



POLITECHNIKA POZNAŃSKA

Wydział Informatyki i Telekomunikacji

mgr inż. Jarosław Synak

**Modelowanie i symulowanie systemów
wieloagentowych opisujących procesy w hipotezie
Świata RNA**

Praca doktorska

Promotor: prof. dr hab. inż. Jacek Błazewicz

Poznań, 2023 r.

Podziękowania

Chciałbym podziękować wszystkim, którzy przyczynili się do powstania tej rozprawy.

Przede wszystkim dziękuję mojemu promotorowi prof. dr. hab. inż. Jackowi Błażewiczowi za pomoc merytoryczną i wsparcie. Dr. hab. inż. Agnieszce Rybarczyk chciałem podziękować za nieocenione porady dotyczące treści poszczególnych rozdziałów.

Dziękuję również Ani za cierpliwość i wyrozumiałość oraz wszystkim moim przyjaciołom i rodzinie za to, że byliście przy mnie w tym ważnym okresie.

Spis treści

Streszczenie	7
Short summary	9
1 Wstęp	11
1.1 Wkład metod informatycznych w badanie początków życia	12
1.2 Cel i zakres pracy	13
2 Podstawy matematyczne i informatyczne	15
2.1 Równania różniczkowe zwyczajne	15
2.1.1 Liniowe równania różniczkowe zwyczajne	17
2.2 Równania różniczkowe cząstkowe.	17
2.3 Złożoność obliczeniowa	18
2.4 Algorytmy wielowątkowe.	19
2.5 Automaty komórkowe.	20
2.6 Systemy wieloagentowe	20
3 Podstawy biologiczne i chemiczne	23
3.1 Podstawy dynamiki molekularnej	23
3.2 Kinetyka reakcji chemicznych	23
3.3 Problem początków życia	24
3.4 Hipoteza Świata RNA	25
3.4.1 Niedoskonała replikacja i paradoks Eigena	26
3.4.2 Model RP	26
3.4.3 Kompartmenty	28
3.5 Ślady geologiczne pozostawione przez wczesne stadia ewolucji komórek	28
3.6 Struktura przestrzenna RNA	29
3.7 Aktualny stan badań in vitro	29
3.7.1 Replikacja za pomocą tripletów	29
3.7.2 Powstawanie cegiełek do budowy RNA - rybonukleotydów	31
4 Metody analitycznego badania Świata RNA	33
4.1 Równanie quasi-gatunku.	33
4.2 Równanie replikatorowe	34
4.3 Równania różniczkowe oparte o hiper cykl	35
4.4 Automaty komórkowe.	36

5	Model wykorzystujący systemy wieloagentowe	37
5.1	Model biologiczny	37
5.2	Model matematyczny	37
5.2.1	Podstawowe parametry	37
5.2.2	Dyfuzja	38
5.2.3	Reakcje pierwszorzędowe	38
5.2.4	Reakcje drugorzędowe	39
5.2.5	Wartości parametrów	39
5.2.5.1	Dyfuzja	40
5.2.5.2	Stała replikacji	41
5.3	Algorytm	41
5.4	Implementacja	41
5.5	Wyniki	43
5.5.1	Mutacja pasożytów	44
5.5.2	Mutacja pasożytów i replikaz	48
5.6	Dyskusja	50
6	Model wieloagentowy z sekwencjami	51
6.1	Model biologiczny	51
6.2	Model matematyczny	52
6.2.1	Aktywność enzymatyczna	52
6.2.2	Prawdopodobieństwo zwinięcia - parametr l	52
6.3	Algorytm	52
6.4	Implementacja	53
6.5	Wyniki	55
6.5.1	Scenariusz 1	56
6.5.2	Scenariusz 2	56
6.5.3	Scenariusz 3	58
6.5.4	Scenariusz 4	58
6.6	Dyskusja	63
7	Model wykorzystujący równania różniczkowe cząstkowe	65
7.1	Model biologiczny	65
7.2	Model matematyczny	65
7.2.1	Replikacja	66
7.2.2	Rozpad	66
7.2.3	Powstawanie zasobów	66
7.2.4	Dyfuzja	66
7.2.5	Równania różniczkowe cząstkowe	67
7.2.6	Rozwiązania omawianych równań różniczkowych	67
7.2.6.1	Dobrze wymieszany roztwór	67
7.3	Algorytm	74
7.4	Implementacja	77
7.5	Wyniki	77
7.6	Dyskusja	89
8	Model wykorzystujący ewolucję syntaz	91
8.1	Model biologiczny	91
8.2	Model matematyczny	92
8.2.1	Aktywność katalityczna	92
8.2.2	Mutacja	92
8.2.3	Ocena populacji	93

8.2.4	Generowanie losowych sekwencji	93
8.3	Algorytm	93
8.4	Implementacja	94
8.5	Wyniki	94
8.6	Dyskusja	94
9	Model wykorzystujący inhibitory kontrolujące populację	97
9.1	Model biologiczny	97
9.2	Model matematyczny	98
9.2.1	Agent i inhibitor	98
9.2.2	Agent, inhibitor i antyinhibitor.	100
9.3	Algorytm	106
9.4	Implementacja	106
9.5	Wyniki	107
9.5.1	Symulacja w arkuszu kalkulacyjnym.	107
9.5.2	Symulacja inhibitorów działających na replikazy (bez pasożytów)	107
9.5.3	Symulacja wieloagentowa	107
9.6	Dyskusja	109
10	Podsumowanie	113
	Bibliografia	117

Streszczenie

Jednym z najbardziej prawdopodobnych scenariuszy powstania życia na Ziemi jest hipoteza Świata RNA. Sugeruje ona, że zanim pojawiły się białka i DNA, cząsteczki RNA mogły być zarówno nośnikami informacji jak i enzymami. Cząsteczki zdolne do pełnienia wspomnianych funkcji musiały być w jakiś sposób replikowane, żeby przekazać swoje funkcjonalne sekwencje następnemu pokoleniu. Zgodnie z opisywaną tu hipotezą, niektóre populacje łańcuchów RNA zdołały w ten sposób przetrwać i z czasem ewoluowały. Proces ten doprowadził w końcu do powstania pierwszej współczesnej komórki, z informacją genetyczną zapisaną w DNA oraz wydajnymi białkowymi enzymami.

Celem niniejszej rozprawy jest odkrycie praw rządzących ewolucją populacji cząsteczek RNA oraz udowodnienie, że ich przetrwanie i ewolucja są możliwe. W tym celu wykorzystano wiele metod z których najważniejsze to równania różniczkowe i symulacje wieloagentowe. Za ich pomocą testowane są liczne scenariusze z różnymi zestawami parametrów, co pozwala na obserwację wielu procesów w najdrobniejszych szczegółach i wyciągnięcie istotnych wniosków.

Rozprawa rozpoczyna się stosownym wstępem, w którym wprowadzona jest hipoteza Świata RNA oraz bioinformatyka jako dziedzina badań. Następnie przedyskutowany jest główny cel przedstawionych w dalszej części badań.

Kolejne dwa rozdziały stanowią podstawę teoretyczną, wprowadzając pojęcia, które są ważne z punktu widzenia prowadzonych przez autora badań. Ze względu na interdyscyplinarność niniejszej rozprawy, zaprezentowane tematy należą do czterech dziedzin: matematyki, informatyki, biologii i chemii.

W następnym rozdziale przedstawione są poprzednie modele teoretyczne Świata RNA. Są one ważne, ponieważ wiele modeli przedstawionych w dalszej części rozprawy zostało nimi zainspirowane, szczególnie przez automat komórkowy zaproponowany przez Takeuchiego i Hogeweg, wspomniany niezliczoną liczbę razy w kolejnych rozdziałach.

Po wprowadzeniu do teorii, opisana jest pierwsza symulacja - symulacja wieloagentowa. Była ona w dużej mierze bazowana na automacie komórkowym Takeuchiego - cząsteczki RNA zostały podzielone na dwie grupy: replikazy zdolne do katalizowania replikacji innych cząsteczek oraz pasożyty pozbawione tej zdolności. Pasożyty mogły też zmieniać swoją efektywność w wyniku mutacji. Była jednak istotna różnica między dwoma modelami - symulowana przestrzeń była ciągła, co lepiej odpowiadało rzeczywistości niż dyskretne komórki w automacie komórkowym. Wyniki okazały się być podobne dla obu opisanych tutaj podejść - pasożyty zwiększały swoją efektywność jako wzorce do replikacji, ale częściej też przebywały w stanie nieaktywnym (w którym nie mogły być replikowane). Te dwa procesy działały przeciwstawnie, co pozwalały przeżyć populacji - zbyt efektywne pasożyty mogą wyprzeć replikazy, dlatego są rzadziej aktywne aby te drugie również miały szansę na zreplikowanie się.

Kolejny model został stworzony w oparciu o wspomnianą wyżej symulację wieloagentową, rozszerzając ją o sekwencje - wszystkie parametry pojedynczego agenta były teraz na jej podstawie

obliczane. Naturalną konsekwencją tej zmiany była możliwość powstawania pasożytów ze zmutowanych replikaz, co nie mogło mieć miejsca w poprzednim modelu. Najważniejszym wnioskiem z przeprowadzonych w dalszej kolejności symulacji był fakt, że replikazy mogły samoczynnie wyłonić się z grupy losowych sekwencji.

W następnym rozdziale opisano odmienne podejście od poprzednich - populacja RNA została zamodelowana za pomocą równań różniczkowych cząstkowych. Wzięcie pod uwagę aspektu przestrzennego pozwoliło autorowi na uwzględnienie dyfuzji. Opis zaczyna się od analizy teoretycznej scenariusza w którym nie ma mutacji RNA, przedyskutowane są potencjalne rozwiązania i ich stabilność. Następnie wprowadzona jest mutacja pasożytów, a wspomniane powyżej równania zostają rozwiązane numerycznie. Otrzymane wyniki zostają dodatkowo zweryfikowane za pomocą systemu wieloagentowego, co pokazuje, że rezultaty są uniwersalne i nie zależą od konkretnej metody. Na końcu rozdziału autor formułuje ważną hipotezę - populacje cząsteczek RNA mogą ochronić się przed zagrożeniem ze strony pasożytów, przechodząc naprzemiennie przez fazy wzrostu i wymierania. Zbyt aktywne pasożyty giną w drugiej z nich, pozostawiając tylko RNA nie zagrażające replikazom.

Kolejny rozdział skupia się na innym procesie, który prawdopodobnie miał miejsce w Świecie RNA - nabywanie dodatkowych funkcji przez pasożyty. Autor proponuje syntezę nukleotydów jako wartą zbadania. Ważnym założeniem jest tutaj obecność błon lipidowych oddzielających poszczególne populacje RNA, które nie występowały w poprzednich modelach. Cały powyższy system został zasymulowany za pomocą programu w języku C++, stworzonym specjalnie do tego celu. Wszystko modelowane jest z użyciem jednowymiarowej tablicy, która przechowuje sekwencje wszystkich cząsteczek w symulacji. Wyniki pokazują, że cząsteczki mogą być wielofunkcyjne, co pomaga zminimalizować całkowitą ilość informacji genetycznej, którą populacja musi utrzymywać, ułatwiając przetrwanie mimo częstych mutacji RNA.

Ostatnim zbadanym w tej rozprawie modelem jest Świat RNA z inhibitorami. Autor proponuje krótkie sekwencje RNA jako dodatkowy mechanizm kontrolujący stężenia poszczególnych typów RNA w populacji. Mogą one ograniczać replikację niektórych enzymów, wiążąc się z nimi na zasadzie komplementarności. Problem jest analizowany z punktu widzenia kinetyki chemicznej, udowodnione jest istnienie i stabilność rozwiązania. Otrzymane wyniki są później weryfikowane za pomocą trzech różnych metod symulacji: numerycznego rozwiązania równań różniczkowych, arkusza kalkulacyjnego oraz systemu wieloagentowego (bazującego na poprzednich rozdziałach). Z symulacji zostaje wyciągnięty wniosek, że inhibitory mogą z powodzeniem stabilizować populację łańcuchów RNA i chronić ją przed wymarciem.

Niniejsza rozprawa kończy się podsumowaniem w którym zestawione są wszystkie analizowane wcześniej modele oraz otrzymane wyniki.

Short summary

One of the most probable scenarios for the emergence of first life on Earth is RNA World. This hypothesis suggests that before DNA and proteins evolved - RNA molecules functioned as both genetic information carriers and enzymes. RNA capable of performing the aforementioned functions had to replicate to pass their functional sequences to next generations of molecules. According to the hypothesis, some populations of RNAs survived and started to evolve. This led to the eventual emergence of the first modern cell, with its genetic information contained in DNA and utilising protein enzymes.

The goal of this thesis is to discover the rules governing RNA population evolution and to prove that its survival and long-term evolution are possible. Various methods are utilised to achieve this objective, the most important being differential equations and multi-agent simulations. Many different scenarios with different sets of parameters are tested in order to observe various processes in great detail and draw meaningful conclusions.

The thesis begins with an introduction to the subject, where the RNA World hypothesis and the field of bioinformatics itself are presented and the overarching goal of the author's research is discussed.

Next two chapters are the theoretical basis, introducing the most important subjects relevant to the research presented in the following chapters. Since this thesis is inherently interdisciplinary in scope, the concepts which are discussed belong to four different areas of knowledge: mathematics, computer science, biology and chemistry.

In the following chapter the previous theoretical models of RNA World are described. The importance of these stems from the fact that many models developed by the author are inspired by them, especially the Takeuchi and Hogeweg cellular automata, which is mentioned countless times in the second half of the thesis.

After the introduction to the theory, the first simulation is described - multi-agent system. It was heavily based on Takeuchi's cellular automata - RNA molecules were divided into two groups: replicases capable of catalysing replication of other molecules and parasites which did not possess this ability. Parasites could change their effectiveness due to mutation. The main difference however was the continuous simulation space, which was closer to the reality than discrete cells in cellular automata. The results were still similar for the two mentioned approaches - parasites became more efficient replication templates, but also spent more time being inactive (unable to be replicated). These two processes worked in opposition to each other to keep the RNA population alive - too efficient parasites can drive replicases to extinction, so they are rarely active to give replicases a chance to replicate themselves.

The next model which was developed was based on the previous multi-agent system, but with the addition of sequences - all parameters of a single agent were derived from its explicitly represented RNA sequence. As a natural consequence of this change, parasites could arise from mutated replicases, which was impossible earlier. The most important conclusion of the simula-

tions which were carried out was that replicases could arise spontaneously from a group of random RNA molecules.

In the next chapter, a different approach is described - modeling an RNA molecule population using partial differential equations. Taking spacial aspect into account allowed the author to include diffusion. The chapter starts with a theoretical analysis of the scenario without RNA mutation - potential solutions and their stability is discussed. In the following part, parasite mutation is introduced and the equations are solved numerically. The results are later confirmed using multi-agent systems to show that the conclusions are general and not specific to a particular method. At the end of the section an important hypothesis is formulated - RNA populations can mitigate a threat posed by growing groups of parasites by going through recurring phases of growth and extinction. Too efficient parasites die out during the extinction, leaving behind only RNAs which do not disrupt replicases.

The following chapter focuses on a different process (probably) occurring in the RNA World - acquiring additional functions by parasites. The author proposes nucleotide synthesis as one worth investigating. The important assumption is the presence of lipid membranes separating RNA molecule populations, which were not present in previous models. The whole system is then simulated using a program in C++ written specifically for that purpose. Everything is modeled using a one-dimensional array, which stores the sequences of all molecules in the simulation. The results show that molecules can be multifunctional, which helps to minimize the total amount of genetic information which has to be kept in the population, making it easier to survive against rapid RNA mutations.

The last model which is investigated in this thesis is the RNA World with inhibitors. The author proposes short RNA sequences as an additional means of control of the concentration of different types of RNA in the population. They can limit replication of some enzymes by binding to them based on Watson-Crick pairing. The problem is analysed from the point of view of chemical kinetics, proving the existence of a stable solution. The results are verified using three different simulation methods - numerical solutions of differential equations, by using spreadsheets and with a mutli-agent simulation based on previous chapters. The conclusion is that inhibitors can succesfully stabilise an RNA molecule population and save it from extinction.

The thesis is ended with a summary, summing up all the models presented and the results obtained.

Wstęp

Pytania dotyczące powstania świata towarzyszą ludzkości od zarania dziejów. Przez tysiąclecia każda kultura postulowała własną wersję wydarzeń, wierzone, że Ziemia została wyłoniona z chaosu, stworzona z ciała upadłego giganta lub powstała dzięki oddzieleniu *yin* i *yang*. Przez większą część historii chrześcijaństwa (do XVIII a nawet XIX wieku) dominował kreacjonizm, czyli pogląd mówiący, że Bóg w pewnym momencie powołał świat do życia w obecnej formie. Jednak początki rozwoju geologii i współczesnej teorii ewolucji nie potwierdzały tej wersji. Naukowcy zaczęli znajdować coraz więcej dowodów na to, że Ziemia nie pojawiła się z dnia na dzień, a raczej została ukształtowana przez miliardy lat najróżniejszych procesów fizykochemicznych, podobnie zresztą jak cały znany wszechświat. Ostatecznie dzięki ogromnemu postępowi wielu nauk, stworzono wiarygodne teorie tłumaczące powstawanie pierwiastków, formowanie planet, pochodzenie poszczególnych minerałów, kształtowanie się kontynentów, zmienność klimatu oraz ewolucję. Jednak jeden z etapów powstawania świata do dziś pozostaje zagadką - pojawienie się pierwszych żywych organizmów. W pewnym momencie z materii nieożywionej musiał samoczynnie powstać system zdolny do samopowieliania się, który w toku ewolucji przekształcił się w pierwsze komórki. Samoorganizacja, która musiała temu towarzyszyć, doprowadzała do lokalnego spadku entropii. Na pierwszy rzut oka jest to niezgodne z drugą zasadą termodynamiki, która postuluje że entropia zamkniętego systemu nigdy nie maleje, jednak żywe organizmy nie są takimi. Podobnie jak komputery (a dokładniej ich pamięć), mogą one zwiększać stopień wewnętrznego uporządkowania kosztem emisji ciepła na zewnątrz, przez co entropia całego systemu wzrasta, w ten sposób zasady termodynamiki nie są łamane. Jedną z ważnych hipotez dotyczących problemów powstania i ewolucji życia jest Świat RNA.

Temat Świata RNA jest rozpatrywany przez wielu badaczy z bardzo różnych perspektyw odkąd pierwszy raz w literaturze pojawiła się wzmianka o nukleotydach jako pierwszych cegiełkach życia [91]. Bardzo ważną częścią są analizy bazujące na kosmologii w celu oszacowania prawdopodobieństwa zajścia procesów prowadzących do powstania życia, przykładowo prowadzone przez Totaniego [117]. Informacja biologiczna musiała mieć swój początek w procesie abiogenezy, jednak z czysto statystycznego punktu widzenia powstanie jej w sposób losowy było mało prawdopodobne, przynajmniej biorąc pod uwagę obserwowalny wszechświat, jednym z rozwiązań mogłoby oczywiście być założenie, że jego rzeczywisty rozmiar jest znacznie większy [117]. Analizy statystyczne mają też ścisły związek z teorią informacji i spostrzeżeniem, że metody stosowane do analizy sekwencji bitów aplikują się również do sekwencji genetycznych [3, 69, 74, 97]. Rozważania na temat sposobu przekazywania informacji biologicznej, koniecznego dla przetrwania życia, doprowadziły do sformułowania tzw. „paradoksu Eigena”. Jego istotą jest fakt, że efektywna maszyna kopiująca sama wymaga odpowiedniej instrukcji do jej zbudowania - czyli odpowiedniej ilości informacji. Jak w takim razie wyjaśnić możliwość stopniowej ewolucji takiej maszyny [41]? Smith podał przypuszczalną wartość progę potrzebną do efektywnego powielania informacji i dalszej ewolucji - około 100 nukleotydów [104]. Sam Eigen [39] proponował, żeby obejść ten problem używając tzw.

„hipercyklu” w którym istniałoby kilka osobnych fragmentów sekwencji, które nawzajem by się kopiowały [39, 41, 111]. Możliwość ta była szczegółowo badana, ustalono że chociaż w modelach opartych na równaniach różniczkowych rozwiązanie to nie działa zbyt dobrze, ograniczona dyfuzja może prowadzić do utworzenia stabilizujących system struktur przestrzennych [18, 19, 34, 111]. Zaobserwowano m.in. tendencję do formowania się spiral [18, 19]. Uproszczoną wersją hipercykli był model RP (replikaza-pasożyt) wykorzystany wielokrotnie w tej pracy, a zaproponowany przez Hogeweg i Takeuchiego [115]. Badano zachowanie takiego systemu w oparciu o równania różniczkowe zwyczajne [11, 115], cząstkowe [106], automaty komórkowe [107, 115, 118], systemy wieloagentowe [105, 106, 110] oraz eksperymenty *in vitro* [9]. Ważnym wnioskiem ze wszystkich wymienionych podejść był fakt, że dużą rolę w przetrwaniu systemu odgrywa klastrowanie (spontaniczne lub wymuszone), dzięki któremu replikazy mogą oprzeć się działaniu pasożytów. Warto wspomnieć, że przeprowadzono wiele innych eksperymentów badających dynamikę ewolucyjną i tworzenie funkcjonalnego RNA, nieopartych stricte o model RP [7, 42, 77, 121, 125]. Wykorzystanie RNA do tego typu eksperymentów jest dodatkowo uzasadnione jego różnorodnymi właściwościami katalitycznymi [25, 48, 121, 123] oraz przesłankami wskazującymi na to, że współczesne polimerazy DNA mogą pochodzić od polimeraz RNA [37], a te drugie mogły naturalnie wyeowoluować od wspólnego przodka [42, 47, 80]. Udało się również zaobserwować *in vitro*, że kwas rybonukleinowy jest w stanie się dość szybko różnicować i specjalizować, jednak maszyna replikacyjna musiała być cały czas wprowadzana zewnątrz [47, 80]. Przeprowadzono także eksperyment z systemem CATCH zdolnym do samoreplikacji w celu zbadania w jaki sposób cząsteczki mogą ewoluować w kooperacji, jednak RNA było jedynie jedną z cegiełek, razem z DNA i białkowymi enzymami. Niemniej udało się pokazać, że taki układ wyraźnie dąży do jak największej wydajności [42].

Badaniom nad populacjami cząsteczek towarzyszą także badania ścieżek reakcji chemicznych, które wyjaśniłyby obecność odpowiednich związków organicznych (np. rybonukleotydów) i proces spontanicznej polimeryzacji w prebiotycznej zupie. Zanim powstały enzymy zdolne do katalizowania potrzebnych reakcji, musiały one przebiegać spontanicznie (lub innymi drogami). Pearce et al. [88] teoretyzują (popierając to stosownym modelem teoretycznym), że zasady azotowe potrzebne do syntezy nukleotydów były dostarczane przez meteoryty, łącząc się z resztami fosforowymi (pochodzącymi z minerałów) oraz rybozą (będącą wynikiem tzw. reakcji formozowej) dostępnymi w ciepłych sadzawkach na Ziemi. Wysokie temperatury z kolei mogły promować polimeryzację i powstanie pierwszych RNA [88]. Na drodze eksperymentów laboratoryjnych zaproponowano także możliwość syntezy odpowiednich reszt azotowych na Ziemi, podając możliwe ścieżki reakcji [64].

Ostatecznie można wymienić trzy najważniejsze, według autora, argumenty przemawiające za istnieniem w pewnym momencie Świata RNA:

1. w trakcie ekspresji genu, informacja z DNA musi zostać najpierw przepisana na RNA (mRNA) [5],
2. istnieją wirusy RNA, które nie zawierają w ogóle DNA (retrowirusy) [134],
3. RNA jest ważnym składnikiem rybozymów, pełni również funkcje regulatorowe we współczesnych komórkach [122].

1.1 Wkład metod informatycznych w badanie początków życia

Współcześnie naukę zajmującą się analizowaniem danych biologicznych z wykorzystaniem metod informatycznych nazywa się bioinformatyką. Jest to jedna z prężnie rozwijających się dziedzin, a dokonywane w niej odkrycia mają duże znaczenie zarówno dla informatyki jak i biologii. Dobrym przykładem jest chociażby sekwencjonowanie DNA, które polegało na składaniu wyjściowej sekwencji ze stosunkowo krótkich fragmentów. Genomy niektórych organizmów liczą nawet mi-

liardy par zasad, co uniemożliwia ich poprawne złożenie bez pomocy komputerów. Problem ten wymagał stworzenia odpowiednich algorytmów [71]. Duże znaczenie miała tutaj teoria grafów, gdyż to za ich pomocą można było reprezentować zależności między poszczególnymi odczytami. Zainicjowało to znaczący rozwój w tej dziedzinie, zdefiniowano nowe rodzaje grafów (grafy DNA, grafy etykietowalne)[16, 59], a badanie ich pozwoliło na odkrycie wielu interesujących zależności, np. okazało się, że cykl Hamiltona w skierowanych grafach liniowych można znaleźć w czasie wielomianowym [4, 16, 17].

Powyższy przykład nie jest odosobniony i oddaje ogromne znaczenie bioinformatyki w nauce, w ciągu kilkudziesięciu lat jej istnienia stworzono niezliczone narzędzia informatyczne mające wspomóc badania biologiczne, ich zastosowania to między innymi:

1. wyszukiwanie podobieństw w sekwencjach DNA (algorytm Needlemana-Wunscha i BLAST) [71],
2. predykcja struktury 3D białek na podstawie sekwencji aminokwasów [75],
3. predykcja struktury 3D RNA na podstawie sekwencji rybonukleotydów [15],
4. zrozumienie mechanizmu działania chorób [71],
5. predykcja interakcji grupy białek (problem dokowania) [71].

W samych badaniach nad Światem RNA metody informatyczne pozwoliły na symulację dynamiki ewolucyjnej (np. [114]) oraz poszukiwanie możliwych ścieżek reakcji abiotycznych za pomocą metod chemii kwantowej [76].

1.2 Cel i zakres pracy

Celem niniejszej rozprawy jest wykorzystanie podejść informatycznych oraz matematycznych w celu modelowania oraz zbadania założeń, ograniczeń i zakresu stosowalności hipotezy Świata RNA. Ważną częścią jest poszukiwanie ogólnych praw rządzących populacjami cząsteczek RNA, które mogłyby aplikować się dla jak najszerszego zakresu parametrów i które działałyby dla wielu przyszłych modeli.

Dużym problemem jest oczywiście brak bezpośrednich danych eksperymentalnych, gdyż w chwili pisania tej pracy nikomu nie udało się odtworzyć w laboratorium procesów prowadzących do powstania żywych komórek. Można opierać się jednak na wynikach częściowych, dotyczących interakcji między łańcuchami RNA, ich wrażliwości na mutacje oraz trwałości przy braku enzymów, które potrafią rozkładać kwasy rybonukleinowe. Równie użyteczną informacją jest sposób funkcjonowania współczesnych organizmów, a konkretnie sposób przekazywania informacji genetycznej, jak również aktywność enzymatyczna oraz regulatorowa cząsteczek RNA. Wszystkie procesy zachodzące dzisiaj wewnątrz komórki musiały mieć kiedyś swój początek, a część z nich mogła być obecna w swojej pierwotnej wersji już w Świecie RNA.

Główną metodą badań stosowaną w tej rozprawie są symulacje komputerowe, które mają na celu badanie wpływu poszczególnych parametrów na przeżywalność populacji łańcuchów RNA. Ważną informacją jest też ewolucja takiego systemu w czasie oraz rozkład poszczególnych typów RNA w przestrzeni, daje to wgląd w procesy które mogły zachodzić i które zwiększały szansę na przetrwanie oraz późniejszą ewolucję w bardziej złożone formy życia. Najczęściej stosowanym w tej pracy modelem symulacji są systemy wieloagentowe. Główną ich zaletą jest duża elastyczność, gdyż symulowane obiekty (agenty) mogą dzielić się na wiele rodzajów, a ich zachowanie można opisywać dość złożonymi algorytmami. Ze względu na ciągłą przestrzeń symulacji to podejście jest znacznie bliższe rzeczywistości niż np. automaty komórkowe [110]. W przypadku Świata RNA, agenty mogą reprezentować pojedyncze nukleotydy (i inne mniejsze cząsteczki), całe łańcuchy RNA lub wręcz protokomórki/kompartmenty, wszystko zależy od wyboru skali. Opisana wyżej metoda badawcza nie jest jedyną zastosowaną tutaj, autor próbował też niekiedy stosować podejście analityczne,

modelując dynamikę systemu za pomocą równań różniczkowych cząstkowych i wyprowadzać z nich ogólne zależności. Ważnym punktem była też weryfikacja zgodności wspomnianych równań z modelem wieloagentowym (jeśli dany model uwzględniał rozkład w przestrzeni) - dla poprawnego modelu każda metoda powinna dawać podobne wyniki. Warto podkreślić, że wyniki symulacji mogą stanowić przekonujący argument za daną hipotezą, nawet jeśli mają mniejszą wartość od wyników doświadczalnych lub znalezisk archeologicznych.

Układ pracy jest następujący. Rozdział 2 jest wstępem teoretycznym i wprowadza podstawowe definicje matematyczne i informatyczne stosowane w późniejszej części. Rozdział 3 ma podobną funkcję - zawarte są w nim najważniejsze informacje z zakresu chemii i biologii, które są konieczne do zrozumienia badań przedstawionych w niniejszej pracy. Rozdział 4 jest poświęcony już *stricto* Światu RNA, stanowi on krótki przegląd wcześniejszych metod wykorzystywanych do jego modelowania. Opis przeprowadzonych przez autora badań i ich wyników rozdzielony jest na: rozdział 5 (systemy wieloagentowe), rozdział 6 (systemy wieloagentowe z uwzględnieniem sekwencji), rozdział 7 (równania różniczkowe cząstkowe), rozdział 8 (wprowadzenie syntaz) oraz rozdział 9 (Świat RNA z inhibitorami). Całość zakończona jest stosownym podsumowaniem w którym omówione są uzyskane wyniki.

Podstawy matematyczne i informatyczne

Niniejszy rozdział zawiera podstawowe informacje matematyczne i informatyczne, które zostaną wykorzystane w dalszej części pracy. Przede wszystkim obowiązuje tu pewien podział - pierwsze dwa podrozdziały poświęcone są układom ciągłym, a dokładniej podstawowym równaniom różniczkowym wykorzystywanym do ich badania, podczas gdy dalsza część dotyczy układów dyskretnych. W badaniach opisanych w niniejszej rozprawie wykorzystywane są modele oparte o oba te pojęcia. Opisy układu oparte jednocześnie o modele ciągłe oraz dyskretne nie są niczym nowym, np. w fizyce kwantowej jedną z podstaw jest dualizm korpuskularno-falowy - obiekty w małej skali mogą zachowywać się jak cząstki z wyraźnie określonymi porcjami energii (układ dyskretny), a jednocześnie posiadają też własności fali (układu ciągłego) [24].

2.1 Równania różniczkowe zwyczajne

W nauce szerokie zastosowanie znalazł rachunek różniczkowy, który został rozwinięty niezależnie przez Isaaca Newtona i Gottfrieda W. Leibniza w XVII wieku. Dał on narzędzia do badania i opisu szerokiej klasy zjawisk i stał się nieodłączną częścią m.in. fizyki, chemii, biologii i ekonomii. Dzięki niemu możliwe stało się np. modelowanie ruchu promów kosmicznych lub ciśnienia wywieranego na tamę przy podnoszącym się poziomie wody [12]. Współcześnie zaawansowane metody rachunku różniczkowego i całkowego są stosowane w ogólnej teorii względności, a równanie Einsteina będące jej podstawą, sprowadza się do szeregu równań różniczkowych cząstkowych:

$$G_{\alpha\beta} + \Lambda g_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta} \quad (2.1)$$

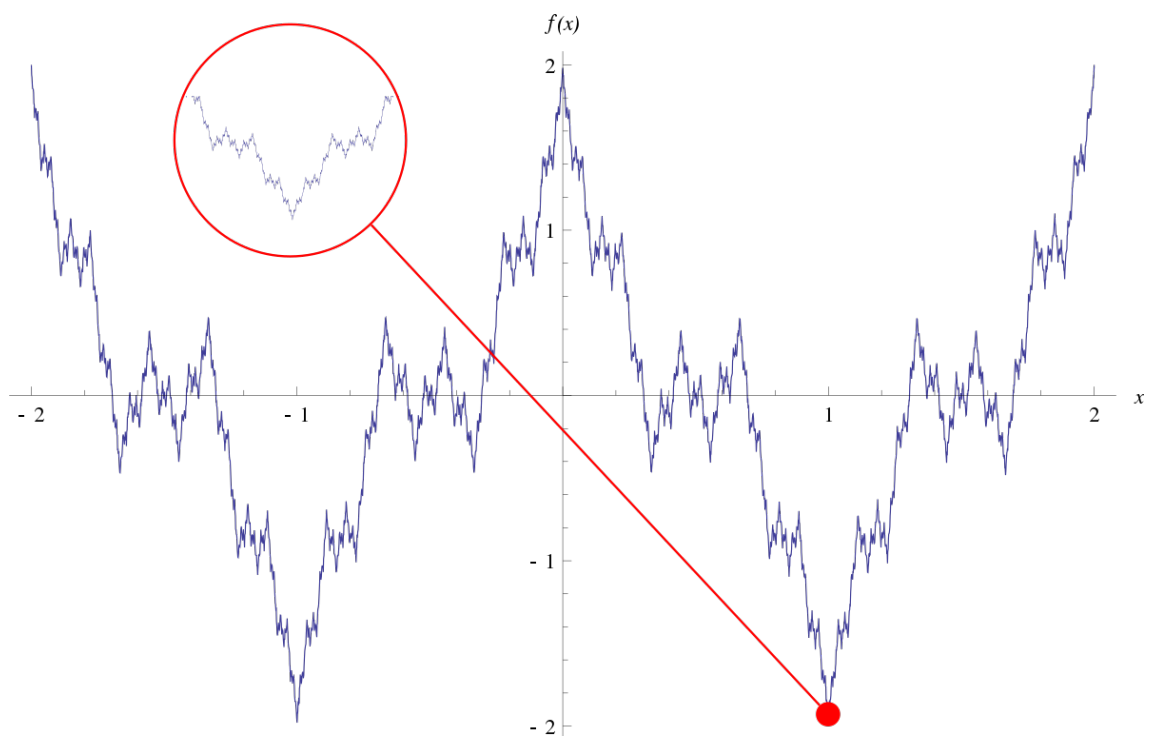
gdzie $G_{\alpha\beta}$ to tensor Einsteina, $g_{\alpha\beta}$ to tensor metryczny, $T_{\alpha\beta}$ - tensor napięć-energii, Λ - stała kosmologiczna, G - stała grawitacji, a c jest prędkością światła [51].

Formalnie pochodną funkcji $f(x)$ można zdefiniować jako granicę:

$$f'(x) = \dot{f}(x) = \frac{df(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (2.2)$$

Jeśli powyższa granica istnieje i ma skończoną wartość, mówi się, że funkcja f jest różniczkowalna w punkcie x [65]. Warto podkreślić, że każda funkcja różniczkowalna jest ciągła, ale twierdzenie odwrotne nie zachodzi. Dobrym kontrprzykładem jest chociażby funkcja Weierstrassa (Rys. 2.1), która jest określona na zbiorze liczb rzeczywistych i w każdym punkcie swojej dziedziny jest ciągła, ale nie różniczkowalna [58].

Można obliczyć pochodną z pochodnej w danym punkcie, powtarzając tę operację otrzymuje



Rys. 2.1: Funkcja Weierstrassa jest fraktalem, oznacza to, że dowolny jej fragment jest nieskończenie skomplikowany. Pochodna nie istnieje w żadnym jej punkcie, można jednak udowodnić ciągłość [58].

Źródło: <https://commons.wikimedia.org/wiki/File:WeierstrassFunction.svg> [dostęp 2022-08-18], Domena publiczna

się pochodne wyższych rzędów. Pochodną n -tego rzędu oznacza się zwyczajowo $f^{(n)}$ [65]. W zastosowaniach wartości pochodnych są często określane równaniami, które ograniczają dopuszczalną przestrzeń rozwiązań tylko do tych, które są interesujące z punktu widzenia danego modelu. Stopień równania jest równy najwyższemu rzędowi pochodnej, który w nim występuje. Równanie różniczkowe zwyczajne n -tego rzędu dla funkcji $f(x)$ można zapisać w sposób ogólny:

$$F(x, f(x), f^{(1)}(x), f^{(2)}(x), \dots, f^{(n)}(x)) = 0 \quad (2.3)$$

F jest oczywiście dowolną funkcją. Warto podkreślić, że rozwiązaniem powyższego równania nie są liczby, ale rodzina funkcji, zwykle określona parametrycznie. W celu uzyskania jednoznacznego rozwiązania, definiuje się tzw. warunki brzegowe, najczęściej są to wartości funkcji oraz niektórych jej pochodnych w ustalonych punktach [103]. Przykładowo jeśli równanie modeluje zmianę jakiejś wartości w czasie - naturalnymi warunkami brzegowymi jest wartość funkcji oraz jej pochodnych na samym początku.

W bardziej skomplikowanych przypadkach ma się do czynienia z układami równań różniczkowych zwyczajnych [103], wtedy funkcje f oraz F stają się wektorami:

$$\mathbf{F}(x, \mathbf{f}(x), \mathbf{f}^{(1)}(x), \mathbf{f}^{(2)}(x), \dots, \mathbf{f}^{(n)}(x)) = 0 \quad (2.4)$$

2.1.1 Liniowe równania różniczkowe zwyczajne

Spośród opisanych powyżej równań różniczkowych wyróżnia się pewną szczególną grupę, której elementy mają dość ważną właściwość - dowolna kombinacja liniowa rozwiązań jest również poprawnym rozwiązaniem. Zależności tego typu znajdują zastosowanie np. w mechanice kwantowej, gdzie równanie Schrödingera, określające ewolucję stanu układu w czasie, jest zawsze liniowe [28]. Formalnie liniowe równanie różniczkowe można zapisać jako [103]:

$$a_n(x)f^{(n)}(x) + \dots + a_2(x)f^{(2)}(x) + a_1(x)f^{(1)}(x) + a_0(x)f(x) = b(x) \quad (2.5)$$

Należy wspomnieć na koniec, że dobrym kursem wprowadzającym do tematyki równań różniczkowych zwyczajnych jest [2].

2.2 Równania różniczkowe cząstkowe

Dla funkcji wielu zmiennych definiuje się tzw. pochodne cząstkowe:

$$\frac{\partial f(x_1, \dots, x_n)}{\partial x_i} = \lim_{\Delta x \rightarrow 0} \frac{f(x_1, \dots, x_i + \Delta x, \dots, x_n) - f(x_1, \dots, x_n)}{\Delta x} \quad (2.6)$$

Są to pochodne po konkretnym argumentie, pozostałe argumenty są traktowane przy obliczaniu pochodnej cząstkowej jak parametry. Użytecznym pojęciem przy przekształceniach jest też tzw. różniczka zupełna funkcji $f(x_1, \dots, x_n)$:

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n \quad (2.7)$$

Równaniem różniczkowym cząstkowym m -tego rzędu jest każde równanie postaci:

$$\mathbf{F}(\mathbf{x}, \mathbf{f}, \frac{\partial \mathbf{f}}{\partial x_1}, \dots, \frac{\partial \mathbf{f}}{\partial x_n}, \frac{\partial^2 \mathbf{f}}{\partial x_1^2}, \frac{\partial^2 \mathbf{f}}{\partial x_1 \partial x_2}, \dots, \frac{\partial^2 \mathbf{f}}{\partial x_1 \partial x_n}, \dots, \frac{\partial^2 \mathbf{f}}{\partial x_n^2}, \dots, \frac{\partial^m \mathbf{f}}{\partial x_1^m}, \dots, \frac{\partial^m \mathbf{f}}{\partial x_n^m}) = 0 \quad (2.8)$$

gdzie $\mathbf{x} = (x_1, \dots, x_n)$, a \mathbf{F} tak jak poprzednio jest dowolną funkcją [20]. W powyższym zapisie użyto wektorów, co oznacza, że może to być w rzeczywistości układ równań. Wygodnie jest

również wprowadzić operator nabra, który działa na funkcjach różniczkowalnych i w przestrzeni euklidesowej (w kartezjańskim układzie współrzędnych) ma postać [124]:

$$\nabla f = \frac{\partial f}{\partial x} \mathbf{e}_x + \frac{\partial f}{\partial y} \mathbf{e}_y + \frac{\partial f}{\partial z} \mathbf{e}_z \quad (2.9)$$

gdzie \mathbf{e}_x , \mathbf{e}_y i \mathbf{e}_z są wektorami jednostkowymi, stanowiącymi bazę układu współrzędnych. Jest oczywiste, że funkcja będąca wynikiem działania operatora jest funkcją wektorową. Dla przestrzeni funkcji dwukrotnie różniczkowalnych definiuje się również operator Laplace'a, który w przestrzeni Euklidesowej zapisuje się jako:

$$\Delta f = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \quad (2.10)$$

Warto podkreślić, że zarówno operator Laplace'a jak i nabla są liniowe. Używając powyższego operatora, można napisać jedno z najbardziej podstawowych równań różniczkowych cząstkowych, zwane równaniem Laplace'a:

$$\Delta f(x, y, z) = 0 \quad (2.11)$$

Jego rozwiązania nazywane są funkcjami harmonicznymi, mają one tę zdumiewającą właściwość, że ich wartość w środku dowolnej sfery jest równa średniej z jej wartości na powierzchni tejże sfery. Funkcje tego typu wykorzystywane są np. w opisie pól elektromagnetycznych czy propagacji ciepła [1]. Z wykorzystaniem operatora Laplace'a zapisuje się też w postaci równania drugie prawo Ficka, które opisuje zmianę stężenia na skutek dyfuzji:

$$\frac{\partial c}{\partial t} = -\Delta c \quad (2.12)$$

gdzie $c(x, y, z, t)$ jest stężeniem, a t to czas [49].

Czytelnik zainteresowany powyższym zagadnieniem może sięgnąć po [81].

2.3 Złożoność obliczeniowa

W informatyce dość ważna jest ocena szybkości działania poszczególnych algorytmów w zależności od danych wejściowych. Sortowanie miliona liczb można wykonywać stulecia lub ułamek sekundy, zależy to m.in. właśnie od zastosowanego algorytmu. Niestety dokładna funkcja określająca długość działania danego programu musiałaby uwzględniać wszystkie jego instrukcje, charakterystykę zbioru danych (np. niektóre algorytmy sortowania działają szybciej jeśli ciąg wejściowy jest już wstępnie uporządkowany), rodzaj procesora oraz obciążenie systemu operacyjnego. Z tego powodu potrzebna była ogólniejsza miara, która byłaby niezależna od większości czynników wymienionych wcześniej i charakteryzowałaby sam algorytm. Przykładem takiej miary jest powszechnie stosowana asymptotyczna złożoność funkcji, która bazuje na istotnym fakcie - dla bardzo dużych danych, o czasie wykonania decyduje w większości sposób rozwiązania problemu, a nie maszyna czy efektywność kodu źródłowego. Jedynym czynnikiem, który nie daje się w ten sposób wyeliminować jest charakterystyka danych wejściowych, dlatego zazwyczaj analizuje się przypadek optymistyczny, pesymistyczny i średni (do tego ostatniego jest oczywiście potrzebny rozkład możliwych danych wejściowych).

Formalnie dla nieujemnej funkcji liczb naturalnych $f(N)$ definiuje się zbiory:

1. $O(f)$ - zbiór funkcji nieujemnych dla których funkcja f jest asymptotyczną górną granicą. Nieujemna funkcja $g(N)$ należy do tego zbioru, jeśli istnieją dodatnie stałe rzeczywiste c oraz N_0 , takie że:

$$\forall N \geq N_0 g(N) \leq cf(N) \quad (2.13)$$

2. $\Omega(f)$ - zbiór funkcji nieujemnych dla których funkcja f jest asymptotyczną dolną granicą. Nieujemna funkcja $g(N)$ należy do tego zbioru, jeśli istnieją dodatnie stałe rzeczywiste c oraz N_0 , takie że:

$$\forall N \geq N_0. g(N) \geq cf(N) \quad (2.14)$$

3. $\Theta(f)$ - zbiór funkcji nieujemnych dla których funkcja f jest zarówno asymptotyczną dolną jak i górną granicą. Warunek ten można zapisać jako:

$$\Theta(f) = O(f) \cap \Omega(f) \quad (2.15)$$

W informatyce przyjęto, że należenie funkcji do powyższych zbiorów oznacza się symbolem "=", mimo że formalnie powinno się pisać $g \in O(f)$, zwykle zapisuje się $g = O(f)$. Naturalnie wydzielono całą listę klas złożoności dla poszczególnych algorytmów, które często też dzieli się na wielomianowe (ich czas działania jest asymptotycznie ograniczony z góry przez wielomian) i niewielomianowe [30].

Więcej na temat złożoności można dowiedzieć się z [8].

2.4 Algorytmy wielowątkowe

Już całe dziesięciolecie temu pojawiły się komputery zdolne do przeprowadzania kilku operacji równoległe (obliczeń wielowątkowych), a obecnie procesory wielowątkowe są już standardem w komputerach osobistych, z tego powodu powstało wiele algorytmów wykorzystujących ten fakt. Projektowanie ich napotyka jednak na pewien poważny problem - producenci maszyn nie uzgodnili jednej obowiązującej architektury i z tego powodu na rynku można spotkać komputery oparte o bardzo różne paradygmaty. Środowiskiem do przeprowadzania obliczeń wielowątkowych może być nawet zestaw komputerów połączonych siecią. Wprowadza się więc dwa podstawowe pojęcia charakteryzujące daną architekturę:

1. wątki statyczne/dynamiczne - w programowaniu równoległym definiuje się pojęcie „wątku”, czyli jednostki mogącej wykonywać operacje niezależnie od innych „wątków”, posiadającej własny, odrębny licznik instrukcji. Jeśli dany program ma dwa wątki oznacza to, że może on teoretycznie wykonywać dwa zestawy poleceń równoległe. Zależnie jednak od środowiska, mogą panować różne zasady dotyczące zarządzania liczbą wątków. Jeśli ich liczba musi być określona na początku i nie zmienia się w czasie działania programu, mówimy wtedy o wątkach statycznych. W przeciwnym razie, kiedy programista może swobodnie tworzyć lub niszczyć wątki - są one dynamiczne. W praktyce nawet jeśli dany system operacyjny pozwala na zmianę ich liczby, często jest to bardzo czasochłonna operacja i choć teoretycznie jest to dozwolone - próbuje się tego unikać.
2. pamięć wspólna/rozproszona - klasyfikacja architektur ze względu na dostęp do pamięci. Jeśli każdy wątek może się swobodnie odwołać do fragmentu pamięci w którym dane udostępnia inny wątek - mówimy o pamięci wspólnej. Natomiast w przypadku gdy dostęp do danych przetwarzanych przez inny wątek wymaga wymiany komunikatów - mamy do czynienia z pamięcią rozproszoną.

Dane środowisko można też scharakteryzować analizując sposób rozdziału pracy między wątki. Każdy z nich może dostać z góry określony zestaw instrukcji/danych, co może bardzo łatwo skutkować nierównym czasem obliczeń i wydłużeniem działania programu (niektóre wątki skończą pracę szybciej i będą marnować czas czekając na pozostałe). Zapobiec temu można implementując narzędzie, które może dynamicznie zarządzać przydziałem i na bieżąco delegować obowiązki tak, aby każdy wątek był równomiernie obciążony.

Kiedy pojawia się potrzeba oceny czasu wykonania algorytmów równoległych - problem staje

się dość złożony, ponieważ pojawia się pytanie, jak traktować segmenty wykonywane równolegle. Obok rzeczywistego czasu wykonania (który zależy między innymi od liczby dostępnych wątków w systemie), pojawia się pojęcie „pracy” czyli sumy czasów wykonania wszystkich wątków, co można traktować jako czas działania „gdyby program był algorytmem jednowątkowym (sekwencyjnym)”. Drugim istotnym wskaźnikiem jest „rozpiętość”, która mówi o długości działania najdłużej wykonującego się wątku, co można też nazwać czasem wykonania tzw. „ścieżki krytycznej”. Mierząc czas wykonania sekwencyjnego i wielowątkowego algorytmu można też obliczyć „przyspieszenie”, które mówi jak efektywnie wykorzystana jest wielowątkowość środowiska, oblicza się je dzieląc po prostu czas wykonania sekwencyjnego przez czas wykonania równoległego. W idealnym przypadku wskaźnik ten powinien być równy liczbie wątków. W praktyce jednak niektóre fragmenty kodu muszą być wykonywane sekwencyjnie, występują też punkty w których następuje „synchronizacja” - wszystkie wątki muszą dotrzeć do tego samego miejsca w kodzie zanim będzie można wykonać dalszą część [31].

Algorytmy wielowątkowe stanowią bardzo ważną dziedzinę informatyki, dlatego zostały szczególnie opisane w wielu książkach, np. [10].

2.5 Automaty komórkowe

Automaty komórkowe zostały pierwszy raz wprowadzone przez von Neumanna i Ulama jako model matematyczny mogący być idealizacją układów biologicznych. Definiuje się je jako siatkę złożoną z komórek/pól, dla której również zdefiniowano sąsiedztwo - każda komórka posiada zbiór „sąsiadów” (innych komórek), najczęściej są to po prostu pola przylegające. Każda komórka ma określony stan należących do zbioru dostępnych stanów określonych w danym modelu. Kroki czasowe również są dyskretne - w każdym następuje aktualizacja stanów wszystkich komórek zgodnie ze zdefiniowaną wcześniej funkcją. Bierze ona pod uwagę stan danej komórki oraz wszystkich jej sąsiadów.

Modele oparte na automatach komórkowych mogą prezentować zachowanie bardzo odmienne od rzeczywistego - np. spontanicznie tworzyć złożone, uporządkowane struktury, co w realnym świecie bez interakcji z otoczeniem znajdującym się poza systemem jest niemożliwe ze względu na prawa rządzące entropią. Niemniej z powodzeniem wykorzystuje się je także do symulacji faktycznych zjawisk w fizyce, chemii (reakcje chemiczne z dyfuzją przestrzenną) oraz biologii (symulacja wzrostu organizmów) [126].

Bardzo dobre wprowadzenie do tematyki automatów komórkowych stanowi [127].

2.6 Systemy wieloagentowe

Systemy wieloagentowe są oparte na zupełnie innej zasadzie niż automaty komórkowe, w tym modelu mamy do czynienia ze zbiorem inteligentnych agentów, które mogą wchodzić w interakcję ze środowiskiem i na podstawie informacji zwrotnej podejmują decyzję o dalszym działaniu. Wszystkie obserwowane zachowania systemu jako całości są wynikiem wielu takich interakcji, które wynikają z lokalnie podejmowanych decyzji a nie z dążenia do wspólnego celu. Środowisko w którym działają agenty można scharakteryzować ze względu na kilka ważnych cech:

1. deterministyczne/niedeterministyczne - stan otoczenia jest w pełni zdeterminowany przez akcje agentów i stan poprzedni (środowisko deterministyczne) lub nie (niedeterministyczne)
2. dostępne/niedostępne - w „dostępnym” środowisku agent może otrzymać pełną informację o otoczeniu, w przeciwnym razie jest to środowisko niedostępne,
3. statyczne/dynamiczne - czy jego stan może się zmieniać w czasie?
4. dyskretne/ciągłe - czy zbiór akcji, które może wykonać agent jest dyskretny [128]?

Symulacje wieloagentowe mają zastosowanie w wielu dziedzinach, warto chociażby wspomnieć o testowaniu algorytmów służących do samodzielnego formowania klastrów przez roboty przy braku jakiegokolwiek centralnej jednostki zarządzającej [55] lub wykorzystaniu ich w biologii obliczeniowej. Corrêa et al. [32] za pomocą algorytmu *ab initio* opartego na zbiorze autonomicznych agentów oraz informacji z bazy danych, stworzyli metodę zdolną do przewidywania struktury 3D białek, której wyniki były topologicznie kompatybilne z eksperymentem. Innym przykładem jest wykorzystanie systemów wieloagentowych na gruncie teorii gier w celu obserwacji zachowań globalnych [100].

Więcej na temat systemów wieloagentowych można dowiedzieć się z [102].

Podstawy biologiczne i chemiczne

W celu zrozumienia opisanych w dalszej części badań, należy zapoznać się z odpowiednimi pojęciami z zakresu chemii, biologii a także bioinformatyki. Najważniejsze informacje podane są poniżej, zainteresowany czytelnik może też sięgnąć po odpowiednie książki:

1. [98] - opis jednej z hipotez powstania Świata RNA poprzedzony dobrym wprowadzeniem do tematyki,
2. [41] - współtworzony przez samego Eigena szczegółowy opis hipercykladu,
3. [79] - wprowadzenie do tematyki początków życia, rozważania na temat słabości teorii mówiących o losowych procesach prowadzących do powstania żywych komórek.

3.1 Podstawy dynamiki molekularnej

Modelowanie zachowania cząsteczek ma bardzo ważne zastosowania praktyczne, np. przy analizie reakcji zachodzących w laserach, a także w symulacjach opisanych w dalszych rozdziałach. Najbardziej zbliżonym do rzeczywistości podejściem byłoby wyliczenie potencjału dla stanu początkowego każdej cząsteczki i wszystkich stanów, które może ona przyjąć, bazując na mechanice kwantowej. Niestety z powodu trudności jakie napotyka się przy próbie zastosowania opisanego podejścia (np. ze względu na wysoki koszt obliczeń dla nawet najprostszyc przypadków), przybliża się badany układ za pomocą mechaniki klasycznej. Okazuje się, że w większości przypadków jest to wystarczające dla osiągnięcia pożądanej dokładności [66]. Uproszczonym modelem, który pozwala na znaczne przyspieszenie obliczeń jest dynamika Browna, wykorzystywana np. w symulacjach biologicznych. W tym podejściu (zależną od czasu) pozycję każdej cząsteczki $\mathbf{X}(t)$ aktualizuje się zgodnie ze wzorem:

$$\mathbf{X}(t + \Delta t) = \mathbf{X}(t) + \sqrt{2D\Delta t}\boldsymbol{\xi} \quad (3.1)$$

gdzie Δt oznacza krok czasowy, D stałą dyfuzji, a $\boldsymbol{\xi}$ wektor o rozkładzie Gaussowskim ze średnią 0 i wariancją równą 1 [43].

3.2 Kinetyka reakcji chemicznych

Opisując kinetykę definiuje się tzw. szybkość reakcji, jest to ilość produktu powstająca (alternatywnie pochłaniana ilość substratu) w jednostce czasu i w jednostce objętości. Dla przykładowej reakcji:



Szybkości reakcji można przedstawić jako: $\frac{d[A]}{dt}$, $\frac{d[B]}{dt}$, $\frac{d[Z]}{dt}$, $\frac{da}{dt}$, $\frac{db}{dt}$ oraz $\frac{dz}{dt}$, gdzie $[A]$, $[B]$ oraz $[C]$ wyrażają stężenia substancji, podczas gdy a , b oraz z to ich łączna ilość. Wymienione wartości pochodnych nie są sobie równe, ale są ze sobą ściśle powiązane. Wpływ na szybkość przebiegu reakcji ma m.in. temperatura, typ reakcji (reakcje z udziałem jonów i pojedynczych atomów są na ogół szybsze niż te w których przerywane są wiązania kowalencyjne), ciśnienie oraz stężenia reagentów. Cały proces można również przyspieszać dodając odpowiedni katalizator [68].

Jednym ze sposobów obliczenia szybkości reakcji jest skorzystanie z „prawa działania mas” sformułowanego w XIXw. przez norweskiego naukowca Cato M. Guldberga. Głosi ono, że szybkość reakcji jest proporcjonalna do iloczynu mas substratów, a każda z nich jest podniesiona do potęgi równej liczbie cząsteczek biorącej udział w reakcji. Dla podanej wcześniej reakcji szybkość zgodnie z tym prawem wyniosłaby:

$$\frac{d[A]}{dt} = k[A][B]^3 \quad (3.3)$$

gdzie k to stała szybkości reakcji [67, 68].

Wpływ temperatury na szybkość reakcji można uwzględnić stosując równanie Arrheniusa:

$$k = A \exp\left(-\frac{E}{RT}\right) \quad (3.4)$$

gdzie A reprezentuje częstotliwość zderzeń atomów i cząsteczek prowadzących do zajścia reakcji, E - energię aktywacji (energię potrzebną do zajścia reakcji), R jest stałą gazową (ok. $8,314 \frac{J}{K \cdot mol}$), a T to temperatura bezwzględna [23].

3.3 Problem początków życia

Pytanie jak powstało życie na Ziemi nie zostało do dzisiaj rozstrzygnięte. Wyróżnia się dwa główne sposoby w jakie mogą formować się żywe organizmy:

1. biogeneza - nowe życie powstaje dzięki reprodukcji innego życia,
2. abiogeneza - powstawanie życia z materii nieożywionej. Zakłada się przy tym, że powstałe w ten sposób organizmy muszą być na początku dosyć proste, stopniowo zwiększając swoją złożoność z czasem.

Jakakolwiek biogeneza musiała być prawdopodobnie poprzedzona przez abiogenezę, jednak przy współczesnym składzie atmosfery uznaje się tę drugą za niemożliwą. Jeśli miała miejsce - zdarzyło się to w czasie kiedy na Ziemi panowały zupełnie inne warunki (por. rys. 3.1).

Rozwój współczesnych hipotez zakładających abiogenezę rozpoczął się w latach 20, kiedy angielski naukowiec J. B. S. Haldane oraz rosyjski biochemik A. Oparin niezależnie od siebie zapropono-



Rys. 3.1: Młoda Ziemia była nieustannie bombardowana przez meteory, na powierzchni znajdowały się prawdopodobnie liczne wulkany [132].

Źródło: NASA/Goddard Image Lab

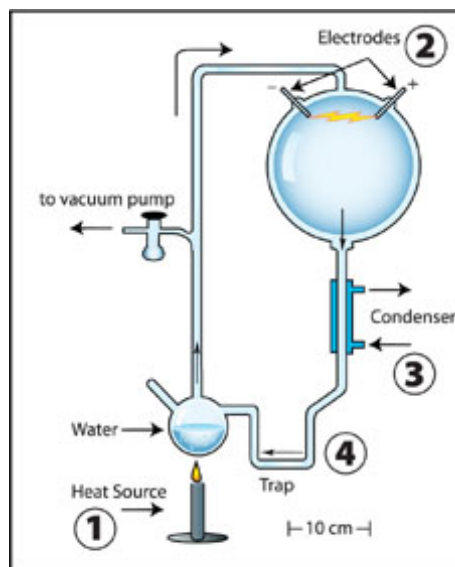
nowali, że związki organiczne mogą samoczynnie powstawać z substancji nieorganicznych. Zgodnie z ich propozycją, działałoby się tak przy obecności zewnętrznego źródła energii (takiego jak np. promieniowanie ultrafioletowe) oraz redukującej atmosfery, czyli przede wszystkim prawie całkowicie pozbawionej wolnego tlenu. Życie mogłoby wtedy powstać w ciepłych oceanach i odżywiać się substancjami, które obecne były w otoczeniu (niezdolne jeszcze do samodzielnej ich syntezy). W tym punkcie jednak wizje Haldane'a i Oparina zaczynają się rozchodzić. Pierwszy z nich uważał (na podstawie swoich badań), że pierwsze enzymy działały w „bąblach”, otoczone sferyczną membraną lipidową, formującą się samoczynnie i utrzymywaną przez siły elektrostatyczne. Takie konstrukty mogły później dać początek komórkom. Oparin natomiast skłaniał się ku hipotezie, że w oceanach ewoluowały coraz bardziej skomplikowane cząsteczki, ostatecznie formując pierwsze komórki (bez etapu pośredniego w postaci „bąbli”).

Kolejnym kamieniem milowym w dziedzinie był eksperyment przeprowadzony w 1953r. przez dwóch badaczy - Harolda C. Ureya oraz Stanleya Millera. Chcieli oni zweryfikować hipotezę Oparina i Haldana, odtwarzając warunki i skład chemiczny młodej Ziemi. Wypełnili oni naczynie ciepłą wodą oraz czterema gazami: parą wodną, metanem, amoniakiem i wodorem. Tak powstałą mieszaninę (mającą składem przypominać dawną atmosferę i ocean) traktowali impulsami elektrycznymi, które miały być wspomnianym wcześniej źródłem zewnętrznej energii. Schemat eksperymentu przedstawiony jest na rys. 3.2. W ciągu tygodnia udało się otrzymać wiele podstawowych związków organicznych, w tym aminokwasy [94].

Od lat 50 pojawiły się liczne teorie, które bazują z grubsza na opisanych wcześniej założeniach, warto wspomnieć chociażby o stworzonej przez grupę niemieckich naukowców hipotezie, zgodnie z którą pierwsze życie powstało w gorących źródłach, które funkcjonowały w sposób cykliczny. Takie warunki doprowadzały do periodycznej zmiany środowiska dla niektórych związków i umożliwiały zajście reakcji, które nie mogłyby zaistnieć w zwykłym roztworze wodnym [98]. Wszystkie próby zmierzają do wyjaśnienia powstania ostatniego wspólnego przodka wszystkich organizmów na Ziemi - LUCA (z ang. „Last Universal Common Ancestor”). Wiadomo o nim na pewno, że posiadał już kod genetyczny zapisany w DNA [45].

3.4 Hipoteza Świata RNA

W latach 60 opisana w poprzednim rozdziale teoria abiogenezy była obiektem wielu badań. Zastanawiano się jak pierwsze (powstałe przypadkowo) funkcjonalne układy biologiczne mogłyby działać. Wiedziano już wtedy, że żywe komórki korzystają przede wszystkim z RNA, DNA oraz białek, jednak nawet przy dużych uproszczeniach tego typu maszyna byłaby zbyt skomplikowana, aby powstać samoczynnie. Zwrócono przy tym uwagę, że kwasy nukleinowe posiadają możliwość naturalnego kopiowania niesionej przez siebie informacji dzięki łączeniu się w pary komplementarnych zasad azotowych, cecha której nie mają białka. RNA i DNA stały się zatem kandydatami do bycia cegiełkami pierwszych form życia. Wybór ostatecznie padł na RNA, ze względu na jego duży udział w translacji (matrycowe, transportowe i rybosomalne RNA), teoretyzowano że mogłoby ono w przeszłości pełnić funkcje współczesnych enzymów białkowych, a dzisiejsze koenzymy mające nukleotydy jako części swojej struktury to pozostałości po tych czasach. Wiarygodności opisanej hipotezie przysporzyło późniejsze (lata 80) odkrycie rybozymów, będących realnym przykładem katalitycznych możliwości RNA [33, 87]. Należałoby tutaj wspomnieć, że choć sama koncepcja została pierwszy raz zaproponowana przez Alexandra Richa w 1962r. [91], nazwa „Świat RNA” pochodzi z lat 80 [26, 48]. Hipoteza do dzisiaj pozostaje obiektem intensywnych badań, zarówno w laboratorium [92, 129, 133] jak i za pomocą metod obliczeniowych [101, 106, 107].



A.

Rys. 3.2: Schemat oryginalnego eksperymentu Millera-Ureya. Woda jest najpierw zagotowywana, a otrzymana para przepuszczana przez łuk elektryczny razem z pozostałymi związkami nieorganicznymi. Następnie produkty są skraplane i wpadają do "pułapki".

Źródło: NASA https://www.nasa.gov/topics/earth/features/volcanic_life_origin.html [dostęp 2022-08-24]

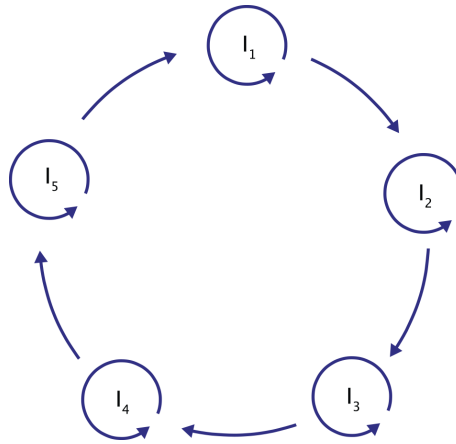
3.4.1 Niekonająca replikacja i paradoks Eigena

Rozważając prosty system zdolny do kopiowania informacji genetycznej, należy zadać sobie pytanie, jakie ma on ograniczenia. Bez wykorzystania współczesnych enzymów białkowych, replikacja musiałaby być podatna na błędy. Jak pokazał Eigen [39] - narzuca to ograniczenie na maksymalną długość sekwencji, jaka może być kopiowana i utrzymywana w roztworze. Smith [104] teoretyzował, że wartość ta dla prostych systemów wynosi około 100 nukleotydów. Jedyny sposób, żeby zwiększyć ilość informacji, to stworzenie bardziej zaawansowanej maszyny replikującej, jednak wymaga to przechowywania instrukcji do jej budowy - czyli dodatkowej informacji. Chociaż taki system mógłby teoretycznie funkcjonować - jaki był jego początek? Zbudowanie odpowiedniej maszyny wymaga danych, których przechowywanie z kolei od samego początku musiało być przez nią wspomaganie. Problem ten nosi nazwę „paradoksu Eigena”, a możliwość jego rozwiązania to jedno z najtrudniejszych pytań w badaniach początków życia [54].

Jedną z możliwych odpowiedzi podał już sam Eigen, kiedy w 1971r. poruszył kwestię wpływu mutacji na przetrwanie informacji genetycznej. Był to tzw. „hipercykl” (rys. 3.3), którego koncepcja zakładała, że powielana sekwencja nukleotydów przetrwa rozdzielona między kilka cząsteczek. Zdolności katalityczne tychże mogą wspierać się nawzajem, co miałyby teoretycznie pozwolić na rozwiązanie paradoksu [111]. Niestety analiza teoretyczna jednoznacznie wskazuje, że taki system nie jest ewolucyjnie stabilny [35].

3.4.2 Model RP

Ze względu na dość rozbudowane modele teoretyczne systemów zdolnych do przeprowadzania replikacji (takie jak np. hipercykle), pojawiła się prostsza wersja - najmniej skomplikowany model, w którym istniałby rozdział na cząsteczki katalizujące kopiowanie innych (replikazy) oraz na te, które nie są do tego zdolne, ale same wykorzystują replikazy do przetrwania (pasożyty). Ponadto proces replikacji musiałby z określonym prawdopodobieństwem kończyć się błędem i wyprodukować



Rys. 3.3: Przykładowy hipercykl złożony z 5 cząsteczek, każda zdolna do katalizowania własnej replikacji oraz replikacji następnej cząsteczki w cyklu.

Źródło: [111]

waniem pasożyta w procesie kopiowania cząsteczki ze zdolnościami replikacyjnymi. Najprostszy model mający opisane wyżej właściwości można opisać szeregiem reakcji:



W powyższych równaniach X oznacza replikazę, Y pasożyta, Q prawdopodobieństwo poprawnej replikacji, r jest stałą szybkości replikacji pasożytów, a d stałą rozpadu. Całość została znormalizowana tak, żeby stała replikacji replikazy wynosiła 1 [46, 78].

Pewną modyfikacją powyższego schematu było wprowadzenie kompleksów, które modelują *explicite* trwającą replikację. Nowy wariant nazwany został przez autorów „modelem CXY” aby odróżnić go od wcześniej opisanego („modelu XY”) [46]:



C_{XX} oraz C_{XY} są kompleksami związku X odpowiednio z inną cząsteczką X i z cząsteczką Y . Kompleks jest formowany z szybkością k_1 , a dysocjuje z szybkością k_{-1} , co odpowiada przedwczesnemu zakończeniu replikacji bez stworzenia nowej cząsteczki. Kolejnym wprowadzonym parametrem jest k_2 - stała szybkości replikacji w kompleksie [46].

Model CXY został wykorzystany przez Takeuchiego i Hogeweg do symulacji Świata RNA. Założyli oni, że X i Y są odpowiednimi cząsteczkami RNA, a następnie za pomocą równań różniczkowych zwyczajnych jak i automatów komórkowych zbadali zachowanie populacji łańcuchów RNA. Pokazali oni, że dzięki tworzeniu odpowiednich struktur przestrzennych system może przetrwać, podczas gdy w dobrze wymieszanym roztworze populacja staje się bardzo niestabilna [112]. Opi-

sany powyżej model Świata RNA będzie w tej pracy nazywany „modelem RP” (replikaza-pasożyt), poza analizą Takeuchiego i Hogeweg, był on również szczegółowo badany w pracach [105, 106, 110] (patrz rozdziały 5, 6 i 7).

3.4.3 Kompartmenty

Jak wspomniano wcześniej, koncepcja pierwszych enzymów ewoluujących w zamkniętych bąblach pojawiła się już u samych początków współczesnych badań nad początkiem życia - u Haldane'a. Nie ulega wątpliwości, że pojawienie się protokomórek (bąbli lub pęcherzyków otoczonych membraną, w których działałyby enzymy) było ważnym etapem na drodze ewolucji prowadzącej do pierwszych żywych komórek. Głównymi problemami napotykanymi przez badaczy jest źródło materiału, który tworzyłby błonę oraz sposób w jaki protokomórki mogłyby się dzielić i w ten sposób zwiększać swoją liczebność. Protokomórki których informacja genetyczna dawałaby przewagę miałyby większą szansę na propagację, co z kolei stałoby się motorem napędowym ewolucji. Mechanizmy, który pozwoliłyby na zachodzenie powyższych procesów są od lat badane metodami laboratoryjnymi [110]. Należy tutaj wspomnieć, że opisana tutaj selekcja naturalna całych pęcherzyków wprowadzałaby dodatkową korekcję stochastyczną populacji w nich zamkniętych. Prowadziłoby to teoretycznie do zmniejszenia częstości mutacji, a co za tym idzie - kolejnego sposobu na rozwiązanie paradoksu Eigena [53]. Takeuchi i Hogeweg [114] badali wpływ kompartmentalizacji na ewolucję w modelu RP za pomocą automatów komórkowych, założyli oni, że budulec pozwalający rozciągać membranę (i w efekcie powiększać pęcherzyki) jest produkowany przez pasożyty. Do podziału dochodziło po przekroczeniu przez protokomórkę określonej w parametrach objętości. Ważnym wnioskiem otrzymanym w toku badań był fakt, że podział przestrzeni na pęcherzyki (lub "kompartmenty" jak nazwane są w omawianej publikacji) ogranicza naturalną ewolucję pasożytów w kierunku coraz skuteczniejszego wykorzystywania istniejącej maszynierii replikacyjnej i chroni przez to populację replikatorów przez wymarciem.

3.5 Ślady geologiczne pozostawione przez wczesne stadia ewolucji komórek

Przyjmuje się, że na Ziemi warunki konieczne dla istnienia życia panują od 4,3 miliarda lat. Należy przez to rozumieć, że skorupa była już wówczas ukształtowana, a na jej powierzchni (albo przy powierzchni) znajdowała się woda w stanie ciekłym. Trzeba pamiętać jednak, że dowody zachowane w skałach sięgają najwyżej 4 miliardy lat wstecz, a najstarsze odnalezione i potwierdzone ślady istnienia życia są 600 milionów lat młodsze i wskazują na istnienie już żywych komórek w tamtym czasie. Wszystkie wcześniejsze stadia muszą być przez to badane innymi metodami, ponieważ są bardzo małe szanse na istnienie jakichkolwiek pozostałości geologicznych ewolucji z tak wczesnego okresu. Powodem dla których analizowanie i potwierdzanie śladów pozostawionych przez aktywność biologiczną stwarza tak ogromne trudności jest między innymi fakt, że skała istniejąca miliardy lat poddawana była najróżniejszym procesom geologicznym i należy je wszystkie uwzględnić przy badaniach. Poważną przeszkodą jest również możliwość powstawania związków organicznych w sposób całkowicie abiotyczny, co udowodnił już Wöhler w 1828r., otrzymując mocznik w laboratorium z substancji nieorganicznych. Jeśli jednak wykluczy się wszystkie alternatywy i udowodni się, że badana próbka zawiera ślady aktywności biologicznej, pozostaje kwestia poprawnego datowania. Wiek samej skały nie musi zgadzać się z momentem pozostawienia w niej materiału pochodzenia organicznego. Mikroorganizmy mogły dostać się do wnętrza znacznie później przez szczeliny i pory razem z wodą [57].

3.6 Struktura przestrzenna RNA

Liczne właściwości regulatorowe i katalityczne RNA wynikają z jego zdolności do formowania skomplikowanych, trójwymiarowych struktur. Bezpośrednie ich ustalanie za pomocą krystalografii jest bardzo pracochłonne i wymaga ogromnej wiedzy, dlatego w celu usprawnienia badania właściwości kolejnych cząsteczek - rozwinięto liczne metody bioinformatyczne (obliczeniowe), korzystające nierzadko z danych eksperymentalnych [72]. Dużo informacji daje już sama wiedza o strukturze drugorzędowej cząsteczki (czyli wiązaniach kanonicznych między rybonukleotydami), przykładowa struktura drugorzędowa przedstawiono na rys. 3.4. Przewidzieć ją można metodami porównawczymi, jednak to podejście wymaga bazy danych z ogromną liczbą sekwencji (homologicznych do analizowanej) o znanym parowaniu. Alternatywą jest wykorzystanie metod opartych o termodynamikę i minimalizowanie energii swobodnej lub maksymalizacja sumy prawdopodobieństw utworzenia poszczególnych par zasad [90].

Struktura drugorzędowa i 3D okazuje się bardzo istotna w przypadku ryboprzełączników, które potrafią wiązać metabolity i w ten sposób regulować ekspresję genów. Ułożenie rybonukleotydów w przestrzeni musi dobrze odpowiadać rozpoznawanej substancji, aby osiągnąć wysoką specyficzność, niestety choć z reguły struktura danego ryboprzełącznika jest zakonserwowana, sekwencje należące do różnych klas bardzo różnią się między sobą. Ten fakt w połączeniu z potrzebą wysokiej dokładności (aby osiągnąć specyficzność) czyni postawiony problem bardzo trudnym [89].

3.7 Aktualny stan badań *in vitro*

3.7.1 Replikacja za pomocą tripletów

Głównym problemem przy badaniach Świata RNA jest brak cząsteczek zdolnych do przeprowadzania replikacji w odpowiednich warunkach (replikaz), które musiały zaginąć w toku ewolucji. Próba ich konstrukcji *in vitro* napotyka na pewien paradoks - replikaza musi mieć odpowiednią strukturę aby funkcjonować, ale kopiowanie jej wymaga już „rozplątania” łańcucha. Attwater et al. [7] zaproponowali replikację opartą na trójkach nukleotydów, które silniej wiążą się ze wzorcem przez co same mogą promować rozplatanie struktury kopiowanego RNA. W celu skonstruowania rybozemu o powyższych własnościach, autorzy wyszli od rybozemu Z RPR (zdolnego do ligacji pojedynczych nukleotydów) i skrócili go, żeby zachować wyłącznie część katalityczną (Z Core). Warto podkreślić, że otrzymana cząsteczka miała już na wstępie zdolność do ligacji oligonukleotydów przy kopiowaniu, jednak w bardzo małym stopniu. W kolejnym etapie eksperymentu dodano do końca 3' każdego RNA 30 losowych nukleotydów i przeprowadzono 7 rund selekcji w eutektycznej mieszaninie wody z lodem (która zwiększała stabilność RNA i skuteczność rybozemu), co doprowadziło do powstania nowego rybozemu, nazwanego w artykule „typem 0”. Typ ten był zdolny do przeprowadzania replikacji z udziałem tripletów, jednak wciąż w mało wydajny sposób. Po kolejnych 14 rundach selekcji wyodrębniono już 6 nowych typów (1-6), które całkowicie wyparły typ 0. Największą część populacji stanowił typ 1, który nie był aktywny katalitycznie z powodu dużej ilości różnorodnych mutacji w swoim centrum. Pozostałe typy (2-6), posiadały już pewną aktywność, ale mniejszą niż cała mieszanina. Dalsze badania wykazały, że typy 2-6 radziły sobie znacznie lepiej, jeśli każdy z nich wymieszano w stosunku molowym 1:1 z typem 1. Ustalono, że ten ostatni mimo braku aktywności enzymatycznej był zdolny do pełnienia roli kofaktora i w ten sposób usprawniał replikację przeprowadzaną przez typy 2-6. Odpowiadał za to fragment sekwencji znajdujący się przy końcu 5' aktywnych rybozemu, nazwany „CAP+”. Przeniesienie „CAP+” do oryginalnych rybozemu (Z RPR i Z Core) również powodowało, że typ 1 zwiększał ich aktywność. Kontynuacja eksperymentu doprowadziła do wyewoluowania z typu 5 (posiadającego największą aktywność w kompleksie z typem 1) nowego, skuteczniejszego rybozemu - „t5⁺¹”. Niestety nie

był on w stanie pełnić wszystkich wymaganych funkcji replikazy - nie potrafił kopiować samego siebie ze względu na swoją długość. Udało się osiągnąć częściowy sukces dzieląc „t5⁺” na łatwiejsze do zreplikowania części (które enzym przy odpowiednim wspomaganiu był w stanie skopiować i poprawnie złożyć), ale osiągnięcie w pełni skutecznego cyklu samoreplikacji wymaga dalszych badań.

3.7.2 Powstawanie cegiełek do budowy RNA - rybonukleotydów

Abiotyczna synteza rybonukleotydów jest bardzo ważna dla hipotezy Świata RNA, ponieważ tłumaczyłaby częściowo spontaniczne pojawienie się kwasów rybonukleinowych. Nam et al. [83] wykorzystali ścieżkę reakcji opartą na Rib-1-P (D-rybozo-1-fosforan) jako związku pośrednim. Wspomniane reakcje przeprowadzono w mikrokropelkach wody z wykorzystaniem kationu magnezu (Mg^{2+}) jako katalizatora. Jak argumentują autorzy, mgły, mgielki i chmury były prawdopodobnie obecne na Ziemi kiedy powstawał Świat RNA, zatem zaproponowane przez nich reakcje chemiczne mogły wtedy zachodzić. Mikrokropelki (w dużej liczbie) dają ogromną powierzchnię styku wody z powietrzem, co jak przytaczają autorzy, uznaje się za środowisko sprzyjające powstawaniu cząsteczek o znaczeniu biologicznym. Ostatecznie z mieszanki D-rybozy, kwasu fosforowego oraz zasad azotowych (adeniny, cytozyny oraz hipoksantyny) udało się uzyskać adenozyne, cytydine oraz inozyne. Ta ostatnia została otrzymana jako zamiennik guanozyny, ze względu na niską rozpuszczalność w wodzie guaniny. Warto tutaj dodać, że autorom udało się wcześniej otrzymać w podobny sposób urydine [82].

Nie tylko bariera woda-powietrze może pomóc przy zachodzeniu potrzebnych reakcji, Chu i Zhang [29] zaproponowali hipotezę, zgodnie z którą ATP mogło powstawać samoczynnie dzięki aktywności wulkanicznej (która jak wspomniano wcześniej była dość duża kilka miliardów lat temu). Same wulkany miały generować wokół siebie kwas cyjanowodorowy, który ulegał częściowej hydrolizie do formamidu i w chłodniejszej wodzie stawał się substratem do abiotycznej syntezy adeniny. Kolejny ważny składnik ATP, czyli ryboza, formował się z formaldehydu samoczynnie powstającego w powietrzu dzięki reakcjom fotochemicznym i elektrochemicznym. Cząsteczki formaldehydu przenikały do stawów u stóp wulkanu razem z fosforanami wydobywającymi się razem z gazami wulkanicznymi. Ostatecznie całość (razem z powstającą równolegle adeniną) łączyła się samoczynnie w ATP dzięki naprzemiennym suchym i mokrym cyklom.

Metody analitycznego badania Świata RNA

4.1 Równanie quasi-gatunku

Koncepcja quasi-gatunku została wprowadzona przez Eigena, McCaskilla i Schustera w 1988 roku [40] w celu stworzenia modelu matematycznego teorii Darwina. Całość opiera się o rywalizację o przetrwanie pomiędzy polimerami. Podstawowe założenia tego modelu były następujące:

1. Wszystkie cząsteczki reprezentowane w systemie muszą mieć zdolność do samoreplikacji, co w równaniach dynamiki reakcji jest wyrażone jako zdolności autokatalityczne. Chociaż szybkość takiej reakcji w rzeczywistości zależy od wielu czynników i powinna uwzględniać wszystkie kroki pośrednie, tutaj przyjmuje się, że jest proporcjonalna do stężenia replikującej się cząsteczki, gdzie stała proporcjonalności jest własnością fenotypową. Oznacza to, że określa ją sekwencja, genotyp oraz własności chemiczne środowiska.
2. Replikacja jest niedoskonała i co jakiś czas występują mutacje. Dla uproszczenia zakłada się jedynie możliwość substytucji, a w równaniach uwzględnia się to jako produkcję przez daną cząsteczkę innych rodzajów. O tym jakie cząsteczki będą najczęściej wynikiem mutacji decyduje odległość Hamminga, czyli liczba pozycji na których występują różnice.
3. System znajduje się w stanie dalekim od jakiegokolwiek punktu równowagi. Reakcje powstawania i rozpadu muszą być nieodwracalne. Celem jest jak największe uzależnienie ewolucji systemu od własności poszczególnych cząsteczek.
4. Nie zakłada się żadnej interakcji między badanymi cząsteczkami, nie mogą one wzajemnie pomagać albo przeszkadzać sobie w replikacji.

Sekwencje rywalizujących molekuł są reprezentowane jako ciągi ν znaków pewnego alfabetu o rozmiarze κ :

$$I_i = (s_1^{(i)} s_2^{(i)} s_3^{(i)} \dots s_\nu^{(i)}) \quad (4.1)$$

Dla sekwencji polinukleotydowych: $s_j^{(i)} \in \{G, A, C, U\}$ oczywiste jest że $\kappa = 4$. Razem z sekwencjami definiuje się także macierz mutacji \mathbf{Q} w której pojedynczy element Q_{ij} oznacza prawdopodobieństwo, że przy kopiowaniu sekwencji j powstanie sekwencja i . Z tego opisu jasno wynika, że:

$$\sum_{k=1}^n Q_{ki} = 1 \quad (4.2)$$

gdzie n to liczba sekwencji w modelu. W celu uproszczenia analizy można przyjąć, że szansa na substitucję w dowolnym miejscu w sekwencji jest zawsze identyczna i wynosi $1 - q$. Nukleotydy zostają wtedy zastąpiony dowolnym innym z równym prawdopodobieństwem. Wzór na poszczególne elementy macierzy \mathbf{Q} wygląda wtedy następująco:

$$Q_{ji} = \left(\frac{1 - q}{q(\kappa - 1)} \right)^{d(i,j)} q^\nu \quad (4.3)$$

Funkcja $d(i, j)$ to odległość Hamminga między sekwencjami i oraz j . W celu zapisania odpowiedniego równania, wygodnie jest zdefiniować dla każdej sekwencji stałą nadprodukcji, będącą różnicą między stałą replikacji i rozpadu:

$$E_i = A_i - D_i \quad (4.4)$$

Oprócz tego wprowadza się także macierz \mathbf{W} , która na przekątnej zawiera stałe nadprodukcji uwzględniające mutację, a poza nią stałe produkcji zmutowanych sekwencji z odpowiednich sekwencji wyjściowych. Matematycznie można powyższą definicję zapisać następująco:

$$W_{ii} = A_i Q_{ii} - D_i \quad (4.5)$$

$$W_{ki} = A_i Q_{ki} \quad (4.6)$$

Jeśli przez $c_i(t)$ zdefiniujemy stężenie sekwencji i w zależności od czasu, można wprowadzić (znormalizowane) stężenia względne jako wektor kolumnowy (\mathbf{x}):

$$x_i(t) = c_i(t) / \sum_{k=1}^n c_k(t) \quad (4.7)$$

$$\sum_{k=1}^n x_k(t) = 1 \quad (4.8)$$

Ostatnim koniecznym składnikiem jest średnia stałych nadprodukcji, dzięki której można upewnić się, że stężenia względne będą sumować się do jedynki:

$$\bar{E}(t) = \sum_{k=1}^n x_k(t) E_k \quad (4.9)$$

Ostateczny układ równań różniczkowych można zapisać jako:

$$\frac{d\mathbf{x}}{dt} = \mathbf{W}\mathbf{x} - \bar{E}\mathbf{x} \quad (4.10)$$

Najważniejszym wnioskiem, który wynika z tegoż równania jest istnienie progu Eigena wspomnianego w rozdziale 3 [40].

4.2 Równanie replikatorowe

W przeciwieństwie do równania quasi-gatunku, równanie replikatorowe zostało zdefiniowane na gruncie teorii gier. Pierwszy raz podał je P. D. Taylor w 1978 [116], opisywało ono ewolucję systemu z n graczami w którym skuteczność każdej ze strategii zależała od strategii przyjętej przez pozostałych, a nagrodą w grze była oczywiście możliwość reprodukcji. W przeciwieństwie do poprzednio omówionego modelu nie zakłada się mutacji oraz nie wyklucza interakcji między cząsteczkami. Formalnie jeśli za wektor \mathbf{x} przyjmą zależne od czasu względne stężenia poszczególnych molekuł (lub w innej interpretacji - odsetki graczy przyjmujących poszczególne strategie), wtedy równanie replikatorowe (w wersji nieliniowej) przyjmuje postać:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) - \bar{f}(\mathbf{x})\mathbf{x} \quad (4.11)$$

gdzie \mathbf{f} jest funkcją przyporządkowującą każdej strategii (na podstawie stanu systemu) jej skuteczność. Z kolei \bar{f} jest średnią ważoną wartości \mathbf{f} , która ma zapewnić, że stężenia względne zawsze będą sumowały się do jedynki:

$$\bar{f} = \sum_{i=1}^n f_i(\mathbf{x})x_i \quad (4.12)$$

Wygodnie wprowadzić tutaj pojęcie „stabilności ewolucyjnej”. Stan ewolucyjnie stabilny to taki stan do którego system zawsze będzie powracał przy niewielkich zaburzeniach. W ujęciu bardziej biologicznym można stwierdzić, że żadna mała grupa mutantów nie może zmusić reszty systemu do zmiany stężeń poszczególnych cząsteczek, zawsze to ta nowa grupa zostanie wyparta [52]. Można zauważyć, że równanie replikatorowe jest bardzo ogólne i zarówno równanie quasi-gatunku jak i model hipercyklu opisany w następnym podrozdziale są tak naprawdę jego szczególnymi przypadkami.

4.3 Równania różniczkowe oparte o hipercykl

Hipercykl opisany w rozdziale 3 może być zamodelowany następującym układem równań:

$$\dot{x}_i = x_i(k_i + \sum_j k_{i,j}x_j - \frac{1}{x}\phi) \quad (4.13)$$

gdzie x_i jest znormalizowanym stężeniem cząsteczki i (stężenia te jak w poprzednich podrozdziałach zawsze muszą sumować się do tej samej wartości - x), k_i to współczynnik określający nadprodukcję cząsteczki i (różnica między stałą samoreplikacji i rozpadu). Stałe $k_{i,j}$ są stałymi replikacji cząsteczki i katalizowanej przez cząsteczkę j (jest to wspomniany wcześniej element kooperacji w hipercyklu). Współczynnik ϕ z kolei gwarantuje, że wszystkie zmienne x_i zawsze będą się poprawnie sumowały.

Powyższe równania można uprościć jeśli wprowadzi się dwa dodatkowe założenia:

1. Replikację cząsteczki może katalizować tylko ona sama oraz poprzednia w cyklu.
2. Stężenia x_i są na tyle duże, że współczynniki k_i mogą zostać pominięte.

Należy pamiętać, że analizowany model jest cyklem, zatem:

$$k_{i,0} = k_{i,n} \quad (4.14)$$

$$x_0 = x_n \quad (4.15)$$

gdzie n to liczba cząsteczek w cyklu. Ostatecznie uproszczone równania przybierają następującą postać:

$$\dot{x}_i = x_i(k_{i,i-1}x_{i-1} - \frac{1}{x}\phi) \quad (4.16)$$

Najważniejszym wnioskiem z analizy powyższego układu równań jest fakt, że cykl dla $n > 4$ jest niestabilny [99, 111].

4.4 Automaty komórkowe

Automaty komórkowe zostały pokrótce omówione w rozdziale 2, w tym akapicie opisane zostanie ich możliwe wykorzystanie do badania Świata RNA. Przykładowa symulacja tego typu została stworzona przez Takeuchiego i Hogeweg [115], oparli ją oni o model RP (opisany w rozdziale 3). Podstawowe równania różniczkowe stanowiące bazę całego systemu mają następującą postać:

$$\dot{R} = k_R R^2 \theta - dR \quad (4.17)$$

$$\dot{P} = k_P R P \theta - dP \quad (4.18)$$

R oraz P to odpowiednio stężenia replikaz i pasożytów, θ reprezentuje zasoby potrzebne do przeprowadzenia replikacji, d to stała rozpadu, a k_R oraz k_P to stałe replikacji odpowiednio dla replikaz i pasożytów. Dla $k_R < k_P$ niemal natychmiast wynika, że pasożyty wyprą replikazy i doprowadzą do śmierci systemu. Okazuje się jednak, że z wykorzystaniem automatów komórkowych, za pomocą których można reprezentować struktury przestrzenne (podczas gdy powyższe równania nie zawierają aspektu przestrzennego), można otrzymać zupełnie odmienne wyniki. Obszarem symulacji u Takeuchiego i Hogeweg była płaszczyzna podzielona na kwadratowe komórki. Każda z nich mogła być pusta, co reprezentowało zasoby, lub zawierać pojedyncze RNA (pasożyta oraz replikazę). W każdym kroku wielokrotnie wybierano losowy punkt na planszy, po czym uwzględniając sąsiedztwo, odpowiedni algorytm wybierał jedną z możliwych reakcji: dyfuzję (zamianę dwóch sąsiednich komórek miejscami), rozpad (opróżnienie komórki) oraz replikację (pod warunkiem sąsiedztwa odpowiedniego RNA oraz pustej komórki, czyli zasobów). Symulacje przeprowadzone dla tak zdefiniowanego modelu pokazały, że system jest w stanie przeżyć dla $k_R < k_P$ dzięki odpowiednim strukturom przestrzennym. W tym przypadku miały one postać poruszających się fal, które składały się z szybko rozrastającej się populacji replikaz z przodu, oraz grupy pasożytów z tyłu, która „pożerała populację przed sobą”, co ostatecznie dawało złudzenie przemieszczających się struktur. Zderzenia dwóch grup często prowadziły do zniszczenia ich obu, ale całkowita populacja nie malała, ponieważ od fal czasem oddzielały się małe grupki replikaz, które szybko rozrastały się i po „infekcji” pasożytami z zewnątrz - dawały początek nowej fali [115].

Wadą tego podejścia było oczywiście ograniczenie powstających struktur i możliwości dyfuzji przez sztywny podział obszaru na kwadratowe komórki. W następnych rozdziałach poświęconych badaniom autora, problem ten zostanie rozwiązany przez wykorzystanie systemów wieloagentowych w których przestrzeń jest już ciągła.

Model wykorzystujący systemy wieloagentowe

Jak wspomniano we wcześniejszych rozdziałach, Świat RNA był wcześniej badany za pomocą automatów komórkowych. W tej części opisany zostanie wierniejszy rzeczywistości model, oparty na systemach wieloagentowych (opisane poniżej metody i rezultaty zostały opublikowane przez autora w [110]).

5.1 Model biologiczny

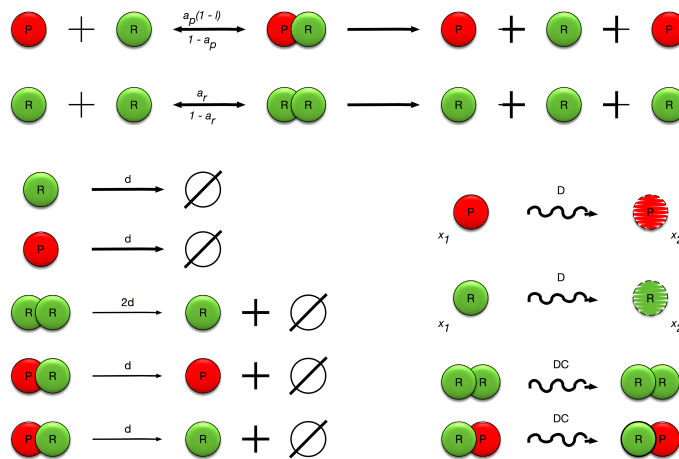
Podstawą dla symulacji wieloagentowych (i automatów komórkowych na których były wzorowane, np. [114]) był opisany w rozdziale 3 model RP. Badane cząsteczki RNA dzielą się na dwa rodzaje - replikazy (replikujące inne cząsteczki) oraz pasożyty (nie replikujące innych RNA, ale korzystające z możliwości replikacji w celu zwiększenia swojej liczebności). Każda cząsteczka może przebywać w stanie zwiniętym z prawdopodobieństwem l (które może być indywidualnie określone dla każdego RNA) co odpowiada naturalnej zdolności RNA do tworzenia skomplikowanych struktur przestrzennych. Zwinięta cząsteczka nie może być poddana replikacji, a jej stan musi być sprawdzany każdorazowo, ponieważ zakłada się że może zmieniać się on z tak dużą częstotliwością, że nie ma potrzeby reprezentowania go *explicite*. Kolejnym ważnym parametrem cząsteczki jest jej powinowactwo do replikaz a , które może być określone indywidualnie lub zbiorczo dla danego typu RNA. Wszystkie zachodzące w systemie reakcje przedstawione są na rys. 5.1.

5.2 Model matematyczny

5.2.1 Podstawowe parametry

W tym podrozdziale wykorzystane są następujące stałe:

1. D - stała dyfuzji pojedynczego RNA,
2. D_C - stała dyfuzji kompleksu,
3. d - stała szybkości rozpadu,
4. K - stała szybkości replikacji,
5. μ_a - prawdopodobieństwo mutacji parametru a agenta,
6. μ_l - prawdopodobieństwo mutacji parametru l agenta.



Rys. 5.1: W modelu biologicznym wykorzystanym do stworzenia symulacji wieloagentowej zachodzą cztery podstawowe rodzaje reakcji: dyfuzja, rozpad, formowanie kompleksów, replikacja oraz dysocjacja.

Źródło: [110]

5.2.2 Dyfuzja

W celu stworzenia algorytmu do symulacji, potrzebne były konkretne metody przemieszczania agentów. W automacie komórkowych obejmowałyby to jedynie przemieszczenie cząsteczki na odpowiednie pole, jednak w systemie wieloagentowym należy uwzględnić ciągłą przestrzeń, zatem wektor przemieszczenia również powinien być ciągły. Taką własność ma dynamika Browna, która była wspomniana w rozdziale 3 i w ramach której wektor przemieszczenia można było wyznaczyć ze wzoru 3.1, co wymagało jedynie podprocedury do generowania zmiennych losowych z rozkładem normalnym.

5.2.3 Reakcje pierwszorzędowe

W modelu dozwolone są trzy typy reakcji pierwszorzędowych: replikacja, rozpad oraz dysocjacja. Zakłada się, że stan agenta nie ma wpływu na moment zajścia powyższych reakcji (zachodzą one losowo), dlatego nie mają w nich udziału żadne zmienne reprezentowane w modelu teoretycznym *explicite*. Oznacza to, że prawdopodobieństwo ich wystąpienia nie zależy również od czasu, czyli rozkład prawdopodobieństwa reakcji pierwszorzędowych posiada własność braku pamięci, co wskazuje że jest on wykładniczy. Biorąc wszystkie fakty powyżej pod uwagę - należałoby sprawdzić co krok czy dana reakcja nie zaszła, ale generowanie każdorazowo losowej zmiennej dla wielu agentów obecnych w symulacji mogłoby być czasochłonne, dlatego w celu usprawnienia na wstępie wyliczany jest czas do zajścia danej reakcji, a potem dekrementuje się go co krok, aż do osiągnięcia 0 (co inicjuje odpowiedni proces). Wspomniany czas można wyliczyć zgodnie z rozkładem prawdopodobieństwa tak, żeby z zewnątrz nie było widać żadnej różnicy między tym podejściem a losowaniem co krok:

$$t_{akcji} = -\ln(X)/k \quad (5.1)$$

gdzie X to losowa liczba rzeczywista z przedziału $(0; 1]$, a k określa stałą szybkości danej reakcji pierwszorzędowej. Wartość k oraz moment w którym wyznacza się tak zdefiniowany przedział czasowy zależy od wybranej reakcji:

1. rozpad - czas wyliczany jest w momencie powstania nowego agenta (w wyniku replikacji, dysocjacji kompleksu, a także na etapie inicjalizacji symulacji), za k podstawiane jest d ,

2. replikacja - t_{akcji} wyznacza się kiedy tworzy się kompleks, wtedy za k podstawiane jest K ,
3. dysocjacja - podobnie jak powyżej, czas obliczany jest przy powstaniu kompleksu, k przyjmuje wtedy wartość $(1 - a)$, gdzie a to powinowactwo wzorca replikowanego RNA) do replikaz. Do dysocjacji kompleksu dochodzi też natychmiast po udanej replikacji.

Każdy agent powstały w wyniku replikacji z prawdopodobieństwem może mieć trochę inną wartość parametrów niż wzorec (mutacja). Powinowactwo do replikaz a modyfikowane jest z prawdopodobieństwem μ_a , a odsetek czasu w stanie zwiniętym l - z prawdopodobieństwem μ_l .

5.2.4 Reakcje drugorzędowe

W modelu zdefiniowane są dwa typy reakcji drugorzędowych (wymagających dwóch cząsteczek): tworzenie kompleksu replikaza-pasożyt oraz kompleksu replikaza-replikaza. Aby doszło do reakcji - molekuly muszą znaleźć się w odpowiednio bliskiej odległości, w omawianej symulacji jest ona zdefiniowana następująco:

$$r_c = r_i + r_j \quad (5.2)$$

gdzie r_i oraz r_j to promienie agentów reprezentujących substraty (są one reprezentowane jako koła), dla uproszczenia wszystkie agenty są takiej samej wielkości (określonej globalnym parametrem). Dla danej cząsteczki wszystkie inne, które znajdują się w odległości umożliwiającej reakcję są uznane za jej sąsiadów. Każdy agent posiadający zbyt dużą liczbę sąsiadów jest usuwany w celu ograniczenia maksymalnego stężenia. Ze względu na to, że nie każde spotkanie substratów kończy się reakcją - wprowadzone zostało odpowiednie prawdopodobieństwo ω , które jest bazowane na pracy Takeuchiego i Hogeweg [114]. W celu jego wyznaczenia brane są pod uwagę parametry wzorca: a oraz l , prawdopodobieństwo można wtedy obliczyć jako:

$$\omega = (1 - l)a \quad (5.3)$$

Można powyższy wzór interpretować jako zmodyfikowaną wartość a , biorącą pod uwagę fakt zwijania się RNA w strukturę. W każdym kroku dla każdego agenta w losowej kolejności przeglądani są jego sąsiedzi i z prawdopodobieństwem ω (jeśli możliwe jest stworzenie kompleksu) dochodzi do reakcji. Jeśli cząsteczka nie zareaguje z żadną inną - jest ona pozostawiana w takim samym stanie w jakim była na początku.

5.2.5 Wartości parametrów

Opisywany w tym rozdziale model miał być z założenia bardzo podobny do podejścia Takeuchiego i Hogeweg [114], co umożliwiłoby łatwe porównanie wyników. Z tego powodu podjęto wysiłek, żeby odpowiednio dostosować parametry symulacji tak, żeby odpowiadała ona tej opartej na automatach komórkowych. Warto podkreślić, że teoretycznie nie było to konieczne, ponieważ oba modele skupiają się na analizie jakościowej, a nie ilościowej. Udało się zapisać odpowiednie wzory dla dyfuzji, niestety większość pozostałych stałych musiała być wyznaczana eksperymentalnie, ponieważ w dużej mierze odpowiednie wartości parametrów zależały od struktury przestrzennej populacji agentów, zatem przeliczenie ich w prosty (analityczny) sposób byłoby niemal niewykonalne. Pojawił się również problem z krokiem czasowym Δt , gdyż taka wartość nie występowała w modelu Takeuchiego, musiała zostać zatem przyjęta arbitralnie - dla uproszczenia nadano jej wartość 1.

Poniżej zaprezentowano metodę wyznaczenia stałej dyfuzji dla modelu wieloagentowego, wartości parametrów z symulacji Takeuchiego wyróżniono indeksem dolnym CA .

5.2.5.1 Dyfuzja

Pojedynczy agent

W celu zanalizowania procesu dyfuzji, należy wziąć pod uwagę jedynie interakcje, które agent może mieć w ciągu swojego czasu życia. Z tego powodu wykluczone są z analizy reakcja kompleksowania i rozpad, ponieważ pierwsza z nich powoduje zniknięcie cząsteczki z symulacji *implicite*, a druga *explicite*. Ostatecznie dyfuzja jest jedyną braną pod uwagę akcją, zatem prawdopodobieństwo jej zajścia ma wartość:

$$p_{D_{CA}} = 1 \quad (5.4)$$

Porównanie stałych dyfuzji w dwóch omawianych modelach wymaga wzięcia pod uwagę średnich kwadratów długości wektorów przemieszczenia (równą ich wariancji, ponieważ średnia wynosi 0). W modelu komórkowym wektor przemieszczenia może przyjmować jedynie skończoną liczbę wartości, zatem konieczne jest obliczenie średniej ważonej z uwzględnieniem odpowiednich prawdopodobieństw. Z przyjętej przez Takeuchiego definicji sąsiedztwa wynika, że z równym prawdopodobieństwem, wektor przemieszczenia może przyjąć jedną z 8 wartości: 4 dla komórek przyległych oraz 4 dla komórek obecnych w rogach. Ostatecznie wariancję można zapisać jako:

$$\sigma_{CA}^2 = \frac{1}{8}(4 \cdot 1^2 + 4 \cdot \sqrt{2}^2) \cdot p_{D_{CA}} = \frac{3}{2}p_{D_{CA}} \quad (5.5)$$

W systemie wieloagentowym wariancję można wyliczyć bazując na dynamice Browna:

$$\sigma^2 = 2D\Delta t \quad (5.6)$$

gdzie D to oczywiście stała dyfuzji w symulacji wieloagentowej.

Jeśli dwa modele mają zachowywać się w sposób podobny, oczywiste jest że:

$$\sigma_{CA}^2 = \sigma^2 \quad (5.7)$$

Podstawiając (równania 5.5 i 5.6):

$$\frac{3}{2}p_{D_{CA}} = 2D\Delta t \quad (5.8)$$

$$D = \frac{3p_{D_{CA}}}{4\Delta t} \quad (5.9)$$

Kompleks

W modelu Takeuchiego kompleksy poruszają się z dwukrotnie niższym prawdopodobieństwem niż pojedyncze cząsteczki. Podobnie jak w poprzedniej sekcji, należy skupić się na akcjach, które kompleks może wykonywać w ciągu swojego czasu życia. Podobnie jak poprzednio należy wykluczyć rozpad jednej ze skompleksowanych cząsteczek, a także dysocjację. W tym przypadku oprócz dyfuzji pozostaje także replikacja. Prawdopodobieństwo dyfuzji wyraża się w takim razie wzorem:

$$p_{D'_{CA}} = \frac{\frac{D_{CA}}{2}}{\frac{D_{CA}}{2} + K_{CA}} \quad (5.10)$$

gdzie D_{CA} to stała dyfuzji, a K_{CA} to stała replikacji w modelu komórkowym.

Należy tutaj podkreślić, że kompleks składa się w tym przypadku z dwóch sąsiednich komórek, zatem przy obliczaniu wariancji wektora przemieszczenia należy wziąć pod uwagę środek masy. Dyfuzja skompleksowanej cząsteczki polega na wybraniu z równym prawdopodobieństwem jednej z 8 sąsiednich komórek (w identyczny sposób jak dla nieskompleksowanej cząsteczki) i przemieszczenia jej tam. Druga część kompleksu zostaje wtedy przesunięta do nowo powstałego pustego

poła. Jeśli przy dyfuzji wybrana zostanie druga część kompleksu - dojdzie do zamiany obu części i środek masy się wtedy nie przesunie. W pozostałych 7 przypadkach przesunięcie nie będzie zerowe - 4 przypadki dają długość wektora $\sqrt{2}/2$, 1 przypadek odpowiada długości 1, a pozostałe 2 - $\sqrt{10}/2$. Uwzględniając powyższe fakty, można napisać równanie dla wariancji:

$$\sigma_{CA}^2 = \frac{1}{8} \left(4 \cdot \left(\frac{\sqrt{2}}{2} \right)^2 + 1^2 + 2 \cdot \left(\frac{\sqrt{10}}{2} \right)^2 \right) \cdot (2p_{D'_{CA}}) = 2p_{D'_{CA}} \quad (5.11)$$

Prawdopodobieństwo $p_{D'_{CA}}$ należy pomnożyć przez 2, ponieważ kompleks składa się z dwóch części i każda z nich może zainicjować ruch całości.

W modelu wieloagentowym ponownie można łatwo wyznaczyć wariancję z dynamiki Browna:

$$\sigma_C^2 = 2D'\Delta t \quad (5.12)$$

gdzie D' to stała dyfuzji dla kompleksów.

Dwie analizowane symulacje powinny zachowywać się w podobny sposób jeśli:

$$\sigma_C^2 = \sigma_{CA}^2 \quad (5.13)$$

Podstawiając (równania 5.11 i 5.12):

$$2p_{D'_{CA}} = 2D'\Delta t \quad (5.14)$$

$$D' = \frac{p_{D'_{CA}}}{\Delta t} \quad (5.15)$$

5.2.5.2 Stała replikacji

W symulacji Takeuchiego kompleks może zdyfundować, zdysocjować, zreplikować lub może rozpaść się któraś z jego składowych. Replikacja, podobnie jak rozpad i dysocjacja, zachodzi zgodnie z rozkładem wykładniczym, który determinuje czas zajścia odpowiedniej reakcji. Powyższe reakcje pierwszorzędowe są kompletnie niezależne od siebie. Warto wspomnieć, że replikacja jako wielostopniowy proces może być modelowana za pomocą delty Diraca. Takie podejście powoduje, że czas jej zajścia staje się deterministyczny. Chociaż w algorytmie dla replikacji użyty jest również rozkład wykładniczy, można udowodnić że dla bardzo dużej wartości stałej K faktycznie dąży on do delty Diraca. Z tego powodu K przypisana jest ekstremalnie duża wartość, co powoduje, że replikacja zachodzi niemal natychmiast po utworzeniu się kompleksu. Takie podejście do replikacji nie oznacza jednak, że zawsze do niej dochodzi - kompleks wciąż może wcześniej zdysocjować.

5.3 Algorytm

Poniżej przedstawiony jest pseudokod algorytmu (algorytm 1) zastosowanego do symulacji, warto wspomnieć, że rozkład początkowy pasożytów i replikaz w przestrzeni może być losowy lub z góry narzucony. Należałoby jeszcze skomentować fakt, że przy dysocjacji kompleksu agenty nie mają wyliczanych nowych czasów życia - w związku z własnością braku pamięci odpowiedniego rozkładu - nie jest to konieczne. Domyślne parametry podane są w tabeli 5.1.

5.4 Implementacja

Pierwszym autorem wykorzystanego w badaniach symulatora był p. Iwo Błądek, później pracę nad nim kontynuował autor niniejszej rozprawy. Program był ogólnego przeznaczenia, a poszcze-

Algorytm 1 Algorytm zastosowany do symulacji wieloagentowej

Zainicjalizuj początkową liczbę i pozycję agentów z pliku lub losowo
Zainicjalizuj wszystkie replikazy z równymi wartościami a_R oraz l_R
Zainicjalizuj wszystkie pasożyty z równymi wartościami a_P oraz l_P
while (czas symulacji < maksymalny czas symulacji) **and**(obecne są zarówno pasożyty jak i replikazy (skompleksowane lub wolne)) **do**
 Zwiększ obecny czas symulacji o Δt
 Zmniejsz czas życia (RLL) każdego agenta (w przypadku kompleksów jest to czas do dysocjacji)
 for all nieskompleksowanych agentów x_i z $RLL = 0$ **do**
 Usuń x_i z symulacji
 end for
 Ustaw agenty w kolejności losowej
 for all Agentów x_i **do**
 if x_i jest skompleksowany **and** $RLL(x_i) = 0$ **then**
 Zdysocjuj x_i na dwa agenty składowe i umieść je w tym samym miejscu w którym znajdował się kompleks
 end if
 for all Agentów $x_j \neq x_i$, który sąsiaduje z x_i (patrz równanie 5.2) **do**
 Dodaj x_j do zbioru sąsiadów N
 end for
 if $|N| > N_{max}$ **then**
 Usuń agenta x_i z symulacji
 end if
 Ustaw elementy zbioru N w losowej kolejności
 Zdyfunduj x_i (równanie 3.1)
 if x_i jest kompleksem **then**
 Zmniejsz czas do replikacji TTR
 if $TTR = 0$ **then**
 Stwórz nowego agenta x' będącego kopią wzorca w kompleksie x_i
 Podaj x' mutacji
 Zdysocjuj x_i (stwórz agenty r oraz t bazując na replikazie i wzorcu, umieść je w miejscu kompleksu, przywróć im czas życia sprzed stworzenia kompleksu oraz usuń x_i)
 end if
 else
 for all $n_j \in N$ **do**
 if n_j jest nieskompleksowany **and** n_j lub x_i jest replikazą **and** zgodnie z prawdopodobieństwem zachodzi reakcja (równanie 5.3) **then**
 Stwórz kompleks x'
 Wylicz $TTR(x')$ (równanie 5.1)
 Wylicz $RLL(x')$ (równanie 5.1)
 Zainicjalizuj pozycję x' jako średnią z pozycji n_j oraz x_i
 Usuń n_j oraz x_i z symulacji
 end if
 end for
 end if
 end for
end while

Nazwa	Wartość domyślna	Opis
d	0,182	stała rozpadu
D	0,75	stała dyfuzji
D'	0,0476	stała dyfuzji kompleksu
K	239800	szybkość mutacji pasożyta
Δt	1	długość kroku czasowego
r	0,5	promień agenta

Tab. 5.1: Domyślne parametry w symulacji wieloagentowej Świata RNA.

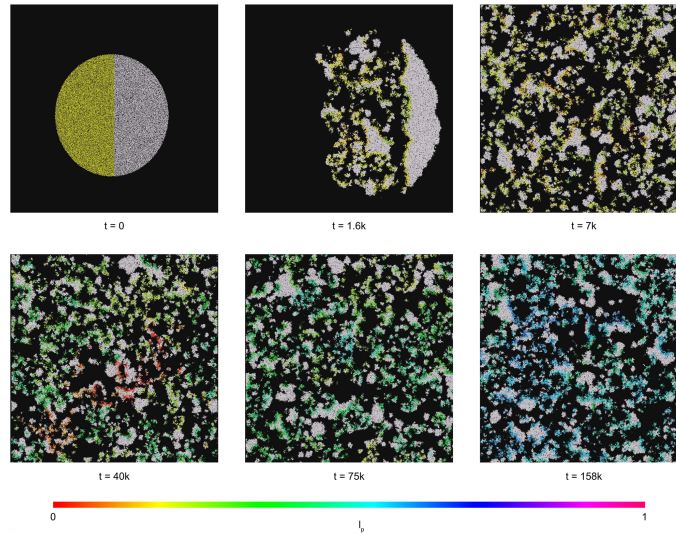
gólne modele były w nim zaimplementowane jako moduły w języku C#, co umożliwiło łatwe rozszerzanie jej w przyszłości o dodatkowe symulacje. Na funkcjonalność opisywanej aplikacji składały się także:

1. podgląd stanu symulacji w czasie rzeczywistym,
2. możliwość wstrzymania symulacji oraz kontynuowania jej przez z góry określoną liczbę kroków,
3. eksport i import parametrów do pliku XML,
4. generowanie plików wideo, wykresów oraz tabel ze szczegółowymi danymi na temat przebiegu symulacji,
5. wczytywanie plików z początkowym rozstawieniem agentów (oprócz możliwości wyboru rozstawienia losowego).

Program był przystosowany do uruchamiania w systemie operacyjnym Windows, głównie od Windowsa 7 w górę. Wykorzystywał on tylko jeden wątek procesora, co niestety ograniczało maksymalną liczbę agentów (niektóre symulacje trwały kilka dni), jednak zmiana tego stanu okazała się zbyt czasochłonna, w związku z czym stworzono nowy symulator oparty o interfejs MPI, opisany w dalszej części rozprawy.

5.5 Wyniki

Opisany powyżej algorytm uruchamiano wielokrotnie w celu przetestowania dużej liczby różnych scenariuszy oraz zbadania jak na zachowanie się systemu wpływa wartość poszczególnych parametrów. Ustawienie początkowe było losowe lub z góry narzucone - w niektórych testach inicjalizowano populację jako koło podzielone na dwie równe części, jedną stanowiły replikazy, a drugą pasożyty. Populacje obu typów agentów zawsze były na początku równe. Mutację symulowano w dwóch wariantach, w pierwszym podlegały jej jedynie pasożyty (podobnie jak w modelu Takeuchiego i Hogeweg [114]), a w drugim również replikazy. We wszystkich symulacjach mierzono średnie wartości parametrów w każdym kroku, prowadzono je aż do ustabilizowania się systemu. Za stan stabilny uznawany moment w którym średnie wartości parametrów nie zmieniały się, a rozkład przestrzenny poszczególnych populacji agentów był równomierny, najczęściej działo się to po ok. 150 tysięcy krokach. Warto wspomnieć, że obserwowane zachowanie było podobne do wyników Takeuchiego i Hogeweg [114] - również dostrzeżono fale, które przemieszczały się po obszarze symulacji, z aktywnymi replikazami na czole oraz pasożytami powodującymi lokalne wymieranie populacji z tyłu. Przykładowy przebieg symulacji przedstawiono na rys. 5.2



Rys. 5.2: Zrzuty ekranu zrobione podczas pierwszej symulacji (skupiającej się na mutacji pasożytów) po odpowiednio 7, 40, 75 oraz 158 tysiącach kroków.

Źródło: [110]

5.5.1 Mutacja pasożytów

W celu sprawdzenia jaki wpływ ma mutacja pasożytów na przebieg symulacji sprawdzono cztery scenariusze, we wszystkich rozmieszczeniem początkowym było opisane wcześniej koło. Parametry zastosowane przy każdym ze scenariuszy podano w tabeli 5.2.

Podczas pierwszej symulacji zarówno a_P jak i l_P mogły powoli ewoluować. W trakcie następnym (scenariusze 2 - 4) mutował już tylko jeden parametr (w celu zbadania czy szybkość mutacji l_P ma inny wpływ na system niż szybkość mutacji a_P), a szybkość mutacji została zwiększona aby łatwiej było obserwować efekty. W ramach scenariuszy 2, 3 i 4 sprawdzano też różne wartości a_{P_0} oraz l_{P_0} (zakres od 0 do 1 z krokiem 0,1). Powinowactwo replikazy do replikazy (a_R) ustawiono na 0,7, była to wartość trochę większa niż 0,6 w symulacji Takeuchiego i Hogeweg [114], jednak testy pokazały że ta zmiana nie wpływa na ogólne zachowanie systemu, a za to prowadzi do stabilniejszej koegzystencji pasożytów i replikaz, co ułatwiło obserwację wyników. Jeśli w danym scenariuszu parametr a_P miał stałą wartość - była ona ustawiana na 0,55, podobnie jak l_P było w analogicznym przypadku równe 0,2, liczby te pochodzą z analizy równań różniczkowych wykonanej przez Takeuchiego i Hogeweg [114]. Wymienieni autorzy pokazali, że przy braku mutacji takie wartości pozwalają na koegzystencję pasożytów i replikaz. Zasada ta przenosiła się na model bazujący na automatach komórkowych i okazała się też nadal obowiązywać w przypadku symulacji wieloagentowych.

Badanie poszczególnych scenariuszy pokazało, że system może przetrwać dla bardzo szerokiego

Scenariusz	a_{P_0}	l_{P_0}	μ_{a_P}	μ_{l_P}
1	0,55	0,2	0,01	0,01
2	0,55	0,0-1,0 (krok 0,1)	N/A	0,19
3	0,0-1,0 (krok 0,1)	0,2	0,19	N/A
4	0,0-1,0 (krok 0,1)	0,2	N/A	0,19

Tab. 5.2: Parametry wykorzystane przy symulowaniu mutacji pasożytów. a_{P_0} to początkowo wartość powinowactwa do replikaz wszystkich pasożytów, l_{P_0} to początkowa prawdopodobieństwo przebywania przez pasożyta w stanie zwiniętym, a μ_{a_P} oraz μ_{l_P} to odpowiednie prawdopodobieństwa mutacji obu wymienionych parametrów.

zakresu parametrów. Wzięcie struktury przestrzennej pod uwagę pozwoliło na stabilną koegzystencję nawet pomimo przewidzianej teoretycznie niestabilności ewolucyjnej układu pasożytów i replikaz [38]. Dość zaskakującą obserwacją (scenariusz 1, rys. 5.2) jest tendencja systemu do zwiększania zarówno a_P oraz l_P , ponieważ zmiany te działają przeciwstawnie. Pierwsza z nich sprawia, że pasożyty mają większą szansę wiązania się z replikazami (przy kolizjach), ale rosnąca wartość l_P z kolei powoduje, że przez większość czasu są nieaktywne i niezdolne do tworzenia kompleksów. Bardzo podobny trend zaobserwowali Takeuchi i Hogeweg [114], pokazali też oni, że populacja przechodzi przez dwie fazy: na początku średnie wartości a_P oraz l_P nie zmieniają się, jednak w drugiej fazie zaczynają one rosnać. W symulacji wieloagentowej pierwszy etap trwał mniej więcej 25 tysięcy kroków. Ilustrują to dość dobrze rys. 5.3 i 5.4.

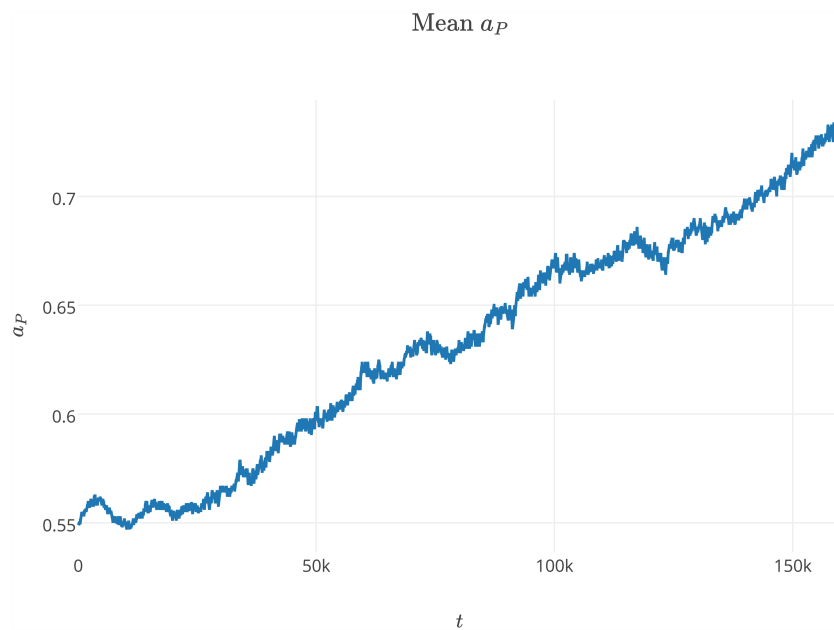
Po scenariuszu 1, który częściowo odtworzył wyniki uzyskane przez Takeuchiego i Hogeweg, pozwolono mutować jedynie zmiennej l_P (scenariusz 2). W celu jak najlepszego zrozumienia wpływu wartości początkowej na rezultaty - przeprowadzono symulację dla 11 różnych wartości początkowych l_P w przedziale 0,0 - 1,0 (z krokiem 0,1). Okazało się, że system mógł zachować się w zupełnie inny sposób dla różnych badanych przypadków:

1. Dla $l_{P_0} = 0$ pasożyty były aktywne przez 100% czasu, przez ich liczebność szybko wzrastała i wypierały one replikazy doprowadzając do śmierci całego systemu (patrz rys. 5.5).
2. Dla $0,1 \leq l_{P_0} \leq 0,6$ wartość l_P szybko ewoluowała i średnia ostatecznie stabilizowała się między 0,2 oraz 0,3. Cały system również osiągał wtedy stan stabilny (patrz rys. 5.6). Obserwacje pokazały jednak, że formowały się fale w obrębie których pasożyty zawsze dążyły do maksymalizacji l_P . Fakt, że mimo tego średnia wartość l_P dla populacji może pozostać stała został wyjaśniony przez Takeuchiego i Hogeweg [114]. Okazuje się, że chociaż lokalnie ewolucja przebiega w kierunku wyższego l_P , na poziomie całych fal presja selekcyjna działa w drugą stronę i większą szansę przetrwania i propagacji mają fale z mniejszą wartością l_P pasożytów.
3. Dla $l_{P_0} \geq 0,7$ pasożyty były zbyt mało aktywne, żeby osiągnąć odpowiedni przyrost i wymierały (patrz rys. 5.5).

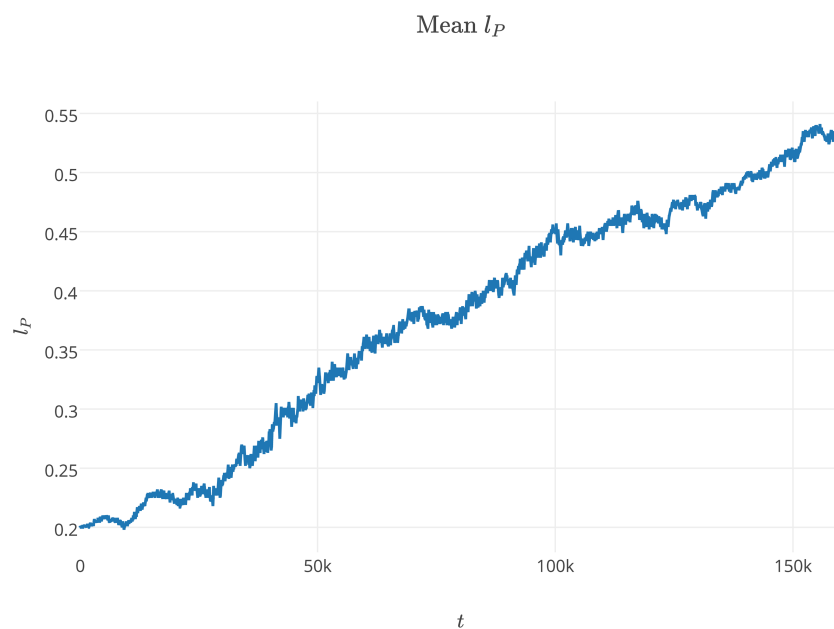
Symulacje w scenariuszu 3 były niejako przeciwieństwem scenariusza 2. W tym przypadku parametr a_P podlegał mutacji i miał inną wartość początkową dla każdej symulacji, podczas gdy l_P było zawsze równe 0,2. Okazało się, że system nie był w stanie przetrwać w żadnym przypadku. Dla $a_{P_0} \leq 0,3$ pasożyty nie były replikowane wystarczająco często i całkowicie wymierały. Z kolei dla $a_{P_0} \geq 0,4$ wartość a_P szybko rosła przez co replikazy przegrywały z pasożytami i były coraz rzadziej replikowane, co doprowadzało do ich zniknięcia i śmierci całego systemu. Ważną obserwacją było też to, że pasożyty zaczynają wypierać replikazy dla $a_P > 0,4$ i jeśli nie było to rekompensowane przez jednoczesny wzrost średniej wartości l_P - doprowadzały one do wymarcia całej populacji. Ewolucja średniej wartości a_P dla różnych symulacji jest przedstawiona na rys. 5.7.

Ostatni scenariusz (4) miał jeszcze lepiej zbadać rolę stabilizującą l_P . Dla różnych wartości a_P (nie podlegających mutacji), badano czy parametr l_P może dostosować się i umożliwić koegzystencję pasożytów i replikaz, nawet wtedy kiedy a_P jest duże. Podobnie jak w scenariuszu 2, zaobserwowano trzy różne możliwości zachowania się systemu:

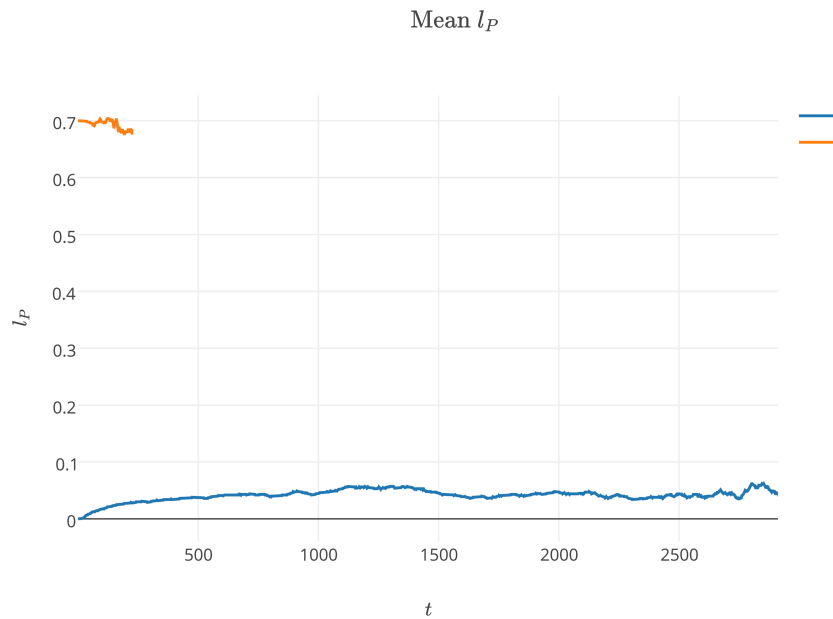
1. Dla $a_{P_0} \leq 0,3$ pasożyty zbyt rzadko wiązały się z replikazami i szybko wymierały. Warto dodać tutaj, że ten przypadek jest najmniej realistyczny, ponieważ utrata powinowactwa w kierunku enzymu w wyniku mutacji jest znacznie mniej prawdopodobna niż utrata zdolności katalitycznych wymagających często określonej sekwencji. Argumentem przemawiającym za tym może być fakt, że fragmenty RNA rozpoznawane przez (albo wiążące) katalizator są zwykle słabiej zakonserwowane [109].
2. Dla $0,4 \leq a_{P_0} \leq 0,6$ system osiąga stan stabilny, a wartość l_P szybko stabilizuje się, przy-



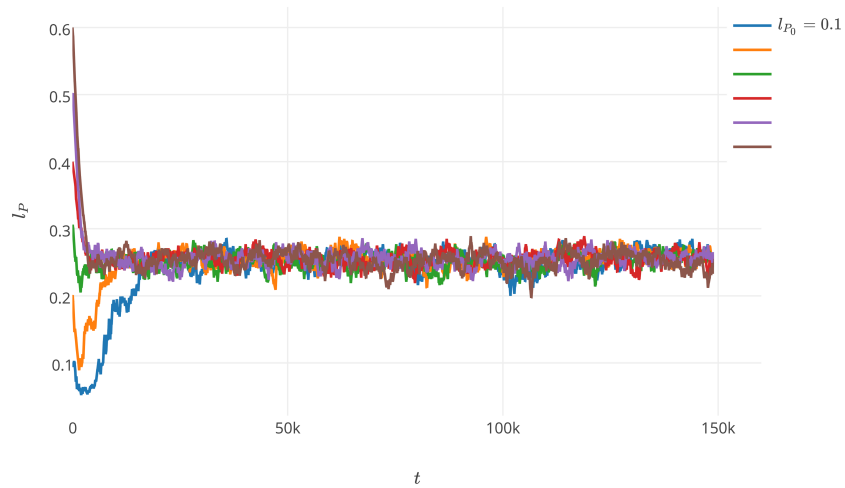
Rys. 5.3: Ewolucja parametru a_P w scenariuszu 1.
Źródło: [110]



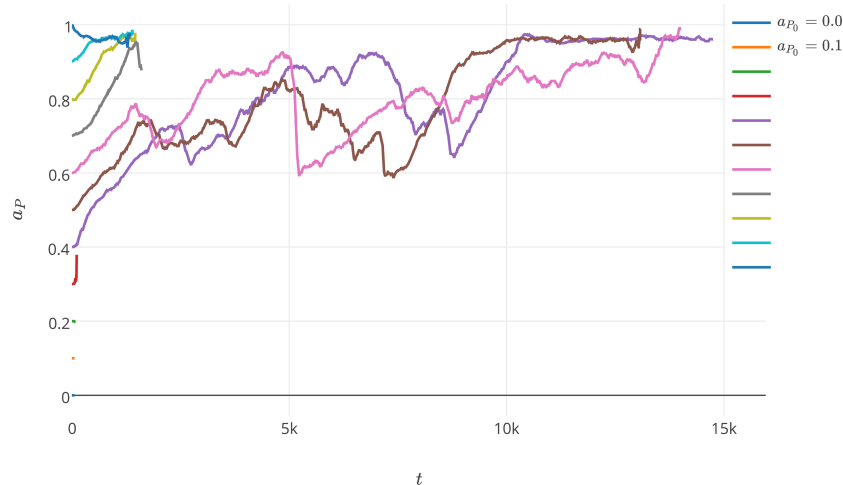
Rys. 5.4: Ewolucja parametru l_P w scenariuszu 1.
Źródło: [110]



Rys. 5.5: Ewolucja parametru l_P w scenariuszu 2 dla różnych wartości początkowych 0,0 oraz 0,7. W pierwszym przypadku pasożyty doprowadzają po czasie do śmierci replikaz, a w drugim same zostają wyparte z systemu.
Źródło: [110]



Rys. 5.6: Ewolucja parametru l_P w scenariuszu 2 dla różnych wartości początkowych z przedziału 0,1 – 0,6.
Źródło: [110]



Rys. 5.7: Ewolucja parametru a_P w scenariuszu 3 dla różnych wartości początkowych z przedziału 0,0 – 1,0.

Źródło: [110]

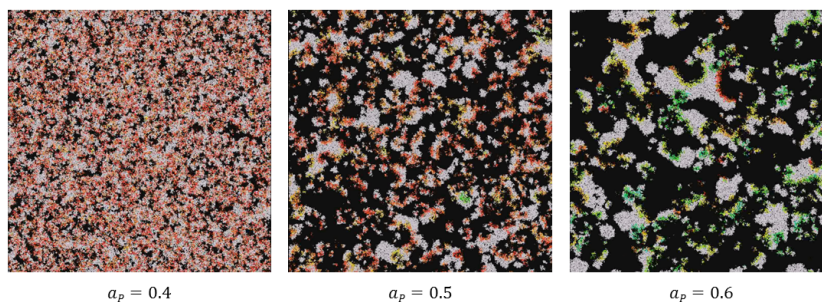
mując wartość ok. 0,05, 0,1, 0,3 dla odpowiednio $a_P = 0,4$, 0,5 oraz 0,6. Należy tutaj dodać, że wartość a_P ma wpływ na wielkość i strukturę powstających fal, które są tym mniejsze i mają tym mniej widoczną strukturę im mniejsza jest wartość a_P (patrz rys. 5.8 oraz rys. 5.9).

3. Dla $a_{P_0} \geq 0,7$ pasożyty doprowadzały do wyparcia replikaz i śmierci całej populacji.

Uzyskane wyniki można podsumować następująco - parametr l_P ma kluczową rolę w stabilizacji systemu i w dużej mierze dzięki niemu możliwa jest koegzystencja replikaz i pasożytów.

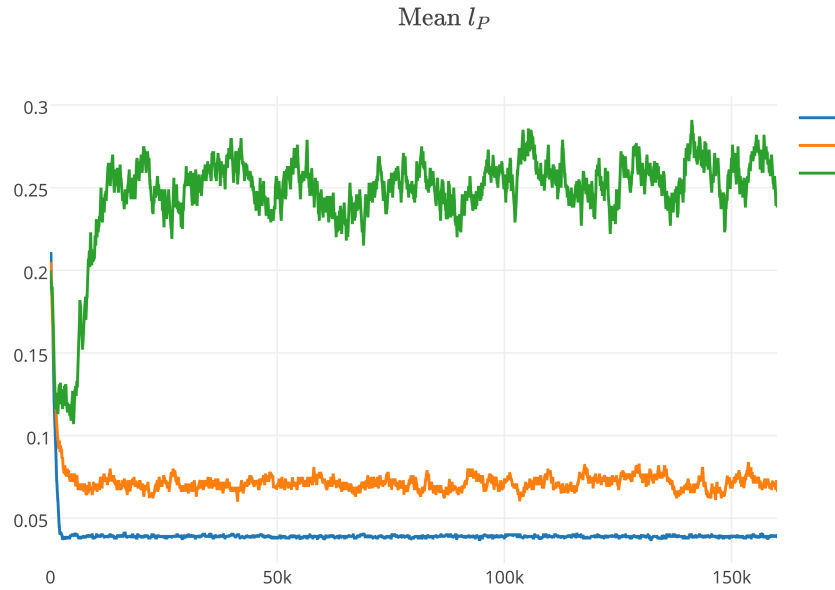
5.5.2 Mutacja pasożytów i replikaz

Poprzednia sekcja poświęcona była badaniom nad wpływem mutacji pasożytów na system. Logicznym krokiem było rozszerzenie modelu i wprowadzenie również mutacji replikaz, co mogło rzucić nieco więcej światła na prawa rządzące koegzystencją replikaz i pasożytów. Eksperymenty prowadzone w tym kierunku można zbiorczo nazwać scenariuszem piątym, mutować mogły w nim cztery parametry: a_P , l_P , a_R i l_R . Testy wykonano dla różnych wartości początkowych a_P oraz a_R w dwóch wariantach - agentów rozmieszczonych losowo oraz w opisanym wcześniej kole. Parametry l_{P_0} i l_{R_0} ustawione były zawsze na 0,2, prawdopodobieństwo mutacji μ dla wszystkich czterech mutujących zmiennych równe było 0,01, a wszystkie inne parametry symulacji miały swoją



Rys. 5.8: Zrzuty ekranu z symulacji wykonanych dla scenariusza 4 dla różnych wartości a_{P_0} . Zostały one zrobione po ok. 80 tysiącach kroków.

Źródło: [110]



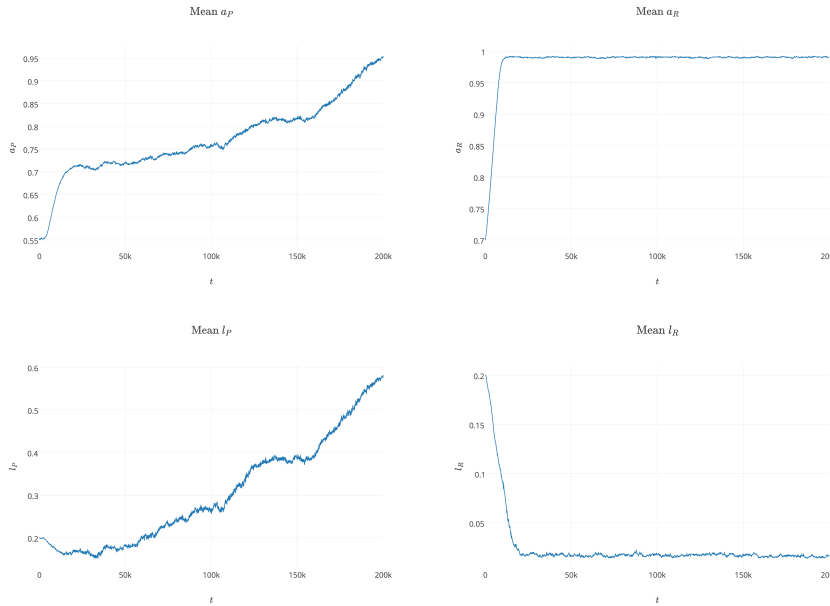
Rys. 5.9: Ewolucja parametru l_P dla różnych wartości początkowych a_P w scenariuszu 4.
Źródło: [110]

wartość domyślną. Wyniki dotyczące przeżywalności systemu zaprezentowane są w tabeli 5.3.

Okazało się, że system przy losowym rozmieszczeniu początkowym nie był w stanie przetrwać w większości przypadków, z kolei dla koła osiągnięcie stanu stabilnego było możliwe częściej. System mógł przeżyć jeśli wartość początkowa a_P była znacznie mniejsza od wartości początkowej a_R , przy czym obie musiały być na tyle duże, żeby odpowiednio pasożyty i replikazy mogły utrzymać swoją populację. Ewolucję przykładowej populacji, która była zdolna do przetrwania, ilustruje rys. 5.10. Wartość a_P szybko wzrosła z 0,55 do 0,7 na początku symulacji, po czym tempo jej wzrostu zwolniło. Średnia wartość l_P najpierw minimalnie spadła, żeby potem zacząć powoli stale się zwiększać, trend ten zaobserwowali także Takeuchi i Hogeweg w swojej symulacji [114]. Dla replikaz ewolucja przebiegała błyskawicznie, wartość a_R dość szybko osiągnęła swoją maksymalną wartość bliską 1, jednocześnie średnie l_R spadło poniżej 0,5. Oba te parametry zmieniały się tak, aby replikazy zwiększały swoją liczebność jak najszybciej, rosnące a_R zwiększyło prawdopodobieństwo tworzenia kompleksów replikaza-replikaza, podczas gdy malejące l_R sprawiło, że replikazy znacznie mniej czasu spędzały w stanie zwiniętym, przez co lepiej służyły jako wzorce do replikacji (co jeszcze bardziej zwiększyło powstającą liczbę kompleksów RR).

a_{P_0}	a_{R_0}	koło	losowe
0,3	0,5	wymiera	wymiera
0,4	0,6	przeżywa	wymiera
0,4	0,7	przeżywa	wymiera
0,55	0,6	wymiera	wymiera
0,55	0,7	przeżywa	przeżywa
0,8	0,6	wymiera	wymiera
0,8	0,7	wymiera	wymiera

Tab. 5.3: Wyniki testów przeprowadzonych w ramach scenariusza 5 dla różnych wartości a_{P_0} i a_{R_0} i dwóch wariantów rozmieszczenia początkowego.



Rys. 5.10: Ewolucja średnich wartości parametrów w scenariuszu 5 dla $a_{P_0} = 0,55$, $a_{R_0} = 0,7$, $l_{P_0} = 0,2$ i $l_{R_0} = 0,2$.
Źródło: [110]

5.6 Dyskusja

Jak wspomniano wcześniej, opisana tutaj wersja modelu RP bazuje na tym wykorzystanym wcześniej przez Takeuchiego i Hogeweg [114]. Obszar symulacji był dwuwymiarowy, co wynika z obserwacji, że proces polimeryzacji rybonukleotydów zachodzi znacznie wydajniej na powierzchniach mineralnych i że produkty reakcji mogą pozostawać na wspomnianej powierzchni, co lokalnie pozwoliłoby utrzymać odpowiednie stężenia [86, 87]. Wnioski otrzymane z symulacji (głównie fakt grupowania się replikaz) nie są zaskakujące, jeśli weźmie się pod uwagę, że są one altruistyczne - replikują cząsteczki innych rodzajów, przez co mają mniejszą możliwość powielania siebie nawzajem [85]. Wciąż takie skupiska mogą być infekowane przez pasożyty, które lokalnie doprowadzają do wymarcia populacji. Okazuje się, że ten problem jest rozwiązywany przez selekcję wyższego poziomu, w której całe populacje (w tym przypadku fale) konkurują ze sobą o zasoby (tutaj jest to oczywiście przestrzeń). Przy obecności wyższych poziomów selekcji, to w jakim stopniu dany fenotyp zwiększa szansę przeżycia i propagacji zależy zarówno od przystosowania osobnika rozpatrywanego indywidualnie, jak i od cech całej ich grupy. Zjawisko to może doprowadzać do faworyzowania cech, które w ujęciu lokalnym (i krótkoterminowo) są niekorzystne, takich jak np. wspomniany już altruizm [36]. W opisanych powyżej symulacjach dokładnie na tej zasadzie działało zwiększanie się parametru l_P z czasem, chociaż lokalnie takie pasożyty były mniej powielane, nie doprowadzały one dzięki temu do śmierci całej grupy i przez to mogły przetrwać.

Model wieloagentowy z sekwencjami

Dalsze badania opisane w tym rozdziale skupiły się bardziej na przekazywaniu i ewolucji informacji genetycznej. Opisany tutaj algorytm oraz wyniki zostały wcześniej opublikowane przez autora w [105].

6.1 Model biologiczny

Symulacje opisane poniżej miały na celu zejście do niższego poziomu w Świecie RNA i zbadanie jak selekcja w prebiotycznej zupie mogłaby wpływać na sekwencje poszczególnych cząsteczek RNA. Model RP wykorzystany (i symulowany w sposób wieloagentowy) w rozdziale 5 został więc rozszerzony o informację genetyczną reprezentowaną dla każdego agenta *explicite*. Parametry a oraz l , podobnie jak rola (pasożyt/replikaza), nie są już określone arbitralnie dla danej cząsteczki, ale są funkcjami jej sekwencji nukleotydów. Prawdopodobieństwo przebywania w stanie zwiniętym w modelu określane jest na podstawie struktury drugorzędowej. Wartość a , która w poprzednim modelu determinowała powinowactwo wzorca do replikazy, stała się niejako przeciwną cechą - teraz reprezentuje aktywność enzymatyczną agenta, zerowa wartość oznacza pasożyta, a jakakolwiek inna - replikazę. Parametr a jest wyznaczany na podstawie podobieństwa sekwencji RNA do ustalonego wzorca. Ze względu na fakt, że w momencie prowadzenia opisanych tutaj rozważań (2020 rok) nie udało się otrzymać w pełni funkcjonalnej replikazy w laboratorium, na wzorec wybrano sekwencję z końca 5' sztucznie zaprojektowanego rybozomu tC19Z (*UCAUUGAAAAAAAAAAGACAA*) [125]. Kolejną zmianą w stosunku do rozdziału 5 jest wymaganie, żeby replikaza była zwinięta jeśli ma działać jako enzym, oczywiście będąc wzorcem replikacji musiała być rozwinięta. Ostatnia modyfikacja polegała na wykluczeniu dyfuzji oraz dysocjacji kompleksów (oprócz tej będącej konsekwencją udanej replikacji) w celu ułatwienia stworzenia algorytmu równoległego do symulacji.

Wszystkie RNA w symulacji miały założoną stałą długość - 50 nukleotydów, a ich struktura drugorzędowa (potrzebna do wyliczenia l) wyznaczana była za pomocą programu IPknot, narzędzia wykorzystującego paczkę ViennaRNA jako bazę i uzupełniającego wynik dodatkowo o pseudowęzły [96].

Warto jeszcze wspomnieć, że jeden z rozważanych w późniejszej części scenariuszy uwzględnia fakt, że wiązania fosfodiesterowe pomiędzy różnymi parami rybonukleotydów mają różny czas życia, obliczając szybkość rozpadu dla każdej cząsteczki bazując na jej sekwencji. Dane dotyczące trwałości każdego 16 możliwych rodzajów wiązań fosfodiesterowych (biorąc pod uwagę jedynie rozpad nieenzymatyczny) zostały zaczerpnięte z prac Kierzka i in. [13, 14, 60, 61]. W przytoczonych publikacjach badano stabilność wiązań w roztworze o następującym składzie:

1. 0,1 pmol fosforylowanych (izotopem fosforu 32) na końcu 5' oligonukleotydów,
2. 1 mM EDTA,

3. 1 mM spermidyny,
4. 50 mM chlorku sodu,
5. 0,1% PVP (poliwinylpiperolidonu),
6. 50 mM buforu Tris-HCL.

Wzorcowe pH całości wynosiło 7,5, a temperatura - 37° C. Są to warunki, które panują wewnątrz wielu komórek eukariotycznych i nie ma pewności, czy są one zgodne ze środowiskiem wodnym w którym ewoluował Świat RNA, niestety w momencie powstawania opisanego w tym rozdziale artykułu (2020r.) nie udało się autorom znaleźć analogicznych analiz, które zostałyby przeprowadzone w innych warunkach. Dalszy ciąg eksperymentu Kierzka polegał na rozdzielaniu produktów reakcji za pomocą elektroforezy w 20% żelu poliakrylamidowym i pomiarze za pomocą autoradiografii (w celu wykrycia wspomnianego wcześniej fosforu 32). Badacze zauważyli, że szybkość rozpadu wiązań fosfodiesterowych faktycznie zależała od połączonych rybonukleotydów, a co więcej była większa dla fragmentów jednoniciowych [61].

6.2 Model matematyczny

Dyfuzja, reakcje pierwszorzędowe oraz prawdopodobieństwo utworzenia kompleksu są wyliczane w identyczny sposób jak w rozdziale 5. Nowym elementem jednak jest sposób wyznaczania parametrów a oraz l dla każdego agenta, jak napisano wcześniej - wyznacza się je na podstawie sekwencji nukleotydów.

6.2.1 Aktywność enzymatyczna

Parametr a wyliczany jest w dwóch krokach:

1. Najpierw wyznaczana jest nieujemna stała szybkości reakcji replikacji k_R , w tym celu analizowane jest pierwsze 20 nukleotydów w sekwencji agenta (koniec 5'). Właściwa wartość k_R to liczba nukleotydów zgadzających się ze wzorcem replikazy pomniejszona o 6 i następnie pomnożona przez 10.
2. Następnie stała szybkości reakcji przeliczana jest na prawdopodobieństwo a za pomocą wzoru 6.1. Równanie to opiera się na założeniu, że w każdym momencie prawdopodobieństwo zajścia reakcji jest identyczne i zależy jedynie od długości rozważanego kroku czasowego dt (patrz tab. 6.1). Powyższy fakt wskazuje, że mamy do czynienia z rozkładem Poissona, co daje już wprost wzór 6.1 [119].

$$a = 1 - e^{-k_R dt} \quad (6.1)$$

6.2.2 Prawdopodobieństwo zwinięcia - parametr l

Parametr l był wyliczany w prosty sposób - po wyznaczeniu struktury drugorzędowej RNA była to liczba sparowanych nukleotydów podzielona przez długość całej sekwencji.

6.3 Algorytm

Opisany w tym rozdziale algorytm (algorytmy 2, 3 i 4) w dużej mierze bazuje na podejściu z rozdziału 5, przy czym początkowy rozkład agentów zawsze jest losowy i jak wspomniano wcześniej - z przyczyn implementacyjnych ograniczono pewne akcje skompleksowanych cząsteczek. Wartości

parametrów wykorzystane w symulacjach (chyba, że w danym przypadku wprost podane są inne) podane są w tabeli 6.1. Należy jeszcze dodać, że poniższy pseudokod jest wykonywany równoległe przez wiele wątków, a każdy zarządza swoim fragmentem obszaru symulacji. Z opisanego wyżej powodu dodano komendę „synchronizuj”, której wykonanie blokuje dany wątek do momentu kiedy wszystkie inne wątki jej nie wykonają - wtedy możliwe jest dalsze działanie programu.

Algorytm 2 Główna pętla programu

```

for step ← 1 TO max_steps do
  Reakcje pierwszorzędowe (algorytm 3)
  synchronizuj
  Symuluj interakcje (algorytm 4)
  synchronizuj
end for

```

Algorytm 3 Reakcje pierwszorzędowe

```

Stwórz listę list wszystkich agentów zarządzanych przez dany wątek w kolejności losowej
while list nie jest pusta do
  Weź pierwszego agenta A z listy list
  if A jest kompleksem then
    Zmniejsz czas do replikacji A
    if Czas do replikacji A wynosi 0 then
      Przeprowadź replikację oraz mutację nowopowstałego agenta
      Zdysocjuj A
    end if
  else
    Zmniejsz czas życia A
    if Czas życia A wynosi 0 then
      Usuń A z symulacji
    else
      Podдай A dyfuzji
    end if
  end if
end while

```

6.4 Implementacja

Symulator wykorzystany w opisywanych tutaj symulacjach został zaimplementowany od zera w języku C++ z wykorzystaniem interfejsu OpenMPI [56]. W celu pomiaru wydajności programu - wykonano testy na dwóch platformach:

1. Power AI firmy IBM
2. serwer Auriga - PUT z procesorem Intel Xeon E5-4640

Oprócz czasu wykonania (który w praktyce silnie zależy od szybkości użytego procesora) zmierzono również efektywność zrównoleglenia, czyli przyspieszenie (por. poprozdział 2.4) podzielone przez liczbę wątków. W teorii wartość ta nie powinna przekraczać 100%, w praktyce jednak może się to zdarzyć choćby ze względu na optymalizacje równoległego dostępu do pamięci we współczesnych procesorach. Należy tutaj wspomnieć, że platforma PowerAI miała wyraźną przewagę nad procesorem Intel, szczególnie dla liczby wątków większej niż 7 (rys. 6.1 oraz 6.2).

Algorytm 4 Interakcje agentów

Stwórz listę *list* wszystkich agentów zarządzanych przez dany wątek w kolejności losowej

while *list* nie jest pusta **do**

 Weź pierwszego agenta *A* z listy *list*

if *A* nie jest kompleksem **then**

 Stwórz listę *list2* sąsiadów *A* w kolejności losowej

if Liczba elementów w *list2* jest większa niż parametr *neigh* **then**

 Usuń *A* z symulacji

else

while *list2* nie jest pusta **do**

 Weź pierwszego agenta *n* z listy *list2*

if *n* nie jest kompleksem **then**

$reaction_probability = A.a * A.l * (1 - n.l)$

 Weź losową liczbę *p* z przedziału [0; 1) (rozkład równomierny)

if $p < reaction_probability$ **then**

 Stwórz kompleks agentów *A* oraz *n*

end if

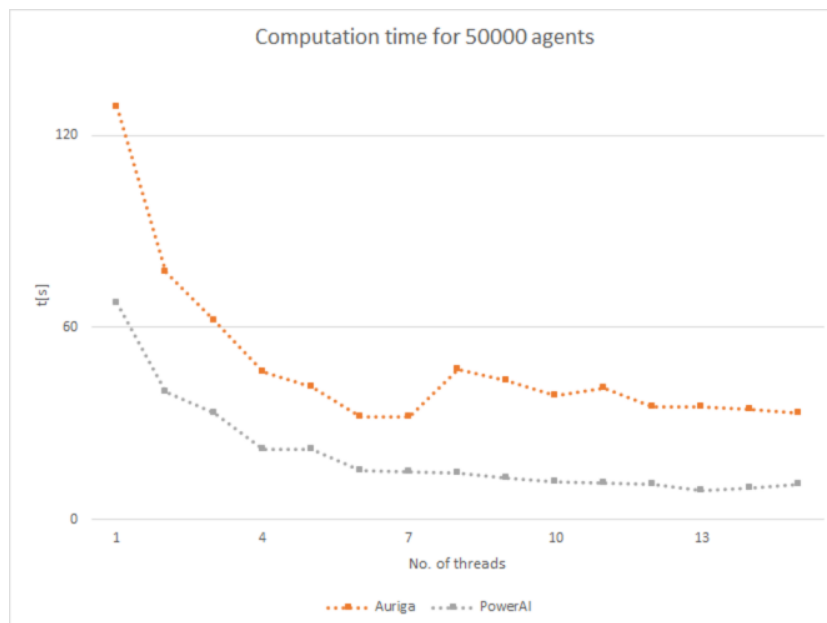
end if

end while

end if

end if

end while



Rys. 6.1: Czas wykonania symulacji dla 50000 agentów w zależności od liczby wątków.

Źródło: [105]

Nazwa	Wartość domyślna	Opis
<i>int_radius</i>	3	promień agenta
<i>sizeX, sizeY</i>	1000	wymiary obszaru symulacji
<i>seq_length</i>	50	długość sekwencji (każda sekwencja zawiera dokładnie tyle nukleotydów)
<i>d</i>	0,01	podstawowa stała rozpadu
<i>seq_mut</i>	0,01	prawdopodobieństwo mutacji nukleotydu
<i>dt</i>	0,1	długość kroku czasowego
<i>D</i>	4	stała dyfuzji
<i>K</i>	∞	stała replikacji
<i>selProb</i>	0,01	Kiedy symulacja się kończy, losowo wybierane są sekwencje, które znajdują się w pliku wynikowym. Parametr <i>selProb</i> określa prawdopodobieństwo wyboru każdego agenta.
<i>initR</i>	zależy od scenariusza	początkowa liczba replikaz (agentów z predefiniowaną sekwencją replikującą)
<i>initP</i>	zależy od scenariusza	początkowa liczba pasożytów (agentów z losową sekwencją, część z nich może w wyniku przypadku być funkcjonalnymi replikazami)
<i>neigh</i>	4	maksymalna liczba sąsiadów agenta
<i>d(NN)</i>	4	Stała szybkości hydrolizy wiązania fosfodiesterowego między dwoma rybonukleotydami. <i>N</i> może oznaczać dowolny nukleotyd: <i>A, C, G</i> lub <i>U</i> .

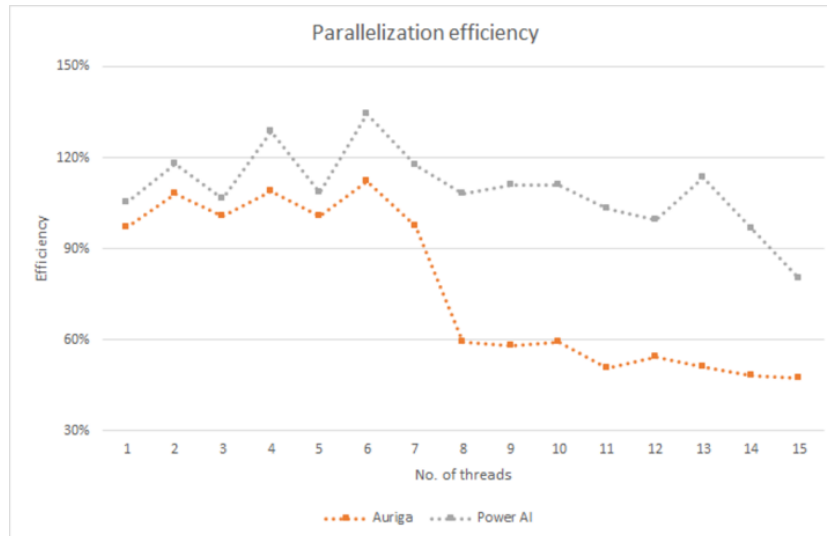
Tab. 6.1: Domyślne parametry w symulacji Świata RNA z sekwencjami.

Źródło: [105]

6.5 Wyniki

Z wykorzystaniem opisanego powyżej modelu przeprowadzono szereg symulacji dla różnych przypadków. Sprawdzono jak ewoluuje populacja złożona początkowo jedynie z idealnych replikaz, gdzie pasożyty powstają samoczynnie na skutek mutacji. Porównano to do zbioru przypadkowych sekwencji, w którym część cząsteczek może katalizować replikację na skutek przypadku. Na końcu wykorzystano także badania Kierzka i jego zespołu o których wspomniano na początku rozdziału i przesymulowano jak różna trwałość wiązań fosfodiesterowych między parami poszczególnych rybonukleotydów wpływa na ewolucję Świata RNA oraz preferowane sekwencje. Wszystkie symulacje były prowadzone do momentu wymarcia systemu (kiedy liczba cząsteczek spadała do 0) lub osiągnięcia stanu równowagi - kiedy liczebność populacji oraz inne śledzone zmienne nie zmieniały się z czasem.

Należy wspomnieć jeszcze, że domyślne wartości parametrów podane w tabeli 6.1 zostały wyznaczone eksperymentalnie. Głównym celem była możliwość symulacji dużej populacji w realistycznym czasie, np. szybkość rozpadu oraz mutacji musiała być duża, żeby można było obserwować kierunek ewolucji, która inaczej mogłaby przebiegać zbyt wolno i nie dałoby się jej stosownie opisać w symulowanym przedziale czasowym. Wartości stałych szybkości hydrolizy poszczególnych wiązań fosfodiesterowych (jak wspomniano wcześniej) pochodzą z danych eksperymentalnych [13, 14, 60, 61].



Rys. 6.2: Efektywność zrównoleglenia symulacji dla 50000 agentów w zależności od liczby wątków.
Źródło: [105]

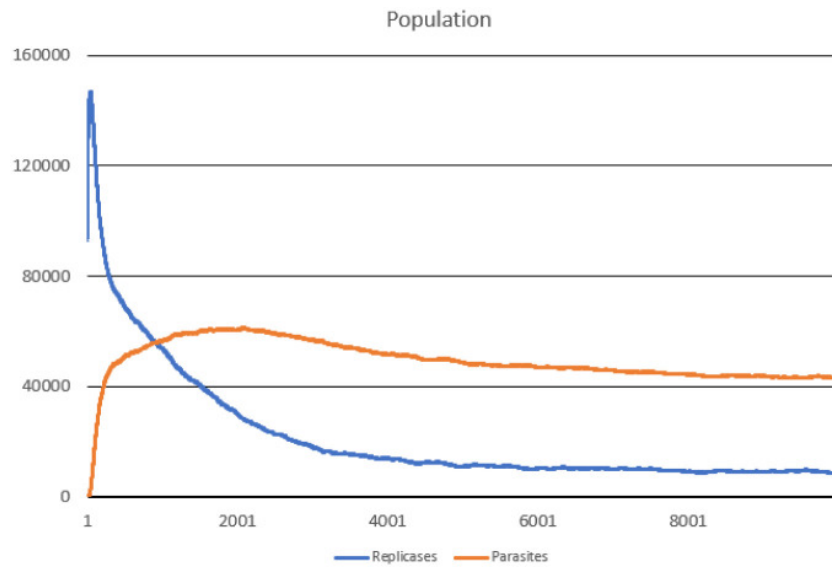
6.5.1 Scenariusz 1

W celu zbadania czy pasożyty powstające z replikaz na skutek mutacji mogą zagrozić przetrwaniu, symulację rozpoczęto rozmieszczając losowo 100 tysięcy idealnych replikaz (ich pierwsze 20 nukleotydów dokładnie zgadzało się ze wzorcem). Obserwowano przede wszystkim ewolucję w czasie liczebności obu rodzajów RNA (replikaz i pasożytów), średniej aktywności replikaz a oraz skłonności do zwijania się l (oddzielnie dla replikaz i pasożytów). Na wykresach 6.3, 6.4 oraz 6.5 pokazano pierwsze 10 tysięcy kroków symulacji.

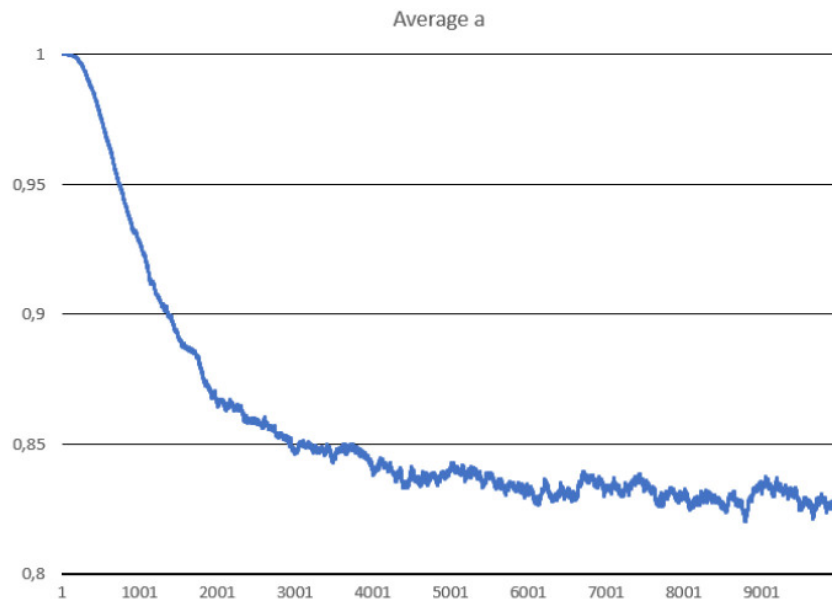
Przede wszystkim można dostrzec, że system jest w stanie przetrwać i stabilizuje się po ok. 8 tys. kroków. W początkowej fazie następuje bardzo szybki wzrost liczebności replikaz, ponieważ nic go nie ogranicza poza dostępną przestrzenią (zasoby nie są symulowane *explicite*). Pasożyty, które pojawiają się w międzyczasie na skutek mutacji, także zaczynają szybko zwiększać swoją liczbę i w pewnym momencie odwracają wspomniany wcześniej trend - populacja replikaz zaczyna się kurczyć. Po niedługim czasie również liczebność pasożytów zaczyna spadać (choć nie tak gwałtownie jak w przypadku replikaz), gdyż są one uzależnione od replikaz - nie potrafią one replikować się samodzielnie. Ostatecznie spadek liczby obu typów RNA mocno zwalnia, co prowadzi do ustabilizowania się całości. Ciekawą obserwacją jest, że replikazy przez całą symulację mają średnio większą wartość parametru l niż pasożyty, co teoretycznie czyni je gorszymi kandydatami do replikacji, nie wymierają one jednak, ponieważ stężenie pasożytów jest przez system utrzymywane na odpowiednim poziomie. Jeśli chodzi o średnią wartość a to nieco spada ona na skutek przypadkowych mutacji, co oznacza że populacja nie jest w stanie utrzymać pierwszych 20 nukleotydów replikaz w stanie niezmiennym.

6.5.2 Scenariusz 2

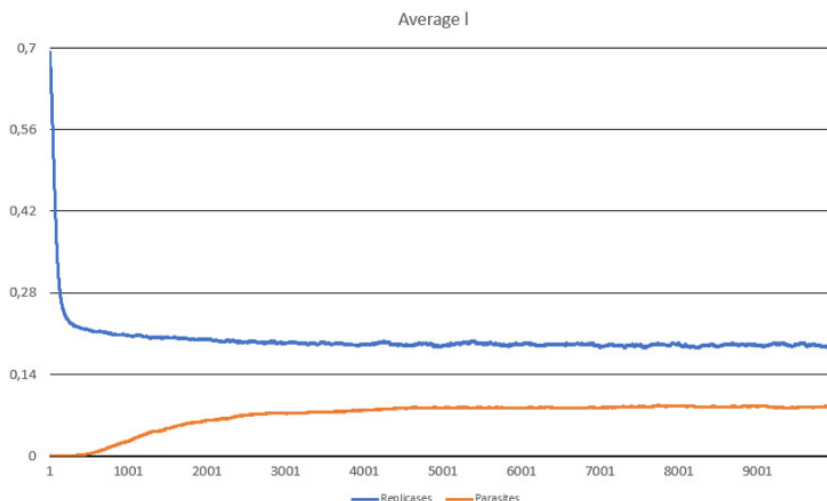
Ten scenariusz jest niejako przeciwieństwem poprzedniego - w tym przypadku początkowa populacja składa się ze 100 tys. całkowicie losowych sekwencji, z których część ma zupełnie przypadkowo zdolność do katalizowania replikacji. Celem było stwierdzenie czy ze zbioru losowych sekwencji może powstać populacja zdolna do przetrwania i ewolucji. Sytuacja symulowana w tym podrozdziale odpowiada samemu początkowi Świata RNA - kiedy żadne funkcjonalne sekwencje nie zostały jeszcze wyodrębnione. Wyniki uzyskane z symulacji 10 tys. kroków przedstawiono na



Rys. 6.3: Zmiana liczebności populacji replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 1.
 Źródło: [105]



Rys. 6.4: Zmiana średniej aktywności replikaz przez pierwsze 10 tys. kroków w scenariuszu 1.
 Źródło: [105]



Rys. 6.5: Zmiana średniej wartości parametru l replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 1.

Źródło: [105]

wykresach 6.6, 6.7 oraz 6.8.

Rezultaty jasno pokazują, że replikazy są zdolne do przetrwania oraz zwiększenia swoich zdolności katalitycznych w toku ewolucji, co pozwala całej populacji osiągnąć stabilność. W początkowej fazie następowało masowe wymieranie cząsteczek, które nie miały dostępu do wystarczająco aktywnej replikazy. W wyniku tego bardzo szybko wzrastała też średnia wartość parametru a - grupy replikaz o większej aktywności enzymatycznej miały znacznie większą szansę na przetrwanie. Po pewnym czasie trend się zmieniał i populacja replikaz osiągała wystarczającą średnią wartość a , żeby zacząć rosnąć. Pasożyty, które są kopiowane przez replikazy (jak również powstają z nich na skutek mutacji) również zaczęły zwiększać swoją liczebność. Po pewnym czasie wzrost ustął, a system ustabilizował się. Parametr l dla replikaz (podobnie jak poprzednio) miał na końcu średnio większą wartość niż dla pasożytów, co również i tutaj utrudniało replikazom powielanie się, jednak mimo to nie doszło do ich wymarcia.

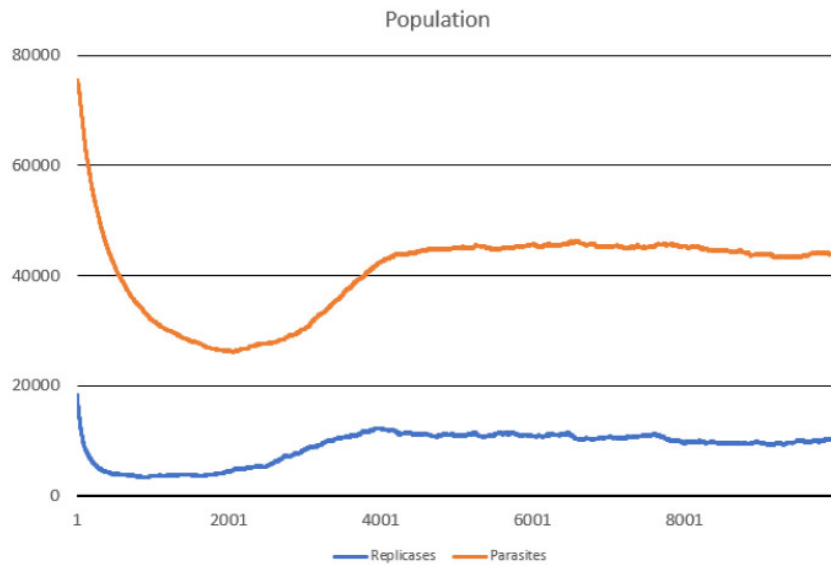
6.5.3 Scenariusz 3

Jak wspomniano wcześniej, w trzecim scenariuszu wzięto pod uwagę podatności poszczególnych rodzajów wiązań fosfodiesterowych na hydrolizę (parametry $d(NN)$). Wartości potrzebnych stałych podano w tabelce 6.2. Dla każdej sekwencji stała rozpadu była teraz wyliczana jako suma podstawowej stałej d oraz stałych hydrolizy wszystkich kolejnych wiązań fosfodiesterowych. Oznaczało to, że niektóre sekwencje były związane z mniejszą trwałością cząsteczki. Symulację zainicjalizowano identycznie jak w scenariuszu 1 - losowo rozmieszczono 100 tys. idealnych replikaz, wyniki pokazane są na wykresach 6.9, 6.10 oraz 6.11.

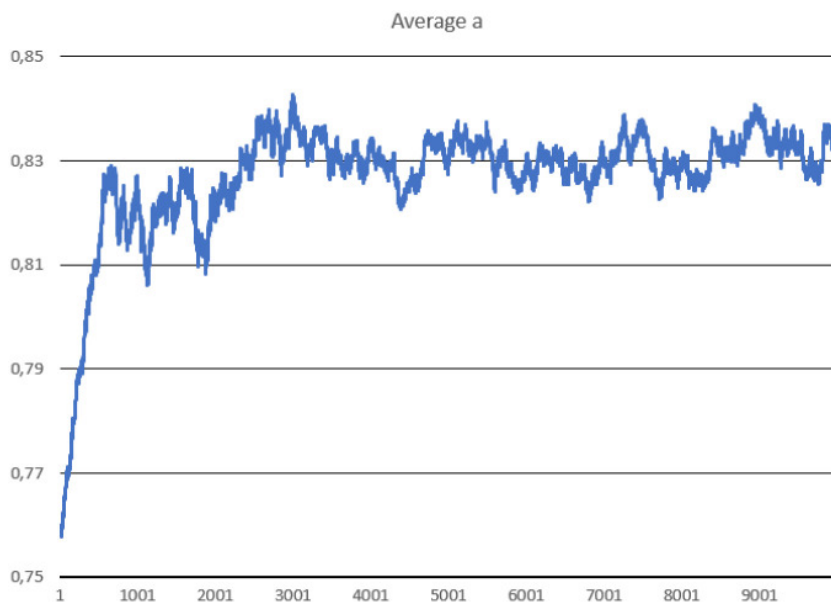
Mimo częstszego rozpadu cząsteczek, populacja była w stanie przetrwać. W wyniku selekcji, uracyl został w dużej mierze wyeliminowany z sekwencji występujących w populacji, występował znacznie rzadziej od innych rybonukleotydów. Poza tym szczegółem, rezultaty były zbieżne ze scenariuszem 1.

6.5.4 Scenariusz 4

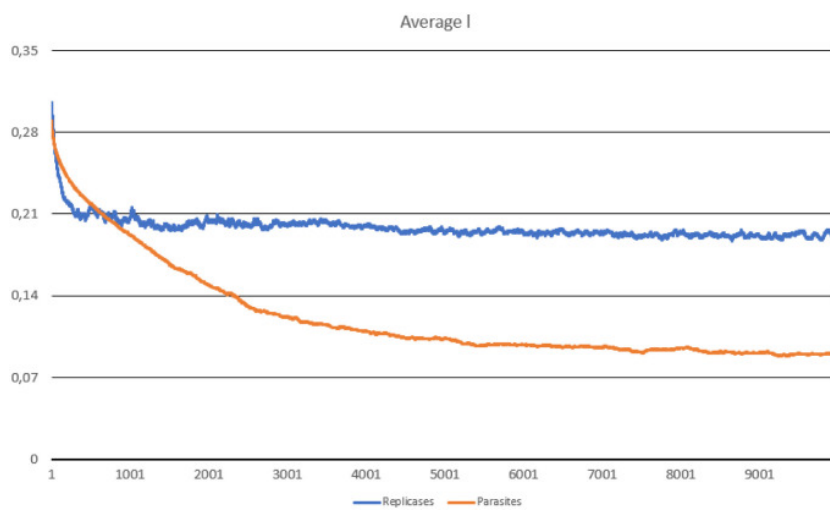
W celu uzupełnienia informacji uzyskanych w poprzednich symulacjach - sprawdzono jak zmiana wartości poszczególnych parametrów wpływa na przeżywalność populacji łańcuchów RNA. Należy



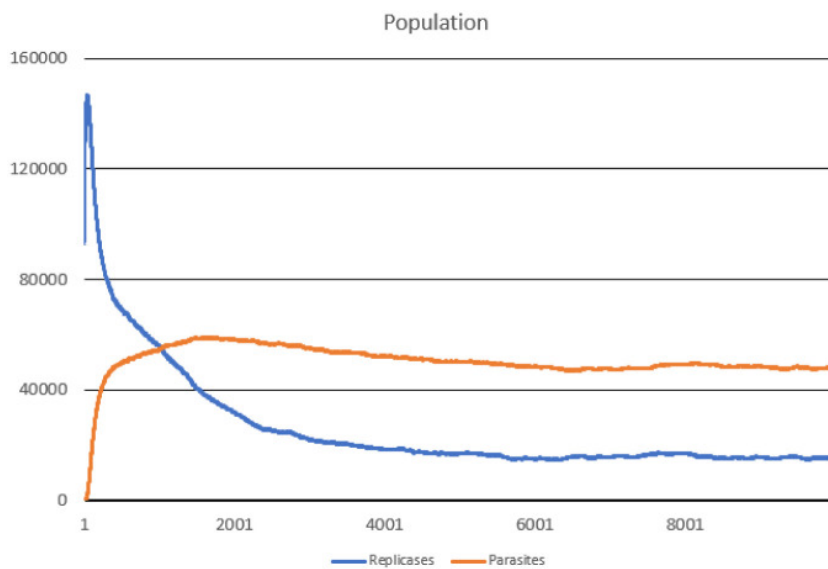
Rys. 6.6: Zmiana liczebności populacji replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 2.
Źródło: [105]



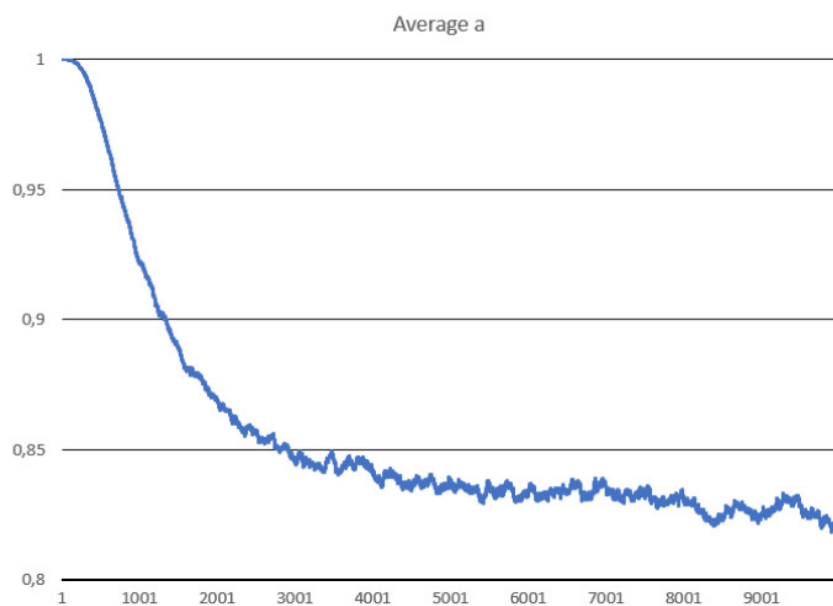
Rys. 6.7: Zmiana średniej aktywności replikaz przez pierwsze 10 tys. kroków w scenariuszu 2.
Źródło: [105]



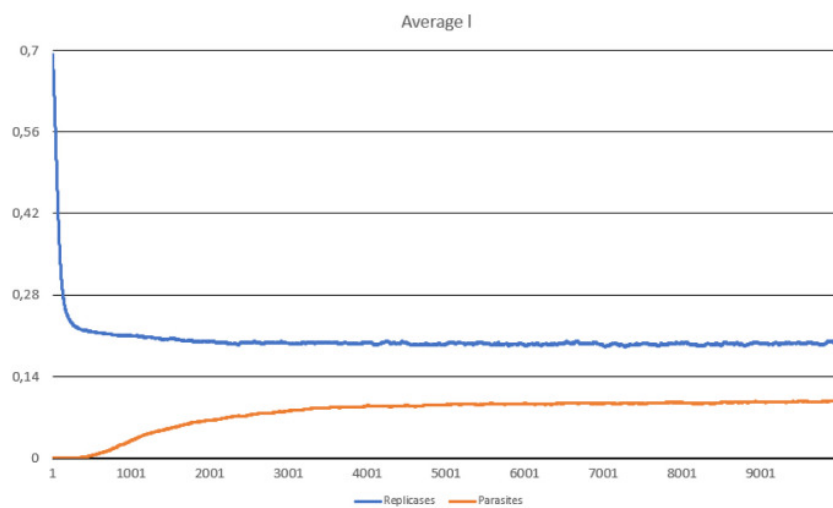
Rys. 6.8: Zmiana średniej wartości parametru l replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 2.
Źródło: [105]



Rys. 6.9: Zmiana liczebności populacji replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 3.
Źródło: [105]



Rys. 6.10: Zmiana średniej aktywności replikaz przez pierwsze 10 tys. kroków w scenariuszu 3.
Źródło: [105]



Rys. 6.11: Zmiana średniej wartości parametru l replikaz (kolor niebieski) i pasożytów (kolor pomarańczowy) przez pierwsze 10 tys. kroków w scenariuszu 3.
Źródło: [105]

Phosphodiester bonds	Hydrolysis rate of the phosphodiester bonds $d(NN)$
UA	0,000953
CA	0,000932
UC	0,000846
CC	0,000846
UG	0,000100
CG	0,000100
UU	0,000100
CU	0,000100

Tab. 6.2: Stałe szybkości hydrolizy $d(NN)$ wiązań fosfodiesterowych między poszczególnymi parami nukleotydów. Dla tych nieuwzględnionych w powyższej tabeli - stała szybkości wynosi 0.

Źródło: [105]

wspomnieć, że symulowana populacja była w tym przypadku mniejsza niż we wcześniejszych scenariuszach, mimo to nadal symulowano 10 tys. kroków. W symulacjach skupiono się na dwóch parametrach (uzyskane wyniki w stanach równowagi przedstawiono w tabelach 6.3 i 6.4):

1. podstawowa stała rozpadu (d) - wartość ta determinuje czas życia pojedynczej cząsteczki RNA, wpływa to m.in. na to jak szybko jedno pokolenie cząsteczek musi zostać zastąpione przez drugie. Determinuje to np. szybkość ewolucji.
2. stała dyfuzji (D) - parametr ten określa jak duży wpływ ma aspekt przestrzenny, przy wysokich wartościach system można traktować jako dobrze wymieszany, przy małych - coraz większe znaczenie zaczyna mieć rozmieszczenie poszczególnych grup w obszarze symulacji.

Podstawowa stała rozpadu d	Wielkość populacji	Średnie a	Średnie l (replikazy)	Średnie l (pasożyty)
0,001	390	0,83	0,19	0,09
0,01	680	0,82	0,17	0,12
0,6	167	0,86	0,26	0,13
0,7	0	n/a	n/a	n/a

Tab. 6.3: Wpływ różnych wartości d na parametry systemu w stanie równowagi w scenariuszu 4.

Źródło: [105]

Stała dyfuzji D	Wielkość populacji	Średnie a	Średnie l (replikazy)	Średnie l (pasożyty)
0	0	n/a	n/a	n/a
4	680	0,82	0,17	0,12
40	351	0,65	0,15	0,08
50	651	0,77	0,16	0,08
100	616	0,79	0,17	0,07
400	619	0,81	0,18	0,08

Tab. 6.4: Wpływ różnych wartości D na parametry systemu w stanie równowagi w scenariuszu 4.

Źródło: [105]

Z przeprowadzonych symulacji wynika, że dla małych wartości parametru d ewolucja (jak przewidziano) przebiega wolniej, populacja jest wtedy również mniejsza, co wydaje się mniej oczywistym faktem. Dla dużych wartości podstawowej stałej rozpadu cząsteczki RNA są tak niestabilne, że albo szybko wymierają albo egzystują na krawędzi wymarcia. Jeśli chodzi o dyfuzję to szybkie przemieszczanie się cząsteczek powoduje, że populacja w stanie równowagi jest mniejsza, z kolei niskie wartości D prowadzą to większej wartości średniej a i jest to jedyna zauważalna różnica. Zbyt małe wartości D mogą jednak poważnie ograniczać możliwości ekspansji populacji, przez co będzie ona skazana na wymarcie.

6.6 Dyskusja

Opisany w tym rozdziale model jest bezpośrednią kontynuacją symulacji wieloagentowej, której poświęcony został rozdział 5 i która została opisana w [110]. Do wprowadzenia niektórych zmian konieczne były dane eksperymentalne, takie jak stałe hydrolizy otrzymane przez Kierzka i zespół [13, 14, 60, 61] oraz dane dotyczące rybozomu tC19Z [125], który posłużył jako wzorzec dla replikaz.

Podobne badania prowadzili Takeuchi i Hogeweg, wykorzystywali oni strukturę drugorzędową RNA, żeby na jej podstawie odróżniać replikazy od pasożytów. Podstawową różnicą jednak było wykorzystanie automatów komórkowych, a także inny cel badań - obserwacja różnicowania się poszczególnych gatunków w populacji. W przypadku replikaz zostało to umożliwione detekcją dwóch struktur typu „stem” połączonych jedną strukturą typu „multiloop”, która w symulacji dawała cząsteczce zdolności katalityczne. Powodowało to, że katalizować replikację mogły RNA o bardzo różnych sekwencjach [113]. Dla porównania w opisanym w tym rozdziale modelu o zdolnościach katalitycznych decydowała bardzo konkretna sekwencja.

Można zauważyć, że wprowadzono tutaj (i również w rozdziale 5) bardzo istotne uproszczenie - w wyniku replikacji RNA powinna powstawać nić komplementarna i dopiero ponowna replikacja powinna dać kopię tej wyjściowej. Wyniki jednak są zgodne z tym co otrzymano np. w [21], badając różnicowanie RNA w protokomórkach na rybozomy oraz nośniki informacji genetycznej. We wspomnianej publikacji uwzględniono już podział na nici „+” i „-” czyli te pełniące konkretną funkcję oraz RNA do nich komplementarne.

Warto na koniec jeszcze wspomnieć, że opisywany w tym rozdziale model można zaklasyfikować jako przede wszystkim „selekcję wielopoziomową typu 1” (MSL1) [36].

Model wykorzystujący równania różniczkowe cząstkowe

Podstawowym celem tego rozdziału jest wprowadzenie modelu opartego o inną filozofię niż systemy wieloagentowe. Próba opisanie populacji cząsteczek RNA formalnie, za pomocą równań różniczkowych, pozwala na analityczne wyprowadzenie ogólnych zasad rządzących ewolucją systemu. Przedstawiony tutaj model i wyniki zostały przez autora opublikowane wcześniej w [106].

7.1 Model biologiczny

Założenia związane z biologicznymi właściwościami badanego systemu są podobne do tych opisanych w poprzednim rozdziale. Przede wszystkim badane interakcje zachodzą w roztworze wodnym, a stężenia reagentów są na tyle niskie, żeby dyfuzja przebiegała swobodnie. Łańcuchy RNA dzielą się na replikazy, które są w stanie kopiować inne cząsteczki oraz pasożyty, które mogą jedynie służyć jako wzorzec. Każde RNA ma przypisany parametr a , który określa łatwość z jaką jest kopiowane przez replikazę. W tym przypadku wartość ta jest stałą szybkości reakcji chemicznej, a nie prawdopodobieństwem. Replikacja ze względu na brak odpowiednich enzymów jest bardzo niestabilnym procesem i może prowadzić do powstania uszkodzonej kopii, w ten sposób podczas powielania replikazy może powstać pasożyt. Poza wyżej opisanymi założeniami dokonano także kilku zmian i uproszczeń:

1. w roztworze obecne są zasoby, w tym przypadku głównie rybonukleotydy, których obecność znacząco wpływa na szybkość replikacji, powstają one samoczynnie ze stałą szybkością,
2. wszystkie reakcje przebiegają zgodnie z prawem działania mas,
3. replikacja zachodzi bez etapu pośredniego (kompleksowania)

7.2 Model matematyczny

Liczebność poszczególnych rodzajów molekuł reprezentowana jest przez ich stężenia molowe w roztworze, zależą one od współrzędnych (x, y, z) oraz czasu (t) . Odpowiednie funkcje oznaczone są następująco:

1. r - stężenie replikaz,
2. p - stężenie pasożytów,
3. n - stężenie zasobów,

Formalnie zależność od współrzędnych i czasu można zapisać:

$$r = r(x, y, z, t) \quad (7.1)$$

$$p = p(x, y, z, t) \quad (7.2)$$

$$n = n(x, y, z, t) \quad (7.3)$$

Teraz należałoby omówić po kolei zachodzące w roztworze reakcje.

7.2.1 Replikacja

Zgodnie z przedstawionym wcześniej modelem biologicznym, w replikacji biorą udział dwa łańcuchy RNA (replikaza i wzorzec), a także zasoby które dla uproszczenia są tutaj reprezentowane jako jedna cząsteczka. Jest to zatem reakcja trzeciorzędowa, czyli zgodnie z prawem działania mas, szybkość zachodzenia reakcji określa wzór:

$$v = k[A][B][C] \quad (7.4)$$

gdzie k jest stałą, a $[A]$, $[B]$ i $[C]$ to stężenia odpowiednich substratów.

Przy każdej sukcesywnej replikacji powstaje jedna, nowa kopia wzorca. Oznacza to, że szybkość reakcji może być interpretowana jako ilość nowo wytworzonych cząsteczek w jednostce czasu. Za stałą k podstawiany jest wspomniany wcześniej parametr a wzorca, a substratami są: wzorzec, replikaza oraz zasoby.

7.2.2 Rozpad

Każda bardziej skomplikowana cząsteczka biologiczna podlega procesowi spontanicznego rozpadu, łańcuchy RNA w omawianym modelu nie są wyjątkiem. Warto podkreślić, że omawiany tutaj proces przebiega nieenzymatycznie i opiera się jedynie na naturalnej, ograniczonej trwałości wiązań chemicznych.

W jednostce czasu każda cząsteczka ma określane prawdopodobieństwo rozpadu, dla uproszczenia wartość ta jest identyczna dla wszystkich RNA danego typu (w rozdziale 5 część scenariuszy uzależniała tę wartość od sekwencji). Szybkość reakcji rozpadu można opisać następującym wzorem:

$$v = k[A] \quad (7.5)$$

gdzie ponownie k jest stałą rozpadu, a $[A]$ stężeniem badanej cząsteczki. Każda reakcja rozpadu kończy się zniszczeniem wyjściowego RNA, zatem szybkość reakcji jednoznacznie przekłada się na liczbę ubywających cząsteczek w założonej jednostce czasu.

7.2.3 Powstawanie zasobów

Zakłada się, że zasoby są tworzone na skutek działania promieniowania słonecznego i/lub innych stabilnych reakcji chemicznych ze stałą szybkością - n_0 .

7.2.4 Dyfuzja

Ze względu na założenie o niezbyt wysokich stężeniach badanych cząsteczek, dyfuzję w pełni może opisać drugie prawo Ficka:

$$\frac{d[A]}{dt} = D\vec{\nabla}^2[A] \quad (7.6)$$

D jest oczywiście stałą dyfuzji, $[A]$ stężeniem molowym danej cząsteczki, a $\vec{\nabla}^2$ to operator Laplace'a, który uzależnia proces od rozkładu $[A]$ w przestrzeni.

7.2.5 Równania różniczkowe cząstkowe

Napisanie ostatecznych równań wymaga zdefiniowania kilku ważnych (dodatnich) stałych:

1. D_n - stała dyfuzji dla zasobów, ze względu na swój niewielki rozmiar, cząsteczki dyfundują szybciej niż RNA,
2. D - stała dyfuzji dla RNA,
3. m - prawdopodobieństwo, że kopiowanie replikazy doprowadzi do powstania pasożyta (ze względu na dość częste błędy replikacji),
4. a_R - powinowactwo replikazy do replikazy (jak łatwa do skopiowania jest replikaza),
5. a_P - powinowactwo pasożyta do replikazy (jak łatwy do skopiowania jest pasożyt),
6. d_R - stała rozpadu replikaz,
7. d_P - stała rozpadu pasożytów.

Wykorzystując powyższe definicje oraz opisy przebiegu poszczególnych reakcji, można rozpisać odpowiednie równania różniczkowe:

$$\frac{\partial n}{\partial t} = n_0 - nr(a_{Rr} + a_{Pp}) + D_n \nabla^2 n \quad (7.7)$$

$$\frac{\partial r}{\partial t} = (1 - m)a_{Rr}nr^2 + D \nabla^2 r - d_R r \quad (7.8)$$

$$\frac{\partial p}{\partial t} = ma_{Rr}nr^2 + a_{Pn}nrp + D \nabla^2 p - d_P p \quad (7.9)$$

Wartościami brzegowymi są funkcje reprezentujące stężenia n , p oraz r w ustalonym czasie t_0 w każdym punkcie przestrzeni.

7.2.6 Rozwiązania omawianych równań różniczkowych

7.2.6.1 Dobrze wymieszany roztwór

Pierwszym krokiem jest zbadanie czy istnieje rozwiązanie dla stężeń będących stałymi, co można rozumieć jako dobrze wymieszany roztwór, którego parametry nie zmieniają się z czasem. Prowadzi to do równań:

$$\frac{\partial n}{\partial t} = 0 \quad (7.10)$$

$$\frac{\partial r}{\partial t} = 0 \quad (7.11)$$

$$\frac{\partial p}{\partial t} = 0 \quad (7.12)$$

$$\nabla^2 n = 0 \quad (7.13)$$

$$\nabla^2 r = 0 \quad (7.14)$$

$$\nabla^2 p = 0 \quad (7.15)$$

Podstawienie powyższych zależności do wyjściowych równań różniczkowych pozwala otrzymać następujące wzory:

$$0 = n_0 - nr(a_{Rr} + a_{Pp}) \quad (7.16)$$

$$0 = (1 - m)a_{Rn}nr^2 - d_Rr \quad (7.17)$$

$$0 = ma_{Rn}nr^2 + a_{Pn}nrp - d_Pp \quad (7.18)$$

Równanie 7.17 może być zapisane jako:

$$0 = r[(1 - m)a_{Rn}nr - d_R] \quad (7.19)$$

Jedną z możliwości jest $r = 0$, ale w takim przypadku nie może istnieć także populacja pasożytów (brak replikacji), co oznaczałoby ciągły przyrost zasobów, które nie byłyby zużywane w żaden sposób (funkcja n nie mogłaby być stała). Alternatywą jest $r > 0$, co prowadzi do:

$$d_R = (1 - m)a_{Rn}nr \quad (7.20)$$

W celu wyznaczenia r z powyższej równości, konieczne są dwa założenia: $m < 1$ oraz $n > 0$. Oba muszą być spełnione, ponieważ zarówno $m = 1$ jak i $n = 0$ prowadziłyby do wymarcia replikaz (przypadek $r = 0$ został omówiony wyżej). Zatem:

$$r = \frac{d_R}{(1 - m)a_{Rn}} \quad (7.21)$$

Zależność wyznaczającą p można otrzymać z równania 7.18:

$$p = \frac{ma_{Rn}nr^2}{d_P - a_{Pn}nr} \quad (7.22)$$

Mianownik jest różny od zera, ponieważ w przeciwnym przypadku równanie 7.18 byłoby niemożliwe do spełnienia (ze względu na fakt, że $m, a_{Rn}, n, r > 0$). Za pomocą podstawienia (równanie 7.21) można usunąć r z mianownika:

$$p = \frac{ma_{Rn}nr^2}{d_P - \frac{a_{Pn}nd_R}{(1-m)a_{Rn}}} \quad (7.23)$$

Podobnie należy usunąć jedno r z licznika:

$$p = \frac{ma_{Rn}nr d_R}{(d_P - \frac{a_{Pn}nd_R}{(1-m)a_{Rn}})(1 - m)a_{Rn}} \quad (7.24)$$

Powyższy wzór można uprościć:

$$p = \frac{ma_{Rn}d_R}{d_P(1 - m)a_{Rn} - a_{Pn}d_R}r \quad (7.25)$$

Z faktu $m > 0$ wynika, że licznik jest dodatni, zatem aby wartość p była poprawna, mianownik także musi być większy od zera:

$$d_P(1 - m)a_{Rn} - a_{Pn}d_R > 0 \quad (7.26)$$

Powyższą nierówność można także zapisać jako:

$$1 - m > \frac{a_{Pn}d_R}{a_{Rn}d_P} \quad (7.27)$$

Jest to warunek konieczny istnienia poprawnego rozwiązania. Ostatnie co należy udowodnić to istnienie wartości n , która pozwoli na spełnienie dotychczas nieanalizowanego równania 7.16. W

takim przypadku dałoby to możliwość otrzymania odpowiednich wartości r oraz p i jednoznacznie udowodniłoby, że nierówność 7.27 jest też warunkiem dostatecznym. Wprowadźmy oznaczenie:

$$A = \frac{ma_R d_R}{d_P(1-m)a_R - a_P d_R} \quad (7.28)$$

Z nierówności 7.26, wynika że $A > 0$. W skrócie, równanie 7.25 można zapisać jako:

$$p = Ar \quad (7.29)$$

Teraz można wrócić do równania 7.16. Wykorzystując wzór $p = Ar$:

$$0 = n_0 - nr(a_R r + a_P Ar) \quad (7.30)$$

Co można też zapisać jako:

$$0 = n_0 - nr^2(a_R + a_P A) \quad (7.31)$$

Wykorzystując równanie 7.21, żeby wyeliminować r :

$$0 = n_0 - n \frac{d_R^2}{[(1-m)a_R n]^2} (a_R + a_P A) \quad (7.32)$$

Rozwiązując powyższe w celu otrzymania n :

$$n = \frac{d_R^2}{[(1-m)a_R]^2 n_0} (a_R + a_P A) \quad (7.33)$$

Nietrudno zauważyć, że tak otrzymana wartość n będzie zawsze dodatnia, zatem jeśli nierówność 7.27 jest spełniona (co wymaga, żeby $m < 1$), zawsze istnieje poprawny stan równowagi. Można wspomnieć jeszcze, że interesujący przypadek zachodzi, kiedy m wbrew założeniom jest równe 0, pasożyty i replikazy mogą koegzystować w dowolnych proporcjach, ale tylko jeśli $a_R d_P = a_P d_R$.

Liniowa stabilność

Dla otrzymanego wcześniej rozwiązania należy sprawdzić czy układ będzie do niego dążył. Wciąż obowiązuje założenie o dobrze wymieszanym roztworze, jednak dopuszczona jest ewolucja zmiennych w czasie. Jedną z podstawowych metod dowodzenia stabilności jest linearyzacja układu w otoczeniu punktu równowagi, jednak to podejście napotyka na poważne przeszkody, kiedy zastosuje się je do stężenia replikaz:

$$\frac{\partial \frac{\partial r}{\partial t}}{\partial r} = 2(1-m)a_R n r - d_R = [(1-m)a_R n r - d_R] + (1-m)a_R n r \quad (7.34)$$

Jeśli stężenie replikaz jest dodatnie, wtedy można zastosować równanie 7.20, które musi być spełnione w stanie równowagi.

$$\frac{\partial \frac{\partial r}{\partial t}}{\partial r} = (1-m)a_R n r \quad (7.35)$$

Pomijając przypadki kiedy $m = 1$ lub $a_R = 0$ (replikazy nie mogą replikować się w tych warunkach), prawa strona równania jest dodatnia. Wynika z tego, że jakkolwiek zmiana stężenia replikaz w punkcie równowagi, podlega dodatniemu sprzężeniu zwrotnemu. Oznacza to, że w celu udowodnienia stabilności układu w punkcie równowagi - należy zastosować bardziej zaawansowane metody.

Model oscylatora

Jak napisano powyżej, udowodnienie stabilności wymaga innego podejścia niż linearyzacja wokół punktu równowagi. Założenia są tożsame z tymi podanymi w poprzednim podrozdziale, wszystkie stężenia są identyczne w każdym punkcie roztworu, ale mogą zmieniać się z czasem:

$$\nabla^2 n = 0 \quad (7.36)$$

$$\nabla^2 r = 0 \quad (7.37)$$

$$\nabla^2 p = 0 \quad (7.38)$$

Ewolucja systemu jest określona w sposób deterministyczny (i nie zależy od współrzędnych), zatem wartości wszystkich zmiennych pozostaną identyczne w całej objętości roztworu.

Równania 7.7 oraz 7.8 przyjmują postać:

$$\frac{\partial n}{\partial t} = n_0 - nr(a_{Rr} + a_{Pp}) \quad (7.39)$$

$$\frac{\partial r}{\partial t} = (1 - m)a_{Rn}r^2 - d_R r \quad (7.40)$$

Dzieląc oba równania (obustronnie) przez r^2 i wprowadzając oznaczenie $y = \frac{1}{r}$:

$$\frac{\partial n}{\partial t} y^2 = n_0 y^2 - n a_R - n a_{Pp} y \quad (7.41)$$

$$- \frac{\partial y}{\partial t} = (1 - m)a_{Rn} - d_R y \quad (7.42)$$

Równanie 7.42 można przekształcić w celu otrzymania n :

$$n = \frac{d_R y - \frac{\partial y}{\partial t}}{(1 - m)a_R} \quad (7.43)$$

Biorąc pochodną po czasie:

$$\frac{\partial n}{\partial t} = \frac{d_R \frac{\partial y}{\partial t} - \frac{\partial^2 y}{\partial t^2}}{(1 - m)a_R} \quad (7.44)$$

Równania 7.43 i 7.44 można wykorzystać w celu wyeliminowania ze wzoru 7.41 funkcji n i jej pochodnej:

$$\frac{d_R \frac{\partial y}{\partial t} - \frac{\partial^2 y}{\partial t^2}}{(1 - m)a_R} y^2 = n_0 y^2 - \frac{d_R y - \frac{\partial y}{\partial t}}{(1 - m)a_R} a_R - \frac{d_R y - \frac{\partial y}{\partial t}}{(1 - m)a_R} a_{Pp} y \quad (7.45)$$

co można zapisać też jako:

$$\frac{\partial^2 y}{\partial t^2} = (d_R a_{Pp} - n_0(1 - m)a_R + \frac{d_R a_R}{y}) + (d_R - \frac{a_{Pp}}{y} - \frac{a_R}{y^2}) \frac{\partial y}{\partial t} \quad (7.46)$$

W ogólności (traktując p jako parametr), powyższe równanie jest postaci:

$$\frac{\partial^2 y}{\partial t^2} = P(y) + Q(y) \frac{\partial y}{\partial t} \quad (7.47)$$

Można je zinterpretować z punktu widzenia fizyki jako równanie wahadła z tarcieniem. $P(y)$ jest przyspieszeniem zależnym od aktualnego wychylenia, a $Q(y)$ reprezentuje opory ruchu (działające w przeciwną stronę do prędkości). Podana interpretacja ma sens pod dwoma istotnymi warunkami:

- $P(y)$ ma miejsce zerowe (punkt równowagi), ponadto wartość funkcji jest ujemna dla każdej wartości y , która jest większa, a dodatnia dla mniejszych y ,

- $Q(y)$ jest ujemne,

Znalezienie miejsca zerowego y_E (co też automatycznie dowodzi jego istnienia) funkcji $P(y)$ jest trywialne:

$$d_R a_{PP} - n_0(1-m)a_R + \frac{d_R a_R}{y_E} = 0 \quad (7.48)$$

$$y_E = \frac{d_R a_R}{n_0(1-m)a_R - d_R a_{PP}} \quad (7.49)$$

y_E musi być większe od 0, ponieważ y jako odwrotność stężenia replikaz nie może być ujemne, oznacza to że:

$$n_0(1-m)a_R - d_R a_{PP} > 0 \quad (7.50)$$

$$p < \frac{n_0(1-m)a_R}{d_R a_P} \quad (7.51)$$

Warunek pozwalający na zastosowanie analogii z wahadłem jest spełniony, ponieważ $P(y)$ jest funkcją ściśle malejącą (dla $y > 0$), zatem jest ujemne dla $y > y_E$ i dodatnie dla $y < y_E$.

Pozostała jeszcze kwestia znaku funkcji $Q(y)$:

$$d_R - \frac{a_{PP}}{y} - \frac{a_R}{y^2} \leq 0 \quad (7.52)$$

Mnożąc przez y^2 :

$$d_R y^2 - a_{PP} y - a_R \leq 0 \quad (7.53)$$

Jest to nierówność kwadratowa, którą łatwo można rozwiązać (y musi oczywiście być dodatnie):

$$y \leq \frac{a_{PP} + \sqrt{a_P^2 p^2 + 4a_R d_R}}{2d_R} \quad (7.54)$$

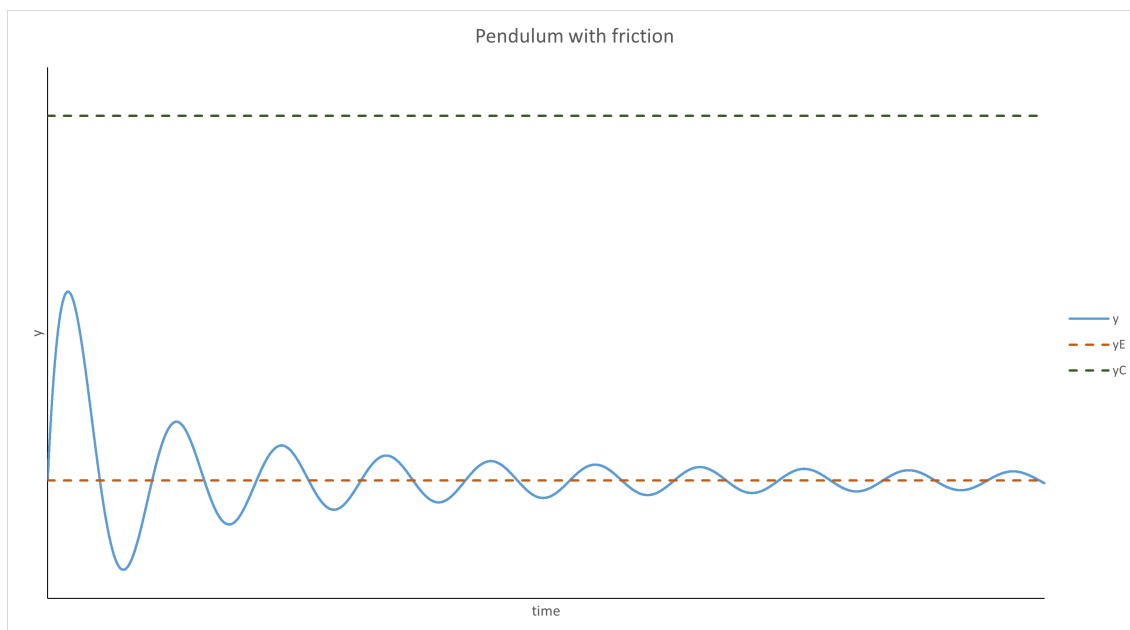
Największa wartość y spełniająca powyższą nierówność będzie nazywana „punktem krytycznym”.

$$y_C = \frac{a_{PP} + \sqrt{a_P^2 p^2 + 4a_R d_R}}{2d_R} \quad (7.55)$$

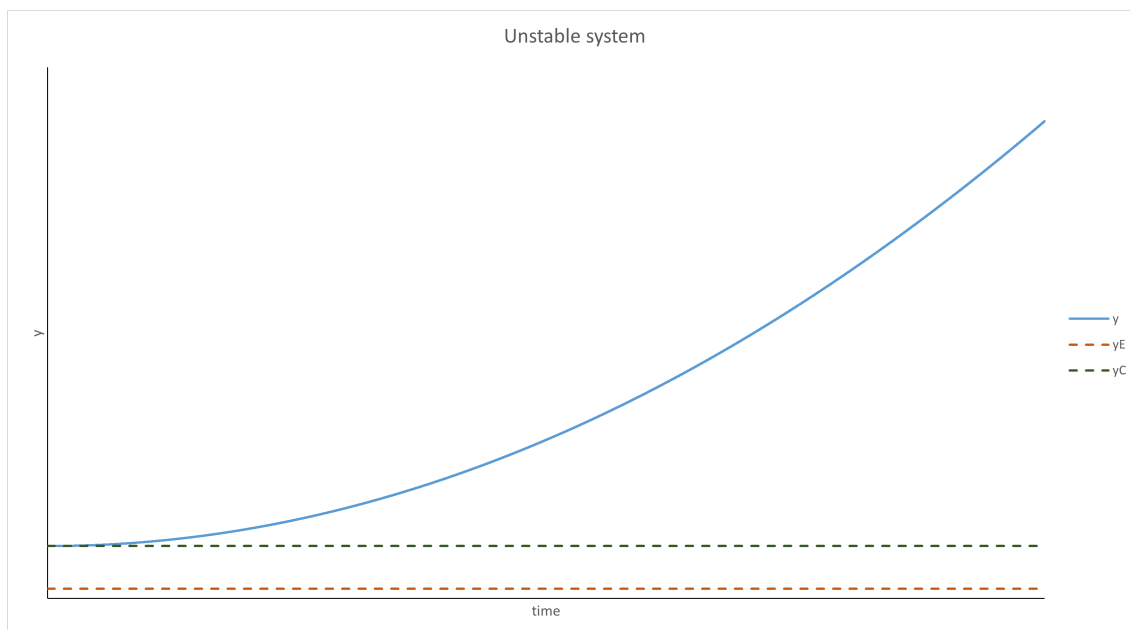
Modelując zachowanie za pomocą wahadła z tarciem, można je w łatwy sposób przewidzieć, należy oczywiście pamiętać, że wspomniana analogia ma sens tylko jeśli $y_E < y_C$ oraz $y < y_C$. Najważniejszym wnioskiem, który można wyciągnąć z takiego przedstawienia problemu jest fakt, że y będzie oscylowało wokół wartości y_E , jednak z coraz mniejszą amplitudą, ostatecznie zbiegając do tego punktu. Hipotetyczną ewolucję zmiennej y w czasie przedstawiają rys. 7.1 i 7.2. Należy przypomnieć, że p jest w tej analizie traktowane jako parametr, zatem jeśli ta wartość zmienia się zbyt gwałtownie - zastosowana analogia nie jest poprawna.

Mutacja pasożytów

Jeśli nierówność 7.27 nie jest spełniona, nie istnieje punkt równowagi dla populacji. Należy tutaj także wspomnieć, że pasożyty w przeciwieństwie do replikaz nie mają istotnej funkcji (przynajmniej na wczesnym etapie), więc jeśli pozwoli się na mutację, naturalnie będą dążyć do wyższej wartości a_P , ponieważ gwarantuje ona pasożytowi większą szansę na skopiowanie. Oznacza to, że nierówność 7.27 sama w sobie nie gwarantuje przetrwania systemu. Poprzednie modele opisane w tej pracy oraz te stworzone przez N. Takeuchiego [115] pokazały, że mutacje nie muszą powodować śmierci całej



Rys. 7.1: Dla $y < y_C$, system będzie zachowywał się jak wahadło z tarcie, czyli wystąpią słabnące oscylacje wokół punktu równowagi. Ostatecznie wartość y się ustabilizuje.



Rys. 7.2: Jeśli y przekroczy punkt krytyczny, funkcja Q zmieni znak na dodatni, czyli zamiast wygaszać drgania - zacznie je wzmacniać. Funkcja P dalej może reprezentować siłę próbującą sprowadzić wahadło do punktu równowagi, jednak jej wartość jest ograniczona i przy odpowiedniej prędkości czynnik $Q(y) \frac{\partial y}{\partial t}$ może przekroczyć jej możliwości. W takim przypadku y będzie dążyć do nieskończoności, co oznacza że replikazy całkowicie wymrą (y jest odwrotnością stężenia replikaz).

populacji. W celu zamodelowania procesów związanych z mutacją, musi być ona reprezentowana *explicite*, zatem parametr a_P należy zdefiniować na nowo - jako funkcję:

$$a_P = a_P(x, y, z, t) \quad (7.56)$$

Funkcja ta reprezentuje średnią wartość parametru a_P w danym punkcie w przestrzeni, ze względu na infinitesimalnie mały rozmiar objętości branej pod uwagę, można przyjąć że średnia odpowiada po prostu powinowactwu wszystkich pasożytów do replikazy. Uwzględnienie zmian w modelu wymaga określenia reguł rządzących ewolucją a_P w czasie. Wpływ na nią mają trzy czynniki:

1. dyfuzja,
2. losowe mutacje,
3. nowe pasoży powstające z błędnie skopiowanych replikaz.

Dla uproszczenia przyjęto założenie, że w trzecim przypadku nowe pasożyty mają wartość a_P równą pozostałym, czyli nie wpływają one na wartość funkcji w danym punkcie. Przed zapisaniem ostatecznego równania trzeba pamiętać, że wartość a_P nie podlega dyfuzji bezpośrednio, tylko polega na nośnikach - pasożytach. Powoduje to, że bezpośrednio zastosowanie drugiego prawa Ficka jest niemożliwe. Jednak wartość $a_P p$ będąca odpowiednikiem sumy wartości powinowactwa dla wszystkich pasożytów podlega już temu prawu:

$$\frac{\partial(a_P p)}{\partial t} = D_P \nabla^2(a_P p) + \frac{\sigma(x, y, z, t)}{\sqrt{p}} p \quad (7.57)$$

gdzie σ jest funkcją losową, która reprezentuje zmiany w wartościach a_P . Zgodnie z Centralnym Twierdzeniem Granicznym powinna być ona określona rozkładem Gaussa, gdyż jest średnią zmian wartości a_P wszystkich pasożytów w danym punkcie (są one niezależne i mają skończoną wariancję ze względu na ograniczoną w praktyce wartość a_P). Ze względu na własności średniej zmiennych losowych, wpływ mutacji na zmiany a_P powinien być dodatkowo proporcjonalny do $\frac{1}{\sqrt{p}}$ (odchylenie standardowe średniej jest odwrotnie proporcjonalne do pierwiastka z liczby próbek). Jest to zgodne z praktycznymi obserwacjami w populacjach biologicznych - im więcej osobników w grupie, tym wolniej zachodzi ewolucja. Równanie 7.57 można rozpisać jako:

$$p \frac{\partial a_P}{\partial t} + a_P \frac{\partial p}{\partial t} = D_P \nabla^2 a_P p + D_P a_P \nabla^2 p + 2D_P \vec{\nabla} a_P \cdot \vec{\nabla} p + \sigma \sqrt{p} \quad (7.58)$$

Jako, że rozpad i tworzenie się nowych pasożytów nie zmienia wartości a_P , drugie prawo Ficka wciąż można stosować do pasożytów w czystej postaci (rozważamy tutaj tylko zmiany w liczebności pasożytów wynikające z dyfuzji):

$$\frac{\partial p}{\partial t} = D_P \nabla^2 p \quad (7.59)$$

Równanie 7.58 przyjmuje postać:

$$p \frac{\partial a_P}{\partial t} + a_P D_P \nabla^2 p = D_P \nabla^2 a_P p + D_P a_P \nabla^2 p + 2D_P \vec{\nabla} a_P \cdot \vec{\nabla} p + \sigma \sqrt{p} \quad (7.60)$$

Przekształcając w celu otrzymania $\frac{\partial a_P}{\partial t}$:

$$\frac{\partial a_P}{\partial t} = D_P \nabla^2 a_P + 2 \frac{D_P}{p} \vec{\nabla} a_P \cdot \vec{\nabla} p + \frac{\sigma}{\sqrt{p}} \quad (7.61)$$

Zgodność modelu z rzeczywistością

Wszystkie analizy zostały przeprowadzone przy pewnych założeniach i uproszczeniach, które dały możliwość wyciągnięcia wniosków w sposób analityczny. Poniżej omówione jest kilka hipotetycznych scenariuszy, w których nie wszystkie poczynione założenia byłyby spełnione.

1. Dyfuzja - obecność dyfuzji pozwala na interakcje w przestrzeni między różnymi fragmentami roztworu. Jeśli istnieje stabilny punkt równowagi, powinien być on osiągnięty w całej objętości. Jednak w bardziej skomplikowanych przypadkach (obecności mutacji i/lub braku wspomnianego punktu równowagi) mogą formować się nietrywialne struktury przestrzenne pozwalające na przetrwanie systemu, podobnie jak w rozdziale 5 oraz w pracy N. Takeuchiego [115]. Przypadki te zostały zbadane za pomocą symulacji komputerowych i są omówione w dalszej części rozdziału.
2. Tworzenie się kompleksów - w tym rozdziale replikacja RNA jest traktowana jak reakcja przebiegająca natychmiastowo, jednak w rzeczywistości trwa to pewien czas. Fakt ten można uwzględnić dodając do modelu nowy typ cząsteczki - kompleks replikazy ze wzorcem, który oznacza aktywną replikację. Zostało to zrobione w poprzednich modelach (rozdziały 5 i 6), zachowanie populacji było bardzo podobne do otrzymanego w dalszej części tego rozdziału. Kompleksy zdają się zatem nie mieć dużego wpływu na ewolucję populacji.
3. Inna dynamika reakcji chemicznych - w tym modelu zakłada się, że wszystkie reakcje chemiczne przebiegają zgodnie z prawem działania mas. Ze względu na dość skomplikowaną naturę złożonych cząsteczek takich jak RNA, replikazy mogą zachowywać się w zupełnie inny sposób w zależności od dostępności zasobów i struktury wzorca. Przyjęto najprostszemu scenariusz, ponieważ w chwili pisania tej pracy nie udało się w żadnym laboratorium stworzyć funkcjonalnej replikazy, nie wiadomo zatem jakie dokładnie byłyby właściwości i zachowanie takiej cząsteczki. Jakikolwiek większe zmiany we właściwościach chemicznych RNA w modelu mogłyby radykalnie zmienić wynik, np. replikazy rozpoznające i powielające jedynie inne replikazy byłyby w stanie zapewnić populacji przetrwanie bez żadnych dodatkowych mechanizmów. Jeśli jednak próbujemy badać pierwsze, najbardziej prymitywne replikazy, założenie że działają one w prosty sposób może być uzasadnione.

7.3 Algorytm

Model z mutacją pasożytów okazał się bardzo trudny do analizy teoretycznej, żeby zbadać ewolucję populacji przy takich założeniach, należało się wspomóc innymi dostępnymi metodami - symulacjami komputerowymi. W tym celu wykorzystano dwa podejścia - pierwsze bazowało na numerycznym rozwiązaniu opisanych wcześniej równań różniczkowych, a drugie wykorzystywało zmodyfikowany algorytm wieloagentowy bazowany na tym z rozdziału 5. Dwie odrębne metody symulacji pozwalały na sprawdzenie otrzymanych rezultatów przez porównanie wyników.

Algorytm wykorzystany do symulacji równań różniczkowych

Jak wspomniano powyżej, autor podjął się numerycznego rozwiązania opisanych w tym rozdziale równań różniczkowych (równania 7.7, 7.8, 7.9 i 7.61). Konstrukcja odpowiedniego algorytmu wymagała dyskretyzacji problemu, dlatego nieskończone małe wartości dx , dy , dz i dt zastąpiono (małymi) liczbami rzeczywistymi Δx , Δy , Δz oraz Δt . Stan systemu reprezentowany jest za pomocą dwuwymiarowej tablicy, w której każdy element reprezentuje wycinek przestrzeni i przechowuje cztery zmienne:

- p - stężenie pasożytów
- r - stężenie replikaz

- a - średnie a_P pasożytów w danym punkcie
- n - stężenie zasobów potrzebnych do replikacji

Algorytm został zaprezentowany poniżej w pseudokodzie (algorytm 5), wszystkie parametry globalne są podane w tabeli 7.1. Należy podkreślić, że dla uproszczenia tempo mutacji pasożytów nie było zależne od ich liczebności w przeciwieństwie do tego co wynikało z równania 7.61.

Algorytm 5 Symulacja równań różniczkowych cząstkowych (patrz sekcja 7.2)

```

Zainicjalizuj dwie tablice dwuwymiarowe o rozmiarze  $sizeX$  na  $sizeY$  -  $current$  i  $next$ 
for all elementów  $f$  tablicy  $current$  do
  Ustaw  $f.p$ ,  $f.r$ ,  $f.a$  i  $f.n$  na 1.
end for
for all kroków symulacji nr  $step$  do
  for all elementów  $f$  tablicy  $current$  do
     $newreplicases = a_R * \Delta t * f.r * f.r * (1 - m_{rp}) * f.n$ 
     $newparasites = \Delta t * f.r * (a_R * f.r * m_{rp} + f.a * f.p) * f.n$ 
     $resourcesused = newreplicases + newparasites$ 
    if  $resourcesused > 0$  and  $resourcesused > n$  then
       $newreplicases = newreplicases * n / resourcesused$ 
       $newparasites = newparasites * n / resourcesused$ 
    end if
    Oblicz średnią każdej zmiennej z pól przylegających do  $f$ :  $avg\_r$ ,  $avg\_p$ ,  $avg\_n$  and  $avg\_a$  (średnia wartość  $a * p$ ).
    Znajdź pole o takich samych współrzędnych jak  $f$  w tablicy  $next$  -  $f2$ .
     $f2.r = f.r + newreplicases - d * \Delta t * f.r + D * \Delta t * (avg\_r - f.r)$ 
     $f2.p = f.p + newparasites - d * \Delta t * f.p + D * \Delta t * (avg\_p - f.p)$ 
     $new\_a = f.a * f.p + D * \Delta t * (avg\_a - f.a * f.p)$ 
    if  $f.p > 0$  then
       $f2.a = new\_a / f.p$ 
    else
       $f2.a = f.a$ 
    end if
     $f2.n = f.n + n_0 * \Delta t - newreplicases - newparasites + D_n * \Delta t * (avg\_n - f.n)$ 
    Losowo zmień  $f2.a$  zgodnie z rozkładem prawdopodobieństwa mutacji - z równym prawdopodobieństwem dodaj lub odejmij  $m * \Delta t$ .
    Wszystkie ujemne zmienne w  $f2$  ustaw na 0.
  end for
  Zapisz wyniki obecnego kroku symulacji.
  Zamień  $current$  z  $next$ .
end for

```

Algorytm wykorzystany w symulacji wieloagentowej (MAS)

Model oparty na równaniach różniczkowych cząstkowych opisuje populacje cząsteczek RNA za pomocą funkcji o wartościach rzeczywistych, naturalną alternatywą jest wykorzystanie dyskretnego sposobu modelowania systemu. Jak wspomniano powyżej, w tym celu wykorzystano lekko zmodyfikowany algorytm z rozdziału 5.

Parametry algorytmu MAS

W symulacji zdefiniowano szereg parametrów globalnych (tabela 7.2). Pojedynczy agent ma ponadto też własne parametry:

- pozycja

Nazwa	Wartość domyślna	Opis
$sizeX, sizeY$	1000	rozmiar obszaru symulacji
d	0,01	stała rozpadu
a_R	1,0	stała szybkości reakcji replikacji replikazy
Δt	0,1	czas trwania jednego kroku (Δt w równaniach)
D	5,0	stała dyfuzji dla RNA
D_n	10,0	stała dyfuzji dla zasobów
n_0	1,0	tempo powstawania zasobów
m_{rp}	0,0	prawdopodobieństwo niepoprawnego skopiowania replikazy
m	1,0	tempo mutacji pasożytów

Tab. 7.1: Parametry globalne w symulacji równań różniczkowych cząstkowych (rozwiązaniu numerycznym).

- typ - replikaza/pasożyt
- k_P (tylko pasożyty) - prawdopodobieństwo replikacji

Nazwa	Wartość domyślna	Opis
$sizeX, sizeY$	800	rozmiar obszaru symulacji
$agent_size$	3,0	promień koła reprezentującego agenta
N_{max}	4	maksymalna liczba sąsiadów
d	0,1	stała rozpadu
k_R	1,0	powinowactwo replikazy do replikazy
Δt	1,0	czas trwania pojedynczego kroku
D	15,0	stała dyfuzji
δ	0,1	szybkość mutacji pasożyta
m_P	0,1	prawdopodobieństwo mutacji pasożyta

Tab. 7.2: Parametry globalne w symulacji wieloagentowej.

Reakcje pierwszorzędowe

Jedyną reakcją pierwszego rzędu w symulacji jest spontaniczny rozpad cząsteczek. Prawdopodobieństwo rozpadu zgodnie z wcześniejszymi założeniami jest określone rozkładem wykładniczym, można je dla danej wartości stałej rozpadu d obliczyć ze wzoru:

$$p_d = 1 - e^{-d\Delta t} \quad (7.62)$$

Najprostszym podejściem jest sprawdzanie co krok, czy zaszła reakcja rozpadu, jednak w celu zwiększenia efektywności wybrano podobne podejście jak w rozdziale 5. Każdy agent ma na wstępie ustawiany czas życia („RLT” - z ang. „Remaining Life Time”), który jest dekrementowany co krok, a kiedy dojdzie do 0 - dochodzi do rozpadu. Powyższe dwa podejścia są statystycznie równoważne, RLT można wyznaczyć ze wzoru:

$$RLT = \lfloor -\frac{\ln X}{d\Delta t} \rfloor + 1 \quad (7.63)$$

gdzie X zmienną losową o rozkładzie równomiernym z przedziału $(0; 1]$.

Rekcje drugorzędowe

Tego typu reakcje mogą zajść kiedy dochodzi do zetknięcia się dwóch agentów (przecięcia kół, które je reprezentują). Reakcja może zajść z określonym prawdopodobieństwem zależnym od ich typów:

- dwa pasożyty - brak reakcji
- pasożyt i replikaza - pasożyt staje się wzorcem replikacji, prawdopodobieństwo zajścia reakcji - k_P pasożyta
- dwie replikazy - replikacja zachodzi z prawdopodobieństwem k_R (parametr globalny), dowolna z nich staje się wzorcem

W wyniku reakcji tworzona jest kopia wzorca, jeśli jest on pasożytem - nowy agent zostaje dodatkowo poddany mutacji. Warto podkreślić, że k_P oraz k_R nie są tymi samymi parametrami co a_P i a_R , ponieważ te pierwsze są wartościami prawdopodobieństwa, a te drugie - stałymi szybkości reakcji.

Mutacja

Każdy nowo powstający pasożyt poddawany jest mutacji z prawdopodobieństwem m_P . Jeśli do niej dojdzie, wtedy parametr k_P nowego agenta modyfikuje się o losową wartość (rozkład równomierny) z przedziału $[-\frac{\delta}{2}; \frac{\delta}{2}]$.

Dyfuzja

Dyfuzja jest symulowana za pomocą dynamiki Browna (BD). Na początku każdego kroku symulacji, każdy agent zmienia swoje położenie o losowy wektor $\vec{\psi}$. Wartość tego ostatniego oblicza się ze wzoru:

$$\vec{\psi} = \sqrt{2D\Delta t}\vec{\xi} \quad (7.64)$$

gdzie ξ jest wektorem losowym zgodnym z rozkładem Gaussa o jednostkowej wariancji i średniej 0.

7.4 Implementacja

Algorytm symulujący równania różniczkowe zaimplementowano w języku C++ na systemie operacyjnym Ubuntu, główny problem jaki napotkano to konieczność podmiany tablic *current* z *next*, gdyż byłoby to bardzo kosztowne czasowo. Problem ten rozwiązano zamieniając miejscami jedynie wskaźniki.

Dla symulacji wieloagentowej, nowy algorytm zaimplementowano jako kolejną opcję w symulatorze opisanym w rozdziale 5, czyli użyto języka C# i zaprogramowano jedynie rodzaje agentów, zasady na których wchodzi w interakcję oraz ich wygląd na obrazie planszy - logika symulacji, interfejs oraz generowanie plików wyjściowych były częściami wykorzystanego programu już wcześniej.

7.5 Wyniki

Głównym celem symulacji było zbadanie najbardziej interesującego przypadku, kiedy nierówność 7.27 nie jest spełniona i nie można wykazać przeżywalności systemu analizując dobrze wymieszany roztwór. Aby zbadać różne przypadki, zdefiniowano kilka scenariuszy. Do części użyto

Algorytm 6 Algorytm MAS

Zainicjuj początkowe pozycje agentów losowo i odpowiednio ustaw każdemu RLT .

Ustaw k_P wszystkich pasożytów na ustaloną wartość początkową.

while (czas symulacji < limit czasu) **and** (w symulacji są zarówno replikazy jak i pasożyty) **do**

 Zmniejsz RLT każdego agenta o 1,

for all agentów z $RLT = 0$ **do**

 Usuń agenta z symulacji.

end for

 Stwórz kolejkę agentów w kolejności losowej.

for all agentów x_i w kolejce **do**

 Zinicjalizuj pusty zbiór sąsiadów N .

for all agentów $an \neq x_i$, które przecinają się z x_i **do**

 Dodaj an do N .

end for

if $|N| > N_{max}$ **then**

 Usuń agenta x_i z symulacji.

end if

 Ustaw agentów w N w kolejności losowej.

 Przesuń x_i o losowy wektor (rozkład Gaussa z wariancją $2D\Delta t$).

for all $n_j \in N$ **do**

if (n_j lub x_i jest replikazą) **and** zaszła reakcja (prawdopodobieństwo k_P/k_R) **then**

 Stwórz kopię wzorca i umieść na tej samej pozycji.

 Zinicjalizuj wartość RLT nowego agenta.

if Nowy agent jest pasożytem **then**

 Sprawdź czy zaszła mutacja i odpowiednio zmodyfikuj k_P .

end if

end if

end for

end for

end while

modelu wieloagentowego (scenariusze 1-8), a do pozostałych numerycznego przybliżenia równań różniczkowych. Wszystkie otrzymane wyniki zostały przedyskutowane na końcu tej sekcji.

Scenariusze 1-8

Pierwsze 8 scenariuszy przesymulowano za pomocą systemu wieloagentowego, nierówność 7.27 nie musiała być w nich spełniona. Stała rozpadu była identyczna dla pasożytów i replikaz, ale te pierwsze miały średnio większe powinowactwo. Wszystkie parametry z wyjątkiem D i m_P miały wartości domyślne (tab. 7.2). Celem symulacji było ustalenie czy system może przetrwać przy różnych zestawach parametrów i przy aktywnej mutacji pasożytów. Wartości D i m_P dla poszczególnych scenariuszy zostały przedstawione w tabeli 7.3, przetestowano cztery wartości stałej dyfuzji (odpowiadające czterem różnym szybkościom) oraz dwie wartości prawdopodobieństwa mutacji odpowiadające wolnej i szybkiej mutacji.

Za każdym razem kiedy systemowi udało się przetrwać, widoczne były „fale” - nierównomierności w rozmieszczeniu cząsteczek (wykresy 7.3, 7.4, 7.5 i 7.6). Replikazy mają na ilustracjach kolor szary, a pasożyty oznaczone są barwą zależną od ich parametru k_P , skala zaczyna się od czerwonego (0,0), przechodzi przez zielony (0,37) i niebieski (0,74) a kończy na magencie (1,0).

Można zaobserwować, że lokalnie pasożyty zawsze ewoluowały w kierunku wyższej wartości k_P , jednak powodowało to śmierć populacji w danym obszarze. Przetrwanie systemu było możliwe dzięki losowemu charakterowi dyfuzji i mutacji - powstawały wyspy gdzie pasożytów było za mało lub były za słabe, żeby wyprzeć replikazy i doprowadzić do śmierci populacji. Te miejsca stawały się zaczątkiem nowych, rozrastających się populacji i cały proces ewolucji pasożytów zaczynał się od początku. Ważnym parametrem jest liczba kroków T , która stanowi okres trwania opisanego przed chwilą cyklu i może zależeć od wszystkich zastosowanych w symulacji parametrów. Niestety funkcja pozwalająca wyliczyć T może nie być możliwa do analitycznego wyprowadzenia, prawdopodobnie do jej wyznaczenia będzie trzeba wykorzystać obliczenia numeryczne.

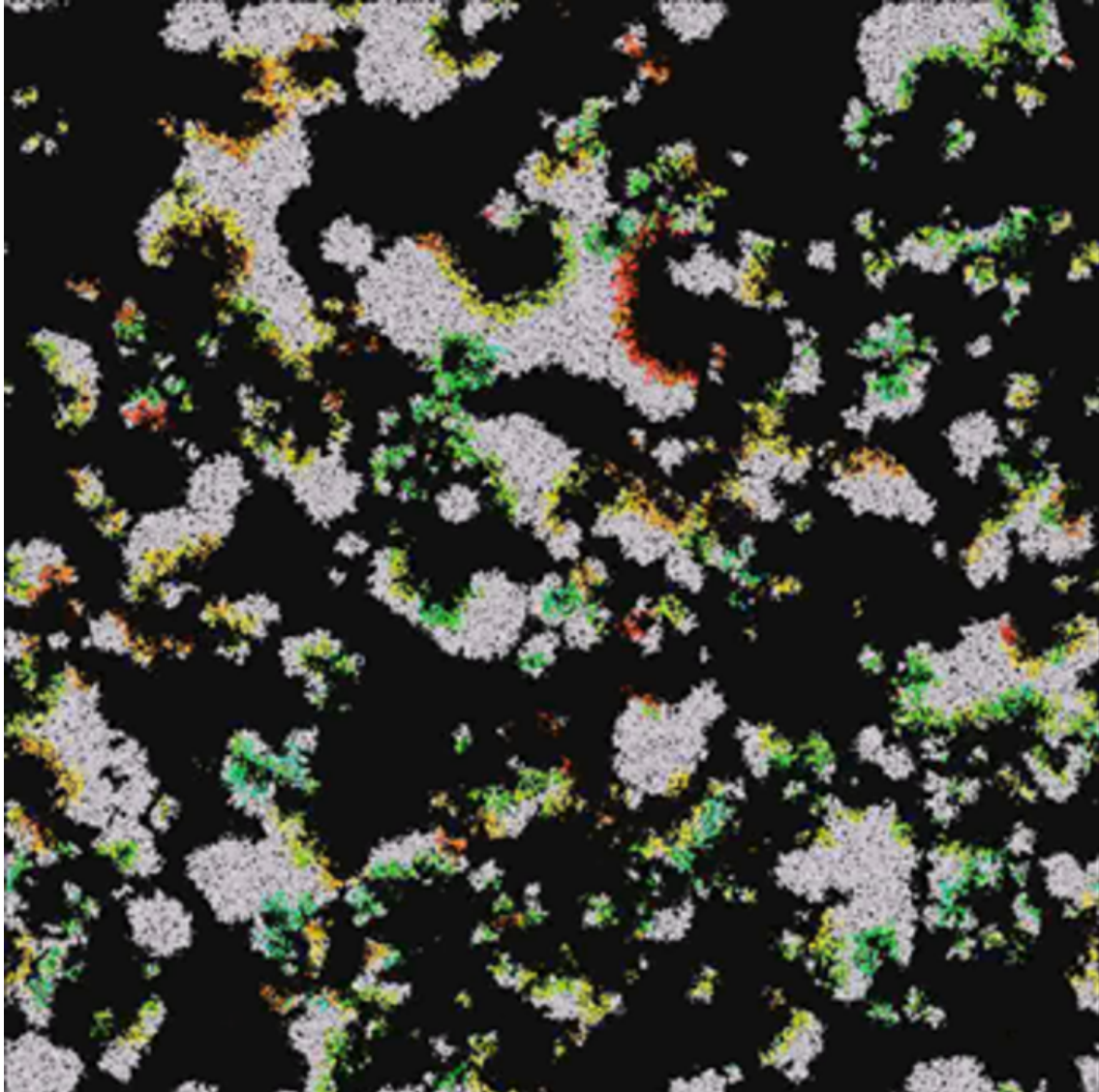
Jeśli oprócz T zdefiniuje się jeszcze dwa parametry - prędkość przemieszczania się granicy rozrastającej się populacji v oraz minimalna powierzchnia gwarantująca losowe pojawienie się „wysp” mogących dać początek nowym populacjom, można zapisać nierówność będącą warunkiem przetrwania systemu:

$$\pi(vT)^2 \geq P \quad (7.65)$$

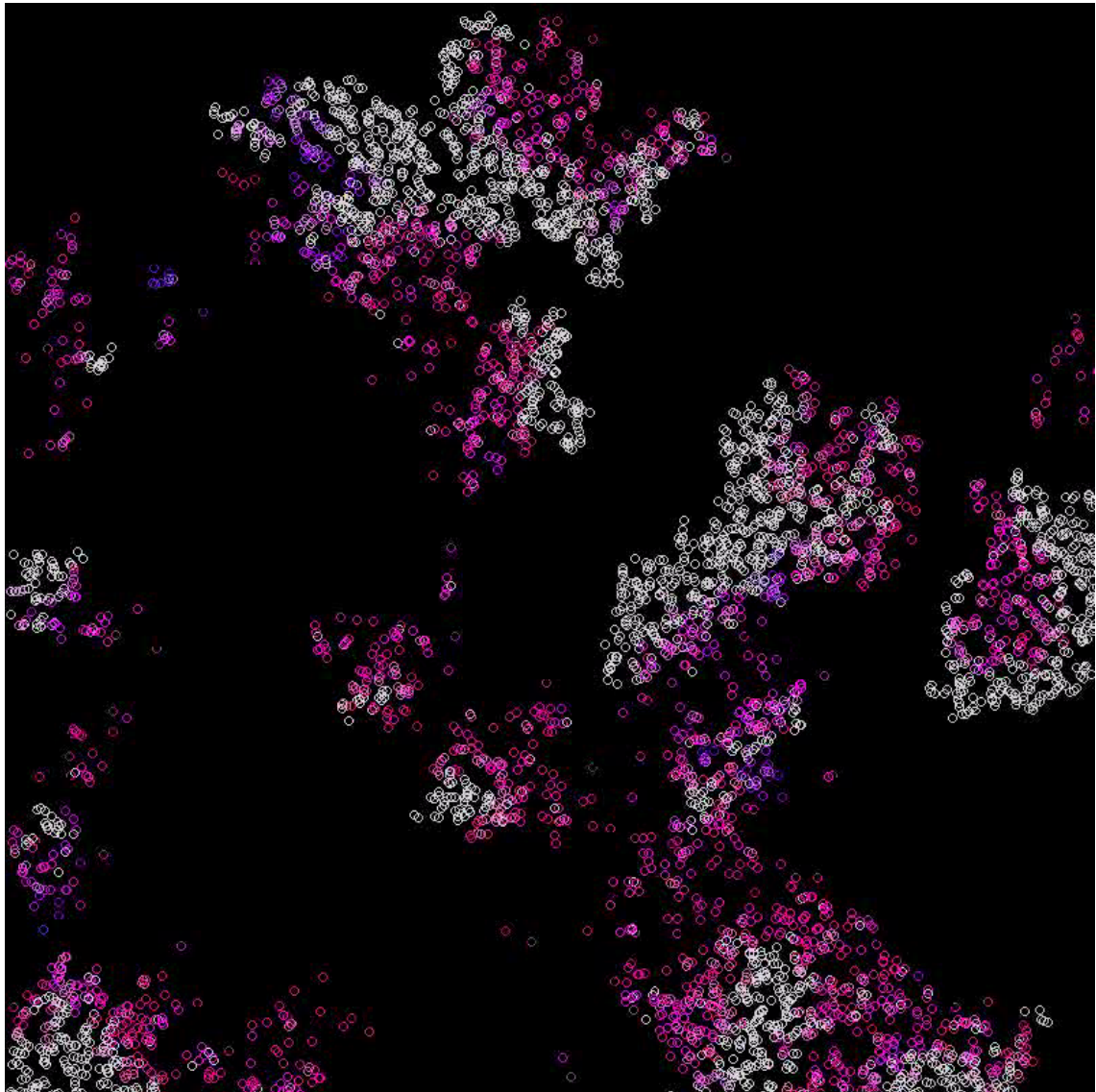
Wyznaczenie parametrów v , P oraz T może być przedmiotem przyszłych badań.

Stała dyfuzji	Prawdopodobieństwo mutacji	Stan systemu
5	0,1	Żyje (wykres 7.4)
5	0,2	Żyje (wykres 7.5)
10	0,1	Żyje (wykres 7.6)
10	0,2	Wymarł
15	0,1	Wymarł
15	0,2	Wymarł
20	0,1	Wymarł
20	0,2	Wymarł

Tab. 7.3: Scenariusze 1-8 symulowane za pomocą systemu wieloagentowego.



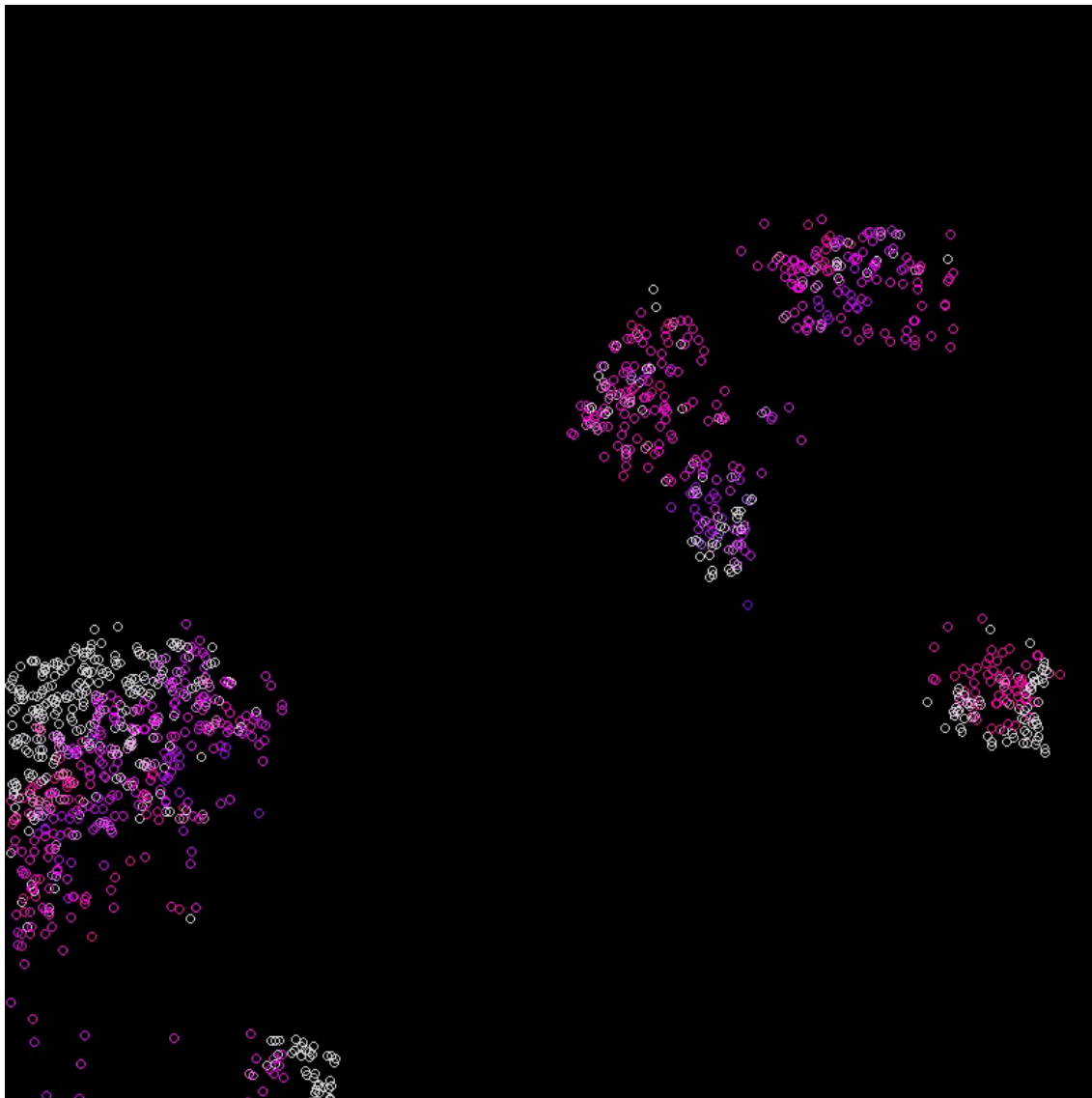
Rys. 7.3: Fale powstające w symulacji wieloagentowej. Reprezentują one różne etapy ewolucji przez które przechodzi populacja. Replikazy oznaczone są kolorem szarym, podczas gdy pasożyty mają kolor zależny od wartości powinowactwa do replikaz. Skala zaczyna się od czerwonego (duża wartość), przechodzi przez żółty, zielony i niebieski, kończąc się na purpurowym dla bardzo małych wartości [110].



Rys. 7.4: Fale powstające przy wartości stałej dyfuzji 5 i prawdopodobieństwu mutacji 0,1. Replikazy są szare, pasożyty purpurowe.



Rys. 7.5: Fale powstające przy wartości stałej dyfuzji 5 i prawdopodobieństwu mutacji 0,2. Replikazy są szare, pasożyty purpurowe.



Rys. 7.6: Fale powstające przy wartości stałej dyfuzji 10 i prawdopodobieństwu mutacji 0,1. Replikazy są szare, pasożyty purpurowe.

Scenariusz 9

Głównym celem tego scenariusza było zasymulowanie dobrze wymieszanego systemu za pomocą algorytmu wieloagentowego. Prawdopodobieństwo mutacji ustawiono na 0, a pozostałe parametrom nadano wartości pozwalające na spełnienie nierówności 7.27. Dyfuzję ustawiono na 5,0, a k_P dla wszystkich pasożytów na 0,6. Pozostałe parametry miały swoje domyślne wartości. Wynik symulacji przedstawione są na rys. 7.7. Najważniejszą obserwacją jest koegzystencja pasożytów i replikaz, populacja przetrwała mimo że rozkład populacji był w przybliżeniu równomierny. Jest to kolejny dowód na to, że jeśli pasożyty są za słabe do wyparcia replikaz - system może bez problemu przetrwać. Jeśli chodzi o ewolucję symulacji w czasie (przed ustaleniem się równowagi), populacja zachowywała się jak wahadło, zgodnie z równaniem 7.46, fakt ten jest dobrze widoczny na rys. 7.8.

Scenariusz 10

Scenariusz 10 w przeciwieństwie do poprzednich bazował na symulacji równań różniczkowych. Wszystkim parametrom zostały przydzielone domyślne wartości, a stężenia replikaz, pasożytów i zasobów w każdym miejscu symulacji przyjęto jako 1,0.

Symulacja oparta na równaniach różniczkowych pokazała, że a_P (mierzone w losowo wybranym punkcie przestrzeni) podlegała oscylacjom (rys. 7.9) między wartościami 0,4 i 0,6. Wynikały one ze zmian w liczebności populacji oraz cykli opisanych przy omawianiu Scenariuszy 1-8. Ważną obserwacją jest fakt, że zmienna a_P nie przekraczała wartości 0,6 mimo, że lokalnie dobór naturalny powinien cały czas dążyć do jej zwiększenia. W symulacji tak jak w poprzednich scenariuszach pojawiły się fale, są one przedstawione na rys. 7.10.

Scenariusz 11

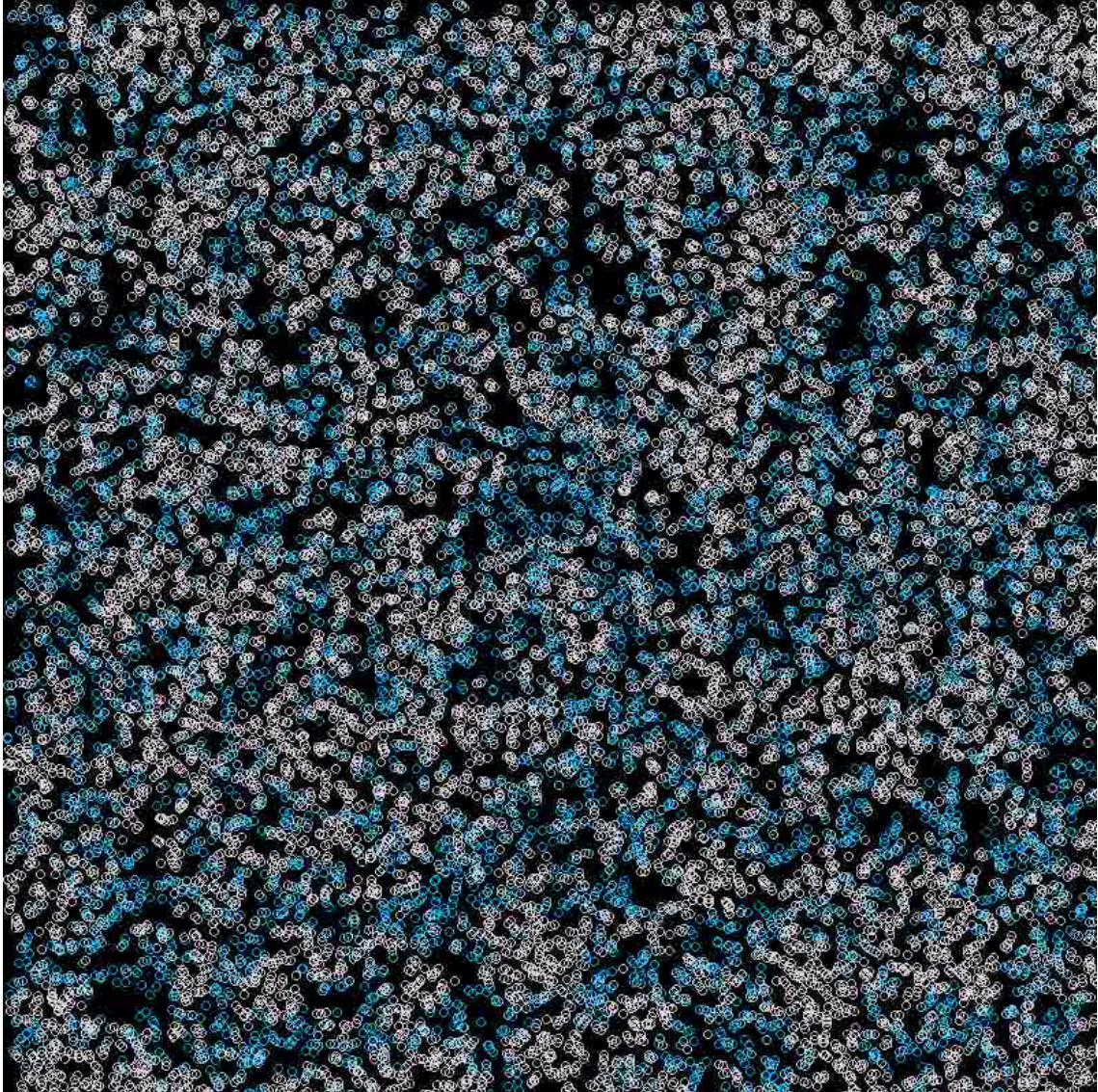
Ten scenariusz był prawie identyczny z poprzednim, jedyną różnicę stanowiła inna wartość parametru m_{rp} (0,1), co pozwalało na powstawanie pasożytów w wyniku replikacji replikazy. Wyniki były bardzo podobne - tutaj także formowały się fale, ale obserwowane różnice w zagęszczeniu populacji były wyraźnie większe.

Scenariusze 12 i 13

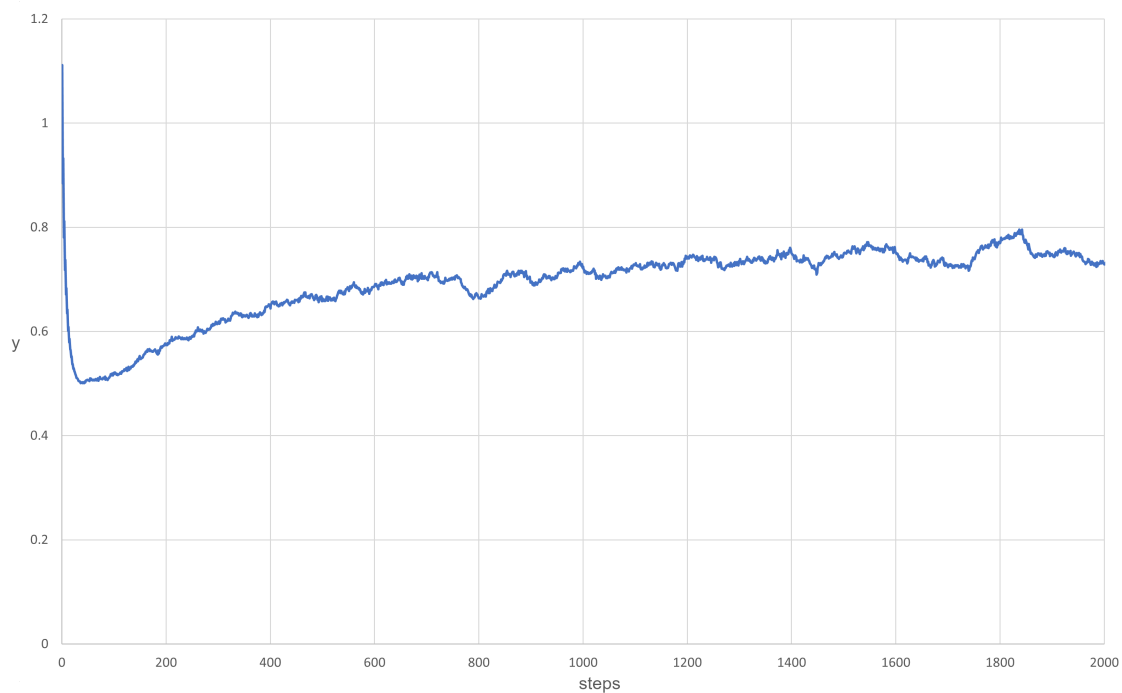
Scenariusze 12 i 13 są odpowiednikami Scenariusza 9, ale zostały zamodelowane z wykorzystaniem równań różniczkowych. Ich celem było oczywiście zweryfikowanie zachowania systemu w momencie kiedy nierówność 7.27 jest w całym rozważanym obszarze spełniona. Mutacje pasożytów oczywiście wyłączono, zerując parametr m . Zgodnie z przewidywaniami - system osiągnął stan równowagi, pasożyty i replikazy koegzystowały. W Scenariuszu 12 (wykres 7.11) stała reakcji dla pasożytów była ustawiona na 0,9 (wszystkie pozostałe parametry poza m miały wartości domyślne), co doprowadziło do silnego zdominowania systemu przez replikazy. Scenariusz 13 różnił się wartością stałej reakcji replikacji pasożytów - wynosiła ona 1,0, pozwoliło to na stan równowagi w którym stężenia replikaz i pasożytów były równe (wykres 7.12).

Podsumowanie scenariuszy 1-13

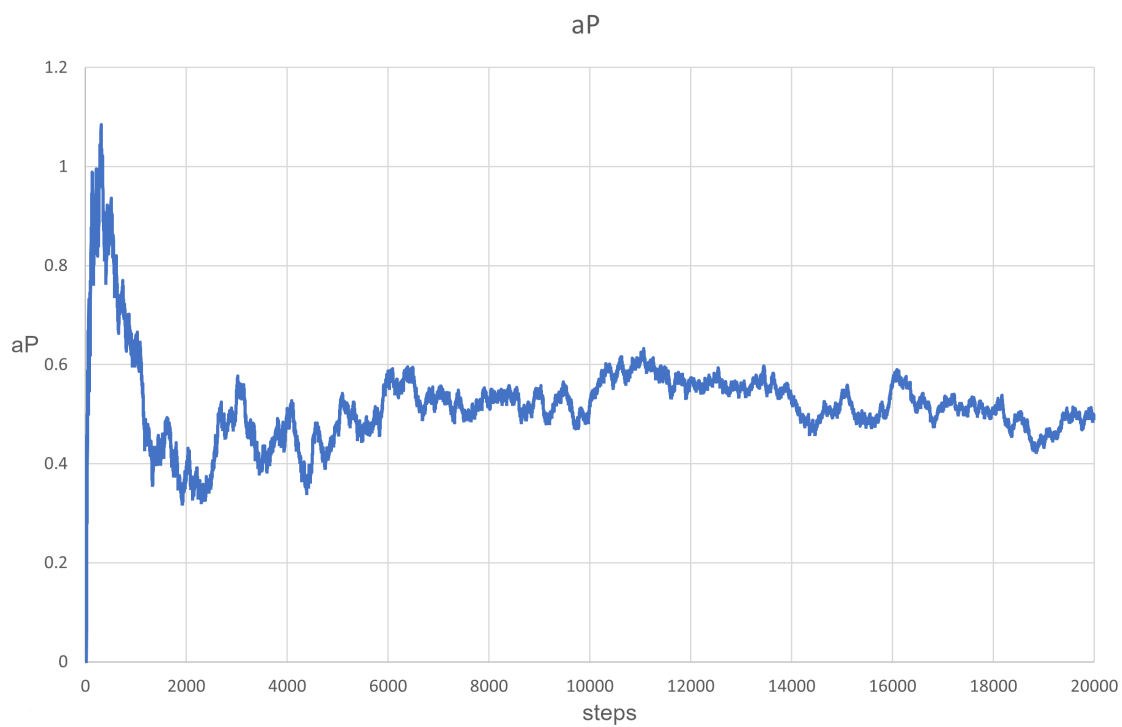
Analiza scenariuszy 1-13 wskazuje, że zarówno w przypadku równań różniczkowych i systemów wieloagentowych system jest lokalnie niestabilny. Selekcja naturalna faworyzuje pasożyty z większą wartością powinowactwa [105, 110]. Populacja rozrasta się z łatwością jeśli lokalnie nierówność 7.27 jest spełniona (populacja w każdym miejscu dąży do stanu równowagi, m.in. zwiększając stężenie replikaz na obrzeżach, wypełniając pustą przestrzeń), jednak z czasem zwiększające się



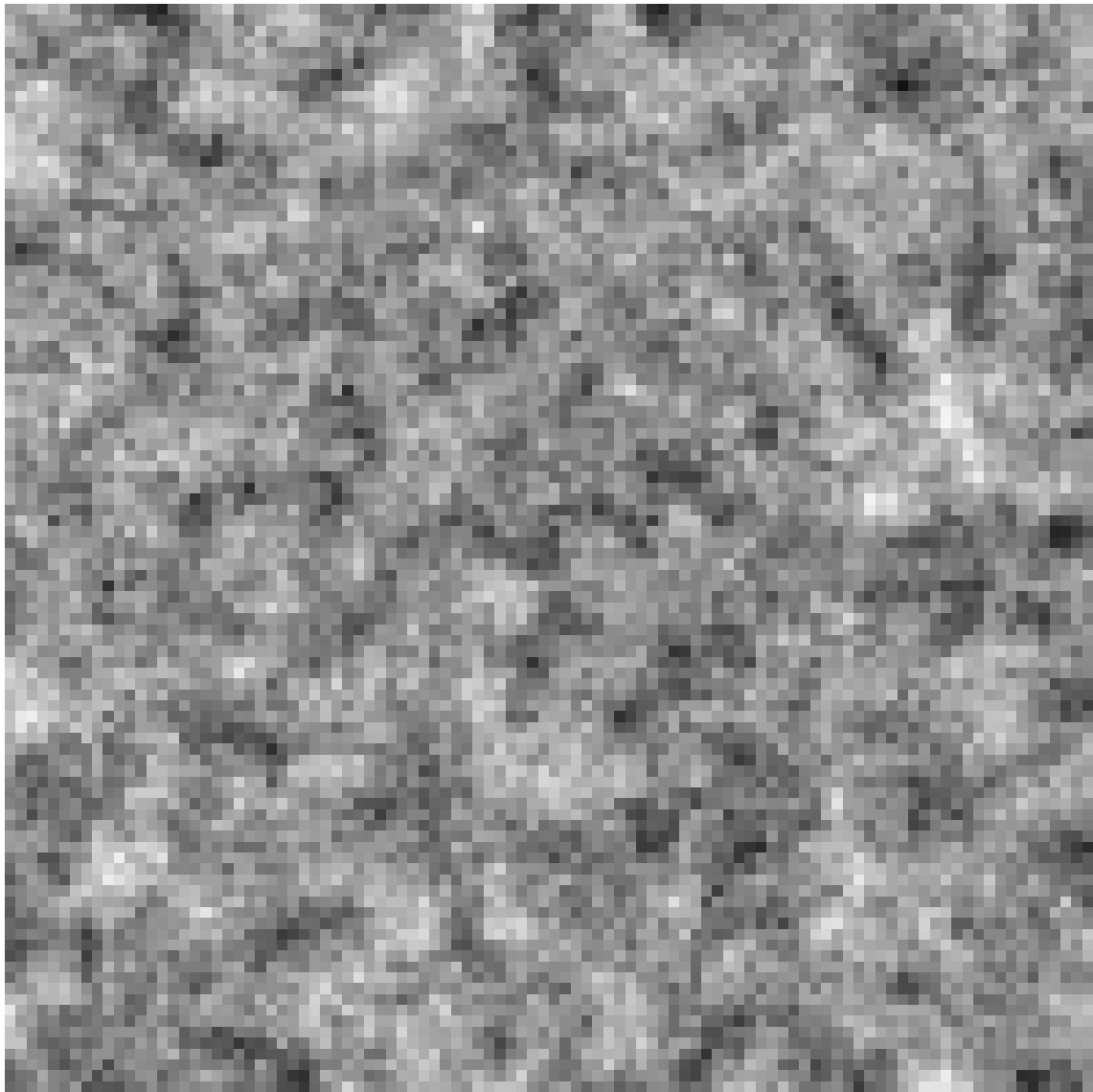
Rys. 7.7: Dobrze wymieszany system w którym możliwa jest koezystencja replikaz i pasożytów.
Na szaro zaznaczono replikazy, na niebiesko - pasożyty.



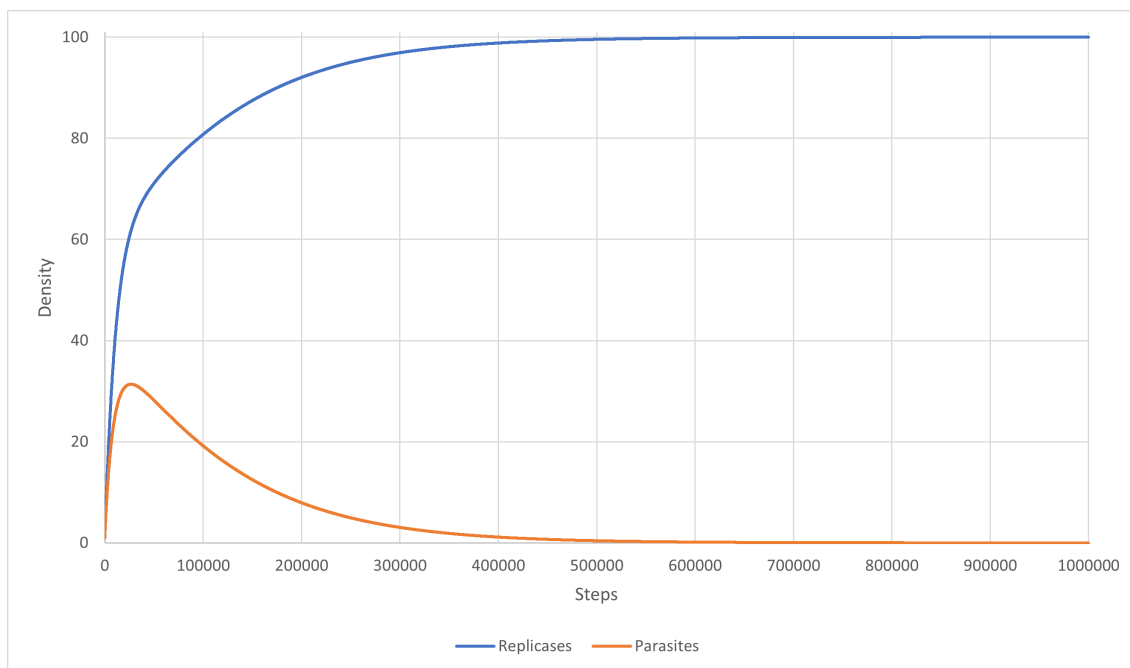
Rys. 7.8: Ewolucja zmiennej y (odwrotności stężenia replikaz pomnożonej przez 10.000 dla czytelności) w czasie jest bardzo podobna do przewidywanej na podstawie równania 7.46. Pojawiają się oscylacje, które zostają szybko wygaszone.



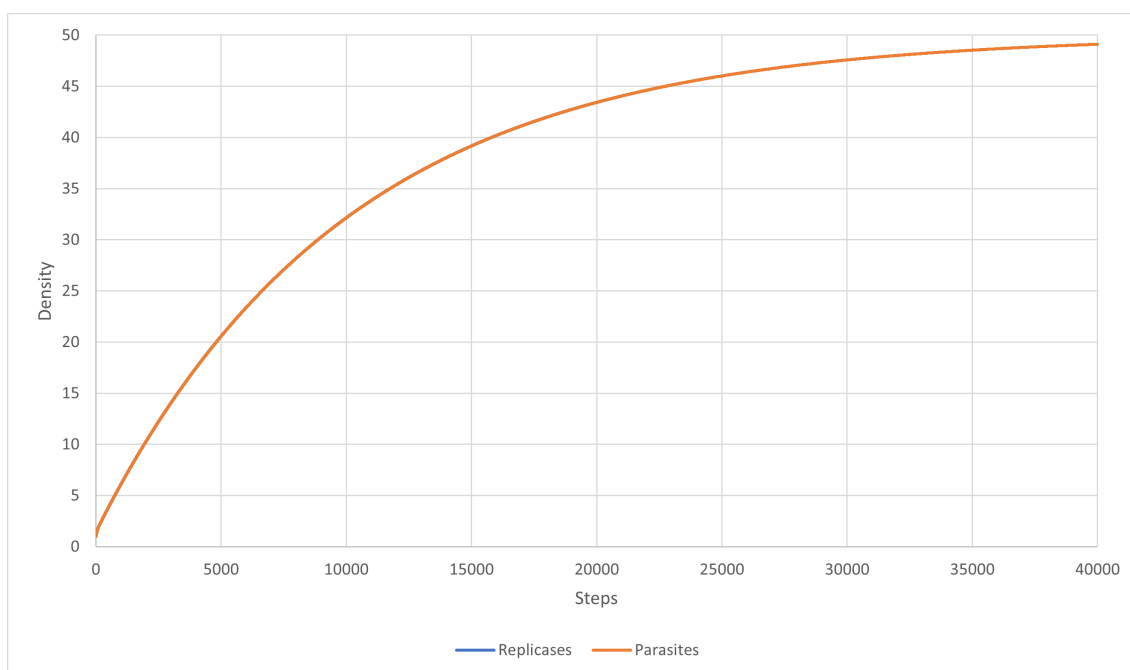
Rys. 7.9: Ewolucja zmiennej a_P w wybranym punkcie przestrzeni. Oscylowała ona między wartościami 0,4 i 0,6.



Rys. 7.10: Fale formujące się w symulacji równań różniczkowych, kiedy nierówność 7.27 nie była w ogólności spełniona. Ciemnym kolorem oznaczono obszary z bardzo małym stężeniem replikaz, a na biało te z dużymi ich skupiskami.



Rys. 7.11: Scenariusz 12, stężenie replikaz i pasożytów. Replikazy zdominowały system ze względu na niskie powinowactwo pasożytów.



Rys. 7.12: Scenariusz 13, stężenia replikaz i pasożytów były równe, dlatego wykresy nakładają się na siebie.

powinowactwo pasożytów będzie powodowało przejście maszynerii replikacyjnej i wyparcie replikaz (nierówność przestanie być spełniona). Ze względu jednak na losowy charakter mutacji, mogą istnieć „wysepki” w obrębie których populacja wciąż jest zdolna do ekspansji, podczas gdy wszystkie replikazy wokół giną (nie mogąc się replikować). Takie miejsca mogą dać początek nowym populacjom, zaczynając cały opisany cykl od początku. Periodyczne zachowanie się systemu w połączeniu z faktem, że populacja może w różnych miejscach być na różnych etapach, może powodować widoczne formowanie się „fal” widocznych na ilustracjach.

7.6 Dyskusja

Zgodnie z wiedzą autora, w odróżnieniu od równań różniczkowych zwyczajnych, równania różniczkowe cząstkowe nie były wcześniej wykorzystane do analizy stabilności ewolucyjnej w kontekście modelu RP. Należy jednak podkreślić, że równania różniczkowe zwyczajne (a także cząstkowe) nie są niczym nowym w szerszej dziedzinie - badaniach początków życia. Eigen [40] za ich pomocą analizował możliwość samoczynnego wyłonienia się maszynerii zdolnej do replikacji w hipotetycznej populacji nośników informacji. Boerlijst razem z Hogeweg [19] pokazali, że z równań różniczkowych zwyczajnych wynika niestabilność hipercykłów liczących więcej niż 4 gatunki. Warto wspomnieć, że w tej samej publikacji analizowano również struktury przestrzenne tworzone przez hipercykl za pomocą równań różniczkowych cząstkowych. W innych badaniach Hogeweg i Takeuchi [114] wykorzystali równania różniczkowe zwyczajne do udowodnienia, że dobrze zmieszany system replikatorowy jest niestabilny ewolucyjnie. W [130] autorzy za pomocą układu równań badali przebieg polimeryzacji pierwszych cząsteczek RNA.

Modele wieloagentowe oraz automaty komórkowe w przeciwieństwie do równań różniczkowych cząstkowych były stosowane co najmniej kilkukrotnie do badań modelu RP [105, 107, 110, 114, 115]. Ich niewątpliwą przewagą jest możliwość zejścia na poziom pojedynczych cząsteczek i definiowania ich właściwości w małej skali, obserwując później jak konkretne założenia przekładają się na właściwości makroskopowe roztworu. Symulacje tego typu pozwalają np. na reprezentowanie informacji genetycznej pojedynczego osobnika *explicite* [105].

Eksperymenty w tym rozdziale pozwoliły na połączenie podejścia makroskopowego reprezentowanego właśnie przez równania różniczkowe zwyczajne oraz równania różniczkowe cząstkowe (pozwalające wprowadzić aspekt przestrzenny) z podejściem mikroskopowym charakterystycznym dla automatów komórkowych i systemów wieloagentowych. Tak jak się spodziewano - otrzymano podobne wyniki, ponieważ oba modele bazowały na podobnych założeniach i zostały stworzone w celu zbadania tego samego zjawiska, mimo że z dwóch różnych perspektyw. Porównanie zebranych danych pozwoliło na otrzymanie wiarygodnych rezultatów.

Modele RP oparte na równaniach różniczkowych spotykane w literaturze [11, 53, 112, 115] z oczywistych względów bardzo upraszczały interakcje między cząsteczkami. W tym rozdziale (i w odpowiedniej publikacji [106]) autor uwzględnił aspekt przestrzenny (w tym dyfuzję) oraz mutację pasożytów, co odróżnia opisywany tutaj model od podejść zastosowanych w artykułach wymienionych wyżej. Niestety i tutaj konieczny był szereg uproszczeń, replikację traktowano jako natychmiastową, założono też że reakcje chemiczne przebiegają zgodnie z prawem działania mas, co dla skomplikowanych cząsteczek biologicznych nie musi być prawdą. Samo traktowanie steżeń jako funkcji ciągłych jest również pewną idealizacją, jednak ze względu na bardzo małe rozmiary cząsteczek chemicznych jest to uzasadnione.

Ważną częścią opisanych w tym rozdziale analiz jest nierówność 7.27, która zgodnie z obliczeniami powinna gwarantować stabilną koegzystencję replikaz i pasożytów. Wniosek ten został dodatkowo potwierdzony przez autora symulacyjnie. W ogólnym przypadku jednak nierówność ta nie jest spełniona, co pokazały np. symulacje wieloagentowe [105, 110]. Konieczne było uwzględnienie tego faktu, np. przez wprowadzenie mutacji pasożytów - głównego czynnika pozwalającego na ich większe powinowactwo do replikazy, a który musiał wystąpić w rzeczywistości ze względu na

prymitywną, podatną na błędy maszynę replikacyjną w Świecie RNA. W toku badań okazało się, że populacje rozwijają się w sposób cykliczny i jedną z faz jest stan w którym koegzystencja gwarantowana przez nierówność 7.27 jest możliwa. Pozwala to na swobodną ekspansję i kolonizację pustej przestrzeni wokół. Lokalnie jednak dobór naturalny powoduje powstawanie coraz łatwiej kopiowalnych pasożytów, a to prowadzi do wypierania replikaz i śmierci populacji (nierówność 7.27 przestaje być spełniona). Ze względu jednak na losowy charakter mutacji oraz dyfuzji - mogą ostać się miejsca pozbawione wydajnych pasożytów, spełniają one warunek koegzystencji co powoduje odrodzenie się populacji i w ten sposób cykl wraca do początku.

Model wykorzystujący ewolucję syntaz

Modele Świata RNA oparte na RP opisane we wcześniejszych rozdziałach nie przewidywały żadnej konkretnej funkcji dla pasożytów poza ewentualną kontrolą populacji replikaz. W celu zbadania możliwości nabywania przez pasożyty użytecznych dla populacji funkcji, wprowadzono dodatkowy typ enzymu - syntazy, które mają zdolność syntezy poszczególnych rybonukleotydów.

8.1 Model biologiczny

W komórkach biologicznych rybonukleotydy są syntezowane przy pomocy odpowiednich enzymów [83]. W Świecie RNA jednak musiało to mieć miejsce samoczynnie, gdyż na samym początku nie istniały jeszcze funkcjonalne enzymy. Model oparty na syntazach skupia się na przejściu z syntezy abiotycznej nukleotydów do znacznie wydajniejszej, katalizowanej przez odpowiednie cząsteczki - w tym przypadku RNA.

Łańcuchy RNA w rozważanej populacji nie mają z góry narzuconych ról, decyduje o nich struktura drugorzędowa przewidywana za pomocą pakietu ViennaRNA [44]. Z tego powodu sekwencja każdej cząsteczki musi być reprezentowana *explicite*. Należy tutaj dodać, że poszczególne funkcje się nie wykluczają, RNA może katalizować replikację i przy okazji pomagać w syntezie nukleotydów. Na szybkość replikacji w modelu mają wpływ następujące czynniki:

1. produkcja nukleotydów - jest to suma nukleotydów powstających w wyniku procesów abiotycznych oraz tych produkowanych przez syntazy. Wartość ta liczona jest oddzielnie dla każdego z 4 rybonukleotydów.
2. skład nukleotydowy replikowanej cząsteczki - określa on zapotrzebowanie na poszczególne nukleotydy w czasie replikacji,
3. aktywność enzymatyczna replikazy a - zdolność replikazy do katalizowania replikacji wynikająca z jej struktury drugorzędowej,
4. dopasowanie znacznika - sekwencja przy końcu 3' wzorca musi pasować (przez komplementarność) do końca 5' replikazy. Potencjalna rola tego typu znaczników w RNA World została omówiona w [131].

Kolejną ważną modyfikacją w stosunku do poprzednich rozdziałów jest wprowadzenie kompartmentów, które zachowują się jak protokomórki. Rywalizacja o przetrwanie nie odbywa się między poszczególnymi cząsteczkami, ale między całymi populacjami. Podobna koncepcja została wprowadzona w pracy Takeuchiego i Hogeweg [114]. Takie podejście wprowadza wyższy poziom selekcji - „selekcję wielopoziomową typu 2” (MSL2) [36]. Z założenia lepiej przystosowane populacje, jeśli osiągną wystarczającą liczebność - dzielą się na dwie mniejsze populacje potomne.

Prawdopodobieństwa poszczególnych mutacji były sprawdzane co nukleotydu i na danej pozycji wzajemnie się wykluczały. Należy tutaj dodać, że substytucja obejmowała zastąpienie danego nukleotydu nowym, całkowicie losowym, mógł to zatem być ten sam.

8.2.3 Ocena populacji

Jak napisano wcześniej, symulacja opierała się na rywalizujących między sobą populacjach. Konkurowały one o przestrzeń reprezentowaną tutaj jako tablica jednowymiarowa w której jedna komórka zawierała jedno RNA, do każdej populacji należał zdefiniowany ciągły przedział tejże tablicy. Co pokolenie każda populacja dostawała względnie tyle komórek ile wynikało z jej oceny podzielonej przez łączną ocenę wszystkich populacji obecnych w symulacji. Przykładowo jeśli dana populacja miała ocenę 100, a suma wszystkich ocen wynosiła 200 to w następnym pokoleniu populacja ta otrzymywała 50% wszystkich komórek w tablicy. W praktyce żaden z kompartmentów nie osiągał takiej wielkości, ponieważ wcześniej podzieliliby się na mniejsze kompartmenty potomne.

Wyznaczanie oceny rozpoczynało się od obliczenia kilku ważnych statystyk:

1. ile razy nukleotyd X łącznie występuje we wszystkich sekwencjach w populacji (n_X),
2. suma aktywności enzymatycznych wszystkich syntaz nukleotydu X (K_X),
3. łączna aktywność replikacyjna (R), jest to suma wartości wyliczanych dla każdej pary cząsteczek (również takich gdzie dwa razy występuje to samo RNA), przy czym kolejność w parze ma znaczenie, dlatego dla dwóch różnych cząsteczek i oraz j zarówno (i, j) jak (j, i) będą rozważane. Dla pary (i, j) jest brana pod uwagę aktywność replikacyjna cząsteczki i , która następnie jest dzielona przez dwa razy tyle na ilu pozycjach znacznik j (ostatnie 6 nukleotydów czytanych od tyłu) nie jest komplementarne do znacznika wiążącego i (pierwsze 6 nukleotydów).
4. wielkość populacji (S).

Następnie powyższe wartości były podstawiane do odpowiedniego wzoru na ocenę (M):

$$M = \frac{R \prod_i (1 + K_i)}{S^2 \prod_i (1 + n_i)} \quad (8.6)$$

S^2 w mianowniku ma na celu znormalizowanie R , ponieważ jak łatwo sprawdzić - R skaluje się z kwadratem wielkości populacji.

8.2.4 Generowanie losowych sekwencji

Sekwencje generowane były jako ciąg losowych nukleotydów, przy czym po wygenerowaniu każdego nukleotydu było 1% szans na zakończenie sekwencji. Dzięki temu tak utworzone sekwencje miały różną długość, ale było to średnio 100 nukleotydów.

8.3 Algorytm

Większość procedur opisano już wcześniej, tutaj warto dodać, że dla przyspieszenia wykonania niektóre pętle były uruchamiane wielowątkowo (bez zmiany implementacji). Pseudokod przedstawiony jest poniżej (algorytm 7).

Algorytm 7 Główna pętla programu

```

Utwórz tablicę  $M$  komórek -  $tab$ 
Inicjalizuj losowe populacje, dzieląc dostępną przestrzeń po równo (każda komórka wypełniana
jest losową sekwencją RNA).
for  $step \leftarrow 1$  to  $max\_steps$  do
  Oceń istniejące populacje
  Stwórz pustą tablicę  $new\_tab$  o identycznej wielkości co  $tab$ 
  Wyznacz liczbę komórek, która przypadnie każdej populacji w następnym pokoleniu i podziel
  odpowiednio tablicę  $new\_tab$ .
  for all populacji  $P$  do
    Przestrzeń przypadającą  $P$  wypełnij w ten sposób, że każda nowa komórka będzie kopią
    losowego RNA z tej populacji poddaną mutacji (jedno RNA może być skopiowane wiele razy).
    if Liczba komórek przypadająca populacji  $P > 200$  then
      Podziel  $P$  na dwie populacje potomne tak, że pierwsza stanowi pierwszą połowę przedziału
      w tablicy  $new\_tab$ , a druga - drugą.
    end if
  end for
  Zamień tablicę  $tab$  z  $new\_tab$ .
end for

```

8.4 Implementacja

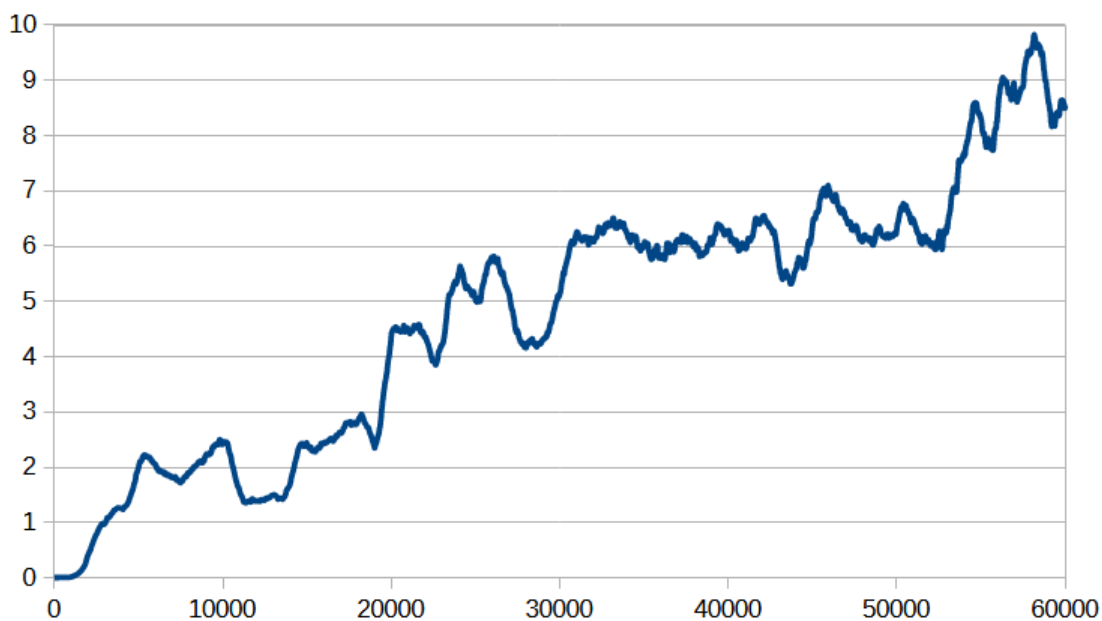
Całość zaimplementowano w języku C++ w systemie operacyjnym Linux. Do łatwego zrównoleglania niektórych fragmentów użyto API OpenMP [6]. Projekt budowany był za pomocą pliku Makefile i uruchamiany z konsoli.

8.5 Wyniki

Symulację wykonano dla 1000 komórek rozdzielonych początkowo między 10 populacji, trwała ona 65.000 pokoleń. Ewolucję średniej sumy ocen populacji w jednej z symulacji przedstawiono na rys. 8.1. Od pewnego momentu wszystkie populacje są potomkami tylko jednej z początkowych dziesięciu. Najważniejszą obserwacją był fakt, że w populacjach obecny był tylko jeden rodzaj RNA, posiadający wszystkie opisane wcześniej funkcje. Wielofunkcyjność została osiągnięta kosztem znacznie gorszych zdolności katalitycznych, co widać na rys. 8.2, na którym struktura sekwencji wyewoluowanej w symulacji została porównana z wzorcami. Połączenie różnych rodzajów enzymów w jeden pozwoliło prawdopodobnie na obejście progu Eigena - ilość informacji genetycznej, która musiała być utrzymywana w systemie była trzykrotnie mniejsza.

8.6 Dyskusja

Synteza nukleotydów jest jednym z ważniejszych tematów rozważanych w Świecie RNA, z tego powodu poszukiwane są ścieżki reakcji pozwalające na ich spontaniczne powstanie, np. w [83] oraz polimeryzację bez udziału enzymów, np. w [93]. W pewnym momencie jednak musiały powstać cząsteczki zdolne do przyspieszenia tego procesu, ponieważ rybonukleotydy były potrzebne do coraz większej liczby replikacji w rozrastającej się populacji, a także przez wzgląd na fakt, że tak odbywa się to współcześnie [83]. Dość dobrymi kandydatami okazały się pasożyty, którym już Takeuchi i Hogeweg przypisali konkretną funkcję - synteza substancji budujących błonę okalającą kompartmenty [114]. Warto tutaj też wspomnieć, że kompartmenty były wcześniej badane przez zespół Jacka Szostaka [50].



Rys. 8.1: Suma ocen populacji w symulacji w każdym kroku. Ze względu na duże wahania wartości - zastosowano średnią ruchomą z 1000 kroków. Wykres dotyczy tej samej symulacji z której otrzymano tylko jedną sekwencję.

```

AUACCUUUGUUAUAUCUUUGUCUAUUCACAUUGGCCAUCCAGUUGGCCGAAUCCUGGCUAAAUAUCAUCGCGAGGCAUCCCAUCCGCAAGCUGUUGUCUGUAGCCUAGUCCUAGGUGCCGAAACCUUAUUGUAUUCGAGAUCCGAGGUCAGAUAGGAU
...((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
(((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((
(((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((((

```

Rys. 8.2: Sekwencja, która wyewoluowała w systemie. Pełniła ona wszystkie trzy zdefiniowane w systemie role, chociaż w żadnej nie była efektywna. W pierwszym wierszu podana jest sekwencja, w drugim jej struktura drugorzędowa. Trzeci wiersz to dopasowana struktura wzorcowej replikazy, czwarty i piąty wiersz to odpowiednio wzorcowa syntaza A i G i wzorcowa syntaza C i U. Różnice w strukturze zaznaczono na czerwono.

Warto wspomnieć, że w przeszłości był już rozważany model Świata RNA w którym syntazy powstały przed replikazami [73]. Autorzy argumentują, że funkcjonalna replikaza RNA musiała być dość długa (ok. 160-200 nukleotydów), a replikacja mogła zachodzić z wystarczającą skutecznością nieenzymatycznie. Wysłanęli oni hipotezę, że pierwszymi aktywnymi katalitycznie cząsteczkami mogły być te zdolne do syntezy rybonukleotydów. Za pomocą symulacji metodą Monte Carlo na dwuwymiarowej siatce pokazali oni, że syntazy mogą bardzo łatwo rozprzestrzenić się w całym badanym obszarze, ponieważ dają one wyraźną przewagę ewolucyjną populacjom RNA.

W [63] do analizy możliwości współistnienia różnych enzymów RNA zastosowano autorski, rozwijany od lat model - „Metabolicznie Sprzężony System Replikatorowy” (z ang. MCRS) jako część symulacji przeprowadzonej na dwuwymiarowej siatce reprezentującej powierzchnię mineralną. W tej wersji MCRS trzy enzymy wspólnie pozwalały na syntezę wszystkich czterech rybonukleotydów koniecznych do syntezy RNA. Aktywność enzymatyczna wymagała miejsca aktywnego o odpowiedniej sekwencji i strukturze drugorzędowej, jedna cząsteczka mogła pełnić więcej niż jedną funkcję, ale taka wszechstronność oznaczała niższą efektywność, a w danym momencie enzym mógł katalizować reakcję tylko jednego typu. Długość cząsteczek była ograniczona przez czas replikacji - duża liczba nukleotydów oznaczała dłuższy czas kopiowania co było niekorzystne ewolucyjnie. Wyniki okazały się nie zależeć od zmiany parametrów (z wyjątkiem ekstremalnych ustawień). W toku ewolucji enzymy wielofunkcyjne praktycznie zawsze znikaly, a każda cząsteczka posiadała tylko jeden rodzaj aktywności katalitycznej. Jest to rezultat na pozór sprzeczny z otrzymanym w tym rozdziale, ale w omawianej publikacji, ze względu na małą długość miejsc aktywnych, współistnienie trzech różnych enzymów nie powodowało przekroczenia progu Eigena. Z tego powodu wyspecjalizowanie mogło okazać się w takich warunkach korzystniejsze niż wielofunkcyjność.

Kim i Higgs [62] zaproponowali podejście w którym replikazy i syntazy koegzystują. Wykorzystali oni dwuwymiarowy automat komórkowy w którym dopuścili 7 stanów każdego pola: puste, polimeraza, nić komplementarna do polimerazy, syntaza, nić komplementarna do syntazy, niefunkcjonalna polimeraza (mutant) oraz niefunkcjonalna syntaza (mutant). Kopiowanie nici prowadzi do powstania nici komplementarnej (z wyjątkiem mutantów, gdzie nie wprowadza się tego rozróżnienia). Dla każdego typu RNA narzuca się z góry długość oraz prawdopodobieństwo mutacji. Rolą syntaz w modelu jest produkcja monomerów, które stanowią zasób konieczny do replikacji, innym ich źródłem jest rozpad RNA. Najważniejszym rezultatem otrzymanym przez autorów są powstające samoczynnie struktury przestrzenne - zgrupowania replikaz (polimeraz) otoczone syntazami. W przeciwieństwie do badań przeprowadzanych w tym rozdziale - nukleotydy nie są rozróżniane, a sekwencja nie jest reprezentowana *explicite*.

Model wykorzystujący inhibitory kontrolujące populację

9.1 Model biologiczny

Wraz z ewolucją Świata RNA na pewnym etapie musiało dojść do powstania kompartmentów - fragmentów przestrzeni otoczonych błoną lipidową, która stanowiła barierę dla cząsteczek RNA [120]. W toku ewolucji nośnikiem informacji stało się DNA, co wymagało wykształcenia całego szeregu mechanizmów, które umożliwiły jego syntezę oraz efektywny odczyt. Skomplikowane mechanizmy biologiczne wymagały jakiegoś sposobu regulacji, który utrzymałby poprawne stężenie poszczególnych substancji aktywnych. Teoria opisana w tym rozdziale tłumaczy w jaki sposób regulacja mogła odbywać się na bardzo wczesnym etapie, czyli w kompartmentach zawierających roztwór cząsteczek RNA oraz zasobów potrzebnych do ich powielania.

Założenia teorii można streścić następująco:

1. Jednym z etapów tworzenia się Świata RNA była tranzycja $\beta \rightarrow \gamma$, czyli spontaniczne łączenie się pojedynczych rybonukleotydów w łańcuchy [110]. Oprócz funkcjonalnych molekuł, które później przeszły samoorganizację ($\gamma \rightarrow \delta$), najpewniej powstawała też duża ilość krótszych cząsteczek RNA, które nie pełniły na początku żadnej konkretnej funkcji. Ich powstawanie można wyjaśnić chociażby niedoskonałością nieenzymatycznej polimeryzacji, która łatwo mogła zostać zakłócona na wczesnym etapie.
2. Krótkie cząsteczki RNA mogą łączyć się z odsłoniętymi fragmentami dłuższych na zasadzie komplementarności.
3. Cząsteczka RNA z doczepionym krótszym fragmentem nie może zostać powielona, ponieważ replikaza nie ma dostępu do sekwencji. Krótkie cząsteczki blokujące dłuższe nazywane są „inhibitorami”.
4. Inhibitory mogą też łączyć się z komplementarnymi, krótkimi fragmentami RNA - „antyinhibitorami”.
5. Opisane wyżej reakcje chemiczne są odwracalne, co znaczy, że mogą biec w obu kierunkach.

Przesłankami co do prawdziwości tej teorii są następujące fakty:

1. Współczesne komórki posiadają cały szereg mechanizmów, które kontrolują wszystkie procesy - od ekspresji genów po stężenie poszczególnych jonów w cytoplazmie. Mechanizmy te musiały mieć gdzieś swój początek.
2. W mRNA występują regiony zwane „ryboprzełącznikami”, do których mogą przyłączać się inne cząsteczki i w ten sposób regulować ekspresję. Co więcej „*Charakterystyka niektórych ry-*

boprzełączników sugeruje, że mogą być one współczesnymi wersjami starodawnych systemów sensorycznych i regulatorowych, które prawdopodobnie funkcjonowały przed pojawieniem się enzymów oraz czynników genetycznych bazujących na białkach (Nahvi et al. 2002; Vitreschak et al. 2004; Breaker 2006). Jeśli jest to prawdą, niektóre struktury i funkcje rybo-przełączników, które tak dobrze służą współczesnym komórkom, mogą wiernie odzwierciedlać możliwości sensorów i przełączników RNA, które istniały w Świecie RNA.” [22].

9.2 Model matematyczny

W celu zbadania potencjalnych właściwości omawianego systemu, najpierw rozważymy prostszy przypadek bez antyinhibitora, a dopiero potem wprowadzimy antyinhibitory.

9.2.1 Agent i inhibitor

Rozważmy roztwór w którym obecne są agent (A) oraz inhibitor (X), które mogą tworzyć nieaktywny kompleks:



gdzie k_1 i k_2 to (dodatnie) stałe szybkości reakcji w odpowiednich kierunkach. Jeśli przyjmiemy, że stężenia poszczególnych reagentów zależą od czasu - obowiązują następujące równania różniczkowe (zgodnie z prawem działania mas):

$$\frac{d[A]}{dt} = k_2[AX] - k_1[A][X] \quad (9.2)$$

$$\frac{d[X]}{dt} = -\frac{d[A]}{dt} \quad (9.3)$$

$$\frac{d[AX]}{dt} = \frac{d[A]}{dt} \quad (9.4)$$

Jeśli czynnik po prawej stronie równania 9.2 jest dodatni, stężenia $[A]$ i $[X]$ będą wzrastać, a stężenie $[AX]$ maleć. Oznacza to, że wartość wspomnianego wcześniej czynnika będzie spadać. W przeciwnym wypadku, jeśli jest ujemny, jego wartość będzie rosła. Prowadzi to do wniosku, że system zawsze dąży do stanu, w którym prawa strona równania 9.2 jest równa 0, czyli do równowagi:

$$k_2[AX] - k_1[A][X] = 0 \quad (9.5)$$

$$\frac{k_2}{k_1}[AX] = [A][X] \quad (9.6)$$

Od tego momentu traktujemy stężenia $[A]$, $[X]$ oraz $[AX]$ nie jako funkcje czasu, ale właśnie jako odpowiednie stężenia w stanie równowagi. Oznaczmy stałą $\frac{k_2}{k_1}$ jako K - stałą równowagi reakcji, co prowadzi ostatecznie do równania:

$$K[AX] = [A][X] \quad (9.7)$$

W celu badania właściwości omawianego systemu wygodnie jest wprowadzić dwie dodatnie stałe: c_A (całkowite stężenie agenta w roztworze) oraz c_X (całkowite stężenie inhibitora w roztworze):

$$c_A = [A] + [AX] \quad (9.8)$$

$$c_X = [X] + [AX] \quad (9.9)$$

Po przekształceniu:

$$[AX] = c_A - [A] \quad (9.10)$$

$$[X] = c_X - c_A + [A] \quad (9.11)$$

Podstawiając do równania (9.7) i stosując oznaczenie $x = [A]$:

$$K(c_A - x) = x(c_X - c_A + x) \quad (9.12)$$

$$x^2 + (c_X - c_A + K)x - Kc_A = 0 \quad (9.13)$$

W celu zbadania potencjalnych rozwiązań powyższego równania, można zdefiniować lewą stronę jako funkcję $f(x)$:

$$f(x) = x^2 + (c_X - c_A + K)x - Kc_A \quad (9.14)$$

Wykresem funkcji $f(x)$ jest parabola skierowana ramionami do góry (współczynnik przy x^2 jest dodatni), łatwo można wykazać, że:

$$f(0) = -Kc_A < 0 \quad (9.15)$$

$$f(c_A) = c_X c_A > 0 \quad (9.16)$$

Powyższe nierówności oznaczają, że 0 znajduje się pomiędzy miejscami zerowymi oraz, że dodatnie miejsce zerowe jest mniejsze od c_A . Można z tego wywnioskować, że istnieje dokładnie jedno x spełniające równanie (9.13) i położone w przedziale $(0; c_A)$. Łatwo udowodnić, że odpowiednie $[AX]$ oraz $[X]$ zawsze będą dodatnie. Rozwiązanie można traktować w takim razie jako funkcję parametrów c_A , c_X i K , można je też wyznaczyć w sposób analityczny - metodą delty:

$$\Delta = (c_X - c_A + K)^2 + 4Kc_A > 0 \quad (9.17)$$

$$x = \frac{-(c_X - c_A + K) \pm \sqrt{\Delta}}{2} \quad (9.18)$$

Pierwiastkiem leżącym w oczekiwanym przedziale jest większe z rozwiązań, zatem przy pierwiastku z delty powinien być znak +.

$$x = \frac{-(c_X - c_A + K) + \sqrt{\Delta}}{2} \quad (9.19)$$

Dla danego rozwiązania można zdefiniować pojemność roztworu (analogicznie do roztworów buforowych), która mówi jak mała zmiana c_A wpłynie na wartość x :

$$\beta = 1 / \frac{\partial x}{\partial c_A} \quad (9.20)$$

Można oczywiście do powyższej równości podstawić jawny wzór na x , ale prowadzi to do dość skomplikowanego wzoru i przez to nie widać najważniejszych własności β . Zastosujemy inną strategię i zaczniemy od zróżniczkowania równania 9.13 po c_A :

$$2x \frac{\partial x}{\partial c_A} + (c_X - c_A + K) \frac{\partial x}{\partial c_A} - x - K = 0 \quad (9.21)$$

$$\beta = 1 / \frac{\partial x}{\partial c_A} = \frac{2x + c_X - c_A + K}{x + K} \quad (9.22)$$

Mnożąc licznik i mianownik przez x :

$$\beta = \frac{2x^2 + c_X x - c_A x + Kx}{x^2 + Kx} \quad (9.23)$$

Z równania 9.13 wynika, że:

$$c_X x = -x^2 + (c_A - K)x + Kc_A \quad (9.24)$$

Podstawiając powyższą zależność do równości 9.23:

$$\beta = \frac{x^2 + Kc_A}{x^2 + Kx} \quad (9.25)$$

Dla ustalonego K i c_A wszystkie wartości x między 0 i c_A są osiągalne, wystarczy odpowiednio dobrać c_X . Dowód tego faktu wynika z równania 9.24:

$$c_X x = -x^2 + (c_A - K)x + Kc_A = (c_A - x)(x + K) \quad (9.26)$$

$$c_X = \frac{(c_A - x)(x + K)}{x} > 0 \quad (9.27)$$

Odpowiedni parametr c_X zawsze ma wartość dodatnią, czyli spełnia założenia modelu.

Wracając do wzoru 9.25 - z poprzednich założeń wynika, że $x < c_A$, czyli licznik jest większy od mianownika, co daje ograniczenie dolne na β - 1. Ograniczeniem górnym jest c_A/x , co można łatwo udowodnić:

$$\beta = \frac{c_A}{x} \frac{x \frac{x}{c_A} + K}{x + K} \quad (9.28)$$

Jako, że $x/c_A < 1$, licznik prawego czynnika jest mniejszy od jego mianownika, co kończy dowód. Ostatecznie możemy zapisać:

$$1 < \beta < \frac{c_A}{x} \quad (9.29)$$

Największe wartości β można uzyskać dla $x \ll c_A$, czego można dowieść za pomocą odpowiedniej granicy wyrażenia 9.25 dla ustalonego K i c_A :

$$\lim_{x \rightarrow 0} \beta = \lim_{x \rightarrow 0} \frac{x^2 + Kc_A}{x^2 + Kx} = \infty \quad (9.30)$$

Dowolnie małe dodatnie x jest osiągalne, co udowodniono wcześniej (wystarczy odpowiednio dobrać c_X).

9.2.2 Agent, inhibitor i antyinhibitor

W tej wersji modelu obecny jest agent (A), inhibitor (X) oraz antyinhibitor (Y). Zachodzą dwie reakcje (w obu kierunkach):





Odpowiednie równania różniczkowe dla poszczególnych stężeń mają postać:

$$\frac{d[A]}{dt} = k_2[AX] - k_1[A][X] \quad (9.33)$$

$$\frac{d[AX]}{dt} = -\frac{d[A]}{dt} \quad (9.34)$$

$$\frac{d[Y]}{dt} = k_4[XY] - k_3[X][Y] \quad (9.35)$$

$$\frac{d[XY]}{dt} = -\frac{d[Y]}{dt} \quad (9.36)$$

$$\frac{d[X]}{dt} = k_2[AX] - k_1[A][X] + k_4[XY] - k_3[X][Y] \quad (9.37)$$

Podobnie jak w uproszczonej wersji modelu, wygodnie jest wprowadzić dodatnie stałe będące całkowitymi stężeniami poszczególnych reagentów:

$$c_A = [A] + [AX] \quad (9.38)$$

$$c_X = [X] + [AX] + [XY] \quad (9.39)$$

$$c_Y = [Y] + [XY] \quad (9.40)$$

Analogicznie jak poprzednio definiujemy odpowiednie równania stanu równowagi (bazując na równaniach 9.33, 9.35, 9.36 oraz 9.37):

$$K[AX] = [A][X] \quad (9.41)$$

$$k[XY] = [X][Y] \quad (9.42)$$

gdzie $K = \frac{k_2}{k_1}$ oraz $k = \frac{k_4}{k_3}$.

Stabilność

Układ równań różniczkowych 9.33, 9.35, 9.36 oraz 9.37 na pierwszy rzut oka wydaje się dość skomplikowany i nie jest jasne, czy system dąży do jakiegokolwiek stanu równowagi ani czy taki stan istnieje. W tym podrozdziale zakładamy jego istnienie (fakt, który zostanie udowodniony w następnym), co pozwoli nam udowodnić, że układ zawsze do niego dąży.

Stan jest reprezentowany przez 5 zmiennych: $[A]$, $[X]$, $[Y]$, $[AX]$ i $[XY]$. Pierwsza rzecz, którą należy zbadać to czy wartości zmiennych zawsze pozostają w dozwolonym zakresie - wszystkie muszą być nieujemne, co wynika z definicji problemu. Wiadomo, że są to funkcje różniczkowalne, więc są one także ciągłe. Oznacza to, że jeśli zmienna osiąga kiedykolwiek wartość ujemną - musi w pewnym momencie mieć wartość 0. Uwzględniając równania na stałe c_A , c_X i c_Y , łatwo zauważyć, że pochodna po czasie zmiennej o wartości 0 jest nieujemna. Przykładowo, jeśli $[A] = 0$ to $[AX] = c_A$, czyli prawa strona równania 9.33 ma wartość nieujemną. Podobnie można uzasadnić ten fakt dla pozostałych zmiennych, ponieważ w każdym z równań na pochodną, czynnik ujemny jest proporcjonalny do wartości samej zmiennej.

Przed przystąpieniem do dalszych rozważań warto udowodnić, że wszystkie stężenia w stanie równowagi (na razie tylko zakładamy jego istnienie) są dodatnie. Istotnie, jeśli którekolwiek byłyby równe 0, oznaczałoby to, że jedno z równań 9.41 i 9.42 również musiałoby mieć 0 po obu stronach. Jednak wtedy c_A , c_X lub c_Y także byłyby równe 0, co przeczy założeniu, że całkowite stężenia są dodatnie.

Jak wspomniano wcześniej - stan układu jest opisywany przez 5 zmiennych, jednak nie są to zmienne niezależne od siebie, wiążą je ze sobą przede wszystkim wykorzystane już wcześniej równania na stałe c_A , c_X oraz c_Y . Z ich przekształcenia można łatwo otrzymać:

$$[AX] = c_A - [A] \quad (9.43)$$

$$[XY] = c_Y - [Y] \quad (9.44)$$

$$[X] = c_X - c_A + [A] - c_Y + [Y] \quad (9.45)$$

Wynika z tego, że wartości zmiennych $[A]$ oraz $[Y]$ wystarczą, żeby w pełni określić stan układu. Dla wygody zastosujemy oznaczenia:

$$x + x_0 = [A] \quad (9.46)$$

$$y + y_0 = [Y] \quad (9.47)$$

Gdzie x_0 oraz y_0 to stężenia $[A]$ oraz $[Y]$ w wybranym stanie równowagi (na tym etapie nie wiem czy zawsze istnieje tylko jeden).

Podstawiając do równań różniczkowych 9.33 i 9.35, równania 9.43, 9.44, 9.45, 9.46 oraz 9.47, mamy układ:

$$\dot{x} = k_2(c_A - x - x_0) - k_1(x + x_0)(c_X - c_A + x + x_0 - c_Y + y + y_0) \quad (9.48)$$

$$\dot{y} = k_4(c_Y - y - y_0) - k_3(y + y_0)(c_X - c_A + x + x_0 - c_Y + y + y_0) \quad (9.49)$$

Z definicji stanu równowagi wynika, że prawe strony obu powyższych równań powinny być równe 0 dla x i y równych 0. Powyższy układ można zapisać też jako (wykorzystujemy tutaj równanie 9.45 dla skrócenia zapisu):

$$\dot{x} = k_2(c_A - x_0) - k_1(x_0)(c_X - c_A + x_0 - c_Y + y_0) - k_2x - k_1x_0(x + y) - k_1x[X] \quad (9.50)$$

$$\dot{y} = k_4(c_Y - y_0) - k_3(y_0)(c_X - c_A + x_0 - c_Y + y_0) - k_4y - k_3y_0(x + y) - k_3y[X] \quad (9.51)$$

Korzystając z własności stanu równowagi (równania 9.41 i 9.42) można wyeliminować dwa pierwsze składniki w każdym z obu równań (są one równe 0):

$$\dot{x} = -k_2x - k_1x_0(x + y) - k_1x[X] \quad (9.52)$$

$$\dot{y} = -k_4y - k_3y_0(x + y) - k_3y[X] \quad (9.53)$$

Posłużymy się teraz metodą Lyapunova - zdefiniujmy teraz pewną funkcję rzeczywistą V , którą wykorzystamy w dalszej części:

$$V(x, y) = k_3y_0x^2 + k_1x_0y^2 \quad (9.54)$$

Jest ona dodatnia poza punktem równowagi $(0, 0)$, ponieważ $x_0, y_0 > 0$, co uzasadniono wcześniej. Następnym krokiem jest wyznaczenie gradientu $V(x, y)$:

$$\frac{\partial V}{\partial x} = 2k_3y_0x \quad (9.55)$$

$$\frac{\partial V}{\partial y} = 2k_1x_0y \quad (9.56)$$

Funkcję $V(x, y)$ można również traktować jako funkcję stanu układu, jej pochodna zupełna po czasie jest równa:

$$\frac{dV}{dt} = \frac{\partial V}{\partial x} \dot{x} + \frac{\partial V}{\partial y} \dot{y} \quad (9.57)$$

Podstawiając otrzymujemy:

$$\frac{dV}{dt} = -2k_2k_3y_0x^2 - 2k_1k_4x_0y^2 - 2k_1k_3x_0y_0(x+y)^2 - 2k_1k_3(y_0x^2 + x_0y^2)[X] \quad (9.58)$$

Wszystkie składniki równania są niedodatnie, pochodna zupełna jest ujemna wszędzie poza stanem równowagi, gdzie $x, y = 0$. Oznacza to, że system zawsze dąży do minimalizacji $V(x, y)$ dopóki nie zostanie osiągnięta minimalna wartość $V(0, 0) = 0$. Dowodzi to, że jeśli istnieje punkt równowagi - system zawsze będzie do niego dążył, wynika stąd też kolejny ważny wniosek - istnieje co najwyżej jeden taki punkt.

Wyznaczanie stanu równowagi

Od tego momentu stężenia nie będą już traktowane jako funkcje zależne od czasu, ale odpowiednie stężenia w stanie równowagi. Celem tego rozdziału jest jego wyznaczenie, przy okazji udowadniając jego istnienie. Przekształcając równanie 9.38:

$$[AX] = c_A - [A] \quad (9.59)$$

i podstawiając do wzoru 9.41, dostajemy:

$$K(c_A - [A]) = [A][X] \quad (9.60)$$

Dla wygody oznaczmy $[A] = x$:

$$[X] = K \frac{c_A - x}{x} \quad (9.61)$$

Wzór 9.40 można przekształcić w następujący sposób:

$$[XY] = c_Y - [Y] \quad (9.62)$$

Podstawiając powyższą zależność oraz równość 9.59 do równania 9.39:

$$[Y] = [X] + c_A - x - c_X + c_Y \quad (9.63)$$

Wykorzystując równania 9.62 oraz 9.63 można zapisać zależność 9.42 jako:

$$k(c_X + x - [X] - c_A) = [X]([X] + c_A - x - c_X + c_Y) \quad (9.64)$$

Mnożąc obie strony przez x^2 :

$$kx(c_Xx + x^2 - [X]x - c_Ax) = [X]x([X]x + c_Ax - x^2 - c_Xx + c_Yx) \quad (9.65)$$

i podstawiając zależność 9.61 oraz porządkując, otrzymujemy:

$$(k-K)x^3 + [(k-K)(c_X + K - c_A) + K(c_A + c_Y)]x^2 + Kc_A(2K - k - c_A + c_X - c_Y)x - K^2c_A^2 = 0 \quad (9.66)$$

Ze względu na długość powyższego wzoru, wygodnie jest wprowadzić sobie oznaczenia dla

współczynników przy kolejnych potęgach x :

$$a = k - K \quad (9.67)$$

$$b = (k - K)(c_X + K - c_A) + K(c_A + c_Y) \quad (9.68)$$

$$c = Kc_A(2K - k - c_A + c_X - c_Y) \quad (9.69)$$

$$d = -K^2c_A^2 \quad (9.70)$$

Oznaczmy powyższy wielomian jako funkcję $L(x)$, stan równowagi odpowiada miejscu zerowemu w przedziale $(0, c_A)$, trzeba w takim razie sprawdzić jego istnienie. Można to zrobić korzystając z faktu, że funkcja $L(x)$ jako wielomian jest ciągła oraz:

$$L(0) = -K^2c_A^2 < 0 \quad (9.71)$$

$$L(c_A) = kc_Xc_A^2 > 0 \quad (9.72)$$

Z własności Darboux wynika, że w przedziale $(0, c_A)$ istnieje co najmniej jedno miejsce zerowe. Z dowodu na stabilność wynika, że istnieje co najwyżej jeden punkt równowagi, czyli funkcja $L(x)$ ma we wspomnianym przedziale zawsze dokładnie jedno miejsce zerowe.

Tutaj milcząco skorzystaliśmy z założenia, że każde miejsce zerowe w przedziale $(0, c_A)$ odpowiada poprawnemu stanowi równowagi. Teraz udowodnimy, że tak jest rzeczywiście. Pierwsze pytanie to czy rozwiązanie równania 9.66 nie powoduje sprzeczności w jakimkolwiek innym równaniu, którego spełnienie jest konieczne. Nasz pierwotny układ równań to równości 9.41, 9.42, 9.38, 9.39 i 9.40. Równanie 9.66 to przekształcone 9.42 po podstawieniach i wyznacza ono stężenie $[A]$. Układ pozostałych czterech równań można przekształcić tak, żeby wyznaczały pozostałe 4 zmienne ($[X]$, $[XY]$, $[AX]$ oraz $[Y]$), są to wzory 9.59, 9.61, 9.62 i 9.63. Pozostaje jeszcze kwestia znaku wszystkich zmiennych, ponieważ zgodnie z założeniami muszą być one nieujemne, a ten fakt nie został bezpośrednio uwzględniony w układzie równań. Stężenie $[A]$ jest w sposób oczywisty dodatnie i dodatkowo mniejsze od c_A , co oznacza, że $[AX]$ oraz $[X]$ również są dodatnie. $[Y]$ i $[XY]$ powinny być albo oba ujemne albo oba dodatnie ze względu na znak prawej strony równania 9.42, jednak z równości 9.40 wynika, że tylko drugi wariant jest możliwy.

Pojemność

Analogicznie do prostszej wersji modelu - definiujemy pojemność systemu. W tym celu należy x (stężenie w stanie równowagi) traktować jako funkcję pięciu parametrów: K , k , c_A , c_X oraz c_Y :

$$\beta = 1 / \frac{\partial x}{\partial c_A} \quad (9.73)$$

Podobnie jak poprzednio - wartość ta mówi jak mała zmiana c_A wpływa na zmianę x . W podobny sposób możemy też uzyskać wzór na β z równania 9.66, różniczkując obie strony po c_A :

$$(3ax^2 + 2bx + c) \frac{\partial x}{\partial c_A} - ex^2 - fx - g = 0 \quad (9.74)$$

gdzie:

$$e = -\frac{\partial b}{\partial c_A} = k - 2K \quad (9.75)$$

$$f = -\frac{\partial c}{\partial c_A} = -K(2K - k - 2c_A + c_X - c_Y) \quad (9.76)$$

$$g = -\frac{\partial d}{\partial c_A} = 2K^2c_A \quad (9.77)$$

Zatem β można zapisać jako:

$$\beta = \frac{3ax^2 + 2bx + c}{ex^2 + fx + g} \quad (9.78)$$

Mnożąc licznik i mianownik przez c_Ax :

$$\beta = \frac{c_A}{x} \frac{3ax^3 + 2bx^2 + cx}{c_Aex^2 + c_Afx + c_Ag} \quad (9.79)$$

Zmienna x określa rozwiązanie równania $L(x) = 0$, zatem można dodać do mianownika i odjąć od licznika $L(x)$:

$$\beta = \frac{c_A}{x} \frac{2ax^3 + bx^2 - d}{ax^3 + (b + c_Ae)x^2 + (c + c_Af)x + (d + c_Ag)} \quad (9.80)$$

Łatwo można się przekonać, że:

$$b + c_Ae = (k - K)(c_X + K) + Kc_Y \quad (9.81)$$

$$c + c_Af = Kc_A^2 \quad (9.82)$$

$$d + c_Ag = K^2c_A^2 \quad (9.83)$$

Czyli β można przedstawić też jako:

$$\beta = \frac{c_A}{x} \frac{2ax^3 + bx^2 + K^2c_A^2}{ax^3 + (b + c_Ae)x^2 + (c + c_Af)x + K^2c_A^2} \quad (9.84)$$

Teraz udowodnimy, że drugi czynnik jest zawsze dodatni i mniejszy od 1. Zdefiniujemy najpierw wielomian będący różnicą licznika i mianownika:

$$G(x) = ax^3 - c_Aex^2 - Kc_A^2x = (k - K)x^3 + c_A(2K - k)x^2 - Kc_A^2x \quad (9.85)$$

Można go rozłożyć:

$$G(x) = -x(c_A - x)[kx + K(c_A - x)] \quad (9.86)$$

Wartość x leży w przedziale $(0, c_A)$, zatem $G(x) < 0$. Oznacza to, że w równaniu 9.84 licznik jest zawsze mniejszy od mianownika. Ponadto licznik jest pochodną wielomianu $L(x)$ po x , jako że w miejscu zerowym następuje zmiana znaku z ujemnego na dodatni - pochodna ta musi być nieujemna. Co prawda od licznika odjęto jeszcze $L(x)$, ale jest oczywiste, że w miejscu zerowym $L(x) = 0$. Z powyższych rozważań wynika jednoznacznie, że:

$$0 \leq \beta < \frac{c_A}{x} \quad (9.87)$$

Wyznaczanie stałych dla założonego c_A , K , k oraz $[A]$

Rozważając zastosowania praktyczne modelu można spróbować rozwiązać inny problem. Dotychczas analiza skupiała się na zachowaniu systemu dla określonych stałych K , k , c_A , c_X oraz c_Y . Czy gdyby narzucić z góry oczekiwane stężenie $[A]$, czy może być ono stanem równowagi dla odpowiednich całkowitych stężeń inhibitora i anty-inhibitora (c_X i c_Y)?

Przekształcając równanie 9.66 można uzyskać:

$$Ac_X - Bc_Y = C \quad (9.88)$$

gdzie

$$A = x[kx + K(c_A - x)] \quad (9.89)$$

$$B = Kx(c_A - x) \quad (9.90)$$

$$C = (K - k)x^3 + [(K - k)(K - c_A) - Kc_A]x^2 - Kc_A(2K - k - c_A)x + K^2c_A^2 \quad (9.91)$$

Łatwo zauważyć, że wartości A i B są dodatnie. Odpowiednio dobierając c_X oraz c_Y zawsze można doprowadzić do spełnienia powyższej równości, co oznacza że można dowolnie ustalić docelowe stężenie $[A]$, a następnie dodać odpowiednią ilość inhibitora i anty-inhibitora, żeby je osiągnąć.

9.3 Algorytm

Opisany w tym rozdziale układ modelowano na trzy sposoby - bezpośrednio symulując opisane tutaj równania różniczkowe w arkuszu kalkulacyjnym, modelując je w programie w języku C++ (z wprowadzeniem replikaz) oraz za pomocą zmodyfikowanego modelu wieloagentowego z rozdziału 7. Algorytm i implementacja pierwszego i drugiego podejścia zostały opisane pokrótce w wynikach, dlatego ten i następny podrozdział zostaną poświęcone wyłącznie modelowi wieloagentowemu.

Modyfikacja wyżej wspomnianego modelu polegała na wprowadzeniu czterech nowych typów agentów:

1. inhibitora
2. antyinhibitora
3. kompleksu inhibitor-antyinhibitor
4. kompleksu pasożyt-inhibitor

Dyfuzja przebiega podobnie do pasożytów i replikaz - pierwsze dwa agenty dyfundują zgodnie z parametrem D dla pojedynczych agentów, a kompleksy zgodnie z parametrem D' . Należy tutaj wspomnieć, że inhibitory i antyinhibitory nie podlegają mutacji ani rozpadowi, ich liczba jest stała w czasie symulacji (licząc łącznie cząsteczki w stanie wolnym i skompleksowanym). Inhibitory mogą reagować z antyinhibitorami oraz z pasożytami, prawdopodobieństwa poszczególnych reakcji określane są na podstawie parametrów k_1 , k_2 , k_3 oraz k_4 (które odpowiadają analogicznym parametrom w równaniach 9.33 - 9.37):

$$p_X = 1 - e^{-k_X \Delta t} \quad (9.92)$$

W powyższym wzorze przyjęto, że reakcja jest zdarzeniem określonym rozkładem Poissona.

9.4 Implementacja

Opisanych w poprzednim podrozdziale zmian dokonano w kodzie w języku C#. Nowe typy agentów dodano jako klasy dziedziczące po szablonach odpowiednio dla pojedynczych agentów jak i dla kompleksów. Na zrzutach ekranu z symulacji inhibitory oraz antyinhibitory są rysowane podobnie jak pozostałe agenty, jednak z wyraźnie określonymi w kodzie kolorami.

9.5 Wyniki

9.5.1 Symulacja w arkuszu kalkulacyjnym

W celu zobrazowania skuteczności opisanego w tym rozdziale mechanizmu, autor zdyskretyzował równania różniczkowe (9.33 - 9.37), zamieniając infitezymalnie małe wartości dt , $d[A]$, ... na skończone wartości Δt , $\Delta[A]$, ... Zasyulowano 5000 kroków, a wszystkie wykorzystane parametry podane są w tabeli 9.1, zostały one dobrane tak, żeby osiągnąć w miarę dużą pojemność (63,86). Początkowo wszystkie cząsteczki były w stanie wolnym ($[XY] = 0$ oraz $[AX] = 0$), dlatego przez pierwsze kilkaset kroków system osiągał dopiero stan równowagi. Układ mógł być w każdym kroku zaburzony z prawdopodobieństwem 0,2%, zaburzenie polegało na zmianę wartości stężenia $[A]$ o losową wartość wybraną z przedziału $[-0,075; 0,075]$ (rozkład równomierny). Wyniki obrazuje Rys. 9.1.

9.5.2 Symulacja inhibitorów działających na replikazy (bez pasożytów)

Jednym z wykonanych testów była prosta symulacja równań różniczkowych w języku C++. Różnicą w stosunku do poprzedniego podejścia (z wykorzystaniem arkusza kalkulacyjnego) było wprowadzenie replikaz. Pasożyty nie zostały w tym podejściu uwzględnione. Równanie różniczkowe modelujące zmiany stężenia replikaz jest zmodyfikowanym równaniem 9.33, które rozszerzono o replikację, rozpad oraz losowe zaburzenia:

$$\frac{d[A]}{dt} = k_{RR}[A]^2 - d_R[A] + \sigma(t) + k_2[AX] - k_1[A][X] \quad (9.93)$$

gdzie $\sigma(t)$ jest funkcją losową, która była obliczana z wykorzystaniem zmiennej losowej ν będącej losową liczbą całkowitą z przedziału $[-5; 5]$:

$$\sigma(t) = 1000\nu[A] \quad (9.94)$$

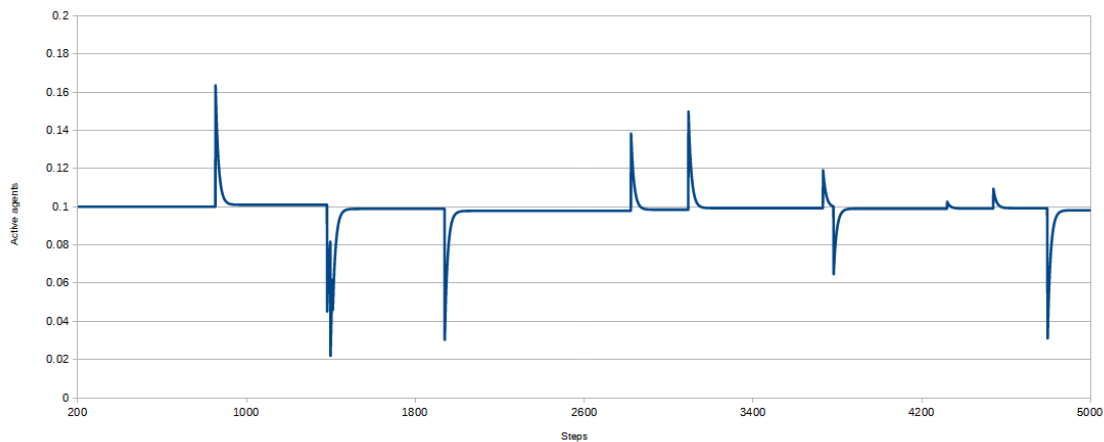
W symulacji wartość ta była wyliczana co krok, co oznaczało, że faktyczna zmiana stężenia replikaz w danym kroku mogła wynosić co najwyżej 50%. Pozostałe wykorzystane parametry podane są w tabeli 9.2. Zmiany stężenia replikaz w czasie przedstawione są na wykresie 9.2.

9.5.3 Symulacja wieloagentowa

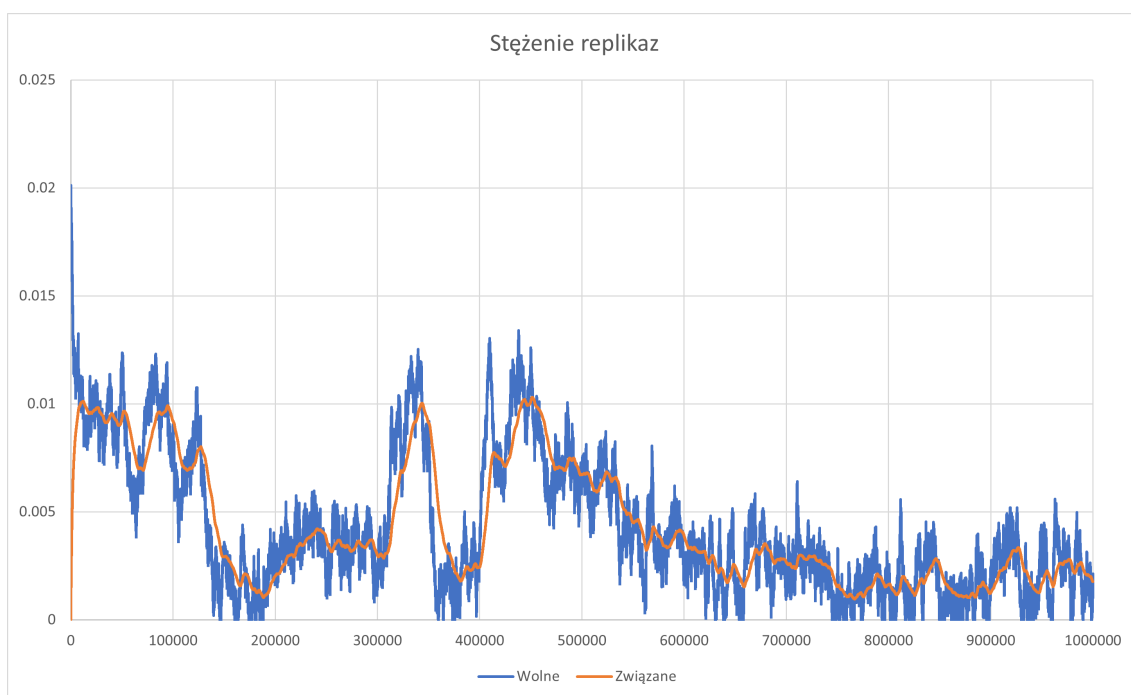
Ważnym założeniem poczynionym w tym modelu był brak reakcji replikaz na inhibitory - kontrolowały one jedynie pasożyty. Biologicznie można to uzasadnić założeniem, że spontanicznie

Parameter	Value
Δt	0.001
k_1	1
k_2	1
k_3	1
k_4	5
c_X	100
c_Y	25,85
c_A	7

Tab. 9.1: Scenarios simulated using multi-agent system.



Rys. 9.1: Wykres przedstawiający stężenie cząsteczek A w czasie. W ciągu pierwszych 200 kroków system osiągał stan równowagi, dlatego go pominięto. Wyraźnie widać, że mimo zaburzeń, stężenie zawsze wracało do pierwotnego poziomu.



Rys. 9.2: Zmiany stężenia replikaz w czasie z podziałem na te nieaktywne (związane z inhibitorem) oraz aktywne (niezwiązane). Przesymulowano łącznie milion kroków, początkowo system przechodzi przez ogromne wahania stężenia, ale po jakimś czasie osiąga względną stabilność, mimo ciągłych zaburzeń wprowadzanych przez funkcję σ .

Nazwa	Wartość domyślna	Opis
$[A]_0$	0,02	początkowe stężenie replikaz
c_X	2,5	całkowite stężenie inhibitora
$c_{\bar{X}}$	3,0	całkowite stężenie anty-inhibitora
k_{RR}	1	stała szybkości replikacji
d_R	0,01	stała rozpadu
k_1	1	stała k_1
k_2	1	stała k_2
k_3	1	stała k_3
k_4	1	stała k_4
dt	0,0001	długość kroku czasowego

Tab. 9.2: Parametry globalne w symulacji równań różniczkowych z replikazami.

powstające krótkie fragmenty RNA mają za małe stężenie i jest ono zwielokrotniane dopiero przez replikazy, jednak nie są one w stanie kopiować inhibitora, który na nie działa, ponieważ jego przyłączenie się do większego łańcucha wyłącza wszelkie funkcje.

Działanie systemu regulacji sprawdzono dla przypadku w którym przy braku inhibitorów i antyinhibitorów populacja wymierała. Parametry wykorzystane w symulacji podane są w tabeli 9.3. Symulacja pokazała, że dla tego zestawu parametrów system wymiera (rys. 9.3), jednak wprowadzenie inhibitorów (dokładnie 13,5 tys. inhibitorów oraz taką samą liczbę anty-inhibitorów) pozwalało na przetrwanie (rys. 9.4). Ciekawym zjawiskiem są wahania obserwowane na wykresach, wynikające prawdopodobnie z opóźnionej reakcji systemu kontrolującego stężenia na zmiany.

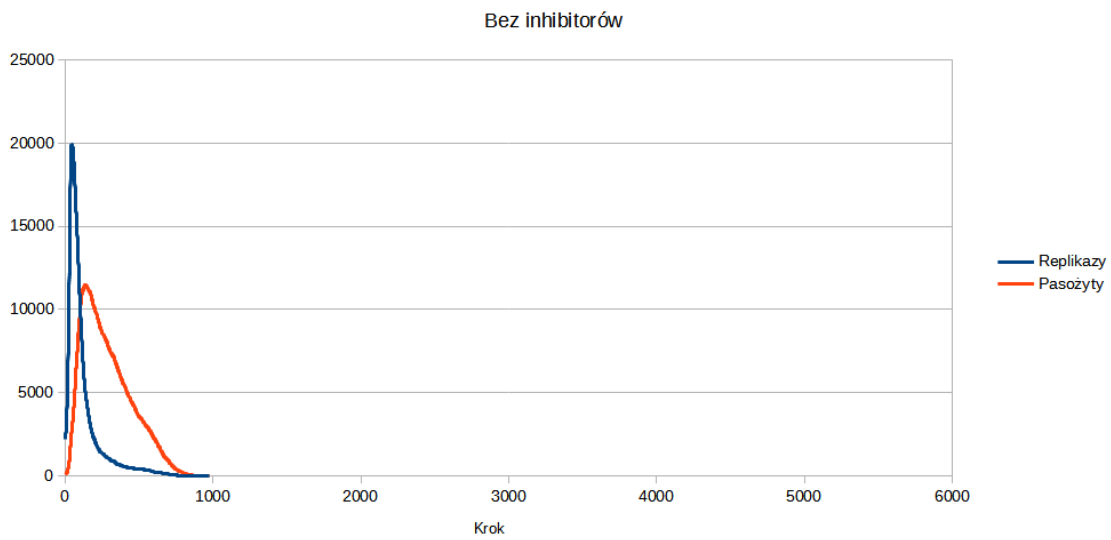
9.6 Dyskusja

Kontrola różnorodności w Świecie RNA jest bardzo istotnym problemem. Istnienie w systemie cząsteczek o różnej sekwencji może w dużej mierze zapewnić tzw. „ekologiczną stabilność”. Próbowano już pokazać w jaki sposób mogłaby ona być osiągnięta przez czynniki środowiskowe oraz modele dynamiki populacji, takie jak: równanie quasigatunku, hipercykl oraz jego wersja przestrzenna i pęcherzykowa, model replikatorów parabolicznych oraz OCF (z ang. „Open Chaotic Flow”) [108].

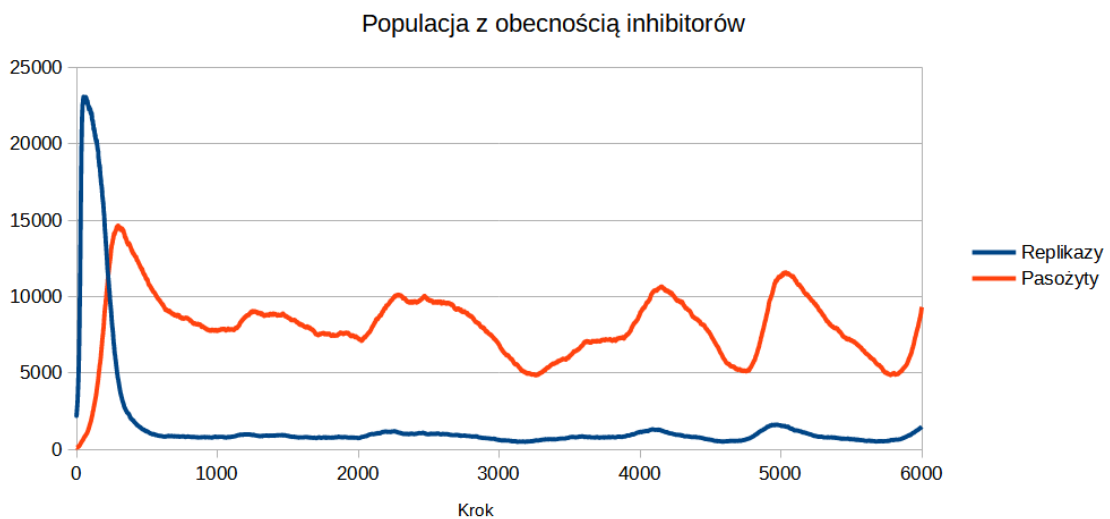
Zgodnie z wiedzą autora, modele oparte na inhibitorach nie były rozważane wcześniej w kontekście Świata RNA. Są one dodatkowym mechanizmem gwarantującym, że populacja nie zostanie zdominowana przez małą grupę najlepiej radzących sobie gatunków. Każda cząsteczka nadmiernie zwiększająca swoją liczebność jest częściej blokowana przez inhibitory, które również będą wtedy częściej uwalniane z kompleksów z anty-inhibitorami. W przeciwnym przypadku - kiedy liczebność któregoś gatunku spada - następuje uwolnienie nieaktywnych cząsteczek związanych z inhibitorem. Wyniki symulacji oraz analiza teoretyczna potwierdzają skuteczność inhibitorów w utrzymywaniu liczebności populacji na odpowiednim poziomie. Sugeruje to nowy, potencjalny szlak ewolucyjny w Świecie RNA, który może być ciekawym kierunkiem dalszych badań.

Jednym z koronnych argumentów za istnieniem tego typu mechanizmów w Świecie RNA są rybo-przełączniki. Są to regulatorowe fragmenty mRNA, które potrafią wiązać mniejsze cząsteczki i w ten sposób regulować transkrypcję i translację. Rozpoznawanie odpowiednich molekuł jest możliwe dzięki dopasowanej strukturze przestrzennej większej nici. Odkrycie sekwencji w RNA pełniących funkcje wcześniej przypisywane białkom świadczy o wszechstronności kwasów rybonukleinowych i jest kolejną przesłanką za istnieniem w przeszłości Świata RNA [22, 95].

Istnieją inne prawdopodobne pozostałości ewolucyjne po Świecie RNA. W [84] autorzy identyfikują kilka kolejnych cząsteczek i mechanizmów, które mogły zostać zakonserwowane i przetrwać



Rys. 9.3: Bez inhibitorów pasożyty szybko osiągały zbyt dużą liczebność i doprowadzały do wyginięcia replikaz i śmierci całej populacji.



Rys. 9.4: Pasożyty były szybko wiązane, co ograniczyło ich liczbę i pozwoliło replikazom osiągnąć odpowiednią liczebność. Później stężenie pasożytów było utrzymywane na stałym poziomie. Można także zaobserwować wahania liczebności populacji.

Nazwa	Wartość domyślna	Opis
$sizeX$	1050	szerokość obszaru symulacji
$sizeY$	675	wysokość obszaru symulacji
$init_R$	2250	początkowa liczba replikaz
$init_P$	112	początkowa liczba pasożytów
$agent_size$	3.0	promień koła reprezentującego agenta
N_{max}	4	maksymalna liczba sąsiadów
d	0,1	stała rozpadu
a_R	0,8	powinowactwo replikazy do replikazy
a_P	0,8	powinowactwo pasożyta do replikazy
l_P	0,2	prawdopodobieństwo przebywania pasożyta w stanie zwiniętym
Δt	1,0	czas trwania pojedynczego kroku
D	400	stała dyfuzji
D'	400	stała dyfuzji kompleksów
δ	0,1	szybkość mutacji pasożyta
dt	1	długość kroku czasowego
K	1	stała replikacji
k_1	10	stała k_1
k_2	5	stała k_2
k_3	10	stała k_3
k_4	5	stała k_4

Tab. 9.3: Parametry globalne w symulacji wieloagentowej.

wewnątrz komórek. Dużo uwagi poświęcają oni nukleotydom cyklicznym. Są to cząsteczki powstające z łatwo dostępnego budulca - trifosforanu nukleotydu (głównie adeniny i guaniny), odpowiednia reakcja powoduje, że reszta fosforanowa wiąże się jednocześnie do atomu 5' jak i 3' rybozy. Okazuje się, że takie molekuly razem z ich pochodnymi, a także cyklicznymi dinukleotydami (dwoma nukleotydami połączonymi w pętlę) są jednymi z najczęstszych mechanizmów komunikacyjnych w przyrodzie. Jeśli dodać do tego fakt, że do tego celu wykorzystywane są praktycznie tylko rybonukleotydy, można podejrzewać, że nukleotydy cykliczne wyprzedzają ewolucyjnie DNA, mając swoje początki jeszcze w Świecie RNA.

Podobnym zagadnieniem do omawianych w tym rozdziale inhibitorów są tagi, które również dotyczyłyby wybranych cząsteczek w Świecie RNA. Pełniłyby one jednak przeciwną funkcję - zamiast wiązania inhibitora potencjalnie hamującego replikację, byłyby rozpoznawane przez replikazy i pozwalałyby na rozpoczęcie procesu kopiowania. W [131] autorzy zademonstrowali z wykorzystaniem metody Monte Carlo na dwuwymiarowej siatce, że mechanizm taki może skutecznie zapobiegać wypieraniu replikaz przez pasożyty, ale tylko przy ograniczonej dyfuzji. Przy swobodnym ruchu cząsteczek powstawały tzw. super-pasożyty - krótkie sekwencje przystosowane do eksploatawania opisanego systemu, które dodatkowo szybko kolonizowały (infekowały) kolejne zgrupowania cząsteczek na planszy.

Podsumowanie

Jak wspomniano w pierwszych rozdziałach, problem początków życia nurtuje naukowców od dawna. Hipoteza RNA od swoich początków w latach 60 XX w. (choć jeszcze nie pod tą nazwą) była przedmiotem intensywnych badań, a kolejne przełomowe odkrycia w biologii (szczególnie opisanie rybozymów w latach 80) pozwoliły jej zyskać na wiarygodności [93].

Niniejszą pracę rozpoczęto od podstaw teoretycznych, rozdziału wprowadzającego do metod matematycznych i informatycznych stosowanych przez autora w badaniach. Następnie wprowadzono podstawowe pojęcia z chemii i biologii, szczególnie podstawowe równania wykorzystywane potem w przeprowadzanych analizach, takie jak wektor przemieszczenia w dynamice Browna lub równowaga reakcji chemicznych. W tym momencie można było sprawnie przejść do samego Świata RNA oraz modeli wykorzystywanych do jego badania przez ostatnie dziesięciolecie. Wprowadzono je, ponieważ projektowane w tej pracy symulacje w dużej mierze opierały się na nich.

W dalszych rozdziałach analizowane były kolejne modele i podejścia do symulacji Świata RNA. Największym problemem był brak danych na temat budowy i dokładnego mechanizmu działania replikaz ze względu na brak danych eksperymentalnych - nie udało się w laboratorium zsyntezować (albo odkryć) RNA o oczekiwanych właściwościach. Powodowało to, że ważne parametry takie jak częstość mutacji, szybkość replikacji, stała kompleksowania oraz skłonność do zwijania się RNA w struktury musiały być określone na podstawie znanych cząsteczek pełniących podobne funkcje (jak np. rybozemu F [121]) lub całkowicie arbitralnie. Podobny problem dotyczył samego środowiska - ilości rybonukleotydów i tempa ich powstawania, obecności innych substancji, które mogłyby mieć wpływ na dyfuzję, replikację i rozpad RNA. Ten ostatni problem wynika m.in. z różnych hipotez dotyczących miejsca w którym istniał Świat RNA: małe stawy lub podwodne kominy geotermalne [88]. Powyższe trudności sprawiają, że nie można traktować opisanych tutaj wyników jako ilościowych, nie taki zresztą był cel badań - cała analiza opierała się na cechach jakościowych systemów i ogólnych prawach rządzących ich ewolucją. Należy tutaj jeszcze wspomnieć, że Świat RNA wciąż pozostaje hipotezą, dlatego nowe wyniki mogą podważyć przedstawione tutaj rezultaty, opierające się (jak każde badania) na pewnych istotnych założeniach.

Pierwszą omawianą symulacją był podstawowy model wieloagentowy opublikowany w 2017r. [110]. Badania te stanowiły niejako kontynuację tych prowadzonych przez Takeuchiego i Hogeweg [114]. Głównym celem było zbadanie jak ewoluuje populacja złożona z pasożytów i replikaz RNA w warunkach ciągłej przestrzeni, bliższej rzeczywistości niż dyskretna przestrzeń w automatach komórkowych we wspomnianym modelu Takeuchiego i Hogeweg. Przede wszystkim udało się otrzymać podobny trend - pasożyty zwiększały swoje powinowactwo w stosunku do replikaz, ale jednocześnie więcej czasu spędzały w stanie uniemożliwiającym ich replikację. Zmiany te były w opozycji do siebie, pierwsza czyniła replikację pasożytów częstszą, podczas gdy druga - rzadszą. Można wysnuć teorię, że w ten sposób system dążył do pewnej równowagi, a aktywność pasożytów była ograniczona przez co nie mogły one doprowadzić do wymarcia replikaz. W symulacji podobnie jak u Takeuchiego formowały się fale, ale z uwagi na ciągłość przestrzeni miały one nieco mniej

regularny kształt. Opisywany eksperyment powtórzono, ale z małą modyfikacją (nieobecną u Takeuchiego) - replikazy również podlegały mutacji (poprzednio ich powinowactwo do innych replikaz było z góry określone dla całej symulacji). Celem było sprawdzenie, czy zmiana ta ma wpływ na zachowanie się populacji, jednak oprócz szybkiej ewolucji w kierunku większej aktywności replikaz - wyniki były identyczne jak poprzednio. Warto dodać, że początkowe ustawienie agentów miało wpływ na przeżywalność - pogrupowanie replikaz oddzielnie z pasożytami dawało większe szanse na przetrwanie systemu niż rozmieszczenie losowe.

Omówiony model zmodyfikowano, wprowadzając *explicite* sekwencje dla każdej cząsteczki. Agent nie był już charakteryzowany przez parametry, ale przez sekwencję z której je wyznaczano. Obliczenia opierały się m.in. na wyznaczaniu struktury drugorzędowej programem IPknot. Replikazy były rozpoznawane na podstawie odpowiedniego fragmentu sekwencji na końcu 5', jego uszkodzenie powodowało osłabienie lub utratę zdolności katalitycznych. Wprowadzało to nową ścieżkę reakcji - pasożyty mogły samoczynnie powstawać w wyniku błędnej replikacji replikaz. Kolejną zmianą było zrezygnowanie z powinowactwa na rzecz miary aktywności replikazy - szansa na replikację nie zależała już od wzorca, ale od enzymu. Ostatnia modyfikacja wprowadzała wymóg, żeby replikaza była zwinięta jeśli ma działać jako enzym. Pierwsza przeprowadzona symulacja polegała na rozpoczęciu procesu od idealnych replikaz i pozwoleniu pasożytom powstać samoczynnie na skutek przypadkowych mutacji. System okazał się zdolny do koegzystencji, co ciekawe replikazy zauważalnie częściej przebywały w stanie zwiniętym niż pasożyty, przez co były gorszym wzorcem, ale zwiększało to ich aktywność enzymatyczną. W drugim eksperymencie zbadano przeciwny przypadek - czy populacja replikaz była w stanie wyłonić się samoczynnie ze zbioru losowych sekwencji. Okazało się, że tak a system osiągał stan identyczny z pierwszą symulacją. Po tym postanowiono wprowadzić dodatkowy mechanizm - wykorzystując dane eksperymentalne, uzależniono czas życia RNA od sekwencji. Jednak poza zmianą częstości występowania niektórych nukleotydów (szczególnie obniżenia się zawartości uracylu), nie zmieniło to zachowania populacji. Eksperymenty zakończono analizą wpływu dyfuzji i czasu życia na ewolucję systemu. Głównym wnioskiem jest fakt, że im szybsza dyfuzja tym system jest lepiej wymieszany i bardziej podatny na wymieranie, zgodnie z wnioskami otrzymanymi z równań różniczkowych (patrz podrozdział 3.4.2).

Następnym krokiem w badaniach była próba opisu systemu RP za pomocą równań różniczkowych cząstkowych, z których dałoby się otrzymać analitycznie ogólne prawa rządzące ewolucją z uwzględnieniem aspektu przestrzennego. Przeprowadzone analizy, które zostały dodatkowo poparte symulacjami wieloagentowymi pozwoliły na zweryfikowanie wniosków otrzymanych analitycznie w praktyce. Jednym z najważniejszych rezultatów była nierówność rządząca kierunkiem rozwoju populacji i w zależności od tego czy była spełniona populacja się rozrastała lub ginęła na skutek nadmiernej dominacji pasożytów. Ostatecznie sformułowano teorię, że fale zaobserwowane w poprzednich modelach są objawem poszczególnych grup przechodzących przez kolejne fazy cyklu związanego właśnie ze wspomnianą nierównością. Grupy cząsteczek RNA najpierw rozrastają się, po czym wymierają, pozostawiając małe wysepki, które w sposób losowy przeżyły i które dają zaczątek nowym populacjom.

Pewnym pobocznym modelem rozwiniętym w międzyczasie była analiza funkcji syntaz w Świecie RNA. Celem było badanie selekcji cząsteczek RNA na wyższym poziomie (MSL2), na którym rywalizacja o zasoby odbywa się między całymi grupami, a nie pojedynczymi osobnikami. W takim środowisku ewolucyjnie korzystne mogłoby się okazać rozwinięcie przez pasożyty dodatkowych funkcji ze względu na ich swobodną ewolucję - w tym przypadku byłaby to synteza rybonukleotydów. RNA były pogrupowane w populacje reprezentujące protokomórki z sekwencjami reprezentowanymi *explicite*. Oprócz dynamiki ewolucyjnej ważna była zdolność systemu do utrzymania informacji genetycznej przez wiele pokoleń. Symulacje pokazały, że wygrywały populacje z dwoma rodzajami cząsteczek, z których jedna potrafiła pełnić więcej niż jedną funkcję. Wielofunkcyjność okazała się w tym modelu jednym ze sposobów na obejście progu Eigena - przechowywanie kilku informacji w jednym miejscu znacznie skraca łączną długość kopiowanych przez system sekwencji.

Ostatnim rozważanym przypadkiem był potencjalny udział tzw. inhibitorów, które miałyby

zdolność selektywnego blokowania niektórych cząsteczek w Świecie RNA. System inspirowany był roztworami buforowymi w chemii, z tego powodu inhibitory same mogły podlegać inhibicji, a równowaga między wszystkimi obecnymi w systemie molekułami miała z założenia utrzymywać stały poziom regulowanego w ten sposób RNA. Analiza równań matematycznych pozwoliła oszacować pojemność (zdolność do tłumienia wahań) takiego systemu. Następnie przeprowadzono symulacje wieloagentowe, oparte na modelu opisanym w rozdziale 5. Ważnym wynikiem była faktyczna zdolność inhibitorów do utrzymywania stabilności populacji, które przy ich braku wymierały.

Podsumowując, badania przeprowadzone w ramach niniejszej pracy pozwoliły rzucić nieco światła na procesy ewolucyjne w Świecie RNA. Stworzono w tym celu nowe modele, skupiające się na jego poszczególnych aspektach. Systemy wieloagentowe pozwoliły na wprowadzenie ciągłej przestrzeni, bliższej rzeczywistości niż automaty komórkowe, co skutkowało bardziej realistyczną dyfuzją. Udało się w ten sposób potwierdzić wyniki Takeuchiego dotyczące długoterminowej ewolucji populacji cząsteczek RNA, chociaż wygląd fal w symulacji różnił się - były bardziej rozmyte ze względu na swobodniejsze przemieszczanie się molekuł, nie ograniczała ich już sztywna siatka automatu komórkowego. Rozszerzenie wspomnianego modelu o sekwencje pozwoliło udowodnić, że przy określonych warunkach replikazy mogły wyłonić się samoczynnie z grupy losowych łańcuchów RNA i utrzymać swoją liczebność. Ma to ogromną wagę, gdyż wyjaśnia jak Świat RNA mógł w ogóle powstać, dodatkowo uwiarygodniając tę hipotezę. Kolejnym krokiem w badaniach autora była próba opisu procesów zachodzących w prebiotycznej zupie za pomocą równań różniczkowych cząstkowych i porównanie wyników z symulacją wieloagentową. Poszukując ogólnych praw rządzących ewolucją w Świecie RNA, ostatecznie sformułowano hipotezę wyjaśniającą obserwowane w poprzednich modelach fale jako różne etapy cyklicznego rozwoju populacji wynikające ze spełnienia odpowiedniej nierówności. Po sformułowaniu wspomnianej teorii zbadano Świat RNA z innej perspektywy - skupiono się na dodatkowych funkcjach pasożytów (syntezie nukleotydów), gdzie każdy rodzaj aktywności enzymatycznej zależał *explicite* od sekwencji (a dokładniej struktury drugorzędowej). Rezultaty były zaskakujące - symulowanym populacjom udało się obejść próg Eigena przez tworzenie struktur pośrednich: cząsteczek zdolnych do pełnienia więcej niż jednej funkcji. W ten sposób łączna ilość informacji genetycznej zakodowana w jednej sekwencji o ustalonej długości była znacznie większa. Ostatnim opisanym w tej rozprawie modelem była populacja RNA z inhibitorami. Analiza teoretyczna oraz trzy rodzaje symulacji pokazały, że mechanizm oparty na krótkich RNA blokujących replikację mógł w niektórych warunkach skutecznie stabilizować system, chroniąc go przed wymarciem. Otwiera to nowy kierunek badań - dalsze analizy działania potencjalnych inhibitorów połączone z eksperymentami laboratoryjnymi.

Bibliografia

- [1] „harmonic function”. Encyclopedia of Mathematics, 5 czerwca 2020r., http://encyclopediaofmath.org/index.php?title=Harmonic_function&oldid=47180, dostęp 28 lutego 2023r.
- [2] M. L. L. Abell i J. P. Braselton. *Introductory Differential Equations*. Academic Press, 2018.
- [3] C. Adami. The use of information theory in evolutionary biology. *Ann N Y Acad Sci*, 1256:49–65, 2012.
- [4] A. M. Alanazi, G. Muhiuddin, D. A. Al-Balawi, i S. Samanta. Different DNA Sequencing Using DNA Graphs: A Study. *Applied Sciences*, 12(11):5414, 2022.
- [5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, i P. Walter. *Molecular Biology of the Cell. 4th edition.*, chapter From DNA to RNA. New York: Garland Science, 2002.
- [6] T. O. ARB. „the openmp api specification for parallel programming”, <https://www.open-mpi.org/>, dostęp 26 kwietnia 2023r.
- [7] J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, i P. Holliger. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLIFE*, 7:e35255, 2018.
- [8] J. Błażewicz. *Złożoność obliczeniowa problemów kombinatorycznych*. Wydawnictwo Naukowo-Techniczne, Warszawa, 1988.
- [9] Y. Bansho, T. Furubayashi, N. Ichihashi, i T. Yomo. Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc Natl Acad Sci U S A*, 113:4045–4050, 2016.
- [10] M. Ben-Ari. *Podstawy programowania współbieżnego i rozproszonego*. Wydawnictwo Naukowo-Techniczne, Warszawa, 2016.
- [11] F. Berezovskaya, G. P. Karev, M. I. Katsnelson, Y. I. Wolf, i E. V. Koonin. Corresponding. Stable coevolutionary regimes for genetic parasites and their hosts: you must differ to coevolve. *Biology direct*, 13:27, 2018.
- [12] J. L. Berggren. „calculus”. Encyclopedia Britannica, 20 września 2022r., <https://www.britannica.com/science/calculus-mathematics>, dostęp 24 lutego 2023r.
- [13] A. Bibillo, M. Figlerowicz, i R. Kierzek. The non-enzymatic hydrolysis of oligoribonucleotides VI. The role of biogenic polyamines. *Nucleic Acids Research*, 27(19):3931–3937, 1999.

- [14] A. Bibillo, M. Figlerowicz, K. Ziomek, i R. Kierzek. The Nonenzymatic Hydrolysis of Oligoribonucleotides VII. Structural Elements Affecting Hydrolysis. *Nucleosides, Nucleotides & Nucleic Acids*, 19(5–6):977–994, 2000.
- [15] M. Biesiada, K. J. Purzycka, M. Szachniuk, J. Blazewicz, i R. W. Adamiak. Automated RNA 3D Structure Prediction with RNAComposer. *Methods in Molecular Biology*, 1490:199–215, 2016.
- [16] J. Blazewicz, A. Hertz, D. Kobler, i D. de Werra. On some properties of DNA graphs. *Discrete Applied Mathematics*, 98(1–2):1–19, 1999.
- [17] J. Blazewicz i M. Kasprzak. Graph reduction and its application to DNA sequence assembly. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, 56(1):65–70, 2008.
- [18] M. C. Boerlijst i P. Hogeweg. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasite. *Physica D: Nonlinear Phenomena*, 48(1):17–28, 1991.
- [19] M. C. Boerlijst i P. Hogeweg. Spatial gradients enhance persistence of hypercycles. *Physica D: Nonlinear Phenomena*, 88(1):29–39, 1995.
- [20] D. Borowski, J. Mikołajski, i J. Morchało. *Równania różniczkowe cząstkowe w zastosowaniach*, strona 9. Wydawnictwo Politechniki Poznańskiej, Poznań, 1995.
- [21] G. Boza, A. Szilágyi, Á. Kun, M. Santos, i E. Szathmáry. Evolution of the division of labor between genes and enzymes in the RNA world. *PLoS Computational Biology*, 10(12):e1003936, 2014.
- [22] R. R. Breaker. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(2):a003566, 2012.
- [23] R. E. Britannica. „arrhenius equation”. Encyclopedia Britannica, 8 marca 2023r., <https://www.britannica.com/science/Arrhenius-equation>, dostęp 14 marca 2023r.
- [24] R. E. Britannica. „wave-particle duality”. Encyclopedia Britannica, 10 lutego 2023r., <https://www.britannica.com/science/wave-particle-duality>, dostęp 8 maja 2023r.
- [25] T. R. Cech. Structural biology. the ribosome is a ribozyme. *Science*, 289(5481):878–9, 2000.
- [26] T. R. Cech. The RNA worlds in context. *Cold Spring Harbor Perspectives in Biology*, 4(7):a006742, 2012.
- [27] K. E. Chapple, D. P. Bartel, i P. J. Unrau. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA*, 9(10):1208–1220, 2003.
- [28] R. E. Christoffersen. *Basic Principles and Techniques of Molecular Quantum Mechanics*, strony 181–182. Springer-Verlag, Nowy Jork, USA, 1989.
- [29] X.-Y. Chu i H.-Y. Zhang. Prebiotic Synthesis of ATP: A Terrestrial Volcanism-Dependent Pathway. *Life*, 13(3):731, 2023.
- [30] T. H. Cormen, C. E. Leiserson, R. L. Rivest, i C. Stein. *Wprowadzenie do Algorytmów*, strony 24–28, 43–53, 949. Wydawnictwo Naukowe PWN, Warszawa, 2012.
- [31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, i C. Stein. *Wprowadzenie do Algorytmów*, strony 791–832. Wydawnictwo Naukowe PWN, Warszawa, 2012.
- [32] L. Corrêa, L. Arantes, P. Sens, M. Inostroza-Ponta, i M. Dorn. A dynamic evolutionary multi-agent system to predict the 3d structure of proteins. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, strony 1–8, 2020.

- [33] F. H. C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38:367–379, 1968.
- [34] M. B. Cronhjort. Hypercycles versus parasites in the origin of life: model dependence in spatial hypercycle systems. *Origins of Life and Evolution of the Biosphere*, 25:227–233, 1995.
- [35] T. Czárán, B. Könnyű, i E. Szathmáry. Metabolically coupled replicator systems: Overview of an rna-world model concept of prebiotic evolution on mineral surfaces. *Journal of Theoretical Biology*, 381:39–54, 2015.
- [36] J. Damuth i I. L. Heisler. Alternative formulations of multilevel selection. *Biology and Philosophy*, 3:407–430, 1988.
- [37] A. P. Dos Santos Junior, M. V. José, i S. T. de Farias. From RNA to DNA: Insights about the transition of informational molecule in the biological systems based on the structural proximity between the polymerases. *Biosystems*, 206:104442, 2021.
- [38] R. Durrett. The Importance of Being Discrete (and Spatial). *Theoretical Population Biology*, 46(3):363–394, 1994.
- [39] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 58(10):465–523, 1971.
- [40] M. Eigen, J. McCaskill, i P. Schuster. Molecular quasi-species. *The Journal of Physical Chemistry*, 92(24):6881–6891, 1988.
- [41] M. Eigen i P. Schuster. *The Hypercycle: A Principle of Natural Self-Organization*. New York: Springer, 1979.
- [42] T. Ellinger, R. Ehricht, i J. S. McCaskill. In vitro evolution of molecular cooperation in CATCH, a cooperatively coupled amplification system. *Chemistry & Biology*, 5:729–741, 1998.
- [43] R. Erban. From molecular dynamics to brownian dynamics. In *Proceedings of the Royal Society A*, volume 470, strona 20140036, 2014.
- [44] I. for Theoretical Chemistry. „viennarna web services”, <http://rna.tbi.univie.ac.at/>, dostęp 26 kwietnia 2023r.
- [45] P. Forterre. Archaea: what can we learn from their sequences? *Current Opinion in Genetics & Development*, 7(6):764–770, 1997.
- [46] R. M. Füchslin, S. Altmeyer, i J. S. McCaskill. Evolutionary stabilization of generous replicases by complex formation. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38:103–110, 2004.
- [47] T. Furubayashi, K. Ueda, Y. Bansho, D. Motooka, S. Nakamura, R. Mizuuchi, i N. Ichihashi. Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. *Elife*, 9:e56038, 2020.
- [48] W. Gilbert. Origin of life: The RNA world. *Nature*, 319:618, 1986.
- [49] A. N. Gorban, H. P. Sargsyan, i H. A. Wahab. Quasichemical models of multicomponent nonlinear diffusion. *Mathematical Modelling of Natural Phenomena*, 6(5):184–262, 2011.
- [50] M. M. Hanczyc, S. M. Fujikawa, i J. W. Szostak. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science*, 302(5645):618–622, 2003.

- [51] J. B. Hartle. *Grawitacja: Wprowadzenie do ogólnej teorii względności Einsteina*, strony 541–542. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa, 2020.
- [52] J. Hofbauer i K. Sigmund. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, 1978.
- [53] P. Hogeweg i N. Takeuchi. Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. *Origins of Life and Evolution of the Biosphere*, 33(4–5):375–403, 2003.
- [54] E. C. Holmes. On being the right size. *Nature Genetics*, 37(9):923–924, 2005.
- [55] J. Hu, P. Bhowmick, I. Jang, F. Arvin, i A. Lanzon. A decentralized cluster formation containment framework for multirobot systems. *IEEE Transactions on Robotics*, 37(6):1936–1955, 2021.
- [56] S. in the Public Interest. „open mpi: Open source high performance computing”, 3 kwietnia 2023r., <https://www.open-mpi.org/>, dostęp 24 kwietnia 2023r.
- [57] E. J. Javaux. Challenges in evidencing the earliest traces of life. *Nature*, 572:451–460, 2019.
- [58] J. Johnsen. Simple proofs of nowhere-differentiability for weierstrass’s function and cases of slow growth. *Journal of Fourier Analysis and Applications*, 16:17–33, 2010.
- [59] M. Kasprzak. On the link between DNA sequencing and graph theory. *Computational Methods in Science and Technology*, 10(1):39–47, 2004.
- [60] R. Kierzek. Hydrolysis of oligoribonucleotides: influence of sequence and length. *Nucleic Acids Research*, 20(19):5073–5077, 1992.
- [61] R. Kierzek. Nonenzymatic cleavage of oligoribonucleotides. *Methods in enzymology*, 341:657–675, 2001.
- [62] Y. E. Kim i P. G. Higgs. Co-operation between Polymerases and Nucleotide Synthetases in the RNA World. *PLoS Computational Biology*, 12(11):e1005161, 2016.
- [63] B. Könyvű, A. Szilágyi, i T. Czárán. In silico ribozyme evolution in a metabolically coupled RNA population. *Biology Direct*, 10:30, 2015.
- [64] F. M. Kruse, J. S. Teichert, i O. Trapp. Prebiotic nucleoside synthesis: The selectivity of simplicity. *Chemistry European Journal*, 26(65):14776–14790, 2020.
- [65] K. Kuratowski. *Rachunek różniczkowy i całkowy*, strony 101–134. Wydawnictwo Naukowe PWN, Warszawa, 2012.
- [66] K. J. Laidler. „chemical kinetics”. Encyclopedia Britannica, 16 kwietnia 2020r., <https://www.britannica.com/science/chemical-kinetics>, dostęp 15 marca 2023r.
- [67] K. J. Laidler. „law of mass action”. Encyclopedia Britannica, 3 stycznia 2022r., <https://www.britannica.com/science/law-of-mass-action>, dostęp 14 marca 2023r.
- [68] K. J. Laidler. „reaction rate”. Encyclopedia Britannica, 5 grudnia 2022r., <https://www.britannica.com/science/reaction-rate>, dostęp 14 marca 2023r.
- [69] R. Landauer. Information is Physical. *Physics Today*, 44:23–29, 1991.
- [70] M. W. Lau, K. E. Cadieux, i P. J. Unrau. Isolation of Fast Purine Nucleotide Synthase Ribozymes. *Journal of the American Chemical Society*, 126(48):15686–15693, 2004.

- [71] A. M. Lesk. „bioinformatics”. Encyclopedia Britannica, 16 lutego 2023r., <https://www.britannica.com/science/bioinformatics>, dostęp 28 kwietnia 2023r.
- [72] B. Li, Y. Cao, E. Westhof, i Z. Miao. Advances in RNA 3D Structure Modeling Using Experimental Data. *Frontiers in Genetics*, 11:574485, 2020.
- [73] W. Ma, C. Yu, W. Zhang, i J. Hu. Nucleotide synthetase ribozymes may have emerged first in the RNA world. *RNA*, 13(11):2012–2019, 2007.
- [74] D. M. MacKay. *Information, Mechanism and Meaning*. MIT Press: Cambridge, 1969.
- [75] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, i C. Sander. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE*, 6(12):e28766, 2011.
- [76] J. F. C. Mayen i J. Blazewicz. Recent Results on Computational Molecular Modeling of The Origins of Life. *Foundations of Computing and Decision Sciences*, 45(1):35–46, 2020.
- [77] J. S. McCaskill i G. J. Bauer. Images of evolution: origin of spontaneous RNA replication waves. *Proceedings of the National Academy of Sciences*, 90(9):4191–4195, 1993.
- [78] J. S. McCaskill, R. M. Füchslin, i S. Altmeyer. The stochastic evolution of catalysts in spatially resolved molecular systems. *Biological Chemistry*, 382(9):1343—1363, 2001.
- [79] S. C. Meyer. *Signature in the Cell: DNA and the Evidence for Intelligent Design*. HarperOne, Nowy Jork, USA, 2009.
- [80] R. Mizuuchi, T. Furubayashi, i N. Ichihashi. Evolutionary transition from a single RNA replicator to a multiple replicator network. *Nat Commun*, 13:1460, 2022.
- [81] J. D. Moore. *Introduction to Partial Differential Equations*. Kendall Hunt Publishing, 2003.
- [82] I. Nam, J. K. Lee, H. G. Nam, i R. N. Zare. Abiotic production of sugar phosphates and uridine ribonucleoside in aqueous microdroplets. *PNAS*, 114(47):12396–12400, 2017.
- [83] I. Nam, H. G. Nam, i R. N. Zare. Abiotic synthesis of purine and pyrimidine ribonucleosides in aqueous microdroplets. *PNAS*, 115(1):36–40, 2017.
- [84] J. W. Nelson i R. R. Breaker. The lost language of the RNA World. *Science Signaling*, 10(483):eaam8812, 2017.
- [85] L. Nunney. Group Selection, Altruism, and Structured-Deme Models. *The American Naturalist*, 126(2):212–230, 1998.
- [86] L. E. Orgel. Polymerization on the rocks: theoretical introduction. *Origins of life and evolution of the biosphere*, 28(3):227–234, 1998.
- [87] L. E. Orgel. Prebiotic chemistry and the origin of the RNA world. *Critical Reviews in Biochemistry and Molecular Biology*, 39(2):99–123, 2004.
- [88] B. K. D. Pearce, R. E. Pudritz, D. A. Semenov, i T. K. Henning. Fate of nucleobases in warm little ponds. *PNAS*, 114(43):11327–11332, 2017.
- [89] K. J. Purzycka, M. Popena, M. Szachniuk, M. Antczak, P. Lukasiak, J. Blazewicz, i R. W. Adamiak. Automated 3D RNA structure prediction using the RNAComposer method for riboswitches. *Methods in Enzymology*, 553:3–34, 2015.
- [90] J. S. Reuter i D. H. Mathews. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11(129), 2010.

- [91] A. Rich. On the problems of evolution and biochemical information transfer. *Horizons in biochemistry*, strony 103–126, 1962.
- [92] P. B. Rimmer, J. Xu, S. J. Thompson, E. Gillen, J. D. Sutherland, i D. Queloz. The origin of RNA precursors on exoplanets. *Science Advances*, 4(8):eaar3302, 2018.
- [93] M. P. Robertson i G. F. Joyce. The origins of the rna world. *Cold Spring Harb Perspect Biol*, 4:a003608, 2012.
- [94] K. Rogers. „abiogenesis”. Encyclopedia Britannica, 15 lutego 2023, <https://www.britannica.com/science/abiogenesis>, dostęp 3 marca 2023.
- [95] A. Roth i R. R. Breaker. The structural and functional diversity of metabolite-binding riboswitches. *Annual Review of Biochemistry*, 78:305–334, 2009.
- [96] K. Sato, Y. Kato, M. Hamada, T. Akutsu, i K. Asai. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics*, 27(13):i85–i93, 2011.
- [97] T. D. Schneider. A brief review of molecular information theory. *Nano Commun Netw*, 1(3):173–180, 2010.
- [98] U. C. Schreiber i C. Mayer. *The First Cell: The Mystery Surrounding the Beginning of Life*. Springer, Cham, Szwajcaria, 2020.
- [99] P. Schuster, K. Sigmund, i R. Wolff. Dynamical systems under constant organization I. Topological analysis of a family of non-linear differential equations - A model for catalytic hypercycles. *Bulletin of Mathematical Biology*, 40(6):743–769, 1978.
- [100] F. Seredynski. Competitive Coevolutionary Multi-Agent Systems: The Application to Mapping and Scheduling Problems. *Journal of Parallel and Distributed Computing*, 47(1):39–57, 1997.
- [101] V. Shah, J. de Bouter, Q. Pauli, A. S. Tupper, i P. G. Higgs. Survival of RNA Replicators is much Easier in Protocells than in Surface-Based, Spatial Systems. *Life*, 9:65, 2019.
- [102] Y. Shoham. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, Anglia, 2008.
- [103] G. F. Simmons. *Differential Equations with Applications and Historical Notes*, strony 4–9,487. Chapman and Hall/CRC, Nowy Jork, USA, 2016.
- [104] J. M. Smith. Models of evolution. *Proceedings of the Royal Society of London B*, 219:315–325, 1983.
- [105] J. Synak, A. Rybarczyk, i J. Blazewicz. Multi-agent approach to sequence structure simulation in the RNA world hypothesis. *PLoS ONE*, 15(8):e0238253, 2020.
- [106] J. Synak, A. Rybarczyk, i J. Blazewicz. RNA world modeling: A comparison of two complementary approaches. *Entropy*, 24(4):536, 2022.
- [107] A. Szilágyi, B. Könnyű, i T. Czárán. Dynamics and stability in prebiotic information integration: an RNA World model from first principles. *Scientific Reports*, 10:51, 2020.
- [108] A. Szilágyi, I. Zachar, I. Scheuring, Á. Kun, B. Könnyű, i T. Czárán. Ecology and Evolution in the RNA World Dynamics and Stability of Prebiotic Replicator Systems. *Life*, 7:48, 2017.

- [109] N. Szostak, F. Royo, A. Rybarczyk, M. Szachniuk, J. Blazewicz, A. del Sol, i J. M. Falcon-Perez. Sorting signal targeting mRNA into hepatic extracellular vesicles. *RNA Biology*, 11(7):836–844, 2014.
- [110] N. Szostak, J. Synak, M. Borowski, S. Wasik, i J. Blazewicz. Simulating the origins of life: The dual role of RNA replicases as an obstacle to evolution. *PLoS ONE*, 12(7):e0180827, 2017.
- [111] N. Szostak, S. Wasik, i J. Blazewicz. Hypercycle. *PLOS Computational Biology*, 12(4):e1004853, 2016.
- [112] N. Takeuchi i P. Hogeweg. The Role of Complex Formation and Deleterious Mutations for the Stability of RNA-Like Replicator Systems. *Journal of Molecular Evolution*, 65:668–686, 2007.
- [113] N. Takeuchi i P. Hogeweg. Evolution of complexity in RNA-like replicator systems. *Biology Direct*, 3:11, 2008.
- [114] N. Takeuchi i P. Hogeweg. Multilevel Selection in Models of Prebiotic Evolution II: A Direct Comparison of Compartmentalization and Spatial Self-Organization. *PLoS Comput Biol*, 5(10):e1000542, 2009.
- [115] N. Takeuchi i P. Hogeweg. Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Physics of Life Reviews*, 9(3):219–263, 2012.
- [116] P. D. Taylor i L. B. Jonker. Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences*, 40(1–2):145–156, 1978.
- [117] T. Totani. Emergence of life in an inflationary universe. *Scientific Reports*, 10(1):1671, 2020.
- [118] A. S. Tupper i P. G. Higgs. Error thresholds for rna replication in the presence of both point mutations and premature termination errors. *Journal of Theoretical Biology*, 428:34–42, 2017.
- [119] A. Viti, A. Terzi, i L. Bertolaccini. A practical overview on probability distributions. *Journal of thoracic disease*, 7(3):E7–E10, 2015.
- [120] A. Vlassov. How was membrane permeability produced in an RNA world? *Origins of life and evolution of the biosphere*, 35(2):135–149, 2005.
- [121] F. Wachowius i P. Holliger. Non-Enzymatic Assembly of a Minimized RNA Polymerase Ribozyme. *ChemSystemsChem*, 1:1–4, 2019.
- [122] Y. Wan i K. Chatterjee. „RNA”. Encyclopedia Britannica, 10 listopada 2022r., <https://www.britannica.com/science/RNA>, dostęp 28 kwietnia 2023r.
- [123] C. E. Weinberg, Z. Weinberg, i C. Hammann. Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res*, 47(18):9480–9494, 2019.
- [124] E. B. Wilson. *Vector Analysis*, strony 137–138. Yale University Press, New Haven, USA, 1913.
- [125] A. Wochner, J. Attwater, A. Coulson, i P. Holliger. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science*, 332(6026):209–212, 2011.
- [126] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55(3):601–644, 1983.

- [127] S. Wolfram. *A New Kind of Science*. WOLFRAM MEDIA INC, Champaign, USA, 2019.
- [128] M. Wooldridge. *An Introduction to Multiagent Systems*, strony 15–45. John Wiley & Sons, Chichester, Anglia, 2002.
- [129] L.-F. Wu, Z. Liu, S. J. Roberts, M. Su, J. W. Szostak, i J. D. Sutherland. Template-free assembly of functional RNAs by loop-closing ligation. *Journal of the American Chemical Society*, 144(30):13920–13927, 2022.
- [130] M. Wu i P. G. Higgs. The origin of life is a spatially localized stochastic transition. *Biology Direct*, 7(42), 2012.
- [131] S. Wu, C. Yu, W. Zhang, S. Yin, Y. Chen, Y. Feng, i W. Ma. Tag mechanism as a strategy for the RNA replicase to resist parasites in the RNA world. *PLoS ONE*, 12(3):e0172702, 2017.
- [132] K. Zahnle, L. Schaefer, i B. Fegley. Earth’s earliest atmospheres. *Cold Spring Harbor Perspectives in Biology*, 2(10):a004895, 2010.
- [133] S. J. Zhang, D. Duzdevich, D. Ding, i J. W. Szostak. Freeze-thaw cycles enable a prebiotically plausible and continuous pathway from nucleotide activation to nonenzymatic RNA copying. *PNAS*, 119(17):e2116429119, 2022.
- [134] J. Zheng, Y. Wei, i G.-Z. Han. The diversity and evolution of retroviruses: Perspectives from viral „fossils”. *Virologica Sinica*, 37(1):11–18, 2022.