Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2021, **19**, 8947

Duplex *vs.* folding: tuning the self-assembly of synthetic recognition-encoded aniline oligomers†

Daniele Rosa-Gastaldo,[a] Vytautas Pečiukėnas,[b] Christopher A. Hunter *[b] and Luca Gabrielli *[a]

Received 21st July 2021,
Accepted 28th September 2021

DOI: 10.1039/d1ob01882e

rsc.li/obc

One of the challenges in the realization of synthetic oligomers capable of sequence-selective duplex formation is intramolecular folding interaction between complementary recognition units. To assess whether complementary hetero-oligomers can assemble into high fidelity duplex structures, the competing folding equilibria must be carefully considered. A family of recognition-encoded aniline oligomers were assembled *via* reductive amination of dianiline linkers and dialdehyde monomers, which were equipped with either a 2-trifluoromethylphenol or a phosphine oxide H-bond recognition unit. To test the possibility of 1,2-folding in mixed sequence oligomers, the self-assembly properties of the homo- and hetero-dimers were characterised by ^{19}F and ^{1}H NMR titration and dilution experiments in toluene and in chloroform. Three different systems were investigated with variations in the steric bulk around the H-bond acceptor unit and the length of the dianiline linker. For two systems, the hetero-dimers folded with intramolecular H-bonding in the monomeric state, reducing stability of the intermolecular duplex by two to three orders of magnitude compared with the corresponding homo-oligomers. However, the use of a long rigid linker as the backbone connecting two monomer units successfully prevents 1,2-folding and leads to the formation of a stable mixed sequence duplex in toluene.

Introduction

Nature accomplishes exquisite control of the macroscopic properties of molecular systems by encoding information into matter. The sequence of monomer units in linear copolymers determines the three-dimensional structures and functional properties of biomacromolecules. Duplex formation is the key architecture that allows Nature to store, copy and translate information,[1] and intramolecular folding of single stranded oligomers is responsible for receptor and catalyst activity. The possibility to combine both duplex-forming and folding properties gives nucleic acids unique properties that can be used in directed evolution experiments to obtain functional polymers.[2,3]

Inspired by nucleic acids, a number of synthetic supramolecular systems that form duplex structures *via* non-covalent interactions have been developed.[4–6] A modular approach for the design of synthetic molecules that form duplexes using multiple cooperative H-bonding interactions allows the independent optimisation of the three key design elements: the chemistry used for the synthesis, the conformational properties of the backbone, and the recognition modules used for base-pairing.[7] We recently reported a two-component design, in which oligomers were synthesised through reductive amination between aldehyde monomers and dianiline linkers.[8] Information can be encoded in an oligomer as a sequence of H-bond donor (trifluoromethylphenol, D) and acceptor (phosphine oxide, A) units, and this H-bonded base-pair is sufficiently stable to allow duplex formation in non-polar solvents like toluene and in more competitive solvents like chloroform.[7h,8] The observation of imine polymerase activity in one of these single-stranded oligomers suggests that more interesting analogies between natural biopolymers and the chemistry of synthetic recognition-encoded oligomers will come to light.[9] To date, duplex formation in this system has only been studied for homo-oligomers.[8] Here we explore duplex formation in mixed sequence oligomers and take advantage of the two-component design to tune the interplay of folding and duplex formation.

In mixed-sequence oligomers the possibility of intramolecular H-bonding can lead to folding equilibria that compete with duplex formation (Fig. 1). If the number of H-bonds formed in the folded state and in the duplex state are identical, the intramolecular process will dominate. Therefore,

[a]*Department of Chemistry, University of Padova, via Marzolo 1, Padova, 35131, Italy. E-mail: luca.gabrielli@unipd.it*
[b]*Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: herchelsmith.orgchem@ch.cam.ac.uk*

†Electronic supplementary information (ESI) available: Materials and methods, synthetic procedures, full characterization of all compounds and NMR dilution and titration experiments. See DOI: 10.1039/d1ob01882e

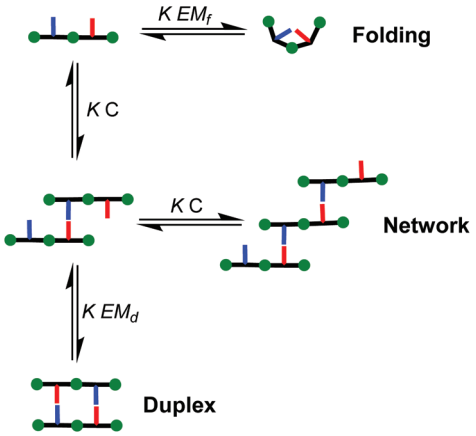


Fig. 1 Competing self-assembly channels mixed sequence oligomers. Synthesis is based on reductive amination of dianiline linkers (green) with dialdehydes monomer building blocks (black) bearing recognition units (blue and red). The first base-pairing interaction can take place intramolecularly leading to the folding channel or in an intermolecular way. The second base-pairing interaction can be intramolecular to initiate duplex formation or intermolecular, leading to the higher order complexes. The outcome depends on the concentration, C, the association constant for the intermolecular base-pairing interaction, K, and the effective molarities for folding, EM_f and duplex initiation EM_d.

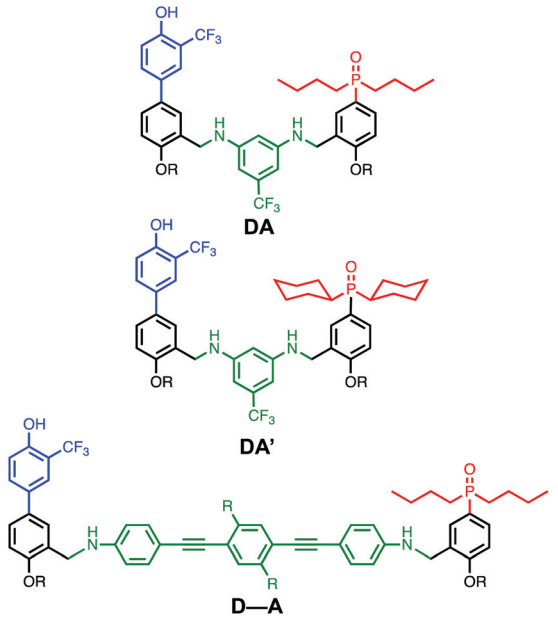


Fig. 2 Structures of hetero-dimers with a short-long phenol·phosphine oxide recognition system (**DA**), with a sterically hindered phosphine oxide (**DA'**), and with an extended dianiline linker (**D–A**). (R = 2-ethylhexyl.)

if two adjacent recognition units in an oligomer can form a base-pair, the folding pathway will be favoured over the duplex assembly channel. The key parameters that determine the distribution of the different species are the association constant for the intermolecular interaction between two complementary H-bonding sites (K), the operating concentration (C), and the effective molarities for folding (EM_f) and duplex formation (EM_d). The competing network channel that leads to supramolecular polymerisation can be avoided by operating at a concentration C lower than EM_d. The undesired 1,2-folding channel can be reduced by minimising EM_f, for example by increasing the rigidity of the backbone.[7g] However, when a rigid backbone is used, the precise geometry becomes critical for allowing extended duplex propagation. Since the design of rigid backbones is not straightforward, it would be preferable to adopt a more flexible system. A second strategy for preventing 1,2-folding is inspired by the size difference between purines and pyrimidines used in the base-pairing system in nucleic acids. Indeed, by using two recognition units of different length, it was possible to minimise EM_f in oligoesters with a very flexible backbone.[7h]

The same long-short phenol-phosphine oxide base-pair has been adopted in the aniline oligomers that are the subject of this work. The self-assembly properties of the homo-oligomers were reported previously, and length-complementary oligomers were found to form stable H-bonded duplexes in chloroform solution. Here we investigate the competing equilibria between duplex formation and 1,2-folding in hetero-oligomers using different dialdehyde monomers and dianiline linkers. The use of a long rigid linker in the backbone is found to effectively hold adjacent recognition sites apart and abolish 1,2-folding, leading to stable duplex formation between hetero-oligomers.

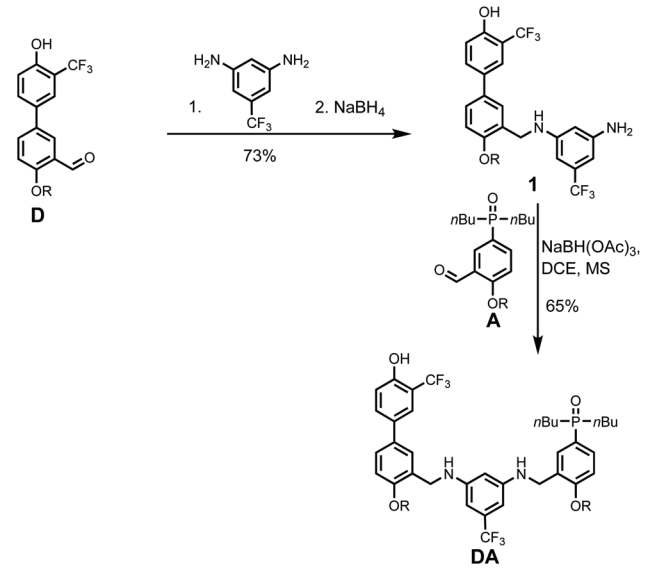
Fig. 2 shows the three different hetero-dimers that were targeted in this work. **DA** is the hetero-dimer of the previously reported oligomers. In **DA'**, the *n*-butyl groups present on the H-bond acceptor unit of **DA** are substituted with bulkier cyclohexyl moieties. In **D–A**, an extended dianiline linker is used instead of the shorter 1,3-phenylenediamine. Comparison of the properties of these systems with the corresponding homo-dimers allows precise characterisation of the folding behaviour of these different oligomer architectures.

Results and discussion

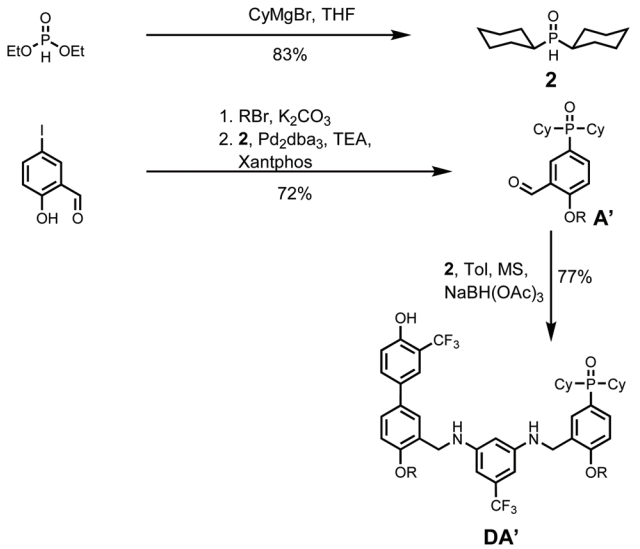
Synthesis

The synthetic route to the **DA** dimer is shown in Scheme 1. We recently reported the synthesis of mono-aldehyde building blocks **A** and **D**.[8] Treatment of the mono-aldehyde donor (**D**) with five equivalents of 5-(trifluoromethyl)-1,3-phenylenediamine gave the imine, which was reduced with $NaBH_4$ to give **1**. Subsequent reaction with acceptor monomer **A**, followed by portion-wise addition of $NaBH(OAc)_3$ as the reducing agent gave the desired **DA** dimer.

The synthesis of the **DA'** dimer, which was equipped with a more hindered phosphine oxide, is shown in Scheme 2. Treatment of diethyl phosphite with cyclohexylmagnesium bromide gave **2**. The mono-aldehyde phosphine oxide **A'** was synthesised by alkylating 5-iodosalicylaldehyde with racemic 2-ethyl-



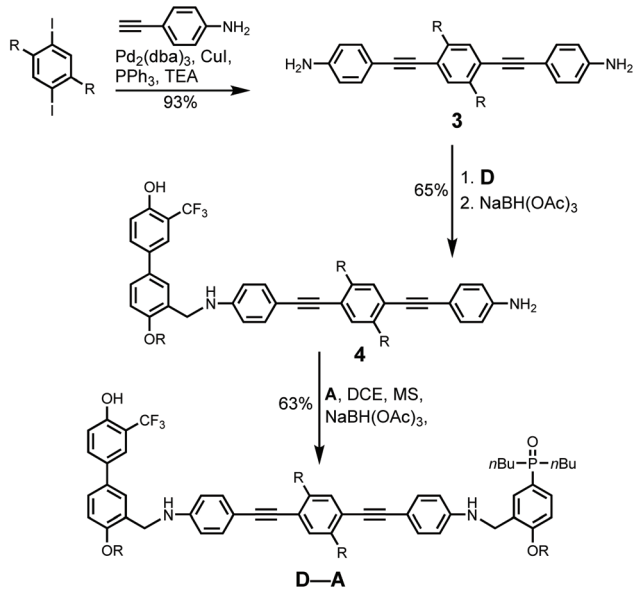
Scheme 1 Synthesis of **DA** (R = 2-ethylhexyl).



Scheme 2 Synthesis of **DA′** (R = 2-ethylhexyl).

hexyl bromide, followed by coupling with **2**, using palladium catalyst and XantPhos. Reaction of **A′** with aniline **1**, followed by portion-wise addition of $NaBH(OAc)_3$ gave the desired **DA′** dimer.

The backbone modification was performed substituting the 5-(trifluoromethyl)-1,3-phenylenediamine linker with the longer dianiline **3**, which was prepared by coupling 4-ethynylaniline with racemic 1,4-bis(2-ethylhexyl)-2,5-diiodobenzene using palladium, triphenylphosphine and copper(I) iodide to give **3** (Scheme 3). Compound **3** was equipped with two additional solubilising groups, in order to ensure solubility in non-competing solvents such as toluene. These groups are remote from the recognition units and are unlikely to affect duplex formation. In all cases, racemic 2-ethylhexyl solubilising groups were used, because the diastereoisomeric mixture



Scheme 3 Synthesis of heterodimer **D–A** (R = 2-ethylhexyl).

obtained improves solubility.[10] Treatment of donor aldehyde **D** with five equivalents of dianiline **3**, gave the corresponding mono-functionalised imine, which was reduced with $NaBH_4$ to give compound **4**. Subsequent reaction with acceptor monomer **A**, followed by portion-wise addition of $NaBH(OAc)_3$, gave the desired **D–A** dimer.

The corresponding homo-dimers are needed as reference systems to evaluate the folding properties of the hetero-dimers. Homo-dimers **DD** and **AA** were reported previously (Fig. 3).[8] Compound **A′A′** was synthesised *via* reductive amination of **A′** with 5-(trifluoromethyl)-1,3-phenylenediamine and $NaBH(OAc)_3$ (Scheme 4). Similarly, the homo-dimers equipped with the longer dianiline linker, **A–A** and **D–D**, were synthesised by treating the corresponding donor (**D**) or acceptor (**A**) aldehyde with dianiline **3**, followed by portion-wise addition of $NaBH(OAc)_3$ (Scheme 4).

NMR binding studies

Complexation of the complementary homo-dimers and self-association of the hetero-dimers were studied in both *d*-chloroform and *d*-toluene through ^1H and ^{19}F NMR titration and dilution experiments. Fig. 4 shows representative ^1H-NMR

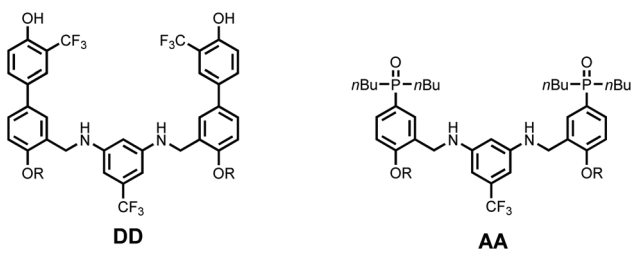
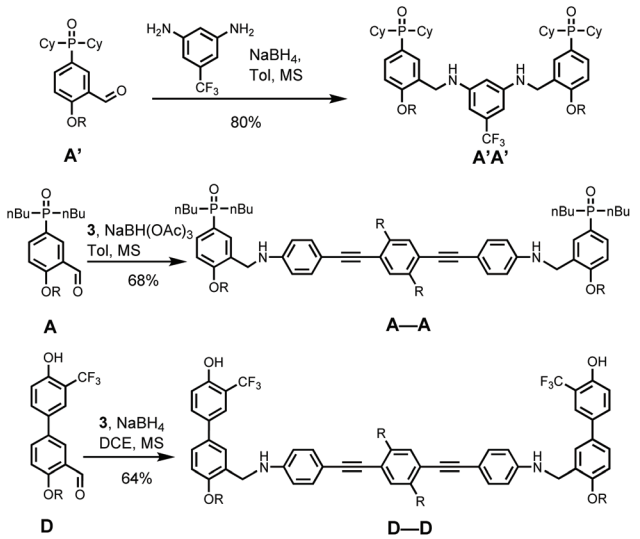


Fig. 3 Structure of **AA** and **DD**. (R = 2-ethylhexyl.)



Scheme 4 Synthesis of **A'A'**, **A—A** and **D—D**. (R = 2-ethylhexyl.)

spectra for the dilution of hetero-dimer **DA** and for the titration of homo-dimer **AA** into **DD** in *d*-chloroform. In the titration experiment, the signal due to the **DD** phenol OH protons (highlighted in blue in Fig. 4b) moves from below 6 ppm to 10 ppm when **AA** is added. This change in chemical shift is characteristic of H-bond formation between the phenol and phosphine oxide recognition units, and the magnitude of the change suggests that **AA** and **DD** form a duplex with two fully bound H-bonds. In contrast, the signal due to the phenol OH proton of **DA** (highlighted in blue in Fig. 4a) appears at a chemical shift close to 10 ppm when it is in the free monomeric state at very low concentrations (0.1 mM). At higher concentrations, only a small increase in the chemical shift of the OH signal was observed. This result suggests that the two recognition units interact in the monomeric state of **DA** and the intramolecular H-bond leads to 1,2-folding. At higher concentrations, intermolecular interactions can compete with the intramolecular interactions, and the increase in the chemical shift of the OH signal suggests formation the **DA·DA** duplex or higher order complexes.

Changes observed in the appearance of the ^{1}H NMR signals due the 1,3-dianiline linker (highlighted in green in Fig. 4a) confirm this hypothesis. At low concentrations when **DA** is in the folded monomeric state, the linker signals are superimposed at around 6.2 ppm, but on increasing of the concentration, three distinct signals are observed for the linker protons. The appearance of the **DA** linker signals at high concentrations (100 mM) is very similar to the distinct linker signals observed for the duplex in the homo-dimer titration experiment (highlighted in green in Fig. 4b). This observation suggests that **DA** populates the duplex conformation at high concentrations, but at low concentrations, there is a change in backbone conformation, which is consistent with 1,2-folding.

The titration data for all combinations of complementary homo-dimers fit well to 1 : 1 binding isotherms, and the dilution data for all of the hetero-dimers fit well to dimerization isotherms. Table 1 summarises the results. The association constants for formation of a single intermolecular H-bond were also measured by carrying out ^{1}H and ^{19}F NMR titrations using the aldehydes **D**, **A** and **A'**.

The association constants of complementary monomers ($K_{D\cdot A}$) and homo-dimers ($K_{DD\cdot AA}$) were used to determine the effective molarity for duplex formation EM_d.

$$EM_d = \frac{K_{DD\cdot AA}}{2K_{D\cdot A}{}^2} \quad (1)$$

We have reported that the **DD·AA** can assemble into stable duplexes in chloroform with an EM_d of 35 mM.[8] However in the case of the heterodimer **DA**, there is the possibility of 1,2-folding, so the observed dimerization constant K_{obs} depends on the relative populations of the open ($\mathbf{DA}_{open}$) and folded ($\mathbf{DA}_{fold}$) conformations present in the single-stranded monomeric state. The concentration of $\mathbf{DA}_{open}$ can be expressed as a function of the equilibrium constant for folding K_{fold}:

$$[DA] = [DA]_{open}\ (1 + K_{fold}) \quad (2)$$

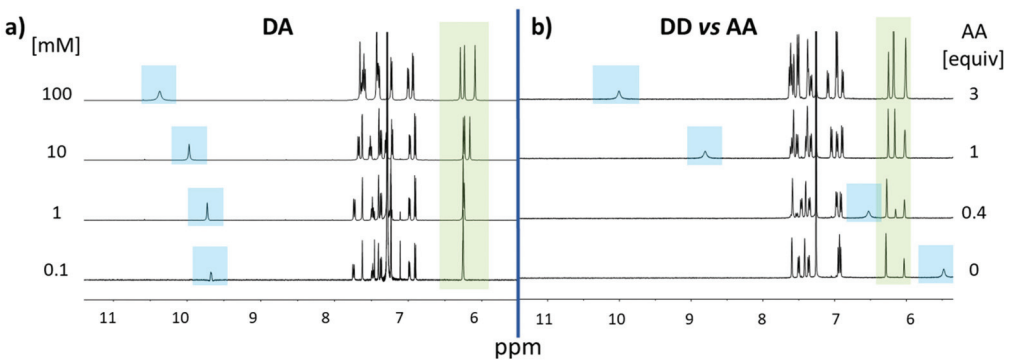


Fig. 4 Partial 500 MHz ^{1}H-NMR spectra for (a) dilution of **DA** (0.1–100 mM) and (b) titration of **AA** (0–3 equiv.) into **DD** (2.3 mM) in *d*-chloroform at 298 K. The signal due to the OH group on the phenol recognition unit D is highlighted in blue, and the signals due to the dianiline linker are highlighted in green.

Table 1 Association constants (K), effective molarities (EM), limiting NMR chemical shifts (δ_{free} and δ_{bound}) and limiting complexation-induced changes in chemical shift ($\Delta\delta$) measured by NMR titrations and dilutions in toluene-d8 and $CDCl_3$ at 298 K[a]

							^{19}F-NMR[b] (ppm)			^{1}H-NMR[b] (ppm)		
Solvent	Complex	$\log(K_{obs}/M^{-1})$	EM_d (mM)	K_{fold}	EM_f (mM)	χ_{fold}	δ_{free}	δ_{bound}	$\Delta\delta$	δ_{free}	δ_{bound}	$\Delta\delta$
$CDCl_3$	**D·A**	2.3 ± 0.1					−61.1	−62.8	−1.7	5.7	11.3	5.6
	DD·AA	3.5 ± 0.1	35 ± 9				−61.0	−62.3	−1.2	5.5	10.3	4.8
	DA·DA	1.5 ± 0.1		3.8	19 ± 2	0.79	−62.0	−62.0	−0.1	9.7	10.8	1.1
	D·A′	2.5 ± 0.1					−60.8	−62.4	−1.4	5.5	11.4	5.3
	DD·A′A′	3.5 ± 0.1	19 ± 5				−60.9	−62.2	−1.3	5.9	10.7	4.8
	DA′·DA′	1.5 ± 0.1		4.2	14 ± 1	0.81	−62.4	−62.4	0.0	10.0	10.6	0.6
	D·A	2.3 ± 0.1					−61.1	−62.8	−1.7	5.7	11.3	5.6
	D–D·A–A	2.7 ± 0.1	n.d.				−61.0	−62.5	−1.5	5.4	10.9	5.5
	D–A·D–A	2.3 ± 0.1		n.d.	n.d.	n.d.	−61.0	−62.2	−1.6	5.5	9.7	4.2
Toluene	**D·A**	3.5 ± 0.1					−61.8	−62.3	−0.5	4.9	11.6	6.7
	DD·AA[c]	5.7 ± 0.1	31 ± 4				−61.5	−61.8	−0.3			
	DA·DA	2.5 ± 0.1		19.4	7 ± 1	0.95	−61.8	−61.8	0.0	11.2	12.0	0.9
	D·A′	3.7 ± 0.1					−61.3	−61.8	−0.4	4.7	12.0	7.4
	DD·A′A′[c]	6.0 ± 0.1	17 ± 4				−61.2	−61.6	−0.4			
	DA′·DA′	2.8 ± 0.1		18.9	3 ± 1	0.95	−61.6	−61.5	0.1	11.4	12.3	0.9
	D·A	3.5 ± 0.1					−61.8	−62.3	−0.5	4.9	11.6	6.7
	D–D·A–A[c]	5.5 ± 0.1	21 ± 5				−61.5	−61.9	−0.4			
	D–A·D–A[c]	5.0 ± 0.1		n.d.	n.d.	n.d.	−61.1	−61.6	−0.5			

[a] Each titration was repeated twice and the average value is reported with errors at the 95% confidence limit. [b] Data for the signals due to the OH and CF_3 groups on the phenol recognition units. [c] Signals due to the OH protons were too broad to be detected. n.d. = not detected.

Eqn (2) allows expression of K_{obs} as a function of the equilibrium constants for folding (K_{fold}) and duplex formation (K_{duplex}):

$$K_{obs} = \frac{[DA \cdot DA]}{[DA]^2} = \frac{[DA \cdot DA]}{[DA]_{open}{}^2(1 + K_{fold})^2} = \frac{K_{duplex}}{(1 + K_{fold})^2} \quad (3)$$

Assuming that the **DD·AA** and **DA·DA** duplexes have similar effective molarities (EM_d), we can use the value of K_{duplex} measured for the homo-dimers ($K_{AA\cdot DD}$) in eqn (3) to determine the equilibrium constant for folding (K_{fold}) in the heterodimer.

$$K_{fold} = \sqrt{\frac{K_{DD\cdot AA}}{4K_{obs}}} - 1 \quad (4)$$

where the factor of 4 describes the difference in the degeneracies of the homo- and hetero-duplexes.

The folding constant (K_{fold}) can be used to determine the effective molarity for folding EM_f (eqn (5)) and the population of the folded state (χ_{fold}) (eqn (6)):

$$EM_f = \frac{K_{fold}}{K_{D\cdot A}} \quad (5)$$

$$\chi_{fold} = 1 - \frac{1}{1 + K_{fold}} \quad (6)$$

For **DA** in chloroform, the intramolecular process dominates in the monomeric state, and the folded state is highly populated ($\chi_{fold} \approx 80\%$), which means that the **DA·DA** duplex is two orders of magnitude less stable than the **AA·DD** duplex. The population of the folded state can also be estimated using ^{1}H- and ^{19}F-NMR chemical shifts.[7e] In the fast exchange regime, the chemical shift of the monomeric state (δ_{free}) is the population-weighted average of the chemical shifts of the folded and open conformations (eqn (7)). Assuming that the chemical shift of $\mathbf{DA}_{fold}$, which is H-bonded, is similar to the chemical shift of the **A·D** bound state ($\delta_{A\cdot D}$) and that the chemical shift of $\mathbf{DA}_{open}$ is similar to chemical shift of the free state of **D** (δ_D), we can use the limiting chemical shift of the monomeric state calculated from the **DA** dilution data (δ_{free}) to estimate the population of the folded conformation.

$$\delta_{free} = \chi_{fold}\delta_{D\cdot A} + (1 - \chi_{fold})\delta_D \quad (7)$$

The limiting complexation-induced change in chemical shift ($\Delta\delta_{DA\cdot DA} = \delta_{bound} - \delta_{free}$) observed in the dilution experiment can be expressed as:

$$\Delta\delta_{DA\cdot DA} = \delta_{D\cdot A} - \chi_{fold}\delta_{D\cdot A} - (1 - \chi_{fold})\delta_D \quad (8)$$

Rearranging eqn (8) allows expression of the folded population in the monomeric state (χ_{fold}) as a function of the limiting complexation-induced changes in chemical shift reported in Table 1.

$$\chi_{fold} = \frac{\Delta\delta_{D\cdot A} - \Delta\delta_{DA\cdot DA}}{\Delta\delta_{D\cdot A}} = 1 - \frac{\Delta\delta_{DA\cdot DA}}{\Delta\delta_{D\cdot A}} \quad (9)$$

Applying eqn (9) to the data for **DA** in chloroform confirms that the monomeric state is largely folded ($\chi_{fold} \approx 87\%$). Analysis of the ^{1}H- and ^{19}F-NMR chemical shifts therefore represents a straightforward method for evaluating the folding properties of these oligomers. The data in Table 1 show that mixed sequence oligomers based on acceptor **A**, donor **D** and a 1,3-phenylenediamine linker are not good candidates for the formation of stable duplexes and encoding of information.

^{1}H- and ^{19}F-NMR titration experiments on the formation of the **D·A′** and **DD·A′A′** complexes in chloroform indicate that steric bulk around the H-bond acceptor **A′** does not hamper the cooperative duplex formation in chloroform. There is a small drop in the effective molarity for duplex formation EM_d, and the chelate cooperativity associated with duplex formation, which is quantified by the product $K_{D\cdot A}$ EM, is reduced from 7 to 5. However, the increased steric bulk did not significantly affect the folding properties of these oligomers, and the folded conformation dominates in the monomeric state of **DA′**: $\chi_{fold} \approx$ 81% using eqn (6), and analysis of the ^{1}H and ^{19}F-NMR chemical shifts of **DA′** show that δ_{free} is very similar to the δ_{bound} of the **D·A′** and **DD·A′A′** complexes, giving a value of $\chi_{fold} \approx$ 90%.

The oligomers with the longer dianiline linker gave quite different behaviour in chloroform, and there was no clear evidence of duplex formation for either the homo- or hetero-dimers. The association constants for formation of **D–D·A–A** and **D–A·D–A** are very similar to the value measured formation of a single intermolecular H-bond. These results are difficult to rationalise, especially as stable duplex formation was observed for this system in toluene (see below).

The results in Table 1 show that switching from chloroform to toluene leads to one order of magnitude increase of the association constant for the intermolecular base-pairing interaction ($K_{D\cdot A}$ and $K_{D\cdot A'}$), without modifying the folding properties of the system. The addition of a second recognition unit in **DD·AA** causes an increase of about two orders of magnitude in the association constant in toluene. The complementary homo-dimers form a very stable duplex with an EM_d of 31 mM, which implies that the doubly H-bonded form is almost exclusively populated (98%), and the chelate cooperativity associated with duplex formation expressed as the product $K_{D\cdot A}$ EM is 80.[8] However in the case of the hetero-dimer **DA**, the folding channel dominates the behaviour in toluene, and the folded population $\chi_{fold} \approx$ 95% according to both eqn (6) and (9).

The system containing **A′** gave similar results in toluene. Table 1 shows that comparing **D·A′** to **DD·A′A′** the association constant increases by two order of magnitude. The effective molarity for duplex formation (EM_d = 17 mM) is similar to the value measured for **DD·AA**, and the chelate cooperativity ($K_{A'\cdot D}$ EM = 54) is so high that the doubly H-bonded form is almost exclusively populated (98%). Dilution studies performed on the **DA′** dimer show that the folded conformation dominates in the monomeric state ($\chi_{fold} \approx$ 95% according to eqn (6) and (9)).

The system equipped with the longer dianiline linker displayed quite different behaviour in toluene from that observed in chloroform. Comparing **D·A** to **D–D·A–A**, the association constant increased by two orders of magnitude. The effective molarity associated with duplex formation is about 21 mM, which corresponds to a chelate cooperativity factor of $K_{A\cdot D}$ EM = 66. Thus in toluene, the closed form of the duplex with two intermolecular H-bonds is almost fully populated (98%). In the case of the hetero-dimer **D–A**, the folding channel does not compete with duplex formation. The observed association constant is similar to the value measured for the **D–D·A–A** duplex and is two orders of magnitude higher than the value measured for a single H-bond in **D·A**. In addition, the limiting ^{19}F-NMR chemical shift of monomeric **D–A** shows that δ_{free} is equal to δ_{free} of both **D** and **D–D**, confirming that there is no intramolecular H-bonding ($\chi_{fold} \approx$ 0% according to eqn (6) and (9)).

Diffusion ordered spectroscopy

Next, we explored how the folding properties influence diffusion of the hetero-dimers. DOSY experiments were used to determine the diffusion coefficients of the folding hetero-dimer **DA** and the extended hetero-dimer **D–A** at 0.4 and 33 mM in toluene solution (Fig. 5). According to the Stoke–Einstein equation, the diffusion coefficient of a species is inversely proportional to the diameter. Hence, the self-assembly of dimers into duplexes or high order networks will cause a decrease in the rate of diffusion. At 33 mM about 72% of **DA** and 99% of **D–A** are assembled through intermolecular H-bonds. Since this concentration is comparable to EM_d, it is likely that both duplexes and higher order aggregates are present. At 0.4 mM, only 8% of **DA** is present as the duplex, and the rest is in the folded monomeric state, which leads to a significant increase in the diffusion coefficient (from 4×10^{-10} to 1.8×10^{-9} m^2 s^{-1}). However at 0.4 mM, 88% of **D–A** populates the duplex, because there is no competing folding

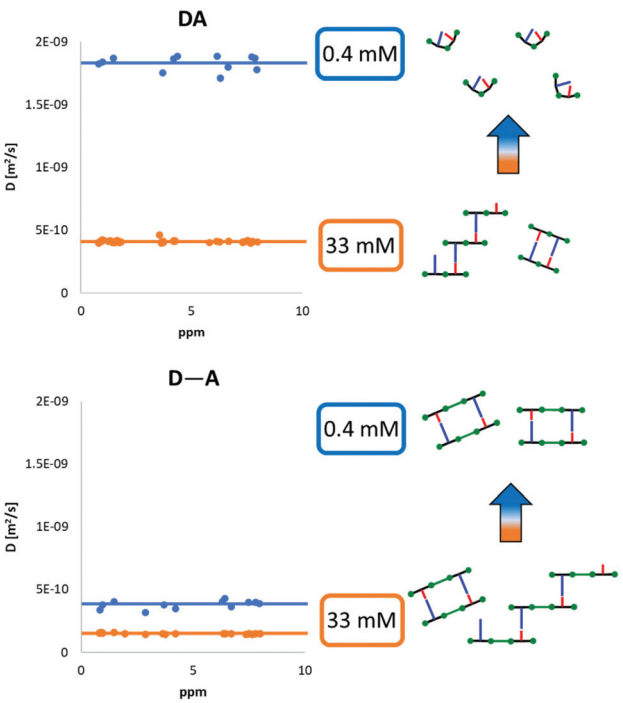


Fig. 5 Diffusion coefficients measured in DOSY experiments on toluene-*d*8 solutions of DA (top) and D–A (bottom) at 298 K. The blue (0.4 mM) and orange (33 mM) bars represent the average diffusion coefficients. A schematic representation of the different assemblies present is shown: duplex and higher order complexes at high concentration; duplex or folded monomer at low concentration.

channel, and consequently the diffusion coefficient does not change significantly (from 2×10^{-10} to 4×10^{-10} m^2 s^{-1}).

Conclusions

Complementary homo-oligomers based on phenol and phosphine oxide recognition units and a 1,3-dianiline linker have been shown to form extended duplexes. We investigated here whether these properties extend to hetero-oligomers. Heterodimer equipped with one H-bond donor and one H-bond acceptor recognition unit (**DA**) was synthesised, and dilution experiments showed that the observed association constant for duplex formation is orders of magnitude lower (two in chloroform and three in toluene) than the association constant for the corresponding duplex formed between two homo-dimers **DD·AA**. Analysis of the limiting complexation-induced changes in ^{19}F- and ^{1}H-NMR chemical shift ($\Delta\delta_{AD\cdot AD} \approx 0.1$ ppm) confirm that intramolecular folding, which competes with duplex formation, is highly populated in the monomeric state of **DA**.

This result suggests that folding will compromise duplex formation in mixed sequence oligomers of this architecture and prompted us investigate whether 1,2-folding could be avoided by increasing the steric bulk around the H-bond acceptor unit or by increasing the aniline linker length. Two new series of homo- and hetero-dimers were synthesised, and the competing equilibria between duplex formation and folding were analysed by NMR titration and dilution experiments.

Cyclohexyl moieties on the phosphine oxide H-bond acceptor (**A′**) slightly reduced the effective molarity for duplex formation but did not change the folding properties. Complementary homo-dimers were still able to assemble into duplex structures, with an increase of 1–2 orders of magnitude in association constant compared with the singly H-bonded complex **D·A′**. However, the observed association constant for duplex formation in the hetero-dimer **DA′** is substantially lower. In the monomeric state, the folded conformation with an intramolecular H-bond between the recognition units dominates in both chloroform and toluene.

The greatest changes in the self-assembly properties of the oligomers were achieved by replacing the 1,3-phenylendiamine linker with a longer phenylacetylene di-aniline. Surprisingly, there was no evidence of duplex formation for either the homo- or hetero-dimers in chloroform. However in toluene for both the homo- and hetero-dimers, the observed association constant for duplex formation is two orders of magnitude higher than for formation of the singly H-bonded complex **D·A**. The effective molarity associated with the intramolecular H-bonds that lead to duplex formation is 21 mM, which implies that the chelate cooperativity factor of 66 and the doubly H-bonded duplex is fully populated (98%). The limiting complexation-induced change in chemical shift found in ^{19}F-NMR dilution experiments on **D–A** ($\Delta\delta_{D-A\cdot D-A} \approx -0.5$ ppm) is equal to the value observed for formation of the **D·A** and **D–D·A–A** complexes. These data indicate that the folded conformation is not populated for **D–A** in the monomeric single stranded state. Thus the extended linker effectively prevents 1,2-folding without affecting duplex formation in toluene, suggesting that this system is a good candidate for the development of longer duplex forming hetero-oligomers.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

L. G. thanks UNIPD STARS-StG (DyNAseq) for funding.

Notes and references

1 (*a*) J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 964; (*b*) W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984. ISBN: 978-1-4612-5190-3.

2 (*a*) A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818; (*b*) R. Stoltenburg, C. Reinemann and B. Strehlitz, *Biomol. Eng.*, 2007, **24**, 381; (*c*) W. Tan, K. Wang and T. J. Drake, *Curr. Opin. Chem. Biol.*, 2004, **8**, 547.

3 (*a*) Z. Chen, P. A. Lichtor, A. P. Berliner, J. C. Chen and D. R. Liu, *Nat. Chem.*, 2018, **10**, 420; (*b*) L. D. Usanov, A. I. Chan, J. P. Maianti and D. R. Liu, *Nat. Chem.*, 2018, **10**, 704.

4 (*a*) R. Kramer, J.-M. Lehn and A. Marquis-Rigault, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5394; (*b*) A. Marquis, V. Smith, J. Harrowfield, J.-M. Lehn, H. Herschbach, R. Sanvito, E. Leize-Wagner and A. van Dorsselaer, *Chem. – Eur. J.*, 2006, **12**, 5632; (*c*) H. L. Anderson, *Inorg. Chem.*, 1994, **33**, 972; (*d*) P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.*, 1999, **121**, 11538; (*e*) V. Berl, I. Huc, R. G. Khoury, M. J. Krische and J.-M. Lehn, *Nature*, 2000, **407**, 720; (*f*) V. Berl, I. Huc, R. G. Khoury, M. J. Krische and J.-M. Lehn, *Chem. – Eur. J.*, 2001, **7**, 2810.

5 (*a*) Y. Tanaka, H. Katagiri, Y. Furusho and E. Yashima, *Angew. Chem.*, 2005, **117**, 3935; (*b*) H. Ito, Y. Furusho, T. Hasegawa and E. Yashima, *J. Am. Chem. Soc.*, 2008, **130**, 14008; (*c*) H. Yamada, Y. Furusho, H. Ito and E. Yashima, *Chem. Commun.*, 2010, **46**, 3487; (*d*) E. Yashima, N. Ousaka, D. Taura, K. Shimomura, T. Ikai and K. Maeda, *Chem. Rev.*, 2016, **116**, 13752.

6 (*a*) J. Sanchez-Quesada, C. Seel, P. Prados, J. de Mendoza, Í. Dalcol and E. Giralt, *J. Am. Chem. Soc.*, 1996, **118**, 277; (*b*) A. P. Bisson, F. Carver, D. S. Eggleston, R. C. Haltiwanger, C. A. Hunter, D. L. Livingstone, J. F. McCabe, C. Rotger and A. E. Rowan, *J. Am. Chem. Soc.*, 2000, **122**, 8856; (*c*) A. P. Bisson and C. A. Hunter, *Chem. Commun.*, 1996, 1723–1724; (*d*) B. Gong, Y. Yan, H. Zeng, E. SkrzypczakJankunn, Y. W. Kim, J. Zhu and H. Ickes, *J. Am. Chem. Soc.*, 1999, **121**, 5607; (*e*) B. Gong, *Synlett*, 2001, 582; (*f*) H. Zeng, R. S. Miller, R. A. Flowers and

B. Gong, *J. Am. Chem. Soc.*, 2000, **122**, 2635; (*g*) B. Gong, *Acc. Chem. Res.*, 2012, **45**, 2077; (*h*) Y. Yang, Z.-Y. Yang, Y.-P. Yi, J.-F. Xiang, C.-F. Chen, L.-J. Wan and Z.-G. Shuai, *J. Org. Chem.*, 2007, **72**, 4936; (*i*) W.-J. Chu, Y. Yang and C.-F. Chen, *Org. Lett.*, 2010, **12**, 3156; (*j*) W.-J. Chu, J. Chen, C.-F. Chen, Y. Yang and Z. Shuai, *J. Org. Chem.*, 2012, **77**, 7815; (*k*) A. E. Archer and M. J. Krische, *J. Am. Chem. Soc.*, 2002, **124**, 5074; (*l*) H. Gong and M. J. Krische, *J. Am. Chem. Soc.*, 2005, **127**, 1719.

7 (*a*) G. Iadevaia, A. E. Stross, H. Neumann and C. A. Hunter, *Chem. Sci.*, 2016, **7**, 1760; (*b*) A. E. Stross, G. Iadevaia and C. A. Hunter, *Chem. Sci.*, 2016, **7**, 5686; (*c*) A. E. Stross, G. Iadevaia and C. A. Hunter, *Chem. Sci.*, 2016, **7**, 94; (*d*) D. Núñez-Villanueva and C. A. Hunter, *Chem. Sci.*, 2017, **8**, 206; (*e*) D. Núñez-Villanueva, G. Iadevaia, A. E. Stross, M. A. Jinks, J. A. Swain and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 6654; (*f*) A. E. Stross, G. Iadevaia, D. Núñez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 12655; (*g*) J. A. Swain, G. Iadevaia and C. A. Hunter, *J. Am. Chem. Soc.*, 2018, **140**, 11526; (*h*) F. T. Szczypiński and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 2444; (*i*) F. T. Szczypiński, L. Gabrielli and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 5397.

8 L. Gabrielli, D. Núñez-Villanueva and C. A. Hunter, *Chem. Sci.*, 2020, **11**, 561.

9 L. Gabrielli and C. A. Hunter, *Chem. Sci.*, 2020, **11**, 7408–7414.

10 The presence of multiple diastereoisomers does not affect the behaviour of the oligomers, because these solubilising groups are remote from the polar sites.

