



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Sequence-Specific Synthesis  
with Artificial Molecular Machines**

by

**Marcus Papmeyer**

Degree of Doctor of Philosophy

School of Chemistry

The University of Edinburgh

2014

## Table of Contents

Abstract and Layout of Thesis	IV
Declaration	V
Meetings Attended and Presentations Given	VI
Acknowledgment	VII
List of Abbreviations	IX
<b>Chapter I: Sequence-Specific Synthesis</b>	<b>1</b>
1.1 Synopsis	2
1.2 Introduction	3
1.3 Strategies for Sequence Control	9
1.4 Conclusion	28
1.5 References	29
<b>Chapter II: The Road towards Sequence-Specific Synthesis with Artificial Synthetic Machines</b>	<b>37</b>
2.1 Synopsis	38
2.2 Introduction	39
2.3 Results and Discussion	55
2.4 Conclusion	63
2.5 Experimental Section	64
2.6 References	96
<b>Chapter III: Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine</b>	<b>98</b>
3.1 Synopsis	99
3.2 Introduction	100
3.3 Results and Discussion	107
3.4 Conclusion	116
3.5 Experimental Section	117
3.6 References	145

<b>Chapter IV: Rotaxane Elongation as a New Strategy for the Construction</b>	<b>147</b>
<b>of Highly Complex Molecular Machines</b>	
4.1 Synopsis	148
4.2 Introduction	149
4.3 Results and Discussion	154
4.4 Conclusion	159
4.5 Experimental Section	160
4.6 References	178
<b>Outlook</b>	<b>181</b>
<b>Appendix</b>	<b>182</b>

## Abstract and Layout of Thesis

Sequence-specific synthesis is essential to life. In nature, information-rich polymers such as polynucleotides, polypeptides and polysaccharides are responsible for virtually all vital processes. As opposed to nature's optimised approach towards sequence control employing sophisticated molecular machines such as ribosomes and nucleotide polymerases, the synthetic chemist's toolbox for the generation of highly ordered monomeric sequences is limited in scope.

In this thesis, the realisation of the first artificial small-molecule machines capable of synthesising peptides, translating information that is encoded in a molecular strand, is described. The chemical structure of such machines is based on a rotaxane architecture: a molecular ring threaded onto a molecular axle. The ring carries a reactive arm, a thiolate group that iteratively removes amino acids from the strand that block the path of the macrocycle. The acyl monomers are transferred to a peptide-elongation site through native chemical ligation, thereby translating the information encoded in the track into a growing peptide strand. The synthesis is demonstrated with  $\sim 10^{18}$  molecular machines acting in parallel; this process generates milligram quantities of a peptide with a single sequence as confirmed by tandem mass spectrometry.

**Chapter I** describes previous strategies that have been employed to realise sequence specific synthesis and gives an overview about relevant literature in the field.

**Chapter II** describes the concept, previous work and model studies which lay the ground work for the more advanced machines. The first generation design of a molecular machine based on transacylation catalysis as well as the second generation design based on native chemical ligation are discussed. The successful operation of a single-barrier rotaxane capable of elongating its reactive arm by a single amide bond formation and subsequent self-immolation is described.

**Chapter III** describes the first small-molecule molecular machine capable of sequence-specific assembly of a tripeptide. The sequence-integrity of the operation product is determined by tandem mass spectrometry and comparison with an authentic sample.

**Chapter IV** describes a novel synthetic approach towards highly complex molecular machines. Using this rotaxane elongation strategy, a molecular machine with four aminoacyl monomers on the strand is reported. The successful operation afforded the expected product resulting from four amide bond forming events without any detectable sequence scrambling.

## Declaration

The scientific work described in the present thesis was carried out in the School of Chemistry at the University of Edinburgh between September 2009 and July 2012, and between September 2012 and July 2013 at the University of Manchester. Unless otherwise stated, it is the work of the author and has not been submitted in whole or in part in support of an application for another degree or qualification at this or any other University or institute of learning.

Signed:

Date:

## Meetings Attended and Presentations Given

- 1. Organic Research Seminars**, School of Chemistry, University of Edinburgh, UK, 2009-2012.
  - a. Towards a Molecular Peptide Synthesiser
  - b. NHC-Assisted Active-Template [2]Rotaxane Formation
  - c. Macrocyclic NHCs in Catalysis – Synthesis of [2]Rotaxanes
- 2. 38th Scottish Regional Meeting of the Organic Division of the Royal Society of Chemistry**, Dundee, December 2009.
- 3. Fircush Symposium Organic Section**, School of Chemistry, Fircush Point Centre, University of Edinburgh, UK, April 2012.

Poster presentation: Towards the Palladium(0)-Catalysed Synthesis of [2]Rotaxanes.
- 4. Fircush Symposium Organic Section**, School of Chemistry, Fircush Point Centre, University of Edinburgh, UK, April 2012.

Oral presentation: A Molecular Machine for Sequence-Specific Synthesis.
- 5. University of Manchester PhD Symposium**, School of Chemistry, Organic Section, Manchester, University of Manchester, UK, May 2013.

Oral presentation: Sequence-Specific Synthesis with Artificial Molecular Machines.

## Acknowledgment

Above all I want to thank my supervisor David A. Leigh for his continuous support and supervision, for his encouragement and his positive attitude towards all problems along the way. I thank him for allowing me to work in the best laboratory environment possible, on the fascinating projects presented in this thesis.

I was glad to learn from amazing and also amazingly patient co-workers: Dr. Daniel D'Souza who taught me (almost) all secrets of practical organic chemistry and became a good friend over my three years in Edinburgh, Dr. Philipp Gramlich who showed me the humorous side of chemical hardship, Dr. Bartosz Lewandowski, who taught me patience and lead by example, and Dr. Guillaume De Bo with his sense of humour and good advice that allowed me to perform as well as I did.

I would like to thank everybody else that contributed to the molecular peptide synthesiser story, might it be the early years or the late stage completion of the project: Dr. María J. Aldegunde, Dr. Anthony E. Fernandes, Dr. Stephen M. Goldup, Dr. Dominik Heckmann, Sonja Kuschel, and Dr. John W. Ward.

I would like to thank everybody else I had the pleasure working with over the last four years: Adam, Alan, Alina, Alex, Anneke, Antonio, Armando, Augustinas, Barry, Barney, Bea, Chris, Chris Martin, Craig, Daniela, Dave H, Francesca, Francesco, Jack, Jeff, Jhenyi, Jon Beves, Jonathan, Jordi, Kathleen, Kevin, Leo, Louise, Malcolm, Marius, Matt, Max, Michael, Miguel, Miriam, Mustafa, Patrick, Paul, Philipp, Romen, Stewart, Sundus, Tug, Ula, Vanesa, Valerie, and Victor.

Also, I want to thank Dr. Matthew O. Kitching for proofreading the entire manuscript as well as Dr. Daniel M. D'Souza for correcting Chapter 1, Miriam Wilson for Chapter 2, Steffen Woltering for Chapter 3, and Dr. John W. Ward for Chapter 4.

Special thanks go to everyone with whom I had the pleasure to work on various other projects in Edinburgh and Manchester: Jason, Matt, Sonja, Steffen, Steven, and Sundus. I would also like to thank Max von Delius, Dave Howgego and Augustinas Markevicius for taking care of the mass spectrometers in our group. Adam Wilson, Bartosz Lewandowski, Miriam Wilson and Sonja Kuschel are acknowledged for maintaining the Leigh group NMR spectrometer. Big thanks go to my peers: Chris, Dave, Patrick and Ula who shared all stages of a postgraduate degree with me and special thanks go to my lunch-mates, Steven and Steffen.

I would further like to thank several members of the Edinburgh School of Chemistry staff which made working in Edinburgh as pleasant as it was: Annette, Amanda, Martin and Manolo for their administrative work and especially everybody from the Stores for doing a terrific job (Derek, Tim, Raymond, John).

In particular I want to thank my wonderful wife Martina for her love, cheerfulness and understanding during the hard times of this work.

Furthermore, I would like to thank my parents, Andrea Dreyhaupt and Karsten Pappmeyer for their encouragement, help and love on every step of my education and career. Finally, I want to acknowledge my grandparents Elfriede and Hartwig Dreyhaupt for all the support they gave me, past and present.

## List of Abbreviations

$\delta$	chemical shift
$\mu w$	microwave
DNA	deoxyribonucleic acid
A : T	adenine : thymine base pair
AcOH	acetic acid
AFM	atomic force microscopy
Ala	alanine
APCI	atmospheric-pressure chemical ionisation
aq.	aqueous
Boc	<i>tert</i> -butyloxycarbonyl
calc.	calculated
Cbz	carboxybenzyl
COSY	correlation spectroscopy
CPK	Corey-Pauling-Koltun
CuAAC	Cu(I)-catalysed azide-alkyne cycloaddition
Cys	cysteine
DCC	<i>N,N'</i> -dicyclohexylcarbodiimide
DIBAL-H	diisobutylaluminium hydride
DIPEA	<i>N,N</i> -diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	<i>N,N</i> -dimethylformamide
DMSO	dimethylsulfoxide
DTS	DNA-templated synthesis
<i>E</i>	entgegen
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
EF	elongation factor

equiv.	equivalents
ESI	electrospray ionisation
Fmoc	fluorenylmethyloxycarbonyl
G : C	guanine : cytosine base pair
GDP	guanosine diphosphate
Gly	glycine
GTP	guanosine triphosphate
HMBC	heteronuclear multiple bond correlation spectroscopy
HOBt	hydroxybenzotriazole
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence spectroscopy
Hz, MHz	Hertz, megahertz
Leu	leucine
LRMS	low resolution mass spectrometry
m.p.	melting point
<i>m/z</i>	mass-to-charge ratio
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
mRNA	messenger RNA
MS	mass spectrometry
NCL	native chemical ligation
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
Phe	phenylalanine
Piv	pivaloyl
PNA	peptide nucleic acid
ppm	parts per million
PTC	peptidyl transferase centre
PyBrOP	bromotrispyrrolidinophosphonium hexafluorophosphate

RNA	ribonucleic acid
RT	room temperature
SPS	solid-phase synthesis
TBAF	tetra- <i>n</i> -butylammonium fluoride
TBTA	tris([1-benzyl-1 <i>H</i> -1,2,3-triazol-4-yl]methyl)amine
<i>t</i> BuOH	<i>tert</i> -butanol
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	trimethylsilyl
tRNA	transfer RNA
Trt	trityl
Z	zusammen

Note: conventional abbreviations for units, constants and physical quantities are not included.

## Chapter I

### Sequence-Specific Synthesis

#### **ACKNOWLEDGMENTS**

Chapter I was written in its entirety by Marcus Pappmeyer. Dr. Daniel M. D'Souza and Steffen Woltering are gratefully acknowledged for their fruitful advice, discussion and proofreading of Chapter I.

## 1.1 Synopsis

Sequence-specific synthesis is essential to life. In nature, information-rich polymers such as DNA, RNA, peptides, and carbohydrates are responsible for information storage and replication, translation of information, catalysis of essential reactions, energy storage, and are employed as structural components. As opposed to nature's approach towards sequence-control using sophisticated molecular machines such as ribosomes and nucleotide polymerases, the chemists' synthetic toolbox for the generation of highly information-rich monomeric sequences is limited in scope, but has developed considerably in recent years. The efficient and selective synthesis of sequence-specific polymers holds the promise for the development of new materials with unprecedented micro- and macroscopic properties. In this first chapter, the ribosome as gold standard for sequence-specific catalysis is analysed and assessed, and the various synthetic strategies developed, ranging from the solid-phase synthesis of oligomers to the synthesis of information-rich synthetic polymers over DNA-templated synthesis that finally lead to the development of small-molecule ribosome mimetics, are discussed.

## 1.2 Introduction

Sequence defined polymers are essential to life and are found in every single living organism. Nature's solution for the chemical challenges all organisms are exposed to is the incorporation of sequence-defined polymers in all processes of life.<sup>1</sup> In order to cope with fundamental problems such as information storage and replication, translation of information, catalysis of essential reactions, energy storage and structure three major classes of information-rich biopolymers have evolved: polynucleotides, polypeptides and polysaccharides.

The individual properties of the incorporated monomers and their sequence in the primary structure of the single-chain polymer strand allow the formation of kinetically stable three-dimensional structures under physiological conditions. Only the ability to form well-defined architectures with the possibility of precise positioning of active groups in space enables the unique catalytic properties found in enzymes.

In nature, sequence control of polymers is realised through the operation of complex molecular machines consisting of sequence-defined polymers themselves.<sup>2</sup> In case of DNA, the synthesis of a new strand is achieved through the action of DNA polymerases, enzymes that transcribe the information contained in a parental DNA strand by connecting the four nucleotide building blocks in the complementary order (DNA replication). Similarly, the sequence-specific synthesis of RNA is realised by RNA polymerases, like the DNA polymerases molecular machines capable of transcribing a single DNA strand to a complementary RNA strand incorporating uracil in all instances where thymine would have occurred (transcription). Polypeptides on the other hand are most commonly synthesised by the action of ribosomes, extraordinary molecular devices that translate the genetic code into a sequence-defined polypeptide by moving directionally along an mRNA strand (translation). In contrast to polypeptides and polynucleotides, the sequence and structure of polysaccharides is commonly less well-controlled.<sup>3</sup> Their structure can range from linear to highly branched, and the monomeric sequence can vary, sometimes containing only a single monomeric repeating unit, in other instances consisting of a variety of monosaccharides with different linkages and substitution patterns. Their moderately sequence-controlled synthesis is achieved *in vivo* by the complex interplay of several cooperatively interacting enzymes but is only partly understood.<sup>4</sup>

In sharp contrast to nature's ability to generate information-rich monodisperse polymers, most approaches used by chemists today are limited in scope. Solution phase synthesis requires elaborate protection/deprotection sequences in order to built up sequence-specific

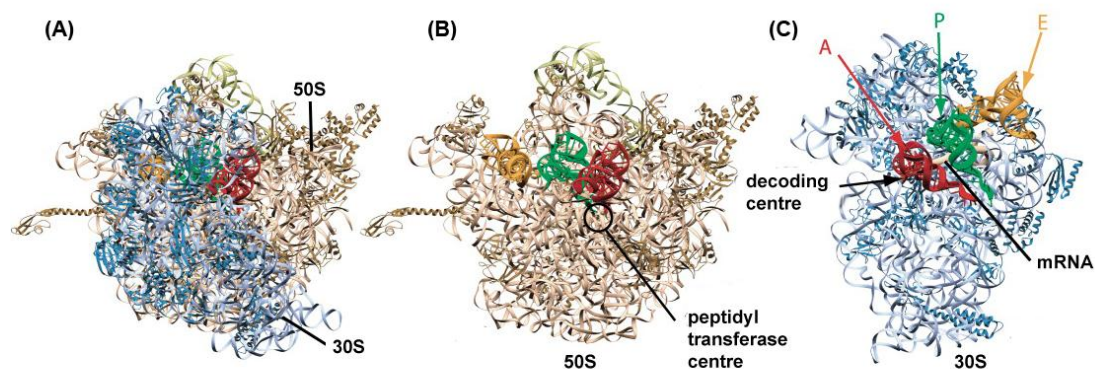
chains and in case of classical step and chain growth polymerisation a precise positional control of co-monomers in the growing polymer chain is not possible and usually a statistical distribution of monomers is found in synthetic copolymers.<sup>5</sup> Although polymers with narrow molecular weight distributions<sup>6</sup> and well-defined tacticities<sup>7</sup> and macromolecular architectures<sup>8</sup> have been realised, the field of sequence-controlled synthesis of polymers remains largely unexplored. It is obvious that the construction of single polymer chains with well-defined primary structure would allow the synthesis of materials with novel micro- and macromolecular properties as found in nature. New techniques using oligonucleotide templates or molecular machines based on synthetic small-molecules are emerging in order to overcome those difficulties but remain thus far limited to short oligomeric products.

This introductory chapter aims to provide an overview of the different approaches that have been applied for sequence control in chemical synthesis. The function of the ribosome as the prime example and benchmark for a sequence-control in polymerisation is assessed and discussed from a synthetic chemist's perspective. Approaches towards sequence control that have appeared in literature, ranging from the use of a solid support over DNA templates to sophisticated ribosome mimetics, are described with a focus on the underlying concepts.

### 1.2.1 The Gold Standard of Sequence Control: the Ribosome

In all living cells, the translation of information from the genetic code into sequence-specific polypeptides is achieved through the action of the ribosome, a highly complex molecular machine.<sup>2</sup> Initially discovered in the 1940s, ribosome research took off at the end of the 1970s with extensive biochemical studies on function and components.<sup>9</sup> Since the publication of the first atomic resolution crystal structure of ribosomal subunits in 2000<sup>10</sup> the understanding of the function and structure of the ribosome has made remarkable progress and was crowned in 2009 with the Nobel Prize in Chemistry for Ramakrishnan, Steitz and Yonath.<sup>11</sup>

Structural studies revealed that the ribosome is composed of two subunits of unequal size (Figure 1.1, A).<sup>12</sup> In bacteria the 70S<sup>13</sup> ribosome consists of a large 50S and a small 30S subunit (Figure 1, A and C respectively), which together form a 2.3 MDa protein-RNA assembly. In their eukaryotic counterparts, the 80S ribosome is considerably larger and is composed of subunits with sedimentation coefficients of 60S and 40S and a total weight of 4.2 MDa. The surfaces of the two subunits responsible for the self-assembly and most parts of the core architecture consist mainly of RNA whereas in peripheral parts of the ribosome proteins are incorporated.

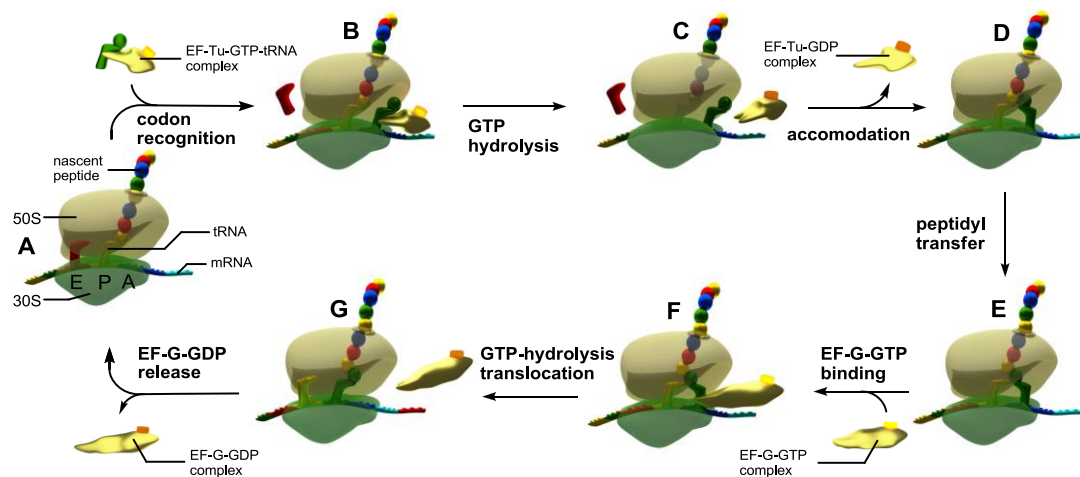


**Figure 1.1.** Crystal structures of the ribosome. (A) 70S ribosome complexed with tRNA and mRNA. (B) Exploded view of the 50S subunit, and (C) exploded view of the 30S subunit; tRNA shown in red (A-site), green (P-site) and orange (E-site). Adapted with permission from Elsevier.

For the ribosomal translation of the genetic code two RNA components are essential: (i) a messenger RNA (mRNA) which is a transcript of a DNA strand and carries the actual information that is translated by the ribosome and (ii) a transfer RNA (tRNA), the building block that carries an activated aminoacyl residue and a three-nucleotide recognition site referred to as a codon that enables the interaction with the mRNA in the translation cycle.

The mRNA that is read-out during translation fits neatly into a cleft of the smaller subunit, which mediates the codon/anti-codon interactions of the incoming tRNA. The peptidyl transferase centre (PTC) responsible for the actual catalysis of the peptide chain transfer is located in the large subunit (Figure 1.1, B) in close proximity to the three binding sites for the tRNA found in the ribosome (red, green, orange structure respectively, Figure 1.1, C): (i) the A-site that interacts with the incoming aminoacyl tRNA, (ii) the P-site that is occupied by the tRNA loaded with the emerging peptide strand, and (iii) the E-site (exit) where the deacylated, “spent” tRNA is transferred to from the P-site after peptide chain elongation.

The process of translation is commonly divided into three stages: initiation, elongation and termination.<sup>12a</sup> During initiation the actual ribosome mRNA complex is assembled from its subunits around the start codon of the mRNA through the action of several protein initiation factors.<sup>14</sup> In the elongation cycle, the assembled ribosome then initiates the translation and consequently the synthesis of the polypeptide (Figure 1.2). At the start of this sequence (Figure 1.2, A), both E- and P-sites are occupied with tRNA exhibiting codon-anticodon binding, whereas the A-site remains empty. The P-site tRNA is carrying the nascent polypeptide chain and the deacylated tRNA remains in the E-position.



**Figure 1.2.** Schematic overview of the ribosomal elongation pathway. For simplicity, not all intermediate steps deduced from kinetic experiments are shown.

For a successful translation, elongation factors (EF), comparatively small proteins that facilitate the translation and mediate interactions between the tRNA, mRNA and the ribosome, are essential. The EFs responsible for the translation cycle (EF-Tu and EF-G) both have a GTP hydrolysis site that allows the conversion to GDP. Upon correct codon/anticodon recognition at the A-site with the corresponding ternary EF-Tu-GTP-tRNA complex (Figure. 2, B), GTP is hydrolysed to give the corresponding EF-Tu-GDP fragment. This fragment can now dissociate from the ribosome, while the tRNA building block carrying the amino acid is delivered to the A-position.<sup>15</sup> Now a series of interconnected events happen: the deacylated, “spent” tRNA in the E-site dissociates and triggers the accommodation of the aminoacyl-end of the “loaded” A-site mRNA in the PTC (Figure. 2, D). The nascent peptide chain is transferred from the middle P-site tRNA to the A-site tRNA (Figure.2, E) and the new amide bond is formed.<sup>16</sup>

In order to achieve a reset of the machine, the spent tRNA blocking the P-site has to be transferred to the E-site. This is achieved through the action of the EF-G-GTP complex and subsequent GTP hydrolysis (Figure. 2, F).<sup>17</sup> Accordingly, the peptidyl-carrying A-site tRNA is moved to the P-site, and the whole ribosome moves along the mRNA in the 3' direction, leaving the A-site empty for the next EF-Tu-GTP-tRNA binding event. The elongation continues until the ribosome encounters a stop codon. Aided by ribosome release factors the ribosome mRNA complex dissociates and the formed polypeptide chain is released to the cellular pool (termination).<sup>18</sup>

Looking at the schematic description of the ribosomal translation from a chemist's perspective attempting to learn about sequence-specific synthesis from nature, two (interconnected) key questions remain unanswered:

(i) How exactly does the ribosome achieve its near-perfect sequence-specificity even for very similar codons and/or aminoacyl residues?

(ii) What mechanism is responsible for uni-directional movement of the ribosome along the mRNA strand?

The most obvious answer to the correct decoding of the mRNA strand is the free energy difference of a matching codon-anticodon pair of mRNA and tRNA compared to the mismatched case. However, with respect to the remarkably high fidelity of decoding leading to error rates in the order of  $10^{-4}$ , the remarkably high accuracy of tRNA selection cannot be solely accounted for by the favourable Watson-Crick base pairing binding energies of a cognate codon.<sup>19</sup> A free energy difference of 18-24 kJ/mol would be required to explain the error rates which are observed both *in vitro* and *in vivo* but the energy differences of cognate (two complementary trinucleotide sequences) versus near-cognate (one out of three nucleotide pairs mismatching) base pairs are estimated to be the order of only 2 to 6 kJ/mol.<sup>20</sup> So if not realised through energetically favourable recognition (alone) how is the remarkably low error rate achieved?

One process that supports the selection of the cognate RNA is the so-called kinetic proofreading<sup>21</sup> which happens after the codon-independent recognition of the ternary EF-Tu-GTP-tRNA complex by the ribosome. In most cases, mismatching tRNA is expelled before GTP hydrolysis by EF-Tu but in case near-cognate RNA is bound and GTP hydrolysis as well as EF-Tu dissociation are triggered, incorrect tRNA can be expelled before the accommodation step takes place (faster dissociation of near-cognate compared to cognate tRNA).<sup>22</sup>

A second process aiding the differentiation between similar codon-anticodon pairs is the induced-fit mechanism<sup>19a,23</sup>, which accelerates the accommodation of the cognate tRNA over the near-cognate tRNA in the PTC, and therefore does not only make use of the differences in rates of dissociation but also in rates of association of tRNA to mRNA.<sup>24</sup> By interaction of ribonucleotide bases of the ribosome with the minor groove of the first two base pairs of the codon-anticodon pairs in the A-site, a domain closure in the 30S subunit is induced and the incorporation of cognate versus non-cognate tRNA in the PTC is accelerated.<sup>25</sup>

This multiplicative exploitation of differences enhances the sequence fidelity significantly compared to a hypothetical mechanism based solely on Watson-Crick base pairing of codon and anticodon.<sup>26</sup> Nature makes use of several processes that are mutually supportive in order to generate an optimal result.

One sometimes overlooked essential requirement for the remarkably small error rates in the translation process is the synthesis of tRNA loaded with the correct aminoacyl residue,

which is carried out independently from the action of the ribosome. If tRNA with a specific codon was loaded with a mixture of several amino acyl residues, even the most efficient ribosome would fail to deliver a peptide with a single sequence.

The aminoacyl-tRNA synthetases that realise the loading of the tRNA can select even from structurally very closely related aliphatic amino acids such as valine, leucine and isoleucine specifically.<sup>27</sup> For example, in case of isoleucyl-tRNA synthetases, it has been found that a double sieve mechanism is in operation.<sup>28</sup> In the first step, both valine and isoleucine are activated and loaded on the corresponding isoleucyl-tRNA, but the slightly larger leucine is excluded (“coarse” sieve). In a second editing step, the enzyme selectively hydrolyses all valylated products (“fine” sieve). As in the case of the selection process of the ribosome, the realisation of sequence fidelity achieved by the complex interplay of different mechanisms and based on amino acyl residues loaded onto—the error rate is typically in the order of  $10^{-4}$ —their corresponding carrier.<sup>29</sup>

Finally, the directional movement of the individual ribosomal components is essential for the successful translation of the mRNA strand. But how is directional movement of the ribosome achieved?

Interestingly, movement of the ribosome is possible in both directions in the absence of elongation factor EF-G and GTP.<sup>30</sup> Both translocation and retro-translocation are observed, albeit at relatively low reaction rates; tRNAs can move through the ribosome at physiological temperature in order to find the energetically most favourable position (tRNAs have, depending on their nature, a preferred site of occupation, i.e. E, P or A). The observation of bidirectional movement strongly suggests that a Brownian ratchet mechanism is in operation.<sup>31</sup>

In order to “pay” the thermodynamic penalty of directional movement a fuel needs to be consumed by the ribosome. This input of energy could be the formation of the energetically more favourable peptide bond or one of the two GTP hydrolysis events during the elongation cycle (or a combination of both).

In reality, the directional movement occurs after EF-G-GTP binding (the second GTP hydrolysis) and seems to be independent from the amide bond formation.<sup>17b,32</sup> The rate-limiting step during the translocation reaction is believed to be the movement of the tRNA anticodon stem-loops on the 30S subunit.<sup>17b</sup> Binding of EF-G-GTP lowers the energy barrier of this transition probably through stabilisation of an otherwise unfavourable transition state.<sup>33</sup> It is not entirely clear whether the decisive translocation step that follows is a power-stroke mechanism, i.e. that the chemical energy from GTP hydrolysis is converted to mechanical energy and therefore biases the forward movement or whether GTP hydrolysis is

just triggered by the forward movement of the ribosome and prevents the reverse reaction from happening.<sup>34</sup> Further experiments will have to indicate which mechanism is in operation.

## 1.3 Strategies for Sequence Control

Nature's capability to control complex monomer sequences remains unrivalled by modern polymer chemistry standards. Consequently, the most straightforward approach towards the synthesis of information-rich polymers is the exploitation of ribosomes and polymerases *in vitro*. Advanced techniques have been developed in recent years ranging from the amplification of DNA (polymerase chain reaction [PCR]) to the incorporation of unnatural amino acids or even the creation of new polymer linkages such as polyesters using nature's machinery.<sup>1,35</sup> But although the use of enzymes *in vitro* for the synthesis of sequence-defined polymers represents a fascinating (and powerful) strategy, the ultimate aim is the synthesis of sequence-defined polymers of any set of co-monomers—natural or synthetic—in order to generate novel functional materials not limited by reaction conditions and a narrow substrate scope.<sup>5c</sup> In the following paragraphs several strategies towards the synthesis of sequence controlled oligomers and polymers will be highlighted and discussed.

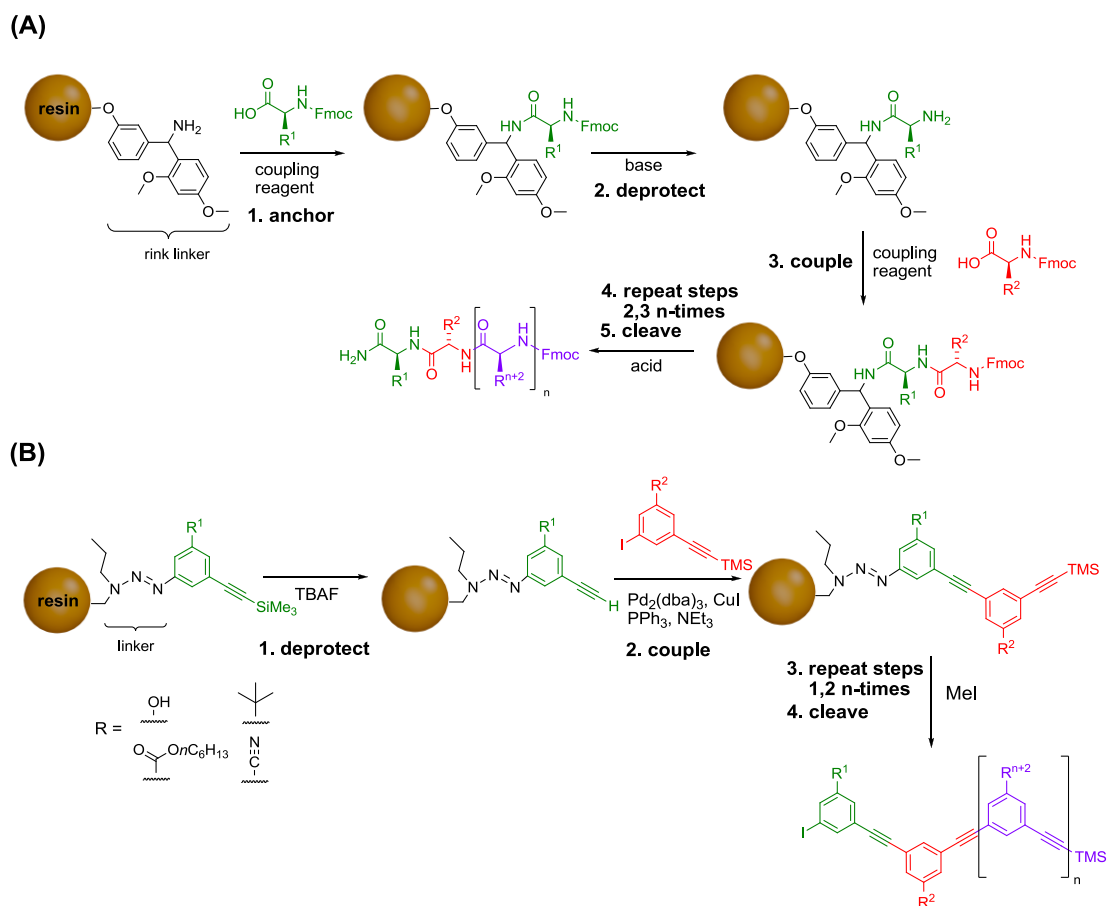
### 1.3.1 Solid Phase Synthesis

Today, the most powerful and most often applied method for the synthesis of sequence-specific biomolecules is the solid-phase synthesis (SPS) introduced in 1963 by Merrifield,<sup>36</sup> who was subsequently awarded the Nobel prize in Chemistry in 1984 for his groundbreaking invention.<sup>37</sup>

In the literature, a vast number of reviews on SPS have appeared, dealing with peptide synthesis specifically<sup>38</sup> or covering other aspects and reactions on solid support<sup>39</sup>. The following paragraph will therefore only highlight selected approaches that focus on sequence-specificity aided by polymer matrix support.

Typically for solid-phase peptide synthesis<sup>40</sup> an *N*-protected amino acid (Scheme 1.1, A) is connected *via* a cleavable linker to a poly(styrene) resin. After deprotection, a second *N*-protected amino acid can be attached. This deprotection/chain-extension cycle is then repeated until the desired peptide sequence is obtained and the target molecule is cleaved from the solid support.

This process allows the facile synthesis of peptides up to 50 amino acids and has been used for the total synthesis of proteins pioneered by Hirschmann and Merrifield as early as 1969.<sup>41</sup>



**Scheme 1.1.** Polymer matrix assisted synthesis of sequence-defined oligomers. (A) Solid phase peptide synthesis originally developed by Merrifield.<sup>42</sup> (B) Sequence-specific synthesis of oligophenylacetylenes introduced by Moore.<sup>43</sup>

The main advantages that lead to the widespread use of SPS<sup>40</sup> are the ease of procedure and especially the rapid purification by filtrating and washing off by-products from the matrix-bound polypeptide. Additionally, the high yield compared to the solution phase approach for multiple peptide couplings and the possibility for automation of the process are major advantages. Using peptide-synthesising machines, large libraries of peptides are accessible with ease.

In principle, SPS can be applied to all step growth polymerisations as long as a suitable, high yielding and clean protection/deprotection strategy is available. Advanced approaches have been developed for the solid phase synthesis of oligonucleotides<sup>44</sup> and oligosaccharides<sup>45</sup>. The well-established peptide bond formation methodology together with fluorenylmethoxycarbonyl (Fmoc)- or *tert*-butyloxycarbonyl (Boc)-protecting groups for the *N*-terminus has been employed for the synthesis of  $\beta$ -peptides<sup>46</sup>,  $\gamma$ -peptides<sup>47</sup>, peptoids<sup>48</sup>, glycopeptides<sup>49</sup> peptide nucleic acids<sup>50</sup>, peptidosulfonamides<sup>51</sup> and unnatural polyamides<sup>52</sup>.

Because of the interest in synthetic foldamers with their potential biological applications<sup>53</sup>, SPS-based strategies for the synthesis of sequence-specific oligoureas<sup>54</sup> and oligocarbamates<sup>55</sup> have been developed as well.

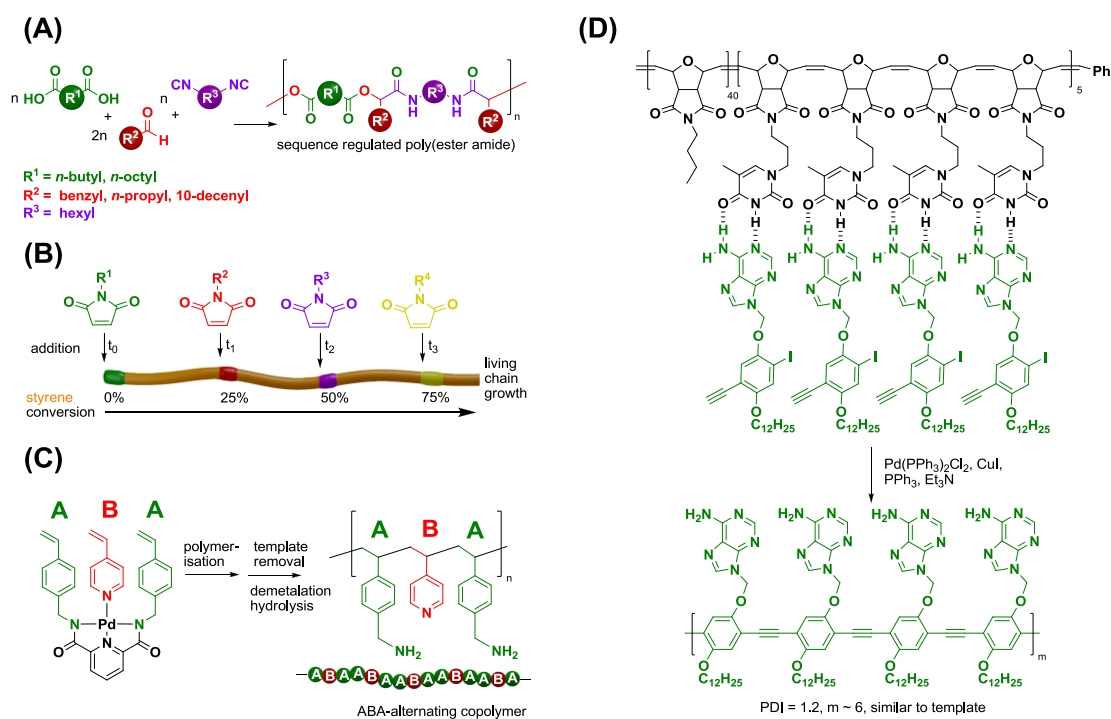
In addition, C-C bond forming reactions have been employed to synthesise polymers *via* the solid phase approach. The use of SPS for the synthesis of conjugated sequence-specific oligomers was demonstrated in 1994 (Scheme 1.1, B).<sup>52c,56</sup> Employing TMS-protected 3-iodo phenylacetylenes, the synthesis starts with a resin bound TMS-protected phenylacetylene derivative. Liberation of the alkyne moiety using TBAF and subsequent addition of a TMS-protected 3-iodo phenylacetylene under Sonogashira coupling conditions allows the stepwise built-up of sequence-defined oligomers. Cleavage from the resin with methyl iodide, ipso-substitution the triazene group for iodine, delivers the target molecule.

Subsequently the SPS approach has been extended to the synthesis of a large range of oligomers incorporating different reactions and substrates. Oligothiophenes can be obtained by alternating bromination and Stille coupling<sup>57</sup>. Other reactions such as Suzuki coupling<sup>58</sup>, reductive amination<sup>59</sup>, esterification<sup>60</sup> and nucleophilic substitution<sup>61</sup> have also been employed.

Despite the evident advantages of SPS, certain limitations remain. In contrast to the biopolymers synthesised by nature, all monomers have to be added sequentially and the sequence selectivity is achieved by step-wise deprotection and elongation which requires individual work-up and purification after each monomer addition. Furthermore, to obtain sequence selectivity, yield or the bond forming reaction, subsequent deprotection and final purification are crucial.

### 1.3.2 Liquid-Phase Step- and Chain Growth Polymerisations

In order to generate high molecular weight polymers with sequence control, bulk polymerisation techniques are desirable especially considering the limitation of iterative synthesis approaches to generate oligomers with more than a few monomeric units. Chain and step growth polymerisations of co-monomers have been used for the synthesis of copolymers in many instances, but sequence-specificity rather than a random incorporation of different units is problematic to realise in practice. In the last years numerous approaches towards the sequence-control of polymers have been developed in order to overcome those inherent problems and sequence-controlled polymerisations have been an emerging field of scientific interest.<sup>5b</sup> For further information, the excellent reviews and perspectives from Lutz are highly recommended.<sup>5b-d</sup>



**Scheme 1.2.** Different approaches towards sequence-controlled polymers. (A) Step-growth Passerini polymerisation of terminal functionalised dicarboxylic acids, bisocyanides and an aldehyde from Li.<sup>62</sup> (B) Sequential addition of *N*-functionalised maleimide co-monomers to a living radical styrene polymerisation introduced by Lutz.<sup>63</sup> (C) Sequence-regulated radical polymerisation using a Pd-templated monomer from Ouchi and Sawamoto.<sup>64</sup> (D) Nucleobase-templated Sonogashira polymerisation pioneered by Sleiman.<sup>64</sup>

Step-growth polymerisation represents the most straightforward strategy for the synthesis of alternating copolymers, the simplest case of a defined monomer sequence. Sequence control is possible using monomers of the A-A and B-B or A-B type.<sup>65</sup> On the other hand, more complex sequences are difficult to realise.

Representing a notable exception, employing the Passerini reaction, a step-growth ABAC-type polymer was synthesised by Li and co-workers (Scheme 1.2, A).<sup>66</sup> The classical three-component Passerini reaction allows the synthesis of a depsipeptide from an isocyanide, an aldehyde (or ketone), and a carboxylic acid.<sup>67</sup> By employing a dicarboxylic acid, an aldehyde, and a diisocyanide it was possible to incorporate the sequential information into the structure of the polymer chain affording a sequence-regulated poly(ester-ester-amide-amide) in a one-pot process.

In the case of chain growth polymerisation, where the chain propagation is achieved by the reaction of highly reactive active centres, i.e. radicals or ions, monomer sequence control is inherently more problematic and in most cases as a result a statistical copolymer is obtained.

One general strategy that allows the synthesis of sequence-defined polymers using chain growth polymerisations exploits reactivity differences of co-monomers. For example, the metal-catalysed copolymerisation of carbon monoxide with epoxides leads to the formation of poly( $\beta$ -hydroxyalkanoate)s since neither carbon monoxide nor epoxides have a tendency to homopropagate under the reaction conditions.<sup>68</sup> Catalyst control is another method for the selective alternating incorporation of monomers and suppression of homopropagation in the growing polymer chain. This method was used by Thomas and co-workers in a very elegant synthesis of syndiospecific alternating polyesters.<sup>69</sup> In the reaction mixture, enantiomerically pure but differently functionalised  $\beta$ -lactone monomers were polymerised using an yttrium complex as initiator and delivering a highly alternating polyester.

But also chain-growth polymerisations of non-metal catalysed reactions can be employed for the generation of alternating polymers. Because of its high tendency for cross-propagation over homopropagation, maleic anhydride or *N*-substituted maleimides have been used for the synthesis of A-B type polymers.<sup>70</sup> The controlled radical copolymerisation of maleic anhydride or *N*-substituted maleimide derivatives and styrene gives highly alternating copolymers, since the electron-poor maleic anhydride is preferentially incorporated into the living polymer chain over styrene, even if a large excess of styrene is present, but has a low tendency to homopropagate.

In a particular example by Kamigaito and co-workers another aspect of maleimide reactivity was exploited for the generation of an AAB-type polymer. In case of the radical copolymerisation of phenylmaleimide (A) and D-limonene (B), it was observed that the intermediate adduct AA of two phenylmaleimide units strongly favours the incorporation of monomer B over A, whereas an AB terminus exclusively reacts with another maleimide monomer A.<sup>71</sup> The solvent choice is critical for the outcome of this polymerisation, as fluorinated alcohols like  $\text{PhC}(\text{CF}_3)_2\text{OH}$  favour the incorporation of another maleimide monomer to an AB-unit, supposedly through hydrogen bonding to the maleimide's carbonyl groups.

A different approach towards sequence-defined polymers that makes use of the sequential addition of monomers to a growing polymer chain was introduced by Lutz and co-workers in 2007 and subsequently extended in the following years allowing the incorporation of individually functionalised maleimide monomers in a polystyrene chain with high precision (Scheme 1.2, B).<sup>72</sup> The position and order of differently functionalised monomers is achieved by addition of differently *N*-substituted maleimides during the living chain growth at different conversions of the styrene co-monomer. Adding exactly one equivalent of maleimide with respect to the number of the living polystyrene chains, precisely one of the

more reactive co-monomer units is incorporated in the growing polystyrene chains due to the fact that maleimides have no tendency to homopropagate. Since a living polymerisation is employed, the length of the simultaneously growing chains is very similar and therefore the positional control of the insertion is relatively accurate. Tuning the reaction conditions and adding the maleimide co-monomers at high styrene conversions, the incorporation in a zone of uncertainty of only 1 or 2 styrene units in average has been achieved.<sup>72a</sup> This sequential addition of co-monomers to a growing polymer chain in order to generate sequence-regulated polymers can be applied, if (i) in the reaction mixture all polymer chains grow simultaneously, (ii) the added co-monomer is more reactive than the co-monomer present in the reaction mixture and therefore incorporated shortly after addition, and (iii) cross-propagation of the added monomer is disfavoured. Unfortunately, with the exception of the styrene/maleimide pair introduced by Lutz no other examples for the sequential addition approach have been reported to date.

Finally, chain-growth sequence-regulated polymers have been synthesised through the use of templates that arrange the co-monomers in a certain sequence. Pioneered by Ouchi and Sawamoto it was possible to obtain AB alternating sequences by controlled radical cyclopolymerisation of methylmethacrylate and a methacrylate held in position by ester linkages onto a naphthalene scaffold.<sup>73</sup> After hydrolysis, an alternating methylmethacrylate-acrylate polymer was obtained.

This strategy was extended to ABA-alternating copolymers (Scheme 1.2, C) exploiting  $\pi$ - $\pi$ -stacking interactions between two aromatic side groups (A) and a pyridine ligand (B) held in place by a palladium atom.<sup>74</sup> Demetallation of the polymer and cleavage of the amide bonds to remove the template lead to an ABA-sequence regulated copolymer.

Ideally, monomers held in place by supramolecular interactions would allow the translation of templates without the need of covalent linkage and post-polymerisation cleavage of the templated polymer strand. First studies towards the template-assisted radical polymerisation relying on supramolecular interactions with the ultimate aim of sequence control have been reported by Ouchi and Sawamoto, but the successful implementation into a sequence-controlled polymerisation has yet to be demonstrated.<sup>75</sup>

Sleiman and co-workers reported an interesting nucleobase-templated polymerisation that allowed to inherit the chain length and polydispersity of a parental template strand with narrow molecular weight distribution to the daughter strand, although no sequence control was possible (Scheme 1.2, D).<sup>76</sup> Usually, step-growth polymerisations suffer from poor molecular weight control, but the preassembly of nucleobase-functionalised monomers on a template through Watson-Crick base pairing enabled the synthesis of daughter strands

through Sonogashira polymerisation with narrow molecular weight distribution. Without template, short conjugated oligomers with a significantly higher polydispersity index and lower average molecular weight are formed.

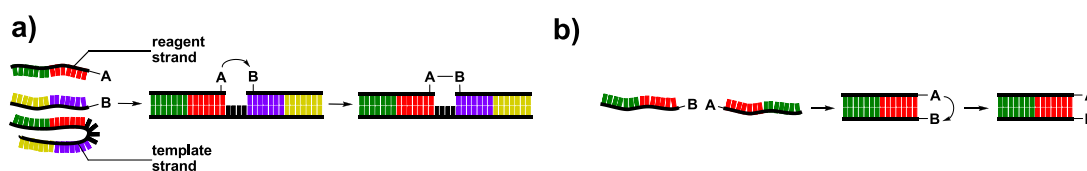
Similarly, Luh and co-workers demonstrated the possibility of polynorbornene replication. Norbornene monomers were assembled on a polynorbornene strand through ester linkages. After ring opening metathesis polymerisation, the daughter strand was removed by saponification, transferring the chain length of the parental template to the replicated strand.<sup>77</sup>

### 1.3.3 Introduction to DNA-Templated Methods

Although step- and chain growth polymerisation allow the synthesis of high molecular weight polymers with narrow weight distributions, to date the sequence-specificity of the polymer chains is limited to small repetitive segments. One reason for the limitation of synthetic polymer chemistry to simple repetitive sequences of monomers is the use of statistical reactions and the lack of information-rich templates that feature selective interactions between monomers and the template strand.

The use of information-rich biopolymers as templates for the generation of synthetic sequence-specific polymers seems therefore a promising approach towards the ultimate goal of precise monomer positioning. Self-evidently, early on DNA has been identified as potential template for polymerisations. Introduced more than 40 years ago by Naylor<sup>78</sup>, a focus of research has been laid upon DNA templated synthesis (DTS) over the last 15 years and the field has been subject to several reviews.<sup>79</sup> DTS relies on the sequence-specific hybridisation of longer oligonucleotide strands; the most commonly used DTS approaches are shown in Figure. 3. Typically, a DTS reagent consists of a DNA oligonucleotide with the sole function of mediating the interactions with the template and a linker that connects the reactive group selected for the templated reaction with the DNA strand (Figure 1.3, A).<sup>79a</sup> The tethered oligonucleotide template itself only interacts with the DNA strands connected to the reactive groups but does not participate in the reaction. Alternatively, an oligonucleotide strand linked to a reactive group can be hybridised directly with a complementary DNA-linked reagent without the need of an additional template strand (Figure 1.3, B).

Due to the high affinity of the DNA oligonucleotides to the template a high effective molarity of reactants is realised, which results in an enhanced reaction rate between the substrates. Accordingly, with DTS it is possible to direct reactions between reactants which would lead to significant cross-over reactions in classical solution phase synthesis.<sup>80</sup>



**Figure 1.3.** General strategies utilised in DNA templated synthesis. (A) Two reactive groups A and B tethered to a DNA-oligonucleotide strand brought together mediated by a DNA-template strand. (B) Reactive groups A and B held together by mutual interactions of their DNA tethers; matching colours indicate complementary sequences between strands; lines indicating base pairing do not represent a particular number of bases.

Furthermore, libraries of DTS reagents can be exposed to *in vitro* selection and PCR amplification. The powerful approach of DTS for controlling the effective molarity of different reactants in solution represents an ideal tool for the realisation of sequence-specific synthesis of either complementary DNA strands or for the translation of information into chemically unrelated synthetic sequences.

### 1.3.4 DNA-Templated Transcription

Initially, DTS was mainly investigated with view on *in vitro* transcription. In the search for ways to non-enzymatic molecular replication—a prerequisite for the emergence of life<sup>81</sup>—DNA templated coupling of activated oligonucleotides has been attempted as early as 1966 by Naylor and Gilham.<sup>82</sup> Seminal work from Orgel and co-workers showed the possibility to generate a complementary (-)-strand to a short RNA-template using suitably activated nucleotide substrates.<sup>83</sup>

Even more intriguing, an autocatalytic full replication cycle of an RNA-hexamer in the presence of two activated trinucleotides, generating the original (+)-RNA strand has been realised by von Kiedrowski in 1986.<sup>84</sup> But although impressive examples of DNA-templated synthesis have been realised, the use DNA-templated polymerisation of nucleotides has faced two substantial limitations.

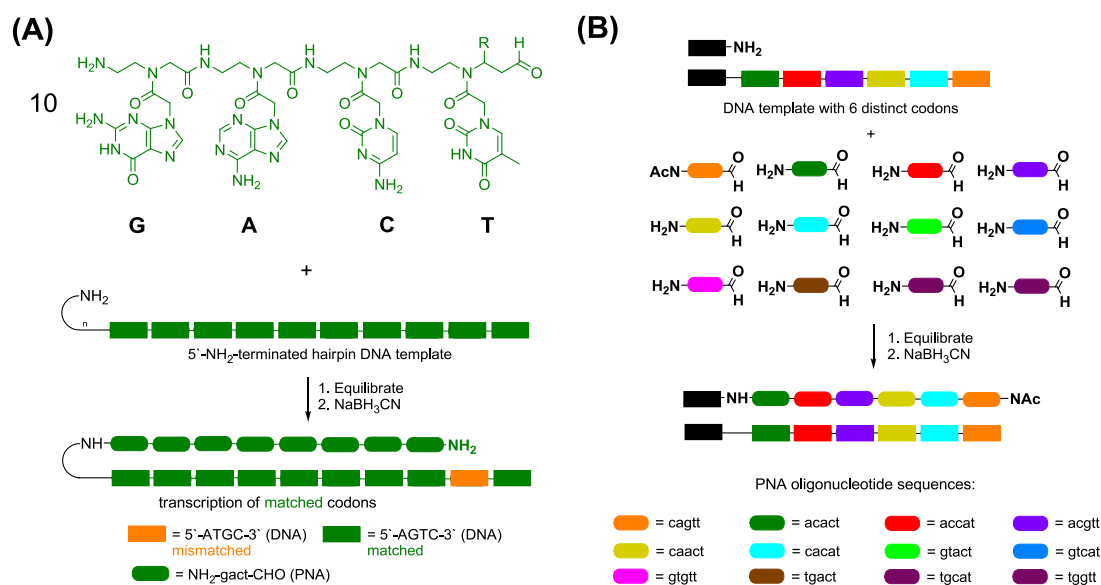
First, all non-enzymatic replication reactions using mononucleotide interactions have been low-yielding with respect to the generated oligomers. One reason for the low yield of the desired strands is the efficiency of the coupling steps. Even for relatively high-yielding reactions a low overall yield is obtained if many connections have to be formed. For example, assuming a coupling efficiency of 75% for each individual step and 10 coupling steps in total, the theoretical yield of the resulting 11-mer would be only 5.6%. Accordingly, in the quite impressive synthesis of an RNA 14-mer by transcribing a DNA strand with mononucleotides from Acevedo and Orgel, the desired target strand could only be isolated in 2% overall yield.<sup>83b</sup>

Second, the sequence accuracy of the strands synthesised was generally not determined in the early work by Orgel and von Kiedrowski.<sup>79a</sup> Later studies carried out by Rojas, Stütz, and Richard showed that error rates of DNA-templated mononucleotide coupling reactions can be greater than 50% for A:T and around 30% for G:C pairs, indicating that sequence-specificity is difficult to achieve with individual mononucleotide interactions.<sup>85</sup> Attempts to overcome problems with yield and sequence fidelity have been undertaken, e.g. by replacing the problematic phosphodiester formation step by using a phosphoramidate backbone,<sup>86</sup> or the use of modified nucleobases that show enhanced interactions<sup>86b,86d</sup>, but to date no enzyme-free copying mechanism showing high sequence fidelity and high yield has been reported.

From these limitations conclusions for the application of templated reactions for the synthesis of non-natural polymers can be drawn. The requirements for the successful templated enzyme-free sequence-specific synthesis of oligomers are: (i) a highly efficient reaction for the coupling of the individual monomer units, (ii) a reaction that is distance-dependent, i.e. successful reactions are only possible with the nearest neighbour of each monomer, (iii) the interactions between the template and the monomers are efficiently strong to hold components in place, and (iv) recognition of each monomer is achieved by a selective interaction able to distinguish between the different monomers in solution.<sup>87</sup>

Liu and co-workers realised the synthesis of peptide nucleic acid (PNA) oligomers containing up to 40 nucleobases (representing 10 consecutive coupling reactions) on a 5'-NH<sub>2</sub>-terminated DNA-hairpin template employing reductive aminations of complementary  $\alpha$ -aldehyde  $\omega$ -amine terminated PNA-tetranucleotides (Scheme 1.3, A).<sup>87</sup> Crucial for the success of this approach is the use of the high-yielding reductive amination strategy that guarantees the selective, i.e. distance dependent, coupling between neighbouring monomeric units assembled on the substrate. Additionally, the PNA-tetramers show strong and selective binding to their respective anticodon on the DNA-hairpin template. In their initial study the authors explored the synthesis of homopolymers of different length. Quite impressively, when mismatched codons only different in one nucleotide (e.g. ATTC instead of AGTC) were incorporated into the sequence of the template at a specific position, only products with a length up to the mismatched codon were produced. In a later study Liu and co-workers showed the possibility to employ side-chain-functionalised PNA tetra- and pentameric aldehydes<sup>89</sup> and, even more remarkably, the synthesis of sequence-defined hexamers in the presence of 12 PNA-monomers from a DNA-hairpin template containing 6 codons (Scheme 1.3, B).<sup>90</sup> With the presented systems, Liu and co-workers successfully performed six iterated cycles of transcription, selection and amplification (a biotinylated, streptavidin-

binding PNA segment was incorporated) and observed a >1,000,000-fold overall enrichment of the parental DNA strand, displaying the possibility of PNA evolution *in vitro*.



**Scheme 1.3.** Synthesis of PNA oligomers assembled on a DNA-template by codon-anticodon recognition and subsequent connection by a reductive amination strategy introduced by Liu.<sup>88</sup> (A) Polymerisation of a PNA-40mer by ten consecutive coupling reactions between oligonucleotide tetramers. (B) Sequence-specific synthesis of a PNA-30mer from six different PNA-pentanucleotides; matching colours represent complementary strands, DNA oligomers are depicted as rectangles, PNA-oligomers are depicted as rectangles with rounded edges.

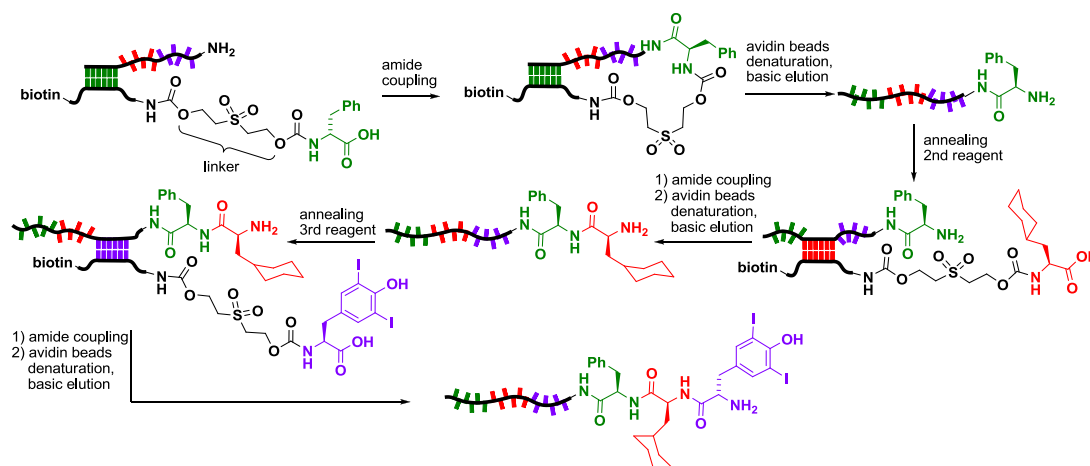
### 1.3.5 Sequence-Specific Translation with DNA-Templated Synthesis

The possibility to control the effective molarity of specific reactants mediating chemical reactions that would lead to considerable cross-reactivity under standard solution phase conditions by transcribing information from a DNA strand has led to the realisation of DNA templated multi-step sequence-specific syntheses.

In their groundbreaking work, Liu and co-workers demonstrated the synthesis of a tripeptide based on information encoded in a DNA-template (Scheme 1.4).<sup>91</sup> In the first step, a biotinylated 10-base oligonucleotide strand loaded with an amino acid ('reagent' strand) is annealed to a specific region of a 30-base DNA-oligomer ('template' strand), which features a terminal amino group, allowing the formation of a peptide bond after treatment with a coupling reagent. One problem the authors faced was the removal of the reagent strand after amide coupling. In order to allow subsequent extension of the peptide chain on the DNA-template strand, the authors used a cleavable, base-labile carbamoylethyl sulfone linker connecting the amino acids to the DNA-oligonucleotide adapter. Therefore, removal of the waste reagent-DNA strand was possible in the second step by capturing the template-reagent adduct with magnetic avidin beads and basic elution, traceless cleaving the linker and releasing the template strand loaded with the first amino acid in solution. Repeating the procedure of reagent annealing, peptide bond formation, avidin capturing and linker cleavage/elution for another two cycles allowed the isolation of the desired sequence-specific tripeptide.

Based on this work, Liu and co-workers demonstrated the translation of a library of template strands into the corresponding sequence-specific synthetic small-molecule macrocycles incorporating various functionalised amino acids.<sup>92</sup> The 65-member library containing one known macrocycle-DNA conjugate possessing target affinity was subjected to *in vitro* selection and amplification and the positive control could be isolated and identified after one cycle. Similarly, the translation of DNA into synthetic *N*-acyloxazolidines has been reported by the same group.<sup>93</sup>

In all three examples, reagents have to be added sequentially and purification has to be performed after each monomer addition. Liu and co-workers showed that a one-pot oligomer synthesis can be realised by simultaneous hybridisation of three reactive monomer strands onto a DNA-template featuring three complementary binding sites.<sup>94</sup> Controlling the secondary structure of the template strand with temperature variations, the effective molarity of the annealed reagents could be fine-tuned and a sequence-ordered triene was obtained from three successive Wittig olefinations.



**Scheme 1.4.** DNA-templated multistep synthesis of a tripeptide reported by Liu.<sup>95</sup> Sequence-control is achieved by sequential addition of oligonucleotide tethered *N*-protected amino acids, which are held in place by the complementary sequence on the amine-functionalised template strand. Amide coupling, avidin capturing, and basic elution allow the transfer of the monomer to the template and the removal of the spent oligonucleotide tether; matching colours indicate complementary sequences between strands; lines indicating base pairing do not represent a particular number of bases.

Instead of using linear templates, Gothelf, Hansen and co-workers introduced multibranch DNA-junctions for the three-step synthesis of a sequence-defined pentapeptide with the aim to overcome limitations of linear templates, e.g. the intrinsic disadvantage of decreasing coupling efficiency with increasing distance between reactive centre on the template and the monomer attached to the reagent strand.<sup>96</sup> The reactive monomeric unit is attached close to the three-way junction, thereby allowing the control of the effective molarity during the assembly of the nanoscale reactor, enabling the reaction of reactants in a confined space. Furthermore, the system allows target affinity screening by repeated rounds of selection, amplification, and translation.

Another approach towards DNA-templated multistep synthesis that overcomes the disadvantages of a simple linear template exploits “toehold” displacement.<sup>97</sup> In an asymmetric DNA duplex  $A^{\backslash}B^{\backslash}:B$ , where  $B$  is a strand complementary but shorter than  $A^{\backslash}B^{\backslash}$  and  $A^{\backslash}$  a “toehold”, i.e. a short sequence of unpaired nucleotides,  $B$  can be replaced by an incoming strand  $AB$  that hybridises with  $A^{\backslash}B^{\backslash}$ , after initial recognition of  $A:A^{\backslash}$ .

Turberfield and co-workers were the first group to demonstrate a strand displacement mechanism for the sequence-specific multistep synthesis of an olefin by consecutive Wittig-reactions.<sup>98</sup> The reaction is based on the hybridisation of the DNA strand bound to the growing oligomer with a reagent strand. After group transfer to the reagent strand the spent DNA strand can be displaced by addition of a fully complementary remover strand.

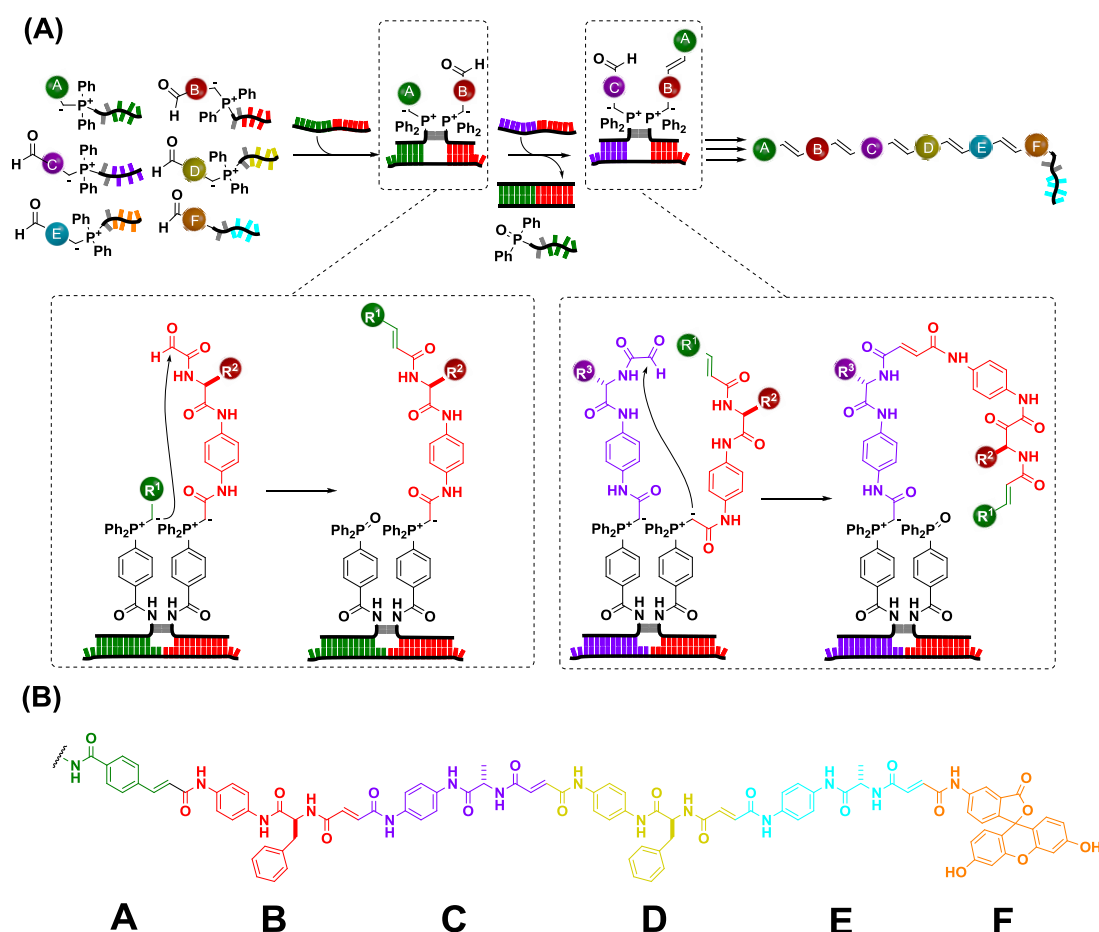
This cycle of reagent addition and strand removal can be continued – in theory- infinitely, making it the first example of multistep DTS where the length of the target oligomer is only limited by the intrinsic yield of the coupling reaction, but not by the length and information content of the template strand. Building on these initial results, Turberfield and co-workers demonstrated the synthesis of a sequence-defined 10-mer, showing the possibility to realise relatively long sequences with DTS.<sup>99</sup>

Using a similar approach, Liu developed the DNA-templated synthesis of a heptamer, featuring six elongation reactions.<sup>97</sup> Instead of adding remover strands, a linear template bearing six different binding sites is employed and displacement of the spent adapters is achieved by addition of a reagent strand that substitutes the former strand through increasingly stronger interactions with the template. The reaction proceeded in 35% overall yield, corresponding to ~83% average yield per step, much higher than for other comparable DNA-templated multi-step syntheses.

Combining the use of branched DNA-junctions and the strand displacement strategy, Turberfield and co-workers recently reported the first example of a templated oligomer synthesis in which all reactive monomers are present in solution simultaneously (Scheme 1.5).

The adapters used for the synthesis consist of oligomers which feature reactive monomer groups at either the 5' or the 3' end. The DNA sequence consists of one individual adapter-specific (Scheme 1.5, coloured lines) and one of two mutually complementary universal domains (Scheme 1.5, grey lines). The universal domain enables the interaction with all complementary adapters and allows the reactive units to be aligned in an end-of-helix configuration during the transient T-way junction that is formed when two adapters interact with a template strand. The template strand itself features two toeholds at either end, allowing its displacement by addition of a fully complementary remover strand. The sequence of the oligomer can therefore be controlled by the alternating addition of a DNA-templated strand and subsequent addition of a remover strand. During each Wittig reaction the growing oligomer is transferred to the incoming adapter, making it possible to synthesise products not limited by the information content of a linear template. Additionally, the approach from Turberfield allows the parallel synthesis of different sections, allowing access to highly complex product libraries and short synthetic sequences.

Extending their DNA-PNA oligomer transcription approach, Liu and co-workers demonstrated in a remarkable example the possibility to translate the information encoded in the DNA-parental strand into structurally unrelated synthetic polymers.



**Scheme 1.5.** Mechanism for the DNA-templated sequence-controlled oligomer synthesis reported by Turberfield.<sup>100</sup> (A) Reactive monomers A to F are tethered to nucleotide sequences featuring individual (coloured) and general sub-sequences (grey) that allow interaction with the complementary template strand (coloured) and complementary adapters (grey), respectively. Addition of a template strand brings two adapters together and enables the Wittig reaction and transfer of the growing oligomer strand. Sequential addition of template and removal strands allows sequence-control of the growing oligomer; matching colours indicate complementary sequences between strands; lines indicating base pairing do not represent a particular number of bases. (B) Hexameric reaction product successfully isolated by Turberfield and co-workers (tethered oligonucleotide not shown).

Therefore, the authors employed PNA-pentamers to which oligomer fragments such as polyethylene glycol,  $\alpha$ -(D)-peptides or  $\beta$ -peptides were connected *via* two disulfide linkages. The macrocyclic tRNA analogues functionalised with either two terminal azides or two terminal alkynes were then assembled on the DNA template and coupled using the CuAAC reaction in an AA-BB manner. Cleaving the disulfide bonds with the reducing agent dithiothreitol liberated the synthetic polymer products. In their most impressive example, the authors reported a polymer with a molecular weight of 26 kDa arising from 16 consecutive substrate couplings and containing 90  $\beta$ -amino acid building blocks.

### 1.3.6 Artificial Molecular Machines for Sequence-Specific Synthesis

All highly sequence-controlled polymerisations in nature are realised with the help of molecular machines. Designing synthetic molecular machines for the sequence-defined assembly of monomeric units therefore holds the promise to eventually achieve sequence control similar to that of DNA-polymerases and ribosomes.

The dynamic behaviour of molecular devices capable of synthesising sequence-specific polymers based on the movement along a template strand, e.g. the ribosome or DNA-polymerase, can be described by fundamental characteristics exhibited by the system similar to that of molecular walkers<sup>101</sup>:

- (i) Processivity: the ability of the machine to remain attached to the template during the operation.
- (ii) Directionality: the tendency of the device to preferentially or exclusively migrate in a given direction, i.e. towards the end of a molecular template.
- (iii) Iterative operation: the capability of the molecular machine to be reset after a complete cycle of operation, e.g. to repeat the synthesis of a given oligomer using the same template.
- (iv) Repetitive operation: the ability of the molecular machine to be reset at the end of each synthetic step without undoing the task i.e. bond formation that was originally performed.
- (v) Autonomous operation: the ability of the molecular motor to continually function as long as an energy input is present, i.e. no external intervention such as the application of a sequence of stimuli is required.

Those characteristics are typically fulfilled for nature's machinery and are generally desirable, but not necessarily required for the successful realisation of a multistep synthesis aided by an artificial molecular machine. A subset of characteristics, e.g. a machine that processively moves along a template destroying the information encoded in the strand with each monomer addition shows processivity, repetitive operation and directionality through the burnt bridges mechanism<sup>101</sup> it follows but is not capable of iterative operation. Also it does not necessarily have to be autonomous. Similarly, other molecular machines exhibiting a subset of characteristics can be successful in the synthesis of sequence-specific polymers. Three molecular machines which feature some of the aforementioned characteristics will be presented in detail in the following section.

### 1.3.7 DNA-Based Molecular Walkers for Sequence Control

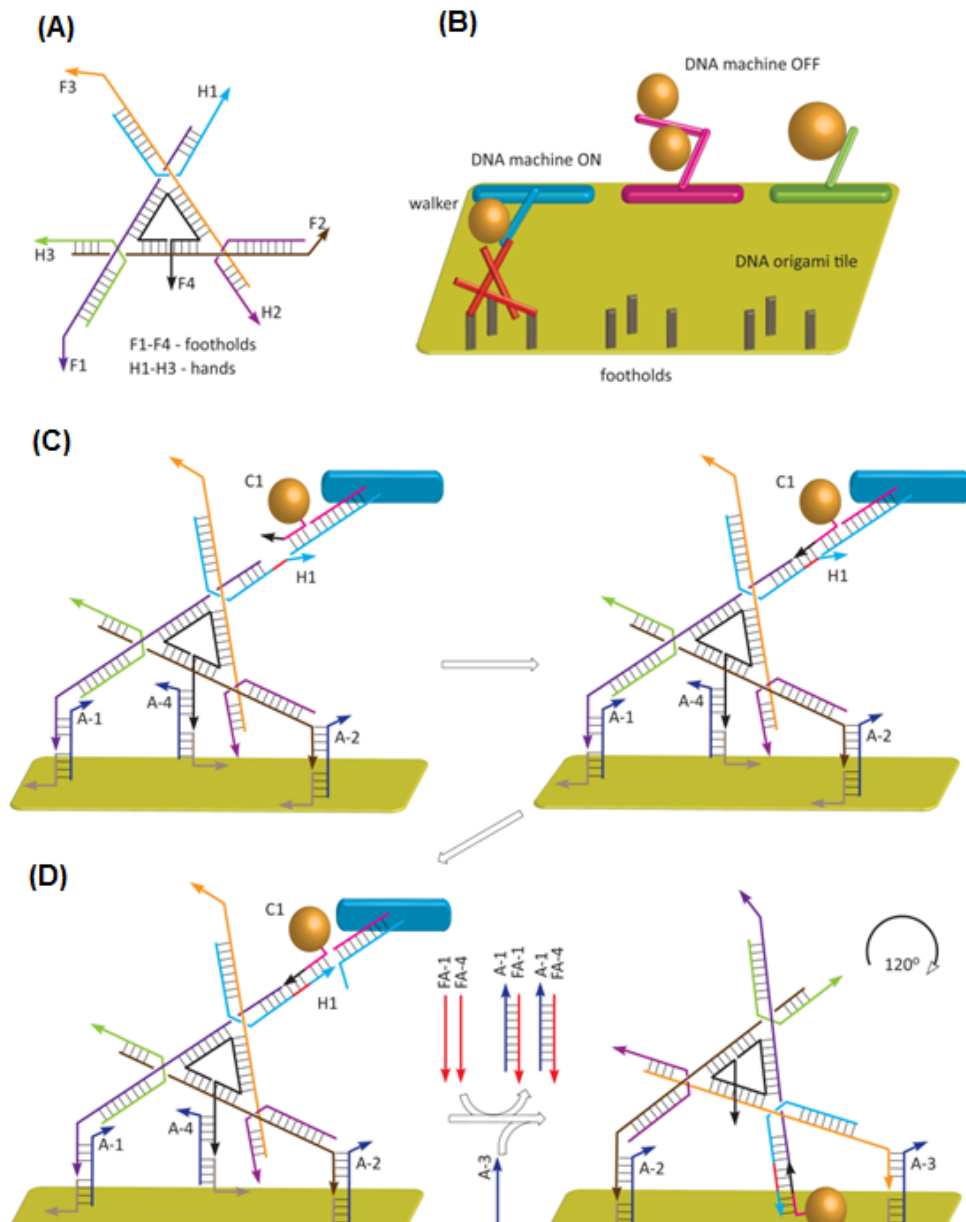
Inspired by motor proteins like kinesin, dynein and certain myosin motor proteins that transport cargoes within the cell by “walking” along polymeric filaments,<sup>101-102</sup> several artificial molecular transport systems based on DNA<sup>103</sup> or small molecules<sup>104</sup> have been successfully synthesised and operated.

Generally, a molecular walker consists of multiple binding points (‘feet’) that can associate with corresponding ‘footholds’ on the track. Movement along a track is achieved by sequential detachment and reattachment of feet that—if biased—result in directionality.

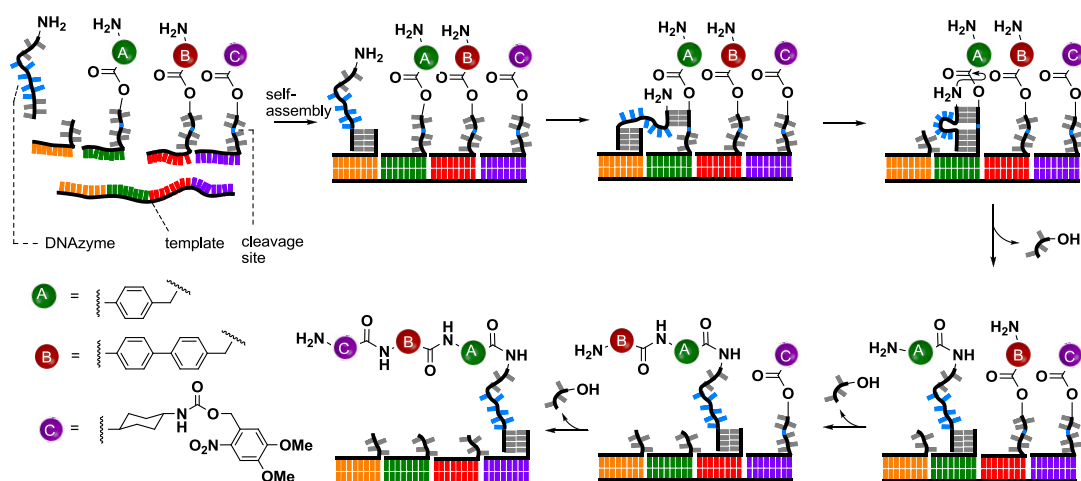
The first example of a sequence-specific synthesis performed by a molecular machine was reported by Seeman and co-workers (Figure 1.4).<sup>103i</sup> In their seminal work, the authors employed a sequentially fuelled DNA-based molecular walker capable of picking up differently sized gold nanoparticles along its path (Figure 1.4, A). The triangular walker moiety itself is constructed from seven oligonucleotide strands that form a tensegrity architecture (Figure 1.4, B). The walker features seven single stranded DNA-sequences: four feet, responsible for the directional motion, and three hands, capable of binding to the cargo. As track a large (ca. 300 nm wide) two-dimensional DNA-origami tile assembled from a total of 202 DNA oligonucleotides and featuring 18 single-stranded protruding footholds was used. Additionally, three cassettes containing three independently controlled two-state DNA machines donating the cargo to the walker in its “on” state and holding it back in the “off” state, were attached to the origami framework.

The walker unit performs a 120° rotation during each step (Figure 1.4, C). Movement is achieved upon addition of complementary fuel strands that trigger branch-migration of the targeted feet by toehold-binding and substitution. The fourth foot is only used for ensuring the proper orientation of the triangle towards the cargo-donating devices and is therefore only bound to the track when cargo is to be transferred. Two steps are needed to move the walker from one cargo-donating station to the next. Reaching a station in its “on” state, the cargo can be transferred to the hand of the walker *via* strand displacement; the decoding strand attached to the cargo features a toehold therefore allowing preferential binding to the hand of the walking moiety (Figure 1.4, D). During its linear trajectory, the DNA-walker encounters three delivery devices, either in an “on” or “off” state, resulting in 8 distinguishable cargo combinations.

The comparatively large size of the system allowed the analysis of product formation by atomic force microscopy (AFM) and scanning electron microscopy (SEM) which was also used for the statistical analysis of the yield of the process.



**Figure 1.4.** Nanoscale assembly line published by Seeman and co-workers.<sup>105</sup> (A) Schematic illustration of the walker moiety featuring seven single-stranded domains: four feet (F1 to F4) and three hands (H1 to H3). (B) Schematic illustration of the cargo transfer from the DNA-adapter to the walker moiety. (C) Handover of DNA-bound cargo (C1) from the DNA machine to the walker. (D) One 'stride' of the walker moiety, requiring the sequential addition of two fuel strands and one anchor strand (walker rotated by 120°). Figure adapted with permission from Urszula Lewandowska.



**Scheme 1.6.** Sequence-defined synthesis of a tripeptide performed by an autonomous molecular walker capable of performing three consecutive intramolecular acylation reactions.<sup>106</sup> The light-blue lines indicate the DNAzyme sequence incorporated in the walker unit, which catalyses the hydrolysis of the ribonucleotide linkage in the three adapters. The amino acids are connected to their oligonucleotide strands *via* reactive succinimidyl ester linkages (not shown). Matching colours indicate complementary sequences between strands; lines indicating base pairing do not represent a particular number of bases.

The yield depends directly on the number of cargo additions to the walker, incorporation of nanoparticles from the “off”-state was practically not detected. For the triple addition, a yield of 43% was observed: programmed single products can be obtained in 90% yield with negligible formation of incorrect products.

A second example of a sequence-specific synthesis aided by a DNA-based molecular walker was reported shortly after Seeman’s work by Liu and co-workers (Scheme 1.6).<sup>103j</sup> The sequence-specific synthesis of a tripeptide was successfully realised with an autonomous DNA walker that performs a series of peptide bond formations through its station to station movement along its track. The translocation mechanism is similar to a previous reported walker by Mao and co-workers.<sup>103c,d</sup> Liu’s device consists of a DNA-track featuring four distinct binding sites that can coordinate to three corresponding footholds (label 2,3,4 in Scheme 1.6) functionalised with activated esters of three different non-natural amino acids and a fourth foothold (1) called the initiator that hybridises with the walker unit during the self-assembly of the machine, resulting in the starting state A depicted in Scheme 1.5. The use of rigid spacers between the *N*-hydroxysuccinimidyl (NHS) ester and the amine functionality in the amino acids was necessary in order to prevent intramolecular cyclisation. The single-stranded part of the walker can now hybridise with toehold 2 and—*via* a competitive strand displacement—move to the second station. Resulting from the high

effective molarity of the amine-functionalised walker and the activated ester an amine transfer is triggered. In order to allow translocation to the third foothold, a cleavage of the ribonucleotide linkage built in foothold 2 occurs triggered by the DNAzyme moiety present in the walker. After dissociation of the short DNA fragment, the now single stranded DNA-part of the walker can hybridise with foothold 3, repeating the series of events and elongating the growing chain by one monomeric unit. After a final cycle of walker migration, acylation and DNAzyme catalysed foothold cleavage, the walker remains attached to foothold 4 together with the assembled tripeptide on its N-terminus.

The reaction mixture was analysed by high-resolution mass spectrometry and by comparison with mass spectra of authentic samples with known composition. The estimated yield of the targeted product was 45%, accompanied by walker-monoamide and walker-diamide side-products. If RNA cleavage of the footholds occurs before the acyl transfer takes place, truncated products result. Therefore, in an ideal system the rate of peptide bond formation should be significantly higher than the DNAzyme catalysed cleavage of the footholds. In order to assess the sequence-specificity of the synthesised peptide, Liu and co-workers introduced an *N*-protected amino acyl residue in the position of foothold 2. As a result, only monoamide product was detected, suggesting the system works processively and the products are synthesised with sequence-specificity.

Liu showed, in contrast to Seeman's triangular walker, that it is possible to design systems that operate not only directionally and processively, but also autonomously although lacking the ability of repetitive operation.

## 1.4 Conclusion

Although all examples using enzyme-free sequence-specific reactions with the various methods demonstrated to date have shown various possibilities for achieving sequence-control in synthesis, the ability of nature to generate peptides with a length of over 30,000 monomeric units<sup>107</sup> in a defined order remains unmatched. Albeit considerable progress towards the realisation of sequence-specific synthesis has been made, there is still a long way to go until the capability of nature to generate information-rich monodisperse polymers can be achieved with the tools of synthetic chemistry. With sequential synthesis on solid phase, polymerisations that exploit reactivity differences of monomers and (mainly DNA-)templated polymerisation capable of assembling monomers in order, it remains to be seen which technique will allow the most versatile and practical control of sequence-specificity. From the author's point of view, the use of complex artificial molecular machines – associated with new designs – offers possibilities that might enable the total synthesis of proteins or polymers with unprecedented micro- and macromolecular properties in the future. The design, synthesis and operation of such novel molecular machines will be discussed in the following chapter.

## 1.5 References

- (1) Brudno, Y.; Liu, D. R. *Chem. Biol.* **2009**, *16*, 265.
- (2) (a) Voet, D.; Voet, J. G. *Biochemistry*; fourth edition; John Wiley & Sons: New York, 2011 (b) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; sixth edition; W. H. Freeman: New York, 2006.
- (3) Rehm, B. H. A. *Nat. Rev. Micro.* **2010**, *8*, 578.
- (4) (a) Whitfield, C. *Annu. Rev. Biochem.* **2006**, *75*, 39 (b) Vollmer, W.; Blanot, D.; De Pedro, M. A. *FEMS Microbiol. Rev.* **2008**, *32*, 149 (c) Becker, A.; Katzen, F.; Pühler, A.; Ielpi, L. *Appl. Microbiol. Biotechnol.* **1998**, *50*, 145 (d) Jahn, M.; Withers, S. G. *Biocatal. Biotransform.* **2003**, *21*, 159.
- (5) (a) Badi, N.; Chan-Seng, D.; Lutz, J. F. *Macromol. Chem. Phys.* **2013**, *214*, 135 (b) Ouchi, M.; Badi, N.; Lutz, J.-F.; Sawamoto, M. *Nat. Chem.* **2011**, *3*, 917 (c) Lutz, J.-F. *Polym. Chem.* **2010**, *1*, 55 (d) Badi, N.; Lutz, J.-F. *Chem. Soc. Rev.* **2009**, *38*, 3383.
- (6) Hiemenz, P. C.; Lodge, T. P. *Polymer Chemistry*; second edition; CRC Press: Boca Raton, 2007.
- (7) Lutz, J. F.; Neugebauer, D.; Matyjaszewski, K. *J. Am. Chem. Soc.* **2003**, *125*, 6986.
- (8) Pasparakis, G.; Krasnogor, N.; Cronin, L.; Davis, B. G.; Alexander, C. *Chem. Soc. Rev.* **2010**, *39*, 286.
- (9) Bashan, A.; Yonath, A. In *Ribosomes*; Rodnina, M., Wintermeyer, W., Green, R., Eds.; Springer Vienna, 2011.
- (10) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
- (11) (a) Yonath, A. *Angew. Chem. Int. Ed.* **2010**, *49*, 4340 (b) Yonath, A. *Angew. Chem.* **2010**, *122*, 4438 (c) Ramakrishnan, V. *Angew. Chem. Int. Ed.* **2010**, *49*, 4355 (d) Ramakrishnan, V. *Angew. Chem.* **2010**, *122*, 4454 (e) Steitz, T. A. *Angew. Chem. Int. Ed.* **2010**, *49*, 4381 (f) Steitz, T. A. *Angew. Chem.* **2010**, *122*, 4482.
- (12) (a) Ramakrishnan, V. *Cell* **2002**, *108*, 557 (b) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242.
- (13) The unit S that characterises the subunits is called Svendberg and is characteristic for the rate of sedimentation of the respective particle.
- (14) (a) Pain, V. M. *Eur. J. Biochem.* **1996**, *236*, 747 (b) Gualerzi, C. O.; Pon, C. L. *Biochemistry* **1990**, *29*, 5881.
- (15) Nissen, P.; Kjeldgaard, M.; Thirup, S.; Polekhina, G.; Reshetnikova, L.; Clark, B. F. C.; Nyborg, J. *Science* **1995**, *270*, 1464.

- (16) Leung, E. K. Y.; Suslov, N.; Tuttle, N.; Sengupta, R.; Piccirilli, J. A. *Annu. Rev. Biochem.* **2011**, *80*, 527
- (17) (a) Wilden, B.; Savelsbergh, A.; Rodnina, M. V.; Wintermeyer, W. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13670 (b) Rodnina, M. V.; Savelsbergh, A.; Katunin, V. I.; Wintermeyer, W. *Nature* **1997**, *385*, 37 (c) Seo, H.-S.; Abedin, S.; Kamp, D.; Wilson, D. N.; Nierhaus, K. H.; Cooperman, B. S. *Biochemistry* **2006**, *45*, 2504.
- (18) Agrawal, R. K.; Sharma, M. R.; Kiel, M. C.; Hirokawa, G.; Booth, T. M.; Spahn, C. M. T.; Grassucci, R. A.; Kaji, A.; Frank, J. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8900.
- (19) (a) Zaher, H. S.; Green, R. *Cell* **2009**, *136*, 746 (b) Kramer, E. B.; Farabaugh, P. J. *RNA* **2007**, *13*, 87 (c) Rodnina, M. V.; Wintermeyer, W. *Annu. Rev. Biochem.* **2001**, *70*, 415.
- (20) Ogle, J. M.; Ramakrishnan, V. *Annu. Rev. Biochem.* **2005**, *74*, 129.
- (21) (a) Hopfield, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4135 (b) Rodnina, M. V.; Wintermeyer, W. *Trends Biochem. Sci.* **2001**, *26*, 124 (c) Thompson, R. C.; Karim, A. M. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 4922.
- (22) (a) Blanchard, S. C.; Gonzalez, R. L.; Kim, H. D.; Chu, S.; Puglisi, J. D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1008 (b) Lee, T.-H.; Blanchard, S. C.; Kim, H. D.; Puglisi, J. D.; Chu, S. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 13661.
- (23) (a) Rodnina, M. V.; Gromadski, K. B.; Kothe, U.; Wieden, H.-J. *FEBS Lett.* **2005**, *579*, 938 (b) Koshland, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98.
- (24) Pape, T.; Wintermeyer, W.; Rodnina, M. *EMBO J.* **1999**, *18*, 3800.
- (25) Ogle, J. M.; Murphy Iv, F. V.; Tarry, M. J.; Ramakrishnan, V. *Cell* **2002**, *111*, 721.
- (26) (a) Jenner, L.; Demeshkina, N.; Yusupova, G.; Yusupov, M. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1072 (b) Ninio, J. *Biochimie* **1975**, *57*, 587.
- (27) (a) Freist, W.; Pardowitz, I.; Cramer, F. *Biochemistry* **1985**, *24*, 7014 (b) Baldwin, A. N.; Berg, P. *J. Biol. Chem.* **1966**, *241*, 839 (c) Ferscht, A. *Structure and Mechanism in Protein Science*; third edition; W.H.Freeman: New York, 1999.
- (28) Nureki, O.; Vassylyev, D. G.; Tateno, M.; Shimada, A.; Nakama, T.; Fukai, S.; Konno, M.; Hendrickson, T. L.; Schimmel, P.; Yokoyama, S. *Science* **1998**, *280*, 578.
- (29) Ibba, M.; Söll, Dieter *Science* **1999**, *286*, 1893.

- (30) (a) Frank, J.; Gonzalez, R. L. *Annu. Rev. Biochem.* **2010**, *79*, 381 (b) Shoji, S.; Walker, S. E.; Fredrick, K. *Mol. Cell* **2006**, *24*, 931 (c) Konevega, A. L.; Fischer, N.; Semenkov, Y. P.; Stark, H.; Wintermeyer, W.; Rodnina, M. V. *Nat. Struct. Mol. Biol.* **2007**, *14*, 318.
- (31) Shoji, S.; Walker, S. E.; Fredrick, K. *ACS Chem. Biol.* **2009**, *4*, 93.
- (32) (a) Fei, J.; Kosuri, P.; MacDougall, D. D.; Gonzalez Jr, R. L. *Mol. Cell* **2008**, *30*, 348 (b) Frank, J.; Gao, H.; Sengupta, J.; Gao, N.; Taylor, D. J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19671.
- (33) Katunin, V. I.; Savelsbergh, A.; Rodnina, M. V.; Wintermeyer, W. *Biochemistry* **2002**, *41*, 12806.
- (34) Moore, P. B. *Annu. Rev. of Biophys.* **2012**, *41*; 1.
- (35) (a) Keefe, A. D.; Cload, S. T. *Curr. Opin. Chem. Biol.* **2008**, *12*, 448 (b) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225 (c) Pinheiro, V. B.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245 (d) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360.
- (36) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.
- (37) Merrifield, R. B. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 799.
- (38) (a) Fields, G. B.; Noble, R. L. *Int. J. Pept. Protein Res.* **1990**, *35*, 161 (b) Amblard, M.; Fehrentz, J. A.; Martinez, J.; Subra, G. *Mol. Biotechnol.* **2006**, *33*, 239.
- (39) (a) Hermkens, P. H. H.; Ottenheijm, H. C. J.; Rees, D. *Tetrahedron* **1996**, *52*, 4527 (b) Dax, S. L.; McNally, J. J.; Youngman, M. A. *Curr. Med. Chem.* **1999**, *6*, 255 (c) Lonnberg, H. *Bioconjugate Chem.* **2009**, *20*, 1065 (d) Kingsbury, C. L.; Mehrman, S. J.; Takacs, J. M. *Curr. Org. Chem.* **1999**, *3*, 497 (e) Boas, U.; Brask, J.; Jensen, K. *J. Chem. Rev.* **2009**, *109*, 2092 (f) Delgado, M.; Janda, K. D. *Curr. Org. Chem.* **2002**, *6*, 1031 (g) Banfi, L.; Guanti, G.; Riva, R.; Basso, A. *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 704 (h) Nandy, J. P.; Prakesch, M.; Khadem, S.; Reddy, P. T.; Sharma, U.; Arya, P. *Chem. Rev.* **2009**, *109*, 1999 (i) Dörwald, F. Z. *Organic Synthesis on Solid Phase: Supports, Linkers, Reactions*; second edition.; Wiley-VHC: Weinheim, 2002.
- (40) Merrifield, B. *Methods Enzymol.* **1997**, *289*, 3.
- (41) (a) Hirschmann, R.; Nutt, R. F.; Veber, D. F.; Vitali, R. A.; Varga, S. L.; Jacob, T. A.; Holly, F. W.; Denkewalter, R. G. *J. Am. Chem. Soc.* **1969**, *91*, 507 (b) Gutte, B.; Merrifield, R. B. *J. Am. Chem. Soc.* **1969**, *91*, 501.
- (42) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.

- (43) (a) Nelson, J. C.; Young, J. K.; Moore, J. S. *J. Org. Chem.* **1996**, *61*, 8160 (b) Young, J. K.; Nelson, J. C.; Moore, J. S. *J. Am. Chem. Soc.* **1994**, *116*, 10841.
- (44) (a) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992**, *48*, 2223 (b) Caruthers, M. H. *Science* **1985**, *230*, 281.
- (45) (a) Plante, O. J.; Palmacci, E. R.; Seeberger, P. H. *Science* **2001**, *291*, 1523 (b) Seeberger, P. H. *Chem. Soc. Rev.* **2008**, *37*, 19.
- (46) (a) Murray, J. K.; Gellman, S. H. *Org. Lett.* **2005**, *7*, 1517 (b) Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. *Chem. Rev.* **2001**, *101*, 3219.
- (47) Bhaskara Reddy, K. M.; Mallikarjunasarma, D.; Bulliraju, K.; Sreelatha, V.; Bharathi Kumari, Y.; Dandala, R.; Ananda, K. *Int. J. Pept. Res. Ther.* **2011**, *17*, 113.
- (48) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. *J. Am. Chem. Soc.* **1992**, *114*, 10646.
- (49) (a) Cohen-Anisfeld, S. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1993**, *115*, 10531 (b) Bejugam, M.; Flitsch, S. L. *Org. Lett.* **2004**, *6*, 4001.
- (50) (a) Christensen, L.; Fitzpatrick, R.; Gildea, B.; Petersen, K. H.; Hansen, H. F.; Koch, T.; Egholm, M.; Buchardt, O.; Nielsen, P. E.; Coull, J.; Berg, R. H. *J. Pept. Sci.* **1995**, *1*, 175 (b) Debaene, F.; Winssinger, N. *Org. Lett.* **2003**, *5*, 4445.
- (51) (a) de Jong, R.; Rijkers, D. T. S.; Liskamp, R. M. J. *Helv. Chim. Acta* **2002**, *85*, 4230 (b) Humljan, J.; Kotnik, M.; Boniface, A.; Šolmajer, T.; Urleb, U.; Blanot, D.; Gobec, S. *Tetrahedron* **2006**, *62*, 10980.
- (52) (a) Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201 (b) König, H. M.; Gorelik, T.; Kolb, U.; Kilbinger, A. F. M. *J. Am. Chem. Soc.* **2006**, *129*, 704 (c) Cormack, P. A. G.; Moore, B. D.; Sherrington, D. C. *Chem. Commun.* **1996**, *32*, 353.
- (53) Guichard, G.; Huc, I. *Chem. Commun.* **2011**, *47*, 5933.
- (54) (a) Douat-Casassus, C.; Pulka, K.; Claudon, P.; Guichard, G. *Org. Lett.* **2012**, *14*, 3130 (b) Violette, A.; Averlant-Petit, M. C.; Semetey, V.; Hemmerlin, C.; Casimir, R.; Graff, R.; Marraud, M.; Briand, J.-P.; Rognan, D.; Guichard, G. *J. Am. Chem. Soc.* **2005**, *127*, 2156 (c) Burgess, K.; Ibarzo, J.; Linthicum, D. S.; Russell, D. H.; Shin, H.; Shitangkoon, A.; Totani, R.; Zhang, A. J. *J. Am. Chem. Soc.* **1997**, *119*, 1556 (d) Kim, J.-M.; Bi, Y.; Paikoff, S. J.; Schultz, P. G. *Tetrahedron Lett.* **1996**, *37*, 5305 (e) Boeijen, A.; van Ameijde, J.; Liskamp, R. M. J. *J. Org. Chem.* **2001**, *66*, 8454.
- (55) Cho, C.; Moran, E.; Cherry; Stephans, J.; Fodor, S.; Adams, C.; Sundaram, A.; Jacobs, J.; Schultz, P. *Science* **1993**, *261*, 1303.

- (56) Young, J. K.; Nelson, J. C.; Moore, J. S. *J. Am. Chem. Soc.* **1994**, *116*, 10841.
- (57) R. L. Malenfant, P.; M. J. Frechet, J. *Chem. Commun.* **1998**, *34*, 2657.
- (58) Turner, D. J.; Anemian, R.; Mackie, P. R.; Cupertino, D. C.; Yeates, S. G.; Turner, M. L.; Spivey, A. C. *Org. Biomol. Chem.* **2007**, *5*, 1752.
- (59) Semetey, V.; Moustakas, D.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **2006**, *45*, 588.
- (60) (a) Fyles, T. M.; Hu, C.-W.; Luong, H. J. *Org. Chem.* **2006**, *71*, 8545 (b) Kuisle, O.; Quiñoá, E.; Riguera, R. *J. Org. Chem.* **1999**, *64*, 8063.
- (61) Martos, V.; Castreño, P.; Royo, M.; Albericio, F.; Mendoza, J. *J. Comb. Chem.* **2009**, *11*, 410.
- (62) Deng, X.-X.; Li, L.; Li, Z.-L.; Lv, A.; Du, F.-S.; Li, Z.-C. *ACS Macro Lett.* **2012**, *1*, 1300.
- (63) Lutz, J.-F.; Schmidt, B. V. K. J.; Pfeifer, S. *Macromol. Rapid Commun.* **2011**, *32*, 127.
- (64) Lo, P. K.; Sleiman, H. F. *J. Am. Chem. Soc.* **2009**, *131*, 4182.
- (65) Ueda, M. *Prog. Polym. Sci.* **1999**, *24*, 699.
- (66) Deng, X.-X.; Li, L.; Li, Z.-L.; Lv, A.; Du, F.-S.; Li, Z.-C. *ACS Macro Lett.* **2012**, *1*, 1300.
- (67) Passerini, M.; Simone, L. *Gazz. Chim. Ital.* **1921**, *51*, 126.
- (68) Allmendinger, M.; Molnar, F.; Zintl, M.; Luinstra, G. A.; Preishuber-Pflügl, P.; Rieger, B. *Chem. – Eur. J.* **2005**, *11*, 5327.
- (69) Kramer, J. W.; Treitler, D. S.; Dunn, E. W.; Castro, P. M.; Roisnel, T.; Thomas, C. M.; Coates, G. W. *J. Am. Chem. Soc.* **2009**, *131*, 16042.
- (70) (a) Benoit, D.; Hawker, C. J.; Huang, E. E.; Lin, Z.; Russell, T. P. *Macromolecules* **2000**, *33*, 1505 (b) Kirci, B.; Lutz, J.-F.; Matyjaszewski, K. *Macromolecules* **2002**, *35*, 2448.
- (71) Satoh, K.; Matsuda, M.; Nagai, K.; Kamigaito, M. *J. Am. Chem. Soc.* **2010**, *132*, 10003.
- (72) (a) Zamfir, M.; Lutz, J.-F. *Nat. Commun.* **2012**, 1138 (b) Lutz, J.-F.; Schmidt, B. V. K. J.; Pfeifer, S. *Macromol. Rapid Commun.* **2011**, *32*, 127 (c) Pfeifer, S.; Lutz, J.-F. *Chem. – Eur. J.* **2008**, *14*, 10949 (d) Pfeifer, S.; Lutz, J.-F. *J. Am. Chem. Soc.* **2007**, *129*, 9542 (e) Kakuchi, R.; Zamfir, M.; Lutz, J.-F.; Theato, P. *Macromol. Rapid Commun.* **2012**, *33*, 54.
- (73) Hibi, Y.; Tokuoka, S.; Terashima, T.; Ouchi, M.; Sawamoto, M. *Polym. Chem.* **2011**, *2*, 341.

- (74) Hibi, Y.; Ouchi, M.; Sawamoto, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 7434.
- (75) (a) Ida, S.; Terashima, T.; Ouchi, M.; Sawamoto, M. *J. Am. Chem. Soc.* **2009**, *131*, 10808 (b) Ida, S.; Ouchi, M.; Sawamoto, M. *J. Am. Chem. Soc.* **2010**, *132*, 14748.
- (76) Lo, P. K.; Sleiman, H. F. *J. Am. Chem. Soc.* **2009**, *131*, 4182.
- (77) Lin, N.-T.; Lin, S.-Y.; Lee, S.-L.; Chen, C.-h.; Hsu, C.-H.; Hwang, L. P.; Xie, Z.-Y.; Chen, C.-H.; Huang, S.-L.; Luh, T.-Y. *Angew. Chem. Int. Ed.* **2007**, *46*, 4481.
- (78) Naylor, R.; Gilham, P. T. *Biochemistry* **1966**, *5*, 2722.
- (79) (a) Li, X.; Liu, D. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 4848 (b) Silverman, S. K. *Angew. Chem. Int. Ed.* **2010**, *49*, 7180 (c) Percivalle, C.; Bartolo, J.-F.; Ladame, S. *Org. Biomol. Chem.* **2013**, *11*, 16.
- (80) Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R. *Angew. Chem. Int. Ed.* **2002**, *41*, 4104.
- (81) Mann, S. *Angew. Chem. Int. Ed.* **2013**, *52*, 155.
- (82) Naylor, R.; Gilham, P. T. *Biochemistry* **1966**, *5*, 2722.
- (83) (a) Inoue, T.; Joyce, G. F.; Grzeskowiak, K.; Orgel, L. E.; Brown, J. M.; Reese, C. B. *J. Mol. Biol.* **1984**, *178*, 669 (b) Acevedo, O. L.; Orgel, L. E. *J. Mol. Biol.* **1987**, *197*, 187 (c) Orgel, L. E. *Acc. Chem. Res.* **1995**, *28*, 109 (d) Kozlov, I. A.; Orgel, L. E. *Mol. Biol.* **2000**, *34*, 781.
- (84) (a) von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 932 (b) Sievers, D.; von Kiedrowski, G. *Nature* **1994**, *369*, 221.
- (85) Rojas Stütz, J. A.; Richert, C. *J. Am. Chem. Soc.* **2001**, *123*, 12718.
- (86) (a) Kaiser, A.; Richert, C. *J. Org. Chem.* **2013** (b) Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *135*, 924 (c) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2 (d) Schrum, J. P.; Ricardo, A.; Krishnamurthy, M.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 14560.
- (87) Rosenbaum, D. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 13924.
- (88) Kleiner, R. E.; Brudno, Y.; Birnbaum, M. E.; Liu, D. R. *J. Am. Chem. Soc.* **2008**, *130*, 4646.
- (89) Kleiner, R. E.; Brudno, Y.; Birnbaum, M. E.; Liu, D. R. *J. Am. Chem. Soc.* **2008**, *130*, 4646.
- (90) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. *Nat. Chem. Biol.* **2010**, *6*, 148.
- (91) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304.
- (92) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601.

- (93) Li, X.; Gartner, Z. J.; Tse, B. N.; Liu, D. R. *J. Am. Chem. Soc.* **2004**, *126*, 5090.
- (94) Snyder, T. M.; Liu, D. R. *Angew. Chem. Int. Ed.* **2005**, *44*, 7379.
- (95) Snyder, T. M.; Liu, D. R. *Angew. Chem. Int. Ed.* **2005**, *44*, 7379.
- (96) Hansen, M. H.; Blakskjær, P.; Petersen, L. K.; Hansen, T. H.; Højfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. V. *J. Am. Chem. Soc.* **2009**, *131*, 1322.
- (97) He, Y.; Liu, D. R. *J. Am. Chem. Soc.* **2011**, *133*, 9972.
- (98) McKee, M. L.; Milnes, P. J.; Bath, J.; Stulz, E.; Turberfield, A. J.; O'Reilly, R. K. *Angew. Chem. Int. Ed.* **2010**, *49*, 7948.
- (99) Milnes, P. J.; McKee, M. L.; Bath, J.; Song, L.; Stulz, E.; Turberfield, A. J.; O'Reilly, R. K. *Chem. Commun.* **2012**, *48*, 5614.
- (100) McKee, M. L.; Milnes, P. J.; Bath, J.; Stulz, E.; O'Reilly, R. K.; Turberfield, A. J. *J. Am. Chem. Soc.* **2012**, *134*, 1446.
- (101) von Delius, M.; Leigh, D. A. *Chem. Soc. Rev.* **2011**, *40*, 3656.
- (102) (a) *Molecular Motors*; Wiley-VHC: Weinheim, 2003 (b) Vale, R. D. *Cell* **2003**, *112*, 467.
- (103) (a) Bath, J.; Green, S. J.; Turberfield, A. J. *Angew. Chem.* **2005**, *117*, 4432 (b) Bath, J.; Green, S. J.; Turberfield, A. J. *Angew. Chem. Int. Ed.* **2005**, *44*, 4358 (c) Tian, Y.; He, Y.; Chen, Y.; Yin, P.; Mao, C. *Angew. Chem.* **2005**, *117*, 4429 (d) Tian, Y.; He, Y.; Chen, Y.; Yin, P.; Mao, C. *Angew. Chem. Int. Ed.* **2005**, *44*, 4355 (e) Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. *Nature* **2008**, *451*, 318 (f) Green, S. J.; Bath, J.; Turberfield, A. J. *Phys. Rev. Lett.* **2008**, *101*, 238101 (g) Omabegho, T.; Sha, R.; Seeman, N. C. *Science* **2009**, *324*, 67 (h) Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. *Nature* **2010**, *465*, 206 (i) Gu, H.; Chao, J.; Xiao, S.-J.; Seeman, N. C. *Nature* **2010**, *465*, 202 (j) He, Y.; Liu, D. R. *Nat. Nanotechnol.* **2010**, *5*, 778 (k) Wickham, S. F. J.; Endo, M.; Katsuda, Y.; Hidaka, K.; Bath, J.; Sugiyama, H.; Turberfield, A. J. *Nat. Nanotechnol.* **2011**, *6*, 166 (l) Muscat, R. A.; Bath, J.; Turberfield, A. J. *Nano Lett.* **2011**, *11*, 982 (m) You, M.; Huang, F.; Chen, Z.; Wang, R.-W.; Tan, W. *ACS Nano* **2012**, *6*, 7935 (n) Wickham, S. F. J.; Bath, J.; Katsuda, Y.; Endo, M.; Hidaka, K.; Sugiyama, H.; Turberfield, A. J. *Nat. Nanotechnol.* **2012**, *7*, 169.

- (104) (a) von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Nat. Chem.* **2010**, *2*, 96(b) Otto, S. *Nat. Chem.* **2010**, *2*, 75(c) von Delius, M.; Geertsema, E. M.; Leigh, D. A.; Tang, D.-T. D. *J. Am. Chem. Soc.* **2010**, *132*, 16134 (f) Barrell, M. J.; Campaña, A. G.; von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Angew. Chem. Int. Ed.* **2011**, *50*, 285 (e) Barrell, M. J.; Campaña, A. G.; von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Angew. Chem.* **2011**, *123*, 299 (f) Campaña, A. G.; Carlone, A.; Chen, K.; Dryden, D. T. F.; Leigh, D. A.; Lewandowska, U.; Mullen, K. M. *Angew. Chem. Int. Ed.* **2012**, *51*, 5480 (g) Campaña, A. G.; Carlone, A.; Chen, K.; Dryden, D. T. F.; Leigh, D. A.; Lewandowska, U.; Mullen, K. M. *Angew. Chem.* **2012**, *124*, 5576.
- (105) Gu, H.; Chao, J.; Xiao, S.-J.; Seeman, N. C. *Nature* **2010**, *465*, 202.
- (106) He, Y.; Liu, D. R. *Nat. Nanotechnol.* **2010**, *5*, 778.
- (107) Labeit, S.; Kolmerer, B. *Science* **1995**, *270*, 293.

## Chapter II

# The Road towards Sequence-Specific Synthesis with Artificial Synthetic Machines

### ACKNOWLEDGMENTS

All experiments and syntheses in Chapter II were carried out by Marcus Pappmeyer. Dr. Bartosz Lewandowski is acknowledged for the initial development of the active metal template synthesis of related one-barrier rotaxane systems and the optimisation of the hydrazone exchange reactions. Guillaume De Bo is acknowledged for his redesign of the synthesis of barrier **29**. Sonja Kuschel is acknowledged for numerous text reactions not presented in this Chapter. Dr. John Ward is acknowledged for the resynthesis of macrocycle **14**. Dr. Philipp M. E. Gramlich is gratefully acknowledged for his idea to incorporate Native Chemical Ligation into a molecular machine. Dr. María J. Aldegunde, Dr. Daniel M. D'Souza, Anthony E. Fernandes, Dr. Stephen M. Goldup, Dr. Philipp M. E. Gramlich and Dr. Dominik Heckmann are gratefully acknowledged for their work on previous designs that ultimately lead to the development of the system that is presented in this Chapter.

## 2.1 Synopsis

In biological systems, molecular machines are employed to perform a variety of complex tasks essential for the functioning of living organisms and have always been an intriguing inspiration for scientists. The realisation of artificial molecular analogues that mimic the function of nature's nanomachinery would ultimately allow access to complex task performance and new ways to control chemical reactions on a molecular level.

In order to achieve such an ambitious goal, model studies and extensive tests are required before the synthesis of a more complex machine—ultimately a synthetic ribosome analogue—can be undertaken. Initial model studies were conducted in order to explore the possibility to synthesise a molecular machine capable of sequence-specific peptide synthesis employing an operation mechanism based on native chemical ligation. A rotaxane containing a macrocycle bearing a reactive arm and stoppered by a leucine phenolic ester derived barrier was successfully synthesised. Pick-up of the aminoacyl residue and subsequent transfer onto the terminal amine functionality of the growing peptide chain attached to the macrocycle through native chemical ligation allowed the self-immolation of the macrocycle with the newly grown peptide attached. The operation product was identified by comparison with an extended reference sample by mass spectrometry.

## 2.2 Introduction

Sequence-specific synthesis aided by molecular machines is found in every living cell—ribosomes, DNA- and RNA-polymerases are essential for all biological processes.<sup>1</sup> Without the capability of synthesising biopolymers with high sequence-specificity life would not exist. Despite the abundance of information-rich polymers in nature with translation and transcription aided by molecular machines as examples for ultimate sequence control, approaches towards the goal of synthesising defined monomer sequences without repetitive protection/deprotection steps or subsequent addition of substrates have remained limited to few examples.<sup>2</sup>

The aim of this work is the development of a molecular machine capable of controlling monomer order in an oligomer chain. Using a synthetic organic approach would allow for the preparation of significant amounts of material, the possibility to translate information for the synthesis of a range of oligomers of different chemical nature and the operation of the machine in a variety of non-physiological conditions, e.g. in organic solvents or at a wide temperature range. Sequence-specific synthesis with artificial machines would give organic chemists a new tool formerly reserved to biological systems: the possibility to translate a given template into a new strand with a chemically distinct nature.

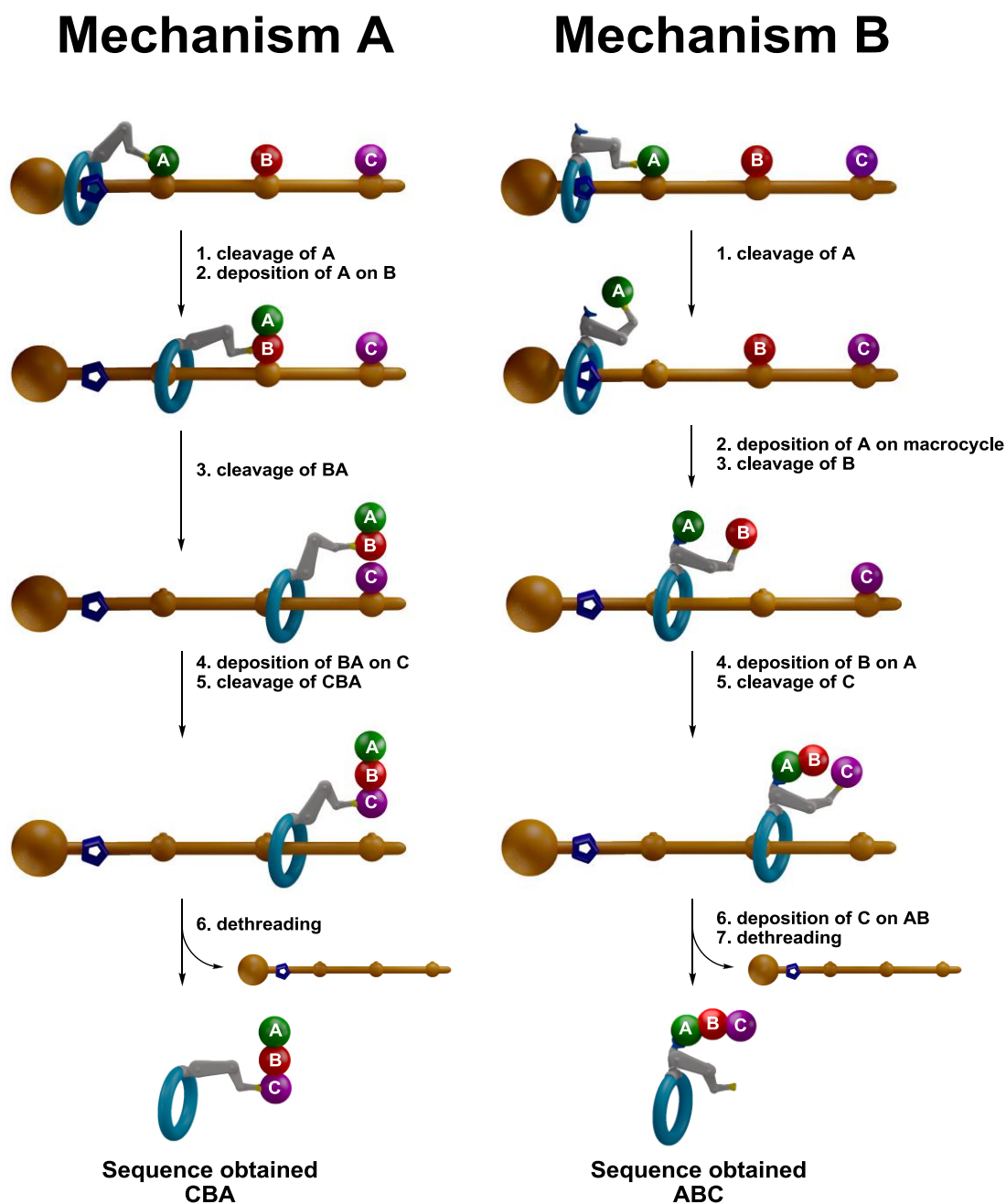
The ribosome and nucleotide polymerases can be seen as ideal and their characteristics represent a benchmark for any primitive first generation organic system. Both types of machines assemble monomers extremely fast and in a highly sequence-specific manner. They do not detach from their substrates, i.e. operate with high processivity and at the same time read out the template strand non-destructively. Furthermore, the enzyme-assisted biopolymerisation works autonomous and the machines are able to be recycled after completed synthesis. Fully repetitive and progressive operation is therefore achieved by natural systems. Certainly not all of these features are essential for an initial proof of concept in a synthetic system but they are of course desirable for the development of more advanced artificial molecular machines.

The initial idea to synthesise a molecular system capable of translation of a molecular template was introduced to the Leigh group in 2007 by Stephen Goldup (Figure 2.1).<sup>3</sup> The general concept involves a rotaxane architecture consisting of a macrocycle bearing a catalytic moiety interlocked on a relatively rigid track. The monomeric units to be assembled in a sequence are attached to the track by cleavable linkages allowing for reaction with the catalyst. The interlocked nature of the molecular machine ensures processivity of the system since the macrocycle cannot detach from the track without breakage of covalent bonds and

the rigidity of the system prevents folding and ensures that a reaction can only take place between catalyst and monomers that are in close proximity. The mechanically interlocked architecture with a macrocycle that can diffuse along a molecular strand is essential for another reason. In order to enable the sequence-specific synthesis the size of the macrocyclic cavity has to be small enough to prevent the macrocycle to move over the monomeric attachment points when a monomeric unit is connected to it which therefore functions as a “barrier” for the catalyst. On the other hand the macrocycle is designed in such a way that diffusion over the attachment point is possible after the pick up of the monomers. Different operation mechanisms can be imagined for such a molecular machine. The initial design involved a nucleophilic catalyst that can attack the first monomeric unit (Figure 2.1, Mechanism A). After cleavage of the bond between the track and monomeric unit, the activated monomer on the macrocycle is deposition onto a reactive group on the next barrier the macrocycle encounters. Transfer of the dimer to the third monomeric unit allows a specific sequence to be built up. Possible side reactions include the transfer of the activated monomer on a barrier that is not in proximity of the macrocycle. Carefully designing the track to prevent folding and utilising a catalytic unit that ensures the exclusive pick-up of the next monomer in sequence, scrambling can be avoided. After translation is complete, the macrocycle can slip off from the track and the operation product can be obtained.

Different from the initially proposed design, a second operation mechanism can be envisioned, where instead of the activated growing strand being transferred onto the next monomer, a reaction occurs with a second functional group on the macrocycle itself, thereby extending the oligomeric chain in a sequence-specific manner (Figure 2.1, Mechanism B). This pick-up and chain-elongation is then continued until the macrocycle can slip off the thread, similarly to operation mechanism A.

A molecular machine based on either mechanism would allow the sequence-specific synthesis of molecules through its processive operation. No stimuli or mechanisms for the directional movement of the macrocycle are required. The purely diffusive motion of the catalyst bearing macrocycle on the thread is sufficient to enable the reaction with the next monomeric unit that is encountered. An autonomous as well as a non-autonomous operation could be realised with such a system depending on the chemical nature of the transformation carried-out by the machine. The speed of the translation depends solely on the reactivity of the molecular machine under the operation conditions.

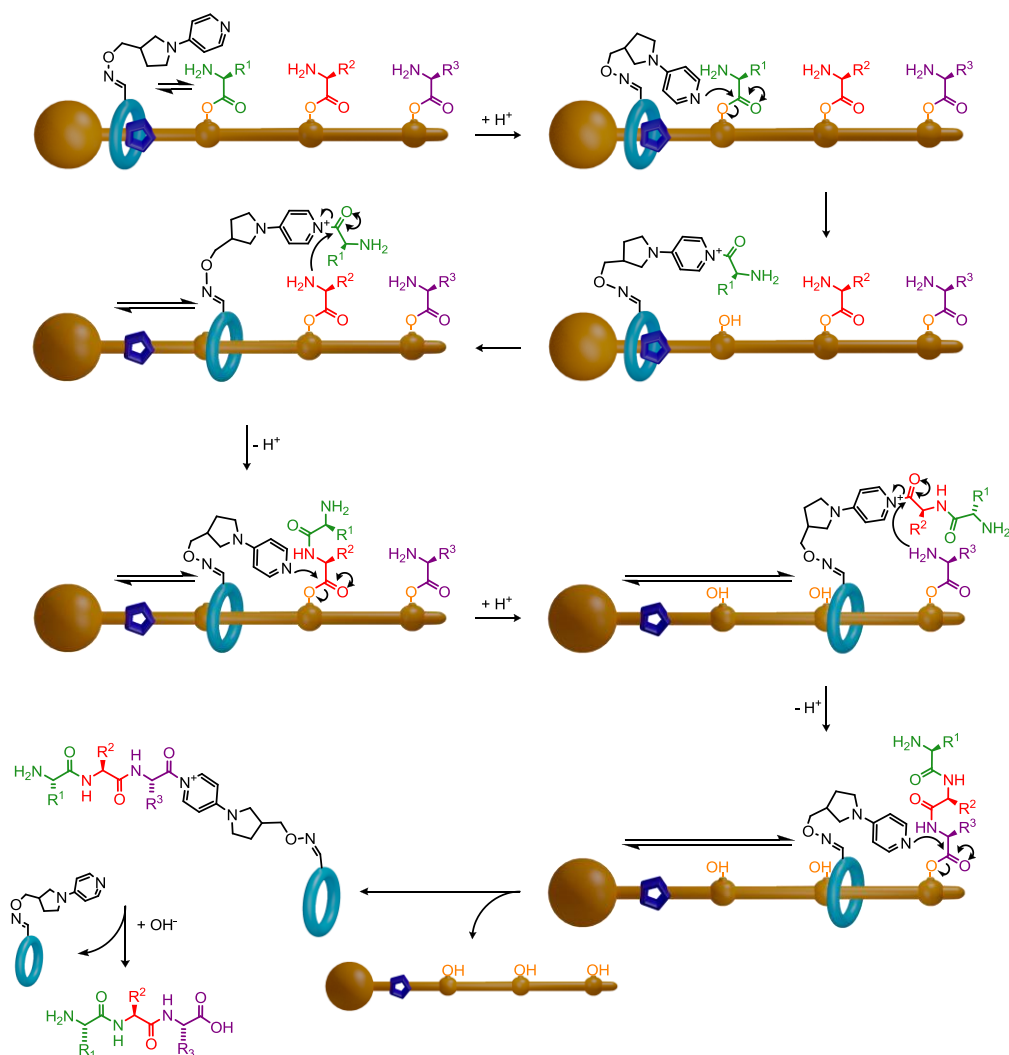


**Figure 2.1.** Schematic illustration of two possible operation mechanisms of a rotaxane-based molecular machine capable of sequence-specific assembly of three monomeric units based on their order on a track. Mechanism A: a nucleophilic catalyst (grey) attached to a macrocycle (blue) picks up monomer A. After cleaving monomer from the thread, the macrocycle can diffuse over the attachment point and deposit A on the next monomeric unit B it encounters. Continuation of this process leads to the sequence CBA. Mechanism B: a bifunctional catalyst (grey) is attached to a macrocycle (blue). Monomer A is cleaved from the track and deposited on a second nucleophilic site on the catalytic unit. Continuation of this process allows the built-up of the sequence ABC.

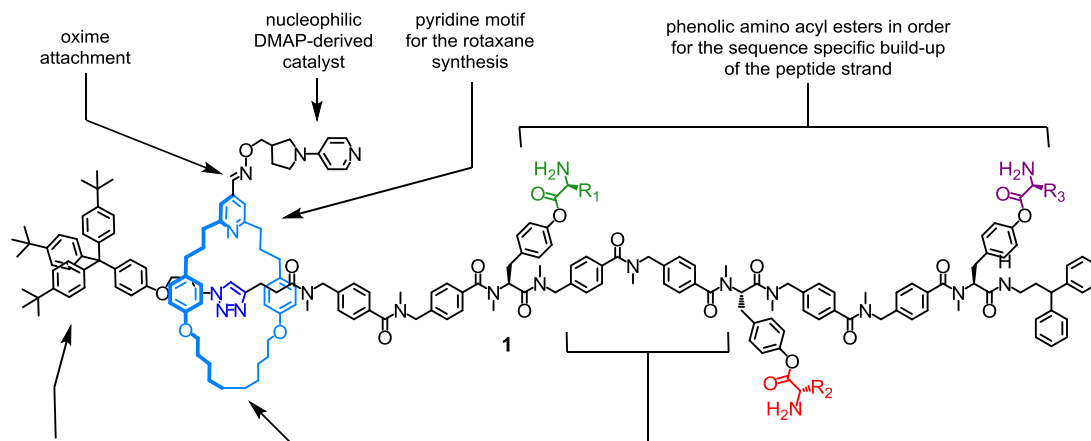
Such artificial molecular machines are based on several elements that have analogues in ribosomal protein synthesis: reactive building blocks (the role played by tRNA-bound amino acids) are delivered in a sequence determined by a molecular strand (the role played by mRNA). The macrocycle ensures processivity during the machine's operation (reminiscent of the way that subunits of the ribosome clamp the mRNA strand) and bears a catalyst that detaches the building blocks from the strand and passes them on to another site. Therefore, even these primitive molecular machines can be seen as a ribosome mimics. However, one major drawback compared to biological systems remains: the destruction of the information incorporated in the strand during the operation of the system.

### 2.2.1 First Generation Design of a Molecular Machine for Sequence-Specific Synthesis

The realisation of a first generation design of a molecular machine capable of sequence-specific synthesis was attempted by Dr. Daniel M. D'Souza, Dr. Anthony Fernandes, Dr. Stephen Goldup, and Dr. Dominik Heckmann in the Leigh group in 2007 and 2008.<sup>3</sup> Based on operation mechanism A (Figure 2.1) the synthesis of an oligopeptide was envisaged using aminoacyl groups attached to the track *via* cleavable ester linkages and a transacylation catalyst connected to the macrocycle *via* a reversibly cleavable attachment (Scheme 2.1). The nucleophilic attack and pick-up of the first amino acyl moiety allows the macrocycle to proceed to the second amino acyl-loaded ester and after amide bond formation with the free amino group of the second amino acyl moiety. Subsequent cleavage of the dipeptide allows the macrocycle to repeat the procedure until it picks up the terminal aminoacyl residue and slips off the thread. The proposed chemical structure of the first-generation molecular machine **1** is shown in Figure 2.2. Rotaxane **1** features a track with a moderately rigid *N*-methylated 4-(aminomethyl)benzoic acid-based spacer unit and the aminoacyl residues are connected to the track through labile tyrosine-derived ester linkages. The pyrrolidinopyridine-derived transacylation catalyst is connected *via* an oxime linkage to the aldehyde-functionality of an endtopic pyridine group that directs the threading of the half-threads during the active-metal template assembly of the rotaxane. A sterically demanding stopper unit ensures that the diffusion of the mechanically interlocked macrocycle along the strand results in a net movement in the direction of the aminoacyl residues and enables the translation of the sequence information in the track to the corresponding peptide.



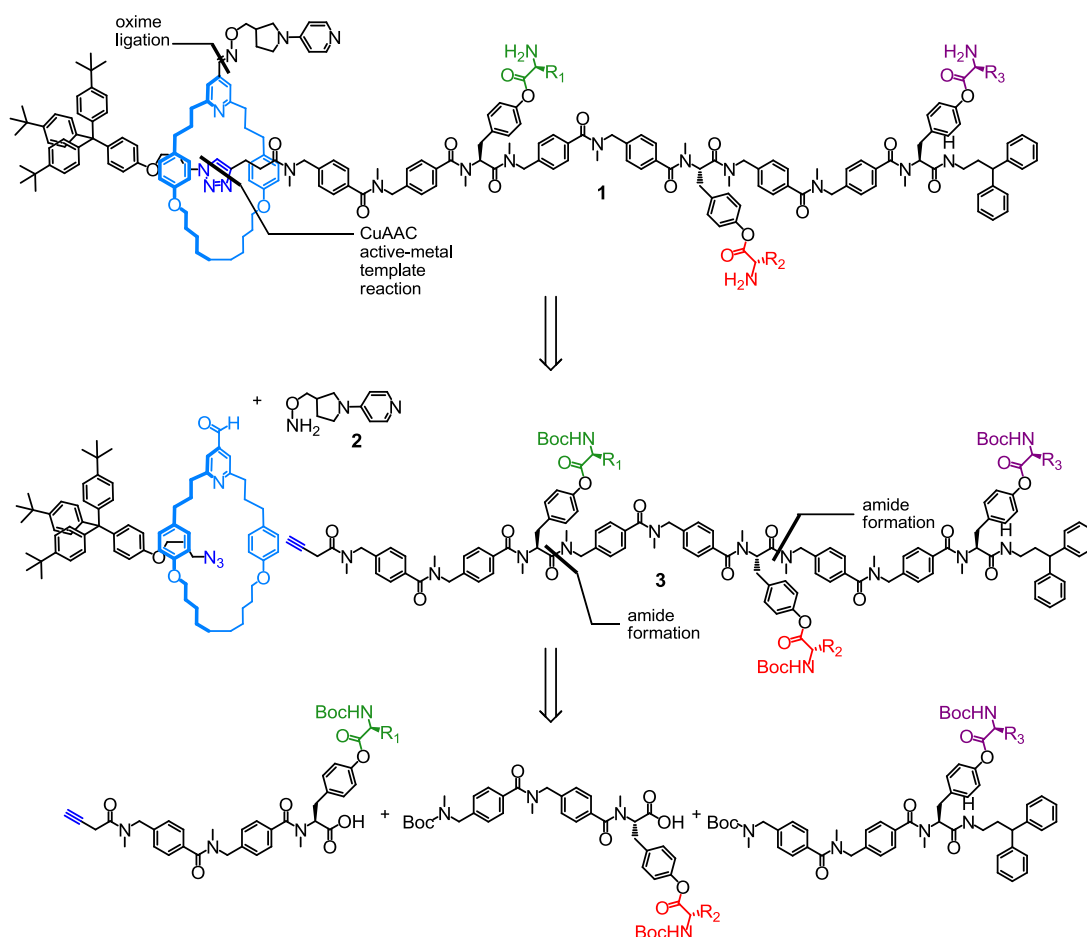
**Scheme 2.1.** Proposed mechanism for sequence-specific peptide synthesis by a molecular machine based on transacylation catalysis. After activation of the machine, successive acyl transfer reactions allow the coupling of the activated amino acid building blocks to the next amino acid the macrocycle encounters. Thereby a built up of the sequence irreversibly transcribed from the template strand is possible. Once the final amino acid is cleaved, the macrocycle bearing the synthesised oligopeptide detaches from the strand. Hydrolysis delivers the desired tripeptide.



**Figure 2.2.** First generation design of a molecular machine capable of sequence-specific synthesis.

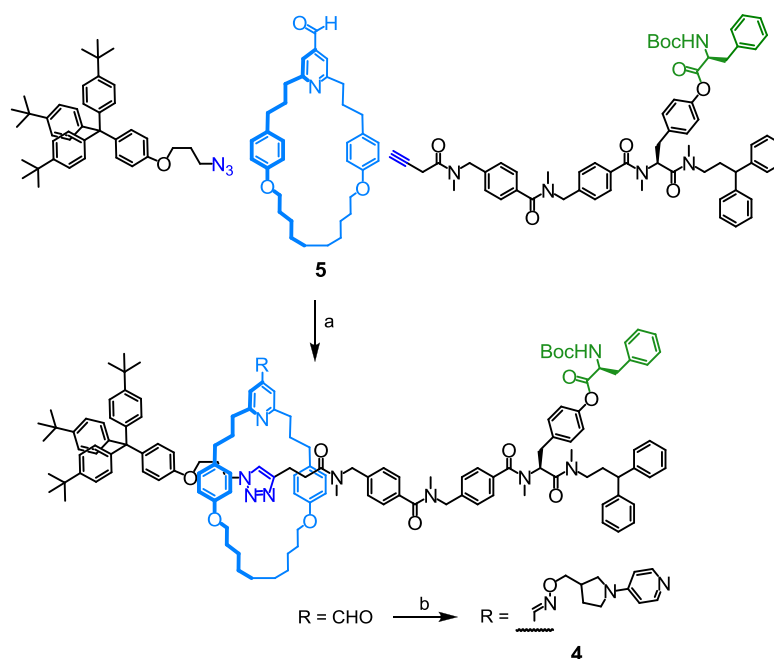
The retrosynthetic analysis of the molecular machine **1** is shown in Scheme 2.2. For such a complex molecule a fast access to all building blocks and a relatively modular design are paramount for the success of the whole project. The functional group tolerating and relatively high yielding Cu(I)-catalysed azide-alkyne (CuAAC) active-metal template reaction<sup>4</sup> was chosen as rotaxane forming reaction. Since the pyrrolidinopyridine-motif could act as a competing ligand for the Cu(I)-ions in solution, incorporation of the catalytic unit **2** after the rotaxane formation step would largely avoid detrimental non-interlocked product formation outside the cavity of the macrocycle. Choosing a suitable deprotection/elongation strategy, the aminoacyl-bearing alkyne **3** could be built up in an efficient convergent fashion through usually high-yielding amide couplings. A backbone containing *N*-methyl amide linkages offers good solubility and the facile interconversion between *cis* and *trans*-amides offers moderate flexibility allowing that the catalytic unit to react with all amino acyl residues.

In order to test the feasibility of the proposed operation principle of the molecular machine, rotaxane **4** with a phenylalanine phenolic ester as acyl-donor was synthesised and after successful rotaxane formation, the catalytic unit **2** was introduced through the proposed oxime ligation strategy (Scheme 2.3). Unfortunately the intramolecular pyrrolidinopyridine catalyst did not cleave the phenolic ester under the reaction conditions and no dethreading of the macrocycle could be observed (Scheme 2.4).

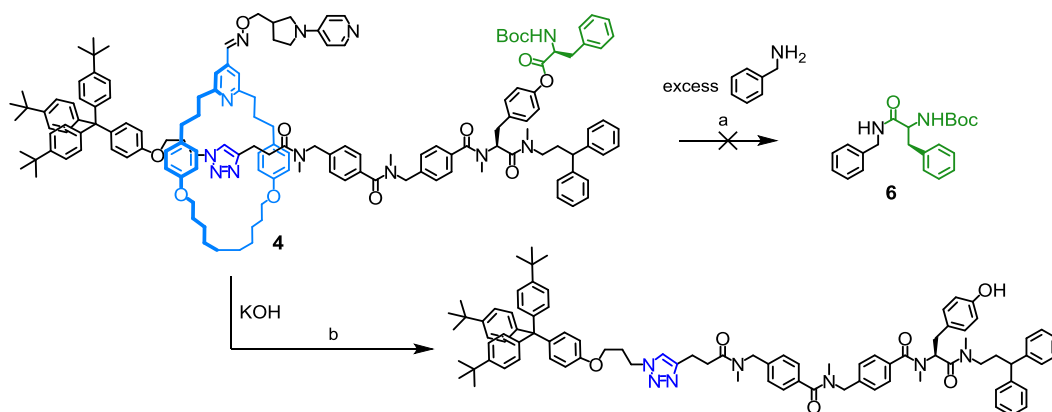


**Scheme 2.2.** Retrosynthetic analysis of molecular machine **1**.

Also, the addition of an external nucleophile (benzylamine) in high concentration did not allow for the interception of the acyl-pyridinium intermediate and the formation of 2-amino-*N*-benzyl-3-phenylpropanamide **6**. Addition of potassium hydroxide led to the saponification of the ester bond of **4** and the dethreading of the macrocycle. The most likely reason for failure of the system was assumed to be the unfavourable equilibrium between the acyl intermediate and the phenolic ester, which lies on the side of the ester. In addition, the peptide bond formation *via* the pyridinium intermediate is too slow to occur on a reasonable time scale.<sup>3</sup> Other disadvantages of the first generation design were problems with the purification even for the relatively small single-barrier model due to the incorporation of a relatively large number of polar groups (four amide linkages, one triazole), broadening of the <sup>1</sup>H-NMR-spectra due to the *cis-trans*-isomerisation of the *N*-methyl amides, and the inability to find a more reactive transacylation catalyst.



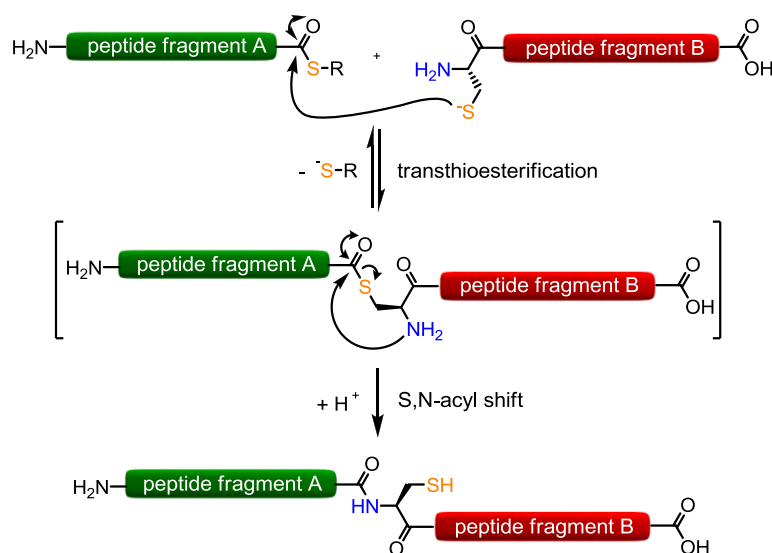
**Scheme 2.3.** Synthesis of rotaxane test substrate **4**. Reaction conditions: a)  $\text{Cu}(\text{CH}_3\text{CN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2$ , RT, 48 h; b) **2**, TFA,  $\text{CDCl}_3$ , RT, 24 h.<sup>3</sup>



**Scheme 2.4.** Operation attempt of model rotaxane **4**. Reaction conditions: a) benzyl amine,  $\text{CDCl}_3$ , 50 °C, 24 h; b) KOH, MeOH, RT, 24 h.<sup>3</sup>

## 2.2.2 Native Chemical Ligation as Operation Mechanism for a Second Generation Machine

Opposed to the first design based on a classical DMAP-derived transacylation catalyst (corresponding to Mechanism A, Figure 2.1), a second generation machine could be envisioned to work with a cysteine derivative as a catalytic unit—the sequence-specific peptide synthesis in such a machine would be achieved through native chemical ligation (NCL, the final machine would operate according to Mechanism B, Figure 2.1).<sup>5</sup> Initially developed for the condensation of unprotected peptide fragments, typically in a classical NCL reaction, the thiol-functionality of cysteine moiety on the *N*-terminus of a peptide fragment reacts with a thioester group on the *C*-terminus of a second fragment (Scheme 2.5).<sup>6</sup>



**Scheme 2.5.** Mechanism of native chemical ligation.

Through a thiol-thioester exchange both peptide strands are linked. In the presence of external thiolate nucleophiles in the reaction mixture, this thiol–thioester exchange is highly reversible. In the presence of the free amino group on the *N*-terminal cysteine, an amide bond is formed through a rapid, irreversible 1,5-*S,N*-acyl shift.<sup>7</sup> NCL has proven to be highly chemoselective (virtually no byproducts are formed), high yielding, and versatile, and therefore has been employed in many examples for the total synthesis of fully functional proteins and enzymes.<sup>6b</sup> Also, the thioester moiety is relatively stable towards hydroxyl nucleophiles, especially if the NCL is carried out in a pH-range between 6 and 8. During NCL typically no racemisation of the moderately activated peptide-thioester in the ligation

process is observed, since the reactions are typically carried out at moderate pH and room temperature.<sup>8</sup>

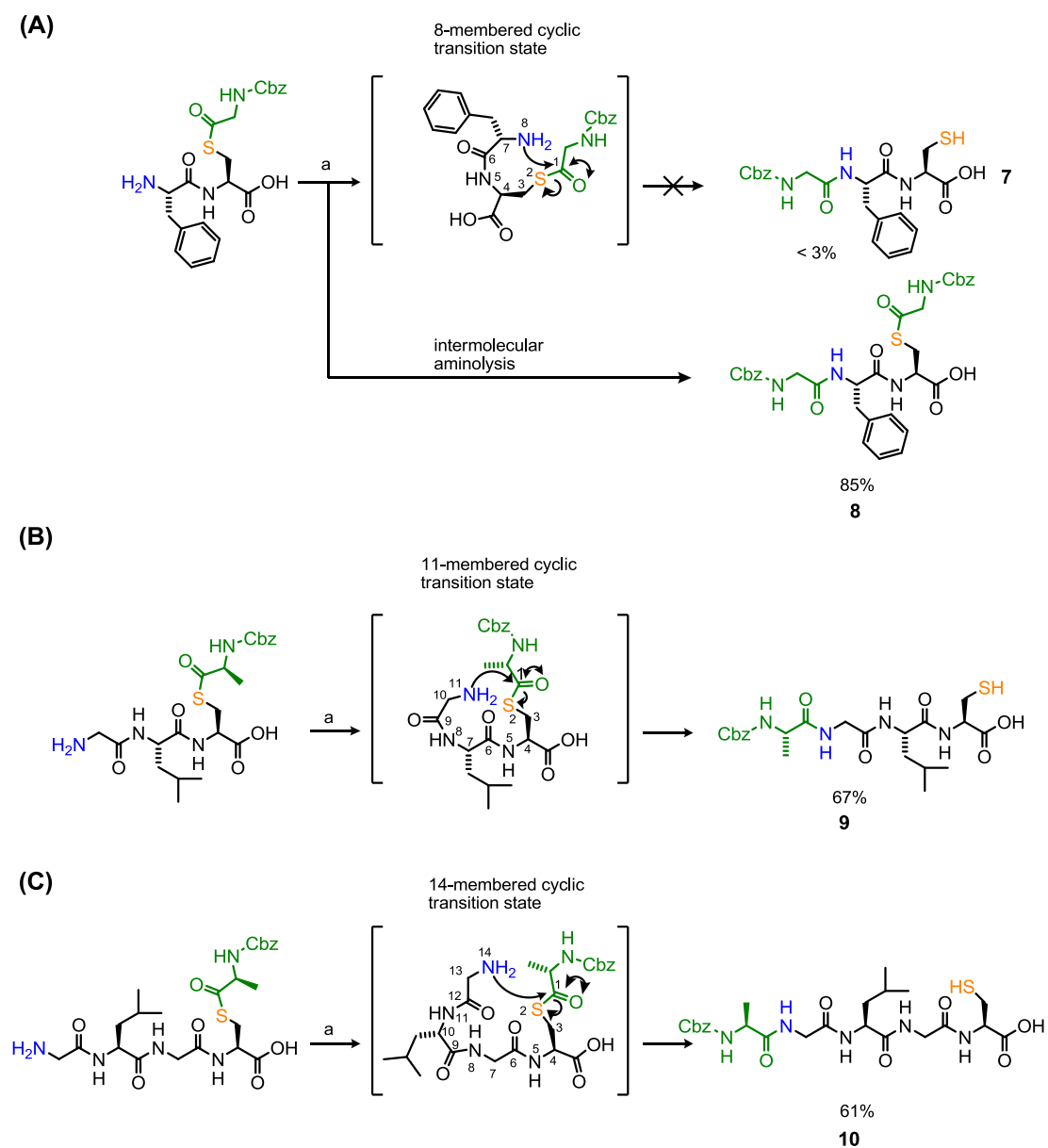
The chemoselectivity in addition to mild reaction conditions and high yield make the NCL a very good candidate for the incorporation in a complex molecular machine capable of sequence-specific synthesis, but two further requirements for a successful molecular machine based on a cysteine-derived reactive arm have to be met: (i) the catalytic unit has to be able to cleave the bond between the acyl residues and the track, and (ii) the *S,N*-shift has to work also with large transition states since otherwise only one amide bond forming reaction could be realised and the molecular machine would stall after the initial pick-up.

Since one major limitation of NCL was its restriction to peptide chains with an *N*-terminal cysteine, several research groups investigated the possibility of long range *S,N*-acyl migrations. While the 1,5-*S,N*-acyl shift is the most facile, 11- up to 29-membered transition states have been reported as feasible in literature.<sup>9</sup> On the contrary, the 8-membered transition state for the *S,N*-acyl shift is unfavoured.<sup>9a,10</sup> Instead of the expected ligation product **7** (Scheme 2.6), Katritzky and co-workers showed that disproportionation product **8** resulting from an intermolecular reaction was mainly produced during an attempted 1,8-*S,N*-acyl shift. A similar outcome was observed for cyclic 9- and 10-membered transition states, which are energetically disfavoured compared to larger transition states.<sup>10</sup> However, both NCL-products **9** resulting from a 11-membered *S,N*-shift and **10** resulting from an 14-membered *S,N*-transfer are formed in 67% and 61% yield, respectively.

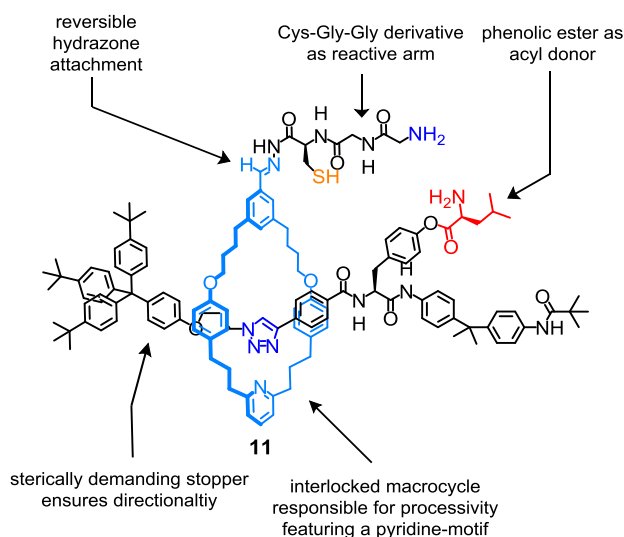
This observation has implications for the design of a molecular machine incorporating several successive *S,N*-transfers. In order to avoid an 8-membered transition state during the course of operation, the incorporation of a tripeptide with cysteine on the *C*-terminus circumvents the necessity of 8, 9 or 10-membered *S,N*-acyl transfers during the operation.

In principle, the aminoacyl moieties could be attached *via* thioester bonds to the molecular scaffold, which would favour the attack of the catalytic unit and would lead to a fast and efficient reaction. But phenolic esters are in general more synthetically accessible and more stable towards hydro- and aminolysis. Indeed, NCL has been reported to proceed with phenolic esters as acyl donors making them therefore desirable motifs for a molecular machine.<sup>11</sup>

Essential information for constructing a molecular machine capable of sequence-specific peptide synthesis based on NCL can be acquired through the synthesis and operation of a rotaxane featuring a single aminoacyl residue and a macrocycle bearing a cysteine-containing tripeptide as catalytic unit. The chemical structure of the proposed test-system is shown in Figure 2.3.



**Scheme 2.6.** Intramolecular *S,N*-acyl shift through cyclic transition states of different sizes. (A) 8-membered transition state, (B) 11-membered transition state, and (C) 14-membered transition state. Reaction conditions: a) phosphate buffer ( $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ , pH 7.8)/MeCN (4:1 v/v),  $\mu\text{W}$ ,  $50^\circ\text{C}$ . Crude yield given based on  $^1\text{H-NMR}$  as reported by Katritzky and co-workers.<sup>9a</sup>

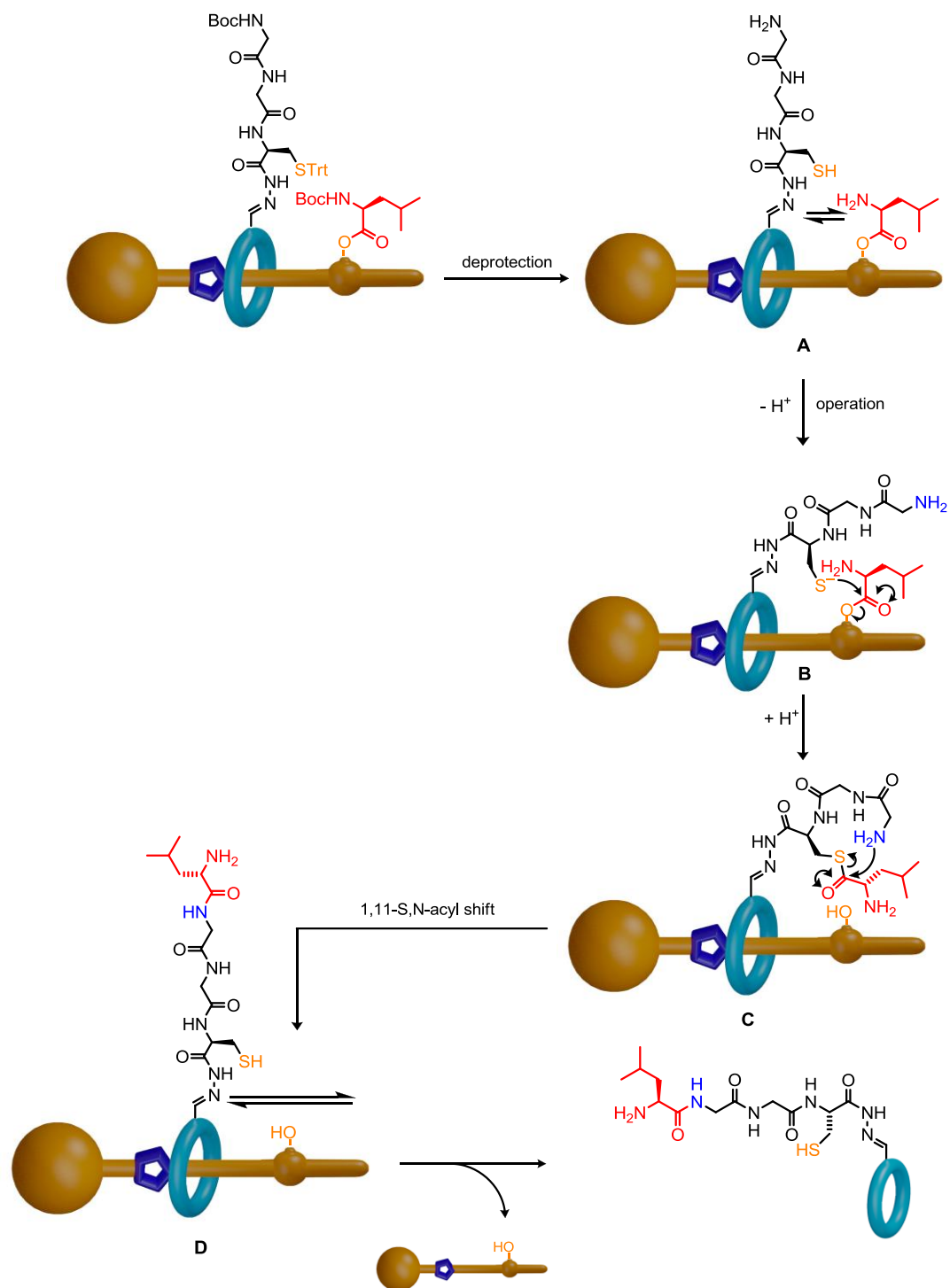


**Figure 2.3.** Rotaxane design of a test substrate for the development of a second generation molecular machine capable of sequence-specific synthesis.

Similar to the first generation design, the strand of rotaxane **11** bears an amino acyl residue attached to the track by weak phenolic ester linkage. In the second generation design, instead of using 4-(aminomethyl)benzoic acid-based spacers, a more rigid diphenylpropane unit is incorporated. In the final machine, the rigidity of such spacer groups would minimise the possibility of the reactive arm of the machine coming into contact and reacting with a building block out of sequence while still maintaining solubility of the whole system. The macrocycle in **11** is closely related to macrocycle **5** of the first generation design featuring an endotopic pyridine moiety needed for the rotaxane forming step. The attachment point for the catalytic unit was chosen to be on the opposite hemisphere of the macrocycle, since electron-withdrawing groups on the pyridine moiety lower the yield of the active metal template reaction required for the assembly of the machine.

Instead of employing an oxime connection for the catalytic unit, a hydrazone linkage was chosen since it performed well in other systems developed by our group.<sup>12</sup> As a catalytic unit a cysteine-hydrazide derivative with an attached glycylglycine chain seemed to be the most promising candidate in order to circumvent the unfavourable 8-membered transition state and realise a 11-membered transition state for the *S,N*-transfer after the pick-up of the first aminoacyl residue.

The proposed mechanism for the NCL-based operation is shown in Scheme 2.7. The design of the machine is such that once the thiol residue of the cysteine group is deprotected (A), it is poised to undergo a transacylation reaction with the first amino acid phenolic ester that blocks the macrocycle's path on the track (B).



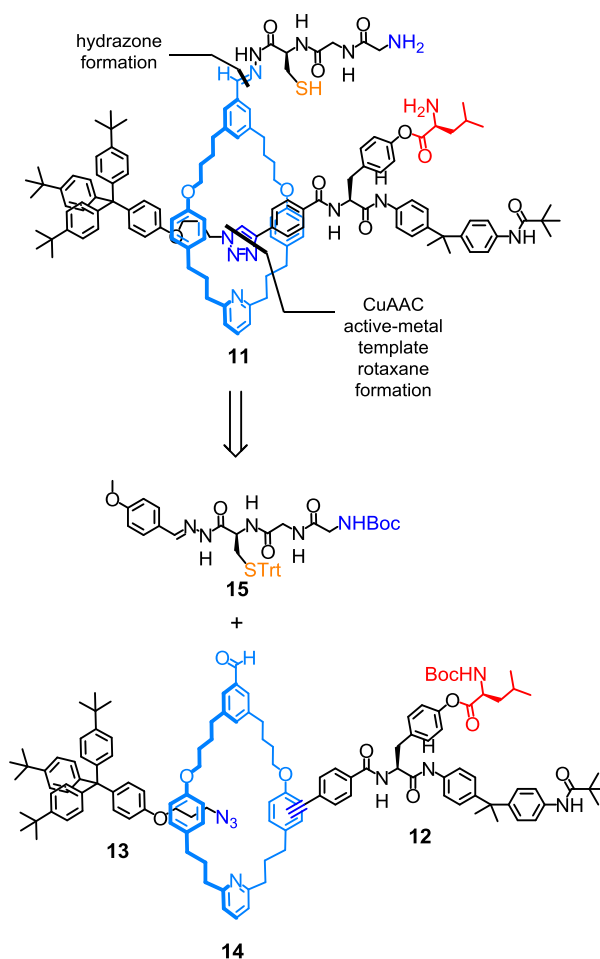
**Scheme 2.7.** Proposed mechanism for the peptide synthesis with a single-barrier model rotaxane. After activation of the machine (A) the thiol functionality is deprotonated (B) and the resulting thiolate cleaves the ester bond (C). The thioester formed undergoes a 1,11-S,N-shift (D) and extends the peptide chain by the leucine moiety. The macrocycle then detaches from the thread.

The subsequently formed aminoacyl thioester (C) reacts further, transferring the amino acid by native chemical ligation to the glycylglycine amine group by an 11-membered cyclic transition state (D). Now that the blocking group is removed, the macrocycle can slip off the thread with the loaded tetrapeptide attached. The information gained from the synthesis of a rotaxane containing a catalytic unit attached to a macrocycle are therefore:

- (i) Suitable sterics of the system: is the size of the macrocycle small enough to be mechanically interlocked before the aminoacyl residue is picked up and large enough to diffuse over the attachment point after cleavage?
- (ii) Suitable reactivity: is the Cys-Gly-Gly derivative reactive enough to cleave the phenolic ester attachment and at the same time, does the *S,N*-transfer happen on a reasonable time scale? Do side reactions interfere?
- (iii) Synthesis and stability: are all components mutually compatible? Is the protecting group-strategy suitable for the assembly of the system?
- (iv) Isolation: is it possible to recover the individual parts of the rotaxane after operation?

### 2.2.3 Retrosynthetic Analysis of a Single-Barrier Model Rotaxane

The proposed retrosynthesis of molecular machine **11** is shown in Scheme 2.8. The crucial interlocking of the components can be achieved by the active metal template Cu(I)-catalysed cycloaddition of the terminal alkyne loaded with the leucine residue **12** with the azide-bearing stopper group **13** mediated by macrocycle **14**.

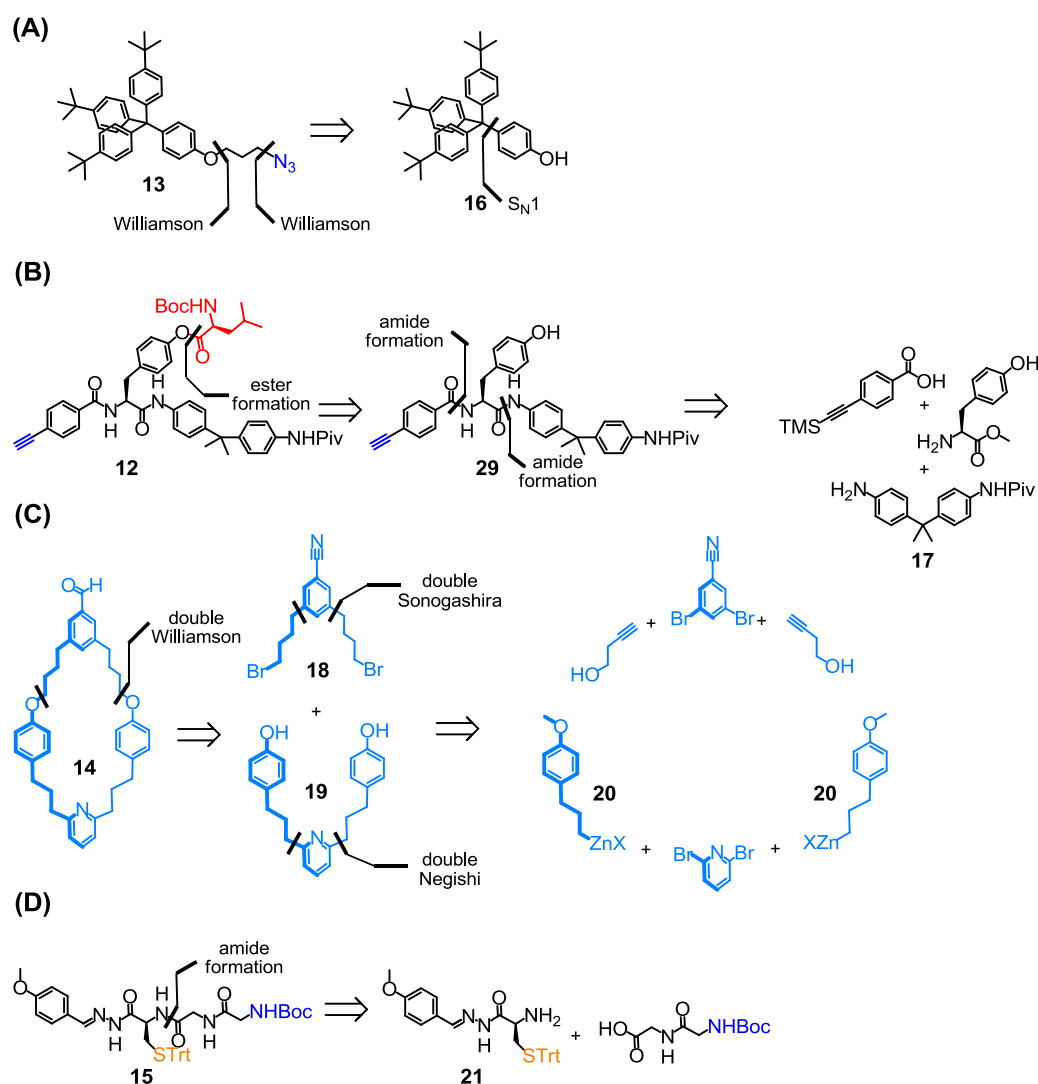


**Scheme 2.8.** Retrosynthetic analysis of rotaxane **11**.

The reactive unit containing a Cys-Gly-Gly motif is incompatible with the active-metal template reaction and therefore should be ideally incorporated after the rotaxane formation. An exchange reaction with anisaldehyde-derived hydrazone such as **15** as reported by Greaney and co-workers is particularly mild and tolerates ester functionalities.<sup>13</sup>

The individual building blocks can be synthesised in a relatively straight-forward manner (Scheme 2.9). Azide functionalised stopper **13** can be derived from the literature known phenol **16** using two Williamson reactions (Scheme 2.9, A).<sup>14</sup> Pivaloyl-protected barrier **12** is

derived from tyrosine methyl ester by subsequent amide coupling, hydrolysis and coupling with aniline **17** (Scheme 2.9, B). Macrocycle **14** can be synthesised by a double Williamson macrocyclisation of terminal dibromide **18** and diphenol U-shape **19**. The northern hemisphere is built up from Sonogashira-couplings of 3,5-dibromobenzonitrile with homopropargyl alcohol, subsequent hydrogenation and Appel-reaction of the alcohol functionalities. The southern hemisphere can be disconnected into 2,6-dibromopyridine and readily accessible zincate **20** (Scheme 2.9, C). Hydrazone exchange reagent **15** can be synthesised by amide coupling of Boc-protected glycyl-glycine with cysteine hydrazone **21** (Scheme 2.9, D).

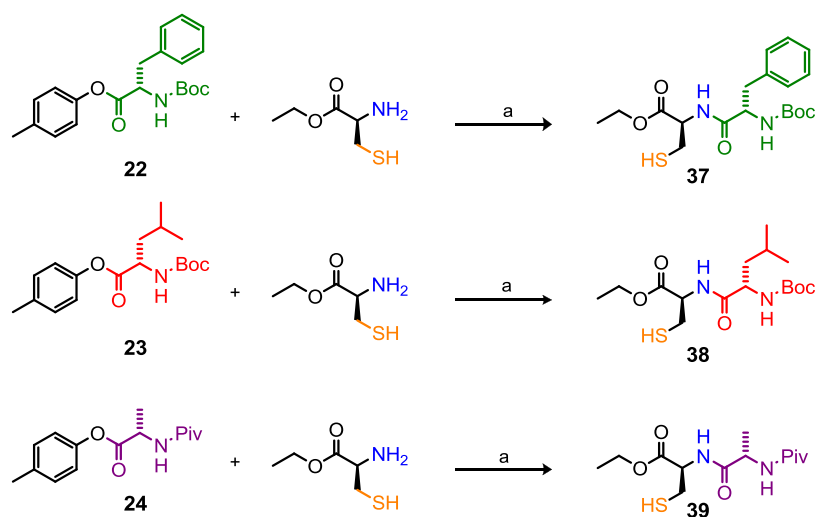


**Scheme 2.9.** Retrosynthetic analysis of the building blocks required for rotaxane **11**. (A) Azide-alkyl stopper **13**. (B) *N*-Boc-leucine-loaded barrier **12**. (C) Aldehyde macrocycle **14**. (D) Hydrazone exchange reagent **15**.

## 2.3 Results and Discussion

### 2.3.1 Intermolecular Model Studies

As a first step we decided to conduct test reactions determining suitable conditions for NCL with phenolic ester derivatives. Essential information obtained from those test reactions is: (i) identification of suitable amino acids, (ii) investigation of stability of the phenolic ester and catalytic unit under the reaction conditions, and (iii) general information on reaction conditions of NCL in non-aqueous solvents. The choice of amino acids that can be coupled to cysteine is limited to sterically non-demanding substrates and generally  $\beta$ -branching is not tolerated in the acyl donor-component of the NCL (e.g. valine and proline thioesters give low conversion under standard NCL conditions).<sup>8</sup> Amino acids with charged side-chains might be problematic for synthetic reasons and it is known that strongly ligating groups such as amines or nitrogen-containing heterocycles lower the yield of the active metal template reaction significantly, leaving the aliphatic amino acids leucine and alanine (the steric demand of glycine is too low for incorporation into the molecular machine as CPK-models suggest that the macrocycle could pass over the amino acid attachment point without cleavage) as well as phenylalanine as most promising substrates for a molecular machine. Therefore, cysteine methyl ester was reacted with three *p*-cresol amino acid esters derived from phenylalanine, leucine, and alanine (Scheme 2.10). It was realised that a reducing agent was needed for successful reactions since disulfide formation, that readily occurs when trace amounts of oxygen are present, renders the cysteine ethyl ester inactive.



**Scheme 2.10.** Investigation of the suitability of phenolic esters in native chemical ligation. Reaction conditions: a) TCEP·HCl, DIPEA, CH<sub>3</sub>CN, 60 °C,  $\mu$ w, 24h.

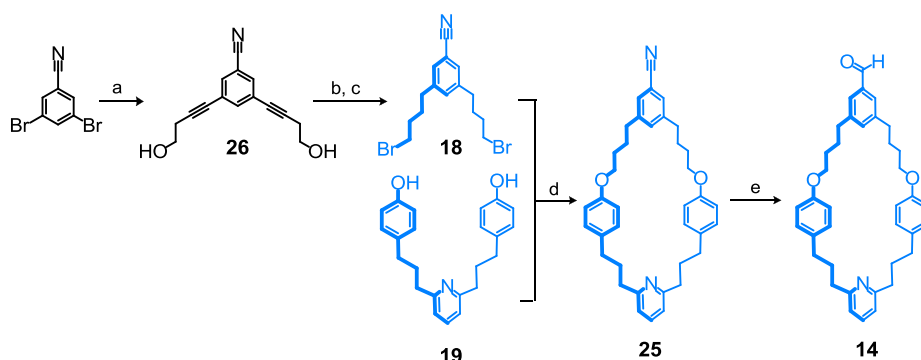
A commonly used additive for this purpose in NCL-chemistry is tris(2-carboxyethyl)phosphine (TCEP).<sup>15</sup> Since the final molecular machine was envisioned to be relatively hydrophobic, reaction conditions in organic solvents instead of aqueous buffers typically used for NCL-chemistry seemed most promising. Good results were obtained in acetonitrile and microwave irradiation. Heating cysteine ethylester with excess DIPEA at 60 °C for 24 h resulted in the complete consumption of all *p*-cresol aminoacyl esters **22**, **23**, and **24** as monitored by TLC, with the concomitant appearance of a new peak in the mass spectrum corresponding to the respective dipeptide. Therefore, phenylalanine derivative **22**, leucine derivative **23**, and alanine derivative **24** seemed to be suitable for NCL. Since chromatographic isolation of the resulting peptides proved to be low yielding, somewhat irreproducible and did not seem to reflect the actual amount of successful peptide formation that occurred in the reaction mixture, a second stability test submitting the *p*-cresol esters to the reaction conditions in the absence of cysteine ethyl ester to check for possible hydrolysis or other decomposition pathways was necessary. In all cases, after 24 h heating at 60 °C in the microwave no decomposition could be observed by TLC or mass spectrometry analysis.

### 2.3.2 Synthesis of a *N*-Boc-Leucine Containing Single-Barrier Rotaxane

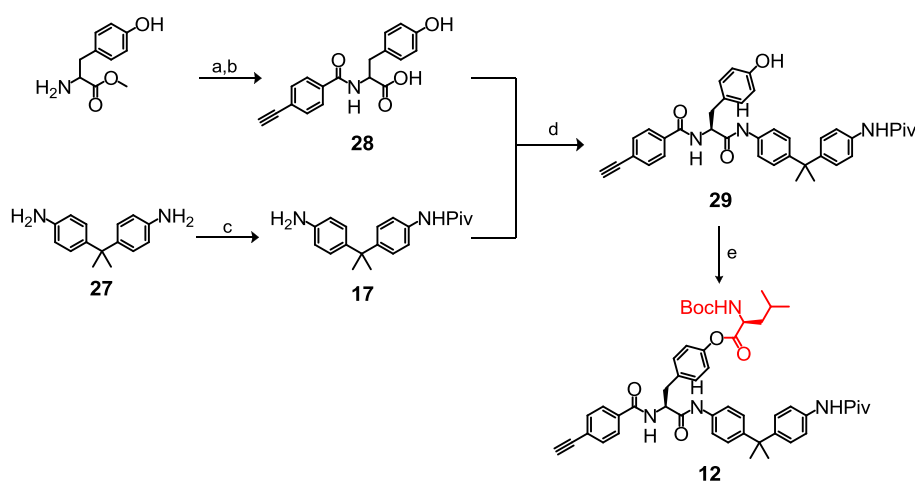
Intermolecular tests can only be a starting point for the exploration of reaction conditions for a molecular machine. Since the effective concentration and conformation of the catalytic unit are very different in the final architecture compared to the situation in intermolecular test systems, the first aim of the project was the synthesis of a rotaxane loaded with a single aminoacyl residue. This model system should give information on whether (i) the macrocycle is interlocked on the thread when the amino acyl residue is loaded as well as passes over the attachment point after deacylation and (ii) whether the envisioned NCL-mechanism is in operation or if side reactions (e.g. hydrolysis of the ester) predominate. The key intermediates for the synthesis of such a system are azide stopper **13**, aldehyde macrocycle **14**, hydrazone exchange reagent **15** and *N*-Boc-leucine containing alkyne **12** (Scheme 2.9).

Bulky azide stopper **13** was synthesised following a literature procedure.<sup>14</sup> The key step for the synthesis of the macrocycle **14** is the double Williamson macrocyclisation to give **25** (Scheme 2.11). The synthesis of the upper hemisphere of macrocycle **14** was started by Sonogashira-coupling of 3,5-dibromobenzonitrile with homopropargyl alcohol to deliver alcohol **26**. Hydrogenation of the alkyne bonds and subsequent Appel reaction gave desired

bromide **18** in good yield. The union of dibromide **18** with literature known diphenol **19**<sup>4a</sup> under high dilution conditions afforded nitrile macrocycle **25**. A final reduction of the nitrile to reveal the aldehyde functionality gave access to target macrocycle **14** in good yield.



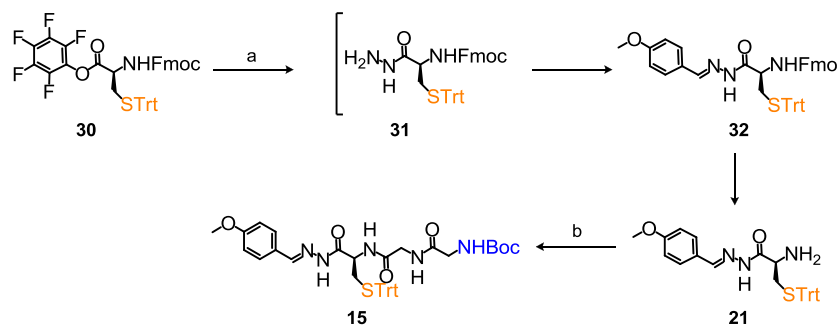
**Scheme 2.11.** Synthesis of aldehyde macrocycle **3**. Reaction conditions: a) but-3-yn-1-ol, Pd(PPh<sub>3</sub>)<sub>4</sub>, CuI, THF/Et<sub>3</sub>N (2.5:1 v/v), 80 °C, 12 h, 65%; b) Pd/C, H<sub>2</sub>, THF, 12 h, 98%; c) PPh<sub>3</sub>, CBr<sub>4</sub>, CH<sub>2</sub>Cl<sub>2</sub>, RT, 12 h, 52%; d) Cs<sub>2</sub>CO<sub>3</sub>, DMF, 50 °C, 72 h, 40%; e) DIBAL-H, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 12 h, 72%.



**Scheme 2.12.** Synthesis of alkyne barrier **12**. Reaction conditions: a) 4-[(trimethylsilyl)ethynyl]benzoic acid, HOBt, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v), RT, 12 h, 62%; b) LiOH, THF/H<sub>2</sub>O (4:1 v/v), 12 h, 99%; c) pivaloyl chloride, DMAP, Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub>, RT, 12 h, 55%; d) HOBt, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v), RT, 12 h, 74%; e) L-N-Boc-leucine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 52%.

The amino acyl loaded tyrosine barriers can be synthesised through a variety of different approaches. Using four steps for the longest linear sequence and avoiding amine protecting groups, monofunctionalisation of diamine **27** with pivaloyl chloride to give pivaloyl-derivative **17** was chosen as initial step (Scheme 2.12). For the left half of the molecule,

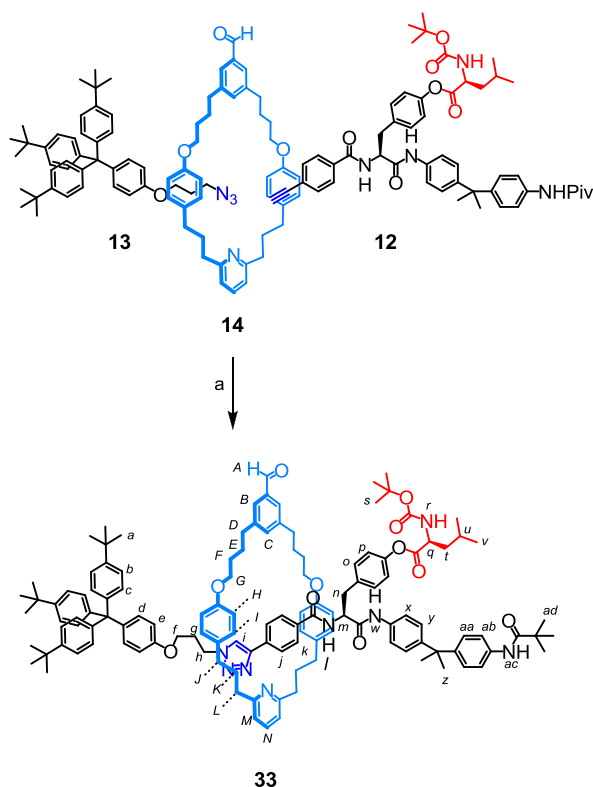
tyrosine methyl ester was selectively coupled to 4-([trimethylsilyl]ethynyl)benzoic acid in presence of the phenol functionality using hydroxybenzotriazole as additive. Hydrolysis with LiOH removed the trimethylsilane group as well as cleaving the methyl ester to deliver carboxylic acid **28**. Amide coupling to unite **17** and **28** gave phenol **29**. *N*-Boc-leucine was incorporated using DMAP/EDCI coupling conditions and target molecule **12** was obtained in moderate yield.



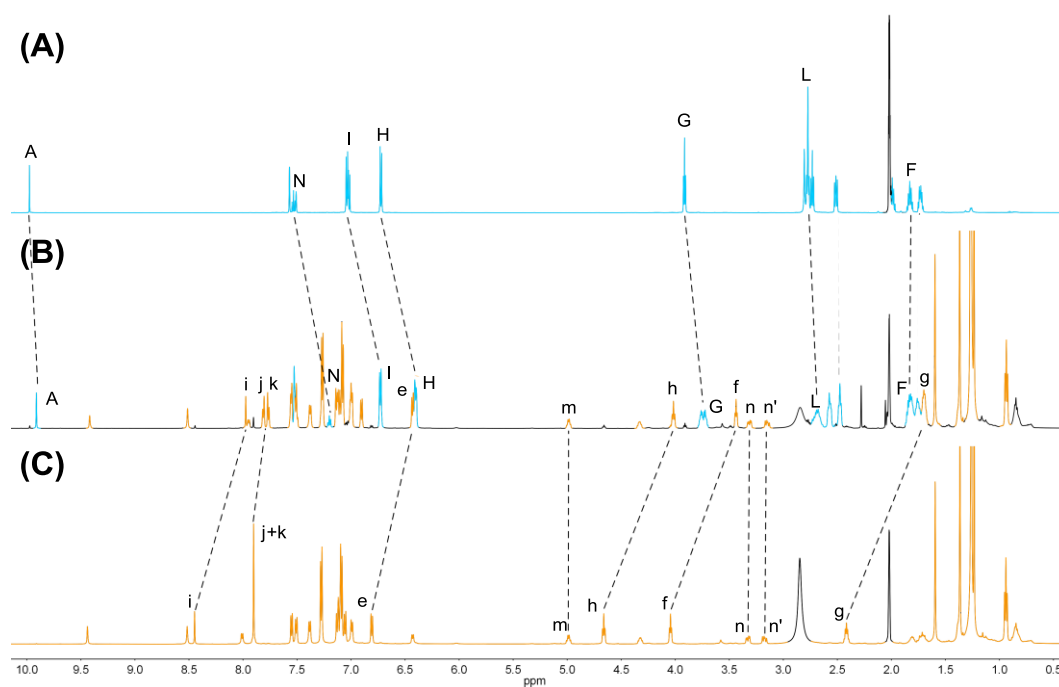
**Scheme 2.13.** Synthesis of hydrazone exchange reagent **15**. Reaction conditions: a)  $\text{NH}_2\text{NH}_2$ , MeOH, RT, 2 h, then 4-methoxybenzaldehyde, AcOH (cat.), MeOH, RT, 12 h, then piperidine, DMF, RT, 2 h, 75% over three steps; b) *N*-Boc-Gly-Gly-OH, 1,1'-carbonyldiimidazole, THF, RT, 12 h, 47%.

The synthesis of hydrazone exchange reagent **15** required for loading the catalytic unit onto the rotaxane was started from commercially available pentafluorophenyl cysteine ester **30** (Scheme 2.13). Addition of hydrazine at room temperature yielded hydrazide **31** and left the Fmoc-group untouched. Direct hydrazone formation on the crude material to give **32** went smoothly, but isolation of amine derivative **21** proved problematic as it hydrolysed on silica. Best results were obtained by submitting the crude reaction mixture after deprotection directly to the coupling with *N*-Boc-Gly-Gly-OH to deliver hydrazone **15**.

For the rotaxane formation step, it is essential to stir aldehyde macrocycle **14** with tetrakis(acetonitrile)copper(I) hexafluorophosphate in  $\text{CH}_2\text{Cl}_2$  for 30 minutes before addition of azide stopper **13** and aminoacyl-loaded alkyne **12** in a mixture of  $\text{CH}_2\text{Cl}_2$  and *t*BuOH (Scheme 2.14). Purification by preparative thin layer chromatography delivered rotaxane **33** in 30% yield. The stacked plot of rotaxane **33** can be seen in Figure 2.4.



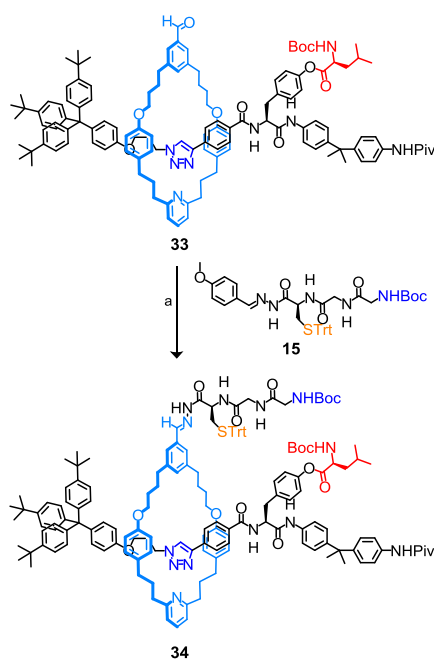
**Scheme 2.14.** Synthesis of rotaxane **33**. Reaction conditions: a)  $\text{Cu}(\text{MeCN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v), RT, 48 h, 30%. The italicised letters indicate signals in the  $^1\text{H}$  NMR spectrum shown in Figure 2.4.



**Figure 2.4.**  $^1\text{H}$  NMR spectrum of (A) aldehyde macrocycle **14**, (B) rotaxane **33**, and (C) corresponding free thread **36**, in acetone- $d_6$  (600 MHz, 298 K). The assignments correspond to the lettering shown in Scheme 2.14.

The interlocked nature of rotaxane **33** becomes apparent from the significant upfield shift of all proton signals in proximity to the central triazole moiety as well as moderate upfield shifts for characteristic macrocyclic proton signals. Such shielding is typical for interlocked architectures in which the aromatic rings of one component (the para-substituted phenyl rings of macrocycle **14**) are face-on to another component (here the triazole of the thread). The most significant upfield shifts can therefore be observed for H<sub>H</sub> and H<sub>I</sub> of macrocycle as well as signals H<sub>f</sub> to H<sub>k</sub> of the thread. However, signals corresponding to the tyrosine moiety and the leucine residue remain largely unaffected. The limitation of the upfield shifts to certain proton signals of the thread implies that the macrocycle is largely confined between the sterically demanding tyrosine barrier and the bulky stopper on the other side as expected from this rotaxane.

The loading of the catalytic moiety was achieved by using hydrazone exchange conditions developed by Greaney and co-workers.<sup>13</sup> The main concern was the stability of the acid labile protecting groups at low pH. Stirring aldehyde rotaxane **33** with hydrazone exchange reagent **15** in DMSO/2-(*N*-morpholino)ethanesulfonic acid (MES) buffer (pH 6.0) for 72 h at 60 °C with aniline as nucleophilic catalyst delivered target rotaxane **34** in moderate yield (Scheme 2.15) and left the ester and *N*-Boc-functionalities unaffected.



**Scheme 2.15.** Synthesis of hydrazone rotaxane **34**. Reaction conditions: a) aniline, DMSO/MES buffer (3:1 *v/v*, pH 6.0), 60 °C, 72 h, 45%.

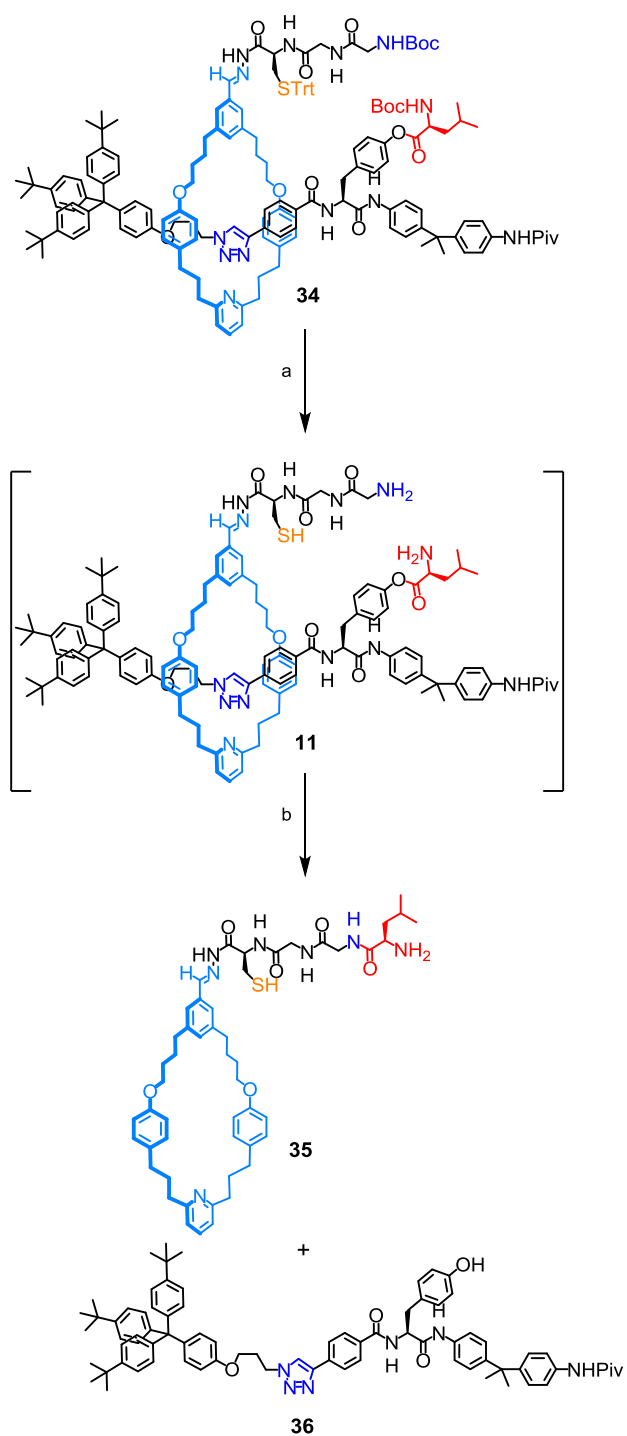
### 2.3.3 Operation of Model Rotaxane 11

For the operation of the rotaxane in the first step the acid labile Boc- and trityl-protecting groups were removed with  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v, Scheme 2.16). Since the purification of the unprotected rotaxane proved difficult, the crude reaction mixture was submitted to the previously established operation conditions obtained from the intermolecular model studies directly. For reasons of solubility, the solvent was altered from acetonitrile as used for the intermolecular test reactions to a mixture of MeCN/DMF (3:1 v/v). After 24 h heating in the microwave under a protective atmosphere, the thiol-containing macrocyclic operation product was oxidised by stirring the reaction mixture in air to ease purification. The resulting disulfide was purified by preparative thin layer column chromatography. Even with this optimised purification conditions only trace amounts of the expected macrocyclic product **35** (or the corresponding disulfide dimer) could be obtained as opposed to 81% yield for the free thread **36**. This discrepancy in yield between **35** and **36** is most likely due to hydrolysis and loss of the hydrazone operation product **35** on silica. The isolated macrocycle could be identified by mass spectrometry and comparison to reference material prepared through conventional synthesis.

This result indicates that the macrocycle is interlocked on the thread when the amino acyl residue is loaded and can dissociate from the thread after amino acyl pick-up. The Cys-Gly-Gly unit seems to be a suitable catalytic unit for the cleavage of the phenolic esters and subsequent *S,N*-transfer to form the amide bond occurs under the reaction conditions.

The operation under high dilution conditions (the concentration of the molecular machine under the operation conditions is smaller than 2 mmol/l) make intermolecular reactions highly unlikely as compared to the intramolecular reaction of the molecular machine as is envisioned by the design. The effective molarity of the catalytic unit can be assumed to be much higher than the concentration of molecular machines in solution.

This result therefore suggests that the synthesis and operation of a larger molecular machine capable of the sequence-specific synthesis of peptides is in reach.



**Scheme 2.16.** Operation of molecular machine **34**. Reaction conditions: a) CH<sub>2</sub>Cl<sub>2</sub>/TFA (5:1 v/v), RT, 2 h 52%; b) DIPEA, TCEP·HCl, MeCN/DMF (3:1 v/v), 60 °C μw, 24 h.

## 2.4 Conclusion

First model studies were conducted to explore the possibility to synthesise a molecular machine capable of the sequence-specific peptide synthesis. The envisioned operation mechanism is based on NCL. The macrocycle bearing a Cys-Gly-Gly-containing catalytic unit is envisioned to pick-up the aminoacyl residues it encounters and deposits them through an *S,N*-transfer on the terminal amine functionality on the growing peptide strand. In order to investigate suitable operation conditions for such a machine, several test reactions were conducted, showing that phenolic esters are suitable acyl donors for the *in-situ* formation of the thioester intermediate. As amino acids, phenyl alanine, leucine and pivaloyl protected alanine were identified as promising substrates. In order to show an initial proof of concept of such an molecular machine, a rotaxane containing the aminoacyl residue leucine and the catalytic unit Cys-Gly-Gly attached to the macrocycle by a reversible hydrazone formation was successfully synthesised. The expected operation product could be isolated and was identified by comparison with reference material that was prepared by conventional synthesis. These promising results indicate that the synthesis of the full molecular machine bearing three different aminoacyl residues is indeed possible with an extended system based on the scaffold presented here.

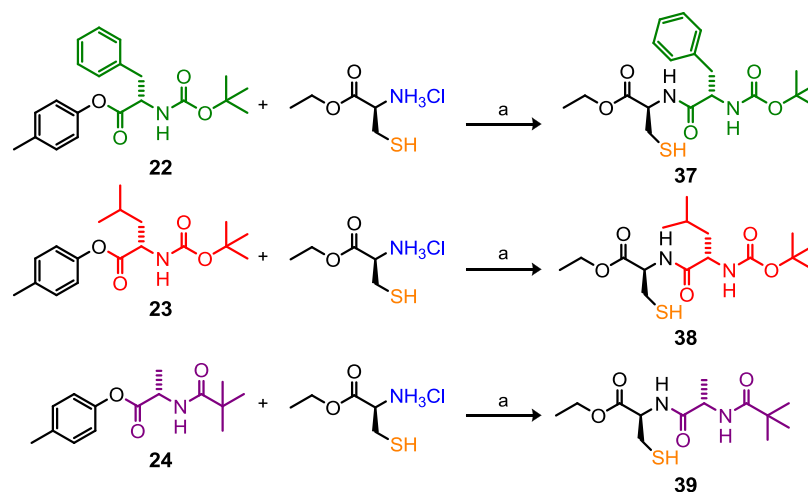
## 2.5 Experimental Section

### 2.5.1 General Methods and Abbreviations

Unless stated otherwise, all reagents and solvents were purchased from Aldrich Chemicals and used without further purification. Compounds **13**<sup>4b</sup> and **19**<sup>4a</sup> were prepared according to literature procedures. Compound **E3** was prepared by Dr. Bartosz Lewandowski. All reactions requiring microwave heating were carried out in a Biotage Initiator 2.5.3. Dry DMF, CH<sub>2</sub>Cl<sub>2</sub>, CHCl<sub>3</sub> and THF was obtained by passing the solvent (HPLC grade) through an activated alumina column on a PureSolv™ solvent purification system (InnovativeTechnologies Inc., MA). Flash column chromatography was carried out using Geduran® Si 60 (particle size 40-63 µm, Merck, Germany) as the stationary phase, and analytical TLC was performed on precoated silica gel plates (0.25 mm thickness, 60 F254, Merck, Germany) and observed under UV light or stained with a phosphomolybdic acid solution in ethanol. Preparative TLC was performed on precoated silica gel plates (2 mm, UNIPLATE GF, Analtech Inc., DE or 2 mm, Tapered plate, UNIPLATE-T, Analtech Inc., DE). NMR spectra were recorded on Bruker AV 300 MHz, Bruker AV 400 MHz, Bruker AV 500 MHz (equipped with a cryoprobe) or Bruker Avance III (equipped with a cryoprobe) with an Oxford AS 600 MHz magnet. Chemical shifts are reported in parts per million (ppm) from high to low frequency and referenced to the residual solvent resonance. Coupling constants (*J*) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintuplet, m = multiplet, br = broad. <sup>1</sup>H-NMR assignments were made using 2D NMR methods (COSY, HSQC, HMBC). Melting points (m.p.) were determined using a Sanyo Gallenkamp apparatus and are reported uncorrected. Low resolution ESI mass spectrometry was performed with a Finnigan LCQ-MS mass spectrometer. High resolution ESI (electrospray ionisation) and APCI (atmospheric-pressure chemical ionisation) mass spectrometry was carried out by the mass spectrometry services at the EPSRC National Mass Spectrometry Service Centre, Swansea, UK or the Mass Spectrometry Service of the University of Manchester, UK. Optical rotations were recorded using a Bellingham and Stanley ADP 220 Polarimeter using a 1 dm cell.

## 2.5.2 Test reactions

### 2.5.2.1 General Procedure for the Initial NCL Test-Reactions



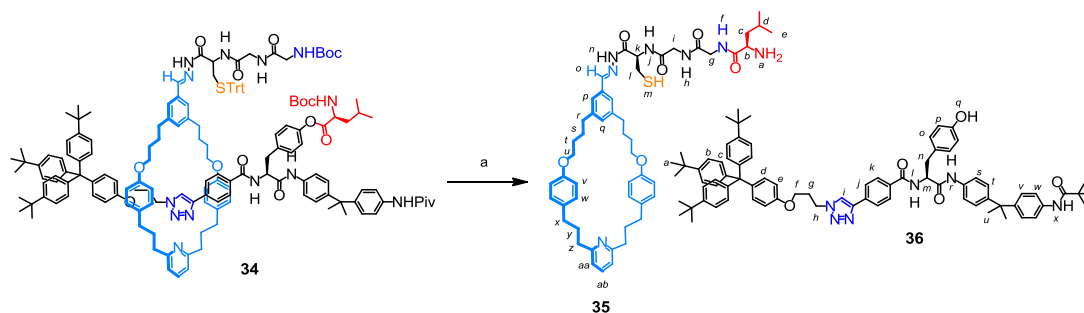
**Scheme 2.17.** Reaction conditions: a) TCEP·HCl, DIPEA, CH<sub>3</sub>CN, 60 °C,  $\mu$ w, 24 h.

A phenolic aminoacyl ester (0.11 mmol, 1.00 equiv.), L-cysteine methyl ester hydrochloride (28 mg, 0.11 mmol, 1.00 equiv.) and TCEP·HCl (62 mg, 0.22 mmol, 2.00 equiv.) were added to a flame dried microwave vial. The vial was evacuated and refilled with nitrogen three times. A separately prepared, degassed solution of 1M DIPEA in MeCN (1.00 mmol, 1.00 ml, 9.10 equiv.) was added and the reaction mixture was heated at 60 °C in the microwave for 24 h. TLC analysis revealed the complete disappearance of phenolic ester (petroleum ether/EtOAc, 7:3 v/v). Mass spectrometric analysis revealed peaks corresponding to the expected dipeptide. LRESI(+)-MS (MeOH) of **37**:  $m/z$  419.3 [M+ Na]<sup>+</sup>; LRESI(+)-MS (MeOH) of **38**:  $m/z$  385.3 [M+ Na]<sup>+</sup>; LRESI(+)-MS (MeOH) of **39**:  $m/z$  327.4 [M+ Na]<sup>+</sup>.

### 2.5.2.2 Stability Tests of Phenolic Esters

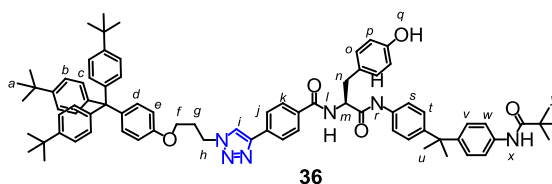
A phenolic ester aminoacyl ester (0.11 mmol, 1.00 equiv.) and TCEP·HCl (62 mg, 0.22 mmol, 2.00 equiv.) were added to a flame dried microwave vial. The vial was evacuated and refilled with nitrogen three times. A separately prepared degassed solution of 1M DIPEA in MeCN (1.00 mmol, 1.00 ml, 9.10 equiv.) was added and the reaction mixture was heated at 60 °C in the microwave for 24 h. TLC analysis revealed a visible spot corresponding to phenolic ester in the reaction mixture (petroleum ether/EtOAc, 7:3 v/v) and no visible spot corresponding to *p*-cresol. Mass spectrometric analysis revealed peaks corresponding to the phenolic esters starting materials. LRESI(+)-MS (MeOH) of **22**:  $m/z$  378.3 [M+ Na]<sup>+</sup>; LRESI(+)-MS (MeOH) of **23**:  $m/z$  322.2 [M+ Na]<sup>+</sup>; LRESI(+)-MS (MeOH) of **24**:  $m/z$  264.4 [M+ Na]<sup>+</sup>.

### 2.5.3 Operation of Molecular Machine 34



**Scheme 2.18.** Operation of molecular machine. Reaction conditions: a)  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v), RT, 2 h, then DIPEA, TCEP·HCl, MeCN/DMF (3:1 v/v),  $\mu\text{w}$ , 60 °C, 48 h.

Molecular machine **1** (20 mg, 7.8  $\mu\text{mol}$ ) was stirred at room temperature in a mixture of  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v, 2 ml) for 2 h. Toluene (5 ml) was added and the solvent removed under reduced pressure. The crude mixture was then dissolved in degassed MeCN/DMF (3:1 v/v, 4 ml) and both *N,N*-diisopropylethylamine (100  $\mu\text{l}$ , 0.58 mmol, 74 equiv.) and TCEP·HCl (3.0 mg, 11  $\mu\text{mol}$ , 1.4 equiv.) were added and the reaction mixture was stirred under microwave heating at 60 °C for 24 h. The reaction mixture was allowed to stir on air for 5 h. Preparative TLC ( $\text{SiO}_2$  [UNIPLATE-T],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  5%) of the crude residue afforded **35** (1.0 mg) as a colourless film. The residue was analysed by LRESI-MS (MeOH):  $m/z$  906.6  $[\text{M}+\text{H}]^+$ ;  $m/z$  952.6  $[\text{M}+\text{COOH}+\text{H}]^+$ ,  $m/z$  985.4  $[\text{M}+\text{COOH}+\text{Na}]^+$ .

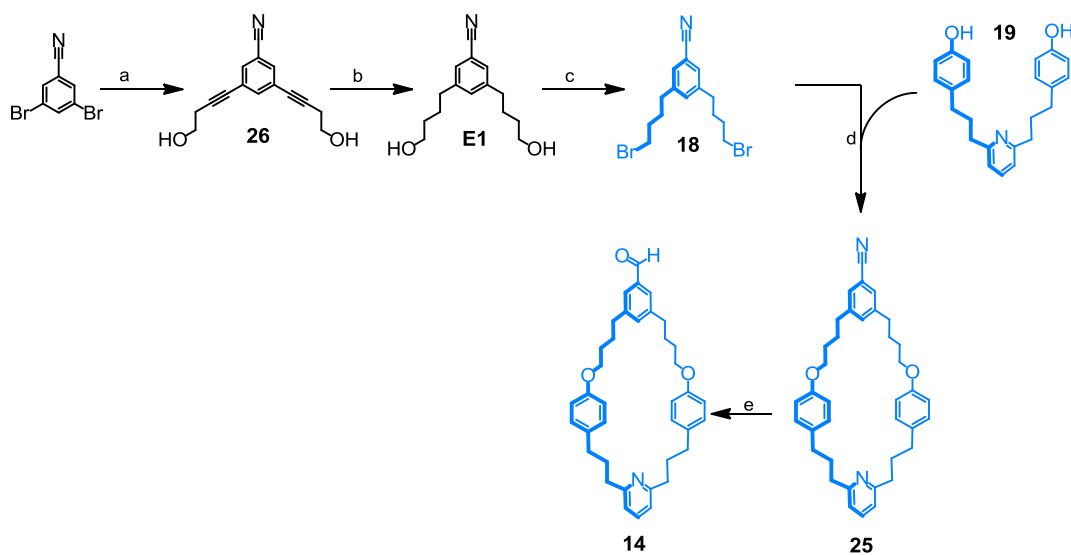


Title compound **36** was isolated as a side product from the reaction mixture for operation of **34** as a colourless film (7.5 mg, 81%).

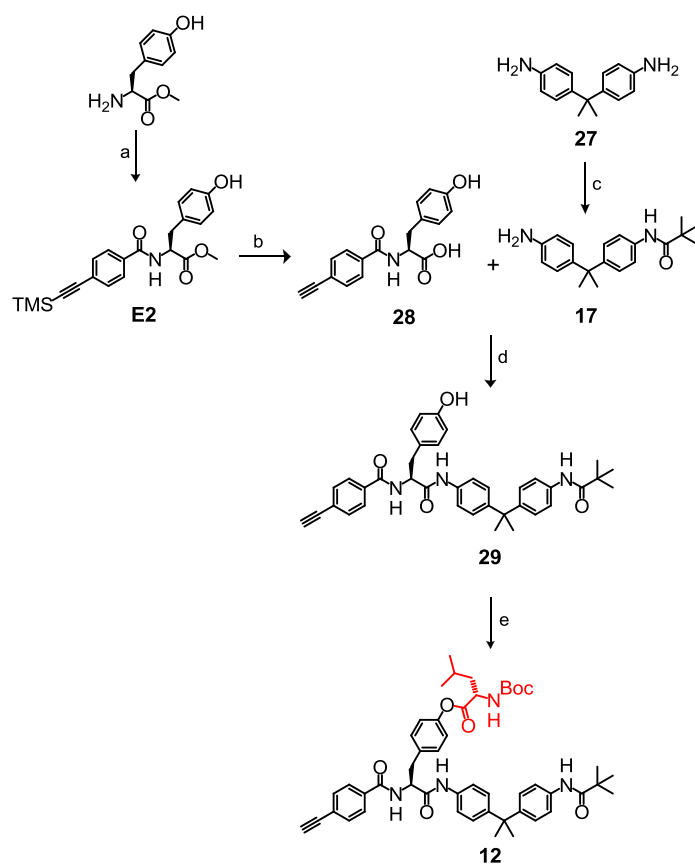
$^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ )  $\delta$  = 8.65 (s, 1H,  $\text{H}_r$ ), 7.83 (s, 1H,  $\text{H}_i$ ), 7.76 (s, 4H,  $\text{H}_{j+k}$ ), 7.46 (s, 1H,  $\text{H}_x$ ), 7.44 (d,  $J$  = 7.8 Hz, 1H,  $\text{H}_l$ ), 7.36 (d,  $J$  = 8.7 Hz, 2H,  $\text{H}_y$ ), 7.27 (d,  $J$  = 8.6 Hz, 2H,  $\text{H}_w$ ), 7.24 (d,  $J$  = 8.5 Hz, 6H,  $\text{H}_b$ ), 7.14 – 7.08 (m, 8H,  $\text{H}_{t+c}$ ), 7.07 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_v$ ), 7.04 (d,  $J$  = 8.2 Hz, 2H,  $\text{H}_o$ ), 6.77 (d,  $J$  = 8.8 Hz, 2H,  $\text{H}_e$ ), 6.68 (d,  $J$  = 8.0 Hz, 2H,  $\text{H}_p$ ), 4.62 (t,  $J$  = 7.0 Hz, 2H,  $\text{H}_h$ ), 4.00 (t,  $J$  = 5.6 Hz, 2H,  $\text{H}_j$ ), 3.23 – 3.09 (m, 2H,  $\text{H}_n$ ), 2.42 (p,  $J$  = 6.4 Hz, 2H,  $\text{H}_g$ ), 1.59 (s, 6H,  $\text{H}_u$ ), 1.31 (s, 27H,  $\text{H}_a$ ), 1.28 (s, 9H,  $\text{H}_s$ );  $^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  = 177.2, 169.9, 167.3, 156.3, 155.6, 148.5, 147.1, 147.0, 146.7, 144.2, 140.3, 135.4, 135.0, 133.9, 133.0, 132.5, 130.8, 130.6, 128.0, 127.4, 127.3, 125.8, 124.2, 121.1, 120.5, 120.3, 120.2, 115.9, 113.1, 67.5, 64.1, 63.2, 56.3, 47.6, 42.4, 37.8, 34.4, 31.5, 30.2, 27.7; LRESI(+)-MS:  $m/z$  1189.7  $[\text{M}+\text{H}]^+$ .

## 2.5.4 Synthetic Schemes

### 2.5.4.1 Synthesis of Aldehyde-Functionalised Pyridyl-Macrocycle 14

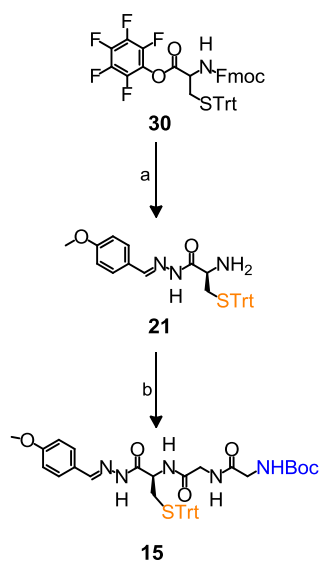


**Scheme 2.19.** Reaction conditions: a) but-3-yn-1-ol, Pd(PPh<sub>3</sub>)<sub>4</sub>, CuI, THF/Et<sub>3</sub>N (2.5:1 v/v), 80 °C, 12 h, 65%; b) Pd/C, H<sub>2</sub>, THF, 12 h, 98%; c) PPh<sub>3</sub>, CBr<sub>4</sub>, CH<sub>2</sub>Cl<sub>2</sub>, RT, 12 h, 52%; d) Cs<sub>2</sub>CO<sub>3</sub>, DMF, 50 °C, 72 h, 40%; e) DIBAL-H, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 12 h, 72%.

2.5.4.2 Synthesis of *N*-Boc-Protected Leucine-Barrier 12

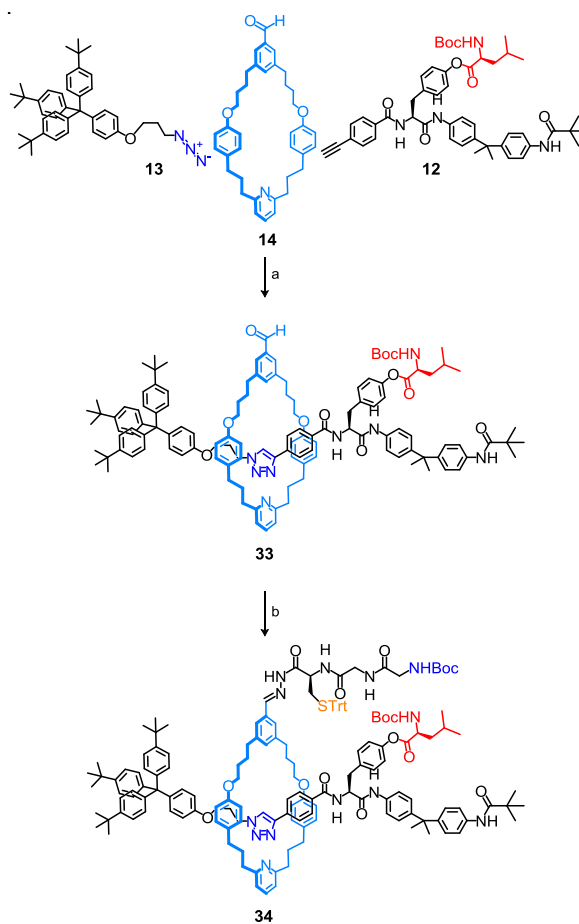
**Scheme 2.20.** Reaction conditions: a) 4-[(trimethylsilyl)ethynyl]benzoic acid, HOBT, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v), RT, 12 h, 62%; b) LiOH, THF/H<sub>2</sub>O (4:1 v/v) 12 h, 99%; c) pivaloyl chloride, DMAP, Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub>, RT, 12 h, 55%; d) HOBT, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v), RT, 12 h, 74%; e) *L*-*N*-Boc-leucine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 52%.

## 2.5.4.3 Synthesis of Hydrazone Exchange Reagent 15



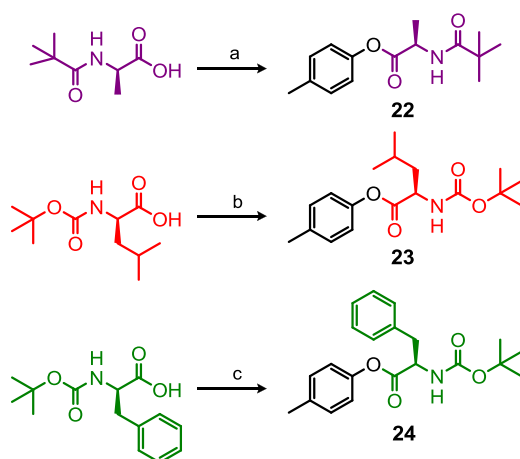
**Scheme 2.21.** Reaction conditions: a) NH<sub>2</sub>NH<sub>2</sub>, MeOH, RT, 2 h, then 4-methoxybenzaldehyde, AcOH (cat.), MeOH, RT, 12 h, then piperidine, DMF, RT, 2 h, 75% over three steps; b) *N*-Boc-Gly-Gly-OH, 1,1'-carbonyldiimidazole, THF, RT, 12 h, 47%.

## 2.5.4.4 Synthesis of Single-Barrier Rotaxane 33 and 34

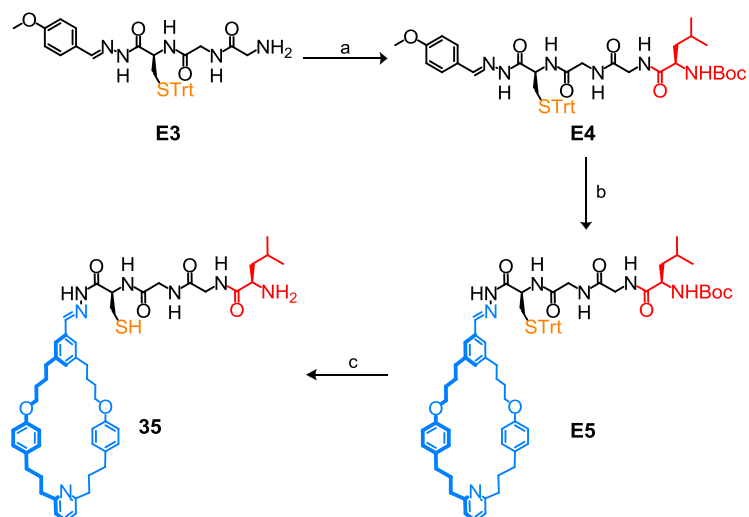


**Scheme 2.22.** Reaction conditions: a)  $\text{Cu}(\text{MeCN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v), RT, 48 h, 45%; b) aniline, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 72 h, 52%.

## 2.5.4.5 Synthesis of Phenolic Esters 22, 23, and 24



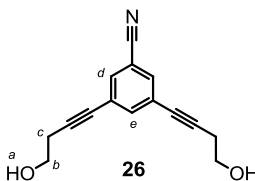
**Scheme 2.23.** Reaction conditions: a) L-N-Piv-alanine, 1,1'-carbonyldiimidazole, THF, 50 °C, 12 h, 52%; b) L-N-Boc-leucine, 1,1'-carbonyldiimidazole, THF, 50 °C, 12 h, 62%; c) L-N-Boc-phenylalanine, 1,1'-carbonyldiimidazole, THF, 50 °C, 12 h, 58%.

2.5.4.6 Synthesis of Authentic Reference Macrocycle **35**

**Scheme 2.24** Reaction conditions: a) L-N-Boc-leucine, PyBrOP, DMF, RT, 24 h, 60%; b) **13**, aniline, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 72 h, 52%; c)  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v), RT, 2 h.

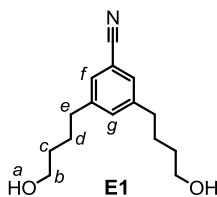
## 2.5.5 Synthetic Procedures and Characterisation Data

### 3,5-Bis(4-hydroxybut-1-ynyl)benzonitrile **26**



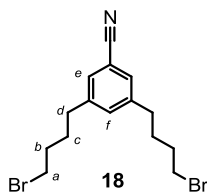
To a stirred solution of 3,5-dibromobenzonitrile (6.00 g, 22.9 mmol, 1.00 equiv.), Pd(PPh<sub>3</sub>)<sub>3</sub>Cl<sub>2</sub> (1.61 g, 2.29 mmol, 0.10 equiv.) and CuI (0.87 g, 4.59 mmol, 0.20 equiv.) in THF/Et<sub>3</sub>N (2.5:1 v/v, 168 ml) at room temperature was added 3-butyn-1-ol (6.96 ml, 91.9 mmol, 4.00 equiv.). The reaction mixture was then heated at 80 °C for 12 h. The reaction mixture was allowed to cool to room temperature and filtered over celite. The filtrate was diluted with EtOAc (200 ml) and washed with aqueous saturated NH<sub>4</sub>Cl (2 × 200 ml). The organic layer was dried over MgSO<sub>4</sub>, filtered, and the solvent removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 2:8) of the residue afforded **26** (5.10 g, 93%) as an off-white solid.

m.p. 65 °C; <sup>1</sup>H NMR (400 MHz, acetone-*d*<sub>6</sub>) δ = 7.70 (d, *J* = 1.5 Hz, 2H, H<sub>d</sub>), 7.66 (s, 1H, H<sub>e</sub>), 4.08 – 4.01 (m, 2H, H<sub>a</sub>), 3.74 (q, *J* = 6.4 Hz, 4H, H<sub>b</sub>), 2.63 (t, *J* = 6.6 Hz, 4H, H<sub>c</sub>); <sup>13</sup>C NMR (100 MHz, acetone-*d*<sub>6</sub>) δ = 139.1, 134.5, 126.7, 118.1, 114.2, 92.3, 79.5, 61.2, 24.4; LRESI(+)-MS (MeOH): *m/z* 262.1 [M+ Na]<sup>+</sup>; HRFT(+)-MS: *m/z* 240.1016 [M+H]<sup>+</sup>, 240.1019 calc. for C<sub>15</sub>H<sub>14</sub>NO<sub>2</sub>.

**3,5-Bis(4-hydroxybutyl)benzonitrile E1**

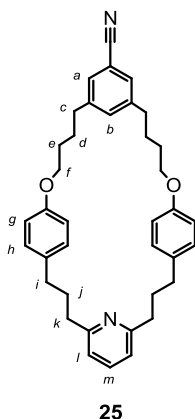
To a stirred solution of **26** (5.10 g, 21.3 mmol, 1.00 equiv.) in THF (100 ml) at room temperature was added palladium on carbon 10% wt. loading (0.50 g) in one portion and the reaction mixture was stirred under a hydrogen atmosphere at room temperature for 12 h. The reaction mixture was filtered over celite and the solvent removed under reduced pressure to afford **E1** (5.20 g, 98%) as a colourless oil.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.30 (d,  $J$  = 1.5 Hz, 2H,  $\text{H}_f$ ), 7.25 (s, 1H,  $\text{H}_g$ ), 3.65 (t,  $J$  = 6.4 Hz, 4H,  $\text{H}_b$ ), 2.65 (t,  $J$  = 7.5 Hz, 4H,  $\text{H}_e$ ), 2.05 (s, 2H,  $\text{H}_a$ ), 1.75 – 1.65 (m, 4H,  $\text{H}_c$ ), 1.63 - 1.54 (m, 4H,  $\text{H}_d$ );  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  = 143.7, 133.4, 129.5, 119.3, 112.1, 62.4, 35.1, 32.0, 27.2; LRESI(+)-MS (MeOH):  $m/z$  270.1  $[\text{M}+\text{Na}]^+$ ; HRFT(+)-MS:  $m/z$  248.1643  $[\text{M}+\text{H}]^+$ , 248.1645 calc. for  $\text{C}_{15}\text{H}_{22}\text{NO}_2$ .

**3,5-Bis(4-bromobutyl)benzonitrile 18**

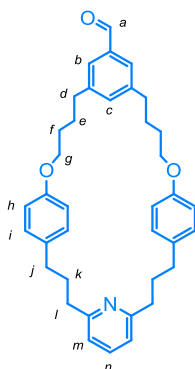
To a stirred solution of **E1** (5.20 g, 21.0 mmol, 1.00 equiv.) and tetrabromomethane (27.9 g, 84.2 mmol, 4.00 equiv.) in  $\text{CH}_2\text{Cl}_2$  (500 ml) at 0 °C was added triphenylphosphine (16.6 g, 63.2 mmol, 3.00 equiv.) in small portions and then the reaction mixture was allowed to warm to room temperature. The reaction mixture was stirred at room temperature for 12 h and the solvent was removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ , petroleum ether/EtOAc 95:5) of the residue afforded **18** (4.11 g, 52%) as a colourless oil.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.30 (d,  $J$  = 1.4 Hz, 2H,  $\text{H}_e$ ), 7.22 (br s, 1H,  $\text{H}_f$ ), 3.42 (t,  $J$  = 6.5 Hz, 4H,  $\text{H}_d$ ), 2.64 (t,  $J$  = 7.6 Hz, 4H,  $\text{H}_c$ ), 1.84 – 1.80 (m, 4H,  $\text{H}_b$ ), 1.82 – 1.72 (m, 4H,  $\text{H}_a$ );  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  = 143.3, 133.2, 129.6, 119.1, 112.4, 34.5, 33.3, 32.0, 29.5; LRESI(+)-MS (MeOH):  $m/z$  371.0  $[\text{M}+\text{H}]^+$ ; HRFT(+)-MS:  $m/z$  370.9870  $[\text{M}+\text{H}]^+$ , 370.9878 calc. for  $\text{C}_{15}\text{H}_{19}\text{NBr}_2$ .

Nitrile-Functionalised Pyridylmacrocycle **25**

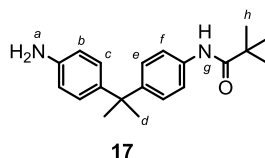
To a stirred solution of  $\text{Cs}_2\text{CO}_3$  (17.9 g, 55.1 mmol, 5.00 equiv.) in DMF (2.50 L) at 50 °C was added a solution of **18** (4.11 g, 11.0 mmol, 1.00 equiv.) in DMF (50 ml) and diphenol **19** (3.82 g, 11.0 mmol, 1.00 equiv.) in DMF (50 ml) at a rate of 2.5 ml/h over a period of 20 h with the aid of two syringe pumps. After completed addition, the reaction mixture was stirred at this temperature for a further 48 h and then the solvent was removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ , petroleum ether/EtOAc 7:3) of the residue afforded **25** (2.46 g, 40%) as a white solid.

m.p. 108 °C;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.49 (t,  $J$  = 7.6 Hz, 1H,  $\text{H}_m$ ), 7.30 (br s, 3H,  $\text{H}_{a+b}$ ), 7.05 (d,  $J$  = 8.6 Hz, 4H,  $\text{H}_h$ ), 6.96 (d,  $J$  = 7.6 Hz, 2H,  $\text{H}_l$ ), 6.78 (d,  $J$  = 8.6 Hz, 4H,  $\text{H}_g$ ), 3.94 (t,  $J$  = 5.5 Hz, 4H,  $\text{H}_j$ ), 2.81 (t,  $J$  = 7.3 Hz, 4H,  $\text{H}_k$ ), 2.67 (t,  $J$  = 7.2 Hz, 4H,  $\text{H}_c$ ), 2.60 - 2.53 (m, 4H,  $\text{H}_i$ ), 2.06 - 1.96 (m, 4H,  $\text{H}_j$ ), 1.84 - 1.73 (m, 8H,  $\text{H}_{e+d}$ );  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  = 161.5, 157.2, 143.8, 136.6, 134.7, 133.1, 129.8, 129.4, 120.2, 119.4, 114.4, 112.2, 67.5, 38.0, 35.0, 34.8, 32.5, 28.6, 28.1; LRESI(+)-MS (MeOH):  $m/z$  559.9  $[\text{M}+\text{H}]^+$ ; HRFT(+)-MS:  $m/z$  559.3323  $[\text{M}+\text{H}]^+$ , 559.3319 calc. for  $\text{C}_{38}\text{H}_{43}\text{N}_2\text{O}_2$ .

Aldehyde-Functionalised Pyridylmacrocycle **14****14**

To a stirred solution of **25** (0.50 g, 0.90 mmol, 1.00 equiv.) in  $\text{CH}_2\text{Cl}_2$  (8 ml) at 0 °C was added a 1.0 M solution of diisobutylaluminum hydride in hexanes (1.80 ml, 1.80 mmol, 2.00 equiv.) dropwise. The reaction mixture was allowed to warm to room temperature and stirred at room temperature for 12 h. Methanol (2 ml) was added to quench the reaction and the reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (10 ml). The organic layer was washed with  $\text{H}_2\text{O}$  (20 ml). The layers were separated and the aqueous layer was extracted with  $\text{CH}_2\text{Cl}_2$  (2 × 10 ml). The combined organic layers were dried over  $\text{MgSO}_4$ , filtered, and the solvent removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ , petroleum ether/EtOAc 7:3) of the residue afforded **14** (0.36 g, 72%) as a white solid.

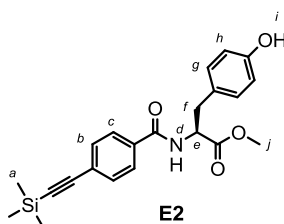
m.p. 99-100 °C;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  = 9.97 (s, 1H,  $\text{H}_a$ ), 7.52 (d,  $J$  = 1.5 Hz, 2H,  $\text{H}_b$ ), 7.48 (t,  $J$  = 7.6 Hz, 1H,  $\text{H}_n$ ), 7.34 (br s, 1H,  $\text{H}_c$ ), 7.05 (d,  $J$  = 8.6 Hz, 4H,  $\text{H}_i$ ), 6.96 (d,  $J$  = 7.6 Hz, 2H,  $\text{H}_m$ ), 6.75 (d,  $J$  = 8.6 Hz, 4H,  $\text{H}_h$ ), 3.94 (t,  $J$  = 5.7 Hz, 4H,  $\text{H}_g$ ), 2.80 (t,  $J$  = 7.4 Hz, 4H,  $\text{H}_l$ ), 2.73 (t,  $J$  = 7.4 Hz, 4H,  $\text{H}_d$ ), 2.59 (t,  $J$  = 8.0 Hz, 4H,  $\text{H}_j$ ), 2.07 – 1.97 (m, 4H,  $\text{H}_k$ ), 1.88 – 1.74 (m, 8H,  $\text{H}_{e+f}$ );  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  = 192.9, 161.5, 157.3, 143.5, 136.9, 136.6, 134.8, 134.7, 129.5, 127.7, 120.2, 114.5, 67.6, 38.0, 35.1, 34.8, 32.5, 28.7, 28.2; LRESI(+)-MS (MeOH):  $m/z$  562.7  $[\text{M}+\text{H}]^+$ ; HRFT(+)-MS:  $m/z$  562.3310  $[\text{M}+\text{H}]^+$ , 562.3316 calc. for  $\text{C}_{38}\text{H}_{44}\text{NO}_3$ .

***N*-(4-(2-(4-Aminophenyl)propan-2-yl)phenyl)pivalamide **17****

To a stirred solution of 4,4'-(propane-2,2-diyl)dianiline (6.78 g, 30.0 mmol, 1.00 equiv.) and Et<sub>3</sub>N (1.67 ml, 12 mmol, 1.20 equiv.) in CH<sub>2</sub>Cl<sub>2</sub> (50 ml) at 0 °C was added pivaloyl chloride (3.69 ml, 36.0 mmol, 1.20 equiv.) dropwise. The reaction mixture was allowed to warm to room temperature and stirred at room temperature for 12 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 6:4) of the residue afforded **17** (5.12 g, 55%) as a pale yellow solid.

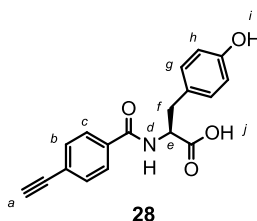
m.p. 42 °C; <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ = 7.41 (d, *J* = 8.7 Hz, 2H, H<sub>f</sub>), 7.27 (s, 1H, H<sub>g</sub>), 7.18 (d, *J* = 8.7 Hz, 2H, H<sub>e</sub>), 7.00 (d, *J* = 8.6 Hz, 2H, H<sub>c</sub>), 6.61 (d, *J* = 8.6 Hz, 2H, H<sub>b</sub>), 5.32 (br s, 2H, H<sub>a</sub>), 1.61 (s, 6H, H<sub>d</sub>), 1.30 (s, 9H, H<sub>h</sub>); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ = 176.6, 147.3, 143.7, 141.3, 135.6, 127.8, 127.4, 119.7, 115.1, 42.0, 39.6, 31.0, 27.8; LRESI(+)-MS (MeOH): *m/z* 311.20 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 311.2124 [M+H]<sup>+</sup>, 311.2118 calc. for C<sub>20</sub>H<sub>27</sub>N<sub>2</sub>O<sub>1</sub>.

**(S)-Methyl-3-(4-hydroxyphenyl)-2-(4-[(trimethylsilyl)ethynyl]benzamido)propanoate**  
**E2**



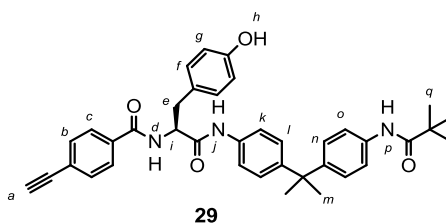
Tyrosine methylester (1.27 g, 6.49 mmol, 1.00 equiv.) and 4-[(trimethylsilyl)ethynyl]benzoic acid (1.70 g, 7.79 mmol, 1.20 equiv.) were dissolved in dry THF/CHCl<sub>3</sub> (2:1 v/v, 60 ml). Then EDCI·HCl (1.49 g, 7.79 mmol, 1.20 equiv.) and HOBt·H<sub>2</sub>O (1.19 g, 7.79 mmol, 1.20 equiv.) were added and the reaction mixture was stirred at room temperature for 12 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 7:3, then 6:4) of the residue afforded **E2** (1.60g, 62%) as an off-white solid.

m.p. 111 °C; [α]<sub>D</sub> = -0.19° (c 1.21, MeOH); <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ = 7.66 (d, 2H, *J* = 8.3 Hz, H<sub>b</sub>), 7.49 (d, 2H, *J* = 8.3 Hz, H<sub>c</sub>), 6.95 (d, 2H, *J* = 8.4 Hz, H<sub>g</sub>), 6.71 (d, 2H, *J* = 8.4 Hz, H<sub>h</sub>), 6.64 (d, 1H, *J* = 7.7 Hz, H<sub>d</sub>), 6.00 (s, 1H, H<sub>i</sub>), 5.04 (dd, 1H, *J* = 13.1 Hz, 5.6 Hz, H<sub>e</sub>), 3.77 (s, 3H, H<sub>j</sub>), 3.20 (dd, 1H, *J* = 14.0 Hz, 5.5 Hz, H<sub>f</sub>), 3.20 (dd, 1H, *J* = 14.0 Hz, 5.5 Hz, H<sub>f</sub>), 0.26 (s, 9H, H<sub>a</sub>); <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ = 172.2, 166.3, 155.2, 133.1, 132.1, 130.4, 127.2, 126.9, 126.8, 115.5, 103.9, 97.3, 53.7, 52.6, 37.1, -0.2; LRESI(+)-MS: *m/z* 396.2 [M+H]<sup>+</sup>; HRESI(+)-MS: *m/z* 418.1447 [M+Na]<sup>+</sup>, 418.1445 calc. for C<sub>22</sub>H<sub>25</sub>O<sub>4</sub>N<sub>1</sub>Na<sub>17</sub>Si<sub>1</sub>.

**(S)-2-(4-Ethynylbenzamido)-3-(4-hydroxyphenyl)propanoic Acid 28**

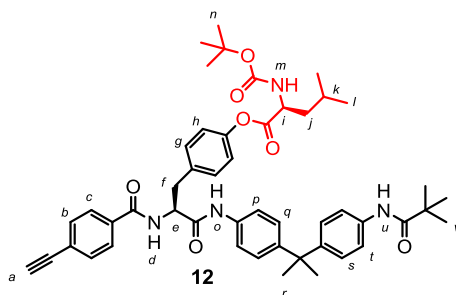
LiOH (120 mg, 5.00 mmol, 5.00 equiv.) was dissolved in THF/H<sub>2</sub>O (4:1 v/v, 12.5 ml). Methyl ester **E2** (395 mg, 1.00 mmol, 1.00 equiv.) was then added and the resulting reaction mixture was stirred at room temperature for 12 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (20 ml) and the pH was adjusted to pH 1 by addition of 1M HCl (20 ml). The solution was extracted with CH<sub>2</sub>Cl<sub>2</sub> (3 × 30 ml). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure to afford **28** (310 mg, 99%) as a white solid. *Please note: Racemisation of the chiral centre occurs during the reaction.*

m.p. 175 °C; <sup>1</sup>H NMR (600 MHz, acetonitrile-*d*<sub>3</sub>) δ = 7.84 (d, *J* = 8.1 Hz, 2H, H<sub>c</sub>), 7.81 (d, *J* = 8.2 Hz, 1H, H<sub>d</sub>), 7.55 (d, *J* = 8.3 Hz, 2H, H<sub>b</sub>), 7.16 (d, *J* = 8.3 Hz, 2H, H<sub>g</sub>), 6.75 (d, *J* = 8.4 Hz, 2H, H<sub>h</sub>), 4.84 (td, *J* = 8.6 Hz, 5.3 Hz, 1H, H<sub>e</sub>), 3.82 (s, 1H, H<sub>a</sub>), 3.24 (dd, *J* = 14.0 Hz, 4.9 Hz, 1H, H<sub>f</sub>), 3.07 (dd, *J* = 14.0 Hz, 9.1 Hz, 1H, H<sub>f</sub>); <sup>13</sup>C NMR (150 MHz, acetonitrile-*d*<sub>3</sub>) δ = 173.5, 167.8, 156.6, 134.6, 132.9, 131.3, 128.9, 128.3, 126.2, 116.1, 83.3, 81.3, 55.3, 36.8; LRESI(+)-MS: *m/z* 308.3 [M-H]<sup>-</sup>; HRESI(+)-MS: *m/z* 332.0915 [M+Na]<sup>+</sup>, 332.0899 calc. for C<sub>18</sub>H<sub>15</sub>O<sub>4</sub>N<sub>1</sub>Na<sub>1</sub>.

**Ethynyl-Tyrosine-*N*-Pivaloyl Derivative 29**

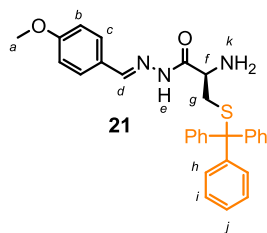
Tyrosine derivative **28** (309 mg, 1.00 mmol, 1.00 equiv.) and amine **17** (342 mg, 1.10 mmol, 1.10 equiv.) were dissolved in THF/CHCl<sub>3</sub> (2:1 v/v, 12 ml). Then, EDCI·HCl (200 mg, 1.05 mmol, 1.05 equiv.) and HOBt·H<sub>2</sub>O (161 mg, 1.05 mmol, 1.05 equiv.) were added and the reaction mixture was stirred at room temperature for 12 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 6:4) of the residue afforded **29** (445 mg, 74%) as a colourless solid.

m.p. 183 °C; <sup>1</sup>H NMR (400 MHz, acetone-*d*<sub>6</sub>) δ = 9.34 (s, 1H, H<sub>j</sub>), 8.53 (s, 1H, H<sub>p</sub>), 7.98 - 7.90 (m, 1H, H<sub>d</sub>), 7.91 - 7.81 (m, 2H, H<sub>c</sub>), 7.58 (d, *J* = 8.8 Hz, 2H, H<sub>k</sub>), 7.54 (d, *J* = 8.3 Hz, 2H, H<sub>b</sub>), 7.52 (d, *J* = 8.5 Hz, 2H, H<sub>o</sub>), 7.20 - 7.11 (m, 6H, H<sub>n+l+f</sub>), 6.73 (d, *J* = 8.5 Hz, 2H, H<sub>g</sub>), 4.92 (td, *J* = 8.2 Hz, 6.0 Hz, 1H, H<sub>i</sub>), 3.81 (s, 1H, H<sub>a</sub>), 3.23 (dd, *J* = 13.9 Hz, 5.9 Hz, 1H, H<sub>e</sub>), 3.07 (dd, *J* = 13.9 Hz, 8.4 Hz, 1H, H<sub>e</sub>), 1.63 (s, 6H, H<sub>m</sub>), 1.27 (s, 9H, H<sub>q</sub>); <sup>13</sup>C NMR (125 MHz, acetone-*d*<sub>6</sub>) δ = 179.9, 172.1, 169.2, 157.4, 148.2, 148.1, 137.3, 136.8, 135.3, 133.0, 131.4, 128.9, 128.6, 128.1, 128.0, 127.1, 122.5, 121.4, 116.3, 83.6, 81.1, 57.7, 43.3, 40.4, 38.5, 31.2, 27.8; LRESI(+)-MS (MeOH): *m/z* 602.3 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 602.3011 [M+H]<sup>+</sup>, 602.3019 calc. for C<sub>38</sub>H<sub>41</sub>N<sub>3</sub>O<sub>4</sub>.

Ethynyl-*N*-Boc-Leucine-*N*-Pivaloyl Barrier **12**

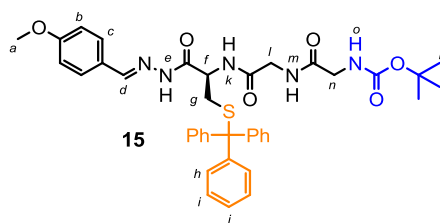
Phenol derivative **29** (100 mg, 166  $\mu\text{mol}$ , 1.00 equiv.), L-*N*-Boc-leucine (46.3 mg, 200  $\mu\text{mol}$ , 1.20 equiv.), EDCI·HCl (38.3 mg, 192  $\mu\text{mol}$ , 1.15 equiv.), DCC (41.3 mg, 206  $\mu\text{mol}$ , 1.24 equiv.) and DMAP (3 mg, 33  $\mu\text{mol}$ , 0.20 equiv.) were stirred in  $\text{CH}_2\text{Cl}_2$  (3 ml) at room temperature for 48 h. The solvent was removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ , petroleum ether/EtOAc 6:4) of the residue afforded **12** (70.0 mg, 52%) as a white solid.

m.p. 120-122  $^\circ\text{C}$ ;  $^1\text{H}$  NMR (600 MHz,  $\text{CDCl}_3$ )  $\delta$  = 7.96 (s, 1H,  $\text{H}_o$ ), 7.72 (d,  $J$  = 7.6 Hz, 2H,  $\text{H}_c$ ), 7.55 (d,  $J$  = 8.0 Hz, 2H,  $\text{H}_b$ ), 7.43 (d,  $J$  = 8.6 Hz, 2H,  $\text{H}_g$ ), 7.32 (d,  $J$  = 8.0 Hz, 3H,  $\text{H}_{t+u}$ ), 7.26 – 7.31 (m, 2H,  $\text{H}_p$ ), 7.17 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_h$ ), 7.15 (d,  $J$  = 8.0 Hz, 2H,  $\text{H}_q$ ), 7.09 (d,  $J$  = 7.3 Hz, 1H,  $\text{H}_d$ ), 7.06 (d,  $J$  = 8.0 Hz, 2H,  $\text{H}_s$ ), 5.05 (d,  $J$  = 8.5 Hz, 1H,  $\text{H}_m$ ), 5.01 – 4.94 (m, 1H,  $\text{H}_e$ ), 4.57 – 4.50 (m, 1H,  $\text{H}_i$ ), 3.31 (dd,  $J$  = 13.9 Hz, 6.1 Hz, 1H,  $\text{H}_j$ ), 3.20 – 3.25 (m, 2H,  $\text{H}_{a+f}$ ), 1.88 – 1.76 (m, 2H,  $\text{H}_j$ ), 1.71 – 1.66 (m, 1H,  $\text{H}_k$ ), 1.64 (s, 6H,  $\text{H}_r$ ), 1.47 (s, 9H,  $\text{H}_n$ ), 1.33 (s, 9H,  $\text{H}_v$ ), 1.03 (d,  $J$  = 6.2 Hz, 6H,  $\text{H}_l$ );  $^{13}\text{C}$  NMR (150 MHz,  $\text{CDCl}_3$ )  $\delta$  = 176.7, 172.4, 168.8, 166.8, 155.7, 149.8, 147.4, 146.6, 135.7, 134.7, 134.4, 133.5, 132.5, 130.6, 127.5, 127.4, 127.2, 126.0, 122.0, 120.0, 119.8, 80.2, 80.0, 55.9, 52.5, 42.5, 41.6, 39.7, 37.9, 30.9, 28.5, 27.8, 25.1, 23.1, 22.0; LRESI(+)-MS (MeOH):  $m/z$  837.4  $[\text{M}+\text{Na}]^+$ ; HRESI(+)-MS:  $m/z$  851.3913  $[\text{M}+\text{Na}]^+$ , 851.3923 calc. for  $\text{C}_{47}\text{H}_6\text{O}_6\text{N}_1\text{Na}_1\text{Si}_1$ .

**(R)-2-Amino-N'-(4-methoxybenzylidene)-3-(tritylthio)propanehydrazide 21**

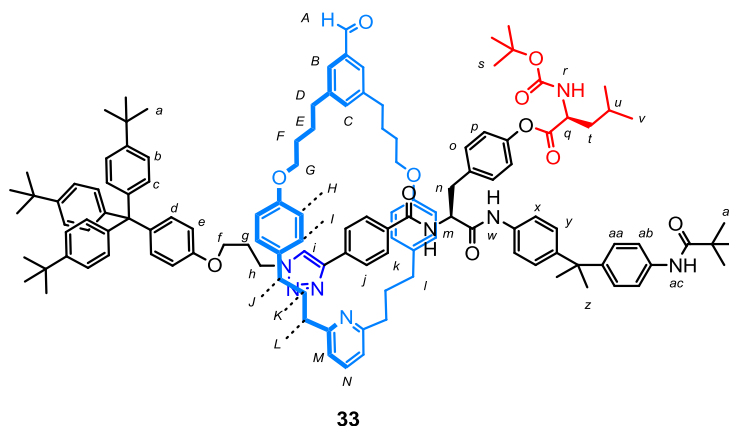
To a stirred solution of Fmoc-Cys(STrt)-OPfp (1.00 g, 1.33 mmol, 1.00 equiv.) in MeOH (20 ml) at room temperature was added hydrazine monohydrate (47  $\mu$ l, 1.46 mmol, 1.10 equiv.). After 2 h the white precipitate that formed was filtered and dried under vacuum. To a suspension of this crude hydrazide (750 mg, 1.25 mmol, 1.00 equiv.) in MeOH (5 ml) at room temperature was added *p*-anisaldehyde (0.17 ml, 1.38 mmol, 1.10 equiv.) and acetic acid (15  $\mu$ L, 0.25 mmol, 0.20 equiv.) and the resulting suspension was stirred at room temperature for 12 h. The solvent was removed under reduced pressure and the residue dissolved in CH<sub>2</sub>Cl<sub>2</sub> (10 ml). The organic layer was washed with H<sub>2</sub>O (10 ml) and dried over MgSO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. The crude hydrazone (753 mg, 1.05 mmol) was then dissolved in a mixture of DMF/piperidine (4:1 v/v, 5 ml) and stirred at room temperature for 2 h before the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 1:1) of the residue afforded **21** (494 mg, 75% over 3 steps) as a pale yellow oil. *Please Note: The compound is unstable and should be used within a week after preparation.*

$[\alpha]_D^{25} = +127^\circ$  (c 0.13, MeOH); <sup>1</sup>H NMR (500 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 9.35 (s, 1H, H<sub>e</sub>), 7.70 (d,  $J$  = 8.8 Hz, 2H, H<sub>c</sub>), 7.48 – 7.43 (m, 6H, H<sub>h</sub>), 7.3 – 7.30 (m, 7H, H<sub>d+i</sub>), 7.28 – 7.24 (m, 3H, H<sub>j</sub>), 6.99 (d,  $J$  = 8.8 Hz, 2H, H<sub>b</sub>), 3.84 (s, 3H, H<sub>a</sub>), 3.52 (dd,  $J$  = 7.7 Hz, 3.9 Hz, 1H, H<sub>f</sub>), 2.80 – 2.73 (m, 1H, H<sub>g</sub>), 2.56 (dd,  $J$  = 15.8 Hz, 6.1 Hz, 1H, H<sub>g</sub>); <sup>13</sup>C NMR (125 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 169.7, 162.7, 154.1, 145.8, 130.5, 130.5, 129.7, 127.7, 127.7, 115.1, 67.4, 55.8, 32.3, 23.3; LRESI(+)-MS (MeOH):  $m/z$  496.2 [M+H]<sup>+</sup>.

**N-Boc-Gly-Gly-Cys(STrt)-4-methoxybenzylidene Acyl Hydrazone 15**

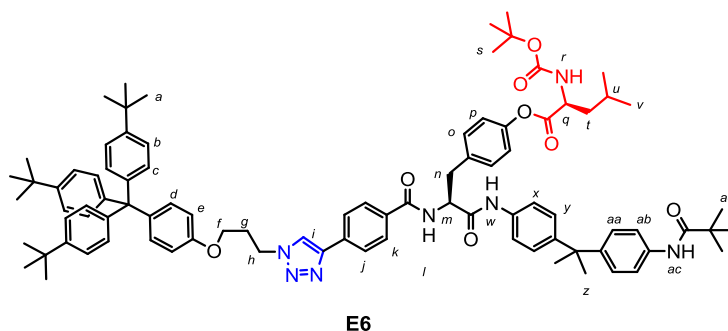
*N*-Boc-Gly-Gly-OH (975 mg, 4.20 mmol, 1.50 equiv.) was suspended in anhydrous THF (40 ml) and 1,1'-carbodiimidazole (680 mg, 4.20 mmol, 1.50 equiv.) was added in small portions. The reaction mixture was allowed to stir at room temperature for 15 min and amine **E3** (1.23 g, 2.80 mmol, 1.00 equiv.) was added. The reaction mixture was allowed to stir at room temperature for a further 12 h. The resulting white precipitate was filtered and washed with Et<sub>2</sub>O (2 × 100 ml) and dissolved in *i*PrOH/CHCl<sub>3</sub> (1:3 v/v, 100 ml). The organic layer was washed with aqueous 1M KHSO<sub>4</sub> (100 ml), aqueous saturated K<sub>2</sub>CO<sub>3</sub> (100 ml) and brine (50 ml). The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent removed under reduced pressure to afford **15** (920 mg, 1.32 mmol, 47%) as a white powder.

10:9 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone. m.p. = 172 - 174 °C; [α]<sub>D</sub> = -133° (c 0.12, MeOH); <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) δ = 11.35 (s, 1H, H<sub>e1</sub>), 11.31 (s, 1H, H<sub>e2</sub>), 8.38 (d, *J* = 8.2 Hz, 1H, H<sub>k1</sub>), 8.33 (d, *J* = 8.6 Hz, 1H, H<sub>k2</sub>), 8.19 (s, 1H, H<sub>d2</sub>), 8.10 (t, *J* = 5.7 Hz, 1H, H<sub>m1</sub>), 7.96 (t, *J* = 5.7 Hz, 1H, H<sub>m2</sub>), 7.88 (s, 1H, H<sub>d1</sub>), 7.64 (d, *J* = 8.3 Hz, 2H, H<sub>c2</sub>), 7.55 (d, *J* = 8.3 Hz, 2H, H<sub>c1</sub>), 7.38 – 7.16 (m, 31H, H<sub>h1+h2+i1+i2+j1+j2+o1</sub>), 7.05 (t, *J* = 6.1 Hz, 1H, H<sub>o2</sub>), 7.04 – 6.99 (m, 4H, H<sub>b1+b2</sub>), 5.38 (td, *J* = 8.4 Hz, 4.8 Hz, 1H, H<sub>f2</sub>), 4.46 (q, *J* = 7.5 Hz, 1H, H<sub>fl</sub>), 3.83 (s, 3H, H<sub>a1</sub>), 3.81 (s, 3H, H<sub>a2</sub>), 3.79 – 3.69 (m, 4H, H<sub>l1+l2</sub>), 3.60 - 3.56 (m, 4H, H<sub>n1+n2</sub>), 2.48 – 2.31 (m, 4H, H<sub>g1+g2</sub>), 1.39 (s, 9H, H<sub>p1</sub>), 1.36 (s, 9H, H<sub>p2</sub>); <sup>13</sup>C NMR (150 MHz, DMSO-*d*<sub>6</sub>) δ = 171.1, 170.5, 170.0, 169.2, 169.0, 166.1, 161.4, 161.2, 156.3, 156.2, 148.3, 144.6, 144.3, 129.5, 129.4, 129.3, 129.0, 128.6, 128.5, 127.3, 127.2, 127.0, 127.0, 114.8, 66.5, 66.3, 55.8, 51.4, 48.7, 43.7, 43.6, 42.4, 42.1, 34.1, 33.9, 28.7, 28.6. Please Note: Due to the presence of two stereo isomers some <sup>13</sup>C-signals coincide. LRESI(+)-MS: *m/z* 710.3 [M+ Na]<sup>+</sup>; HRFT(+)-MS: *m/z* 710.2990 [M+H]<sup>+</sup>, 710.3012 calc. for C<sub>39</sub>H<sub>44</sub>N<sub>5</sub>O<sub>6</sub>S<sub>1</sub>.

Aldehyde *N*-Boc-Leucine Single-Barrier Rotaxane **33**

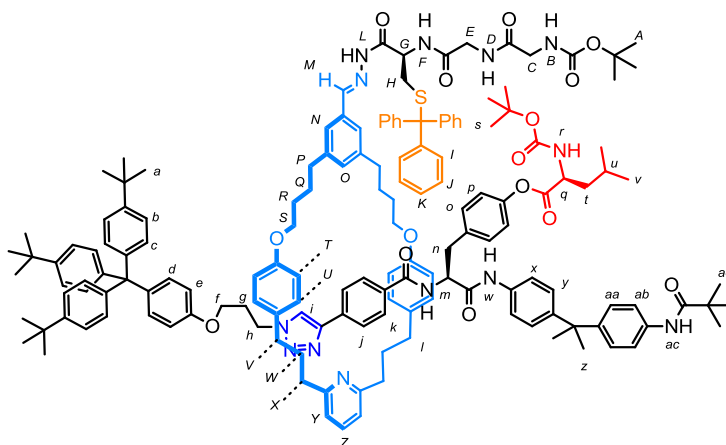
Macrocyclic **14** (55 mg, 98  $\mu\text{mol}$ , 2.00 equiv.) and  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (9.1 mg, 24.5  $\mu\text{mol}$ , 0.50 equiv.) were stirred in degassed  $\text{CH}_2\text{Cl}_2$  (3 ml) at room temperature for 30 minutes. Then, a solution of alkyne **12** (40 mg, 49  $\mu\text{mol}$ , 1.00 equiv.) and azide **13** (58 mg, 98  $\mu\text{mol}$ , 2.00 equiv.) in degassed  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v, 3 ml) was added. The reaction mixture was stirred at room temperature for 8 h. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (20 ml) and the organic layer was washed with aqueous 0.1M  $\text{Na}_4\text{EDTA}$  solution ( $3 \times 20$  ml). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [Merck],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) of the crude residue afforded **33** (43 mg, 45%) as a colourless film.

$^1\text{H}$  NMR (600 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 9.96 (s, 1H, H<sub>A</sub>), 9.47 (s, 1H, H<sub>w</sub>), 8.57 (s, 1H, H<sub>ac</sub>), 8.03 (s, 1H, H<sub>i</sub>), 8.00 (d,  $J$  = 8.3 Hz, 1H, H<sub>l</sub>), 7.86 (d,  $J$  = 8.0 Hz, 2H, H<sub>j</sub>), 7.82 (d,  $J$  = 8.0 Hz, 2H, H<sub>k</sub>), 7.63 – 7.53 (m, 6H, H<sub>ab+x+c</sub>), 7.43 (d,  $J$  = 8.0 Hz, 2H, H<sub>o</sub>), 7.32 (d,  $J$  = 7.9, 6H, H<sub>b</sub>), 7.26 (t,  $J$  = 7.1 Hz, 1H, H<sub>N</sub>), 7.21 – 7.11 (m, 12H, H<sub>c+y+aa+M</sub>), 7.08 – 7.02 (m, 4H, H<sub>p+B</sub>), 6.96 (d, 2H, H<sub>d</sub>), 6.78 (d,  $J$  = 7.9 Hz, 4H, H<sub>l</sub>), 6.50 – 6.43 (m, 6H, H<sub>e+H</sub>), 5.04 (q,  $J$  = 7.4 Hz, 1H, H<sub>r</sub>), 4.42 – 4.35 (m, 1H, H<sub>m</sub>), 4.07 (t,  $J$  = 7.0 Hz, 2H, H<sub>h</sub>), 3.84 – 3.75 (m, 4H, H<sub>G</sub>), 3.49 (t,  $J$  = 6.1 Hz, 2H, H<sub>f</sub>), 3.37 (dd,  $J$  = 14.0 Hz, 5.9 Hz, 1H, H<sub>n</sub>), 3.20 (dd,  $J$  = 14.0 Hz, 8.5 Hz, 1H, H<sub>n'</sub>), 2.80 – 2.70 (m, 4H, H<sub>D</sub>), 2.65 – 2.59 (m, 4H, H<sub>L</sub>), 2.56 – 2.50 (m, 4H, H<sub>J</sub>), 1.92 – 1.71 (m, 15H, H<sub>t+u+E+F+k</sub>), 1.65 (s, 6H, H<sub>z</sub>), 1.42 (s, 9H, H<sub>s</sub>), 1.32 (s, 27H, H<sub>a</sub>), 1.29 (s, 9H, H<sub>ad</sub>), 0.99 (t,  $J$  = 7.6 Hz, 6H, H<sub>v</sub>);  $^{13}\text{C}$  NMR (150 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 193.1, 176.9, 172.7, 170.30, 167.1, 162.5, 158.0, 157.2, 156.6, 150.6, 149.0, 147.0, 146.9, 146.4, 145.3, 144.7, 140.1, 138.1, 137.9, 137.6, 137.2, 136.4, 135.6, 135.2, 134.2, 132.6, 131.4, 130.0, 129.0, 128.7, 127.9, 127.9, 127.8, 127.5, 125.9, 125.0, 122.2, 120.6, 120.4, 120.0, 114.9, 114.1, 79.4, 67.9, 64.8, 63.8, 53.3, 47.5, 42.7, 41.0, 40.1, 38.4, 35.7, 35.6, 34.9, 32.9, 31.7, 31.1, 30.4, 29.0, 28.6, 27.8, 25.6, 23.3, 21.8; LRESI(+)-MS:  $m/z$  1965.3 [M+H]<sup>+</sup>.

**N-Boc-Leucine Thread E6**

Title compound **E6** (14 mg, 35%) was isolated as a side product from the reaction mixture for the synthesis of **33** as a colourless film after preparative TLC (SiO<sub>2</sub> [Merck], CH<sub>2</sub>Cl<sub>2</sub>/MeOH 2%).

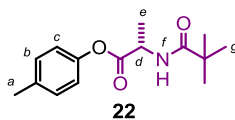
<sup>1</sup>H NMR (600 MHz, acetone-*d*<sub>6</sub>) δ = 9.47 (s, 1H, H<sub>w</sub>), 8.55 (s, 1H, H<sub>ac</sub>), 8.48 (s, 1H, H<sub>l</sub>), 8.04 (d, *J* = 8.1 Hz, 1H, H<sub>l</sub>), 7.93 (s, 4H, H<sub>j+k</sub>), 7.58 (d, *J* = 8.1 Hz, 2H, H<sub>ab</sub>), 7.54 (d, *J* = 8.2 Hz, 2H, H<sub>x</sub>), 7.41 (d, *J* = 8.0 Hz, 2H, H<sub>o</sub>), 7.31 (d, *J* = 7.4 Hz, 6H, H<sub>b</sub>), 7.16 (d, *J* = 8.2 Hz, 2H, H<sub>aa</sub>), 7.14 – 7.10 (m, 8H, H<sub>c+y</sub>), 7.08 (d, *J* = 7.8 Hz, 2H, H<sub>d</sub>), 7.02 (d, *J* = 8.0 Hz, 2H, H<sub>p</sub>), 6.84 (d, *J* = 8.2 Hz, 2H, H<sub>e</sub>), 6.46 (d, *J* = 8.2 Hz, 1H, H<sub>r</sub>), 5.02 (q, *J* = 7.6 Hz, 1H, H<sub>m</sub>), 4.69 (t, *J* = 6.9 Hz, 2H, H<sub>h</sub>), 4.39 – 4.32 (m, 1H, H<sub>q</sub>), 4.07 (t, *J* = 5.9 Hz, 2H, H<sub>f</sub>), 3.36 (dd, *J* = 14.0 Hz, 5.7 Hz, 1H, H<sub>n</sub>), 3.20 (dd, *J* = 13.9 Hz, 8.7 Hz, 1H, H<sub>n'</sub>), 2.44 (p, *J* = 6.2 Hz, 2H, H<sub>g</sub>), 1.88 – 1.80 (m, 1H, H<sub>u</sub>), 1.80 – 1.69 (m, 2H, H<sub>t</sub>), 1.39 (s, 9H, H<sub>s</sub>), 1.29 (s, 27H, H<sub>a</sub>), 1.26 (s, 9H, H<sub>ad</sub>), 0.97 (t, *J* = 7.4 Hz, 6H, H<sub>v</sub>); <sup>13</sup>C NMR (150 MHz, acetone-*d*<sub>6</sub>) δ = 177.0, 172.7, 170.5, 170.4, 167.2, 167.1, 157.6, 156.7, 150.6, 149.1, 147.0, 147.0, 146.4, 145.2, 140.5, 138.1, 137.3, 136.4, 135.2, 134.1, 132.8, 131.4, 131.2, 128.9, 127.8, 127.5, 125.9, 125.1, 122.5, 122.2, 120.4, 120.1, 114.1, 79.4, 63.9, 56.8, 53.4, 47.9, 41.0, 37.9, 34.8, 31.6, 31.1, 28.5, 27.8, 23.3, 21.8; LRESI(+)-MS: *m/z* 1424.7 [M+Na]<sup>+</sup>.

***N*-Boc-Gly-Gly-Cys(STrt)-Acyl Hydrazone *N*-Boc-Leucine Single-Barrier Rotaxane 34****34**

A mixture of rotaxane **33** (20 mg, 10.8  $\mu\text{mol}$ , 1.00 equiv.), *p*- $\text{CH}_3\text{OC}_6\text{H}_4\text{CH}=\text{NNH-Cys(STrt)-Gly-Gly-N-Boc}$  **15** (20 mg, 28.2  $\mu\text{mol}$ , 2.60 equiv.) and aniline (50  $\mu\text{l}$ ) in DMSO/2-(*N*-morpholino)ethanesulfonic acid buffer (3:1 v/v, 1 ml, pH 6.0) was stirred at 60  $^\circ\text{C}$  for 72 h. The reaction mixture was allowed to cool to room temperature and then diluted with  $\text{CH}_2\text{Cl}_2$  (10 ml) and the organic layer was washed with  $\text{H}_2\text{O}$  (10 ml). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [UNIPLATE-T],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%, 3 elutions) of the residue afforded **34** (15 mg, 55%) as a colourless solid.

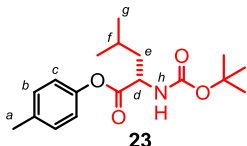
1:1 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone.  $^1\text{H}$  NMR (600 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 10.58 – 10.36 (m, 2H,  $\text{H}_{L1+L2}$ ), 9.62 – 9.46 (m, 2H,  $\text{H}_{w1+w2}$ ), 8.60 – 8.52 (m, 2H,  $\text{H}_{ac1+ac2}$ ), 8.11 (d,  $J$  = 8.2 Hz, 1H,  $\text{H}_{D1}$ ), 8.03 – 7.98 (m, 1H,  $\text{H}_{D2}$ ), 7.95 – 7.90 (m, 2H,  $\text{H}_{M1+M2}$ ), 7.89 – 7.82 (m, 4H,  $\text{H}_{j1+j2}$ ), 7.80 – 7.70 (m, 6H,  $\text{H}_{k1+k2+i1+i2}$ ), 7.62 – 7.48 (m, 12H,  $\text{H}_{o1+o2+n1+n2+o1+o2+z1+z2}$ ), 7.45 – 7.33 (m, 16H,  $\text{H}_{K1+K2+x1+x2+ab1+ab2+F1+F2}$ ), 7.34 – 7.27 (m, 24H,  $\text{H}_{b1+b2+l1+l2}$ ), 7.23 – 7.11 (m, 32H,  $\text{H}_{y1+y2+aa1+aa2+j1+j2+c1+c2}$ ), 7.05 – 6.98 (m, 6H,  $\text{H}_{p1+p2+y1+y2}$ ), 6.99 – 6.93 (m, 4H,  $\text{H}_{d1+d2}$ ), 6.80 – 6.73 (m, 8H,  $\text{H}_{U1+U2}$ ), 6.53 (t,  $J$  = 5.6 Hz, 1H,  $\text{H}_{B1}$ ), 6.50 – 6.40 (m, 12H,  $\text{H}_{r1+r2+e1+e2+T1+T2}$ ), 6.31 – 6.25 (m, 1H,  $\text{H}_{B2}$ ), 5.57 – 5.48 (m, 1H,  $\text{H}_{G1}$  or  $\text{G2}$ ), 5.07 – 4.97 (m, 2H,  $\text{H}_{m1+m2}$ ), 4.57 – 4.50 (m, 1H,  $\text{H}_{G1}$  or  $\text{G2}$ ), 4.39 – 4.32 (m, 2H,  $\text{H}_{q1+q2}$ ), 4.07 – 3.96 (m, 5H,  $\text{H}_{E1}$  or  $\text{E2}$ ,  $h1+h2$ ), 3.92 (d,  $J$  = 5.4 Hz, 1H,  $\text{H}_{E1}$  or  $\text{E2}$ ), 3.82 – 3.72 (m, 8H,  $\text{H}_{C1+C2+s1+s2}$ ), 3.43 – 3.37 (m, 4H,  $\text{H}_{f1+f2}$ ), 3.34 (dd,  $J$  = 14.0 Hz, 6.0 Hz, 2H,  $\text{H}_{n1+n2}$ ), 3.19 (dt,  $J$  = 13.6 Hz, 8.9 Hz, 2H,  $\text{H}_{n1+n2}$ ), 2.80 – 2.69 (m, 4H,  $\text{H}_{H1+H2}$ ), 2.66 – 2.54 (m, 16H,  $\text{H}_{X1+X2+p1+p2}$ ), 2.54 – 2.49 (m, 8H,  $\text{H}_{V1+V2}$ ), 1.91 – 1.79 (m, 16H,  $\text{H}_{W1+W2+r1+r2}$ ), 1.80 – 1.75 (m, 4H,  $\text{H}_{g1+g2}$ ), 1.75 – 1.67 (m, 8H,  $\text{H}_{Q1+Q2}$ ), 1.62 (s, 12H,  $\text{H}_{z1+z2}$ ), 1.40 – 1.38 (m, 36H,  $\text{H}_{A1+A2+s1+s2}$ ), 1.37 – 1.33 (m, 6H,  $\text{H}_{u1+u2+t1+t2}$ ), 1.29 (s, 27H,  $\text{H}_{a1}$  or  $a2$ ), 1.29 (s, 27H,  $\text{H}_{a1}$  or  $a2$ ), 1.26 (s, 18H,  $\text{H}_{ad1+ad2}$ ), 0.97 (d,  $J$  = 6.5 Hz, 6H,  $\text{H}_{v1+v2}$ ), 0.95 (d,  $J$  = 6.6 Hz, 6H,  $\text{H}_{v1+v2}$ );

$^{13}\text{C}$  NMR (150 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 177.0, 172.7, 171.4, 170.6, 170.5, 170.5, 169.8, 169.3, 167.3, 167.3, 166.8, 162.5, 162.4, 158.0, 158.0, 157.2, 150.6, 149.6, 149.1, 149.0, 147.1, 147.1, 146.8, 146.8, 146.4, 145.6, 145.5, 145.3, 144.1, 144.1, 144.0, 144.0, 140.1, 138.1, 137.6, 137.5, 137.3, 136.4, 135.3, 135.1, 135.1, 135.0, 134.2, 133.8, 132.6, 131.4, 131.3, 131.2, 130.4, 130.3, 130.0, 128.9, 128.8, 127.8, 127.8, 127.6, 127.5, 126.0, 125.1, 122.2, 120.6, 120.6, 120.4, 120.2, 120.2, 115.0, 115.0, 114.2, 68.0, 68.0, 64.8, 64.8, 63.9, 56.9, 56.8, 55.8, 55.7, 53.4, 53.0, 50.2, 47.6, 47.6, 45.0, 44.7, 44.1, 43.2, 42.7, 41.3, 41.0, 38.4, 38.4, 37.9, 37.9, 36.1, 36.0, 36.0, 35.6, 35.6, 34.9, 32.9, 32.9, 31.7, 31.1, 30.3, 29.2, 28.7, 28.6, 28.6, 27.8, 25.6, 23.3, 21.8; LRESI(+)-MS:  $m/z$  2536.9 [M+Na]<sup>+</sup>.

**(R)-*p*-Tolyl 2-pivalamidopropanoate **22****

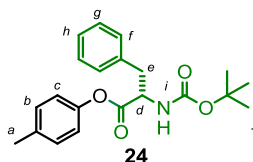
To a solution of *N*-Piv-alanine (500 mg, 2.89 mmol, 2.00 equiv.) in anhydrous THF (10 ml) at room temperature was added *N,N'*-carbonyldiimidazole (469 mg, 2.89mmol, 2.00 equiv.) in small portions and the reaction mixture was stirred at room temperature for 1 h. To the reaction mixture *p*-cresol (156 mg, 1.45 mmol, 1.00 equiv.) was added and stirring was continued at 50 °C for 24 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (50 ml) and the organic layer was washed with aqueous 1M NaOH (2 × 25ml), aqueous 1M KHSO<sub>4</sub> (2 × 25ml) and brine (25 ml). The organic layer was dried over MgSO<sub>4</sub>, filtered, and the solvent removed under reduced pressure to afford **22** (197 mg, 52%) as a white solid.

m.p. 86 – 87 °C;  $[\alpha]_D = +6.7^\circ$  (c 0.42, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>)  $\delta$  = 7.19 (d,  $J$  = 7.9 Hz, 2H, H<sub>b</sub>), 6.99 (d,  $J$  = 8.0, 2H, H<sub>c</sub>), 6.26 (br d,  $J$  = 7.6 Hz, 1H, H<sub>f</sub>), 4.81 (m, 1H, H<sub>d</sub>), 2.36 (s, 3H, H<sub>a</sub>), 1.59 (d,  $J$  = 7.3 Hz, 3H, H<sub>e</sub>), 1.25 (s, 9H, H<sub>g</sub>); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>)  $\delta$  = 178.3, 172.4, 148.3, 135.9, 130.1, 121.0, 48.3, 38.7, 27.5, 21.0, 18.4; LRESI(+)-MS (MeOH):  $m/z$  264.1 [M+H]<sup>+</sup>; HRESI(+)-MS:  $m/z$  286.1407 [M+Na]<sup>+</sup>, 286.1419 calc. for C<sub>15</sub>H<sub>21</sub>N<sub>1</sub>O<sub>3</sub>Na<sub>1</sub>.

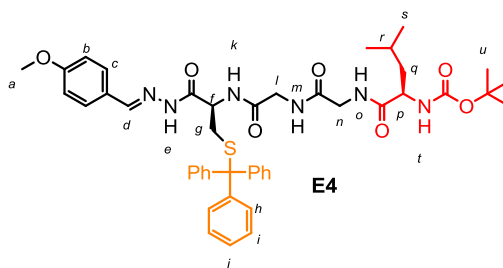
**(R)-*p*-Tolyl 2-([*tert*-butoxycarbonyl]amino)-4-methylpentanoate **23****

To a solution of *N*-Boc-leucine (2.00 g, 6.23 mmol, 2.00 equiv.) in anhydrous THF (30 ml) at room temperature was added *N,N'*-carbonyldiimidazole (1.01 g, 6.23 mmol, 2.00 equiv.) in small portions and the reaction mixture was stirred at room temperature for 1 h. To the reaction mixture was added *p*-cresol (337 mg, 3.12 mmol, 1.00 equiv.) and stirring was continued at 50 °C for 24 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (100 ml) and the organic layer was washed with aqueous 1M NaOH (2 × 50 ml), aqueous 1M KHSO<sub>4</sub> (2 × 50 ml) and brine (50 ml). The organic layer was dried over MgSO<sub>4</sub>, filtered, and the solvent removed under reduced pressure to afford **23** (1.24 g, 62%) as a white solid.

m.p. 52 °C;  $[\alpha]_D = -24.2^\circ$  (c 0.91, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>)  $\delta = 7.18$  (d,  $J = 9.6$  Hz, 2H, H<sub>b</sub>), 6.99 (d,  $J = 9.6$  Hz, 2H, H<sub>c</sub>), 5.15 (d,  $J = 10.3$  Hz, 1H, H<sub>h</sub>), 4.56 (q,  $J = 9.3$  Hz, 1H, H<sub>d</sub>), 2.35 (s, 3H, H<sub>a</sub>), 1.92-1.77 (m, 2H, H<sub>e</sub>), 1.76 – 1.64 (m, 1H, H<sub>f</sub>), 1.49 (s, 9H, H<sub>i</sub>), 1.03 (d,  $J = 7.9$  Hz, 6H, H<sub>g</sub>); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>)  $\delta = 172.4, 155.5, 148.3, 135.6, 129.91, 121.0, 52.3, 41.5, 28.3, 24.9, 22.9, 21.9, 20.8$ .; LRESI(+)-MS (MeOH):  $m/z$  322.2 [M+H]<sup>+</sup>; HRESI(+)-MS:  $m/z$  344.1832 [M+Na]<sup>+</sup>, 344.1832 calc. for C<sub>18</sub>H<sub>27</sub>N<sub>1</sub>Na<sub>1</sub>.

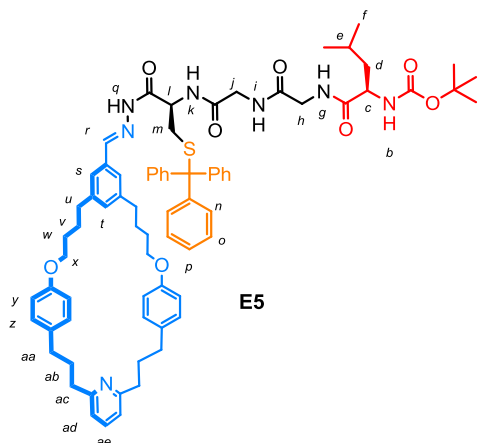
**(R)-*p*-Tolyl 2-([*tert*-butoxycarbonyl]amino)-3-phenylpropanoate **24****

To a solution of *N*-Boc-phenylalanine (1.00 g, 3.77 mmol, 2.00 equiv.) in anhydrous THF (25 ml) at room temperature was added *N,N'*-carbonyldiimidazole (0.61 g, 3.77 mmol, 2.00 equiv.) in small portions and the reaction mixture was stirred at room temperature for 1 h. To the reaction mixture *p*-cresol (204 mg, 1.89 mmol, 1.00 equiv.) was added and stirring was continued at 50 °C for 24 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (100 ml) and the organic layer was washed with aqueous 1M NaOH (2 × 50 ml), aqueous 1M KHSO<sub>4</sub> (2 × 50 ml) and brine (50 ml). The organic layer was dried over MgSO<sub>4</sub>, filtered, and the solvent removed under reduced pressure to afford **24** (389 mg, 58%) as a white solid. m.p. 93-95 °C; [α]<sub>D</sub> = -8.9° (c 0.45, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ = 7.38 (t, *J* = 7.4 Hz, 2H, H<sub>g</sub>), 7.32 (t, *J* = 7.2 Hz, 1H, H<sub>h</sub>), 7.29 (d, *J* = 7.3 Hz, 2H, H<sub>f</sub>), 7.19 (d, *J* = 8.0 Hz, 2H, H<sub>b</sub>), 6.91 (d, *J* = 8.0 Hz, 2H, H<sub>c</sub>), 5.15 (d, *J* = 8.2 Hz, 1H, H<sub>i</sub>), 4.89 – 4.83 (m, 1H, H<sub>d</sub>), 3.28 (br d, *J* = 6.1 Hz, 2H, H<sub>e</sub>), 2.37 (s, 3H, H<sub>a</sub>), 1.49 (s, 9H, H<sub>j</sub>); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ = 170.8, 155.2, 148.1, 135.8, 130.0, 129.5, 128.7, 127.2, 121.0, 115.2, 54.6, 38.4, 28.4, 20.9; LRESI(+)-MS (MeOH): *m/z* 378.3 [M+Na]<sup>+</sup>; HRESI(+)-MS: *m/z* 378.1678 [M+Na]<sup>+</sup>, 378.1676 calc. for C<sub>21</sub>H<sub>25</sub>N<sub>1</sub>Na<sub>1</sub>.

***N*-Boc-Leu-Gly-Gly-Cys(STrt)-4-methoxybenzylidene Acyl Hydrazone **E4****

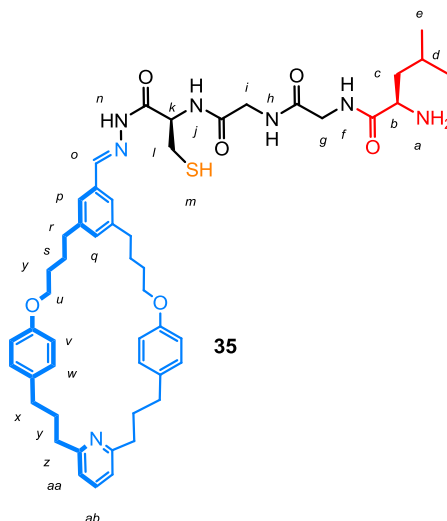
Amine **E3** (30.0 mg, 49  $\mu\text{mol}$ , 1.00 equiv.) was dissolved in DMF (2 ml) and DIPEA (0.10 ml), and L-*N*-Boc-leucine (17.0 mg, 73.5  $\mu\text{mol}$ , 1.50 equiv.) and PyBrop (34.0 mg, 73.5  $\mu\text{mol}$ , 2.00 equiv.) were added. The reaction mixture was allowed to stir at room temperature for 24 h and the reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (20 ml). The organic layer was washed with  $\text{H}_2\text{O}$  (20 ml), brine (20 ml) and the organic layer was dried over  $\text{MgSO}_4$ , filtered, and the solvent removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [UNIPLATE-T],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  3%) gave hydrazone **E4** (19.5 mg, 47%) as colourless oil.

1:1 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone. m.p. decomp. over 217°;  $[\alpha]_{\text{D}} = +0.05^\circ$  (c 0.2, MeOH);  $^1\text{H NMR}$  (600 MHz,  $\text{DMSO-}d_6$ )  $\delta = 11.36 - 11.30$  (br s, 2H,  $\text{H}_{e1}$ ), 8.43 – 8.26 (m, 1H,  $\text{H}_{k2}$ ), 8.16 (s, 1H,  $\text{H}_{dl}$ ), 8.13 – 7.96 (m, 4H,  $\text{H}_{m1+m2+o1+o2}$ ), 7.63 (d,  $J = 8.4$  Hz, 1H,  $\text{H}_{c1}$ ), 7.54 (d,  $J = 8.4$  Hz, 1H,  $\text{H}_{c2}$ ), 7.39 – 7.15 (m, 30H,  $\text{H}_{i1+i2+j1+j2+h1+h2}$ ), 7.04 – 6.94 (m, 4H,  $\text{H}_{b1+b2+l1+l2}$ ), 5.38 (q,  $J = 7.4$  Hz, 1H,  $\text{H}_{f1}$ ), 4.44 (q,  $J = 7.6$  Hz, 1H,  $\text{H}_{j2}$ ), 4.01 – 3.92 (m, 2H,  $\text{H}_{p1+p2}$ ), 3.83 (s, 3H,  $\text{H}_{a1}$  or  $a_2$ ), 3.83 (s, 3H,  $\text{H}_{a1}$  or  $a_2$ ), 3.78 – 3.69 (m, 4H,  $\text{H}_{l1+l2+n1+n2}$ ), 2.46 – 2.30 (m, 4H,  $\text{H}_{g1+g2}$ ), 1.59 (q,  $J = 9.0$  Hz, 2H,  $\text{H}_{r1+r2}$ ), 1.47 – 1.38 (m, 2H,  $\text{H}_{q1+q2}$ ), 1.37 (d,  $J = 5.4$  Hz, 18H,  $\text{H}_{u1+u2}$ ), 0.92 – 0.80 (m, 12H,  $\text{H}_{s1+s2}$ );  $^{13}\text{C NMR}$  (150 MHz,  $\text{DMSO-}d_6$ )  $\delta = 170.7, 169.4, 160.9, 160.8, 155.5, 147.8, 144.2, 143.8, 129.1, 128.8, 128.5, 128.1, 128.0, 126.9, 126.7, 126.5, 114.3, 55.4, 52.8, 51.0, 48.2, 42.1, 42.0, 41.8, 41.6, 40.6, 40.1, 39.9, 33.7, 33.4, 28.2, 24.2, 23.0, 21.5$ ; LRESI(+)-MS (MeOH):  $m/z$  845.4  $[\text{M}+\text{Na}]^+$ ; HRESI(+)-MS:  $m/z$  845.3668  $[\text{M}+\text{Na}]^+$ , 845.3672 calc. for  $\text{C}_{45}\text{H}_{54}\text{N}_6\text{O}_7\text{SNa}_1$ .

**N-Boc-Leu-Gly-Gly-Cys(STrt)-Acyl Hydrazone Macrocycle E5**

A mixture of macrocycle **14** (10 mg, 18  $\mu\text{mol}$ , 1.0 equiv.), *p*-CH<sub>3</sub>OC<sub>6</sub>H<sub>4</sub>CH=NNH-Cys(STrt)-Gly-Gly-Leu-*N*-Boc (20 mg, 26  $\mu\text{mol}$ , 1.5 equiv.) and aniline (50  $\mu\text{l}$ ) in DMSO/2-(*N*-morpholino)ethanesulfonic acid buffer (3:1 *v/v*, 0.5 ml, pH 6.0) was stirred at 60 °C for 72 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (20 ml). The organic layer was washed with H<sub>2</sub>O (2  $\times$  20 ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent removed under reduced pressure. Preparative TLC (SiO<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>/MeOH 4%) of the crude residue afforded **E5** (14 mg, 75%) as a colourless film.

2:3 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone.  $[\alpha]_{\text{D}} = +0.02^{\circ}$  (c 0.10, MeOH); <sup>1</sup>H NMR (600 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 10.40 (s, 1H, H<sub>q1</sub>), 10.38 (s, 1H, H<sub>q2</sub>), 8.36 (s, 1H, H<sub>r2</sub>), 8.23 (t, *J* = 5.8 Hz, 1H, H<sub>il</sub>), 8.21 – 8.15 (m, 1H, H<sub>i2</sub>), 8.03 (s, 1H, H<sub>r1</sub>), 7.81 – 7.73 (m, 2H, H<sub>g1+g2</sub>), 7.58 (t, *J* = 7.7 Hz, 2H, H<sub>ae1+ae2</sub>), 7.47 – 7.13 (m, 36H, H<sub>n1+n2+o1+o2+p1+p2+s1+s2+t1+t2</sub>), 7.10 – 7.05 (m, 12H, H<sub>z1+z2+ad1+ad2</sub>), 6.79 – 6.73 (m, 4H, H<sub>y1+y2</sub>), 6.35 (d, *J* = 6.7 Hz, 1H, H<sub>b1</sub>), 6.30 (d, *J* = 7.3 Hz, 2H, H<sub>b2</sub>), 5.59 – 5.53 (m, 1H, H<sub>il</sub>), 4.39 – 4.32 (m, 1H, H<sub>i2</sub>), 4.18 – 4.09 (m, 2H, H<sub>c1+c2</sub>), 4.02 – 3.81 (m, 14H, H<sub>x1+x2+j1+h1+h2</sub>), 3.64 – 3.59 (m, 1H, H<sub>j2</sub>), 2.78 (t, *J* = 7.4 Hz, 8H, H<sub>ac1+ac2</sub>), 2.76 – 2.70 (m, 6H, H<sub>u1+u2+m1+m2</sub>), 2.59 – 2.55 (m, 2H, H<sub>m'1+m'2</sub>), 2.06 – 2.00 (m, 8H, H<sub>ab1+ab2</sub>), 1.88 – 1.81 (m, 8H, H<sub>v1+v2</sub>), 1.79 – 1.74 (m, 8H, H<sub>w1+w2</sub>), 1.69 – 1.55 (m, 6H, H<sub>d1+d2+e1+e2</sub>), 1.44 – 1.39 (m, 18H, H<sub>a1+a2</sub>), 0.93 – 0.85 (m, 12H, H<sub>f1+f2</sub>); <sup>13</sup>C NMR (150 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 174.7, 171.2, 168.8, 168.3, 165.7, 161.1, 161.0, 157.4, 157.3, 148.1, 144.8, 144.7, 144.5, 143.0, 142.8, 136.4, 134.8, 134.3, 130.5, 130.3, 129.6, 129.6, 129.5, 129.4, 129.2, 128.4, 128.0, 127.9, 127.9, 126.7, 126.7, 126.6, 125.1, 125.1, 120.0, 120.0, 120.0, 114.1, 67.3, 53.7, 52.7, 43.6, 43.3, 42.6, 40.5, 40.5, 34.8, 34.7, 34.2, 28.7, 28.3, 28.0, 27.7, 24.6, 22.5, 21.1, 13.5; LRESI(+)-MS (MeOH): *m/z* 1248.5 [M+H]<sup>+</sup>.

**N-Boc-Gly-Gly-Cys(SH)-Acyl Hydrazone Macrocycle 35**

**E5** (15 mg, 9.4  $\mu\text{mol}$ , 1.0 equiv.) was dissolved in  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (4:1 v/v, 1 ml) and triisopropylsilane (5.0  $\mu\text{l}$ , 24  $\mu\text{mol}$ , 2.6 equiv.) was added. The mixture was stirred at room temperature for 90 minutes. The solvent was removed under reduced pressure and the residue washed with  $\text{Et}_2\text{O}$  ( $3 \times 10\text{ml}$ ) and analysed without further purification. LRESI-MS (MeOH):  $m/z$  906.6  $[\text{M}+\text{H}]^+$ ;  $m/z$  952.6  $[\text{M}+\text{COOH}+\text{H}]^+$ ,  $m/z$  985.4  $[\text{M}+\text{COOH}+\text{Na}]^+$ .

## 2.6 References

- (1) Voet, D.; Voet, J. G. *Biochemistry*; fourth edition; John Wiley & Sons: New York, 2011.
- (2) A full overview over different approaches to sequence-specific synthesis is given in Chapter 1. Two examples for molecular machines capable of sequence-specific synthesis are: (a) He, Y.; Liu, D. R. *Nat. Nanotechnol.* **2010**, *5*, 778 (b) Gu, H.; Chao, J.; Xiao, S.-J.; Seeman, N. C. *Nature* **2010**, *465*, 202.
- (3) Fernandes, A., Synthetic Molecular Nanodevices for Selective Peptide-Based Therapy, PhD Thesis, University of Edinburgh, 2009.
- (4) (a) Aucagne, V.; Berná, J.; Crowley, J. D.; Goldup, S. M.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Ronaldson, V. E.; Slawin, A. M. Z.; Viterisi, A.; Walker, D. B. *J. Am. Chem. Soc.* **2007**, *129*, 11950 (b) Aucagne, V.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Walker, D. B. *J. Am. Chem. Soc.* **2006**, *128*, 2186.
- (5) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science* **1994**, *266*, 776.
- (6) (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776 (b) Kent, S. B. H. *Chem. Soc. Rev.* **2009**, *38*, 338.
- (7) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923.
- (8) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10068.
- (9) (a) Katritzky, A. R.; Abo-Dya, N. E.; Tala, S. R.; Abdel-Samii, Z. K. *Org. Biomol. Chem.* **2010**, *8*, 2316 (b) Katritzky, A. R.; Tala, S. R.; Abo-Dya, N. E.; Ibrahim, T. S.; El-Feky, S. A.; Gyanda, K.; Pandya, K. M. *J. Org. Chem.* **2011**, *76*, 85 (c) Payne, R. J.; Ficht, S.; Tang, S.; Brik, A.; Yang, Y.-Y.; Case, D. A.; Wong, C.-H. *J. Am. Chem. Soc.* **2007**, *129*, 13527.
- (10) Bol'shakov, O.; Kovacs, J.; Chahar, M.; Ha, K.; Khelashvili, L.; Katritzky, A. R. *J. Pept. Sci.* **2012**, *18*, 704.
- (11) Wan, Q.; Danishefsky, S. J. *Angew. Chem. Int. Ed.* **2007**, *46*, 9248.
- (12) (a) Barrell, M. J.; Campaña, A. G.; von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Angew. Chem. Int. Chem.* **2011**, *123*, 299 (b) von Delius, M.; Geertsema, E. M.; Leigh, D. A. *Nat. Chem.* **2010**, *2*, 96 (c) von Delius, M.; Geertsema, E. M.; Leigh, D. A.; Tang, D.-T. D. *J. Am. Chem. Soc.* **2010**, *132*, 16134.
- (13) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. *Nat. Chem.* **2010**, *2*, 490.

- (14) Zheng, H.; Zhou, W.; Lv, J.; Yin, X.; Li, Y.; Liu, H.; Li, Y. *Chem. – Eur. J.* **2009**, *15*, 13253.
- (15) Getz, E. B.; Xiao, M.; Chakrabarty, T.; Cooke, R.; Selvin, P. R. *Anal. Biochem.* **1999**, *273*, 73.

## Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine

### ACKNOWLEDGMENTS

All reactions presented in this Chapter with exception of the operation of molecular machine **18** (carried out by Dr. Bartosz Lewandowski) were conducted by Marcus Pappmeyer. Dr. Bartosz Lewandowski, Dr. Guillaume De Bo, Sona Kuschel and Dr. John W. Ward are acknowledged for their contributions to the development of synthetic routes and continuous resynthesis of building blocks during the course of the project. Dr. Philipp M. E. Gramlich is gratefully acknowledged for his idea to incorporate Native Chemical Ligation into a molecular machine. Dr. María J. Aldegunde, Dr. Daniel M. D'Souza, Anthony E. Fernandes, Dr. Stephen M. Goldup, Dr. Philipp M. E. Gramlich and Dr. Dominik Heckmann are gratefully acknowledged for work on previous designs that ultimately lead to the development of the system that is presented in this Chapter.

### 3.1 Synopsis

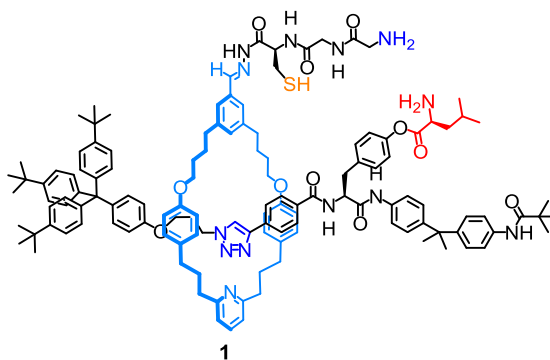
The ribosome builds proteins by joining together amino acids in an order determined by messenger RNA. The design, synthesis and operation of the first molecular small-molecule machine that travels along a molecular strand, picking up amino acids that block its path, to synthesise a peptide in a sequence-specific manner is described in this Chapter. The chemical structure is based on a rotaxane, a molecular ring threaded onto a molecular axle. The ring carries a thiolate group that iteratively removes three amino acids in order from the strand and transfers them to a peptide elongation site through native chemical ligation. The synthesis is demonstrated using  $\sim 10^{18}$  molecular machines acting in parallel and generates milligram quantities of a peptide with a single sequence confirmed by tandem mass spectrometry.

## 3.2 Introduction

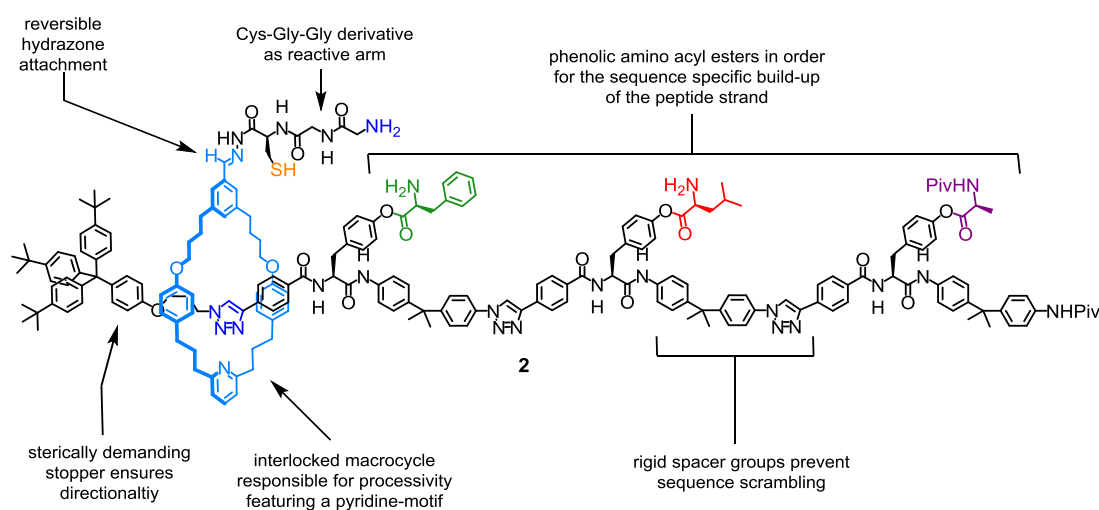
Artificial small-molecule machines<sup>1</sup> have previously been used to store information<sup>2</sup>, to do mechanical work<sup>3</sup>, and employed in synthesis to processively epoxidise an unsaturated polymer<sup>4</sup>, switch ‘on’ and ‘off’ catalytic activity<sup>5</sup> and to change the handedness of a reaction product<sup>6</sup>. Large synthetic DNA molecules have been used to guide the formation of bonds between unnatural building blocks<sup>7</sup> and assemble gold nanoparticles in particular sequences<sup>8</sup>. The sequence-specific synthesis of a peptide consisting of naturally occurring amino acid building blocks has not been realised so far.

Here, the design, synthesis and operation of a rotaxane-based small-molecule machine in which a functionalised macrocycle operates on a thread containing building blocks in a pre-determined order to achieve sequence-specific peptide synthesis is reported. The design of the artificial molecular machine is based on several elements that have analogues in either ribosomal<sup>9</sup> or nonribosomal<sup>10</sup> protein synthesis: reactive building blocks (the role played by tRNA-bound amino acids) are delivered in a sequence determined by a molecular strand (the role played by mRNA). A macrocycle ensures processivity during the machine’s operation (reminiscent of the way that subunits of the ribosome clamp the mRNA strand) and bears a catalyst—a tethered thiolate group—that detaches the amino acid building blocks from the strand and passes them on to another site at which the resulting peptide oligomer is elongated in a single specific sequence, through chemistry reminiscent of nonribosomal peptide synthesis<sup>10</sup>.

The operation principles and essential design features are based on the work outlined in Chapter 2. As iterative chemical transformation required for the assembly of three aminoacyl units in sequence, native chemical ligation was chosen. NCL has proven to be high yielding, chemoselective, can proceed through large transition states and the suitability of a cysteine-glycylglycine reactive arm for the incorporation in a rotaxane has been established in the successful operation of model rotaxane **1**<sup>11</sup> (Figure 3.1). The chemical structure of molecular machine **2** featuring three amino acyl residues in sequence is shown in Figure 3.2. Analogous to the test-system **1**, molecular machine **2** is envisioned to be built-up from a 30-membered macrocycle incorporating a pyridyl unit required for the assembly of the rotaxane based on Cu(I)-catalysed azide-alkyne cycloaddition (CuAAC) active-metal template strategy and a reversible hydrazone attachment point for the reactive arm. The track features a sterically demanding stopper unit ensuring that the macrocycle can only diffuse in the direction of the amino acyl residues which are connected *via* cleavable phenolic ester linkages.



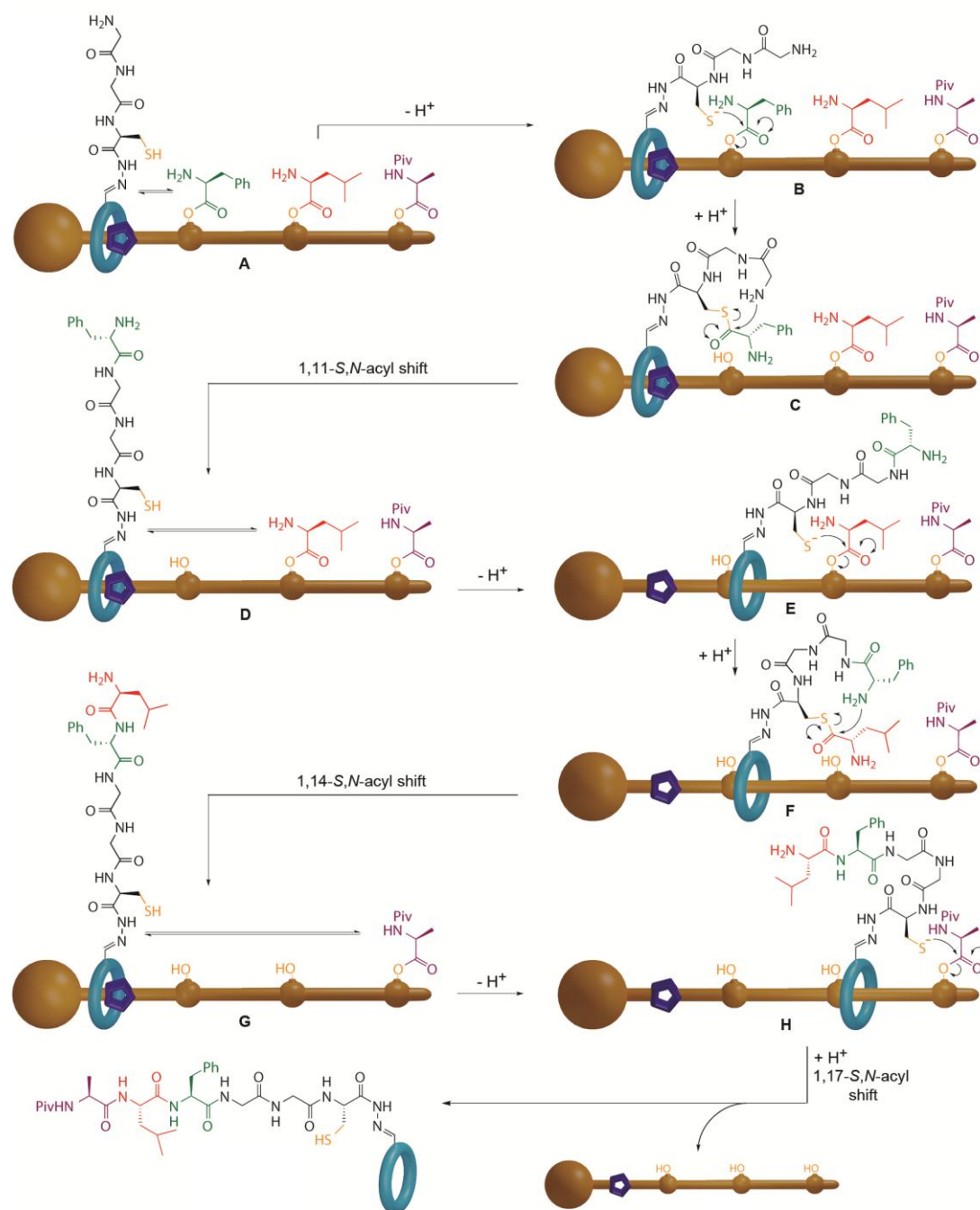
**Figure 3.1.** Rotaxane design of a test substrate for the development of a second generation molecular machine capable of sequence-specific synthesis.



**Figure 3.2.** Second generation design of a molecular machine capable of sequence-specific synthesis.

Relatively rigid 2,2-diphenylpropane-based spacer units prevent the reactive arm of the macrocycle from overreaching and reacting with monomers out of sequence. The mechanically interlocked nature of the machine ensures processivity and prevents the reactive arm from dissociation from the track.

The suggested operation mechanism of machine **2** is shown in Scheme 3.1. After global deprotection of the corresponding machine in its protected form, machine **2** is allowed to operate in the presence of base which deprotonates the thiolate catalyst (Scheme 3.1, A). The *in-situ* formed thiolate functionality of the catalytic arm is poised to cleave the ester bond of the first amino acid phenolic ester that blocks the path of the macrocycle (Scheme 3.1, B). An *S,N*-acyl transfer *via* a 11-membered transition state transfers the amino acyl residue to the glycyglycine amine group (Scheme 3.1, C).



**Scheme 3.1.** Proposed mechanism for sequence-specific peptide synthesis by molecular machine 2. After activation of the machine by acidic cleavage of the Boc and Trt protecting groups under basic conditions, successive native chemical ligation reactions transfer the amino acid building blocks to the peptide elongation site on the macrocycle in the order that they appear on the thread. Once the final amino acid is cleaved, the macrocycle bearing the synthesised oligopeptide dethreads from the strand.

The thiolate catalyst on the reactive arm is regenerated (Scheme 3.1, D) and the macrocycle—now able to diffuse over the first attachment point—comes in close proximity to the second aminoacyl residue (Scheme 3.1, E). Formation of the thioester and subsequent *S,N*-transfer *via* a 14-membered cyclic transition state (Scheme 3.1, F) allows the elongation of the peptide chain by another unit (Scheme 3.1, G). In its final step the thiolate catalyst picks up the third amino acyl residue (Scheme 3.1, H) and dissociates from the thread with the hexapeptide attached. Hydrolysis of the hydrazone linkage would allow the isolation of the newly formed, full length peptide. The artificial molecular machine synthesises the peptide from the *C*- to the *N*-terminus, the opposite direction to ribosomal translation and opposite to the first generation machine presented in Chapter 2.

The features exhibited by this machine are<sup>12</sup>:

- (i) **Processive Operation:** the reactive arm covalently attached to a interlocked macrocycle prevents dissociation until the operation is completed.
- (ii) **Directional Movement:** the macrocycle, although no bias is applied and the motion is purely diffusional, moves along the molecular strand away from the bulky stopping group.
- (iii) **Repetitive Elongation:** after each pick-up of an amino acyl moiety the catalytic thiolate unit is regenerated without undoing the task that was performed.
- (iv) **Autonomous Operation:** once the protective groups are removed and the molecular machine is submitted to operation conditions, no interference is required and the hexapeptide is synthesised autonomously.

Those four key characteristics underpin the similarity of the molecular machine to the ribosome, which features processivity<sup>13</sup> (the molecular clamp made up from the 30S and 50S subunits does not dissociate from the mRNA strand before a stop codon is encountered), directional movement<sup>14</sup> (the ribosome moves along the mRNA strand in the 5'-to-3' direction), repetitive elongation<sup>15</sup> (up to 10000 amino acids are joined together through a single operation) and, of course, autonomous operation. The fundamental difference between both machines—leaving the vastly different levels of performance aside—is the inability of the artificial system to reuse the information encoding strand the oligomer is translated from. The information contained is destroyed as the macrocycle moves along whereas a ribosome, after termination, can again synthesise peptides using the same or a different mRNA-template.

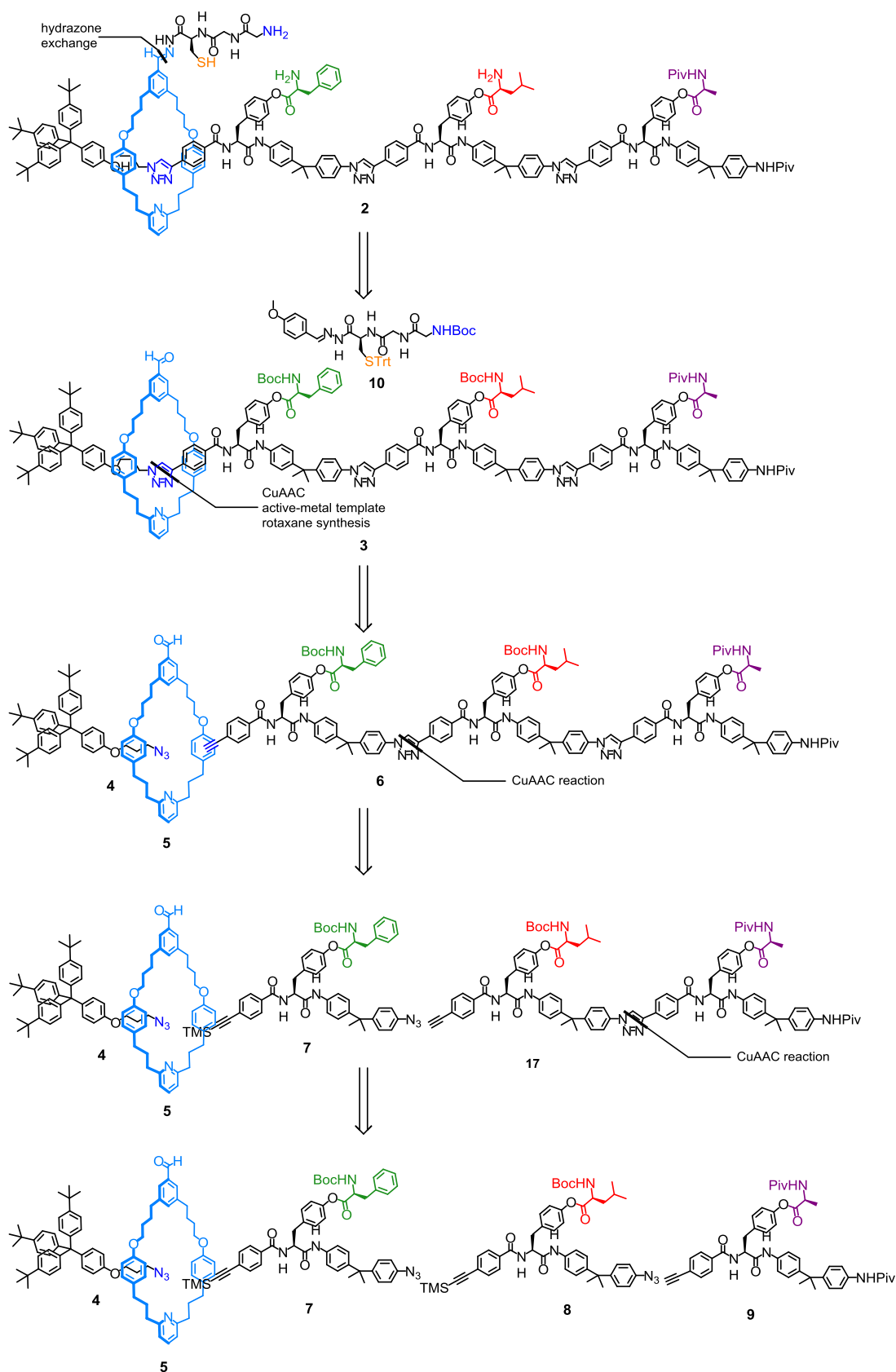
### 3.2.1 Retrosynthetic Analysis of Three-Barrier Rotaxane 2

The importance of a fast and efficient synthetic route to a highly complex molecular machine cannot be overemphasised and is crucial for the success of the whole project, since often late-stage modifications are necessary in order to overcome unforeseen problems. Therefore, a short and convergent route, as well as a modular built-up that allows modification of key components in the end game of the synthesis (e.g. change of amino acids or the reactive arm), are equally important.

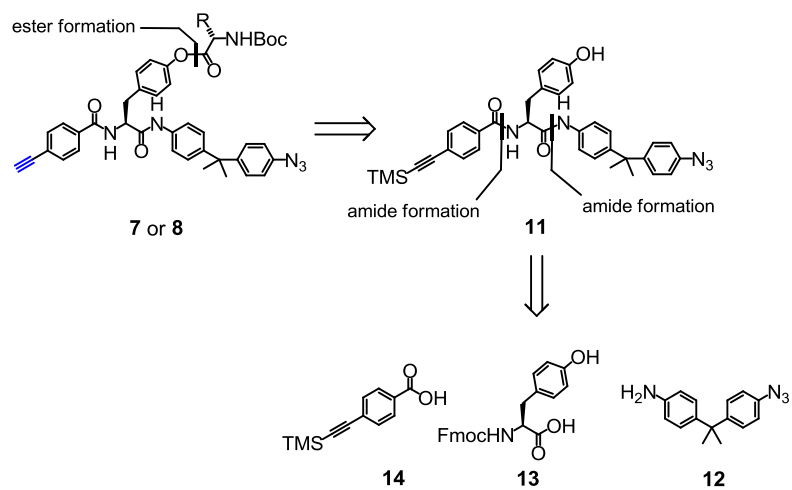
The retrosynthetic analysis of molecular machine **2** is shown in Scheme 3.2. The reactive arm containing the tethered thiolate catalyst can be attached *via* a mild and selective hydrazone exchange reaction starting from aldehyde functionalised rotaxane **3**, potentially allowing the screening of different catalytic units in the last step of the synthetic route. As protecting groups, acid labile Boc-groups for the amine-functionality and a trityl group for the thiolate masking allow a global deprotection in a single step. Rotaxane **3** can be derived from the CuAAC active metal template reaction of azide functionalised stopper **4**, 2,6-substituted pyridyl macrocycle **5** and amino acid-loaded track **6** containing a terminal alkyne functionality. This axle is derived from successive CuAAC “click” reactions of monomers featuring an azide group and a terminal trimethylsilyl (TMS)-protected alkyne from building blocks **7**, **8** and **9**. Interlocking the components at a late stage in the synthesis allows the change of amino acids or the length of the track with relative ease.

Crucial for the success of this elongation/deprotection strategy required for the built-up of thread **6** are conditions that remove the TMS-protecting group without cleaving the loaded amino acyl residues. One particular attractive set of TMS-deprotection conditions makes use of silver(I)nitrate and allows the mild removal of TMS in the presence of base labile groups.<sup>16</sup>

The synthesis of the remaining building blocks is straightforward. The synthesis of azide stopper **4** is reported in literature<sup>17</sup>, the syntheses of hydrazone exchange reagent **10** and macrocycle **5** is described in Chapter 2.<sup>11</sup> External barrier **9** can be synthesised analogous to *N*-Boc-Leu-derivative outlined in the previous Chapter. Internal barrier **7** and **8** can be derived from phenol **11** *via* standard ester formation. Two amide disconnections leave monofunctionalised 2,2-diphenylpropane unit **12**, Fmoc-tyrosine **13** and TMS-protected *p*-acetylene benzoic acid **14** as starting materials.



**Scheme 3.2.** Retrosynthetic analysis of molecular machine 2.

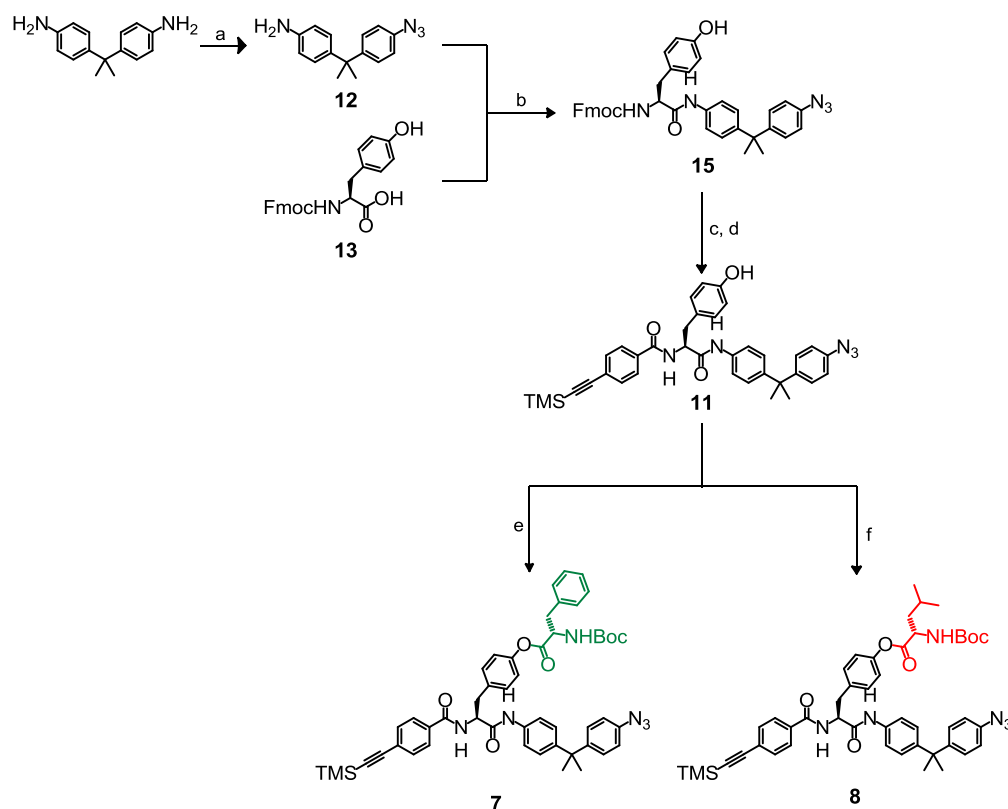


**Scheme 3.3.** Retrosynthetic analysis of building block **7** and **8**.

### 3.3 Results and Discussion

#### 3.3.1 Synthesis of Molecular Machine 2

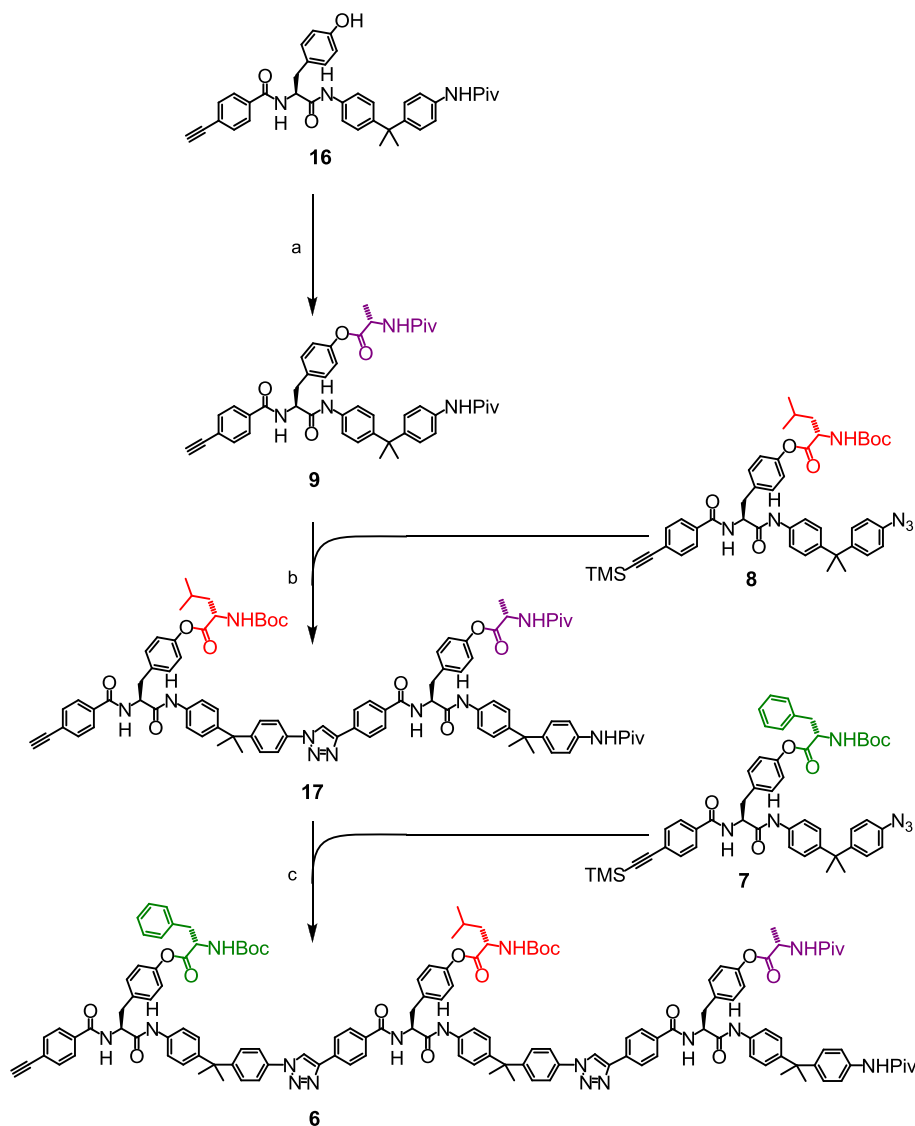
For the synthesis of molecular machine **2** key intermediates are macrocycle **5**, external barrier loaded with pivaloyl-protected alanine **9**, internal barriers loaded with boc-protected phenyl alanine **7** and loaded with boc-protected leucine **8** as well as the hydrazone exchange reagent **10** required for loading on the reactive arm. The synthesis of internal barriers **7** and **8** differs in one key aspect from the route to the related external barrier **9**: the TMS-protecting group has to remain on the alkyne at the end of the synthesis. Therefore the synthesis commences with the amide coupling using HOBt/EDCI of Fmoc-protected tyrosine **13** with *p*-azidophenyl derivative **12** to give amide **15** (Scheme 3.4). After Fmoc-deprotection an amide coupling yields TMS-protected acetylene **11**. Coupling with the desired amino acid gives access to **7** and **8**.



**Scheme 3.4.** Synthesis of barriers **7** and **8**. Reaction conditions: a) *iso*-amyl nitrite,  $\text{NaN}_3$ , AcOH, RT, 2 h, 45%; b) L-*N*-Fmoc-tyrosine, HOBt, EDCI·HCl, THF/ $\text{CHCl}_3$  (2:1 v/v), 12 h, 60%; c) DMF/piperidine (4:1 v/v), RT, 3 h, 97%; d) 4-((trimethylsilyl)ethynyl)benzoic acid, HOBt, EDCI·HCl, THF/ $\text{CHCl}_3$  (2:1 v/v), RT, 12 h, 71%; e) L-*N*-Boc-phenylalanine, EDCI·HCl, DCC, DMAP,  $\text{CH}_2\text{Cl}_2$ , RT, 48 h, 42%; f) L-*N*-Boc-leucine, EDCI·HCl, DCC, DMAP,  $\text{CH}_2\text{Cl}_2$ , RT, 48 h, 63%.

External Barrier **9** loaded with *N*-Piv-alanine is derived from an esterification of the corresponding carboxylic acid with phenol **16**.

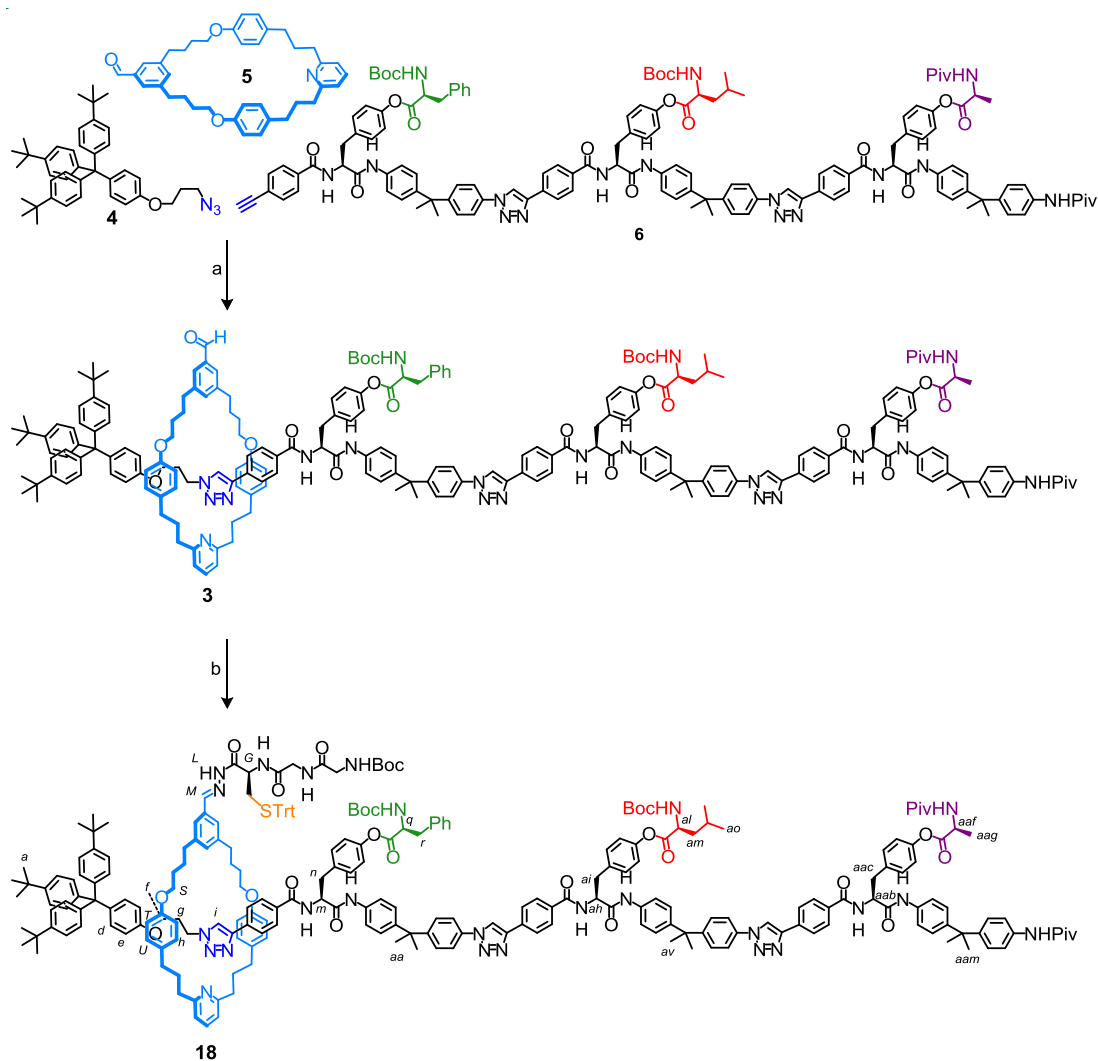
The synthesis of alkyne-thread **5** starts with a CuAAC reaction and subsequent TMS-deprotection to give **17**. Polymer-bound tris([1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl)amine (Tentagel TBTA) gave best results for the azide-alkyne cycloaddition since it allowed the removal of otherwise difficult to separate TBTA by a simple filtration (Scheme 3.5).



**Scheme 3.5.** Synthesis of three-barrier alkyne **6**. Reaction conditions: a) *L*-*N*-Piv-alanine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 56%; b) Tentagel TBTA, Cu(MeCN)<sub>4</sub>PF<sub>6</sub>, CH<sub>2</sub>Cl<sub>2</sub>/*t*BuOH (1:1 v/v), RT, 48 h, then AgNO<sub>3</sub>, acetone/H<sub>2</sub>O (10:1 v/v), RT, 18 h, 43% over two steps; c) Tentagel TBTA, Cu(MeCN)<sub>4</sub>PF<sub>6</sub>, CH<sub>2</sub>Cl<sub>2</sub>/*t*BuOH (1:1 v/v), RT, 48 h, then AgNO<sub>3</sub>, acetone/H<sub>2</sub>O (10:1 v/v), RT, 18 h, 37% over two steps.

Base-free deprotection conditions employing silver(I)nitrate have been chosen in order to circumvent saponification of the ester bonds that would be expected from basic hydrolysis which is commonly used for TMS-removal. Similarly, tetra-*n*-butylammonium fluoride (TBAF) as strongly nucleophilic reagent was not compatible with the substrate and cleavage of the esters was observed, whereas deprotection with Ag(I)NO<sub>3</sub> in acetone delivers alkyne **17** in good yield, which is then extended by a second CuAAC reaction. Again, a base-free deprotection of the terminal TMS-group allowed access to terminal alkyne **6**.

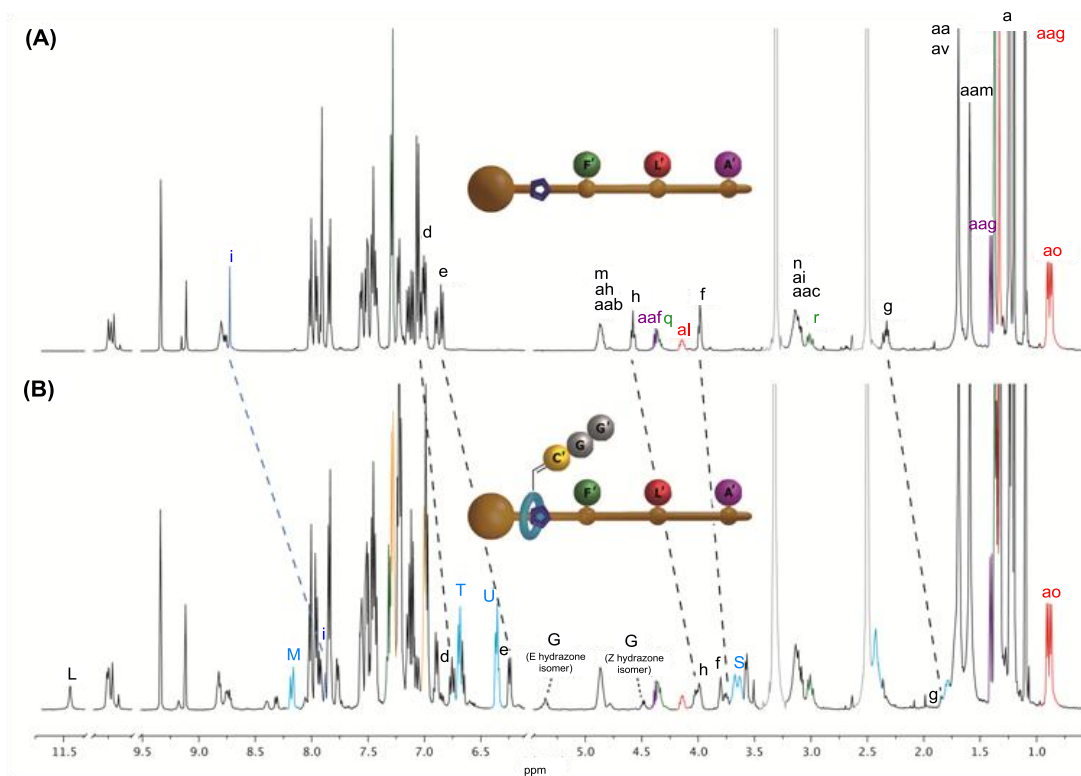
The crucial interlocking of pyridyl-macrocycle **5**, azide stopper **4** and aminoacyl-loaded alkyne-thread **6** was successfully accomplished using 0.5 equiv. Cu(MeCN)<sub>4</sub>PF<sub>6</sub> in a mixture of CH<sub>2</sub>Cl<sub>2</sub> and *t*BuOH (4:1 v/v). Stirring the reaction mixture for seven days gave access to rotaxane **3** in 30% yield. Analogous to the simplified test-system from Chapter 2, the reactive arm was introduced using the mild hydrazone exchange conditions from Greaney and co-workers<sup>18</sup> to give the fully assembled molecular machine **18** in good yield.



**Scheme 3.6.** Synthesis of rotaxane-based molecular machine **18**. Reaction conditions:

a)  $\text{Cu}(\text{CH}_3\text{CN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (2:1 v/v), RT, 4 d, 30%; b) aniline, *N*-Boc-Gly-Gly-Cys(STrt)NHN=CHC<sub>6</sub>H<sub>4</sub>OCH<sub>3</sub>, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 2 d, 90%. The italicised letters indicate key signals in the <sup>1</sup>H NMR spectrum shown in Figure 3.3.

The fully assembled machine is stable in its protected form **18** and the interlocked nature of the molecule becomes apparent from the upfield shifts of the proton signal  $\text{H}_i$  corresponding to the triazole on account of shielding from the phenyl rings of the macrocycle, confirming that the ring is trapped in the region of the strand between the terminal stopper and the *N*-Boc-phenylalanine ester (Figure 3.3). Similarly, the signal  $\text{H}_e$  stemming from the two protons on the phenolic ring of the stopper group as well as signals  $\text{H}_f$ ,  $\text{H}_g$ , and  $\text{H}_h$  connecting the triazole moiety with the stopping group are shifted significantly upfield in the rotaxane compared to the free thread. Signals further away from the shielding effects of the macrocycle remain largely unaffected.



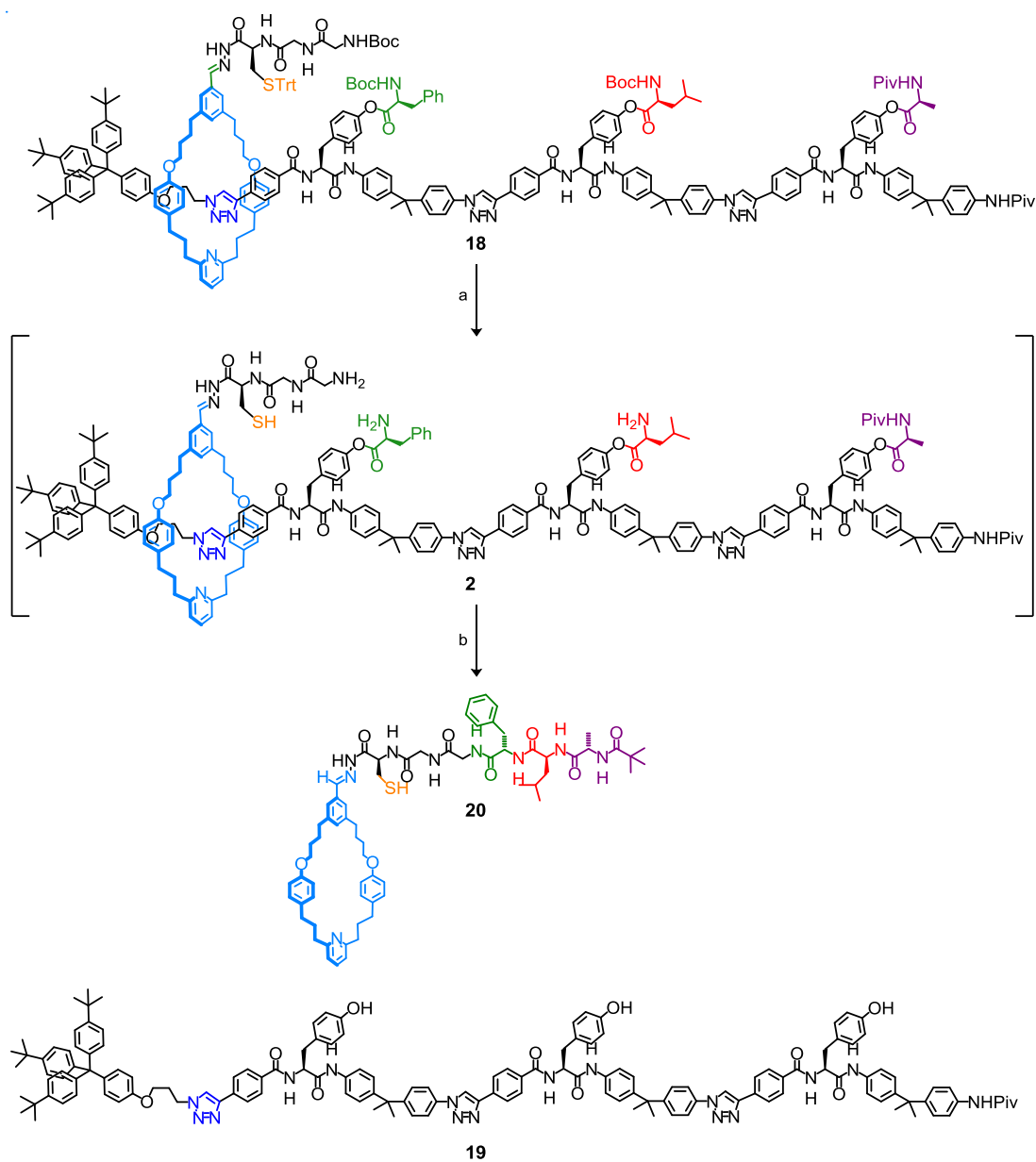
**Figure 3.3.**  $^1\text{H}$  NMR spectrum of (A) the non-interlocked thread and (B) rotaxane **18**, in  $\text{DMSO-}d_6$  (500 MHz, 298 K). Rotaxane **18** exists in both *E*- and *Z*-hydrazone forms. The assignments correspond to the lettering shown in Scheme 3.6.  $\text{C}' = \text{L-S-Trt-cysteine}$ ,  $\text{G}' = \text{L-N-Boc-glycine}$ ,  $\text{F}' = \text{L-N-Boc-phenylalanine}$ ,  $\text{L}' = \text{L-N-Boc-leucine}$ ,  $\text{A}' = \text{L-N-Piv-alanine}$ .

### 3.3.2 Operation of Molecular Machine 18

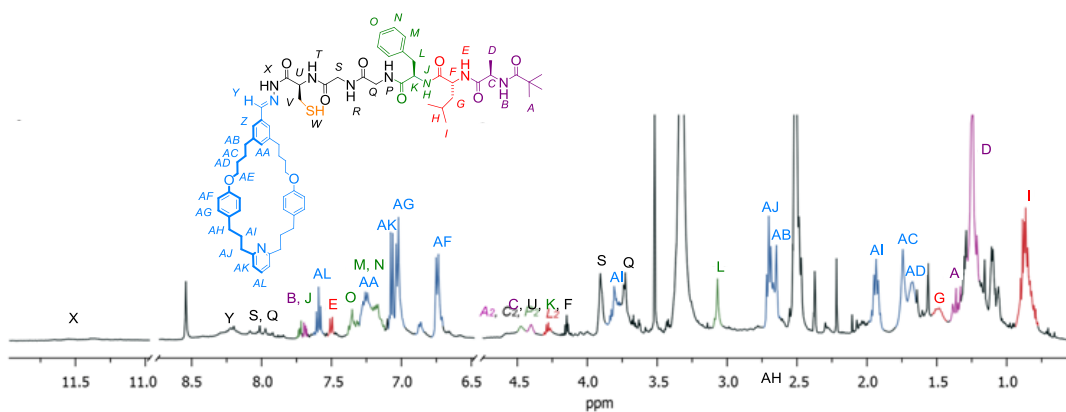
Operation conditions were chosen to be analogous to the previously established single-barrier system: an excess of non-nucleophilic DIPEA in order to deprotonate the thiolate catalyst on the reactive arm as well as the amine functionalities, two equivalents reducing agent TCEP (tris[2-carboxyethyl]phosphine) to break-up disulfide bonds as soon as they are formed and acetonitrile/DMF (3:1 *v/v*) as polar solvent mixtures ensuring a homogeneous solution of all components (Scheme 3.7). After heating at 60 °C for 36 h in the microwave no starting material remained, as evidenced by thin layer chromatography. Two major products were isolated from the reaction mixture. One product was identified by <sup>1</sup>H NMR spectroscopy and mass spectrometry as the completely deacylated thread **19**. The other product had a <sup>1</sup>H NMR spectrum and molecular weight (Figure 3.4) consistent with the macrocycle bearing the hydrazone linked to the hexapeptide *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH) **20**.

The <sup>1</sup>H NMR spectra of the operation product with the correct sequence *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH) **20** and the sequence-scrambled analogue *N*-Piv-Ala-Phe-Leu-Gly-Gly-Cys(SH) **21** are very similar as determined by comparison with an authentic sample and an isomer in which the order of the Phe and Leu residues was reversed, each prepared unambiguously by conventional peptide synthesis.

Therefore, tandem mass spectrometry<sup>19</sup> was chosen to elucidate the peptide sequence of the operation product. In an MS/MS-run based on a product ion scan commonly applied for sequencing of peptide chains, a precursor ion is selected in the first analyser and then allowed to fragment. The generation of product ions is achieved by collision induced dissociation<sup>20</sup> through which the parental ion collides with neutral helium atoms after passing the first analyser. With a second analyser and a detector in line, then the fragments, i.e. the product ions, can be studied. Figure 3.5, A shows the MS/MS spectrum of one isotope of a 2+ ion of **20** formed during ionisation through the cysteine as an *S,N*-acetal, a species that gave a sufficient signal-to-noise ratio for the MS/MS experiment (the adduct is present in the unambiguously prepared reference substrate as well). Figure 3.5, B shows the superimposition of the MS/MS spectra of a similar isotope and ion from the authentic samples prepared unambiguously by synthesis of the macrocycle bearing the sequence *N*-Piv-Ala-Phe-Leu-Gly-Gly-Cys(SH) **20** (red) and *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH) **21** (blue).



**Scheme 3.7.** Operation of molecular machine **18**. Reaction conditions: a)  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1  $v/v$ ), RT, 2 h; b) DIPEA, TCEP·HCl, MeCN/DMF (3:1  $v/v$ ), 60 °C,  $\mu\text{w}$ , 24 h.



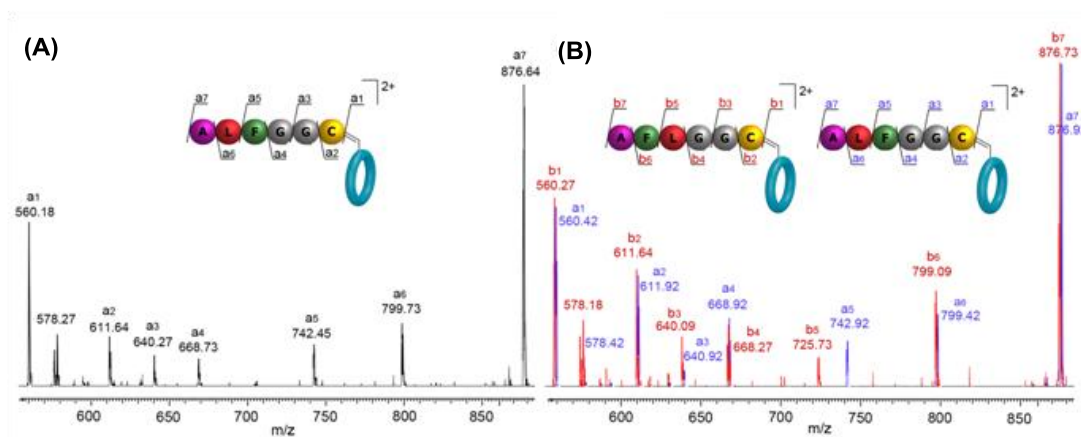
**Figure 3.4.**  $^1\text{H}$  NMR spectrum of molecular machine operation product **20** in  $\text{DMSO-}d_6$  (500 MHz, 298 K).

For both **20** and **21** a fragment with mass  $m/z$  799 is observed as expected from the loss of the terminal pivaloyl-protected alanine (Figure 3.5, B). The difference in the fragmentation masses of the two sequence isomers is apparent from the loss of the next amino acid ( $m/z$  726 for Leu-Gly-Gly-Cys(SH)-macrocycle;  $m/z$  743 for Phe-Gly-Gly-Cys(SH)-macrocycle). Since the MS/MS of operation product **20** shows only a peak  $m/z$  743 it was confirmed as corresponding to the intended sequence isomer.

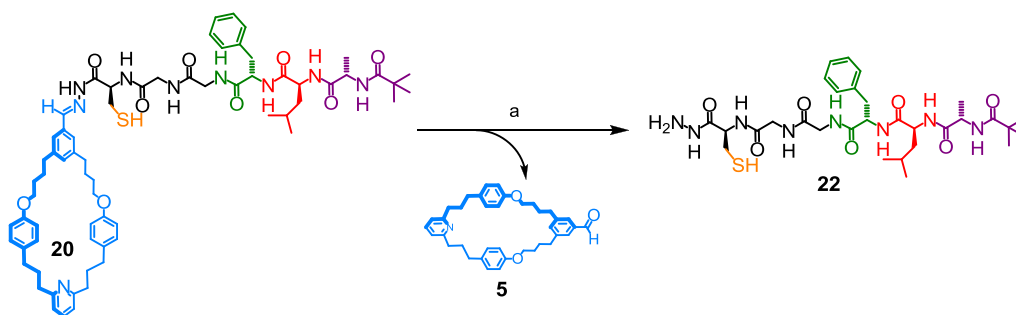
No products corresponding to other peptide compositions (neither different sequences nor peptides with more or less than one Phe, Leu or Ala residue) were detected by preparative thin layer chromatography and subsequent MS analysis as well as performing an MS of the reaction mixture after the operation of **18**, indicating that the peptide synthesis occurs overwhelmingly within the confines of the molecular machine.

In contrast, a control reaction carried out under identical conditions but using the non-threaded strand and macrocycle gave several products, including strands with one or more amino acid groups cleaved, but there was no evidence for the formation of **20** under these conditions. Thus the threaded architecture of the molecular machine, encompassing the catalytic unit, elongation site and the building block strand, is essential for the sequential peptide synthesis and the mode of operation of the molecular machine is consistent with the mechanism shown in Scheme 3.1.

The peptide **22**, still bearing the Cys-Gly-Gly unit at the C-terminus, could subsequently be cleaved from the macrocycle by hydrolysis as evidenced by MS (Scheme 3.8).



**Figure 3.5.** (A) Tandem mass spectrum of a single isotope of the 2+ ion ( $m/z$  876.64) of *S,N*-acetal-derivatised **20**. (B) Superimposed tandem mass spectra of 2+ ions of *S,N*-acetal-derivatised macrocycles bearing the peptide sequences *N*-Piv-Ala-Phe-Leu-Gly-Gly-Cys(SH) (**21**, red; 2+ ion isotope selected  $m/z$  876.73) and *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH) (**20**, blue; 2+ ion isotope selected  $m/z$  876.92), each prepared unambiguously by conventional peptide synthesis. Derivatisation occurs during the sample preparation in MeOH and HCOOH for both operation product and reference material (see Scheme 3.10).



**Scheme 3.8.** Hydrolysis of operation product **20**. Reaction conditions: a) CH<sub>2</sub>Cl<sub>2</sub>/TFA/H<sub>2</sub>O (3:1:1 v/v/v), RT, 12 h.

### 3.4 Conclusion

The synthesis of the small peptide through autonomous multi-step synthesis by artificial small-molecule machine **2** was carried out on a scale of tens of milligrams, corresponding to parallel synthesis by  $\sim 10^{18}$  machines. Once its operation is initiated, the synthetic tasks performed by **2** proceed autonomously, requiring no further intervention. As the catalytic thiolate is constrained by the threaded architecture of the machine from reacting with building blocks out of sequence, balancing the rate of reactions with the speed that templates rearrange is unnecessary for a rotaxane-based machine.

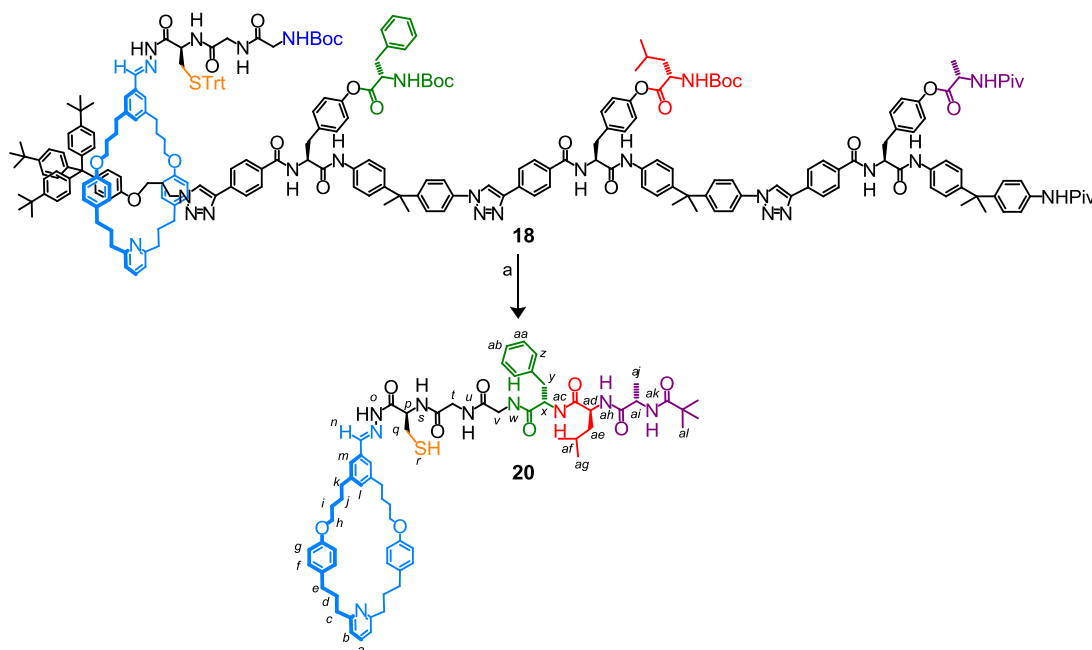
Rotaxane **2** is a (very) primitive analogue of the ribosome. Limitations of the second generation artificial system include slow kinetics (1 takes  $\sim 12$  h to make each amide bond, compared to the 10 to 15 amide bonds synthesised per second by a ribosome) and loss of the sequence information on the strand as it is translated into the product. Furthermore, the length of an oligopeptide that can be produced may ultimately be restricted by the size of the cyclic transition states involved in *S*-to-*N* acyl transfer (although peptide ligation has been successfully utilised with up to 29-membered cyclic transition states<sup>27</sup>). Nevertheless **2** demonstrates that relatively small, highly modular, artificial molecular machines can be designed to autonomously perform iterative tasks in synthesis. Potentially, the principles employed in the design and operation of **2** should be broadly applicable to other types of monomers and chemical reactions<sup>28</sup>.

## 3.5 Experimental Section

### 3.5.1 General Methods and Abbreviations

Unless stated otherwise, all reagents and solvents were purchased from Aldrich Chemicals and used without further purification. Compound **4**<sup>17</sup> was prepared according to the literature procedure. Dr. Guillaume De Bo conducted a control reaction carried out under identical conditions as depicted in Scheme 3.12 but using the non-threaded strand and macrocycle. Compounds **21**, **E4**, **E6**, **E9**, and **E10** were prepared by Dr. Bartosz Lewandowski. Compounds **E5** and **E8** were prepared using solid phase peptide synthesis by Dr. Guillaume De Bo. The synthesis of **1**, **5**, **10**, **16** and **E3** is described in Chapter 2.<sup>11</sup> All reactions requiring microwave heating were carried out in a Biotage Initiator 2.5.3. Dry DMF, CH<sub>2</sub>Cl<sub>2</sub>, CHCl<sub>3</sub> and THF was obtained by passing the solvent (HPLC grade) through an activated alumina column on a PureSolv™ solvent purification system (InnovativeTechnologies Inc., MA). Flash column chromatography was carried out using Geduran® Si 60 (particle size 40-63 μm, Merck, Germany) as the stationary phase, and analytical TLC was performed on precoated silica gel plates (0.25 mm thickness, 60 F254, Merck, Germany) and observed under UV light or stained with a phosphomolybdic acid solution. Preparative TLC was performed on precoated silica gel plates (2 mm, UNIPLATE GF, Analtech Inc., DE or 2 mm, Tapered plate, UNIPLATE-T, Analtech Inc., DE). NMR spectra were recorded on Bruker AV 400 MHz, Bruker AV 500 MHz (equipped with a cryoprobe) or Bruker Avance III (equipped with a cryoprobe) with an Oxford AS 600 MHz magnet. Chemical shifts are reported in parts per million (ppm) from high to low frequency and referenced to the residual solvent resonance. Coupling constants (*J*) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintuplet, m = multiplet, br = broad. <sup>1</sup>H-NMR assignments were made using 2D NMR methods (COSY, HSQC, HMBC). Melting points (m.p.) were determined using a Sanyo Gallenkamp apparatus and are reported uncorrected. Low resolution ESI mass spectrometry was performed with a Finnigan LCQ-MS mass spectrometer. High resolution ESI (electrospray ionisation) and APCI (atmospheric-pressure chemical ionisation) mass spectrometry was carried out by the mass spectrometry services at the EPSRC National Mass Spectrometry Service Centre, Swansea, UK or the Mass Spectrometry Service of the University of Manchester. Optical rotations were recorded using a Bellingham and Stanley ADP 220 Polarimeter using a 1 dm cell.

## 3.5.1.1 Operation of Molecular Machine 18



**Scheme 3.9.** Operation of molecular machine **18**. Reaction conditions: a)  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v), RT, 2 h, then DIPEA, TCEP·HCl, MeCN/DMF (3:1 v/v), 60 °C,  $\mu\text{w}$ , 36 h.

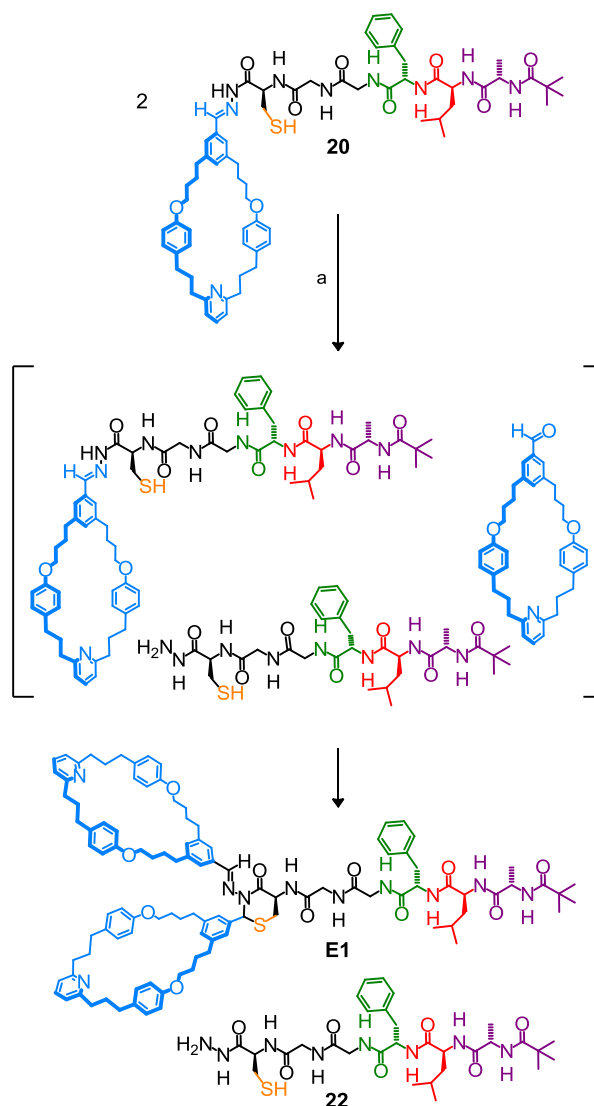
Molecular machine **18** (20 mg, 5.0  $\mu\text{mol}$ ) was stirred at room temperature in a mixture of  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (5:1 v/v, 2 ml) for 2 h. Toluene (5 ml) was added and the solvent was removed under reduced pressure. The residue was then dissolved in degassed MeCN/DMF (3:1 v/v, 4 ml) and both *N,N*-diisopropylethylamine (43  $\mu\text{l}$ , 0.25 mmol) and tris(2-carboxyethyl)phosphine hydrochloride (2 mg, 8  $\mu\text{mol}$ ) were added and the reaction mixture was stirred under microwave heating at 60 °C for 36 h. Preparative TLC ( $\text{SiO}_2$  [UNIPLATE-T],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  5%) of the residue afforded **20** (2.0 mg, 30%) as a colourless film.

$^1\text{H}$  NMR (500 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 11.40 (m, 1H,  $\text{H}_o$ ), 8.23 (m, 1H,  $\text{H}_n$ ), 8.00 (m, 2H,  $\text{H}_{u+w}$ ), 7.73 (m, 1H,  $\text{H}_{ac}$ ), 7.69 (m, 1H,  $\text{H}_{ak}$ ), 7.59 (t,  $J$  = 5.9 Hz, 1H,  $\text{H}_a$ ), 7.48 (m, 1H,  $\text{H}_{ah}$ ), 7.26 (m, 7H,  $\text{H}_{m+z+aa+ab}$ ), 7.07 (d,  $J$  = 5.7 Hz, 2H,  $\text{H}_b$ ), 7.03 (d,  $J$  = 8.4 Hz, 4H,  $\text{H}_j$ ), 6.87 (d,  $J$  = 7.5 Hz, 1H,  $\text{H}_l$ ), 6.74 (d,  $J$  = 8.3 Hz, 4H,  $\text{H}_g$ ), 4.58 (bs, 1H,  $\text{H}_p$ ), 4.47 (bs, 1H,  $\text{H}_x$ ), 4.41 (bs, 1H,  $\text{H}_{al}$ ), 4.26 (m, 1H,  $\text{H}_{ad}$ ), 3.91 (m, 2H,  $\text{H}_t$ ), 3.80 (m, 4H,  $\text{H}_h$ ), 3.72 (m, 2H,  $\text{H}_v$ ), 3.07 (m, 2H,  $\text{H}_y$ ), 2.69 (m, 8H,  $\text{H}_{k+c}$ ), 1.93 (m, 4H,  $\text{H}_d$ ), 1.74 (m, 4H,  $\text{H}_i$ ), 1.67 (m, 5H,  $\text{H}_{i+af}$ ), 1.49 (m, 2H,  $\text{H}_{ae}$ ), 1.36 (m, 3H,  $\text{H}_{aj}$ ), 1.25 (s, 9H,  $\text{H}_{al}$ ), 0.86 (m, 6H,  $\text{H}_{ag}$ ); LRESI(+)-MS (MeOH):  $m/z$  1208.7  $[\text{M}+\text{H}]^+$ .

*Please note: The mechanism for epimerisation of the amino acid units during the machine's operation would cause significant hydrolysis of the acyl amino units from the thread leading to loss of sequence integrity in the product, which is not observed in the operation of **18**.*

*Thus, it is very likely that the molecular machine operates with high stereochemical integrity.*

## 3.5.1.2 Mass-spectrometric Analysis of the Operation Product 20

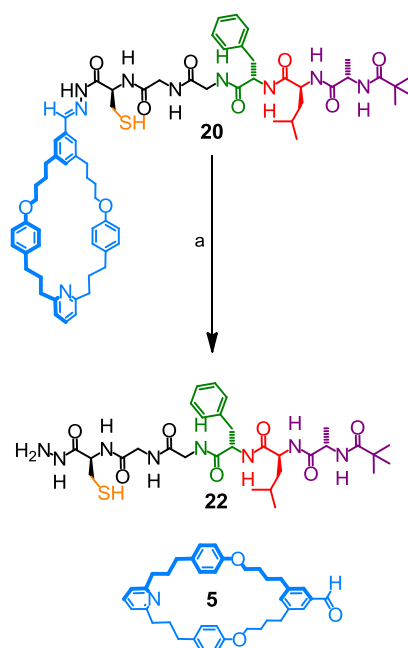


**Scheme 3.10.** *S,N*-Acetal formation observed during sample preparation of hexapeptide macrocycle for mass spectrometry. Reaction conditions: HCO<sub>2</sub>H/CH<sub>2</sub>Cl<sub>2</sub>/H<sub>2</sub>O/MeOH (1:5:10:84 v/v/v/v).

Analysis of the operation of molecular machine **18** obtained has a mass spectrum corresponding to the desired product **20** (LRESI(+)-MS [MeOH]:  $m/z$  1208.7 [M+H]<sup>+</sup>). Upon addition of formic acid during sample preparation (0.1 mg of **20** in 1ml of HCO<sub>2</sub>H:CH<sub>2</sub>Cl<sub>2</sub>:H<sub>2</sub>O:MeOH [1:5:10:84 v/v/v/v]) intended for multiply-charged ion tandem mass spectrometry experiments (MS/MS experiments normally work better with multiply charged ions which are typically produced by adding acid to the sample) the rapid formation (within 5 min, complete conversion after 1 h according to MS) of signals corresponding to the free aldehyde macrocycle (**5**,  $m/z$  562.3 [M+H]<sup>+</sup>) and the formation of a derivative of **20**

was observed. It was rationalised, that *S,N*-acetal formation with another molecule of **5** occurred to give **E1** (Scheme 3.10).

Similarly, *S,N*-acetal could also be observed by dissolving unambiguously prepared macrocycle **20** or **21** upon addition of formic acid during sample preparation (0.1 mg of **7** in 1ml of HCO<sub>2</sub>H:CH<sub>2</sub>Cl<sub>2</sub>:H<sub>2</sub>O:MeOH [1:5:10:84 v/v/v/v]). The *S,N*-acetal accepts multiple charges more readily than the underivatised counterparts and accordingly was used in the MS/MS peptide sequencing experiments where typically M<sup>2+</sup> ions give best results.

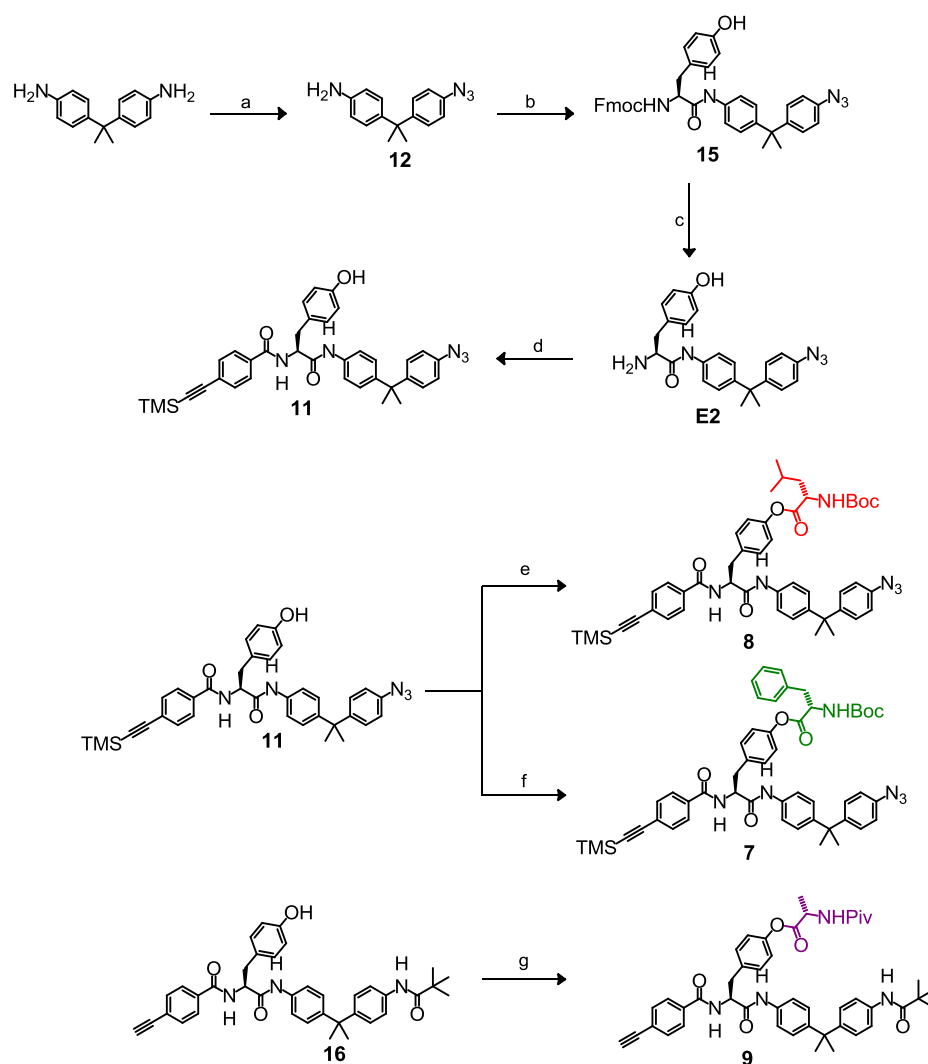
3.5.1.3 Hydrolysis of **20**

**Scheme 3.11.** Hydrolysis of the hexapeptide macrocycle. Reaction conditions: a) CH<sub>2</sub>Cl<sub>2</sub>/TFA/H<sub>2</sub>O (3:1:1 v/v/v), RT, 12 h.

Macrocycle **20** (2.0 mg, 1.7 μmol) was dissolved in a mixture of CH<sub>2</sub>Cl<sub>2</sub>/TFA/H<sub>2</sub>O (3:1:1 v/v/v, 1 ml). The reaction mixture was stirred at room temperature overnight. Subsequently toluene (5 ml) was added and the solvent was removed under reduced pressure. The residue was purified by preparative TLC (SiO<sub>2</sub> [Analtech GF], CH<sub>2</sub>Cl<sub>2</sub>/MeOH 5%) and analysed by LRESI-MS (MeOH): *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH)-NHNH<sub>2</sub> *m/z* 563.25 [M+H]<sup>+</sup>; *N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH)-OH *m/z* 664.23 [M+H]<sup>+</sup>, *m/z* 673.00 [M+Na]<sup>+</sup>.

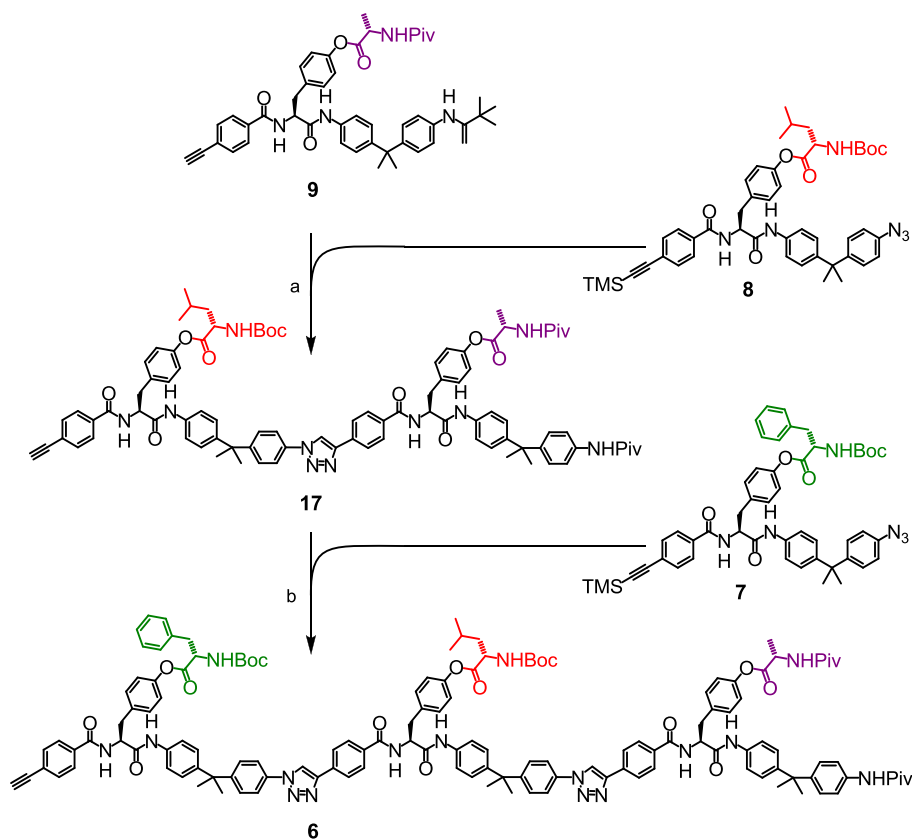
## 3.5.2 Synthetic Schemes

## 3.5.2.1 Synthesis of Barriers 7, 8 and 9



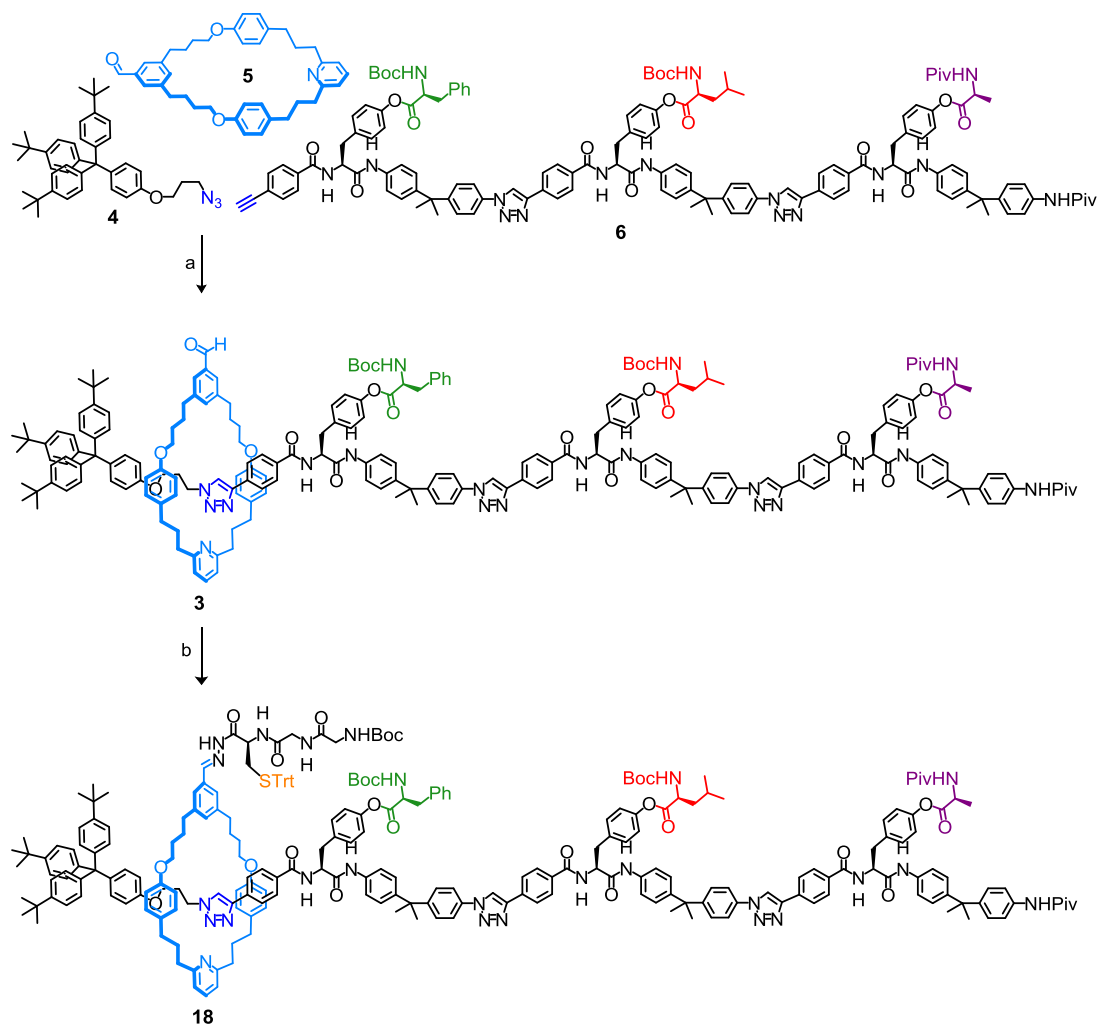
**Scheme 3.12.** Reaction conditions: a) *iso*-amyl nitrite, NaN<sub>3</sub>, AcOH, RT, 2 h, 45%; b) L-*N*-Fmoc-tyrosine, HOBt, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v) 12 h, 60%; c) DMF/piperidine (4:1 v/v), RT, 3 h, 97%; d) 4-([trimethylsilyl]ethynyl)benzoic acid, HOBt, EDCI·HCl, THF/CHCl<sub>3</sub> (2:1 v/v), RT, 12 h, 71%; e) L-*N*-Boc-leucine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 42%; f) L-*N*-Boc-phenylalanine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 63%; g) L-*N*-Piv-alanine, EDCI·HCl, DCC, DMAP, CH<sub>2</sub>Cl<sub>2</sub>, RT, 48 h, 56%.

## 3.5.2.2 Synthesis of Three-Barrier Alkyne 6



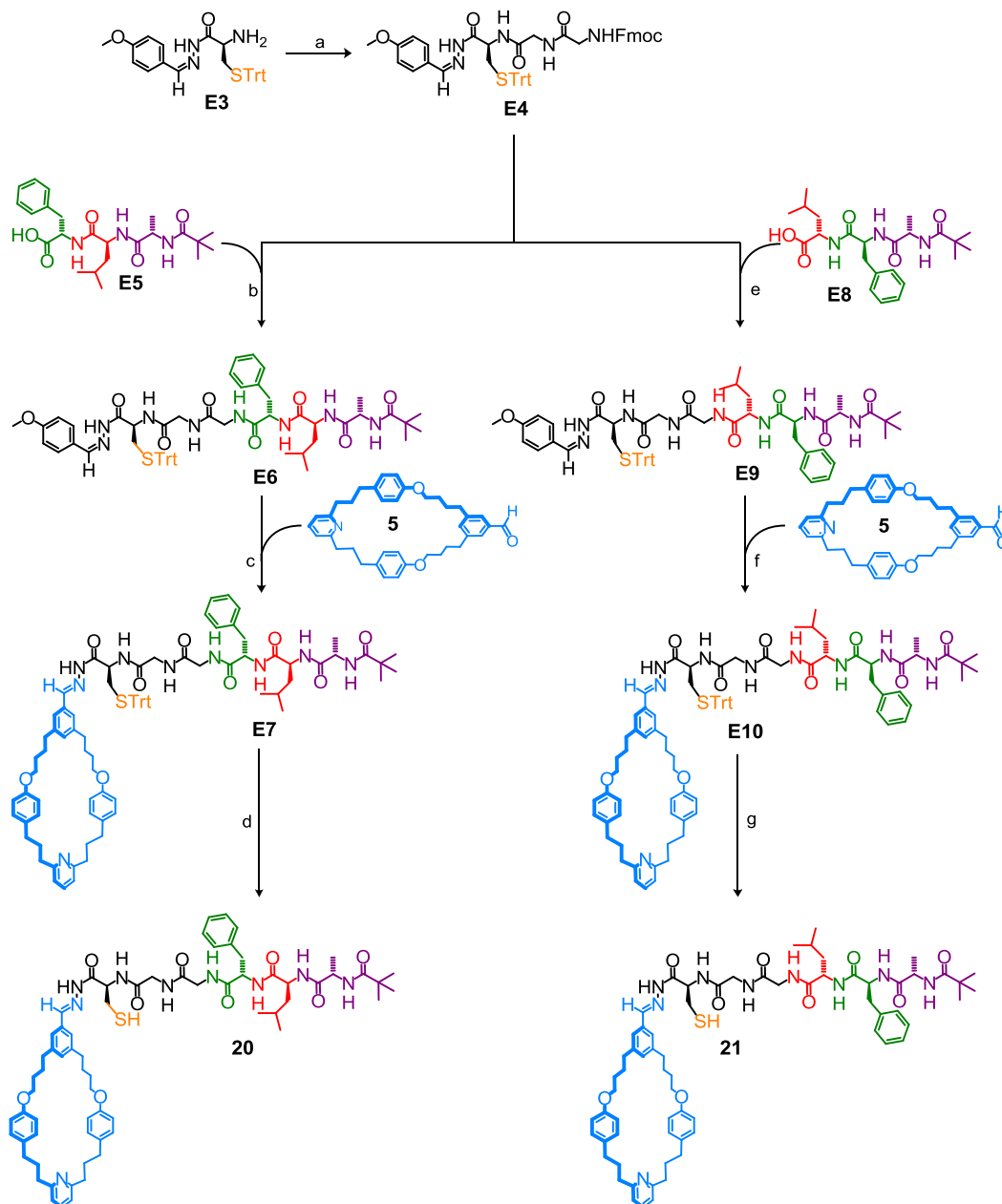
**Scheme 3.13.** Reaction conditions: a) Tentagel TBTA,  $\text{Cu}(\text{MeCN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (1:1 v/v), RT, 48 h, then  $\text{AgNO}_3$ , acetone/ $\text{H}_2\text{O}$  (4:1 v/v), RT, 18 h, 43% over two steps; b) Tentagel TBTA,  $\text{Cu}(\text{MeCN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (1:1 v/v), RT, 48 h, then  $\text{AgNO}_3$ , acetone/ $\text{H}_2\text{O}$  (4:1 v/v), RT, 18 h, 37% over two steps.

## 3.5.2.3 Synthesis of Three-Barrier Rotaxane 18



**Scheme 3.14** Reaction conditions: a)  $\text{Cu}(\text{MeCN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v), RT, 7 d, 32%; b) aniline, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 72 h, 90%.

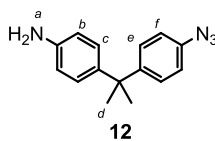
## 3.5.2.4 Synthesis of Reference Macrocyclic 20 and 21



**Scheme 3.15.** Reaction conditions: a) Fmoc-Gly-Gly-OH, DMF, PyBrop, DIPEA, RT, 12 h, 84%; b) DMF/piperidine (4:1 v/v), RT, 2 h, then Piv-Ala-Leu-Phe-OH, DIPEA, DMF, RT, 12 h, 88% over two steps; c) aniline, DMSO/MES (3:1 v/v, pH 6.0), 60 °C, 3 d, 75%; d) TIPS, CH<sub>2</sub>Cl<sub>2</sub>/TFA (4:1 v/v), RT, 2 h, 99%; e) DMF/piperidine (4:1 v/v), RT, 2 h, then Piv-Ala-Phe-Leu-OH, DIPEA, DMF, RT, 12 h, 70% over two steps; f) aniline, DMSO/MES (3:1 v/v, pH 6.0), 60 °C, 3 d, 75%; g) TIPS, CH<sub>2</sub>Cl<sub>2</sub>/TFA (4:1 v/v), RT, 2 h, 99%.

## 2.4.5 Synthetic Procedures and Characterisation Data

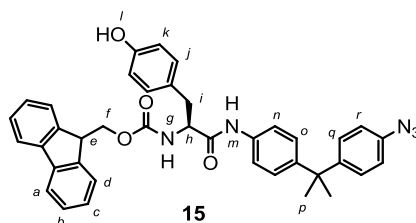
### 4-(2-[4-Azidophenyl]propan-2-yl)aniline **12**



To a stirred solution of 4,4'-(propane-2,2-diyl)dianiline (1.48 g, 6.50 mmol, 1.00 equiv.) in acetic acid at 0 °C was added dropwise *iso*-amyl nitrite (674 mg, 6.50 mmol, 1.00 equiv.). The reaction mixture was allowed to warm to room temperature and after 30 min NaN<sub>3</sub> (850 mg, 13.0 mmol, 2.00 equiv.) was added in small portions. Stirring was continued at room temperature for 2 h. The solvent was removed under reduced pressure and the residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (250 ml). The organic layer was washed with aqueous saturated Na<sub>2</sub>CO<sub>3</sub> (200 ml), brine (50 ml), dried over MgSO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 7:3) of the residue afforded **12** (735 mg, 45%) as a pale yellow solid.

m.p. 47 °C; <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ = 7.23 (d, *J* = 8.7 Hz, 2H, H<sub>f</sub>), 7.00 (d, *J* = 8.6 Hz, 2H, H<sub>c</sub>), 6.92 (d, *J* = 8.7 Hz, 2H, H<sub>e</sub>), 6.61 (d, *J* = 8.6 Hz, 2H, H<sub>b</sub>), 3.58 (s, 2H, H<sub>a</sub>), 1.63 (s, 6H, H<sub>d</sub>); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ = 148.4, 144.3, 140.7, 137.3, 128.4, 127.8, 118.7, 115.0, 42.1, 31.1; LRAPCI(+)-MS (CH<sub>2</sub>Cl<sub>2</sub>): *m/z* 252.9 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 253.1451 [M+H]<sup>+</sup>, 253.1448 calc. for C<sub>15</sub>H<sub>17</sub>N<sub>4</sub>.

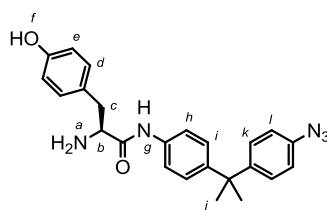
**(S)-(9H-Fluoren-9-yl)methyl 1-(4-[2-(4-Azidophenyl)propan-2-yl]phenylamino)-3-(4-hydroxyphenyl)-1-oxopropan-2-ylcarbamate **15****



To a stirred solution of L-N-Fmoc-tyrosine (1.10 g, 2.74 mmol, 1.00 equiv.), HOBt·H<sub>2</sub>O (443 mg, 2.89 mmol, 1.05 equiv.) and EDCI·HCl (554 mg, 2.89 mmol, 1.05 equiv.) in THF/CHCl<sub>3</sub> (2:1 v/v, 15 ml) at room temperature was added amine **12** (730 mg, 2.89 mmol, 1.05 equiv.) in THF/CHCl<sub>3</sub> (2:1 v/v, 6 ml) dropwise. The reaction mixture was stirred at room temperature for 12 h. The solvent was removed under reduced pressure. The residue dissolved in CH<sub>2</sub>Cl<sub>2</sub> (250 ml). The organic layer was washed with H<sub>2</sub>O (200 ml), dried over MgSO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 7:3) of the residue afforded **15** (1.06 g, 60%) as a pale yellow solid.

m.p. 114–116 °C;  $[\alpha]_D = 13.6^\circ$  (c 1.02, CHCl<sub>3</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ = 7.99 (s, 1H, H<sub>m</sub>), 7.72 (d, *J* = 7.6 Hz, 2H, H<sub>a</sub>), 7.49 (t, *J* = 7.9 Hz, 2H, H<sub>c</sub>), 7.35 (t, *J* = 7.5 Hz, 2H, H<sub>b</sub>), 7.29 – 7.19 (m, 4H, H<sub>d+r</sub>), 7.13 (d, *J* = 8.6 Hz, 2H, H<sub>n</sub>), 7.05 (d, *J* = 8.3 Hz, 2H, H<sub>q</sub>), 7.03 – 6.96 (m, 2H, H<sub>j</sub>), 6.88 (d, *J* = 8.6 Hz, 2H, H<sub>o</sub>), 6.66 (d, *J* = 8.3 Hz, 2H, H<sub>k</sub>), 6.00 – 5.94 (m, 1H, H<sub>l</sub>), 5.80 – 5.70 (m, 1H, H<sub>g</sub>), 4.57 – 4.46 (m, 1H, H<sub>h</sub>), 4.42 – 4.25 (m, 2H, H<sub>f</sub>), 4.14 (t, *J* = 7.0 Hz, 1H, H<sub>e</sub>), 3.06 – 2.97 (m, 2H, H<sub>i</sub>), 1.58 (s, 6H, H<sub>p</sub>); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ = 169.6, 156.5, 155.1, 147.5, 147.0, 143.7, 143.6, 141.4, 137.5, 134.8, 130.6, 129.2, 128.3, 127.9, 127.3, 127.3, 125.1, 120.2, 118.7, 115.8, 67.5, 57.3, 47.1, 42.4, 38.1, 30.8; LRAPCI(+)-MS (CH<sub>2</sub>Cl<sub>2</sub>): *m/z* 637.7 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 638.2757 [M+H]<sup>+</sup>, 638.2762 calc. for C<sub>39</sub>H<sub>36</sub>N<sub>5</sub>O<sub>4</sub>.

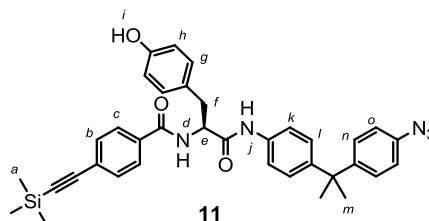
**(S)-2-Amino-N-(4-[2-(4-azidophenyl)propan-2-yl]phenyl)-3-(4-hydroxyphenyl)propanamide E2**



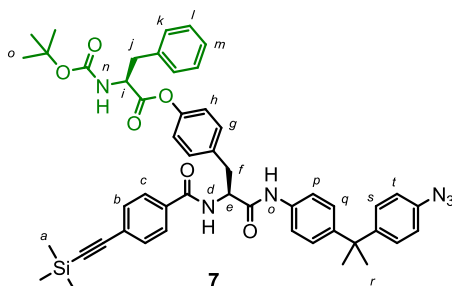
**E2**

Fmoc-derivative **15** (1.04 g, 1.64 mmol, 1.00 equiv.) was dissolved in DMF/piperidine (4:1 v/v, 20 ml) and the reaction mixture was stirred at room temperature for 3 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>/MeOH 5%) of the residue afforded **E2** (665 mg, 97%) as a pale yellow solid.

m.p. 102 °C; <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ = 9.35 (s, 1H, H<sub>g</sub>), 7.48 (d, *J* = 8.6 Hz, 2H, H<sub>l</sub>), 7.23 – 7.13 (m, 4H, H<sub>h+k</sub>), 7.10 (d, *J* = 8.4 Hz, 2H, H<sub>d</sub>), 6.92 (d, *J* = 8.6 Hz, 2H, H<sub>i</sub>), 6.80 (d, *J* = 8.4 Hz, 2H, H<sub>e</sub>), 3.68 (dd, *J* = 9.2 Hz, 4.0 Hz, 1H, H<sub>b</sub>), 3.24 (dd, *J* = 14.0 Hz, 4.0 Hz, 1H, H<sub>c</sub>), 2.73 (dd, *J* = 13.9 Hz, 9.2 Hz, 1H, H<sub>c</sub>), 1.65 (s, 6H, H<sub>j</sub>); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ = 172.7, 154.9, 147.7, 146.4, 137.5, 135.4, 130.6, 129.5, 128.3, 127.4, 119.5, 118.7, 115.8, 57.0, 42.5, 39.9, 30.9; LRESI(+)-MS (MeOH): *m/z* 416.2 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 416.2081 [M+H]<sup>+</sup>, 416.2081 calc. for C<sub>24</sub>H<sub>26</sub>N<sub>5</sub>O<sub>2</sub>.

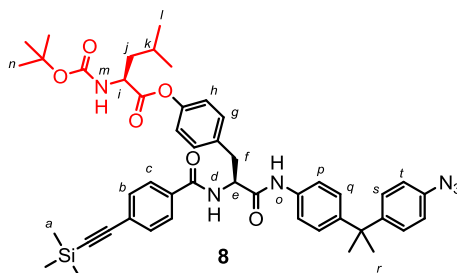
**TMS-Ethynyl-Tyrosine-Azide Derivative 11**

To a solution of amine **E2** (630 mg, 1.50 mmol, 1.00 equiv.), 4-([trimethylsilyl]ethynyl)benzoic acid (397 mg, 1.81 mmol) and EDCI·HCl (349 mg, 1.81 mmol, 1.20 equiv.) in THF/CHCl<sub>3</sub> (2:1 *v/v*, 20 ml) was added HOBt·H<sub>2</sub>O (279 mg, 1.81 mmol, 1.20 equiv.) and the reaction mixture was stirred at room temperature for 12 h. The solvent was removed under reduced pressure and the residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (250 ml). The organic layer was washed with H<sub>2</sub>O (150 ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 7:3) of the residue afforded **11** (666 mg, 71%) as a pale yellow solid. m.p. 132–134 °C; [α]<sub>D</sub> = 13.6 (c 1.03, CHCl<sub>3</sub>); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ = 8.48 (s, 1H, H<sub>j</sub>), 7.67 (d, *J* = 8.5 Hz, 2H, H<sub>b</sub>), 7.45 (d, *J* = 8.3 Hz, 2H, H<sub>c</sub>), 7.34 – 7.26 (m, 3H, H<sub>d+o</sub>), 7.15 (d, *J* = 8.6 Hz, 2H, H<sub>k</sub>), 7.12 – 7.03 (m, 4H, H<sub>g+n</sub>), 6.89 (d, *J* = 8.6 Hz, 2H, H<sub>l</sub>), 6.68 (d, *J* = 8.5 Hz, 2H, H<sub>h+i</sub>), 5.04 (q, *J* = 7.2 Hz, 1H, H<sub>e</sub>), 3.17 (br d, *J* = 6.3 Hz, 2H, H<sub>f</sub>), 1.60 (s, 6H, H<sub>m</sub>), 0.26 (s, 9H, H<sub>a</sub>); <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>) δ = 169.7, 167.2, 155.2, 147.5, 147.0, 137.5, 135.0, 132.9, 132.3, 130.7, 128.3, 128.0, 127.3, 127.2, 120.2, 118.8, 115.9, 110.1, 104.0, 97.7, 56.2, 42.5, 38.1, 30.8, 0.0; LRESI(+)-MS (MeOH): *m/z* 638.3 [M+Na]<sup>+</sup>; HRFT(+)-MS: *m/z* 616.2738 [M+H]<sup>+</sup>, 616.2738 calc. for C<sub>36</sub>H<sub>38</sub>N<sub>5</sub>O<sub>3</sub>Si.

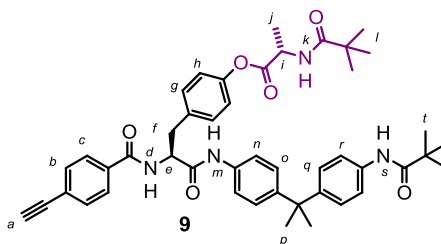
**TMS-Ethynyl-*N*-Boc-Phenylalanine-Azide Barrier 7**

Phenol **11** (500 mg, 0.81 mmol, 1.00 equiv.), L-*N*-Boc-phenylalanine (430 mg, 1.62 mmol, 2.00 equiv.), EDCI·HCl (171 mg, 0.90 mmol, 1.11 equiv.), DCC (184 mg, 0.90 mmol, 1.11 equiv.) and DMAP (9 mg, 0.08 mmol, 0.10 equiv.) were stirred in CH<sub>2</sub>Cl<sub>2</sub> (10 ml) at room temperature for 24 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 6:4) of the residue afforded **7** (300 mg, 42%) as a pale yellow solid.

m.p. 92-95 °C;  $[\alpha]_D = +0.01^\circ$  (c 4.10, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  = 8.22 (s, 1H, H<sub>o</sub>), 7.65 (d, *J* = 8.4 Hz, 2H, H<sub>b</sub>), 7.47 (d, *J* = 8.4 Hz, 2H, H<sub>c</sub>), 7.33 – 7.24 (m, 7H, H<sub>t+p+k+m</sub>), 7.22 (d, *J* = 6.8, 2H, H<sub>m</sub>), 7.19 – 7.15 (m, 2H, H<sub>g</sub>), 7.17 – 7.10 (m, 3H, H<sub>d+s</sub>), 6.93 – 6.88 (m, 4H, H<sub>h+q</sub>), 5.09 (d, *J* = 8.0, 1H, H<sub>n</sub>), 5.02 (q, *J* = 7.3, 1H, H<sub>e</sub>), 4.83 – 4.73 (m, 1H, H<sub>i</sub>), 3.31 - 3.17 (m, 4H, H<sub>f+j</sub>), 1.62 (s, 6H, H<sub>r</sub>), 1.43 (s, 9H, H<sub>o</sub>), 0.29 – 0.24 (m, 9H, H<sub>a</sub>); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  = 170.8, 169.1, 167.1, 157.0, 155.4, 149.7, 147.6, 147.0, 137.6, 135.9, 135.1, 134.6, 133.1, 132.4, 130.7, 129.7, 128.9, 128.4, 127.5, 127.4, 127.2, 121.9, 120.2, 118.8, 104.1, 97.7, 80.4, 55.9, 42.6, 38.5, 38.0, 34.0, 30.9, 28.5, -0.1; LRESI(+)-MS (MeOH): *m/z* 863.6 [M+H]<sup>+</sup>; HRFT(+)-MS: *m/z* 863.3931 [M+H]<sup>+</sup>, 863.3946 calc. for C<sub>50</sub>H<sub>55</sub>N<sub>6</sub>O<sub>6</sub>Si.

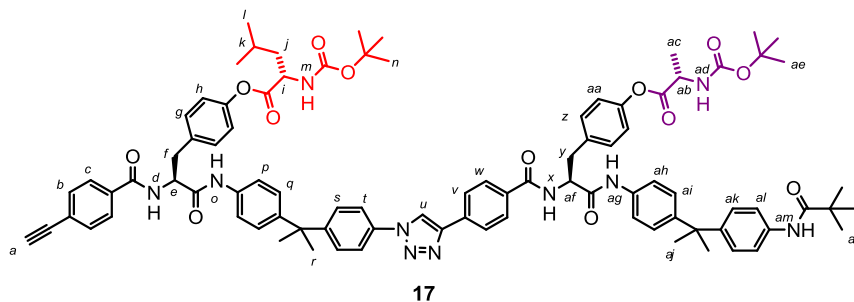
**TMS-Ethynyl-*N*-Boc-Leucine-Azide Barrier 8**

Phenol **11** (400 mg, 0.64 mmol, 1.00 equiv.), L-*N*-Boc-leucine (300 mg, 1.30 mmol, 2.00 equiv.), EDCI·HCl (124 mg, 0.64 mmol, 1.00 equiv.), DCC (134 mg, 0.64 mmol, 1.00 equiv.), and DMAP (16 mg, 0.12 mmol, 0.19 equiv.) were stirred in CH<sub>2</sub>Cl<sub>2</sub> (4 ml) at room temperature for 24 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (8 ml). The organic layer was washed with H<sub>2</sub>O (150 ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 6:4) of the residue afforded **8** (424 mg, 63%) as a pale yellow solid. m.p. 111 °C; [α]<sub>D</sub> = +0.05° (c 0.50, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ = 7.91 (s, 1H, H<sub>o</sub>), 7.69 (d, *J* = 8.3, 2H, H<sub>b</sub>), 7.53 (d, *J* = 8.4, 2H, H<sub>c</sub>), 7.34 (d, *J* = 8.4, 2H, H<sub>i</sub>), 7.32 – 7.28 (m, 2H, H<sub>p</sub>), 7.20 (d, *J* = 8.6, 2H, H<sub>g</sub>), 7.14 (d, *J* = 8.6, 2H, H<sub>q</sub>), 7.07 (d, *J* = 8.2, 2H, H<sub>s</sub>), 7.02 – 6.98 (m, 1H, H<sub>d</sub>), 6.94 (d, *J* = 8.6, 2H, H<sub>h</sub>), 5.00 – 4.90 (m, 2H, H<sub>e+m</sub>), 4.54 (s, 1H, H<sub>i</sub>), 3.33 (dd, *J* = 13.8, 6.0, 1H, H<sub>j</sub>), 3.24 (dd, *J* = 13.8, 8.2, 1H, H<sub>j'</sub>), 1.90 – 1.76 (m, 2H, H<sub>j</sub>), 1.73 – 1.60 (m, 7H, H<sub>r+k</sub>), 1.48 (s, 9H, H<sub>n</sub>), 1.04 (d, *J* = 6.3, 6H, H<sub>l</sub>), 0.26 (s, 9H, H<sub>a</sub>); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ = 172.3, 168.8, 167.0, 155.6, 149.9, 147.6, 147.0, 137.6, 134.9, 134.4, 133.0, 132.3, 130.6, 128.3, 127.4, 127.2, 127.1, 122.0, 120.2, 118.8, 104.0, 97.7, 80.3, 55.9, 52.5, 41.7, 42.5, 37.9, 30.9, 28.5, 23.1, 0.0; LRESI(+)-MS (MeOH): *m/z* 852.42 [M+Na]<sup>+</sup>; HRFT(+)-MS: *m/z* 529.4103 [M+H]<sup>+</sup>, 529.4136 calc. for C<sub>47</sub>H<sub>57</sub>N<sub>6</sub>O<sub>6</sub>Si.

**Ethynyl-*N*-Piv-Alanine-*N*-Pivaloyl Barrier 9**

Phenol **16** (300 mg, 0.50 mmol, 1.00 equiv.), *L*-*N*-piv-alanine (173 mg, 1.00 mmol, 2.00 equiv.), EDCI·HCl (96 mg, 0.50 mmol, 1.00 equiv.), DCC (103 mg, 0.50 mmol, 1.00 equiv.) and DMAP (12 mg, 0.10 mmol, 0.20 equiv.) were allowed to stir in CH<sub>2</sub>Cl<sub>2</sub> (3 ml) at room temperature for 48 h. The solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, petroleum ether/EtOAc 6:4) of the residue afforded **9** (328 mg, 56%) as a colourless solid.

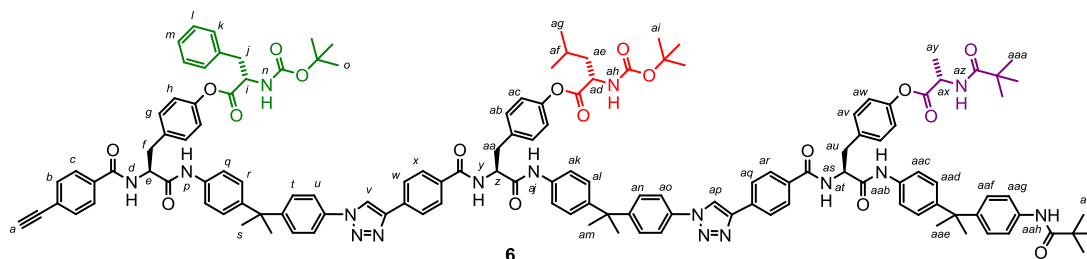
*Please Note: The title compound is a 1:1 mixture of two diastereoisomers.* m.p. 104 °C; [α]<sub>D</sub> = +0.0° (c 0.60, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ = 8.82 – 8.74 (m, 1H, H<sub>i</sub>), 7.68 (d, *J* = 8.4, 2H, H<sub>c</sub>), 7.56 – 7.50 (m, 1H, H<sub>d</sub>), 7.46 (d, *J* = 8.4 Hz, 2H, H<sub>b</sub>), 7.41 (d, *J* = 8.8 Hz, 2H, H<sub>g</sub>), 7.31 – 7.27 (m, 2H, H<sub>n</sub>), 7.20 – 7.16 (m, 2H, H<sub>q</sub>), 7.15 – 7.08 (m, 4H, H<sub>o+h</sub>), 6.92 – 6.88 (m, 2H, H<sub>h</sub>), 6.27 – 6.22 (m, 1H, H<sub>k</sub>), 5.17 – 5.08 (m, 1H, H<sub>e</sub>), 4.76 – 4.68 (m, 1H, H<sub>i</sub>), 3.24 – 3.15 (m, 3H, H<sub>a+f</sub>), 1.60 (s, 6H, H<sub>p</sub>), 1.51 (d, *J* = 7.2 Hz, 3H, H<sub>j</sub>), 1.29 (s, 9H, H<sub>l</sub>), 1.21 (s, 9H, H<sub>l</sub>); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ = 178.3, 176.7, 172.0, 169.3, 166.8, 149.4, 147.1, 146.4, 135.7, 135.7, 134.9, 134.4, 134.4, 133.3, 132.3, 130.5, 130.4, 127.3, 127.2, 127.2, 125.9, 121.5, 119.9, 119.9, 82.7, 79.9, 55.7, 48.2, 42.3, 39.5, 38.6, 38.0, 30.8, 27.6, 27.4, 18.2. LRESI(+)-MS (MeOH): *m/z* 779.90 [M+Na]<sup>+</sup>; HRFT(+)-MS: *m/z* 757.3964 [M+H]<sup>+</sup>, 757.3959 calc. for C<sub>46</sub>H<sub>53</sub>N<sub>4</sub>O<sub>6</sub>.

**Ethynyl- *N*-Boc-Leucine-*N*-Piv-Alanine-*N*-Pivaloyl Barrier 17**

Tentagel TBTA (750 mg) resin was loaded with  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (0.12 mmol/g, 0.20 equiv.) in  $\text{CH}_2\text{Cl}_2$  and washed with  $\text{Et}_2\text{O}$ . After drying, the resin was swollen in degassed  $\text{CH}_2\text{Cl}_2$  (2 ml) for 30 min. A degassed solution of **8** (365 mg, 0.44 mmol, 1.00 equiv.) and **9** (333 mg, 0.44 mmol, 1.00 equiv.) in a mixture of  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (1:1 v/v, 7 ml) was added and the reaction mixture was shaken at room temperature for 48 h, then filtered and the resin was washed with a mixture of  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  (2:1 v/v, 6 ml). The combined organic filtrates were washed with aqueous 0.01M  $\text{Na}_4\text{EDTA}$  ( $2 \times 20$  ml), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. The residue was dissolved in acetone (150 ml) and a solution of  $\text{AgNO}_3$  (85 mg, 0.50 mmol, 1.14 equiv.) in a mixture of acetone/ $\text{H}_2\text{O}$  (4:1 v/v, 60 ml) was added. The mixture was stirred in the dark at room temperature for 18 h. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (200 ml). The organic layer was washed with aqueous 1M  $\text{NaCN}$  (100 ml) and  $\text{H}_2\text{O}$  ( $2 \times 200$ ml), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [Analtech GF],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) of the residue afforded **17** (296 mg, 43%) as a pale yellow film.

$^1\text{H}$  NMR (500 MHz, acetone- $d_6$ )  $\delta$  = 9.48 (bs, 1H,  $\text{H}_o$ ), 9.43 (bs, 1H,  $\text{H}_{ag}$ ), 9.03 (bs, 1H,  $\text{H}_u$ ), 8.52 (bs, 1H,  $\text{H}_{am}$ ), 8.08 (d,  $J$  = 8.1 Hz, 1H,  $\text{H}_d$ ), 8.04 (d,  $J$  = 8.4 Hz, 3H,  $\text{H}_{v+x}$ ), 7.98 (d,  $J$  = 8.4 Hz, 2H,  $\text{H}_w$ ), 7.88 (d,  $J$  = 8.4 Hz, 2H,  $\text{H}_c$ ), 7.86 (d,  $J$  = 8.6 Hz, 2H,  $\text{H}_i$ ), 7.61 - 7.56 (m, 4H,  $\text{H}_{p+al}$ ), 7.56 - 7.52 (m, 4H,  $\text{H}_{b+ai}$ ), 7.49 (d,  $J$  = 8.4 Hz, 2H,  $\text{H}_s$ ), 7.24 - 7.41 (bd,  $J$  = 8.2 Hz, 4H,  $\text{H}_{g+z}$ ), 7.25-7.19 (m, 3H,  $\text{H}_{q+ad}$ ), 7.17 (d,  $J$  = 8.6 Hz, 2H,  $\text{H}_{ai}$ ); 7.14 (d,  $J$  = 8.7 Hz, 2H,  $\text{H}_{ak}$ ); 7.03 (bd, 4H,  $\text{H}_{h+aa}$ ), 6.44 (d,  $J$  = 8.1 Hz, 1H,  $\text{H}_m$ ), 5.05-4.99 (m, 2H,  $\text{H}_{e+af}$ ), 4.54 (p,  $J$  = 7.2 Hz, 1H,  $\text{H}_{ab}$ ), 4.36 (td,  $J$  = 9.1 Hz, 5.6 Hz, 1H,  $\text{H}_j$ ), 3.82 (s, 1H,  $\text{H}_a$ ), 3.38-3.34 (m, 2H,  $\text{H}_{f+y}$ ), 3.23-3.18 (m, 2H,  $\text{H}_{f+y}$ ), 1.91-1.81 (m, 1H,  $\text{H}_k$ ), 1.80-1.74 (m, 2H,  $\text{H}_j$ ), 1.73 (s, 6H,  $\text{H}_r$ ), 1.63 (s, 6H,  $\text{H}_{ai}$ ), 1.48 (d,  $J$  = 7.3 Hz, 3H,  $\text{H}_{ac}$ ), 1.40 (s, 9H,  $\text{H}_n$ ), 1.27 (s, 9H,  $\text{H}_{am}$ ), 1.17 (s, 9H,  $\text{H}_{ae}$  A32), 0.97 (t,  $J$  = 6.7 Hz, 6H,  $\text{H}_l$ );  $^{13}\text{C}$  NMR (125 MHz, acetone- $d_6$ )  $\delta$  = 178.6, 176.9, 172.7, 172.4, 170.4, 170.4, 167.2, 166.8, 156.7, 152.6, 150.8, 50.7, 147.8, 147.1, 146.4, 146.2, 138.1, 137.7, 137.3, 136.3, 136.3, 135.8, 135.3, 134.7, 134.6, 132.7, 131.2, 131.2, 129.0, 129.0, 128.5, 127.9, 127.8, 127.5, 126.18 2, 126.1, 122.3, 120.7, 120.4,

120.4, 120.3, 120.2, 83.5, 81.4, 79.4, 56.0, 53.4, 49.5, 43.3, 42.8, 41.0, 40.1, 38.9, 37.9, 31.1, 31.0, 28.6, 27.8, 27.7, 25.6, 23.3, 21.8, 17.2; LRESI(+)-MS (MeOH): 1535.5 [M+H]<sup>+</sup>.

**Ethynyl- *N*-Boc-Phenylalanine-*N*-Boc-Leucine-*N*-Piv-Alanine-*N*-Pivaloyl Barrier 6**

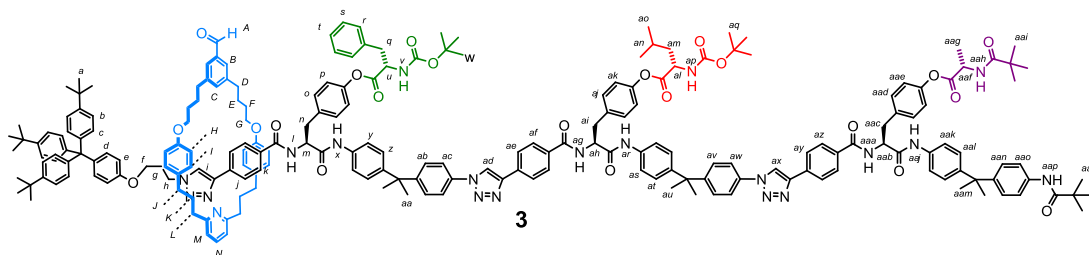
Tentagel TBTA (750 mg) resin was loaded with  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (0.12 mmol/g, 0.28 equiv.) in  $\text{CH}_2\text{Cl}_2$  and washed with  $\text{Et}_2\text{O}$ . After drying, the resin was swollen in degassed  $\text{CH}_2\text{Cl}_2$  (5 ml) for 30 min. A degassed solution of **7** (282 mg, 0.32 mmol, 1.00 equiv.) and **17** (494 mg, 0.32 mmol, 1.00 equiv.) in a mixture of  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (1:1 v/v, 7 ml) was added. The mixture was shaken at room temperature for 48 h, filtered, and the resin was washed with a mixture of  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  (2:1 v/v, 6 ml). The combined organic filtrates were washed with aqueous 0.01M  $\text{Na}_4\text{EDTA}$  ( $2 \times 20$  ml), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. The residue was dissolved in acetone (150 ml) and a solution of  $\text{Ag}(\text{I})\text{NO}_3$  (66 mg, 0.39 mmol, 1.20 equiv.) in a mixture of acetone/ $\text{H}_2\text{O}$  (4:1 v/v, 60 ml) was added. The reaction mixture was stirred in the dark at room temperature for 18 h. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (200 ml) and the organic layer was washed with aqueous 1M  $\text{NaCN}$  (100 ml) and  $\text{H}_2\text{O}$  ( $2 \times 200\text{ml}$ ), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [Merck],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) of the residue afforded **6** (280 mg, 37%) as pale yellow film.

$^1\text{H}$  NMR (500 MHz, acetone- $d_6$ )  $\delta$  = 9.53 (bs, 1H,  $\text{H}_{aj}$ ), 9.49 (bs, 1H,  $\text{H}_p$ ), 9.45 (bs, 1H,  $\text{H}_{aab}$ ), 9.04 (s, 2H,  $\text{H}_{aah+ap}$ ), 8.55 (s, 1H,  $\text{H}_{az}$ ), 8.12 (bs, 1H,  $\text{H}_{as}$ ), 8.10 (bs, 1H,  $\text{H}_d$ ), 8.09 (bs, 1H,  $\text{H}_y$ ), 8.05 (d,  $J = 8.0$  Hz, 4H,  $\text{H}_{w+aq}$ ), 8.00 (d,  $J = 8.0$  Hz, 2H,  $\text{H}_x$ ), 7.99 (d,  $J = 8.0$  Hz, 2H,  $\text{H}_{ar}$ ), 7.89 (d,  $J = 8.3$  Hz, 2H,  $\text{H}_b$ ), 7.87 (d,  $J = 8.5$  Hz, 4H,  $\text{H}_{u+ao}$ ), 7.67 – 7.55 (m, 4H,  $\text{H}_{q+ak}$ ), 7.56 – 7.51 (m, 6H,  $\text{H}_{aac+aag+c}$ ), 7.50 (d,  $J = 8.3$  Hz, 4H,  $\text{H}_{t+an}$ ), 7.45 (d,  $J = 8.3$  Hz, 2H,  $\text{H}_g$ ), 7.43 – 7.36 (m, 4H,  $\text{H}_{ab+av}$ ), 7.36 – 7.32 (m, 4H,  $\text{H}_{k+l}$ ), 7.32 – 7.20 (m, 6H,  $\text{H}_{r+m+al}$ ), 7.18 (d,  $J = 8.1$  Hz, 2H,  $\text{H}_{aad}$ ), 7.16 (d,  $J = 8.6$  Hz, 2H,  $\text{H}_{aaf}$ ), 7.06 (d,  $J = 8.6$  Hz, 2H,  $\text{H}_h$ ), 7.04 (d,  $J = 8.5$  Hz, 2H,  $\text{H}_{ac}$ ), 6.97 (d,  $J = 8.3$  Hz, 2H,  $\text{H}_{aw}$ ), 6.90 (d,  $J = 8.4$ , 1H,  $\text{H}_{ah}$ ), 6.66 (d,  $J = 8.5$ , 1H,  $\text{H}_{ah}$ ), 5.10 – 5.02 (m, 3H,  $\text{H}_{e+z+ai}$ ), 4.62 – 4.51 (m, 2H,  $\text{H}_{i+ax}$ ), 4.42 – 4.33 (m, 1H,  $\text{H}_{ad}$ ), 3.84 (s, 1H,  $\text{H}_a$ ), 3.36 – 3.22 (m, 8H,  $\text{H}_{f+j+aa+au}$ ), 1.89 – 1.82 (m, 1H,  $\text{H}_{af}$ ), 1.80 – 1.75 (m, 2H,  $\text{H}_{ae}$ ), 1.74 (s, 12H,  $\text{H}_{s+am}$ ), 1.64 (s, 6H,  $\text{H}_{aee}$ ), 1.50 (d,  $J = 7.3$  Hz, 3H,  $\text{H}_{ay}$ ), 1.41 (s, 9H,  $\text{H}_o$  or  $\text{ai}$ ), 1.38 (s, 9H,  $\text{H}_o$  or  $\text{ai}$ ), 1.27 (s, 9H,  $\text{H}_{aai}$ ), 1.19 (s, 9H,  $\text{H}_{aaa}$ ), 0.99 (t,  $J = 6.6$  Hz, 6H,  $\text{H}_{ag}$ );  $^{13}\text{C}$  NMR (125 MHz, acetone- $d_6$ )  $\delta$  = 177.7, 176.1, 171.8, 171.5, 170.7, 169.7, 169.6, 169.6, 166.4, 166.3, 166.0, 155.8, 155.5, 151.7, 149.8, 149.8, 149.7, 146.9,

146.2, 145.5, 145.4, 137.2, 137.2, 136.4, 135.6, 135.4, 135.4, 134.9, 134.4, 133.8, 133.8, 133.7, 133.6, 133.6, 131.8, 130.4, 130.3, 129.4, 128.4, 128.1, 128.1, 127.6, 127.0, 126.9, 126.7, 125.3, 121.4, 119.8, 119.6, 119.5, 119.3, 82.6, 80.5, 78.7, 78.5, 56.0, 55.9, 55.6, 52.6, 48.6, 42.4, 41.9, 40.1, 39.2, 38.0, 37.2, 37.0, 37.0, 30.2, 30.1, 27.7, 27.6, 26.9, 26.8, 24.7, 22.4, 20.9, 16.3; LRESI(-)-MS (MeOH): LRESI(-)MS (MeOH):  $m/z$  2337.9 [M-H]<sup>-</sup>.

Aldehyde *N*-Boc-Phenylalanine-*N*-Boc-Leucine-*N*-Piv-Alanine Three-Barrier Rotaxane

3

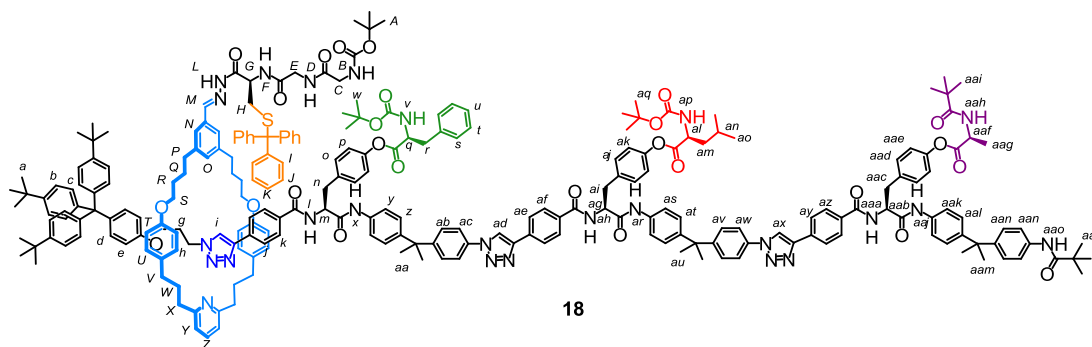


Macrocycle **5** (89 mg, 0.16 mmol, 5.00 equiv.) and  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (5.9 mg, 16  $\mu\text{mol}$ , 0.5 equiv.) were stirred in degassed  $\text{CH}_2\text{Cl}_2$  (4 ml) at room temperature for 30 minutes. Then, a solution of alkyne **6** (73 mg, 0.032 mmol, 1.00 equiv.) and azide **4** (37 mg, 0.063 mmol, 2.00 equiv.) in degassed  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v, 38 ml) was added. The reaction mixture was stirred at room temperature for 7 d. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (20 ml). The organic layer was washed with 0.1M  $\text{Na}_4\text{EDTA}$  (20 ml) and the aqueous layer was extracted with  $\text{CH}_2\text{Cl}_2$  ( $3 \times 20$  ml). The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [Merck],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) of the residue afforded **3** (35 mg, 32%) as a colourless solid.

$^1\text{H}$  NMR (500 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 9.96 (s, 1H,  $\text{H}_A$ ); 9.54 (s, 1H,  $\text{H}_{ar}$ ), 9.49 (s, 1H,  $\text{H}_x$ ), 9.47 (s, 1H,  $\text{H}_{aaj}$ ); 9.02 (s, 2H,  $\text{H}_{ad+ax}$ ); 8.54 (s, 1H,  $\text{H}_{aap}$ ); 8.11 (bs, 3H,  $\text{H}_{aaa+l+ag}$ ); 8.04 (d,  $J = 7.7$  Hz, 4H,  $\text{H}_{af+az}$ ), 8.00 (d,  $J = 7.4$  Hz, 4H,  $\text{H}_{ae+ay}$ ), 7.94 – 7.88 (m, 2H,  $\text{H}_j$ ), 7.86 (d,  $J = 8.4$  Hz, 4H,  $\text{H}_{ac+av}$ ), 7.62 (d,  $J = 9.1$  Hz, 2H,  $\text{H}_{as}$ ); 7.60 (d,  $J = 9.1$  Hz, 4H,  $\text{H}_{y+aak}$ ), 7.61 – 7.53 (m, 7H,  $\text{H}_{aao+i+k+B}$ ), 7.49 (d,  $J = 7.7$  Hz, 4H,  $\text{H}_{ab+av}$ ), 7.54 – 7.38 (m, 7H,  $\text{H}_{o+aj+aad+N}$ ), 7.35 – 7.30 (m, 3H,  $\text{H}_{s+t}$ ), 7.34 – 7.28 (m, 6H,  $\text{H}_b$ ); 7.24 (bd, 7H,  $\text{H}_{z+at+r+aaH}$ ), 7.18 (d,  $J = 9.2$  Hz, 3H,  $\text{H}_{aan+C}$ ), 7.15 (d,  $J = 9.2$  Hz, 2H,  $\text{H}_{aal}$ ), 7.12 (m, 6H,  $\text{H}_c$ ); 7.06 (d,  $J = 9.0$  Hz, 2H,  $\text{H}_p$ ), 7.04 (d,  $J = 9.0$  Hz, 3H,  $\text{H}_{ak+M}$ ), 6.98 (d,  $J = 8.1$  Hz, 2H,  $\text{H}_{aae}$ ), 6.99 – 6.93 (m, 2H,  $\text{H}_d$ ); 6.77 (d,  $J = 8.4$  Hz, 4H,  $\text{H}_H$ ); 6.53 – 6.47 (m, 8H,  $\text{H}_{v+ap+e+l}$ ); 5.07 (bs, 3H,  $\text{H}_{m+ah+aab}$ ); 4.61 – 4.51 (m, 2H,  $\text{H}_{u+aaF}$ ); 4.41 – 4.34 (m, 1H,  $\text{H}_{al}$ ); 4.10 – 4.03 (m, 2H,  $\text{H}_p$ ); 3.82 – 3.76 (m, 4H,  $\text{H}_G$ ); 3.53 – 3.46 (m, 2H,  $\text{H}_h$ ); 3.42 – 3.36 (m, 3H,  $\text{H}_{n+ai+aac}$ ); 3.27 – 3.18 (m, 5H,  $\text{H}_{n'+ai'+aac'+q}$ ); 2.76 – 2.71 (m, 4H,  $\text{H}_D$ ); 2.65 – 2.58 (m, 4H,  $\text{H}_L$ ); 2.55 – 2.49 (m, 4H,  $\text{H}_J$ ); 1.91 – 1.84 (m, 6H,  $\text{H}_{an+K}$ ); 1.74 – 1.84 (m, 12H,  $\text{H}_{g+am+E+F}$ ); 1.74 (s, 12H,  $\text{H}_{aa+au}$ ); 1.64 (s, 6H,  $\text{H}_{aam}$ ); 1.49 (d,  $J = 7.3$  Hz, 3H,  $\text{H}_{aaG}$ ); 1.41 (s, 9H,  $\text{H}_{aq}$  or  $w$ ); 1.37 (s, 9H,  $\text{H}_{aq}$  or  $w$ ); 1.31 (s, 27H,  $\text{H}_A$ ); 1.28 (s, 9H,  $\text{H}_{aaq}$ ); 1.19 (s, 9H,  $\text{H}_{aai}$ ); 0.99 (dd,  $J = 5.9$  Hz, 12.6 Hz, 2H,  $\text{H}_{an}$ ), 6H,  $\text{H}_{ao}$ );  $^{13}\text{C}$  NMR (125 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 192.3, 177.8, 176.1,

171.8, 171.5, 170.7, 169.7, 169.7, 169.7, 166.4, 166.4, 166.4, 157.2, 155.9, 155.5, 151.7, 149.9, 149.8, 149.6, 148.2, 146.9, 146.2, 146.0, 145.6, 145.4, 145.3, 144.4, 143.8, 137.2, 137.2, 137.1, 136.8, 136.4, 135.5, 135.5, 135.4, 134.9, 134.7, 133.9, 133.8, 133.6, 133.6, 133.6, 133.3, 132.9, 131.7, 130.5, 130.4, 130.3, 130.3, 129.4, 129.1, 129.1; 128.4, 128.1, 127.9, 127.1, 127.0, 126.9, 126.7, 126.7, 125.3, 125.0, 124.2, 121.4, 119.8, 119.6, 119.5, 119.5, 119.3, 114.1, 113.2, 78.7, 78.5, 67.1, 64.0, 63.0, 56.0, 55.9, 55.9, 55.6, 52.6, 48.6, 46.7, 42.4, 41.9, 40.2, 39.2, 38.0, 37.2, 37.0, 37.0, 34.9, 34.0, 32.0, 31.7, 30.8, 30.2, 30.1, 29.4, 29.2, 28.3, 27.7, 27.6, 26.9, 26.8, 24.7, 22.4, 20.9, 16.3; LRESI(+)MS (MeOH):  $m/z$  3478.4 [M+Na]<sup>+</sup>.

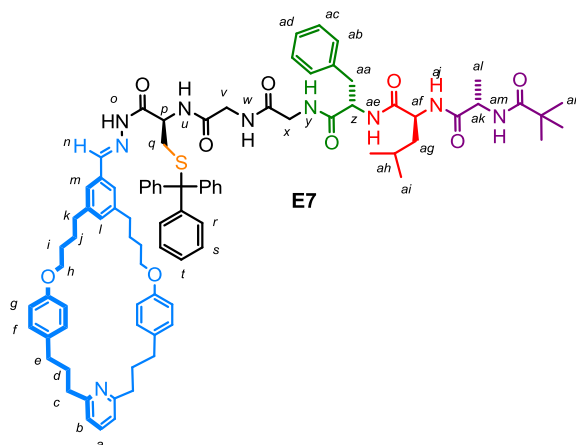
***N*-Boc-Gly-Gly-Cys(STrt)-Acyl Hydrazone *N*-Boc-Phenylalanine-*N*-Boc-Leucine-*N*-Piv-Alanine Three-Barrier Rotaxane **18****



A mixture of rotaxane **3** (30 mg, 7.5  $\mu\text{mol}$ , 1.0 equiv.), *p* CH<sub>3</sub>OC<sub>6</sub>H<sub>4</sub>CH=NNH-Cys(STrt)-Gly-Gly-*N*-Boc **10** (8.0 mg, 11  $\mu\text{mol}$ , 1.5 equiv.) and aniline (50  $\mu\text{l}$ ) in DMSO/2-(*N*-morpholino)ethanesulfonic acid buffer (3:1 *v/v*, 1 ml, pH 6.0) was stirred at 60 °C for 72 h. The reaction mixture was allowed to cool to room temperature and then diluted with CH<sub>2</sub>Cl<sub>2</sub> (10 ml) and the organic layer was washed with H<sub>2</sub>O (10 ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent removed under reduced pressure. Preparative TLC (SiO<sub>2</sub> [UNIPLATE-T], CH<sub>2</sub>Cl<sub>2</sub>/MeOH 5%) of the residue afforded **18** (21 mg, 90%) as a colourless solid.

<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  = 11.45 (s, 1H, H<sub>L</sub>), 10.33 – 10.12 (m, 3H, H<sub>x+ar+aa</sub>), 9.35 (s, 2H, H<sub>ad+ax</sub>), 9.14 (s, 1H, H<sub>aap</sub>), 8.83 (t, 1H, *J* = 7.04 Hz, H<sub>D</sub>), 8.80 – 8.70 (m, 2H, H<sub>l+aa</sub>), 8.41 (d, *J* = 8.4 Hz, 1H, H<sub>Fmin</sub>), 8.32 (d, *J* = 8.4 Hz, 1H, H<sub>Fmaj</sub>), 8.18 (s, 1H, H<sub>Mmin</sub>), 8.17 (s, 1H, H<sub>Mmaj</sub>), 8.05 – 7.90 (m, 12H, H<sub>i+ae+af+ay+az+ag+aaa</sub>), 7.85 (d, *J* = 8.3 Hz, 8H, H<sub>p+ak+aae+k</sub>), 7.81 – 7.75 (m, 2H, H<sub>j</sub>), 7.60 – 7.42 (m, 15H, H<sub>Z+N+y+ac+as+aw+aak+aaa</sub>), 7.37 – 7.07 (m, 36, H<sub>K+J+l+z+ab+at+av+aal+aan+b+Y+O</sub>), 7.18 – 6.96 (m, 17H, H<sub>c+o+aj+aad+r+s+t</sub>), 6.94 – 6.88 (m, 2H, H<sub>aad</sub>), 6.80 – 6.74 (m, 2H, H<sub>d</sub>), 6.73 – 6.44 (m, 4H, H<sub>U</sub>), 6.40 – 6.33 (m, 4H, H<sub>T</sub>), 6.26 (d, *J* = 9.0 Hz, 2H, H<sub>e</sub>), 5.40 – 5.33 (m, 1H, H<sub>Gmaj</sub>), 4.92 – 4.82 (m, 3H, H<sub>m+ah+aab</sub>), 4.83 – 4.76 (m, 1H, H<sub>Gmin</sub>), 4.43 – 4.31 (m, 2H, H<sub>u+aa</sub>), 4.19 – 4.12 (m, 1H, H<sub>al</sub>), 4.06 – 3.96 (m, 2H, H<sub>h</sub>), 3.85 – 3.48 (m, 8H, H<sub>C+E+S+h</sub>), 3.22 – 2.93 (m, 10H, H<sub>n+ai+aac+H+q</sub>), 2.49 – 2.31 (m, 12H, H<sub>X+V+P</sub>), 1.83 – 1.53 (m, 32H, H<sub>aa+au+aaam+R+Q+W</sub>), 1.44 – 1.19 (m, 66H, H<sub>a+A+w+aq+aaq+an+am</sub>), 1.11 (s, 9H, H<sub>aa</sub>), 0.90 (dd, *J* = 14.9 Hz, 6.3 Hz, 6H, H<sub>ao</sub>); <sup>13</sup>C NMR (125 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  = 178.2, 176.7, 172.4, 172.1, 171.5, 170.7, 169.2, 166.5, 161.5, 156.9, 156.3, 156.1, 156.0, 156.0, 151.8, 149.6, 149.4, 148.1, 146.9, 146.0, 146.0, 145.6, 145.2, 144.6, 144.6, 144.4, 144.4, 143.2, 139.1, 137.7, 137.4, 137.1, 136.8, 136.3, 134.7, 133.8, 133.5, 131.5, 130.7, 130.7, 130.4, 129.7, 129.5, 129.4, 129.3, 128.7, 128.7, 128.6, 128.6, 128.5, 128.5, 128.4, 127.3, 127.2, 126.8, 125.4, 125.1, 124.8, 124.6, 121.6, 121.5, 120.9, 120.3, 120.3, 120.1,

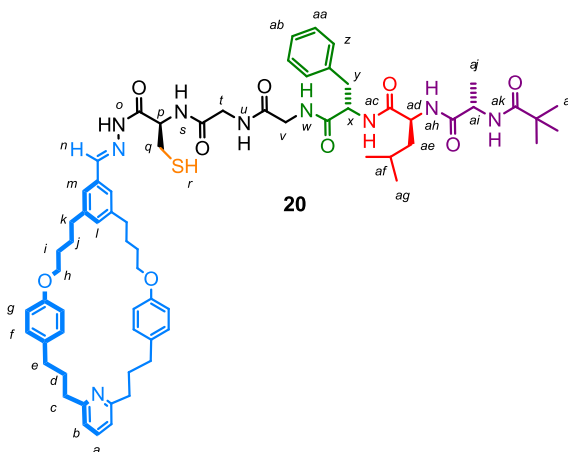
119.7, 119.6, 119.5, 115.4, 114.4, 113.5, 79.0, 78.9, 78.5, 67.3, 66.5, 66.1, 64.1, 62.9, 56.2, 56.0, 52.7, 48.7, 46.8, 43.7, 42.7, 42.1, 38.3, 37.6, 37.0, 36.7, 35.0, 34.7, 34.5, 34.5, 32.0, 31.6, 31.0, 30.8, 29.4, 28.7, 28.7, 28.6, 28.6, 28.1, 27.7, 27.7, 26.8, 24.8, 23.3, 21.7, 16.8, 0.6. *Please Note: MS analysis (LRESI, MALDI) was unsuccessful.*

***N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(STrt)-Acyl Hydrazone Macrocycle **E7****

A mixture of macrocycle **5** (10 mg, 18  $\mu\text{mol}$ , 1.0 equiv.), *p*-CH<sub>3</sub>OC<sub>6</sub>H<sub>4</sub>CH=NNH-Cys(STrt)-Gly-Gly-Phe-Leu-Ala-*N*-Piv **E6** (18 mg, 18  $\mu\text{mol}$ , 1.0 equiv.) and aniline (50  $\mu\text{l}$ ) in DMSO-2-(*N*-morpholino)ethanesulfonic acid buffer (3:1 v/v, 4 ml, pH 6.0) was stirred at 60 °C for 2 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (20 ml) and the organic layer was washed with H<sub>2</sub>O (2  $\times$  20 ml). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent removed under reduced pressure. Preparative TLC (SiO<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>/MeOH 4%) of the residue afforded **E7** (14 mg, 75%) as a colourless film.

9:10 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone.  $[\alpha]_{\text{D}} = -0.26^\circ$  (c 0.30, MeOH); <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>)  $\delta = 11.47$  (s, 1H, H<sub>o1</sub>), 11.45 (s, 1H, H<sub>o2</sub>), 8.41 – 8.19 (m, 6H, H<sub>w1+w2+y1+y2+ae1+ae2</sub>), 8.18 (s, 1H, H<sub>n1</sub>), 8.05 – 8.00 (m, 2H, H<sub>u1+u2</sub>), 7.94 (s, 1H, H<sub>n2</sub>), 7.69 – 7.65 (m, 2H, H<sub>am1+am2</sub>), 7.60 – 7.55 (m, 2H, H<sub>a1+a2</sub>), 7.42 – 7.38 (m, 2H, H<sub>aj1+aj2</sub>), 7.37 – 7.29 (m, 12H, H<sub>m1+m2+ab1+ab2+ac1+ac2+ad1+ad2</sub>), 7.28 – 7.19 (m, 18H, H<sub>r1+r2+t1+t2</sub>), 7.18 – 7.09 (m, 8H, H<sub>l1+l2+s1+s2</sub>), 7.07 (d, *J* = 7.5 Hz, 4H, H<sub>b1+b2</sub>), 7.03 (d, *J* = 8.7 Hz, 4H, H<sub>f1</sub>), 7.01 (d, *J* = 8.6 Hz, 4H, H<sub>f2</sub>), 6.74 (d, *J* = 8.7 Hz, 4H, H<sub>g1</sub>), 6.69 (d, *J* = 8.7 Hz, 4H, H<sub>g2</sub>), 5.42 – 5.35 (m, 1H, H<sub>p1</sub>), 4.54 – 4.43 (m, 3H, H<sub>p2+z1+z2</sub>), 4.27 – 4.12 (m, 4H, H<sub>ak1+ak2+af1+af2</sub>), 3.90 (t, *J* = 5.8 Hz, 4H, H<sub>v1+v2</sub>), 3.86 – 3.69 (m, 8H, H<sub>h1+h2+x1+x2</sub>), 3.08 – 3.00 (m, 2H, H<sub>aal+aa2</sub>), 2.86 – 2.78 (m, 2H, H<sub>aal'+aa2'</sub>), 2.70 (t, *J* = 7.3 Hz, 8H, H<sub>k1+k2</sub>), 2.67 – 2.61 (m, 8H, H<sub>c1+c2</sub>), 2.48 – 2.41 (m, 8H, H<sub>e1+e2</sub>), 2.40 – 2.28 (m, 4H, H<sub>q1+q2</sub>), 1.96 – 1.87 (m, 8H, H<sub>d1+d2</sub>), 1.76 – 1.68 (m, 8H, H<sub>j1+j2</sub>), 1.67 – 1.60 (m, 10H, H<sub>i1+i2+ah1+ah2</sub>), 1.52 – 1.43 (m, 4H, H<sub>ag1+ag2</sub>), 1.37 – 1.31 (m, 6H, H<sub>al1+al2</sub>), 1.09 (s, 9H, H<sub>an1</sub>), 1.08 (s, 9H, H<sub>an2</sub>), 0.84 – 0.75 (m, 12H, H<sub>ai</sub>); <sup>13</sup>C NMR (100 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  161.9, 158.2, 145.7, 145.6, 143.8, 143.7, 138.5, 137.3, 135.2, 130.5, 130.4, 130.4, 130.3, 130.3, 130.2, 130.1, 129.7, 129.1, 129.1,

129.0, 128.8, 128.7, 127.6, 127.5, 127.2, 126.1, 120.9, 115.0, 115.0, 68.1, 68.1, 41.3, 41.3, 37.9, 37.9, 35.7, 35.6, 35.1, 32.6, 30.6, 28.6, 27.8, 27.7, 27.0, 27.0, 21.4; *Please Note: Due to the presence of two stereoisomers, some <sup>13</sup>C-signals coincide.* LRESI(+)-MS (MeOH):  $m/z$  1451.2 [M+H]<sup>+</sup>,  $m/z$  1473.3 [M+Na]<sup>+</sup>.

***N*-Piv-Ala-Leu-Phe-Gly-Gly-Cys(SH)-Acyl Hydrazone Macrocycle **20****

Hexapeptide macrocycle **E7** (14 mg, 9.7  $\mu\text{mol}$ , 1.0 equiv.) was dissolved in  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (4:1 v/v, 2 ml). Triisopropylsilane (50  $\mu\text{l}$ ) was added and the mixture was stirred at room temperature for 48 h. The reaction mixture was diluted with toluene (5 ml) and the solvent was removed under reduced pressure. The residue was purified by triturating with  $\text{Et}_2\text{O}$  ( $2 \times 5$  ml) to afford authentic reference sample **20** as colourless solid (11 mg, 99%).

m.p. 117 – 120  $^\circ\text{C}$ ;  $[\alpha]_{\text{D}} = -1.50^\circ$  (c 0.05, MeOH);  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 11.50 (m, 1H,  $\text{H}_o$ ), 8.27 (m, 1H,  $\text{H}_u$ ), 8.20 (m, 1H,  $\text{H}_s$ ), 8.06 (m, 1H,  $\text{H}_s$ ), 7.97 (m, 1H,  $\text{H}_w$ ), 7.69 (m, 2H,  $\text{H}_{ak+ac}$ ), 7.42 (m, 1H,  $\text{H}_a$ ), 7.36 (m, 1H,  $\text{H}_{ah}$ ), 7.21 (m, 7H,  $\text{H}_{m+z+aa+ab}$ ), 7.10 (m, 2H,  $\text{H}_b$ ), 7.03 (d,  $J = 8.5$  Hz, 4H,  $\text{H}_j$ ), 6.88 (bs, 1H,  $\text{H}_i$ ), 6.75 (d,  $J = 8.4$  Hz, 4H,  $\text{H}_g$ ), 4.53 (m, 1H,  $\text{H}_p$ ), 4.47 (m, 1H,  $\text{H}_x$ ), 4.26 (m, 1H,  $\text{H}_{ai}$ ), 4.21 (m, 1H,  $\text{H}_{ad}$ ), 3.90 (m, 2H,  $\text{H}_i$ ), 3.80 (m, 4H,  $\text{H}_h$ ), 3.75 (m, 2H,  $\text{H}_v$ ), 3.06 (m, 2H,  $\text{H}_y$ ), 2.68 (m, 8H,  $\text{H}_{k+c}$ ), 2.38 (m, 2H,  $\text{H}_q$ ), 1.95 (m, 4H,  $\text{H}_d$ ), 1.72 (m, 4H,  $\text{H}_j$ ), 1.68 (m, 5H,  $\text{H}_{l+af}$ ), 1.51 (m, 2H,  $\text{H}_{ae}$ ), 1.36 (m, 3H,  $\text{H}_{aj}$ ), 1.09 (s, 9H,  $\text{H}_{al}$ ), 0.82 (dd,  $J = 18.7$  Hz, 5 Hz, 6H,  $\text{H}_{ag}$ );  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO}-d_6$ )  $\delta$  177.2, 172.5, 171.8, 171.5, 168.9, 156.8, 142.0, 141.7, 138.0, 130.1, 129.2, 128.1, 128.0, 128.0, 126.9, 126.9, 126.7, 126.3, 126.3, 126.2, 114.2, 113.8, 99.5, 67.1, 55.1, 54.1, 53.9, 51.1, 50.8, 50.6, 48.6, 48.2, 45.9, 42.2, 41.9, 40.9, 40.1, 37.9, 37.5, 37.3, 31.2, 28.0, 27.6, 27.3, 24.0, 22.8, 21.7; LRESI(+)-MS:  $m/z$  1208.6  $[\text{M}+\text{H}]^+$ .

### 3.6 References

- (1) Kay, E. R.; Leigh, D. A.; Zerbetto, F. *Angew. Chem. Int. Ed.* **2007**, *46*, 72.
- (2) (a) Collier, C. P.; Wong, E. W.; Belohradský, M.; Raymo, F. M.; Stoddart, J. F.; Kuekes, P. J.; Williams, R. S.; Heath, J. R. *Science* **1999**, *285*, 391 (b) Green, J. E.; Wook Choi, J.; Boukai, A.; Bunimovich, Y.; Johnston-Halperin, E.; DeIonno, E.; Luo, Y.; Sheriff, B. A.; Xu, K.; Shik Shin, Y.; Tseng, H.-R.; Stoddart, J. F.; Heath, J. R. *Nature* **2007**, *445*, 414.
- (3) (a) Berna, J.; Leigh, D. A.; Lubomska, M.; Mendoza, S. M.; Perez, E. M.; Rudolf, P.; Teobaldi, G.; Zerbetto, F. *Nat. Mater.* **2005**, *4*, 704 (b) Eelkema, R.; Pollard, M. M.; Vicario, J.; Katsonis, N.; Ramon, B. S.; Bastiaansen, C. W. M.; Broer, D. J.; Feringa, B. L. *Nature* **2006**, *440*, 163 (c) Muraoka, T.; Kinbara, K.; Aida, T. *Nature* **2006**, *440*, 512 (d) Juluri, B. K.; Kumar, A. S.; Liu, Y.; Ye, T.; Yang, Y.-W.; Flood, A. H.; Fang, L.; Stoddart, J. F.; Weiss, P. S.; Huang, T. J. *ACS Nano* **2009**, *3*, 291.
- (4) (a) Thordarson, P.; Bijsterveld, E. J. A.; Rowan, A. E.; Nolte, R. J. M. *Nature* **2003**, *424*, 915 (b) Monnereau, C.; Ramos, P. H.; Deutman, A. B. C.; Elemans, J. A. A. W.; Nolte, R. J. M.; Rowan, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 1529.
- (5) (a) Würthner, F.; Rebek, J. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 446 (b) Yoon, H. J.; Kuwabara, J.; Kim, J.-H.; Mirkin, C. A. *Science* **2010**, *330*, 66 (c) Schmittel, M.; De, S.; Pramanik, S. *Angew. Chem. Int. Ed.* **2012**, *51*, 3832 (d) Blanco, V.; Carlone, A.; Hänni, K. D.; Leigh, D. A.; Lewandowski, B. *Angew. Chem. Int. Ed.* **2012**, *51*, 5166.
- (6) Wang, J.; Feringa, B. L. *Science* **2011**, *331*, 1429.
- (7) (a) He, Y.; Liu, D. R. *Nat. Nanotechnol.* **2010**, *5*, 778 (b) McKee, M. L.; Milnes, P. J.; Bath, J.; Stulz, E.; O'Reilly, R. K.; Turberfield, A. J. *J. Am. Chem. Soc.* **2012**, *134*, 1446 (c) He, Y.; Liu, D. R. *J. Am. Chem. Soc.* **2011**, *133*, 9972.
- (8) Gu, H.; Chao, J.; Xiao, S.-J.; Seeman, N. C. *Nature* **2010**, *465*, 202.
- (9) (a) Yonath, A. *Angew. Chem. Int. Ed.* **2010**, *49*, 4340 (b) Ramakrishnan, V. *Angew. Chem. Int. Ed.* **2010**, *49*, 4355 (c) Steitz, T. A. *Angew. Chem. Int. Ed.* **2010**, *49*, 4381.
- (10) Sieber, S. A.; Marahiel, M. A. *Chem. Rev.* **2005**, *105*, 715.
- (11) Molecules labelled **1**, **4**, **5**, **10**, **16**, and **E3** correspond to molecules **11**, **13**, **14**, **15**, **29** and **21** in Chapter 2, respectively. Their syntheses and characterisation data are reported in the Experimental Section 2.5.
- (12) von Delius, M.; Leigh, D. A. *Chem. Soc. Rev.* **2011**, *40*, 3656.

- (13) Frank, J.; Gonzalez, R. L. *Annu. Rev. Biochem.* **2010**, *79*, 381.
- (14) Moore, P. B. *Annu. Rev. Biophys.* **2012**, *41*, 1.
- (15) Labeit, S.; Kolmerer, B. *Science* **1995**, *270*, 293.
- (16) Yamamoto, Y. *Chem. Rev.* **2008**, *108*, 3199.
- (17) Zheng, H.; Zhou, W.; Lv, J.; Yin, X.; Li, Y.; Liu, H.; Li, Y. *Chem. –Eur. J.* **2009**, *15*, 13253.
- (18) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. *Nat. Chem.* **2010**, *2*, 490.
- (19) (a) Louris, J. N.; Wright, L. G.; Cooks, R. G.; Schoen, A. E. *Anal. Chem.* **1985**, *57*, 2918 (b) Coon, J. J. S., J. E. P.; Shabanowitz, J.; Hunt, D. F. *BioTechniques* **2005**, *38*, 519.
- (20) Sleno, L.; Volmer, D. *J. Am. Chem. Soc.* **2006**, *128*, 2186.

## Chapter IV

# Rotaxane Elongation as a New Strategy for the Construction of Highly Complex Molecular Machines

### ACKNOWLEDGMENTS

Dr. Guillaume De Bo and Dr. Bartosz Lewandowski are acknowledged for the final steps of the synthesis of rotaxane **11**. Sonja Kuschel is acknowledged for the synthesis of barrier **6**. All other syntheses presented in this Chapter were carried out by Marcus Pappmeyer. Dr. John Ward is acknowledged for the resynthesis of macrocycle **7**.

## 4.1 Synopsis

The ribosome can construct polypeptides with a length up to 10000 amino acid units. However, artificial ribosome mimics have been limited to three individual coupling events. In Chapter 4, the design, synthesis and operation of a molecular machine capable of assembling four individual amino acid units with complete sequence-specificity is described. A more versatile synthetic route based on the elongation of a preformed rotaxane structure was crucial to the success of the project and allows in principle the incorporation of a wide range of otherwise difficult to introduce substrates.

## 4.2 Introduction

In biological systems molecular machines are employed to perform a variety of complex tasks essential for the functioning of living organisms (protein and nucleic acid synthesis, carrying cargo within cells, muscle movements, performing propelling motion etc.).<sup>1</sup> These highly complex molecular machines have been inspiring scientists for more than twenty years to create synthetic analogues in order to better understand the mechanisms behind their operation and to mimic their function in an efficient way.<sup>2</sup>

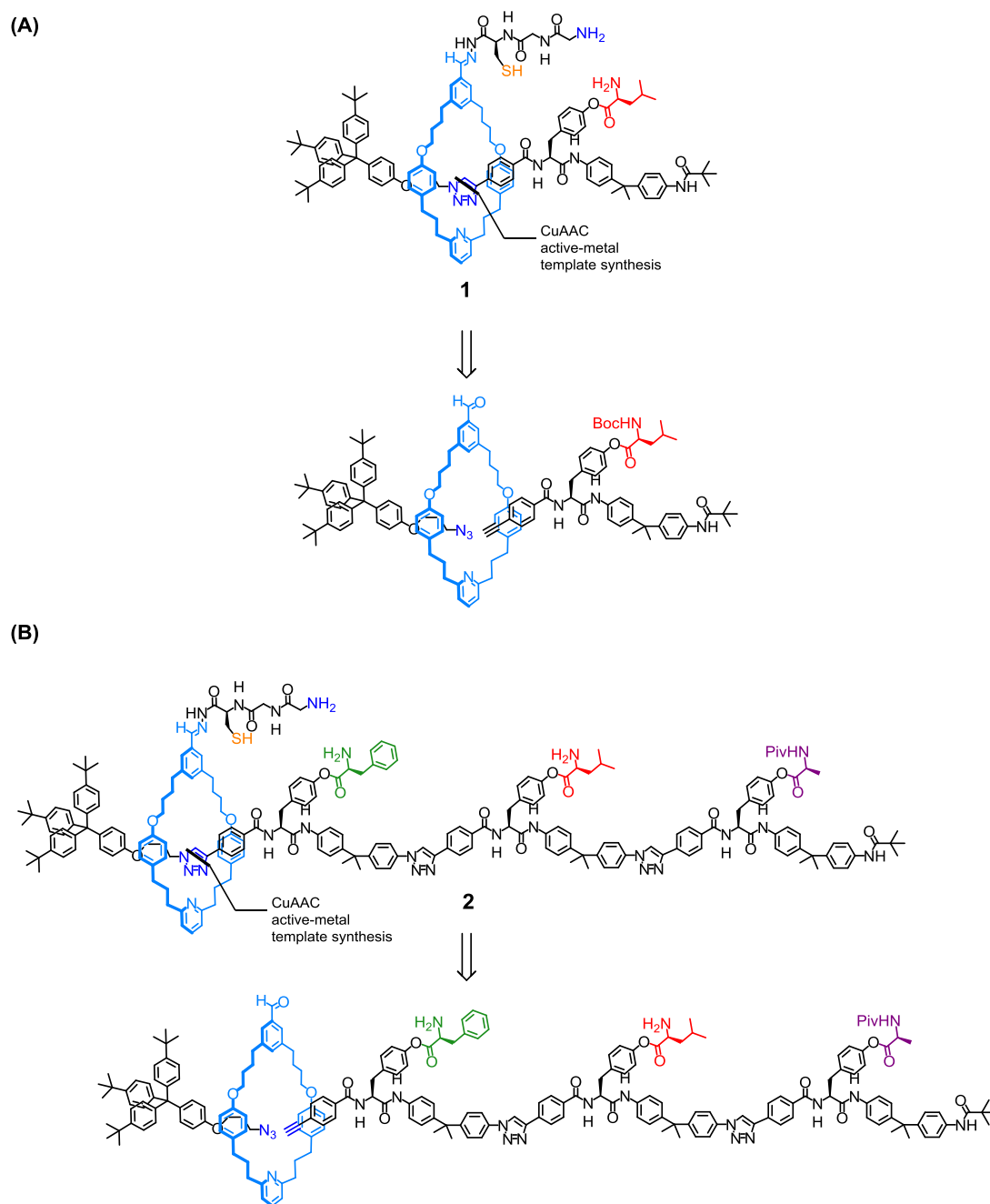
Since then, various artificial small-molecule machines have been successfully prepared and utilised, amongst others, to do mechanical work<sup>3</sup> or as switchable catalytic systems<sup>4</sup>. Processive catalysts have been reported to operate on oligomeric<sup>5</sup> and polymeric<sup>6</sup> tracks but only randomly transform the functional groups within their confines. Small-molecule artificial molecular machines have been prepared making use of systems exhibiting axial chirality and isomerisable alkene moieties or allosteric regulation of organometallic catalysts.<sup>7</sup> Large synthetic DNA molecules have been used to guide the formation of bonds between unnatural building blocks<sup>8</sup> and assemble gold nanoparticles sequence-specifically<sup>9</sup>. In particular, mechanically interlocked systems offer great potential for the synthesis of complex molecular machines and therefore have been incorporated in a large number of molecular devices.<sup>10</sup> In Chapter 3 a rotaxane-based molecular machine capable of performing three peptide forming reactions in a particular order was reported. This very simple synthetic mimic of a ribosome operates processively due to the mechanical bond between the macrocycle containing the catalyst and the axle containing the amino acids and synthesises a single peptidic product in a sequence-specific manner.

In order to explore the limitations of this machine design, the construction of a molecular machine with more than three monomeric units would be desirable. The ultimate target for a system capable of sequence-specific synthesis would be the incorporation of a polymeric thread. Therefore, the synthesis of a molecular machine which generates oligomers with a higher number of monomeric units should not only point the way towards the successful synthesis of increasingly complex molecular machines but also demonstrate the possibility to construct and operate machines with an even higher number of monomeric units. Since the synthesis with the previously employed active metal template approach for the late stage interlocking is impractical for larger systems, a new strategy towards the realisation of machines with longer threads is described.

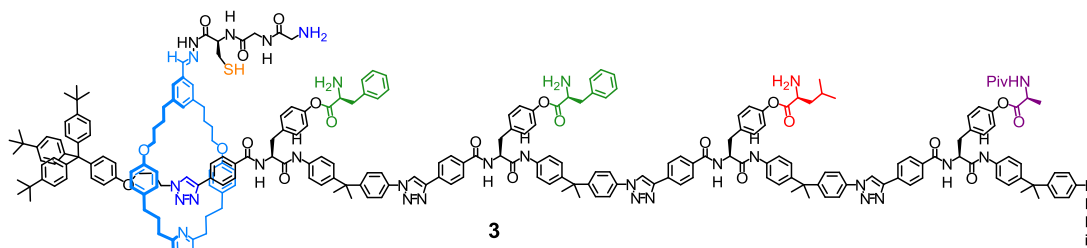
### 4.2.1 Rotaxane Elongation as a Novel Synthetic Tool

The key step in the assembly of either model rotaxane **1** reported in Chapter 2 or fully functional molecular machine **2** reported in Chapter 3 was the interlocking of the macrocycle, azide half-thread and an alkyne bearing the amino acyl residues to be assembled through an active-metal template synthesis (Figure 4.1).<sup>11</sup> In case of single-barrier rotaxane **1**, the yield of this Cu(I)-catalysed azide-alkyne cycloaddition was 45% and complete conversion was achieved after two days. For three-barrier rotaxane **2** the best yield obtained was 30% after seven days of reaction time. Target rotaxane **3** featuring four amino acyl groups (Figure 4.2) would pose an even bigger challenge for an active metal template reaction. A lower yield and significantly longer reaction times are expected. Although the active metal template methodology usually provides interlocked structures in good to excellent yields<sup>12</sup>, in this case the complexity of the alkyne substrate, containing multiple functionalities and a large number of heteroatoms, would be the most likely reason for the diminished efficiency. Hydrolysis of the amino acyl functional groups in the presence of Lewis acidic copper(I)- and the inevitably formed copper(II)-ions due to trace amounts of oxygen in the reaction media during the long reaction time (probably greater than 7 days) are likely side-reactions. The active metal template reaction as the final interlocking step represents therefore a bottleneck for the synthesis of more complex molecular machines.

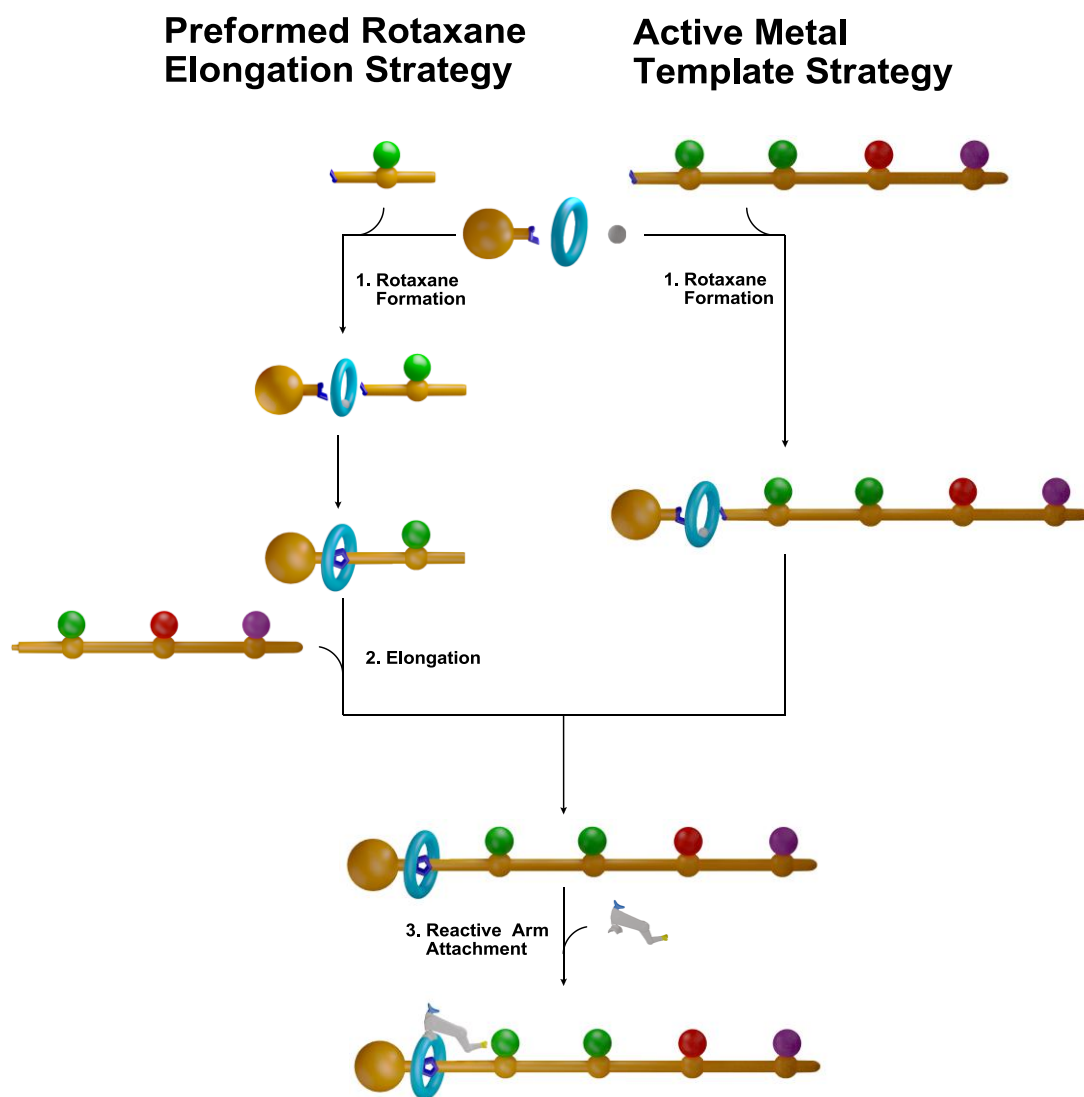
To circumvent these problems, an approach involving the formation of a simpler rotaxane containing a single blocking group on the axle that could in principle be derivatised with threads of any length was envisioned (Figure 4.3).<sup>13</sup> This preformed rotaxane elongation strategy would give access to a wide range of machines that could be utilised to obtain a variety of oligomeric products. Also, this approach would allow the incorporation of functional groups in the thread that are otherwise incompatible with the active metal template strategy such as strongly coordinating groups or polymeric substrates. Moreover, the formerly difficult separation of the desired rotaxane from the corresponding un-interlocked thread would be more readily achieved.



**Figure 4.1** Key disconnection in the synthesis of molecular machines (A) single-barrier system **1** and (B) three-barrier system **2**.



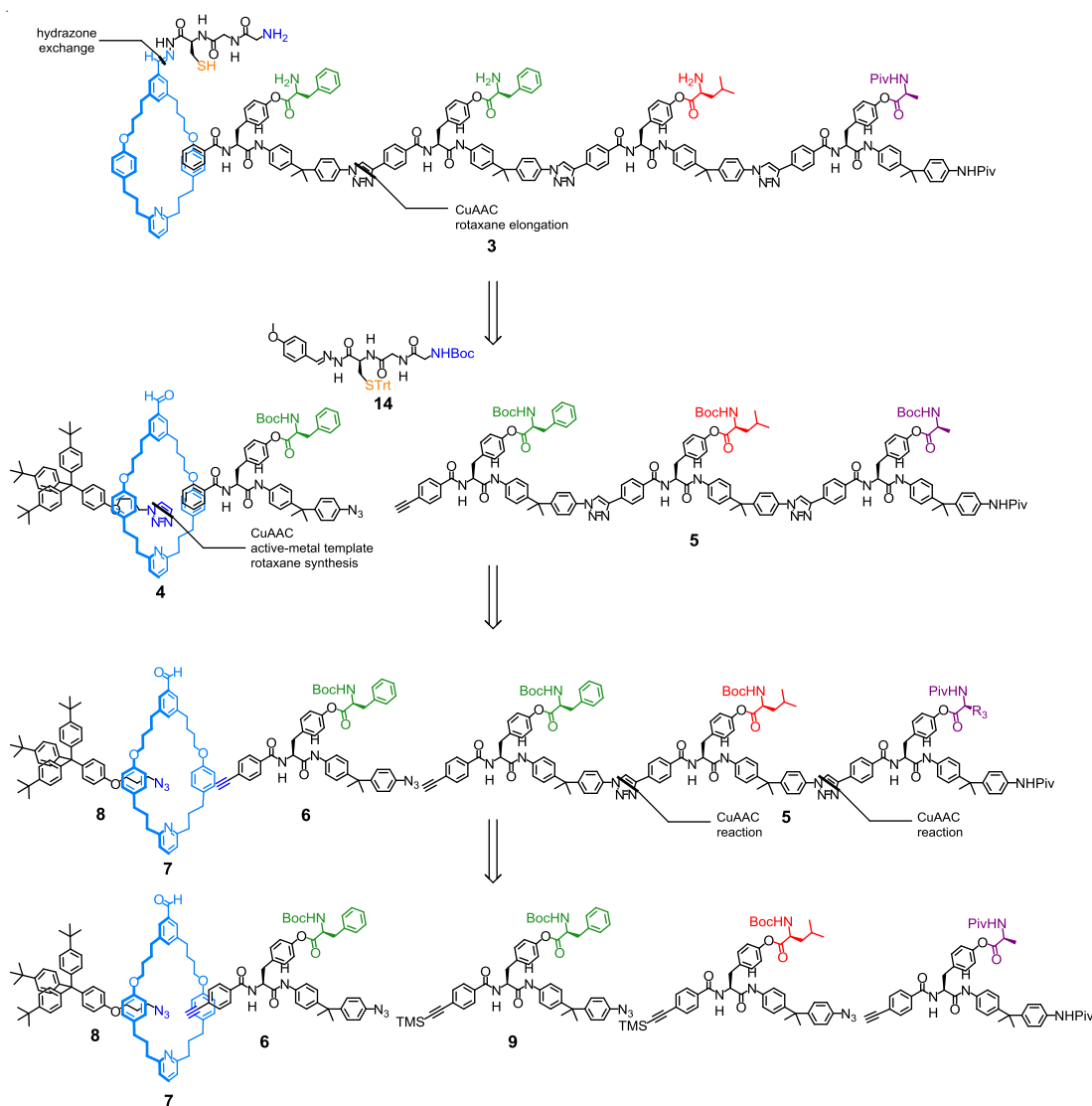
**Figure 4.2.** Structure of molecular machine **3**.



**Figure 4.3.** Two Strategies towards the synthesis of a four barrier molecular machine: the *preformed rotaxane elongation strategy* (left) and the *active metal template synthesis* (right).

### 4.2.2 Retrosynthetic Analysis of the Four-Barrier Rotaxane 3

The retrosynthesis of molecular machine **3** is depicted in Scheme 4.1. The incorporation of the reactive arm *via a* hydrazone exchange reaction in the last step has proven to be successful previously for molecular machines **1** and **2** and leaves therefore the corresponding four-barrier aldehyde rotaxane as starting compound. The aldehyde rotaxane can be disconnected into azide functionalised single-barrier rotaxane **4** featuring only an individual blocking group and alkyne thread **5**. Single-barrier rotaxane **4** can be derived from an active metal template reaction of *N*-Boc-phenylalanine-loaded alkyne barrier **6**, macrocycle **7** and a large excess of alkyl-azide stopper **8**.

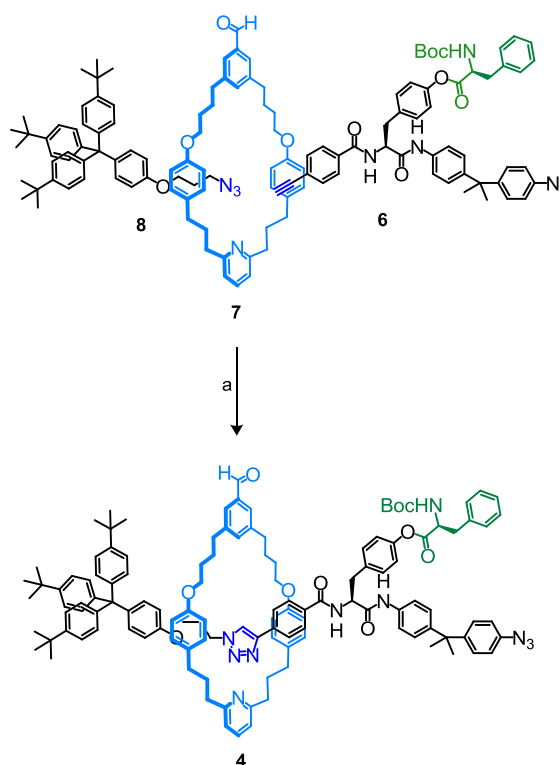


**Scheme 4.1.** Retrosynthetic analysis of molecular machine **3**.

## 4.3 Results and Discussion

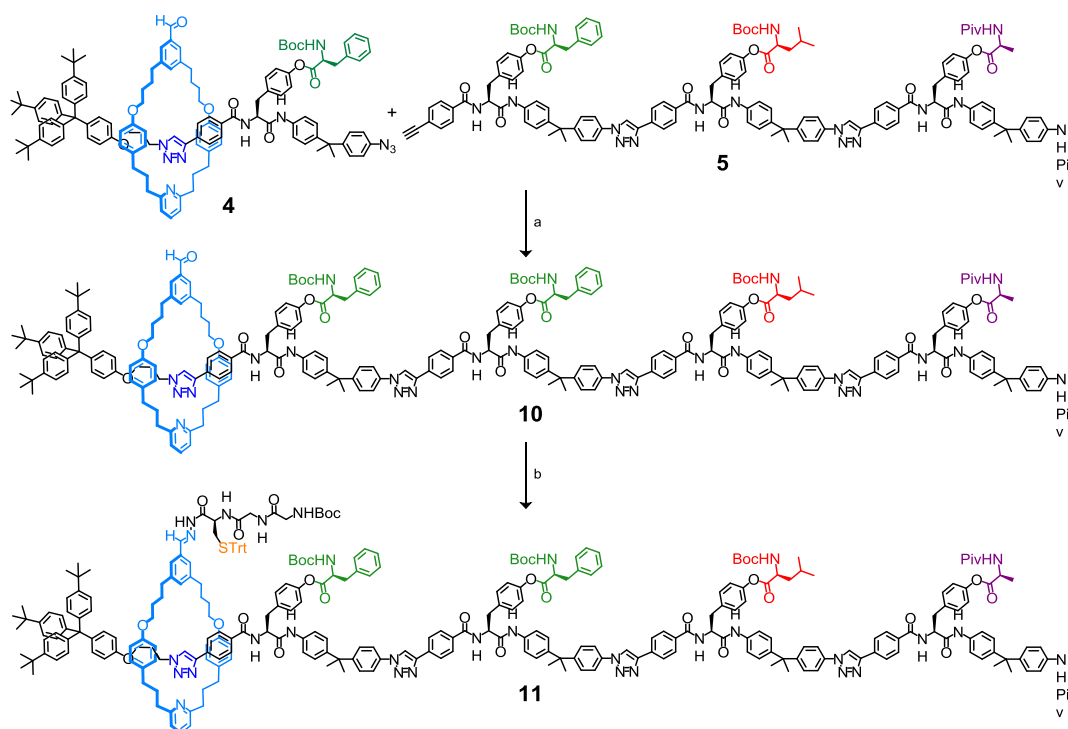
### 4.3.1 Synthesis of Molecular Machine 11

The synthesis of molecular machine **3** commenced with the synthesis of rotaxane **4** (Scheme 4.2). The CuAAC active-metal synthesis of barrier **6**, derived from its TMS-protected analogue **9** by an Ag(I)NO<sub>3</sub>-mediated TMS-deprotection, two equivalents of macrocycle **7** and ten equivalents of azide stopper **8** with Cu(I)(CH<sub>3</sub>CN)<sub>4</sub>PF<sub>6</sub> as the copper source gave the desired rotaxane **4** in 40% yield. Gratifyingly the excess azide stopper **8** as well as non-interlocked macrocycle **7** could be regenerated after column chromatography.



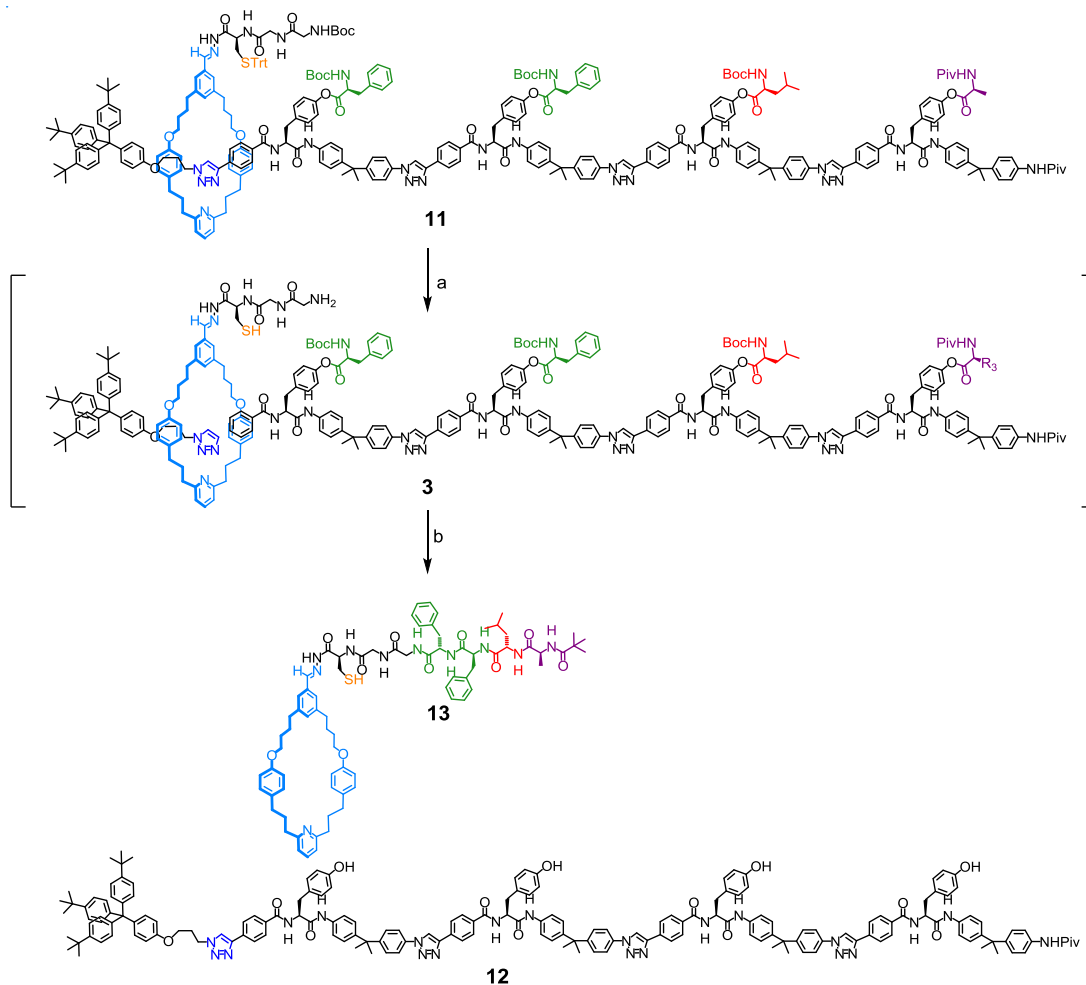
**Scheme 4.2.** Synthesis of the azide terminated single-barrier rotaxane **4**. Reaction conditions: a) **8** (10 equiv.), **6** (2 equiv.), **7** (1 equiv.), Cu(CH<sub>3</sub>CN)<sub>4</sub>.PF<sub>6</sub> (0.5 equiv.), CH<sub>2</sub>Cl<sub>2</sub>/*t*BuOH (3:1 v/v), RT, 24 h, 40%.

Subsequently rotaxane **4** was reacted with an alkyne-terminated molecular strand containing three amino acids in its structure (Scheme 4.3). The reaction proceeded smoothly allowing the isolation of aldehyde rotaxane **10** in 40% yield. Following the previously reported hydrazone exchange conditions, the reactive-arm based on the Cys-Gly-Gly unit was attached to the rotaxane giving access to **11** in good yield.



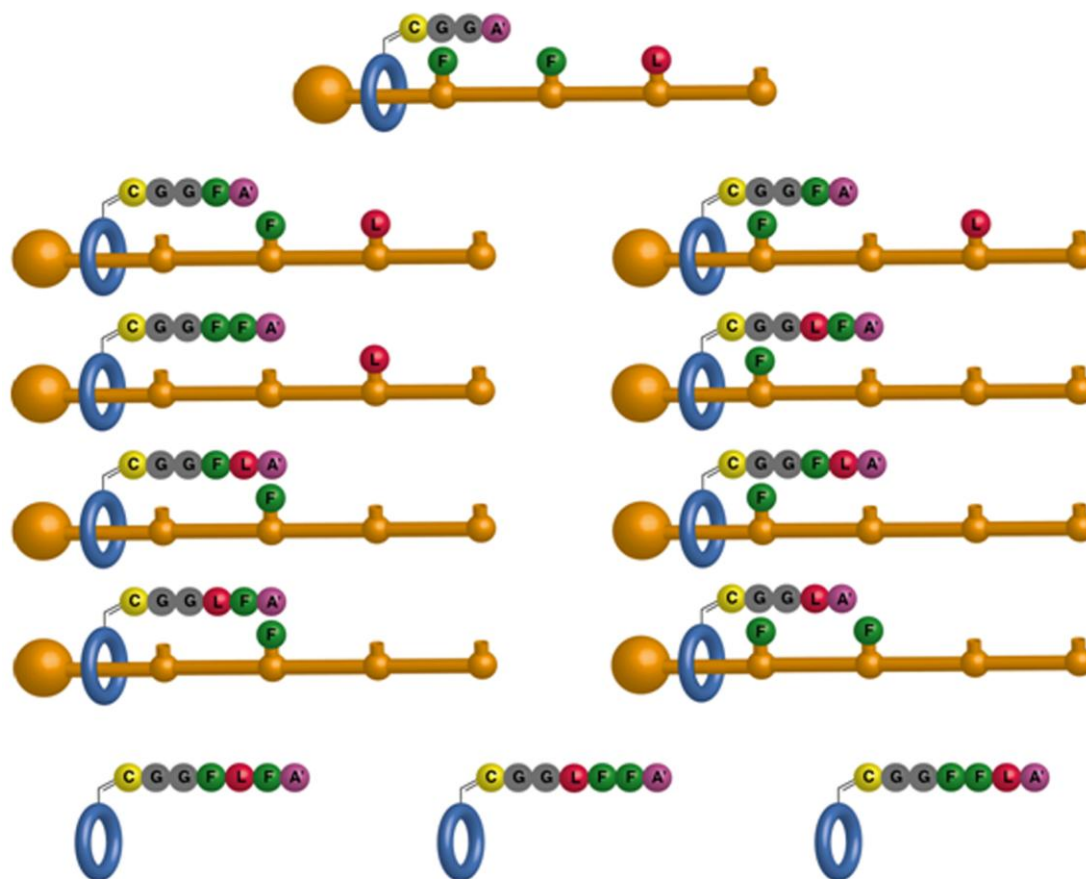
**Scheme 4.3.** Synthesis of 4-barrier molecular machine **11**. Reaction Conditions: a) Tentagel TBTA/Cu(CH<sub>3</sub>CN)<sub>4</sub>.PF<sub>6</sub>, CH<sub>2</sub>Cl<sub>2</sub>/tBuOH, RT, 48 h, 40%; b) aniline, *N*-Boc-Gly-Gly-Cys(STrt)NHN=CHC<sub>6</sub>H<sub>4</sub>OCH<sub>3</sub>, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 48 h, 90%.

## 4.3.2 Operation of Molecular Machine 11



**Scheme 4.4.** Operation of molecular machine **11**. Reaction conditions: a) CH<sub>2</sub>Cl<sub>2</sub>/TFA (5:1 v/v), RT, 2 h; b) DIPEA, TCEP·HCl, MeCN/DMF (3:1 v/v), 60 °C μw, 24 h.

Molecular machine **11** was operated under the previously established conditions (see Chapter 2 and 3, Scheme 4.4). As in this case the operation involved the formation of four peptide bonds the reaction time was extended by additional 12 h (studies on the original system suggested that this is the approximate time required to form one peptide bond). Upon analysis of the reaction mixture it was confirmed that the operation reached completion at this point and that two main products were present in the post-operation mixture. One product was identified as the uninterlocked thread with four ‘free’ hydroxyl groups **12** (no amino acids attached), the other as a macrocycle that by TLC analysis corresponded to the expected operation product **13**.



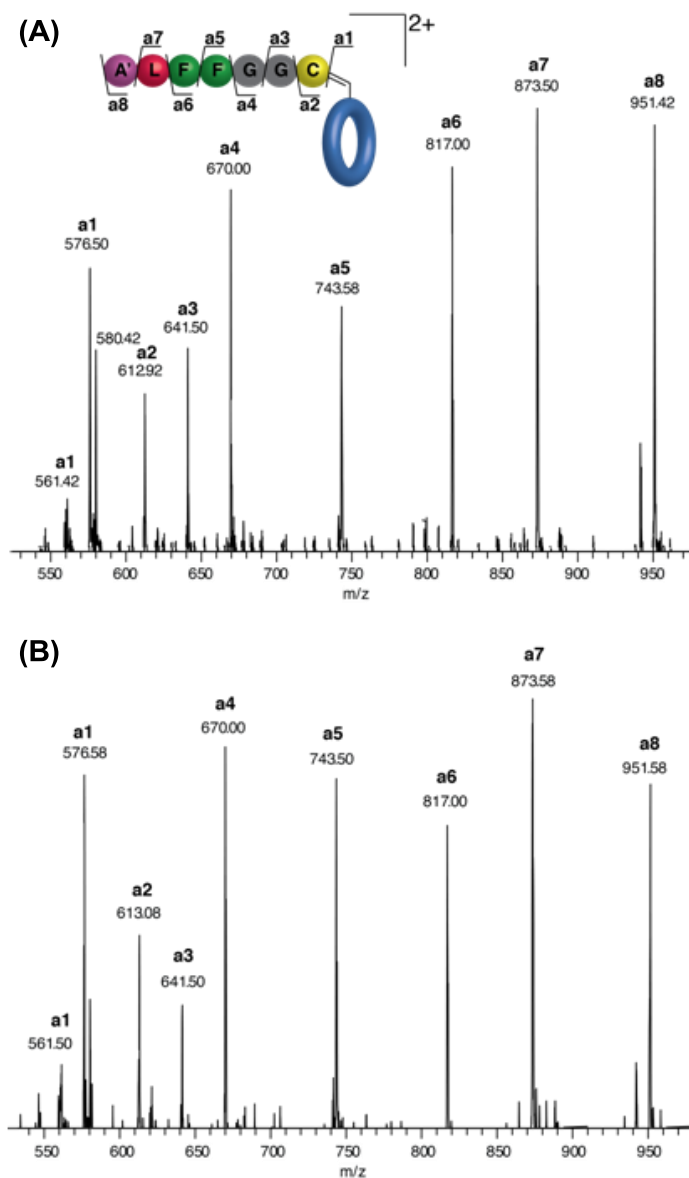
**Figure 4.4.** The 12 possible operation products of molecular machine 3.

Considering only intra-machine reactions, 12 products can be formed during the operation assuming that the machine does not proceed sequence-specifically (Figure 4.4).

In such a highly complex system side reactions can prove detrimental for the function of the molecular machine, since the yield of the final oligomeric operation product decreases exponentially with the increasing number of individual coupling steps. Even a relatively high yielding coupling should therefore lead to considerable byproduct formation after four coupling steps if the transformation is not quantitative and sequence-selective.

Since no rotaxane products have been found after preparative TLC and  $^1\text{H-NMR}$  analysis, the sequence of the isolated operation product is reduced to three possibilities (macrocycles shown in Figure 4.4). The identity of the operation product with the desired macrocycle carrying a Phe-Phe-Leu-Ala-Piv-peptide chain was confirmed by MS analysis which revealed the presence of expected ions in the spectrum ( $m/z$  1355  $[\text{M}+\text{H}]^+$ ,  $m/z$  678  $[\text{M}+2\text{H}]^+$ ,  $m/z$  951  $[\text{MD}+2\text{H}]^{2+}$ , D being the S,N-acetal derivative of the operation product – probably formed during the ionisation required for the ESI mass spectrometric analysis and observed in previous studies as well). Tandem mass spectrometry was then used to determine the

sequence of the peptide attached to the macrocycle (as reported for the original system in Chapter 3). Comparison of the fragmentation pattern of the isolated product with the above mentioned authentic sample of the macrocycle carrying a heptapeptide with identical amino-acid sequence confirmed that the operation product was obtained with complete sequence-specificity (Figure 4.4). This result indicates that the molecular machine operates with a high degree of sequence integrity and side reactions are practically not observed.



**Figure 4.5.** Tandem mass spectra of 2+ ions of *S,N*-acetal-derivatised macrocycles bearing the peptide sequences *N*-Piv-Ala-Leu-Phe-Phe-Gly-Gly-Cys(SH): (A) operation product (2+ ion isotope-selected  $m/z$  951.42), (B) reference compound prepared unambiguously by conventional peptide synthesis (2+ ion isotope-selected  $m/z$  951.58). Fragment a1 is observed as a singly-charged ion adduct of the aldehyde macrocycle ( $m/z$  562) or the corresponding methanol or hydrazone adduct ( $m/z$  576).

## 4.4 Conclusion

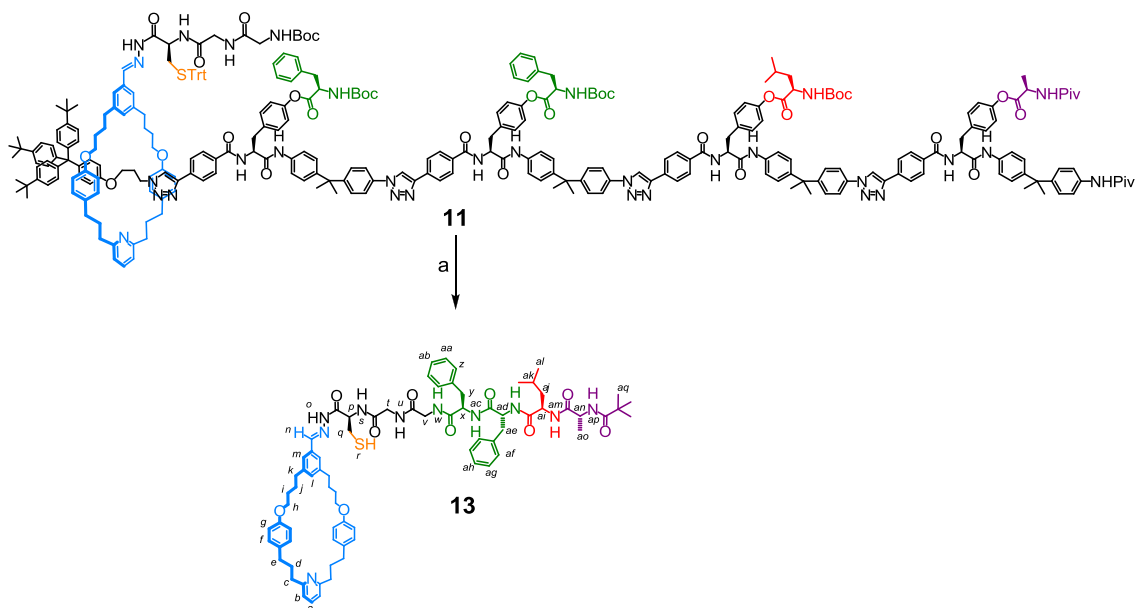
A novel approach towards complex molecular machines designed to perform synthesis of oligomers from monomeric building blocks was developed. The synthetic route involves preparation of a simple rotaxane, functionalised with a terminal azide moiety which can then in principle be extended by an alkyne-terminated thread of any given length and functionality. Having attached an alkyne that contains three amino acids derived building blocks successfully, a molecular machine for tetrapeptide synthesis was obtained. The operation of the machine using conditions established in our previous study yielded the desired product with complete sequence-specificity. These results open a new path for the development of complex systems intended to function as sophisticated molecular devices. The versatility of the synthetic approach and efficiency of machine's operation constitute another big step towards creating even more elaborate systems acting as molecular factories for preparation of useful, functional materials in the future.

## 4.5 Experimental Section

### 4.5.1 General Methods and Abbreviations

Unless stated otherwise, all reagents and solvents were purchased from Aldrich Chemicals and used without further purification. Compound **8**<sup>15</sup> was prepared according to the literature procedure. Compounds **E5** and **E6** were prepared by Dr. Bartosz Lewandowski. Compound **E2** was prepared by Dr. Bartosz Lewandowski. The syntheses of **1** and **7** are described in Chapter 2.<sup>11a</sup> The syntheses of **2**, **5**, and **9** are described in Chapter 3.<sup>11b</sup> All reactions requiring microwave heating were carried out in a Biotage Initiator 2.5.3. Dry DMF, CH<sub>2</sub>Cl<sub>2</sub>, CHCl<sub>3</sub> and THF was obtained by passing the solvent (HPLC grade) through an activated alumina column on a PureSolv™ solvent purification system (InnovativeTechnologies Inc., MA). Flash column chromatography was carried out using Geduran® Si 60 (particle size 40-63 μm, Merck, Germany) as the stationary phase, and analytical TLC was performed on precoated silica gel plates (0.25 mm thickness, 60 F254, Merck, Germany) and observed under UV light or stained with a phosphomolybdic acid solution. Preparative TLC was performed on precoated silica gel plates (2 mm, UNIPLATE GF, Analtech Inc., DE or 2 mm, Tapered plate, UNIPLATE-T, Analtech Inc., DE). NMR spectra were recorded on Bruker AV 400 MHz, Bruker AV 500 MHz (equipped with a cryoprobe) or Bruker Avance III (equipped with a cryoprobe) with an Oxford AS 600 MHz magnet. Chemical shifts are reported in parts per million (ppm) from high to low frequency and referenced to the residual solvent resonance. Coupling constants (*J*) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintuplet, m = multiplet, br = broad. <sup>1</sup>H-NMR assignments were made using 2D NMR methods (COSY, HSQC, HMBC). Melting points (m.p.) were determined using a Sanyo Gallenkamp apparatus and are reported uncorrected. Low resolution ESI mass spectrometry was performed with a Finnigan LCQ-MS mass spectrometer. High resolution ESI (electrospray ionisation) and APCI (atmospheric-pressure chemical ionisation) mass spectrometry was carried out by the mass spectrometry services at the EPSRC National Mass Spectrometry Service Centre, Swansea, UK or the Mass Spectrometry Service of the University of Manchester. Optical rotations were recorded using a Bellingham and Stanley ADP 220 Polarimeter using a 1 dm cell.

## 4.5.1.1 Operation of Molecular Machine 11

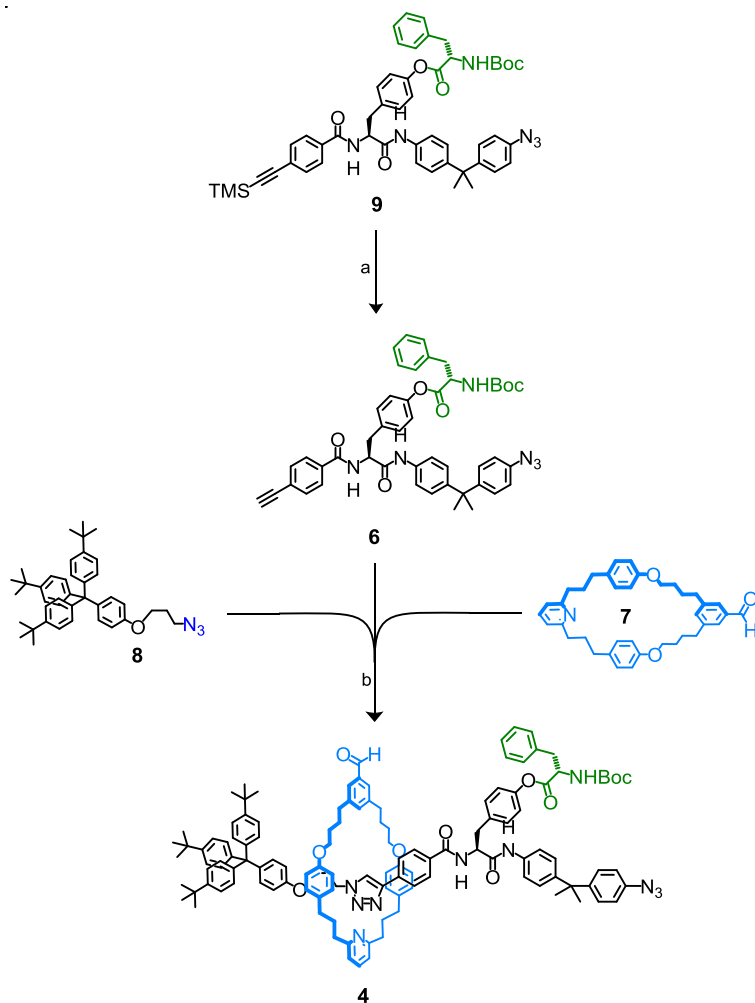


**Scheme 4.5.** Operation of molecular machine. Reaction conditions: a) **1**. CH<sub>2</sub>Cl<sub>2</sub>/TFA (5:1 v/v), RT, 2 h, then DIPEA, TCEP·HCl, MeCN/DMF (3:1 v/v),  $\mu$ w, 60 °C, 48 h.

Molecular machine **11** (17 mg, 3.5  $\mu$ mol, 1.0 equiv.) was stirred at room temperature in a mixture of CH<sub>2</sub>Cl<sub>2</sub>/TFA (5:1 v/v, 2 ml) for 2 h. Toluene (5 ml) was added and the solvent removed under reduced pressure. The residue was then dissolved in degassed MeCN/DMF (3:1 v/v, 4 ml) and both *N,N*-diisopropylethylamine (43  $\mu$ l, 250  $\mu$ mol 71 equiv.) and *tris*(2-carboxyethyl)phosphine hydrochloride (2.0 mg, 7.1  $\mu$ mol, 2.0 equiv.) were added and the reaction mixture was stirred under microwave heating at 60°C for 48 h. Preparative TLC (SiO<sub>2</sub> [UNIPLATE-T], CH<sub>2</sub>Cl<sub>2</sub>/MeOH 5%) of the residue afforded **13** (2.5 mg, 53%) as a colorless film.

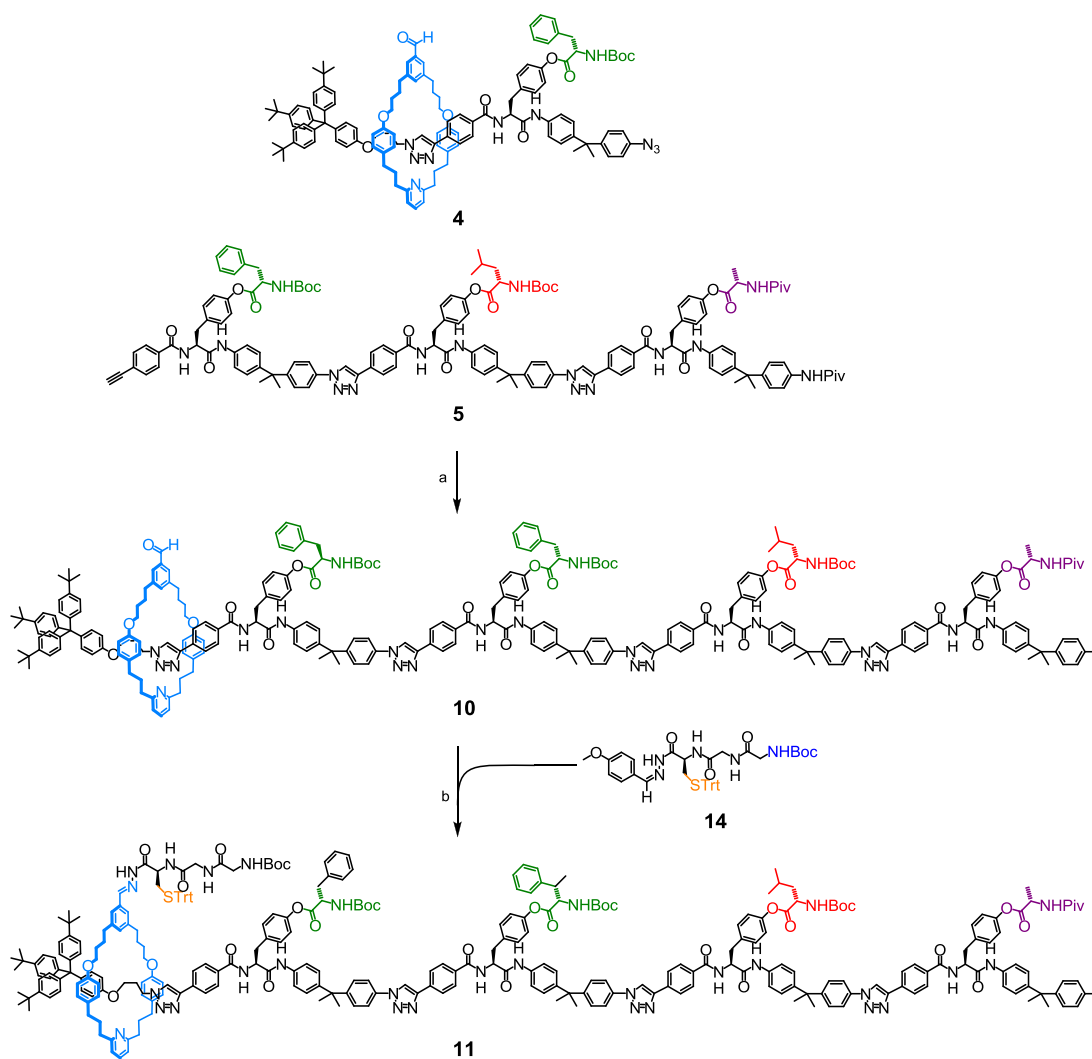
2:3 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone. <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  = 11.54 (s, *J* = 7.7 Hz, 2H, H<sub>o1+o2</sub>), 8.47 – 8.00 (range of signals from different isomers, 10H, H<sub>s1+s2+u1+u2+w1+w2+ac1+ac2+ad1+ad2</sub>), 8.20 (s, 1H, H<sub>o1</sub>), 7.97 (s, 1H, H<sub>o2</sub>), 7.93 – 7.60 (m, 4H, H<sub>am1+am2+ap1+ap2</sub>), 7.32 – 7.11 (m, 26H, H<sub>a1+a2+l1+l2+m1+m2+z1+z2+aa1+aa2+ab1+ab2+af1+af2+ag1+ag2+ah1+ah2</sub>), 7.06 (d, *J* = 7.3 Hz, 2H, H<sub>b1+b2</sub>), 7.03 (dd, *J* = 8.4 Hz, 2.2 Hz, 8H, H<sub>fl+f2</sub>), 6.75 (dq, *J* = 8.1, 2.5 Hz, 8H, H<sub>g1+g2</sub>), 4.65 – 4.39 (m, 6H, H<sub>p1+p2+x1+x2+ad1+ad2</sub>), 4.24 (m, 4H, H<sub>ai1+ai2+an1+an2</sub>), 3.89 (t, *J* = 5.8 Hz, 8H, H<sub>h1+h2</sub>), 3.85 – 3.73 (m, 8H, H<sub>tl+t2+v1+v2</sub>), 3.12 - 2.98 (m, 8H, H<sub>y1+y2+ae1+ae2</sub>), 2.83 - 2.73 (m, 8H, H<sub>c1+c2</sub>), 2.69 – 2.61 (m, 8H, H<sub>k1+k2</sub>), 2.45 – 2.39 (m, 4H, H<sub>q1+q2</sub>), 1.94 (p, *J* = 7.6 Hz, 8H, H<sub>d1+d2</sub>), 1.75 – 1.63 (m, 16H, H<sub>il+i2+j1+j2</sub>), 1.51 – 1.45 (m, 2H, H<sub>ak1+ak2</sub>), 1.35 – 1.29 (m, 4H, H<sub>aj1+aj2</sub>), 1.17 (d, *J* = 7.2 Hz, 3H, H<sub>ao1</sub>), 1.14 (d, *J* = 7.1 Hz, 3H, H<sub>ao2</sub>), 1.11 – 1.05 (m, 18H, H<sub>aq1+aq2</sub>), 0.82 – 0.70 (m, 12H, H<sub>al1+al2</sub>).

## 4.5.2 Synthetic Schemes

4.5.2.1 Synthesis of Single-Barrier Rotaxane **4**

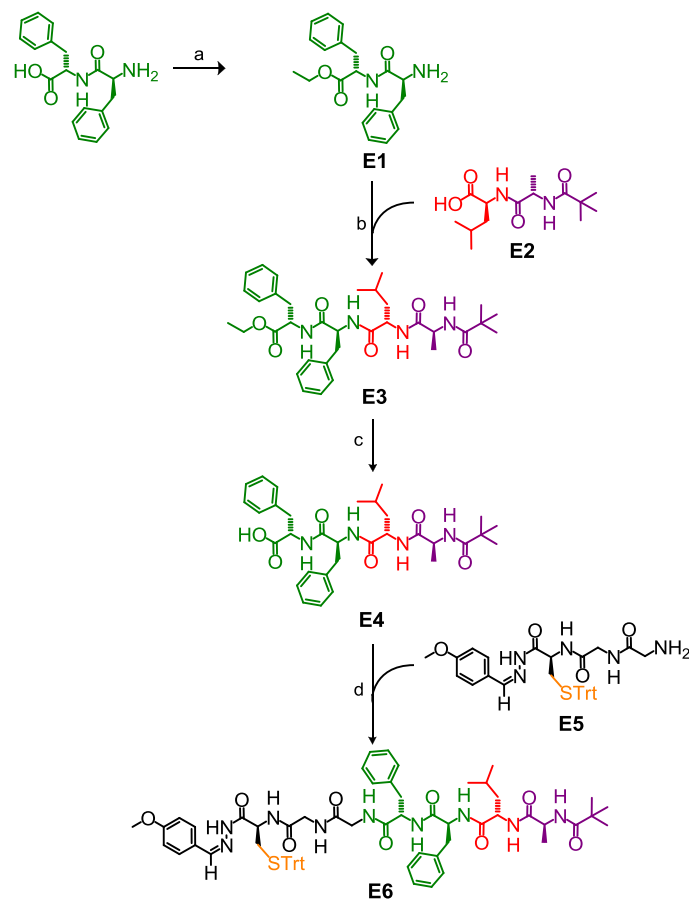
**Scheme 4.6.** Reaction conditions: a)  $\text{AgNO}_3$ , acetone/ $\text{H}_2\text{O}$  (4:1 v/v), RT, 77%; b) **8** (10 equiv.), **7** (2 equiv.), **6** (1 equiv.),  $\text{Cu}(\text{CH}_3\text{CN})_4\text{PF}_6$ ,  $\text{CH}_2\text{Cl}_2$ /*t*BuOH (3:1 v/v), RT, 24 h, 37%.

## 4.5.2.2 Synthesis of Rotaxane 11



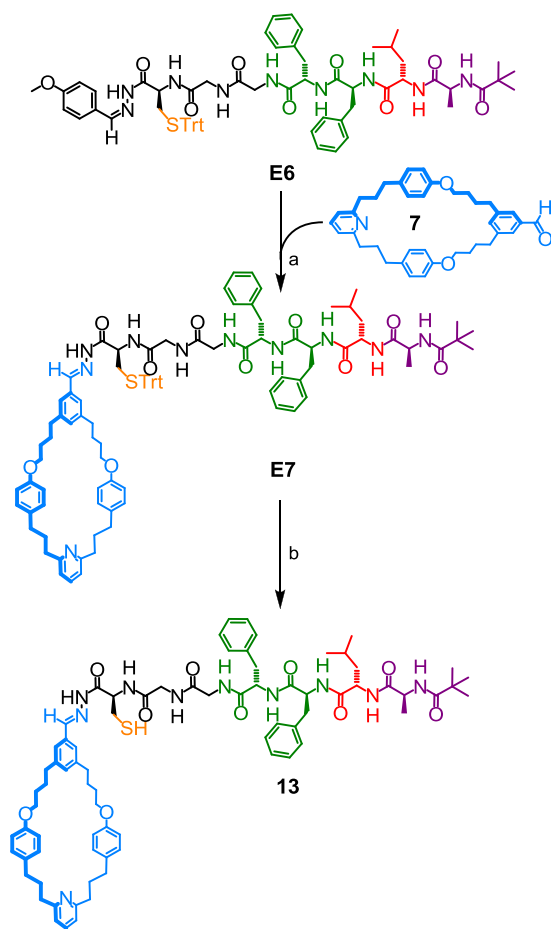
**Scheme 4.7.** Reaction conditions: a) Tentagel-TBTA/Cu(MeCN)<sub>4</sub>PF<sub>6</sub>, CH<sub>2</sub>Cl<sub>2</sub>/*t*BuOH (4:1 v/v), RT, 4 d, 40%; b) aniline, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 48 h, 90%.

## 4.5.2.3 Synthesis of E6



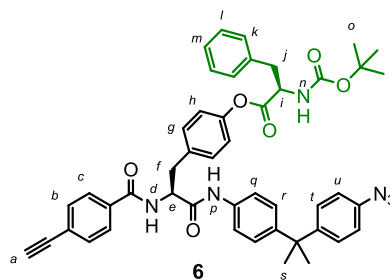
**Scheme 4.8.** Reaction conditions: a)  $\text{H}_2\text{SO}_4$ , EtOH, reflux, 24 h, 94%; b) PyBroP, DIPEA, DMF, RT, 24 h, 90%; c) NaOH, MeOH, RT, 2 h, 94%; d) PyBroP, DIPEA, DMF, RT, 48 h, 60%.

## 4.5.2.4 Synthesis of 13



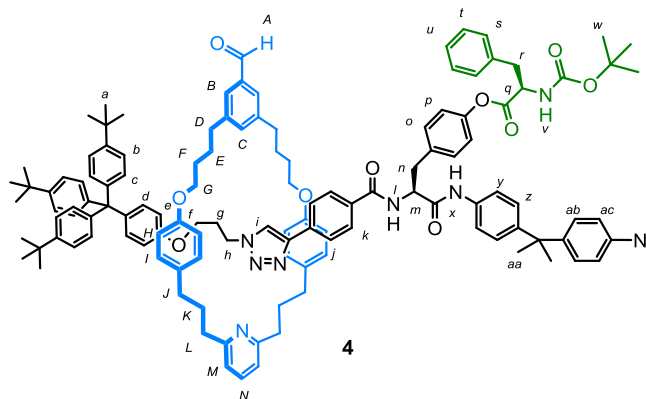
**Scheme 4.9.** Reaction conditions: a) aniline, DMSO/MES buffer (3:1 v/v, pH 6.0), 60 °C, 72 h, 56%;  
b) CH<sub>2</sub>Cl<sub>2</sub>/TFA (4:1 v/v), RT, 90 min.

## 4.5.3 Synthetic Procedures and Characterisation Data

Ethynyl-*N*-Boc-Phenylalanine-Azide Barrier **6**

*N*-Boc-protected phenylalanine-loaded barrier **9** (480 mg, 0.56 mmol, 1.00 equiv.) was dissolved in acetone (120 ml) and a solution of AgNO<sub>3</sub> (95.0 mg, 0.56 mmol, 1.00 equiv.) in H<sub>2</sub>O (4 ml) was added. The reaction mixture was stirred in the dark at room temperature for 18 h. The reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (200 ml). The organic layer was washed with aqueous 1M NaCl (100 ml) and H<sub>2</sub>O (2 × 200ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. Flash column chromatography (SiO<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>/MeOH 1.5%) of the residue afforded **6** (340 mg, 77%) as a colourless film.

[ $\alpha$ ]<sub>D</sub> = -0.01° (c 2.20, CH<sub>2</sub>Cl<sub>2</sub>); <sup>1</sup>H NMR (600 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 9.43 (s, 1H, H<sub>p</sub>), 8.06 (d, *J* = 8.0 Hz, 1H, H<sub>d</sub>), 7.87 (d, *J* = 8.0 Hz, 2H, H<sub>c</sub>), 7.55 (d, *J* = 8.1 Hz, 2H, H<sub>b</sub>), 7.54 (d, *J* = 8.2 Hz, 2H, H<sub>q</sub>), 7.38 (d, *J* = 8.1 Hz, 2H, H<sub>g</sub>), 7.37 – 7.30 (m, 4H, H<sub>k + l</sub>), 7.28 (d, *J* = 8.3 Hz, 2H, H<sub>i</sub>), 7.24 (t, *J* = 7.2 Hz, 1H, H<sub>m</sub>), 7.17 (d, *J* = 8.3 Hz, 2H, H<sub>r</sub>), 6.99 (d, *J* = 8.3 Hz, 2H, H<sub>u</sub>), 6.96 (d, *J* = 8.1 Hz, 2H, H<sub>h</sub>), 6.43 (d, *J* = 7.9 Hz, 1H, H<sub>n</sub>), 4.94 (q, *J* = 7.7 Hz, 1H, H<sub>e</sub>), 4.57 (q, *J* = 7.6 Hz, 1H, H<sub>j</sub>), 3.83 (s, 1H, H<sub>a</sub>), 3.35 (dd, *J* = 14.0 Hz, 5.7 Hz, 1H, H<sub>f</sub>), 3.26 (dd, *J* = 13.9 Hz, 6.0 Hz, 1H, H<sub>j</sub>), 3.18 (dd, *J* = 14.7, 8.7 Hz, 1H, H<sub>f</sub>), 3.15 (dd, *J* = 14.7 Hz, 8.9 Hz, 1H, H<sub>j</sub>), 1.65 (s, 6H, H<sub>s</sub>), 1.37 (s, 9H, H<sub>o</sub>); <sup>13</sup>C NMR (150 MHz, acetone-*d*<sub>6</sub>)  $\delta$  = 171.6, 170.3, 166.8, 156.3, 150.5, 148.7, 146.5, 138.1, 138.1, 137.5, 137.4, 136.3, 135.3, 132.7, 131.1, 130.2, 129.3, 129.2, 128.5, 127.8, 127.6, 126.2, 122.2, 120.2, 119.4, 83.5, 81.4, 79.6, 56.8, 56.8, 56.5, 38.1, 37.8, 31.0, 28.5; LRESI(-)-MS: *m/z* 825.36 [M+Cl]<sup>-</sup>; HRESI(+)-MS: 813.3372 [M+Na]<sup>+</sup>, 813.3377 calc. for C<sub>47</sub>H<sub>46</sub>O<sub>6</sub>N<sub>6</sub>Na.

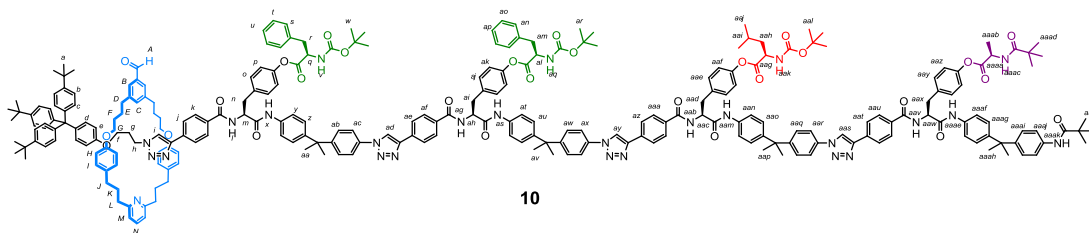
Aldehyde *N*-Boc-Phenylalanine Single-Barrier Rotaxane **4**

Macrocycle **7** (135 mg, 0.24 mmol, 2.00 equiv.) and  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (22 mg, 60  $\mu\text{mol}$ , 0.50 equiv.) were allowed to stir in degassed  $\text{CH}_2\text{Cl}_2$  (2 ml) at room temperature for 30 minutes. Then a solution of alkyne **6** (95 mg, 120  $\mu\text{mol}$ , 1.0 equiv.) and azide **8** (706 mg, 1.2 mmol, 10 equiv.) in degassed  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v, 5 ml) was added. The reaction mixture was stirred at room temperature for 40 hours. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (50 ml) and the organic layer was washed with aqueous 0.01M  $\text{Na}_4\text{EDTA}$  ( $2 \times 20$  ml). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ ,  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) of the residue allowed the recovery of the excess of the azide **1** and macrocycle **2**. A combined fraction of rotaxane **4** and the corresponding uninterlocked thread was purified on preparative TLC ( $\text{SiO}_2$  [Merck],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  2%) affording **4** (87 mg, 37%) as a colourless film.

$[\alpha]_{\text{D}} = -1.18^\circ$  (c 0.06, acetone);  $^1\text{H}$  NMR (600 MHz, acetone- $d_6$ )  $\delta = 9.97$  (s, 1H,  $\text{H}_A$ ), 9.48 (s, 1H,  $\text{H}_x$ ), 8.04 (s, 1H,  $\text{H}_i$ ), 8.01 (d,  $J = 8.2$  Hz, 1H,  $\text{H}_j$ ), 7.87 (d,  $J = 8.1$  Hz, 2H,  $\text{H}_j$ ), 7.83 (d,  $J = 8.1$  Hz, 2H,  $\text{H}_k$ ), 7.63 – 7.53 (m, 6H,  $\text{H}_{ac+B+N}$ ), 7.42 (d,  $J = 8.0$  Hz, 2H,  $\text{H}_o$ ), 7.37 – 7.29 (m, 13H,  $\text{H}_{b+s+t+y+C}$ ), 7.25 (t,  $J = 7.2$  Hz, 1H,  $\text{H}_u$ ), 7.20 (d,  $J = 8.2$  Hz, 2H,  $\text{H}_{ab}$ ), 7.13 (d,  $J = 8.3$  Hz, 6H,  $\text{H}_c$ ), 7.05 (d,  $J = 7.7$  Hz, 2H,  $\text{H}_M$ ), 7.01 (d,  $J = 8.4$  Hz, 2H,  $\text{H}_z$ ), 6.98 (d,  $J = 8.1$  Hz, 2H,  $\text{H}_p$ ), 6.95 (d,  $J = 8.4$  Hz, 2H,  $\text{H}_d$ ), 6.78 (d,  $J = 7.9$  Hz, 4H,  $\text{H}_l$ ), 6.48 (d,  $J = 8.6$  Hz, 2H,  $\text{H}_e$ ), 6.47 – 6.43 (m, 5H,  $\text{H}_{H+v}$ ), 5.02 (q,  $J = 7.2$  Hz, 1H,  $\text{H}_m$ ), 4.59 (q,  $J = 7.2$  Hz, 1H,  $\text{H}_q$ ), 4.08 (t,  $J = 7.1$  Hz, 2H,  $\text{H}_h$ ), 3.85 – 3.74 (m, 4H,  $\text{H}_G$ ), 3.49 (t,  $J = 6.0$  Hz, 2H,  $\text{H}_p$ ), 3.36 (dd,  $J = 14.1$  Hz, 6.2 Hz, 1H,  $\text{H}_n$ ), 3.27 (dd,  $J = 13.8$  Hz, 6.0 Hz, 1H,  $\text{H}_r$ ), 3.23 – 3.13 (m, 2H,  $\text{H}_{n+r}$ ), 2.79 – 2.69 (m, 4H,  $\text{H}_D$ ), 2.66 – 2.59 (m, 4H,  $\text{H}_L$ ), 2.55 – 2.51 (m, 4H,  $\text{H}_j$ ), 1.93 – 1.84 (m, 6H,  $\text{H}_{g+K}$ ), 1.84 – 1.78 (m, 4H,  $\text{H}_E$ ), 1.77 – 1.72 (m, 4H,  $\text{H}_F$ ), 1.67 (s, 6H,  $\text{H}_{aa}$ ), 1.38 (s, 9H,  $\text{H}_w$ ), 1.32 (s, 27H,  $\text{H}_a$ );  $^{13}\text{C}$  NMR (150 MHz, acetone- $d_6$ )  $\delta = 193.2$ , 171.6, 170.4, 167.2, 162.5, 158.0, 157.2, 156.3, 150.5, 149.1, 148.7, 146.9, 146.5, 145.3, 144.7, 140.1, 138.1, 138.1, 137.9, 137.6, 137.6, 136.4, 135.6, 135.2, 134.2, 133.9, 132.6,

131.4, 131.2, 130.2, 130.0, 129.3, 129.2, 128.7, 127.9, 127.8, 127.6, 125.9, 125.1, 122.2, 122.2, 120.6, 120.2, 119.4, 115.0, 114.1, 79.5, 67.9, 64.8, 63.9, 56.8, 56.5, 47.5, 42.9, 38.4, 38.1, 37.8, 35.7, 35.6, 34.9, 32.9, 31.7, 31.0, 30.4, 29.6, 29.1, 28.5; LRESI(+)-MS:  $m/z$  1939.75 [M+H]<sup>+</sup>.

**Aldehyde**      ***N*-Boc-Penylalanine-*N*-Boc-Phenylalanine-*N*-Boc-Leucine-*N*-Piv-Alanine**  
**Four-Barrier Rotaxane 10**

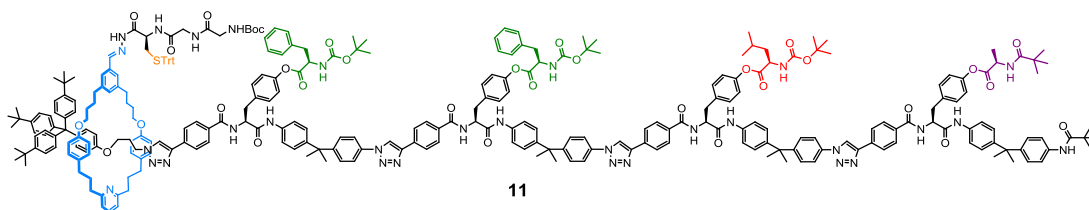


Tentagel TBTA (500 mg) resin was loaded with  $\text{Cu}(\text{MeCN})_4\text{PF}_6$  (0.12 mmol/g, 0.28 equiv.) in  $\text{CH}_2\text{Cl}_2$  and washed with  $\text{Et}_2\text{O}$ . After drying, a degassed solution of rotaxane **4** (54 mg, 28  $\mu\text{mol}$ , 2.0 equiv.) and alkyne **5** (32 mg, 14  $\mu\text{mol}$ , 1.0 equiv.) in  $\text{CH}_2\text{Cl}_2/t\text{BuOH}$  (4:1 v/v, 5 ml) was added. The resulting suspension was shaken at room temperature for 4 d, filtered, and the resin was washed with a mixture of  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  (2:1 v/v, 6 ml). The solvent was removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [Merck],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  3%, 3 elutions) of the residue afforded **10** (24 mg, 40%) as a colourless film.

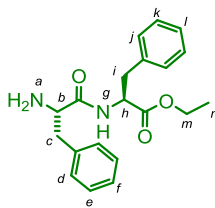
$^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 10.30 – 10.14 (m, 4H,  $\text{H}_{x+as+aam+aaae}$ ), 9.87 (s, 1H,  $\text{H}_A$ ), 9.35 (bs, 3H,  $\text{H}_{ad+ay+aay}$ ), 9.13 (bs, 1H,  $\text{H}_{aaak}$ ), 8.87 – 8.81 (m, 2H,  $\text{H}_{l+aab}$ ), 8.77 – 8.71 (m, 2H,  $\text{H}_{aav+ag}$ ), 8.16 (s, 1H,  $\text{H}_i$ ), 8.01 (d,  $J$  = 8.4 Hz, 6H,  $\text{H}_{af+aaa+aau}$ ), 7.96 (d,  $J$  = 8.1 Hz, 6H,  $\text{H}_{ae+az+aat}$ ), 7.93 (bs, 1H,  $\text{H}_{aaac}$ ), 7.84 (d,  $J$  = 8.3 Hz, 6H,  $\text{H}_{ac+ax+aar}$ ), 7.82 (d,  $J$  = 8.4 Hz, 2H,  $\text{H}_{af}$ ), 7.74 (d,  $J$  = 8.0 Hz, 2H,  $\text{H}_j$ ), 7.58 – 7.48 (m, 16H,  $\text{H}_{y+aj+at+aae+aan+aay+aaaf+B}$ ), 7.48 – 7.41 (m, 14H,  $\text{H}_{aak+aq+ab+aw+aaq+o+N}$ ), 7.33 (s, 1H,  $\text{H}_C$ ), 7.31 – 7.25 (m, 8H,  $\text{H}_{s+an+t+ao}$ ), 7.24 – 7.18 (m, 14H,  $\text{H}_{b+aaag+ao+z+au}$ ), 7.17 – 7.13 (m, 2H,  $\text{H}_{ar}$ ), 7.11 (d,  $J$  = 8.3 Hz, 2H,  $\text{CH}_{ar}$ ), 7.08 – 7.05 (m, 2H,  $\text{H}_{u+ap}$ ), 7.03 – 6.95 (m, 8H,  $\text{H}_c$ ,  $\text{H}_{ar}$ ), 6.90 (t,  $J$  = 8.1 Hz, 2H,  $\text{H}_M$ ), 6.73 (d,  $J$  = 8.7 Hz, 2H,  $\text{H}_d$ ), 6.70 (d,  $J$  = 7.5 Hz, 2H,  $\text{H}_l$ ), 6.65 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_{ar}$ ), 6.35 (dd,  $J$  = 8.5 Hz, 3.5 Hz, 4H,  $\text{H}_H$ ), 6.23 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_e$ ), 4.91 – 4.82 (m, 3H,  $\text{H}_{aaw+aac+ah}$ ), 4.82 – 4.73 (m, 1H,  $\text{H}_m$ ), 4.42 – 4.28 (m, 2H,  $\text{H}_{aaa+q}$ ), 4.22 – 4.04 (m, 2H,  $\text{H}_{aag+al}$ ), 3.98 (t,  $J$  = 7.1 Hz, 2H,  $\text{H}_f$ ), 3.72 – 3.56 (m, 4H,  $\text{H}_G$ ), 3.51 (bs, 2H,  $\text{H}_h$ ), 3.21 – 3.06 (m, 6H,  $\text{H}_{aax+aad+ai+n+r+am}$ ), 3.05 – 2.95 (m, 6H,  $\text{H}_{aax'+aad'+ai'+n'+r'+am'}$ ), 2.60 – 2.40 (m, 12H,  $\text{H}_{D+J+L}$ ), 1.80 – 1.70 (m, 7H,  $\text{H}_{aai+g+K}$ ), 1.69 (bs, 18H,  $\text{H}_{aap+av+aa}$ ), 1.63 – 1.54 (m, 14H,  $\text{H}_{aaah+E+F}$ ), 1.40 (d,  $J$  = 7.3 Hz, 3H,  $\text{H}_{aaad}$ ), 1.37 – 1.32 (m, 27,  $\text{H}_{w+ar+aal}$ ), 1.23 (bs, 27H,  $\text{H}_a$ ), 1.20 (s, 9H,  $\text{H}_{aaad}$ ), 1.09 (s, 9H,  $\text{H}_{aaal}$ ), 0.92 – 0.84 (m, 6H,  $\text{H}_{aaj}$ );  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 193.1, 177.8, 176.3, 174.4, 172.0, 171.7, 171.5, 171.0, 170.4, 170.3, 166.1, 165.9, 161.0, 156.4, 155.8, 155.7, 155.6, 155.5, 151.3, 149.1, 149.0, 148.9, 147.7, 147.6, 146.4, 145.6, 145.5, 145.5, 145.1, 144.7, 144.0, 143.3, 138.6, 137.3, 136.9, 136.7, 136.4, 135.9, 134.3, 134.3, 133.5, 133.4, 133.3, 133.1, 133.0, 133.0, 131.6, 131.4, 131.0, 130.3, 130.2, 130.2, 130.0, 129.3, 129.1, 128.9, 128.3, 128.1, 128.0, 126.9, 126.8, 126.6, 126.4, 126.3, 125.0, 124.6, 124.4, 124.2, 121.1, 121.1, 120.5, 119.9, 119.8, 119.7, 119.2, 119.1, 119.0, 114.9,

113.9, 113.0, 78.5, 78.4, 78.0, 69.8, 67.4, 66.8, 62.5, 56.6, 56.2, 55.8, 55.5, 55.2, 52.2, 48.2, 47.6, 45.9, 45.9, 42.2, 41.7, 39.5, 38.1, 37.8, 37.1, 36.6, 36.3, 34.4, 34.3, 34.1, 34.0, 31.6, 31.1, 30.5, 30.3, 28.2, 28.2, 28.1, 27.6, 27.3, 27.2, 27.2, 26.0, 25.9, 24.4, 22.8, 21.3, 16.4;  
LRESI(+)-MS:  $m/z$  2132.8  $[M+H+Na]^{2+}$ .

***N*-Boc-Gly-Gly-Cys(Trt)-Acyl Hydrazone Aldehyde *N*-Boc-Penylalanine-*N*-Boc-Phenylalanine-*N*-Boc-Leucine-*N*-Piv-Alanine Four-Barrier Rotaxane **11****

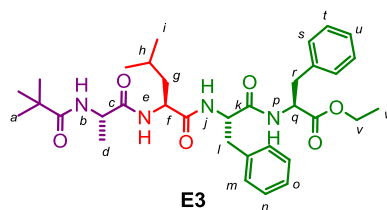


A mixture of **10** (17 mg, 4.0  $\mu\text{mol}$ ), **14** (4.0 mg, 6.0  $\mu\text{mol}$ ) and aniline (50  $\mu\text{l}$ ) in DMSO/2-(*N*-morpholino)ethanesulfonic acid buffer (3:1 *v/v*, 1 ml, pH 6.0) was stirred at 60  $^{\circ}\text{C}$  for 72 h. The reaction mixture was diluted with  $\text{CH}_2\text{Cl}_2$  (10 ml) and the organic layer was washed with  $\text{H}_2\text{O}$  (10 ml). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent removed under reduced pressure. Preparative TLC ( $\text{SiO}_2$  [UNIPLATE-T],  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  5%) of the residue afforded **11** (17 mg, 90%) as a colourless solid. The title compound was operated without further purification.

**NH<sub>2</sub>-Phe-Phe-OEt E1****E1**

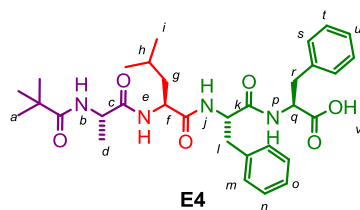
*N*-Boc-Phe-Phe-OEt (0.31 g, 1.00 mmol) was dissolved in EtOH (10 ml). Aqueous H<sub>2</sub>SO<sub>4</sub> (5%, 0.50 ml) was added and the reaction mixture was heated to reflux for 24 h. The reaction mixture was allowed to cool to room temperature and diluted with EtOAc (25 ml). The organic layer was washed with aqueous 1M NaHCO<sub>3</sub> (10 ml) and the aqueous layer was extracted with EtOAc (3 × 25 ml). The combined organic layers were washed with H<sub>2</sub>O (25 ml), brine (25 ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure to afford **E1** (0.32 g, 94% yield) as amorphous white solid.

m.p. 296 ° C; [α]<sub>D</sub> = -0.03° (c 0.43, MeOH); <sup>1</sup>H NMR (600 MHz, CDCl<sub>3</sub>) δ = 7.78 (d, *J* = 8.4 Hz, 1H, H<sub>g</sub>), 7.34 (t, *J* = 7.4 Hz, 2H, H<sub>e</sub>), 7.30 – 7.24 (m, 4H, H<sub>f + k + l</sub>), 7.22 (d, *J* = 7.7 Hz, 2H, H<sub>j</sub>), 7.08 (d, *J* = 7.3 Hz, 2H, H<sub>d</sub>), 4.89 (q, *J* = 7.0 Hz, 1H, H<sub>h</sub>), 4.24 – 4.13 (m, 2H, H<sub>m</sub>), 3.62 (dd, *J* = 9.5 Hz, 3.9 Hz, 1H, H<sub>b</sub>), 3.18 (dd, *J* = 13.8 Hz, 4.0 Hz, 1H, H<sub>c</sub>), 3.12 (dd, *J* = 8.6 Hz, 6.3 Hz, 2H, H<sub>i</sub>), 2.63 (dd, *J* = 13.7 Hz, 9.2 Hz, 1H, H<sub>c'</sub>), 1.29 – 1.21 (m, 3H, H<sub>n</sub>); <sup>13</sup>C NMR (150 MHz, CDCl<sub>3</sub>) δ = 173.8, 171.6, 137.7, 136.1, 129.4, 129.4, 128.7, 128.5, 127.0, 126.9, 126.9, 126.9, 61.4, 56.3, 52.8, 40.8, 38.2, 14.2; HRESI(+)-MS: *m/z* 363.1682 [M+Na]<sup>+</sup>, 363.1679 calc. for C<sub>20</sub>H<sub>24</sub>O<sub>3</sub>N<sub>2</sub>Na.

**Piv(NH)-Ala-Leu-Phe-Phe-OEt E3**

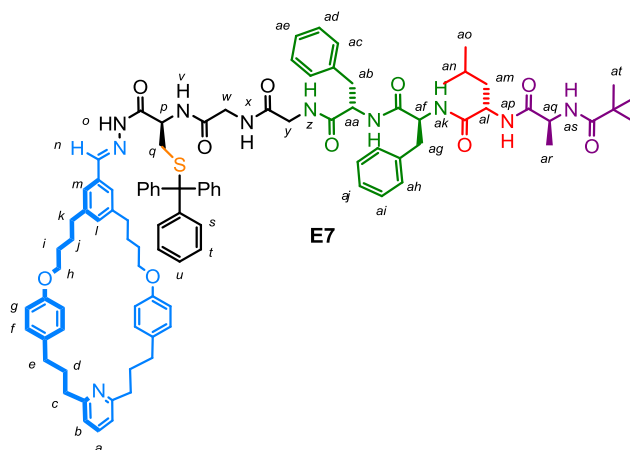
**E1** (0.17 g, 0.50 mmol, 1.00 equiv.) and **E2** (0.15 g, 0.52 mmol, 1.04 equiv.) were dissolved in DMF (5 ml). PyBrOP (0.35 mg, 0.75 mmol, 1.50 equiv.) and *N,N*-diisopropylethylamine (0.20 ml, 1.15 mmol, 2.30 equiv.) were added and the mixture was stirred at room temperature for 24 h. The reaction mixture was diluted with EtOAc (20 ml) and the organic layer was washed with H<sub>2</sub>O (20 ml). The aqueous layer was extracted with EtOAc (2 × 20 ml). The combined organic layers were washed with H<sub>2</sub>O (40 ml), brine, dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. The residue was purified by trituration with Et<sub>2</sub>O (5 ml) to give **E3** (0.27 g, 90% yield) as a white powder.

m.p. 232 °C;  $[\alpha]_D = -0.16^\circ$  (c 0.16, MeOH); <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>)  $\delta = 8.45$  (d,  $J = 8.1$  Hz, 1H, H<sub>j</sub>), 7.89 (d,  $J = 8.1$  Hz, 1H, H<sub>p</sub>), 7.62 (d,  $J = 8.9$  Hz, 1H, H<sub>b</sub>), 7.43 (d,  $J = 8.2$  Hz, 1H, H<sub>e</sub>), 7.30 – 7.13 (m, 10H, H<sub>m+n+o+s+t+u</sub>), 4.58 – 4.50 (m, 1H, H<sub>q</sub>), 4.43 (q,  $J = 6.9$  Hz, 1H, H<sub>k</sub>), 4.27 – 4.16 (m, 2H, H<sub>c+f</sub>), 4.01 (q,  $J = 6.1$  Hz, 1H, H<sub>v</sub>), 3.04 – 2.91 (m, 3H, H<sub>l+r</sub>), 2.76 (dd,  $J = 14.5$  Hz, 8.5 Hz, 1H, H<sub>r</sub>), 1.52 – 1.42 (m, 1H, H<sub>h</sub>), 1.30 (t,  $J = 6.9$  Hz, 2H, H<sub>g</sub>), 1.16 (d,  $J = 6.7$  Hz, 3H, H<sub>d</sub>), 1.11 – 1.04 (m, 12H, H<sub>a+w</sub>), 0.82 (d,  $J = 6.4$  Hz, 3H, H<sub>i</sub>), 0.78 (d,  $J = 6.4$  Hz, 3H, H<sub>i</sub>); <sup>13</sup>C NMR (150 MHz, DMSO-*d*<sub>6</sub>)  $\delta = 177.7, 172.6, 172.0, 171.6, 171.5, 138.0, 137.4, 129.6, 129.5, 128.7, 128.4, 127.0, 126.7, 61.0, 54.2, 53.6, 51.4, 48.6, 41.5, 38.4, 38.0, 37.2, 27.7, 24.4, 23.5, 22.1, 17.9$ ; HRESI(+)-MS:  $m/z$  609.3634 [M+H]<sup>+</sup>, 609.3647 calc. for C<sub>34</sub>H<sub>49</sub>O<sub>6</sub>N<sub>4</sub>.

**Piv-Ala-Leu-Phe-Phe-OH E4**

To a solution of **E3** (0.20 g, 0.33 mmol, 1.00 equiv.) in MeOH (10 ml) was added aqueous 1M NaOH (2.00 ml, 2.00 mmol, 6.00 equiv.) and the reaction mixture was stirred at room temperature for 2 h. The solvent was removed under reduced pressure. The residue was dissolved in EtOAc (25 ml) and the aqueous layer was washed with H<sub>2</sub>O (25 ml). The aqueous layer was acidified with aqueous 1M HCl to pH 3. The aqueous layer was extracted with EtOAc (2 × 20 ml). The combined organic layers were washed with H<sub>2</sub>O (25 ml), brine (25 ml), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure to afford **E4** (180 mg, 94%) as white solid.

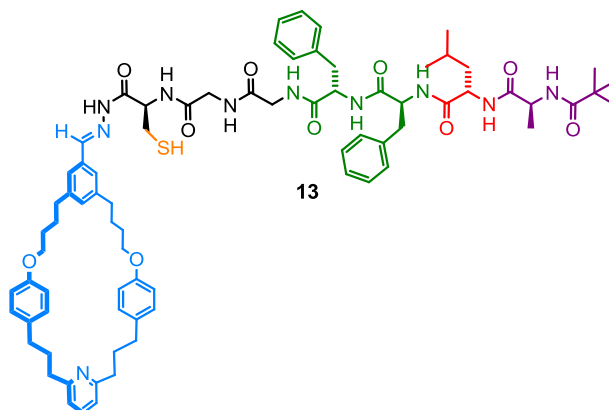
m.p. 123 °C;  $[\alpha]_{\text{D}} = -0.09^{\circ}$  (c 0.17, MeOH); <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  12.77 (bs, 1H, H<sub>v</sub>), 8.27 (d,  $J = 7.7$  Hz, 1H, H<sub>p</sub>), 7.88 (d,  $J = 8.3$  Hz, 1H, H<sub>j</sub>), 7.64 (d,  $J = 8.2$  Hz, 1H, H<sub>e</sub>), 7.43 (d,  $J = 7.5$  Hz, 1H, H<sub>b</sub>), 7.22 (m, 10H, H<sub>m+n+o+s+t+u</sub>), 4.53 (td,  $J = 9.5$  Hz, 4.7 Hz, 1H, H<sub>k</sub>), 4.44 (q,  $J = 7.1$  Hz, 1H, H<sub>q</sub>), 4.22 (m, 2H, H<sub>b+f</sub>), 3.09 – 2.97 (m, 2H, H<sub>l+r</sub>), 2.92 (dd,  $J = 14.0$  Hz, 8.6 Hz, 1H, H<sub>r</sub>), 2.77 (dd,  $J = 14.0$  Hz, 9.4 Hz, 1H, H<sub>l</sub>), 1.52 – 1.44 (m, 1H, H<sub>h</sub>), 1.31 (t,  $J = 7.3$  Hz, 2H, H<sub>g</sub>), 1.17 (d,  $J = 7.1$  Hz, 3H, H<sub>d</sub>), 1.09 (s, 9H, H<sub>a</sub>), 0.83 (d,  $J = 6.5$  Hz, 3H, H<sub>i</sub>), 0.79 (d,  $J = 6.5$  Hz, 3H, H<sub>i</sub>); <sup>13</sup>C NMR (150 MHz, DMSO-*d*<sub>6</sub>)  $\delta$  = 177.7, 173.1, 172.5, 171.9, 171.3, 138.0, 137.8, 129.6, 129.6, 128.7, 128.4, 126.9, 126.7, 53.9, 53.7, 51.4, 48.6, 41.5, 38.4, 37.9, 37.1, 27.7, 24.4, 22.1, 17.9; HRESI(+)-MS:  $m/z$  603.3140 [M+Na]<sup>+</sup>, 603.3153 calc. for C<sub>32</sub>H<sub>44</sub>O<sub>6</sub>N<sub>4</sub>Na.

**Piv-Ala- Leu-Phe-Phe-Gly-Gly-Cys(STrt)-Macrocyclic Hydrazone E7**

**E6** (23 mg, 20  $\mu\text{mol}$ , 1.00 equiv.) and aldehyde macrocycle **7** (33 mg, 59  $\mu\text{mol}$ , 3.0 equiv.) were dissolved in DMSO (1.5 ml). Aqueous 2-(*N*-morpholino)ethanesulfonic acid -buffer (0.5ml, pH 6.0) and aniline (50  $\mu\text{l}$ ) was added. The reaction mixture was stirred at 60  $^{\circ}\text{C}$  for 72 h. The reaction mixture was allowed to cool to room temperature and then diluted with EtOAc (20 ml). The organic layer was washed with  $\text{H}_2\text{O}$  (20 ml) and the aqueous layer extracted with EtOAc (2  $\times$  20 ml). The combined organic layers were washed with  $\text{H}_2\text{O}$  (20 ml), brine (20 ml), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and the solvent was removed under reduced pressure. Flash column chromatography ( $\text{SiO}_2$ ,  $\text{CH}_2\text{Cl}_2/\text{MeOH}$  4%) of the residue afforded **E7** (18 mg, 56% yield) as yellowish oil.

1:2 mixture of *E* (labelled 1)- and *Z* (labelled 2)-hydrazone. Several amide-rotamers are present.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 11.51 (s, 1H,  $\text{H}_{o1}$ ), 11.48 (s, 1H,  $\text{H}_{o2}$ ), 8.44 - 7.96 (m, 5H,  $\text{H}_{v1+v2+x1+x2+z1+z2+ap1+ap2}$ ), 8.18 (s, 1H,  $\text{H}_{s1}$ ), 7.93 (s, 1H,  $\text{H}_{s2}$ ), 7.86–7.63 (m, 2H,  $\text{H}_{as1+as2+au1+au2}$ ), 7.59 (t,  $J$  = 7.6 Hz, 1H,  $\text{H}_{al+a2}$ ), 7.37 – 7.12 (m, 28H,  $\text{H}_{j1+j2+m1+m2+j1+j2+s1+s2+t1+t2+u1+u2+ac1+ac2+ad1+ad2+ae1+ae2+ai1+ai2+aj1+aj2+ak1+ak2}$ ), 7.06 (d,  $J$  = 7.7 Hz, 2H,  $\text{H}_{b1+b2}$ ), 7.02 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_{f1}$ ), 7.00 (d,  $J$  = 8.6 Hz, 2H,  $\text{H}_{f2}$ ), 6.74 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_{g1}$ ), 6.68 (d,  $J$  = 8.5 Hz, 2H,  $\text{H}_{g2}$ ), 5.41 (td,  $J$  = 8.5 Hz, 4.1 Hz, 1H,  $\text{H}_{p2}$ ), 4.64 – 4.41 (m, 3H,  $\text{H}_{v1+aa1+aa2+ag1+ag2}$ ), 4.28 – 4.19 (m, 2H,  $\text{H}_{am1+am2+ar1+ar2}$ ), 3.89 (t,  $J$  = 5.8 Hz, 4H,  $\text{H}_{l1}$ ), 3.79 (m, 8H,  $\text{H}_{l2+w1+w2+y1+y2}$ ), 3.10 – 3.04 (m, 4H,  $\text{H}_{ab2+ah2}$ ), 2.88 – 2.74 (m, 4H,  $\text{H}_{ab1+ah1}$ ), 2.70 (t,  $J$  = 7.4 Hz, 4H,  $\text{H}_{c1+c2}$ ), 2.67 – 2.63 (m, 4H,  $\text{H}_{k1+k2}$ ), 2.47 (q,  $J$  = 7.6 Hz, 4H,  $\text{H}_{e1+e2}$ ), 2.43 – 2.38 (m, 2H,  $\text{H}_{q1+q2}$ ), 1.93 (p,  $J$  = 7.7 Hz, 4H,  $\text{H}_{d1+d2}$ ), 1.77 – 1.69 (m, 4H,  $\text{H}_{j1+j2}$ ), 1.69 – 1.61 (m, 4H,  $\text{H}_{i1+i2}$ ), 1.51 – 1.45 (m, 1H,  $\text{H}_{ao1+ao2}$ ), 1.32 – 1.27 (m, 2H,  $\text{H}_{an1+an2}$ ), 1.19 (dd,  $J$  = 7.1 Hz, 1.6 Hz, 3H,  $\text{H}_{ar1}$ ), 1.14 (dd,  $J$  = 7.1 Hz, 1.5 Hz, 3H,  $\text{H}_{ar1}$ ), 1.11 – 1.05 (m, 9H,  $\text{H}_{au1+au2}$ ), 0.82 – 0.75 (m, 6H,  $\text{H}_{ap1}$ ), 0.72 – 0.67 (m, 6H,  $\text{H}_{ap2}$ );  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ )  $\delta$  = 177.6, 173.0, 172.8, 172.6, 172.2, 171.5, 169.3, 169.1, 161.0, 157.2, 148.6, 144.7, 144.6, 143.0, 138.4, 138.1, 137.1, 134.4, 134.4, 129.7, 129.7, 129.6, 129.7, 129.5,

128.6, 128.5, 128.4, 128.4, 128.3, 127.3, 127.1, 129.4, 125.4, 125.2, 120.6, 114.6, 114.5,  
66.5, 66.1, 54.5, 54.4, 54.4, 54.2, 54.2, 51.2, 51.1, 49.1, 48.9, 42.5, 42.3, 41.7, 41.6, 41.2,  
38.4, 38.4, 38.4, 38.3, 37.0, 37.0, 34.8, 34.7, 34.2, 34.1, 31.8, 28.3, 28.3, 28.1, 27.7, 24.5,  
24.3, 23.6, 23.5, 23.4, 23.3, 18.6, 18.6.

**Piv-Ala- Leu-Phe-Phe-Gly-Gly-Cys(SH)-Macrocyclic Hydrazone 13**

**E7** (15 mg, 9.4  $\mu\text{mol}$ , 1.0 equiv.) was dissolved in  $\text{CH}_2\text{Cl}_2/\text{TFA}$  (4:1 v/v, 1 ml) and triisopropylsilane (5.0  $\mu\text{l}$ , 24  $\mu\text{mol}$ , 2.6 equiv.) was added. The mixture was stirred at room temperature for 90 minutes. The solvent was removed under reduced pressure and the residue washed with  $\text{Et}_2\text{O}$  (3  $\times$  10ml) and analysed without further purification. LRESI(+)-MS:  $m/z$  1355.50  $[\text{M}+\text{H}]^+$ .

## 4.6 References

- (1) Voet, D.; Voet, J. G. *Biochemistry*; fourth edition; John Wiley & Sons: New York, 2011.
- (2) Kay, E. R.; Leigh, D. A.; Zerbetto, F. *Angew. Chem. Int. Ed.* **2007**, *46*, 72.
- (3) (a) Berna, J.; Leigh, D. A.; Lubomska, M.; Mendoza, S. M.; Perez, E. M.; Rudolf, P.; Teobaldi, G.; Zerbetto, F. *Nat. Mater.* **2005**, *4*, 704 (b) Liu, Y.; Flood, A. H.; Bonvallet, P. A.; Vignon, S. A.; Northrop, B. H.; Tseng, H.-R.; Jeppesen, J. O.; Huang, T. J.; Brough, B.; Baller, M.; Magonov, S.; Solares, S. D.; Goddard, W. A.; Ho, C.-M.; Stoddart, J. F. *J. Am. Chem. Soc.* **2005**, *127*, 9745 (c) Eelkema, R.; Pollard, M. M.; Vicario, J.; Katsonis, N.; Ramon, B. S.; Bastiaansen, C. W. M.; Broer, D. J.; Feringa, B. L. *Nature* **2006**, *440*, 163 (d) Lussis, P.; Svaldo-Lanero, T.; Bertocco, A.; Fustin, C.-A.; Leigh, D. A.; Duwez, A.-S. *Nat. Nanotechnol.* **2011**, *6*, 553.
- (4) (a) Yoon, H. J.; Kuwabara, J.; Kim, J.-H.; Mirkin, C. A. *Science* **2010**, *330*, 66 (b) Sohtome, Y.; Tanaka, S.; Takada, K.; Yamaguchi, T.; Nagasawa, K. *Angew. Chem. Int. Ed.* **2010**, *49*, 9254 (c) Zirngast, M.; Pump, E.; Leitgeb, A.; Albering, J. H.; Slugovc, C. *Chem. Commun.* **2011**, *47*, 2261 (d) Wang, J.; Feringa, B. L. *Science* **2011**, *331*, 1429 (e) Blanco, V.; Carlone, A.; Hänni, K. D.; Leigh, D. A.; Lewandowski, B. *Angew. Chem. Int. Ed.* **2012**, *51*, 5166 (f) Mortezaei, S.; Catarineu, N. R.; Canary, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 8054 (g) Neilson, B. M.; Bielawski, C. W. *J. Am. Chem. Soc.* **2012**, *134*, 12693 (h) Schmittel, M.; De, S.; Pramanik, S. *Angew. Chem. Int. Ed.* **2012**, *51*, 3832 (i) Schmittel, M.; De, S.; Pramanik, S. *Angew. Chem.* **2012**, *124*, 3898 (j) Wilson, D.; Branda, N. R. *Angew. Chem. Int. Ed.* **2012**, *51*, 5431 (k) Wilson, D.; Branda, N. R. *Angew. Chem.* **2012**, *124*, 5527.
- (5) Miyagawa, N.; Watanabe, M.; Matsuyama, T.; Koyama, Y.; Moriuchi, T.; Hirao, T.; Furusho, Y.; Takata, T. *Chem. Commun.* **2010**, *46*, 1920.
- (6) (a) Thordarson, P.; Bijsterveld, E. J. A.; Rowan, A. E.; Nolte, R. J. M. *Nature* **2003**, *424*, 915 (b) Monnereau, C.; Ramos, P. H.; Deutman, A. B. C.; Elemans, J. A. A. W.; Nolte, R. J. M.; Rowan, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 1529.
- (7) (a) Fletcher, S. P.; Dumur, F.; Pollard, M. M.; Feringa, B. L. *Science* **2005**, *310*, 80 (b) Yoon, H. J.; Kuwabara, J.; Kim, J.-H.; Mirkin, C. A. *Science* **2010**, *330*, 66.

- (8) (a) He, Y.; Liu, D. R. *Nat. Nanotechnol.* **2010**, *5*, 778 (b) Wickham, S. F. J.; Bath, J.; Katsuda, Y.; Endo, M.; Hidaka, K.; Sugiyama, H.; Turberfield, A. J. *Nat. Nanotechnol.* **2012**, *7*, 169 (c) He, Y.; Liu, D. R. *J. Am. Chem. Soc.* **2011**, *133*, 9972 (d) Niu, J.; Hili, R.; Liu, D. R. *Nat. Chem.* **2013**, *5*, 282.
- (9) Gu, H.; Chao, J.; Xiao, S.-J.; Seeman, N. C. *Nature* **2010**, *465*, 202.
- (10) (a) Kay, E. R.; Leigh, D. A. *Pure Appl. Chem.* **2008**, *80*, 17 (b) Yang, W.; Li, Y.; Liu, H.; Chi, L.; Li, Y. *Small* **2012**, *8*, 504.
- (11) (a) Molecules labelled **1**, **7**, and **8** correspond to molecules **11**, **14**, and **13** in Chapter 2, respectively. Their syntheses and characterisation data are reported in the Experimental Section 2.5. (b) Molecules labelled **2**, **5**, and **9** correspond to molecules **2**, **6**, and **4** in Chapter 3, respectively. Their syntheses and characterisation data are reported in the Experimental Section 3.5.
- (12) (a) Aucagne, V.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Walker, D. B. *J. Am. Chem. Soc.* **2006**, *128*, 2186 (b) Berná, J.; Crowley, J. D.; Goldup, S. M.; Hänni, K. D.; Lee, A.-L.; Leigh, D. A. *Angew. Chem. Int. Ed.* **2007**, *46*, 5709 (c) Aucagne, V.; Berná, J.; Crowley, J. D.; Goldup, S. M.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Ronaldson, V. E.; Slawin, A. M. Z.; Viterisi, A.; Walker, D. B. *J. Am. Chem. Soc.* **2007**, *129*, 11950 (d) Berná, J.; Goldup, S. M.; Lee, A.-L.; Leigh, D. A.; Symes, M. D.; Teobaldi, G.; Zerbetto, F. *Angew. Chem.* **2008**, *120*, 4464 (e) Berná, J.; Goldup, S. M.; Lee, A.-L.; Leigh, D. A.; Symes, M. D.; Teobaldi, G.; Zerbetto, F. *Angew. Chem. Int. Ed.* **2008**, *47*, 4392 (f) Crowley, J. D.; Goldup, S. M.; Lee, A.-L.; Leigh, D. A.; McBurney, R. T. *Chem. Soc. Rev.* **2009**, *38*, 1530 (g) Goldup, S. M.; Leigh, D. A.; Long, T.; McGonigal, P. R.; Symes, M. D.; Wu, J. *J. Am. Chem. Soc.* **2009**, *131*, 15924 (h) Goldup, S. M.; Leigh, D. A.; McBurney, R. T.; McGonigal, P. R.; Plant, A. *Chem. Sci.* **2010**, *1*, 383 (i) Crowley, J. D.; Hänni, K. D.; Leigh, D. A.; Slawin, A. M. Z. *J. Am. Chem. Soc.* **2010**, *132*, 5309 (j) Crowley, J. D.; Goldup, S. M.; Gowans, N. D.; Leigh, D. A.; Ronaldson, V. E.; Slawin, A. M. Z. *J. Am. Chem. Soc.* **2010**, *132*, 6243 (k) Barran, P. E.; Cole, H. L.; Goldup, S. M.; Leigh, D. A.; McGonigal, P. R.; Symes, M. D.; Wu, J.; Zengerle, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 12280 (l) Barran, P. E.; Cole, H. L.; Goldup, S. M.; Leigh, D. A.; McGonigal, P. R.; Symes, M. D.; Wu, J.; Zengerle, M. *Angew. Chem.* **2011**, *123*, 12488.

- (13) (a) Fernandes, A.; Viterisi, A.; Coutrot, F.; Potok, S.; Leigh, D. A.; Aucagne, V.; Papot, S. *Angew. Chem. Int. Ed.* **2009**, *48*, 6443 (b) Fernandes, A.; Viterisi, A.; Coutrot, F.; Potok, S.; Leigh, D. A.; Aucagne, V.; Papot, S. *Angew. Chem.* **2009**, *121*, 6565 (c) Fernandes, A.; Viterisi, A.; Aucagne, V.; Leigh, D. A.; Papot, S. *Chem. Commun.* **2012**, *48*, 2083.
- (14) Aucagne, V.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Walker, D. B. *J. Am. Chem. Soc.* **2006**, *128*, 2186.
- (15) Zheng, H.; Zhou, W.; Lv, J.; Yin, X.; Li, Y.; Liu, H.; Li, Y. *Chem. –Eur. J.* **2009**, *15*, 13253.

## Outlook

The molecular machines capable of sequence-specific synthesis of oligopeptides presented in this thesis represent a significant advance in the development of novel molecular task performance inspired by nature. A rotaxane architecture performing highly complex tasks, generating an operation product that would not be possible with standard solution phase chemistry, therefore paves the way to the realisation for even more complicated systems.

The ultimate goal of sequence-specific synthesis with artificial molecular machines would be the generation of polymers completely unrelated to the template that holds the information to be translated. Ideally, several reactions including C-C bond formations or a mixture of different co-polymers would be compatible with such a hypothetical machine. The individual bond forming reactions should be high yielding, fast and translate information with great accuracy, allowing access to very long information-rich polymer strands. Also desirable would be a device that—similar to the ribosome—can be reused in another translation event.

The initial design proves that enzyme-free synthetic translation can be achieved with interlocked architectures, but still the level of performance of nature's nanomachinery is out of reach. The system presented in this work destroys the information that is encoded in the track and is limited inherently to the synthesis of short oligomers. The individual bond forming NCL-steps are relatively slow—12 h per amide bond formation event—and also the hypothetical extension of the systems to longer polymers would be problematic because of the increasing size of the transition states that are encountered as the thiolate catalyst advances. Furthermore, the thread component has to be synthesised in a lengthy protection/deprotection sequence.

Several projects are currently ongoing in the Leigh group that address shortcomings of this initial design. The use of tracks which incorporate aminoacyl residues synthesised in a single polymerisation by sequential addition of aminoacyl succinimide derivatives would allow the extension of the system and easy access to longer polymer templates. A second project investigates the use of transacylation catalysts similar to the first generation design as discussed in Chapter 2 and would allow for translation without increasing the size of the transition states. Finally, a third project attempts the incorporation of reloadable barriers into a rotaxane architecture, therefore allowing the machine to be reused for a second or even third operation, thereby maintaining the information on the thread. These innovative new systems could bring us even closer towards the development of artificial ribosomes—a fascinating prospect with still a long way to go.

## Appendix

### Published Paper

*Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine*, Lewandowski, B.; De Bo, G.; Ward, J. W.; Pappmeyer, M.; Kuschel, S.; Aldegunde, M. J.; Gramlich, P. M. E.; Heckmann, D.; Goldup, S. M.; D'Souza, D. M.; Fernandes, A. E.; Leigh, D. A. *Science* **2013**, 339, 189-193.

spectively.  $\rho$  and  $M$  are the density and molecular weight of water, respectively.  $\Delta G_{\text{cycle}}$  is the molar Gibbs free-energy change of absorbed water during one expansion/contraction cycle.  $f_{\text{ad}}$  is the adhesive force coefficient between PEE-PPy films and moist substrates. Given a certain  $E$ ,  $R$ , and  $f_{\text{ad}}$ , Eqs. 1 and 2 roughly define a theoretical maximum and minimum limit on the required thickness of the actuator to perform fast locomotion. In practice, we found that the optimal thickness for PEE-PPy actuators on a moist paper was roughly 15 to 40  $\mu\text{m}$ . Actuators thinner than 15  $\mu\text{m}$  tended to stick to the moist paper, whereas actuators thicker than 40  $\mu\text{m}$  showed significantly slower locomotion.

Because this PEE-PPy actuator could continuously extract chemical potential energy out of ambient water gradients to perform mechanical work, it should be able to drive a piezoelectric element to convert the mechanical energy into electrical energy. A 9- $\mu\text{m}$ -thick piezoelectric polyvinylidene difluoride (PVDF) film was metallized, wired, and insulated on both faces (Fig. 4A). A 27- $\mu\text{m}$ -thick PEE-PPy actuator was attached to one face of the PVDF element. When placed on a moist substrate with the actuator facing down, the actuator bent and stretched the PVDF element repeatedly (movie S5), generating an open-circuit voltage up to 3 V. A 10-megohm resistor was loaded onto this generator (Fig. 4B), and the peak output reached  $\sim 1.0$  V (Fig. 4D). Analysis indicated that the frequency of the alternating voltage signal was  $\sim 0.3$  Hz (fig. S12), which matched the motion frequency of the generator (movie S5). The average power output was 5.6 nW (fig. S13), which corresponded to a power density of 56  $\mu\text{W}/\text{kg}$  for the 100-mg generator. In contrast, the same PVDF element did not move on the moist substrate, and the recording showed only noise (fig. S14), with analysis of this background noise giving an average power output of 0.015 nW (fig. S15). The

generated alternating electrical pulses by the generator were rectified using a commercial full-wave bridge rectifier, then stored in a 2.2  $\mu\text{F}$  capacitor (Fig. 4C). Within 7 min of charging, the voltage of the capacitor was saturated to  $\sim 0.66$  V (Fig. 4E). This was lower than the peak output voltage of the generator, possibly due to voltage drop across the rectifying diodes and/or current leakage of the capacitor.

This PEE-PPy polymer composite system features an interpenetrating network of a rigid polymer with a soft, hydrolytically sensitive polymer that can perform water-gradient-induced displacement, converting the chemical potential energy in water gradients to mechanical work. Besides mechanical vibration energy, the generator based on this powerful actuator can use ubiquitous low-temperature water gradients as its energy source, in contrast to state-of-the-art piezoelectric energy scavengers that rely solely on mechanical vibration energy (26). Thus, the water-gradient-driven actuator and generator demonstrated potential applications as sensors, switches, and power sources for ultralow-power devices.

#### References and Notes:

1. S. Minko, *Responsive Polymer Materials: Design and Applications* (Blackwell, Ames, IA, 2006).
2. E. W. H. Jager, E. Smela, O. Inganäs, *Science* **290**, 1540 (2000).
3. Y. Osada, H. Okuzaki, H. Hori, *Nature* **355**, 242 (1992).
4. A. Lendlein, H. Y. Jiang, O. Jünger, R. Langer, *Nature* **434**, 879 (2005).
5. A. Lendlein, R. Langer, *Science* **296**, 1673 (2002).
6. M. Camacho-Lopez, H. Finkelmann, P. Palfy-Muhoray, M. Shelley, *Nat. Mater.* **3**, 307 (2004).
7. C. L. van Oosten, C. W. M. Bastiaansen, D. J. Broer, *Nat. Mater.* **8**, 677 (2009).
8. Y. Yu, M. Nakano, T. Ikeda, *Nature* **425**, 145 (2003).
9. J. Kim, J. A. Hanna, M. Byun, C. D. Santangelo, R. C. Hayward, *Science* **335**, 1201 (2012).
10. A. Sidorenko, T. Krupenkin, A. Taylor, P. Fratzl, J. Aizenberg, *Science* **315**, 487 (2007).
11. D. J. Beebe *et al.*, *Nature* **404**, 588 (2000).

12. G. H. Chen, A. S. Hoffman, *Nature* **373**, 49 (1995).
13. J. M. Bennis, J. S. Choi, R. I. Mahato, J. S. Park, S. W. Kim, *Bioconjug. Chem.* **11**, 637 (2000).
14. M. E. Calderera-Moore, W. B. Liechty, N. A. Peppas, *Acc. Chem. Res.* **44**, 1061 (2011).
15. P. Fratzl, F. G. Barth, *Nature* **462**, 442 (2009).
16. E. Smela, *Adv. Mater.* **15**, 481 (2003).
17. R. H. Baughman, *Science* **308**, 63 (2005).
18. H. Okuzaki, T. Kuwabara, T. Kunugi, *J. Polym. Sci. B Polym. Phys.* **36**, 2237 (1998).
19. T. Ushiki, *Arch. Histol. Cytol.* **65**, 109 (2002).
20. M. Shibayama, M. Sato, Y. Kimura, H. Fujiwara, S. Nomura, *Polymer (Guildf.)* **29**, 336 (1988).
21. T. Park, S. C. Zimmerman, *J. Am. Chem. Soc.* **128**, 13986 (2006).
22. X. M. Ren, P. G. Pickup, *J. Phys. Chem.* **97**, 5356 (1993).
23. M. Szafran, Z. Degaszfran, *J. Mol. Struct.* **321**, 57 (1994).
24. S. Hara, T. Zama, W. Takashima, K. Kaneto, *Polym. J.* **36**, 151 (2004).
25. J. D. Madden, P. G. Madden, P. A. Anquetil, I. W. Hunter, in *Electroactive Polymers and Rapid Prototyping* (Materials Research Society, Warrendale, PA, 2002), vol. 698, p. 137.
26. Y. Qin, X. Wang, Z. L. Wang, *Nature* **451**, 809 (2008).

**Acknowledgments:** M.M. and R.L. conceived the idea and designed the experiments. M.M. and L.G. performed the experiments. M.M., L.G., D.G.A., and R.L. contributed materials and/or tools, analyzed the data, and wrote the paper. This research was supported in part by a National Heart, Lung, and Blood Institute Program of Excellence in Nanotechnology (PEN) Award, contract no. HHSN268201000045C; National Cancer Institute grant CA151884; and Armed Forces Institute of Regenerative Medicine Award no. W81XWH-08-2-0034. Experimental procedures, additional data for materials characterization, and the device test are presented in the supplementary materials. We thank D. Bong for thoughtful discussion and N. Zhang for help in the preparation of figures.

#### Supplementary Materials

www.sciencemag.org/cgi/content/full/339/6116/186/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S16  
Tables S1 and S2  
Eqs. S1 to S3  
Movies S1 to S5  
References (27–32)

17 September 2012; accepted 9 November 2012  
10.1126/science.1230262

## Sequence-Specific Peptide Synthesis by an Artificial Small-Molecule Machine

Bartosz Lewandowski,<sup>1</sup> Guillaume De Bo,<sup>1</sup> John W. Ward,<sup>1</sup> Marcus Pappmeyer,<sup>1</sup> Sonja Kuschel,<sup>1</sup> María J. Aldegunde,<sup>2</sup> Philipp M. E. Gramlich,<sup>2</sup> Dominik Heckmann,<sup>2</sup> Stephen M. Goldup,<sup>2</sup> Daniel M. D'Souza,<sup>2</sup> Antony E. Fernandes,<sup>2</sup> David A. Leigh<sup>1,2\*</sup>

The ribosome builds proteins by joining together amino acids in an order determined by messenger RNA. Here, we report on the design, synthesis, and operation of an artificial small-molecule machine that travels along a molecular strand, picking up amino acids that block its path, to synthesize a peptide in a sequence-specific manner. The chemical structure is based on a rotaxane, a molecular ring threaded onto a molecular axle. The ring carries a thiolate group that iteratively removes amino acids in order from the strand and transfers them to a peptide-elongation site through native chemical ligation. The synthesis is demonstrated with  $\sim 10^{18}$  molecular machines acting in parallel; this process generates milligram quantities of a peptide with a single sequence confirmed by tandem mass spectrometry.

Cells achieve the sequence-specific synthesis of information-rich oligomers and polymers through the operation of complex

molecular machines that transcribe information from the genetic code (1). The most extraordinary of these is the ribosome (2–4), a  $\sim 2.6$ -MD

(bacterial) to  $\sim 4.3$ -MD (eukaryotic) molecular machine found in all living cells that assembles amino acids from tRNA building blocks into a peptide chain with an order defined by the sequence of the mRNA strand that it moves along. Artificial small-molecule machines (5) have previously been used to store information (6, 7) and do mechanical work (8–11); others have been employed in synthesis to processively epoxidize an unsaturated polymer (12, 13), switch “on” and “off” catalytic activity (14–17), and change the handedness of a reaction product (18). Large synthetic DNA molecules have been used to guide the formation of bonds between unnatural building blocks (19–22) and assemble

<sup>1</sup>School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK. <sup>2</sup>School of Chemistry, University of Edinburgh, The King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK.

\*To whom correspondence should be addressed. E-mail: david.leigh@manchester.ac.uk



sequence, through chemistry related to non-ribosomal peptide synthesis (24).

The chemical structure of the artificial molecular machine, **1**, is shown in Fig. 1. Strand **2** bears three amino acids attached to the track by weak phenolic ester linkages (25) and separated from each other by rigid spacers that minimize the possibility of the reactive arm of the machine coming into contact and reacting with a building block out of sequence (26). Macrocycle **3** contains an endocyclic pyridine group that directs the

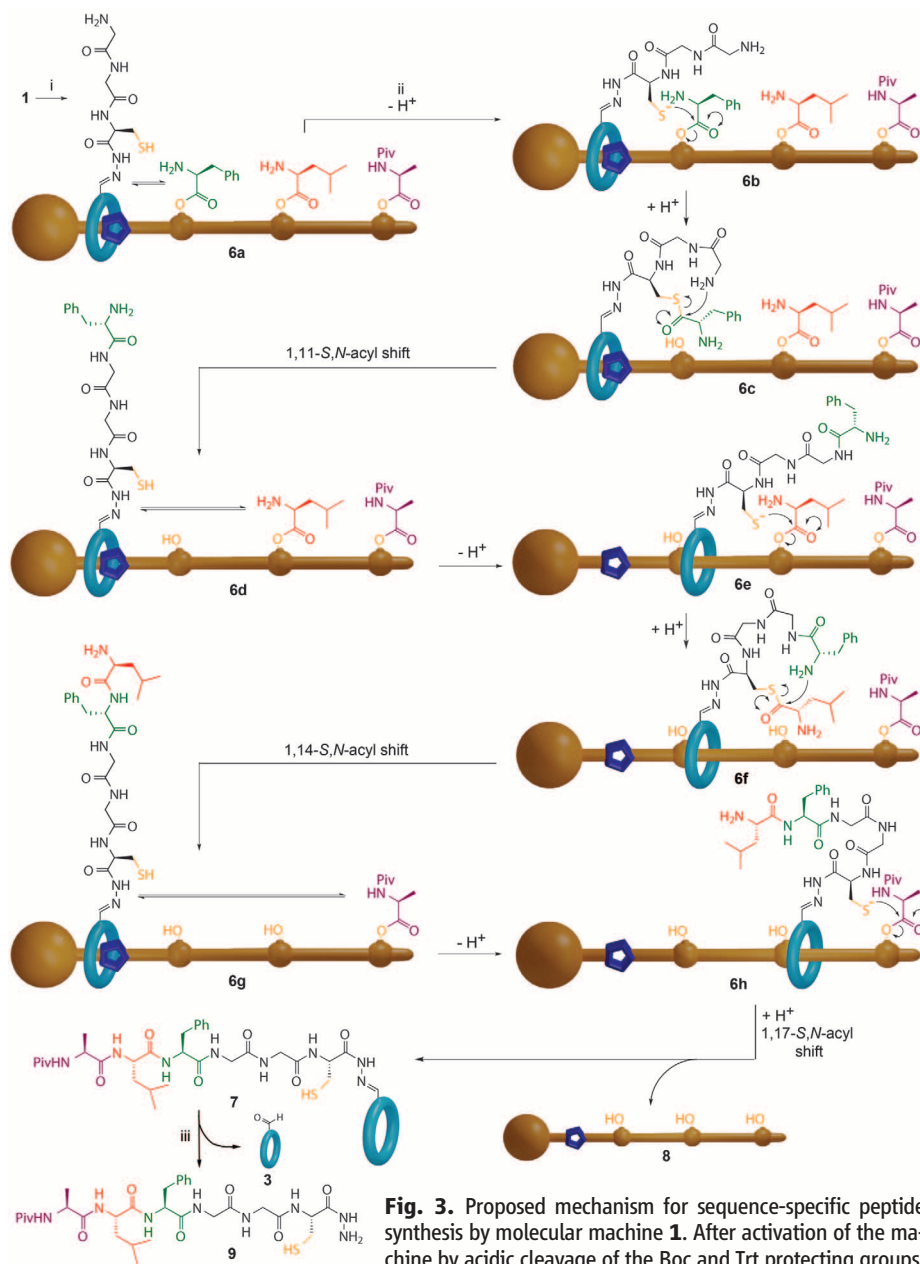
threading of strand **2** during the Cu(I)-catalyzed cycloaddition of the terminal alkyne with the azide-bearing stopper group **5**, leading to the assembly of rotaxane **4** in 30% yield (Fig. 1, i). This active template (27, 28) strategy ensured that the resulting threaded structure did not have residual attractive intercomponent interactions that would tend to localize the position of the ring rather than allow it to move freely up and down the strand between blocking groups. Once we assembled the macrocycle-strand-stopper

conjugate **4**, we used reversible hydrazone exchange to introduce a cysteine derivative bearing the reactive arm [a trityl (Trt)-protected thiol group] and the site for peptide elongation [a *tert*-butoxycarbonyl carbamate (Boc)-protected amine at the end of a glycylglycine residue] (Fig. 1, ii). The fully assembled machine **1** is stable in its protected form, with upfield shifts of the  $H_{P1}$  triazole and nearby  $H_{S12-S14}$  proton signals evident in the  $^1H$  nuclear magnetic resonance (NMR) spectrum on account of shielding from the phenyl rings of the macrocycle, confirming that the ring is trapped in the region of the strand between the terminal stopper and the Boc-phenylalanine ester (Fig. 2).

We used acid-catalyzed cleavage of the Boc and trityl protecting groups (Fig. 3, i) to activate the molecular machine and then allowed it to operate (Fig. 3, ii) at 60°C under microwave heating in a 3:1 acetonitrile:dimethylformamide solution in the presence of *N,N*-diisopropylethylamine (a non-nucleophilic base) and *tris*(2-carboxyethyl)phosphine (a reducing agent that cleaves any disulfide bonds formed through thiol oxidation). The design of the machine is such that once the thiolate residue of the cysteine group (**6a**) is deprotected, it is poised to undergo a transacylation reaction with the first amino acid phenolic ester that blocks the macrocycle's path on the track (**6b**). We hypothesized that the subsequently formed phenylalanine thioester (**6c**) would be able to react further, transferring the amino acid by native chemical ligation (29) to the glycylglycine amine group by an 11-membered-ring transition state [the dipeptide spacer between the cysteine residue and the amine of the peptide-elongation site was introduced because native chemical ligation is reported to be very slow via 8-membered-ring transition states (30)]. This sequence simultaneously transfers the amino acid to the end of the growing peptide (**6d**) and regenerates the catalytic thiolate group, ready for the cleavage and transfer of further building blocks. *S-N* acyl transfer is a key feature of nonribosomal peptide synthesis (24).

Once the covalent bond connecting an amino acid to the strand is broken, the macrocycle is able to move further along the track until its path is blocked by the next amino acid group (**6d**). The *O-S* acyl transfer/*S-N* acyl transfer/catalyst regeneration/ring movement process continues (**6e** to **6h**) until the last amino acid on the track is cleaved (**6h**) and the macrocycle detaches from the strand (**8**) with the newly formed, full length, peptide attached (**7**). The artificial molecular machine synthesizes the peptide from the C terminus to the N terminus, the opposite direction of ribosomal translation (2–4).

After a 36-hour operation of **6a** at 60°C, no starting material remained, as shown by high-performance liquid chromatography (HPLC), and two major products were isolated from the reaction mixture (26). We used  $^1H$  NMR spectroscopy and mass spectrometry to identify one product as the completely deacylated thread, **8**. The other product had a  $^1H$  NMR spectrum and molecular



**Fig. 3.** Proposed mechanism for sequence-specific peptide synthesis by molecular machine **1**. After activation of the machine by acidic cleavage of the Boc and Trt protecting groups, under basic conditions successive native chemical ligation reactions transfer the amino acid building blocks to the peptide-elongation site on the macrocycle in the order they appear on the thread. Once the final amino acid is cleaved, the macrocycle bearing the synthesized oligopeptide **7** detaches from the strand. The hydrazide peptide **9** is subsequently released from the macrocycle by hydrolysis. Reaction conditions: (i) 20%  $CF_3CO_2H$  in dichloromethane, room temperature, 2 hours, 100%. (ii)  $((CH_3)_2CH)_2NEt$ ,  $(HO_2CCH_2CH_2)_3P$  in 3:1 acetonitrile:dimethylformamide, 60°C, 36 hours. Et, ethyl. (iii) 30%  $CF_3CO_2H$  in 3:1 dichloromethane:water, room temperature, 18 hours.

reactions transfer the amino acid building blocks to the peptide-elongation site on the macrocycle in the order they appear on the thread. Once the final amino acid is cleaved, the macrocycle bearing the synthesized oligopeptide **7** detaches from the strand. The hydrazide peptide **9** is subsequently released from the macrocycle by hydrolysis. Reaction conditions: (i) 20%  $CF_3CO_2H$  in dichloromethane, room temperature, 2 hours, 100%. (ii)  $((CH_3)_2CH)_2NEt$ ,  $(HO_2CCH_2CH_2)_3P$  in 3:1 acetonitrile:dimethylformamide, 60°C, 36 hours. Et, ethyl. (iii) 30%  $CF_3CO_2H$  in 3:1 dichloromethane:water, room temperature, 18 hours.

weight (Fig. 4) consistent with the macrocycle bearing the hydrazone linked to the hexapeptide (Piv)AlaLeuPheGlyGlyCys, **7** [Piv, COC(CH<sub>3</sub>)<sub>3</sub>].

To confirm that the product of the molecular machine's operation had the amino acids assembled in the correct order, we used tandem mass spectrometry (MS/MS) to determine the peptide sequence. We then validated the sequence by comparing it with an authentic sample and an isomer in which the order of the Phe and Leu residues was reversed, each prepared unambiguously by conventional peptide synthesis (26). Figure 4A shows the <sup>1</sup>H NMR spectrum of the molecular machine product **7**, and Fig. 4B shows the MS/MS spectrum of one isotope of a 2+ ion of **7** derivatized through the cysteine as an *S,N*-acetal, a species that gave a sufficient signal-to-noise ratio for the MS/MS experiment. Figure 4C shows the superimposition of the MS/MS spectra of a similar isotope and ion from the authentic samples of the macrocycle bearing the sequence (Piv)AlaPheLeuGlyGlyCys (red peaks) and (Piv)AlaLeuPheGlyGlyCys (blue peaks). The difference in the fragmentation masses of the two sequence isomers is

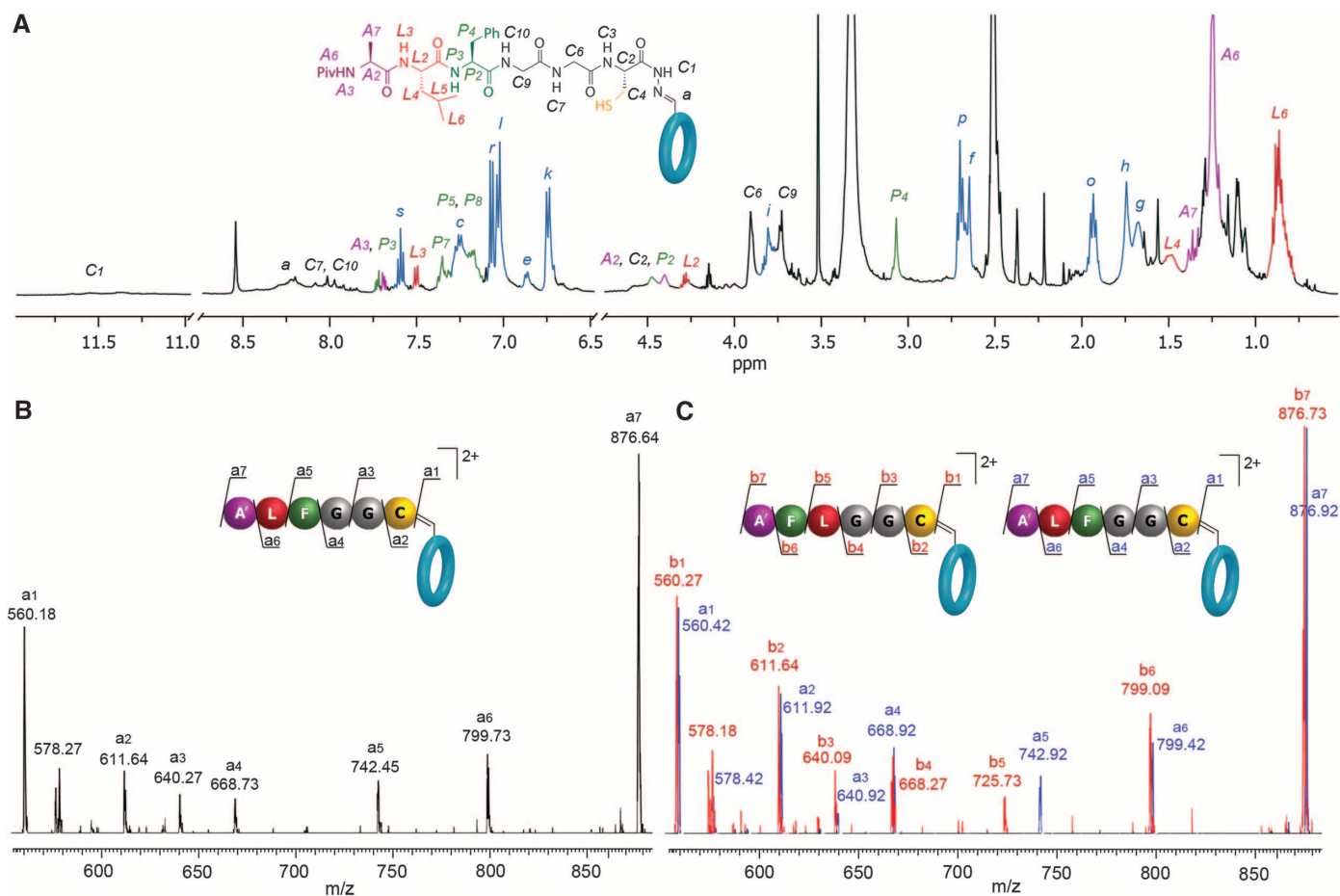
apparent in Fig. 4C (725.73 for the LeuGlyGlyCys-macrocycle, 742.92 for the PheGlyGlyCys-macrocycle), and product **7** was confirmed as corresponding to the intended sequence isomer.

With the use of HPLC-MS analysis of the reaction mixture from the operation of **1**, we did not detect any products corresponding to other peptide compositions (neither different sequences nor peptides with more or less than one Phe, Leu, or Ala residue), indicating that the peptide synthesis occurs overwhelmingly within the confines of the molecular machine. In contrast, a control reaction carried out under identical conditions but using the nonthreaded strand and macrocycle yielded several products, including strands with one or more amino acid groups cleaved, but there was no evidence for the formation of **7** under these conditions. Thus, the threaded architecture of the molecular machine—encompassing the catalytic site, elongation site, and the building block strand—is essential for the sequential peptide synthesis, and the mode of operation of the molecular machine is consistent with the mechanism shown in Fig. 3. The peptide (**9**), still bearing the GlyGlyCys unit at the C terminus, could

subsequently be cleaved from the macrocycle by hydrolysis (Fig. 3, iii).

On a scale of tens of milligrams, we performed the synthesis of the small peptide through autonomous multistep production by artificial small-molecule machine **1**, corresponding to parallel synthesis by ~10<sup>18</sup> machines. Once operation is initiated, the synthetic tasks performed by **1** proceed automatically, requiring no further intervention. As the catalytic thiolate is constrained by the threaded architecture of the machine from reacting with building blocks out of sequence, the act of balancing the rate of reactions with the speed that templates rearrange (19–22) is unnecessary for a rotaxane-based machine.

Rotaxane **1** is a (very) primitive analog of the ribosome. Limitations of the first-generation artificial system include slow kinetics (**1** takes ~12 hours to make each amide bond, compared to the 15 to 20 amide bonds synthesized per second by a ribosome) and loss of the sequence information on the strand as it is translated into the product. Furthermore, the size of oligopeptide that can be produced may ultimately be restricted by the size of the cyclic transition states involved



**Fig. 4.** (A) Proton NMR spectrum of molecular machine operation product **7** in *d*<sub>6</sub>-dimethylsulfoxide (500 MHz, 298 K). (B) Tandem mass spectrum of a single isotope of the 2+ ion [mass/charge ratio (*m/z*) = 876.64] of *S,N*-acetal-derivatized **7**. (C) Superimposed tandem mass spectra of 2+ ions of *S,N*-acetal-

derivatized macrocycles bearing the peptide sequences (Piv)AlaPheLeuGlyGlyCys (red; 2+ ion isotope-selected *m/z* = 876.73) and (Piv)AlaLeuPheGlyGlyCys (blue; 2+ ion isotope-selected *m/z* = 876.92), each prepared unambiguously by conventional peptide synthesis.

in *S*-to-*N* acyl transfer [although peptide ligation has been successfully used with up to 29-membered cyclic transition states (31)]. Nevertheless, **1** demonstrates that relatively small, highly modular, artificial molecular machines can be designed to autonomously perform iterative tasks in synthesis. The principles employed in the design and operation of **1** should be broadly applicable to other types of monomer and chemical reactions (32).

#### References and Notes

- J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry* (W. H. Freeman, New York, ed. 6, 2006).
- A. Yonath, *Angew. Chem. Int. Ed.* **49**, 4340 (2010).
- V. Ramakrishnan, *Angew. Chem. Int. Ed.* **49**, 4355 (2010).
- T. A. Steitz, *Angew. Chem. Int. Ed.* **49**, 4381 (2010).
- E. R. Kay, D. A. Leigh, F. Zerbetto, *Angew. Chem. Int. Ed.* **46**, 72 (2007).
- C. P. Collier *et al.*, *Science* **285**, 391 (1999).
- J. E. Green *et al.*, *Nature* **445**, 414 (2007).
- J. Berná *et al.*, *Nat. Mater.* **4**, 704 (2005).
- R. Eelkema *et al.*, *Nature* **440**, 163 (2006).
- T. Muraoka, K. Kinbara, T. Aida, *Nature* **440**, 512 (2006).
- B. K. Juluri *et al.*, *ACS Nano* **3**, 291 (2009).
- P. Thordarson, E. J. Bijnsterveld, A. E. Rowan, R. J. M. Nolte, *Nature* **424**, 915 (2003).
- C. Monnereau *et al.*, *J. Am. Chem. Soc.* **132**, 1529 (2010).
- F. Würthner, J. Rebek Jr., *Angew. Chem. Int. Ed. Engl.* **34**, 446 (1995).
- H. J. Yoon, J. Kuwabara, J.-H. Kim, C. A. Mirkin, *Science* **330**, 66 (2010).
- M. Schmittel, S. De, S. Pramanik, *Angew. Chem. Int. Ed.* **51**, 3832 (2012).
- V. Blanco, A. Carlone, K. D. Hänni, D. A. Leigh, B. Lewandowski, *Angew. Chem. Int. Ed.* **51**, 5166 (2012).
- J. Wang, B. L. Feringa, *Science* **331**, 1429 (2011).
- Y. He, D. R. Liu, *Nat. Nanotechnol.* **5**, 778 (2010).
- M. L. McKee *et al.*, *Angew. Chem. Int. Ed.* **49**, 7948 (2010).
- M. L. McKee *et al.*, *J. Am. Chem. Soc.* **134**, 1446 (2012).
- Y. He, D. R. Liu, *J. Am. Chem. Soc.* **133**, 9972 (2011).
- H. Gu, J. Chao, S.-J. Xiao, N. C. Seeman, *Nature* **465**, 202 (2010).
- S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **105**, 715 (2005).
- S*,*S*-acyl transfer is readily reversible, so employing thioesters (the most commonly used acyl source for native chemical ligation) to attach the building blocks to the strand would risk an amino acid being returned to the track after the macrocycle had passed by, potentially reducing the sequence integrity of the peptide synthesis.
- Materials and methods are available as supplementary materials on *Science* Online.
- V. Aucagne, K. D. Hänni, D. A. Leigh, P. J. Lusby, D. B. Walker, *J. Am. Chem. Soc.* **128**, 2186 (2006).
- J. D. Crowley, S. M. Goldup, A.-L. Lee, D. A. Leigh, R. T. McBurney, *Chem. Soc. Rev.* **38**, 1530 (2009).
- P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **266**, 776 (1994).
- F. K. Hansen *et al.*, *Org. Biomol. Chem.* **9**, 7162 (2011).
- R. J. Payne *et al.*, *J. Am. Chem. Soc.* **129**, 13527 (2007).
- N. Badi, J.-F. Lutz, *Chem. Soc. Rev.* **38**, 3383 (2009).

**Acknowledgments:** We thank V. Aucagne for useful suggestions during early versions of the molecular machine design. This research was funded by the Engineering and Physical Sciences Research Council (UK). We are grateful to the following organizations for postdoctoral fellowships: Fundacja na Rzecz Nauki Polskiej (to B.L.), Fonds de la Recherche Scientifique and Wallonie-Bruxelles International (to G.D.B.), the European Union 7th Framework Marie Curie Intra European Fellowship Program (to M.J.A.), Deutscher Akademischer Austausch Dienst (to P.M.E.G. and D.H.), and Deutsche Akademie der Naturforscher Leopoldina and Peter und Traudl Engelhorn-Stiftung (to D.M.D.).

#### Supplementary Materials

www.sciencemag.org/cgi/content/full/339/6116/189/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S37  
References (33–35)

5 September 2012; accepted 20 November 2012  
10.1126/science.1229753

## Shape-Memory Nanopores Induced in Coordination Frameworks by Crystal Downsizing

Yoko Sakata,<sup>1,2</sup> Shuhei Furukawa,<sup>1,2\*</sup> Mio Kondo,<sup>1,2</sup> Kenji Hirai,<sup>3</sup> Nao Horike,<sup>2</sup> Yohei Takashima,<sup>2</sup> Hiromitsu Uehara,<sup>2</sup> Nicolas Louvain,<sup>1,2</sup> Mikhail Meilikhov,<sup>2</sup> Takaaki Tsuruoka,<sup>2,5</sup> Seiji Isoda,<sup>1</sup> Wataru Kosaka,<sup>2</sup> Osami Sakata,<sup>4</sup> Susumu Kitagawa<sup>1,2,3\*</sup>

Flexible porous coordination polymers change their structure in response to molecular incorporation but recover their original configuration after the guest has been removed. We demonstrated that the crystal downsizing of twofold interpenetrated frameworks of  $[\text{Cu}_2(\text{dicarboxylate})_2(\text{amine})]_n$  regulates the structural flexibility and induces a shape-memory effect in the coordination frameworks. In addition to the two structures that contribute to the sorption process (that is, a nonporous closed phase and a guest-included open phase), we isolated an unusual, metastable open dried phase when downsizing the crystals to the mesoscale, and the closed phase was recovered by thermal treatment. Crystal downsizing suppressed the structural mobility and stabilized the open dried phase. The successful isolation of two interconvertible empty phases, the closed phase and the open dried phase, provided switchable sorption properties with or without gate-opening behavior.

Shape-memory materials alter their morphological appearance in response to an external stimulus (for example, mechanical stress created by macroscopic structural deformation), hold their new temporary shape after the stimulus has been removed, and return to their original morphology in the presence of another external stimulus (1, 2). For instance, a metal alloy can exhibit shape-memory effect if it has two phases that can interconvert reversibly; that is, without requiring atoms to diffuse through the structure. Here, we describe a molecular-scale shape-memory effect (MSME) in nanoporous

framework materials in which the application of an adsorption stress deforms the original shape of the nanopore into a temporary shape, which is maintained even after desorption, and thermal treatment then recovers the original shape.

Our design takes advantage of the flexibility of crystalline porous coordination polymers (PCPs) (3–8), which are assembled from organic spokes and inorganic joints. These flexible PCPs cooperatively reconfigure their framework structures in response to the incorporation of molecules into the nanopores; this adsorption process triggers the deformation of the pore shape. Most flexible

PCPs recover the original structure after the removal of the adsorption stress (that is, the desorption of the guest molecules), which leads to the so-called framework elasticity property (9). The MSME requires that any structural transformation during desorption should be suppressed. We show that crystal downsizing influences the structural mobility, because a reduction in the number of repeating units should be sufficient to regulate the cooperative nature of the structural transformation and the effect of stress.

We fabricated MSME nanopores by crystal downsizing, which regulated the flexibility of the framework, and demonstrated the switchable sorption events based on the presence of two interconvertible pore shapes (Fig. 1). Among the variety of flexible PCPs, we chose a PCP with a twofold interpenetrated framework (10–12)—namely,  $[\text{Cu}_2(\text{bdc})_2(\text{bpy})]_n$  (**1**, bdc = 1,4-benzenedicarboxylate, bpy = 4,4'-bipyridine) (13)—that exhibits a cooperative structural transformation from the

<sup>1</sup>World Premier International Research Initiative—Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan. <sup>2</sup>Exploratory Research for Advanced Technology (ERATO) Kitagawa Integrated Pores Project, Japan Science and Technology Agency (JST), Kyoto Research Park Building #3, Shimogyo-ku, Kyoto 600-8815, Japan. <sup>3</sup>Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan. <sup>4</sup>Synchrotron X-ray Station at SPring-8, National Institute for Materials Science (NIMS), Kouto, Sayo, Hyogo 679-5148, Japan. <sup>5</sup>Frontiers of Innovative Research in Science and Technology (FIRST), Konan University, 7-1-20, Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan.

\*To whom correspondence should be addressed. E-mail: shuhei.furukawa@icems.kyoto-u.ac.jp (S.F.); kitagawa@icems.kyoto-u.ac.jp (S.K.)