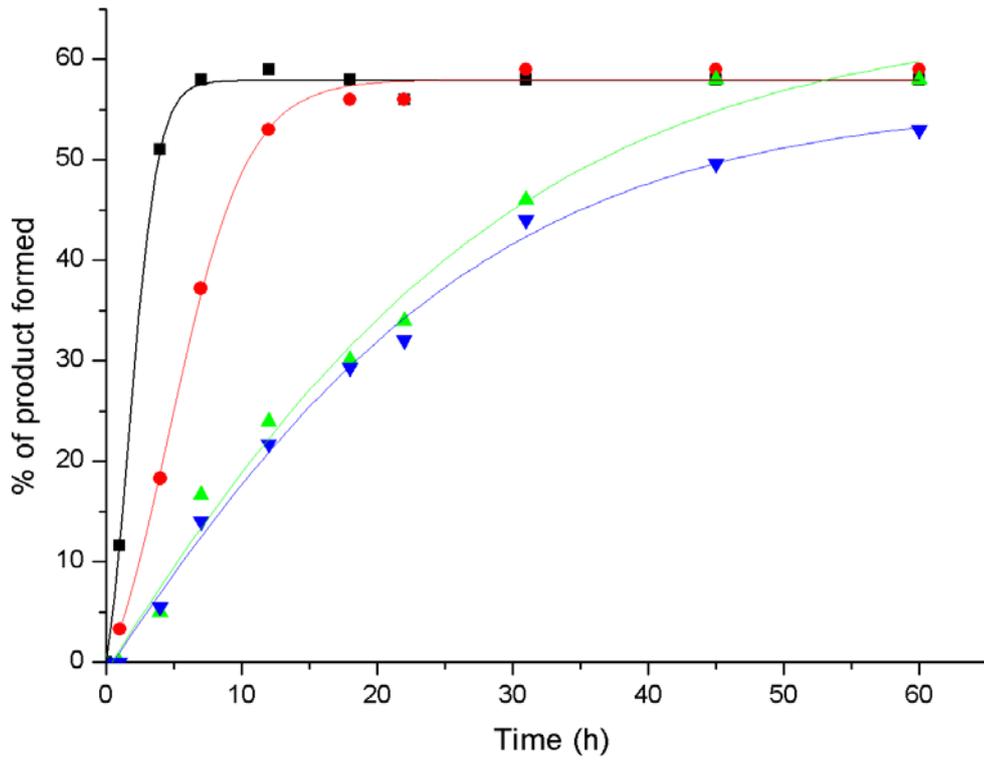
	2a								2b							
	Proline		Pro. ester		Piperidine		Blank		Proline		Pro. ester		Piperidine		Blank	
	F	H	F	H	F	H	F	H	F	H	F	H	F	H	F	H
t_{1/2} (h)	1.2	-	2	-	2	-	2.1	-	2	-	6	-	13	-	19	-
V_i (% per h)	19	9	15.5	1.5	18.5	1.5	12	1.5	13.1	9.2	5	6	1.25	4.5	1.25	2
	2c								3a							
	Proline		Pro. ester		Piperidine		Blank		Proline		Pro. ester		Piperidine		Blank	
	F	H	F	H	F	H	F	H	F	H	F	H	F	H	F	H
t_{1/2} (h)	2	-	3	-	4	-	4	-	2	1	6	13	-	18	-	-
V_i (% per h)	16	10	10.5	10	4	10	2.7	10	28	1.3	1.3	0.4	7	0.06	1.5	0
	3b								3c							
	Proline		Pro. ester		Piperidine		Blank		Proline		Pro. ester		Blank and Piperidine			
	F	H	F	H	F	H	F	H	F	H	F	H	F		H	
t_{1/2} (h)	0.1	0.1	0.2	0.5	0.1	0.1	0.2	0.1	-	-	-	-	-		-	
V_i (% per h)	70	11	47	5.5	68	5.5	35	5.5	-	-	-	-	-		-	

Table S1: Initial rate and half-live of benzylidene conjugates **2a-c** and **3a-c** as determined by ¹H NMR spectroscopy. (The quoted values are an average of three independent measurements). F= formation, H= hydrolysis; (-)= very fast.

Formation of compound 2a



Formation of compound 2b



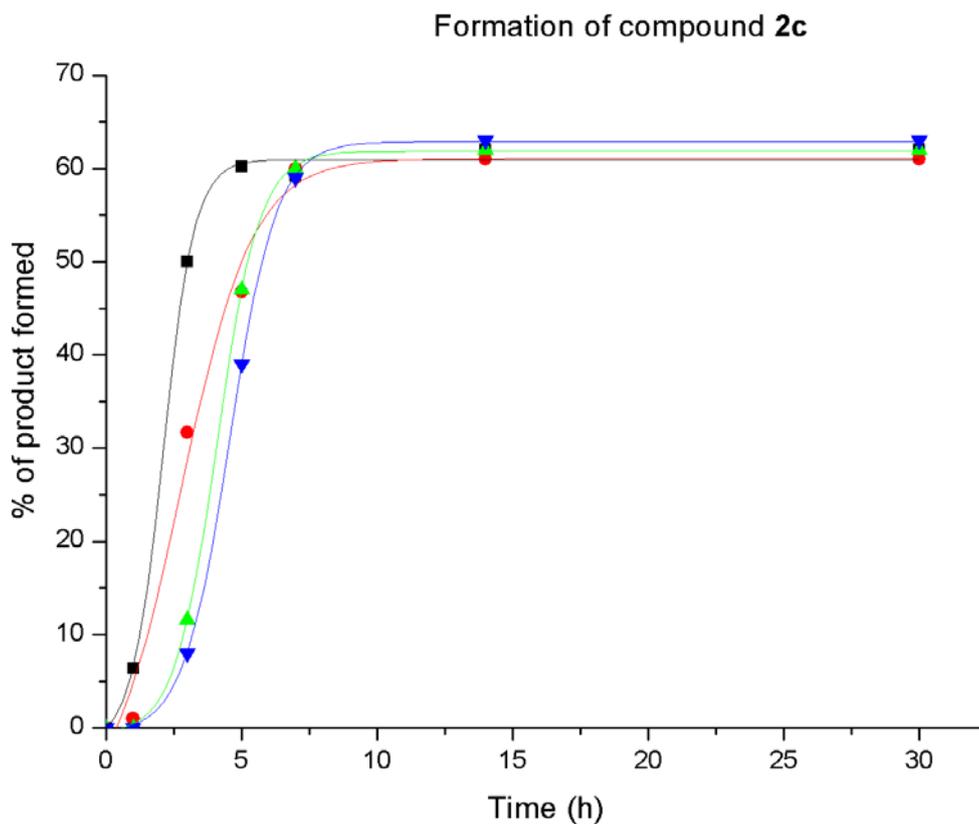
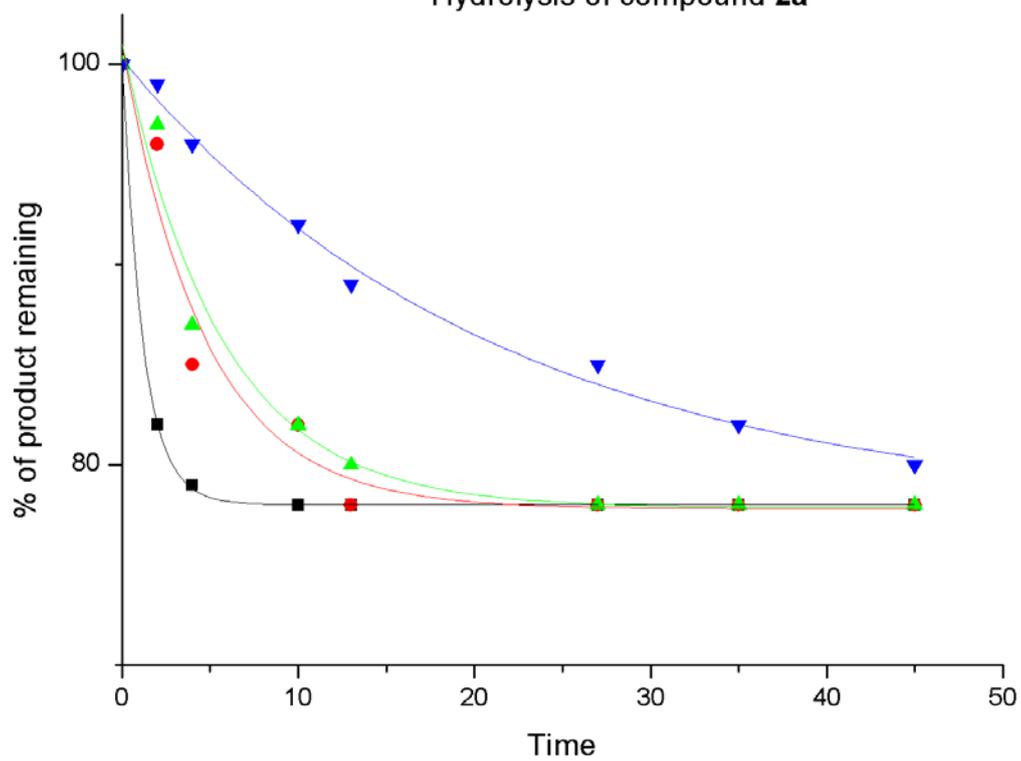
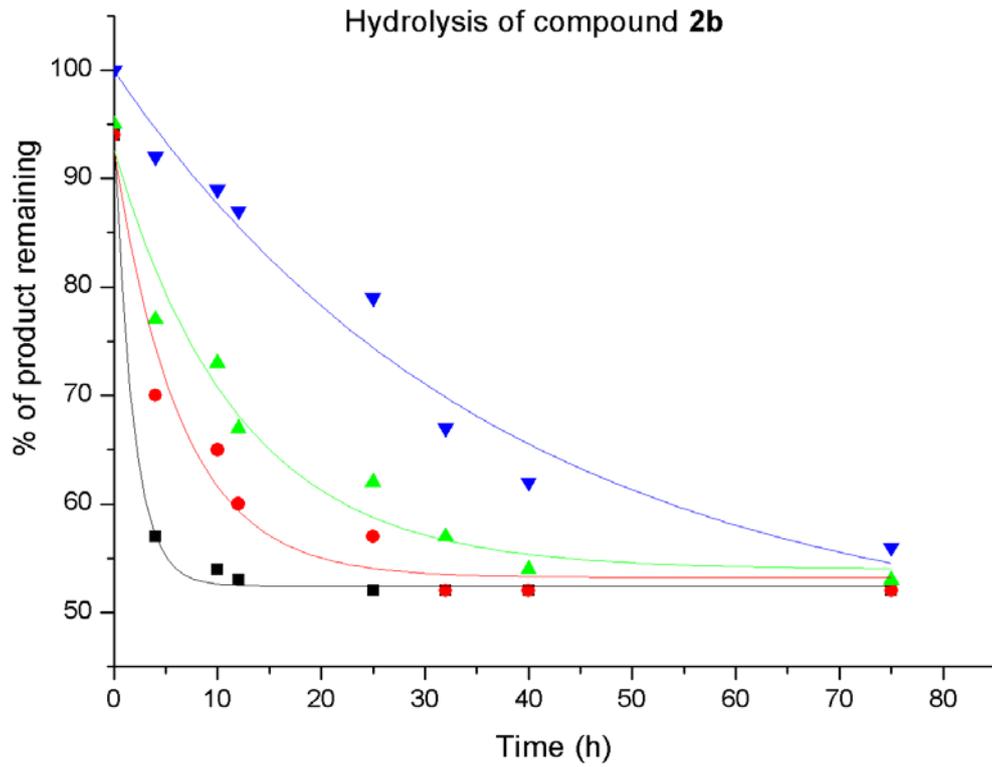


Figure S1: Synthesis of barbiturate benzylidene from barbituric acid and aldehydes **1a-c** in the presence or in the absence of base. Blue (no base); Black (L-proline); Red (L-proline methyl ester); Green (piperidine). Reactions were carried out in triplicate and the standard deviation was < 5%. The curves were drawn through the experimental points for clarity.

Hydrolysis of compound **2a**



Hydrolysis of compound **2b**



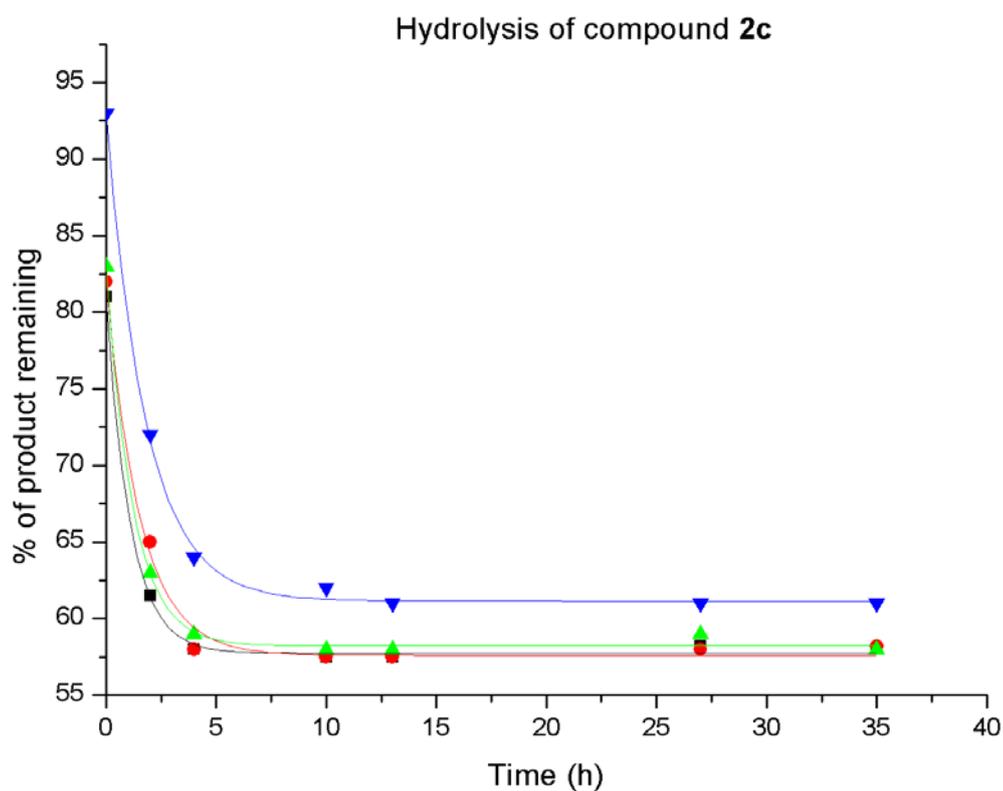
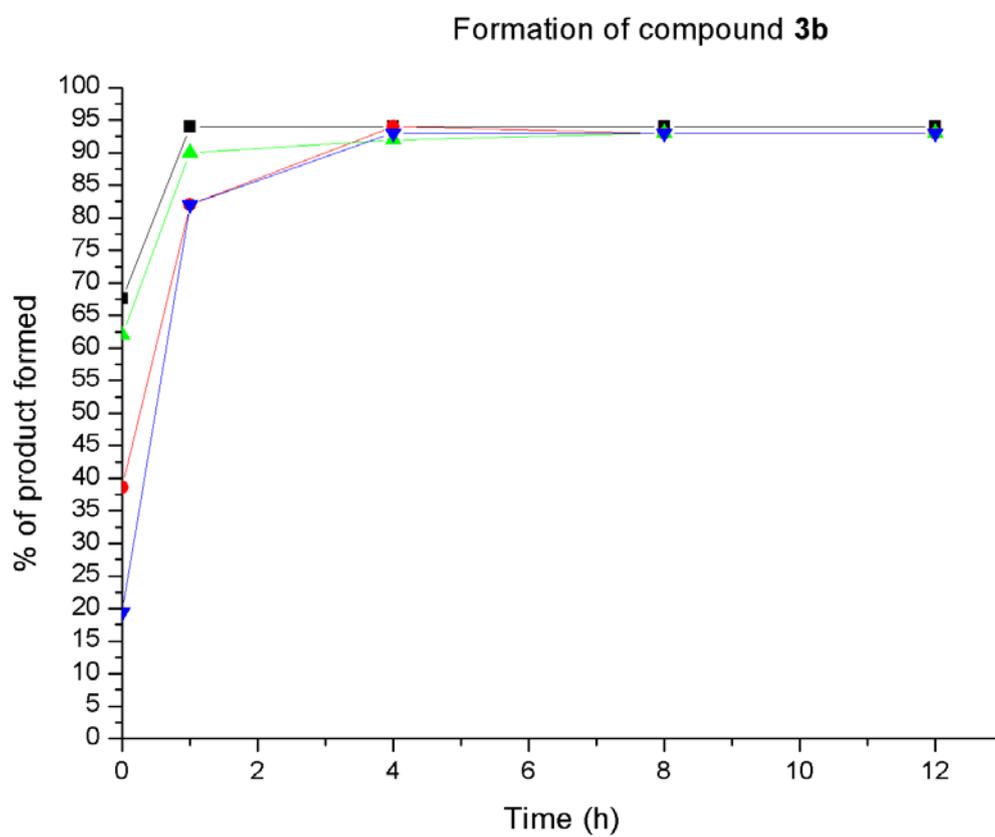
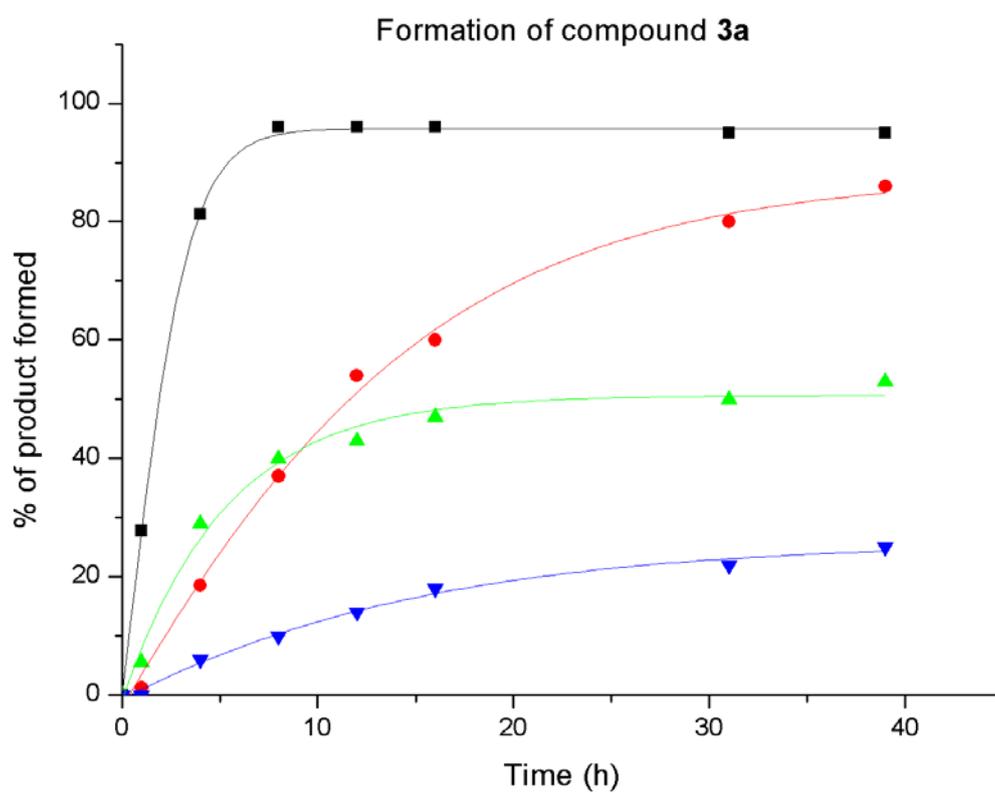


Figure S2: Hydrolysis of barbiturate benzylidene **2a-c** in the presence or in the absence of base. Blue (no base); Black (L-proline); Red (L-proline methyl ester); Green (piperidine). Reactions were carried out in triplicate and the standard deviation is < 5%. The curves were drawn through the experimental points for clarity.



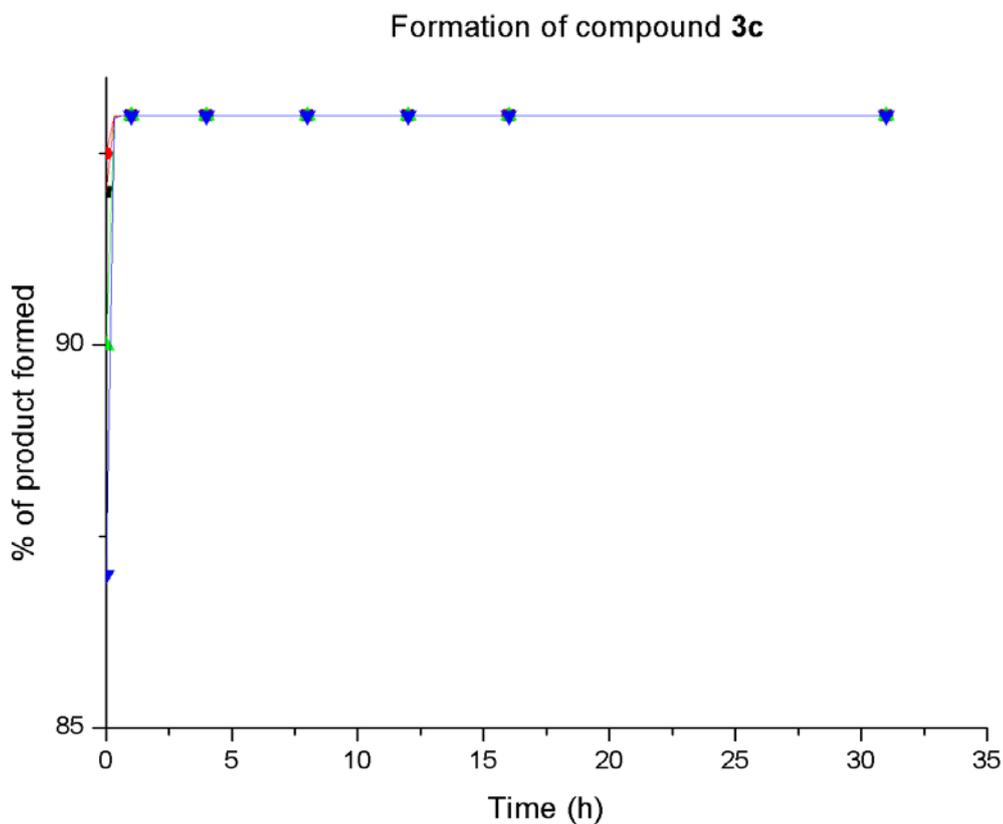
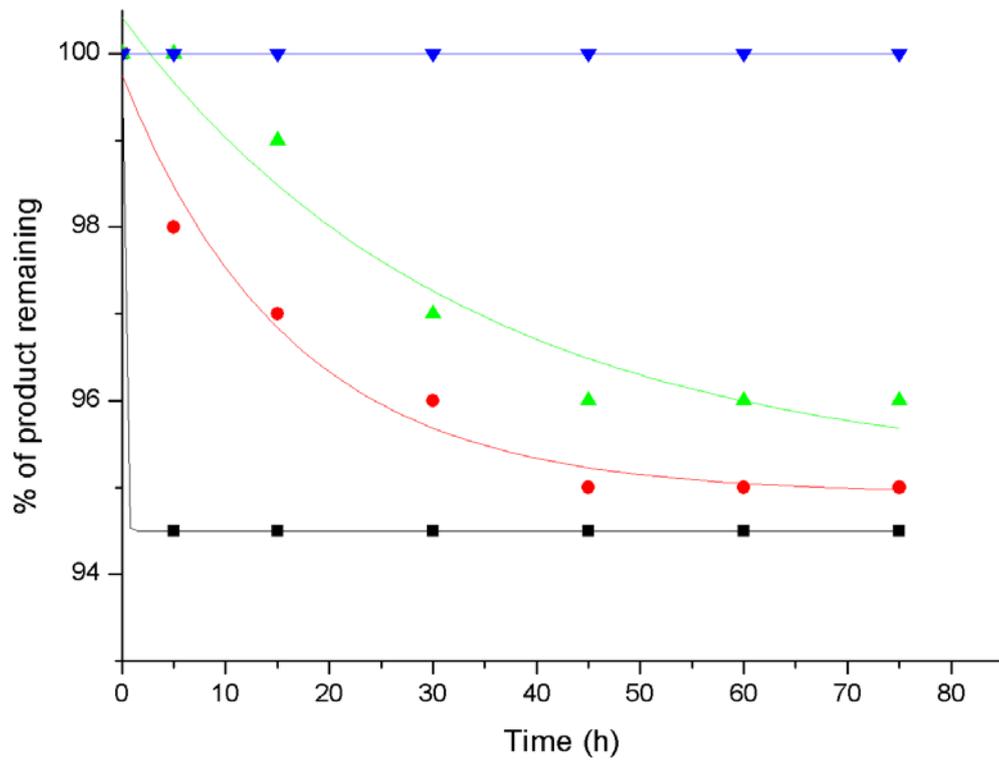
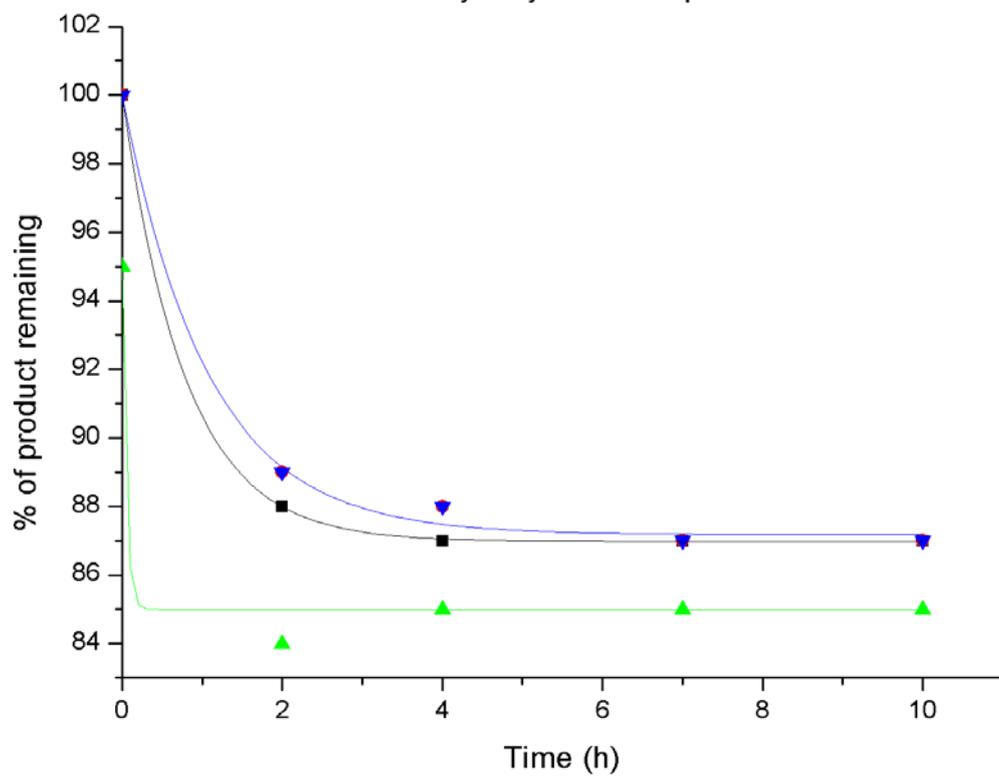


Figure S3: Synthesis of malonitrile benzylidene from malonitrile and aldehydes **1a-c** in the presence or in the absence of base. Blue (no base); Black (L-proline); Red (L-proline methyl ester); Green (piperidine). Reactions were carried out in triplicate and the standard deviation is < 3%. The curves were drawn through the experimental points for clarity.

Hydrolysis of compound 3a



Hydrolysis of compound 3b



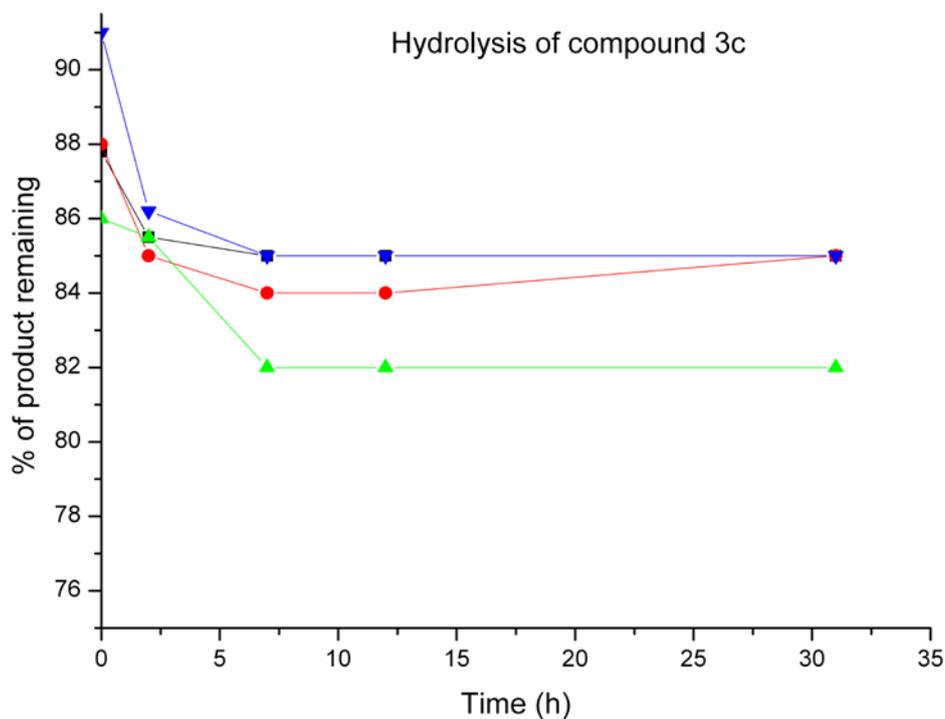
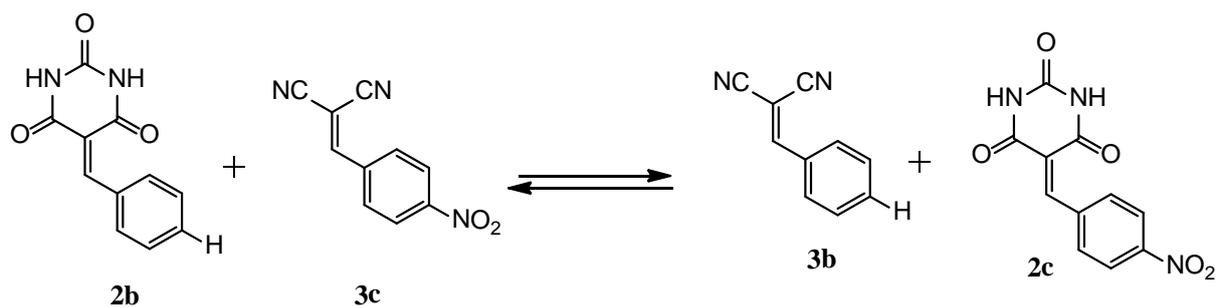
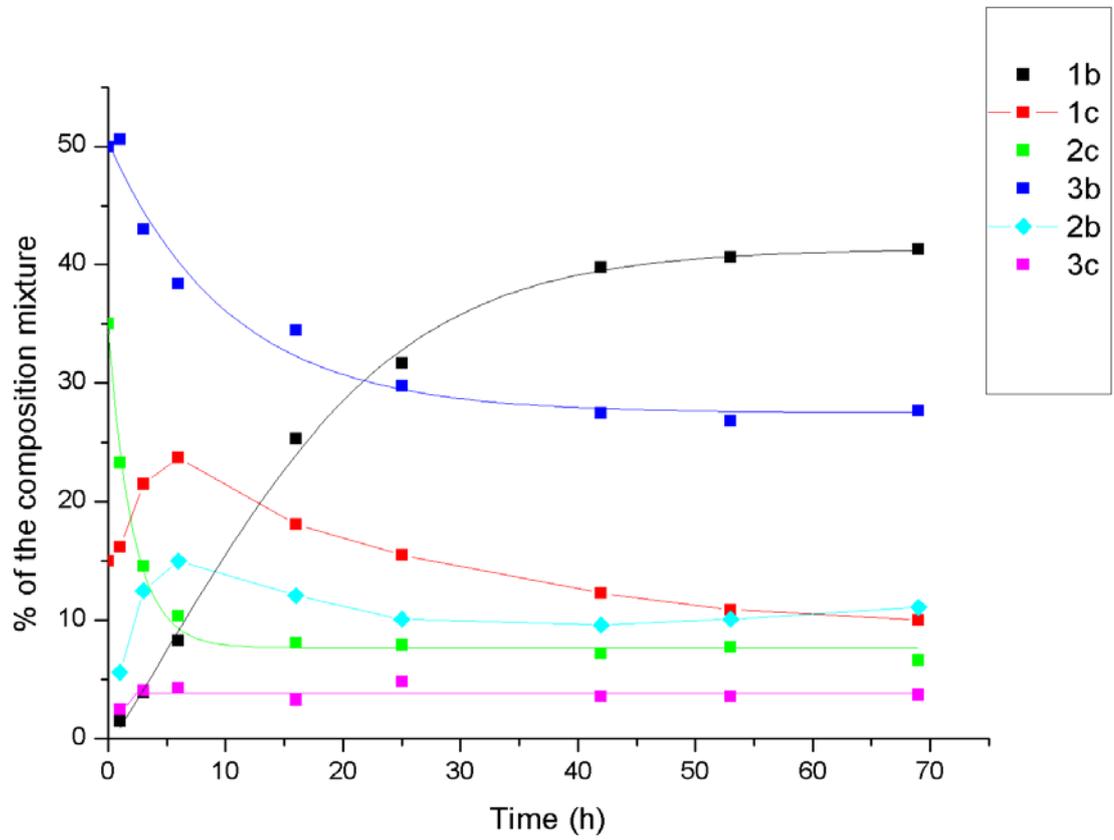


Figure S4: Hydrolysis of malononitrile benzylidenes **2a-c** in the presence or in the absence of base. Blue (no base); Black (L-proline); Red (L-proline methyl ester); Green (piperidine). Reactions were carried out in triplicate and the standard deviation is < 5%. The curves were drawn through the experimental points for clarity.

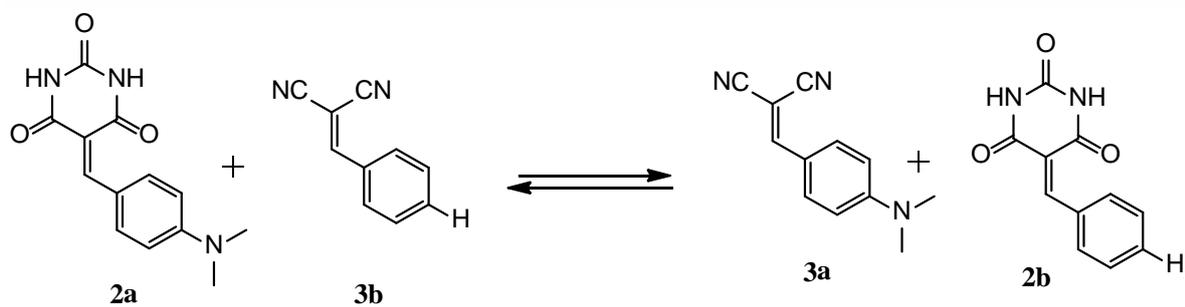
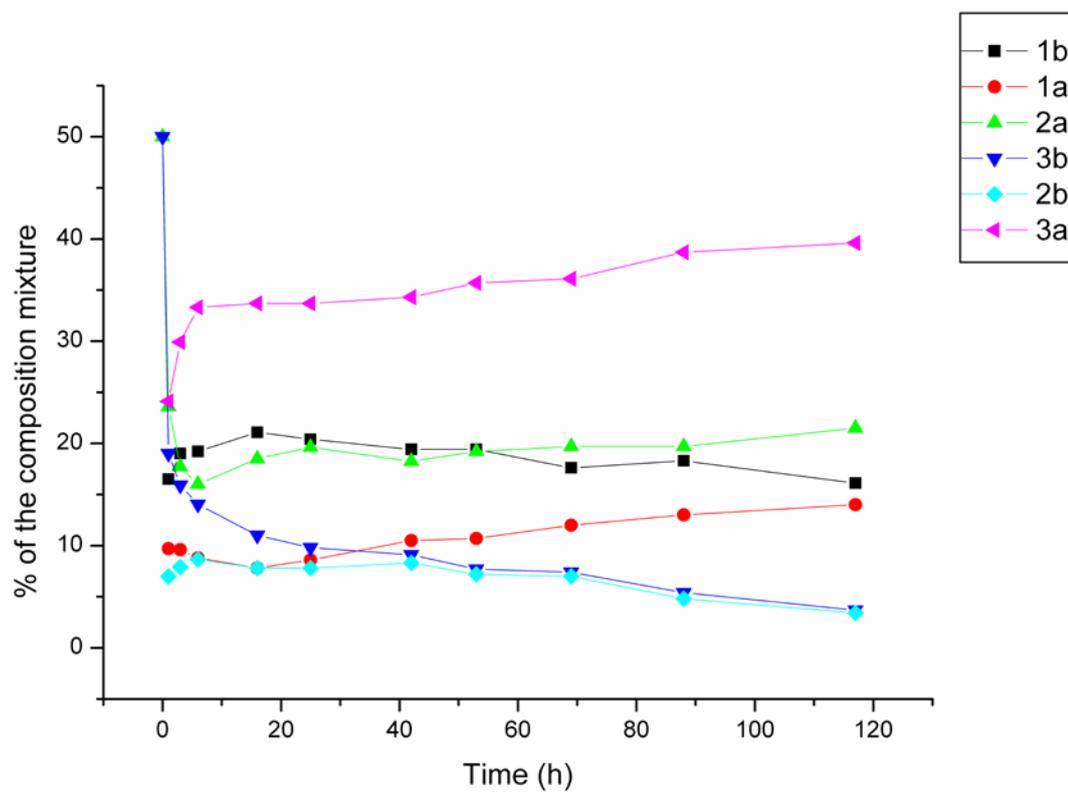
Cross-Metathesis reactions:



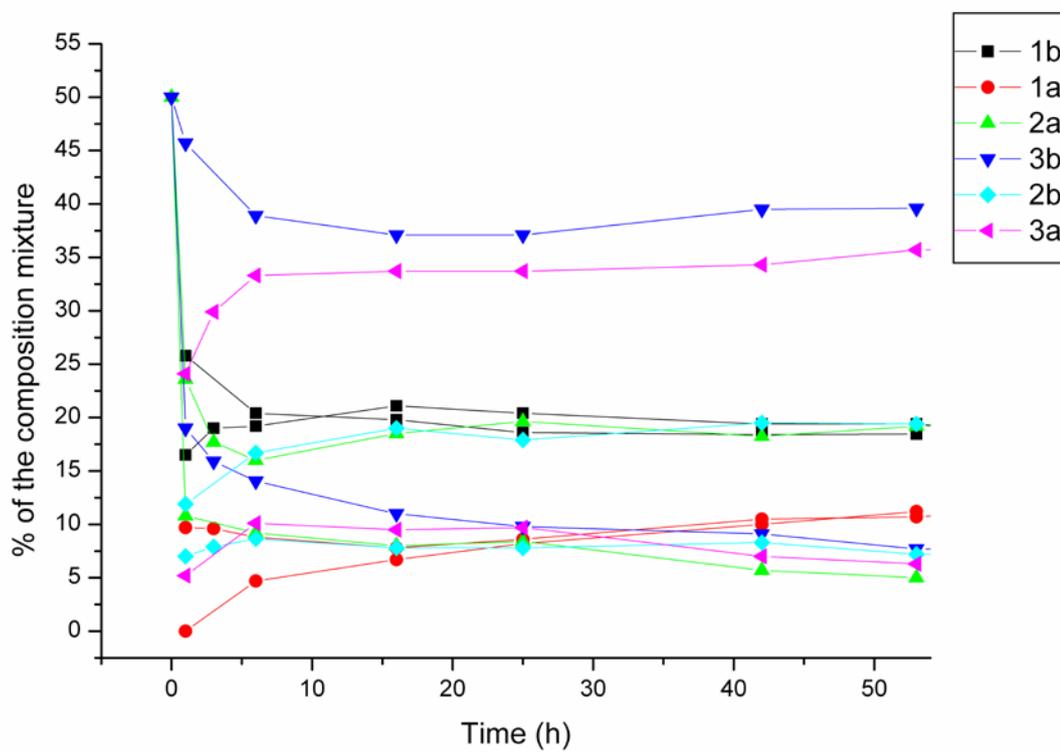
Forward reaction :



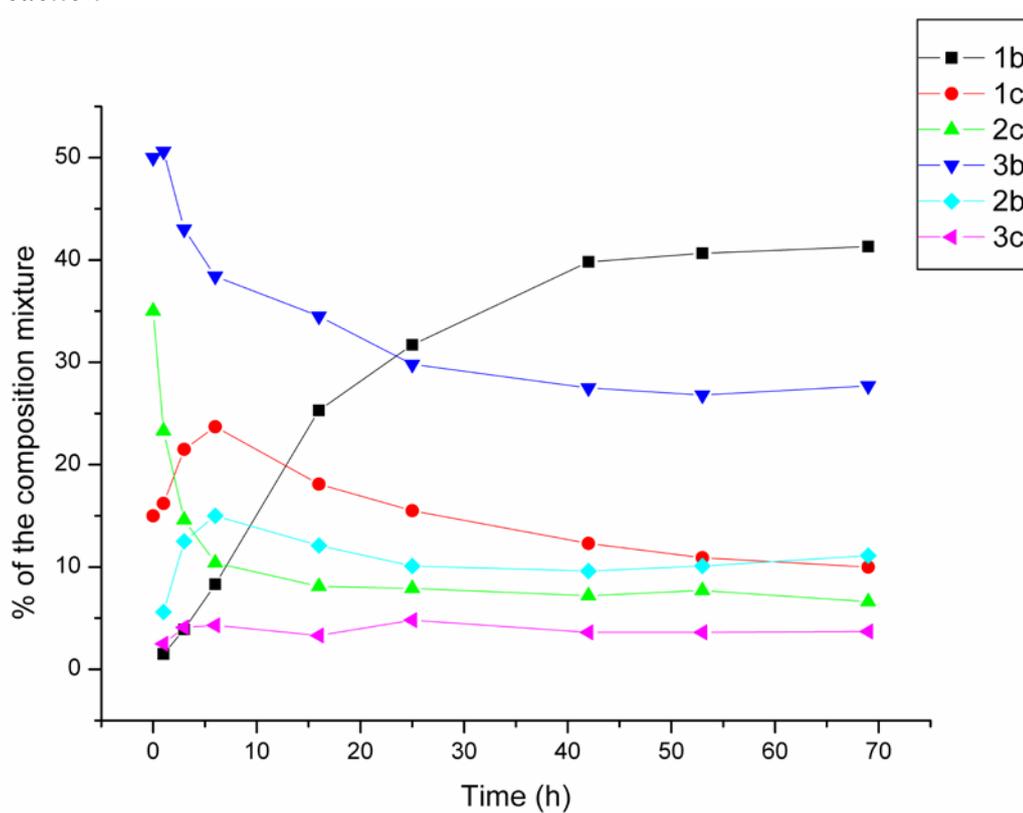
Reverse reaction :



Forward reaction



Reverse reaction



- [1] J.-M. Lehn, *Chem Comm.* 2007.
- [2] a) J.-M. Lehn, *Chem. Eur. J.* **1999**, *5*, 2455; b) A. V. Eliseev, J.-M. Lehn, *Science* **2001**, *291*, 2331; c) O. Ramström, J.-M. Lehn, *Nat. Rev. Drug Discovery* **2001**, *1*, 26.
- [3] a) G. R. L. Cousins, S.-A. Poulsen, J. K. M. Sanders, *Curr. Opin. Chem. Biol.* **2000**, *4*, 270; b) S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Curr. Opin. Chem. Biol.* **2002**, *6*, 321; c) B. Klekota, B. L. Miller, *TIBTECH* **1999**, *17*, 205; (d) V. Wittmann, *Nachr. Chem.* **2002**, *50*, 1364.
- [4] For dynamic materials, see: a) J.-M. Lehn, in *Supramolecular Science: Where It Is and Where It Is Going*; Ungaro, R., Dalcanale, E., Eds.; Kluwer: Dordrecht, The Netherlands, **1999**; p 287; b) J.-M. Lehn, *Polym. Int.* **2002**, *51*, 825; c) Polyacylhydrazones: W. G. Skene, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8270.
- [5] a) I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2106; b) J. R. Nitschke, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11970; c) R. Nguyen, I. Huc, *Chem. Commun.* **2003**, *8*, 942.
- [6] P. A. Brady, R. P. Bonar-Law, S. J. Rowan, C. J. Suckling, J. K. M. Sanders, *Chem. Commun.* **1996**, 319.
- [7] a) O. Ramström, J.-M. Lehn, *ChemBioChem* **2000**, *1*, 41; b) S. Otto, R. L. E. Furlan, J. K. M. Sanders, *J. Am. Chem. Soc.* **2000**, *122*, 12063; c) H. Hioki, W. C. Still, *J. Org. Chem.*, **1998**, *63*, 904.
- [8] P. G. Swann, R. A. Casanova, A. Desai, M. M. Fraunhoff, M. Urbancic, U. Slomczynska, A. J. Hopfinger, G. C. Le Breton, D. L. Venton, *Biopolymers* **1996**, *40*, 617.
- [9] I. Nakazawa, S. Suda, M. Masuda, M. Asai, T. Shimizu, *Chem. Commun.* **2000**, *10*, 881.
- [10] a) C. Brändli, T. R. Ward, *Helv. Chim. Acta* **1998**, *81*, 1616; (b) K. C. Nicolaou, R. Hughes, S. Y. Cho, N. Winssinger, C. Smethurst, H. Labischinski, R. Enderman, *Angew. Chem., Int. Ed.* **2000**, *39*, 3832.
- [11] P. J. Boul, P. Reutenauer, and J.-M. Lehn, *Org. Lett.* **2005**, *7*, 15.
- [12] B. List, *Angew. Chem. Int. Ed.* **2010**, *49*, 1730 – 1734.
- [13] a) E. Knoevenagel, *Ber. Dtsch. Chem. Ges.* **1896**, *29*, 172; b) E. Knoevenagel, *Ber. Dtsch. Chem. Ges.* **1898**, *31*, 738; c) E. Knoevenagel, *Ber. Dtsch. Chem. Ges.* **1898**, *31*, 2585; d) E. Knoevenagel, *Ber. Dtsch. Chem. Ges.* **1898**, *31*, 2596; for an excellent review, see e) L. F. Fieser, U. Beifuss in *Comprehensive Organic Synthesis* (Ed.: B. M. Trost), Pergamon, Oxford, 1991.
- [14] P. Wipf, S. G. Mahler, K. Okumura, *Org. Lett.* **2005**, *7*, 4483.
- [15] O. Kaumanns, R. Lucius, H. Mayr, *Chem. Eur. J.* **2008**, *14*, 9675.
- [16] M. K. Haldar, M. D. Scott, N. Sule, D. K. Srivastava and S. Mallik, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2373-2376.
- [17] S. Wanga, Z. Rena, W. Cao, W. Tonga, *Synthetic Communication*, **2001**, *31*, 673-677.

CHAPTER THREE

IMPROVING THE PREDICTION OF MAXIMAL ABSORPTION WAVELENGTHS OF CYANINE DYES: A COMPUTATIONAL STUDY USING TD-DFT

I. Introduction to TD-DFT

A major difference between quantum and classical mechanics is that classical mechanics is deterministic while quantum mechanics is probabilistic (more correctly, quantum mechanics is also deterministic, but the interpretation is probabilistic). Deterministic means that Newton's equation can be integrated over time (forward or backward) and can predict where the particles are at a certain time. This, for example, allows prediction of where and when solar eclipses will occur many thousands of years in advance, with an accuracy of meters and seconds. Quantum mechanics, on the other hand, only allows calculation of the probability of a particle being at a certain place at a certain time. The probability function is given as the square of a wave function, $P(\mathbf{r},t) = \Psi^2(\mathbf{r},t)$, where the wave function Ψ is obtained by solving either the Schrödinger (non-relativistic) or Dirac (relativistic) equation.

Schrödinger equation is an equation that describes how the quantum state of a physical system changes in time. It is as central to quantum mechanics as Newton's laws are to classical mechanics. In the standard interpretation of quantum mechanics, the quantum state, also called a wavefunction or state vector, is the most complete description that can be given to a physical system. Solutions to Schrödinger's equation describe not only molecular, atomic and subatomic systems, but also macroscopic systems, possibly even the whole universe. The equation is named after Erwin Schrödinger, who constructed it in 1926¹.

The general form of the Schrodinger equation is described as follows:

$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi$$

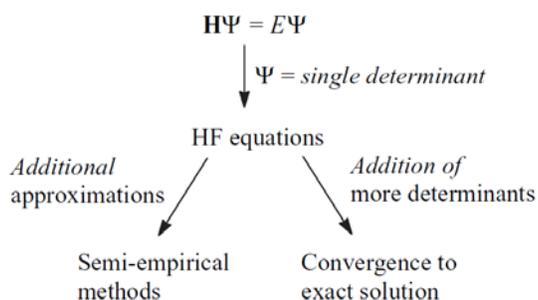
Ψ is the wave function; the probability amplitude for different configurations of the system at different times.

$i\hbar \frac{\partial}{\partial t}$ is the energy operator (here, i is the imaginary unit and \hbar is the reduced Planck constant),

\hat{H} is the Hamiltonian operator.

If solutions are generated without reference to experimental data, the methods are usually called *ab initio* (Latin: “from the beginning”), in contrast to semi-empirical models. An essential part of solving the Schrödinger equation is the Born–Oppenheimer approximation, where the coupling between the nuclei and electronic motion is neglected.

Within electronic structure theory, only the latter has an acceptable accuracy, and is called Hartree–Fock (HF) theory. In the HF model, each electron is described by an orbital and the total wave function is given as a product of orbitals.



Semi-empirical methods are derived from the HF model by neglecting all integrals involving more than two nuclei in the construction of the Fock matrix. Since the HF model by itself is only capable of limited accuracy, such approximations will by themselves lead to a poor model. The success of semi-empirical methods relies on turning the remaining integrals into parameters, and fitting these to experimental data, especially molecular energies and geometries. Such methods are computationally much more efficient than the *ab initio* HF method, but are limited to systems for which parameters exist. HF theory only accounts for the average electron–electron interactions, and consequently neglects the correlation between electrons. Methods that include electron correlation require a multi-determinant wave function, since HF is the best single determinant wave function. Multi-determinant methods are computationally much more involved than the HF model but can generate results that systematically approach the exact solution of the Schrödinger equation.

Density Functional Theory (DFT) introduced by Kohn-Sham² is an improvement of the HF theory, where the many-body effect of electron correlation is modeled by a function of the electron density. In other words, the ground state electronic energy is determined completely by the electron density. DFT is, analogously to HF, an

independent-particle model, and is comparable to HF computationally, but provides significantly better results. In its most commonly applied approximations to the electronic exchange and correlation, i.e. in the local density approximation (LDA) and the semi local generalized gradient approximation (GGA); DFT is highly efficient and surprisingly accurate. The main disadvantage of DFT is that there is no systematic approach to improving the results towards the exact solution.

One of the approximations inherent in essentially all ab initio methods is the introduction of a basis set. Expanding an unknown function, such as a molecular orbital, in a set of known functions is not an approximation if the basis set is complete. The smaller the basis set, the poorer the representation. The type of basis functions used also influence the accuracy.

Basis-sets: The basic-sets used in computational chemistry were designed by Pople *et al.* and are of the split valence type, with the k in front of the dash indicating how many Gaussian-type orbitals are used for representing the core orbitals. In other words, they are used to create molecular orbitals. All basic-set are represented as k-nlmG basis sets, the letter G indicating that a Gaussian calculation was performed. There exist hundreds of basis sets composed of Gaussian-type orbitals (GTOs). The “nlm” term after the dash indicates both how many functions the valence orbitals are split into, and how many PGTOs are used for their representation. All quantum calculations are performed using basis-sets.

Two values (nl) indicate a split valence, while three values (nlm) indicate a triple split valence. The values before the G (for Gaussian) indicate the s- and p-functions in the basis; the polarization functions are placed after the G. These types of basis sets have the further restriction that the same exponent is used for both the s and p-functions in the valence. This increases the computational efficiency, but of course decreases the flexibility of the basis set. The exponents and contraction coefficients have been optimized by variational procedures at the HF level for atoms.

3-21G: This is a split valence basis, where the core orbitals are a contraction of three Gaussian-type orbitals, the inner part of the valence orbitals is a contraction of two

PGTOs and the outer part of the valence is represented by one PGTO. The designation of the carbon 3-21G basis is $(6s3p) \rightarrow [3s2p]$.

6-31G: This is also a split valence basis, where the core orbitals are a contraction of six PGTOs, the inner part of the valence orbitals is a contraction of three PGTOs and the outer part of the valence is represented by one PGTO. The designation of the carbon 6-31G basis is $(10s4p) \rightarrow [3s2p]$. In terms of contracted basis functions it contains the same number as 3-21G, but the representation of each function is better since more PGTOs are used.

6-311G: This is a triple split valence basis, where the core orbitals are a contraction of six PGTOs and the valence split into three functions, represented by three, one and one PGTOs, respectively, i.e. $(11s5p) \rightarrow [4s3p]$.

To each of these basis sets can be added diffuse and/or polarization functions. Diffuse functions are normally s- and p-functions and consequently go before the G. They are denoted by + or ++, with the first + indicating one set of diffuse s- and p-functions on heavy atoms, and the second + indicating that a diffuse s-function is added also to hydrogen. The argument for only adding diffuse functions on non-hydrogen atoms is the same as for only adding polarization functions on non-hydrogens. Polarization functions are indicated after the G, with a separate designation for heavy atoms and hydrogen. The 6-31+G(d) is a split valence basis with one set of diffuse sp-functions on heavy atoms only and a single d-type polarization function on heavy atoms. A 6-311++G(2df,2pd) is similarly a triple split valence with additional diffuse sp-functions, two d-functions and one f-function on heavy atoms, and diffuse s- and two p- and one d-functions on hydrogen. The largest standard Pople style basis set is 6-311++G(3df, 3pd). These types of basis set have been derived for hydrogen and the first row elements, and some of the basis sets have also been derived for second and higher row elements.

All exchange and correlations effects of the many-electron system are incorporated into the so-called exchange-correlation (xc) energy functional, which is not known exactly and must be approximated in practice. Hybrid functionals are an approximation of the exchange-correlation mixing the Hartree-Fock theory with the DFT. The most famous Hybrid functionals are PBE0, HSE03, and B3LYP³. HF/DFT hybrid

functionals are known to present an improved description of the thermochemistry of molecular systems⁴. Indeed, since several decades, a detailed evaluation of the description of structural, thermochemical, and electronic properties of extended systems using the PBE0, HSE03, and B3LYP⁵.

The exact exchange energy functional is expressed in terms of the Kohn-Sham orbitals rather than the density, so is termed an implicit density functional⁶. The parameters determining the weight of each individual functional are typically specified by fitting the functional's predictions to experimental or accurately calculated thermochemical data. The B3LYP is the most successful method in the study of thermochemistry⁷. The B3LYP hybrid was originally developed to improve the ground state of small compounds⁸. The Perdew-Wang gradient-corrected correlation energy, which was used in the original work of Becke⁹, is replaced by Lee-Yang-Parr correlation energy. The popular B3LYP (Becke, three-parameters, Lee-Yang-Parr¹⁰) exchange-correlation functional is defined as:

$$E_{xc}^{\text{B3LYP}} = E_{xc}^{\text{LDA}} + a_0(E_x^{\text{HF}} - E_x^{\text{LDA}}) + a_x(E_x^{\text{GGA}} - E_x^{\text{LDA}}) + a_c(E_c^{\text{GGA}} - E_c^{\text{LDA}}),$$

where $a_0=0.2$, $a_x=0.72$ and $a_c=0.81$, are the three empirical parameters¹¹. E_x^{GGA} and E_c^{GGA} are generalized gradient approximations: the Becke 88 exchange functional¹² and the correlation functional of Lee, Yang and Parr, and E_c^{LDA} is the VWN local-density approximation to the correlation functional.

II. Using TD-DFT to predict the maximum absorption wavelength of cyanine dyes.

In this chapter, we present a novel approach to improve significantly the prediction of maximal absorption wavelengths of cyanine dyes from time-dependent density functional theory (TD-DFT) calculations. In recent years, TD-DFT has been applied increasingly to calculate spectroscopic properties of dyes and important

methodological improvements have been achieved, for example through the inclusion of solvent effects. Herein, we take the theory to the next level by introducing an ad hoc correction which accounts for vibronic contributions. Its application to cyanine dyes shows that the correction term decreases the average error about three fold, which permits a more general application of the TD-DFT method. It can now be applied not only to symmetrical and asymmetrical trimethine and pentamethine cyanine dyes but also to unusual imino analogues of cyanine dyes. This work was carried out under the supervision of Dr Martin Spichy (ISIS, Strasbourg) and has recently been submitted for publication in *Dyes and Pigments*.

REFERENCES

- ¹ Schrödinger, E. *Phys. Rev.* **1926**, *28*, 1049.
- ² W. Kohn and L. J. Sham *Phys. Rev.* **1965**, *140*, 1133.
- ³ (a) Paier J., Marsman M., Hummer K., Kresse G., Gerber I.C. and Angyan J.G. *J. Chem. Phys.* **2006**, *124*, 154709; (b) Paier J., Marsman M. and Kresse G. *J. Chem. Phys.* **2007**, *127*, 24103.
- ⁴ (a) Curtiss L.A., Raghavachari K., Redfern P.C. and Pople J.A. *J. Phys. Chem.* **1997**, *106*, 1063; (b) Paier J., Hirschl R., Marsman M. and Kresse G. *J. Chem. Phys.* **2005**, *122*, 234102
- ⁵ Heyd J. and Scuseria G.E. *J. Chem. Phys.* **2004**, *121*, 1187.
- ⁶ http://en.wikipedia.org/wiki/Hybrid_functional
- ⁷ Chen Z.Y., Yang J.L. *Front. Phys. China* **2006**, *3*, 339.
- ⁸ Becke A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- ⁹ See reference 8
- ¹⁰ (a) Kim and K. D. Jordan, *J. Phys. Chem.* **1994**, *98*, 10089; (b) P.J. Stephens, F.J. Devlin, C.F. Chabalowski and M.J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
- ¹¹ Becke, A.D. *J. Chem. Phys.*, **1993**, *98*, 5648.
- ¹² Becke, A.D. *Phys. Rev. A*, **1988**, *38*, 3098.

A conceptually improved TD-DFT approach for predicting the maximum absorption wavelength of cyanine dyes

Kamel Meguellati[§], Sylvain Ladame[§], and Martin Spichy^{§,§,*}

[§]Institut de Science et d'Ingénierie Supramoléculaires, Université de Strasbourg, 8 allée Gaspard Monge, B.P. 70028, 67083 Strasbourg Cedex;

[§]Current address: Laboratoire de Biologie Moléculaire de la Cellule, Différenciation et Cycle Cellulaire, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07.

*To whom correspondence should be addressed, Email: martin.spichy@ens-lyon.fr.

ABSTRACT. Cyanine dyes have found valuable applications in modern bioresearch because of their biocompatibility, high molar absorptivity and moderate fluorescence quantum yield. Of particular interest for sensing and labeling applications is the fact they can cover a very large spectral range (from blue to Infra-Red). To design and select the most appropriate dyes for a given application the computational prediction of the absorption wavelength (prior to the costly chemical synthesis) serves as a valuable guidance. However, predicting absorption and emission wavelengths of such compounds remains a challenging task. Herein, we report a fast and highly accurate computational approach which allows the prediction of the maximum absorption wavelength for a wide range of cyanine dyes, including symmetrical and unsymmetrical, trimethine and pentamethine cyanine dyes but also unusual imino-based analogues. In addition to the vertical excitation energy (calculated from time-dependent density functional theory), the approach makes use of a novel correction term that is based on the ground-state zero-point vibrational energy (ZPVE). The correction term is statistically significant (*F*-test), and it reduces the average error and maximal error of the prediction by a factor of two. We anticipate that the concept of including the ZPVE into the calculation of the maximum absorption wavelength can be used also for other families of dyes to improve their predictability.

KEYWORDS: quantum chemical calculations, time-dependent density functional theory, linear scaling approach, zero-point vibrational energy, empirical correction.

1. Introduction

Dyes are molecules that display light absorption in the long-wavelength region and thereby give rise to color perception [1]. Of particular interest in modern bioresearch are dyes capable of emitting a fluorescent signal upon excitation with a single (or multiple) photon(s). Fluorescence arises when the relaxation corresponding to the transition of an electron from the first singlet excited state (S_1) to the singlet electronic ground state (S_0) occurs via photon emission. Since the discovery of the first natural fluorophore in 1845 [2], numerous fluorescent small molecules have been engineered that cover a very broad spectral range (from blue to infra-red) [3]. Among them, cyanine dyes have found valuable applications (*e.g.* for optical data storage, proteomic, labeling of biomolecules) [4]. They generically consist of a conjugated system based on a polymethine chain linking two nitrogen-containing heterocycles (*e.g.* indoles, benzothiazoles) [5].

(Insert Figure 1 here)

We have recently introduced a new class of dyes differing from the well known cyanine dyes by one or two C→N substitutions within the polymethine chain (Figure 1) [6,7]. While the potential of polymethine cyanine dyes for labeling purposes is now well-established [8], the possible use of their imino analogues for either sensing malondialdehyde (MDA) or alkylating DNA/RNA nucleobases has also been recently demonstrated [7].

Depending on the process or metabolite that needs to be sensed, fluorescent molecules absorbing and emitting in specific regions of the spectrum are necessary. Therefore, understanding and tuning the spectroscopic properties of cyanine dyes remains of vital interest. The prediction of spectroscopic properties from numerical calculations provides valuable input in this regard; *e.g.*, the computational

appraisal of the absorption wavelength prior to the costly chemical synthesis may serve as guidance for the design and selection of the most appropriate dyes for a given application.

Absorption wavelengths of dyes have been calculated by various approaches ranging from simplistic π -electron models and semi-empirical all-valence-electron methods to computationally intensive *ab initio* model chemistries (for a concise overview, see reference [9]). In recent years time-dependent density functional theory (TD-DFT)[10] has increasingly been used [9, 11-20] to fill the gap between the semi-empirical and high-level *ab initio* methods, *i.e.*, TD-DFT is a first-principles method that overcomes short-comings of the former methods (*e.g.*, electron correlation) and yet it is still applicable to much larger systems (about 100 second-row atoms) than the latter [9].

Despite considerable efforts in TD-DFT, the accurate prediction of maximum absorption wavelengths from computations remains, however, a challenging task. There are several reasons for this. Beside the Hamiltonian approximations (*e.g.*, exchange correlation functional, basis set, treatment of solvent) the calculations usually differ also conceptually from the experiment, *i.e.*, a discrete value for the vertical excitation energy of a ground-state minimum-energy structure is calculated (ΔE_v), whereas the reported value of the experiment corresponds to the energy value of the absorption maximum (ΔE_{\max}) which depends on the vibronic structure of the dye; see also discussion by Champagne *et al.* [14].

Herein, we report the development of an optimized approach for predicting with high accuracy the maximum absorption wavelength of polymethine cyanine dyes as well as that of their imino analogues. As a first test, we use a linear scaling approach[14] that estimates the maximum absorption wavelength ΔE_{\max} solely from the TD-DFT-calculated ΔE_v . We then propose a conceptual refinement of the standard approach by introducing an empirical correction term that is based on the zero-point vibrational energy. The statistical significance of the ZPVE-based correction is verified (*F*-test), and the portability of this methodological development to other families of dyes is discussed.

2. Material and methods

Compounds **1-14** were synthesized following experimental procedures previously reported by us [6,7] and others [21]. Ultraviolet absorption spectra of analytically pure samples were recorded on a V-670 UV-Visible spectrophotometer from JASCO in a 10 mm pathlength cuvette.

3. Theory and calculations

We build on a linear scaling approach [14]: ΔE_{\max} , is obtained from the TD-DFT-based ΔE_v through an empirical equation:

$$\Delta E_{\max} = \alpha \Delta E_v + \beta, \quad (1)$$

where α and β are two parameters that are fitted from a series of dyes with known experimental maximum absorption wavelengths. We used the cyanine dyes of Table 1 for which ΔE_{\max} has been measured under the same experimental conditions, *i.e.*, same solvent (DMSO), counter ion, and temperature [6,7].

Variants of Eq. 1 are known where additional terms for solvent effects are included [15,16], or where the absorption wavelengths of maximal intensity is calculated based on multiple QSAR descriptors [22]. Here we used the following extension of Eq. 1:

$$\Delta E_{\max} = \alpha \Delta E_v + \beta + \gamma(\text{ZPVE}/n_{\text{DOF}} - \delta)^2, \quad (2)$$

with ZPVE the zero-point vibrational energy (in the harmonic approximation) [23], n_{DOF} the number of vibrational degrees of freedom ($3N-6$ for non-linear molecules, with N the number of atoms), and γ , δ are two additional parameters to be fitted.

Several methods of calculating ΔE_v were tested to fit Eq.1, the “best” method (highest R^2 and Q^2 , see below) was then used to fit Eq. 2. The minimum-energy structure was calculated on the B3LYP/6-31G* level of theory either in the gas phase or in solution (using the PCM solvation model [24]); in both cases a tight minimization was performed and the localization of the stationary points was verified with a frequency calculation that yielded ZPVE. Two different functionals, B3LYP [25,26] and PBE0 [27] (also named PBE1PBE [28]), in combination with three different basis sets, 6-31G*, 6-311G*, and 6-311+G**, were then used to determine the vertical excitation energy ΔE_v by a single-point TD-DFT calculation (with and without solvation model) [29]. All calculations were carried out with the program Gaussian09 [28].

For each prediction model we calculated the squared correlation coefficient, R^2 , between the calculated and experimental ΔE_{\max} values, as well as the leave-one-out cross-validated R^2 , named Q^2 ; the former parameter is a measure of the goodness-of-fit, the latter of the goodness-of-prediction [30]. When two models are compared that are based on different equations (with a different number of parameters, such as Eq. 1, Eq. 2 and variants of the latter; see section Results and Discussion), an increase in R^2 when going from the smaller, nested equation to the larger equation does not always indicate a statistically significant improvement of the model. We therefore performed a likelihood-ratio test (F -test) [31]. Under the null hypothesis that the larger model does not fit the data better than the smaller, nested model, the value

$$F = [(R_1^2 - R_s^2)/(1 - R_1^2)] \times [(p_1 - p_s)/(N - p_1)], \quad (3)$$

will have a F -distribution with $p_1 - p_s$ numerator and $N - p_1$ denominator degrees of freedom. R_1^2 is the correlation coefficient for the model with the larger number of parameters (p_1); R_s^2 and p_s are the respective quantities of model with the simpler, nested equation. N is the number of data points available for the fit. To reject the null hypothesis, the value of Eq. 3 needs to be greater than the value of

the F -distribution for some desired false-rejection probability (P value). For the comparison of the fits with Eq. 2 and Eq. 1, for example, the value of the F -distribution (two numerator and ten denominator degrees of freedom) is 4.10 for the critical P value of 0.05.

(Insert Table 1 here)

4. Results and discussion

It has been noted previously that the use of a solvation model can improve significantly the agreement between experimental ΔE_{\max} and predicted values from Eq. 1 [14]. When a pure gas phase method, *i.e.*, both geometry optimization and single-point calculation are carried out in gas phase (see values in parentheses of Table 2), is compared with a method where the latter step is carried out with the PCM solvation model, the correlation coefficient R^2 and the leave-one-out cross-validated correlation coefficient Q^2 increase by about 20 %. When both steps are carried out in solution, R^2 and Q^2 further increase; the effect is, however, smaller. Improvement is also possible when the functional B3LYP is replaced by PBE0 [15]. Furthermore, increasing the basis set from 6-31G* to 6-311G* increases R^2 and Q^2 , but 6-311G* and 6-311++G** yield essentially identical results. The highest goodness-of-fit and goodness-of-prediction are obtained for the method PBE0/6-311++G**(PCM)//B3LYP/6-31G*(PCM) (Figure 2a). With this model the error in the prediction of ΔE_{\max} ranges from -0.07 to $+0.12$ eV (-28 to $+21$ nm), the root-means-square error is 0.06 eV (15 nm).

(Insert Table 2 & Figure 2 here)

The quality of the prediction with Eq. 1 (correlation coefficient $R^2=0.968$, maximal error = 0.12 eV) is notably worse than previously reported for analogous polymethine cyanine dyes ($R^2=0.988$, maximal error = 0.04 eV) [14]. In this study, however, we used a much more heterogeneous data set for the fitting of Eq. 1 including mono- and di-imino dyes. Especially di-imino derivatives show large

deviations between prediction and experiment as indicated by the arrows in Figure 2a. Thus for this heterogeneous set the simple linear scaling approach of Eq. 1 fails to reproduce the experimental order of the ΔE_{\max} values.

We note significant differences in the zero-point vibrational energy (ZPVE) among the studied dyes. The total ZPVE depends strongly on the number of atoms, so that its comparison between molecules with different number of atoms is not very informative. Normalization (*i.e.*, division) by the number of degrees of freedom (n_{DOF}) allows, however, the identification of changes in the distribution of normal mode frequencies. Such changes may influence the “sharp” vibrational coupling of cyanine dyes with the solvent [32]; thereby alternate the strength of the solute-solvent interaction (*e.g.*, change the solvation contribution to the free energy values in the Marcus model [33]) and slightly shift the absorption maximum. The vertical excitation energy ΔE_v (calculated from an energy-minimized structure with an implicit solvation model) does not capture such solvent coupling effects.

(Insert Figure 3 here)

The ZPVE is also of relevance in terms of vibronic contributions (when calculating ΔE_{\max} from ΔE_v). For cyanine dyes the band with the highest intensity corresponds to a transition between the vibrational ground states of S_0 ($v''=0$) and S_1 ($v'=0$) [34]. Consider for the moment a (0,0)-transition (Figure 3) of a hypothetical molecular model system with n_{DOF} identical vibrations of frequency ν and $\text{ZPVE} = \frac{1}{2} n_{\text{DOF}} h\nu$. The frequency does not change upon excitation, but the equilibrium geometry of the excited state, r_e' , differs slightly from that of the ground-state, r_e'' , where the difference in reduced (dimensionless) coordinates is denoted by d . Since it is $\Delta E_{\max} = D_0$ (see Figure 3) we find $\Delta E_{\max} = \Delta E_v - n_{\text{DOF}} \frac{1}{2} h\nu d^2 = \Delta E_v - \text{ZPVE} d^2$, where the last term, $\text{ZPVE} d^2$, is the vibrational reorganization energy [33]. For systems with large n_{DOF} the value d should be smaller than for systems with small n_{DOF} because for large molecules the relative geometric change per degree of freedom is smaller upon excitation than for small

molecules. If we assume that the total geometric change in reduced coordinates is constant among dyes of the same family (*i.e.*, $n_{\text{DOF}} d^2 = \text{const}$), we have

$$\Delta E_{\text{max}} - \Delta E_{\text{v}} \propto \text{ZPVE}/n_{\text{DOF}}, \quad (4)$$

i.e., the difference between the maximum absorption energy and the vertical excitation energy is proportional to $\text{ZPVE}/n_{\text{DOF}}$. Eq. 4 is, of course, a drastically oversimplified representation for realistic molecular systems. But it may reflect to some extent relevant information when comparing different members of dyes from the same family (*e.g.*, cyanine dyes) relative to each other. Using the calculated ZPVE values of the studied dyes the hypothetical frequency $\nu = (2/h) \text{ZPVE}/n_{\text{DOF}}$ ranges between 1180 and 1273 cm^{-1} . Interestingly this is within the typical frequency span of vibrations that get excited upon $S_0 \rightarrow S_1$ transitions in cyanine dyes ($1200 \pm 200 \text{ cm}^{-1}$ [35]).

We tried to include possible solvent coupling effects and vibronic contributions into the calculation of ΔE_{max} by complementing Eq. 1 with an empirical $\text{ZPVE}/n_{\text{DOF}}$ -based correction term (Eq. 2). The improvement of the prediction with Eq. 2 in comparison to Eq. 1 is obvious (Figure 2b). R^2 and Q^2 increase from 0.968 to 0.994 and from 0.956 to 0.989, respectively. The error in the prediction of ΔE_{max} ranges now from -0.04 to $+0.06$ eV (-10 to $+12$ nm), the root-mean-square error is 0.03 eV (7 nm). Thus, both the maximal and root-mean-square error drop by a factor of two in comparison with the best model of Eq. 1. The model based on Eq. 2 yields also the correct order of ΔE_{max} values for the set of studied cyanine dyes; an exception is the couple **8** and **9** for which the experimental ΔE_{max} differs by less than 0.02 eV (*i.e.*, the difference is smaller than the root-mean-square error of the model).

When we compare the models of Eq. 2 and Eq. 1 in a likelihood-ratio test under the null hypothesis that Eq. 2 does not fit the data better than Eq. 1, we find an F value of 21.3. This is much larger than the value of the F -distribution (4.10) for a critical false-rejection probability of 0.05. In fact, the null hypothesis can be rejected with a certainty of 99.97 % (P value = 2.48×10^{-4}). This underlines the statistical significance of the correction term in Eq. 2.

We tested also a linear form of the correction term ($\gamma [\text{ZPVE}/n_{\text{DOF}}]$) with only one additional parameter

to fit) and a cubic form ($\gamma[\text{ZPVE}/n_{\text{DOF}}] + \delta[\text{ZPVE}/n_{\text{DOF}}]^2 + \varepsilon[\text{ZPVE}/n_{\text{DOF}}]^3$, *i.e.*, a third-order polynomial with three additional parameters). The former fits the data notably worse ($R^2 = 0.985$ and $Q^2 = 0.978$) than the quadratic form (Eq. 2), the latter fits marginally better ($R^2 = 0.995$, $Q^2 = 0.989$). When comparing the models of the quadratic correction term and the nested, linear term in a likelihood-ratio test, we find an F value of 14.3 which corresponds to a P value of 0.004 (one numerator and ten denominator degrees of freedom). The null hypothesis (*i.e.*, the quadratic correction term does not fit the data better than the linear correction term) can be rejected with a certainty of 99.6 %. The comparison of the models of the cubic correction term and the nested, quadratic term, yields, however, an F value of only 1.2 which corresponds to a P value of 0.302 (one numerator and nine denominator degrees of freedom). This is six times larger than the critical false-rejection probability of 0.05. The quadratic form of the correction term is therefore sufficient to describe the dependence of ΔE_{max} on $\text{ZPVE}/n_{\text{DOF}}$.

5. Conclusions

A more accurate TD-DFT prediction of the maximum absorption wavelength of symmetrical or unsymmetrical cyanine dyes is possible by introducing a correction term that potentially accounts for vibronic contributions and solvent coupling effects. Interestingly, with this correction term the prediction is also successful for “unusual” cyanine dyes that contain one or two imino linkage(s) within their polymethine chain. The maximal error in the prediction is 12 nm with correction term, instead of 28 nm without correction term; the improvement of the agreement (between prediction and experiment) is statistically significant. The correction term is based on the zero-point vibrational energy (ZPVE) normalized by the number of vibrational degrees of freedom. With the aid of a simplified model system it was shown that the correction term possibly accounts for the conceptual difference between the maximum absorption energy (as obtained from experiments, see ΔE_{max} in Figure 3) and the vertical excitation energy (as obtained from quantum chemical calculations, ΔE_{v}). The model is not limited to cyanine dyes but it is generally applicable to dyes with a dominant (0,0)-transition. For dyes with larger

geometric changes upon excitation (*i.e.*, for dyes where the most favorable transition leads to a vibrationally excited S_1 state) the difference between ΔE_{\max} and ΔE_v also depends on ZPVE: the energy for the most probable transition corresponds to the zero-point corrected vertical transition energy (Franck-Condon principle); the normalization of ZPVE by $3N-6$ could be less important though. We anticipate therefore that ZPVE-based correction terms can be used also for many families of dyes to improve the prediction of their maximum absorption wavelengths. The investigation of the theoretical basis of the correction term, especially its quadratic nature (*e.g.*, anharmonicity effects) is subject to future studies.

Acknowledgement. We thank Prof. Martin Karplus for valuable discussions, and GENCI (Grand Equipement National de Calcul Intensif, France) for computational resources. MS thanks ISIS and the University of Strasbourg for hosting him as guest assistant professor. SL and KM are grateful for financial support from Centre National de la Recherche Scientifique (CNRS) and the International Center for Frontier Research in Chemistry (FRC).

Supplementary data. Chemical structures of compounds **1-14**.

[1] Lotto RB, Purves D. The empirical basis of color perception. *Conscious Cogn* 2002; 11: 609-29.

[2] Herschel JFW. On a case of superficial colour presented by a homogeneous liquid internally colourless. *Phil Trans R Soc* 1845; 135: 143-5.

[3] Lavis LD, Raines RT. Bright ideas for chemical biology. *ACS Chem Biol* 2008; 3: 142-55.

[4] Mishra A, Behera RK, Behera PK, Mishra BK, Behera GB. Cyanines during the 1990s: a review.. *Chem Rev* 2000; 100: 1973-2011.

[5] Benson RC, Kues HA. Absorption and fluorescence properties of cyanine dyes. *J Chem Engin Data* 1977; 22: 379-83.

- [6] Meguellati K, Spichty M, Ladame S. Reversible Synthesis and Characterization of Dynamic Imino Analogues of Trimethine and Pentamethine Cyanine Dyes. *Org Lett* 2009; 11: 1123-6.
- [7] Meguellati K, Spichty M, Ladame S. Synthesis, spectroscopic and DNA alkylating properties of malondialdehyde (MDA) bis-imine fluorescent adducts. *Mol BioSyst* 2010; 6: 1694-9.
- [8] Gonçalves MS. Fluorescent Labeling of Biomolecules with Organic Probes. *Chem Rev* 2009; 109: 190-212.
- [9] Fabian J. TDDFT-calculations of Vis/NIR absorbing compounds. *Dyes Pigm* 2010; 84: 36-53.
- [10] Burke K, Werschnik J, Gross EKV. Time-dependent density functional theory: Past, present, and future. *J Chem Phys* 2005; 123: 062206.
- [11] Bamgbelu A, Wang J, Leszczynski J. TDDFT Study of the Optical Properties of Cy5 and Its Derivatives. *J Phys Chem A* 2010; 114: 3551-5.
- [12] Jacquemin D, Wathélet V, Perpète EA, Adamo C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J Chem Theory Comput* 2009; 5: 2420-35.
- [13] Hagberg DP, Marinado T, Karlsson KM, Nonomura K, Qin P, Boschloo G, Brinck T, Hagfeldt A, Sun L. Tuning the HOMO and LUMO Energy Levels of Organic Chromophores for Dye Sensitized Solar Cells. *J Org Chem* 2007; 72: 9550-6.
- [14] Champagne B, Guillaume M, Zutterman F. TDDFT investigation of the optical properties of cyanine dyes. *Chem Phys Lett* 2006; 425: 105-9.
- [15] Wang LY, Chen QW, Zhai GH, Wen ZY, Zhang ZX. Investigation of the structures and absorption spectra for some hemicyanine dyes with pyridine nucleus by TD-DFT/PCM approach. *J Mol Struct (Theochem)* 2006; 778: 15-20.
- [16] Fu YL, Huang W, Lia CL, Wang LY, Wei YS, Huang Y, Zhang XH, Wen ZY, Zhan ZX.

- Monomethine cyanine dyes with an indole nucleus: Microwave-assisted solvent-free synthesis, spectral properties and theoretical studies. *Dyes Pigm* 2009; 82: 409-15.
- [17] Dierksen M, Grimme S. The Vibronic Structure of Electronic Absorption Spectra of Large Molecules: A Time-Dependent Density Functional Study on the Influence of “Exact” Hartree–Fock Exchange. *J Phys Chem A* 2004; 108: 10225-37.
- [18] Luňák Jr. S, Vyňuchala J, Valac M, Havel L, Hrdina R. The synthesis, absorption and fluorescence of polar diketo-pyrrolo-pyrroles. *Dyes Pigm* 2009; 82: 102-8.
- [19] Guillaumont D, Nakamura S. Calculation of the absorption wavelength of dyes using time-dependent density-functional theory (TD-DFT). *Dyes Pigm* 2000; 46: 85-92.
- [20] Guillaume M, Liégeois V, Champagne B, Zutterman F. Time-dependent density functional theory investigation of the absorption and emission spectra of a cyanine dye. *Chem Phys Lett* 2007; 446: 165-9.
- [21] Hamer FM. *The Cyanine dyes and related compounds*. Interscience Publishers: New York, London; 1964.
- [22] Fayet G, Jacquemin D, Wathelet V, Perpète EA, Rotureau P, Adamo C. Excited-state properties from ground-state DFT descriptors: A QSPR approach for dyes. *J Mol Graph Modell* 2010; 28: 465-71.
- [23] McQuarrie DA, Simon JD. *Molecular Thermodynamics*. University Science Books: Sausalito; 1999.
- [24] Tomasi J, Mennucci B, Cancès E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J Mol Struct (Theochem)* 1999; 464: 211-26.
- [25] Becke AD. new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 1993;

98: 1372-7.

[26] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 1988; 37: 785-9.

[27] Adamo C, Barone V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J Chem Phys* 1999; 110: 6158-69.

[28] Frisch M. et al. Gaussian 09 Revision A.02. Gaussian, Inc.: Wallingford; 2009.

[29] Stratmann RE, Scuseria GE, Frisch MJ. An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. *J Chem Phys* 1998; 109: 8218-24.

[30] Leach AR, Gillet VJ. *An introduction to Chemoinformatics*. Revised Edition. Springer: Dordrecht; 2007.

[31] Allen MP. *Understanding Regression Analysis*. Plenum Press: New York; 1997.

[32] Sheppard SE, Brigham HR. Some Effects of Solvents upon the Absorption Spectra of Dyes. III. Temperature and Organic Solutions of Cyanine Dyes. *J Am Chem Soc* 1944; 66: 380-4.

[33] Marcus RA. Relation between charge transfer absorption and fluorescence spectra and the inverted region. *J Phys Chem* 1989; 93: 3078–86.

[34] Mustroph H, Reiner K, Mistol J, Ernst S, Keil D, Hennig L. Relationship between the molecular structure of cyanine dyes and the vibrational fine structure of their electronic absorption spectra. *ChemPhysChem* 2009; 10: 835-40.

[35] Pouradier J. Remarque sur le spectre d'absorption des cyanines en solution. *J Chim Phys* 1964; 61: 1107-14.

Figure Legends.

Figure 1. General structure of polymethine ($X = \text{CH}$) and polyimine ($X = \text{N}$) cyanine dyes covering a spectral range from blue to red. Pictures of selected dyes in solution (DMSO) are also given for illustration.

Figure 2. Plot of calculated and experimental ΔE_{max} values for cyanine dyes of Table 1 using Eq. 1 (a) and Eq. 2 (b). The fitted parameters are $\alpha = 1.2104$, $\beta = -0.8334$ eV for Eq. 1, and $\alpha = 1.1985$, $\beta = -0.8687$ eV, $\gamma = 1.2771 \times 10^4$ (eV)⁻², and $\delta = 7.7045 \times 10^{-2}$ eV, for Eq. 2, respectively. The value of δ is close to the value of $\text{ZPVE}/n_{\text{DOF}}$ averaged over all dyes ($\langle \text{ZPVE}/n_{\text{DOF}} \rangle = 7.6113 \times 10^{-2}$ eV). The arrows in a) indicate di-imino derivatives with large deviations between experiment and prediction using Eq. 1.

Figure 3. Schematic representation of the potential energy surfaces of the ground-state, S_0 , and the first excited state, S_1 . The value ΔE_v (thin arrow) corresponds to the potential energy difference between S_1 and S_0 at the equilibrium geometry of the ground state (r_{eq}''). Note that ΔE_v is a purely theoretical quantity (which has no experimental equivalence); it corresponds to the vertical electronic transition energy traditionally calculated in quantum-chemical calculations, *e.g.*, single-point TD-DFT calculations. For fluorescent molecules such as cyanine dyes the equilibrium geometry of S_1 (r_{eq}') differs only slightly from that of S_0 [34], so that the most intensive transition occurs between $v''=0$ and $v'=0$. When the vibrational frequency ν is identical in S_0 and S_1 the excitation energy for the most probable $S_0 \rightarrow S_1$ transition, ΔE_{max} (bold arrow), corresponds to D_0 (the energy difference between the minima of the two potential energy surfaces). Thus, the difference $\Delta E_{\text{max}} - \Delta E_v$ depends on the zero-point vibrational energy, $\text{ZPVE} = \frac{1}{2} h\nu$, and the difference between r_{eq}' and r_{eq}'' in reduced coordinates, $d = a^{1/2} r_{\text{eq}}'' - a^{1/2} r_{\text{eq}}'$ (with $a = 4\pi^2 \mu \nu / h$; h the Planck constant and μ the reduced mass of the vibration).

Table 1. Absorption properties of cyanine dyes from experiments and calculations, and zero-point vibration energy per vibrational degree of freedom. Chemical structures of compounds **1-14** are shown in the supplementary data.

Compound	Chemical Structure		Experiment			Calculation			
	Het1	Het2	X	Y	n	λ_{\max} (nm)	ΔE_{\max} (eV)	ΔE_v^a (eV)	ZPVE/ n_{DOF}^b (10^{-2} eV)
1	Benzothiazole	Benzothiazole	N	N	0	412	3.00	3.13	7.40
2	Quinoline	Quinoline	N	N	0	424	2.92	3.16	7.81
3	Naphtobenzothiazole	Naphtobenzothiazole	N	N	0	460	2.69	2.81	7.32
4	Benzothiazole	3,3-dimethylindole	N	CH	0	469	2.64	2.86	7.84
5	Benzothiazole	Benzothiazole	N	N	1	512	2.42	2.66	7.42
6	Quinoline	Quinoline	N	N	1	520	2.38	2.67	7.79
7	Benzothiazole	3,3-dimethylindole	CH	CH	0	542	2.29	2.63	7.90
8	Naphtobenzothiazole	Naphtobenzothiazole	N	N	1	557	2.23	2.44	7.34
9	Benzothiazole	Benzothiazole	CH	CH	0	562	2.21	2.58	7.57
10	Benzothiazole	3,3-dimethylindole	N	CH	1	579	2.14	2.51	7.82
11	Naphtobenzothiazole	Naphtobenzothiazole	CH	CH	0	592	2.09	2.40	7.45
12	Benzothiazole	3,3-dimethylindole	CH	CH	1	643	1.93	2.31	7.88
13	Benzothiazole	Benzothiazole	CH	CH	1	660	1.88	2.29	7.58
14	Naphtobenzothiazole	Naphtobenzothiazole	CH	CH	1	683	1.82	2.15	7.46

a) PBE0(PCM)/6-311++G**/B3LYP(PCM)/6-31G*

b) B3LYP(PCM)/6-31G*

Table 2. The correlation coefficient R^2 for the fit of the experimental ΔE_{\max} data using Eq. 1 and different methods of calculating ΔE_v . The leave-one-out cross validated correlation coefficient Q^2 is shown in italics. For selective cases the single-point TD-DFT calculations were carried out without solvation model (see values in parentheses). The best result is marked in bold.

Geometry	Functional	Basis set			
		6-31G*	6-311G*	6-311++G**	
<i>Gas phase</i>	B3LYP	0.925	0.935	0.937	
		<i>(0.776)</i>	<i>0.903</i>	<i>0.918</i>	
	PBE0	<i>(0.717)</i>	0.953	0.960	0.962
		<i>(0.843)</i>	<i>0.937</i>	<i>0.946</i>	<i>0.948</i>
		<i>(0.802)</i>			
<i>Solution</i>	B3LYP	0.935	0.944	0.946	
		<i>0.916</i>	<i>0.926</i>	<i>0.929</i>	
	PBE0	0.960	0.967	0.968	
		<i>0.947</i>	<i>0.955</i>	0.956	

Figure 1

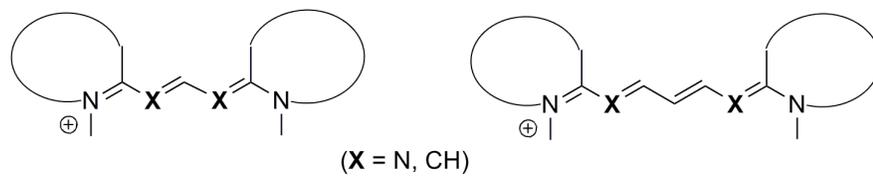
**CH****N****CH****N**

Figure 2

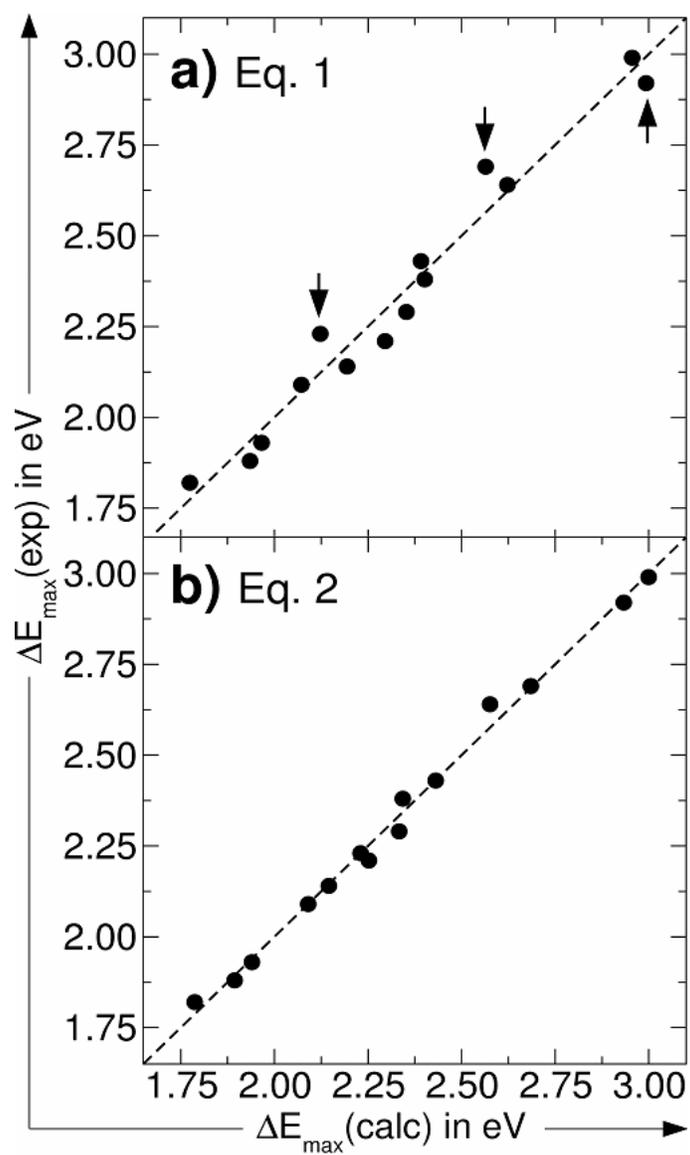
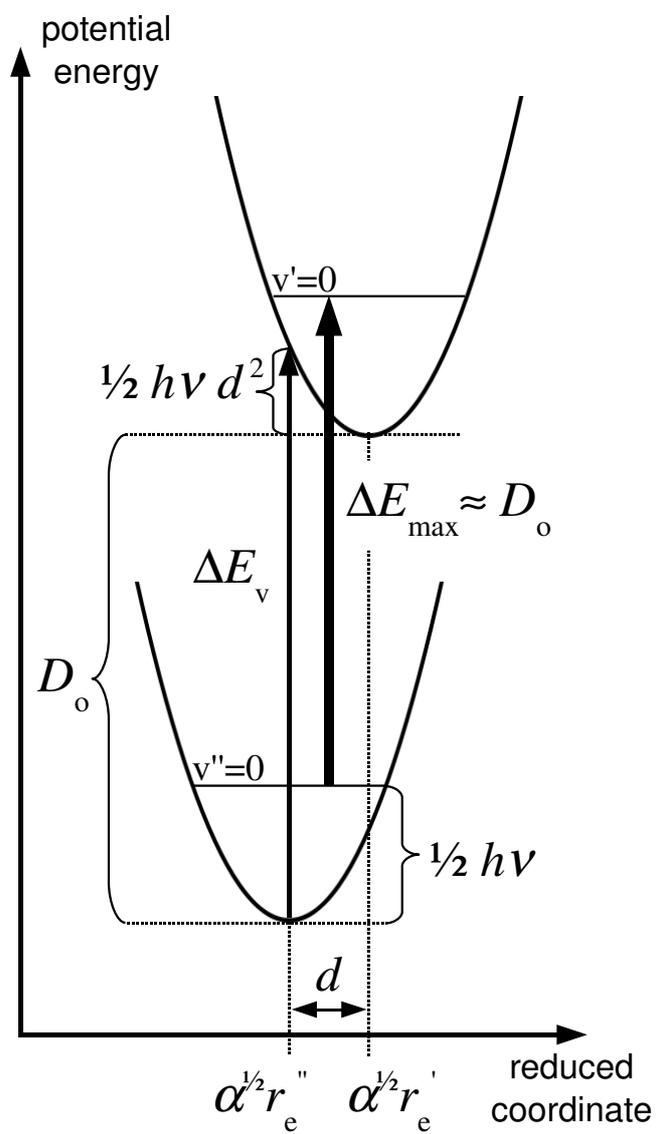


Figure 3



A conceptually improved TD-DFT approach for predicting the maximal absorption wavelength of cyanine dyes.

Supplementary data

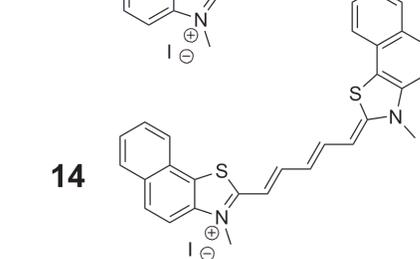
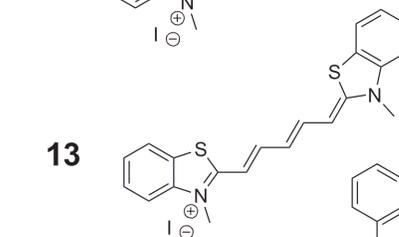
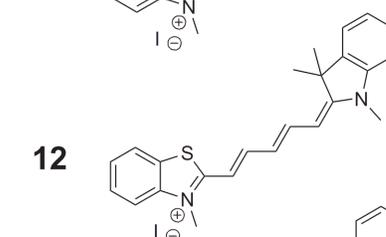
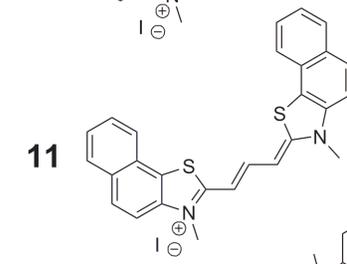
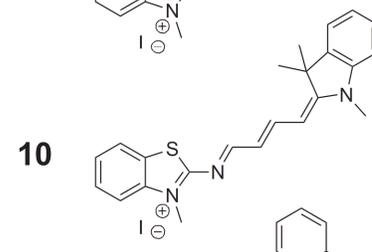
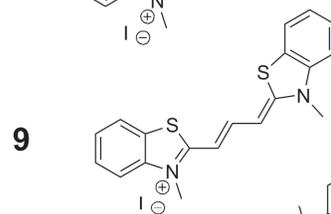
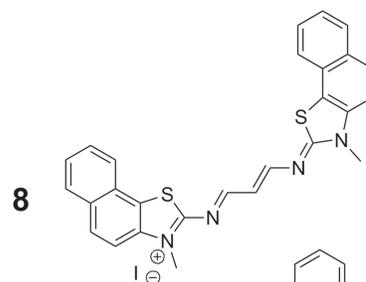
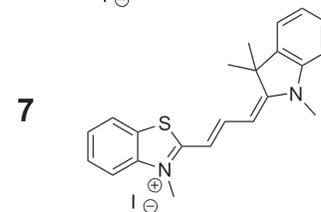
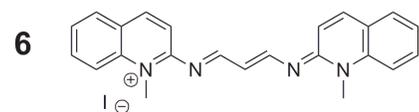
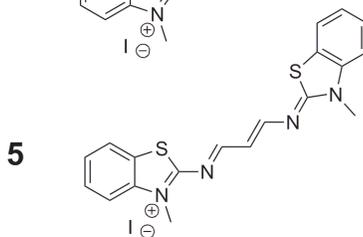
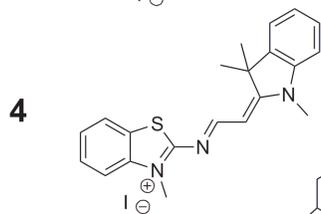
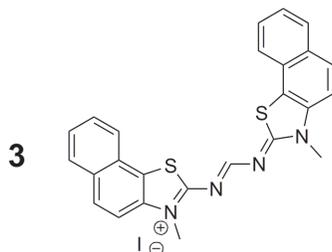
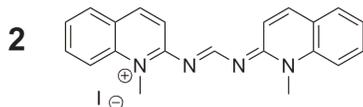
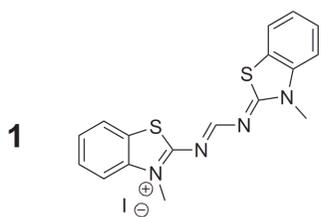
Kamel Meguellati[§], Sylvain Ladame[§], and Martin Spichy^{§,*}

[§]Institut de Science et d'Ingénierie Supramoléculaires, Université de Strasbourg, 8 allée Gaspard Monge, B.P. 70028, 67083 Strasbourg Cedex,

[§]Current address: Laboratoire de Biologie Moléculaire de la Cellule, Différenciation et Cycle Cellulaire, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07.

*To whom correspondence should be addressed, Email: martin.spichy@ens-lyon.fr.

Chemical structures of studied cyanine dyes:



CHAPTER FOUR

REVERSIBLE FORMATION OF IMINO CYANINE DYES IN WATER-IN-OIL DROPLETS

I. Introduction to microfluidics

Biotechnology is a science that is not only dedicated to assist biologists in their desire to understand the complexity of life. Since few decades, miniaturization of devices has been considered as very important. The fluids and their properties describe a big part of the life. For example in leaves, the sap flows through micro-channels. Fluids commonly used in microfluidic devices include whole blood samples, bacterial cell suspensions, protein or antibody solutions and various buffers. Microfluidic is the science and technology of fluids compartmentalized in micro-channels of 5 to 500 μm diameter. Compartmentalization of individual samples in aqueous droplets dispersed in an oil phase is now becoming a powerful method for high-throughput assays in chemistry and biology.

Microfluidics is a fairly recent concept that emerged in the beginning of the 1980s. In vitro compartmentalization was then developed by Griffiths and Tawfik for directed in vitro evolution in the late 1990s.¹ Water-in-oil droplets can be easily made by mixing oil and water using a stirrer, homogenizer or extruder. Droplets generated in this way have proven successful in many applications. These droplets can be as small as bacteria, with diameters of $\geq 1 \mu\text{m}$, and have volumes of less than a femtolitre, although emulsions can also be made with droplets of diameters up to 100 μm and volumes of nearly 1 nanolitre. The high capacity ($>10^{10}$ droplets in 1 ml of emulsion), the ease of preparing emulsions and their high stability over a broad range of temperatures, pH and salt concentrations render them an ideal means of compartmentalizing biochemical and genetic assays.

In recent years, microfluidic systems have been used extensively for the development of inkjet print-heads, DNA chips, lab-on-a-chip technology, micro-propulsion, and micro-thermal technologies². Microfluidic applications are very broad and diverse. For instance, microfluidic systems have been developed for the determination of molecular diffusion coefficients³, fluid viscosity, pH⁴, chemical binding coefficients and enzyme reaction kinetics⁵. Other applications for microfluidic devices include capillary electrophoresis⁶, isoelectric focusing⁷, immuno-assays⁸, flow cytometry⁹, sample injection of proteins for analysis via mass spectrometry¹⁰, crystallization of proteins¹¹, PCR amplification¹², DNA analysis¹³, cell manipulation¹⁴, cell separation¹⁵, cell patterning¹⁶ and chemical gradient formation¹⁷.

Basic principles of microfluidic:

The flow of a fluid through a microfluidic channel can be characterized by the Reynolds number, defined as¹⁸:

$$Re = \frac{\frac{\rho U^2 L}{\mu U}}{\frac{L^2}{L^2}} = \frac{\rho L U}{\mu}$$

with L represents the most relevant length scale, μ represents the viscosity, U represents the average of velocity, ρ is the fluid density.

Re is a dimensionless number interpreted as the ratio of inertial forces to viscous forces in the fluid. The inertial forces are proportional to the flow diameter, velocity, and fluid density. The viscous force is represented by the fluid's absolute viscosity. Due to the small dimensions of microchannels, Re is often much less than 100 and even less than 1. The number of Reynolds is very small because of the small scale and the high speed of the liquid flow within the channel. Flow is completely laminar and out of turbulence. In a laminar flow, all layers are parallel and directed by diffusion without convection or advection. Laminar flow in microfluidic permits to predict the manner that compounds are transported through the microchannel.

Devices design and preparation:

The technique used for the preparation of microfluidic devices is the soft lithography using PDMS^{19,20} (Polydimethylsiloxane). PDMS has become the most widely used material in microfluidic device fabrication due to its many advantages in fabrication ease, physical properties and economy. PDMS, as an elastomer, is known for its mold-release properties and ability to replicate features down to the nanoscale, with low shrinkage during cure (around 1%) and excellent elastic properties. It is soft, flexible but not highly resistant and permeable with gases. Another material used for the fabrication of microfluidic devices is silica glass or Plexiglas. The figure below summarizes all the steps for the preparation of PDMS-based devices:

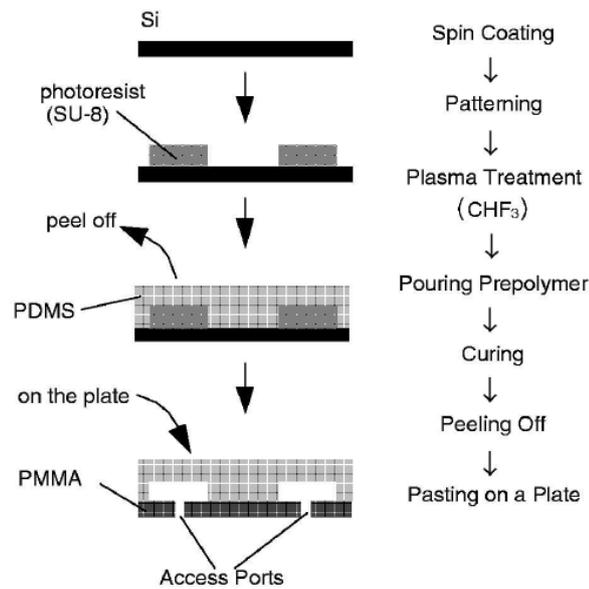
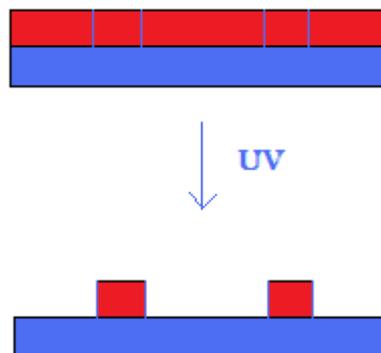


Figure 1: Synthesis of PDMS-based micro device extracted from reference ²¹

Briefly, this multi-step procedure involves:

- (1) Formation of a mold for casting the PDMS structure. The mold can be made of a variety of materials but SU-8 is the most common molding media used for PDMS-based microfluidic structure fabrication. Mold formation via photolithography requires the use of a mask and a light source. Typically, the mold is a “negative” mold with the PDMS poured into it and filling the regions left open by the mold: The irradiation with UV permits to sculpt the surface and only the photoresist SU-8 stays on the silica.



- (2) Casting, curing and releasing PDMS from the mold. The pre-polymer of PDMS is then poured onto the previously formed mold. The procedure for casting of PDMS

replicas from an etched Si master with positive relief structures (in Si) has been developed for chemical analysis. A mixture of PDMS/AIBN 9/1 is poured on a silicium surface containing on the surface the positive structure of the microchannel under appropriate atmosphere. Then, after heating the silicium box which contains the PDMS, all is heated at 100 °C for at least 2 days. The PDMS is peeled off very carefully from the silica surface and flushed with nitrogen.

- (3) Interfacing and integration. Naturally, the structure will require inlet and outlet connections. For this purpose, holes can be drilled in appropriate sites. The PDMS is finally pasted on a Si plate in glass using a plasma oven. The plasma oxidizes the silica surface by activation of the Si-OH in SiO. After that, the microchannels are treated with silane to increase the hydrophobicity of the surface.

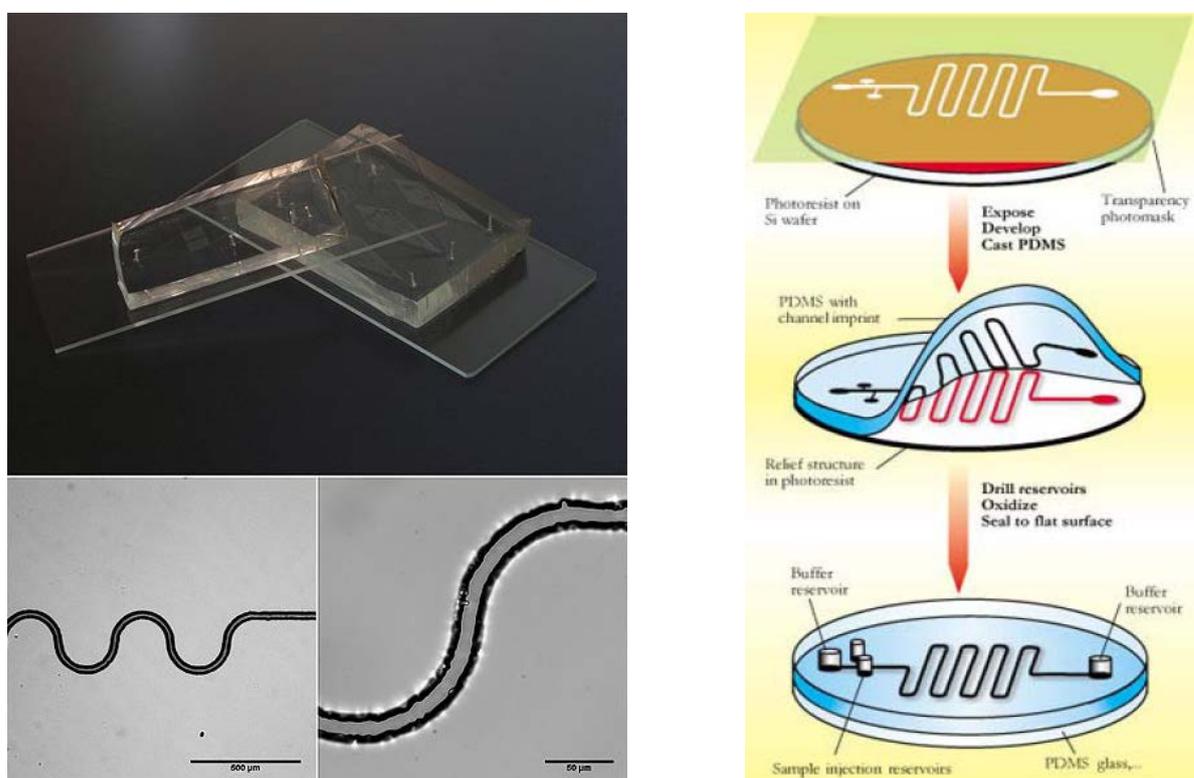


Figure 2: on the left: one picture of a typical device and on the right the detail description.

Just to illustrate the power of the microfluidic technology, Wang recently used an integrated microfluidic device to perform 1000 Huisgen's cycloaddition reactions in on time²²

thus demonstrating the potential of microfluidic to create chemical and structural diversity in a very high-throughput manner.



Figure 3: Device developed for click chemistry

Droplet production:

The first step in the microfluidic life cycle of a droplet is its production. The majority of microfluidic methods produce droplet volumes ranging from femtolitres to nanolitres. This is achieved through passive techniques which generate a uniform, evenly spaced, continuous stream. Not only should devices for making drops produce a regular and stable monodisperse droplet stream, they also need to be flexible enough to provide droplets of prescribed volume at a prescribed rate. To this end, three main approaches have emerged based on different physical mechanisms; they are best described by the flow field topology in the vicinity of the drop production zone: (i) breakup in co-flowing streams, (ii) breakup in cross-flowing streams and (iii) breakup in elongational strained flows. The junction where the two fluids meet is designed to optimize the reproducibility of droplet production. Indeed, the geometry of the junction, together with the flow rates and the physical properties of the fluids (interfacial tension, viscosities) determine the local flow field, which in turn deforms the interface and eventually leads to droplet formation. The size of the droplet is set by a competition between the pressure due to the external flow and viscous shear stresses, on the one hand, and the capillary pressure resisting deformation on the other. A commonly used system for the production of water-in-oil droplets is by using a flow-focusing device (see Figure below). A flow of water comes perpendicular to the oil flow which can then cut the water flow and generate droplets in the nozzle. The droplets are stabilized in the presence of surfactants in order to ensure they are homogeneous and mono-disperse.

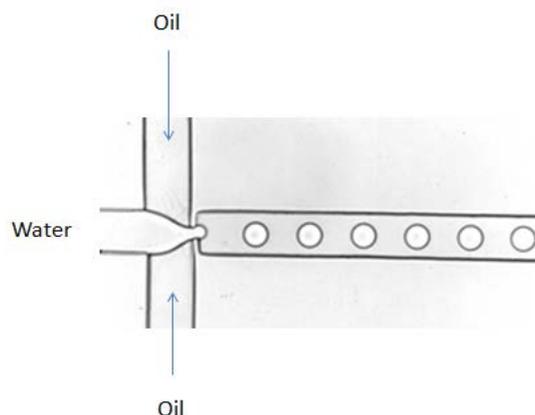


Figure 4: Generation of water droplets in oil phase by flow-focusing (extracted from ref ²³)

Another approach, the co-flow, consists in the mixing of two fluids just before the generation of the droplets as described in the bottom figure²⁴:

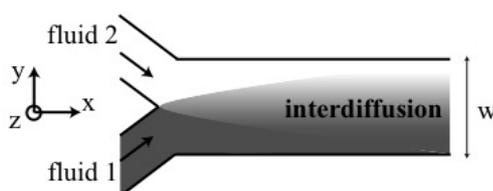


Figure 5: Generation of droplets by co-flow.

II. Chemical reactions in droplets

To perform reactions within microfluidic devices, at least two criteria should be met. First, the microfluidic tool should be able to perform typical procedures that are conducted for reactions on the macro-scale. These procedures include the controlled addition of reagents to a reaction mixture, the thorough mixing of reagents, control of the reaction time, the combining and splitting of reaction mixtures for multiple-step reactions, and analysis over the course of a reaction. Second, the microfluidic tool should provide a characteristic advantage, for example, the ability to perform more reactions under more reaction conditions. Within the last decade, considerable progress has been made in order to do chemistry in droplets: different techniques have been developed to introduce reagents into droplets for different droplet-based microfluidic applications; Rapid and controlled mixing of the reagents inside the droplet has been achieved; control of the liquid-liquid interface has been achieved using

for instance fluorinated surfactants; combining and splitting reactions is now possible in order to carry multi-step syntheses; analysis of the content of single droplets is now possible, although requiring ultra-sensitive detection techniques (e.g. fluorescence spectroscopy).

III. PUBLICATION

In this chapter, we have investigated the effect of compartmentalization into water-in-oil microdroplets on the kinetics and thermodynamics of our fluorogenic and reversible reaction of imine formation. Briefly, two non-fluorescent and water-soluble amine and aldehyde building blocks were synthesized. Their ability to form a fluorescent imino cyanine dye in bulk solvent and in droplets of different sizes has been investigated by fluorescence spectroscopy. Interestingly, we discovered that although this reversible reaction of imine formation is highly unfavorable in bulk water, it can be catalyzed when sequestered into droplets. Even more interesting is the observation that the initial rate and the product yield at equilibrium are proportional to the ratio of droplet surface area to volume, increasing as the droplet volume decreased, indicating that the observed effect is mediated by the droplet interface. Although changes in activity coefficient due to the microenvironment at the phase interface could play a role, this effect can be explained by a decrease in entropy of the reactants due to low affinity Van der Waals and/or hydrophobic interactions of the reactants with the droplet interface. Indeed, at the interface, the dimensionality of the reaction space is effectively reduced from 3 to 2 dimensions resulting in high reactant concentrations, thus increasing the rate of product formation and shifting the equilibrium towards product formation. This suggests that compartments the size of bacterial cells can be efficient environments for synthetic chemistry and that droplets or other systems with high surface to volume ratios may have played an important role in synthetic chemistry in the prebiotic world.

This work was carried out in collaboration with the group of Pr Andrew Griffiths (ISIS, Strasbourg). The experimental details, main results and result interpretations are given below. They will be submitted for publication within the next few weeks and they are presented in a publication format.

REFERENCES

- ¹ Tawfik D.S., Griffiths A.D. *Nature Biotechnol.* **1998**, *16*, 652.
- ² <http://en.wikipedia.org/wiki/Microfluidics>
- ³ (a) Kamholz, A. E., Weigl, B. H., Finlayson, B. A. & Yager, P. *Anal. Chem.* **1999**, *71*, 5340; (b) Kamholz, A. E., Schilling, E. A. & Yager, P., *Biophys. J.* **2001**, *80*, 1967.
- ⁴ (a) Weigl, B. H. & Yager, P., *Sensors and Actuators B-Chemical* **1997**, *39*, 452; (b) Macounova, K., Cabrera, C. R., Holl, M. R. & Yager, P., *Anal. Chem.* **2000**, *72*, 3745.
- ⁵ (a) Duffy, D. C., Gillis, H. L., Lin, J., Sheppard, N. F. & Kellogg, G. J., *Anal. Chem.* **1999**, *71*, 4669; (b) Hadd, A. G., Jacobson, S. C. & Ramsey, J. M., *Anal. Chem.* **1999**, *71*, 5206.
- ⁶ Kameoka, J., Craighead, H. G., Zhang, H. W. & Henion, J., *Anal. Chem.* **2001**, *73*, 1935.
- ⁷ Macounova, K., Cabrera, C. R. & Yager, P., *Anal. Chem.* **2001**, *73*, 1627.
- ⁸ Hatch, A. *et al.*, *Nature Biotechnol.* **2001**, *19*, 461.
- ⁹ Sohn, L. L. *et al.*, *Proc. Nat. Acad. Sci. USA* **2000**, *97*, 10687.
- ¹⁰ Jiang, Y., Wang, P. C., Locascio, L. E. & Lee, C. S. , *Anal. Chem.* **2001**, *73*, 2048.
- ¹¹ Hansen, C. L., Skordalakes, E., Berger, J. M. & Quake, *Proc. Natl Acad. Sci. USA* **2002**, *99*, 16531.
- ¹² Belgrader, P., Okuzumi, M., Pourahmadi, F., Borkholder, D. A. & Northrup, M. A., *Biosensors & Bioelectronics* **2000**, *14*, 849.
- ¹³ Buchholz, B. A. *et al.*, *Anal. Chem.* **2001**, *73*, 157.
- ¹⁴ Glasgow, I. K. *et al.*, *Transactions On Biomedical Engineering* **2001**, *48*, 570.
- ¹⁵ Yang, J., Huang, Y., Wang, X. B., Becker, F. F. & Gascoyne, P. R. C. , *Anal. Chem.* **1999**, *71*, 911.
- ¹⁶ Chiu, D. T. *et al.* , *Proc. Natl Acad. Sci. USA* **2000**, *97*, 2408.
- ¹⁷ Dertinger, S. K. W., Chiu, D. T., Jeon, N. L. & Whitesides, G. M., *Anal. Chem.* **2001**, *73*, 1240.
- ¹⁸ Squires, T. M. & Quake, S. R., *Rev. Mod. Phys.* **2005**, *77*, 977.
- ¹⁹ Xia, Y. and Whitesides, G. M., *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 551.
- ²⁰ Rogers, J. A. & Nuzzo, R. G. *Materials today* **2005**, *8*, 50.
- ²¹ Teruo F., *Microelectronic Engineering* **2002**, *61–62*, 907.
- ²² Y. Wang, W.Y. Lin, K. Liu, R.J. Lin, M. Selke, H.C. Kolb, N. Zhang, X.Z. Zhao, M.E. Phelps, C.K.F. Shen, K.F. Faull and H.R. Tseng, *Lab Chip*, **2009**, *9*, 2281.
- ²³ <http://www-isis.u-strasbg.fr/lbc/index2.html>
- ²⁴ <http://www.math.u-bordeaux1.fr/~jdambrin/english/img/interdiffusion.jpg>

Manuscript in preparation:

Enhanced efficiency of synthetic organic reactions in picolitre volume droplets

Abstract

A fluorogenic reaction was used to monitor imine bond formation from amine and aldehyde building blocks in monodisperse aqueous droplets dispersed in fluorinated oil using a microfluidic system and stabilized by surfactant. The reaction is slow and thermodynamically disfavored in bulk solvent, but in droplets of less than 40 pl volume ($\leq 42 \mu\text{m}$ diameter) the initial rate of product formation and equilibrium is substantially shifted towards the formation of the fluorescent imino product. In 2.5 pl droplets (17 μm diameter), the initial rate of product formation was increased by a factor of 100 and the yield of product at equilibrium increased by a factor of 29 compared to bulk solvent. Both the initial rate and the product yield at equilibrium were proportional to the ratio of droplet surface area to volume, increasing as the droplet volume decreased, indicating that the observed effect is mediated by the droplet interface. Although changes in activity coefficient due to the microenvironment at the phase interface may play a role, this effect can be explained solely by a decrease in entropy of the reactants due to low affinity Van der Waals and/or hydrophobic interactions of the reactants with the droplet interface: at the interface, the dimensionality of the reaction space is effectively reduced from three to two dimensions resulting in high reactant concentrations, increasing the rate of product formation and shifting the equilibrium towards product formation. This suggests that compartments the size of biological cells can be efficient environments for synthetic chemistry and that droplets, or other systems with high surface to volume ratios may have played an important role in synthetic chemistry in the prebiotic world.

Introduction

In solution, entropic factors tend to thermodynamically disfavor the formation of larger molecules from smaller ones. A chemical reaction can proceed if the change in Gibbs free energy of reaction (ΔG) is negative: $0 > \Delta G = \Delta H - T\Delta S = G_{\text{products}} - G_{\text{reactants}} = H_{\text{products}} - H_{\text{reactants}} - T(S_{\text{products}} - S_{\text{reactants}})$. The enthalpy of the reaction (ΔH) depends mainly on the change of bond energy and solvation energy. The entropy of the reaction (ΔS) is a measure of the change in possibility of movement of the constituents of the reaction: the change in the degrees of freedom of overall translational movement and rotational movement is most significant whereas the contribution by vibrations and internal rotations is small (1). The formation of one product molecule from two reactant molecules ($2 \rightarrow 1$ reaction) in solution results in a high loss of mobility (and hence a loss of entropy) equivalent to 6 degrees of freedom (3 translational and 3 rotational). Hence, for a synthetic reaction in solution to have an equilibrium position ($\Delta G = 0$) favoring product formation, highly reactive functional groups are required so that a high negative value of ΔH can compensate for a high negative value of ΔS . This thermodynamic unfavorability has led to strong criticism (2-4) of the prebiotic broth theory for the origin of life of Darwin, Oparin and Haldane (5-7). Synthetic reactions are, however, favored, both kinetically and thermodynamically, by increasing the concentration of the reactants: in $2 \rightarrow 1$ reactions of type $A + A \rightarrow B$, the initial rate $v =$

$k_1[A]^2$ and at equilibrium $[B] = [A]^2/K_{eq}$ (where k_1 = the forward rate constant and K_{eq} = the equilibrium constant). One hypothesis to explain the generation of sufficiently high concentrations of reactants for efficient prebiotic synthetic chemistry is based on atmospheric aerosols (8). The water in atmospheric aerosol droplets of $\sim 1 \mu\text{m}$ diameter coated with organic surfactant films, originally formed at the ocean surface, is proposed to evaporate until a complete and compact surfactant layer is formed, at which point water would continue to transfer across the complete layer, but at a slower rate (9). In addition to concentrating aqueous solution, these droplets have a number of other attractive features for prebiotic chemistry, including mobility through a wide range of temperature and radiation fields, their frequent, widespread and continuous creation and their ability to merge and divide, so sharing contents and information. The potentially important role of the droplet interface has, however, been largely overlooked.

In 1968, Adam and Delbruck proposed that the efficiency of molecular collision (and hence reaction efficiency) could be improved by reducing the dimensionality in which diffusion takes place from a three dimensional space to a two-dimensional surface (10). Indeed, the idea that life may have arisen on the two-dimensional surface of minerals dates as far back as 1949, with the clay theory of Bernal (11), and other theories based on chemistry on surfaces have been proposed since, e.g. (12).

More recently, it has been demonstrated that several uni- and bimolecular reactions are greatly accelerated when carried out in vigorously stirred aqueous suspension (an emulsion) with high interfacial area (13). The mechanistic basis for the acceleration seen with these “on water” reactions has not been elucidated, but it is clear that the properties of the phase boundary between the water and insoluble hydrophobic oils play a key role.

We show here that in small aqueous droplets, in which the surface to volume ratio is high, for a synthetic $2 \rightarrow 1$ reaction the equilibrium of a reaction which is thermodynamically disfavored in bulk solvent can be substantially shifted towards the formation of product and the rate of product formation greatly enhanced. Although factors such as the unique properties of molecules at the interface between water and insoluble hydrophobic oils (14-18), or the redistribution of surface species driven by surface-tension energetics (19) may play a role, this effect can be explained solely by low affinity Van der Waals and/or hydrophobic interactions of the reactants with the droplet interface and indicated that droplets, or other spaces, in which the surface to volume ratio is large, can be efficient environments for synthetic chemistry, and that such systems with high surface to volume ratios may have played an important role in synthetic chemistry in the prebiotic world.

Results and discussion

We recently achieved the synthesis, in organic solvent, of a fluorescent imino analogue of a trimethine cyanine dye from a non-fluorescent amine (i.e. N-methyl-2-aminobenzothiazolium iodide) and a very weakly fluorescent (fluorescence quantum yield $\phi < 0.0002$) indoline Fisher's base aldehyde. This dye differs from the well known trimethine cyanine dyes by one C \rightarrow N substitution in the polymethine chain, thus making its formation reversible and thermodynamically controlled (20). Due to its fluorogenic nature, this reaction offers the advantage of easy and highly sensitive monitoring by fluorescence spectroscopy.

Herein, we investigated the effect of external stimuli (microenvironment and compartmentalization) on the formation of an imino-based cyanine dye in water. Two cationic and water-soluble amine and aldehyde building blocks were synthesized as described in the Supplementary Materials. Formation of the fluorescent imine **3** from aldehyde **1** and amine **2** (**Fig. 1A**) was first investigated in bulk solvent and its properties determined by UV-absorption and fluorescence spectroscopy. As anticipated for a reversible reaction proceeding

via formation of an hemiaminal intermediated followed by water elimination, the formation of imine **3** proved more efficient in organic solvents than in water: at a given concentration of amine and aldehyde (5 mM each), twice as much imine was formed at equilibrium in DMSO compared to water. In aqueous solution, formation of the imine **3** was barely detectable with stoichiometric reactant concentrations ≤ 0.5 mM, but when working at a ≥ 5 mM concentration of both amine and aldehyde, formation of imine **3** was easily detectable and increased over a period of 20-24h until equilibrium was reached (**Fig. 1B**). Nevertheless, imine **3** formation was both slow ($k_1 = 4.89 \times 10^{-4} \text{ M}^{-1} \text{ min}^{-1}$) and thermodynamically unfavorable ($K_{\text{eq}} = 2.62 \times 10^{-2} \text{ M}^{-1}$).

Both the rate of imine **3** formation, and its equilibrium concentration could, however, be greatly enhanced by performing the reaction compartmentalized in aqueous droplets in an inverse emulsion. We used a microfluidic system to create highly monodisperse droplets, allowing a detailed investigation of the role of the interface on the thermodynamic properties of our chemical system. Each droplet contained a stoichiometric mixture of aldehyde **1** and amine **2** (15 mM each), with volumes ranging from 2.5 to 160 μl (corresponding to spherical droplets of diameter 17 to 67 μm) (**Fig. 2A** and **Table 1**). Droplets were produced by flow focusing of the aqueous phase (**21**) with a fluorinated oil phase (HFE7500, 3M) containing a PEG-PFPE tri-block copolymer surfactant (**22**). With droplets of 160 μl volume, there was only a small difference compared to the reaction in bulk: the apparent second order rate constant for the formation of imine, $k_{\text{app}} = 3.18 \times 10^{-4} \text{ M}^{-1} \text{ min}^{-1}$ and the apparent equilibrium constant, $K_{\text{app}} = 2.37 \times 10^{-2} \text{ M}^{-1}$). However, as droplet volume was reduced, both the rate of formation, and equilibrium concentration of imine **3** were greatly increased: in 2.5 μl droplets, $k_{\text{app}} = 4.86 \times 10^{-2} \text{ M}^{-1} \text{ min}^{-1}$ and $K_{\text{app}} = 0.759 \text{ M}^{-1}$, representing a factor of 100-fold and 29-fold increase compared to k_1 and K_{eq} in bulk (**Fig. 2B** and **2C** and **Table 1**).

The ratio of surface area a , to volume v , of a spherical droplet is inversely proportional to the radius r ($a/v = 3/r$), hence, as the droplets become smaller, a/v increases. Both k_{app} and K_{eq} increased linearly as a/v of the droplets increased, indicating that the observed effects are mediated by the droplet interface (**Fig. 2C** and **2D**).

In an attempt to provide a better understanding of the chemical phenomena responsible for the reaction acceleration and increased reaction efficiency observed in μl volume droplets (when compared to similar reactions carried out in bulk) reactions were carried out in bulk in the presence of various self-organized systems like surfactants (cationic, anionic and non ionic) or polyelectrolytes. Various concentrations, above and below the critical micelle concentration (CMC), were examined to identify the effect of the monomers organization on reaction efficiency (**Table 2**).

Only the anionic molecules (SDS, PAA250 and PAA805) proved capable of substantially enhancing the initial rate and equilibrium concentration of imine **3**. In contrast, the addition of cationic (CTAB) and non-ionic (Pluronic F-127, PVP, Triton X-100, Zonyl and Synperonic F-108) molecules resulted in little or no change in imine production. The enhancement was more marked at high (0.8% w/w) than at low (0.03% w/w) surfactant concentrations, but comparable trends were obtained in both cases. At 0.03%, all of the surfactants tested are below the CMC, and no supramolecular auto-organization of the surfactant was possible. However, at 0.8%, with the exception of PAA250, PAA805 and Pluronic F-108, all the surfactants are above the CMC and were shown to form micellar structures, the size of which was determined by dynamic light scattering (DLS) (see Supporting Materials).

Considering the cationic nature of both reactants and of the final imino dye **3**, it is not surprising to see that both the initial rate and the equilibrium constant of the reaction increased upon addition of the anionic surfactants only. Moreover, at 0.8% concentration, of the three anionic molecules tested, SDS induced the highest increase in reaction efficiency.

This could be correlated with the fact that it is also the only one capable of micellar auto-organization, forming fairly large micelles of 390 nm diameter (as determined by DLS, see Supporting Materials) with a highly charged surface and a hydrophobic core. In order to study the effect of self-organization of SDS into micelles, kinetic experiments were carried out by varying the SDS concentration from 0.15 times to 2.5 times its CMC (CMC = 9 mM) (**Fig. 3A** and **3B**). When working at SDS concentrations below the CMC, the amount of imine formed at equilibrium only slowly increased with SDS concentration, while a dramatic shift of the position of equilibrium was observed when working above the CMC (**Fig. 3A**). In contrast, while the initial reaction rate increases linearly with SDS concentration when working below the CMC, it reaches a plateau when the $[SDS] = CMC$ (**Fig. 3B**). Increasing the SDS concentration further does not accelerate the reaction, the optimal acceleration being obtained when working at the CMC. This indicates that below the CMC, SDS in solution catalyses the reaction, probably as a result of favorable electrostatic interactions between the anionic surfactant and the transition state. However, below the CMC, SDS only subtly effects the position of equilibrium, likely due to hydrophobic interactions between the SDS tails (in the absence of micelle formation) inducing a local increase in reactant concentration. The catalytic effect seems, however, to be less important for SDS in the form of micelles, since the initial rate does not increase further with increasing SDS above the CMC, where extra SDS is incorporated spontaneously into micelles. However, above the CMC, the equilibrium position of the reaction is substantially shifted towards the formation of imine, probably largely due to the increase in local concentration of reactants due to sequestration in the micelles (23, 24).

Indeed, the rate and yield of many reactions is enhanced in micellar systems (23, 24), and although micellar systems (which are thermodynamically favorable, and form spontaneously) are fundamentally different from emulsions (which are thermodynamically unfavorable and do not form spontaneously) (25), there are clearly similarities between reactions in micellar systems and the effect observed here in droplets. In some cases, micelles function as catalysts since they possess catalytically active groups and/or the transition state of the reaction is stabilized (reduction of the activation energy) by interaction with the polar head groups of the surfactant. In other cases the enhancement is thought to be due to the micellar microenvironment (activity coefficient effects). However, for bimolecular reactions, the dominant effect in many cases is thought to be due to reactants being concentrated relative to the surrounding water phase through interaction with the micelle surface or insertion into the micelle itself (entropy effects) (23, 24).

We therefore modeled the droplet system, taking into account only entropy effects resulting from interactions of reactants with the droplet interface. The model is based on the assumption that the system is governed by four reversible processes: 1) reaction of aldehyde **1** and amine **2** to form imine **3** in aqueous solution, 2) binding of **1** and **2** to the droplet interface, 3) reaction of **1** and **2** to form **3** at the interface, and 4) dissociation of **3** from the interface (**Fig. 4A**).

Figures and Tables

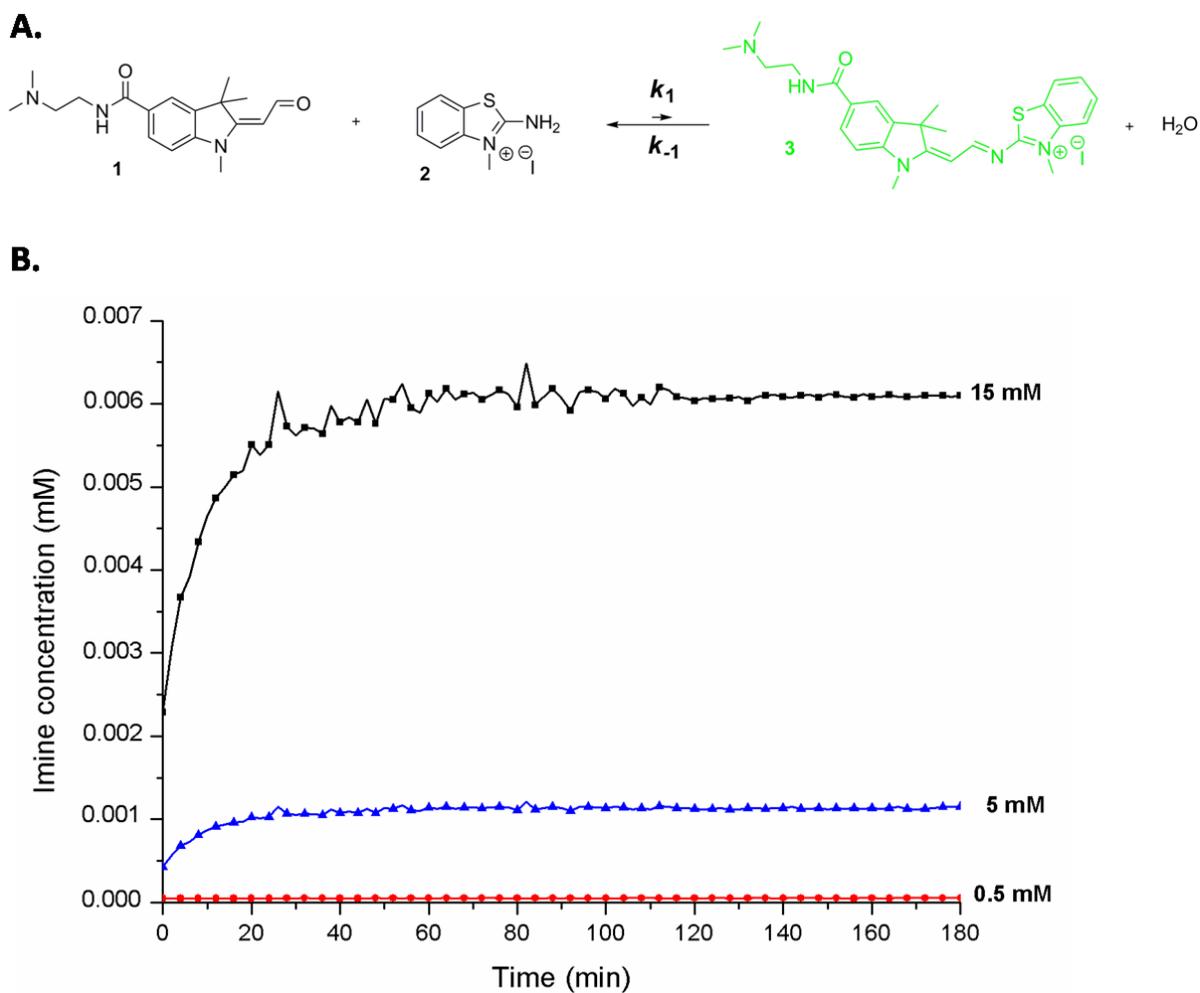


Figure 1. (A) Fluorogenic and reversible reaction of imine **3** formation from aldehyde **1** and amine **2**. (B) Kinetics of formation of imine **3** in bulk solvent (water, pH7) at 25°C from a stoichiometric mixture of 500 μ M (red), 5 mM (blue) and 15 mM (black) of both aldehyde **1** and amine **2**. Reactions were monitored by fluorescence spectroscopy ($\lambda_{em} = 520$ nm).

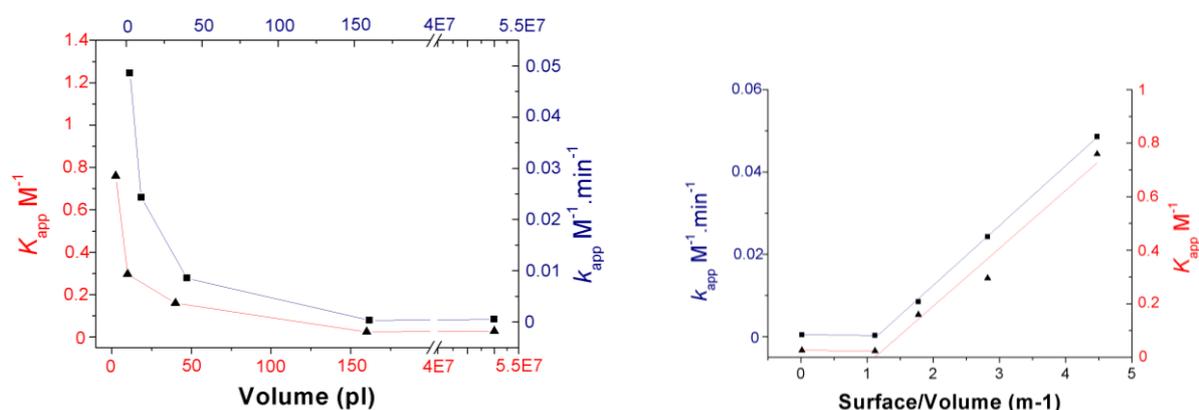
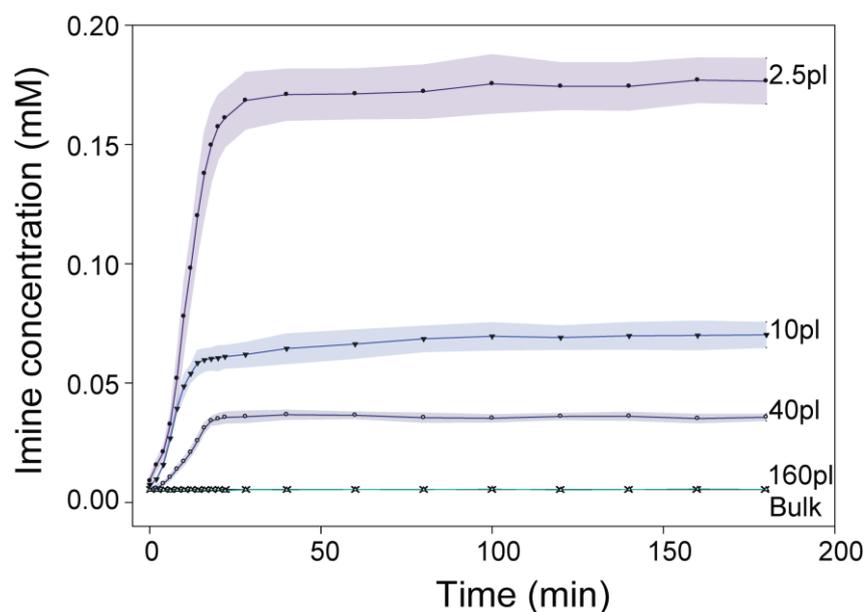
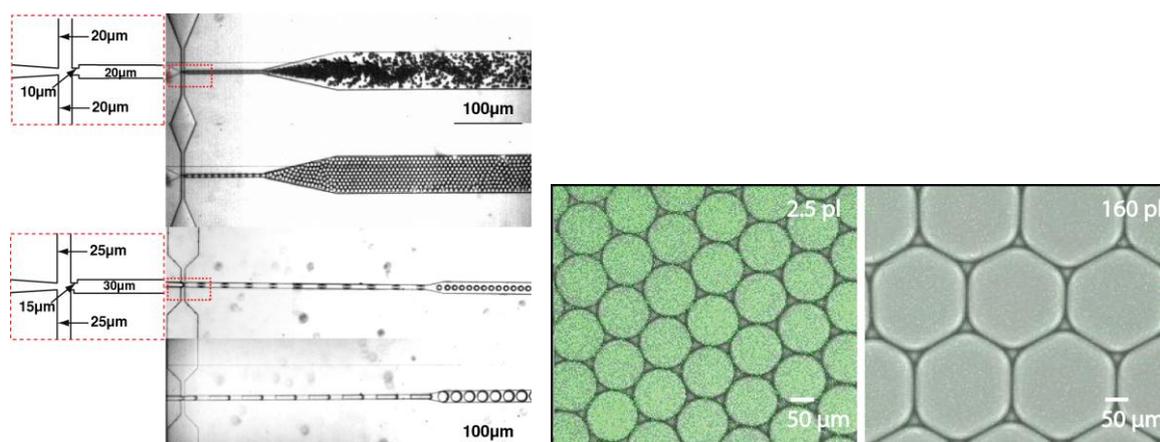


Figure 2. Kinetics and thermodynamics of imine **3** synthesis in droplets of different size. **(A)** Formation of droplets of 2.5, 10, 40 and 160 pl volume, containing a stoichiometric mixture of 15 mM aldehyde **1** and amine **2** in water (pH7), by flow focusing in microfluidic devices. The design of the nozzle (red box) for creation of 2.5 and 10 pl droplets (Device 1), and for the creation of 40 and 160 pl droplet (Device 2) is shown on the left. **(B)** Detection of imine **3** synthesized in 2.5 and 160 pl droplets after 2 hours incubation at 25°C by epifluorescence confocal microscopy ($\lambda_{em} = 520$ nm). The droplets are flattened beneath a cover slip in the

z-axis resulting in larger diameters in the x-y plane than calculated for spherical droplets **(C)** The droplets are flattened beneath a cover slip in the z-axis resulting in larger diameters in the x-y plane than calculated for spherical droplets. Kinetics of imine **3** formation in droplets of different volume, or in bulk at 25°C, monitored by fluorescence spectroscopy ($\lambda_{em} = 505$ to 550 nm). In each case 25 μ l of aqueous phase was monitored. **(D)** Plots of apparent second order rate constants (k_{app} , blue) and apparent equilibrium constants (K_{app} , red) versus volume and surface/volume ratio.

Table 1. Effect of droplet volume on the apparent second order rate constants (k_{app}) and apparent equilibrium constants (K_{app}) of the reaction of imine formation from stoichiometric amounts of amine and aldehyde (15 mM each).

Volume (pl)	Surface/Volume (m^{-1})	k_{app} ($M^{-1}min^{-1}$)	K_{app} (M^{-1})
2.5	4.47×10^{-3}	4.86×10^{-2}	0.759
10	2.82×10^{-3}	2.43×10^{-2}	0.296
40	1.78×10^{-3}	8.52×10^{-3}	0.159
160	1.12×10^{-3}	3.18×10^{-4}	2.37×10^{-2}
Bulk (50000000)	Bulk (0.016×10^{-3})*	1.54×10^{-3}	2.71×10^{-2}

*The surface to volume ratio of the bulk reaction was approximated assuming that the 50 μ l volume was spherical.

Table 2. Effect of surfactants on the equilibrium constants (K_{eq}) and second order rate constants (k_1) of the imine formation reaction.

	0.03 % w/w surfactant*		0.8 % w/w surfactant*	
	K_{eq} (M^{-1})	k_1 ($M^{-1}min^{-1}$)	K_{eq} (M^{-1})	k_1 ($M^{-1}min^{-1}$)
SDS	0.054	2.40×10^{-3}	1.520	0.0309
Triton X-100	0.026	5.89×10^{-4}	0.055	9.96×10^{-4}
CTAB	0.027	8.12×10^{-4}	0.048	7.56×10^{-4}
F-108	0.026	9.51×10^{-4}	0.041	9.83×10^{-4}
F-127	0.025	8.49×10^{-4}	0.043	1.16×10^{-3}
PVP	0.027	1.02×10^{-3}	0.045	5.90×10^{-4}
Zonyl	0.027	8.63×10^{-4}	0.065	7.62×10^{-4}
PAA250	0.028	7.58×10^{-4}	0.350	8.93×10^{-3}
PAA805	0.031	1.05×10^{-3}	0.140	2.95×10^{-3}

*Reactions were carried out with stoichiometric quantities of amine and aldehyde (15 mM each) and in the presence of 0.03% or 0.8% w/w surfactant and monitored by fluorescence spectroscopy ($\lambda_{em} = 520$ nm). The surfactants tested were sodium dodecylsulfate (SDS), Triton X-100, cetyl trimethylammonium bromide (CTAB), Pluronic F-108, Pluronic F-127, Polyvinylpyrrolidone (PVP), Zonyl and Polyacrylic acids (PAA250 and PAA805).

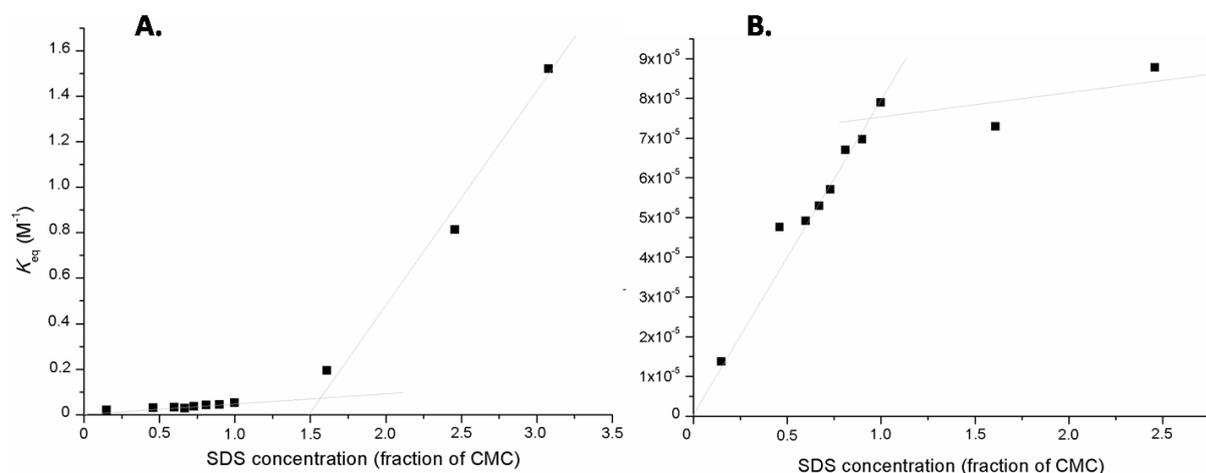


Figure 3. Kinetics and thermodynamics of imine **3** synthesis in bulk but in the presence of different concentrations of SDS. **(A)** Plot of the equilibrium constants (K_{eq}) calculated from the concentration of imine **3** formed at equilibrium ($t = 3\text{h}$) when reacting a stoichiometric mixture of amine **1** and aldehyde **2** (15 mM each) in the presence of various concentrations of SDS. **(B)** Plot of the second order rate constants (k_1) calculated from the initial rate of formation of imine **3** when reacting a stoichiometric mixture of amine **1** and aldehyde **2** (15 mM each) in the presence of various concentrations of SDS. SDS concentrations are given as a molar fraction of CMC. Experimental data were fitted by linear regression (grey lines). Reactions were monitored by fluorescence spectroscopy ($\lambda_{em} = 520\text{ nm}$).

References

1. M. Page, W. Jencks, *Proc. Natl. Acad. Sci. USA* **68**, 1678 (1971).
2. A. G. Cairns-Smith, Genetic takeover and the mineral origin of life. (Cambridge University Press, New York, 1982).
3. C. B. Thaxton, W. L. Bradley, R. L. Olsen, The mystery of life's origin. (Philosophical Library Inc., New York, 1984).
4. C. R. Woese, *J Mol Evol* **13**, 95 (1979).
5. F. Darwin, The life and letters of Charles Darwin. (John Murray, London, 1887), vol. 3.
6. J. B. S. Haldane, *Rationalist Annu.* **1929**, 148 (1929).
7. A. I. Oparin, Proiskhozhdenie zhizny. Izd. Mosk. Rabochii (Moscow, 1924).
8. C. M. Dobson, G. B. Ellison, A. F. Tuck, V. Vaida, *Proc Natl Acad Sci USA* **97**, 11864 (2000).
9. A. Tuck, *Surv Geophys* **23**, 379 (2002).
10. G. Adam, M. Delbrück, in *Structural chemistry and molecular biology*, A. Rich, N. Davidson, Eds. (W.H. Freeman and Co., New York, 1968), pp. 198-215.
11. J. Bernal, *Proc. Phys. Soc. B* **62**, 597 (1949).

12. G. Wächtershäuser, *Microbiol Rev* **52**, 452 (1988).
13. S. Narayan *et al.*, *Angew Chem Int Edit* **44**, 3275 (2005).
14. R. U. Lemieux, *Accounts Chem Res* **29**, 373 (1996).
15. S. K. Pal, J. Peon, A. H. Zewail, *Proc Natl Acad Sci USA* **99**, 1763 (2002).
16. L. R. Pratt, A. Pohorille, *Chem Rev* **102**, 2671 (2002).
17. D. N. Shin, J. W. Wijnen, J. B. F. N. Engberts, A. Wakisaka, *J Phys Chem B* **105**, 6759 (2001).
18. D. N. Shin, J. W. Wijnen, J. B. F. N. Engberts, A. Wakisaka, *J Phys Chem B* **106**, 6014 (2002).
19. J. T. Koberstein, *J Polym Sci Pol Phys* **42**, 2942 (2004).
20. K. Meguellati, M. Spichy, S. Ladame, *Org Lett* **11**, 1123 (2009).
21. S. L. Anna, N. Bontoux, H. A. Stone, *Applied Physics Letters* **82**, 364 (2003).
22. C. Holtze *et al.*, *Lab on a chip* **8**, 1632 (2008).
23. E. H. Cordes, *Pure Appl Chem* **50**, 617 (1978).
24. T. Dwars, E. Paetzold, G. Oehme, *Angew Chem Int Edit* **44**, 7174 (2005).
25. F. Leal-Calderon, V. Schitt, J. Bibette, *Emulsion Science. Basic Principles*. (Springer Verlag, New York, ed. 2nd, 2007).

Enhanced efficiency of synthetic organic reactions in picolitre volume droplets: Supplementary Online Material

*To whom correspondence should be addressed. E-mail: griffiths@unistra.fr

MATERIALS AND METHODS

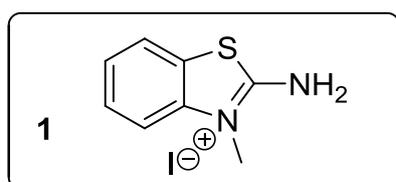
Synthesis

General Experimental

Reagents and solvents were purchased from Sigma-Aldrich and used without further purification. Compounds were characterized using ^1H and ^{13}C NMR that were recorded on a Bruker Avance DRX 400 spectrometer at 400 and 100.6 MHz, respectively. Chemical shifts are reported as δ values (ppm) with reference to the residual solvent peaks. Products were purified by flash column chromatography on silica gel.

Synthesis of amine and aldehyde building blocks

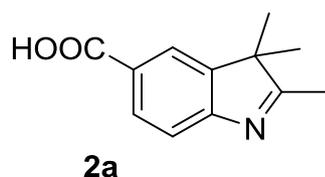
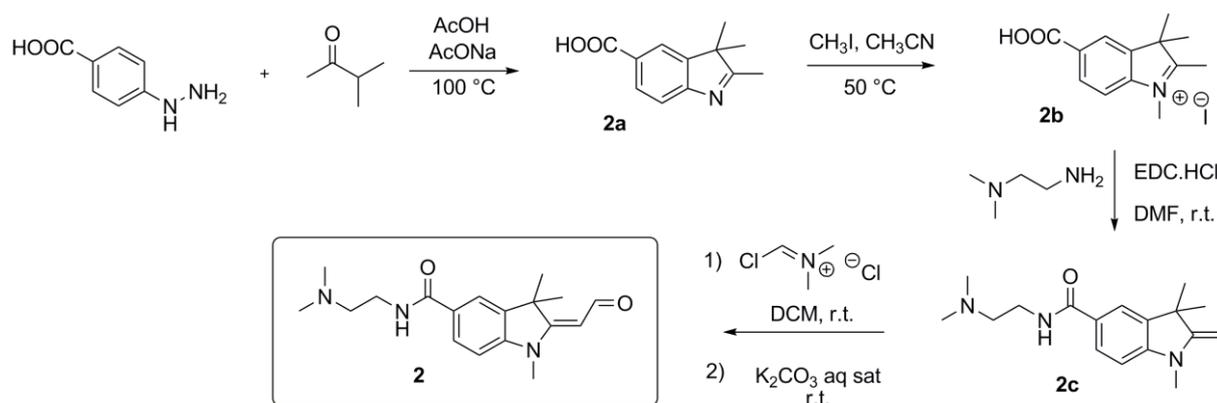
Amine synthesis



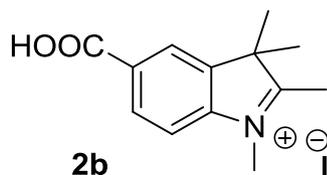
2-Aminobenzothiazole (3.5 g, 23 mmol) was dissolved in acetonitrile (50 mL) and iodomethane (2 mL) was added. The solution was stirred at 45 °C for 12 hours. After cooling to room temperature, Et₂O (50 mL) was added and the white suspension was filtered. The precipitate was washed with diethylether and dried under vacuum. Compound **1** was obtained pure as a white solid (6.75 g, 100%, mp 227-228 °C). ^1H NMR (400 MHz, DMSO) δ 10.0 (s, 2 H), 8.0 (d, $J = 7.9$ Hz, 1 H), 7.69 (d, $J = 7.9$ Hz, 1 H), 7.6 (t, $J = 7.5$ Hz, 1H), 7.44 (t, $J = 7.5$ Hz, 1 H), 3.75 (s, 3 H). ^{13}C NMR (100 MHz, DMSO) δ 167.9, 138.9, 127.7, 125.1, 123.4,

122.2, 122.1, 113.3, 32.2. HRMS (positive ES): m/z : calcd for $C_8H_9N_2S^+$: 165.0481; found 165.0475.

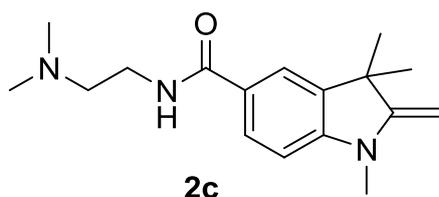
Synthesis of water-soluble Fisher's base aldehyde **2**



4-Hydrazinobenzoic acid (5.00 g, 32.9 mmol), 3-Methyl-2-butanone (4.24 g, 5.30 mL, 49.5 mmol) and NaOAc (5.40 g, 66 mmol) were dissolved in AcOH (70 mL) and the resulted brown mixture was heated to 100°C and stirred at the same temperature for 16 h. After cooling the reaction mixture to room temperature, acetic acid was removed under reduced pressure, cooled the resulted residue to 0°C and a saturated solution of K_2CO_3 (200 mL) was added slowly. The aqueous layer was extracted with CH_2Cl_2 (3 x 200 mL) before it was acidified to pH = 4 with conc. HCl at 0°C. The aqueous layer was extracted with CH_2Cl_2 (3 x 200 mL). The combined organic layers were dried over Na_2SO_4 , filtered and concentrated in *vacuo* to give the title compound **2a** (6.5 g, 97%) as a brown oil. The crude product was directly used in the next step without further purification. 1H NMR (400 MHz, DMSO): δ = 7.99 (d, J = 1.6 Hz, 1H), 7.91 (dd, J = 8.0, 1.6 Hz, 1H), 7.50 (d, J = 8.0 Hz, 1 H), 2.25 (s, 3H), 1.27 (s, 6H). ^{13}C NMR (100 MHz, DMSO): δ = 191.7, 167.5, 157.4, 146.1, 129.6, 127.3, 122.7, 119.1, 53.5, 22.3 (x 2), 15.3. HRMS (negative ES) m/z : calcd for $C_{12}H_{13}NO_2^-$ 202.086; found 202.084



To a solution of compound **2a** (6.50 g, 32.0 mmol) in acetonitrile (100 mL) iodomethane (11.3 g, 5.00 mL, 80 mmol) was added at room temperature and the resulted mixture was heated to 45°C and stirred at the same temperature for 12 h. After cooling to room temperature, Et₂O (100 mL) was added and the precipitate was filtered off. The precipitate was washed with Et₂O (3 x 30 mL) and dried under vacuum. Compound **2b** (8.03 g, 72%) was obtained as a red solid. ¹H NMR (400 MHz, DMSO): d = 8.38 (s, 1H), 8.18 (d, *J* = 8.3 Hz, 1H), 8.02 (d, *J* = 8.5, 1H), 3.99 (s, 3H), 2.81 (s, 3H), 1.57 (s, 6H). ¹³C NMR (100 MHz, DMSO): d = 199.0, 166.4, 145.2, 141.9, 131.6, 130.3, 124.2, 115.3, 54.2, 35.0, 21.5 (x 2), 14.5. HRMS (positive ES) *m/z*: calcd for C₁₃H₁₆NO₂⁺ 218.118; found 218.114



Compound **2b** (1.00 g, 2.90 mmol), *N,N*-dimethylethylenediamine (358 mg, 443 μL, 4.06 mmol) and EDC·HCl (610 mg, 3.19 mmol) were dissolved in anhydrous DMF (45 mL) and the resulted red colored solution was stirred at room temperature for 10 hours. DMF was then evaporated off under reduced pressure and the residue was dissolved in CH₂Cl₂ (200 mL). The organic layer was washed with a saturated solution of K₂CO₃ (2 x 100 mL), dried over Na₂SO₄, filtered and concentrated in *vacuo*. The residue was purified by column chromatography on SiO₂ (EtOAc/NEt₃ 50:1, EtOAc/MeOH/NEt₃ 200:5:2, 200:10:2; R_f = 0.3) to give the title compound **2c** (510 mg, 61 %) as a red oil. ¹H NMR (400 MHz, DMSO): d = 8.06 (t, *J* = 5.7 Hz, 1H), 7.68-7.66 (m, 2H), 6.69 (d, *J* = 8.0 Hz, 1H), 3.96 (d, *J* = 1.8 Hz, 1H), 3.95 (d, *J* = 1.8 Hz, 1H), 3.32 (q, *J* = 6.6 Hz, 2H), 3.05 (s, 3H), 2.37 (t, *J* = 7.0 Hz, 2H), 2.16 (s, 6H), 1.29 (s, 6H). ¹³C NMR (100 MHz, DMSO): d = 166.1, 161.5, 148.4, 136.7, 127.9, 124.5, 120.9, 104.4, 75.8, 58.5, 45.3 (x 2), 43.2, 37.3, 29.6 (x 2), 28.6. HRMS (positive ES) *m/z*: calcd for C₁₇H₂₆N₃O⁺: 288.207; found 288.206

Imine quantification

Imine quantification was achieved by monitoring simultaneously, and at a given reactant concentration (10 mM aldehyde + 10 mM amine), the formation of the fluorescent imine ($\lambda_{\text{exc}} = 480$ nm and $\lambda_{\text{em}} = 520$ nm) and the disappearance of the weakly fluorescent aldehyde ($\lambda_{\text{exc}} = 340$ nm and $\lambda_{\text{em}} = 400$ nm). Briefly, a mixture of amine **1** and aldehyde **2** (10 mM each) was incubated in 0.1% w/w SDS solution at 25°C. After 2h, equilibrium was reached which corresponded to a 5% conversion of aldehyde **1**. Since aldehyde **1** can only be converted into imine **3**, it was possible to determine the concentration of imine **3** at equilibrium as being 0.5 mM (5% of 10 mM). Since the fluorescence intensity for imine **3** ($\lambda_{\text{exc}} = 480$ nm and $\lambda_{\text{em}} = 520$ nm) at equilibrium reached 450 Rfu, we concluded that, under the conditions of our experiment, 0.5 mM of Imine **3** corresponds to 450 Rfu (see **Fig. S1**).

Surfactant effect

The effect of different surfactants on the kinetics and thermodynamics of the reaction of imine formation were determined by fluorescence spectroscopy. A stoichiometric mixture of amine **1** and aldehyde **2** (15 mM each) in water were reacted in the presence of different concentrations (0.03% w/w and 0.8% w/w) of various surfactants (CTAB, Triton X-100, SDS, F-127, F-108, Zonyl, PVP, PAA250, PAA805). The reaction time course were determined by fluorescence spectroscopy ($\lambda_{\text{exc}} = 480$ nm and $\lambda_{\text{em}} = 520$ nm) over three hours (when equilibrium was reached).

The ability of each surfactant to form micelles at a 0.8% w/w concentration was confirmed by DLS studies using a Zetasizer Nano-S instrument from Malvern Instruments (**Fig. S2**). Under these conditions, all surfactants formed detectable micelles except for F-108 and PAA250 and PAA805.

Microfluidic devices

Microfluidic devices were fabricated by patterning 10 μm deep channels for the 2.5 pl and 10 pl droplet generator device (**Fig. S3a**), 25 μm deep channels for 40 pl and 160 pl droplet generator device (**Fig. S3b**) in poly(dimethylsiloxane) (PDMS) using soft lithography (*1*). The effect of different surfactants on the kinetics and thermodynamics of the reaction of imine formation were determined by fluorescence spectroscopy. Briefly, a mould of SU-8 resist (MicroChem Corp.) was fabricated on a silicon wafer (Siltronix) by UV exposure (MJB3 contact mask aligner; SUSS MicroTec) through a photolithography mask (Selba SA) and

subsequent development (SU-8 developer; MicroChem Corp.). Curing agent was added to PDMS base (Sylgard 184 silicone elastomer kit; Dow Corning Corp.) to a final concentration of 10% (v/v), mixed, and poured over the mould to a depth of 5 mm. Following cross-linking at 65°C for ~12 hours, the PDMS was peeled off the mould and the input and output ports were punched with a 0.75 mm-diameter Harris Uni-Core biopsy punch (Electron Microscopy Sciences). Particles of PDMS were cleared from the ports using pressurized nitrogen gas. The structured side of the PDMS slab was bonded to a 76 x 26 x 1 mm glass microscope slide (Paul Marienfeld GmbH & Co. KG) by exposing both parts to an oxygen plasma (PlasmaPrep 2 plasma oven; GaLa Instrumente GmbH) and pressing them together. The device was flushed with a solution of 1% of 1H, 1H, 2H, 2H-Perfluorodecyltrichlorosilane (ABCR) in the fluorinated oil HFE7500 (3M), then flushed with HFE7500, and finally with nitrogen.

Liquids were pumped into microfluidic devices using Standard-Pressure Infuse/Withdraw PHD 22/2000 syringe pumps (Harvard Apparatus Inc.). Syringes were connected to the microfluidic device using 0.6 × 25 mm Neolus needles (Terumo Corporation) and PTFE tubing with an internal diameter of 0.56 mm and an external diameter of 1.07 mm (Fisher Bioblock Scientific).

Reactions in droplets

The effect of compartmentalization in droplets of different sizes on the kinetics and thermodynamics of the reaction of imine **3** formation were determined by fluorescence spectroscopy. Two 30 mM solutions of compound **1** and **2** were heated at 40°C for compound **1** (which is relatively heat sensitive) and 70°C for compound **2** to avoid any precipitates. The two solutions were loaded into separate syringes, then injected onto the microfluidic device at the same flow rate. On chip, the two solutions were brought together, co-flowed and then compartmentalized into droplets by flow focusing (2) of the aqueous phase with a fluorinated oil phase (HFE7500, 3M) containing a PEG-PFPE tri-block copolymer surfactant (3). In order to produce different droplet we used two different devices. For 2.5 pl and 10 pl droplets we used device 1 and for 40 pl and 160 pl droplets we used device 2 (**Fig. S3**). Each emulsion was prepared with 25 µl of aqueous phase during 15 min and collected in a 384-well plate. The plate was centrifuged at 100rpm for one minute in order to accelerate the creaming of the emulsion. Four emulsions were produced in the exact same conditions and the fluorescence measured with the fluorescence plate reader over 3 hours at 25°C. Each droplet size was assayed in triplicate.

Confocal microscopy imaging

Droplets prepared using the microfluidic devices were placed between two 0.17 mm thick microscope coverslips and images were acquired with a Zeiss (Jena, Germany) LSM510 laser-scanning confocal microscope equipped with a C-Apochromat 20x (n.a. 0.8) objective. Imine **3** was excited at 488 nm using an argon laser, and emission spectra were recorded from 505 to 550 nm. Images were processed with the Zeiss LSM Image Browser Version 2.50.0929 software.

SUPPLEMENTARY FIGURES

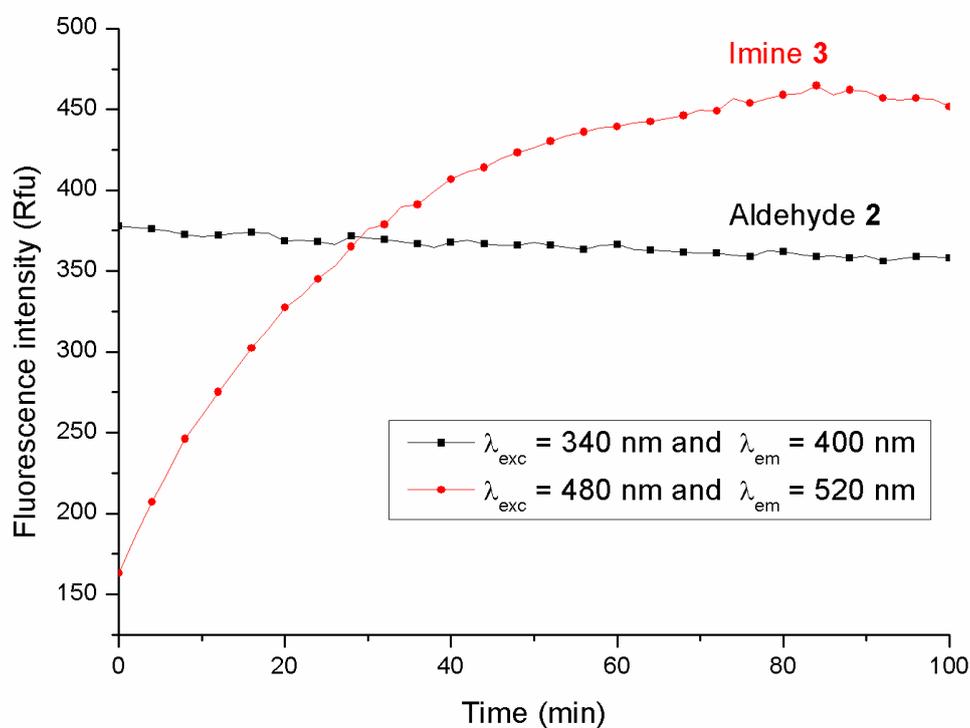


Figure S1. Formation of imine **3** and disappearance of aldehyde **2**, monitored by fluorescence spectroscopy.

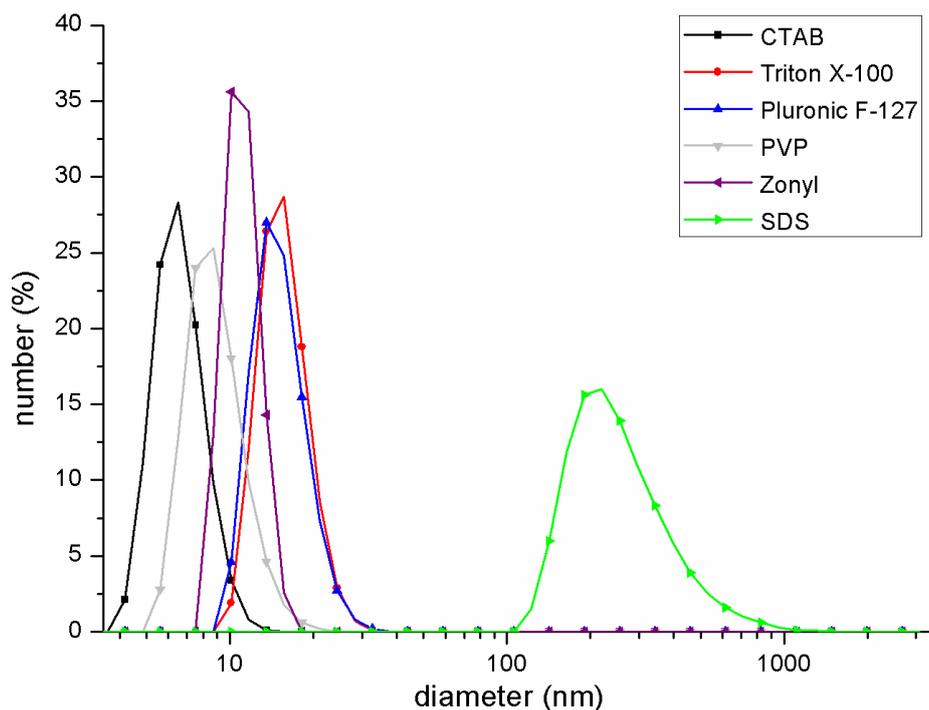


Figure S2.

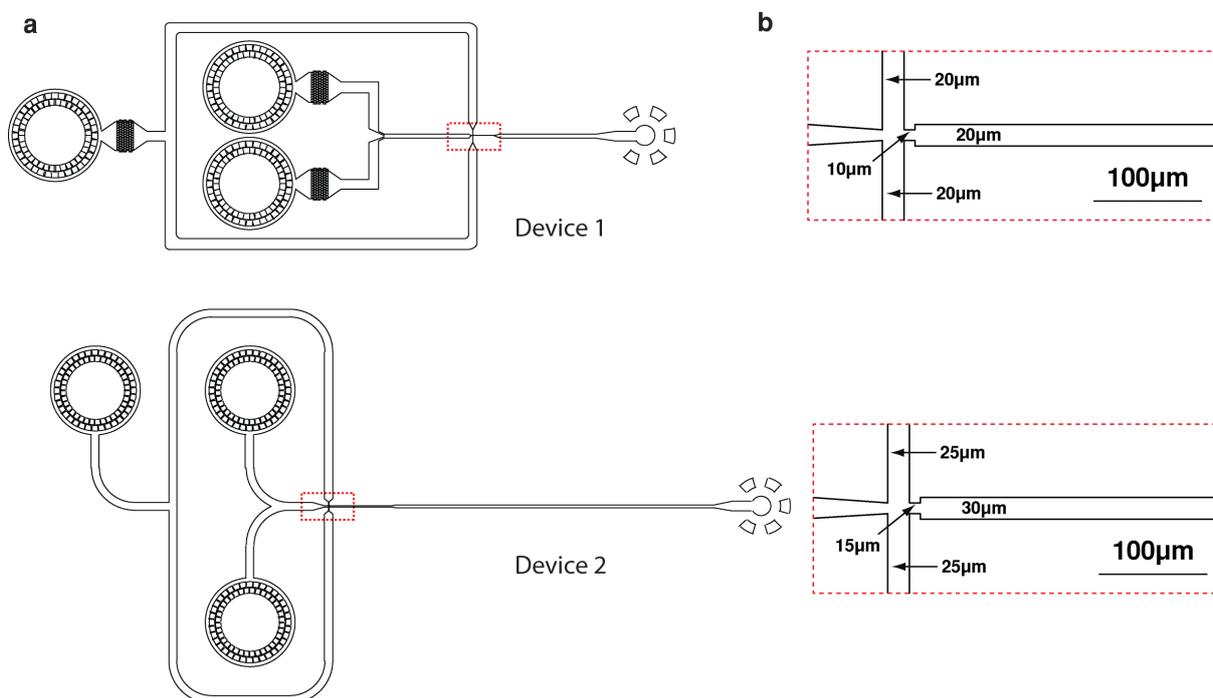


Figure S3. Design of the droplet generation devices. The devices contained one oil and two aqueous inlets (Aq 1 and Aq2) (A). Droplets were generated at a flow-focussing junction (red square) using HFE7500 fluorinated oil containing 2% (w/w) PEG-PFPE tri-block copolymer surfactant (3). 2.5 pl droplets (~17 μm diameter) and 10 pl droplets (~27 μm diameter) were

generated using device 1 (depth 10 μm). 40 pl droplets (~ 42 μm diameter) and 160 pl droplets (~ 67 μm diameter) were generated using device 2 (depth 25 μm). **(B)** Magnified view of the flow-focusing junction used to produce droplets (red box). The corresponding movies of droplet production are available (**Movies S1 to S4**).

SUPPLEMENTARY REFERENCES

1. D. C. Duffy, J. C. McDonald, O. J. A. Schueller, W. G.M., *Anal Chem* **70**, 4974 (1998).
2. S. L. Anna, N. Bontoux, H. A. Stone, *Appl Phys Lett* **82**, 364 (2003).
3. C. Holtze *et al.*, *Lab on a chip* **8**, 1632 (2008).

CONCLUSION

PERSPECTIVES

We have developed fluorescent and fluorogenic probes which have multiple applications in modern bio research. We have demonstrated the potential of these systems for sensing nucleic acid structures (e.g. DNA G-quadruplexes, DNA hairpins), nucleic acid sequences (detection of single mutations) and biometabolites (e.g. Malondialdehyde). Most systems rely on the templated synthesis of cyanine dyes (or cyanine dye imino analogues) via either an aldolisation-elimination irreversible reaction or a reversible reaction of imine formation.

Of particular interest is the original family of imine-containing cyanine dyes obtained by reaction between a Fisher's base aldehyde and a 2-amino-benzothiazolium derivative. These compounds were shown to form under reversible conditions compatible with the concept of dynamic combinatorial chemistry (DCC). This fluorogenic system is also responsive to external stimuli and we demonstrated its potential use for monitoring reactions within water-in-oil microdroplets.

Dynamic cyanine dyes were also obtained from the reversible reaction between two molecules of 2-amino-benzothiazolium derivative and one molecule of malondialdehyde (MDA). We demonstrated that these compounds would serve as prodrugs for the selective alkylation of cytosines *in vitro*. Further work will be required to demonstrate whether the same strategy can be applied *in vivo*.

We also used TD-DFT calculations to predict the maximum absorption wavelength of these modified cyanine dyes as well as that of the classical dyes. A very good correlation between theory and calculation was obtained by introducing a zero-point vibrational energy term.

Finally, we have demonstrated that the retro-Knoevenagel reaction was reversible and suitable for the concept of DCC. We have proven the reversibility of this C=C exchange reaction and in the future we plan to use this chemistry for developing new dynamic polymers.