



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Un modello statistico per l’origine della vita

Relatore

Prof. Amos Maritan

Laureando

Enrico Chiesa

Anno Accademico 2021/2022

Indice

| | | |
|----------|---|-----------|
| 1 | Teorie sull'origine della vita | 4 |
| 1.1 | Introduzione storica | 4 |
| 1.2 | Teorie recenti | 5 |
| 1.2.1 | Oparin | 5 |
| 1.2.2 | Schrödinger e Von Neumann | 8 |
| 1.2.3 | Eigen, Orgel e la teoria del mondo ad RNA | 8 |
| 1.2.4 | Dyson e la doppia origine della vita | 10 |
| 2 | Il modello di Dyson | 11 |
| 2.1 | Teoria generale dei sistemi molecolari | 11 |
| 2.2 | "A toy model of the Oparin Theory" | 12 |
| 3 | Lavori successivi | 23 |
| 3.1 | Nowak e Chen | 23 |
| 3.2 | Hilhorst | 25 |

It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us. [...] There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved

The Origin of Species
CHARLES DARWIN

Introduzione

Questo lavoro di tesi nasce da un interesse di lungo corso per le discipline biologiche; rinnovato nell'ultimo anno dal corso extracurricolare di biologia evolutiva, ambientale e dello sviluppo, che è stato capace di far intravedere come ogni organismo sia intimamente connesso – *"entangled"*, direbbe Darwin con un termine particolarmente caro ai fisici – con tutto ciò che lo circonda. L'incredibile complessità dei processi e delle relazioni molecolari riscontrabile anche nel più semplice dei sistemi viventi, unito alla *forma mentis* del fisico, mi ha portato a domandarmi quale potrebbe essere il modello più semplice possibile per un essere vivente. Mi è sembrato sensato che questo "organismo minimale" potesse essere anche il primo, antenato comune di tutte le specie viventi. Il mio interesse in questo senso ha preso forma in un testo, redatto per il corso di storia della fisica, incentrato sul lavoro di Oparin[9], di cui il primo capitolo di questo elaborato è debitore. Ho poi trovato, grazie al prof. Maritan, nel lavoro di Freeman Dyson un modello intelligente ed essenziale in grado di rappresentare un primo abbozzo di vita.

In questa tesi analizzeremo quindi il modello che il fisico e matematico statunitense ha elaborato, nei primi anni ottanta, per cercare di costruire un impianto statistico attraverso cui affrontare il tema dell'origine della vita. Tale modello, *"that is far too simple to be true"* (Dyson, 1999, p.90), non ha la pretesa di affrontare la questione nella sua totale complessità fisica e chimica. L'autore, per enfatizzare questo punto, chiama il suo *"a toy model"* dichiarando l'obbiettivo di stimolare riflessioni ed esperimenti a riguardo piuttosto che di fornire una soluzione al problema: *"If my remarks have any value, it is only insofar as they suggest new experiments"* (Dyson, 1999, p.91). Il modello è quindi molto generale, puramente statistico e, come vedremo, abbastanza semplice da poter essere affrontato senza simulazioni al computer, fornendo comunque risultati precisi.

Il paragrafo con cui si apre questo elaborato è quello che conclude *L'origine delle specie* di Charles Darwin, e non è certo scelto a caso. Proprio la darwiniana immagine di un *"entangled bank"* è infatti quella che Dyson ha in mente pensando ad una cellula primitiva. Egli immagina un insieme di diverse specie molecolari intrecciate e interagenti, come le piante e gli insetti del microcosmo di Darwin. A quest'immagine Dyson attribuisce il merito dell'idea cardine del suo modello: la capacità del sistema di tollerare errori. Questa caratteristica va però contro ciò che sappiamo sulla replicazione di acidi nucleici, che molti (vedi Eigen, Orgel, Nowak) considerano come sinonimo di vita, e che necessita, per funzionare, di una grande precisione. Dyson rigetta, quindi, l'idea dominante (allora come oggi) di un mondo ad RNA come culla della vita, abbracciando la visione opariniana di nascita della vita come nascita del metabolismo. Il protagonista di questo lavoro è una protocellula che sviluppa funzioni metaboliche, vale a dire contiene polimeri che catalizzano la formazione di altri polimeri, ed ha la capacità di riprodursi ma non di replicarsi. L'organismo primitivo di Dyson infatti non dispone di un apparato replicativo con acidi nucleici, che, secondo l'autore, si svilupperà solo in seguito, come parassita e poi simbiote all'interno della protocellula proteica già formata.

Questo aspetto verrà chiarito dalla trattazione del primo capitolo che, dopo aver fornito una panoramica storica delle idee con cui la comunità scientifica ha affrontato e sta affrontando la questione dell'origine della vita, tenta di spiegare le idee di Dyson a riguardo.

Il secondo capitolo è dedicato all'analisi approfondita del modello di Dyson per come l'autore l'ha presentato nel suo articolo[4] e nel successivo libro[3].

Nel terzo capitolo si affrontano brevemente due lavori più recenti ad opera di Nowak e Chen[1] e di Hilhorst[6] che riprendono alcuni dei concetti alla base del modello di Dyson, sviluppandoli in modi diversi.

Tutti questi modelli sono ovviamente semplicistici e non puntano a rispecchiare una situazione reale, ma piuttosto ad organizzare idee e concetti per avvicinarci sempre un po' di più a capire, non solo come è nata, ma anche cosa è, la vita.

Capitolo 1

Teorie sull'origine della vita

In questo primo capitolo verranno passate in rassegna alcune delle teorie riguardo l'origine della vita. Verrà fatto innanzitutto un excursus sulla visione della nascita della vita dall'antichità fino al XIX secolo, seguendo il percorso tracciato in questo senso da Oparin nel suo libro "The origin of life" (Oparin, 1953). Nella seconda sezione l'attenzione si sposterà su alcune teorie recenti, ed alcuni personaggi, che hanno particolarmente influenzato le idee di Dyson riguardo a questo tema. In particolare, si parlerà diffusamente di Oparin e della sua teoria, a cui Dyson tenta di dare un impianto matematico quantitativo; delle idee di Schrödinger e soprattutto di Von Neumann; e della teoria del mondo ad RNA sostenuta tra gli altri da Eigen ed Orgel. Si chiarirà infine la sintesi che Dyson opera di queste teorie, trasformandole di fatto in pezzi diversi di un'unica teoria della doppia origine della vita.

1.1 Introduzione storica

Per un periodo di tempo molto lungo, il problema dell'origine della vita è stato affrontato da un punto di vista prettamente religioso e filosofico. Il fatto che semplici forme di vita – insetti, muffe e perfino roditori – potessero emergere dalla materia inanimata era considerato un'evidenza dettata dall'esperienza di tutti i giorni. L'idea della generazione spontanea, quindi, era alla base di tutte le teorie antiche relative all'origine della vita: la più fortunata tra queste è stata senz'altro quella sviluppata da Aristotele ed espressa in particolar modo nel testo *Historia animalium*. Secondo il filosofo greco, la vita non deriva solo da altra vita ma anche da materia inanimata: in particolare le piante sono generate dalla terra, le creature marine dall'acqua e gli animali terrestri dall'aria; mentre il fuoco è la fonte primaria dei corpi celesti. Una visione dello stesso periodo che si avvicina di più a quella moderna è quella di Democrito che, come Anassimandro, credeva che la vita fosse iniziata in acqua e gli organismi avessero attraversato un lungo periodo di evoluzione fino ai giorni nostri. Inoltre, seguendo l'idea atomista, Democrito era convinto che i viventi fossero composti dagli stessi atomi di cui era composta la materia inanimata, organizzati però in modo tale da formare veri e propri organismi viventi. La teoria aristotelica, pur subendo diverse modifiche, è rimasta, per diversi secoli, quella a cui il mondo scientifico ha riferito; in questo periodo sono stati prodotti innumerevoli esperimenti che dimostravano che insetti, vermi e piccoli animali potevano generarsi spontaneamente da ogni sorta di materiale inanimato. Nel 1688, Francesco Redi ha pubblicato il trattato *Esperienze attorno alla generazione degli insetti* in cui riporta per primo prove in contrasto con la tesi della generazione spontanea. Il naturalista toscano ha osservato che i vermi bianchi che si formavano nella carne in putrefazione erano larve di mosca e la deposizione delle uova era una condizione necessaria per la loro comparsa. Questi ed altri esperimenti non hanno scardinato, però, il pensiero dominante.

Con la scoperta di batteri ed organismi microscopici, la generazione spontanea sembrava essere sempre più l'unico meccanismo in grado di spiegare un gran numero di esperimenti in cui i microorganismi comparivano in ambienti teoricamente sterili.

Nel 1862 Louis Pasteur ha posto fine in maniera definitiva a questa disputa che animava il mondo scientifico da qualche secolo. In una serie di lavori ed esperimenti egli dimostrò che aria e ogni tipo

di superficie pullulano di vita microscopica, ciò gli permise di identificare le fonti di contaminazione di tutti gli esperimenti precedenti, invalidandone i risultati. In seguito Pasteur, conservando materiali organici in ambienti ben sterilizzati senza che questi si decomponessero, ha dimostrato che la decomposizione dei materiali organici non è causa ma piuttosto conseguenza della presenza dei microorganismi. Le scoperte di Pasteur hanno provocato un deciso cambio nell'opinione generale, insinuando l'idea che la vita non potesse generarsi dalla materia inanimata, ma che fosse eterna e fondamentale quanto la materia stessa. Si sono sviluppate quindi diverse teorie legate all'eternità della vita e particolare fascino ha suscitato l'idea che la vita fosse arrivata sulla terra dallo spazio, sfruttando meteoriti o comete. Studi sulle radiazioni interstellari ad alta energia hanno però invalidato completamente tali ipotesi, costringendo sulla terra il luogo di origine della vita.

Tuttavia, le conclusioni di Pasteur non ci forniscono informazioni sulla comparsa della vita in un ambiente ed un periodo completamente diversi da quelli presenti. Nella seconda metà del diciannovesimo secolo sono stati fatti numerosi tentativi di risolvere il problema della genesi della vita, emblematica della diffusa attenzione a questo tema è una lettera di Charles Darwin a J.D.Hooker datata 1 febbraio 1871:

"It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present.— But if (& oh what a big if) we could conceive in some warm little pond with all sorts of ammonia & phosphoric salts,—light, heat, electricity & c present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter wd be instantly devoured, or absorbed, which would not have been the case before living creatures were formed.—" [2]

Darwin elenca, in queste poche frasi, quelli che poi saranno in effetti gli elementi basilari delle teorie successive, in particolare quella di Oparin: acqua, calore e, soprattutto, la mutabilità di un sistema complesso di elementi organici, da cui, ad un certo punto si è formata, a scapito dell'entropia, la vita.

1.2 Teorie recenti

Ci occuperemo ora di affrontare le teorie e idee, tra quelle diffuse nel secolo scorso sull'origine della vita, più rilevanti per il lavoro di Dyson e la sintesi che egli ne fa.

1.2.1 Oparin

Alexander Ivanovič Oparin è stato un biochimico, nato in Russia nel 1894[5], interrogatosi per tutta la sua esistenza sul problema dell'origine della vita. Nel 1922, ad un meeting della *Russian Botanical Society*, ha introdotto per la prima volta la sua idea di un organismo primordiale eterotrofo, formatosi in un brodo di composti organici da cui la protocellula avrebbe potuto trarre nutrimento. Oparin ha raccolto le sue idee in un volume monografico nel 1924 e le ha riorganizzate ed ampliate poi in quella che è la sua opera magna: "Origin of life"¹[9]. Il libro di Oparin è stato pubblicato in russo nel 1934 e trasposto in inglese nel 1936; nonostante qualche resistenza nel mondo anglosassone alle idee di uno scienziato sovietico, ha avuto subito grande risonanza internazionale, tanto che, due anni dopo la pubblicazione, il *New York Times* lo definì "a landmark for discussion for a long time to come"[7]. Previsione che appare confermata dal fatto che il modello di Dyson, elaborato nei primi anni '80, tenta di dare un impianto matematico alla teoria esposta proprio in questo testo.

Nel volume, Oparin, basandosi su osservazioni in diversi ambiti, dall'astronomia alla chimica, presenta un dettagliato resoconto del lungo processo che avrebbe portato all'emergere del primo organismo vivente. L'idea fondamentale di Oparin è che la vita sia emersa dopo un lungo periodo di "evoluzione prebiotica" delle sostanze inorganiche, in contrasto con la teoria di una generazione casuale ed improvvisa del primo organismo vivente. In effetti, il capitolo conclusivo del libro si apre con una frase piuttosto perentoria in questo senso:

...one must first of all categorically reject every attempt to renew the old arguments in favour of a sudden and spontaneous generation of life (Oparin, 1953, p.246)

¹Dyson, quasi settant'anni dopo, prenderà in prestito questo titolo per il suo di libro[3].

L'autore cerca innanzitutto di capire in che forma si trovavano, in quel periodo così distante, gli elementi che oggi riconosciamo come fondamentali per la vita. Studiando la composizione di meteoriti, spettri di stelle e di pianeti del sistema solare, Oparin arriva ad affermare che il carbonio sulla terra fosse inizialmente non nella sua forma ossidata ma, al contrario, in quella ridotta, sotto forma di idrocarburi. Similmente, anche l'azoto sarebbe stato presente sulla terra primordiale principalmente sotto forma di ammoniaca. Sebbene l'ipotesi di un'atmosfera riducente non sia oggi considerata plausibile, ambienti chimicamente simili a quello descritto da Oparin potrebbero essere state zone delle profondità oceaniche riscaldate da fonti geotermiche. Le successive derivazioni di Oparin valgono quindi per questo tipo di ambienti, considerati oggi tra i più probabili per l'origine della vita.

Successivamente l'autore si concentra sull'origine delle sostanze organiche e delle prime proteine, usando gli strumenti della chimica organica e biologica. Più che concentrarsi sui dettagli delle reazioni possibili, tenta di individuare *"a definite conception of the fundamental tendencies, of the basic trends of the behavior of organic matter in its continual evolution on the Earth's surface"* (Oparin, 1953, p.106). Oparin immagina quindi che nella calda atmosfera primordiale (o nelle calde profondità oceaniche) si siano formate grandi quantità di idrocarburi insaturi che, reagendo con l'acqua ed ossidandosi, avrebbero originato una considerevole varietà di alcoli, aldeidi, chetoni e acidi organici. Questa pletora di composti si sarebbe poi depositata in un mare caldo e primordiale, dove le reazioni sarebbero continuate per un lunghissimo periodo di tempo in mezzo acquoso, dando origine a composti organici sempre più complessi. In una lunga digressione, Oparin illustra varie reazioni che tendono a mimare quelli che, allora, erano riconosciuti come i principali processi chimici cellulari. Tramite reazioni di polimerizzazione, condensazione e ossidoriduzione, si arriva a spiegare la formazione di zuccheri, cellulosa, carboidrati complessi, proteine e molti altri composti necessari alla vita. In questo modo *"in these primitive waters materials were created out of which living organisms were to be built out subsequently"* (Oparin, 1953, p.127). Oparin dedica inoltre una sezione alla spiegazione dell'asimmetria nell'abbondanza di enantiomeri di molecole chirali, caratteristica onnipresente negli esseri viventi. Citando a proposito degli esperimenti in cui quantità diverse di enantiomeri vengono formate sotto l'azione di luce polarizzata o, in generale, di forze fisiche asimmetriche.

L'attenzione si sposta poi sulla formazione delle proteine: i primi proto-amminoacidi si sarebbero infatti formati dall'interazione di ammoniaca con ossiacidi ottenuti da aldeidi, tramite la reazione di Cannizzaro. Le proteine hanno per l'autore un'importanza fondamentale: nel 1936 era concezione diffusa che fossero proprio le proteine, e non il DNA, molecola ancora piuttosto sconosciuta, a racchiudere le informazioni genetiche degli individui. L'esistenza stessa dei geni era messa in discussione da molti nell'ambiente scientifico del tempo, e a questi dubbi si aggiungeva lo scontro, anche ideologico, tra Darwinismo e Mendelismo, teorie secondo Oparin assolutamente inconciliabili[7]. In particolare egli era sostenitore della prima corrente di pensiero e il Darwinismo era all'epoca parte integrante dell'ideologia ufficiale dell'URSS. Seguendo l'idea di Haeckel, Oparin era convinto che, almeno per i primi organismi, l'ereditarietà potesse essere spiegata in termini di crescita e divisione del protoplasma, tramite quindi un processo riproduttivo, senza necessità di replicazione. Una caratteristica fondamentale dei composti proteici è poi la capacità di formare sistemi colloidali, grado di organizzazione della materia più elevato che, secondo Oparin, *"...bridges the gap between organic compounds and living things."* (Oparin, 1953, p.136). I sistemi colloidali sono di fondamentale importanza per la teoria di Oparin, poiché in tali miscele eterogenee emergono proprietà che non sono rintracciabili nei singoli elementi che le compongono. In particolare, egli si concentra sulle proprietà dei colloidi idrofili, che hanno la caratteristica di dividersi in due strati: un sedimento fluido, ricco di molecole colloidali, in equilibrio con uno strato liquido formato quasi esclusivamente da solvente. Considerando l'idrosfera primitiva come un enorme sistema colloidale, è "inevitabile", dice Oparin, che si siano formati coacervati di molecole organiche tenuti insieme da forze elettrostatiche. Tali raggruppamenti di macromolecole sarebbero stati separati dal solvente da strati di molecole di acqua, ordinatamente disposte tutt'attorno alle molecole colloidali. Questo passaggio della teoria Opariniana è oggi messo molto in discussione, si ritiene in effetti che tali sistemi siano troppo instabili e delicati per permettere uno sviluppo soddisfacente della struttura interna degli stessi. Un'idea più moderna è che questi coacervati fossero racchiusi all'interno di membrane fosfolipidiche, simili a quelle delle cellule odierne, generatesi spontaneamente. Questi sistemi sono in grado di assorbire varie sostanze dal medium in cui

sono immersi, crescendo in dimensioni e cambiando struttura chimica. In questo fondamentale passaggio, quindi, si vengono a creare delle nicchie di materiale organico nettamente separato dal mezzo in cui era disciolto, all'interno delle quali si può ritrovare un qualche grado di ordine nella disposizione reciproca delle molecole.

È proprio questa la fase che il modello di Dyson intende analizzare, cercando di fornire stime quantitative della probabilità con cui monomeri polimerizzano in molecole catalizzatrici ed enzimi. Passando all'origine dei primi esseri viventi propriamente detti, Oparin si concentra dapprima sulla comparsa degli enzimi, elementi catalitici irrinunciabili per la vita. L'autore passa in rassegna diversi esperimenti recenti, volti a dimostrare che gli enzimi sono in effetti sistemi complessi, formati da composti chimici diversi, che devono essere tutti presenti per garantire l'azione catalitica. Non basta però la vicinanza spaziale di tutti i componenti, questi devono anche avere orientazioni e posizioni reciproche precise perché l'enzima raggiunga la capacità catalitica osservata in natura. Tali catalizzatori devono quindi essere il risultato di una lunga evoluzione e non di un fortuito evento puntuale, per dirla con le parole dell'autore: "*The inner chemical organization became strengthened in the process of natural selection, insuring a gradual evolution which finally culminated in those highly perfected enzyme systems existing at the present time.*" (Oparin, 1953, p.174). Oparin trova conferma dell'ultima affermazione in uno studio che evidenzia come la proteasi di organismi più primitivi sembri una sorta di "prototipo" per quella di organismi di più recente apparizione. Un enzima, per quanto ottimo catalizzatore, non è però che un singolo passaggio di una lunga catena di reazioni che una cellula mette in atto per ottenere un prodotto finale. La mera presenza di tutti gli enzimi e gli elementi necessari ad un processo biologico non è garanzia che questo possa avvenire, tutte le reazioni devono essere ben organizzate spazio-temporalmente per dare il risultato voluto. È quindi chiaro che l'organizzazione spaziale del protoplasma è un elemento fondamentale per l'attività enzimatica.

Oparin tenta quindi di tracciare un possibile percorso che a partire da delle fragili e instabili "gocce" di coacervati possa portare ad una proto-cellula. Egli immagina che, in condizioni particolarmente favorevoli, alcuni dei sistemi precedentemente descritti avrebbero potuto mantenersi stabili assorbendo elementi dal medium esterno ed assimilandoli attraverso lente reazioni, "*In this way a natural selection of coacervates originated in the most primitive and simplest form*" (Oparin, 1953, p.191). Viene quindi nominato il concetto di selezione naturale, basato non tanto sulla competizione tra individui ancora inesistenti, ma sulle caratteristiche proprie dei coacervati che, in qualche fortunato caso, sarebbero stati abbastanza stabili da permettere la formazione al loro interno di enzimi primitivi. Questi, a loro volta, avrebbero aumentato la dinamicità e la capacità di accrescimento del sistema: "*Each coacervate droplet increased in weight and volume the more rapidly the more perfect its physico-chemical organization was adapted to this end.*" (Oparin, 1953, p.193). Ogni ulteriore cambiamento nell'"organizzazione" citata da Oparin, costituisce un elemento positivo o negativo per la stabilità e l'evoluzione della goccia considerata, determinandone l'esistenza o la degradazione. La velocità delle reazioni all'interno dei singoli coacervati diventa il parametro più importante per la loro crescita e quindi per il loro successo. Non si può però certo pensare che un coacervato cresca indefinitamente, l'autore immagina che ad un certo punto una goccia si divida in gocce figlie, che conservano, in qualche misura, la stessa organizzazione della madre e possono continuare a crescere per poi dividersi nuovamente raggiunta una dimensione "critica". Questo processo non porta solo ad un aumento della quantità di materiale "organizzato" ma anche ad una progressiva maggiore qualità dell'organizzazione stessa. Sebbene l'autore riconosca che strutture come quelle descritte siano troppo instabili e destinate a scomparire in un lasso di tempo piuttosto breve, egli lascia aperta la possibilità che nella loro breve esistenza alcune gocce possano aver subito tanti e tali cambiamenti di organizzazione e composizione interna da costituire un sistema effettivamente stabile, in grado di accrescersi e dividersi conservando la propria struttura e preservando la propria esistenza. Un sistema di questo tipo è ciò che, secondo Oparin, cerchiamo quando pensiamo al primo organismo vivente.

L'aspetto più importante dell'opera opariniana è però probabilmente il livello di dettaglio con cui l'autore tratta vari problemi dell'origine della vita; creando, per la prima volta, la possibilità di una ricerca sperimentale in merito. Stimolare la ricerca sperimentale riguardo a questo tema è in effetti anche l'obbiettivo dichiarato che Dyson avrebbe voluto raggiungere attraverso la formulazione del suo modello. In effetti gli esperimenti a supporto della teoria di Oparin non tardarono ad arrivare dai

laboratori, tra gli altri, di Urey e Miller[8]. Le idee di Oparin hanno segnato uno stacco netto con il passato, inaugurando una stagione di esperimenti e proposte teoriche nuove ad un problema antico.

1.2.2 Schrödinger e Von Neumann

Nel 1944, Erwin Schrödinger, pubblica un libretto dal titolo *What is life?*[10] che si diffonde moltissimo e genera grande interesse attorno al processo, ancora poco compreso, della replicazione. Sebbene il fisico austriaco non tratti per nulla il tema dell'origine della vita, Dyson gli riconosce il merito di aver posto delle domande fondamentali e di non averne poste affatto delle altre. Schrödinger, infatti, dedica ben quattro capitoli alla replicazione, e solamente uno, "meno lucido", al metabolismo; seguendo l'idea che la prima sia in effetti l'elemento fondamentale per poter parlare di un "essere vivente". Negli anni successivi si diffonde l'idea che l'origine della vita sia di fatto stata un'origine dei processi replicativi. Per Dyson però non è questo il caso, egli propone l'ipotesi che i primi organismi viventi fossero in grado di riprodursi, ma non di replicarsi. Per una cellula, riprodursi vuol dire dividersi in modo che la cellula figlia erediti grossomodo la metà dei costituenti della cellula madre, e quindi in qualche modo il suo metabolismo. Per una molecola, replicarsi significa ottenere una precisa copia di sé; negli organismi moderni la riproduzione è sempre accompagnata da replicazione di molecole, non è detto però che sia sempre stato così. Quattro anni dopo il libro di Schrödinger, i rapporti logici tra replicazione e metabolismo sono chiariti da John von Neumann. Egli traccia un'analogia tra organismi viventi e automi, sostenendo che la parte proteica, e quindi metabolica, e quella di acidi nucleici, cioè replicativa, degli esseri viventi corrispondano rispettivamente ad hardware e software in un automa. L'hardware, cioè il metabolismo, è logicamente precedente rispetto al software: è quindi possibile immaginare un organismo che riesca a mantenere il proprio ordine interno e addirittura a riprodursi (cioè in qualche modo a propagare quell'ordine), senza che sia necessario alcun meccanismo di replicazione. D'altro canto, un automa composto unicamente da software non potrà che essere un parassita obbligato, questo è in effetti il caso dei virus, che possiedono del materiale genetico ma difettano quasi del tutto di metabolismo. Questo passaggio apre alla possibilità, caldeggiata da Dyson, di una doppia origine della vita: la prima che riguarda organismi quasi esclusivamente proteici, in grado di crescere e riprodursi ma non di replicarsi, e la seconda che vede la nascita di catene di nucleotidi che si comportano dall'inizio in modo parassitico, sfruttando le strutture proteiche già presenti per replicarsi. Le prime creature proteiche potrebbero essere esistite per un lungo periodo di tempo evolvendo da sole un apparato metabolico anche molto sofisticato, e, dice Dyson, è più facile immaginare che due eventi molto improbabili come la formazione delle prime proteine e quella dei primi acidi nucleici avvengano in momenti diversi piuttosto che contemporaneamente, come vorrebbe una teoria della singola origine della vita.

1.2.3 Eigen, Orgel e la teoria del mondo ad RNA

Abbiamo detto prima che, di norma, la funzione di hardware è attribuita a proteine e quella di software ad acidi nucleici, esistono però eccezioni a questa regola, e una in particolare ha un ruolo centrale in alcune teorie sull'origine della vita: L'RNA. Questa molecola è presente nelle cellule in forme diverse: l'RNA genomico, che costituisce il codice genetico di molti virus ed ha indubbia funzione di software; l'RNA ribosomiale e di trasporto, con sicura funzione di hardware; e l'RNA messaggero, a cui era attribuita funzione di software fino al 1982, quando Thomas Cech scoprì che alcune molecole di questo tipo di RNA, ribattezzate ribozimi, si comportano come enzimi, catalizzando reazioni di altre molecole di mRNA. Per questa sua straordinaria varietà di funzioni, negli anni '80 prese piede l'ipotesi, tra le più condivise ancora oggi, di un primitivo "Mondo ad RNA", popolato da organismi in cui sia le funzioni metaboliche che quelle replicative fossero svolte proprio da questa molecola.

In questo contesto si collocano gli esperimenti, in un certo senso complementari, di Eigen e Orgel. Entrambi i chimici sperimentali si sono concentrati sul problema di capire se, in condizioni ambientali favorevoli, una molecola di RNA possa formarsi spontaneamente da una collezione di nucleotidi in soluzione. I loro esperimenti hanno effettivamente portato alla polimerizzazione di acidi nucleici, in grado poi di evolvere e mutare, come si evince dal grafico in figura 1.1. Tuttavia, le condizioni necessarie al successo degli esperimenti non sono quelle che si potrebbero trovare in un ambiente prebiotico.

Nel caso di Engel, nella soluzione di nucleotidi di partenza, è presente anche l'enzima polimerasi, un catalizzatore proteico estratto da un batteriofago. È solamente grazie all'azione di questo enzima che i nucleotidi riescono a polimerizzare, formando molecole di RNA.

Orgel, invece, non si avvale di enzimi biologici per indurre la polimerizzazione dei nucleotidi, ma aggiunge alla soluzione ioni di zinco e delle molecole di RNA modello, già formate, che i nucleotidi possono "copiare". Agli occhi di Dyson, l'esperimento di Orgel risulta in qualche modo tautologico: si forma RNA in presenza di altro RNA, ma allora come si è formata la prima molecola di RNA? I risultati di Engel invece, sempre secondo Dyson, sono facilmente collocabili in seno alla teoria della doppia origine della vita, la polimerizzazione di nucleotidi potrebbe essere avvenuta all'interno di organismi unicamente metabolici, in un ambiente protetto e già popolato da una varietà di enzimi proteici in grado, magari, di svolgere un lavoro simile a quello della polimerasi.

La teoria del mondo ad RNA di Eigen, dando per scontato che in qualche modo si formino le prime molecole di RNA per polimerizzazione di nucleotidi, si basa su due concetti principali: quasi-specie e ipercicli. Una quasi-specie è una popolazione di molecole di RNA geneticamente imparentate, ma non identiche, che fungono da modello per la polimerizzazione di nuovi elementi della quasi-specie. Nel corso delle generazioni avvengono mutazioni casuali, che però sono, secondo Engel, limitate dalla selezione Darwiniana che tende a mantenere le quasi-specie omogenee.

L'iperciclo è una forma di organizzazione più complessa, che emerge nel momento in cui più quasi-specie di RNA cominciano ad interagire con popolazioni di enzimi proteici. Nell'idea di Engel, gli enzimi associati ad una quasi-specie dovrebbero catalizzare la formazione di un'altra quasi-specie e viceversa, mantenendo il sistema in un equilibrio stabile.

Dyson è piuttosto scettico riguardo l'applicabilità di questa teoria, e riscontra la sua maggiore criticità nel fatto che la replicazione, per funzionare, deve essere da subito quasi perfetta. Se così non fosse, nel giro di poche generazioni, si accumulerebbe un numero di errori tale da causare un progressivo deterioramento del sistema, fino alla scomparsa dell'ordine che lo caratterizza. Questo processo viene indicato con il nome di "catastrofe degli errori" e lo stesso Eigen fornisce una formula per valutarne l'impatto. Si supponga di avere un sistema con N bits di informazione, in cui ogni volta che un bit viene copiato la probabilità di errore è ϵ ; e per cui la selezione naturale penalizzi gli errori tramite un fattore di selezione S . Cioè un sistema senza errori ha un vantaggio selettivo S rispetto ad uno con un solo errore, e così via. Il criterio di sopravvivenza in queste condizioni è:

$$N\epsilon < \log S \quad (1.1)$$

Se la condizione (1.1) è soddisfatta, il vantaggio selettivo è sufficiente a mantenere il numero di errori sotto controllo, altrimenti in breve tempo il sistema andrà deteriorandosi. In termini di teoria dell'informazione, il membro sinistro della disuguaglianza rappresenta il numero di bits di informazione persi a causa degli errori in ogni generazione. Il termine di destra invece è il numero di bits di informazione forniti dall'attività selettiva dell'ambiente. Se l'informazione fornita ad ogni generazione è minore di quella persa, il deterioramento del sistema è inevitabile.

Poiché il vantaggio selettivo di un sistema senza errori non può essere enorme, un valore biologicamente sensato per $\log S$ è l'unità, ciò implica che, per soddisfare (1.1), bisognerebbe avere al più $\epsilon \simeq N^{-1}$. Questa condizione è a mala pena rispettata negli organismi viventi moderni, che hanno N dell'ordine di 10^8 ed ϵ dell'ordine di 10^{-8} , e possono raggiungere frequenze di errore così basse solo grazie ad un

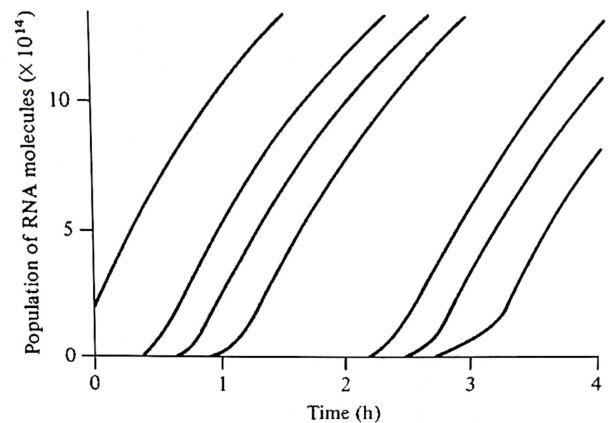


Figura 1.1: Esperimento Biebricher-Eigen-Luce, 1981. Evoluzione delle molecole di RNA a partire da soluzione di nucleotidi ed enzima polimerasi. Le quattro curve sulla destra sono relative ad esperimenti in cui sono state aggiunte, rispettivamente, da dx a sx, 10^{14} , 10^6 , 10^3 e 1 molecola di RNA modello. Le tre curve sulla sinistra rappresentano esperimenti senza RNA modello.

(Dyson, 1999, p.12)

complessissimo apparato di replicazione e controllo. La condizione (1.1), quindi, impone un vincolo piuttosto stringente a qualsiasi teoria dell'origine della vita che si basi sulla replicazione. Tutti gli esperimenti di replicazione dell'RNA in condizioni prebiotiche hanno dato valori di ϵ dell'ordine di 10^{-2} , imponendo la massima dimensione del sistema alle centinaia di bits di informazione, decisamente troppo pochi per descrivere una chimica catalitica complessa.

In conclusione, la teoria di Eigen richiederebbe un meccanismo di replicazione allo stesso tempo estremamente semplice e preciso, a livelli che è difficile immaginare per un sistema prebiotico. Principalmente per questo motivo, Dyson decide di concentrarsi su un'altra teoria: quella della doppia origine della vita.

1.2.4 Dyson e la doppia origine della vita

Come già accennato in precedenza, Dyson si fa promotore della teoria della doppia origine della vita. L'idea fondamentale è che le funzioni metaboliche e quelle genetiche, che appaiono oggi perfettamente legate ed intersecate, abbiano in realtà avuto due origini diverse e temporalmente molto distanti. In un certo senso, Dyson, dà ragione sia ad Oparin che ad Eigen, immaginando una prima origine di organismi proteici opariniani e, in un secondo momento, all'interno di questi, una seconda origine di sistemi di acidi nucleici. Questi ultimi, essendo solo software e peraltro piuttosto delicati, sarebbero stati costretti a vivere da parassiti all'interno dei primi, sfruttando strutture proteiche, formatesi nel corso di molti anni di evoluzione, per la propria replicazione. Ad un certo punto, poi, le molecole di acidi nucleici sarebbero diventate parte integrante della struttura cellulare, ed avrebbero acquisito il ruolo che hanno oggi.

Questo passaggio ricorda da vicino la teoria endosimbiontica proposta, tra gli altri, da Margulis, e oggi largamente accettata, per spiegare la nascita delle cellule eucariotiche. Il punto essenziale dell'idea di Margulis è che il parassitismo e la simbiosi siano le vere forze che guidano l'evoluzione della complessità cellulare. L'ipotesi della doppia origine della vita implica che gli acidi nucleici siano stati i primi, e i più efficaci, parassiti cellulari (intendo qui con cellula anche un organismo come quelli descritti da Oparin). Dyson immagina una situazione in cui organismi puramente metabolici sono esistiti per milioni di anni, sviluppando processi e cicli catalitici sempre più complessi, ed arrivando a sintetizzare molecole di ATP (Adenosin trifosfato). Le cellule con ATP sono più efficienti e quindi prevalgono nei processi di selezione naturale, nel corso del tempo ci si ritrova con una popolazione di cellule piene di molecole di ATP e di sottoprodotti del loro utilizzo, come l'AMP (adenosin monofosfato) che è uno dei nucleotidi che costituisce l'RNA. In questo ambiente, protetto e ricco di nucleotidi e strutture proteiche, accade quindi quel che succede nell'esperimento di Eigen: si formano le prime molecole di RNA. In questo modo l'RNA fa la sua comparsa come una malattia parassitica all'interno della cellula. Come per la teoria endosimbiontica, alcune cellule sarebbero state in grado di sopravvivere all'infezione, tramutando il parassita in simbiote. Sebbene il fisico americano ci tenga a specificare che "*This view of RNA as the oldest and most incurable of our parasitic diseases is only a poetic fancy, not yet a serious scientific theory*" (Dyson, 1999, p.16), questa idea è interessante per almeno tre ragioni. Innanzitutto, viene in qualche modo rispettata la precedenza logica dell'hardware rispetto al software, tanto cara a Von Neumann. Secondariamente, in un ambiente prebiotico, gli amminoacidi sono monomeri di relativa facile sintesi, ne è prova il fatto che sono stati trovati anche all'interno di rocce meteoritiche. Mentre gli acidi nucleici sono molecole molto più difficili da sintetizzare e i nucleotidi sono meno stabili degli amminoacidi. È quindi più facile, per i nucleotidi, formarsi e reagire all'interno di un ambiente stabile e in presenza di enzimi. La terza ragione per cui questa teoria è interessante è che potrebbe essere sperimentalmente testabile. Dyson contempla infatti la possibilità di trovare dei microfossili di cellule, abbastanza ben conservati ed antichi, che potrebbero avere tracce di sistemi metabolici ma non di acidi nucleici. Tuttavia reperti di questo tipo non sono ancora stati trovati, e non è chiaro se tali informazioni potrebbero in ogni caso essere ottenute da microfossili. Una volta formati i primi organismi in cui parte metabolica e parte genetica lavorano in modo sinergico, questi avrebbero avuto un notevole vantaggio nei confronti della popolazione unicamente metabolica, venendo quindi premiati dalla selezione naturale.

Passate in rassegna queste teorie, abbiamo un contesto generale in cui inserire il modello di Dyson, che ora andiamo ad analizzare.

Capitolo 2

Il modello di Dyson

In questo capitolo andremo ad analizzare nel dettaglio il "toy model" proposto da Dyson nel suo articolo (Dyson, 1982) e nel successivo libro (Dyson, 1999). Il modello di Dyson è un tentativo di inquadrare la teoria di Oparin in una forma matematica sufficientemente precisa da poterne trarre conclusioni definite. L'autore propone di riuscirci in due step: il primo è individuare delle equazioni precise, che consentano di trattare la popolazione di molecole come un sistema dinamico classico, disinteressandosi completamente delle leggi di interazione tra i monomeri. Questa teoria generale dei sistemi molecolari permette di definire cosa si intende per "origine del metabolismo", senza però fornire alcuna informazione sulle condizioni a ciò necessarie. Il secondo step consiste invece nel ridurre tale teoria generale ad un modello vero e proprio, assumendo regole definite per il calcolo delle probabilità di interazione molecolare. In questo modo il modello si comporta di fatto saltando, con probabilità definite, tra due stati che differiscono per il grado di organizzazione metabolica.

2.1 Teoria generale dei sistemi molecolari

Vediamo ora di affrontare il primo step: si consideri una popolazione di molecole confinate in una goccia, che evolve e cambia attraverso reazioni chimiche di vario genere, rappresentate da una singola matrice di probabilità M . Il sistema può essere rappresentato su uno spazio multidimensionale astratto, in cui ogni punto corrisponde ad una particolare configurazione molecolare della goccia. La popolazione evolve per passi discreti e stocastici, che rappresentano singole reazioni chimiche, in modo che la distribuzione di probabilità della popolazione dopo k "passi" sia data da:

$$P(k) = MP(k-1) \quad \text{quindi} \quad P(k) = M^k P(0) \quad (2.1)$$

dove $P(0)$ è la distribuzione di probabilità in uno stato iniziale arbitrario. Siamo in particolare interessati a distribuzioni che persistono nel tempo, che chiamiamo quasi-stazionarie ed indichiamo con la lettera Q , che soddisfano l'equazione

$$MQ - Q = F \quad (2.2)$$

Tutte le distribuzioni quasi-stazionarie hanno un loro bacino di attrazione nello spazio delle configurazioni, F rappresenta il flusso in uscita dal bacino di Q , definisce la probabilità che una popolazione esca dal bacino cui appartiene e "scollini" in uno adiacente. F è in genere piuttosto piccolo, dell'ordine di m^{-1} per una distribuzione che si conserva per m passi del sistema.

La configurazione essenziale per il modello di Dyson consta di due stati Q e Q' con bacini di attrazione separati da un'alta barriera. Q è relativo ad una distribuzione disordinata, che potremmo definire "morta", in cui la popolazione viene a trovarsi facilmente partendo da una condizione iniziale casuale. Q' , invece, si presenta più strutturato, magari con cicli biochimici attivi e comunque con un più alto

tasso metabolico. Il raro salto statistico, per cui una popolazione nello stato Q va incontro ad una serie di modificazioni, che la spingono contro il gradiente di probabilità, fino ad entrare nel bacino di Q' , è l'evento che chiamiamo "origine del metabolismo". In una teoria più completa dell'origine della vita è probabile che ci siano molti di questi passaggi, ognuno dei quali porta la popolazione di molecole in uno stato quasi-stazionario più ordinato. Tuttavia, nel suo modello, Dyson si concentra unicamente sul primo di questi salti. L'obiettivo è quindi quello di trovare una probabilità da associare a tale salto che sia trascurabile per una singola goccia ma abbastanza grande da permettere che tali passaggi avvengano occasionalmente in una popolazione abbastanza grande e stabile di gocce.

2.2 "A toy model of the Oparin Theory"

Lo scopo dichiarato di Dyson è di fornire "...an idealized picture of molecular evolution that resembles in a qualitative fashion the Oparin picture of the origin of life." (Dyson, 1999, p.53). Il modello sfrutta equazioni proprie di biologia delle popolazioni e si basa su 10 assunti¹, ordinati dai più generali a quelli più specifici, che lo definiscono univocamente. In questo modo, dice l'autore, può essere più facile generalizzare il modello modificando solo le ultime affermazioni, senza sconvolgerlo totalmente. Gli assunti sono:

1. Teoria di Oparin: le cellule vengono prima, gli enzimi poi, i geni molto dopo.
2. Chiamiamo cellula un volume di fluido confinato, contenente componenti molecolari, che chiameremo genericamente monomeri, i quali vengono scambiati lentamente con l'ambiente circostante, che serve da fonte di energia per le reazioni all'interno della cellula.
3. Ogni cellula contiene N monomeri di $(n+1)$ tipi diversi, ugualmente abbondanti, alcuni di questi saranno liberi e altri combinati tra loro a formare polimeri.
4. All'interno delle cellule, l'evoluzione della popolazione di monomeri procede in modo esclusivamente randomico. Non sono presenti meccanismi di selezione naturale poiché le cellule non interagiscono tra di loro.
5. La popolazione di monomeri evolve per passi discreti, che chiameremo genericamente mutazioni e che occorrono ad una distanza di un tempo Δt l'una dall'altra. Viene interessato un solo monomero per volta, che può venire aggiunto, tolto o sostituito in un polimero.
6. Ogni monomero all'interno di una cellula muta con uguale probabilità $\frac{1}{N}$ nell'unità di tempo Δt .
7. I monomeri possono trovarsi in due soli stati: attivo, quando il monomero è correttamente inserito in un polimero che catalizzi la sintesi di altre strutture catalitiche; o inattivo, altrimenti. In questo modo si riduce un random walk multidimensionale ad uno unidimensionale.
8. In una data posizione di un polimero solo una specie di monomeri potrà essere attiva, tutte le altre saranno inattive. Indichiamo: con np la probabilità per unità di tempo Δt (rate) che si aggiunga ad un polimero in formazione un monomero sbagliato, che risulterà quindi inattivo; con qp , dove q è una costante², il rate con cui un monomero, attivo o meno, si slega da un polimero; con $\Psi(x)p$, dove x è la frazione di monomeri attivi sul totale, la probabilità che in un'unità di tempo ad un polimero si leghi un monomero corretto, che quindi diventa attivo.

La funzione $\Psi(x)$ rappresenta l'efficacia della popolazione di monomeri esistente nel catalizzare la formazione di nuovi monomeri attivi. Considerando ogni mutazione come un passaggio da una popolazione madre ad una popolazione figlia, il fatto che $\Psi(x)$ dipende da x significa che esiste un fenomeno

¹Nell'articolo (Dyson, 1982) in realtà i punti sono nove e nel libro (Dyson, 1999) sono 10; le informazioni necessarie a definire il modello sono circa le stesse ma distribuite in modo diverso. In questo lavoro si sono mantenuti 10 punti e si è cercato di fare una sintesi tra i due approcci, piuttosto diversi, usati dall'autore nei due testi

²Dyson ipotizza che il valore di q possa dipendere da fattori ambientali, come la temperatura, e dalle specifiche forze di attrazione e legame tra i vari monomeri, ma non scende in dettagli a riguardo.

di ereditarietà, nel senso che la nuova generazione dipenderà dalla composizione della generazione precedente³. Dyson sottolinea come la condizione 8 sia una drastica approssimazione, che egli chiama "approssimazione di campo medio", poiché analoga a quella presente nel modello di ferromagnete di Curie-Weiss. Essa infatti sostituisce la media delle efficienze di una popolazione di catalizzatori con l'efficienza di un catalizzatore medio. L'effetto di ciò è la riduzione del sistema ad un random walk ad un singolo parametro x . L'autore fa notare poi che le previsioni ottenute con un'approssimazione del genere sono di solito, sia in fisica che in biologia delle popolazioni, pessimiste; nel senso che predicono un livello di ordine minore di quello effettivamente osservato.

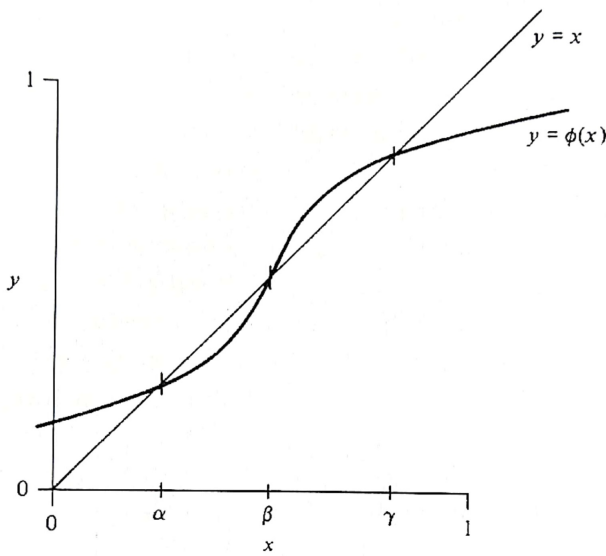
Come conseguenza del punto 8, si ha che le frazioni di monomeri attivi, inattivi e liberi sono proporzionali, secondo il fattore p , a , rispettivamente, $\Psi(x)$, n e q . La somma di tali frazioni dà l'unità; per cui, data $a = n + q^4$, si ha

$$(\Psi(x) + a)p = 1 \quad \text{quindi} \quad \phi(x) = \Psi(x)p = \left(1 + \frac{a}{\Psi(x)}\right)^{-1} \quad (2.3)$$

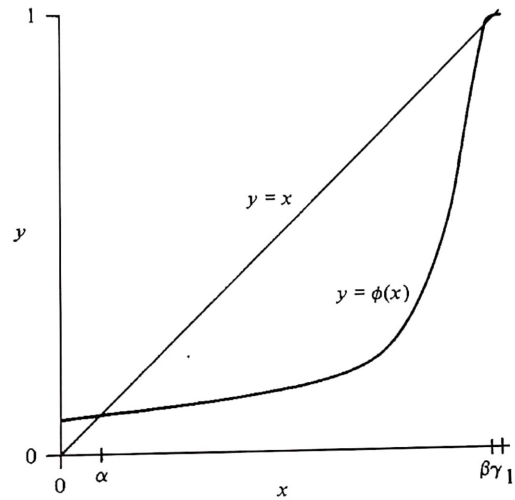
Con l'introduzione del parametro a si può considerare una nuova quantità ap che rappresenta il rate con cui un monomero risulta inattivo.

Per uno stato stazionario, in cui la popolazione figlia eredita la stessa attività catalitica media della popolazione madre, si avrà $\phi(x) = x$.

9. La curva $y = \phi(x)$ intercetta la retta $y = x$ in tre punti α , β e γ , tali che $0 < \alpha < \beta < \gamma < 1$. Ci sono quindi tre valori di x per cui il sistema si trova in equilibrio.



(a) Andamento "ad S" di $y = \phi(x)$.
(Dyson, 1999, p.57)



(b) Andamento di $y = \phi(x)$ per una cellula moderna.
(Dyson, 1999, p.70)

La condizione 9, unita al fatto che $\phi(0) > 0$, poiché anche in assenza di attività catalitica è sempre possibile che una mutazione porti alla formazione di un monomero attivo, impone che la curva $y = \phi(x)$ abbia la forma rappresentata nel grafico in figura 2.1a. Si ha quindi che gli equilibri posti in α e γ sono stabili mentre quello in β è instabile⁵, infatti $\phi'(\alpha) < 1$ e $\phi'(\gamma) < 1$ e $\phi'(\beta) > 1$. In particolare, α rappresenta lo stato di equilibrio disordinato o "morto" e γ quello più ordinato e "vivo".

Si noti che nella figura 2.1a α , β e γ sono circa equidistanti, la curva appropriata per descrivere una cellula moderna sarebbe però fortemente asimmetrica, con β e γ molto vicini tra loro e spostati sulla

³In biologia delle popolazioni, la funzione $\Psi(x)$ prende il nome di "viability" e rappresenta la probabilità di sopravvivenza (e quindi di riproduzione) in funzione di un parametro fenotipico x .

⁴Si potrebbe in effetti considerare la situazione di "monomero non legato" come se ci fosse una specie in più di monomero, con peso statistico q , a differenza di 1 per tutti gli altri, che risulta sempre inattivo.

⁵In effetti, α e γ rappresentano gli stati stazionari Q e Q' , citati nella teoria generale dei sistemi molecolari mentre β rappresenta il punto più alto della barriera che li divide.

destra verso uno e α molto vicino a zero, come in figura 2.1b. In questo modo il modello riflette il fatto che è possibile, ma non troppo facile, per una cellula morire; mentre è praticamente impossibile che una cellula morta torni in vita per fluttuazione statistica. L'estrema asimmetria è conseguenza dell'enorme complessità e dell'estrema precisione del metabolismo di una cellula moderna. Si potrebbe ipotizzare, ed è proprio ciò che fa Dyson, che l'attuale $\phi(x)$ asimmetrica si sia evoluta da una $\phi(x)$ primitiva e più simmetrica, simile a quella in figura 2.1a.

Fino a questo punto, il modello è generale ed astratto e non ha nessuna connessione con i meccanismi chimici che portano a cambiamenti nella popolazione. Tutti i dettagli a riguardo sono nascosti nella precisa forma della funzione $\phi(x)$, che quindi dovrebbe essere scelta basandosi su una teoria completa della chimica prebiotica. Dal momento che però tale teoria non esiste, la scelta va fatta in modo piuttosto arbitrario, questo sarà l'oggetto del decimo assunto, ora ci concentriamo però sul trarre delle conseguenze dal modello per come lo abbiamo definito fino a questo punto.

Nel suo articolo Dyson dichiara che le equazioni che descrivono l'evoluzione della popolazione nel modello sono le seguenti:

$$P_j(k+1) - P_j(k) = \psi_j(k) - \psi_{j-1}(k)^6 \quad (2.4)$$

$$\psi_j(k) = f\left(\frac{j+1}{N}\right) P_{j+1}(k) - g\left(\frac{j}{N}\right) P_j(k) \quad \text{dove} \quad (2.5)$$

$$f(x) = x(1 - \phi(x)); \quad g(x) = (1 - x)\phi(x) \quad (2.6)$$

Dove $P_j(k)$ è la probabilità che ci siano j monomeri attivi nella popolazione dopo k mutazioni. La funzione $f(x)$, con x che indica la frazione di monomeri attivi sul totale N , rappresenta la probabilità che un monomero attivo diventi inattivo, mentre la funzione $g(x)$ il contrario. La funzione $\psi_j(k)$, quindi, indica la differenza tra la probabilità che il sistema passi da $j+1$ a j monomeri attivi e la probabilità che lo stesso passi da j a $j+1$:

$$\psi_j(k) = Prob(j+1 \rightarrow j) - Prob(j \rightarrow j+1)$$

La differenza $\psi_j(k) - \psi_{j-1}(k)$ dell'equazione 2.4 quindi risulta positiva se il sistema è attratto verso il valore j e negativa se invece tende ad allontanarsi da tale valore:

$$\psi_j(k) - \psi_{j-1}(k) = (Prob(j+1 \rightarrow j) + Prob(j-1 \rightarrow j)) - (Prob(j \rightarrow j+1) + Prob(j \rightarrow j-1)) \quad (2.7)$$

L'equazione 2.4 si applica per $j = 0, 1, 2, \dots, N$ con le condizioni al contorno $\psi_N(k) = \psi_{-1}(k) = 0$ che riflettono il fatto che non è possibile avere $N+1$ o -1 monomeri attivi.

L'unica soluzione stazionaria dell'equazione 2.4 è ottenibile imponendo $P_j(k) = P_j$, indipendente da k . La condizioni appena mostrate, insieme all'equazione 2.4, impongono, in situazione stazionaria, $\psi_j(k) = 0$. Vediamo ora di ricavare una forma esplicita per P_j partendo dall'equazione 2.5 unita alle condizioni appena enunciate:

$$0 = f\left(\frac{j}{N}\right) P_j(k) - g\left(\frac{j-1}{N}\right) P_{j-1}(k) \Rightarrow \quad (2.8)$$

$$P_j = \frac{g((j-1)/N)}{f(j/N)} P_{j-1} = a_j P_{j-1} \quad \text{dove} \quad a_j = \frac{g((j-1)/N)}{f(j/N)} \quad (2.9)$$

Dato un P_0 iniziale quindi, si trova ricorsivamente

⁶Si noti che la funzione ψ , che compare in queste equazioni, rappresenta una quantità diversa dalla funzione Ψ dell'assunto 8.

$$P_j = \prod_{l=1}^j a_l P_0 = P_0 \exp \left(\sum_{l=1}^j \ln a_l \right) = P_0 \exp(-W(l)) \quad \text{dove} \quad (2.10)$$

$$W(l) = \left(\sum_{l=1}^j \ln (a_l)^{-1} \right) = \sum_{l=1}^j \left[\ln \left(\frac{l}{(N+1-l)} \right) - \ln \left(\frac{\phi((l-1)/N)}{(1-\phi(l/N))} \right) \right] \quad (2.11)$$

Si ha in effetti

$$a_l^{-1} = \frac{f(l/N)}{g((l-1)/N)} = \frac{\frac{l}{N}(1-\phi(l/N))}{(1-\frac{l-1}{N})\phi((l-1)/N)} = \left(\frac{l}{N-l+1} \right) \frac{(1-\phi(l/N))}{\phi((l-1)/N)} \quad (2.12)$$

Ci si aspetta di dover lavorare con valori di N piuttosto grandi, è quindi lecito usare un'approssimazione continua per 2.11 che diventa

$$W(j) = \int_1^j \left[\ln \left(\frac{l}{(N+1-l)} \right) - \ln \left(\frac{\phi((l-1)/N)}{(1-\phi(l/N))} \right) \right] dl \quad \text{sostituendo } \frac{l}{N} = y \quad (2.13)$$

$$W(j) = N \int_{1/N}^{j/N} \left[\ln \left(\frac{y}{(1+1/N-y)} \right) - \ln \left(\frac{\phi(y-1/N)}{(1-\phi(y))} \right) \right] dy \quad \text{ponendo } 1/N \simeq 0 \text{ si ottiene} \quad (2.14)$$

$$W(j) = N \int_0^{j/N} \left[\ln \left(\frac{y}{(1-y)} \right) - \ln \left(\frac{\phi(y)}{(1-\phi(y))} \right) \right] dy = NU(j/N) \quad (2.15)$$

Dove

$$U(x) = \int_0^x \left[\ln \left(\frac{y}{(1-y)} \right) - \ln \left(\frac{\phi(y)}{(1-\phi(y))} \right) \right] dy = \int_0^x \ln \left(\frac{f(y)}{g(y)} \right) dy \quad (2.16)$$

Rappresenta il potenziale del sistema. La funzione $U(x)$ ha infatti due minimi (equilibri stabili) in $x = \alpha, \gamma$ ed un massimo (equilibrio instabile) in $x = \beta$.

Si avrà quindi alla fine ⁸

$$W(j) = NU(j/N) + 1/2 \ln [2\pi N u(j/N)] \quad (2.17)$$

Dove $U(x)$ ha l'espressione in 2.16 e

$$u(x) = \frac{(x-x^2)\phi(x)(1-\phi(x))}{\phi(0)(1-\phi(0))} \quad (2.18)$$

La distribuzione di probabilità stazionaria P_j sarà concentrata attorno ai due minimi α e γ con forma approssimativamente gaussiana: infatti, approssimando $P_j = P_0 \exp -NU(x)$ ed espandendo $U(x)$ in serie di Taylor attorno ad x_0 , con la condizione $U'(x_0) = 0$, si ottiene

$$P_j = P_0 \exp(-NU(x_0)) \exp \left(-(NU''(x_0)) \frac{(x-x_0)^2}{2} \right)$$

⁷Nell'articolo di Dyson (Dyson, 1982), tra le due parentesi che compongono l'argomento del secondo logaritmo della formula 2.11, è indicato un prodotto in luogo di un quoziente.

⁸Si riporta nella formula 2.17 l'espressione per $W(j)$ che compare nell'articolo di Dyson; si noterà tuttavia che dai conti precedenti si è derivato solo il primo addendo a destra dell'uguaglianza. Nonostante numerosissimi tentativi, specialmente rimaneggiando la formula di Eulero-McLaurin, non siamo infatti riusciti ad ottenere la parte che comprende $u(j/N)$; tali inconcludenti calcoli non sono stati quindi riportati. Dyson stesso, nel suo libro (Dyson, 1999, p.62, Eq.3.12), riporta, "... after making some unimportant approximation", l'espressione $P_j = K \exp(-NU(j/N))$, senza quindi considerare per nulla il termine in cui compare $u(j/N)$, arrivando poi comunque alle medesime conclusioni.

La distribuzione centrata in x_0 ha quindi varianza $V(x) = (NU''(x_0))^{-1}$.

Calcoliamo ora le derivate di $U(x)$:

$$U'(x) = \ln\left(\frac{x}{(1-x)}\right) - \ln\left(\frac{\phi(x)}{(1-\phi(x))}\right) \quad (2.19)$$

$$U''(x) = \frac{1}{(x-x^2)} - \frac{\phi'(x)}{\phi(x) - \phi(x)^2} \quad (2.20)$$

È facile verificare che, per $x = \phi(x)$, quindi $x = \alpha, \beta, \gamma$, si ha $U'(x) = 0$ e $U''(x) = \frac{1-\phi'(x)}{(x-x^2)}$; si ottiene quindi

$$V(N\alpha) = N^2V(\alpha) = (N/U''(\alpha)) = N \frac{(\alpha - \alpha^2)}{1 - \phi'(\alpha)} \quad (2.21)$$

$$V(N\gamma) = N^2V(\gamma) = (N/U''(\gamma)) = N \frac{(\gamma - \gamma^2)}{1 - \phi'(\gamma)} \quad (2.22)$$

La quantità fondamentale per il fenomeno dell'"origine della vita" è il tasso di transizione dallo stato disordinato α a quello ordinato γ . Per calcolarlo consideriamo una situazione di flusso stazionario "artificiale" in cui le popolazioni che arrivano nello stato con $j = N\gamma$ vengono immediatamente rimosse e sostituite da popolazioni disordinate a $j = 0$. L'equazione 2.4 diviene quindi

$$\begin{cases} \psi_j - \psi_{j-1} = \epsilon & \text{per } j = N\gamma \\ \psi_j - \psi_{j-1} = -\epsilon & \text{per } j = 0 \\ \psi_j - \psi_{j-1} = 0 & \text{altrimenti} \end{cases} \quad (2.23)$$

Dove ϵ è il flusso di transizione, cioè la probabilità per mutazione che una popolazione attraversi la barriera dal disordine all'ordine. Con riferimento alla relazione 2.7 si nota come $x = \gamma$ (cioè $j = N\gamma$) è un polo attrattivo, con flusso $\epsilon > 0$, mentre nello stato con $j = x = 0$ il flusso sia negativo, e il sistema tende quindi ad allontanarsene.

Si ha inoltre

$$P_j = 0 \quad \forall j \geq N\gamma \quad (2.24)$$

le condizioni 2.23 e 2.24, trasposte in 2.5 danno, per $j = N\gamma$

$$\psi_{N\gamma} - \psi_{N\gamma-1} = \epsilon = g\left(\frac{N\gamma-1}{N}\right) P_{N\gamma-1} \quad \Rightarrow \quad P_{N\gamma-1} = g\left(\frac{N\gamma-1}{N}\right)^{-1} \epsilon \quad (2.25)$$

Poiché tutti gli altri termini contengono P_j con $j \geq N\gamma$; procedendo a ritroso, per $j = N\gamma - 1$, si ha

$$0 = -g\left(\frac{N\gamma-1}{N}\right) P_{N\gamma-1} - f\left(\frac{N\gamma-1}{N}\right) P_{N\gamma-1} + g\left(\frac{N\gamma-2}{N}\right) P_{N\gamma-2} \quad \Rightarrow \quad (2.26)$$

$$P_{N\gamma-2} = g\left(\frac{N\gamma-2}{N}\right)^{-1} \epsilon \left(1 + \frac{f\left(\frac{N\gamma-1}{N}\right)}{g\left(\frac{N\gamma-1}{N}\right)}\right) = g\left(\frac{N\gamma-2}{N}\right)^{-1} \epsilon + g\left(\frac{N\gamma-1}{N}\right) \epsilon \left(\frac{f\left(\frac{N\gamma-1}{N}\right)}{g\left(\frac{N\gamma-2}{N}\right)}\right) \quad (2.27)$$

Per i calcoli visti in 2.11 e 2.12 l'equazione precedente si può scrivere

$$P_{N\gamma-2} = g \left(\frac{N\gamma-2}{N} \right)^{-1} \epsilon + g \left(\frac{N\gamma-1}{N} \right) \epsilon \exp [W(N\gamma-1) - W(N\gamma-2)] \quad (2.28)$$

Ricorsivamente si trova quindi

$$P_j = \epsilon \sum_{l=j}^{N\gamma-1} g \left(\frac{l}{N} \right)^{-1} \exp [W(l) - W(j)] \quad (2.29)$$

E poiché la probabilità totale è normalizzata ad 1 si ha

$$\epsilon^{-1} = \sum_{j=0}^{N\gamma-1} \sum_{l \geq j} g \left(\frac{l}{N} \right)^{-1} \exp [W(l) - W(j)] \quad (2.30)$$

Sostituendo poi il risultato trovato in 2.17 per l'approssimazione nel continuo con $N \gg 1$ si ottiene

$$\epsilon^{-1} = N^2 \int_0^\gamma dx \int_x^\gamma g(y)^{-1} \left(\frac{u(y)}{u(x)} \right)^{1/2} \exp [N(U(y) - U(x))] \quad (2.31)$$

Dove N^2 deriva dai cambi di variabile $l \rightarrow y = l/N$ e $j \rightarrow x = j/N$. Il contributo principale all'integrale è dato per $x \simeq \alpha$ e $y \simeq \beta$, con un'approssimazione gaussiana dell'integrando vicino a questi valori otteniamo

$$\epsilon = \frac{1}{2\pi N} \left[(1 - \phi'(\alpha)) (\phi'(\beta) - 1) \frac{\alpha - \alpha^2}{\beta - \beta^2} \right]^{1/2} \exp [-N(U(\beta) - U(\alpha))] \quad (2.32)$$

Il tasso di transizione dallo stato disordinato a quello ordinato quindi dipende esponenzialmente dalla differenza $\Delta = (U(\beta) - U(\alpha))$ tra il massimo ed il minimo disordinato. Questa è infatti l'altezza della barriera di potenziale che la popolazione deve oltrepassare per diventare viva. Si nota che ϵ non dipende da $U(\gamma)$, infatti, una volta passato il massimo $U(\beta)$, il tempo necessario a scivolare nello stato ordinato è trascurabile.

Similmente si può calcolare il tasso di transizione inverso, la formula sarà la 2.32 con γ al posto di α .

Calcolate queste quantità fondamentali in completa generalità, ci concentriamo ora sulla forma analitica di $\phi(x)$. Scegliamo la funzione in modo che dipenda da due soli parametri a e b , che rappresentano, rispettivamente, la varietà delle specie di monomeri e l'efficienza dei catalizzatori polimerizzanti. Assumendo che la popolazione contenga $(a+1)$ specie di monomeri ugualmente abbondanti⁹ si ha che

$$\phi(0) = (1+a)^{-1} \quad (2.33)$$

è la probabilità che un sito qualsiasi di un polimero venga correttamente occupato da un monomero, in totale assenza di catalisi. Ogni catalizzatore produce una differenza tra l'energia di attivazione per il posizionamento di un monomero attivo e di uno inattivo. Se ogni monomero è attivo, la probabilità che un nuovo monomero sia a sua volta attivo è data da

$$\phi(1) = \left(1 + \frac{a}{b} \right)^{-1} \quad (2.34)$$

In particolare, detta k_B la costante di Boltzmann e T la temperatura assoluta, la differenza nell'energia di attivazione sarà una quantità

$$d = k_B T \ln b \quad \Rightarrow \quad b = \exp \left(\frac{d}{k_B T} \right) \quad (2.35)$$

⁹Si sta qui facendo un'ulteriore approssimazione rispetto a quanto riportato nell'assunto 8, ponendo di fatto $q \simeq 0$; questa situazione si ottiene con monomeri molto reattivi, in modo tale che appena uno si slega da un polimero un altro prende immediatamente il suo posto.

In una popolazione con una frazione x di monomeri attivi, la differenza tra le energie di attivazione per una corretta o incorretta acquisizione sarà xd . Stiamo qui di nuovo usando l'approssimazione di campo medio, assumiamo cioè che l'effetto medio di una popolazione di catalizzatori variamente imperfetti, sia uguale all'effetto di un singolo catalizzatore con efficacia proporzionale all'attività media x della popolazione. Siamo ora pronti per enunciare l'ultimo assunto necessario alla caratterizzazione del modello:

10. Si ha che il rate di formazione di un legame "attivo" è maggiore rispetto a quello di uno "inattivo" per un fattore $\Psi(x) = b^x$; si ottiene quindi, grazie a 2.3:

$$\phi(x) = (1 + ab^{-x})^{-1} = (1 + \exp(A - Bx))^{-1} \quad (2.36)$$

Dove $A = \ln a$ e $B = \ln b$.

Con $\phi(x)$ come in 2.36, possiamo calcolare esplicitamente il potenziale $U(x)$: la formula 2.16 diviene

$$U(x) = \int_0^x \ln \left(\frac{y}{(1-y)} \right) dy - \int_0^x \ln \left(\frac{(1 + \exp(A - By))^{-1}}{(1 - (1 + \exp(A - By))^{-1})} \right) dy \quad (2.37)$$

$$U(x) = y \ln \frac{y}{1-y} - \int_0^x \left(\frac{1}{(1-y)} \right) dy - \int_0^x \ln \left(\frac{1}{(1 + \exp(A - By))} \cdot \frac{(1 + \exp(A - By))}{\exp(A - By)} \right) dy \quad (2.38)$$

$$U(x) = y \ln \frac{y}{1-y} - \int_0^x \left(\frac{1}{(1-y)} \right) dy + \int_0^x (A - By) dy \quad (2.39)$$

$$U(x) = \left[y \ln \frac{y}{1-y} + \ln(1-y) + Ay - \frac{1}{2}By^2 \right]_0^x \quad (2.40)$$

$$U(x) = x \ln x + (1-x) \ln(1-x) + Ax - \frac{1}{2}Bx^2 \quad (2.41)$$

Perchè le condizioni 9 e 10 siano rispettate contemporaneamente sarà necessario imporre alcune condizioni sui parametri a e b . Una condizione necessaria affinché la curva $y = \phi(x)$ intersechi la retta $y = x$ è che la derivata prima $\phi'(x)$ sia maggiore di uno nel punto in cui la funzione inverte la propria concavità (cioè dove $\phi''(x)=0$). Date quindi le derivate di $\phi(x)$

$$\phi'(x) = \frac{B \exp(A - Bx)}{(1 + \exp(A - Bx))^2} \quad \text{e} \quad \phi''(x) = 2 \frac{(B \exp(A - Bx))^2}{(1 + \exp(A - Bx))^3} - \frac{B^2 \exp(A - Bx)}{(1 + \exp(A - Bx))^2} \quad (2.42)$$

Imponendo $\phi''(x) = 0$ e moltiplicando per $\frac{(1 + \exp(A - Bx))^3}{B^2}$ si ottiene

$$0 = 2 \exp[2(A - Bx)] - (1 + \exp(A - Bx)) \exp(A - Bx) \quad \Rightarrow \quad (2.43)$$

$$0 = \exp[2(A - Bx)] - \exp(A - Bx) \quad \Rightarrow \quad \exp(A - Bx) = 0 \vee 1 \Rightarrow \quad (2.44)$$

$$0 = A - Bx \quad \Rightarrow \quad x = \frac{A}{B} \quad (2.45)$$

Dove si è scartata la soluzione con l'esponenziale uguale a zero. Sostituendo questo risultato nella disequazione $\phi'(x) > 1$ si ottiene

$$\phi' \left(\frac{A}{B} \right) = \frac{B \exp(0)}{(1 + \exp(0))^2} = \frac{B}{4} > 1 \quad \Rightarrow \quad B > 4 \quad (2.46)$$

Il modello riproduce quindi risultati sensati, nel senso che esistono i due stati ordinati tra cui è possibile avvengano salti, per valori di $b > e^4 \simeq 55$. I catalizzatori devono quindi avere un fattore di discriminazione dell'ordine delle centinaia. I moderni enzimi hanno fattori di discriminazione con valori tra 10^3 e 10^4 , il modello quindi funziona anche con protoenzimi dalle capacità catalitiche piuttosto ridotte. Un'altra condizione che si può sfruttare, per ottenere indicazioni sul valore di a , è il fatto

che il punto di flesso di cui sopra è l'equilibrio instabile del sistema β e quindi soddisfa $\phi(x) = x$, sostituendo in questa relazione la 2.45 si ha

$$\phi\left(\frac{A}{B}\right) = (1 + \exp 0)^{-1} = \frac{A}{B} \Rightarrow B = 2A \Rightarrow A > 2 \quad (2.47)$$

Il modello contempla l'esistenza di due equilibri stabili, e quindi la possibilità di origine della vita, solo per valori di $a > e^2 \simeq 7.5$; non è applicabile perciò a sistemi che contengano meno di 9 specie diverse di monomeri. È significativo, in particolare, che il modello descritto dai 10 assunti non sia applicabile per $a = 3$, che potrebbe essere il caso che descrive la formazione di polimeri di nucleotidi, cioè di acidi nucleici.

Per valori di a e b che non rispettano le condizioni in 2.46 e 2.47, il modello non consente la transizione tra ordine e disordine: nel sistema c'è un solo stato di equilibrio, che può essere ordinato o meno, ma non è possibile che i due equilibri stabili coesistano. Una prova di ciò si può trovare nel fatto che per $B < 4$ si ha $U''(x) > 0 \forall x \in [0, 1]$, infatti

$$U'(x) = \ln x - \ln(1-x) + A - Bx \quad ; \quad U''(x) = \frac{1}{x} + \frac{1}{1-x} - B \quad (2.48)$$

$$U''(x) = \frac{1 - B(x - x^2)}{(x - x^2)^2} \leq 4 \left(1 - \frac{B}{4}\right) \quad \forall x \in [0, 1] \quad (2.49)$$

In questo caso il modello rappresenta una situazione in cui una cellula è incapace di vivere o incapace di morire. In effetti, esiste un range di valori di a e b che permette transizioni dall'ordine al disordine e viceversa, come illustrato in figura 2.2.

Il grafico rappresenta uno spazio delle fasi con coordinate (a, b) , in cui ogni punto rappresenta una composizione chimica diversa, in termini di numero di specie e di efficienza dei catalizzatori, del sistema. La linea tratteggiata è la curva $b = a^2$ e rappresenta lo stato simmetrico per cui $\beta = 1/2$ e $\gamma = 1 - \alpha$. Lo spazio è diviso in tre regioni: quella centrale, chiamata "Transition region", è quella di interesse biologico, in cui il modello descrive protocellule in grado di passare tra stati con livelli di ordine diverso. La regione in alto a sinistra rappresenta i casi in cui l'efficienza catalitica dei composti è talmente elevata, e la varietà di monomeri così ridotta, che il sistema è costretto nello stato più ordinato, senza possibilità di morire. Dyson chiama questa regione "...Garden of Eden" because this phrase has been used to describe an alternative theory of the origin of life" (Dyson, 1999, p.67). La regione in basso a destra invece comprende i modelli che presentano solo stato disordinato e non quello ordinato, a causa di una troppo debole attività catalitica e/o di una eccessiva varietà di monomeri. Tale sezione dello spazio prende il nome di "hot sulphide soup" in quanto questa espressione è stata usata per caratterizzare i venti idrotermali nelle profondità oceaniche, dove si pensa che la vita potrebbe effettivamente aver avuto origine.

Il range preciso dei valori dei parametri A e B è dato da

$$A_- \leq A \leq A_+ \quad \text{dove} \quad A_{\pm} = 1 + e^{\pm\theta} \mp \theta \quad (2.50)$$

$$\text{con } \theta \text{ definito da} \quad B = A_- + A_+ = 2 + 2 \cosh \theta \quad (2.51)$$

Le condizioni 2.50 e 2.51 definiscono nel piano (a, b) la cuspidi del grafico in figura 2.2.

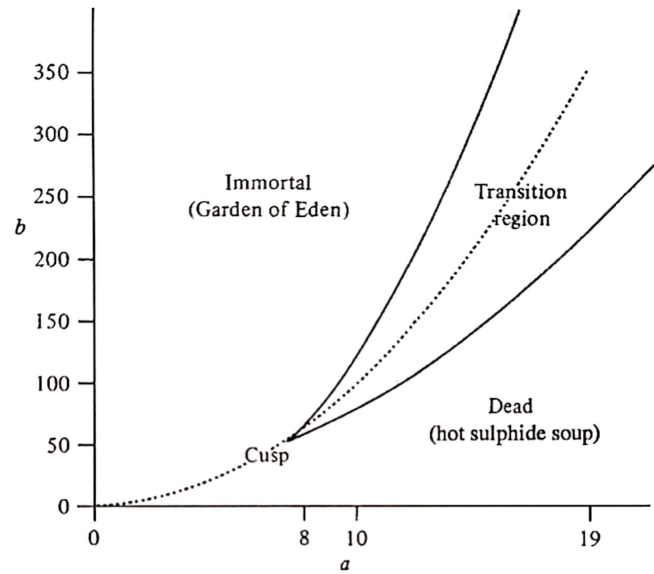


Figura 2.2: Spazio delle fasi (a, b) , diviso in tre regioni. (Dyson, 1999, p.65)

Nel suo articolo Dyson prende in considerazione tre casi:

- Caso "simmetrico", per cui $A = \frac{B}{2}$; $b = a^2$
- Caso "*marginally immortal*", per cui $A = A_-$
- Caso "*marginally alive*", per cui $A = A_+$

Come detto precedentemente, il caso simmetrico è caratterizzato dalla curva $\phi(x)$ che è, appunto, simmetrica rispetto al punto di inflessione in $\phi(x) = x = 1/2$; i punti di intersezione della curva con la bisettrice del primo quadrante sono:

$$\alpha, \quad \beta = 1/2, \quad \gamma = 1 - \alpha$$

e il potenziale $U(x)$ ha minimi uguali in α e γ , è infatti anch'esso simmetrico rispetto al massimo β . Ovviamente non vi è alcuna ragione fisica o chimica che ci faccia pensare che il sistema sia perfettamente simmetrico, tuttavia è lecito pensare che l'efficienza dei catalizzatori aumenti con il numero di specie di monomeri. Per $A = A_-$ e $A = A_+$ ci troviamo nel caso limite in cui l'equazione $\phi(x) = x$ è risolta per solo due valori, cioè si ha rispettivamente $\alpha = \beta = [1+e^\theta]^{-1}$ e $\beta = \gamma = [1+e^{-\theta}]^{-1}$. Il potenziale $U(x)$ in questa situazione presenta un minimo (rispettivamente in γ e α) ed un punto di flesso, con $U'(x) = U''(x) = 0$ in β . Si capisce quindi come questi siano effettivamente i casi limite per cui è possibile la transizione tra ordine e disordine: per valori di a e b che non rispettano le condizioni 2.50, 2.51 il potenziale avrà un solo punto di minimo e non sarà possibile alcun salto. I due casi rappresentano, rispettivamente, il più alto ed il più basso livello di efficienza dei catalizzatori che permette la transizione tra stati a diverso grado di ordine, dato un numero di specie di monomeri. Vediamo ora di calcolare analiticamente alcune quantità che ci permetteranno di valutare l'applicabilità del modello per i tre casi sopra illustrati.

Per quanto riguarda il caso simmetrico, conviene esprimere tutte le quantità in termini di un parametro q così definito:

$$\alpha = (q+1)^{-1} \quad ; \quad \gamma = q(1+q)^{-1} \quad ; \quad q > 1 \quad (2.52)$$

Si ha quindi, per $x = \alpha$ e $2A = B$

$$\phi(\alpha) = \alpha \quad \Rightarrow \quad 1 + \exp\left(A - \frac{B}{q+1}\right) = q+1 \quad \Rightarrow \quad A - \frac{2A}{q+1} = \ln q \quad \Rightarrow \quad (2.53)$$

$$A = \frac{q+1}{q-1} \ln q \quad (2.54)$$

Avendo poi imposto $\beta = 1/2$, si ottiene che il coefficiente di $(-N)$ nella formula 2.32 è

$$\Delta = U(\beta) - U(\alpha) = \ln \frac{1}{2} + \frac{q+1}{4(q-1)} \ln q - \frac{\ln(1+q)}{1+q} - \frac{q}{1+q} \ln\left(\frac{q}{1+q}\right) - \frac{\ln q}{q-1} + \frac{\ln q}{q^2-1} = \quad (2.55)$$

$$= \ln\left(\frac{1+q}{2}\right) - \left(\frac{3q+1}{4q+4}\right) \ln q \quad (2.56)$$

Il tempo medio che una popolazione impiega ad effettuare la transizione da ordine a disordine sarà:

$$t = F\tau \exp(N\Delta) \quad (2.57)$$

dove F è un fattore numerico che dipende da q , τ è il tempo medio che intercorre tra due mutazioni successive, N è il numero totale di monomeri della popolazione e Δ è definito dall'equazione 2.56.

Nel caso *marginally alive* è conveniente esprimere tutto di nuovo in termine di un parametro positivo ω , tale che

$$A = \frac{\omega^2}{p} \ln \frac{p}{r} \quad ; \quad B = \omega^2 \left(\frac{1}{p} + \frac{1}{r}\right) \quad (2.58)$$

$$\text{Dove } p = e^\omega - 1 - \omega \quad ; \quad r = e^{-\omega} - 1 + \omega \quad (2.59)$$

Le soluzioni dell'equazione $\phi(x) = x$ sono allora:

$$\alpha = \frac{r}{p+r} \quad , \quad \beta = \gamma = \frac{r}{\omega(\omega-r)} \quad (2.60)$$

Si ha quindi che la formula 2.57 vale con

$$\Delta = \frac{1}{2} \left(\frac{\omega^2}{p+r} - 1 - \ln \frac{\omega^2}{p+r} \right) \quad (2.61)$$

Per il caso *marginally immortal* si ha che B ha la stessa espressione del caso precedente, mentre

$$A = \frac{\omega^2}{r} \ln \frac{r}{p} \quad e \quad (2.62)$$

$$\alpha = \beta = \frac{p}{\omega(\omega-p)} \quad , \quad \gamma = \frac{p}{p+r} \quad (2.63)$$

In questo caso si ha $\Delta = 0$ per definizione, dato che $\alpha = \beta$; questo sta ad indicare il fatto che la transizione dal disordine all'ordine diventa infinitamente rapida con l'approcciarsi di A al valore A_- .

Ragioniamo ora sui valori di Δ e N che permettono di ottenere valori dell'esponenziale in 2.57 che non siano impossibilmente grandi. Dyson pone un limite superiore al valore di N :

$$N < N_c = 30\Delta^{-1} \quad (2.64)$$

in modo tale che l'esponenziale in 2.57 sia al più $e^{30} \simeq 10^{13}$. La scelta di 30 nell'equazione 2.64 è arbitraria, Dyson arriva a questo numero pensando ad una situazione in cui 10^{10} gocce, ciascuna con una popolazione di N_c monomeri, coesistono in un ambiente favorevole per un tempo di $10^3\tau$. In questa situazione, secondo l'equazione 2.57, un numero sostanziale di popolazioni avrebbe potuto effettuare la transizione dal disordine all'ordine.

Nella tabella 2.1 sono riportati i valori di b , α , β , γ , Δ e N_c corrispondenti a tre diversi valori di a , per i casi *marginally alive*, simmetrico e *marginally immortal*.

| a | b | α | β | γ | Δ | N_c |
|----|-------|----------|---------|----------|------------------------|----------|
| 8 | 62.9 | 0.32 | 0.59 | 0.59 | 0.002695 | 11131 |
| 8 | 64 | 0.33 | 0.50 | 0.67 | 0.001129 ¹⁰ | 26566 |
| 8 | 65.7 | 0.39 | 0.39 | 0.70 | 0 | ∞ |
| 10 | 89.4 | 0.19 | 0.67 | 0.67 | 0.0298 | 1006 |
| 10 | 100.0 | 0.20 | 0.50 | 0.80 | 0.0145 | 2070 |
| 10 | 128.0 | 0.29 | 0.29 | 0.87 | 0 | ∞ |
| 19 | 219.3 | 0.07 | 0.75 | 0.75 | 0.1906 | 157 |
| 19 | 361 | 0.08 | 0.50 | 0.92 | 0.1051 | 285 |
| 19 | 3195 | 0.14 | 0.14 | 0.99 | 0 | ∞ |

Tabella 2.1: Nella tabella sono riportati i valori delle varie quantità per 3 valori di a per i casi (dall'alto verso il basso) *marginally alive*, simmetrico e *marginally immortal*. Tabella presa da (Dyson, 1982; Table 1).

Il caso con $a = 19$ è riportato per considerare il sistema in cui i monomeri siano in effetti i 20 amminoacidi che compongono le moderne proteine; è evidente che il nostro modello non si adatta benissimo a tale situazione. La popolazione critica perché la transizione avvenga con probabilità accettabile è infatti piuttosto contenuta, dell'ordine di 200 monomeri. D'altro canto, se le specie di monomeri sono circa una decina, il modello ammette transizioni dallo stato disordinato a quello ordinato per popolazioni di monomeri decisamente più nutrite: dell'ordine di $2 \cdot 10^3 - 10^4$. I risultati

¹⁰Sia nell'articolo che nel libro, Dyson fa notare che questo valore di Δ corrisponde alla differenza $(\ln 3 - (19/12) \ln 2)$, che è nota ai musicisti in quanto è la differenza frazionaria tra l'intonazione di una vera quinta e di una equitemperata.

suggeriscono quindi che il modello è particolarmente adatto a descrivere situazioni che trovano posto nella regione vicino alla cuspidale del grafico 2.2. Le condizioni che producono risultati ottimali sono le seguenti

- $a \in (8, 10)$
- $b \in (60, 100)$
- $N \in (2000, 20000)$

Questi range sembrano anche sensati da un punto di vista chimico-fisico, è infatti plausibile che una decina di monomeri, magari antenati degli odierni amminoacidi, potessero polimerizzare dando origine a catalizzatori con efficienza di poco superiore a quella di alcuni catalizzatori inorganici. La dimensione della popolazione di monomeri all'interno della protocellula, è anch'essa plausibile: qualche migliaio di monomeri legati a formare qualche centinaio di polimeri fornirebbero sufficiente varietà di strutture per permettere l'esistenza di cicli catalitici interessanti. Un valore di N dell'ordine delle decine migliaia, infatti, è contemporaneamente abbastanza grande da fornire la complessità chimica tipica della vita e abbastanza piccolo da permettere una fluttuazione statistica che consenta il salto da disordine ad ordine. È notevole poi il fatto che, per a e b nei range riportati sopra, i valori di Δ sono molto bassi ($10^{-2} / 10^{-3}$). Ciò indica il fatto che il potenziale $U(x)$ in questo range è quasi piatto, e la barriera che separa i due equilibri stabili in α e γ è molto bassa.

La vera ragione per cui il modello funziona è, secondo Dyson, il fatto che esso tollera un'alta frequenza di errori. In questo modo il modello evita la "catastrofe degli errori" legata al concetto di replicazione: non vi è necessità di un preciso controllo sulle strutture molecolari. In effetti il tasso di errore nello stato ordinato è tra il 20 ed il 30 per cento se a e b sono nei range sopra indicati. Un tasso di errore del 20% significa che quattro su cinque monomeri in ogni polimero sono correttamente piazzati. Un catalizzatore con un sito attivo composto da 5 monomeri ha quindi una possibilità su tre di essere completamente funzionante. Questo livello di errori è tollerabile per un sistema che si basi sulla riproduzione ma non sulla replicazione, ed è il principale motivo per cui il livello ordinato è statisticamente accessibile nel nostro modello.

In conclusione si può affermare che il modello, pur senza scendere in nessun dettaglio per quel che riguarda la natura chimico-fisica delle interazioni tra i monomeri, ci consegna una situazione applicabile a amminoacidi o antenati degli stessi. Viene invece scartata la possibilità che questo modello possa rappresentare una situazione in cui i monomeri sono nucleotidi e i polimeri acidi nucleici. Si ha quindi un risultato che sembra a favore della teoria di Oparin dell'origine puramente metabolica della vita. Ciò è compatibile anche con una teoria della doppia origine della vita, preferita da Dyson.

Capitolo 3

Lavori successivi

In questa sezione si affronteranno due lavori successivi a quello di Dyson, uno del 2011, di Martin A. Nowak e Irene A. Chen [1] e uno del 2020, di H. J. Hilhorst[6], che presentano altri semplici modelli per spiegare il passaggio *"From prelife to life"*, come suggerito dal titolo di entrambi i papers.

3.1 Nowak e Chen

Il lavoro di Nowak e Chen propone un approccio piuttosto diverso da quello dell'articolo di Dyson, che in effetti non è nemmeno citato nella bibliografia. Il modello presentato in questo articolo si inserisce in un contesto che potrebbe essere stato quello agli albori di un mondo ad RNA. È da subito molto chiaro che l'idea di origine della vita degli autori è quella di Eigen e Orgel: *"Life emerges with the ability of replication"* (Nowak e Chen, 2011), per cui vita è sinonimo di replicazione. Nella sezione *"Discussions"* dell'articolo è comunque fatto presente che questo modello potrebbe anche essere utile per la descrizione di polimeri replicanti diversi dagli acidi nucleici, che però, essendo appunto replicanti, comunque non sono sovrapponibili a quelli descritti dal modello di Dyson. L'articolo analizza quindi due situazioni diverse: una di *"prelife"* in cui non vi è replicazione ma semplicemente specie di monomeri polimerizzano, e una invece *"viva"*, in cui alcuni di questi polimeri acquisiscono la capacità di replicarsi. Queste due situazioni sono diverse soprattutto per il diverso peso della selezione nelle stesse: senza la replicazione piccole differenze nella composizione dei polimeri (e quindi nelle costanti di rate che definiscono il modello) portano a piccole differenze nella loro abbondanza; con la replicazione, invece, il ruolo della selezione si fa decisamente più incisivo.

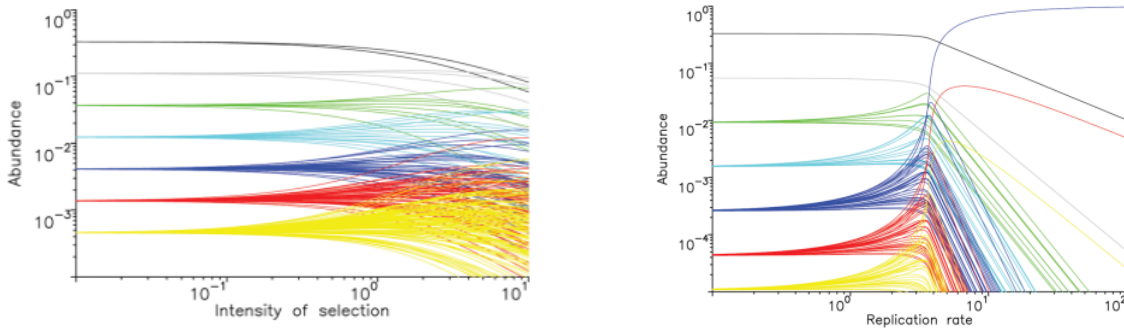
La prima fondamentale semplificazione operata dal modello è quella di considerare solo due tipi diversi di monomeri 0^* e 1^* ¹ che possono polimerizzare in sequenze binarie secondo le reazioni $i + \sigma^* \rightarrow i\sigma$, dove i è una sequenza binaria qualsiasi e $\sigma = 0, 1$. Si assume che i polimeri possano crescere solo in una direzione. Ogni sequenza i ha quindi un precursore i' e due possibili successori $i1$ e $i0$. Questa situazione è rappresentabile con un albero binario, in cui ogni percorso rappresenta un possibile polimero. Il rate con cui il polimero i si forma dal suo precursore i' è dato da a_i ; tale parametro dipende dalla sequenza binaria di i , si avranno quindi polimeri che crescono più velocemente di altri. Per evitare che la popolazione cresca all'infinito è necessario un qualche meccanismo di rimozione dei polimeri in formazione. Assumiamo quindi che le sequenze siano rimosse ad un rate d , inizialmente uguale per tutti. Si ha quindi che il sistema è descritto da un sistema infinito di equazioni differenziali:

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1}) x_i \quad (3.1)$$

Dove x_i denota l'abbondanza di polimeri di tipo i e \dot{x}_i è la derivata rispetto al tempo; l'indice i rappresenta tutte le combinazioni binarie finite. Per semplicità si assume $x_{0'} = x_{1'} = 1$ e che le concentrazioni di 0^* e 1^* siano costanti e comprese all'interno delle costanti di rate a_i . Se tutte le costanti di rate sono positive si ha un equilibrio stabile in cui sono presenti tutte le sequenze. Indicando

¹Nell'articolo i monomeri sono rappresentati con l'asterisco: 0^* se sono liberi, e senza: 0 se sono legati ad altri in un polimero, si manterrà qui la stessa notazione.

con $b_i = a_i/(d + a_{i0} + a_{i1})$, l'abbondanza della sequenza i all'equilibrio è data da $x_i = b_i \cdot b_{i'} \cdot b_{i''} \cdot \dots \cdot b_{i^{(k-1)}}$ e la dimensione della popolazione totale è $\sum_i x_i = (a_0 + a_1)/d$. Questa situazione è rappresentata in figura 3.1a, considerando $a_0 = a_1 = d = 1$ e, per $i \neq 0, 1$, $a_i = 1 + s\xi_i$, dove s è un parametro che rappresenta "l'intensità della selezione" e ξ_i è un numero randomico scelto tra 0 e 1.



(a) Condizione di equilibrio in "prelife" (per l'eq 3.1) in (b) Condizione di equilibrio per l'eq 3.2, in cui polimeri con scenario randomico: ξ_i è scelta casualmente nell'intervallo $[0, 1]$. $k=3, 4, 5$ possono replicarsi con f_i scelta randomicamente nell'intervallo $[0, 1]$.

Figura 3.1: Abbondanza delle popolazioni di 2^k diversi polimeri di lunghezza k , con $k=1$ (nero), $k=2$ (grigio), $k=3$ (verde), $k=4$ (azzurro), $k=5$ (blu), $k=6$ (rosso), $k=7$ (giallo). (Nowak e Chen, 2011)

La figura 3.1b, invece, rappresenta l'abbondanza dei polimeri nello stato di equilibrio nel caso in cui siano presenti fenomeni replicativi, in funzione del rate globale di replicazione r . Ogni sequenza avrà poi un rate di replicazione relativo (fitness) f_i . Anche in figura 3.1b si è posto $a_0 = a_1 = d = 1$ con però ogni altro $a_i = 0.5$. Questa dinamica è descritta dal sistema di equazioni differenziali

$$\dot{x}_i = a_i x_{i'} - (d + a_{i0} + a_{i1})x_i + (r f_i - \phi)x_i \quad (3.2)$$

Dove ϕ è un rate di morte dei polimeri che dipende dalla densità degli stessi. Si impone $\phi = a_0 + a_1 - 1 - d + r \sum_i x_i f_i$ così da avere la dimensione della popolazione uguale a 1. Confrontando i due grafici in figura 3.1a e 3.1b si nota come lo scenario prebiotico contempli una coesistenza delle popolazioni diverse di polimeri, mentre lo scenario con la replicazione porta a competizione esclusiva. Infatti, per grandi valori di r , si ha che una sola specie di lunghezza 5 monomeri, che evidentemente ha il valore di fitness f_i maggiore, costituisce la quasi totalità degli individui.

Gli autori traggono delle conseguenze dal loro semplice modello: la prima, piuttosto controintuitiva è che sequenze lunghe possono essere molto più abbondanti di sequenze più corte. Questo vale sia in condizioni di "prelife", se s e ξ_i sono particolarmente grandi, che, soprattutto, con l'introduzione del meccanismo replicativo. In questo caso il replicatore più lungo ed efficiente può dominare in modo drammatico sfruttando l'abbondanza di precursori, dovuta alla capacità anche degli stessi di replicarsi. Questa tendenza è particolarmente interessante perché sembra puntare ad una evoluzione nel senso di una maggiore complessità. Un altro aspetto interessante è che in questo modello la vita replicante prevale in modo schiacciante sui polimeri incapaci di replicazione. Ciò è dovuto essenzialmente al fatto che i polimeri

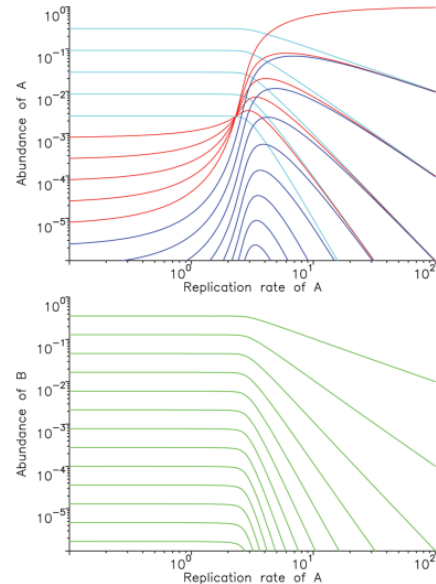


Figura 3.2: Nella figura in alto è rappresentata l'abbondanza delle stringhe di tipo A, quelle con lunghezza da 6 a 10 monomeri (rosse) possono replicarsi; Nella figura in basso è rappresentata l'abbondanza delle stringhe di tipo B, incapaci di replicazione. Si è imposto $a_A = a_B = d_A = 1$, $d_B = 0.5$, $f_i = 1 \forall i \in [6, 10]$ e $f_i = 0$ altrimenti. (Nowak e Chen, 2011).

non replicanti, anche se catalitici, dovrebbero "andare per tentativi", e produrrebbero molti polimeri diversi da loro prima di uno uguale. Questa situazione di competizione tra un sistema replicante A e uno non replicante B è rappresentata in figura 3.2, in cui, per semplicità si sono raggruppate in una sola linea tutte le sequenze della stessa lunghezza, senza tenere conto delle differenze di composizione delle stesse. Si vede dal grafico che, per valori di r abbastanza grandi, il polimero replicante più lungo, quello che appartiene ad A con 10 monomeri surclassa tutti gli altri e a maggior ragione i polimeri non replicanti. Queste argomentazioni potrebbero andare a supporto della seconda origine della vita nella teoria della doppia origine, spiegando come, una volta emersi dei sistemi replicanti, questi avrebbero avuto un grosso vantaggio evolutivo, eliminando di fatto i loro antenati.

3.2 Hilhorst

Il lavoro di Hilhorst prende molto da quello, precedente di quasi quarant'anni, di Dyson. La trattazione matematica del modello è piuttosto lunga e complessa, ci si limita quindi a presentare i risultati principali del lavoro, senza scendere nei dettagli tecnici. In particolare la situazione immaginata è quella di un reticolo unidimensionale con N siti² a cui si possono legare polimeri, anche in questo caso immaginati come sequenze binarie randomiche, che si sviluppano poi in lunghezza perpendicolarmente allo stesso. La lunghezza media dei polimeri è controllata da un parametro z , chiamato fugacità. Due polimeri su siti adiacenti hanno un fattore di peso ω per ogni monomero uguale; tale parametro rappresenta l'interazione polimero-polimero. Il diagramma di fase nel piano $z\omega$ in figura 3.3 mostra una linea critica $z = z_c(\omega)$, che divide la regione I dalla regione III.

Nella regione I esiste un vero stato di equilibrio con una lunghezza media dei polimeri finita e precisa $\langle n \rangle$; nelle regioni II e III, invece l'equilibrio è instabile e la crescita dei polimeri potrebbe essere illimitata. In particolare, nella regione II, per cui $z > 1/2$, la crescita illimitata è dovuta alla sola pressione della fugacità z ; nella regione III invece, tale condizione è dovuta all'effetto combinato della fugacità z e dell'interazione polimero-polimero ω . La linea che separa I da III è data dall'equazione $\omega = (2z^q)^{\frac{-1}{q-1}}$, con q che rappresenta il numero di polimeri, o equivalentemente di siti, presenti. Gli stati nella regione III sono analoghi a quelli che negli altri modelli abbiamo chiamato autocatalitici o replicativi, quindi quelli "vivi". La caratteristica di questo modello che più lo distingue dagli altri affrontati in questo elaborato è la presenza di una forma di interazione polimero-polimero. Nel modello si prende in considerazione il fatto che, perché un polimero favorisca in modo catalitico la creazione di un altro, le loro sequenze di monomeri devono essere in una relazione precisa. Una conseguenza di questo aspetto è che la crescita dei polimeri è possibile solo come fenomeno collettivo, e porta alla proliferazione di un set limitato di specie. È notevole il fatto che, in questo senso, l'articolo di Hilhorst arriva, come quello di Nowak e Chen, alla conclusione che in uno stato di prelife è possibile la convivenza di diverse specie, mentre la vita porta ad una situazione di competizione esclusiva. La conclusione principale del lavoro è, secondo gli stessi autori, che la selezione di alcune specie a discapito di altre potrebbe essere il risultato di un'interazione tra le stesse; e che un set di parametri random, anche se chimicamente realistici, non è una condizione necessaria perché ciò avvenga. L'aspetto però forse più rivoluzionario del lavoro di Hilhorst è che l'esistenza della regione III, e quindi di una "fase viva", non è dovuta ad un termine autocatalitico specificatamente inserito, come per i lavori di Dyson e Nowak, ma emerge come conseguenza dell'interazione a livello dei singoli monomeri di diverse specie di polimeri.

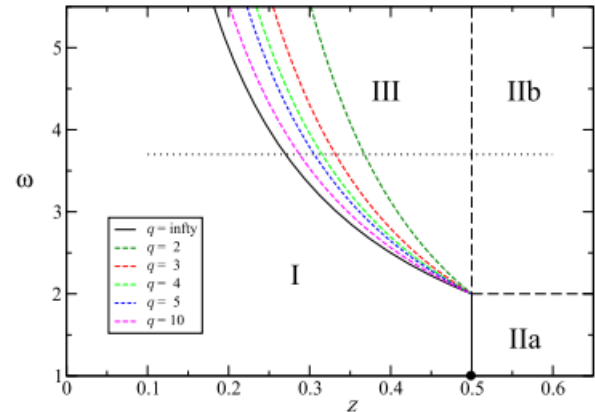


Figura 3.3: Spazio delle fasi nel piano $z\omega$ diviso in regioni (Hilhorst, 2020)

²L'idea di un reticolo con i polimeri che si sviluppano legati a questo è la stessa con cui Dyson presenta il suo modello nel libro[3].

Bibliografia

- [1] Nowak MA. Chen IA. “From prelife to life: how chemical kinetics become evolutionary dynamics.” In: *Acc Chem Res.* 45(12) (Dec 2012), pp. 2088–96. DOI: 10.1021/ar2002683.
- [2] Charles Darwin. *Letter to Hooker*. URL: <https://www.darwinproject.ac.uk/letter/DCP-LETT-7471.xml>. (accessed: 15.11.2021).
- [3] Freeman J Dyson. *Origin of life*. Cambridge university press, 1999.
- [4] Freeman J. Dyson. “A Model for the Origin of Life”. In: *J. Mol. Evol.* 18 (1982), pp. 344–350. DOI: <https://doi.org/10.1007/BF01733901>.
- [5] Sidney W. Fox. *Aleksandr Oparin*. URL: <https://www.britannica.com/biography/Aleksandr-Oparin>. (accessed: 23.11.2021).
- [6] H J Hilhorst. “From prelife to life: a bio-inspired toy model”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2020.7 (lug. 2020), p. 073203. DOI: 10.1088/1742-5468/ab99c2. URL: <https://doi.org/10.1088/1742-5468/ab99c2>.
- [7] A. Lazcano. “Alexandr I. Oparin and the Origin of Life: A Historical Reassessment of the Heterotrophic Theory.” In: *J. Mol. Evol.* 83 (2016), pp. 214–222. DOI: <https://doi.org/10.1007/s00239-016-9773-5>.
- [8] Lazcano A. Miller S. Schopf J. “Oparin’s “Origin of Life”: Sixty Years Later.” In: *J. Mol. Evol.* 44 (1997), pp. 351–353. DOI: <https://doi.org/10.1007/PL00006153>.
- [9] Aleksandr Ivanovič Oparin e Sergius Morguli. *The Origin of life*. Dover, 1953.
- [10] Erwin Schrödinger. *What is life?* Cambridge university press, 1944.