Chapter 1 Folding and Replication in Complex Chemical Systems

The origin of life and the de novo synthesis of life are among the greatest challenges in contemporary science. It is well known that life consists of many complex processes in which self-replication and folding play key roles. One of the most important features of a living system is its ability to replicate itself. Self-replication is one of the most important ingredients in the origin of life and self-replicating molecules are a promising starting point for the de novo synthesis of life. Folding is the process by which proteins and nucleic acid strands acquire a three-dimensional structure with a biologically functional conformation in a fast and reproducible manner. The most critical metabolic processes in organisms rely on the correct folding of biomolecules such as proteins and DNA. Therefore, the study of self-replication and folding may not only help us uncover the mystery of the origin of life, but may also guide us to synthesize life.

Dynamic combinatorial chemistry (DCC) is a promising tool for creating and studying chemical complexity. DCC not only allows us to simulate and control the process of self-replication, but also to study the process of folding. More importantly, as demonstrated in this thesis, DCC provides a simple way to combine these two complex processes in a single system. In this chapter, we first briefly highlight the current state of the art of synthetic folding and self-replicating systems, and then we summarize the folding and self-replicating systems constructed by using DCC. Finally, the contents of this thesis are outlined.
Chapter 1

The central principle of molecular biology includes two important parts: the replication of genetic information and its transcription into proteins. Replication of genetic information is a biological process that occurs in all organisms and is the basis of biological inheritance. Proteins, as the basic substances of cells, play an important role in all living systems. In order to achieve functions, the protein needs to be correctly folded into a three-dimensional structure. The synthesis of proteins in the cell starts with messenger RNA which is formed through the transcription of DNA, which is then translated into unfolded or randomly coiled peptide chain. Finally, the linear peptide chain is folded into a three-dimensional structure borne from the primary sequence. The primary sequence of a protein plays a decisive role in its folding, but environmental changes and chaperones may also affect the spatial structure of the protein (secondary, tertiary and quaternary structures) and biological activity. In the majority of the protein structures hydrophobic residues are concealed inside the protein, minimally exposed to the solution, and hydrophilic residues are exposed to the outside and interact with the solution to stabilize the protein conformation.

1.1 Systems chemistry and the de-novo synthesis of life

For many years, chemists have been more inclined to study isolated substances than complex mixtures of molecules that can interact and react with each other. Nowadays, this situation is likely to change because of the great interest in systems biology and the availability of analytical techniques.¹,² Unlike other areas of chemistry that focus on simple systems, systems chemistry tends to study multiple variables simultaneously. Over the past decade, scientists have established three thermodynamic models for characterizing complex chemical systems: (i) systems that achieve minimum energy states under thermodynamic control, and (ii) systems that can be trapped in local minimum energy states under kinetics control, and (iii) systems that are kept away from equilibrium by continuous energy input.³⁵

The emergence of systems chemistry has spurred effects directed towards the de-novo synthesis of life and the origin of life. A reasonable approach toward constructing the first protocell, from the highly diverse components available on prebiotic Earth, involves the integration of primitive metabolism, self-replication and membrane subsystems through different physicochemical mechanisms and reaction pathways.⁶ The prebiotic Earth can be seen as a huge reactor that contains complex, different types of small molecules that engage in a huge variety of possible interactions and reactions. When such systems are maintained far from equilibrium, this complex collection can explore an extremely large number of possible reaction pathways. Attempts at identifying thermodynamically sound chemical pathways leading to life often involves studying a mixture of
Folding and replication in complex chemical systems

potential molecular components and studying the chemical and physical interactions between them (eg, interconversion, condensation, and polymerization).

DNA and proteins are two of the most important molecules in all known living systems. However, the DNA-protein tandem seems too complicated to appear spontaneously. There is a “chicken and egg” problem: which is first, chicken (protein, phenotype) or egg (DNA, genotype)? The most widely accepted solution to this problem is the existence of an RNA world before DNA and proteins. RNA is ubiquitous and plays a different role in nature: in addition to storing genetic information, it is also involved in gene expression, catalysis and translation of some steps in the flow of genetic information. The chemical basis of this versatility relies on the fact that RNA is usually a single-stranded molecule, thus facilitating intramolecular base pairing to generate more types of three-dimensional structure/functional motifs than double-stranded DNA. The secondary structure of RNA also provides a simplified and appropriate phenotype of the genotype, which is useful for solving related evolutionary problems. Therefore, the clear relationship between sequence, structure and function makes RNA the best model for molecular evolution experiments and computational studies. Although the possibilities of the RNA world are supported by some experiments, a "spontaneously generated" RNA world that produces proteins and DNA is not obvious. Perhaps more importantly, the initial emergence of the RNA world itself remains an open question.

Self-replication can be considered as a key process for a protocell. Autocatalysis is a relatively rare and complex behavior in chemistry. In addition, autocatalysis is also the basis for oscillatory behavior in several reactions. Autocatalysis is a fundamental concept for all living systems that make more copies of themselves, and it produces a series of potentially more complex systems through evolution. Without autocatalysis, the transition from a chemical system to a biological system does not seem to be feasible. Furthermore, the process of self-replication makes information transfer between molecules and systems possible, which is essential for Darwinian evolution.

Metabolism is another important ingredient. The metabolic network in life is a large chemical reaction system with amazing complexity and adaptability. It has two main purposes: the first is to convert energy in the environment into a form of energy that is useful to the organism; the second is to obtain nutrients from the environment to synthesize the small molecules needed for cell growth. These small molecules include DNA nucleotides, RNA nucleotides, sugars, lipids and amino acids. Most reactions in an organism are controlled by enzymes, which determine the rates and the selectivities. An enzyme must be folded into a specific configuration before it can perform its function. So the process of folding is the key to bring functions to biomolecules.
Chapter 1

As part of building a synthetic protocell, we must find a reasonable way to combine these two processes of self-replication (genotypes) and folding (phenotypes). In past studies, scientists have established separate systems that show either folding or self-replication. In the following sections (1.2 and 1.3) we will briefly summarize synthetic folding and replicating systems.

1.2 Synthetic folded structures - foldamers

The word ‘foldamer’ was used for the first time in 1996 by Gellman. He proposed the definition: "Any polymer with a strong tendency to adopt a specific compact conformation". In 2001, Moore added on this definition as "A foldamer is any oligomer that folds into a conformationally ordered state in solution, the structures of which are stabilized by a collection of non-covalent interactions between non-adjacent monomer units". In recent years, after continuous experimentation and exploration, chemists have successfully synthesized various helical structures and functional folds.

According to the different backbones of the synthetic folded structures, foldamers can be roughly classified into two classes: aliphatic peptides and aromatic foldamers. In this section we will briefly highlight the most important achievements for synthetic foldamers.

1.2.1 Aliphatic peptide foldamers

Folded peptides are the structural basis for proteins to achieve their functions. The way in which the 20 natural amino acids are arranged dictates dynamicity and the functional shape. Synthetic peptide foldamers can also adopt well-defined conformations. In most cases these are helical structures. Strategies for the modification of folded structures have focused on modifying the amino-acid side chains or the peptide backbones. Modifying the backbone of peptides has proven to be particularly effective and involves introduction of non-natural peptides. Figure 1.1 summarizes the backbones of the most popular aliphatic peptides foldamers.

![Figure 1.1. The backbones of aliphatic peptides foldamers](image-url)
β-Peptide foldamers are determined by 12 and 14 helices. Gellman et al. constructed a peptide foldamer consisting of cyclic β amino-acids to form a stable 12-helical conformation featuring hydrogen bonds between sites separated by 12 atoms (Figure 1.2a). Seebach et al. designed β-peptides that adopt a 14-helix with intramolecular hydrogen bonds between amides that are 14 atoms apart (Figure 1.2b).

![Figure 1.2](image1.png)

**Figure 1.2.** The β-peptide (a) 12 and (b) 14 helical folds. Adapted from ref.22

Analogous γ-peptide foldamers are more likely to acquire conformational stability by hydrogen bonding between adjacent amide bonds. The NMR of the cis-γ-amino-1-proline γ-peptide in water shows that it is folded into a C9-related structure. In the secondary structure, two amide bonds that bind to α and γ positions of proline are in the same plane as the proline (Figure 1.3a). Sharma and Kunvar used carbofuranosyl as a conformational stabilizing side chain to force the γ-peptide to form an ordered 9-helix structure (Figure 1.3b).

![Figure 1.3](image2.png)

**Figure 1.3.** C9 H-bonded rings in γ-peptide foldamers. Adapted from ref.24,25
Chapter 1

1.2.2 Aromatic foldamers

The field of aromatic foldamers started with the synthesis of a helical structure in which aromatic rings were connected directly to each other. These molecules can adopt different conformations due to the rotatability of the C-C bond. Simpkins et al. synthesized a biphenyl oligomer by direct bonding of an ortho-benzene subunit resulting in an angle of 60° between adjacent benzene rings (Figure 1.4a).26 Yashima et al. synthesized another type of biphenyl structure using a meta-substituted benzene ring, giving rise to an angle of 120° between adjacent aromatic rings (Figure 1.4b). Steric hindrance limits its conformational freedom, leading to the formation of a helical structure of about five benzene rings, which can form a double helix in water.27 Due to the structural similarity with benzene, six-membered aromatic heterocyclic rings such as pyridine derivatives can also be included to obtain biphenyl-like molecules, folding into various structures (Figure 1.4c).28-33 In addition, five-membered heterocycles such as thiophene, furan, triazole and other conjugated structures and combinations thereof can also form different helical structures (Figure 1.4d).34-38 Acetylene units can form a biphenyl-like structure when they connect two benzene rings. The m-phenylene acetylene oligomers constructed by Moore can not only form a cyclic six-membered ring but can also be further extended to form a helical configuration (Figure 1.4e).39,40

![Molecular structures](image)

**Figure 1.4.** Molecular structures with different aromatic rings folded into different helical shapes. Adapted from ref.26-28,34,39

Apart from direct bonding, the use of amides to link different aromatic units is one of the most common methods for obtaining aromatic foldamers.41 The amide bond has a planar rigid structure
Folding and replication in complex chemical systems

and is also a good hydrogen bond donor and acceptor. It can be introduced into the aromatic system and can form a hydrogen bond with a neighboring hydrogen bond donor or acceptor to limit bond rotation, which facilitates the interaction between aromatic rings and the adoption of a specific conformation. Recently, rapid progress has been made in constructing aromatic folded structures using amide bonds.42-46

A series of hydrogen-bonding-driven non-heterocyclic aromatic amide foldamers were reported after the pioneering work of Gellman on peptide-based foldamers. Gong's group reported an aromatic amide foldamer, which consists of an amide-linked benzene ring that forms a folded structure driven by hydrogen bonds (Figure 1.5a).47 After that, Li and Zeng et al. prepared similar oligomers capable of forming folded structures using a methoxy group at the 2-position of the benzene ring (Figure 1.5b).48,49 In addition, the same research group synthesized novel aromatic hydrazide foldamers by replacing the amide bonds with hydrazide bonds (Figure 1.5c).50,51

![Figure 1.5. Non-heterocyclic aromatic amide foldamers. Adapted from ref.47,48,50](image)

In the design and preparation of amide-based aromatic foldamers, aromatic heterocyclic molecules are widely used to control the conformation of the foldamers through hydrogen bonds.52-54 Huc and Lehn reported a pyridine-based amide foldamer in 2000 (Figure 1.6a). They synthesized a series of oligomeric molecules with helical conformations constructed by pyridine amide units. The results showed that the helical oligomer exhibited a dynamic exchange between a single and a double stranded helix in solution. In a subsequent study, they found that the δ-amino acid derived from quinoline could generate oligomers that possessed a compact folded conformation through a stable tricentric hydrogen bond (Figure 1.6b).55 Apart from the pyridine and quinoline, other conjugated rings such as derivatives of pyridine and quinoline, naphthalene and anthracene can also be used as building blocks to construct various helical structures (Figure 1.6c, d and e). Huc et al. constructed different types of helical oligomers by using quinoline derivatives,56 hydrazine derivatives57,58 and
Chapter 1

their hybrids.\textsuperscript{42,59} Chen et al. constructed a new class of helical oligomers using heterozygous phenanthroline and ortho-benzene rings.\textsuperscript{60}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.6.png}
\caption{Heterocyclic aromatic amide foldamers. Adapted from ref.\textsuperscript{55-57,60,61}}
\end{figure}

1.2.3 Aliphatic and aromatic hybrid foldamers

The hybrids of aromatic and aliphatic building blocks have the potential to possess the characteristics of both the aromatic and the aliphatic folded structures.\textsuperscript{62-64} In 1995, Lokey and Iverson described the first such hybrid that folded in an aqueous environment.\textsuperscript{65} Huc et al. introduced a methylene group in a rigid helical structure, which, upon fine-tuning gives rise to a new helical structure—the herringbone spiral.\textsuperscript{66} For molecules with alternating aromatic and aliphatic units, the intramolecular structure is formed by π-π interaction between adjacent or non-adjacent aromatic rings and van der Waals interactions or hydrogen bonds between adjacent or non-adjacent aliphatic units.\textsuperscript{67-69}

1.3 Synthetic self-replicating systems

Understanding the essence of the process of self-replication and establishing self-replication theory consistent with living systems requires the development of minimal synthetic self-replicating systems. During the past two decades, a large number of such self-replicating systems have been developed
Folding and replication in complex chemical systems

from simple self-replicators operating in isolation to complex replicator networks. In this section, we will summarize the most significant developments and achievements for synthetic self-replicating systems based on biological molecules, including DNA, RNA and peptides and small organic molecules.

In general, a molecular self-replicating system transfers structural information through an autocatalytic process. The rate of the autocatalytic reaction is directly related to the amount of catalytically active template. Self-replicating systems usually feature two reaction processes. First, the self-replicators must form spontaneously at the beginning of the reaction. Secondly, the product formed during the early stages of the reaction catalyzes the synthesis of copies of itself via autocatalysis. Figure 1.7 shows a minimal replicating system, in which several reaction channels need to be considered.

![Figure 1.7](image)

Figure 1.7. Minimal model of self-replication. Adapted from ref.74

The first channel is the non-catalytic bimolecular reaction between components A and B, the product of which is the template T of the self-replicating system. In the second channel, the product is a complementary binary complex [AB], which has no active recognition site. Although this channel can speed up the initial reaction rate, the resulting template is catalytically inert. The third reaction channel is autocatalytic. In this channel, first, starting materials A and B are bound to the template T by molecular recognition to form a catalytic ternary complex [A-B-T]. Then A and B react to form the noncovalent dimer of the product [T-T]. Its dissociation produces two separate template molecules,
Chapter 1

each of which can participate in another replication cycle. In an ideal self-replication system, the number of templates in the system is doubled in each cycle, so their growth is exponential. However, as components A and B are depleted, the growth rate diminishes and eventually stops. Overall this leads to the sigmoidal growth curve, typical of autocatalytic reactions in closed systems.

1.3.1 DNA-based self-replicating systems

In 1986, Günter von Kiedrowski developed the first non-enzymatic chemical self-replicating system based on an oligonucleotide strand with a palindromic sequence (Figure 1.8).75 In the replication cycle, a trinucleotide CCG (protected at the 5’ end) was coupled with another trinucleotide CGG (protected at the 3’ end) in the presence of EDC to generate a template hexanucleotide CCGCGG. The resulting 5’ and 3’ protected hexanucleotide CCGCGG facilitated the formation of a reaction product that was both complementary and identical to the template via Watson-Crick base pairing. The resulting double-stranded product can then dissociate into two single-stranded molecules that can be used as templates for next catalytic cycle.

![Figure 1.8. The first chemical self-replicating system reported by von Kiedrowski in 1986.](image)

The existence of an autocatalytic pathway in this system was demonstrated by the addition of a small amount of template at the beginning of the reaction which was found to accelerate product formation. However, the uncatalyzed background reaction contributed significantly to the overall reaction rate. Another limitation of this system is that the reaction rate is slow and only 12 % product formed after 4 days reaction.

In a subsequent experiment, von Kiedrowski and colleagues found that the problem of low efficiency can be solved by replacing the phosphodiester bond in the DNA strand with a phosphoramidate bond.
linkage. Furthermore, the rate of self-replication relative to the background reaction was increased in this system, and they therefore observed the first synthetic self-replicator with a sigmoidal growth profile. However, the growth of the replicator was parabolic rather than exponential. After that, they found that this approach can also be applied to a self-replicating system consisting of three nucleotide building blocks. Zielinski and Orgel also established a self-replicating system based on 3'-amino-3'-deoxynucleotides. However, the replication of the resulting tetranucleotide was, again, hampered by product inhibition.

To obtain a self-replicating system with exponential growth properties, von Kiedrowski and colleagues designed a method called SPREAD (surface-promoted replication and exponential amplification of DNA analogs). In their method, the single-stranded template was first immobilized onto the surface of a solid support and then the complementary nucleotide fragments were bound to the template. The nucleotide fragments were linked by a coupling reagent and the product was then liberated from the template at elevated temperature to free the template and product for a second replication cycle. The liberated product was bound to free sites on the surface of the solid support and the above process was repeated to obtain exponential growth.

1.3.2 RNA-based self-replicating systems

Paul and Joyce successfully developed the first RNA-based self-replicating system in 2002 (Figure 1.9). Their experiments employed a modified R3C ligase that catalyzes the formation of 3',5'-phosphodiester bonds between two separate RNA molecules. The RNA ribozyme template T is capable of achieving its own precise replication by ligating two RNA subunits A and B through a ternary complex. The addition of a pre-formed template to the reaction revealed a significant increase in the initial rate of template formation, indicating that template formation was an autocatalytic process. However, the increase in reaction rate occurs only in the initial stage of the reaction, which indicates that product self-inhibition occurs. Kinetic fitting revealed that the reaction contained two phases. The increase in replication rate observed early in the reaction is attributed to the formation of the [A_B_T] complex. In contrast, the second, slower phase is the bimolecular reaction of A and B without a template. The authors suggested that the inefficiency of the designed RNA system is due to the similarity in the nucleotide sequences of components A and B, which leads to the formation of the inactive binary complex [A_B], which does not dissociate even upon addition of the template. Subsequently, they found that the deleterious effects on the replication due to the formation of the stable complex [A_B] can be avoided by premixing T with B before adding A or by adding an excess of A to the reaction mixture.
They have further modified the structure of the ligases in order to generate two RNA enzymes that were capable of catalyzing each other’s synthesis. This modification resulted in cross-replication, but the new self-replicating system still exhibited ‘burst phase’ kinetics similar to earlier ones. In subsequent work, they developed two mutually replicative RNA enzyme systems that can catalyze the synthesis of each other from a mixture of four different structural units by template guidance. These cross-replicating RNA enzymes undergo sustained amplification in the absence of proteins or other biological materials and can persist indefinitely. In addition, they also made a series of modifications to get further insight into the RNA-based self-replication mechanism.

1.3.3 Peptide-based self-replicating systems

Ghadiri et al. reported the first peptide-based self-replicating system in 1996. Their system used a simple protein consisting of a seven-peptide repeat (abcdefg)\textsubscript{n} containing 32 residues, which can be assembled into two entangled coiled-coil structures as a result of hydrophobic and electrostatic interactions (Figure 1.10). Amino-acid residues at positions a and d of the peptide sequence are capable of driving the recognition between the helices by hydrophobic interactions and determine the stability and helical orientation of the coiled-coil structure. Residues at positions e and g in the heptad repeats drive intramolecular recognition by electrostatic interactions. The realization of self-replication of the peptides relies on native chemical ligation. Under the direction of the template, two peptide building blocks containing a thiobenzyl ester and free cysteine, respectively, and the template are assembled into a ternary complex with a coiled-coil structure. Then the ligation occurred to produce a stable amide bond by intramolecular rearrangement at the junction site. Kinetic experiments showed that the addition of template at the beginning of the reaction promoted
the formation of product, showing the ability of the designed coiled-coil peptide to self-replicate. However, due to the high stability of the double-stranded coiled-coil structure, the replication process shows parabolic growth. This is the same problem as the inhibition of self-replication by the dissociation of the double strands previously observed in nucleic acid-based self-replicating systems.

![Image](image_url)

**Figure 1.10.** Peptide based self-replicating system. Adapted from ref.70

Subsequently, Chmielewski and colleagues developed a similar peptide self-replicating system.91 Although the system exhibited a sigmoidal growth curve, the replication process was still inhibited by the stability of the duplex. To solve this problem, Chmielewski reduced the stability of the coiled-coil structure by subtly modifying the sequence of the template peptide.92 This simple modification partially solved the problem and resulted in close to exponential growth. In addition, they found that the incorporation of proline into the peptide sequence can achieve the same effect, because the addition of proline makes the peptide sequence more distorted and reduces the stability of the double-stranded helix structure.93 Ashkenasy et al. constructed a β-sheet-based self-replicating system consisting of alternating hydrophobic and hydrophilic amino acids that can be assembled into nanosheet-like structures in water.94 The assembled nanostructures can act as a template to facilitate the replication reaction. A kinetic analysis shows that the system initially grows exponentially. However, upon aging, the assembly morphology changes and the replication efficiency diminishes.

1.3.4. Non-biological self-replicating systems

In 1990, Rebek and colleagues developed the first self-replicating system based on fully synthetic organic molecules.95 This system utilizes the amide bond formation between an amine and an
Chapter 1

activated ester as a strategy for forming template 3 from adenine derivative 2 (Figure 1.11). Although no sigmoidal growth curve was observed, the initial reaction rate was increased by adding a small amount of amide template to the reaction mixture, confirming the existence of an autocatalytic process. However, in 1994, Menger and his colleagues found that the formation of template 3 was catalyzed by the addition of a simple amide, which caused doubts about the self-replicating nature of the Rebek system.\(^9^6\) This controversy was finally resolved by Reinhoudt et al. through detailed kinetic analysis of the Rebek system.\(^9^7\) Complete kinetic analysis shows that there are five different reaction pathways that promote the formation of the final product, and the contribution of the various reaction pathways depends on the concentration of the reactant. The binary complex mediated catalytic pathway is the primary pathway for product formation in this self-replicating system (pathway II). Subsequently, Rebek and colleagues redesigned a new self-replicating template to make the length of the self-replicating precursors mismatched, hindering the formation of binary complexes, and finally observed the S-shaped growth curve.\(^9^8\) They also developed other similar self-replication systems based on molecular recognition in later studies.\(^9^9\)-\(^1^0^5\)

![Figure 1.11. Non-biological self-replicating systems. Adapted from ref.\(^7^0\)]

In 1997, Sutherland’s group reported a non-biological self-replicating system based on Diels-Alder reaction between maleimide and cyclohexadiene.\(^1^0^6\) The self-replication characteristics of the system were confirmed by the addition of a pre-formed template. Philp and colleagues used the Diels-Alder reaction to construct two structurally similar furan and maleimide-based replicating platforms for studying the effects of structural variation on replication efficiency and other recognition-mediated effects.\(^1^0^7\)-\(^1^1^2\) In addition to utilizing the Diels-Alder reaction, Philp and colleagues pioneered the development of a self-replicating system based on the 1,3-dipolar cycloaddition reaction.\(^1^1^3\)-\(^1^1^5\) Their research found that the construction of efficient self-replicating systems requires a certain degree of
rigidity of the components and a suitable spatial arrangement of recognition sites to form catalytically active ternary complexes.116

1.4 Dynamic combinatorial chemistry

In the previous sections, we briefly summarized the state of the art in the fields of foldamers and self-replicating systems. Although great achievements have been made in these fields, most of the systems developed relied on multi-step organic synthesis and elaborate design. In addition, no connections have been developed between these two fields. Dynamic combinatorial chemistry not only enables the processes of folding and self-replication to occur with a relaxed demand for multi-step synthesis, but also offers the possibility of linking and merging these two processes, as will be shown in this thesis.

The concepts and principles of dynamic combinatorial chemistry (DCC) were pioneered by Sanders and Lehn in the mid-1990s.117-120 DCC relies on a reversible process to spontaneously produce many possible combinations of a set of building blocks. Using a reversible reaction to form a dynamic combinatorial library (DCL), all components in the library can continuously interconvert by exchanging building blocks with each other.

The most commonly used reversible covalent bonds for constructing DCLs are imines, hydrazones and disulfides. Disulfide bonds play an important role in life’s chemistry.121 Proteins contain disulfide bonds, and thiols and disulfides maintain the redox state of cells. Disulfide bonds have the following characteristics:122

(1) In solution, thiols are easily oxidized to disulfides by oxygen in the air (Figure 1.12a).

(2) The exchange of disulfide bonds can occur in the presence of a catalytic amount of thiolate anion (Figure 1.12b).

(3) The disulfide formation and exchange reactions are typically carried out under neutral-weakly basic conditions and slow down under acidic conditions.

(4) The oxidation and exchange reactions can be carried out in aqueous solution, including under physiological conditions.

(5) The oxidation and exchange reactions can be carried out at room temperature with quantitative conversion.
The applications of DCC include: (a) selection of a host or guest; (b) selection of self-replicating molecules and (c) selection of foldamers. In this section, we will summarize the most significant developments and achievements for self-assembly directed self-replication and folding by using DCC.

1.4.1 Folding in DCLs

Balasubramanian’s group found that folding can lead to self-sorting of peptide assemblies by the formation of secondary structure. They used two peptide building blocks that can form β-sheets, which consisted of Leu-Lys repeat sequences of different lengths and featured a cysteine at the ends. The results show that self-recognition did not occur upon fast oxidation of the DCLs. However, homodimers of the long-chain sequence, directed by the formation of β-sheets, were obtained when disulfide exchange was enabled by using a redox buffer based on glutathione. The same strategy has also been applied to tertiary structure-directed peptide self-sorting. Kumar et al. synthesized two different peptides that can form α-helices, which are designed to form parallel homodimeric coiled coil assemblies. Both peptides contain the same sequence, except that all leucines in one of the sequences were replaced with hexafluoroleucine residues. Each peptide is equipped with a flexible Gly-Gly-Gly linker at one end, followed by a cysteine residue, allowing for the formation of a peptide dimer by oxidizing the cysteine thiol group to a disulfide bond (Figure 1.13). The author found that heterodimers were transferred into homodimer in redox buffer.

In addition, Balasubramanian’s group also reported dynamic self-assemblies based on the formation of G-quadruplexes. They chose PNA in their research, because PNA has good water solubility, and
Folding and replication in complex chemical systems

is easily functionalized with amino acids to introduce thiol groups for the exchange reaction. Under kinetic control, dimers TSST, GSST and GSSG are formed in approximately statistical ratio. In contrast, self-sorting was observed under thermodynamic control in the presence of potassium ions, which stabilize the G-quadruplex structure.

Inspired by folding directed self-assembly of biomolecules, chemists have also developed folding-induced dynamic assemblies of synthetic systems. Moore's group developed a series of methods for assembling monomers into folded dimers and oligomers using reversible chemical bonds. In most of their studies, they used imine metathesis catalyzed by oxalic acid in organic solvents. In their initial study, they used a mono-functionalized building block to study the effect of oligomer length on the folding characteristics. They synthesized two types of m-phenylene-acetylene-based building blocks with different lengths, one of which was functionalized by an amino group and the other by an aldehyde group, each capped by an imine (Figure 1.14). Simulations show that six aromatic units are necessary to form a complete helix, while additional aromatic units stabilize this folded structure. Indeed, the results of NMR studies indicate that imines containing two or five aromatic units didn’t form a helical structure, whereas stable helical structures were observed when the oligomers have more than six aromatic units. Folding caused the equilibrium of the reaction to be shifted towards the formation of longer oligomers that can form stable helical structures. Furthermore, the experimental results showed that no helix formed in chloroform, while a helical structure was observed in the more polar acetonitrile. In addition, they also found that the position of the imine bond in the helix did not significantly affect the stability of the helical structure, confirming that the imine is a suitable structural analog of the alkyne linker. Also, if the length of the building block is extended such that it forms a stable helix by itself, further increasing the chain length does not affect the position of the imine equilibrium.
Chapter 1

The authors also studied folding-directed polymerization. They used two m-phenylene-ethynyl building blocks with two blocked amine or aldehyde functional groups. When both building blocks only contained two aromatic units, no polymer formation was observed, but cyclic dimers based on the two building blocks were formed, which were stacked on each other and formed a columnar assembly. Folding-directed polymerization did occur upon changing the polarity of the solvent and the length of the monomer. As expected, the polymer length increases with increasing polarity because spiral formation is more advantageous in polar solvents.

1.4.2 Emergence of self-replicators from DCLs

Philip's group reported the first self-replicating system in a dynamic combinatorial library in 2008. The DCLs they studied contained four different components: two aldehydes, one of which contained an amide pyridine unit with recognition function while the other aldehyde was a simple benzaldehyde, a p-fluoroaniline and fluorohydroxylamine (Figure 1.15). After the library reached equilibrium, two imines and two nitrones were formed. Then a maleimide is added to the reservoir. At the end of the reaction, it was found that only one of the four possible products was selectively produced, and this product was found to have autocatalytic properties. The autocatalytic nature of the process was confirmed by the addition of a small amount of template which was found to accelerate the rate of the reaction.
Subsequently, Giuseppone’s group developed an autocatalytic system using imine-based DCLs in 2009. They found that amphiphilic imines can reversibly assemble into spherical micelles and cylindrical micelles (Figure 1.16). Since the amphiphilic imine is stabilized by forming a supramolecular assembly, the formation of micelles promotes further formation of the imine, resulting in the growth of the aggregates. As the micelles grow, they become unstable, which causes them to split into smaller aggregates. In this case, the reproducing entity is the entire micelle. In the dynamic combinatorial library the imines that were able to form nanostructures were formed selectively.

Figure 1.15. 1,3-dipolar cycloaddition triggered autocatalytic amplification from DCLs. Adapted from ref. 70
Our group serendipitously discovered self-replicating molecules in DCLs made from dithiol functionalized building blocks in 2010.\textsuperscript{137} The building blocks are functionalized with short peptide chains composed of alternating hydrophobic and hydrophilic amino-acid residues to facilitate the generation of β-sheets and an aromatic dithiol core for thiol-disulfide exchange (Figure 1.17). Upon oxidation of these building blocks in water, first a mixture of macrocyclic disulfides forms. Then one of the macrocyclic compounds stacks into fibers due to the formation of β-sheets. When the fibers grow long enough, they become susceptible to shear stress. Mechanically induced breakage of the fibers creates more fiber ends, thereby propelling the replication process in the dynamic combinatorial library toward a single product. The fiber-growth-breakage cycle enables exponential replication. The autocatalytic nature of this process was confirmed by adding a small amount of template product which was found to accelerate the rate of the reaction. The emergence of self-replicators from DCLs can be affected by templates,\textsuperscript{138} mechanical agitation,\textsuperscript{137} solvent environments\textsuperscript{139} and pre-existing replicators.\textsuperscript{140,141} Moreover, these dynamic self-replication systems can exhibit parasitic phenomena similar to biological systems.\textsuperscript{142}
1.5 Aim and outline

So far, many synthetic self-replicating and folding systems have been developed in the past years. However, most of these systems rely on complex multi-step synthesis. In addition, most synthetic self-replicating systems are unable to grow exponentially which is particularly important for Darwinian evolution. In the field of foldamers, only few skeleton structures can form folded molecules. After many years of development, DCC has become a powerful tool for constructing and studying complex systems. It relaxes the demand for multi-step synthesis, providing a direct and efficient method for constructing complex structures and dynamic processes. So far, many applications have been developed by using DCC and there is still much room for new developments. One of these involves the use of DCC to achieve the synthesis of complex folded structures. Although the principle has been proven many years ago, there have been no reports that go beyond proof of principle. In addition, since DCC has the ability to link multiple subsystems, another potential application is to merge subsystems with different characteristics to build new complex systems with more interesting dynamic behaviors and properties.
Chapter 1

Self-replication is one of the most important ingredients in the origin of life, and self-replicating molecules are a promising starting point for the de-novo synthesis of life. Folding is the process by which proteins and nucleic acid strands acquire a three-dimensional structure required for function. This thesis intends to capture the processes of self-replication and folding individually, or combined by using dynamic combinatorial chemistry. We first describe how to use dynamic combinatorial chemistry to construct complex folded structures. By making use of simple building blocks, we have achieved the selective assembly of remarkably complex folded structures. We also combined the processes of folding and self-replication in a single system, giving rise to the first such example in synthetic systems.

In Chapter 2, we describe how a dynamic combinatorial selection approach allows access to a foldamer of remarkable complexity constituted by 15 identical peptide-nucleobase building blocks. The folded structure has a complex secondary and tertiary structure and can emerge autonomously and spontaneously from a dynamic combinatorial library. Folding drives the highly selective (95% yield) synthesis of this remarkable stable folded structure from a mixture of interconverting molecules of different ring sizes in a one-step process. Structural characterization reveals that noncovalent interactions such as π-π stacking and hydrogen bonding play a key role in the stabilization of this complex molecule.

Following the results in Chapter 2, we synthesized a series of new building blocks by replacing the peptide-nucleobase motif with a simple dipeptide subunit. In Chapter 3, we describe a family of complex folded structures that emerged from dynamic combinatorial libraries made from these building blocks. Like the peptide-nucleobase foldamer described in Chapter 2, the formation of dipeptide foldamers is also spontaneous and selective. The introduction of hydrogen bonding sites on the structure of the dipeptide building blocks can significantly affect the formation of folded structures, leading to the emergence of macrocycles of different sizes (9-23mers).

Having established the method for the selective formation of complex folded structures from dynamic combinatorial libraries, we tried to operate the processes of folding and self-replication in a single system. In Chapter 4, a new dynamic combinatorial library is set up by mixing the building block that is capable of forming folded structures with another building block that can undergo self-replication. The results show no self-sorting between self-replicator and foldamer in this mixed system. Instead a series of mixed-building-block self-replicators was obtained. Unlike other nucleic-acid self-replicating systems that rely on base pairing, the emergence of the new self-replicators depends on the ratio of the building blocks and the assembly into ordered supramolecular...
Folding and replication in complex chemical systems

nanostructures. In addition, selective auto- and cross-catalysis in multicomponent systems are also described in this chapter.

In Chapter 5, we further explored the possibility of self-sorting of self-replicators and foldamers in mixed building block systems. By simply changing the structure of the peptide building block, the processes of self-replication and folding were found to occur simultaneously in a single dynamic library. The results show that the emergence of self-replicator or foldamer is determined by the ratio of the building blocks that make up the dynamic library. The emergence of the self-replicator can promote the formation of the complex folded structures. Furthermore, transient formation of a self-replicator can also be achieved by adjusting the ratio of the building blocks.

Finally, Chapter 6 gives a summary of this thesis and places the results in a broader perspective.

1.6 References

Chapter 1


Folding and replication in complex chemical systems


(84) Ferretti, A. C.; Joyce, G. F. Biochemistry 2013, 52, 1227.


(99) Hong, J. I.; Feng, Q.; Rotello, V.; Rebek, J. Science 1992, 255, 848.


Chapter 1


