

LA SCIENZA CHE SALVERA'IL MONDO

Il 2020 non è stato solo l'anno della pandemia. La biologia molecolare (grazie ricercatori!), la sintesi più moderna del pensiero evoluzionistico e la teoria dell'informazione (grazie Google!) si sono incontrate per darci il vaccino anti Covid e la confer

Il Foglio Quotidiano · 25 gen 2021 · 1 · Di Enrico Bucci

Il 2020 sarà probabilmente ricordato come l'anno della pandemia da Sars-CoV-2. Il virus e la malattia hanno forgiato l'immaginario collettivo, modificato le nostre vite, seminato disastri e messo in dubbio quello che ragionevolmente potevamo immaginare del nostro futuro, sia a livello individuale che collettivo. Eppure, questo anno è stato notevole anche per gli avanzamenti scientifici realizzati. Guardando all'indietro, vorrei provare a trasmettere al lettore il senso di meraviglia per quello che i ricercatori hanno raggiunto nel mio specifico campo di interesse, quello della biologia molecolare. Io credo non sia fuori luogo affermare che nel 2020 abbiamo assistito a conferme spettacolari di una nuova visione nelle scienze della vita, un campo che forse potremmo chiamare "information biology", locuzione non perfettamente



sovrapponibile alla sua versione italiana "biologia dell'informazione". Ho parlato di conferme spettacolari, e non di emersione improvvisa, perché ritengo che si tratti del culmine di un processo che ha portato a più stretta convergenza discipline preesistenti e consolidate diverse, quali la biologia molecolare, la sintesi più moderna del pensiero evoluzionistico, la teoria dell'informazione. Questo processo, in corso da vari decenni, è stato ben identificato nei suoi obiettivi e nella sua ampiezza dai più acuti e interessanti studiosi di biologia evoluzionistica; credo che fra le parole più appropriate, in questo senso, si possano ricordare quelle scritte da John Maynard Smith nel 1995, le quali recitano pressappoco così:

“Un’idea centrale nella biologia contemporanea è quella dell’informazione. La biologia dello sviluppo può essere vista come lo studio di come le informazioni nel genoma vengono tradotte nella struttura adulta e la biologia evolutiva di come le informazioni sono arrivate lì in primo luogo” (Szathmáry & Maynard Smith, 1995).

Qui si evidenzia come un tema unificante per tutte le scienze della vita sia connesso alla trasmissione, alla variazione e all’utilizzo dell’informazione da parte dei sistemi biologici; un tipo di informazione, come giustamente scrive Maynard Smith, codificata nel genoma e in grado di originare per intero un organismo, ma anche di cambiare nel tempo, definendo l’evoluzione delle specie.

Questa visione, la visione cioè che assegna non al veicolo più o meno occasionale rappresentato dagli organismi viventi, ma al contenuto informativo un ruolo di primo piano nel plasmare tutto ciò che vive, sia per ciò che è che per ciò che nel tempo diventa, nel 2020 ha trovato alcune spettacolari conferme. I risultati e le scoperte ottenuti in questo anno hanno infatti contribuito a stabilire basi molto salde e ad allargare l’orizzonte e le applicazioni della “information biology”; e, per far capire immediatamente al lettore l’importanza della cosa, credo che non ci sia modo migliore che partire dal primo di questi risultati, e cioè la realizzazione e il successo sempre più probabile dei vaccini a Rna.

Vaccini a Rna e informazione molecolare

Facciamo un passo indietro. Nel 1976, in suo famosissimo saggio intitolato “Il gene egoista”, Richard Dawkins scrisse quanto segue: “La darwiniana ‘sopravvivenza del più adatto’ è in realtà un caso speciale di una legge più generale di sopravvivenza di ciò che è stabile. L’universo è popolato da cose stabili. Una cosa stabile è un insieme di atomi che è abbastanza permanente o comune da meritare un nome”. Qui l’autore intende dire che ciò che per Darwin è “più adatto” condivide una proprietà generale: durare di più nel tempo, cioè rimanere stabile. Andando avanti, egli identifica una classe speciale di “oggetti stabili”, che chiama “replicatori”: si tratta di oggetti che durante il periodo in cui esistono come individui sono in grado di creare copie di sé stessi – raggiungendo quindi una diversa sorta di stabilità, che potremmo definire “stabilità del tipo”, consistente nella conservazione nel tempo di quella speciale varietà di replicatore in modo più o meno invariato, senza che gli individui necessariamente sopravvivano. Se gli oggetti in questione sono organismi viventi, questa particolare stabilità è quella caratteristica per esempio delle specie biologiche; ma come Dawkins ha specificato a chiare lettere, le leggi di selezione darwiniana dei replicatori sono universali, e valgono per qualunque tipo di oggetto che, nel creare nuove copie di sé stesso, introduca per disegno o per caso un numero di differenze tali nei suoi discendenti che individui diversi abbiano probabilità fra loro diverse di lasciare a loro volta un determinato numero di discendenti in un intervallo di tempo prefissato.

Ma quali sono le caratteristiche minime che devono avere questi replicatori per esistere? Sostanzialmente ogni replicatore, come individuo, deve possedere, codificata in qualche

modo, l'informazione necessaria a creare una copia di sé stesso, che ridotta ai minimi termini significa l'informazione contenente la descrizione per creare e mantenere una macchina in grado di autocopiarci utilizzando materia ed energia a disposizione nell'ambiente e considerate le condizioni fisiche generali in cui esso è immerso.

E' il culmine di un processo che ha portato a più stretta convergenza discipline preesistenti e consolidate diverse

I "replicatori" di Dawkins, che durante il periodo in cui esistono come individui sono in grado di creare copie di sé stessi

Enrico Bucci, Ph.D. in Biochimica e Biologia molecolare (2001), è professore aggiunto alla Temple University di Filadelfia. Si occupa di dati biomedici, frodi scientifiche e biologia dei sistemi complessi. "Cattivi scienziati" è la sua rubrica quotidiana sul Foglio dall'inizio della pandemia.

Qualunque informazione ausiliaria posseduta, poi, che serva a mantenere più a lungo la stabilità del replicatore nel tempo oppure a permettergli di creare un numero di copie maggiori per unità di tempo, conferirà a quel replicatore la possibilità di lasciare più discendenti rispetto ad altri; ed è evidente che, in essenza, se le risorse di materia ed energia per generare nuove copie sono limitate, la selezione darwiniana agirà favorendo queste caratteristiche. L'informazione per replicatori più efficienti in termini di discendenza (date le condizioni ambientali), in altre parole, quando emerge porta alla scomparsa di quei replicatori codificati da informazione meno efficiente, a causa delle limitazioni imposte dall'ambiente al numero di copie di sé stesso che ciascun individuo può generare.

Se a questo punto volgiamo lo sguardo agli organismi viventi, scopriamo che essi possiedono tutte le caratteristiche dei replicatori darwiniani di Dawkins – e infatti, come tutti sappiamo, la selezione naturale di Darwin è un processo che ben spiega molti dei cambiamenti e della varietà delle forme biologiche. In particolare, tutti gli organismi viventi sono alla fine delle macchine che creano copie di sé stesse; l'informazione che specifica il progetto della "macchina copiatrice", ovvero dei replicatori darwiniani che chiamiamo organismi viventi, è come sappiamo specificata sotto forma di lunghissime molecole di Dna o di Rna.

Il modo in cui l'informazione è specificata in Dna e Rna è presto detto: queste molecole corrispondono a catene di componenti chimici di soli quattro tipi diversi, le cosiddette "basi", che sono un po' come quattro tipi diversi di pietre in una collana. La sequenza specifica di pietre preziose in una collana differenzia un gioiello dall'altro, e così la sequenza specifica di basi differenzia una molecola di Dna o Rna da un'altra; ogni organismo possiede una o più "collane" di Dna o Rna, che chiamiamo cromosomi, e l'insieme di tutte queste collane corrisponde al patrimonio genetico o genoma di quel dato organismo vivente. In un modo che vedremo più avanti, ciascuna diversa sequenza di Dna o Rna, e a maggior ragione ciascun diverso genoma, conferisce proprietà diverse al "replicatore" biologico che la porta; proprietà che in ultima analisi sono in grado di influenzare, a volte in modo molto sottile, il numero di discendenti che quel dato replicatore lascerà, sotto la pressione combinata

dell'ambiente e del caso.

Da miliardi di anni, innumerevoli codici genetici fatti di Rna e Dna si trasmettono attraverso il sistema della replicazione biologica, competono per le limitate risorse disponibili e mutano continuamente per abbondanza relativa, perché casualmente emergono delle varianti più adatte a lasciare discendenti nel momento e nell'ambiente specifico in cui si trovano, oppure per semplici effetti neutrali – eventi cioè che portano alla ribalta un certo genoma perché improvvisamente esposto a condizioni ambientali molto favorevoli, oppure che ne sopprimono uno più promettente attraverso catastrofi di ogni genere, oppure ancora che lo proteggono dalla competizione con altre varianti attraverso l'isolamento geografico o temporale.

I “geni egoisti” di Dawkins compaiono per caso, fioriscono per caso o per selezione darwiniana e per caso o selezione darwiniana scompaiono dalla faccia della terra; e, di fatto, noi siamo immersi in questo pulsante e caotico proliferare di informazione genetica, che attraverso la modifica di tutti gli organismi intorno a noi, può esserci indifferente, può costituire una risorsa o può, come nel caso di Sars-CoV-2, costituire un pericolo serio alla nostra stessa integrità vitale.

Sapendo tutto questo – acquisito in oltre cinquanta anni di esperimenti e di teoria – è spontaneo fare una considerazione: se è tutto vero, o per meglio dire nella misura in cui tutto questo è vero, se noi davvero funzioniamo così, allora fornendo la giusta informazione – ovvero l'informazione in grado di conferire una certa proprietà a un organismo, sotto forma di codice genetico – dovremmo riuscire a ottenere da qualsiasi essere vivente, e quindi anche dal nostro stesso corpo, una risposta predeterminata.

Per esempio, potremmo fornire al nostro organismo le istruzioni necessarie per combattere un virus come Sars-CoV-2, se sapessimo quali sono sia in termini di “parole” da utilizzare, sia in termini di “processi” da attivare. Nello specifico, si potrebbero fornire ai nostri corpi le istruzioni per fare una proteina virale, insegnando così al nostro sistema immune a riconoscere un pezzo essenziale del virus Sars-CoV-2 qualora si sia attaccati, accompagnate dalle istruzioni necessarie a specificarne la quantità molto elevata desiderata e dalle istruzioni che specificano dove vogliamo che questa proteina sia “sistemata” non appena fabbricata, in modo che si arrivi alla fine a una robusta risposta immune. Il tutto con un “messaggio genetico” che, una volta adempiuto il proprio mestiere, si distrugga da solo, per evitare di lasciare acceso un processo non più necessario (e che consumerebbe inutilmente risorse delle nostre cellule) quando abbia raggiunto lo scopo di generare la memoria immune che serve.

Bene: sapendo che l'istruzione per fare la proteina virale era contenuta nel patrimonio genetico del virus Sars-CoV-2, in meno di un mese dall'isolamento i ricercatori hanno “letto” grazie al processo di sequenziamento genetico l'informazione necessaria a codificare la proteina per “aizzare” il nostro sistema immune. Dalle conoscenze di biologia cellulare e molecolare acquisite nei decenni appena trascorsi, sapevano quale altro pezzetto di Rna aggiungere a quello della proteina Spike per fare in modo che il nostro organismo la produ-

cesse in quantità elevata e come modificare la stessa sequenza di Rna, rispetto a quella naturale del virus, sempre allo stesso scopo; e altri pezzetti ancora di Rna sono stati aggiunti per specificare dove dovesse poi essere posizionata la proteina, per ottenere il miglior effetto immunizzante possibile. Gli scienziati sapevano inoltre che avrebbero ottenuto quell'effetto, grazie a decenni di studio di immunologia molecolare e cellulare, che ci avevano insegnato come quella proteina sarebbe stata riconosciuta da cellule specializzate del sistema immunitario umano, e a cosa avrebbe portato questo riconoscimento. Ancora, conoscendo il modo in cui i messaggi di Rna (o Rna messenger) sono naturalmente letti e utilizzati dalle nostre cellule, i ricercatori sapevano cosa ci si poteva aspettare in termini di eliminazione del messaggio, una volta che esso avesse adempiuto il suo compito. Infine, ancora altri ricercatori, fondandosi sulle cono

Siamo immersi in questo pulsante e caotico proliferare di informazione genetica, che attraverso la modifica di tutti gli organismi intorno a noi, può esserci indifferente, può costituire una risorsa o può, come nel caso di Sars-CoV-2, costituire un pericolo serio alla nostra stessa integrità vitale

Nel vasto mondo di informazione genetica che ci avvolge, o ci sostiene o ci aggredisce o semplicemente ci ignora, abbiamo immesso in quantità crescente un pezzetto di informazione in grado di modificare la traiettoria evolutiva di almeno due specie di replicatori diversi – noi e il virus Sars-CoV-2

scienze di chimica farmaceutica, sapevano come impacchettare in una piccola gocciolina di grasso, sale e zuccheri il messaggio di Rna, così che potesse arrivare a destinazione ed essere utilizzato come voluto.

Tutto questo è stato realizzato, testato in decine di migliaia – ormai centinaia di migliaia – di esseri umani, e sembra proprio che funzioni: abbiamo i nostri “vaccini” a Rna, ma sarebbe meglio dire che abbiamo pronta in milioni di dosi l'informazione giusta perché il nostro corpo si vaccini da solo.

L'informazione genetica che abbiamo confezionato nelle fiale pronte per l'iniezione ha una caratteristica fondamentale che la distingue da quella che muove i replicatori darwiniani di Dawkins, contro cui essa è diretta: è frutto non del caso e della selezione successiva, ma del disegno intelligente da parte di artefici umani. Nello specifico, non abbiamo faticosamente isolato ciò che ci serviva, provando e riprovando varianti diverse, ma lo abbiamo creato sulla base della nostra conoscenza; e nel vasto mondo di informazione genetica che ci avvolge, o ci sostiene o ci aggredisce o semplicemente ci ignora, abbiamo immesso in quantità crescente un pezzetto di informazione in grado di modificare la traiettoria evolutiva di almeno due specie di replicatori diversi – noi e il virus Sars-CoV-2.

Abbiamo dimostrato come sia possibile tradurre la nostra conoscenza in codice genetico, utilizzandolo per interagire con altri codici genetici – virali, cancerogeni o di qualsiasi altra natura – che ci minacciano, facendolo tuttavia in maniera temporanea e reversibile grazie all'uso di Rna, invece che modificando il Dna

Leggere l'informazione genetica che ci serve per acquisire conoscenza, e poi codificare in

informazione genetica altre conoscenze utili a competere contro l'informazione genetica di uno o più organismi: come specie, abbiamo portato direttamente informazione frutto di cognizione (potremmo dire memi) a competere con informazione frutto di mutazione casuale e selezione (i replicatori darwiniani biologici che chiamiamo virus), per di più utilizzando informazione genetica "biodegradabile" che non lascia tracce dopo aver spostato l'ago della bilancia nella competizione tra noi e il virus.

Questo abbiamo fatto; e, a vedere ciò che è in corso di sperimentazione clinica anche avanzata, possiamo aspettarci che ove il successo sia confermato nei prossimi mesi, si cominceranno a usare altri pezzi di conoscenza umana, codificati sotto forma di Rna, per modificare per esempio il decorso di qualche tipo di cancro, o per riaggiustare un cuore malato, o anche semplicemente per produrre farmaci o altri materiali in appositi bioreattori.

Come aveva preconizzato Dawkins nel suo libro visionario, è l'informazione che conta e il modo in cui essa è codificata; e nel caso della vita, davvero il grosso di questa informazione è racchiusa nel genoma.

Ma vi è di più.

L'informazione codificata nel genoma è come abbiamo visto quello che consente a un organismo, come il virus Sars-CoV-2, di sopravvivere e moltiplicarsi. Abbiamo anche detto che tanto più quella informazione è in grado di conferire un vantaggio per la sua stessa replicazione all'individuo che la porta, tanto più si diffonderà in presenza di risorse limitate; ma quali sono le condizioni in cui tale vantaggio si realizza? In sostanza, quando quell'informazione è ben adattata all'ambiente in cui il replicatore è immerso; cioè quando contiene elementi che permettono di reagire nella maniera migliore possibile alle variazioni ambientali, preservando l'integrità dell'organismo (o del replicatore). Possedere questo tipo di informazione significa, in ultima analisi, possedere elementi predittivi circa gli stati che è possibile incontrare nell'ambiente e circa la loro evoluzione, codificati direttamente nel genoma. In un certo senso, le informazioni sulla nicchia ambientale in cui un organismo passa la sua esistenza sono memorizzate nei geni e rappresentano previsioni sui possibili stati futuri dell'ambiente. Poiché il successo riproduttivo di un organismo è più elevato in ambienti ben predetti (semplicemente perché è più facile sfruttare le caratteristiche dell'ambiente per la riproduzione se sono prevedibili), gli organismi con più informazioni circa la loro nicchia ambientale dovrebbero prevalere su quelli con meno informazioni, suggerendo una relazione diretta tra contenuto informativo del genoma e idoneità al proprio ambiente che favorisce la codifica di specifiche informazioni su un ambiente all'interno del genoma di un organismo. Nel genoma di Sars-CoV-2, ad esempio, troviamo codificate le caratteristiche dell'ambiente in cui il virus si replica sotto forma di mutazioni in risposta all'attacco del nostro sistema immunitario; oppure di mutazioni che rendono meno efficace il distanziamento sociale, aumentando la trasmissibilità; e così via.

Informazioni di questo tipo sono direttamente accessibili tramite il sequenziamento genetico (che ne rivela la presenza) e i test del significato funzionale delle mutazioni (che ne interpretano il significato); e una volta comprese, le informazioni contenute nel genoma del-

le varianti che via via emergono sono sia utili per cambiare il messaggio che diamo al nostro corpo sotto forma di vaccini a Rna, sia per modificare il messaggio che diamo alle nostre popolazioni circa i comportamenti da adottare per evitare il peggio. Ecco perché è importantissimo continuare il processo di “lettura” mediante sequenziamento dei genomi virali durante una pandemia; ed ecco come, dallo studio dei genomi di ogni specie di organismo, continuiamo a imparare qualcosa circa la storia delle specie viventi – compresa la nostra – e dell’ambiente in cui vivono o vivevano.

Ma perché e in che modo una semplice “collana di basi di Dna” racchiude nella sua sequenza tutta l’informazione che serve? Per capirlo, lasciate che io introduca il secondo dei risultati scientifici che considero della massima importanza, raggiunto nel 2020.

L’intelligenza artificiale al lavoro

Se è vero, come è vero, che la sequenza del genoma (di Dna o Rna) di ogni organismo è ciò che racchiude tutte le informazioni necessarie a quel particolare replicatore biologico per funzionare e creare copie di sé stesso, vuol dire che ci deve essere un meccanismo attraverso cui quella informazione “legge sé stessa” e dà origine all’organismo che la racchiude. Fino al 2020, questo meccanismo, come ho appreso già ai tempi dei miei studi universitari, si basava sostanzialmente sui seguenti fatti, connessi da un’ipotesi importante non ancora dimostrata.

La sequenza di basi nel genoma di un organismo corrisponde all’insieme di tutte le sequenze di proteine che compongono quell’organismo. Le proteine sono altre “collanine di parti chimiche”, le cui pietre preziose componenti non sono le quattro basi di Dna e Rna, ma 21 tipi diversi di aminoacidi. Il codice genetico di Dna e Rna viene tradotto in codice proteico con delle apposite nanomacchine biologiche chiamate ribosomi, le quali “sgranano” la collana di basi, e a ogni tripletta di basi aggiungono uno e un solo specifico aminoacido a una “collana” di aminoacidi che è una proteina. In questo modo, ogni volta che viene “sgranata” da un ribosoma la stessa sequenza di basi componenti il genoma, si ottiene la stessa proteina, in un processo che letteralmente “traduce” la catena chimica di basi che forma il genoma in una catena chimica di aminoacidi che forma una proteina. Il genoma tipico degli organismi viventi contiene la sequenza di moltissime proteine, un po’ come un intero libro in inglese (il genoma, “scritto” in sequenze di basi) è l’insieme di tanti paragrafi diversi, ciascuno dei quali può essere tradotto indipendentemente in italiano (le proteine, “scritte” in sequenze di aminoacidi).

Le proteine sono le macchine che muovono il nostro corpo, i sensori dei nostri occhi, i trasportatori dell’ossigeno, gli enzimi del nostro sistema digerente, e in breve il grosso di tutte le componenti di cui siamo fatti; come abbiamo visto, il genoma ne codifica la sequenza, e il processo di traduzione da una sequenza di basi a una di aminoacidi è quello che serve a generare una proteina. Siccome cambi della sequenza di una certa porzione del genoma di Dna sono in grado di causare cambi nella sequenza della proteina corrispondente, e visto che quando ciò avviene la proteina in questione può perdere la sua funzione specifica nel nostro organismo o (in rarissimi casi) acquisire qualche nuova funzione (come avviene per

certe mutazioni che inducono il cancro), ben presto ci si è accorti che la funzione delle proteine deve dipendere dalla loro sequenza. Ma in che modo? Qui arriviamo all'ipotesi dimostrata, che funzionava così bene, e che ha costituito il fulcro di ciò che è stato insegnato a generazioni di studenti di biologia molecolare.

L'idea è questa: siccome i 21 aminoacidi, le "pietre preziose" che vengono usate per costruire la "collana" corrispondente alla sequenza di ogni proteina, hanno forma tridimensionale e proprietà chimiche diverse che conferiscono a ciascun aminoacido la capacità di attirare o respingere altri aminoacidi nella stessa sequenza proteica, in realtà le proteine non assomigliano a una collana distesa, ma piuttosto a un gomito molto stretto, che assume quindi una forma tridimensionale specifica per ogni proteina. La sequenza delle proteine, cioè, si ripiega su sé stessa, in un modo che è specifico e determinato per ogni particolare proteina che ha quella data sequenza di aminoacidi; questo avviene perché le "perline" aminoacidiche che compongono la "collana" proteica si attraggono, si respingono, si incastrano o si respellono fra loro in maniera che dipende dalle specifiche proprietà chimiche di ciascuna di esse. Ogni specifica perline è in grado di attirare e legare strettamente o respingere altre perline di un tipo preciso, e solo di quello: questo provoca ripiegamenti specifici a seconda di come le perline sono disposte.

Dunque l'ipotesi di fondo è che la sequenza in aminoacidi di una proteina ne determini la forma tridimensionale attraverso il ripiegamento spontaneo della catena di aminoacidi che la costituisce; e siccome la sequenza di una proteina è ottenuta in ogni essere biologico dalla traduzione di una particolare parte della sequenza di Dna o Rna che costituisce il genoma dell'organismo in questione, ecco che l'informazione contenuta nel genoma influenza come siano fatte dal punto di vista della loro forma tridimensionale tutte le proteine che compongono un dato organismo.

La forma tridimensionale assunta da ciascuna proteina, a sua volta, è quella che ne determina la funzione: lo sappiamo da tempo, sia perché abbiamo osservato come le proteine si organizzano tra loro in macchine molecolari ben funzionanti, di cui ciascuna proteina costituisce un pezzo specifico (proprio come in un'automobile ogni componente ha una forma e una funzione precise), sia perché sappiamo che se cambiamo, attraverso una mutazione della sua sequenza aminoacidica, la forma di una proteina, essa tende a perdere la sua funzione.

Dunque, riassumiamo: fino al 2020, a meno di alcune complicazioni specifiche che non discuterò qui, sapevamo che una sequenza di basi di Dna o Rna contiene l'informazione per ogni proteina che ci compone; che tale informazione è tradotta in una sequenza di aminoacidi, che costituisce una proteina specifica; che il ripiegamento in tre dimensioni di questa catena aminoacidica conferisce alla proteina la sua funzione specifica. Ipotizzavamo, inoltre, che la forma assunta dipendesse proprio dalla sequenza specifica, e quindi in ultima analisi dalla informazione genetica contenuta nel genoma di un organismo, corrispondente a quella proteina.

Ma la prova? Qual era la prova che l'informazione genetica, quella che abbiamo discusso a

proposito del mondo dei replicatori di Dawkins, davvero fosse in grado di conferire la forma e dunque la funzione alle proteine?

Ottenere questa prova significa una cosa precisa: significa saper prevedere come si ripiegherà tridimensionalmente una specifica “collana” di aminoacidi, ovvero quale struttura assumerà una proteina che possiede una certa sequenza, ovvero a quale forma proteica tridimensionale corrisponde una certa sequenza di basi in un pezzo di genoma di un qualunque organismo vivente.

Fino al 2020, noi non siamo stati in grado di riuscire a fare questa previsione. Non si riusciva cioè a trovare un modo per cui, avendo come unico dato la sequenza di una proteina, si ottenesse una previsione accurata della sua struttura tridimensionale.

Senza questa prova, non potevamo essere certi che il mondo dell’informazione genetica e dei replicatori in guerra fra loro descritto da Dawkins, fosse davvero il mondo in cui emergono le varianti e le mutazioni sulle quali agisce poi la selezione naturale: in teoria, si sarebbe potuto verificare che le mutazioni del Dna di un organismo non influenzassero poi così tanto il suo successo riproduttivo, quanto invece fenomeni sconosciuti o poco caratterizzati che avvenivano a livello delle sue proteine – cioè a livello delle sue parti costituenti, in grado di influenzarne il funzionamento (e quindi anche il successo riproduttivo) – i quali, indipendentemente da ciò che avveniva a livello di genoma, avrebbero potuto pesantemente determinare il successo del “replicatore”.

Invece, nel 2020, si è avuta la svolta, e uno dei pezzi più importanti della nostra teoria su come funziona la vita è andato a posto. Un programma di intelligenza artificiale, creato da Google e chiamato

“alphafold”, ha fatto ciò che le reti neurali sanno fare meglio: alimentato con un grande database di sequenze di proteine la cui struttura tridimensionale è nota, è riuscito a individuare delle caratteristiche – dei “pattern” – nell’informazione fornita correlati alla forma posseduta da ciascuna sequenza. Questi pattern, a noi sconosciuti, consistono certamente nella disposizione di specifici aminoacidi in “nuclei locali” in punti specifici della sequenza che influenzano l’inizio del ripiegamento della “collana” di aminoacidi costituenti una proteina; a sua volta, la vicinanza di un “nucleo locale” con tipi diversi di altri nuclei, cioè con altri “pattern” di sequenza identificati dalla rete neurale di Google ripiegati in forme specifiche all’inizio del processo, porta in maniera prevedibile a ulteriori ripiegamenti tridimensionali e alla formazione del “gomitolo” che ha quella forma specifica per quella sequenza specifica.

Si è trattato di un successo senza precedenti, che ha l’unica pecca di non fornire una spiegazione basata sulla chimica e sulla fisica del perché una certa sequenza si ripieghi in un certo modo; ma che tuttavia dimostra in modo inequivocabile che, data una certa sequenza, si ottiene un certo specifico ripiegamento della catena proteica nello spazio, un po’ come una sorta di origami che, data una certa sequenza ripiegamenti, porta a una certa forma tridimensionale.

Decenni di teorizzazioni e di sforzi, in un colpo solo, sono andati a posto: davvero la se-

quenza nel nostro genoma determina la forma delle nostre componenti elementari, le proteine, e attraverso queste, il funzionamento del nostro intero organismo.

Ecco perché l'istruzione che abbiamo fornito alle nostre cellule attraverso i vaccini a Rna per ottenere una reazione immune utile contro il virus SarsCoV-2 funziona: perché quella informazione corrisponde a una data sequenza proteica, che adotta la stessa forma tridimensionale della proteina virale; e quella forma è ciò che il sistema immunitario impara a riconoscere "tastandola" fino a selezionare quegli anticorpi che a essa meglio si adattano e che sono quindi meglio in grado di ricoprirli neutralizzandone la funzione indispensabile per il virus.

Il 2020, anno della "information biology"

Possiamo tirare a questo punto le fila del mio ragionamento. Nell'anno appena passato, abbiamo contrastato un replicatore darwiniano quasi perfetto – Sars-CoV-2, il vettore di un pezzo di informazione genetica prodotto dalla selezione naturale – utilizzando una informazione genetica creata mediante "disegno intelligente" per mitigare i danni a nostro carico. La competizione è appena iniziata, perché sappiamo che quel replicatore, obbedendo alle leggi darwiniane, sta mutando proprio per evadere la nostra risposta immune; ma la procedura di lettura dell'informazione genetica virale, accoppiata alla rapida "scrittura" in milioni di copie di controinformazione utile da mettere sotto forma di mRNA nelle fiale dei vaccini Pfizer e Moderna, ci conferisce la speranza di poter reagire per la prima volta quasi alla stessa velocità per rincorrere un processo di selezione naturale a noi sfavorevole.

Il tutto funziona perché, come è stato definitivamente dimostrato nel 2020 grazie all'intelligenza artificiale di un programma creato da Google, ciò che serve per determinare la funzione delle proteine – cioè degli "ingranaggi" molecolari che in massima parte compongono gli organismi viventi – è la loro sequenza; dunque la lettura della sequenza dei virus mutanti è tutto ciò che serve per creare messaggi in grado di stimolare il nostro stesso corpo a riprodurre per il nostro sistema immunitario la forma tridimensionale delle proteine virali mutate, così che possa essere addestrato a riconoscere anche quelle.

Alzando il livello del ragionamento, possiamo dire che abbiamo definitivamente dimostrato come sia possibile per noi tradurre le nostre idee e la nostra conoscenza in codice genetico, utilizzando direttamente questo per interagire con altri codici genetici – virali, cancerogeni o di qualsiasi altra natura – che ci minacciano, facendolo tuttavia in maniera temporanea e reversibile grazie all'uso di Rna, invece che modificando il Dna. Potremmo dire che per la prima volta abbiamo trovato il modo di scrivere memi non permanenti in codice genetico, dimostrando davvero, come aveva predetto Dawkins nel suo "il gene egoista", che ciò che conta nel mondo dei replicatori darwiniani (cioè della vita) al fine del successo almeno temporaneo nella competizione fra le specie sono la pura informazione e il suo utilizzo, non la sua origine; e se l'origine è il disegno dell'essere umano, invece che la selezione operata su varianti casuali, per la prima volta entra nel piano della competizione fra replicatori un nuovo tipo di informazione, dotata di scopo, e non prodotta dal caso.

Come specie, abbiamo portato direttamente informazione frutto di cognizione (potremmo

dire memi) a competere con informazione frutto di mutazione casuale e selezione (i replicatori darwiniani biologici che chiamiamo virus), per di più utilizzando informazione genetica che poi non lascia tracce

Con un programma di intelligenza artificiale creato da Google, decenni di teorizzazioni e di sforzi, in un colpo solo, sono andati a posto: davvero la sequenza nel nostro genoma determina la forma delle nostre componenti elementari, le proteine, e attraverso queste, il funzionamento del nostro intero organismo