PERSPECTIVES

Is the transition from chemistry to biology a mystery?

Hans Kuhn

Abstract

Today most chemists think that the answer to how life on earth emerged is still unknown. They assume a gap between chemistry and biology that is still unbridged. For chemists, understanding the origin of life requires the experimental modeling of a process that bridges this gap. They will not consider the problem solved before they are able to perform such tasks. No gap appears when we are pursuing a less ambitious goal, namely, to present a sequence of hypothetical processes that lead to an apparatus with the basic structure and fundamental feature of the genetic apparatus of biosystems but strongly simplified. The modeled apparatus has the basic machinery of living entities. Its fundamental feature is Darwinian behavior. Living individuals have the power to evolve toward ever increasing complexity and intricacy if appropriate conditions are given. The task to understand life's origin as a rational process is closely related to the earlier attempts of the present author to design and construct supra-molecular machines. The skill of the experimentalist has to be replaced by the presence of very particular conditions given by chance in a very particular location. The resulting apparatus has a distinct basic structure and function. The essence of what happens is inevitable, not accidental. Thus the emergence of life is assumed to be described by a distinct theory. Today's great challenge is experimentally investigating chemical systems with the goal of creating artificial chemical life and the given theory provides a powerful stimulus. Life, from the perspective of physics, is the living state of matter and this view calls for a theory describing the fundamental requirements for the appearance of such a living state of matter (on the early earth and in the universe). The approach given here is an attempt in this direction. According to that approach the appearance of an entity with Darwinian behavior is instantaneous and linked with the creation of matter that carries information. Thus, Information (measured in bits according to Shannon) takes a meaning with that instant, the appearance of the first entity that evolves by multiplication, variation, selection and keeps that meaning during the entire evolution of the living (Information carrying) state of matter. Another consequence of this initial event is a spontaneous symmetry breaking due to the equal probabilities that the oligomer starting the process is right handed or left handed.

1. Introduction

The early assumption, that life's emergence is related to the second law of thermodynamics, has its case as follows. Eugene Wigner [1] believed to have demonstrated that the origin of life is in contradiction to the second law. Ilya Prigogine and Peter Glansdorff [2] showed that a homogeneous system in a steady state far from equilibrium is unstable and in Manfred Eigen's view [3] this fact is an essential thermodynamic principle in understanding the origin of life bridging the gap between chemistry and biology. Today it is broadly assumed that

"the central issue regarding the very essence of living systems remains unresolved" (Addy Pross [4]). "The center is still mysterious" (Stuart Kauffman [5]). "Most chemists believe, as I do, that life emerged spontaneously from a mixture of molecules in a pre-biotic Earth. How? I have no idea" (George Whitesides [6]) and "We must seek out principles that would explain the transformation of inanimate and animate. And it is on this most fundamental of issues that confusion and controversy stubbornly remain." (Günter Wächtershäuser [7])

1.1 Paradigm: invent sequence of logically consistent hypothetical chemical reactions

This paper is based on another early assumption: life originates as a sequence of distinct chemical reactions.

Correspondence: h-k@netplus.ch
Director Emeritus of the Department Molecular Systems Assembly, Max-Planck-Institute for Biophysical Chemistry, (Karl Friedrich Bonhoeffer Institute), D-37077 Goettingen-Nikolausberg, Germany

Chemical reactions, per se, are in agreement with thermodynamics. Finding a rational approach to understand life's origin is searching for a sequence of chemical reactions that lead to first-life-like entities, thus entities with early-life-like genetic apparatus. It is a particular process consisting of an immense number of causally interlocked steps. This process has been theoretically modeled by strongly simplifying what might be the actual process ([8] and later developed [9-18]). No unbridged gap appeared. The resulting genetic apparatus has the principle structure and the fundamental properties of life's genetic apparatus. In particular, the theoretically modeled entities have the power of evolving toward ever increasing complexity and intricacy if appropriate conditions are given. This is the particular property of life on earth.

1.2 Support theoretical model by experiment

In an early given scenario [8] it was assumed that the origin of life began by enzyme-free replication of a homochiral RNA oligomer that came into being in a heterochiral mixture of nucleodides in a most seldom oligomerization process occurring by chance. This initial template has D or L chirality with equal probability. Either D or L originated. This is a process of spontaneous symmetry breaking determining the succeeding evolution. All events leading to life's origin are driven by distinct environmental conditions.

It has been shown by Gerald Joyce et. al. [19] that a pair of two ligase-ribozymes do exponentially replicate indefinitely. This is in favor of the given scenario.

Albert Eschenmoser et. al. [20] have demonstrated in model experiments how symmetry breaking in oligonucleotides can occur. They demonstrated chiroselectivity in copolymerization forming pyranosyl-RNA from pyranosyl G-C-C-GcP tetramers (where cP means cyclophosphate). The tetramers are heterochiral and the ligation process leads to selection of those tetramers that are homochiral, and exclude all the heterochiral tetramers through their property of not being able to participate in the template-directed ligation efficiently.

John Sutherland et. al. [21] have demonstrated a way to obtain activated pyrimidine ribonucleotides. Beginning with glycolaldehyde and guanidine a condensation product A is obtained in a distinct sequence of reactions. In a second independent step this product reacts with glycerinealdehyde forming a product B. It reacts in another separate step with cyanoacetylene to a product C that reacts with phosphate forming cytidine-2,3-cyclophosphate. This may be imagined to take place in nearby locations where glycolaldehyde and guanidine, glycerinealdehyde, cyanoacetylene are accumulated respectively. A, B and C are imagined to reach the corresponding locations by flow or diffusion. Prebiotically

reasonable chemical conditions allowing such a situation are unknown and it is a challenge to find any experimental approach. Guenter von Kiedrowski et.al. have experimentally demonstrated replication of a short DNA analogue driven by cycles of environmental changes [22].

1.3 Theory: postulates and logical consequences

Here we focus on fundamental aspects of life's origin that have not been emphasized in earlier papers. The aim is to conceptualize a theory. A theory brings experimental facts in a rational order by proposing a set of assumptions (postulates). In the present case the task is to show by logical considerations how the genetic apparatus-the basic machinery of life-might have originated. In a theory, once the postulates are formulated, the task is to find their logical consequences. This is quite different from guesses and intuitions.

The genetic apparatus is the device consisting of DNA that replicates, catalyzed by the polymerase and carries the message to produce proteins. This is performed by synthesis of messenger RNA (using transcriptase) and by translating the message given by the sequence of nucleotides to the message given by the sequence of amino acids in proteins. This is performed by transfer RNA and is assisted by the ribosome.

The genetic apparatus is such a complex machinery that a theory, today, must be restricted to the task of bringing the basic features into a rational order. That is, to identify the causes why the genetic apparatus could have emerged and why basically similar devices are highly probable to emerge anywhere in the universe where appropriate conditions are present. It is assumed that appropriate energy-rich monomers and appropriate environmental conditions are given. These assumptions are set as postulates.

Another fundamental postulate is the assumption that distinct and distinctly changing environmental conditions are present driving the evolutionary process. This postulate is important to render possible the processes, which otherwise and in other views on life's origin are considered to be impossible.

• *Explanatory note on difficulties understanding origin of life*. Difficulties are avoided by the given postulate. For example, a difficulty is frequently seen in experimentally reaching an exponential growth in replication. This difficulty does not appear if the templates and monomers are exposed to a distinctly periodically changing environment inducing and driving replication and selection. Other frequently seen difficulties are the lack of specific selection, the accumulation of too many sequences with similar efficiencies and no clear distinctions between losers

and winners. According to the present postulate only those entities can survive that are adapted to the environment. Adaptation is the result of being selected in a Darwinian process. Small changes can lead to highly specific adaptations, for example the colonization of a region not accessible to the previous existent forms. Therefore, the considered difficulties do not arise in such a scenario.

Such specific conditions are assumed to be given at a very particular location on the early earth. They drive the formation of a simple machine that causes by a Darwinian process the occurrence of a more complex machine and so on. The task is to find by logical considerations conditions that lead in an immense number of distinct steps to a bio-like genetic apparatus. This process is a kind of molecular engineering, where the engineer is replaced by nature's very particular conditions given by chance at a very particular location. Postulating distinct and diversified environmental conditions is fundamental to drive evolution to increasing complexity. Areas with less and less appropriate conditions are colonized with more and more complex entities each having the capacity to get established in the corresponding area.

There is no indication to look at the origin of life as an unknown, still mysterious process. Of course, developing a more detailed rational description of how life may originate should be helpful in performing experiments focused on theory. Such experiments would be very important in expressing more detailed postulates and in this way theory would further develop. The interplay between theory and experiment would be a great challenge.

• *Explanatory note on consistency in theoretical modeling.* An important feature in modeling life's emergence is illustrated here and discussed as a logically consistent process below. Let us consider a scenario beginning with G and C nucleotides separated from contaminations by very specific environmental conditions. This assumption is experimentally supported, indicating that particular conditions at a particular location are required in a prebiotic synthesis of nucleotides (see section 1.2). Peter Strazewski [23] supports this by demonstrating the great importance to search for precise and particular conditions that allow oligonucleotides to concentrate from high dilution onto a two-dimensional interface. Among a great number of oligonucleotides an oligo-nucleotide may occur by chance that is replicable and multiplies initiating a Darwinian behavior. By linking such oligomers further oligomers of larger size are created that aggregate. This requires compartmentation,

e.g. small pores in a porous material. Monomers diffuse into the pores while the more expanded oligomers and aggregates cannot easily leave the pores. In the course of time aggregates appear, that act as catalysts linking amino acids. Oligo-peptides associate building an envelope. A strong evolutionary driving force is present toward independence of the porous region and getting adapted to strongly diversified environmental conditions. Entities of increasing complexity emerge. The envelope evolves to an increasingly refined cell membrane including lipids whereas inside the cell an increasing number of enzymes of increasing sophistication develop. In any scenario attempting to understand the origin of life each step must be a consequence of the preceding step. The construction of lipid vesicles containing replicating RNA oligomers is a very exciting experiment [24-27], but having a lipid cell membrane already in the beginning of replication the evolutionary driving force as described above is missing.

1.4 The very basic steps in the given theory

As already mentioned, evolutionary steps are initiated, sooner or later, by the accidental occurrence of an entity with new functionality due to some mismatch in replication. Particular steps inducing an important evolutionary process are inevitable. Thus, in the essence, the evolutionary process is not accidental. Therefore it can be described by a distinct theory. This is attempted in this paper.

1.4.1 First Darwinian behavior

Let us consider the first entity with Darwinian properties, that is, an entity determined by causality but seemingly behaving as if it would have a goal, namely the goal to survive as a species by reproduction and to evolve by variation and selection in the particular environment driving the process. This property is not present in any ancestral form in a pre-biotic world. The Darwinian behavior appears instantaneous by chance when the first entity appears that fulfills the required conditions. We define this ultimate process as "emergence of life". The descendants evolve to increasingly complex and intricate forms related to an increasing precision in reproduction.

1.4.2 First occurrence of an entity that has the capacity of ever increasing complexity if appropriate conditions are given (this is the particular property of all life on earth)

The early process described in section 1.4.1 comes to an end at a certain precision in replication. A further precision cannot be reached with the existent kind of genetic device. A fundamental change in the machinery is required to further evolve. In a distinct immense number of steps a situation is finally reached by the

occurrence of a first entity that allows an evolution to ever increasing complexity provided that appropriate environmental conditions are given. This entity constitutes a primordial form with the particular property of all life on earth. The emergence of such a primordial form is what we define as "emergence of complex life". Its initial occurrence may be called the transition from chemistry to biology. It is reasonable to assume that such entities may evolve, again in an immense number of distinct steps, to entities with biogenetic apparatus. A similar transition can be considered as the process that leads to higher forms of life in the universe.

2. Origin of First Darwinian Entities
2.1 First Fundamental Step. Replicable Oligomer
Can a replicable oligomer appear with reasonable probability? We attempt an estimate by theoretical modeling of simple examples.
2.1.1 Correct linking of monomers
Let us consider de-novo formation of a homo-chiral oligomer in a hetero-chiral solution of two kinds of monomers that are complementary. The environmental conditions are such that a replication takes place if a replicable oligomer appears by chance. To be specific in modeling the process, we assume de-novo formation of oligomers consisting of 11 monomers of uniform chirality (say L). The reason for choosing 11 monomers is given below. The probability of linking two monomers such that a replicable oligomer can be formed is assumed to be 0.01.

> • *Explanatory note on the probability of linking two monomers correctly.* This estimate is obtained assuming a 50% probability that a monomer is of the correct kind (not a contaminant; a contaminant is a molecule, similar to a monomer, that can replace the monomer in de-novo forming the oligomer), and assuming a 50% probability that it links correctly (at the 3' position in a nucleotide). The probability that the monomer has chirality L is 1/2. Thus the probability that the monomer is correct and correctly bonding is 1/8. The probability that a correct bond is formed (at 5' position of the second partner in the case of a nucleotide) is $(1/8) * (1/8) \approx 0.01$. Our aim is only to test conditions required to allow occurrence of a replicable oligomer, we do not intend to give exact numbers.

The monomers must be of the correct kind (assuming a contamination by similar monomers), homochiral and correctly linked. This is a requirement to allow precision in replication. The double stranded region forming in the replication process must be well stabilized by complementary interlocking. The question then arises: how many oligomers consisting of 11 monomers must be formed de-novo until a replicable oligomer actually appears? The probability of de-novo formation of a replicable oligomer (10 links, L-chirality) is $0.01^{10} = 10^{-20}$. A replicable oligomer with D-chirality appears with the same probability. Thus, on the average, $(1/2) * 10^{20}$ oligomers de-novo appear, until a replicable oligomer comes into existence. This accidentally occurring replicable oligomer is either right handed or left handed. A symmetry breaking is induced determining further evolution.

2.1.2 Replicating oligomers can die out by chance
The probability of the replicating form to die out in the critical initial phase of multiplication by statistical reason should be about 0.9.

> • *Explanatory note on the probability of the replicating form to die out.* Assume a periodic change between replication phases (where only one replication of each oligomer takes place) and phases of selection where the probability for a single oligomer to survive is 70% and the probability to die is 30%. Then at the end of the first period the probability to have two oligomers is 0.7 * 0.7 = 0.49, the probability to have one oligomer is 0.7 * + 0.3 * 0.7 = 0.42, and the probability that the oligomer died out (all copies died) is * 0.3 = 0.09. The probability of the oligomer to die out in the second period is smaller and with increasing number of periods it is increasingly less probable that all copies of the oligomer die during a period. The probability that all copies finally die out is a few times 0.09, not larger than 0.9.

Then the probability that the descendants overcome that phase and continuously multiply is 0.1. On the average, $10^{21} = 2^{70}$ oligomers emerge de-novo until a replicating oligomer appears and is successful.

> • *Explanatory note on overcoming the critical initial multiplication phase.* When the probability for overcoming the critical initial multiplication phase drops from 0.1 to 0.01 (99% 'mortality') then $10^{22} = 2^{73}$ oligomers need to emerge de-novo before the first successfully replicating oligomer appears.

2.1.3 Condition for beginning replication
It is reasonable to assume that the particular environmental conditions required to drive a continuous replication of oligomers are given in a small location, say in a volume of a cm^3 or a mm^3. Can the process begin and continue with reasonable probability? Based on the present assumptions the probability to find a correct

oligomer in a cm^3 is 10^{-5}. Assuming disturbing contaminations being absent, a single correct oligomer is expected to be in a mm^3.

• *Explanatory note on the probability of initiating a Darwinian process*. Let us consider a solution of monomers of 10^{-3} mole/Liter in a porous material. We assume that conditions are given that allow replication of an appropriate oligomer in a volume of one cm^3. In this case

$6 * 10^{23}$ (monomers/mole) * 10^{-3} (mole/Liter) * 10^{-3} Liter = $6 * 10^{17}$ monomers are present forming $6 * 10^{16}$ oligomers. The probability that a de-novo originating oligomer is replicable and multiplies is 10^{-21} according to section 2.1.2. The probability to find such an oligomer in one cm^3 is

$10^{-21} * 6 * 10^{16} = 6 * 10^{-5}$. Thus the situation initiating Darwinian evolution is given acceptably often. If contaminations are absent the density of correct oligomers is much higher. The probability of correct bonding is $(1/4) * (1/4) \approx 0.06$ instead 0.01. Thus, the probability of appearance of a correct oligomer is $0.06^{10} = 6 * 10^{-13}$ instead of $0.01^{10} = 10^{-20}$. Then the density of correct oligomers is $6 * 10^{-13}/10^{-20} = 6 * 10^7$ times larger than with contaminant. There are $10^{-5} * 6 * 10^7 = 600$ correct oligomers in 1 cm^3 or about one correct oligomer in a mm^3.

2.1.4 Required beginning with short oligomers
Why do we assume oligomers of 11 monomers? Replicable strands must be sufficient in length to be well interlocked in forming double strands. On the other hand the probability to obtain a correctly linked oligomer decreases exponentially and we can neglect de-novo formation of longer correct strands.

• *Explanatory note on the length of replicating oligomer formed de-novo*. Assuming oligomers of 21 instead of 11 monomers: the probability that a successful de-novo oligomer appears is $0.01^{20} * 0.1 = 10^{-41}$ instead of 10^{-21}. Assuming 15-mers instead of 11-mers the probability that a successful de-novo copy appears is 10^{-29} instead of 10^{-21}. This is unreasonable. It appears as crucial to begin with short oligomer strands.

2.2 Variation and selection
2.2.1 Variation of oligomers
In a solution of monomers and replicable oligomers a multiplication of oligomers takes place. By mismatch in the replication process (incorporating an equal monomer in the replicate instead of a complementary monomer) the sequence of the two kinds of monomers will gradually change, a statistical mixture of oligomers with different sequences (among the 2^{11} possibilities) will be generated.

2.2.2. Selection of oligomers
Let us focus on a region (close to the region already considered) where only oligomers of distinct sequences of monomers can survive, for instance by their particular folding, protecting them from hydrolysis or binding them to a given surface. Such a region might be colonized by diffusion of accidentally occurring oligomers of appropriate sequence of monomers. A selection of oligomers accidentally adapted to the environment takes place resulting from mismatches in replication. Thus, little by little, more and more neighboring regions are colonized.

2.3 Probability of life
George Whitesides's opinion mentioned in the beginning of chapter 1 is based on a detailed consideration of difficulties in trying to approach the origin of life [28]. He strongly emphasizes on the importance "to describe life in physical terms-that is, we will rationalize life satisfactorily in molecular detail based on accepted scientific law and scientific theory using the scientific method" and "This problem deserves our most careful thought." In George Whitesides's words "The origin of life is one of the biggest of the big questions about the nature of existence". We both have the same view on the scientific approach to be distinguished from guesses and intuitions, why do we have such different opinions?

I had a very limited goal: using reasonable conditions as postulates and finding the consequence by systematic thinking, thus showing that a device evolves that has the basic properties and functionalities of the bio-genetic apparatus but is much simpler. It is using the "scientific method", setting reasonable assumption as postulates and finding such a device as the result of an intellectual proceeding. It appears as inevitable. It is reasonable to assume that the emergence of bio-systems is based on the same fundamental conditions as postulated in the given model. I have no hard proof for this generalization but I see no reason against this assumption. Once there is a working Darwinian device (a bio-like genetic apparatus) one can expect evolution of continuously increasing complexity.

3. The living state of matter. Information and Knowledge
3.1 Origin of life and information
A living individual, in the present context, is an entity that carries information. The entity has the property to evolve by multiplication, variation, and selection. It behaves as if it would have the know-how to survive as a species and to improve gradually in structural and functional complexity. The information to reproduce the

entity is transferred from one generation to the next generation and increases in the course of evolution. It is of great interest to have a quantitative measure for the information required for reproduction. According to Shannon [29] the "quantity of information is measured by the number of binary decisions".

In the example given in section 2.1 we considered oligomers consisting of monomers of two kinds (which are complementary to each other) in an arbitrary sequence. To construct replicating oligomers with 11 monomers of arbitrary sequence requires $10^{21} = 2^{70}$ de-novo forming oligomers, in average, until the first replicating oligomer appears. Thus, this oligomer carries 70 bits of information.

Information is linked to Darwinian evolution. It gets a meaning with the incidence of the appearance of the first self-reproducing entity. Information keeps that meaning during the entire evolution of the living state of matter that is a consequence of this initial event. Note that Information grows by multiplication, variation and selection in all stages in the evolution of life and in all that was leading to humanity's constructions such as language, writing, print and electronic means of multiplication, storage, developing human's ideas and realizing them physically. The evolution leading to today's motor-cars is an example.

3.2 Information and Knowledge

What is a measure of the quality of genetically transferred information, how to assess the functional intricacy of the evolving machinery? A large increase in the evolving system's functionality requires a strong effort, that is, it takes many generations until a mutant appears by chance, that brings a revolutionary change in functionality. Thus, a great number of bits must be discarded. For example, in the case considered in section 2.1 and 3.1 about 2^{70} oligomers are formed de-novo and are discarded until a replicating oligomer occasionally occurs by chance. That is, on the average, 70 bits are discarded. We call K = 70 bit the Knowledge of the replicating oligomer [9,11,12]. We define time t = 0 as the time when the first de-novo formed oligomer appears that multiplies, initiating a Darwinian process. At time t = 0, Knowledge K increases instantaneously from 0 to 70 bits in the given model. Shortly after, assuming that an oligomer with 11 monomers of a distinct sequence is selected (one of 2^{11} possible sequences), Knowledge K increases to 70 + 11 = 81 bits.

  • *Explanatory note on the value of the initial step of Knowledge K.* The value is strongly dependent on the probability of correctly linking two monomers. Assuming the unreasonable value 0.001 instead of 0.01 given above, we would obtain $10^{31} = 2^{K}$, K = 103 instead of $10^{21} = 2^{K}$, K = 70.

Then knowledge K increases stepwise in the course of evolution (figure 1). Small functional improvements are reflected by short phases of Knowledge K being constant followed by low steps in K, revolutionary improvements by long phases of K being constant followed by high steps [9].
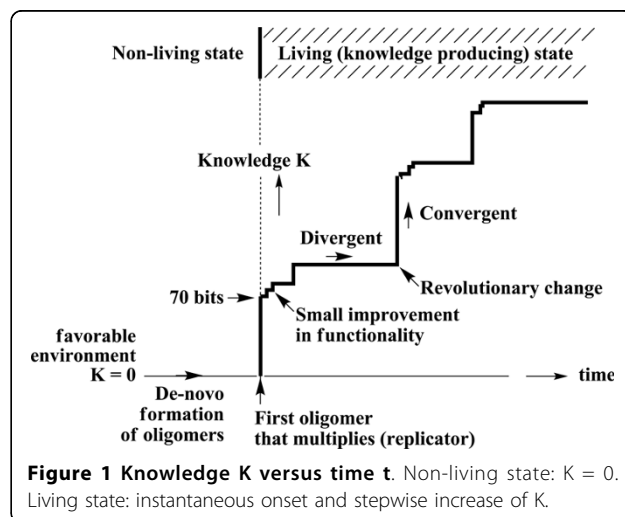
3.3. Entropy S and Knowledge K

We have emphasized that approaching to understand the origin of life is to search for a sequence of logically consistent and chemically reasonable hypothetical steps leading to a genetic apparatus. It is not a problem in thermodynamics. This should be emphasized since there is a remarkable formal similarity between the entropy S at temperature T and knowledge K at time t. [17]

  • Entropy of a body S is zero at temperature T = 0 and increases with T.
  • Knowledge K is zero at time t = 0 and it increases with t.
  • Entropy S increases at phase changes.
  • Knowledge K increases at changes of functionality.
  • Entropy S of an isolated system increases in time t.
  • Knowledge K of a living system increases in time t. (K has no meaning when environmental changes do not any longer allow life to exist)

4. Evolution leading to primordial genetic apparatus (PGA)

4.1 Construction of supra-molecular machines and the driving force in the emergence of the genetic apparatus

The scope in classical preparative chemistry is the synthesis of pure substances, single kinds of molecules. The proposed line of action in chemistry was seen in synthesizing different kinds of molecules purposefully

Figure 1 Knowledge K versus time t. Non-living state: K = 0. Living state: instantaneous onset and stepwise increase of K.

planned to interlock precisely in a distinct way when exposed to distinctly planned changing environmental conditions. The remote goal was fabricating useful complex supra-molecular devices suggested by molecular biology and any complex system arising from chemist's imagination. (For constructing prototypes see [30-32] for important recent work on the same topic named Interfacial Systems Chemistry see [33-39]). Having in mind the emergence of the genetic apparatus, the skill of the experimentalist is replaced by very particular periodically changing environmental properties leading to that outcome. Such conditions are assumed to be given by chance at a very particular location on the pre-biotic earth and elsewhere in the universe. Developing both fields requires thinking in terms of strongly simplifying theoretical models leading to suggestions how to construct supra-molecular machines on the one hand and how a bio-like genetic apparatus can evolve on the other hand. Seeing the mechanism in obtaining complex machines in life's emergence can be useful in future developments of supramolecular engineering.

The close relationship between (i) first experimental achievements to construct prototypes of supramolecular machines, (ii) thoughts on the origin of life, and (iii) today's Interfacial Systems Chemistry, is exciting. The required precision was early reached only by extending the Langmuir-Blodgett (LB) method [30-32,40]. First interfacial molecular devices were obtained in the authors research group [41] by applying Zisman's self-assembly-monolayer method (SAM) [42] and developed by Jacob Sagiv [43-45]. Today, many appropriately interlocking compounds are available to construct increasingly complex SAMs allowing to get interfacial systems that are three-dimensionally precisely ordered and porous. Up to 37 layers can be superimposed [33,46,47]. Most promising is the development of single molecule machines based on changes in conformation [48].

4.2 First molecular aggregate that acts as a supra-molecular machine

Bio-systems are complex aggregates of distinctly interlocking and functionally interacting molecules. Emergence and evolution of such systems is assumed to be initiated by the formation of simple aggregates and the continuous increase in complexity in a Darwinian process. How might the process begin?

Oligomers are assumed to evolve by increasing in length and increasing specificity in folding. Errors in replication produce more and more misfolded structures. This leads to a barrier in the evolution of oligomers at a critical length due to the unavoidable mismatches in replication. The barrier is overcome by selection of replication products that fit into an aggregate of precisely interlocking oligomers (figure 2)
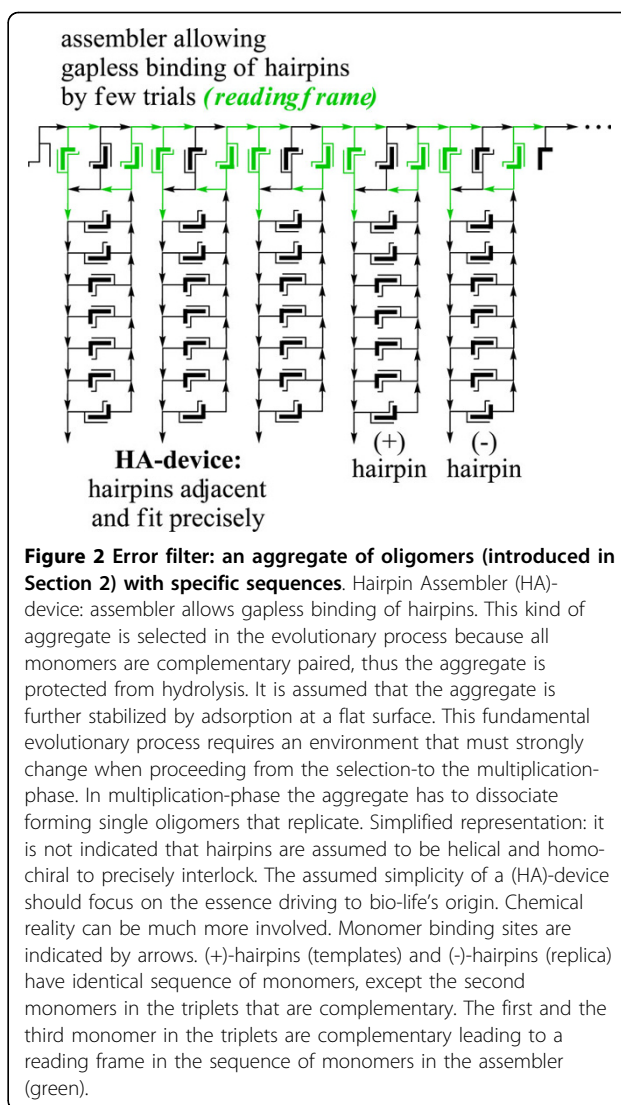

Figure 2 Error filter: an aggregate of oligomers (introduced in Section 2) with specific sequences. Hairpin Assembler (HA)-device: assembler allows gapless binding of hairpins. This kind of aggregate is selected in the evolutionary process because all monomers are complementary paired, thus the aggregate is protected from hydrolysis. It is assumed that the aggregate is further stabilized by adsorption at a flat surface. This fundamental evolutionary process requires an environment that must strongly change when proceeding from the selection-to the multiplication-phase. In multiplication-phase the aggregate has to dissociate forming single oligomers that replicate. Simplified representation: it is not indicated that hairpins are assumed to be helical and homo-chiral to precisely interlock. The assumed simplicity of a (HA)-device should focus on the essence driving to bio-life's origin. Chemical reality can be much more involved. Monomer binding sites are indicated by arrows. (+)-hairpins (templates) and (-)-hairpins (replica) have identical sequence of monomers, except the second monomers in the triplets that are complementary. The first and the third monomer in the triplets are complementary leading to a reading frame in the sequence of monomers in the assembler (green).

[10,11,13,14,16]. Misfolded oligomers do not interlock and disappear (error filter). The effective reproduction error probability is strongly diminished.

• *Explanatory note on the emergence of a HA-device. Evolution of entities that are protected from hydrolysis.* Let us consider G and C like nucleotides enclosed in a pore with positively charged flat walls. A periodically changing temperature causes an alternation between multiplication phases and selection phases. Oligomers evolve that are best protected from hydrolysis during the selection phases and do not easily escape from the pore during the multiplication phases. The result of this evolutionary process is hairpins. They are best protected from hydrolysis by complementary pairing of monomers, and aggregates of hairpins will be attached to the wall by electrostatic forces. We assume that the aggregate is
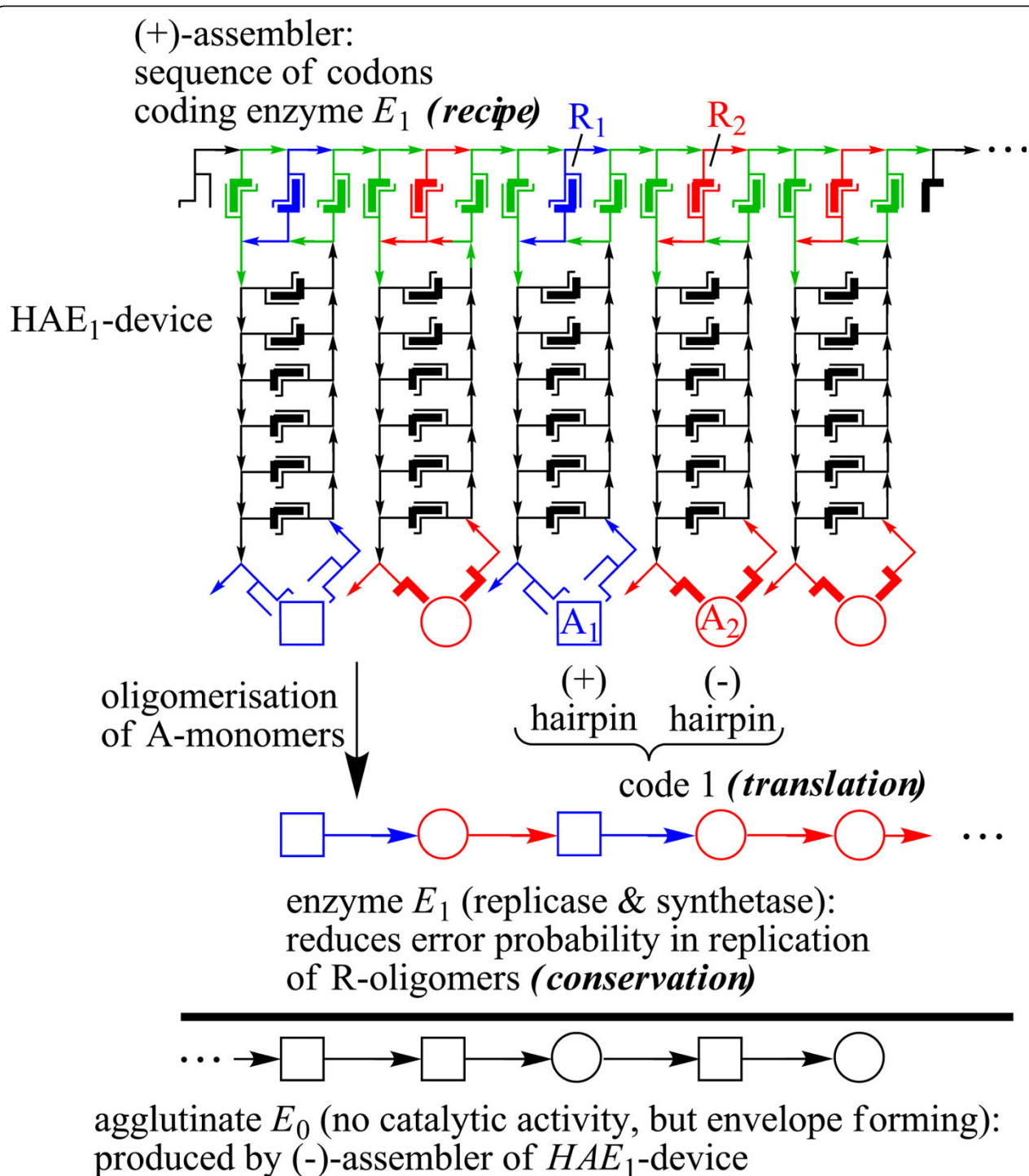
Figure 3 Translation device: a by-product of error filter (shown in figure 2). (+)- hairpin is linked to A1 and R1 (blue), (-)-hairpin is linked to A2 and R2 (read). The sequence of hairpins in the aggregate is determined by the sequence of monomers in the assembler. Stage where the sequence of monomers in the (+)-assembler is translated into enzyme E_1 (specific sequence of A-monomers). The sequence of monomers in the (-)-assembler is translated into E_0 (non-sense sequence). The figure illustrates theoretical modeling, chemical reality will be more sophisticated.

composed of descendants of an initially occurring hairpin, (+) hairpins and (-) hairpins, being complementary to each other (figure 2). Further protection from hydrolysis is reached by attaching an assembler strand. Such a strand may be formed by linking monomers attached to the "anti-codons" (triplets) of the hairpins by complementary binding. The sequence of the assembler allows then binding to hairpins as indicated in figure 2. For another mechanism leading to the formation of a HA-device and used in quantitative modeling see [14,16]. Thus, the assembler has a reading frame (let us say G is at the first position of the anti-codon of each hairpin and C is at the third position). The second position reflects the sequence of (+)-hairpins and (-)-hairpins (G and C respectively; note that GGC codes for glycin, GCC codes for alanine). This scenario might be much too simple to describe what actually happens but it may indicate principles important in life's origin that account in attempts to create artificial chemical life, the chemical synthesis of a simple translation device related to what the given scenario suggests.

4.3 The first translation machine

We discuss a later assumed evolutionary step than in Section 4.2. A translation device has evolved (exemplified, in our model, by a HAE_1-device, figure 3) [10,11,13,14,16]. An evolutionary stage is reached which is modeled by a strand (a (+)-assembler strand) that produces an enzyme E_1 by translation. This enzyme is assumed to have the property to diminish mismatches in replication. The replication product in the model, a (-)-assembler strand, forms HAE_0-devices with a translation product E_0 (with a non-sense sequence).

Initially E_0 is the only product of the HA-device and (with its random sequence) has no enzymatic activity but is capable of forming an envelope by agglutination. With the HAE_1-device coming into existence the HAE_0-device is an indispensable intermediate in the device's replication. With gradually increasing complexity of the evolving entity E_0 turns into a contaminant.

4.4 Evolutionary barrier by contamination. Overcoming barrier initiates emergence of a genetic apparatus

The increasing contamination with increasing complexity of the evolving system leads to a barrier (figure 4). The barrier is overcome when it is assumed that E_1 evolves to enzymes E_{RD} and E_{DR}. E_{RD} is an enzyme catalyzing formation of D-strands complementary to R-templates, and E_{DR} is an enzyme catalyzing formation of R-strands complementary to D-templates. It is assumed that D-strands do not form any HAE-devices (this is a reasonable assumption). Systems of cooperating HAE-devices forming less contaminant E_0 have better survival chances. This leads to the evolution of a cooperating HAE_{RD}-devices [10,11,13] (figure 5) and HAE_{DR}-devices. Contaminant E_0 is no longer produced.

4.5. From primordial genetic apparatus to bio-genetic apparatus

It has been shown by modeling that the given process initiates an evolution toward a bio-like genetic apparatus in a long sequence of reasonable steps [10-13]. In a step of particular importance an enzyme E_{DD} appears catalyzing replication of D-strands complementary to D-templates. D-strands become carriers of genetic information. E_{RD} disappears. This is the breakthrough of the machinery on which the mechanism of the genetic apparatus is based. Modeling the evolution of the genetic code on the basis of the given theoretical approach leads to a code that is in
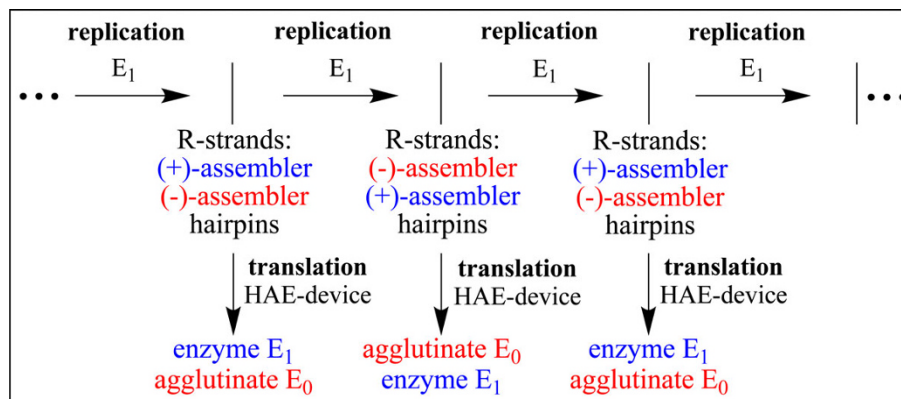
Figure 4 Evolution barrier: agglutinate (shown in figure 3) becoming a contaminant by accumulation. Stage where the (+)-assembler and (-)-assembler are translated into enzyme E_1 (specific sequence) and agglutinate E_0 (non-sense sequence) respectively. With increasing complexity the accumulating agglutinate E_0 becomes a contaminant, which leads to a barrier.
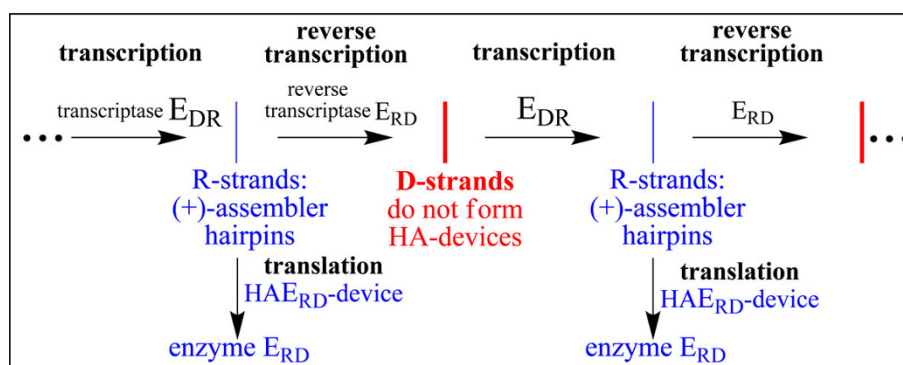
Figure 5 Overcoming evolution barrier (shown in figure 4). Stage where the (+)-R-strand's message is translated by the HAE$_{RD}$- device into enzyme E$_{RD}$ (specific sequence) while the D-strands do not form HAE-devices (as it is in biology: in contrast to RNA, DNA does not form a translation device). It cooperates with a similar device, where the translation product is enzyme E$_{DR}$ (not shown).

good agreement with the actual code [13,18]. This important part of the evolutionary process toward the biogenetic apparatus is not discussed in this paper that is focused on the first appearance of a device that has the capacity of a permanent increase in complexity. Modeling evolutionary processes that lead to the genetic code of bio-systems (linked to a corresponding increase in sophistication of the genetic apparatus) indicate that the essential mechanisms that effect life's origin are unfolding. However, a sequence of distinct, realistic, chemical reactions cannot be given. Experimental chemists have an important future task.

5. Summary

This paper is focused on what is seen as the decisive two steps in theoretical modeling the origin of life. That is the first appearance of an entity that has the capacity to evolve by multiplication, variation and selection, and much later the first appearance of a device that has the capacity of a permanent increase in complexity. These events are seen to be inevitable if appropriate environmental conditions are present that drive the process. Otherwise life cannot emerge. Accidental errors in reproduction are not determinants for what essentially happens. Thus the way of how life comes into being is determined by essential structural condition in space and time. Therefore, basic processes can be described by a theory postulating the required conditions. This is attempted in the present paper. Experiments focused on theory would be very important in expressing postulates in more detail. The improved theory should be helpful in suggesting experiments. This interplay should be important in developing the field. In the present view Shannon's information gets a meaning with the incidence of the appearance of the first self-reproducing and evolving entity.

References

1. Wigner EP: **Symmetries and reflections.** *Scientific essays of Eugene P. Wigner* Bloomington: Indiana University Press 1967, 200.
2. Prigogine I, Glansdorff P: **Non-equilibrium stability theory.** *Physica* 1970, **46**:344-366.
3. Eigen M: **Selforganization of matter and the evolution of biological macro-molecules.** *Naturwiss* 1971, **58**:465-523.
4. Pross A: **Seeking the chemical roots of Darwinism: bridging between chemistry and biology.** *Chemistry* 2009, **15**:8374-8381.
5. Kauffman SA: *Investigations* Oxford: Oxford University Press 2000.
6. Whitesides GM: **Revolutions in chemistry.** *Chem Eng News* 2007, **85**:12-17.
7. Wächtershäuser G: **The origin of life and its methodological challenge.** *J theor Biol* 1997, **187**:483-494.
8. Kuhn H: **Selforganization of molecular systems and evolution of the genetic apparatus.** *Angew Chem Int Ed Engl* 1972, **11**:798-820.
9. Kuhn H: **Model consideration for the origin of life. Environmental structure as stimulus for the evolution of chemical systems.** *Naturwissenschaften* 1976, **63**:68-80.
10. Kuhn H, Waser J: **Molecular self-organization and the origin of life.** *Angew Chem Int Ed Engl* 1981, **20**:500-520.
11. Kuhn H, Waser J: **Self organization of matter and early evolution of life.** *Biophysics* Berlin: Springer-VerlagHoppe W, Lohmann W, Markl H, Ziegler H 1983, 830-874.
12. Kuhn H: **Origin of life and physics: diversified microstructure-inducement to form information carrying and knowledge-accumulating systems.** *IBM Res Develop* 1988, **32**:37-46.
13. Kuhn H, Waser J: **A model of the origin of life and perspectives in supramolecular engineering.** *Lock-and-key principle* Chichester: WileyBehr J-P 1994, 247-306.
14. Kuhn C: **Computer-modeling origin of a simple genetic apparatus.** *PNAS* 2001, **98**:8620-8625.

15. Kuhn H, Kuhn C: **Diversified world: drive of life's origin?!** *Angew Chem Int Ed Engl* 2003, **42**:262-266.
16. Kuhn C: **A computer-glimpse of the origin of life.** *J Biol Phys* 2005, **31**:571-585.
17. Kuhn H: **Origin of life-symmetry breaking in the universe: emergence of homochirality.** *Current Opinion in Colloid and Interface Science* 2008, **13**:3-11.
18. Kuhn H, Waser J: **Hypothesis: on the origin of the genetic code.** *FEBS letters* 1994, **352**:259-264.
19. Lincoln TA, Joyce GF: **Self-sustained Replication of an RNA Enzyme.** *Science* 2000, **323**:1229-1232.
20. Bolli M, Micura R, Eschenmoser A: **Pyranosyl-RNA: chiro-selective selfassembly of base sequences by ligative oligomerization of tetranucleotide-2′,3′-cyclophosphates with a commentary concerning the origin of bio-molecular homochirality.** *Chem Bio* 1997, **4**:309-320.
21. Powner MW, Gerland B, Sutherland JD: **Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions.** *Nature* 2009, **459**:239-242.
22. Luther A, Brandsch R, von Kiedrowski G: **Surface promoted replication and exponential amplification of DNA analogues.** *Nature* 1998, **396**:245-248.
23. Strazewski P: **Adding to Hans Kuhn's thesis on the emergence of the genetic apparatus: of the Darwinian advantage to be neither too soluble nor too insoluble, neither too solid, nor completely liquid.** *Colloids and Surfaces B, Biointerfaces* 2009, **74**:419-425.
24. Mansy SS, Schrum JP, Krishnamurthy M, Tobé S, Treco D, Szostak JW: **Template-directed synthesis of a genetic polymer in a model protocell.** *Nature* 2008, **454**:122-125.
25. Zhu TF, Szostak JW: **A robust pathway for protocell growth and division under plausible prebiotic conditions.** *J Am Chem Soc* 2009, **131**:5705-5713.
26. Szostak JW: **Origins of life: Systems chemistry on early Earth.** *Nature* 2009, **459**:171-172.
27. Ricardo A, Szostak JW: **The origin of life on earth.** *Scientific American* 2009, **301**:54-61.
28. Whitesides GM: **The improbability of life. (Foreword).** *Fitness of the Cosmos for Life: Biochemistry and Fine-Tuning* Cambridge: Cambridge University PressBarrow JD, Morris CS, Freeland SJ, Harper CL 2008.
29. Shannon CE, Weaver W: **The mathematical theory of communication.** Urbana: University of Illinois Press 1949.
30. Kuhn H: **Versuche zur Herstellung einfacher organisierter Systeme von Molekülen.** *Verhandlungen der Schweiz Naturforsch Ges* 1965, 245-266.
31. Kuhn H, Möbius D: **Systems of monomolecular layers-assembling and physicochemical properties.** *Angew Chem Int Ed Engl* 1971, **10**:620-37.
32. Bücher H, Drexhage KH, Fleck M, Kuhn H, Möbius D, Schäfer FP, Sondermann J, Sperling W, Tillmann P, Wiegand J: **Controlled transfer of excitation energy through thin layers.** *Molecular Crystals* 1967, **2**:199-230.
33. Gölzhäuser A, Wöll C: **Interfacial Systems Chemistry: out of the vaccum-through the liquid-into the cell.** *Bunsen-Magazin* 2010, **12**:56-67.
34. Love JC, Estroff LA, Kriebel JK, Nuzzo RG, Whitesides GM: **Self-assembled monolayers of thiolates on metals as a form of nanotechnology.** *Chem Rev* 2005, **105**:1103-1169.
35. London G, Carroll GT, Fernández Landaluce T, Pollard MM, Rudolf P, Feringa BL: **Light-driven altitudinal molecular motors on surfaces.** *Chem Commun* 2009, 1712-1714.
36. Klok M, Boyle N, Pryce MT, Meetsma A, Browne WR, Feringa BL: **MHz Unidirectional Rotation of Molecular Rotary Motors.** *J Am Chem Soc* 2008, **130**:10484-10485.
37. Pantarotto D, Browne WR, Feringa BL: **Autonomous propulsion of carbon nanotubes powered by a multienzyme ensemble.** *Chem Commun* 2008, 1533-1535.
38. Kudernac T, van der Molen SJ, van Wees BJ, Feringa BL: **Uni- and bi-directional light-induced switching of diarylethenes on gold nanoparticles.** *Chem Commun* 2006, 3597-3599.
39. de Jong JJD, Browne WR, Walko M, Lucas LN, Barrett LJ, McGarvey JJ, van Esch JH, Feringa BL: **Raman scattering and FT-IR spectroscopic studies on dithienylethene switches–towards non-destructive optical readout.** *Org Biomol Chem* 2006, **4**:2387-2392.
40. Kuhn H, Försterling H-D, Waldeck DH: **Principles of Physical Chemistry.** Hoboken: Wiley, 2 2009.
41. Polymeropoulos EE, Sagiv J: **Electrical conduction through adsorbed monolayers.** *J Chem Phys* 1978, **69**:1836-1847.
42. Bigelow WC, Pickett DL, Zisman WA: **Oleophobic monolayers. 1. Films adsorbed from solution in non-polar liquids.** *J Colloid Sci* 1946, **1**:513-538.
43. Sagiv J: **Organized monolayers by adsorption. 1. Formation and structure of oleophobic mixed monolayers on solid surfaces.** *J Am Chem Soc* 1980, **102**:92-98.
44. Netzer L, Sagiv J: **A new approach to construction of artificial monolayer assemblies.** *J Am Chem Soc* 1983, **105**:674-676.
45. Zeira A, Chowdhury D, Hoeppener S, Shantang Liu, Berson J, Cohen SR, Maoz R, Sagiv J: **Patterned Organosilane Monolayers as Lyophobic-Lyophilic Guiding Templates in Surface Self-Assembly: Monolayer Self-Assembly versus Wetting-Driven Self-Assembly.** *Langmuir* 2009, **25**:13984-14001.
46. Shekhah O, Wang H, Kowarik S, Schreiber F, Paulus M, Tolan M, Sternemann C, Evers F, Zacher D, Fischer RA, Wöll C: **Step-by step route for synthesis of metal-organic frameworks.** *J Am Chem Soc* 2007, **129**:15118-15119.
47. Shekhah O, Wang H, Paradinas M, Ocal C, Schüpbach B, Terfort A, Zacher D, Fischer RA, Wöll C: **Controlled interpenetration in metal-organic frameworks by liquid phase epitaxy.** *Nature Materials* 2009, **8**:481-484.
48. Puchner EM, Gaub HE: **Exploring the conformation-regulated function of titin kinase by mechanical pump and probe experiments with single molecules.** *Angew Chem Int Ed Engl* 2010, **49**:1147-1150.