



HAL
open science

A statistical view on the origin of homochirality

Gabin Laurent

► **To cite this version:**

Gabin Laurent. A statistical view on the origin of homochirality. Organic chemistry. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLS043 . tel-04057861

HAL Id: tel-04057861

<https://pastel.hal.science/tel-04057861>

Submitted on 4 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Ecole Supérieure de Physique et de Chimie
Industrielles de la ville de Paris, au laboratoire Gulliver

A statistical view on the origin of homochirality

Perspectives statistiques de l'origine de l'homochiralité

Soutenue par

Gabin Laurent

Le 09 novembre 2022

École doctorale n°564

Physique en Île-de-France

Spécialité

Physique

Composition du jury :

Cornelia Meinert Directrice de recherche Université Côte d'Azur - CNRS	<i>Rapportrice Présidente du jury</i>
Jeanne Crassous Directrice de recherche Université de Rennes - CNRS	<i>Rapportrice</i>
Raphaël Plasson Maître de conférences Université d'Avignon	<i>Examineur</i>
Farshid Jafarpour Assistant professor Utrecht University	<i>Examineur</i>
Pierre Gaspard Full Professor Université Libre de Bruxelles	<i>Examineur</i>
David Lacoste Directeur de recherche ESPCI Paris - CNRS	<i>Directeur de thèse</i>
Ulrich Gerland Full Professor Technical University of Munich	<i>Co-encadrant de thèse</i>

Acknowledgments

Durant ces trois années de doctorat, j'ai pu rencontrer, discuter et échanger avec bon nombre de personnes au laboratoire Gulliver, à l'ESPCI, à l'Université Technique de Munich ou encore ailleurs, à qui je souhaite en premier lieu faire honneur. Je souhaite donc remercier :

Cornelia Meinert et Jeanne Crassous, pour avoir accepté de prendre le temps de rapporter cette thèse, ainsi que les examinateurs membres du jury : Farshid Jafarpour, Pierre Gaspard et Raphaël Plasson.

Mon directeur de thèse, David Lacoste, que j'ai rencontré en début d'année 2019 et avec qui j'ai eu la chance de travailler pendant ces trois dernières années, pour la supervision de mon travail de thèse. Je le remercie pour sa présence, ses conseils, sa grande bienveillance et la confiance qu'il m'a accordé dès le début.

Ulrich Gerland, my co-advisor, who welcomed me for one and a half years in his group in Munich with great kindness, for the freedom he gave me to carry my own project, and for his trust, his advice, and comments on my work. I had the unique opportunity to benefit from two complementary supervisions during those three years, and I thank you, Uli and David, for that.

Philippe Nghe, sans qui cette thèse n'aurait jamais vu le jour.

L'Initiative de Recherches Interdisciplinaires et Stratégiques - Origines et Conditions d'Apparition de la Vie (IRIS-OCAV) qui a financé le travail effectué au sein du laboratoire Gulliver à l'ESPCI Paris.

The Origins Excellence Cluster which funded my Ph.D. position and my work at the Technical University of Munich.

Pierre Gaspard une nouvelle fois, pour la chance et le grand plaisir que j'ai eu à collaborer avec pour notre projet sur l'homochiralité.

Luca Peliti, Arthur Duprat, Marina Katava et Alex Blokhuis, pour les nombreuses discussions scientifiques autour de mon travail.

Ludovic Jullien and Zorana Zeravcic, the two members of my thesis committee, for their strong involvement in the smooth running of my thesis from both scientific and human perspectives.

Hyo Jin Cho, Fée Sorrentino, David Noël et Elisa Silveira, secrétaires et gestionnaires du laboratoire Gulliver, pour leur disponibilité et sympathie. I thank the two secretaries of the group in Munich as well, Daniela Neufang and Claudia Öckler, for their help.

Tobias Göppel, who has been of great help during my time in Munich, and the members of the origins of life group: Johannes, Ludwig, and Julio but also all the other members

of the lab.

Armand, Maitane, Yann, Claire et Paul du laboratoire Gulliver, pour les conversations partagées autour de repas, les discussions sur les travaux de chacun qui furent toutes très agréables ainsi que les autres doctorants et post-doctorants du laboratoire, bien trop nombreux pour être tous nommés ici.

Mes professeurs de physique et de chimie, du lycée et de classes préparatoires : Delphine Cossardeaux, Catherine Brune et Jérôme Vimal pour leurs enseignements et leur profonde humanité.

Arthur Genthon, mon ami “co-thésard”, pour sa joie de vivre contagieuse... et tous ces rires, avec qui j’ai eu la grande chance de partager une bonne partie de mon quotidien au laboratoire pendant ces trois années de galère.

Mes très chers amis Maxime, Etienne, Antoine et Rani, pour avoir pollué ma galerie photo de téléphone avec quantité de memes douteux au fil du temps.

Ma mère Isabelle, mon grand-père Bernard et ma sœur Jeanne pour leur soutien.

Enfin, Iris, pour son soutien moral et tout son amour depuis plusieurs années maintenant, et, je l’espère, pour les nombreuses à venir.

Publications

Five publications were produced for the duration of this thesis. Only the work of Refs. 2, 3 and 4 is presented in details in this manuscript.

1. Gabin Laurent, Luca Peliti and David Lacoste
Survival of Self-Replicating Molecules under Transient Compartmentalization with Natural Selection
Life, **2019**, 9, 4
2. Gabin Laurent, David Lacoste and Pierre Gaspard
Emergence of Homochirality in Large Molecular Systems
Proceedings of the National Academy of Sciences, **2021**, 118, 3
3. Gabin Laurent, Pierre Gaspard and David Lacoste
A Robust Transition to Homochirality in Complex Chemical Reaction Networks
Proceedings of the Royal Society A, **2022**, 478, 2257
4. Gabin Laurent, Tobias Göppel, David Lacoste and Ulrich Gerland
Template-Directed Ligation: an Ideal Candidate for Homochirality Emergence?
Under writing process
5. Marina Katava, Gabin Laurent and David Lacoste
Thermodynamics of dimerization in open reactors
Under writing process

Contents

Résumé	7
1 General introduction	13
1.1 Context	13
1.1.1 What is homochirality?	13
1.1.2 What we are specifically interested in	14
1.1.3 Three different and relative conventions	15
1.1.4 The importance to be out of equilibrium	17
1.1.5 Some proposed theories on homochirality	19
1.1.6 Original Frank model	20
1.2 Outline	22
2 A robust transition to homochirality in large and complex chemical networks	23
2.1 Study of the abundance of chiral species in nature	24
2.1.1 Previous results	25
2.1.2 Analysis of the PubChem database	28
2.2 A general model for large reaction networks	31
2.2.1 The continuous stirred tank reactor dynamics	32
2.2.2 About mirror symmetry	33
2.2.3 The stability of the racemic state	35
2.2.4 Random matrices and the stability of dynamical systems	37
2.2.5 A criterion for the racemic state stability	39
2.2.6 Sparse chemical networks	41
2.2.7 What about homochiral fixed points?	43
2.3 Example: A generalization of the Frank model	44
2.3.1 The generalized model	44
2.3.2 Stability of the steady-states	48
2.3.3 How does the sparsity of the network affect the results?	56

2.3.4	Correlations in the Jacobian matrix	57
2.4	Discussion	58
2.A	Additional data from the PubChem analysis	60
2.B	Eigenvalues of a random matrix with non zero mean	61
2.B.1	Isolated eigenvalue of a non zero mean random matrix	61
2.B.2	Perturbation of other eigenvalues	62
2.C	Elements of the Jacobian in the homochiral state	63
2.D	Calculation of the maximal eigenvalue of the Jacobian matrix	65
2.E	Compartmentalized systems with diffusive coupling	68
3	Template-directed ligation: an ideal candidate for homochirality emergence?	71
3.1	Random polymerization	72
3.1.1	A simple model for polymerization	72
3.1.2	The Gillespie algorithm	73
3.1.3	The incapacity of step-growth polymerization to produce a homochiral system under closed conditions	75
3.2	Template-directed ligation	77
3.2.1	Hybridization and dehybridization	77
3.2.2	Ligation and stalling	80
3.2.3	Hydrolysis	81
3.2.4	Temperature cycles	82
3.2.5	Racemization reactions	82
3.2.6	Previous works	82
3.3	Closed reactor	83
3.3.1	Quantities of interest	84
3.3.2	Irrelevance of block energy model details	85
3.3.3	System without racemization	85
3.3.4	Simulations with racemization	86
3.3.5	Effects of the racemization reaction speed	88
3.3.6	Effects of temperature cycles	89
3.3.7	Effects of the chiral stalling amplitude	90
3.4	Open reactor	91
3.4.1	Racemic chemostat	91
3.4.2	Templated ligation with a biased chemostat	92
3.5	Discussion	94

3.A Comparison of block energy models	97
3.B Single trajectories for closed reactors	98
3.C Single trajectories for open reactors	99
4 Conclusion	101
4.A Survival of Self-Replicating Molecules under Transient Compartmentaliza- tion with Natural Selection	104
Bibliography	117

Résumé

L'homochiralité, encore appelée asymétrie des molécules biologiques est un phénomène observé dans la matière vivante sur Terre et qui est à ce jour, encore inexpliqué. Les mécanismes sous-jacents à la vie reposent principalement sur des molécules chirales (comme les acides aminés qui composent les protéines, ou nucléotides qui composent l'ADN ou l'ARN, par exemple) qui sont par définition, des molécules qui ne sont pas superposables à leur image dans un miroir plan. Les deux versions, images l'une de l'autre, de ces molécules, forment alors un couple d'énantiomères. L'homochiralité du vivant signifie précisément que pour chaque molécule chirale, seulement un seul des deux possibles énantiomères est observé alors que l'autre ne l'est pas, ou est présent seulement à l'état de traces [Meierhenrich, 2008]. Par exemple, tous les acides aminés protéinogènes chiraux sont observés sous la forme L, et tous les nucléotides qui forment les acides nucléiques contiennent des D-riboses. D'un point de vue statistique, la synthèse chirale est en majorité symétrique, c'est à dire qu'autant de chaque énantiomère est produit par une réaction chimique impliquant des réactifs achiraux, ce qui semble naturellement incompatible avec l'émergence de l'homochiralité. Il apparaît alors naturel de se demander comment et à travers quel mécanisme biochimique ou physique, l'homochiralité du vivant est apparue, et à quel moment dans les scénarios sur les origines de la vie.

Le degré d'homochiralité d'un système est quantifié par l'écart énantiomérique, qui mesure l'écart normalisé de concentration entre les énantiomères d'une espèce chirale. Les scénarios d'émergence de l'homochiralité comportent trois phases distinctes : (i) l'apparition d'un très faible écart énantiomérique, soit provoqué par un processus physique ou chimique particulier, soit dû à des fluctuations autour de l'état racémique (l'état d'égalité de concentrations entre les deux énantiomères d'une même molécule chirale), (ii) l'amplification de cet écart initial, jusqu'à des valeurs proches d'une homochiralité totale, puis (iii) la maintenance de cet état homochiral et le transfert de cette asymétrie à d'autres groupes de molécules. Dans ce manuscrit, nous nous intéressons principalement à la phase d'amplification de l'écart énantiomérique initial.

Les systèmes chimiques chiraux ont une propriété particulière : la symétrie miroir. Les deux énantiomères d'une espèce chimique chirale ont très peu de différences : ils interagissent de la même manière dans un environnement symétrique et ont les mêmes propriétés physiques (à l'exception de leur interaction avec de la lumière polarisée circulairement), et ont aussi les mêmes propriétés chimiques dans un milieu achiral. Toutefois, cela change en environnement chiral, et un principe actif chiral peut avoir un effet totalement différent sur le corps humain dont les molécules formant les biopolymères présentent

une homochiralité (cf. scandale lié à la Thalidomide aux Etats-Unis dans les années 60, dont l'un des énantiomères soulage les nausées des femmes enceintes alors que l'autre a un effet tératogène et est à l'origine de malformations chez le fœtus). Cependant, pour toute réaction impliquant un composé chiral, une réaction miroir impliquant les énantiomères de chaque réactif et produit existe également, et ce, avec les mêmes constantes réactionnelles. On appelle cela la symétrie miroir des réseaux chimiques chiraux. Cette symétrie a pour effet que l'état d'équilibre d'un système chimique chiral fermé est racémique. En effet, tout système à l'équilibre vérifie le bilan détaillé qui contraint les concentrations des espèces chimiques d'un système à l'équilibre et la symétrie chirale implique des concentrations égales pour chaque couple d'énantiomères. Il apparaît alors nécessaire pour l'émergence de l'homochiralité de considérer des systèmes maintenus hors d'équilibre, soit par des flux de matière entrante et sortante, soit par des chemostats ou un apport énergétique. Dans ce manuscrit, nous présentons deux systèmes distincts et maintenus hors d'équilibre dans lesquels l'homochiralité serait susceptible d'émerger.

Tout d'abord dans le Chapitre 1 introductif, nous présentons et précisons les points évoqués précédemment concernant l'homochiralité et les propriétés des réseaux chimiques chiraux. Nous présentons également brièvement quelques travaux précédemment publiés par divers chercheurs et chercheuses sur l'origine de l'homochiralité. L'un des travaux pionniers est le modèle de Frank [Frank, 1953], qui est un modèle simple décrivant un système chimique contenant les deux énantiomères d'une espèce chirale ainsi qu'une espèce achirale. Les énantiomères peuvent chacun réagir avec l'espèce achirale au sein d'une réaction autocatalytique et stéréosélective, et peuvent également réagir entre eux et s'inhiber. Frank montra en 1953 que ce genre de système, maintenu hors d'équilibre, exhibait une transition de l'état racémique à un état homochiral. Cependant, ce modèle peut-être jugé trop simple : la réaction autocatalytique et stéréosélective qui le compose a été mise en évidence seulement 40 ans après [Soai et al., 1995], et la réaction d'inhibition chirale n'est pas non plus expérimentalement évidente à mettre en œuvre. De plus, le modèle comporte seulement 2 espèces chimiques, alors que le système prébiotique dans lequel l'homochiralité pourrait avoir émergé serait plutôt susceptible d'avoir une composition complexe.

Motivés par ces limitations, au cours du Chapitre 2 nous nous intéressons au cas d'un réacteur parfaitement mélangé maintenu hors d'équilibre par des flux de matière entrante et sortante, dans lequel des réactions chimiques se produisent. En utilisant la propriété de symétrie miroir d'un tel système et en effectuant une analyse de stabilité, il est possible de montrer que l'état racémique d'un tel système perd sa stabilité lorsque le nombre d'espèces chirales qu'il contient devient grand. Ce résultat découle de l'hypothèse que le réseau chimique considéré est complexe et sans hypothèse spécifique sur sa structure.

Il en résulte que le Jacobien du système peut-être considéré comme étant une matrice aléatoire [Girko, 1985], ce qui simplifie la résolution de son spectre et donc de l'étude de la stabilité du système qui dépend des valeurs propres de cette matrice. De cette analyse découle que la stabilité de l'état racémique peut être brisée de deux façons : (i) soit pour un sous-ensemble des espèces chimiques et dans des directions différentes (avec différentes chiralités), (ii) soit pour toutes les espèces chimiques composant le système et ce dans la même direction, et cela dépend des propriétés du système ainsi que du nombre d'espèces chimiques chirales qui le composent. Quoiqu'il en soit, ce type de système est susceptible de présenter une transition vers un état homochiral lorsque le nombre d'espèces chirales augmente. De plus, le mécanisme par lequel la stabilité de l'état racémique se trouve brisée apparaît être robuste vis à vis de la sparsité du réseau chimique ou bien de la matrice Jacobienne, qui caractérise sa stabilité.

Ce modèle générique est ensuite illustré avec une généralisation du modèle de Frank précédemment introduit, qui contient alors un grand nombre d'espèces chirales. En effectuant un développement similaire au cas du modèle plus général, nous montrons que ce modèle admet aussi une transition due à une brisure de symétrie d'un état racémique à un état homochiral lorsque le nombre d'espèces chirales qu'il contient devient grand. Cette transition est aussi possible lorsque la concentration d'espèces achirales entrant dans le réacteur augmente, ce qui correspond à fournir de plus en plus d'énergie au réacteur, soit, à le maintenir plus loin de l'équilibre. La transition observée pour ce modèle de Frank généralisé montre aussi une robustesse par rapport à la sparsité du réseau chimique, ainsi que vis à vis des corrélations dans la matrice Jacobienne, qui modifient ses propriétés spectrales. De plus, nous observons des synchronisations de chiralité pour le modèle général ainsi que le modèle généralisé de Frank dans le cadre de systèmes inhomogènes à deux compartiments, lorsque la diffusion entre les deux réacteurs est suffisamment importante. Aussi, une propriété très importante de ces modèles est qu'ils ne sont basés sur aucune propriété structurale des molécules, et que la chiralité de celles-ci n'est qu'une étiquette, qui peut être interchangé. Cela signifie que ces modèles n'imposent pas une seule et même chiralité pour toutes les molécules composant l'état homochiral, ce qui serait contraire aux observations actuelles de la matière biologique (acides aminés protéinogènes de chiralité L et riboses de chiralité D).

Une analyse de base de données chimiques est aussi menée dans le Chapitre 2, pour quantifier l'abondance des molécules chirales dans la nature. Il paraît logique que le nombre de molécules chirales possibles augmente avec le nombre d'atomes lourds contenus dans une molécule, mais une quantification exacte ou approchée n'est pas disponible dans la littérature, hormis pour certains groupes bien définis de molécules comme les alcanes et

les alcanes monosubstitués. Nous déterminons alors le seuil au delà duquel les molécules chirales existent en abondance par rapport aux molécules achirales, qui a lieu pour $n \simeq 9$ atomes lourds par molécule. Il apparaît donc possible pour un système composé d'un grand nombre d'espèce chirales d'exister dans une chimie composée de molécules qui ne sont pas encore trop complexes.

Ensuite, dans le Chapitre 3, nous présentons un mécanisme basé sur la polymérisation sur « template » de biopolymères tels que l'ARN. Ce mécanisme implique l'extension de brins d'ARN par ligation avec d'autres oligomères ou monomères compatibles par appariement par paires de base. Ce processus est par essence autocatalytique puisque chaque brin d'ARN sert de modèle pour sa réplication. De plus, des expériences ont mis en évidence par le passé un caractère stéréosélectif de la réplication sur template : lorsque la chiralité des monomères au niveau du site de ligation est différente de celle du template, la ligation se trouve ralentie entre 10 et 100 fois [Joyce et al., 1984, Bolli et al., 1997]. Ce phénomène, qu'on appelle « chiral stalling » induit une stéréosélectivité dans le processus de polymérisation sur template.

Dans un premier temps, nous étudions un système dans lequel des oligomères et des monomères se lient par polymérisation classique. En utilisant des simulations numériques stochastiques, nous montrons que la polymérisation classique ne permet pas d'induire une brisure de symétrie chirale significative dans un système initialement racémique, même en présence d'une influence extérieure (comme une surface telle que l'argile de Montmorillonite ou une enzyme chirale) qui impliquerait un biais de sélection chirale lors de la polymérisation de l'ARN. En effet, la polymérisation traditionnelle, même en présence d'un biais de sélection chirale ne présente pas de comportement autocatalytique qui favoriserait une des deux chiralités si un léger déséquilibre dans l'écart énantiomérique se formait par hasard. En revanche, nous montrons qu'un système fermé, initialement inoculé avec des monomères et des dimères de manière racémique, montre une transition vers un état totalement homochiral grâce à la polymérisation sur template lorsque le chiral stalling est pris en compte. Ce phénomène est observé en présence de réactions de racémisation qui interconvertissent la chiralité de nucléotides isolés, ainsi qu'en présence de cycles de températures. En effet, l'appariement entre deux brins d'ARN devient très stable lorsque les brins d'ARN sont longs, et les cycles de températures empêchent un état gelé pour le système, dans lequel tous les brins seraient hybridés à un autre.

Nous étudions aussi l'effet de la rapidité des réactions de racémisation, des cycles thermiques et de l'amplitude du chiral stalling. Il apparaît premièrement que la transition observée dans un système fermé se produit même lorsque les réactions de racémisation des monomères se produisent lentement. De plus, la durée entre les pics de chaleur qui

permettent de séparer les brins d'ARN hybridés influe principalement sur les échelles de temps de la transition, mais, pour des élévations de températures espacées par de longs cycles, le degré d'homochiralité final semble diminuer tout en restant cependant élevé. Enfin nous remarquons que les effets de l'amplitude du chiral stalling sont forts : il semble exister un seuil au dessous duquel, lorsque le ralentissement de la ligation entre deux brins n'est pas assez prononcé en présence d'incompatibilités chirales, la brisure de symétrie ne se produit pas dans le système, ou alors pas dans un temps raisonnable.

Le cas du réacteur ouvert est également étudié pour vérifier s'il tend à exhiber le même type de transition, mais cette fois ci sans la présence de réactions de racémisation. Ce type de réacteur est maintenu hors d'équilibre par un chemostat, c'est à dire que le nombre de monomères et de dimères est constant au cours du temps, et par une dégradation constante des espèces chimiques. Dans le cadre d'un chemostat non biaisé, donc racémique, le système ne semble pas atteindre un écart énantiomérique significatif dans l'état stationnaire. Cependant, un écart énantiomérique transitoire significatif peut-être observé. Dans le cas d'un chemostat biaisé donc chiral, une amplification très prononcée de l'écart énantiomérique initial est observée. Ce résultat semble important puisqu'il a déjà été démontré par le passé la possibilité d'une production asymétrique de D-riboses en présence d'acides aminés ou de dipeptides d'une chiralité donnée qui pourrait être à l'origine d'un biais dans une solution de nucléotides [Pizzarello and Weber, 2004, Weber and Pizzarello, 2006]. Cette amplification, qui est impossible avec un mécanisme de polymérisation classique, produit des écarts énantiomériques finaux jusqu'à 7 fois supérieurs à ceux du chemostat.

Ces résultats montrent que l'émergence de l'homochiralité semble possible dans un système polymérique sujet à la polymérisation sur template, et que le chiral stalling joue un rôle crucial ainsi que les réactions de racémisation (dans le cadre d'un système fermé uniquement). De plus, dans le cas du système fermé, la convergence du système vers un état totalement homochiral permet l'augmentation de la longueur moyenne des polymères puisque le ralentissement de la ligation dû au chiral stalling ne se manifeste plus. Ce phénomène est intéressant pour les scénarios d'origine de la vie puisque l'augmentation de la longueur des polymères permet d'accéder à une plus grande partie de l'espace des séquences théoriquement possible pour les brins d'ARN, et notamment facilite la probabilité d'apparition de séquences fonctionnelles qui nécessite en général des brins d'au moins 100 à 150 monomères de long. Ce mécanisme d'émergence de l'homochiralité trouverait toute sa place dans la théorie du monde ARN [Higgs and Lehman, 2015], qui propose l'idée qu'un biopolymère comme l'ARN a un jour servi de porteur d'information ainsi que de molécule fonctionnelle.

Enfin, le Chapitre 4 présente une conclusion générale au travail mené pendant ces trois

années de thèse sur l'émergence de l'homochiralité, et un article publié sur l'importance des compartiments dans les scénarios d'origine de la vie est présenté en Annexe [4.A](#).

Contents

1.1	Context	13
1.1.1	What is homochirality?	13
1.1.2	What we are specifically interested in	14
1.1.3	Three different and relative conventions	15
1.1.4	The importance to be out of equilibrium	17
1.1.5	Some proposed theories on homochirality	19
1.1.6	Original Frank model	20
1.2	Outline	22

1.1 Context**1.1.1 What is homochirality?**

Homochirality, also called biological asymmetry, is a long-standing problem in the origins of life research. Life on earth relies mainly on chiral molecules, which are, by definition, molecules not superposable to their mirror images. In a more formal way, a chiral molecule is a molecule that does not admit any axis of improper symmetry (*i.e.* a combination of a rotation and a reflection in a plane perpendicular to the rotation axis) [Collet *et al.*,

2006, Harris et al., 1999]. The biological asymmetry [Meierhenrich, 2008] describes the current observation that chiral biomolecules composing biopolymers are present under only one of their two possible forms (D or L) in living matter. For instance, in living matter, proteins are composed of amino acids that are - on earth - observed to be all L-chiral, and nucleic acids such as DNA and RNA contain only nucleotides formed out of D-ribose. Many different mechanisms have been studied to explain the emergence of homochirality [Sallembien et al., 2022]. However, to this date, none has been estimated as a convincing explanation by the majority of the scientific community. In a symmetric environment or an achiral environment, the synthesis of chiral compounds is qualified as “symmetric”, meaning that the products of a reaction involving achiral compounds are racemic (referring to a solution with as many D-enantiomers as L-enantiomers). It is, for example, illustrated in the Miller-Urey experiment [Miller, 1953] whose main aim was to produce the basic building blocks of life, (*e.g.* amino acids) from basic compounds found in the atmosphere under realistic prebiotic conditions. The products of these experiments were present in essentially racemic concentrations. Obviously, considering a finite number of products of such a symmetric reaction [Blackmond, 2010], for an odd number of molecules, the number of D-molecules cannot equal exactly the number of L-molecules, but statistically, this kind of synthesis is qualified symmetric, with small fluctuations about the racemate. Homochirality is thus incompatible with symmetric environments in the absence of any bias. Therefore, several questions emerge concerning homochirality: how did it emerge? When? Before or after the advent of life? Since homochirality is observed in living systems, is it a requirement for life? Or a consequence of it? This thesis will focus on “how” and “when” it emerged. The latter questions are, however, also fundamental issues [Brandenburg, 2021] because if homochirality came to be a consequence of life, it could be one of its signature and could then be used to detect extraterrestrial life.

1.1.2 What we are specifically interested in

The emergence of homochirality is a three-step process: (i) first, a minute imbalance appears in the enantiomeric excess (*e.e.*) of a chiral system. The enantiomeric excess is defined as

$$e.e. = \frac{|c_D - c_L|}{c_D + c_L}, \quad (1.1)$$

where c_D (*resp.* c_L) is the concentration of D-enantiomers (*resp.* L-enantiomers) and is the fundamental quantity that characterizes the degree of homochirality of a chiral system. Then, (ii) the imbalance is amplified over time, until the *e.e.* converges to unity. Finally, (iii) homochirality is maintained over time and transferred to other groups of

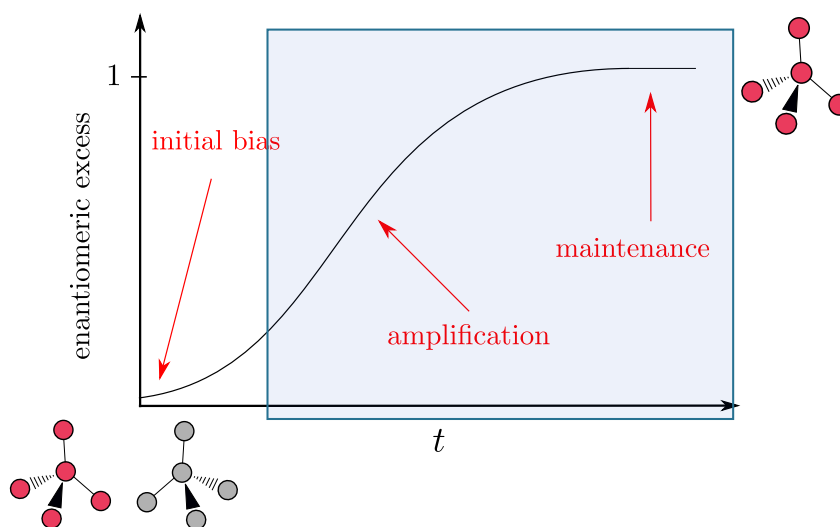


Figure 1.1: Graphical representation of a transition from a system with a slight deviation from the racemic state to a homochiral state. Red and grey molecules represent two enantiomers of a given chiral species. The transparent blue frame highlights the step of the mechanism which are studied in this manuscript.

molecules. A graphical representation of such transition is depicted in Figure 1.1. In general, mechanisms that seek to explain the initial imbalance are distinct from those that aim to explain the amplification process. This manuscript will focus on the amplification mechanism and its maintenance, irrespective of the phenomenon that produces an initial imbalance.

1.1.3 Three different and relative conventions

In 1848, Pasteur discovered [Pasteur, 1848, Vantomme and Crassous, 2021] that salts of synthetic tartaric acids contained two types of crystals, which are mirror images of each other. He observed that the two different types of crystal rotated linearly polarized light in opposite directions, and from this observation, the optical activity became an important method to measure the chirality of a given solution. Substances are called *dextrorotary* and denoted $d(+)$ if the rotation angle is clockwise, and called *levorotary* and denoted $l(-)$ if the rotation angle is anti-clockwise. Although this classification has an important place in history, three major hurdles arise: (i) it is not straightforward to determine the optical activity of a chiral compound by looking at its chemical formula, (ii) the optical activity of a substance depends on the solvent of the solution, and (iii) when multiple chiral centers are present in a chiral molecule, the determination of the chirality by the measurement of the optical rotation becomes inaccurate (however, an additive

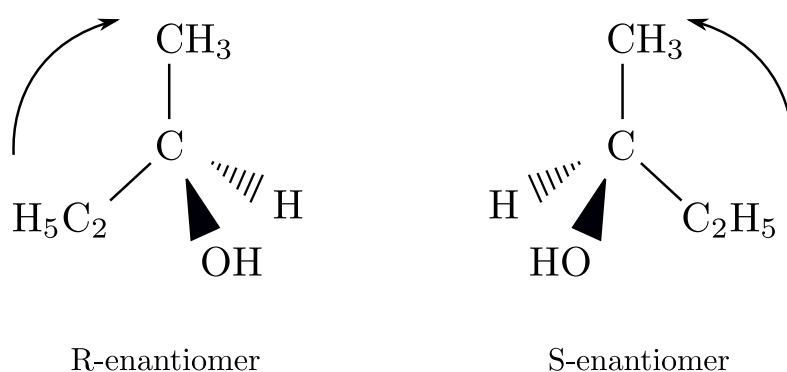


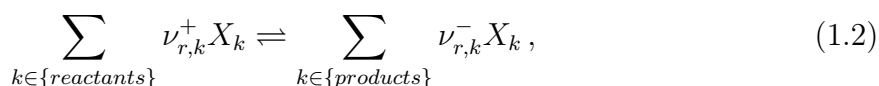
Figure 1.2: Two enantiomers of a given chiral species.

rule exists to obtain the optical rotation of a solution in which different enantiomers coexist in the absence of the Horeau effect [Collet et al., 2006, Horeau and Guetté, 1974]. The only direct link between optical activity and chirality is established for molecules with only a single chirocenter, such as amino acids or simple sugars. A second common convention uses the R/S descriptors for stereocenters, in which R (*resp.* S) stands for Rectus (*resp.* Sinister), which, in Latin, means right-handed (*resp.* left-handed). This convention consists of ordering the groups attached to a stereocenter based on their atomic number, according to a set of priority rules introduced by Cahn, Ingold, and Prelog in 1966 [Cahn et al., 1966]. After assigning the priority to each molecular group, the molecule is oriented such that the lowest priority group points away from the observer, and the rotation direction of the molecule is dictated by the sense of rotation of the trajectory that goes through groups from the lowest to the highest priority. If the rotation is clockwise, the chiral center is denoted R, and if it is anti-clockwise, S (see Figure 1.2). This description has the advantage of being deduced directly from the three-dimensional structure of the molecule. If several stereocenters are present, the isomer is described by the full list of R/S descriptors for each chiral center of the molecule. We add, however, two remarks: first, there is no correlation between the R/S convention and the optical activity of a molecule. Also, this description is not invariant for given functional groups: it is, for example, the case for proteinogenic amino acids that are all L-molecules. L-alanine is also the S-alanine while L-cysteine is also named R-cysteine. Finally, the third convention is the D/L convention where D (*resp.* L) stands for Dexter (*resp.* Laevus), the Latin for right (*resp.* left) and which is also known as the Fisher-Rosanoff convention. Fisher introduced a planar representation for sugars based on their carbon chain, and in 1906, Rosanoff chose the glyceraldehyde (which is a monosaccharide) as the reference for the stereochemistry of carbohydrates and other chiral molecules [Rosanoff, 1906]. He then arbitrarily assigned the descriptor D (*resp.* L) to the (+)-glyceraldehyde (*resp.* (-

)-glyceraldehyde). This notation has been universally accepted and used to obtain the relative configuration of chiral molecules. It allows the division of chiral molecules such as amino acids and carbohydrates into two families, the L and the D series, which are separately conserved for functional groups of molecules forming biopolymers such as amino acids. However, once again, no correlations are observed with the optical activity of chiral molecules or the R/S convention. Those three different conventions aimed to measure chirality are arbitrary. Throughout this manuscript, we will use the D/L convention to denote the enantiomers of chiral species. As stated previously, in the D/L convention, all proteinogenic amino acids are L-oriented, and sugars are D-oriented in biopolymers. Looking at the R/S convention, most amino acids have an S chiral center, but there are some exceptions, such as R-cysteine. Depending on the initial choice for each convention, the descriptor could be the opposite. The fact that conventions are arbitrary implies that the only meaning of homochirality is the dominance of one enantiomer over the other one, for a given chiral species, independently of what happens for the other chiral species. Some models in the literature studied how a group of molecules can become collectively homochiral in which all molecules adopt the same sign (*e.g.* polymeric systems [Liu, 2020, Li et al., 2021]). However, it does not follow from the definition of homochirality that all chiral biomolecules should have the same sign. Understanding the relative signs of different groups of molecules (*e.g.* the opposite signs of amino acids and nucleic acids) is still an open question.

1.1.4 The importance to be out of equilibrium

There are a few fundamental differences between two enantiomers of a given chiral species. They interact similarly in a symmetric environment and have the same physical properties (except for their interaction with circularly polarized light). They also have the same chemical properties within achiral media, with the same rate constants for reactions with achiral molecules. For reactions involving chiral reactants, a mirror-image reaction involves enantiomers of the reactants and products of the first reaction, with the same rate constant. It means that, considering the following reaction involving species X_k that can be, or not, chiral species



of forward (*resp.* backward) rate constant k_{+r} (*resp.* k_{-r}), where $\nu_{r,k}^+$ (*resp.* $\nu_{r,k}^-$) is the stoichiometric coefficient of species k for the reaction r in the forward (*resp.* backward)

reaction direction, the mirror reaction

$$\sum_{k \in \{\text{reactants}\}} \nu_{r,k}^+ \overline{X}_k \rightleftharpoons \sum_{k \in \{\text{products}\}} \nu_{r,k}^- \overline{X}_k, \quad (1.3)$$

exists. Here, \overline{X}_k is the opposite enantiomer if species X_k is chiral (and $\overline{X}_k = X_k$ if the species is achiral), and the rate constants in both reaction directions are the same for the two mirror reactions. This property of chiral reaction networks is called the *mirror symmetry* and has an important consequence for their dynamics within a closed system. In a closed system, the concentrations of the chemical species converge towards a stationary equilibrium solution. At a stationary point, the sum of the production flux for a given chemical species equals the sum of the degradation fluxes through chemical reactions. In other words, at a stationary state,

$$\sum_r \nu_{r,a}^- w_{-r,a}(\mathbf{c}_*) = \sum_r \nu_{r,a}^+ w_{+r,a}(\mathbf{c}_*), \quad (1.4)$$

for every species a . Here, the index r span over all the reactions in which species a is involved, $w_{-r,a}$ (*resp.* $w_{+r,a}$) is the rate at which species a is produced (*resp.* consumed) through reaction r , and \mathbf{c}_* is the stationary concentration vector of all the species in the system on which depend the rates. However, at equilibrium, due to microscopic reversibility, the detailed balance holds, and thus every individual reaction of the network is at thermodynamic equilibrium, *i.e.*

$$w_{-r,a}(\mathbf{c}_{\text{eq}}) = w_{+r,a}(\mathbf{c}_{\text{eq}}), \quad (1.5)$$

for every species a and every reaction r , where \mathbf{c}_{eq} is the equilibrium concentration vector. Assuming mass-action law, this implies compatibility conditions on rate constants

$$\frac{k_{-r}}{k_{+r}} = \frac{\prod_{i \in \{\text{reactants}\}} [X_i]_{\text{eq}}^{\nu_{r,i}^+}}{\prod_{i \in \{\text{products}\}} [X_i]_{\text{eq}}^{\nu_{r,i}^-}}, \quad (1.6)$$

where k_{+r} (*resp.* k_{-r}) is the forward (*resp.* backward) constant rate of reaction r , and $[X_i]_{\text{eq}}$ the equilibrium concentration of the species X_i . The same condition holds for the mirrored reaction involving enantiomers \overline{X}_i . Thus

$$\frac{\prod_{i \in \{\text{reactants}\}} [X_i]_{\text{eq}}^{\nu_r^i}}{\prod_{i \in \{\text{products}\}} [X_i]_{\text{eq}}^{\nu_r^i}} = \frac{\prod_{i \in \{\text{reactants}\}} [\overline{X}_i]_{\text{eq}}^{\nu_r^i}}{\prod_{i \in \{\text{products}\}} [\overline{X}_i]_{\text{eq}}^{\nu_r^i}}, \quad (1.7)$$

for every reaction r . A consequence of relations (1.7) and conservation laws for a closed system is that the equilibrium state is always racemic (when conditions are symmetric). It appears now that closed systems are not particularly interesting in the origins of homochirality research, and it explains why we will be interested in dissipative systems that are driven out of equilibrium in this thesis.

1.1.5 Some proposed theories on homochirality

There are multiple scenarios which seek to explain the emergence of homochirality in the literature. Some of them involve deterministic processes, while other theories are based on random ones. Some scenarios occur in a prebiotic environment, or during the apparition of life or even after the advent of life. Some seek to explain how a first imbalance between the enantiomeric populations could have appeared: for instance, due to the interaction with a mineral surface [Ernst, 2013], due to a difference in thermodynamic properties between two enantiomers caused by the parity violation of the weak force [Quack, 2002], or ionization by cosmic rays [Globus and Blandford, 2020, Globus et al., 2021]. However enantiomeric excesses produced via those physical mechanisms are in general low and cannot explain the ones observed in meteorites: an outer space explanation for an initial enantiomeric excess has also been popular, after the fall of the Murchison meteorite in Australia in 1969 in which amino acids as well as purine and pyrimidine nucleobases were observed. First measurements showed an excess of L-Alanine [Engel and Nagy, 1982], but a post-arrival contamination of the meteorite was considered highly probable. This changed when measurements of an excess of L-Isovaline [Cronin and Pizzarello, 1997, Pizzarello and Cronin, 2000], an isomer of L-Valine, but which is not one of the 22 proteinogenic amino acids, reinforced the idea of an initial excess being brought from outer space. One possible explanation of such enantiomeric enrichment among the amino acids observed on meteorites such as the Murchison one is the asymmetric synthesis of amino acids due to an exposure to circularly polarized ultraviolet light [de Marcellus et al., 2011, Meinert et al., 2011].

The chiral transfer between molecular groups has also been studied, and it could be shown for example that the synthesis of tetroses and pentoses is asymmetric when catalyzed by amino acids exhibiting an enantiomeric excess [Pizzarello and Weber, 2004, Weber and Pizzarello, 2006]. The influence of alanine and isovaline (found in significant amount on meteorites) were specifically studied, and the same researchers could also show significant excesses in the formation of D-ribose from LL-dipeptide catalysis. A chiral transfer could also be imagined due to higher dimensional chiral structures that

would dictate the chirality of their constituting elements, as suggested recently through DFT computations of helices stability [Liu, 2020, Li et al., 2021]

Other approaches, focused on the amplification of the enantiomeric excess showed that a wide class of models of out of equilibrium reaction networks exhibit a transition to homochirality thanks to a chiral symmetry breaking [Frank, 1953, Kondepudi and Nelson, 1983, Plasson et al., 2004, Plasson et al., 2007] where, for instance, an autocatalytic reaction scheme with stochasticity can converge towards a non-deterministic homochiral fixed point without the need for nonlinear interactions such as chiral inhibition reactions [Jafarpour et al., 2015].

Among theories based on amplification processes, step-growth and chain-growth polymerization scenarios were extensively studied [Brandenburg and Multamäki, 2004, Brandenburg et al., 2005a, Brandenburg et al., 2005b, Blanco and Hochberg, 2011, Saito and Hyuga, 2005, Gleiser and Walker, 2008]. They showed that homochirality could emerge in the presence of chiral inhibition during polymerization [Sandars, 2003] or in the presence of emerging homochiral ribozymes that catalyze the production of monomers of the same handedness [Wu et al., 2012] at the polymer level under given conditions. However, some central hypothesis used in these works, *e.g.* enantiomeric cross-inhibition in chain-growth polymerization, are not generally compatible with common biopolymers such as RNA, and the suggested convergence toward homochirality most often relies on enzymes or ribozymes, which suppose either complex chemistry or long biopolymers to allow for functional properties. Recently, Tupper *et al.* [Tupper et al., 2017] and Chen *et al.* [Chen and Ma, 2020] explored template-directed polymerization mechanisms through coarse-grained numerical simulations and showed promising results regarding its capability to drive an RNA system to full homochirality during polymerization. The main advantage of template-directed synthesis over chain-growth polymerization is its natural autocatalytic behavior, as RNA strands act as model for their own replication, as well as its enantiomeric cross-inhibition process which has been experimentally observed. Those two aspects of template-directed ligation were already considered as key parameters in one of the first physical model seeking to explain the emergence of homochirality, namely the Frank model.

1.1.6 Original Frank model

In 1953, Frank proposed a simple model to solve the homochirality problem [Frank, 1953]. It is composed of two autocatalytic and enantioselective reactions involving both enantiomers of a chiral species plus an achiral species acting as a fuel and an inhibition reaction

that suppressed chiral species. The mechanism can be described by the three following reactions



with D and L the enantiomers of the chiral species and A the achiral species. Autocatalytic reactions given by equations (1.8) and (1.9) occur with a rate constant k_a and the inhibition reaction (1.10) with a rate k_i . The dynamics of such a system is simply described by the set of differential equations

$$\frac{dD}{dt} = k_a AD - k_i DL, \quad (1.11)$$

$$\frac{dL}{dt} = k_a AL - k_i DL, \quad (1.12)$$

$$\frac{dA}{dt} = -k_a A(D + L) + 2k_i DL, \quad (1.13)$$

where A , D and L denote the concentrations of achiral and chiral species. Considering an oriented enantiomeric excess

$$w = \frac{D - L}{D + L}, \quad (1.14)$$

its time evolution is given by

$$\frac{dw}{dt} = k_i w \frac{1 - w^2}{2} (D + L). \quad (1.15)$$

It thus appears that the system has three fixed points: a racemic one ($w = 0$), which is unstable, and two stable homochiral ones ($w = \pm 1$). This model, yet well-known, can be seen as too simple. In the first place, the autocatalytic enantioselective reactions used in the reaction scheme are, in fact, not common in nature. It took chemists forty years to find an experimental system to illustrate it. The first demonstration was done by Soai, who could observe experimentally such a reaction scheme [Soai et al., 1995] but with chemical compounds which do not have relevance for prebiotic chemistry [Blackmond, 2004]. It is also not clear how to implement the inhibition reaction (1.10) experimentally. We also argue that the scheme is too simple because it involves only two species (three if we count the two enantiomers as “different species”). The prebiotic system in which homochirality probably emerged is more likely to have been messy and of a complex composition, with many different achiral and chiral species rather than two well-defined species. Those

limitations of the Frank model are the starting point of our work.

1.2 Outline

The thesis is mainly divided into two research chapters in which we present two distinct scenarios that could explain the emergence of homochirality. Firstly, we study in Chapter 2 how complex chiral chemical networks can converge to homochirality when the number of chiral compounds they contain becomes large. We consider the case of an open chemical reactor driven out of equilibrium by an inflow and an outflow of matter which exhibits basic chiral properties. We base our study on random matrix properties which are used to determine the stability of the racemic state of such a system and we show that they are susceptible to undergo a transition from a racemic state to a homochiral state due to a symmetry breaking when the number of chiral compounds they contain becomes large. We then illustrate our model with a specific generalization of the Frank model described previously in the introduction chapter, and show it agrees well with the general prediction. An additional database analysis evaluating the threshold above which chiral species become numerous compared to achiral ones is carried out to contextualize our model. Then, in Chapter 3 we develop another possible scenario based on template-directed ligation of biopolymers such as RNA. After highlighting the difficulties for step-growth polymerization to generate any significant enantiomeric excess in a polymeric system, we investigate the propensity of racemic pools of nucleotides to converge towards homochirality in the presence of racemization reactions and temperature cycles through templated polymerization. Closed system and open reactor with chemostated species and constant degradation are both studied. We show that homochirality can be achieved in closed reactors provided racemization reactions occur in the system, and provided the chiral stalling is sufficiently strong. We also show that open reactors exhibit transient significant enantiomeric excesses in an initially racemic system, and that template-directed ligation can also strongly amplify a bias already present in the chemostat, even when it is weak, without racemization reactions. Finally, a conclusion is given in Chapter 4.

A robust transition to homochirality in large and complex chemical networks

Contents

2.1	Study of the abundance of chiral species in nature	24
2.1.1	Previous results	25
2.1.2	Analysis of the PubChem database	28
2.2	A general model for large reaction networks	31
2.2.1	The continuous stirred tank reactor dynamics	32
2.2.2	About mirror symmetry	33
2.2.3	The stability of the racemic state	35
2.2.4	Random matrices and the stability of dynamical systems	37
2.2.5	A criterion for the racemic state stability	39
2.2.6	Sparse chemical networks	41
2.2.7	What about homochiral fixed points?	43
2.3	Example: A generalization of the Frank model	44
2.3.1	The generalized model	44
2.3.2	Stability of the steady-states	48
2.3.3	How does the sparsity of the network affect the results?	56

2.3.4	Correlations in the Jacobian matrix	57
2.4	Discussion	58
2.A	Additional data from the PubChem analysis	60
2.B	Eigenvalues of a random matrix with non zero mean	61
2.B.1	Isolated eigenvalue of a non zero mean random matrix	61
2.B.2	Perturbation of other eigenvalues	62
2.C	Elements of the Jacobian in the homochiral state	63
2.D	Calculation of the maximal eigenvalue of the Jacobian matrix	65
2.E	Compartmentalized systems with diffusive coupling	68

In this section, we study the propensity of large chiral chemical systems driven out of equilibrium to spontaneously transition from a racemic state towards a homochiral one. We first lead an analysis of chemical databases to study the distribution of chiral compounds as a function of molecule sizes to quantify the threshold above which chiral molecules can be qualified numerous. Then, we study a general out of equilibrium chemical system and show, under some assumption, that it is likely to become homochiral when it contains a large number of chiral species. Finally, we build a generalized modified Frank model with the expected dynamical properties.

2.1 Study of the abundance of chiral species in nature

The idea that the number of chiral species increases with the size of molecules seems trivial to most chemists. Chirality is rare among small molecules, and the number of possible chiral stereoisomers multiplies as the number of atoms in molecules increases, since the bigger the molecule, the greater the possibility of having one, or several asymmetric centers. However, no precise estimation is available in the literature about when, or in other words, for which typical size of a molecule, the number of chiral compounds overcomes the number of achiral ones in nature. In this section, we study first the crossover (see Figure 2.1) between a world of small molecules dominated by achiral species and a world dominated by chiral species. In a second step in Section 2.2, we will demonstrate that a large non-equilibrium chemical network is likely to become homochiral by a spontaneous symmetry breaking. In order to answer the first question, we investigate the fractions of achiral and chiral species as functions of the number of heavy atoms (*i.e.* all atoms except hydrogen) in each molecule. Such fractions can be defined in two different fashions, depending on whether we count each pair of enantiomers once or twice. The number of

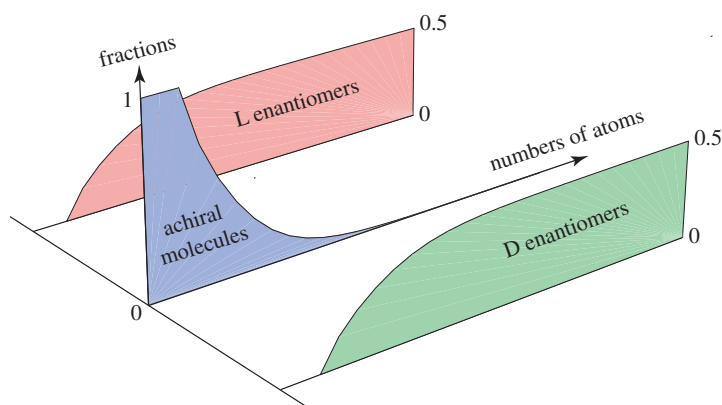


Figure 2.1: A three-dimensional schematic representation of the fractions of possible achiral and chiral molecules as function of their number of atoms. Every chiral molecule appears as mirror-reflected D- and L-enantiomers.

achiral and chiral species with n heavy atoms being respectively denoted A_n and C_n , the fraction of achiral species and the fraction of chiral species defined by counting once each pairs of enantiomers are given by

$$f_n^{(A)} \equiv \frac{A_n}{A_n + C_n} \quad \text{and} \quad f_n^{(C)} \equiv \frac{C_n}{A_n + C_n}, \quad (2.1)$$

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

$$A_{n_1} = C_{n_1}. \quad (2.2)$$

The fractions defined by counting twice the pairs of enantiomers are given by

$$\tilde{f}_n^{(A)} \equiv \frac{A_n}{A_n + 2C_n} \quad \text{and} \quad \tilde{f}_n^{(C)} \equiv \frac{2C_n}{A_n + 2C_n}, \quad (2.3)$$

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

$$A_{n_2} = 2C_{n_2}. \quad (2.4)$$

With these definitions, we should expect that $n_2 < n_1$.

2.1.1 Previous results

Even though no precise estimation of n_1 and n_2 is to be found in the literature, some analysis for specific families of chemical compounds have been carried out in the past, in particular for basic species such as alkanes and monosubstituted alkanes.

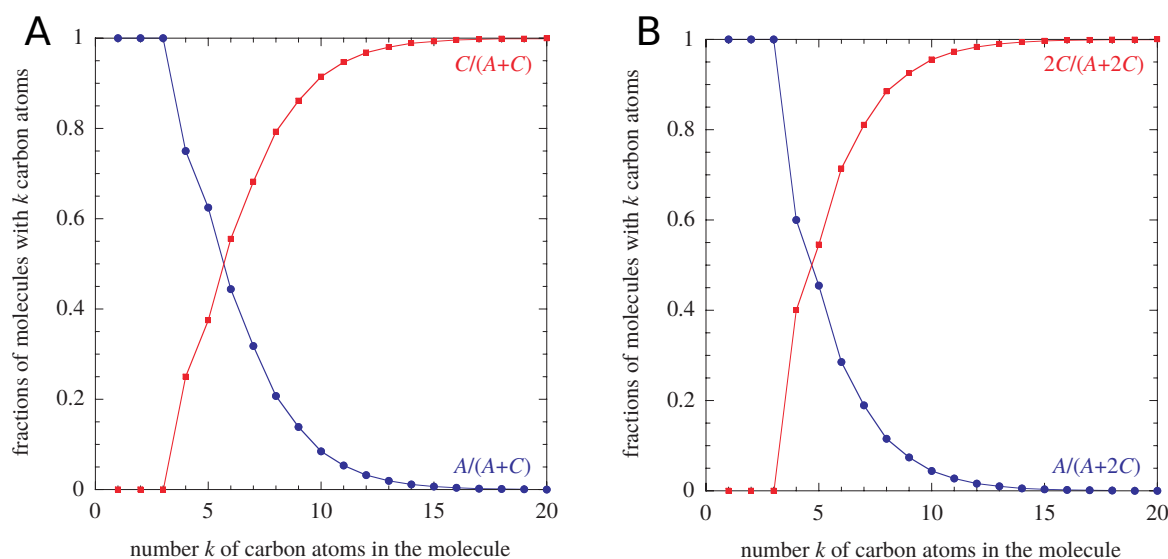


Figure 2.2: Fractions of achiral and chiral stereoisomers of monosubstituted alkanes versus the number k of carbon atoms in the molecule, counting (A) once and (B) twice the pairs of enantiomers. The data are from [Fujita, 2007b].

Monosubstituted alkane stereoisomers

Achiral and chiral monosubstituted alkanes $C_kH_{2k+1}X$ have been enumerated using graph theory and combinatorics [Fujita, 2007b]. Table 1 of [Fujita, 2007b] gives the numbers of achiral A_k and chiral C_k stereoisomers of monosubstituted alkanes versus the number k of carbon atoms they contain (up to $k = 100$). The fractions of achiral and chiral stereoisomers, computed from these data are shown in Figure 2.2 by counting once or twice the enantiomeric pairs. We observe that in this particular case, 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 stated earlier, we have that $k_2 < k_1$. In this case, 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. Fujita also showed that the numbers of chiral C_k and achiral A_k species are growing exponentially with their number k of carbon atoms according to

$$C_k \sim \Lambda^k \quad \text{and} \quad A_k \sim \Lambda^{k/2} \quad \text{with} \quad \Lambda \simeq 3.3 \quad (2.5)$$

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

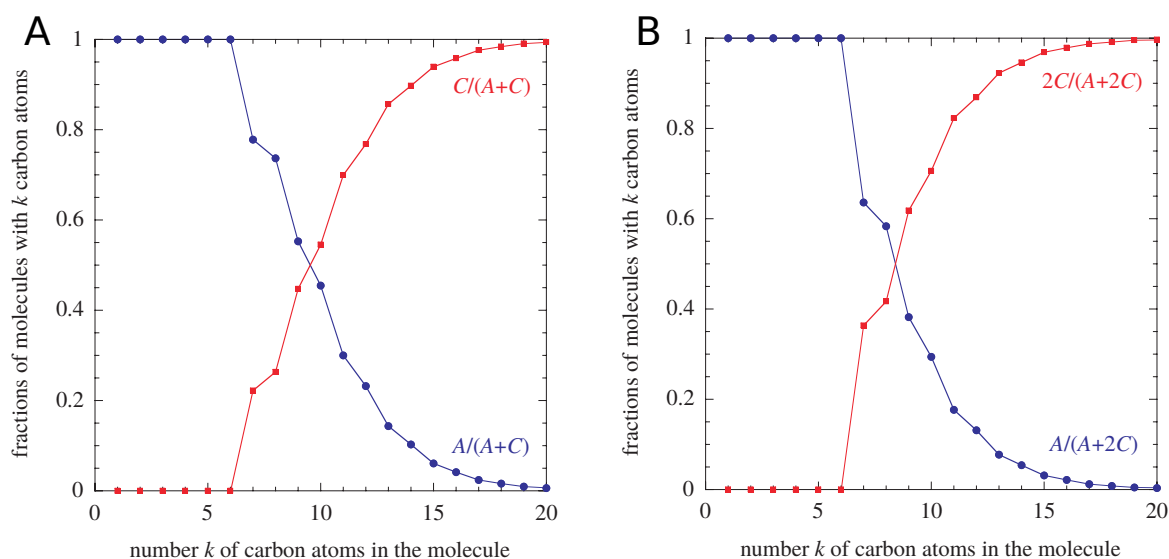


Figure 2.3: Fractions of achiral and chiral stereoisomers of alkanes versus the number k of carbon atoms in the molecule, counting (A) once and (B) twice the pairs of enantiomers. The data are from [Fujita, 2007a].

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 [Fujita, 2007a]. Table 3 of [Fujita, 2007a] gives the numbers of achiral A_k and chiral C_k stereoisomers of alkanes versus the number k of carbon atoms they contain (up to $k = 100$ again). The fractions of achiral and chiral stereoisomers obtained with these data are plotted in Figure 2.3. 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 atoms usually constituting the backbone of the molecule. The crossover happens for somewhat larger molecules than for monosubstituted alkanes because alkanes have molecular structures that are more symmetric than in the presence of one substitution, thus delaying the crossover as the number of atoms in the molecule increases.

Generated chemical universe

Finally, in their search for new drugs for pharmaceutical purposes, Fink and Reymond [Fink and Reymond, 2007], have generated all the possible molecules up to 11 atoms of Carbon, Nitrogen, Oxygen, and Fluor by considering valency, steric effects, chemical stability, and synthetic feasibility rules, and collected them in a database containing 26.4 million molecules and 110.9 million stereoisomers. The Figure 2.4 extracted from [Fink and Rey-

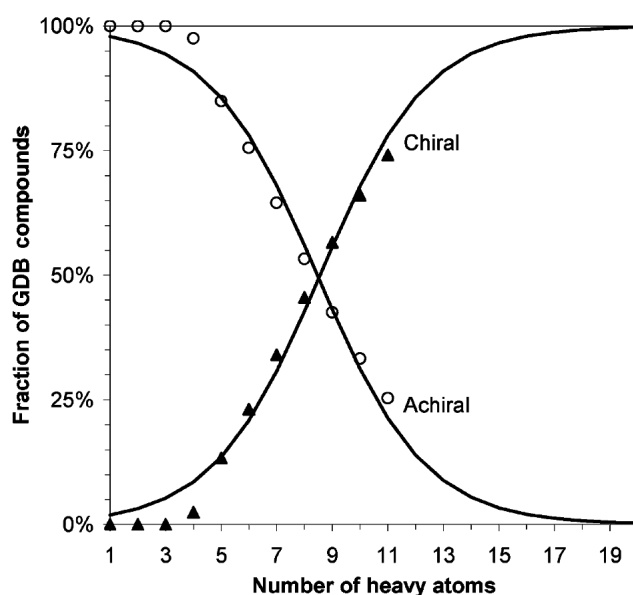


Figure 2.4: Fractions of achiral and chiral compounds as functions of the number of heavy atoms in molecules in the GDB database *i.e.* the database of compounds generated by Fink and Reymond. The figure is extracted from [Fink and Reymond, 2007].

[mond, 2007] 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 exploration of the virtual chemical universe clearly demonstrates the prevalence of chirality for large enough molecules.

2.1.2 Analysis of the PubChem database

From our perspective, it was important to study crossovers in very different databases to reduce statistical biases due to a specific database. One drawback of the GDB database is that the molecules are virtual while the origin of homochirality (or the origin of life) has no reason to be restricted to a well designed chemical system. To carry out a study of real existing molecules in nature, and estimate n_1 and n_2 , we carried out an analysis of the free access PubChem database self-defined as “[...] *the world’s largest collection of freely accessible chemical information.*”. It contains (in late 2019, when the analysis has been carried out) 139 million of chemical species and we restricted the analysis to species containing less than 20 heavy atoms. There are two reasons for this choice, on one hand the statistics becomes more limited for molecules longer than 20 heavy atoms, and on the other hand, there is a discontinuity in the number of achiral and chiral molecules in the PubChem database as shown in Figure 2.12a of Appendix 2.A. After discussion with the curators of the database, there is currently no information available about the

origin of this discontinuity, but it has perhaps to do with the fact that PubChem is itself a patchwork of sub-databases. In any case, we hope to avoid any perturbation from this discontinuity in data by considering only molecules with $n \leq 20$ heavy atoms.

Raw data

From the complete dataset downloaded from PubChem, 91.6 million molecules were analyzed after rejection of compounds with isotopic elements, multiple components (such as complexes) or incomplete data on bond structure. From the 33.6 million molecules with less than 21 heavy atoms, 18.7 million molecules (55.7 %) are chiral, and 1.4 million chiral molecules have no stereocenters (7.4 % of the chiral molecules with less than 21 heavy atoms) thus for these molecules their chirality arise only from their non planar geometry. As shown in Figure 2.12b of Appendix 2.A, the analysis of the 33.6 million 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 $n_{\text{raw}} \simeq 9.4$. A crossover in this region is coherent with an increase in the number of stereoisomers per molecule for molecules of this length.

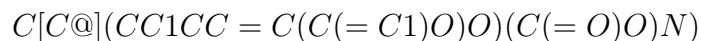
Chirality detection

Chiral species in the PubChem database were detected using the *chiral flag* present in the SDF file which contains the information about the structure of a given molecule in the database. SDF files have a specific format including the MDL Molfile which is a file format in which information about atoms, their relative positions, the different bonds and the information about the molecule's chirality are stored in chemical databases. We did not detect chirality using the sometimes available geometry of the molecule as it would require more complex algorithms, and we assumed the information about the chirality was correct (knowing however the database would contain errors). The entire numerical treatment and analysis depicted in the current Section 2.1.2 have been carried out using Python 3.

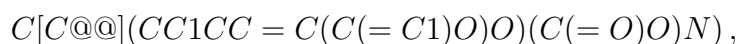
Simple methods of generation of stereoisomers and enantiomers

The chirality analysis was not restricted to the available species present in the database as for some species, only one of the two possible enantiomers or one of the many possible stereoisomers were present. We thus generated numerically all those absent stereoisomers to account for their possible presence in nature. The generation of enantiomers and stereoisomers is based on SMILES (Simplified Molecular Input Line Entry Specification) strings that describe the structure of a molecule in a compact fashion. For instance, the

SMILES string of glyceraldehyde is $C(C(C=O)O)O$. Stereocenters appear in SMILES strings called non-canonical SMILES through the @ or @@ descriptor as in the two enantiomers



and



where @ refers to the S and @@ the R configuration of the chiral center (or the other way around). However, only stereocenters for which the substituents are different at the first order (*i.e.* the first neighbors) are indicated in the SMILES strings from the PubChem databank. Thus it makes it impossible to distinguish two diastereomers or enantiomers in which a stereocenter is bounded to the same element twice, which also is a source of errors. Note that contrarily to non-canonical SMILES, canonical SMILES do not carry the chirality marker.

For the generation of enantiomers, a list of non-canonical SMILES strings was built, and for each chiral molecule with an available SMILES string, we generated its mirrored version, and then searched it in the list. If it was not found, then it was added to the list in the expanded database. However, for chiral centers that do not explicitly appear in SMILES strings, 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, we looked whether the database contained the theoretical maximum number of stereoisomers for a given species, which can be evaluated from the number of stereocenters. If all the stereoisomers are not present, which is frequent for molecules containing several stereocenters, the set of chemical species from the database is expanded so that each chiral species has the maximum possible number of stereoisomers (*i.e.*, 2^{n_s} with n_s 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. Once the stereoisomers have been generated using the procedure described previously, 47 million stereoisomers were added to complete the PubChem dataset to the 33 million present before the expansion of the database. In this case, the intersection occurs (see Figure 2.5) at $n_2 \simeq 6.4$ if both enantiomers are considered and $n_1 \simeq 8.1$ if only one enantiomer is considered. Plots for the raw data (Figure 2.12) and the completed database with generated enantiomers (Figure 2.13) are shown in Appendix 2.A. Other databases were analyzed as

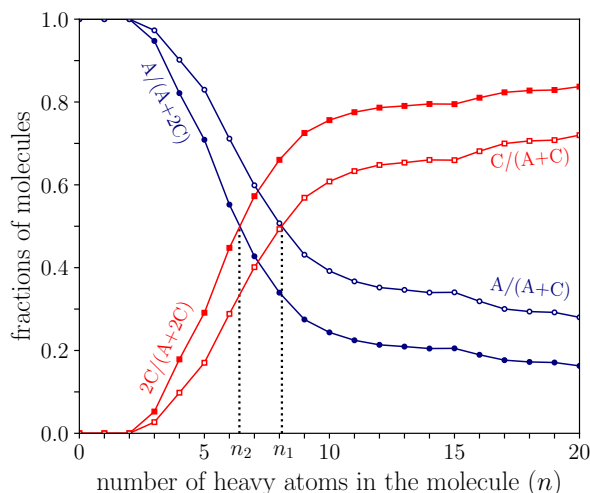


Figure 2.5: A two-dimensional representation of the fractions of achiral (circles) and chiral (squares) molecules in an expanded PubChem database containing all the stereoisomers of molecules with at most $n \leq 20$ heavy atoms. Filled symbols correspond to counting stereoisomers twice, while empty symbols correspond to counting them once in the fractions. Error bars are smaller than symbol sizes and the crossover occurs at $n_2 \simeq 6.4$ and $n_1 \simeq 8.1$ depending on which counting is considered.

well, namely DrugBank and ChemSpider, however they contained much less compounds leading to poor statistics. Also databases such as DrugBank, focused on drug data might be biased toward chiral compounds, which is another reason not to use their data. All the data from the previous work carried out by Fujita, Fink and Reymond alongside our analysis of the PubChem database are gathered in Table 2.1. Remarkably, irrespective of the precise procedure to generate and analyze the database and regardless of the precise composition of the molecules, the crossover between the achiral and chiral worlds occurs for a number of heavy atoms of the order of about 9. This indicates that there is no need to have a hundred or a thousand atoms in the molecule for chiral molecules to become dominant. The main consequence of this crossover 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 and gets bigger than $n \simeq 9$. This emerging bimodality is potentially susceptible to induce a chiral symmetry breaking in a system driven out of equilibrium as we show in the next section.

2.2 A general model for large reaction networks

In this section, we develop a simple description of a reaction network involving a large number of chiral species. As stated in the introduction of this chapter, we will show that

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

Table 2.1: Positions of the crossover 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. * The crossover for PubChem raw data occurs between n_1 and n_2 because not all enantiomers of a given species are present in the database.

large out of equilibrium chiral molecular systems are likely to undergo a transition from a racemic state to a homochiral one as the number of chiral species contained in it becomes large.

2.2.1 The continuous stirred tank reactor dynamics

We first consider our chemical system to be open, and ruled by the Continuous Stirred Tank Reactor (CSTR) dynamics. The system is thus assumed to be homogeneous because well mixed, and driven out of equilibrium by an influx of species. The dynamics of the system is thus described by the following CSTR equation

$$\frac{d\mathbf{c}}{dt} = \mathbf{F}(\mathbf{c}) + \frac{1}{\tau}(\mathbf{c}_0 - \mathbf{c}), \quad (2.6)$$

where \mathbf{c} is the concentration vector for all species in the system, \mathbf{c}_0 is the concentration vector for species supplied into the system by the environment at a rate $1/\tau$, τ is the residence time of species inside the reactor and $\mathbf{F}(\mathbf{c})$ is a function describing the gains and losses of chemical compounds through the reactions occurring in the system. It reads

$$\mathbf{F}(\mathbf{c}) = \boldsymbol{\nu} \cdot \mathbf{w}(\mathbf{c}), \quad (2.7)$$

where $\boldsymbol{\nu}$ is the matrix of stoichiometric coefficients, and $\mathbf{w}(\mathbf{c})$ is the vector of net reaction rates, which do not need to obey mass-action law, for all the possible reactions in the system. For a given reaction r ,

$$w_r(\mathbf{c}) = w_{+r}(\mathbf{c}) - w_{-r}(\mathbf{c}). \quad (2.8)$$

The trajectories of the dissipative dynamical system described by equation (2.6) are in general converging towards an attractor, which may be a steady state, a limit cycle corresponding to periodic oscillations, or a strange attractor sustaining chaotic oscillations. In the present context, the attractors may be assumed to be stationary and they may undergo bifurcations leading to multi-stability. In this kinetics framework, we consider that the concentration vector associated to a system containing chiral and achiral species reads

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}_D \\ \mathbf{c}_A \\ \mathbf{c}_L \end{pmatrix} \quad (2.9)$$

where \mathbf{c}_D (resp. \mathbf{c}_L) denotes the concentration vector for D-enantiomers (resp. L-enantiomers), and \mathbf{c}_A the concentration vector of achiral species. The number of chiral species is denoted N_C and the number of achiral species, N_A , for a total number of species in the system being $N_S = N_A + 2N_C$. The particular case of the closed system is reached in the limit $\tau \rightarrow \infty$.

2.2.2 About mirror symmetry

One particular property of chiral systems is that they admit a mirror symmetry: all the reactions undergone by the i -th D-enantiomer also occurs for the i -th L-enantiomer with the same rate constant, but for mirrored species. The mirror symmetry of the system corresponds to the following exchange of concentrations of D- and L-enantiomers,

$$\mathbf{S}_c \cdot \mathbf{c} = \begin{pmatrix} 0 & 0 & \mathbf{I} \\ 0 & \mathbf{I} & 0 \\ \mathbf{I} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{c}_D \\ \mathbf{c}_A \\ \mathbf{c}_L \end{pmatrix} = \begin{pmatrix} \mathbf{c}_L \\ \mathbf{c}_A \\ \mathbf{c}_D \end{pmatrix}, \quad (2.10)$$

written in terms of the $N_S \times N_S$ matrix such that $\mathbf{S}_c^2 = \mathbf{I}$, 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

$$\mathbf{w}(\mathbf{S}_c \cdot \mathbf{c}) = \mathbf{S}_w \cdot \mathbf{w}(\mathbf{c}) \quad (2.11)$$

with some $N_R \times N_R$ matrix satisfying $\mathbf{S}_w^2 = \mathbf{I}$, where N_R is the number of reactions in the system. As a consequence of the mirror symmetry, the $N_S \times N_R$ matrix of stoichiometric coefficients obeys the following symmetry relation

$$\mathbf{S}_c \cdot \boldsymbol{\nu} \cdot \mathbf{S}_w = \boldsymbol{\nu}. \quad (2.12)$$

The kinetic equation (2.6) can be rewritten as three different equations

$$\frac{d\mathbf{c}_D}{dt} = \mathbf{F}_D(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) + \frac{1}{\tau}(\mathbf{c}_{D0} - \mathbf{c}_D), \quad (2.13)$$

$$\frac{d\mathbf{c}_A}{dt} = \mathbf{F}_A(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) + \frac{1}{\tau}(\mathbf{c}_{A0} - \mathbf{c}_A), \quad (2.14)$$

$$\frac{d\mathbf{c}_L}{dt} = \mathbf{F}_L(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) + \frac{1}{\tau}(\mathbf{c}_{L0} - \mathbf{c}_L). \quad (2.15)$$

In this case, \mathbf{F}_D , \mathbf{F}_A and \mathbf{F}_L are expressed as

$$\mathbf{F}_D(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) = \boldsymbol{\nu}_D \cdot \mathbf{w}(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) \quad (2.16)$$

$$\mathbf{F}_A(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) = \boldsymbol{\nu}_A \cdot \mathbf{w}(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) \quad (2.17)$$

$$\mathbf{F}_L(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) = \boldsymbol{\nu}_L \cdot \mathbf{w}(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_L) \quad (2.18)$$

with $\boldsymbol{\nu}_D$ (*resp.* $\boldsymbol{\nu}_L$) being the $N_C \times N_R$ submatrix relatives to D-enantiomers (*resp.* L-enantiomers), and $\boldsymbol{\nu}_A$ the $N_A \times N_R$ submatrix relatives to achiral species. Within this formalism, the mirror symmetry is expressed as

$$\mathbf{F}_D(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{F}_L(\mathbf{z}, \mathbf{y}, \mathbf{x}) \quad \text{and} \quad \mathbf{F}_A(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{F}_A(\mathbf{z}, \mathbf{y}, \mathbf{x}). \quad (2.19)$$

We note that the concentration vector \mathbf{c}_0 , which is controlling the supply of species in the reactor, may or may not satisfy the symmetry condition $\mathbf{S}_c \cdot \mathbf{c}_0 = \mathbf{c}_0$. The chiral symmetry breaking may happen spontaneously in the former case and explicitly by the external control in the latter case. If the kinetic equation (2.6) admits a stationary solution \mathbf{c}_s , it will be either a racemic mixture if $\mathbf{S}_c \cdot \mathbf{c}_s = \mathbf{c}_s$, or a steady state with some non-vanishing enantiomeric excess if $\mathbf{S}_c \cdot \mathbf{c}_s \neq \mathbf{c}_s$. 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 the case of a racemic steady-state $\mathbf{S}_c \cdot \mathbf{c}_s = \mathbf{c}_s$, the racemic mixture characterized by equal concentrations of D- and L-enantiomers,

$$\mathbf{c}_D = \mathbf{c}_L, \quad (2.20)$$

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

$$\frac{d\mathbf{c}_A}{dt} = \mathbf{F}_A(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_D) + \frac{1}{\tau}(\mathbf{c}_{A0} - \mathbf{c}_A), \quad (2.21)$$

$$\frac{d\mathbf{c}_D}{dt} = \mathbf{F}_D(\mathbf{c}_D, \mathbf{c}_A, \mathbf{c}_D) + \frac{1}{\tau}(\mathbf{c}_{D0} - \mathbf{c}_D), \quad (2.22)$$

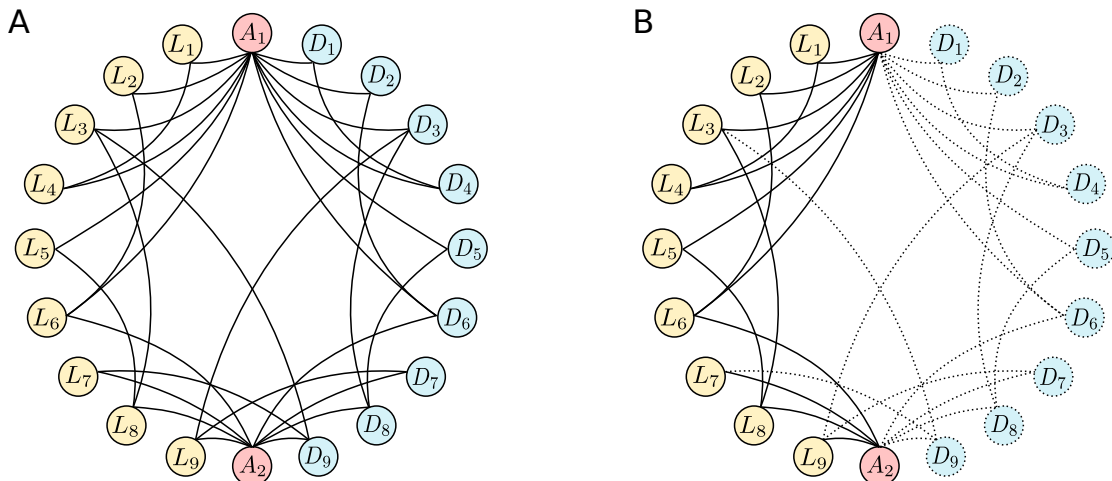


Figure 2.6: Example of a reaction network containing two (resp. nine) achiral (resp. chiral) species. A_i represents achiral species and D_i (resp. L_i) represents the D-enantiomer (resp. L-enantiomer) of the i^{th} chiral species. Solid lines connect random species involved in the same reaction. (A) Illustration of the mirror symmetry, according to which each chiral species linked by a random reaction is involved in the same - but mirrored - reaction involving mirrored species (enantiomers) with identical rate constants. As a result, the reaction network admits an axis of symmetry going through achiral species, here A_1 and A_2 . (B) Illustration of the broken symmetry of a homochiral state. Dotted circles represent species that are no longer present in the system (or at extremely low concentrations) and dotted lines represent the reactions in which they were participating. Here, the simple case of a system undergoing a global symmetry breaking for all chiral species is depicted.

is stable in the racemic subspace described by equation (2.20). The mirror symmetry is depicted in Figure 2.6 which shows that the mirror symmetry is present in the initial racemic state of the reaction network (Figure 2.6a), while the symmetry gets broken in the homochiral state (Figure 2.6b).

2.2.3 The stability of the racemic state

Assuming the system possesses at least one racemic fixed point, we introduce the following parameter, which describes the deviations from the racemate

$$\mathbf{x} = \frac{1}{2}(\mathbf{c}_D - \mathbf{c}_L). \quad (2.23)$$

The time evolution of an infinitesimal perturbation from the racemate thus reads

$$\frac{d\delta\mathbf{x}}{dt} = \frac{1}{2} \frac{d}{dt}(\mathbf{c}_D - \mathbf{c}_L). \quad (2.24)$$

Denoting \mathbf{c}_* the concentration vector of the racemic fixed point such that $\mathbf{c}_D = \mathbf{c}_* - \delta\mathbf{x}$ and $\mathbf{c}_L = \mathbf{c}_* + \delta\mathbf{x}$,

$$\begin{aligned} \frac{d\delta\mathbf{x}}{dt} = \frac{1}{2} \left[\mathbf{F}_L(\mathbf{c}_* - \delta\mathbf{x}, \mathbf{c}_A, \mathbf{c}_* + \delta\mathbf{x}) + \frac{1}{\tau} (\mathbf{c}_0 + \delta\mathbf{x}_0 - \mathbf{c} - \delta\mathbf{x}) \right. \\ \left. - \mathbf{F}_D(\mathbf{c}_* - \delta\mathbf{x}, \mathbf{c}_A, \mathbf{c}_* + \delta\mathbf{x}) - \frac{1}{\tau} (\mathbf{c}_0 - \delta\mathbf{x}_0 - \mathbf{c} + \delta\mathbf{x}) \right], \end{aligned} \quad (2.25)$$

and a Taylor expansion at first order about the racemic fixed point gives

$$\begin{aligned} \frac{d\delta\mathbf{x}}{dt} = \frac{1}{2} \left[\mathbf{F}_L(\mathbf{c}_*, \mathbf{c}_A, \mathbf{c}_*) + \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} \cdot \delta\mathbf{x} - \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*} \cdot \delta\mathbf{x} \right. \\ \left. - \mathbf{F}_D(\mathbf{c}_*, \mathbf{c}_A, \mathbf{c}_*) - \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} \cdot \delta\mathbf{x} + \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*} \cdot \delta\mathbf{x} \right] \\ + \frac{1}{\tau} (\delta\mathbf{x}_0 - \delta\mathbf{x}). \end{aligned} \quad (2.26)$$

where $|_{\mathbf{c}_*}$ means evaluated at \mathbf{c}_* . From the mirror symmetry relations equation (2.19), it is straightforward to show that the equality

$$\mathbf{F}_L(\mathbf{c}_*, \mathbf{c}_A, \mathbf{c}_*) = \mathbf{F}_D(\mathbf{c}_*, \mathbf{c}_A, \mathbf{c}_*) \quad (2.27)$$

holds. The relations

$$\frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*} = \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} \quad (2.28)$$

and

$$\frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} = \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*} \quad (2.29)$$

also arise from the mirror symmetry. Using relations given by equations (2.27)-(2.29), we obtain

$$\frac{d\delta\mathbf{x}}{dt} = (\mathbf{J}_{DD} - \mathbf{J}_{DL}) \cdot \delta\mathbf{x} + \frac{1}{\tau} (\delta\mathbf{x}_0 - \delta\mathbf{x}), \quad (2.30)$$

where

$$\mathbf{J}_{DD} \equiv \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*} = \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} \quad (2.31)$$

and

$$\mathbf{J}_{DL} \equiv \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_L} \Big|_{\mathbf{c}_*} = \frac{\partial \mathbf{F}_L}{\partial \mathbf{c}_D} \Big|_{\mathbf{c}_*}. \quad (2.32)$$

Without loss of generality, we assume that $\delta\mathbf{x}_0 = 0$ meaning that we have a racemic influx of matter in the reactor since we do not want to break the symmetry explicitly.

Equation (2.30) thus reduces to

$$\frac{d\delta\mathbf{x}}{dt} = \left(\mathbf{J} - \frac{1}{\tau}\mathbf{I} \right) \cdot \delta\mathbf{x}, \quad (2.33)$$

where $\mathbf{J} = \mathbf{J}_{\text{DD}} - \mathbf{J}_{\text{DL}}$ and \mathbf{I} is the identity matrix. Here, \mathbf{J} is an effective Jacobian matrix evaluated at the racemic fixed point \mathbf{c}_* . According to stability analysis theory and equation (2.33), the racemic state is stable when all the eigenvalues of the matrix $\mathbf{J} - \frac{1}{\tau}\mathbf{I}$ (which in general are complex numbers) have a negative real part [Strogatz, 1994]. Reciprocally, as soon as one of those eigenvalues gets a positive real part, the racemic state becomes unstable to any perturbation, meaning that any infinitesimal deviation from the racemic state will be amplified exponentially fast.

2.2.4 Random matrices and the stability of dynamical systems

A random matrix (RM) is simply defined as an $n \times m$ matrix, with $n, m \in \mathbb{N}^*$, whose entries are random variables. They first appeared in the literature in 1928 [Wishart, 1928] as null models for covariance matrices in statistics, and soon their use spread to many different research fields [Potters and Bouchaud, 2020]. The main interest in those matrices is to solve eigenvalues problems *i.e.* to determine their spectrum depending on conditions on the entries of the RM. Some of the most famous results are the Wigner semicircle law [Wigner, 1955, Wigner, 1958] which holds for entries distribution in the Gaussian Orthogonal Ensemble (GOE), in the Gaussian Unitary Ensemble (GUE) or for Wigner RM, and the Marchenko-Pastur distribution [Marchenko and Pastur, 1967] for Wishart RM. In 1965, Ginibre considered real, complex and quaternionic valued matrices and showed, *inter alia*, that for an $n \times n$ complex RM, \mathbf{M} , with no orthogonal or unitary constraints but still having Gaussian distributed entries with zero mean [Ginibre, 1965], its eigenvalues were uniformly distributed in the complex plane in a disk of radius

$$R = \sigma\sqrt{n}, \quad (2.34)$$

centered around the origin of the complex plane, where σ is the standard deviation of the matrix \mathbf{M} elements, in the limit $n \rightarrow \infty$. Twenty years later, Girko got free of the Gaussian distribution constraint, and generalized the same result [Girko, 1985] to the spectrum of an $n \times n$ random matrix whose entries are independent and identically distributed (i.i.d.) according to an arbitrary distribution of variance σ^2 and zero mean.

In 1970, Gardner and Ashby got interested in the stability of connected dynamical graphs. Those graphs could represent many different systems such as the traffic in an

airport, slum areas with a high number of inhabitants or neurons in interaction, as they stated in their work [Gardner and Ashby, 1970]. They proposed a simple linearly interacting model described by the dynamical equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{A} \cdot \mathbf{x}, \quad (2.35)$$

where \mathbf{x} is the vector of n dynamical variables (such as population vector for ecological models) and \mathbf{A} an $n \times n$ interaction matrix whose element a_{ij} describes how the variable x_j interacts with and affects the variable x_i , n being the number of dynamical variables in the system. To avoid relying on a specific structure of the interaction matrix, they assumed all entries of \mathbf{A} were distributed evenly in the interval $[-1, 1]$ at random and diagonal elements (self-interaction terms) would be all negatives and evenly distributed in the interval $[-1, -0.1]$ to ensure stability for every variable taken independently. They observed with computer simulations that the probability of stability of the system (equivalent to the probability to have a maximal eigenvalue of \mathbf{A} whose real part is negative) decreased with the increase of the number of variables in the system n , and with the increase of connectance, C , of the system, which characterizes the probability for an \mathbf{A} element a_{ij} to be non zero (*i.e.* “zero” meaning species j does not affect species i). Two years later, May, interested in the stability of large ecosystems, looked at this problem with an analytical approach [May, 1972], extending it to non-linear systems with equation (2.35) representing a linearization about a steady-state. \mathbf{A} elements were also drawn randomly from a statistical distribution with zero mean and standard deviation σ (equivalent here to an average interaction strength) with an additional -1 terms for diagonal elements to ensure individual stability for species. With this choice, the matrix \mathbf{A} could be written as

$$\mathbf{A} = \mathbf{B} - \mathbf{I}, \quad (2.36)$$

where \mathbf{B} is a random matrix and \mathbf{I} the identity matrix. Using results from the Wigner semicircle law, although similar arguments would work for more general RM using the Girko circle law [Girko, 1985], May showed that such systems are likely to be stable if

$$\sigma < \frac{1}{\sqrt{n}}, \quad (2.37)$$

and almost certainly unstable if

$$\sigma > \frac{1}{\sqrt{n}}, \quad (2.38)$$

in the limit $n \gg 1$. For sparse RM, where some entries of \mathbf{A} matrix were equal to

zero, meaning some species would not interact, the result would change with $n \rightarrow nC$ in equations (2.37) and (2.38), where C is the connectance, *i.e.* the percentage of non-zero elements in \mathbf{A} . Those results were in agreements with ecological studies which suggested that large ecosystems are not stable, and the larger are ecosystems, the weaker the interactions strength has to be for stability to remain. Those two papers had a large impact in dynamical studies as they show that random matrix theory had produced enough results on spectral properties of RM to now be applicable to a wide range of large and complex dynamical systems even though May's approach first required some time to be recognized by the scientific community.

2.2.5 A criterion for the racemic state stability

Going back to our model for chemical networks, the elements of the Jacobian matrix \mathbf{J} consist of products of rate constants, stoichiometric coefficients, and concentrations. We will assume, for solving the stability of our system, that \mathbf{J} can be considered as a random matrix. Here the randomness of the elements of the Jacobian matrix comes primarily from the randomness of the rate constants and to a lesser extent from that of concentrations. For instance, in the specific case of enzymatic reactions, it has been shown that rate constants are distributed according to a log-normal distribution [Davidi et al., 2018]. The general observation is that, in known reaction networks such as the Belousov-Zhabotinsky system, the rate constants may take very different values without regularity between them among the different reactions of the network. It is important to note that we do not assume the stoichiometric matrix $\boldsymbol{\nu}$ to be a random matrix. The network has a structure ruled by the mirror symmetry which is required to derive equation (2.33). Once one accepts the idea of treating the Jacobian matrix \mathbf{J} as a random matrix of size $N_C \times N_C$, the simplest possibility is to also assume that the elements of this matrix are independent and identically distributed (i.i.d.) random variables (but not necessarily Gaussian) of mean value μ and variance σ^2 . Gaussian entries are not necessary in principle to deduce spectral properties of the matrix \mathbf{J} since the circle law of Girko holds for an arbitrary distribution of elements in the RM [Girko, 1985] as stated in the previous section. Since the Jacobian matrix \mathbf{J} depends on the concentrations of the racemic steady state, we note that both μ and σ^2 may vary with the mean residence time τ , which controls the steady state of equation (2.6). As explained at the end of Section 2.2.3, we have to compute the spectrum of $\mathbf{J} - \frac{1}{\tau}\mathbf{I}$ to determine the stability of the racemic state. It is first trivial that

$$\mathcal{S}(\mathbf{J} - \frac{1}{\tau}\mathbf{I}) = \mathcal{S}(\mathbf{J}) - \frac{1}{\tau}, \quad (2.39)$$

where $\mathcal{S}(\mathbf{M})$ denotes the spectrum of a matrix \mathbf{M} , since the spectrum of the diagonal matrix $\frac{1}{\tau}\mathbf{I}$ is composed of the eigenvalue $1/\tau$ N_C -fold degenerated. We thus focus now on the matrix \mathbf{J} .

The zero mean case ($\mu = 0$)

When $\mu = 0$, based on Girko's circle law, the complex eigenvalues of \mathbf{J} are uniformly distributed in a disk of radius $\sigma\sqrt{N_C}$ in the limit of large values of N_C . In this scenario, the real part of the largest eigenvalue λ_{\max} scales as the border of the circle, $\Re(\lambda_{\max}) \sim \sigma\sqrt{N_C}$, meaning that, using equation (2.39), the stability of the racemic state is conserved as long as

$$\sigma\sqrt{N_C} - \frac{1}{\tau} < 0, \quad (2.40)$$

i.e. as long as the number of chiral species verifies

$$N_C < \frac{1}{(\sigma\tau)^2}. \quad (2.41)$$

When this inequality is no longer verified, the system is susceptible to converge towards a homochiral steady-state. It is also noticeable that the eigenvector associated to this eigenvalue is a random vector. However, having a zero mean for \mathbf{J} entries is a particular case that has no reason to be verified. In this case, the Girko's circle law gets modified slightly.

The non-zero mean case ($\mu \neq 0$)

When $\mu \neq 0$, it is possible to show that one eigenvalue of the matrix \mathbf{J} gets isolated from the rest of the spectrum (which still verifies Girko's circle law) and is equal to $\lambda_{\text{iso}} = \mu N_C$ and the corresponding eigenvector has uniform components to dominant order (see Appendix 2.B for a demonstration). In the case where λ_{iso} is greater than the maximum eigenvalue from the Girko circle, the stability of the racemic state holds as long as

$$\mu N_C - \frac{1}{\tau} < 0, \quad (2.42)$$

i.e. as long as the number of chiral species verifies

$$N_C < \frac{1}{\mu\tau}. \quad (2.43)$$

As for the previous case where $\mu = 0$, when this inequality does not hold anymore (*e.g.* when N_C grows), the stability of the racemic state is broken and the system is susceptible to converge towards a homochiral state. Two possible mechanisms for the instability of the racemic state then emerge for large N_C . Either (i) it occurs due to an eigenvalue located on the edge of the circle (which may be real or complex valued) otherwise (ii) the instability occurs due to the isolated eigenvalue. It follows from the two criteria in equations (2.41) and (2.43) that when

$$(\sigma/\mu)^2 \geq N_C \geq 1/(\tau\sigma)^2, \quad (2.44)$$

the system becomes unstable by the first (i) mechanism and in this case only a subpart of all the species become unstable with random directions at the transition. Indeed, in the vicinity of the racemic fixed point, $\delta\mathbf{x}$ will be oriented in the same direction as the random eigenvector associated to the leading eigenvalue of the edge of the disk. And when $\mu > 0$ and

$$N_C \geq \max\{1/(\tau\mu), (\sigma/\mu)^2\}, \quad (2.45)$$

the system becomes unstable by the second (ii) mechanism where all species become simultaneously unstable in the same direction, as the eigenvector associated to this eigenvalue is a homogeneous vector. In such cases, random matrix theory predicts that as N_C becomes large, these mechanisms of instability become more and more likely. The two mechanisms are graphically represented in Figure 2.7. Figures 2.7a-c represent the first criterion where the stability is broken by the edge of Girko's circle, and Figs. 2.7d-f describe the second mechanism where the stability of the racemic state is broken by the isolated eigenvalue. Instability probabilities are also depicted in Figure 2.7g. There is however no fundamental reason to assume that the elements of the Jacobian matrix \mathbf{J} of a reaction network should be in general i.i.d. random variables. In fact, we will observe in Section 2.3 that there can be correlations in the Jacobian matrix \mathbf{J} of a generalized Frank model, which arise from the difference between the diagonal and non-diagonal elements of the matrix. In that model, eigenvalues do not fill a Girko circle but fill a domain of different shape, closer to an ellipse.

2.2.6 Sparse chemical networks

Another interesting aspect of the random matrix approach is that it can describe important features of the network such as its sparsity (meaning some elements are zero) as mentioned before. The sparsity of the matrix \mathbf{J} could originate from the sparsity of

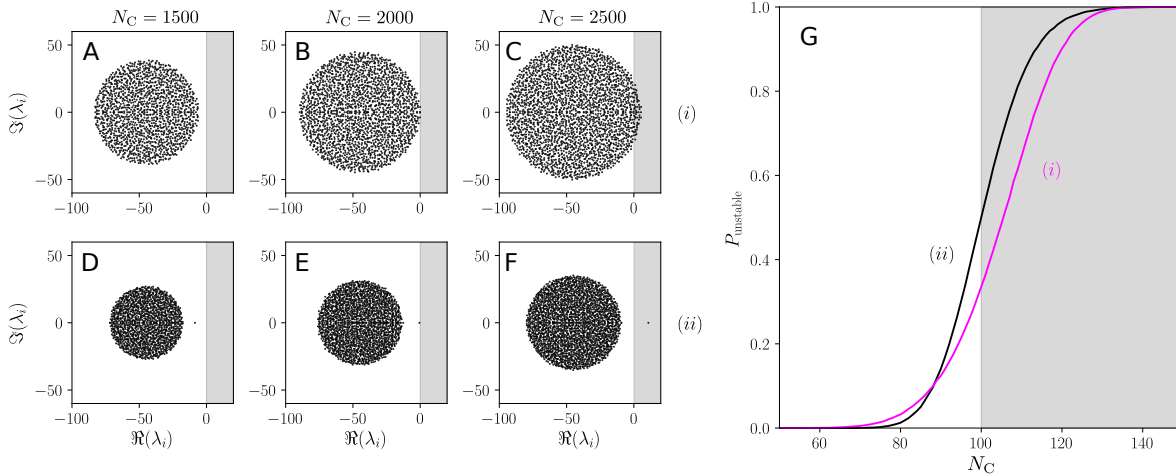


Figure 2.7: Distribution of eigenvalues and corresponding instability criterion. Subfigures (A), (B) and (C) represent the eigenvalues of the random matrix \mathbf{M} with Gaussian entries with $\mu = 0$, $\sigma = 1$ and $\tau = 1/\sqrt{2000}$, where the instability mechanism (i) occurs as N_C increases, *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 (ii) 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 (ii) 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 (i) 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 . The determination of the spectrum of random matrices was performed using linear algebra functions of the Python library Numpy.

the stoichiometric matrix $\boldsymbol{\nu}$, which has already been observed in studies of metabolism of living systems, and which results from the low connectivity of certain species to the rest of the network. A second origin to sparsity lies in the fact that in a typical chemical system, the various rate constants span many orders of magnitude, resulting in certain matrix elements becoming negligible compared to other ones. A third origin to sparsity may be due to the variability of concentrations of species in the network, which could be large (specially if the system is not well mixed). Whatever its origin, the sparsity of the Jacobian matrix \mathbf{J} can be accounted for by the theoretical treatment of the previous section because it only affects the radius of the disk in which the eigenvalues are distributed in. More precisely, this radius which is $\sigma\sqrt{N_C}$ for a non-sparse network becomes $\sigma\sqrt{\alpha N_C}$ in a sparse matrix, where $\alpha \in [0, 1]$ measures the percentage of non-zero elements in \mathbf{J} . Moreover, in the case where $\mu \neq 0$, the isolated eigenvalue is approximately changed to $\mu N_C \rightarrow \alpha\mu N_C$. These two changes imply that the criterion for the system to be unstable by the first mechanism becomes

$$(\sigma/\mu\sqrt{\alpha})^2 \geq N_C \geq 1/(\sigma\tau\sqrt{\alpha})^2, \quad (2.46)$$

and by the second mechanism,

$$N_C \geq \max\{1/(\alpha\tau\mu), (\sigma/\mu\sqrt{\alpha})^2\}. \quad (2.47)$$

In other words, the sparser the matrix \mathbf{J} , the higher the number of chiral species needed for the chiral symmetry breaking to occur in a system with given μ , σ , and τ . In the end, if the number of chiral species is sufficiently high, the results remain unchanged. An interesting consequence of that observation is that the sparsity coefficient of \mathbf{J} is a tuning parameter to control the homochiral transition at a fixed number of chiral species N_C (provided that N_C is in a range that allow the system to be unstable) similarly to what May proved for the stability of ecological systems [May, 1972].

2.2.7 What about homochiral fixed points?

One of the first critic one could make about the previous derivation is that it is very general and it would also apply to any fixed points of the system, including homochiral fixed points. One could then conclude that they are all unstable under the same criterions. This argument is in fact wrong, as one can not construct the same kind of Jacobian matrix as previously done when considering a homochiral fixed point. In the case of a homochiral fixed point (or a non racemic fixed point), one would have to determine the spectrum of

the complete $(2N_C + N_A) \times (2N_C + N_A)$ Jacobian \mathbf{J}_{comp} which reads

$$\mathbf{J}_{\text{comp}} = \begin{pmatrix} \partial_{\mathbf{c}_D} \mathbf{F}_D & \partial_{\mathbf{c}_A} \mathbf{F}_D & \partial_{\mathbf{c}_L} \mathbf{F}_D \\ \partial_{\mathbf{c}_D} \mathbf{F}_A & \partial_{\mathbf{c}_A} \mathbf{F}_A & \partial_{\mathbf{c}_L} \mathbf{F}_A \\ \partial_{\mathbf{c}_D} \mathbf{F}_L & \partial_{\mathbf{c}_A} \mathbf{F}_L & \partial_{\mathbf{c}_L} \mathbf{F}_L \end{pmatrix}, \quad (2.48)$$

evaluated at the homochiral fixed point, associated to the linear dynamical system

$$\frac{d\mathbf{c}}{dt} = \left(\mathbf{J}_{\text{comp}} - \frac{1}{\tau} \mathbf{I} \right) \cdot \mathbf{c}, \quad (2.49)$$

to determine the stability of the fixed point. In this case, the RM treatment performed for racemic fixed points is not expected to work. An example of resolution of this issue for a particular chemical network is shown in the next section using a generalized version of the Frank model.

2.3 Example: A generalization of the Frank model

In this section, we develop an alternative version of the well known Frank model generalized to an arbitrary number of chiral and achiral species. This serves as an example for the general model developed in the previous section.

2.3.1 The generalized model

In order to test the random matrix scenario, we introduce a generalization of the Frank model [Frank, 1953], in which the numbers of chiral and achiral species are significantly increased and we assume an arbitrary assignation L or D to each enantiomer. Reverse reactions are also included in order to guarantee the compatibility with the existence of an equilibrium state even though the system is driven out of equilibrium by an inflow of matter. It is essential that the system is driven out of equilibrium in order for a chiral bias to be maintained. We thus assume that the system is thermodynamically open, due to fluxes of matter in and out of the system. In this generalized model, we 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 and are not supplied to the system (it is just a particular case of a racemic inflow where $D_{i0} = L_{i0} = 0 \forall i$ if D_{i0} (*resp.* L_{i0}) denotes the inflow concentration of the i -th D-enantiomer (*resp.* L-enantiomer)). In this regard, the achiral species $\{A_a\}_{a=1}^{N_A}$ are of high free energy, and the achiral species $\{\tilde{A}_a\}_{a=1}^{\tilde{N}_A}$ of low free energy. The reaction

networks are given by the following reactions:



where the enantiomer species are either $E_m = D_m$ and $\bar{E}_m = L_m$, or $E_m = L_m$ and $\bar{E}_m = D_m$ for each enantiomeric pair $m = i, j, k = 1, 2, \dots, N_C$. The achiral species are labeled with $a = 1, 2, \dots, N_A$; and $b, c = 1, 2, \dots, \tilde{N}_A$. Equations (2.50)-(2.52) define a total of 2^{N_C-1} nonequivalent 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 $E_m = D_m$ and $\bar{E}_m = L_m$ for all the pairs $m = 1, 2, \dots, N_C$ is the direct generalization of Frank model, considered below. Here the reactions (2.50)-(2.52) all have some degree of enantioselectivity as in the original Frank model [Frank, 1953]. The dynamics of the system is described by the following set of equations

$$\dot{A}_d = - \sum_{\substack{aijk \\ j \leq k}} \delta_{da} w_{aijk}^{(D)} - \sum_{\substack{aijk \\ j \leq k}} \delta_{da} w_{aijk}^{(L)} + \frac{1}{\tau} (A_{0d} - A_d), \quad (2.53)$$

$$\dot{D}_m = \sum_{\substack{aijk \\ j \leq k}} \nu_{m,ijk} w_{aijk}^{(D)} - \sum_{\substack{bcij \\ b \leq c}} \delta_{mi} \tilde{w}_{bcij} - \frac{1}{\tau} D_m, \quad (2.54)$$

$$\dot{L}_m = \sum_{\substack{aijk \\ j \leq k}} \nu_{m,ijk} w_{aijk}^{(L)} - \sum_{\substack{bcij \\ b \leq c}} \delta_{mj} \tilde{w}_{bcij} - \frac{1}{\tau} L_m, \quad (2.55)$$

$$\dot{\tilde{A}}_e = \sum_{\substack{bcij \\ b \leq c}} (\delta_{eb} + \delta_{ec}) \tilde{w}_{bcij} - \frac{1}{\tau} \tilde{A}_e, \quad (2.56)$$

where

$$\nu_{m,ijk} \equiv -\delta_{mi} + \delta_{mj} + \delta_{mk} \quad (2.57)$$

and the rates w read, assuming mass-action law,

$$w_{aijk}^{(D)} = k_{+aijk} A_a D_i - k_{-aijk} D_j D_k \quad \text{with} \quad j \leq k, \quad (2.58)$$

$$w_{aijk}^{(L)} = k_{+aijk} A_a L_i - k_{-aijk} L_j L_k \quad \text{with} \quad j \leq k, \quad (2.59)$$

$$\tilde{w}_{bcij} = \tilde{k}_{-bcij} D_i L_j - \tilde{k}_{+bcij} \tilde{A}_b \tilde{A}_c \quad \text{with} \quad b \leq c, \quad (2.60)$$

where the positive sign in the subscripts of the rate constants refers to the direction of the reaction. We note that $\tilde{k}_{\pm bcij} = \tilde{k}_{\pm bcji}$ because of the mirror symmetry. The reversible generalized Frank model described by reactions (2.50)-(2.52) undergoes relaxation to racemic equilibrium in a closed system, corresponding to an infinite mean residence time $\tau = \infty$ in equation (2.6). Here, we consider the generalized Frank model under far-from-equilibrium conditions where a bifurcation towards homochirality happens beyond a threshold in the nonequilibrium driving due to injection of the high free-energy achiral species $\{A_a\}_{a=1}^{N_A}$ with respect to ejected low free-energy ones $\{\tilde{A}_a\}_{a=1}^{\tilde{N}_A}$. In such far-from-equilibrium regimes, the reversed reactions are often playing a negligible role, which can be described by assuming that their reaction rate constants are arbitrarily small (a proof that this assumption is compatible with the existence of an equilibrium state is available in Appendix C of [Laurent et al., 2022]). With this assumption, and assuming that only one achiral species of high free energy and one achiral species of low free energy are present in the system, the reaction scheme we study in the following simplifies as



with k_{+ijk} (*resp.* k_{-ij}) the rate constant associated with reactions (2.61) and (2.62) (*resp.* reaction (2.63)). In this case, the $2 + 2N_C$ kinetics equations are

$$\begin{aligned} \dot{A} &= - \sum_{\substack{ijk \\ j \leq k}} k_{+ijk} A D_i - \sum_{\substack{ijk \\ j \leq k}} k_{+ijk} A L_i + \frac{1}{\tau} (A_0 - A), \\ \dot{D}_m &= - \sum_{\substack{ij \\ i \leq j}} k_{+mij} A D_m + \sum_{\substack{ij \\ m \leq j}} k_{+imj} A D_i + \sum_{\substack{ij \\ j \leq m}} k_{+ijm} A D_i - \sum_i \tilde{k}_{-mi} D_m L_i - \frac{1}{\tau} D_m, \\ \dot{L}_m &= - \sum_{\substack{ij \\ i \leq j}} k_{+mij} A L_m + \sum_{\substack{ij \\ m \leq j}} k_{+imj} A L_i + \sum_{\substack{ij \\ j \leq m}} k_{+ijm} A L_i - \sum_i \tilde{k}_{-im} D_i L_m - \frac{1}{\tau} L_m, \\ \dot{\tilde{A}} &= 2 \sum_{ij} \tilde{k}_{-ij} D_i L_j - \frac{1}{\tau} \tilde{A}. \end{aligned} \quad (2.64)$$

Using a Runge-Kutta algorithm of second order, one can numerically solve the set of equations (2.64) and observe in Figure 2.8 different behaviors for different inflow concentrations of achiral species for a fixed number of chiral species N_C . The numerical integration of the kinetic equations has been performed by setting $\tau = 1$, meaning that we take τ as the time unit. All the simulations of the generalized Frank model were

performed using C programming language. 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 of the species. The rate constants are all supposed to be randomly distributed according to a log-normal distribution whose average and standard deviation are specified in the caption. We can directly observe that depending on whether the inflow concentration A_0 of achiral species of high free energy is low or large, we end up in a racemic or a homochiral state (it is also illustrated in the phase diagram of Figure 2.10a). At N_C fixed, an increase of A_0 can be seen as an increase of energy supplied to the reactor. Here, the more we drive the system far from equilibrium, the more it seems to converge towards a homochiral steady-state. To study further

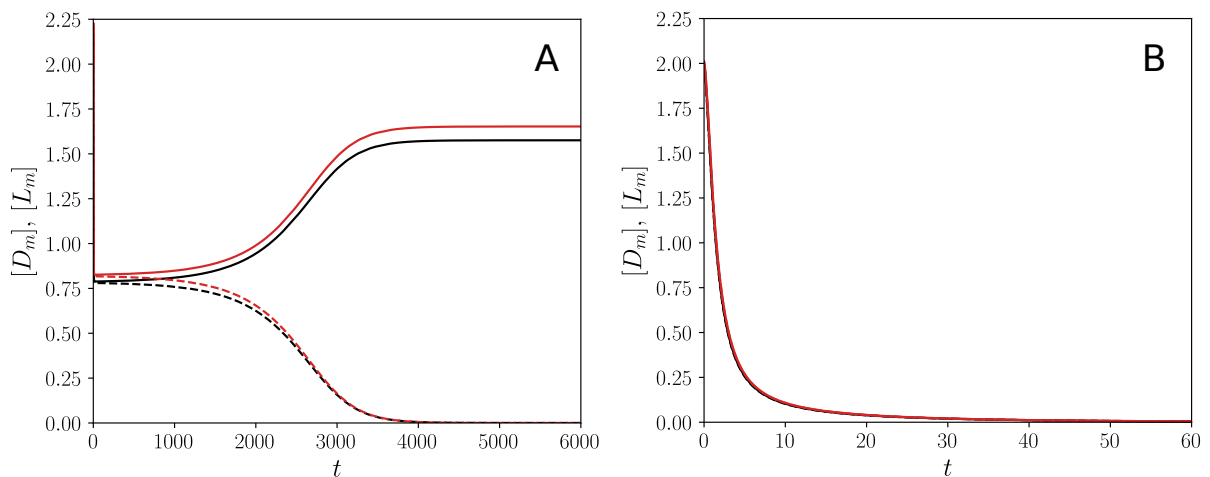


Figure 2.8: Dynamical simulations of the autocatalytic network (2.61)-(2.62)-(2.63). Typical time evolution of two species contained in the autocatalytic network as a function of time (A) above the threshold concentration A_0^* from equation (2.83) and (B) below it. The solid lines represent one of the 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 inactivated achiral species 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. The number of chiral species was set up to $N_C = 20$.

the dynamics of this system, the stability of the stationary states can be determined by solving equations (2.64) analytically using a similar random matrix treatment as for the general model of Section 2.2.

2.3.2 Stability of the steady-states

We introduce the variables $\mathbf{D} = \{D_i\}_{i=1}^{N_C}$ and $\mathbf{L} = \{L_i\}_{i=1}^{N_C}$ which are concentration vectors for D- and L-enantiomers.

Racemic stationary states

Following the same recipe as for the general model of open reactor, and using mirror symmetry, we focus our analysis on the dynamics of infinitesimal deviations from the racemic state

$$\delta\mathbf{X} = \frac{1}{2}(\mathbf{L} - \mathbf{D}), \quad (2.65)$$

which are ruled by the following set of linear equations

$$\frac{d}{dt}\delta\mathbf{X} = \mathbf{M} \cdot \delta\mathbf{X}, \quad (2.66)$$

where the $N_C \times N_C$ matrix \mathbf{M} is given by

$$\mathbf{M} = \left. \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} \right|_{\mathbf{c}_*} - \left. \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{L}} \right|_{\mathbf{c}_*}, \quad (2.67)$$

knowing that the mirror symmetry implies that

$$\left. \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} \right|_{\mathbf{c}_*} = \left. \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} \right|_{\mathbf{c}_*}, \quad (2.68)$$

and

$$\left. \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{L}} \right|_{\mathbf{c}_*} = \left. \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{D}} \right|_{\mathbf{c}_*}, \quad (2.69)$$

where \mathbf{c}_* denotes a racemic steady-state. Deriving the equation for \dot{D}_m in the equation set (2.64) with respect to \dot{D}_n and \dot{L}_n we obtain the matrix elements of \mathbf{M} :

$$M_{mn} = \frac{\partial \dot{D}_m}{\partial D_n} - \frac{\partial \dot{D}_m}{\partial L_n}, \quad (2.70)$$

which read

$$M_{mn} = \sum_{\substack{i \\ m \leq i}} k_{+nmi} A + \sum_{\substack{i \\ i \leq m}} k_{+nim} A + \tilde{k}_{-mn} D_m - \delta_{mn} \left(\sum_{\substack{ij \\ i \leq j}} k_{+nij} A + \sum_i \tilde{k}_{-mi} L_i + \frac{1}{\tau} \right), \quad (2.71)$$

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. The inverse time residence once again appears only on diagonal elements of \mathbf{M} having the effect of shifting its spectrum towards negative values of the complex plane. We can thus decompose \mathbf{M} as

$$\mathbf{M} = \mathbf{J} - \frac{1}{\tau} \mathbf{I}. \quad (2.72)$$

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 Appendix 2.B into a mean value and fluctuations of root mean square. We consider now the particular case of the racemic fixed point where $D_i = L_i = \tilde{A} = 0$ for all species i and $A = A_0$ which we call the trivial racemic fixed point. The elements of the matrix \mathbf{M} thus become

$$M_{mn} = A_0 \left(\sum_{\substack{i \\ m \leq i}} k_{+nmi} + \sum_{\substack{i \\ i \leq m}} k_{+nim} \right) - \delta_{mn} \left(A_0 \sum_{\substack{ij \\ i \leq j}} k_{+nij} + \frac{1}{\tau} \right). \quad (2.73)$$

We observe in equations (2.71) and (2.73) that not only the inverse residence time appears only on diagonal element but also a second term depending on A_0 and a sum of rate constants. We thus separate the matrix \mathbf{M} in three contributions

$$\mathbf{M} = A_0 \mathbf{Q} - A_0 \mathbf{R} - \frac{1}{\tau} \mathbf{I}, \quad (2.74)$$

where \mathbf{Q} is a matrix whose elements are given by

$$Q_{mn} = \sum_{\substack{i \\ m \leq i}} k_{+nmi} + \sum_{\substack{i \\ i \leq m}} k_{+nim}, \quad (2.75)$$

and \mathbf{R} a diagonal matrix whose elements are given by

$$R_{nn} = \sum_{\substack{ij \\ i \leq j}} k_{+nij}. \quad (2.76)$$

The matrix elements of \mathbf{Q} are sums of $N_C + 1$ random variables whose mean and standard deviation are respectively $\langle k_+ \rangle$ and σ_{k_+} . Using the central limit theorem (CLT), we deduce that the elements of \mathbf{Q} are distributed according to a Gaussian distribution of mean $\mu_Q = \langle k_+ \rangle (N_C + 1)$ and standard deviation $\sigma_Q = \sigma_{k_+} \sqrt{N_C + 1}$. From the construction of the matrix \mathbf{R} we deduce that its diagonal elements are also randomly distributed following another Gaussian distribution of mean $\mu_R = \langle k_+ \rangle N_C (N_C + 1) / 2$ and standard deviation $\sigma_R = \sigma_{k_+} \sqrt{N_C (N_C + 1) / 2}$. The spectrum of \mathbf{M} cannot be determined simply by summing the spectrums of \mathbf{Q} and \mathbf{R} as they have no reason to be diagonal in the same basis. However, we can decompose the matrix \mathbf{Q} as an average term plus small deviations:

$$\mathbf{Q} = \mu_Q \mathbf{1} + \sigma_Q \mathbf{G}, \quad (2.77)$$

where $\mathbf{1}$ is a matrix full of ones, and \mathbf{G} is a matrix whose entries are distributed according to a Gaussian statistics of mean zero and unit variance. The matrix \mathbf{R} can also be decomposed into two distinct contributions

$$\mathbf{R} = \mu_R \mathbf{I} + \mathbf{H}, \quad (2.78)$$

where \mathbf{H} is a diagonal matrix with subdominant terms as compared to μ_R (this follows from the law of large numbers). Using decompositions of equations (2.77) and (2.78), the matrix \mathbf{M} can be rewritten as

$$\mathbf{M} = \mathbf{P} + A_0 \sigma_Q \mathbf{G} - A_0 \mathbf{H}, \quad (2.79)$$

where

$$\mathbf{P} = A_0 \mu_Q \mathbf{1} - (A_0 \mu_R + \frac{1}{\tau}) \mathbf{I}. \quad (2.80)$$

The largest eigenvalue of \mathbf{P} thus of \mathbf{M} , if we neglect the contributions from the subdominant terms, is given by

$$\lambda_{\max}^M = A_0 \langle k_+ \rangle \frac{N_C (N_C + 1)}{2} - \frac{1}{\tau}, \quad (2.81)$$

recalling that $\langle k_+ \rangle > 0$ as we supposed that all k_{+ijk} , with $i, j, k \in [1, N_C]$ are distributed according to a log-normal law. For the trivial racemic state to be unstable, the dominant eigenvalue λ_{\max}^M must be positive, thus the threshold above which this state loses its stability is

$$A_0 > A_0^*, \quad (2.82)$$

where

$$A_0^* = \frac{2}{N_C(N_C + 1)\langle k_+ \rangle \tau}. \quad (2.83)$$

This result holds and numerical simulations do not show deviations from the theoretical prediction as long as the standard deviation σ_{k_+} remains low compared to $\langle k_+ \rangle$, as it is shown in Figure 2.9.

Symmetry properties of the Jacobian matrix

We remark that the permutation $D_m \leftrightarrow L_m$ for an enantiomeric pair m implies that the corresponding enantiomeric excess changes sign $\delta X_m \rightarrow -\delta X_m$, and that the eigenvalues of the matrix \mathbf{M} remain unchanged under such transformations. We denote \mathbf{u} an eigenvector (associated to the eigenvalue λ) of the matrix \mathbf{M} , and \mathbf{u}' (associated to the eigenvalue p') the new eigenvector of the transformed matrix \mathbf{M}' obtained after a permutation $D_m \leftrightarrow L_m$, so that

$$\sum_j M_{ij} u_j = p u_i, \quad (2.84)$$

and

$$\sum_j M'_{ij} u'_j = p' u'_i. \quad (2.85)$$

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 equation (2.66), we obtain that $M'_{ij} = (-1)^{s_i + s_j} M_{ij}$. It follows from this that the eigenvectors transform as $u'_i = (-1)^{s_i} u_i$ with the eigenvalue conserved in the permutation $p' = p$. Thus, the result regarding the transition to homochirality holds for the 2^{N_C-1} models considered with the reaction scheme described by equations (2.50)-(2.52). This invariance property is important for the discussion on chiral signs of different groups of molecule in Section 1.1.3. It shows that incompatibilities in the different definitions of chirality are not an issue, nor the fact that different groups of molecules do not share the same chirality.

Existence of a homochiral stationary state

It is also possible to show that the loss of stability for the trivial racemic state coincides with the gain of stability for a homochiral state. As stated in the Section 2.2.7, homochiral fixed points do not have the same symmetry relations than racemic states. Therefore a similar treatment as done previously is not possible and we have to consider a standard approach of the Jacobian to solve their stability. A first observation we can make is that

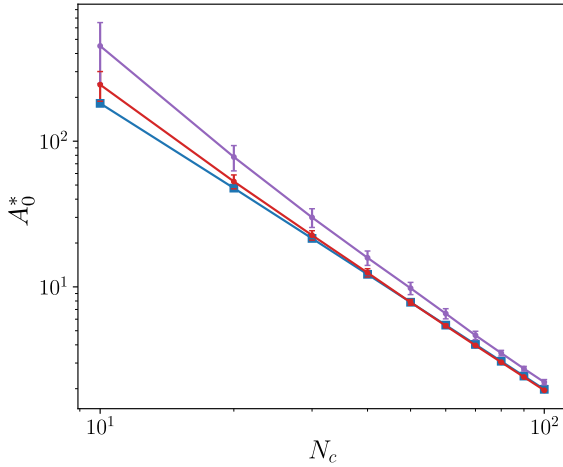


Figure 2.9: Comparison between the observed control parameter value A_0 at the transition with the theoretical prediction given by equation (2.83) (blue solid line) after averaging over 100 realizations of the rate constants for different standard deviation σ_{k_+} and $\sigma_{\tilde{k}_-}$ of rate constants : $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 10^{-3}$ (red), $\sigma_{k_+} = \sigma_{\tilde{k}_-} = 10^{-2}$ (purple), while $\langle k_+ \rangle = \langle \tilde{k}_- \rangle = 10^{-4}$.

the set of equations (2.64) is compatible with a homochiral steady state since the condition $\dot{L}_m = 0$ is verified when $L_i = 0$ for every species i (or the same condition for D-species in the opposite homochiral state). We first obtain an expression for the concentration of the activated achiral species A in the steady state, using the condition that $D_i \neq 0$ and $L_i = 0$ for all species i :

$$A = \frac{A_0}{\tau \sum_{\substack{ijk \\ j \leq k}} k_{+ijk} D_i + 1}. \quad (2.86)$$

The sum in the denominator can be simplified using the central limit theorem (CLT) :

$$\sum_{\substack{ijk \\ j \leq k}} k_{+ijk} D_i = \langle k_+ \rangle \frac{N_C(N_C + 1)}{2} \sum_i D_i, \quad (2.87)$$

which tells us that the concentration A of achiral species in the steady state reads

$$A = \frac{A_0}{\tau \langle k_+ \rangle \frac{N_C(N_C + 1)}{2} \sum_i D_i + 1}. \quad (2.88)$$

Summing the second equation in (2.64) over m , in the steady state, we find that

$$-A \sum_m D_m \sum_{\substack{ij \\ i \leq j}} k_{+mij} + A \sum_i D_i \sum_{\substack{mj \\ m \leq j}} k_{+imj} + A \sum_i D_i \sum_{\substack{mj \\ j \leq m}} k_{+ijm} - \frac{1}{\tau} \sum_m D_m = 0, \quad (2.89)$$

from which, using the CLT to compute the sums, and the property that all D_i should be non zero and positive concentrations, we deduce an alternative expression for A denoted A_0^* in the steady state:

$$A = \frac{2}{\tau \langle k_+ \rangle N_C(N_C + 1)}. \quad (2.90)$$

Looking back at equation (2.83), we recognize the threshold for the stability of the racemic state A_0^* , so the previous equation is equivalent to

$$A = A_0^*. \quad (2.91)$$

Combining the two expressions for A , namely (2.88) and (2.90), we obtain an expression for $\sum_i D_i$ in the steady state

$$\sum_i D_i = A_0 - \frac{2}{\tau \langle k_+ \rangle N_C (N_C + 1)}. \quad (2.92)$$

A solution where $\sum_i D_i < 0$ has no physical meaning here as D_i s are concentrations. Therefore a full homochiral state exists only if $A_0 > A_0^*$. This makes sense because this is precisely the condition for the racemic state to lose its stability we derived in equation (2.82). So far in this derivation, we have used only the global condition $\sum_i D_i > 0$, which does not ensure that $D_i > 0, \forall i$ which is also required. Going back to equation set (2.64) and using once again the CLT, one finds that in the steady state,

$$D_m = \frac{A \sum_i D_i \left(\sum_{m \leq j}^j k_{+imj} + \sum_{j \leq m}^j k_{+ijm} \right)}{A \sum_{i \leq j}^{ij} k_{+mij} + \frac{1}{\tau}}, \quad (2.93)$$

thus

$$D_m = \frac{A \langle k_+ \rangle (N_C + 1) \sum_i D_i}{A N_C (N_C + 1) \langle k_+ \rangle / 2 + 1/\tau} > 0, \quad (2.94)$$

which shows that all D_i 's are positive and approximately equal to each other since $\sum_i D_i > 0$ if $A_0 > A_0^*$. With a mean-field approximation, we can also deduce from equation (2.92) that

$$D = \frac{A_0 - A_0^*}{N_C}. \quad (2.95)$$

So far, we have characterized the homochiral steady state. Now we have to study its stability to show that this state becomes an attractor of the dynamics when the trivial racemic states loses its stability.

Stability of the homochiral fixed point

As stated earlier, this section serves to show how different the study of stability of the homochiral state is, which helps to understand the specificity of the result regarding the stability of the racemic state in equation (2.66). Here we have to consider the $(2N_C +$

1) $\times (2N_C + 1)$ Jacobian matrix \mathbf{N} given by

$$\mathbf{N} = \begin{pmatrix} \frac{\partial \dot{A}}{\partial A} & \frac{\partial \dot{A}}{\partial \mathbf{D}} & \frac{\partial \dot{A}}{\partial \mathbf{L}} \\ \frac{\partial \dot{\mathbf{D}}}{\partial A} & \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} & \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{L}} \\ \frac{\partial \dot{\mathbf{L}}}{\partial A} & \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{D}} & \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} \end{pmatrix}. \quad (2.96)$$

The details of the computation are written in Appendix 2.D but the main ideas are set up here. Especially, all the elements of the matrix \mathbf{N} are given in Appendix 2.C. Since $\partial \dot{\mathbf{L}}/\partial A$ and $\partial \dot{\mathbf{L}}/\partial \mathbf{D}$ are respectively a null vector and a null matrix in this homochiral state, the characteristic determinant of \mathbf{N} can be decomposed into a determinant product of two diagonal square blocks

$$\det(\mathbf{N} - \lambda \mathbf{I}) = \det \begin{pmatrix} \frac{\partial \dot{A}}{\partial A} - \lambda & \frac{\partial \dot{A}}{\partial \mathbf{D}} \\ \frac{\partial \dot{\mathbf{D}}}{\partial A} & \frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}} - \lambda \mathbf{I} \end{pmatrix} \det \left(\frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} - \lambda \mathbf{I} \right). \quad (2.97)$$

The spectrum of \mathbf{N} is thus determined by the eigenvalues of its upper-left block and its lower-right block. The spectrum of the block $\partial \dot{\mathbf{L}}/\partial \mathbf{L}$ and its maximal eigenvalue are given in Appendix 2.C. The spectrum of the upper-left block of matrix \mathbf{N} is determined in Appendix 2.D. The analysis shows that the spectrum of the latter submatrix is given by $N_C - 1$ unperturbed negative eigenvalues of $\partial \dot{\mathbf{D}}/\partial \mathbf{D}$ lying into Girko's circle plus two eigenvalues given by

$$\lambda_+ = -\frac{1}{\tau}, \quad \text{and} \quad \lambda_- = \frac{1}{\tau} \left(1 - \frac{A_0}{A_0^*} \right). \quad (2.98)$$

The first eigenvalue λ_+ is always negative. The second one λ_- is negative when $A_0 > A_0^*$, i.e., when the racemic state becomes unstable and at the same time, a physically acceptable homochiral state with $D_i > 0, \forall i$ becomes possible. As shown in Appendix 2.C, the maximal eigenvalue of the matrix $\partial \dot{\mathbf{L}}/\partial \mathbf{L}$ is

$$\lambda_{\max}^L = -N_C D \langle \tilde{k}_- \rangle \quad (2.99)$$

in the large N_C limit, where $\langle \tilde{k}_- \rangle$ is the mean value of chiral inhibition rate constants. Furthermore, in the mean-field approximation, we have $D = (A_0 - A_0^*)/N_C$ by equation (2.95). Thus, in the regime of the homochiral state (i.e., when $A_0 > A_0^*$), this state is

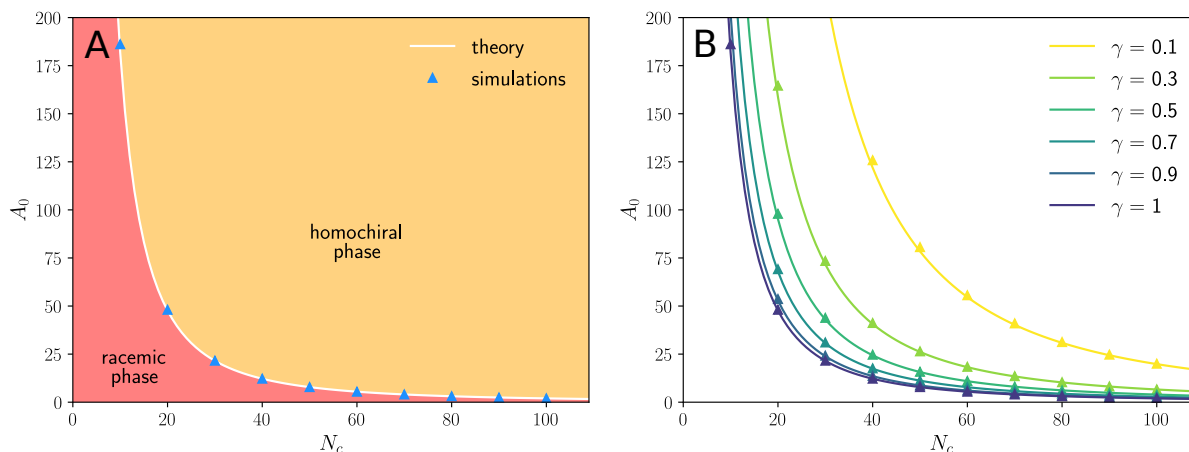


Figure 2.10: Phase diagram showing the border between the racemic and homochiral phases in the (A_0, N_C) plane for the generalized Frank model. A_0 is the inflow concentration of achiral species and N_C the number of chiral species in the reaction network. (A) In the lower domain of the diagram, the racemic state is stable against small perturbations. In the upper domain, the racemic state loses its stability while the homochiral state has gained stability. (B) Phase diagram showing the border between the racemic and homochiral phases for various sparsity coefficient γ . Below each solid line, the racemic phase is stable whereas above it it is unstable and instead the homochiral phase becomes stable. Simulations were performed for 100 realizations of the rate constants k_+ . Triangles represent simulations and solid lines the theory for both figures. The thresholds for both figures were computed by analyzing the stability of the racemic state of the system with simulations using the programming language C.

stable because $\lambda_{\max}^L < 0$. When $A_0 > A_0^*$ the racemic state is unstable but the homochiral state is an attractor of the dynamics. Conversely, when $A_0 < A_0^*$, the nonphysical condition $D < 0$ which is obtained means that there can be no stable homochiral state and in that case, the system should converge towards the racemic state as it is an attractor of the dynamics. The presence of the mean value of chiral inhibition rate constants in λ_{\max}^L underlines its key role in the stability of the homochiral state. Several reaction schemes were studied, and the only ones that admit stable and fully homochiral states seem to be the ones which possess chiral inhibition reactions. This whole stability analysis shows that for this model they are two ways to break the stability of the racemic state: either by increasing the number of chiral species in the system N_C , as proposed in the general model for chiral chemical networks, or by increasing A_0 , the inflow concentration of achiral species of high free energy, which is equivalent to increase the amount of energy supplied to the system and thus driving the system further out of equilibrium. A phase diagram summing up those ideas is shown in Figure 2.10.

2.3.3 How does the sparsity of the network affect the results?

As stated in Section 2.2.6, different mechanisms could be responsible for making the Jacobian matrix sparse, but random matrix results although modified, still hold under such conditions. Instead of only looking at sparsity of the global set of reactions, we introduce a new parameter $\gamma \in [0, 1]$ which quantifies the probability of autocatalytic reactions (reactions (2.61) and (2.62)) to be present in the system, meaning that every autocatalytic reaction has the chance $1 - \gamma$ to have a zero rate constant. In this case, the criterion for the stability of the racemic state to be broken becomes $A_0 > A_0^*/\gamma$, meaning that we need to increase the energy supply or the number of chiral species to achieve homochirality in a sparser chemical system. This result, which goes in the same direction as the general model, is however not straightforwardly linked to the sparsity described in Section 2.2.6. For the general model the sparsity coefficient described the global sparsity of the Jacobian matrix while here, γ accounts for autocatalytic reactions only, and not directly the sparsity of \mathbf{M} . The dependence of the threshold on γ is illustrated in Figure 2.10. In addition, we also introduce the parameter $\beta \in [0, 1]$ which quantifies the probability of chiral inhibition reactions (reactions (2.63)) to be present in the system. It appears that this kind of sparsity, characterized by β , does not affect the stability of the homochiral state as long as every species (together with its enantiomer) is involved in at least one chiral inhibition reaction. It makes sense as inhibition plays a key role in the Frank model to mimic a competition between species. However if β does not affect the stability of the system to some extent, it affects its dynamics. Indeed, one can show that the characteristic time t_c to reach the homochiral state when $A_0 > A_0^*/\gamma$ becomes

$$t_c \propto \frac{1}{\beta(A_0^*/\gamma - A_0)}, \quad (2.100)$$

provided

$$\beta > \frac{\gamma}{\tau A_0^* \langle \tilde{k}_- \rangle}, \quad (2.101)$$

and

$$t_c \propto 1 - \frac{\gamma A_0}{A_0^*}, \quad (2.102)$$

otherwise. Note that there is also a critical slowing down as A_0 approaches the threshold value A_0^*/γ , which was already present in non-sparse networks. In the end, even though the stability might be affected, and the dynamics changed, results still hold but with a higher threshold for the required number of chiral species to obtain a homochiral state.

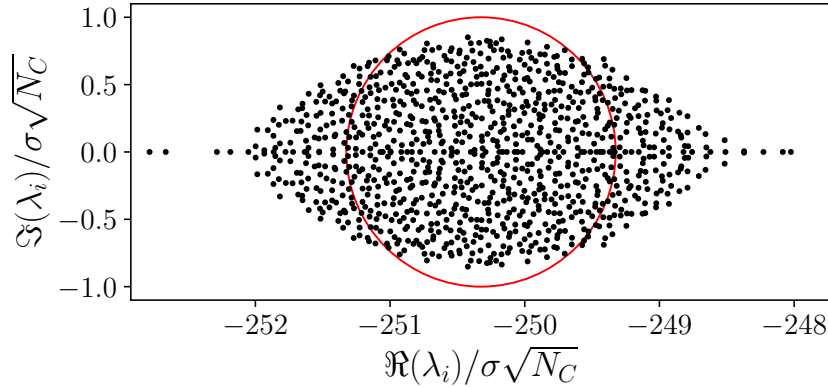


Figure 2.11: Close-up of the non-dominant eigenvalues of the Jacobian matrix \mathbf{M} of the generalized Frank model, which do not fill the Ginibre circle. Here, the matrix \mathbf{M} characterizes a system of $N_C = 1000$ chiral species, with rate constants distributed according to a log-normal distribution 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 $1/\sigma\sqrt{N_C}$ is expressed in terms of $\sigma = A_0\sigma_Q$ with $\sigma_Q = \sigma_{k_+}\sqrt{N_C + 1}$, which is the standard deviation of the matrix \mathbf{Q} elements in the decomposition of the matrix \mathbf{M} in equation (2.74).

2.3.4 Correlations in the Jacobian matrix

In general, it appears that we cannot really suppose the elements of \mathbf{J} from equations (2.33) or (2.66) to be perfect independent and identically distributed random variables. For many reasons (due to the structure of the reaction network) some correlations could appear in \mathbf{J} and in $\mathbf{M} = \mathbf{J} - \mathbf{I}/\tau$, affecting their spectrum and thus the stability of the racemic state. This effect on the spectrum can be observed in Figure 2.11, which displays the eigenvalues of the matrix \mathbf{M} in the case $N_C = 1000$ for the generalized Frank model. In this case the eigenvalues appear to be distributed in an ellipse rather than in a simple disk. Recently, structure dependent interaction matrices were studied for ecological system stability [Allesina and Tang, 2012, Allesina and Tang, 2015] and it has been found that interactions can distort the spectrum of the Jacobian matrix into ellipses. For instance, predator-prey interactions in ecological models imply new symmetry relations for matrix elements of the interaction matrix, leading to transformed eigenvalues and thus modified stability criterion than those presented in Section 2.2.4 (it is to be noted that structure dependence of the spectrum of RM has also been studied by mathematicians [Sommers et al., 1988, Aceituno et al., 2019]). However, even if Girko's circle law is affected, we observe in Figure 2.11 that the stability of the racemic state is still lost as before. The scaling of the isolated eigenvalue appears to be robust as it does not

get affected significantly by correlations of the Jacobian matrix. This also explains why, in our study of the generalized Frank model, we never observed mechanism (i) but only mechanism (ii). This example shows that the elements of Jacobian matrices are in general correlated, but we proved that nevertheless certain features remain. In particular, if the matrix elements are statistically correlated, the non-dominant eigenvalues may have a different distribution, but the isolated eigenvalue behaves similarly. This means that the scenario (ii) is expected to be robust.

2.4 Discussion

In this chapter, we showed that even when they are sparse, chiral molecular systems are susceptible to undergo a general and robust transition to homochirality when the number of chiral species becomes large. It is general in that the details of the reaction network should not matter, to some extent, according to the developments of Section 2.3. Its robustness comes from the fact that the results, when rescaled, still hold when sparsity is added to the model, and from the RM treatment which shows that results still hold in the presence of correlations in the Jacobian matrix, coming from particular structural properties of the reaction networks. Moreover, the argument does not rely specifically on the well-mixed property of the CSTR that implies the homogeneity of the concentrations of species in the system: non well-mixed systems also induce sparsity in Jacobian matrices and the simple case of two diffusively linked compartments also exhibits a transition to homochirality as demonstrated in Appendix 2.E. There, we show that the dominant chiralities in each compartment synchronize or not depending on whether the exchanges due to the diffusion between the compartments are strong or not. The relevance of compartments in origins of life scenarios and prebiotic chemistry has already been appreciated [Matsumara et al., 2016, Bansho et al., 2016, Blokhuis et al., 2018, Furubayashi and Ichihashi, 2018, Laurent et al., 2019] thus we regard as essential that the results still hold for compartmentalized systems. Moreover, our model differs significantly from previous models described in the introduction, because it incorporates a crucial invariance with respect to permutation in the designation D/L of any member of a pair of enantiomers. Thus, the model agrees with the possibility that different groups of chiral molecules adopt different signs of chirality as observed in modern biology and not a global sign for all biomolecules.

It is also important to recall that the model should be considered far from equilibrium. A critique one could make to our model would be that in the infinite residence time limit $\tau \rightarrow \infty$, corresponding to the limit of a closed system, the instability criterion holds

under any conditions, which would mean that the racemic state is always unstable. This would be a major problem for closed chemical chiral systems, because in those systems, an equilibrium state is reached for long times and is always racemic. We address this criticism in two points: on one hand, the elements of the matrix \mathbf{J} depend on τ , since the stationary concentrations \mathbf{c}_* are given by equation (2.6) under the assumption of stationarity. This result could change the behavior of \mathbf{J} elements once $\tau \rightarrow \infty$. On the other hand, it should be understood that the model is relevant only out of equilibrium because this is the regime we expect to be most relevant for homochirality transition. Living beings are dissipative systems that remain living as long as they avoid thermodynamic equilibrium, or in other words, “Living matter evades the decay to equilibrium.” [Schrödinger, 1944]. It is the same for homochirality, which appears to be also characteristic of living systems: closed systems are not relevant since a transition to homochirality is not physically possible in them; thus, our model is to be considered in a far-from-equilibrium regime only. We believe that the general model proposed in this chapter in Section 2.3 is a strong result because it shows a universal behavior: any chiral system whose racemate deviations $\delta\mathbf{x}$ verifies equation (2.33) and whose Jacobian \mathbf{J} verifies the RM assumption should undergo a transition to homochirality given a similar criterion. This idea is illustrated with the generalization of the Frank model, whose stability was analyzed in Section 2.3. This model exhibited similar behavior for $N_C \gg 1$ or when a sufficient supply of energy is provided to the system, even for sparse networks, which is somehow related to the strength of interaction between chemical species through their concentrations (interaction strength and sparsity have also been identified as a control parameter for the stability of ecosystems [Ratzke et al., 2020]) as long as a minimal number of chiral inhibition reactions are present in the system. Chiral inhibition reactions have been thought to play a crucial role, along with autocatalytic behavior, for homochirality emergence, as a way to connect and to establish a competition between the two families of enantiomers in a chiral system. They also play a fundamental role in scenarios involving the polymerization of monomers such as nucleic or ribonucleic acids, as shown in the next chapter. Finally, from the chemical database analysis, we found that the crossover between the chiral and the achiral world occurs around 9 heavy atoms per molecule, which indicates that the transition presented in this section could already occur for prebiotic systems involving not too complex or too long molecules.

Appendix 2.A Additional data from the PubChem analysis

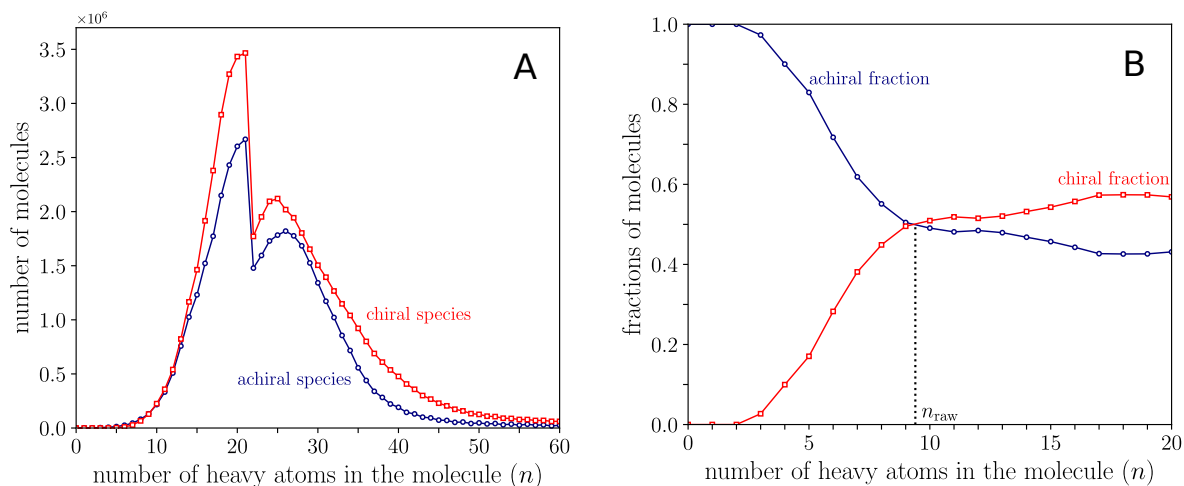


Figure 2.12: (A) Total number of achiral and chiral species in the raw PubChem database containing about 139 million of species. (B) Fractions of chiral and achiral molecules containing $n \leq 20$ heavy atoms. In this case, only 34 million molecules with $n \leq 20$ heavy atoms were analyzed after the specific selection. The figure on the right shows an intersection at $n_{\text{raw}} \simeq 9.4$.

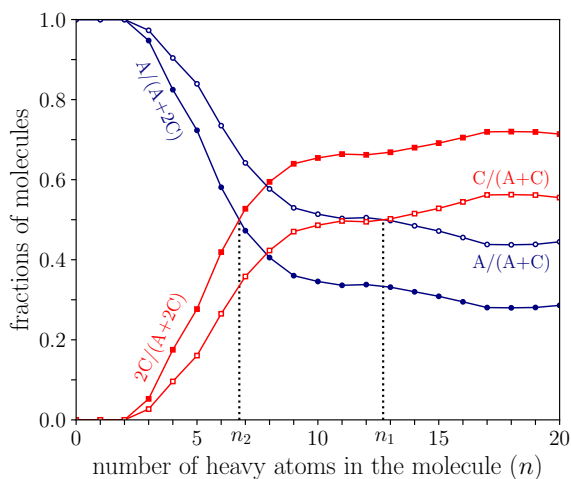


Figure 2.13: Analysis of the database expanded in enantiomers, with 50 million molecules in data (*i.e.*, 17 million enantiomers were generated) with $n \leq 20$ heavy atoms. The intersection occurs at $n_2 \simeq 6.7$ for if both enantiomers are considered and $n_1 \simeq 12.7$ if only one enantiomer is considered.

Appendix 2.B Eigenvalues of a random matrix with non zero mean

In this appendix, we show that the perturbed spectrum of an $N \times N$ random matrix whose entries are i.i.d. with a mean $\mu \neq 0$ and a variance σ^2 is the same as the case with $\mu = 0$ but with one isolated eigenvalue $\lambda_{\text{iso}} = \mu N$.

2.B.1 Isolated eigenvalue of a non zero mean random matrix

Let us suppose \mathbf{M} is a random matrix, with independent and identically distributed elements m_{ij} ($i, j \in [1, N]$) with mean μ and variance σ^2 . We can separate the mean from its fluctuations and rewrite $\mathbf{M} = \mu \mathbf{1} + \sigma \mathbf{F}$, with $\mathbf{1}$ a matrix full of ones (and not the unity matrix \mathbf{I}). The spectrum of $\mathbf{1}$ is constituted by the single top eigenvalue N and $N - 1$ times degenerated eigenvalue 0. The eigenvalue N is associated with the eigenvector $|1\rangle$. The eigenvalues 0 are associated with eigenvectors $(-1, 1, 0, \dots, 0, 0)^T$, $(-1, 0, 1, \dots, 0, 0)^T$, \dots , $(-1, 0, 0, \dots, 0, 1)^T$. Using perturbation theory with the normalized eigenvector $\frac{1}{\sqrt{N}} |1\rangle$, we find that the first-order perturbed single eigenvalue is

$$\lambda_{\text{iso}} = \mu N + \frac{\sigma}{N} \langle 1 | \mathbf{F} | 1 \rangle . \quad (2.103)$$

The perturbation term on the right hand side reads

$$\frac{\sigma}{N} \langle 1 | \mathbf{F} | 1 \rangle = \frac{\sigma}{N} \sum_{ij} f_{ij} , \quad (2.104)$$

and using the Central Limit Theorem,

$$\frac{\sigma}{N} \langle 1 | \mathbf{F} | 1 \rangle = \frac{\sigma}{N} N^2 \mu_{\mathbf{F}} + \frac{\sigma}{N} O(\sigma_{\mathbf{F}} N) , \quad (2.105)$$

and knowing that $\mu_{\mathbf{F}} = 0$ and $\sigma_{\mathbf{F}} = 1$,

$$\frac{\sigma}{N} \langle 1 | \mathbf{F} | 1 \rangle = O(\sigma) , \quad (2.106)$$

we obtain that

$$\lambda_{\text{iso}} = \mu N + O(\sigma) . \quad (2.107)$$

Thus, in the limit $\sigma/\mu N \ll 1$, the isolated eigenvalue of \mathbf{M} is μN .

2.B.2 Perturbation of other eigenvalues

The other eigenvalues are also perturbed. Their perturbation can also be determined by perturbation theory. Let us consider the vector $|k\rangle = (1, -1, 0, \dots, 0)^T$. Using the normalized vector $|k\rangle \rightarrow \frac{1}{\sqrt{2}}|k\rangle$, we find that the perturbation of the associated eigenvalue is given by

$$\frac{\sigma}{2} \langle k | \mathbf{F} | k \rangle = O(\sigma), \quad (2.108)$$

because f_{ij} are i.i.d. random variables with mean 0 and variance 1. Thus, the isolated eigenvalue is also the top eigenvalue (or the lower one if $\mu < 0$), provided $\sigma/\mu N \ll 1$.

Appendix 2.C Elements of the Jacobian in the homochiral state

The elements of the Jacobian matrix \mathbf{N} , evaluated in the fully homochiral state where $L_i = 0, \forall i$ read

$$\frac{\partial \dot{A}}{\partial A} = - \sum_{\substack{ijk \\ j \leq k}} k_{+ijk} D_i - \frac{1}{\tau}, \quad (2.109)$$

$$\frac{\partial \dot{A}}{\partial D_n} = - \sum_{\substack{jk \\ j \leq k}} k_{+njk} A, \quad (2.110)$$

$$\frac{\partial \dot{A}}{\partial L_n} = - \sum_{\substack{jk \\ j \leq k}} k_{+njk} A, \quad (2.111)$$

$$\frac{\partial \dot{D}_m}{\partial A} = - \sum_{\substack{ij \\ i \leq j}} k_{+mij} D_m + \sum_{\substack{ij \\ m \leq j}} k_{+imj} D_i + \sum_{\substack{ij \\ j \leq m}} k_{+ijm} D_i, \quad (2.112)$$

$$\frac{\partial \dot{L}_m}{\partial A} = 0, \quad (2.113)$$

$$\frac{\partial \dot{D}_m}{\partial D_n} = \sum_{\substack{i \\ m \leq i}} k_{+nmi} A + \sum_{\substack{i \\ i \leq m}} k_{+nim} A - \delta_{nm} \left(A \sum_{\substack{ij \\ i \leq j}} k_{+mij} + \frac{1}{\tau} \right), \quad (2.114)$$

$$\frac{\partial \dot{D}_m}{\partial L_n} = -\tilde{k}_{-mn} D_m, \quad (2.115)$$

$$\frac{\partial \dot{L}_m}{\partial D_n} = 0, \quad (2.116)$$

$$\frac{\partial \dot{L}_m}{\partial L_n} = \sum_{\substack{i \\ m \leq i}} k_{+nmi} A + \sum_{\substack{i \\ i \leq m}} k_{+nim} A - \delta_{nm} \left(A \sum_{\substack{ij \\ i \leq j}} k_{+mij} + \sum_i \tilde{k}_{-im} D_i + \frac{1}{\tau} \right) \quad (2.117)$$

Let us denote as λ^D (resp. λ^L) the eigenvalues of the matrix $\frac{\partial \dot{\mathbf{D}}}{\partial \mathbf{D}}$ (resp. $\frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}}$), and as λ_{\max}^D (resp. λ_{\max}^L) the corresponding maximal eigenvalue. Using methods introduced in Section 2.3.2, we obtain from equations (2.114) and (2.117)

$$\lambda_{\max}^D = A \langle k_+ \rangle \frac{N_C(N_C + 1)}{2} - \frac{1}{\tau}, \quad (2.118)$$

and

$$\lambda_{\max}^L = A \langle k_+ \rangle \frac{N_C(N_C + 1)}{2} - \frac{1}{\tau} + S, \quad (2.119)$$

where S is defined as $S = -DN_C \langle \tilde{k}_- \rangle$ and is due to the term $-\sum_i \tilde{k}_{-im} D_i$ in equation (2.117) for the m^{th} diagonal element. Using the mean-field expression for D derived in equation (2.95), and the expression for A from equation (2.90) we obtain

$$\lambda_{\max}^D = 0, \quad \lambda_{\max}^L = S = -N_C D \langle \tilde{k}_- \rangle \quad (2.120)$$

with $\langle \tilde{k}_- \rangle$ the mean value of chiral inhibition reaction constants.

Appendix 2.D Calculation of the maximal eigenvalue of the Jacobian matrix

We first define the upper-left block submatrix of the matrix \mathbf{N} as \mathbf{H} ,

$$\mathbf{H} = \begin{pmatrix} a & \mathbf{v}^T \\ \mathbf{u} & \mathbf{K} \end{pmatrix}, \quad (2.121)$$

where

$$a = \frac{\partial \dot{A}}{\partial A}, \quad (\mathbf{v}^T)_n = \frac{\partial \dot{A}}{\partial D_n}, \quad (\mathbf{u})_m = \frac{\partial \dot{D}_m}{\partial A}, \quad (\mathbf{K})_{mn} = \frac{\partial \dot{D}_m}{\partial D_n}. \quad (2.122)$$

In the large N_C limit, the vector \mathbf{u} converges towards a constant vector. Its elements are given by, according to equation (2.112),

$$(\mathbf{u})_m \underset{N_C \rightarrow \infty}{=} DN_C \frac{N_C + 1}{2} \langle k_+ \rangle, \quad (2.123)$$

thus using equation (2.95),

$$(\mathbf{u})_m \underset{N_C \rightarrow \infty}{=} (A_0 - A_0^*) \frac{N_C + 1}{2} \langle k_+ \rangle. \quad (2.124)$$

All the elements of \mathbf{u} converge to the same value given by the previous equation. We can do the same evaluation for \mathbf{v}^T defined by equation (2.122) using equation (2.90):

$$(\mathbf{v}^T)_n \underset{N_C \rightarrow \infty}{=} -AN_C \frac{N_C + 1}{2} \langle k_+ \rangle, \quad (2.125)$$

thus

$$(\mathbf{v}^T)_n \underset{N_C \rightarrow \infty}{=} -\frac{1}{\tau}. \quad (2.126)$$

Finally, we can compute the value of the scalar element a defined by equation (2.122) in the $N_C \rightarrow \infty$ limit:

$$a \underset{N_C \rightarrow \infty}{=} -D \langle k_+ \rangle N_C^2 \frac{N_C + 1}{2} - \frac{1}{\tau}, \quad (2.127)$$

thus

$$a \underset{N_C \rightarrow \infty}{=} -\frac{A_0}{A_0^* \tau}. \quad (2.128)$$

after simplification with equations (2.90) and (2.95). The characteristic determinant of the matrix \mathbf{H} can be written using the Schur complement of the block a , with λ a com-

plex variable such that $\lambda \notin \sigma_{\mathbf{K}} = \{\lambda_1, \lambda_2, \dots, \lambda_{N_C}\}$, a condition which guarantees the invertibility of $\mathbf{K} - \lambda \mathbf{I}$:

$$\det(\mathbf{H} - \lambda \mathbf{I}) = [a - \lambda - \mathbf{v}^T \cdot (\mathbf{K} - \lambda \mathbf{I})^{-1} \cdot \mathbf{u}] \det(\mathbf{K} - \lambda \mathbf{I}), \quad (2.129)$$

with \mathbf{I} the identity matrix. At this point, it is important to recall that, in the large N_C limit, the isolated eigenvalue of the matrix $\mathbf{K} = \frac{\partial \mathbf{D}}{\partial \mathbf{D}}$, which we here denote λ_1 , is associated to the eigenvector $(1, \dots, 1)^T$. Therefore we see that \mathbf{u} and \mathbf{v}^T are eigenvectors of matrix \mathbf{K} associated to λ_1 , i.e., $\mathbf{K} \cdot \mathbf{u} = \lambda_1 \mathbf{u}$, thus

$$(\mathbf{K} - \lambda \mathbf{I})^{-1} \cdot \mathbf{u} = \frac{1}{\lambda_1 - \lambda} \mathbf{u}. \quad (2.130)$$

A similar development could be done by considering \mathbf{v}^T as a left-eigenvector of \mathbf{K} . Substituting this relation into equation (2.129), we get

$$\det(\mathbf{H} - \lambda \mathbf{I}) = \left(a - \lambda - \frac{\mathbf{v}^T \cdot \mathbf{u}}{\lambda_1 - \lambda} \right) \det(\mathbf{K} - \lambda \mathbf{I}), \quad (2.131)$$

thus

$$\det(\mathbf{H} - \lambda \mathbf{I}) = \frac{(a - \lambda)(\lambda_1 - \lambda) - \mathbf{v}^T \cdot \mathbf{u}}{\lambda_1 - \lambda} \det(\mathbf{K} - \lambda \mathbf{I}). \quad (2.132)$$

We use the decomposition of $\det(\mathbf{K} - \lambda \mathbf{I})$ as a polynomial function of roots $\sigma_{\mathbf{K}} = \{\lambda_1, \lambda_2, \dots, \lambda_{N_C}\}$ (i.e., the eigenvalues of \mathbf{K}) to find

$$\det(\mathbf{H} - \lambda \mathbf{I}) = [(a - \lambda)(\lambda_1 - \lambda) - \mathbf{v}^T \cdot \mathbf{u}] (\lambda_2 - \lambda) \times \dots \times (\lambda_{N_C} - \lambda). \quad (2.133)$$

By continuity (since initially we had assumed that $\lambda \notin \sigma_{\mathbf{K}}$), we deduce from equation (2.133) that the spectrum of matrix \mathbf{H} is given by the $N_C - 1$ eigenvalues of \mathbf{K} distributed in the circle (from Girko's theorem), and two eigenvalues solutions of $(a - \lambda)(\lambda_1 - \lambda) - \mathbf{v}^T \cdot \mathbf{u} = 0$. As a result, the isolated eigenvalue λ_1 is modified and becomes one of the two solutions of the previous equation to solve. From equation (2.120), the isolated eigenvalue λ_1 should be identified with the largest eigenvalue λ_{\max}^D . This means that $\lambda_1 \rightarrow 0$ in the large N_C limit. The equation that remains to be solved to fully determine the spectrum of \mathbf{H} is then

$$\lambda^2 - a\lambda - \mathbf{v}^T \cdot \mathbf{u} = 0. \quad (2.134)$$

Using the expressions (2.124) and (2.126) for \mathbf{u} and \mathbf{v}^T , as well as equations (2.90), we determine their dot product

$$\mathbf{v}^T \cdot \mathbf{u} = -\frac{A_0 - A_0^*}{\tau} N_C \frac{N_C + 1}{2} \langle k_+ \rangle, \quad (2.135)$$

or simply

$$\mathbf{v}^T \cdot \mathbf{u} = -\frac{A_0 - A_0^*}{\tau^2 A_0^*}, \quad (2.136)$$

where the additional N_C factor comes from the summation. With equation (2.128), the discriminant of equation (2.134) can thus be written

$$\Delta = a^2 + 4 \mathbf{v}^T \cdot \mathbf{u} = \frac{1}{\tau^2} \left(\frac{A_0}{A_0^*} - 2 \right)^2. \quad (2.137)$$

We observe that $\Delta > 0$ when $A_0 > A_0^*$, but $\Delta = 0$ in the particular case where $A_0 = 2A_0^*$. When $A_0 \neq 2A_0^*$, the two real solutions of equation (2.134) are

$$\lambda_+ = \frac{1}{2} \left[-\frac{A_0}{A_0^* \tau} + \frac{1}{\tau} \left(\frac{A_0}{A_0^*} - 2 \right) \right] = -\frac{1}{\tau}, \quad (2.138)$$

and

$$\lambda_- = \frac{1}{2} \left[-\frac{A_0}{A_0^* \tau} - \frac{1}{\tau} \left(\frac{A_0}{A_0^*} - 2 \right) \right] = \frac{1}{\tau} \left(1 - \frac{A_0}{A_0^*} \right). \quad (2.139)$$

Finally, in the particular case where $A_0 = 2A_0^*$, there is a unique solution,

$$\lambda_+ = \lambda_- = -1/\tau. \quad (2.140)$$

Appendix 2.E Compartmentalized systems with diffusive coupling

In this appendix, we study the case of two diffusively coupled compartments for the general model developed in Section 2.3. In this case, species are exchanged through compartments with a diffusion coefficient κ such that the set of equations (2.13)-(2.15) becomes

$$\frac{d\mathbf{c}_{D,i}}{dt} = \mathbf{F}_D(\mathbf{c}_{D,i}, \mathbf{c}_{A,i}, \mathbf{c}_{L,i}) + \frac{1}{\tau}(\mathbf{c}_{D0,i} - \mathbf{c}_{D,i}) + \kappa(\mathbf{c}_{D,j} - \mathbf{c}_{D,i}), \quad (2.141)$$

$$\frac{d\mathbf{c}_{A,i}}{dt} = \mathbf{F}_A(\mathbf{c}_{D,i}, \mathbf{c}_{A,i}, \mathbf{c}_{L,i}) + \frac{1}{\tau}(\mathbf{c}_{A0,i} - \mathbf{c}_{A,i}) + \kappa(\mathbf{c}_{A,j} - \mathbf{c}_{A,i}), \quad (2.142)$$

$$\frac{d\mathbf{c}_{L,i}}{dt} = \mathbf{F}_L(\mathbf{c}_{D,i}, \mathbf{c}_{A,i}, \mathbf{c}_{L,i}) + \frac{1}{\tau}(\mathbf{c}_{L0,i} - \mathbf{c}_{L,i}) + \kappa(\mathbf{c}_{L,j} - \mathbf{c}_{L,i}), \quad (2.143)$$

where $(i, j) = (1, 2)$ or $(2, 1)$ are the indices referring to compartments 1 and 2. Linearizing the time evolution of infinitesimal deviations from the racemate in compartment i ,

$$\frac{d\delta\mathbf{x}_i}{dt} = \frac{1}{2} \frac{d}{dt}(\mathbf{c}_{D,i} - \mathbf{c}_{L,i}), \quad (2.144)$$

we obtain the equivalent of equation (2.33) for each compartment

$$\frac{d\delta\mathbf{x}_i}{dt} = \left(\mathbf{J}_i - \frac{1}{\tau} \mathbf{I} - \kappa \mathbf{I} \right) \cdot \delta\mathbf{x}_i + \kappa \cdot \delta\mathbf{x}_j, \quad (2.145)$$

provided $\delta\mathbf{x}_{0,i} = 0$ for each compartment, meaning that the inflow of species in both compartments is racemic, where \mathbf{J}_i is given by

$$\mathbf{J}_i = \left. \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_D} \right|_{\mathbf{c}_{*,i}} - \left. \frac{\partial \mathbf{F}_D}{\partial \mathbf{c}_L} \right|_{\mathbf{c}_{*,i}}, \quad (2.146)$$

and depends only on the racemic state considered in the compartment i , as the structure of the reaction network is the same in both compartments. In the special case where the same racemic state is considered in both compartments $\mathbf{c}_{*,1} = \mathbf{c}_{*,2}$, we deduce from equation (2.145) that the deviations of the whole system are described by the linear set of equations

$$\frac{d}{dt} \begin{pmatrix} \delta\mathbf{x}_{0,1} \\ \delta\mathbf{x}_{0,2} \end{pmatrix} = \begin{pmatrix} \mathbf{M} - \kappa \mathbf{I} & \kappa \mathbf{I} \\ \kappa \mathbf{I} & \mathbf{M} - \kappa \mathbf{I} \end{pmatrix} \cdot \begin{pmatrix} \delta\mathbf{x}_{0,1} \\ \delta\mathbf{x}_{0,2} \end{pmatrix}, \quad (2.147)$$

where $\mathbf{M} = \mathbf{J} - \frac{1}{\tau} \mathbf{I}$ and $\mathbf{J} = \mathbf{J}_1 = \mathbf{J}_2$. 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 $\sqrt{N_C}$ 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} the eigenvector of the matrix \mathbf{M} and the corresponding eigenvalue λ . A simple calculation provides the eigenvectors \mathbf{u}' and the eigenvalues λ' of the matrix \mathbf{L} as function of \mathbf{u} and λ . In the case where the dominant eigenvalue of \mathbf{M} is the isolated eigenvalue $\lambda = \mu N_C - 1/\tau + O(\sigma)$, one finds that the dominant contribution to λ' equals either $\mu N_C - 1/\tau$ or $\mu N_C - 1/\tau - 2\kappa$. The corresponding dominant eigenvectors have respectively uniform components across both compartments: $\mathbf{u}' = (1, \dots, 1)^T$ or opposite components on each compartment: $\mathbf{v}' = (1, \dots, 1, -1, \dots, -1)^T$. If opposite deviations in each compartment emerge, the synchronization towards a global homochiral state occurs when the contribution of \mathbf{u}' wins over that of \mathbf{v}' on long times. For weak couplings $\kappa \rightarrow 0$, the contribution of the top two eigenvalues are similar and in the vicinity of the racemic state, the vector $(\delta \mathbf{x}_{0,1}, \delta \mathbf{x}_{0,2})^T$ is aligned with a linear combination of eigenvectors \mathbf{u}' and \mathbf{v}' , depending on the initial deviation. However, when κ increases, the gap between the two eigenvalues grows, and the contribution of the eigenvector \mathbf{u}' becomes strongly dominant, indicating a synchronization of the deviations from the racemic state between the two compartments, thus a synchronization of the directions of homochirality in both compartments. This is observed numerically with the Generalized Frank model. When diffusion is weak for $\kappa \rightarrow 0$, we recover the previous scenario for a transition to homochirality, separately holding in each compartment in a well-mixed system. 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 (favoring L-species for instance), then one ends up with a global homochiral state (which is L-oriented in that case). The interesting case is therefore when the two compartments are initially given opposite small biases. Then, as shown in Figure 2.14, 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.

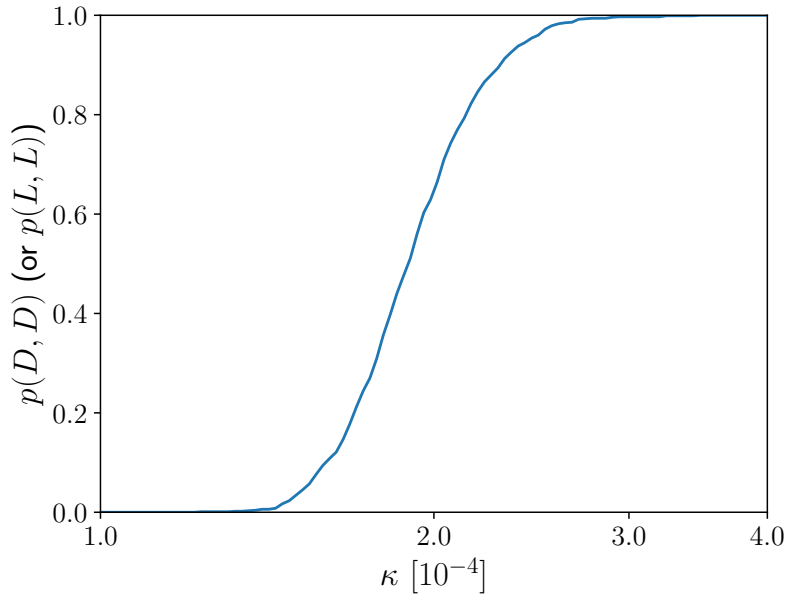


Figure 2.14: Probability to have two homochiral compartments of the same chirality on long times (either both are D or both are L) as function of the transition probability of transfer of molecules between the two compartments κ . The initial condition is such that there is an opposite chiral bias in 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 inactivated 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. The number of chiral species was set up to $N_C = 20$.

Template-directed ligation: an ideal candidate for homochirality
emergence?

Contents

3.1	Random polymerization	72
3.1.1	A simple model for polymerization	72
3.1.2	The Gillespie algorithm	73
3.1.3	The incapacity of step-growth polymerization to produce a ho- mochiral system under closed conditions	75
3.2	Template-directed ligation	77
3.2.1	Hybridization and dehybridization	77
3.2.2	Ligation and stalling	80
3.2.3	Hydrolysis	81
3.2.4	Temperature cycles	82
3.2.5	Racemization reactions	82
3.2.6	Previous works	82
3.3	Closed reactor	83
3.3.1	Quantities of interest	84
3.3.2	Irrelevance of block energy model details	85
3.3.3	System without racemization	85

3.3.4	Simulations with racemization	86
3.3.5	Effects of the racemization reaction speed	88
3.3.6	Effects of temperature cycles	89
3.3.7	Effects of the chiral stalling amplitude	90
3.4	Open reactor	91
3.4.1	Racemic chemostat	91
3.4.2	Templated ligation with a biased chemostat	92
3.5	Discussion	94
3.A	Comparison of block energy models	97
3.B	Single trajectories for closed reactors	98
3.C	Single trajectories for open reactors	99

In this section, we study how homochirality could emerge through template-directed polymerization of nucleic acids like RNA or prebiotic equivalents. In the case of RNA, template-directed synthesis of polymers consists in the extension of an RNA polymer by ligation with another polymer or monomer while being attached to a template strand through base pairing. In modern biology, as stated in the introduction, riboses and deoxyriboses, hence all nucleotides are D-oriented. Based upon experimental evidences that template-directed synthesis of RNA admits some degree of chiroselectivity, we explore here the propensity of a racemic pool of nucleotides to converge toward a homochiral pool of RNA strands in the presence, or not, of racemization reactions. Two kind of systems and their effects are studied in this chapter: (i) closed systems with a conserved number of RNA monomers and (ii) open reactors where species are being degraded over time and some are being chemostated.

3.1 Random polymerization

3.1.1 A simple model for polymerization

RNA strands are chains of ribonucleotides attached to each other through phosphodiester bonds, which are covalent bonds between a phosphate group and a carbon of two different nucleotides. Ribonucleotides are themselves composed of three different structures: a nucleobase which can be either a purine (adenine or guanine) or a pyrimidine (cytosine or uracile), a ribose (*i.e.* a 5 carbon simple sugar) and a phosphate group. Ribonucleotides get their chirality from the ribose which contains three different stereocenters. Recent

studies established prebiotic plausible pathways for the synthesis of purines and pyrimidine [Powner et al., 2009, Becker et al., 2019].

We first consider a system in which ribonucleotides undergo *step-growth* polymerization as depicted in Figure 3.1.a, *i.e.* template free polymerization, to see if this simple mechanism can already exhibit any transition to a homochiral state. The system is initially composed of two enantiomeric families of activated monomers that go through polymerization events with a rate constant k_{lig} , therefore it is a binary system where we do not consider any chemical differences between the four standard ribonucleotides AGCU but their two possible chiralities. The two possible monomers are therefore denoted D and L, referring to their respective chirality. Those strands can be cleaved by hydrolysis with a rate constant k_{cut} for *every phosphodiester bond* in the polymer (*i.e.* a strand of length l has the rate $(l - 1)k_{\text{cut}}$ to be cleaved at one of its bonds). This is described by the following mechanism



where S_i is a strand of length i , with $i, j > 0$. There is no evidence for a chiroselective behavior during step-growth polymerization of RNA. Therefore, RNA polymerization from a racemic pool of monomers produces strands composed of nucleotides of a random chirality. However, it has been shown experimentally that this process become chiroselective through specific chiral ribozyme catalysis or when performed on a mineral surface such as montmorillonite clays [Joshi et al., 2006, Joshi et al., 2011]. In this case, the addition of a monomer with a different chirality than at the 3' end of the extended strand is slowed by a factor $\alpha < 1$, *i.e.* the ligation rate becomes $k'_{\text{lig}} = \alpha k_{\text{lig}}$ as depicted in Figure 3.1.c.

3.1.2 The Gillespie algorithm

Deterministic rate equations, often used to model biochemical system are in fact only valid under the assumption that the system contains a very large number of molecules of each species (or in other words, it works for large concentrations). When modelling biochemical systems with low number of particles, one cannot neglect the stochasticity of the collisions and thus the chemical reaction process become stochastic. In this case, deterministic equations give a poor description of the dynamics of the system. One should then solve the dynamics using the Gillespie algorithm [Gillespie, 1976] which is an exact implementation of the master equation. This stochastic algorithm, introduced by Gillespie in 1977 [Gillespie, 1977, Gillespie, 2007] allows to simulate chemical systems with a few reactants since every reaction involving two species is explicitly simulated. The direct Gillespie algorithm works as follow:

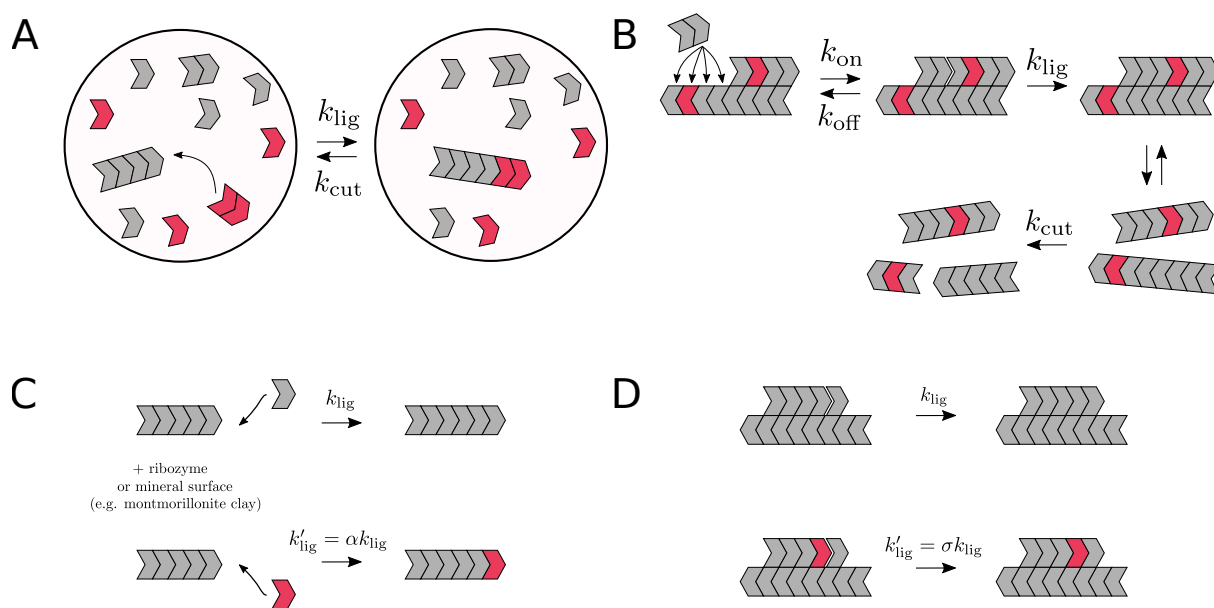


Figure 3.1: Polymerization schemes. Grey and red colors represent the two D and L chiralities. (A) Step-growth polymerization of a single-stranded polymer in a chiral closed system. (B) Template-directed ligation of activated nucleotides. The arrow with four ends depicts the four possibilities for the dimer to bind to the longer template. (C) Step-growth polymerization of a single-stranded polymer is slowed during the incorporation of a monomer of the wrong chirality in the presence of a chiroselective ribozyme or a mineral surface such as Montmorillonite clay. (D) The ligation of two strands on a template is stalled if, at the ligation site, there is a chirality mismatch between one of the two ligated strands and the template.

1. Set the initial copy number for all species and set time to $t = 0$.
2. Compute the propensity for all reactions of the system, and the total propensity of the system (it is the sum of propensities for all reactions in the system). The propensity of a reaction is a rescaled rate, whose scaling depends on its molecularity, *i.e.* whether the reaction is monomolecular (an outflow or a cleavage) or bimolecular (*e.g.* $A + B \rightarrow C$) reaction. The propensity for a given reaction controls its probability to occur between two Gillespie steps.
3. Draw dt the time interval during which a reaction will occur, from an exponential distribution, whose mean is the inverse of the sum of propensities.
4. Draw the next reaction that occurs in the system at random. Each possible reaction has a chance proportional to its propensity to be the chosen one.
5. Update the number of particles depending on which reaction occurred in the system, and update the new simulation time $t = t + dt$. The simulation stops when it reaches the maximum time limit.

3.1.3 The incapacity of step-growth polymerization to produce a homochiral system under closed conditions

Numerical simulations of the previously described system were performed using a Gillespie algorithm coded in C++, considering the case of a closed reactor. The distribution of homochiral strands as a function of the length of the strand produced in the reactor shows a deviation from the random case in the presence of a chiroselective enzyme (see Figure 3.2a). With no chiroselection, each monomer within a strand has an equal probability to be of D- or L-chirality. Therefore the probability for a strand of a given length k to be homochiral is 2^{1-k} because there are two possibilities for homochiral strands (either full D or full L). The system reaches a state where all strands taken individually are homochiral in the case of perfect chiral selection ($\alpha \rightarrow 0$), while the pool remains generally racemic. If we now introduce a racemization reaction (racemization reactions will be specified in Section 3.2.5) at the *free monomer* level, which is here assumed very fast (the pool at the free monomer level is racemic at any time), we observe in Figure 3.2.b a bias in the enantiomeric excess (*e.e.*) at the monomer level (all monomers of the system, polymerized and non-polymerized ones). The enantiomeric excess is defined here as

$$e.e. = \frac{|D_m - L_m|}{D_m + L_m}, \quad (3.2)$$

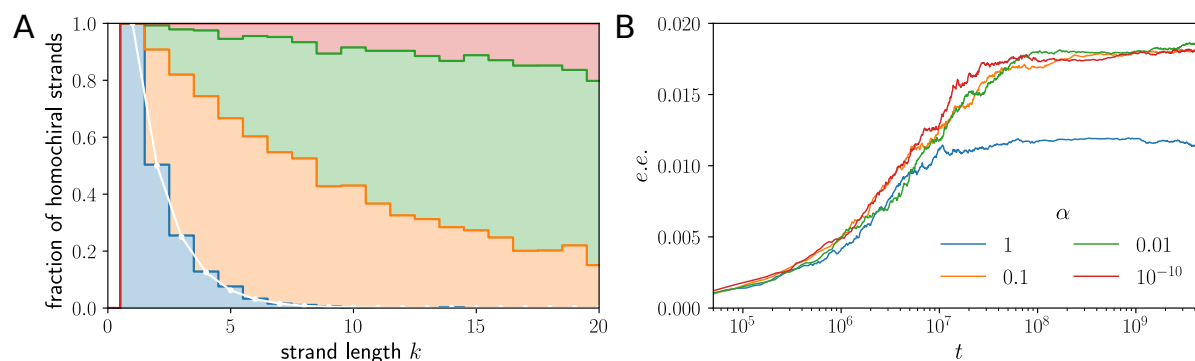


Figure 3.2: Stochastic simulations of step-growth polymerization. (A) Fraction of homochiral strands as a function of their length k in the steady state, without any racemization reactions. The white curve is the theoretical prediction for unbiased homochiral and heterochiral ligation. In this case, the fraction of homochiral strands is 2^{1-k} . (B) Time evolution of the enantiomeric excess at the monomer level. The simulations were performed with instantaneous racemization reactions for monomers that are not polymerized. α is the ligation chiral bias $\alpha = k'_{\text{lig}}/k_{\text{lig}}$. Data are averaged over 100 independent numerical realizations starting from an initial racemic pool composed of 2500 D- and 2500 L-monomers. Parameter set is : $k_{\text{lig}} = 5.27 \times 10^{-9}$, $k_{\text{cut}} = 10^{-10}$ and $V = 4.69 \times 10^{-4} \mu\text{m}^3$.

where D_m (*resp.* L_m) is the total number of D-monomers (*resp.* L-monomers) in the system. We observe convergence to a non-racemic steady-state but at a small *e.e.* without chiroselective bias (*e.e.* $\sim 1\%$). With an effective chiral selection, the system ends up in a steady state with a slightly larger *e.e.* but which hardly go beyond 2% even with perfect chiral selection. In this case, this increase of *e.e.* depends on the chiroselective effect but saturates for value of $\alpha > 0.1$. Also, if the catalyst would disappear from the system at some point, the *e.e.* would relax to the non-biased steady-state value.

Eventhough catalyzed step-growth polymerization can favor homochiral strands, no significant imbalance appears in the system due to the lack of autocatalytic dynamics, and further production of biased homochiral strands relies entirely on an external factor such as chiral enzymes or mineral surfaces. Here, even though it is closed, the system never reaches a racemic equilibrium because racemization drives it out of equilibrium. Equivalently, racemization breaks the conservation law between D- and L-monomers. Basically, even with chiroselection and racemization reactions, simple RNA step-growth fails to produce any significant *e.e.* in a closed system. The main reason here is that no autocatalytic behavior emerges from step-growth polymerization that would select only one chirality and amplify it, at the expense of the other.

3.2 Template-directed ligation

We now consider a chemical system where, starting from a pool of nucleotides, template-directed ligation involving RNA polymers and monomers occurs, as represented in Figure 3.1.b. RNA strands have the ability to attach to each other through base pairing, a mechanism in which face to face nucleobases contained in two attached RNA strands link through hydrogen bonding, resulting in a RNA duplex. For simplicity we once again consider a binary model composed of two ribonucleotides of different chiralities D and L, which can self-pair. One of the requirements for ligation to occur is that the phosphate group of nucleotides is activated, as the formation of a covalent bond during ligation requires energy provided by the environment. Thus, an activation chemistry is assumed to be supplied in a sufficient amount to assume that all nucleotides are always activated (example of activation chemistry are given in [Kervio et al., 2016, Walton and Szostak, 2017, Sosson and Richert, 2018]). Activation reactions are not modeled explicitly in this work.

3.2.1 Hybridization and dehybridization

Hybridization and dehybridization of two strands are assumed to be single-step processes. Two complexes collide with a rate k_{coll} and hybridize with equal probabilities for all the possible binding configurations. The rates of those two reactions, namely k_{on} and k_{off} , are constrained by the thermodynamic relation

$$\frac{k_{\text{off}}}{k_{\text{on}}} = (V N_A c^\circ) \cdot \exp\left(\frac{\Delta G_{\text{hyb}}}{k_B T}\right), \quad (3.3)$$

where V is the reactor volume, N_A the Avogadro number, $c^\circ = 1 \text{ mol.L}^{-1}$ the reference concentration, k_B the Boltzmann constant, T the temperature, here fixed at $T = 37^\circ\text{C}$, and ΔG_{hyb} the hybridization energy associated with the part of sequence which is hybridized. The hybridization rate k_{on} between two strands for one specific combination is related to the constant collision rate k_{coll} through the relation

$$k_{\text{on}} = \frac{1}{\theta} k_{\text{coll}}, \quad (3.4)$$

where θ is the number of possible duplex combinations they can form. Two strands of length l_1 and l_2 have $\theta = l_1 + l_2 - 1$ possible ways to attach to each other through base pairing. This expression changes if one of the strands is already part of a complex, but the total rate for the hybridization of two strands still sum up to k_{coll} . The computation

of ΔG_{hyb} depends of the choice of an energy model for the stability of complexes. In the first place, it is important to note that ΔG_{hyb} does not just depend on the length of the hybridized part within a complex. As base pair mismatches in DNA and RNA destabilize a complex, chiral mismatches also destabilize [Damha et al., 1991, Damha et al., 1994, Urata et al., 2003, Kawakami et al., 2005, Hauser et al., 2006, Szabat et al., 2016] heterochiral DNA and RNA complexes.

Naive pairwise energy model

A straightforward way to construct the hybridization energy of a complex is to sum up contributions to hybridization energy for all base pairs in the duplex (a duplex being a complex composed of two hybridized strands). We consider in a first approximation that base pairs consisting of two monomers of the same chirality in a complex (*i.e.* homochiral base pairs) contribute negatively to the free energy (hence stabilizing the complex). In contrast, heterochiral base pairs interact in a way that destabilizes the complex and contributes positively to the hybridization energy (see Fig 3.3). The hybridization energy

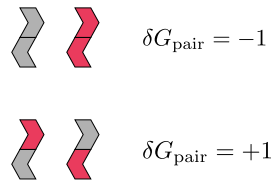
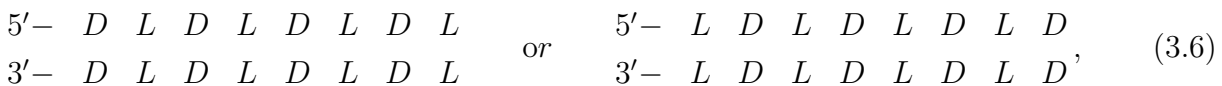


Figure 3.3: The four different possible base pairs.

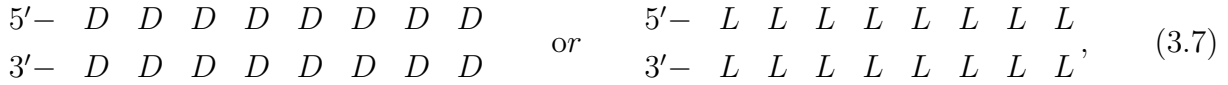
of a given complex would result in the sum over all of its base pairs

$$\Delta G_{\text{hyb}} = \sum_{i \in \{\text{pairs}\}} \delta G_i^p, \quad (3.5)$$

where δG_i^p is the energy contribution of the i -th base pair in the complex. Even though this energy model could produce interesting results, it does not properly describe how heterochiral internal bonding affects complexes' stability. It is crucial to consider neighbor interactions within each complex strand [Liu, 2020] as it will disrupt the helix rotation of the complex. For instance, with a base-pair model, a purely alternating duplex with homochiral base pairs



would have the same stability as the associated fully homochiral duplex



which has been disproved experimentally [Hauser et al., 2006] for DNA for instance. Therefore a block-wise model instead of a pair-wise energy model [Turner and Mathews, 2009] appears more relevant for our study.

Block-wise energy model

The block-wise energy model considers the contributions of blocks of two base pairs rather than individual base pairs. It accounts for base pairs and lateral interactions within each strand, when a duplex structure is formed during hybridization. The total hybridization energy is summed up over all blocks in the complex

$$\Delta G_{\text{hyb}} = \sum_{i \in \{\text{blocks}\}} \delta G_i^b, \quad (3.8)$$

where δG_i^b is the free energy contribution of the i -th block. Turner and SantaLucia [SantaLucia and Hicks, 2004, Turner and Mathews, 2009] already measured all hybridization block energy contributions for every possible base pair combination for D-DNA or D-RNA in a duplex. The same values would be measured in theory for L-DNA and L-RNA, but unfortunately, block energies have not been measured extensively for DNA/RNA complexes of mixed chiralities. One can however circumvent the lack of experimental data by building a minimal model to describe heterochiral interactions within a block of four monomers (see Figure 3.4). We denote $\delta G_1 = E < 0$ the hybridization block energy associated with a fully homochiral block (the other hybridization block energies are defined graphically on Figure 3.4). Averaging over all blocks from the 2004 Turner library (for RNA), we set $E = -1.59 \text{ kcal.mol}^{-1}$. We denote u the vertical penalty associated with a mismatch of chiralities in a base pair, and v is the lateral penalty associated with heterochiral patterns in one of the strands of the complex (u and v are positive quantities). The block energies are thus ordered in amplitude as follows

$$\delta G_1 < \left\{ \begin{array}{l} \delta G_2 \\ \delta G_3 \end{array} \right\} \leq \delta G_5 \leq \left\{ \begin{array}{l} \delta G_3 \\ \delta G_2 \end{array} \right\} < \delta G_4. \quad (3.9)$$

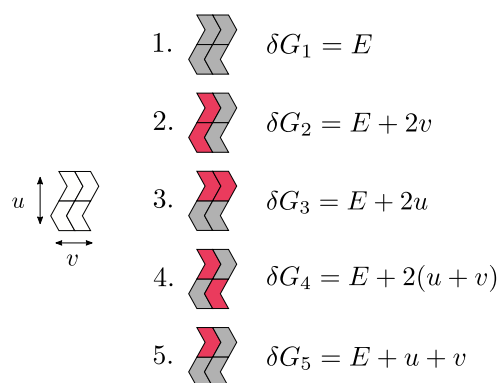


Figure 3.4: The five different kinds of 4-monomer blocks and their associated free energies as a function of the basal free energy of a fully homochiral block, E , and the vertical and lateral penalties u and v due to chiral mismatching or heterochiral patterns within the two distinct strands.

The double diagonal mismatch destabilizes the most the complex due to an incorrect base pairing and incompatible rotational twists in the duplex that is supposed to form a helix. We differentiate three different cases depending on the relative values of u and v (namely $u < v$, $v < u$, $u \sim v$) describing the strength of a lateral or a vertical mismatch penalty. From the experimental literature [Urata et al., 2003], we finally deduce an approximate value for $u + v$ amplitude, $u + v \sim 1.26 \text{ kcal.mol}^{-1}$. The three different couples of values for (u, v) will be 1. $(0.42, 0.84) \text{ kcal.mol}^{-1}$, 2. $(0.84, 0.42) \text{ kcal.mol}^{-1}$ and 3. $(0.63, 0.63) \text{ kcal.mol}^{-1}$. For the case of terminal blocks formed out of three monomers instead of four, when the end of a strand does not coincide with the end of the template in a complex, we use the same model but with a different energy for the homochiral three-monomer block: $E_{3\text{-block}} = -0.47 \text{ kcal.mol}^{-1}$ (this value is also the result of an average of the data from the 2004 Turner library for RNA). In the simulations, all the block energies and penalties are rescaled by a factor 0.49 for optimization purposes.

3.2.2 Ligation and stalling

In this model, we neglect *random ligation*, *i.e.* ligation occurring without templates, as it is slower than templated one. Two strands can ligate on a template if they share a common ligation site (see Figure 3.1b), with the rate constant k_{lig} . During the template-directed ligation of RNA, the ligation speed depends on the chiral configuration near the ligation site. Experiments have shown [Joyce et al., 1984, Bolli et al., 1997] that the ligation of two strands on a template is slowed when the chirality of the nucleotide at the ligated end of a strand does not match the chirality of the template (see Figure 3.1.d). The amplitude of the stalling slows the ligation from one up to two orders of magnitude. Indeed,

in 1984, a key experiment [Joyce et al., 1984] where G-ribonucleotides were polymerized onto a poly(C_D) template (an oligomer composed of D-cytosine nucleobases) showed such a chiroselective behaviour. Joyce *et al.* noted first that G_D (which are activated D-guanosine nucleobases) were incorporated more readily during templating onto the poly(C_D) oligomer than G_L ribonucleotides, and that G_L acted as chain terminator that prevents further extension of the elongated poly(G) strand, describing both a chiral selection and a perfect chiral cross-inhibition. They still observed G_L ribonucleotides at non-terminal positions indicating that there is still the possibility to incorporate G_D nucleotides after a chiral mismatch. The cross-inhibition is thus strong, but definitely not perfect. In 1997, other experiments [Bolli et al., 1997] were performed using pyranosyl-RNA instead of furanosyl-RNA (which result in a 4'-2' phosphodiester bonding between ribonucleotides of a given polymer rather than 5'-3' bonding) for templated ligation. It could be observed that the ligation speed was reduced to 10% its basal speed (*i.e.* when no chiral mismatches are present at the ligation site) when a chiral mismatch involving a purine was present at a ligation site. For chiral mismatches involving pyrimidines, the ligation speed was reduced up to two order of magnitudes. All in all, when chiral mismatches between a template and an elongated strand appears at a ligation site, the ligation is stalled. For homochiral binding sites, the ligation occurs at its basal rate, $k_{\text{lig}} = \lambda$. Therefore, the ligation rate for a given complex reads

$$k_{\text{lig}} = \begin{cases} \lambda \sigma^2 & \text{if } C_S^{-1} \neq C_T^{-1} \text{ and } C_S^{+1} \neq C_T^{+1} \\ \lambda \sigma & \text{if } C_S^{\pm 1} \neq C_T^{\pm 1} \text{ and } C_S^{\mp 1} = C_T^{\mp 1} \\ \lambda & \text{otherwise.} \end{cases} \quad (3.10)$$

where C_S^i (*resp.* C_T^i) describe the chirality of the i -th monomer of a strand participating in a ligation event (*resp.* the template strand), which can be either D or L and $\sigma \leq 1$ the stalling factor which we introduce here to quantify the effect of the stalling observed in [Joyce et al., 1984, Bolli et al., 1997]. Monomers on both sides of the ligation site are labeled -1 and $+1$. This phenomenon induces a powerful chiroselective effect during template-assisted RNA ligation that will reduce chirality errors during its replication.

3.2.3 Hydrolysis

RNA strands can be cleaved by hydrolysis, during which a phosphodiester bond of their backbone is broken. The cleavage rate is denoted k_{cut} and is the same for all phosphodiester bonds in the system. Therefore any strand of length l is susceptible to be cleaved at one of its junctions with a rate $(l - 1)k_{\text{cut}}$ resulting in two shorter strands. However

we note that hydrolysis does not occur when strands are part of a duplex.

3.2.4 Temperature cycles

From the thermodynamic relation of equation (3.3), it appears that the hybridization energy of long complexes becomes arbitrarily large in negative values and results in frozen complexes that do not dehybridize in a finite time. To circumvent this well known issue, known as product inhibition, which would lead to frozen dynamics in which all strands would be part of highly stable complexes, we impose thermal cycling to the system that melts the complexes periodically [Mast and Braun, 2010, Salditt et al., 2020]. Instantaneous temperature peaks with a given frequency $1/\tau_{\text{cycle}}$ periodically split all complexes apart. We assume here that these temperature changes are quick enough not to affect the other processes in the system, eventhough, in theory, all chemical process should be affected by the temperature change. Another possibility to force the melting of the complexes would be pH changes or increases of salts concentration [Tkachenko and Maslov, 2015, Keil et al., 2017, Mariani et al., 2018, Ianeselli et al., 2019]. The dynamics of these cycles is not described in details in the model but it is essential to solve the issue of product inhibition.

3.2.5 Racemization reactions

In simulations where it is explicitly stated, we will consider that a racemization reaction occurs in the system. We assume that this racemization occurs only for unpolymerized monomers and that it is a one-step process that can convert D-monomers into L-monomers and *vice versa*. Experimentally, racemization could occur as the result of the production of achiral free radicals from riboses' chiral centers when exposed to UV light for instance.

3.2.6 Previous works

Two different works [Tupper et al., 2017, Chen and Ma, 2020] studied so far templated replication of RNA as a proxy for homochirality emergence. In 2017, Tupper *et al.*, performed numerical simulations of a pool of monomers undergoing templated polymerization. They assumed that RNA polymers could elongate through step-growth polymerization, be cleaved by hydrolysis and also could ligate once bounded on a template. The two short strands that would hybridize onto the template should have a uniform sequence and an exact chiral compatibility with the template. In other words, no chiral mismatches were allowed during hybridization. There was also no description of the sta-

bility of duplexes formed through RNA hybridization as no description of duplexes was really carried out. In this context, they assumed that duplexes would dehybridize directly after the ligation of the two elongated strands. Note that we do not seek to judge the decisions taken by the authors here. They themselves stated in their work that they aimed for a simple model to describe template-directed ligation, and we will see that such a *simple* description could already give satisfying results. Considering only closed reactors, a racemization reaction could interconvert single monomers into the other chirality. With this set of assumptions, they could observe a symmetry breaking in the dynamics of the enantiomeric excess for sufficiently large values of the ligation rate between two strands sitting next to each other on a template. In 2020, Chen and Ma also performed numerical simulations of a similar ribonucleic system. Here only monomers of the same chirality as the template could be incorporated into the extended hybridized chain. This assumption was further relaxed in their work but still, the incorporation of a monomer from the opposite chirality acted as a terminator of the template-directed synthesis. As for Tupper *et al.* their model does not propose any thermodynamic description of the RNA duplexes formed by hybridization. They were also able to show with their model that such a closed system undergoes a symmetry breaking leading to the convergence of the enantiomeric excess to values close to one in presence of racemization reactions for single monomers. The main differences with our present model is first, the thermodynamic description of RNA hybridization we propose and that we allow any feasible hybridization, and second, the fact that chiral mismatches are possible and do not formally stop the ligation process but only stall it as observed experimentally. Also we come up with a physical explanation for RNA duplex melting with periodic heat peaks that separate all the RNA complexes.

3.3 Closed reactor

We focus now on the results of the model of template-directed ligation described in Section 3.2. The model is also numerically simulated with a Gillespie algorithm in a well-mixed system using a modified version of the numerical toolbox developed in C++ by Gerland's group in Munich [Rosenberger *et al.*, 2021, Göppel *et al.*, 2022]. The system is initially inoculated with a total of 5,000 nucleotides distributed in 4,920 monomers (2,460 D- and 2,460 L-monomers) and 40 dimers (10 DD-, 10 LL-, 10 DL- and 10 LD-dimers) in a racemic fashion. Chiral biases will emerge naturally in those stochastic simulations. All parameters used for the simulations are given in Table 3.1.

reaction type	parameter	expression	value (in k_{coll} units)
collision	k_{coll}	1	1
hybridization	k_{on}	$\frac{1}{\theta}k_{\text{coll}}$	
ligation	k_{lig}		1.5×10^{-4}
temperature cycle	$1/\tau_{\text{cycle}}$	$k_{\text{coll}} \cdot \exp(\delta G_1 L_{\text{cycle}})$	1.8×10^{-3}
racemization	k_{rac}		
hydrolysis	k_{cut}		10^{-10}
outflow	k_{out}		10^{-8}
chiral inhibition	σ		[0.05, 1]

Table 3.1: Parameters set used in the simulations. Parameters are specified if different values are used. Note that σ has no units. τ is parametrized with L_{cycle} which is defined in Section 3.3.6 in equation (3.14).

3.3.1 Quantities of interest

Without racemization of the system at the free monomer level, the two enantiomeric populations remain constant over time. As a consequence, there is no interest in looking at the overall *e.e.* for such systems. Therefore, we introduce the chirality parameter χ , which describes the chirality of a given strand

$$\chi = 2f_D - 1, \quad (3.11)$$

where f_D is the fraction of D-monomers in a given strand. With this, $\chi = 0$ describes a strand with half its monomers being D-chiral and the second half, L-chiral (the *e.e.* of such a strand equals 0), and $\chi = 1$ characterizes a homochiral strand of D-monomers while $\chi = -1$ a homochiral strand of L-monomers. We also introduce a second parameter, that is, the fraction of homochiral motifs in the strand i , ξ_i such that

$$\xi_i = \frac{h_i}{l_i - 1}, \quad (3.12)$$

where h_i is the number of homochiral motifs (DD or LL) in the strand i of length l_i . The fraction of homochiral motifs in the system Ξ thus reads

$$\Xi = \frac{\sum_i \xi_i (l_i - 1)}{\sum_i (l_i - 1)}. \quad (3.13)$$

A system composed of strands with only homochiral motifs will be characterized by a parameter $\Xi = 1$ and $\Xi = 0$ in the case of heterochiral motifs only.

3.3.2 Irrelevance of block energy model details

In the first place, simulations data indicate that there is no fundamental difference of dynamics between the three different parameter sets for the block energy model ($u < v, v < u$ or $u \sim v$), assuming $u > 0$ and $v > 0$. For every quantity studied here, the dynamics are similar (see Appendix 3.A), indicating the details of the block energy model do not matter as long as there is a stability difference between fully homochiral blocks and heterochiral ones. The simulations presented throughout this section will be carried out in the case $u \sim v$ from now on (meaning that base pair mismatches in chirality destabilize duplex formation as much as chirality alternation inside a strand during hybridization).

3.3.3 System without racemization

Without racemization reactions, the system reaches a steady state after a transient regime of exponential growth (Figure 3.5a) which is necessarily racemic regarding its general composition. The number of monomers of both chiralities is conserved over time. However, patterns can still emerge, and we could end up with a system composed of two mirrored populations of homochiral strands. Looking at Figure 3.5b, we observe in the case with no kinetic stalling ($\sigma = 1$) that there seems to be almost no bias in the steady state of the system (the fraction of homochiral motif Ξ has a value close to 0.55 which describes a system with almost as many heterochiral motifs than homochiral ones). Even though the bias emerging from the thermodynamics of RNA hybridization is only mild, when we include the stalling in the simulation ($\sigma = 0.05$ here), we observe a final steady state that is strongly biased and contains an important fraction of homochiral motifs ($\Xi \sim 75\%$). This observation first demonstrates that the thermodynamical bias in the energy model cannot alone induce a significant bias in the final composition of the system. It needs the presence of kinetic stalling to do so (as illustrated and explained in detail in [Göppel et al., 2022]). As it will also be shown in the simulations with racemization, the stalling effect is a crucial component for homochirality emergence in such systems where template-directed ligation takes place.

The nonmonotonous shape of Ξ is explained by multiple-step growth in the system. There is first a slight growth phase where dimers will be elongated with dimers or monomers that will consume most of the free monomers initially present in the system. Thermodynamic discrimination is intense for complexes involving short strands (such as monomers and dimers), and therefore, mainly homochiral bonds will be formed during this phase. The fraction of homochiral motifs in the system, Ξ , reaches its maximal value. Once almost all monomers have been consumed in the system, the ligation will occur

with longer strands, and the average length of the system will significantly grow. As strands become longer, the thermodynamic discrimination of complexes involving chiral mismatches will be weakened by the long hybridized interface stabilizing complexes, and replication with mismatches is eased, increasing the number of heterochiral motifs in the system. Eventually, the system reaches a steady state ruled by ligation and hydrolysis events.

Looking at the distribution of the chiral component χ in the steady state (Figure 3.5c), we observe that it is mainly centered around $\chi = 0$. Two peaks indicate a significant number of homochiral strands (either D or L), but only short strands contribute to these peaks. Indeed, they are only present in the blue histogram that describes the whole system and disappear within the red histogram taking into account only strands that are 20–monomers long and more. Therefore, the natural thermodynamic bias plus the inherent stalling of template-directed ligation of RNA are not sufficient to produce a pool of homochiral polymers only, even though a significant portion of short strands presents a chiral bias with an *e.e.* $\geq 50\%$.

3.3.4 Simulations with racemization

We now include a fast racemization reaction in the simulation. At any time, the system is racemized at the monomer level. In this case, a second growth phase is observed for the average length when $\sigma = 0.05$, *i.e.* when we have an effective kinetic stalling during ligation events (Figure 3.5d). The mean length reaches first a transient plateau (corresponding to the steady-state of the experiment without racemization) and then undergoes a second growth phase towards a final stationary state. This second growth phase is explained by the increase of one enantiomer population while the other decreases. Thus more monomers are available for non-stalled ligation. Indeed, the *e.e.* (Figure 3.5f) increases significantly in the racemized experiment and eventually reaches *e.e.* $\sim 100\%$, which characterizes a complete homochiral system. While the system is converging towards a single chirality state, ligation events are less and less stalled since chiral mismatches at ligation sites are less likely to occur as the chiral homogeneity of the system grows. We believe that this second growth could have major relevance for the origins of life, as it will be discussed in Section 3.5. Comparing experiments with and without racemization, we observe that the case without chiral stalling ($\sigma = 1$) gives similar results. With racemization but no stalling, the system reaches a *e.e.* $\sim 4 - 5\%$ due only to the thermodynamic discrimination of heterochiral phosphodiester bonds or base pairs during strand hybridization. In the case of chiral stalling (observed experimentally by Bolli, Joyce and al. [Bolli

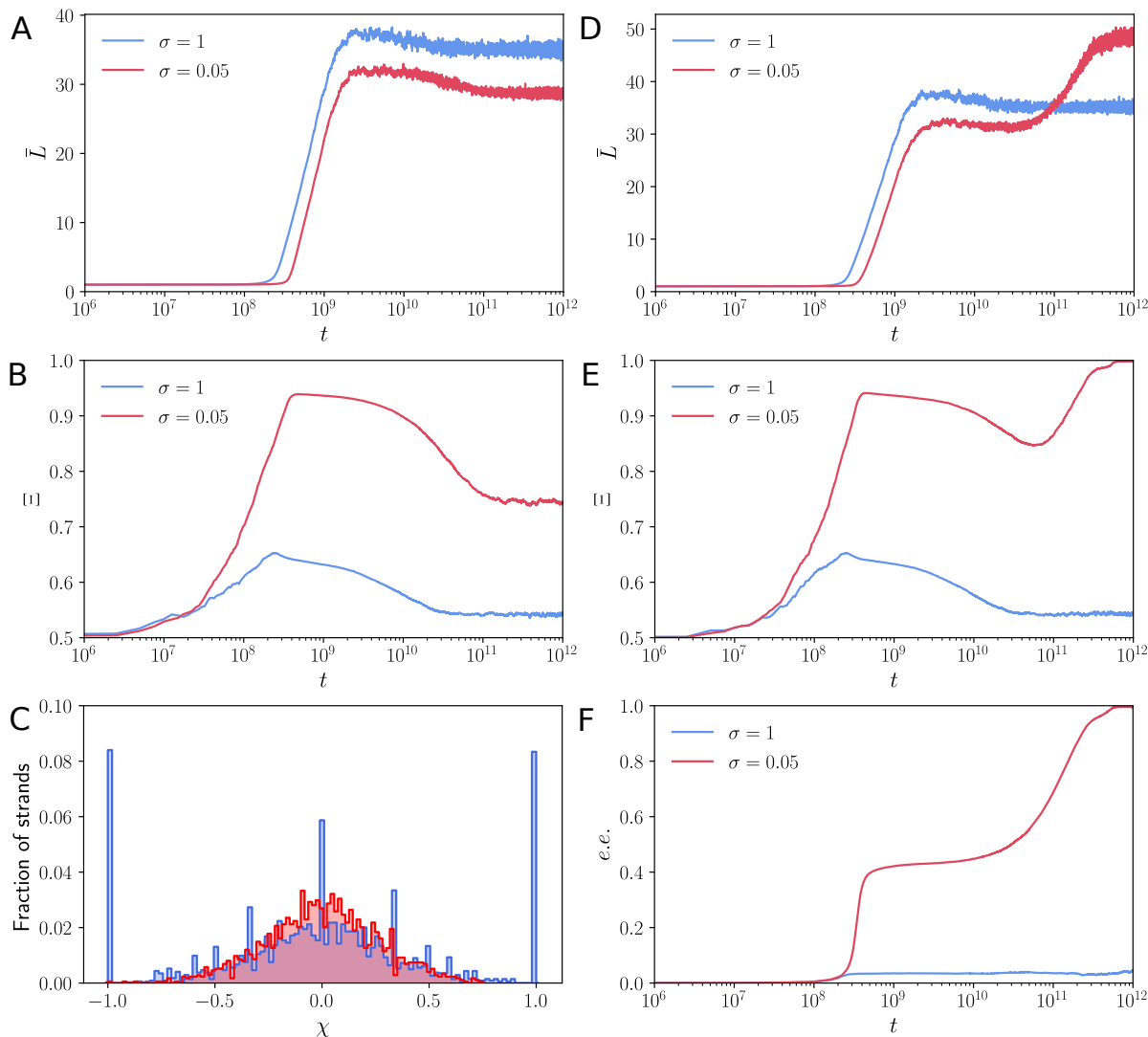


Figure 3.5: Simulations of template-directed ligation in a closed reactor. Panels (A), (B), and (C) are data from simulations performed without any racemization reaction, while panels (D), (E), and (F) are extracted from simulations with fast racemization. (A) Average strand length as a function of simulation time (without racemization). (B) Fraction of homochiral motifs in the system as a function of time (without racemization). (C) Histograms of the fraction of strands in the system as a function of the chiral parameter χ . This data is extracted from a single snapshot at a time step located in the steady state. Blue histogram considers all strands in the system. The red histogram only considers strands with a length $L > 20$ monomers. (D) Average strand length as a function of simulation time in simulations, including fast racemization. (E) Fraction of homochiral motifs in the system as a function of time (with racemization). (F) Enantiomeric excess (*e.e.*) at the monomer level (and not only the *free* monomers level) as a function of time. All data shown here are averaged over 20 independent realizations.

et al., 1997, Joyce et al., 1984]) adding racemization induces a convergence towards a total homochiral system. In this case, the fraction of homochiral motifs converges to $\Xi = 1$ (Figure 3.5e).

3.3.5 Effects of the racemization reaction speed

To ensure that the observed results are not specific to the particular case of instantaneous racemization for non-polymerized monomers, we investigated the effects of the racemization reaction speed on the *e.e.* dynamics. We model this reaction assuming mass-action law, *i.e.* with $k_{\text{rac}}D_1$ (*resp.* $k_{\text{rac}}L_1$) being the reaction rate to convert D-monomers into L-monomers (*resp.* L-monomers into D-monomers), where D_1 (*resp.* L_1) is the concentration of D-monomers (*resp.* L-monomers). Figure 3.6 shows first that the *e.e.* saturates to

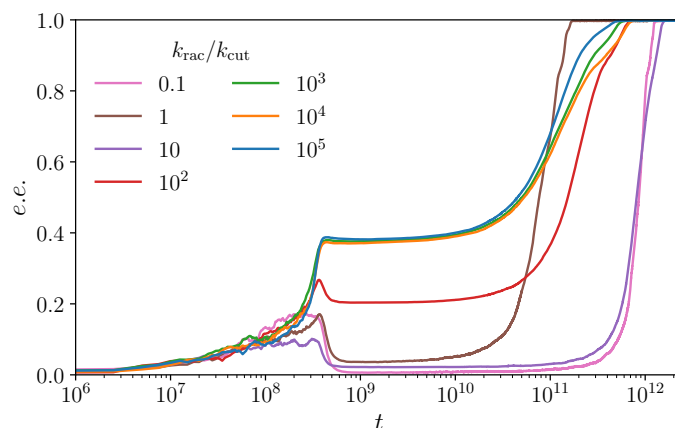


Figure 3.6: Enantiomeric excess as a function time in a template-directed ligating closed system in the presence of chiral stalling $\sigma = 0.05$. The enantiomeric excess is computed over all polymerized monomers, thus excluding non-polymerized monomers. Simulations were carried out with $k_{\text{cut}} = 10^{-10}$, however, for computational time purposes, brown and pink curves were simulated with $k_{\text{cut}} = 10^{-9}$ (which explains the difference in timescales for the transition to occur). Data shown here are averaged over 20 independent realizations starting from an initial racemic pool of 5,000 nucleotides.

a limit value when the racemization rate constant exceeds $k_{\text{rac}} = 10^{-7}$ and also that the system always converges towards a homochiral state, even when $k_{\text{rac}} < k_{\text{cut}}$, which means that the racemization reaction has the lowest rate constant of the system. However, we observe that the timescale for the system to converge increases as the racemization dampens, and it also affects the height of the transient plateau in the simulation. This makes sense as for rapid racemization, a slight imbalance of polymerized nucleotides appears stochastically and gets amplified, while the system is racemized at the non-polymerized

monomers level. When racemization slows down, non-polymerized monomers are less racemized, and strands are formed out of the chirality that dominates free monomers. Looking at the *e.e.* for polymerized monomers only, this effect tends to decrease the chiral asymmetry before racemization reactions start to occur.

3.3.6 Effects of temperature cycles

The influence of the period of temperature cycles on the dynamics of the systems has also been studied. Temperature cycles are characterized by the time period τ parametrized by the length L_{cycle} such that

$$\tau = \frac{1}{k_{\text{coll}}} \exp(-\delta G_1 L_{\text{cycle}}). \quad (3.14)$$

This choice of parametrization means that on average, duplexes formed by strands of length $l < L_{\text{cycle}}$ will have time to dehybridize between two temperature peaks while duplexes formed by strands of length $l > L_{\text{cycle}}$ will not and are for the most part, melted during the temperature elevation, *i.e.* the higher L_{cycle} , the longer the time for duplexes to be dehybridized. This phenomenon can be observed in Figure 3.7, where the dynamics

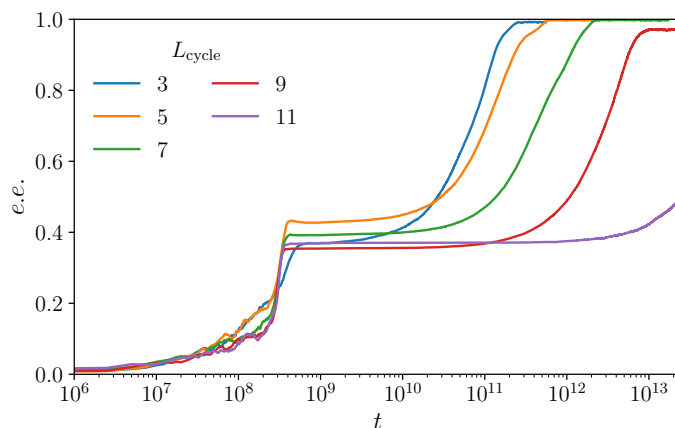


Figure 3.7: Enantiomeric excess as a function time in a template-directed ligating closed system in the presence of chiral stalling $\sigma = 0.05$ for different temperature cycles. The enantiomeric excess is computed over polymerized monomers only. Data shown here are averaged over 20 independent realizations starting from an initial racemic pool of 5,000 nucleotides with instantaneous racemization for non polymerized monomers.

spans over more orders of magnitude of time as the length L_{cycle} increases and thus the time τ between two temperature peaks, increase (note that $\delta G_1 < 0$). With the increase of τ the dynamics slows down but the system still end up in a homochiral state in the

end. However, we notice that for typical lengths $L_{\text{cycle}} > 7$, the degree of homochirality of the steady state starts to slightly decrease. This could be due to the fact that hybridized strands are protected against cleavage as hydrolysis is only possible for free RNA strands. This induces that remaining monomers of the disappearing chirality are protected against the conversion mechanism that would require first the hydrolysis of a strand resulting in the production of a free monomer that ends up converted to the other chirality, because strands remain hybridized for longer times as L_{cycles} increases. Eventually in the limit $L_{\text{cycle}} \rightarrow \infty$ and thus $\tau \rightarrow \infty$, the dynamics would end up frozen before any state with a significant *e.e.* would appear, as duplexes formed by long strands would be highly stable (see equation (3.3)) and the typical time to separate the duplex would be too large to see the system evolve on a reasonable timescale.

3.3.7 Effects of the chiral stalling amplitude

In this chapter, we generally used the reference value $\sigma = 0.05$ for the chiral stalling as it corresponds to a reduction between one and two order of magnitude of the ligation speed when chiral mismatches occur around ligation sites. We investigated the effect of changes in the chiral stalling amplitude and found out that it affects both the timescales of the dynamics and also the final stationary state. Indeed, Figure 3.8 shows that for

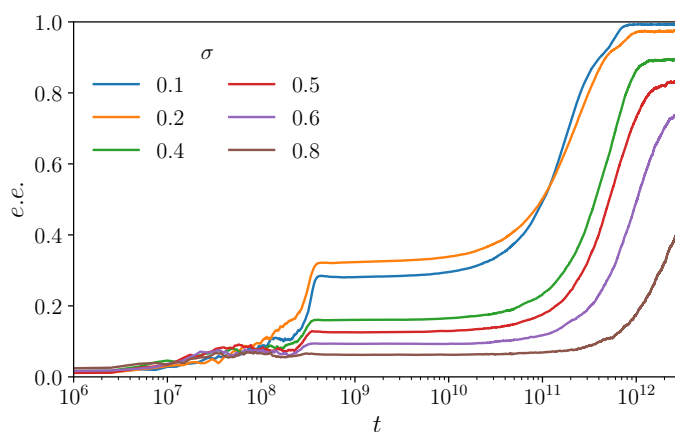


Figure 3.8: Enantiomeric excess as a function time in a template-directed ligating closed system for different kinetic stallings. The enantiomeric excess is computed over polymerized monomers only. Data shown here are averaged over 20 independent realizations starting from an initial racemic pool of 5,000 nucleotides with instantaneous racemization for non-polymerized monomers.

chiral stalling with values above 0.2, the system do not end up in a full homochiral state but into a partial one, with enantiomeric excesses that decreases as σ gets closer to 1.

Eventually, for $\sigma = 1$ the system reaches an almost racemic steady-state (see Figure 3.5f). Looking at the graph, it is unclear if these partially homochiral stationary states are really steady states or if it is just that the time required for the changes in the *e.e.* gets longer for high values of σ . However it is important to notice that the log scale compresses the right part of the plot associated with long times t . With a closer look, we observe that the simulations indeed reach steady states (except for $\sigma = 0.8$ which did not have the time to converge). In any case, we observe that there seems to be a threshold above which no significant enantiomeric excess can appear in a reasonable time. We also can be confident with the value $\sigma = 0.05$ used in our study as the stalling observed by Joyce and Bolli was strong in both cases. Also the case $\sigma = 0.2$ gives similar results to the ones found earlier, and it corresponds to a 80% reduction of the ligation speed.

3.4 Open reactor

3.4.1 Racemic chemostat

Open reactors have already been studied in the context of homochirality, as they can be easily driven out of equilibrium by an inflow of matter to avoid relaxation towards an equilibrium racemic state [Laurent et al., 2021]. Here we consider an open system chemostated with monomers and dimers, meaning their particle numbers remain constant over time, and where species can be depleted due to an outflow of rate ϕ . The chemostat is racemic, so the symmetry of the system is not broken explicitly. The system is inoculated with a total of 2,800 nucleotides distributed in 2,000 monomers (1,000 D- and 1,000 L-monomers) and 400 dimers (100 DD-, 100 LL-, 100 DL- and 100 LD-dimers) in a racemic fashion. In this experiment, we investigate whether we can free the system from the racemization constraint to reach a homochiral state. It appears first that the steady state reached by the system is not homochiral (Figure 3.9.a). In the case with chiral stalling, the final *e.e.* is low ($\sim 4 - 5\%$). The fraction of homochiral motifs reaches a stationary value of $\Xi \sim 75\%$; however, it does not mean much about the chiral parameter of a given strand. Indeed, a strand composed of 50% of D-monomers and 50% of L-monomers can have only one heterochiral motif while being perfectly heterochiral ($\chi = 0$). Initially, due to statistical fluctuations, slightly more shorts strands of one chirality will be produced, then they will start to produce more of their kind autocatalytically, and thus one of the two populations will reach longer and longer strands. This explains the increase in the *e.e.* observed in Figure 3.9.a. Unlike the closed reactor case, there is no conservation law for monomers that would slow the growth of the strands. Strands mainly composed of one

chirality will then quickly reach a steady state for their average length. As no racemization is present in the system, there is no real negative interaction between monomers of different chirality (except during hybridization and ligation), so the two populations are in some ways independent. Therefore, strands composed of the dominated chirality will start to grow, and their concentration will exhibit a similar but retarded growth phase as the dominant chirality. Eventually, the initially dominated population will catch up with the dominant one, inducing the vanishing of the enantiomeric excess. However, the fraction of homochiral motifs in the system remains high (Figure 3.9b) in the steady state. Looking at the distribution of the chiral parameter χ in Figure 3.9c, we see, as in the situation of the closed reactor without racemization reactions, that homochiral strands are mainly short strands. And that long strands are not that homochiral. However, if we focus on the transient behavior of the system, and more specifically at the time where the *e.e.* is at its maximum, the distribution is shifted towards homochiral strands composed of the dominant species in the system. Indeed, if we define the chiral parameter of the dominant species in a given strand, χ_E as

$$\chi_E = 2f_E - 1, \quad (3.15)$$

with f_E , the fraction of monomers of the dominant chirality (equals to f_D or f_L) in a strand, we see in Figure 3.9d that during the transient state where the *e.e.* is significant (*e.e.* $\sim 20\%$), a lot of long strands are built from almost only one chirality in the system.

3.4.2 Templated ligation with a biased chemostat

So far, we only studied initially racemic systems because a minute imbalance can arise from statistical fluctuations due to the stochasticity of the polymerization process. However, there is also the possibility that the initial system is not racemic in the first place if the synthesis of riboses, thus nucleotides, was not symmetric. It has been shown, for instance, that the synthesis of pentoses catalyzed by homochiral LL-dipeptides in prebiotic conditions is not symmetric and produced in particular a greater amount of D-ribose than its enantiomer [Pizzarello and Weber, 2004, Weber and Pizzarello, 2006, Pizzarello and Weber, 2010]. In a closed reactor, as demonstrated in Section 3.3, initially racemic systems converge towards a homochiral state. Hence there is no interest in investigating a biased system as it will favor one chirality in the homochiral state. In the case of the open reactor without racemization reactions, however, we saw that even though a transient state with a significant *e.e.* is achieved, the steady-state lies around the racemic state. Therefore, we simulated initially biased open reactors to study whether significant amplification of the small bias present in the chemostat is possible. It turns out, as one could expect that

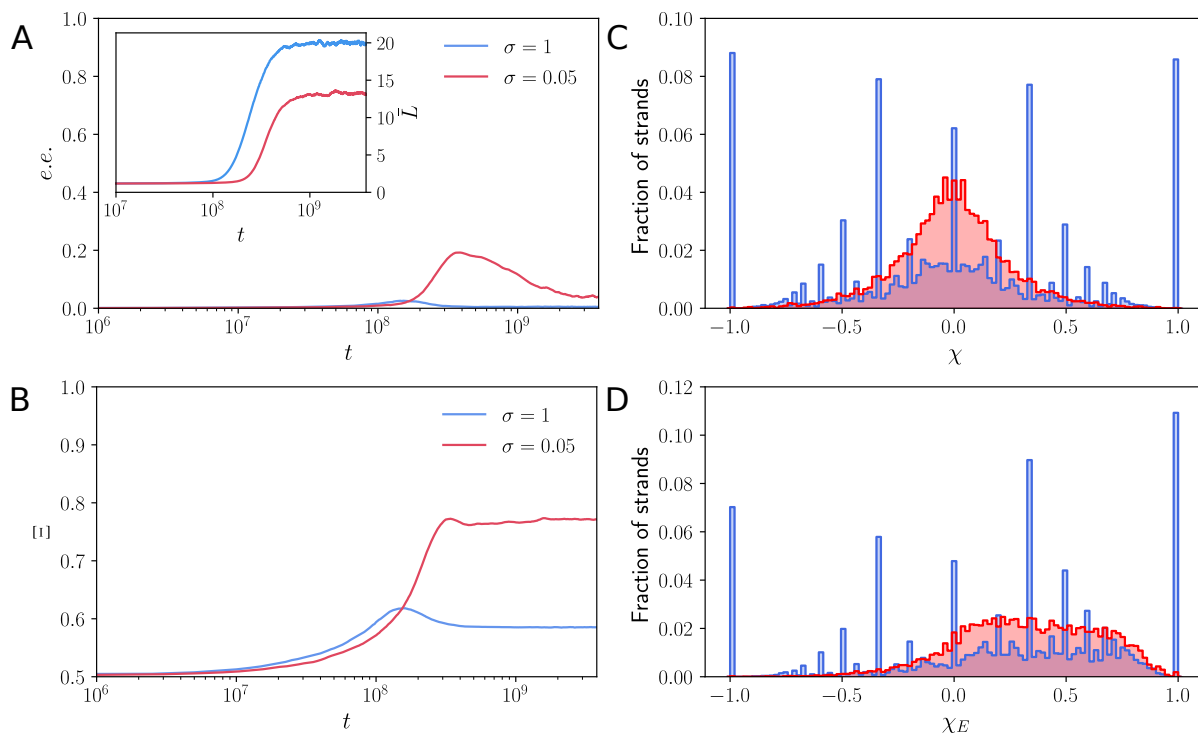


Figure 3.9: Template-directed ligation in a chemostated open reactor with outflow. Simulations are carried out without any racemization reaction. (A) $e.e.$ at the monomer level as a function of time. Inset: average length in the system over time. (B) Fraction of homochiral motifs Ξ in the system as a function of time. (C) Histograms of the fraction of strands in the system as a function of the chiral parameter χ . This data is extracted from a single snapshot at a time step located in the steady state. (D) Histograms of the fraction of strands in the system as a function of the chiral parameter χ_E . This data is extracted from a single snapshot at the time where for each realization, the $e.e.$ is at its maximum value. Blue histogram considers all strands in the system. The red histogram only considers strands with a length of $L > 20$ monomers.

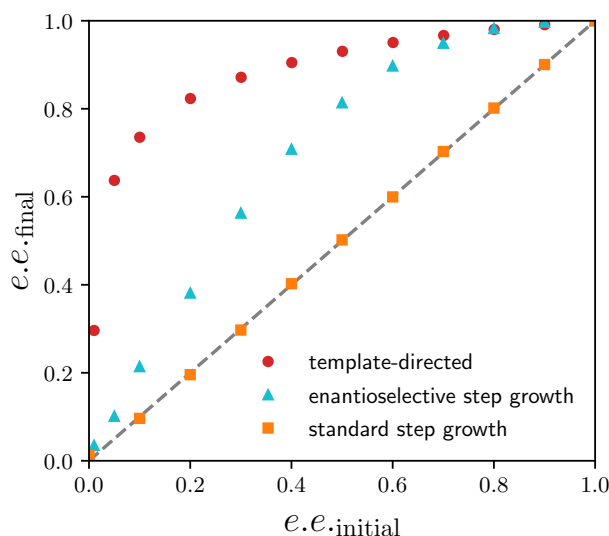


Figure 3.10: Final enantiomeric excess as a function of the initial imbalance, for an open system. The simulations were carried out with a chemostat formed out of 2,800 nucleotides except for the standard step-growth simulations where 700 nucleotides are chemostated. Chiral stalling is set to $\sigma = 0.05$ in the template-directed experiment. The enantioselective experiment has a bias $\alpha = 10^{-10}$, and the standard one no bias. The outflow rate is $k_{\text{out}} = 10^{-8}$ and ligation rate for the step-growth polymerization is $k_{\text{lig}} = 5.27 \times 10^{-9}$. Each data point results from the average of 20 independent trajectories.

unbiased step-growth polymerization conserves the enantiomeric excess of the chemostat. The bias is amplified in the case of a perfectly chiroselective step-growth process. Still, this mechanism's lack of autocatalytic property keeps the amplification low compared to the case of templated ligation. Indeed, template-directed ligation in the presence of chiral stalling (here $\sigma = 0.05$) amplifies best the bias of the chemostat, especially at low chemostat bias where a bias of $e.e. = 10\%$ can be boosted up to values as high as $e.e. = 75\%$ without the need of any racemization reactions. These results, depicted in Figure 3.10, might change depending on the concentration of the chemostat or ligation rates and should be considered indicative only. Still, it is to be noted that the selective step-growth polymerization results do not vary much when changing the concentration of the chemostat or the ligation rate.

3.5 Discussion

The main result of the closed reactor experiment is that RNA template-directed synthesis in the presence of racemization reactions, starting from a racemic pool of monomers and dimers, can converge toward a homochiral system of RNA polymers. The cross-inhibition

(in practice, the chiral stalling) during the ligation observed by Joyce *et al.* is a crucial component of the mechanism because, without it, the system remains closed to the racemic state (Figure 3.5.b). In addition, this transition appears robust and does not require fast racemization. Even in the case of slow racemization, the results still hold. A second significant result concerns the length of the strands in the transition to homochirality. During this process, one of the two chirality will be washed out from the system. The disappearing chirality is random: a minute imbalance due to the stochasticity of chemical reactions, or to other causes, will amplify. The fact that one chirality disappears from the system can be seen as reducing the sequence space. Indeed, in the final homochiral steady-state, heterochiral strands and homochiral strands of the chirality that has been erased are no longer formed. A reduction of the phase space could be seen a disadvantage: a smaller variety of sequences will be explored, and thus fewer functional sequences might be discovered. However, as discussed in Section 3.3.4, the disappearance of one chirality increases the average strand length in the homochiral steady state (Figure 3.5.d). It occurs naturally as more nucleotides from the dominant chirality become available to join in template-directed events which are not stalled anymore. This phenomenon has significant consequences: reducing the sequence space during the transition to homochirality benefits the length of the strands in the system, which is a crucial parameter as the longer the strands, the higher the chances of reaching a first functional catalytic sequence in the system. This idea is depicted in Figure 3.11 and could be significant in the context of the origins of life. The length of polymers that can be replicated without loss of information is fundamentally limited by the error threshold [Eigen, 1971, Smith and Szathmary, 1995] which constrains scenarios on origin of life. Recent studies show that this threshold can be relaxed with template-directed polymerization [Matsubara *et al.*, 2021]. Our studies seem to confirm this since our mechanism can favor the formation of longer polymers.

Regarding the open reactor experiment with chiral stalling, even though no significant *e.e.* is reached during the steady state, we observe an interesting transient behavior when the *e.e.* reaches its maximum value. At this point, one chirality is much more present in the system, and a large part of the system is composed of monomers of this dominant chirality. Plus a large proportion of those strands have an *e.e.* > 50%. From this transient state could emerge a first ribozyme composed of the dominant chirality, accelerating the replication of its kind and thus amplifying definitely the enantiomeric excess and eventually driving the system towards a homochiral system. In addition, template-directed synthesis appears to be a good amplification mechanism of pre-existing chemostat bias in open systems, especially when initial biases are low. The amplification of the bias allows one enantiomeric population to explore a much wider variety of sequences for strands of

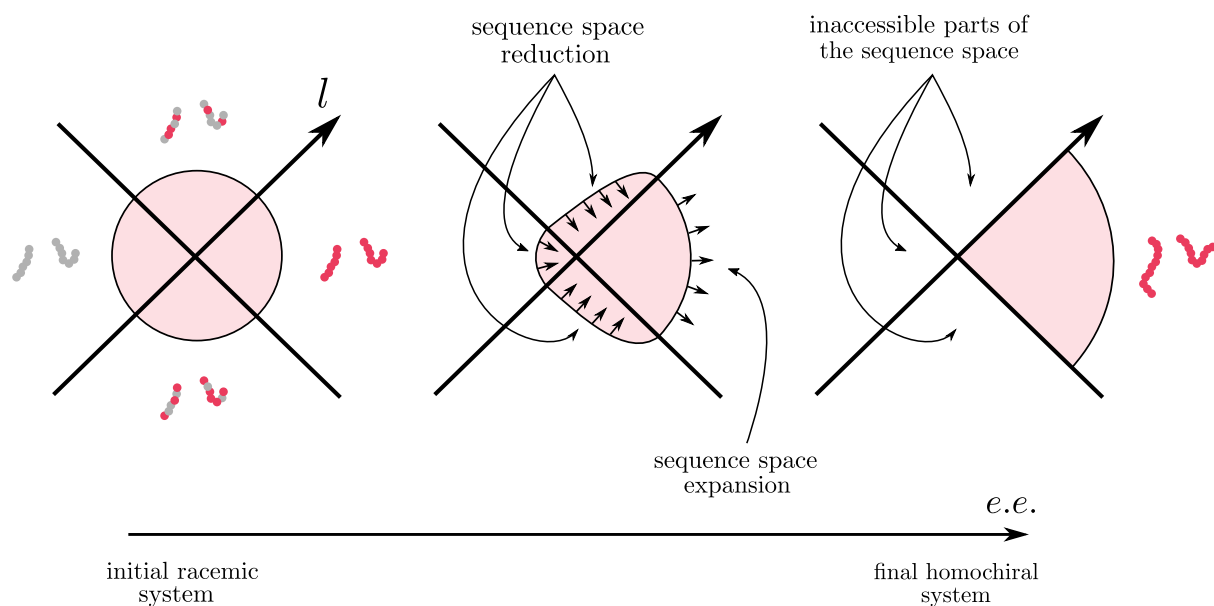


Figure 3.11: Graphical representation of the reduction of the sequence space during the transition toward a homochiral closed system. The reduction of the sequence space resulting in the inaccessibility to heterochiral RNA polymers also results in the widening of the sequence space of the chirality that is now dominating.

their chirality, making it easier for the first functional sequences to appear in solution and thus helping the system to transition eventually to complete homochirality.

In this chapter, we showed that a closed system initially inoculated with a racemic mixture of RNA monomers and dimers, performing template-directed ligation converges towards a homochiral state in the presence of racemization reactions. This mechanism is robust because there is no dependence on external factors, and the self-replication and chiral selectivity components are inherent to RNA template-directed synthesis, apart from the racemization reactions whose speed does not matter. This idea is satisfying and is compatible with the RNA-world scenario [Higgs and Lehman, 2015]. We also understand here the emergence of homochirality as a progressive reduction of the RNA sequence space in early life that allows longer RNA strands to emerge and facilitate reaching the first functional sequences. Our detailed study reinforces recent studies [Chen and Ma, 2020, Tupper et al., 2017] that homochirality could have emerged alongside the first functional polymers thanks to template-directed polymerization.

Appendix 3.A Comparison of block energy models

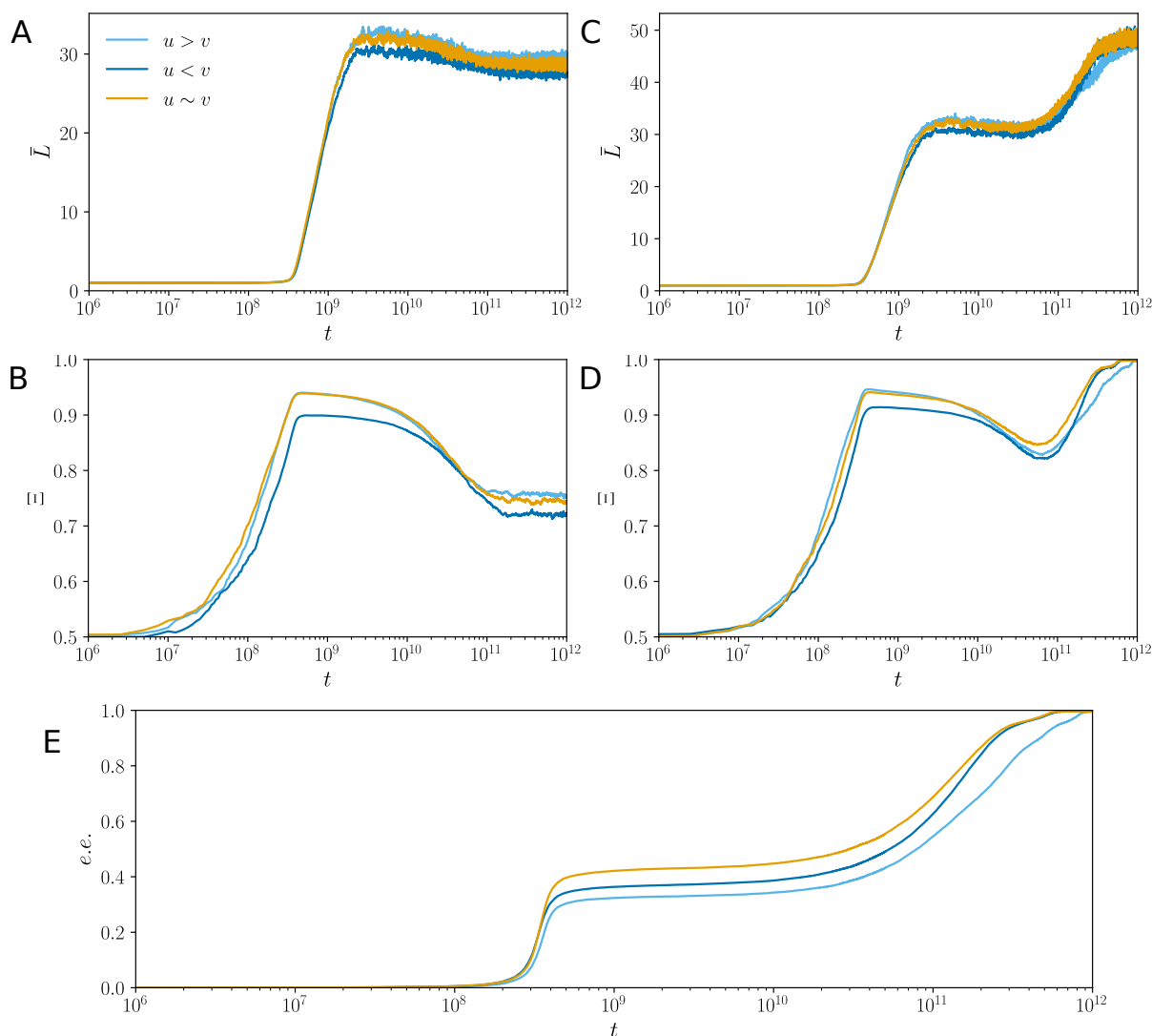


Figure 3.12: Template-directed ligation in a closed reactor simulations for different values of the block energy model. Panels (A) and (B) are data from simulations performed without any racemization reaction, while panels (C), (D) and (E) are extracted from simulations with fast racemization. (A) Average strand length as a function of simulation time (without racemization). (B) Fraction of homochiral motifs in the system as a function of time (without racemization). (C) Average strand length as a function of simulation time in simulations including fast racemization. (D) Fraction of homochiral motifs in the system as a function of time (with racemization). (E) Enantiomeric excess ($e.e.$) at the monomer level (and not only the *free* monomers level) as a function of time. All data shown here are averaged over 20 independent realizations.

Appendix 3.B Single trajectories for closed reactors

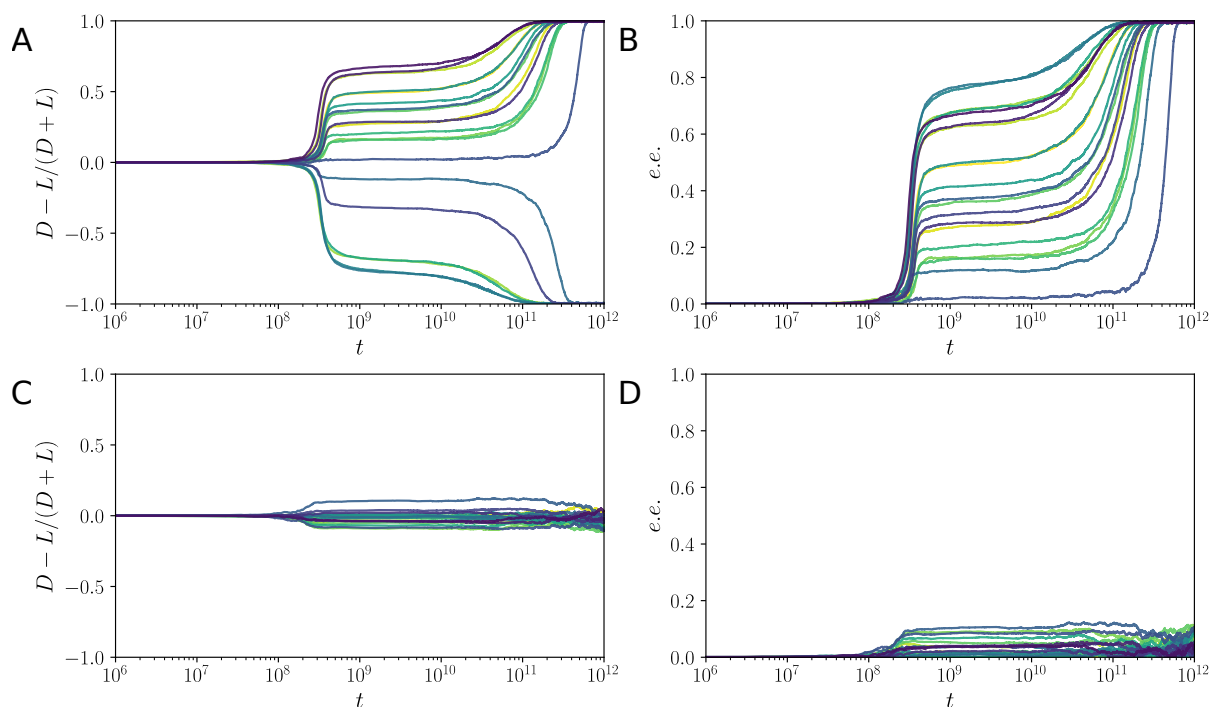


Figure 3.13: Single trajectories for template-directed polymerization in a closed reactor. Panels (A) and (B) show 20 single trajectories which were averaged to obtain the curve of Figure 3.5f, *i.e.*, simulations performed in the presence of racemization reactions and with $\sigma = 0.05$. Panels (C) and (D) show 20 single trajectories of simulations performed in the same conditions but with $\sigma = 1$. Panels (A) and (C) shows an oriented version of the enantiomeric excess, in which the numerator does not have an absolute value. We observe in both cases that trajectories goes in favor either of the D- or L-configuration randomly, due to stochastic fluctuations.

Appendix 3.C Single trajectories for open reactors

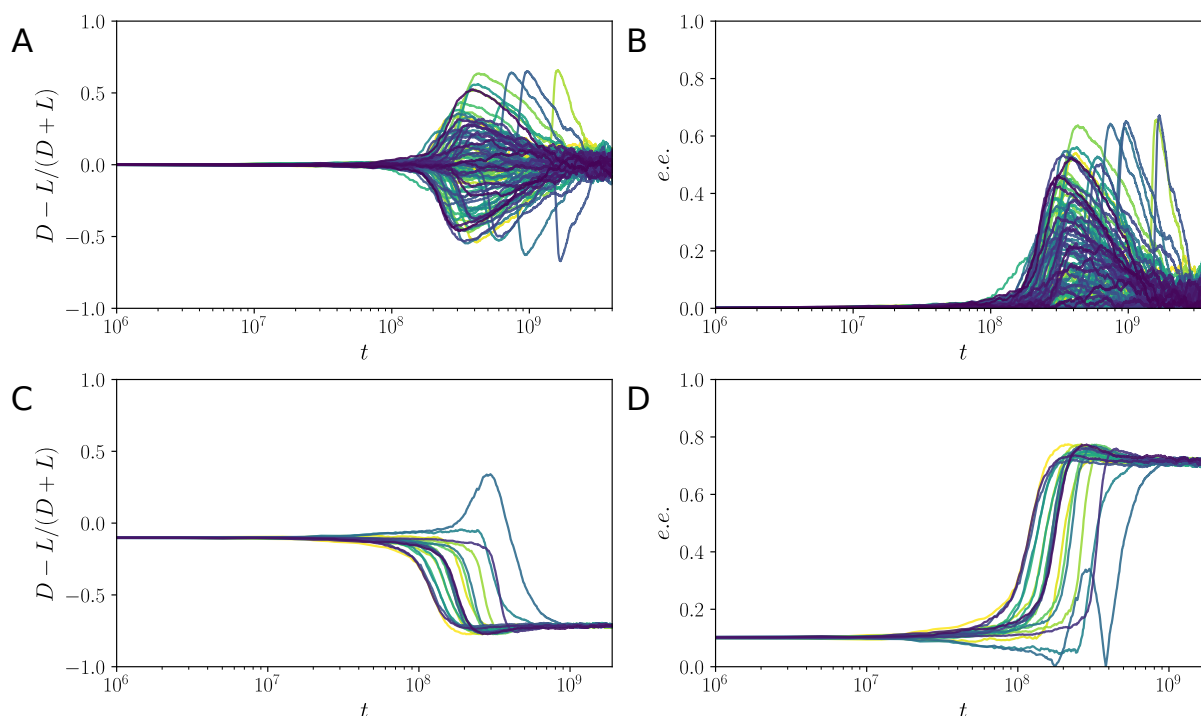


Figure 3.14: Single trajectories for template-directed polymerization in an open reactor. Panels (A) and (B) show 100 single trajectories which were averaged to obtain the curve of Figure 3.9a, *i.e.*, simulations performed without racemization reactions and with $\sigma = 0.05$. Panels (C) and (D) show 20 single trajectories of simulations performed in the presence of an initial enantiomeric excess $e.e. = 10\%$, and with $\sigma = 0.05$. These trajectories have been used to construct Figure 3.10. Panels (A) and (C) shows an oriented version of the enantiomeric excess, in which the numerator does not have an absolute value. We observe in subfigure (A) that trajectories goes in favor either of D- or L-configuration randomly, due to stochastic fluctuations. In the case of the subfigure (C) however, almost all trajectories goes in the same direction, imposed by the initial enantiomeric excess of the chemostat. We however observe than one trajectory goes in the opposite direction for early times, and eventually join the other curve as the chemostat is biased.

CHAPTER 4

Conclusion

In this thesis, we proposed two distinct scenarios involving out of equilibrium physical systems in which homochirality could have emerged. The first scenario involves large chemical systems. We showed, based on results of random matrix theory, that chiral chemical networks in an open reactor (CSTR) are susceptible to undergo a symmetry breaking transition from a racemic state to a homochiral steady state when the number of chiral species in the system becomes large. The likelihood of this transition increases as the number of chiral species involved in the chemical network increases. The stability of the racemic state can be broken into two different fashions, depending on the spectral properties of the Jacobian of the system: the symmetry can be broken for all the species in the same direction or in different directions for a random subset of species. One major advantage of this scenario is that the model does not use any physical or chemical property of enantiomer pairs except for the mirror symmetry, which is a fundamental characteristic of chiral reaction networks. It has the consequence that D and L assignments are totally arbitrary and they just play the role of labels here. We do not rely on and do not use any structural property of D- and L-enantiomers to show the transition. It also means that when we say the symmetry is broken in the same direction for all species, the consequence is not that all dominant species will be part of the D-series or the L-series. The system adopts no collective sign in the homochiral state, which is consistent with the current observation that functional groups do not have the same sign in biology. Moreover, the mechanism has proven to be robust and applicable to sparse chemical networks. To illustrate this general argument, which as described in Section 2.4, should be considered

as a generic and universal model, we introduced a generalization of the Frank model. The generalization includes an arbitrary number of chiral species, which take part in autocatalytic and enantioselective reactions. It also includes chiral inhibition reactions, and exhibits a transition to a homochiral stationary state when (i) the number of chiral species becomes sufficiently large or when (ii) the inflow concentration of achiral species (which act as food) of high free energy becomes sufficiently large, which corresponds to an increase of the supply of energy in the dissipative system. Finally, using an analysis of large chemical databases, we could quantify a threshold above which chiral species become abundant and dominate achiral species in nature. This threshold occurs for a number of heavy atoms $n \simeq 9$, indicating that the previous transition could occur in a system including rather small molecules.

The second scenario is based on template-directed ligation of RNA, a particular form of polymerization which occurs through the ligation of two RNA strands hybridized onto a third one that acts as a template. Using detailed stochastic simulations, we first put in evidence the difficulties for a system undergoing step-growth polymerization to produce any significant enantiomeric excess starting from a racemic pool of activated nucleotides. We then demonstrated that a similar system with template-directed ligation can instead converge towards homochirality assuming racemization reactions at the level of free nucleotides and temperature cycles to impose a periodic melting of RNA complexes. We analyzed the effect of temperature cycles and chiral stalling to confirm that the former does not affect too much the final stationary state (despite the slight decrease of the enantiomeric excess observed) but affects the dynamics, while chiral stalling needs to be sufficiently strong for the system to undergo a symmetry breaking. We also ensure that the symmetry breaking was not just an artifact of the fast racemization at the single nucleotide level by showing that results are conserved even for extremely slow racemization of the system. In addition to the convergence towards a homochiral polymeric system, we also observe an increase in the average polymer length as the enantiomeric excess grows in the system. This logical observation could have major importance for origins of life research as the accessibility to longer biopolymers allows the exploration of functional sequences more easily. In a second time, open chemostated reactors have also been studied, exhibiting a transient increase in the enantiomeric excess without the need for any racemization reactions. Initially biased systems were also analyzed and showed a strong amplification of the initial bias in an open and chemostated reactor, which is unachievable with step-growth polymerization (for instance, a final *e.e.* of 75% starting from an initial excess of 10% has been observed *in silico*). Template-directed ligation appears again to be a key mechanism in the origins of life research. It was recently shown for RNA that

it allows for sequence selection in enzyme-free out of equilibrium reactors [Göppel et al., 2022], and has also been demonstrated as a mechanism for chiroselective self-replication of peptides [Saghatelian et al., 2001] even though proteins usually do not polymerize through templating. We hope, with this work, to have contributed to the research on the emergence of homochirality but also to possibly inspire researchers to pursue the use of random matrix theory for studies of the stability of chemical networks. The next natural step of these works would be to test the scenarios with experimental setups. In the introduction of this thesis, we raised several open questions regarding the emergence of homochirality, and specifically “when did homochirality emerge?”. By construction, the model developed in the first scenario seems well suited for messy and complex prebiotic chemical networks. The second mechanism, however, based on templating, could take place alongside the emergence of the first functional RNA sequences, in the RNA world. The chiral stalling, inherent to RNA template-directed ligation, which tends to inhibit the growth of RNA polymers, plus the possibility for a polymeric system, to increase its average length through its transition to a homochiral steady-state, strongly suggest that homochirality should have appeared at the latest, alongside life.

Appendix 4.A Survival of Self-Replicating Molecules under Transient Compartmentalization with Natural Selection

In this appendix, we present the publication of a work on the effect of transient compartmentalization on the persistence of self-replicating systems subject to competition with fast-replicating parasites.

Article

Survival of Self-Replicating Molecules under Transient Compartmentalization with Natural Selection

Gabin Laurent ¹, Luca Peliti ²  and David Lacoste ^{1,*} 

¹ UMR CNRS Gulliver 7083, ESPCI, 10 rue Vauquelin, CEDEX 05, 75231 Paris, France; gabin.laurent@u-psud.fr

² Santa Marinella Research Institute, 00052 Santa Marinella, Italy; luca@peliti.org

* Correspondence: david.lacoste@espci.fr or david.lacoste@gmail.com

Received: 30 August 2019; Accepted: 27 September 2019; Published: 3 October 2019



Abstract: The problem of the emergence and survival of self-replicating molecules in origin-of-life scenarios is plagued by the error catastrophe, which is usually escaped by considering effects of compartmentalization, as in the stochastic corrector model. By addressing the problem in a simple system composed of a self-replicating molecule (a replicase) and a parasite molecule that needs the replicase for copying itself, we show that transient (rather than permanent) compartmentalization is sufficient to the task. We also exhibit a regime in which the concentrations of the two kinds of molecules undergo sustained oscillations. Our model should be relevant not only for origin-of-life scenarios but also for describing directed evolution experiments, which increasingly rely on transient compartmentalization with pooling and natural selection.

Keywords: origin of life; error catastrophe; parasites

1. Introduction

Research on the origins of life is plagued by several chicken-and-egg problems [1]. One central problem concerns the emergence of functional self-replicating molecules. To be a functional replicator, a molecule must be long enough to carry sufficient information, but if it is too long it cannot be replicated accurately, because shorter non-functional molecules called parasites may replicate faster and take over the system. This was experimentally observed many years ago by Spiegelman [2]. This observation was then rationalized using the notion of error threshold [3], which plays a key role in research on the origins of life [4].

Several theoretical solutions have been proposed to address this issue, among which the stochastic corrector model [5,6] is prominent. In this model, small groups of replicating molecules grow in compartments, to a fixed final size called the carrying capacity. Then, the compartments are divided and their contents are stochastically partitioned between the two daughter compartments. Thanks to the variability introduced by this stochastic division, and to the selection acting on the compartments, functional replicators can be maintained in the presence of parasites.

For a long time, these theoretical ideas have lacked an experimental illustration. This changed in 2016 thanks to progresses in the manipulation of in vitro molecular systems. That year using such systems, Matsumura et al. demonstrated that transient compartmentalization is indeed able to maintain RNA replicators despite the presence of RNA parasites [7], as predicted by the Stochastic corrector model. Then, Bansho et al. built an in vitro molecular ecosystem, based on a different compartmentalized RNA, which was able to display host-parasite oscillatory dynamics and evolution [8].

Inspired by the work of S. Matsumura, we recently explored general scenarios of transient compartmentalization that are also able to maintain information in replicating systems [9,10]. We proposed a transient compartmentalization dynamics with no cell division, which should be achievable when only prebiotic chemistry is available. In our framework, there are no specific requirements regarding the chemical composition or the topology of the compartment boundaries: transient compartmentalization can result from environment fluctuations due to day-night cycles [11], tides cycles [12] or cycles of confinement and release of chemicals from pores [13].

In this paper, we extend the framework in [10] to the case of transient compartmentalization of self-replicating molecules. The main new element in this extension is that the selective pressures acting on this system are not externally imposed, as in our previous work, but stem from the system intrinsic dynamics acting as a form of natural selection. Therefore, the assumptions of this new model are agnostic about the environment and its interaction with the system, which is a desirable feature for scenarios on the origins of life. Besides, this extension may be also pertinent for certain in vitro evolution experiments [14]. Indeed, in vitro evolution experiments based on external selection are often more difficult and cumbersome to carry out than those based on natural evolution, which we consider here.

In one version of our model, we find oscillatory behavior in the population size of replicators. These oscillations are different from the ones observed with bulk hypercycles [15], because they only exist due to the transient compartmentalization dynamics. The oscillations present in our model are similar to the ones which have been observed in the molecular ecosystems mentioned above [8]. In the Section 4 we compare the predictions of our model to these experiments and to the theoretical model [16] developed to analyze them. While our framework is applicable to such experimental systems, it is important to appreciate that it has a wide generality. It could equally well describe many other forms of compartmentalized hypercycles or coupled autocatalytic sets, because the self-replicating molecules which we consider need not be RNA replicases.

2. Materials and Methods

Here, we introduce two models describing a transient compartmentalization process in which self-replicating molecules (the replicase) may coexist with non-self replicating ones (the parasites) which may be replicated by the first ones. These models are amenable to mathematical analysis. In Section 1, we describe a model of transient compartmentalization where the compartments are populated at each round with an inoculum which has a fixed average size λ as shown in Figure 1a. In Section 2, the size of the inoculum $\lambda(t)$ is allowed to vary in time as a result of successive dilution steps as shown in Figure 1b.

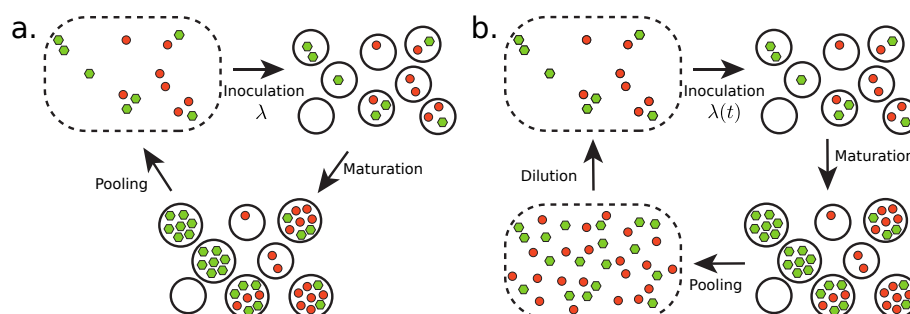


Figure 1. (a) Transient compartmentalization at fixed average number of molecules per compartment, and (b) with a variable average number of molecules. In (a), the cycle splits into steps of inoculation, with a fixed average number of molecules per compartment λ , maturation and then pooling, while in (b) the inoculation step is done with a variable average number of molecules per compartment $\lambda(t)$ because the cycle contains in addition a dilution step. The green and red circles represent the replicators and their parasites, respectively.

2.1. Transient Compartmentalization with a Fixed Inoculum Size

In this subsection, we describe the behavior of a compartmentalized self-replicating system made of two species: self-replicating molecules (replicases) and parasites. Replicases can make copies of themselves and of other parasite molecules, while the parasites can only be copied by the replicases. This is a different case from that discussed in [10], where both the molecules of interests (in that case, the ribozymes) and the parasites could be replicated by externally provided enzymes. In further contrast, in the present case, there is no externally applied selection. Thus, the main steps of the replicating cycle in this case are as shown in Figure 1a:

- Inoculate the compartments.
- Maturate the compartments.
- Pool compartment contents.

In the inoculation step, as in [10], one chooses a number n of molecules from the pool, where n is Poisson distributed with average λ . The resulting inoculum then contains m replicases and $y = n - m$ parasites, which are distributed according to a Binomial distribution of parameter x , where x is the initial fraction of replicases in the pool. We denote by $P_\lambda(n, m, x)$ the resulting probability distribution. This follows closely the corresponding steps in [10]. However, the dynamics of the maturation step is different and is described by the following equations:

$$\begin{aligned} \dot{m}(t) &= \alpha m(t)^2, \\ \dot{y}(t) &= \gamma m(t) y(t), \end{aligned} \quad (1)$$

where $m(t)$ and $y(t)$ are, respectively, the self-replicating and the parasitic species populations at time t , while α and γ are their respective replication rates. The analytical solution described in Appendix A yields the compartment composition $(m(T), y(T))$ at the stopping time T as a function of the initial composition, denoted by $(m = m(0), y = y(0))$. The stopping time T is itself fixed by the condition

$$m(T) + y(T) = K + n, \quad (2)$$

where $n = m + y$ denotes the initial number of molecules in the compartment and K is a parameter that represents the number of new strands that can be created during the replication process, due to the finite amount of nutrients present in the compartment. We call it the carrying capacity. We use in the following the shorthands: $\bar{m} = m(T)$ and $\bar{n} = m(T) + y(T)$. Moreover, the ratio $\Lambda = \gamma/\alpha$ of the replicating constants of both species is another important parameter of the dynamics.

After the maturation step, the contents of the compartments are pooled. The fraction x' of replicases in the pool is expressed in terms of its value x at the beginning of the round by the following equation:

$$x'(x, \lambda) = \frac{\langle \bar{m} \rangle}{\langle \bar{n} \rangle} = \frac{\sum_{n,m} \bar{m}(n, m) P_\lambda(n, m, x)}{\sum_{n,m} \bar{n}(n, m) P_\lambda(n, m, x)}, \quad (3)$$

where $\langle \dots \rangle$ denotes the average with respect to the probability distribution $P_\lambda(n, m, x)$. Note that the number of molecules in the compartments at the end of the maturation step is not uniform (in particular, compartments which are pure in parasites contain at the end the same number of molecules as in the beginning). Thus, we cannot directly average x over the compartments as done in [10].

In Appendix B, we show that, in the limit $\Lambda \gg 1$ where the parasites are much more aggressive than the self-replicating molecules, the recursion equation (3) can be simplified, yielding

$$x'(\lambda, x) = \frac{\lambda x + K e^{-\lambda} (e^{\lambda x} - 1)}{\lambda + K(1 - e^{-\lambda x})}. \quad (4)$$

The behavior of this model is described in Section 3.1.

2.2. Transient Compartmentalization with Variable Inoculum Size

We now consider a model of transient compartmentalization with a variable inoculum size λ . In experiments based on serial transfers, a fraction of the solution is transferred into a new fresh medium repeatedly [17]. This can be described theoretically by adding a dilution step in the replicating cycle, as shown in Figure 1b. Then, λ can change because a given amount of the pooling solution can contain a variable number of replicating molecules, depending on their average concentration. The dynamics is now described by a pair of equations for the evolution of the fraction x and of the parameter λ :

$$x' = \frac{\langle \bar{m} \rangle}{\langle \bar{n} \rangle}, \quad \lambda' = \frac{\langle \bar{n} \rangle}{d}. \quad (5)$$

where d is the dilution factor and \bar{m} , \bar{n} are given by the same equations as above, evaluated with the current value of λ .

Using the same approximations used to derive Equation (4), we obtain the following set of equations, valid for $\Lambda \gg 1$:

$$\begin{aligned} x'(\lambda, x) &= \frac{\lambda x + Ke^{-\lambda}(e^{\lambda x} - 1)}{\lambda + K(1 - e^{-\lambda x})}, \\ \lambda'(\lambda, x) &= \frac{\lambda + K(1 - e^{-\lambda x})}{d}. \end{aligned} \quad (6)$$

These equations can be more easily manipulated than the corresponding equations for the global compartmentalization process.

The behavior of this model is described in Section 3.2.

3. Results

We first describe the results of the model with fixed inoculum size (Figure 1a, Section 2.1), and then those of the variable inoculum-size model (Figure 1b, Section 2.2).

3.1. Fixed Inoculum Size

By studying the stability of the fixed points of the recursion of Equation (4), we obtain the phase diagram shown in Figure 2, which represent the compositions that are accessible to the system on long times. In contrast to the phase diagram obtained in [10], we find a large region of coexistence between the self-replicating molecules and the parasites and no pure parasite phase. The absence of the pure parasite phase is expected, since parasites can not grow without replicators. Thus, there are only two phases: a pure replicator phase and a coexistence phase, in which compartments remain of mixed composition. Although the coexistence region appears large, in fact, in the main part of it, self-replicating molecules are maintained at a very small concentration, as shown by the color scale in Figure 2. Therefore, to maintain replicators at a significant concentration, one can not escape the condition that the average size of compartments be of the order of one molecule per compartment.

By evaluating the derivative of x' with respect to x at $x = 1$, one obtains the equation of the vertical asymptote of the phase diagram separating the pure replicator phase and the coexistence region, which is given by $\lambda = 1$. The same condition used at $x = 0$ shows that this fixed point is always unstable at finite value of λ , which confirms that there is no pure parasite phase. In the coexistence region, a family of vertical asymptotes can be obtained by solving the equation $x'(\lambda, x) = x$ for $0 < x < 1$ in terms of λ . No simple expression has been found for the equations of the corresponding horizontal asymptotes, which separate the pure replicator phase and the coexistence region when $\lambda \rightarrow \infty$.

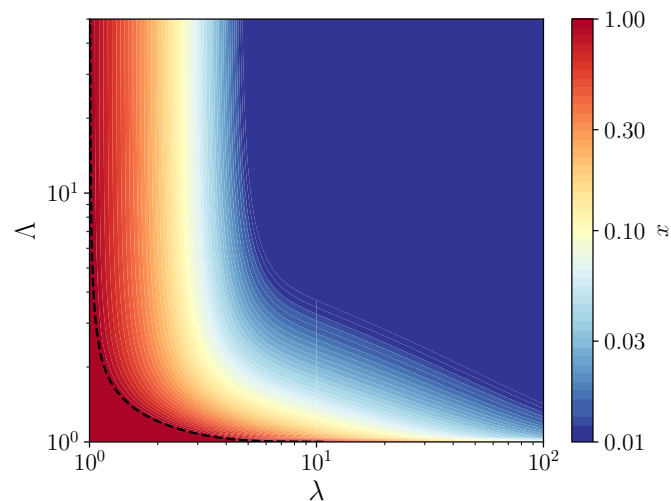


Figure 2. Contour map of the fraction x of replicators as a function of (λ, Λ) , for a carrying capacity $K = 100$, where λ denotes the average number of molecules per compartment and Λ the relative growth rates of the parasites with respect to the host. The dotted line is the contour of $x = 1$, which marks the border of the pure replicators phase (the red region). Above this line, a coexistence region exists between the two species at a fraction of replicators indicated by the color scale.

3.2. Variable Inoculum Size

The most striking feature of the model with variable inoculum size, described by Equation (5), is the appearance of oscillations in both x and λ . They are similar to the ones observed in experiments with host-parasite RNAs [8] and modeled numerically in [16].

When the value of K and of d are not too large, this system exhibits oscillations in the populations of replicators and parasites shown in Figure 3a. These oscillations can also be seen when representing the fraction x of replicators as function of the average compartment size λ , as shown in Figure 3b. This behavior can be explained as follows: after the first inoculation, parasites are being replicated quickly by replicators, and the dilution does not counterbalance this increase in population. At some point, the fraction of parasites in the population is so high that there are not enough self-replicators to contribute to their replication. Then, the dilution has an important effect, since it decreases the population per compartment λ , until its average reaches values around $\lambda = 1$ (see Figure 3b). At this point, compartments contain on average only a single molecule, which can be either a parasite or a self-replicator. The population in empty compartments or compartments containing a single parasite molecule does not grow, therefore only compartments containing a single replicator or containing one replicator and one parasite will contribute substantially to the next round. At this point, the replicator population increases, and starts replicating parasites for several rounds, triggering the process again. Note that the mechanism producing these oscillations is different from the Lotka–Volterra one, where the competition between the two species is the main ingredient [18]; instead, here, transient compartmentalization plays an essential role.

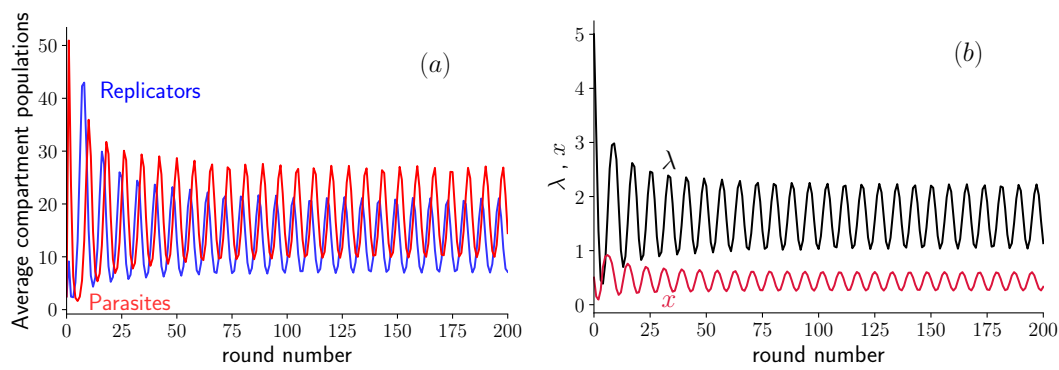


Figure 3. Oscillations in the average amount of self-replicating and parasite molecules per compartment as a function of the round number for $d = 19$, $K = 60$, and $\Lambda = 5$. **(a)** Average population size $\langle \bar{n} \rangle$ of replicators and $\langle \bar{n} - \bar{m} \rangle$ of parasites after the growth step plotted vs. round number. **(b)** Fraction x of replicators and average λ of inoculum size. Notice that the oscillations rebound close to the line $\lambda = 1$.

To delve deeper in the analysis of these oscillations, let us proceed with the equations in Equation (6), which are valid in the limit $\Lambda \gg 1$. In a simulation of these equations at a given value of K , we observe an abrupt transition when varying the dilution factor. Indeed, when $K = 60$ and $d = 18$, the two average populations oscillate steadily, as shown in Figure 4a, while when $d = 22$, oscillations quickly die out as shown in Figure 4b. This abrupt transition is the sign of a bifurcation, which we identify as a supercritical Hopf bifurcation (see Appendix C for more details). The bifurcation occurs at $d = 20.74$ given that $K = 60$. Below this value, the system shows unstable spirals and converges to a limit cycle, while above this value, the system shows stable spirals which converge towards a fixed point (cf. [19], Section 8.2).

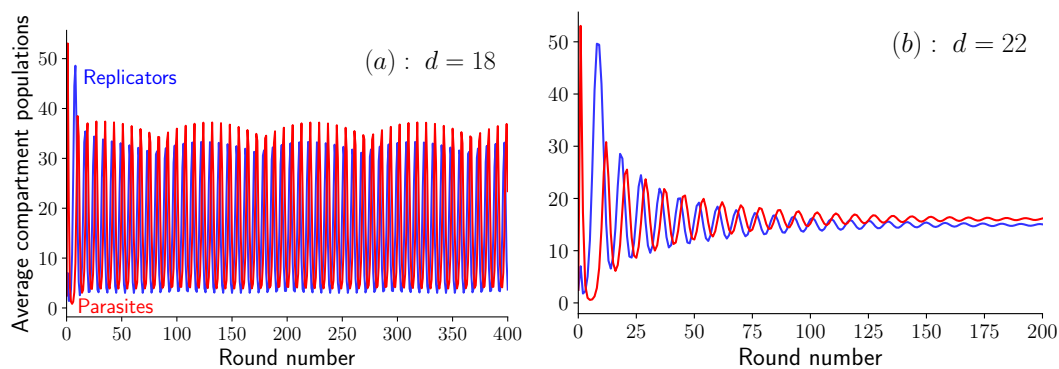


Figure 4. Oscillations in the average amount of self-replicating and parasites molecules per compartment as a function of the round number for $K = 60$ and $\Lambda \gg 1$. **(a)** Steady oscillations at $d = 18$ (unstable spirals), and **(b)** damped oscillations at $d = 22$ (stable spirals). Note the beating pattern in the oscillations visible in **(a)**.

When the parameter d is further increased still keeping K fixed, we find a second transition at $d = 37.15$. At this point, the system no longer oscillates or spirals around a fixed point, but instead converges towards this fixed point monotonically, a case identified as stable node in the literature ([19], p. 128).

Another interesting feature in these oscillations is the beating pattern which is visible on Figure 4a as a modulation in the amplitude of the oscillations. This pattern results from the interplay between two frequencies, the sampling frequency fixed by the duration of a single round, and the intrinsic frequency of the oscillations. By changing the sampling frequency, the beating pattern is accordingly modified.

To summarize all these results, we build the phase diagram in the plane (K, d) shown in Figure 5a. As can be seen in this figure, there are three phases in this diagram, which are separated by line

boundaries in the plane (K, d) . The upper region corresponds to a phase of unstable spirals [19], where steady oscillations are present; an intermediate region of stable spirals [19], where damped oscillations are present; and the lower region of stable node [19], where oscillations are absent. The presence of these line boundaries can be understood from the following argument. For large values of K , the equations which determine the fixed point coordinates (x^*, λ^*) deduced from Equation (6) can be simplified to yield :

$$\begin{aligned} x^* &= \frac{e^{-\lambda^*}(e^{\lambda^* x^*} - 1)}{1 - e^{-\lambda^* x^*}}, \\ \lambda^* &= \frac{K(1 - e^{-\lambda^* x^*})}{d}. \end{aligned} \tag{7}$$

The second equation above can be written as

$$\frac{K}{d} = \frac{\lambda^*}{1 - e^{-\lambda^* x^*}}, \tag{8}$$

which shows that the coordinates of the fixed point (x^*, λ^*) only depend on the ratio K/d in this large K limit. It follows that the boundary between the region of unstable and stable spirals, where the Hopf bifurcation occurs is a straight line as shown in Figure 5a. A similar argument holds for the boundary between the stable spirals and the stable node, which is also a straight line in this diagram.

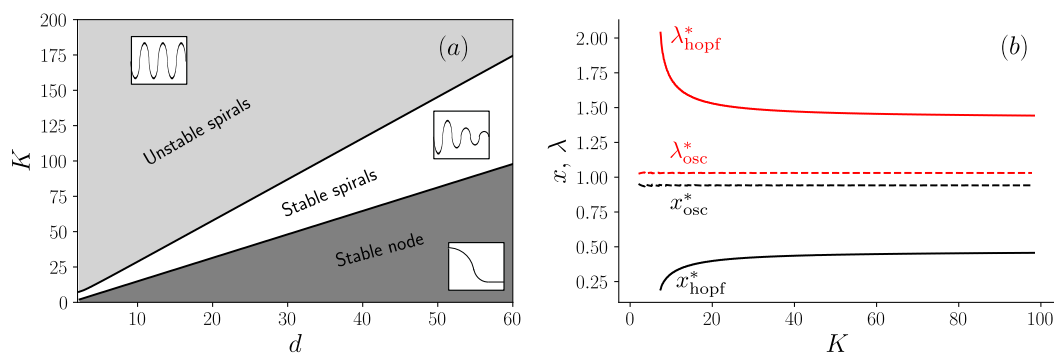


Figure 5. (a) Phase diagram in the plane (K, d) in the limit $\Lambda \gg 1$ containing three regions: unstable spirals (with an inset representing steady oscillations), stable spirals (with an inset representing damped oscillations) and stable node (with an inset representing a curve with no oscillations); and (b) evolution of the fixed point coordinates (x^*, λ^*) as a function of K , on the Hopf bifurcation (solid line) and on the transition line between the stable node and stable spirals (dashed line).

To confirm this interpretation, we show in Figure 5b the values of (x^*, λ^*) as a function of K , evaluated either on the boundary of the Hopf bifurcation and denoted with the subscript “hopf”, or on the stable node-stable spirals boundary and denoted with the subscript “osc”. When reporting the asymptotic values of (x^*, λ^*) obtained for large K in Equation (8), one recovers the values of the slopes of two lines in Figure 5a.

4. Discussion

We studied a simple system composed of a self-replicating molecule (a replicase) and a parasite molecule that needs the replicase for copying itself. In the case of a fixed inoculum size (i.e., for a fixed value of the parameter λ), we found that this system is able to maintain the replicase molecules against the take-over of parasites in the absence of artificial selection. Although the phase diagram contains a large coexistence region, only in a small part of it, when λ is close to one, are the replicase molecules maintained at a significant concentration. This may explain why experiments on directed evolution using compartmentalized self-replicating molecules such as DNA or RNA are usually carried

out in this regime for these molecules, while all other required chemical species (nucleotides, other intermediates, etc.) are typically in excess.

The theoretical framework we developed for the case of a variable inoculum size has many similarities with the model proposed in [16] to explain experiments on host-parasite RNAs [8]. There are however some differences: we consider an infinite number of compartments instead of a finite one, we do not include mutations which could turn the replicase into a parasite, and we do not include local mixing, which means that our model corresponds to the infinite mixing limit of ref [16]. Despite these differences, we also find a regime of values of the parameters (in particular for the dilution factor or the carrying capacity) in which sustained oscillations are possible in agreement with Furubayashi and Ichihashi [16]. In [8], the ratio of catalysis rate constants of the parasite with respect to host, which we denote Λ , was about 5: this can be obtained by extracting, from Table I in that reference, values of $\alpha = 0.29, \gamma = 1.5$ and using $\Lambda = \gamma/\alpha \simeq 5$. We find that oscillations are indeed present in our model in this range of the parameters, and oscillation periods comparable to the value reported in [8] can be recovered from this estimate. These similarities suggest that our assumptions of an infinite number of compartments may be reasonable and that other differences related to local mixing and pooling may not be essential. To test these points more precisely, it would be useful to redo experiments similar to the ones of Bansho et al. [8] but with a pooling step done as in [7].

It appears that, for both fixed and variable inoculum sizes, the regime of pure compartments (one molecule per compartment on average) has a particular importance: for a fixed inoculum size, only in this regime can a significant average fraction of self-replicating molecules be maintained, and for a variable inoculum size, only in this regime can a rebound occur in the populations of molecules, allowing oscillations. We surmise that this regime could have a specific significance for the origin of life. To elaborate a bit on this point, we recall that the emergence of special molecules bearing the genetic information is an essential step in the origin of life as emphasized in the RNA world. These molecules are typically found in minority with respect to other species, yet this minority has control of the entire cell [20]. This is a form of information control, which is thought to be one of the key parameters in the origins of life [21]. Fluctuations of this minority species therefore have a special role due to their small number. In contrast, many other chemical species, which are not information carriers, are found in large numbers, with fluctuations statistically obeying the law of large numbers. In our model, we see a clear illustration of this mechanism: the replicase behaves as a genome-like molecule, present at the lowest non-zero possible concentration of one molecule per compartment, while all other molecules, which depend on the genome molecules for their own making, are available in the protocell in large concentration.

5. Conclusions

Without considering complex chemistry, we have proposed a model which is able to capture important features for origins of life research, such as the ability to maintain self-replicating molecules using transient compartmentalization and natural selection. An interesting feature of the model with constant inoculum size is the maintenance of the self-replicating molecule by a form of information control, at the critical level of one molecule per compartment. A striking feature of the model with variable inoculum size is the appearance of oscillations, which are similar to the ones observed in experiments with compartmentalized self-replicating RNAs [8].

Naturally, Bansho et al. [8] presented much more than the mere observation of these oscillations. By studying the sequence information of the replicase and its parasites, they suggested that parasites can take an active part in the evolution of their host and not just in their own. Different sub-populations of parasites can appear, forming an ecosystem [22], which accelerates evolution. Future studies are needed to quantify these co-evolutionary mechanisms, and perhaps our model could help in that task.

Another important direction for future work would be to consider a large number of interacting chemical species, a situation frequently encountered in statistical physics [23]. In this case, we expect that the basic unit of description may no longer be that of single chemical species, but could become

collective excitations of the composition, similar to quasi-species [24] or composomes [25]. A general theory of non-equilibrium chemical networks, constrained by conservation laws and symmetries, has recently been put forward [26,27]. One attractive feature of such a framework for describing complex chemical systems is that it relies mainly on stoichiometry; therefore, the explicit knowledge of the kinetics, which is often missing, is not needed [17].

New types of emergent behaviors could arise by enlarging further the dynamics of compartmentalization. One possibility would be to consider loose compartments [28] or a continuous automated in vitro evolution [29], which will require introducing spatially dependent parameters in our transient compartmentalization dynamics. Besides the relevance for the origins of life, we hope that our work could trigger new research directions on applications of transient compartmentalization for chemistry or biochemistry. Perhaps, these new research directions could help overcome practical and fundamental hurdles associated with the synthesis of complex molecules, and facilitate the making of new catalysts or artificial cells [30].

Author Contributions: Conceptualization, L.P. and D.L.; methodology, G.L.; software, G.L.; validation, G.L., D.L. and L.P.; formal analysis, G.L.; investigation, G.L.; resources, D.L.; data curation, G.L.; writing—original draft preparation, D.L.; writing—review and editing, L.P. and G.L.; visualization, G.L.; supervision, D.L.; and project administration, D.L.; funding acquisition, D.L.

Funding: L.P. was supported by the Agence Nationale de la Recherche (No. ANR-10-IDEX- 0001-02, IRIS OCAV).

Acknowledgments: L.P. is grateful to the Gulliver lab of the ESPCI for a most pleasant hospitality. We would like to thank T. Furubayashi for many insightful discussions and a critical reading of this work. D.L. would like to thank A. Blokhuis for a particularly fruitful collaboration on which this work was built.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Exact Solution of the Maturation Equations

The maturation Equation (1) can be solved analytically to give

$$\begin{aligned} m(t) &= \frac{m}{1 - \alpha t m}, \\ y(t) &= \frac{y}{(1 - \alpha t m)^\Lambda}, \end{aligned} \quad (\text{A1})$$

where m and y are the initial population sizes of both species, n their sum, t the time, and $\Lambda = \gamma/\alpha$ the ratio of the two replicating constants. The dynamics of these equations is hyper-exponential, and exhibits finite-time divergences. However, these divergences are not relevant for the model, since the carrying capacity cut this dynamics off at finite values of m and y , given the stopping condition given by Equation (2). Introducing the quantity $u = 1 - \alpha t m$, we express the stopping condition in Equation (2) as follows:

$$\frac{m}{u} + \frac{(n - m)}{u^\Lambda} = K + n. \quad (\text{A2})$$

We can then solve this equation in terms of u to obtain the final population sizes denoted by \bar{m} and \bar{n} .

Appendix B. Derivation of the Equations in the $\Lambda \gg 1$ Limit

The expression of x' in the $\Lambda \gg 1$ limit is evaluated by splitting averages in multiple parts. The denominator $\langle \bar{n} \rangle$ of the recursion in Equation (3) is given by

$$\sum_{n,m} \bar{n} P_\lambda(n, m, x) = \sum_n \bar{n} P_\lambda(n, 0, x) + \sum_{n>0, m>0} \bar{n} P_\lambda(n, m, x), \quad (\text{A3})$$

where the first sum of the right hand side corresponds to compartments without self-replicators ($m = 0$) and the second one corresponds to all other compartments containing replicases. Their final populations remain equal to n molecules after the maturation in the former compartments without replicators and grow to $K + n$ molecules in compartment containing replicases. Thus, $\langle \bar{n} \rangle$ can be expressed by

$$\langle \bar{n} \rangle = \sum_n n P_\lambda(n, 0, x) + \sum_{n>0, m>0} (K + n) P_\lambda(n, m, x), \tag{A4}$$

and, using the definition of $P_\lambda(n, m, x)$ introduced in Section 2.1, yields the exact equation

$$\langle \bar{n} \rangle = \lambda + K(1 - e^{-\lambda x}). \tag{A5}$$

The numerator $\langle \bar{m} \rangle$ is evaluated in a similar way, starting by splitting the average to give

$$\sum_{n,m} \bar{m} P_\lambda(n, m, x) = \sum_{n>0} \bar{m} P_\lambda(n, n, x) + \sum_{n>0, 0<m<n} \bar{m} P_\lambda(n, m, x), \tag{A6}$$

where the first sum of the right hand side corresponds to compartments with only self-replicating molecules ($m = n$), and the second sum corresponds to compartments with mixed populations. Empty and pure parasitic compartments do not contribute to the average because in them $\bar{m} = 0$. In the former case, the final self-replicator population verifies $\bar{m} = K + n$. In the latter case, with mixed population, no exact solution for \bar{m} can be determined. We assume that self-replicators do not have the time to replicate during the maturation in presence of aggressive parasites, i.e., that $\bar{m} = m$ if $\Lambda \gg 1$. Equation (A6) can thus be rewritten

$$\langle \bar{m} \rangle = \sum_{n>0} (K + n) P_\lambda(n, n, x) + \sum_{n>0, 0<m<n} m P_\lambda(n, m, x), \tag{A7}$$

and gives after similar derivations as for Equation (A5),

$$\langle \bar{m} \rangle = K e^{-\lambda} (e^{\lambda x} - 1) + \lambda x, \tag{A8}$$

as long as $\Lambda \gg 1$. Finally, by combining Equations (A5) and (A8), one obtains the following approximate expression, valid for $\Lambda \gg 1$, of the recursion Equation (3):

$$x'(\lambda, x) = \frac{\lambda x + K e^{-\lambda} (e^{\lambda x} - 1)}{\lambda + K(1 - e^{-\lambda x})}. \tag{A9}$$

The approximate expression of Equation (5) straightforwardly follows.

Appendix C. Analysis of The Bifurcation

To analyze the nature of the bifurcation, one could a priori use either the coordinates $(\langle \bar{m} \rangle, \langle \bar{y} \rangle)$ or the coordinates (x, λ) , since there exists a simple bijection between the two sets of coordinates defined by Equation (5). In the following, we study numerically the Jacobian of the system of equations in the coordinates (x, λ) given by Equation (6). We denote the two eigenvalues of this Jacobian by ϕ_k with $k = 1, 2$. The behavior of these quantities is shown in Figure A1. By evaluating the two eigenvalues at the fixed point (x^*, λ^*) , we observe that their modulus moves from above to below 1 as d changes from $d < 20.74$ to $d > 20.74$. This means that the fixed point (x^*, λ^*) with $0 < x^* < 1$, which is unstable for $d < 20.74$, becomes stable as d increases beyond this value. At the transition point for $d = 20.74$, the eigenvalues are complex, which is the indication of a Hopf bifurcation. Below $d < 20.74$, the system spirals up from the unstable fixed point to a stable limit cycle, which encloses the fixed point, while for $d > 20.74$, the system spirals down towards the stable fixed point. We checked that the amplitude of the oscillations decreases smoothly to zero as the bifurcation point is approached, and that therefore the bifurcation is supercritical. There is a second transition at $d = 37.15$, where the imaginary parts of

the eigenvalues vanish. This means that the system no longer oscillates or spirals around the fixed point, but instead converges to it monotonically. In this regime, the fixed point is therefore a stable node ([19], p. 128).

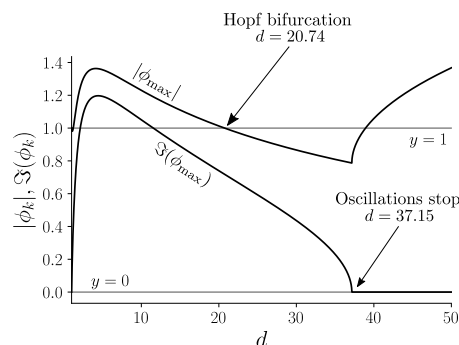


Figure A1. Maximal modulus and maximal imaginary part of eigenvalues of the Jacobian corresponding to Equation (6) for a carrying capacity $K = 60$.

References

- Dyson, F. *Origins of Life*; Cambridge University Press: Cambridge, UK, 1985.
- Spiegelman, S.; Haruna, I.; Holland, I.B.; Beaudreau, G.; Mills, D. The synthesis of a self-propagating infectious nucleic acid with a purified enzyme. *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 919–927. [[CrossRef](#)] [[PubMed](#)]
- Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523. [[CrossRef](#)] [[PubMed](#)]
- Tupper, A.S.; Higgs, P.G. Error thresholds for RNA replication in the presence of both point mutations and premature termination errors. *J. Theor. Biol.* **2017**, *428*, 34–42. [[CrossRef](#)] [[PubMed](#)]
- Grey, D.; Hutson, V.; Szathmáry, E. A re-examination of the stochastic corrector model. *Proc. R. Soc. Lond. B* **1995**, *262*, 29–35.
- Szathmáry, E.; Demeter, L. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **1987**, *128*, 463–486. [[CrossRef](#)]
- Matsumura, S.; Kun, A.; Ryckelynck, M.; Coldren, F.; Szilágyi, A.; Jossinet, F.; Rick, C.; Nghe, P.; Szathmáry, E.; Griffiths, A.D. Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* **2016**, *354*, 1293–1296. [[CrossRef](#)] [[PubMed](#)]
- Bansho, Y.; Furubayashi, T.; Ichihashi, N.; Yomo, T. Host–parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4045–4050. [[CrossRef](#)]
- Blokhuis, A.; Nghe, P.; Peliti, L.; Lacoste, D. The generality of transient compartmentalization and its associated error thresholds. *BioRxiv* **2019**, submitted.
- Blokhuis, A.; Lacoste, D.; Nghe, P.; Peliti, L. Selection dynamics in transient compartmentalization. *Phys. Rev. Lett.* **2018**, *120*, 158101. [[CrossRef](#)]
- Tkachenko, A.V.; Maslov, S. Spontaneous emergence of autocatalytic information-coding polymers. *J. Chem. Phys.* **2015**, *143*, 045102. [[CrossRef](#)]
- Geyrhofer, L.; Brenner, N. Coexistence and cooperation in structured habitats. *bioRxiv* **2018**, 1–23. [[CrossRef](#)]
- Baaske, P.; Weinert, F.M.; Duhr, S.; Lemke, K.H.; Russell, M.J.; Braun, D. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9346–9351. [[CrossRef](#)]
- Zadorin, A.S.; Rondelez, Y. Selection strategies for randomly partitioned genetic replicators. *Phys. Rev. E* **2019**, *99*, 062416. [[CrossRef](#)] [[PubMed](#)]
- Takeuchi, N.; Hogeweg, P. Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Phys. Life Rev.* **2012**, *9*, 219–263. [[CrossRef](#)] [[PubMed](#)]
- Furubayashi, T.; Ichihashi, N. Sustainability of a compartmentalized host-parasite replicator system under periodic washout-mixing cycles. *Life* **2018**, *8*, 3. [[CrossRef](#)]
- Blokhuis, A.; Lacoste, D.; Gaspard, P. Reaction Kinetics in open reactors and serial transfers between closed reactors. *J. Chem. Phys.* **2018**, *148*, 144902. [[CrossRef](#)]

18. Fujii, T.; Rondelez, Y. Prey Molecular Ecosystems. *ACS Nano* **2013**, *7*, 27–34. [[CrossRef](#)]
19. Strogatz, S.H. *Nonlinear Dynamics and Chaos*; Perseus Books Publishing: New York, NY, USA, 1994.
20. Kaneko, K.; Yomo, T. On a kinetic origin of heredity: Minority control in a replicating system with mutually catalytic molecules. *J. Theor. Biol.* **2002**, *214*, 563–576. [[CrossRef](#)]
21. Nghe, P.; Hordijk, W.; Kauffman, S.A.; Walker, S.I.; Schmidt, F.J.; Kemble, H.; Yeates, J.A.; Lehman, N. Prebiotic network evolution: Six key parameters. *Mol. Biosyst.* **2015**, *11*, 3206–3217. [[CrossRef](#)]
22. Furubayashi, T.; Ueda, K.; Bansho, Y.; Motooka, D.; Nakamura, S.; Ichihashi, N. Evolutionary arms-races between host and parasitic RNA replicators in an artificial cell-like system. *bioRxiv* **2019**, 728659. [[CrossRef](#)]
23. Goldenfeld, N.; Woese, C. Life is physics: Evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 375–399. [[CrossRef](#)]
24. Eigen, M.; Schuster, P. The Hypercycle. *Naturwissenschaften* **1978**, *65*, 7–41. [[CrossRef](#)]
25. Lancet, D.; Zidovetzki, R.; Markovitch, O. Systems protobiology: origin of life in lipid catalytic networks. *J. R. Soc. Interface* **2018**, *15*, 20180159. [[CrossRef](#)] [[PubMed](#)]
26. Rao, R.; Esposito, M. Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics. *Phys. Rev. X* **2016**, *6*, 041064. [[CrossRef](#)]
27. Poletini, M.; Esposito, M. Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. *J. Chem. Phys.* **2014**, *141*, 024117. [[CrossRef](#)]
28. Kamimura, A.; Matsubara, Y.J.; Kaneko, K.; Takeuchi, N. Horizontal transfer between loose compartments stabilizes replication of fragmented ribozymes. *PLoS Comput. Biol.* **2019**, *15*, e1007094. [[CrossRef](#)]
29. Yoshiyama, T.; Ichii, T.; Yomo, T.; Ichihashi, N. Automated in vitro evolution of a translation-coupled RNA replication system in a droplet flow reactor. *Sci. Rep.* **2018**, *8*, 11867. [[CrossRef](#)]
30. Beneyton, T.; Krafft, D.; Bednarz, C.; Kleineberg, C.; Woelfer, C.; Ivanov, I.; Vidaković-Koch, T.; Sundmacher, K.; Baret, J.C. Out-of-equilibrium microcompartments for the bottom-up integration of metabolic functions. *Nat. Commun.* **2018**, *9*, 2391. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Bibliography

- [Aceituno et al., 2019] Aceituno, P. V., Rogers, T., and Schomerus, H. (2019). Universal Hypotrochoidic Law for Random Matrices with Cyclic Correlations. *Physical Review E*, 100(1).
- [Allesina and Tang, 2012] Allesina, S. and Tang, S. (2012). Stability Criteria for Complex Ecosystems. *Nature*, 483(7388):205–208.
- [Allesina and Tang, 2015] Allesina, S. and Tang, S. (2015). The Stability–Complexity Relationship at Age 40: a Random Matrix Perspective. *Population Ecology*, 57(1):63–75.
- [Bansho et al., 2016] Bansho, Y., Furubayashi, T., Ichihashi, N., and Yomo, T. (2016). Host–Parasite Oscillation Dynamics and Evolution in a Compartmentalized RNA Replication System. *Proceedings of the National Academy of Sciences*, 113(15):4045–4050.
- [Becker et al., 2019] Becker, S., Feldmann, J., Wiedemann, S., Okamura, H., Schneider, C., Iwan, K., Crisp, A., Rossa, M., Amatov, T., and Carell, T. (2019). Unified Prebiotically Plausible Synthesis of Pyrimidine and Purine RNA Ribonucleotides. *Science*, 366:76–82.
- [Blackmond, 2004] Blackmond, D. G. (2004). Asymmetric autocatalysis and its implications for the origin of homochirality. *Proceedings of the National Academy of Sciences*, 101(16):5732–5736.
- [Blackmond, 2010] Blackmond, D. G. (2010). The Origin of Biological Homochirality. *Cold Spring Harbor Perspectives in Biology*, 2.

- [Blanco and Hochberg, 2011] Blanco, C. and Hochberg, D. (2011). Chiral Polymerization: Symmetry Breaking and Entropy Production in Closed Systems. *Physical Chemistry Chemical Physics*, 13:839–849.
- [Blokhuys et al., 2018] Blokhuys, A., Lacoste, D., Nghe, P., and Peliti, L. (2018). Selection Dynamics in Transient Compartmentalization. *Physical Review Letters*, 120(15):158101.
- [Bolli et al., 1997] Bolli, M., Micura, R., and Eschenmoser, A. (1997). Pyranosyl-RNA: Chiroselective Self-Assembly of Base Sequences by Ligative Oligomerization of Tetranucleotide-2',3'-cyclophosphates (with a commentary concerning the origin of biomolecular homochirality). *Chemistry & Biology*, 4(4):309–320.
- [Brandenburg, 2021] Brandenburg, A. (2021). Homochirality: A Prerequisite or Consequence of Life? In *Prebiotic Chemistry and the Origin of Life*, Advances in Astrobiology and Biogeophysics, pages 87–115. Springer.
- [Brandenburg et al., 2005a] Brandenburg, A., Andersen, A. C., Höfner, S., and Nilsson, M. (2005a). Homochiral Growth Through Enantiomeric Cross-Inhibition. *Origins of Life and Evolution of Biospheres*, 35:225–241.
- [Brandenburg et al., 2005b] Brandenburg, A., Andersen, A. C., and Nilsson, M. (2005b). Dissociation in a Polymerization Model of Homochirality. *Origins of Life and Evolution of Biospheres*, 35:507–521.
- [Brandenburg and Multamäki, 2004] Brandenburg, A. and Multamäki, T. (2004). How Long can Left and Right Handed Life Forms Coexist? *International Journal of Astrobiology*, 3(3):209–219.
- [Cahn et al., 1966] Cahn, R. S., Ingold, C., and Prelog, V. (1966). Specification of Molecular Chirality. *Angewandte Chemie International Edition in English*, 5(4):385–415.
- [Chen and Ma, 2020] Chen, Y. and Ma, W. (2020). The Origin of Biological Homochirality Along with the Origin of Life. *PLOS Computational Biology*, 16(1).
- [Collet et al., 2006] Collet, A., Crassous, J., Dutasta, J.-P., and Guy, L. (2006). *Molécules chirales : Stéréochimie et propriétés*. EDP Sciences/CNRS Editions.
- [Cronin and Pizzarello, 1997] Cronin, J. R. and Pizzarello, S. (1997). Enantiomeric Excesses in Meteoritic Amino Acids. *Science*, 275(5302):951–955.

- [Damha et al., 1994] Damha, M. J., Giannaris, P. A., and Marfey, P. (1994). Antisense L/D-Oligodeoxynucleotide Chimeras: Nuclease Stability, Base-Pairing Properties, and Activity at Directing Ribonuclease H. *Biochemistry*, 33:7877–7885.
- [Damha et al., 1991] Damha, M. J., Giannaris, P. A., Marfey, P., and Reid, L. S. (1991). Oligodeoxynucleotides Containing Unnatural L-2'-deoxyribose. *Tetrahedron Letters*, 32(23):2573–2576.
- [Davidi et al., 2018] Davidi, D., Longo, L. M., Jabłońska, J., Milo, R., and Tawfik, D. S. (2018). A Bird's-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations. *Chemical Reviews*, 118(18):8786–8797.
- [de Marcellus et al., 2011] de Marcellus, P., Meinert, C., Nuevo, M., Filippi, J.-J., Danger, G., Deboffe, D., Nahon, L., Le Sergeant d'Hendecourt, L., and Meierhenrich, U. J. (2011). Non-Racemic Amino Acid Production by Ultraviolet Irradiation of Achiral Interstellar Analogs with Circularly Polarized Light. *The Astrophysical Journal Letters*, 727(2).
- [Eigen, 1971] Eigen, M. (1971). Selforganization of Matter and the Evolution of Biological Macromolecules. *Naturwissenschaften*, 58:465–523.
- [Engel and Nagy, 1982] Engel, M. H. and Nagy, B. (1982). Distribution and Enantiomeric Composition of Amino Acids in the Murchison Meteorite. *Nature*, 296:837–840.
- [Ernst, 2013] Ernst, K.-H. (2013). Molecular Chirality in Surface Science. *Surface Science*, 613:1–5.
- [Fink and Reymond, 2007] Fink, T. and Reymond, J.-L. (2007). Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F. *Journal of Chemical Information and Modeling*, 47:342–353.
- [Frank, 1953] Frank, F. C. (1953). On Spontaneous Asymmetric Synthesis. *Biochimica et Biophysica Acta*, 11:459–463.
- [Fujita, 2007a] Fujita, S. (2007a). Alkanes as Stereoisomers. Enumeration by the Combination of Two Dichotomies for Three-Dimensional Trees. *MATCH Communications in Mathematical and in Computer Chemistry*, 57:299.
- [Fujita, 2007b] Fujita, S. (2007b). Numbers of Monosubstituted Alkanes as Stereoisomers. *Journal of Computer Chemistry, Japan*, 6:59–72.

- [Furubayashi and Ichihashi, 2018] Furubayashi, T. and Ichihashi, N. (2018). Sustainability of a Compartmentalized Host-Parasite Replicator System under Periodic Washout-Mixing Cycles. *Life*, 8(1).
- [Gardner and Ashby, 1970] Gardner, M. R. and Ashby, W. R. (1970). Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability. *Nature*, 228(5273):784–784.
- [Gillespie, 1976] Gillespie, D. T. (1976). A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics*, 22(4):403–434.
- [Gillespie, 1977] Gillespie, D. T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Chemical Chemistry*, 81(25):2340–2361.
- [Gillespie, 2007] Gillespie, D. T. (2007). Stochastic Simulation of Chemical Kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55.
- [Ginibre, 1965] Ginibre, J. (1965). Statistical Ensembles of Complex, Quaternion, and Real Matrices. *Journal of Mathematical Physics*, 6(3):440–449.
- [Girko, 1985] Girko, V. L. (1985). Circular Law. *Theory of Probability & Its Applications*, 29(4):694–706. Publisher: Society for Industrial and Applied Mathematics.
- [Gleiser and Walker, 2008] Gleiser, M. and Walker, S. I. (2008). An Extended Model for the Evolution of Prebiotic Homochirality: A Bottom-Up Approach to the Origin of Life. *Origins of Life and Evolution of Biospheres*, 38:293–315.
- [Globus and Blandford, 2020] Globus, N. and Blandford, R. (2020). The Chiral Puzzle of Life. *The Astrophysical Journal Letters*, 895(L11).
- [Globus et al., 2021] Globus, N., Fedynitch, A., and Blandford, R. D. (2021). Polarized Radiation and the Emergence of Biological Homochirality on Earth and beyond. *The Astrophysical Journal*, 910(85).
- [Göppel et al., 2022] Göppel, T., Rosenberger, J., Altaner, B., and Gerland, U. (2022). Thermodynamic and Kinetic Sequence Selection in Enzyme-Free Polymer Self-Assembly inside a Non-equilibrium RNA Reactor. *Life*, 12(4):567–606.
- [Harris et al., 1999] Harris, A. B., Kamien, R. D., and Lubensky, T. C. (1999). Molecular Chirality and Chiral Parameters. *Reviews of Modern Physics*, 71(5):1745–1757.

- [Hauser et al., 2006] Hauser, N. C., Martinez, R., Jacob, A., Rupp, S., Hoheisel, J. D., and Matysiak, S. (2006). Utilising the left-helical conformation of L-DNA for analysing different marker types on a single universal microarray platform. *Nucleic Acids Research*, 34(18):5101–5111.
- [Higgs and Lehman, 2015] Higgs, P. G. and Lehman, N. (2015). The RNA World: Molecular Cooperation at the Origins of Life. *Nature Reviews*, 16.
- [Horeau and Guetté, 1974] Horeau, A. and Guetté, J. P. (1974). Interactions Diastéréoisomères d’Antipodes en Phase Liquide. *Tetrahedron*, 30:1923–1931.
- [Ianeselli et al., 2019] Ianeselli, A., Mast, C. B., and Braun, D. (2019). Periodic Melting of Oligonucleotides by Oscillating Salt Concentrations Triggered by Microscale Water Cycles Inside Heated Rock Pores. *Angewandte Chemie International Edition*, 58:13155.
- [Jafarpour et al., 2015] Jafarpour, F., Biancalani, T., and Goldenfeld, N. (2015). A Noise-Induced Mechanism for Biological Homochirality of Early Life Self-Replicators. *Physical Review Letters*, 115:158101.
- [Joshi et al., 2011] Joshi, P. C., Aldersley, M. F., and Ferris, J. P. (2011). Progress in Demonstrating Total Homochiral Selection in Montmorillonite-Catalyzed RNA Synthesis. *Biochemical and Biophysical Research Communications*, 413:594–598.
- [Joshi et al., 2006] Joshi, P. C., Pitsch, S., and Ferris, J. P. (2006). Selectivity of Montmorillonite Catalyzed Prebiotic Reactions of D, L-Nucleotides. *Origins of Life and Evolution of Biospheres*, 37:3–26.
- [Joyce et al., 1984] Joyce, G. F., Visser, G. M., van Boeckel, C. A. A., van Boom, J. H., Orgel, L. E., and van Westrenen, J. (1984). Chiral Selection in Poly(C)-Directed Synthesis of Oligo(G). *Nature*, 310:602–604.
- [Kawakami et al., 2005] Kawakami, J., Kazuhiko, T., and Sugimoto, N. (2005). Thermodynamic Analysis of Duplex Formation of the Heterochiral DNA with L - Deoxyadenosine. *Analytical Sciences*, 21:77–82.
- [Keil et al., 2017] Keil, L. M. R., Möller, F. M., Kiess, M., Kudella, P. W., and Mast, C. B. (2017). Proton Gradients and pH Oscillations Emerge from Heat Flow at the Microscale. *Nature Communications*, 8.

- [Kervio et al., 2016] Kervio, E., Sosson, M., and Richert, C. (2016). The Effect of Leaving Groups on Binding and Reactivity in Enzyme-Free Copying of DNA and RNA. *Nucleic Acids Research*, 44(12):5504–5514.
- [Kondepudi and Nelson, 1983] Kondepudi, D. K. and Nelson, G. W. (1983). Chiral Symmetry Breaking in Nonequilibrium Systems. *Physical Review Letters*, 50:1023–1026.
- [Laurent et al., 2022] Laurent, G., Gaspard, P., and Lacoste, D. (2022). A Robust Transition to Homochirality in Complex Chemical Reaction Networks. *Proceedings of the Royal Society A*, 478.
- [Laurent et al., 2021] Laurent, G., Lacoste, D., and Gaspard, P. (2021). Emergence of Homochirality in Large Molecular Systems. *Proceedings of the National Academy of Sciences*, 118(3).
- [Laurent et al., 2019] Laurent, G., Peliti, L., and Lacoste, D. (2019). Survival of Self-Replicating Molecules under Transient Compartmentalization with Natural Selection. *Life*, 9(4).
- [Li et al., 2021] Li, M., He, X., Chen, J., Wang, B., Liu, S., and Rong, C. (2021). Density Functional Theory and Information-Theoretic Approach Study on the Origin of Homochirality in Helical Structures. *Journal of Physical Chemistry A*, 125(5):1269–1278.
- [Liu, 2020] Liu, S. (2020). Homochirality Originates from the Handedness of Helices. *The Journal of Physical Chemistry Letters*, 11:8690–8696.
- [Marchenko and Pastur, 1967] Marchenko, V. A. and Pastur, L. A. (1967). Distribution of Eigenvalues for some Sets of Random Matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483.
- [Mariani et al., 2018] Mariani, A., Bonfio, C., Johnson, C. M., and Sutherland, J. D. (2018). pH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. *Biochemistry*, 57(45):6382–6386.
- [Mast and Braun, 2010] Mast, C. B. and Braun, D. (2010). Thermal Trap for DNA Replication. *Physical Review Letters*, 104:188102.
- [Matsubara et al., 2021] Matsubara, Y. J., Takeuchi, N., and Kaneko, K. (2021). Error Catastrophe Can Be Avoided by Proofreading Innate to Template-Directed Polymerization. *arXiv e-prints*, page arXiv:2108.09961.

- [Matsumara et al., 2016] Matsumara, S., Kun, A., Ryckelynck, M., Coldren, F., Szilágyi, A., Jossinet, F., Rick, C., Nghe, P., Szathmáry, E., and Griffiths, A. D. (2016). Transient Compartmentalization of RNA Replicators Prevents Extinction due to Parasites. *Science*, 354(6317):1293–1296.
- [May, 1972] May, R. (1972). Will a Large Complex System be Stable? *Nature*, page 413.
- [Meierhenrich, 2008] Meierhenrich, U. (2008). *Amino Acids and the Asymmetry of Life*. Springer.
- [Meinert et al., 2011] Meinert, C., de Marcellus, P., Le Sergeant d’Hendecourt, L., Nahon, L., Jones, N. C., Soren V., H., Hendrik Bredehöft, J., and Meierhenrich, U. J. (2011). Photochirogenesis: Photochemical Models on the Absolute Asymmetric Formation of Amino Acids in Interstellar Space. *Physics of Life Reviews*, 8(3):307–330.
- [Miller, 1953] Miller, S. L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science*, 117(3046):528–529.
- [Pasteur, 1848] Pasteur, L. (1848). Mémoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisation rotatoire. *Comptes Rendus Hebdomadaires des Séances de L’Academie des Sciences*, 21(26):535–538.
- [Pizzarello and Cronin, 2000] Pizzarello, S. and Cronin, J. R. (2000). Non-Racemic Amino Acids in the Murray and Murchison Meteorites. *Geochemica et Cosmochimica Acta*, 64(2):329–338.
- [Pizzarello and Weber, 2004] Pizzarello, S. and Weber, A. L. (2004). Prebiotic Amino Acids as Asymmetric Catalysts. *Science*, 303:1151.
- [Pizzarello and Weber, 2010] Pizzarello, S. and Weber, A. L. (2010). Stereoselective Syntheses of Pentose Sugars Under Realistic Prebiotic Conditions. *Origins of Life and Evolution of Biospheres*, 40:3–10.
- [Plasson et al., 2004] Plasson, R., Bersini, H., and Commeyras, A. (2004). Recycling Frank: Spontaneous Emergence of Homochirality in Noncatalytic Systems. *Proceedings of the National Academy of Sciences*, 101:16733–16738.
- [Plasson et al., 2007] Plasson, R., Kondepudi, D. K., Bersini, H., Commeyras, A., and Asakura, K. (2007). Emergence of Homochirality in Far-From-Equilibrium Systems: Mechanisms And Role in Prebiotic Chemistry. *Chirality*, 19:589–600.

- [Potters and Bouchaud, 2020] Potters, M. and Bouchaud, J.-P. (2020). *A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists*. Cambridge University Press.
- [Powner et al., 2009] Powner, M. W., Gerland, B., and Sutherland, J. D. (2009). Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature*, 459:239–242.
- [Quack, 2002] Quack, M. (2002). How Important is Parity Violation for Molecular and Biomolecular Chirality? *Angewandte Chemie International Edition*, 41(24):4618–4630.
- [Ratzke et al., 2020] Ratzke, C., Barrere, J., and Gore, J. (2020). Strength of Species Interactions Determines Biodiversity and Stability in Microbial Communities. *Nature Ecology & Evolution*, 4:376–383.
- [Rosanoff, 1906] Rosanoff, M. A. (1906). On Fischer’s Classification of Stereo-Isomers 1. *Journal of the American Chemical Society*, 28(1):114–141.
- [Rosenberger et al., 2021] Rosenberger, J. H., Göppel, T., Kudella, P. W., Braun, D., Gerland, U., and Altaner, B. (2021). Self-Assembly of Informational Polymers by Templated Ligation. *Physical Review X*, 11:031055.
- [Saghatelian et al., 2001] Saghatelian, A., Yokobayashi, Soltani, K., and Ghadari, M. R. (2001). A Chiroselective Peptide Replicator. *Nature*, 409:797–801.
- [Saito and Hyuga, 2005] Saito, Y. and Hyuga, H. (2005). Chirality Selection in Open Flow Systems and in Polymerization. *Journal of the Physical Society of Japan*.
- [Salditt et al., 2020] Salditt, A., Keil, L. M. R., Horning, D. P., Mast, C. B., Joyce, G. F., and Braun, D. (2020). Thermal Habitat for RNA Amplification and Accumulation. *Physical Review Letters*, 125:048104.
- [Sallembien et al., 2022] Sallembien, Q., Bouteiller, L., Crassous, J., and Matthieu, R. (2022). Possible Chemical and Physical Scenarios Towards Biological Homochirality. *Chemical Society Reviews*.
- [Sandars, 2003] Sandars, P. G. H. (2003). A Toy Model for the Generation of Homochirality During Polymerization. *Origins of Life and Evolution of the Biosphere*, 33:575–587.
- [SantaLucia and Hicks, 2004] SantaLucia, Jr., J. and Hicks, D. (2004). The Thermodynamics of DNA Structural Motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440.

- [Schrödinger, 1944] Schrödinger, E. (1944). *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press.
- [Smith and Szathmáry, 1995] Smith, J. M. and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford University Press.
- [Soai et al., 1995] Soai, K., Shibata, T., Morioka, H., and Choji, K. (1995). Asymmetric Autocatalysis and Amplification of Enantiomeric Excess of a Chiral Molecule. *Nature*, 378(6559):767–768.
- [Sommers et al., 1988] Sommers, H. J., Crisanti, A., Sompolinsky, H., and Stein, Y. (1988). Spectrum of Large Random Asymmetric Matrices. *Physical Review Letters*, 60(19):1895–1899.
- [Sosson and Richert, 2018] Sosson, M. and Richert, C. (2018). Enzyme-Free Genetic Copying of DNA and RNA Sequences. *Beilstein Journal Of Organic Chemistry*, 14:603–617.
- [Strogatz, 1994] Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos*. Perseus Books Publishing.
- [Szabat et al., 2016] Szabat, M., Gudanis, D., Kotkowiak, W., Gdaniec, Z., Kierzek, R., and Pasternak, A. (2016). Thermodynamic Features of Structural Motifs Formed by beta-L-RNA. *PLOS One*, 11(2).
- [Tkachenko and Maslov, 2015] Tkachenko, A. V. and Maslov, S. (2015). Spontaneous Emergence of Autocatalytic Information-Coding Polymers. *The Journal of Chemical Physics*, 143:045102.
- [Tupper et al., 2017] Tupper, A. S., Shi, K., and Higgs, P. G. (2017). The Role of Templating in the Emergence of RNA from the Prebiotic Chemical Mixture. *Life*, 7(41).
- [Turner and Mathews, 2009] Turner, D. H. and Mathews, D. H. (2009). NNDB: The Nearest Neighbor Parameter Database for Predicting Stability of Nucleic Acid Secondary Structure. *Nucleic Acids Research*, (38).
- [Urata et al., 2003] Urata, H., Shimizu, H., Hiroaki, H., Kohda, D., and Akagi, M. (2003). Thermodynamic Study of Hybridization Properties of Heterochiral Nucleic Acids. *Biochemical and Biophysical Communications*, 309:79–83.

- [Vantomme and Crassous, 2021] Vantomme, G. and Crassous, J. (2021). Pasteur and Chirality: A Story of How Serendipity Favors the Prepared Minds. *Chirality*, 33(10):597–601.
- [Walton and Szostak, 2017] Walton, T. and Szostak, J. W. (2017). A Kinetic Model of Nonenzymatic RNA Polymerization by cytidine-5'-phosphoro-2-aminoimidazole. *Biochemistry*, 56(43):5739–5747.
- [Weber and Pizzarello, 2006] Weber, A. L. and Pizzarello, S. (2006). The Peptide-Catalyzed Stereospecific Synthesis of Tetroses: A Possible Model for Prebiotic Molecular Evolution. *Proceedings of the National Academy of Sciences*, 103(34):12713–12717.
- [Wigner, 1955] Wigner, E. P. (1955). Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564.
- [Wigner, 1958] Wigner, E. P. (1958). On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics*, 67(2):325–327.
- [Wishart, 1928] Wishart, J. (1928). The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A(1-2):32–52.
- [Wu et al., 2012] Wu, M., Walker, S. I., and Higgs, P. G. (2012). Autocatalytic Replication and Homochirality in Biopolymers: Is Homochirality a Requirement of Life or a Result of It? *Astrobiology*, 12(9):818–828.

RÉSUMÉ

Les systèmes vivants sur Terre sont homochiraux. Cela signifie que pour chaque espèce chimique chirale qu'ils contiennent, la fraction d'un des deux énantiomères est très largement supérieure à celle de l'autre énantiomère. Depuis la découverte de la chiralité par Pasteur, l'homochiralité reste un mystère puisque aucun mécanisme jugé entièrement satisfaisant justifiant son émergence n'a été proposé à ce jour et que l'on ne sait pas non plus si l'homochiralité est un prérequis ou bien une conséquence de la vie. Dans cette thèse, nous proposons deux scénarios physiques au sein desquels l'homochiralité pourrait avoir émergé avant ou bien conjointement avec la vie. Dans un premier temps, nous montrons que les grands réseaux chimiques chiraux complexes sont susceptibles de subir une transition d'un état racémique à un état homochiral due à une brisure de symétrie spontanée, lorsque le nombre d'espèces chirales qu'ils contiennent devient important. Ce mécanisme robuste repose sur les propriétés mathématiques de grandes matrices aléatoires et n'impose que peu de contraintes sur le réseau chimique. Nous illustrons ce mécanisme avec une généralisation du modèle de Frank, avec un grand nombre d'espèces chimiques. Enfin, nous évaluons quantitativement l'abondance des molécules chirales dans la nature ainsi que le seuil au-delà duquel les espèces chirales dominent en nombre celles achirales à l'aide d'une analyse de bases de données moléculaires. Dans un second temps, nous présentons un mécanisme basé sur la polymérisation sur « template » de biopolymères tels que l'ARN, qui implique l'extension de brins d'ARN par ligature avec d'autres oligomères ou monomères compatibles par appariement par paire de base. Ce processus est autocatalytique et montre une inhibition chirale, qui sont tous deux des éléments clés pour induire une brisure de symétrie provoquant une transition vers un état homochiral. À l'aide de simulations stochastiques détaillées de polymérisation sur « template » dans des systèmes de polymères chiraux, nous étudions la propension de systèmes initialement composés d'un mélange racémique de monomères à évoluer vers des systèmes homochiraux de polymères en présence de réactions de racémisation. Deux types de réacteurs chimiques sous différentes conditions sont étudiés dans ce travail : d'abord des réacteurs fermés mais maintenus hors d'équilibre, avec un nombre constant de monomères, puis des réacteurs ouverts dans lesquels les espèces sont dégradées et certaines sont chémostatées. Nous supposons aussi que les systèmes sont soumis à des cycles de températures dans les deux cas. Nous trouvons qu'un état parfaitement homochiral est atteint par les systèmes fermés en présence de réactions de racémisation grâce au « chiral stalling » qui ralentit la ligature entre deux brins d'ARN lorsque deux monomères de chiralités opposées sont appariés proche de la position de la ligature. Il apparaît que cette transition homochirale permet au système d'atteindre une longueur moyenne plus élevée de polymères qui le composent, ce qui est difficile en général pour la polymérisation non-enzymatique. Enfin, les simulations de réacteurs ouverts montrent que seul un écart énantiomérique partiel et transitoire peut être atteint mais sans que des réactions de racémisation soient nécessaires. Le travail présenté dans cette thèse se focalise principalement sur le processus d'amplification d'un faible écart énantiomérique initial produit soit par un phénomène physique ou chimique particulier, soit simplement par des fluctuations statistiques.

MOTS CLÉS

Homochiralité, physique statistique, physique hors d'équilibre, matrices aléatoires, polymérisation sur template, origines de la vie

ABSTRACT

Living systems on earth are homochiral. This means that for every chiral species they contain, one of the two possible enantiomers is present in much higher fraction than its mirrored counterpart. Homochirality has continuously puzzled scientists ever since the discovery of chirality by Pasteur, because a mechanism for its emergence is not yet solved, nor is the question of whether homochirality is a prerequisite or a consequence of life. In this thesis, we propose two physical scenarios in which homochirality could have emerged prior to or alongside life. We first show that large and complex chiral chemical networks are subject to a symmetry breaking transition from a racemic state to a homochiral one as the number of chiral compounds they contain becomes large. This robust mechanism relies on properties of large random matrices and requires only a few constraints on the chemical network. It is illustrated with a generalization of the famous Frank model which contains a large number of chemical species. We also quantify how abundant chiral molecules are in nature through an analysis of molecular databanks which shows a threshold above which chiral compounds dominate achiral ones. In a second part, we present a scenario based on template-directed ligation of biopolymers such as RNA, which involves the extension of RNA polymers by ligation with other oligomers or monomer compatible with base pairing. This process presents autocatalysis and chiral inhibition which are two key ingredients for a symmetry breaking transition leading to a homochiral state. Using detailed stochastic simulations of template-directed ligation of chiral polymeric systems, we thus investigate the propensity of systems inoculated initially with a racemic mixture of RNA monomers to evolve towards a homochiral polymer system in the presence of racemization reactions. Two kinds of reactors and their different conditions are studied in this work: closed out-of-equilibrium reactors with a conserved number of RNA monomers and open reactors in which species are being degraded over time and some are chemostated. In addition, temperature cycles or dry-wet cycles are assumed to be present in both cases. We find that full homochirality is reached for closed systems in presence of racemization reactions due to chiral stalling, which slows ligation when opposite chiralities are paired closed to the ligation site. Remarkably, the homochirality transition helps the system to reach longer average polymer length, which is typically difficult in non-enzymatic polymerization. Open reactor simulations can only reach partial and transient enantiomeric excesses but without the need of racemization reactions. The work presented in this thesis focuses on the amplification process of a small initial enantiomeric excess imbalance generated by a particular physical or chemical phenomenon or simply by statistical fluctuations.

KEYWORDS

Homochirality, statistical physics, out of equilibrium physics, random matrices, template-directed polymerization, origins of life