

# DALL'INFORMAZIONE ALLA COMPLESSITÀ: UNA METAFORA DEI SISTEMI CHIMICI, BIOLOGICI E INGEGNERISTICI

La struttura del DNA ha evidenziato che la biologia costituisce un aspetto della scienza dell'informazione, aprendo interessanti prospettive nella descrizione delle caratteristiche e dell'evoluzione dei sistemi viventi a partire dalle cellule. Il confronto dei risultati così ottenuti con quelli sperimentali concernenti la sintesi di cellule con minimo genoma (JCVI-syn3.0) lascia prevedere importanti sviluppi nelle ricerche in corso nei settori della biologia sintetica e dell'ingegneria metabolica.



## La termodinamica dell'informazione

“Chiamala entropia. Nessuno sa cosa sia per cui nei dibattiti ti troverai avvantaggiato”. Con queste parole John von Neumann, il leggendario matematico, commentava una formula proposta da Claude Shannon nel 1948 per misurare l'informazione contenuta in un messaggio. Ma che cosa è l'informazione, e soprattutto che cosa la rende così importante?

Shannon, ingegnere alla Bell, aveva osservato che il contenuto di un messaggio si può esprimere mediante una serie di decisioni binarie, come si procede nel gioco in cui mediante una successione di sì e di no si arriva ad identificare la carta di un mazzo o una persona presente in un gruppo. La formula assume una forma semplice se riguarda  $\Omega$  elementi aventi la stessa probabilità. Identificando l'informazione  $I$  con il numero delle scelte consegue che  $\Omega=2^I$ . Ovvero  $I=\log_2 \Omega$ , relazione che richiama l'equazione  $S=k_B \ln \Omega$ ,

( $k_B=1,3805 \times 10^{-23} \text{ JK}^{-1}$ ), introdotta nella seconda metà dell'Ottocento da Boltzmann, per valutare l'entropia  $S(E)$  di un sistema termodinamico la cui energia  $E$  è distribuita fra  $\Omega$  stati equiprobabili. Ricordando che l'entropia è espressa dal rapporto fra l'energia e la temperatura assoluta  $T$ , ne consegue che [1]:

$$I = \log_2 \Omega = S(E)/k_B \ln 2 = E/k_B T \ln 2$$

Questa relazione evidenzia che l'informazione ha le stesse radici dell'entropia che misura la nostra ignoranza sulla distribuzione dell'energia fra i diversi stati di un sistema termodinamico del quale conosciamo il valore di grandezze macroscopiche, quali la temperatura e il volume. Da essa risulta anche che l'energia associata all'unità di informazione o bit, corrispondente a  $I=1$ , ammonta a  $k_B T \ln 2$ .

Rolf Landauer, un matematico che operava presso l'IBM, ha successivamente evidenziato che i bit sono sempre connessi a particolari situazioni fisiche, ad esempio i buchi di una carta perforata, lo stato di un neurone e al limite lo spin *up* o *down* di una particella. Pertanto l'informazione non costituisce un'astrazione, anche perché la sua cancellazione richiede un'energia, pari a  $k_B T \ln 2$  per ogni bit.

## Un calcolatore chiamato Universo

L'Universo si può assimilare ad una sfera avente un raggio espresso dal prodotto della velocità della luce per il tempo trascorso dalla sua nascita, ovvero dal Big Bang. Moltiplicandone il volu-

me per la densità della materia si ricava la massa, che, moltiplicata a sua volta per il quadrato della velocità della luce, fornisce il suo contenuto di energia come ci ha insegnato Einstein [2].

Poiché l'inversione di un bit da *up* a *down* o viceversa comporta una variazione di energia  $E$ , Norman Margolus e Lev Levitin, memori del principio di indeterminazione di Heisenberg, hanno dimostrato che la transizione avviene in un intervallo di tempo  $\tau=h/4E$ , essendo  $h$  la costante di Planck. Applicando questo risultato a tutta l'energia presente nell'Universo si ricava che è in grado di fare  $10^{103}$  operazioni al secondo, per cui nei suoi 14 miliardi di anni di vita può aver eseguito circa  $10^{122}$  operazioni. Numero enorme, che sancisce il limite delle elaborazioni entro le quali hanno operato le leggi della fisica.

Metaforicamente l'universo si può pertanto assimilare ad un gigantesco calcolatore, l'unico in grado di rispondere alla domanda: “Si potrà ricostruire il Big Bang?”, con: “Fiat lux!”. Ma che cosa ha fabbricato fruendo di tale gargantuesca capacità di calcolo? L'attenzione viene subito posta alla sintesi di quei composti chimici che giocano un ruolo essenziale nei processi vitali, quali le proteine, fondamentali non solo per le loro proprietà strutturali, ma anche per le loro funzioni catalitiche. È improbabile che nell'intero Universo esistano molecole con proprietà superiori a quelle delle proteine. Trattandosi di polimeri costituiti da 20 tipi di amminoacidi ne consegue che ad una catena lunga  $n$  unità corrispondono  $20^n$  isomeri [3]. Per un valore medio ragionevole di  $n=200$ , si ottengono  $20^{200}=10^{260}$  molecole. Si

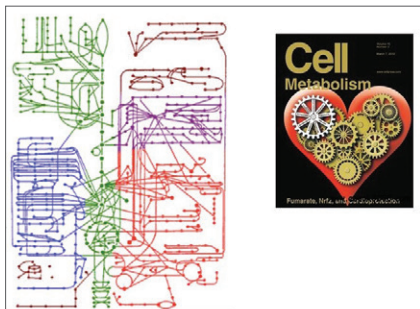


Fig. 1 - Gli organismi trasformano materiali organici dell'ambiente e li convertono in altri utili per la loro sopravvivenza. Gli attori di tali operazioni sono i cicli metabolici, nel cui ambito si svolgono le corrispondenti reazioni chimiche, articolati in complessi network

tratta di un numero impressionante, molto più alto di quello dei bit messi a disposizione dall'Universo durante la sua vita. Il fatto che le osservazioni spettroscopiche evidenzino la presenza nelle nubi galattiche fredde di molti e complicati composti della chimica organica, suggerisce allora che le loro sintesi facciano intervenire dei complessi meccanismi organizzatori in grado al limite di generare la vita.

La complessità costituisce però un concetto elusivo poiché Seth Lloyd ne ha trovato in letteratura quarantadue definizioni, ovvero un numero superiore alle ricette che vengono propagandate per fabbricare una pizza! Nei sistemi biologici si manifesta nelle interazioni fra una moltitudine di molecole organiche, organizzate in reti, o *networks*, nel cui ambito fluiscono le informazioni. Tipici sono i cicli metabolici presenti nelle cellule grazie ai quali si produce energia, acidi nucleici, proteine e lipidi, il cui controllo viene esercitato influenzando sul comportamento dei singoli enzimi che catalizzano le varie reazioni (Fig. 1).

Ma quali sono i meccanismi evolutivi in grado di favorire l'emergenza di strutture chimiche capaci di incorporare un contenuto di informazione che permetta la loro autoriproduzione?

### Il cammino dell'evoluzione

Circa 150 anni fa Charles Darwin spiegò come l'evoluzione abbia trasformato le prime forme di vita comparse sulla Terra nella presente ricca panoplia di specie viventi mediante un processo che si è protratto per circa 3,5 miliardi di anni (Fig. 2). Attualmente si ritiene che sia esistito un precursore chiamato IDA (*Initial Darwinian Ancestor*) costituito da molecole non ancora identificate, che si è trasformato in LUCA, acronimo di "Last Universal Common Ancestor", ovvero di

un organismo primordiale in grado di accumulare le informazioni che hanno generato le diverse forme di vita esistenti sulla Terra [4]. Ma in che modo sono state trasmesse?

Horace Judson [5] ha definito ottavo giorno della creazione quello della scoperta della struttura del DNA, perché ha trasformato profondamente la biologia, evidenziando che costituisce un aspetto della scienza dell'informazione grazie alla disposizione sequenziale dei nucleotidi, che codifica le proteine, e alla complementarità che lega perfettamente un filo dell'elica con il suo partner, che trasmette l'eredità. Il DNA è il blueprint per la costruzione di un organismo in grado di sopravvivere nel migliore dei modi nel suo ambiente nativo e di trasferire informazioni alla sua progenie. Ne consegue che la vita dovrebbe essere pensata non come un evento chimico, ma come una continua elaborazione di informazioni. Il genoma ne è depositario mediante una collezione di bit, accumulati durante il processo evolutivo. La sua complessità si identifica pertanto con l'ammontare delle informazioni che vengono trasferite alla sintesi delle proteine.

La termodinamica ci insegna che un sistema, a temperatura e pressione costanti, nel raggiungere l'equilibrio con l'ambiente con il quale è a contatto, è in grado di produrre un lavoro massimo uguale alla variazione di energia libera di Gibbs, che a sua volta si dimostra essere proporzionale all'informazione [6]:

$$\Delta G = T(S_{eq} - S) = T(k_B \ln 2)$$

L'elaborazione di tale informazione associata al flusso di exergia proveniente dal Sole, eseguita dal calcolatore Terra ha contribuito ad aumentare la complessità della biosfera durante un processo che perdura dall'inizio della storia del pianeta. Il tasso del flusso di exergia sancisce pertanto il limite superiore dello sviluppo. Esistono comunque elevati margini, poiché il flusso solare ammonta a 140.000 TW, quello impiegato dalla fotosintesi 90 TW e quello impiegato nelle attività umane a 15 TW. Purtroppo i processi naturali di trasformazione hanno un rendimento molto basso, per cui gran parte di tale energia che interviene anche nei processi atmosferici e superficiali del nostro inquieto pianeta, viene disperso in processi irreversibili.

Mediante un programma di calcolo, battezzato Avida [7], si possono condurre esperimenti numerici su una moltitudine di singoli programmi in grado ciascuno di replicarsi ed evolvere in un

ambiente soggetto a fluttuazioni. Si tratta di uno strumento destinato a ricerche sull'evoluzione della complessità non solo dei sistemi biologici, ma anche di quelli tecnologici e delle società. Le trasformazioni di ciascun organismo digitale, gestite da istruzioni che costituiscono il loro genoma, riflettono i tre ingredienti tipici dell'evoluzione darwiniana, ovvero:

- le repliche, nelle quali gli elementi copiano le istruzioni del loro codice in una memoria fresca, allungando così la loro memoria prima di dividersi;
- le mutazioni, dovute agli errori digitali di trascrizione che intervengono durante il processo di copiatura, per cui i figli differiscono dai loro genitori;
- la selezione, che elimina le mutazioni sfavorevoli alla sopravvivenza, per cui si può assimilare ad un filtro che agisce come se fosse un diavoleto di Maxwell. In realtà tende a eliminare gli organismi portatori di mutazioni che diminuiscono l'adattamento, o *fitness*, con l'ambiente circostante.

Dalle simulazioni risulta che gli organismi digitali si replicano, mutano e si riproducono per diverse generazioni dando origine a trasformazioni che talora manifestano comportamenti inaspettati [8]. Nella Fig. 3 viene illustrato l'esempio di una popolazione soggetta a trasformazioni analoghe a quelle che i biologi osservano su un disco di Petri.

### La cellula in una scheggia di silicio

Grazie all'impiego di algoritmi evolutivi è possibile approfondire mediante modelli matematici i processi che si svolgono in natura cogliendone l'analogia con quelli produttivi della chimica e dell'ingegneria, come appare dai seguenti esempi:



Fig. 2 - Caricatura di Darwin riportata nel 1882 su Punch

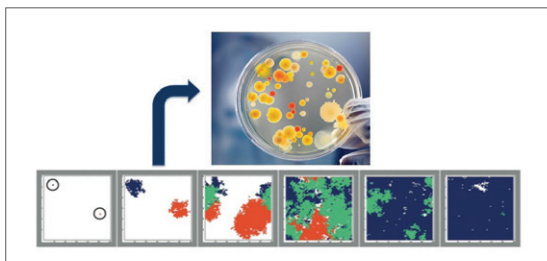


Fig. 3 - Nell'evoluzione di una popolazione arvidiana simulata al computer si avvicendano diverse specie, simili a quelle che si riscontrano sperimentalmente in un disco di Petri [rif. (7)]

- l'insieme delle reazioni chimiche dei cicli metabolici che avvengono nelle cellule per sostenere la vita, può essere assimilato ad un laboratorio chimico che converte materie prime in prodotti ricchi di energia o in grado di conferire una struttura alle funzioni cellulari;
- nell'industria i prodotti chimici vengono ottenuti mediante processi che comprendono diverse reazioni integrate in strutture la cui complessità rivaleggia con quella delle cellule viventi.

In sostanza la biochimica e l'ingegneria chimica grazie alla scienza dell'informazione stanno alla base di una nuova rivoluzione scientifica ispirata dalla simulazione del comportamento dei sistemi biologici, ponendo in primo piano quello delle cellule (Fig. 4). Si tratta di un laboratorio virtuale che permette di evidenziare le risposte di una cellula a stimoli esterni e le sue possibili riprogettazioni su basi razionali. Queste indagini stanno acquistando un particolare rilievo con lo sviluppo della biologia sintetica che combina diverse discipline quali le biotecnologie, la biologia molecolare, la biofisica, la biologia evolutiva ed altre, svolgendo ricerche volte alla progettazione di sistemi biologici che, attraverso l'elaborazione dell'informazione, siano in grado di sintetizzare nuove molecole e nuovi materiali. Dal punto di vista matematico tale modellazione comporta l'integrazione di sistemi di equazioni differenziali non lineari che riflettono i bilanci materiali dei diversi componenti coinvolti in schemi complessi delle reazioni metaboliche [9]. Per tenere conto del traffico molecolare è necessario far intervenire i processi diffusivi ai quali contribuiscono le fluttuazioni stocastiche per cui è talora opportuno impiegare l'equazione di Langevin, introdotta per la descrizione dei moti browniani. Il modello deve contemplare il meccanismo di trascrizione delle informazioni contenute nel genoma per la sintesi delle proteine che opera come viene illustrato [10] nella Fig. 5.

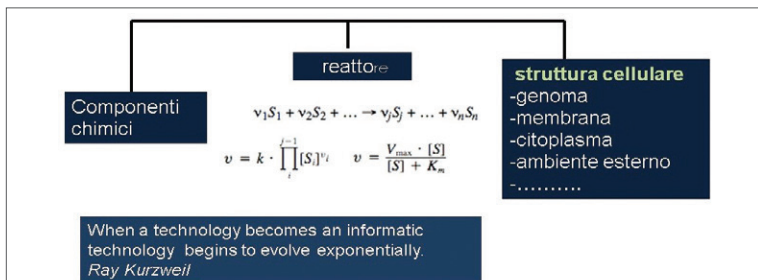


Fig. 4 - Una sfida per l'ingegneria delle reazioni chimiche: la formulazione di un modello in silicio di una cellula

In realtà le attività presenti in una cellula presentano ancora comportamenti ignoti sulle cui origini vengono svolte indagini sperimentali, mentre il lavoro dei computer si svolge in direzione opposta intesa a simulare determinate funzioni. Un interessante programma, chiamato *Whole Cell Simulation* formulato da Markus W. Covert e collaboratori [11], descrive l'intero ciclo del batterio procariotico *Mycoplasma genitalium*. Nel modello sono inclusi i più importanti processi vitali (Fig. 6), quali la trascrizione del DNA nell'RNA e quindi nelle proteine, il metabolismo volto alla produzione di energia e dei componenti strutturali, la replicazione del genoma ed, infine, la scissione della cellula. I risultati dei calcoli si accordano soddisfacentemente con le informazioni sperimentali.

**Verum est factum**

Le concettualizzazioni precedentemente illustrate, anche se offrono degli strumenti con i quali è possibile sviluppare calcoli e simulazioni sul modo in cui operano i sistemi biologici, non ci informano sulle caratteristiche dell' "Initial Darwinian Ancestor (IDA)". Si tratta di una zona ancora buia che mette in evidenza il fatto che nei sistemi fisici concernenti la gestione dell'informazione potrebbero intervenire variazioni del flusso simili alle transizioni di fase. Allo stato attuale, comunque resta insoluto il quesito: "Che cosa è la vita?", anche se in realtà non esiste una formulazione adeguata di questa domanda, per cui solitamente viene elusa presentando una lista delle caratteristiche che contraddistinguono gli organismi viventi, quali il movimento, la respirazione, la nutrizione, la crescita e la riproduzione. Una risposta funzionale compatibile con le leggi della chimica-fisica è stata offerta da Erwin Schroedinger [12], uno degli artefici della meccanica quantistica, che ha scritto: "Un vivente

è un sistema termodinamico aperto in grado di mantenersi in modo autonomo in uno stato energetico stazionario, ed in grado di dirigere una serie di reazioni chimiche verso la sintesi di se stesso". Ineccepibile, si potrebbe dire, ma manca di un ingrediente fondamentale: la capacità di trasferire l'informazione. Gerald Joyce ha affermato che la caratteristica peculiare della vita è quella di subire un'evoluzione darwiniana che prenda origine dalle molecole: "La chimica non ha una storia a differenza della biologia, per cui la nascita della vita si identifica con quella della biologia scritta

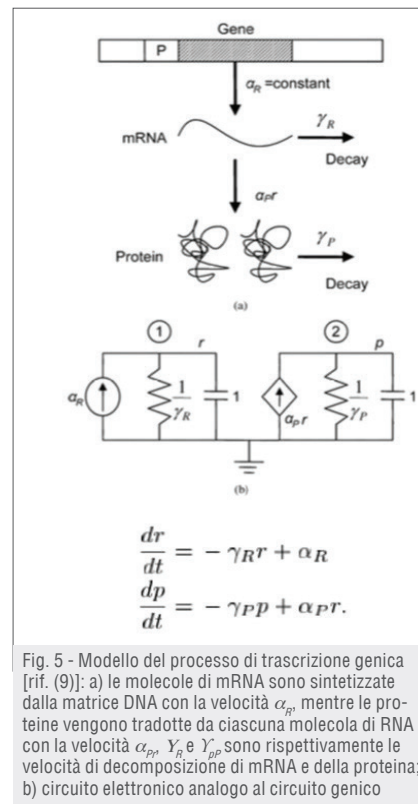


Fig. 5 - Modello del processo di trascrizione genica [rif. (9)]; a) le molecole di mRNA sono sintetizzate dalla matrice DNA con la velocità  $\alpha_R$ , mentre le proteine vengono tradotte da ciascuna molecola di RNA con la velocità  $\alpha_P$ ;  $\gamma_R$  e  $\gamma_P$  sono rispettivamente le velocità di decomposizione di mRNA e della proteina; b) circuito elettronico analogo al circuito genico

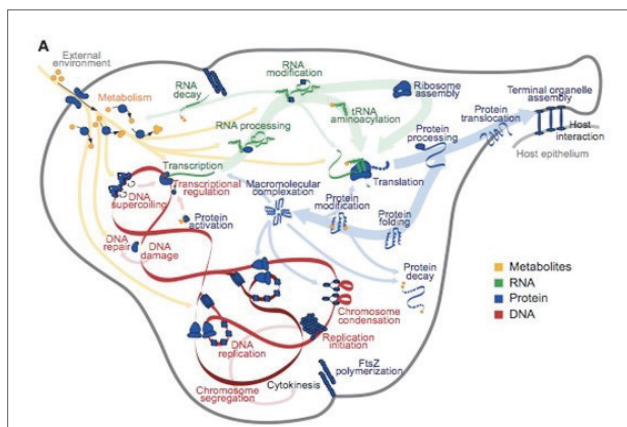


Fig. 6 - Descrizione schematica dei 28 sotto modelli, indicati con colori diversi, con cui è stato costruito il modello *Whole cell* [metabolico (arancio), RNA (green), proteinico (blu), DNA (rosso)] [rif. (10)]

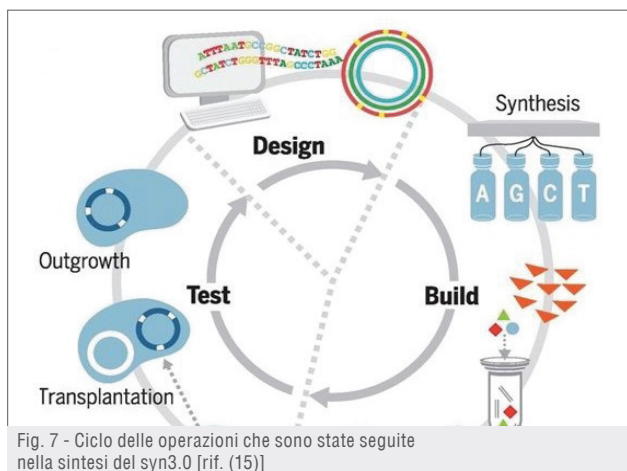


Fig. 7 - Ciclo delle operazioni che sono state seguite nella sintesi del syn3.0 [rif. (15)]

nelle molecole genetiche" [13]. Ovvero nel DNA, un codice, come visto, che può essere copiato e ricopiato più volte con errori che creano nuove informazioni.

Esiste però un approccio alternativo per affrontare il problema, che riflette l'atteggiamento di Richard Feynman, uno degli artefici dell'elettrodinamica quantistica, che lasciò scritto sulla sua lavagna: "What I cannot create I do not understand". In questa frase riecheggia l'insegnamento di Gianbattista Vico il quale sosteneva che per comprendere qualcosa bisogna averlo fatto. È stata ripresa e modificata da Craig Venter, uno dei pionieri della decifrazione del genoma umano, affermando: "What I cannot build, I cannot understand" [14]. Si tratta di un messaggio denso di significati che ci porta agli esordi di un'avventura, che emerge dalla biologia sintetica: la creazione di una cellula artificiale. In un primo tempo, ovvero

sino al 1984 si riteneva che la più semplice cellula in grado di avere una crescita autonoma fosse il batterio *Mycoplasma mycoides*, perché depositario del più piccolo genoma richiesto per comprendere i principi basilari della vita. Ovvero tale da sancire la soglia di complessità richiesta perché essa si inneschi. Venter ed i suoi collaboratori si sono impegnati in tali indagini con lo scopo di produrre una cellula sintetica minima più semplice di quelle naturali. In una recente pubblicazione viene annunciata la scoperta della sintesi di una catena di oligonucleotidi in grado di riprodursi dopo essere stati installati in un'opportuna membrana [15]. Il tutto seguendo il ciclo di operazioni illustrato nella Fig. 7. Battezzato "JCVI-syn 3.0" o in breve "syn3.0", contiene circa 531.000 coppie di basi organizzate in 473 geni trapiantati nelle cellule di un sottile batterio *Mycoplasma mycoides*.

Si tratta del più piccolo genoma sintetico sino ad ora ottenuto. In realtà questo successo ha evidenziato però che il concetto di minimo gene risulta più complesso di quanto fosse apparso nelle prime indagini poiché accanto ai geni essenziali devono esserne presenti altri, quasi essenziali perché richiesti per una crescita robusta resistente alle perturbazioni. Pertanto il syn3.0 costituisce un'approssimazione del lavoro di

esplorazione in corso. Tuttavia non si può fare a meno di prendere atto che la sua creazione, oltre che essere densa di contenuti e di significati scientifici e filosofici, ispirerà gli sforzi che verranno condotti nella biologia sintetica. Inoltre conferirà slancio alle attività in corso intese ad utilizzare le cellule batteriche quali fabbriche per la bioconversione di risorse rinnovabili a prodotti chimici di importanza applicativa. La produzione di nuovi cammini di reazione costituisce infatti il cuore della ingegneria metabolica intesa a creare funzioni biologiche non esistenti in natura combinando la biologia, la chimica, la teoria dell'informazione e l'ingegneria.

## BIBLIOGRAFIA

- [1] S. Lloyd, *Nature*, 2000, **406**, 1047.
- [2] S. Lloyd, *Programming the Universe*, Alfred A. Knopf, New York, 2007.
- [3] P.C.W. Davies, *Complexity*, 2004, **10**, 11.
- [4] F.M. Harold, *In Search of Cell History*, The University of Chicago Press, 2014.
- [5] H.F. Judson, *L'ottavo giorno della creazione*, Editori Riuniti, 1982.
- [6] G. Wall, M. Gong, *Exergy Int. J.*, 2001, **1**(3), 128.
- [7] C. Adami, C. Ofria, T. C. Collier, *PNAS*, 2000, **97**, 4463.
- [8] P. McKinley *et al.*, *Computer*, 2000, 54.
- [9] C. Tan *et al.*, *Molecular Bio Systems*, 2007, **3**, 343.
- [10] M.L. Simpson *et al.*, *Engineering in the Biological Substrate: Information Processing in Genic Circuits*, Proc. of the IEEE, 2004, **92**, 849.
- [11] J.R. Karr *et al.*, *Cell*, 2012, **150**, 389.
- [12] E. Schroedinger, *What is Life?*, Cambridge University Press, 1944.
- [13] A. Rutherford, *The Origin of Life*, Penguin, 2013.
- [14] J. Craig Venter, *Life at the speed of light*, Little Brown, 2013.
- [15] C.A. Hutchison *et al.*, *Science*, 2016,

### From Information to Complexity: A Metaphor For Chemical, Biological and Engineering Systems

The structure of DNA has evidenced that biology constitutes an aspect of the information science by opening interesting perspectives in the description of the characteristics and of the evolution of living systems, starting from the cells. The comparison of the obtained results with the experimental ones concerning the synthesis of cells with minimal genome (JCVI-syn3.0) foresees important developments in the ongoing researches on synthetic biology and on metabolic engineering