COMMUNICATION

ChemComm
www.rsc.org/chemcomm

New building blocks for the assembly of sequence selective molecular zippers

Christopher A. Hunter,*[a] Philip S. Jones,[b] Pascale M. N. Tiger[a] and Salvador Tomas[a]

[a] *Centre for Chemical Biology, Krebs Institute for Biomolecular Science, Department of Chemistry, University of Sheffield, Sheffield, UK S3 7HF*

[b] *Roche Discovery Welwyn, 40 Broadwater Road, Welwyn Garden City, Herts, UK AL7 3AY*

Received (in Cambridge, UK) 11th April 2003, Accepted 13th May 2003
First published as an Advance Article on the web 16th June 2003

Synthetic H-bonded molecular zippers contain no sequence information that can be used to engineer the selective binding interactions characteristic of biopolymers; reversing the sense of the amide bonds in the two binding partners generates a new orthogonal recognition motif and the mutually complementary binding partners form complexes an order of magnitude more stable than the corresponding mismatch complexes.

Oligomeric materials composed of a linear array of recognition sites possess some unique properties, as typified by the nucleic acids. DNA stores genetic information as a sequence of covalently-linked nucleotides, and its ability to reproduce this information is based on the self-assembly of two complementary linear strands into a double-stranded complex. H-Bonding interactions between the bases allow one strand to 'read' the sequence of another strand, and this property is responsible for the high specificity observed in the self-assembly and self-replication of double-stranded nucleic acids.[1] This property has been exploited in molecular and macromolecular construction for the formation of complex topological objects and the encoded organisation of nanoparticles.[2] The assembly of linear proteins into multi-stranded complexes through encoded recognition sites is the basis of the behaviour of biological fibres such as muscle.[3] Although there are examples of synthetic systems which self-assemble into double-stranded complexes,[4–8] at present, there are no examples in which sequence recognition information is encoded, and so these systems do not exhibit any of the more interesting and potentially exploitable properties of their biological counterparts.

We have developed a synthetic system where hydrogen-bonding and aromatic interactions direct the assembly of double-stranded complexes of amide oligomers, 'molecular zippers' (Fig. 1).[9,10] The association constant increases by an order of magnitude for each increment in the length of the oligomer, indicating that the interactions along the zipper complexes are highly cooperative.[11] However, these systems rely entirely on length for their recognition properties and the selectivity between complementary and non-complementary arrangements is poor. Here, we describe a new strategy for introducing sequence information into the zipper structures.

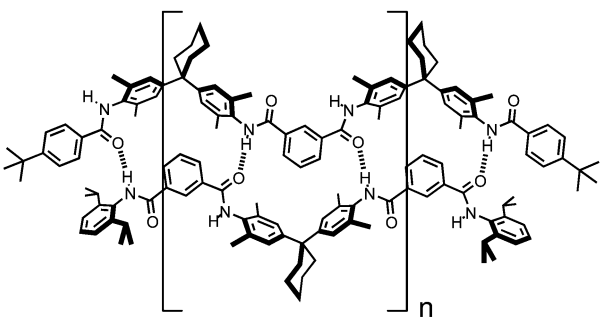


Fig. 1 Molecular zipper complexes formed between bisaniline–isophthalic acid oligomers in chloroform.

The bisaniline–isophthalic acid recognition motif used in the zippers shows excellent complementarity which is not dissipated in long chain lengths. Thus the best chance for success in designing the new recognition elements required to encode sequence information is to make a minimal structural change, *i.e.* reverse the sense of the amide bonds. The approach is illustrated in Fig. 2. The geometry of interaction is identical to the zipper complexes reported previously (Fig. 1), and the expectation is that the stability will be similar. However, mixing the recognition motifs in Figs. 1 and 2 should lead to a geometric mismatch in which the hydrogen bonding groups will not be optimally positioned or aligned. This should generate selective mutually exclusive binding interactions between complementary partners. To test the validity of this approach, we have studied the properties of simple model compounds containing the two different recognition motifs and investigated all possible pairwise binding interactions (Fig. 3).

Compounds **1** and **2** have been described elsewhere.[12] Compound **3** was prepared from bisaniline **5** (Scheme 1). Reaction with $NaNO_2$ and HCl gave the diazonium salt which was subsequently treated with KI to give the diiodide, **6**. Treatment of **6** with ^{n}BuLi followed by solid CO_2 gave the dicarboxylic acid **7** in 44% overall yield from **5**. The diacid was converted to the corresponding diacid chloride with oxalyl chloride and coupled with 4-*tert*-butyl aniline to give the target compound **3**.† Compound **4** was prepared in a similar fashion by converting 2,6-diisopropyl aniline to the corresponding benzoic acid **9** and then coupling with 1,3-diaminobenzene (Scheme 1).†

The selectivity of the binding interactions between the four compounds was assessed by ^{1}H NMR titrations, and the results are summarised in Table 1. The two complementary pairs of binding partners, **1**·**2** and **3**·**4**, form complexes that are nearly an order of magnitude more stable than the mismatched pairs, **1**·**4** and **3**·**2**. The dimerisation constants and association constants for the complexes between the isophthalic acid and bisaniline pairs, **1**·**3** and **2**·**4**, are all too low to be measured ($< 1\ M^{-1}$). In these systems, the only possible mode of interaction is *via* a single hydrogen bond. Thus the larger association constants observed for the mismatched complexes, **1**·**4** and **3**·**2**, indicate that there is some specific interaction taking place in these systems, *i.e.* there are two intermolecular hydrogen bonds in these complexes.

The complexation-induced changes in chemical shift indicate that all four complexes, **1**·**2**, **3**·**4**, **1**·**4** and **3**·**2**, have similar

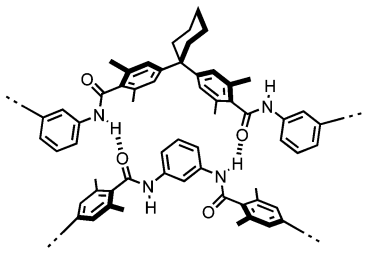


Fig. 2 Reversing the sense of the amide bonds generates a new complementary recognition motif.

DOI: 10.1039/b304041k

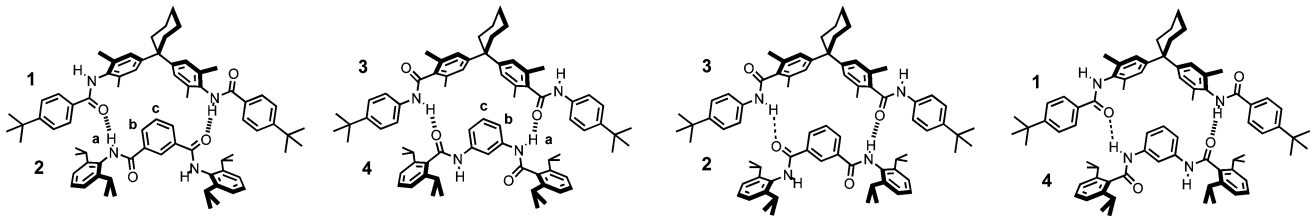
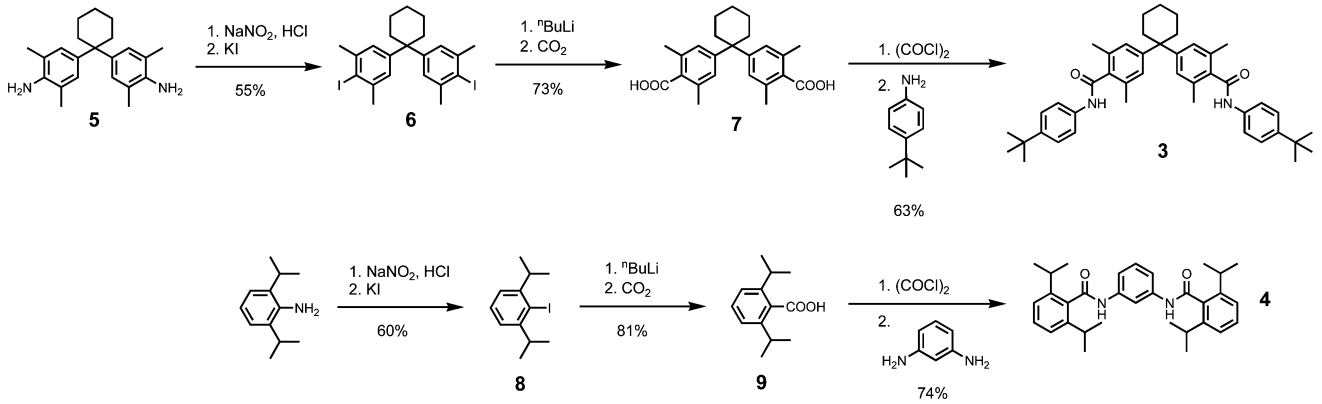


Fig. 3 The **1**·**2** complex represents the parent recognition motif that forms the basis for the zipper complexes shown in Fig. 1. In the **3**·**4** complex, the orientation of all of the amide bonds has been reversed to generate the new recognition motif shown in Fig. 2. The **3**·**2** and **1**·**4** complexes represent the mismatched pairs: when one hydrogen bond is optimally positioned as shown, the amide groups on the other side of the complex are too far apart and incorrectly oriented for optimal hydrogen bonding. The ^{1}H NMR proton labelling used in Table 2 is shown for compounds **2** and **4**.



Scheme 1

structures (Table 2). The downfield shifts of the amide protons, **a**, show that the amides are involved in H-bonding interactions. The upfield shifts of protons **b** and **c** show that the *meta* substituted aromatic rings are located in similar orientations in the diarylcyclohexane binding pockets. Thus the subtle differences in geometry associated with reversing the orientation of the amide bonds are not enough to completely disrupt complex formation. However, they are sufficient to significantly alter the relative stabilites of the complexes. In these model complexes, the ends of the molecules are free to distort to maximise the intermolecular interactions, but in longer oligomers, the geometrical constraints imposed by neighbouring interactions in the chain are likely to amplify the differences in stability between the matched and mismatched pairs.

This system therefore looks like a promising candidate for the construction of mutually complementary oligomers that carry sequence specific binding information. We are currently developing the synthesis of such oligomers using unsymmetrical amino acid building blocks to connect the two recognition motifs described here. These compounds will open the door to the development of new synthetic molecular information systems.

This work was supported by the Roche (P. M. N. T.) and the European Union (S. T.).

Table 1 Association constants (K_a/M^{-1}) in $CDCl_3$ at 295 K[a]

Compound	**1**	**2**	**3**	**4**
1	[b]	35 ± 2	[b]	5 ± 1
2	35 ± 2	[b]	6 ± 1	[b]
3	[b]	6 ± 1	[b]	31 ± 3
4	5 ± 1	[b]	31 ± 3	[b]

[a] Average values for at least two separate experiments. Titration data for 4–6 signals were used to determine the association constants in each experiment. Errors are quoted as twice the standard error from weighted mean (weighting based on the observed change in chemical shifts). [b] No binding interactions were detected which means that these complexes have association constants less than 1 M^{-1}.

Table 2 Complexation-induced changes in ^{1}H NMR chemical shift ($\Delta\delta$, ppm) in $CDCl_3$ at 295 K[a]

	Proton		
Complex	**a**	**b**	**c**
1·**2**	1.2	−1.6	−0.4
1·**4**	1.2	−1.3	−0.5
3·**2**	1.2	−1.3	−0.5
3·**4**	1.2	−1.5	−0.4

[a] Errors are of the order of 20%. See Fig. 2 for the proton labelling scheme.

Notes and references

† All new compounds gave satisfactory spectroscopic data.

1 J. H. G. Vanroode and L. E. Orgel, *J. Mol. Biol.*, 1980, **144**, 579.
2 C. Mao, W. Sun, Z. Shen and N. C. Seeman, *Nature*, 1999, **397**, 144.
3 H. E. Huxley, *Science*, 1969, **164**, 1356.
4 U. Koert, M. M. Harding and J.-M. Lehn, *Nature*, 1990, **346**, 339.
5 R. Krämer, J.-M. Lehn, A. D. Cain and J. Fischer, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 703.
6 J. Sanchez Quesada, C. Seel, P. Prados, J. deMendoza, I. Dalcol and E. Giralt, *J. Am. Chem. Soc.*, 1996, **118**, 277.
7 H. L. Anderson, *Inorg. Chem.*, 1994, **33**, 972.
8 P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.*, 1999, **121**, 11538.
9 A. P. Bisson, F. J. Carver, C. A. Hunter and J. P. Waltho, *J. Am. Chem. Soc.*, 1994, **116**, 10292.
10 A. P. Bisson, F. J. 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.
11 A. P. Bisson and C. A. Hunter, *Chem. Commun.*, 1996, 1723.
12 A. P. Bisson, C. A. Hunter, J. C. Morales and K. Young, *Chem. Eur. J.*, 1998, **4**, 845.