ChemComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: H. Mattelaer, C. Mattelaer, N. Papastavrou, W. Dehaen and P. Herdewijn, *Chem. Commun.*, 2017, DOI: 10.1039/C7CC00831G.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

View Article Online
DOI: 10.1039/C7CC00831G

Chemical Communications

COMMUNICATION

Oligonucleotide promoted peptide bond formation using a tRNA mimicking approach†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

H.-P. Mattelaer,[a,b] C.-A. Mattelaer,[a] Nikolaos Papastavrou,[a] W. Dehaen[b] and P. Herdewijn[a]*

TransferRNA's role in protein translation is the prime example of an Informational Leaving Group (ILG). A simplified model produced oligophenylalanine with a modified uracil as ILG in presence of specific oligonucleotides. Our preliminary studies contribute to the importance of hybrid species in bridging the gap between peptides and nucleic acids.

Translating genetic information into proteins plays a central role in all vital functions and is operated by a complex, cellular machinery. Simplified, this process represents an oligonucleotide templated and catalysed production of peptides. Templated synthesis has seen great advances over the past years. Starting from non-enzymatic oligonucleotide synthesis by prebiotic chemists[1]; oligonucleotides have been used to catalyse many different type of chemistries, including amide formation.[2] However early attempts to catalyse peptide formation only achieved dimerization and subsequent cyclization of amino acids when using 2'(3')-esters of mononucleotides.[3] More successful were enzyme mimicking approaches for (amino) acyl transfer reactions by programmed proximity[4], with the artificial ribosome as prime example.[5] However those designed molecular systems do not start from a separate template. The search for a non-enzymatic informational link between oligonucleotides and peptides remains a challenge. A link that could be bridged by hybrids, such as PNA.[6] This oligonucleotide analogue combines a stable peptide backbone with nucleobase pairing and has proven to be a valuable example of templated amide bond formation.[7] Another interesting take came from Robertson and Miller[8] who drew their inspiration from tRNA which undergoes extensive post-transcriptional modifications[9] including functional groups of present-day amino acids. Such nucleobase-peptide hybrids could potentially increase the catalytic functionality of nucleic acids.[8] One of those uracil derivatives (5-(4-hydroxybenzyl)uracil) caught our eye, as the phenolic moiety could activate carboxylic acids and thus initiate peptide synthesis.[10] It can be considered as a uracil bearing a tyrosine sidechain and as the chemical simplification of tRNA, which acts as a leaving group (LG) and information carrier during protein synthesis (Figure 1A, 1B). The compound was loaded with an amino acid, N-protected with a photocleavable group (NVOC). This protection strategy is known in amino acyl-tRNA synthesis[11] as irradiation avoids the use of harsh reagents incompatible with the oligonucleotide in question. Here it allows us to start the oligomerization with minimal interference. Furthermore, phenylalanine was chosen for ease of follow-up during synthesis and analysis (Figure 1C). Further synthetic details can be found in the Supplementary Information (SI).

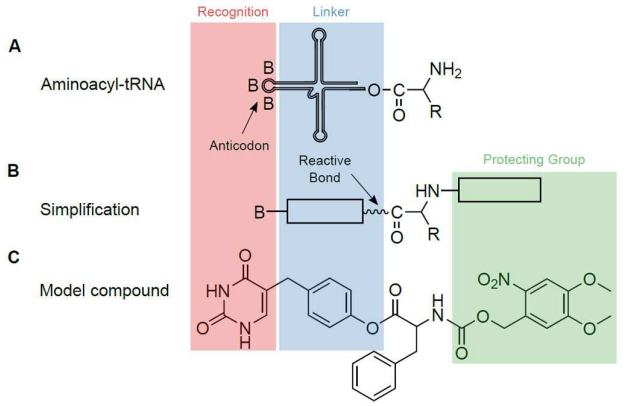
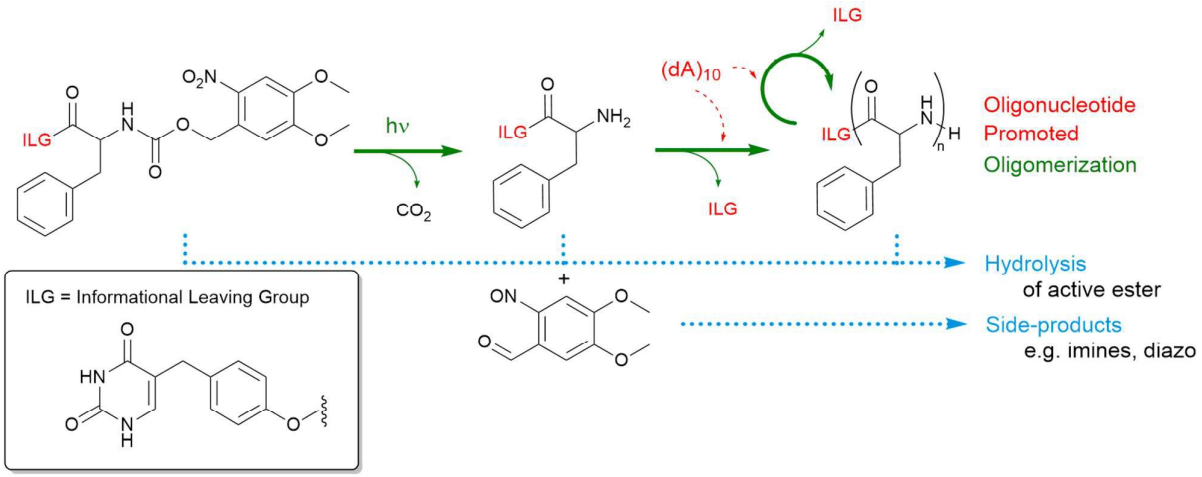


Figure 1. The Informational Leaving Group approach: A) aminoacyl-tRNA, depicted in the typical cloverleaf representation with anticodon and 3' amino acyl detailed; B) the simplification of tRNA into Recognition, Linker and Amino Acid with Reactive Bond, introducing a Protecting Group as synthetic necessity; C) the model compound envisaged based upon the prebiotic available 5-(4-hydroxybenzyl)uracil, using phenylalanine as amino acid and N-(4,5-dimethoxy-2-nitrobenzyloxycarbonyl) as photocleavable protecting group.

[a] *KU Leuven, Rega Institute for Medical Science, Medicinal Chemistry Herestraat 49, 3000 Leuven, Belgium*
E-mail: Piet.Herdewijn@kuleuven.be.
[b] *KU Leuven, Department of Chemistry, Molecular Design and Synthesis, Celestijnenlaan 200F, 3000 Leuven, Belgium.*
*Author to whom correspondence should be addressed:
E-mail: Piet.Herdewijn@kuleuven.be.
†Electronic Supplementary Information (ESI) available: Synthetic details and additional figures S1-S10. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C7CC00831G



Scheme 1. Studied model of the ILG approach: The green pathway describes the oligomerisation of free activated amino acid upon irradiation. By using an ILG complementary to the used oligonucleotide (insert), the efficiency of this process could be increased (red). Expected side-reactions, competing with this process, are depicted in blue. These consists of hydrolysis of the phenolic ester and side-products emanating from the photolytical removal of the amine protection group.

Before studying oligomerization reactions of the model compound (Scheme 1); any influence of the LG on amide formation was studied. Kinetic measurements were designed to compare 5-(4-hydroxybenzyl)uracil and *p*-cresol as LG; to determine any influence of the uracil moiety on the amide bond formation, as pyrimidinones are known for their catalytic effect.[12] Indications for a small catalytic effect were found but the potential influence of LG is minimal in comparison to solvent effects on peptide bond formation (SI, kinetics).

Spontaneous formation of peptides, with or without uracil in the LG, was studied by deprotecting a solution of monomer in DMSO by UV irradiation, followed by the addition of an equal volume of either water, buffer and/or a solution of oligonucleotide. After a set time of incubation, the crude reaction mixture was analysed by HPLC. An example can be seen in Figure 2, with an overlay of monomer elongation in the absence or presence of $(dA)_{10}$. The complex mixture seen is most likely caused by the photolytic deprotection of the amine, which results in reactive side-products (Scheme 1).[13] However, their low concentration and poor resolution limited the definitive identification. Fortunately, this was not the case for the oligomerisation products. Oligomers of phenylalanine up to tetramers were identified, capped by the 5-(4-hydroxybenzyl)uracil LG. Furthermore, a few hydrolysis product could be identified (Figure 3C). All products were identified using simultaneous MS spectra (LC-MS, Figure 3A,3B) or with retention time derived from commercial references (Figure 3D).

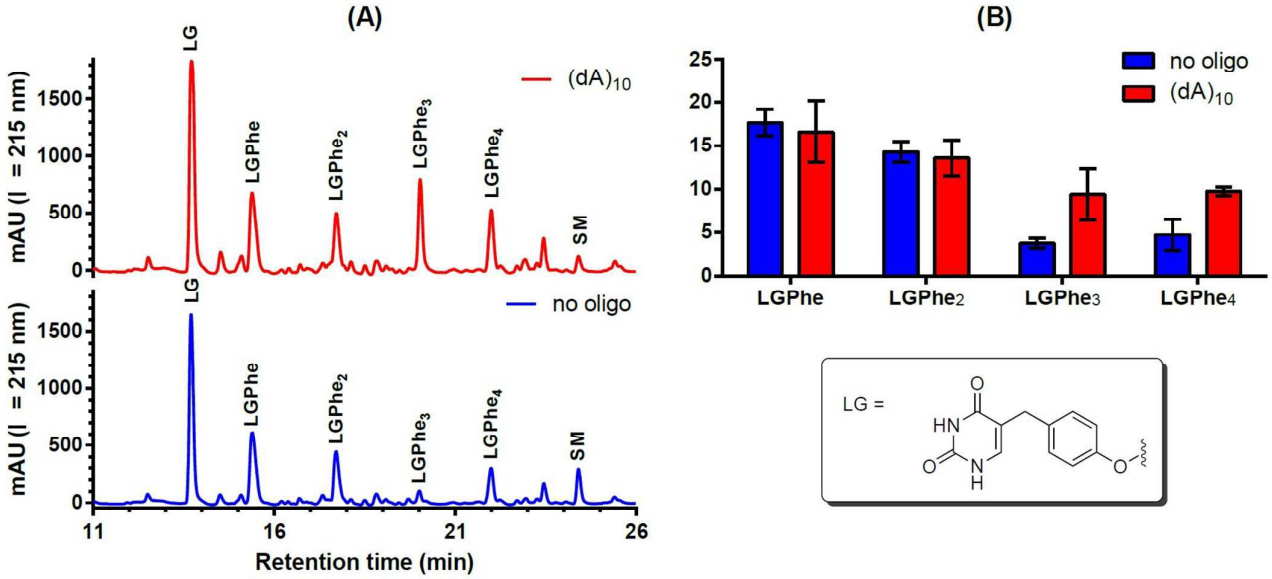


Figure 2. The photolytical cleavage of the model compound: A) representative example. Photolysis of the model compound of Figure 1C followed by incubating 1h at room temperature, with either 5 mol% of oligodeoxyadenosine (red) or water (H_2O). UV detection at 215 nm, zoom with identification of species. Full spectrum and other data in the SI. B) AUC analysis of identified species. The mean Area Under the Curve (AUC) of peptide species with (red) or without (blue) added oligonucleotide, as identified in A, are plotted as a relative percentage of the total AUC. Error values of three independent experiments; LG = 5-(4-hydroxybenzyl)uracil (insert), SM = starting material.

View Article Online
DOI: 10.1039/C7CC00831G

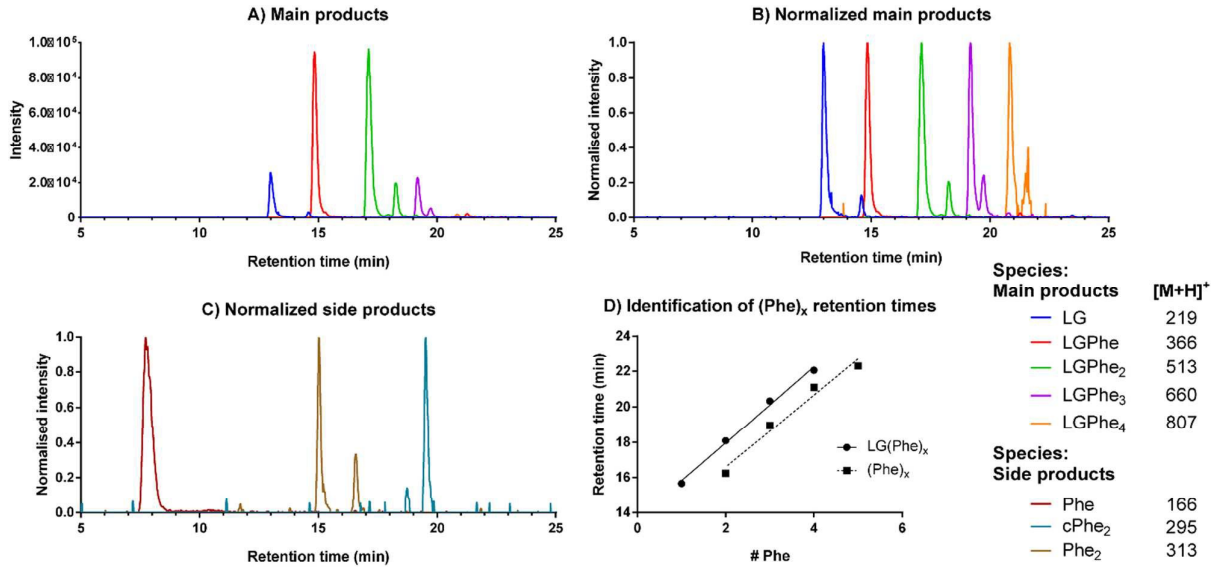


Figure 3. Identification of investigated species: A) Ion extract chromatograms (IEC) of the main products: the LG and LG-capped amino acid/peptides. B) Normalized IEC of main products illustrating increasing amounts of diasteroisomers with polymerisation and the increasing difficulty to observe higher polymerized species. C) Normalized IEC of side products detected. D) Plot of retention times of LG capped oligomers (experimental data) versus uncapped, linear oligomers of phenylalanine (commercial references) illustrating the chemical analogy and predictable retention times. LG = (5-(4-hydroxybenzyl)uracil)

The mass spectrum in Figure 3B additionally revealed secondary peaks with equal mass for oligomerized species at different retention times and with increasing relative intensity at greater chain length. This observation of presumed diastereoisomers agrees with previously obtained results on peptide formation of activated species. [14]

Another interesting find is the relatively low amounts of diketopiperazine (DKP) in all of the conducted analyses. Whether it be a with or without an oligonucleotide additive; the cyclized product is only a minor side product of this system in contrary to other known systems[3], even when using phenyl esters.[14] However, most important is the increase of trimer ($LGPhe_3$) and tetramer ($LGPhe_4$) by adding the oligonucleotide (Figure 2). This was confirmed by plotting the main products as a relative percentage towards the total absorbance from three independent experiments (with or without oligonucleotide) in Figure 2B. It should be noted that this remains a qualitative confirmation against the control experiments as the growing chain adds to the molecular extinction coefficient, leading to overestimation of larger species when comparing them individually in one experiment. Furthermore, larger oligomers (n > 4) were currently not found. Figure 3A shows the rapid decay of MS signal intensity with larger species, probably due to the combination of low concentration and difficult ionisation under the conditions used e.g. supramolecular assembly[15] and gas-phase adducts might results in assemblies too large to detect. Additionally, the extrapolation of Figure 3D for $LGPhe_5$ led to an area of several small peaks ($t_R \approx 24$min) which might also overlap with the remaining starting material.

By switching the LG to *p*-cresol, no noticeable change in reaction mixture composition was observed upon addition of $(dA)_{10}$ (Figure S8).A brief screening of different oligonucleotides can be seen in Figure 4. Oligodeoxythymidine $(dT)_{10}$ gave no significant improved oligomerisation compared to control conditions, as did the alternating sequence of deoxyadenosine and -cytosine $d(CA)_5$. However, $(dC)_{10}$ did increase oligomerisation with a similar effect on $LGPhe_3$, but with lower efficiency on producing $LGPhe_2$ and $LGPhe_4$ as compared to $(dA)_{10}$.

In light of previous results, the influence of pH catalysis was investigated. Unless specified, the reaction pH was uninfluenced, resulting in an observed value of 6.1-6.5 after chromatographic analysis. A slight acidic pH (pH = 6) favoured the product distribution towards oligomerized species (Figure S4, S5) in control conditions (without oligonucleotide). This distribution is influenced by the acidic additive if we compare both. Nevertheless, with $(dA)_{10}$, higher oligomers ($LGPhe_3$ and $LGPhe_4$) followed the same increasing trend as seen in aqueous solutions. Buffering at pH 7.5 follows this trend, although the effects are more pronounced (Figure S6).

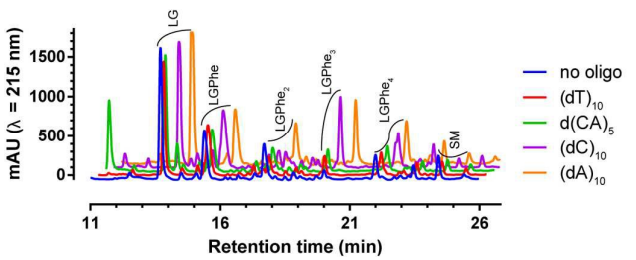


Figure 4. Testing different oligonucleotides: Overlay of standard photolysis experiment while using various oligonucleotides. UV detection at 215 nm, zoom with identification of studied species.

View Article Online
DOI: 10.1039/C7CC00831G

An alkaline pH of 8.4 results in a distinctively different pattern with many new and unresolved peaks (Figure S7). The oligomers could be determined, but no effect of the oligonucleotide was witnessed. Furthermore, incubating the samples for a prolonged period of time did not resulted in higher yields of oligomers (Figure S8, S9). On the contrary, the samples degraded and the mixtures became more complex presumably caused by side reactions with the (previously mentioned) reactive intermediates of the photolytical deprotection.

Our preliminary studies on peptide oligomerisation reactions shows promise as implementation of an Informational Leaving Group (ILG)[16], inspired by the chemical simplification of tRNA. However some issues still need to be addressed. First of all, the *in situ* activation of the reaction by irradiation causes more analytical problems than the issues we avoid in producing the starting material. Secondly, the lipophilicity of the starting material forced the use of DMSO/aqueous solutions. The resulting oligomers are equally poorly soluble and prone to supramolecular effects[15], hampering analysis. A different method for reaction initiation (e.g. pH shift of the free ammonium salt in solution) and a different amino acid could help eliminating organic solvents, increase signal-to-noise and, in result, should lead to witnessing longer oligomers. Thirdly, the catalytic effect of the oligonucleotide is not solely mediated by base pairing interactions. Although lacking an uracil moiety in the LG does not produce any catalytic effect, it is not only $(dA)_{10}$ that influences the reaction. The presence of an exocyclic amine seems to favour the peptide bond formation[17], which could be mediated by other interactions with the ILG such as dipole-dipole interaction or salt-like bridges. A similar oligomerisation effect is seen at acidic pH. However, the oligonucleotide supplements the latter and acts as organic catalyst at neutral pH. Additionally, it probably provides a scaffold for directing the catalytic fragments, as an alternating sequence of deoxyadenosine and -cytosine was not as efficient. Therefore, NMR studies and titration experiments will be part of future studies to confirm the specific nature of the observed catalytic effect. Bridging the information gap between nucleic acids and amino acid is not only important to expand chemistry templated and catalysed by oligonucleotides, but contributes to the open-ended quest in search of the origin of translation. Current ideas on the origin of life emphasize the intertwining chemistry of both biomolecules involved.[18] Early attempts at peptide formation on nucleic acid backbones, including PNA, were inefficient.[3, 6, 19] A more successful approach was improving self-replicating peptides[20] by introducing nucleobase side chains.[21] By changing the oligonucleotide in our model to RNA, our model may become an example of the catalytic prowess modified nucleobases and short oligonucleotides might have held in the prebiotic world.

Acknowledgements

H.P.M. is a doctoral fellow of FWO Vlaanderen (11B6313N), Belgium. This research received funding from FWO Vlaanderen, Belgium (G078014N). We thank Guy Schepers for providing the oligonucleotides and Chantal Biernaux for editorial help

Notes and references

1 T. Inoue and L. E. Orgel, *Science*, 1983, **219**, 859-862.
2 K. Gorska and N. Winssinger, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 6820-6843.
3 (*a*) A. L. Weber and L. E. Orgel, *J. Mol. Evol.*, 1981, **17**, 190; (*b*) A. L. Weber and L. E. Orgel, *J. Mol. Evol.*, 1979, **13**, 185.
4 (*a*) M. L. McKee, A. C. Evans, S. R. Gerrard, R. K. O'Reilly, A. J. Turberfield and E. Stulz, *Org. Biomol. Chem.*, 2011, **9**, 1661; (*b*) K. M. Wilcoxen, L. J. Leman, D. A. Weinberger, Z. Z. Huang and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2007, **129**, 748; (*c*) W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542-548.
5 B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes and D. A. Leigh, *Science*, 2013, **339**, 189.
6 P. E. Nielsen, *Chem. Biodivers.*, 2007, **4**, 1996.
7 X. Piao, X. Xia, J. Mao and D. Bong, *J. Am. Chem. Soc.*, 2015, **137**, 3751-3754.
8 M. P. Robertson and S. L. Miller, *Science*, 1995, **268**, 702.
9 J. A. McCloskey and P. F. Crain, *Nucleic Acids Res.*, 1998, **26**, 196.
10 J. A. Walder, R. Y. Walder, M. J. Heller, S. M. Freier, R. L. Letsinger and I. M. Klotz, *Proc. Natl. Acad. Sci. USA*, 1979, **76**, 51.
11 S. A. Robertson, J. A. Ellman and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 2722-2729.
12 X.-Q. Liu, L. Jin, C. K. Kim and Y. Xue, *J. Mol. Catal. A: Chem.*, 2012, **355**, 102.
13 A. Patchornik, B. Amit and R. B. Woodward, *J. Am. Chem. Soc.*, 1970, **92**, 6333.
14 D. Beaufils, S. Jepaul, Z. Liu, L. Boiteau and R. Pascal, *Orig. Life Evol. Biosph.*, 2016, **46**, 19.
15 Z. Shlomo, T. P. Vinod, R. Jelinek and H. Rapaport, *Chem. Commun.*, 2015, **51**, 3154-3157.
16 (*a*) X.-P. Song, M. Maiti and P. Herdewijn, *J. Syst. Chem.*, 2011, **2**, 3; (*b*) E. Bigan, H.-P. Mattelaer and P. Herdewijn, *J. Mol. Evol.*, 2016, **82**, 93.
17 (*a*) F. M. Menger, A. V. Eliseev, N. A. Khanjin and M. J. Sherrod, *J. Org. Chem.*, 1995, **60**, 2870; (*b*) D. N. Reinhoudt, D. M. Rudkevich and F. de Jong, *J. Am. Chem. Soc.*, 1996, **118**, 6880.
18 (*a*) B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy and J. D. Sutherland, *Nat. Chem.*, 2015, **7**, 301-307; (*b*) H. Griesser, P. Tremmel, E. Kervio, C. Pfeffer, U. E. Steiner and C. Richert, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 1219-1223; (*c*) H. Griesser, M. Bechthold, P. Tremmel, E. Kervio and C. Richert, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 1224-1228.
19 A. Bencheikh and L. E. Orgel, *J. Mol. Evol.*, 1990, **30**, 315.
20 D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin and M. R. Ghadiri, *Nature*, 1996, **382**, 525.
21 S. Matsumura, T. Takahashi, A. Ueno and H. Mihara, *Chem. Eur. J.*, 2003, **9**, 4829.