Author/s:
Saker Neto, N;Jones, DJ;Wong, WWH

Title:
Theoretical Aspects of Iterative Coupling for Linear Oligomers and Polymers

Date:
2020-03-01

Citation:
Saker Neto, N., Jones, D. J. & Wong, W. W. H. (2020). Theoretical Aspects of Iterative Coupling for Linear Oligomers and Polymers. Macromolecular Theory and Simulations, 29 (2), https://doi.org/10.1002/mats.201900048.

Persistent Link:
https://hdl.handle.net/11343/275280

Theoretical Aspects of Iterative Coupling for Linear Oligomers and Polymers

Nicolau Saker Neto, David J. Jones, Wallace W. H. Wong*

Nicolau Saker Neto, Dr. David J. Jones, Dr. Wallace W. H. Wong

Australian Centre for Advanced Photovoltaics, School of Chemistry, Bio21 Institute, The University of Melbourne, Parkville, Victoria, 3010, Australia.

E-mail: nsaker@student.unimelb.edu.au

Dr. Wallace W. H. Wong

ARC Centre of Excellence in Exciton Science, School of Chemistry, The University of Melbourne, Parkville, Victoria, 3010, Australia.

Keywords: iterative coupling, sequence-controlled polymers, stoichiometry, dispersity, mixture distributions

A conceptual study of iterative coupling (IC) is performed, providing a unified description and new research directions. IC chain growth rates and functional group choice are analyzed, guiding construction of efficient schemes. The concept of cycle efficiency is defined as a more complete metric of experimental implementations of IC, and then applied to the main linear and exponential IC processes. The mathematical relations between individual reactions, cycles, and the iterative process as a whole are studied. Finally, macromolecule IC is proposed as a strikingly complementary process to standard IC, with potential to reduce the dispersity of non-uniform samples. Due to its connection with the central limit theorem of statistics, it provides an unusually robust, powerful and general method for scalable production of polymer samples with narrow distribution. In all, this

contribution assists development of improved IC processes targeting low dispersity linear oligomers and polymers.

1. Introduction

The controlled synthesis of polymers is a subject of ever increasing importance, focusing on the discovery of general and reliable techniques for obtaining polymers with target sequences and molecular weight ranges with low dispersity. One of the techniques available for achieving these uniform (or monodisperse) materials is the stringent purification of low dispersity polymer samples from controlled polymerization processes, such as described by Lawrence et al.[1,2] This "top-down" approach is complemented by iterative coupling (IC), a "bottom-up" process where a planned sequence of reactions and purifications can result in a single product with precisely known structure.[3]

Though there are many published implementations of IC, these can largely be represented by the two schemes shown in **Figure 1.** With the adequate use of protected functional groups, it is possible to establish iterative *cycles*, whereby the chemical functionality of a chain is recovered after a sequence of reactions and allows a set of transformations to be repeated. Figure 1a represents the main linear IC cycle, while Figure 1b represents the main exponential IC cycle. Here, the descriptors linear and exponential refer to the growth rate of the chains as they undergo multiple cycles.

Linear IC is by far the most popular IC method. The simplicity of the chemistry used and the absolute control over sequence composition and length is well-suited to the synthesis of biomolecules, including oligopeptides[4] and oligonucleotides,[5] and has promise in the emerging field of sequence-controlled and sequence-defined polymers.[6-10] Advances in cross-coupling reactions, especially C-C

bond-forming reactions, may lead to linear IC becoming a fundamental part of synthetic chemistry in general.[11]

In contrast, exponential IC (also known as repeated molecular doubling,[12] divergent/convergent approach,[13] iterative binomial synthesis,[14] and iterative exponential growth[15]) is most suited to the fast synthesis of homopolymer chains, or other chains with simple periodicity (e.g. alternating or diblock copolymers).[9,16] Though exponential IC can, in principle, also furnish sequence-defined polymers of arbitrary complexity, very complex chains can require many additional synthetic steps to be accommodated, eventually making the process hardly distinguishable from a linear IC process. It is noteworthy, however, that linear IC is a subset of the possibilities allowed by the chemistry developed for an exponential IC process.

The work published in IC has so far been largely empirical. An excellent recent review by Solleder et al. covers a wide range of synthetic implementations of IC in general,[3] while additional reviews exist focusing on linear IC chemistry[17,18] or on exponential IC chemistry.[19,20]

In a landmark review,[21] the growing importance of IC in polymer synthesis has recently lead to the classification of "multistep-growth synthesis" as a third major approach for producing polymers, on par with chain-growth and step-growth polymerizations, and with distinct and complementary properties. However, the basic theoretical properties of IC have not been fully described in the literature. In this work, we attempt to mitigate this deficiency. We believe the framework provided will be beneficial to guide and interpret experimental results, provide a clearer comparison of the efficiency of different chemical processes, and allow the discovery of new implementations and applications of IC.

The fundamental question of IC is the following: *how can a collection of monomers be coupled into the target oligomer/polymer in as few steps as possible?* We shall start with a very general analysis

centered on linear polymers, and then introduce chemical considerations to infer the most promising processes. Throughout this paper, for simplicity we generally assume the use of pure reagents in each iterative cycle, and therefore assume the cycle products are purified stringently during iteration. The adequate purification of intermediates in an iterative process aiming for a uniform polymer is a challenging aspect in the field,[3] but we do not address the matter of impurities quantitatively here.

2. Iterative Growth Rates

To simplify matters as much as possible, any functional groups present are initially neglected, and linear homopolymer chains are considered solely as a concatenation of monomer repeating units. A concise diagram can then be drawn for proposed syntheses of a polymer chain (**Figure 2**).

As a concrete example, consider the multiple methods of reaching an octamer from a sample of monomers (**Figure 3**). The chain may be built by coupling one unit at a time, as shown in Figure 3a. The total number of coupling reactions performed in diagram is directly visualized as the number of points where lines meet. Therefore, the simplest approach would require a total of seven distinct coupling reactions to be performed in sequence (or in general, one less than the desired *i*-mer).

Many alternative syntheses may be proposed by changing the order of reactions. At a first glance the total number of coupling reactions to reach the target octamer in any reaction sequence remains the same (Figure 3b). However, it is important to note that not all coupling reactions need be distinct. In this case, instead of performing a same coupling reaction multiple times, it may be performed only

once and the product used wherever else it is required. Each time a particular material is reused in another branch, a coupling reaction may be omitted (Figure 3c).

Therefore, it is clear that designing a reaction scheme containing as few distinct coupling reactions is highly advantageous. In the extreme where this is exploited at every possible opportunity, the strikingly simple diagram in Figure 3d is obtained. The effect builds upon itself, and is the fundamental origin of the fast growth rate seen in exponential IC. The best-case scenario is achieved whenever the target chain is a 2^N-mer, the longest chain for which only N distinct coupling reactions are required.

More generally, it is not necessary that only two chain segments join during a coupling step; three, four or even more chains can theoretically be coupled together in a single reaction. These can described within the diagrams by having additional lines meet simultaneously (Figure 3e and 3f). Unsurprisingly, when more units are allowed to couple in a single step, even faster chain growth rates can be achieved. If k segments are allowed to join in a single step, then the longest chain achievable from a sample of monomer after N reactions contains k^N segments.

However, is should be noted that this theoretical faster growth rate comes at a very steep increase in the difficulty of practical implementation. Any scheme with k^N growth rate can be emulated by a 2^N growth process with $log_2 k$-times as many coupling reactions, while being substantially more accessible. Furthermore, increasing the base of an exponential curve (e.g. from 2^N to 3^N) is a comparatively small effect with respect to the change in paradigm when going from linear growth to any exponential growth. Combining these facts, there are diminishing returns when aiming for a process with greater than a 2^N growth rate.

In more complex schemes, the base of the exponential growth may not be an integer. For example, Kock, Smith and Heeney construct a "Fibonacci" IC process with a sequence of interleaving coupling

reactions between two segments,[22] which possesses an asymptotic growth rate of φ^N, where $\varphi \approx 1.618$ is the golden ratio.

Lastly, it should be noted that fundamentally faster growth rates are achievable under different assumptions. For example, if the target polymer need not be linear, then a double-exponential (2^{2^N}) growth scheme has been previously exemplified for the synthesis of dendrimers, and a discrete 255-mer was obtained after only three iterative cycles starting from a monomer.[23]

During implementation of IC, the fastest path achievable towards a particular target chain with *i* monomeric units depends sensitively on its composition and sequence. Fully linear IC and fully exponential IC represent the respective upper ($i - 1$) and lower bounds (practically, $log_2 i$) on the number of unique coupling steps required. For low-complexity target sequences, such as homopolymers or alternating co-polymers, the overall fastest process is dominated by exponential IC, and therefore can match or come close to the lower bound. Meanwhile, at the limit of completely random sequences with many distinct monomer units (e.g. proteins in general), it is difficult if not impossible to make use of exponential IC effectively, and even the fastest convergent synthesis will contain close to the maximum number of unique coupling reactions.

In summary, regardless of the exact characteristics of an IC scheme, the fastest chain growth rates are achieved when *all* the reactants in a particular coupling step are synthesized in the step immediately prior. This is the source of exponential chain growth in exponential IC. With this general result, we now turn towards a more detailed analysis, including the role of functional groups.

3. Functional Group Considerations

A partial theoretical survey of functional group choices suitable for linear IC h as been performed previously by Trinh, Laure and Lutz.[18] Here we provide a more global overview and study different possible outcomes.

To begin analyzing the impact of functional groups towards the feasibility of an IC scheme, we consider two chains (two components) with explicit functional groups joining together (**Figure 4**). Importantly, in every individual coupling event, one of the functionalities of each chain is retained, whereas the other reacting group is lost. While there is no restriction on the identity of the retained functionality, by definition the reacting functionality cannot be in a protected form (A^p or B^p).

In the case of two-component reactions, a plus symbol will be used to represent a coupling reaction between reagents with functionality specified on each side of the plus symbol. For example, Figure 4 displays a reaction mixture containing A^pB and AB chains, and can be written as an A^pB+AB reaction. This concise notation also suggests the intended type of reaction; A^pB+AB implies the use of BA cross-coupling reaction conditions (equivalent to AB cross-coupling). Note also that side-reactions are possible; it would be hard to stop the AB chains from undergoing self-condensation under BA cross-coupling conditions, irrespective of the presence of A^pB chains. Therefore an A^pB+AB reaction would likely undergo uncontrolled polymerization and create a mixture of products.

With the distinction between a reacting and a retained functionality, all combinations of functional groups for a linear chain can be listed. For completeness, chains where only one end partakes in chemical reactions shall also be considered. This can be understood as a lack of functionality on the other terminus. More generally, we consider a spectating functional group S which does not interact with the functional groups A or B, and therefore is necessarily retained. Therefore, there are five

total options for the retained functionality (*A*, *B*, A^p, B^p and *S*). To simplify matters, it is initially assumed the reacting functionalities can only be either *A* or *B.* This yields a total of 5×2 = 10 combinations of functional groups for a chain (namely *SA*, *AA*, *AB*, AA^p, AB^p, *SB*, *BB*, *BA*, BA^p and BB^p). Consequently, there are a total of 10×10 = 100 possible two-component reaction mixtures with these chains, which can react in *AA*- or *BB*-homocoupling, or *AB* cross-coupling conditions. The general outcome of all these reactions can be determined and the results are displayed in **Figure 5**. Here it is assumed that that both reagents have the same chain length. Clearly if reagents with different chain sizes are used, there is greater potential for multiple products to be obtained. Some examples of interpreting the table are shown.

The symmetry of Figure 5 arises from swapping the identity of *A* and *B*, and of reagent 1 and 2, which are arbitrarily defined. This symmetry allows the same information to be condensed into 30 entries at the expense of a slightly more cumbersome lookup procedure. The table is clearly divided into four quadrants of 5×5 = 25 cells. The top-left and bottom-right quadrants represent homocoupling reaction mixtures, while top-right and bottom-left quadrants represent cross-coupling reaction mixtures.

Certainly, uncontrolled polymerization and formation of multiple products represent the most undesirable outcomes. These are shown in the black, red, orange and yellow cells in Figure 5. Yellow cells represent situations which have multiple products with the same functionalities, but in a different state of protection; for example, the products in the reaction mixture are a combination of *AA*, A^pA and A^pA^p-functionalized chains, which could be converted to a single product by full protection or deprotection of the functionalities. In cells marked with an asterisk, multiple regioisomeric products can be formed, depending on the molecular structure of the chains linked. Altogether, these generally undesirable outcomes are most preponderant in the regions of the table

describing homocoupling reactions, and therefore it can be concluded that homocoupling reactions are in general poorly suited for IC schemes.

Meanwhile, cross-coupling reactions permit more interesting behaviour. Even in these regions, it is notable that most of the possible outcomes are still undesirable, but finally there appear situations with a high likelihood of forming a single product in a controlled fashion.

Orange cells can also potentially result in multiple products, unless a large molar excess of one of the reagents is used. In this case, either a significant amount of reagent is expended to favour the formation of a single product and the excess should be recovered to avoid a wasteful process, or the excess reagent is incorporated twice to form a functionally symmetric product. Interestingly, in the latter scenario, the orange cells have a unique property; they are the only ones where a *two*-component reaction can result in *three* chains combining deterministically to form a single product. This has been exploited in "bidirectional" linear IC,[3,7] an adaptation of the scheme in Figure 1a where the growing chain contains two symmetric functionalities. This variant of linear IC is largely similar in its properties to the main process, and will not be discussed in detail here.

In grey are shown reactions where a pair of functionally desymmetrized (*AB*, A^pB, *BA* or B^pA) or differentially protected reagents (*AA*^{p}, A^pA, *BB*^{p} or B^pB) result in a functionally symmetric product (*AA*, *BB*, A^pA^p or B^pB^p). This causes a considerable decrease in the scope of chemical diversification possible unless a new functional desymmetrization is performed, which is in general difficult. In green are highlighted reactions which are natively compatible with linear IC. Indeed, the reaction pair *SA*+*BA*^{p} (or equivalently *SB*+*AB*^{p}) describes the standard linear IC scheme shown in Figure 1a. The *SA*+*AB*^{p} and *SB*+*BA*^{p} combination could be used in a form of "alternating" linear IC, where the active functionality alternates between *A* and *B* in each iterative cycle.

Out of all the previously discussed combinations of functional groups and coupling reactions, only the blue cells formed by the combinations A^pB+AB^p or A^pA+BB^p are generally suitable for exponential IC without additional considerations. In fact, attempting iteration of the latter pair reduces to the former after one cycle, after which iteration can be performed with no overall change in the functionalities used. In this sense, use of the A^pB+AB^p combination furnishes the simplest generally applicable implementation of exponential IC. We reiterate that linear IC is also easily performed in the regions which allow exponential IC.

This theoretical analysis can be extended by also considering the cases where additional unique functional groups *C* and *D* are involved. These functional groups may be related to *A* and *B* in multiple ways. For example, functional group *C* may be distinct from *A* while also reacting with *B*, or functional group *D* could be a reactive functional group while also being convertible to A^p. Given the number of possible relations between *A*, *B*, *C* and *D*, the complexity of the analysis quickly increases, with many additional cases to consider. No particular correlation between all the functional groups will be determined, and we seek only to provide a general overview.

Once again, the number of possible combinations of functional groups in a chain can be counted. Without loss of generality, there exist five total options for the retained functionality, as before, but now there are four reacting functionalities available (*A*, *B*, *C* and *D*). It is unnecessary to consider *C*, C^p, *D* and D^p among the retained functionalities, as there would be a suitable swap with the identity of *A* or *B* such that they would fall in the above cases. For a similar reason, the combinations *SC* and *SD* are superfluous. The addition of a fifth functional group *E* is not required when analyzing a single two-component reaction of terminally-functionalized linear chains, as a maximum of four unique groups may be simultaneously present.

Therefore, a new total of 18 combinations of functional groups can be written, and these could be paired up in 18×18 = 324 two-component reaction mixtures. Evaluation of all the reaction outcomes results in **Figure 6**, of which Figure 5 is a subset, visible on the top-left. Notably, even with the use of more functional groups, homocoupling reactions (shown in the diagonal blocks) still result in largely unfavourable outcomes. Meanwhile, a far greater variety of combinations using cross-coupling can potentially be used for exponential IC. However, the true suitability of these combinations depends on further details of how the functional groups *C* and *D* relate to *A* and *B*. A new type of outcome, where differentially-protected products are generated, can also become suitable if conversion between *A* and *B* is possible. As examples of prior work in the expanded region of Figure 6, the submonomer method for synthesis of oligopeptoids by Zuckermann et al., [24] and Lutz and co-workers' synthesis of oligo(triazole amide)s[25] rely on linear IC using an alternation of *SA*+*CB* and *SB*+*DA* reactions to achieve sequence-defined materials. Lutz and co-workers have also explored linear IC based on iterated *AC*+*DB* reactions, including the use of a support.[26,27]

It is worth briefly contemplating reaction mixtures where three or more components all combine in a single step. The calculation process here described can be generalized to these situations. However, attempting to define a three-component analogue of Figure 6 leads to a three-dimensional table containing 26×36×26 = 24336 cells. This could be reduced dramatically by exploiting all available symmetries, but the unique cases remaining would still number in the thousands. Furthermore, effective exploitation of this space requires use of multiple unique functional groups, making chemical implementation difficult; when only a few are used, most reaction mixtures result in uncontrolled polymerization or multiple products. Nevertheless, three-component and even four-component reactions have been investigated in IC processes resulting in highly functionalized uniform oligomers.[28-30]

It is important to stress that the information provided in Figure 5 and 6 is only meant as a guide for finding effective IC processes towards the controlled synthesis of linear polymers. In reality, additional considerations or real-world behaviour (e.g. reaction kinetics) may result in different outcomes. As an extreme example, the work of Izumi et al.[31] represents a remarkable method for approximately constructing an exponential IC scheme and obtaining discrete polythiophenes by iterating an *AA+AA* reaction, which formally should result in uncontrolled polymerization. Another example is given by Hartmann et al.,[32] who use a sequence of alternating *SA+BB* and *SB+AA* reactions on a solid support with excess functionally symmetric reagents to synthesize poly(amidoamine)s via a linear IC protocol without protecting group chemistry.

Lastly, due to the importance of solid and soluble supports in IC, we explicitly consider their integration into the functional group analysis. Take, for example, a linear IC process characterised by iteration of the $SA+BA^P$ coupling reaction. The generic nature of the spectating functionality *S* and its non-participation in the overall process allows it to be directly replaced by any compatible support, in principle. Since the functional group *S* remains unchanged throughout the IC process, the substrate stays attached to the support. This is, of course, abundantly exemplified by solid-phase peptide synthesis. Additionally, if a functional group of a chain is inactive when attached to a support, but is activated (directly or subsequently) upon cleavage of the support from the substrate, then it is entirely possible to use the support-bound functionality as a protected functional group A^p or B^p, which can be activated to an unbound functionality *A* or *B*. It is even possible to consider processes where the binding to the support *causes* activation of a protected functionality, but it less clear whether this mode of action can be effectively used for development of IC processes. In this way, the entirety of Figure 6 is also applicable to support-based IC.

Having performed a theoretical survey of the functional group landscape, further focus is given to the main linear and exponential IC schemes of interest, as shown in Figure 1.

4. Stoichiometric Analysis of the Main Iterative Coupling Schemes

As with any chemical procedure, it is essential to compare different IC scheme implementations with regards to their synthetic efficiency. Often this can be adequately determined by looking at the numerical product of reaction yields based on the limiting reagent, which is generally the main substrate of the reaction sequence. However, it shall be made evident that this is not sufficient to accurately describe an IC process, which makes comparison of published works less straightforward.

We perform a more in-depth study of stoichiometry in iterative processes. The main goal is to establish the concept of a *cycle efficiency* (CE), which accurately keeps track of how many repeating units present in reagents during the course of an iterative cycle can be found in the product of the cycle. Thus, the cycle efficiency for any N^{th} cycle (CE_N) is given by the ratio of the overall cumulative yield at the end of the N^{th} cycle (Y_N^{cumul}) and the end of the prior cycle (Y_{N-1}^{cumul}), as per Equation (1).

$$CE_N = \frac{Y_N^{cumul}}{Y_{N-1}^{cumul}} \quad ; \quad CE_1 = Y_1^{cumul} \tag{1}$$

This definition is very general, and is applicable for any IC process. The challenge then becomes to establish the cumulative yields at any point based on the stoichiometric details of the individual reactions. We propose that calculation of CE_N for each cycle furnishes more direct and complete measures of efficiency in an IC process, and should be the standard by which published works are compared.

4.1. Linear Iterative Coupling

We begin by evaluating the main linear IC scheme, which makes use of the two-component *SA+BA*^{p} reaction. For simplicity, we neglect the spectating functional group *S*. The important quantities and terminology used for this cycle are detailed in **Figure 7**. It is always assumed that the substrate is the limiting reagent ($R \geq 1$).

In all yield calculations in this article, it is assumed no excess reagent can be recovered. For example, if ten molar equivalents of linker are mixed with substrate during the coupling reaction and no recovery of this material is performed, then $R = 10$. If ten molar equivalents are mixed, but four equivalents are recovered after purification, then for the purposes of the equations here defined, it is equal to only having used six equivalents ($R = 6$). If a deprotection reaction results in the activation of 80% of molecules in a sample, and 15% of the starting material is recovered unchanged, then the deprotection yield applicable to the formulas is 0.8/(1-0.15)=94%. It is worth noting that if recovery of reagents is very facile, especially excess reagent from the coupling reactions, powerful IC procedures with high cumulative yields can be constructed.

Stipulating an initial molar amount n_{Sub} of *k*-mer substrate allows calculation of the molar amount of repeating units at the start, kn_{Sub}. By keeping track of the number of repeating units using the deprotection and coupling yields in each cycle, the amount of substrate can be easily calculated at any point through the process. The amount of linker consumed is then determined based on the amount of limiting reagent (substrate) of a coupling reaction. Altogether, this results in the information displayed in **Table S1**.

One may be tempted to focus solely on the substrate. It can easily be determined that the yield with respect to the substrate $Y_{Sub,N}$ for any single N^{th} cycle is simply the product of the deprotection yield

Y_{D_N} and coupling yield Y_{C_N} for that cycle, as per Equation (2). The cumulative substrate-based yield $Y_{Sub,N}^{cumul}$ up to the end of the N^{th} cycle is simply given by the product of all the deprotection yields and coupling yields, given by Equation (3).

$$Y_{Sub,N} = Y_{D_N} Y_{C_N} \quad (2)$$

$$Y_{Sub,N}^{cumul} = \prod_{i=1}^{N} Y_{D_i} Y_{C_i} = \prod_{i=1}^{N} Y_{Sub,i} \quad (3)$$

This definition is simple, but entirely inadequate on its own, and cannot be used as a reliable metric of the overall efficiency of a linear IC process. As an extreme example of how its use is misleading, consider a linear iterative cycle where the deprotection step is performed in 100% yield. The activated substrate is then coupled with the linker. However, to ensure every chain grows by one unit during the coupling step, ten times the stoichiometric quantity of linker is added and the excess is not recovered ($R = 10$). Following standard practice, the chemical yield for the coupling step is reported based on the limiting reagent, and is therefore equal to 100%.

Based on the yield metrics above, this process would be considered perfectly efficient, when in truth it is obvious that the process is in fact extremely wasteful. This discrepancy between the calculated values and the actual outcome indicate the poor suitability of this metric in describing the efficiency of the cycle. Even if a large amount of the excess is recovered, any losses will accumulate.

It is therefore imperative to include information about the usage of linker, which introduces the coupling ratio R to the equations. Analyzing the consumption of the linker alone, it quickly becomes

apparent that calculations based on the linker are very different from those of the substrate. This is because from the perspective of the substrate, it undergoes a straight sequence of reactions. However, from the perspective of the linker, it undergoes a multitude of reactions in parallel, and thus the individual yields cannot be multiplied throughout. The linker yield in each cycle $Y_{Link,N}$ and the cumulative yield based on the linker $Y_{Link,N}^{cumul}$ are given by Equation (4) and (5), where we make use of the definition of an "empty product" $\prod_{j=2}^{1} Y_{D_j} Y_{C_{j-1}} = \prod_{j=p}^{q<p} f(j) = 1$; a product of no terms is equal to 1, the multiplicative identity.

$$Y_{Link,N} = \frac{R}{Y_{C_N}} \tag{4}$$

$$Y_{Link,N}^{cumul} = \frac{N Y_{C_1} \prod_{i=2}^{N} Y_{D_i} Y_{C_i}}{\sum_{i=1}^{N} \left(R_i \prod_{j=2}^{i} Y_{D_j} Y_{C_{j-1}} \right)} \neq \prod_{i=1}^{N} Y_{Link,i} \tag{5}$$

Finally, both the substrate and linker are taken into account simultaneously to obtain *combined* yields Y_N for any N^{th} cycle and cumulative combined yields Y_N^{cumul} from the first to the N^{th} cycle, shown in Equation (6) and (7), respectively. The form of the equations are largely similar to those considering only the linker, and as such, once again the individual yields cannot be multiplied to generate the cumulative yield. However, at last, Equation (7) truly characterizes the entirety of the iterative coupling process accurately.

$$Y_N = \frac{(k+Nl) Y_{D_N} Y_{C_N}}{l Y_{D_N} R_N + (N-1) l + k} \tag{6}$$

$$Y_N^{cumul} = \frac{(k+Nl)\prod_{i=1}^{N} Y_{D_i} Y_{C_i}}{lY_{D_1}\left[\sum_{i=1}^{N}\left(R_i \prod_{j=2}^{i} Y_{D_j} Y_{C_{j-1}}\right)\right]+k} \neq \prod_{i=1}^{N} Y_i \quad (7)$$

Using Equation (7), we can now explicitly write a formula for CE_N in linear IC processes, given by Equation (8).

$$CE_N = \frac{k+Nl}{k+(N-1)l} Y_{D_N} Y_{C_N} \frac{\theta}{\theta + lY_{D_1} R_N \prod_{j=2}^{N} Y_{D_j} Y_{C_{j-1}}} \quad (8)$$

$$\theta = lY_{D_1}\left[\sum_{i=1}^{N-1}\left(R_i \prod_{j=2}^{i} Y_{D_j} Y_{C_{j-1}}\right)\right] + k \quad (9)$$

Unfortunately, mathematical simplification of Equation (8) is limited, and it appears daunting upon expansion. However, its evaluation is unproblematic when performed stepwise from the first cycle, as is necessary in any case to fully characterize the iterative process.

Interestingly, while Equation (6) accurately describes the overall yield of an individual cycle in linear IC, it is not equal to the cycle efficiency as defined in Equation (1). This is ultimately because linear IC consists of "open" cycles, where new material is introduced in each cycle (the linker). In every cycle, the fraction of linker consumed, relative to the initially available pool, is therefore dependent not only on the cycle itself, but all those before it. This leads to the complexity of Equation (8).

It is interesting to consider what happens to the cycle efficiency at very large cycle numbers. By taking the limit and assuming that the coupling ratios do not grow without bound, Equation (10) is obtained.

$$\lim_{N \to +\infty} CE_N = Y_{D_N} Y_{C_N} = Y_{Sub.N} \quad (10)$$

Remarkably, this is the exact formula for yield based on the substrate shown in Equation (3), which indicates that the amount of linker consumed becomes less important at high cycle numbers. One way to understand this is to realize that in this limit, the extended substrate will have a molecular weight far larger than the linker, and therefore even a large molar excess of linker in the coupling reaction represents a small amount of matter. In this view, most of the consumption of the linker throughout the whole process happens in the initial stages of the iteration. Another interpretation is that at very high cycle numbers, the substrate carries a far greater number of repeating units than the linker. Therefore, the loss of some substrate represents a much greater waste of repeating units than the loss of the same molar amount of linker. It is thus important to maximize the coupling yield, even if at the expense of using a large ratio of linker to substrate.

4.2. Exponential Iterative Coupling

Having fulfilled the important goal of determining the cycle efficiency in linear IC, we now move on to the study of exponential IC. Similarly to before, the exponential iterative cycle is redrawn and the parameters involved are pointed out (**Figure 8**). Now it becomes important to differentiate the two kinds of deprotection reaction. Also, in this exponential IC cycle, the distinction between substrate and linker disappears. Therefore, the terminology is not carried over, and instead the reagents in the coupling reaction are differentiated by their active functionality.

Once again, the stoichiometric analysis begins by keeping track of the number of repeating units present in each species in the cycle, and the amount consumed in reagents. Immediately, a profound difference between the main linear and exponential IC cycles becomes apparent. In exponential IC, all of the outputs of a cycle are used as inputs in the next cycle, and there is no external source of repeating unit. In this sense, exponential IC cycles can be viewed as "closed" cycles. This has the effect of dramatically simplifying calculations, as exponential IC cycles can be analyzed completely independently from each other.

The exponential IC cycle is studied starting from a *k*-mer and resulting in a 2*k*-mer. Here, it is easiest to start from the molar amount of the product of the cycle n_{Prod}, and work backwards to calculate the amount of reagents consumed. There are in fact two cases to consider, one valid for $R_{A/B} \geq 1$ and another for $R_{A/B} \leq 1$, as this changes the limiting reagent in the coupling reaction. The molar amount of each reagent used and the portion incorporated into the product of a cycle are shown in **Table S2**.

With the molar quantities calculated, it is possible to determine the combined yields Y_N via Equation (11) and (12), which take into account all sources of repeating units and are valid for any N^{th} cycle. If $R_{A/B} = 1$, both equations are equivalent. Here, we assume without loss of generality that $Y_{DA} \geq Y_{DB}$. Whenever this statement is not true, then all that is required is for the identities of *A* and *B* to be swapped, along with the fact that $R_{A/B} = (R_{B/A})^{-1}$.

$$Y_N = \frac{2Y_{C_N}}{\frac{R_{A/B_N}}{Y_{DA_N}} + \frac{1}{Y_{DB_N}}} \qquad (R_{A/B} \geq 1) \tag{11}$$

$$Y_N = \frac{2Y_{C_N}}{\frac{1}{Y_{DA_N}} + \frac{1}{R_{A/B_N} Y_{DB_N}}} \qquad (R_{A/B} \leq 1) \tag{12}$$

We then consider performing multiple exponential IC cycles in sequence. For simplicity $Y_{DA} \geq Y_{DB}$ and $R_{A/B} \geq 1$ is assumed to hold for all cycles, allowing Equation (13) to be obtained. Pleasingly, this product can be described as a product of the individual Y_N. This result is general and holds even if the limiting reagent changes in a particular exponential IC sequence of reactions.

$$Y_N^{cumul} = \frac{2^N}{\prod_{i=1}^{N} \frac{1}{Y_{C_i}} \left(\frac{R_{A/B_i}}{Y_{DA_i}} + \frac{1}{Y_{DB_i}} \right)} = \prod_{i=1}^{N} Y_i \tag{13}$$

The N^{th} cycle efficiency CE_N is once again defined according to Equation (1). Substitution of Equation (13) results in the cancellation of all terms but the last, such that the cycle efficiency for the N^{th} cycle may be defined by the very simple expressions shown in Equation (14) and (15). It is important to note that, as with the equations for linear IC, the calculations implicitly assume that any spent material which was not incorporated into the cycle product is unrecoverable.

$$CE_N = \frac{2Y_{C_N}}{\frac{R_{A/B_N}}{Y_{DA_N}} + \frac{1}{Y_{DB_N}}} \qquad (R_{A/B} \geq 1) \tag{14}$$

$$CE_N = \frac{2Y_{C_N}}{\frac{1}{Y_{DA_N}} + \frac{1}{R_{A/B_N} Y_{DB_N}}} \qquad (R_{A/B} \leq 1) \tag{15}$$

Compared to linear IC, the simplicity of the equations of exponential IC is inviting to analysis. First of all, it can be seen that the cycle efficiency is more sensitive to changes in coupling yield than to changes in individual deprotection yields. This is easily understood, as each individual deprotection reaction encompasses only a fraction of the material undergoing exponential IC, while the coupling reaction encompasses the entirety. Choosing effective cross-coupling chemistry for exponential IC is therefore particularly important to achieve high cycle efficiencies, and should be the priority when proposing a new process.

Graphs varying the parameters involved in Equation (14) and (15) are shown in **Figure 9**. It is imperative to note that even with quantitative reaction yields, a coupling ratio different from unity inherently limits the maximum achievable cycle efficiency, even with otherwise ideal reactions. Contrary to linear IC, variations of the coupling ratio produce effects of equal magnitude regardless of the cycle number.

Straying from the ideal ratio will likely cause a net loss of product in a cycle even if the coupling reaction itself has an improved yield, though some other advantages may exist, such as facilitated purification. Furthermore, as noted previously, if one of the coupling reagents is easily recoverable after the reaction, it can be highly beneficial if used in excess, as it would increase the coupling yield while keeping the effective coupling ratio used close to unity.

Lastly, there is a more subtle effect when varying the coupling ratio. When the deprotection reactions have equal yield, the cycle efficiency will ideally be the same whether a particular ratio or its inverse is used. Plotting this behaviour generates the graphs in Figure 9c, which are notably symmetric with respect to the vertical line $R_{A/B} = 1$. However, when the deprotection reactions have different yields, it is slightly more advantageous to perform a coupling reaction with excess

reagent generated by the higher yielding deprotection reaction. This leads to the graphs in Figure 9d, where the previous symmetry is now gone.

Overall, the calculations and graphs show how the importance of the coupling ratio is on par with the deprotection and coupling yields, and must be reported with similar visibility. This is rarely the case in currently published literature investigating exponential IC schemes.

4.3. Cycle Iteration

For any process with a cycle efficiency determined as per Equation (1), including the main linear and exponential IC processes, the effect of cycle iteration is shown in **Figure 10**. The ratio between two curves of different cycle efficiency grows exponentially with the number of cycles. This clearly illustrates one of the most fundamental requirements of any successful iterative process: the stringent optimization of a cycle to maximize its yield and allow high cycle numbers to be achieved in acceptable cumulative yield. The benefit of optimization itself grows exponentially further into the iterative process, and due to the repeated nature of the deprotection and coupling reactions, optimizing any cycle can contribute to the optimization of all cycles.

5. Macromolecule Iterative Coupling

IC provides a comparatively simple and general route towards the synthesis of discrete linear oligomers and polymers. However, multiple cycles are required to reach relatively large chains, even

with exponential IC. Thus, the synthesis of large amounts of narrow, high molecular weight materials for applications other than fundamental studies appears problematic.

In principle, one method to reach high molecular weights faster with an IC process is to start with an already large molecule. For example, if a monomer could be converted in a single step to a 16-mer, then this single step would be equivalent to four exponential IC cycles, and as many as 15 linear IC cycles. However, a general procedure capable of synthesizing large discrete oligomers and polymers in a single step does not currently exist. There are methods to generate polymers with somewhat controlled chain lengths and narrow weight distributions, such as living chain-growth polymerization reactions,[33] but control typically becomes more difficult when targeting higher molecular weights. Therefore, IC could use relatively narrow macromolecule dispersions with lower molecular weights as input and increment the weight by performing iterative cycles while maintaining a relatively low dispersity.

This raises a new question: *what is the effect of sample dispersity in an iterative coupling process?* We must focus on the coupling reactions, where significant changes to dispersity may occur. Here, the situation is much more complicated than in the uniform case for either linear or exponential IC. Instead of a single kind of coupling reaction, there are potentially thousands, between *i*-mers of all sizes.

Viewed statistically, a cross-coupling reaction between two samples in linear or exponential IC can be thought of as a sum of two distributions, one for each sample containing the active functional group *A* or *B*. We therefore track the number-average degree of polymerization μ and number-average chain length variance σ^2 of the distributions prior to and after the coupling reaction. These quantities are related to the dispersity Đ of a distribution as shown by Equation (16).[34]

$$Đ = 1 + \frac{\sigma^2}{\mu^2} \quad (16)$$

The sum of distributions is a well-understood concept in statistics. It is important to note that in general the sum of two distributions varies depending on their correlation, which modifies the shape of the resulting distribution. However, according to Flory's *principle of equal reactivity*,[35] to a good approximation the coupling kinetics of chains in a reaction mixture is independent of the size of the *i*-mers involved. In other words, a coupling reaction can be visualized as *randomly* selecting a member of each distribution and adding them together. This simplifies the description of the resulting distribution, whose parameters can be calculated as shown in Equation (17) and (18).

$$\mu_{Sum} = \mu_A + \mu_B \quad (17)$$

$$\sigma_{Sum}{}^2 = \sigma_A{}^2 + \sigma_B{}^2 \quad (18)$$

We begin by considering the main linear IC process using a sample of macromolecules, shown in Figure 1a. The macromolecules can be used as the substrate, the linker, or both. The latter case is of most interest, and shall be described here. For simplicity, we assume the linker and substrate macromolecule samples are both described by the same initial distribution parameters μ_0, $\sigma_0{}^2$ and $Đ_0$. An ideal coupling reaction is then performed, where the only substances present are the linker and activated substrate in equimolar ratio, and the reaction is quantitative. In this case, the resulting

distribution has the parameters $\mu_1 = 2\mu_0$ and ${\sigma_1}^2 = 2{\sigma_0}^2$, according to Equation (17) and (18). Substituting this information into Equation (16), we find the following remarkable result.

$$Đ_1 = \frac{Đ_0+1}{2} \tag{19}$$

After an ideal linear IC cycle using a sample of macromolecule as both the substrate and linker, the dispersity of the product is greatly reduced. In fact, the higher the initial value of the dispersity is, the stronger the reduction. Moreover, this result is completely independent of the shape of the distribution.

We then consider performing consecutive ideal linear IC cycles with the sample. In every N^{th} cycle, the substrate is initially determined by a new set of distribution parameters $\mu_{N-1} = (N-1)\mu_0$, ${\sigma_{N-1}}^2 = (N-1){\sigma_0}^2$ and $Đ_{N-1}$, while the linker remains described by μ_0, ${\sigma_0}^2$ and $Đ_0$. The following formulas for the dispersity of the product at the end of the N^{th} cycle are obtained.

$$Đ_N = \frac{1+NĐ_{N-1}}{N+1} = \frac{N+Đ_0}{N+1} \tag{20}$$

While initially there is a strong reduction in dispersity, at later cycles it decreases much more slowly. This is because the extended substrate, which now has a much narrower distribution, is continuously reacted with the same comparatively broad linker. Even so at the limit of infinite cycles, the dispersity slowly approaches unity, no matter the characteristics of the initial distribution.

We now consider the main exponential IC process, shown in Figure 1b. The initial macromolecule sample is once again described by the parameters μ_0, ${\sigma_0}^2$ and $Đ_0$, and it is not necessary to consider the distinction between substrate and linker. As previously, the cycles are performed ideally. The reaction mixture in the coupling step contains only singly-activated chains, in equimolar ratio. Interestingly, after the first exponential IC cycle, the dispersity of the sample is also described by Equation (19). The difference appears in the second cycle, as now all the reagents come from distributions characterized by μ_1, ${\sigma_1}^2$ and $Đ_1$. After N cycles, the dispersity obeys the following relations.

$$Đ_N = \frac{1+Đ_{N-1}}{2} = \frac{2^N-1+Đ_0}{2^N} \quad (21)$$

While the first cycle in the linear and exponential IC processes considered is similar, over a longer process the latter has the potential for an outstanding decrease in dispersity, exponentially converging to unity. Once again, this result is independent of the nature of the initial distribution.

A comparison between the main a linear and exponential IC processes is shown in **Figure 11**, starting from a sample with very high dispersity. We stress that the reduction in dispersity originates solely from joining chains at random, with no size fractionation employed at any point.

To further understand how the dispersity decreases in an IC process, **Figure 12** displays multiple simulated exponential IC cycles on various initial distributions containing 1 million elements each, along with their dispersities. There is a clear tendency for clustering around the mean of the distributions. This is because, when randomly summed with a copy of itself, the low and high extremes of any distribution are most likely to add to elements larger or smaller than themselves,

respectively. The behaviour displayed in Figure 11 and 12 is in fact a manifestation of the central limit theorem of statistics.

Though enlightening due to their simplicity, the ideal cases described above, with quantitative coupling reactions, are clearly not representative of a real process. It is entirely possible that the reduction in dispersity depends very sensitively on achieving a quantitative reaction, and that even small reductions in coupling yield would present a much less appealing result.

Therefore, we consider the important generalization where the coupling yield Y_C is allowed to vary between 0% and 100%. This can be viewed as a "partial" sum of two distributions. The result is a weighted combination of the initial distributions and the product distribution from an ideal reaction. The statistics of mixture distributions is used to describe this process. [36]

We analyze only the first iterative cycle of a process with a non-ideal coupling reaction, though the ideal equimolar ratio of coupling reagents is maintained. Under these conditions, the previously described linear and exponential IC processes using macromolecule inputs are equivalent. For the sum of two distributions with normalized mixing coefficients $p_0 + p_1 = 1$, the statistical parameters of the mixture distribution are given by Equation (22) and (23).

$$\mu_{Mix} = p_0\mu_0 + p_1\mu_1 \tag{22}$$

$$\sigma_{Mix}{}^2 = p_0{\sigma_0}^2 + p_1{\sigma_1}^2 + p_0p_1(\mu_0 - \mu_1)^2 \tag{23}$$

The relationship between the mixing coefficients and the coupling yield Y_C must be found. These are determined by Equation (24) and (25). The factor of two in the numerator of p_0 arises because two reacting chains are consumed to create one product chain. This also has the effect of altering the overall number of chains present, meaning the normalization factor of the mixing coefficients depends on Y_C, generating the denominators in both equations.

$$p_0 = \frac{2(1-Y_C)}{2-Y_C} \tag{24}$$

$$p_1 = \frac{Y_C}{2-Y_C} \tag{25}$$

The mean and variance of the mixed distribution can then be obtained as a function of Y_C. Performing the appropriate substitutions, Equation (26) and (27) are obtained.

$$\mu_{Mix} = \frac{2}{2-Y_C}\mu_0 \tag{26}$$

$$\sigma_{Mix}{}^2 = \frac{2}{2-Y_C}\sigma_0{}^2 + \frac{2Y_C(1-Y_C)}{(2-Y_C)^2}\mu_0{}^2 \tag{27}$$

Equation 15 is now applied to the mixed distribution, to obtain the expression for $Đ_{Mix}$ shown in Equation (28).

$$Đ_{Mix} = \left(1 - \frac{Y_C}{2}\right)(Đ_0 + Y_C) \quad (28)$$

Again, we are presented with a remarkably simple formula. As expected, if $Y_C = 1$, Equation (19) is recovered. It is possible to directly plot Equation (28) and analyse the dependencies between the variables. However, it is more illuminating to study the relative change in dispersity after the coupling reaction compared to the starting materials. For this, both sides of Equation (28) are divided by the initial dispersity to obtain Equation (29), and then plotted to generate **Figure 13.**

$$\frac{Đ_{Mix}}{Đ_0} = \left(1 - \frac{Y_C}{2}\right)\left(1 + \frac{Y_C}{Đ_0}\right) \quad (29)$$

We obtain a number of surprising results. First of all, if the initial dispersity is sufficiently high ($Đ_0 \geq 2$), then no matter how inefficient the coupling reaction is, the product will have a lower dispersity. Interestingly, if $Đ_0 < 2$ the dispersity of the product may increase if the coupling yield is sufficiently low. The closer the starting distribution is to a uniform sample, the higher the coupling yield must be to avoid the increase in dispersity. If the starting materials are indeed uniform, only an ideal coupling reaction can avoid an increase in dispersity, as intuitively expected.

The critical line $Đ_0 + Y_C = 2$ separates the regions in the plot where the dispersity remains constant after the coupling reaction (as well as the trivial solution $Y_C = 0$). Above this line lies the vast region where the dispersity decreases after coupling, and as expected the reduction is greatest when the coupling yield is as high as possible and when the starting distribution is very broad.

However, below the line is a triangular region where the dispersity increases. Remarkably, not only is this region comparatively small, but the ratio of final to initial dispersities never rises much above 1. The global maximum of Equation (29) within the space of allowed parameters ($Đ_1 \geq 1$ and $0 \leq Y_C \leq 100\%$) returns the value 1.125 for $Đ_0 = 1$ and $Y_C = 50\%$. In relative terms, this is the worst-case scenario for a sample undergoing the first IC cycle. These findings suggest that the mechanism by which an IC cycle reduces dispersity is in fact unexpectedly robust with respect to variations in the coupling yield.

It is also important to mention how the coupling reaction affects the number-average degree of polymerization of the sample, which is described by Equation (26). Here again it is useful to study the relative change, dividing both sides by the initial value to obtain Equation (30). The plot is shown in **Figure 14.**

$$\frac{\mu_{Mix}}{\mu_0} = \frac{2}{2-Y_C} \tag{30}$$

For the description of further cycles, would be necessary to distinguish between linear and exponential IC. However, more crucially, the nature of the functionalities present must be analyzed. The deduction of Equation (20) and (21) is simplified because iteration of ideal cycles means the coupling reaction in later cycles contains only chains with a single active functional group, as is the case in the first cycle.

In contrast, the product of a non-ideal coupling reaction will contain residual chains with an active functionality. If these residual active chains are selectively removed by purification, then the product distribution becomes exactly the same as in an ideal IC cycle, with the loss of material being the only

side-effect. In practice, this may be difficult to implement efficiently. Another possibility is to perform *deactivation* reactions after the coupling reaction, to ensure that all functional groups in a sample are in their protected form prior to the start of the next cycle.

If the residual active functionalities are not removed after a non-ideal coupling reaction, then during the deprotection steps of the subsequent cycle, there is necessarily the formation of doubly-activated chains. These can trigger partial uncontrolled polymerization, most likely reducing the benefits of macromolecule IC. Mathematical treatment of this scenario will not be sought here, but we note that this form of contamination may be tolerable at low levels.

One may also ask about the effect of straying from the ideal equimolar ratio of reagents in the coupling reaction. Unless the difference is only slight, this will almost certainly lead to a poorer result, as the benefit of increased coupling yield is outstripped by the contamination with unreacted chains.

In a practical implementation of macromolecule IC, we believe it may be most beneficial to avoid purifying the polymer sample as it undergoes multiple cycles using an equimolar coupling ratio at all times (though IC on a solid support may be an exception). It remains to be tested if the inherent properties of the IC process, particularly exponential IC, are able to maintain a low sample dispersity while increasing average molecular weight in spite of real-world factors. However, we are hopeful that a process using optimized chemistry would be capable of producing low dispersity polymer samples in multi-gram scale with minimal purification. This would be a strikingly complementary outcome to the use of IC in generating truly uniform polymer samples.

6. Conclusion

Iterative coupling (IC) is a powerful process which has seen abundant experimental work in the synthesis of linear oligomers and polymers, but whose theoretical description has seemingly been neglected. With this work, we have attempted to address this deficiency. Starting with a broad view, upper bounds were established on the maximum possible growth rate of any IC process. The fastest (exponential) growth rates were shown to be the result of a simple and synthetically economical idea; producing all coupling reagents in a particular step by recombining the product of the immediately prior coupling reaction. The simplest practical implementation of exponential IC leads to a 2^N-mer after N coupling reactions. Faster growth rates can be achieved, but at perhaps too great an increase in complexity.

The general effect of the functional groups and coupling reactions chosen was then described, allowing focus on more promising combinations. Studying two-component reactions, it was determined that homocoupling reactions are overall less suitable as the basis of gainful IC schemes. Cross-coupling reactions involving at least one differentially protected or functionally desymmetrized species are more versatile, and can provide the control required for efficient IC schemes. These are important applications of these classes of materials, which are often overlooked due to their reduced synthetic accessibility.

In sequence, the main schemes used for linear and exponential IC were investigated in depth, and the concept of cycle efficiencies was established as a more effective method of describing the overall efficiency of an IC process. The relevant formulas for each process were explicitly provided, based on the individual reaction yields and coupling ratios, and some properties of the equations were

investigated. This information is useful to better characterize work with IC chemistry in general, and assists in guiding the proposal of new and efficient methodologies.

Lastly, in what appears to be a novel proposal, the effect of using IC with initially non-uniform samples was explored. We have shown that inherent properties of IC result in a surprisingly strong tendency for the dispersity of a sample to greatly decrease after completion of the first IC cycle. Importantly, this result was extended for arbitrary coupling yields, showing the effect is very robust. Further cycles are most beneficial using exponential IC schemes. Experimental verification of these results is an alluring target, and is currently underway.

Supporting Information.
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was made possible by support from the Australian Renewable Energy Agency which funds the project grants within the Australian Centre for Advanced Photovoltaics (ACAP). W.W.H.W. is also supported by the ARC Centre of Excellence in Exciton Science (CE170100026). N.S.N. thanks the Albert Shimmins Fund for a scholarship while preparing this work for publication.

Conflict of Interest

The authors declare no conflict of interest.

Received:
Revised:
Published online:

References

[1] J. Lawrence, S.-H. Lee, A. Abdilla, M. D. Nothling, J. M. Ren, A. S. Knight, C. Fleischmann, Y. Li, A. S. Abrams, B. V. K. J. Schmidt, M. C. Hawker, L. A. Connal, A. J. McGrath, P. G. Clark, W. R. Gutekunst, C. J. Hawker, *J. Am. Chem. Soc.* **2016**, *138*, 6306.

[2] J. Lawrence, E. Goto, J. M. Ren, B. McDearmon, D. S. Kim, Y. Ochiai, P. G. Clark, D. Laitar, T. Higashihara, C. J. Hawker, *J. Am. Chem. Soc.* **2017**, *139*, 13735.

[3] S. C. Solleder, R. V Schneider, K. S. Wetzel, A. C. Boukis, M. A. R. Meier, *Macromol. Rapid Commun.* **2017**, *38*, 1600711.

[4] R. B. Merrifield, *Science* **1965**, *150*, 178.

[5] M. Caruthers, *Science* **1985**, *230*, 281.

[6] J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, *Science* **2013**, *341*, 1238149.

[7] M. A. R. Meier, C. Barner-Kowollik, *Adv. Mater.* **2019**, *31*, 1.

[8] L. Hartmann, H. G. Börner, *Adv. Mater.* **2009**, *21*, 3425.

[9] J.-F. Lutz, *Macromol. Rapid Commun.* **2017**, *38*, 1700582.

[10] J.-F. Lutz, *Sequence-Controlled Polymers*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2018**.

[11] J. W. Lehmann, D. J. Blair, M. D. Burke, *Nat. Rev. Chem.* **2018**, *2*, 0115.

[12] O. I. Paynter, D. J. Simmonds, M. C. Whiting, *J. Chem. Soc. Chem. Commun.* **1982**, 1165.

[13] D. L. Pearson, J. S. Schumm, J. M. Tour, *Macromolecules* **1994**, *27*, 2348.

[14] J. Alzeer, C. Cai, A. Vasella, *Helv. Chim. Acta* **1995**, *78*, 242.

[15] J. C. Barnes, D. J. C. Ehrlich, A. X. Gao, F. A. Leibfarth, Y. Jiang, E. Zhou, T. F. Jamison, J. A. Johnson, *Nat. Chem.* **2015**, *7*, 810.

[16] T. Aoki, T. Kaneko, M. Teraguchi, *Polymer (Guildf).* **2006**, *47*, 4867.

[17] S. A. Hill, C. Gerke, L. Hartmann, *Chem. - An Asian J.* **2018**, *13*, 3611.

[18] T. T. Trinh, C. Laure, J.-F. Lutz, *Macromol. Chem. Phys.* **2015**, *216*, 1498.

[19] S. Binauld, D. Damiron, L. A. Connal, C. J. Hawker, E. Drockenmuller, *Macromol. Rapid Commun.* **2011**, *32*, 147.

[20] J. M. Tour, *Chem. Rev.* **1996**, *96*, 537.

[21] J.-F. Lutz, J. M. Lehn, E. W. Meijer, K. Matyjaszewski, *Nat. Rev. Mater.* **2016**, *1*, 16024.

[22] F. P. V. Koch, P. Smith, M. Heeney, *J. Am. Chem. Soc.* **2013**, *135*, 13695.

[23] T. Kawaguchi, K. L. Walker, C. L. Wilkins, J. S. Moore, *J. Am. Chem. Soc.* **1995**, *117*, 2159.

[24] R. Zuckermann, J. Kerr, S. Kent, W. Moos, *J. Am. Chem. Soc.* **1992**, *114*, 10646.

[25] S. Pfeifer, Z. Zarafshani, N. Badi, J.-F. Lutz, *J. Am. Chem. Soc.* **2009**, *131*, 9195.

[26] S. Pfeifer, Z. Zarafshani, N. Badi, J.-F. Lutz, *J. Am. Chem. Soc.* **2009**, *131*, 9195.

[27] T. T. Trinh, L. Oswald, D. Chan-Seng, L. Charles, J.-F. Lutz, *Chem. - A Eur. J.* **2015**, *21*, 11961.

[28] S. C. Solleder, M. A. R. Meier, *Angew. Chemie - Int. Ed.* **2014**, *53*, 711.

[29] O. E. Vercillo, C. K. Z. Andrade, L. A. Wessjohann, *Org. Lett.* **2008**, *10*, 205.

[30] A. C. Boukis, M. A. R. Meier, *Eur. Polym. J.* **2018**, *104*, 32.

[31] T. Izumi, S. Kobashi, K. Takimiya, Y. Aso, T. Otsubo, *J. Am. Chem. Soc.* **2003**, *125*, 5286.

[32] L. Hartmann, E. Krause, M. Antonietti, H. G. Börner, *Biomacromolecules* **2006**, *7*, 1239.

[33] R. B. Grubbs, R. H. Grubbs, *Macromolecules* **2017**, *50*, 6979.

[34] P. C. Hiemenz, T. P. Lodge, *Polymer Chemistry*, CRC Press, Boca Raton, USA, **2007**.

[35] P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, USA, **1953**.

[36] B. Everitt, *Stat. Methods Med. Res.* **1996**, *5*, 107.

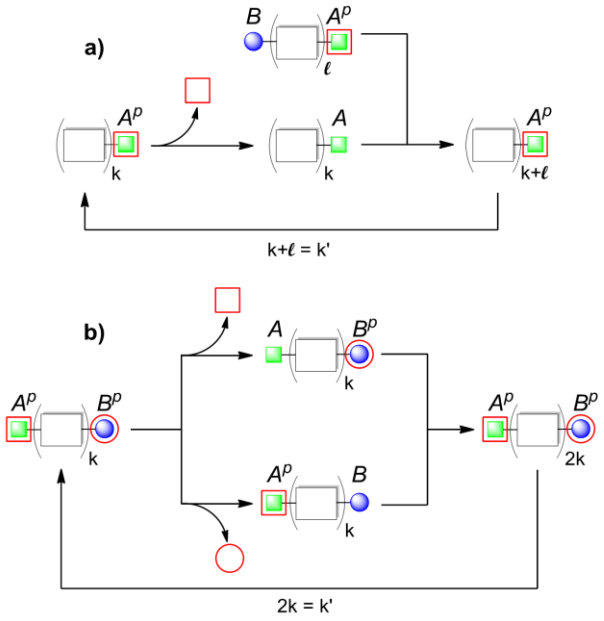

Figure 1. Representation of the deprotection and coupling reactions used in the main iterative coupling cycles towards linear polymers. (a) Linear IC. (b) Exponential IC. Capital letters in italic and

coloured shapes indicate generic functional groups. The superscript ^{p} and the red outline indicate the functional group is in a protected or deactivated form.

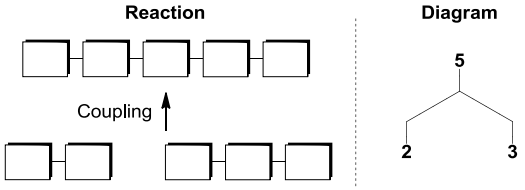

Figure 2. Example of a diagram showing the coupling of a dimer and trimer to a pentamer. Functional groups are omitted.

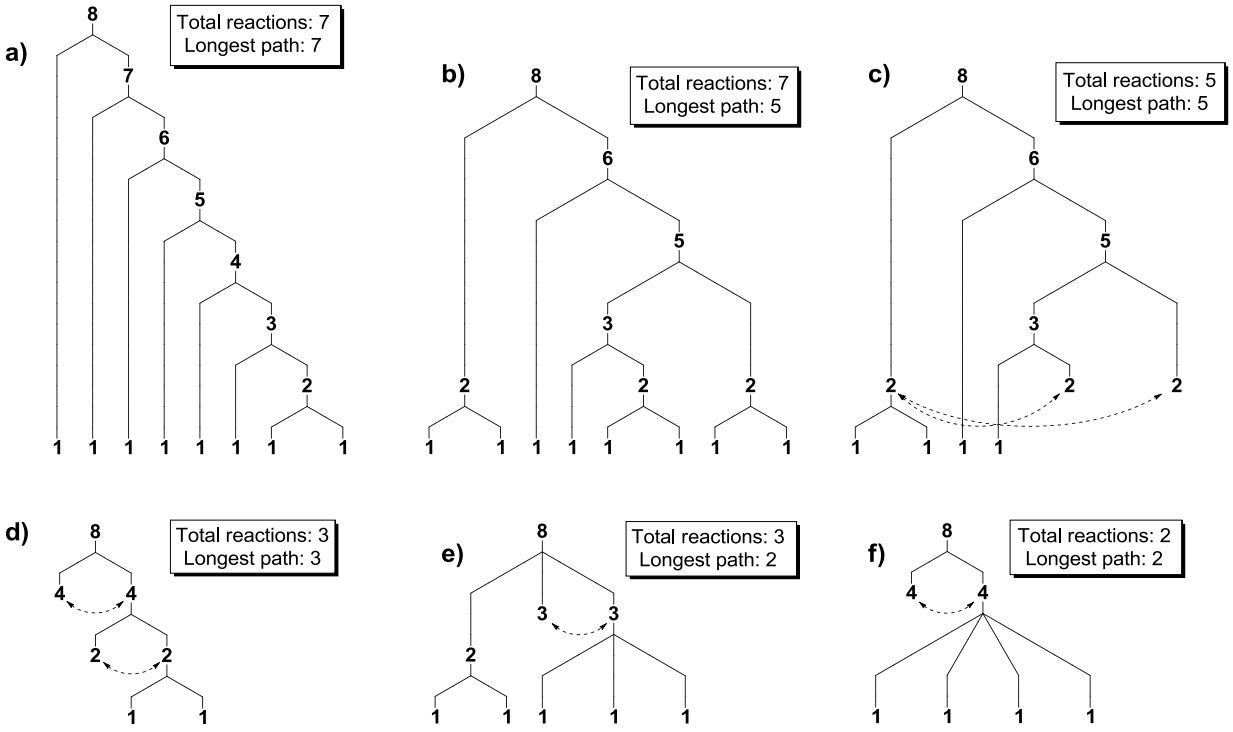

Figure 3. Multiple possible syntheses of an octamer via IC. a) The longest possible diagram, coupling each monomer sequentially. b) A more branched diagram, but with the same overall number of

reactions. c) A simplification of the prior diagram, without redundant reactions. d) The most concise (exponential) IC diagram possible, assuming only two chains couple per reaction. e) and f) Even more concise diagrams, possible only if more than two chains couple in a single reaction.

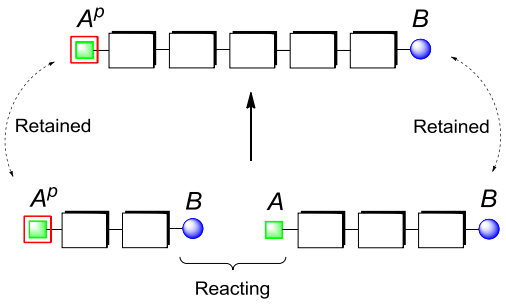

Figure 4. Example of one of the possible coupling events between A^pB and AB chains under BA cross-coupling conditions, represented succinctly as A^pB+AB. The functional groups present can be classified as either reacting or retained functionalities. Note that though the rightmost B functionality is retained in this particular coupling event, it may react in a further coupling event in the same reaction mixture (i.e. with additional AB chains).

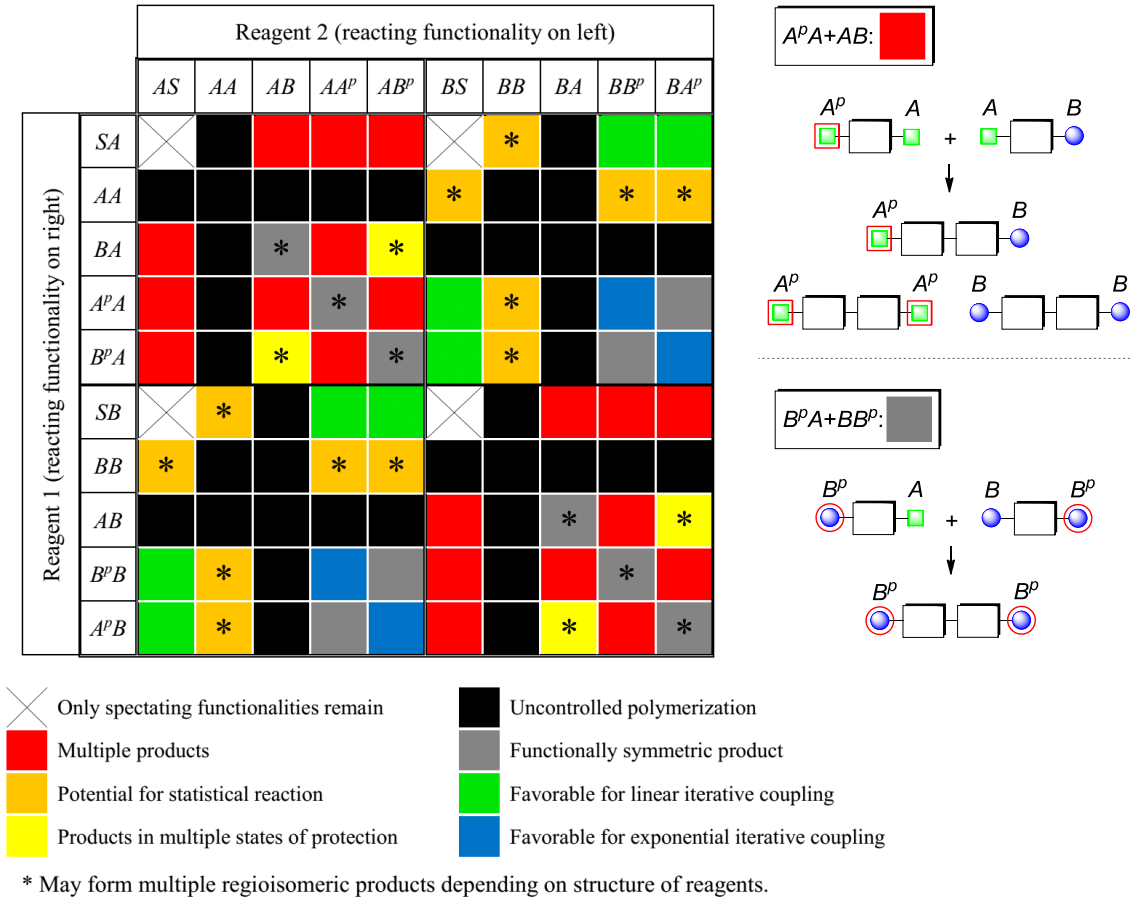

Figure 5. Description of the general outcomes of two-component reactions, with variations of the functionalities *A* and *B*. Examples of reactions given to the right.

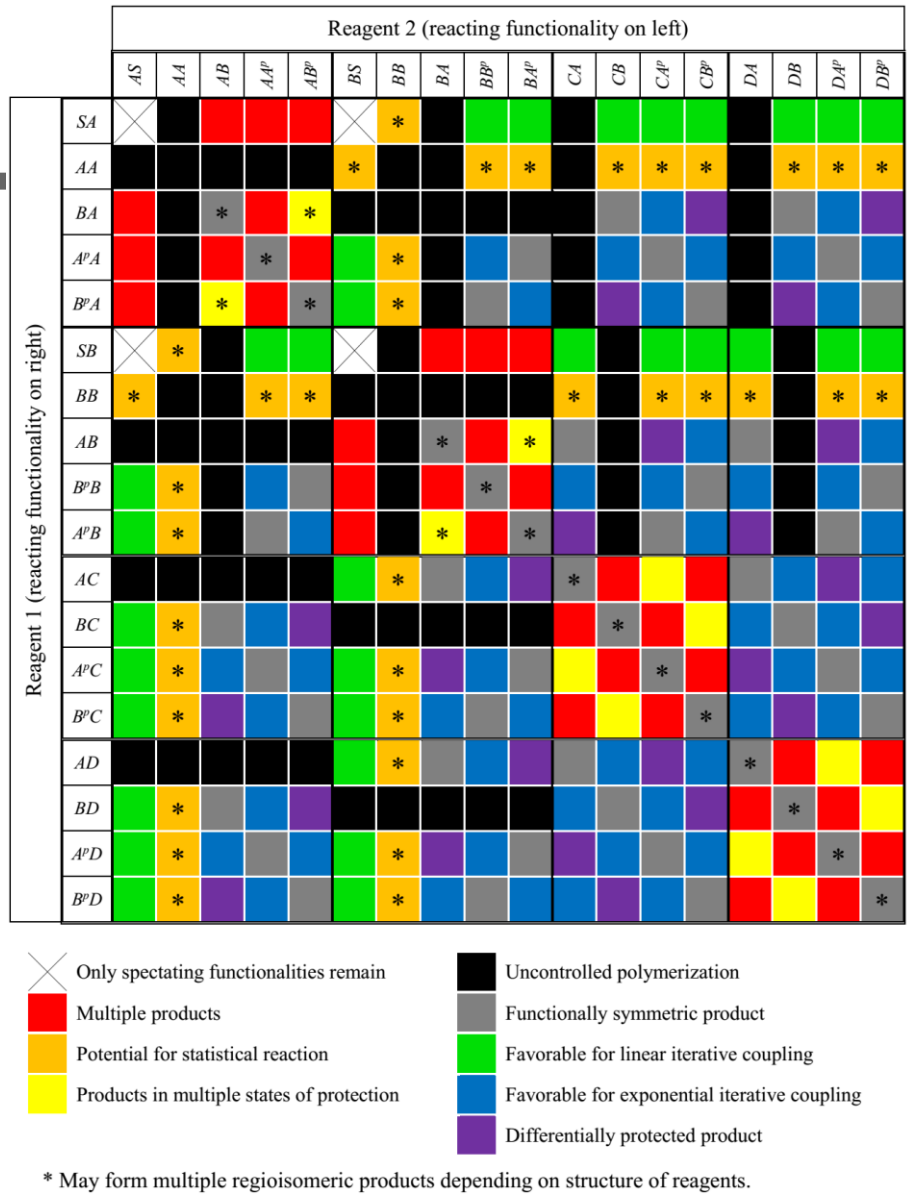

Figure 6. General outcome of all possible two-component reactions.

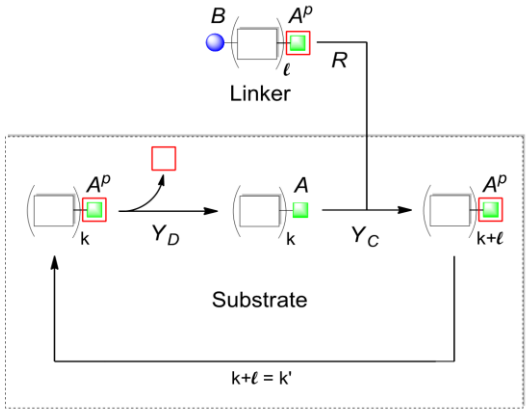

Figure 7. The main linear IC cycle, along with the necessary definitions for calculating cycle efficiencies. Y_D is the deprotection yield for functional group *A*, Y_C is the coupling yield based on limiting reagent (substrate), and R is the molar ratio of linker to substrate in the coupling reaction.

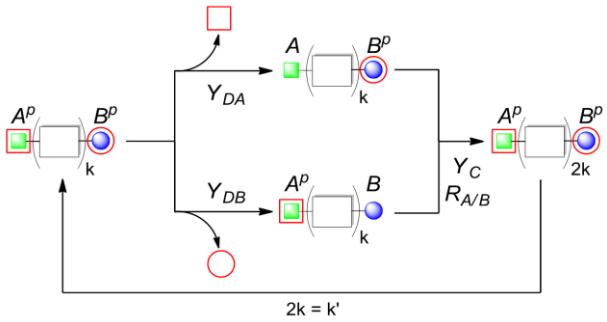

Figure 8. The main exponential IC cycle and the relevant parameters for calculation of cycle efficiencies. Y_{DA} and Y_{DB} are the deprotection yields for respective functional groups *A* and *B*, Y_C is the coupling yield based on limiting reagent, and $R_{A/B}$ is the molar ratio of chains containing *A* and *B* in the coupling reaction.

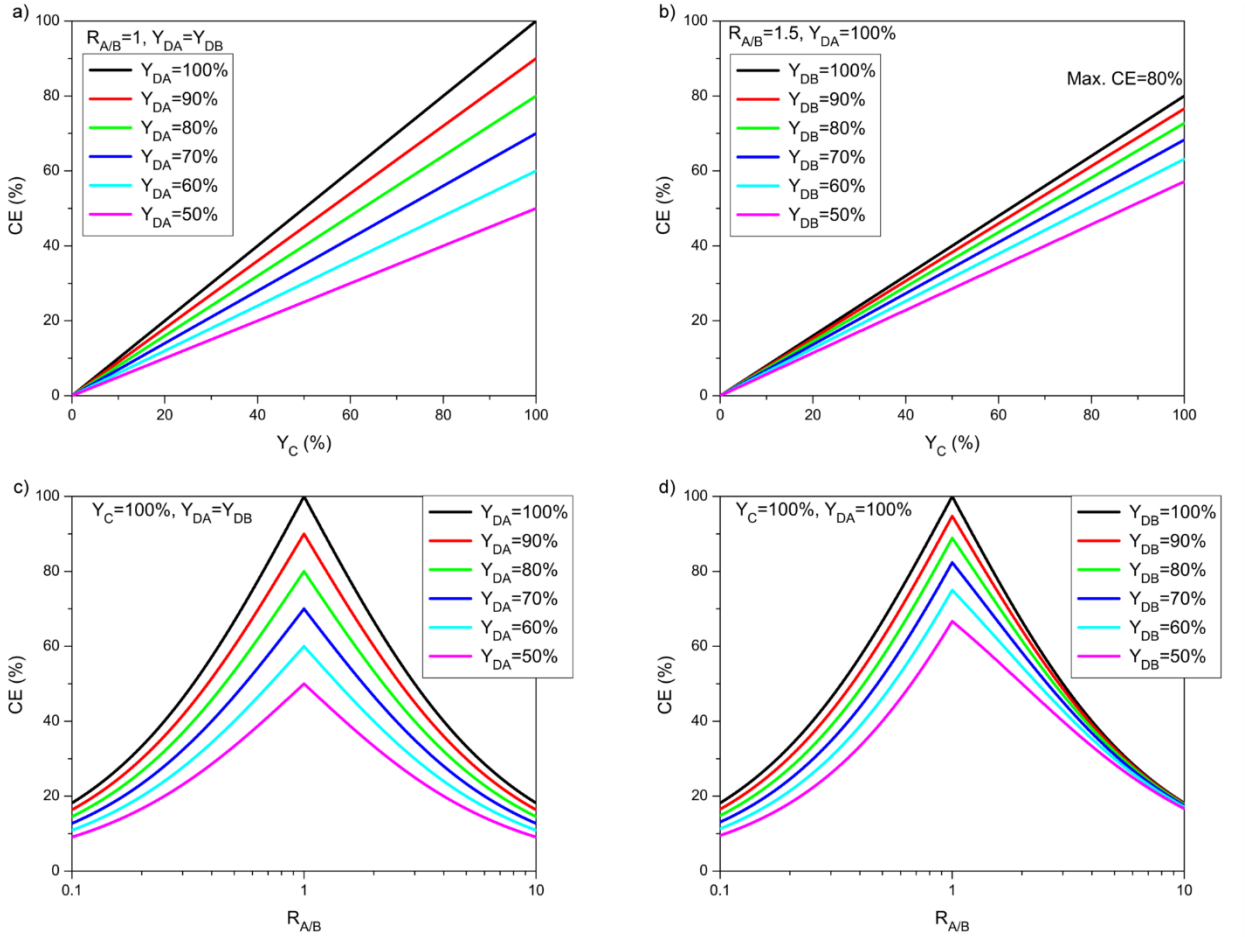

Figure 9. Comparison of cycle efficiencies in exponential IC varying the stoichiometric parameters of the individual steps. a) Variation of coupling yield in equimolar coupling reactions and equal deprotection yields. b) Variation of coupling yield non-equimolar coupling reactions and different deprotection yields. c) Variation of the coupling ratios with equal deprotection yields. d) Variation of the coupling ratios with different deprotection yields.

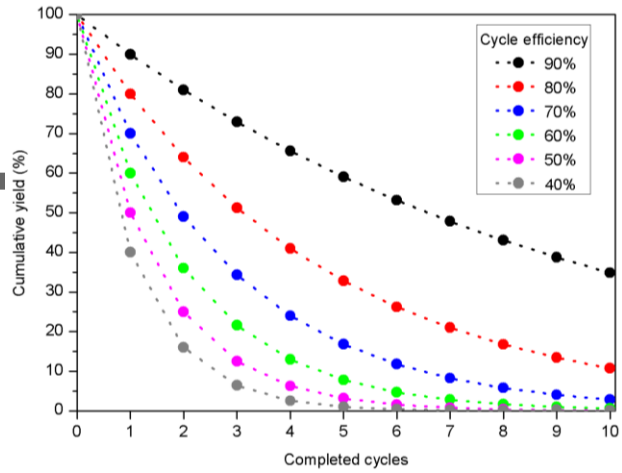

Figure 10. Plots of overall cumulative yield in an IC process with a constant cycle efficiency. Lines added for visual guidance.

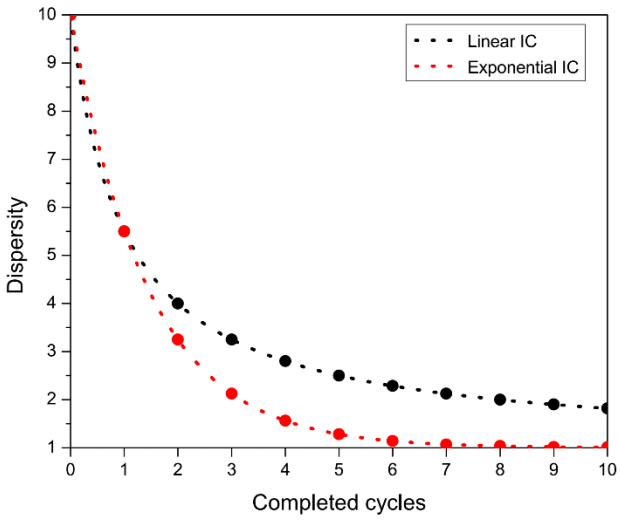

Figure 11. Effect of ideal linear and exponential IC cycles on the dispersity of any distribution with initial dispersity $Đ_0 = 10$. The lines represent the graphs of Equation (20) and (21).

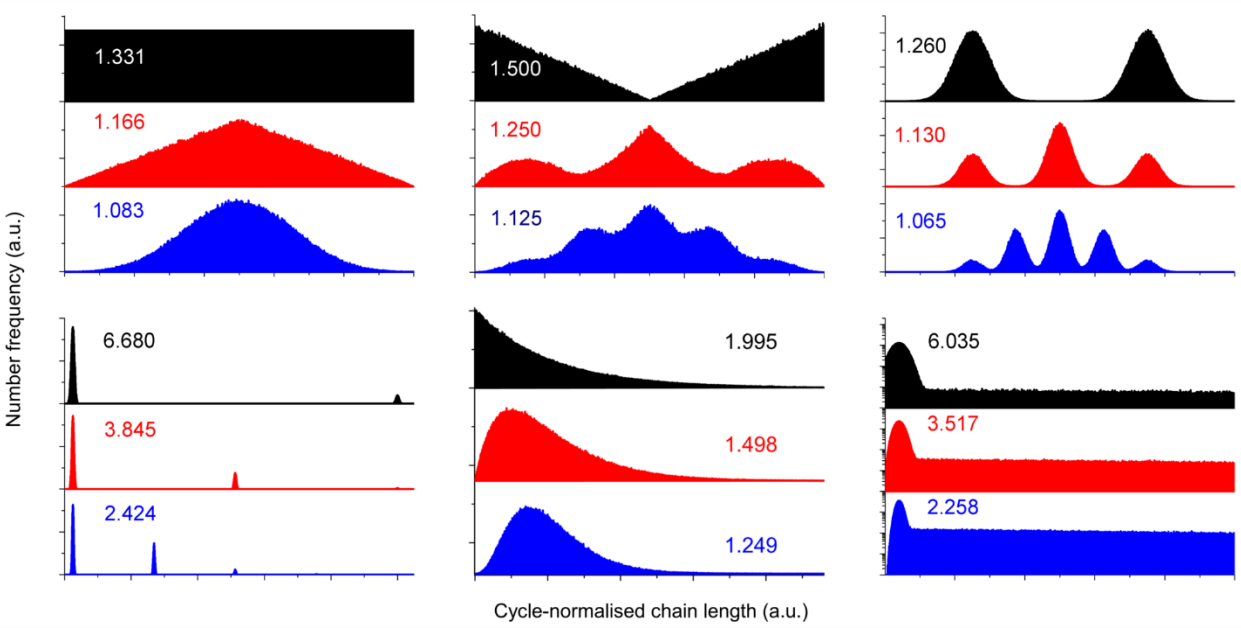

Figure 12. Simulated distributions undergoing consecutive ideal exponential IC cycles, from black to red to blue. The bottom-right distributions have a logarithmic y-axis, and long tails which are not fully shown. After every cycle, the x-axis is contracted by a factor of 2. The numbers indicate distribution dispersities. Values after a cycle match predictions to better than 0.2%.

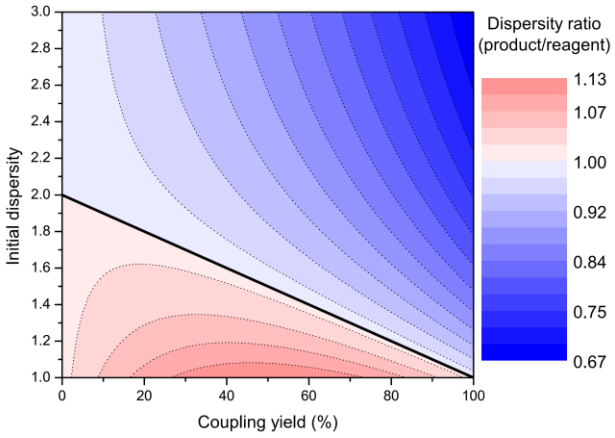

Figure 13. Contour plot of Equation (29), representing the outcome of the first IC cycle with non-uniform macromolecule samples. The solid critical line represents unchanged dispersity. Below the line the dispersity increases, and above the line it decreases.

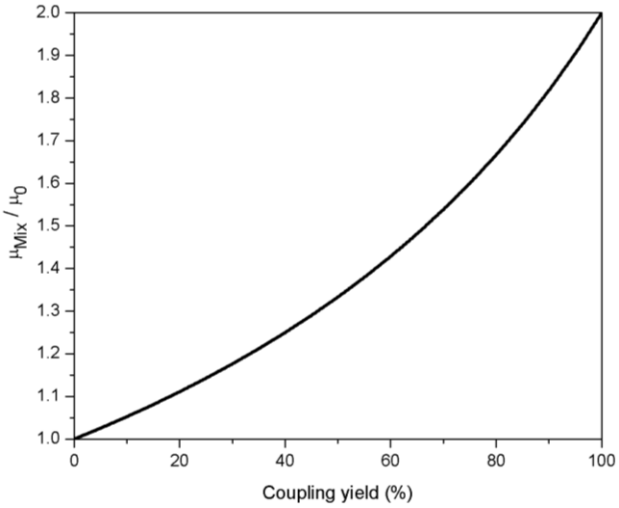

Figure 14. Increase in the number-average degree of polymerization of a non-uniform sample, after the coupling reaction in the first cycle of macromolecule IC with an equimolar ratio of active functional groups.

The fundamental properties of iterative coupling are analyzed to guide further experimental work. Polymer chain growth rates and functional group choices are assessed. Cycle efficiencies are defined to allow detailed stoichiometric characterization. Lastly, the effect of sample dispersity is studied, leading to the proposal of macromolecule iterative coupling as a general method for controlled polymer synthesis.

sequence-controlled polymers

Nicolau Saker Neto,* Dr. David J. Jones, Dr. Wallace W. H. Wong

Theoretical Aspects of Iterative Coupling for Linear Oligomers and Polymers