

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

TESI DI LAUREA

Simulazioni di modelli di protocellule

Author:
Marco PANARELLI

Supervisor:
Marco VILLANI

Tesi laurea triennale

DIPARTIMENTO DI Scienze Fisiche, Informatiche e Matematiche
Corso di Laurea in INFORMATICA

ANNO ACCADEMICO 2020-21

«Not all those who wander are lost.»

J.R.R. Tolkien

Ringraziamenti

Ringrazio il professore Marco Villani per avermi dato la possibilità di svolgere l'esperienza di ricerca che ha portato alla scrittura di questa tesi.

I restanti ringraziamenti non sono rivolti a nessun altro se non alla mia persona, completa dei suoi lati virtuosi e viziosi. Sono le singole persone con le loro azioni, i loro ideali e le loro conoscenze a dover essere elogiate o criticate per le gesta che compiono.

Tuttavia spetta un doveroso ringraziamento ad una persona che finora nella sua vita ha dato tutto per me ...

Grazie mamma.

Capitolo 1

Introduzione

Tutte le forme di vita oggi conosciute sono composte da cellule, unità costituenti fondamentali capaci di auto-replicazione ed evoluzione. È opinione condivisa che le prime forme di vita emersero sulla Terra circa quattro miliardi di anni fa in un ambiente prebiotico e molto differente da quello attuale. Le protocellule sono considerate un passaggio fondamentale nella transizione tra materia inanimata e vita. La protocellula è intesa come un sistema auto-organizzato, guidato dalle leggi della chimica e della fisica, esibente poche funzioni semplificate: una struttura primitiva, un proto-metabolismo e una genetica rudimentale. La protocellula è quindi una struttura in grado di mantenersi, crescere e duplicarsi, producendo sistemi simili ai predecessori. Uno dei modelli più semplici, ma fino a questo momento anche più efficaci, la descrive come una vescicola formata da una membrana lipidica e contenente molecole auto-replicanti, in grado di influenzare la crescita del contenitore che, raggiunte dimensioni troppo grandi, diventa instabile e può dividersi in più cellule figlie, le quali in questo modo condividono il materiale del "genitore". Questa combinazione di eventi di crescita, divisione ed ereditarietà potrebbe essere stata in grado di dar luogo al formarsi di popolazioni di protocellule in grado di evolvere. La protocellula tuttavia è ancora a livello di proposta teorica in quanto non vi sono ancora state conferme sperimentali. A prescindere dall'interesse motivato dal problema dell'origine della vita sul nostro pianeta, la ricerca sulle protocellule trova importanti sbocchi applicativi in diversi campi, come ad esempio la sintesi di farmaci. L'obiettivo di questo scritto è descrivere un software che simula chimiche di protocellule, mostrando anche la sua applicabilità in ambito sperimentale. La protocellula è una struttura, molto simile alle odierne cellule, formata da materiale inorganico racchiuso in uno spazio isolato dall'esterno da una membrana di fosfolipidi che è capace di autoreplicarsi. Nella teoria darwiniana la protocellula si colloca come il predecessore del "singolo antenato comune" dal quale si sarebbe diramata la vita a partire da materia considerata non vivente. In questo campo la maggior difficoltà è riprodurre in laboratorio le condizioni prebiotiche che avrebbero favorito la formazione spontanea di una protocellula. Nel capitolo 2 verranno descritte queste condizioni considerando la Terra nei suoi stadi antecedenti l'emergenza della vita e si ripercorre l'evoluzione della conoscenza umana in questo campo partendo dagli esperimenti di Pasteur fino alla chimica prebiotica di Miller. Il modello da noi utilizzato (uno dei possibili, e tuttora uno dei più accurati) è presentato nel capitolo 3. Tramite alcune assunzioni è possibile ottenerne una forma utile ad essere implementata in un software simulativo (4) che ha l'obiettivo di effettuare esperimenti ed analisi su chimiche di protocellule predefinite o generate randomicamente basandosi su di una chimica artificiale che implementa le strutture alla base della chimica dei polimeri. Infine nel capitolo 5 verranno analizzate delle simulazioni effettuate su chimiche in ambiente statico e dinamico dimostrando come il software scritto possa essere utile per effettuare diversi esperimenti ed analisi.

Capitolo 2

A Brief History of Life

In questo capitolo si vuole raccontare la storia della vita sul nostro pianeta passando per le sue fasi più importanti e ponendo l'attenzione sui contributi degli scienziati che hanno plasmato le basi per lo studio dell'origine della vita. Il racconto è strutturato per creare interesse su quello che si pensa sia stato il primo stadio della vita sul nostro pianeta: la *Protocellula*.

2.1 La formazione della Terra

Si può datare la "nascita" della Terra intorno a 4.54 miliardi di anni fa (Ma), poco dopo la sua formazione si pensa che un corpo planetario, comunemente chiamato Theia e dalle dimensioni di Marte, abbia impattato la superficie terrestre andando così a formare il nostro satellite, la Luna. La così detta *ipotesi dell'impatto gigante* è quella scientificamente più favorita nello spiegare la formazione della Luna nonostante ci siano ancora alcuni punti aperti. Il periodo che va dai 4.5 Ma ai 4 Ma viene chiamato Adeano, questo nome deriva dal dio greco Ade e descrive l'aspetto infernale del pianeta, dominato da un'intensa attività vulcanica e con una temperatura superficiale di 230 °C, a queste condizioni va aggiunta una fase, che va dai 4.1 Ma ai 3.8 Ma, di intenso bombardamento meteorico [Britt, 2002]. Dato lo stato del pianeta in questo periodo si pensa che la vita non avesse ancora i requisiti necessari per emergere, tuttavia alcune prove [Wilde S.A., 2013] suggeriscono la presenza di acqua allo stato liquido intorno ai 4.4 Ma e per questo si specula che la vita potrebbe aver avuto inizio già nel periodo Adeano [Julie, 2009] .

2.2 Le prime forme di vita

Alla fine di questo periodo la Terra si raffreddò abbastanza da permettere la formazione dei continenti e degli oceani. Siamo nell'Archeano, che spazia dai 4 Ma ai 2.5 Ma, ed è proprio in questo periodo che vediamo i primi segni di vita sul nostro pianeta. Attualmente il reperto più antico che testimonia la prima forma di vita risale ad approssimativamente 3.42 Ma, si tratta di microorganismi fossilizzati ritrovati in fondo a sorgenti idrotermali e si ritiene che siano proprio le condizioni generate da queste sorgenti ad aver favorito la proliferazione di queste primordiali forme di vita [Cavalazzi Barbara, 2021]. Ci sono diverse evidenze [Dodd Matthew S., 2017] che fanno pensare che siano proprio questi sistemi idrotermali ad aver favorito la formazione dei primi microorganismi, essi garantivano dei flussi continui di acqua marina mista a materiale chimico, derivante dall'interazione tra l'acqua e la roccia, a temperature relativamente basse ($\leq 150^{\circ}\text{C}$). Come si vedrà più avanti, l'apporto continuo di materiale e la presenza di determinate temperature sono condizioni necessarie

per permettere la formazione di rudimentali strutture cellulari. Se del materiale viene racchiuso da una membrana in un spazio relativamente piccolo e in questo spazio le sostanze, reagendo tra loro, esibiscono dei comportamenti abbastanza complessi, allora vi è una possibilità di ottenere strutture resistenti, ed in qualche grado aventi la possibilità di crescere e replicarsi, queste strutture vengono chiamate *Protocellule*. Le caratteristiche della protocellula e della sua formazione verranno ampiamente discusse successivamente, per ora è importante sottolineare il suo ruolo nel passaggio da una prima fase di semplici reazioni chimiche (*fase chimica*) ad una successiva (*fase biologica*) in cui sono maggiormente presenti le caratteristiche della vita, come l'evoluzione e l'interazione tra individui.

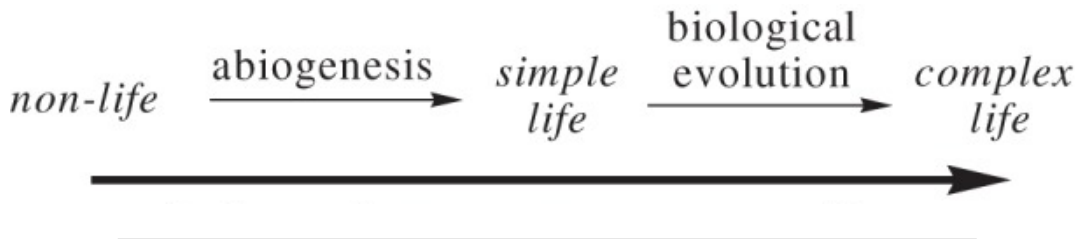


FIGURA 2.1: Transizione da non-vita a forme di vita complessa.

Seguendo lo schema dell'immagine 2.1 la protocellula si colloca nel processo di *abiogenesi*, un processo attraverso il quale la vita emerge a partire da materiale non vivente e da eventi di natura chimico-fisica. Da questo insieme di materiali e reazioni chimiche si sarebbero poi create strutture e comportamenti via via più complessi, fino ad arrivare ai primi organismi unicellulari. Benché l'abiogenesi sia ampiamente accettata come spiegazione dell'origine della vita, sono ancora molti i meccanismi ai quali non si riesce a dare una spiegazione soddisfacente. Successivamente si vedrà come il processo che ha portato generale fiducia nella teoria abiogenica è ben strutturato, molteplici infatti sono gli apporti di scienziati, in diversi campi delle scienze naturali, che a partire da Pasteur da metà dell'800, e ancora prima da Lazzaro Spallanzani, hanno portato all'accettazione di questa teoria.

2.3 Dagli organismi unicellulari all'esplosione Cambriana

Siamo nel periodo Archeano (4 Ma - 2.5 Ma) e si può affermare, con dovuta cautela, che verso la sua fine le scintille della vita fossero accese già da diverso tempo. Dalle protocellule si è passati, in un lasso di tempo non ben definito, ai primi organismi unicellulari i quali hanno dominato la terra probabilmente per 3 miliardi di anni, la maggior parte della sua esistenza. Il come si è arrivati a organismi pluricellulari è, se possibile, ancora più nebuloso dell'emergenza della vita unicellulare. Alcuni reperti [El Albani Abderrazak, 2010] datati tra i 2.3 Ma e i 2.1 Ma fanno pensare alla presenza di vita pluricellulare già in questo periodo, nonostante ciò una datazione più accurata colloca la presenza di organismi pluricellulari 600 milioni di anni fa [Chen L., 2014]. Quello che è certo è che circa 541 milioni di anni fa tutti i maggiori phyla iniziano ad apparire tra i reperti fossili in quella che è chiamata *esplosione Cambriana* [Maloof A. C., 2014]. Seguendo la linea temporale di 2.2 si può notare come a partire da questo periodo l'enorme diversificazione ha portato la vita ad essere come la conosciamo oggi, infatti la maggior parte dei phyla animali odierni è apparsa proprio in questo periodo [G.E, 2014].

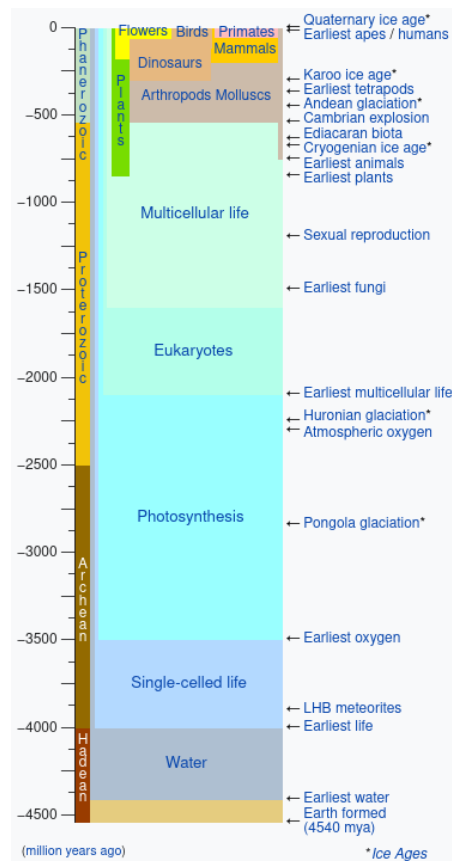


FIGURA 2.2: La timeline prende in considerazione tutta la vita della Terra.

2.4 Una storia di estinzioni

Fino ad ora è stata raccontata la storia della vita sulla Terra partendo dalla formazione del pianeta e passando attraverso le fasi principali: i primi organismi unicellulari, discendenti di semplici reazioni chimico-fisiche avvenute all'interno di una membrana, che hanno dominato il pianeta per un lungo periodo si sarebbero evoluti in agglomerati di cellule capaci di collaborare tra loro generando così forme di vita più complesse, gli organismi pluricellulari. Questi si sarebbero continuamente diversificati fino ad arrivare all'era Cambriana in cui emersero la maggior parte delle forme di vita che oggi conosciamo. In questo percorso sono state ovviamente omesse diverse sottofasi, non meno importanti nel dipingere il quadro della vita, per non divergere troppo dall'obiettivo dello scritto. Tuttavia è necessario aggiungere ancora un tassello alla storia costruita finora: la vita sulla Terra è stata caratterizzata da continue e cicliche estinzioni che hanno spazzato via la maggior parte delle forme di vita che hanno abitato il nostro pianeta. Si stima infatti che più del 99% delle specie esistite sulla Terra si siano estinte [Kunin W.E., 1996]. Per portare un esempio l'estinzione di massa del Permiano-Triassico [Wikipedia, n.d.] ha portato alla scomparsa dell'81% delle specie marine e del 70% delle specie di vertebrati terrestri, inoltre si ritiene che questa fu l'unica estinzione di massa di insetti.

Sono almeno una ventina le estinzioni, tra maggiori e minori, avvenute sulla Terra [Ritchie e Roser, 2021], tra queste nell'immagine 2.3 sono evidenziate le 5 principali. Le ipotesi sulle cause delle estinzioni sono diverse e coinvolgono agenti esterni

al pianeta, come asteroidi o esplosioni di supernove, ed eventi interni come cambiamenti climatici, acidificazione degli oceani e vulcanismo. Ad ogni modo è giusto prendere in considerazione anche un diverso punto di vista sui motivi che avrebbero portato ciclicamente alla scomparsa di diverse forme di vita. Come propone S. Kauffman nel libro *At Home in the Universe* le estinzioni sembrano essere il riflesso della dinamica di una comunità di specie. La necessità di adattarsi e sopravvivere ad un ambiente continuamente influenzato e modificato da altre specie porterebbe alcune di queste a sopperire in favore di altre. Sempre nel suo lavoro egli propone un paragone con l'avanzamento tecnologico: con l'emergere di nuove e più performanti tecnologie quelle vecchie vengono necessariamente soppiantate. Le estinzioni quindi non sarebbero altro che una caratteristica emergente ed endemica del sistema complesso che governa la vita.

'Big Five' Mass Extinctions in Earth's History

A mass extinction is defined by the loss of at least 75% of species within a short period of time (geologically, this is around 2 million years).

Our World
in Data

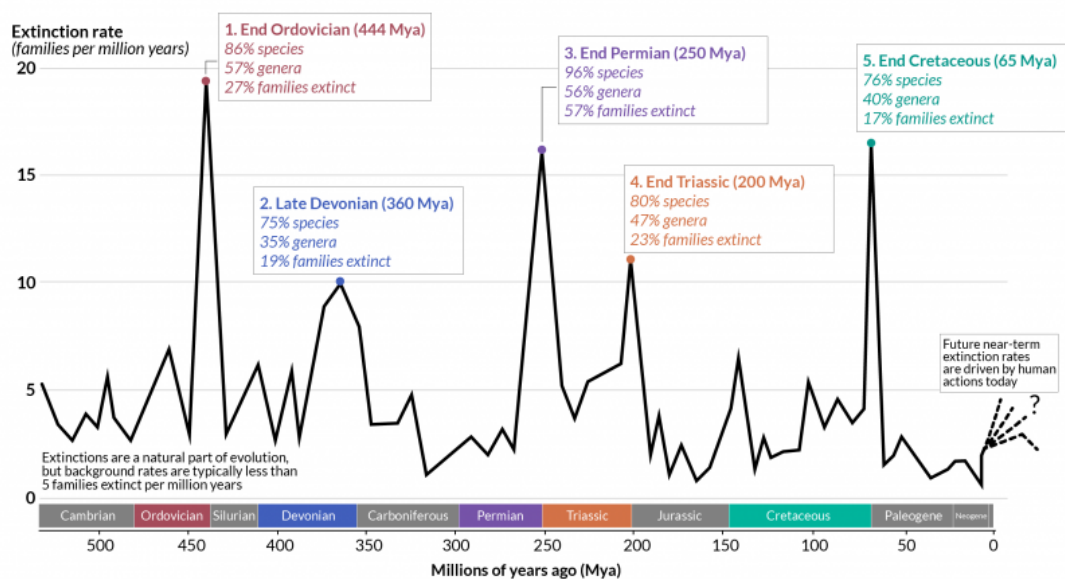


FIGURA 2.3: Le 5 estinzioni più importanti a partire dal Cambriano [Ritchie e Roser, 2021].

2.5 Da Pasteur all'abiogenesi

Nella sua storia l'essere umano ha sempre provato a costruire teorie unificatrici, è innata infatti la necessità di mettere i vari pezzi della conoscenza in un unico costruito in grado di spiegare tutti i fenomeni e gli eventi che ci si presentano dinnanzi. Paradossalmente, più la nostra conoscenza aumenta, più è complicato raggiungere questo risultato.

2.5.1 Pasteur e la biogenesi

Aristotele, già nel 300 A.C., volle spiegare tutto, e in realtà ci riuscì per più di mille anni, fino all'avvento di Galileo, Copernico ed in generale del metodo scientifico. Il completo cambio di paradigma, perlomeno nel campo della fisica e dell'astronomia, che si ha a partire da Galileo lo si può paragonare all'ascesa dell'abiogenesi a

partire dagli incontestabili esperimenti di Pasteur. Fino al XVIII secolo l'unica teoria accettata era quella della generazione spontanea: questa teoria, anch'essa costruita da Aristotele [André, 1998], presupponeva che la vita potesse generarsi "spontaneamente" da materia inanimata. Per esempio larve e insetti potevano nascere in maniera autonoma da fango o da carcasse in putrefazione. Questa teoria resse, come il modello tolemaico, per due millenni, ma con gli esperimenti di Francesco Redi, nel XVII secolo sulla generazione spontanea degli insetti [Levine R, 2008], e con quelli di Lazzaro Spallanzani [Spallanzani, 1765] nel XVIII secolo fu evidente che quella teoria non poteva durare ancora a lungo. Tuttavia fu grazie a L. Pasteur, nel 1859, che si poté abbandonare definitivamente l'idea della generazione spontanea, grazie ad un semplice e brillante esperimento [Pasteur, 1859]. Egli usò un matraccio a collo d'oca, la cui forma permetteva il passaggio dell'ossigeno ma impediva che agenti contaminanti, come spore e batteri, entrassero a contatto con il liquido interno. Una volta bollito questo liquido, per uccidere ogni forma di vita al suo interno, dimostrò che i microorganismi apparivano soltanto se si rompeva il collo del matraccio, permettendo così ai microorganismi di entrare.

Spodestata la teoria della generazione spontanea si arrivò alla conclusione che la vita per emergere deve partire quantomeno da altra vita, questo concetto è chiamato *biogenesi* e prende piede proprio a partire dagli esperimenti di Pasteur. Tuttavia è evidente come questo assunto preservi una fallacia logica, poiché quella che definiamo vita, in un primo momento, deve essersi generata da qualcos'altro. Come è stato esposto precedentemente (2.2) questo "qualcos'altro" potrebbe essere proprio la protocellula, o un'evoluzione di essa. La strada per arrivare ad essa è ancora lunga, ma già dalla teoria dell'evoluzione delle specie, elaborata indipendentemente da A.R. Wallace e da C.R. Darwin nel 1858, emerge l'idea che tutte le forme di vita debbano discendere da un singolo antenato, che in origine deve essere stato molto semplice dal punto di vista biologico. La vita si sarebbe poi diramata da questo antenato comune per mezzo della selezione naturale nei miliardi di anni successivi.

2.5.2 Oparin e Miller

Fino ad ora ci siamo riferiti al termine "vita" dandone per scontato il valore semantico, se si chiede ad una persona la differenza tra un sasso e una pianta si avrà come risposta che il sasso è "inanimato" mentre la pianta è un essere vivente. Tuttavia abbiamo già visto precedentemente (2.1) che potrebbero essere proprio le componenti che noi consideriamo informalmente "inanimate" a costituire la base dell'origine della vita, mediante il processo di *abiogenesi*. È importante porre l'accento sul termine *processo*: l'ipotesi prevalente è che la transizione da "non-vita" a "vita" sia appunto il frutto di un processo di complessità crescente piuttosto che la conseguenza di un singolo evento. Il biochimico A. Oparin e lo scienziato britannico J. Haldane nel 1929 arrivarono indipendentemente alle stesse conclusioni sui macro-eventi che avrebbero composto questo processo, successivamente il biochimico Robert Shapiro condensò le loro ipotesi, in maniera più "matura", nei seguenti punti (ipotesi del brodo primordiale).

1. La Terra, ai suoi albori, aveva un'atmosfera ricca di componenti quali metano, ammoniacale, idrogeno e vapore acqueo.
2. In questa atmosfera si sarebbero prodotti semplici componenti organici grazie a diverse forme di energia.
3. Questi componenti si riversarono in un cosiddetto "brodo primordiale" concentrandosi poi in vari punti come le sorgenti idrotermali.

4. In seguito a continue trasformazioni si generarono componenti organici via via più complessi, fino ad arrivare alle prime forme di vita.

Grazie ai contributi di questi scienziati è possibile approcciarsi allo studio dell'origine della vita in modo metodologico e multidisciplinare piuttosto che in maniera soltanto speculativa [Antonio, 2010]. Le ipotesi stilate da Oparin e Haldane mettono chiarezza sul "come" le protocellule possano essersi formate, tuttavia è ancora poco chiaro l'origine del loro contenuto, ovvero i "semplici componenti organici" citati nel 2° punto del precedente schema.

Nel 1952 il neolaureato in chimica Stanley Miller propose un esperimento che avrebbe posto le basi della chimica prebiotica, portò anche la prima evidenza sperimentale sulla teoria del brodo primordiale. Come nel caso di Pasteur, si trattò di un esperimento piuttosto semplice; l'apparato costruito da Miller voleva simulare l'interazione tra un'atmosfera ed un oceano, per far ciò lo riempì con varie sostanze quali metano, ammoniaca, idrogeno nonché dell'acqua portata ad ebollizione. Per attivare la reazione simulò dei fulmini, che si pensa fossero un'importante fonte di energia nella Terra primordiale, ed aspettò una settimana prima di analizzare i risultati. In seguito ad analisi constatò che sul fondo e sui lati del contenitore si formarono diversi amminoacidi ed altri semplici composti organici dimostrando così che i componenti base della vita potrebbero essersi formati nei primi stadi della Terra.

Arrivati a questo punto sembra che la storia raccontata sia abbastanza autoconclusiva: sulla Terra, nei primi stadi in seguito alla sua formazione, erano presenti le giuste componenti e le giuste condizioni per l'emergere delle protocellule che nel tempo avrebbero esibito comportamenti e strutture sempre più complesse. Da esse si sarebbero generati i primi organismi monocellulari dai quali si differenziò la vita così come la conosciamo oggi, spodestando in questo modo la teoria della generazione spontanea in favore dell'abiogenesi (2.5). Nonostante tutto il racconto sia coerente sono molti i passaggi e i fenomeni ancora non chiari. Come citato precedentemente è ancora oscuro il passaggio da vita unicellulare a vita pluricellulare, ma ancora più centrale è il ruolo della protocellula, come si è formata, in che condizioni e come essa si sia successivamente evoluta nel nostro "singolo antenato comune".

Capitolo 3

La Protocellula

Finora (2.2) si è insistito molto sull'idea che la vita sia emersa grazie al processo di abiogenesi, in questo capitolo andremo ad analizzare più nel dettaglio questo processo, caratterizzando meglio le condizioni e i fenomeni che a partire da materiale "non-vivente" possono favorire la formazione della protocellula, un costrutto primordiale capace di ospitare la vita.

3.1 Le caratteristiche della vita

Già nel capitolo 2.5.2 ci si è interrogati sulla definizione del termine "vita"; benché informalmente sia facile definire cosa è vivente e cosa no, quando si comincia a scendere nei dettagli è necessario provare a dare una definizione il più ampia possibile per delineare i confini dentro i quali si vuole operare. Ad oggi esistono più di 120 definizioni di vita [Trifonov, 2011], questo numero così elevato è auto esplicativo sulle difficoltà che filosofi e scienziati devono affrontare nel cercare un'insieme di parole che descrivano in maniera completa questo processo. Ad oggi è generalmente accettato definire un essere vivente come un'entità che ha le seguenti caratteristiche:

1. un **metabolismo** che permette all'entità di crescere e di produrre i propri componenti
2. un modo attraverso il quale **riprodursi** generando nuovi esseri simili a se stesso
3. la presenza di variabilità, con la conseguente possibilità di **evoluzione**

Questa definizione sembra essere in linea con il nostro senso comune, tuttavia è facile trovare un'entità che non rispetta le condizioni poste, nonostante verrebbe comunemente definita come "vivente". I virus non soddisfano il primo punto, infatti essi hanno bisogno di un organismo ospitante per poter svolgere le loro funzioni. Tuttavia per gli obiettivi di questo scritto, ed in generale per la biologia, queste tre caratteristiche sono più che sufficienti come base dalla quale studiare l'argomento.

3.2 La struttura della Protocellula

Nel pensare alla struttura di una protocellula non bisogna discostarsi troppo dall'immagine che si ha di una comune cellula biologica, entrambe hanno una membrana che li protegge dall'ambiente esterno e al loro interno hanno dei componenti che contribuiscono alla vita e alle funzioni della cellula stessa.

3.2.1 Membrana

Nell'analizzare la membrana bisogna collocarsi in quello che precedentemente (2.5.2) è stato chiamato brodo primordiale, una distesa acquosa in cui sono immerse varie sostanze chimiche capaci di reagire tra loro. In questo contesto è possibile osservare una prima forma di auto-organizzazione. Le molecole anfipatiche sono composti chimici formati da una parte idrofila e una idrofoba, si è osservato [McCaskill, 2007] che in ambiente acquoso esse tendono ad aggregarsi tra di loro, poiché non è vantaggioso per la coda idrofoba trovarsi a contatto con un liquido. Tra gli anfipati più rilevanti si citano i fosfolipidi, essi aggregandosi formano una membrana chiusa a doppio strato in cui le code sono a contatto tra di loro mentre le teste idrofile sono rivolte verso il substrato acquoso, una testa verso quello interno e una verso l'esterno. La struttura creata si chiama liposoma ed è in grado di ospitare del materiale al suo interno che viene raccolto dall'ambiente esterno durante la formazione del liposoma stesso.

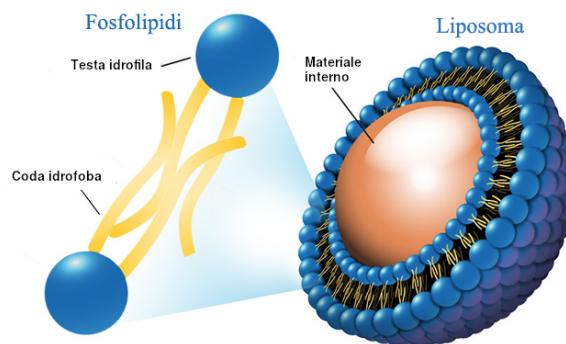


FIGURA 3.1: I fosfolipidi aggregandosi formano una membrana a doppio strato che si chiude su stessa, formando il liposoma

Questo fenomeno ha bisogno di determinate condizioni per potersi verificare, in primis è necessario che la concentrazione dei fosfolipidi in acqua superi una certa soglia (*concentrazione micellare critica*) sotto la quale non si osserva la formazione della membrana a doppio strato. Altro fattore determinante è la temperatura; sopra una determinata temperatura critica (*temperatura di fusione*) la membrana si trova in una fase detta *liquido-cristallina*, in questa fase si ha una membrana semipermeabile, condizioni necessaria per l'avvenimento di una serie di fenomeni. In questa condizione la membrana permette il passaggio di materiale da e verso l'esterno, inoltre i fosfolipidi godono di una mobilità che permette alla membrana di crescere. Il ruolo della membrana può sembrare ovvio, ma è necessario un ulteriore approfondimento. Essa durante la sua formazione raccoglie del materiale dall'ambiente che viene racchiuso in uno spazio limitato, all'interno di questo spazio le reazioni possono avvenire più facilmente poiché la concentrazione delle sostanze è sostanzialmente maggiore rispetto all'ambiente esterno [Serra e Villani, 2017] dove la diluizione avverrebbe molto più velocemente. Quindi, una membrana semipermeabile che fa passare solo alcune sostanze, da e verso l'esterno, fa sì che al suo interno ci siano condizioni più favorevoli all'avvenire di reazioni chimiche, queste reazioni possono portare alla crescita della membrana stessa e successivamente alla sua duplicazione.

3.2.2 Materiale interno

Arrivati a questo punto è chiaro il ruolo centrale che svolge la membrana nella creazione della protocellula, un'altra parte importante è cosa accade all'interno di quest'ultima. Finora ci si è riferiti ad una singola protocellula ma è logico pensare che, date le giuste condizioni, dove se ne forma una se ne possano formare molte altre, parliamo quindi di una popolazione di protocellule. Se però la popolazione è abbastanza piccola da non influenzare l'ambiente in cui opera e non si tiene conto dell'interazione tra individui (benché in generale non sia da trascurare), in questo modo si può studiare il singolo individuo solo relativamente all'ambiente e alle sue condizioni interne. È ora importante capire cosa differenzia un individuo da un altro, il suo materiale genetico. È stato detto che al momento della sua formazione la membrana raccoglie dall'ambiente delle sostanze chimiche, alcune di queste possono contribuire direttamente alle caratteristiche della cellula, determinandone quindi il comportamento. Queste specie vengono chiamate *genetic memory molecules* (GMM) e rappresentano il materiale genetico della protocellula, in questo scritto si definiscono GMMs le specie che hanno un accoppiamento esplicito con la membrana e che quindi ne determinano la velocità di crescita, ci sarà sempre almeno una GMM altrimenti la cellula non potrà crescere. La chimica su cui ci focalizzeremo nel seguito è un'astrazione della chimica dei polimeri (molto importante per quanto riguarda la chimica cellulare), e ne conserva gli aspetti essenziali. Le molecole sono costituite da monomeri (molecole semplici) o polimeri, ovvero catene di monomeri. Per semplicità i monomeri vengono rappresentati mediante una singola lettera dell'alfabeto (A, B, ...) mentre i polimeri da una stringa (ABBA).

3.2.3 Reazioni

Le specie chimiche nello spazio confinato dalla membrana interagiscono tra loro tramite reazioni di vario tipo, l'insieme delle specie chimiche e delle reazioni che avvengono all'interno di una protocellula viene chiamato *chimica*. Le reazioni interne alla protocellula corrispondono, in un certo senso, al metabolismo della protocellula poiché da esse dipende la crescita della cellula e la sua duplicazione. Quando si parla di reazioni si considerano quasi sempre reazioni catalizzate, in esse il catalizzatore è una specie chimica che facilita notevolmente l'avvenire di una reazione. La reazione $A + B + C \rightarrow AB + C$ unisce A e B in un unico composto AB, per come è scritta C sembra quasi non svolgere alcun ruolo, in questo caso C è proprio il catalizzatore e può essere visto come un supporto che facilita l'incontro tra A e B. Volendo essere più precisi, in chimica le reazioni tri-molecolari sono molto rare, infatti vengono sempre scomposte in più reazioni bi-molecolari, nell'esempio precedente i passi sarebbero i seguenti:

1. $A + C \rightarrow \text{Complesso}$
2. $\text{Complesso} + B \rightarrow AB + C$

Il complesso è una sostanza intermedia debolmente-stabile che è composta da uno dei due substrati, in questo caso A, e il catalizzatore, C non viene consumato dalla reazione proprio perché il suo ruolo è quello di favorire l'avvenire della reazione. In teoria la reazione *spontanea* $A + B \rightarrow AB$ può avvenire in natura senza l'aiuto di un catalizzatore ma è ritenuta molto meno probabile, in quanto più lenta, della corrispettiva catalizzata e per questo viene trascurata.

Non tutte le chimiche possono portare alla crescita del contenitore e quindi alla duplicazione della protocellula, tra le più importanti strutture che è bene siano presenti si discutono gli ACS (autocatalytic sets) e i RAF (reflexive autocatalytic food-generated sets). Gli ACS, o insiemi autocatalitici, sono insiemi di molecole le cui reazioni creano altre molecole appartenenti allo stesso insieme, si crea quindi una sorta di economia circolare in cui la creazione di una specie chimica è garantita da altre specie che a loro volta sono create da altre reazioni e così via. Ovviamente un ACS non può essere completamente autonomo, infatti per quanto le specie chimiche siano continuamente rigenerate dal ciclo di reazioni è necessario del "cibo" che mantenga in vita questo ciclo. La definizione di RAF [Serra e Villani, 2017] formalizza la presenza della necessità del cibo, evidenziando l'importanza di tenere in considerazione sia le specie chimiche che la forma delle reazioni che le collegano. Questo cibo è rappresentato da specie chimiche la cui esistenza è garantita dall'esterno del sistema. In molte simulazioni assumeremo che l'ambiente esterno alla protocellula sia stabile ed in grado di rimpiazzare le sostanze che la protocellula stessa assorbe tramite la membrana, queste sostanze saranno poi consumate internamente. In tal caso, la concentrazione di tali sostanze può essere assunta come costante, in gergo chimico si dice che le sostanze sono *bufferizzate*. In simulazioni successive rilasceremo tale vincolo, permettendo variazioni imposte dall'esterno per rappresentare un ambiente variabile. Anche in assenza di alcune specie i RAF riescono a mantenere la loro capacità di auto-riproduzione poiché ci sono altre specie che garantiscono la creazione delle sostanze mancanti. Questa caratteristica di robustezza è affiancata alla forte dipendenza dal substrato proveniente dall'esterno, se questo viene a mancare allora anche l'insieme autocatalitico si distrugge [Serra e Villani, 2017]. Nel modello di Kauffman [Kauffman, 1995] sulla generazione di chimiche di tipo polimerico, che non differisce molto da quanto descritto finora, emerge come la creazione di un ACS è quasi inevitabile se si parte da un insieme di molecole abbastanza variegato, infatti più specie chimiche sono presenti più aumenta la densità delle reazioni fino ad un punto in cui una struttura autocatalitica viene sicuramente a formarsi. Questo suggerisce come l'emergere di queste strutture sia un evento abbastanza comune se sono presenti le giuste condizioni.

3.2.4 Duplicazione

L'autoreplicazione è una caratteristica comune a diversi esseri viventi: il DNA durante la divisione cellulare viene replicato e passato alla cellula figlia. La duplicazione di una protocellula [Serra e Villani, 2017] non differisce molto da un processo di autoreplicazione, con la distribuzione delle sostanze chimiche interne a giocare il ruolo delle caratteristiche ereditarie. Come è stato definito in 3.1, *riproduzione* e *evoluzione* sono due caratteristiche fondamentali per la vita e verrà visto come la protocellula, o quanto meno il modello delineato finora, le detiene entrambe. Prima che la duplicazione possa avvenire è necessario che la cellula cresca, il contenitore si ingrandisce grazie all'aumento del numero di fosfolipidi nella membrana ma questo dipende molto dalle GMMs e in generale dalla chimica della cellula; il ruolo dei RAF "utili" consiste quindi nel garantire la presenza di GMMs. La duplicazione della protocellula ha caratteristiche prettamente fisiche, la cellula crescendo entra in una fase instabile che viene compensata con la divisione in due cellule figlie che hanno approssimativamente la metà dei materiali della cellula madre [Villani e Serra, 2014]. Per semplicità nel modello si assume che le figlie abbiano le stesse GMMs della madre, tuttavia è logico pensare che nel processo di divisione del materiale possa essere perso o modificato così come parte dei fosfolipidi che compongono la

membrana, questo fa sì che si possano introdurre delle *mutazioni* che differenziano la prole dai genitori. Considerando una popolazione di protocellule si può vedere come il processo di selezione Darwiniana abbia la sua importanza, dati i presupposti solo le cellule con la giusta chimica potranno duplicarsi e tra queste coloro che si duplicano più velocemente prenderanno il sopravvento sulla popolazione, diventando le più numerose.

3.2.5 Sincronizzazione

È già stato sottolineato come un prerequisito per la duplicazione è la presenza del giusto materiale genetico, a questo si affianca un'altra proprietà molto importante, la *sincronizzazione*. In una protocellula avvengono principalmente due fenomeni, la crescita del materiale lipidico del contenitore e la crescita delle GMM. Perché la duplicazione possa avvenire è necessario che questi due processi avvengano alla stessa velocità, la crescita troppo veloce del materiale genetico causerebbe l'esplosione della cellula per l'eccessiva presenza di materiale al suo interno, mentre se viceversa è la membrana a crescere più velocemente le GMMs diverrebbero troppo diluite. È quindi necessario che i due processi si sincronizzino, questo fenomeno di sincronizzazione va analizzato tenendo conto delle successive generazioni di una protocellula, è infatti possibile che la stabilità venga raggiunta dopo diverse divisioni. Per chiudere il quadro si sottolinea come sia strettamente necessaria la presenza di un RAF perché la sincronizzazione possa avvenire [Villani e Serra, 2014], rimarcando nuovamente l'importanza di questo tipo di strutture.

Capitolo 4

Implementazione della protocellula in un software di simulazione

Nel capitolo precedente si è descritto genericamente cosa sia una protocellula e da cosa è composta, inoltre si sono analizzati alcuni dei suoi comportamenti più importanti. Parte del materiale interno, ovvero le GMMs, può alimentare la membrana e farla crescere di dimensioni, sopra una certa soglia la cellula si divide generando due cellule figlie con lo stesso patrimonio genetico, in particolare la divisione è possibile se la crescita delle GMMs e quella della membrana avviene allo stesso ritmo, senza sincronizzazione è difficile che la protocellula continui il suo sviluppo. In questo capitolo si andranno a formalizzare queste proprietà della protocellula per delineare un modello [Villani e Serra, 2014] utilizzabile successivamente in un programma, tramite esso si potranno simulare chimiche di protocellule ed effettuare esperimenti su di esse. Ricordando che la chimica di una protocellula è composta dalle sue specie chimiche interne e dalle reazioni di queste, l'obiettivo del software è quello di simulare l'andamento della concentrazione delle specie nel tempo e al susseguirsi delle divisioni della cellula.

4.1 Geometria

La geometria del modello è abbastanza semplice, poiché si assume che la membrana mantenga una forma sferica durante tutto il processo, inoltre rimangono anche invariati spessore (δ) e densità molecolare del materiale lipidico (ρ) che compone la membrana. Con queste premesse usando la geometria elementare si può ricavare il volume interno della protocellula in metri cubi:

$$V_m = \frac{4}{3}\pi(r + \delta)^3 - \frac{4}{3}\pi r^3 \quad (4.1)$$

È tuttavia possibile utilizzare un altro metodo per ricavarlo considerando la quantità di lipide (C) espressa in moli:

$$V(C)_l = \frac{1000}{6}\pi\delta^3\left(\sqrt{\frac{C}{\rho\pi\delta^3}} - \frac{1}{3} - 1\right)^3 \quad (4.2)$$

Da qui in avanti si userà il termine *lipide* per indicare la membrana della cellula, questo perché si presuppone che la membrana sia composta da questo tipo di materiale, infatti anche la densità molecolare ρ sarà sempre uguale a $2640 \text{ mol}/\text{m}^3$ ovvero quella del colesterolo.

4.2 Materiale interno e reazioni

Come già anticipato nei capitoli 3.2.2 le specie chimiche interne si esprimono mediante un alfabeto composto da lettere, di queste specie verrà sempre tenuta in considerazione la concentrazione espressa in mol/l , la nomenclatura A indica "la specie chimica A ", mentre $[A]$ indica la sua concentrazione. Per quanto riguarda le reazioni se ne prevedono 8 tipi elencati di seguito:

- **"Aggiunta"** $NULL \rightarrow A$: una specie viene alimentata da situazioni esterne al modello
- **"Scomparsa"** $A \rightarrow NULL$: una specie scompare (viene cioè espulsa dal sistema modellato)
- **"Cleavage spontaneo"** $A \rightarrow B + C$: una specie si scinde spontaneamente senza l'aiuto di catalizzatori, spesso usata per rappresentare la *scissione* di un complesso
- **"Condensazione spontanea"** $A + B \rightarrow C$: due specie si condensano spontaneamente senza l'aiuto di catalizzatori, spesso usata per rappresentare la *creazione* di un complesso
- **"Scambio"** $AA + BB \rightarrow AB + BA$: due specie si mescolano tra loro scambiandosi alcuni loro componenti
- **"Cleavage"** $AB + C \rightarrow A + B + C$: una specie si scinde con l'aiuto di un catalizzatore
- **"Condensazione"** $A + B + C \rightarrow AB + C$: due specie si condensano con il contributo di un catalizzatore, come già discusso in 3.2.3 questa reazione può essere scomposta in più reazioni bi-molecolari mediante la creazione di un *complesso*
 - "Creazione del complesso" : $A + C \rightarrow Compl$
 - "Rottura del complesso" : $Compl \rightarrow A + C$
 - "Condensazione finale" : $Compl + B \rightarrow AB + C$
- **"Passaggio di membrana"** $A_{ext} \longleftrightarrow A$: una specie viene alimentata da una sostanza esterna bufferizzata

In quest'ultima reazione A_{ext} rappresenta una sostanza esterna alla protocellula che riesce ad attraversare la membrana per alimentare le specie interne, solitamente le sostanze esterne sono bufferizzate per consentire un continuo apporto di cibo. Si presuppone inoltre che tutte le reazioni avvengano istantaneamente. Tutte le reazioni sono modellate attraverso equazioni differenziali ordinarie del primo ordine, in un cleavage ($AB + C \rightarrow A + B + C$) la variazione di concentrazione di una specie AB che si scinde in A e B mediante un catalizzatore C è:

$$\frac{d[AB]}{dt} = -K_B^A [AB][C] \quad (4.3)$$

dove K_B^A è la costante relativa alla singola reazione e il prodotto $K_B^A [AB][C]$ è il *flusso* della reazione, con questa definizione di flusso è facile stabilire la variazione delle

restanti specie della reazione:

$$\begin{cases} \frac{d[C]}{dt} = -flusso \\ \frac{d[A]}{dt} = flusso \\ \frac{d[B]}{dt} = flusso \\ \frac{d[C]}{dt} = flusso \end{cases} \quad (4.4)$$

essendo C il catalizzatore la sua variazione è uguale a 0. La condensazione invece può essere espressa nella sua forma bi-molecolare, mediante il complesso, o tri-molecolare, si vedrà che in base a questa scelta le chimiche si comportano in maniera differente. I restanti tipi di reazione possono essere elaborati usando questo semplice schema, soltanto il "*passaggio di membrana*" presenta uno schema leggermente diverso, in questo caso viene usata la legge di Fick relativamente al trasporto di materia attraverso membrane biologiche []:

$$\frac{d[A]}{dt} = K_A \frac{([A_{ext}] - [A])}{\sqrt[3]{V}} \quad (4.5)$$

Definito questo schema è possibile calcolare l'andamento delle concentrazioni delle sostanze nel tempo, ora è necessario mettere in relazione queste sostanze con la membrana della protocellula, definendo come essa viene alimentata e sotto quali condizioni si può duplicare.

4.3 Alimentazione della membrana e Duplicazione

Le GMM sono le specie chimiche che determinano la velocità di crescita della membrana e fungono da catalizzatori per la produzione di lipidi, indicando quindi con C la quantità totale di lipidi della membrana l'equazione del tasso di crescita della protocellula ha la seguente forma:

$$\frac{dC}{dt} = \sum_{i=1}^N K_i [X_i] V \quad (4.6)$$

dove N è il numero totale di specie, V è il volume interno della protocellula e $[X_i]$ è la concentrazione della specie interna; i coefficienti K_i saranno uguali a 0 quando la sostanza non alimenta la membrana. Anche in questo caso, come nelle reazioni, si presuppone che i lipidi prodotti internamente si aggregino istantaneamente alla membrana. La 4.6 presuppone un volume costante e va ad agire direttamente sulla quantità di lipide anziché sulla sua concentrazione.

Ad ogni istante di tempo vengono applicate le equazioni delle reazioni, le concentrazioni delle sostanze variano e la membrana viene alimentata dalle GMMs, questo accade fino a quando la protocellula non è cresciuta abbastanza per potersi dividere in due cellule figlie. La condizione di duplicazione all'istante di tempo t_x , in questo modello, è definita dal rapporto tra la quantità di lipide al tempo t_x e la quantità di lipide al tempo t_0 :

$$\frac{C_{t_x}}{C_{t_0}} \geq \text{coefficiente} \quad (4.7)$$

il *coefficiente* è arbitrario ma solitamente si fa avvenire la duplicazione quando il lipide raddoppia (*coefficiente* = 2) rispetto alla quantità iniziale. Quando questa condizione viene verificata la quantità di lipide viene dimezzata ($C_{new} = C_{t_x}/2$) e viene ricalcolato il volume usando la 4.2 con $C = C_{new}$.

4.4 Struttura generale del software

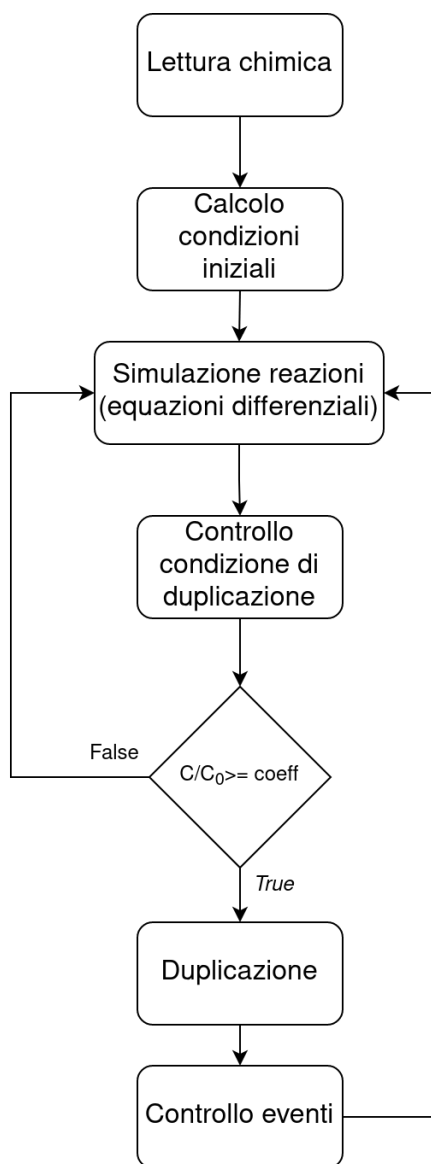


FIGURA 4.1: Diagramma di flusso del programma

In figura 4.1 viene mostrato il diagramma di flusso del programma, passando per le operazioni principali necessarie a simulare la chimica di una protocellula. Nel diagramma tuttavia si omettono alcuni dettagli che vengono discussi di seguito.

4.4.1 Lettura chimica

La lettura dei dati avviene da un file testuale contenente una lista di specie con relativa concentrazione iniziale e una lista di reazioni, questi due formano la chimica della protocellula che si vuole simulare. Per ogni specie si indica inoltre se essa è bufferizzata, quindi se ha concentrazione costante, e il coefficiente di interazione con la membrana ovvero il K_i della 4.6. Ogni reazione scritta su file segue la seguente nomenclatura:

- 23 BBB + ABB > B + BB + ABB ; 1000.0

il numero intero iniziale è un codice che definisce il tipo di reazione, in questo caso si tratta di un cleavage e il "2" indica i 2 reagenti mentre il "3" i 3 prodotti, il numero finale è il coefficiente di reazione, il K della 4.3. Non c'è un riferimento esplicito per il catalizzatore, anche se è facile identificarlo in una reazione guardando se la stessa specie compare sia nei reagenti che nei prodotti. Maggiori dettagli sono presenti in appendice A.2.

4.4.2 Calcolo condizioni iniziali

Sempre dal file della chimica vengono lette le informazioni necessarie per calcolare la geometria della cellula e la quantità di lipide iniziale. I dati iniziali sono lo spessore della membrana (δ), il raggio della cellula (r) e la densità (ρ) del materiale lipidico, da questi si può ricavare il volume della cellula e la quantità iniziale di lipide con la seguente formula:

$$C_0 = \rho \left(\frac{4}{3} \pi (r + \delta)^3 - \frac{4}{3} \pi r^3 \right) \quad (4.8)$$

4.4.3 Simulazione reazioni

Questa è la parte in cui effettivamente inizia la simulazione della chimica, il calcolo delle equazioni viene effettuato mediante un'apposita libreria esterna di calcolo differenziale che prende in ingresso un array contenente le concentrazioni iniziali di tutte le specie e una funzione di cambiamento. L'array rappresenta lo stato iniziale del problema (y_0) e oltre alle concentrazioni delle specie contiene la quantità di lipide della cellula, infatti anch'esso deve variare seguendo la 4.6. La funzione di cambiamento incorpora la logica di simulazione delle reazioni e ad ogni passo effettua le seguenti operazioni:

1. Viene calcolato il volume attuale della cellula V_t usando la 4.2 e viene anche calcolato il volume al passo precedente V_{t-1} , l'array y_0 infatti contiene la quantità di lipide al tempo t e quella al tempo $t - 1$.
2. Si calcola il rapporto $\frac{V_{t-1}}{V_t}$ e si ricalcolano le concentrazioni delle sostanze in base a esso, questa operazione corrisponde alla diluizione delle sostanze in seguito ad un aumento di volume.
3. Si simulano tutte le reazioni, l'ordine in cui queste vengono applicate non è rilevante poiché l'output di questa funzione è un vettore contenente le variazioni totali di tutte le specie.
4. Infine si calcola la variazione della quantità di lipide con la 4.6.

4.4.4 Duplicazione

La simulazione delle reazioni va avanti fino a quando non si verifica la condizione di duplicazione 4.7, questa viene verificata tramite una struttura di controllo degli eventi messa a disposizione dalla libreria esterna. In questo caso si tratta di un evento terminale che mette fine al calcolo delle reazioni ma è possibile anche inserire degli eventi non terminali. Verificata quindi la condizione si dimezza la quantità di lipide e si ricalcola il volume. Benché si parli di *duplicazione* non vengono effettivamente generate due cellule figlie ma si prosegue col simulare una cellula sola, questo perché l'obiettivo principale del software non è simulare una *popolazione* di protocellule ma bensì la chimica di singoli lignaggi (linee di discendenza genitore-figlio). Il risultato della duplicazione sarà quindi una cellula con la stessa chimica della madre (stesse reazioni e stesse concentrazioni) ma metà del materiale lipidico, spesso si indica anche come *generazione* di una cellula l'intervallo di tempo tra l'inizio della simulazione delle sue reazioni e la duplicazione.

4.4.5 Controllo eventi

In seguito alla duplicazione si interrompe l'esecuzione della libreria esterna di calcolo differenziale, viene quindi impostato un nuovo problema in cui le condizioni iniziali y_0 equivalgono ad un array contenente le concentrazioni della cellula appena divisa e la quantità dimezzata di materiale lipidico. Prima di riprendere la simulazione delle reazioni si controlla se ci sono da applicare degli eventi, questi vengono letti all'inizio da file e ne sono previsti due tipi: la modifica dei coefficienti di reazione e la variazione della concentrazione delle specie. Per ogni simulazione si possono introdurre un numero arbitrario di eventi indicando l'istante di tempo t in secondi nel quale si vuole far accadere i cambiamenti, tuttavia l'evento non avverrà esattamente in t ma verrà applicato alla prima divisione successiva a t . Questo modo di gestire gli eventi è legato al fatto che la libreria esterna non prevede la possibilità di modificare a runtime i valori delle concentrazioni, si è quindi optato per effettuare gli eventi tra una generazione e l'altra. Per esempio se un evento dovrebbe avvenire a $t = 300s$ ed una generazione dura da 0 a 450 esse verrà applicato al secondo 450.

In questo capitolo si è descritto a grandi linee il modello della protocellula utilizzato e il software che lo gestisce, per ulteriori dettagli relativi al programma si rimanda all'appendice A.

Capitolo 5

Esperimenti e risultati

Lo scopo principale di questi esperimenti è mostrare la possibilità di utilizzare il software per simulare chimiche di protocellule in ambienti statici e dinamici, per far ciò si prendono in esame due scenari e si usano le simulazioni per provare a capire i comportamenti delle chimiche.

5.1 Chimica bi e tri-molecolare

Il primo esperimento proposto è il confronto tra chimiche bi (DUO) e tri-molecolari (TRI). Come è stato descritto nella sezione 3.2.3 le reazioni tri-molecolari vengono scomposte in più reazioni poiché è più probabile la collisione tra due molecole rispetto a tre, è tuttavia utile analizzare il comportamento di una medesima chimica scritta nei due modi diversi. Per fare questo si è partiti da una chimica abbastanza semplice (12 specie e 11 reazioni - si noti che le chimiche con reazioni bimolecolari introducono un più alto numero di specie chimiche rispetto all'approccio trimolecolare: le entità aggiuntive sono i "complessi" intermedi, tipicamente oggetti a emivita molto breve) e la si è simulata in entrambi i modi per 20 generazioni.

A_EXT	0.01	1	0	A_EXT	0.01	1	0
B_EXT	0.01	1	0	B_EXT	0.01	1	0
AA_EXT	0.01	1	0	AA_EXT	0.01	1	0
BB_EXT	0.01	1	0	BB_EXT	0.01	1	0
A	0.01	0	0	A	0.01	0	0
B	0.01	0	0	B	0.01	0	0
AA	0.01	0	0	AA	0.01	0	0
BB	0.01	0	0	BB	0.01	0	0
AAA	0.01	0	0.01	AAA	0.01	0	0.01
BBB	0.01	0	0.01	BBB	0.01	0	0.01
*AABBB	0	0	0				
*BBAAA	0	0	0	210	A_EXT / A ; 8.00E-17		
				210	B_EXT / B ; 8.00E-17		
210	A_EXT / A ; 8.00E-17			210	AA_EXT / AA ; 8.00E-17		
210	B_EXT / B ; 8.00E-17			210	BB_EXT / BB ; 8.00E-17		
210	AA_EXT / AA ; 8.00E-17			32	AA + A + BBB > AAA + BBB ; 2000		
210	BB_EXT / BB ; 8.00E-17			32	BB + B + AAA > BBB + AAA ; 2000		
21	AA + BBB > *AABBB ; 2000			23	BBB + BBB > BB + B + BBB ; 1000		
12	*AABBB > AA + BBB ; 14.86						
22	*AABBB + A > AAA + BBB ; 1000						
21	BB + AAA > *BBAAA ; 2000						
12	*BBAAA > BB + AAA ; 14.86						
22	*BBAAA + B > BBB + AAA ; 1000						
23	BBB + BBB > BB + B + BBB ; 1000						

FIGURA 5.1: La chimica in forma bi-molecolare (a sinistra) e tri-molecolare (a destra)

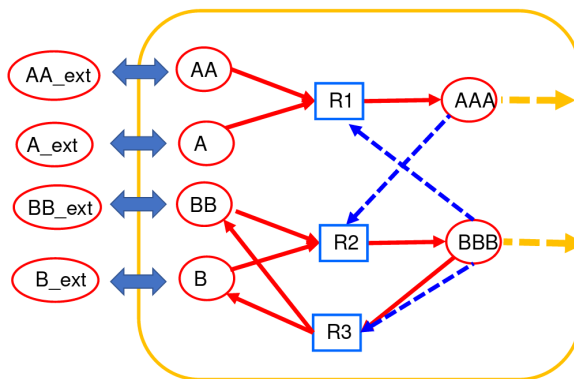


FIGURA 5.2: Schema della chimica

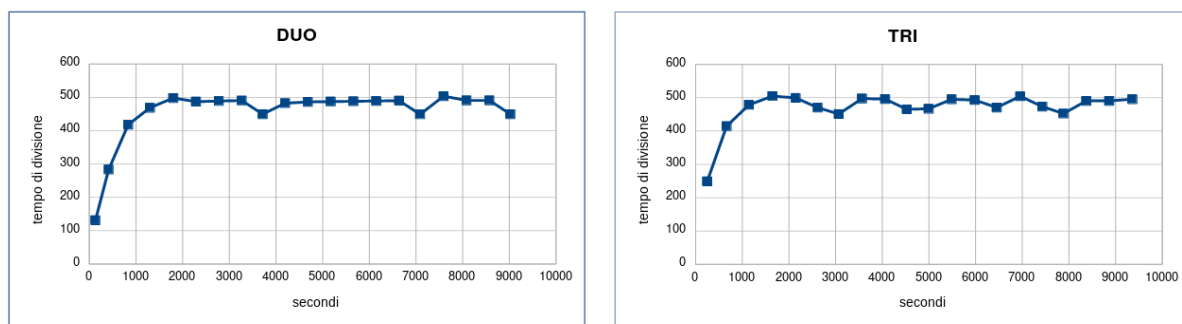


FIGURA 5.3: Tempi di divisione per i casi bi e tri molecolare

I parametri dei due approcci simulativi (basati rispettivamente su reazioni bimolecolari e trimolecolari) sono stati settati in modo da ottenere andamenti del tempo di divisione simili, a regime più o meno stabili nell'intervallo $[400, 500]$ secondi (da qui ed in seguito per comodità useremo il termine "secondi" - si tratta in realtà di unità arbitrarie di tempo). In figura 5.4 sono invece presenti i tempi di divisione per una chimica più complessa (142 specie e 265 reazioni), in cui abbiamo utilizzato i parametri trovati nella chimica precedente. In questo caso la versione TRI si estingue mentre la DUO sebbene dopo molte generazioni si stabilizza. Questo lo si può vedere bene osservando gli andamenti delle concentrazioni delle specie (5.5) che sono state suddivise in base alla loro lunghezza. In queste chimiche e in tutte quelle seguenti le specie di lunghezza 1 e 2 hanno il rispettivo bufferizzato esterno e vengono quindi continuamente alimentate.

Nella figura 5.6 si è invece usata una chimica scritta dal generatore di chimiche discusso nell'appendice B, anche in questo caso i tempi di divisione sono diversi sia nei valori che nell'andamento ed anche le specie (5.7) si comportano in modo differente.

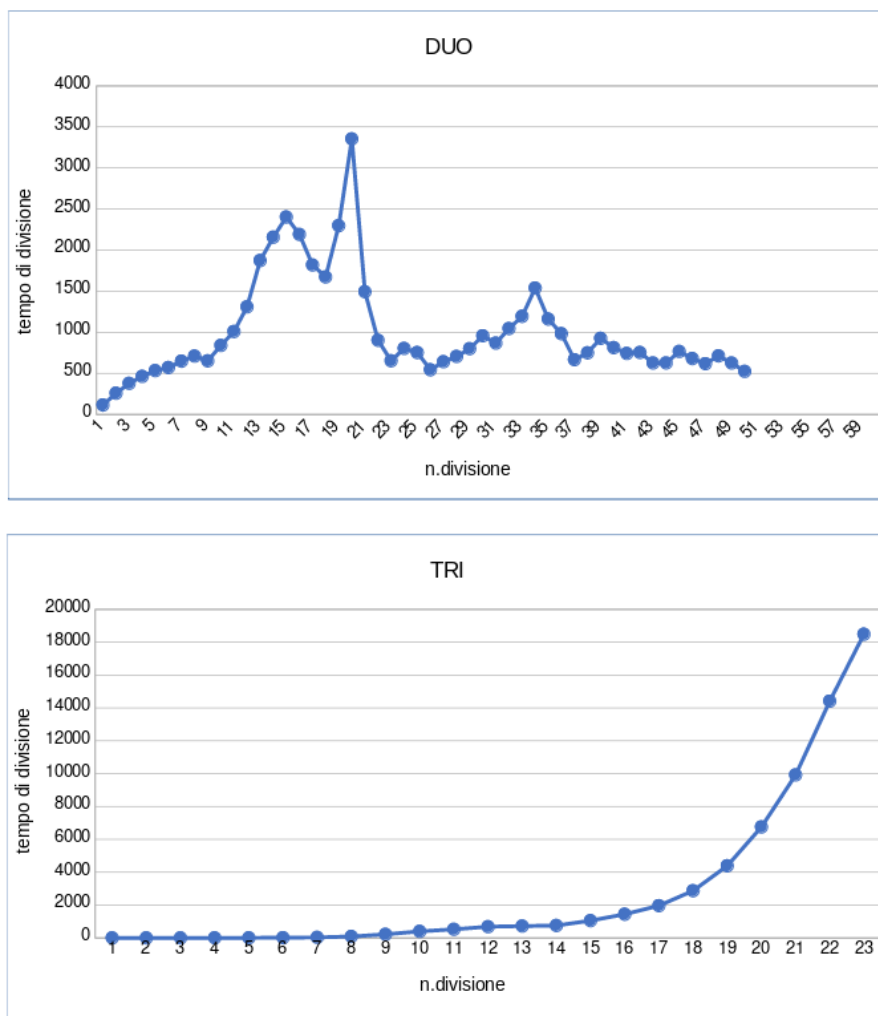


FIGURA 5.4: Tempi di divisione per i casi bi e tri molecolare

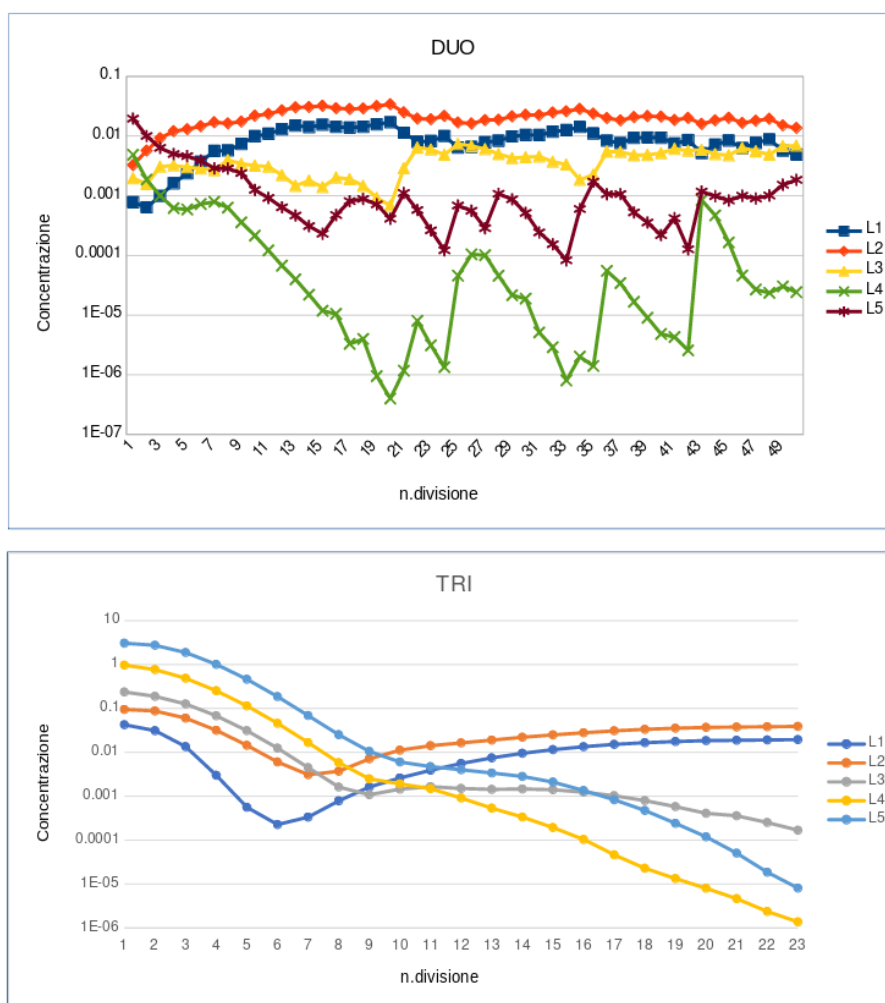


FIGURA 5.5: Concentrazione delle specie suddivise per lunghezza, in ogni lunghezza sono state sommate le concentrazioni delle corrispondenti specie chimiche (scala logaritmica)

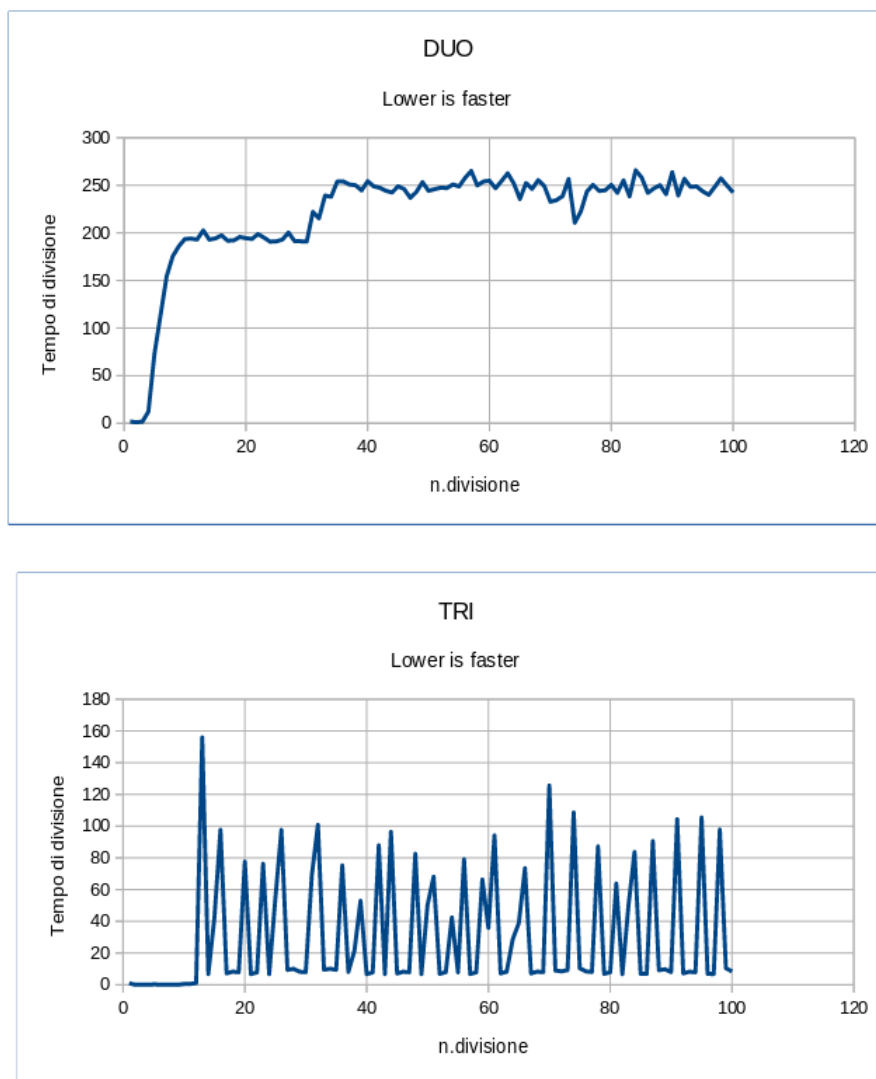


FIGURA 5.6: Tempi di divisione per i casi bi e tri molecolare della chimica 0

Quest'ultima chimica, che per semplicità viene chiamata *chimica_0* (76 specie e 86 reazioni), verrà analizzata maggiormente nella sezione successiva dove verranno introdotti gli eventi di cambio dei coefficienti di reazione, va tuttavia sottolineato come essa presenti delle forti oscillazioni nell'andamento dei tempi di divisione e nell'andamento delle specie, nonostante riesca comunque a trovare una sua stabilità. Ciò che appare finora è che in presenza di chimiche pressoché stabili le versioni bi e tri-molecolari seguono gli stessi andamenti, quando invece si hanno situazioni più complesse esse divergono fortemente. Le cause di questo fenomeno possono essere diverse:

1. Si tratta di un fenomeno numerico introdotto dalla libreria di integrazione nello svolgere il sistema di equazioni differenziali
2. Il fenomeno è reale ma è dovuto ad una particolare scelta delle condizioni iniziali

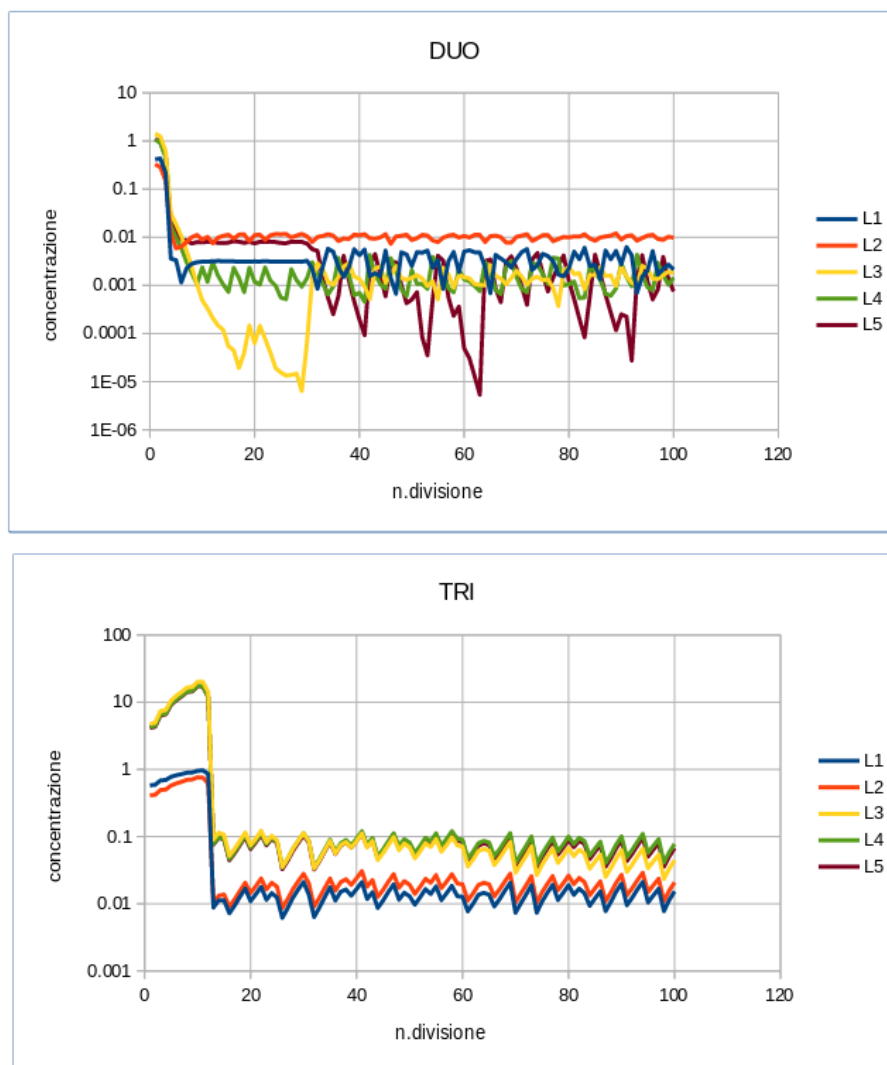


FIGURA 5.7: Concentrazione delle specie della chimica 0 suddivise per lunghezza, in ogni lunghezza sono state sommate le concentrazioni delle corrispondenti specie chimiche (scala logaritmica)

3. Si tratta di un fenomeno reale dovuto agli effetti di sequestro svolti dai complessi chimici
4. È un fenomeno reale dovuto alla composizione di reazioni chimiche aventi diverso ordine di reazione

Tra queste conclusioni le ultime due sarebbero le più interessanti, tuttavia servirebbero più simulazioni per avere un quadro migliore della situazione. Come fatto preliminare sembra che in generale le chimiche DUO siano più stabili e quindi più facili da studiare, per questo nei seguenti esperimenti si è scelto di usare solamente questa categoria nonostante le chimiche risultino più grandi e più pesanti da simulare.

5.2 Chimiche in ambiente variabile

In questa sezione vengono discussi i risultati delle simulazioni di due chimiche, entrambe generate con il software discusso in appendice B, con l'introduzione di un ambiente dinamico. Per semplicità verranno chiamate rispettivamente *chimica_0* e *chimica_1*. Entrambe sono state generate nella loro forma bi-molecolare con 20 cleavage, 20 condensazioni e una lunghezza massima delle specie pari a 5, la configurazione del generatore è la stessa mostrata in figura B.1. Tutte le simulazioni sono state effettuate usando il metodo di integrazione RK45. La prassi che è stata seguita per effettuare gli esperimenti è la seguente:

1. Sono state generate una serie di chimiche e le si sono simulate per 100 generazioni.
2. Dopo queste 100 generazioni si sono analizzati i risultati con l'obiettivo di cercare delle chimiche con tempi di divisione stabili.
3. Trovate queste chimiche (*chimica_0* e *chimica_1*) sono state simulate altre 100 generazioni in modo da avere una base di confronto per i successivi esperimenti. Non si è voluto simulare 200 generazioni in un unico lancio per avere una mole di dati più facilmente trattabile.
4. A partire dal punto stabile trovato nel secondo passaggio si sono introdotti gli eventi di cambio dei coefficienti di reazione, simulando sempre 100 generazioni.

Per questi eventi si sono effettuati una serie di tentativi per capire quali fossero i giusti coefficienti da applicare, un'altra variabile è invece l'intervallo di tempo che intercorre tra i vari eventi. A riguardo gli intervalli di tempo testati sono 500, 1000 e 2000 secondi. La modifica dei coefficienti di reazione rappresenta un cambiamento nelle condizioni dell'ambiente, si è quindi intervenuti aumentando e diminuendo in opposizione i coefficienti dei cleavage e delle condensazioni. Anche in questo caso si sono effettuate diverse prove:

- I coefficienti raddoppiano e si dimezzano ad ogni evento (*variazione piccola*)
- I coefficienti aumentano e diminuiscono di 4 volte ad ogni evento (*variazione intermedia*)
- I coefficienti aumentano e diminuiscono di 8 volte ad ogni evento (*variazione sostanziale*)

Quindi, si alternano eventi in cui vengono favorite le condensazioni e disincentivati i cleavage (corrispondenti a situazioni di penuria di acqua: "siccità") e viceversa (la presenza di acqua disincentiva i legami peptidici). Per quanto riguarda le condensazioni, trattandosi di chimiche bi-molecolari, è stato modificato solo l'ultimo coefficiente, per esempio in una reazione del tipo 5.8:

```

21      B + AAA > *BAAA ; 2000.0
12      *BAAA > B + AAA ; 14.86
22      *BAAA + BAA > BBAA + AAA ; 1000.0

```

FIGURA 5.8: Reazione bi-molecolare

che equivale alla tri-molecolare $B + BAA + AAA \rightarrow BBAA + AAA$ solo l'ultimo coefficiente viene modificato. In totale sono stati introdotti 21 eventi per ogni simulazione e si è scelto di focalizzarsi sulla variazione intermedia dei coefficienti, il ventunesimo evento riporta i coefficienti delle sostanze ai valori originari.

5.2.1 Chimica 1

La prima chimica analizzata è la *chimica_1* poiché il tempo di duplicazione risulta molto più stabile della *chimica_0* ed è quindi più facile fare le considerazioni sull'andamento delle specie. In figura 5.9 sono presenti i risultati della *chimica_1* in assenza di eventi, i tempi di divisione sono abbastanza stabili e le specie di lunghezza 5 sono praticamente assenti, si ricorda che le specie di lunghezza 1 e 2 (*L1* e *L2*) sono il cibo della chimica e vengono continuamente alimentate dalle rispettive specie esterne bufferizzate con concentrazione costante 0,01.

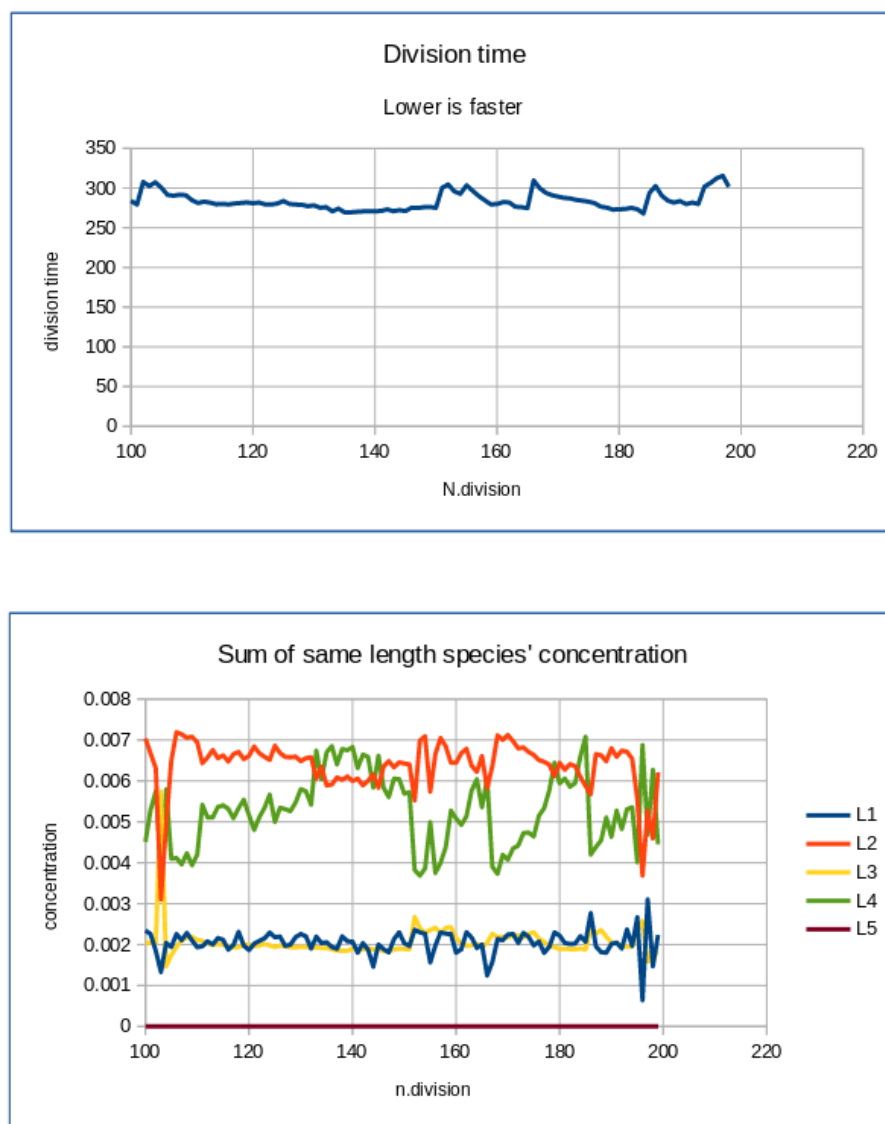


FIGURA 5.9: Tempi di divisione e andamento delle concentrazioni della chimica 1

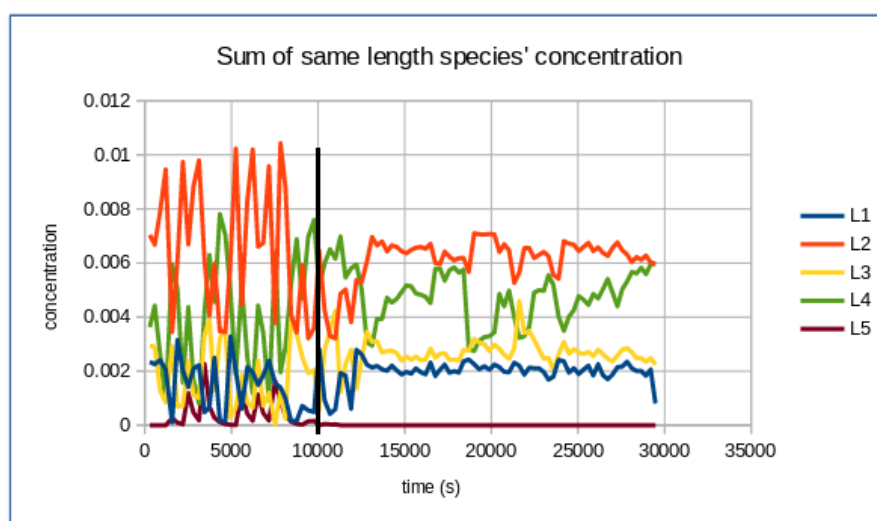
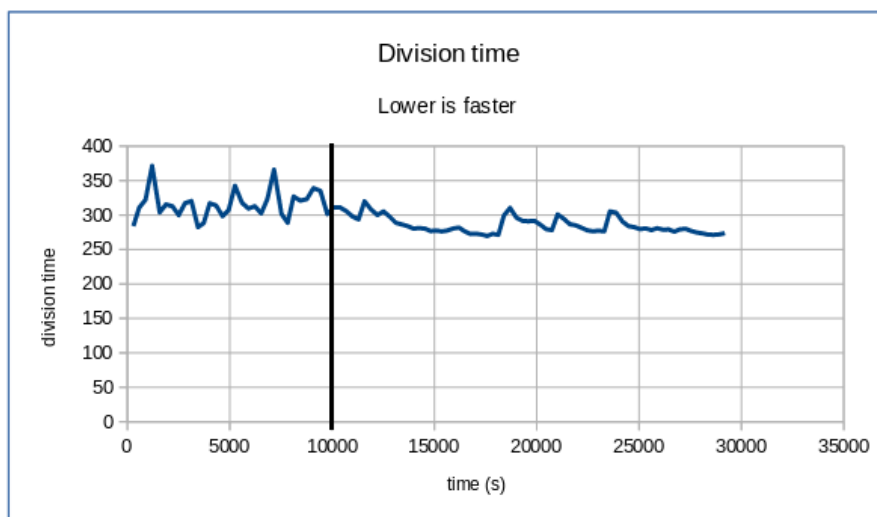


FIGURA 5.10: Tempi di divisione e andamento delle concentrazioni della chimica 1 con eventi ogni 500 secondi, in corrispondenza della linea verticale terminano gli eventi, esattamente al secondo 10500

L'introduzione di eventi ogni 500 secondi apporta evidenti cambiamenti sia ai tempi di divisione (5.10), che ora oscillano leggermente, che all'andamento delle sostanze le quali seguono le oscillazioni imposte dagli eventi esterni, inoltre le specie L5 riescono a rimanere a concentrazioni apprezzabili. Dal secondo 10500 cessano gli eventi e il sistema torna approssimativamente allo stato base.

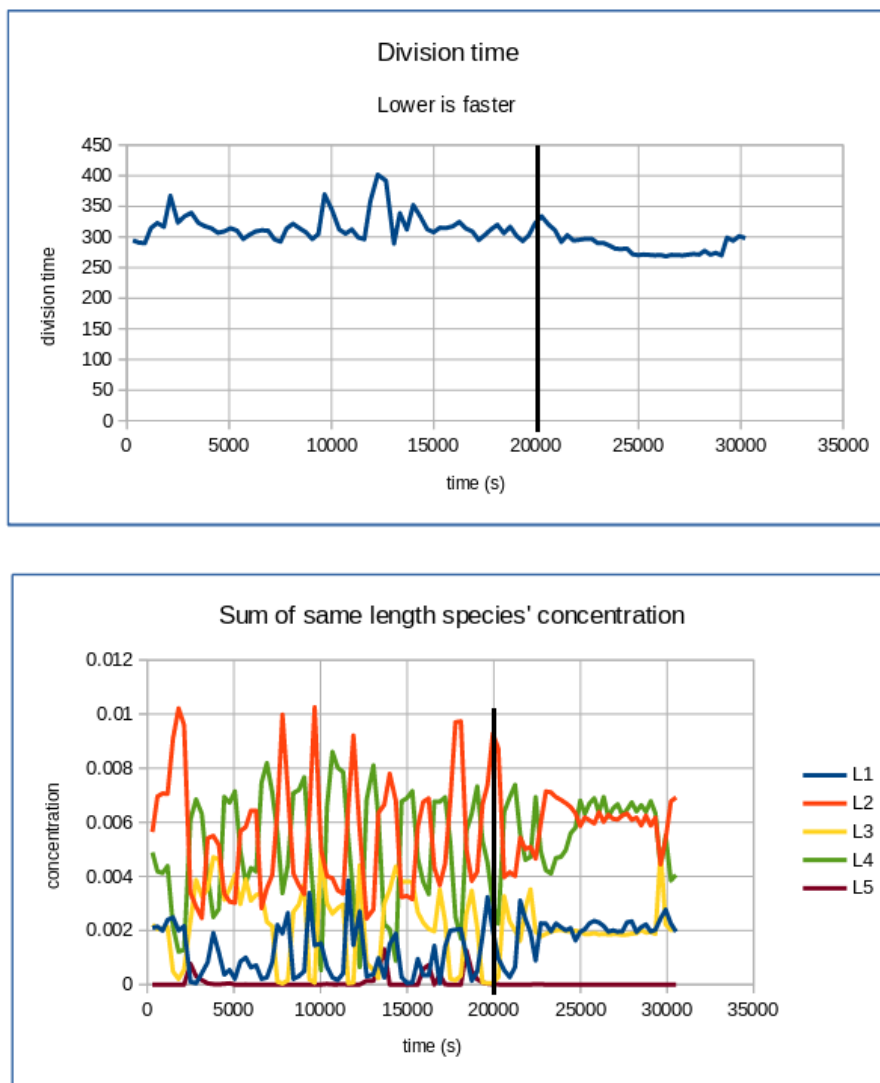


FIGURA 5.11: Tempi di divisione e andamento delle concentrazioni della chimica 1 con eventi ogni 1000 secondi, in corrispondenza della linea verticale terminano gli eventi, esattamente al secondo 21000

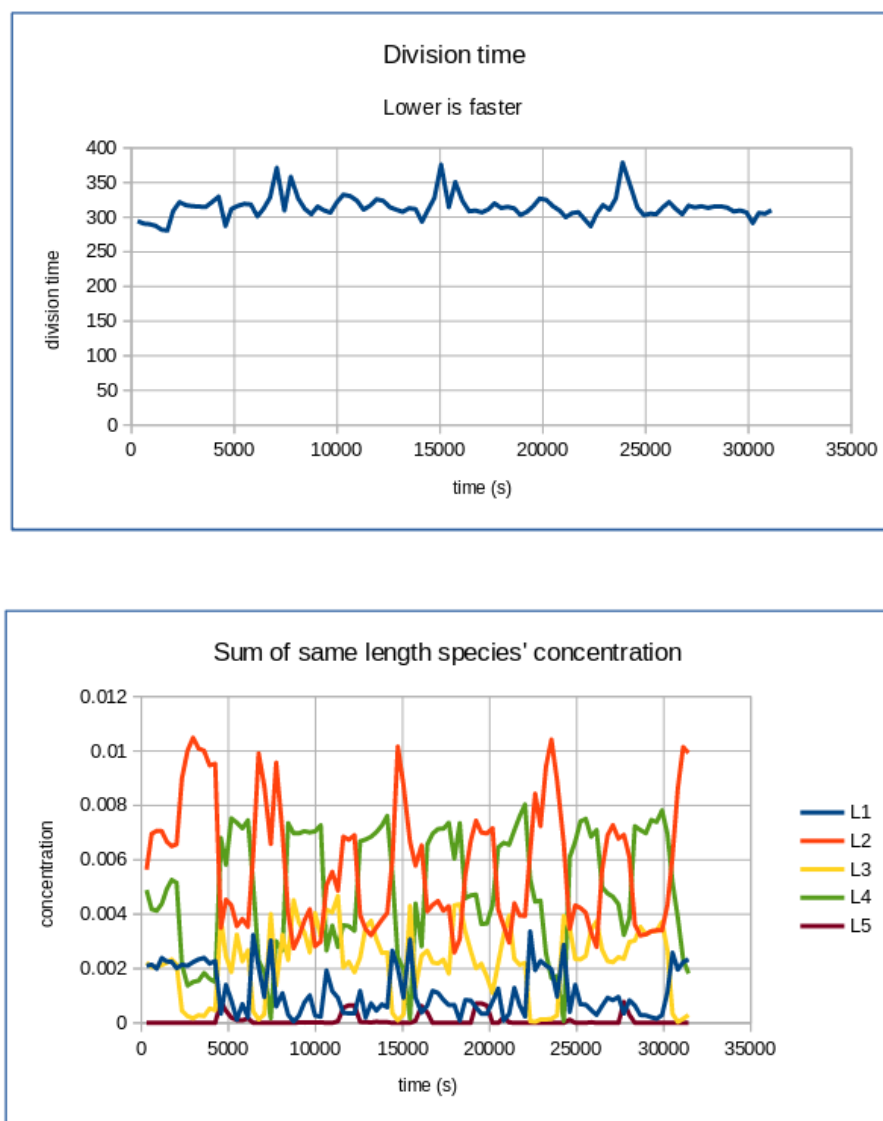


FIGURA 5.12: Tempi di divisione e andamento delle concentrazioni della chimica 1 con eventi ogni 2000 secondi, in questo caso gli eventi non terminano ed il sistema non torna allo stato base

Le considerazioni rimangono in linea generale le stesse anche passando ad intervalli di tempo più ampi, infatti sia per gli intervalli di 1000 (5.11) e 2000 (5.12) secondi i tempi di divisione rimangono stabili nell'intervallo $[250, 400]$ secondi, va sottolineato che in quest'ultimo caso non tutti gli eventi vengono applicati poiché la simulazione finisca prima. Si può fare un'ulteriore osservazione guardando le chimiche in un ambiente in cui vengono costantemente favorite le condensazioni o i cleavage. Questi due casi sono stati chiamati *wet* e *dry*, nel primo i coefficienti delle condensazioni precedentemente indicati sono stati impostati costantemente a 4000 mentre i cleavage a 250, nel *dry* avviene l'opposto. L'idea è che le concentrazioni delle specie chimiche corrispondenti a questi due ambienti estremi dovrebbero essere costantemente ai livelli massimi, per il *wet*, e minimi, per il *dry*, delle oscillazioni viste nell'ambiente variabile.

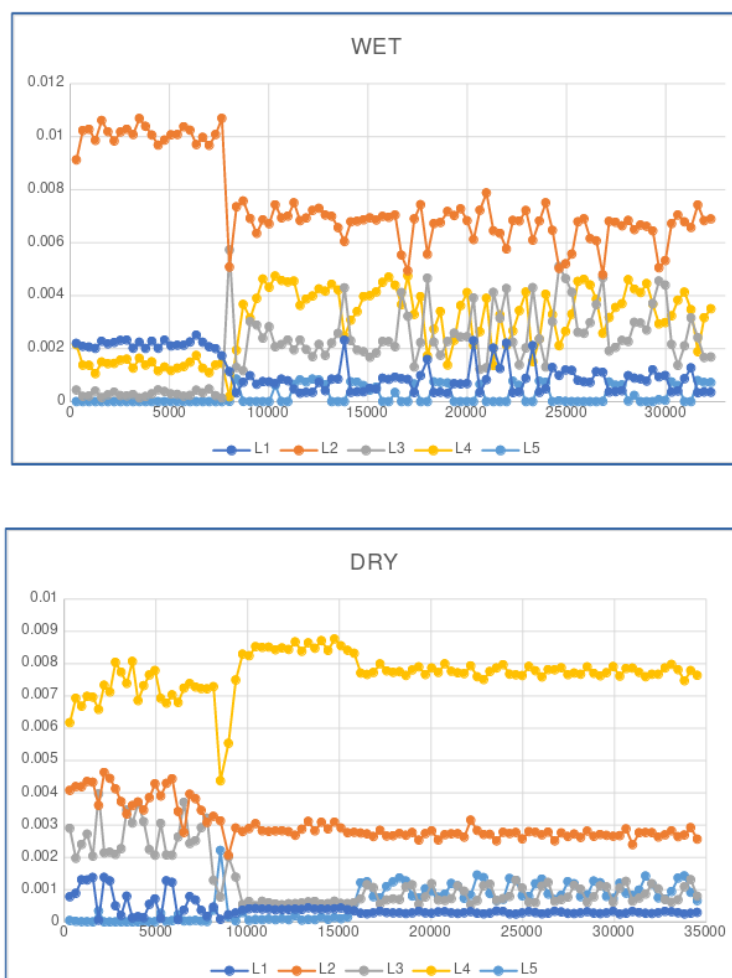


FIGURA 5.13: Andamento delle concentrazioni della chimica 1 con ambiente wet e dry

La situazione è tuttavia più complessa, infatti si osserva (5.13) come in entrambi i casi si vengano a creare più situazioni stabili, la prima nel periodo $[0, 8000]$ secondi e la seconda da 8000 in avanti. La considerazione precedente è comunque in parte vera, infatti guardando le L2 nel caso *wet* hanno un massimo in 0,01 nel primo intervallo, proprio lo stesso valore visto nei picchi delle simulazioni con eventi. Una considerazione simile la si può fare per le L4 che nel caso *dry* si stabilizzano intorno

a 0,008 come i massimi locali osservabili bene in figura 5.12.

5.2.2 Chimica 0

Come detto precedentemente la *chimica_0* è più complessa, presenta infatti delle forti oscillazioni già nelle 100 generazioni senza eventi (5.14). Questo lo si può vedere bene nei tempi di divisione e soprattutto nell'andamento delle specie che compongono il cibo (L1 e L2), queste oscillazioni si estendono oltre la vita di una singola protocellula ed a differenza della *chimica_1* le L5 sono molto presenti. Un'altra particolarità di questa chimica è che il cibo di lunghezza 2 (L2) assume talvolta concentrazioni maggiori al corrispettivo esterno bufferizzato, che ha concentrazione costante uguale a 0,01, quando ciò accade il cibo interno viene rigettato all'esterno per effetto delle reazione di *passaggio di membrana*. Questo può avere diversi significati, in primis può essere che le specie L2 siano maggiormente sintetizzate dalla chimica stessa ma è anche possibile che quest'ultima faccia in generale meno affidamento su questo cibo e più su L1.

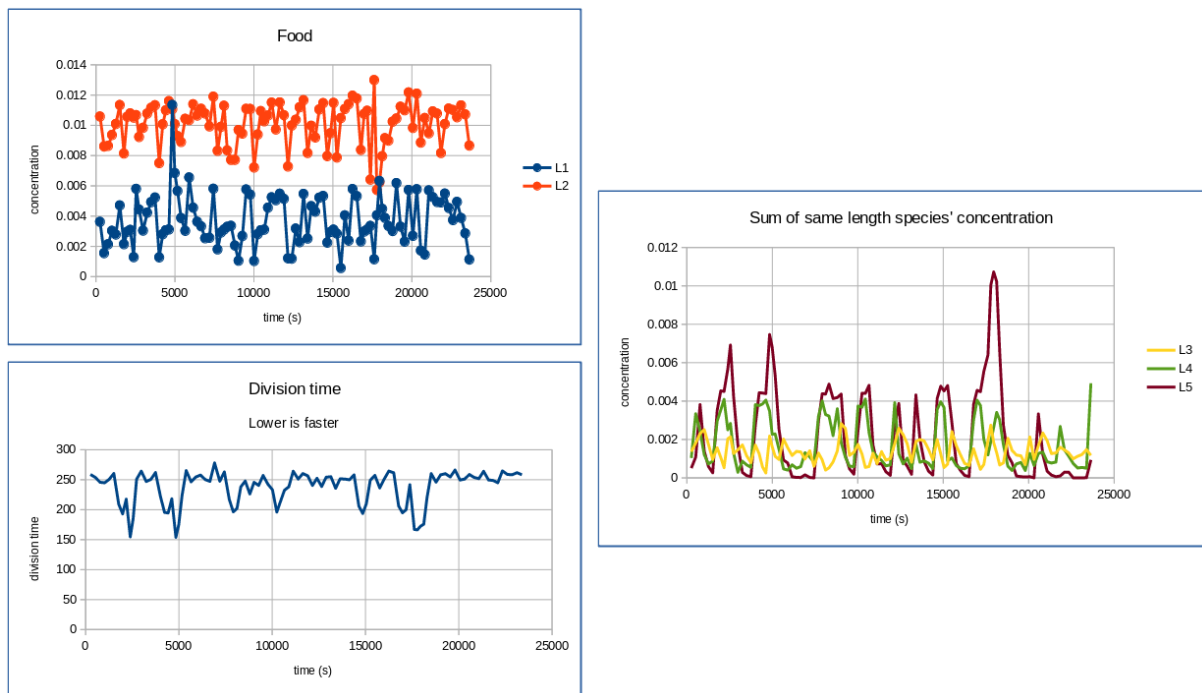


FIGURA 5.14: Andamento dei tempi di divisione e delle concentrazioni della chimica 0

Le considerazioni che si possono fare in seguito all'introduzione degli eventi sono a grandi linee le stesse fatte per la *chimica_1*, infatti si può ben vedere come il cambiamento dell'ambiente alteri l'andamento delle concentrazioni (5.15). La chimica recupera sempre il suo stato base, questo lo si può notare bene osservando come nelle L2 ci sia un forte picco intorno al secondo 17000 sia nel caso base che in quello con eventi ogni 500 secondi.

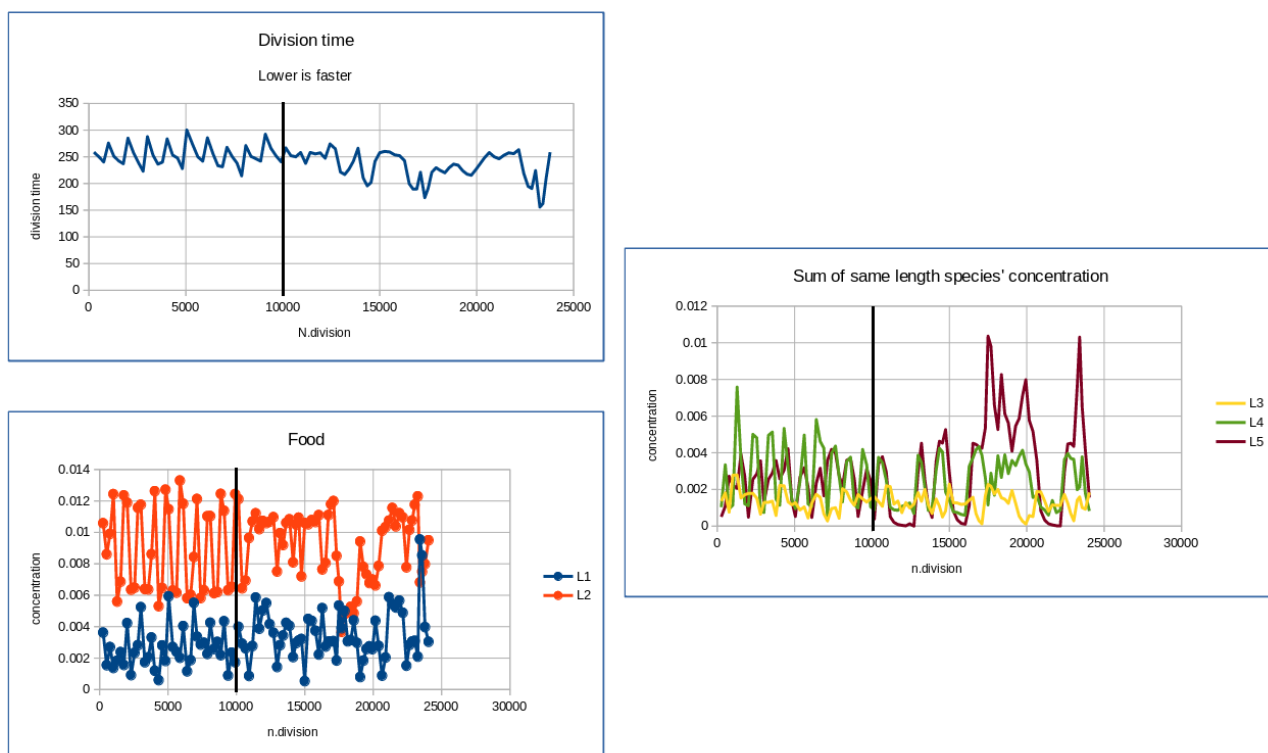


FIGURA 5.15: Tempi di divisione e andamento delle concentrazioni della chimica 0 con eventi ogni 500 secondi, in corrispondenza della linea verticale terminano gli eventi, esattamente al secondo 10500

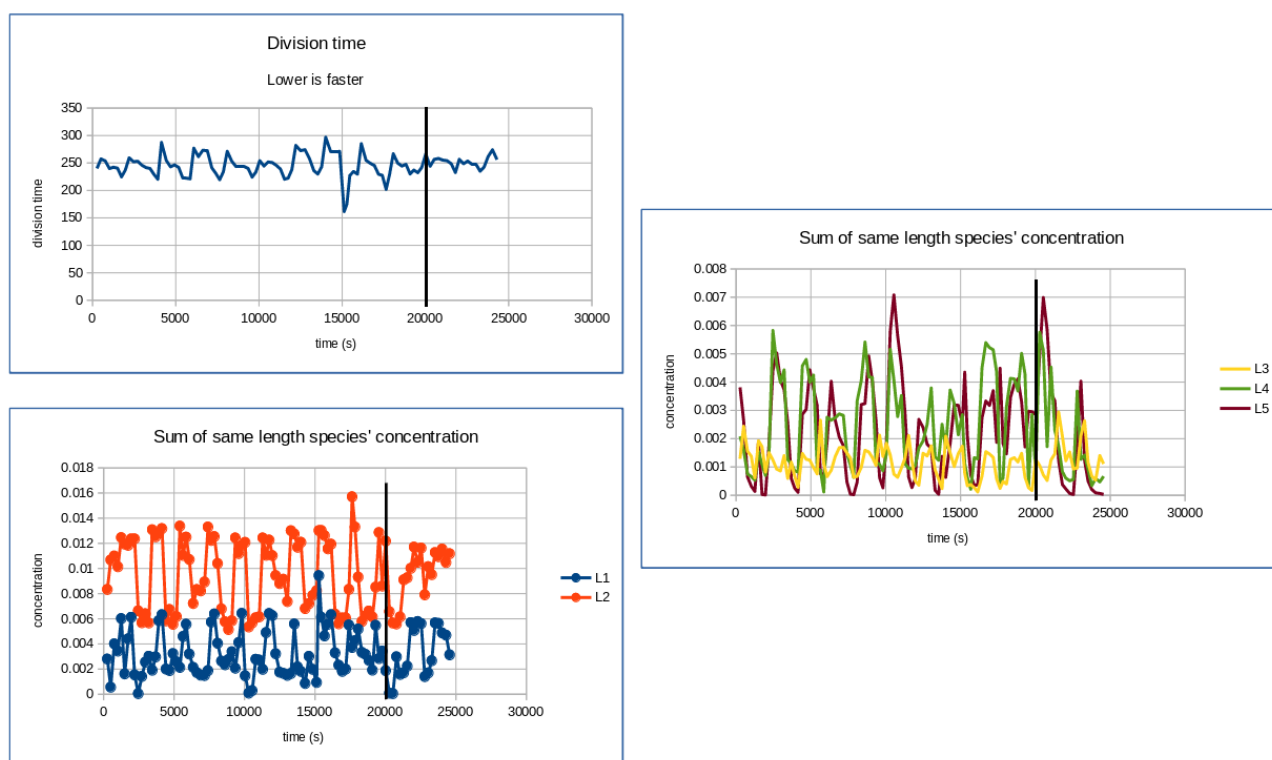


FIGURA 5.16: Tempi di divisione e andamento delle concentrazioni della chimica 0 con eventi ogni 1000 secondi, in corrispondenza della linea verticale terminano gli eventi, esattamente al secondo 21000

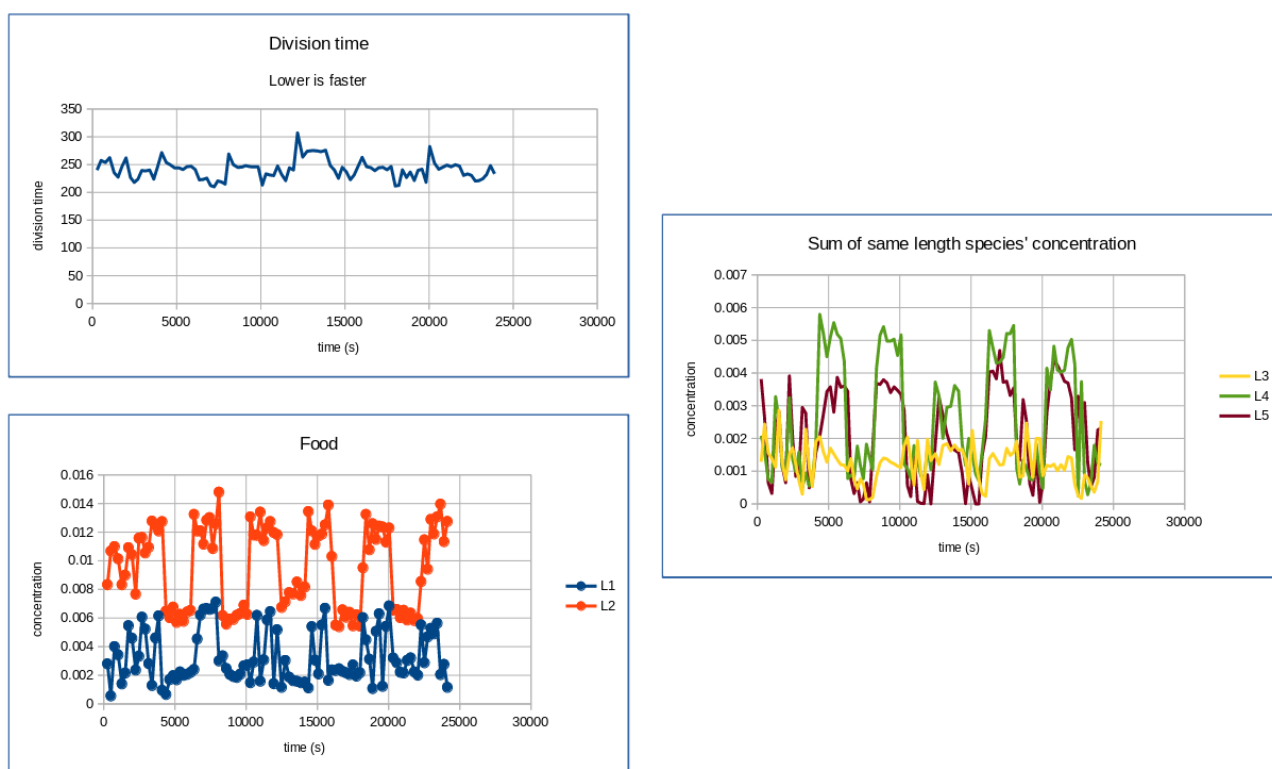


FIGURA 5.17: Tempi di divisione e andamento delle concentrazioni della chimica 0 con eventi ogni 2000 secondi, in questo caso gli eventi non terminano ed il sistema non torna allo stato base

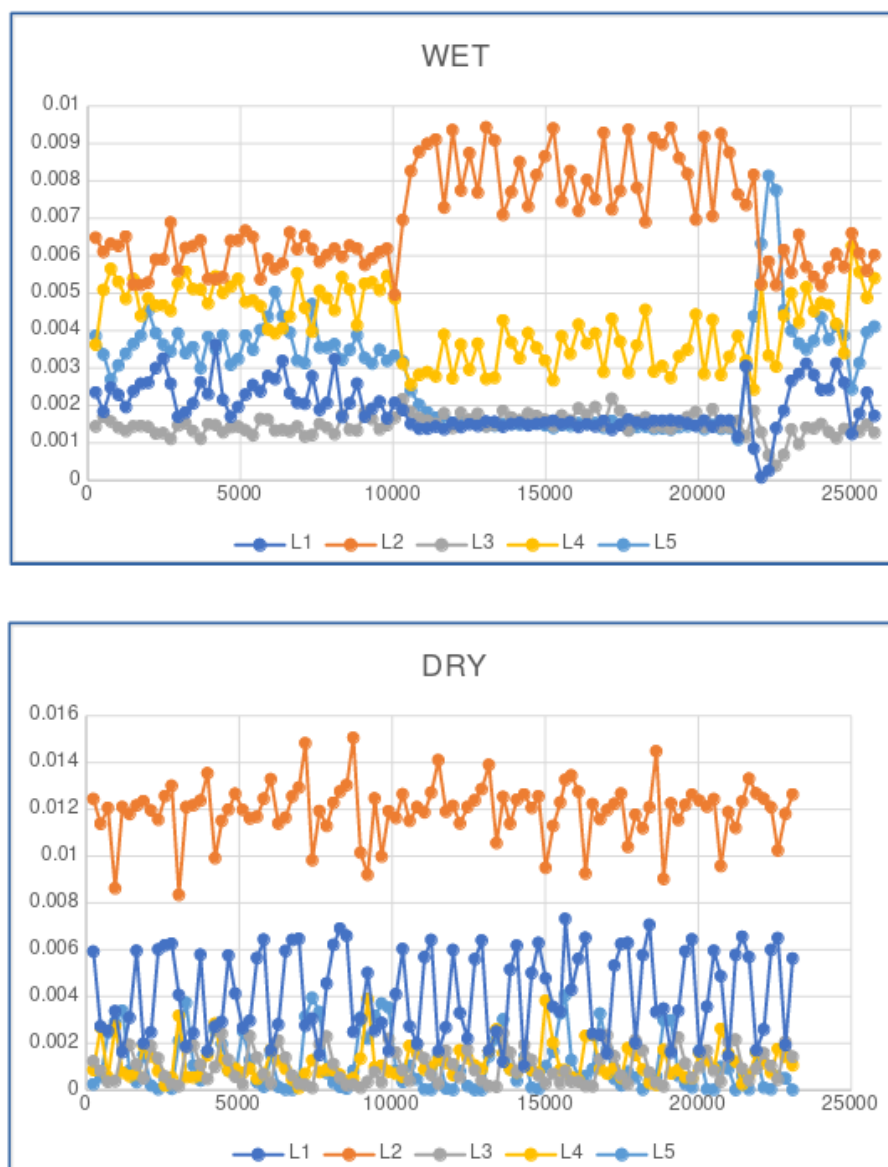


FIGURA 5.18: Andamento delle concentrazioni della chimica 0 con ambiente wet e dry

Le simulazioni *wet* e *dry* sono state fatte con gli stessi presupposti della *chimica_1*. Nel caso *dry* non si osservano comportamenti particolarmente stabili, infatti similmente al caso base le oscillazioni continuano ad essere molto forti per tutte le specie. Guardando il caso *wet* si delineano due situazioni diverse che si alternano tra di loro, una prima in cui persistono le solite oscillazioni e una seconda dove cambiano nettamente i valori assoluti delle concentrazioni con le L2 che aumentano in controtendenza alle altre specie. Le osservazioni riguardo l'influenza dell'ambiente sulla chimica interna alla protocellula fatte nel caso della *chimica_1* sembrano quindi essere confermate.

Nel caso della *chimica_0* non è chiara l'origine delle oscillazioni del periodo di duplicazione e delle concentrazioni delle specie chimiche interne, presenti anche nella situazione senza eventi. Potrebbe essere un fenomeno realmente presente all'interno delle retroazioni presenti nel sistema, oppure essere semplicemente un fenomeno

dovuto alle approssimazioni fatte e/o alle scelte delle variazioni del passo di integrazione effettuate dall'algoritmo utilizzato. In questi ultimi casi, le conseguenze di tali approssimazioni hanno un significato molto simile a quello di variazioni casuali, un fenomeno comunque presente anche nella situazione reale. La presenza di queste oscillazioni sembra quindi poter corrispondere ad eventi reali. Un'ulteriore interessante osservazione riguarda la presenza di cambi nei livelli di stabilità della situazione *wet* vi sono infatti diversi equilibri che si alternano dopo una presenza durata molte generazioni. Occorrerà indagare sulle motivazioni (e sugli eventuali inneschi) dei cambi di stabilità.

Capitolo 6

Conclusioni

In questo scritto si è posta l'attenzione sulle protocellule, sulla loro struttura e come un modello di esse possa essere implementato tramite un software ad hoc per effettuare esperimenti e simulazioni. L'efficacia e la flessibilità dell'applicativo sviluppato sono dimostrate dall'efficienza con cui è stato possibile realizzare simulazioni di diverse chimiche in ambienti e situazioni differenti.

Il software è quindi collaudato (testato in situazioni chiave note e semplici, ed efficace nell'approccio a sistemi chimici di dimensioni relativamente rilevanti). Le indicazioni di tipo scientifico che si possono trarre sono per ora preliminari, poiché non è ovviamente possibile fare affidamento su un numero ridotto di chimiche per estrarre informazioni generali. Quanto fatto finora costituisce però un importante punto di partenza per futuri studi.

Un primo punto interessante è il fatto che le chimiche descritte tramite reazioni con base bi e trimolecolare esibiscono comportamenti differenti, con queste ultime che sono più propense ad estinguersi o ad accentuare le instabilità intrinseche della chimica stessa. Molto probabilmente non si tratta di un puro fenomeno numerico (ipotesi contrastata anche dal fatto che la differenza si ripete indipendentemente dal tipo di integratore utilizzato): l'ipotesi su cui si sta lavorando è che la descrizione trimolecolare, non modellando il sequestro dei reagenti e dei catalizzatori delle varie reazioni compiuto dai complessi, stia trascurando un fattore cinetico rilevante, foriero di stabilità. La *chimica_0* ad esempio presenta forti oscillazioni che vengono espresse in maniera molto più marcata nella sua versione tri-molecolare. Suggeriamo quindi che la rappresentazione bi-molecolare possa risultare più adeguata nel seguire il fenomeno reale.

Si sono osservati sistemi in cui le protocellule tendono ad assumere comportamenti costanti, ma anche sistemi che mostrano andamenti oscillanti quasi-periodici (nei tempi di duplicazione e/o nei livelli di concentrazione delle specie chimiche interne), con variazioni comprendenti al loro interno molte generazioni differenti. Il "rumore" causato dall'approssimazione numerica (dovuta da vari fattori inclusi la natura discreta dell'elaboratore, le "scelte" di passo dei metodi utilizzati, i feedback presenti nel sistema chimico) non è sufficiente a eliminare la regolarità del fenomeno.

La simulazione di un ambiente variabile tra situazioni di umidità e siccità altera notevolmente il comportamento della protocellula. Qui vi sono diverse osservazioni interessanti, le principali riguardando l'influenza della frequenza di tali variazioni:

- in caso di variazioni poco frequenti la protocellula ha il tempo di assestarsi nelle configurazioni presenti nei periodi "solo umidi" (concentrazioni basse

di polimeri lunghi) o “solo secchi” (concentrazioni relativamente più alte di polimeri lunghi)

- in caso di variazioni frequenti la protocellula può assumere configurazioni con relativamente alte concentrazioni di polimeri lunghi, nonostante l’alternanza con i periodi secchi
- la frequenza esterna può avere interazioni complesse con la pseudo-frequenza propria di alcuni sistemi protocellulari

La simulazione di ambienti differenti (differenti set di parametri cinetici) ha anche portato alla scoperta di situazioni in cui le protocellule sembrano passare bruscamente da una situazione apparentemente stabile - a questo punto, meta-stabile - ad un’altra situazione metastabile. Dopo numerose generazioni con relativa stabilità, il sistema recupera la situazione precedente. Nel futuro occorrerà approfondire le organizzazioni in grado di esprimere il fenomeno, e le cause di tali cambi di stabilità.

Non da ultimo è stato realizzato uno script in grado di generare chimiche con alfabeti e dimensioni arbitrarie, decisi da file di configurazione, basato sullo schema delle reazioni di polimerizzazione.

6.1 Sviluppi Futuri

Il software, così come è scritto, è pienamente utilizzabile per condurre studi di vario genere sulle chimiche di protocellule. Per come è stato strutturato, e per gli obiettivi di questo lavoro, il programma è essenzialmente uno script che prende in input una serie di file testuali e restituisce in output ulteriori file che vanno poi elaborati separatamente. L’idea è quella trasformare l’applicativo in una piattaforma tramite la quale è possibile impostare degli esperimenti e ricevere direttamente i dati di cui si ha bisogno. Tramite il generatore di chimiche si producono un numero elevato di chimiche che vengono passate direttamente al simulatore, alla fine di tutte le simulazioni la piattaforma elabora i risultati e fornisce in output i dati aggregati richiesti inizialmente, senza necessariamente dover intervenire manualmente. I vantaggi di questo approccio sarebbero molteplici:

- Ci si potrebbe discostare dalla lettura da file.
- Non si dovrebbe necessariamente spostare i risultati in un foglio di calcolo separato.
- Il generatore e il simulatore sarebbero direttamente collegati.

Quest’ultimo passaggio si potrebbe implementare in maniera provvisoria tramite uno script bash che automaticamente prende il risultato del generatore e lo passa al simulatore, per farlo sarebbero necessarie modifiche minime ai programmi.

Sempre seguendo l’idea della piattaforma unificata si potrebbe impostare il programma per simulare popolazioni di protocellule, sviluppando l’interazione tra individui e la modifica attiva dell’ambiente, questo tuttavia richiederebbe modifiche più massive al programma e renderebbe quasi necessario discostarsi dalla libreria Scipy per il calcolo differenziale. Infatti, nonostante la libreria performi abbastanza bene, vi sono dettagli non controllabili, come ad esempio la gestione di eventi particolari (cambi di concentrazione di sostanze chimiche imposti dall’esterno). Questo

fatto a sua volta può portare ad una non agevole gestione delle interazioni fra protocellule presenti nello stesso ambiente (popolazioni di protocellule).

Il lavoro ha quindi portato a risultati notevoli, ed ha generato una serie di domande interessanti.

La tesi ha quindi raggiunto i propri obiettivi.

Appendice A

Dettagli implementativi del software

La struttura del software è stata discussa generalmente nel capitolo 4.4, in questa sezione si analizzerà più nel dettaglio il programma ponendo l'attenzione su alcuni suoi dettagli implementativi.

A.1 Esempio di utilizzo

Il programma è scritto completamente in Python e può essere reperito dalla repository `git`. L'ambiente di lavoro è gestito da `pipenv` che provvede ad installare le librerie necessarie, una volta configurato il programma può essere avviato con il comando `python start.py`. Le configurazioni della simulazione si effettuano mediante il file `conf.txt` nel quale si possono specificare i seguenti dati:

- *t0*: istante di tempo in cui far iniziare la simulazione
- *tn*: durata massima di ogni generazione, si ricorda che il calcolo differenziale viene reimpostato dopo ogni duplicazione
- *n_divisioni*: il numero massimo di divisioni che si vogliono simulare
- *div_coeff*: il coefficiente di divisione della 4.7
- *div_history* e *full_history*: se si vogliono avere i file di output del programma, il primo stampa lo stato della cellula dopo ogni divisione mentre il secondo stampa lo storico completo
- *method*: il metodo di integrazione che si vuole utilizzare, maggiori dettagli in A.2
- *threshold*: soglia minima per le concentrazioni, è plausibile che durante la simulazione alcune sostanze assumano concentrazioni molto basse, questa soglia serve quindi ad impostare un limite inferiore per le concentrazioni. Le concentrazioni al di sotto di questa soglia verranno messe a 0.0 al momento delle duplicazioni.

La prima cosa che viene chiesta è il file della chimica che ha la struttura mostrata in A.1, di seguito viene chiesto se sono previsti degli eventi i quali vengono letti dalla directory `\eventi`, la simulazione inizia e va avanti fino a quando non viene effettuato il numero di divisioni specificato nel file di configurazione. Durante la simulazione vengono creati dei file intermedi che corrispondono agli output delle singole divisioni, questo è utile per monitorare l'andamento della simulazione ed

interromperla se si verifica che i risultati sono errati o poco interessanti. Una simulazione di una chimica non troppo complessa (71 specie e 86 reazioni) può durare dai 50 ai 75 minuti su un computer di fascia media.

```

1 num_specie      6
2 num_reac        5
3 memb_thick      1e-08
4 cont_rad        1e-06
5 density         2635.778
6
7 Aext  0.05  1      0
8 Bext  0.05  1      0
9 A     0.01  0      0
10 B    0.01  0      0
11 C    0.01  0      1
12 *Compl 0.0  0      0
13
14 210      Aext > A ; 8e-017
15 210      Bext > B ; 8e-017
16 21      A + C > *Compl ; 2000.0
17 12      *Compl > A + C ; 14.86
18 22      *Compl + B > C + C ; 1000.0

```

FIGURA A.1: Esempio di file di chimica

L'immagine A.1 rappresenta una chimica di esempio molto semplice, il file testuale è diviso in tre aree, la prima contiene delle informazioni generali come il numero di specie, il numero di reazioni, lo spessore della membrana, il raggio della cellula e la densità del materiale lipidico. Nella seconda area si elencano le specie chimiche della protocellula, per ogni specie si indica il nome, la *concentrazione iniziale*, un flag intero che specifica se la specie è bufferizzata o meno ed un float per il coefficiente di accoppiamento con la membrana. Per convenzione si indicano prima le specie bufferizzate, poi tutte le specie interne ed infine i complessi, tutto in ordine lessicografico. Nell'ultima area si elencano le reazioni che devono rispettare la nomenclatura spiegata in 4.4.1, si fa notare come i coefficienti di reazione vadano scritti con la notazione scientifica.

A.2 Simulazione reazioni

La struttura del programma è abbastanza semplice, esistono tre classi:

- **Reazione:** rappresenta una reazione e contiene il tipo di reazione, una lista di reagenti, una lista di prodotti ed una costante di reazione
- **Specie:** rappresenta una specie chimica, contiene il nome della specie, la sua concentrazione ed un dizionario per le varie proprietà della specie, finora si memorizza solo il *coefficiente* di accoppiamento con la membrana e se la specie è *bufferizzata*
- **Proto:** è la classe principale e contiene tutta la logica di funzionamento della protocellula

I tipi di reazione supportati sono quelli elencati nella sezione 4.2, a questi si associa un codice numerico:

- "Aggiunta" $\text{NULL} \rightarrow \text{A} : 01$
- "Scomparsa" $\text{A} \rightarrow \text{NULL} : 10$
- "Cleavage spontaneo" $\text{A} \rightarrow \text{B} + \text{C} : 12$

- "Condensazione spontanea" $A + B \rightarrow C$: 21
- "Scambio" $AA + BB \rightarrow AB + BA$: 22
- "Cleavage" $AB + C \rightarrow A + B + C$: 23
- "Condensazione" $A + B + C \rightarrow AB + C$: 32
- "Passaggio di membrana" $A_{ext} \longleftrightarrow A$: 210

La classe `Proto` è il cuore del programma e svolge tutti i passaggi elencati nel diagramma 4.1, di seguito ci si concentra sulla simulazione delle reazioni. Le reazioni vengono simulate mediante una libreria esterna di calcolo differenziale chiamata `Scipy` che, tra le tante funzionalità, permette la risoluzione numerica di un sistema di equazioni differenziali ordinarie a partire da delle condizioni iniziali.

$$\begin{cases} \frac{dy}{dt} = f(t, y) \\ y(t_0) = y_0 \end{cases} \quad (A.1)$$

Per impostare un problema di calcolo differenziale, `Scipy` mette a disposizione il seguente metodo:

- `sol1 = solve_ivp(fun, t_span, y0, method=RK45, events=[...])`

In `sol1` vengono memorizzati i risultati del calcolo, `t_span` è l'intervallo di integrazione, `y0` è l'array delle condizioni iniziali, in `method` si può indicare il metodo di integrazione mentre `events` contiene la lista degli eventi che si voglio monitorare. Per quanto riguarda i metodi di integrazione `Scipy` mette a disposizione una serie di metodi espliciti ed impliciti, in generale è consigliato usare il metodo di default `RK45` ovvero il metodo esplicito Runge-Kutta del 5° ordine. Il parametro `fun` invece è la funzione di cambiamento, nonché la f della A.1 ed è implementata all'interno della classe `Proto`. Le operazioni svolte dalla funzione sono quelle riportate nella sezione 4.4.3 ma è tuttavia necessario sottolineare alcune particolarità. La funzione `fun` è definita da R^N a R^N e restituisce un array *delta* che indica la variazione delle concentrazioni a seguito delle reazioni, come parametri in ingresso ha l'istante di tempo t ed un vettore y contenente le concentrazioni di tutte le specie, si ricorda che y contiene anche la quantità di lipide al passo t e quella al passo $t - 1$ rispettivamente in ultima e penultima posizione. Prima di effettuare le reazioni le concentrazioni devono essere ricalcolate poiché è plausibile che ad ogni passo dell'integratore il volume cambi. Calcolando il volume vecchio e nuovo con la 4.2 si ottiene il rapporto $v_{rapp} = \frac{V_{t-1}}{V_t}$ che serve a ricalcolare le concentrazioni con la seguente formula:

$$delta_i = delta_i - y_i * \frac{1 - v_{rapp}}{dt} \quad (A.2)$$

dove $delta_i$ e y_i sono rispettivamente la variazione e la concentrazione della i -esima specie. Tutto questo viene fatto prima di simulare le reazioni che vengono elaborate usando lo schema descritto nella 4.4. Prima che la funzione termini si calcola la variazione della quantità di lipide (4.6), la variazione della quantità di lipide al passo $t - 1$ viene calcolata nel modo seguente:

$$\Delta C_{t-1} = \frac{C_t - C_{t-1}}{dt} \quad (A.3)$$

Sia la A.2 che la A.3 utilizzano dt , ovvero l'intervallo di tempo scelto dall'integratore per effettuare il passo successivo, tuttavia questo non viene nativamente reso

esplicito da Scipy ed è stato infatti necessario apportare delle modifiche alla libreria. Nel file `scipy_patch.py` vengono definite due funzioni che vengono sostituite a quelle originarie della libreria Scipy, questo è effettuato mediante *monkey patching* ovvero la modifica a runtime di una classe o modulo, questa modifica tuttavia non influisce sul codice sorgente originale. Questa operazione è solitamente possibile nei linguaggi dinamici come Python poiché le classi sono *mutable*, ovvero possono essere ridefinite a runtime, essendo i metodi degli attributi di una classe anch'essi possono essere modificati. In questo modo alla funzione `fun` viene passato anche l'intervallo di tempo dt , tuttavia queste modifiche sono valide solo per i metodi di integrazione `RK45` e `RK23`.

A.3 Gestione degli eventi

Scipy dà la possibilità di inserire degli eventi durante la simulazione, nel programma si vuole controllare la condizione di duplicazione ed interrompere la simulazione se questa viene soddisfatta. Dopo l'interruzione si reperiscono i valori da `sol1`, che verranno successivamente scritti su file, e si reimposta il problema richiamando `solve_ivp` dove `y0` sarà l'array delle concentrazioni in seguito alla divisione. Prima di riavviare l'integrazione si controlla se ci sono degli eventi da applicare, questi eventi non vengono gestiti da Scipy poiché esso non dà la possibilità di modificare la concentrazione delle specie durante l'integrazione. Come detto precedentemente gli eventi sono di due tipi, il cambio dei coefficienti di reazione e la variazione delle concentrazioni delle specie. Essi possono essere definiti mediante file testuali memorizzati nella directory `./eventi` e seguono la struttura presente in figura A.2.

```
1 t_abs 200
2
3 n_reazione/specie      costante/concentrazione
4 A                      0.05
5 B                      0.01
6 5                      2.0
7 7                      5.4
8 20                     3.0
9 32                     4.0
```

FIGURA A.2: Esempio di file evento

Nella prima riga si indica in che istante di tempo si vuole far accadere l'evento, nel resto del file ogni riga corrisponde ad un singolo evento, per il cambio di concentrazione si indica il nome della specie chimica e la nuova concentrazione, per il cambio di coefficienti va indicata la posizione (partendo da 0) in cui compare la reazione nel file della chimica e il nuovo coefficiente. Entrambi i tipi di evento possono essere inseriti in uno stesso file. Dato il modo in cui vengono applicati è possibile che alcuni eventi non vengano realmente utilizzati, se per esempio si hanno due eventi di cambio concentrazione rispettivamente a 300 e 400 secondi ed una generazione che dura da 200 a 500 secondi è probabile che il secondo evento sovrascriva alcuni cambiamenti previsti dal primo, questo perché vengono applicati in cascata a partire da quello col tempo minore.

Appendice B

Generatore di chimiche

Il software descritto in questa appendice serve a generare un file di chimica da poter dare in input al programma di simulazione, infatti nonostante sia possibile creare il file descritto nell'immagine A.1 da zero, risulta più efficiente e corretto ai fini di effettuare degli esperimenti generare questi file in modo randomico ed automatico. La chimica creata da questo generatore seguirà lo stesso modello di protocellula descritto nel capitolo 4 con alcune leggere differenze.

Anche in questo caso la configurazione del generatore vengono lette da un file testuale *generator_conf.txt* (B.1) le cui voci sono abbastanza intuitive. La seconda voce è il *seed* che verrà dato in input alla libreria random di Python, la lista di specie in fondo al file rappresenta tutte le specie che fungono da cibo per la protocellula, esse infatti avranno il corrispettivo esterno bufferizzato, solitamente le specie di lunghezza 1 e 2 sono quelle considerate abbastanza piccole da attraversare la membrana.

```
chem_gen99 //nome file output
34519256 //seme random
5 //massima lunghezza
2 //cardinalità alfabeto
3 //minima lunghezza catalizzatori
5 //massima lunghezza catalizzatori
20 //numero condensazioni (esempio condensazione: 32 AA + B + BB > AAB + BB)
20 //numero cleavage (esempio cleavage: 23 AAB + AA > A + AB + AA)
0.01 //concentrazione iniziale specie chimiche
1000 //coefficiente diretto per cleavage
20000 //coefficiente diretto per condensazione
1 //genera reazioni bi-molecolari
0 //genera reazioni simmetriche
8e-17 //coefficiente per passaggio membrana
1e-08 //spessore membrana in metri
1e-06 //raggio iniziale contenitore sferico in metri
2635.778 //densità molecolare contenitore in mol/m^3

Elenco specie che attraversano la membrana (con in più la concentrazione delle specie esterne)
A 0.01
B 0.01
AA 0.01
AB 0.01
BA 0.01
BB 0.01
```

FIGURA B.1: Esempio di file di configurazione

B.1 Generazione Alfabeto

Il primo passo per impostare un file di chimica è definire le specie che ne fanno parte, in questo caso l'insieme di queste specie è chiamato *alfabeto*. L'alfabeto viene generato a partire dalla massima lunghezza delle specie L e dalla cardinalità dell'alfabeto X , per esempio con cardinalità 2 si avranno specie chimiche composte dalle lettere A e B. Con questi due dati vengono generate tutte le combinazioni di X simboli di lunghezza $[1, \dots, L]$. Di default tutte le specie, tranne quelle bufferizzate ed i complessi, contribuiscono all'alimentazione della membrana con contributo 0,01.

B.2 Generazione Reazioni

Le reazioni previste dal generatore sono solo 3: *passaggio di membrana*, *condensazione* e *cleavage*. Le prime dipendono strettamente dalle specie elencate nel file di configurazione, infatti verrà generata una reazione del tipo $Aext \longleftrightarrow A$ per ogni specie elencata. Per le condensazioni e i cleavage si parla sempre di reazioni catalizzate, infatti in base alla minima e massima lunghezza dei catalizzatori, specificata nel file, si crea un pool di catalizzatori, sottoinsieme dell'alfabeto di base descritto sopra. Per generare un cleavage o una condensazione si seguono i seguenti passaggi:

1. Si sceglie una specie a caso dall'alfabeto (es. AAB)
2. Si sceglie a caso un punto di taglio (es. $A|AB$)
3. Si prende randomicamente un catalizzatore dal rispettivo pool (es. AA)
4. Se si tratta di un cleavage AAB è un reagente e A , AB sono i prodotti
 - $AAB + AA \rightarrow A + AB + AA$
5. Se si tratta di una condensazione AAB è il prodotto e A , AB sono i reagenti
 - $A + AB + AA \rightarrow AAB + AA$

Per la condensazione si distingue tra reazioni bi-molecolari e tri-molecolari, l'esempio scritto sopra è una reazione tri-molecolare, l'equivalente bi-molecolare è:

- $A + AA \rightarrow *AAA$
- $*AAA \rightarrow A + AA$
- $*AAA + AB \rightarrow AAB + AA$

dove le specie che iniziano con un $*$ rappresentano i complessi. Le chimiche bi-molecolari sono più grandi ed in generale più lente da simulare. Il software non inibisce in alcun modo reazioni in cui una sostanza catalizza la propria produzione ($B + AA + BAA \rightarrow BAA + BAA$), in questo caso BAA è sia un catalizzatore che un prodotto. Per quanto riguarda la condensazione c'è anche la possibilità di generare reazioni simmetriche, sempre considerando l'esempio di prima, la rispettiva reazione simmetrica sarebbe:

- $AB + AA \rightarrow *ABAA$
- $*ABAA \rightarrow AB + AA$
- $A + *ABAA \rightarrow AAB + AA$

Bibliografia

- André, Brack (1998). «Aristotle gathered the different claims into a real theory». In: *Cambridge University Press*. URL: https://assets.cambridge.org/97805215/64755/excerpt/9780521564755_excerpt.pdf.
- Antonio, Lazcano (nov. 2010). «Historical Development of Origins Research». In: *Cold Spring Harbor perspectives in biology vol. 2*. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2964185/>.
- Britt, Robert Roy (lug. 2002). «New Insight into Earth's Early Bombardment». In: *Space.com*. URL: <http://www.space.com/2299-insight-earths-early-bombardment.html>.
- Cavalazzi Barbara, et al (lug. 2021). «Cellular remains in a 3.42-billion-year-old sub-seafloor hydrothermal environment». In: *Science Advances*. URL: <https://www.science.org/doi/10.1126/sciadv.abf3963>.
- Chen L. Xiao S., Pang K. (2014). «Cell differentiation and germ-soma separation in Ediacaran animal embryo-like fossils». In: *Nature* 516.
- Dodd Matthew S. Papineau Dominic, Grenne Tor (mar. 2017). «Evidence for early life in Earth's oldest hydrothermal vent precipitates». In: *Nature* 543. URL: http://eprints.whiterose.ac.uk/112179/1/ppnature21377_Dodd_for%20Symplectic.pdf.
- El Albani Abderrazak, et al (lug. 2010). «Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago». In: *Nature* 466.
- G.E, Budd (2014). «The Cambrian Fossil Record and the Origin of the Phyla». In: *Integrative and Comparative Biology* 43.
- Julie, Steenhuysen (mag. 2009). «Study turns back clock on origins of life on Earth». In: *Reuters*. URL: <https://www.reuters.com/article/us-asteroids-idUSTRE54J5PX20090520?pageNumber=1&virtualBrandChannel=0>.
- Kauffman, S. (1995). «At Home in the Universe The Search for the Laws of Self Organization and Complexity». In: *Santa Fe Institute*.
- Kunin W.E., Gaston Kevin (1996). «The Biology of Rarity: Causes and consequences of rare-common differences». In: *Springer Science*. URL: https://books.google.it/books?id=4LHnCAAAQBAJ&pg=PA110&redir_esc=y#v=onepage&q&f=false.
- Levine R, Evers C. (2008). «The Slow Death of Spontaneous Generation (1668-1859)». In: URL: https://web.archive.org/web/20080426191204/http://www.accessexcellence.org/RC/AB/BC/Spontaneous_Generation.php.
- Maloof A. C. Porter S. M., Moore J. L. (2014). «The earliest Cambrian record of animals and ocean geochemical change». In: *Nature* 516.
- McCaskill, et al (2007). «The organization of groups of amphiphilic molecules in water or other solvents». In.
- Pasteur, Louis (1859). «Pasteur's "col de cygnet"». In: *British Society for Immunology*. URL: <https://gutenberg.beic.it>.
- Ritchie, Hannah e Max Roser (2021). «Biodiversity». In: *Our World in Data*. URL: <https://ourworldindata.org/biodiversity>.
- Serra, R. e M. Villani (2017). «Modelling Protocells». In: *Springer Science*.

- Spallanzani, Lazzaro (1765). «Saggio di osservazioni microscopiche concernenti il sistema della generazione de' signori di Needham e Buffon». In: URL: <https://gutenberg.beic.it>.
- Trifonov, Edward N. (mar. 2011). «Vocabulary of Definitions of Life Suggests a Definition». In: *Journal of Biomolecular Structure and Dynamics* 29. URL: <https://doi.org/10.1080/073911011010524992>.
- Villani, M. e R. Serra (2014). «Growth and Division in a Dynamic Protocell Model». In: *Artificial Life*.
- Wikipedia (n.d.). «Estinzione di massa del Permiano Triassico». In: (). URL: https://it.wikipedia.org/wiki/Estinzione_di_massa_del_Permiano-Triassico.
- Wilde S.A. Valley J.W., Peck W.H. I& Graham C.M. (mag. 2013). «Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago». In: *Nature* 409. URL: http://www.geology.wisc.edu/zircon/Wilde_et_al.PDF.