Emergence of homochirality in large molecular systems

Gabin Laurent[a,1], David Lacoste[a], and Pierre Gaspard[b]

[a]Gulliver, UMR CNRS 7083, École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, Paris Sciences et Lettres Research university, F-75231 Paris, France
[b]Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, B-1050 Brussels, Blegium.
[1]gabin.laurent@espci.fr

Abstract

The selection of a single molecular handedness, or homochirality across all living matter, is a mystery in the origin of life. Frank's seminal model showed in the fifties how chiral symmetry breaking can occur in non-equilibrium chemical networks. However, an important shortcoming in this classic model is that it considers a small number of species, while there is no reason for the prebiotic system, in which homochirality first appeared, to have had such a simple composition. Furthermore, this model does not provide information on what could have been the size of the molecules involved in this homochiral prebiotic system. Here, we show that large molecular systems are likely to undergo a phase transition towards a homochiral state, as a consequence of the fact that they contain a large number of chiral species. Using chemoinformatics tools, we quantify how abundant are chiral species in the chemical universe of all possible molecules of a given length. Then, we propose that Frank's model should be extended to include a large number of species, in order to possess the transition towards homochirality as confirmed by numerical simulations. Finally, using random matrix theory, we prove that large non-equilibrium reaction networks possess a generic and robust phase transition towards a homochiral state.

Keywords: homochirality, origin of life, prebiotic chemistry, random matrices, statistical physics

Life on Earth relies on chiral molecules−that is, species not superposable on their mirror images. A given biological molecule forms with its mirror image a pair of enantiomers. Homochirality precisely means the dominance of one member of the pair across the entire biosphere. For instance in our cells, biochemical reaction networks only involve left-handed (L-chiral) amino-acids and right-handed (D-chiral) sugars, but the reason for this absolute specificity escapes us and is one of the most fascinating questions in the origin of life.

The origin of homochirality comes with two questions and related observations: what caused the initial biais of one enantiomer over the other in the presumably racemic environment of the prebiotic world and how was this bias sustained and maintained as in today's biological world [1] ? It is believed that mineral surfaces on earth [2] or circularly polarized light in interstellar space [3] could explain the first observation, while models based on non-equilibrium reaction networks can explain the second observation [4–15].

However, there is an important shortcoming in the common discussions addressing the issue of homochirality−namely that they only consider a small number of chiral species, as in Frank's classic model [4], or in its first experimental realization more than forty years later by Soai et al. [16]. There is no reason to expect that the prebiotic world, in which homochirality first emerged, had such a simple and homogeneous chemical composition. Instead, it is more natural to assume that this composition was complex, heterogeneous and included a large number of chiral and achiral species. We show in this paper that generic non-equilibrium reaction networks possess a phase transition towards a homochiral state as a consequence of the fact that the number of chiral species becomes large.

Cross-Over between chiral and achiral chemical worlds

With this aim, we first ask how abundant are chiral species in the chemical universe of all possible molecules? It turns out that chirality is rare among molecules with a small number of atoms, but that possible chiral stereoisomers multiply as their number of atoms increases. Accordingly, we should expect a cross-over between the achiral world of small molecules and the chiral world of large molecules involved in chemical reaction networks. The cross-over should be characterized by some specific number of atoms where the fractions of achiral and chiral molecules become equal, as schematically depicted in (Fig. 1a). Beyond the cross-over, the chiral molecules are dominant over achiral ones. The issue of this cross-over is important, in particular, because it associates the emergence of homochirality with some molecular size. Starting with monosubstituted alkanes and alkanes for which an exact enumeration of stereoisomers is available [17, 18], we find that the cross-over measured in number of carbon atoms in these molecules, lies between 4.7 and 5.7 for monosubstituted alkanes, and between 8.4 and 9.5 for alkanes.

In the chemical universe of Fink and Reymond [19], which contains all virtual molecules of a given number of heavy atoms (*i.e.* atoms heavier than hydrogen) and satisfying some basic set of rules of chemistry, we find a cross-over at 8.5 heavy atoms. Since a similar study was not available for real molecules, we turn to the chemical database PubChem [20]. From the raw data of this large database, we find that a cross-over occurs for molecules of 9.4 heavy atoms (SI Appendix, section S1). However, since many chiral molecules do not have all their enantiomers or stereoisomers, we have also analyzed an expanded database, in which every chiral molecule contains either all its enantiomers or all its stereoisomers. In Fig. 1b, the fractions of chiral and achiral species are shown for the case of the database expanded in stereoisomers. Our results for the various estimates for the cross-over are gathered in Table 1 and extracted from Fig. 1b and SI Appendix, Figs. S1, S2, S3 and S4. Remarkably, irrespective of the precise procedure to generate and analyze the database and regardless of the precise composition of the molecules, the cross-over between the achiral and chiral worlds occurs for a number of heavy atoms of the order of about 10. The main consequence of this cross-over is that the stereoisomer distribution goes from unimodal (with a maximum for achiral molecules) to bimodal (with maxima for opposite enantiomers) as the length of molecules increases. This emerging bimodality is potentially susceptible to induce a chiral symmetry breaking.

Table 1: Positions of the cross-over n_1 and n_2, measured in terms of number of carbon atoms, in the study of alkanes and monosubstituted alkanes, or in terms of number of heavy atoms in the other studies. The estimate n_1 (resp. n_2) is obtained by counting once (resp. twice) the pairs of enantiomers.

Data	n_1	n_2
monosubstituted alkanes stereoisomers	5.7	4.7
alkanes stereoisomers	9.5	8.4
Chemical Universe	8.5	–
PubChem database using raw data	9.4*	–
PubChem database using generated enantiomers	12.7	6.7
PubChem database using generated stereoisomers	8.1	6.4

* The cross-over for PubChem raw data occurs between n_1 and n_2 because not all enantiomers of a given species are present in the database.

Spontaneous symmetry breaking into a chiral state

We now come to our central point−namely, on how to explain the emergence of homochirality from the multiplication of chiral species in non-equilibrium reaction networks. Specifically, we consider a reaction network involving achiral and chiral species described by the concentration vector $\mathbf{c}$, which contains the vector $\mathbf{c}_\mathrm{D}$ (resp. $\mathbf{c}_\mathrm{L}$) for the N_C D-enantiomers (resp. for the N_C L-enantiomers) and the vector $\mathbf{c}_\mathrm{A}$ for the remaining N_A achiral species. In such a system, the evolution of the concentrations is ruled by

$$\frac{d\mathbf{c}}{dt} = \mathbf{F}(\mathbf{c}) + \frac{1}{\tau}(\mathbf{c}_0 - \mathbf{c})\,, \qquad (1)$$

where $\mathbf{c_0}$ is the concentration vector of the species supplied from the environment at the rate $1/\tau$ and responsible for driving the system out of equilibrium and $\mathbf{F}(\mathbf{c}) = \boldsymbol{\nu}\cdot\mathbf{w}(\mathbf{c})$ are the reaction rates with specific chiral symmetry (Eq. (14) in Materials and Methods), which need not obey mass-action law. In this expression, $\boldsymbol{\nu}$ is the matrix of stoichiometric coefficients, $\mathbf{w}(\mathbf{c})$ the set of net reaction rates. After reaction, the species in excess are flowing out of the system at the same rate $1/\tau$ as for the supply, so that τ represents the mean residence time of the species in the system.

The stability of these equations may be characterized by linearizing them about the racemic state, which is

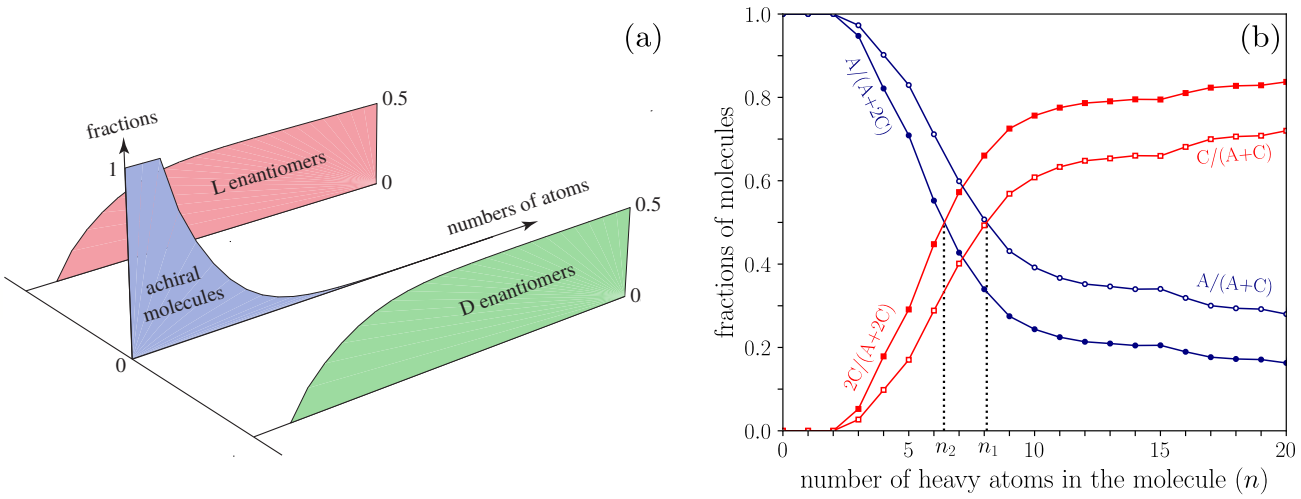


Figure 1: Cross-over between the achiral and chiral chemical worlds. (a) A three-dimensional schematic representation of the fractions of possible achiral and chiral molecules as function of their number of atoms. Every chiral molecule appears as mirror-reflected D- and L-enantiomers. (b) A two-dimensional representation of the fractions of achiral (circles) and chiral (squares) molecules in an expanded PubChem database containing all the stereoisomers of molecules with at most $n \leq 20$ heavy atoms. Filled symbols correspond to counting stereoisomers twice, while empty symbols correspond to counting them once in the fractions. Error bars are smaller than symbol sizes and the cross-over occurs at $n_2 \simeq 6.4$ and $n_1 \simeq 8.1$ depending on which counting is considered (SI Appendix, section S1).

defined by the condition $\mathbf{c}_\mathrm{D} = \mathbf{c}_\mathrm{L}$, and is assumed to exist in a steady state. With the small parameter $\delta\mathbf{x}$, where $\mathbf{x}$ denotes the chiral enantiomeric excess $\mathbf{x} \equiv \frac{1}{2}(\mathbf{c}_\mathrm{L} - \mathbf{c}_\mathrm{D})$, we obtain

$$\frac{d}{dt}\delta\mathbf{x} = \left(\mathbf{J} - \frac{1}{\tau}\right) \cdot \delta\mathbf{x} + \frac{1}{\tau}\,\delta\mathbf{x}_0\,, \qquad (2)$$

where $\mathbf{J}$ represents the Jacobian matrix deduced from the kinetic equations (1). The racemic mixture is unstable if at least one of the eigenvalues of the matrix $\mathbf{M} = \mathbf{J} - \frac{1}{\tau}\mathbf{I}$ (with $\mathbf{I}$ the identity matrix) has a positive real part. In a large reaction network, the reaction rates may take very different values, so that the matrix $\mathbf{J}$ may be treated as a random matrix [21–23]. The simplest model is to assume that the elements of this matrix are independent and identically distributed real numbers (but not necessarily Gaussian distributed) of mean value μ and variance σ^2 [24]. When $\mu = 0$, random matrix theory shows that the complex eigenvalues are uniformly distributed in a disk of radius $\sigma\sqrt{N_\mathrm{C}}$ in the limit of large values of N_C [25]. When $\mu \neq 0$, we find that there exists a single and isolated eigenvalue, which is equal to μN_C, and the corresponding eigenvector has uniform components to dominant order (SI Appendix, section S2). Two possible mechanisms for the instability of the racemic state then emerge for large N_C (SI Appendix, section S3). Either (i) the instability occurs due to the isolated eigenvalue as illustrated in Fig. 2d,e and f; otherwise (ii) it occurs due to the eigenvalues located on the edge of the circle (which may be real or complex valued) as illustrated in Fig. 2a,b and c. It follows from this that when $\mu > 0$ and $N_\mathrm{C} \geq \max\{1/(\tau\mu), (\sigma/\mu)^2\}$, the system becomes unstable by the first mechanism where all species become simultaneously unstable, and when $(\sigma/\mu)^2 \geq N_\mathrm{C} \geq 1/(\tau\sigma)^2$, the system becomes unstable by the second mechanism and in this case only a subpart of all the species become unstable at the transition. In such cases, random matrix theory predicts that as N_C becomes large, these mechanisms of instability become more and more likely. This is confirmed by the shape of the probability for the racemic state to be unstable versus N_C shown in Fig. 2g for both mechanisms. If the matrix elements are statistically correlated, the non-dominant eigenvalues may have a different distribution, but the isolated eigenvalue behaves similarly.

The effect of chiral species multiplication

In order to show that this general scenario can be realized in practice in a non-equilibrium reaction network, we now introduce generalizations of Frank's model, in which we have multiplied the number of chiral and achiral species

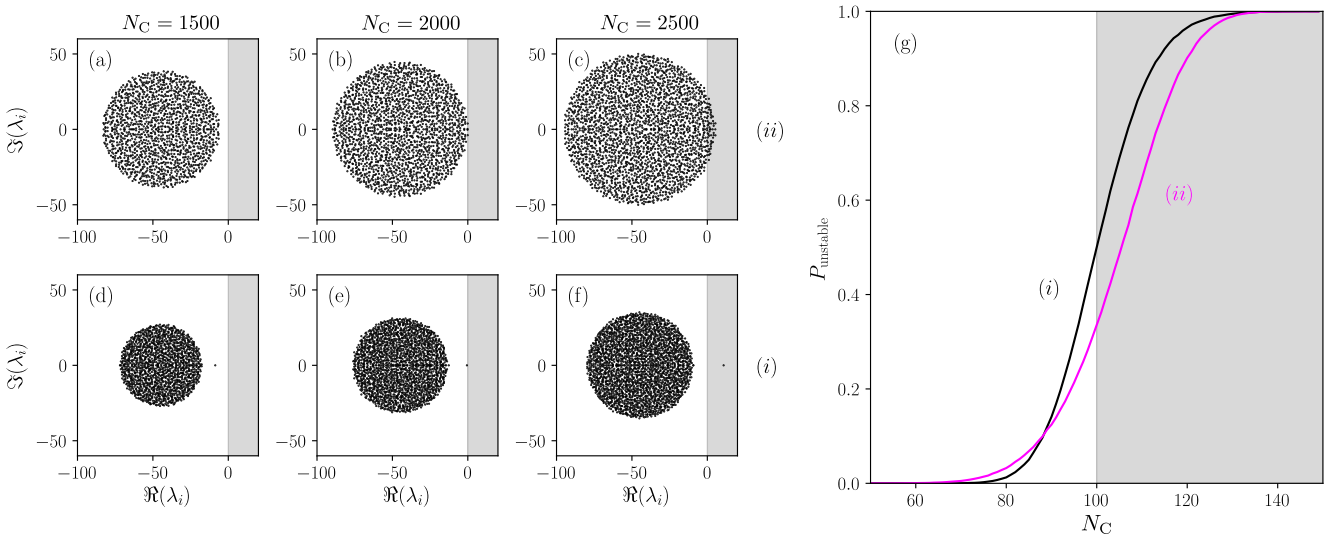


Figure 2: Distribution of eigenvalues and corresponding instability criterion. Subfigures (a), (b) and (c) represent the eigenvalues of the random matrix $\mathbf{M}$ with Gaussian entries with $\mu = 0$, $\sigma = 1$ and $\tau = 1/\sqrt{2000}$, where the instability mechanism (ii) occurs as N_C increases, *i.e.* the edge of the Girko's circle densely filled by eigenvalues crosses the zero real axis. Subfigures (d), (e) and (f) depict the mechanism (i) where the zero real axis is crossed by an isolated eigenvalue. In this case, the matrix $\mathbf{M}$ is a random matrix with Gaussian entries of parameters $\mu = 1/\sqrt{2000}$, $\sigma = 0.7$ and $\tau = 1/\sqrt{2000}$. Grey colored area represents the positive real part zone of the complex plane, *i.e.* the area in which an eigenvalue causes the instability. Sizes of random matrices $\mathbf{M}$ are $N_C = 1500$ for subfigures (a) and (d), $N_C = 2000$ for subfigures (b) and (e) and $N_C = 2500$ for subfigures (c) and (f). Subfigure (g) represents the probability for the racemic state of the chemical system to be unstable as a function of N_C. Black curve represents the mechanism (i) for a Gaussian random matrix $\mathbf{M}$ with $\mu = 1/\sqrt{100}$, $\sigma = 0.7$ and $\tau = 1/\sqrt{100}$. Magenta curve represents the mechanism (ii) for a Gaussian random matrix $\mathbf{M}$ with $\mu = 0$, $\sigma = 1$ and $\tau = 1/\sqrt{100}$. Grey colored area represents the theoretical instability area, which is $N > 100$ for both mechanisms here, given the choice of parameters μ, σ and τ. The statistics has been carried out over $100,000$ realizations of random matrices. Note also that the two curves in subfigure (g) for the two mechanisms, were drawn for different parameters μ and σ.

and we have assumed an arbitrary assignation L or D to each enantiomer. We also include reverse reactions in order to guarantee the compatibility with the existence of an equilibrium state even though the system is driven out of equilibrium. It is essential that the system be driven out of equilibrium in order for chirality to be maintained. We thus assume that the system is thermodynamically open, due to fluxes of matter in and out of the system.

Let us also suppose that species entering the autocatalytic system are achiral but of high free energy, while the achiral species produced by the reactions involving the two D- and L-enantiomers have a lower free energy. In this regard, the achiral species $\{A_a\}_{a=1}^{N_A}$ are of high free energy, and the achiral species $\{\tilde{A}_a\}_{a=1}^{\tilde{N}_{\tilde{A}}}$ of low free energy. The reaction networks are given by the following reactions:

$$\mathrm{A}_a + \mathrm{E}_i \rightleftharpoons \mathrm{E}_j + \mathrm{E}_k \,, \tag{3}$$

$$\mathrm{A}_a + \bar{\mathrm{E}}_i \rightleftharpoons \bar{\mathrm{E}}_j + \bar{\mathrm{E}}_k \,, \tag{4}$$

$$\mathrm{E}_i + \bar{\mathrm{E}}_j \rightleftharpoons \tilde{\mathrm{A}}_b + \tilde{\mathrm{A}}_c \,, \tag{5}$$

where the enantiomer species are either $\mathrm{E}_m = \mathrm{D}_m$ and $\bar{\mathrm{E}}_m = \mathrm{L}_m$, or $\mathrm{E}_m = \mathrm{L}_m$ and $\bar{\mathrm{E}}_m = \mathrm{D}_m$ for each enantiomeric pair $m = i, j, k = 1, 2, \ldots, N_C$; $a = 1, 2, \ldots, N_A$; and $b, c = 1, 2, \ldots, \tilde{N}_{\tilde{A}}$. Equations (3)-(4)-(5) define a total of $2^{N_C - 1}$ inequivalent reaction networks differing by the permutations of D- and L-enantiomers for some enantiomeric pairs. For given reaction rates, all these networks manifest similar dynamical behaviors. Among them, the network with $\mathrm{E}_m = \mathrm{D}_m$ and $\bar{\mathrm{E}}_m = \mathrm{L}_m$ for all the pairs $m = 1, 2, \ldots, N_C$ is the direct generalization of Frank's model, considered below.

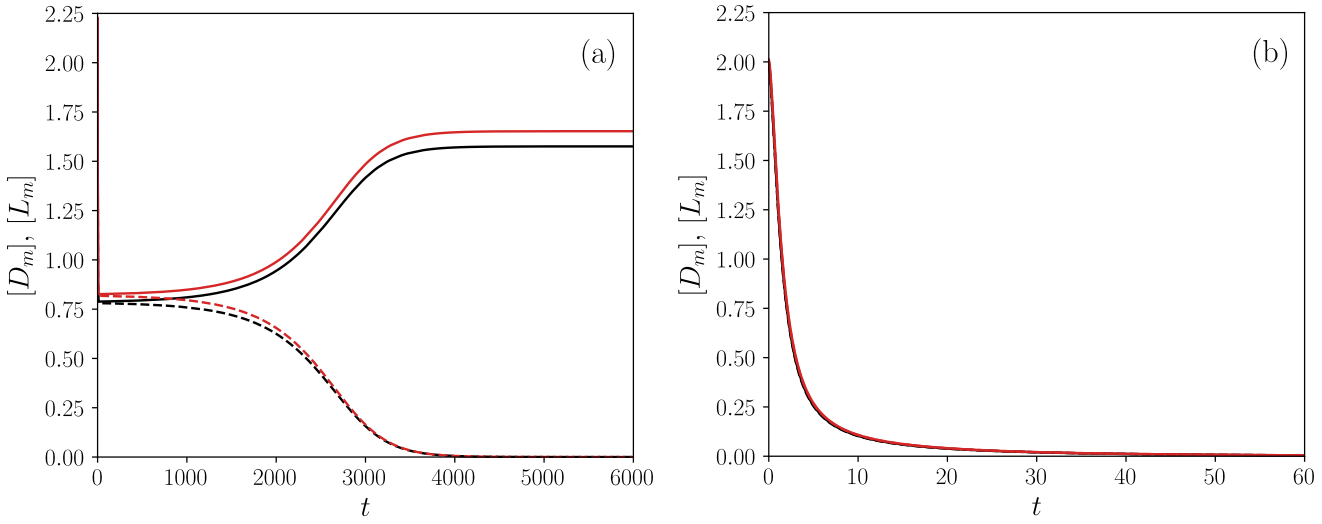


Figure 3: Dynamical simulations of the autocatalytic network (21)-(22)-(23). Typical time evolution of two species contained in the autocatalytic network as a function of time (a) above the threshold concentration A_0 and (b) below it. The solid lines represent one of the two enantiomers for a given species and the dashed line, the other enantiomer. Both simulations were carried out with an initial enantiomeric excess $\epsilon = 10^{-2}$ and D- and L-enantiomers concentrations of all chiral species were initialized at $D_0 = 2 + \epsilon$ and $L_0 = 2 - \epsilon$. The unactivated achiral specie was initialized at $\tilde{A}_0 = 0$ and the activated one at $A_0 = 80$ in (a) and $A_0 = 45$ in (b). All the constants k_{+ijk} and $\tilde{k}_{-ij}$ follow a log-normal distribution of parameters $\mu = -10.02$ and $\sigma = 1.27$ (*i.e.*, corresponding to a log-normal distribution with $\langle k_+ \rangle = \langle \tilde{k}_- \rangle = 10^{-4}$ and $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 2 \times 10^{-4}$), with $\tilde{k}_{ij} = \tilde{k}_{ji}$ to satisfy the mirror symmetry described in Eq. (14). The number of chiral species was set up to $N_C = 20$.

For our numerical implementation of this model, we focus on the fully irreversible regime, in which reactions (3), (4), and (5) only proceed forward due to the supply of achiral species with high free energy at the same concentration A_0. Thus, there are two main control parameters in this model: the supply concentration A_0 and the residence time τ.

For one particular realization of these rate constants, Fig. 3a shows the evolution of the concentrations of the species present in the system as function of time above the threshold concentration A_0, while Fig. 3b shows the case below threshold. In Fig. 3a, we see that on long times the system converges towards a steady state, which is homogeneous and chiral. Only two species have been shown in these figures for clarity but the time evolution of their concentrations is typical of the evolution of all the other species: on long times, only one enantiomer is present, which is of the same chirality for all the species, while the other enantiomer reaches a vanishing concentration. Instead, in Fig. 3b, all the species converge on long times towards a vanishing concentration.

The case where all the rate constants would be identical can be treated analytically as done in (SI Appendix, section S5), so let us now instead assume that the rate constants k_+ of reactions (3) and (4) are taken according to a log-normal distribution [26]. We find that the spontaneous chiral symmetry breaking happens if the following criterion is satisfied,

$$\langle k_+ \rangle \, \tau \, N_A \, A_0 > \frac{2}{N_C(N_C + 1)} \,, \tag{6}$$

i.e., the residence time multiplied by the total concentration $N_A \, A_0$ of achiral species supplied to the system must exceed a threshold determined by the average rate constant $\langle k_+ \rangle$ of autocatalysis and the number N_C of chiral species in the reaction network. We have tested this result by performing a linear stability analysis of the racemic steady state. Simulations show that instability is due to an isolated top eigenvalue, and confirm the criterion (6) when the distribution of the rate constants is not too broad as shown in the inset of Fig. 4. For a very broad distribution, the threshold is pushed towards higher value than predicted by Eq. (6), but the transition still occurs at sufficiently large N_C (SI Appendix, Fig. S5). Importantly, the transition becomes sharp as the number of chiral species increases as shown in Fig. 4.

Thus, the random matrix theory argument holds and the mechanism (i) is confirmed, although the eigenvalues of the Jacobian matrix do not cover uniformly a circle (SI Appendix, Fig. S6) due to the difference of statistics between the diagonal and the off-diagonal elements (SI Appendix, section S5). In this case, the mechanism (ii) is found not to be relevant. Furthermore, the multiplication of chiral species is also multiplying the number of reaction networks manifesting similar chiral symmetry breaking, but with either the D- or the L-enantiomer for the different chiral species.

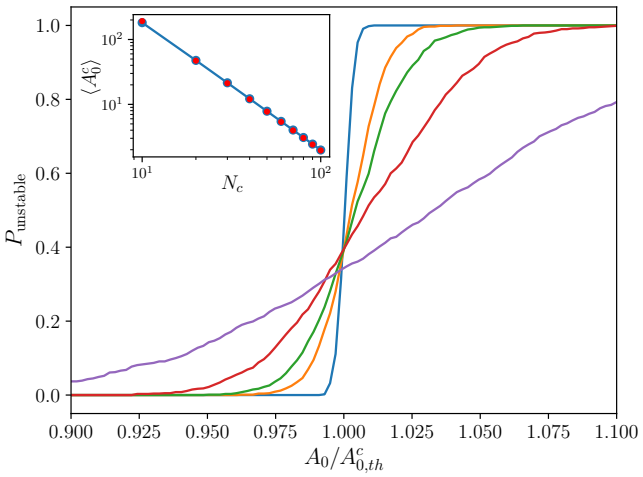


Figure 4: Probability of instability of the racemic state for the generalized Frank model (21)-(22)-(23). Probability of the initial racemic state to be unstable by mechanism (i) as function of the normalized value of the control parameter A_0 for the expanded Frank's model in the irreversible regime and for different values of number of chiral species N_C: $N_C = 10$ (magenta), $N_C = 20$ (red), $N_C = 30$ (green), $N_C = 40$ (yellow) and $N_C = 100$ (blue). The control parameter A_0 has been normalized by the theoretical threshold at the transition, defined by the equality in the inequality of Eq. (6). An average over a thousand realizations of the rate constants following a log-normal distribution such that $\langle k_+ \rangle = \langle \tilde{k}_- \rangle = 10^{-4}$ and $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 2 \times 10^{-4}$ has been performed. Inset: comparison between the observed control parameter value A_0 at the transition (red circles) with the theoretical prediction given by Eq. (6) (blue solid line) after averaging over 100 realizations of the rate constants.

Discussion

Let us now come back to our evaluation of the cross-over to chirality in the context of the origin of life. First of all, we observe that among the 20 natural amino acids, the only one that is achiral−*i.e.*, glycine−is also the smallest containing only 5 heavy atoms. Secondly, the first chiral molecule observed in space is propylene oxide, containing 4 heavy atoms [27]. Thirdly, among the 11 carboxylic acids of the Krebs cycle, the majority of them, 9 lie in the range between 8 and 13 heavy atoms, only two of them are smaller, acetate and pyruvate. In fact, the Krebs cycle appears to function precisely at the border between the world of achiral and small molecules and that of large and chiral molecules [28]. The emergence of the Krebs cycle thus represents a major step, which facilitates the synthesis of a large number of long chiral molecules [29].

We presented a scenario that explains why a large complex molecular system tends to become chiral. Based on fundamental properties of phase transitions, confirmed by numerical simulations of a non-equilibrium reaction network, our scenario is both general and robust. Details of the reaction network should not matter, nor the precise way in which the system is driven out of equilibrium, provided the system is large enough. In addition, the homochiral state of our model does not need to be all D or L across all species, in agreement with the observation that for instance, amino-acids are L-chiral, but sugars are D-chiral. Moreover, our reaction scheme needs not satisfy mass-action law, and there is also no requirement that the system be fully well-mixed, it could be compartmentalized. As an illustration of this idea, we study two diffusively coupled chemical reactors, identical to the one considered so far (SI Appendix, section S6). For low coupling, the two compartments undergo separately the homochiral transition, while at high coupling, they reach the same homochiral state [13]. Compartmentalized systems enrich the scenarios for the transition to homochirality, because on one hand, as shown recently, compartmentalization significantly broadens the diversity of available autocatatalytic networks which can be built with a limited number of compounds [30], and on the other hand, in such systems, the number of species is effectively increased, which favors our mechanism.

Our scenario thus offers a universal pathway towards homochirality potentially unifying many previous approaches on this issue. In the context of the origin of life, we also find that there is no need for very long and complex molecules for this homochiral state to emerge. The transition can already occur in a prebiotic world containing molecules with about 10 heavy atoms.

Materials and Methods

Chemoinformatics of chirality A study of the scaling of the number of chiral and achiral species with the

number of atoms for monosubstituted alkanes and alkanes stereoisomers is provided in SI Appendix section S1, together with details on the chemoinformatic analysis of the PubChem database [20].

Chiral symmetry breaking for a general reaction scheme The spontaneous symmetry breaking of chirality can be described in the framework of kinetics. The reaction network is supposed to involve achiral and chiral species at the concentrations

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}_\mathrm{D} \\ \mathbf{c}_\mathrm{A} \\ \mathbf{c}_\mathrm{L} \end{pmatrix}, \tag{7}$$

which respectively denote the concentrations of D-enantiomers, achiral species, and L-enantiomers. If N_D, N_A, and N_L denote the respective numbers of these species, the system is described in terms of $N_\mathrm{S} = N_\mathrm{D} + N_\mathrm{A} + N_\mathrm{L} = N_\mathrm{A} + 2N_\mathrm{C}$ concentrations with $N_\mathrm{C} \equiv N_\mathrm{D} = N_\mathrm{L}$.

In an open system, the time evolution of these concentrations is ruled by the N_S kinetic equations in Eq. (1) with $\mathbf{F}(\mathbf{c}) = \boldsymbol{\nu} \cdot \mathbf{w}(\mathbf{c})$, expressed in terms of the matrix $\boldsymbol{\nu}$ of stoichiometric coefficients, the set $\mathbf{w}(\mathbf{c})$ of net reaction rates $w_r(\mathbf{c}) = w_{+r}(\mathbf{c}) - w_{-r}(\mathbf{c})$ with $r = 1, 2, ..., N_\mathrm{R}$, the supply concentrations $\mathbf{c}_0$, and the residence time τ of the species in the system. The system is closed if the residence time is infinite, in which case the last term drops in Eq. (1.

The mirror symmetry of the system corresponds to the following exchange of concentrations of D- and L-enantiomers,

$$\mathbf{S_c} \cdot \mathbf{c} = \begin{pmatrix} 0 & 0 & \mathbf{I} \\ 0 & \mathbf{I} & 0 \\ \mathbf{I} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{c}_\mathrm{D} \\ \mathbf{c}_\mathrm{A} \\ \mathbf{c}_\mathrm{L} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_\mathrm{L} \\ \mathbf{c}_\mathrm{A} \\ \mathbf{c}_\mathrm{D} \end{pmatrix}, \tag{8}$$

written in terms of the $N_\mathrm{S} \times N_\mathrm{S}$ matrix such that $\mathbf{S}_\mathbf{c}^2 = \mathbf{I}$, where $\mathbf{I}$ denotes the corresponding identity matrix. Since the rate constants take equal values for mirror-symmetric reactions, the reaction rates have the symmetry

$$\mathbf{w}(\mathbf{S_c} \cdot \mathbf{c}) = \mathbf{S_w} \cdot \mathbf{w}(\mathbf{c}) \tag{9}$$

with some $N_\mathrm{R} \times N_\mathrm{R}$ matrix satisfying $\mathbf{S}_\mathbf{w}^2 = \mathbf{I}$. As a consequence of the mirror symmetry, the $N_\mathrm{S} \times N_\mathrm{R}$ matrix of stoichiometric coefficients obeys the following symmetry relation

$$\mathbf{S_c} \cdot \boldsymbol{\nu} \cdot \mathbf{S_w} = \boldsymbol{\nu}\,. \tag{10}$$

We note that the kinetic equations may also be written in the following form,

$$\frac{d\mathbf{c}_\mathrm{D}}{dt} = \mathbf{F}_\mathrm{D}(\mathbf{c}_\mathrm{D}, \mathbf{c}_\mathrm{A}, \mathbf{c}_\mathrm{L}) + \frac{1}{\tau}(\mathbf{c}_\mathrm{D0} - \mathbf{c}_\mathrm{D})\,, \tag{11}$$

$$\frac{d\mathbf{c}_\mathrm{A}}{dt} = \mathbf{F}_\mathrm{A}(\mathbf{c}_\mathrm{D}, \mathbf{c}_\mathrm{A}, \mathbf{c}_\mathrm{L}) + \frac{1}{\tau}(\mathbf{c}_\mathrm{A0} - \mathbf{c}_\mathrm{A})\,, \tag{12}$$

$$\frac{d\mathbf{c}_\mathrm{L}}{dt} = \mathbf{F}_\mathrm{L}(\mathbf{c}_\mathrm{D}, \mathbf{c}_\mathrm{A}, \mathbf{c}_\mathrm{L}) + \frac{1}{\tau}(\mathbf{c}_\mathrm{L0} - \mathbf{c}_\mathrm{L})\,, \tag{13}$$

where the mirror symmetry is expressed as

$$\mathbf{F}_\mathrm{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{F}_\mathrm{L}(\mathbf{z}, \mathbf{y}, \mathbf{x}) \qquad \text{and} \qquad \mathbf{F}_\mathrm{A}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{F}_\mathrm{A}(\mathbf{z}, \mathbf{y}, \mathbf{x})\,. \tag{14}$$

The symmetry can be explicitly broken by input concentrations such that $\mathbf{S_c} \cdot \mathbf{c}_0 \neq \mathbf{c}_0$. However, the equations remain symmetric if the condition $\mathbf{S_c} \cdot \mathbf{c}_0 = \mathbf{c}_0$ holds. In this case, the racemic mixture characterized by equal concentrations of D- and L-enantiomers,

$$\mathbf{c}_\mathrm{D} = \mathbf{c}_\mathrm{L}\,, \tag{15}$$

is maintained during the time evolution of the reaction network, if the dynamics

$$\frac{d\mathbf{c}_\mathrm{A}}{dt} = \mathbf{F}_\mathrm{A}(\mathbf{c}_\mathrm{D}, \mathbf{c}_\mathrm{A}, \mathbf{c}_\mathrm{D}) + \frac{1}{\tau}(\mathbf{c}_\mathrm{A0} - \mathbf{c}_\mathrm{A})\,, \tag{16}$$

$$\frac{d\mathbf{c}_\mathrm{D}}{dt} = \mathbf{F}_\mathrm{D}(\mathbf{c}_\mathrm{D}, \mathbf{c}_\mathrm{A}, \mathbf{c}_\mathrm{D}) + \frac{1}{\tau}(\mathbf{c}_\mathrm{D0} - \mathbf{c}_\mathrm{D})\,, \tag{17}$$

is stable in the racemic subspace (15). In order to investigate this issue, we introduce the variables

$$\mathbf{x} \equiv \frac{1}{2}(\mathbf{c}_\mathrm{L} - \mathbf{c}_\mathrm{D}) \tag{18}$$

characterizing deviations with respect to the racemic subspace and we perform the linear stability analysis for infinitesimal deviations $\delta\mathbf{x}$ with respect to the racemic subspace. These deviations are ruled by the following set of linear equations:

$$\frac{d}{dt}\delta\mathbf{x} = \left(\mathbf{J}_\mathrm{DD} - \mathbf{J}_\mathrm{DL} - \frac{1}{\tau}\right) \cdot \delta\mathbf{x} + \frac{1}{\tau}\,\delta\mathbf{x}_0\,, \tag{19}$$

where

$$\mathbf{J}_\mathrm{DD} \equiv \frac{\partial \mathbf{F}_\mathrm{D}}{\partial \mathbf{c}_\mathrm{D}} = \frac{\partial \mathbf{F}_\mathrm{L}}{\partial \mathbf{c}_\mathrm{L}} \qquad \text{and} \qquad \mathbf{J}_\mathrm{DL} \equiv \frac{\partial \mathbf{F}_\mathrm{D}}{\partial \mathbf{c}_\mathrm{L}} = \frac{\partial \mathbf{F}_\mathrm{L}}{\partial \mathbf{c}_\mathrm{D}}\,. \tag{20}$$

Note that the chiral symmetry conditions, namely Eq. (14) have been used to derive these equations. We suppose that $\delta\mathbf{x}_0 = 0$, so that there is no explicit symmetry breaking. Moreover, the dynamics in the racemic subspace is assumed to have a steady state. We use the notations $\mathbf{J} = \mathbf{J}_\mathrm{DD} - \mathbf{J}_\mathrm{DL}$ and $\mathbf{M} \equiv \mathbf{J} - \mathbf{I}/\tau$. The $N_\mathrm{C} \times N_\mathrm{C}$ $\mathbf{M}$ matrix controls the linear stability of the

racemic steady state. It is asymptotically stable if all the eigenvalues $\{\lambda_i\}$ of that matrix have a negative real part: $\Re\,\lambda_i < 0$ for all $i = 1, 2, ..., N_{\rm C}$. The racemic mixture is unstable if at least one of its eigenvalues has a positive real part.

Generalized Frank's model For numerical investigations, we consider the model (3)-(4)-(5) with $\mathrm{E}_i = \mathrm{D}_i$ and $\bar{\mathrm{E}}_i = \mathrm{L}_i$ for all $i = 1, 2, \ldots, N_{\rm C}$ in the irreversible regime with $N_{\rm A} = \tilde{N}_{\tilde{\rm A}} = 1$ and $N \equiv N_{\rm C} \gg 1$. Moreover, we suppose that the initial concentrations $D_{m0} = L_{m0} = \tilde{A}_0 = 0$ for all species m. Since $D_{m0} = L_{m0}$, there is no explicit chiral symmetry breaking caused by non-racemic inflow from the environment. In this regime, the reaction network reads

$$\mathrm{A} + \mathrm{D}_i \to \mathrm{D}_j + \mathrm{D}_k\,, \tag{21}$$

$$\mathrm{A} + \mathrm{L}_i \to \mathrm{L}_j + \mathrm{L}_k\,, \tag{22}$$

$$\mathrm{D}_i + \mathrm{L}_j \to 2\,\tilde{\mathrm{A}}\,, \tag{23}$$

where $i, j, k = 1, 2, \ldots, N_{\rm C}$.

Now, the net reaction rates are given by

$$w_{ijk}^{(\mathrm{D})} = k_{+ijk}\,A\,D_i \qquad \text{with} \qquad j \le k\,, \tag{24}$$

$$w_{ijk}^{(\mathrm{L})} = k_{+ijk}\,A\,L_i \qquad \text{with} \qquad j \le k\,, \tag{25}$$

$$\tilde{w}_{ij} = \tilde{k}_{-ij}\,D_i\,L_j\,, \tag{26}$$

where $\tilde{k}_{-ij} = \tilde{k}_{-ji}$ because of the mirror symmetry. Rate constants are randomly distributed according to a log-normal distribution as explained in SI Appendix, section S4.

The kinetic equations have thus the following form,

$$\dot{A} = -\sum_{\substack{ijk\\ j\le k}} w_{ijk}^{(\mathrm{D})} - \sum_{\substack{ijk\\ j\le k}} w_{ijk}^{(\mathrm{L})} + \frac{1}{\tau}(A_0 - A)\,, \tag{27}$$

$$\dot{D}_m = \sum_{\substack{ijk\\ j\le k}} \nu_{m,ijk}\,w_{ijk}^{(\mathrm{D})} - \sum_i \tilde{w}_{mi} - \frac{1}{\tau}D_m\,, \tag{28}$$

$$\dot{L}_m = \sum_{\substack{ijk\\ j\le k}} \nu_{m,ijk}\,w_{ijk}^{(\mathrm{L})} - \sum_i \tilde{w}_{im} - \frac{1}{\tau}L_m\,, \tag{29}$$

$$\dot{\tilde{A}} = 2\sum_{ij} \tilde{w}_{ij} - \frac{1}{\tau}\tilde{A}\,, \tag{30}$$

with $\nu_{m,ijk} \equiv -\delta_{mi} + \delta_{mj} + \delta_{mk}$.

Numerical simulations of the reaction network The above equations for the fully irreversible model, have been simulated using a Runge-Kutta algorithm of second order. The numerical integration of the kinetic equations has been performed by setting $\tau = 1$, meaning that we take τ as time unit. At the initial time, we assume that there is a very small imbalance between the two enantiomers of given species, characterized by the small parameter ϵ, which is homogeneous among all the species.

The integration of the ordinary differential equations allows us to determine the threshold of instability, as well as every asymptotically stable solution−in particular, the racemic solution with $D_i = L_i$ below the threshold of instability. This threshold can be determined by increasing the control parameter A_0 until the solution of the equations is no longer racemic, giving the critical value of the threshold $A_{0\rm c}$ for the transition-breaking chiral symmetry inside the system.

With this dynamics, we observe that the system never converges towards a non-trivial racemic state where the concentrations of the two enantiomers of a given species would be non-zero and equal to each other. Thus, starting with a state with a small enantiomeric excess, we either reach the trivial racemic state or a homochiral state. For this reason, Fig. 4 has been made by studying the stability of the trivial racemic state using many random realizations of the rate constants k_{+ijk} following a log-normal distribution, which is more efficient numerically than a time integration of the equations of motion.

Further materials

In SI Appendix, section S2, we study the properties of the Jacobian matrix, and deduce from them in section S3 a general instability criterion of the racemic state based on random matrix theory. In section S4, we present the reversible generalized Frank model and we analyze its properties for uniform rate constants. In section S5, we derive the instability criterion for the generalized Frank model with random rate constants. In section S6, we study two diffusively coupled compartments.

Acknowledgments

The authors acknowledge fruitful discussions with A. Blokhuis and A. Duprat. L. Leibler and Y. Geerts are thanked for their helpful comments. PG acknowledges the financial support of the Université Libre de Bruxelles (ULB) and the Fonds de la Recherche Scientifique-FNRS under the grant PDR T.0094.16 for the project "SYMSTATPHYS". DL acknowledges support from Agence Nationale de la Recherche (ANR-10-IDEX-0001-02, IRIS OCAV) and (ANR-11-LABX-0038, ANR-10-IDEX- 0001-02).

Bibliography

[1] D. G. Blackmond, "The origin of biological homochirality," *CSH Perspect. Biol.*, vol. 2, pp. a002147–a002147, May 2010.

[2] K.-H. Ernst, "Molecular chirality in surface science," *Surf. Sci.*, vol. 613, pp. 1–5, 2013.

[3] C. Meinert, P. de Marcellus, L. Le Sergeant d'Hendecourt, L. Nahon, N. C. Jones, S. V. Hoffmann, J. H. Bredehöft, and U. J. Meierhenrich, "Photochirogenesis: Photochemical models on the absolute asymmetric formation of amino acids in interstellar space," *Physics of Life Reviews*, vol. 8, pp. 307–330, Oct. 2011.

[4] F. C. Frank, "On spontaneous asymmetric synthesis," *Biochim. Biophys. Acta*, vol. 11, pp. 459–463, 1953.

[5] D. K. Kondepudi and G. W. Nelson, "Chiral symmetry breaking in nonequilibrium systems," *Phys. Rev. Lett.*, vol. 50, pp. 1023–1026, 1983.

[6] R. Plasson, D. K. Kondepudi, H. Bersini, A. Commeyras, and K. Asakura, "Emergence of homochirality in far-from-equilibrium systems: Mechanisms and role in prebiotic chemistry," *Chirality*, vol. 19, pp. 589–600, 2007.

[7] D. K. Kondepudi and K. Asakura, "Chiral autocatalysis, spontaneous symmetry breaking, and stochastic behavior," *Acc. Chem. Res.*, vol. 34, pp. 946–954, 2001.

[8] J. A. D. Wattis and P. V. Coveney, "Symmetry-breaking in chiral polymerisation," *Origins Life Evol. Bios.*, vol. 35, pp. 243–273, 2005.

[9] R. Kafri, O. Markovitch, and D. Lancet, "Spontaneous chiral symmetry breaking in early molecular networks," *Biol. Direct*, vol. 5, p. 38, 2010.

[10] Y. Saito and H. Hyuga, "Homochirality: Symmetry breaking in systems driven far from equilibrium," *Rev. Mod. Phys.*, vol. 85, pp. 603–621, 2013.

[11] M. Stich, J. M. Ribó, D. G. Blackmond, and D. Hochberg, "Necessary conditions for the emergence of homochirality *via* autocatalytic self-replication," *J. Chem. Phys.*, vol. 145, p. 074111, 2016.

[12] D. Hochberg, R. D. Bourdon Garcia, J. A. Agreda Bastidas, and J. M. Ribó, "Stoichiometric network analysis of spontaneous mirror symmetry breaking in chemical reactions," *Phys. Chem. Chem. Phys.*, vol. 19, p. 17618, 2017.

[13] F. Jafarpour, T. Biancalani, and N. Goldenfeld, "Noise-induced symmetry breaking far from equilibrium and the emergence of biological homochirality," *Phys. Rev. E*, vol. 95, p. 032407, Mar. 2017.

[14] F. Jafarpour, T. Biancalani, and N. Goldenfeld, "Noise-induced mechanism for biological homochirality of early life self-replicators," *Phys. Rev. Lett.*, vol. 115, p. 158101, 2015.

[15] R. Plasson, H. Bersini, and A. Commeyras, "Recycling Frank: Spontaneous emergence of homochirality in noncatalytic systems," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 16733–16738, 2004.

[16] K. Soai, T. Shibata, H. Morioka, and K. Choji, "Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule," *Nature*, vol. 378, pp. 767–768, Dec. 1995.

[17] S. Fujita, "Alkanes as stereoisomers. enumeration by the combination of two dichotomies for three-dimensional trees," *MATCH Commun. Math. Comput. Chem.*, vol. 57, p. 299, 2007.

[18] S. Fujita, "Numbers of monosubstituted alkanes as stereoisomers," *J. Comput. Chem. Jpn.*, vol. 6, pp. 59–72, 2007.

[19] T. Fink and J.-L. Reymond, "Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F," *J. Chem. Inf. Model*, vol. 47, pp. 342–353, 2007.

[20] "PubChem, https://pubchem.ncbi.nlm.nih.gov/ (National center for Biotechnology Information, Bethesda MD)," 2019.

[21] S. Allesina and S. Tang, "Stability criteria for complex ecosystems," *Nature*, vol. 483, pp. 205–208, Mar. 2012.

[22] R. May, "Will a large complex system be stable?," *Nature*, vol. 238, p. 413, 1972.

[23] M. R. Gardner and W. R. Ashby, "Connectance of large dynamic (cybernetic) systems: Critical values for stability," *Nature*, vol. 228, pp. 784–784, Nov. 1970.

[24] J. Ginibre, "Statistical ensembles of complex, quaternion, and real matrices," *J. Math. Phys.*, vol. 6, pp. 440–449, Mar. 1965.

[25] V. L. Girko, "Circular law," *Theory Probab. Appl.*, vol. 29, pp. 694–706, Jan. 1985.

[26] D. Davidi, L. M. Longo, J. Jabłońska, R. Milo, and D. S. Tawfik, "A bird's-eye view of enzyme evolution: Chemical, physicochemical, and physiological considerations," *Chem. Rev.*, vol. 118, pp. 8786–8797, Sept. 2018.

[27] B. A. McGuire, P. B. Carroll, R. A. Loomis, I. A. Finneran, P. R. Jewell, A. J. Remijan, and G. A. Blake, "Discovery of the interstellar chiral molecule propylene oxide (CH_3CHCH_2O)," *Science*, vol. 32, no. 6292, pp. 1449–1452, 2016.

[28] E. Smith and H. J. Morowitz, *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere*. New York, NY: Cambridge University Press, Mar. 2016.

[29] H. J. Morowitz, J. D. Kostelnik, J. Yang, and G. D. Cody, "The origin of intermediary metabolism," *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 7704–7708, July 2000.

[30] A. Blokhuis, D. Lacoste, and P. Nghe, "Universal motifs and the diversity of autocatalytic systems," *Proc. Natl. Acad. Sci. USA*, vol. 117, no. 41, pp. 25230–25236, 2020.

Supplementary Materials

S1 The multiplication of chiral molecules with their number of atoms

The number of species that are chiral are observed to increase faster with their number of atoms, than the number of achiral species, in spite of the fact that diatomic and triatomic molecules are achiral and only tetratomic molecules can be chiral (in their ground electronic state). In particular, studies of alkane and monosubstituted alkane stereoisomers [18] show that the numbers of chiral C_k and achiral A_k species are growing exponentially with their number k of carbon atoms according to

$$C_k \sim \Lambda^k \qquad \text{and} \qquad A_k \sim \Lambda^{k/2} \qquad \text{with} \qquad \Lambda \simeq 3.287112 \tag{S1}$$

for $k \gg 1$. As a consequence, chiral molecules become overwhelmingly dominant for a large enough number of atoms.

Now, the question is to determine when the crossover occurs between a world of small molecules dominated by achiral species and a world dominated by chiral molecules, which is likely to become homochiral by spontaneous symmetry breaking induced in the non-equilibrium reaction network.

In order to answer this question, we have investigated the fractions of achiral and chiral species as functions of the number of atoms in each molecule. Such fractions can be defined by counting the pairs of enantiomers either once or twice. The numbers of achiral and chiral species with n heavy atoms being respectively denoted A_n and C_n, on the one hand, the fractions defined by counting once the pairs of enantiomers are given by

$$f_n^{(\mathrm{A})} \equiv \frac{A_n}{A_n + C_n} \qquad \text{and} \qquad f_n^{(\mathrm{C})} \equiv \frac{C_n}{A_n + C_n}, \tag{S2}$$

such that $f_n^{(\mathrm{A})} + f_n^{(\mathrm{C})} = 1$, in which case, the crossover happens for n_1 atoms in the molecule such that

$$A_{n_1} = C_{n_1} \, . \tag{S3}$$

On the other hand, the fractions defined by counting twice the pairs of enantiomers are given by

$$\tilde{f}_n^{(\mathrm{A})} \equiv \frac{A_n}{A_n + 2C_n} \qquad \text{and} \qquad \tilde{f}_n^{(\mathrm{C})} \equiv \frac{2C_n}{A_n + 2C_n}, \tag{S4}$$

such that $\tilde{f}_n^{(\mathrm{A})} + \tilde{f}_n^{(\mathrm{C})} = 1$ in which other case, the crossover happens for n_2 atoms in the molecule such that

$$A_{n_2} = 2C_{n_2} \, . \tag{S5}$$

Since C_n becomes larger than A_n as n increases, we should expect that $n_2 < n_1$.

S1.1 Monosubstituted alkane stereoisomers

Achiral and chiral monosubstituted alkanes $\mathrm{C}_k\mathrm{H}_{2k+1}\mathrm{X}$ have been enumerated in Ref. [18]. In particular, Table 1 of Ref. [18] gives the numbers of achiral A_k and chiral C_k monosubstituted alkanes as stereoisomers versus the number k of carbon atoms they contain and this up to $k = 100$. Using these data, the fractions of achiral and chiral stereoisomers have been obtained and they are shown in Fig. S1 by counting once or twice the enantiomeric pairs.

We observe that the crossover occurs at $k_1 \simeq 5.7$ or $k_2 \simeq 4.7$, depending on whether the pairs of enantiomers are counted once or twice. As expected, we have that $k_2 < k_1$. Now, the result is that the crossover happens for a relatively small number of carbon atoms. Here, the carbon atoms and the substituted atom X determine the geometry of the molecule. Here, the temperature is supposed to be high enough such that hydrogen atoms rotate and vibrate fast enough that the chirality is determined by the skeleton of the carbon and X atoms.

S1.2 Alkane stereoisomers

The stereoisomers of alkanes C_kH_{2k+2} have also been studied and the enumeration of achiral and chiral alkanes is given in Ref. [17]. Table 3 of Ref. [17] gives the numbers of achiral A_k and chiral C_k alkanes as stereoisomers versus the number k of carbon atoms they contain and this up to $k = 100$. The fractions of achiral and chiral stereosiomers obtained with these data are plotted in Fig. S2.

Here, we see that the crossover occurs at $k_1 \simeq 9.5$ if the pairs of enantiomers are counted once and at $k_2 \simeq 8.4$ if they are counted twice. Again, the crossover happens for a relatively small number of carbon atoms (which are the atoms determining the molecular geometry). The crossover happens for somewhat larger molecules because alkanes have molecular structures that are more symmetric than in the presence of one substitution, thus delaying the crossover as the number of determining atoms increases.

S1.3 Chemical Universe

In Ref. [19], all the possible molecules up to 11 atoms of C, N, O, and F were generated by considering simple valency, chemical stability, and synthetic feasibility rules, and they were collected in a database containing 26.4 million molecules and 110.9 million stereoisomers. Fig. 5 of Ref. [19] shows the fractions of achiral and chiral molecules in the database as a function of their size characterized by the number of heavy atoms. Here, the crossover happens for $n_1 \simeq 8.5 \pm 0.1$. This virtual exploration of the chemical universe clearly demonstrates the prevalence of chirality for large enough molecules.

S1.4 Analysis of the PubChem database

S1.4.1 Raw data

The raw database of PubChem contains 139 millions of species. In the following, we will restrict our analysis to species which contain less than 20 heavy atoms (*i.e.*, atoms heavier than hydrogen). There are two reasons for this choice, on one hand the statistics becomes more limited for molecules much longer than 20, and on the other hand, there is a discontinuity in the number of achiral and chiral molecules in the PubChem database as shown in Fig. S3a. We have contacted the curators of the database, but there is no information currently available about the origin of this discontinuity. In any case, we should avoid this problem by staying below 20.

From the complete dataset downloaded from PubChem, $91,606,016$ molecules were analyzed after rejection of compounds with isotopic elements, multiple components or incomplete data on bond structure. From the $33,563,343$ molecules with less than 21 heavy atoms, $18,705,878$ molecules (55.7 %) are chiral, and $1,376,672$ chiral molecules have no stereocenters (7.4 % of the chiral molecules with less than 21 heavy atoms) thus for these molecules their chirality depends only on their non planar geometry.

As shown in Fig. S3b, the analysis of the $33,563,343$ molecules with less than 21 heavy atoms of the database in terms of their fraction of chiral and non-chiral species shows a crossover around $n_{\rm raw} \simeq 9.4$. A crossover in this region is coherent with an increase in the number of stereoisomers per molecule for molecules of this length.

S1.4.2 Methods of generation of stereoisomers and enantiomers

Chiral species in the PubChem database were detected using the chiral flag present in the list of SDF files which contains information about the structure of the molecules in the database.

For the generation of enantiomers, a list of non-canonical SMILES formulas was built, which contains information about defined stereocenters. For each chiral molecule with an available SMILES formula (*e.g.*, C[**C@@**](CC1CC=C(C(=C1)O)O)(C(=O)O)N), one generates a mirror image of it (*i.e.*, C[**C@**](CC1CC=C(C(=C1)O)O)(C(=O)O)N for the latter example), and then one searches it in the list. If it is not found, then it is added to the list in the expanded database. However, for chiral centers that do not explicitly appears in the SMILES formula, one cannot find them and reverse them with a simple method and this creates an uncertainty in the final number of chiral molecules due to missed generated enantiomers. Thus a fraction of enantiomers cannot be generated due to incomplete data in the PubChem database.

For the generation of stereoisomers, one looks whether the database contains the theoretical maximum number of stereoisomers for a given species, which can be evaluated from the number of defined stereocenters. If all the stereoisomers are not present, which is frequent for molecules containing several stereocenters, the database is expanded so that each chiral species has the maximum possible number of stereoisomers (*i.e.*, 2^{n^*} with n^* the number of stereocenters in the molecule). However, this procedure does not count properly the meso forms, which should be labeled as achiral although they contain stereocenters due to an internal symmetry.

Now, after generating enantiomers as explained previously, one obtains Fig. S4.

Then, stereoisomers were generated using the procedure described previously, giving Fig. 1a of the main article where 47,452,700 steoisomers have been added to complete the PubChem dataset. In this case, the intersection occurs at $n_2 \simeq 6.4$ for if both enantiomers are considered and $n_1 \simeq 8.1$ if only one enantiomer is considered.

S1.4.3 Error bars

In this subsection, we explain how error bars were obtained in the graph of achiral and chiral fractions. For molecules with i atoms, the number of chiral molecules is denoted $\mathcal{N}_i$ and the one of achiral molecules $\mathcal{M}_i$. These numbers are taken as independent Poisson distributions with a parameter given by their mean number, *i.e.*, by $\mathcal{N}_i$ and $\mathcal{M}_i$ themselves. The fraction of chiral molecules is given by

$$x_i = \frac{\mathcal{N}_i}{\mathcal{M}_i + \mathcal{N}_i} \,. \tag{S6}$$

The error on x_i thus reads

$$\sigma_{x_i} = x_i(1-x_i)\left(\frac{1}{\sqrt{\mathcal{M}_i}} + \frac{1}{\sqrt{\mathcal{N}_i}}\right). \tag{S7}$$

The model does not capture systematic errors, but only the statistical errors in the counting of $\mathcal{N}_i$ or $\mathcal{M}_i$.

S2 Separating the mean from fluctuations in the Jacobian matrix

We now study the properties of the $N \times N$ Jacobian matrix $\mathbf{J}$ with $N = N_{\rm C}$ introduced in the Materials and Methods of the main text. For large complex networks, this matrix may be supposed to be random because of fluctuations in the values of its elements for the different reactions and species. Let us separate the mean of the elements of $\mathbf{J}$ from their fluctuations in the following way:

$$\mathbf{J} = \mu\,\mathbf{1} + \sigma\,\mathbf{G} \tag{S8}$$

where $\mathbf{1}$ is the matrix full of elements equal to 1 and $\mathbf{G}$ has elements distributed according to independent Gaussian distributions of zero mean and unit variance:

$$\langle G_{ab}\rangle = 0\,, \qquad \langle G_{ab}G_{cd}\rangle = \delta_{ac}\delta_{bd}\,. \tag{S9}$$

Accordingly, the mean value of Eq. (S8) gives

$$\langle \mathbf{J}\rangle = \mu\,\mathbf{1}\,, \tag{S10}$$

allowing us to determine the parameter μ as the mean value of the elements of the matrix $\mathbf{J}$:

$$\mu = \frac{1}{N^2}\sum_{a,b=1}^{N} J_{ab}\,. \tag{S11}$$

Moreover, the parameter σ can be evaluated by the root mean square of the matrix elements,

$$\sigma = \frac{1}{N}\left[\sum_{a,b=1}^{N}(J_{ab}-\mu)^2\right]^{1/2}, \tag{S12}$$

as a consequence of Eq. (S9).

The matrix full of ones has the eigenvalues $\{N, 0, 0, ..., 0\}$ and it can be diagonalized with an orthogonal transformation $\mathbf{O}$ composed of the eigenvectors $\mathbf{v} = \{\cos[2\pi k(m-1)/N]\}_{k=1}^{N}$ with $m = 1, 2, ..., N$ (after their normalization). The eigenvector corresponding to the eigenvalue equal to N is thus given by $m = 1$ in the expression of the latter eigenvectors. Accordingly,

$$\begin{aligned} \mathbf{J}' = \mathbf{O}^{\mathrm{T}} \cdot \mathbf{J} \cdot \mathbf{O} &= \mu\, \mathbf{O}^{\mathrm{T}} \cdot \mathbf{1} \cdot \mathbf{O} + \sigma\, \mathbf{O}^{\mathrm{T}} \cdot \mathbf{G} \cdot \mathbf{O} \\ &= \mu \begin{pmatrix} N & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} + \sigma\, \mathbf{G}' \end{aligned} \tag{S13}$$

with the matrix

$$\mathbf{G}' = \mathbf{O}^{\mathrm{T}} \cdot \mathbf{G} \cdot \mathbf{O}\,. \tag{S14}$$

This latter is again a random matrix with elements

$$G'_{ij} = \sum_{a,b} (\mathbf{O}^{\mathrm{T}})_{ia} G_{ab} (\mathbf{O})_{bj} \tag{S15}$$

distributed according to independent Gaussian distribution of zero mean and unit variance. Indeed, we have that

$$\langle G'_{ij} \rangle = \sum_{a,b} (\mathbf{O}^{\mathrm{T}})_{ia} \langle G_{ab} \rangle (\mathbf{O})_{bj} = 0 \tag{S16}$$

and

$$\begin{aligned} \langle G'_{ij} G'_{kl} \rangle &= \sum_{a,b,c,d} (\mathbf{O}^{\mathrm{T}})_{ia} (\mathbf{O})_{bj} (\mathbf{O}^{\mathrm{T}})_{kc} (\mathbf{O})_{dl} \langle G_{ab} G_{cd} \rangle \\ &= \sum_{a} (\mathbf{O}^{\mathrm{T}})_{ia} (\mathbf{O})_{ak} \sum_{b} (\mathbf{O}^{\mathrm{T}})_{jb} (\mathbf{O})_{bl} = \delta_{ik}\, \delta_{jl}\,. \end{aligned} \tag{S17}$$

If $\mu \neq 0$ and $\sigma \neq 0$, the eigenvalue problem can thus be solved by perturbation theory in the small parameter $\zeta = \sigma/(\mu N)$. Therefore, in the limit $N \to \infty$, the spectrum of the matrix $\mathbf{J}$ is composed of the eigenvalue

$$\lambda^{(1)} = \mu\, N + O(N^0) \tag{S18}$$

and $N - 1$ eigenvalues $\{\lambda^{(m)}\}_{m=2}^{N}$ contained inside the disk of radius $\sigma\sqrt{N}$ in the complex plane $(\Re\,\lambda, \Im\,\lambda)$ with probability one in the limit $N \to \infty$. The probability density of these latter eigenvalues is similar as for the real Ginibre ensemble in the limit $N \to \infty$ [24]. In the general case where the matrix elements have non-Gaussian distributions and/or are statistically correlated, the dominant eigenvalue behaves as described by Eq. (S18), but the $N - 1$ non-dominant eigenvalues may have different kinds of probability distribution.

S3 General instability criterion

Now using the results of section S2 for random matrices with Gaussian independent and identically distributed elements, we conclude that when $\mu N \geq \sigma\sqrt{N}$, which is equivalent to $N \geq (\sigma/\mu)^2$, an isolated eigenvalue will be dominant as in mechanism (i). If this condition is satisfied, the instability occurs when $\mu N \geq 1/\tau$, in other words when the number of chiral species $N = N_{\mathrm{C}}$ is such that $N_{\mathrm{C}} \geq \max\{1/(\tau\mu), (\sigma/\mu)^2\}$ with $\mu > 0$.

Instead when $N \leq (\sigma/\mu)^2$, the previous eigenvalue is no longer dominant and the instability can only occur due to eigenvalues that are located at the edge of the Girko circle [25], which is the second mechanism (ii). This edge, which corresponds to eigenvalues with a maximum real part, can be made of either a single real eigenvalue or to a

pair of conjugated complex valued eigenvalues. The instability then occurs when $N \geq 1/(\tau\sigma)^2$, or taken together when $(\sigma/\mu)^2 \geq N \geq 1/(\tau\sigma)^2$.

We note that the first mechanism (i) holds with $\mu > 0$ even for a vanishing root mean square $\sigma = 0$, but the second mechanism (ii) requires that the root mean square σ is not equal to zero.

According to this analysis, a non-equilibrium reaction network with sufficiently many chiral species is likely to undergo spontaneous chiral symmetry breaking. We now discuss a specific implementation for a chemical reaction network.

S4 Generalized Frank's model

The basic idea is that there exist more chiral than achiral species. Frank's model [4] is thus generalized by multiplying the species, especially, the chiral species. Moreover, the reverse reactions are included in order to possibly satisfy microreversibility. The reaction network is given by Eqs. (3)-(4)-(5) of the main text with $\mathrm{E}_i = \mathrm{D}_i$ and $\bar{\mathrm{E}}_i = \mathrm{L}_i$, reading

$$\mathrm{A}_a + \mathrm{D}_i \rightleftharpoons \mathrm{D}_j + \mathrm{D}_k \,, \tag{S19}$$

$$\mathrm{A}_a + \mathrm{L}_i \rightleftharpoons \mathrm{L}_j + \mathrm{L}_k \,, \tag{S20}$$

$$\mathrm{D}_i + \mathrm{L}_j \rightleftharpoons \tilde{\mathrm{A}}_b + \tilde{\mathrm{A}}_c \,, \tag{S21}$$

with $a = 1, 2, \ldots, N_\mathrm{A}$; $b, c = 1, 2, \ldots, \tilde{N}_{\tilde{\mathrm{A}}}$; and $i, j, k = 1, 2, \ldots, N_\mathrm{C}$. The net reaction rates are given by

$$w^{(\mathrm{D})}_{aijk} = k_{+aijk} A_a D_i - k_{-aijk} D_j D_k \qquad \text{with} \qquad j \leq k \,, \tag{S22}$$

$$w^{(\mathrm{L})}_{aijk} = k_{+aijk} A_a L_i - k_{-aijk} L_j L_k \qquad \text{with} \qquad j \leq k \,, \tag{S23}$$

$$\tilde{w}_{bcij} = \tilde{k}_{-bcij} D_i L_j - \tilde{k}_{+bcij} \tilde{A}_b \tilde{A}_c \qquad \text{with} \qquad b \leq c \,, \tag{S24}$$

where the positive sign in the subscripts of the rate constants refers to the direction of chirality generation. We note that $\tilde{k}_{\pm bcij} = \tilde{k}_{\pm bcji}$ because of the mirror symmetry (14). The kinetic equations are thus given by

$$\dot{A}_d = -\sum_{\substack{aijk \\ j \leq k}} \delta_{da}\, w^{(\mathrm{D})}_{aijk} - \sum_{\substack{aijk \\ j \leq k}} \delta_{da}\, w^{(\mathrm{L})}_{aijk} + \frac{1}{\tau}(A_{0d} - A_d) \,, \tag{S25}$$

$$\dot{D}_m = \sum_{\substack{aijk \\ j \leq k}} \nu_{m,ijk}\, w^{(\mathrm{D})}_{aijk} - \sum_{\substack{bcij \\ b \leq c}} \delta_{mi}\, \tilde{w}_{bcij} + \frac{1}{\tau}(D_{m0} - D_m) \,, \tag{S26}$$

$$\dot{L}_m = \sum_{\substack{aijk \\ j \leq k}} \nu_{m,ijk}\, w^{(\mathrm{L})}_{aijk} - \sum_{\substack{bcij \\ b \leq c}} \delta_{mj}\, \tilde{w}_{bcij} + \frac{1}{\tau}(L_{m0} - L_m) \,, \tag{S27}$$

$$\dot{\tilde{A}}_e = \sum_{\substack{bcij \\ b \leq c}} (\delta_{eb} + \delta_{ec})\, \tilde{w}_{bcij} + \frac{1}{\tau}(\tilde{A}_{0e} - \tilde{A}_e) \,, \tag{S28}$$

where

$$\nu_{m,ijk} \equiv -\delta_{mi} + \delta_{mj} + \delta_{mk} \,. \tag{S29}$$

The rate constants can be taken according to log-normal distributions [26]. If the rate constants are distributed around some mean values with relatively small root mean squares, the leading behavior can be determined by replacing the rate constants with their mean value. In this respect, we may assume that all the rate constants are equal,

$$\tilde{k}_{\pm abij} = \tilde{k}_\pm \,, \qquad k_{\pm aijk} = k_\pm \qquad (\forall\, a, b, i, j, k) \,, \tag{S30}$$

and the concentrations of achiral species, D-, and L-enantiomers can also be supposed to be equal

$$A_a = A \,, \qquad D_i = D \,, \qquad L_i = L \,, \qquad \text{and} \qquad \tilde{A}_a = \tilde{A} \,, \qquad (\forall\, a, i) \,. \tag{S31}$$

Thus, the kinetic equations become

$$\dot{A} = -\frac{N_{\rm C}}{N_{\rm A}}\left[(K_+ A\, D - K_- D^2) + (K_+ A\, L - K_- L^2)\right] + \frac{1}{\tau}(A_0 - A)\,, \tag{S32}$$

$$\dot{D} = (K_+ A\, D - K_- D^2) - (\tilde{K}_- DL - \tilde{K}_+ \tilde{A}^2) + \frac{1}{\tau}(D_0 - D)\,, \tag{S33}$$

$$\dot{L} = (K_+ A\, L - K_- L^2) - (\tilde{K}_- DL - \tilde{K}_+ \tilde{A}^2) + \frac{1}{\tau}(L_0 - L)\,, \tag{S34}$$

$$\dot{\tilde{A}} = 2\,\frac{N_{\rm C}}{\tilde{N}_{\tilde{\rm A}}}\,(\tilde{K}_- DL - \tilde{K}_+ \tilde{A}^2) + \frac{1}{\tau}(\tilde{A}_0 - \tilde{A})\,, \tag{S35}$$

with the effective rate constants:

$$K_\pm \equiv \frac{N_{\rm A} N_{\rm C}(N_{\rm C}+1)}{2}\,k_\pm \qquad \text{and} \qquad \tilde{K}_\pm \equiv \frac{\tilde{N}_{\tilde{\rm A}}(\tilde{N}_{\tilde{\rm A}}+1)N_{\rm C}}{2}\,\tilde{k}_\pm\,. \tag{S36}$$

As a consequence of these kinetic equations, we have that

$$\begin{aligned} N_{\rm A}\,\dot{A} + N_{\rm C}\,\dot{D} + N_{\rm C}\,\dot{L} + \tilde{N}_{\tilde{\rm A}}\dot{\tilde{A}} = \frac{1}{\tau}\big[&(N_{\rm A}\,A_0 \;+\; N_{\rm C}\,D_0 \;+\; N_{\rm C}\,L_0 \;+\; \tilde{N}_{\tilde{\rm A}}\,\tilde{A}_0) \\ &-(N_{\rm A}\,A \;+\; N_{\rm C}\,D \;+\; N_{\rm C}\,L \;+\; \tilde{N}_{\tilde{\rm A}}\,\tilde{A})\big], \end{aligned} \tag{S37}$$

implying

$$\lim_{t\to\infty}(N_{\rm A}\,A + N_{\rm C}\,D + N_{\rm C}\,L + \tilde{N}_{\tilde{\rm A}}\,\tilde{A}) = N_{\rm A}\,A_0 + N_{\rm C}\,D_0 + N_{\rm C}\,L_0 + \tilde{N}_{\tilde{\rm A}}\,\tilde{A}_0\,. \tag{S38}$$

The model is compatible with the existence of equilibrium. Indeed, the detailed balance conditions give the following Guldberg-Waage equilibrium relations,

$$\frac{D_{\rm eq}}{A_{\rm eq}} = \frac{L_{\rm eq}}{A_{\rm eq}} = \frac{K_+}{K_-} \qquad \text{and} \qquad \frac{\tilde{A}_{\rm eq}}{A_{\rm eq}} = \frac{K_+}{K_-}\sqrt{\frac{\tilde{K}_-}{\tilde{K}_+}}\,. \tag{S39}$$

Therefore, equilibrium exists for any positive value of the rate constants and two rate constants can independently take arbitrarily small values.

In the fully irreversible regime with $K_- = 0$ and $\tilde{K}_+ = 0$, we further suppose that the system is only supplied with the achiral species of high free energy: $D_0 = L_0 = \tilde{A}_0 = 0$. Since Eq. (S38) holds for long enough time, we have that

$$A = A_0 - \frac{N_{\rm C}}{N_{\rm A}}\,(D + L) - \frac{\tilde{N}_{\tilde{\rm A}}}{N_{\rm A}}\,\tilde{A}\,. \tag{S40}$$

Therefore, the kinetic equations reduce to the three following equations:

$$\dot{D} = K_+ D\left[A_0 - \frac{N_{\rm C}}{N_{\rm A}}\,(D + L) - \frac{\tilde{N}_{\tilde{\rm A}}}{N_{\rm A}}\,\tilde{A}\right] - \tilde{K}_- DL - \frac{1}{\tau}\,D\,, \tag{S41}$$

$$\dot{L} = K_+ L\left[A_0 - \frac{N_{\rm C}}{N_{\rm A}}\,(D + L) - \frac{\tilde{N}_{\tilde{\rm A}}}{N_{\rm A}}\,\tilde{A}\right] - \tilde{K}_- DL - \frac{1}{\tau}\,L\,, \tag{S42}$$

$$\dot{\tilde{A}} = 2\,\frac{N_{\rm C}}{\tilde{N}_{\tilde{\rm A}}}\,\tilde{K}_- DL - \frac{1}{\tau}\,\tilde{A}\,. \tag{S43}$$

Setting

$$\Delta \equiv K_+ A_0 - \frac{1}{\tau} = \frac{N_{\rm C}(N_{\rm C}+1)}{2}\,k_+\,N_{\rm A}\,A_0 - \frac{1}{\tau}\,, \tag{S44}$$

the steady states and their eigenvalues $\{\xi_1, \xi_2, \xi_3\}$ of linear stability are here given by

$$D = L = \tilde{A} = 0 : \qquad \xi_1 = \xi_2 = \Delta\,,\ \xi_3 = -1/\tau\,, \tag{S45}$$

$$D = 0\,,\ L = \frac{N_{\rm A}\,\Delta}{N_{\rm C}\,K_+}\,,\ \tilde{A} = 0 : \qquad \xi_1 = -\frac{N_{\rm A}\,\tilde{K}_-\,\Delta}{N_{\rm C}\,K_+}\,,\ \xi_2 = -\Delta\,,$$
$$\xi_3 = -1/\tau\,, \tag{S46}$$

$$D = \frac{N_{\rm A}\,\Delta}{N_{\rm C}\,K_+}\,,\ L = 0\,,\ \tilde{A} = 0 : \qquad \xi_1 = -\Delta\,,\ \xi_2 = -\frac{N_{\rm A}\,\tilde{K}_-\,\Delta}{N_{\rm C}\,K_+}\,,$$
$$\xi_3 = -1/\tau\,, \tag{S47}$$

$$D = L = \frac{N_{\rm A}\,\Delta}{2N_{\rm C}\,K_+ + N_{\rm A}\,\tilde{K}_-} + O(\Delta^2)\,,\ \tilde{A} = O(\Delta^2) : \qquad \xi_1 = \frac{N_{\rm A}\,\tilde{K}_-\,\Delta}{2N_{\rm C}\,K_+ + N_{\rm A}\,\tilde{K}_-} + O(\Delta^2)\,,$$
$$\xi_2 = -\Delta + O(\Delta^2)\,,$$
$$\xi_3 = -1/\tau + O(\Delta^2)\,. \tag{S48}$$

In these expressions, the terms $O(\Delta^2)$ are negligible if

$$\frac{N_{\rm A}\,N_{\rm C}\,K_+\,\tilde{K}_-\,\tau\,\Delta}{(2N_{\rm C}\,K_+ + N_{\rm A}\,\tilde{K}_-)^2} \ll 1\,. \tag{S49}$$

The behavior is determined by the parameter (S44). Since concentrations are always non-negative, the only steady state that exists if $\Delta < 0$ is the trivial racemic state (S45), which is an attractor because $\xi_1, \xi_2, \xi_3 < 0$ in this case. If $\Delta > 0$, three new steady states emerge, which are the L-homochiral attractor (S46), the D-homochiral attractor (S47), and the non-trivial racemic state (S48). This latter is unstable since $\xi_1 > 0$ for this new steady state. The threshold of instability towards homochirality is thus found at $\Delta = 0$. Therefore, spontaneous chiral symmetry breaking happens if the following criterion is satisfied,

$$k_+\,\tau\,N_{\rm A}\,A_0 > \frac{2}{N_{\rm C}(N_{\rm C}+1)}\,, \tag{S50}$$

which is Eq. (6) of the main text in the case $\langle k_+ \rangle = k_+$ where all the rate constants are equal.

Now, if the rate constants were not all equal as in Eq. (S30), but if they were statistically distributed, the analysis carried out here above would provide the mean behavior of the system. However, the statistical distribution of the rate constants would introduce further effects that should also be analyzed. In particular, for every steady state, the matrix of linear stability could be decomposed in a similar way as in Eq. (S8) into a mean value that would be given by Eq. (S11) and fluctuations of root mean square (S12). The leading eigenvalue of this random matrix could thus be evaluated as in section S2, giving an estimation comparable to the eigenvalues ξ_i obtained here above and this for every stable or unstable steady state.

Similar results hold for the other models with either $\mathrm{E}_m = \mathrm{D}_m$ and $\bar{\mathrm{E}}_m = \mathrm{L}_m$ or $\mathrm{E}_m = \mathrm{L}_m$ and $\bar{\mathrm{E}}_m = \mathrm{D}_m$ for each enantiomeric pairs $m = 1, 2, \ldots, N_{\rm C}$.

S5 Instability criterion of the trivial racemic state for the generalized Frank model

Spontaneous chiral symmetry breaking can be investigated by considering the linear stability analysis of any racemic solution with the stationary concentrations $\mathbf{D} = \mathbf{L}$ where $\mathbf{D} = \{D_i\}_{i=1}^N$ and $\mathbf{L} = \{L_i\}_{i=1}^N$. For this purpose, we introduce the variables

$$\delta\mathbf{X} \equiv \frac{1}{2}\,(\delta\mathbf{L} - \delta\mathbf{D})\,, \tag{S51}$$

characterizing infinitesimal deviations with respect to the racemic subspace. These deviations are ruled by the following set of linear equations:

$$\frac{d}{dt}\delta\mathbf{X} = \mathbf{M}\cdot\delta\mathbf{X}\,, \tag{S52}$$

with the matrix

$$\mathsf{M} = \mathsf{J} - \frac{1}{\tau}\,\mathsf{I} = \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} - \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{L}}\,, \tag{S53}$$

since the fundamental chiral symmetry of the kinetic equations implies that

$$\frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} = \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} \qquad \text{and} \qquad \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{L}} = \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{D}}\,. \tag{S54}$$

For the irreversible model we consider, we have the matrix elements

$$\begin{aligned} M_{mn} &= J_{mn} - \frac{1}{\tau}\,\delta_{mn} = \frac{\partial \dot{D}_m}{\partial D_n} - \frac{\partial \dot{D}_m}{\partial L_n} = \sum_{\substack{i \\ m \le i}} k_{+nmi}\,A + \sum_{\substack{i \\ i \le m}} k_{+nim}\,A \\ &+ \tilde{k}_{-mn}\,D_m - \delta_{mn}\left(\sum_{\substack{ij \\ i \le j}} k_{+nij}\,A + \sum_i \tilde{k}_{-mi}\,L_i + \frac{1}{\tau}\right), \end{aligned} \tag{S55}$$

where A and $D_i = L_i$ are the concentrations of the stationary racemic solution. Since the rate constants are supposed to be statistically distributed, this is also the case for the stationary concentrations $\{D_i\}_{i=1}^N$ and $\{L_i\}_{i=1}^N$ and thus for the matrix elements J_{mn}. The statistical distribution of the matrix elements J_{mn} depends on the reaction network and may be complicated, but they could be decomposed as explained in section S2 into a mean value given by Eq. (S11) and fluctuations of root mean square (S12).

At the trivial racemic fixed point such that $D_i = L_i = \tilde{A} = 0$ for all species i and $A = A_0$, the elements (S55) of the Jacobian matrix associated with the evolution of the enantiomeric excess are evaluated by

$$M_{mn} = J_{mn} - \frac{1}{\tau}\,\delta_{mn} = A_0\left(\sum_{\substack{i \\ m \le i}} k_{+nmi} + \sum_{\substack{i \\ i \le m}} k_{+nim}\right) - \delta_{mn}\left(A_0 \sum_{\substack{ij \\ i \le j}} k_{+nij} + \frac{1}{\tau}\right). \tag{S56}$$

Thus the matrix M can be decomposed into three matrices Q and R as

$$\mathsf{M} = \mathsf{J} - \frac{1}{\tau}\,\mathsf{I} = A_0\,\mathsf{Q} - A_0\,\mathsf{R} - \mathsf{I}/\tau\,. \tag{S57}$$

where I is the identity, the elements of matrix Q are given by

$$Q_{mn} = \sum_{\substack{i \\ m \le i}} k_{+nmi} + \sum_{\substack{i \\ i \le m}} k_{+nim}\,, \tag{S58}$$

which are sums of $N_\mathrm{C} + 1$ random variables of mean $\langle k_+ \rangle$ and standard deviation σ_{k_+} (the element k_{+nmm} occurs twice, once in each sum). Finally, R is a diagonal matrix of elements

$$R_{nn} = \sum_{\substack{ij \\ i \le j}} k_{+nij}\,. \tag{S59}$$

According to the central limit theorem, in the large N_C limit, the elements of Q are distributed following a Gaussian distribution of mean $\mu_Q = \langle k_+ \rangle (N_\mathrm{C} + 1)$ and standard deviation $\sigma_Q = \sigma_{k_+}\sqrt{N_\mathrm{C} + 1}$. Moreover, the elements of the matrix R are also randomly distributed according to a Gaussian of mean $\mu_R = \langle k_+ \rangle N_\mathrm{C}(N_\mathrm{C} + 1)/2$ and standard deviation $\sigma_R = \sigma_{k_+}\sqrt{N_\mathrm{C}(N_\mathrm{C} + 1)/2}$. Unfortunately, although the spectra of the matrices Q and R are known, it is not possible to deduce immediately from these the spectrum of M, because these matrices are not diagonal in the same base.

One can however still use perturbation theory. For the matrix Q, we use the same decomposition in terms of a full matrix of ones plus a correction G:

$$\mathsf{Q} = \mu_Q\,\mathbf{1} + \sigma_Q\,\mathsf{G}\,, \tag{S60}$$

and we decompose the matrix $\mathbf{R}$ as

$$\mathbf{R} = \mu_R \, \mathbf{I} + \mathbf{H} \,, \tag{S61}$$

where $\mathbf{H}$ is a diagonal matrix with subdominant terms as compared to μ_R (this follows from the law of large numbers). In the end, this means we can decompose $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{P} + A_0 \, \sigma_Q \, \mathbf{G} - A_0 \, \mathbf{H} \,. \tag{S62}$$

where $\mathbf{P} = A_0 \mu_Q \mathbf{1} - \left(A_0 \mu_R + \tau^{-1} \right) \mathbf{I}$. The largest eigenvalue of $\mathbf{P}$ is $A_0(N_\mathrm{C}\mu_Q - \mu_R) - \tau^{-1} = A_0 \langle k_+ \rangle N_\mathrm{C}(N_\mathrm{C}+1)/2 - \tau^{-1}$, and this eigenvalue can be shown to be dominant using the same perturbation calculation as done before.

For the system to be unstable, this dominant eigenvalue must be positive. Thus, the threshold above which the system is unstable is

$$\langle k_+ \rangle \, \tau \, A_0 > \frac{2}{N_\mathrm{C}(N_\mathrm{C}+1)} \,, \tag{S63}$$

which is Eq. (S50) with the number of achiral species equal to $N_\mathrm{A} = 1$. Moreover, Eq. (S50) is recovered when all the rate constants are equal. We note that a deviation from the prediction (S63) is observed when σ_{k_+} becomes large compare to $\langle k_+ \rangle$, as depicted in Fig. S5.

One observes that the other eigenvalues of $\mathbf{M}$ do not stay within a Girko circle as shown in Fig. S6. There is no contradiction since the random matrix $\mathbf{M}$ does not have the same statistics for its diagonal and off-diagonal elements, therefore the assumptions of the Girko theorem do not hold anymore [25].

We note that the permutation $\mathrm{D}_m \leftrightarrow \mathrm{L}_m$ for some enantiomeric pair m implies that the corresponding enantiomeric excess changes sign, $\delta X_m \to -\delta X_m$. However, the eigenvalues of the matrix $\mathbf{M}$ remain unchanged under such transformations. This can be shown by the following calculation. Let us denote the eigenvector $\mathbf{u}$ (with eigenvalue p) of the original matrix $\mathbf{M}$, and the new eigenvector $\mathbf{u'}$ (with eigenvalue p') of the transformed matrix $\mathbf{M'}$, obtained after such a permutation, so that

$$\sum_j \mathsf{M}_{ij} \, u_j = p \, u_i \qquad \text{and} \qquad \sum_j \mathsf{M}'_{ij} \, u'_j = p' u'_i \,, \tag{S64}$$

Now, the permutation of the enantiomers means that $\delta X'_i = (-1)^{s_i} \delta X_i$ with $s_i = 1$ if the i^{th} enantiomers are permuted and $s_i = 0$ otherwise. Using Eq. (S52) for the matrices $\mathbf{M}$ and $\mathbf{M'}$, we obtain $M'_{ij} = (-1)^{s_i + s_j} M_{ij}$. It follows from this that the eigenvectors transform as $u'_j = (-1)^{s_j} u_j$ with no change in the eigenvalues $p' = p$. Accordingly, all our results hold for the $2^{N_\mathrm{C}-1}$ models considered.

S6 Two diffusively coupled compartments

We consider here two diffusively coupled compartments containing the same chemical network considered before in a well-mixed situation. In addition, we assume the irreversible regime with $N_\mathrm{A} = \tilde{N}_{\tilde{\mathrm{A}}} = 1$ and $N_\mathrm{C} \gg 1$. The reactions within each compartment read:

$$\mathrm{A}_a + \mathrm{D}_i \to \mathrm{D}_j + \mathrm{D}_k \,, \tag{S65}$$

$$\mathrm{A}_a + \mathrm{L}_i \to \mathrm{L}_j + \mathrm{L}_k \,, \tag{S66}$$

$$\mathrm{D}_i + \mathrm{L}_j \to 2\tilde{\mathrm{A}}_b \,, \tag{S67}$$

where $a, b = 1$ and $i, j, k = 1, \ldots, N_\mathrm{C}$ for species in the first compartment, and $a, b = 2$ and $i, j, k = (N_\mathrm{C} + 1), \ldots, (2N_\mathrm{C})$ for species in the second compartment. In addition, there are transfer reactions between compartments for all the species present:

$$\mathrm{D}_i \rightleftharpoons \mathrm{D}_{i+N_\mathrm{C}}, \tag{S68}$$

$$\mathrm{L}_i \rightleftharpoons \mathrm{L}_{i+N_\mathrm{C}}, \tag{S69}$$

$$\mathrm{A}_a \rightleftharpoons \mathrm{A}_{a+1}, \tag{S70}$$

$$\tilde{\mathrm{A}}_b \rightleftharpoons \tilde{\mathrm{A}}_{b+1}, \tag{S71}$$

which we assume are characterized by the same transition probability κ (where the convention that $\mathrm{E}_i = \mathrm{E}_{i+2N_\mathrm{C}}$ for the concentrations of enantiomers, $\mathrm{A}_a = A_{a+2}$, and $\tilde{\mathrm{A}}_b = \tilde{\mathrm{A}}_{b+2}$ is adopted). In the end, the kinetic rate equations of the first reactor are:

$$\dot{A}_d = -\sum_{\substack{aijk \\ j\leq k}} \delta_{da}\, w^{(\mathrm{D})}_{aijk} - \sum_{\substack{aijk \\ j\leq k}} \delta_{da}\, w^{(\mathrm{L})}_{aijk} + \frac{1}{\tau}(A_{0d} - A_d) + \kappa(A_{d+1} - A_d)\,, \tag{S72}$$

$$\dot{D}_m = \sum_{\substack{aijk \\ j\leq k}} \nu_{m,ijk}\, w^{(\mathrm{D})}_{aijk} - \sum_{\substack{bcij \\ b\leq c}} \delta_{mi}\, \tilde{w}_{bcij} + \frac{1}{\tau}(D_{m0} - D_m) + \kappa(D_{m+N} - D_m)\,, \tag{S73}$$

$$\dot{L}_m = \sum_{\substack{aijk \\ j\leq k}} \nu_{m,ijk}\, w^{(\mathrm{L})}_{aijk} - \sum_{\substack{bcij \\ b\leq c}} \delta_{mj}\, \tilde{w}_{bcij} + \frac{1}{\tau}(L_{m0} - L_m) + \kappa(L_{m+N} - L_m)\,, \tag{S74}$$

$$\dot{\tilde{A}}_e = \sum_{\substack{bcij \\ b\leq c}} (\delta_{eb} + \delta_{ec})\, \tilde{w}_{bcij} + \frac{1}{\tau}(\tilde{A}_{0e} - \tilde{A}_e) + \kappa(\tilde{A}_{e+1} - \tilde{A}_e), \tag{S75}$$

where $d, e = 1$; $m = 1, \ldots, N_\mathrm{C}$; and κ is the diffusive coupling parameter. Similar equations hold for the other reactor, where $d, e = 2$ and $m = (N_\mathrm{C}+1), \ldots, (2N_\mathrm{C})$. From these equations one can proceed by using the enantiomeric excess (S51) which obeys as before the equation:

$$\frac{d}{dt}\delta\mathbf{X} = \mathbf{M}\cdot\delta\mathbf{X}\,, \tag{S76}$$

with the matrix

$$\mathbf{M} \equiv \frac{\partial\dot{\mathbf{D}}}{\partial\mathbf{D}} - \frac{\partial\dot{\mathbf{D}}}{\partial\mathbf{L}}\,. \tag{S77}$$

Now the matrix $\mathbf{M}$ has the following block structure

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 - \kappa\mathbf{I} & \kappa\mathbf{I} \\ \kappa\mathbf{I} & \mathbf{M}_2 - \kappa\mathbf{I} \end{pmatrix},$$

where $\mathbf{M}_{1,2}$ represent the Jacobian matrix of compartments $1, 2$ respectively and $\mathbf{I}$ is the identity matrix of same dimension. In the limit of small κ, we can treat the effect of diffusion as a small perturbation. This perturbation will introduce a correction of the order of κ on the eigenvalues of the uncoupled case ($\kappa = 0$). Since the dominant eigenvalues in the uncoupled case are of the order of $\sqrt{N_\mathrm{C}}$ or N_C, depending on whether the scenario (i) or (ii) is relevant, this correction should have a small effect on the threshold of instability.

Let us call $\mathbf{u}_1$ (resp. $\mathbf{u}_2$) the eigenvectors of the matrix $\mathbf{M}_1$ (resp. $\mathbf{M}_2$) and the corresponding eigenvalues p_1 and p_2. A simple calculation provides the eigenvectors $\mathbf{u}$ and the eigenvalues p of the matrix $\mathbf{M}$ as function of $\mathbf{u}_{1,2}$ and $p_{1,2}$. In the case where the dominant eigenvalues of $\mathbf{M}_{1,2}$ are $p_1 = p_2 = \mu N_\mathrm{C} \pm O(\sigma)$ for both submatrices, one finds that the dominant contribution to p equals either μN_C or $\mu N_\mathrm{C} - 2\kappa$. The corresponding dominant eigenvectors have respectively uniform components across both compartments: $\mathbf{u} = (1, \ldots, 1)^\mathrm{T}$ or opposite components on each compartment: $\mathbf{v} = (1, \ldots, 1, -1, \ldots, -1)^\mathrm{T}$. The synchronization towards a global homochiral state occurs when the contribution of $\mathbf{u}$ wins over that of $\mathbf{v}$ on long times. Therefore, one then finds that such a synchronization should occur approximately when $\kappa \simeq \sigma/2$.

Another way to look at the synchronization of the states of the two compartments is to consider the evolution of the averaged enantiomeric excess in both compartments, defined by

$$\delta\bar{X}_i = \frac{1}{2}\left(\delta X_i + \delta X_{i+N_\mathrm{C}}\right). \tag{S78}$$

The equation of evolution of that quantity is controlled by a matrix $\bar{M}_{mn}$ such that

$$\bar{M}_{mn} = \frac{1}{2}\left(M_{mn} + M_{m+N_\mathrm{C},n}\right). \tag{S79}$$

In the notation of section S5, this may be written as

$$\bar{M}_{mn} = A_0 Q_{mn} - \frac{1}{2}\delta_{mn}(A_0 R_{mm} + 1/\tau) - \frac{1}{2}\delta_{m+N_C,n}(A_0 R_{mm} + 1/\tau), \tag{S80}$$

where we have used that $Q_{mn} = Q_{m+N_C,n}$ and $R_{m+N_C,m+N_C} = R_{mm}$. This property holds since the rate constants take exactly the same values in both compartments because the chemical composition and reactions in the two compartments are exactly the same. Although the concentrations of species take different values in the two compartments, their values do not enter in the stability of the trivial racemic fixed point. It follows from Eq. (S80), that the eigenvalues of the matrix $\bar{M}$ are exactly the ones we had before in the well-mixed case. From our study of the well-mixed case, we expect that the average enantiomeric excess should undergo an instability when the driving is sufficiently large. If the average enantiomeric excess reaches extremal values 1 or -1 at long times, then the two compartments must be both homochiral of the same chirality.

Using numerical simulations, we have confirmed this scenario. Firstly, when diffusion is weak for $\kappa \to 0$, we recover the previous scenario for a transition to homochirality, separately holding in each compartment As κ increases, so does the coupling between the two compartments. If there is a small bias present which is the same in the two compartments (L for instance), then one ends up with a homochiral state which is L in that case. The interesting case is therefore when the two compartments are given an opposite small bias initially. Then, as shown in Fig. S7, we find that as κ increases, we go from a global racemic state at small values of κ towards a global homochiral state when the coupling is sufficiently strong. The threshold of instability is found not to be significantly changed as compared to the well-mixed case in agreement with the theoretical argument given above. In addition, the threshold where the transition occurs is indeed of the order of $\sigma = 2 \times 10^{-4}$.

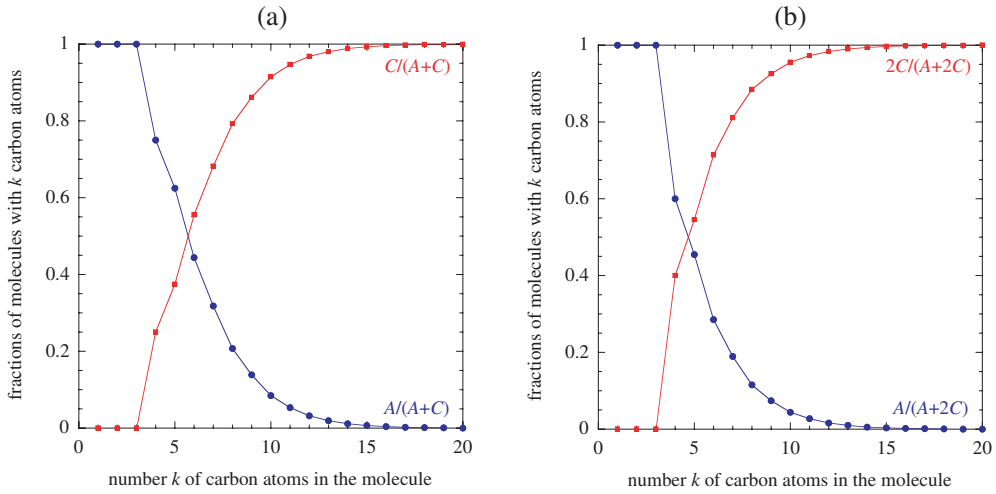


Figure S1: Fractions of achiral and chiral stereoisomers of monosubstituted alkanes versus the number k of carbon atoms in the molecule, counting (a) once and (b) twice the pairs of enantiomers. The data are from Ref. [18].

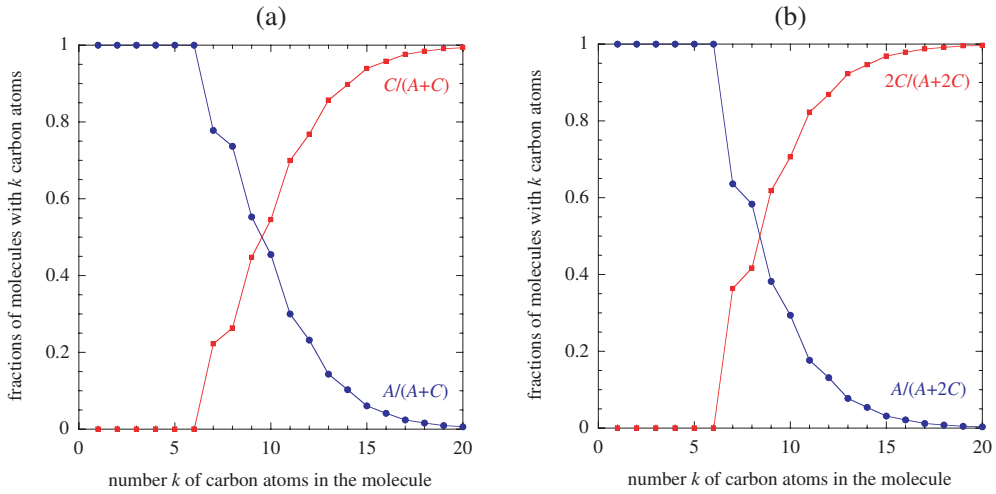
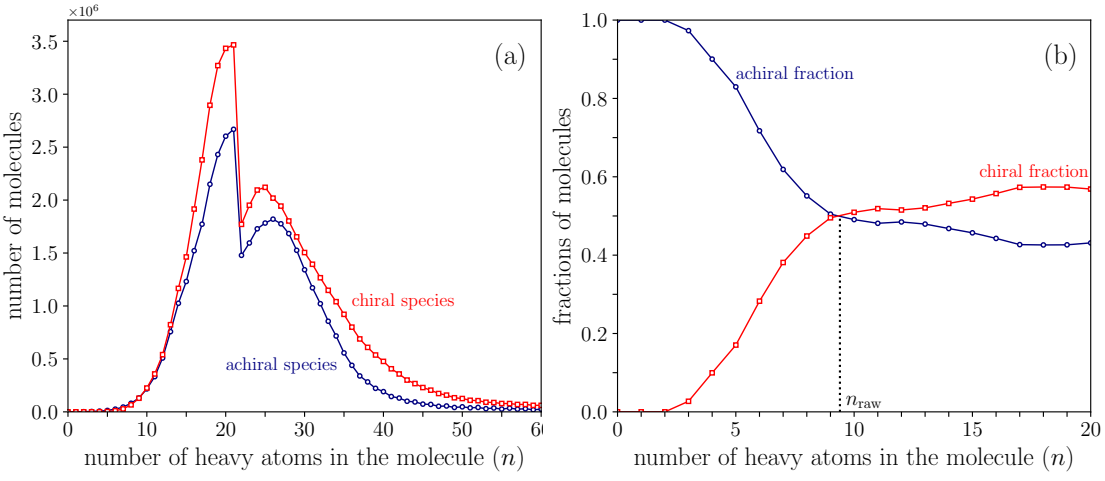


Figure S2: Fractions of achiral and chiral stereoisomers of alkanes versus the number k of carbon atoms in the molecule, counting (a) once and (b) twice the pairs of enantiomers. The data are from Ref. [17].



(a) (b)

Figure S3: (a) Total number of achiral and chiral species in the raw PubChem database containing about 139 millions of species. (b) Fractions of chiral and achiral molecules containing $n \leq 20$ heavy atoms. In this case, only $N = 33,563,343$ molecules with $n \leq 20$ heavy atoms were analyzed after the specific selection. The figure on the right shows an intersection at $n_{\text{raw}} \simeq 9.4$.

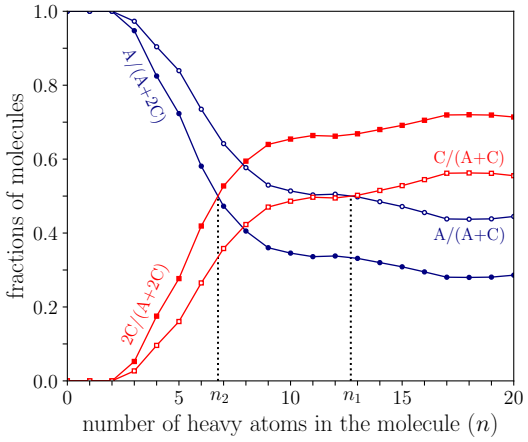


Figure S4: Analysis of the database expanded in enantiomers, with $N = 50,252,957$ molecules in data (*i.e.*, $16,689,614$ enantiomers were generated) with $n \leq 20$ heavy atoms. The intersection occurs at $n_2 \simeq 6.7$ for if both enantiomers are considered and $n_1 \simeq 12.7$ if only one enantiomer is considered.

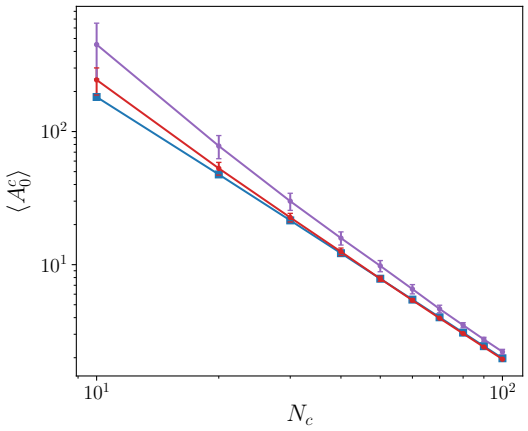


Figure S5: Comparison between the observed control parameter value A_0 at the transition with the theoretical prediction given by Eq. (6) of the main text (blue solid line) after averaging over 100 realizations of the rate constants for different standard deviation σ_{k_+} and $\sigma_{\tilde{k}_-}$ of rate constants : $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 10^{-3}$ (red), $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 10^{-2}$ (purple), while $\langle k_+ \rangle = \langle \tilde{k}_- \rangle = 10^{-4}$.

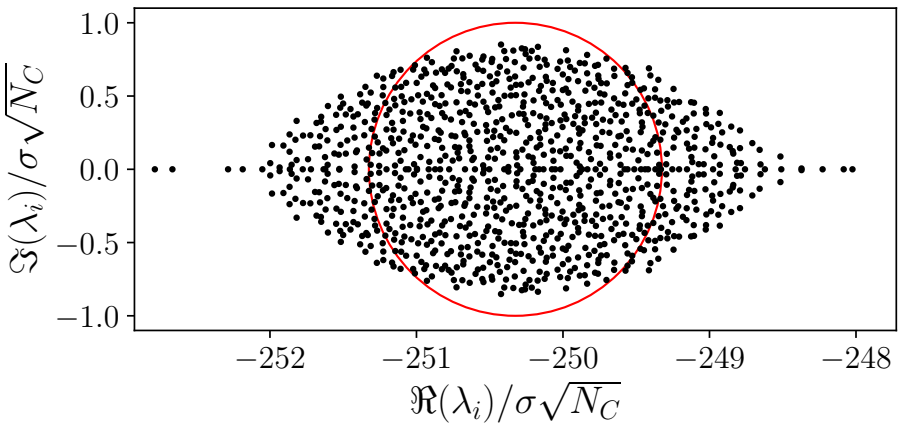


Figure S6: Close-up of the non-dominant eigenvalues of the Jacobian matrix **M** of the generalized Frank model, which do no fill the Ginibre circle. Here, the matrix **M** characterizes a system of $N_{\mathrm{C}} = 1000$ chiral species, with rate constants distributed according to a log-normal distribution of parameters $\langle k_+ \rangle = 10^{-4}$ and $\sigma_{k_+} = 2 \times 10^{-4}$ and $\tau = 1$. The parameter $A_0 = 25$, is far beyond the instability threshold for the trivial racemic state. The dominant eigenvalue is around 250 and lies outside the field of view for these parameters. The red circle is the unit circle, and the normalization factor $1/\sigma\sqrt{N_{\mathrm{C}}}$ is expressed in term of $\sigma = A_0\sigma_Q$ with $\sigma_Q = \sigma_{k_+}\sqrt{N_{\mathrm{C}}+1}$, which is the standard deviation of the matrix **Q** elements in the decomposition of the matrix **M** in section S5.

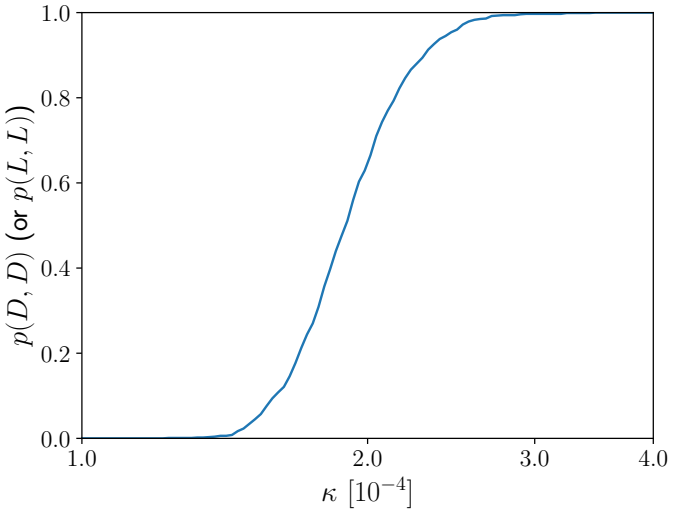


Figure S7: Probability to have two homochiral compartments of the same chirality on long times (either both are D or both are L) as function of the transition probability of transfer of molecules between the two compartments κ. The initial condition is such that there is an opposite chiral bias in both compartments, so that in the absence of diffusion coupling, the global state will be racemic. For one realization of the rate constants, the parameter κ is varied; then this procedure is repeated for different realizations of the rate constants. The curve has been done for a value of the driving force which is above the instability threshold. Simulations were carried out with an initial enantiomeric excess $\epsilon_1 = 10^{-2}$ and D- and L-enantiomers concentrations of all chiral species were initialized at $D_0 = 2 + \epsilon_1$ and $L_0 = 2 - \epsilon_1$ in the first compartment and with $\epsilon_2 = 1.5 \times 10^{-2}$ in the second compartment but favoring L-enantiomers. The unactivated achiral specie was initialized at $\tilde{A}_0 = 0$ and the activated one at $A_0 = 80$, far above the homochirality threshold in each compartment. All the constants k_{+ijk} and $\tilde{k}_{-ij}$ follow a log-normal distribution of parameters $\mu = -10.02$ and $\sigma = 1.27$ (*i.e.*, corresponding to a log-normal distribution with $\langle k_+ \rangle = \langle \tilde{k}_- \rangle = 10^{-4}$ and $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 2 \times 10^{-4}$), with $\tilde{k}_{ij} = \tilde{k}_{ji}$ to satisfy the mirror symmetry described in Eq. (14). The number of chiral species was set up to $N_C = 20$.