

## Title Page

Konstantin K Konstantinov, Alisa F Konstantinova

# Chiral Symmetry Breaking in Peptide Systems During Formation of Life on Earth

**Affiliation:**

A.V. Shubnikov Institute of Crystallography, Russian Academy of Sciences  
Leninskii pr. 59, Moscow, 119333, Russia.

**Corresponding author:** Konstantin K Konstantinov

E-mail: [kkk@jfcglobal.net](mailto:kkk@jfcglobal.net)

Phone: +1-514-467-5423

Fax: +1-866-669-5557

K. K. Konstantinov, A.F. Konstantinova

# Chiral Symmetry Breaking in Peptide Systems During Formation of Life on Earth

## Contents

Abstract.....	4
Keywords.....	4
Introduction .....	5
Definitions .....	6
"Closure" of models .....	7
Averaging of enantioselectivity in catalytic models .....	8
Models with peptides up to length 5, catalytic synthesis, and crystallization of insoluble pair .....	10
Reversible synthesis of amino acids .....	11
Catalytic synthesis of amino acids .....	11
Activation .....	12
Polymerization / Depolymerization .....	12
Sedimentation of insoluble pairs .....	13
Classes of considered models .....	14
Solution method .....	14
Symmetry breaking in the models with peptides up to length 5, catalytic synthesis of amino acids and simplified sedimentation .....	14
Symmetry breaking without catalytic synthesis .....	16
Symmetry breaking in the presence of weakly enantioselective forward catalytic synthesis .....	19
Symmetry breaking in epimerization based models .....	21
Symmetry breaking in APED models in the presence of sedimentation .....	26
Symmetry breaking in the models with peptides up to length 3, catalytic synthesis of amino acids, pair formation and Noyes – Whitney sedimentation .....	27
Proposed Experiment.....	27
Conclusion.....	29
References .....	30
Supplementary Materials.....	32

Source code..... 32

Main results of the article..... 32

Naming and Folder conventions..... 33

## Abstract

Chiral symmetry breaking in complex chemical systems with a large number of amino acids and a large number of similar reactions was considered. It was shown that effective averaging over similar reaction channels may result in very weak effective enantioselectivity of forward reactions, which does not allow most of the known models to result in chiral symmetry breaking during formation of life on Earth.

Models with simple and catalytic synthesis of a single amino acid, formation of peptides up to length five, and sedimentation of insoluble pair of substances were considered. It was shown that depending on the model and the values of the parameters, chiral symmetry breaking may occur in up to about 10% out of all possible unique insoluble pair combinations even in the absence of any catalytic synthesis. If weak enantioselective catalytic synthesis of amino acids is present, then the number of possible variants, in which chiral symmetry breaking may occur, increases substantially. It was shown that the most interesting catalysts have zero or one amino acid of “incorrect” chirality. If the parameters of the model are adjusted in such a way to result in an increase of concentration of longer peptides, then catalysts with two amino acids of incorrect chirality start to appear at peptides of length five.

Models of chiral symmetry breaking in the presence of epimerization were considered for peptides up to length three. It was shown that the range of parameters in which chiral symmetry breaking could occur significantly shrinks in comparison to previously considered models with peptides up to length two.

An experiment of chiral symmetry breaking was proposed. The experiment consists of a three-step cycle: reversible catalytic synthesis of amino acids, reversible synthesis of peptides, and irreversible sedimentation of insoluble substances.

## Keywords

Chiral symmetry breaking, Origin of life

## Introduction

The problem of chiral symmetry breaking during formation of life on Earth has been puzzling researchers for over 150 years starting from the discovery of enantiomers by L. Pasteur. The number of works in this area could be made a subject of a separate research. Vast number of models of chiral symmetry breaking were proposed over the years and numerous experiments were performed, for example (Frank 1953; Avetisov and Goldanskii 1996; Plasson et al. 2004; Saito and Hyuga 2004; Weissbuch et al. 2005; Brandenburg et al. 2007; Plasson et al. 2007; van der Meijden et al. 2009; Noorduyn et al. 2010; Blackmond 2011; Coveney et al. 2012; Hein et al. 2012; Ribó et al. 2013; Konstantinov and Konstantinova 2013, 2014), and many others. However, most, if not all, theories and experiments related to chiral symmetry breaking during formation of life on Earth considered only a very limited number of substances and reactions, whereas the actual number of possible substances and reactions among them must have been extremely large, as we will soon discuss. Therefore, it is a variety of both substances and reactions that substantially alters the picture.

There are only three possible ways how enantiomeric content of a closed chemical system can change over time. The chiral molecules can be created, destroyed, or converted into each other. Subsequently all models of chiral symmetry breaking have to utilize one or several of these channels. The models based on creation and destruction utilize reversible autocatalytic or catalytic synthesis of amino acids (Avetisov and Goldanskii 1996; Plasson et al. 2007) and in addition often require a “mutual antagonism” as was originally proposed by Frank (Frank 1953). Alternative models of chiral symmetry breaking are based on inter-conversion of enantiomers. These models can be grouped into two classes: the models based on epimerization (Plasson et al. 2004; Weissbuch et al. 2005; Brandenburg et al. 2007; Plasson et al. 2007) and the models based on the deracemization process (Hein et al. 2012; van der Meijden et al. 2009; Noorduyn et al. 2010).

In order to have chiral symmetry breaking, autocatalytic models (Avetisov and Goldanskii 1996) require substantial enantioselectivity of a catalyst. This can be called “fine-tuning” or described as needing a “robust system”. However, the abundance of various substances on a prebiotic Earth means effective averaging of the rates of catalytic reactions over all available catalysts for a given reaction and having the large value of enantioselectivity of the averaged rates seems questionable due to the law of large numbers. We will consider this in more detail below. In addition, it is often argued that amino acids do not possess autocatalytic properties. However, there are known catalysts for asymmetric synthesis of amino acids already at the dipeptide level (Iyer et al. 1996) and therefore it is important to extend catalytic models to longer peptide chains.

However, what is even more important is that some of the autocatalytic models mentioned above utilized incorrect assumptions about how catalysts change the reaction. Due to thermodynamic reasons (because the catalyst is not consumed), the presence of a catalyst cannot change thermodynamic equilibrium. Only the speed at which the equilibrium is reached is increased. This fact imposes very strict constraints on some of the parameters of catalytic reactions: total enantioselectivity, which is defined as a sum of forward and backward enantioselectivities, must be identically zero. This effectively eliminates all the models, which require any non-zero value of total enantioselectivity in order to result in chiral symmetry breaking. We will also consider this in more detail below.

Current epimerization based models (Activation, Polymerization, Epimerization, Depolymerization – APED) also have some fine-tuning requirements, which are similar to fine-tuning in catalytic and autocatalytic models. Certain parameters in such models must be placed far away from the expected average values (Plasson et al. 2004; Plasson et al. 2007).

On the opposite side, deracemization utilizes a phase transition process of crystallization and nonlinearity due to the rate of crystallization being proportional to the surface area of the crystals (Hein et al. 2012; van der Meijden et al. 2009; Noorduyn et al. 2010). In contrast to basic chemical reactions, which are polynomial functions of concentrations, a phase transition, if viewed as a mathematical function, exhibits some discontinuity. It is therefore not subject to the averaging considerations above, because a small change in the parameters near a phase transition point results in substantial changes in the output. However, deracemization requires a high racemization rate in the solution and enantiopure substance(s) accumulate in the sediment. While it is relatively easy to imagine how this process could have been realized on a geological scale, it poses some difficulties to chiral symmetry breaking on a prebiotic Earth. A high racemization rate in the solution means that most, if not all, complex

molecules, which could have been formed in the solution, would have had a racemic content. Rather, an enantiopure sediment must be taken away and only then could complex enantiopure substances be formed. In addition, deracemization also requires that the enantiomer substances crystallize as conglomerates, and there is only about 5-10% of amino acids, which exhibit that property. That means that, if the number of substances in the solution is very large, then a substantial number of substances will not form conglomerates. Subsequently, that results in sediment containing an enantiopure mixture of some substances and a racemic mixture of a far more substantial amount of some other substances.

From the third side, it was experimentally shown in (Blackmond 2011) that different properties of enantiomers and, in particular, formation of sediment of insoluble diastereomeric pairs may result in substantial enantiomeric excess (sometimes called chiral polarization) even in the absence of any autocatalysis. The application of this process to autocatalytic reactions was studied theoretically in (Konstantinov and Konstantinova 2013) where it was shown how the differences between physical properties of diastereomers may result in additional enantioselective factors. However, that is not sufficient due to the absence of autocatalytic properties of amino acids.

In addition, Earth is not a thermodynamically closed system as it constantly receives light (high frequency photons) from the Sun and emits heat (low frequency photons) into the outside space. This is most likely the most important factor influencing chiral symmetry breaking on Earth as it places its chemical system far from equilibrium state. The importance of far from thermodynamic equilibrium states for chiral symmetry breaking was considered previously in (Plasson et al. 2007) and they require a constant inflow of energy to operate.

It is therefore important to consider if there are models, which correctly treat thermodynamic constraints on the properties of catalysts, do not encounter fine-tuning issues in case of a very large number of substances, consider far from thermodynamic equilibrium systems, and take into account that amino acids may form longer peptide chains.

## Definitions

Current work considers large number of complex models, which, in turn, have substantial number of parameters and /or various definitions. The main parameters and definitions used throughout the article are as follows:

- $\rho_X$  is the concentration of a substance  $X$ .
- $\hat{E}(X)$  is the enantiomer of some chiral substance  $X$ .
- $\theta_X$  and  $\Delta_X$  variables. For any chiral substance  $X$  we denote its enantiomer as  $\hat{E}(X)$ , their concentrations as  $\rho_X$  and  $\rho_{\hat{E}(X)}$  correspondingly, and introduce a change of variables:

$$\begin{aligned}\theta_X &= \rho_X + \rho_{\hat{E}(X)} \\ \Delta_X &= \rho_X - \rho_{\hat{E}(X)}\end{aligned}\tag{1}$$

- We consider models with only one achiral ("food") substance, which we denote as  $Y$  and its concentration as  $\rho_Y$  and only one left amino acid, which we denote as  $A$  and its enantiomer as  $a$ .
- Letter  $C$  is used to denote a catalyst.
- When we consider peptide chains, we will use capital letters for left amino acids and small letters for right amino acids. So, for example:

$$\hat{E}(AAaA) = aaAa\tag{2}$$

where  $A$  is some left amino acids, and  $a$  is its enantiomer (the letter is not related to standard one letter amino acid coding).

- For any peptide chain  $P$ , for example,  $P = AAaA$ , we denote  $\hat{I}(P)$  as an inverse of  $P$ , which is a peptide chain in a reverse order, for example:

$$\hat{I}(AAaA) = AaAA \quad (3)$$

Some of the models, which we considered, are invariant to this inversion transformation of the equations.

- $\rho_A^{tot} = \sum_s n_A(s) \rho_s$ ,  $\rho_a^{tot} = \sum_s n_a(s) \rho_s$ , where  $n_A(s)$  is a number of left amino acids in a given substance  $s$ ,  $n_a(s)$  is a number of right amino acids in a given substance  $s$ ,  $\rho_s$  is the final concentration of substance  $s$ , are the total equivalent concentrations of left and right amino acids,
- $\eta = \left| \frac{\rho_A^{tot} - \rho_a^{tot}}{\rho_A^{tot} + \rho_a^{tot}} \right|$  is the enantiomeric excess,
- $\rho_0 = \rho_Y + \rho_A^{tot} + \rho_a^{tot}$  is the integral of motion and it describes the total amount of matter (per unit of volume) present in the system.
- For any reaction  $R$  we denote  $k(R)$  as the rate of such a reaction, for example  $k(A + C \rightarrow Y + C)$  is the rate of forward part of reaction  $Y + C \rightleftharpoons A + C$ .
- $c_+ = k(Y \rightarrow A) = k(Y \rightarrow a)$ ,  $c_- = k(A \rightarrow Y) = k(a \rightarrow Y)$  are the rates of forward and backward amino acid synthesis reactions.
- $a_+ = k(A \rightarrow A^*) = k(a \rightarrow a^*)$ ,  $a_- = k(A^* \rightarrow A) = k(a^* \rightarrow a)$  are the rates of forward and backward activation reactions (when used).
- $\Gamma_+ = \left( k(Y + C \rightarrow A + C) + k(Y + \hat{E}(C) \rightarrow A + \hat{E}(C)) \right)$  is the total rate of forward catalytic synthesis of amino acid reaction in the presence of catalyst  $C$  and its enantiomer  $\hat{E}(C)$ .
- $\Gamma_- = \left( k(A + \hat{E}(C) \rightarrow Y + \hat{E}(C)) + k(A + C \rightarrow Y + C) \right)$  is the total rate of backward catalytic synthesis of amino acid reaction in the presence of catalyst  $C$  and its enantiomer  $\hat{E}(C)$ .
- $\gamma_+ = \frac{k(Y+C \rightarrow A+C) - k(Y+\hat{E}(C) \rightarrow A+\hat{E}(C))}{k(Y+C \rightarrow A+C) + k(Y+\hat{E}(C) \rightarrow A+\hat{E}(C))}$  is the enantioselectivity of forward catalytic synthesis of amino acid.
- $\gamma_- = \frac{k(A+\hat{E}(C) \rightarrow Y+\hat{E}(C)) - k(A+C \rightarrow Y+C)}{k(A+\hat{E}(C) \rightarrow Y+\hat{E}(C)) + k(A+C \rightarrow Y+C)}$  is the enantioselectivity of backward catalytic synthesis of amino acid.
- $\Lambda_+$  is the rate of forward ligation reactions, for example:  $A + aa \rightarrow Aaa$ . The same value of  $\Lambda_+$  was used regardless enantiomeric content and length of peptides.
- $\Lambda_-$  is the rate of backward ligation reactions, for example:  $Aaa \rightarrow A + aa$ . The same value of  $\Lambda_-$  was used regardless enantiomeric content and length of peptides.
- $\varepsilon_- = k(P_1 + P_2 \rightarrow nY)$ , where  $P_1$  and  $P_2$  are the substances (peptides), which form an insoluble pair, and  $n$  is their combined length, is the rate of sedimentation of insoluble pair. Please, refer below to the discussion below why such "reactions" were used.
- $a = k(A \rightarrow A^*) \equiv a_+$ ,  $b = k(A^* \rightarrow A) \equiv a_-$ ,  $e = k(Aa \rightarrow aa)$ ,  $h = k(AA \rightarrow A + A) \equiv \Lambda_-$ , and  $p = k(A^* + A \rightarrow AA)$  are parameters of activation, polymerization, epimerization, depolymerization (APED) model. Letters  $a, b, e, h$  were chosen for describing the rates of reactions to match the terminology used in APED model.

## "Closure" of models

Consider some general isotropic and homogeneous model of chemical reactions. Suppose that we have  $N$  different chemical substances ( $S_1, \dots, S_N$ ) and  $K$  possible reactions among them ( $R_1, \dots, R_K$ ). The state of the system is determined by the number of molecules of each substance:  $\mathbf{X}(t) = (X_1(t), \dots, X_N(t))$  at a given time, or equivalently by their concentrations:

$$\boldsymbol{\rho}(t) = (\rho_1(t), \dots, \rho_N(t)) \quad (4)$$

where  $\rho_\alpha(t) = \frac{1}{\Omega} X_\alpha(t)$  and  $\Omega$  is a volume of the system and  $\alpha = 1, 2, \dots, N$ . Evolution of such a system under the assumption that  $X_i \gg 1$  can be written as a system of ordinary differential equations (Andrews et al. 2009):

$$\frac{d\rho_\alpha(t)}{dt} = \sum_{i=1}^K v_{\alpha i} a_i(\boldsymbol{\rho}(t)), \quad (5)$$

where  $v_{\alpha i}$  is a stoichiometric matrix and  $a_i$  is the rate of  $i$ -th reaction. In the general case  $a_i$  could be some complicated functions of  $\boldsymbol{\rho}(t)$ .

A system of equations (5) expressed in  $\rho_Y$ ,  $\theta_X$ , and  $\Delta_X$  variables will be called a **full system**. A **reduced system** is defined as a system (5) with all  $\Delta_X \equiv 0$  for all chiral  $X$ . Studying a symmetry breaking in a full system (or a bifurcation from a racemic steady-state) is equivalent to studying the eigenvalues of a matrix of coefficients of all  $\Delta_X$ . This matrix is obtained by linearizing a full system near stable fixed point(s) of a reduced system. This linearization effectively makes many models (for example, pass-through models and cyclic models) equivalent or nearly equivalent to each other, provided that the steady state concentrations do not change under such closing. For example, consider a pass-through model where some achiral substances flow into the system at some rate and some substances flow out of the system. A good example is a system of a geological scale, in which a sediment constantly forms and falls to the ocean floor, thus effectively leaving the reaction area. Suppose that the system is in some stable fixed point of a reduced system and we are interested in the symmetry breaking in a full system. We can "close" such a pass-through model and make it a cyclic model by adding fictitious irreversible decay reactions with the properly chosen rate coefficients, so that all output (and effectively non-chiral) substances decay into input (non-chiral) substances. The linearized matrix of coefficients of  $\Delta_X$  for such a completed cyclic model will be the same as for a pass-through model, from which we started. This can be illustrated by the following example. Consider two irreversible reactions, which terminate a pass-through model with some rate coefficient  $k$ :



where  $A$  and  $B$  are some chiral substances and  $Z$  is some discarded substance. We can view  $Z$  as an achiral substance because it leaves the reaction area. Then in  $\theta$  and  $\Delta$  variables we will have:

$$\frac{d\rho_Z}{dt} = k\rho_A \rho_B + k\rho_a \rho_b = \frac{k}{2}(\theta_A \theta_B + \Delta_A \Delta_B) \approx \frac{k\theta_A \theta_B}{2} \quad (7)$$

because we ignore quadratic term  $\Delta_A \Delta_B$  in the linearized model. The requirement that the steady state concentrations do not change under closing requires that some of the rate coefficients need to be adjusted in order to fulfill that under closing. It is important to note that both pass-through models and "closed" cyclic models are open systems: pass-through models require constant inflow of "food" matter and a "sink" to remove waste products and "closed" cyclic models require constant inflow of energy to "recycle" the waste products.

## Averaging of enantioselectivity in catalytic models

We would like first to consider how averaging of various rates of reactions due to large number of substances may affect enantioselectivity of different types of reactions. Consider some reversible reaction:





and its catalytic version:



where  $C$  is a catalyst and  $A$ ,  $B$ , and  $D$  are some substances (not necessarily amino acids in these equations). Such reactions or their autocatalytic versions are usually a basis of analysis for chiral symmetry breaking (Avetisov and Goldanskii 1996; Saito and Hyuga 2004; Plasson et al. 2007). Forward ( $R^+$ ) and backward ( $R^-$ ) reaction rates for Eq. (9) are then:

$$\begin{aligned} R^+ &= k^+ \rho_A \rho_B \rho_C \\ R^- &= -k^- \rho_D \rho_C \end{aligned} \quad (10)$$

where  $\rho_i$  is a concentration of the  $i$ -th substance,  $k^+$  and  $k^-$  are forward and backward rate coefficients. However, if the number of catalysts is very large (which must be the case if the number of substances is very large), then at each moment of time forward and backward rates effectively become weighted averages of the rates of reactions with all relevant catalysts:

$$\begin{aligned} R^+ &= \rho_A \rho_B \sum_{\{C\}} k_C^+ \rho_C \equiv k_{eff}^+ \rho_A \rho_B \rho_{tot} \\ R^- &= -\rho_D \sum_{\{C\}} k_C^- \rho_C \equiv -k_{eff}^- \rho_D \rho_{tot} \end{aligned} \quad (11)$$

where  $\{C\}$  is a set of all possible catalysts for a given reaction at a given time and:

$$\begin{aligned} \rho_{tot} &= \sum_{\{C\}} \rho_C \\ k_{eff}^\pm &= \sum_{\{C\}} k_C^\pm \frac{\rho_C}{\rho_{tot}} \end{aligned} \quad (12)$$

If catalyst  $C$  is chiral, then its enantiomer  $\hat{E}(C)$  is also a catalyst.



and, therefore, both will contribute to averaged reaction rates as expressed by Eq. (11). In the current work, we are only interested in the analysis of the stability of the symmetric state. Subsequently, we ignore any possible asymmetry between enantiomers, for example, global asymmetry due to weak interaction (Rein 1974; Letokhov 1975). Therefore, we consider that the rates of reaction (9) and its equivalent for all enantiomers:



are exactly the same.

In addition, thermodynamic considerations require that the equilibrium state is unchanged for Eq. (9) and (13) in comparison to Eq. (8). This well-known fact means that forward and backward rates of the reaction (8) are equally affected by the catalyst and, therefore, forward and backward rates (in the presence of catalyst) must just be multiplied by exactly the same number. Subsequently, that means that total enantioselectivity, which is defined by Eq. (21) below, must be identically zero for any catalyst / enantiomer pair. Surprisingly, total enantioselectivity was “allowed” to deviate from that identically zero value in a substantial number of models and for quite a large number of years. However, since current work is not about the analysis of the previous mistakes, we will not perform any further review of that matter.

Apart from some very exotic cases, like some specifically constructed cyclic peptides, which are invariant under simultaneous application of operator  $\hat{E}$  and some nontrivial rotation around the cycle, the enantiomer  $\hat{E}(C)$  is a different substance and, in general, will have different catalytic properties when applied to reaction (8). Therefore, multiplication factors could be different for Eq. (9) and (13) and that will result in non-zero enantioselectivities of forward and backward reactions. It is then interesting that some of the well-known catalysts for chiral separation, asymmetric synthesis, or even ligation (Fujima et al. 2006; Sczepanski and Joyce 2014) have more stereo centers of opposite chirality in comparison to the substances on which they operate.

## Models with peptides up to length 5, catalytic synthesis, and crystallization of insoluble pair

Modeling of a system with a large number of amino acids and long peptides is an NP-complete problem (Cook 1971) because the number of possible peptides and reactions grows exponentially as the number of amino acids and length of peptides increases. Therefore, it is important to understand if there are models, which can produce chiral symmetry breaking with a substantially limited number of amino acids and short peptides, but without fine-tuning. Models with autocatalytic synthesis of a single amino acid (which could be called a “peptide” of length one) were considered in great detail in the last 60 years (Frank 1953; Avetisov and Goldanskii 1996) and APED-like models considered peptides of length two but without catalytic synthesis of amino acids. Therefore, the minimum interesting length of peptides is three. From one side, further increase in peptide length and / or number of amino acids in the model will result in exponentially increasing difficulties. From the other side, if peptide formation is reversible, which is applicable in this case because we will be considering polymerization and depolymerization reactions, then it is straightforward to illustrate that the distribution of concentration of peptides by length forms decreasing geometric progression with some common ratio  $\alpha < 1$ . Therefore, the longer the peptide lengths, the smaller concentration they will have and therefore the smaller their contribution will be. As such, we would like now to limit our consideration to just one amino acid, which we denote here as  $A$  for left enantiomer and as  $a$  for right enantiomer. We combine several well-known models described in the literature (Frank 1953; Avetisov and Goldanskii 1996; Plasson et al. 2004; Saito and Hyuga 2004; Weissbuch et al. 2005; Brandenburg et al. 2007; Plasson et al. 2007; van der Meijden et al. 2009; Noorduyn et al. 2010; Blackmond 2011; Coveney et al. 2012; Hein et al. 2012; Ribó et al. 2013; Konstantinov and Konstantinova 2013, 2014) and then study which combinations can possibly result in the symmetry breaking under the conditions of weak enantioselectivity of all relevant reactions. Even though the minimum interesting peptide length is three, we consider peptide chains up to length five in order to study how the increase in peptide length influences the symmetry breaking. Therefore, the possible substances are:

$$\begin{aligned}
 &Y, A^*, a^*, A, a, AA, Aa, aA, aa, AAA, AAa, AaA, aAA, Aaa, aAa, aaA, aaa, AAAA, AAAa, AAaA, \\
 &AaAA, aAAA, AAaa, AaAa, Aaaa, aAAa, aAaA, aaAA, Aaaa, aAaa, aaAa, aaaA, aaaa, AAAAA, \\
 &AAAAa, AAAaA, AAaAA, AaAAA, aAAAA, AAAaa, AAaAa, AaaaA, AaAaA, AaaAA, aAAAA, aAAaA, aAaAA, aaAAA, AAaaa, AaAaa, AaaAa, AaaaA, aAaaa, aAaAa, aAaaa, aaAaa, aaaAa, aaaaA, aaaaa \quad (15)
 \end{aligned}$$

where  $A^*$ ,  $a^*$  are activated left and right amino acids as used in some models. We would like to study all possible variants when no more than one substance (and its enantiomer) are catalysts with low enantioselectivity and exactly one pair of substances (and their enantiomers) forms some insoluble substances. If we have many substances in the solution, then as the amount of solvent decreases and concentrations increase, there is always the least soluble substance, which will be the first to start forming sediment. In the current simplified model, we assume that the least soluble substance is insoluble. However, a more realistic model must take into account non-zero solubilities of all substances in the system and their influence on each other.

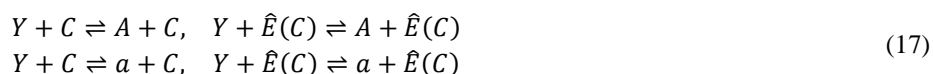
It is important to note that, while sedimentation / dissolution is, of course, a reversible process, there are certain conditions under which it could be viewed as irreversible. For example, a sediment could fall into the ocean floor where the temperature is much lower. Thus it will effectively leave the reaction area. We consider here sedimentation as an effectively irreversible process and that imposes a cycle of how chiral matter moves in the system. The following reactions are included in our consideration:

### Reversible synthesis of amino acids



It was shown (Avetisov and Goldanskii 1996) that the coefficients  $k(Y \rightarrow A)$  and  $k(A \rightarrow Y)$  must be small in comparison to the rate of catalytic synthesis. However, the maximum of  $\frac{1}{k(Y \rightarrow A)}$  and  $\frac{1}{k(A \rightarrow Y)}$  determines time, which is needed for the system to reach equilibrium if only these two reactions are considered. Therefore, the evolution time must be  $t \gg \max\left(\frac{1}{k(Y \rightarrow A)}, \frac{1}{k(A \rightarrow Y)}\right)$ . If these coefficients are chosen too small, then the evolution time will be very large and subsequently the calculation time will increase drastically. We used the values  $k(Y \rightarrow A) = k(Y \rightarrow a) = c_+ = 10^{-2}$  and  $k(A \rightarrow Y) = k(a \rightarrow Y) = c_- = 10^{-3}$  in our calculations.

### Catalytic synthesis of amino acids



where  $C$  is a catalyst (any one of  $A, a, AA, Aa, aA, aa, AAA, AAa, Aaa, aAA, aAa, aaA, aaa, \dots, aaaaa$ ).

We are interested in the scenario when the forward coefficient is much larger than the backward one. However, backward coefficients must be also chosen so that the characteristic evolution time is not too large. Given that the estimated minimum concentration of a potential catalyst in the models, which we considered, is  $\sim 10^{-3}$  and the used evolution time, backward coefficient should not less than 0.1. If catalytic reactions were used, then given a catalyst  $C$ , we used the following values of the parameters:

$$\Gamma_+ = 10, \quad \Gamma_- = 0.1 \quad (18)$$

Alternatively, if catalysis was not used, then we used:

$$\Gamma_+ = \Gamma_- \equiv 0 \quad (19)$$

As discussed above, thermodynamic considerations require that the catalyst does not change the equilibrium. This condition means that coefficients of enantioselectivity of forward and inverse catalysis satisfy the following condition:

$$\gamma_+ \equiv -\gamma_- \quad (20)$$

Subsequently that means that total enantioselectivity of a catalyst,  $\gamma$ , is identically zero:

$$\gamma = \gamma_+ + \gamma_- \equiv 0 \quad (21)$$

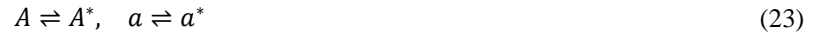
For the cases when  $\Gamma_+ > 0$  and  $\Gamma_- > 0$  we considered two distinct variants of enantioselectivity:

$$\begin{aligned} \gamma_+ &\equiv -\gamma_- = 0 \\ \gamma_+ &\equiv -\gamma_- = \pm\delta_\gamma \end{aligned} \quad (22)$$

We used the value  $\delta_\gamma = 0.1$  in our calculations.

## Activation

It is important to note that two primary scenarios can be realized experimentally in regards to the formation of peptides: peptides can be considered as growing only in one direction by adding one amino acid at only one end of the peptide or arbitrary reactions of peptide formation can be considered. That loosely corresponds to two distinct classes of models: the models with activation and the models without activation. The first scenario was considered in (Plasson et al. 2004; Brandenburg et al. 2007) and therefore we consider the class of models where an amino acid must first be activated before it can attach to another amino acid of a peptide chain, as well as the models when the peptides can grow from both sides. The activation reaction (if used) can be written as follows:



We used the values  $k(A \rightarrow A^*) = k(a \rightarrow a^*) = a_+ = 1$  and  $k(A^* \rightarrow A) = k(a^* \rightarrow a) = a_- = 10^{-3}$  in our calculations. This corresponds to the situation when most of the amino acids are in an activated state.

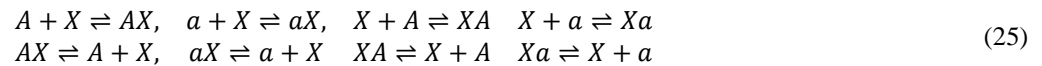
## Polymerization / Depolymerization

If activation was taken into account, then we consider the models where the peptide chains can only grow to the left:



where  $X$  is any of the allowed by the model substances ( $A, a, AA, Aa, aA, aa, \dots, aaaa$ ).

If activation was not considered, then the following reactions are used:



where  $X$  is any of the allowed by the model substances ( $A, a, AA, Aa, aA, aa, \dots, aaaa$ ).

The following reactions were also considered for the models with peptides longer than three and when activation was not used:

$$P_1 + P_2 \rightleftharpoons P_1 P_2, \quad (26)$$

where  $P_1$  and  $P_2$  are some peptide chains with a combined length not exceeding five. With the exception of APED-like models coefficients for all forward polymerization reactions were considered to have the same value  $\Lambda_+ = 1$ . Similarly, all depolymerization reactions were also considered as having the same coefficient with the value  $\Lambda_-$  for all models except APED-like ones. We considered several values of  $\Lambda_-$  in order to study how it affects the symmetry breaking.

### Sedimentation of insoluble pairs

As discussed above, any pass-through models, including models with sedimentation can be "closed" by short-circuiting them back to the beginning of the models (synthesis of amino acids). As we are interested only in the symmetry breaking, we can utilize this to short-circuit pair formations with subsequent sedimentation. That allows us to use a simplified model of irreversible sedimentation:

$$P_1 + P_2 \rightarrow nY, \quad \hat{E}(P_1) + \hat{E}(P_2) \rightarrow nY \quad (27)$$

where  $P_1$  and  $P_2$  are the substances (peptides), which form an insoluble pair, and  $n$  is their combined length. Given two substances  $P_1$  and  $P_2$  there are two other possible combinations, which may form insoluble pairs:

$$P_1 + \hat{E}(P_2) \rightarrow nY, \quad \hat{E}(P_1) + P_2 \rightarrow nY \quad (28)$$

In the simplified model of sedimentation, we assume that only one of the reactions Eq. (27) or Eq. (28) holds, while the other pair is soluble.

While the computational complexity substantially decreases in such a case, it is important to compare the results with a more realistic model. If we assume that a pair of peptides may form a temporary (not peptide) bond and then subsequently some of such pairs are insoluble, then we have to consider all possible combinations. That creates additional reaction channels:

$$P_1 + P_2 \rightleftharpoons (P_1)(P_2), \quad (29)$$

which are different from peptide formation reaction (26), and then some of these pairs form a sediment. In addition, we can no longer consider only one combination Eq. (27) or Eq. (28). Rather, it is important to take both into account and also consider possible mutual influence on the solubility: if we have many substances in the solvent, then it becomes a very complicated process. Moreover, the sedimentation process  $C_{\rho_i \rightleftharpoons \rho_{C_i}}(\rho_i, \rho_{C_i})$ , even in its simplest form, is a discontinuous function of two variables  $\rho_i$  (the concentration of the substance in the solution) and  $\rho_{C_i}$  (the total amount of the sediment divided by the total volume of the system):

$$C_{\rho_i \rightleftharpoons \rho_{C_i}}(\rho_i, \rho_{C_i}) = \begin{cases} 0, & \rho_i \leq \rho_i^{(max)} \text{ and } \rho_{C_i} = 0 \\ k_{sol}(\rho_i - \rho_i^{(max)}), & \rho_i \leq \rho_i^{(max)} \text{ and } \rho_{C_i} > 0, \\ k_{cryst}(\rho_i - \rho_i^{(max)}), & \rho_i > \rho_i^{(max)} \end{cases} \quad (30)$$

where  $\rho_i^{(max)}$  is a maximum solubility,  $k_{cryst}$  is the rate of sedimentation if the concentration exceeds the maximum solubility,  $k_{sol}$  is the rate at which the sediment dissolves if the concentration is less than the maximum solubility and there is some sediment present, and a simplified Noyes - Whitney equation (Dokoumetzidis and Macheras 2006) was assumed. The fact that

function  $C_{\rho_i \rightleftharpoons \rho_{C_i}}(\rho_i, \rho_{C_i})$  is a discontinuous function results in significant difficulties in numerical simulations. The major one is that the solver engine will routinely overshoot when approaching zero and will calculate some concentrations as negative. While Filippov's continuation (Filippov 1988) could be used to remedy the situation, it will require creating some artificial constraints for non-physical negative concentrations. A simpler solution is to smooth out the function  $C_{\rho_i \rightleftharpoons \rho_{C_i}}(\rho_i, \rho_{C_i})$  near discontinuities and also wrap all concentrations into Heaviside theta functions to ensure that unphysical negative concentrations (if they occurred due to the solver's overshooting) will be treated as zeros. While the solution method then becomes stable and does not produce unphysical results, it also slows down very significantly. Therefore, we compare the simplified model described by Eq. (27) and the more realistic model described by Eq. (29, 30) for the models with peptides up to length three and use only simplified equations for the models with peptides up to length five.

## Classes of considered models

Table 1 below contains classification of the models, which we considered.

Table 1. Classes of considered models

No	Max Peptide Length	Activation	Catalysis	APED	Sedimentation
1	5	No	No / NE / WE	No	No / Simplified
2	5	Yes	No / NE / WE	No	No / Simplified
3	3	Yes	No	Yes	No / Simplified
4	3	No	No / NE / WE	No	Noyes - Whitney

where column **Max Peptide Length** contains the maximum length of the peptides allowed in the model, column **Activation** indicates whether activation was considered or not, column **Catalysis** contains types of considered catalytic synthesis of amino acids (**No** – no catalytic synthesis, **NE** – catalytic synthesis is not enantioselective  $\gamma_+ = \gamma_- = 0$ , **WE** – catalytic synthesis is weakly enantioselective:  $0 < |\gamma_+| = |\gamma_-| = \delta_\gamma \ll 1$ ) column **APED** indicates whether the model is APED-like, and column **Sedimentation** contains type of sedimentation considered (**No** uses no sedimentation, **Noyes – Whitney** uses Eq. (29, 30) and **Simplified** uses Eq. (27)).

## Solution method

The solution method is as follows. First, we choose a class of model from Table 1. Then for each combination of the catalyst and insoluble pair we generate a system of ordinary differential equations which describe such a model. The initial conditions were set to have slightly larger amounts of left amino acids in the system. Then the system of ordinary differential equations was solved for some period of time ( $t = 10^5$ ) and we recorded the final enantiomeric excess  $\eta$ .

The algorithm was written in Wolfram and it runs under Mathematica 10 or above. The source code is provided under GNU GPLv3 license or any later version (GNU General Public License 2007) and is available on GitHub at <https://github.com/kkmail/CLM/>.

## Symmetry breaking in the models with peptides up to length 5, catalytic synthesis of amino acids and simplified sedimentation

We first start with the models with peptide length up to five with or without activation and simplified sedimentation. These models are described by lines one and two of Table 1. We study the models with catalytic synthesis in three steps. First, we consider the models, in which catalytic synthesis is completely turned off, then we add catalytic synthesis, which has no enantioselectivity  $\gamma_+ = \gamma_- = 0$  for exactly one pair of enantiomers, and then we study the models with weak forward enantioselectivity of catalytic synthesis:  $\gamma_+ = -\gamma_- = \pm\delta_\gamma$  ( $\delta_\gamma > 0$ ). The value  $\gamma_+ = \delta_\gamma$  corresponds to the slightly positively enantioselective forward reaction  $Y + C \rightarrow A + C$  and the value  $\gamma_+ = -\delta_\gamma$  corresponds to the slightly positively

enantioselective forward reaction  $Y + \hat{E}(C) \rightarrow A + \hat{E}(C)$ . Therefore, there are a total of 62 possible catalysts for the peptides of length up to five.

We considered that exactly two substances (and their enantiomers) form insoluble substances. Given two substances  $F$  and  $G$  and their enantiomers the following distinct cases are possible. We can choose either pairs  $F + G$  and  $\hat{E}(F) + \hat{E}(G)$  or pairs  $F + \hat{E}(G)$  and  $\hat{E}(F) + G$  as insoluble. Even though there are  $62 \times 62 = 3,844$  pairs, there are only 992 possible unordered distinct pairs for peptides of length up to five. Therefore, all possible combinations of catalysts and pairs form a  $62 \times 992$  matrix. We will call the models with no sedimentation and no catalytic synthesis as **base models**.

As mentioned earlier, the distribution of concentration of peptides by length for all base models forms a geometric progression with some scale factor, which we denote as  $\rho_1$ , and some common ratio, which we denote as  $\alpha$ . It is difficult to directly compare the models without activation (NA) and with activation (A) because they have some differences in the equations. In particular, with the chosen coefficients of activation and deactivation most of the amino acids in the considered models with activation are in an activated state ( $A^*$ ). Nevertheless, if we add a concentration of not-activated and activated molecules then the geometric progression still holds. In order to compare such models, we chose to adjust coefficients  $\rho_0$  and  $\Lambda_-$  for relevant models with activation, so that the base models with and without activation have the same values of  $\rho_1$  and  $\alpha$ . Table 2 summarizes the values of  $\Lambda_+$ ,  $\Lambda_-$ , and  $\rho_0$ , which we considered, as well as the resulting values of  $\rho_1$  and  $\alpha$ . The model name was introduced for convenience and is used in subsequent results.

Table 2. Considered values of  $\Lambda_+$ ,  $\Lambda_-$ ,  $\rho_0$ ,  $\rho_{A_0}$ , and  $\alpha$

Model Name	Type	$\Lambda_+$	$\Lambda_-$	$\rho_0$	$\alpha$	$\rho_1$
NA100	NA	1	1	1	0.519	0.259
NA030	NA	1	0.3	1	0.788	0.118
A100	A	1	1	1.261	0.519	0.259
A044	A	1	0.437	1.174	0.788	0.118

The base model NA030 has a larger common ratio ( $\alpha = 0.788$ ) than model NA100 ( $\alpha = 0.519$ ). As the total concentration is the same for both of these models, this results in model NA030 having a larger concentration of peptides at length 4 and 5, approximately the same at length 3, and a smaller concentration at lengths 1 and 2 in comparison to model NA100. This is illustrated on Figure 1. Therefore, the influence of longer peptides should increase and the influence of shorter peptides should decrease in model NA030 in comparison to model NA100. Similar results should be expected in comparing models A100 and A044. Nevertheless, it is important to note that such an adjustment introduced some extra matter into the systems described by models A100 and A044 and, therefore, this may affect some output values of these models.

There is still some ambiguity left in comparing models with different common ratios of geometric progression while limiting the length of considered peptides up to 5. The problem is that by limiting our consideration to peptides of length 5 we effectively cut the tail of longer peptides. If the common ratio is larger, then the ignored tail is bigger. Therefore, it could be argued that the models NA030 and A044 have a larger effective total concentration  $\rho_0$  if it were adjusted for that missing tail.

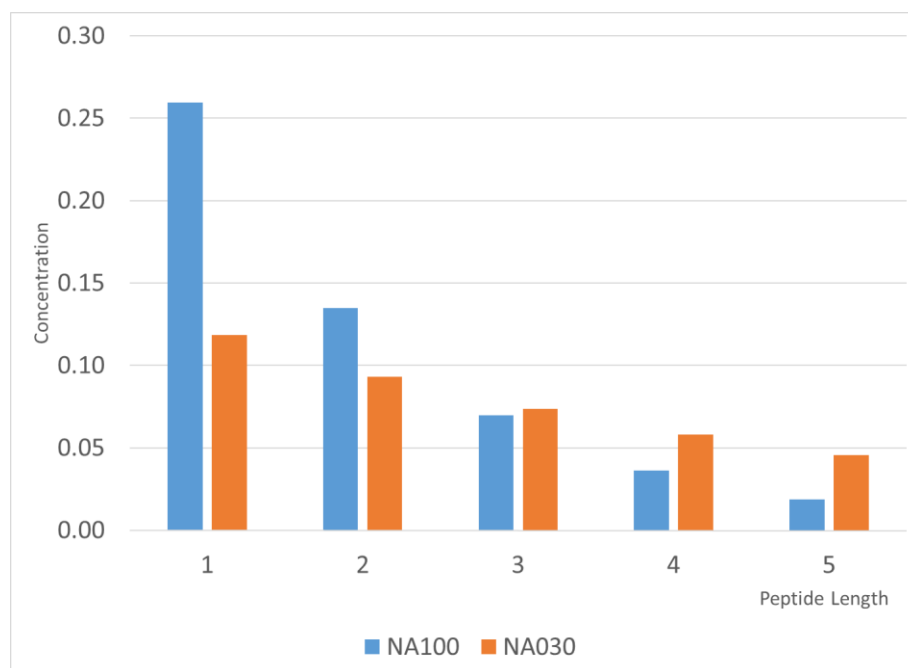


Figure 1. Total concentration of peptides of different length for the models NA100 and NA030.

Taking into account that the models that we considered are symmetric under permutation of the elements of the pairs that form a sediment, all models with sedimentation and no catalytic synthesis can be described either by  $62 \times 62$  matrix of values of output enantiomeric excess  $\eta$  for each combination of two possible peptides, which form insoluble pairs in the model, or by  $1 \times 992$  column if only unique variants are considered. Similarly, the models with sedimentation and catalytic synthesis can be described either by a  $62 \times 62 \times 62$  cube of values of output enantiomeric excess  $\eta$  for each combination of a catalyst and an insoluble pair or by a  $62 \times 992$  matrix if only unique insoluble pairs are considered.

### Symmetry breaking without catalytic synthesis

We start from the analysis of the results of the models with sedimentation and no catalytic synthesis. All non-zero results are summarized in Table 3. The table utilizes 3-color gradient color coding (red-yellow-green) to assist in identifying the pairs that result in a substantial symmetry breaking: the greener is the cell, the larger is the output enantiomeric excess and subsequently the redder is the cell, the lower is the output enantiomeric excess. Yellow cells mark pairs with intermediate enantiomeric excess output and white cells correspond to the pairs that produce no symmetry breaking at all.

Table 3. Enantiomeric excess in the absence of catalytic synthesis

Pair	NA100	NA030	A100	A044
AA + Aaa	85%	76%	90%	89%
A + AAaa			70%	80%
AA + AAaa			73%	82%
AAA + AAaa				36%
AAa + AAaa			40%	71%
AaA + AAaa			40%	71%

Pair	NA100	NA030	A100	A044
AAA + aaaA	68%	68%	64%	61%
AAA + aaAa	66%	64%	64%	61%
Aaa + aaAA			57%	74%
AAA + aAaa	66%	64%	64%	62%
Aaa + aAaA				54%
Aaa + aAAa		49%		54%



A + AaAa			62%	75%
AA + AaAa			62%	75%
AAA + AaAa				21%
AAa + AaAa				58%
AaA + AaAa			47%	71%
A + AaaA	36%	77%	62%	75%
AA + AaaA		67%	62%	75%
AAA + AaaA				21%
AAa + AaaA		49%	47%	71%
AaA + AaaA				58%
AAA + Aaaa	68%	68%	69%	71%
Aaaa + AAAaa		53%		
Aaaa + AAaAa		53%		31%
Aaaa + AAaaA		46%		31%
AA + AAaaa	73%	80%	70%	80%
AAA + AAaaa	54%	70%	63%	72%
AAAA + AAaaa		20%	16%	36%
AAAAa + AAaaa		53%	34%	58%
AAaA + AAaaa		55%	34%	58%
AaAA + AAaaa		56%	34%	57%
Aaaa + AaAAa		54%		38%
Aaaa + AaAaA		52%		38%
AA + AaAaa	73%	80%	69%	80%
AAA + AaAaa	53%	70%	61%	70%
AAAA + AaAaa		20%	4%	32%
AAAAa + AaAaa		53%	31%	56%
AAaA + AaAaa		55%	31%	56%
AaAA + AaAaa		57%	37%	60%
Aaaa + AaaAA		47%		40%
AA + AaaAa	73%	81%	69%	79%
AAA + AaaAa	53%	70%	61%	70%
AAAA + AaaAa		19%		31%
AAAAa + AaaAa		54%	29%	55%
AAaA + AaaAa		58%	37%	60%
AaAA + AaaAa		56%	30%	54%
AA + AaaaA	74%	82%	69%	79%
AAA + AaaaA	57%	72%	61%	70%
AAAA + AaaaA		26%		31%
AAAAa + AaaaA		59%	37%	60%
AAaA + AaaaA		59%	29%	55%
AaAA + AaaaA		59%	30%	54%
AAAA + Aaaaa	41%	53%	43%	52%
AA + aaA	85%	76%	87%	82%
AA + aAa	82%	72%	87%	82%

AAAA + aaaaA	41%	53%	38%	43%
AAAA + aaaAa	38%	50%	38%	43%
AA + aaaAA	73%	80%	54%	60%
AAA + aaaAA	54%	70%	51%	58%
AAAA + aaaAA		20%		
AAAAa + aaaAA		56%		36%
AAaA + aaaAA		56%		36%
AaAA + aaaAA		55%		
AAAA + aaAaa	37%	50%	38%	43%
AA + aaAaA	73%	80%	57%	69%
AAA + aaAaA	53%	70%	54%	63%
AAAA + aaAaA		20%		14%
AAAAa + aaAaA		54%		40%
AAaA + aaAaA		57%		6%
AaAA + aaAaA		55%		38%
AA + aaAAa	71%	78%	57%	69%
AAA + aaAAa	49%	67%	54%	63%
AAAA + aaAAa		5%		14%
AAAAa + aaAAa		47%		6%
AAaA + aaAAa		50%		40%
AaAA + aaAAa		49%		38%
Aaaa + aaAAA		56%	38%	58%
AAAA + aAaaa	38%	50%	39%	44%
AA + aAaaA	73%	81%	61%	74%
AAA + aAaaA	53%	70%	56%	65%
AAAA + aAaaA		19%		21%
AAAAa + aAaaA		54%		47%
AAaA + aAaaA		56%		46%
AaAA + aAaaA		58%		45%
AA + aAaAa	71%	79%	61%	74%
AAA + aAaAa	51%	69%	56%	65%
AAAA + aAaAa		19%		21%
AAAAa + aAaAa		52%		46%
AAaA + aAaAa		53%		47%
AaAA + aAaAa		53%		45%
Aaaa + aAaAA		54%	28%	48%
AA + aAAaa	71%	78%	62%	75%
AAA + aAAaa	49%	67%	56%	66%
AAAA + aAAaa		5%		24%
AAAAa + aAAaa		46%		48%
AAaA + aAAaa		49%		48%
AaAA + aAAaa		50%		48%
Aaaa + aAAaA		54%	27%	48%
Aaaa + aAAAA		59%	27%	48%

There are several interesting facts that can be observed from that table:

1. Total number of left enantiomers is never equal to total number of right enantiomers in the pairs that produce symmetry breaking. This is subtle but very crucial. It is due to the following fact. Consider two peptides,  $P_1$  and  $P_2$ , which form insoluble pairs, and which have symmetric state concentrations:

$$\rho(P_1) = \rho(\hat{E}(P_1)) = \theta_1, \quad \rho(P_2) = \rho(\hat{E}(P_2)) = \theta_2 \quad (31)$$

and consider that total numbers of left and right amino acids in  $P_1$  and  $P_2$  are:

$$N_A^{tot} = \sum_{s=1,2} n_A(s), \quad N_a^{tot} = \sum_{s=1,2} n_a(s) \quad (32)$$

where  $n_A(s)$  is a number of left amino acids in a given substance  $s$ ,  $n_a(s)$  is a number of right amino acids in a given substance  $s$ . And then consider a variation:

$$\begin{aligned} \rho(P_1) &\rightarrow \frac{\theta_1 + \Delta_1}{2}, & \rho(\hat{E}(P_1)) &\rightarrow \frac{\theta_1 - \Delta_1}{2} \\ \rho(P_2) &\rightarrow \frac{\theta_2 + \Delta_2}{2}, & \rho(\hat{E}(P_2)) &\rightarrow \frac{\theta_2 - \Delta_2}{2} \end{aligned} \quad (33)$$

The variation in the difference between the number of left and right amino acids turns out to be proportional to  $(N_A^{tot} - N_a^{tot})$  and, therefore, when this quantity is zero symmetry breaking cannot possibly occur in the considered models.

2. Minimum total combined length of pairs, which produce symmetry breaking, is 5, for example,  $AA + Aaa$ . There are no pairs of length 1 or 2, which could result in chiral symmetry breaking in the absence of catalytic synthesis. There are very interesting pairs at length 3 ( $AA + Aaa$ ,  $AA + aAa$ ,  $AA + aaA$ ), which result in substantial enantiomeric excess both for models with and without activation. Similar pairs appear at length 4 ( $AAA + aaaaA$ ,  $AAA + aaAa$ ,  $AAA + aAaa$ ,  $AAA + Aaaa$ ) and length 5 ( $AA + aaaaA$ ,  $AA + aaAaA$ ,  $AA + aaAAa$ ,  $AA + aAaaa$ ,  $AA + aAaaA$ ,  $AA + aAAaa$ ,  $AA + AaaaaA$ ,  $AA + AaaAa$ ,  $AA + AaAaa$ ,  $AA + AAaaa$ ).
3. An interesting question becomes why shorter pairs don't work. The models with peptides with length up to two can be considered analytically. Consider, for example, a model with the following insoluble pair:  $A + Aa$ . Using  $\theta_X$  and  $\Delta_X$  variables and performing a variation around symmetric state the following matrix of coefficients of small variations  $\Delta_X = (\Delta_A, \Delta_{AA}, \Delta_{Aa})$  around the fixed point, the following matrix of coefficients can be obtained:

$$M = \begin{pmatrix} -\frac{\theta_A \Lambda_+ (5 \varepsilon_- \theta_A + 8 \Lambda_-)}{4 (\varepsilon_- \theta_A + 2 \Lambda_-)} & \Lambda_- & -\frac{\varepsilon_- \theta_A}{2} \\ \frac{\theta_A \Lambda_+}{2} & -\frac{\Lambda_-}{2} & 0 \\ -\frac{\varepsilon_- \theta_A^2 \Lambda_+}{4 (\varepsilon_- \theta_A + 2 \Lambda_-)} & 0 & -\frac{\varepsilon_- \theta_A + 2 \Lambda_-}{4} \end{pmatrix} \quad (34)$$

Performing analysis of the roots of characteristic equation, it is straightforward to show that all roots have negative real part for any values of parameters and, therefore, no symmetry breaking can occur in such model.

4. Models NA100 and A100 have much less possible insoluble pairs which can produce a symmetry breaking without any catalytic synthesis, in comparison to models NA030 and A044. This is due to the fact that models NA030 and A044 have larger concentrations of longer peptides and, therefore, longer peptides can have larger possible contribution to the symmetry breaking in these models.
5. It appears that model A100 has more possible pairs, which produce a symmetry breaking in the absence of any catalytic synthesis, than model NA100. Similarly, model A044 seems to have more possible pairs, which produce such symmetry breaking in comparison to model NA030. While it might be tempting to say that models with activation can produce symmetry breaking for the larger number of possible pairs, this discrepancy could be attributed to the higher total concentration  $\rho_0$  in models A100 and A044 in comparison to models NA100 and NA030.
6. Models without activation (NA100, NA030) are invariant to inversion transformation: the values of the output enantiomeric excess for the pairs  $F + G$  and  $\hat{I}(F) + \hat{I}(G)$  are the same. On the opposite side, models with activation are not invariant to inversion transformation. As a result, the values of output enantiomeric excess may be different for the pairs  $F + G$  and  $\hat{I}(F) + \hat{I}(G)$ .
7. The models with and without activation are nearly identical if sedimentation occurs at peptide length no more than 3. However, as the length of the peptides increases, the models with activation start to differ more and more from the models without activation. Again, this should be attributed to the invariance / non-invariance of the equations to the inversion transformation for models without / with activation.

## Symmetry breaking in the presence of weakly enantioselective forward catalytic synthesis

We now "turn on" slightly enantioselective forward catalytic synthesis and compare the differences with the models without any catalytic synthesis. In particular, we are interested if there are any insoluble pairs which will result in significant symmetry breaking in the presence of catalytic synthesis but did not work without it.

If forward catalytic synthesis is not enantioselective at all ( $\gamma_+ = \gamma_- = 0$ ) then it mostly does not result in any additional pairs which can result in symmetry breaking. There are a few pairs, which produce weak output enantioselectivity:  $AA + Aaaa$ ,  $Aaaa + aAAAA$ , and  $AAAA + AaaaA$ . Only the first pair ( $AA + Aaaa$ ) produces a reasonably large output of enantiomeric excess ( $\eta = 0.36 - 0.46$ ) for some catalysts and some models and the other two produce an insignificant output of enantiomeric excess. We do not consider models with non-enantioselective forward catalytic synthesis further, as they do not produce significant chiral symmetry breaking beyond what is produced by the models without catalytic synthesis. Some catalysts may actually work against symmetry breaking and therefore some combinations do not produce symmetry breaking in the presence of catalysts in comparison to models without catalytic synthesis.

If forward catalysis is weakly enantioselective ( $\gamma_+ \neq 0$ , but, as mentioned earlier,  $\gamma_+ + \gamma_- \equiv 0$ ) then additional variants of symmetry breaking occur in comparison to base models. The first question is what happens with the pairs that produced symmetry breaking without any catalytic synthesis, as specified in Table 3. It turns out that the "strong" pairs, which produce substantial symmetry breaking, are mostly unaffected by the catalysts, even by the ones that inhibit symmetry breaking rather than amplify it. We will not consider that in details as it does not add anything interesting to the results described above. Much more interesting question is if there are pairs, which did not produce a symmetry breaking without catalytic synthesis but produce in its presence. The number of combinations where this happens is still quite large. Therefore, we only look at a subset where enantiomeric excess  $\eta \geq 0.6$  and the number of catalysts, which works for at least one of the models is also sufficiently large. Given that only 31 catalyst at maximum can potentially increase the symmetry breaking (the remaining 31 are their enantiomers and, therefore, they inhibit the symmetry breaking), we only looked at combinations where at least half of potential

catalysts (15 or more) could work. Table 4 shows all such pairs along with the number of catalysts that work for a particular model / pair combination. This table also utilizes 3-color gradient color coding (red-yellow-green) to assist in identifying the pairs that work with a substantial number of catalysts: the greener is the cell, the larger is the number of catalysts, which produce symmetry breaking, and subsequently the redder is the cell, the smaller is the number of such catalysts. Yellow cells mark pairs with intermediate number of catalysts and white cells correspond to the pairs that produce no symmetry breaking at all for any catalysts.

Table 4. Number of catalysts producing symmetry breaking with  $\eta \geq 0.6$ 

Pair	NA100	NA030	A100	A044
A + Aa	9	25	16	19
Aa + Aa	11	27	26	26
A + Aaa	16	29	22	23
AAa + Aaa	20	25	25	27
AaA + Aaa	18	28	25	27
Aa + AAaa	21	27	22	22
Aaa + AAaa	13	22	6	5
AAaa + AAaa	11	24	10	17
Aa + AaAa	22	24	21	22
Aaa + AaAa	14	24	11	19
AAaa + AaAa	17	25	16	24
AaAa + AaAa	11	24	10	18
Aa + Aaaa	21	26	21	22
Aaa + Aaaa	11	13	11	19
AAaa + Aaaa	17	26	16	24
AaAa + Aaaa	17	25	16	24
Aaaa + Aaaa	11	24	10	18
AA + Aaaa	22	27	23	25
AAaa + Aaaa	10	20	5	9
AAaA + Aaaa	11	23	5	9
AaAA + Aaaa	11	21	3	9
Aaa + AAaaA	10	22		8
Aaa + AAaAa	10	22		17
Aaa + AAaaaA	10	20		17
A + AAaaa	10	19		18
AAa + AAaaa	10	22	5	19
AaA + AAaaa	10	23	5	19
Aaa + AaAAa	10	22		17
Aaa + AaAaA	10	20		17
A + AaAaa	10	19		18
AAa + AaAaa	10	22	5	18
AaA + AaAaa	10	20	5	18
Aaa + AaaaA	10	21		17
A + AaaaA	10	19		18
AA + aaAa	22	27	23	26
AAaA + aaAa	10	19		
A + aaAA	21	26	10	6
AA + aaAA	13	20		
Aa + aaAA	21	27	22	22
AAa + aaAA	14	24	10	10
AaA + aaAA	15	23	10	10
AAaa + aaAA	11	24	9	19
AA + aAaa	22	27	23	24
AAaa + aAaa	11	23	7	17
AAaA + aAaa	11	23	7	17
AaAA + aAaa	10	19		
A + aAaA	22	24	10	10
AA + aAaA	11	20	1	5
Aa + aAaA	22	24	21	22
AAa + aAaA	14	24	11	15
AaA + aAaA	12	21	10	9
AAaa + aAaA	17	25	13	24
AaAa + aAaA	11	24	10	19
A + aAAa	15	21	10	10
Aa + aAAa	21	26	21	22
AaA + aAAa	15	18	11	15
AAaa + aAAa	17	26	13	24
AaAa + aAAa	17	25	12	24
Aaaa + aAAa	11	25	10	19
AAA + aaaaA	10	20	5	11
AAA + aaaAa	10	20	5	11
A + aaaAA	10	19		6
AAa + aaaAA	10	22		17
AaA + aaaAA	10	23		17
AAA + aaAaa	10	20	5	8
A + aaAaA	10	19		6
AAa + aaAaA	10	22		17

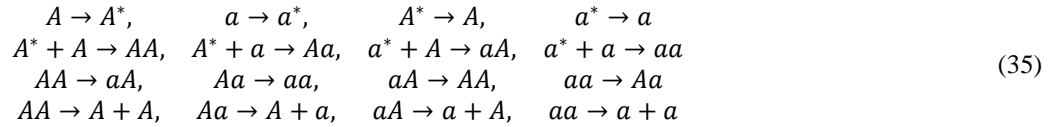
AAa + AaaAa	10	22	5	17	AaA + aaAaA	10	20		11
AaA + AaaaAa	10	23	5	17	A + aaAAa	10	19		6
A + AaaaA	10	19		18	AAa + aaAAa	10	21		11
AAa + AaaaA	10	23	5	17	AaA + aaAAa	10	22		17
AaA + AaaaA	10	23	5	17	Aaa + aaAAA	10	22	5	17
AAA + Aaaaa	10	20	5	7	AAA + aAaaa	10	20	5	7
A + a	3	4	27	28	A + aAaaA	10	19		10
AA + aa	11	26	26	26	AAa + aAaaA	10	23		18
A + aA	9	25	16	19	AaA + aAaaA	10	23		17
Aa + aA	11	27	26	26	A + aAaAa	10	19		10
AAA + aaa	14	19	6	3	AAa + aAaAa	10	20		17
A + aaA	16	29	22	23	AaA + aAaAa	10	19		18
AAa + aaA	23	27	24	25	Aaa + aAaAA	10	22	5	17
A + aAa	16	25	22	23	A + aAAaa	10	19		10
AAa + aAa	18	28	25	27	AAa + aAAaa	10	20		18
AaA + aAa	22	25	24	25	AaA + aAAaa	10	22		18
Aaa + aAA	23	27	24	22	Aaa + aAAaA	10	23	5	17
AA + aaaA	22	27	23	26	Aaa + aAAaA	10	23	5	17

The results are very interesting (some of them are not possible to show in the table above but follow from the analysis of all data):

1. Out of 104 pairs in Table 4, 89 have equal total number of right and left amino acids. Such pairs cannot possibly produce symmetry breaking without catalytic synthesis, as discussed above. What happens is that the catalyst is responsible for symmetry breaking in such models and sedimentation "assists" the catalyst by eliminating either the catalyst directly or some peptides, from which it could be assembled.
2. All catalysts which work with a substantial number of insoluble pairs for models without and with activation, share the same property. They have either zero or one amino acid of the opposite chirality: *A, aA, Aa, AA, aAA, AaA, Aaa, AAA, aAAA, Aaaa, AAaA, AAAa, AAAA, aAAAA, AaAAA, AAaAA, AAAaA, AAAAA, AAAAAA*. The catalysts with length 1 and 2 work only with a very limited number of possible insoluble pairs. The catalysts with length 3 to 5 work almost with the same pairs for the models without activation. The catalysts with length 3 work with a slightly smaller number of pairs for the models with activation.
3. All the models considered in the current article are far from equilibrium. Given that considered models enforce a unidirectional flow of matter, this effectively "turns off" the effect of enantioselectivity of reverse catalysis. Subsequently it is only the value of  $\gamma_+$  that affects symmetry breaking, whereas models in thermodynamic equilibrium are governed by the value of total enantioselectivity  $\gamma = (\gamma_+ + \gamma_-) \equiv 0$  for catalytic models.

## Symmetry breaking in epimerization based models

The models based on activation, polymerization, epimerization, depolymerization cycle and peptides up to length 2 (APED-2) (Plasson et al. 2004; Weissbuch et al. 2005; Brandenburg et al. 2007; Plasson et al. 2007) are based on the following set of reactions:



where  $A$  and  $a$  are left and right enantiomers of some amino acid and  $A^*$  and  $a^*$  are activated amino acids. The APED-2 model utilizes parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  to measure enantioselectivity of polymerization, depolymerization, and epimerization correspondingly:

$$\alpha = \frac{k(A^* + a \rightarrow Aa)}{k(A^* + A \rightarrow AA)}, \quad \beta = \frac{k(Aa \rightarrow A + a)}{k(AA \rightarrow A + A)}, \quad \gamma = \frac{k(AA \rightarrow aA)}{k(Aa \rightarrow aa)} \tag{36}$$

In order to compare the APED-2 model with the APED-like model for peptides up to length three (APED-3) we need to express both models using the same parameters. Starting from the APED-2 model and transforming to  $\theta_X$ , and  $\Delta_X$  variables, it is straightforward to obtain the system of equations for a reduced system:

$$\left\{ \begin{array}{l} \rho_Y + \theta_A + \theta_{A^*} + 2(\theta_{Aa} + \theta_{AA}) = 0 \\ a \theta_A - \frac{1}{2}(2b + p(1 + \alpha)\theta_A)\theta_{A^*} = 0 \\ \frac{1}{2}(2e \theta_{Aa} - 2(h + e \gamma)\theta_{AA} + p \theta_A \theta_{A^*}) = 0 \\ \frac{1}{2}(-2(e + h \beta)\theta_{Aa} + 2e \gamma \theta_{AA} + p \alpha \theta_A \theta_{A^*}) = 0 \end{array} \right. \tag{37}$$

where  $a = k(A \rightarrow A^*)$ ,  $b = k(A^* \rightarrow A)$ ,  $e = k(Aa \rightarrow aa)$ ,  $h = k(AA \rightarrow A + A)$ , and  $p = k(A^* + A \rightarrow AA)$  are parameters of the APED model. The equations are linear in  $\rho_Y$ ,  $\theta_A$ ,  $\theta_{Aa}$ , and  $\theta_{AA}$ . Therefore, they can be solved to express these variables via  $\theta_{A^*}$ :

$$\theta_A = \frac{2b\theta_{A^*}}{C_2}, \quad \theta_{AA} = \frac{bpD\theta_{A^*}^2}{hBC_2}, \quad \theta_{Aa} = \frac{bpE\theta_{A^*}^2}{hBC_2} \tag{38}$$

where:

$$\begin{array}{l} B = e(1 + \beta\gamma) + h\beta \\ C_2 = 2a - (1 + \alpha)p\theta_{A^*} \\ D = e(1 + \alpha) + h\beta \\ E = h\alpha + e(1 + \alpha)\gamma \end{array} \tag{39}$$

Condition  $C_2 > 0$  imposes a constraint on the maximum possible value of dimensionless quantity  $\epsilon = \left(\frac{p \theta_{A^*}}{a}\right)$  in such a system:

$$\epsilon^{max} = \left(\frac{p \theta_{A^*}}{a}\right)^{max} = \frac{2}{(1 + \alpha)} \tag{40}$$

Linearizing near the fixed point of a reduced system produces the following  $4 \times 4$  matrix  $M$  of coefficients of small variations  $\Delta_X = (\Delta_A, \Delta_{A^*}, \Delta_{AA}, \Delta_{A\alpha})$  around the fixed point:

$$M = \begin{pmatrix} -a - \frac{1}{2}p(1+\alpha)\theta_{A^*} & b - \frac{1}{2}p(1-\alpha)\theta_A & 2h & 0 \\ a - \frac{1}{2}p(1-\alpha)\theta_{A^*} & -b - \frac{1}{2}p(1+\alpha)\theta_A & 0 & 0 \\ \frac{p\theta_{A^*}}{2} & \frac{p\theta_A}{2} & -h - e\gamma & -e \\ -\frac{1}{2}p\alpha\theta_{A^*} & \frac{p\alpha\theta_A}{2} & -e\gamma & -e - h\beta \end{pmatrix} \quad (41)$$

We only need the determinant of matrix  $M$  to determine the boundary conditions for  $\alpha$ ,  $\beta$ , and  $\gamma$ . The boundary is determined when the value of the determinant crosses zero. The following equation describes the boundary:

$$\beta\gamma = \frac{(1-\alpha)p\theta_{A^*}}{4a - (1-\alpha)p\theta_{A^*}} \equiv \frac{\alpha(1-\alpha)}{p(1+3\alpha) + \left(\frac{b}{p\theta_A}\right)} \quad (42)$$

The case when there is no enantioselectivity in the APED-2 model corresponds to values  $\alpha = \beta = \gamma = 1$ . For the values  $\beta = \gamma$ , the maximum value of  $(\beta\gamma)$  is reached at  $\alpha = \beta = \gamma = \frac{1}{3}$ . The reported experimental value  $\alpha \approx 0.35$  (Plasson 2003) for some amino acids is close to the optimal and estimated value for  $\gamma \approx 0.3$  interestingly points to a systematic shift (Plasson et al. 2004; Saint-Martin and Julg 1991). However, there seems to be no solid experimental data on the value of parameter  $\beta$ . Therefore, fine-tuning in APED-like models may also require specific values of enantioselectivity of polymerization, epimerization, and depolymerization reactions.

If the number of amino acids is very large, then instead of, for example, reaction  $A^* + A \rightarrow AA$  we will have many reactions like  $A^* + B \rightarrow AB$  where  $B$  is any of the available in the solution amino acids. We can therefore view the enantioselectivity of polymerization:

$$\tilde{\lambda}_{A,B}^+ = \frac{1}{4} \left( \frac{k(A^* + B \rightarrow AB)}{k(A^* + b \rightarrow Ab)} - \frac{k(A^* + b \rightarrow Ab)}{k(A^* + B \rightarrow AB)} \right) \quad (43)$$

as a function of two discrete variables,  $A$  and  $B$ . Consider first  $\tilde{\lambda}_{A,A}^+$ . Measured for some amino acids value of  $\alpha \approx 0.35$  corresponds to a value of  $\tilde{\lambda}_{A,A}^+ \approx 0.63$ , which is far enough from zero. Consider now some amino acid  $A$ , for which  $\tilde{\lambda}_{A,A}^+$  has appropriate for APED value and consider what are the values of  $\tilde{\lambda}_{A,B}^+$  when we look at many different  $B$  values. Two extreme cases are as follows. In one case, the value of  $\tilde{\lambda}_{A,B}^+$  depends solely on  $A$  amino acid. In this case weighted averaging over  $B$  does not change the effective averaged value of  $\tilde{\lambda}_{A,B}^+$  and it will be equal to  $\tilde{\lambda}_{A,A}^+$ . Thus, an APED-like model may result in the symmetry breaking for amino acid  $A$  in this case. In another case the value of  $\tilde{\lambda}_{A,B}^+$  depends solely on  $B$  amino acid (values of  $\tilde{\lambda}_{B,B}^+$  for all  $B$ ). In this case weighted averaging over  $B$  will produce a value, which is close to the mean value of distribution of  $\tilde{\lambda}_{B,B}^+$ . Therefore, if the mean is close to zero, then an APED-like model will not work in this case. The same considerations apply to epimerization and depolymerization. Therefore, experimental measurements of enantioselectivity of polymerization, epimerization, and depolymerization of a large number of pairs of amino acids are needed in order to make a conclusion on applicability of APED-like models to chiral symmetry breaking in complex chemical systems.

We now consider the APED-3 model. Similar to the APED-2, we consider that interconversion between  $A$  and  $a$  enantiomers occurs only on N-terminal residue of the peptides (Plasson et al. 2004) and therefore the following epimerization reactions are included in APED-3 model:



The equations in  $\theta$  variables for a reduced system in the APED-3 model and expressed in APED-model parameters (Eq. (36)) are as follows:

$$\left\{ \begin{aligned} &\theta_A + \theta_{A^*} + 2(\theta_{Aa} + \theta_{AA}) + 3(\theta_{Aaa} + \theta_{AaA} + \theta_{AAa} + \theta_{AAA}) = \rho^{tot} \\ &-a\theta_A + \left(b - \frac{p(1+\alpha)\theta_A}{2}\right)\theta_{A^*} + h(2\theta_{AA} + \beta(2\theta_{Aa} + \theta_{Aaa} + \theta_{AAa}) + \theta_{AaA} + \theta_{AAA}) = 0 \\ &a\theta_A - \theta_{A^*} \left(b + \frac{p(1+\alpha)(\theta_A + \theta_{Aa} + \theta_{AA})}{2}\right) = 0 \\ &\frac{p\theta_{A^*}(\theta_A - (1+\alpha)\theta_{AA})}{2} + e(\theta_{Aa} - \gamma\theta_{AA}) + h(\beta\theta_{Aaa} + \theta_{AaA} - \theta_{AA}) = 0 \\ &\frac{p\theta_{A^*}(\alpha\theta_A - (1+\alpha)\theta_{Aa})}{2} + e(\gamma\theta_{AA} - \theta_{Aa}) + h(\beta\theta_{Aaa} + \theta_{AaA} - \beta\theta_{Aa}) = 0 \\ &\frac{p\theta_{A^*}\theta_{AA}}{2} + e\theta_{Aaa} - h\theta_{AaA} - e\gamma\theta_{AAa} = 0 \\ &\frac{p\theta_{A^*}\theta_{Aa}}{2} + e\theta_{AAa} - h\theta_{AaA} - e\gamma\theta_{AAa} = 0 \\ &\frac{p\alpha\theta_{A^*}\theta_{Aa}}{2} - e\theta_{AAa} - h\beta\theta_{AaA} + e\gamma\theta_{AAa} = 0 \\ &\frac{p\alpha\theta_{A^*}\theta_{AA}}{2} - e\theta_{Aaa} - h\beta\theta_{Aaa} + e\gamma\theta_{AAa} = 0 \end{aligned} \right. \quad (45)$$

The equations are linear in  $\theta_A, \theta_{AA}, \theta_{Aa}, \theta_{AAA}, \theta_{AaA}, \theta_{AaA}, \theta_{Aaa}$  and the solution is:

$$\begin{aligned} \theta_A &= \frac{4bhB\theta_{A^*}}{C}, & \theta_{AA} &= \frac{2bpD\theta_{A^*}^2}{C}, & \theta_{Aa} &= \frac{2bpE\theta_{A^*}^2}{C}, \\ \theta_{AAA} &= \frac{bp^2D^2\theta_{A^*}^3}{hBC}, & \theta_{AAa} &= \frac{bp^2DE\theta_{A^*}^3}{hBC}, & \theta_{AaA} &= \frac{bp^2E^2\theta_{A^*}^3}{hBC}, & \theta_{Aaa} &= \frac{bp^2DE\theta_{A^*}^3}{hBC} \end{aligned} \quad (46)$$

where:

$$\begin{aligned} B &= e(1 + \beta\gamma) + h\beta \\ C &= 4ahB - p(1 + \alpha)\theta_{A^*}(2h(e(1 + \beta\gamma) + h\beta) + p(h(\alpha + \beta) + e(1 + \alpha)(1 + \gamma))\theta_{A^*}) \\ D &= e(1 + \alpha) + h\beta \\ E &= h\alpha + e(1 + \alpha)\gamma \end{aligned} \quad (47)$$



Condition  $\rho^{tot} \rightarrow \infty$  is equivalent to condition  $C \rightarrow 0$  and that determines the maximum possible value of dimensionless quantity  $\epsilon = \left(\frac{p \theta_{A^*}}{a}\right)$  in such a system:

$$\epsilon^{max} = \left(\frac{p \theta_{A^*}}{a}\right)^{max} = \frac{4}{(1 + \alpha)(1 + \sqrt{1 + \delta})} \quad (48)$$

where  $\delta = \left(\frac{4a(D+E)}{hB(1+\alpha)}\right)$ . The value of  $\epsilon^{max} \approx \frac{2}{(1+\alpha)}$  for small values of  $\delta$  and  $\epsilon^{max}$  decreases  $\sim \frac{1}{\sqrt{\delta}}$  for large values of  $\delta$ . Comparing this expression with a similar expression for APED-2 model (Eq. (40)), we can note that  $\epsilon^{max}$  for the APED-3 model is smaller than that for the APED-2 model by a factor of  $\left(\frac{1+\sqrt{1+\delta}}{2}\right) > 1$ . Linearizing near the fixed point of a reduced system produces the following  $8 \times 8$  matrix  $M$  of coefficients of small variations  $\Delta_x = \{\Delta_A, \Delta_{A^*}, \Delta_{AA}, \Delta_{Aa}, \Delta_{AAA}, \Delta_{AAa}, \Delta_{AaA}, \Delta_{Aaa}\}$  around the fixed point:

$$M = \begin{pmatrix} \frac{p_+ \theta_{A^*}}{2} + a & -b + \frac{p_- \theta_A}{2} & -2h & 0 & -h & -h & -h\beta & -h\beta \\ \frac{p_- \theta_{A^*}}{2} - a & b + \frac{p_+ (\theta_A + \theta_{Aa} + \theta_{AA})}{2} & \frac{p_- \theta_{A^*}}{2} & \frac{p_- \theta_{A^*}}{2} & 0 & 0 & 0 & 0 \\ -\frac{p \theta_{A^*}}{2} & -\frac{p(\theta_A - (1 - \alpha)\theta_{AA})}{2} & h + e\gamma + \frac{p_+ \theta_{A^*}}{2} & e & -h & 0 & 0 & h\beta \\ \frac{p\alpha\theta_{A^*}}{2} & \frac{1}{2}p(\theta_{Aa} - \alpha(\theta_A + \theta_{Aa})) & e\gamma & e + h\beta + \frac{p_+ \theta_{A^*}}{2} & 0 & -h & h\beta & 0 \\ 0 & -\frac{p\theta_{AA}}{2} & -\frac{p\theta_{A^*}}{2} & 0 & h + e\gamma & 0 & 0 & e \\ 0 & -\frac{p\theta_{Aa}}{2} & 0 & -\frac{p\theta_{A^*}}{2} & 0 & h + e\gamma & e & 0 \\ 0 & -\frac{p\alpha\theta_{Aa}}{2} & 0 & \frac{p\alpha\theta_{A^*}}{2} & 0 & e\gamma & e + h\beta & 0 \\ 0 & -\frac{p\alpha\theta_{AA}}{2} & \frac{p\alpha\theta_{A^*}}{2} & 0 & e\gamma & 0 & 0 & e + h\beta \end{pmatrix} \quad (49)$$

where  $p_+ = p(1 + \alpha)$  and  $p_- = p(1 - \alpha)$ . Similar to the APED-2 model, the boundary conditions for  $\alpha$ ,  $\beta$ , and  $\gamma$  is determined when the value of the determinant of matrix  $M$  crosses zero. However, the resulting expression for the APED-3 model is substantially more complicated than that for the APED-2 model:

$$\begin{aligned} & p \theta_{A^*} (-1 + \alpha)(\alpha + \beta\gamma) \left( h(-p \theta_{A^*} \alpha + 2h\beta + p \theta_{A^*} \beta) \right. \\ & \quad + e(-p \theta_{A^*} (1 + \alpha)(-1 + \gamma) + 2(h + h\beta\gamma)) \left( h(2h\beta + p \theta_{A^*} (\alpha + \beta)) \right. \\ & \quad \left. \left. + e(p \theta_{A^*} (1 + \alpha)(1 + \gamma) + 2(h + h\beta\gamma)) \right) \right. \\ & \quad + 4ah(e + h\beta + e\beta\gamma) \left( hp \theta_{A^*} \alpha(\alpha - \beta) + h\beta(4h\beta + p \theta_{A^*} (\alpha + 3\beta))\gamma \right. \\ & \quad \left. \left. + e(4h\beta\gamma(1 + \beta\gamma) + p \theta_{A^*} ((-1 + \alpha)\alpha + (3 + \alpha)(\alpha + \beta)\gamma + (-1 + \alpha)\beta\gamma^2)) \right) \right) = 0 \end{aligned} \quad (50)$$

The equation can be solved numerically, for example, for  $\gamma$  as a function of  $\alpha$  and  $\beta$ , and it determines the boundary, provided that the constraint  $\frac{p\theta_{A^*}}{a} \leq \epsilon^{max}$  is satisfied. Figures 2 and 3 illustrate the values of  $\gamma$  for the values  $p\theta_{A^*} = a = e = h = 1$  for the APED-2 and the APED-3 models correspondingly.

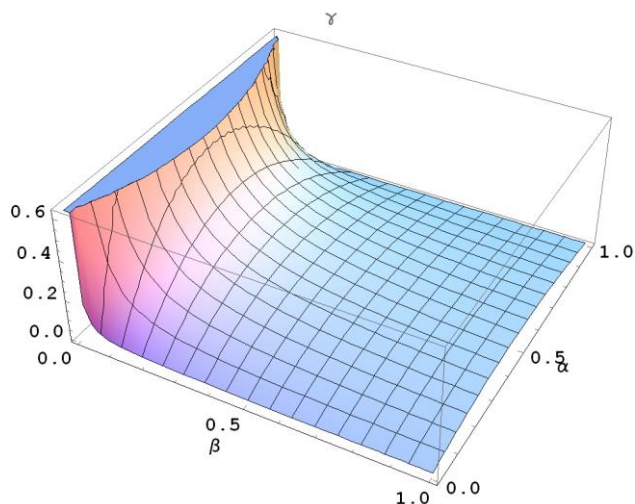


Figure 2. Boundary of symmetry breaking in the APED-2 model.

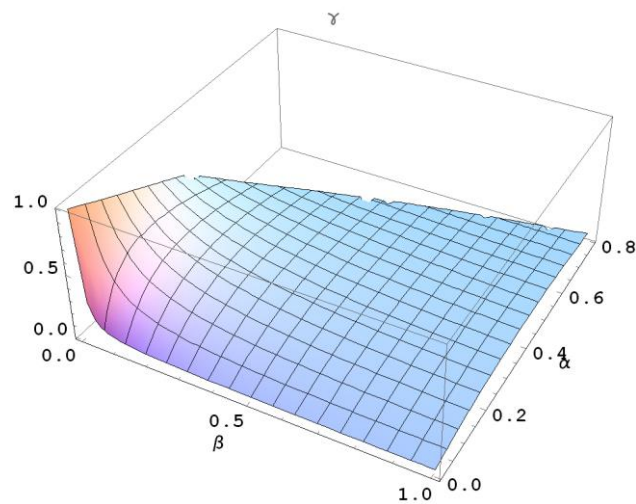


Figure 3. Boundary of symmetry breaking in the APED-3 model.

That edge of the graph as  $\alpha$  goes toward one on Figure 3 illustrates the end of possible values due to violation of the  $\frac{p\theta_{A^*}}{a} \leq \epsilon^{max}$  constraint (the value  $p\theta_{A^*} = 1$  cannot be achieved in the area where the graph is absent). As the figures 2 and 3 illustrate, the area, where the symmetry breaking can occur, shrinks in  $\alpha\beta\gamma$  parameter space when we go from the APED-2 to the APED-3 model. In addition, the situation of weak enantioselectivity in APED models corresponds to the area near point  $\alpha = \beta = \gamma = 1$ . Neither the APED-2, nor the APED-3 could produce a symmetry breaking near that point.

### Symmetry breaking in APED models in the presence of sedimentation

According to our calculations, adding sedimentation of an insoluble pair to APED-2 and APED-3 models unfortunately did not produce any substantial extension of the symmetry breaking area toward weak enantioselectivity values for  $\alpha$ ,  $\beta$ , and  $\gamma$ . We calculated several dependencies of maximum real part of all eigenvalues of linearized matrix of coefficients for the model APED-3 near the fixed point of the system and in the presence of sedimentation of the most promising pair ( $AA + Aaa$ ). That pair produces symmetry breaking with  $\eta \approx 0.90$  in model A100 without any catalytic synthesis. It appears that the combination of systems A100 and APED-3 produces symmetry breaking in some disjoint areas in the space of parameters, so that in one area the system must be like the A100 model with almost no APED-related enantioselectivity of  $\alpha$ ,  $\beta$ , and  $\gamma$  parameters and no epimerization, and in another area the system must be like the APED-3 with large enantioselectivity of  $\alpha$ ,  $\beta$ , and  $\gamma$  parameters and epimerization but indifferent to sedimentation in a certain window of sedimentation rates. We do not present the figures as they seem to provide no additional useful information.

## Symmetry breaking in the models with peptides up to length 3, catalytic synthesis of amino acids, pair formation and Noyes – Whitney sedimentation

We now want to consider a more realistic scenario when any pair of peptides may form a temporary (not peptide) bond as described by Eq. (29) and subsequently where some of the pairs are insoluble. We also would like to apply a more realistic reversible sedimentation process as described by Eq. (30). While it could be argued that formation of temporary pairs is proportional to collision probability and therefore the larger the peptide molecule the higher is the probability of its collision with some other peptide molecule, we did not want to take this into account in the current work. Rather we assumed that all the rates for temporary pair formations are the same. A model is fixed by assuming that a certain pair of substances is insoluble. However, following the discussion above we consider both Eq. (27) and Eq. (28) and set the parameters so that one combination is nearly insoluble and the other has good solubility. Nevertheless, it is important to note that the more realistic model of sedimentation process where there are many substances involved is needed before more precise results can be produced.

The results are as follows. First, the pairs (up to length three), which can produce the symmetry breaking in the absence of any catalytic synthesis, in these models are the same as in the models with simplified sedimentation and peptides up to length 5 ( $AA + aaA$ ,  $AA + aAa$ ,  $AA + aaA$ ). Second, if both Eq. (27) and Eq. (28) are taken into account, then the symmetry breaking can occur only within a certain window of concentrations. A similar result was obtained in (Konstantinov and Konstantinova 2013) for the model with peptide length up to two. We did not want to pursue this research further in the current work because more precise model of sedimentation of solution of many substances is needed in order to produce meaningful results.

It is also important to contrast the difference between fine tuning of some chemical rate coefficients, as required in some models, and adjusting the concentrations. The rate coefficients are what they are and as such they cannot be changed. Requiring that they have some specific values for the model to operate is fine-tuning. The concentrations, on the other side, can be easily adjusted by adding or removing solvent and, therefore, there is no fine-tuning in such models. For example, a standard attrition-enhanced deracemization also has some constraints on concentrations: if we add too much solvent, then there will be no sediment at all and attrition-enhanced deracemization will not work.

## Proposed Experiment

Given the calculations above, it would be interesting to devise an experiment, which would attempt to study chiral symmetry breaking along the considered models. Figure 4 shows the sketch of a proposed experiment of chiral symmetry breaking.

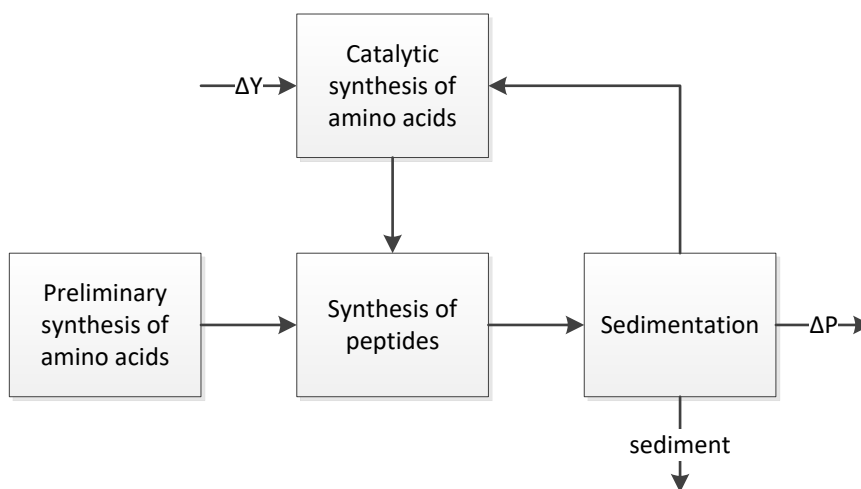


Figure 4. Layout of proposed experiment of chiral symmetry breaking

As discussed in (Avetisov and Goldanskii 1996), the rate of non-catalytic synthesis must be small in comparison to the rate of catalytic synthesis. Otherwise, non-catalytic synthesis will substantially dampen symmetry breaking. Therefore, a one-time non-catalytic synthesis of amino acids might be necessary in order to start the cycle. There are three major blocks in the main cycle of proposed experiment.

The first step is the catalytic synthesis of amino acids with some constant inflow of achiral matter  $\Delta Y$  and in the presence of a mixture of various peptides, some of which may be catalysts. It is crucial that non-catalytic synthesis is severely limited on this step.

The second step is the reversible synthesis of peptides. Control parameters on this step include forward and backward reaction rates (which affect the common ratio of geometric progression in the base model) as well as whether or not activation is used. We can look at the equivalent total concentration of amino acids in peptides of length  $n$ . If we consider that the total amount of amino acids in the system is the same ( $\rho_0$ ) then  $\Lambda_+$  and  $\Lambda_-$  determine  $\rho_1$  and  $\alpha$ . The total amount of amino acids in peptides of length  $n$  is  $\rho_n = n \rho_1 \alpha^n$  and then:

$$\sum_{n=1}^{\infty} \rho_n = \sum_{n=1}^{\infty} n \rho_1 \alpha^n = \rho_0 \quad (51)$$

Therefore:

$$\frac{\rho_n}{\rho_0} = n \alpha^{n-1} (1 - \alpha)^2 \quad (52)$$

Figure 5 shows the distribution of  $\frac{\rho_n}{\rho_0}$  for various  $n$  and  $\alpha$  with the same color corresponding to the same value of  $\alpha$ .

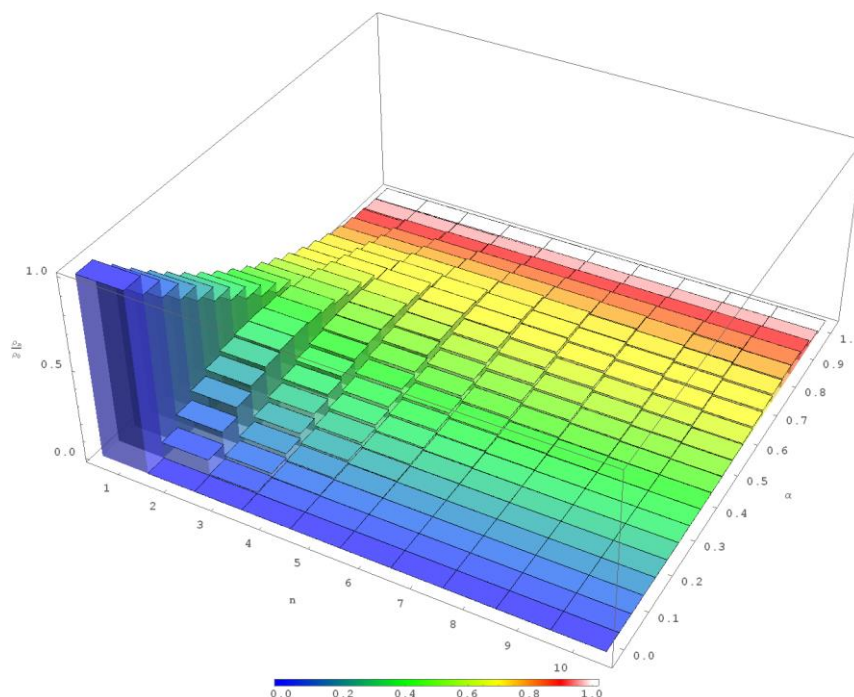


Figure 5. Total amount of amino acids in peptides of different length.

Following the result above, since there are no insoluble pairs, which can produce symmetry breaking for "peptides" of length one in the absence of catalytic synthesis, we can see from Figure 5 that the maximum amount of amino acids for peptides of length two and three is in the range  $\alpha \approx 0.3 - 0.6$  and for peptides with length four and five is in the range  $\alpha \approx 0.4 - 0.7$ . Concentrations of individual peptides are even smaller by a factor of  $2^{-n}$ . In addition, if there are many possible amino acids in the solution, then concentrations of individual peptides of length  $n$  further decrease by approximately  $\sim M^{-n}$  where  $M$  is the number of amino acids, which are present in the solution in substantial amounts. However, we would expect that there will be many similar pairs and catalysts with various amino acid content in that case. It is therefore our view that parameter  $\alpha$  should first be probed experimentally in the range  $\alpha \approx 0.3 - 0.7$ .

The third step is sedimentation process. It is very easy to discard the sediment both from experimental and geological points of views. For example, a device with much lower temperature at the bottom may allow continuous operation of the sedimentation step of the system. In this case the bottom layer with much lower temperature could be slowly washed away whereas the ingredients from the previous step are flowing to the much warmer top layer. As discussed in (Konstantinov and Konstantinova 2013), if more than one substance can form a sediment then it is possible that there is a certain window of concentrations in which the sedimentation may work as an additional enantioselective factor. Therefore, the control parameters of this step are the factors which affect the solubility, like temperature, total amount of solvent, some additives which may affect the overall solubility, etc.

The control parameters of the feedback are the percentages of what is recycled back to the catalytic synthesis block and what is considered as output  $\Delta P$ .

## Conclusion

It was argued that the number of substances and possible reactions among them must have been very large on prebiotic Earth and therefore that resulted in effective averaging over similar reaction channels including the reactions related to chiral symmetry breaking in amino acids. Effective averaging for catalytic synthesis of amino acids, epimerization, polymerization,

depolymerization, and sedimentation was considered. It was shown that such averaging results in weak effective enantioselectivity of catalytic synthesis, does not affect sedimentation, and needs additional research to assess the effect on epimerization, polymerization, and depolymerization.

A classes of models with formation of peptides up to length five, catalytic synthesis of amino acids, and a pair of substances, which form an insoluble substance, were considered. The Wolfram Mathematica program, which is capable of solving an arbitrary system of homogeneous chemical equations, was created. Using the program, all the matrix of 62 possible catalysts by 992 possible unique insoluble pairs was analyzed and all combinations, which result in substantial chiral symmetry breaking in the presence of very weak enantioselectivity of catalytic synthesis, were found.

It was shown that catalysts, which result in significant chiral symmetry breaking in the large number of possible insoluble pair combinations, have either zero or one amino acid of the opposite chirality: *A, aA, Aa, AA, aAA, AaA, Aaa, AAA, aAAA, Aaaa, AaaaA, AAAa, AAAAA, Aaaaa, AaaaaA, AaaaaA, AAAAAA*. It was shown that catalysts with length one and two work only with very limited number of possible insoluble pairs and therefore it was shown that the minimum interesting length of potential catalysts is three.

It was shown that an increase in peptide length from two to three in models based on activation, polymerization, epimerization depolymerization cycle results in a decrease of area in parameter space where chiral symmetry breaking can occur.

An experiment of chiral symmetry breaking is proposed. The experiment consists of a three-step cycle where the first step is catalytic synthesis of amino acids, the second step is reversible synthesis of peptides, and the third step is sedimentation of substances.

## References

- Andrews SS, Dinh T, Arkin AP (2009) Stochastic Models of Biological Processes, Encyclopedia of Complexity and System Science 9: 8730-8749
- Avetisov VA, Goldanskii VI (1996) Physical aspects of mirror symmetry breaking of the bioorganic world. Advances in Physical Sciences 166 (8): 873 - 891
- Blackmond DG (2011) The origin of biological homochirality. Phil Trans R Soc B 366: 2878-2884.
- Axel Brandenburg, Harry J. Lehto, Kirsi M. Lehto (2007) Homochirality in an Early Peptide World, Astrobiology 7 (5): 725-732
- Cook SA (1971) The Complexity of Theorem Proving Procedures. Proc Third Annual ACM Symposium on the Theory of Computing. ACM, New York, p 151-158
- Peter V. Coveney, Jacob B. Swadling, Jonathan A. D. Wattis, and H. Christopher Greenwell, Chem. Soc. Rev., 2012, 41, 5430-5446
- Dokoumetzidis A, Macheras P (2006) A century of dissolution research: From Noyes and Whitney to the Biopharmaceutics Classification System. Int J Pharm 321 (1-2): 1-11.
- Filippov AF (1988) Differential Equations with Discontinuous Righthand Sides, Kluwer Academic Publishers, Dordrecht
- Frank FC (1953) On spontaneous asymmetric synthesis. Biochim Biophys Acta 11: 459-463
- Yoshito Fujima, Masaya Ikunaka, Toru Inoue, and Jun Matsumoto (2006) Synthesis of (S)-3-(N-Methylamino)-1-(2-thienyl) propan-1-ol: Revisiting Eli Lilly's Resolution-Racemization-Recycle Synthesis of Duloxetine for Its Robust Processes. Org Process Res Dev 10 (5): 905-913

GNU General Public License, <https://gnu.org/licenses/gpl.html>

Hein JE, Huynh Cao B, Viedma C, Kellogg RM, Blackmond DG (2012) Pasteur's Tweezers Revisited: On the Mechanism of Attrition-Enhanced Deracemization and Resolution of Chiral Conglomerate Solids. *J Am Chem Soc* 134 (30): 12629-12636.

Iyer MS, Gigstad KM, Namdev ND, Lipton M (1996) Asymmetric catalysis of the Strecker amino acid synthesis by a cyclic dipeptide. *Amino Acids* 11 (3-4): 259-268

Konstantinov KK, Konstantinova AF (2013) New Concept of the Origin of Life on Earth. *Crystallography Reports* 58 (5): 697–709

Konstantinov KK, Konstantinova AF Konstantinova (2014), Influence of crystallization on formation of biomolecular homochirality on Earth. *Acta Cryst* 70: 1656

Letokhov, V. S.: 1975, On Difference of Energy Levels of Left and Right Molecules due to Weak Interactions, *Phys. Lett. A* 53, 275

Noorduyn WL, van der Asdonk P, Bode AAC, Meekes H, van Enkevort WJP, Vlieg E, Kaptein B, van der Meijden MW, Kellogg RM, Deroover G (2010) Scaling Up Attrition-Enhanced Deracemization by Use of an Industrial Bead Mill in a Route to Clopidogrel (Plavix). *Org Proc Res & Dev* 14: 908-911

Plasson R (2003) Dissertation. Université Montpellier II, Montpellier, France.

Plasson R, Bersini H, Commeyras A (2004) Recycling Frank: Spontaneous emergence of homochirality in noncatalytic systems. *PNAS* 101 (48): 16733–16738.

Plasson R, Kondepudi DK, Bersini H, Commeyras A, Asakura K (2007) Emergence of Homochirality in Far-From-Equilibrium Systems: Mechanisms and Role in Prebiotic Chemistry. Wiley InterScience, *Chirality* 19: 589–600.

Rein, D. W.: 1974, Some Remarks on Parity Violating Effects of Intramolecular Interactions, *J. Mol. Evol.* 4, 15.

Ribó JM, Crusats J, El-Hachemi Z, Moyano A, Blanco C, Hochberg D (2013) Spontaneous mirror symmetry breaking in the limited enantioselective autocatalysis model: abyssal hydrothermal vents as scenario for the emergence of chirality in prebiotic chemistry. *Astrobiology* 13(2): 132-142

Saint-Martin B, Julg A (1991) Influence of the interaction between asymmetry centers on the kinetics of racemization. *Journal of Molecular Structure: THEOCHEM*, 251 (6): 375–383.

Yukio Saito, Hiroyuki Hyuga, Complete Homochirality Induced by Nonlinear Autocatalysis and Recycling, *Phys Soc Jap* 2004, vol 73, no 1, pp 33-35.

Sczepanski JT, Joyce GF (2014) A cross-chiral RNA polymerase ribozyme. *Nature* 515: 440–442.

van der Meijden MW, Leeman M, Gelens E, Noorduyn WL, Meekes H, van Enkevort WJP, Kaptein B, Vlieg E, and Kellogg RM (2009) *Org Proc Res & Dev*, 13: 1195-1198

Isabelle Weissbuch, Leslie Leiserowitz, Meir Lahav, Stochastic "Mirror Symmetry Breaking" via Self-Assembly, Reactivity and Amplification of Chirality: Relevance to Abiotic Conditions *Top Curr Chem* (2005) 259: 123-165.

## Supplementary Materials

### Source code

All the source code along with results of numerical calculations, which were used to produce the results of the current article are available on GitHub at <https://github.com/kkkmail/CLM/>.

The system uses various methods for finding solutions. Two major methods are called “Evolution Run” and “Bifurcation Run”. The first method uses Mathematica differential equation solver (NDSolve) to find an evolutionary solution. The second method performs bifurcation analysis first by finding the fixed points of the system and then performing stability analysis by finding eigenvalues of the system linearized near fixed point(s). This method then attempts to maximize the real part of the eigenvalues by varying the parameters of the model, thus trying to find the values where bifurcation can occur. However, as the complexity of the system increases, numerical bifurcation analysis becomes less and less stable due to the following reasons. First, as the complexity increases, the overall polynomial power of the system grows rapidly. Finding the roots of polynomials numerically becomes a challenging task and it requires a substantial increase in precision, sometimes up to 50-100 digits. However, the worst part comes from the fact that a linearized matrix often has at least one zero or nearly zero eigenvalue. That makes the system insusceptible to the change of variables. Subsequently the minimization problem becomes too complex too fast and all together the solver often cannot find the solution. It is our current view that it is close to impossible to utilize bifurcation analysis for the systems with peptide lengths larger than three, especially if many parameters are used. Therefore, we used the evolutionary method in the current work.

Homogeneous and isotropic chemical systems can be fully described by specifying all the reactions in such systems. Each reaction is described by input and output reagents as well as the coefficient of the reaction. Therefore, given the description of the reactions, a system of ordinary differential equations, which describe such a system, can be generated. The system core does exactly that and it allows adding arbitrary reaction types by specifying such descriptors of reactions. A descriptor is effectively a stoichiometric matrix along with the coefficients of forward and backward reactions.

However, crystallization is not a chemical reaction but rather a complicated physical process, which depends on many conditions. Therefore, some simplifications were utilized to keep the solution time within reasonable limits while making sure that the obtained results have meaning. Two different models of crystallization are implemented in the system. One uses a simplified Noyes-Whitney equation, which treats the “rates” of forward and backward crystallization as a unified function of concentration of the relevant substance and a total amount of its sediment. This process is reversible, and it is slow to compute. An alternative method, which we called the “direct crystallization” utilizes the fact that crystallization is relatively fast in comparison to chemical reactions in consideration and that the system is designed to remove any sediment. The required assumption is that a substance, for which we apply “direct crystallization” method, must have very low solubility. A comparison between Noyes-Whitney type sedimentation with subsequent removal of the sediment and “direct crystallization” for peptides up to level three was performed to ensure that the latter method produces meaningful results. No significant discrepancies were found.

In addition, there are various statistical models used in the system. The purpose of these models is to generate some distributions of coefficients in large systems. This statistical part was not used in the current article.

System core is in subfolder **Kernel**. The modules to run various models are in subfolder **Run**. A detailed description has not yet been prepared but comments can be provided upon request.

### Main results of the article

The results used for the preparation of the current article are in the folder:

[https://github.com/kkkmail/CLM/tree/master/Results/ISSOL\\_201612](https://github.com/kkkmail/CLM/tree/master/Results/ISSOL_201612)



and below.

File [https://github.com/kkkmail/CLM/blob/master/Results/ISSOL\\_201612/CLM\\_ResultInfo.m](https://github.com/kkkmail/CLM/blob/master/Results/ISSOL_201612/CLM_ResultInfo.m) plots figures, which graphically represents the result of the current article. Parameter **PathList** must be set to point to the correct location of the data (described below) and then parameter **resultID** should be varied from 1 to 4 to plot the relevant figures. The naming conventions are described below.

The code for APED-related part of the article is in the folder:

[https://github.com/kkkmail/CLM/tree/master/Results/ISSOL\\_201612/Calc/APED](https://github.com/kkkmail/CLM/tree/master/Results/ISSOL_201612/Calc/APED).

There are only two files there and, therefore, no substantial naming convention was introduced.

## Naming and Folder conventions

All the raw calculations results are in folder **Calc** and precomputed data files are in **Data**.

Convention of naming files and folders is easier to explain using some examples. Consider a file (Name 1):

**CLM\_ER\_A\_Cmp\_Lzz\_En\_L5\_P0\_324\_Res\_03.m**

in the folder:

[https://github.com/kkkmail/CLM/tree/master/Results/ISSOL\\_201612/Calc/Evolution/L5/A/Cmp/g%3D010](https://github.com/kkkmail/CLM/tree/master/Results/ISSOL_201612/Calc/Evolution/L5/A/Cmp/g%3D010)

and a file (Name 2):

**CLM\_ER\_NA\_Cpm\_Lzz\_En\_L5\_lm030\_P0\_323\_R1.m**

in the folder:

[https://github.com/kkkmail/CLM/tree/master/Results/ISSOL\\_201612/Calc/Evolution/L5/NA/Cpm/g%3D010](https://github.com/kkkmail/CLM/tree/master/Results/ISSOL_201612/Calc/Evolution/L5/NA/Cpm/g%3D010)

Combined with the relevant part of the path they look like:

**Evolution/L5/A/Cmp/g=010/CLM\_ER\_A\_Cmp\_Lzz\_En\_L5\_P0\_324\_Res\_03.m**

**Evolution/L5/NA/Cpm/g=010/CLM\_ER\_NA\_Cpm\_Lzz\_En\_L5\_lm030\_P0\_323\_R1.m**

Enantioselective reactions are coded by a capital letter of the reaction type (**C** – catalysis, **L** – ligation, **E** – epimerization), followed by one or two lower case letter coding for enantioselectivity (**p** – positive, **m** – negative, **z** – no enantioselectivity, **n** – no reaction). Two letter coding is used when there are distinct forward and backward reactions. Epimerization effectively does not have a reverse reaction because it is symmetric to the relabeling of molecules.

The underscores serve as separators and so these folders/files have the following parts:

Name 1	Name 2	Description
Evolution	Evolution	Folder where all evolution models are stored
L5	L5	Models with peptides up to length 5
A	NA	<b>A</b> means models with activation. Models without activation have value <b>NA</b> .

Cmp	Cpm	Forward enantioselectivity of a given catalyst is slightly negative ( <b>minus</b> ) and backward enantioselectivity is slightly positive ( <b>plus</b> ). The other thermodynamically allowed combinations are <b>pm</b> , <b>zz</b> , or <b>nn</b> .
g=010	g=010	$ \gamma_+  =  \gamma_-  = 0.10$ , whereas the signs are determined by <b>Cmp</b> / <b>Cpm</b> coding.
CLM	CLM	Overall name of the model: <b>Chiral Life Model</b> .
ER	ER	Prefix to distinguish that the “ <b>Evolution Run</b> ” engine was used to produce results. The other available engine is “ <b>Bifurcation Run</b> ”.
A	NA	<b>A</b> means models with activation. Repeats folder coding for convenience. Models without activation have value <b>NA</b> .
Cmp	Cpm	Repeats folder coding of signs for forward / backward enantioselectivity for convenience.
Lzz	Lzz	<b>L</b> igation has <b>z</b> ero forward and <b>z</b> ero backward enantioselectivity.
En	En	<b>E</b> pimerization is <b>n</b> ot used.
L5	L5	Repeats folder coding for maximum peptide length (5) for convenience.
	lm030	Optional. If used, then encodes the value of backward ligation coefficient (0.30). Helps to distinguish the files where only that parameter was varied.
P0	P0	Page 0. Evolution models with length 5 are calculated in 9 pages (pages 0 through 8) because Mathematica's parallel engine does not go well with internally generated code. Therefore, instead of making parallel evaluations, 9+ independent Mathematica sessions were run simultaneously to speed up computations.
324	323	Major version number. The changes between versions 3.22 through 3.24 were mostly due to some restructuring and, therefore some files may have a run version inside slightly different from the code in the file name.
Res_03	R1	<b>Res_03</b> is the result number 3. Some of the results were discarded due to errors or other reasons. Each subsequent run of the same model was assigned a consecutive result number. Alternative coding is just the letter “R” followed by the result number, like <b>R1</b> .

Raw data files are combined in the relevant Excel files, which follow similar naming conventions. Each file has several sheets (the sheets that are not described were not used in the current work and may contain some obsolete data):

1. Sheet **Params** contains the parameters of the model.
2. Results of the calculations were copied into sheets **Cnn+**, **Cnn-** (combined into **CnnAll**), **Cpm+**, **Cpm-**, **Cmp+**, **Cmp-** (combined in **CpmAll**).
3. The result from sheets **CnnAll** and **CpmAll** were then copied into Mathematica modules, which are in the folder **Data**. These Mathematica modules are then used by **CLM\_ResultInfo.m** as a source for plotting various figures.